JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 14 (2026) ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Viewpoints

Systematic and Collaborative Approach to Learning and Educational Content Development (SCALED) for Health Apps: An Experience-Informed Conceptual Framework (e83442)
Qi Ong, Anita Pienkowska, Emina Obarcanin, Anne-Claire Stona, Andy Khong, Josip Car, Andy Ho. 23

Intensive, Repeated Self-Report Measures: Should We Be Concerned About Changes in Data Quality Over Time? (e68735)
Arthur Stone, Stefan Schneider, Meynard Toledo, Raymond Hernandez. 35

A Beginner's Guide to Applying Large Language Models in Behavioral Interventions (e79302)
Nirali Shah, Lorraine Buis, Derek Papierski, Alexis Castellanos, Marvin Mlakha, Susan Murphy. 44

Reviews

Applications and Current Landscape of Digital Health Technologies in the Continuity of Care for Liver Transplant Candidates and Recipients: Scoping Review (e106470)
Wanqi Shi, Jiao Sun, Jiahao Li, Yuanyuan Li, Feiyan Yuan, Ruiya Li, Zijing Chu. 54

Reporting Gaps in mHealth Intervention Studies for Adults With Diabetes: Systematic Review (e95863)
Xin Yue, Jiacheng Shen, Junke Peng, Xiaoting Huang, Mingyue Hu, Xing Hu. 76

Stakeholder Perspectives on Digital Health Technologies in Poststroke Self-Management and Rehabilitation: Systematic Review Based on the NASSS Framework (e95197)
Jing Zhao, Jie Zhao, Xiaohan Liu, Yongxin Cao, Xuan Wang, Yanhong Dong, Hong Chang. 96

The Representation of Different Populations in Studies Assessing the Validity of Consumer Wearable Photoplethysmography-Based Measurements: Scoping Review (e87324)
Rebecca Schipper, Fatime Oumar Djibrillah, Meyke Roosink, Laura Winkens, Arlene John, Eric Hazebroek, Annemieke Witteveen, Agnes Berendsen. 112

Effects of Self-Guided Digital Mindfulness Interventions on the Mental Health of Health Care Workers: Systematic Review and Meta-Analysis of Randomized Controlled Trials (e74272)
Natsu Sasaki, Akiomi Inoue, Mako Iida, Kazuhiro Watanabe, Yuka Kobayashi, Asuka Sakuraya, Hisashi Eguchi, Kanami Tsuno, Yu Komase, Yasumasa Otsuka, Mai Iwanaga, Reiko Inoue, Kazuto Kuribayashi, Ayako Hino, Hiroki Asaoka, Satoru Kanamori, Yuta Inagawa, Takenori Yamauchi, Hiroki Ikeda, Go Muto, Rika Kato, Takuro Okuyama, Naomichi Tani, Noriko Kojimahara, Takeshi Ebara, Akihito Shimazu, Akizumi Tsutsumi, Norito Kawakami, Kotaro Imamura. 137

XSL•FO
RenderX

Just-in-Time Adaptive Interventions to Reduce Sedentary Behavior and Promote Physical Activity in Patients With Chronic Disease: Scoping Review (e81378)
Junchen Guo, Furong Chen, Chan Chen, Yunyun Dai, Xianghua Xu, Yongyi Chen, Mengying Huang. 162

Patient-Reported and Technology-Assisted Monitoring in Orthotic Management of Adolescents With Idiopathic Scoliosis: Scoping Review (e88880)
Anna Mathew, Kamila Sykorova, Nenad Pavel, Marianne Johnsen, Minna Pikkarainen, Sadeeq Ali, Mette Kvammen, Parisa Gazerani. 188

Barriers and Enablers to Integrating Patient-Generated Health Data in Shared Decision-Making From Health Care Professional and Patient Perspectives: Scoping Review (e85197)
Pavithren V S Pakianathan, Devender Kumar, Jayathissa Prabath, Rada Hussein, Josef Niebauer, Albrecht Schmidt, Jan Smeddinck. 230

Patient-Facing AI-Enabled Digital Health Technologies and Quality of Life in Cancer: Systematic Review and Exploratory Meta-Analysis (e94793)
Anica Ilic, Lene Juvet, Patrick Cairns, Kristen Thornton, Hanne Lie. 253

Conceptualization From the Sensors to Suicide-Related Outcomes: Scoping Review Based on Layered Hierarchical Sensemaking Framework (e98616)
Sohee Kim, Jinyeong Kim, Wai Chien, Tzu Luk, Yu Zhang, Zhen Tan, Eunju Park, Heejung Kim. 268

Mobile Health in Home-Based Palliative Care: Realist Review of Contextual Factors and Mechanisms Influencing Outcomes (e80321)
Nuhamin Gebre, Oladayo Afolabi, Nahla Gafer, Kennedy Nkhoma, Nicola Ayers, Charlotte Hanlon, Richard Harding. 291

Engagement of Users in Digital Health Applications: Scoping Review (e66002)
Melek Aktas, Linda Cambon, Olivier Aromatario. 327

Heart Rate Monitors for the Estimation of Physical Activity in Patients With Cardiovascular Disease: Systematic Review (e79995)
Paulien Vermunicht, Christophe Buyck, Sebastiaan Naessens, Wendy Hens, Emeline Van Craenenbroeck, Kris Laukens, Lien Desteghe, Hein Heidbuchel. 340

mHealth Technologies for the Care of Children With Congenital Heart Disease: Scoping Review (e83385)
Mingzhu Wang, Qiuyin Pan, Shengrong Tan, Yuanchun Kong, Zhengling Dai, Jiao Cai, Duoqin Bi, Jiali Zhou. 384

Effects of Remote, Virtual, or Hybrid Cardiac Rehabilitation Supported by mHealth in Patients With Heart Failure: Systematic Review and Meta-Analysis (e90422)
Kaidong Shao, Chunqiu Liu, Tianshu Li, Huiyan Qu, Hua Zhou. 395

Mechanisms of Engagement With Mobile Health Apps for Adults With Long-Term Conditions: Overview of Systematic Reviews (e88382)
Jeni Baykoca, Goretti Hurtado Barbeito, Christina Pearce, Madison Milne-Ives, Joanna Hudson, Sam Norton, Rona Moss-Morris. 411

Perceptions of Adults Aged 50 Years and Older Regarding the Use of Wearable mHealth Technologies to Promote Physical Activity: Systematic Review and Meta-Ethnography (e67157)
Beatriz Rodríguez-Martín, Diego González-Arroyo, Susana Priego-Jiménez, María Lopez-González, María Medrano-Echeverría, Marta Ruiz-Grao, Celia Álvarez-Bueno. 430

Patient Acceptance and Barriers to IoT Usage in Health Care: Systematic Literature Review (e81260)
Zaenal Arifin, Putu Handayani. 469

Machine Learning Frameworks for Wearable-Based Stress Modeling in Naturalistic Settings: Scoping Review (e76632)
Shifali Sharma, Aswin Janakiraman, Lujie Chen. 494

XSL•FO
RenderX

Wearable Devices for Monitoring and Management of Comorbid Obstructive Sleep Apnea and Hypertension: Scoping Review (e84506)
Hongyi Wang, Anran Cheng, Yusheng Li, Haining Zhang, Shun Fan, Rong Shen, Wei Zhang, Huanan Li, Jingui Wang. 521

Effectiveness of Mobile Health Interventions in Pediatric Cancer: Systematic Review and Meta-Analysis of Randomized Controlled Trials (e86836)
Huilu Yao, Yiting Wen, Hongxiu Wang, Ying Xiao, Meiling Deng, Wei Yang, Yuqin Li, Xiaorong Mao. 538

Effectiveness of mHealth-Based Nutritional Interventions on Iron Status of Pregnant Women: Systematic Review of Randomized Controlled Trials (e81001)
Saba Belay, Afework Bezabih, Wim Petegem, Christophe Matthys. 561

mHealth-Based Gamification Interventions to Promote Health Among Older Adults: Scoping Review (e82368)
Xi Chen, Yishu Zhu, Dian Jiang, Jundan Huang, Qi Xie, Yishi Chen, Jing Chang, Weiping Huang, Hongting Ning, Hui Feng. 582

Effectiveness of Text Messaging Interventions on BMI Among Adults With Prediabetes: Systematic Review and Meta-Analysis (e78521)
Arkers Wong, Shun Choi, Tiffany Cheng, Chelsea Cheng, Ivy Kwong, Cheuk Fong, Chi Wong, Ka Chan. 597

Evaluating Effectiveness of mHealth Interventions for Rehabilitation in Patients With Head and Neck Cancer: Systematic Review and Meta-Analysis of Randomized Controlled Trials (e78109)
Yang Xiao, Qiuyu Zhou, Feiruo Hong, Yuwei Lin, Enhong Li, Xuefen Yu. 609

Wearable Devices for Remote Monitoring of Chronic Diseases: Systematic Review (e74071)
Masresha Tegegne, Sharareh Niakan Kalhori, Paulo Haas, Viktor Sobotta, Joana Warnecke, Thomas Deserno. 630

Video-Based Motion Capture Smartphone Apps for Testing Human Motor Performance Skills: Scoping Review (e65474)
Clara Zoeller, Claudia Niessner, Manuel Fleps, Thorsten Klein, Anke Hanssen-Doose, Alexander Burchartz, Alexander Woll, Thorsten Stein. .
6 5 2

Digital Health Interventions in Children and Adolescents With Type 1 Diabetes Mellitus and Their Impact on Clinical and Behavioral Outcomes: Scoping Review (e79338)
Eirena Beh, Jiaying Lin, Geraldyn Leong, Pei Loh, Ee Poh, Daniel Chan. 666

Examining the Use of Consumer Wearable Devices and Digital Tools for Stress Measurement in College Students: Scoping Review of Methods (e64144)
Aarti Sathyanarayana, Ohida Amin, Jennie An, Jukka Onnela. 682

Mental Health Apps Implemented in the Workplace: Scoping Review of Trends and Gaps in Evaluation Research (e57046)
Sheila Addanki, Luciana Macedo, Joy MacDermid, Sandra Moll. 728

Exploiting Unsupervised Free-Living Data for Cardiorespiratory Fitness Estimation: Systematic Review and Meta-Analysis (e69996)
Alexios Dosis, Aron Syversen, Mikolaj Kowal, Daniel Grant, Jim Tiernan, David Wong, David Jayne. 750

Digital Health Interventions to Support Chronic Disease Management: Systematic Scoping Review (e63742)
Abdullah Al Mahmud, Shane Joachim, Prem Jayaraman, Caitlin Learmonth, Shivani Tyagi, Abdur Forkan, Muhammad Shuakat, Nilmini Wickramasinghe, Jack Wheeler, Stephanie Best, Alison Trainer. 774

Mobile Apps to Improve Health Parameters in Healthy Adults: Systematic Review (e66881)
Gaia Leuzzi, Mirko Job, Carola Cosentino, Riccardo Testa, Aldo Scafoglieri, Marco Testa. 800

XSL•FO
RenderX

Advancements in Wearable Sensor Technologies for Health Monitoring in Terms of Clinical Applications, Rehabilitation, and Disease Risk Assessment: Systematic Review (e76084)
Bonsang Gu, Hyeon Kim, HyunBin Kim, Jun-Il Yoo. . . . 815

Evaluating the Effectiveness of Digital Interventions for Stress Management in Pregnant Women: Systematic Review and Meta-Analysis (e66267)
Jeung-Im Kim, Joo Lee, So-Hee Park. . . . 830

Calorie-Counting Apps for Monitoring and Managing Calorie Intake in Adults Living With Weight-Related Chronic Diseases: Decade-Long Scoping Review (2013-2024) (e64139)
Kaylee Dugas, Marie-Andrée Giroux, Abdelatif Guerroudj, Jazna Leger, Asal Rouhafzay, Ghazal Rouhafzay, Jalila Jbilou. . . . 852

Key Features of Engagement Strategies in Nutrition Apps for Adults: Scoping Review (e82276)
Maria Vasiloglou, Zoë van der Heijden, Eric Scuccimarra, Alberto Conde Freniche, Desiree Lucassen, Nienke de Vlieger, Frédéric Ronga, Elske Brouwer-Brolsma, Tamara Bucher. . . . 880

Cost-Effectiveness of Virtual Emergency Care Models: Systematic Review (e82143)
Ravi Shankar, Linda Wang, Soon Ho, Mei Liew, Satya Kumar Gollamudi, Tze Wong, Serene Wong. . . . 892

Application of Mobile Health Technologies to Prevent Unintentional Childhood Injuries: Scoping Review (e94038)
Huan Wang, Fengjiao Pu, Dongjun Jing, Hangcheng Liu, Miao Li, Xixi Li. . . . 914

Mechanical Contact Conditions in Wearable Reflectance Photoplethysmography: Scoping Review (e99333)
Chenxi Yang, Jiahang Xie, Zifei He, Jianqing Li, Chengyu Liu. . . . 926

Continuous Temperature Monitoring via Wearable Devices for Fever or Infection Management in Acute Care Hospitals: Scoping Review of Clinical Implementation and Data Analytics (e89773)
Yuwei Liu, Qimeng Zhao, Ka Li, Dawn Dowding. . . . 950

Effectiveness of Wrist-Worn Wearable–Guided Cardiac Rehabilitation Delivered Through Fully Remote or Hybrid Programs in Patients With Cardiovascular Disease: Systematic Review and Meta-Analysis of Randomized Controlled Trials (e101306)
Jun Choi, So Lee, Kyung Seo, Sang Lee, Hyo Lee, Dongbin Heo, Hyun Ryu, Jong Kim, Jeong Kwon, Bo Kim, Michael Lee. . . . 971

Measuring Perceived Stress in Healthy Adults in Daily Life: Scoping Review of Smartphone-Based Self-Assessment (e94204)
Toon Colman, Tom Van Daele, Bert Bonroy, Jean-Marie Aerts, Glen Debard, Jasper Gielen, Romy Sels, Catharina Van Oost, Maria Faurholt-Jepsen, Nele De Witte. . . . 997

Distinguishing Common Digital Phenotyping and Self-Report Parameters for Monitoring and Predicting Depression: Scoping Review (e70840)
Lisa Busshart, Milica Petrovic, Rebeka Amin, Ulrich Hegerl. . . . 1043

New Insights and Effectiveness of Digital Interventions for Health Anxiety: Systematic Review and Meta-Analysis (e91108)
Jing Xu, Jifeng Shen, Guanghui Huang, Jue Wang. . . . 1139

Mobile Health Apps for Individuals Who Are Blind or Have Low Vision: Systematic Review of Evidence, Accessibility, and App Quality (e85809)
Carlos Mora, Ester Solé, Pere Llorens-Vernet, Jordi Miró. . . . 1460

XSL•FO
RenderX

Safer Sleep Guidance in Standalone UK Smartphone Apps for Parents of Newborns and Infants: Systematic mHealth App Review (e95642)
Senna Yousef, Niamh Meaney, Dawn Dowding, Norina Gasteiger. 3332

Postoperative Telerehabilitation in Patients With Hip Fracture: Systematic Review and Meta-Analysis (e77341)
Can-xin Cai, Penny Qin, Chuan-yao Liu, Peng Cai, Xijun Wei. 3651

Effects of Telehealth Interventions for People With Parkinson Disease: Systematic Review and Meta-Analysis of Randomized Controlled Trials (e70994)
Minyue Sun, Fuyou Tang, Luo min, Shiyu Wen, Shuang Wang, Huiping Jiang. 3681

Clinical Improvements From Telemedicine Interventions for Managing Type 2 Diabetes Compared With Usual Care: Systematic Review, Meta-Analysis, and Meta-Regression (e70429)
Shujie Jiang, Xianru Gao, Haiqing Diao, Yang Zhang, Guangyu Lu, Xiaoguang Liu, Yuping Li. 3703

Original Papers

Daily Negative Affect and Reaction Time Inconsistency in Emerging Adults: Ecological Momentary Assessment Study (e64397)
Lauren Rutter, Pei-Ying Chen, Prabhvir Lakhan, Laura Germine, Zoe Hawks. 1014

Objective Smartphone Screen Time Monitoring Among Caregivers of Preschool-Aged Children: Reliability Study of Monitoring Duration (e82846)
Joshua Culverhouse, Taylor Adair, Meghan Restino, Heidi Weeks, Jenny Radesky, R Weaver, Sarah Burkart, Hannah Parker, Olivia Finnegan, Anthony Holmes, Keagan Kiely, Rahul Ghosal, Srihari Nelakuditi, Michael Beets, Bridget Armstrong. 1030

Evaluation of Smartphone Camera Positioning on Artificial Intelligence Pose Estimation Accuracy for Exercise Detection: Observational Study (e82412)
Eduarda Oliosi, Soraia Ferreira, Ana Giordano, Guilherme Viveiros, José Parraca, Paulo Pereira, Federico Guede-Fernández, Salomé Azevedo.
1 0 6 1

Investigating Children's Exposure to Outdoor Food Marketing in 2 Swedish Cities Using a Smartphone App: Cross-Sectional Study (e70192)
Sofia Spolander, Petter Fagerberg, Billy Langlet, Ioannis Sarafis, Ioannis Ioakimidis. 1075

The Impact of Sunlight and Artificial Light at Night on Sleep Stages: Evidence From a 7-Day Sensor-Based Observational Study (e75898)
Andrea Montanari, Li Wang, Amit Birenboim, Basile Chaix. 1090

Smartwatch-Measured Clinician Perceptions of Emergency Department Crowding and Their Association With Objective Metrics: Prospective Observational Study (e103317)
Ji Yoon, Sookyung Park, K Muir, Pamela DeGuzman, So Shin, Hansol Chang, Taerim Kim, Se Lee, Sungjun Hong, Won Cha, Junsang Yoo. .
1 1 0 3

Platform Differences in Smartphone Sensing Features for Mental Health Across Four Annual College Cohorts: Secondary Analysis (e103711)
Ariz Naqvi, Farah Naqvi. 1121

Gender-Affirming Voice Training Delivered by a Novel Mobile App: Mixed Methods Study (e96102)
Isaac Alter, Christopher Kruglik, Sara Albert, Rachel Coleman, Mary Pitti, Alexandra Li, Katerina Andreadis, David Bayne, Eseosa Odigie, Iyra Chandra, Valentina Mocchetti, David Liao, Keith Chadwick, Anaïs Rameau. 1164

XSL•FO
RenderX

Dynamics of Digital Support Platform Use Among Family Caregivers of Children With Disabilities: Longitudinal Study (e92481)
Xiaojiao Yuan, Xi Zhao, Meng Zhang, Yurou Duan, Yizhong Shui. 1184

User-Reported Issues With Mental Health Apps: Machine-Assisted Topic Analysis of Social Media Posts (e85575)
Jack Bolter, Trisevgeni Papakonstantinou, Paulina Bondaronek. 1197

Mobile Phone Access, Usage Patterns, and Perceptions of Adolescents Living With HIV on the Use of Gamified Interventions to Improve Antiretroviral Therapy Adherence in Eswatini: Qualitative Study (e74207)
Londiwe Hlophe, Peter Nyasulu, Constance Shumba. 1209

Vaccination-Related Applications and Health Care Professionals' Observed Changes in Human Papillomavirus Vaccine Hesitancy: Cross-Sectional Survey (e77778)
Manali Desai, Shreela Sharma, Joel Fokom-Domgue, Robert Yu, Wenyaw Chan, Onyema Chido-Amajuoyi, Charles Darkoh, Sanjay Shete. . .
1 2 2 2

Patient and Clinician Attitudes Toward Mobile Health Apps: Qualitative Study (e77519)
Cornelius James, Sarah Krein, Sarah Yon, Michael Dorsch, John Piette, Brahmajee Nallamothu. 1235

Spanish Version of the mHealth App Usability Questionnaire (S-MAUQ): Translation, Adaptation, and Validation Study (e64787)
Antonio Carrillo-León, Juan Falgueras-Cano, Esther Díaz-Mohedo. 1248

Trust and Use of Recommendations for Health Apps Among European Residents: Cross-Sectional Survey (e64468)
Mariam Shokralla, Romy Willemsen, Marise Kasteleyn, Niels Chavannes, Esther Talboom-Kamp. 1262

Reluctance to Use a Psycho-Oncology Mobile App Among Patients With Primary Breast Cancer: Retrospective Cross-Sectional Survey (e71412)
Marta Pawełczak-Szastok, Anna Syska-Bielak, Aleksandra Krzywon, Michał Jarz b. 1272

Features of Mobile Health Apps for Tobacco Cessation That Appeal to Black Adults Who Use Tobacco Products: Focus Group Study (e63340)
Sonia Clark, Remi Philips, Christine Kistler, Adam Goldstein, Chineme Enyioha. 1287

Designing mHealth Apps for Substance Use Recovery Through Real-World Co-Design and Deployment: Mixed Methods Study (e83984)
Teale Masrani, David Hodgins, Hyoun Kim, Katherine Rittenbach, Erika Johnson, Geoffrey Messier. 1302

Adapting and Validating Tools to Assess the Usability and Acceptability of mHealth Tools Among Community Health Workers in Rural Settings: Development and Usability Study (e64916)
Jonathan Nkurunziza, Sarah Nuss, Eve Estrada, Marthe Kubwimana, Adeline Boatin, Laban Bikorimana, Richard Fletcher, Nissi Byiringiro, Bethany Hedt-Gauthier, Vincent Cubaka. 1321

Lived Experiences of Older Adults Using Wearables With Real-Time Feedback: Phenomenological Study (e71509)
Tien-Ying Lu, Aurora Rosato, Seraina Dual, Sanna Kuoppamäki. 1332

Smartphone Apps for Cannabis Cessation: Quality Assessment and Content Analysis (e58908)
Siddharth Seth, Sumedha Kushwaha, Reshma Prashad, Michael Chaiton. 1350

XSL•FO
RenderX

Mobile Health App Acceptance in Japan's Aging Society: Multigroup Structural Equation Modeling Based on the Extended Unified Theory of Acceptance and Use of Technology and eHealth Literacy Frameworks (e87832)
Yasue Fukuda, Koji Fukuda. . . . 1360

Chinese Version of the mHealth App Usability Questionnaire for Health Care Providers: Cross-Sectional Translation and Validation Study (e86783)
Mengxia Chen, Wenjun Gao, Yijie Pan, Mengdi Wang, Bing Yu, Dongmei Li, Huanhuan Hu, Jing Zheng, Lingjuan Zhang, Yuan Gao. . . . 1383

Design Preferences for Mental Health Apps Among Nondigitally Native Adults With Chronic Pain: Qualitative Analysis (e87358)
Abby Cheng, Christine Gou, Adriana Martin, Sarah Hartz, Joanna Abraham. . . . 1397

Accessing Germany's Electronic Patient Record Through Health Insurer–Provided Apps: Qualitative Interview Study (e95294)
Susann May, Manuela Marquardt, Gina Barzen, Frances Seifert, Mary Nyarko, Antonia Niemeyer, Sonja Gscheidle, Liane Schenk, Sebastian Spethmann, Felix Muehlensiepen. . . . 1414

Mobile Heat Early Warning System for Pregnant and Postpartum Women in South Africa: Prospective Observational Evaluation Study (e102863)
Clara Heil, Shobna Sawry, Susanne Frennert, Jean Le Roux, Pascalia Munyewende, Sibusiso Mkwananzi, Ijeoma Solarin, Perkins Watambwa, Koen van der Sanden, Isabelle Lange, Günter Alce, Veronique Filippi, Stanley Luchters, Chuansi Gao. . . . 1428

Tablet-Based Neurocognitive Tests for Ugandan Adolescents Living With HIV: Cross-Sectional Validation Study (e103288)
Reuben Robbins, Angel Nanteza, Joy Gumikiriza-Onoria, Anthony Santoro, Christopher Ferraris, Christine Karungi, Daphne Tsapalas, Nana Asiedu, Courtney Kirsch, Mei Tan, Jun Liu, Itziar Familiar-Lopez, Victor Musiime, Sahera Dirajlal-Fargo. . . . 1446

Frontline Health Workers' Perspectives of the World Health Organization Skin Neglected Tropical Diseases App in Kenya: Qualitative Study on AI-Embedded mHealth Implementation (e81829)
Emily Quilter, Ruth Nyangacha, Esther Kinyeru, Aïna Fuster Casanovas, Amberly Brigden, Kenton O'Hara, Diana Atieno, José Ruiz-Postigo, Carme Carrion. . . . 1489

Feasibility and Acceptability of a Mobile App and Wearable Device for Collecting Mental Health Survey and Passively Sensed Data Among Health Care Workers in Kenya: Mixed Methods Pilot Study (e77761)
Linda Khakali, Andrew Aballa, Dorcas Mwigereri, Eileen Weinheimer-Haus, Willie Njoroge, Rachel Maina, Amos Bunde, Peris Musitia, Moses Nyongesa, James Orwa, Jasmit Shah, Zhenke Wu, Anthony Ngugi, Lukoye Atwoli, Srijan Sen, Candace Kolars, Akbar Waljee, Amina Abubakar, Elena Frank, Zul Merali. . . . 1511

Feasibility and Acceptability of mPallCare, a Digital Health Intervention for People Living With Advanced Cancer in a Refugee Settlement in Uganda: Mixed Methods Study (e73483)
Eve Namisango, Agatha Aduro, William Goodman, Shaunna Burke, Raphael Ryabonye, Nickson Mutaasa, Timothy Muyami, Elizabeth Nabirye, Dennis Olodi, Viola Ederu, Bassey Ebenso, Omolola Salako, Kehinde Okunade, Desiree Azizoddin, Mhoira Leng, Karl Lorenz, Felix Muehlensiepen, Richard Powell, Matthew Allsop. . . . 1524

Terminal Digit Preference and Threshold Avoidance in Digital Blood Pressure Measurements During Pregnancy: Secondary Analysis of Data From the CLIP and PRECISE Cohorts (e73307)
Peter von Dadelszen, Akshdeep Sandhu, Jeffrey Bone, Rahat Qureshi, Olukayode Dada, Ashalata Mallapur, Marleen Temmerman, Marie-Laure Volvert, Joseph Waiswa, Moses Mukhanya, Hiten Mistry, Shivaprasad Goudar, Hawanatu Jah, Angela Koech, Charfudin Sacoor, Anifa Vala, Khatia Munguambe, Anna Roca, Olelekan Adetoro, Esperanca Sevene, Marianne Vidler, Mrutyunjaya Bellad, John Sotunsa, Hannah Blencowe, Zulfiqar Bhutta, Umberto d'Alessandro, John Ansermino, Guy Dumont, Laura Magee, CLIP Trials Study Group, PRECISE Network. . . . 1547

A Community-Based Usability Study of an AI-Enabled Oral Cancer Screening App Operated by Village Health Volunteers: Mixed Methods Study (e83738)
Mansuang Wongsapai, Kornwipa Wudtijureepun, Thawatchai Suthachai, Decha Tamdee, Jitjiroj Ittichaicharoen, Patiwet Wuttisarnwattana, Siriwan Suebnukarn. . . . 1560

XSL•FO
RenderX

Timed Up-and-Go Test Using Instrumented Insoles to Monitor Tibial or Malleolar Fracture Healing in Patients With Healing Fractures or Nonunion: Longitudinal and Cross-Sectional Exploratory Study (e92493)
Elke Warmerdam, Lea-Marie Burger, Bergita Ganse. 1575

Feasibility of a Stool State Check App Using AI During Bowel Preparation Before Colonoscopy: Multicenter Prospective Study (SCAN Study) (e91274)
Atsushi Inaba, Kensuke Shinmura, Shozo Osera, Takahiro Yamada, Hiroaki Kon, Maki Kanazawa, Naoki Sugimura, Yasushi Sano, Hiroko Hosaka, Toshio Uraoka, Daiki Sato, Yusuke Yoda, Hiroyuki Takamaru, Yutaka Saito, Toshihiko Gocho, Atsushi Katagiri, Kazuhisa Yamaguchi, Takahisa Matsuda, Atsuki Imai, Hitomi Fujimoto, Hiroki Matsuzaki, Nobuyoshi Takeshita, Masashi Wakabayashi, Hiroaki Ikematsu, Tomonori Yano. . . .
1 5 9 0

Outcomes of Mobile Health Use in Sinonasal Surgery: Retrospective Cohort Study (e75403)
Heli Majeethia, Akshay Prabhakar, Justina Varghese, Najm Khan, Roshan Dongre, Vincent Provasek, Faizaan Khan, Zain Mehdi, Omar Ahmed, Masayoshi Takashima. 1605

Long-Term Engagement With a Physical Activity App Among Gynecologic Cancer Survivors in the LETSGO Trial: 12-Month Prospective, Multicenter, Quasi-Experimental Study (e84838)
Sindre Fosstveit, Ingvild Vistad, Mette Skorstad, Sveinung Berntsen. 1614

Predictors of Mobile Health App Acceptance Among Nurses in Kashan, Iran: Cross-Sectional Study (e93888)
Narjes Mirabootalebi, Felix Holl, Walter Swoboda, Zahra Tagharrobi, Fatemeh Rangraz Jeddi, Hossein Akbari. 1631

Evaluation of the Impact of a Mobile App (LoAD Calc) on the Calculation of Maximum Safe Doses of Local Anesthetics: Randomized Controlled Trial (e89236)
Alexandra Stefani, Pietro Fubini, Tal Beckmann, Caroline Samer, Georges Savoldelli, Mélanie Suppan. 1647

A Deep Neural Network for Interpreting Wearable Electrocardiogram Data in Atrial Fibrillation: Prospective Observational Diagnostic Accuracy Study (e82475)
Olli Rantula, Jukka Lipponen, Jari Halonen, Helena Jäntti, Tuomas Rissanen, Mika Tarvainen, Noora Naukkarinen, Eemu-Samuli Seljola, Onni Santala, Jagdeep Sedha, Tero Martikainen, Juha Hartikainen. 1658

Definition and Test-Retest Reliability of a Monitoring Method Integrating Accelerometric Actigraphy and Bluetooth Indoor Location Tracking Applied in a Long-Term Residential Unit for Persons With Dementia: Longitudinal Observational Study (e70188)
Marco Rabuffetti, Pietro Trimarchi, Alessia Gallucci, Ilaria Carpinella, Elena Kisel, Maria Andriani, Ennio De Giovannini, Gaia Bailo, Fabrizio Giunco, Maurizio Ferrarin. 1674

Failed Implementation of Mobile Access to Electronic Health Records in Home Care: Qualitative Study in Sweden (e69590)
Lovisa Jäderlund Hagstedt, Helena Hvitfeldt, Maria Hägglund. 1694

Evaluating the Feasibility of an Electronic Patient-Reported Outcomes Platform Integrating Electronic Health Records and a Mobile Messaging App in Breast Cancer Radiotherapy: Retrospective Cross-Sectional Study (e67514)
Jong Baek, Haeyoung Kim, Won Cho, Nalee Kim, Tae Lee, Won Cha. 1711

Use of an AI Image Recognition–Based Smartphone App in Personalized Bowel Preparation Before Colonoscopy: Prospective Randomized Controlled Trial (e77701)
Fei Han, Shilin Gao, Hanqiang Zhan, Nan Wang, Zhaoxiang Song, Jinlin Xie, Yuhan Hou, Yimeng Deng, Jianning Yao. 1723

Dietary and Exercise Relapse Management in Mobile Health Apps for Weight and Glycemic Control: Descriptive Qualitative Study (e94792)
Ahlam Alhajri, Afnan Alandanoosi, Hala Al-Otaibi. 1734

XSL•FO
RenderX

Effects of a Hypertension Management Mobile App on Urinary Sodium Excretion in Patients With Chronic Kidney Disease: Randomized Controlled Trial (e68447)
Takayuki Kawaoka, Yusuke Sakaguchi, Tatsufumi Oka, Yohei Doi, Ryohei Yamamoto, Isao Matsui, Masayuki Mizui, Jun-Ya Kaimori, Yoshitaka Isaka. . . . 1749

A Mobile App–Based Individualized Nonpharmacological Intervention for Behavioral and Psychological Symptoms in Dementia: Pilot Randomized Controlled Trial (e79469)
Eunhee Cho, Minhee Yang, Sinwoo Hwang, Eunkyo Kim, Jungwon Cho, Min Kim. . . . 1762

Contribution of Longitudinal Mobile Health Measures in the Dynamic Track of Patients With Major Depressive Disorder: Multiple Centers, Prospective Cohort Study Using Functional Data Analysis and Machine Learning (e81397)
Rou Zhong, Nanxi Li, Le Xiao, Lei Feng, Yuan Feng, Gang Wang, Xuequan Zhu. . . . 1779

Digital Structured Education With Behavioral Nudge Tools for Adults With Type 2 Diabetes: Multicenter Randomized Controlled Trial (e91407)
Yan Lin, Yingchun Zeng, Kaining Chen, Zongcun Chen, Caihua Ye, Ying Zhou, Qiwei Zhou, Chengying Yu, Vivien Wu, Samuel Seidu, Bin Li, Xinjun Jiang. . . . 1794

Radiographic Knee Osteoarthritis Identification Using Wearable Plantar Pressure: Secondary Cross-Sectional Analysis of Interpretable Framework Integrating Biomechanical Descriptors and the Multiscale Synergy Index (e101026)
Guanyu Xin, Hongbo Yao, Jingyuan Qiao, Huilin Yao, Siyuan Li, Lu Liu, Jingyi Zhang, Ke Qin, Lin Shu. . . . 1811

The Mediating Role of Sleep Quality in Associations of Smartphone Addiction and Time-Specific Use With Prenatal Depression and Anxiety: Cross-Sectional Observational Study (e100064)
Xiangxin Zhang, Mixia Ouyang, Xiaoli Wu, Xinghua Yang, Xinyi Wang, Yuxinqiao Cao, Mingmin Xu, Yanling Shu, Mingyang Wu. . . . 1831

Multidomain Outcomes of a Digital Self-Management Program in Rural Adults With Chronic Conditions: Single-Arm Prospective Interventional Study (e95283)
Nathan Smith, Gabriela Ilie, Cody MacDonald, James Langley, Robert Rutledge. . . . 1850

Evaluating City-Cut Geographic Risk Classification for a Pandemic Travel-Record System in Shanghai, China: Retrospective Observational Analysis (e64941)
Rowan Wang, Baofeng Shi, Shenghan Li, Cong Xie, Qingpeng Zhang. . . . 1872

Impact of Mobilization Facilitated by Wearable Device Enhanced Patient Monitoring/Electrophysiology Pod–Based Feedback on Postoperative Complications Following Colorectal Cancer Surgery: Randomized Controlled Trial (e70534)
Yang Meng, Fengyan Fan, Yumei Ma, Nong Yan, Huan Wang, Zhen Zhang, Yiting Wang, Hailong Dong, Huang Nie. . . . 1886

Real-World Long-Term Engagement With a Mobile App Intervention to Improve Self-Management of Type 2 Diabetes Mellitus in China (SMARTDiabetes): Mixed Methods Study (e76699)
Xuanchen Tao, David Peiris, Yuxia Ma, Yaoshen Zhang, Hongyi Song, Limin Mao, Puhong Zhang. . . . 1901

Digital Cognitive Behavioral Therapy for Chronic Insomnia in South Korea: Cost-Effectiveness Analysis Using Decision Tree and Markov Modeling Based on a Secondary Analysis of a Randomized Clinical Trial (e71750)
Jung Kim, Minjeong Sohn, Jimin Woo, Jaeyong Shin, Euna Han. . . . 1927

Efficacy and Safety of Mobile App–Based Metamemory Cognitive Training for Mild Cognitive Impairment: Multicenter Randomized Clinical Trial (e73464)
Sunyoung Kang, Jung-In Lim, Lukas Stenzel, Keun Kim, Eosu Kim, Hong Jeon, Doo-Heum Park, Hyun Lim, YongSoo Shim, Jae-Won Jang, Yeshin Kim, Sanghoon Lee, Kee Park. . . . 1942

XSL•FO
RenderX

Development and Validation of a Novel Lymphedema Monitoring System Using a Bodysuit and a Smartphone: Prospective Comparative Study (e77935)
Ryo Karakawa, Hidehiko Yoshimatsu, Katsu Karakawa, Kuniko Utsugi, Kengo Ono, Eri Kojima, Takayasu Yamada, Duncan Eason, Francois Goffinet, Tomoyuki Yano. . . . 1961

Associations Between Technology Attitudes, App Use, and Health Outcomes in a Diabetes Self-Management Education and Support Intervention: Secondary Analysis of a Randomized Controlled Trial (e105830)
King Oware, Gang Han, Samuel Towne Jr, Christi Esquivel, SangNam Ahn, Shinduk Lee, Matthew Smith, Marcia Ory. . . . 1972

The Usage Effects, Effect Modifiers, and Experiences of a Web-Based App for Healthy Habit Formation in Adults: Exploratory Analysis of a Quasi-Experimental Real-World Intervention (e84076)
Eeva Rantala, Mikko Valtanen, Adil Umer, Suvi Parikka, Jussi Pihlajamäki, Ilona Ruotsalainen, Jaana Lindström. . . . 1987

Caregiver App Use and Child Behavioral Adherence in Relation to Obesity-Related Outcomes in an mHealth App-Assisted Childhood Obesity Intervention: Secondary Analysis of a Cluster Randomized Trial (e89548)
Shiyu Yan, Jinlang Lyu, Chunjie Yin, Wenhao Li, Yan Li, Hai-jun Wang. . . . 2006

How People Integrate Mental Health Apps Into Everyday Self-Care: Qualitative Study (e95414)
Vít Suchý, Zuzana Scott, Barbora Boštíková, Marek Janoušek, Eliška Josková, Anna Šev íková. . . . 2020

Daily Reminders Increase Adherence to an Electronic Headache Diary (DMKG-App): Retrospective Longitudinal Cohort Study (e73259)
Ruth Ruscheweyh, Lars Neeb, Bianca Raffaelli, Torsten Kraya, Tobias Freilinger, Thomas Dresler, Gudrun Gossrau, Tim Jürgens, Victoria Ruschil, Jörg Scheidt, Yannic Siebenhaar. . . . 2035

Weight Loss and Engagement Among Active Users of a Dietitian-Led Digital Nutrition Platform: Retrospective Cohort Study (e94732)
Frederick Goddard, Jessica Garay, Francisco Teixeira Neves, Alexandra Ligeti, Liudmila Zhaunova, Manuela Abreu. . . . 2048

Feasibility and Preliminary Efficacy of a “Sit Less” Program Leveraging Fitbit Tracking and Tailored Text Messages in Cardiometabolic Disease: Findings From 2 Parallel Randomized Controlled Trials in Coronary Artery Disease and Type 2 Diabetes (e56497)
Chorong Park, Mary Dietrich, Britta Larsen, Lindsay Mayberry, Abigail Doyle, Soojung Ahn, Jason Jean, Kevin Maquiling, Mulubrhan Mogos, James Muchira, Shelagh Mulvaney. . . . 2061

Immediate and Sustained Improvements in Mood and Stress Associated With Yuna, an AI-Powered Digital Mental Health Intervention: Real-World Retrospective Study (e94070)
Kelsey McAlister, Courtney Jewell, Chad Stecher, Jennifer Huberty. . . . 2080

Self-Monitoring of Weight Loss Over Time: Secondary Data Analysis of a Randomized Controlled Trial (e71604)
Renata de Oliveira, Sharon Nabwire, Heta Merikallio, Markku J Savolainen, Janne Hukkanen, Harri Oinas-Kukkonen. . . . 2096

Behavioral Mechanisms of a Digital Health Intervention for Self-Management in Type 2 Diabetes Mellitus: Prospective Longitudinal Cohort Study (e79081)
Yibo Wu, Yang Ni, Yang Jiang, Zijie Xu, Hewei Min, Ping Chen, Xinbao Gu, Bingyang Kong, Yadi Gan, Pei Li, Mingzi Li, Xiaohui Guo, Xuxi Zhang, Aijuan Ma, Xinying Sun. . . . 2108

Feasibility, Usability, and Preliminary Effectiveness of an mHealth App to Promote Screening Behaviors Among High-Risk Populations for Breast Cancer: Randomized Controlled Pilot Study (e86429)
Linping Zhang, Mengjiao Xu, Xian Zhang, Xiaoxu Li, Shuo Wang, Fang Zhou, Jing Han. . . . 2124

XSL•FO
RenderX

The ManageHF Just-in-Time Adaptive Mobile App Interventions to Promote Self-Management and Improve Outcomes in Heart Failure: Randomized Controlled Trial (e74121)
Michael Dorsch, Mohamed Ali, Amy Krambrink, Giselle Kolenic, Sabah Ganai, Juan Arzac, Xutong Zhang, Kaitlyn Greer, Amit Shah, Jennifer Cowger, Gregory Ewald, Jo Rodgers, Dave Dixon, Todd Koelling, Scott Hummel. . . . 2144

Adoption of App-Based Peer Support by Canadian Health Care Providers: Mixed Methods Organizational Implementation Study (e85658)
Sandra Moll, Melissa Parker, Peter Smith, Edward Sykes. . . . 2157

Action Planning for Reducing Sugar-Sweetened Beverage Intake in Appalachian Adults: Longitudinal Process Evaluation of a Digital Behavioral Health Intervention (e71241)
Annie Reid, Lee Ritterband, Kathleen Porter, Donna-Jean Brock, Christina Frederick, Jamie Zoellner. . . . 2175

A Metaverse Program to Prevent E-Cigarette Use Among Adolescents: Pilot Mixed Methods Study (e65817)
Hyunmi Son, Gyumin Han. . . . 2193

Enhancing Cognitive Functions of Older Adults With Software Robot: Longitudinal Exploratory Field Study (e75308)
Byunghun Yun, Bohee Kim, Hyunjeong Ko, Jinsung Kim, Bori Kim, Whani Kim, Soyoon Park, Jinwoo Kim, Jee Lee, Geon Kim. . . . 2204

Long-Term Effectiveness of Unguided Internet-Based Cognitive Behavioral Therapy on Major Depressive Disorder in Chinese Adults: Randomized Controlled Trial With a 12-Month Follow-Up (e68394)
Wenjing Zhou, Huimin Zhang, Yunbin Jiang, Yanzhi Li, Guangduoji Shi, Hao Zhao, Wanxin Wang, Yuhua Liao, Yifeng Liu, Jiejing Hao, Roger McIntyre, Beifang Fan, Ciyong Lu. . . . 2225

Efficacy of a Cognitive Behavioral Therapy–Based Online Self-Help Group for Depression and Suicide Ideation: Randomized Controlled Trial (e76028)
Minkyung Yim, Haeun Kim, Soo-Eun Lee, Eunjin Jo, Ji-Won Hur. . . . 2246

Alleviating Nurse Burnout With an Artificial Intelligence–Selected Mobile Cognitive Behavioral Therapy–Based Intervention: Mixed Methods Randomized Controlled Trial (e85986)
Yeongeun Kim, Chiyoung Cha, Gumhee Baek. . . . 2266

Effects, Acceptability, and Use of a Dynamically Tailored Mobile What Do You Drink Intervention to Reduce Excessive Drinking Among Adolescents and Young Adults in the Netherlands: Randomized Controlled Trial (e68468)
Hilde van Keulen, Carmen Voogt, Marloes Kleinjan, Arjan Huizing, Rosa Andree, Pepijn van Empelen. . . . 2282

Adherence to and Engagement With an mHealth Physical Activity Intervention After Mild Stroke or Transient Ischemic Attack: Secondary Analysis of a Feasibility Randomized Controlled Trial (e75662)
Hanna Lagerlund, Lucian Bezuidenhout, Sophia Humphries, Lisa Holmlund, Lydia Kwak, Charlotte Häger, David Moulaee Conradsson. . . . 2303

Comparison Between Browser- and App-Based Versions of a Program for Self-Management of Mild to Moderate Depression: Log Data Analysis of a Convenience Sample (e58835)
Katharina Scholze, Caroline Allenhof, Ulrich Hegerl. . . . 2319

Promoting Self-Regulated Social Media Use on Smartphones With a Mobile Intervention App (Wellspent): Randomized Controlled Trial (e56824)
Lea Mertens, Lina Brockmeier, Christina Roitzheim, Theda Radtke, Tilman Dingler, Jan Keller. . . . 2327

Short-Term Effects of an mHealth Intervention on Healthy Behaviors and Cardiometabolic Health in Sedentary Employees: Quasi-Experimental Study (e70074)
Yun-Ping Lin, Shu-Hua Lu, Kwo-Chen Lee, Wei-Fen Ma, Ya-Fang Ho, Wen-Chun Liao, Hui-Ting Yang, OiSaeng Hong. . . . 2344

XSL•FO
RenderX

One-Year Trajectory of Step Counts and Weight Loss in Adults With Overweight/Obesity: Retrospective Cohort Study (e80339)
Kenshiro Taguchi, Asuka Oyama, Jun'ichi Kotoku, Hiroshi Toki, Ryohei Yamamoto. . . . 2367

Traditional Social Sports Games and Mental Training for Smartphone Addiction and Psychological Distress in School-Aged Adolescents: Randomized Controlled Trial (e85353)
Mohamed Yaakoubi, Ahmed Ghorbel, Hiba Abdelkafi, Liwa Masmoudi, Omar Trabelsi, Safaa Elkholi, Fatma Yagin, Georgian Badicu, Adnene Gharbi. . . . 2383

Multimodal Personalized Mobile Health Just-in-Time Adaptive Intervention for Occupational Stress Management: Pilot Study (e79642)
Beatrix Kunas, Oliver Jung, Christoph Schranz, Mathias Schmoigl-Tonis, Julian Mehlis, Anton-Rupert Laireiter. . . . 2397

A Theory-Based Digital Intervention to Improve Maternal Oral Health Behaviors for Young Children: Quasi-Experimental Study (e79002)
Mei Zhao, Min Liu, Shiyu Wang, Xiaoyue Zhang, Chun Chang, Qingping Yun. . . . 2413

Effectiveness of Step Goal Personalization Strategies on Physical Activity in a Mobile Health App: A Field Study (e81779)
Xia Liu, Tammo Bijmolt, Marijke Leliveld, Ernst Noppers. . . . 2425

Assessing Wearable mHealth Adherence in Underserved Adolescents and its Associations With Physical Activity, Sports, and Safety Perceptions: Prospective Cohort Study (e80465)
Annabel Nunez-Gaunaurd, Michele Raya. . . . 2437

Efficacy of a Digital Peer Support Program on Weight Management and Mental Health in University Students With Preobesity: Randomized Controlled Trial (e78960)
Xingyu Liu, Ting Liu, Tao Chen, Ruisi Ma. . . . 2453

Analysis of Training Behavior in Users of a Fitness App: Cross-Sectional Study (e72201)
Andrea Fuente-Vidal, Roger Prat, Juan Arribas-Marin, Oscar Bastidas-Jossa, Myriam Guerra-Balic, Begonya Garcia-Zapirain, Joel Montane, Javier Jerez-Roig. . . . 2470

Effectiveness of a Hybrid Community-Based Heart-Healthy Lifestyle Intervention: Three-Arm Randomized Controlled Trial Integrating mHealth and Motivational Interviewing (e76521)
Jina Choo, Yura Shin, Songwhi Noh, Juneyoung Lee. . . . 2492

A Brief Web-Based and Mobile Intervention of Intermittent Fasting With Meal Support for Weight Loss Among Adults With Overweight and Obesity in Japan: Pilot Randomized Controlled Trial (e58930)
Takashi Noda, Tomonari Shimamoto, Kosuke Kiyohara, Mie Imanaka, Norihiro Nishioka, Yukiko Tateyama, Taku Iwami. . . . 2511

Effectiveness of Smartphone-Based Dyadic Interventions to Increase Physical Activity in Romantic Couples: Microrandomized Trial (e67136)
Patrick Höhener, Robert Tobias, James Allen, Pascal Küng, Urte Scholz. . . . 2528

Adolescents' Engagement With an mHealth Multiple Health Behavior Change Intervention (LIFE4YOUth): Mixed Methods and Qualitative Comparative Analysis (e88054)
Anna Seiterö, Pontus Henriksson, Kristin Thomas, Marie Löf, Marcus Bendtsen, Ulrika Müssener. . . . 2548

Assessing Mental Health and Emotional States by Using Smartphone Photoplethysmography–Based Digital Pulse Waveform Analysis: Cross-Sectional Observational Study (e81301)
Ivan Liu, Luming Hu, Jing Luo, Chang Liu, Qi Zhong, Shiguang Ni. . . . 2565

XSL•FO
RenderX

Reducing Sedentary Time After Knee Replacement Using a Multicomponent mHealth Intervention: Randomized Controlled Trial (e83148)
Christine Pellegrini, Clare Kennerley, Sara Wilcox, Jungwha Lee, Katherine DeVivo, Kailyn Horn, Scott Jamieson, Jeffrey Hopkins, Harley Davis, J Jackson III. . . . 2585

Automated Physical Activity Support for Adults and Youth From Low-Income Communities: Single-Arm Pilot Study (e76991)
Jordan Carlson, Frank Materia, Mallory Moon, Suryeon Ryu, Cory Yeager, Chelsea Steel, Kacee Shields, Harpreet Gill, Jannette Berkley-Patton, Delwyn Catley. . . . 2599

Tobacco Use Trajectories and Associated Changes in Biometrics and Sleep During the First 72 Weeks of Wearable Membership: Observational Cohort Study (e98116)
Dylan Curran, Josh Leota, William Hippel, Finnbarr Fielding, Christopher Chapman, Jenna Cohen, David Presby, Reha Jhunjhunwala, Kristen Holmes, Gregory Grosicki. . . . 2614

Week-Ahead Prediction of High-Risk Drinking Episodes Among Young Adults Using Wearable Biosignals and Psychological Vulnerabilities: Prospective Observational Machine Learning Study (e88223)
Jae Kwak, Hae Lee, Sun-Jin Jo, Jun Kwon, Sun Kwon, Yena Kim, Haejung Lee. . . . 2631

Brief Web-Based Psychological Intervention for Depressive Symptoms Among Chinese Adolescents: Randomized Controlled Trial (e86134)
Pei Liu, Xinwei Wang, Xiaoxia Duan, Mingjing Situ, Yujie Tao, Zhaozhi Yang, Shengnan Zhao, Lei Li, Sixun Li, Yuchu Jiang, Manxue Zhang, Di Jing, Xinyi Yu, Yi Huang. . . . 2652

A Mobile Virtual Reality–Based Intervention for Stress Reduction Among Young Adults: Pilot Randomized Controlled Trial (e83438)
Houlin Hong, Matthew Lootens, Rebecca Bachman, Luis Olivera Leon, Ann Dunlap, Emilio Goldenhersch, Nicolas Rosencovich, Cristian Waitman, Katarzyna Wyka, Terry Huang. . . . 2670

User Engagement and Feature Preferences in an AI-Powered mHealth Intervention for Diabetes Prevention: Secondary Analysis of a Randomized Controlled Trial (e92981)
Benjamin Lalani, Gabriela Siew, Yllka Valdez, Aliyah Shehadeh, Daniel Zade, Kristin Riekert, Nestoras Mathioudakis, AI-DPP Study Group. . 2 6 8 3

Engagement With a Relaxation-Based Mobile Health Intervention for Perioperative Anxiety: Prospective Longitudinal Cohort Mixed Methods Study (e64278)
Xinghui Yan, Afton Hassett, Jennifer Waljee, Mark Newman, Sun Park, Rongqi Bei, Noelle Carlozzi. . . . 2697

Association Between Intradialytic Foot Temperature Gradients and Severe Peripheral Artery Disease in Patients Undergoing Hemodialysis: Prospective Cohort Study (e96782)
Donna Lin, Chih-Cheng Wu, Chung-Wei Yang, Jen-Kuang Lee, Chien-Yi Hsu. . . . 2716

Mobile AI-Assisted Oral Lesion Triage for Community Oral Cancer Screening: Prospective Field Evaluation Study (e101721)
Hao-Yun Liu, Yu-Cheng Huang, Si-Wei Chen, Ling-Ying Wei, Cheng-Ying Chou, Jang-Jaer Lee, Shyh-Jye Chen, Chien-Chang Lee. . . . 2735

CogniPaz, an Ecologically Oriented Gamified mHealth Tool for Detecting Mild Cognitive Impairment: Prospective Cohort Diagnostic Accuracy Study (e99688)
Carlos Estebas-Armas, Alejandro Medina-López, Daniel Blanco-de-Miguel, Itsaso Losantos-García, María Hernández-Barral, Elisa Alonso-López, Sergio Henche-Rodríguez, Oscar Ocampo-Pérez, Gerardo Zmork-Martínez, María Alonso-de-Leciñana, Ana Frank-García, Ángel Martín-Montes. . . . 2748

Passive Screening for Depressive Symptoms Using Daily Wrist Actigraphy and Deep Learning: Model Development and Validation Study (e91479)
Doljinsuren Enkhbayar, Somin Oh, Jinhee Lee, Min-Hyuk Kim, Erdenebayar Urtnasan, Jaehong Key. . . . 2772

XSL•FO
RenderX

Smartphone-Based Physical Performance and Multidimensional Determinants of Self-Reported Knee Pain in Community-Dwelling Older Adults: Cross-Sectional Machine Learning and Network Analysis Study (e92466)
Ui-jae Hwang, Peng Xia, Tianxiang Fan, Arnold Wong, Oh-yun Kwon, Siu-ngor Fu. 2791

Utility of Tablet-Based Eye Tracking for Early Screening of Poststroke Cognitive Impairment: Diagnostic Cohort Study (e83194)
Boxu Xie, Ronghui Ren, Yuting Zhang, Xuan Wang, Junli Zhen. 2808

Facilitators of and Barriers to Implementation of a Tablet-Based Digital Health Program for Colorectal Cancer Screening in Primary Care: Qualitative Pragmatic Implementation Study (e74206)
Mark Cromo, Aliza Randazzo, David Miller Jr, Kristie Foley, Ajay Dharod, Anna Snavely, Elena Wright, Holly Brower, Mark Dignan. 2828

Gait Analysis for Identifying Normal Cognition, Subjective Cognitive Decline, and Mild Cognitive Impairment in Parkinson Disease: Diagnostic Study (e69273)
Juan Huang, Lingyu Wu, Hui Wang, Lin Chen, Binbin Hu, Fei Zhang, Kang Ren, Yun Ling, Zhonglue Chen, Wei Huang. 2837

Co-Designed Text Messaging Intervention for Parents of Children With Early Childhood Caries: Randomized Controlled Trial (e91720)
Rebecca Chen, Michelle Irving, Harleen Kumar, Neeta Prabhu, Bradley Christian, Woosung Sohn, Heiko Spallek, Simone Marschner, Clara Chow. 2855

Time to Cancer Screening Completion Following an SMS Reminder Intervention in a Large Federally Qualified Health Center Network: Secondary Data Analysis (e93912)
Tonghui Xu, Summer Chavez, Ben King, Gonzalo Ramirez-Pulido, Chinedum Ojinnaka, Daniel Osayi, Valery Kounga, Omolola Adepoju. . . . 2 8 6 8

Individualized Treatment Effects of a Digital Smoking Cessation Intervention Among Individuals Looking Online for Help: Secondary Analysis of a Randomized Controlled Trial (e63578)
Joel Crawford, Jenny Blomqvist, Katarina Ulfsdotter Gunnarsson, Preben Bendtsen, Marcus Bendtsen. 2880

WeChat-Based Intervention for Glycemic Control in Patients With Type 2 Diabetes Mellitus: Multicenter Randomized Controlled Trial (e80738)
Chuanfen Zheng, Xiong Dou, Xiaotao Xiong, En-Yu Lei, Ming Jiang, Yulin Wu, Jing Yu, Xianjun Wang, Ling Zhang, Honghui Rong, Lu Lu, Fengju Li, Ting Luo, Xiangyu Ma, Ji-An Chen. 2894

WeChat-Delivered Mobile Medical Nutrition Therapy Intervention in Gestational Diabetes Mellitus: Randomized Controlled Trial (e67410)
XiaoRong Wang, MiaoMiao Bian, Fang Li, Chang Zhou, LiNing Mu. 2909

A Pilot mHealth Text Messaging Program Targeting Parents During the First 2000 Days: Nonrandomized Repeat Cross-Sectional Analysis to Evaluate Feasibility, Engagement, Acceptability, and Potential Effectiveness (e83162)
Jacklyn Jackson, Nayerra Hudson, Tessa Delaney, Alison Brown, Sienna Kavalec, Jessica Pinfold, Rebecca Sewter, Hannah McCormick, Rebecca Liackman, Christophe Lecathelinais, Lynda Davies, Luke Wolfenden, Daniel Groombridge, Belinda Tully, Sinead Redman, Tauri Smart, Nafiseh Ghafournia, Paul Craven, Rachel Sutherland. 2927

Efficacy of a Mobile Phone–Based Life-Skills Training Program for Addiction Prevention Optimized Among Adolescents With Low Program Engagement: Cluster-Randomized Controlled Trial (e78081)
Severin Haug, Nikolaos Boumparis, Olivia Studhalter, Andreas Wenger, Michael Schaub, Nikolai Kiselev. 2946

Characteristics of Tailored Text Messages Associated With Increased Physical Activity Among Cardiac Rehabilitation Enrollees: Secondary Analysis of a Microrandomized Trial (e79792)
Namratha Atluri, Kashvi Gupta, Tanima Basu, Evan Luff, Jieru Shi, Thomas Boyden, Bhramar Mukherjee, Sachin Kheterpal, Predrag Klasnja, Walter Dempsey, Brahmajee Nallamothu, Jessica Golbus. 2962

XSL•FO
RenderX

Personalized Mobile App–Based Messaging and Step Tracking for Improving Walking Behavior in Adults: Randomized Controlled Trial (e84430)
Sarang Jang, Aeree Sohn. . . . 2974

Behavioral Interactions Shaping the Use of a Mobile Health Intervention for Diabetic Foot Self-Care Among People With Diabetes and Caregivers: COM-B and TDF Qualitative Analysis (e101195)
Sofia Rebull, Guillermo Almeida-Huanca, Josselyn Jauregui Mantilla, Lourdes Cruzado-Castro, Sebastian García Cruz, Pablo Perel, Juan Miranda, Juned Siddique, Jacqueline Seiglie, Hueiming Liu, Stephen Jan, Laura Downey, David Beran, David Armstrong, German Comina, María Lazo-Porras, Jessica Zafra-Tanaka. . . . 2992

Development and Launch of a Dutch Mobile App (MediMama) on Over-the-Counter Medication Safety During Pregnancy and Breastfeeding: Development and Usability Study (e85948)
Veronique Maas, Maud de Feijter, Anneke Passier, Miranda van Tuyl, Agnes Kant, Maartje Conijn. . . . 3008

Acceptability, Feasibility, and Outcome Responsiveness of the Joint Effort Mobile App for Promoting Lower-Risk Cannabis Use Among Young Adults: Pilot Randomized Controlled Trial (e71957)
José Côté, Gabrielle Chicoine, Patricia Auger, Billy Vinette, Geneviève Rouleau, Marc-André Maheu-Cadotte, M Pagé, Judith Lapierre, Shalini Lal, Christine Genest, Guillaume Fontaine, Sylvie Cossette, Jinghui Cheng, Didier Jutras-Aswad. . . . 3024

A Personal Health App and Wearable Co-Design Framework for Rare and Complex Diseases: User-Centered, Collaborative Co-Design Study (e85425)
Sarah Goodday, Emma Karlin, Madeleine Sorensen, Robin Yang, Paul Gordon, Caresse Opoku, Daniel Vuong, Jenee Wilson, Diane McKenzie, Jules Piccotti, Massimiliano Tavanti, Megan Fitzgerald, Yochai Re'em, Hannah Wei, Megan Golden, Daniel Morgan, Michele Manion, Ricardo Mosquera, Tricha Shivas, Elise Hoover, Allison Peck, Nathan Peck, Zollie Yavarow, Heidi Bjornson-Pennell, Andra Stratton, Samali Anova Sahoo, Douglas Arbittier, Elizabeth Arbittier, Renee Goff, Donna Harvin-Graham, Nancy Howard, Camille Knudsen, Mary Oldham, Stephen Friend. .
3 0 4 6

Exploring the Design Space of Glanceable Smartwatch Feedback Displays: Experimental Study (e81972)
Yuxuan Li, Mark Newman, Predrag Klasnja. . . . 3064

Configurations of Quality Dimensions Associated With Researcher and Patient Ratings of HIV/AIDS Mobile Health Apps in China: Fuzzy-Set Qualitative Comparative Analysis (e88896)
Peng Liu, Fuzhi Wang, Changzheng Yang. . . . 3082

Privacy Policy Compliance of Mobile Sports and Health Apps in China: Scale Development, Data Analysis, and Prospects for Regulatory Reform (e73651)
Rengui Guo, Fanhong Chen. . . . 3109

Perceived Sensitivity of Sensor-Based Digital Health Data: Qualitative Interview Study (e78788)
Christine Deeney, Anika Sonig, Meghan Hurley, Birkan Tunç, Eric Storch, John Herrington, Jennifer Blumenthal-Barby, Kristin Kostick-Quenet.
3 1 3 1

Functionality Review of Mobile Apps for the Tracking and Self-Management of Fatigue: Systematic Search in App Stores and Content Analysis (e84755)
Amr Abdulla, Corina Sas, Gavin Doherty. . . . 3153

Smartphone Apps for Preventing Adolescent Health Problems Among Health Care Professionals: Systematic Search and Quality Assessment (e83908)
Elodie Million, Marijn Kuijper, Carole Bautista, Lola Pirat, Bruno Falissard, Béatrice Lognos, François Carbonnel. . . . 3177

Mobile Apps Addressing Feeding of Children With Autism: Systematic Search and Content and Quality Analysis (e94908)
Fabrice Mukarage, Christopher Cofie Kuzagbe, Skye Nandi Adams, N'guessan Yves-Roland Douha, Edith Talina Luhanga. . . . 3195

Mobile App–Supported Self-Management for Chronic Low Back Pain: Realist Evaluation (e66435)
Rebecca Hunter, Trish Gorely, Michelle Beattie. . . . 3209

XSL•FO
RenderX

Identifying and Evaluating Patient-Centered Mobile and Web Apps for Patients With Chronic Spontaneous Urticaria: Systematic Search and Content Analysis (e71435)
Caroline Glatzel, Bernadette Glatzel, Tassilo Dege, Lena Brückel, Matthias Goebeler, Astrid Schmieder. . . . 3229

Clinical Usability of Exercise Prescription Apps for Professional Use: Systematic Review and Multidimensional Evaluation (e77616)
Cheng-Hao Wu, Che-Ning Chang, Chu-Fang Chang, Ming-Hwai Lin, Hsing-Yu Chen, Yu-Chun Chen. . . . 3241

Mobile Apps for Tinnitus: Systematic Search in App Stores and Review of Intervention Components and Behavior Change Techniques (e66151)
Alina Rinn, Sarah Goetsch, Sandy Hannibal, Dirk Lehr, Cornelia Weise. . . . 3259

Patient-Centered Lupus Erythematosus Mobile Apps: Systematic Search and Cross-Sectional Evaluation by Patients and Physicians (e73019)
Tassilo Dege, Antonia Ullmann, Caroline Glatzel, Janik Fleißner, Vanessa Borst, Patrick-Pascal Strunz, Marc Schmalzing, Matthias Goebeler, Astrid Schmieder. . . . 3274

Digital Health Apps and Web-Based Platforms to Support the Prevention and Management of Snakebite Envenoming: Scoping Review (e83744)
Deborah Hosemann, Oliver Gries, Jade Rae, Thao Tran, Philipp Sprengholz, Lars Korn, Thi Pham, Thi Dang, Benno Kreuels. . . . 3284

Quality and Multifunctionality in Mobile Apps for Gestational Diabetes: Systematic App Review (e76862)
Qimeng Zhao, Alison Cooke, Lishan Huang, Yimin Tang, Dawn Dowding. . . . 3299

Mobile Apps for Oncology Health Care Professionals: Mapping and Assessment Study (e71203)
David Liñares, Iolie Nicolaidou, Andreas Charalambous, Daniela Cabutto, Deborah Moreno-Alonso, Clara Madrid Alejos, Norbert Couespel, Noemí López-Rey, María Fernández-Domínguez, Carme Carrion, Ana Clavería. . . . 3319

Improving Pre-Exposure Prophylaxis Adherence in People at Risk for HIV: Secondary Analysis of a Digital Health Intervention to Enhance User Engagement (e66837)
Lisa Hightow-Weidman, Seul-Ki Choi, Ibrahim Yigit, Crissi Rainer, Victoria Phillips, Kathryn Muessig. . . . 3346

Evidence of Efficacy of the My Personal Health Guide Mobile Phone App on Antiretroviral Therapy Adherence Among Young African American Men Who Have Sex With Men at 1 Month: Randomized Controlled Trial (e75005)
Mark Dworkin, Kara Herrera, Sierra Upton, Casey Luc, Jeb Jones, Paul Burns, Li Liu, Antonio Jimenez, Ruiqi Ren, Meaghan Woody, Robert Garofalo, Sangyoon Lee. . . . 3357

Use of Software as a Medical Device to Improve Therapeutic Adherence in Patients With Hematological Malignancies: Prospective Interventional MargheRITA Study (e59662)
Serena Dalto, Antonella Romanelli, Martina Violati, Carlotta Galeone, Davide Gaudesi, Francesca Sacchi, Massimo Beccaria, Mauro Moroni, Alessandro Ferri, Vittorio Montefusco. . . . 3374

Digital Health Support for Cataract Surgery With the Sharp Health Companion CareKit App: Randomized Controlled Trial (e78710)
Tommy Korn, Armand Assissini, Lesli Slater, Janet Holland, Lauren Doone, Elvis Lam, Joy Fox, Kristina Greenwood, Eric Linebarger. . . . 3388

Support Strategies to Enhance Adherence to a Prescription Digital Therapeutic for Erectile Dysfunction: Retrospective Quasi-Experimental Cohort Study (e76724)
Leo Dieter, Mara Haschke, Kurt Miller, Laura Wiemer. . . . 3409

Real-World Meditation App Engagement: Longitudinal Study of the Medito Meditation App (e79366)
Julia Adams, Jonathan Davies, Prai Wattanatakulchat, Julieta Galante, Simon D'Alfonso, Nicholas Van Dam. . . . 3424

XSL•FO
RenderX

Managing BMI and Emotional Distress Using mHealth: Nationally Representative Survey Study (e75499)
Ranran Mi, Fei Shen, Julia Strugala. . . . 3445

The Impact of Telehealth Use Experience on Recruitment of Underserved Populations Into a Telehealth-Delivered Mindfulness-Based Trial for Patients With Chronic Low Back Pain: Cross-Sectional Survey (e82093)
Marva Foster, Phuong Nguyen, Janice Weinberg, Keturah Faurot, Natalia Morone. . . . 3457

Unlimited Smartphone Data Plans in Older Adults With Data Deprivation: Quasi-Experimental Study (e68930)
Seoyeong Choi, Eunjeong Choi, Suk-Yong Jang. . . . 3473

Impact of Different Onboarding Strategies on Low Adoption and Engagement With a Self-Monitoring and Management App for Chronic Musculoskeletal Pain: Prospective Study (e78827)
Cinja Koller, Marc Blanchard, Johanna Mettler, Tiffany Prétat, Pedro Ming Azevedo, Thomas Hügle. . . . 3485

Effect of Teledermatologist Supervision in AI Skin Cancer Apps on Patients' Trust and Adherence Intentions: Randomized Controlled Trial (e93991)
Ritwik Swain, Mihaela Masic, Justine Staal, Natalie Dangerman, Elise Barendse, Marlies Wakkee, Laura Zwaan. . . . 3499

Insights on Postimplementation and Delivery of Telemedicine in Remote Areas of Rajasthan: Qualitative Study (e93146)
Ramesh Huda, Rahul Chowhan, Sumit Kalra, Ramesh Sangwan, Mukti Khetan. . . . 3515

Effectiveness of a Multicomponent Digital Therapeutic Program on Inhaled Medication Adherence in Chronic Obstructive Pulmonary Disease: Pilot Randomized Controlled Trial (e98870)
Hanna Wu, Fen Dong, Yanan Cui, Danjing Hu, Yushi Zhang, Tianqi Peng, Yaodie Peng, Xiang Li, Yong Li, Xiaopan Li, Yongxia Chen, Qingjie Yu, Hongtao Niu, Ting Yang. . . . 3536

Home-Based Digital Exercise Program for Patients After Open Repair of Acute Achilles Tendon Rupture: Noninferiority Randomized Controlled Trial (e78100)
Hui Wang, Tianyi Wu, Jiye He, Lihua Huang, Yanhong Ma, Manrong Xu, Shengdi Lu, Jian Zou. . . . 3553

Remote Patient Monitoring Program Components and Short-Term Hypertension Control: Retrospective Cohort Study (e69546)
Sarah LaPointe, Michael Merrill. . . . 3569

Effects of the mHealth Supportive Care Program for Family Caregivers of Individuals With Dementia and Diabetes: Pilot Randomized Controlled Trial (e72012)
Xinying Lin, Miaoqian Huang, Tuantuan Guo, Rong Lin, Chenshan Huang, Yuanjiao Yan, Hong Li. . . . 3580

Effectiveness of a Telecare-Based Intervention Program in Supporting Informal Caregivers of Community-Dwelling Older Adults in Reducing Stress Levels: Randomized Controlled Trial (e70791)
Arkers Wong, Jonathan Bayuo, Nga Ng, Matthew Yau, Ka Yu, Rose Lin, Jing Su, Vivian Hui, Jed Montayre. . . . 3600

The Role of Videoconferencing Teleconsultation in Improving Transfer Efficiency and Functional Outcomes in Rural Stroke Care: Retrospective Cohort Study (e86436)
Chi Wang, Yi-Ju Chen, Tzu-Chieh Lin, Hui-Mei Huang, Pei-Ru Tu, Po-Lin Chen, Jin-An Huang. . . . 3614

Influence of Online Sessions via a Deep Brain Stimulation Device: Prospective, Single-Arm, Longitudinal, Nonrandomized Self-Controlled Cohort Study (e80223)
Ryoma Morigaki, Hideo Mure, Kouki Takasuka, Joji Fujikawa, Toshiro Shinmen, Shin Tarui, Masataka Hatanaka, Keisuke Shinohara, Taku Matsuda, Kazuhisa Miyake, Yasushi Takagi. . . . 3625

XSL•FO
RenderX

Inclusive Contactless Monitoring for Older Adults From Diverse Backgrounds: Mixed Methods Study (e79892)
Titilola Yakubu, Nooshin Jafari, Samya Torres, Michael Lim, David Rivest-Henault, Thomas Vaughan, Catherine Proulx, Linda Pecora, Di Jiang, Kendall Ho. . . . 3640

Acceptability of Telehealth as the Default Modality for Multiple Sclerosis Care in Switzerland: Cross-Sectional Study (e84447)
Sintieh Ekongefeyin, Paola Daniore, Vasileios Nittas, Stefania Iaquinto, Enriqueta Vallejo-Yagüe, Christian Kamm, Pasquale Calabrese, Claudia Baum, Claudio Gobbi, Chiara Zecca, Andrew Chan, Milo Puhan, Viktor von Wyl. . . . 3666

A Telemedicine App for Nonrigid Facial Rehabilitation Training Enhanced by Efficient Fully Convolutional Neural Network With Residual Network (EffiFCNN-ResNet) to Improve Accessibility for Patients With Nasopharyngeal Carcinoma Cancer: Randomized Controlled Trial (e72560)
Tong Wu, Ting Han, Xiaoju Zhang, Yumei Dai, Xiaoyan Meng. . . . 3727

Tailored and Interactive Mobile Telehealth Contraceptive Counseling Compared With In-Person Care: Systematic Review and Meta-Analysis of Randomized Controlled Trials (e88887)
Maja Weinryb, Wilhelmus Grooten, Andrea Koris, Margit Endler. . . . 3753

Home-Based, Digitally Assisted Dyadic Tai Chi Intervention for Patients With Chronic Heart Failure and Their Caregivers: Development of an Intervention and Pilot Randomized Controlled Trial (e95689)
Yan Zhang, Fei Ngai, Lihua Cao, Xiaohong Lu, Wenzhong Zhang, Lili Wei, Yao Xie. . . . 3770

mHealth–Supported Perioperative Care for Family Caregivers of Children Undergoing Tonsillectomy and/or Adenoidectomy (TONAPP): Randomized Controlled Trial (e93989)
Raffaella Dobrina, Chiara De Vita, Laura Brunelli, Giulia Galvani, Margherita Dal Cin, Manuela Giangreco, Milena Ciampechini, Silvana Schreiber, Giada Ferrari, Paola Di Rocco, Sara Zaccariotto, Maria Saija, Ilaria del Giorno, Sara Zanchiello, Anja Starec, Andrea Cassone. . . . 3786

Families Moving Forward Connect mHealth Intervention for Caregivers of Children With Fetal Alcohol Spectrum Disorders: Randomized Controlled Trial (e73647)
Christie Petrenko, Madeline Rockhold, Julianne Myers, Cristiano Tapparello, Carson Kautz-Turnbull, Emily Speybroeck, Zhi Li, Lynn Cole, Heather Olson. . . . 3803

Long-Term Effectiveness of a Game-Based Mobile App for Training in Cardiopulmonary Resuscitation and Automated External Defibrillator Use: Nonrandomized Controlled Trial (e78689)
Ruting Gu, Yueshuai Pan, Jing Han, Min Wang, Xiaomin Liu, Xia Liu, Qianqian Li, Jingyuan Wang, Shanshan Ji, Changfang Shi, Haiqing Zhou, Lili Wei. . . . 3823

WhatsApp-Supported Teledentistry to Reinforce Oral Health Promotion Among Older Adults Residing in Rural and Urban Areas: Randomized Controlled Trial (e71251)
Fernanda Muñoz-Sepúlveda, Víctor Beltrán, Alfredo von Marttens, Pablo Navarro, Claudia Acevedo, Rodrigo Mariño, Alejandra Chaparro, Leonardo López. . . . 3838

Mobile-First Web Access and Captioned Video in Francophone Cardiology Education: Multicountry Ecological Learning Analytics Study (e90345)
Thomas Roussel, Elodie Benhaim, Louis Perrard, Benoit Merat, Shamir Vally, Remi Girerd, Pierre Deharo, Louis-Marie Desroche, École Numérique de Cardiologie. . . . 3864

App-Based Training Module on Guiding Physicians' Prescription for Antibiotic Treatment of Gonorrhea: Cluster Randomized Controlled Trial (e63736)
Ting-Ting Jiang, Xiao-Bin Zhang, Pei-Zhen Zhao, Li-Juan Fei, Hui-Min Chen, Yun-Qing Yang, Man-Hong Jia, Cheng Wang, Yun-Liang Shen, Yu-Jun Xu, Yan Han, Yue-Ping Yin, Xiang-Sheng Chen. . . . 3878

XSL•FO
RenderX

Real-World Functional Gait Recovery Trajectories After Total Hip and Total Knee Arthroplasty via Smartphone Sensors: Observational Longitudinal Study (e97657)
Levi Teitz, Einat Granot-Hershkovitz, Pei-An Lee, On-Yee Lo, Brad Manor, Yuval Naveh. 3890

Patient Engagement in a Virtual Cardiac Rehabilitation Platform: Secondary Analysis of the ERA-Nudge Trial (e64502)
Prerna Gupta, Josh Durfee, Joy Waughtal, Dana Fletcher, Vincenzo Polsinelli, Harsh Vathsangam, Joan Hilton, Alexis Beatty, P Ho, Larry Allen, Sheana Bull, Pamela Peterson. 3908

The Efficiency and Cost-Effectiveness of Wearable Sensors in a Digital Physiotherapeutic Total Hip Arthroplasty–Specific Training System for Patients After Total Hip Arthroplasty: Randomized Controlled Trial (e93050)
Chenyi Jiang, Yun Shen, Lihua Huang, Yanhong Ma, Shengdi Lu, Jimin Yin. 3921

Evaluation of the Impact of a Smartphone App on Adherence to an Exercise Program in People With Chronic Low Back Pain: Randomized Controlled Trial (e77736)
Jean-Baptiste Lechauve, Lech Dobija, Bruno Pereira, Maxime Grolier, Mathilde Pelletier-Visa, Charlotte Lanhers, Emmanuel Coudeyre. 3939

Smartphone Cardiac Rehabilitation, Assisted Self-Management (SCRAM) Versus Usual Care: Multicenter Randomized Controlled Trial (e66074)
Ralph Maddison, Narayan Subedi, Peixuan Li, Karen Lamb, Kylie Ball, Brian Oldenburg, Clara Chow, Sarah McNaughton, Elena George, Lan Gao, Marj Moodie, John Amerena, Voltaire Nadurata, Christopher Neil, Stuart Cameron, Jonathan Rawstorn. 3956

Effect of a Comprehensive Mobile-Based Respiratory Training Program on Respiratory Function in Survivors of Acute Stroke: Randomized Controlled Trial (e78637)
Zong-Ke Ma, Han-Hong Jiang, Yan-Hua Tang, Jia Yang, Qing-Chuan Wei, Qiang Gao. 3973

Effectiveness of a Personalized Digital Exercise and Nutrition Rehabilitation Program in Postoperative Patients With Gastric Cancer: Randomized Controlled Trial (e85795)
Inah Kim, Ji Lim, Sun Kim, Jun Lee, Tae Sohn, Sungsoo Park, Seok Kang, Ji Lee, Ji Hwang. 3994

mHealth-Supported Exercise Rehabilitation to Reverse Frailty After Autologous Transplantation in Multiple Myeloma: Randomized Controlled Trial (e87628)
Kyuwan Lee, Justin Shamunee, Haehyun Lee, Xinyi Du, Lanie Lindenfeld, Amrita Krishnan, Nitya Nathwani, F Wong, Saro Armenian. 4018

Digital Physiotherapeutic Elbow-Specific Training System for Patients After Arthroscopic Release of Elbow Contracture: Noninferiority Randomized Controlled Trial (e87459)
Yanmao Wang, Lihua Huang, Ruixin Wang, Zhihao Xu, Yanhong Ma, Shiyang Yu, Jian Ding, Yun Shen, Shengdi Lu. 4029

Effects of Telerehabilitation Based on Motion Recognition Technology on Exercise Endurance of Patients With Non–Small Cell Lung Cancer After Surgery: Single-Center, Prospective, Open-Label, Randomized Controlled Trial (e82447)
Ling Chen, Xiaoyi Xu, Xiaoliang Yang, Siying Long, Na Chen, Zifan Li, Tingting Yang, Qianru Li, Meng Li, Hai Zhong, Guozhi Huang, Qing Zeng. 4052

Smartphone-Based Interpretable Machine Learning for Classifying Single-Leg Squat Performance Using Trunk, Pelvic, and Knee Kinematics: Cross-Sectional Study (e85126)
Sihyun Kim, Kyuenam Park. 4074

Mobile App Rating Scale for Health Care Professionals to Assess the Quality of mHealth Apps: Questionnaire Development and Psychometric Analysis (e48828)
Albie Sharpe, Yiong Chan, Amarasinghe Saparamadu. 4088

The Landscape of Mobile Apps for Healthy Eating: Case Study for a Systematic Review and Quality Assessment (e68737)
Garlene Zamora Zamorano, Alejandro Déniz-García, Alezandra Torres-Castaño, María Álvarez-Malé, Inger Gram, Guri Skeie, Ana Wägner. .
4 1 0 3

XSL•FO
RenderX

Robust Assessment of Free-Living Physical Behaviors and Activity Intensity Using Dual-Wearable Multitask Learning: Development and Evaluation Study From the Multicenter WEALTH Project (e94302)
Luis Sigcha, Annika Swenne, Grainne Hayes, Jitka Kuhnova, Richard Cimler, Steriani Elavsky, Tomas Vetrovsky, Léopold Fezeu Kamedjie, Jérôme Bouchan, Jean-Michel Oppert, Janas Harrington, Greet Cardon, Antje Hebestreit, Alan Donnelly, Pepijn Van de Ven, Christoph Buck, WEALTH consortium. 4116

An Information-Driven Approach for the Early Health Technology Sustainability Assessment and the Frugal Design of the Internet of Medical Things: Exploratory Study of Wearable Activity Monitoring Devices (e88237)
Ernesto Quisbert-Trujillo, Nicolas Vuillerme. 4137

Noninvasive Interstitial Glucose Estimation Using Wearables and Machine Learning in Healthy Individuals and Individuals With Obesity: Observational Cohort Study (e91724)
Franziska Schmelter, Xinyu Huang, Annika Heidmann, Paul Beier, Friederike Rundfeldt, Christina Plöger, Artur Piet, Md Hasan, Yuanheng Zhang, Lennart Jablonski, Oliver Witt, Torsten Schröder, Stefanie Derer, Chris Easthope Awai, Marcin Grzegorzek, Christian Sina. 4159

Reducing Low Anterior Resection Syndrome After Low Rectal Cancer Surgery Using an Integrated Intra-Anal Balloon Training and Reminder Device: Feasibility Nonrandomized Controlled Trial (e91332)
Qing Zhang, Yanjun Wang, Meiling Wang, Haiyan Hu, Quan Wang, Yuchen Guo, Jianan Sun, Xuan Sun. 4177

Monitoring Respiratory Health in Children With Acute Asthma Using Wearable Electrical Bioimpedance and Breath Sounds: Observational Case-Control Study (e72979)
Jesus Sanchez-Perez, John Berkebile, Natalie Jordan, Kevin Maher, Omer Inan, Jocelyn Grunwell. 4190

Evaluation of the Importance of Stopping Elderly Accidents, Deaths, and Injuries (STEADI)–Based Factors in Wearable Fall Risk Assessment: Secondary Data Analysis (e93877)
Nozhan Ghoreishi, Stella Ansah, Jiayu Lu, Wei Lu, Sanghee Moon, Ferdinand Delgado, Diliang Chen. 4208

Determining a Likely Mechanism of Missingness in Repeated Measures Sleep Data From Wearable Fitness Trackers: Longitudinal Analysis (e81123)
Kate Mobley, Kevin Gittner, Karen Nielsen, Lauren Matheny, Gita Taasoobshirazi, Charles Natuhamya, Monica Swahn. 4226

Determining Cluster-Specific Differences in the Number of Days Required to Reliably Predict Habitual Physical Activity: Intraclass Correlation Resampling Analysis (e64323)
Conor Murphy, Gabriel Jouan, Katrin Friðgeirsdóttir, Anna Islind, Jose Saavedra, María Óskarsdóttir, Erna Arnardóttir. 4239

A Cardiovascular Care Pathway Supported by Remote Patient Management in Dutch Primary Care: Cost-Effectiveness, Budget Impact, and Workload Analysis (e79512)
Jan Heijdra Suasnabar, Margot Rakers, Nicoline van Hattem, Eric Hiddink, Marcel Haas, Xuerui Zhang, Douwe Atsma, Tobias Bonten, M van den Akker-van Marle, Hendrikus van Os. 4295

Efficacy and Safety of a Telemedicine System in Patients With Gestational Diabetes Mellitus (TELEGLAM): Single-Center, 2-Arm, Randomized, Open-Label, Parallel-Group Study (e72242)
Kazuki Aoyama, Yuya Nakajima, Shu Meguro, Yasunori Sato, Rei Goto, Mariko Hida, Takeshi Arimitsu, Yoshifumi Kasuga, Mamoru Tanaka, Hiroshi Itoh, Kaori Hayashi. 4313

Cost-Effectiveness of Internet of Things–Based Management of Home Noninvasive Positive Pressure Ventilation in Patients With Chronic Obstructive Pulmonary Disease and Hypercapnic Chronic Respiratory Failure: Trial-Based Economic Evaluation (e71340)
Ruihua Feng, Jiu Cheng, Yueying Cui, Xi Wang, Yinghao Lv, Weipeng Jiang, Chenjun Zhou, Yuanlin Song, Hui Liu. 4327

Incremental Value of Smartphone Sensing for Monitoring Momentary Affect Intensity in Adults Using Transformer-Based Models: Observational Study (e90970)
Yiqin Zhu, Yuyi Yang, Renee Thompson. 4354

XSL•FO
RenderX

Predicting Affective Episodes in Bipolar Disorder Using Statistical Process Control Analysis of GPS-Based Mobility Patterns: Quantitative Study (e77272)
Marvin Guth, Carl Bittendorf, Clemens Krug, Vera Ludwig, Esther Muehlbauer, Lisa-Marie Hartnagel, Emanuel Severus, Michael Bauer, Philipp Ritter, Ulrich Ebner-Priemer. . . . 4375

Noninvasive Electrophysiological Biomarkers of Olfactory Responses Across Cognitive States in Alzheimer Dementia: Cross-Sectional Study (e76245)
Juchan Ha, Junsoo Bok, Yunjin Lee, Yong Kim, June Kim, Hee-Jin Kim, Yongwoo Jang. . . . 4391

Passive Monitoring for Detecting Hypomanic Symptoms in Bipolar Disorder Using Explainable Machine Learning: Longitudinal Observational Study (e105647)
Ramzi Halabi, Martin Alda, Mirkamal Tolend, Claire O'Donovan, Arend Hintze, Benoit Mulsant, Abigail Ortiz. . . . 4412

Exploring Consumers' Situational Snacking Behaviors Using a Mobile App: Longitudinal Cohort Study (FOODLOOP) Among Millennials in the Netherlands (e71858)
Mariëlle de Vaal, Vincenzo Fogliano, Ruud Verkerk, Bea Steenbekkers. . . . 4435

A Smartphone-Based Ecological Momentary Intervention for Workplace Mental Health: Randomized Controlled Trial (e72435)
Sisi Li, Lik Lo, Chiu Lau, Megan Lam, Meanne Chan. . . . 4460

Identifying Symptom Dynamics and Profiles of Subgroups at Risk for Major Depression and Suicidal Ideation Among Korean Adults: 2-Week Ecological Momentary Assessment Study (e80332)
Kyungmi Chung, Jin Park. . . . 4476

Associations Between 24-Hour Physical Behavior, Self-Perceived Stress, and Coping Self-Efficacy in Everyday Life: Ambulatory Assessment Study (e81502)
Katrin Bonn, Doreen Wohlfarth, Irina Timm, Oliver Bender, Ulrich Ebner-Priemer, Marco Giurgiu. . . . 4500

Integration of Ecological Momentary Assessment and Intervention With Wearable-Based Digital Phenotypes to Treat Anxiety in Young Adults: Mixed Methods Experimental Study (e88919)
Hyunsil Song, Yoobin Choi, Jinho Park, Hyunchan Hwang, Jinwoo Kim, Chanmi Park. . . . 4516

Exploring College Students' Compliance During a 60-Day Alcohol-Related Ecological Momentary Assessment Using Location- and Time-Based Questions: Prospective Observational Study (e81713)
Lucy Napper, Bilal Khan, Gavin Fox, Mohamed Saad, Laura Wolter, Dennis McChargue. . . . 4539

Behavioral Modification as a Putative Mediator of Digital Therapeutic Response in Temporomandibular Disorders: Secondary Analysis of a Multicenter Sham-Controlled Randomized Trial (e104264)
Soo-Hwan Byun, Sung-Woon On, Byong-Eun Yang, Sung-Ah Che, Sang-Min Yi, Yongjin Park, Yeolib Kim, Sang-Yoon Park. . . . 4551

Research Letter

Alcohol Use Trajectories During the First 72 Weeks of WHOOP Wearable Platform Membership: Observational Cohort Study (e91288)
Gregory Grosicki, William Hippel, Finnbarr Fielding, Christopher Chapman, David Presby, Josh Leota, Kristen Holmes. . . . 2360

Tutorials

Designing a Self-Guided Digital Intervention for Self-Management of Shoulder Pain in People Living With Spinal Cord Injury: Tutorial on Using a Person-Based Approach (e66678)
Verna Stavric, Nicola Saywell, Nicola Kayes. . . . 4250

XSL•FO
RenderX

Intensive Longitudinal Data Collection Methods in Health Research: Tutorial for Selecting Measures and Designing a Sampling Protocol (e81290)
Jimikaye Courtney, Genevieve Dunton. 4274

Corrigenda and Addendas

Correction: Trustworthiness of Web-Based Pharmacy Apps in Pakistan Based on the Mobile App Rating Scale: Content Analysis and Quality Evaluation (e88648)
JMIR Publications Editorial Office. 4341

Correction: Real-World Functional Gait Recovery Trajectories After Total Hip and Total Knee Arthroplasty via Smartphone Sensors: Observational Longitudinal Study (e116089)
Levi Teitz, Einat Granot-Hershkovitz, Pei-An Lee, On-Yee Lo, Brad Manor, Yuval Naveh. 4343

Correction: Efficacy and Safety of a Telemedicine System in Patients With Gestational Diabetes Mellitus (TELEGLAM): Single-Center, 2-Arm, Randomized, Open-Label, Parallel-Group Study (e113548)
Kazuki Aoyama, Yuya Nakajima, Shu Meguro, Yasunori Sato, Rei Goto, Mariko Hida, Takeshi Arimitsu, Yoshifumi Kasuga, Mamoru Tanaka, Hiroshi Itoh, Kaori Hayashi. 4344

Correction: Perceptions of Adults Aged 50 Years and Older Regarding the Use of Wearable mHealth Technologies to Promote Physical Activity: Systematic Review and Meta-Ethnography (e111082)
Beatriz Rodríguez-Martín, Diego González-Arroyo, Susana Priego-Jiménez, María Lopez-González, María Medrano-Echeverría, Marta Ruiz-Grao, Celia Álvarez-Bueno. 4346

Correction: Short-Term Effects of an mHealth Intervention on Healthy Behaviors and Cardiometabolic Health in Sedentary Employees: Quasi-Experimental Study (e106593)
Yun-Ping Lin, Shu-Hua Lu, Kwo-Chen Lee, Wei-Fen Ma, Ya-Fang Ho, Wen-Chun Liao, Hui-Ting Yang, OiSaeng Hong. 4348

Addendum: Evaluating Characteristics and Quality of Mental Health Apps Available in App Stores for Indian Users: Systematic App Search and Review (e91934)
Seema Mehrotra, Ravikesh Tripathi, Pramita Sengupta, Abhishek Karishiddimath, Angelina Francis, Pratiksha Sharma, Paulomi Sudhir, T Srikanth, Girish Rao, Rajesh Sagar. 4350

Correction: Effects of Telehealth Interventions for People With Parkinson Disease: Systematic Review and Meta-Analysis of Randomized Controlled Trials (e108495)
Minyue Sun, Fuyou Tang, Luo min, Shiyu Wen, Shuang Wang, Huiping Jiang. 4352

XSL•FO
RenderX

Systematic and Collaborative Approach to Learning and Educational Content Development (SCALED) for Health Apps: An Experience-Informed Conceptual Framework

Qi Chwen Ong[1*], MBBS, MPH; Anita Pienkowska[1*], PhD; Emina Obarcanin[1], PhD; Anne-Claire Stona[2,3], MD, MSc; Andy W H Khong[1,4], PhD; Josip Car[5], MD, PhD; Andy Hau Yan Ho[1,6], PhD, EdD

[1]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[2]SingHealth Duke-NUS Global Health Institute, Duke-NUS Medical School, Singpapore, Singapore

[3]Centre for Regulatory Excellence, Duke-NUS Medical School, Singapore, Singapore

[4]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore

[5]King's Population Health Institute, King's College London, London, United Kingdom

[6]Psychology, School of Social Sciences, College of Humanities, Arts and Social Sciences, Nanyang Technological University, SHHK-04-03, 48 Nanyang Drive, Singapore, Singapore

[*]these authors contributed equally

Corresponding Author:
Andy Hau Yan Ho, PhD, EdD
Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

Abstract

Mobile health (mHealth) apps are widely used for noncommunicable disease prevention and self-management. However, their effectiveness and safety are undermined by substantial variation in content quality. Existing guiding frameworks primarily focus on user interface, functionality, and intervention delivery, with limited emphasis on content creation. In this viewpoint, we introduce Systematic and Collaborative Approach to Learning and Educational Content Development (SCALED), a conceptual framework designed to guide a systematic, collaborative, and evidence-based approach to mHealth educational content development, intended for app developers, researchers, health care providers, and the wider mHealth community. Developed and refined across 3 phases, the SCALED framework consists of 11 components organized into 3 sequential stages: planning and conceptualization, development of textual content, and finalization into delivery format. We discuss the rationale behind each component and illustrate its applicability through 2 mHealth use cases. The framework integrates real-world experience from the development of 3 mHealth apps, qualitative findings from 2 studies, and insights from key stakeholders. By offering a structured and replicable methodology for content development, SCALED addresses a critical gap in current mHealth frameworks and provides a practical guide to improve content veracity, with potential for adaptation across a range of medical conditions.

(JMIR Mhealth Uhealth 2026;14:e83442) doi:10.2196/83442

KEYWORDS

patient education; mHealth; mobile health; health information; conceptual framework; health apps

Introduction

Mobile health (mHealth) apps have emerged as a promising tool for disease prevention and management, providing accessible, scalable, and personalized health care interventions. Defined by the World Health Organization as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices" [1], mHealth apps are widely tested for patient education, remote monitoring, behavioral interventions, and personalized care [2].

As of 2024, it was estimated that more than 35,000 mHealth apps were available on both the Apple App Store and the Google Play Store [3,4]. Despite their ubiquity and growing adoption, the effect of these apps across various health outcomes remains inconsistent [2,5,6]. Evidence from systematic reviews indicates that most mHealth apps lacked clinical credibility [7], were developed without health care professionals' (HCPs) involvement [8], and did not report theory or conceptual framework underpinning the development of educational materials [9]. Moreover, a scoping review found that 67 of 80 identified safety concerns with consumer-facing mHealth apps were related to the information quality, such as incorrect or incomplete information, variation in content quality, and inappropriate response to consumer needs [10].

These gaps could have a negative impact on their uptake, posing grave concerns for patient safety. Without adequate quality control measures and regulation for apps entering the market

XSL•FO
RenderX

[11], health care providers and organizations may become more cautious and reserved about prescribing or recommending mHealth apps to patients [12]. From the user's perspective, exposure to low-quality or misleading content may result in inappropriate lifestyle choices or self-management decisions, introducing potential clinical harms [13]. Given the dynamic and rapidly expanding mHealth landscape, it is imperative to establish and promote good practices in the design and development of educational content to ensure its effectiveness, safety, and trustworthiness.

mHealth Frameworks and Instruments

A growing number of frameworks have been proposed for mHealth apps. These frameworks primarily focus on aspects related to technical and functional design [14,15], usability and user-centeredness [16,17], as well as risk and trustworthiness [13,18]. In parallel, many instruments and tools developed for mHealth apps are intended for quality assessment and rating purposes [19,20]. Notable examples include the Mobile App Rating Scale (MARS) [21], which is a multidimensional instrument for assessing app quality across domains such as engagement, functionality, aesthetics, and information quality, and the THESIS tool, which was developed to assess the risks and benefits of health apps across a broader set of dimensions [22]. Although most included criteria are related to the veracity of content, such considerations are often addressed in a generic sense only [19]. A systematic review of mHealth assessment frameworks synthesized evaluation criteria into 7 overarching categories, one of which pertains to information and content. Key elements in this category are credibility, accuracy, quality, and quantity of information [20]. This reflects the recognition of content quality, alongside technical design, usability, and data privacy, as a key factor in how mHealth apps are evaluated and ultimately used.

The importance of content veracity in mHealth has also been formally acknowledged. The CONSORT (Consolidated Standards of Reporting Trials)-eHealth checklist, which extends reporting standards for randomized controlled trials of web-based and mHealth interventions, explicitly recommends reporting on quality assurance measures taken to ensure accuracy and quality of information in eHealth interventions [23]. While tools for evaluating readability and quality of consumer-facing health information such as Flesch-Kincaid and DISCERN do exist [24,25], there remains a critical gap: the lack of standardized methodologies for developing mHealth educational materials that are evidence-based, clinically accurate, and theory-informed. In response, some articles have outlined multistage content development processes involving initial drafting, multiple rounds of edits, professional editing, and final reviews by clinical experts [26,27]. These offer useful starting points for shaping best practices in mHealth content development, but fall short of offering a structured, stepwise approach.

To address this gap, we propose a Systematic and Collaborative Approach to Learning and Educational Content Development (SCALED), a framework informed by real-world experience in developing 3 mHealth apps, qualitative insights from end-users, and iterative input from a multidisciplinary team. By integrating educational and behavioral change theories, clinical guidelines, and user-centered design principles, SCALED aims to provide a comprehensive guide for improving content credibility and accessibility. In this viewpoint, we describe the framework's conceptual foundation, its application in 3 use cases, and its implications for mHealth research and practice.

The SCALED Framework

Conceptualization of the Framework

SCALED was conceptualized and developed in 3 phases. First, an initial framework informed by the development of a comprehensive mHealth education program was evaluated and improved in a co-design study with patients (GLOW) [28]. Second, application in 2 use cases (WellFeet and CADENCE D-PHA) to demonstrate its usability in practice and refine for improvement [29,30]. Third, finalization of the framework through discussion and consensus building with stakeholders involved in content creation. A glossary of key terms is provided in Textbox 1.

Textbox 1. Glossary of Terms.

Theme: broad, overarching topic that serves as a unifying thread for organizing curriculum content and learning experiences.

Module: self-contained units of educational content designed to address specific learning objectives within a theme.

Bite-size: small, manageable portions of content intended to facilitate focused learning and reduce cognitive load.

Conversational agent (chatbot): "Computer programs that use text, speech, and other input modalities to enable communication with users" [14].

Action-oriented: learning or intervention content designed to prompt users to perform specific behaviors, tasks, or decision-making steps that reinforce learning objectives.

Research Context

The experience of developing a diabetes education app (GLOW, Nanyang Technological University Singapore) and findings from a qualitative co-design study with patients with lived experiences of diabetes guided the conceptualization of the initial mHealth content development framework [28]. GLOW was designed to provide users with verified health information on general knowledge related to Type 2 diabetes mellitus, including complications, foot health, medication, and blood glucose monitoring, and lifestyle behaviors including exercise, diet, and sleep. It consisted of 10 modules, comprising 63 units with learning cards, quizzes, mini-challenges, and rules-based action-oriented, frequently asked questions (FAQ) chatbot conversations. The content was guided by 172 knowledge-based, 54 skills-based, and 106 attitude-based learning objectives.

XSL•FO
RenderX

Additionally, 232 audio recordings of the text, 711 static images, and 55 motion graphics were integrated to complement the textual content of the modules. Although patients were not directly involved, GLOW app content was co-created with lay professionals. These educational modules underwent review by HCPs, including endocrinologists, nurses, podiatrists, and dieticians, from a major tertiary hospital in Singapore. This process served as the basis for the initial structure of the SCALED framework.

We subsequently applied the initial framework to two use cases: (1) WellFeet, a multilingual diabetes foot care education app for individuals at risk of diabetic foot ulcers [29], and (2) CADENCE D-PHA (digital personal health assistant) app, a multicomponent mobile health app for improving medication adherence in adults with hyperlipidemia [30].

WellFeet content consists of 4 themes, 21 topics, and 54 bite-sized videos in English, Chinese, and Malay, accompanied by quiz questions and rules-based chatbot conversations. They were guided by 21 general, 54 knowledge-based, 40 skill-based, and 38 attitude-based learning objectives. Bite-sized animations with a voice-over and a synthetically generated artificial intelligence (AI) avatar-presenter were accompanied by quiz questions, polls, reflections, and rules-based action-oriented FAQ chatbot conversations framed by a personalized and adaptive learning experience. The content was reviewed by project co-investigators and HCPs from multiple major health care institutions in Singapore. The app was tested in a single-arm, pre-post mixed-methods feasibility study [29].

CADENCE D-PHA's content (phase 1) comprises 5 themes, 13 topics, and 63 bite-sized videos accompanied by quiz questions and rules-based chatbot conversations. They were guided by 40 knowledge-based, 35 skill-based, and 20 attitude-based learning objectives. Bite-sized videos are animations with a voice-over and video recordings of HCPs and patients, accompanied by quiz questions, polls, free-text responses, and reflections. Similar to WellFeet, the education content is supported by rules-based, action-oriented FAQ chatbot conversations. Moreover, the content was reviewed by project co-investigators, collaborating HCPs from multiple tertiary and primary health care institutions in Singapore, as well as patient advocates. Phase 2 content is being developed and progressively rolled out with app updates. CADENCE D-PHA is currently being tested in an ongoing multicenter randomized controlled trial.

The framework was modified to meet the needs of each individual project, with components added or omitted as appropriate, considering factors such as target user groups, team composition, technological infrastructure, available funding, and project timeline. Applying the initial framework to these 2 use cases allowed us to assess the framework's overall applicability and refine its components. Across the 3 use cases, several cross-cutting insights emerged that informed the refinement of the SCALED framework. First, the nature of deliverables differed across projects, ranging from text-based modules to multilingual videos and interactive chatbot features. Second, variations in available resources, including team composition and funding, shaped how specific components were operationalized, particularly in decisions related to outsourcing multimedia production. Third, differing technological requirements, such as the integration of conversational agents, multimedia elements, and adaptive features, introduced additional complexity in translating textual content into user-facing formats. Despite these differences, early-stage components such as team composition, theoretical grounding, and curriculum definition remained consistently critical, while challenges frequently arose at the interface between content creation and delivery. These cross-case insights highlight the importance of a structured yet adaptable framework and inform the final configuration of the SCALED components.

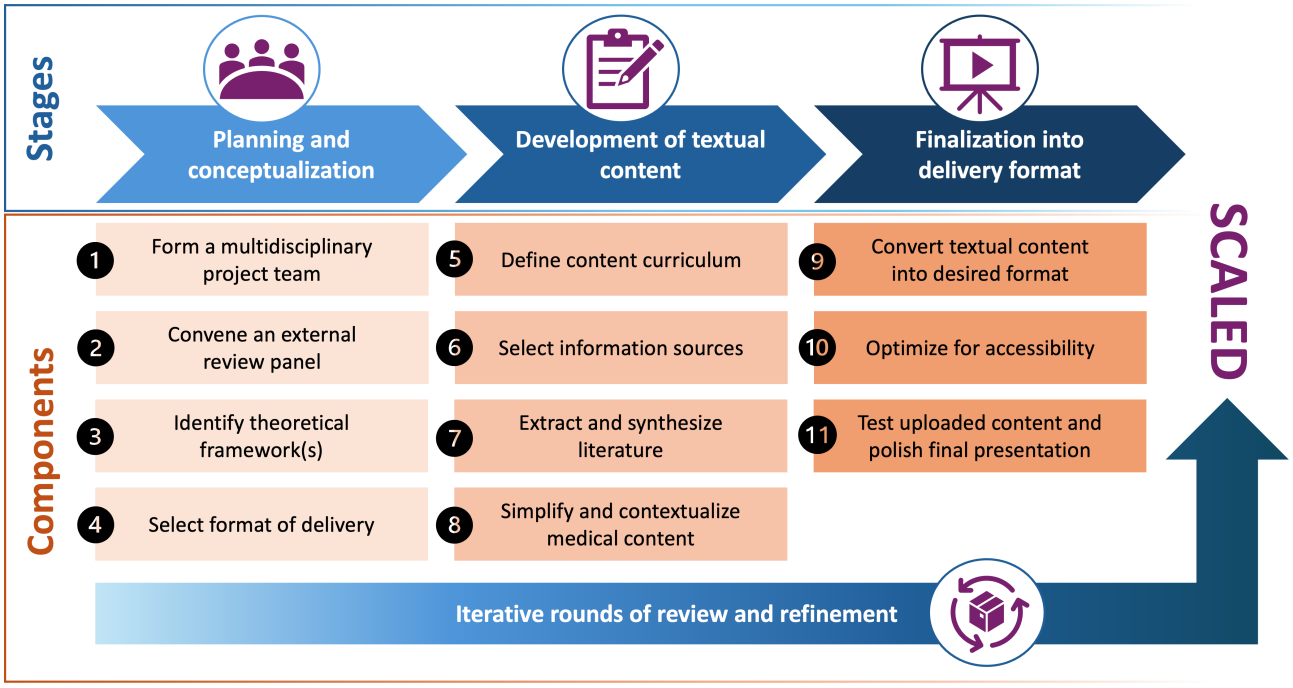
SCALED Pillars

Drawing on real-world experience of developing 3 mHealth apps, 2 studies with end-users, and iterative discussions within the research team, the final SCALED framework (Figure 1) consists of eleven components embedded within 3 sequential stages. The framework moves from planning and conceptualization, to the development of textual content, and finally to its translation into delivery formats, reflecting the practical workflow observed across our 3 app development cases. The first stage establishes the conceptual, theoretical, and collaborative foundations of the intervention. The second focuses on generating, synthesizing, and tailoring the educational content, whereas the third addresses its conversion into accessible, polished, and implementation-ready materials.

The eleven components within SCALED were core recurring needs identified from our empirical experience, and they do not constitute an exhaustive taxonomy of all conceivable content-development activities. Although presented as a structured sequence, the components are not intended to function as rigid requirements to be applied identically in every context. Instead, they are core yet adaptable elements that may be modified, emphasized, or omitted depending on the intervention aims, target users, technological infrastructure, and available resources. Together, these components provide a structured pathway for developing mHealth educational content from initial conception to final deployment.

XSL•FO
RenderX

Figure 1. The Systematic and Collaborative Approach to Learning and Educational Content Development (SCALED) framework.



Planning and Conceptualization

Form a Multidisciplinary Project Team

The creation of content in mHealth apps necessitates a multidisciplinary team to ensure clinical accuracy, behavioral and cultural relevance, and user-centered delivery. Team members from diverse professional backgrounds often bring differing expectations around project management, role responsibilities, content delivery, and even use distinct "language" (using different terminologies and methodologies), which can pose communication challenges [31]. However, the benefits of this diversity offset the potential difficulties, as varied backgrounds enable team members to complement one another's knowledge and skills, ultimately enhancing the quality of the project.

HCPs contribute clinical expertise to ensure content alignment with clinical practice guidelines. Their involvement in mHealth development has been recommended to mitigate safety concerns [10]. Psychologists and behavioral scientists incorporate principles of behavior change to enhance the chances of healthy lifestyle modification. Learning designers provide input on content structure, educational framework, and delivery format to improve comprehension and accessibility. Although clearly defined roles should be established prior to initiating content development, in practice, many tasks require a flexible approach and a diverse set of competencies. For instance, content writers will often have responsibilities beyond script writing, such as adapting materials into animation-ready formats and coordinating the logistical aspects of video production. Depending on the scope and organizational capacity, specialized expertise may be acquired through external collaborators such as animators for creating visually engaging educational materials.

In all 3 cases, textual content was produced by a multidisciplinary team with capabilities to assess clinical information, conceptualize copywriting approaches, write relatable content, and perform project management duties. Given the multinational composition of the team, we made a deliberate effort to ensure our patient-centered message considered a balanced cultural power dynamic, avoiding dominance of any single cultural viewpoint. Multimedia components (eg, static graphics, animations, and short video recordings) were produced in-house by the project team in GLOW, drawing on existing expertise within the team and allowing close alignment with the pedagogical objectives. In contrast, for WellFeet and CADENCE D-PHA, a multimedia production unit was contracted to ensure professional quality, specialized technical skills, and efficient turnaround. The choice between in-house development and external contracting was guided by the availability of internal expertise, the complexity of the materials required, and project resources.

Convene an External Review Panel

It is important to engage stakeholders early in the process [32]. An external expert review panel is essential to ensure that consumer-facing health education materials are medically accurate, up-to-date, aligned with local (or regional) clinical practice and relevant to patients. In addition to the development phase, the review process should be repeated periodically as guidelines and patient needs evolve. The panel should include not just physicians, but also HCPs who are involved in the continuum of care, such as pharmacists, nurses, health psychologists, dietitians, and exercise physiologists. This allows every aspect of (chronic) disease management to be comprehensively reviewed from different perspectives.

It is equally important to actively involve a patient representative or advocate to ensure user-centered design [33]. Participatory action research (PAR) together with patient and public involvement (PPI) are increasingly recognized for their role in enhancing the quality and appropriateness of research [34]. In the context of mHealth, PAR and PPI could help identify

XSL•FO
RenderX

end-user priorities, select pertinent health education topics, ensure linguistic and cultural appropriateness, and inform user-centered delivery formats. It is also recommended to involve individuals with lived experience of the condition to strengthen the impact of PAR and PPI [35].

In all 3 use cases, the content was reviewed by collaborators from partner health care institutions. For CADENCE D-PHA, we established a formal review panel independent of the content-writing team. It comprised primary care physicians, cardiologists, clinical pharmacists, health psychologists, dietitians, and patient advocates with lived experience of hyperlipidemia and cardiovascular disease (CVD). In the case of GLOW, we conducted a co-design study with a group of patients and caregivers with diverse socioeconomic and cultural backgrounds.

Identify Theoretical Framework

Establishing the theoretical underpinnings of an intervention is a fundamental step in the development of complex programs [32]. Despite this, most mHealth apps lacked a theoretical grounding in their development, including an educational component [8,9] and behavior changes. Identifying and selecting appropriate behavioral theory is essential for supporting the adoption of healthy behaviors, while educational frameworks guide a structured and outcomes-driven approach to learning. With a plethora of behavioral theories and health promotion strategies [36,37], selecting those most relevant to the intervention's goals and the target population's needs is critical to maximizing the effectiveness of mHealth interventions.

In our use cases, we selected the health belief model and the transtheoretical model because both were well-aligned with the educational and behavioral aims of the interventions [38,39]. The health belief model was useful for structuring content around risk perception, perceived benefits and barriers, and cues to action, whereas the Transtheoretical Model supported the stepwise tailoring of content to users' readiness for behavior change. For instance, themes and topics in CADENCE D-PHA were systematically organized according to the core constructs of the health belief model. The content was designed to address the increased risk of developing CVD among patients with hyperlipidemia (perceived susceptibility), the impact of untreated hyperlipidemia and CVD on quality of life (perceived severity), misbeliefs and misconceptions about adverse effects of statins (perceived barriers), and the benefits of statin therapy in CVD prevention (perceived benefits).

However, theory selection within SCALED is not intended to be prescriptive. Rather, the framework encourages developers to identify and justify theoretical approaches that best fit the intervention goals, target behaviors, and characteristics of the intended user group. The theories used in our projects are therefore illustrative examples of how behavioral and educational frameworks can be applied within SCALED, rather than mandatory theoretical foundations of the framework.

Select Format of Delivery

The choice of instructional design and delivery format can influence user outcomes by engaging different learning styles and improving memory retention [40-42]. Compared to static images and text, animations have been shown to enhance procedural skills acquisition [43], with a positive impact on knowledge, attitudes, and cognitive engagement [44]. Conversational agents are also emerging as an innovative tool for patient education [14]. Nonetheless, the selection of delivery formats must balance the intervention's goals with practical considerations, such as funding availability and access to skilled, experienced personnel.

Across our use cases, we progressively expanded the range of formats to support engagement and learning. Beginning with text, audio, and static images in GLOW, we subsequently incorporated animations in WellFeet. These animated videos were accompanied by an AI-generated avatar-presenter that delivered the scripts. In CADENCE D-PHA, we further included transcribed videos featuring HCPs and patients, as well as audio recordings for mindfulness and meditation practice. Conversational agents in the form of rules-based chatbots remained largely consistent across 3 mHealth apps, with adaptations in their placement within the overall learning sequence to suit the specific context of each intervention. The chatbots were designed to engage users in action-oriented conversations with branching logics, enabling some degrees of personalization of content based on users' responses.

Development of Textual Content

Define Content Curriculum

Learning objectives serve as a compass for educational content developers [45]. As content creation is iterative and often collaborative, a clearly defined curriculum acts as a blueprint to ensure logical flow, alignment with intervention goals, and consistency across contributors. Topics identified through clinical expertise and literature review can be organized into thematic groups to shape the learning outcomes that reflect the overarching goals of the intervention. Learning objectives and topics can be evaluated for completeness and quality using tools such as the DISCERN instrument [25]. Feedback from an external review panel further allows content writers to prioritize practical, relevant content and remove concepts that are overly theoretical or clinically complex.

In all 3 cases, we used Bloom's taxonomy to structure learning progression and scaffold educational content by organizing it across two key dimensions: cognitive complexity and behavioral focus [46]. This involved guiding users from foundational knowledge acquisition (eg, recalling the symptoms of diabetes) to higher-order thinking skills such as creating and evaluating action-oriented goals. Complementing this, we adapted the competency-based education model to further categorize learning outcomes into knowledge, skills, and attitudes, ensuring a competency-driven approach with a strong emphasis on building efficacy in self-care skills and behaviors [47,48]. This scaffolding informed the development of learning objectives and encouraged deeper reflection on the complexity of targeted behaviors and their component tasks. It also supported more realistic decision-making about the cognitive and behavioral demands placed on users, ensuring an appropriate balance between expectation, ambition, and feasibility.

XSL•FO
RenderX

Select Information Sources

An evidence-based approach to content creation is critical to ensuring users receive health information that is up-to-date and clinically accurate. In line with principles of evidence-based medicine [49], information sources should be systematically identified and appraised. Primary sources should include an informed integration of guidelines from international or regional professional bodies, health agencies, and local clinical practice protocols. This is because although widely accepted international guidelines are useful, they might not serve the needs of certain populations or align with local standards. For instance, the recommendations for management of hyperlipidemia are known to differ between guidelines from Singapore's Agency for Care Effectiveness [48], the European Society of Cardiology [50], and the American Heart Association/American College of Cardiology [51]. Research publications, such as randomized controlled trials and systematic reviews, should be referenced with caution as individual studies may yield conflicting conclusions, vary in methodological robustness, and lack clear recommendations for real-world patient care [52,53]. Local clinical practice protocol that integrates evidence from high-quality studies with expert consensus could render them more appropriate to inform patient education. If additional information on resources such as community care support and services is needed [54], websites of health authorities, organizations, and institutions such as the World Health Organization, the UK's National Health Service, and the Centers for Disease Control and Prevention could be invaluable [48,55,56]. Once compiled, these guidelines serve as a repository to inform the entire content writing process. For each project, we systematically documented all sources used in content development, both for internal transparency and to provide users with credible references for further reading. This practice also aligns with the requirements of Google Play, which mandates the inclusion of references prior to publishing mHealth apps [57].

Extract and Synthesize Medical Literature

After creating a repository of information sources, the next step involves extracting information for the proposed topics and distilling key findings. Synthesis goes beyond merely compiling information. Given the variability in clinical guidelines and emerging evidence, it is essential to critically appraise recommendations and study findings, reconciling differences when necessary. It also involves integrating insights across multiple sources to present a balanced, contextually relevant perspective. This ensures that content remains aligned with best practices while being adaptable to specific context.

To minimize discrepancies among content writers and ensure consistency of outputs, we developed a structured template for content extraction and synthesis. This template was applied uniformly across all topics, detailing the topic description, learning outcomes, sources of evidence, raw information extracted, lay-adapted content, and the content review process. This standardizes the internal process of how and from where clinical recommendations, evidence summaries, and practical applications are documented, ensuring uniformity in tone, depth, and language. Apart from strengthening the overall coherence, the template also streamlines the review process, enabling the review panel to identify any misinterpretation of clinical recommendations by comparing the lay-adapted content with the original references and raw information extracted.

Simplify and Contextualize Medical Content

Translating complex medical information into accessible and relatable language is essential for effective health promotion [58]. Achieving this requires both simplification and contextualization. Simplification improves readability, which is a key determinant of health information accessibility, especially for people with low health literacy [59]. However, research indicates that the content of many mHealth apps was written at reading-grade levels that exceed recommended standards [60,61].

Adapting medical content into lay language involves not only enhancing readability by replacing medical jargon with simpler terms or simplifying sentence structure [59,62], but also adopting strategies to improve understandability and actionability [63], minimize linguistic stigma [64], and improve cultural sensitivity through enhancement of appropriateness and representativeness [65,66]. For example, recommendations such as the Mediterranean diet or olive oil use may hold little relevance for patients in Asia, as they do not align with local dietary norms.

For all use cases, we developed a copywriting guide with suggestions for simplified, actionable, and empathetic language. For WellFeet and CADENCE D-PHA, we adapted content into a conversational tone to suit the video scripts. We also used copywriting techniques, such as analogies and metaphors, to clarify complex ideas. When medical terms were necessary, they were supported by a built-in glossary, allowing users to click on highlighted terms for plain-language definitions. In CADENCE D-PHA, we also provided a short video introducing and explaining medical terminologies that were commonly used, such as low-density lipoprotein cholesterol, unsaturated fats, statins, and cardiovascular disease.

Finalization Into Delivery Format

Convert Textual Content into Desired Format

Design principles differ across instructional formats, with videos requiring their own specific approach [67]. Adapting text-based content for video involves more than converting it into a script. It also calls for additional guidance and materials, especially since animators usually do not have a medical background and may need support in visualizing some of the medical concepts accurately. An iterative process of providing input, reviewing storyboards, and refining video drafts is essential to ensure clarity, accuracy, and effective communication.

For both WellFeet and CADENCE D-PHA, we collaborated closely with the appointed multimedia units, not only to guide the creative process in choosing color schemes, typography, and voice-overs, but also to ensure the visual content accurately reflected the intended message. Providing carefully selected reference images was essential for ensuring accurate depiction of anatomical structures and biological processes, such as cholesterol deposition and plaque formation in arteries. All

XSL•FO
RenderX

videos underwent additional rounds of review by the content development team to verify accuracy, ensure consistency, and maintain alignment with the educational goals before being finalized and uploaded. After receiving the storyboards from animators, we optimized the characters and images. The attire, skin tone, and facial expression of animated characters were revised to better align with local context and reflect the diversity of the local population. By incorporating these considerations, mHealth apps could deliver inclusive, equitable, and accessible content to a broad range of users.

Optimize for Accessibility

Optimizing mHealth app content for better accessibility ensures that users with varying levels of digital literacy can effectively engage with the materials. Beyond app functionalities such as implementing adjustable font sizes or using a larger font as the default [68], content designers should consider including alternative text descriptions for images, high-contrast color schemes, and scalable fonts to accommodate users with visual impairments. For quizzes with text response in CADENCE D-PHA, we integrated a dictation tool to enable voice input, reducing the reliance on typing for individuals with fine motor impairments, visual disabilities, or cognitive difficulties such as dyslexia. For videos, we provided subtitles and transcripts to support individuals with hearing difficulties or impairments. This could also cater to users of a reading-based learning style. Voice-over texts were carefully refined for appropriate speed, tone, and accent suited for the target population.

Test Uploaded Content and Polish Final Presentation

Pilot testing of the final content is essential to optimize content delivery and ensure fitting with user interface before rolling out the mHealth app. The layout of content can impact visual perception and usability, often requiring refinement due to misalignment between content design and software development, which are frequently managed by separate teams. Factors such as user settings (eg, dark vs light themes and adjustable font sizes) can further influence how content is displayed and received. At this stage, established app evaluation and usability frameworks may also be used as complementary quality-assurance tools to assess broader aspects of app usability and overall quality beyond content development alone. In all 3 use cases, we observed such misalignments and addressed them during the refinement phase. For seamless presentation and usability, pilot testing of the app and its content should be conducted with end users and other stakeholders. Usability testing should gather feedback on navigation, content comprehension, visual compatibility, and overall user satisfaction. At this stage, established app evaluation and usability frameworks may also be used as complementary quality-assurance tools to assess broader aspects of app usability and overall quality beyond content development alone. Importantly, this feedback loop must be maintained continuously throughout the content development process to identify and address issues arising at every stage, ensuring that the final product meets user needs and delivers a polished, engaging experience.

Discussion

SCALED is a conceptual framework that integrates insights from real-world experience of developing mHealth apps, findings from 2 studies with end-users, and adaptations based on stakeholders' input. With eleven components grouped into 3 sequential stages, the framework offers a practical guide for mHealth content development. Unlike existing frameworks that primarily focus on app usability and intervention design, SCALED provides a structured approach to producing mHealth content that is evidence-based, clinically accurate, theoretically grounded, and user-centered. This is particularly important given that many mHealth apps currently on the market lack health care professional involvement and fail to integrate behavioral science principles, potentially limiting their effectiveness in promoting sustained behavior change.

Concerns about the content quality and safety of mHealth apps are not new. Previous research has attempted to overcome this challenge by proposing various tools and frameworks to establish standards for evaluating the content, effectiveness, and usability of mHealth apps [69]. A range of frameworks has also emerged to guide specific aspects of mHealth, such as smartphone-delivered conversational agents [14], behavior intervention technology design [70], user-centered technology design [17,33], and data security and privacy [71,72]. In parallel, broader frameworks such as the Medical Research Council framework and the Behavioral Intervention Technology model provide high-level guidance on the development and evaluation of complex interventions and on aligning clinical aims with behavior change strategies and technological features [32,70]. SCALED is intended to complement, rather than supersede, these frameworks by addressing a more specific but underdeveloped aspect of mHealth intervention design: the systematic development of educational content. This directly tackles the root cause of variability in content quality.

SCALED may have broad implications for stakeholders in mHealth apps. For developers, it provides a structured methodology to create clinically accurate, theoretically grounded, and user-centered educational content, which could help build user trust and support regulatory compliance [73]. This could be highly relevant in commercial settings, where a multidisciplinary team of HCPs might not be readily available within the organizations. In such cases, SCALED's recommendation to engage an external review panel could ensure content credibility and alignment with clinical standards. Users could benefit from more reliable, accessible, and evidence-based health information, enhancing their health literacy and informed decision-making [74]. Health care providers could use SCALED as a reference to appraise the rigor of the mHealth content development process and to assess their appropriateness before recommending them to patients. Additionally, for health systems actors and policymakers, the framework could serve as a foundation for establishing quality standards and guiding endorsement of digital health tools. By promoting a more consistent and transparent approach to content development, SCALED may ultimately support improvements in the effectiveness, credibility, and uptake of mHealth apps across different levels of the digital health ecosystem.

XSL•FO
RenderX

A key strength of the SCALED framework is its empirical foundation, having been refined based on real-world lessons from 3 mHealth apps, making it more practical and applicable than purely theoretical frameworks. Additionally, SCALED benefited from iterative feedback loops involving key stakeholders, including patients, HCPs, and content designers. Furthermore, its high-level structure provides a flexible foundation, allowing room for modification and customization of the more intricate steps in developing content across different formats and disease contexts.

A few limitations should be acknowledged. One major limitation is its untested generalizability, as the framework is based on a limited number of education-focused mHealth apps targeting chronic disease management, which may limit its direct applicability to other contexts. For instance, interventions in acute care settings, mental health, or those incorporating highly adaptive or real-time decision-making may require additional considerations beyond those captured in SCALED. Similarly, while the framework is designed to be flexible, its implementation in low-resource settings or by commercial developers without access to multidisciplinary academic teams may present practical challenges. SCALED is currently most directly applicable to interventions with a strong educational component, particularly those involving structured or rule-based content and human-reviewed materials. Its applicability to systems that rely heavily on generative AI may be limited, thus requiring further adaptation to address issues such as dynamic content generation, validation, and real-time personalization.

Another limitation of the framework is that it does not take into account regulatory checkpoints or alignments, as digital health policies vary globally. While SCALED does not address formal requirements set by regulatory bodies such as the United States Food and Drug Administration and CE (Conformité Européene) marking processes in Europe, it may serve as a complementary, preregulatory quality assurance approach by supporting the development of evidence-based and transparent educational content. It is worth noting that the framework was designed to serve as an initial step in response to the urgent need for improved quality of content in mHealth apps. Future iterations could be strengthened by having stronger engagement of target user groups and incorporating findings from systematic reviews, qualitative research, and structured stakeholder consensus methods, such as a Delphi study, to enhance framework comprehensiveness and applicability.

Conclusions

The SCALED framework offers a systematic, collaborative, and evidence-based approach to mHealth content development, addressing a critical gap in existing mHealth offerings. By ensuring alignment with clinical guidelines, theoretical frameworks, behavioral science principles, and user-centered design, SCALED demonstrates a practical and stepwise approach to improving the quality of educational content and safety of mHealth apps. Future work may focus on assessing its usability across diverse development settings, evaluating its influence on content quality and user learning, and exploring its potential for adaptation in different health contexts and digital ecosystems.

Acknowledgments

The authors specifically thank Dr Daniel Ek Kwang Chew, Dr Huiling Liew, Prof Bernhard Otto Boehm, Dr Daniel Muhammad Mahadzir, Dr Victor Jin Xue Goh, Ms Sui Jing Lim, and Mr Josh Jialong Quah for their contributions to the Glow and WellFeet app content development process. The authors are grateful to all members of the content review panels, content writers, animators, and project team members of GLOW, WellFeet, and CADENCE D-PHA apps, as well as patients who participated in the qualitative studies. We used a generative artificial intelligence (AI) tool (ChatGPT-5.2; OpenAI) solely for language refinement and grammar checks to improve the clarity and readability of the manuscript. No generative AI tools were used to generate any scientific content or figures.

Funding

This article is based on work supported by the Joint Platform 3 under the Cardiovascular Disease National Collaborative Enterprise (CADENCE) National Clinical Translational Program (MOH-001277-01). Development of WellFeet was supported by the Singapore Ministry of Health's National Medical Research Council under Health Services Research Grant (HSRG-DB17Nov002). Development of GLOW was supported by the Singapore Ministry of Health (MOH/NIC/CDM2/2018). The funders had no role in the conceptualization, development, or writing of this viewpoint.

Data Availability

This manuscript is a viewpoint, and thus no primary data were used.

Authors' Contributions

QCO, AP, EO, and AHYH conceived this viewpoint. QCO, AP, EO, AS, JC, and AHYH contributed to the methodology. All authors contributed to project administration. QCO and AP wrote the first draft of the manuscript. EO, AS, AWHK, JC, and AHYH critically reviewed and revised the manuscript for important intellectual content. AWHK, JC, and AHYH supervised this work. All authors approved the final manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Ryu S. Book Review: mHealth: New Horizons for Health through Mobile Technologies: Based on the Findings of the Second Global Survey on eHealth (Global Observatory for eHealth Series, Volume 3). Healthc Inform Res 2012;18(3):231. [doi: 10.4258/hir.2012.18.3.231]
2. Sun S, Simonsson O, McGarvey S, Torous J, Goldberg SB. Mobile phone interventions to improve health outcomes among patients with chronic diseases: an umbrella review and evidence synthesis from 34 meta-analyses. Lancet Digit Health 2024 Nov;6(11):e857-e870. [doi: 10.1016/S2589-7500(24)00119-5] [Medline: 39332937]
3. Number of mHealth apps available in the Google Play Store from 1st quarter 2015 to 2nd quarter 2024. Statista. 2024. URL: https://www.statista.com/statistics/779919/health-apps-available-google-play-worldwide/?srsltid=AfmBOopDdfGMJSd0IEaFrSr2qvDsA9oZMnU7GKH105-xj6Hmv4VSBkL3 [accessed 2026-03-10]
4. Number of mHealth apps available in the Apple App Store from 1st quarter 2015 to 2nd quarter 2024. Statista. 2024. URL: https://www.statista.com/statistics/779910/health-apps-available-ios-worldwide/?srsltid=AfmBOorTWm2KO2THUa-koNNNPu0breco4bzKbVl75D-QzYKzhJbsuNip [accessed 2026-03-10]
5. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019;2(1):118. [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
6. Iribarren SJ, Akande TO, Kamp KJ, Barry D, Kader YG, Suelzer E. Effectiveness of mobile apps to promote health and manage disease: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2021 Jan 11;9(1):e21563. [doi: 10.2196/21563] [Medline: 33427672]
7. Ali S, Alizai H, Hagos DJ, et al. mHealth apps for dementia, Alzheimer disease, and other neurocognitive disorders: systematic search and environmental scan. JMIR Mhealth Uhealth 2024 Jul 3;12:e50186. [doi: 10.2196/50186] [Medline: 38959029]
8. Alessa T, Hawley MS, Hock ES, de Witte L. Smartphone apps to support self-management of hypertension: review and content analysis. JMIR Mhealth Uhealth 2019 May 28;7(5):e13645. [doi: 10.2196/13645] [Medline: 31140434]
9. Bashi N, Fatehi F, Fallah M, Walters D, Karunanithi M. Self-management education through mhealth: review of strategies and structures. JMIR Mhealth Uhealth 2018 Oct 19;6(10):e10771. [doi: 10.2196/10771] [Medline: 30341042]
10. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020 Feb 1;27(2):330-340. [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
11. Shuren J, Patel B, Gottlieb S. FDA regulation of mobile medical apps. JAMA 2018 Jul 24;320(4):337-338. [doi: 10.1001/jama.2018.8832] [Medline: 29971339]
12. Sarradon-Eck A, Bouchez T, Auroy L, Schuers M, Darmon D. Attitudes of general practitioners toward prescription of mobile health apps: qualitative study. JMIR Mhealth Uhealth 2021 Mar 4;9(3):e21795. [doi: 10.2196/21795] [Medline: 33661123]
13. Lewis TL, Wyatt JC. mHealth and mobile medical apps: a framework to assess risk and promote safer use. J Med Internet Res 2014 Sep 15;16(9):e210. [doi: 10.2196/jmir.3133] [Medline: 25223398]
14. Dhinagaran DA, Martinengo L, Ho MHR, et al. Designing, developing, evaluating, and implementing a smartphone-delivered, rule-based conversational agent (DISCOVER): development of a conceptual framework. JMIR Mhealth Uhealth 2022 Oct 4;10(10):e38740. [doi: 10.2196/38740] [Medline: 36194462]
15. Wilhide Iii CC, Peeples MM, Anthony Kouyaté RC. Evidence-based mhealth chronic disease mobile app intervention design: development of a framework. JMIR Res Protoc 2016 Feb 16;5(1):e25. [doi: 10.2196/resprot.4838] [Medline: 26883135]
16. Mandracchia F, Llauradó E, Tarro L, Valls RM, Solà R. Mobile phone apps for food allergies or intolerances in app stores: systematic search and quality assessment using the mobile app rating scale (MARS). JMIR Mhealth Uhealth 2020 Sep 16;8(9):e18339. [doi: 10.2196/18339] [Medline: 32936078]
17. Farao J, Malila B, Conrad N, Mutsvangwa T, Rangaka MX, Douglas TS. A user-centred design framework for mHealth. PLoS ONE 2020;15(8):e0237910. [doi: 10.1371/journal.pone.0237910] [Medline: 32813711]
18. van Haasteren A, Gille F, Fadda M, Vayena E. Development of the mHealth app trustworthiness checklist. Digit Health 2019;5:2055207619886463. [doi: 10.1177/2055207619886463] [Medline: 31803490]
19. Azad-Khaneghah P, Neubauer N, Miguel Cruz A, Liu L. Mobile health app usability and quality rating scales: a systematic review. Disabil Rehabil Assist Technol 2021 Oct;16(7):712-721. [doi: 10.1080/17483107.2019.1701103] [Medline: 31910687]
20. Nouri R, R Niakan Kalhori S, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inform Assoc 2018 Aug 1;25(8):1089-1098. [doi: 10.1093/jamia/ocy050] [Medline: 29788283]

XSL•FO
RenderX

21. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
22. Levine DM, Co Z, Newmark LP, et al. Design and testing of a mobile health application rating tool. NPJ Digit Med 2020;3(1):74. [doi: 10.1038/s41746-020-0268-9] [Medline: 32509971]
23. Eysenbach G, Group CE. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
24. Kapoor K, George P, Evans MC, Miller WJ, Liu SS. Health literacy: readability of ACC/AHA online patient education material. Cardiology 2017;138(1):36-40. [doi: 10.1159/000475881] [Medline: 28571004]
25. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111. [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
26. Sun K, Goodfellow H, Konstantara E, et al. The multidisciplinary, theory-based co-design of a new digital health intervention supporting the care of oesophageal cancer patients. Digit Health 2021;7:20552076211038410. [doi: 10.1177/20552076211038410] [Medline: 34873450]
27. Broderick J, Devine T, Lemerise AJ, et al. Designing health literate mobile apps. : NAM Perspectives; 2014. [doi: 10.31478/201401a]
28. Pienkowska A, Ang CS, Mammadova M, Mahadzir MDA, Car J. A diabetes education app for people living with type 2 diabetes: co-design study. JMIR Form Res 2023 Sep 18;7(1):e45490. [doi: 10.2196/45490] [Medline: 37721799]
29. Liew H, Pienkowska A, Ang CS, et al. Empowering foot care literacy among people living with diabetes and their carers with an mhealth app: protocol for a feasibility study. JMIR Res Protoc 2023 Nov 21;12(1):e52036. [doi: 10.2196/52036] [Medline: 37988150]
30. CADENCE D-PHA, Nanyang Technological University. Google Play. URL: https://play.google.com/store/apps/details?id=com.ntu.cadence&hl=en_SG [accessed 2026-03-10]
31. Vermeulen J, Verwey R, Hochstenbach LMJ, van der Weegen S, Man YP, de Witte LP. Experiences of multidisciplinary development team members during user-centered design of telecare products and services: a qualitative study. J Med Internet Res 2014 May 19;16(5):e124. [doi: 10.2196/jmir.3195] [Medline: 24840245]
32. Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061. [doi: 10.1136/bmj.n2061] [Medline: 34593508]
33. Schnall R, Rojas M, Bakken S, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251. [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
34. Brett J, Staniszewska S, Mockford C, et al. Mapping the impact of patient and public involvement on health and social care research: a systematic review. Health Expect 2014 Oct;17(5):637-650. [doi: 10.1111/j.1369-7625.2012.00795.x] [Medline: 22809132]
35. Boivin A, Richards T, Forsythe L, et al. Evaluating patient and public involvement in research. BMJ 2018:k5147. [doi: 10.1136/bmj.k5147]
36. Cho YM, Lee S, Islam SMS, Kim SY. Theories applied to m-health interventions for behavior change in low- and middle-income countries: a systematic review. Telemed J E Health 2018 Oct;24(10):727-741. [doi: 10.1089/tmj.2017.0249] [Medline: 29437546]
37. Sharma M. Theoretical Foundations of Health Education and Health Promotion: Jones & Bartlett Learning; 2021.
38. Champion VL, Skinner CS. The health belief model. In: Health Behavior and Health Education: Theory, Research, and Practice: Jossey-Bass; 2008, Vol. 4:45-65.
39. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
40. Sankey M, Birch D, Gardiner M. The impact of multiple representations of content using multimedia on learning outcomes across learning styles and modal preferences. IJEDICT 2011;7(3):18-35 [FREE Full text]
41. Doolittle PE, Tech V. Multimedia learning: empirical results and practical applications. Proc Irish Educ Technol Users' Conf 2002 [FREE Full text]
42. Mayer RE. Multimedia learning. In: Psychology of Learning and Motivation: Elsevier; 2002:85-139. [doi: 10.1016/S0079-7421(02)80005-6]
43. Höffler TN, Leutner D. Instructional animation versus static pictures: a meta-analysis. Learn Instr 2007 Dec;17(6):722-738. [doi: 10.1016/j.learninstruc.2007.09.013]
44. Moe-Byrne T, Evans E, Benhebil N, Knapp P. The effectiveness of video animations as information tools for patients and the general public: a systematic review. Front Digit Health 2022;4:1010779. [doi: 10.3389/fdgth.2022.1010779] [Medline: 36777110]
45. Melton R. Objectives, Competencies and Learning Outcomes: Developing Instructional Materials in Open and Distance Learning: Routledge; 2014. [doi: 10.4324/9781315041926]
46. Anderson LW, Krathwohl DR. A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives: Addison Wesley Longman, Inc; 2001.

XSL•FO
RenderX

47. Su WM, Herron B, Osisek PJ. Using a competency-based approach to patient education: achieving congruence among learning, teaching and evaluation. Nurs Clin North Am 2011 Sep;46(3):291-298. [doi: 10.1016/j.cnur.2011.05.003] [Medline: 21791264]
48. Frank JR, Snell LS, Cate OT, et al. Competency-based medical education: theory to practice. Med Teach 2010;32(8):638-645. [doi: 10.3109/0142159X.2010.501190] [Medline: 20662574]
49. Guyatt GH, Haynes RB, Jaeschke RZ, et al. Users' guides to the medical literature: XXV. Evidence-based medicine: principles for applying the users' guides to patient care. Evidence-based medicine working group. JAMA 2000 Sep 13;284(10):1290-1296. [doi: 10.1001/jama.284.10.1290] [Medline: 10979117]
50. Mach F, Baigent C, Catapano AL, et al. 2019 ESC/EAS guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. Eur Heart J 2020 Jan 1;41(1):111-188. [doi: 10.1093/eurheartj/ehz455] [Medline: 31504418]
51. Grundy SM, Stone NJ, Bailey AL, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines. J Am Coll Cardiol 2019 Jun 25;73(24):e285-e350. [doi: 10.1016/j.jacc.2018.11.003] [Medline: 30423393]
52. Mulder R, Singh AB, Hamilton A, et al. The limitations of using randomised controlled trials as a basis for developing treatment guidelines. Evid Based Mental Health 2018 Feb;21(1):4-6. [doi: 10.1136/eb-2017-102701]
53. Li W, van Wely M, Gurrin L, Mol BW. Integrity of randomized controlled trials: challenges and solutions. Fertil Steril 2020 Jun;113(6):1113-1119. [doi: 10.1016/j.fertnstert.2020.04.018] [Medline: 32387277]
54. Care services. Agency for Integrated Care. URL: https://www.aic.sg/care-services/ [accessed 2026-06-10]
55. Health A to Z. National Health Service. URL: https://www.nhs.uk/conditions/ [accessed 2026-03-10]
56. Health topics A-Z United States of America. Center for Disease Control and Prevention. URL: https://www.cdc.gov/health-topics.html [accessed 2026-03-10]
57. Play console help. Google Help. URL: https://support.google.com/googleplay/android-developer [accessed 2026-03-10]
58. Stableford S, Mettger W. Plain language: a strategic response to the health literacy challenge. J Public Health Policy 2007;28(1):71-93. [doi: 10.1057/palgrave.jphp.3200102] [Medline: 17363939]
59. McInnes N, Haglund BJA. Readability of online health information: implications for health literacy. Inform Health Soc Care 2011 Dec;36(4):173-189. [doi: 10.3109/17538157.2010.542529] [Medline: 21332302]
60. Dunn Lopez K, Chae S, Michele G, et al. Improved readability and functions needed for mHealth apps targeting patients with heart failure: an app store review. Res Nurs Health 2021 Feb;44(1):71-80. [doi: 10.1002/nur.22078] [Medline: 33107056]
61. Ayyaswami V, Padmanabhan DL, Crihalmeanu T, Thelmo F, Prabhu AV, Magnani JW. Mobile health applications for atrial fibrillation: a readability and quality assessment. Int J Cardiol 2019 Oct 15;293:288-293. [doi: 10.1016/j.ijcard.2019.07.026] [Medline: 31327518]
62. Sand-Jecklin K. The impact of medical terminology on readability of patient education materials. J Community Health Nurs 2007;24(2):119-129. [doi: 10.1080/07370010701316254] [Medline: 17563283]
63. Shoemaker SJ, Wolf MS, Brach C. Development of the patient education materials assessment tool (PEMAT): a new measure of understandability and actionability for print and audiovisual patient information. Patient Educ Couns 2014 Sep;96(3):395-403. [doi: 10.1016/j.pec.2014.05.027] [Medline: 24973195]
64. Language Matters Diabetes. URL: https://www.languagemattersdiabetes.com/ [accessed 2026-03-10]
65. Ayonrinde O. Importance of cultural sensitivity in therapeutic transactions: considerations for healthcare providers. Dis Manage Health Outcomes 2003;11:233-248. [doi: 10.2165/00115677-200311040-00004]
66. Brooks LA, Manias E, Bloomer MJ. Culturally sensitive communication in healthcare: a concept analysis. Collegian 2019 Jun;26(3):383-391. [doi: 10.1016/j.colegn.2018.09.007]
67. Fyfield M, Henderson M, Phillips M. Improving instructional video design: a systematic review. AJET 2022;38(3):150-178. [doi: 10.14742/ajet.7296]
68. Hou G, Anicetus U, He J. How to design font size for older adults: a systematic literature review with a mobile device. Front Psychol 2022;13:931646. [doi: 10.3389/fpsyg.2022.931646] [Medline: 35978796]
69. Muro-Culebras A, Escriche-Escuder A, Martin-Martin J, et al. Tools for evaluating the content, efficacy, and usability of mobile health apps according to the consensus-based standards for the selection of health measurement instruments: systematic review. JMIR Mhealth Uhealth 2021 Dec 1;9(12):e15433. [doi: 10.2196/15433] [Medline: 34855618]
70. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 5;16(6):e146. [doi: 10.2196/jmir.3077] [Medline: 24905070]
71. Hussain M, Zaidan AA, Zidan BB, et al. Conceptual framework for the security of mobile health applications on Android platform. Telemat Inform 2018 Aug;35(5):1335-1354. [doi: 10.1016/j.tele.2018.03.005]
72. Kotz D, Avancha S, Baxi A. A privacy framework for mobile health and home-care systems. 2009 Nov 13 Presented at: CCS '09: 16th ACM Conference on Computer and Communications Security 2009; Nov 13, 2009. [doi: 10.1145/1655084.1655086]
73. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mHealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817. [doi: 10.2196/13817] [Medline: 31293246]

XSL•FO
RenderX

74. Lin YH, Lou MF. Effects of mHealth-based interventions on health literacy and related factors: a systematic review. J Nurs Manag 2021 Apr;29(3):385-394. [doi: 10.1111/jonm.13175] [Medline: 33040429]

Abbreviations

AI: artificial intelligence
CONSORT: Consolidated Standards of Reporting Trials
CVD: cardiovascular disease
FAQ: frequently asked questions
HCP: health care professional
MARS: Mobile App Rating Scale
mHealth: mobile health
PAR: participatory action research
PPI: patient and public involvement
SCALED: Systematic and Collaborative Approach to Learning and Educational Content Development

Edited by A Stone, L Buis; submitted 05.Sep.2025; peer-reviewed by D Jones, S Narayan; revised version received 13.Apr.2026; accepted 04.May.2026; published 02.Jul.2026.

Please cite as:
Ong QC, Pienkowska A, Obarcanin E, Stona AC, Khong AWH, Car J, Ho AHY
Systematic and Collaborative Approach to Learning and Educational Content Development (SCALED) for Health Apps: An Experience-Informed Conceptual Framework
JMIR Mhealth Uhealth 2026;14:e83442
URL: https://mhealth.jmir.org/2026/1/e83442
doi:10.2196/83442

© Qi Chwen Ong, Anita Pienkowska, Emina Obarcanin, Anne-Claire Stona, Andy W H Khong, Josip Car, Andy Hau Yan Ho. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 2.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Intensive, Repeated Self-Report Measures: Should We Be Concerned About Changes in Data Quality Over Time?

Arthur A Stone[1*], PhD; Stefan Schneider[1*], PhD; Meynard J Toledo[1*], PhD; Raymond Hernandez[1], PhD

Department of Psychology, University of Southern California, Los Angeles, CA, United States

[*]these authors contributed equally

Corresponding Author:
Arthur A Stone, PhD
Department of Psychology
University of Southern California
McClintock Street
Los Angeles, CA, 90290
United States
Phone: 1 6316820014
Email: arthur.stone@usc.edu

Abstract

Intensive, repeated self-report measures are an important tool for behavioral and medical researchers and practitioners who are concerned with the dynamic interplay among variables at a granular level. Many mobile health applications rely on accurate measurement of immediate states and environments for both assessment and intervention delivery. Techniques for capturing repeated momentary assessments yield data with several salutary qualities: recall bias is minimized relative to assessments that rely on much longer recall periods; measurements are taken in individuals' everyday environments; and dense, repeated measures allow a new window into the processes transpiring between individuals and their environments. In this paper, we highlight several features of repeatedly completing momentary assessments that may change the nature or quality of the data collected over time. Several lines of inquiry are discussed that call into question the presumption that there is invariance in how people complete repeated assessments over time. A result of this possibility could be a reduction in data quality. We present 4 phenomena, with selected results, that may induce noninvariance in repeated measures: the amount of time required to complete assessments, the rate of missing data, the degree of careless responding, and the presence of several components of reactivity. In each of these areas, we found evidence that changes could occur over time, and we consider how data might be affected by such changes. Our conclusion is that researchers should be aware that changes can occur over time and that these changes may affect data quality.

(JMIR Mhealth Uhealth 2026;14:e68735) doi:10.2196/68735

KEYWORDS

ecological momentary assessment; EMA; repeated measurements; invariance; self-report

Introduction

Repeated self-report measures are central to intensive longitudinal methods that use momentary assessments, including ecological momentary assessment (EMA) [1,2], the experience sampling method [3], and ambulatory assessment methods [4]. Methodological strategies derived from momentary assessment studies form the backbone of optimal mobile health (mHealth) data acquisition for participant monitoring, in-the-moment treatment delivery (eg, just-in-time adaptive interventions), and real-time feedback on behavior and thoughts. As such, there is considerable value in recognizing the strengths and the potential limitations of these methods.

The advantages of repeated measures over static measures are numerous: they provide the opportunity to track trajectories of outcomes over time, enable exploration of lagged associations that may confer near-causal status, and reduce problems associated with retrospection [2,5-7]. Repeated-measures study designs are also part and parcel of many mHealth practices. Moreover, constructs often measured using momentary methods are pertinent for mHealth researchers; they include internal states (eg, pain intensity and quality, emotions, fatigue, perceived stress, symptoms, and cognitive status) and external (often observable) behaviors (eg, daily activities, social engagement, consumption, substance use, exogenous events, and location). Real-time self-reports may also be linked to ambulatory measures of physiological function (eg, heart rate, blood pressure, cortisol levels, and blood glucose) and to data

XSL•FO
RenderX

routinely collected by smartphones (eg, location, step counts, and time taken to respond to questions) to provide insights into the dynamics of these variables. Finally, intensive, repeated measures are indispensable tools for monitoring populations, evaluating treatments in medical research, analyzing economic patterns, and addressing behavioral science questions. Many review and position papers [6,8-11] are available for the interested reader.

A generally unspoken belief about repeated self-report measures is that repeated assessments conducted throughout a study—regardless of the study duration or the momentary measurement frequency—are thought to be invariant regarding how questions are interpreted, how response scales are construed, and how internal reference standards are applied to ratings. All of these are important considerations for achieving reliable and valid data. That is, we assume that the quality and integrity of responses to all assessments are not affected by repeated measurement processes. However, if this assumption of invariance does not hold, it becomes challenging to disentangle true within-person changes from shifts in measurement quality over time. For example, observed within-person patterns could be erroneously attributed to genuine change when they instead reflect evolving interpretations of questions or changes in response scale use, thereby compromising the validity of inferences about within-person dynamics and processes.

Our primary intention in this viewpoint is to highlight several features of repeated momentary assessments that we think deserve attention, given the possibility that they are associated with changes in momentary data quality. We present selected results to illustrate the point. Because the findings presented later have been reported previously, it is reasonable to assume that they may be present in at least some repeated momentary studies and, importantly, *may* systematically impact data quality and warrant further research. In the same spirit, we do not argue that the features presented are ubiquitous in EMA studies.

We now describe how intensive longitudinal momentary measurements may be distorted by repeatedly asking the same or similar questions over many repetitions. Theoretically, these distortions may be particularly germane when assessments occur many times a day, with only hours between assessments [12], but they may also occur at longer intervals. There is already substantial evidence that responses are not invariant over time, given that repeated measures can create practice effects, a phenomenon supported by abundant research on memory and performance [13-15]. Practice effects could be salutary, leading to faster and more accurate responses [16], or pernicious, leading to unwanted measurement effects such as increased bias and error [17]. In either case, the key point is that repeated exposure to the content can change how questions are answered. To date, evidence on the *adverse* consequences of repeated assessments as they pertain to EMA has been available in a piecemeal fashion spanning several content domains, making it difficult to fully appreciate the scope and importance of the issue.

Response Times Decrease for EMA Question Completion Over Repeated Measures

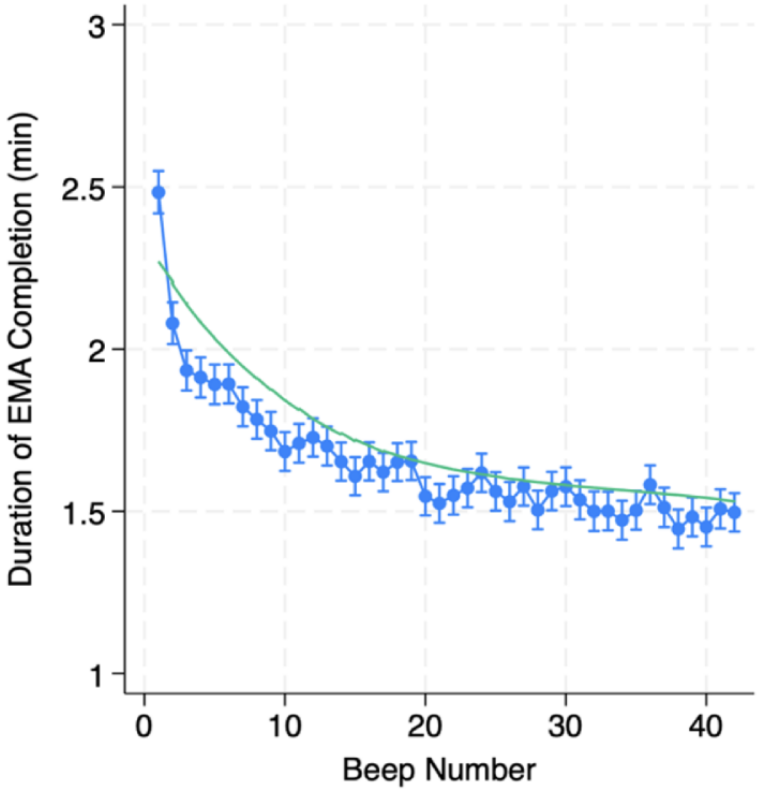
Overview

Although perhaps not well known to EMA researchers, studies examining the time taken to complete an EMA assessment have reported decreases in response time (time it takes individuals to complete questions) over repeated measures, consistent with the practice effects mentioned previously [18-20]. Response time reductions are likely due to respondents' increasing familiarity with the task, with the wording and meaning of questions, and with the response options over repeated assessments [21]. These decrements can be striking, with reductions of more than 50%. Aggregated response times usually follow a negative exponential pattern, with the steepest decrements occurring early during repeated measures.

To illustrate this pattern of response times over time in an EMA study, data from the Understanding American Study conducted at University of Southern California are shown in Figure 1. A total of 22,531 prompts were recorded from 706 community-dwelling adults, and up to 42 prompts were completed per week. Average response time is plotted for each prompt, revealing a dramatic drop across the first 9 prompts followed by a modest decline over the remaining prompts. Over the course of a week, there was a 32% decline in response times. The magnitude of the decrease in survey completion time is influenced by multiple factors, including the burden of the interview, incentives for completion, and the composition of the participant sample.

XSL•FO
RenderX

Figure 1. Response times over prompts. EMA: ecological momentary assessment.



Implications

We conclude that the time to complete assessments may decrease over repeated measures. We speculate that if such changes are associated with, or caused by, any of the factors discussed later, data quality may change over repeated measures. For example, increasing missingness and carelessness over time may generate faster response times and, consequently, compromise data quality. Given this possibility, we suggest that researchers examine the response time changes over the course of repeated measures, though we also acknowledge that response time may or may not be associated with validity.

Rate of Data Missingness May Increase Over Repeated Measures

Overview

One plausible explanation for faster survey completion with repeated measures is that individuals gradually disengage from the study procedures and, therefore, spend less time completing assessments. At the start of a study, participants may be motivated to complete all questions and generate optimal answers but may lose interest and become less invested as they repeatedly answer the same questions [22]. When this happens, participants may show increasing noncompliance with the study protocol and may miss assessments when prompted, a concern that has previously been acknowledged by EMA researchers [23].

To assess the magnitude of change in missed EMA prompts over time, we examined many EMA studies with a wide range of samples and variety of assessment schedules. Some studies showed a drop in compliance rates over time, across both shorter and longer studies, and with various frequencies of EMA prompting. Therefore, we suggest that repeated measures can induce *some level* of missingness in EMA studies.

We now turn to the question of whether changes in missing data are serious contenders for inducing bias and how such missingness may affect data quality. If shifts in missingness over time are minimal or if compliance rates remain high overall despite an increase in missed assessments, there is likely little cause of concern. However, in the presence of substantial increases in missingness, missing data can introduce a range of problems, including reduced statistical power, unrepresentative momentary samples (or changes in sample representation), and potential distortions of longitudinal effects. Increasing rates of missing values imply that missing values do not occur completely at random, and researchers may wish to routinely incorporate study day (or prompt number) as an auxiliary variable or covariate in statistical analyses to reduce potential bias due to systematic missingness patterns (ie, to increase the plausibility of the assumption that data are missing at random. Although statistical methods (eg, multilevel multiple imputation [24]) can help account for missing-at-random mechanisms, it is not possible to fully protect against potential patterns that are missing not at random, and only a few studies have developed or evaluated missing data methods specifically for intensive longitudinal contexts, such as EMA data [25].

Implications

Repeated measures may create bias when missingness increases to substantial levels over the course of a study, and we suggest that missingness over time be considered in intensive repeated-measures studies. Unfortunately, the precise proportion

XSL•FO
RenderX

of missingness that would signal caution depends on the cause of this missingness (eg, simple burden or a change in activities that encourage missingness, such as more exercise). Our recommendation is to implement as many safeguards as possible to encourage the most complete data throughout an entire study. This could entail a variety of methodological maneuvers: creating EMA designs that reduce respondent burden related to questionnaire length, daily sampling frequency and scheduling, and study duration; incentive structures that encourage high compliance (eg, payment being contingent on prespecified completion rules from the investigators); and implementing procedures for monitoring compliance over time and providing appropriate feedback to encourage compliance (eg, real-time or daily signaling to the research team about missed prompts). Further research would be welcome to determine which of these strategies are most effective, so that efforts to reduce missingness could be implemented efficiently.

Frequency of Careless Responding Over Time

Overview

Careless responding (or insufficient effort responding) is defined as respondents providing answers without regard to the content of the questions, and it can occur when they do not read an item or do not pay attention to what the item is asking [26,27]. Careless responding can also occur as a more subtle form of participant disengagement compared with the overt noncompliance discussed previously, in which a respondent misses prompts entirely. To reduce the burden associated with repeated assessments, respondents may minimize the effort expended by doing the least possible to satisfy study requirements [22]. Careless responding can present as invariable responses (eg, *straight-lining*, in which the same answer is given to every question) or as inconsistent or random responses. A recent small study interviewed individuals at the close of a many-month EMA study to better understand the burden and other issues that could affect data quality [28]. The authors found that respondents tried to counter the burden by responding quickly and that repeated measures over time led to more neutral responses.

Many studies have shown the detrimental effects that careless responding can have on measurement accuracy and reliability [26]. Careless responding can occur when respondents are asked to complete long questionnaires. For instance, toward the end of lengthy questionnaires with dozens or hundreds of items, respondents often give more random, uniform, and fast responses, suggesting that their motivation has waned over time [29,30]. An intuitive extrapolation of these findings is that the repetitive nature of completing brief surveys may similarly lead to more careless responding. Ganzach and Bulmash [31] documented less variable and less complex patterns of self-reporting over the course of daily repeated measures on affect and stress. These findings demonstrated that increased carelessness could occur over repeated measures [32]. Whether decreases in survey completion time can be attributed to increases in extremely fast responses (an indicator of carelessness) has also been shown to occur. In another EMA study, the frequency of speed responding increased from 3% to 8% across 49 assessments over 7 days [33]. Again, we are not saying that all studies, or even most studies, show this pattern—only that it can occur in repeated-measures studies. Another method for assessing changes in carelessness over time is by comparing self-report ratings with objective assessments of the same concept. With increased carelessness, associations over repeated measures would be predicted to decrease because of the increased measurement error. One study found a decreasing association between self-reported and objectively measured time spent on social media in adolescents, which could have been generated through carelessness [34].

Implications

Available evidence shows instances of increasing carelessness over densely repeated measures. We recommend future work to examine possible changes in careless responding over time. The addition of attention check questions might also improve engagement and reduce carelessness [28]. A fruitful direction for future research could be the development of advanced statistical models, including machine learning techniques, to detect careless responding in datasets in which EMA self-reports can be compared with objective measures. Furthermore, the longitudinal aspect of intensive repeated-measures data may prove a fertile ground for extending carelessness detection approaches, for instance, by examining shifts in within-person parameters (eg, variability, autocorrelation, and outliers) over time.

Reactivity to Question Content May Increase Over Repeated Measures

Overview

Reactivity, or reactive arrangements, has been recognized since the 1930s, when workers at a manufacturing facility in Hawthorne, New Jersey, altered their behavior in response to their actions being observed by scientists [35], even though an alternative interpretation has been offered [36]. Here, we restrict our inquiry to reactivity associated with repeated measures in intensive momentary designs, namely, the repeated answering of the same or similar questions many times. This section is divided into 3 subparts: reactivity producing a change in a targeted behavior or cognition, reactivity producing a change in the reporting of a behavior or cognition (but without corresponding change in the targeted phenomenon), and specific reactivity processes associated with changes in question meaning or scale recalibration. These are not exclusive designations, and it is possible that studies exhibit blends of these types of reactivity.

Reactivity in Behavior, Cognition, and Emotion

This type of reactivity is defined as a change in actual behavior, cognition, and emotion over time due to repeated responses. In fact, some interventions have intentionally incorporated this type of reactivity to induce desired outcomes, such as the use of self-monitoring to affect physical activity behaviors [37,38]. Despite its potential utility in intervention contexts, this phenomenon poses a threat to the internal validity of an observational study as it signifies that measurement is not

XSL•FO
RenderX

invariant over time. For example, repeated assessments may cause participants to become more conscious of their behaviors, which may, in turn, trigger self-regulatory mechanisms [39]. The same process may happen with behaviors and emotions. Unfortunately, very few studies are available on whether and how this phenomenon unfolds in EMA studies over time (eg, potential for cumulative effects) [40].

Reactivity in Reporting

A second type of reactivity is defined by respondents altering the way they *report* behaviors, cognitions, and emotions, independent of whether those behaviors or states have actually changed over time (the previous type of reactivity). In repeated-measures studies, this phenomenon could manifest as a shift in the validity of the reported variables over repeated assessments. For example, reports of affect might be accurate at the onset of a study but gradually shift to become inaccurate or less reliable later; the converse might also be true, as discussed later [18]. A problem for reporting reactivity is that most studies cannot distinguish between real change and reporting change because doing so requires objective measurement of the variable of interest, which is very uncommon in the constructs being measured in EMA. This requirement presents a challenge for the assessment of internal states, where changes over time may indicate actual change, reporting change, or a combination of both.

Reactivity Processes

The third section on reactivity discusses several ways that actual levels of, or reports of, behavior, cognition, and emotion might change over time through specific psychological processes that may be inherent in repeated measures.

Initial elevation is defined as higher levels of a variable when measured early in a study compared with levels observed at later measurements when there is no apparent reason for a shift in levels over time [41-43]. Although recognized for many years, a paper by Shrout et al [43] has renewed interest in the phenomenon. They found an initial elevation effect for momentary negative states and for momentary internal states in 4 studies with repeated measures, with small to medium effect sizes. Furthermore, Anvari et al [44] (2023) studied several thousand college students and reported a strong initial elevation bias. Thus, initial elevation has been observed, although there is some competing evidence [45] consistently. Biasing of overall levels of variables can result from such effects, for example, by producing (apparently) incorrect downward trajectories in variables over time.

Another facet of reactivity that has not received much attention is that the *meaning of questions* may evolve over repeated administrations of questions [46]. Following up on previous work by Windle [47] (1994), Knowles et al [48] (1996) explicitly explored the possibility that the interpretation of questions changes over time. Results indicated that over multiple exposures to items, individuals gained knowledge about the construct being studied (an anxiety questionnaire). Moreover, they observed shifts in question meaning within a single test administration. Later interpretations of anxiety items demonstrated that a higher standard for endorsing anxiety was evident; if all else remained constant, this would result in diminished reported anxiety later in the study. The upshot of such shifts in question meaning over repeated measures over time is clear: what we learn at the beginning of a study might not be comparable with what we learn later.

Another aspect of question meaning is *responsescale recalibration*, defined as a shift in how participants use question response options that occurs as a reaction to events experienced throughout the study [46]. Exposures to events that elicit very intense emotions or other strong reactions (eg, pain of childbirth or trauma) may alter the way individuals interpret a rating scale. Such recalibration could result in a stimulus that was once rated at a certain intensity to later receive a lower intensity rating because the upper end of the scale was redefined by the extreme emotion or reaction. Alternatively, recalibration could be induced in certain treatment modalities intended to reduce reported symptomatology through cognitive and social manipulations [49]. Therefore, recalibration could have a major impact on interpreting results over time.

Implications

It is fundamental to know if repeatedly measuring a variable creates any of the reactivity processes, given that they would be a threat to internal validity. If there is a high likelihood for reactivity, we recommend preliminary investigations to assess the potential magnitude of these effects and to incorporate design strategies to mitigate them, while acknowledging the extra financial and investigator burden inherent in this recommendation. More broadly, if empirical evidence suggests that reactivity is likely (eg, respondents having access to feedback that has been previously shown to induce reactivity), steps could be taken to eliminate the reactivity-inducing aspect of the protocol.

An existing barrier to the application of this approach is that we simply do not have an adequate way to identify reactivity-inducing circumstances. Of particular importance is the possibility that processes and events with the potential of shifting means or recalibrating scales are not routinely measured in longitudinal studies; we therefore recommend routine collection of such data to signal possible distortion. In addition, experimental manipulations of item administration (eg, item order) have been recommended as useful methods to detect possible biases from measurement reactivity [19]. Advancing methodologies to statistically control for these shifts is also recommended; however, some efforts to accomplish this have not resulted in reduced bias, such as using the *then-test* to control for shifts from pretest to posttest [49].

Conclusions

The aim of this viewpoint was to consider the possibility that studies using intensive repeated self-report measures could distort interpretations of data collected over time. We found evidence supporting the potential for all factors considered: response times can be considerably reduced, missing data rates can increase, carelessness rates can rise, and reactive effects (including reference standards and question meaning) may manifest over repeated measures. Response time reductions

XSL•FO
RenderX

alone may or may not impact data quality, as practice effects could be positive or negative. However, when reduced response times lead to increases in missingness, carelessness, and reactivity, there is the potential for compromised data quality. Heightened awareness of these possibilities is warranted for researchers using repeated assessments. We also offered suggestions for monitoring and reducing these possible effects.

We view this discussion as a supplement to prior papers that suggested *guidelines* for conducting, reporting, and analyzing EMA studies. Those papers have guidelines that could be helpful for examining some threats to validity across repeated measures, although not with an eye toward invariance over time. As early as 2002, an EMA guidelines publication [50] indicated the need for comprehensive reporting of procedures and enumerated ways of calculating prompt compliance, which are useful for understanding the threats to validity discussed here. The early set of guidelines did not mention tracking the duration of prompt completion or methods for assessing careless responding. A guidance document [51] included the recommendation to record prompt duration; these data would be pertinent for evaluating the issues raised in this paper. However, their suggestions did not mention tracking careless responding. Overall, these recommendations are positive developments for the field that should result in more robust EMA studies.

In summary, we believe a cautious position is warranted regarding the possibility that repeated questionnaire measurements affect how questions are answered. It is plausible that some results from repeated-measures studies have been affected by the processes discussed previously. This paper calls for research directed at understanding whether reactivity distorts data collected through intensive repeated measures. A speculative extension of this conclusion concerns repeated measures taken at longer intervals, such as every week or month: such longitudinal data may also be susceptible to the biases reported here, with potentially far-reaching implications.

Funding

This study was supported by award R37AG057685 (AAS, principal investigator).

Conflicts of Interest

AAS is a senior scientist with the Gallup Organization and a consultant for Lore Contagious, Inc and AstraZeneca, Inc. All other authors declare no other conflicts of interest.

References

1. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008 Apr 01;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
2. Stone AA, Shiffman S. Ecological momentary assessment (EMA) in behavioral medicine. Ann Behav Med 1994;16(3):199-202. [doi: 10.1093/abm/16.3.199]
3. deVries MW. Investigating mental disorders in their natural settings: introduction to the special issue. J Nerv Ment Dis 1987 Sep;175(9):509-513. [doi: 10.1097/00005053-198709000-00001] [Medline: 3655774]
4. Trull TJ, Ebner-Priemer UW. Ambulatory assessment in psychopathology research: a review of recommended reporting guidelines and current practices. J Abnorm Psychol 2020 Jan;129(1):56-63. [doi: 10.1037/abn0000473] [Medline: 31868388]
5. Stone AA, Turkkan JS, Jobe JB, Bachrach C, Kurtzman HS, Cain VS. The Science of Self-Report: Implications for Research and Practice. London, UK: Lawrence Erlbaum Associates; 2000.
6. Bolger N, Laurenceau JP. Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research. New York, NY: Guilford Press; 2013.
7. Trull TJ, Ebner-Priemer UW. Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: introduction to the special section. Psychol Assess 2009 Dec;21(4):457-462 [FREE Full text] [doi: 10.1037/a0017653] [Medline: 19947780]
8. Smyth JM, Stone AA. Ecological momentary assessment research in behavioral medicine. J Happiness Stud 2003 Mar;4:35-52. [doi: 10.1023/a:1023657221954]
9. Ebner-Priemer UW, Trull TJ. Ecological momentary assessment of mood disorders and mood dysregulation. Psychol Assess 2009 Dec;21(4):463-475. [doi: 10.1037/a0017075] [Medline: 19947781]
10. Stone AA, Broderick JE, Schneider S, Schwartz JE. Expanding options for developing outcome measures from momentary assessment data. Psychosom Med 2012 May;74(4):387-397 [FREE Full text] [doi: 10.1097/PSY.0b013e3182571faa] [Medline: 22582336]
11. Stone AA, Obbarius A, Junghaenel DU, Wen CK, Schneider S. High-resolution, field approaches for assessing pain: ecological momentary assessment. Pain 2021 Jan;162(1):4-9 [FREE Full text] [doi: 10.1097/j.pain.0000000000002049] [Medline: 32833794]
12. Robinson MD, Clore GL. Belief and feeling: evidence for an accessibility model of emotional self-report. Psychol Bull 2002;128(6):934-960. [doi: 10.1037//0033-2909.128.6.934]

XSL•FO
RenderX

13. Bartels C, Wegrzyn M, Wiedl A, Ackermann V, Ehrenreich H. Practice effects in healthy adults: a longitudinal study on frequent repetitive cognitive testing. BMC Neurosci 2010 Sep 16;11:118 [FREE Full text] [doi: 10.1186/1471-2202-11-118] [Medline: 20846444]
14. Holm SP, Wolfer AM, Pointeau GH, Lipsmeier F, Lindemann M. Practice effects in performance outcome measures in patients living with neurologic disorders - a systematic review. Heliyon 2022 Aug;8(8):e10259 [FREE Full text] [doi: 10.1016/j.heliyon.2022.e10259] [Medline: 36082322]
15. Duff K, Callister C, Dennett K, Tometich D. Practice effects: a unique cognitive variable. Clin Neuropsychol 2012;26(7):1117-1127. [doi: 10.1080/13854046.2012.722685] [Medline: 23020261]
16. van den Broek GS, Segers E, Takashima A, Verhoeven L. Do testing effects change over time? Insights from immediate and delayed retrieval speed. Memory 2014 Sep 02;22(7):803-812. [doi: 10.1080/09658211.2013.831455] [Medline: 23998337]
17. Schnell PM, Wascher M, Rempala GA. Overcoming repeated testing schedule bias in estimates of disease prevalence. J Am Stat Assoc 2023 Sep 06;119(545):1-13. [doi: 10.1080/01621459.2023.2238943]
18. Hernandez R, Schneider S, Pinkham AE, Depp CA, Ackerman R, Pyatak EA, et al. Comparisons of self-report with objective measurements suggest faster responding but little change in response quality over time in ecological momentary assessment studies. Assessment 2025 Apr;32(3):335-355. [doi: 10.1177/10731911241245793] [Medline: 38634454]
19. Arslan RC, Reitz AK, Driebe JC, Gerlach TM, Penke L. Routinely randomize potential sources of measurement reactivity to estimate and adjust for biases in subjective reports. Psychol Methods 2021 Apr;26(2):175-185. [doi: 10.1037/met0000294] [Medline: 32584065]
20. Schneider S, Hernandez R, Junghaenel DU, Orriens B, Lee PJ, Stone AA. Response times in ecological momentary assessment (EMA): shedding light on the response process with a drift diffusion model. Curr Psychol 2024 Feb 27;43(7):5868-5886. [doi: 10.1007/s12144-023-04773-0] [Medline: 40969298]
21. McCabe D, Langer KG, Borod JC, Bender HA. Practice effects. In: Kreutzer JS, DeLuca J, Caplan B, editors. Encyclopedia of Clinical Neuropsychology. New York, NY: Springer; 2011.
22. Krosnick JA. Response strategies for coping with the cognitive demands of attitude measures in surveys. Appl Cognit Psychol 2006 Feb 13;5(3):213-236. [doi: 10.1002/acp.2350050305]
23. Wrzus C, Neubauer AB. Ecological momentary assessment: a meta-analysis on designs, samples, and compliance across research fields. Assessment 2023 Apr 11;30(3):825-846 [FREE Full text] [doi: 10.1177/10731911211067538] [Medline: 35016567]
24. Ji L, Li Y, Potter LN, Lam CY, Nahum-Shani I, Wetter DW, et al. Multiple imputation of missing data in multilevel ecological momentary assessments: an example using smoking cessation study data. Front Digit Health 2023 Nov 10;5:1099517 [FREE Full text] [doi: 10.3389/fdgth.2023.1099517] [Medline: 38026834]
25. Goldberg SB, Bolt DM, Davidson RJ. Data missing not at random in mobile health research: assessment of the problem and a case for sensitivity analyses. J Med Internet Res 2021 Jun 15;23(6):e26749 [FREE Full text] [doi: 10.2196/26749] [Medline: 34128810]
26. Ward MK, Meade AW. Dealing with careless responding in survey data: prevention, identification, and recommended best practices. Annu Rev Psychol 2023 Jan 18;74(1):577-596 [FREE Full text] [doi: 10.1146/annurev-psych-040422-045007] [Medline: 35973734]
27. Huang JL, Liu M, Bowling NA. Insufficient effort responding: examining an insidious confound in survey data. J Appl Psychol 2015 May;100(3):828-845. [doi: 10.1037/a0038510] [Medline: 25495093]
28. Wang SD, Hatzinger L, Morales J, Hewus M, Intille S, Dunton GF. Burden and inattentive responding in a 12-month intensive longitudinal study: interview study among young adults. JMIR Form Res 2024 Aug 02;8:e52165 [FREE Full text] [doi: 10.2196/52165] [Medline: 39093606]
29. Bowling NA, Gibson AM, Houpt JW, Brower CK. Will the questions ever end? Person-level increases in careless responding during questionnaire completion. Organ Res Methods 2020 Aug 26;24(4):718-738. [doi: 10.1177/1094428120947794]
30. Galesic M, Bosnjak M. Effects of questionnaire length on participation and indicators of response quality in a web survey. Public Opin Q 2009 Jun;73(2):349-360. [doi: 10.1093/poq/nfp031]
31. Ganzach Y, Bulmash B. The effect of serial day on the measurement of positivity and emotional complexity in diary studies. Eur J Soc Psychol 2021 Dec;51(7):1213-1225. [doi: 10.1002/ejsp.2809]
32. Gochmann V, Ohly S, Kotte S. Diary studies, a double-edged sword? An experimental exploration of possible distortions due to daily reporting of social interactions. J Organ Behav 2022 Sep;43(7):1209-1223. [doi: 10.1002/job.2633]
33. Ulitzsch E, Nestler S, Lüdtke O, Nagy G. A screen-time-based mixture model for identifying and monitoring careless and insufficient effort responding in ecological momentary assessment data. Psychol Methods 2024 Feb 29. [doi: 10.1037/met0000636] [Medline: 38421768]
34. Verbeij T, Pouwels JL, Beyens I, Valkenburg PM. The accuracy and validity of self-reported social media use measures among adolescents. Comput Hum Behav Rep 2021;3:100090. [doi: 10.1016/j.chbr.2021.100090]
35. Gillespie R. Manufacturing Knowledge: A History of the Hawthorne Experiments. Cambridge, UK: Cambridge University Press; 1993.

XSL•FO
RenderX

36. Adair JG. The Hawthorne effect: a reconsideration of the methodological artifact. J Appl Psychol 1984 May;69(2):334-345. [doi: 10.1037/0021-9010.69.2.334]
37. Compernolle S, DeSmet A, Poppe L, Crombez G, De Bourdeaudhuij I, Cardon G, et al. Effectiveness of interventions using self-monitoring to reduce sedentary behavior in adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2019 Aug 13;16(1):63 [FREE Full text] [doi: 10.1186/s12966-019-0824-3] [Medline: 31409357]
38. Kanejima Y, Kitamura M, Izawa KP. Self-monitoring to increase physical activity in patients with cardiovascular disease: a systematic review and meta-analysis. Aging Clin Exp Res 2019 Feb;31(2):163-173. [doi: 10.1007/s40520-018-0960-7] [Medline: 29714027]
39. Maher JP, Arigo D, Baga K, Salvatore GM, Pasko K, Hudgins B, et al. Measurement reactivity in ecological momentary assessment studies of movement-related behaviors. J Meas Phys Behav 2024 Jan;7(1) [FREE Full text] [doi: 10.1123/jmpb.2023-0035]
40. Eisele G, Vachon H, Lafit G, Tuyaerts D, Houben M, Kuppens P, et al. A mixed-method investigation into measurement reactivity to the experience sampling method: the role of sampling protocol and individual characteristics. Psychol Assess 2023 Jan;35(1):68-81. [doi: 10.1037/pas0001177] [Medline: 36174163]
41. Johar O, Sackett AM. The self-contaminating nature of repeated reports of negative emotions. Basic Appl Soc Psychol 2018;40(5):293-307. [doi: 10.1080/01973533.2018.1496336] [Medline: 38634454]
42. French DP, Sutton S. Reactivity of measurement in health psychology: how much of a problem is it? What can be done about it? Br J Health Psychol 2010 Sep;15(Pt 3):453-468. [doi: 10.1348/135910710X492341] [Medline: 20205982]
43. Shrout PE, Stadler G, Lane SP, McClure MJ, Jackson GL, Clavél FD, et al. Initial elevation bias in subjective reports. Proc Natl Acad Sci U S A 2018 Jan 02;115(1):E15-E23 [FREE Full text] [doi: 10.1073/pnas.1712277115] [Medline: 29255039]
44. Anvari F, Efendić E, Olsen J, Arslan RC, Elson M, Schneider IK. Bias in self-reports: an initial elevation phenomenon. Soc Psychol Pers Sci 2022 Oct 07;14(6):727-737. [doi: 10.1177/19485506221129160]
45. Cerino ES, Schneider S, Stone AA, Sliwinski MJ, Mogle J, Smyth JM. Little evidence for consistent initial elevation bias in self-reported momentary affect: a coordinated analysis of ecological momentary assessment studies. Psychol Assess 2022 May;34(5):467-482 [FREE Full text] [doi: 10.1037/pas0001108] [Medline: 35175074]
46. Ubel PA, Peeters Y, Smith D. Abandoning the language of "response shift": a plea for conceptual clarity in distinguishing scale recalibration from true changes in quality of life. Qual Life Res 2010 May 29;19(4):465-471. [doi: 10.1007/s11136-010-9592-x] [Medline: 20112000]
47. Windle C. Test-retest effect on personality questionnaires. Educ Psychol Meas 1954 Dec 01;14(4):617-633. [doi: 10.1177/001316445401400404]
48. Knowles ES, Coker MC, Scott RA, Cook DA, Neville JW. Measurement-induced improvement in anxiety: mean shifts with repeated assessment. J Pers Soc Psychol 1996 Aug;71(2):352-363. [doi: 10.1037//0022-3514.71.2.352] [Medline: 8765486]
49. Nolte S, Elsworth GR, Sinclair AJ, Osborne RH. The inclusion of 'then-test' questions in post-test questionnaires alters post-test responses: a randomized study of bias in health program evaluation. Qual Life Res 2012 Apr 28;21(3):487-494. [doi: 10.1007/s11136-011-9952-1] [Medline: 21710355]
50. Stone AA, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med 2002 Aug;24(3):236-243. [doi: 10.1207/s15324796abm2403_09]
51. Dao KP, De Cocker K, Tong HL, Kocaballi AB, Chow C, Laranjo L. Smartphone-delivered ecological momentary interventions based on ecological momentary assessments to promote health behaviors: systematic review and adapted checklist for reporting ecological momentary assessment and intervention studies. JMIR Mhealth Uhealth 2021 Nov 19;9(11):e22890 [FREE Full text] [doi: 10.2196/22890] [Medline: 34806995]

Abbreviations

EMA: ecological momentary assessment
mHealth: mobile health

Edited by L Buis; submitted 12.Nov.2024; peer-reviewed by A Jamal, C Baxter; comments to author 24.Apr.2025; revised version received 17.Oct.2025; accepted 16.Dec.2025; published 14.Jan.2026.

Please cite as:
Stone AA, Schneider S, Toledo MJ, Hernandez R
Intensive, Repeated Self-Report Measures: Should We Be Concerned About Changes in Data Quality Over Time?
JMIR Mhealth Uhealth 2026;14:e68735
URL: https://mhealth.jmir.org/2026/1/e68735
doi:10.2196/68735
PMID:

XSL•FO
RenderX

©Arthur A Stone, Stefan Schneider, Meynard J Toledo, Raymond Hernandez. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

A Beginner's Guide to Applying Large Language Models in Behavioral Interventions

Nirali Shah[1], PT, PhD; Lorraine Buis[1], PhD; Derek Papierski[1], BS; Alexis Castellanos[1], MS; Marvin Mlakha[1], BS; Susan Murphy[1], SCD, OTR

Department of Physical Medicine and Rehabilitation, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Nirali Shah, PT, PhD
Department of Physical Medicine and Rehabilitation
University of Michigan
Ann Arbor, MI
United States
Phone: 1 8573101084
Email: niralish@med.umich.edu

Abstract

Digital behavioral interventions are increasingly used to support chronic disease self-management, yet many systems rely on predetermined content that limits personalization and sustained engagement. Large language models (LLMs) offer new opportunities to deliver conversational behavioral support. However, integrating LLMs into behavioral interventions requires careful architectural, methodological, and ethical planning, which may be challenging for researchers without formal training in artificial intelligence. This viewpoint provides a structured introduction to LLMs tailored to behavioral science. We describe foundational concepts in natural language processing and transformer-based architectures, outline the core components of LLM-based systems, including prompting strategies, context management, retrieval-augmented generation, and guardrails, and illustrate these principles through our experience integrating a proprietary LLM into a mobile self-management intervention for individuals with systemic sclerosis. Building on this case example, we propose a phased design workflow to guide early-stage development and responsible implementation, along with a decision framework to help researchers navigate scientific and logistical trade-offs between proprietary models and other alternatives. The considerations presented here are informed by formative implementation efforts and are intended to support early-stage design decisions for LLM-based behavioral interventions. As these interventions continue to evolve, rigorous evaluation and interdisciplinary collaboration will be important to ensure that these systems improve personalization and scalability while maintaining safety and scientific rigor.

(JMIR Mhealth Uhealth 2026;14:e79302) doi:10.2196/79302

KEYWORDS

behavior; digital health; intervention; large language model; LLM; behavioral intervention; self-management

Introduction

Behavioral interventions are treatment approaches designed to help individuals change behavior and develop skills to manage their chronic health conditions [1]. These interventions typically involve an interactive process between the individual and a health care provider and draw on evidence-based behavior change strategies, such as education, identification of patterns, motivation, goal setting, self-monitoring, evaluation, and reinforcement [1,2]. Despite evidence supporting their effectiveness in improving health outcomes [3-5], behavioral interventions are inconsistently implemented in primary care settings due to a shortage of trained providers, limited access, implementation costs, and insurance coverage constraints [6-8]. These challenges hinder scalability, personalization, and sustained engagement, thereby limiting the broader impact of behavioral interventions [9].

While digital technologies, including SMS text messaging, mobile health (mHealth), and web-based platforms, have been increasingly adopted for disease management [10,11], many systems rely on predetermined content or predefined if-then rules that do not adapt to changes in user behavior, engagement patterns, or contextual factors over time [12-14]. Digital interventions with limited tailoring or personalization are associated with lower engagement and ultimately reduced effectiveness in supporting sustained behavior change and disease management [12,15]. These limitations underscore the

XSL•FO
RenderX

need for digital behavioral interventions that are both scalable and responsive to individual needs as they evolve.

Natural Language Processing and Large Language Models

Artificial Intelligence (AI) refers to technologies that simulate human intelligence, including learning, reasoning, and decision-making [16]. Within this landscape, natural language processing (NLP) plays a central role in enabling machines to analyze and generate human language [17]. Historically, NLP relied on rule-based systems that used predefined grammars and pattern matching to classify text [17,18]. While effective for well-defined tasks, these systems struggled with the nuance of human conversations and were limited in their ability to generalize across different conversational contexts. For example, in the statement, "I'm exhausted, so I didn't go for a walk today," a rule-based system might flag the keywords "didn't go" and simply classify the statement as an instance of nonadherence. Subsequent statistical models, such as recurrent neural networks, improved on this by processing text sequentially, one word at a time [19]. However, these models often suffered from the "vanishing gradient" problem, where the mathematical influence of earlier words weakened by the time the model reached the end of a long or complex sentence [20]. In the example statement "I'm exhausted, so I didn't go for a walk today," the recurrent neural network model might lose the semantic link between "exhausted" and "didn't go," resulting in a fragmented analysis that overlooks the underlying cause of the missed physical activity. Within the context of a behavioral intervention, understanding the underlying factors of nonadherence, such as low motivation, emotional burnout, lifestyle demands, or symptoms (eg, pain or fatigue), is crucial for delivering appropriate behavioral support, yet such nuance is difficult to capture using fixed rules or narrowly trained models.

Recent advances in NLP center on transformer architectures, which use self-attention mechanisms to process all words in a sequence simultaneously, allowing the model to weigh the relevance of each word relative to every other word in the message, thus preserving the causal and contextual information central to behavioral interventions [21]. Thus, in the example stated earlier, "I'm exhausted, so I didn't go for a walk today," the transformer-based model is able to capture the relationship between "exhausted" and "didn't go for a walk," preserving the understanding that the individual did not go for a walk because they were exhausted. Training transformer architecture models on massive datasets has given rise to large language models (LLMs) with emergent capabilities [22]. A defining feature of LLMs is "in-context learning," which allows the model to adapt its interpretation and performance via specific instructions or examples through a technique called "prompting." This includes "zero-shot prompting" (following instructions without examples) and "few-shot prompting" (guided by a small number of task-relevant examples) [22]. This flexibility allows LLMs to be adapted to specific clinical contexts through instruction rather than requiring the model to be built or retrained from scratch. While this can be highly beneficial in behavioral health settings, applying LLMs to interventions requires careful design and oversight because model outputs are generated probabilistically and may occasionally produce inaccurate, misleading, or clinically inappropriate responses if not properly constrained. Therefore, responsible integration of AI systems, including the use of grounding strategies, guardrails, and systematic evaluation, is essential when using LLMs for health-related interventions.

Although LLMs are increasingly adopted in behavioral and nonpharmacological treatments [23], few resources provide practical guidance on the foundational concepts and their responsible integration into interventions. This viewpoint paper aims to bridge this gap by providing a perspective-driven, practice-informed, structured introduction to core concepts, informed by our firsthand experience integrating LLMs into an mHealth intervention for individuals with systemic sclerosis (SSc) called resilience-building energy management to enhance well-being (RENEW) [24].

RENEW Intervention

The RENEW intervention is a 12-week mHealth, peer-supported fatigue self-management intervention developed for individuals with SSc [24]. The original intervention was grounded in a positive psychology framework and paired participants with trained peer health coaches who provided individualized guidance to support behavior change across domains, such as physical activity, activity pacing, relaxation techniques, adaptive thinking, self-care, nutrition, and sleep, through structured, real-time video-based coaching sessions [24]. The health coaches were trained in motivational interviewing and followed a manualized protocol with ongoing supervision to support intervention fidelity. Although the original RENEW intervention demonstrated clinical benefit [24], sustaining a peer-delivered coaching model presented challenges related to personalization, scalability, cost, and long-term availability of trained coaches. These constraints motivated the integration of LLM-based systems into the RENEW platform through the RENEW-AI and RENEW via Artificial Language (RENEWAL) projects.

RENEW-AI and RENEWAL Projects

In this paper, we describe key lessons learned from integrating an LLM into the RENEW intervention and iteratively refining it to deliver behavioral coaching content aligned with our program goals. This work was carried out in 2 phases: the RENEW-AI and RENEWAL projects. In RENEW-AI, we incorporated a chat-based interface within the RENEW mobile app that used a proprietary LLM (ie, GPT-4o; OpenAI) to provide disease-related education and behavioral health coaching consistent with the principles of the original peer-led RENEW program. In this initial iteration, responses of the AI health coach were shaped through text-based prompts such that the chat-based health coach reflected the tone, scope, and boundaries of peer health coaching provided in the original RENEW intervention. We emphasized concepts of motivational interviewing, empathy, encouragement, goal reflection, and self-management strategies, while explicitly excluding medical advice, diagnosis, or treatment recommendations.

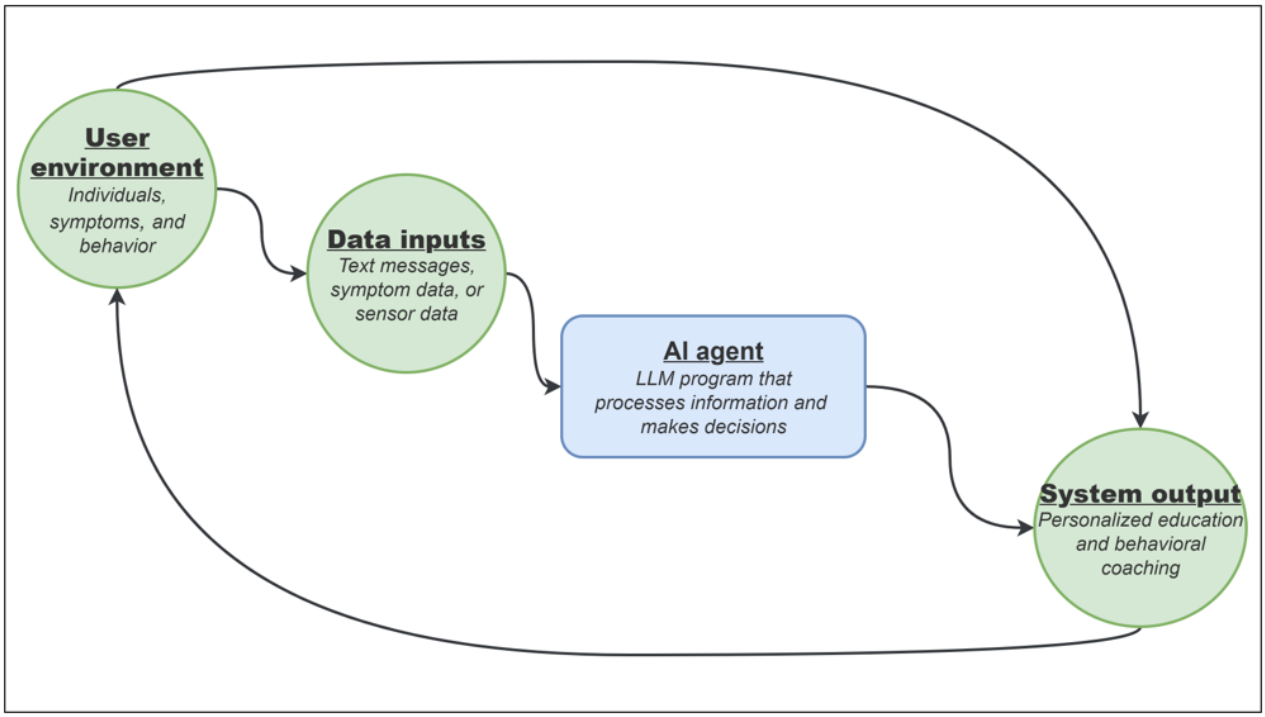
XSL•FO
RenderX

Building on this foundation, the RENEWAL project extends the system architecture with a curated, evidence-based knowledge base of disease-specific materials. This approach allowed us to ground the model's responses in program-specific content, ensuring that conversational outputs were clinically accurate. In addition to providing standardized education, the RENEWAL framework delivers personalized behavioral health coaching tailored to individuals' reported symptoms. Specifically, the system adapts behavioral goals based on the participants' reported symptoms at baseline. To ensure safety, the system is governed by guardrails that prevent the LLM from engaging in clinical decision-making or medical diagnosis.

Overview of LLM Systems

An LLM-based behavioral intervention typically consists of a user environment, data inputs, the AI agent, and system outputs (Figure 1).

Figure 1. Overview of a large language model (LLM)–based behavioral intervention. This closed loop illustrates the dynamic interaction between the user, data inputs, the artificial intelligence (AI) agent, its underlying LLM, and system outputs.



User Environment

The user environment refers to the real-world context in which an individual interacts with the LLM system [25]. The environment includes contextual factors (eg, time of day, social setting, and symptoms) and behavioral factors (eg, physical activity, daily routines, and medication adherence) that influence both health status and device use patterns. In the RENEW-AI and RENEWAL projects, the user environment includes individuals living with SSc and their lived experience of the disease, including fluctuating symptoms and behaviors, which influence their interaction patterns with the AI health coach.

Data Inputs

Data inputs refer to the information collected from the user environment that the LLM system processes to generate system outputs [25]. The inputs may be structured (eg, survey responses) or unstructured (eg, text messages). In cases where passive data streams (eg, wearable sensor data) are incorporated, a preprocessing or translation layer is required for the LLMs to convert raw time-series data into structured features or text-based summaries that can be interpreted by the model as conversational context.

In RENEW-AI, primary data inputs consist of text messages entered by individuals with SSc through the chat functionality of the RENEW mobile app. In RENEWAL, additional input includes user-reported symptoms collected using survey data, which the system uses to provide disease-specific education and personalized health coaching.

AI Agent

An AI agent is an interactive system that receives data input, processes information, and generates output in response to user interactions [25].

In both RENEW-AI and RENEWAL, the LLM agent is a conversational health coach integrated into the RENEW mobile app. The function of the agent is determined by its underlying LLM program, components of which are described in the following subsection.

System Output

System output refers to the responses generated by the LLM agent and delivered to the user [25]. In LLM-based systems, system outputs are typically conversational text messages.

In both RENEW-AI and RENEWAL, system outputs include health coaching text responses designed to support reflection,

XSL•FO
RenderX

reinforce self-efficacy, and guide self-management behaviors. The system outputs of the LLM agent influence the user environment by shaping users' perceptions, motivation, and actions, thereby completing a feedback loop in which subsequent interactions generate new data inputs for the system.

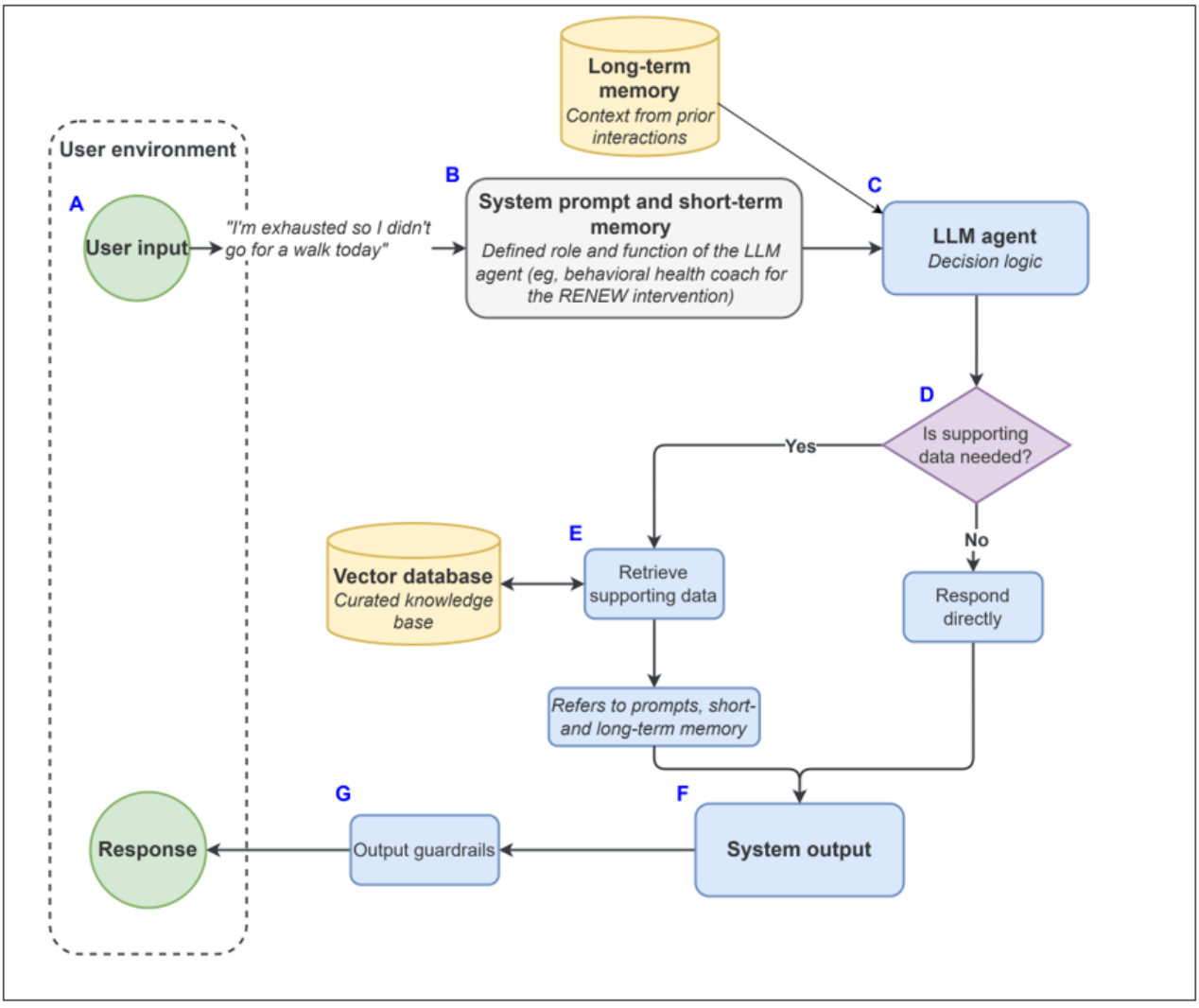
LLM Agent and Programs

Overview

The operational architecture of an LLM-based behavioral intervention is driven by the LLM agent, the system's decision-making logic, and its underlying program, which provides the computational framework to generate natural language. While the agent serves as the system's "reasoning" layer, the underlying LLM program generates language by estimating the probability of the next unit of text (token) based on the preceding user context [21,26].

To describe the role of LLM agents and their underlying programs, we present the architectural configuration adopted in the RENEW-AI and RENEWAL projects. This configuration serves as a reproducible template for the design and implementation of future LLM-based behavioral interventions. As shown in Figure 2, the system workflow progresses through a series of components, from user input (A) to system output and guardrails (F-G).

Figure 2. Architectural workflow of the large language model–based behavioral intervention. Labeled components (A-F) correspond to key stages in the system, including (A) user input, (B) memory integration, (C) large language model (LLM) agent, (D) decision-making, (E) retrieval-augmented generation, (F) system output, and (G) output guardrails. RENEW: Resilience-Building Energy Management to Enhance Well-being.



User Input

This process is initiated within the user environment when a user provides input (eg, "I'm exhausted so I didn't go for a walk today"). This input serves as the primary data source for ongoing processing by the LLM system.

Prompting and Memory Integration

The user input is combined with system-level instructions and contextual information.

System-level instructions are given through a process called "prompting." Prompting refers to the specification of instructions, roles, tone, boundaries, and output expectations provided to an LLM [27,28]. Prompting can be done either

XSL•FO
RenderX

using instructions alone (zero-shot prompting) or using a combination of instructions and examples of input-output pairs (few-shot prompting). This approach allows the model to respond in a consistent manner without modifying the model's underlying parameters [22]. For behavioral interventions, prompt design may include the LLM's role (eg, behavioral health coach using motivational interviewing principles), interaction style (eg, empathic and reflective), and explicit boundaries (eg, avoiding medical advice or diagnosis). In the RENEW-AI and RENEWAL projects, the system prompts defined the model as a supportive peer health coach that used motivational interviewing techniques and reflective listening. The prompts also included explicit instructions to avoid making medical diagnosis or treatment recommendations.

Contextual information is provided to the system using short-term memory (current conversation history) and long-term memory (user summaries from prior interactions). The context window in an LLM system refers to the model's working memory, that is, the total volume of information the model can consider at a given time during a conversation. This typically includes data inputs, prompt instructions, context from the current conversation, and context from accumulated conversation history. Information that exceeds the context window is no longer available to the model's immediate processing unless it is explicitly reintroduced or stored through system-level mechanisms. For behavioral interventions, the standard context window (short-term memory) may be insufficient. Consequently, LLM systems may incorporate long-term memory mechanisms, such as external databases to store past interactions, allowing the agent to selectively retrieve and use relevant historical context in conversations [26]. While user summaries can improve contextual continuity in conversations, overreliance on user summaries may introduce the risk of information degradation over time. Repeated summarization of prior interactions may lead to loss of nuance, emotional context, or specific behavioral details, which are often critical in behavioral health settings. Therefore, systems incorporating long-term memory should consider strategies to preserve important contextual information, such as periodic reference to current interaction data or structured storage of key variables [26].

In the RENEWAL project, the model's context window includes the system prompt (defining the health coach's role and persona), the current session's conversation history (short-term memory), and retrieval-augmented generation (RAG)–retrieved documents (disease-specific information pulled from a curated knowledge base). This is supplemented by long-term memory in the form of user summaries, which are a running update of the user's history, health goals, symptoms, and recurring topics. This summary is stored in the database and included in the context window for future sessions.

LLM Agent and Decision-Making

The user input integrated with system-level prompts and contextual memory is processed by a large neural network (the LLM agent), which uses its transformer architecture to generate contextually relevant text responses [21]. At this stage, the system determines whether additional supporting information is required to generate an appropriate response.

In the RENEW-AI and RENEWAL projects, rather than training a model from scratch, we used a proprietary GPT model available through our university (University of Michigan; UM). While the full architectural background of these models is not yet disclosed, it is known that the GPT family is built upon transformer-based systems described above. To integrate UM-GPT4o into our intervention, we accessed it through application programming interfaces (APIs), which act as a secure data bridge between the mobile app or mHealth intervention and the institutional AI platform. Using a proprietary model via API allowed our study participants to have bidirectional conversations with our LLM-based health coach while substantially reducing the technical complexity, development time, and computational cost for our project. Because the base model's internal weights were fixed, we focused on safety, clinical relevance, and narrowing the model's general capabilities to the specific goals of our intervention through refinement of instructional prompts, grounding data, and output guardrails.

RAG and the Vector Database

RAG is an architectural approach that supplements an LLM's internal knowledge with an external library or curated knowledge base [29]. This mechanism allows the system to ground its responses in verified intervention materials, thus maintaining clinical accuracy and mitigating the risk of hallucinations (when a model produces factually incorrect information). When the LLM agent receives data input from the user and determines that additional supporting information is needed, it initiates a multistep retrieval process. First, the system converts the user's text into a numerical representation called an embedding. This embedding is then used to perform a similarity search within a vector database containing the preindexed RENEW program materials and user summaries. The system then retrieves the most relevant "chunks" of text, often referred to as the "top-k" units, and appends them to the prompt. This grounding process increases the likelihood that responses are informed by established clinical protocols and historical user context rather than general probabilistic patterns. However, RAG does not fully eliminate hallucinations, as models may still misinterpret retrieved information or fail to appropriately incorporate provided context. Therefore, RAG can be viewed as a mitigation strategy that improves response reliability but does not replace the need for additional safeguards, such as prompt constraints and output guardrails.

System Output

Based on the user input and integrated context (prompt, memory, and optional retrieved content), the LLM generates a response or system output. In the RENEW framework, the system output was intended to support reflection, reinforce self-efficacy, and guide behavioral decision-making.

Output Guardrails

The output guardrails refer to mechanisms designed to constrain the system's response to ensure safety, clinical appropriateness, and adherence to ethical boundaries. Because LLMs generate

XSL•FO
RenderX

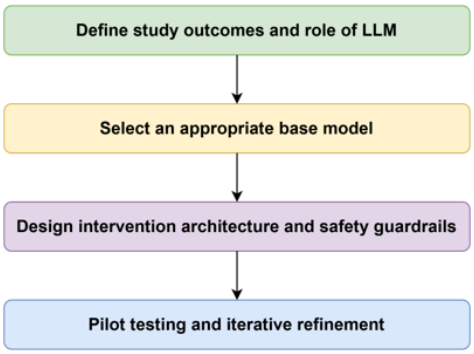
responses probabilistically, they are prone to hallucinations, but using guardrails at the levels of data input, system prompts, and output can help reduce the risk of hallucinations.

In the RENEW framework, data inputs are scanned for indicators of self-harm (guardrails at the level of data input). If high-risk language is detected, the system directs the user to appropriate professional resources or emergency services. At the prompt level, safety is maintained through prompts (ie, role-definition and RAG-grounding). Finally, at the output level, system responses are flagged if clinical recommendations fall outside RENEW's scope (eg, specific dosage recommendations or diagnostic statements). Together, this ensures that the system output remains ethically sound and clinically safe. The system output completes the feedback loop by providing a grounded, supportive coaching response, which influences the user's behavior and thereby the user's next interaction with the LLM.

A Practical Design Workflow for LLM-Based Behavioral Interventions

Building on our experience with the RENEW-AI and RENEWAL projects, we propose a 4-phase design workflow to guide the development of LLM-based behavioral interventions (Figure 3).

Figure 3. Design workflow for large language model (LLM)–integrated behavioral interventions.



At the time of writing, our work reflects early implementation efforts, including system design, pilot testing, and initial participant engagement, rather than completed efficacy trials. Accordingly, we present this workflow as a structured approach to early-stage development and responsible integration.

Define Study Outcomes and the Role of LLM

The first step in integrating an LLM into a behavioral intervention is to clarify its intended behavioral role. Researchers should define the specific function of the model, such as providing reflective coaching, delivering structured education, reinforcing goal setting, or supporting motivational processes. Equally important is explicitly stating what the model is not intended to do, including offering medical diagnoses or treatment recommendations. In parallel, researchers should define the study's primary and secondary outcomes and articulate how LLM integration is hypothesized to influence those outcomes. These may include direct clinical outcomes, such as, fatigue severity, physical activity, and causal mechanisms (eg, engagement, self-efficacy, or adherence).

Selection of an Appropriate Base Model

Once the LLM's role and outcomes are clearly defined, researchers can determine the appropriate technical strategy. This includes selecting the appropriate LLM for the behavioral intervention. As described earlier, our team adapted a proprietary LLM to function as a digital health coach for individuals with SSc [24]. One of the most consequential decisions in this process was selecting the type of LLM to use.

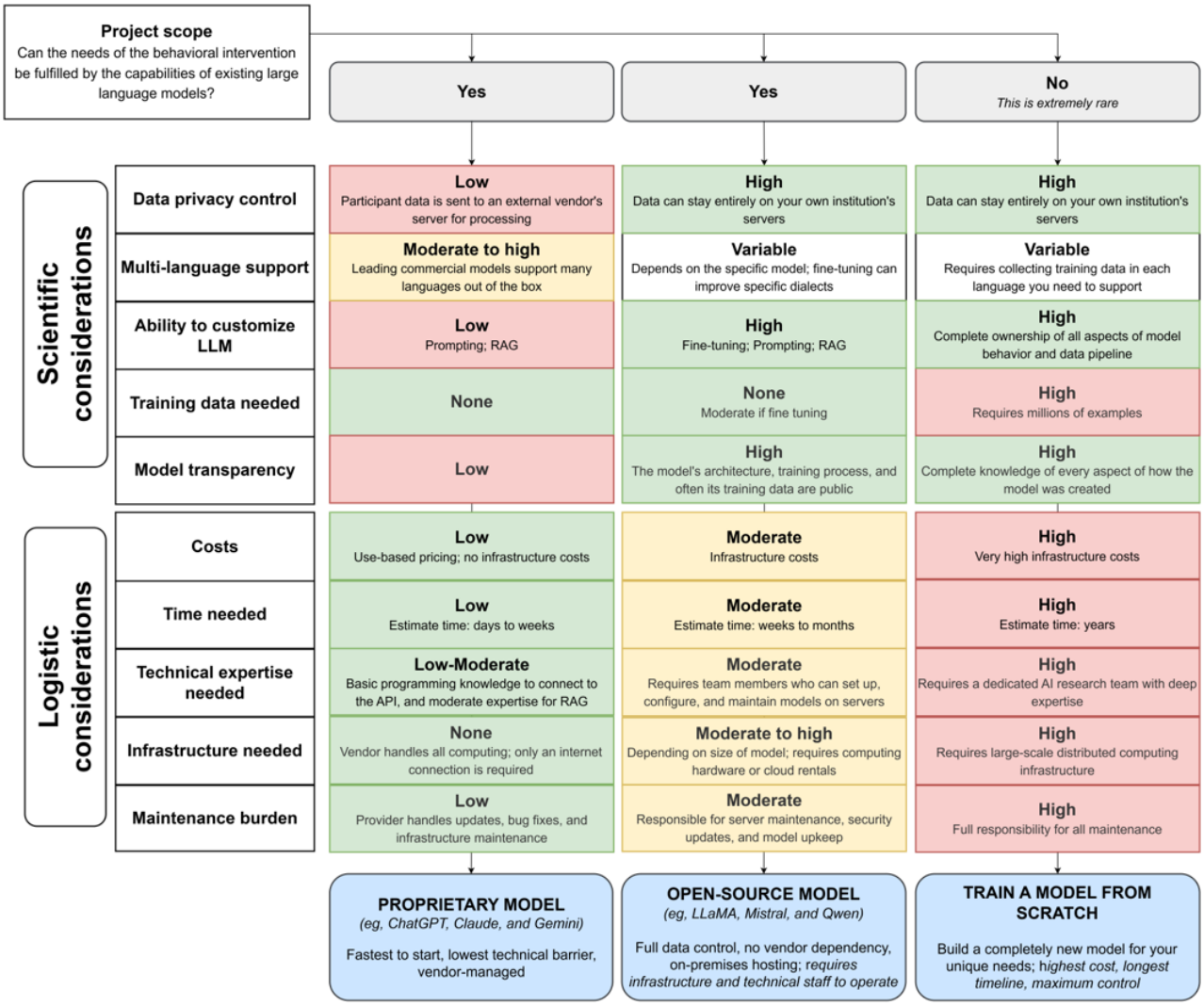
For our project, we began by asking, "Can the needs of our behavioral intervention be fulfilled by the capabilities of existing LLMs?" From there, we reflected on a series of scientific and logistical considerations. Scientifically, we considered issues such as data privacy, control over participant interaction data, the language and tone required for the intervention, the specific role the LLM was expected to play, the extent to which the model would need to be customized, training data needed to refine the model's response, and finally, implications for transparency and reproducibility of our scientific methods. Logistically, we considered practical constraints including cost and available research funding, development timelines, access to technical expertise, infrastructure requirements, and our capacity to host and maintain an LLM-based intervention over time. Together, these considerations shaped our decision-making process and informed how we ultimately integrated UM-GPT4o into the RENEW platform.

As shown in Figure 4, integrating LLMs into behavioral interventions typically involves choosing among 3 approaches: using a proprietary model (eg, GPT-4o and Google Gemini), fine-tuning and hosting an open-source model (eg, Llama and Mistral), or training a foundational model from scratch. The appropriate choice depends on the intervention's goals, privacy requirements, available expertise, and resource constraints. For many behavioral research teams, adapting an existing proprietary or open-source model is the most practical and scientifically reasonable option. Although training a model from scratch offers the highest degree of control over model behavior and data pipelines, it generally requires substantial data, infrastructure, and specialized expertise. In contrast, hosting and fine-tuning an open-source model allows data to remain on institutional servers and provides greater control over model versioning and updates, which may support reproducibility in research settings. This approach, however, requires institutional infrastructure and technical personnel capable of deploying, maintaining, and securing the system over time. Finally, proprietary models have

XSL•FO
RenderX

the lowest technical barrier and minimal infrastructure requirements. They can typically be integrated into existing mHealth platforms relatively quickly and without significant upfront investment in computing resources, which makes them well-suited for pilot studies or early-stage development. However, this convenience may come with reduced transparency regarding model architecture and training data, and less direct control over how updates are managed. Therefore, selecting an appropriate LLM is a methodological decision that balances scientific and logistical considerations within the realities of a given intervention.

Figure 4. Considerations for deciding what type of large language model (LLM) is most suited for a behavioral intervention. AI: artificial intelligence; API: application programming interface; RAG: retrieval-augmented generation.



Design Intervention Architecture and Safety Guardrails

After selecting the model, attention shifts to architectural design. This includes developing prompting strategies that define the LLM's role, tone, and boundaries, as well as implementing guardrails to mitigate risks such as inappropriate medical advice or failure to detect crisis language. Furthermore, decisions related to short-term context windows, long-term memory mechanisms (eg, user summaries), and RAG need to be made. These architectural choices directly influence personalization, safety, and reproducibility.

Regarding safety guardrails, researchers should consider safety risks related to inappropriate or out-of-scope responses and the potential for algorithmic bias arising from training data used in both proprietary and open-source models. Because LLMs are trained on large-scale datasets that may reflect historical and societal biases, model outputs may differentially represent or respond to users based on factors such as language, culture, gender, or health status. In behavioral health contexts, such biases may influence user engagement, perceived empathy, and the relevance of behavioral recommendations, potentially contributing to inequities in intervention delivery and outcomes. Therefore, researchers should consider strategies to identify and mitigate bias, including diverse evaluation samples, human review processes, and ongoing monitoring of model performance across user subgroups.

Pilot Testing and Iterative Refinement

The final phase involves pilot testing, evaluation, and iterative refinement of the intervention. Depending on the stage of development, researchers may consider feasibility studies, pilot randomized trials, or hybrid effectiveness-implementation

XSL•FO
RenderX

designs to assess clinical outcomes and implementation factors, such as adoption, usability, and sustainability. Aligning early system evaluation with these study designs may support the translation of LLM-based interventions into rigorous behavioral and clinical research. A human-in-the-loop review may be important in early implementation to assess response quality, empathic tone, clinical alignment, and safety adherence. Additionally, user-centered design, such as usability testing and iterative feedback from users, may help ensure that these interventions are acceptable, understandable, and responsive to the needs of the target population.

While evaluation metrics for LLM-based behavioral interventions are still evolving, a combination of automated technical assessment with human-centered evaluation is increasingly used in practice [30,31]. For the RENEW intervention, automated evaluation included a structured assessment of response quality to examine relevance, hallucination, and completeness of retrieved data. In parallel, human evaluation included clinician review of responses for fidelity to behavioral strategies (eg, alignment with motivational interviewing principles), as well as assessment of empathic tone, clarity, and clinical appropriateness.

Conclusions

LLMs offer new opportunities to enhance personalization and adaptability in digital behavioral interventions. However, integrating LLMs into behavioral interventions requires deliberate architectural, methodological, and ethical decision-making. This viewpoint provides a structured synthesis of LLM foundations tailored to behavioral science, translates technical concepts into intervention-relevant design considerations, and grounds these principles in an applied case example. By combining conceptual explanation with practical implementation experience, we aim to offer behavioral researchers a clear and actionable framework for early-stage LLM integration. Along with these strengths, several limitations should be acknowledged. First, the RENEW-AI and RENEWAL projects are in the early stages of implementation, with formal evaluations of effectiveness, safety, and long-term engagement ongoing and therefore not included in this paper. Second, this framework focuses specifically on LLM-based systems and may not generalize to other AI approaches, such as reinforcement learning or unsupervised modeling. Third, although we describe architectural safeguards and grounding mechanisms, ongoing monitoring for model drift, hallucinations, and potential biases across demographic groups is not elaborated on in this paper.

Acknowledgments

The authors would like to thank Dr Amy Bucher, Chief Behavioral Officer at Lirio, for generously reviewing our manuscript and providing valuable feedback on its technical accuracy. ChatGPT (GPT-4o; OpenAI) was used to assist with improving the clarity and flow of the written text.

Data Availability

The manuscript discusses the author's experiences in integrating artificial intelligence (AI) into a behavioral intervention. No datasets were used in the preparation of this work. There are no associated datasets available for this manuscript.

Funding

NS is funded by the Rheumatology Research Foundation Scientist Development Award (Identifier No. 1346603). No other external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Authors' Contributions

NS conceptualized the paper and drafted the original manuscript. DP and MM assisted with the visualizations. AC contributed to the technical accuracy of the content, and SM and LB reviewed and provided feedback on the overall manuscript. All authors have read and approved the final version of the manuscript.

Conflicts of Interest

LB, the second author of this manuscript, serves as editor-in-chief of *JMIR mHealth* and *uHealth*. All other authors declare no other conflicts of interest.

References

1. Benzo RP. Self-management programs and the pursuit of behavior change. Respir Care 2024 May 28;69(6):678-685. [doi: 10.4187/respcare.11987] [Medline: 38806226]
2. Erhard P, Wong T, Barnett M, Falcomata TS, Lang R. Self-management skills and applied behavior analysis. In: Matson JL, Sturmey P, editors. Handbook of Autism and Pervasive Developmental Disorder. Cham, Switzerland: Springer; 2022.

XSL•FO
RenderX

3. Hobby J, Crowley J, Barnes K, Mitchell L, Parkinson J, Ball L. Effectiveness of interventions to improve health behaviours of health professionals: a systematic review. BMJ Open 2022 Sep 26;12(9):e058955 [FREE Full text] [doi: 10.1136/bmjopen-2021-058955] [Medline: 36167392]
4. Mair JL, Salamanca-Sanabria A, Augsburger M, Frese BF, Abend S, Jakob R, et al. Effective behavior change techniques in digital health interventions for the prevention or management of noncommunicable diseases: an umbrella review. Ann Behav Med 2023 Sep 13;57(10):817-835 [FREE Full text] [doi: 10.1093/abm/kaad041] [Medline: 37625030]
5. Silva CC, Presseau J, van Allen Z, Schenk PM, Moreto M, Dinsmore J, et al. Effectiveness of interventions for changing more than one behavior at a time to manage chronic conditions: a systematic review and meta-analysis. Ann Behav Med 2024 May 23;58(6):432-444 [FREE Full text] [doi: 10.1093/abm/kaae021] [Medline: 38721982]
6. Carroll KM. Lost in translation? Moving contingency management and cognitive behavioral therapy into clinical practice. Ann N Y Acad Sci 2014 Oct;1327(1):94-111 [FREE Full text] [doi: 10.1111/nyas.12501] [Medline: 25204847]
7. Krist AH, Baumann LJ, Holtrop JS, Wasserman MR, Stange KC, Woo M. Evaluating feasible and referable behavioral counseling interventions. Am J Prev Med 2015 Sep;49(3 Suppl 2):S138-S149 [FREE Full text] [doi: 10.1016/j.amepre.2015.05.009] [Medline: 26296548]
8. Rapfogel N. The behavioral health care affordability problem. Center for American Progress. 2022 May 26. URL: https://www.americanprogress.org/article/the-behavioral-health-care-affordability-problem/ [accessed 2026-04-15]
9. Hagger MS, Weed M. DEBATE: do interventions based on behavioral theory work in the real world? Int J Behav Nutr Phys Act 2019 Apr 25;16(1):36 [FREE Full text] [doi: 10.1186/s12966-019-0795-4] [Medline: 31023328]
10. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
11. Moschonis G, Siopis G, Jung J, Eweka E, Willems R, Kwasnicka D, et al. Effectiveness, reach, uptake, and feasibility of digital health interventions for adults with type 2 diabetes: a systematic review and meta-analysis of randomised controlled trials. Lancet Digit Health 2023 Mar;5(3):e125-e143 [FREE Full text] [doi: 10.1016/S2589-7500(22)00233-3] [Medline: 36828606]
12. Lee JA, Choi M, Lee SA, Jiang N. Effective behavioral intervention strategies using mobile health applications for chronic disease management: a systematic review. BMC Med Inform Decis Mak 2018 Feb 20;18(1):12 [FREE Full text] [doi: 10.1186/s12911-018-0591-0] [Medline: 29458358]
13. Murray E. Web-based interventions for behavior change and self-management: potential, pitfalls, and progress. Med 2 0 2012 Aug 14;1(2):e3 [FREE Full text] [doi: 10.2196/med20.1741] [Medline: 25075231]
14. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
15. Tong HL, Quiroz JC, Kocaballi AB, Fat SC, Dao KP, Gehringer H, et al. Personalized mobile technologies for lifestyle behavior change: a systematic review, meta-analysis, and meta-regression. Prev Med 2021 Jul;148:106532 [FREE Full text] [doi: 10.1016/j.ypmed.2021.106532] [Medline: 33774008]
16. Stryker C, Kavlakoglu E. What is artificial intelligence (AI)? IBM. URL: https://www.ibm.com/think/topics/artificial-intelligence [accessed 2026-04-15]
17. Chowdhary KR. Natural language processing. In: Fundamentals of Artificial Intelligence. Berlin, Germany: Springer; 2020:603-649.
18. Martin JH, Jurafsky D. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. London, UK: Pearson Prentice Hall; 2008.
19. Graves A. Supervised sequence labelling. In: Supervised Sequence Labelling with Recurrent Neural Networks. Berlin, Germany: Springer; Feb 06, 2012.
20. Goldberg Y. Neural Network Methods for Natural Language Processing. San Rafael, CA: Morgan & Claypool Publishers; Apr 17, 2017.
21. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. ArXiv. Preprint posted online on June 12, 2017 . [doi: 10.48550/arXiv.1706.03762]
22. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language models are few-shot learners. ArXiv. Preprint posted online on May 28, 2020 . [doi: 10.48550/arXiv.2005.14165]
23. Shah N, Castellanos A, Chen Y, Piette J, Bucher A, Murphy S. A scoping review on artificial intelligence-supported interventions for nonpharmacologic management of chronic rheumatic diseases. Arthritis Care Res (Hoboken) 2026 Feb;78(2):270-281. [doi: 10.1002/acr.25612] [Medline: 40654107]
24. Murphy SL, Chen YT, Alore M, Hicks S, Pape A, Hassett AL, et al. Effects of a resilience-building energy management program on fatigue and other symptoms in systemic sclerosis: a randomized controlled trial. Arthritis Care Res (Hoboken) 2024 Mar;76(3):318-327 [FREE Full text] [doi: 10.1002/acr.25253] [Medline: 37846437]
25. Russell SJ, Norvig P. Artificial Intelligence: A Modern Approach. Upper Saddle River, NJ: Prentice Hall Press; 2009.
26. Weng L. LLM powered autonomous agents. Lil'Log. 2023. URL: https://lilianweng.github.io/posts/2023-06-23-agent/ [accessed 2026-04-15]

XSL•FO
RenderX

27. Liu P, Yuan W, Fu J, Jiang Z, Hayashi H, Neubig G. Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing. ACM Comput Surv 2023 Jan 16;55(9):1-35. [doi: 10.1145/3560815]
28. Reynolds L, McDonell K. Prompt programming for large language models: beyond the few-shot paradigm. In: Proceedings of the Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI EA '21; May 8-13, 2021; Yokohama, Japan. [doi: 10.1145/3411763.3451760]
29. Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. ArXiv. Preprint posted online on May 22, 2020 . [doi: 10.48550/arXiv.2005.11401]
30. Abbasian M, Khatibi E, Azimi I, Oniani D, Shakeri Hossein Abad Z, Thieme A, et al. Foundation metrics for evaluating effectiveness of healthcare conversations powered by generative AI. NPJ Digit Med 2024 Mar 29;7(1):82 [FREE Full text] [doi: 10.1038/s41746-024-01074-z] [Medline: 38553625]
31. Lai X, Lai Y, Chen J, Huang S, Gao Q, Huang C. Evaluation strategies for large language model-based models in exercise and health coaching: scoping review. J Med Internet Res 2025 Oct 14;27:e79217 [FREE Full text] [doi: 10.2196/79217] [Medline: 41086432]

Abbreviations

AI: artificial intelligence
API: application programming interface
LLM: large language model
mHealth: mobile health
NLP: natural language processing
RAG: retrieval-augmented generation
RENEW: resilience-building energy management to enhance well-being
RENEWAL: RENEW via Artificial Language
SSc: systemic sclerosis
UM: University of Michigan

Edited by A Coristine; submitted 18.Jun.2025; peer-reviewed by R Kravitz, J Kullgren, C Emezue; comments to author 20.Oct.2025; revised version received 08.Apr.2026; accepted 08.Apr.2026; published 29.Apr.2026.

Please cite as:
Shah N, Buis L, Papierski D, Castellanos A, Mlakha M, Murphy S
A Beginner's Guide to Applying Large Language Models in Behavioral Interventions
JMIR Mhealth Uhealth 2026;14:e79302
URL: https://mhealth.jmir.org/2026/1/e79302
doi:10.2196/79302
PMID:

©Nirali Shah, Lorraine Buis, Derek Papierski, Alexis Castellanos, Marvin Mlakha, Susan Murphy. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 29.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Applications and Current Landscape of Digital Health Technologies in the Continuity of Care for Liver Transplant Candidates and Recipients: Scoping Review

Wanqi Shi[1*], MSc; Jiao Sun[2*], PhD; Jiahao Li[2], MSc; Yuanyuan Li[2], MSc; Feiyan Yuan[2], MSc; Ruiya Li[2], MSc; Zijing Chu[1], MSc

[1]Department of Hepatology, The First Hospital of Jilin University, 1 Xinmin Street, Chaoyang District, Changchun, Jilin, China

[2]School of Nursing, Jilin University, Changchun, Jilin, China

[*]these authors contributed equally

Corresponding Author:
Zijing Chu, MSc
Department of Hepatology, The First Hospital of Jilin University, 1 Xinmin Street, Chaoyang District, Changchun, Jilin, China

Abstract

Background: Liver transplant candidates and recipients require continuous care throughout the transplant trajectory, including pretransplant prehabilitation and long-term posttransplant management. Conventional care models are constrained by time, distance, and limited health care resources and may therefore be insufficient to meet patients' needs across the continuum of care. Digital health technologies offer new approaches to remote care delivery and self-management support in liver transplantation.

Objective: This scoping review aimed to map the technological modalities, intervention components, outcome measures, and implementation issues associated with digital health technologies used across the care continuum for adult liver transplant candidates and recipients and to identify gaps for future digital continuity-of-care research.

Methods: Nine English- and Chinese-language databases—PubMed, Embase, CINAHL, Web of Science, the Cochrane CENTRAL, CNKI, Wanfang Data, VIP Database, and SinoMed—were searched from inception through July 10, 2026. Original quantitative, qualitative, and mixed methods studies involving adult liver transplant candidates or recipients and the use of digital health technologies for care beyond conventional in-person settings were eligible. The review followed the Arksey and O'Malley framework, was refined using the Joanna Briggs Institute methodology for scoping reviews, and was reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines.

Results: The search identified 7746 records, and 28 studies were included. Telehealth platforms or remote management systems were used in 24 studies, and remote education was included in 22 studies. From the 28 studies, 5 focused on pretransplant/candidacy, 9 focused on early posttransplant recovery, 4 focused on long-term survivorship, 8 were cross-stage/mixed, and 2 had unspecified posttransplant timing. No study focused only on the perioperative/inpatient-to-home transition stage. A total of 9 randomized intervention studies were identified. Among these studies, positive findings were more common for medication or treatment adherence, self-management, health behaviors, psychological outcomes, and selected health service outcomes. Results were less consistent for quality of life, physiological outcomes, and functional outcomes. From the 28 studies, 15 reported implementation-related outcomes. Reported barriers included limited device access, privacy concerns, limited uptake, and difficulties with sustained engagement.

Conclusions: Current evidence is concentrated in early posttransplant recovery. Evidence across transplant stages and outcome domains is uneven. Studies focused on the perioperative/inpatient-to-home transition stage are lacking. Future studies should use multicenter designs and standardized outcome measures; include longer follow-up; and evaluate implementation, equitable access, and costs.

(JMIR Mhealth Uhealth 2026;14:e106470) doi:10.2196/106470

KEYWORDS

digital health; mobile health; mHealth; eHealth; telemedicine; liver transplantation; mobile apps; continuity of patient care; scoping review; Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PRISMA

XSL•FO
RenderX

Introduction

Background

Liver transplantation (LT) is a definitive treatment for a range of end-stage liver diseases [1,2]. Advances in surgical techniques, perioperative management, and immunosuppressive therapy have substantially improved posttransplant survival [3]. A 50-year dataset from the European Liver Transplant Registry, covering 169 centers in 32 countries, showed overall survival rates of 83% at 1 year and 61% at 10 years after LT [4]. Despite its lifesaving potential, LT entails lifelong immunosuppressive therapy and persistent risks of acute rejection and multisystem complications [5,6]. Many recipients also face substantial financial burdens, barriers to resuming social roles, and other psychosocial and practical challenges [7]. Long-term health management across the posttransplant trajectory is therefore essential. Continuity-of-care interventions may strengthen self-management, support rehabilitation and social reintegration, and improve quality of life (QoL) among liver transplant recipients [8].

Continuity-of-care extends professional health care services from the hospital to patients' homes and communities, enabling patients to receive coordinated and ongoing support after discharge and helping to address their out-of-hospital care needs [9,10]. Evidence suggests that high-quality continuity-of-care interventions can improve self-management and, ultimately, health outcomes and QoL [11]. In LT, the care continuum spans the pretransplant waiting period, the early posttransplant transition, out-of-hospital rehabilitation, and lifelong follow-up [12]. Liver transplant candidates may experience advanced disease, complications, and psychological distress, requiring continued monitoring and support [13,14]. After transplantation, recipients remain at risk of rejection and other short-term and long-term complications and require continued follow-up and self-management support [15]. Including both candidates and recipients allows digital health applications to be examined across the full transplant trajectory. Current continuity-of-care models for liver transplant candidates and recipients continue to rely primarily on outpatient follow-up, supplemented by telephone follow-up and home visits [6]. These approaches are constrained by high outpatient workloads, geographic and temporal barriers, incomplete follow-up information, and substantial time and financial costs [16].

Digital health technologies, including telehealth, mobile apps, wearable devices, and remote monitoring systems, provide new ways to deliver care outside conventional clinical settings [17-19]. In transplantation, these technologies can support remote monitoring, follow-up, education, and communication between patients and health care professionals [20]. Evidence from solid organ transplantation suggests that digital health may support medication adherence, self-management, and selected clinical outcomes [21]. A recent scoping review by Gao et al [22] summarized the modalities, intervention components, and outcomes of telehealth used in home-based care for adult liver transplant recipients, providing an initial evidence base for the use of digital technologies in this field.

Despite growing evidence of the potential value of mobile health (mHealth) in the continuity of care for LT, existing research has focused predominantly on posttransplant recipients and has largely consisted of single-center studies with small samples. The literature is further characterized by limited use of theoretical frameworks, fragmented technological approaches, substantial heterogeneity in intervention components, and insufficient coverage of the full transplant trajectory.

Objectives

This scoping review aimed to map the current use of digital health technologies in continuity of care across the liver transplant trajectory. Specifically, we sought to (1) identify the principal technological modalities, intervention components, and outcome measures reported for pretransplant candidates and posttransplant recipients; (2) examine the advantages, implementation limitations, and appropriate care contexts of different digital health approaches; and (3) identify gaps in theoretical grounding, intervention design, outcome evaluation, and clinical translation to inform the development of specialty-specific digital continuity-of-care models for LT.

Methods

Study Design

This scoping review systematically mapped and synthesized evidence on the use of digital health technologies to support continuity of care across the liver transplant trajectory. The review was conducted according to the framework developed by Arksey and O'Malley [23], informed by updated methodological guidance from the Joanna Briggs Institute (JBI) [24], and reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guideline (Checklist 1) [25]. A review protocol was not prospectively registered.

Search Strategy

An initial PubMed search was used to identify relevant terms and refine the search strategy. Reference lists of relevant studies were also hand searched. A total of 9 electronic databases were searched from inception through July 10, 2026: PubMed, Embase, CINAHL, Web of Science, Cochrane CENTRAL, CNKI, Wanfang Data, VIP Database, and SinoMed. Search concepts were developed using the PCC (population, concept, and context) framework. MeSH and free-text terms were combined using Boolean operators, truncation, and wildcard symbols. The complete database-specific search strategies are provided in Multimedia Appendix 1.

Inclusion and Exclusion Criteria

The inclusion criteria were developed according to the PCC framework recommended by the JBI as follows:

- Population: studies involving liver transplant candidates or recipients aged 18 years or older were included.
- Concept: original empirical studies evaluating the application of digital health technologies in continuity of care for LT were included. Relevant technologies included mobile apps, telehealth or web-based platforms,

XSL•FO
RenderX

internet-based communication technologies, remote monitoring or wearable devices, intelligent robots, and other digital technologies used to support continuity of care. These technologies were used for education, monitoring, consultation, self-management, rehabilitation, or psychosocial support.

- Context: studies conducted across the liver transplant care continuum in clinical, home, community, or virtual settings were eligible. Care could be delivered remotely or through hybrid online and face-to-face models and could occur before or after transplantation. Eligible contexts included pretransplant care, hospital-to-home transition, postoperative recovery, and long-term follow-up.

Original empirical studies using quantitative, qualitative, or mixed methods were included. Eligible study designs included randomized controlled trials (RCTs), nonrandomized intervention studies, observational studies, feasibility studies, pilot studies, technology development and user evaluation studies, and qualitative studies.

We excluded studies that (1) were reviews, study protocols, case reports, conference abstracts, editorials, commentaries, or letters; (2) involved only liver transplant recipients younger than 18 years or did not allow separate extraction of data for adult participants; (3) involved multiple organ transplant populations and did not allow separate extraction of liver transplant–specific data; (4) used digital technologies solely for research data collection rather than for the provision of patient care, health management, or continuity of care; (5) used digital technologies only for medical imaging, disease diagnosis, surgical navigation, organ allocation, risk prediction, or other technological applications unrelated to continuity of care in LT; or (6) were duplicate publications.

Study Selection

After duplicate records were removed using EndNote 2025 (Clarivate), 2 reviewers (SWQ and LYY) independently screened titles and abstracts and subsequently assessed the full texts against the eligibility criteria. A third reviewer (LJH) verified the search strategy and the final inclusion and exclusion decisions. Disagreements were resolved through discussion among the 3 reviewers.

Data Analysis and Synthesis

Two reviewers (SWQ and LRY) independently extracted and cross-checked study-level data using a standardized Microsoft Excel worksheet developed by the review team. Extracted data included study characteristics, participants, transplant timing, intervention and comparator characteristics, digital technologies, intervention content, outcomes, study findings, and implementation-related information. Two additional reviewers (LJH and CZJ) checked the extracted data for completeness and consistency. Disagreements were resolved by discussion, with SJ acting as an adjudicator when required.

The analytical categories for technology modalities, intervention components, outcome domains, and transplant trajectory stages were not prespecified. Consistent with JBI guidance, they were developed iteratively after review of all included studies and relevant literature [24,26]. Recurring characteristics were grouped into conceptually similar categories through team discussion. Technology modalities were classified by digital delivery medium, intervention components by clinical or supportive function, and outcomes by the constructs measured. Multiple coding was allowed because studies could include more than one technology, component, or outcome.

For stage-specific mapping, an additional coding framework was developed to locate each intervention along the liver transplant trajectory. Stage assignment was based primarily on participants' transplant status at the initiation of the intervention and the intended context in which the intervention was delivered, rather than on the final follow-up time point. The pretransplant/candidacy stage included transplant evaluation, wait-list management, decision support, and prehabilitation before LT. The perioperative/inpatient-to-home transition stage encompassed interventions actively delivered during the index transplant hospitalization, postoperative inpatient recovery, discharge planning, and transition to home. Technology onboarding or training performed immediately before discharge was not, by itself, considered a perioperative intervention when the substantive digital care began after discharge. Early posttransplant recovery was operationally defined as interventions initiated after discharge and targeting care within the first posttransplant year. Long-term posttransplant survivorship was defined as interventions initiated at or beyond 1 year after LT or interventions clearly targeting stable long-term recipients whose reported posttransplant timing was consistent with a survivorship context. The 1-year threshold was used as an analytical anchor rather than a biological cutoff and was informed by posttransplant survivorship literature [27,28]. The perioperative/inpatient-to-home transition stage was defined by clinical events rather than a fixed postoperative interval [29,30]. Studies in which the intervention or study population spanned more than one stage and stage-specific findings could not be separated were classified as cross-stage/mixed. If a study involved posttransplant recipients but did not report sufficient information to determine intervention timing, the stage was coded as not specified rather than inferred.

No formal critical appraisal or risk-of-bias assessment was performed. Evidence from different study designs was synthesized descriptively and interpreted according to the study design. The findings were not used to establish evidence certainty or rank intervention effectiveness.

Ethical Considerations

Ethical approval was not required because this scoping review used data from published studies and did not involve the collection of primary data from human participants.

Results

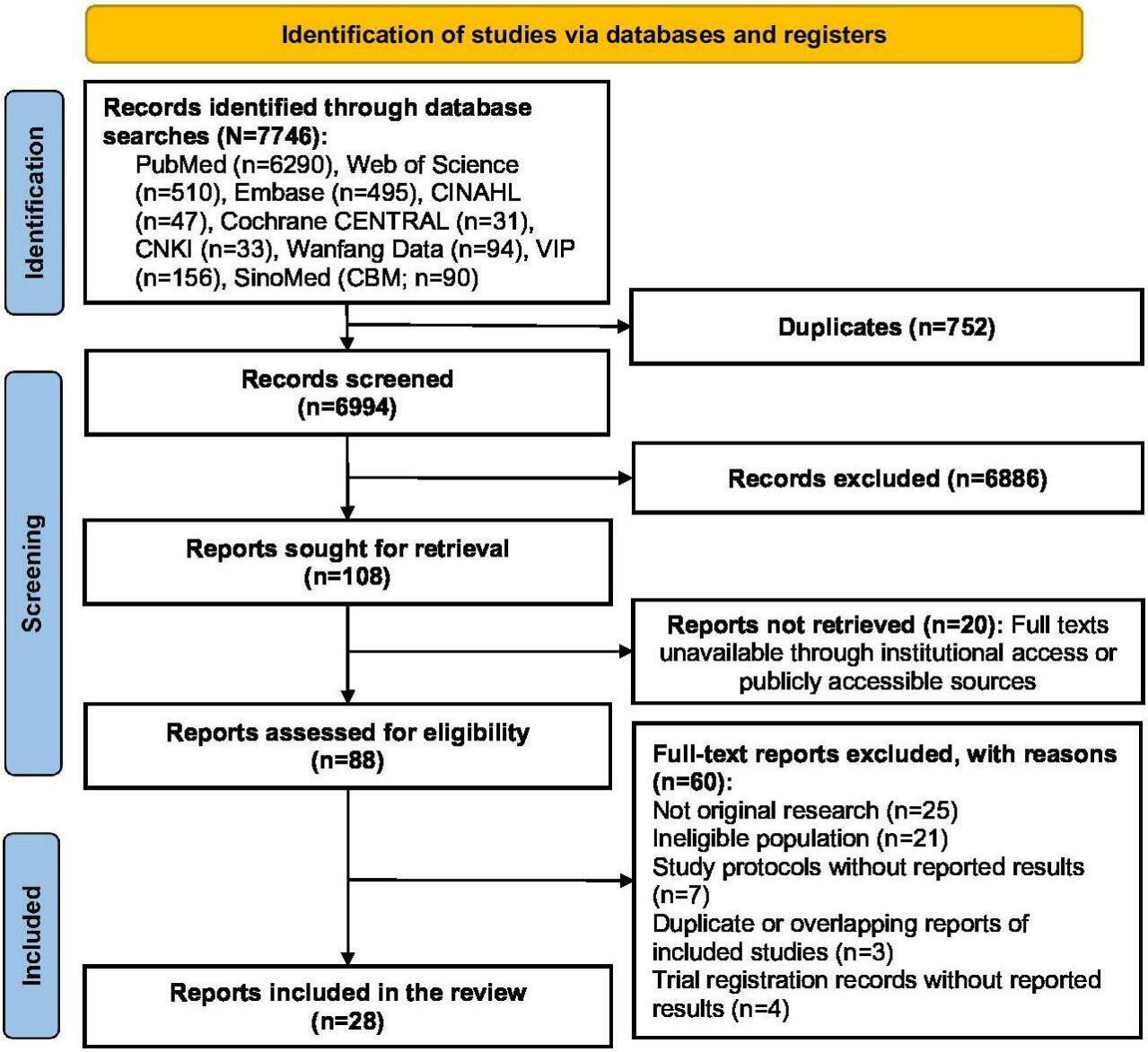
Study Selection

The database searches identified 7746 records. All records were imported into EndNote 2025, and 752 (9.71%) duplicates were removed, leaving 6994 records for title and abstract screening. Two reviewers (SWQ and LYY) independently screened these records against the eligibility criteria, and 6886 (98.46%) records were excluded. Of the 108 reports sought for retrieval, 20

XSL•FO
RenderX

full-text reports could not be obtained through the institutional or publicly accessible sources available at the time of screening, leaving 88 reports for full-text eligibility assessment. Of these, 60 were excluded because they were not original research (n=25), involved an ineligible population (n=21), were study protocols without reported results (n=7), were duplicate or overlapping reports of included studies (n=3), or were trial registration records without reported results (n=4). A total of 28 studies were included in the review. Figure 1 presents the PRISMA-ScR flow diagram.

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) flow diagram.



Study Characteristics

The extracted data were synthesized to characterize the included studies by author, publication year, country, study design, sample size, participant age, comparator, intervention technology, and reported outcomes. Because of substantial heterogeneity in study designs and outcome reporting, a meta-analysis was not appropriate; findings were therefore synthesized narratively and summarized in tables.

The characteristics of the 28 included studies are summarized in Table 1. The 28 included studies were published between 2012 and 2026, with 16 (57.1%) published in 2021 or later. The studies were conducted in seven countries. The United States and China contributed the largest number of studies (n=11 each, 39.3%), followed by Australia (n=2, 7.1%). Austria, Belgium, Brazil, and Turkey each contributed 1 (3.6%) study.

XSL•FO
RenderX

Table . General characteristics of the included studies (N=28).

Study/year/country	Study design	Transplant trajectory stage	Population	Duration	Sample size	Age
Knihs et al [31], 2024, Brazil	Technological product development and validation study	Posttransplant timing unspecified	Patients who underwent LT[a], nurses, and IT professionals	Not applicable	28 participants (17 patients, 7 nurses, and 4 IT professionals)	Patients: 16 - 76 years; professionals: 23 - 47 years
Hickman et al [32], 2019, Australia	Qualitative study	Long-term survivorship	Adult LT recipients ≥12 months posttransplant	Not applicable	28	Mean 53 (SD 13; range 24 - 68) years
Goswami et al [33], 2024, United States	Pilot study	Cross-stage/mixed	LT candidates and recipients with ALD[b]/AUD[c]	12 weeks	42 (23 pretransplant/evaluation; 19 posttransplant)	Median 49 (IQR 39 - 63) years
Andrä et al [34], 2022, Austria	Single-center cohort study	Cross-stage/mixed	Liver transplant recipients enrolled in an outpatient aftercare program	2 months	124 (42 app users; 82 nonusers)	Entire cohort: 63.2 (range 19 - 76) years; app users: 55.4 years
Delman et al [35], 2021, United States	Retrospective cohort study	Cross-stage/mixed	Adult OLT[d] recipients	6-month follow-up	344 (274 pre–COVID-19; 70 COVID-19 era)	Pre–COVID-19: mean 53.2 (SD 17.1) years; COVID-19 era: mean 56.2 (SD 10.9) years
Lee et al [36], 2019, United States	Randomized controlled trial	Early posttransplant	Adult liver transplant recipients aged 18 years or older	90-day intervention; follow-up to 180 days	106 randomized; 100 completed and analyzed (50 intervention; 50 control)	Intervention: median 58.5 (IQR 51 - 65) years; control: median 60 (IQR 56 - 65) years
Tian et al [37], 2021, China	Cohort study	Early posttransplant	Adult OLT recipients	2-week intervention; 12-month follow-up	110 enrolled; 102 analyzed (52 telemedicine; 50 usual care)	Overall: mean 46.65 (SD 9.66) years; telemedicine: mean 45.35 (SD 10.44) years; usual care: mean 48.00 (SD 8.68) years
Koc et al [38], 2022, Belgium	Prospective cohort study	Long-term survivorship	Stable adult liver transplant recipients at least 12 months posttransplantation	Median follow-up approximately 2 years	115 (87 remote monitoring; 28 standard follow-up)	Median: autonomous remote monitoring 59.1 years; nonautonomous remote monitoring 67.2; standard follow-up 66.0
Volk et al [39], 2014, United States	Pilot pre-post study	Pretransplant/candidacy	Adult liver transplant candidates	Mean completion time 15 (range 7 - 28) minutes	56 enrolled; 53 completed	Median 56 (range 20 - 67) years
Bailey et al [40], 2017, United States	Randomized controlled trial	Pretransplant/candidacy	Adult patients awaiting LT and caregivers	8-week intervention; follow-up to 12 weeks	115 (56 SMI[e]; 59 LDE[f])	Self-management intervention: mean 56.4 (SD 9.9) years; education: mean 55.6 (SD 10.0) years
Konjeti et al [41], 2019, United States	Retrospective cohort study	Pretransplant/candidacy	Adult veterans referred for OLT evaluation	Not applicable	190 (91 telehealth evaluation; 99 direct referral)	Telehealth: 60.1 years; direct referral: 60.0 years
Hickman et al [42], 2021, Australia	Randomized feasibility study	Long-term survivorship	LT recipients ≥6 months posttransplant	12 weeks	35 (23 intervention; 12 delayed-intervention control)	Intervention: 50±15 years; control: 51±15 years

XSL•FO
RenderX

Study/year/country	Study design	Transplant trajectory stage	Population	Duration	Sample size	Age
Dolanbay et al [43], 2025, Turkey	Randomized controlled trial	Cross-stage/mixed	Adult LT recipients 2 months–2 years posttransplant	1 month	90 (45 intervention; 45 control)	Intervention: mean 43.64 (SD 13.51) years; control: mean 53.24 (SD 10.52) years
Ertel et al [44], 2016, United States	Nonrandomized observational pilot study	Early posttransplant	Consecutive adult liver transplant recipients discharged home	90 days	20	Mean 56 (SD 7; range 42 - 70) years
Duarte-Rojo et al [45], 2023, United States	Single-arm pilot and feasibility trial	Pretransplant/candidacy	Adult liver transplant candidates with cirrhosis, aged 40 - 70 years	14 weeks (2-week run-in plus 12-week intervention)	31 consented; 21 started the intervention; 15 completed final assessment	Consented cohort: mean 61 (SD 7) years; intervention cohort: mean 60 (SD 7) years
DeMartini et al [46], 2018, United States	Randomized pilot feasibility trial	Pretransplant/candidacy	Prelisting LT candidates with alcohol-related liver disease	8 weeks	15 randomized (8 intervention; 7 control); 14 analyzed	Overall: mean 50.80 (SD 7.86) years; intervention: mean 49.88 (SD 6.71) years; control: mean 51.86 (SD 9.44) years
Wang et al [47], 2023, United States	Cross-sectional model-development and internal-validation study	Cross-stage/mixed	Adults with cirrhosis in LT setting, listed or history of LT	Not applicable	145	Median 60 years; frail: median 61 (IQR 56 - 67); nonfrail: median 60 (IQR 52 - 65)
Catalli et al [48], 2026, United States	Single-arm quasi-experimental pre-post study	Long-term survivorship	Adults with MASLD[g]/cirrhosis; includes post-LT subgroup	12 weeks (six 90-minute sessions every 2 weeks)	70 (45 nontransplant; 25 post–liver transplant)	Overall: median 57 (IQR 46 - 66) years; posttransplant subgroup: median 60 (IQR 50 - 65) years
Lu et al [49], 2023, China	Randomized controlled study	Cross-stage/mixed	Adult patients after LT	3 months	46 (23 intervention; 23 control)	Intervention: mean 49.1 (SD 3.7; range 41 - 55) years; control: mean 49.3 (SD 3.2; range 40 - 56) years
Wang et al [50], 2012, China	Single-group pre-post implementation study	Posttransplant timing unspecified	Liver transplant recipients using an internet information platform for follow-up management	8 months	340	Not reported
Li et al [51], 2024, China	Technological system-development and preliminary application study	Cross-stage/mixed	Organ transplant recipients of all types, including liver transplant recipients	March 2019-May 2024	1730 total, including 488 LT	Not reported; 1648 adults and 82 minors
Liu et al [52], 2019, China	Quasi-experimental controlled study	Cross-stage/mixed	Adult first-time liver transplant recipients at least 6 months after LT	6-week intervention; follow-up at 3, 6, and 12 months	63 (32 intervention; 31 control)	Not reported
Zhai et al [53], 2021, China	Randomized controlled trial	Early posttransplant	Patients after LT	6 months	58 (29 intervention; 29 control)	Intervention: mean 47.31 (SD 9.24; range 26 - 66) years; control: mean 46.97 (SD 0.40; range 25 - 67) years

XSL•FO
RenderX

Study/year/country	Study design	Transplant trajectory stage	Population	Duration	Sample size	Age
Pan et al [54], 2023, China	Randomized controlled study	Early posttransplant	Patients after LT	1 month	80 (45 intervention; 35 control)	Intervention: mean 49.38 (SD 4.86; range 35 - 63) years; control: mean 49.12 (SD 5.02; range 36 - 62) years
Sun et al [55], 2019, China	Single-group pre-post study	Early posttransplant	Patients discharged after LT	3 months	60	Not reported
Tian et al [56], 2019, China	Prospective nonrandomized controlled cohort study	Early posttransplant	Recipients of DCD[h] liver transplants	3 months	100 (50 telemedicine robot; 50 telephone follow-up)	Telemedicine robot: mean 39 (SD 5) years; telephone follow-up: mean 38 (SD 6) years
Wang et al [57], 2023, China	Randomized controlled trial	Early posttransplant	Adult first-time liver transplant recipients after surgery	6 months	92 (46 intervention; 46 control)	Overall: mean 49.88 (SD 12.95; range 37 - 65) years; intervention: mean 49.95 (SD 13.64) years; control: mean 49.31 (SD 13.05) years
Xue et al [58], 2019, China	Single-arm feasibility study	Early posttransplant	Liver transplant recipients scheduled for outpatient follow-up after discharge	2 weeks	30 enrolled; 26 completed	22 - 67 years

[a]LT: liver transplantation.

[b]ALD: alcohol-related liver disease.

[c]AUD: alcohol use disorder.

[d]OLT: orthotopic liver transplantation.

[e]SMI: self-management intervention.

[f]LDE: liver disease education.

[g]MASLD: metabolic dysfunction-associated steatotic liver disease.

[h]DCD: donation after cardiac death.

A wide range of study designs was represented, including 7 (25.0%) RCTs, 2 (7.1%) randomized feasibility/pilot feasibility trials, 6 (21.4%) cohort studies, 4 (14.3%) quasi-experimental or single-group pre-post studies, 3 (10.7%) pilot studies, 2 (7.1%) single-arm feasibility studies, 2 (7.1%) technology development and preliminary application or validation studies, 1 (3.6%) cross-sectional model development and internal validation study, and 1 (3.6%) qualitative study.

With respect to the comparative structure, 17 (60.7%) studies included a distinct comparator group, such as a randomized concurrent control group, a nonrandomized concurrent comparator group, or a historical comparator group. An additional 5 (17.9%) studies used a single-group pre-post design. The remaining 6 (21.4%) studies included neither an external comparator group nor a formal pre-post comparison and primarily focused on technology development and validation, qualitative exploration, cross-sectional model development, or single-arm feasibility and implementation assessment.

According to the transplant trajectory classification, 5 studies focused on the pretransplant/candidacy stage [39-41,45,46], 9 on early posttransplant recovery [36,37,44,53-58], 4 on long-term survivorship [32,38,42,48], 8 were classified as cross-stage/mixed [33-35,43,47,49,51,52], and 2 involved posttransplant recipients for whom intervention timing was not specified [31,50]. No study focused exclusively on the perioperative/inpatient-to-home transition stage.

Modalities and Application Contexts of Digital Health Technologies for Liver Transplant Candidates and Recipients

Among the included studies, 5 principal digital health technology modalities were identified: mHealth apps (n=8), telehealth platforms or remote management systems (n=24), intelligent robots (n=3), portable wearable or mobile devices (n=4), and internet-based communication technologies (n=20). These categories were not mutually exclusive because individual interventions could use more than one modality. Telehealth platforms or remote management systems were the most frequently reported modalities. The identified technologies were used mainly for communication, monitoring, education, follow-up, rehabilitation, and self-management support (Table 2). One qualitative study [32] did not directly implement a digital health intervention and was excluded from the modality

XSL•FO
RenderX

counts; it was retained as complementary evidence on user needs and implementation context.

Table . Digital health technology modalities and core functions.

Technology modality	Number of studies, n	Forms	Core functions
Mobile health apps	8 [31,34,36,37,45,51,54,56]	Smartphone-based or tablet-based applications, including dedicated apps and WeChat-based apps	Education, medication management, diet and exercise guidance, health monitoring; rehabilitation, and self-management support
Telehealth platforms or remote management systems	24 [33-39,41-45,47-58]	Web-based or platform-based systems integrating video visits, communication, monitoring, assessment, and feedback	Communication, remote monitoring, alerts, clinical decision support, education, follow-up, and multidisciplinary support
Intelligent robots	3 [37,56,58]	Telemedicine or remote monitoring robots	Physiological data collection, video communication, individualized guidance, and abnormality feedback
Portable wearable or mobile devices	4 [36,42,44,45]	Smartwatches, activity trackers, Bluetooth-enabled monitoring devices, and home-based exercise devices	Physiological and activity monitoring, data transmission, reminders, goal setting, and remote feedback and rehabilitation
Internet-based communication technologies	20 [33,35-37,40-42,45,46,48-58]	Videoconferencing, web portals, WeChat, text messaging, and telephone calls	Consultation, education, test-result feedback, medication support, rehabilitation and lifestyle guidance, and psychosocial support

Core Intervention Components of Digital Health Technologies for Liver Transplant Candidates and Recipients

Five major intervention components were identified: remote education (n=22), remote data monitoring or health assessment (n=16), teleconsultation (n=14), psychosocial support (n=10), and remote rehabilitation or exercise guidance (n=8). These categories were not mutually exclusive, and their distribution varied across the transplant trajectory (Table 3). Early posttransplant recovery contained the broadest combination of components: all 9 studies included remote education, and 8 each included monitoring or health assessment and teleconsultation [36,37,44,53-58]. Pretransplant/candidacy interventions focused on education and decision support, psychosocial support, prehabilitation, and remote transplant assessment [39-41,45,46]. Long-term intervention studies emphasized remote monitoring, lifestyle management, and rehabilitation [38,42,48], whereas cross-stage/mixed studies used combinations of education, monitoring, teleconsultation, and psychosocial support [33-35,43,47,49,51,52]. No study focused exclusively on the perioperative/inpatient-to-home transition stage.

XSL•FO
RenderX

Table . Intervention components across the liver transplant trajectory.

Main intervention components	Number of studies, n	Main functions/content	Transplant trajectory distribution
Remote education	22 [31,33-37,39,40,42-46,48,49,52-58]	Transplant/disease education, medication use, diet and nutrition, complication prevention, self-monitoring, lifestyle management, decision support, relapse prevention, and peer/family education	• Early posttransplant: n=9 [36,37,44,53-58] • Cross-stage/mixed: n=6 [33-35,43,49,52] • Pretransplant/candidacy: n=4 [39,40,45,46] • Long-term survivorship: n=2 [42,48] • Posttransplant timing unspecified: n=1 [31]
Remote data monitoring or health assessment	16 [34,36-38,42,44,45,47,49-51,53,54,56-58]	Vital signs, glucose and weight, physical activity, laboratory/imaging results, immunosuppressant levels, liver function, frailty, and longitudinal health records	• Early posttransplant: n=8 [36,37,44,53,54,56-58] • Cross-stage/mixed: n=4 [34,47,49,51] • Long-term survivorship: n=2 [38,42] • Pretransplant/candidacy: n=1 [45] • Posttransplant timing unspecified: n=1 [50]
Teleconsultation	14 [35-38,41,49-51,53-58]	Remote assessment and follow-up, triage, test-result feedback, medication adjustment, and individualized clinical guidance	• Early posttransplant: n=8 [36,37,53-58] • Cross-stage/mixed: n=3 [35,49,51] • Pretransplant/candidacy: n=1 [41] • Long-term survivorship: n=1 [38] • Posttransplant timing unspecified: n=1 [50]
Psychosocial support	10 [33,40,46,48,49,52-54,57,58]	Illness uncertainty, emotional support, relapse prevention, coping and behavior change, and peer/family support	• Early posttransplant: n=4 [53,54,57,58] • Cross-stage/mixed: n=3 [33,49,52] • Pretransplant/candidacy: n=2 [40,46] • Long-term survivorship: n=1 [48]
Remote rehabilitation/exercise guidance	8 [37,42,45,48,49,54,56,57]	Exercise and prehabilitation, lifestyle/weight management, rehabilitation follow-up, individualized activity goals, and family-supported rehabilitation	• Early posttransplant: n=4 [37,54,56,57] • Long-term survivorship: n=2 [42,48] • Pretransplant/candidacy: n=1 [45] • Cross-stage/mixed: n=1 [49]

Outcome Measures and Effects of Digital Health Technologies

Across the 28 included studies, reported outcomes were mapped into 6 nonmutually exclusive domains: physiological or functional indicators, psychosocial outcomes, clinical and health service utilization outcomes, self-management-related outcomes, health-related QoL, and implementation-related outcomes or patient experience (Table 4). Self-management-related outcomes were most frequently assessed (17/28, 60.7%), followed by implementation-related outcomes or patient experience (15/28, 53.6%) and clinical or health service utilization outcomes (10/28, 35.7%). Physiological or functional indicators and health-related QoL were each reported in 7 (25.0%) studies, whereas psychosocial outcomes were reported in 5 (17.9%) studies. Because individual studies often assessed more than one outcome domain, studies could contribute to multiple categories.

XSL•FO
RenderX

Table . Outcome domains and their distribution across the liver transplant trajectory.

Outcome domain	Number of studies, n	Main measures/outcomes	Transplant trajectory distribution	Design representation
Physiological/functional indicators	7 [38,42,44,45,47,48,54]	Tacrolimus concentration, serum creatinine, vital-sign thresholds, body weight/BMI, fasting and postprandial glucose, Liver Frailty Index, 6-minute walk test, physical activity, and metabolic indicators	• Long-term survivorship: 3 • Early posttransplant: 2; Pretransplant/candidacy: 1 • Cross-stage/mixed: 1	• Randomized: 2 • Nonrandomized: 5
Psychosocial outcomes	5 [40,46,48,53,54]	Illness uncertainty, depression, anxiety, perceived stress, alcohol craving, abstinence self-efficacy, hope, confidence in behavior change, and emotional self-efficacy	• Pretransplant/candidacy: 2 • Early posttransplant: 2 • Long-term survivorship: 1	• Randomized: 4 • Nonrandomized: 1
Clinical and health service utilization outcomes	10 [33,35-38,41,44,46,50,54]	Readmission, length of stay, hospitalization/follow-up costs, outpatient utilization, complications, adverse events, rejection, graft/patient survival, alcohol relapse, and transplant evaluation/triage	• Early posttransplant: 4 • Pretransplant/candidacy: 2 • Cross-stage/mixed: 2 • Long-term survivorship: 1 • Posttransplant timing unspecified: 1	• Randomized: 3 • Nonrandomized: 7
Self-management-related outcomes	17 [34,36,39,40,42,43,45,46,48-50,52-57]	Disease/medication knowledge, organ-quality decision knowledge, medication and follow-up adherence, treatment adherence, self-efficacy, self-care, self-management behavior, and behavior-change confidence	• Early posttransplant: 6 • Pretransplant/candidacy: 4 • Cross-stage/mixed: 4 • Long-term survivorship: 2 • Posttransplant timing unspecified: 1	• Randomized: 9 • Nonrandomized: 8
Health-related quality of life	7 [36,40,42,49,52,53,55]	FACT-G[a], SF-12v2[b], SF-36[c], liver transplant-specific quality-of-life instruments, multidimensional physical, psychological, and social domains	• Early posttransplant: 3 • Cross-stage/mixed: 2; Pretransplant/candidacy: 1 • Long-term survivorship: 1	• Randomized: 5 • Nonrandomized: 2
Implementation-related outcomes and patient experience	15 [31,33,34,36,38,39,42,44-46,48,50,51,56,58]	Participation, attendance, completion, system mastery and use, consultations/messages, satisfaction, acceptability, usability, accessibility, and privacy and digital-access concerns	• Early posttransplant: 4 • Pretransplant/candidacy: 3 • Long-term survivorship: 3 • Cross-stage/mixed: 3 • Posttransplant timing unspecified: 2	• Randomized: 3 • Nonrandomized: 12

[a]FACT-G: Functional Assessment of Cancer Therapy-General.

[b]SF-12v2: 12-Item Short Form Health Survey, Version 2.

[c]SF-36: 36-Item Short Form Health Survey.

The distribution of outcomes varied across the transplant trajectory. In the pretransplant/candidacy stage, the evidence focused primarily on decision-related knowledge, self-management, psychosocial adjustment, alcohol-related behavior, frailty, and physical fitness. Early posttransplant studies covered the broadest range of outcomes, particularly medication adherence, self-care and self-management, readmission and other health service outcomes, physiological monitoring, psychological adjustment, and QoL. Long-term survivorship studies more often assessed lifestyle behavior, cardiometabolic or functional outcomes, self-management, and implementation-related experience. Cross-stage/mixed studies contributed mainly to self-management, clinical/service utilization, QoL, and technology-supported follow-up outcomes. No included study focused exclusively on the

XSL•FO
RenderX

perioperative/inpatient-to-home transition stage; this phase was represented only within broader cross-stage care models.

The strength and type of evidence also differed across outcome domains. Randomized studies accounted for 9 of the 17 studies assessing self-management-related outcomes and 5 of the 7 studies assessing health-related QoL. Psychosocial outcomes were also predominantly represented by randomized designs (4/5). In contrast, clinical and health service utilization outcomes (3/10, randomized), physiological or functional outcomes (2/7, randomized), and implementation-related outcomes or patient experience (3/15, randomized) were more frequently reported in nonrandomized, feasibility, pilot, cohort, or development-oriented studies. Accordingly, Table 4 is intended as an evidence map rather than an effectiveness summary. Detailed quantitative findings from randomized intervention studies are presented separately in the *Findings From Randomized Intervention Studies* section and Table 5, whereas quantitative findings from nonrandomized studies are summarized in the *Findings From Nonrandomized Studies* section and Table 6.

XSL•FO
RenderX

Table . Key findings from randomized intervention studies (n=9).

Study	Study design and transplant trajectory stage	Intervention and comparator	Main outcome(s)	Key quantitative and statistical findings
Lee et al [36], 2019	• Randomized controlled trial • Early posttransplant	Telemedicine-based home monitoring program vs standard care	Hospital readmission and HRQoL[a]	90-day readmission was lower with telemonitoring (28% vs 58%; *P*=.004). Physical functioning at 90 days also favored intervention (75 vs 60; *P*=.02), whereas most other SF-36[b] domains did not differ significantly
Bailey et al [40], 2017	• Randomized controlled trial • Pretransplant/candidacy	Telephone uncertainty self-management intervention vs liver-disease education	Illness uncertainty, depression, anxiety, uncertainty management, self-efficacy, and HRQoL	No significant between-group differences were found in illness uncertainty, depression, anxiety, self-efficacy, or FACT-G[c] QoL[d] at either follow-up (all *P*>.05)
Hickman et al [42], 2021	• Randomized feasibility study • Long-term survivorship	Telehealth-delivered Mediterranean diet + exercise program vs delayed-intervention control	Dietary adherence, anthropometric/metabolic outcomes, functional status, and HRQoL	MEDAS[e] change favored intervention (+2.4 vs −0.1; *P*=.004), as did MCS-12[f] (+2.8 vs −2.0; *P*=.03). Most anthropometric, biochemical, and functional outcomes did not differ significantly
Dolanbay et al [43], 2025	• Randomized controlled trial • Cross-stage/mixed	Web-based education vs control	Immunosuppressive medication-related symptoms, symptom distress, and medication adherence	Symptom occurrence (90.96 vs 112.49; *P*=.009) and symptom distress (19.98 vs 53.51; *P*=.009) were lower, and medication adherence was higher (49.33 vs 46.18; *P*=.001) in the intervention group
DeMartini et al [46], 2018	• Randomized pilot feasibility trial • Pretransplant/candidacy	Tailored text-messaging relapse-prevention intervention + standard care vs standard care	Alcohol use/relapse, stress, abstinence self-efficacy, craving, and feasibility	At 8 weeks, positive urine EtG[g] tests occurred in 0/8 intervention participants vs 2/6 controls; no formal between-group test was reported. Self-reported drinking did not differ (*P*=.19). Stress showed a condition × time interaction (*P*<.05), without significant between-group differences at weeks 4 or 8
Lu et al [49], 2023	• Randomized controlled study • Cross-stage/mixed	Internet-enabled hospital-community-family continuity-of-care model vs conventional continuity care	Self-care, self-management, and HRQoL	At 3 months, self-care, self-management, and multiple LT[h]-specific QoL domains favored intervention (all reported *P*<.05); medication-management scores were 22.87 vs 17.29
Zhai et al [53], 2021	• Randomized controlled trial • Early posttransplant	Mobile health platform–based continuity of care vs routine care	Psychological outcomes; hope, adherence, self-care, self-management, HRQoL, and satisfaction	Medication adherence (12.14 vs 8.69; *P*<.001) and self-management (44.86 vs 34.45; *P*<.001) were higher, while anxiety and depression were lower. Total QoL also favored intervention, although not all QoL domains differed significantly

XSL•FO
RenderX

Study	Study design and transplant trajectory stage	Intervention and comparator	Main outcome(s)	Key quantitative and statistical findings
Pan et al [54], 2023	• Randomized controlled study • Early posttransplant	WeChat educational app + supervised family intervention vs routine care	Glycemic control, medication adherence, self-management, anxiety/depression, and complications	Glycemic control, medication adherence, self-management, and psychological outcomes favored intervention (all *P*<.001). Postoperative complications were lower (2.22% vs 17.14%; *P*=.02)
Wang et al [57], 2023	• Randomized controlled trial • Early posttransplant	Remote rehabilitation management vs routine postoperative follow-up	Treatment adherence and self-care capacity	Treatment adherence (85.03 vs 64.58; *P*<.001) and self-care capacity (95.38 vs 72.23; *P*<.001) were higher with remote rehabilitation management

[a]HRQoL: health-related quality of life.

[b]SF-36: 36-Item Short Form Health Survey.

[c]FACT-G: Functional Assessment of Cancer Therapy-General.

[d]QoL: quality of life.

[e]MEDAS: Mediterranean Diet Adherence Screener.

[f]MCS-12: Mental Component Summary of the 12-Item Short Form Health Survey.

[g]EtG: ethyl glucuronide.

[h]LT: liver transplantation.

XSL•FO
RenderX

Table . Key findings from nonrandomized quantitative studies (n=18).

Study	Study design and transplant trajectory stage	Digital application/intervention	Key quantitative findings
Knihs et al [31], 2024	• Technology product development and validation • Posttransplant timing unspecified	Educational mobile game for home care	Patient SUS[a] mean=80; professional usability ratings were >75%. No clinical-effect comparison was performed
Goswami et al [33], 2024	• Pilot study • Cross-stage/mixed	Tailored virtual alcohol use disorder treatment program	76% (32/42) participants completed the program. Relapse occurred in 22% of completers vs 80% of noncompleters (*P*=.002); this comparison was observational
Andrä et al [34], 2022	• Single-center cohort study • Cross-stage/mixed	Mobile app for outpatient LT[b] aftercare	91.9% (124/135) agreed to participate, 57.3% (71/124) had a compatible device, and 33.8% (42/124) used the app. Privacy concerns were reported by 59.5% of users
Delman et al [35], 2021	• Retrospective cohort study • Cross-stage/mixed	Telehealth-centered transplant service during COVID-19	Telehealth accounted for up to 85% of clinic visits. Hospital stay was shorter in the COVID-era cohort (8.5 vs 13.1 d; *P*=.02); patient and graft survival did not differ significantly
Tian et al [37], 2021	• Cohort study • Early posttransplant	Telemedicine follow-up management	Telemedicine was associated with shorter initial hospitalization (16.31 vs 19.12 days; *P*=.03), lower initial cost (*P*=.049), and lower 30-day readmission (*P*=.02); 12-month survival did not differ significantly
Koc et al [38], 2022	• Prospective cohort study • Long-term survivorship	Telemedicine remote monitoring program	Remote monitoring users had fewer outpatient visits (2 vs 5; *P*<.001) and more TAC[c] determinations (8 vs 6; *P*=.001), with no significant difference in graft rejection (2.6% vs 3.6%; *P*=.34)
Volk et al [39], 2014	• Pilot pre-post study • Pretransplant/candidacy	Web-based organ-quality decision aid	Knowledge, risk perception, and willingness to consider less-than-perfect donor organs improved after the intervention (all *P*<.001); decision confidence did not change significantly
Konjeti et al [41], 2019	• Retrospective cohort study • Pretransplant/candidacy	SCAN-ECHO[d] telehealth triage for LT evaluation	Rejection after full transplant workup was lower with telehealth triage (23.1% vs 55.6%; *P*<.0001), suggesting fewer potentially futile evaluations
Ertel et al [44], 2016	• Nonrandomized observational pilot • Early posttransplant	Video education plus telehealth home monitoring	95% viewed all educational videos and mean satisfaction was 29/30. 30-day and 90-day readmission rates were 20% and 30%, respectively; no comparative statistical test was performed
Duarte-Rojo et al [45], 2023	• Single-arm pilot and feasibility trial • Pretransplant/candidacy	Wearable-supported and app-supported home prehabilitation with remote coaching	21/31 participants entered the intervention and training adherence was 57%. Among completers, LFI[e] improved (*P*=.03) and 6MWT[f] increased from 318 to 358 m (*P*=.005)

XSL•FO
RenderX

Study	Study design and transplant trajectory stage	Digital application/intervention	Key quantitative findings
Wang et al [47], 2023	• Cross-sectional model development/internal validation • Cross-stage/mixed	TeLeFI[g]	TeLeFI models showed cross-validated *C*-statistics of 0.78 and 0.76; the model incorporating virtual chair stands improved frailty classification (NRI[h]=0.70; *P*=.002)
Catalli et al [48], 2026	• Single-arm quasi-experimental pre-post study • Long-term survivorship	Telehealth shared medical appointments	Several psychosocial and self-efficacy measures improved among post-LT participants (*P*=.001-.03), whereas BMI reduction was not significant (32.0 to 30.1 kg/m²; *P*=.06)
Wang et al [50], 2012	• Single-group pre-post implementation study • Posttransplant timing unspecified	Internet information platform for follow-up	Follow-up laboratory-result time decreased from 240 to 5 minutes, scheduled follow-up adherence increased from 82.3% to 95.6%, and adverse events decreased from 11.8% to 4.4%
Li et al [51], 2024	• Technology system development/preliminary application • Cross-stage/mixed	Whole-lifecycle digital health management system (SWOR[i])	The system was implemented in 1730 organ transplant recipients, including 488 LT recipients. No comparative LT-specific patient-effect estimate was reported
Liu et al [52], 2019	• Quasi-experimental controlled study • Cross-stage/mixed	WeChat-based peer education	Medication-adherence measures improved at selected follow-ups, but total adherence was not significantly different at 12 months (*P*=.07). Between-group QoL[j] differences were not significant
Sun et al [55], 2019	• Single-group pre-post study • Early posttransplant	WeChat-based multidisciplinary continuity care	Self-care (*P*=.002), transplant knowledge (*P*<.001), and medication-adherence distribution (*P*<.001) improved at 3 months compared with discharge
Tian et al [56], 2019	• Prospective nonrandomized controlled cohort • Early posttransplant	Telemedicine robot follow-up	Total adherence was higher with robot-based follow-up (84.0 vs 61.4; *P*<.001), with shorter follow-up time and higher satisfaction (both *P*<.001)
Xue et al [58], 2019	• Single-arm feasibility study • Early posttransplant	Remote follow-up management system	81.1% (30/37) eligible recipients participated, system mastery was 100%, and 86.7% (26/30) completed 2 weeks; satisfaction was generally >4/5

[a]SUS: System Usability Scale.

[b]LT: liver transplantation.

[c]TAC: tacrolimus.

[d]SCAN-ECHO: Specialty Care Access Network–Extension for Community Healthcare Outcomes.

[e]LFI: Liver Frailty Index.

[f]6MWT: 6-minute walk test.

[g]TeLeFI: Tele-Liver Frailty Index.

[h]NRI: net reclassification improvement.

[i]SWOR: system for the whole life-cycle management of organ transplant recipient.

[j]QoL: quality of life.

Findings From Randomized Intervention Studies

Among the included studies, 7 were conventional RCTs and 2 additional studies used randomized feasibility or pilot feasibility designs. This review separately examined these 9 randomized intervention studies to provide a more specific account of the reported intervention effects and statistical findings (Table 5). According to the transplant trajectory classification used in this

XSL•FO
RenderX

review, 4 studies focused on early posttransplant recovery [36,53,54,57], 2 on the pretransplant/candidacy stage [40,46], 2 spanned multiple stages of posttransplant care [43,49], and 1 focused predominantly on long-term posttransplant survivorship [42]. No randomized study focused exclusively on the perioperative/inpatient-to-home transition stage.

Across randomized studies, favorable findings were more consistently reported for proximal outcomes, including medication or treatment adherence, self-management or self-care, dietary behavior, psychological outcomes, and selected health service outcomes. For example, telemedicine-based home monitoring reduced 90-day readmission (28% vs 58%; *P*=.004) [36], and a WeChat-based intervention reported fewer postoperative complications (2.22% vs 17.14%; *P*=.02), together with improvements in glycemic control, adherence, self-management, and psychological outcomes [54]. Findings for broader outcomes were less consistent. The long-term diet and exercise feasibility trial improved Mediterranean diet adherence and mental HRQoL but not most anthropometric, biochemical, or functional outcomes [42], whereas the 2 pretransplant randomized studies reported limited or inconsistent between-group effects [40,46].

Several interventions with favorable findings used repeated digital contact together with education, monitoring, professional feedback, rehabilitation guidance, or family support. However, the studies differed in population, transplant stage, intervention intensity, comparator, follow-up duration, and outcome measurement, and 2 were designed primarily to assess feasibility. The findings were therefore synthesized descriptively and were not used to infer the superiority of specific intervention modalities or components.

Findings From Nonrandomized Studies

Of the 19 nonrandomized studies, 18 reported extractable quantitative findings and are summarized in Table 6; the remaining study was a qualitative study of long-term liver transplant recipients [32] and therefore did not contribute quantitative effect estimates. The nonrandomized evidence encompassed cohort comparisons, quasi-experimental and single-group pre-post studies, pilot and feasibility studies, and technology-development or model-development studies. Because these designs differed substantially in their susceptibility to confounding, selection effects, and secular trends, the numerical findings are presented descriptively and are not interpreted as causal estimates of intervention effectiveness.

Pretransplant studies addressed decision support, transplant triage, and prehabilitation. A web-based decision aid improved several knowledge and risk-perception measures [39], telehealth triage reduced the proportion of patients completing a full evaluation before being judged ineligible [41], and home-based prehabilitation improved Liver Frailty Index and 6-minute walk distance among completers [45]. Early posttransplant studies focused mainly on monitoring and follow-up. Telemedicine follow-up was associated with shorter initial hospitalizations, lower costs, and lower 30-day readmission rates without significant survival differences [37], while other studies reported improvements in adherence, self-care, follow-up efficiency, or short-term feasibility [50,55,56,58].

Among long-term and cross-stage populations, remote monitoring reduced outpatient visits while maintaining comparable graft-rejection outcomes [38], whereas lifestyle and peer-support interventions produced mixed behavioral, psychosocial, and QoL findings [33,48,52]. Several studies evaluated technology usability, implementation, or model performance rather than therapeutic effectiveness [31,47,51]. Overall, favorable findings frequently arose from single-arm comparisons, self-selected users, historical controls, or small pilot cohorts and should therefore be interpreted cautiously.

Implementation-Related Outcomes of Digital Health Technologies

Structured implementation-related outcomes or patient-experience measures were reported in 15 of the 28 included studies [31,33,34,36,38,39,42,44-46,48,50,51,56,58]. These outcomes primarily assessed participation and uptake, intervention completion, system mastery and frequency of use, satisfaction, acceptability, usability, accessibility, and user concerns related to digital technology. From the 15 studies, 3 used randomized designs [36,42,46], whereas the remaining 12 used nonrandomized cohort, pilot, single-arm feasibility, pre-post, or technology-development designs. Accordingly, these measures were interpreted as indicators of implementation feasibility and user experience rather than as evidence of clinical effectiveness.

Uptake and engagement varied across studies. In 1 outpatient app study, 91.9% of eligible recipients agreed to participate, but only 57.3% had a compatible device and 33.8% ultimately used the application; 59.5% of users reported privacy concerns [34]. Other studies reported generally favorable short-term completion, usability, or satisfaction [31,33,44,45,58]. The qualitative study [32] provided complementary evidence on recipients' preferences and perceived barriers. Implementation evaluations were distributed across early posttransplant recovery (n=4), pretransplant/candidacy (n=3), long-term survivorship (n=3), cross-stage/mixed care (n=3), and unspecified posttransplant timing (n=2); none focused exclusively on the perioperative/inpatient-to-home transition stage. Overall, willingness to participate, technological access, actual uptake, and sustained engagement emerged as related but distinct implementation dimensions.

Discussion

Principal Findings

This scoping review mapped digital health technologies across the liver transplant trajectory and identified substantial variation in technological modalities, intervention components, study designs, and outcomes. Telehealth platforms or remote management systems were the most frequently reported technological modalities, and remote education was the most common intervention component. Evidence was unevenly distributed across the transplant trajectory. Early posttransplant recovery was the most frequently studied stage, whereas no included study focused exclusively on the perioperative/inpatient-to-home transition stage. Compared with the previous scoping review of telehealth in home-based care

XSL•FO
RenderX

for adult liver transplant recipients [22], the present review included both transplant candidates and recipients and examined intervention components, outcomes, study designs, and implementation evidence across different stages of the transplant trajectory.

Randomized studies more often reported favorable findings for outcomes closely related to the intervention, including medication or treatment adherence, self-management and self-care, dietary behavior, psychological outcomes, and selected health service outcomes [36,42,43,49,53,54,57]. Findings were less consistent for broader outcomes such as overall QoL, physiological or functional measures, and longer-term clinical outcomes [40,42,46]. This pattern may reflect differences in how directly these outcomes can be influenced by short-term digital interventions. Behavioral and self-management outcomes may respond more quickly to repeated education, reminders, monitoring, or feedback, whereas broader clinical and QoL outcomes are influenced by multiple clinical and psychosocial factors and may require longer follow-up.

Several randomized interventions with favorable findings combined repeated digital contact with education, monitoring, professional feedback, rehabilitation guidance, or family support [36,42,49,53,54,57]. These interventions generally extended beyond one-way information delivery and maintained communication between patients and health care professionals. However, this pattern should not be interpreted as evidence that multicomponent interventions are superior. The included randomized studies differed substantially in population, transplant stage, intervention intensity, comparator, follow-up duration, and outcome measurement, and this review did not compare the effects of individual intervention components. The observed pattern therefore identifies a potential direction for intervention design rather than a specific effective combination of components.

Stage-specific mapping showed that digital health was used to address different care needs across the transplant trajectory. Pretransplant interventions addressed decision support, psychological and behavioral support, transplant assessment, and prehabilitation [39-41,45,46], but the evidence remained limited and heterogeneous. Early posttransplant recovery had the broadest range of intervention components and the largest concentration of studies, with particular emphasis on education, monitoring, teleconsultation, medication management, rehabilitation, and follow-up [36,37,44,53-58]. In contrast, long-term survivorship studies focused more on remote monitoring, lifestyle and metabolic management, rehabilitation, and psychosocial support [32,38,42,48]. Cross-stage interventions addressed broader continuity-of-care needs through combinations of education, monitoring, consultation, peer support, and longitudinal management [33-35,43,47,49,51,52].

The absence of studies focused exclusively on the perioperative/inpatient-to-home transition represents an important evidence gap. Some cross-stage interventions included elements relevant to discharge and follow-up, but no intervention was designed and evaluated specifically for this transition period. This gap should not be interpreted as indicating a lower need for digital support; rather, it reflects the current distribution of research across the transplant trajectory.

Implementation findings further showed that initial willingness to participate should not be equated with successful implementation. Across studies, acceptability, technological access, actual uptake, and sustained engagement varied and represented distinct implementation dimensions [32,34]. High satisfaction or short-term completion therefore does not necessarily indicate that an intervention can be sustained or integrated into routine transplant care. Overall, the current evidence suggests that digital health may be most useful when incorporated into structured follow-up pathways that support communication, monitoring, education, self-management, and professional feedback. However, much of the evidence for physiological, health service, and implementation outcomes came from nonrandomized, pilot, feasibility, cohort, or technology-development studies. These findings should therefore be interpreted as evidence of potential clinical utility and implementation feasibility rather than definitive comparative effectiveness.

Strengths and Limitations

This review has 5 key strengths. First, the search covered 9 English-language and Chinese-language databases, capturing evidence from both international and Chinese settings, and included both liver transplant candidates and recipients across a range of clinical contexts including donor organ decision support [39], alcohol-related behavioral risk management [33,46], and long-term recipients' care preferences [32]. Second, qualitative evidence was included to complement quantitative findings on patient needs and implementation context. Third, evidence was organized by a transplant trajectory stage framework, clearly identifying evidence gaps such as the lack of perioperative transition studies. Fourth, randomized and nonrandomized evidence were synthesized separately to avoid conflating heterogeneous study designs, and implementation-related outcomes (uptake, usability, acceptability, and so on) were incorporated to support real-world clinical translation. Finally, the review followed established scoping review guidance (Arksey and O'Malley framework, JBI guidance, and PRISMA-ScR) to ensure transparency and reproducibility.

Several limitations should be considered when interpreting the findings. First, the scoping review protocol was not prospectively registered. Second, no formal critical appraisal or risk-of-bias assessment of included studies was performed, so findings should not be interpreted as establishing the certainty of evidence or comparative effectiveness of specific interventions. Third, 20 potentially relevant reports could not be retrieved in full text, which may have introduced selection bias and limited the comprehensiveness of the evidence base. Fourth, substantial heterogeneity in study designs, intervention characteristics, follow-up periods, and outcome measures precluded quantitative synthesis. Many included studies were small, single-center pilot/feasibility designs with limited longitudinal evidence [33,39,42,44-46,58], and their findings should be considered exploratory. These limitations indicate that the field remains at an early stage of development, requiring

XSL•FO
RenderX

further evidence to establish the effectiveness of digital health interventions in liver transplant care.

Recommendations for Future Research and Development

Future research should strengthen the methodological rigor and comparability of digital health studies in LT. The current evidence includes relatively few randomized trials, while many studies were small, single-center, pilot, feasibility, cohort, or pre-post studies. Studies evaluating effectiveness should use multicenter comparative designs with justified sample sizes and longer follow-up. Greater consistency in outcome selection and measurement is also needed. Researchers should use validated measures where possible, prespecify primary and secondary outcomes, and report intervention content, delivery mode, intensity, duration, adherence, and attrition in sufficient detail. Reporting guidance such as CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) may improve transparency and facilitate future evidence synthesis [59].

Research priorities should also reflect the transplant trajectory. Pretransplant studies should further evaluate digital approaches for decision support, behavioral and psychological support, and prehabilitation. Early posttransplant research should examine whether improvements in adherence and self-management are sustained and translate into clinical or health service outcomes. Long-term studies should place greater emphasis on lifestyle management, functional recovery, cardiometabolic risk, psychosocial adjustment, and sustained engagement with follow-up care. The perioperative/inpatient-to-home transition requires particular attention because no included study focused exclusively on this stage. Future interventions could address discharge preparation, early symptom monitoring, medication management, and communication between patients and the transplant team.

Intervention development should also clarify which components contribute to meaningful and sustained outcomes. Several randomized interventions with favorable findings combined repeated digital contact with education, monitoring, professional feedback, rehabilitation guidance, or family support. However, the current evidence does not establish which individual components or combinations are responsible for these findings. Future studies should describe intervention components, delivery processes, professional contact, personalization, caregiver involvement, and intervention fidelity more consistently. Comparative or optimization designs may help examine the contribution of individual components and distinguish the effects of technology from the clinical support delivered through it.

Implementation should be treated as a core research outcome. Future studies should examine not only satisfaction or usability but also access, uptake, engagement, workflow integration, maintenance, and scalability. The effects of digital care on the workload of transplant nurses and other health care professionals should also be assessed, including how patient-generated data are reviewed and how responsibilities are distributed within multidisciplinary teams. User-centered design should remain central [60], and privacy and data security should be considered throughout intervention development and implementation [61].

Equity and economic sustainability require parallel evaluation. Digital interventions may improve access to specialist care but may also disadvantage patients with limited device access, internet connectivity, digital skills, health literacy, or confidence in using technology. Older adults may face additional usability and accessibility barriers, although age alone should not be treated as a proxy for inability to use digital care [62,63]. Future studies should assess digital access and literacy, include patients and caregivers from diverse socioeconomic and geographic backgrounds, and consider training, caregiver support, alternative access methods, and hybrid care models. Economic evaluations should include development, equipment, maintenance, staffing, training, data management, and patient-side costs and assess whether reductions in travel, outpatient visits, or unplanned service use offset implementation costs. Effectiveness, implementation, equity, and economic outcomes should be considered together when assessing the long-term role of digital health in routine liver transplant care.

Conclusions

This scoping review mapped digital health technologies across the liver transplant trajectory. Most studies focused on early posttransplant recovery. Randomized studies more often reported positive findings for adherence, self-management, health behaviors, and selected psychosocial or health service outcomes, while findings for QoL, physiological outcomes, functional outcomes, and long-term clinical outcomes were less consistent. No study focused exclusively on the perioperative/inpatient-to-home transition stage. Much of the current evidence came from nonrandomized, pilot, feasibility, or small single-center studies. Future studies should use stage-specific and multicenter designs, improve the consistency of intervention and outcome reporting, and assess long-term implementation, access, and costs.

Generative AI Use

No generative AI tools were used for manuscript preparation, writing, editing, or data-related analyses in this study.

Funding

This work was supported by the Scientific Research Project of the Education Department of Jilin Province (JJKH20250227KJ) and the Jilin Provincial Department of Finance (J1SRCZX2025-074). The funders had no role in the design or conduct of the review; literature searching; study selection; data extraction, analysis, or interpretation; preparation of the manuscript; or the decision to submit the manuscript for publication.

XSL•FO
RenderX

Authors' Contributions

Conceptualization: WS, JS
Data curation: WS, JL, YL
Formal analysis: WS, RL, FY
Investigation: WS, JL, YL
Supervision: JS
Validation: JL, ZC
Visualization: WS, JL, YL
Writing – original draft: WS, RL
Writing – review & editing: JS, FY

Conflicts of Interest

None declared.

Multimedia Appendix 1
Complete database-specific search strategies for 5 international and 4 Chinese bibliographic databases, covering database inception through July 10, 2026.
[PDF File, 163 KB - mhealth_v14i1e106470_app1.pdf]

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 69 KB - mhealth_v14i1e106470_app2.docx]

References

1. Sugawara Y, Hibi T. Recent trends and new developments in liver transplantation. Biosci Trends 2024 Jul 9;18(3):206-211. [doi: 10.5582/bst.2024.01176] [Medline: 38945855]
2. Department of Medical Administration, National Health Commission of the People's Republic of China. Guidelines for the diagnosis and treatment of primary liver cancer (2026 Edition) [Article in Chinese]. Med J Peking Union Med Coll Hosp 2026;17(3):735-770. [doi: 10.12290/xhyxzz.2026-0487]
3. Bernal W, Bittermann T, Sheth R, et al. Survival with and without liver transplantation in critically ill patients with cirrhosis: a 20-year experience. Clin Gastroenterol Hepatol 2026 Feb;24(2):432-442. [doi: 10.1016/j.cgh.2025.04.029] [Medline: 40473089]
4. Adam R, Karam V, Cailliez V, et al. 2018 annual report of the European Liver Transplant Registry (ELTR) - 50-year evolution of liver transplantation. Transpl Int 2018 Dec;31(12):1293-1317. [doi: 10.1111/tri.13358] [Medline: 30259574]
5. Kaplan A, Korenjak M, Brown RS Jr. Post-liver transplantation patient experience. J Hepatol 2023 Jun;78(6):1234-1244. [doi: 10.1016/j.jhep.2023.01.008] [Medline: 37208108]
6. Shen S, Su G, Liang H. Long-term follow-up and complication management after curative treatment for liver cancer [Article in Chinese]. J Clin Intern Med 2026;43(2):105-108. [doi: 10.3969/j.issn.1001-9057.2026.02.004]
7. Wang J, Zhou D, Zhou Y, et al. Patients waiting for a liver transplantation from the perspective of psychological resilience: a qualitative study [Article in Chinese]. Chin J Nurs 2025;60(15):1864-1871. [doi: 10.3761/j.issn.0254-1769.2025.15.011]
8. Hoppmann N, Massoud O. Medical care of liver transplant patients. Expert Rev Gastroenterol Hepatol 2020 Oct;14(10):901-918. [doi: 10.1080/17474124.2020.1797483] [Medline: 32684007]
9. Xu J. Advances in research on continuity-of-care models in China [Article in Chinese]. Chin Nurs Manage 2012;12(9):18-19. [doi: 10.3969/j.issn.1672-1756.2012.09.005]
10. Guo H, Sha L, Pu C, Cheng Y. Application of "internet+" to continuous nursing of patients after surgery: a literature review [Article in Chinese]. Chin Nurs Manage 2019;19(7):1045-1049. [doi: 10.3969/j.issn.1672-1756.2019.07.018]
11. Davis KM, Eckert MC, Hutchinson A, et al. Effectiveness of nurse-led services for people with chronic disease in achieving an outcome of continuity of care at the primary-secondary healthcare interface: a quantitative systematic review. Int J Nurs Stud 2021 Sep;121:103986. [doi: 10.1016/j.ijnurstu.2021.103986] [Medline: 34242979]
12. Lu M, Yu X, Zhang S, et al. A longitudinal qualitative study on the rehabilitation needs of liver transplant patients at different stages [Article in Chinese]. Chin J Nurs 2026;61(5):660-666. [doi: 10.3761/j.issn.0254-1769.2026.05.012]
13. Artru F, Trovato F, Morrison M, Bernal W, McPhail M. Liver transplantation for acute-on-chronic liver failure. Lancet Gastroenterol Hepatol 2024 Jun;9(6):564-576. [doi: 10.1016/S2468-1253(23)00363-1] [Medline: 38309288]
14. Saracino RM, Jutagir DR, Cunningham A, et al. Psychiatric comorbidity, health-related quality of life, and mental health service utilization among patients awaiting liver transplant. J Pain Symptom Manage 2018 Jul;56(1):44-52. [doi: 10.1016/j.jpainsymman.2018.03.001] [Medline: 29526612]

XSL•FO
RenderX

15. Zhou H, Su M, Yang Z, Huang Y, He H, Xu L. Current status and attribute analysis of continuous care service needs for liver transplantation recipients [Article in Chinese]. Chin Nurs Manage 2024;24(11):1738-1742. [doi: 10.3969/j.issn.1672-1756.2024.11.027]
16. Lara AN, Sartori AMC, de Soárez PC, et al. Healthcare resource utilization and costs of outpatient follow-up after liver transplantation in a university hospital in São Paulo, Brazil: cost description study. Sao Paulo Med J 2015;133(3):171-178. [doi: 10.1590/1516-3180.2013.7000011] [Medline: 26039536]
17. Hong W. Advances and opportunities of mobile health in the postpandemic era: smartphonization of wearable devices and wearable deviceization of smartphones. JMIR mHealth uHealth 2024 Jan 22;12:e48803. [doi: 10.2196/48803] [Medline: 38252596]
18. Portz J, Moore S, Bull S. Evolutionary trends in the adoption, adaptation, and abandonment of mobile health technologies: viewpoint based on 25 years of research. J Med Internet Res 2024 Sep 27;26:e62790. [doi: 10.2196/62790] [Medline: 39331463]
19. Duettmann W, Naik MG, Zukunft B, et al. eHealth in transplantation. Transpl Int 2021 Jan;34(1):16-26. [doi: 10.1111/tri.13778] [Medline: 33111393]
20. Ji R, Yan W, Li Z, et al. Application and prospects of mobile health applications in the health management of organ transplant recipients [Article in Chinese]. Organ Transplant 2025;16(3):474-481. [doi: 10.12464/j.issn.1674-7445.2025071]
21. Tang J, James L, Howell M, Tong A, Wong G. eHealth interventions for solid organ transplant recipients: a systematic review and meta-analysis of randomized controlled trials. Transplantation 2020 Aug;104(8):e224-e235. [doi: 10.1097/TP.0000000000003294] [Medline: 32732828]
22. Gao R, Feng SY, Zheng J, Zhai LJ, Liu R. Telehealth and telemedicine in the management of adult patients after liver transplantation: a scoping review. Telemed J E Health 2025 Apr;31(4):375-385. [doi: 10.1089/tmj.2024.0359] [Medline: 39630539]
23. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
24. Peters MDJ, Marnie C, Tricco AC, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020 Oct;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
25. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
26. Pollock D, Peters MDJ, Khalil H, et al. Recommendations for the extraction, analysis, and presentation of results in scoping reviews. JBI Evid Synth 2023 Mar 1;21(3):520-532. [doi: 10.11124/JBIES-22-00123] [Medline: 36081365]
27. Issa DH, Alkhouri N. Long-term management of liver transplant recipients: a review for the internist. Cleve Clin J Med 2015 Jun;82(6):361-372. [doi: 10.3949/ccjm.82a.14072] [Medline: 26086495]
28. Lieber SR, Jones AR, Gowda P, et al. Survivorship interventions after liver transplantation (LT): a systematic review of interventions targeting physical and psychosocial well-being after LT. Liver Transpl 2026 Feb 1;32(2):195-221. [doi: 10.1097/LVT.0000000000000682] [Medline: 40674286]
29. Wachholz LF, Knihs NDS, Sens S, Paim SMS, Magalhães ALP, Roza BDA. Good practices in transitional care: continuity of care for patients undergoing liver transplantation. Rev Bras Enferm 2021;74(2):e20200746. [doi: 10.1590/0034-7167-2020-0746] [Medline: 34037143]
30. Brustia R, Monsel A, Skurzak S, et al. Guidelines for perioperative care for liver transplantation: Enhanced Recovery After Surgery (ERAS) recommendations. Transplantation 2022 Mar 1;106(3):552-561. [doi: 10.1097/TP.0000000000003808] [Medline: 33966024]
31. Knihs NDS, Silva AMD, Grespi LDO, et al. Mobile game: educational technology for home care of patients undergoing liver transplantation. Texto Contexto - Enferm 2024;33. [doi: 10.1590/1980-265x-tce-2023-0162pt]
32. Hickman IJ, Coran D, Wallen MP, et al. 'Back to Life'-using knowledge exchange processes to enhance lifestyle interventions for liver transplant recipients: a qualitative study. Nutr Diet 2019 Sep;76(4):399-406. [doi: 10.1111/1747-0080.12548] [Medline: 31083821]
33. Goswami A, Weinberg E, Coraluzzi L, et al. A tailored virtual program for alcohol use disorder treatment among liver transplant candidates and recipients is feasible and associated with lower post-transplant relapse. Clin Transplant 2024 Jul;38(7):e15381. [doi: 10.1111/ctr.15381] [Medline: 38940230]
34. Andrä M, Sibinovic M, Pfeiffer K, Kniepeiss D. Implementation of a mobile application for outpatient care after liver transplantation. Digit Health 2022;8:20552076221145855. [doi: 10.1177/20552076221145855] [Medline: 36601283]
35. Delman AM, Turner KM, Jones CR, et al. Keeping the lights on: telehealth, testing, and 6-month outcomes for orthotopic liver transplantation during the COVID-19 pandemic. Surgery 2021 Jun;169(6):1519-1524. [doi: 10.1016/j.surg.2020.12.044] [Medline: 33589248]
36. Lee TC, Kaiser TE, Alloway R, Woodle ES, Edwards MJ, Shah SA. Telemedicine based remote home monitoring after liver transplantation: results of a randomized prospective trial. Ann Surg 2019 Sep;270(3):564-572. [doi: 10.1097/SLA.0000000000003425] [Medline: 31356267]
37. Tian M, Wang B, Xue Z, et al. Telemedicine for follow-up management of patients after liver transplantation: cohort study. JMIR Med Inform 2021 May 17;9(5):e27175. [doi: 10.2196/27175] [Medline: 33999008]

XSL•FO
RenderX

38. Koc Ö, Pierco M, Remans K, et al. Telemedicine based remote monitoring after liver transplantation: feasible in a select group and a more stringent control of immunosuppression. Clin Transplant 2022 Jan;36(1):e14494. [doi: 10.1111/ctr.14494] [Medline: 34596918]
39. Volk ML, Roney M, Fagerlin A. Pilot test of a patient decision aid about liver transplant organ quality. Liver Transpl 2014 Jul;20(7):850-855. [doi: 10.1002/lt.23882] [Medline: 24706409]
40. Bailey DE, Hendrix CC, Steinhauser KE, et al. Randomized trial of an uncertainty self-management telephone intervention for patients awaiting liver transplant. Patient Educ Couns 2017 Mar;100(3):509-517. [doi: 10.1016/j.pec.2016.10.017] [Medline: 28277289]
41. Konjeti VR, Heuman D, Bajaj JS, et al. Telehealth-based evaluation identifies patients who are not candidates for liver transplantation. Clin Gastroenterol Hepatol 2019 Jan;17(1):207-209. [doi: 10.1016/j.cgh.2018.04.048] [Medline: 29723691]
42. Hickman IJ, Hannigan AK, Johnston HE, et al. Telehealth-delivered, cardioprotective diet and exercise program for liver transplant recipients: a randomized feasibility study. Transplant Direct 2021;7(3):e667. [doi: 10.1097/TXD.0000000000001118] [Medline: 33564717]
43. Dolanbay N, Ozkan M. The effect of web-based education on symptoms and adherence to immunosuppressive medications in liver transplant patients. Transplant Proc 2025 Jun;57(5):803-809. [doi: 10.1016/j.transproceed.2025.03.005] [Medline: 40300900]
44. Ertel AE, Kaiser TE, Abbott DE, Shah SA. Use of video-based education and tele-health home monitoring after liver transplantation: results of a novel pilot study. Surgery 2016 Oct;160(4):869-876. [doi: 10.1016/j.surg.2016.06.016] [Medline: 27499142]
45. Duarte-Rojo A, Bloomer PM, Grubbs RK, et al. Use of a mobile-assisted telehealth regimen to increase exercise in transplant candidates: a home-based prehabilitation pilot and feasibility trial. Clin Transl Gastroenterol 2023 Nov 1;14(11):e00601. [doi: 10.14309/ctg.0000000000000601] [Medline: 37477616]
46. DeMartini KS, Schilsky ML, Palmer A, et al. Text messaging to reduce alcohol relapse in prelisting liver transplant candidates: a pilot feasibility study. Alcohol Clin Exp Res 2018 Apr;42(4):761-769. [doi: 10.1111/acer.13603] [Medline: 29498753]
47. Wang M, Shui AM, Barry F, et al. The tele-liver frailty index (TeLeFI): development of a novel frailty tool in patients with cirrhosis via telemedicine. Am J Transplant 2023 Jul;23(7):966-975. [doi: 10.1016/j.ajt.2023.04.012] [Medline: 37061188]
48. Catalli LL, Ha NB, Miller SA, et al. A shared medical appointment program to improve self-management of metabolic dysfunction-associated steatotic liver disease. Gastro Hep Adv 2026;5(4):100903. [doi: 10.1016/j.gastha.2026.100903] [Medline: 41868859]
49. Lu XP, Jing XL, Gu SS. Application value of internet plus hospital-community-family integrated continuous nursing model in patients after liver transplantation [Article in Chinese]. Mod J Integr Tradit Chin West Med 2023;32(13):1868-1872. [doi: 10.3969/j.issn.1008-8849.2023.13.024]
50. Wang Y, Wang L, Tian Y, Bai L. Application of online information platform in the follow-up management for liver transplantation patients. Chin J Mod Nurs 2012;18(34):4176-4178. [doi: 10.3760/cma.j.issn.1674-2907.2012.34.027]
51. Li ZX, Zhang Q, Zhang DY, et al. Preliminary experience in the construction and application of a digital health management system for the whole lifecycle of organ transplant recipients [Article in Chinese]. Pract J Organ Transplant 2024;12(6):539-542. [doi: 10.3969/j.issn.2095-5332.2024.06.011]
52. Liu J, Xie J, Liu L, et al. Effects of WeChat-based peer supports upon medication adherence in liver transplant recipients [Article in Chinese]. Chin J Organ Transplant 2019;40(8):484-488. [doi: 10.3760/cma.j.issn.0254-1785.2019.08.008]
53. Zhai MM, Li L, Zhou YX, Li SY, Liu SS, Wang DH. Study on the effect of continuous nursing based on mobile medical platform in patients after liver transplantation [Article in Chinese]. Lab Med Clin 2021;18(5):644-648. [doi: 10.3969/j.issn.1672-9455.2021.05.019]
54. Pan T, Wei L, Zhou Q, Yu J. The effect of WeChat education app combined with supervised family intervention on patients after liver transplantation [Article in Chinese]. J Qilu Nurs 2023;29(4). [doi: 10.3969/j.issn.1006-7256.2023.04.010]
55. Sun M, Li N, Liu G, Huang WH, Liu JL, Ren YW. Evaluation of the effectiveness of multidisciplinary participation in continuity of care after liver transplantation through a WeChat platform [Article in Chinese]. Mod Dig Interv 2019;24(S2):2616-2617 [FREE Full text]
56. Tian BY, Lu HP, Zhang JQ, et al. Application of telemedicine robot in follow-up after liver transplantation from donation after cardiac death [Article in Chinese]. Organ Transplant 2019;10(1):79-83. [doi: 10.3969/j.issn.1674-7445.2019.01.012]
57. Wang MM, Wang D, Huang XK. Application of remote rehabilitation management mode in postoperative follow-up of liver transplantation patients [Article in Chinese]. Chin J Mod Nurs 2023;29(3):375-379. [doi: 10.3760/cma.j.cn115682-20220303-00992]
58. Xue Z, Tian M, Du YY, et al. Application of remote follow-up management system in transitional care of liver transplant recipients [Article in Chinese]. J Nurs Sci 2019;34(21):73-76. [doi: 10.3870/j.issn.1001-4152.2019.21.073]
59. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]

XSL•FO
RenderX

60. Schnall R, Rojas M, Bakken S, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251. [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
61. Rezaee R, Khashayar M, Saeedinezhad S, Nasiri M, Zare S. Critical criteria and countermeasures for mobile health developers to ensure mobile health privacy and security: mixed methods study. JMIR Mhealth Uhealth 2023 Mar 2;11:e39055. [doi: 10.2196/39055] [Medline: 36862494]
62. Nieman CL, Oh ES. Connecting with older adults via telemedicine. Ann Intern Med 2020 Nov 17;173(10):831-832. [doi: 10.7326/M20-1322] [Medline: 32777187]
63. Nimmanterdwong Z, Boonviriya S, Tangkijvanich P. Human-centered design of mobile health apps for older adults: systematic review and narrative synthesis. JMIR Mhealth Uhealth 2022 Jan 14;10(1):e29512. [doi: 10.2196/29512] [Medline: 35029535]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
JBI: Joanna Briggs Institute
LT: liver transplantation
mHealth: mobile health
PCC: population, concept, and context
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
QoL: quality of life
RCT: randomized controlled trial

Edited by A Stone; submitted 07.Jul.2026; peer-reviewed by B Breznen, T Matsusaki; revised version received 08.Sep.2026; accepted 22.Sep.2026; published 16.Nov.2026.

Please cite as:
Shi W, Sun J, Li J, Li Y, Yuan F, Li R, Chu Z
Applications and Current Landscape of Digital Health Technologies in the Continuity of Care for Liver Transplant Candidates and Recipients: Scoping Review
JMIR Mhealth Uhealth 2026;14:e106470
URL: https://mhealth.jmir.org/2026/1/e106470
doi:10.2196/106470

© Wanqi Shi, Jiao Sun, Jiahao Li, Yuanyuan Li, Feiyan Yuan, Ruiya Li, Zijing Chu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 16.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Reporting Gaps in mHealth Intervention Studies for Adults With Diabetes: Systematic Review

Xin Yue[1], MSc; Jiacheng Shen[1], MSc; Junke Peng[1], BSc; Xiaoting Huang[2], PhD; Mingyue Hu[2], PhD; Xing Hu[3], BSc

[1]Xiangya School of Nursing, Central South University, Changsha, Hunan, China

[2]Central South University, Changsha, Hunan, China

[3]Sanzhen Rehabilitation Hospital, 448 Fenglin 3rd Road, Yuelu District, Changsha, Hunan, China

Corresponding Author:
Xing Hu, BSc
Sanzhen Rehabilitation Hospital, 448 Fenglin 3rd Road, Yuelu District, Changsha, Hunan, China

Abstract

Background: Diabetes self-management education and support requires scalable digital solutions, making mobile health (mHealth) interventions increasingly vital. Despite their proven clinical efficacy, the real-world translation and scalability of these interventions are severely hindered by fragmented and opaque reporting. To address this methodological gap, the World Health Organization developed the mHealth evidence reporting and assessment (mERA) checklist to standardize reporting transparency.

Objective: This study aimed to systematically evaluate the reporting completeness of published randomized controlled trials (RCTs) on mHealth interventions for adult diabetes using the mERA checklist; identify critical reporting gaps; analyze how these deficiencies undermine feasibility, scalability, and sustainability; and propose targeted strategies for future optimization.

Methods: A PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)–compliant systematic review was conducted. Five major databases (PubMed, Embase, Web of Science, MEDLINE, and the Cochrane Library) were systematically searched for relevant RCTs published from inception to July 2026. Data extraction and reporting completeness assessments were conducted using the 16-item mERA core checklist. Nonparametric and categorical analyses were used to evaluate factors associated with reporting completeness.

Results: A total of 144 RCTs conducted across 49 countries were included in the final analysis. The mean overall mERA reporting score was 6.74 out of 16 (SD 2.71). While basic characteristics such as intervention delivery (n=128, 88.9%) and technology platform (n=117, 81.3%) were well reported, critical system-level domains were largely omitted. Specifically, interoperability (n=12, 8.3%), data security (n=28, 19.4%), compliance with guidelines (n=41, 28.5%), and intervention fidelity (n=53, 36.8%) were severely underreported. Additionally, reporting completeness was significantly higher in studies published between 2016 and 2026 (*P*=.01) and in multicomponent interventions compared to single-component ones (*P*=.04).

Conclusions: Empirical research on mHealth interventions for adult diabetes exhibits systemic structural flaws in reporting, particularly concerning system integration, data security, and regulatory compliance. These omissions significantly undermine the transparency, reproducibility, and real-world scalability of digital diabetes care solutions. Future trials must rigorously adopt standardized reporting frameworks such as the mERA checklist to bridge the gap between clinical efficacy and large-scale public health policy translation.

(JMIR Mhealth Uhealth 2026;14:e95863) doi:10.2196/95863

KEYWORDS

mobile health; mHealth; systematic review; diabetes; mobile app; reporting guidelines

Introduction

Diabetes is a chronic metabolic disease characterized by elevated blood glucose levels resulting from insufficient insulin production or ineffective insulin use [1]. Its global burden continues to rise rapidly. According to the Noncommunicable Disease Risk Factor Collaboration, 828 million adults were living with diabetes in 2022, an increase of 630 million since 1990 [2]. Nearly 59% of these patients—approximately 450 million adults—remain untreated, with the largest treatment gaps concentrated in low- and middle-income countries (LMICs) [2]. Concurrently, disparities in treatment coverage are widening between high-income Western countries and regions with low treatment coverage, including Pacific island nations, South Asia, and sub-Saharan Africa, reinforcing the need for delivery models that can reach populations with limited access to conventional services [2]. The economic burden is similarly substantial, with global costs projected to increase from US $1.3 trillion in 2015

XSL•FO
RenderX

to US $2.2 trillion by 2030, representing up to 2.2% of the global gross domestic product [3].

Traditional health care facilities are increasingly inadequate to accommodate the growing demand for chronic disease care. Consequently, traditional diabetes self-management education and support services require integration with highly scalable digital health systems [4,5]. Mobile health (mHealth) interventions—including smartphone apps, videoconferencing, and SMS text messaging—can provide remote care, health education, and self-management support [4]. By 2023, global smartphone penetration reached 6.92 billion users, representing approximately 86% of the world population [6]. This widespread adoption positions mobile technology as a viable infrastructure to overcome geographical and resource barriers, facilitating a continuous, technologically supported self-management feedback loop for patients [6].

Evidence from randomized controlled trials (RCTs) and meta-analyses supports the clinical potential of mHealth for adults with type 2 diabetes mellitus [7-9]. A meta-analysis of 41 studies found that app-based interventions significantly improved glycemic control compared with standard care, with a mean hemoglobin A_{1c} reduction of 0.49% (95% CI –0.65% to –0.32%) [7]. Real-world evidence has also shown clinically relevant benefits. In 2023, Heald et al [9] evaluated the Healum collaborative care planning app in the UK National Health Service primary care and reported, over 6 months, a 7.4% mean percentage reduction in hemoglobin A_{1c}, a 0.7% reduction in BMI, and a mean EQ-5D-5L score increase of 0.0464.

Clinical effectiveness, however, does not by itself ensure that an intervention can be reproduced, implemented, or scaled. Translation into routine and reimbursable care is hindered when intervention reporting is fragmented or insufficiently transparent [10,11]. Vague descriptions of delivery processes can limit evidence synthesis and practical application [10-12], whereas failure to report the frequency and duration of app use makes dose-response relationships difficult to assess [6]. Furthermore, critical parameters such as implementation feasibility and technological adaptability across diverse health care settings remain underreported [7]. As a result, comparisons across studies become less reliable, and decision-makers lack key evidence for judging real-world feasibility, scalability, and long-term sustainability [6,13].

To improve reporting transparency, the World Health Organization mHealth Technical Evidence Review Group developed the 16-item mHealth evidence reporting and assessment (mERA) core checklist, which covers intervention content, implementation context, and technical features [14]. Rather than measuring clinical efficacy, the mERA checklist assesses whether reports contain the information needed to understand and replicate an intervention. Previous applications of the mERA checklist reveal substantial gaps: data security and privacy protection have been reported in 0% of studies, cost assessments have been reported in only 13%, and compliance with national clinical guidelines such as HIPAA (Health Insurance Portability and Accountability Act) or diabetes self-management education and support standards have been reported in only 20% [15]. A 2022 systematic review further found that none of the assessed mobile diabetes interventions in LMICs had applied mERA guidelines [16], reflecting a widespread tendency to prioritize clinical outcomes over process evaluation. These limitations ground the rationale for this study. Given that adult diabetes management demands rigorous data interaction, privacy protection, and cost control, systematically auditing current RCT literature via the mERA checklist is methodologically imperative to address the bottlenecks hindering large-scale digital health policy translation.

Therefore, this study used the mERA checklist as a systematic evaluation framework to analyze the reporting quality of published RCTs evaluating adult diabetes mHealth interventions. By analyzing compliance rates across mERA items, this study identified systemic reporting blind spots in current digital health literature. Furthermore, it explored how these deficiencies undermine the feasibility, scalability, and sustainability of digital diabetes care solutions. Finally, this study proposes targeted strategies to optimize future digital health interventions.

Methods

Study Design

This systematic review was conducted and reported in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist [17]. The completed PRISMA 2020 checklist is provided in Checklist 1. This review was not registered.

Inclusion and Exclusion Criteria

The eligibility criteria for this review were structured as follows.

Study Design

Only peer-reviewed RCTs evaluating intervention efficacy were included. Conversely, feasibility studies, pilot trials, conference abstracts, and unpublished studies were excluded.

Participants

Eligible participants were adults (aged ≥18 years) with a confirmed diagnosis of type 1 or type 2 diabetes mellitus. Studies evaluating patients with gestational diabetes mellitus were excluded.

Interventions

All mHealth interventions designed for health promotion or disease management were eligible for inclusion. Acceptable delivery modalities encompassed SMS text messaging, telephone calls, emails, wearable devices, mobile apps, web portals, and hybrid interventions.

Search Strategy

This study systematically searched PubMed, Embase, Web of Science, MEDLINE, and the Cochrane Library to identify relevant studies available from the inception of each database to July 2026. Only articles published in English were included in the analysis. Keywords related to mHealth, diabetes, effectiveness, and RCTs were used (detailed in Multimedia Appendix 1). Within each database, terms were combined using the “OR” operator, followed by the integration of results across databases using the “AND” operator. The comprehensive search

XSL•FO
RenderX

strategies and specific keywords are thoroughly described in Multimedia Appendix 1, ensuring transparency and reproducibility of the retrieval process.

Study Selection

Duplicate entries were removed using the EndNote software (version 21; Clarivate Analytics), followed by a manual cross-check to ensure data integrity. The corresponding authors of 17 research reports were contacted to request the complete manuscripts. Of these 17 authors, 15 (88.2%) responded and provided the required data, and the remaining 2 (11.8%) studies were excluded because the corresponding authors did not reply, rendering the necessary data unavailable. Subsequently, 2 independent reviewers (XY and JS) screened all titles, abstracts, and full texts according to the predefined inclusion criteria. Discrepancies were resolved through consensus-seeking discussions with a third reviewer (MH). Reasons for excluding full-text articles were meticulously documented. The detailed search strategies can be found in Multimedia Appendix 1.

Data Extraction

Three researchers (XY, JP, and JS) completed data extraction. Prior to the formal process, the research team underwent standardized training, during which XY introduced the mERA core checklist and clarified its 16 items. A pilot assessment was then conducted to evaluate interrater reliability, after which the Fleiss κ coefficient was calculated, and all discrepancies were discussed iteratively to reach a consensus. During the formal extraction phase, 2 researchers (JS and JP) independently extracted data from each study using a standardized form, with any unresolved disagreements adjudicated by a third researcher (XY). The extracted variables included the first author, publication year, country, mHealth intervention type, participant digital literacy, compliance with mERA core items, and funding sources.

Data Synthesis and Analysis

Descriptive statistics were used to summarize study characteristics and mERA reporting completeness. Continuous variables (eg, total mERA scores) are presented as medians and IQRs due to nonnormal distribution, whereas categorical variables are presented as frequencies and percentages. The normality of continuous data was evaluated using the Shapiro-Wilk test.

To explore factors associated with reporting completeness, nonparametric tests were used. Specifically, the Mann-Whitney *U* test was used to compare total mERA scores between 2 independent groups. The year 2016 was selected a priori as the temporal cutoff because the mERA reporting guideline was published in that year [14]; studies were therefore grouped as 2009 to 2015 and 2016 to 2026. The Kruskal-Wallis *H* test was used for comparisons across more than 2 groups. For categorical variables (ie, item-by-item adherence rates), differences between groups were analyzed using the Fisher exact test or Pearson chi-square test as appropriate based on expected cell counts. All statistical analyses were performed using R (version 4.5.0; R Foundation for Statistical Computing). A 2-tailed *P* value below .05 was considered statistically significant.

Classification of mHealth Intervention Types

To categorize mHealth intervention types, we applied a component-based classification approach. Each intervention was classified according to the primary technological modality used to deliver the intervention. When an intervention incorporated additional digital components beyond the primary modality, it was classified as a multicomponent intervention rather than being assigned to a single technology category.

The intervention categories were defined as follows: (1) mobile app–based interventions, referring to interventions primarily delivered through smartphone or tablet apps; (2) SMS text messaging–based interventions, referring to interventions mainly delivered through SMS text messages; (3) website-based interventions, referring to interventions delivered through web-based platforms; (4) wearable device–based interventions, referring to interventions involving wearable sensors or devices for monitoring or feedback; and (5) combined interventions, referring to interventions that integrated 2 or more technological modalities (eg, mobile app plus SMS text messaging or mobile app plus wearable devices).

Therefore, categories such as “mobile app,” “mobile app and wearable devices,” and “mobile app and SMS” represent mutually distinguishable intervention strategies rather than overlapping categories.

Evaluation of Evidence Reporting

To assess the quality of evidence reporting in each study, we adopted the mERA core checklist in accordance with the guidelines proposed by Iribarren et al [18]. The mERA core checklist consists of 16 items, which represent the minimum set of information that should be reported to ensure the replicability of an intervention. These items cover the intervention content, technical characteristics, and implementation context, as well as participants’ engagement and experience with the intervention. An item was classified as “reported” if it was adequately described in the included papers and their supplementary files. Conversely, if an item was absent from the included papers, it was classified as “unreported.” In this process, the evidence related to mHealth interventions in each study was evaluated based on the criteria specified in the mERA core checklist. During the evaluation, reviewers assigned a score of either 1 or 0 for each criterion to each study, where a score of 1 indicated that a specific mERA assessment criterion was met and a score of 0 indicated that it was not met. The assignment table for each evaluation indicator of each study is provided in the Multimedia Appendix 2.

Interrater reliability measures the degree of agreement among evaluators. Given that our assessment process involved multiple evaluators, testing their interrater reliability was not only crucial but also essential. To this end, we conducted a preassessment in which 3 studies were randomly selected following the recommendations put forward by Agarwal et al [14]. During the scoring process, evaluators assigned a value of either 1 or 0 to each study, where 1 indicated that a specific assessment criterion of the mERA checklist was met and 0 indicated that it was not met. To evaluate the interrater reliability, we used the Fleiss κ coefficient [19]. It should be noted that in the

XSL•FO
RenderX

context of grading health care research, a κ value as low as 0.41 is considered acceptable [20]. The results of the preassessment showed that for the 3 randomly selected studies, the interrater reliability values were 0.65, 0.83, and 0.82, which are far higher than the recommended acceptable threshold.

Results

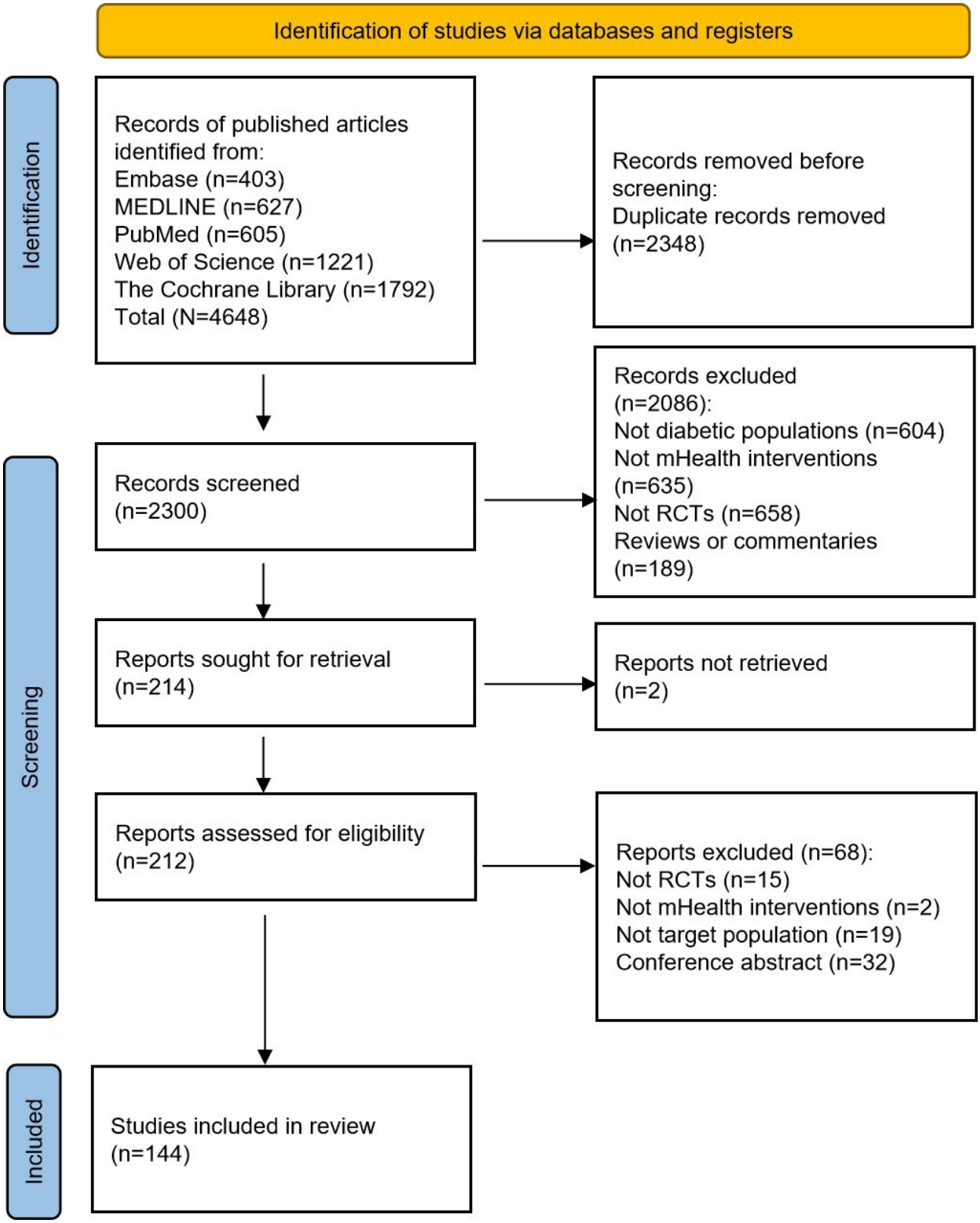
Overview

Figure 1 shows the rigorously updated PRISMA 2020–compliant flow diagram outlining the study selection process. The initial systematic literature search across 5 databases yielded a total of 4648 records (Embase: n=403; MEDLINE: n=627; PubMed: n=605; Web of Science: n=1221; Cochrane Library: n=1792). Following the removal of 2348 duplicate records before screening, 2300 records remained for title and abstract evaluation. This initial screening phase resulted in the exclusion of 2086 records. The specific categorical reasons for exclusion at this stage were no diabetic populations (n=604), no mHealth interventions (n=635), no RCTs (n=658), and reviews or commentaries (n=189).

Full-text retrieval was sought for the remaining 214 reports, of which 2 could not be retrieved, resulting in 212 reports assessed for detailed full-text eligibility. Upon careful examination, 68 reports were excluded for the following reasons: not the target population (n=19, 27.9%), not RCTs (n=15, 22.1%), not mHealth interventions (n=2, 2.9%), and conference abstracts (n=32, 47.1%). The final sample consisted of 144 studies included in the review [9,21-163].

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the study selection process. mHealth: mobile health; RCT: randomized controlled trial.



Study Characteristics

Table S1 in Multimedia Appendix 2 delineates the comprehensive characteristics of the 144 included trials. The included studies were conducted across 49 countries. The leading single-country contributors were China (n=24, 16.7%), the United States (n=22, 15.3%), South Korea (n=9, 6.3%), India (n=6, 4.2%), and the United Kingdom (n=6, 4.2%); Spain and Iran also contributed 4.2% (n=6) of the studies each. Most studies (n=76, 52.8%) included digital literacy or device ownership as an eligibility criterion.

The number of mHealth intervention RCTs increased progressively from 2009 to 2026 (Figure 2), with studies

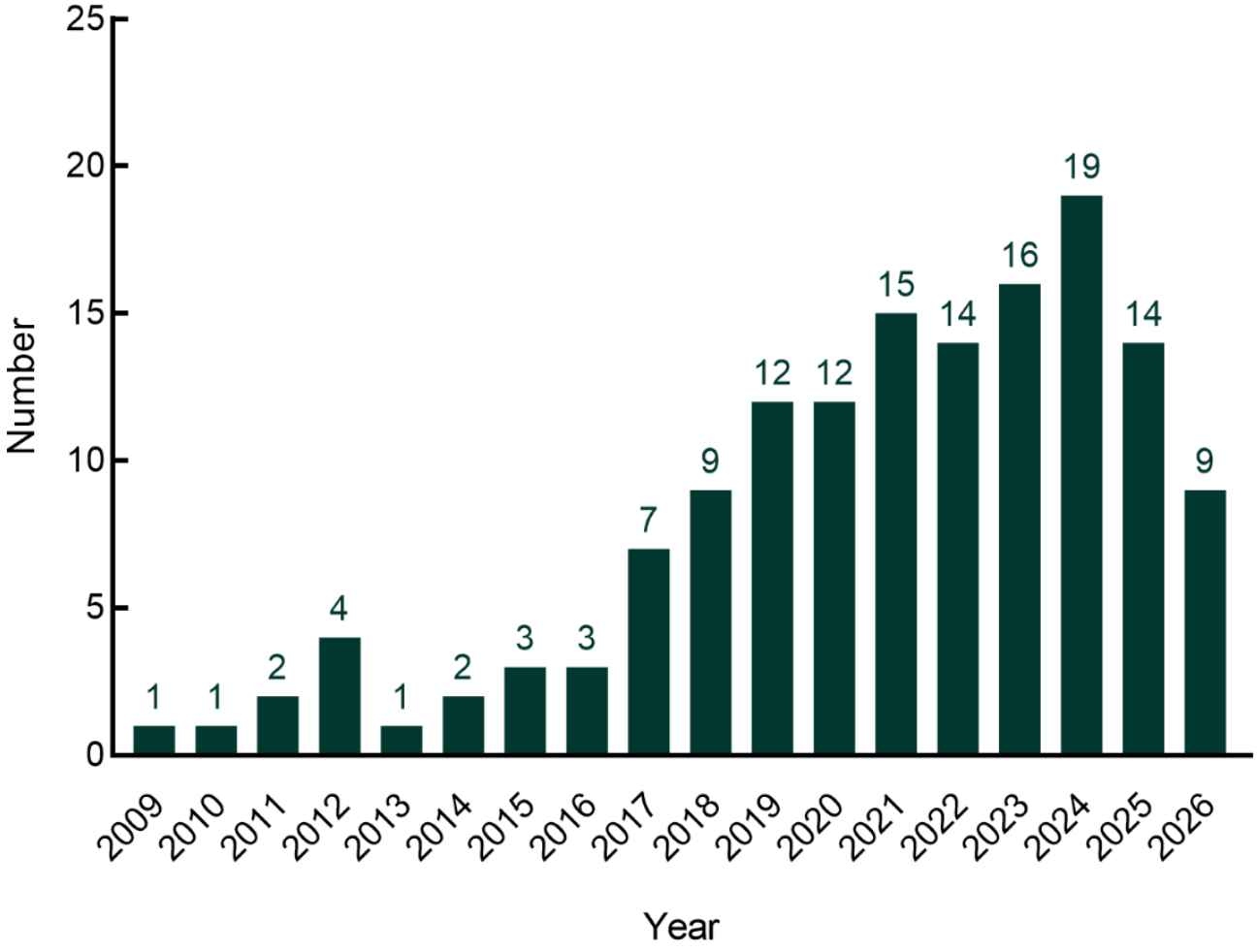
XSL•FO
RenderX

published from 2016 to 2026 accounting for 90.3% (130/144) of the included evidence. Studies published from 2020 to 2024 represented 52.8% (76/144) of all included trials.

Table 1 shows the distribution of mHealth intervention types. Single digital modalities were classified as single-component interventions, whereas combinations of technologies were classified as multicomponent interventions. Mobile app–based interventions were most frequent (74/144, 51.4%), followed by websites (20/144, 13.9%) and SMS text messaging (20/144, 13.9%). Multicomponent interventions constituted 13.9% (20/144), including mobile apps plus wearable devices (15/20, 75%), mobile apps plus SMS text messaging (4/20, 20%), and websites plus SMS text messaging (1/20, 5%).

Intervention modalities also evolved over time. Early studies (2009-2015; 14/144, 9.7%) mainly used single-channel approaches. After 2016 (130/144, 90.3% of the studies), mobile app–based interventions remained dominant, and all multicomponent interventions were from this period.

Figure 2. Number of randomized controlled trials on mobile health interventions for adult diabetes.



XSL•FO
RenderX

Table . Distribution of mobile health intervention types for adult patients with diabetes (N=144).

Intervention category	Studies, n (%)
Single-component interventions	
Mobile app	74 (51.4)
Website	20 (13.9)
SMS text messaging	20 (13.9)
Voice call	4 (2.8)
Video	5 (3.5)
Email	1 (0.7)
Multicomponent interventions	
Mobile app and wearable devices	15 (10.4)
Mobile app and SMS text messaging	4 (2.8)
Website and SMS text messaging	1 (0.7)

Quantitative Assessment of mERA Reporting Completeness

An overall reporting completeness score was calculated for each study based on the 16-item WHO mERA checklist (0 to 16 points). Across the 144 included RCTs, the mean overall mERA score was 6.74 (SD 2.71), with a median score of 6 (IQR 5-8; range 1-15). No study fulfilled all 16 criteria.

Factors Associated With mERA Reporting Completeness

Nonparametric and categorical analyses were conducted to evaluate factors associated with overall scores and item-specific reporting completeness.

Reporting completeness increased over time. Studies published after the mERA guidelines (2016-2026; 130/144, 90.3%) had higher overall reporting scores than pre-2016 publications (2009-2015; 14/144, 9.7%; median: 7 (IQR 5–8.75) vs 5 (IQR 4–5.75); mean: 6.90 (SD 2.74) vs 5.21 (SD 1.89); Mann-Whitney *U* test=528.5; *P*=.01).

After excluding 8 multicountry studies, high-income countries (HICs; 78/136, 57.4%) and LMICs (58/136, 42.6%) showed no significant difference in total reporting scores (median: 6 (IQR 5–8) vs 7 (IQR 4.25–9); mean: 6.50 (SD 2.26) vs 6.78 (SD 3.22); Mann-Whitney *U* test=2219.5; *P*=.85). Item-level analyses showed unadjusted *P*<.05 differences for items 1, 2, 6, and 12. Detailed item-level results are presented in Table 2.

XSL•FO
RenderX

Table . Item-by-item analysis of mHealth evidence reporting and assessment (mERA) reporting completeness across different intervention modalities and socioeconomic settings (N=144).

mERA core items	Overall reporting rate, n (%)	Single component (n=124), n (%)	Multicomponent (n=20), n (%)	*P* value[a]	HICs[b] (n=78), n (%)	LMICs[c] (n=58), n (%)	*P* value[d]
Item 1	21 (14.6)	19 (15.3)	2 (10)	.74	6 (7.7)	15 (25.9)	.004
Item 2	117 (81.3)	99 (79.8)	18 (90)	.37	69 (88.5)	41 (70.7)	.009
Item 3	12 (8.3)	10 (8.1)	2 (10)	.67	5 (6.4)	5 (8.6)	.74
Item 4	128 (88.9)	110 (88.7)	18 (90)	>.99	69 (88.5)	51 (87.9)	.92
Item 5	78 (54.2)	65 (52.4)	13 (65)	.30	39 (50)	31 (53.4)	.69
Item 6	23 (16.0)	19 (15.3)	4 (20)	.53	6 (7.7)	14 (24.1)	.007
Item 7	40 (27.8)	35 (28.2)	5 (25)	.77	23 (29.5)	14 (24.1)	.49
Item 8	80 (55.6)	67 (54.0)	13 (65)	.36	41 (52.6)	36 (62.1)	.27
Item 9	21 (14.6)	17 (13.7)	4 (20)	.50	11 (14.1)	8 (13.8)	.96
Item 10	70 (48.6)	58 (46.8)	12 (60)	.27	39 (50)	26 (44.8)	.55
Item 11	90 (62.5)	75 (60.5)	15 (75)	.21	48 (61.5)	35 (60.3)	.89
Item 12	78 (54.2)	69 (55.6)	9 (45)	.38	36 (46.2)	38 (65.5)	.03
Item 13	90 (62.5)	73 (58.9)	17 (85)	.03	50 (64.1)	34 (58.6)	.52
Item 14	28 (19.4)	24 (19.4)	4 (20)	>.99	18 (23.1)	6 (10.3)	.05
Item 15	41 (28.5)	32 (25.8)	9 (45)	.08	17 (21.8)	21 (36.2)	.06
Item 16	53 (36.8)	43 (34.7)	10 (50)	.19	30 (38.5)	18 (31.0)	.37

[a]*P* value derived from chi-square or Fisher exact test for the comparison between single-component and multicomponent interventions.

[b]HIC: high-income country.

[c]LMIC: low- or middle-income country.

[d]*P* value derived from the chi-square or Fisher exact test for the comparison between HICs and LMICs (excluding 8 multinational studies).

Multicomponent interventions had higher overall reporting scores than single-component interventions (mean: 7.75, SD 2.94 vs 6.57, SD 2.65; median: 7.5, IQR 5.75-9.25 vs 6, IQR 5-8; Mann-Whitney *U* test=900.5; *P*=.04). Item 13 showed an unadjusted *P*<.05 difference, which did not remain significant after Benjamini–Hochberg false discovery rate (BH-FDR) correction.

To reduce confounding from multinational frameworks, we restricted the geographical comparison to 5 leading publishing nations (China, the United States, South Korea, India, and the United Kingdom; 66/144, 45.8% of the studies). Median reporting scores were 7.5 for China, 6.0 for the United States, 7.0 for South Korea, 5.0 for India, and 5.5 for the United Kingdom. The Kruskal-Wallis *H* test did not show a significant overall difference (*H*=8.21; *P*=.08).

Findings on the Essential mHealth Criteria

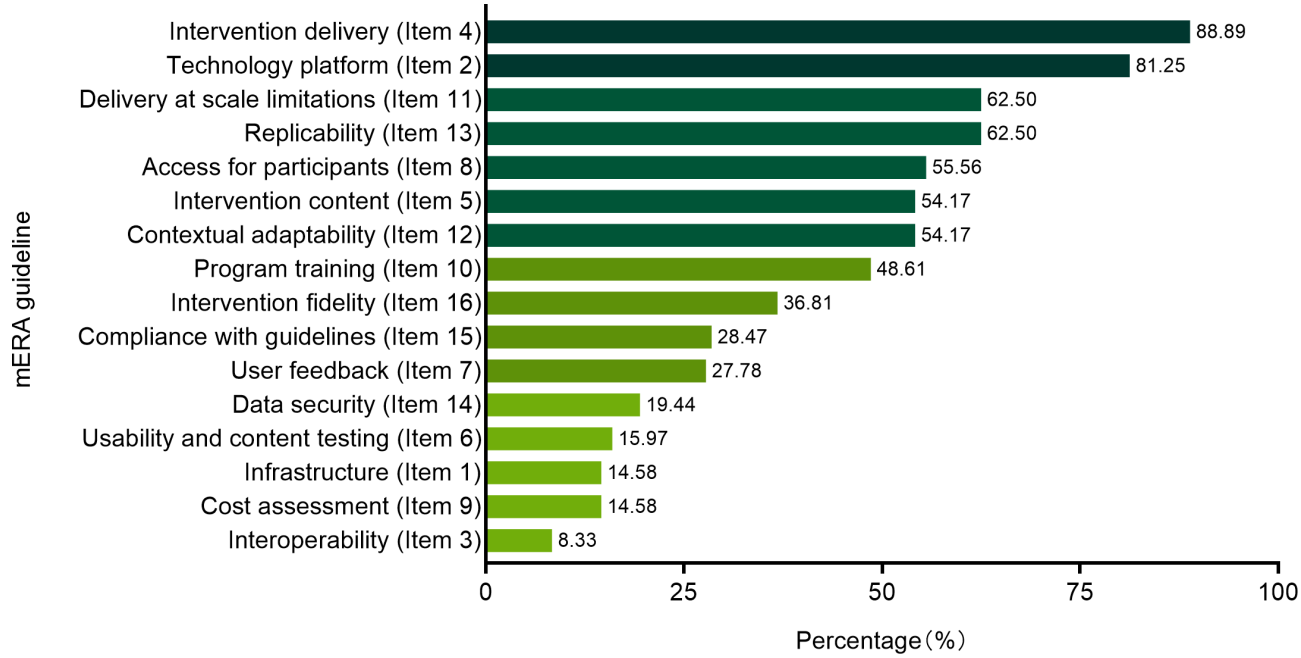
Figure 3 shows the evaluation results of the 16 basic mHealth criteria. Technology platform (item 2) was reported in 81.3% (117/144) of the studies, whereas intervention delivery (item 4) was reported in 88.9% (128/144).

Interoperability (item 3) was reported in only 8.3% (12/144) of the studies. Data security (item 14) was reported in 19.4% (28/144), and compliance with guidelines (item 15) was reported in 28.5% (41/144).

Cost assessment (item 9) was present in 14.6% (21/144) of the trials. Program training (item 10), scalability limitations (item 11), contextual adaptability (item 12), and intervention fidelity (item 16) were reported in 48.6% (70/144), 62.5% (90/144), 54.2% (78/144), and 36.8% (53/144) of the studies, respectively.

XSL•FO
RenderX

Figure 3. Reporting completeness of mobile health interventions for the 16 items of the mHealth evidence reporting and assessment (mERA) checklist.



Discussion

Principal Findings

This systematic review shows that reporting of adult diabetes mHealth RCTs remains uneven. Basic intervention delivery and technology platform information is generally described, whereas system-level and implementation domains—including interoperability, data security, regulatory compliance, and intervention fidelity—are much less consistently reported. These gaps limit readers' ability to interpret, reproduce, and translate interventions into routine practice. The updated analyses further suggest that reporting improved in more recent studies and tends to be more complete for multicomponent interventions, although substantial omissions remain. Together, these findings reinforce the central purpose of mERA: reports should provide sufficient information on intervention content, context, and technical characteristics to support interpretation and replication [14].

Comparison With Prior Work and the Real-World Consequences of Omission

Our findings extend prior evidence that evidence of clinical effectiveness in diabetes mHealth interventions does not necessarily indicate that their technical and implementation characteristics have been reported comprehensively. Previous systematic reviews have shown that mHealth interventions can improve glycemic and patient-centered outcomes while also demonstrating substantial heterogeneity in intervention delivery and implementation reporting [16]. In LMICs, for example, a systematic review of 30 studies found that implementation outcomes such as acceptability, appropriateness, and cost were reported inconsistently and that the mERA checklist had not been routinely applied [16]. Against this background, the present review shows that the reporting problem persists at the level of intervention specification itself: although core delivery characteristics are usually described, technical features needed to understand how an intervention functions and how it could be reproduced remain much less visible. The mERA framework was specifically developed to address this gap by establishing a minimum reporting set for the content, context, and technical features of mHealth interventions [14].

The low reporting of privacy, security, and regulatory information is particularly important because mHealth interventions routinely generate, transmit, or store sensitive health information. In our review, these features were frequently not described. This does not indicate that the interventions were insecure or noncompliant; rather, it indicates that these characteristics could not be established from the published reports. Previous reviews of mHealth apps have similarly identified privacy and security as important weaknesses requiring explicit attention during app design and evaluation [164]. Moreover, digital health data quality is multidimensional and closely linked to downstream clinical, research, clinician, and organizational outcomes [165]. Therefore, reporting security architecture, data governance, storage arrangements, authentication, and relevant regulatory requirements should be regarded as part of intervention characterization rather than as optional technical detail. Such information is essential for readers to judge whether an intervention can be safely transferred from a research setting into routine care.

Interoperability emerged as the most consistently overlooked core technical domain in the updated dataset. This finding is consequential because the value of many mHealth interventions depends on whether patient-generated data can move reliably among the mobile intervention, clinical information systems, and other digital health infrastructure. Standard-based approaches such as the SMART (Standard Based, Machine Readable, Adaptive, Requirement Based, and Testable) guidelines based on the Fast Healthcare Interoperability Resources demonstrate the feasibility of developing interoperable applications across heterogeneous health information systems but also highlight the practical challenges involved in moving from prototype-level interoperability to

XSL•FO
RenderX

production deployment [166]. Thus, the absence of interoperability reporting should not be interpreted simply as evidence that interoperability was absent. Instead, it represents an important reporting gap that prevents assessment of data exchange mechanisms, interface standards, integration with electronic health records, and the extent to which manual data transfer was required. Similarly, intervention fidelity was often insufficiently reported, limiting the ability to determine whether observed effects were attributable to the intended intervention or to deviations in delivery, engagement, or implementation intensity. These omissions are particularly important when evaluating complex, multicomponent interventions.

The updated analyses also modify the interpretation of temporal change and intervention complexity. Studies published from 2016 onward had significantly higher reporting scores than those published from 2009 to 2015, suggesting an improvement in reporting completeness after the introduction and dissemination of the mERA checklist and related digital health reporting standards [14,138]. However, the overall median score of 6 out of 16 (IQR 5-8) indicates that this improvement did not translate into comprehensive reporting across all domains. Multicomponent interventions also tended to be more completely reported than single-component interventions. This association should be interpreted cautiously: it does not establish that technological complexity itself causes better reporting. Multicomponent interventions may instead include more detailed methodological descriptions because they involve more components, delivery pathways, or implementation procedures. Importantly, the present data do not support the earlier claim of an inverse relationship between technological complexity and reporting transparency. More broadly, the distinction between methodological quality and reporting completeness remains essential. Reporting guidelines such as CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) complement rather than replace design-specific risk-of-bias assessment by encouraging authors to describe intervention characteristics, context, and implementation sufficiently for readers to evaluate applicability and reproducibility [138].

Limitations

Several limitations should be considered when interpreting these findings. First, reporting completeness was assessed from the information available in published articles and supplementary materials; a low reporting score, therefore, indicates lack of documented information rather than confirmed absence of the underlying technical feature. Second, the analysis was designed as a meta-research assessment of reporting quality rather than a clinical effectiveness meta-analysis, so the relationship between reporting completeness and treatment effect could not be examined. Third, although the dataset was expanded to 144 trials and included studies published up to 2026, the evidence base remains heterogeneous with respect to technologies, intervention components, countries, and publication periods. Fourth, the HIC vs LMIC comparison excluded 8 multinational studies, and the country-level Kruskal-Wallis analysis was limited to 5 selected high-contributing countries; these analyses, therefore, should be interpreted as exploratory rather than definitive explanations of geographic differences. Finally, the present study focused on RCTs published in English, which may limit generalizability to pragmatic evaluations, observational studies, implementation studies, and non–English-language evidence, particularly from settings with fewer research and publication resources.

Conclusions

This updated review of 144 adult diabetes mHealth RCTs published from 2009 to 2026 demonstrates a persistent gap between reporting of intervention delivery and reporting of the technical and implementation characteristics needed for replication and translation. Overall mERA reporting remained modest (median 6/16, IQR 5-8). Although technology platform and intervention delivery were reported in 81.3% (117/144) and 88.9% (128/144) of the studies, respectively, interoperability (12/144, 8.3%), data security (28/144, 19.4%), regulatory compliance (41/144, 28.5%), and intervention fidelity (53/144, 36.8%) remained incompletely reported. Reporting completeness was significantly higher in studies published from 2016 to 2026 and in multicomponent compared with single-component interventions, but these associations do not eliminate substantial deficiencies in system-level reporting.

These findings suggest that improving digital health evidence requires reporting standards to be implemented alongside conventional trial reporting requirements. The mERA checklist provides a structured minimum set of items for describing mHealth interventions, whereas the CONSORT-EHEALTH checklist provides complementary guidance for reporting randomized evaluations of eHealth and mHealth interventions [14,138]. Future diabetes mHealth trials should explicitly report interoperability mechanisms, data security and governance, regulatory considerations, intervention fidelity, and implementation procedures. Such reporting would improve reproducibility, facilitate evidence synthesis, and allow health care organizations to assess whether interventions can be transferred safely and sustainably into routine practice. The priority should therefore shift from documenting only whether an intervention works to documenting sufficiently how, where, and under what technical and implementation conditions it works.

Acknowledgments

The authors declare the use of generative AI (GenAI) in the research and writing process. According to the Generative AI Delegation Taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: proofreading and editing and translation. The GenAI tool used was Google Gemini 3.1 Pro. Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

XSL•FO
RenderX

Funding

The authors declared no financial support was received for this work.

Authors' Contributions

XY, X Huang, and MH contributed to the systematic review's conception and research question. XY contributed to the database searches. XY, X Huang, and MH participated in the selection of included studies. XY, JS, and JP extracted data from the selected studies. XY, X Hu, and MH were involved in statistical analysis and manuscript drafting. All authors gave final approval for the version to be published and agree to be accountable for all aspects of the work, ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies used for each database in this systematic review..
[DOCX File, 18 KB - mhealth_v14i1e95863_app1.docx]

Multimedia Appendix 2
Item level mERA assessment data and supplementary results for the included studies.
[XLSX File, 24 KB - mhealth_v14i1e95863_app2.xlsx]

Checklist 1
PRISMA 2020 checklist.
[DOCX File, 82 KB - mhealth_v14i1e95863_app3.docx]

References

1. Tang Z, Zhao L, Li J, et al. Prognostic effectiveness of interactive vs. non-interactive mobile app interventions in type 2 diabetes: a systematic review and meta-analysis. Arch Public Health 2024 Nov 22;82(1):221. [doi: 10.1186/s13690-024-01450-x] [Medline: 39578910]
2. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes prevalence and treatment from 1990 to 2022: a pooled analysis of 1108 population-representative studies with 141 million participants. Lancet 2024 Nov 23;404(10467):2077-2093. [doi: 10.1016/S0140-6736(24)02317-1] [Medline: 39549716]
3. Bommer C, Sagalova V, Heesemann E, et al. Global economic burden of diabetes in adults: projections from 2015 to 2030. Diabetes Care 2018 May;41(5):963-970. [doi: 10.2337/dc17-1962] [Medline: 29475843]
4. American Diabetes Association Professional Practice Committee. 1. Improving care and promoting health in populations: Standards of Care in Diabetes-2025. Diabetes Care 2025 Jan 1;48(1 Suppl 1):S14-S26. [doi: 10.2337/dc25-S001] [Medline: 39651974]
5. Beck J, Greenwood DA, Blanton L, et al. 2017 National standards for Diabetes Self-Management Education and Support. Sci Diabetes Self Manag Care 2021 Feb;47(1):14-29. [doi: 10.1177/0145721720987926] [Medline: 34078205]
6. Moschonis G, Siopis G, Jung J, et al. Effectiveness, reach, uptake, and feasibility of digital health interventions for adults with type 2 diabetes: a systematic review and meta-analysis of randomised controlled trials. Lancet Digit Health 2023 Mar;5(3):e125-e143. [doi: 10.1016/S2589-7500(22)00233-3] [Medline: 36828606]
7. Yu X, Wang Y, Liu Z, Jung E. Technological functionality and system architecture of mobile health interventions for diabetes management: a systematic review and meta-analysis of randomized controlled trials. Front Public Health 2025;13:1549568. [doi: 10.3389/fpubh.2025.1549568] [Medline: 40051514]
8. Hilmarsdóttir E, Sigurðardóttir Á, Arnardóttir RH. A digital lifestyle program in outpatient treatment of type 2 diabetes: a randomized controlled study. J Diabetes Sci Technol 2021 Sep;15(5):1134-1141. [doi: 10.1177/1932296820942286] [Medline: 32680441]
9. Heald AH, Roberts S, Albeda Gimeno L, et al. A randomised control trial to explore the impact and efficacy of the Healum collaborative care planning software and app on condition management in the type 2 diabetes mellitus population in NHS primary care. Diabetes Ther 2023 Jun;14(6):977-988. [doi: 10.1007/s13300-023-01404-6] [Medline: 37079268]
10. Joiner KL, Nam S, Whittemore R. Lifestyle interventions based on the diabetes prevention program delivered via eHealth: a systematic review and meta-analysis. Prev Med 2017 Jul;100:194-207. [doi: 10.1016/j.ypmed.2017.04.033] [Medline: 28456513]
11. Yu X, Chau JP, Huo L. The effectiveness of traditional Chinese medicine-based lifestyle interventions on biomedical, psychosocial, and behavioral outcomes in individuals with type 2 diabetes: a systematic review with meta-analysis. Int J Nurs Stud 2018 Apr;80:165-180. [doi: 10.1016/j.ijnurstu.2018.01.009] [Medline: 29471267]

XSL•FO
RenderX

12. Hoque MR, Rahman MS, Nipa NJ, Hasan MR. Mobile health interventions in developing countries: a systematic review. Health Informatics J 2020 Dec;26(4):2792-2810. [doi: 10.1177/1460458220937102] [Medline: 32691659]
13. White BK, Burns SK, Giglia RC, Scott JA. Designing evaluation plans for health promotion mHealth interventions: a case study of the Milk Man mobile app. Health Promot J Austr 2016 Feb;27(3):198-203. [doi: 10.1071/HE16041] [Medline: 27784506]
14. Agarwal S, LeFevre AE, Lee J, et al. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
15. Bashi N, Fatehi F, Fallah M, Walters D, Karunanithi M. Self-management education through mHealth: review of strategies and structures. JMIR Mhealth Uhealth 2018 Oct 19;6(10):e10771. [doi: 10.2196/10771] [Medline: 30341042]
16. Mokaya M, Kyallo F, Vangoitsenhoven R, Matthys C. Clinical and patient-centered implementation outcomes of mHealth interventions for type 2 diabetes in low-and-middle income countries: a systematic review. Int J Behav Nutr Phys Act 2022 Jan 6;19(1):1. [doi: 10.1186/s12966-021-01238-0] [Medline: 34991593]
17. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097. [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
18. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017;12(2):e0170581. [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
19. Burla L, Knierim B, Barth J, Liewald K, Duetz M, Abel T. From text to codings: intercoder reliability assessment in qualitative content analysis. Nurs Res 2008;57(2):113-117. [doi: 10.1097/01.NNR.0000313482.33917.7d] [Medline: 18347483]
20. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282. [Medline: 23092060]
21. Clarke J, Sanatkar S, Baldwin PA, et al. A web-based cognitive behavior therapy intervention to improve social and occupational functioning in adults with type 2 diabetes (the Springboard trial): randomized controlled trial. J Med Internet Res 2019 May 21;21(5):e12246. [doi: 10.2196/12246] [Medline: 31115345]
22. Baldwin PA, Sanatkar S, Clarke J, et al. A web-based mental health intervention to improve social and occupational functioning in adults with type 2 diabetes (the Springboard trial): 12-month outcomes of a randomized controlled trial. J Med Internet Res 2020 Dec 1;22(12):e16729. [doi: 10.2196/16729] [Medline: 33258790]
23. Middleton T, Constantino M, McGill M, et al. An enhanced SMS text message-based support and reminder program for young adults with type 2 diabetes (TEXT2U): randomized controlled trial. J Med Internet Res 2021 Oct 21;23(10):e27263. [doi: 10.2196/27263] [Medline: 34524102]
24. Newby J, Robins L, Wilhelm K, et al. Web-based cognitive behavior therapy for depression in people with diabetes mellitus: a randomized controlled trial. J Med Internet Res 2017 May 15;19(5):e157. [doi: 10.2196/jmir.7274] [Medline: 28506956]
25. Islam SM, George ES, Maddison R. Effectiveness of a mobile phone text messaging intervention on dietary behaviour in patients with type 2 diabetes: a post-hoc analysis of a randomised controlled trial. Mhealth 2021;7:10. [doi: 10.21037/mhealth-2020-3] [Medline: 33634193]
26. Banu B, Ko KC, Khan M, et al. Effects of traditional versus m-Health educational interventions for diabetic patients: a randomised controlled trial in peripheral district of Bangladesh. Diabetes Epidemiol Manag 2023;9:100106. [doi: 10.1016/j.deman.2022.100106]
27. Araya R, Menezes PR, Claro HG, et al. Effect of a digital intervention on depressive symptoms in patients with comorbid hypertension or diabetes in Brazil and Peru: two randomized clinical trials. JAMA 2021 May 11;325(18):1852-1862. [doi: 10.1001/jama.2021.4348] [Medline: 33974019]
28. Ferreira J, Cruvinel-Júnior RH, da Silva EQ, et al. Effectiveness of a web-based foot-ankle exercise program for treating ulcer risk factors in diabetic neuropathy in a randomized controlled trial. Sci Rep 2024 Nov 8;14(1):27291. [doi: 10.1038/s41598-024-78188-7] [Medline: 39516524]
29. Câmara de Souza AB, Toyoshima MT, Cukier P, et al. Electronic glycemic management system improved glycemic control and reduced complications in patients with diabetes undergoing coronary artery bypass surgery: a randomized controlled trial. J Diabetes Sci Technol 2026 Mar;20(2):308-316. [doi: 10.1177/19322968241268352] [Medline: 39096188]
30. Steinman L, van Pelt M, Hen H, et al. Can mHealth and eHealth improve management of diabetes and hypertension in a hard-to-reach population? -Lessons learned from a process evaluation of digital health to support a peer educator model in Cambodia using the RE-AIM framework. Mhealth 2020;6:40. [doi: 10.21037/mhealth-19-249] [Medline: 33437836]
31. Van Olmen J, Kegels G, Korachais C, et al. The effect of text message support on diabetes self-management in developing countries - a randomised trial. J Clin Transl Endocrinol 2017;7:33-41. [doi: 10.1016/j.jcte.2016.12.005] [Medline: 29067248]
32. Lemelin A, Godbout A, Paré G, Bernard S. Improved glycemic control through the use of a telehomecare program in patients with diabetes treated with insulin. Diabetes Technol Ther 2020 Apr;22(4):243-248. [doi: 10.1089/dia.2019.0324] [Medline: 31657625]

XSL•FO
RenderX

33. Agarwal P, Mukerji G, Desveaux L, et al. Mobile app for improved self-management of type 2 diabetes: multicenter pragmatic randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 10;7(1):e10321. [doi: 10.2196/10321] [Medline: 30632972]
34. Logan AG, Irvine MJ, McIsaac WJ, et al. Effect of home blood pressure telemonitoring with self-care support on uncontrolled systolic hypertension in diabetics. Hypertension 2012 Jul;60(1):51-57. [doi: 10.1161/HYPERTENSIONAHA.111.188409] [Medline: 22615116]
35. Jia W, Zhang P, Zhu D, et al. Evaluation of an mHealth-enabled hierarchical diabetes management intervention in primary care in China (ROADMAP): a cluster randomized trial. PLoS Med 2021 Sep;18(9):e1003754. [doi: 10.1371/journal.pmed.1003754] [Medline: 34547030]
36. Wang X, Liu D, Du M, Hao R, Zheng H, Yan C. The role of text messaging intervention in Inner Mongolia among patients with type 2 diabetes mellitus: a randomized controlled trial. BMC Med Inform Decis Mak 2020 May 14;20(1):90. [doi: 10.1186/s12911-020-01129-7] [Medline: 32410608]
37. Huo X, Krumholz HM, Bai X, et al. Effects of mobile text messaging on glycemic control in patients with coronary heart disease and diabetes mellitus: a randomized clinical trial. Circ Cardiovasc Qual Outcomes 2019 Sep;12(9):e005805. [doi: 10.1161/CIRCOUTCOMES.119.005805] [Medline: 31474119]
38. Zhang B, Kalampakorn S, Powwattana A, Sillabutra J, Liu G. A transtheoretical model-based online intervention to improve medication adherence for Chinese adults newly diagnosed with type 2 diabetes: a mixed-method study. J Prim Care Community Health 2024;15:21501319241263657. [doi: 10.1177/21501319241263657] [Medline: 39077970]
39. Zhang L, He X, Shen Y, et al. Effectiveness of smartphone app-based interactive management on glycemic control in Chinese patients with poorly controlled diabetes: randomized controlled trial. J Med Internet Res 2019 Dec 9;21(12):e15401. [doi: 10.2196/15401] [Medline: 31815677]
40. Xia SF, Maitiniyazi G, Chen Y, et al. Web-based TangPlan and WeChat combination to support self-management for patients with type 2 diabetes: randomized controlled trial. JMIR Mhealth Uhealth 2022 Mar 30;10(3):e30571. [doi: 10.2196/30571] [Medline: 35353055]
41. Kang J, Chen Y, Zhao Y, Zhang C. Effect of remote management on comprehensive management of diabetes mellitus during the COVID-19 epidemic. Prim Care Diabetes 2021 Jun;15(3):417-423. [doi: 10.1016/j.pcd.2020.12.004] [Medline: 33422431]
42. Yang L, Xu J, Kang C, et al. Effects of mobile phone-based telemedicine management in patients with type 2 diabetes mellitus: a randomized clinical trial. Am J Med Sci 2022 Mar;363(3):224-231. [doi: 10.1016/j.amjms.2021.09.001] [Medline: 34534510]
43. Luo E, Wan J, Su M, et al. Efficacy of a basal insulin dose management smartphone application for controlling fasting blood glucose in patients with type-2 diabetes mellitus: a single-centre, randomised clinical study. Clin Endocrinol (Oxf) 2023 Oct;99(4):361-369. [doi: 10.1111/cen.14956] [Medline: 37528510]
44. Fang R, Deng X. Electronic messaging intervention for management of cardiovascular risk factors in type 2 diabetes mellitus: a randomised controlled trial. J Clin Nurs 2018 Feb;27(3-4):612-620. [doi: 10.1111/jocn.13962] [Medline: 28700102]
45. Feng Y, Zhao Y, Mao L, et al. The effectiveness of an eHealth family-based intervention program in patients with uncontrolled type 2 diabetes mellitus (T2DM) in the community via WeChat: randomized controlled trial. JMIR Mhealth Uhealth 2023 Mar 20;11(1):e40420. [doi: 10.2196/40420] [Medline: 36939825]
46. Dong Y, Wang P, Dai Z, et al. Increased self-care activities and glycemic control rate in relation to health education via WeChat among diabetes patients: a randomized clinical trial. Medicine (Baltimore) 2018;97(50):e13632. [doi: 10.1097/MD.0000000000013632] [Medline: 30558051]
47. Zhang W, Zhu X, Que X, Zhang X. The effects of shared decision-making informed dietary intervention based on digital health technology in older adults with type 2 diabetes mellitus: a randomized controlled trial. Digit Health 2024;10:20552076241272514. [doi: 10.1177/20552076241272514] [Medline: 39403713]
48. Xu Z, Geng J, Zhang S, et al. A mobile-based intervention for dietary behavior and physical activity change in individuals at high risk for type 2 diabetes mellitus: randomized controlled trial. JMIR Mhealth Uhealth 2020 Nov 3;8(11):e19869. [doi: 10.2196/19869] [Medline: 33141092]
49. Zhai Y, Yu W. A mobile app for diabetes management: impact on self-efficacy among patients with type 2 diabetes at a community hospital. Med Sci Monit 2020 Nov 16;26:e926719. [doi: 10.12659/MSM.926719] [Medline: 33196634]
50. Yan M, Yu Y, Li S, Zhang P, Yu J. Effectiveness of King's Theory of Goal Attainment in blood glucose management for newly diagnosed patients with type 2 diabetes: randomized controlled trial. J Med Internet Res 2024 Oct 31;26:e59142. [doi: 10.2196/59142] [Medline: 39481094]
51. Guo M, Meng F, Guo Q, et al. Effectiveness of mHealth management with an implantable glucose sensor and a mobile application among Chinese adults with type 2 diabetes. J Telemed Telecare 2023 Sep;29(8):632-640. [doi: 10.1177/1357633X211020261] [Medline: 34152238]
52. María Gómez A, Cristina Henao D, León Vargas F, et al. Efficacy of the mHealth application in patients with type 2 diabetes transitioning from inpatient to outpatient care: a randomized controlled clinical trial. Diabetes Res Clin Pract 2022;189:109948. [doi: 10.1016/j.diabres.2022.109948] [Medline: 35700926]

XSL•FO
RenderX

53. Moravcová K, Sovová M, Ožana J, et al. Comparing the efficacy of digital and in-person weight loss interventions for patients with obesity and glycemic disorders: evidence from a randomized non-inferiority trial. Nutrients 2024 May 16;16(10):1510. [doi: 10.3390/nu16101510] [Medline: 38794747]
54. Thorsen IK, Yang Y, Valentiner LS, et al. The effects of a lifestyle intervention supported by the InterWalk smartphone app on increasing physical activity among persons with type 2 diabetes: parallel-group, randomized trial. JMIR Mhealth Uhealth 2022 Sep 28;10(9):e30602. [doi: 10.2196/30602] [Medline: 36170002]
55. Abaza H, Marschollek M. SMS education for the promotion of diabetes self-management in low & middle income countries: a pilot randomized controlled trial in Egypt. BMC Public Health 2017 Dec 19;17(1):962. [doi: 10.1186/s12889-017-4973-5] [Medline: 29258499]
56. Franc S, Joubert M, Daoudi A, et al. Efficacy of two telemonitoring systems to improve glycaemic control during basal insulin initiation in patients with type 2 diabetes: the TeleDiab-2 randomized controlled trial. Diabetes Obes Metab 2019 Oct;21(10):2327-2332. [doi: 10.1111/dom.13806] [Medline: 31173451]
57. Hermanns N, Ehrmann D, Finke-Groene K, et al. Use of smartphone application versus written titration charts for basal insulin titration in adults with type 2 diabetes and suboptimal glycaemic control (My Dose Coach): multicentre, open-label, parallel, randomised controlled trial. Lancet Reg Health Eur 2023 Oct;33:100702. [doi: 10.1016/j.lanepe.2023.100702] [Medline: 37954005]
58. Ebert DD, Nobis S, Lehr D, et al. The 6-month effectiveness of internet-based guided self-help for depression in adults with type 1 and 2 diabetes mellitus. Diabet Med 2017 Jan;34(1):99-107. [doi: 10.1111/dme.13173] [Medline: 27334444]
59. Fountoulakis S, Papanastasiou L, Gryparis A, Markou A, Piaditis G. Impact and duration effect of telemonitoring on HbA1c, BMI and cost in insulin-treated diabetes mellitus patients with inadequate glycemic control: a randomized controlled study. Hormones (Athens) 2015;14(4):632-643. [doi: 10.14310/horm.2002.1603] [Medline: 26188234]
60. Bajaj HS, Aberle J, Davies M, et al. Once-weekly insulin icodec with dosing guide app versus once-daily basal insulin analogues in insulin-naive type 2 diabetes (ONWARDS 5): a randomized trial. Ann Intern Med 2023 Nov;176(11):1476-1485. [doi: 10.7326/M23-1288] [Medline: 37748181]
61. Shamanna P, Joshi S, Thajudeen M, et al. Personalized nutrition in type 2 diabetes remission: application of digital twin technology for predictive glycemic control. Front Endocrinol (Lausanne) 2024;15:1485464. [doi: 10.3389/fendo.2024.1485464] [Medline: 39634180]
62. Prabhakaran D, Jha D, Prieto-Merino D, et al. Effectiveness of an mHealth-based electronic decision support system for integrated management of chronic conditions in primary care: the mWellcare cluster-randomized controlled trial. Circulation 2019 Jan 15;139(3):380-391. [doi: 10.1161/CIRCULATIONAHA.118.038192] [Medline: 30586732]
63. Patnaik L, Panigrahi SK, Sahoo AK, Mishra D, Muduli AK, Beura S. Effectiveness of mobile application for promotion of physical activity among newly diagnosed patients of type II diabetes - a randomized controlled trial. Int J Prev Med 2022;13:54. [doi: 10.4103/ijpvm.IJPVM_92_20] [Medline: 35706879]
64. Vinitha R, Nanditha A, Snehalatha C, et al. Effectiveness of mobile phone text messaging in improving glycaemic control among persons with newly detected type 2 diabetes. Diabetes Res Clin Pract 2019 Dec;158:107919. [doi: 10.1016/j.diabres.2019.107919] [Medline: 31711858]
65. Kumar D, Raina S, Sharma SB, Raina SK, Bhardwaj AK. Effectiveness of randomized control trial of mobile phone messages on control of fasting blood glucose in patients with type-2 diabetes mellitus in a Northern State of India. Indian J Public Health 2018;62(3):224-226. [doi: 10.4103/ijph.IJPH_199_17] [Medline: 30232974]
66. Sunil Kumar D, Prakash B, Subhash Chandra BJ, et al. Technological innovations to improve health outcome in type 2 diabetes mellitus: a randomized controlled study. Clin Epidemiol Glob Health 2021;9:53-56. [doi: 10.1016/j.cegh.2020.06.011]
67. Budhiwardoyo N, Kusumo MP. Transforming diabetes management: a study on the effectiveness of smartphone applications in enhancing knowledge, attitudes, and behavior of people with diabetes mellitus. Open Access Maced J Med Sci 2023;11(E):388-393. [doi: 10.3889/oamjms.2023.11601]
68. Noviani L, Diantini A, Subarnas A. Collaboration between interprofessional healthcare and patients to improve quality of life type 2 diabetes via smartphone application. Int J Pharm Res 2021;13(1):645. [doi: 10.31838/ijpr/2021.13.01.126]
69. Goodarzi M, Ebrahimzadeh I, Rabi A, Saedipoor B, Jafarabadi MA. Impact of distance education via mobile phone text messaging on knowledge, attitude, practice and self efficacy of patients with type 2 diabetes mellitus in Iran. J Diabetes Metab Disord 2012 Aug 31;11(1):10. [doi: 10.1186/2251-6581-11-10] [Medline: 23497632]
70. Mirzaei-Alavijeh M, Hosseini SN, Niksirt M, Hashemian AH, Khashij S, Jalilian F. The efficacy of theory driven treatment adherence promotion program among type 2 diabetic patients: application of intervention mapping and mHealth. J Diabetes Metab Disord 2023 Dec;22(2):1609-1615. [doi: 10.1007/s40200-023-01291-5] [Medline: 37975125]
71. Lari H, Noroozi A, Tahmasebi R. Comparison of multimedia and short message services education on the physical activity of diabetic patients: an application of health promotion model. Iran Red Crescent Med J 2018;20(S1):e59800. [doi: 10.5812/ircmj.59800]
72. Zanjanchi Neko F, Mohammadi Zeidi I, Morshedi H, Mohammadi Zeidi B, Maleki MR, Pakpoor Hajia A. Effectiveness of theory-based intervention on knowledge level, psychological constructs, metabolic index and physical activity status in patients with type 2 diabetes: application of the Health Action Process Approach (HAPA) model. Clin Diabetol 2023;12(5):290-300. [doi: 10.5603/cd.96203]

XSL•FO
RenderX

73. Castelnuovo G, Manzoni GM, Cuzziol P, et al. TECNOB study: ad interim results of a randomized controlled trial of a multidisciplinary telecare intervention for obese patients with type-2 diabetes. Clin Pract Epidemiol Ment Health 2011 Mar 4;7(1):44-50. [doi: 10.2174/1745017901107010044] [Medline: 21559233]
74. Leung N, Waki K, Nozoe S, et al. Efficacy of Save Medical Corporation (SMC)-01, a smartphone app designed to support type 2 diabetes self-management based on established guidelines: randomized controlled trial. J Med Internet Res 2024 Sep 10;26:e53740. [doi: 10.2196/53740] [Medline: 39255478]
75. Waki K, Nara M, Enomoto S, et al. Effectiveness of DialBetesPlus, a self-management support system for diabetic kidney disease: randomized controlled trial. NPJ Digit Med 2024;7(1):104. [doi: 10.1038/s41746-024-01114-8] [Medline: 38678094]
76. Waki K, Aizawa K, Kato S, et al. DialBetics with a multimedia food recording tool, FoodLog: smartphone-based self-management for type 2 diabetes. J Diabetes Sci Technol 2015;9(3):534-540. [doi: 10.1177/1932296815579690] [Medline: 25883164]
77. Kitazawa M, Takeda Y, Hatta M, et al. Lifestyle intervention with smartphone app and isCGM for people at high risk of type 2 diabetes: randomized trial. J Clin Endocrinol Metab 2024 Mar 15;109(4):1060-1070. [doi: 10.1210/clinem/dgad639] [Medline: 37931069]
78. Cho JH, Kim HS, Yoo SH, et al. An internet-based health gateway device for interactive communication and automatic data uploading: clinical efficacy for type 2 diabetes in a multi-centre trial. J Telemed Telecare 2017 Jul;23(6):595-604. [doi: 10.1177/1357633X16657500] [Medline: 27381040]
79. Yoo HJ, Park MS, Kim TN, et al. A Ubiquitous Chronic Disease Care system using cellular phones and the internet. Diabet Med 2009 Jun;26(6):628-635. [doi: 10.1111/j.1464-5491.2009.02732.x] [Medline: 19538239]
80. Lee DY, Park J, Choi D, Ahn HY, Park SW, Park CY. The effectiveness, reproducibility, and durability of tailored mobile coaching on diabetes management in policyholders: a randomized, controlled, open-label study. Sci Rep 2018 Feb 26;8(1):3642. [doi: 10.1038/s41598-018-22034-0] [Medline: 29483559]
81. Lee DY, Yoo SH, Min KP, Park CY. Effect of voluntary participation on mobile health care in diabetes management: randomized controlled open-label trial. JMIR Mhealth Uhealth 2020 Sep 18;8(9):e19153. [doi: 10.2196/19153] [Medline: 32945775]
82. Yang Y, Lee EY, Kim HS, Lee SH, Yoon KH, Cho JH. Effect of a mobile phone-based glucose-monitoring and feedback system for type 2 diabetes management in multiple primary care clinic settings: cluster randomized controlled trial. JMIR Mhealth Uhealth 2020 Feb 26;8(2):e16266. [doi: 10.2196/16266] [Medline: 32130172]
83. Cho JH, Choi YH, Kim HS, Lee JH, Yoon KH. Effectiveness and safety of a glucose data-filtering system with automatic response software to reduce the physician workload in managing type 2 diabetes. J Telemed Telecare 2011;17(5):257-262. [doi: 10.1258/jtt.2011.101006] [Medline: 21628421]
84. Lee EY, Cha SA, Yun JS, et al. Efficacy of personalized diabetes self-care using an electronic medical record-integrated mobile app in patients with type 2 diabetes: 6-month randomized controlled trial. J Med Internet Res 2022 Jul 28;24(7):e37430. [doi: 10.2196/37430] [Medline: 35900817]
85. Lee SE, Park SK, Park YS, Kim KA, Choi HS, Oh SW. Effects of short-term mobile application use on weight reduction for patients with type 2 diabetes. J Obes Metab Syndr 2021 Dec 30;30(4):345-353. [doi: 10.7570/jomes21047] [Medline: 34875628]
86. Farmer A, Bobrow K, Leon N, et al. Digital messaging to support control for type 2 diabetes (StAR2D): a multicentre randomised controlled trial. BMC Public Health 2021 Oct 21;21(1):1907. [doi: 10.1186/s12889-021-11874-7] [Medline: 34674688]
87. Hidrus A, Kueh YC, Norsaádah B, et al. Effects of brain breaks videos on the motives for the physical activity of Malaysians with type-2 diabetes mellitus. Int J Environ Res Public Health 2020 Apr 6;17(7):2507. [doi: 10.3390/ijerph17072507] [Medline: 32268601]
88. Derkaoui N, Benyakhlef S, Rami I, et al. Effectiveness of a new smartphone application on type 1 diabetes control and self-management in times of Covid-19: prospective comparative study. Rev Diabet Stud 2023;19(2):71-76. [doi: 10.1900/RDS.2023.19.71]
89. Kooiman TJ, de Groot M, Hoogenberg K, Krijnen WP, van der Schans CP, Kooy A. Self-tracking of physical activity in people with type 2 diabetes: a randomized controlled trial. Comput Inform Nurs 2018 Jul;36(7):340-349. [doi: 10.1097/CIN.0000000000000443] [Medline: 29742550]
90. Vluggen S, Candel M, Hoving C, Schaper NC, de Vries H. A web-based computer-tailored program to improve treatment adherence in patients with type 2 diabetes: randomized controlled trial. J Med Internet Res 2021 Feb 23;23(2):e18524. [doi: 10.2196/18524] [Medline: 33620321]
91. Kirwan M, Vandelanotte C, Fenning A, Duncan MJ. Diabetes self-management smartphone application for adults with type 1 diabetes: randomized controlled trial. J Med Internet Res 2013 Nov 13;15(11):e235. [doi: 10.2196/jmir.2588] [Medline: 24225149]
92. McLeod M, Stanley J, Signal V, et al. Impact of a comprehensive digital health programme on HbA1c and weight after 12 months for people with diabetes and prediabetes: a randomised controlled trial. Diabetologia 2020 Dec;63(12):2559-2570. [doi: 10.1007/s00125-020-05261-x] [Medline: 32886192]

XSL•FO
RenderX

93. Burnside MJ, Lewis DM, Crocket HR, et al. Open-source automated insulin delivery in type 1 diabetes. N Engl J Med 2022 Sep 8;387(10):869-881. [doi: 10.1056/NEJMoa2203913] [Medline: 36069869]
94. Ihekoronye MR, Osemene KP, Oamen TE. Pharmacist-led intervention to improve treatment outcomes in type 2 diabetes: a randomized controlled trial. J Pharm Health Serv Res 2024 Jun 1;15:rmae005. [doi: 10.1093/jphsr/rmae005]
95. Shahid M, Mahar SA, Shaikh S, Shaikh ZU. Mobile phone intervention to improve diabetes care in rural areas of Pakistan: a randomized controlled trial. J Coll Physicians Surg Pak 2015 Mar;25(3):166-171. [Medline: 25772954]
96. Sani M, Makeen A, Albasheer OB, Solan YM, Mahfouz MS. Effect of telemedicine messages integrated with peer group support on glycemic control in type 2 diabetics, Kingdom of Saudi Arabia. Int J Diabetes Dev Ctries 2018 Oct;38:495-501. [doi: 10.1007/s13410-018-0608-3]
97. Lim SL, Ong KW, Johal J, et al. Effect of a smartphone app on weight change and metabolic outcomes in Asian adults with type 2 diabetes: a randomized clinical trial. JAMA Netw Open 2021 Jun 1;4(6):e2112417. [doi: 10.1001/jamanetworkopen.2021.12417] [Medline: 34081137]
98. Finkelstein EA, Gardner DS, Tham KW, et al. Effectiveness and cost-effectiveness of an app and rewards-based intervention in type 2 diabetes: a randomised controlled trial. Diabetes Obes Metab 2025 Feb;27(2):729-739. [doi: 10.1111/dom.16067] [Medline: 39638769]
99. Jiang Y, Ramachandran HJ, Teo JY, et al. Effectiveness of a nurse-led smartphone-based self-management programme for people with poorly controlled type 2 diabetes: a randomized controlled trial. J Adv Nurs 2022;78(4):1154-1165. [doi: 10.1111/jan.15178] [Medline: 35170786]
100. Tan NC, Tyagi S, Lee CS, et al. Effectiveness of an algorithm-driven home telemonitoring system on the metabolic control and self-care behaviour of Asian adults with type-2 diabetes mellitus: a randomised controlled trial. J Telemed Telecare 2025;31(4):523-533. [doi: 10.1177/1357633X231203490] [Medline: 37753721]
101. Aguilera A, Arévalo Avalos M, Xu J, et al. Effectiveness of a digital health intervention leveraging reinforcement learning: results from the diabetes and mental health adaptive notification tracking and evaluation (DIAMANTE) randomized clinical trial. J Med Internet Res 2024 Oct 8;26:e60834. [doi: 10.2196/60834] [Medline: 39378080]
102. Alonso-Domínguez R, García-Ortiz L, Patino-Alonso MC, Sánchez-Aguadero N, Gómez-Marcos MA, Recio-Rodríguez JI. Effectiveness of a multifactorial intervention in increasing adherence to the Mediterranean diet among patients with diabetes mellitus type 2: a controlled and randomized study (EMID study). Nutrients 2019 Jan 14;11(1):162. [doi: 10.3390/nu11010162] [Medline: 30646500]
103. Zamanillo-Campos R, Fiol-Deroque MA, Serrano-Ripoll MJ, et al. Effectiveness of DiabeText, a mHealth intervention to support diabetes self-management: randomized controlled trial in primary care. medRxiv. Preprint posted online on Feb 29, 2024. [doi: 10.1101/2024.02.28.24303489]
104. Sáenz A, Brito M, Morón I, Torralba A, García-Sanz E, Redondo J. Development and validation of a computer application to aid the physician's decision-making process at the start of and during treatment with insulin in type 2 diabetes: a randomized and controlled trial. J Diabetes Sci Technol 2012 May 1;6(3):581-588. [doi: 10.1177/193229681200600313] [Medline: 22768889]
105. Gunawardena KC, Jackson R, Robinett I, et al. The influence of the Smart Glucose Manager mobile application on diabetes management. J Diabetes Sci Technol 2019 Jan;13(1):75-81. [doi: 10.1177/1932296818804522] [Medline: 30264583]
106. Bonn SE, Hummel M, Peveri G, et al. Effectiveness of a smartphone app to promote physical activity among persons with type 2 diabetes: randomized controlled trial. Interact J Med Res 2024 Mar 21;13:e53054. [doi: 10.2196/53054] [Medline: 38512333]
107. Höchsmann C, Infanger D, Klenk C, Königstein K, Walz SP, Schmidt-Trucksäss A. Effectiveness of a behavior change technique-based smartphone game to improve intrinsic motivation and physical activity adherence in patients with type 2 diabetes: randomized controlled trial. JMIR Serious Games 2019;7(1):e11444. [doi: 10.2196/11444] [Medline: 30758293]
108. Höchsmann C, Müller O, Ambühl M, et al. Novel smartphone game improves physical activity behavior in type 2 diabetes. Am J Prev Med 2019;57(1):41-50. [doi: 10.1016/j.amepre.2019.02.017] [Medline: 31128953]
109. Chiaranai C, Chularee S, Prawatwong W, Srithongluang S. Two-way SMS reminders for medication adherence and quality of life in adults with type 2 diabetes: a randomized controlled trial. PRIJNR 2023;27(3):457-471. [doi: 10.60099/prijnr.2023.262244]
110. Poonprapai P, Lerkiatbundit S, Saengcharoen W. Family support-based intervention using a mobile application provided by pharmacists for older adults with diabetes to improve glycaemic control: a randomised controlled trial. Int J Clin Pharm 2022 Jun;44(3):680-688. [doi: 10.1007/s11096-022-01389-5] [Medline: 35247147]
111. Dincer B, Bahçecik N. The effect of a mobile application on the foot care of individuals with type 2 diabetes: a randomised controlled study. Health Educ J 2021;80(4):425-437. [doi: 10.1177/0017896920981617]
112. Hayıroğlu M, Çinier G, Yüksel G, et al. Effect of a mobile application and smart devices on heart rate variability in diabetic patients with high cardiovascular risk: a sub-study of the LIGHT randomized clinical trial. Kardiol Pol 2021;79(11):1239-1244. [doi: 10.33963/KP.a2021.0112] [Medline: 34599495]
113. Murray E, Sweeting M, Dack C, et al. Web-based self-management support for people with type 2 diabetes (HeLP-Diabetes): randomised controlled trial in English primary care. BMJ Open 2017 Sep 27;7(9):e016009. [doi: 10.1136/bmjopen-2017-016009] [Medline: 28954789]

XSL•FO
RenderX

114. Unsworth R, Armiger R, Jugnee N, et al. Safety and efficacy of an adaptive bolus calculator for type 1 diabetes: a randomized controlled crossover study. Diabetes Technol Ther 2023 Jun;25(6):414-425. [doi: 10.1089/dia.2022.0504] [Medline: 37017468]
115. Baron JS, Hirani S, Newman SP. A randomised, controlled trial of the effects of a mobile telehealth intervention on clinical and patient-reported outcomes in people with poorly controlled diabetes. J Telemed Telecare 2017 Feb;23(2):207-216. [doi: 10.1177/1357633X16631628] [Medline: 26880694]
116. Baron JS, Hirani SP, Newman SP. Investigating the behavioural effects of a mobile-phone based home telehealth intervention in people with insulin-requiring diabetes: results of a randomized controlled trial with patient interviews. J Telemed Telecare 2017 Jun;23(5):503-512. [doi: 10.1177/1357633X16655911] [Medline: 27377790]
117. Heald A, Roberts S, Albelda Gimeno L, et al. Enhancing type 2 diabetes treatment through digital plans of care - a randomized controlled trial: evaluation of change in patient reported outcome measures. Expert Rev Endocrinol Metab 2024 Jul;19(4):385-391. [doi: 10.1080/17446651.2024.2334220] [Medline: 38567714]
118. Sachmechi I, Salam S, Amini M, Khan R, Spitznogle A, Belen T. Frequent monitoring of blood glucose levels via a remote patient monitoring system helps improve glycemic control. Endocr Pract 2023;29(6):441-447. [doi: 10.1016/j.eprac.2023.03.270] [Medline: 36965657]
119. Kerfoot BP, Gagnon DR, McMahon GT, Orlander JD, Kurgansky KE, Conlin PR. A team-based online game improves blood glucose control in veterans with type 2 diabetes: a randomized controlled trial. Diabetes Care 2017;40(9):1218-1225. [doi: 10.2337/dc17-0310] [Medline: 28790131]
120. Young HM, Miyamoto S, Dharmar M, Tang-Feldman Y. Nurse coaching and mobile health compared with usual care to improve diabetes self-efficacy for persons with type 2 diabetes: randomized controlled trial. JMIR Mhealth Uhealth 2020 Mar 2;8(3):e16665. [doi: 10.2196/16665] [Medline: 32130184]
121. Mitchell SE, Bragg A, De La Cruz BA, et al. Effectiveness of an immersive telemedicine platform for delivering diabetes medical group visits for African American, Black and Hispanic, or Latina women with uncontrolled diabetes: the women in control 2.0 noninferiority randomized clinical trial. J Med Internet Res 2023 May 10;25:e43669. [doi: 10.2196/43669] [Medline: 37163341]
122. Spierling Bagsic SR, Fortmann AL, San Diego ER, et al. Outcomes of the Dulce Digital-COVID Aware (DD-CA) discharge texting platform for US/Mexico border Hispanic individuals with diabetes. Diabetes Res Clin Pract 2024 Apr;210:111614. [doi: 10.1016/j.diabres.2024.111614] [Medline: 38484985]
123. Hsia J, Guthrie NL, Lupinacci P, et al. Randomized, controlled trial of a digital behavioral therapeutic application to improve glycemic control in adults with type 2 diabetes. Diabetes Care 2022 Dec 1;45(12):2976-2981. [doi: 10.2337/dc22-1099] [Medline: 36181554]
124. Katz LB, Aparicio M, Cameron H, Ceppa F. Use of a meter with color-range indicators and a mobile diabetes management app improved glycemic control and patient satisfaction in an underserved Hispanic population: “Tu Salud”-a randomized controlled partial cross-over clinical study. Diabetes Spectr 2022 Feb 15;35(1):86-94. [doi: 10.2337/ds20-0101] [Medline: 35308153]
125. Lee JL, Foschini L, Kumar S, et al. Digital intervention increases influenza vaccination rates for people with diabetes in a decentralized randomized trial. NPJ Digit Med 2021 Sep 17;4(1):138. [doi: 10.1038/s41746-021-00508-2] [Medline: 34535755]
126. Quinn CC, Butler EC, Swasey KK, et al. Mobile diabetes intervention study of patient engagement and impact on blood glucose: mixed methods analysis. JMIR Mhealth Uhealth 2018 Feb 2;6(2):e31. [doi: 10.2196/mhealth.9265] [Medline: 29396389]
127. Mahoney E, Glezer S, Baccari L, Lebowitz J, Yue W, Klonoff DC. Use of a diabetes self-management application in combination with a 4 mm pen needle and its impact on glycemic variability and patient-reported outcomes in people with type 2 diabetes using basal-bolus insulin therapy. Clin Diabetol 2022;11(3):156-164. [doi: 10.5603/DK.a2022.0012]
128. Ramallo-Fariña Y, García-Bello MA, García-Pérez L, et al. Effectiveness of internet-based multicomponent interventions for patients and health care professionals to improve clinical outcomes in type 2 diabetes evaluated through the INDICA study: multiarm cluster randomized controlled trial. JMIR Mhealth Uhealth 2020 Nov 2;8(11):e18922. [doi: 10.2196/18922] [Medline: 33136059]
129. Gerber BS, Biggers A, Tilton JJ, et al. Mobile health intervention in patients with type 2 diabetes: a randomized clinical trial. JAMA Netw Open 2023 Sep 5;6(9):e2333629. [doi: 10.1001/jamanetworkopen.2023.33629] [Medline: 37773498]
130. Lorig K, Ritter PL, Laurent DD, et al. Online diabetes self-management program: a randomized study. Diabetes Care 2010 Jun;33(6):1275-1281. [doi: 10.2337/dc09-2153] [Medline: 20299481]
131. Kumar S, Moseson H, Uppal J, Juusola JL. A diabetes mobile app with in-app coaching from a certified diabetes educator reduces A1C for individuals with type 2 diabetes. Diabetes Educ 2018;44(3):226-236. [doi: 10.1177/0145721718765650] [Medline: 29575982]
132. Heisler M, Choi H, Mase R, Long JA, Reeves PJ. Effectiveness of technologically enhanced peer support in improving glycemic management among predominantly African American, low-income adults with diabetes. Diabetes Educ 2019;45(3):260-271. [doi: 10.1177/0145721719844547] [Medline: 31027477]

XSL•FO
RenderX

133. Adepoju OE, Bolin JN, Phillips CD, et al. Effects of diabetes self-management programs on time-to-hospitalization among patients with type 2 diabetes: a survival analysis model. Patient Educ Couns 2014 Apr;95(1):111-117. [doi: 10.1016/j.pec.2014.01.001] [Medline: 24468198]
134. Levy N, Moynihan V, Nilo A, et al. The Mobile Insulin Titration Intervention (MITI) for insulin adjustment in an urban, low-income population: randomized controlled trial. J Med Internet Res 2015 Jul 17;17(7):e180. [doi: 10.2196/jmir.4716] [Medline: 26187303]
135. Felker GM, Sharma A, Mentz RJ, et al. A randomized controlled trial of mobile health intervention in patients with heart failure and diabetes. J Card Fail 2022;28(11):1575-1583. [doi: 10.1016/j.cardfail.2022.07.048] [Medline: 35882260]
136. Bell AM, Fonda SJ, Walker MS, Schmidt V, Vigersky RA. Mobile phone-based video messages for diabetes self-care support. J Diabetes Sci Technol 2012 Mar 1;6(2):310-319. [doi: 10.1177/193229681200600214] [Medline: 22538140]
137. Joshi R, Joshi D, Cheriyath P. Improving adherence and outcomes in diabetic patients. Patient Prefer Adherence 2017;11(11):271-275. [doi: 10.2147/PPA.S122490] [Medline: 28243070]
138. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
139. Alòs F, Colomer M, Bort-Roig J, et al. Impact of a primary care-based mobile health intervention to “sit less and move more” on HbA1c, blood pressure, and other clinical outcomes in office employees with type 2 diabetes: a randomized controlled trial. Prim Care Diabetes 2025;19(5):434-445. [doi: 10.1016/j.pcd.2025.05.010] [Medline: 40494751]
140. Asante E, Carter G, McAneney H, Bam V, Sarfo-Kantanka O, Prue G. Nurse-led mobile phone intervention to promote self-management in type 2 diabetes in Ghana: a randomized controlled trial. Sci Diabetes Self Manag Care 2025;51(1):85-99. [doi: 10.1177/26350106241293113] [Medline: 39588726]
141. Bretschneider MP, Kolasińska AB, Šomvárska L, Klásek J, Mareš J, Schwarz PE. Evaluation of the impact of mobile health app Vitadio in patients with type 2 diabetes: randomized controlled trial. J Med Internet Res 2025 May 9;27:e68648. [doi: 10.2196/68648] [Medline: 40344662]
142. Chen J, Lv L, Zhao X, et al. The effectiveness of a community-based online low-glycaemic index diet and lifestyle recommendations intervention for people with type 2 diabetes: a randomized controlled trial. Arch Public Health 2025 Mar 6;83(1):61. [doi: 10.1186/s13690-025-01552-0] [Medline: 40045437]
143. Cui J, Liu Q, Huang L, Yu H. Digital outdoor exercise program for obese patients with type 2 diabetes mellitus: a non-inferiority randomized controlled trial. Front Endocrinol 2025;16:1654129. [doi: 10.3389/fendo.2025.1654129]
144. Hangaard S, Kronborg T, Cohen SR, et al. Effectiveness and safety of telemonitoring compared with standard of care in people with type 2 diabetes treated with insulin: a national multicenter randomized controlled trial. Eur J Intern Med 2025 Jul;137:112-118. [doi: 10.1016/j.ejim.2025.05.017] [Medline: 40414784]
145. Nguyen PV, Thach PT, Tran HK, et al. Effectiveness of a digital diabetes self-management education and support program in Vietnamese adults with type 2 diabetes. Sci Rep 2025 Nov 21;15(1):41369. [doi: 10.1038/s41598-025-25245-4] [Medline: 41271901]
146. Ólafsdóttir AF, Sveen KA, Wijkman M, et al. Systematic intensive therapy in addition to continuous glucose monitoring in adults with type 1 diabetes: a multicentre, open-label, randomised controlled trial. Lancet Reg Health Eur 2025 Dec;59:101485. [doi: 10.1016/j.lanepe.2025.101485] [Medline: 41142656]
147. Philis-Tsimikas A, Fortmann AL, Clark T, et al. Dulce Digital-Me: results of a randomized comparative trial of static versus adaptive digital interventions for Latine adults with diabetes. Ann Behav Med 2025 Jan 4;59(1):kaae077. [doi: 10.1093/abm/kaae077] [Medline: 39707158]
148. Rodríguez Montes OE, Gogeascoechea Trejo MC, Bermúdez-Tamayo C. Efficacy of a mobile application for glycemic control in type 2 diabetes: a randomized clinical trial. Diabetes Epidemiol Manag 2025;19-20:100279. [doi: 10.1016/j.deman.2025.100279]
149. Shabani Bazneshin F, Mohammadi Zeidi I, Morshedi H. The effectiveness of theory based educational intervention on health literacy, medication adherence and self-manag. J Prev Med Hyg 2025 Jun;66(2):E235-E247. [doi: 10.15167/2421-4248/jpmh2025.66.2.3149] [Medline: 41235049]
150. Wongdama S, Paemueang W, Sriphrapradang C. A tailor-made mobile app with a local cuisine database for self-management of type 2 diabetes mellitus: randomized controlled trial. JMIR Diabetes 2025 Dec 29;10:e83685. [doi: 10.2196/83685] [Medline: 41461045]
151. Wu Y, Min H, Li M, et al. Effect of Artificial Intelligence-based Health Education Accurately Linking System (AI-HEALS) for type 2 diabetes self-management: protocol for a mixed-methods study. BMC Public Health 2023;23(1):1325. [doi: 10.1186/s12889-023-16066-z] [Medline: 37434126]
152. Zamani S, Abdoli S, Ghadirian F, Mardanian Dehkordi L. The impact of digital storytelling on psychosocial well-being in individuals with type 1 diabetes: a randomized clinical trial study in Iran. BMC Psychol 2025 Aug 19;13(1):942. [doi: 10.1186/s40359-025-03258-y] [Medline: 40830533]
153. Zhang PZ, Guo D, Liu CQ, et al. Mobile app-based intervention and cardiovascular risk factors in patients with uncontrolled type 2 diabetes: a randomized clinical trial. JAMA Netw Open 2025 Sep 2;8(9):e2529762. [doi: 10.1001/jamanetworkopen.2025.29762] [Medline: 40892411]

XSL•FO
RenderX

154. Balbinot GS, Costódio RS, Vicentini GE. Impact of a remote nutrition education on low-carbohydrate diet based in type 2 diabetes management: findings from a Brazilian primary care randomized controlled trial. Diabetol Int 2026 Jul;17(3):42. [doi: 10.1007/s13340-026-00898-2] [Medline: 42040054]
155. Crowley MJ, Lewinski AA, Yang Q, et al. Expanding technology-enabled, nurse-delivered chronic disease care: a pragmatic, randomized, effectiveness-implementation trial. Ann Intern Med 2026;179(7):968-979. [doi: 10.7326/ANNALS-26-00132] [Medline: 42330500]
156. Huda RK, Singh J, Kumar P, et al. Effectiveness of mHealth interventions for improving diabetes management outcomes in patients with type 2 diabetes: a cluster randomized crossover trial in Western Rajasthan. Clin Epidemiol Glob Health 2026 Jan;37:102264. [doi: 10.1016/j.cegh.2025.102264]
157. Jang MK, Cho YK, Moon JY, Min SH, Hwang JH, Jung CH. Impact of smart watch mobile application on the risk treatment of type 2 diabetes mellitus (iSMART-DM). Prim Care Diabetes 2026 Feb;20(1):47-52. [doi: 10.1016/j.pcd.2025.10.008] [Medline: 41145321]
158. Lu CH, Chu CH, Ou HY, et al. Effectiveness and safety of connected solutions among people with type 2 diabetes uncontrolled on oral anti-diabetic drugs: a randomized real-world study on Insultrate. Prim Care Diabetes 2026 Aug;20(4):411-417. [doi: 10.1016/j.pcd.2026.04.012] [Medline: 42067440]
159. Sebastian-Valles F, Sager-La Ganga C, Justel Enriquez A, et al. Analysis of glycaemic control with a connected smart pen cap in adults with type 1 diabetes: a randomised, open-label, parallel-group trial. Diabetologia 2026 May;69(5):1191-1204. [doi: 10.1007/s00125-026-06674-w]
160. Wang M, Lee K, Thew HZ, et al. Effectiveness of the mobile-based Diabetes Little Helper video intervention on medication adherence among older adults living with type 2 diabetes mellitus, Henan, China: randomized controlled trial. J Med Internet Res 2026 Feb 17;28:e78731. [doi: 10.2196/78731] [Medline: 41701967]
161. Xiao J, Wang Q, Tan S, et al. Effect of a digital-driven physician-pharmacist collaborative model for diabetes in primary health care: cluster randomized trial. J Med Internet Res 2026 Mar 13;28:e77470. [doi: 10.2196/77470] [Medline: 41824949]
162. Zhou L, Wu C, Wang C, et al. Smart insulin dosage software based on trend arrows: an effective method for controlling hyperglycemia in hospitalized diabetic patients in non-endocrine departments (a randomized controlled trial). Endocrine 2026 Feb 16;91(1):86. [doi: 10.1007/s12020-025-04509-z] [Medline: 41697460]
163. Saetang T, Greeviroj P, Thavaraputta S, Santisitthanon P, Houngngam N, Laichuthai N. The effectiveness of telemonitoring and integrated personalized diabetes management in people with insulin-treated type 2 diabetes mellitus. Diabetes Obes Metab 2024 Nov;26(11):5233-5238. [doi: 10.1111/dom.15870] [Medline: 39161074]
164. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Privacy and security in mobile health apps: a review and recommendations. J Med Syst 2015 Jan;39(1):181. [doi: 10.1007/s10916-014-0181-3] [Medline: 25486895]
165. Syed R, Eden R, Makasi T, et al. Digital health data quality issues: systematic review. J Med Internet Res 2023 Mar 31;25:e42615. [doi: 10.2196/42615] [Medline: 37000497]
166. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. J Am Med Inform Assoc 2016 Sep;23(5):899-908. [doi: 10.1093/jamia/ocv189] [Medline: 26911829]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
HIC: high-income country
HIPAA: Health Insurance Portability and Accountability Act
LMIC: low- or middle-income country
mERA: mHealth evidence reporting and assessment
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

Edited by A Stone; submitted 22.Mar.2026; peer-reviewed by H Ranjani, HL Tam; revised version received 27.Aug.2026; accepted 27.Aug.2026; published 01.Oct.2026.

Please cite as:
Yue X, Shen J, Peng J, Huang X, Hu M, Hu X
Reporting Gaps in mHealth Intervention Studies for Adults With Diabetes: Systematic Review
JMIR Mhealth Uhealth 2026;14:e95863
URL: https://mhealth.jmir.org/2026/1/e95863
doi:10.2196/95863

XSL•FO
RenderX

© Xin Yue, Xiaoting Huang, Xing Hu, Mingyue Hu, Jiacheng Shen, Junke Peng. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 1.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Stakeholder Perspectives on Digital Health Technologies in Poststroke Self-Management and Rehabilitation: Systematic Review Based on the NASSS Framework

Jing Zhao[1], MM; Jie Zhao[1], MM; Xiaohan Liu[1], BSc; Yongxin Cao[1], BSc; Xuan Wang[1], BSc; Yanhong Dong[2], PhD; Hong Chang[1], MM

[1]Department of Neurology, Xuanwu Hospital, Capital Medical University, No. 45 Changchun Street, Beijing, China

[2]Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

Corresponding Author:
Hong Chang, MM
Department of Neurology, Xuanwu Hospital, Capital Medical University, No. 45 Changchun Street, Beijing, China

Abstract

Background: Digital health technologies (DHTs) offer opportunities to support poststroke self-management and rehabilitation, but their implementation may be hindered by mismatches between technological design, stakeholder needs, and care contexts. Synthesizing stakeholder perspectives can clarify implementation priorities and gaps in the evidence.

Objective: This study aimed to synthesize stakeholders' perceptions, experiences, and recommendations regarding DHT implementation in poststroke self-management and rehabilitation and to map the findings to the NASSS (Nonadoption, Abandonment, Scale-Up, Spread, and Sustainability) framework.

Methods: Six databases (PubMed, MEDLINE, Embase, Scopus, CINAHL, and Web of Science) were searched from inception through October 2025. Eligible studies were English-language, peer-reviewed primary studies using qualitative, quantitative, or mixed methods designs to report stakeholder perspectives on DHTs in poststroke care. Two reviewers independently screened studies, extracted data, and assessed methodological quality using the Mixed Methods Appraisal Tool (MMAT). A convergent integrated synthesis was conducted: quantitative findings were transformed into textual descriptors, combined with qualitative findings, coded inductively, and subsequently mapped deductively to NASSS domains and subdomains. Disagreements were resolved through consensus or by third-reviewer adjudication.

Results: Of 15,262 records identified, 42 studies published between 2019 and 2025 were included: 24 qualitative, 8 quantitative, and 10 mixed methods studies. Patients, health care professionals, caregivers, and members of the public were represented in 32, 19, 8, and 3 studies, respectively, with some studies including multiple groups. Evidence concentrated on the condition, technology, value proposition, and adopter-system domains. Stakeholders valued personalized functions, intuitive interfaces, timely feedback, and digital support that complemented human care. Barriers included stroke-related motor, cognitive, communication, and sensory impairments; unequal digital access and literacy; caregiver demands; staff workload and training needs; and poor integration with clinical routines. Data privacy and unclear payment arrangements also raised concerns. Evidence on organizational readiness and wider-system determinants was limited, and no eligible study directly examined interaction and adaptation over time. Methodological concerns included nonresponse bias and insufficient interpretation of integrated findings in mixed methods studies.

Conclusions: Stakeholder perspectives indicate that DHT implementation in poststroke care should combine accessible, personalized design with support for patients, caregivers, and health care professionals, alongside integration into clinical workflows and clear data governance and financing arrangements. The NASSS mapping identified priorities beyond usability, but current evidence does not establish sustained adoption, scale-up, or spread. Longitudinal studies in routine care are needed to evaluate adaptation and sustainability.

(JMIR Mhealth Uhealth 2026;14:e95197) doi:10.2196/95197

KEYWORDS

stroke; digital health technologies; stroke care; systematic review; self-management; adoption

XSL•FO
RenderX

Introduction

Background

Stroke remains a major public health challenge worldwide and a leading contributor to mortality and disability [1]. According to the 2021 Global Burden of Disease (GBD) study, there were 93.8 million prevalent cases of stroke, 11.9 million new stroke events, and 7.3 million deaths from stroke [2]. Without strengthened prevention, treatment, and long-term care strategies, the global stroke burden is projected to continue increasing substantially by 2050 [3]. Nearly half of patients who are poststroke experience long-term disabilities that compromise independence and quality of life, resulting in ongoing needs for rehabilitation, self-management support, and long-term care [4].

The rapid development and adoption of digital health technologies (DHTs) have created unprecedented opportunities for supporting chronic disease management and home-based rehabilitation [5]. DHTs refer to technology-enabled tools that support health information delivery, monitoring, and care management, such as patient portals, mobile apps, wearable devices, and telemedicine platforms [6]. By making care more accessible and flexible and potentially reducing costs, DHTs could extend stroke support beyond traditional clinical settings and facilitate clinical decision-making, personalized care, and service delivery for populations with diverse needs, including those in low-resource settings [7].

DHTs are increasingly used to support poststroke monitoring, self-management, rehabilitation, and remote consultations [8]. Growing evidence suggests that digital health interventions can improve several clinically relevant outcomes among patients who are poststroke. For example, digital health–enabled stroke management has been associated with improved blood pressure control and reduced stroke recurrence [9]. In addition, meta-analytic evidence indicates potential benefits for poststroke cognitive functioning [10]. Home-based mobile health interventions have also shown potential to support risk-factor control and continuity of poststroke care [11]. By offering alternative and personalized interventions, DHTs can help overcome barriers to care and mitigate workforce shortages. DHTs may offer opportunities for scale-up, although successful dissemination and maintenance of intervention fidelity depend on the implementation context and sustained organizational support [12].

DHTs can deliver their potential benefits only when they are successfully implemented in clinical practice [7,13]. Real-world implementation requires not only evidence of clinical effectiveness but also a clear understanding of how DHTs are perceived, adopted, and integrated by relevant stakeholders [7]. Stakeholder perspectives from patients, caregivers, health care professionals (HCPs), organizations, funders, and policymakers are therefore essential for identifying the practical needs, contextual barriers, and system-level conditions that shape the implementation of DHTs in stroke care [14]. The NASSS (Nonadoption, Abandonment, Scale-Up, Spread, and Sustainability) framework was developed to support the evaluation and implementation of technology-supported health care programs [15]. Its seven domains cover the condition, technology, value proposition, adopters, organization, wider system, and interaction and adaptation over time, providing a structured lens for analyzing the implementation complexity of DHTs in stroke care [15,16].

Objectives

An increasing number of primary studies have examined the use of DHTs to support self-management and rehabilitation in patients who are poststroke [17-20]. However, few systematic reviews have synthesized this evidence through a structured implementation framework such as the NASSS framework. Applying the NASSS framework can help identify implementation determinants across diverse clinical and care settings and support more comparable, theory-informed implementation research [13]. The aims of this mixed methods systematic review were (1) to explore the perceptions, experiences, and recommendations of key stakeholders regarding the implementation of DHT in poststroke self-management and rehabilitation and (2) to map these findings onto the NASSS framework. This review provides implementation-oriented evidence to guide the future integration and scaling of DHT interventions in stroke care.

Methods

Search Strategy

We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [21]. The review protocol was registered in PROSPERO (International Prospective Register of Systematic Reviews; CRD420251112158). We searched 6 databases (PubMed, MEDLINE, Embase, Scopus, CINAHL, and Web of Science) from inception through October 2025. Reviews were excluded, but their reference lists and those of eligible studies were checked for additional records. Controlled vocabulary and free-text terms were tailored to each database. The full search strategies for all databases are available in Multimedia Appendix 1.

Eligibility Criteria

DHTs were defined as technology-based interventions that use data collection, automated analysis, feedback, or remote communication to support rehabilitation or disease management [8]. Studies were eligible if they (1) investigated a DHT, including therapy or education app, digital rehabilitation services, automated reminders, self-administered online services, or objective monitoring tools; (2) involved patients with stroke or stakeholders relevant to their care, including patients, family caregivers, family members, and HCPs; (3) reported qualitative or quantitative evidence on stakeholder perspectives, experiences, recommendations, or implementation challenges; and (4) were published in English. The intended end users had to include patients with stroke. We excluded technologies used only to deliver therapist-led exercises without data storage, feedback, or interactive functionality and communication-only services (eg, email, chat, or videoconferencing) without an autonomous or scalable digital component.

XSL•FO
RenderX

Study Selection

After deduplication in EndNote X9 (version 9.3.3; Clarivate Analytics), 2 reviewers independently screened the titles and abstracts of retrieved articles from the search strategy that potentially adhered to the study eligibility criteria. Full texts were then reviewed by the 2 reviewers for inclusion in the review. Disagreements were resolved through discussion and, when consensus could not be reached, through adjudication by a third reviewer.

Quality Appraisal

Two reviewers independently appraised the included studies using the 2018 Mixed Methods Appraisal Tool (MMAT) [22]. No study was excluded on the basis of the MMAT appraisal, and no overall numerical quality score or statistical weighting was applied. Criterion-level methodological concerns were considered when interpreting the breadth and strength of the NASSS-mapped findings.

Data Extraction

A shared Microsoft Excel extraction form, aligned with Joanna Briggs Institute guidance, was piloted on a subset of studies and refined before full extraction [23]. Two reviewers independently extracted publication details, country, design, data collection and analysis methods, participant characteristics, setting, DHT types, implementation context, and stakeholder findings. Discrepancies and unclear reporting were discussed by 3 authors, and the agreed decisions were recorded in the final extraction file.

Data Synthesis and Integration

We used a convergent integrated synthesis to analyze qualitative, quantitative, and mixed methods findings together [24]. Quantitative results were transformed into textual descriptors that retained the direction, magnitude, population, and context of the original result; qualitative findings and mixed methods results were extracted as author interpretations supported, where available, by participant quotations. The analysis proceeded in 3 stages. First, 2 reviewers independently applied open codes to the integrated findings without using the NASSS framework as an a priori coding structure. Second, related codes were compared and refined inductively into clinically meaningful categories through team discussion. Third, these categories were mapped deductively to the 7 NASSS domains and their subdomains using MAXQDA 2020 (VERBI Software) [25]. The NASSS framework was selected because it provides a common structure for comparing interacting implementation determinants across heterogeneous technologies, stakeholders, and care contexts [16,26,27]. As NASSS was originally developed primarily for evaluating individual implementation cases, we used it as an organizing framework. A category could be mapped to more than 1 domain when it represented distinct mechanisms. Findings that did not fit an existing subdomain were retained and used to develop 2 adapted technology subdomains. Disagreements were resolved through consensus or adjudication by a third reviewer.

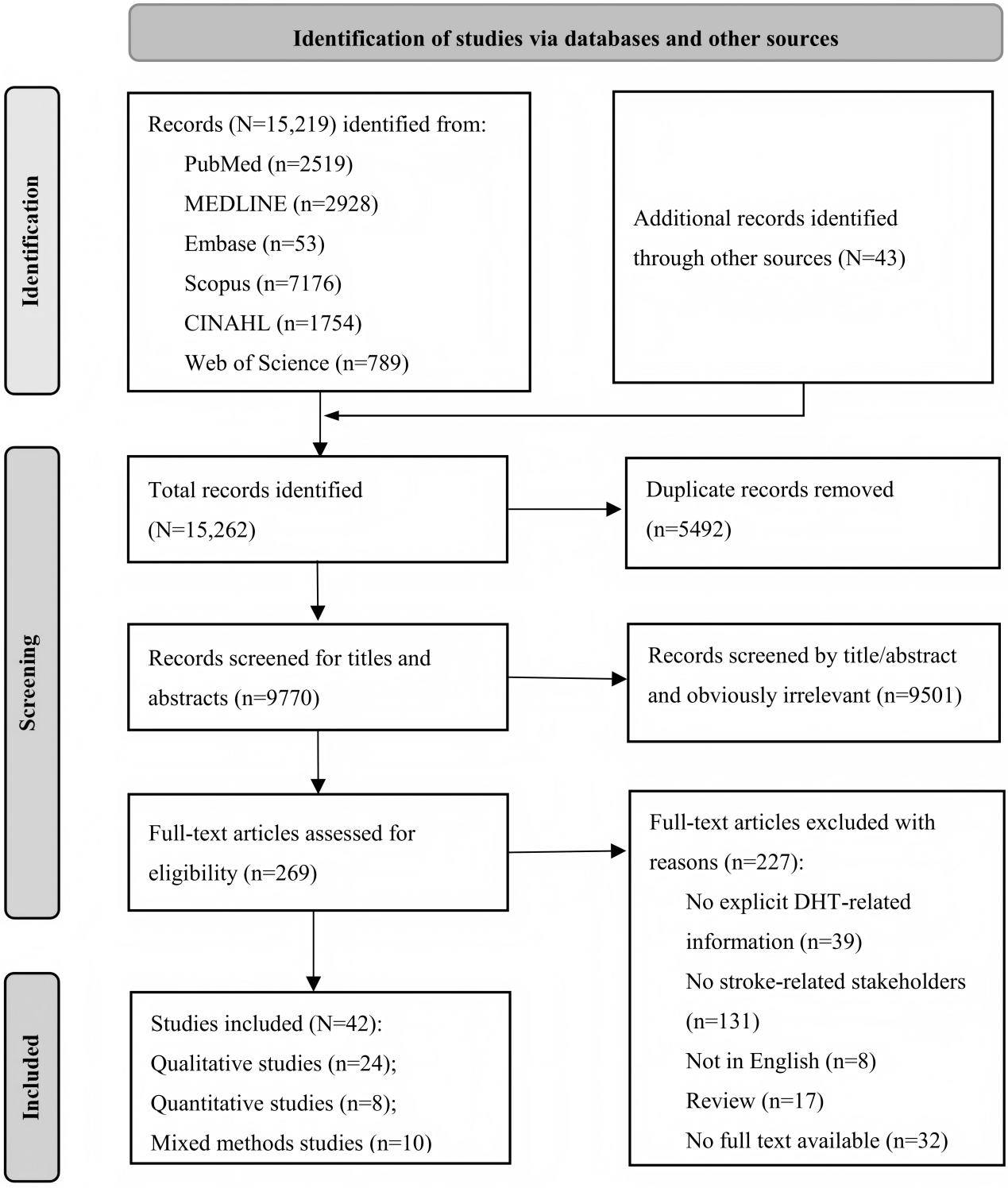
Results

Literature Search Results

The searches identified 15,262 records. After 5492 duplicates were removed, 9770 records underwent title and abstract screening, and 269 articles were assessed in full text. Of these, 227 were excluded and 42 studies were included (Figure 1): 24 qualitative studies [17,18,20,28-48], 8 quantitative studies [19,49-55], and 10 mixed methods studies [56-65]. Devittori et al [55,66] published related studies in 2024 and 2025. Because the latter extended the former, only the 2024 study was included in the synthesis [55]. One included study used NASSS in its primary analysis [31]. Four stakeholder groups were represented: patients with stroke, caregivers, HCPs (including physicians, nurses, occupational therapists, and physiotherapists), and members of the public.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the systematic review. DHT: digital health technologies.



Study Characteristics

Detailed characteristics of the 42 included studies are presented in Multimedia Appendix 2. The included studies were published between 2019 and 2025, with the number of publications increasing from 2022 onward. Studies were conducted in 21 countries: Australia (n=7), the United States (n=5), the United Kingdom (n=4), and China (n=3) were the most represented, while the remaining 23 studies were distributed across 17 countries. Settings included hospitals (n=14), rehabilitation centers (n=8), communities (n=4), homes (n=2), combined hospital-community settings (n=2), and multiple settings (eg, homes, hospitals, and communities; n=4). Two studies were conducted online, and 1 analyzed published user reviews of commercial stroke apps. The remaining 5 studies reported only geographical coverage without specifying the study setting. Six studies reported no funding, 18 received national research funding, 8 received institutional or association grants, 5 reported other funding, and 5 did not report funding.

XSL•FO
RenderX

Participant Characteristics

Of the 42 studies, 32 included patients, 19 included HCPs, 8 included caregivers, and 3 included members of the public. Because individual studies could include more than one stakeholder group, these stakeholder-study counts exceed 42. Sixteen of the 32 patient-related studies included only patients with stroke. By study design, the 24 qualitative studies included 441 participants, the 8 quantitative studies included 830 participants, and the 10 mixed methods studies included 1550 participants. Overall, studies reported 1197 men and 1319 women participants. Gender was not reported for 305 participants. Sample sizes ranged from 4 to 48 for qualitative studies, 13 to 486 for quantitative studies, and 19 to 1194 for mixed methods studies. Technologies included mobile apps, wearable devices, telemedicine platforms, virtual reality systems, rehabilitation robots, and AI-enabled tools used for remote monitoring, personalized rehabilitation, and disease management.

Methodological Quality of Included Studies

Quality appraisal results are detailed in Multimedia Appendix 3. Among the 24 qualitative studies, most showed coherence between their aims, data collection, analysis, and interpretation, although 2 did not clearly report data-collection procedures [29,44]. Among the 8 quantitative studies, measures and analyses were generally appropriate; 5 had a low risk of nonresponse bias [49-52,55], whereas 3 had a higher or unclear risk because response rates or differences between respondents and nonrespondents were insufficiently reported [19,53,54]. Only 4 of the 10 mixed methods studies adequately interpreted the outputs of qualitative-quantitative integration [56,58,60,63].

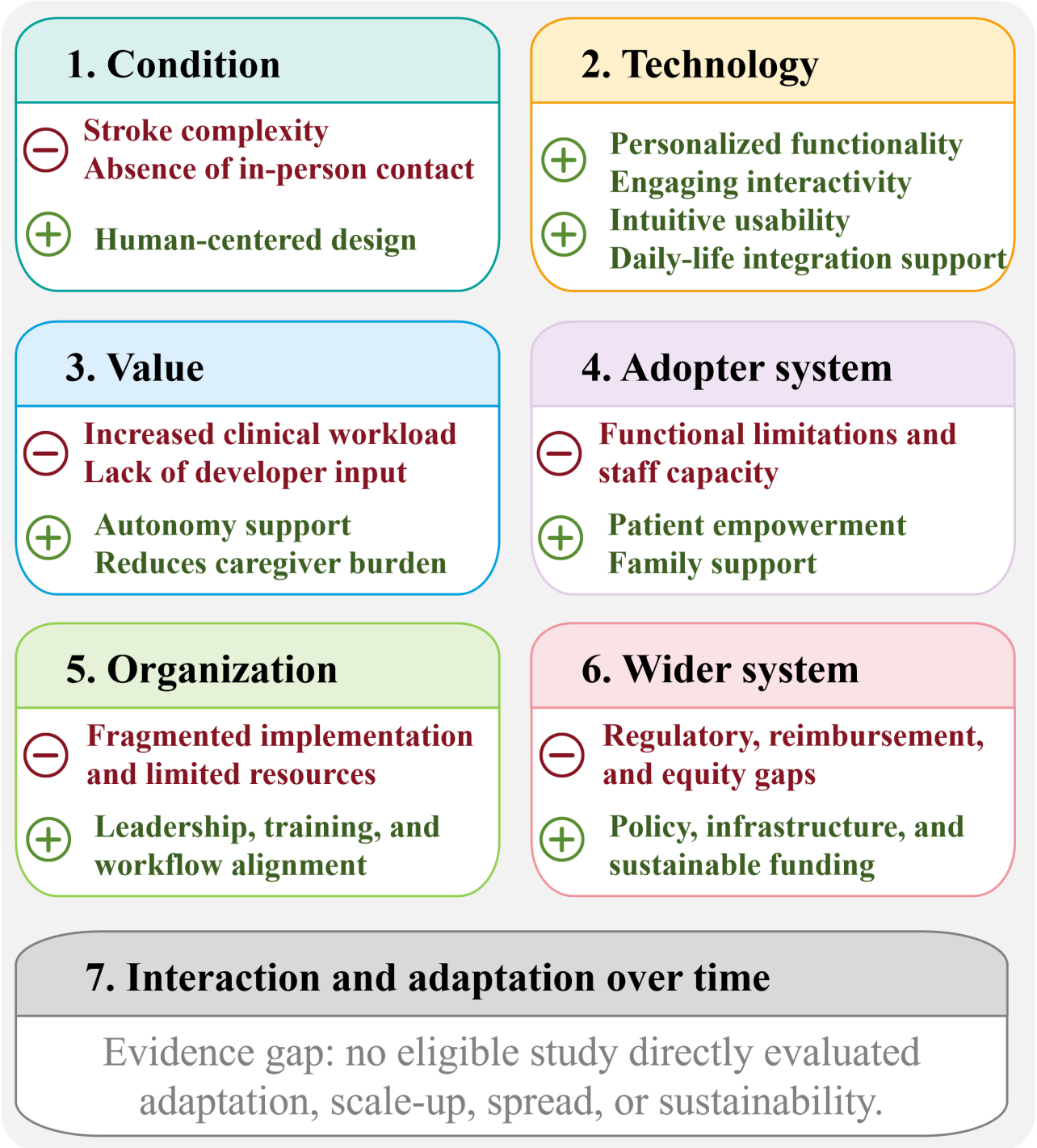
Themes

Overview

Inductive content analysis generated clinically meaningful categories, which were subsequently mapped to the NASSS domains and subdomains across the 42 studies. Evidence was concentrated in the condition, technology, value proposition, and adopter-system domains. No relevant evidence was identified for organizational capacity to innovate, the nature of organizational adoption or funding decisions, the professional context, the sociocultural context at the wider system level, or interaction and adaptation over time. Two adapted subdomains were added within the material features to distinguish personalization and accessibility from usability and simplicity. Figure 2 summarizes facilitators, barriers, and evidence gaps across the 7 domains. Figure 3 shows the study-level distribution of stakeholder evidence.

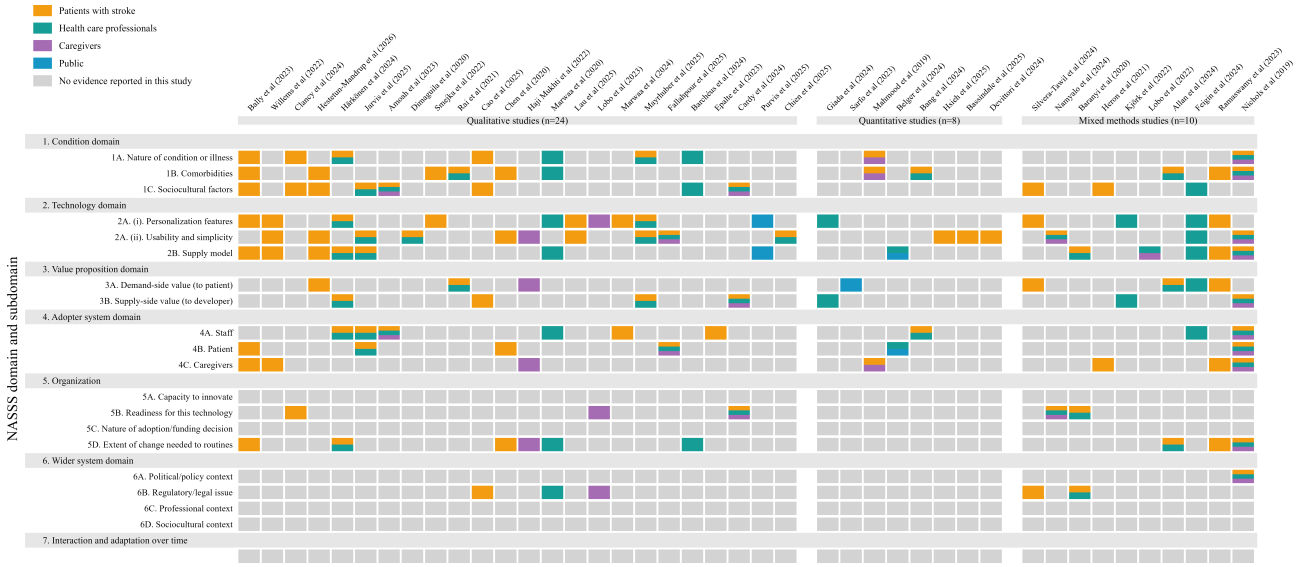
XSL•FO
RenderX

Figure 2. Synthesis of implementation determinants across the 7 NASSS (Nonadoption, Abandonment, Scale-Up, Spread, and Sustainability) domains. Green plus signs denote facilitators and red minus signs denote barriers reported by stakeholders. Domain 7 is shown in gray because no eligible study directly examined adaptation, scale-up, spread, or sustainability. The facilitators and barriers shown summarize key findings rather than providing an exhaustive list; detailed findings are presented in the Results section.



XSL•FO
RenderX

Figure 3. Matrix summarizing stakeholder perspectives reported across NASSS (Nonadoption, Abandonment, Scale-Up, Spread, and Sustainability) domains and subdomains. Columns represent 42 studies grouped by design, and rows represent NASSS domains and subdomains. Stacked colors indicate multiple stakeholder groups within the same study and subdomain [17-20,28-65].



Condition Domain

The first NASSS domain focuses on the illness-related context in which an intervention or technology is implemented. Across the included studies, stakeholders emphasized the context-dependent nature of DHTs, highlighting how the characteristics of stroke, together with the scope and setting of implementation, shaped technological requirements and constraints. Consideration of the stroke or clinical context was reported in 22 studies.

Nature of Condition or Illness

Across the included studies, perspectives on DHTs were predominantly patient-centered, reflecting the complex and heterogeneous nature of stroke [48]. Stroke-related impairments, including motor, cognitive, communication, emotional, visual, attentional, and fatigue-related difficulties, were frequently reported to constrain patients' capacity to engage with digital technologies and self-management activities [30,35]. Stakeholders viewed DHTs as supportive tools to facilitate poststroke self-management and rehabilitation rather than as standalone solutions [18]. The long-term and evolving recovery trajectory of stroke was identified as a challenge for the implementation of DHTs. While digital support was increasingly regarded as necessary to sustain rehabilitation and ongoing care beyond the acute phase [38,44], multiple studies emphasized that patients remained central to their recovery journey, with technology serving a complementary role [20]. Maintaining person-centered care, preserving human connection, and tailoring the balance between digital and face-to-face support to individual needs were consistently identified as key facilitators of effective implementation in poststroke contexts [42,51].

Comorbidities

Stroke commonly co-occurs with other chronic conditions [36]. Common comorbidities include cognitive, memory, language, and attention deficits, as well as psychomotor, fine motor, and sensory impairments [18,38]. Motivation to use DHTs was influenced by the perceived severity and relevance of the target problem, such as sleep disturbance [33]. Management of hypertension, diabetes, HIV/AIDS, and coronary heart disease was also considered important [34]. As noted by a participant, "Stroke is often a consequence of other chronic diseases such as hypertension, diabetes, HIV/AIDS, and coronary heart disease. Many with these underlying risk factors remain at risk if not well managed" [57]. Barriers therefore extended beyond clinical impairment to limited digital literacy, inadequate home infrastructure, and low motivation [29,34]. Many studies excluded people with moderate-to-severe cognitive impairment or aphasia, which limits the representativeness and generalizability of the evidence [19,20,52].

Sociocultural Factors

Stroke recovery demands longitudinal, human-centered technologies rather than one-off interventions [40]. Patients, caregivers, and clinicians consistently regarded in-person contact as important for safety, respect, and therapeutic relationships [18,20,35]. One patient noted: "It has to be personal by talking to each other in person and not via video call" [18]. Digital rehabilitation could not replace human interaction and care [29]. Acceptance was facilitated when DHTs complemented face-to-face care, incorporated culturally appropriate features, and involved family members [29,31]. Moreover, some HCPs have expressed concerns that DHTs could be mistaken for a substitute for clinical input, raising fears of professional displacement [44]. Acceptance of technology is further influenced by age, cultural norms such as familial caregiving expectations, and disability-related stigma [56]. Older adults required interfaces adapted to physical, cognitive, and social needs [35]. In younger adults, the reported underuse concerned stroke-specific digital tools rather than general digital literacy; limited age-targeted design and the perception that existing tools were intended for older populations may reduce relevance and engagement [17].

XSL•FO
RenderX

Technology Domain

The technology domain refers to the materials, data, knowledge, and supply-related characteristics of DHTs. Most included studies (30/42, 71.4%) primarily focused on proposed functions, technical features, and user interface design. Based on the findings, 2 key aspects related to material characteristics were identified. In contrast, issues concerning knowledge development and intellectual property were addressed far less frequently.

Personalization Features

Material features of DHTs were reported as facilitators. Personalization was a central material feature of DHTs. It addressed heterogeneity in patients' impairments, goals, and recovery trajectories [18,28], particularly through adaptable interface design and prioritization of core functions [64]. Features such as adjustable text size, zoom functions, and simplified layouts were perceived as enhancing accessibility for users with sensory or motor impairments [40,42]. Establishing an individualized information library enables autonomous retrieval of relevant information, covering topics like stroke causes, medication, prognosis, and follow-up care [18,49]. In addition, providing opportunities to explore advanced information through external resources and reliable project-specific websites is considered beneficial [47]. Patient-entered health data enabled tailored alerts and reminders [30,33,40,56]. Continuous customized support can maintain motivation, self-management, and empowerment [41]. Personalization also enhances user-friendliness and promotes long-term adherence to ambulatory monitoring [47,60,64]. The interactive features such as chatbots, scoring, rewards, animations, and motivational toolkits were also considered important to successful DHT design [38,39].

Usability and Simplicity

Across studies, simplicity and intuitive navigation were central to adoption. Patients preferred systems that required little cognitive effort and remained stable, free from advertisements and disruptive updates [18,28]. As an HCP noted, "Keep it as simple as possible; these are all terms that people, especially with cognitive impairments, find difficult to understand" [28]. Stable interfaces with clear navigation, free from advertisements and disruptive updates, were regarded as essential for usability and cognitive accessibility [38]. "I think it needs to be easy to use. Some people I find are surprisingly lazy, and if it's not easy to use the first time, they just won't go back and try again," a health expert noted [47]. Systems integrating medication reminders, activity-based goal setting, structured daily routines, and self-monitoring functions (eg, symptoms, sleep, and physical activity) within user-friendly diaries or checklists were perceived as highly relevant and supportive of daily use [30,31]. Incorporating stroke education and rehabilitation guidance boosted relevance and immersion [29]. These features increased the likelihood that a technology could be incorporated into everyday life rather than used only during a short study period [31].

Supply Model

Across studies, relevant and timely feedback was identified as a central facilitator of engagement with DHT [31,32]. DHTs commonly generated personalized behavioral reports [28], providing feedback on progress, adherence, and outcomes [19,54]. Patients particularly valued positive and corrective feedback to sustain motivation and to make slow recovery trajectories visible [42]. One patient noted, "I need more feedback. If I have doubts, I can use it as a reference" [37]. DHTs generated behavioral health reports based on individualized health data and on users' performance [55]. Patients received feedback after submitting their daily performance scores [43]. As a patient explained, "Tracking your recovery so that you can actually see or know you're recovering. It's such a slow process. You want something that gives you feedback each day on how you're going" [17]. Clinicians also relied on real-time or near-real-time feedback to monitor patient activity and inform follow-up, valuing automated updates that required minimal manual oversight. For example, staff using the VERA (Virtual Engagement Rehabilitation Assistant) system valued automatic updates: "as long as it gives me feedback, like how many times this patient exercised over the weekend" [31]. Feedback delivered through visual [39], auditory [55], interactive [54], and rapid real-time formats [19] was associated with improved user experience, with rapid feedback and embedded encouragement further supporting motivation and emotional resilience [19,36,54].

Value Proposition Domain

The value proposition domain concerns whether the technology is worth developing or introducing for patients, clinicians, and suppliers. Across stakeholder groups, DHTs were broadly perceived as adding value beyond usual care, particularly in enhancing efficiency, accuracy, and accessibility [49].

Demand-Side Value (to the Patient)

Patients reported that DHTs could improve education, monitoring, and access to follow-up support. Quantitative research has indicated that more than 80% of users are satisfied with the technical performance and perceived health benefits of DHTs [50,56]. Patients valued reminders for home exercises, timely health information, access to therapists, and the ability to monitor progress [63,64]. Nevertheless, trust in clinicians and health care organizations generally exceeded trust in digital systems, and infrequent phone use, limited flexibility, and low perceived treatment relevance reduced demand [56]. Caregivers anticipated reduced workload [37], whereas HCPs were concerned that poorly designed monitoring and messaging systems could generate anxiety and additional clinical work [34].

Supply-Side Value (to the Developer)

Developer-related perspectives were reported in 7 studies, but they mainly came from HCPs involved in co-design rather than from commercial technology developers. Systematically eliciting the "voice of stroke patients," prioritizing requirements, and providing reliable educational resources were considered essential [31,40]. From the demand-side perspective, patients similarly emphasized that developers ensure easy access to

XSL•FO
RenderX

reliable educational resources [35]. Automation was valued for providing consistent monitoring and prompts, but it did not remove the need for clinical oversight, validation, or accountability [29]. The absence of direct industry perspectives limited conclusions about intellectual property, commercial incentives, interoperability, and scalability.

Adopter System Domain

Clinicians, patients, and family caregivers were identified as key stakeholder groups whose roles are expected to evolve with the adoption of DHTs. The adopter domain focuses on these groups' capacity, motivation, and willingness to adopt and sustain the use of DHTs.

Staff

Within the adopter system, HCPs linked patient-generated data with clinical decision-making. Medical staff emphasized the importance of timely, transparent access to patient data and digital resources as a prerequisite for engagement [31]. DHTs enable clinicians to monitor patients' health status longitudinally, manage structured clinical records, and use decision-support tools to guide clinical care [30]. These capabilities supported phased goal setting, remote monitoring, and continuity of care during transitions to community-based rehabilitation teams [17]. However, barriers to adoption were also prominent. Many clinicians reported high workload, limited training and digital literacy, and insufficient organizational support [38]. Clinicians were also reluctant to recommend technologies that lacked public funding, institutional approval, or clear evidence of benefit [41].

Patient

Intuitive, adaptable, and intrinsically motivating DHTs facilitated patients' transition from passive care recipients to active self-managers and increased awareness of their health status and recovery progress [32]. Patients valued the autonomy to determine when and how to engage with DHTs, which influenced both adoption and sustained use [43]. DHTs also supported home-based recovery by promoting independence in daily activities and rehabilitation, thereby fostering a sense of control that was often less evident in clinic-centered care [38]. This emerging digital awareness reflected patients' increasing recognition of their poststroke capabilities and energy limitations in everyday life. However, physical, cognitive, and emotional consequences of stroke, together with fatigue and low confidence in using DHTs, constrained sustained engagement for some patients [20,45].

Caregivers

Family caregivers were important implementation partners who provided technical assistance, emotional encouragement, and structure for daily routines. Family and peer support could reduce isolation and reinforce engagement [28]. However, caregivers also faced limited digital literacy, distrust of technology, competing demands, and family tension [18]. Sustainable implementation therefore requires explicit assessment of caregiver capacity, training and support for those who choose to participate, and safeguards against transferring clinical or technical responsibilities to families without adequate resources [28,37].

Organization

This domain addresses organizational capacity, readiness, adoption decisions, and the changes required to integrate DHTs into routine stroke care. Evidence was reported in 10 qualitative and 4 mixed methods studies but was limited to readiness for this technology and the extent of change needed to routines; no relevant data were identified for capacity to innovate or for the nature of organizational adoption or funding decisions.

Capacity to Innovate

No eligible study directly examined an organization's systemic capacity to innovate, such as its ability to absorb repeated technological change, maintain innovation infrastructure, or learn across implementation cycles.

Readiness for This Technology

Organizational readiness was constrained by fragmented implementation processes, poor resource allocation, and limited integration with clinical information systems [37,38]. An analysis of user reviews found that 39 of 46 stroke apps had been developed by nonmedical or noneducational entities, often with limited clinical validation or alignment with health-system standards [40]. A patient noted, "With the stroke, once you actually leave hospital, that's it. That's all. I've had absolutely no follow-up, none whatsoever...They, they just toss you out the door, and that's it; you're on your own" [20]. Facilitators included leadership support, protected implementation time, clinical validation, staff training, and clearly assigned responsibility for monitoring and follow-up.

Nature of the Adoption or Funding Decision

No eligible study directly evaluated how organizations made formal adoption, procurement, or internal funding decisions. References to reimbursement and public funding were coded under the wider-system domain because they concerned policy-level arrangements rather than a documented organizational decision-making process.

Extent of Change Needed to Routines

Because stroke care is multidisciplinary and crosses organizational boundaries, DHT implementation may require substantial changes to routines [18]. Studies identified a need for standardized digital care pathways, validated repositories, and mechanisms for sharing information across settings [30]. Dedicated implementation coordinators could support interoperability, training, and communication between technology teams and clinical services [36]. Caregivers should be recognized as optional partners with defined roles and support, not as substitutes for professional care [38]. One HCP said, "I believe that non-governmental organizations and hospitals can work together to help stroke patients, primarily to involve those responsible" [37]. Technologies that did not align with person-centered, integrated care risked superficial use or abandonment [44]. Organizational readiness depends on favorable conditions, such as leadership support, clear roles, and protected time for adaptation [65].

XSL•FO
RenderX

Wider System Domain

The wider system domain encompasses political, economic, and regulatory contexts that shape the implementation of DHTs and influence their scalability and long-term sustainability. Only 6 included studies reported determinants operating at the broader health care, social, or policy system level.

Political/policy Context

Despite the proliferation of DHTs, evidence to guide policy remained limited [35,65]. While technology-supported models may succeed temporarily through local improvisations or pilot projects, they ultimately falter without supportive regulatory frameworks and sustainable financing [38]. Long-term viability demands aligned policy, standardized protocols, and systemic investment, not just technical innovation [40].

Regulatory/legal Issues

While participants acknowledged that data sharing enables personalized feedback, it was also perceived as requiring trade-offs in personal privacy. As a clinical manager noted, “when you want to understand somebody, data must be collected in several areas, and based on these data you can provide feedback. So, you give up some privacy to be coached" [28]. The use of cameras or images within apps raised additional privacy concerns for some users [56]. Another requirement involves obtaining patient consent for sharing sensitive data [38]. The issue is the authenticity of the app, as users felt they were being scammed into paying for features within the app [35]. Moreover, users who contacted customer care services for assistance did not receive a reply, which increased their fears [40]. Unclear reimbursement and payment arrangements created uncertainty for providers and could exclude people unable or unwilling to pay. Trust therefore depended on transparent data practices, credible governance, responsive support, and equitable financing [56].

Professional Context

No eligible study directly examined how professional regulation, accreditation, liability standards, or interprofessional jurisdiction shaped implementation.

Sociocultural Context at the Wider-System Level

No eligible study directly examined population-level sociocultural norms or public discourse as wider-system determinants. Patient-level and family-level cultural findings were retained in the sociocultural factors domain.

Interaction and Adaptation Over Time

No study directly evaluated the seventh NASSS domain. The evidence focused on anticipated use, co-design, cross-sectional experience, or short-term feasibility; it did not track how DHTs were adapted, abandoned, embedded, scaled, spread, or sustained in routine care. This absence limits conclusions about long-term sustainability.

Discussion

Principal Findings

In this review, we synthesized findings from 42 studies to examine stakeholder perspectives on DHTs in poststroke care. The findings were organized within the NASSS framework. The DHT-specific subdomains identified may inform the future development and optimization of DHTs [26]. The earliest included study was published in 2019, and the number of publications increased from 2022 onward, reflecting the rapid development of stroke-related DHTs in recent years [4]. Our findings indicate that the 4 stakeholder groups recognize the importance of the proposed functions, features, and user interface design elements of DHTs, such as personalization, feedback mechanisms, and alert functions. These are seen as essential for the effective use of DHTs in stroke care. A key concern raised by stakeholders related to data security, its ownership, and payment for DHTs. Despite the evident benefits that DHTs could offer patients with stroke in terms of their health, patients emphasized that safety and reliability must be guaranteed as a prerequisite for use. In particular, the supply side adopts a supportive role, focusing on aligning technological solutions with user needs. Our review also revealed that there is limited discussion of clinical and technical accountability among regulators and policymakers. Evidence concerning the adoption of regulatory standards was limited [67]. This suggests that further research should focus on regulatory aspects, particularly in the context of clinical implementation [14].

Personalization, Usability, and Feedback

The perceived usefulness and ease of use of technology are the most significant determinants of its adoption, providing users with a sense of agency in managing their stroke condition [40]. Although designing tailored DHTs is more time-consuming than adopting a one-size-fits-all approach, this approach enables consideration of the unique characteristics, needs, and preferences of different stroke populations, which may enhance user experience, engagement, and adherence [18,32]. Therefore, tailoring both functionality and interface design to specific stroke subgroups is essential [35]. Patients emphasized the need for design elements that favor simplicity and are easy to use and intuitive. For older adults in particular, providing technical support tailored to their needs can enhance their digital skills and address usability barriers [68]. The number of tailored versions required depends on the nature of the condition, associated comorbidities, and sociocultural factors related to stroke. Moreover, by integrating wearable DHTs into clinical care, personalized, data-driven intervention effects can lead to results that are superior to those achieved with the use of generalized clinical data [20]. DHTs can be used to provide personalized health recommendations based on specific patient data, which has the potential to motivate patients, modify their behavior, and tailor their medication and treatment plans [20,69]. Therefore, features of DHTs should facilitate a personalized approach to meet patients’ needs. The findings highlight the critical role of timely and relevant feedback within DHTs in supporting stroke survivors’ engagement and facilitating the monitoring of disease progression. Both patients and clinicians value personalized, frequent feedback mechanisms, as these

XSL•FO
RenderX

help to sustain motivation and enhance recovery tracking [29,59]. Patient portals have the potential to enhance patient engagement in managing their health by allowing access, for example, to discharge summaries, medications, laboratory results, and secure patient-provider communication. Providing visual, auditory, automated, and real-time feedback enhances the user experience, promotes emotional resilience, and supports long-term recovery [70].

Digital Literacy and Human Support

To reduce the "digital divide," ongoing education and programs are essential to enhance digital literacy among both patients and HCPs, ensuring they can effectively use DHTs [19]. DHTs can be delivered either as a purely self-help tool or with human guidance, and incorporating human guidance has been shown to enhance treatment adherence and improve outcomes. In addition, software developers should involve end users in the design and testing phases to ensure that DHT systems align with users' contextual needs and seamlessly integrate into existing routines. Empowering patients with access to their health data not only fosters a sense of control over their own health management but also enhances patient satisfaction [38]. This active involvement can strengthen the patient-provider relationship by promoting shared decision-making and encouraging greater engagement in the management of their condition. The success of such use is often dependent on coordination across home, clinic, and rehabilitation center networks. The role of technology is to facilitate this, rather than to create fragmentation. Caregivers can be better supported through structured training programs, such as toolkits, workshops, or digital courses that build their competence in using DHTs and understanding stroke-related needs [37,46]. Additionally, integrating peer support groups can reduce isolation and foster resilience. Ensuring caregivers have ongoing access to professional guidance and emotional support empowers them to provide consistent, person-centered care [71].

Financing, Regulation, and Data Privacy

Consistent with previous systematic reviews, this study identified a scarcity of health economic evaluations of DHTs in stroke care [72]. There remains a recognized need for ongoing regulation to govern the clinical use of DHTs and the contexts in which they are implemented, underscoring the importance of further research into regulatory frameworks, particularly regarding clinical integration [14]. National-level research funding accounted for the largest proportion of support for the included studies, highlighting its critical role in driving high-quality research on DHTs in stroke care. To ensure equitable access, policy-level decisions are required on reimbursement, welfare support, and public investment, with the aim of positioning digital health as a public good rather than a consumer product. Safety and confidentiality are important ethical issues in DHTs. Studies suggest that users often hesitate to share their personal information with unfamiliar suppliers due to fear of data misuse. Consequently, this affects how users interact with the technology [61]. The amount of personal and health data collected varies depending on the purpose of the intervention; therefore, it is essential to inform users about what data are collected, how they are used, and where they are stored [12].

Implications for Practice

To our knowledge, this is the first systematic review–level synthesis of stroke-related DHT evidence organized using the NASSS framework. This review summarizes the key priorities for optimizing DHTs. The general positive attitude toward DHTs among stakeholders creates a foundation of trust [4]. To maximize impact, future practice should prioritize delivering specific, personalized information and interventions tailored to the diverse characteristics and preferences of patients with stroke. The involvement of family members in the implementation and management of DHTs is important to enhance social and mental support. To ensure the effectiveness and quality of management programs, it is essential for health care providers and other program stakeholders to strengthen training and education, clearly define responsibilities, establish communication platforms, and create supervision and feedback processes [67]. Moreover, given the identified resource limitations, future research should better define the cost-effectiveness and equitable distribution of stroke-related DHTs to ensure long-term sustainability.

Recommendations and Future Directions

Future implementation of DHTs in poststroke care should be treated as a system-level program jointly owned by patients, caregivers, HCPs, developers, organizations, and policymakers. From the outset, implementation plans should define clinical ownership, patient and caregiver support, staff training, workflow integration, escalation pathways, technical maintenance, interoperability, data governance, and sustainable financing. Particular attention should be given to equitable access and to accommodating stroke-related motor, cognitive, communication, and sensory limitations [38]. The absence of evidence regarding interaction and adaptation over time in the NASSS framework also highlights a critical research priority. Longitudinal mixed methods studies embedded in routine care should combine repeated stakeholder assessments with usage, clinical, organizational, and economic data to examine adaptation, scale-up, and sustainability across settings and service transitions. Consistent reporting of implementation processes and contextual changes would help distinguish initial acceptability from sustained adoption and identify which support mechanisms remain effective over time.

Limitations

Several limitations should be acknowledged. First, restricting the review to English-language, peer-reviewed publications and excluding gray literature, conference proceedings, and trial registries may have introduced language and publication bias. Second, although the search combined controlled vocabulary and free-text terms, the diversity and rapidly evolving terminology of DHTs may have led to the omission of relevant terms or studies. Third, interrater agreement coefficients were not prospectively calculated or retained for study screening, quality appraisal, or NASSS mapping. Fourth, the evidence primarily represented patients, caregivers, and HCPs, with limited input from software developers, industry representatives,

XSL•FO
RenderX

and policymakers. This restricted consideration of commercialization, regulation, and scalability. Finally, substantial heterogeneity in DHT functions, apps, settings, and target populations may have contributed to variability across findings and limited their generalizability.

Conclusions

Successful implementation of DHTs in poststroke care requires more than technical functionality. Technologies should accommodate stroke-related impairments, digital access and literacy, caregiver capacity, and the routines of multidisciplinary services. Organizations should define clinical ownership, training, interoperability, data governance, and sustainable financing before deployment. NASSS can be used prospectively to identify implementation complexity and evidence gaps, but the present literature provides insufficient longitudinal evidence to infer sustained use, scale-up, or spread. Future studies should therefore evaluate adaptation and outcomes over time while reporting how patient, caregiver, professional, organizational, and policy priorities are addressed.

Acknowledgments

The authors would like to express their sincere gratitude to Dr Yuanxi Jia from the National University of Singapore for her invaluable guidance and expertise on the methodology of this systematic review. The authors confirm that no generative AI tools were used at any stage in the preparation of this manuscript. The authors take full responsibility for the accuracy, originality, and integrity of the manuscript.

Funding

This work was supported by the National Key Research and Development Program of China (grants 2023YFC3605200 and 2023YFC3605201), the Beijing "Young Seedling Program: for High-Level Innovative and Entrepreneurial Leading Talents (grant 2-1-008-0260), and the Clinical Medical Research Promotion Program of China Medical Foundation (grant 2025CMFA04).

Data Availability

All data analyzed during this study are included in this published article and in its Multimedia Appendix information files.

Authors' Contributions

Jing Zhao and HC conceived and designed the study. Jing Zhao performed the literature search, conducted data extraction, and drafted the manuscript. Jie Zhao, XL, YC, and XW conducted the study selection, quality assessment, and data curation. Jie Zhao and YD provided critical revisions and edited the manuscript for intellectual content. HC supervised the project and finalized the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies.
[DOCX File, 32 KB - mhealth_v14i1e95197_app1.docx]

Multimedia Appendix 2
Study characteristics.
[DOCX File, 59 KB - mhealth_v14i1e95197_app2.docx]

Multimedia Appendix 3
Quality appraisal.
[DOCX File, 25 KB - mhealth_v14i1e95197_app3.docx]

Checklist 1
PRISMA checklist.
[DOCX File, 27 KB - mhealth_v14i1e95197_app4.docx]

References

1. GBD 2021 Causes of Death Collaborators. Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2024 May 18;403(10440):2100-2132. [doi: 10.1016/S0140-6736(24)00367-2] [Medline: 38582094]

XSL•FO
RenderX

2. GBD 2021 Stroke Risk Factor Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet Neurol 2024 Oct;23(10):973-1003. [doi: 10.1016/S1474-4422(24)00369-7] [Medline: 39304265]
3. Feigin VL, Owolabi MO, World Stroke Organization–Lancet Neurology Commission Stroke Collaboration Group. Pragmatic solutions to reduce the global burden of stroke: a World Stroke Organization-Lancet Neurology Commission. Lancet Neurol 2023 Dec;22(12):1160-1206. [doi: 10.1016/S1474-4422(23)00277-6] [Medline: 37827183]
4. Prust ML, Forman R, Ovbiagele B. Addressing disparities in the global epidemiology of stroke. Nat Rev Neurol 2024 Apr;20(4):207-221. [doi: 10.1038/s41582-023-00921-z] [Medline: 38228908]
5. Allan LP, Beilei L, Cameron J, et al. A scoping review of mHealth interventions for secondary prevention of stroke: implications for policy and practice. Stroke 2023 Nov;54(11):2935-2945. [doi: 10.1161/STROKEAHA.123.043794] [Medline: 37800373]
6. What is digital health? US Food and Drug Administration. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence/what-digital-health [accessed 2026-06-20]
7. Narayan SM, Chung MK, Adedinsewo D, et al. Access to digital health technologies: personalized framework and global perspectives. Nat Rev Cardiol 2026 Jan;23(1):9-22. [doi: 10.1038/s41569-025-01184-5] [Medline: 40670723]
8. Cavero-Redondo I, Saz-Lara A, Sequí-Dominguez I, et al. Comparative effect of eHealth interventions on hypertension management-related outcomes: a network meta-analysis. Int J Nurs Stud 2021 Dec;124:104085. [doi: 10.1016/j.ijnurstu.2021.104085] [Medline: 34601205]
9. Tan J, Gong E, Gallis JA, et al. Primary care-based digital health-enabled stroke management intervention: long-term follow-up of a cluster randomized clinical trial. JAMA Netw Open 2024 Dec 2;7(12):e2449561. [doi: 10.1001/jamanetworkopen.2024.49561] [Medline: 39671199]
10. Wang C, Liu M. Digital interventions for cognitive dysfunction in patients with stroke: systematic review and meta-analysis. J Med Internet Res 2025 Jul 24;27:e73687. [doi: 10.2196/73687] [Medline: 40705404]
11. Liang Q, Tao Y, He J, Bo Y, Xu L, Zhao F. Effects of home-based telemedicine and mHealth interventions on blood pressure in stroke patients: a systematic evaluation and meta-analysis of randomized controlled trials. J Stroke Cerebrovasc Dis 2024 Nov;33(11):107928. [doi: 10.1016/j.jstrokecerebrovasdis.2024.107928] [Medline: 39187214]
12. Ginsburg GS, Picard RW, Friend SH. Key issues as wearable digital health technologies enter clinical care. N Engl J Med 2024 Mar 21;390(12):1118-1127. [doi: 10.1056/NEJMra2307160] [Medline: 38507754]
13. Berwanger O, Machline-Carrion MJ. Digital health–enabled clinical trials in stroke: ready for prime time? Stroke 2022 Sep;53(9):2967-2975. [doi: 10.1161/STROKEAHA.122.037378] [Medline: 35770670]
14. Ling Kuo RY, Freethy A, Smith J, et al. Stakeholder perspectives towards diagnostic artificial intelligence: a co-produced qualitative evidence synthesis. EClinicalMedicine 2024 May;71:102555. [doi: 10.1016/j.eclinm.2024.102555] [Medline: 38549586]
15. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367. [doi: 10.2196/jmir.8775] [Medline: 29092808]
16. Reddy S. Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. Implement Sci 2024 Mar 15;19(1):27. [doi: 10.1186/s13012-024-01357-9] [Medline: 38491544]
17. Amoah D, Prior S, Mather C, Schmidt M, Bird ML. Exploring the unmet needs of young adults with stroke in Australia: can technology help meet their needs? A qualitative study. Int J Environ Res Public Health 2023 Jul 26;20(15):6450. [doi: 10.3390/ijerph20156450] [Medline: 37568991]
18. Bally ELS, Cheng D, van Grieken A, et al. Patients' perspectives regarding digital health technology to support self-management and improve integrated stroke care: qualitative interview study. J Med Internet Res 2023 Apr 4;25:e42556. [doi: 10.2196/42556] [Medline: 37014677]
19. Bassindale K, Golus S, Horder J, et al. The feasibility and user experience of a program of progressive cued activity to promote functional upper limb activity in the inpatient rehabilitation setting with follow-up at home. Appl Sci (Basel) 2025 Mar 2;15(6):3010. [doi: 10.3390/app15063010] [Medline: 40727538]
20. Clancy B, Bonevski B, English C, Guillaumier A. The online health information-seeking behaviors of people who have experienced stroke: qualitative interview study. JMIR Form Res 2024 Oct 18;8:e54827. [doi: 10.2196/54827] [Medline: 39423365]
21. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
22. Hong QN, Fàbregues S, Bartlett G, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Educ Inf 2018;34(4):285-291. [doi: 10.3233/EFI-180221]
23. Stern C, Lizarondo L, Carrier J, et al. Methodological guidance for the conduct of mixed methods systematic reviews. JBI Evid Synth 2020 Oct;18(10):2108-2118. [doi: 10.11124/JBISRIR-D-19-00169] [Medline: 32813460]
24. Hong QN, Pluye P, Bujold M, Wassef M. Convergent and sequential synthesis designs: implications for conducting and reporting systematic reviews of qualitative and quantitative evidence. Syst Rev 2017 Mar 23;6(1):61. [doi: 10.1186/s13643-017-0454-2] [Medline: 28335799]

XSL•FO
RenderX

25. Lizarondo L, Stern C, Carrier J, et al. Mixed methods systematic reviews. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis: JBI; 2020:272-310 URL: https://jbi-global-wiki.refined.site/space/MANUAL/355863557/Previous+versions?attachment=/download/attachments/355863557/JBI_Reviewers_Manual_2020June.pdf&type=application/pdf&filename=JBI_Reviewers_Manual_2020June.pdf#page=270 [accessed 2026-09-24]
26. Abell B, Naicker S, Rodwell D, et al. Identifying barriers and facilitators to successful implementation of computerized clinical decision support systems in hospitals: a NASSS framework-informed scoping review. Implement Sci 2023 Jul 26;18(1):32. [doi: 10.1186/s13012-023-01287-y] [Medline: 37495997]
27. Hogg HDJ, Al-Zubaidy M, Technology Enhanced Macular Services Study Reference Group, et al. Stakeholder perspectives of clinical artificial intelligence implementation: systematic review of qualitative evidence. J Med Internet Res 2023 Jan 10;25:e39742. [doi: 10.2196/39742] [Medline: 36626192]
28. Willems EMG, Vermeulen J, van Haastregt JCM, Zijlstra GAR. Technologies to improve the participation of stroke patients in their home environment. Disabil Rehabil 2022 Nov;44(23):7116-7126. [doi: 10.1080/09638288.2021.1983041] [Medline: 34607474]
29. Hestetun-Mandrup AM, Hamre C, Lund A, Martinsen ACT, He HG, Pikkarainen M. Exploring people with stroke's perceptions of digital technologies in post-stroke rehabilitation - a qualitative study. Disabil Rehabil 2026 Jan;48(2):359-368. [doi: 10.1080/09638288.2025.2504615] [Medline: 40382688]
30. Härkönen H, Myllykangas K, Kärppä M, et al. Perspectives of clients and health care professionals on the opportunities for digital health interventions in cerebrovascular disease care: qualitative descriptive study. J Med Internet Res 2024 Dec 2;26:e52715. [doi: 10.2196/52715] [Medline: 39622027]
31. Jarvis K, Cook J, Bavikatte G, et al. A pilot exploration of staff and service-user perceptions of a novel digital health technology (Virtual Engagement Rehabilitation Assistant) in complex inpatient rehabilitation. Disabil Rehabil Assist Technol 2025 Jan;20(1):64-74. [doi: 10.1080/17483107.2024.2351499] [Medline: 38743465]
32. Dimaguila GL, Batchelor F, Merolli M, Gray K. 'We are very individual': anticipated effects on stroke survivors of using their person-generated health data. BMJ Health Care Inform 2020 Sep;27(3):e100149. [doi: 10.1136/bmjhci-2020-100149] [Medline: 32928779]
33. Smejka T, Henry AL, Wheatley C, Espie CA, Johansen-Berg H, Fleming MK. A qualitative examination of the usability of a digital cognitive behavioral therapy for insomnia program after stroke. Brain Inj 2022 Jan 28;36(2):271-278. [doi: 10.1080/02699052.2022.2034182] [Medline: 35108134]
34. Rai T, Morton K, Roman C, et al. Optimizing a digital intervention for managing blood pressure in stroke patients using a diverse sample: integrating the person-based approach and patient and public involvement. Health Expect 2021 Apr;24(2):327-340. [doi: 10.1111/hex.13173] [Medline: 33316120]
35. Cao W, Li A, Kadir AA, et al. Older adult stroke survivors' needs and perspectives of a stroke app: a qualitative study. BMC Geriatr 2025 Mar 29;25(1):212. [doi: 10.1186/s12877-025-05849-1] [Medline: 40158132]
36. Chen Y, Chen Y, Zheng K, et al. A qualitative study on user acceptance of a home-based stroke telerehabilitation system. Top Stroke Rehabil 2020 Mar;27(2):81-92. [doi: 10.1080/10749357.2019.1683792] [Medline: 31682789]
37. Haji Mukhti MI, Ibrahim MI, Tengku Ismail TA, et al. Exploring the need for mobile application in stroke management by informal caregivers: a qualitative study. Int J Environ Res Public Health 2022 Oct 10;19(19):12959. [doi: 10.3390/ijerph191912959] [Medline: 36232257]
38. Marwaa MN, Kristensen HK, Guidetti S, Ytterberg C. Physiotherapists' and occupational therapists' perspectives on information and communication technology in stroke rehabilitation. PLoS One 2020;15(8):e0236831. [doi: 10.1371/journal.pone.0236831] [Medline: 32857781]
39. Lau SCL, Bright L, Connor LT, Baum CM. Experiences with mobile health-enabled ambulatory monitoring among stroke survivors: a qualitative study. OTJR (Thorofare N J) 2025 Apr;45(2):232-240. [doi: 10.1177/15394492241238948] [Medline: 38491760]
40. Lobo EH, Abdelrazek M, Frølich A, et al. Detecting user experience issues from mHealth apps that support stroke caregiver needs: an analysis of user reviews. Front Public Health 2023;11:1027667. [doi: 10.3389/fpubh.2023.1027667] [Medline: 37304084]
41. Marwaa MN, Guidetti S, Ytterberg C, Kristensen HK. Acceptability of two mobile applications to support cross-sectoral, person-centred and empowering stroke rehabilitation - a process evaluation. Ann Med 2024 Dec;56(1):2302979. [doi: 10.1080/07853890.2024.2302979] [Medline: 38466794]
42. Mayrhuber L, Roffler L, Easthope CA, Gassert R, Lambercy O. Arm activity feedback for stroke survivors - a user needs analysis for the design of a mobile application. IEEE Int Conf Rehabil Robot 2025 May;2025:876-881. [doi: 10.1109/ICORR66766.2025.11062943] [Medline: 40644278]
43. Fallahpour M, Gustavsson M, Guidetti S. Experiences of F@ce - a team-based, person-centred intervention supported by information and communication technology for rehabilitation after a stroke. Disabil Rehabil 2025 Dec;47(25):6681-6691. [doi: 10.1080/09638288.2025.2487556] [Medline: 40219731]
44. Barchéus IM, Ranner M, Månsson Lexell E, Larsson-Lund M. Occupational therapists' experiences of using a new internet-based intervention - a focus group study. Scand J Occup Ther 2024 Jan;31(1):2247029. [doi: 10.1080/11038128.2023.2247029] [Medline: 37708913]

XSL•FO
RenderX

45. Epalte K, Tomsone S, Vētra A, Bērziņa G. Patient experience using digital therapy "Vigo" for stroke patient recovery: a qualitative descriptive study. Disabil Rehabil Assist Technol 2023 Feb;18(2):175-184. [doi: 10.1080/17483107.2020.1839794] [Medline: 33155507]
46. Cardy N, Hunter A, Carter D, et al. Perspectives of people with stroke, caregivers and healthcare professionals on an adaptive mHealth intervention for physical activity in the prevention of secondary stroke: a qualitative study. J Multidiscip Healthc 2024;17:2677-2688. [doi: 10.2147/JMDH.S420530] [Medline: 38831803]
47. Purvis T, Burns C, Barker S, et al. Co-designing a digital stroke prevention platform: leveraging lived experience and expert advice. Health Expect 2025 Jun;28(3):e70293. [doi: 10.1111/hex.70293] [Medline: 40405455]
48. Chien SY. A usability evaluation framework for a mobile application in supporting home-based rehabilitation for stroke patients: a qualitative study. Digit Health 2025;11:20552076251340183. [doi: 10.1177/20552076251340183] [Medline: 42186542]
49. Devittori G, Peduzzi M, Dinacci D, et al. Health knowledge after stroke in Switzerland: a survey among health professionals on current practice and suggestions for the implementation of a technology-based educational program for stroke survivors. BMC Health Serv Res 2024 Oct 19;24(1):1259. [doi: 10.1186/s12913-024-11735-0] [Medline: 39425149]
50. Sarfo FS, Obiako R, Nichols M, et al. Knowledge and perspectives of community members on risk assessment for stroke prevention using mobile health approaches in Nigeria. J Stroke Cerebrovasc Dis 2023 Sep;32(9):107265. [doi: 10.1016/j.jstrokecerebrovasdis.2023.107265] [Medline: 37487320]
51. Mahmood A, Blaizy V, Verma A, et al. Acceptability and attitude towards a mobile-based home exercise program among stroke survivors and caregivers: a cross-sectional study. Int J Telemed Appl 2019;2019:5903106. [doi: 10.1155/2019/5903106] [Medline: 31186627]
52. Belger J, Wagner S, Gaebler M, et al. Application of immersive virtual reality for assessing chronic neglect in individuals with stroke: the immersive virtual road-crossing task. J Clin Exp Neuropsychol 2024 Apr;46(3):254-271. [doi: 10.1080/13803395.2024.2329380] [Medline: 38516790]
53. Bang M, Kim MA, Kim SS, Kim HS. Cognitive training using virtual reality: an assessment of usability and adverse effects. Arch Rehabil Res Clin Transl 2024;6(4):100378. [doi: 10.1016/j.arrct.2024.100378] [Medline: 39822192]
54. Hsieh YW, Howe TH, Lee MT, Tai RY, Chen CC. Design and usability evaluation of an immersive virtual reality mirrored hand system for upper limb stroke rehabilitation. Sci Rep 2025 Feb 17;15(1):5781. [doi: 10.1038/s41598-025-90698-6] [Medline: 39962160]
55. Devittori G, Dinacci D, Romiti D, et al. Unsupervised robot-assisted rehabilitation after stroke: feasibility, effect on therapy dose, and user experience. J Neuroeng Rehabil 2024 Apr 9;21(1):52. [doi: 10.1186/s12984-024-01347-4] [Medline: 38594727]
56. Silvera-Tawil D, Cameron J, Li J, et al. Multicomponent support program for secondary prevention of stroke using digital health technology: co-design study with people living with stroke or transient ischemic attack. J Med Internet Res 2024 Aug 22;26:e54604. [doi: 10.2196/54604] [Medline: 39172512]
57. Namyalo PK, Setekera R, Nakazibwe P. Adopting information and communications technology in the control, prevention, and management of stroke: perspectives from patients and providers in Uganda. Front Stroke 2024;3:1440047. [doi: 10.3389/fstro.2024.1440047] [Medline: 41542257]
58. Baranyi R, Czech P, Hofstatter S, Aigner C, Grechenig T. Analysis, design, and prototypical implementation of a serious game Reha@Stroke to support rehabilitation of stroke patients with the help of a mobile phone. IEEE Trans Games 2020;12(4):341-350. [doi: 10.1109/TG.2020.3017817]
59. Heron N, O'Connor SR, Kee F, et al. Development of a digital lifestyle modification intervention for use after transient ischaemic attack or minor stroke: a person-based approach. Int J Environ Res Public Health 2021 May 2;18(9):4861. [doi: 10.3390/ijerph18094861] [Medline: 34063298]
60. Kjörk EK, Sunnerhagen KS, Lundgren-Nilsson Å, Andersson AK, Carlsson G. Development of a digital tool for people with a long-term condition using stroke as a case example: participatory design approach. JMIR Hum Factors 2022 Jun 3;9(2):e35478. [doi: 10.2196/35478] [Medline: 35657650]
61. Lobo EH, Kensing F, Frølich A, et al. mHealth intervention for carers of individuals with a history of stroke: heuristic evaluation and user perspectives. Digit Health 2022;8:20552076221089070. [doi: 10.1177/20552076221089070] [Medline: 35371533]
62. Allan LP, Silvera-Tawil D, Cameron J, et al. Novel multicomponent digital care assistant and support program for people after stroke or transient ischaemic attack: a pilot feasibility study. Sensors (Basel) 2024 Nov 13;24(22):7253. [doi: 10.3390/s24227253] [Medline: 39599030]
63. Feigin VL, Krishnamurthi R, Medvedev O, et al. Usability and feasibility of PreventS-MD web app for stroke prevention. Int J Stroke 2024 Jan;19(1):94-104. [doi: 10.1177/17474930231190745] [Medline: 37485871]
64. Ramaswamy S, Gilles N, Gruessner AC, et al. User-centered mobile applications for stroke survivors (MAPPS): a mixed-methods study of patient preferences. Arch Phys Med Rehabil 2023 Oct;104(10):1573-1579. [doi: 10.1016/j.apmr.2023.05.009] [Medline: 37295706]

XSL•FO
RenderX

65. Nichols M, Singh A, Sarfo FS, et al. Post-intervention qualitative assessment of mobile health technology to manage hypertension among Ghanaian stroke survivors. J Neurol Sci 2019 Nov 15;406:116462. [doi: 10.1016/j.jns.2019.116462] [Medline: 31610382]
66. Devittori G, Dinacci D, Petrillo C, Rossi P, Gassert R, Lambercy O. Unsupervised robot-assisted therapy at home after stroke: a pilot feasibility study. IEEE Int Conf Rehabil Robot 2025 May;2025:166-171. [doi: 10.1109/ICORR66766.2025.11062934] [Medline: 40644292]
67. Niyomyart A, Ruksakulpiwat S, Benjasirisan C, et al. Current status of barriers to mHealth access among patients with stroke and steps toward the digital health era: systematic review. JMIR mHealth uHealth 2024 Aug 22;12:e54511. [doi: 10.2196/54511] [Medline: 39173152]
68. Ge H, Li J, Hu H, Feng T, Wu X. Digital exclusion in older adults: a scoping review. Int J Nurs Stud 2025 Aug;168:105082. [doi: 10.1016/j.ijnurstu.2025.105082] [Medline: 40279791]
69. Misterka J, Gidner B, Johnson T, Gohel P, Sheehan L, Wertheimer J. Transforming stroke intervention through digital health and person-centered navigation: a deeper dive into the Kandu Health pilot study. Arch Phys Med Rehabil 2025 Apr;106(4):e48. [doi: 10.1016/j.apmr.2025.01.124]
70. Wen X, Li Y, Zhang Q, et al. Enhancing long-term adherence in elderly stroke rehabilitation through a digital health approach based on multimodal feedback and personalized intervention. Sci Rep 2025 Apr 23;15(1):14190. [doi: 10.1038/s41598-025-95726-z] [Medline: 40268986]
71. Rivera BD, Nurse C, Shah V, et al. Do digital health interventions hold promise for stroke prevention and care in Black and Latinx populations in the United States? A scoping review. BMC Public Health 2023 Dec 21;23(1):2549. [doi: 10.1186/s12889-023-17255-6] [Medline: 38129850]
72. Valenzuela Espinoza A, Steurbaut S, Dupont A, et al. Health economic evaluations of digital health interventions for secondary prevention in stroke patients: a systematic review. Cerebrovasc Dis Extra 2019;9(1):1-8. [doi: 10.1159/000496107] [Medline: 30616238]

Abbreviations

DHT: digital health technology
GBD: Global Burden of Disease
HCP: health care professional
MMAT: Mixed Methods Appraisal Tool
NASSS: Nonadoption, Abandonment, Scale-Up, Spread, and Sustainability
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
VERA: Virtual Engagement Rehabilitation Assistant

Edited by L Buis; submitted 12.Mar.2026; peer-reviewed by A Hamdan, FJ Yang; revised version received 31.Aug.2026; accepted 01.Sep.2026; published 09.Oct.2026.

Please cite as:
Zhao J, Zhao J, Liu X, Cao Y, Wang X, Dong Y, Chang H
Stakeholder Perspectives on Digital Health Technologies in Poststroke Self-Management and Rehabilitation: Systematic Review Based on the NASSS Framework
JMIR Mhealth Uhealth 2026;14:e95197
URL: https://mhealth.jmir.org/2026/1/e95197
doi:10.2196/95197

© Jing Zhao, Jie Zhao, Xiaohan Liu, Yongxin Cao, Xuan Wang, Yanhong Dong, Hong Chang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 9.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Representation of Different Populations in Studies Assessing the Validity of Consumer Wearable Photoplethysmography-Based Measurements: Scoping Review

Rebecca M Schipper[1], MSc; Fatime Oumar Djibrillah[2,3], MSc; Meyke Roosink[3,4], PhD; Laura Winkens[5], PhD; Arlene John[2,3], PhD; Eric Hazebroek[1,6], PhD; Annemieke Witteveen[2,3], PhD; Agnes A M Berendsen[1], PhD

[1]Division of Human Nutrition and Health, Wageningen University & Research, P.O. Box 17, Wageningen, The Netherlands

[2]Department of Biomedical Signals and Systems, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, The Netherlands

[3]TechMed Institute, University of Twente, Enschede, The Netherlands

[4]Faculty of Behavioral, Management and Social Sciences, University of Twente, Enschede, The Netherlands

[5]Consumption and Healthy Lifestyles Group, Wageningen University & Research, Wageningen, The Netherlands

[6]Vitalys Obesity Clinic, Rijnstate Hospital, Elst, The Netherlands

Corresponding Author:
Rebecca M Schipper, MSc
Division of Human Nutrition and Health, Wageningen University & Research, P.O. Box 17, Wageningen, The Netherlands

Abstract

Background: Consumer wearables are increasingly being integrated into health research for data collection. Although they are attractive to use, the accuracy of their photoplethysmography (PPG)-based measurements can be influenced by user characteristics such as sex, age, BMI, and skin tone. However, our knowledge regarding the validity of these measurements in certain populations, such as those with darker skin tones, seems limited. This is concerning as uncorrected differences in measurement accuracy can lead to health disparities when consumer wearable measurements are used more frequently. A potential cause for the gap in our knowledge regarding consumer wearable validity is the underrepresentation of certain population groups in studies validating PPG-based consumer wearables.

Objective: This scoping review aimed to map the representation of different sex, age, BMI, and skin tone groups in studies assessing the validity of PPG-based pulse rate, heart rate variability, blood pressure, peripheral blood oxygen saturation (SpO_2), and respiratory rate measurements of consumer wearables.

Methods: A literature search was conducted in Scopus, PubMed, and IEEE Xplore in July 2025. Papers were eligible if they assessed the validity of consumer wearable PPG-based measurements, expressed as the agreement with a reference method. From the included papers, the study population distribution of sex, age, BMI, and Fitzpatrick scale was extracted. To evaluate the representation, percentages of people in specific age, BMI, and skin tone groups were estimated based on reported means and SDs. The median percentage of participants in each population group, as well as the total percentage, is reported.

Results: After the removal of duplicates, 734 papers were screened for eligibility. Following title and abstract screening, 238 papers remained, of which 186 passed full-text screening and were included in the review. Most of the studies (n=160) focused on pulse rate. Sex, age, BMI, and Fitzpatrick scale were reported by 179 (96.0%), 178 (96.0%), 101 (54.0%), and 35 (19.0%) out of 186 studies, respectively. While the median representation was 0% (IQR 0%-8%) for both older adults (>65 y) and individuals with obesity (BMI>30 kg/m^2; IQR 0%-13%), aggregate participation across all studies was higher (1290/6367, 20.0% and 473/3428, 14.0%, respectively). Individuals with underweight (BMI<18.5 kg/m^2) remained rare (median 3%, IQR 0%-7%), and the aggregate was 7.0% (225/3428). The median percentage of people with darker skin tones (Fitzpatrick type V and VI) participating in a study was 0%.

Conclusions: Based on our results, it can be concluded that older adults and people with underweight, obesity, or darker skin tones are generally underrepresented in studies assessing the validity of consumer wearable PPG-based measurements. Future validation studies should focus more on the representativeness of the study population. This can be achieved by setting a benchmark for representativeness and including study population representatives during the study design process.

(JMIR Mhealth Uhealth 2026;14:e87324) doi:10.2196/87324

KEYWORDS

wearable; PPG; photoplethysmography; representation; heart rate; heart rate variability; SpO2; blood pressure; respiration rate

XSL•FO
RenderX

Introduction

In 2012, the first commercial wrist-worn pulse rate sensor was released via a Kickstarter campaign [1]. It was developed for athletes to keep track of their pulse rate during training without having to wear a chest strap. Now, a little more than a decade later, the sensing functionalities of consumer wearables have expanded significantly. In addition to pulse rate, these wearables measure, among others, heart rate variability (HRV), peripheral blood oxygen saturation (SpO_2), respiratory rate, blood pressure, skin temperature, and skin conductance. Using these measurements, composite measures such as stress and sleep quality scores are also derived. The added sensing functionalities have broadened the use of consumer wearables from a tool to enhance exercise performance in athletes to health and well-being monitoring for consumers. It is forecasted that the number of smartwatch users will reach 740.53 million in 2029 [2].

Not only have consumers embraced wearables, but the use of consumer wearables in health research is also increasing. A scoping review published in 2022, mapping the use of wearables in health research, identified 179 studies, of which 163 used a consumer wearable [3]. An increase in the number of studies using wearables published per year was observed, rising from 1 in 2013 to 21 in 2017 and doubling to 56 in 2020. In these studies, consumer wearables were implemented for various purposes, ranging from examining correlations between wearable-derived outcomes and other physiological outcomes to developing risk prediction models, disease screening, and monitoring.

The integration of consumer wearables into health research is attractive, as they enable continuous and remote measurements of diverse outcomes while being more comfortable for participants and often more affordable than medical-grade or research-grade sensors. It should be noted, however, that many consumer wearable measurements, including pulse rate, HRV, SpO_2, blood pressure, and respiratory rate, are based on photoplethysmography (PPG). PPG is an optical measurement technique in which light is used to detect changes in blood volume in the skin microvascular bed [4]. The signal produced by this technique is sensitive to external factors, such as environmental light and movement, as well as user characteristics that influence the optical properties of the skin, such as sex, age, BMI, and skin tone [5-7]. For example, melanin in the skin, a determinant of skin tone, absorbs and scatters light, which influences the PPG signal quality [5]. The validity of PPG-based measurements thus varies between measurement conditions and users.

In the last few years, concerns have been raised about our limited knowledge regarding the validity of PPG-based consumer wearable measurements in certain population groups, such as people with darker skin tones or older individuals [8-10]. As manufacturers often do not report the validity of their consumer wearable measurements, researcher-initiated validation studies are being conducted to compare consumer wearable measurements with a reference method. Thus far, these validation studies have reported the accuracy of consumer wearable measurements under different conditions, such as during rest, physical activity, or sleep, in both controlled laboratory settings and free-living conditions [11,12]. However, the influence of characteristics such as sex, age, BMI, and skin tone on measurement validity remains largely unclear, and studies investigating this influence appear to be scarce [13-15]. If potential differences in measurement validity remain unidentified, these differences could contribute to variations in the accuracy of consumer wearable–based risk prediction, disease screening, and monitoring tools across populations. In turn, this may lead to an increase in health disparities—defined as preventable differences in the opportunity to achieve optimal health—as the integration of consumer wearables into health research and, eventually, health care progresses [8-10].

It is hypothesized that the underrepresentation of certain population groups in studies assessing the validity of PPG-based consumer wearable measurements contributes to the gaps in our knowledge about the validity of these measurements in different populations. To improve our understanding of these topics, the potential gaps in the representation of different population groups need to be identified and addressed. The aim of this scoping review is to systematically map the representation of different sex, age, BMI, and skin tone population groups in studies assessing the validity of PPG-based consumer wearable measurements, including pulse rate, HRV, blood pressure, blood oxygen saturation, and respiratory rate. Additionally, we will evaluate the representation of these different population groups to determine which groups require increased representation in future validation studies to enable the investigation of differences in measurement validity between population groups.

Methods

Study Design

This study was performed between May 2024 and August 2025 using the methodological framework for scoping reviews proposed by Arksey and O'Malley [16]. The registration and protocol for this review are available on the Open Science Framework (OSF) [17]. This review is reported according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) extension for scoping reviews (PRISMA-ScR; Checklist 1) guidelines [18].

Eligibility Criteria

Papers eligible for inclusion in this scoping review were those assessing the criterion validity, the agreement of a new measurement method with a criterion measure, of one or more commercial wearable sensors for measuring pulse rate, HRV, blood pressure, respiratory rate, and/or oxygen saturation in humans. The following exclusion criteria were applied:

- The wearable sensor that was validated is a chest strap, as chest straps are considered a criterion measure [19].
- The wearable sensor that was validated is not and was not at any point commercially available to consumers, as this scoping review focuses on consumer wearables only.
- The wearable sensor that was validated is not meant to be worn continuously during daily activities, as we consider

XSL•FO
RenderX

this an important characteristic of a true consumer wearable device.
- The wearable measurement that was validated was measured using a method other than PPG, as this scoping review focuses on PPG-based measurements only.
- N=1 or case studies are excluded as this review has a focus on the distribution of characteristics within study populations.
- Animal studies are excluded, as this scoping review focuses on human participants.
- Reviews, editorials, opinion pieces, and protocols are excluded, as these are not primary research papers.
- No full text could be obtained, or insufficient information was provided in the full text. At least the sample size, the reference method used, and the name and brand of the wearable used should be mentioned to be able to verify if other exclusion criteria apply.
- Publication languages other than English are not considered, as English is the language in which all involved reviewers are proficient.

Search Strategy

A literature search was conducted on June 11, 2024, in the databases Scopus, PubMed, and IEEE Xplore. To update the results before publication, the same search was repeated on July 16, 2025. The main concepts included in the search strategy were “PPG-based measurements,” “validity,” and “consumer wearable.” The keywords for these concepts were selected through pilot searches. The search strategy for all 3 databases can be found in Multimedia Appendix 1. Additionally, the reference lists of reviews returned by the database search were reviewed to identify additional papers.

Screening and Selection

The database search results were uploaded to EndNote 20 (Clarivate) for duplicate removal, after which the remaining papers were uploaded to Rayyan for screening. Prior to initiating the official screening process, the screening guide was tested by 2 reviewers (RMS and FOD) during a pilot screening round with 10 papers. There were no conflicts between the 2 reviewers, and no adjustments were made to the screening guide. The reviewers, RMS and FOD, then proceeded to independently screen the papers for eligibility based on title and abstract, after which a full-text screening took place. Conflicts were resolved through discussion between the 2 reviewers.

Data Charting

Extracted data included information on the type of publication, wearable, study context, recruitment and eligibility, and study population. A complete overview of the data items extracted is presented in Table 1. Most studies do not report how they assessed sex; therefore, it is often unclear whether they report sex or gender. For consistency within this review, we will refer to this characteristic as sex. For skin tone, it was decided to extract the Fitzpatrick scale as this is how most studies report skin tone [13], and as it is the measure recommended by several guidelines, including the US Food and Drug Administration (FDA) guideline on pulse oximeter evaluation, the guidelines for the evaluation of consumer wearable pulse rate validity published by the INTERLIVE network, and the ANSI/CTA (American National Standards Institute/Consumer Technology Association) standard [19-21].

The data charting form was created by RMS in Microsoft Excel. This form was used independently by 2 data extractors in a data charting pilot (RMS and FOD) with 10% of the included papers. For the main data items of interest, for example, study population characteristics, the percentage agreement between the 2 extractors was 95.5%. However, the agreement on the level of physical activity was low (40%). Therefore, additional guidelines for extracting the level of physical activity were added to the data extraction form, after which RMS proceeded with charting the data of the remaining papers independently.

Table . Extracted data items from 186 studies assessing the validity of consumer wearable photoplethysmography-based measurements.

Category	Extracted information
Publication information	Title, year of publication, first author, and DOI
Wearable	Brand, model, version of operating system, and intended wear location
Context	Study location, reference device, and level of physical activity
Recruitment and eligibility	Recruitment location, recruitment channels, and eligibility criteria
Study population	Sample size, distribution of sex, age, height, weight, BMI, and Fitzpatrick scale, and disease status

Synthesis of Results

With the extracted data, the number of studies per PPG-based measurement, the number of wearable brands and models, studies per year, and per study location were determined. Additionally, the recruitment locations and eligibility criteria were summarized.

Evaluating Representation

To provide an overview of the representation of population groups, 2 different approaches were applied. First, for the characteristic sex, the percentage of women in a study was extracted or calculated based on the number of women and the sample size. Based on this percentage, studies were categorized as equal distribution (at least 40% men or women), men-dominant or women-dominant (more than 60% men or women, respectively), men only, or women only. As none of the studies reported the participation of intersex individuals, this review can therefore only report the distribution of men and women.

XSL•FO
RenderX

Second, for age, weight, height, BMI, and Fitzpatrick scale, the weighted mean (SD) was calculated to capture the distribution of these characteristics in the included studies. Additionally, for age, BMI, and Fitzpatrick scale, the percentage of participants in commonly used groups of these characteristics was estimated to better visualize the representation of these different groups (Textbox 1) [22-24]. The percentages were estimated based on probability, using the mean and SD reported by the included studies [25]. First, the *z*-scores for the boundaries of each group were calculated using the provided mean and SD. Second, to estimate the probability of a value equal to or less than the boundary *z*-scores occurring, known as the cumulative distribution function, the *pnorm* function was used in RStudio (version 2023.6.1.524) with R (version 4.4.1). The percentage of participants in a group was then estimated by subtracting the cumulative distribution function of the lower boundaries from the cumulative distribution function of the upper boundaries. A threshold was applied to ensure that percentages lower than the contribution of 1 participant were rounded to 0. For age, the distribution was often truncated at 18 years of age, as studies only included adults. In these cases, the *ptruncnorm* function with a lower bound of 18 was used instead of the *pnorm* function.

Textbox 1. Age, BMI, and skin tone categories used in a scoping review of 186 studies assessing the validity of consumer PPG-based measurements.

Age categories (years)

- Children: <13
- Adolescents: 13 - 17
- Adults: 18 - 65
- Older adults: ≥65

BMI categories (kg/m^2)

- Underweight: <18.5
- Normal weight: 18.5 - 25
- Overweight: 25 - 30
- Obesity: ≥30

Skin tone categories

- Fitzpatrick type I: always burns, never tans
- Fitzpatrick type II: usually burns, tans less than average
- Fitzpatrick type III: sometimes burns mildly, tans about average
- Fitzpatrick type IV: rarely burns, tans more than average
- Fitzpatrick type V: rarely burns, brown tone
- Fitzpatrick type VI: rarely burns, dark brown or black tone

To visualize the distribution of the representation across studies, the median (IQR) of the percentage of participants within each category was reported. To quantify the representation in all studies combined, the total number of participants within each category was calculated based on the estimated percentages. All plots and quantitative analyses were performed in RStudio (version 2023.6.1.524) with R (version 4.4.1).

To evaluate whether a population group was sufficiently represented, the median and total percentage of participants in this population group were compared to a predetermined benchmark [26]. For the purpose of this review, the selected benchmark is equal representation of all population groups. For example, for BMI with 4 groups, the percentage of participants in each BMI group should approach 25%. This benchmark was selected as equal representation of all population groups is desirable when statistically comparing group results.

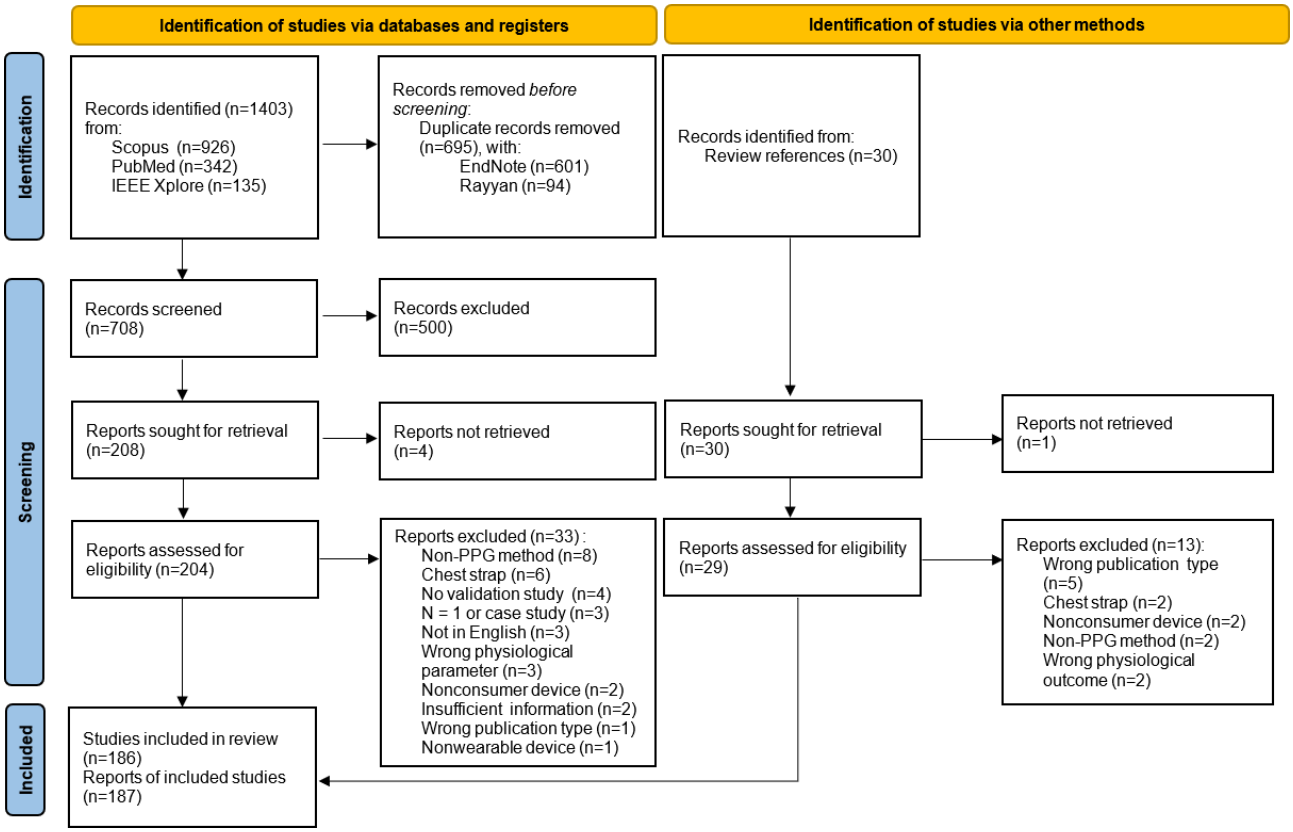
Results

Study Selection

A total of 1403 papers were identified and extracted from PubMed, Scopus, and IEEE Xplore (Figure 1), of which 695 were identified as duplicates and removed, leaving 708 papers for title and abstract screening. Based on title and abstract screening, 500 papers were excluded, and an additional 5 papers were excluded because the full text could not be retrieved. Screening the reference lists of relevant reviews identified 29 additional papers for full-text screening. In the end, 233 papers were included in full-text screening. Based on full-text screening, 187 papers remained for inclusion, providing results from 186 individual studies.

XSL•FO
RenderX

Figure 1. PRISMA flowchart. PPG: photoplethysmography; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

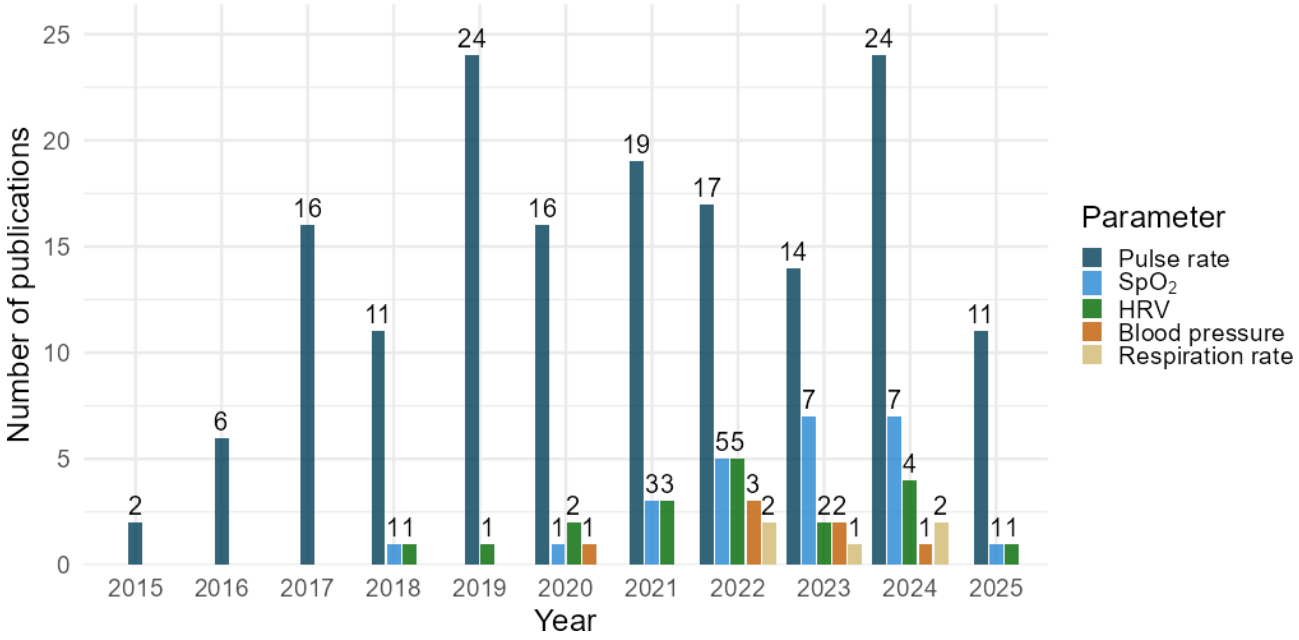


General Study Characteristics

Of the 186 included studies (n=8786 participants), 160 (n=6461 participants) examined pulse rate, 25 (n=2053) oxygen saturation, 19 (n=887) HRV, 7 (n=326) blood pressure [27-33], and 5 (n=329) respiration rate [34-38]. The median (IQR) sample size was 30 (19-49) participants [19,39]. The publication year ranged from 2015 to 2025 (Figure 2), and studies were performed in 32 countries (Table 2). Across all studies, 33 wearable brands and 129 different models were used. Most of the examined wearables were wrist-worn (119 models), and some were worn on the upper arm (3 models), the finger (2 models) [40-44], or in the ear (5 models) [42,45-47]. Most studies (106 studies) validated 1 wearable device; the remaining 80 studies validated 2 to 7 wearable devices. Additional information about the study characteristics of studies for each examined PPG-based measurement can be found in Multimedia Appendix 2. The extracted data for each paper can be found in Multimedia Appendix 3.

XSL•FO
RenderX

Figure 2. Number of studies validating consumer wearable measurements per photoplethysmography-based measurement per year, including studies published until July 16, 2025. HRV: heart rate variability; SpO_2: oxygen saturation.



XSL•FO
RenderX

Table . Number of studies validating consumer wearable photoplethysmography-based measurements by study location at the country level.

Country	Number of studies, n	References
United States	54	[6,32,39,42,43,45,47-94]
Italy	13	[27,30,35,95-104]
Germany	12	[34,36,46,105-113]
Australia	11	[41,114-123]
Canada	10	[124-133]
Republic of Korea	9	[37,134-141]
United Kingdom	9	[142-150]
Spain	8	[151-158]
Belgium	6	[28,38,159-162]
Brazil	6	[29,33,163-166]
Denmark	6	[167-172]
Switzerland	6	[44,173-177]
Finland	4	[40,178-180]
France	3	[181-184]
Ireland	3	[185-187]
Singapore	3	[188-190]
Austria	3	[100,191,192]
China	2	[31,193]
Czech Republic	2	[194,195]
Netherlands	2	[196,197]
Saudi Arabia	2	[198,199]
Taiwan	2	[200,201]
Turkey	2	[202,203]
India	1	[204]
Japan	1	[205]
Malaysia	1	[206]
New Zealand	1	[207]
Norway	1	[208]
Philippines	1	[209]
Poland	1	[210]
Sweden	1	[211]
Thailand	1	[212]

Participant Eligibility and Recruitment

Eligibility Criteria

Of the 186 studies, 150 studies provided eligibility criteria for their study populations, and 69 (37.0%; n=3640 participants) studies stated that adults were eligible for participation. Moreover, there were 31 of 186 studies (7.0%; n=1001 participants) that only included younger adults, 10 (5.0%; n=684 participants) studies that only included children and/or adolescents, 4 (2.0%; n=97 participants) that only included older adults [38,92,163,185], and 1 (n=60 participants) that included children, adolescents, and adults [178]. Eight studies used BMI-based exclusion criteria, either excluding people who were underweight (3 studies) [78,116,188], people with overweight (1 study) [176], people with obesity (2 studies) [77,177], people with clinical obesity (BMI≥40 kg/m^2, 1 study) [108], or both people who were underweight and overweight (1 study) [131].

Health Status

According to the description of the populations, out of 186 studies, most of them (n=137, 74.0%) included healthy populations. There were 47 (25.0%) studies that used the Physical Activity Readiness Questionnaire to screen participants for inclusion. Exclusion criteria for cardiovascular diseases

XSL•FO
RenderX

(n=81, 44.0%), musculoskeletal injuries, diseases, or disabilities (n=46, 25.0%), and pregnancy (n=11, 6.0%) were common. Of 186 studies, 49 (26.0%; n=4921 participants) were performed on patient populations with various conditions. These conditions included cardiovascular diseases (21 studies), respiratory diseases (8 studies) [38,49,53,82,101,166,169,202], surgeries (4 studies) [33,36,98,108], disabilities (3 studies) [167,197,208], breast cancer (1 study) [92], COVID-19 (1 study) [123], chronic musculoskeletal pain (1 study) [211], and cystic fibrosis (1 study) [116].

Recruitment Location

Of the 186 studies, 91 (49.0%) studies provided information on the recruitment location; 46 (25.0%) studies recruited participants from a hospital, and 26 (14.0%) studies recruited participants from a university or research group exclusively. Other recruitment settings included local communities and associations (n=9, 5.0%) [55,80,92,132,136,140,142,185,187], patient organizations (n=3, 2.0%) [167,197,208], professional environments such as the military (n=3, 2.0%) [121,174,176], and a physical activity challenge (n=1, 1.0%) [188]. One study included participants from a previous trial [148].

Reporting of Study Population Characteristics

In the 186 studies reviewed, the most reported study population characteristics were age and sex (n=178, 96.0%, and n=179, 96.0%, respectively), followed by BMI (n=101, 54.0%), weight (n=102, 55.0%), height (n=98, 53.0%), and finally the Fitzpatrick scale (n=35, 19.0%; Table 3). Over time, the percentage of studies reporting weight and height remained stable at ~50% (Figure 3). The percentage of studies reporting BMI fluctuated over time, with a minimum of 0% (0/2) in 2015 and a maximum of 77.0% (10/13) in 2025. An upward trend in the reporting of the Fitzpatrick scale was observed from 2019 to 2025 (1/24, 4.0% to 5/13, 38.0%).

Table . Number (%) of consumer wearable photoplethysmography (PPG)-based measurement validation studies reporting sex, age, BMI, weight, height, and Fitzpatrick scale.

Characteristics	All	PR[a]	HRV[b]	SpO_2[c]	BP[d]	RR[e]
Studies, N	186	160	19	25	7	5
Sex, n (%)	179 (96.0)	157 (99.0)	16 (84.0)	23 (92)	7 (100)	5 (100)
Age, n (%)	178 (96.0)	156 (98.0)	16 (84.0)	22 (88)	7 (100)	4 (80)
BMI, n (%)	101 (54.0)	87 (54.0)	7 (37.0)	13 (52)	3 (43.0)	4 (80)
Weight, n (%)	102 (55.0)	93 (58.0)	6 (32.0)	12 (48)	1 (14.0)	1 (20)
Height, n (%)	98 (53.0)	89 (56.0)	5 (26.0)	12 (48)	1 (14.0)	1 (20)
Fitzpatrick scale, n (%)	35 (19.0)	31 (19.0)	3 (16.0)	9 (36)	1 (14.0)	2 (40)
All, n (%)	15 (8.0)	14 (9.0)	2 (11.0)	3 (12)	0 (0)	1 (20)

[a]PR: pulse rate.

[b]HRV: heart rate variability.

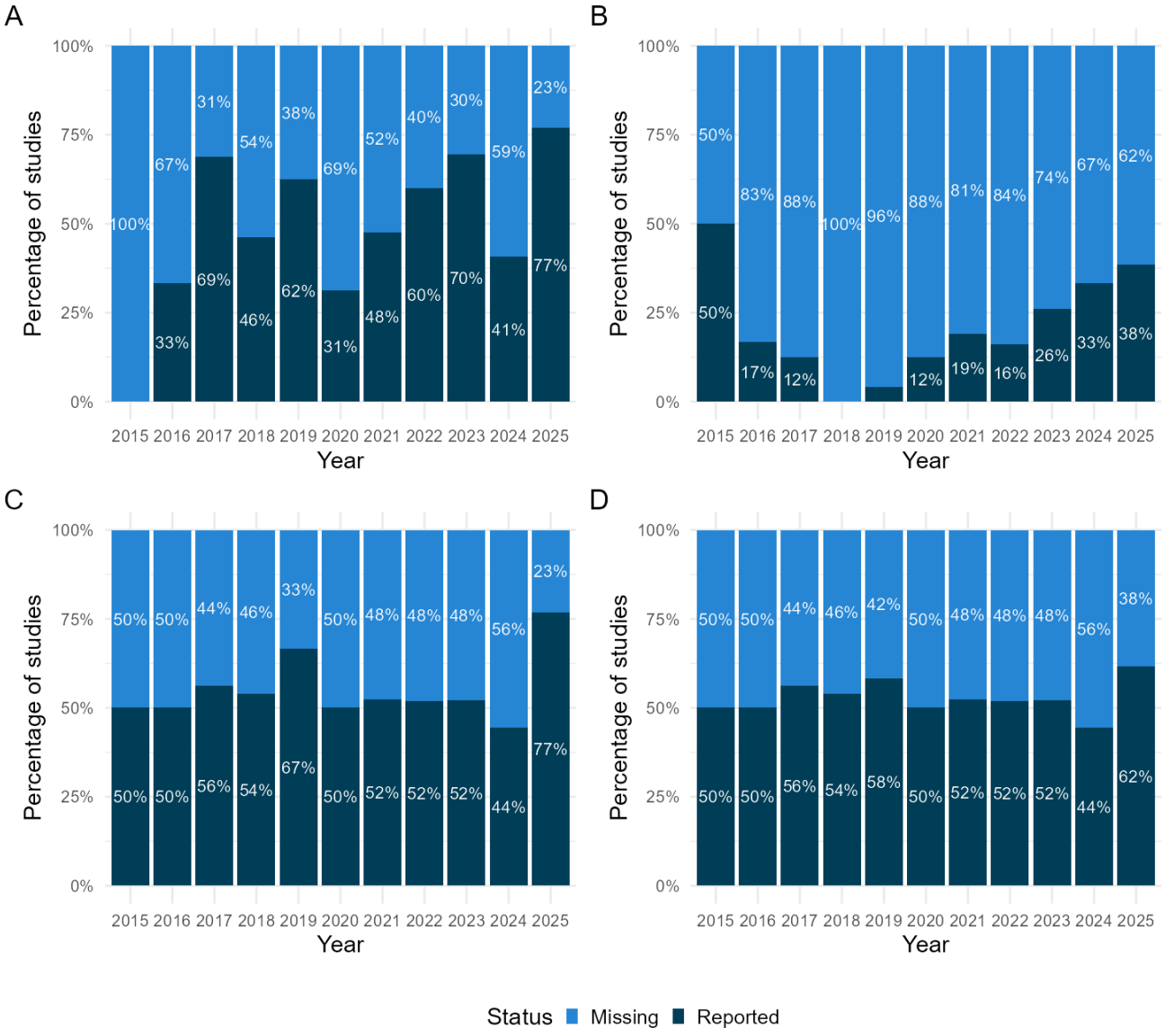
[c]SpO_2: oxygen saturation.

[d]BP: blood pressure.

[e]RR: respiratory rate.

XSL•FO
RenderX

Figure 3. The percentage of consumer wearable photoplethysmography-based measurement validation studies per year reporting BMI (A), Fitzpatrick scale (B), weight (C), and height (D).



Age, BMI, weight, and height were mostly reported as mean or median combined with an SD, SE, or IQR (age: 158/178, 89.0%; BMI: 98/101, 97.0%; weight: 98/102, 96.0%; height: 94/98, 96.0%). A lower percentage of studies additionally reported the range (age: 44/178, 25.0%; BMI: 25/101, 25.0%; weight: 25/102, 25.0%; height: 23/98, 23.0%). Reporting of the Fitzpatrick scale was highly variable, with studies either reporting a combination of mean, median, SD, or IQR (12/35, 34.0%), the percentage of participants in each level or in a combination of at least 2 levels (6/35, 17.0%, for both), the range (5/35, 14.0%), or a combination of these options (6/35, 17.0%).

Representation of Study Population Characteristics

Representation of Men and Women

In 179 studies (n=8203 participants), the percentage of women in a study was 43±19%. Most of the studies (102 studies, n=4570 participants) included an approximately equal percentage of men and women. There were 46 men-dominant studies (n=2637 participants), 17 women-dominant studies (n=688 participants), 11 men-only (n=277 participants) [59,121,135,146,157,164,167,174,199], and 3 women-only (n=31 participants) studies [35,136,213].

Representation of Age Groups

The weighted mean (SD) of age across 149 studies (n=6367) reporting these statistics was 44 (11.9) years. The lowest reported age was 1.2 months [127], and the highest reported age was 94 years [126,202]. In the 112 studies (n=5088), which did not exclude older adults and reported the mean (SD), the median (IQR) percentage of older adults in a study was 0% (0%-8%). The total estimated number of older adults participating across the 149 studies that reported the mean (SD) was 1290 of 6367 participants (20.0%). The results of each examined PPG-based measurement separately are presented in Table 4 and Multimedia Appendix 4. Of these 1290 estimated older adults, 1247 (97.0%) participated in studies with a focus on patient populations. From 2021 to 2024, the total percentage of older adults participating each year increased from 23% to 38% (Multimedia Appendix 5).

XSL•FO
RenderX

Table . Percentage (median, IQR) of older adults and participants belonging to BMI and skin tone groups in consumer wearable photoplethysmography-based measurement validation studies.

Participant characteristics	PR[a,b], median (IQR)	HRV[c,d], median (IQR)	BP[e,f], median (IQR)	SpO_2[g,h], median (IQR)	RR[i,j], median (IQR)
Age					
Older adults	0 (0 - 0)	0 (0 - 0)	11 (2-28)	6 (0 - 27)	11 (6-14)
BMI					
Underweight	4 (0 - 7)	6 (3-7)	4 (2-6)	3 (0 - 3)	3 (1-3)
Normal weight	58 (41-72)	42 (36-46)	23 (30-37)	34 (26-50)	37 (26-68)
Overweight	31 (17-36)	34 (27-36)	34 (30-37)	40 (28-41)	27 (13-35)
Obesity	0 (0 - 12)	13 (9-19)	37 (36-38)	20 (0 - 33)	15 (8-34)
Fitzpatrick scale					
I	15 (7-24)	18 (16-20)	—[k]	15 (3-21)	—
II	35 (19-45)	32 (24-41)	—	34 (26-46)	—
III	25 (14-30)	20 (20,21)	—	21 (15-30)	—
IV	9 (3-21)	9 (4-13)	—	10 (2-18)	—
V	0 (0 - 9)	9 (4-13)	—	1 (0 - 3)	—
VI	0 (0 - 0)	9 (5-14)	—	0 (0 - 2)	—

[a]PR: pulse rate.

[b]Age: n=132; BMI: n=78; Fitzpatrick scale: n=16.

[c]HRV: heart rate variability.

[d]Age: n=12; BMI: n=5; Fitzpatrick scale: n=2.

[e]BP: blood pressure.

[f]Age: n=6; BMI: n=2.

[g]SpO_2: peripheral blood oxygen saturation.

[h]Age: n=16; BMI: n=8; Fitzpatrick scale: n=6.

[i]RR: respiratory rate.

[j]Age: n=5; BMI: n=3.

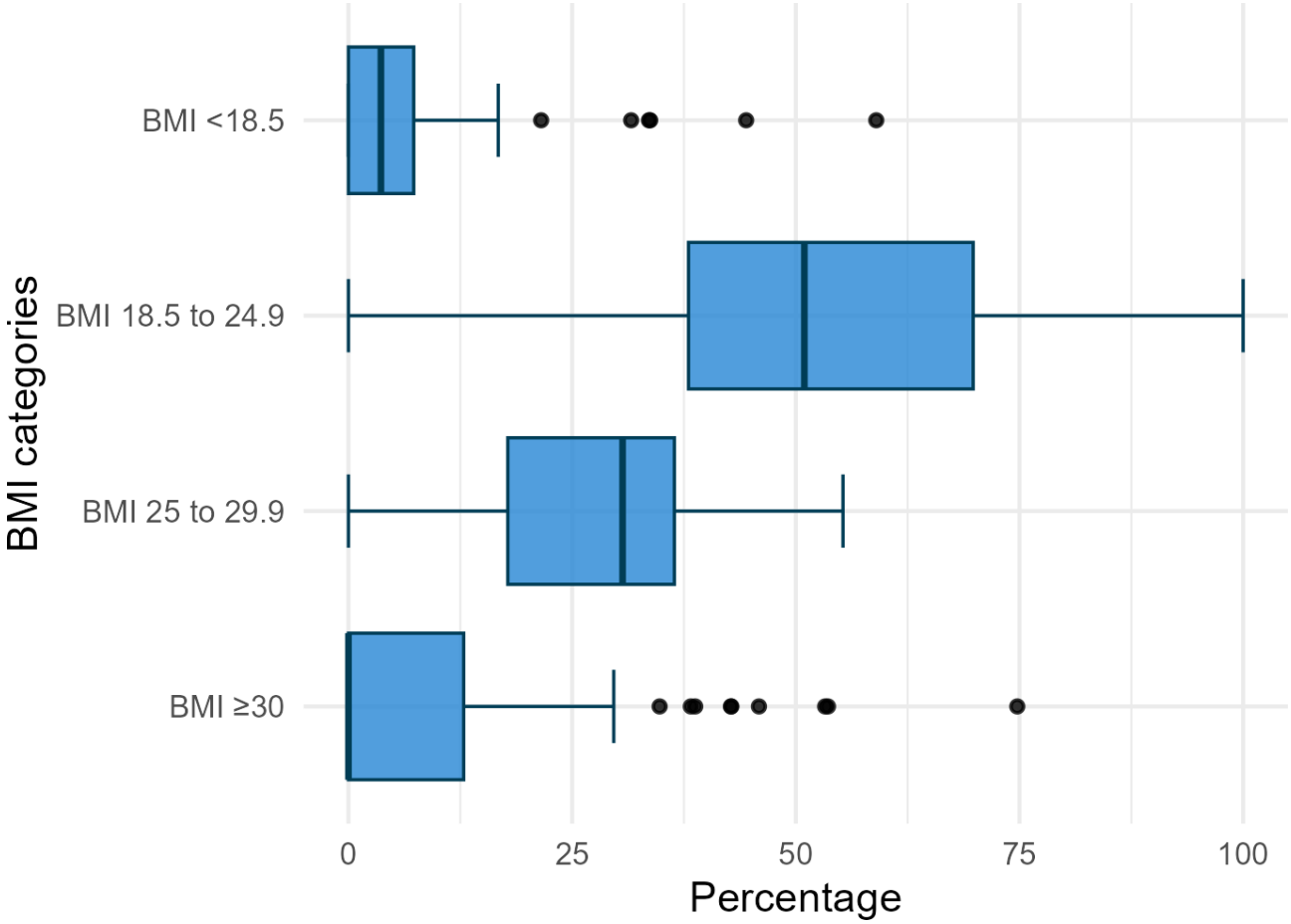
[k]Not applicable.

Representation of BMI Categories

Based on 89 studies (n=3428), the weighted mean (SD) of BMI was 24.8 (4.2) kg/m^2, with BMIs ranging from 16.3 [118] to 52.0 kg/m^2 [34]. In these studies, the median (IQR) percentage of people with underweight, normal weight, overweight, and obesity in a study was 3% (0%-7%), 51% (38%-70%), 31% (18%-37%), and 0% (0%-13%), respectively (Figure 4). Of the 3428 participants, it was estimated that 225 (7.0%) were underweight, 1602 (47.0%) had normal weight, 1048 (31.0%) were overweight, and 473 (14.0%) had obesity. Of the 473 estimated participants with obesity, 364 (77.0%) participated in studies that focused on patient populations. The representation of people with obesity increased over time from 4% in 2016 to 22% in 2022, after which it decreased to 15% in both 2023 and 2024 (Multimedia Appendix 5).

XSL•FO
RenderX

Figure 4. Boxplot of the distribution of the estimated percentage of participants in 4 BMI (kg/m^2) categories in 89 studies (n=3428 participants) validating consumer wearable photoplethysmography-based measurements.

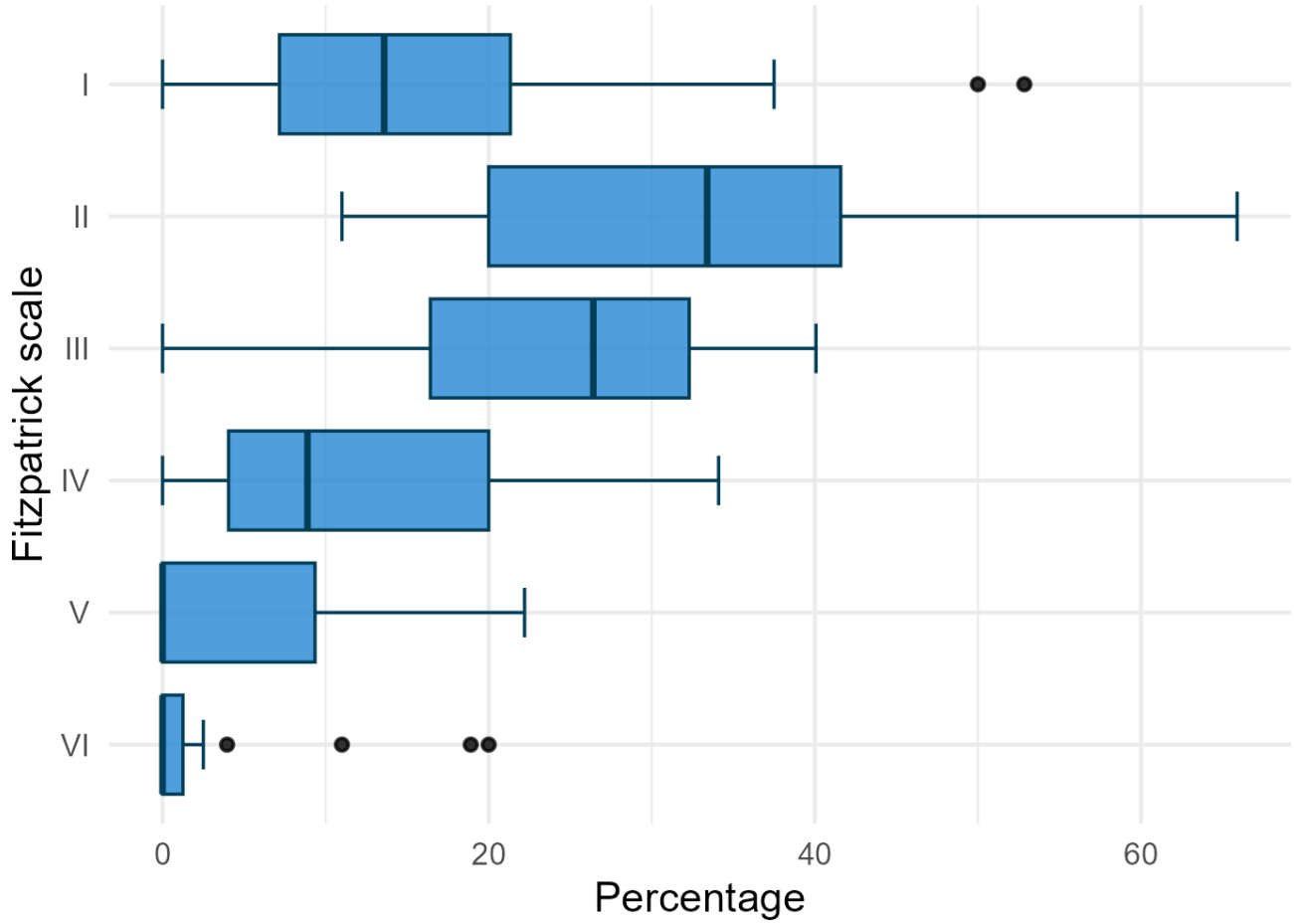


Representation of Fitzpatrick Scale Categories

In the 19 studies (n=876) that reported the mean (SD) or percentage of participants in each Fitzpatrick scale category, the majority of participants fell in type II (median 33%, IQR 20%-42%), while the minority were in types V and VI (median 0%, IQR 0%-9% and median 0%, IQR 0%-1%, respectively; Figure 5). Over the years, the representation of the different categories fluctuated, with no clear trend toward increasing or decreasing representation of any category (Multimedia Appendix 5).

XSL•FO
RenderX

Figure 5. Boxplot of the distribution of the estimated percentage of participants in the Fitzpatrick scale categories in 19 studies (n=876 participants) validating consumer wearable photoplethysmography-based measurements.



Discussion

Principal Findings

This scoping review aimed to provide an overview of the representation of different sex, age, BMI, and skin tone groups in studies assessing the validity of PPG-based consumer wearable measurements. Based on the inclusion of 186 studies, our results indicate that the representation of the different population groups is not equal. In particular, older adults, people who were underweight, people living with obesity, and people with darker skin tones (Fitzpatrick type V or VI) are underrepresented compared to other groups. When interpreting our findings, it is important to consider the following points of discussion: (1) the reporting of user characteristics by the included studies; (2) the benchmark used for evaluating representation; and (3) the relevance of evaluating both study-level and total representation.

First, the results of this study revealed that a large percentage of studies assessing the validity of consumer wearables' PPG-based measurements did not report BMI and skin tone. Not only does this limit our ability to fully evaluate the representation of these characteristics in this review, but it also indicates a critical gap in the available evidence. The nonreporting of these characteristics provides a barrier to investigating the influence of these characteristics on measurement validity. To increase the reporting of study population characteristics like BMI and skin tone, reporting guidelines for consumer wearable validation studies, such as those provided by the INTERLIVE network and the ANSI/CTA, could help. These reporting guidelines recommend the reporting of study populations' sex, age, weight, height, BMI, and skin tone distribution for studies validating pulse rate measurements and could be extended toward other PPG-based validation studies [19,20]. To truly achieve improved reporting, guidelines like those from the INTERLIVE network should additionally be endorsed by journals that publish validation studies [214,215].

Second, the benchmark used to evaluate representation impacts what is considered sufficient representation and underrepresentation. For example, using equal representation as a benchmark, we consider people with underweight, with a median representation of 3%, to be underrepresented. However, using the prevalence of underweight in the general population, namely 1.6%, as a benchmark, the representation of people who are underweight would be considered sufficient [216,217]. The most suitable benchmark depends on the study's aim. When the aim is to investigate the influence of sex, age, BMI, and skin tone on wearable measurement validity, equal representation of these groups is a suitable benchmark. However, if a validation study aims to reflect performance in a specific target population or the general population, then equal representation of all groups would not be a suitable benchmark. For this purpose, the prevalence of sex, age, BMI, and skin tone groups in the target or general population should be used as a benchmark. The

XSL•FO
RenderX

Clinical Trial Diversity Rating framework, developed by the Institute for Clinical and Economic Review, can then be used to evaluate if this benchmark is sufficiently met [26]. Selecting a suitable benchmark is an important step toward achieving study population representativeness.

Third, in this review, we evaluated representation at the level of individual studies as well as total representation. There are population groups that had low representation at both levels, such as children and adolescents, people who are underweight, and people with darker skin tones (Fitzpatrick types V and VI). However, for older adults and people with obesity, although the representation at the study level was very low, the overall representation was higher. This discrepancy can occur if there is a particular subset of studies in which a group is highly represented, while in other studies, the representation is low. For example, in this scoping review, there was a subset of studies with large sample sizes conducted in patient populations in which older adults were highly represented, while in the remaining studies with smaller sample sizes, almost no older adults participated. In these validation studies, the older adults represented are thus mostly older adults living with disease, who might not be an accurate representation of healthy older adults. This pattern was only observed because we evaluated representation both at the level of individual studies and at the overall level, indicating the relevance of evaluating both levels. In addition to evaluating representation at these 2 levels, it would have been valuable to examine the representation of demographic characteristics intersectionally. Based on the information provided by the included studies, this was not feasible. In reality, the 4 characteristics investigated in this review coexist in one individual, and their cumulative effect is what influences signal quality. An older individual with obesity and dark skin, for example, could present an "intersectional stress test" that current algorithms frequently fail. This necessitates stratified sampling and reporting in future validation studies.

Increasing the Representation of Underrepresented Population Groups

There are 2 explanations for the underrepresentation observed in this review. The first cause is the active exclusion of some population groups through exclusion criteria. For example, many of the validation studies investigating pulse rate measurement validity included in this review had a vigorous exercise component, which can be harmful to some population groups, including older adults. Likely due to this ethical consideration, 31 studies excluded older adults. The second cause is barriers to participation related to other elements of the study. These elements included the research question, the composition of the research team, the recruitment methods, the complexity of research processes, and the required investment of time and resources from the participants [218]. For example, none of the studies had skin tone-based eligibility criteria, yet the representation of people with darker skin tones was very low, which suggests that some other study elements could have prevented their participation.

Irrespective of the cause, the underrepresentation of older adults, people who are underweight, people with obesity, and people with darker skin tones currently provides a barrier for us to gain a better understanding of the differences in measurement validity between population groups. If there are relevant differences in consumer wearable measurement validity between these groups, as suspected, we need to take them into account when analyzing consumer wearable measurement data and interpreting the results. If not, differences in the accuracy of consumer wearable-based risk prediction, disease screening, and monitoring tools between populations may be created, exacerbating already existing health disparities as the implementation of these tools progresses. Additionally, a reduced accuracy of these tools in older adults and people with obesity poses a public health risk, as these tools can be used to screen for conditions that are more prevalent in these population groups (eg, atrial fibrillation and sleep apnea) [219]. Increasing the representation of the "missing" population groups in consumer wearable validation studies is therefore vital. To increase the representation of these groups, a suitable benchmark for representation should be selected during the study design process, and the representativeness of the study population should be evaluated based on that benchmark throughout the entire study. Ideally, this evaluation should not only be conducted by the research team but also by funding agencies, journals, peer reviewers, and other stakeholders in the research process. Additionally, researchers should involve their study populations in the study design process to reduce barriers to participation.

Strengths and Limitations

This scoping review is the first to provide a systematic overview of the representation of sex, age, BMI, and skin tone groups in consumer wearable validation studies. Through an extensive search in 3 electronic bibliographic databases and the reference lists of 58 reviews, we were able to include 186 relevant papers. To ensure the transparency of our methods, the entire review process was guided by a protocol that is openly available on OSF. Nevertheless, this review has some limitations that need to be considered.

The representation of age, BMI, and skin tone groups was estimated based on the reported mean (SD), assuming a normal distribution. Though this approach allowed us to visualize the representation of these population groups, there is likely some overestimation or underestimation in these percentages. We hypothesize, based on our results, that the distribution of age, BMI, and skin tone is more likely to be skewed to the right in the included validation studies. An overestimation of the percentage of older adults, people with obesity, and people with darker skin tones in this review is therefore more likely. This suggests that the true representation of these population groups could be even lower than the percentages presented in this review. This is important to take into account when interpreting our findings, as is the fact that there is likely still variation in the representation within the categories applied in this review. Especially within the category of adults, which includes both young adults (18 - 40 y) and middle-aged adults (41 - 64 y), the high representation of young adults might mask the underrepresentation of middle-aged adults. Similarly, as no breakdown of the representation of the demographic groups per brand or model is reported in this review, it is important to note

XSL•FO
RenderX

that there may be differences in the adequacy of representation between different devices, which are not reported in this review.

Another limitation of this review is the use of the Fitzpatrick scale to assess the representation of skin tone. The Fitzpatrick scale was included in this review because alternative measures are currently limited, and this scale is recommended by current reporting guidelines. However, it does not accurately depict the true representation of different skin tones with different melanin content, which truly impacts the PPG signal by absorbing and scattering the light [5]. Therefore, we recommend that future validation studies consider the use of melanometry, a method using reflected light that has been suggested as an objective quantitative measurement of skin melanin [220]. Future studies should also consider using arterial blood gas as a reference standard for measuring SpO_2 instead of finger pulse oximeters. It is a critical limitation that many included studies used finger pulse oximeters, as these devices also exhibit pigment-related bias. Comparing wearables against them, rather than arterial blood gas, may obscure the true extent of skin tone inaccuracies. Additionally, it should be noted that BMI, like the Fitzpatrick scale, is a crude proxy for what it aims to measure, namely, adiposity. Although BMI is convenient for reporting and provides a general overview of the representation of different body sizes, future validation studies focusing on the mechanistic determinants of the validity of PPG-based measurements should consider measuring body composition.

Additionally, it is important to consider that this review only included studies that were published in English. This could have resulted in an underrepresentation of studies conducted in Asia, Africa, and South America, thereby reducing the representation of darker skin tones observed in this review. During the screening of the search results, only 3 studies were excluded based on the language criterion. This suggests that our search strategy, in general, was not very effective in retrieving non-English validation studies. If a similar review is performed in the future, we therefore recommend investigating methods to improve the retrieval of non-English studies, such as including local scientific literature databases.

Conclusions

With this scoping review, we aimed to identify underrepresented population groups by mapping the representation of sex, age, BMI, and skin tone groups in 186 consumer wearable PPG-based measurement validation studies. Based on the results, we conclude that older adults, people with underweight or obesity, and people with darker skin tones are generally strongly underrepresented in consumer wearable validation studies. Researchers making use of PPG-based consumer wearable measurements in their studies should, therefore, be aware of the limited knowledge of the measurement validity in these population groups and take this uncertainty into account when interpreting their findings. To avoid the creation of further health disparities, there should be more focus on the representativeness of consumer wearable validation study populations. More transparent reporting of study population characteristics and the use of benchmarks for representation will allow researchers and other stakeholders in the research process to evaluate the representativeness of a study population. To achieve study populations that properly represent the target population, representatives from the target population should be included in the study design process.

Acknowledgments

In the process of this review, ChatGPT (free version) was used in the creation of the R script to provide suggestions for the code. These suggestions were then verified by testing the code and reading other sources. Additionally, Grammarly was used to improve the spelling and grammar of the written text.

Funding

This work was supported by the RECENTRE HTSF program, financed by the 4TU.Federation (grant 4TU-UIT-468).

Data Availability

After the publication of this manuscript, a CSV file with the extracted data and R scripts will be uploaded to the related project in OSF [17]. A CC-BY-4.0 license will apply.

Authors' Contributions

Conceptualization: RMS, AAMB, EH
Data curation: RMS
Formal analysis: RMS
Funding acquisition: AAMB, AW, LW
Investigation: RMS, FOD
Methodology: RMS, AAMB, EH
Project administration: RMS
Resources: AAMB
Software: RMS
Supervision: AAMB, EH
Validation: FOD

XSL•FO
RenderX

Visualization: RMS
Writing – original draft: RMS, AAMB, EH
Writing – review & editing: RMS, FOD, AJ, MR, LW, EH, AW, AAMB

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full search queries for each database.
[DOCX File, 18 KB - mhealth_v14i1e87324_app1.docx]

Multimedia Appendix 2
Study characteristics per photoplethysmography-based measurement.
[DOCX File, 23 KB - mhealth_v14i1e87324_app2.docx]

Multimedia Appendix 3
Study results per source. Sources are presented in alphabetical order.
[DOCX File, 111 KB - mhealth_v14i1e87324_app3.docx]

Multimedia Appendix 4
Representation of the different population groups for each photoplethysmography-based measurement.
[DOCX File, 24 KB - mhealth_v14i1e87324_app4.docx]

Multimedia Appendix 5
Representation of age, BMI, and Fitzpatrick scale categories over time.
[DOCX File, 207 KB - mhealth_v14i1e87324_app5.docx]

Checklist 1
PRISMA-ScR checklist.
[PDF File, 177 KB - mhealth_v14i1e87324_app6.pdf]

References

1. Dickinson L. ALPHA—the Holy Grail of Heart Rate (Look Ma, no hands!). Kickstarter. 2012. URL: https://www.kickstarter.com/projects/alphaheartrate/alpha-the-holy-grail-of-heart-rate-look-ma-no-hand/description [accessed 2024-11-15]
2. Number of user of smartwatches world wide from 2017 to 2029 (in millions). Statista. 2025. URL: https://tinyurl.com/3srbkba4 [accessed 2026-07-14]
3. Huhn S, Axt M, Gunga HC, et al. The impact of wearable technologies in health research: scoping review. JMIR mHealth uHealth 2022 Jan 25;10(1):e34384. [doi: 10.2196/34384] [Medline: 35076409]
4. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas 2007 Mar;28(3):R1-39. [doi: 10.1088/0967-3334/28/3/R01] [Medline: 17322588]
5. Fine J, Branan KL, Rodriguez AJ, et al. Sources of inaccuracy in photoplethysmography for continuous cardiovascular monitoring. Biosensors (Basel) 2021 Apr 16;11(4):126. [doi: 10.3390/bios11040126] [Medline: 33923469]
6. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 2020;3(1):18. [doi: 10.1038/s41746-020-0226-6] [Medline: 32047863]
7. Boonya-Ananta T, Rodriguez AJ, Du Le VN, Ramella-Roman JC. Monte Carlo analysis of optical heart rate sensors in commercial wearables: the effect of skin tone and obesity on the photoplethysmography (PPG) signal. Biomed Opt Express 2021 Dec 1;12(12):7445-7457. [doi: 10.1364/BOE.439893] [Medline: 35003845]
8. Colvonen PJ, DeYoung PN, Bosompra NOA, Owens RL. Limiting racial disparities and bias for wearable devices in health science research. Sleep 2020 Oct 13;43(10):zsaa159. [doi: 10.1093/sleep/zsaa159] [Medline: 32893865]
9. Zinzuwadia A, Singh JP. Wearable devices-addressing bias and inequity. Lancet Digit Health 2022 Dec;4(12):e856-e857. [doi: 10.1016/S2589-7500(22)00194-7] [Medline: 36335031]
10. Ibrahim H, Liu X, Zariffa N, Morris AD, Denniston AK. Health data poverty: an assailable barrier to equitable digital health care. Lancet Digit Health 2021 Apr;3(4):e260-e265. [doi: 10.1016/S2589-7500(20)30317-4] [Medline: 33678589]
11. Doherty C, Baldwin M, Keogh A, Caulfield B, Argent R. Keeping pace with wearables: a living umbrella review of systematic reviews evaluating the accuracy of consumer wearable technologies in health measurement. Sports Med 2024 Nov;54(11):2907-2926. [doi: 10.1007/s40279-024-02077-2] [Medline: 39080098]

XSL•FO
RenderX

12. Fuller D, Colwell E, Low J, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR mHealth uHealth 2020 Sep 8;8(9):e18694. [doi: 10.2196/18694] [Medline: 32897239]
13. Koerber D, Khan S, Shamsheri T, Kirubarajan A, Mehta S. Accuracy of heart rate measurement with wrist-worn wearable devices in various skin tones: a systematic review. J Racial Ethn Health Disparities 2023 Dec;10(6):2676-2684. [doi: 10.1007/s40615-022-01446-9] [Medline: 36376641]
14. Musa SB, Ellis R, Chafe B, et al. Wearable device validity in measuring steps, energy expenditure, and heart rate across age, gender, and body mass index: data analysis from a systematic review. J Phys Act Health 2022 Feb 1;20(2):100-105. [doi: 10.1123/jpah.2022-0160] [Medline: 36535270]
15. Singh S, Bennett MR, Chen C, Shin S, Ghanbari H, Nelson BW. Impact of skin pigmentation on pulse oximetry blood oxygenation and wearable pulse rate accuracy: systematic review and meta-analysis. J Med Internet Res 2024 Oct 10;26:e62769. [doi: 10.2196/62769] [Medline: 39388258]
16. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
17. Schipper RM, Oumar Djibrillah F, Roosink M, et al. The representation of different study populations in studies assessing the validity of consumer wearables for measuring vital signs: a scoping review. OSF. 2024 Nov 27. [doi: 10.17605/OSF.IO/48V7G]
18. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
19. Mühlen JM, Stang J, Lykke Skovgaard E, et al. Recommendations for determining the validity of consumer wearable heart rate devices: expert statement and checklist of the INTERLIVE network. Br J Sports Med 2021 Jul;55(14):767-779. [doi: 10.1136/bjsports-2020-103148] [Medline: 33397674]
20. Physical activity monitoring for heart rate—real world analysis (ANSI/CTA-2065.1). : Consumer Technology Association (CTA); 2023 URL: https://www.cta.tech/standards/ansicta-20651/ [accessed 2026-07-14]
21. FDA executive summary: performance evaluation of pulse oximeters taking into consideration skin pigmentation, race and ethnicity. : U.S. Food and Drug Administration; 2024 URL: https://www.fda.gov/media/175828/download [accessed 2026-07-14]
22. Age groups. National Library of Medicine. 1998. URL: https://www.ncbi.nlm.nih.gov/mesh/68009273 [accessed 2025-10-20]
23. Fitzpatrick TB. The validity and practicality of sun-reactive skin types I through VI. Arch Dermatol 1988 Jun;124(6):869-871. [doi: 10.1001/archderm.124.6.869] [Medline: 3377516]
24. Obesity and overweight. World Health Organisation. 2025. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2025-09-17]
25. Konrat C, Boutron I, Trinquart L, Auleley GR, Ricordeau P, Ravaud P. Underrepresentation of elderly people in randomised controlled trials. The example of trials of 4 widely prescribed drugs. PLoS One 2012;7(3):e33559. [doi: 10.1371/journal.pone.0033559] [Medline: 22479411]
26. Agboola F, Wright AC. A framework for evaluating the diversity of clinical trials. J Clin Epidemiol 2024 May;169:111299. [doi: 10.1016/j.jclinepi.2024.111299] [Medline: 38395092]
27. Antognoli L, Panni L, Cosoli G, Scalise L. Metrological characterization of commercial smartwatches: are these sensors suitable for the assessment of well-being? Presented at: 2023 IEEE International Workshop on Metrology for Living Environment (MetroLivEnv); May 29-31, 2023. [doi: 10.1109/MetroLivEnv56897.2023.10164071]
28. Falter M, Budts W, Goetschalckx K, Cornelissen V, Buys R. Accuracy of Apple Watch measurements for heart rate and energy expenditure in patients with cardiovascular disease: cross-sectional study. JMIR mHealth uHealth 2019 Mar 19;7(3):e11889. [doi: 10.2196/11889] [Medline: 30888332]
29. Lins LFTDS, do Nascimento EGC, da Silva Júnior JA, de Medeiros Fernandes TAA, de Andrade MF, de Mesquita Andrade C. Accuracy of wearable electronic device compared to manual and automatic methods of blood pressure determination. Med Biol Eng Comput 2023 Oct;61(10):2627-2636. [doi: 10.1007/s11517-023-02869-0] [Medline: 37405672]
30. Poli A, Cosoli G, Iadarola G, Spinsante S, Scalise L. Feasibility of blood pressure measurement through wearable devices: analysis of smartwatches performance. Presented at: 2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA); Jun 22-24, 2022. [doi: 10.1109/MeMeA54994.2022.9856533]
31. Yi L, Lv ZH, Hu SY, et al. Validating the accuracy of a multifunctional smartwatch sphygmomanometer to monitor blood pressure. J Geriatr Cardiol 2022 Nov 28;19(11):843-852. [doi: 10.11909/j.issn.1671-5411.2022.11.004] [Medline: 36561062]
32. Hahnen C, Freeman CG, Haldar N, et al. Accuracy of vital signs measurements by a smartwatch and a portable health device: validation study. JMIR mHealth uHealth 2020 Feb 12;8(2):e16811. [doi: 10.2196/16811] [Medline: 32049066]
33. Monteiro R, Rabello GCM, Moreno CR, et al. Enhancing cardiac postoperative care: a smartwatch-integrated remote telemonitoring platform for health screening with ECG analysis. Front Cardiovasc Med 2024;11:1443998. [doi: 10.3389/fcvm.2024.1443998] [Medline: 39380627]
34. Budig M, Stoohs R, Keiner M. Validity of two consumer multisport activity tracker and one accelerometer against polysomnography for measuring sleep parameters and vital data in a laboratory setting in sleep patients. Sensors (Basel) 2022 Dec 6;22(23):9540. [doi: 10.3390/s22239540] [Medline: 36502241]

XSL•FO
RenderX

35. Cosoli G, Antognoli L, Scalise L. Indirect estimation of breathing rate through wearable devices. Presented at: 2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA); Jun 22-24, 2022. [doi: 10.1109/MeMeA54994.2022.9856464]
36. Helmer P, Hottenrott S, Wienböker K, et al. Reliability of continuous vital sign monitoring in post-operative patients employing consumer-grade fitness trackers: a randomised pilot trial. Digit Health 2024;10:20552076241254026. [doi: 10.1177/20552076241254026] [Medline: 38746874]
37. Jung HC, Kang M, Lee NH, Jeon S, Lee S. Impact of placement of Fitbit HR under laboratory and free-living conditions. Sustainability 2020;12(16):6306. [doi: 10.3390/su12166306]
38. Hermans F, Arents E, Blondeel A, et al. Validity of a consumer-based wearable to measure clinical parameters in patients with chronic obstructive pulmonary disease and healthy controls: observational study. JMIR mHealth uHealth 2024 Nov 6;12:e56027. [doi: 10.2196/56027] [Medline: 39504450]
39. Bai Y, Hibbing P, Mantis C, Welk GJ. Comparative evaluation of heart rate-based monitors: Apple Watch vs Fitbit Charge HR. J Sports Sci 2018 Aug;36(15):1734-1741. [doi: 10.1080/02640414.2017.1412235] [Medline: 29210326]
40. Cao R, Azimi I, Sarhaddi F, et al. Accuracy assessment of Oura Ring nocturnal heart rate and heart rate variability in comparison with electrocardiography in time and frequency domains: comprehensive analysis. J Med Internet Res 2022 Jan 18;24(1):e27487. [doi: 10.2196/27487] [Medline: 35040799]
41. Miller DJ, Sargent C, Roach GD. A validation of six wearable devices for estimating sleep, heart rate and heart rate variability in healthy adults. Sensors (Basel) 2022 Aug 22;22(16):6317. [doi: 10.3390/s22166317] [Medline: 36016077]
42. Navalta JW, Montes J, Bodell NG, Salatto RW, Manning JW, DeBeliso M. Concurrent heart rate validity of wearable technology devices during trail running. PLoS One 2020;15(8):e0238569. [doi: 10.1371/journal.pone.0238569] [Medline: 32866216]
43. Stone JD, Ulman HK, Tran K, et al. Assessing the accuracy of popular commercial technologies that measure resting heart rate and heart rate variability. Front Sports Act Living 2021;3:585870. [doi: 10.3389/fspor.2021.585870] [Medline: 33733234]
44. Abdalazim N, Larraza JAA, Alchieri L, Alecci L, Santini S, Gashi S. Heart rate during sleep measured using finger-, wrist- and chest-worn devices: a comparison study. In: Tsanas A, Triantafyllidis A, editors. Pervasive Computing Technologies for Healthcare—16th EAI International Conference, PervasiveHealth 2022, Thessaloniki, Greece, December 12-14, 2022, Proceedings: Springer Nature; 2023:18-32. [doi: 10.1007/978-3-031-34586-9_2]
45. Boudreaux BD, Hebert EP, Hollander DB, et al. Validity of wearable activity monitors during cycling and resistance exercise. Med Sci Sports Exerc 2018 Mar;50(3):624-633. [doi: 10.1249/MSS.0000000000001471] [Medline: 29189666]
46. Passler S, Müller N, Senner V. In-ear pulse rate measurement: a valid alternative to heart rate derived from electrocardiography? Sensors (Basel) 2019 Aug 21;19(17):3641. [doi: 10.3390/s19173641] [Medline: 31438600]
47. Reece JD, Bunn JA, Choi M, Navalta JW. Assessing heart rate using Consumer Technology Association standards. Technologies (Basel) 2021;9(3):46. [doi: 10.3390/technologies9030046]
48. Rozanski GM, Aqui A, Sivakumaran S, Mansfield A. Consumer wearable devices for activity monitoring among individuals after a stroke: a prospective comparison. JMIR Cardio 2018 Jan 4;2(1):e1. [doi: 10.2196/cardio.8199] [Medline: 31758760]
49. Browne SH, Vaida F, Umlauf A, Kim J, DeYoung P, Owens RL. Performance of a commercial smart watch compared to polysomnography reference for overnight continuous oximetry measurement and sleep apnea evaluation. J Clin Sleep Med 2024 Sep 1;20(9):1479-1488. [doi: 10.5664/jcsm.11178] [Medline: 38652502]
50. Bai Y, Tompkins C, Gell N, Dione D, Zhang T, Byun W. Comprehensive comparison of Apple Watch and Fitbit monitors in a free-living setting. PLoS One 2021;16(5):e0251975. [doi: 10.1371/journal.pone.0251975] [Medline: 34038458]
51. Gillinov S, Etiwy M, Wang R, et al. Variable accuracy of wearable heart rate monitors during aerobic exercise. Med Sci Sports Exerc 2017 Aug;49(8):1697-1703. [doi: 10.1249/MSS.0000000000001284] [Medline: 28709155]
52. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR. Accuracy of PurePulse photoplethysmography technology of Fitbit Charge 2 for assessment of heart rate during sleep. Chronobiol Int 2019 Jul;36(7):927-933. [doi: 10.1080/07420528.2019.1596947] [Medline: 30990098]
53. Jiang Y, Spies C, Magin J, Bhosai SJ, Snyder L, Dunn J. Investigating the accuracy of blood oxygen saturation measurements in common consumer smartwatches. PLoS Digit Health 2023 Jul;2(7):e0000296. [doi: 10.1371/journal.pdig.0000296] [Medline: 37437005]
54. Pasadyn SR, Soudan M, Gillinov M, et al. Accuracy of commercially available heart rate monitors in athletes: a prospective study. Cardiovasc Diagn Ther 2019 Aug;9(4):379-385. [doi: 10.21037/cdt.2019.06.05] [Medline: 31555543]
55. Shcherbina A, Mattsson CM, Waggott D, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017 May 24;7(2):2017-2015. [doi: 10.3390/jpm7020003] [Medline: 28538708]
56. Stahl SE, An HS, Dinkel DM, Noble JM, Lee JM. How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? BMJ Open Sport Exerc Med 2016;2(1):e000106. [doi: 10.1136/bmjsem-2015-000106] [Medline: 27900173]
57. Cadmus-Bertram L, Gangnon R, Wirkus EJ, Thraen-Borowski KM, Gorzelitz-Liebhauser J. The accuracy of heart rate monitoring by some wrist-worn activity trackers. Ann Intern Med 2017 Apr 18;166(8):610-612. [doi: 10.7326/L16-0353] [Medline: 28395305]

XSL•FO
RenderX

58. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating accuracy at exercise intensities: a comparative study of self-monitoring heart rate and physical activity wearable devices. JMIR mHealth uHealth 2017 Mar 16;5(3):e34. [doi: 10.2196/mhealth.7043] [Medline: 28302596]
59. El-Amrawy F, Nounou MI. Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? Healthc Inform Res 2015 Oct;21(4):315-320. [doi: 10.4258/hir.2015.21.4.315] [Medline: 26618039]
60. Nash D, Shah MJ, Shehab O, et al. "But for the blind spot": accuracy and diagnostic performance of smart watch cardiac features in pediatric patients. Heart Rhythm 2024 May;21(5):581-589. [doi: 10.1016/j.hrthm.2024.01.021] [Medline: 38246569]
61. Wang R, Blackburn G, Desai M, et al. Accuracy of wrist-worn heart rate monitors. JAMA Cardiol 2017 Jan 1;2(1):104-106. [doi: 10.1001/jamacardio.2016.3340] [Medline: 27732703]
62. Yu H, Kotlyar M, Dufresne S, Thuras P, Pakhomov S. Feasibility of using an armband optical heart rate sensor in naturalistic environment. Pac Symp Biocomput 2022:43-54. [doi: 10.1142/9789811270611_0005] [Medline: 36540963]
63. Dondzila CJ, Lewis C, Lopez JR, Parker T. Congruent accuracy of wrist-worn activity trackers during controlled and free-living conditions. Int J Exerc Sci 2018;11(7):575-584. [doi: 10.70252/UBBT6886]
64. Saghir N, Aggarwal A, Soneji N, Valencia V, Rodgers G, Kurian T. A comparison of manual electrocardiographic interval and waveform analysis in lead 1 of 12-lead ECG and Apple Watch ECG: a validation study. Cardiovasc Digit Health J 2020;1(1):30-36. [doi: 10.1016/j.cvdhj.2020.07.002] [Medline: 35265871]
65. Seshadri DR, Bittel B, Browsky D, et al. Accuracy of the Apple Watch 4 to measure heart rate in patients with atrial fibrillation. IEEE J Transl Eng Health Med 2020;8:2700204. [doi: 10.1109/JTEHM.2019.2950397] [Medline: 32128290]
66. Spierer DK, Rosen Z, Litman LL, Fujii K. Validation of photoplethysmography as a method to detect heart rate during rest and exercise. J Med Eng Technol 2015;39(5):264-271. [doi: 10.3109/03091902.2015.1047536] [Medline: 26112379]
67. Thiebaud RS, Funk MD, Patton JC, et al. Validity of wrist-worn consumer products to measure heart rate and energy expenditure. Digit Health 2018;4:2055207618770322. [doi: 10.1177/2055207618770322] [Medline: 29942628]
68. Thomson EA, Nuss K, Comstock A, et al. Heart rate measures from the Apple Watch, Fitbit Charge HR 2, and electrocardiogram across different exercise intensities. J Sports Sci 2019 Jun;37(12):1411-1419. [doi: 10.1080/02640414.2018.1560644] [Medline: 30657025]
69. Turki A, Behbehani K, Ding K, Zhang R, Li M, Bell K, et al. Estimation of heart rate variability measures using Apple Watch and evaluating their accuracy: estimation of heart rate variability measures using Apple Watch. Presented at: PETRA '21:The 14th PErvasive Technologies Related to Assistive Environments Conference; Jun 29 to Jul 2, 2021. [doi: 10.1145/3453892.3462647]
70. Turki AF, Jani MB, Ding K, Zhang R, Behbehani K. An investigation of heartrate sensing accuracy by wrist-worn fitness tracking devices. Annu Int Conf IEEE Eng Med Biol Soc 2019 Jul;2019:3337-3340. [doi: 10.1109/EMBC.2019.8856375] [Medline: 31946596]
71. Lauterbach CJ, Romano PA, Greisler LA, Brindle RA, Ford KR, Kuennen MR. Accuracy and reliability of commercial wrist-worn pulse oximeter during normobaric hypoxia exposure under resting conditions. Res Q Exerc Sport 2021 Sep;92(3):549-558. [doi: 10.1080/02701367.2020.1759768] [Medline: 32633688]
72. Merrigan JJ, Stovall JH, Stone JD, Stephenson M, Finomore VS, Hagen JA. Validation of Garmin and Polar devices for continuous heart rate monitoring during common training movements in tactical populations. Meas Phys Educ Exerc Sci 2023 Jul 3;27(3):234-247. [doi: 10.1080/1091367X.2022.2161820]
73. Montalvo S, Martinez A, Arias S, et al. Commercial smart watches and heart rate monitors: a concurrent validity analysis. J Strength Cond Res 2023 Sep 1;37(9):1802-1808. [doi: 10.1519/JSC.0000000000004482] [Medline: 36862131]
74. Montes J, Tandy R, Young J, Lee SP, Navalta J. A comparison of multiple wearable technology devices heart rate and step count measurements during free motion and treadmill based measurements. Int J Kinesiol Sports Sci 2019;7(2):30-39. [doi: 10.7575/aiac.ijkss.v.7n.2p.30]
75. Montoye AHK, Mitrzyk JR, Molesky MJ. Comparative accuracy of a wrist-worn activity tracker and a smart shirt for physical activity assessment. Meas Phys Educ Exerc Sci 2017 Oct 2;21(4):201-211. [doi: 10.1080/1091367X.2017.1331166]
76. Navalta JW, Davis DW, Carrier B, et al. Validity and reliability of wearable devices during self-paced walking, jogging and overground skipping. Sport Mont 2023 Oct 1;21(3):23-29. [doi: 10.26773/smj.231004]
77. Nuss KJ, Sanford JL, Archambault LJ, et al. Accuracy of heart rate and energy expenditure estimations of wrist-worn and arm-worn Apple watches. J Meas Phys Behav 2019;2(3):166-175. [doi: 10.1123/jmpb.2018-0061]
78. Pope ZC, Lee JE, Zeng N, Gao Z. Validation of four smartwatches in energy expenditure and heart rate assessment during exergaming. Games Health J 2019 Jun;8(3):205-212. [doi: 10.1089/g4h.2018.0087] [Medline: 31045446]
79. Powierza CS, Clark MD, Hughes JM, Carneiro KA, Mihalik JP. Validation of a self-monitoring tool for use in exercise therapy. PM R 2017 Nov;9(11):1077-1084. [doi: 10.1016/j.pmrj.2017.03.012] [Medline: 28400221]
80. de Zambotti M, Baker FC, Willoughby AR, et al. Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. Physiol Behav 2016 May;158:143-149. [doi: 10.1016/j.physbeh.2016.03.006] [Medline: 26969518]

XSL•FO
RenderX

81. Etiwy M, Akhrass Z, Gillinov L, et al. Accuracy of wearable heart rate monitors in cardiac rehabilitation. Cardiovasc Diagn Ther 2019 Jun;9(3):262-271. [doi: 10.21037/cdt.2019.04.08] [Medline: 31275816]
82. Huynh P, Shan R, Osuji N, et al. Heart rate measurements in patients with obstructive sleep apnea and atrial fibrillation: prospective pilot study assessing Apple Watch's agreement with telemetry data. JMIR Cardio 2021 Feb 8;5(1):e18050. [doi: 10.2196/18050] [Medline: 33555260]
83. Reddy RK, Pooni R, Zaharieva DP, et al. Accuracy of wrist-worn activity monitors during common daily physical activities and types of structured exercise: evaluation study. JMIR mHealth uHealth 2018 Dec 10;6(12):e10338. [doi: 10.2196/10338] [Medline: 30530451]
84. Jo E, Lewis K, Directo D, Kim MJ, Dolezal BA. Validation of biofeedback wearables for photoplethysmographic heart rate tracking. J Sports Sci Med 2016 Sep;15(3):540-547. [Medline: 27803634]
85. Brazendale K, Decker L, Hunt ET, et al. Validity and wearability of consumer-based fitness trackers in free-living children. Int J Exerc Sci 2019;12(5):471-482. [doi: 10.70252/BBTU4889] [Medline: 30899354]
86. AbiMansour JP, Kaur J, Velaga S, Vatsavayi P, Vogt M, Chandrasekhara V. Accuracy and role of consumer facing wearable technology for continuous monitoring during endoscopic procedures. Front Digit Health 2024;6:1422929. [doi: 10.3389/fdgth.2024.1422929] [Medline: 39355612]
87. Bonneval L, Wing D, Sharp S, et al. Validity of heart rate variability measured with Apple Watch Series 6 compared to laboratory measures. Sensors (Basel) 2025 Apr 9;25(8):2380. [doi: 10.3390/s25082380] [Medline: 40285070]
88. Carrier B, Marten Chaves S, Navalta JW. Validation of aerobic capacity (VO2max) and pulse oximetry in wearable technology. Sensors (Basel) 2025 Jan 6;25(1):275. [doi: 10.3390/s25010275] [Medline: 39797066]
89. Chase SE, Liddell RG, McGonagle CL, Ives SJ. Accuracy of the Apple Watch Series 9 for measures of energy expenditure and heart rate at rest and during exercise: impact of skin pigmentation. J Funct Morphol Kinesiol 2024 Dec 17;9(4):275. [doi: 10.3390/jfmk9040275] [Medline: 39728259]
90. Icenhower A, Murphy C, Brooks AK, et al. Investigating the accuracy of Garmin PPG sensors on differing skin types based on the Fitzpatrick scale: cross-sectional comparison study. Front Digit Health 2025;7:1553565. [doi: 10.3389/fdgth.2025.1553565] [Medline: 40212900]
91. Navalta JW, Carrier B, Blank M, et al. Validity and reliability of wearable technology devices during simulated pickleball game play. Sports (Basel) 2024 Aug 28;12(9):234. [doi: 10.3390/sports12090234] [Medline: 39330711]
92. Page LL, Fanning J, Phipps C, Berger A, Reed E, Ehlers D. Heart rate monitoring among breast cancer survivors: quantitative study of device agreement in a community-based exercise program. JMIR Cancer 2024 Jun 20;10:e51210. [doi: 10.2196/51210] [Medline: 38900505]
93. Pierick AR, Burke KJ, Prusi M, et al. Validity of wrist-worn activity tracker heart rate detection in Fontan patients during exercise. Med Sci Sports Exerc 2025 Feb 1;57(2):280-284. [doi: 10.1249/MSS.0000000000003567] [Medline: 39294928]
94. Yasuda M, Silverman J, Best AM. Can caregivers reliably assess their child's heart rate and respiration rate using smartphone or smartwatch applications? Pediatr Emerg Care 2025 Aug 1;41(8):e67-e72. [doi: 10.1097/PEC.0000000000003400] [Medline: 40341066]
95. Benedetti D, Olcese U, Frumento P, et al. A validation study versus portable polysomnography. J Sleep Res 2021;30(6):e13346. [doi: 10.1111/jsr.13346] [Medline: 33837981]
96. Benedetto S, Caldato C, Bazzan E, Greenwood DC, Pensabene V, Actis P. Assessment of the Fitbit Charge 2 for monitoring heart rate. PLoS One 2018;13(2):e0192691. [doi: 10.1371/journal.pone.0192691] [Medline: 29489850]
97. Cosoli G, Antognoli L, Veroli V, Scalise L. Accuracy and precision of wearable devices for real-time monitoring of swimming athletes. Sensors (Basel) 2022 Jun 23;22(13):4726. [doi: 10.3390/s22134726] [Medline: 35808223]
98. Pelizzo G, Guddo A, Puglisi A, et al. Accuracy of a wrist-worn heart rate sensing device during elective pediatric surgical procedures. Children (Basel) 2018 Mar 8;5(3):38. [doi: 10.3390/children5030038] [Medline: 29518020]
99. Rho G, Di Rienzo F, Marinai C, et al. Preliminary assessment of the Samsung Galaxy Watch 5 accuracy for the monitoring of heart rate and heart rate variability parameters. In: Badnjević A, Gurbeta Pokvić L, editors. MEDICON'23 and CMBEBIH'23—Proceedings of the Mediterranean Conference on Medical and Biological Engineering and Computing (MEDICON) and International Conference on Medical and Biological Engineering (CMBEBIH), September 14–16, 2023, Sarajevo, Bosnia and Herzegovina—Volume 1: Imaging, Engineering and Artificial Intelligence in Healthcare: Springer; 2024:22-30. [doi: 10.1007/978-3-031-49062-0_3]
100. Schiefer LM, Treff G, Treff F, et al. Validity of peripheral oxygen saturation measurements with the Garmin Fēnix 5X plus wearable device at 4559 m. Sensors (Basel) 2021 Sep 23;21(19):6363. [doi: 10.3390/s21196363] [Medline: 34640680]
101. Spaccarotella C, Polimeni A, Mancuso C, Pelaia G, Esposito G, Indolfi C. Assessment of non-invasive measurements of oxygen saturation and heart rate with an Apple smartwatch: comparison with a standard pulse oximeter. J Clin Med 2022 Mar 8;11(6):1467. [doi: 10.3390/jcm11061467] [Medline: 35329793]
102. Spinsante S, Porfiri S, Scalise L. Accuracy of heart rate measurements by a smartwatch in low intensity activities. Presented at: 2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA); Jun 26-28, 2019. [doi: 10.1109/MeMeA.2019.8802216]

XSL•FO
RenderX

103. Cosoli G, Antognoli L, Scalise L. Wearable electrocardiography for physical activity monitoring: definition of validation protocol and automatic classification. Biosensors (Basel) 2023 Jan 18;13(2):154. [doi: 10.3390/bios13020154] [Medline: 36831919]
104. Coelli S, Carrara M, Bianchi AM, De Tommaso M, Actis-Grosso R, Reali P. Comparing consumer and research-grade wristbands for inter-beat intervals monitoring. Presented at: 2024 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE); Oct 21-23, 2024. [doi: 10.1109/MetroXRAINE62247.2024.10797108]
105. Budig M, Keiner M, Stoohs R, Hoffmeister M, Höltke V. Heart rate and distance measurement of two multisport activity trackers and a cellphone app in different sports: a cross-sectional validation and comparison field study. Sensors (Basel) 2021 Dec 28;22(1):180. [doi: 10.3390/s22010180] [Medline: 35009723]
106. Düking P, Giessing L, Frenkel MO, Koehler K, Holmberg HC, Sperlich B. Wrist-worn wearables for monitoring heart rate and energy expenditure while sitting or performing light-to-vigorous physical activity: validation study. JMIR mHealth uHealth 2020 May 6;8(5):e16716. [doi: 10.2196/16716] [Medline: 32374274]
107. Hänsel K, Poguntke R, Haddadi H, Alomainy A, Schmidt A. What to put on the user: sensing technologies for studies and physiology aware systems. Presented at: CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems; Apr 21-26, 2018. [doi: 10.1145/3173574.3173719]
108. Helmer P, Hottenrott S, Rodemers P, et al. Accuracy and systematic biases of heart rate measurements by consumer-grade fitness trackers in postoperative patients: prospective clinical trial. J Med Internet Res 2022 Dec 30;24(12):e42359. [doi: 10.2196/42359] [Medline: 36583938]
109. Nissen M, Slim S, Jäger K, et al. Heart rate measurement accuracy of Fitbit Charge 4 and Samsung Galaxy Watch Active2: device evaluation study. JMIR Form Res 2022 Mar 1;6(3):e33635. [doi: 10.2196/33635] [Medline: 35230250]
110. Pätz C, Michaelis A, Markel F, et al. Accuracy of the Apple Watch oxygen saturation measurement in adults and children with congenital heart disease. Pediatr Cardiol 2023 Feb;44(2):333-343. [doi: 10.1007/s00246-022-02987-w] [Medline: 35995951]
111. Mink F, Lutz W, Rafaeli E, Hehlmann M, Schwartz B, Uhl J. Validation and application of wrist-worn wearable physiological measures in psychotherapy. Cogn Ther Res 2025 Oct;49(5):902-919. [doi: 10.1007/s10608-025-10601-5]
112. Weis A, Leroy M, Jux C, Rupp S, Backhoff D. Oxygen saturation measurement in cyanotic heart disease with the Apple Watch. Cardiol Young 2024 Oct 8;34(9):1-3. [doi: 10.1017/S1047951124025216] [Medline: 39376086]
113. Wolf S, Seidel P, Ockenga TA, Schoder D. Heart-to-wear: assessing the accuracy of heart rate sensor measurements of wearable devices in uncontrolled environments. In: Proceedings of the 57th Hawaii International Conference on System Sciences: IEEE Computer Society; 2024:3183-3191. [doi: 10.24251/HICSS.2024.385]
114. Al-Kaisey AM, Koshy AN, Ha FJ, et al. Accuracy of wrist-worn heart rate monitors for rate control assessment in atrial fibrillation. Int J Cardiol 2020 Feb 1;300:161-164. [doi: 10.1016/j.ijcard.2019.11.120] [Medline: 31787389]
115. Climstein M, Alder JL, Brooker AM, et al. Reliability of the Polar Vantage M sports watch when measuring heart rate at different treadmill exercise intensities. Sports (Basel) 2020 Aug 23;8(9):117. [doi: 10.3390/sports8090117] [Medline: 32842476]
116. Gaynor M, Sawyer A, Jenkins S, Wood J. Variable agreement between wearable heart rate monitors during exercise in cystic fibrosis. ERJ Open Res 2019 Oct;5(4):00006-2019. [doi: 10.1183/23120541.00006-2019] [Medline: 31687372]
117. Ge Z, Prasad PWC, Costadopoulos N, Alsadoon A, Singh AK, Elchouemi A. Evaluating the accuracy of wearable heart rate monitors. Presented at: 2016 2nd International Conference on Advances in Computing, Communication, & Automation (ICACCA) (Fall); Sep 30 to Oct 1, 2016. [doi: 10.1109/ICACCAF.2016.7748986]
118. Hettiarachchi IT, Hanoun S, Nahavandi D, Nahavandi S. Validation of Polar OH1 optical heart rate sensor for moderate and high intensity physical activities. PLOS One 2019;14(5):e0217288. [doi: 10.1371/journal.pone.0217288] [Medline: 31120968]
119. Koshy AN, Sajeev JK, Nerlekar N, et al. Smart watches for heart rate assessment in atrial arrhythmias. Int J Cardiol 2018 Sep 1;266:124-127. [doi: 10.1016/j.ijcard.2018.02.073] [Medline: 29887428]
120. Shumate T, Link M, Furness J, Kemp-Smith K, Simas V, Climstein M. Validity of the Polar Vantage M watch when measuring heart rate at different exercise intensities. PeerJ 2021;9:e10893. [doi: 10.7717/peerj.10893] [Medline: 33614295]
121. Uphill A, Kendall KL, Fogarty A, et al. Validity of Apple Watch, Garmin Forerunner 935 and GENEActiv for estimating energy expenditure during close quarter battle training in Special Forces soldiers. Eur J Sport Sci 2024 May;24(5):614-622. [doi: 10.1002/ejsc.12080]
122. Wallen MP, Gomersall SR, Keating SE, Wisløff U, Coombes JS. Accuracy of heart rate watches: implications for weight management. PLoS One 2016;11(5):e0154420. [doi: 10.1371/journal.pone.0154420] [Medline: 27232714]
123. Rajakariar K, Buntine P, Ghaly A, et al. Accuracy of smartwatch pulse oximetry measurements in hospitalized patients with coronavirus disease 2019. Mayo Clin Proc Digit Health 2024 Mar;2(1):152-158. [doi: 10.1016/j.mcpdig.2024.02.001] [Medline: 40206673]
124. Gagnon J, Khau M, Lavoie-Hudon L, Vachon F, Drapeau V, Tremblay S. Comparing a Fitbit wearable to an electrocardiogram gold standard as a measure of heart rate under psychological stress: a validation study. JMIR Form Res 2022 Dec 21;6(12):e37885. [doi: 10.2196/37885] [Medline: 36542432]

XSL•FO
RenderX

125. Kroll RR, Boyd JG, Maslove DM. Accuracy of a wrist-worn wearable device for monitoring heart rates in hospital inpatients: a prospective observational study. J Med Internet Res 2016 Sep 20;18(9):e253. [doi: 10.2196/jmir.6025] [Medline: 27651304]
126. Lee C, Lee C, Fernando C, Chow CM. Comparison of Apple Watch vs KardiaMobile: a tale of two devices. CJC Open 2022 Nov;4(11):939-945. [doi: 10.1016/j.cjco.2022.07.011] [Medline: 36444370]
127. Littell L, Roelle L, Dalal A, et al. Assessment of Apple Watch Series 6 pulse oximetry and electrocardiograms in a pediatric population. PLoS Digit Health 2022 Aug;1(8):e0000051. [doi: 10.1371/journal.pdig.0000051] [Medline: 36812630]
128. Moayedi Y, Abdulmajeed R, Duero Posada J, et al. Assessing the use of wrist-worn devices in patients with heart failure: feasibility study. JMIR Cardio 2017 Dec 19;1(2):e8. [doi: 10.2196/cardio.8301] [Medline: 31758789]
129. Ray I, Liaqat D, Gabel M, de Lara E. Skin tone, confidence, and data quality of heart rate sensing in WearOS smartwatches. Presented at: 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops); Mar 22-26, 2021. [doi: 10.1109/PerComWorkshops51409.2021.9431120]
130. Sequeira N, D'Souza D, Angaran P, Aves T, Dorian P. Common wearable devices demonstrate variable accuracy in measuring heart rate during supraventricular tachycardia. Heart Rhythm 2020 May;17(5 Pt B):854-859. [doi: 10.1016/j.hrthm.2020.02.018] [Medline: 32354450]
131. Horton JF, Stergiou PRO, Fung TS, Katz L. Comparison of Polar M600 optical heart rate and ECG heart rate during exercise. Med Sci Sports Exerc 2017 Dec;49(12):2600-2607. [doi: 10.1249/MSS.0000000000001388] [Medline: 29135785]
132. Hung SH, Serwa K, Rosenthal G, Eng JJ. Validity of heart rate measurements in wrist-based monitors across skin tones during exercise. PLoS One 2025;20(2):e0318724. [doi: 10.1371/journal.pone.0318724] [Medline: 39928630]
133. Held NJ, Perrotta AS, Mueller T, Pfoh-MacDonald SJ. Agreement of the Apple Watch and Fitbit Charge for recording step count and heart rate when exercising in water. Med Biol Eng Comput 2022 May;60(5):1323-1331. [doi: 10.1007/s11517-022-02536-w] [Medline: 35288797]
134. Ahn J, Ra HK, Yoon HJ, Son SH, Ko J. On-device filter design for self-identifying inaccurate heart rate readings on wrist-worn PPG sensors. IEEE Access 2020;8:184774-184784. [doi: 10.1109/ACCESS.2020.3025776]
135. Baek S, Ha Y, Park HW. Accuracy of wearable devices for measuring heart rate during conventional and Nordic walking. PM R 2021 Apr;13(4):379-386. [doi: 10.1002/pmrj.12424] [Medline: 32449291]
136. Choi JH, Kim HJ. Accuracy verification of heart rate and energy consumption tracking devices to develop forest-based customized health care service programs. J People Plants Environ 2019 Apr;22(2):219-229. [doi: 10.11628/ksppe.2019.22.2.219]
137. Hwang J, Kim J, Choi KJ, Cho MS, Nam GB, Kim YH. Assessing accuracy of wrist-worn wearable devices in measurement of paroxysmal supraventricular tachycardia heart rate. Korean Circ J 2019 May;49(5):437-445. [doi: 10.4070/kcj.2018.0323] [Medline: 30808083]
138. Jung H, Kim D, Choi J, Joo EY. Validating a consumer smartwatch for nocturnal respiratory rate measurements in sleep monitoring. Sensors (Basel) 2023 Sep 19;23(18):7976. [doi: 10.3390/s23187976] [Medline: 37766031]
139. Jung H, Kim D, Lee W, et al. Performance evaluation of a wrist-worn reflectance pulse oximeter during sleep. Sleep Health 2022 Oct;8(5):420-428. [doi: 10.1016/j.sleh.2022.04.003] [Medline: 35817700]
140. Kang S, Kim Y, Byun W, Suk J, Lee JM. Comparison of a wearable tracker with actigraph for classifying physical activity intensity and heart rate in children. Int J Environ Res Public Health 2019 Jul 25;16(15):2663. [doi: 10.3390/ijerph16152663] [Medline: 31349667]
141. Kim C, Song JH, Kim SH. Validation of wearable digital devices for heart rate measurement during exercise test in patients with coronary artery disease. Ann Rehabil Med 2023 Aug;47(4):261-271. [doi: 10.5535/arm.23019] [Medline: 37536665]
142. Abt G, Bray J, Benson AC. The validity and inter-device variability of the Apple Watch for measuring maximal heart rate. J Sports Sci 2018 Jul;36(13):1447-1452. [doi: 10.1080/02640414.2017.1397282] [Medline: 29090987]
143. Collins T, Woolley SI, Oniani S, et al. Version reporting and assessment approaches for new and updated activity and heart rate monitors. Sensors (Basel) 2019 Apr 10;19(7):1705. [doi: 10.3390/s19071705] [Medline: 30974755]
144. Hough P, Glaister M, Pledger A. The accuracy of wrist-worn heart rate monitors across a range of exercise intensities. J Phys Act Res 2017 Jan 21;2(2):112-116. [doi: 10.12691/jpar-2-2-8]
145. Jarchi D, Salvi D, Velardo C, Mahdi A, Tarassenko L, Clifton DA. Estimation of HRV and SpO2 from wrist-worn commercial sensors for clinical settings. Presented at: 2018 IEEE 15th International Conference on Wearable and Implantable Body Sensor Networks (BSN); Mar 4-7, 2018. [doi: 10.1109/BSN.2018.8329679]
146. Khushhal A, Nichols S, Evans W, et al. Validity and reliability of the Apple Watch for measuring heart rate during exercise. Sports Med Int Open 2017 Oct;1(6):E206-E211. [doi: 10.1055/s-0043-120195] [Medline: 30539109]
147. Muggeridge DJ, Hickson K, Davies AV, et al. Measurement of heart rate using the Polar OH1 and Fitbit Charge 3 wearable devices in healthy adults during light, moderate, vigorous, and sprint-based exercise: validation study. JMIR mHealth uHealth 2021 Mar 25;9(3):e25313. [doi: 10.2196/25313] [Medline: 33764310]
148. O'Driscoll R, Turicchi J, Hopkins M, et al. The validity of two widely used commercial and research-grade activity monitors, during resting, household and activity behaviours. Health Technol 2020 May;10(3):637-648. [doi: 10.1007/s12553-019-00392-7]
149. Williams K, Jamieson A, Chaturvedi N, Hughes A, Orini M. Validation of wearable derived heart rate variability and oxygen saturation from the Garmin's health snapshot. Presented at: 50th Computing in Cardiology Conference (Computing

XSL•FO
RenderX

in Cardiology 2023, CinC 2023); Sep 24-27, 2023 URL: https://discovery.ucl.ac.uk/id/eprint/10186376/1/CinC2023-237.pdf [accessed 2026-07-14]
150. Jamieson A, Jones S, Chaturvedi N, Hughes AD, Orini M. Accuracy of smartwatches for the remote assessment of exercise capacity. Sci Rep 2024 Oct 3;14(1):22994. [doi: 10.1038/s41598-024-74140-x] [Medline: 39362983]
151. Alfonso C, Garcia-Gonzalez MA, Parrado E, Gil-Rojas J, Ramos-Castro J, Capdevila L. Agreement between two photoplethysmography-based wearable devices for monitoring heart rate during different physical activity situations: a new analysis methodology. Sci Rep 2022 Sep 14;12(1):15448. [doi: 10.1038/s41598-022-18356-9] [Medline: 36104356]
152. Hernando D, Roca S, Sancho J, Alesanco Á, Bailón R. Validation of the Apple Watch for heart rate variability measurements during relax and mental stress in healthy subjects. Sensors (Basel) 2018 Aug 10;18(8):2619. [doi: 10.3390/s18082619] [Medline: 30103376]
153. Martín-Escudero P, Cabanas AM, Dotor-Castilla ML, et al. Are activity wrist-worn devices accurate for determining heart rate during intense exercise? Bioengineering (Basel) 2023 Feb 15;10(2):254. [doi: 10.3390/bioengineering10020254] [Medline: 36829748]
154. Meza C, Juega J, Francisco J, et al. Accuracy of a smartwatch to assess heart rate monitoring and atrial fibrillation in stroke patients. Sensors (Basel) 2023 May 10;23(10):4632. [doi: 10.3390/s23104632] [Medline: 37430546]
155. Ruiz-Malagón EJ, Castro-Infantes S, Ritaco-Real M, Soto-Hermoso VM. Concurrent validity of the Polar Precision Prime photoplethysmographic system to measure heart rate during a tennis training session. Proc Inst Mech Eng P J Sports Eng Technol 2025 Sep;239(3):545-551. [doi: 10.1177/17543371231165102]
156. Ruiz-Malagón EJ, Ruiz-Alias SA, García-Pinillos F, Delgado-García G, Soto-Hermoso VM. Comparison between photoplethysmographic heart rate monitor from Polar Vantage M and Polar V800 with H10 chest strap while running on a treadmill: validation of the Polar Precision Prime photoplethysmographic system. Proc Inst Mech Eng P J Sports Eng Technol 2021 Sep;235(3):212-218. [doi: 10.1177/1754337120976659]
157. Sañudo B, De Hoyo M, Muñoz-López A, Perry J, Abt G. Pilot study assessing the influence of skin type on the heart rate measurements obtained by photoplethysmography with the Apple Watch. J Med Syst 2019 May 22;43(7):195. [doi: 10.1007/s10916-019-1325-2] [Medline: 31119387]
158. Marzano-Felisatti JM, De Lucca L, Priego-Quesada JI, Pino-Ortega J. Heart rate measurement accuracy during intermittent efforts under laboratory conditions: a comparative analysis between chest straps and armband. Appl Sci 2024;14(24):11872. [doi: 10.3390/app142411872]
159. Claes J, Buys R, Avila A, et al. Validity of heart rate measurements by the Garmin Forerunner 225 at different walking intensities. J Med Eng Technol 2017 Aug;41(6):480-485. [doi: 10.1080/03091902.2017.1333166] [Medline: 28675070]
160. Falter M, Scherrenberg M, Driesen K, et al. Smartwatch-based blood pressure measurement demonstrates insufficient accuracy. Front Cardiovasc Med 2022;9:958212. [doi: 10.3389/fcvm.2022.958212] [Medline: 35898281]
161. Van Oost N, Masci F, Schyvens AM, et al. A new perspective on validation of wearables for stress monitoring of occupational drivers. Presented at: 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Jul 15-19, 2024. [doi: 10.1109/EMBC53108.2024.10781880]
162. Vermunicht P, Makayed K, Buyck C, et al. Continuous heart rate measurements in patients with cardiac disease: device comparison and development of a novel artefact removal procedure. Digit Health 2025;11:20552076251337598. [doi: 10.1177/20552076251337598] [Medline: 40547435]
163. Rodrigues E, Lima D, Barbosa P, et al. HRV monitoring using commercial wearable devices as a health indicator for older persons during the pandemic. Sensors (Basel) 2022 Mar 4;22(5):2001. [doi: 10.3390/s22052001] [Medline: 35271148]
164. Damasceno VDO, Costa ADS, Campello MCA, et al. Criterion validity and accuracy of a heart rate monitor. Hum Mov 2022;23(1):60-68. [doi: 10.5114/hm.2021.104188]
165. Lima CS, Bertocco FC, de Oliveira JIV, de Souza TMF, da Silva EP, Von Zuben FJ. Assessment of Samsung Galaxy Watch4 PPG-based heart rate during light-to-vigorous physical activities. IEEE Sens Lett 2024;8(7):1-4. [doi: 10.1109/LSENS.2024.3408089]
166. Pipek LZ, Nascimento RFV, Acencio MMP, Teixeira LR. Comparison of SpO2 and heart rate values on Apple Watch and conventional commercial oximeters devices in patients with lung disease. Sci Rep 2021 Sep 23;11(1):18901. [doi: 10.1038/s41598-021-98453-3] [Medline: 34556765]
167. Jørgensen A, Toftager M, Eghøj M, Ried-Larsen M, Bjørk Petersen C. Heart rate responses, agreement and accuracy among persons with severe disabilities participating in the indirect movement program: Team Twin—an observational study. Front Sports Act Living 2023;5:1213655. [doi: 10.3389/fspor.2023.1213655] [Medline: 37941848]
168. Leth S, Hansen J, Nielsen OW, Dinesen B. Evaluation of commercial self-monitoring devices for clinical purposes: results from the Future Patient Trial, phase I. Sensors (Basel) 2017 Jan 22;17(1):211. [doi: 10.3390/s17010211] [Medline: 28117736]
169. Støve MP, Graversen AH, Sørensen J. Assessment of noninvasive oxygen saturation in patients with COPD during pulmonary rehabilitation: smartwatch versus pulse oximeter. Respir Care 2023 Aug;68(8):1041-1048. [doi: 10.4187/respcare.10760] [Medline: 37193599]
170. Støve MP, Hansen ECK. Accuracy of the Apple Watch Series 6 and the Whoop Band 3.0 for assessing heart rate during resistance exercises. J Sports Sci 2022 Dec;40(23):2639-2644. [doi: 10.1080/02640414.2023.2180160] [Medline: 36803578]

XSL•FO
RenderX

171. Støve MP, Haucke E, Nymann ML, Sigurdsson T, Larsen BT. Accuracy of the wearable activity tracker Garmin Forerunner 235 for the assessment of heart rate during rest and activity. J Sports Sci 2019 Apr;37(8):895-901. [doi: 10.1080/02640414.2018.1535563] [Medline: 30326780]
172. Støve MP, Holm RS, Kjaersgaard AS, Duncker K, Jensen MR, Larsen BT. Measurement latency significantly contributes to reduced heart rate measurement accuracy in wearable devices. J Med Eng Technol 2020 Apr;44(3):125-132. [doi: 10.1080/03091902.2020.1753836] [Medline: 32404012]
173. Barrios L, Oldrati P, Santini S, Lutterotti A. Evaluating the accuracy of heart rate sensors based on photoplethysmography for in-the-wild analysis. Presented at: PervasiveHealth'19: The 13th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 20-23, 2019. [doi: 10.1145/3329189.3329215]
174. Gilgen-Ammann R, Roos L, Wyss T, et al. Validation of ambulatory monitoring devices to measure energy expenditure and heart rate in a military setting. Physiol Meas 2021 Aug 27;42(8). [doi: 10.1088/1361-6579/ac19f9] [Medline: 34340217]
175. Kozak D, Burgermeister S, De Chassey JDB, Naef A, Maringue A, Dietrich D. A functionality, safety and validity study of innovative REACH devices. Gerontechnology 2017 Nov 7;16(3):181-188. [doi: 10.4017/gt.2017.16.3.007.00]
176. Stucky B, Clark I, Azza Y, et al. Validation of Fitbit Charge 2 sleep and heart rate estimates against polysomnographic measures in shift workers: naturalistic study. J Med Internet Res 2021 Oct 5;23(10):e26476. [doi: 10.2196/26476] [Medline: 34609317]
177. Schweizer T, Gilgen-Ammann R. Wrist-worn and arm-worn wearables for monitoring heart rate during sedentary and light-to-vigorous physical activities: device validation study. JMIR Cardio 2025 Mar 21;9:e67110. [doi: 10.2196/67110] [Medline: 40116771]
178. Kinnunen H, Rantanen A, Kenttä T, Koskimäki H. Feasible assessment of recovery and cardiovascular health: accuracy of nocturnal HR and HRV assessed via ring PPG in comparison to medical grade ECG. Physiol Meas 2020 May 7;41(4):04NT01. [doi: 10.1088/1361-6579/ab840a] [Medline: 32217820]
179. Nuuttila OP, Korhonen E, Laukkanen J, Kyröläinen H. Validity of the wrist-worn Polar Vantage V2 to measure heart rate and heart rate variability at rest. Sensors (Basel) 2021 Dec 26;22(1):137. [doi: 10.3390/s22010137] [Medline: 35009680]
180. Sarhaddi F, Kazemi K, Azimi I, et al. A comprehensive accuracy assessment of Samsung smartwatch heart rate and heart rate variability. PLOS One 2022;17(12):e0268361. [doi: 10.1371/journal.pone.0268361] [Medline: 36480505]
181. Hermand E, Coll C, Richalet JP, Lhuissier FJ. Accuracy and reliability of pulse O2 saturation measured by a wrist-worn oximeter. Int J Sports Med 2021 Dec;42(14):1268-1273. [doi: 10.1055/a-1337-2790] [Medline: 34000752]
182. Bretonneau Q, Peruque-Gayou E, Wolfs E, Bosquet L. Accuracy of heart-rate-recovery parameters assessed from a wrist-worn photoplethysmography monitor (Polar Unite). Int J Sports Physiol Perform 2023 Jan 1;19(1):13-18. [doi: 10.1123/ijspp.2023-0120] [Medline: 37917971]
183. Bretonneau Q, Peruque-Gayou E, Wolfs E, Bosquet L. Parameters influencing the accuracy of a wrist photoplethysmography heart-rate monitor (Polar Unite) during exercise. Int J Sports Physiol Perform 2023 Apr 1;18(4):440-443. [doi: 10.1123/ijspp.2022-0288] [Medline: 36805933]
184. Ceugniez M, Devanne H, Hermand E. Reliability and accuracy of the Fitbit Charge 4 photoplethysmography heart rate sensor in ecological conditions: validation study. JMIR mHealth uHealth 2025 Jan 8;13:e54871. [doi: 10.2196/54871] [Medline: 39789790]
185. Tedesco S, Sica M, Ancillao A, Timmons S, Barton J, O'Flynn B. Accuracy of consumer-level and research-grade activity trackers in ambulatory settings in older adults. PLoS One 2019;14(5):e0216891. [doi: 10.1371/journal.pone.0216891] [Medline: 31112585]
186. Giggins OM, Doyle J, Sojan N, et al. Accuracy of wrist-worn photoplethysmography devices at measuring heart rate in the laboratory and during free-living activities. Presented at: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); Nov 1-5, 2021. [doi: 10.1109/EMBC46164.2021.9629522]
187. O'Grady B, Lambe R, Baldwin M, Acheson T, Doherty C. The validity of Apple Watch Series 9 and Ultra 2 for serial measurements of heart rate variability and resting heart rate. Sensors (Basel) 2024 Sep 26;24(19):6220. [doi: 10.3390/s24196220] [Medline: 39409260]
188. Müller AM, Wang NX, Yao J, et al. Heart rate measures from wrist-worn activity trackers in a laboratory and free-living setting: validation study. JMIR mHealth uHealth 2019 Oct 2;7(10):e14120. [doi: 10.2196/14120] [Medline: 31579026]
189. Gorny AW, Liew SJ, Tan CS, Müller-Riemenschneider F. Fitbit Charge HR wireless heart rate monitor: validation study conducted under free-living conditions. JMIR mHealth uHealth 2017 Oct 20;5(10):e157. [doi: 10.2196/mhealth.8233] [Medline: 29055881]
190. Ravanelli N, Lefebvre K, Brough A, Paquette S, Lin W. Validation of an open-source smartwatch for continuous monitoring of physical activity and heart rate in adults. Sensors (Basel) 2025 May 6;25(9):2926. [doi: 10.3390/s25092926] [Medline: 40363363]
191. Theurl F, Schreinlechner M, Sappler N, et al. Smartwatch-derived heart rate variability: a head-to-head comparison with the gold standard in cardiovascular disease. Eur Heart J Digit Health 2023 May;4(3):155-164. [doi: 10.1093/ehjdh/ztad022] [Medline: 37265873]

XSL•FO
RenderX

192. Neudorfer M, Kumar D, Smeddinck JD, et al. Validity of four consumer‐grade optical heart rate sensors for assessing volume and intensity distribution of physical activity. Scand J Med Sci Sports 2024 Nov;34(11). [doi: 10.1111/sms.14756] [Medline: 39508366]
193. Yang J, Lu Z, Li X, Tao K, Qiu J. Validity of Hera Leto consumer earbuds for heart rate monitoring during stationary cycling. Sci Rep 2025;15(1):15944. [doi: 10.1038/s41598-025-00559-5]
194. Rafl J, Bachman TE, Rafl-Huttova V, Walzel S, Rozanek M. Commercial smartwatch with pulse oximeter detects short-time hypoxemia as well as standard medical-grade device: validation study. Digit Health 2022;8:20552076221132127. [doi: 10.1177/20552076221132127] [Medline: 36249475]
195. Walzel S, Mikus R, Rafl-Huttova V, Rozanek M, Bachman TE, Rafl J. Evaluation of leading smartwatches for the detection of hypoxemia: comparison to reference oximeter. Sensors (Basel) 2023 Nov 14;23(22):9164. [doi: 10.3390/s23229164] [Medline: 38005550]
196. Haveman ME, van Rossum MC, Vaseur RME, et al. Continuous monitoring of vital signs with wearable sensors during daily life activities: validation study. JMIR Form Res 2022 Jan 7;6(1):e30863. [doi: 10.2196/30863] [Medline: 34994703]
197. Hoevenaars D, Yocarini IE, Paraschiakos S, et al. Accuracy of heart rate measurement by the Fitbit Charge 2 during wheelchair activities in people with spinal cord injury: instrument validation study. JMIR Rehabil Assist Technol 2022 Jan 19;9(1):e27637. [doi: 10.2196/27637] [Medline: 35044306]
198. Turki AF. An investigation into the accuracy of blood oxygen saturation measurement using smart watch devices among pediatric users in Saudi Arabia. Presented at: 2023 IEEE International Biomedical Instrumentation and Technology Conference (IBITeC); Nov 9-10, 2023. [doi: 10.1109/IBITeC59006.2023.10390911]
199. Khushhal AA, Mohamed AA, Elsayed ME. Accuracy of Apple Watch to measure cardiovascular indices in patients with chronic diseases: a cross sectional study. J Multidiscip Healthc 2024;17:1053-1063. [doi: 10.2147/JMDH.S449071] [Medline: 38496326]
200. Ho WT, Yang YJ, Li TC. Accuracy of wrist-worn wearable devices for determining exercise intensity. Digit Health 2022;8:20552076221124393. [doi: 10.1177/20552076221124393] [Medline: 36081752]
201. Chow HW, Yang CC. Accuracy of optical heart rate sensing technology in wearable fitness trackers for young and older adults: validation and comparison study. JMIR mHealth uHealth 2020 Apr 28;8(4):e14707. [doi: 10.2196/14707] [Medline: 32343255]
202. Arslan B, Sener K, Guven R, et al. Accuracy of the Apple Watch in measuring oxygen saturation: comparison with pulse oximetry and ABG. Ir J Med Sci 2024 Feb;193(1):477-483. [doi: 10.1007/s11845-023-03456-w] [Medline: 37440093]
203. Paslı S, Topçuoğlu H, Yılmaz M, Yadigaroğlu M, İmamoğlu M, Karaca Y. Diagnostic accuracy of Apple Watch ECG outputs in identifying dysrhythmias: a comparison with 12-lead ECG in emergency department. Am J Emerg Med 2024 May;79:25-32. [doi: 10.1016/j.ajem.2024.01.046] [Medline: 38330880]
204. Jaiswal P, Sahu NK, Lone HR. Comparative assessment of smartwatch photoplethysmography accuracy. IEEE Sens Lett 2024;8(1):1-4. [doi: 10.1109/LSENS.2023.3342292]
205. Imai H, Shimada J. Evaluation of telemetric single-lead cardiac transmitter and Apple Watch for heart rate monitoring: implications for heart failure management in home care. Cureus 2025 Mar;17(3):e80232. [doi: 10.7759/cureus.80232] [Medline: 40196089]
206. Ibrahim NS, Rampal S, Lee WL, Pek EW, Suhaimi A. Evaluation of wrist-worn photoplethysmography trackers with an electrocardiogram in patients with ischemic heart disease: a validation study. Cardiovasc Eng Tech 2024 Feb;15(1):12-21. [doi: 10.1007/s13239-023-00693-z] [Medline: 37973701]
207. Flett IL, Su Y, Cameron C, Zhou C, Holder-Pearson L, Chase JG. Assessing the accuracy of a wrist-worn wearable device's heart rate measurement over changing levels of activity. SSRN. Preprint posted online on Sep 10, 2025. [doi: 10.2139/ssrn.5412841]
208. Danielsson ML, Vergeer M, Plasqui G, Baumgart JK. Accuracy of the Apple Watch Series 4 and Fitbit Versa for assessing energy expenditure and heart rate of wheelchair users during treadmill wheelchair propulsion: cross-sectional study. JMIR Form Res 2024 May 7;8:e52312. [doi: 10.2196/52312] [Medline: 38713497]
209. Ferrera LE, Caballero JM, Kurniawan Y, Tedjasaputra JA, Sari E. Health quadrant mHealth application: a performance analysis of the heart rate accuracy vis a vis a pulse oximeter. In: Proceedings of the 4th International Conference on Human-Computer Interaction and User Experience in Indonesia, CHIuXiD '18: Association for Computing Machinery; 2018:54-60. [doi: 10.1145/3205946.3205954]
210. Jachymek M, Jachymek MT, Kiedrowicz RM, Kaźmierczak J, Płońska-Gościniak E, Peregud-Pogorzelska M. Wristbands in home-based rehabilitation-validation of heart rate measurement. Sensors (Basel) 2021 Dec 23;22(1):60. [doi: 10.3390/s22010060] [Medline: 35009603]
211. Sjöberg V, Westergren J, Monnier A, et al. Wrist-worn activity trackers in laboratory and free-living settings for patients with chronic pain: criterion validity study. JMIR mHealth uHealth 2021 Jan 12;9(1):e24806. [doi: 10.2196/24806] [Medline: 33433391]
212. Choksatchawathi T, Ponglertnapakorn P, Ditthapron A, et al. Improving heart rate estimation on consumer grade wrist-worn device using post-calibration approach. IEEE Sensors J 2020;20(13):7433-7446. [doi: 10.1109/JSEN.2020.2979191]

XSL•FO
RenderX

213. Maijala A, Kinnunen H, Koskimäki H, Jämsä T, Kangas M. Nocturnal finger skin temperature in menstrual cycle tracking: ambulatory pilot study using a wearable Oura Ring. BMC Womens Health 2019 Nov 29;19(1):150. [doi: 10.1186/s12905-019-0844-9] [Medline: 31783840]
214. Stahl AC, Tietz AS, Kendziora B, Dewey M. Has the STARD statement improved the quality of reporting of diagnostic accuracy studies published in European Radiology? Eur Radiol 2023 Jan;33(1):97-105. [doi: 10.1007/s00330-022-09008-7] [Medline: 35907025]
215. Plint AC, Moher D, Morrison A, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. Med J Aust 2006 Sep 4;185(5):263-267. [doi: 10.5694/j.1326-5377.2006.tb00557.x] [Medline: 16948622]
216. Fryar CD, Carroll MD, Afful J. Prevalence of Underweight Among Adults Aged 20 and Over: United States, 1960–1962 Through 2017–2018. : National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC); 2020 URL: https://www.cdc.gov/nchs/data/hestat/underweight-adult-17-18/ESTAT-Underweight-Adult-H.pdf [accessed 2026-07-15]
217. Waist circumference and BMI. Australian Bureau of Statistics. 2022. URL: https://www.abs.gov.au/statistics/health/health-conditions-and-risks/waist-circumference-and-bmi/latest-release [accessed 2026-07-15]
218. National Academies of Sciences, Engineering, and Medicine. Barriers to representation of underrepresented and excluded populations in clinical research. In: Improving Representation in Clinical Trials and Research: Building Research Equity for Women and Underrepresented Groups: The National Academies Press; 2022. URL: https://www.ncbi.nlm.nih.gov/books/NBK584407/ [accessed 2026-07-15]
219. Warburton DER, Gledhill N, Jamnik VK, et al. Evidence-based risk assessment and recommendations for physical activity clearance: Consensus Document 2011. Appl Physiol Nutr Metab 2011 Jul;36 Suppl 1:S266-S298. [doi: 10.1139/h11-062] [Medline: 21800945]
220. Vasudevan S, Vogt WC, Weininger S, Pfefer TJ. Melanometry for objective evaluation of skin pigmentation in pulse oximetry studies. Commun Med (Lond) 2024 Jul 11;4(1):138. [doi: 10.1038/s43856-024-00550-7] [Medline: 38992188]

Abbreviations

ANSI: American National Standards Institute
CTA: Consumer Technology Association
HRV: heart rate variability
PPG: photoplethysmography
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: PRISMA extension for scoping review
SpO_2: oxygen saturation

Edited by L Buis; submitted 07.Nov.2025; peer-reviewed by CA Sanches, SH Wu; revised version received 12.Jun.2026; accepted 12.Jun.2026; published 28.Aug.2026.

Please cite as:
Schipper RM, Oumar Djibrillah F, Roosink M, Winkens L, John A, Hazebroek E, Witteveen A, Berendsen AAM
The Representation of Different Populations in Studies Assessing the Validity of Consumer Wearable Photoplethysmography-Based Measurements: Scoping Review
JMIR Mhealth Uhealth 2026;14:e87324
URL: https://mhealth.jmir.org/2026/1/e87324
doi:10.2196/87324

© Rebecca M Schipper, Fatime Oumar Djibrillah, Meyke Roosink, Laura Winkens, Arlene John, Eric Hazebroek, Annemieke Witteveen, Agnes A M Berendsen. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 28.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effects of Self-Guided Digital Mindfulness Interventions on the Mental Health of Health Care Workers: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Natsu Sasaki[1], MD, PhD; Akiomi Inoue[2], PhD; Mako Iida[1], PhD; Kazuhiro Watanabe[3], PhD; Yuka Kobayashi[4], PhD; Asuka Sakuraya[5], PhD; Hisashi Eguchi[6], MD, PhD; Kanami Tsuno[7], PhD; Yu Komase[5,8], PhD; Yasumasa Otsuka[9], PhD; Mai Iwanaga[10], PhD; Reiko Inoue[3], MD; Kazuto Kuribayashi[11], PhD; Ayako Hino[6], MD, PhD; Hiroki Asaoka[1], PhD; Satoru Kanamori[12], PhD; Yuta Inagawa[13], MD, PhD; Takenori Yamauchi[14], PhD; Hiroki Ikeda[15], PhD; Go Muto[16], MD, PhD; Rika Kato[13], MD, PhD; Takuro Okuyama[12], BS; Naomichi Tani[17], PhD; Noriko Kojimahara[18], MD, PhD; Takeshi Ebara[17], PhD; Akihito Shimazu[19], PhD; Akizumi Tsutsumi[3], MD, PhD; Norito Kawakami[5], MD, PhD; Kotaro Imamura[5], PhD

[1]Department of Mental Health, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[10]Department of Community Mental Health and Law, National Institute of Mental Health, National Center of Neurology and Psychiatry, Tokyo, Japan

[11]Department of Psychiatric Nursing, Division of Nursing, Wayo Women's University, Ichikawa, Japan

[12]Graduate School of Public Health, Teikyo University, Tokyo, Japan

[13]Department of Psychiatry, Jichi Medical University, Shimotsuke, Japan

[14]Department of Hygiene, Public Health and Preventive Medicine, Showa Medical University, Tokyo, Japan

[15]National Institute of Occupational Safety and Health, Japan Organization of Occupational Health and Safety, Kawasaki, Japan

[16]Department of Hygiene and Public Health, Tokyo Women's Medical University School of Medicine, Tokyo, Japan

[17]Department of Ergonomics, Institute of Industrial Ecological Sciences, University of Occupational and Environmental Health, Japan, Kitakyushu, Japan

[18]Section of Epidemiology, Shizuoka Graduate University of Public Health, Shizuoka, Japan

[19]Faculty of Policy Management, Keio University, Kanagawa, Japan

[2]Institutional Research Center, University of Occupational and Environmental Health, Japan, Kitakyushu, Japan

[3]Department of Public Health, Kitasato University School of Medicine, Sagamihara, Japan

[4]Department of Clinical Psychology, Faculty of Social Policy and Administration, Hosei University, Tokyo, Japan

[5]Department of Digital Mental Health, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan

[6]Department of Mental Health, Institute of Industrial Ecological Sciences, University of Occupational and Environmental Health, Japan, Kitakyushu, Japan

[7]School of Health Innovation, Kanagawa University of Human Services, Kawasaki, Japan

[8]Healthcare Business Division, Fujitsu Japan Limited, Kawasaki, Japan

[9]Institute of Human Sciences, University of Tsukuba, Tokyo, Japan

Corresponding Author:
Kotaro Imamura, PhD
Department of Digital Mental Health, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan

Abstract

Background: Mindfulness interventions are considered an effective strategy for preventing mental health problems among health care workers (HCWs). While prior reviews have often included facilitator-guided or multicomponent interventions, it is unclear whether self-guided digital mindfulness interventions can also be effective without professional or in-person support.

Objective: This systematic review aimed to investigate the effects of self-guided digital mindfulness interventions on the mental health and well-being of HCWs.

Methods: This work was conducted as part of the DeLiGHT project. We searched electronic databases, including PubMed (MEDLINE), Embase, Cochrane CENTRAL, PsycINFO, PsycARTICLES, and the Japan Medical Abstract Society database, from inception in 2010 to July 25, 2023. To ensure timeliness, we performed an updated search on December 27, 2025. Database searching yielded 37,851 abstracts and 145 studies that examined the effectiveness of digital health interventions for workers in the initial 2023 screen. In the present study, we limited the included studies to those that met the following criteria: randomized

XSL•FO
RenderX

controlled trials (RCTs) using self-guided digital mindfulness interventions (eg, smartphone apps or web-based programs) delivered without face-to-face or professional support compared to waiting list controls or usual care among HCWs. The updated search in 2025 yielded 6328 additional records, of which 11 studies were examined. Study quality was assessed using the Cochrane Risk of Bias tool, with meta-analyses conducted using a random-effects model.

Results: A total of 9 RCTs with 3088 participants were included in the final analysis. The study populations were diverse, covering North America, Europe, Oceania, and Asia. Interventions were delivered via various digital platforms, including commercial mindfulness apps (n=3), social networking service tools (n=3), and web-based programs (n=3), with an average duration of 6.17 weeks (range: 1.5 - 18 wk). Five studies assessed outcomes only postintervention, and 4 included short-term follow-up (4 - 12 wk); no long-term effects were evaluated. The meta-analysis showed a significant beneficial effect on depression (standardized mean difference [SMD]=−0.44, 95% CI −0.88 to −0.00; *P*=.049), anxiety (SMD=−0.29, 95% CI −0.51 to −0.06; *P*=.011), perceived stress (SMD=−0.42, 95% CI −0.77 to −0.06; *P*=.02), and well-being (SMD=0.20, 95% CI 0.09 to 0.30; *P*<.001) immediately postintervention. The adherence rates reported across studies were highly variable, ranging from 19% to 56%. According to the Cochrane Risk of Bias tool, most studies were rated as having some concerns or a high risk of bias. The certainty of evidence included in the analysis ranged from “Low” to “Very Low.”

Conclusions: Overall, low to very low certainty evidence suggests that self-guided digital mindfulness interventions may have short-term beneficial effects on depression, anxiety, perceived stress, and well-being among HCWs. Future high-quality studies are needed to examine the long-term effects and cross-cultural generalizability of these interventions.

Trial Registration: UMIN UMIN000051631; https://center6.umin.ac.jp/cgi-open-bin/ctr/ctr_view.cgi?recptno=R000058861

(JMIR Mhealth Uhealth 2026;14:e74272) doi:10.2196/74272

KEYWORDS

digital intervention; occupational stress; psychological intervention; self-guided intervention; telehealth

Introduction

Health care workers (HCWs) are at high risk of experiencing mental health problems worldwide [1,2]. A systematic review has reported that the prevalence of mental health symptoms is high among HCWs: depression (13%), anxiety (16%), and traumatic stress (22%) [3]. Another systematic review reported that half of HCWs experience burnout [4], characterized primarily by emotional exhaustion and compassion fatigue, which negatively impacts patient care [5]. These mental health challenges affect not only the well-being of HCWs but also the quality of care provided to patients [6], potentially leading to long-term negative consequences on community health care systems. Therefore, effective and accessible interventions that focus on individual-level approaches, which are likely easier to implement, are urgently needed to improve the mental health of HCWs.

Mindfulness interventions are an effective nonpharmacological approach for improving mental health [7]. Mindfulness refers to paying attention to present-moment experiences in a nonjudgmental and accepting manner [8]. In particular, there is a growing interest in self-guided digital mindfulness interventions, which do not require face-to-face delivery or professional support, as scalable and accessible approaches. Systematic reviews have shown the effectiveness of face-to-face mindfulness interventions in improving the mental health of HCWs, including individuals with depression, anxiety [9,10], and emotional exhaustion [11], as well as in promoting their well-being [12]. Mindfulness is the most commonly used psychological group intervention for reducing distress symptoms in HCWs [13]. Mindfulness is an acceptable intervention for HCWs because it has a reasonable rationale and can be done quickly [14]. However, HCWs often face barriers when engaging in mental health interventions, such as limited availability (eg, shift work), high job demands, concerns about stigma, and negative social judgments, as well as issues related to disclosure and confidentiality, which are more common among HCWs than other workers [15]. In this context, self-guided digital interventions, which are not provided by professionals or delivered in person, may have advantages in engaging HCWs in the interventions.

Digital mindfulness interventions can overcome the barriers of logistics and time availability by providing flexible access and standardized content at a low cost [16]. Self-guided digital mindfulness tools have already been widely used, and over 560 mindfulness apps are available on the market [17]. However, the effectiveness of self-guided digital mindfulness interventions on the mental health of HCWs is still unclear. A previous systematic review reported on the effectiveness of web-based mindfulness-based interventions for HCWs. However, it focused primarily on the COVID-19 pandemic, with half of the included studies conducted in a single country, leading to a risk of overestimation and limited generalizability [18].

It did not restrict itself to self-guided interventions and included nonmindfulness components such as cognitive behavioral therapy, music therapy, and positive psychology interventions. Additionally, another review included studies involving students, searched only a single database, and did not conduct any quantitative analyses of the effects of web-based tools and mobile apps [19]. Considering advancements in technology and synthesizing the findings of randomized controlled trials (RCTs) conducted over the past 10 years with participants limited to HCWs, and not restricted to the pandemic period, may provide a more comprehensive and accurate understanding of the effectiveness of digital mindfulness interventions on the mental health of HCWs. If self-guided digital mindfulness interventions are found to be effective, such interventions may be promoted

XSL•FO
RenderX

even in emergency situations like pandemics and in low-resource settings.

This systematic review and meta-analysis thus aimed to address this gap by examining the effectiveness of self-guided digital mindfulness interventions without face-to-face, professional, or personal support components on mental health and well-being among HCWs from 2010 to 2025.

Methods

Study Design

This work was part of an overarching large-scale systematic review project developing Minds-compliant guidelines for general preventive interventions using digital health technologies for mental health (the DeLiGHT project; [20]), funded by the Japan Agency for Medical Research and Development.

The review protocol has been registered in the UMIN registry (registration number: UMIN000051631). The method was reported according to PRISMA (the Preferred Reporting Items for Systematic Review and Meta-Analysis) 2020 guidelines (Checklist 1) [21]. Among the included literature in the DeLiGHT project, this review reported on the effectiveness of fully automated digital mindfulness interventions for HCWs.

Data Sources, Search Strategy, and Study Selection Process

In the DeLiGHT project, we initially searched for publications from 2010 to July 25, 2023, across databases such as PubMed (MEDLINE), Embase, Cochrane CENTRAL, PsycINFO, PsycARTICLES, and the Japan Medical Abstracts Society database. The search was limited to studies published from 2010 onward, reflecting the widespread adoption of smartphones and web-based platforms. The search terms are detailed in Multimedia Appendix 1. Search terms were comprehensively selected, referring to previous reports on digital health technology [22] and outcomes [23]. The search strategy was constructed using combinations of population- and intervention-related terms. In addition, study design-related words were applied to restrict the search to randomized controlled trials. Although outcome-related terms were included to improve feasibility given the broad scope of the literature, we acknowledge that the inclusion of outcome terms may increase the risk of missing relevant studies, in line with the recommendations of the Cochrane Handbook [24,25]. We included all studies that followed the inclusion criteria of participants, interventions, comparisons, and outcomes (PICO): (P) any type of employee or worker; (I) any type of nonpharmacological intervention aiming at primary prevention using digital health technologies (eg, exercises, diet, lifestyle, and psychological interventions such as cognitive behavioral therapy); (C) any type of control condition; and (O) mental health conditions, positive mental health, and work-related outcomes. In the subgroup considerations, HCWs were defined using terms explicitly referring to medical and health professionals (eg, physicians, nurses, pharmacists, psychologists, therapists, and health care providers). In addition, studies using an RCT design and published after 2010 were included. The language was limited to English or Japanese. Exclusion criteria were as follows: (1) studies targeting participants who have any specific disorder, (2) studies including participants who are not employees or workers, (3) studies applying any interventions aimed at secondary or tertiary prevention, and (4) studies aiming to evaluate the treatment effects of interventions on specific disorders or symptoms. One investigator (M Iida) removed duplicate entries, and the remaining articles were shared with all co-authors. During the first screening phase, based on the eligibility criteria, each article's title and abstract were independently assessed by pairs of investigators, with one investigator being a staff member from an external contracting agency. Following this, 2 investigators independently reviewed the full texts of studies that passed the first screening. Any disagreements were resolved through discussion and consensus among all authors. Reasons for study exclusion were documented during the full-text review phase. Microsoft Excel was used to manage all identified studies.

The DeLiGHT project's database search initially retrieved 37,851 abstracts (from 2010 to July 25, 2023). During the initial screening, after removing 15,485 duplicates and 3275 older studies published before 2010, 19,091 records remained for further review, of which 18,806 were excluded. Subsequently, 285 records were forwarded for full-text screening. Following the full-text review, 140 studies were excluded. Finally, 145 RCTs were identified as digital interventions aimed at improving workers' mental health.

An updated search was conducted on December 27, 2025, in PubMed (MEDLINE), Cochrane CENTRAL, PsycINFO, PsycARTICLES, and the Japan Medical Abstracts Society database, using the same search terms (Multimedia Appendix 1). Notably, Embase could not be used for the updated search due to access restrictions. However, even in the original search, which included Embase, all studies ultimately included in this review were indexed in PubMed. Moreover, previous research has suggested that the additional benefit of supplementing PubMed with Embase may be limited in some fields [26]. Taken together, these considerations suggest that the absence of Embase is unlikely to have materially affected the comprehensiveness of the update. The search was limited to studies published from 2023 onward, to update the search of the DeLiGHT project. In the 2025 update, eligible RCTs were directly identified in accordance with the selection criteria of this study, which are described in the next section.

Selection Criteria for the Review of Digital Mindfulness Interventions

In this study, we limited the included studies according to the criteria for this review: (P) HCWs, including workers in clinical hospitals or emergency units; (I) digital mindfulness interventions; and (O) mental health symptoms (ie, depression, anxiety, perceived stress, traumatic stress reaction, and sleep quality), emotional exhaustion, compassion fatigue, and positive mental health and well-being. Digital mindfulness interventions in this review were defined as technology-assisted mindfulness interventions aimed at fostering greater attention to and awareness of present-moment experiences [7,27], delivered via digital platforms without face-to-face components or professional or personal support, as described in previous

XSL•FO
RenderX

reviews of digital mindfulness interventions [28,29]. Following this definition, we excluded interventions primarily focused on other approaches (eg, cognitive behavioral therapy-based skills) or interventions that could not be evaluated for the extent to which they emphasized mindfulness. Complex or multimodal interventions were carefully discussed in terms of eligibility by NS, AI, and KI. For multimodal interventions, eligibility was determined based on whether mindfulness was described as the primary intervention component in the intervention rationale and content descriptions, with decisions made through independent review and consensus among the reviewers. The yoga component was included in this review if it was delivered via a digital tool without professional support, as yoga has been recognized as a primary discipline and practice with the potential to cultivate mindfulness [30].

Data Collection Process and Data Items

Among the 145 records identified in the initial screening (DeLiGHT project, from 2010 to July 25, 2023), 3 investigators (NS, AI, and KI) assessed the studies' eligibility in pairs. In the update, 2 investigators (NS and KI) conducted both the initial and full-text reviews in accordance with the current studies' eligibility criteria (ie, mindfulness for HCWs). They consolidated the data, which included information such as publication year, authors' names, study location, participants' characteristics (eg, HCWs), number of participants, intervention details, control group information, outcome variables, and adherence or dose. Follow-up assessments were defined as outcome measurements conducted after the immediate postintervention assessment, and were used to evaluate the short-term sustainability of intervention effects on the same mental health outcomes.

Risk of Bias Assessment

Six investigators assessed the risk of bias independently in 3 pairs (NS, AI, M Iida, KW, YKob, and KI) using the recently revised Cochrane risk-of-bias tool for randomized trials (RoB 2.0) [31]. This tool comprises 5 bias domains: bias arising from the randomization process (domain 1), bias due to deviations from intended interventions (domain 2), bias due to missing outcome data (domain 3), bias in measuring the outcome (domain 4), and bias in the selection of the reported result (domain 5). For each domain, a series of signaling questions with answers (yes, probably yes, no information, probably no, and no) determine the risk of bias (low risk, some concerns, and high risk). Any discrepancies in the quality assessment were addressed through discussion and consensus among the authors.

Synthesis of Results and Meta-Analysis

In the analysis, we synthesized all of the included interventions and outcomes. A meta-analysis was performed if at least 3 eligible studies had the same outcome. The effect size was calculated based on the difference in the change of baseline scores between the intervention and control. The standardized mean difference (SMD) was used to assess the effect of interventions. Meta-analyses were performed for the mean effect on each outcome of mental health symptoms (eg, depression, anxiety, and perceived stress), emotional exhaustion and compassion fatigue, and each outcome of positive mental health and well-being. Meta-analysis was performed at each time point (immediately postintervention and follow-up) for each outcome. A random-effects model was used to consider differences in the treatment effect from study to study. When the SD of change from baseline was not shown, it was calculated based on the assumption that the correlation between scores at baseline and those at follow-up was 0.5 [32]. Heterogeneity between studies was measured through the I^2 statistic (low, moderate, and high heterogeneity corresponding to I^2 values of 25%, 50%, and 75%, respectively) [33]. Funnel plots and Egger tests were used to assess potential publication bias for the meta-analyses [34,35]. Consistent with the Cochrane Handbook [36], tests for funnel plot asymmetry were interpreted with caution because fewer than 10 studies were included, and statistical power is known to be low under such conditions. For outcomes with I^2 more than 75%, we performed sensitivity analyses by excluding outliers identified through visual inspection of funnel plots.

We used the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach to assess the certainty of the evidence [37].

Statistical significance was set at a 2-sided $P<.05$. Analysis was conducted using IBM SPSS Statistics Version 28.0. (Japanese version). Meta-analysis was performed in R version 4.5.1 (R Foundation for Statistical Computing) with the *metafor* package [38].

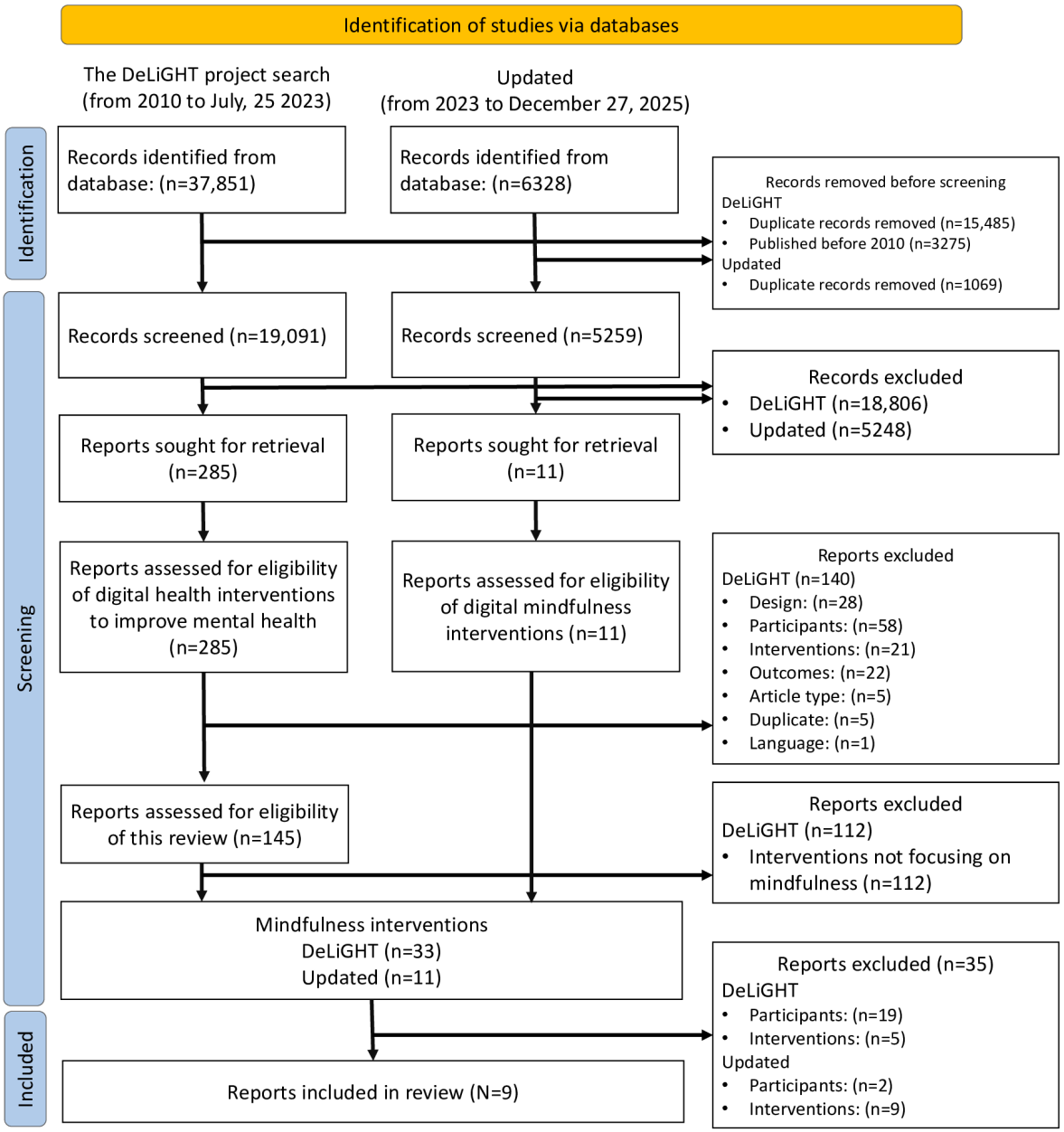
Results

Database Searching

Among the 145 studies identified in the DeLiGHT project, we excluded 112 articles because their interventions did not focus on mindfulness. Among 33 mindfulness intervention studies for workers, we excluded 24 studies for reasons such as participants (ie, not targeting HCWs; n=19) and interventions (ie, digital mindfulness without personal or in-person support; n=5). An updated database search retrieved 6328 abstracts (from 2023 to December 27, 2025), and 5259 abstracts were screened after removing duplicate records. Subsequently, 11 studies were assessed for eligibility for digital mindfulness interventions, but all studies were excluded. Finally, 9 studies [39-47] were included in the systematic review. The study selection process is illustrated in Figure 1. A list of 35 studies excluded at full-text screening, along with reasons for exclusion, is provided in Multimedia Appendix 2.

XSL•FO
RenderX

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 flow diagram of the systematic review search results. DeLiGHT: Developing minds-compliant guidelines for general preventive intervention using digital health technologies for mental health.



Study Characteristics

The characteristics of 9 RCTs are presented in Table 1. The total number of included participants was 3088 (ranging from 41 to 2182 in 1 study) [39-47]. Countries where studies were conducted varied, including Western countries (the United States or Canada [42,45]), England [46], Australia [47], Iran [39,44], and other Asian countries (South Korea [40], Singapore [41], and Thailand [43]). Of the 9 included studies, 3 studies [40,42,46] delivered the intervention before the COVID-19 pandemic.

XSL•FO
RenderX

Table . The details of the included randomized controlled trials (N=9).[a]

First author, year	Intervention year	Country	Participants (total)	Intervention (1-7)[b]	Control	Post FU[c] (from postsurvey)	Outcomes	Adherence or dose
Asadollah et al [39], 2023	• No information but under COVID-19 pandemic	• Iran	• NICU[d] nurses in 2 educational hospitals (N=66)	• 6 weeks • Reduction of compassion fatigue • "Loving-kindness meditation" (a state of self-awareness and repeating phrases to reinforce a positive attitude with compassion for oneself and others) • At least 3 times a week • App (Whatsapp, where audio files for the meditation were distributed) • Participants were organized into 2 WhatsApp groups for interaction with the researchers and fellow participants • N/A[e]	• Miscellaneous files on mental health during the COVID-19 pandemic period	• Post 6 week	• MH[f] symptoms and disorders: N/A • Emotional exhaustion and compassion fatigue compassion fatigue (NCFI[g]) • Positive MH and well-being: N/A	N/A[h]

XSL•FO
RenderX

First author, year	Intervention year	Country	Participants (total)	Intervention (1-7)[b]	Control	Post FU[c] (from postsurvey)	Outcomes	Adherence or dose
Hwang and Jo [40], 2019	• 2018	• South Korea	• Nurses employed at college hospitals (N=56)	• 4 weeks • Stress management • Meditation, breathing methods, and yoga • More than twice per week for at least 10 minutes each session • App • N/A • Music focused on healing, health information for mental health care every week (diet, benefits of exercise, and so on)	• Waitlist	• Post 4 week	• MH symptoms and disorders: Perceived stress (PSS-10), depression (PHQ[i]-9), anxiety (GAD-7) • Emotional exhaustion and compassion fatigue: N/A • Positive MH and well-being: well-being (WHO-5[j]), self-efficacy (Likert scale)	N/A
Keng et al [41], 2022	• No information but under COVID-19 pandemic	• Singapore	• Health care workers (N=80)	• 1.5 weeks • Reduction of mental health problems during the COVID-19 pandemic • Mindful breathing, mindfulness of thoughts, and mindfulness of sounds • 10 minutes per day • App ("Headspace") • N/A • N/A	• Active control ("Lumosity" a daily training component of 3 free games involving problem solving, memory, and attention, totaling 10 minutes, with a 3-week daily practice)	• Post 3 weeks • FU 4 weeks		• 56.41% (22/39) completed ≥80%, 30.76% completed 40% - 80%, and 12.8% completed 20% - 40% of the required practice duration

XSL•FO
RenderX

First author, year	Intervention year	Country	Participants (total)	Intervention (1-7)[b]	Control	Post FU[c] (from postsurvey)	Outcomes	Adherence or dose
							• MH symptoms and disorders: depression and anxiety (DASS-21[k]), PTSD[l] symptoms (PCL-C[m]), fear of COVID-19 (FCV-19S[n]), sleep quality (PSQI[o]) • Emotional exhaustion and compassion fatigue: N/A • Positive MH and well-being: Personal well-being (PWI[p])	
Lilly et al [42], 2019	• 2016-2017	• United States or Canada	• Emergency medical dispatchers (telecommunicators; N=323)		• Waitlist	• Post 7 weeks • FU 12 weeks	• MH symptoms and disorders: stress symptoms (C-SOSI[r]) • Emotional exhaustion and compassion fatigue: N/A • Positive MH and well-being: N/A	• 89/161 (55.3%) completed 6 or 7 modules. Mean number of days/wk: 2.1 (SD=1.5, range 0 - 6.8).

XSL•FO
RenderX

First author, year	Intervention year	Country	Participants (total)	Intervention (1-7)[b]	Control	Post FU[c] (from postsurvey)	Outcomes	Adherence or dose
				• 7 weeks • Reduction of work-related stress for emergency medical dispatchers • 7 modules MBSR[q]-based short videos, text descriptions, and audio-guided meditation exercises • 20 - 30 minutes per module per week, with additional daily practice encouraged (5 - 10 min) • Web ("Destress 9-1-1") • Moderated discussion board • N/A				
Luangapichart et al [43], 2022	• 2021	• Thailand	• Medical personnel with moderate or high levels of burnout and stress from across Thailand (N=90)		• Active control (PSA only)	• Post 4 weeks • FU 4 weeks		• 22.2% listened to the audio files more than 3 times per day for over 60% of the total 4 weeks.

XSL•FO
RenderX

First author, year	Intervention year	Country	Participants (total)	Intervention (1-7)[b]	Control	Post FU[c] (from postsurvey)	Outcomes	Adherence or dose
							• MH symptoms and disorders: stress (ST[s]-5), anxiety (HADS[t]), and depression (HADS) • Emotional exhaustion and compassion fatigue: N/A • Positive MH and well-being: N/A	

XSL•FO
RenderX

First author, year	Intervention year	Country	Participants (total)	Intervention (1-7)[b]	Control	Post FU[c] (from postsurvey)	Outcomes	Adherence or dose
				• 4 weeks • Reduction of burnout and stress • Mindfulness (by using body sensations, surrounding sounds, and front images) • At least 3 times daily • App ("Mindful Senses" an online audio-based program via the LINE application) • Participants were able to send inquiries regarding mindfulness practice to the therapist via the LINE application at any time during the program and received a reply within 1 day • Psychological self-help articles (PSAs) on burnout, stress management, relationship management, and men-				

XSL•FO
RenderX

First author, year	Intervention year	Country	Participants (total)	Intervention (1-7)[b]	Control	Post FU[c] (from postsurvey)	Outcomes	Adherence or dose
				tal health promotion				
Nourian et al [44], 2021	• 2020	• Iran	• Nurses working in the COVID-19 care units (N=41)	• 7 weeks • Reduction of stress • MBSR, including meditations and yoga exercises • At least 1 hour per day, 6 days a week • App (WhatsApp Messenger) • Group and individual guidance was provided through 2-person chats and short messages to resolve any issues related to the exercises for that week • N/A	• Music files or training items on caring for patients with COVID-19	• Post 7 weeks	• MH symptoms and disorders: sleep quality (PSQI) • Emotional exhaustion and compassion fatigue: N/A • Positive MH and well-being: N/A	N/A
Pratt et al [45], 2023	• 2021-2022	• United States	• Nurses working in COVID-19 units at a single hospital (N=102)		• Waitlist	• Post 4 weeks		• 27.5% completed at least one mindfulness session per week, 18.8% completed ≥75% of mindfulness sessions

XSL•FO
RenderX

First author, year	Intervention year	Country	Participants (total)	Intervention (1-7)[b]	Control	Post FU[c] (from postsurvey)	Outcomes	Adherence or dose
				• 4 weeks • Reduction of psychological distress and burnout • Mindfulness ("Lift") • Daily • Mobile app • If participants reported suicidal ideation, they were immediately contacted by phone and connected to mental health service providers • N/A			• MH symptoms and disorders: depression (PHQ-9), anxiety (GAD[u]-7), and stress (PSS-4) • Emotional exhaustion and compassion fatigue • Emotional exhaustion (MBI[v]) • Positive MH and well-being: N/A	
Taylor et al [46], 2022	• 2017-2018	• England	• National Health Service staff (N=2182)		• Active control ("NHS[w] Mood-zone," a psychoeducational digital platform offering evidence-based advice and resources on managing work-related stress and mental health, including videos, audio, and podcasts, with 10 min daily)	• Post 18 weeks	• MH symptoms and disorders: depression, anxiety, and stress (DASS) • Emotional exhaustion and compassion fatigue • Emotional exhaustion (MBI) • Positive MH and well-being: Well-being (SWEMWBS[x])	• Average 3.5 days per week; 66.6% practiced at least 3 days a week

XSL•FO
RenderX

First author, year	Intervention year	Country	Participants (total)	Intervention (1-7)[b]	Control	Post FU[c] (from postsurvey)	Outcomes	Adherence or dose
				• 18 weeks • Reduction of stress, mental health, and work-related outcomes • Brief mindfulness-based practices alongside psychoeducational materials 4. Completing the Take 10 introductory pack (guided 10-min mindfulness practices daily for 10 consecutive days) and at least one 10-minute mindfulness practice daily for the duration of the study • Website or App ("Headspace") • N/A • N/A				
Xu et al [47], 2022	• 2019-2020	• Australia	• Staff in 2 emergency departments (N=148)		• Waitlist	• Post 4 weeks • FU 12 weeks		• Most frequent use was 1 - 3 times per week (41%) of 2 - 10 min sessions (68%)

XSL•FO
RenderX

First author, year	Intervention year	Country	Participants (total)	Intervention (1-7)[b]	Control	Post FU[c] (from postsurvey)	Outcomes	Adherence or dose
				• 4 weeks • Reduction of emergency department work stress • Guided mindfulness meditation with access to over 50 sessions ranging from 3 to 20 minutes • Daily practice of 10 minutes, with the option to choose preferred sessions • App ("Headspace") • N/A • N/A			• MH symptoms and disorders: stress (PSS[y]); emotional exhaustion and compassion fatigue: emotional exhaustion (MBI); positive MH and well-being: well-being (WEMWBS[z])	

[a]All studies allocated participants to the intervention group and the control group in a 1:1 ratio.

[b](1) Intervention period, (2) program objective, (3) content, (4) intensity, (5) platform, (6) interpersonal support, and (7) other elements besides mindfulness.

[c]FU: follow-up.

[d]NICU: neonatal intensive care unit.

[e]N/A: not assessed.

[f]MH: mental health.

[g]NCFI: Nurses' Compassion Fatigue Inventory.

[h]Not applicable.

[i]PHQ: Patient Health Questionnaire.

[j]WHO-5: World Health Organization-5 Well-Being Index.

[k]DASS: Depression Anxiety Stress Scales.

[l]PTSD: posttraumatic stress disorder.

[m]PCL-C: Post-Traumatic Stress Disorder Checklist–Civilian Scale.

[n]FCV-19S: Fear of COVID-19 Scale.

[o]PSQI: Pittsburgh Sleep Quality Index.

[p]PWI: Personal Wellbeing Index.

[q]MBSR: mindfulness-based stress reduction.

[r]C-SOSI: Calgary Symptoms of Stress Inventory.

[s]ST: Stress Test Questionnaire.

[t]HADS: Hospital Anxiety and Depression Scale.

[u]GAD: Generalized Anxiety Disorder.

[v]MBI: Maslach Burnout Inventory.

[w]NHS: National Health Service.

XSL•FO
RenderX

[x]SWEMWBS: Short Warwick Edinburgh Mental Well-Being Scale.

[y]PSS: Perceived Stress Scale.

[z]WEMWBS: Warwick Edinburgh Mental Well-Being Scale.

Characteristics of Interventions

The average intervention period was 6.17 weeks, ranging from 1.5 weeks [41] to 18 weeks [46]. Regarding the platforms used:

- Commercial apps: three RCTs [41,46,47] used interventions involving a commercial mindfulness app ("Headspace").
- Social networking service (SNS) tools: three RCTs [39,43,44] used interventions involving SNS messaging platforms.
- Other platforms: others used original apps [40,45] and a web-based program [42].

Regarding follow-up assessments, 5 studies [39,40,44-46] measured outcomes only immediately postintervention, and 4 [41-43,47] measured outcomes twice: immediately postintervention and at follow-up (ranging from 4 to 12 wk after the postintervention period). No studies evaluated long-term effectiveness beyond this timeframe.

Outcome Measures and Data Synthesis

The included studies assessed a broad range of psychological outcomes. Perceived stress was the most frequently measured outcome (n=6) [40,42,43,45-47], followed by depression and anxiety (n=5) [40,41,43,45,46]. Other measures included emotional exhaustion and compassion fatigue (n=4) [39,45-47], as well as positive mental health constructs such as well-being (n=4) [40,41,46,47].

Certain outcomes, including sleep quality (n=2) [41,44], self-efficacy (n=1) [40], symptoms of PTSD (n=1) [41], and fear of COVID-19 (n=1) [41], were excluded from the meta-analysis due to an insufficient number of studies for robust synthesis.

Adherence to the prescribed dose was reported in 6 out of 9 studies. Program completion rates were documented in 3 studies [41,42,45]. Among these, 56.4% (22/39) of participants completed more than 80% of the required practice duration in one study [41], and 55.3% (89/161) completed 6 or 7 modules in another [42]. In contrast, 1 study reported that 19% (13/69) of participants completed more than 75% of the daily sessions [45]. The average frequency of weekly usage was reported in 5 studies [42,43,45-47]. Specifically, 28% (19/69) of participants completed at least 1 session per week [45], 41% (29/71) used the program 1 to 3 times per week [47], and the average weekly usage was 2.1 (SD 1.5) days [42]. In the Taylor et al [46] study, 66.6% (452/679) of participants practiced 3 or more days per week. Furthermore, in the Luangapichart et al [43] study, 22.2% (10/45) of participants used the program 4 or more days per week, with 3 or more sessions per day.

Risk of Bias

Table 2 presents the risk of bias assessed using RoB 2.0. All studies were evaluated as having a high risk of bias. In many studies, domains 1, 2, and 5 were rated as low or as having some concerns; however, domains 3 and 4 were frequently rated as high. Consequently, the overall bias across all studies was determined to be high.

XSL•FO
RenderX

Table . Summary of the risk of bias assessment for the included studies.

First author, year	Domain 1 (R[a])	Domain 2 (D[b])	Domain 3 (Mi[c])	Domain 4 (Me[d])	Domain 5 (S[e])	Overall
Outcome: depression (meta-analysis conducted)						
Hwang and Jo [40], 2019	?	+[f]	–	+	?[g]	–[h]
Keng et al [41], 2022	?	?	+	–	?	–
Luangapichart et al [43], 2022	+	?	+	–	+	–
Pratt et al [45], 2023	?	?	–	–	+	–
Taylor et al [46], 2022	+	?	–	–	?	–
Outcome: anxiety (meta-analysis conducted)						
Hwang and Jo [40], 2019	?	+	–	+	?	–
Keng et al [41], 2022	?	?	+	–	?	–
Luangapichart et al [43], 2022	+	?	+	–	+	–
Pratt et al [45], 2023	?	?	–	–	+	–
Taylor et al [46], 2022	+	?	–	–	?	–
Outcome: perceived stress (meta-analysis conducted)						
Hwang and Jo [40], 2019	?	+	–	+	?	–
Lilly et al [42], 2019	?	?	–	–	?	–
Luangapichart et al [43], 2022	+	?	+	–	+	–
Pratt et al [45], 2023	?	?	–	–	+	–
Taylor et al [46], 2022	+	?	–	–	?	–
Xu et al [47], 2022	+	?		–	+	–
Outcome: emotional exhaustion and compassion fatigue (meta-analysis conducted)						
Asadollah et al [39], 2023	–	–	+	–	?	–
Pratt et al [45], 2023	?	?	–	–	+	–
Taylor et al [46], 2022	+	?	–	–	?	–
Xu et al [47], 2022	+	?	–	–		–
Outcome: positive mental health and well-being (meta-analysis conducted)						
Hwang and Jo [40], 2019	?	+	–	+	?	–
Keng et al [41], 2022	?	?	+	–	?	–
Taylor et al [46], 2022	+	?	–	–	?	–

XSL•FO
RenderX

First author, year	Domain 1 (R[a])	Domain 2 (D[b])	Domain 3 (Mi[c])	Domain 4 (Me[d])	Domain 5 (S[e])	Overall
Xu et al [47], 2022	+	?	–	–	+	–
Outcome: PTSD (meta-analysis not conducted)						
Keng et al [41], 2022	?	?	+	–	?	–
Outcome: fear of COVID-19 (meta-analysis not conducted)						
Keng et al [41], 2022	?	?	+	–	?	–
Outcome: sleep quality (meta-analysis not conducted)						
Keng et al [41], 2022	?	?	+	–	?	–
Nourian et al [44], 2021	?	–	–	–	?	–
Outcome: self-efficacy (meta-analysis not conducted)						
Hwang and Jo [40], 2019	?	+	–	+	?	–

[a]R: bias arising from the randomization process.

[b]D: bias due to deviations from intended interventions.

[c]Mi: bias due to missing outcome data.

[d]Me: bias in measurement of the outcome.

[e]S: bias in selection of the reported result.

[f]+: low risk of bias.

[g]?: some concerns.

[h]-: high risk of bias.

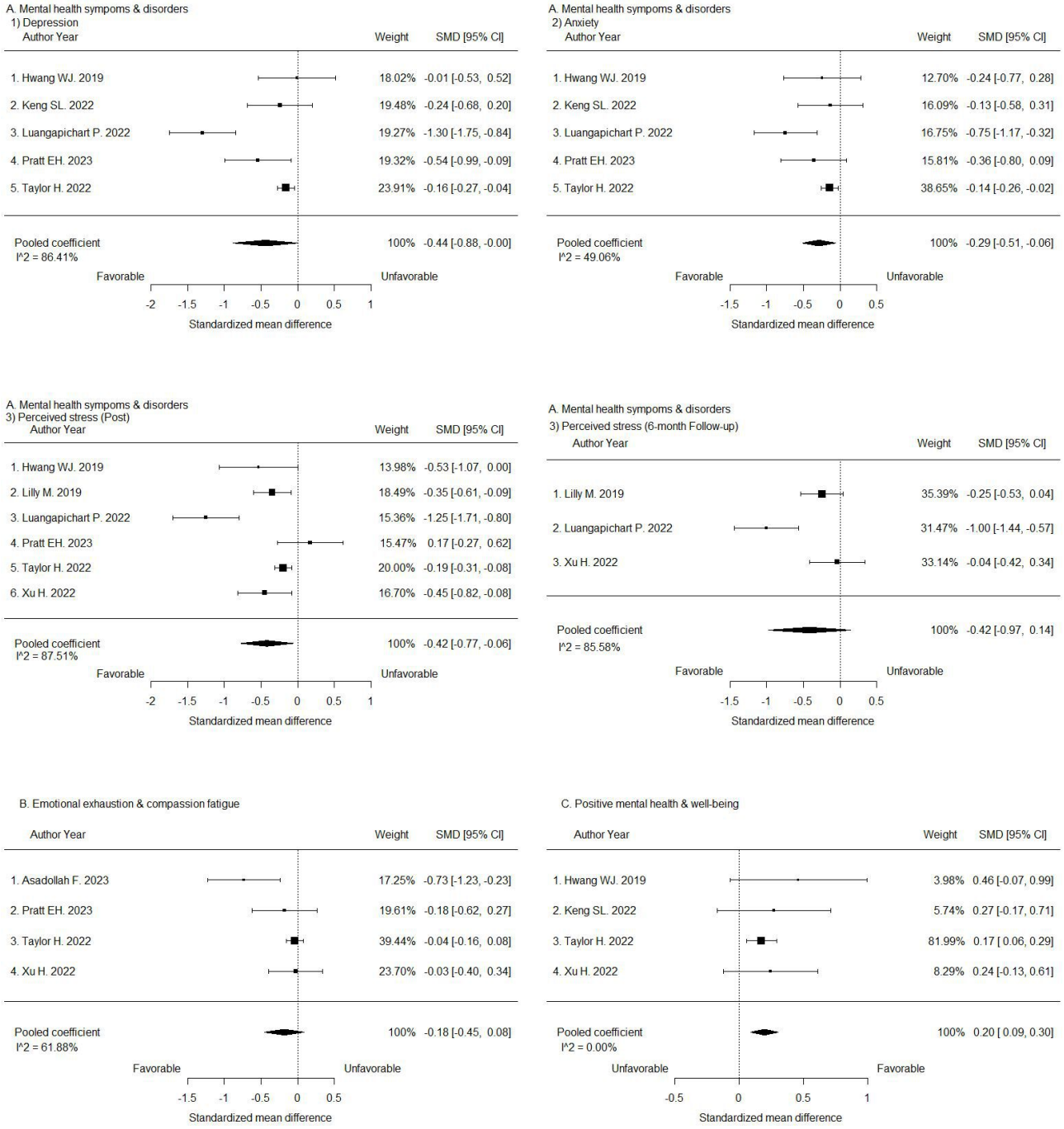
Meta-Analysis

The results of the meta-analysis are shown in Figure 2. Digital mindfulness interventions demonstrated significant beneficial effects in improving the postintervention scores of individuals with depression (k=5, SMD=-0.44, 95% CI −0.88 to −0.003; *P*=.049), anxiety (k=5; SMD=−0.29, 95% CI −0.51 to −0.06; *P*=.011), perceived stress (k=6; SMD=−0.42, 95% CI −0.77 to −0.06; *P*=.021), and well-being (k=4; SMD=0.20, 95% CI 0.09 to 0.30; *P*<.001). No significant effectiveness was observed in addressing emotional exhaustion or compassion fatigue. Additionally, no long-term efficacy was identified for perceived stress during the follow-up periods, which ranged from 4 to 12 weeks after the postintervention period. The heterogeneity varied from low to high (I^2=0% in well-being, 50% - 60% in anxiety and emotional exhaustion or compassion fatigue, and over 80% in depression and perceived stress).

Sensitivity analyses excluding Luangapichart et al [43], which reported a notably larger intervention effect, significantly reduced heterogeneity for depression (k=4; SMD=−0.18, 95%CI −0.28 to −0.07; *P*=.001; I^2=0%) and postintervention perceived stress (k=5; SMD=−0.25, 95% CI −0.40 to −0.10; *P*<.001; I^2=28.1%).

XSL•FO
RenderX

Figure 2. Summary of meta-analysis [39-43,45-47]. (A) Mental health symptoms and disorders: (1) depression and (2) anxiety [40, 41, 43, 45, 46], (3) perceived stress (post) [40, 42, 43, 45, 46, 47], (4) perceived stress (6-mo follow-up) [42, 43, 47]. (B) Emotional exhaustion and compassion fatigue [39, 45, 46, 47]. (C) Positive mental health and well-being [40, 41, 46, 47].



Publication Bias

Multimedia Appendix 3 presents funnel plots of the included studies. We qualitatively and quantitatively assessed publication bias by visually inspecting funnel plot asymmetry. The visual inspection suggested symmetry. Egger tests were not significant for all outcomes. Although no clear evidence of funnel plot asymmetry was observed, these findings should be interpreted cautiously due to the limited number of included studies.

Certainty of Evidence

The GRADE evidence profile is presented in Multimedia Appendix 4. The certainty of evidence for the outcomes included in the analysis ranged from “Low” to “Very Low.”

Discussion

Principal Findings

Overview

This systematic review and meta-analysis evaluated fully automated digital mindfulness interventions for HCWs, ultimately including 9 RCTs with 3088 participants. Adherence

XSL•FO
RenderX

and dose varied widely across the included studies. Results showed that these interventions had moderate, beneficial effects on depression (SMD=−0.44, 95% CI −0.88 to −0.003), anxiety (SMD=−0.29, 95% CI −0.51 to −0.06), and perceived stress (SMD=−0.42, 95% CI −0.77 to −0.06) immediately postintervention, as well as a slight improvement in well-being (SMD=0.20). However, the interventions did not significantly impact emotional exhaustion or compassion fatigue, and there was no evidence of sustained effects on perceived stress during the follow-up period, which ranged from 4 to 12 weeks after the postintervention period. Heterogeneity across studies was low in well-being but moderate to high in other outcomes.

Effects on Depression, Anxiety, and Perceived Stress

Digital mindfulness intervention was effective in improving depression, anxiety, and perceived stress among HCWs. This result was consistent with a previous report that workplace mindfulness interventions improved mental health [9]. While previous meta-analyses of workplace mindfulness interventions, which predominantly included face-to-face or facilitator-guided programs, reported large pooled effects on composite mental health outcomes (eg, SMDs >0.6) [9], the present review demonstrated moderate effects (eg, SMD=−0.44, 95% CI −0.88 to −0.003 for depression) despite focusing exclusively on self-guided digital interventions. This pattern is consistent with the lower intensity of digital interventions. An earlier systematic review reported the effectiveness of web-based mindfulness interventions, including those with professional assistance for HCWs during COVID-19 [18]. However, it included an app with various components (ie, cognitive behavioral approach) other than mindfulness [48] and half of the included studies were from one specific country, leading to a risk of overestimation of effectiveness. In this study, we excluded interventions that did not focus solely on mindfulness or those with professional or personal support. The present findings could provide more precise insights into the effects of digital mindfulness interventions as a primary prevention strategy for the mental health of HCWs.

Effects on Well-Being and Potential Mechanisms

The results demonstrated the effectiveness in improving the well-being of HCWs. This finding was consistent with previous reviews [9,12]. Our review measured well-being using the World Health Organization-Five Well-Being Index (WHO-5) and the Warwick Edinburgh Mental Well-being Scale (WEMWBS). Mindfulness enhances skills to accept negative emotions and to cope with distress [49,50], leading to emotional stability and calmness. Additionally, mindfulness promotes greater attention to and awareness of present moment experiences [27] and may encourage individuals to be active and vigorous in pursuing something that they want to focus on. Positive affects, including relaxed, calm, active, vigorous, and energetic states, were measured by both WHO-5 and WEMWBS. Mindfulness also nurtures self-acceptance and self-compassion; it may reduce self-criticism and improve self-esteem. This contributes to enhancing items on the WEMWBS, such as “feeling good about myself” and “feeling confident.” Digital mindfulness interventions can increase positive aspects of mental health while also reducing negative emotions.

Effects on Emotional Exhaustion and Compassion Fatigue

No significant effectiveness was shown in addressing emotional exhaustion or compassion fatigue. One previous systematic review of mindfulness for nurses reported its effectiveness in improving personal accomplishment (a subscale of burnout), but no effectiveness was observed for emotional exhaustion or depersonalization [51]. Another meta-analysis of interventions for nurses also reported nonsignificant effectiveness across all subscales of burnout [52]. However, these meta-analyses included only 2 studies to estimate the efficacy. More studies are needed to draw firm conclusions. A previous review of RCTs of both individual-based and organization-based interventions for reducing burnout among physicians reported beneficial integrated effectiveness on all burnout subscales [53]. There remains uncertainty about the effectiveness of digital mindfulness interventions on emotional exhaustion. Nevertheless, it may be necessary to consider using such interventions in conjunction with structural or organizational strategies (eg, working hour limitation)

Cultural and Occupational Contexts

Cultural and occupational contexts may also influence the effectiveness of digital mindfulness interventions. The included RCTs were conducted across diverse regions, including Europe and Asia, where cultural norms regarding emotional expression, self-care, and help-seeking may shape engagement with mindfulness practices. In addition, occupational differences among HCWs, such as those between nurses and physicians, may affect baseline stressors, workload, and responsiveness to self-guided interventions. These contextual factors should be considered when interpreting the findings and their generalizability.

Heterogeneity Across Interventions

A substantial heterogeneity observed in this study may also be attributed to the variability in intervention characteristics, including duration, number of modules, and delivery modes. For instance, intervention durations ranged from 1.5 weeks [41] to 18 weeks [46], and content delivery varied between smartphone apps (eg, commercial apps [41,46,47], original apps [40,45], SNS tools [39,43,44]), and web-based platforms [42]. Although the limited number of studies precluded a subgroup analysis to statistically quantify the impact of these variations, our sensitivity analysis indicated that excluding specific outliers significantly reduced statistical heterogeneity. This suggests that while structural differences exist, they may not be the sole driver of the observed inconsistency. Nevertheless, this variability highlights a challenge in synthesizing evidence. To facilitate more precise comparisons and identify optimal intervention components, future research should strive for greater standardization in intervention protocols and reporting.

Limitations

While our meta-analysis indicated a beneficial trend of the intervention on mental health outcomes, the certainty of evidence assessed by the GRADE approach remained “Low” or “Very Low.” These results suggest that the true effect may be substantially different from the estimate of the effect. The

XSL•FO
RenderX

primary factor reducing the certainty was the "very serious risk of bias" identified in all included studies. Methodological limitations, such as a lack of blinding or allocation concealment and reliance on self-reported outcomes, were present in many studies and may have led to an overestimation of the effects. Furthermore, inconsistency across studies for depression and perceived stress outcomes, as well as imprecision of the estimates for emotional exhaustion, further contributed to the lower certainty grades. Therefore, the current findings should be interpreted with caution. To establish more certain evidence, future research requires well-designed, high-quality RCTs that minimize the risk of bias. In addition to these points, this review has several limitations. First, only nine studies met the inclusion criteria, which restricts the generalizability of the findings. The inclusion of outcome-related terms in the search strategy may have increased the risk of missing some relevant studies, as noted in the Cochrane Handbook [25], and should be considered a limitation of this review. Additionally, a high risk of bias was observed in the quality assessment using the RoB2 tool, primarily due to self-reported outcome measures. Due to the nature of digital mindfulness interventions as psychosocial interventions, blinding of participants was not feasible in most studies, which may have introduced performance bias alongside reporting and detection bias. This introduces potential bias and may impact the reliability of the results. Furthermore, the included studies exhibit considerable heterogeneity in intervention content, making it challenging to draw consistent conclusions across studies. In addition, the quality of the included digital mindfulness applications may have varied, as not all interventions were clearly reported to be developed or supervised by qualified mindfulness experts. This variability in app quality should be considered when interpreting the findings. The review also includes several RCTs with small sample sizes, limiting the statistical power of these studies. The results from Egger tests and a visual inspection of the funnel plots provide little indication of publication bias. Yet, because only a small number of studies were included, conclusions regarding publication bias should be drawn with caution. The language restriction to studies published in English or Japanese may have resulted in the exclusion of relevant research published in other languages, thereby narrowing the scope of this review. The long-term effects of the interventions still need to be clarified, as few studies evaluated outcomes over an extended period. Adherence rates were reported in some studies but could not be formally linked to outcomes due to limited and inconsistent reporting. Future individual participant data meta-analyses could examine adherence–outcome relationships in self-guided digital mindfulness interventions. Although classified as self-guided, some interventions allowed minimal asynchronous contact, which may have influenced adherence and effect size. Finally, this study's findings, derived from RCTs, present a limitation because the inherent research design often involves human-mediated interventions, such as reminders and economic rewards, which likely enhance adherence. Therefore, in real-world settings where such self-guided digital mindfulness interventions are used without human facilitation, the opportunity to achieve similar levels of adherence—and thereby comparable effects—may be significantly reduced, potentially resulting in smaller observed effects in real-world data contexts. Although beyond the primary scope of this review, ethical and implementation considerations such as user privacy, data security, and responsible data management are important issues for the real-world deployment of digital mindfulness interventions. These factors should be carefully addressed in future research and implementation efforts.

Conclusions

Given the high risk of bias across all included studies, the findings should be interpreted with caution. Nevertheless, health care organizations may consider integrating digital mindfulness interventions as an accessible, short-term mental health support tool for reducing depression, anxiety, and perceived stress among HCWs. However, additional strategies may be needed to address persistent emotional exhaustion and compassion fatigue.

Acknowledgments

The authors wish to thank Inforesta Co, Ltd for their professional and technical support in sifting the articles at the first screening step of this systematic review. Generative AI tools were used to assist with language editing and with improving the clarity of the manuscript. All scientific content, interpretations, and conclusions were developed by the authors, who take full responsibility for the integrity of the work.

Funding

This study is supported by the Japan Agency for Medical Research and Development (AMED) under grant number JP22rea522006 and The University of Tokyo Occupational Mental Health (TOMH) Research Fund, obtained from the Department of Mental Health, Graduate School of Medicine, The University of Tokyo, Japan. The sponsors had no role in the design and conduct of the study; collection, management, analysis, or interpretation of the data; in the preparation, review, or approval of the manuscript; or in the decision to submit the manuscript for publication.

Data Availability

All data generated or analyzed during this study are included in this published article and its multimedia appendices. The data supporting the findings of this review were extracted from previously published articles included in the systematic review, all of which are cited in the reference list.

XSL•FO
RenderX

Authors' Contributions

Conceptualization: NS, AI, KI
Methodology: NS, AI, M Iida, KW, KI
Investigation (screening, data extraction, and risk of bias assessment): NS, AI, M Iida, KW, Y Kobayashi, A Sakuraya, HE, KT, Y Komase, YO, M Iwanaga, RI, KK, AH, HA, SK, YI, TY, HI, GM, RK, TO, NT, KI
Data curation: NS, M Iida, KI
Formal analysis: KW, NS, KI
Funding acquisition: TE, KI
Project administration: KI
Supervision: N Kojimahara, TE, A Shimazu, AT, N Kawakami
Writing – original draft: NS, KI
Writing – review & editing: all authors

Conflicts of Interest

NS received personal fees from Medilio Inc. and research funding from AI Mental Support Inc. outside the submitted work. The authors ASa, KI, and NK are employed at the Department of Digital Mental Health, an endowment department supported with an unrestricted grant from 15 enterprises [54] outside the submitted work. NK is employed at the Junpukai Foundation. Other co-authors had no statements.

Multimedia Appendix 1
Search terms.
[DOCX File, 42 KB - mhealth_v14i1e74272_app1.docx]

Multimedia Appendix 2
Excluded studies with reasons.
[XLSX File, 14 KB - mhealth_v14i1e74272_app2.xlsx]

Multimedia Appendix 3
Results of funnel plots of the included studies.
[DOCX File, 102 KB - mhealth_v14i1e74272_app3.docx]

Multimedia Appendix 4
GRADE Evidence profile.
[DOCX File, 19 KB - mhealth_v14i1e74272_app4.docx]

Checklist 1
PRISMA checklist.
[PDF File, 141 KB - mhealth_v14i1e74272_app5.pdf]

References

1. Gerada C, Sidhu A, Griffiths F. Doctors and suicide. BMJ 2024 Aug 21;386:q1758. [doi: 10.1136/bmj.q1758] [Medline: 39168508]
2. Guille C, Sen S. Burnout, depression, and diminished well-being among physicians. N Engl J Med 2024 Oct 24;391(16):1519-1527. [doi: 10.1056/NEJMra2302878] [Medline: 39442042]
3. Hill JE, Harris C, Danielle L C, et al. The prevalence of mental health conditions in healthcare workers during and after a pandemic: systematic review and meta-analysis. J Adv Nurs 2022 Jun;78(6):1551-1573. [doi: 10.1111/jan.15175] [Medline: 35150151]
4. Zhang YY, Han WL, Qin W, et al. Extent of compassion satisfaction, compassion fatigue and burnout in nursing: a meta-analysis. J Nurs Manag 2018 Oct;26(7):810-819. [doi: 10.1111/jonm.12589] [Medline: 30129106]
5. Kase SM, Gribben JL, Guttmann KF, Waldman ED, Weintraub AS. Compassion fatigue, burnout, and compassion satisfaction in pediatric subspecialists during the SARS-CoV-2 pandemic. Pediatr Res 2022 Jan;91(1):143-148. [doi: 10.1038/s41390-021-01635-y] [Medline: 34211128]
6. Murthy VH. Confronting health worker burnout and well-being. N Engl J Med 2022 Aug 18;387(7):577-579. [doi: 10.1056/NEJMp2207252] [Medline: 35830683]
7. Kabat-Zinn J. Mindfulness-based interventions in context: past, present, and future. Clin Psychol Sci Pract 2003;10(2):144-156. [doi: 10.1093/clipsy.bpg016]

XSL•FO
RenderX

8. Kabat-Zinn J. Catalyzing movement towards a more contemplative society. Presented at: The Contemplative Mind in Society: Meeting of the Working Group; Sep 29 to Oct 2, 1994 URL: https://contemplativemind.org/admin/wp-content/uploads/2012/09/kabat-zinn.pdf [accessed 2026-07-02]
9. Michaelsen MM, Graser J, Onescheit M, et al. Mindfulness-based and mindfulness-informed interventions at the workplace: a systematic review and meta-regression analysis of RCTs. Mindfulness (N Y) 2023 May 11:1-34. [doi: 10.1007/s12671-023-02130-7] [Medline: 37362186]
10. Cohen C, Pignata S, Bezak E, Tie M, Childs J. Workplace interventions to improve well-being and reduce burnout for nurses, physicians and allied healthcare professionals: a systematic review. BMJ Open 2023 Jun 29;13(6):e071203. [doi: 10.1136/bmjopen-2022-071203] [Medline: 37385740]
11. Luken M, Sammons A. Systematic review of mindfulness practice for reducing job burnout. Am J Occup Ther 2016;70(2):7002250020p1-7002250020p10. [doi: 10.5014/ajot.2016.016956] [Medline: 26943107]
12. Ong NY, Teo FJJ, Ee JZY, et al. Effectiveness of mindfulness-based interventions on the well-being of healthcare workers: a systematic review and meta-analysis. Gen Psychiatr 2024;37(3):e101115. [doi: 10.1136/gpsych-2023-101115] [Medline: 38737894]
13. Sanz S, Valiente C, Espinosa R, Trucharte A. Psychological group interventions for reducing distress symptoms in healthcare workers: a systematic review. Clin Psychol Psychother 2024;31(3):e2980. [doi: 10.1002/cpp.2980] [Medline: 38706143]
14. Banerjee M, Cavanagh K, Strauss C. A qualitative study with healthcare staff exploring the facilitators and barriers to engaging in a self-help mindfulness-based intervention. Mindfulness (N Y) 2017;8(6):1653-1664. [doi: 10.1007/s12671-017-0740-z] [Medline: 29201248]
15. Clement S, Schauman O, Graham T, et al. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. Psychol Med 2015 Jan;45(1):11-27. [doi: 10.1017/S0033291714000129] [Medline: 24569086]
16. Mrazek AJ, Mrazek MD, Cherolini CM, et al. The future of mindfulness training is digital, and the future is now. Curr Opin Psychol 2019 Aug;28:81-86. [doi: 10.1016/j.copsyc.2018.11.012] [Medline: 30529975]
17. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR mHealth uHealth 2015 Aug 19;3(3):e82. [doi: 10.2196/mhealth.4328] [Medline: 26290327]
18. Yang JM, Ye H, Long Y, et al. Effects of web-based mindfulness-based interventions on anxiety, depression, and stress among frontline health care workers during the COVID-19 pandemic: systematic review and meta-analysis. J Med Internet Res 2023 Aug 29;25:e44000. [doi: 10.2196/44000] [Medline: 37527546]
19. Pospos S, Young IT, Downs N, et al. Web-based tools and mobile applications to mitigate burnout, depression, and suicidality among healthcare students and professionals: a systematic review. Acad Psychiatry 2018 Feb;42(1):109-120. [doi: 10.1007/s40596-017-0868-0] [Medline: 29256033]
20. DeLight [Article in Japanese]. URL: https://delight.sanei.or.jp/ [accessed 2026-07-02]
21. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
22. Eguchi H, Kojimahara N, Kanamori S, Imamura K, Tani N, Ebara T. The use of digital health technology to provide mental health services for employees in Japan. Environ Occup Health Pract 2024;6(1):2023-0016. [doi: 10.1539/eohp.2023-0016-CT] [Medline: 40061366]
23. Guidelines on mental health at work. : World Health Organization; 2022 URL: https://iris.who.int/server/api/core/bitstreams/6152a556-6893-4c4e-9ed8-094478bb25eb/content [accessed 2026-07-02]
24. Eriksen MB, Frandsen TF. The impact of patient, intervention, comparison, outcome (PICO) as a search strategy tool on literature search quality: a systematic review. J Med Libr Assoc 2018 Oct;106(4):420-431. [doi: 10.5195/jmla.2018.345] [Medline: 30271283]
25. Lefebvre C, Glanville J, Briscoe S, et al. Searching for and selecting studies. In: Cochrane Handbook for Systematic Reviews of Interventions: John Wiley & Sons; 2019:67-107. [doi: 10.1002/9781119536604.ch4]
26. Frandsen TF, Eriksen MB, Hammer DMG, Christensen JB, Wallin JA. Using Embase as a supplement to PubMed in Cochrane reviews differed across fields. J Clin Epidemiol 2021 May;133:24-31. [doi: 10.1016/j.jclinepi.2020.12.022] [Medline: 33359253]
27. Creswell JD. Mindfulness Interventions. Annu Rev Psychol 2017 Jan 3;68:491-516. [doi: 10.1146/annurev-psych-042716-051139] [Medline: 27687118]
28. Cavanagh K, Strauss C, Forder L, Jones F. Can mindfulness and acceptance be learnt by self-help?: a systematic review and meta-analysis of mindfulness and acceptance-based self-help interventions. Clin Psychol Rev 2014 Mar;34(2):118-129. [doi: 10.1016/j.cpr.2014.01.001] [Medline: 24487343]
29. Spijkerman MPJ, Pots WTM, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: a review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114. [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
30. Lazaridou A, Philbrook P, Tzika AA. Yoga and mindfulness as therapeutic interventions for stroke rehabilitation: a systematic review. Evid Based Complement Alternat Med 2013;2013(1):357108. [doi: 10.1155/2013/357108] [Medline: 23781259]

XSL•FO
RenderX

31. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
32. Balk EM, et al. Empirical assessment of within-arm correlation imputation in trials of continuous outcomes. : Agency for Healthcare Research and Quality (AHRQ); 2012 URL: https://www.ncbi.nlm.nih.gov/books/NBK115797/pdf/Bookshelf_NBK115797.pdf [accessed 2026-07-02]
33. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 6;327(7414):557-560. [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
34. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634. [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
35. Sterne JAC, Sutton AJ, Ioannidis JPA, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011 Jul 22;343:d4002. [doi: 10.1136/bmj.d4002] [Medline: 21784880]
36. Page MJ, Higgins JPT, Sterne JAC. Assessing risk of bias due to missing results in a synthesis. In: Cochrane Handbook for Systematic Reviews of Interventions: John Wiley & Sons; 2019:349-374. [doi: 10.1002/9781119536604.ch13]
37. Schünemann HJ, Higgins JPT, Vist GE, et al. Completing 'summary of findings' tables and grading the certainty of the evidence. In: Cochrane Handbook for Systematic Reviews of Interventions: John Wiley & Sons; 2019:375-402. [doi: 10.1002/9781119536604.ch14]
38. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw 2010;36(3):1-48. [doi: 10.18637/jss.v036.i03]
39. Asadollah F, Nikfarid L, Sabery M, Varzeshnejad M, Hashemi F. The impact of loving-kindness meditation on compassion fatigue of nurses working in the neonatal intensive care unit: a randomized clinical trial study. Holist Nurs Pract 2023;37(4):215-222. [doi: 10.1097/HNP.0000000000000590] [Medline: 37335149]
40. Hwang WJ, Jo HH. Evaluation of the effectiveness of mobile app-based stress-management program: a randomized controlled trial. Int J Environ Res Public Health 2019 Nov 3;16(21):21. [doi: 10.3390/ijerph16214270] [Medline: 31684169]
41. Keng SL, Chin JWE, Mammadova M, Teo I. Effects of mobile app-based mindfulness practice on healthcare workers: a randomized active controlled trial. Mindfulness (N Y) 2022;13(11):2691-2704. [doi: 10.1007/s12671-022-01975-8] [Medline: 36160038]
42. Lilly M, Calhoun R, Painter I, et al. Destress 9-1-1-an online mindfulness-based intervention in reducing stress among emergency medical dispatchers: a randomised controlled trial. Occup Environ Med 2019 Oct;76(10):705-711. [doi: 10.1136/oemed-2018-105598] [Medline: 31138676]
43. Luangapichart P, Saisavoey N, Viravan N. Efficacy and feasibility of the minimal therapist-guided four-week online audio-based mindfulness program 'Mindful Senses' for burnout and stress reduction in medical personnel: a randomized controlled trial. Healthcare (Basel) 2022 Dec 14;10(12):12. [doi: 10.3390/healthcare10122532] [Medline: 36554056]
44. Nourian M, Nikfarid L, Khavari AM, Barati M, Allahgholipour AR. The impact of an online mindfulness-based stress reduction program on sleep quality of nurses working in COVID-19 care units: a clinical trial. Holist Nurs Pract 2021;35(5):257-263. [doi: 10.1097/HNP.0000000000000466] [Medline: 34407023]
45. Pratt EH, Hall L, Jennings C, et al. Mobile mindfulness for psychological distress and burnout among frontline COVID-19 nurses: a pilot randomized trial. Ann Am Thorac Soc 2023 Oct;20(10):1475-1482. [doi: 10.1513/AnnalsATS.202301-025OC] [Medline: 37289650]
46. Taylor H, Cavanagh K, Field AP, Strauss C. Health care workers' need for Headspace: findings from a multisite definitive randomized controlled trial of an unguided digital mindfulness-based self-help app to reduce healthcare worker stress. JMIR mHealth uHealth 2022 Aug 25;10(8):e31744. [doi: 10.2196/31744] [Medline: 36006668]
47. Xu HG, Eley R, Kynoch K, Tuckett A. Effects of mobile mindfulness on emergency department work stress: a randomised controlled trial. Emerg Med Australas 2022 Apr;34(2):176-185. [doi: 10.1111/1742-6723.13836] [Medline: 34378320]
48. Fiol-DeRoque MA, Serrano-Ripoll MJ, Jiménez R, et al. A mobile phone-based intervention to reduce mental health problems in health care workers during the COVID-19 pandemic (PsyCovidApp): randomized controlled trial. JMIR mHealth uHealth 2021 May 18;9(5):e27039. [doi: 10.2196/27039] [Medline: 33909587]
49. Lindsay EK, Chin B, Greco CM, et al. How mindfulness training promotes positive emotions: dismantling acceptance skills training in two randomized controlled trials. J Pers Soc Psychol 2018 Dec;115(6):944-973. [doi: 10.1037/pspa0000134] [Medline: 30550321]
50. Hill CLM, Updegraff JA. Mindfulness and its relationship to emotional regulation. Emotion 2012 Feb;12(1):81-90. [doi: 10.1037/a0026355] [Medline: 22148996]
51. Ramachandran HJ, Bin Mahmud MS, Rajendran P, Jiang Y, Cheng L, Wang W. Effectiveness of mindfulness-based interventions on psychological well-being, burnout and post-traumatic stress disorder among nurses: a systematic review and meta-analysis. J Clin Nurs 2023 Jun;32(11-12):2323-2338. [doi: 10.1111/jocn.16265] [Medline: 35187740]
52. Suleiman-Martos N, Gomez-Urquiza JL, Aguayo-Estremera R, Cañadas-De La Fuente GA, De La Fuente-Solana EI, Albendín-García L. The effect of mindfulness training on burnout syndrome in nursing: a systematic review and meta-analysis. J Adv Nurs 2020 May;76(5):1124-1140. [doi: 10.1111/jan.14318] [Medline: 32026484]
53. West CP, Dyrbye LN, Erwin PJ, Shanafelt TD. Interventions to prevent and reduce physician burnout: a systematic review and meta-analysis. Lancet 2016 Nov 5;388(10057):2272-2281. [doi: 10.1016/S0140-6736(16)31279-X] [Medline: 27692469]

XSL•FO
RenderX

54. Participating companies and cooperating courses [Article in Japanese]. DMH (Department of Digital Mental Health). URL: https://dmh.m.u-tokyo.ac.jp/c [accessed 2026-07-02]

Abbreviations

GRADE: Grading of Recommendations Assessment, Development and Evaluation
HCW: health care worker
PICO: participants, interventions, comparisons, and outcomes
PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analysis
RCT: randomized controlled trial
SMD: standardized mean difference
SNS: social networking service
WEMWBS: Warwick Edinburgh Mental Well-being Scale
WHO-5: World Health Organization-Five Well-Being Index

Edited by L Buis; submitted 21.Mar.2025; peer-reviewed by A Adeoye, M Iwakura, S Markham, T Tanimoto; revised version received 21.May.2026; accepted 26.May.2026; published 21.Aug.2026.

Please cite as:
Sasaki N, Inoue A, Iida M, Watanabe K, Kobayashi Y, Sakuraya A, Eguchi H, Tsuno K, Komase Y, Otsuka Y, Iwanaga M, Inoue R, Kuribayashi K, Hino A, Asaoka H, Kanamori S, Inagawa Y, Yamauchi T, Ikeda H, Muto G, Kato R, Okuyama T, Tani N, Kojimahara N, Ebara T, Shimazu A, Tsutsumi A, Kawakami N, Imamura K
Effects of Self-Guided Digital Mindfulness Interventions on the Mental Health of Health Care Workers: Systematic Review and Meta-Analysis of Randomized Controlled Trials
JMIR Mhealth Uhealth 2026;14:e74272
URL: https://mhealth.jmir.org/2026/1/e74272
doi:10.2196/74272

© Natsu Sasaki, Akiomi Inoue, Mako Iida, Kazuhiro Watanabe, Yuka Kobayashi, Asuka Sakuraya, Hisashi Eguchi, Kanami Tsuno, Yu Komase, Yasumasa Otsuka, Mai Iwanaga, Reiko Inoue, Kazuto Kuribayashi, Ayako Hino, Hiroki Asaoka, Satoru Kanamori, Yuta Inagawa, Takenori Yamauchi, Hiroki Ikeda, Go Muto, Rika Kato, Takuro Okuyama, Naomichi Tani, Noriko Kojimahara, Takeshi Ebara, Akihito Shimazu, Akizumi Tsutsumi, Norito Kawakami, Kotaro Imamura. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 21.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Just-in-Time Adaptive Interventions to Reduce Sedentary Behavior and Promote Physical Activity in Patients With Chronic Disease: Scoping Review

Junchen Guo[1,2], MSN; Furong Chen[3], MSN; Chan Chen[4], PhD; Yunyun Dai[5], PhD; Xianghua Xu[2], PhD; Yongyi Chen[2], PhD; Mengying Huang[6], MSN

[1]Australasian Health Outcomes Consortium, University of Wollongong, Wollongong, NSW, Australia

[2]Department of Palliative Care, Hunan Cancer Hospital, Changsha, Hunan, China

[3]School of Nursing, Guangzhou Medical University, Guangzhou, China

[4]School of Journalism and Communication, Hunan Normal University, Changsha, China

[5]School of Nursing, Guilin Medical University, Guilin, China

[6]Department of Nursing, Hunan Provincial Hospital of Integrated Traditional Chinese and Western Medicine, No. 58 Lushan Road, Yuelu District, Changsha, Hunan, China

Corresponding Author:
Mengying Huang, MSN
Department of Nursing, Hunan Provincial Hospital of Integrated Traditional Chinese and Western Medicine, No. 58 Lushan Road, Yuelu District, Changsha, Hunan, China

Abstract

Background: Patients with chronic diseases often struggle to maintain sufficient physical activity (PA) and reduce sedentary behavior in daily life. Just-in-time adaptive interventions (JITAIs), delivered through mobile health technologies such as mobile apps, wearable devices, sensors, and ecological momentary assessment, offer timely and personalized support based on individuals' changing needs. However, the evidence base for their use in promoting PA and reducing sedentary behavior among patients with chronic diseases remains fragmented.

Objective: This review aimed to describe study characteristics and design features of JITAIs, summarize reported evaluation findings related to PA and sedentary behavior, and identify barriers and facilitators related to engagement and use of JITAIs.

Methods: A comprehensive literature search was conducted in PubMed, Embase, Web of Science, CINAHL, and Scopus from database inception to June 2025. Two researchers independently screened the records, selected eligible studies, and performed data charting to ensure rigor and consistency.

Results: A total of 14 studies were included in this review. Of these, 6 primarily targeted PA, 1 focused solely on sedentary behavior, and 5 addressed both PA and sedentary behavior. Of the remaining, 6 studies provided preliminary evidence supporting the positive effects of JITAIs in promoting PA among individuals with chronic diseases, while 3 studies suggested positive effects on reducing sedentary behavior. Barriers and facilitators influencing the use of JITAIs in this population were identified. Barriers included technological usability and literacy, burden and intrusiveness of the intervention, perceived value and acceptability, external and contextual barriers, and privacy and data security concerns. Facilitators included enhanced motivation and behavioral engagement, increased awareness and self-monitoring, usability and integration into daily life, timely and personalized support, and support from health care professionals and the care environment.

Conclusions: This scoping review identified preliminary evidence suggesting that JITAIs may help reduce sedentary behavior and promote PA among patients with chronic diseases. However, more large-scale, high-quality randomized controlled trial studies are needed to strengthen evidence and generalizability.

(JMIR Mhealth Uhealth 2026;14:e81378) doi:10.2196/81378

KEYWORDS

physical activity; sedentary behavior; just-in-time adaptive intervention; digital health; chronic disease

Introduction

According to the World Health Statistics 2023 report published by the World Health Organization, noncommunicable diseases account for approximately 41 million deaths annually, representing 74% of all deaths worldwide [1]. Chronic diseases, such as cardiovascular diseases, chronic respiratory diseases, and diabetes, are long-term conditions that generally require

XSL•FO
RenderX

ongoing medical care and sustained self-management [2]. These conditions can often be prevented or managed through behavioral changes, particularly by increasing physical activity (PA) and reducing sedentary behavior. PA refers to any bodily movement produced by skeletal muscles that results in energy expenditure, whereas sedentary behavior refers to any waking behavior characterized by low energy expenditure while sitting, reclining, or lying [3,4]. Regular PA is associated with improved cardiovascular health, glycemic control, and mental well-being [5,6]. However, many patients with chronic diseases have difficulty achieving sufficient levels of PA and limiting sedentary time in daily life. Prolonged sedentary behavior has been associated with impaired metabolic function, obesity, anxiety, depression, and increased risks of hypertension, diabetes, cardiovascular disease, and other adverse health outcomes [7-9]. Therefore, reducing sedentary behavior and promoting PA are important components of chronic disease management.

Traditional lifestyle interventions designed to reduce sedentary behavior and promote PA may be effective in controlled settings, but they often rely on standardized, fixed plans that deliver the same type and intensity of support to all patients at predetermined time points [10]. This approach may overlook individual variability and dynamic changes in patients' needs, potentially resulting in insufficient, excessive, or poorly matched support [10]. Digital health, which refers to the use of digital technologies to support health care delivery, monitoring, and self-management, offers opportunities to provide more timely and personalized behavioral support [11]. Within digital health, mobile health (mHealth) interventions use mobile phones, wearable activity trackers, sensors, and other wireless technologies to monitor health behaviors and deliver intervention content in everyday settings [12]. In interventions targeting PA and sedentary behavior, mHealth tools can track steps, activity intensity, sedentary time, and location, as well as deliver feedback, reminders, motivational messages, and activity suggestions [13].

In recent years, just-in-time adaptive interventions (JITAIs) have emerged as a promising mHealth approach to supporting behavior change [14]. A JITAI is designed to provide the right type and amount of support at the right time by adapting intervention delivery to a person's current needs, context, or state of vulnerability and opportunity [14,15]. In practice, JITAIs commonly use data from wearable sensors, mobile apps, ecological momentary assessment (EMA), or contextual information, such as time of day, location, and recent activity patterns, to determine when and how support should be delivered [15,16]. Key JITAIs design components typically include decision points, tailoring variables, intervention options, decision rules, and proximal or distal outcomes [16]. Previous studies have applied JITAIs across a range of health behavior areas, including smoking cessation, alcohol reduction, dietary improvement, and mental health support, suggesting that they may provide more personalized and timely support in daily life than traditional interventions [17,18]. In the field of PA promotion and sedentary behavior reduction, JITAIs have increasingly been applied among individuals with diabetes, cardiovascular disease, and other long-term conditions [19-21].

A key potential advantage of JITAIs is their ability to use data from mobile phones, sensors, and self-reports to deliver context-sensitive support that accounts for daily fluctuations in symptoms, motivation, routines, and opportunities for PA [19-21].

Despite these potential advantages, the evidence base for the use of JITAIs to promote PA and reduce sedentary behavior among patients with chronic diseases remains fragmented. Existing studies also showed their implementation may be affected by potential disadvantages, as JITAIs often depend on patients' access to and ability to use mobile or wearable technologies, which can create barriers related to technical problems, device compatibility, and user burden [15,16]. Frequent or poorly timed prompts may also lead to notification fatigue or disengagement [15,16]. How these advantages and disadvantages influence PA and sedentary behavior outcomes among patients with chronic diseases has not been consistently synthesized across existing studies. Moreover, current studies also differ in their design and evaluation outcomes, making it difficult to evaluate the effectiveness of JITAIs in promoting PA and reducing sedentary behavior in this population. Therefore, we conducted this scoping review focusing on studies in which JITAIs designed to reduce sedentary behavior and/or promote PA among patients with chronic diseases. The review aimed to describe the study characteristics and design features of JITAIs, summarize reported evaluation findings related to PA and sedentary behavior, and identify barriers and facilitators related to engagement and use of JITAIs. By mapping this evidence, the review may inform the development of future JITAIs that are more acceptable and suitable for supporting PA promotion and sedentary behavior reduction in chronic disease self-management.

Methods

Study Design

This scoping review was systematically conducted following the methodological framework proposed by the Joanna Briggs Institute (JBI) [22]. The review was reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist (Checklist 1) [23]. Ethical approval was not required for this scoping review, as no new data were collected.

Research Questions

The research questions of this scoping review were identified as follows:

- What are the key study characteristics and design features of JITAIs designed to reduce sedentary behavior and promote PA in patients with chronic disease?
- What evidence exists regarding the evaluation of these JITAIs in terms of PA and sedentary behavior?
- What barriers and facilitators have been reported in relation to the engagement and use of JITAIs among patients with chronic disease?

XSL•FO
RenderX

Search Methods

Eligibility Criteria

The eligibility criteria were established based on the population-concept-context (PCC) framework proposed by the JBI [22]. The inclusion criteria were as follows: (1) population: studies involving adults (aged 18 years and older) diagnosed with one or more chronic diseases; (2) concept: studies that used JITAIs aimed at reducing sedentary behavior and/or promoting PA, where JITAIs were defined as interventions that use sensor technologies (eg, accelerometers, GPS, and light sensors) embedded in mobile devices to collect real-time behavioral and contextual data and deliver tailored support at moments when behavior change is most likely to occur [17]; (3) context: studies conducted in a variety of settings, including health care, community, and home settings, where JITAIs were used to support behavior change in individuals with chronic conditions; and (4) types of evidence sources: both quantitative and qualitative studies were included. Mixed methods studies were also eligible. Only peer-reviewed articles published in English were considered. Studies were excluded if they focused solely on healthy populations, used nonadaptive or nondigital interventions, or did not provide sufficient detail on the intervention characteristics or outcomes of interest.

Search Strategy

In collaboration with two independent university medical librarians in China, a systematic search was conducted across 5 electronic databases, including PubMed, Embase, Web of Science, CINAHL, and Scopus, covering literature from database inception to June 2025. An initial exploratory search was performed in PubMed to identify relevant keywords and MeSH terms based on titles and abstracts of pertinent studies. Boolean operators were used to capture studies related to the 4 core concepts: “just-in-time adaptive interventions,” “physical activity,” “sedentary behavior,” and “chronic disease.” To define the scope of chronic conditions, 10 diseases were selected based on the list published by the Office of the Assistant Secretary for Health of the United States [24]: hypertension, type 2 diabetes (T2D), chronic obstructive pulmonary disease, stroke, heart failure, coronary artery disease, chronic kidney disease, hyperlipidemia, asthma, and arthritis. Conditions that were psychiatric (eg, schizophrenia), acute, or not conducive to self-management (eg, dementia) were excluded as this review focused on chronic physical conditions in which PA and sedentary behavior can be targeted through patient-led or patient-engaged self-management [25]. These conditions may require different intervention approaches or rely more heavily on management led by clinicians or caregivers, and were therefore beyond the scope of this review. Additionally, a snowballing technique was used to manually screen the reference lists of included studies to identify further relevant literature. The PubMed search strategy is presented in Table 1, while the complete search terms, strategies, and results for other databases are provided in Multimedia Appendix 1.

XSL•FO
RenderX

Table . Search strategy of PubMed.

Procedure	Search strategy
#1	"chronic disease" [MeSH Terms] OR "essential hypertension" [MeSH Terms] OR "heart failure" [MeSH Terms] OR "coronary artery disease" [MeSH Terms] OR "stroke" [MeSH Terms] OR "paresis" [MeSH Terms] OR "arthritis" [MeSH Terms] OR "osteoarthritis" [MeSH Terms] OR "asthma" [MeSH Terms] OR "renal insufficiency, chronic" [MeSH Terms] OR "pulmonary disease, chronic obstructive" [MeSH Terms] OR "diabetes mellitus" [MeSH Terms] OR "insulin resistance" [MeSH Terms] OR "hyperlipidemias" [MeSH Terms]
#2	"chronic disease" [Title/Abstract] OR "chronic illness" [Title/Abstract] OR "chronic condition" [Title/Abstract] OR "non communicable disease" [Title/Abstract] OR "noncommunicable disease" [Title/Abstract] OR "chronically ill" [Title/Abstract] OR "hypertension" [Title/Abstract] OR "hypertensive" OR "high blood pressure" [Title/Abstract] OR "heart failure" [Title/Abstract] OR "CHF" [Title/Abstract] OR "HF" [Title/Abstract] OR "cardiac failure" [Title/Abstract] OR "heart decompensation" [Title/Abstract] OR "coronary artery disease" [Title/Abstract] OR "coronary arteriosclerosis" [Title/Abstract] OR "coronary atherosclerosis" [Title/Abstract] OR "angina pectoris" [Title/Abstract] OR "CAD" [Title/Abstract] OR "heart disease" [Title/Abstract] OR "myocardial infarction" [Title/Abstract] OR "unstable angina*" [Title/Abstract] OR "angor pectoris" [Title/Abstract] OR "coronary thrombosis" [Title/Abstract] OR "acute coronary syndrome" [Title/Abstract] OR "myocardial ischemia" [Title/Abstract] OR "myocardial ischaemia" [Title/Abstract] OR "stroke" [Title/Abstract] OR "hemiplegia" [Title/Abstract] OR "hemiplegia" [Title/Abstract] OR "hemiplegias" [Title/Abstract] OR "paresis" [Title/Abstract] OR "cerebrovascular trauma" [Title/Abstract] OR "cerebrovascular accident" [Title/Abstract] OR "CVA" [Title/Abstract] OR "apoplexy" [Title/Abstract] OR "arthritis" [Title/Abstract] OR "rheuma*" [Title/Abstract] OR "osteoarthritis" [Title/Abstract] OR "arthritides" [Title/Abstract] OR "polyarthritis" [Title/Abstract] OR "polyarthritides" [Title/Abstract] OR "asthma" [Title/Abstract] OR "status asthmaticus" [Title/Abstract] OR "bronchial hyper reactivity" [Title/Abstract] OR "asthmatic" [Title/Abstract] OR "wheez" [Title/Abstract] OR "bronchial" [Title/Abstract] OR "obstructive lung disease" [Title/Abstract] OR "chronic renal insufficiency" [Title/Abstract] OR "chronic kidney failure" [Title/Abstract] OR "chronic renal failure" [Title/Abstract] OR "chronic renal disease" [Title/Abstract] OR "chronic kidney disease" [Title/Abstract] OR "chronic kidney disorder" [Title/Abstract] OR "CKD" [Title/Abstract] OR "ESRD" [Title/Abstract] OR "CRD" [Title/Abstract] OR "chronic kidney insufficiency" [Title/Abstract] OR "chronic obstructive pulmonary disease" [Title/Abstract] OR "chronic bronchitis" [Title/Abstract] OR "COPD" [Title/Abstract] OR "chronic obstructive airway disease" [Title/Abstract] OR "chronic airflow obstruction" [Title/Abstract] OR "chronic obstructive lung disease" [Title/Abstract] OR "emphysema" [Title/Abstract] OR "diabetes mellitus" [Title/Abstract] OR "insulin resistance" [Title/Abstract] OR "DMII" [Title/Abstract] OR "DM2" [Title/Abstract] OR "noninsulin dependent" [Title/Abstract] OR "impaired glucose tolerance" [Title/Abstract] OR "impaired glucose tolerant" [Title/Abstract] OR "hyperlipidemias" [Title/Abstract] OR "overweight" [Title/Abstract] OR "obese" [Title/Abstract]
#3	#1 OR #2
#4	"motor activity" [MeSH Terms] OR "exercise" [MeSH Terms] OR "walking" [MeSH Terms] OR "sports" [MeSH Terms] OR "jogging" [MeSH Terms] OR "swimming" [MeSH Terms] OR "health behaviour" [MeSH Terms] OR "cool-down exercise" [MeSH Terms] OR "gymnastics" [MeSH Terms] OR "muscle stretching exercises" [MeSH Terms] OR "circuit-based exercise" [MeSH Terms] OR "endurance training" [MeSH Terms] OR "high-intensity interval training" [MeSH Terms] OR "plyometric exercise" [MeSH Terms] OR "resistance training" [MeSH Terms] OR "warm-up exercise" [MeSH Terms] OR "weight lifting" [MeSH Terms] OR "screen time" [MeSH Terms]

XSL•FO
RenderX

Procedure	Search strategy
#5	"physical activit*" [Title/Abstract] OR "physical fitness" [Title/Abstract] OR "sedentary behaviour" [Title/Abstract] OR "sitting" [Title/Abstract] OR "sedentariness" [Title/Abstract] OR "sedentar*" [Title/Abstract] OR "walk*" [Title/Abstract] OR "run" [Title/Abstract] OR "sport*" [Title/Abstract] OR "jog" [Title/Abstract] OR "swim*" [Title/Abstract] OR "cool-down exercise" [Title/Abstract] OR "cool down exercise" [Title/Abstract] OR "gymnastics" [Title/Abstract] OR "muscle stretching exercises" [Title/Abstract] OR "circuit-based exercise" [Title/Abstract] OR "endurance training" [Title/Abstract] OR "high-intensity interval training" [Title/Abstract] OR "high intensity interval train*" [Title/Abstract] OR "plyometric exercise" [Title/Abstract] OR "resistance training" [Title/Abstract] OR "resistance train*" [Title/Abstract] OR "warm-up exercise" [Title/Abstract] OR "weight lift*" [Title/Abstract] OR "circuit train*" [Title/Abstract] OR "circuit exercise*" [Title/Abstract] OR "weight train*" [Title/Abstract] OR "aerobic train*" [Title/Abstract] OR "aerobic exercise*" [Title/Abstract] OR "cardio train*" [Title/Abstract] OR "screen behavio*" [Title/Abstract] OR "screen Time" [Title/Abstract] OR "screen watching" [Title/Abstract] OR "TV watching" [Title/Abstract] OR "prolonged sitting" [Title/Abstract] OR "lying time" [Title/Abstract] OR "television viewing" [Title/Abstract] OR "active recreation" [Title/Abstract] OR "active leisure" [Title/Abstract] OR "active living" [Title/Abstract]
#6	#4 OR #5
#7	"Just-in-time adaptive intervention" [Title/Abstract] OR "Just-in-time" [Title/Abstract] OR "JITAI" [Title/Abstract] OR "ecological momentary intervention*" [Title/Abstract] OR "EMI" [Title/Abstract] OR "adaptive intervention" [Title/Abstract] OR "dynamic intervention" [Title/Abstract] OR "real time intervention*" [Title/Abstract] OR "context aware*" [Title/Abstract] OR "context triggered" [Title/Abstract] OR "context tailor*" [Title/Abstract] OR "dynamic tailor*" [Title/Abstract] OR "real time tailor*" [Title/Abstract] OR "sensor triggered" [Title/Abstract] OR "geofenc*" [Title/Abstract] OR "context sens*" [Title/Abstract] OR "real time context*" [Title/Abstract] OR "persuasive technolog*" [Title/Abstract] OR "sensing technolog*" [Title/Abstract]
#8	#3 AND #6 AND #7

Data Management

Data Selection

The search results were imported into EndNote software (version X9.1; Clarivate). After removing duplicate records, two researchers (JG and FC) independently screened the titles and abstracts of all retrieved studies to assess initial eligibility based on the predefined inclusion criteria. Studies that met the inclusion criteria during title and abstract screening were retrieved in full text and further assessed for eligibility. Any discrepancies between the two primary researchers were resolved through discussion or adjudicated by the third reviewer (YC).

Data Extraction

Two researchers (JG and MH) independently extracted and recorded data using a predesigned standardized Excel form (Microsoft, 2022). The data extraction process was guided by the review's research questions and included the following information: (1) characteristics of included studies such as first author, year of publication, country, objectives, study design, population and setting, and sample size; (2) design characteristics of JITAIs, including theoretical framework, target behaviors, sensors/data sources, tailoring variables, decision points, intervention options, decision rules, prompt frequency/timing, adaptation mechanisms, intervention duration, engagement, and implementation-related measures; (3) evaluation of JITAIs in terms of PA and sedentary behavior; and (4) reported barriers and facilitators related to engagement and use of JITAIs. A third reviewer (YD) independently reviewed the extracted data from all included studies. Any discrepancies between the two primary researchers were resolved through discussion or adjudicated by the third reviewer (YD).

Data Synthesis and Analysis

Data synthesis and analysis were conducted using a descriptive and narrative approach, in line with the objectives and research questions of this scoping review. Given the heterogeneity of study designs and target populations, a meta-analysis was not conducted. Data on study characteristics, design features of JITAIs, and evaluation findings were descriptively synthesized and tabulated to summarize the overall profile of the included studies [26]. Where appropriate, similar JITAI design components were grouped into categories. For example, decision rules were categorized as EMA-based if-then rules, sensor-triggered or sedentary-threshold rules, scheduled rules, microrandomization rules, or coach- or staff-supported rules. Additionally, qualitative data and open-ended data were synthesized using inductive thematic analysis to identify reported barriers and facilitators related to the engagement and use of JITAIs among patients with chronic diseases [27]. Two

XSL•FO
RenderX

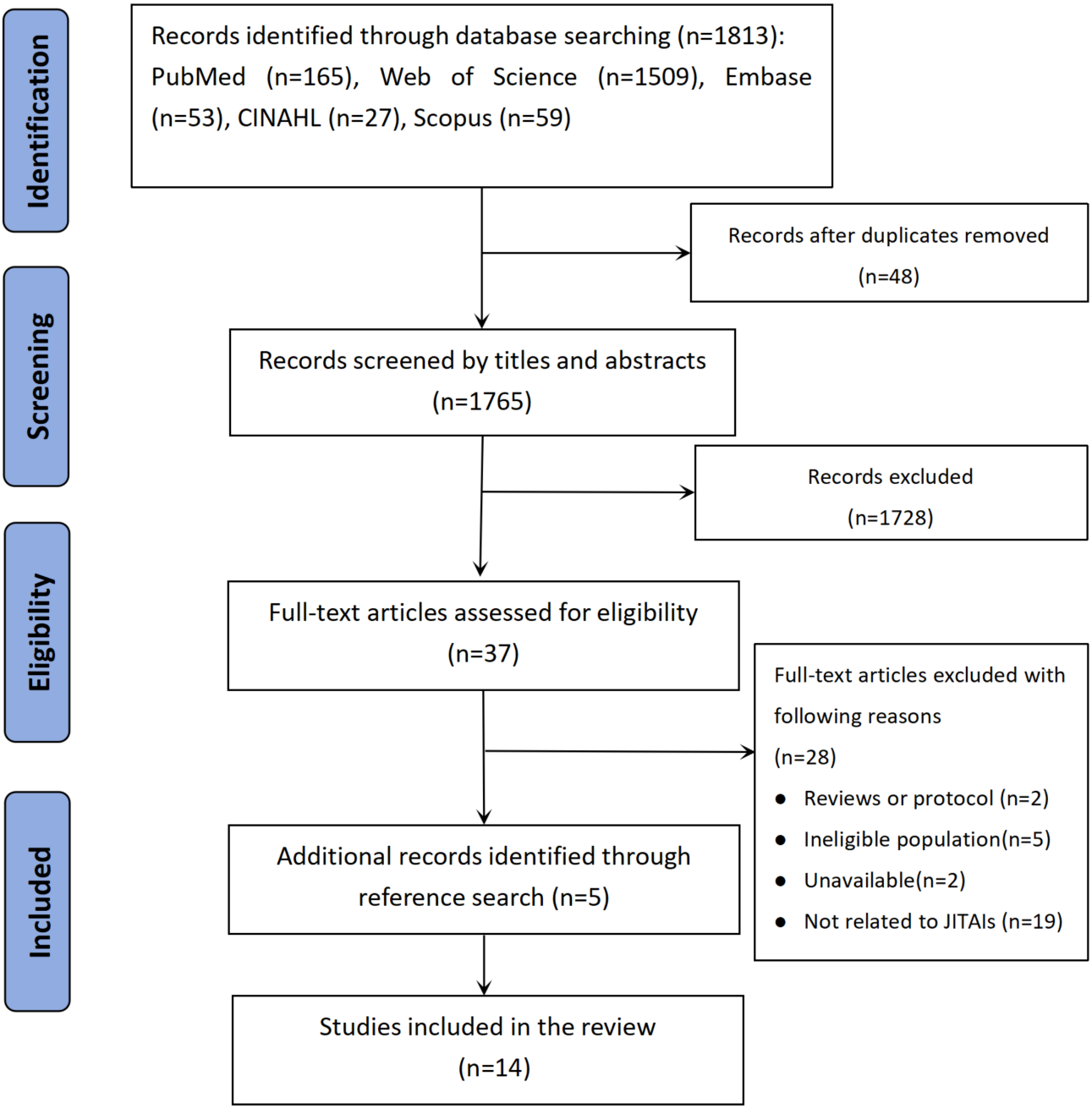
reviewers (JG and FC) independently coded relevant qualitative findings and extracted statements. Codes were compared, refined, and grouped into overarching themes and subthemes. Discrepancies in interpretation were resolved through discussion or adjudicated by the third reviewer (YD). Final themes and findings were reviewed by the full research team to ensure coherence, transparency, and consistency with the review questions.

Results

Search Results

A total of 1813 records were identified through searches of 5 electronic databases. After removing 48 duplicates, 1765 records remained for title and abstract screening, of which 1728 were excluded based on relevance. The full texts of the remaining 37 articles were assessed for eligibility, resulting in the exclusion of 28 studies. An additional 5 relevant studies were identified through reference list screening. Finally, 14 studies were included in this scoping review. The PRISMA flow diagram of study selection is shown in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of study selection. JITAI: just-in-time adaptive intervention.



XSL•FO
RenderX

Characteristics of Included Studies

The publication years of included studies ranged from 2015 to 2025. The majority (n=9) were conducted in the United States [28-36], while the remaining 5 were carried out in Ireland [37], Australia [38], the Netherlands [19], Singapore [39], and the Czech Republic [20], respectively. All studies were described as feasibility or pilot studies, including 4 qualitative studies [19,31,37,38], 3 randomized controlled trial (RCT) studies [28,29,32], 6 single-armed feasibility or pilot studies [20,33-36,39], and 1 prospective cohort study [30]. Sample sizes ranged from 8 to 108 participants. Among those, 4 studies recruited individuals with T2D [19,20,35,38], 3 targeted individuals undergoing or completing cardiac rehabilitation [28,31,33], and 2 included individuals with stroke [30,37]. The remaining studies enrolled participants with hypertension (n=1) [29], spinal cord injury (n=1) [34], acute coronary syndrome or those who had undergone coronary revascularization (n=1) [33], individuals who were overweight or obese (n=1) [36], and older adults with a chronic condition (n=1) [39]. More details of the characteristics of included studies are presented in Table 2.

XSL•FO
RenderX

Table . Characteristics of included studies.

Author	Year of publication	Country	Objectives	Study design	Population and setting	Sample size
Hietbrink et al [19]	2025	Netherlands	To assess the acceptability of EMA[a]-driven, just-in-time adaptive lifestyle support in individuals with T2D[b]	Qualitative study	• Population: people diagnosed with T2D • Age: mean 70.5 (SD 9) years • Setting: hospital	8
Cardy et al [37]	2024	Ireland	To explore the perspectives of people with strokes, their caregivers, and health care providers on the delivery of an adaptive, personalized mHealth[c] intervention to promote PA[d] after stroke	Qualitative study	• Population: people with stroke, caregivers, health care workers • Age: 54.5, 66.5, and 34 years • Setting: hospital and community	People with stroke (n=12); caregivers (n=6); health care workers (n=10)
Golbus et al [28]	2024	USA	To support cardiac rehabilitation patients in increasing and sustaining PA using a mobile health intervention	RCT[e]	• Population: people were enrolled in cardiac rehabilitation based on a qualifying diagnosis • Age: mean 59.5 (SD 10.7) years • Setting: medical and community health care center	108
Novak et al [20]	2024	Czech Republic	To describe the development and piloting of the mHealth intervention and its final version to be evaluated in the ENERGISED trial	Pilot study	• Population: people diagnosed with prediabetes and T2D • Age: between 40 and 76 years • Setting: hospital	10
Bai et al [29]	2022	USA	To determine the effectiveness of wearable activity trackers alone or in combination with behavior change strategies for promoting PA among individuals with prehypertension or hypertension	RCT	• Population: prehypertension or hypertension patients • Age: mean 55 (SD 8.8) years • Setting: medical clinics, community	44 (control: 22; intervention: 22)

XSL•FO
RenderX

Author	Year of publication	Country	Objectives	Study design	Population and setting	Sample size
Daryabeygi-Khotbehsara et al [38]	2022	Australia	To develop an Android mobile app to deliver a JITAI[f] aimed at reducing sitting time and increasing PA in people with T2D	Qualitative study	• Population: people diagnosed with T2D • Age: mean 49.0 (SD 8.6) years • Setting: not reported	10
Lau et al [30]	2022	USA	To evaluate the feasibility and acceptability of using accelerometry and EMA to monitor daily activity and symptoms in stroke survivors	Prospective cohort study	• Population: people diagnosed with mild to moderate stroke • Age: mean 52.8 (SD 7.48) years • Setting: community	40
Mair et al [39]	2022	Singapore	To describe the development of JitaBug designed to support older adults to increase or maintain their PA level	Feasibility study	• Population: community-dwelling older adults with chronic disease • Age: mean 65.5 (SD 5.4) years • Setting: community	31
Elnaggar et al [31]	2021	USA	To evaluate the acceptability of a wearable device, mobile app, and push messages to facilitate PA following cardiac rehabilitation completion	Qualitative study	• Population: people with a previous cardiac event who qualified and were actively participating in cardiac rehabilitation • Age: mean 66.7 (SD 8.6) years • Setting: community	26
Park et al [32]	2021	USA	To examine preliminary effects of an mHealth intervention on group differences in PA and functional capacity to maintain exercise after cardiac rehabilitation	RCT	• Population: people diagnosed with CVD[g] and were within 2 weeks of completing cardiac rehabilitation • Age: mean 66.7 (SD 8.6) years • Setting: community	51 (control: 25; intervention: 26)
Sengupta et al [33]	2020	USA		Feasibility study		10

XSL•FO
RenderX

Author	Year of publication	Country	Objectives	Study design	Population and setting	Sample size
			To examine the usability and health behaviors of a prototypic mHealth intervention designed specifically for women with CHD[h]		• Population: women diagnosed with an acute coronary syndrome or coronary revascularization • Age: mean 64.4 (SD 6.3) years • Setting: medical center	
Hiremath et al [34]	2019	USA	To use an mHealth-based PA measurement system to track PA levels of individuals with spinal cord injury in the community	Pilot study	• Population: people diagnosed with spinal cord injury who were at least 6 months postinjury • Age: mean 39.4 (SD 12.8) years • Setting: community	20
Pellegrini et al [35]	2015	USA	To examine the acceptability of a new technology developed to interrupt prolonged bouts of sedentary time among adults with T2D	Feasibility study	• Population: people diagnosed with T2D • Age: mean 53.1 (SD 10.7) years • Setting: community	8
Thomas et al [36]	2015	USA	To examine prompts and behavioral responses to the B-MOBILE JITAI for reducing sedentary behavior in overweight/obese individuals	Feasibility study	• Population: overweight/obese participants with BMI≥25 kg/m^2 • Age: mean 47.5 (SD 13.5) years • Setting: not reported	30

[a]EMA: ecological momentary assessment.

[b]T2D: type 2 diabetes.

[c]mHealth: mobile health.

[d]PA: physical activity.

[e]RCT: randomized controlled trial.

[f]JITAI: just-in-time adaptive intervention.

[g]CVD: cardiovascular disease.

[h]CHD: coronary heart disease.

Design Characteristics of JITAIs

Among the included studies, 13 reported on the design characteristics of JITAIs, with 2 studies using the same version of the JITAIs [31,32], resulting in a total of 12 unique JITAIs. Additionally, 8 of the JITAIs were grounded in theoretical frameworks [19,20,29,32-34,38,39], with social cognitive theory being the most commonly applied [32,34]. Overall, 6 JITAIs

XSL•FO
RenderX

[28,30,32-34,37] were designed to primarily target PA behavior, 1 focused solely on sedentary behavior [36], and 5 addressed both sedentary behavior and PA [20,29,35,38,39]. Most studies used wearable devices, smartphone apps, or both to collect PA or sedentary behavior data. Wearable or sensor-based data were used in studies (n=11) [20,28,29,31-36,38,38,39] using Fitbit or smartwatch devices, activity trackers, smartphone accelerometers, or external accelerometers. Several studies (n=4) also incorporated EMA or self-report data. Additional contextual data, such as location, weather, time of day, or day of week, were used in Hietbrink et al [19], Golbus et al [28], Daryabeygi-Khotbehsara et al [38], Mair et al [39], and Novak et al [20]. Two studies [31,32] also included app-entered clinical or health data, such as symptoms, blood pressure, heart rate, medication use, or self-reported health measures.

Tailoring variables could be grouped into 5 categories: behavioral variables, contextual variables, self-reported momentary states, goal/progress variables, and clinical or safety-related variables. Behavioral variables included step count, sedentary bout duration, walking status, activity intensity, and recent PA patterns [20,28,29,31-36,38,39]. Contextual variables, including time of day, weather, location, weekday/weekend, or study phase [19,20,28,38,39]. Self-reported momentary states, such as mood, fatigue, restrictions, cravings, readiness, and energy [19,33,34]. Goal- or progress-related tailoring was used in 6 studies [20,29,33,34,38,39]; for example, Novak et al [20] tailored messages according to recent step counts, weekly step goals, and individual action plans. Clinical or safety-related tailoring, such as symptom-triggered safety responses, was mainly reported in Elnaggar et al [31] and Park et al [32].

Decision points also differed across studies. Some interventions (n=4) [20,34-36] used sensor-triggered decision points, where prompts were delivered when a behavioral threshold was detected. For example, Pellegrini et al [35] delivered prompts after 20 minutes of consecutive sedentary time; Thomas et al [36] tested prompts after 30, 60, or 120 minutes of sedentary behavior; Novak et al [20] used Fitbit-detected inactivity or walking patterns; and Hiremath et al [34] triggered feedback when moderate-intensity wheelchair-based activity was detected. Other studies (n=5) [28,31,32,38,39] used scheduled or semi-scheduled decision points. Golbus et al [28] and Daryabeygi-Khotbehsara et al [38] also used microrandomized decision points, where participants could be randomized to receive or not receive an intervention message at predefined times or contexts. EMA-based decision or assessment points were used in Hietbrink et al [19] and Sengupta et al [33].

Intervention options could be grouped into PA prompts (n=9) [19,20,28,31-34,38,38,39] sedentary break prompts (n=5) [20,35,36,38,39], goal-setting or planning support (n=7) [20,28,29,33,34,38,39], feedback and encouragement (n=8) [19,20,29,31-34,39], education or coping support (n=5) [19,20,31-33], and safety prompts, which were specifically reported in Elnaggar et al [31] and Park et al [32], where participants reporting symptoms such as chest pain or shortness of breath received safety-related guidance. Decision rules were mostly simple and rule-based, which could be classified as EMA-based if-then rules, sensor-triggered if-then rules, sedentary-threshold rules, scheduled rules, microrandomization rules, and coach- or staff-supported rules. EMA-based if-then rules were used in 2 studies. Hietbrink et al [19] matched text messages to EMA-reported activity, mood, location, restrictions, weather, or cravings, while Sengupta et al [33] used EMA and goal information to trigger reminders or reinforcement messages. Four studies [20,34-36] adopted sensor-triggered or sedentary-threshold rules, where prompts were delivered after predefined behavioral states such as inactivity, low-intensity walking, moderate-intensity wheelchair activity, or prolonged sedentary time. Scheduled rules were used in 4 studies [20,31,32,39], with messages delivered at fixed daily or weekly times. Probability-based or availability-based microrandomization was used in Golbus et al [28] and Daryabeygi-Khotbehsara et al [38], where participants were randomized to receive or not receive messages at predefined decision points, sometimes conditional on availability. Four studies [29,31-33] used coach- or staff-supported decision rules, where participant data were reviewed by coaches or study staff to support tailored feedback, motivational messages, or safety guidance.

The adaptation mechanisms could be grouped into sensor-based tailoring, contextual tailoring, self-report-based tailoring, goal/progress-based tailoring, and coach- or staff-supported tailoring in this study. Sensor-based tailoring was used in 6 studies [20,34-36,38,39], which adapted messages according to step count, sedentary time, activity intensity, or walking patterns. Contextual tailoring based on time, weather, location, or day of week was used in 5 studies [19,20,28,38,39]. Two studies [19,33] adopted self-report-based tailoring, which used EMA responses such as mood, fatigue, restrictions, cravings, or readiness for activity. Goal/progress-based tailoring (n=6) [20,29,33,34,38,39] adapted feedback according to whether participants were meeting activity or sitting goals. Four studies [29,31-33] used coach- or staff-supported tailoring, which involved study staff or health coaches reviewing participant data and providing individualized feedback. Furthermore, a total of 12 studies reported intervention duration of JITAIs, which ranged from 7 days to 6 months, and was typically 2 to 12 weeks. Further details regarding the design characteristics of the JITAIs are provided in Table 3.

XSL•FO
RenderX

Table . Design characteristics of JITAIs[a].

Study	Theoretical framework	Target behavior	Sensors/data sources	Tailoring variables	Decision points	Intervention options	Decision rules	Prompt frequency/timing	Adaptation mechanisms	Intervention duration
Hietbrink et al [19], 2025	Behavior maintenance	PA[b]	EMA[c] self-report via SMS link/mobile phone. No passive wearable sensor was used. Participants completed EMAs on activity, location, mood, condition/restrictions, weather, and cravings. Baseline web-based survey measured self-efficacy and phase of behavior change.	Activity, location, mood, physical/mental restrictions, weather, cravings; plus baseline self-efficacy and phase of behavior change used to select suitable messages.	Each EMA prompt functioned as a decision point. EMAs were delivered at 1‐2 semirandom time points per day: once in the morning and once in the afternoon/evening.	Lifestyle support messages, including PA prompts, motivational/reflective messages, skill-building tips, and coping support	EMA-based if-then rules	EMA prompts 1‐2 times/day; intervention options delivered by text message, maximum 2 intervention messages/day.	Contextual tailoring and self-report-based tailoring	2 weeks
Golbus et al [28], 2024	Not reported	PA	Smartwatch data from Apple Watch or Fitbit Versa; mobile study app; step count measured after message delivery. Contextual data included weather, time of day, day of week, and time since enrollment.	Weather, temperature/precipitation, time of day, weekday/weekend, duration in study. Some messages could include preferred name, gain/loss framing, emoji, or dashboard hyperlink.	Participants were randomized 4 times/day to receive either no message or a PA-promoting message.	PA-promoting text messages or no message, with additional exercise planning, feedback, and self-monitoring components	Probability-based microrandomization	4 activity-message decision points/day over the trial; expected activity message rate ≈1/day due to 25% probability at each point.	Contextual tailoring	6 months

XSL•FO
RenderX

Study	Theoretical framework	Target behavior	Sensors/data sources	Tailoring variables	Decision points	Intervention options	Decision rules	Prompt frequency/timing	Adaptation mechanisms	Intervention duration
Novak et al [20], 2024	Self-regulation theory	PA and sedentary behavior	Fitbit Inspire 2 activity tracker; HealthReact platform; Fitbit step count, walking cadence proxy, heart rate to confirm wear, and syncing pattern.	Fitbit-detected walking/sitting patterns, step count, heart rate/wear status, recent step counts, weekly step goal progress, individual action plans/routines, baseline step goals, Fitbit syncing reliability, phone type, internet access/mobile data availability.	Main decision points included 5 min walking with 60 - 100 steps/min between 8 AM and 8 PM; 30 min with zero steps plus heart rate detection between 4 and 8 PM. Other decision points were individual routine-based reminders and fixed weekly times.	PA prompts, sedentary break prompts, action-plan reminders, goal review, feedback/encouragement, and health education	Sensor-triggered if-then rules plus scheduled rules	Expected 3 - 6 messages/week. Caps: Stand Up max 1/day; Walk Faster max 2/day with ≥60 min interval. Goal review: Friday 8 - 10 PM; Feedback/encouragement: Sunday 6 - 8 PM; Health education: Tuesday 6 - 8 PM.	Sensor-based tailoring, contextual tailoring, and goal/progress-based tailoring	2 weeks
Bai et al [29], 2022	Motivational interviewing and self-determination theory	PA and sedentary behavior	Fitbit Charge HR 3/Fitbit device for self-monitoring; ActiGraph wGT3X-BT used to assess MVPA[d], steps, and sedentary time.	Fitbit-derived activity progress; step counts; PA goals; goal review information used by the health coach.	Not reported.	Self-monitoring, goal setting/review, adaptive feedback	Coach-led goal/feedback rules	Not reported	Goal/progress-based tailoring	12 weeks

XSL•FO
RenderX

Study	Theoretical framework	Target behavior	Sensors/data sources	Tailoring variables	Decision points	Intervention options	Decision rules	Prompt frequency/timing	Adaptation mechanisms	Intervention duration
Daryabeygi-Khotbehsara et al [38], 2022	Behavior change techniques	PA and sedentary behavior	iMOVE Android app; SORD wearable sensor for real-time sitting, standing, and walking detection; Bluetooth connection; smartphone GPS for location; OpenWeather API for weather; Android time data; cloud server/database.	Sitting/standing/walking status; location, especially home vs workplace; weather conditions; time of day; weekday/weekend; personal sitting and walking goals.	Five daily decision windows included morning, lunchtime, afternoon 1, afternoon 2, and evening, from 8 AM to 8:30 PM. Randomization occurred only when participants were available: not driving, not sleeping, not already active, and connected to the intern.	"Sit less/move more" prompts, goal setting	Availability-based microrandomization	5 decision points/day; 70% intervention probability, split into 35% Sit Less and 35% Move More; 30% no-message control.	Sensor-based tailoring, contextual tailoring, and goal/progress-based tailoring	Not reported
Lau et al [30], 2022	Not reported	PA	Accelerometer plus EMA surveys; retrospective self-report assessments; acceptability questionnaire.	Daily activity and health-related symptoms: mood, cognitive complaints, fatigue, and pain.	Not reported.	Monitoring only	Not reported	8 EMA surveys/day for 7 days.	Not reported	7 days

XSL•FO
RenderX

Study	Theoretical framework	Target behavior	Sensors/data sources	Tailoring variables	Decision points	Intervention options	Decision rules	Prompt frequency/timing	Adaptation mechanisms	Intervention duration
Mair et al [39], 2022	Behavior Change Wheel	PA and sedentary behavior	Fitbit Charge 4; JitaBug smartphone app; Firebase repository; Fitbit data repository; OpenWeather API; smartphone GPS/home postcode for weather; accelerometer/activity tracker data; EMA for mood and well-being; voice memos for contextual PA information.	Real-time PA level from Fitbit, chosen activity goal, step-count vs activity-minute goal, time of day, daily goal status, and weather condition.	Server checked Fitbit data every hour between 9 AM and 8 PM to determine goal achievement; personalized JITAI messages were generated at 12:30 PM, 5:30 PM, and 8:30 PM using Fitbit, Firebase, and weather data.	Goal feedback, PA suggestions, activity-break prompts, planning reminders	Rule-based if-then rules	Hourly goal-achievement checks from 9 AM to 8 PM; personalized messages at 12:30 PM, 5:30 PM, and 8:30 PM.	Sensor-based tailoring, contextual tailoring, and goal/progress-based tailoring	6 weeks
Elnaggar et al [31], 2021	Not reported	PA	Fitbit Charge 2; Fitbit app; Movn mobile app; Fitabase; self-entered daily weight, blood pressure, heart rate, medication use, uncaptured exercise, and CVD[e] symptoms. Fitbit step count was the main PA data used.	Step count; PA progress; self-entered health data; cardiovascular symptom entries; staff knowledge of participant progress; participant-specific feedback needs.	Scheduled push-message times; Fitbit inactivity reminders after >2 hours of inactivity; symptom-triggered safety response when chest pain or shortness of breath was reported.	PA prompts, self-monitoring, education, symptom safety prompt	Scheduled + symptom/device-triggered rules	Push messages 3 times/week on random weekdays between 9 AM and 6 PM; Fitbit inactivity nudge after >2 hours of inactivity.	Coach- or staff-supported tailoring	2 months

XSL•FO
RenderX

Study	Theoretical framework	Target behavior	Sensors/data sources	Tailoring variables	Decision points	Intervention options	Decision rules	Prompt frequency/timing	Adaptation mechanisms	Intervention duration
Park et al [32], 2021	Social cognitive theory	PA	Fitbit Charge 2; Fitbit app; Movn mobile app; self-entered weight, blood pressure, heart rate, medication use, additional exercise, cardiovascular symptoms; and 6-minute walk test.	Step count, PA progress, self-entered health/activity data, cardiovascular symptoms, and study team–delivered motivational/educational content.	Scheduled push-message times; participant entries of health measures; symptom reporting through Movn; shortness of breath or chest pain triggered a safety prompt to call 911; study team triaged entries once daily.	Self-monitoring, PA prompts, education, symptom safety prompt	Scheduled + symptom/device-triggered rules	Push messages 3 times/week on random days between 9 AM and 6 PM; symptom safety prompt when chest pain or shortness of breath was reported.	Coach- or staff-supported tailoring	2 months
Sengupta et al [33], 2020	Behavior theory	PA	HerBeat smartphone app; Moto 360 smartwatch; smartwatch accelerometer/gyroscope-derived step count; heart rate; EMA surveys; cloud/server database; web-based health-coach dashboard. Data were refreshed on the dashboard every 10 min; smartwatch streamed step count and heart rate every 3 min from 6 AM to 10 PM.	Step count; distance walked; heart rate; walking goals; goal completion; readiness to begin PA; energy level; EMA-reported activity, location, mood, eating episodes, and social context; video use; Wi-Fi/Bluetooth disconnections.	EMA prompts 8 random times/day; goal-setting moments; progress-check moments; 4 PM check for no PA goal; goal-achievement points; health-coach dashboard review.	Goal setting, PA prompts, feedback, videos, coach messages	EMA/goal-based if-then rules	8 EMA prompts/day; 4 PM prompt if no PA goal; reinforcing message after goal achievement; coach message about once/week.	Self-report-based tailoring, goal/progress-based tailoring, and coach- or staff-supported tailoring	12 weeks

XSL•FO
RenderX

Study	Theoretical framework	Target behavior	Sensors/data sources	Tailoring variables	Decision points	Intervention options	Decision rules	Prompt frequency/timing	Adaptation mechanisms	Intervention duration
Hiremath et al [34], 2019	Social cognitive theory	PA	Android smartphone; PHIRE app; LG-Urbane smartwatch accelerometer; Bluetooth wheel-rotation monitor; EMA prompts; Firebase cloud storage; custom desktop visualization software; surveys on PA, pain, fatigue, and related information.	Real-time wheelchair-based PA classification, energy expenditure, distance traveled, moderate-intensity PA bouts, prior PA patterns, daily goal, minutes remaining to goal, and participant characteristics entered into the app, including weight, height, age, gender, and injury level.	PA classification and PA-level estimation occurred every minute. JITAI feedback was triggered when ≥3 continuous minutes of moderate-intensity or higher PA were detected; congratulatory messages continued every minute until activity stopped, and additional messages were sent when daily goals were reached or exceeded. EMA assessment prompts were delivered 6 times/day.	PA feedback, positive reinforcement, congratulatory messages, adaptive goal feedback	Sensor/ML[f]-detected PA if-then rules	JITAI prompts after ≥3 continuous minutes of moderate-intensity or higher PA; messages repeated every minute while activity continued; EMA prompts 6 times/day.	Sensor-based and tailoring and goal/progress-based tailoring	3 months

XSL•FO
RenderX

Study	Theoretical framework	Target behavior	Sensors/data sources	Tailoring variables	Decision points	Intervention options	Decision rules	Prompt frequency/timing	Adaptation mechanisms	Intervention duration
Pellegrini et al [35], 2015	Not reported	PA and sedentary behavior	NEAT! smartphone app on participants' Android phones; Shimmer wireless accelerometer worn at the waist for intervention triggering; ActiGraph 7164 accelerometer for sedentary behavior and PA assessment.	Consecutive sedentary time, sit-to-stand transition, active vs sedentary state, participant prompt response, and participant preference for noise or vibration prompt.	A decision point occurred when 20 minutes of consecutive sedentary time was detected. The sedentary counter restarted when a sit-to-stand transition was detected or the participant responded to the phone prompt. If the participant selected Stand but did not stand, reminders were repeated every 2 minutes.	Sedentary break prompt, stand/light-activity reminder, response options	Sedentary-bout-triggered if-then rules	Prompt after 20 minutes of consecutive sedentary time; additional reminders every 2 minutes if the user selected "Stand" but did not stand.	Sensor-based and tailoring	1 month

XSL•FO
RenderX

Study	Theoretical framework	Target behavior	Sensors/data sources	Tailoring variables	Decision points	Intervention options	Decision rules	Prompt frequency/timing	Adaptation mechanisms	Intervention duration
Thomas et al [36], 2015	Not reported	Sedentary behavior	Android smartphone; B-MOBILE app; smartphone onboard accelerometer; validated algorithm classifying behavior as sedentary behavior or nonsedentary behavior in 1-minute epochs; SenseWear Mini Armband/objective multisensor monitor for assessment in the B-MOBILE study.	Continuous sedentary minutes, sedentary behavior vs nonsedentary behavior state, walking-break condition, walking-break completion, total daily reminders after 5 and 10 minutes of continued sedentary behavior minutes, total daily PA minutes, and number of walking prompts met.	Participants reached condition-specific sedentary thresholds: 30, 60, or 120 continuous sedentary minutes. The app also triggered reminders after 5 and 10 minutes of continued sedentary behavior after a walking prompt.	Walking-break prompts, reminder prompts, feedback display	Sedentary-threshold if-then rules	Three conditions were tested: a 3-min walking break after 30 min of sedentary behavior, a 6-min walking break after 60 min of sedentary behavior, or a 12-min walking break after 120 min of sedentary behavior; additional reminders were delivered after 5 and 10 min if sedentary behavior continued.	Sensor-based and tailoring	21 days

[a]JITAI: just-in-time adaptive intervention.

[b]PA: physical activity.

[c]EMA: ecological momentary assessment.

[d]MVPA: moderate to vigorous physical activity.

[e]CVD: cardiovascular disease.

[f]ML: machine learning.

Engagement and implementation-related measures were inconsistently reported across studies. Reported measures included study retention, wearable-device use, EMA completion, app use, goal setting, message exposure and responsiveness, intervention-delivery fidelity, usability, and satisfaction. Four studies [33,34,36,39] reported quantitative engagement and delivery-fidelity data. Hietbrink et al [19] and Daryabeygi-Khotbehsara et al [38] mainly reported qualitative acceptability or development-stage feedback, while Golbus et al [28], Bai et al [29], Elnaggar et al [31], and Park et al [32] provided limited objective usage or response data. Further details are presented in Multimedia Appendix 2.

Effectiveness of JITAIs in Reducing Sedentary Behavior and Promoting PA Among Patients With Chronic Disease

Of the 14 included studies, 7 objectively measured PA. Three RCTs [28,29,32] reported preliminary evidence of a positive effect of JITAIs on promoting PA. In the study by Golbus et al [28], participants' 60-minute step counts significantly increased within the first 30 days following the intervention. In the study by Park et al [32], the intervention group had a higher mean daily step count compared to the control group (P=.02). Similarly, Bai et al [29] found that participants in the Fitbit plus health coaching group increased their daily step count (P<.01) and moderate to vigorous physical activity (MVPA) (P<.05), with an improvement in daily MVPA compared to those in the Fitbit-only group (P<.05). Four nonrandomized studies reported PA outcomes, with three demonstrating positive effects [30,34,35]. Studies by Hiremath et al [34] and Pellegrini et al [35] reported increases in PA levels compared to baseline (both P<.05). The study by Lau et al [30] also showed that a higher proportion of light and moderate to vigorous PA was recorded among participants receiving JITAIs. However, the study conducted by Sengupta et al [33] found no significant change in PA levels compared to baseline.

Three studies [29,35,36] objectively measured sedentary behavior, including 1 RCT and 2 nonrandomized studies. In the RCT conducted by Bai et al [29], participants in the Fitbit plus health coaching group preliminarily demonstrated a reduction in sitting time (P<.01). Among the 2 nonrandomized studies, both reported a reduction in sedentary time or behavior compared with baseline (P<.05) [35,36]. Further details are provided in Multimedia Appendix 3.

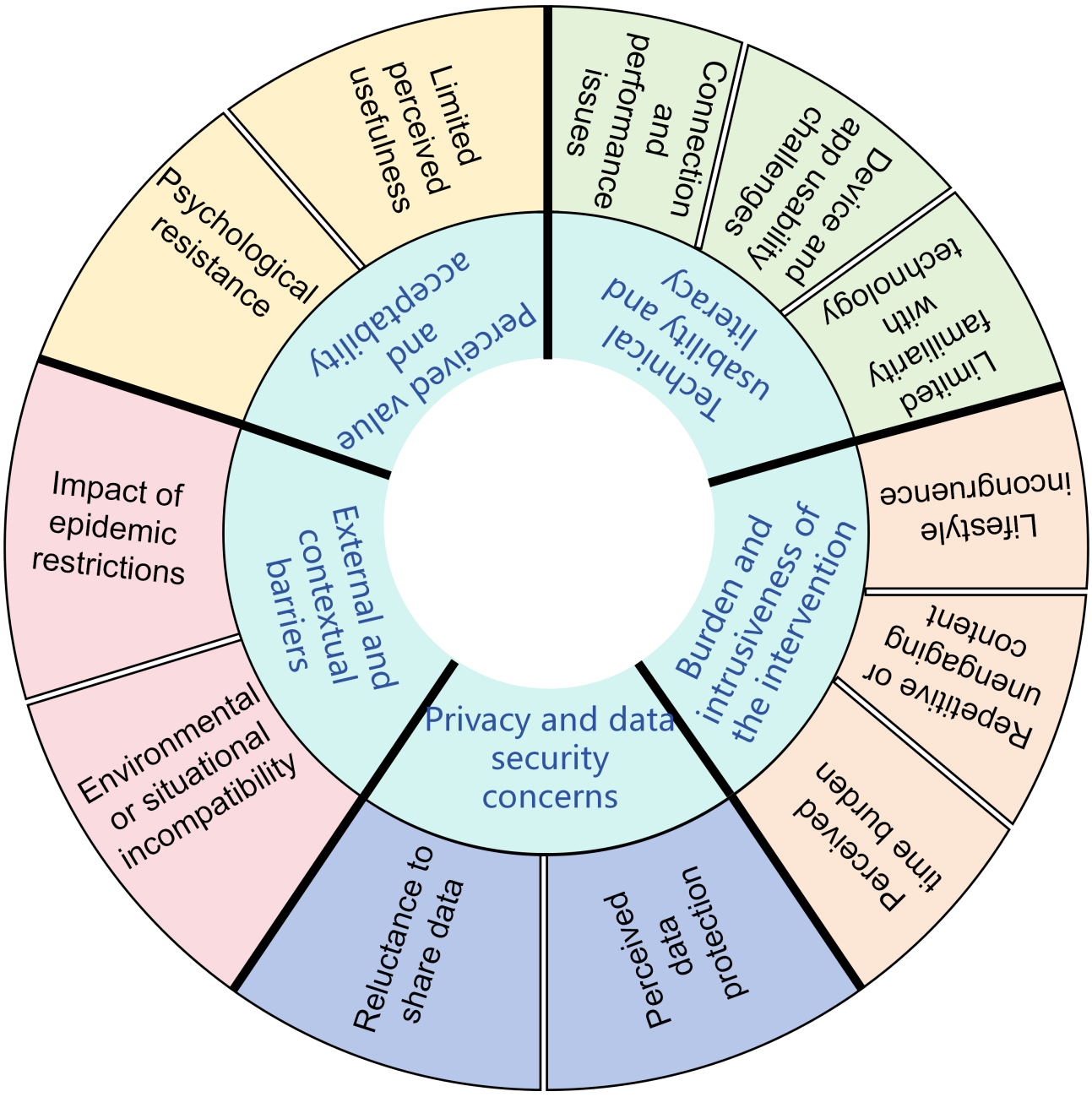
XSL•FO
RenderX

Barriers and Facilitators of JITAIs in Promoting PA and Reducing Sedentary Behavior Among Patients With Chronic Disease

Barriers Related to Engagement and Use of JITAIs

Several barriers were identified that may hinder the use of JITAIs in promoting PA and reducing sedentary behavior among patients with chronic diseases [19,20,31,34,37-39] (Figure 2). These barriers were categorized into five overarching themes: (1) technological usability and literacy, including limited familiarity with technology, device and app usability challenges such as small text or rigid settings, connection and performance issues such as devices unable to detect low-intensity or short-duration activities; (2) burden and intrusiveness of the intervention, including perceived time burden, repetitive or unengaging content, and lifestyle incongruence; (3) perceived value and acceptability, including limited perceived usefulness and psychological resistance; (4) external and contextual barriers, including the impact of epidemic restrictions and environmental or situational incompatibility; and (5) privacy and data security concerns, including reluctance to share data and perceived data protection.

Figure 2. Barriers related to engagement and use of just-in-time adaptive interventions (JITAIs).



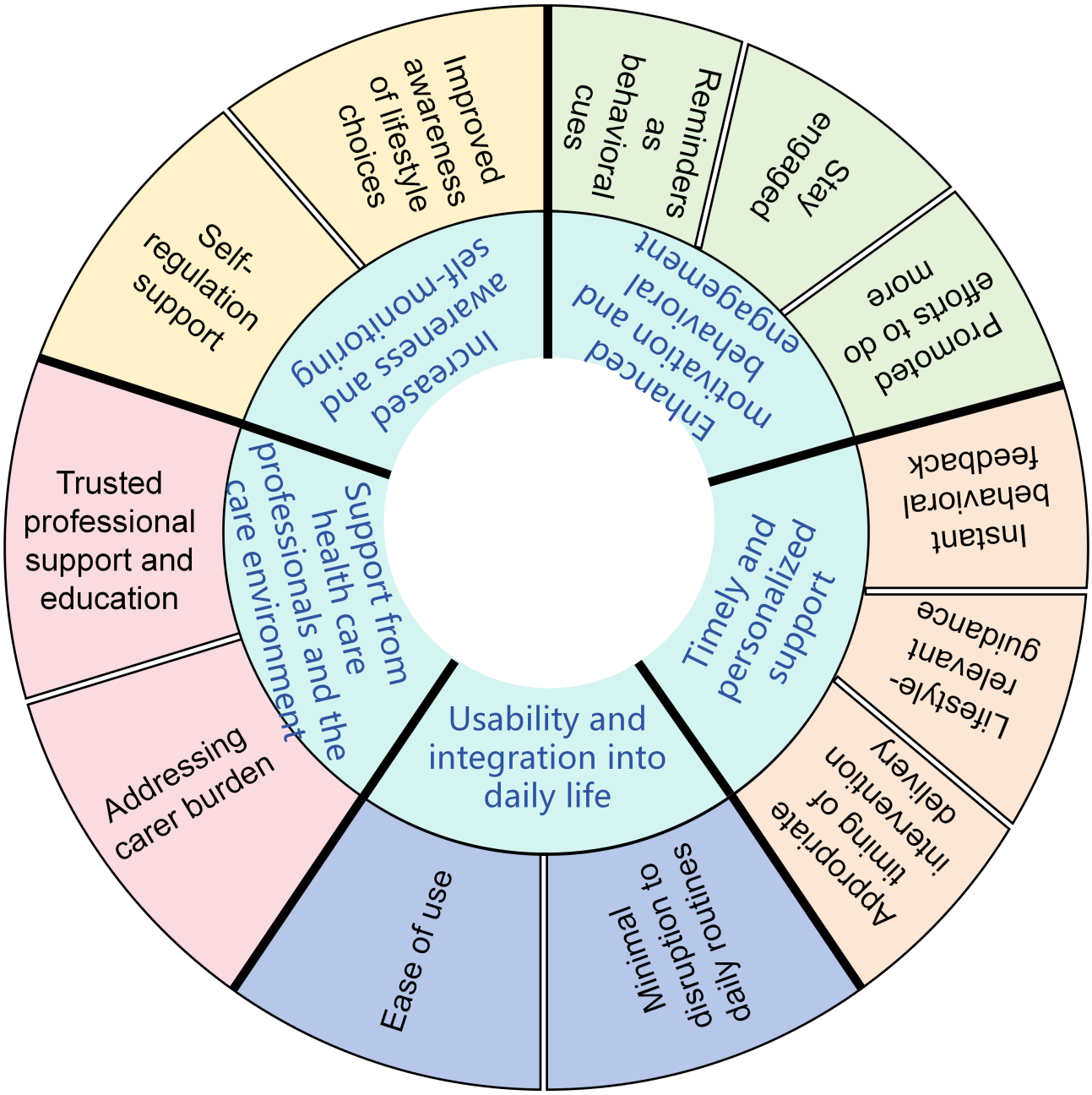
Facilitators Related to Engagement and Use of JITAIs

The facilitators identified across the six included studies [19,31,34,37-39] were synthesized into five major themes: (1) enhanced motivation and behavioral engagement, including promoted efforts to "do more," stay engaged and reminders as behavioral cues (eg, just-in-time reminders prompted timely action and supported formation of new routines and habits); (2) increased awareness and self-monitoring, including improved awareness of lifestyle choices and self-regulation support; (3)

XSL•FO
RenderX

usability and integration into daily life, including ease of use and minimal disruption to daily routines; (4) timely and personalized support, including appropriate timing of intervention delivery (eg, participants felt supported "in the moment"), lifestyle-relevant guidance and instant behavioral feedback; and (5) support from health care professionals and the care environment, including trusted professional support and education, and addressing caregiver burden (Figure 3).

Figure 3. Facilitators related to engagement and use of just-in-time adaptive interventions (JITAIs).



Discussion

Principal Findings

To the best of our knowledge, this is the first scoping review to systematically map and synthesize the existing literature on JITAIs specifically aimed at reducing sedentary behavior and promoting PA in individuals with chronic disease. A total of 14 studies were finally included in this review, and we systematically summarized the key characteristics of delivery methods and components used in JITAIs designed to reduce sedentary behavior and promote PA, and explored the existing evidence regarding the evaluation of these interventions in improving PA levels and reducing sedentary time. Additionally, this review also identified reported barriers and facilitators associated with the use of JITAIs among individuals living with chronic diseases. The findings contribute to a clearer understanding of how JITAIs are currently operationalized and provide a foundation for guiding future research and intervention development in chronic disease management.

Among the included studies, most of them provided participants with instructions on how to operate these technologies. As the integration of wearable activity trackers becomes more common, user training and support are essential to maximize engagement and minimize technological barriers [20,28]. The widespread

XSL•FO
RenderX

use of built-in accelerometers and GPS sensors indicates that sensor-based context detection is a central feature of current JITAIs [30,35]. By identifying timely opportunities for intervention, these technologies may support more personalized and responsive behavioral prompts. Additionally, the intervention options and decision rules identified in this review suggest that most JITAIs used relatively simple designs. The frequent use of PA prompts, goal-setting or planning support, and feedback/encouragement indicates that most interventions focused on motivating participants to initiate or maintain PA, whereas fewer studies specifically targeted sedentary breaks. Safety prompts were reported only in Elnaggar et al [31] and Park et al [32], suggesting that symptom- or risk-related tailoring remains limited despite its relevance for patients with chronic disease. Decision rules were also mostly predefined and rule-based, which, although practical and transparent, indicates that many current JITAIs are adaptive mainly to predefined behavioral or contextual triggers rather than continuously learning from individual responses over time. Future research could therefore explore more dynamic and data-driven adaptation strategies that can continuously learn from users' evolving behaviors and intervention responses.

In this review, most included studies reported preliminary evidence of a positive effect of JITAIs in promoting PA among patients with chronic disease. JITAIs provide personalized, real-time support based on individual context, such as location, time of day, or recent activity levels, which may enhance the relevance and timeliness of prompts and promote behavioral engagement [15,16]. The use of wearable sensors and mobile apps may also support continuous monitoring and feedback, potentially enhancing motivation and self-awareness [15,16]. In addition, many interventions incorporated behavior change techniques such as goal-setting, reminders, and feedback, which have been shown to support PA improvement [40]. However, the findings from one study [33] found no significant change in PA levels compared to baseline. One possible explanation for the lack of significant change in PA levels in the study could be the exclusive inclusion of female participants. Factors such as caregiving responsibilities, differing motivational drivers, and social or cultural expectations may influence women's engagement with PA interventions [41]. Three studies [29,35,36] reported preliminary evidence of a positive effect of JITAIs in reducing sitting time and promoting breaks in sedentary behavior. The effects may be related to the real-time and context-sensitive design of JITAIs, which allows prompts or support to be delivered when individuals are likely to be sedentary, such as after prolonged inactivity or during sitting-related contexts, including watching television or working at a desk [29,35,36]. The timely cues may help interrupt habitual sedentary patterns and encourage individuals to take breaks, stand, or engage in light activity. However, the overall evidence base for JITAIs in promoting PA and reducing sedentary behavior remains limited by small sample sizes, few RCTs, and short intervention durations. In this review, only one study [28] had an intervention duration of 6 months, whereas the remaining studies involved relatively shorter intervention periods. The limited durations may not be sufficient to determine whether JITAIs can produce sustained behavioral changes or maintain long-term user engagement. Future research could therefore use RCT designs adequately, include longer intervention and follow-up periods, and recruit more diverse chronic disease populations across different clinical and community settings to further establish the effectiveness of JITAIs.

The identification of multiple barriers to JITAI use among patients with chronic diseases highlights important challenges that need to be addressed to optimize their utility. Most studies reported that technological usability and literacy emerged as a significant obstacle. Issues such as small text sizes, rigid app settings, and unreliable detection of low-intensity or brief activities can compromise user experience and reduce engagement, aligning with previous findings, which emphasize the digital divide, particularly among older or less technologically experienced populations commonly affected by chronic conditions [42,43]. Future interventions can adopt inclusive design principles that accommodate users with varying levels of digital literacy. This involves simplifying user interfaces, enabling customizable text sizes and display settings, and ensuring device compatibility with a wide range of PA levels, particularly low-intensity movements common among individuals with chronic diseases. Incorporating user-centered design through participatory development processes may enhance both usability and acceptability by aligning intervention features with patient needs and preferences [31,37]. Additionally, the burden and intrusiveness of the intervention represent another critical barrier. Patients reported perceptions of excessive time demands, repetitive or not engaging content, and misalignment with their daily routines, which may undermine adherence [20,34,38,39]. The results underscore the necessity for interventions that are not only effective but also seamlessly integrate into users' lifestyles, minimizing disruption and maximizing relevance. Moreover, perceived value and acceptability also influence intervention uptake. Limited perceived usefulness and psychological resistance, potentially stemming from skepticism toward new technologies or concerns about personal capability, suggest that patient education and motivational strategies are crucial components for successful implementation [19]. Privacy and data security concerns were reported in several studies, reflecting widespread apprehension regarding data sharing and protection, similar to the findings from Muhunzi et al [44]. Future interventions should ensure transparent data policies and robust security measures while granting users control over their personal information to build trust, enhance sustained engagement, and support the further development of JITAIs.

Limitations

This scoping review has several limitations that should be acknowledged. First, although a comprehensive search strategy was used, some relevant studies may have been missed due to publication bias or limitations in database indexing. In addition, only peer-reviewed articles published in English were included, which may have excluded relevant studies published in other languages or gray literature. Second, the heterogeneity of the included studies in terms of population characteristics, intervention designs, and duration limited our ability to quantitatively synthesize findings or make direct comparisons across studies. Third, many included studies had small sample sizes and short follow-up periods, which may limit the

XSL•FO
RenderX

generalizability of the findings and make it difficult to determine whether the reported effects can be sustained over time. Additionally, most studies were conducted in high-income countries, which may limit the applicability of findings to low- and middle-income settings. Therefore, the conclusions should be interpreted with caution. Future systematic reviews with meta-analyses and risk-of-bias assessments are warranted to more rigorously evaluate the effectiveness of JITAIs in promoting PA and reducing sedentary behavior.

Conclusions

This scoping review identified preliminary evidence suggesting that JITAIs may help reduce sedentary behavior and promote PA among patients with chronic diseases. However, the evidence supporting this conclusion remains limited by the small number of RCTs, small sample sizes, short intervention and follow-up periods, and heterogeneity across studies. Therefore, these findings should be interpreted as preliminary. Future research could use RCT designs with sufficient sample sizes, longer follow-up periods, and more diverse chronic disease populations to further establish the effectiveness and generalizability of JITAIs.

Funding

This work was supported by the Hunan Provincial Natural Science Foundation (2026JJ81732) and Scientific Climbing Project of Hunan Cancer Hospital (YF2020010).

Data Availability

The data used during this study are available from the corresponding authors on reasonable request.

Authors' Contributions

Conceptualization: JG, FC, YD, YC, MH
Formal analysis: JG, FC, CC, MH
Methodology: JG, FC, CC, YD, YC, XX, MH
Project administration: YC
Supervision: YC
Validation: YD, YC
Visualization: JG, FC, XX
Writing – original draft: JG
Writing – review & editing: FC, CC, YD, YC, MH

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 32 KB - mhealth_v14i1e81378_app1.docx]

Multimedia Appendix 2
Engagement and implementation-related measures of just-in-time adaptive interventions.
[DOCX File, 18 KB - mhealth_v14i1e81378_app2.docx]

Multimedia Appendix 3
Effectiveness of just-in-time adaptive interventions among patients with chronic disease.
[DOCX File, 17 KB - mhealth_v14i1e81378_app3.docx]

Checklist 1
PRISMA-ScR checklist.
[PDF File, 677 KB - mhealth_v14i1e81378_app4.pdf]

References

1. World health statistics 2023: monitoring health for the SDGs, sustainable development goals. World Health Organization. 2023. URL: https://www.who.int/publications/i/item/9789240074323 [accessed 2026-07-20]
2. Alves AJ, Viana JL, Cavalcante SL, et al. Physical activity in primary and secondary prevention of cardiovascular disease: overview updated. World J Cardiol 2016 Oct 26;8(10):575-583. [doi: 10.4330/wjc.v8.i10.575] [Medline: 27847558]

XSL•FO
RenderX

3. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep 1985;100(2):126-131. [Medline: 3920711]
4. Tremblay MS, Aubert S, Barnes JD, et al. Sedentary Behavior Research Network (SBRN) – Terminology Consensus Project process and outcome. Int J Behav Nutr Phys Act 2017 Jun 10;14(1):75. [doi: 10.1186/s12966-017-0525-8] [Medline: 28599680]
5. Asfaw MS, Dagne WK. Physical activity can improve diabetes patients' glucose control: a systematic review and meta-analysis. Heliyon 2022 Dec;8(12):e12267. [doi: 10.1016/j.heliyon.2022.e12267] [Medline: 36578408]
6. Black SV, Cooper R, Martin KR, Brage S, Kuh D, Stafford M. Physical activity and mental well-being in a cohort aged 60-64 years. Am J Prev Med 2015 Aug;49(2):172-180. [doi: 10.1016/j.amepre.2015.03.009] [Medline: 26070782]
7. Wang X, Li Y, Fan H. The associations between screen time-based sedentary behavior and depression: a systematic review and meta-analysis. BMC Public Health 2019 Nov 14;19(1):1524. [doi: 10.1186/s12889-019-7904-9] [Medline: 31727052]
8. Zink J, Belcher BR, Imm K, Leventhal AM. The relationship between screen-based sedentary behaviors and symptoms of depression and anxiety in youth: a systematic review of moderating variables. BMC Public Health 2020 Apr 10;20(1):472. [doi: 10.1186/s12889-020-08572-1] [Medline: 32272906]
9. Silveira EA, Mendonça CR, Delpino FM, et al. Sedentary behavior, physical inactivity, abdominal obesity and obesity in adults and older adults: a systematic review and meta-analysis. Clin Nutr ESPEN 2022 Aug;50:63-73. [doi: 10.1016/j.clnesp.2022.06.001] [Medline: 35871953]
10. Wang Q, Huang XT, Tang SY, Huang CM. The application progress of just-in-time adaptive intervention in nursing. Chin Jour Nurs 2024;59(4):490-495. [doi: 10.3761/j.issn.0254-1769.2024.04.016]
11. Global strategy on digital health 2020-2025. World Health Organization. 2021 Aug 18. URL: https://www.who.int/publications/i/item/9789240020924 [accessed 2026-07-20]
12. Yerrakalva D, Yerrakalva D, Hajna S, Griffin S. Effects of mobile health app interventions on sedentary time, physical activity, and fitness in older adults: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e14343. [doi: 10.2196/14343] [Medline: 31778121]
13. Daryabeygi-Khotbehsara R, Shariful Islam SM, Dunstan D, McVicar J, Abdelrazek M, Maddison R. Smartphone-based interventions to reduce sedentary behavior and promote physical activity using integrated dynamic models: systematic review. J Med Internet Res 2021 Sep 13;23(9):e26315. [doi: 10.2196/26315] [Medline: 34515637]
14. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: a pragmatic framework. Health Psychol 2015 Dec;34(Suppl):1209-1219. [doi: 10.1037/hea0000306] [Medline: 26651462]
15. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019 Apr 3;16(1):31. [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
16. Hsu TCC, Whelan P, Gandrup J, Armitage CJ, Cordingley L, McBeth J. Personalized interventions for behaviour change: a scoping review of just-in-time adaptive interventions. Br J Health Psychol 2025 Feb;30(1):e12766. [doi: 10.1111/bjhp.12766] [Medline: 39542743]
17. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
18. van Genugten CR, Thong MSY, van Ballegooijen W, et al. Beyond the current state of just-in-time adaptive interventions in mental health: a qualitative systematic review. Front Digit Health 2025;7(7):1460167. [doi: 10.3389/fdgth.2025.1460167] [Medline: 39935463]
19. Hietbrink EAG, Middelweerd A, d'Hollosy W, Schrijver LK, Laverman GD, Vollenbroek-Hutten MMR. Exploring the acceptance of just-in-time adaptive lifestyle support for people with type 2 diabetes: qualitative acceptability study. JMIR Form Res 2025 Feb 19;9:e65026. [doi: 10.2196/65026] [Medline: 39969969]
20. Novak J, Jurkova K, Lojkaskova A, et al. Participatory development of an mHealth intervention delivered in general practice to increase physical activity and reduce sedentary behaviour of patients with prediabetes and type 2 diabetes (ENERGISED). BMC Public Health 2024 Mar 31;24(1):927. [doi: 10.1186/s12889-024-18384-2] [Medline: 38556892]
21. Golbus JR, Dempsey W, Jackson EA, Nallamothu BK, Klasnja P. Microrandomized trial design for evaluating just-in-time adaptive interventions through mobile health technologies for cardiovascular disease. Circ Cardiovasc Qual Outcomes 2021 Feb;14(2):e6760. [doi: 10.1161/CIRCOUTCOMES.120.006760]
22. Peters MDJ, Marnie C, Tricco AC, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020 Oct;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
23. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
24. Goodman RA, Posner SF, Huang ES, Parekh AK, Koh HK. Defining and measuring chronic conditions: imperatives for research, policy, program, and practice. Prev Chronic Dis 2013 Apr 25;10:E66. [doi: 10.5888/pcd10.120239] [Medline: 23618546]

XSL•FO
RenderX

25. Riegel B, Westland H, Iovino P, et al. Characteristics of self-care interventions for patients with a chronic condition: a scoping review. Int J Nurs Stud 2021 Apr;116:103713. [doi: 10.1016/j.ijnurstu.2020.103713] [Medline: 32768137]
26. Li Y, Lu H, Zhao Y, et al. Core competencies of the midwifery workforce in China: a scoping review. J Nurs Manag 2022 Mar;30(2):535-558. [doi: 10.1111/jonm.13531] [Medline: 34894017]
27. Klooster E, Koenders N, Vermeulen-Holsen J, Vos L, van der Wees PJ, Hoogeboom TJ. Healthcare professionals feel empowered by implementing a hospital-based multifaceted intervention: a qualitative study using inductive thematic analysis. BMC Health Serv Res 2022 Jul 12;22(1):903. [doi: 10.1186/s12913-022-08310-w] [Medline: 35820839]
28. Golbus JR, Shi J, Gupta K, et al. Text messages to promote physical activity in patients with cardiovascular disease: a micro-randomized trial of a just-in-time adaptive intervention. Circ Cardiovasc Qual Outcomes 2024 Jul;17(7):e10731. [doi: 10.1161/CIRCOUTCOMES.123.010731]
29. Bai Y, Burns R, Gell N, Byun W. A randomized trial to promote physical activity in adult pre-hypertensive and hypertensive patients. J Sports Sci 2022 Jul;40(14):1648-1657. [doi: 10.1080/02640414.2022.2099179] [Medline: 35830497]
30. Lau SCL, Connor LT, King AA, Baum CM. Multimodal ambulatory monitoring of daily activity and health-related symptoms in community-dwelling survivors of stroke: feasibility, acceptability, and validity. Arch Phys Med Rehabil 2022 Oct;103(10):1992-2000. [doi: 10.1016/j.apmr.2022.06.002]
31. Elnaggar A, von Oppenfeld J, Whooley MA, Merek S, Park LG. Applying mobile technology to sustain physical activity after completion of cardiac rehabilitation: acceptability study. JMIR Hum Factors 2021 Sep 2;8(3):e25356. [doi: 10.2196/25356] [Medline: 34473064]
32. Park LG, Elnaggar A, Lee SJ, et al. Mobile health intervention promoting physical activity in adults post cardiac rehabilitation: pilot randomized controlled trial. JMIR Form Res 2021 Apr 16;5(4):e20468. [doi: 10.2196/20468] [Medline: 33861204]
33. Sengupta A, Beckie T, Dutta K, Dey A, Chellappan S. A mobile health intervention system for women with coronary heart disease: usability study. JMIR Form Res 2020 Jun 3;4(6):e16420. [doi: 10.2196/16420] [Medline: 32348270]
34. Hiremath SV, Amiri AM, Thapa-Chhetry B, et al. Correction: Mobile health-based physical activity intervention for individuals with spinal cord injury in the community: a pilot study. PLoS One 2019;14(11):e0225490. [doi: 10.1371/journal.pone.0225490] [Medline: 31725792]
35. Pellegrini CA, Hoffman SA, Daly ER, Murillo M, Iakovlev G, Spring B. Acceptability of smartphone technology to interrupt sedentary time in adults with diabetes. Transl Behav Med 2015 Sep;5(3):307-314. [doi: 10.1007/s13142-015-0314-3] [Medline: 26327936]
36. Thomas JG, Bond DS. Behavioral response to a just-in-time adaptive intervention (JITAI) to reduce sedentary behavior in obese adults: implications for JITAI optimization. Health Psychol 2015 Dec;34(Suppl):1261-1267. [doi: 10.1037/hea0000304] [Medline: 26651467]
37. Cardy N, Hunter A, Carter D, et al. Perspectives of people with stroke, caregivers and healthcare professionals on an adaptive mHealth intervention for physical activity in the prevention of secondary stroke: a qualitative study. J Multidiscip Healthc 2024;17:2677-2688. [doi: 10.2147/JMDH.S420530] [Medline: 38831803]
38. Daryabeygi-Khotbehsara R, Shariful Islam SM, Dunstan DW, et al. Development of an Android mobile application for reducing sitting time and increasing walking time in people with type 2 diabetes. Electronics (Basel) 2022;11(19):3011. [doi: 10.3390/electronics11193011]
39. Mair JL, Hayes LD, Campbell AK, Buchan DS, Easton C, Sculthorpe N. A personalized smartphone-delivered just-in-time adaptive intervention (JitaBug) to increase physical activity in older adults: mixed methods feasibility study. JMIR Form Res 2022 Apr 7;6(4):e34662. [doi: 10.2196/34662] [Medline: 35389348]
40. Tian SY, Chang H. Research progress on the application of just-in-time adaptive intervention in chronic illness sedentary and physical activity. Chin J Nurs 2025;60(5):619-623. [doi: 10.3761/j.issn.0254-1769.2025.05.016]
41. Ominyi J, Clifton A, Nwedu A. Understanding physical activity participation among underserved women: a mixed-methods cross sectional study using an ecological framework. BMC Public Health 2025 Jul 2;25(1):2178. [doi: 10.1186/s12889-025-23189-y] [Medline: 40604521]
42. Wang Y, Wu Z, Duan L, et al. Digital exclusion and cognitive impairment in older people: findings from five longitudinal studies. BMC Geriatr 2024;24(1):406. [doi: 10.1186/s12877-024-05026-w]
43. Park Y, Kim EJ, Park S, Lee M. Digital health intervention effect on older adults with chronic diseases living alone: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2025 Mar 31;27:e63168. [doi: 10.2196/63168] [Medline: 40163849]
44. Muhunzi D, Kitambala L, Mashauri HL. Big data analytics in the healthcare sector: opportunities and challenges in developing countries. A literature review. Health Informatics J 2024;30(4):1217779593. [doi: 10.1177/14604582241294217] [Medline: 39434249]

Abbreviations

EMA: ecological momentary assessment
JBI: Joanna Briggs Institute
JITAI: just-in-time adaptive intervention

XSL•FO
RenderX

mHealth: mobile health
MVPA: moderate to vigorous physical activity
PA: physical activity
PCC: population-concept-context framework
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
RCT: randomized controlled trial
T2D: type 2 diabetes

Edited by G Badicu; submitted 27.Jul.2025; peer-reviewed by A Gimeno-Miguel, I Jimoh, S Fenton, Z Zandesh; revised version received 22.Jun.2026; accepted 26.Jun.2026; published 07.Aug.2026.

Please cite as:
Guo J, Chen F, Chen C, Dai Y, Xu X, Chen Y, Huang M
Just-in-Time Adaptive Interventions to Reduce Sedentary Behavior and Promote Physical Activity in Patients With Chronic Disease: Scoping Review
JMIR Mhealth Uhealth 2026;14:e81378
URL: https://mhealth.jmir.org/2026/1/e81378
doi:10.2196/81378

© Junchen Guo, Furong Chen, Chan Chen, Yunyun Dai, Xianghua Xu, Yongyi Chen, Mengying Huang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 7.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Patient-Reported and Technology-Assisted Monitoring in Orthotic Management of Adolescents With Idiopathic Scoliosis: Scoping Review

Anna Mathew[1], MSc; Kamila Sykorova[1], MPT; Nenad Pavel[2], PhD; Marianne Bakke Johnsen[1], PhD; Minna Pikkarainen[1,2], PhD; Sadeeq Ali[3], PhD; Mette Fløystad Kvammen[4,5], MSW; Parisa Gazerani[6], PhD

[1]Department of Rehabilitation Science and Health Technology, Faculty of Health Sciences, OsloMet – Oslo Metropolitan University, Oslo, Norway

[2]Department of Product Design, Faculty of Technology, Art and Design, OsloMet – Oslo Metropolitan University, Kjeller, Oslo, Norway

[3]Prosthetics and Orthotics, Department of Rehabilitation Science and Health Technology, Faculty of Health Sciences, OsloMet – Oslo Metropolitan University, Oslo, Norway

[4]Department of Sociology and Social Work, University of Agder, Kristiansand, Agder, Norway

[5]Norwegian Spine and Back Pain Association, Oslo, Norway

[6]ADEPT Research Group, Department of Life Sciences and Health, Faculty of Health Sciences, OsloMet – Oslo Metropolitan University, Pilestredet 50, Oslo, Norway

Corresponding Author:
Parisa Gazerani, PhD
ADEPT Research Group, Department of Life Sciences and Health, Faculty of Health Sciences, OsloMet – Oslo Metropolitan University, Pilestredet 50, Oslo, Norway

Abstract

Background: Orthotic treatment for adolescents diagnosed with idiopathic scoliosis is long-term and often associated with challenges related to treatment adherence. Traditional patient monitoring strategies include clinical visits combined with patient discussions, clinical examination reports, and the use of standard questionnaires. Sensor-assisted monitoring is explored to improve treatment outcomes, with technology influencing different aspects of the health care system.

Objective: The scoping review aimed to map evidence based on 3 research questions (RQs), namely user experience and perception, monitoring tools, and technology-assisted approaches, including smart orthoses with sensors, during orthotic treatment for an adolescent patient (aged 10 to 17.11 y) diagnosed with idiopathic scoliosis.

Methods: A systematic search was conducted in MEDLINE, Embase, PsycINFO, Cochrane Library, CINAHL, Web of Science, and Scopus as 2 complementary search components. The primary search targeted studies on smart orthoses, also called braces, and scoliosis, while the secondary search focused on user experiences with orthotic treatment to capture all 3 RQs. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-analyses extension for Scoping Reviews) guidelines and checklist supported searching, analyzing, and reporting the results systematically. Predefined inclusion and exclusion criteria were followed for screening. All articles published up to May 2024 were included. Moreover, 2 independent reviewers screened 2088 articles in 2 stages of screening. Data coding was conducted using the 3 RQs by data extraction and its synthesis.

Results: A total of 88 articles met the inclusion criteria. Evidence from user experience studies highlighted discomfort, psychosocial burden, and demand for real-time feedback. Monitoring relied primarily on patient-reported outcome measures (PROMs) such as questionnaires and logbooks. Culturally adapting to patients' diverse global needs helps capture richer and more candid user perspectives. Sensor-based systems used temperature, force, or pressure, electromyography, and motion sensors to track orthosis wear time, pressure distribution, activity, and posture. Two overarching monitoring themes emerged: subjective (patient-reported or user experience) and objective (technology-assisted or sensor-based) monitoring. Integrated approaches synthesize both.

Conclusions: Subjective clinical tools and objective technologies help monitor orthosis wear of adolescent patients with idiopathic scoliosis. Co-design of a user-centered scoliosis management system for adolescents that integrates PROMs with measured sensor data could provide a multidimensional view of orthosis adherence along with psychosocial and clinical effectiveness. Future research will focus on user-centered care, with real-time integrated monitoring, to provide clinically meaningful feedback that motivates adolescents to achieve their treatment goals.

(JMIR Mhealth Uhealth 2026;14:e88880) doi:10.2196/88880

XSL•FO
RenderX

KEYWORDS

braces; health technology; health information systems; orthotic devices; patient monitoring; physiologic; sensors; scoliosis; surveys and questionnaires; self report; wearable electronic devices

Introduction

Background

Scoliosis refers to the deformity of the spine, thorax, and trunk. The root cause of scoliosis can differ. About 80% of cases are classified broadly by etiology as idiopathic scoliosis [1,2]. The International Scientific Society on Scoliosis Orthopedic and Rehabilitation Treatment reports that idiopathic scoliosis is diagnosed in seemingly healthy children and tends to progress rapidly at puberty [3].

Nonsurgical treatment options for scoliosis management include observation, physiotherapeutic-specific exercises, special inpatient rehabilitation, and bracing [3]. Bracing combined with exercise is an effective conservative treatment approach for patients with adolescent idiopathic scoliosis [4]. Since the last decade, scoliosis orthoses or braces have been considered the gold standard for decreasing the progression rate of curves, based on the multicenter study "Bracing in Adolescent Idiopathic Scoliosis Trial" from the United States and Canada, which demonstrated that the success of orthotic treatment increases with higher wearing hours [5]. However, treating adolescents with prolonged unpredictable treatment outcomes is challenging [6,7]. The most affected are female individuals with a higher diagnosis rate than male individuals [1], and the rate of deformity is highest at the beginning of puberty, during adolescence [2].

The hours of orthosis wear stand as an integral component of the treatment [8]. These orthoses can be rigid or soft, varying in construction, correction principles, and clinical outcomes. Ali et al [9], in the systematic review, compare 20 different orthoses and highlight their differences. Greater curve correction is obtained with rigid orthoses, while soft orthoses can deliver correction, enhanced comfort, and engage the patient's muscles [9].

The Cobb angle is the clinical measure that quantifies the degree of spinal deformity. At the same time, the Risser stage assesses skeletal maturity on a 0‐5 scale, using either a US or a European scale where 5 means complete maturity. In addition to these clinical approaches, patient-reported questionnaires support with valuable information on the patients' emotional and psychological burden [10]. Different patient monitoring tools support clinicians in long-term scoliosis management in adolescents by assessing the effectiveness of orthoses and enabling patient-focused treatment adjustments [11,12]. Digital health care–based monitoring using sensors can support the treatment journey [2]. As technology advances, there is a need to update and integrate sensors with digitalized monitoring solutions to improve bracing outcomes. Mobile-based health monitoring systems are emerging to track adolescent treatment compliance and well-being [13]. Scoliosis care and treatment protocols have been inconsistent globally, with regional variants. Different health care providers involved in scoliosis treatment identified the need for increased research collaboration and standardization of treatment protocols globally [2]. Accepting these diverse and significant contexts that address user needs within scoliosis, integrating both health-oriented and technology-driven strategies, can be useful for future co-designs.

Context and Significance

User Experience and Perception

In scoliosis orthotic management, patient adherence to treatment goals is significant but challenging [8,10]. Adherence to orthotic treatment is a process that may begin with experiencing symptoms, consulting a therapist, becoming aware of visible spinal deformity, and receiving attention from family members. It may then progress to seeking care by accepting the diagnosis, discussing options with an orthotist, agreeing to treatment and living with an orthosis, and adapting psychosocially. Each step can influence the patient's compliance. Thus, treatment with orthosis is a process that depends on the combined effort of patients, family, physicians, and orthotists to reach treatment goals [14,15]. The treatment aims to enhance the patient's quality of life by addressing current and long-term needs. The target of scoliosis management includes preventing curve progression and thereby reducing physical disability, improving body aesthetics, delivering daily functioning, overall well-being, including mental health, and improving self-esteem [16]. To achieve these broader treatment goals, it is essential to incorporate evaluation and monitoring methods that capture patients' experiences and perceptions of treatment challenges and their preferences [17].

Monitoring Tools

Different questionnaires are designed for monitoring patients. Questionnaires are frequently used to objectively assess patients' subjective experiences, such as physical function and psychological well-being [18-20]. Capturing meaningful patient-reported information enables health personnel to understand treatment adherence better and recommend more effective interventions. Clinicians need to discuss essential aspects of orthosis adherence, such as the actual hours the orthosis is worn compared with the prescribed duration, and the effectiveness of prescribed exercises [21]. This information can support prognosis and guide treatment decisions. Record cards or patient logbooks can give additional information to track patient observations between visits [22]. These methods represent traditional and nontechnological approaches to monitoring. While pain is not typically a primary clinical issue in orthosis-treated adolescent patients with idiopathic scoliosis, some questionnaires include pain subscales because these tools were originally developed for broader scoliosis populations. Pain-related items were therefore reported when present in included studies but interpreted cautiously.

Technology-Assisted Approaches

Sensor-assisted technologies are evolving toward personalized and intelligent orthotic management systems that enable real-time monitoring and communication among patients,

XSL•FO
RenderX

caregivers, and health care professionals. Various wearable sensors have been developed to objectively monitor orthosis use, including temperature, pressure, force, and motion sensors (eg, studies by Fregna et al [8], Rahman et al [23], Zou et al [24], Cordani et al [25], Tymińska et al [26], and Donzelli et al [27]). For instance, in the study by Donzelli et al [27], sensor data provides information on actual orthosis wear hours, aiding clinicians in planning personalized treatment goals. Smart textile sensor systems for scoliosis monitoring are attracting research interest [28]. These technologies can quantify orthosis wear time and adherence, providing clinicians with objective information to support individualized treatment decisions. Data collected from these devices can be displayed through mobile applications or web-based platforms, allowing continuous monitoring of treatment progress [17]. More recently, wearable sensors integrated with AI have expanded opportunities for personalized monitoring and decision support (eg, studies by Cordani et al [25], Dehzangi et al [29], Zhu et al [30], and Li et al [31]). However, further evidence is needed to determine how these technologies can best support clinical decision-making and improve outcomes in adolescents.

Among existing reviews of adolescent idiopathic scoliosis management, the study by Raudenbush et al [32] provides an overview of studies evaluating, diagnosing, and treating adolescent patients with idiopathic scoliosis. Previous scoping reviews have explored evidence of successful treatment outcomes predominantly using the Cobb angle and quality-of-life questionnaires [33], and another study consolidated user needs, particularly regarding body image schema in adolescent patients with idiopathic scoliosis [34]. The scoping review includes studies from 2020 that capture the themes relevant to integrated sensor-based wearables, particularly electromyography systems, AI, and relevant platforms, and reports the persistent gap in laboratory design and long-term clinical validation [35]. There appears to be a lack of scoping review that integrates patient-reported outcome measures (PROMs) with technology-assisted monitoring approaches for adolescent idiopathic scoliosis.

It is important to distinguish between subjective and objective approaches to monitoring orthotic treatment in adolescents with scoliosis. Subjective approaches capture the patient's lived experience and perceptions, including comfort, usability, adherence challenges, psychosocial burden, and satisfaction with treatment, typically assessed through PROMs, questionnaires, or interviews. In contrast, objective approaches rely on technology-assisted monitoring tools, such as embedded sensors, temperature or pressure measurements, wear-time trackers, or other digital systems that quantify orthosis use or biomechanical parameters without relying on self-report. A third category includes integrated approaches, which combine patient-reported measures with objective sensor-based data to provide a more comprehensive understanding of both adherence and user experience. This distinction provides a clearer conceptual framework for synthesizing the existing literature and interpreting how different monitoring strategies contribute to treatment optimization.

Aim and Research Questions

The primary aim of this study was to determine the existing evidence on monitoring approaches spanning user experiences, clinical patient-monitoring tools, and technology-assisted systems for optimizing adherence and outcomes in the treatment and management of adolescent idiopathic scoliosis.

The integration of health and technology requires a detailed review to identify gaps and emerging technologies with the potential to improve patient care, especially for adolescents. This scoping review aims to estimate current evidence on user experience and perception of adolescents on diverse monitoring tools, including the technological and clinically assisted approaches, as part of conservative orthotic management for adolescents with idiopathic scoliosis. The research questions (RQs) are listed in Textbox 1.

Textbox 1. Three research questions (RQs).

RQ-A

- User experience and perception: What is known about user experience, perception, and needs related to orthosis-related monitoring in adolescents with idiopathic scoliosis?

RQ-B

- Monitoring tools: What patient-reported or digital monitoring tools have been used to assess adherence, comfort, or treatment progress in bracing among adolescents with idiopathic scoliosis?

RQ-C

- Technology-assisted approaches: What technology-assisted or sensor-based monitoring systems have been used in the management of bracing for adolescents with idiopathic scoliosis and which parameters do they capture?

Methods

Rationale for Scoping Review

A systematic review addresses defined, narrow questions with a strict assessment of evidence, while scoping reviews are ideal for examining emerging areas, clarifying definitions, and identifying research gaps [36,37]. A scoping review is particularly well-suited to this topic, given the complexity of the RQs and the interdisciplinary nature, allowing exploration of the breadth and depth of available evidence and supporting the identification of key concepts and gaps. Given the diverse study designs on this topic, including both health and technological perspectives, a scoping review enables a more inclusive and exploratory synthesis of available knowledge.

XSL•FO
RenderX

Framework

Arksey and O'Malley [37] provide the methodological framework of the review process that included systematic identification, selection of articles based on inclusion and exclusion, and charting of evidence across predefined RQs [38]. Data extraction into tables, its analysis, and presentation were defined by the Joanna Briggs Institute that uses qualitative content analysis and data visualization approaches [39]. The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) and the checklists were used to report these processes as suggested by [40]. In the review, the PRISMA-ScR guidelines and the PRISMA-ScR checklist were followed. CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) was reviewed for completeness because some included studies involve tech-assisted interventions, but PRISMA-ScR remains the main reporting framework [41].

Protocol and Registration

The protocol was prepared before conducting the search. RQs helped to guide the database search and the data collection in accordance with the scoping review framework. It is registered in INPLASY (ID INPLASY2025100032).

Information Sources and Search

The research team developed the scoping review protocol, drawing on existing literature to initiate discussions with university librarians, after a joint dialogue with specialized librarians and researchers. Search trials of the Medline database were prepared and tested first. A 2-step search strategy was finalized to comprehensively address all RQs. Based on this defined search strategy, librarians conducted the search in the 7 databases, namely MEDLINE, Embase, PsycINFO, Cochrane Library, CINAHL, Web of Science, and Scopus [42,43]. The results from both searches in each database were consolidated to generate the final single EndNote 21 (Clarivate) file that was compiled by the librarians. All articles published up to May 2024 were considered. Both the search strategies for each database are provided in Multimedia Appendix 1.

Eligibility Criteria

The primary criterion is that every article must include at least 1 patient participant, aged 10 to 17 years and 11 months (17.11 y) and diagnosed with idiopathic scoliosis [3]. If studies contain diagnoses other than adolescent idiopathic scoliosis and results are presented together, the article shall be excluded. Textbox 2 illustrates inclusion and exclusion criteria based on the Population, Intervention, Comparison, Outcome, and Study design (PICOS) framework, along with other generic criteria.

XSL•FO
RenderX

Textbox 2. Inclusion and exclusion criteria.

Inclusion Criteria

- Population
  - (1) Diagnosis: Patients diagnosed with adolescent idiopathic scoliosis
  - (2) Age: Between 10 years and 17.11 years.
- Intervention: Idiopathic scoliosis treatments with orthosis, with or without physical activities (including training, physical activity, and exercises).
- Comparators: All comparators are eligible (including no comparator, placebo, or other orthotic treatments).
- Outcomes: Not applicable (scoping review design).
- Study: All articles meeting search criteria considered.
- Publication type: Peer-reviewed articles and conference papers; opinions and medical hypotheses.
- Availability: Full-text accessible through institutional affiliations (eg, OsloMet, Aalborg University).
- Language: English only.

Exclusion Criteria

- Population
  - (1) Diagnosis:
  - Congenital scoliosis, neuromuscular scoliosis, or other nonadolescent idiopathic scoliosis types.
  - Articles with incomplete participant information or without patient participants.
  - Studies with healthy subjects.
  - (2) Age: Below 10 years or above 17.11 years and unclear age information from full text.
- Intervention: Nonorthotic treatments, including surgical interventions.
- Study: Simulation studies without patient involvement (eg, model-based or radiograph-only analysis). Studies focused solely on orthosis or prototype development without clinical or patient testing, without patient-reported outcomes or user perspectives.
- Publication type: Book reviews, editorials, conference abstracts (eg, oral presentations, short abstracts without details), short papers without peer review, commentaries, errata, letters to the editor.
- Availability: Articles not accessible in a readable format; ongoing or incomplete trials without published results.
- Language: Articles not available in English and with English abstracts but foreign-language titles or published in non-English journals.

Selection of Sources of Evidence

Overview

The Rayyan software tool was used to organize and manage the screening process. Each article was compared against its duplicates to retain only 1 version, following priority order based on the following database sources: MEDLINE, Embase, PsycINFO, Cochrane Library, CINAHL, Web of Science, and Scopus. Rayyan's reference comparison features helped highlight differences between duplicate entries. Duplicates were manually removed to ensure accuracy, despite the ability of Rayyan's feature to remove duplicates as groups with a certain percentage of duplication.

Articles published in health sciences typically present patient information clearly, particularly in the Methods section. However, this level of detail was often lacking in technology-focused or prototype-based studies, even when participants were involved. When further clarification was required, articles were annotated using the labeling feature of Rayyan. Such articles with partial or ambiguous information were retained and moved to full-text screening. Through digital meetings, the 2 independent researchers (AM and PG) resolved ambiguity according to the predefined inclusion and exclusion criteria. Most disagreements were resolved through discussions, and when necessary, a third researcher (MP) provided input. For example, disagreements regarding age limits in the eligibility criteria were resolved using the International Scientific Society on Scoliosis Orthopaedic and Rehabilitation Treatment (SOSORT) 2012 guidelines, which specify that adolescent idiopathic scoliosis is for ages 10 to 17 years (specifically 17 y and 11 mo), ages 3 to 9 years as juveniles, and ages 18 years and older as adults [3]. Furthermore, in cases where participant demographics varied from article to article, independent reviewers resolved them by verifying the full text during online meetings. Full texts were retrieved through university affiliations, specifically OsloMet and Aalborg University, and 22 articles were acquired with support from OsloMet University for full-text screening.

Language was considered as a criterion; articles that had English abstracts but were written in foreign languages fell outside the scope of this review. All articles excluded during the abstract and title screening phase used predefined exclusion labels in the Rayyan software. For example, the label "foreign language"

XSL•FO
RenderX

was used to exclude articles. If the title appeared in square brackets (eg, [German] in the Rayyan software tool) or if the journal title was in a non-English language, they were excluded at title and abstract screening. The user-defined label in the Rayyan software helped trace the stage.

Participant demographics were another eligibility criterion. Articles that failed to meet the inclusion criteria, specifically age or disease diagnosis, were the most common and challenging eligibility criteria. If the article included at least 1 participant younger than 10 years or older than 17.11 years, it was excluded despite valuable insights. Articles with unclear age ranges but reporting mean and SD, or referencing Risser stages, were considered on a case-by-case basis. Studies that included both juvenile and adolescent participants were excluded as results cannot be separated (eg, study by Lou et al [44]).

Data Charting

During the charting process, independent reviewers encountered challenges organizing multidisciplinary articles, which were resolved by iteratively refining a table structure based on charted data samples. The final table was refined through multiple meetings. The extracted data items were organized as columns.

Data Items

During data extraction, included studies were categorized according to the primary monitoring approach used: (1) subjective monitoring, focusing on patient-reported outcomes and user experience; (2) objective monitoring, using technology-assisted or sensor-based tools to quantify orthosis use or treatment parameters; or (3) integrated monitoring, combining both subjective and objective measures. This categorization was applied consistently across study characteristics and informed the narrative synthesis.

The data extraction table included columns for study characteristics, including participant demographics related to scoliosis and age, followed by columns for clinical and technological RQs. Due to variability in how age was reported across studies, this information was standardized as “mean age (SD)” supplemented with reported age ranges (lower-upper limits) in Multimedia Appendix 2. These age data are finalized for each selected article by verifying inclusion criteria, demographic tables, or results sections, and are presented as age range (minimum year-maximum year). For example, when the Risser stage is represented as the upper limit, it is presented as range (minimum year-maximum Risser 2/3). Risser stages 2 to 4 correspond to mean ages ranging from 13.1 (SD 1.1) to 14.8 (SD 1.0) years in adolescent patients with idiopathic scoliosis, which is less than 18 years [45]. Other study data extracted as part of this review included publication year, country data, author information, journal data, participant data, including the male-to-female ratio, and study methodology [39]. In alignment with the RQs, additional data were extracted on orthosis types, clinical monitoring instruments (eg, questionnaires), technology-based sensor monitoring parameters, data analysis methods, participant feedback and their perspectives, study strengths, results, and proposed future directions. The result of this phase is a detailed data extraction table, which is provided in Multimedia Appendix 2.

Data Synthesis Approach

The refined table aligns with the 3 RQs, helping condense the 2 main themes that emerged from data synthesis. AI was not used in data synthesis. The detailed data extraction table records key findings from each article. The time-based evolution of selected data items based on the 3 RQs is presented as illustrative figures that are created using these data in Microsoft Excel using Insert Tab/PivotChart and PivotTable features.

Ethical Considerations

The research was conducted independently. The funder (University) played no role in the formulation of the study design, data analysis and interpretation, or writing of this paper.

Results

Overview

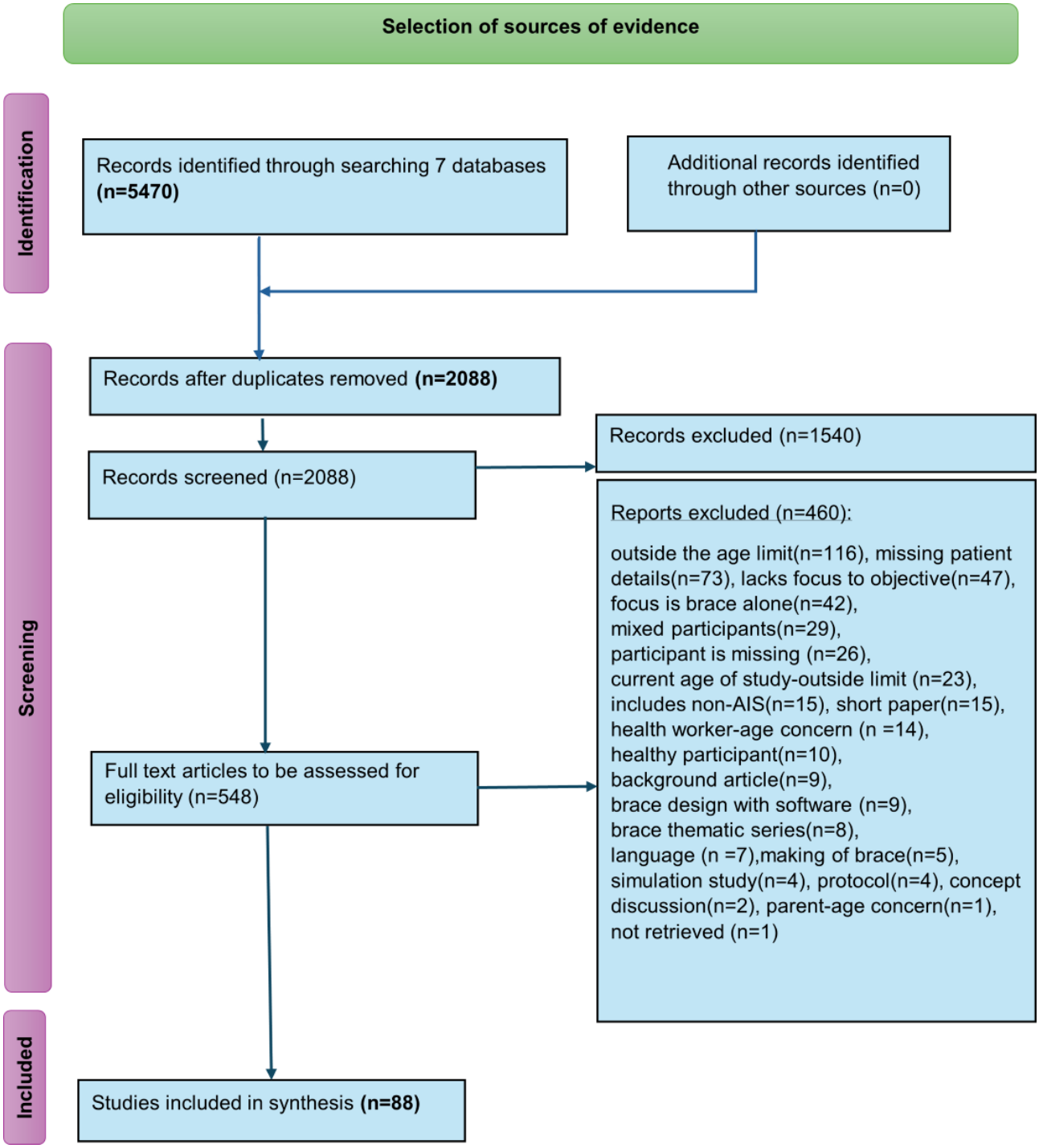
A total of 5470 records were identified from 7 databases: MEDLINE (n=1016), Embase (n=1472), PsycINFO (n=30), Cochrane Library (n=178), CINAHL (n=518), Web of Science (n=798), and Scopus (n=1458). Due to overlap between the 2 search strategies, approximately 61.8% (3382/5470) of the duplicate articles were removed. A total of 2088 articles were identified for further screening based on their titles and abstracts, followed by full-text screening. Exclusion of articles was applied based on the predefined eligibility criteria in Textbox 2, specifically targeting participants diagnosed with idiopathic scoliosis and aged 10-17.11 years.

During full-text screening, articles (n=460) were excluded using user-defined exclusion criteria (as illustrated in Figure 1), which helped identify the stage at which each article was removed in the Rayyan software. For example, 7 articles were excluded at the full-text stage that were not captured at the first phase and were labeled as “language (n=7)” as in Figure 1. Article exclusions were predominantly based on participant demographics and alignment with the review objectives. Some exclusion categories were conceptually related. To avoid duplicate counting, each article was assigned 1 primary reason that was the most appropriate and explanatory reason. Exclusion reasons include age outside the inclusion criteria (n=116), follow-up study design with older participants (n=23), missing patient or participant information (n=73 and n=26), protocol-only publications without participants (n=4), and short papers with missing data (n=15). Studies with ineligible participants were excluded, for instance, studies involving surgical interventions (considered mixed participants; n=29), nonidiopathic scoliosis diagnoses (n=15), healthy participants (n=10), and parents (n=1). Delphi studies involving clinicians only were also excluded (n=14). The remaining exclusions were due to a lack of alignment with the scoping review objectives, including orthosis-focused, design-, simulation-, manufacturing-, thematic-, background-, or concept-based studies. Moreover, 1 article was excluded because it was not available in a readable format. A total of 88 studies (Table 1) were included for data extraction using Microsoft Excel. Each row in the data extraction table corresponds to an article, thereby yielding 88 rows. Table 1 summarizes the study characteristics of each included article based on the research objectives. This process is presented in

XSL•FO
RenderX

Figure 1 as a flow diagram in accordance with PRISMA-ScR guidelines published by Tricco et al in 2018 [40].

Figure 1. Flow diagram based on PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-analyses extension for scoping reviews).



XSL•FO
RenderX

Table . Summary of results.

Research questions	Title of article	Study demographics (Adolescent participants with idiopathic scoliosis)					Themes: subjective (S), objective (O), and integrated (I)		
		Year	Country	Female	Male	Age range (y): (Minimum - Maximum)	Report on orthosis (Brace)	Monitoring tool	S vs O vs I
RQ[a]-A only (Subjective)									
A1	Adolescents' Experience during Brace Treatment for Scoliosis: A Qualitative Study [46]	2022	China	13	2	10-16	Type missing (report brace)	Interview	S
RQ-B only (Subjective)									
B1	Chinese Adaptation of the Bad Sobernheim Stress Questionnaire for Patients With Adolescent Idiopathic Scoliosis Under Brace Treatment [47]	2015	China	71	15	10-16	Brace (Shanghai Delin Artificial Limb Recovering Equipment, China)	BSSQ[b] (Deformity and -Brace), SRS-22[c]	S
B2	Cross-Cultural Adaptation and Validation of the Bad Sobernheim Stress Questionnaire in Iranian Adolescents with Idiopathic Scoliosis Using Thoracolumbar Orthoses [48]	2022	Iran	40	15	12-17	TLSO[d]	BSSQ-Brace, SRS-22r	S
B3	Greek Adaptation and Validation of the Bad Sobernheim Stress Questionnaire-Brace and the Bad Sobernheim Stress Questionnaire-Deformity [49]	2023	Greece	46	1	11-17	Asymmetric Scoliosis Brace by SLC	BSSQ (Deformity and -Brace), SRS-22	S

XSL•FO
RenderX

Research questions	Title of article	Study demographics (Adolescent participants with idiopathic scoliosis)					Themes: subjective (S), objective (O), and integrated (I)		
		Year	Country	Female	Male	Age range (y): (Minimum - Maximum)	Report on orthosis (Brace)	Monitoring tool	S vs O vs I
B4	Persian adaptation of the Bad Sobernheim stress questionnaire for adolescent with idiopathic scoliosis [50]	2020	Iran	38	15	10-16	Milwaukee brace	BSSQ (Deformity and Brace), SRS-22	S
B5	Polish adaptation of Bad Sobernheim Stress Questionnaire-Brace and Bad Sobernheim Stress Questionnaire-Deformity [51]	2009	Poland	35	0	12-16	Cheneau brace	BSSQ (Deformity and -Brace), SRS-22	S
B6	Spanish validation of Bad Sobernheim Stress Questionnaire (BSSQ (brace).es) for adolescents with braces [52]	2010	Spain	33	2	10-16	Rigo System Cheneau Brace	BSSQ-Brace, SRS-22	S
B7	The Turkish version of the Brace Questionnaire in brace-treated adolescents with idiopathic scoliosis [53]	2018	Turkey	25	3	10-17	TLSO	BrQ[e], SRS-22, BSSQ-Brace	S
B8	Validation of a Korean version of the quality-of-life profile for spine deformities (QLPSD) in patients with adolescent idiopathic scoliosis [54]	2022	South Korea	91	13	10.4-17.1	Type missing (report brace)	QLPSD, SRS-22	S

XSL•FO
RenderX

Research questions	Title of article	Study demographics (Adolescent participants with idiopathic scoliosis)					Themes: subjective (S), objective (O), and integrated (I)		
		Year	Country	Female	Male	Age range (y): (Minimum - Maximum)	Report on orthosis (Brace)	Monitoring tool	S vs O vs I
B9	Validation of the Italian spine youth quality of life (ISYQOL) in Korean population [55]	2021	South Korea	120	0	10.4-17.1	Type missing (report brace)	ISYQOL, SRS-22	S
B10	Validation of the Korean version of the Brace Questionnaire [56]	2018	South Korea	93	10	10-16	TLSO	K-BrQ[f], K-SRS-22[g]	S
B11	Revisiting the psychometric properties of the Scoliosis Research Society-22 (SRS-22) French version [57]	2017	Canada	307	45	10-17	Type missing (report brace)	SRS-22fv	S
B12	Whether Orthotic Management and Exercise are Equally Effective to the Patients with Adolescent Idiopathic Scoliosis in Mainland China? [58]	2018	China	19	5	11-14	Rigid TLSO	SRS 22	S
B13	The relationship between quality of life and compliance to a brace protocol in adolescents with idiopathic scoliosis: a comparative study [59]	2009	South Africa	31	0	13-16	Rigo System Cheneau Brace	BrQ	S
B14	The influence of brace on quality of life of adolescents with idiopathic scoliosis [60]	2006	Greece	32	4	12-17	Boston brace	BrQ	S

XSL•FO
RenderX

Research questions	Title of article	Study demographics (Adolescent participants with idiopathic scoliosis)					Themes: subjective (S), objective (O), and integrated (I)		
		Year	Country	Female	Male	Age range (y): (Minimum - Maximum)	Report on orthosis (Brace)	Monitoring tool	S vs O vs I
B15	The importance of trunk perception during brace treatment in moderate juvenile idiopathic scoliosis: What is the impact on self-image? [61]	2017	Italy	54	0	10-17	Cheneau, Boston, and Milwaukee	TAPS[h], SRS-22	S
B16	The efficacy of Schroth exercises combined with the Cheneau brace for the treatment of adolescent idiopathic scoliosis: a retrospective controlled study [62]	2022	China	157	NR	10-16	Cheneau-type TLSO	SRS-22r, EQ-5D	S
B17	The effect of Schroth exercises added to the standard of care on the quality of life and muscle endurance in adolescents with idiopathic scoliosis-an assessor and statistician blinded randomized controlled trial: "SOSORT[i] 2015 Award Winner" [63]	2015	Canada	47	3	12.7-14.2	TLSO, Charleston, Providence Combination of Providence/TLSO brace	SEQ[j], SAQ[k], SRS-22r	S

XSL•FO
RenderX

Research questions	Title of article	Study demographics (Adolescent participants with idiopathic scoliosis)					Themes: subjective (S), objective (O), and integrated (I)		
		Year	Country	Female	Male	Age range (y): (Minimum - Maximum)	Report on orthosis (Brace)	Monitoring tool	S vs O vs I
B18	The effect of compliance to a Rigo System Cheneau brace and a specific exercise program on idiopathic scoliosis curvature: A comparative study: SOSORT 2014 award winner [64]	2014	South Africa	80	0	12-16	Rigo System Cheneau (RSC) brace	BrQ, HSPQ[l], 16PF[m]	S
B19	The Effect of Bracing on Spinopelvic Rotation and Psychosocial Parameters in Adolescents with Idiopathic Scoliosis [65]	2019	Iran	28	2	10-Risser 2	Milwaukee, TLSO	Persian BSSQ and BrQ	S
B20	Quality of life and patient satisfaction in bracing treatment of adolescent idiopathic scoliosis [66]	2018	Argentina	43	0	10.8-14.5	TLSO	BrQ	S
B21	Quality of life after conservative treatment of adolescent idiopathic scoliosis [67]	2008	Greece	32	0	12-16	modified Boston brace	BrQ	S
B22	Psychological Effects of the SRS-22 on Girls With Adolescent Idiopathic Scoliosis [68]	2018	United States	45	0	10-17	Type missing (report braced)	BAS[n], SRS-22	S

XSL•FO
RenderX

Research questions	Title of article	Study demographics (Adolescent participants with idiopathic scoliosis)					Themes: subjective (S), objective (O), and integrated (I)		
		Year	Country	Female	Male	Age range (y): (Minimum - Maximum)	Report on orthosis (Brace)	Monitoring tool	S vs O vs I
B23	Patients' and Parents' Perceptions of Appearance in Scoliosis Treated with a Brace: A Cross-Sectional Analysis [69]	2014	Poland	41	0	10-17	Cheneau brace	SAQ-pl	S
B24	Novel questionnaire to enhance brace wear adherence in patients with adolescent idiopathic scoliosis and the relationship of the quality of life [22]	2021	Thailand	35	5	11-17.3	Type missing (report braced)	Brace questionnaire	S
B25	Mindset correlates with health-related quality of life assessment in patients with adolescent idiopathic scoliosis [70]	2021	United States	94	16	10-Risser 2	TLSO	SRS-30 HRQoL[o] survey and Health Mindset Scale	S
B26	Longer Brace Duration Is Associated with Lower Stress Levels and Better Quality of Life in Adolescents with Idiopathic Scoliosis [71]	2023	Italy	38	8	11-16	Sforzesco, Chêneau	BSSQ, ISYQoL	S
B27	Is higher compliance to brace therapy associated with poorer quality of life and self-image? A 36-months follow-up study [72]	2023	Italy	61	7	10-16	TLSO rigid brace	SRS-22r	S

XSL•FO
RenderX

Research questions	Title of article	Study demographics (Adolescent participants with idiopathic scoliosis)					Themes: subjective (S), objective (O), and integrated (I)		
		Year	Country	Female	Male	Age range (y): (Minimum - Maximum)	Report on orthosis (Brace)	Monitoring tool	S vs O vs I
B28	Flexibility Predicts Curve Progression in Providence Nighttime Bracing of Patients With Adolescent Idiopathic Scoliosis [73]	2016	Denmark	60	3	10-16	Providence brace	Danish version SRS-22	S
B29	Female patients' and parents' assessment of deformity- and brace-related stress in the conservative treatment of adolescent idiopathic scoliosis [74]	2012	Poland	63	0	10-17	Cheneau brace	Polish BSSQ (Deformity and Brace)	S
B30	Exploring mass customization and textile application in medical products: redesigning scoliosis brace for shorter production lead time and better quality of life [75]	2020	China	10	0	11-15	Garment brace & hard brace	BrQ questionnaire	S
B31	Enhancing Idiopathic Scoliosis Patients' Outcomes Through Functional Rehabilitation Training in Combination with Orthosis [76]	2023	China	30	64	10-17	Fabricated orthosis	SF-36[p]	S

XSL•FO
RenderX

Research questions	Title of article	Study demographics (Adolescent participants with idiopathic scoliosis)					Themes: subjective (S), objective (O), and integrated (I)		
		Year	Country	Female	Male	Age range (y): (Minimum - Maximum)	Report on orthosis (Brace)	Monitoring tool	S vs O vs I
B32	Effect of different undergarment designs on the compliance and acceptance of the patients with adolescent idiopathic scoliosis under orthotic treatment [77]	2022	China	37	1	10-15	Spinal orthosis	SRS-22r and Brace Questionnaire (BrQ)	S
B33	Effect of bracing on the quality of life of adolescents with idiopathic scoliosis [78]	2004	United States	173	41	10.1-17.7	Type missing (report braced)	CHQ PF-28[q], PODCI[r]	S
B34	Differences in deformity and bracing-related stress between rural and urban area patients with adolescent idiopathic scoliosis treated with a Cheneau brace [79]	2011	Poland	64	0	10-17	Cheneau brace	Polish BSSQ-brace and BSSQ deformity	S
B35	Core stabilization exercises versus scoliosis-specific exercises in moderate idiopathic scoliosis treatment [80]	2019	Turkey	30	0	12-Risser 3	Type missing (report braced)	WRVAS[s], POTSI[t], SRS-22	S
B36	Body-Related Attentional Bias in Adolescents Affected by Idiopathic Scoliosis [81]	2023	Italy	14	0	12-16	Type missing (report brace treatment)	SRS-22r, Visual Match-to-Sample Task	S

XSL•FO
RenderX

Research questions	Title of article	Study demographics (Adolescent participants with idiopathic scoliosis)					Themes: subjective (S), objective (O), and integrated (I)		
		Year	Country	Female	Male	Age range (y): (Minimum - Maximum)	Report on orthosis (Brace)	Monitoring tool	S vs O vs I
B37	Body Image and Quality-of-Life in Untreated Versus Brace-Treated Females With Adolescent Idiopathic Scoliosis [82]	2016	United States	319	0	10-15	Type missing (report brace)	SAQ, PedsQoL[u]	S
B38	Back Pain Prevalence Is Associated With Curve-type and Severity in Adolescents With Idiopathic Scoliosis: A Cross-sectional Study [83]	2017	Australia	158	23	10-17	Type missing (report brace wearing)	RMDQ[v], BPI[w]	S
B39	An alternative to a randomized control design for assessing the efficacy and effectiveness of bracing in adolescent idiopathic scoliosis [84]	2015	China	63	5	10-15	TLSO	SRS-22, STAI[x], BDI-II[y]	S
RQ-C only (Objective)									
C1	A compliance real-time monitoring system for the management of the brace usage in adolescent idiopathic scoliosis patients: a pilot study [30]	2021	China	23	5	10-15	Chêneau brace	Force	O

XSL•FO
RenderX

Research questions	Title of article	Study demographics (Adolescent participants with idiopathic scoliosis)					Themes: subjective (S), objective (O), and integrated (I)		
		Year	Country	Female	Male	Age range (y): (Minimum - Maximum)	Report on orthosis (Brace)	Monitoring tool	S vs O vs I
C2	Analysis of the corrective forces exerted by a dynamic derotation brace (DDB) [85]	2011	Greece	38	6	10-17.2	Dynamic derotation brace (DDB)	Force	O
C3	A smart point-of-care compliance monitoring solution for brace treatment of adolescent idiopathic scoliosis patients [29]	2021	United States	3	0	10-13	TLSO	Force and motion	O
C4	A wireless sensor network system to determine biomechanics of spinal braces during daily living [86]	2010	Canada	5	1	10-13	TLSO	Force	O
C5	Biomechanical Assessment of Providence Nighttime Brace for the Treatment of Adolescent Idiopathic Scoliosis [87]	2016	Canada	16	2	10-16	Providence brace	Pressure	O
C6	Biomechanical evaluation of the Milwaukee brace [88]	1998	China	9	0	10-14.8	Milwaukee brace	Strap pressure	O
C7	Boston brace correction in idiopathic scoliosis: a biomechanical study [89]	2003	United States	12	0	13-14	Boston brace	Pressure matrix	O

XSL•FO
RenderX

Research questions	Title of article	Study demographics (Adolescent participants with idiopathic scoliosis)					Themes: subjective (S), objective (O), and integrated (I)		
		Year	Country	Female	Male	Age range (y): (Minimum - Maximum)	Report on orthosis (Brace)	Monitoring tool	S vs O vs I
C8	Compliance with nighttime overcorrection bracing in adolescent idiopathic scoliosis: Result from a cohort follow-up [90]	2020	France	19	1	11.2-15.6	"CAEN[z]" brace	Temperature	O
C9	Correlation between quantity and quality of orthosis wear and treatment outcomes in adolescent idiopathic scoliosis [91]	2004	Canada	12	3	12.2-16.8	Boston brace	Force	O
C10	Early intervention in nonoperative management of adolescent idiopathic scoliosis [92]	2000	China	9	0	10.5-14.0	TLSO (underarm brace)	Pressure and tension	O
C11	Effectiveness and biomechanics of spinal orthoses in the treatment of adolescent idiopathic scoliosis (AIS) [93]	2000	China	33	0	10-15	TLSO	Pressure and tension	O
C12	Effect of Compliance Counseling on Brace Use and Success in Patients with Adolescent Idiopathic Scoliosis [94]	2016	United States	154	17	10.2-16.0	TLSO	Temperature	O
C13	Electromyography of scoliotic patients treated with a brace [95]	2003	Canada	11	0	12-16	Boston brace	EMG[aa]	O

XSL•FO
RenderX

Research questions	Title of article	Study demographics (Adolescent participants with idiopathic scoliosis)					Themes: subjective (S), objective (O), and integrated (I)		
		Year	Country	Female	Male	Age range (y): (Minimum - Maximum)	Report on orthosis (Brace)	Monitoring tool	S vs O vs I
C14	Electronic monitoring of scoliosis brace wear compliance [23]	2010	United States	8	2	10.3-16.4	Wilmington scoliosis brace (TLSO)	Temperature	O
C15	Electronic monitoring of orthopedic brace compliance [96]	2015	United States	52	3	10-Risser 2	Wilmington brace	Temperature	O
C16	Forces exerted during exercises by patients with adolescent idiopathic scoliosis wearing fiberglass braces [97]	2006	Italy	14	3	12-17	Fiberglass braces	Pressure	O
C17	How quantity and quality of brace wear affect the brace treatment outcomes for AIS [98]	2016	Canada	52	8	10-Risser 2	TLSO	Force	O
C18	Load compliance monitor system for the treatment of scoliosis [99]	1999	Canada	1	0	13	Boston brace	Force	O
C19	Mechanical and Clinical Evaluation of a Shape Memory Alloy and Conventional Struts in a Flexible Scoliotic Brace [100]	2018	China	2	0	10-14	Garment brace	Pressure	O
C20	Objective compliance of adolescent girls with idiopathic scoliosis in a dynamic SpineCor brace [101]	2010	Switzerland	12	0	10-15.6	Dynamic SpineCor brace	Temperature	O

XSL•FO
RenderX

Research questions	Title of article	Study demographics (Adolescent participants with idiopathic scoliosis)					Themes: subjective (S), objective (O), and integrated (I)		
		Year	Country	Female	Male	Age range (y): (Minimum - Maximum)	Report on orthosis (Brace)	Monitoring tool	S vs O vs I
C21	Outcomes for nighttime bracing in adolescent idiopathic scoliosis based on brace wear adherence [102]	2024	United States	103	19	10-16	Providence nighttime brace	Temperature	O
C22	Relationships between strap tension, interface pressures and spine correction in brace treatment of scoliosis [103]	2002	Canada	35	0	10-16	Boston brace	Pressure from force	O
C23	Study of the pressures applied by a Cheneau brace for correction of adolescent idiopathic scoliosis [104]	2008	France	30	2	10-Risser 2	Cheneau brace	Pressure	O
C24	The association between brace compliance and outcome for patients with idiopathic scoliosis [105]	2005	United States	30	4	10-16	Wilmington jacket	Temperature	O
C25	The biomechanical effectiveness of the Boston brace in the management of adolescent idiopathic scoliosis [106]	1989	United Kingdom	12	2	11-16	Boston brace	Pressure and strap force	O
C26	The effect of time on qualitative compliance in brace treatment for AIS [107]	2008	Canada	0	11	11-14.9	Boston brace	Pressure from force	O

XSL•FO
RenderX

Research questions	Title of article	Study demographics (Adolescent participants with idiopathic scoliosis)					Themes: subjective (S), objective (O), and integrated (I)		
		Year	Country	Female	Male	Age range (y): (Minimum - Maximum)	Report on orthosis (Brace)	Monitoring tool	S vs O vs I
C27	The Influence of Body Habitus on Documented Brace Wear and Progression in Adolescents With Idiopathic Scoliosis [108]	2020	United States	159	16	10-15.8	Boston brace, TLSO	Temperature	O
C28	Validation of a miniature thermochron for monitoring thoracolumbosacral orthosis wear time [109]	2012	United States	7	0	13.3-15.7	TLSO	Temperature	O
C29	Effect of the application of the Boston brace system on the electrical activity of the paravertebral muscles in adolescents with idiopathic scoliosis: preliminary report [110]	1982	Canada	6	0	11-16	Boston brace	EMG	O
C30	Lumbopelvic postural differences in adolescent idiopathic scoliosis: A pilot study [111]	2022	United States	21	6	11-17	Type missing (report braced)	EMG	O
C31	Is daily walking distance affected in adolescent idiopathic scoliosis? An original prospective study using the pedometer on smartphones [112]	2020	France	18	1	11-17	Night time brace	Motion	O
RQ-A ∩ RQ-B (Subjective)									

XSL•FO
RenderX

Research questions	Title of article	Study demographics (Adolescent participants with idiopathic scoliosis)					Themes: subjective (S), objective (O), and integrated (I)		
		Year	Country	Female	Male	Age range (y): (Minimum - Maximum)	Report on orthosis (Brace)	Monitoring tool	S vs O vs I
AB1	What factor induces stress in patients with AIS under brace treatment? Analysis of a specific factor using exploratory factor analysis [113]	2021	Japan	66	3	10-16	Underarm brace or Boston brace	BSSQ, SRS-22	S
AB2	The experience of brace treatment in children/adolescents with scoliosis [114]	2006	Greece	9	3	10-16	Type missing (report braced >12 h)	Interview	S
AB3	The effect of rigid versus flexible spinal orthosis on the clinical efficacy and acceptance of the patients with adolescent idiopathic scoliosis [115]	2008	China	Not reported out of 43	Not reported out of 43	10-14	SpineCor and rigid spinal orthosis	VAS[ab], 4 open Qn	S
AB4	Motivations for Compliance With Bracing in Adolescent Idiopathic Scoliosis [116]	2017	United States	39	0	11-15	Boston brace	Scoliosis Compliance Questionnaire	S
AB5	Psychosocial adaptation to wearing the Milwaukee brace for scoliosis. A pilot study of adolescent females and their mothers [117]	1984	United States	16	0	14-17	Milwaukee brace	Interview	S

XSL•FO
RenderX

Research questions	Title of article	Study demographics (Adolescent participants with idiopathic scoliosis)					Themes: subjective (S), objective (O), and integrated (I)		
		Year	Country	Female	Male	Age range (y): (Minimum - Maximum)	Report on orthosis (Brace)	Monitoring tool	S vs O vs I
AB6	Is self-image, in reference to the gravitational vertical, altered in adolescent idiopathic scoliosis? A multicenter, single-blind, case-control study [118]	2022	France	0	63	12.20-14.84	Type missing (report braced)	SRS-22, WRVAS, SAQ, TAPS	S
RQ-B ∩ RQ-C (Subjective and Objective Integrated)									
BC1	Measuring the compliance behavior of adolescents wearing orthopedic braces [119]	1999	Canada	40	0	10-16	Boston brace, semi-rigid orthopedic brace	Interview	I
BC2	A correlation study between in-brace correction, compliance to spinal orthosis and health-related quality of life of patients with Adolescent Idiopathic Scoliosis [120]	2014	China	42	0	11-15	Hong Kong brace: TLSO (underarm brace)	Chinese: TAPS, BrQ and SRS-22	I
BC3	Accuracy in the prediction and estimation of adherence to bracewear before and during treatment of adolescent idiopathic scoliosis [19]	2008	United States	108	16	10-15	Boston brace	Brace-Beliefs Questionnaire	I

XSL•FO
RenderX

Research questions	Title of article	Study demographics (Adolescent participants with idiopathic scoliosis)					Themes: subjective (S), objective (O), and integrated (I)		
		Year	Country	Female	Male	Age range (y): (Minimum - Maximum)	Report on orthosis (Brace)	Monitoring tool	S vs O vs I
BC4	The Intelligent Automated Pressure-Adjustable Orthosis for Patients With Adolescent Idiopathic Scoliosis: A Bi-Center Randomized Controlled Trial [121]	2020	China	23	0	10-14	Orthosis	SAQ, BrQ (follow up)	I
BC5	Impact of Brace-Related Stress on Brace Compliance in Adolescent Idiopathic Scoliosis: A Single-Center Comparative Study Using Objective Compliance Measurement and Brace-Related Stress [122]	2023	Japan	42	3	10-15	Underarm or Boston brace	Japanese BSSQ-brace	I
BC6	Factors Influencing Optimal Bracing Compliance in Adolescent Idiopathic Scoliosis: A Single Center Prospective Cohort Study [123]	2024	Japan	114	8	10-15	Boston brace	PROMs[ac], SRS-22r, SJ27[ad], JB-SSQ[ae]-brace	I
BC7	Effects of Mindfulness-Based Intervention to Improve Bracing Compliance in Adolescent Idiopathic Scoliosis Patients: a Randomized Controlled Trial [124]	2023	China	62	21	10-15	Type missing (report bracing)	EQ-5D-5L, SRS-22, BrQ, ERQ-CCA[af], SCS[ag], GSE[ah], PSS[ai], FFMQ[aj]	I

XSL•FO
RenderX

Research questions	Title of article	Study demographics (Adolescent participants with idiopathic scoliosis)					Themes: subjective (S), objective (O), and integrated (I)		
		Year	Country	Female	Male	Age range (y): (Minimum - Maximum)	Report on orthosis (Brace)	Monitoring tool	S vs O vs I
BC8	Effects of Bracing in Adolescents with Idiopathic Scoliosis [5]	2013	United States	135	11	10-15.10	Rigid thoracolumbosacral orthosis	PedsQoL	I
BC9	A Randomized Controlled Trial to Evaluate the Clinical Effectiveness of 3D-Printed Orthosis in the Management of Adolescent Idiopathic Scoliosis [125]	2022	China	30	0	10.7-14.5	3D-printed orthosis	SRS-22r, TAPS, BrQ	I
RQ-A ∩ RQ-C (Subjective and Objective Integrated)									
None									
RQ-A ∩ RQ-B ∩ RQ-C (Subjective and Objective -Fully integrated with user perspective)									
ABC1	Exploration of Contributory Factors to an Unpleasant Bracing Experience of Adolescent Idiopathic Scoliosis Patients a Quantitative and Qualitative Research [126]	2022	China	13	4	10-17	Cheneau / Boston brace	Force and SRS-22, GCQ[ak], ODI[al]	I

XSL•FO
RenderX

Research questions	Title of article	Study demographics (Adolescent participants with idiopathic scoliosis)					Themes: subjective (S), objective (O), and integrated (I)		
		Year	Country	Female	Male	Age range (y): (Minimum - Maximum)	Report on orthosis (Brace)	Monitoring tool	S vs O vs I
ABC2	Why Don't Adolescents Wear Their Brace? A Prospective Study Investigating Psychosocial Characteristics That Predict Scoliosis Brace Wear [127]	2023	United States	37	4	10-Risser 2	Custom TLSO orthosis	Temperature, patient information from validated questionnaire	I

[a]RQ: research question.
[b]BSSQ: Bad Sobernheim Stress Questionnaire.
[c]SRS-22: Scoliosis Research Society-22 Patient Questionnaire.
[d]TLSO: Thoracolumbosacral orthoses.
[e]BrQ: Brace Questionnaire.
[f]K-BrQ: Korean version of the Brace Questionnaire.
[g]K-SRS-22: Korean version of the Scoliosis Research Society-22 Outcomes questionnaire.
[h]TAPS: Trunk Appearance Perception Scale.
[i]SOSORT: International Society on Scoliosis Orthopedic and Rehabilitation Treatment.
[j]SEQ:Self-Efficacy Questionnaire
[k]SAQ: Spinal Appearance Questionnaire.
[l]HSPQ: High School Personality Questionnaire
[m]16PF: Sixteen Personality Factor Questionnaire
[n]BAS: Body Appreciation Scale.
[o]HRQoL: Health-Related Quality of Life.
[p]SF-36: 36-Item Short-Form Health Survey.
[q]CHQ PF-28: Child Health Questionnaire – Parent Form, 28 items.
[r]PODCI: Pediatric Outcomes Data Collection Instrument
[s]WRVAS: Walter Reed Visual Assessment Scale.
[t]POTSI: Posterior Trunk Symmetry Index.
[u]PedsQoL: Pediatric Quality of Life Inventory.
[v]RMDQ: Roland-Morris Disability Questionnaire
[w]BPI: Brief Pain Inventory.
[x]STAI: State-Trait Anxiety Inventory.
[y]BDI-II: Beck Depression Inventory-II.
[z]CAEN: corset à appui électif nocturne ("CAEN" brace)
[aa]EMG: electromyography.
[ab]VAS: Visual Analog Scale.
[ac]PROM: Patient-Reported Outcome Measure.
[ad]SJ27: Japanese Questionnaire-27
[ae]JBSSQ: Japanese version of the Bad Sobernheim Stress Questionnaire
[af]ERQ-CCA: Chinese adaptation of the Emotion Regulation Questionnaire for Children.
[ag]SCS: Self-Compassion Scale
[ah]GSE: General Self-Efficacy Scale.
[ai]PSS: Perceived Stress Scale
[aj]FFMQ: Five Facet Mindfulness Questionnaire.
[ak]GCQ: General Comfort Questionnaire

XSL•FO
RenderX

[al]ODI: Oswestry Disability Index

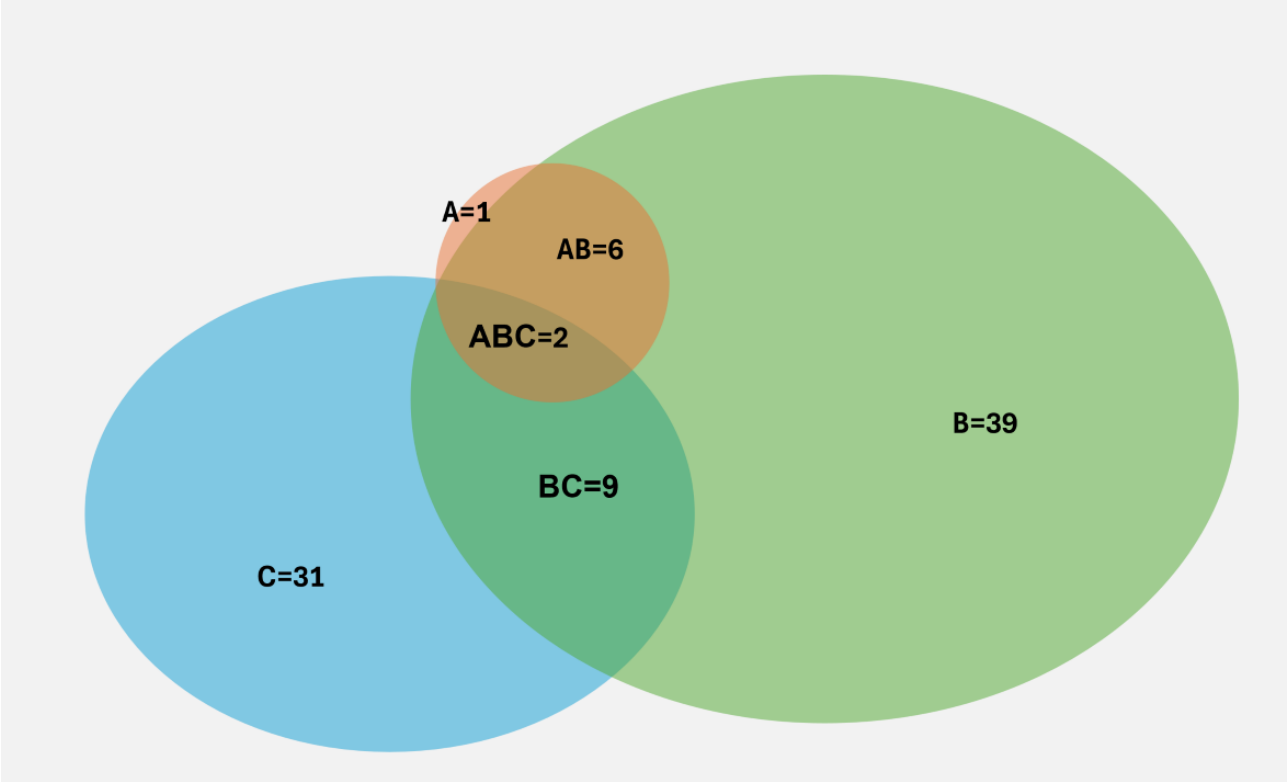
Study Characteristics

Mapping to RQs

All 88 studies for data extraction were classified into 3 predefined RQs (RQs-A, B, or C; A presenting user experience and perception, B presenting monitoring tools, and C presenting technology-assisted approaches. Articles were labeled by relevancy (A, B, C, AB, BC, AC, and ABC) in the extraction table. The Venn diagram (Figure 2) illustrates the thematic overlap among the included studies. The largest group of studies (n=39, B1-B39 as listed in Column 2 of Table 1) addressed only B, focusing primarily on subjective monitoring methods such as questionnaires and self-reports, and the next largest group of 31 studies (C1-C31 as listed in Column 2 of Table 1) addressed C alone, most of which involved small-scale or pilot implementations of technologies, such as sensors and smart orthosis. Only 2 articles [126,127] addressed all 3 RQs, 6 addressed RQs A and B [113-118], and 9 addressed RQs B and C [5,19,119-122,124,125]. These distributions indicate a persistent integration gap between clinical (subjective) and technology-based (objective) monitoring in adolescent idiopathic scoliosis care. Furthermore, the results highlight that no published study meets the eligibility criteria of adolescents diagnosed with idiopathic scoliosis and aged 10-17.11 years, which integrates user experience and perception with technology-assisted approaches (RQs A and C, n=0).

Two main themes that evolved from the findings are (1) subjective (patient-reported or user experience) monitoring using questionnaires with patient-reported outcomes, interviews, and logbooks capturing the patient experience; and (2) objective (technology-assisted or sensor-based) monitoring using sensor-based systems measuring parameters such as temperature, pressure, force, and motion, and a real-time, clinically integrated support system. Integrated approaches combine both. Each included study is coded into these categories S, O, or I, as illustrated in Table 1.

Figure 2. Visual illustration of study characteristics based on 3 RQs. A: User experience and perception (RQ-A), B: Monitoring Tools (RQ-B), and C: Technology-Assisted Approaches (RQ-C).



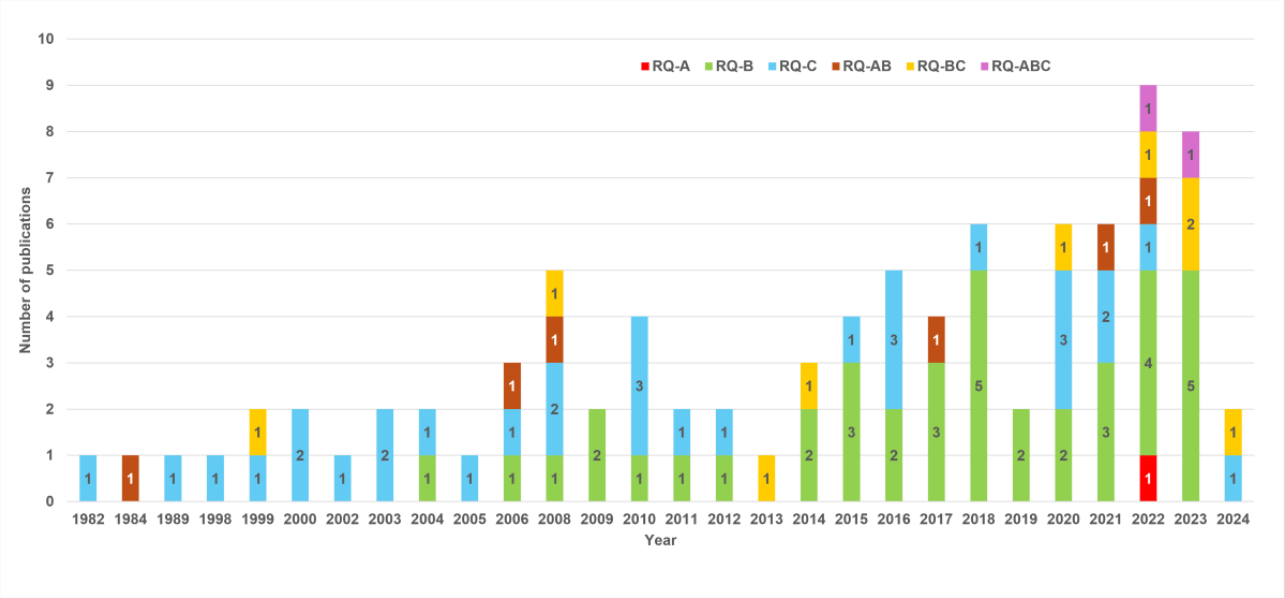
Publication Trends

The number of publications in the focus area has increased over time, as illustrated in Figure 3. Technology-based monitoring has been in use for over 4 decades and accounts for 47.73% (42/88), where 42 studies are categorized into 3 subgroups (C1-C31, BC1-BC9, and ABC1-ABC2), as listed in Column 2 of Table 1. The evolution of articles over time indicates that research on technology-based monitoring has a long history, underscoring the need for updates as technology continues to advance.

Research incorporating patient-reported outcomes emerged approximately a decade later. Once these tools were adopted, an increasing number of studies began to focus on capturing previously unheard patient needs as part of the treatment burden. Articles that combined subjective monitoring tools with sensor-based approaches demonstrate a trend since 1999 toward integrating subjective (RQ A or B) and objective (RQ C) measures. Since 2022, research on user-centered integrated monitoring has gained momentum. These trends are illustrated as publications over time, with research areas and their combinations (Figure 3).

XSL•FO
RenderX

Figure 3. Number of publications over time based on the three research questions A, B, and C.



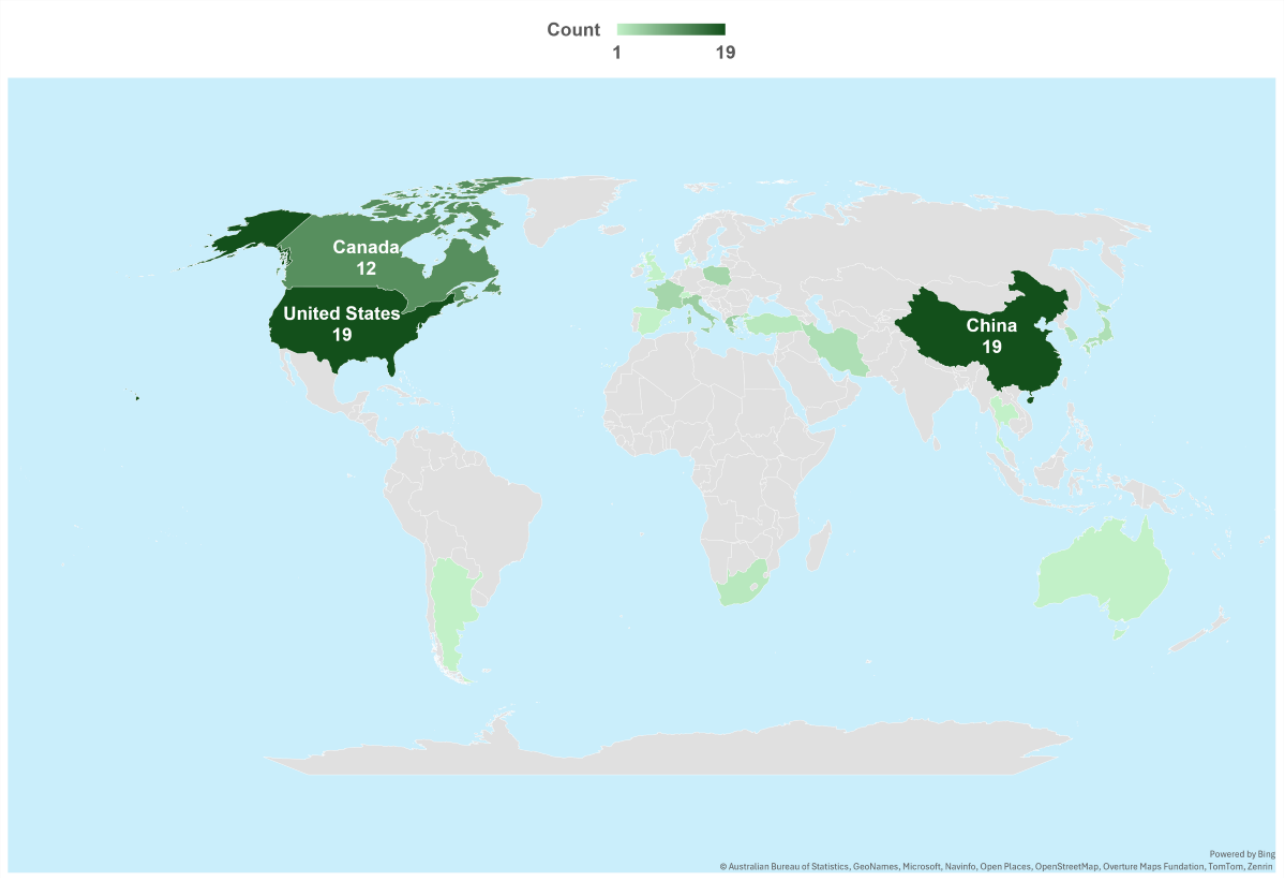
Global Contributors

Research output was concentrated in a few countries: the United States [5,19,23,29,68,70,78,82,89,94,96,102,105,108,109,111,116,117,127] and China [30,46,47,58,62,75,76,77,84,88,92,93,100,115,120-124,125,126] (n=19 each) contributed the most, followed by Canada [57,63,86,87,91,95,98,99,103,107,110,119] (n=12); other contributors included Italy [61,71,72,81,97] and Greece [49,60,67,85,114] (n=5 each), France [90,104,112,118] and Poland [51,69,74,79] (n=4 each), Iran [48,50,65], Japan [113,122,123], and South Korea [54,55,56] (n=3 each), South Africa [59,64], and Turkey [53,80] (n=2 each), and Argentina [66], Australia [83], Denmark [73], Spain [52], Switzerland [101], Thailand [22], and the United Kingdom [106] (n=1 each; Figure 4 and Table 1).

The observed trends indicate that research in multidisciplinary health and technology domains with a user-centered focus is concentrated in a few leading countries, such as China, the United States, and Canada, with small contributions from other regions. Subjective and objective monitoring with the patient as a participant requires approval in accordance with the relevant country's ethical guidelines. Geographic distribution shows the ethical constraints and regulatory challenges in conducting studies that integrate health and technology among participants diagnosed with diseases, such as idiopathic scoliosis.

XSL•FO
RenderX

Figure 4. Visual representation of global contributors among the 88 included publications.



Orthosis Types

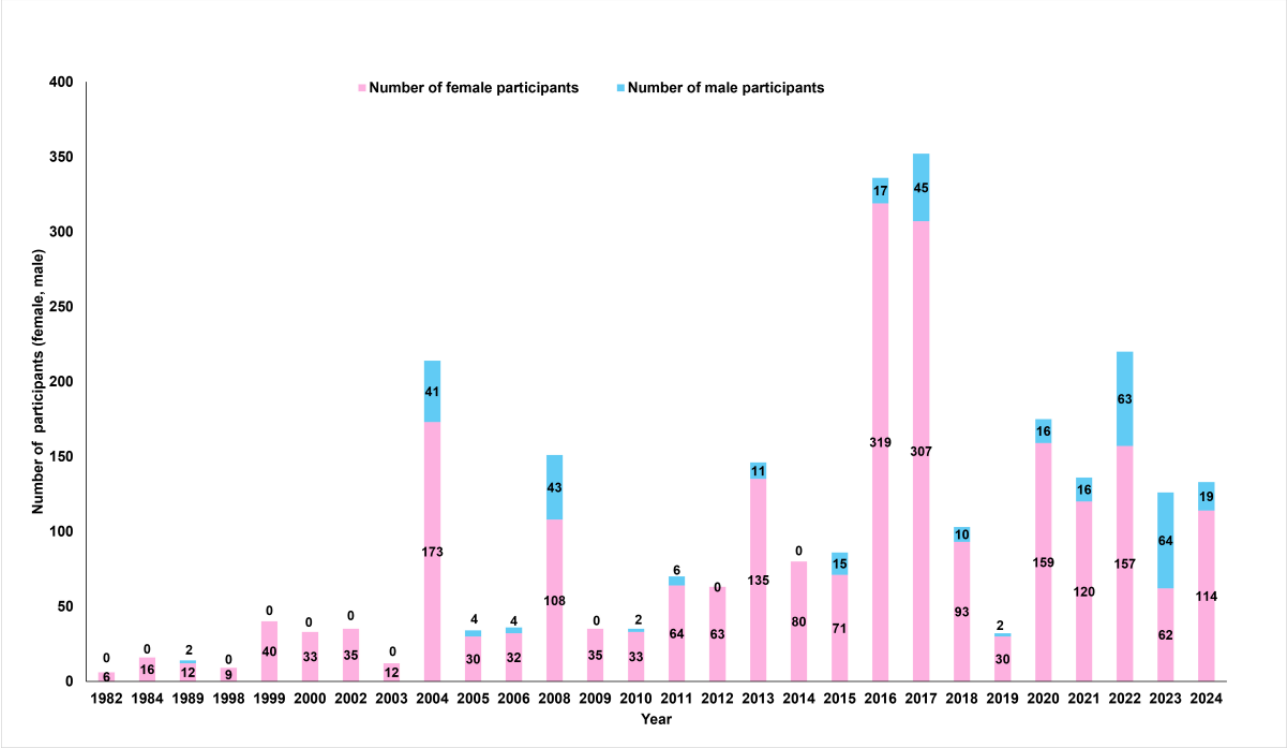
Across the 88 studies, a range of rigid and soft orthoses was reported, including Boston [60,89,103,107,116,119], Milwaukee [65,88], Chêneau [30,71,74,104], and Rigo System Chêneau [59,64], along with various Thoracolumbosacral orthoses, such as rigid [70,72], underarm [120], and other types include the Sforzesco brace [71], Providence [73,102], Charleston, Wilmington [23,96], SpineCor [115], dynamic derotation brace [85], CAEN brace [90] and asymmetric Scoliosis Brace by SLC clinic [49] as well as fiberglass [97], custom-fabricated [76], 3D-printed [125], garment [100], and semirigid [119]. Unfortunately, some articles [114,118] report on the use of braces or orthosis without specifying the type. Some articles report the use of orthosis without reporting the exact type or other specifications (eg, studies by Lin et al [121,125]), while some full-text reports the patient's use of orthosis. For example, the study by Sapountzi-Krepia et al [114] mentions that the orthosis was worn for more than 12 hours, while Bertuccelli et al [81] report on orthotic treatment. Details from all included articles are reported in Table 1. The orthosis used in that study will influence participants' perspectives and patient-reported outcomes. Given the broad scoping focus on analyzing generic challenges in scoliosis orthotic management, we consolidated the observations based on the evolved themes.

Sex

Sex distributions broadly reflect the higher female prevalence [1,2], with negligible male representation until 2004 and increases thereafter (Figure 5). Although male individuals remain underrepresented overall, their contribution to research has improved since 2004. Studies that explore sex-specific needs and perspectives have yet to be reported in detail. The use of patient-reported tools in research has stimulated researchers since 2004 to capture the distinctive needs of male individuals, which differ from those of female individuals, especially during adolescence. In 1 study that used a comparison of bracing versus observation reported no significant overall health-related quality of life (HRQoL) reduction with bracing; however, they reported about family-activity difference favoring male individuals [78]. More focus is needed to know sex-based monitoring preferences in the future, especially for adolescents treated with orthoses.

XSL•FO
RenderX

Figure 5. Sex distribution trend of all included articles.



Subjective Monitoring (RQ A and RQ B)

Subjective monitoring in the management of adolescent patients with idiopathic scoliosis primarily evolved from research on patients' lived experience and perceptions, as reflected in RQ A and traditional clinical monitoring approaches relevant to RQ B. These strategies use different tools to capture patients' needs. For example, clinicians rely on self-reports, relevant questionnaires to address needs (eg, stress or quality of life), and interviews to capture user perceptions. Over time, they have become central to understanding psychosocial and cultural factors influencing orthosis compliance.

Questionnaires as a Subjective Monitoring Tool

Questionnaires and patient logbooks capture user compliance and user-reported data that support subjective scoliosis management for adolescent patients with idiopathic scoliosis [76]. Studies verify that exercise with bracing is associated with improvements in quality of life and functional outcomes, using PROMs or logbooks [58,62,63]. Comparative studies recorded efficacy differences between rigid versus flexible orthoses and captured discomfort and acceptance via a visual analog scale questionnaire, along with open-ended items [128]. The use of "open-ended" items in questionnaires, along with other sensor monitoring in a mixed design, was found beneficial for evaluating patient needs [124]. A combination of subjective data from interviews with sensor-based objective data reports about mal-appearance, inconvenience, and psychological distress [126]. Future studies should integrate PROMs with objective sensor-derived data to provide a more comprehensive understanding of user experiences and treatment adherence.

Among the included studies, subjective monitoring reports the use of Scoliosis Research Society-22 Patient Questionnaire, Brace Questionnaire, Bad Sobernheim Stress Questionnaire (Brace/Deformity), Italian Spine Youth Quality of Life, EQ-5D-5L, and 36-Item Short-Form Health Survey [51,52,55,67,71,75,76,124]. A multicenter study found that adolescents with more severe scoliosis underestimated their deformity and body image. Perception tools Walter Reed Visual Assessment Scale, Spinal Appearance Questionnaire, and Trunk Appearance Perception Scale were also used as a clinical tool to capture user perception [118]. The medium of communication to express influences subjective monitoring. The language of the questionnaire tool, including verbal and pictorial illustrations, is used to capture the user's perspective on the disease condition, and cultural adaptation of questionnaires based on the research focus can influence the expression of that perspective. Other main instruments included Pediatric Quality of Life Inventory [82], Posterior Trunk Symmetry Index [80], Brief Pain Inventory [83], and psychological scales (State-Trait Anxiety Inventory, Beck Depression Inventory-II) [84] to capture often unexpressed psychological aspects of adolescent patients with idiopathic scoliosis burden.

Cross-Cultural Adaptations of Monitoring Tool

Cross-cultural adaptation of the standardized questionnaires is the strategy used to improve user acceptability and relevance among culturally diverse populations. Among the articles included in the review, 6 studies adapted the Bad Sobernheim Stress Questionnaire from the base language to another target language (Japanese, Spanish, Persian, Greek, Polish, and Chinese) [47-49,51,52,113], 2 adapted the Brace Questionnaire (Turkish and Korean) [53,56], and 1 each adapted to Korean Quality of Life Profile for Spine Deformities [54], Korean Italian Spine Youth Quality of Life [55]. Across these studies, the Scoliosis Research Society Patient Questionnaire-22 (and its revised version) was commonly used as a reference instrument for comparative analysis. The adaptation processes follow an

XSL•FO
RenderX

established methodological process, incorporating forward-backward translation, expert review, pilot testing, and evaluation that includes test-retest reliability, internal consistency, and validity with reference tailored to the study objectives (eg, study by Jafarian et al [48]). Collectively, these articles address RQ B by demonstrating the utility of culturally adapted questionnaires for monitoring, and RQ A by capturing patient perceptions within routine care.

User Perceptions

Given the scoping focus on analyzing long-term challenges in scoliosis orthotic management, we consolidate generic user experience and perceptions. In a qualitative study, patient interviews revealed that they were afraid of what the doctors might say during the next feedback session [114]. Orthosis compliance and outcomes vary based on the orthosis used by the participant. For instance, the Milwaukee brace, a type of orthosis, was reported to cause emotional strain and barriers, including issues with clothing fit, restricted movement, and social limitations [117]. Across the included studies, physical and emotional challenges related to orthosis wear were consistently noted: discomfort, restricted movement, heat or climate-related issues, clothing difficulties, and social limitations. The only article from RQ A that used patient-reported data from in-depth semistructured interviews with Chinese adolescents who described activity- and heat-related discomfort and a desire for more personalized, co-designed orthoses. Active patient participation in the intervention is a recommendation [46]. Chinese randomized trial reports the use of a specially designed thin undergarment that improved comfort and compliance [76]. Although the study was conducted in China, similar climatic and seasonal conditions may also influence orthosis use in patients from other geographical regions. Family dynamics (eg, parental reminders) also influenced adherence and contributed positively to scoliosis management among adolescents. These patterns address RQ A user experience and perceived barriers or facilitators, and RQ B, especially the use of clinical tools such as logbooks to monitor adherence.

Psychological or Emotional Well-Being

Bracing in adolescent patients with idiopathic scoliosis has a significant impact on HRQoL, reporting psychological, motor, social, and school-related challenges [66]. Studies use tools, for example, Body Appreciation Scale (BAS) and trunk appearance perception scale to examine body aesthetics and user perception that considers self-image, trunk perception, and psychological outcomes [68,118]. Reduced body-related attentional biases are reported with the use of psychotherapy [81]. Long-term orthotic treatment is demanding. Still, progress in trunk correction and self-visualization offers a sense of positivity. Studies have demonstrated an optimistic view of bracing when studied using the trunk appearance perception scale as orthoses can influence trunk correction, leading to greater satisfaction with orthotic treatment [61]. Likewise, using another visualization score, the BAS gave a similar perspective from the orthosis-treated and control group [68]. This underscores the need for interdisciplinary psychological and social support for adolescent patients with idiopathic scoliosis.

Family Context

The patient's family also played a significant role in handling challenges and managing scoliosis, as reported in the articles included. In Argentina, a study using a brace questionnaire recommended separate patient and parent surveys to reduce response bias [66]. Polish studies using Bad Sobernheim Stress Questionnaire-brace or deformity found comparable stress in adolescents and parents, with parents overestimating deformity-related stress; spinal appearance questionnaire comparisons emphasized the roles of both mothers and fathers [69,74]. Accepting cultural diversity, the roles of both parents, mother and father, and family bonding are recognized as influential parameters for adolescents' emotional coping.

Societal or Socioeconomic Context

Subjective tools identified peer perceptions at school as sources of stress or anxiety [113]. In a Spanish cohort using Bad Sobernheim Stress Questionnaire-brace, no embarrassment was reported among adolescents wearing Rigo System Chêneau orthosis with close family and friends at school [52]. Financial burdens were noted (orthosis and consultation costs) [51]. Rural versus urban comparisons showed similar moderate orthosis-related stress and low deformity stress, with greater appearance concern among rural participants with larger lateral shift [79].

To sum up the findings, addressing RQs A and B by mapping psychosocial parameters relevant to subjective monitoring (eg, sex, well-being, family, and school or society) influences patient perspectives on stress, body image, and daily life.

Objective Monitoring (RQ-C)

Sensor Types and Parameters

Early systems (1980s) were developed in controlled settings with static assessments. Over time, objective monitoring expanded to include electromyography (EMG), pressure or force, temperature, and motion sensors that captured orthosis use, activity, and physiological responses.

EMG

EMG is used to evaluate neuromuscular responses to bracing. There are limited (n=3) studies using EMG. The oldest included study, from 1982 in Canada, reported that it was unable to detect any observable trend among the 6 female individuals [110]. Even though the results were not promising, objective monitoring to understand the effect of orthosis began to evolve. Moreover, 2 other studies that used EMG demonstrated how orthosis improved spinal stability by stimulating muscle engagement [95] and the feasibility of wearable EMG sensor-based posture tracking for daily activities [111].

Motion

Movement tracking represents activity in daily life; 1 study reported no significant difference in walking activity between adolescent patients with idiopathic scoliosis and controls using a smartphone-based approach [112]. Multimodal systems combining accelerometers, gyroscopes, and force are gaining research interest recently [29].

XSL•FO
RenderX

Pressure or Force

Studies monitored strap or pad forces and interface pressures to support fit, personalization, and compliance, ranging from calibrated measurements to embedded or load-monitoring systems and mapping approaches linking tension, pressure, and correction [103,106,107,129]. One system reported a decline in tightness without monitoring and improved compliance with monitoring or refitting [29]. Works characterized different research interests that include corrective force patterns [87,89], explored respiratory mechanics under bracing [93], examined posture-specific force changes [85], and used exercise-related pressure patterns to inform personalized exercise selection [97]. Adjustable orthoses logged pressure and temperature over extended periods [121].

Temperature

Temperature sensors quantified wear time with high accuracy and were used to explore adherence patterns. Prospective cohorts recorded orthosis wear at set intervals over months and reported associations between higher compliance and improved outcomes, with features for time-based feedback, as reported in [23,96,105]. Other studies that use temperature-based reporting monitoring report high accuracy in estimating daily wear, increased daily wear with compliance counseling, and associations between reduced adherence and curve progression in night-time bracing [102,108,109]. Studies compared orthosis types, estimated compliance in dynamic systems, and examined temporal patterns (for eg, higher weekday or night-time wear; relationships with stress and satisfaction) [5,19,101,122,123,125]. Together, these sensor families demonstrate the feasibility of objective monitoring of muscle activity, strap or pad forces and interface pressure, wear time, and movement or activity during orthotic treatment.

Wear Time and Adherence

In temperature-based studies, wear time was derived from periodic logging and compared with diaries and/or clinical outcomes; high accuracy versus diaries was reported, and adherence counseling and visit-time feedback were associated with increased wear [23,96,108,109]. Findings describe differences by time period (eg, weekday or night-time patterns), declines in orthosis tightness over time without monitoring, and improved adherence with feedback-enabled systems [29,122]. In night-time bracing, lower adherence was associated with curve progression [102]. Large cohort data supported a dose-response relationship between longer wear and reduced progression to surgical thresholds [5]. These studies address RQ C by quantifying orthosis wear time and adherence patterns, and by relating adherence to treatment outcomes.

Integration of Subjective and Objective Monitoring

Integrated clinical monitoring, along with relevant data visualization, is gaining importance due to its design strength. Integrated monitoring of adolescents with idiopathic scoliosis orthotic treatment combines subjective patient-reported outcomes with objective sensors to track both perceived burden and real-time adherence. Integrated monitoring combines 9 articles with RQs B and C, which include compliance monitoring and interviews, as well as a randomized controlled trial using automated pressure-adjustable orthoses and variant temperature or heat sensors, linking wear quality with HRQoL outcomes [5,19,119-125]. Moreover, 2 studies that demonstrate the integration of 3 RQs with an all-around monitoring approach that merges user perspectives with clinical and technological approaches, namely qualitative interviews and psychosocial questionnaires, to capture lived experience [126,127]. Integration of subjective and objective monitoring is progressing.

Real-Time Monitoring

This scoping review tries to collect evidence from all selected databases up to the search date, which allowed the inclusion of engineering articles from the 1980s and 1990s. These early studies reflect efforts to measure orthosis adherence objectively; however, they relied on nonportable devices, limiting monitoring to controlled laboratory environments. Research growth is an indicator of how technology and health research are advancing over time, with a growing focus on user-focused solutions that include portable and compact solutions. At the same time, objective sensors and adherence logs provide real-time data. Integration of patient needs, cultural context, and comfort, along with biomechanical and compliance metrics, can provide assurance in scoliosis management and serve as a convergence that enables accurate adherence measurement, personalized feedback, and improved clinical decision-making for patient-centered care.

One configuration involved force sensors with a mobile mini-program for daily uploads, enabling remote clinician monitoring and scheduled follow-ups, and was evaluated in 28 adolescents, with satisfaction assessed at 6 months [30]. The design load-monitoring system for objective monitoring developed by Lou et al [98] tracks both the quantity and quality of orthosis wear. The combination of objective data supported with PROMs both supports adherence and patient-reported experience [124]. These findings demonstrate the feasibility of real-time or connected monitoring with clinic-linked follow-up, complementing subjective tools to provide a more complete view of orthosis use.

These innovations target continuous data collection and personalized solutions that help health personnel and empower adolescent users to play an active role in their treatment, reducing the effort required of health personnel [26,30,31]. Advancements could lead to patient-friendly, effective scoliosis management and improve long-term outcomes with more focus on distinct aspects of monitoring adolescent patients with idiopathic scoliosis.

Discussion

Summary of Key Findings by RQs and Themes

The review identified 2 central themes, namely subjective monitoring and objective monitoring, with limited integration. RQs A and B capture subjective monitoring that relies on patient-reported outcomes and qualitative user-experience data, consistently documenting discomfort, restricted movement, heat or climate concerns, clothing challenges, peer-related anxiety, and the influence of family dynamics. Cross-cultural adaptations were common and used standard translation procedures. The

XSL•FO
RenderX

exercise program in combination with orthotic treatment improves quality of life as well as body function. The RQ C, objective monitoring demonstrated feasibility using temperature (wear-time quantification), pressure or force (fit, personalization, and corrective load), EMG-based neuromuscular response, and motion (activity or posture). Temperature-based monitoring showed high accuracy for wear-time estimation and was associated with increased adherence when coupled with counseling or feedback; dose-response patterns linked greater wear with reduced progression. The strict age restrictions, along with the exclusion of studies involving mixed-age idiopathic scoliosis cohorts, probably narrowed the evidence base and reduced the inclusion of potentially relevant sensor-based research. Connected systems enabling remote follow-up were feasible in small samples.

Clearly distinguishing between subjective and objective monitoring approaches is clinically meaningful. While objective tools offer precise and continuous measurements of orthosis wear and biomechanical parameters, they do not capture how adolescents experience orthotic treatment in daily life. Conversely, patient-reported measures provide critical insight into comfort, usability, and psychosocial impact but may be influenced by recall bias or social desirability. Integrated approaches that combine both perspectives appear particularly promising, as they acknowledge that treatment adherence and effectiveness are shaped by both measurable behaviors and lived experience. This framework helps contextualize current evidence and highlights directions for future digital health development in scoliosis care.

Interpretation and Clinical Implications

Different PROMs and qualitative tools capture adherence to comfort, climate, clothing, and psychosocial context, which are not accessible from imaging alone, while sensor-based systems provide objective quantitative data on wear time and the quality of orthosis wear. A combination of diverse modalities supports individualized adjustments (eg, orthosis fit or tightness and targeted exercises) by providing tailored psychosocial support. Family involvement and feedback mechanisms were associated with better adherence.

Mixed methods designs that link temperature-based adherence with HRQoL instruments illustrate how clinicians can monitor both use and impact to guide timely interventions. Although Thermobrace was excluded (due to participants with a diagnosis of hyperkyphosis along with remaining idiopathic scoliosis), temperature-based approaches, in general, were used and reported accurate wear-time estimation [27]. A recent Chinese pilot study reports the use of an integrated sensor, namely force and temperature, that provides a comprehensive understanding of the compliance by demonstrating data from 12 patients by comparing subjective and objective monitoring for 1 month and found that patients' self-reported data exceeded the sensor's actual average wear time [24]. User acceptance of objective monitoring provides reassurance and perceived helpfulness as a long-term supportive tool [17]. A recent study reports that subjective patient-reported data, along with automated alerts to clinicians at predefined thresholds, can increase usability and engagement, supporting users in meeting prescribed orthosis-wear recommendations during the study period [130].

Implications for Research and Future Directions

Evidence collected over time across multidisciplinary technological and clinical research suggests a gap for integrated user-centered approaches. Studies need to include standardized monitoring protocols, detailed reporting of clinically relevant participant demographics, and harmonized outcomes, which would improve comparability and inclusiveness of research across disciplines. In the future, subjective monitoring questionnaires are to be culturally adapted and validated, especially focusing on orthotic treatment (eg, study by Caronni et al [131]). Research could also expand sex-specific analyses, evaluate acceptability across cultures, and test multimodal systems (eg, temperature, pressure, or motion) in larger, longitudinal cohorts that support patient rehabilitation.

An integrated workflow can combine subjective PROMs with objective technology-based monitoring, supported by structured clinical interventions and timely feedback, to enhance adherence throughout treatment and patient rehabilitation [130]. For example, real-time review could empower adolescent users and their families by enabling earlier identification of nonadherence and supporting personalized treatment adjustments. The system needs to ensure user and patient security and trustworthiness, and provide patient-specific guidance [132], in line with the nation's protocols.

The continued gap in integrating subjective patient experiences with technology-assisted monitoring might be due to structural barriers, such as the stages of technology readiness of emerging innovations for integration into clinical workflows, existing ethical complexities in clinical trial approval processes, and user acceptance and trust in new technologies. The integration of subjective and objective monitoring advances slowly. Technological evidence must be rigorously validated before it can be used in a clinical trial. Thus, innovative technologies are typically evaluated through phased testing to improve technology readiness. Objective sensor data (eg, body-pressure distribution on a scoliosis orthosis) are tested on a mannequin or a laboratory-based test environment with clinically relevant pressure patterns, serving as the preparatory analysis for the future clinical trial [133]. These stages of testing and preparation ensure feasibility, safety, and effectiveness.

Standardized protocols should be established for data handling and clinically aligned technology-based monitoring to support personalized, data-driven care for adolescents with idiopathic scoliosis. Detailed testing has been conducted and reported to prepare for future clinical testing of a real-time contact pressure monitoring solution and highlights that future monitoring solutions for adolescent idiopathic scoliosis during bracing need to consider patient comfort [134].

Future research directions across the 3 RQs highlight critical gaps and opportunities. For RQ A, there is a need to strengthen the focus on lived experiences and user perceptions, which remain the least represented in the current literature. Currently, no existing review integrates (RQ A) user perspectives, wishes, and needs with technology-based monitoring (RQ C) in the

XSL•FO
RenderX

context of adolescent idiopathic scoliosis. For RQ B, monitoring tools should evolve to include questionnaires that address technological perspectives, usability challenges in daily life, concerns related to personal and health data security, and new needs. For RQ C, technology updates should be integrated into the health care system that uses real-time monitoring and time-bound clinically significant feedback. In the future, co-design of systems with stakeholders can help to develop a user-friendly solution that can be widely accepted and suitable for rigorous use.

Challenges

Methodological diversity among the included articles complicated appraisal using established tools Critical Appraisal Skills Program, Joanna Briggs Institute, and Mixed Methods Appraisal Tool, which were not compatible with engineering-focused studies that often lack a clinical structure, underscoring the need for frameworks that bridge engineering and clinical research. Variability was observed in the reporting of CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist items related to methodological details across included studies, with more comprehensive reporting in health science–focused research (eg, cultural adaptation of questionnaires) and comparatively limited descriptions in technology-focused studies, where standardized questionnaires were often used as data collection tools secondary to technological evaluation [135]. Informal appraisal without formal methodological appraisal proceeded alongside iterative data coding, which is acceptable in scoping reviews.

Inconsistent reporting of participant details (age, diagnosis, and orthosis type) was more common in engineering papers; lack of focus on RQs and emphasis on technology alone led to exclusions. Along with orthosis monitoring, soft orthosis, including garment orthosis and vests, were included in the scope. However, due to the strict requirements for patient-based studies, many studies were excluded because they often lacked patient participation criteria. Only patient-tested solutions were retained, which led to the exclusion of several early-stage prototypes under RQ C.

Strengths

Having a specific and focused target group, adolescents diagnosed with idiopathic scoliosis, led to the exclusion of technologically noteworthy studies. On the other hand, this focus strengthened the scoping review's ability to identify gaps and highlight future needs toward interdisciplinary research. The inclusion of diverse article types spanning medical and engineering fields enriched the analysis. The 2 independent reviewers, one with a medical background and the other with a technological background, helped with this process. Screening benefited from Rayyan-supported comparisons; however, final inclusion and exclusion decisions were made manually, using human-judgment-guided selection.

Limitations

This review is limited by sample size constraints, lack of longitudinal data, poor reporting in engineering studies, and insufficient risk of bias evaluation, which may affect the robustness of the evidence base. The review included articles from the listed databases that are digitally available and English-language publications, leading to the exclusion of relevant studies that were published in other language publications, nondigital form, or sourced from unindexed repositories.

Our strict age inclusion criteria (10 - 17.11 y) excluded multiple technology-focused papers with mixed-age samples, likely reducing the number of eligible sensor-based studies. Clinically significant parameter Cobb angle was considered during protocol development. However, a broader approach was adopted to accommodate variability in reporting, especially in nonhealth publications.

Conclusions

This scoping review of 88 included studies reveals the clear growth in adolescent patients with idiopathic scoliosis conservative monitoring from separate subjective and objective approaches toward integrated, patient-centered models. Questionnaire-based tools remain essential for capturing psychosocial burden, body image, and cultural context, but they cannot measure real-time adherence or biomechanical quality. Technological solutions that use sensors provide objective information, progressing from static laboratory setups to portable, connected systems that enable remote oversight and timely feedback. Subjective tools and objective sensors provide complementary views of adolescent orthosis wear. Combining patient-reported outcomes with sensor data offers the greatest potential, as it provides a multidimensional view of adherence, comfort, and clinical effectiveness. However, reporting inconsistencies and limited scalability persist, especially in technology-focused research. In the future, research should better integrate health and technology disciplines, integrate culturally adapted objective and subjective metrics, and co-design a user-centered solution that motivates adolescents to ensure usability. Moving beyond pilot studies to mixed method clinical trials and real-time platforms will enable personalized care, improve adherence, and optimize outcomes. The convergence of subjective and objective monitoring marks the shift toward holistic management of adolescent patients with idiopathic scoliosis, aligning clinical precision with patient motivation.

Acknowledgments

We are grateful to OsloMet librarians Elisabeth Karlsen and Ingjerd Legreid Ødemark for their assistance in developing, documenting, and implementing the tailored search criteria for this scoping review. We are also grateful for support from Klaus Jøran Tollan and Vegard Syrstad, OsloMet librarians for cleaning and exporting DOI, and for the collection of full texts, respectively. The generative AI -powered tools: Microsoft Copilot was used to condense and restructure text for clarity, and Grammarly was used for language correction.

XSL•FO
RenderX

Funding

This work is part of an internally funded PhD fellowship at Oslo Metropolitan University. The authors acknowledge the financial support from OsloMet towards publication. No external funding was received.

Data Availability

Data are made available in Multimedia Appendices 1 and 2.

Authors' Contributions

AM and PG contributed equally to investigation, data curation, resources, software, formal analysis, writing- original draft. AM, PG, MP, and KS participated in methodology, validation, and visualization. PG, NP, MP, and AM contributed to the conceptualization and design. PG, NP, and MBJ contributed to project supervision and project administration. All authors contributed to reviewing and editing the manuscript and approved the final version based on their respective professional backgrounds (AM, engineer; KS and MBJ, physiotherapists; NP, product designer; MP, digitalization expert in the health sector; SA, orthopedic engineer; MFK, scoliosis patient organization representative; and PG, health and technology researcher).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Literature search documentation.
[DOCX File, 413 KB - mhealth_v14i1e88880_app1.docx]

Multimedia Appendix 2
Data items.
[XLSX File, 94 KB - mhealth_v14i1e88880_app2.xlsx]

Checklist 1
PRISMA ScR checklist.
[DOCX File, 112 KB - mhealth_v14i1e88880_app3.docx]

References

1. Janicki JA, Alman B. Scoliosis: review of diagnosis and treatment. Paediatr Child Health 2007 Nov;12(9):771-776. [doi: 10.1093/pch/12.9.771] [Medline: 19030463]
2. Negrini S, Donzelli S, Aulisa AG, et al. 2016 SOSORT guidelines: orthopaedic and rehabilitation treatment of idiopathic scoliosis during growth. Scoliosis Spinal Disord 2018 Jan 10;13(1):3. [doi: 10.1186/s13013-017-0145-8] [Medline: 29435499]
3. Negrini S, Aulisa AG, Aulisa L, et al. 2011 SOSORT guidelines: orthopaedic and rehabilitation treatment of idiopathic scoliosis during growth. Scoliosis 2012 Jan 20;7(1):3. [doi: 10.1186/1748-7161-7-3] [Medline: 22264320]
4. Negrini S, Donzelli S, Lusini M, Minnella S, Zaina F. The effectiveness of combined bracing and exercise in adolescent idiopathic scoliosis based on SRS and SOSORT criteria: a prospective study. BMC Musculoskelet Disord 2014 Aug 6;15(1):263. [doi: 10.1186/1471-2474-15-263] [Medline: 25095800]
5. Weinstein SL, Dolan LA, Wright JG, Dobbs MB. Effects of bracing in adolescents with idiopathic scoliosis. N Engl J Med 2013 Oct 17;369(16):1512-1521. [doi: 10.1056/NEJMoa1307337] [Medline: 24047455]
6. Donzelli S, Zaina F, Negrini S. Predicting scoliosis progression: a challenge for researchers and clinicians. EClinicalMedicine 2020 Jan 6;18:100244. [doi: 10.1016/j.eclinm.2019.100244] [Medline: 31993579]
7. Motyer GS, Kiely PJ, Fitzgerald A. Adolescents' experiences of idiopathic scoliosis in the presurgical period: a qualitative study. J Pediatr Psychol 2022 Feb 14;47(2):225-235. [doi: 10.1093/jpepsy/jsab095] [Medline: 34524430]
8. Fregna G, Rossi Raccagni S, Negrini A, Zaina F, Negrini S. Personal and clinical determinants of brace-wearing time in adolescents with idiopathic scoliosis. Sensors (Basel) 2023 Dec 25;24(1):116. [doi: 10.3390/s24010116] [Medline: 38202978]
9. Ali A, Fontanari V, Fontana M, Schmölz W. Spinal deformities and advancement in corrective orthoses. Bioengineering (Basel) 2020 Dec 25;8(1):2. [doi: 10.3390/bioengineering8010002] [Medline: 33375594]
10. Zaina F, Atanasio S, Negrini S. Clinical evaluation of scoliosis during growth: description and reliability. Stud Health Technol Inform 2008;135:125-138. [Medline: 18401086]
11. Toh ZA, Berg B, Han QYC, et al. Clinical decision support system used in spinal disorders: scoping review. J Med Internet Res 2024 Mar 19;26:e53951. [doi: 10.2196/53951] [Medline: 38502157]
12. Negrini S, Grivas TB, Kotwicki T, Rigo M, Zaina F, International Society on Scoliosis Orthopaedic and Rehabilitation Treatment (SOSORT). Guidelines on "Standards of management of idiopathic scoliosis with corrective braces in everyday

XSL•FO
RenderX

clinics and in clinical research": SOSORT consensus 2008. Scoliosis 2009 Jan 16;4(1):2. [doi: 10.1186/1748-7161-4-2] [Medline: 19149877]
13. Martínez-Borba V, Suso-Ribera C, Díaz-García A, et al. mHealth for the monitoring of brace compliance and wellbeing in adolescents with idiopathic scoliosis: study protocol for a feasibility study. Int J Environ Res Public Health 2021 Jul 22;18(15):7767. [doi: 10.3390/ijerph18157767] [Medline: 34360060]
14. Chan A, Lou E, Hill D. Review of current technologies and methods supplementing brace treatment in adolescent idiopathic scoliosis. J Child Orthop 2013 Oct;7(4):309-316. [doi: 10.1007/s11832-013-0500-0] [Medline: 24432092]
15. Ghorbani F, Kamali M, Ranjbar H, Kamyab M, Razavi H, Babaee T. Brace compliance process in adolescents with spinal deformities: a qualitative study. PLoS One 2024 Aug 2;19(8):e0305754. [doi: 10.1371/journal.pone.0305754] [Medline: 39093888]
16. Negrini S, Grivas TB, Kotwicki T, et al. Why do we treat adolescent idiopathic scoliosis? What we want to obtain and to avoid for our patients. SOSORT 2005 consensus paper. Scoliosis 2006 Apr 10;1(1):4. [doi: 10.1186/1748-7161-1-4] [Medline: 16759352]
17. Donzelli S, Zaina F, Martinez G, Di Felice F, Negrini A, Negrini S. Adolescents with idiopathic scoliosis and their parents have a positive attitude towards the Thermobrace monitor: results from a survey. Scoliosis Spinal Disord 2017 Apr 8;12(12):12. [doi: 10.1186/s13013-017-0119-x] [Medline: 28405634]
18. Vasiliadis E, Grivas TB, Gkoltsiou K. Development and preliminary validation of Brace Questionnaire (BrQ): a new instrument for measuring quality of life of brace treated scoliotics. Scoliosis 2006 May 20;1(7):7. [doi: 10.1186/1748-7161-1-7] [Medline: 16759366]
19. Morton A, Riddle R, Buchanan R, Katz D, Birch J. Accuracy in the prediction and estimation of adherence to bracewear before and during treatment of adolescent idiopathic scoliosis. J Pediatr Orthop 2008;28(3):336-341. [doi: 10.1097/BPO.0b013e318168d154] [Medline: 18362800]
20. Botens-Helmus C, Klein R, Stephan C. The reliability of the Bad Sobernheim Stress Questionnaire (BSSQbrace) in adolescents with scoliosis during brace treatment. Scoliosis 2006 Dec 19;1(1):22. [doi: 10.1186/1748-7161-1-22] [Medline: 17176483]
21. Tavernaro M, Pellegrini A, Tessadri F, Zaina F, Zonta A, Negrini S. Team care to cure adolescents with braces (avoiding low quality of life, pain and bad compliance): a case-control retrospective study. 2011 SOSORT award winner. Scoliosis 2012 Sep 20;7(1):17. [doi: 10.1186/1748-7161-7-17] [Medline: 22995590]
22. Lertudomphonwanit T, Pengrung N, Kriwattanapong C, Angsanuntsukh C, Leelapattana P, Chanplakorn P. Novel questionnaire to enhance brace wear adherence in patients with adolescent idiopathic scoliosis and the relationship of the quality of life. Orthop Rev (Pavia) 2021 Mar 30;13(1):8840. [doi: 10.4081/or.2021.8840] [Medline: 33907613]
23. Rahman T, Borkhuu B, Littleton AG, et al. Electronic monitoring of scoliosis brace wear compliance. J Child Orthop 2010 Aug;4(4):343-347. [doi: 10.1007/s11832-010-0266-6] [Medline: 21804897]
24. Zou Y, Zhou L, Wang J, Lou E, Wong MS. The application of integrated force and temperature sensors to enhance orthotic treatment monitoring in adolescent idiopathic scoliosis: a pilot study. Sensors (Basel) 2025 Jan 23;25(3):686. [doi: 10.3390/s25030686] [Medline: 39943325]
25. Cordani C, Malisano L, Febbo F, et al. Influence of specific interventions on bracing compliance in adolescents with idiopathic scoliosis-a systematic review of papers including sensors' monitoring. Sensors (Basel) 2023 Sep 4;23(17):7660. [doi: 10.3390/s23177660] [Medline: 37688117]
26. Tymińska P, Zaborowska-Sapeta K, Janczak D, Giżewski T. TLSO with graphene sensors-an application to measurements of corrective forces in the prototype of intelligent brace. Sensors (Basel) 2022 May 25;22(11):4015. [doi: 10.3390/s22114015] [Medline: 35684636]
27. Donzelli S, Zaina F, Negrini S. In defense of adolescents: they really do use braces for the hours prescribed, if good help is provided. Results from a prospective everyday clinic cohort using thermobrace. Scoliosis 2012 May 31;7(1):12. [doi: 10.1186/1748-7161-7-12] [Medline: 22651570]
28. Lee KP, Yip J, Yick KL, Lu C, Lu LY, Lei QWE. A novel force-sensing smart textile: inserting silicone-embedded FBG sensors into a knitted undergarment. Sensors (Basel) 2023 May 28;23(11):5145. [doi: 10.3390/s23115145] [Medline: 37299872]
29. Dehzangi O, Bache BA, Iftikhar O, Wensman J, Li Y. A smart point-of-care compliance monitoring solution for brace treatment of adolescent idiopathic scoliosis patients. Smart Health 2021 Jul;21:100179. [doi: 10.1016/j.smhl.2021.100179]
30. Zhu C, Wu Q, Xiao B, et al. A compliance real-time monitoring system for the management of the brace usage in adolescent idiopathic scoliosis patients: a pilot study. BMC Musculoskelet Disord 2021 Feb 5;22(1):152. [doi: 10.1186/s12891-021-03976-5] [Medline: 33546648]
31. Li X, Huo Z, Hu Z, et al. Which interventions may improve bracing compliance in adolescent idiopathic scoliosis? A systematic review and meta-analysis. PLoS One 2022 Jul 20;17(7):e0271612. [doi: 10.1371/journal.pone.0271612]
32. Raudenbush B, Simela A, Joseph H. A review of the evaluation, diagnosis, and nonsurgical treatment of adolescent idiopathic scoliosis. Osteopath Fam Physician 2013 Jul;5(4):158-168. [doi: 10.1016/j.osfp.2013.01.009]
33. Joarder I, Taniguchi S, Mendoza A, Snow ME. Defining "successful" treatment outcomes in adolescent idiopathic scoliosis: a scoping review. Eur Spine J 2023 Apr;32(4):1204-1244. [doi: 10.1007/s00586-023-07592-w] [Medline: 36847911]

XSL•FO
RenderX

34. Bertuccelli M, Cantele F, Masiero S. Body image and body schema in adolescents with idiopathic scoliosis: a scoping review. Adolescent Res Rev 2023 Mar 24;8:97-115. [doi: 10.1007/s40894-022-00187-4]
35. Fazeli Veisari S, Bidari S, Barati K, Atlasi R, Komeili A. Wearable devices in scoliosis treatment: a scoping review of innovations and challenges. Bioengineering (Basel) 2025 Jun 25;12(7):696. [doi: 10.3390/bioengineering12070696] [Medline: 40722388]
36. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143. [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
37. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
38. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69. [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
39. Pollock D, Peters MDJ, Khalil H, et al. Recommendations for the extraction, analysis, and presentation of results in scoping reviews. JBI Evid Synth 2023 Mar 1;21(3):520-532. [doi: 10.11124/JBIES-22-00123] [Medline: 36081365]
40. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
41. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
42. Siripanyakhemakul W, Permpool K, Seng-Iad S. Effectiveness of orthotic treatment for adolescent idiopathic scoliosis: a scoping review protocol of systematic reviews. BMJ Open 2023 Nov 8;13(11):e078064. [doi: 10.1136/bmjopen-2023-078064] [Medline: 37940156]
43. Hoelen TCA, de Bie RA, Arts JJ, Willems PC. Methodology and application of three-dimensional technology for brace design and production for treatment in patients with adolescent idiopathic scoliosis: a scoping review protocol. BMJ Open 2023 Dec 9;13(12):e079673. [doi: 10.1136/bmjopen-2023-079673] [Medline: 38070915]
44. Lou E, Hill D, Hedden D, Mahood J, Moreau M, Raso J. An objective measurement of brace usage for the treatment of adolescent idiopathic scoliosis. Med Eng Phys 2011 Apr;33(3):290-294. [doi: 10.1016/j.medengphy.2010.10.016] [Medline: 21112234]
45. Heegaard M, Tøndevold N, Dahl B, Andersen TB, Gehrchen M, Ohrt-Nissen S. Does Risser stage accurately predict the risk of curve progression in patients with adolescent idiopathic scoliosis treated with night-time bracing? Eur Spine J 2023 Sep;32(9):3077-3083. [doi: 10.1007/s00586-023-07808-z] [Medline: 37314578]
46. Cheung MC, Law D, Yip J, Cheung JPY. Adolescents' experience during brace treatment for scoliosis: a qualitative study. Int J Environ Res Public Health 2022 Aug 25;19(17):10585. [doi: 10.3390/ijerph191710585] [Medline: 36078297]
47. Xu X, Wang F, Yang M, et al. Chinese adaptation of the Bad Sobernheim Stress Questionnaire for patients with adolescent idiopathic scoliosis under brace treatment. Medicine (Baltimore) 2015 Aug;94(31):e1236. [doi: 10.1097/MD.0000000000001236] [Medline: 26252283]
48. Jafarian FS, Yeowell G, Sadeghi-Demneh E. Cross-cultural adaptation and validation of the Bad Sobernheim Stress Questionnaire in Iranian adolescents with idiopathic scoliosis using thoracolumbar orthoses. Adv Biomed Res 2022 May 30;11(1):39. [doi: 10.4103/abr.abr_154_21] [Medline: 35814302]
49. Kastrinis A, Koumantakis G, Tsekoura M, et al. Greek adaptation and validation of the Bad Sobernheim Stress Questionnaire-Brace and the Bad Sobernheim Stress Questionnaire-Deformity. Adv Exp Med Biol 2023;1425:141-149. [doi: 10.1007/978-3-031-31986-0_13] [Medline: 37581788]
50. Rezaei Motlagh F, Pezham H, Babaee T, Saeedi H, Hedayati Z, Kamali M. Persian adaptation of the Bad Sobernheim stress questionnaire for adolescent with idiopathic scoliosis. Disabil Rehabil 2020 Feb;42(4):562-566. [doi: 10.1080/09638288.2018.1503728] [Medline: 30325684]
51. Misterska E, Głowacki M, Harasymczuk J. Polish adaptation of Bad Sobernheim Stress Questionnaire-Brace and Bad Sobernheim Stress Questionnaire-Deformity. Eur Spine J 2009 Dec;18(12):1911-1919. [doi: 10.1007/s00586-009-1126-2] [Medline: 19669802]
52. D'Agata E, Testor CP, Rigo M. Spanish validation of Bad Sobernheim Stress Questionnaire (BSSQ (brace).es) for adolescents with braces. Scoliosis 2010 Jul 15;5(15):15. [doi: 10.1186/1748-7161-5-15] [Medline: 20633253]
53. Gür G, Yakut Y, Grivas T. The Turkish version of the Brace Questionnaire in brace-treated adolescents with idiopathic scoliosis. Prosthet Orthot Int 2018 Apr;42(2):129-135. [doi: 10.1177/0309364617690393] [Medline: 28190379]
54. Park SH, Goh TS, Park YG, Kim DS, Lee JS. Validation of a Korean version of the quality-of-life profile for spine deformities (QLPSD) in patients with adolescent idiopathic scoliosis. Eur Rev Med Pharmacol Sci 2022 Jan;26(1):84-89. [doi: 10.26355/eurrev_202201_27751] [Medline: 35049023]
55. Park SH, Goh TS, Son SM, Kim DS, Lee JS. Validation of the Italian spine youth quality of life (ISYQOL) in Korean population. J Clin Neurosci 2021 Oct;92:165-168. [doi: 10.1016/j.jocn.2021.08.015] [Medline: 34509245]
56. Lim JM, Goh TS, Shin JK, Kim DS, Lee CS, Lee JS. Validation of the Korean version of the Brace Questionnaire. Br J Neurosurg 2018 Dec;32(6):678-681. [doi: 10.1080/02688697.2018.1501464] [Medline: 30146912]

XSL•FO
RenderX

57. Théroux J, Stomski N, Innes S, et al. Revisiting the psychometric properties of the Scoliosis Research Society-22 (SRS-22) French version. Scoliosis Spinal Disord 2017;12:21. [doi: 10.1186/s13013-017-0129-8] [Medline: 28725867]
58. Zheng Y, Dang Y, Yang Y, et al. Whether orthotic management and exercise are equally effective to the patients with adolescent idiopathic scoliosis in mainland China? Spine (Phila Pa 1976) 2018 May 1;43(9):E494-E503. [doi: 10.1097/BRS.0000000000002412] [Medline: 28885287]
59. Rivett L, Rothberg A, Stewart A, Berkowitz R. The relationship between quality of life and compliance to a brace protocol in adolescents with idiopathic scoliosis: a comparative study. BMC Musculoskelet Disord 2009 Jan 14;10:5. [doi: 10.1186/1471-2474-10-5] [Medline: 19144157]
60. Vasiliadis E, Grivas TB, Savvidou O, Triantafyllopoulos G. The influence of brace on quality of life of adolescents with idiopathic scoliosis. Stud Health Technol Inform 2006;123:352-356. [Medline: 17108451]
61. Paolucci T, Piccinini G, Iosa M, et al. The importance of trunk perception during brace treatment in moderate juvenile idiopathic scoliosis: what is the impact on self-image? J Back Musculoskelet Rehabil 2017;30(2):203-210. [doi: 10.3233/BMR-160733] [Medline: 27392847]
62. Fang MQ, Huang XL, Wang W, et al. The efficacy of Schroth exercises combined with the Chêneau brace for the treatment of adolescent idiopathic scoliosis: a retrospective controlled study. Disabil Rehabil 2022 Sep;44(18):5060-5068. [doi: 10.1080/09638288.2021.1922521] [Medline: 33984249]
63. Schreiber S, Parent EC, Moez EK, et al. The effect of Schroth exercises added to the standard of care on the quality of life and muscle endurance in adolescents with idiopathic scoliosis-an assessor and statistician blinded randomized controlled trial: "SOSORT 2015 Award Winner". Scoliosis 2015 Sep 18;10(24):24. [doi: 10.1186/s13013-015-0048-5] [Medline: 26413145]
64. Rivett L, Stewart A, Potterton J. The effect of compliance to a Rigo System Cheneau brace and a specific exercise programme on idiopathic scoliosis curvature: a comparative study: SOSORT 2014 award winner. Scoliosis 2014;9:5. [doi: 10.1186/1748-7161-9-5] [Medline: 24926318]
65. Khoshhal Y, Jalali M, Babaee T, Ghandhari H, Gum JL. The effect of bracing on spinopelvic rotation and psychosocial parameters in adolescents with idiopathic scoliosis. Asian Spine J 2019 Dec;13(6):1028-1035. [doi: 10.31616/asj.2018.0307] [Medline: 31434463]
66. Piantoni L, Tello CA, Remondino RG, et al. Quality of life and patient satisfaction in bracing treatment of adolescent idiopathic scoliosis. Scoliosis Spinal Disord 2018 Dec 14;13(26):26. [doi: 10.1186/s13013-018-0172-0] [Medline: 30564635]
67. Vasiliadis E, Grivas TB. Quality of life after conservative treatment of adolescent idiopathic scoliosis. Stud Health Technol Inform 2008;135:409-413. [Medline: 18401108]
68. Villamor GA, Andras LM, Yang J, Skaggs DL. Psychological effects of the SRS-22 on girls with adolescent idiopathic scoliosis. Spine Deform 2018;6(6):699-703. [doi: 10.1016/j.jspd.2018.04.005] [Medline: 30348346]
69. Misterska E, Glowacki M, Adamczyk K, Jankowski R. Patients' and parents' perceptions of appearance in scoliosis treated with a brace: a cross-sectional analysis. J Child Fam Stud 2014;23(7):1163-1171. [doi: 10.1007/s10826-013-9776-4] [Medline: 25210419]
70. Joseph GP, Segovia NA, Wright RC, Mueller C, Tileston KR. Mindset correlates with health-related quality of life assessment in patients with adolescent idiopathic scoliosis. Spine Deform 2021 Mar;9(2):349-354. [doi: 10.1007/s43390-020-00243-w] [Medline: 33201496]
71. Di Maria F, Testa G, Carnazza M, Testaì M, Pavone V. Longer brace duration Is associated with lower stress levels and better quality of life in adolescents with idiopathic scoliosis. Children (Basel) 2023 Jun 28;10(7):1120. [doi: 10.3390/children10071120] [Medline: 37508617]
72. Cantele F, Posanti E, Pittarello C, Masiero S. Is higher compliance to brace therapy associated with poorer quality of life and self-image? A 36-months follow-up study. J Back Musculoskelet Rehabil 2023;36(5):1163-1169. [doi: 10.3233/BMR-220361] [Medline: 37458018]
73. Ohrt-Nissen S, Hallager DW, Gehrchen M, Dahl B. Flexibility predicts curve progression in providence nighttime bracing of patients with adolescent idiopathic scoliosis. Spine (Phila Pa 1976) 2016 Nov 15;41(22):1724-1730. [doi: 10.1097/BRS.0000000000001634] [Medline: 27076435]
74. Misterska E, Glowacki M, Latuszewska J. Female patients' and parents' assessment of deformity- and brace-related stress in the conservative treatment of adolescent idiopathic scoliosis. Spine (Phila Pa 1976) 2012 Jun 15;37(14):1218-1223. [doi: 10.1097/BRS.0b013e31824b66d4] [Medline: 22310093]
75. Ho-Yi Fung O, Yip J, Cheung MC, et al. Exploring mass customization and textile application in medical products: re-designing scoliosis brace for shorter production lead time and better quality of life. Text Res J 2020 Oct;90(19-20):2304-2321. [doi: 10.1177/0040517520916922]
76. Yuan Z, Guo C, Li Y, Guo X, Wang Z, Wang Z. Enhancing idiopathic scoliosis patients' outcomes through functional rehabilitation training in combination with orthosis. Altern Ther Health Med 2024 Jun;30(6):264-269. [Medline: 37944970]
77. Wong MS, Beygi BH, Wong KW, Sin SW, Kwok WK, Wu HD. Effect of different undergarment designs on the compliance and acceptance of the patients with adolescent idiopathic scoliosis under orthotic treatment. Prosthet Orthot Int 2022 Dec 1;46(6):576-581. [doi: 10.1097/PXR.0000000000000161] [Medline: 36515903]

XSL•FO
RenderX

78. Ugwonali OF, Lomas G, Choe JC, et al. Effect of bracing on the quality of life of adolescents with idiopathic scoliosis. Spine J 2004;4(3):254-260. [doi: 10.1016/j.spinee.2003.12.001] [Medline: 15125845]
79. Misterska E, Głowacki M, Ignyś-O'Byrne A, et al. Differences in deformity and bracing-related stress between rural and urban area patients with adolescent idiopathic scoliosis treated with a Cheneau brace. Ann Agric Environ Med 2011;18(2):410-414. [Medline: 22216821]
80. Yagci G, Yakut Y. Core stabilization exercises versus scoliosis-specific exercises in moderate idiopathic scoliosis treatment. Prosthet Orthot Int 2019 Jun;43(3):301-308. [doi: 10.1177/0309364618820144] [Medline: 30628526]
81. Bertuccelli M, Rubega M, Cantele F, et al. Body-related attentional bias in adolescents affected by idiopathic scoliosis. Eur J Investig Health Psychol Educ 2023 Sep 18;13(9):1909-1919. [doi: 10.3390/ejihpe13090138] [Medline: 37754477]
82. Schwieger T, Campo S, Weinstein SL, Dolan LA, Ashida S, Steuber KR. Body image and quality-of-life in untreated versus brace-treated females with adolescent idiopathic scoliosis. Spine (Phila Pa 1976) 2016 Feb;41(4):311-319. [doi: 10.1097/BRS.0000000000001210] [Medline: 26555827]
83. Théroux J, Le May S, Hebert JJ, Labelle H. Back pain prevalence is associated with curve-type and severity in adolescents with idiopathic scoliosis: a cross-sectional study. Spine (Phila Pa 1976) 2017 Aug 1;42(15):E914-E919. [doi: 10.1097/BRS.0000000000001986] [Medline: 27870807]
84. Fong DYT, Cheung KMC, Wong YW, et al. An alternative to a randomised control design for assessing the efficacy and effectiveness of bracing in adolescent idiopathic scoliosis. Bone Joint J 2015 Jul;97-B(7):973-981. [doi: 10.1302/0301-620X.97B7.35147] [Medline: 26130355]
85. Loukos I, Zachariou C, Nicolopoulos C, Korres D, Efstathopoulos N. Analysis of the corrective forces exerted by a dynamic derotation brace (DDB). Prosthet Orthot Int 2011 Dec;35(4):365-372. [doi: 10.1177/0309364611420477] [Medline: 21960051]
86. Lou E, Hill DL, Raso JV. A wireless sensor network system to determine biomechanics of spinal braces during daily living. Med Biol Eng Comput 2010 Mar;48(3):235-243. [doi: 10.1007/s11517-010-0575-4] [Medline: 20094808]
87. Sattout A, Clin J, Cobetto N, Labelle H, Aubin CE. Biomechanical assessment of providence nighttime brace for the treatment of adolescent idiopathic scoliosis. Spine Deform 2016 Jul;4(4):253-260. [doi: 10.1016/j.jspd.2015.12.004] [Medline: 27927513]
88. Wong MS, Evans JH. Biomechanical evaluation of the Milwaukee brace. Prosthet Orthot Int 1998 Apr;22(1):54-67. [doi: 10.3109/03093649809164457] [Medline: 9604276]
89. Périé D, Aubin CE, Petit Y, Beauséjour M, Dansereau J, Labelle H. Boston brace correction in idiopathic scoliosis: a biomechanical study. Spine (Phila Pa 1976) 2003 Aug 1;28(15):1672-1677. [doi: 10.1097/01.BRS.0000083165.93936.6D] [Medline: 12897490]
90. Antoine L, Nathan D, Laure M, Briac C, Jean-François M, Corinne B. Compliance with night-time overcorrection bracing in adolescent idiopathic scoliosis: result from a cohort follow-up. Med Eng Phys 2020 Mar;77:137-141. [doi: 10.1016/j.medengphy.2020.01.003] [Medline: 31992499]
91. Lou E, Raso JV, Hill DL, Mahood JK, Moreau MJ. Correlation between quantity and quality of orthosis wear and treatment outcomes in adolescent idiopathic scoliosis. Prosthet Orthot Int 2004 Apr;28(1):49-54. [doi: 10.3109/03093640409167925] [Medline: 15171578]
92. Wong MS. Early intervention in non-operative management of adolescent idiopathic scoliosis (Thesis). : Hong Kong Polytechnic University; 2000 URL: https://www.proquest.com/openview/0276fb4055e1500da85e244b3a9344dc/1?pq-origsite=gscholar&cbl=18750&diss=y [accessed 2026-07-20]
93. Wong MS, Mak AF, Luk KD, Evans JH, Brown B. Effectiveness and biomechanics of spinal orthoses in the treatment of adolescent idiopathic scoliosis (AIS). Prosthet Orthot Int 2000 Aug;24(2):148-162. [doi: 10.1080/03093640008726538] [Medline: 11061202]
94. Karol LA, Virostek D, Felton K, Wheeler L. Effect of compliance counseling on brace use and success in patients with adolescent idiopathic scoliosis. J Bone Joint Surg Am 2016 Jan 6;98(1):9-14. [doi: 10.2106/JBJS.O.00359] [Medline: 26738898]
95. Odermatt D, Mathieu PA, Beauséjour M, Labelle H, Aubin CE. Electromyography of scoliotic patients treated with a brace. J Orthop Res 2003 Sep;21(5):931-936. [doi: 10.1016/S0736-0266(03)00038-X] [Medline: 12919883]
96. Rahman T, Sample W, Yorgova P, et al. Electronic monitoring of orthopedic brace compliance. J Child Orthop 2015 Oct;9(5):365-369. [doi: 10.1007/s11832-015-0679-3] [Medline: 26310101]
97. Romano M, Carabalona R, Petrilli S, Sibilla P, Negrini S. Forces exerted during exercises by patients with adolescent idiopathic scoliosis wearing fiberglass braces. Scoliosis 2006 Jul 21;1:12. [doi: 10.1186/1748-7161-1-12] [Medline: 16859544]
98. Lou EHM, Hill DL, Raso JV, Moreau M, Hedden D. How quantity and quality of brace wear affect the brace treatment outcomes for AIS. Eur Spine J 2016 Feb;25(2):495-499. [doi: 10.1007/s00586-015-4233-2] [Medline: 26386869]
99. Load compliance monitor system for the treatment of scoliosis. Presented at: Canadian Conference on Electrical and Computer Engineering; May 9-12, 1999 URL: https://www.scopus.com/pages/publications/0033333669 [accessed 2026-07-20]

XSL•FO
RenderX

100. Chan WY, Yip J, Yick KL, et al. Mechanical and clinical evaluation of a shape memory alloy and conventional struts in a flexible scoliotic brace. Ann Biomed Eng 2018 Aug;46(8):1194-1205. [doi: 10.1007/s10439-018-2016-8] [Medline: 29691786]
101. Hasler CC, Wietlisbach S, Büchler P. Objective compliance of adolescent girls with idiopathic scoliosis in a dynamic SpineCor brace. J Child Orthop 2010 Jun;4(3):211-218. [doi: 10.1007/s11832-010-0249-7] [Medline: 21629374]
102. Zapata KA, Virostek D, Ma Y, et al. Outcomes for nighttime bracing in adolescent idiopathic scoliosis based on brace wear adherence. Spine Deform 2024 May;12(3):643-650. [doi: 10.1007/s43390-024-00835-w] [Medline: 38457029]
103. Beauséjour M, Petit Y, Grimard G, Aubin CE, Dansereau J, Labelle H. Relationships between strap tension, interface pressures and spine correction in brace treatment of scoliosis. Stud Health Technol Inform 2002;88:207-211. [doi: 10.3233/978-1-60750-932-5-207] [Medline: 15456033]
104. Pham VM, Houilliez A, Schill A, Carpentier A, Herbaux B, Thevenon A. Study of the pressures applied by a Chêneau brace for correction of adolescent idiopathic scoliosis. Prosthet Orthot Int 2008 Sep;32(3):345-355. [doi: 10.1080/03093640802016092] [Medline: 18609033]
105. Rahman T, Bowen JR, Takemitsu M, Scott C. The association between brace compliance and outcome for patients with idiopathic scoliosis. J Pediatr Orthop 2005;25(4):420-422. [doi: 10.1097/01.bpo.0000161097.61586.bb] [Medline: 15958887]
106. Chase AP, Bader DL, Houghton GR. The biomechanical effectiveness of the Boston brace in the management of adolescent idiopathic scoliosis. Spine (Phila Pa 1976) 1989 Jun;14(6):636-642. [doi: 10.1097/00007632-198906000-00018] [Medline: 2749381]
107. Mak I, Lou E, Raso JV, et al. The effect of time on qualitative compliance in brace treatment for AIS. Prosthet Orthot Int 2008 Jun;32(2):136-144. [doi: 10.1080/03093640802015839] [Medline: 18569881]
108. Karol LA, Wingfield JJ, Virostek D, Felton K, Jo C. The influence of body habitus on documented brace wear and progression in adolescents with idiopathic scoliosis. J Pediatr Orthop 2020 Mar;40(3):e171-e175. [doi: 10.1097/BPO.0000000000001420] [Medline: 31259783]
109. Benish BM, Smith KJ, Schwartz MH. Validation of a miniature thermochron for monitoring thoracolumbosacral orthosis wear time. Spine (Phila Pa 1976) 2012 Feb 15;37(4):309-315. [doi: 10.1097/BRS.0b013e31821e1488] [Medline: 21540779]
110. Walmsley RP, Galpin P, Lowther K, Ashworth MA. Effect of the application of the Boston brace system on the electrical activity of the paravertebral muscles in adolescents with idiopathic scoliosis: preliminary report. Physiother Can 1982;34.
111. Jamison M, Glover M, Peterson K, et al. Lumbopelvic postural differences in adolescent idiopathic scoliosis: a pilot study. Gait Posture 2022 Mar;93:73-77. [doi: 10.1016/j.gaitpost.2022.01.002] [Medline: 35093665]
112. Gaume M, Pietton R, Vialle R, Chaves C, Langlais T. Is daily walking distance affected in adolescent idiopathic scoliosis? An original prospective study using the pedometer on smartphones. Arch Pediatr 2020 Aug;27(6):333-337. [doi: 10.1016/j.arcped.2020.04.002] [Medline: 32563619]
113. Asada T, Kotani T, Sunami T, et al. What factor induces stress in patients with AIS under brace treatment? Analysis of a specific factor using exploratory factor analysis. J Orthop Sci 2021 Nov;26(6):999-1003. [doi: 10.1016/j.jos.2020.10.024] [Medline: 33309133]
114. Sapountzi-Krepia D, Psychogiou M, Peterson D, et al. The experience of brace treatment in children/adolescents with scoliosis. Scoliosis 2006 May 22;1(1):8. [doi: 10.1186/1748-7161-1-8] [Medline: 16759368]
115. Wong MS, Cheng JCY, Lam TP, et al. The effect of rigid versus flexible spinal orthosis on the clinical efficacy and acceptance of the patients with adolescent idiopathic scoliosis. Spine (Phila Pa 1976) 2008 May 20;33(12):1360-1365. [doi: 10.1097/BRS.0b013e31817329d9] [Medline: 18496349]
116. Brigham EM, Armstrong DG. Motivations for compliance with bracing in adolescent idiopathic scoliosis. Spine Deform 2017 Jan;5(1):46-51. [doi: 10.1016/j.jspd.2016.09.004] [Medline: 28038693]
117. Gratz RR, Papalia-Finlay D. Psychosocial adaptation to wearing the Milwaukee brace for scoliosis. a pilot study of adolescent females and their mothers. J Adolesc Health Care 1984 Oct;5(4):237-242. [doi: 10.1016/s0197-0070(84)80124-2] [Medline: 6490478]
118. Catanzariti JF, Darchicourt D, Chevutschi A, Le Berre M, Tanche L, Guyot MA. Is self-image, in reference to the gravitational vertical, altered in adolescent idiopathic scoliosis? A multicenter, single-blind, case-control study. J Pediatr Rehabil Med 2022;15(3):477-486. [doi: 10.3233/PRM-200689] [Medline: 36031914]
119. Vandal S, Rivard CH, Bradet R. Measuring the compliance behavior of adolescents wearing orthopedic braces. Issues Compr Pediatr Nurs 1999;22(2-3):59-73. [doi: 10.1080/014608699265293] [Medline: 10786513]
120. Chan SL, Cheung KM, Luk KD, Wong KW, Wong MS. A correlation study between in-brace correction, compliance to spinal orthosis and health-related quality of life of patients with adolescent idiopathic scoliosis. Scoliosis 2014 Feb 22;9(1):1. [doi: 10.1186/1748-7161-9-1] [Medline: 24559234]
121. Lin Y, Lou E, Lam TP, et al. The intelligent automated pressure-adjustable orthosis for patients with adolescent idiopathic scoliosis: a bi-center randomized controlled trial. Spine (Phila Pa 1976) 2020 Oct 15;45(20):1395-1402. [doi: 10.1097/BRS.0000000000003559] [Medline: 32453223]
122. Asada T, Kotani T, Sakuma T, et al. Impact of brace-related stress on brace compliance in adolescent idiopathic scoliosis: a single-center comparative study using objective compliance measurement and brace-related stress. Spine Surg Relat Res 2023 Mar 13;7(4):377-384. [doi: 10.22603/ssrr.2022-0246] [Medline: 37636154]

XSL•FO
RenderX

123. Asada T, Kotani T, Sakuma T, et al. Factors influencing optimal bracing compliance in adolescent idiopathic scoliosis: a single-center prospective cohort study. Spine (Phila Pa 1976) 2024 Dec 15;49(24):1708-1715. [doi: 10.1097/BRS.0000000000004997] [Medline: 38597189]
124. Li X, Lau ENS, Chan SKC, et al. Effects of mindfulness-based intervention to improve bracing compliance in adolescent idiopathic scoliosis patients: a randomized controlled trial. Mindfulness 2023 Jan 16;14(2):322-334. [doi: 10.1007/s12671-022-02021-3]
125. Lin Y, Cheung JPY, Chan CK, et al. A randomized controlled trial to evaluate the clinical effectiveness of 3D-printed orthosis in the management of adolescent idiopathic scoliosis. Spine (Phila Pa 1976) 2022 Jan 1;47(1):13-20. [doi: 10.1097/BRS.0000000000004202] [Medline: 34392277]
126. Wang H, Meng X, Tetteroo D, et al. Exploration of contributory factors to an unpleasant bracing experience of adolescent idiopathic scoliosis patients a quantitative and qualitative research. Children (Basel) 2022 Apr 28;9(5):635. [doi: 10.3390/children9050635] [Medline: 35626812]
127. Gornitzky AL, England P, Kiani SN, Yellin JL, Flynn JM. Why don't adolescents wear their brace? A prospective study investigating psychosocial characteristics that predict scoliosis brace wear. J Pediatr Orthop 2023 Jan 1;43(1):51-60. [doi: 10.1097/BPO.0000000000002272] [Medline: 36194756]
128. Wong MS, Cheng CY, Ng BKW, et al. The effect of rigid versus flexible spinal orthosis on the gait pattern of patients with adolescent idiopathic scoliosis. Gait Posture 2008 Feb;27(2):189-195. [doi: 10.1016/j.gaitpost.2007.03.007] [Medline: 17466520]
129. Lou E, Hill DL, Raso VJ, Durdle NG. Continuous assessment of posture in adolescent idiopathic scoliosis. Can Conf Electr Comput Eng 1999;3:1511-1511. [doi: 10.1109/CCECE.1999.804935]
130. Sánchez-Raya J, Salat-Batlle J, Castilla D, Zaragozá I, García-Palacios A, Suso-Ribera C. Utilizing mobile health technology to enhance brace compliance: feasibility and effectiveness of an app-based monitoring system for adolescents with idiopathic scoliosis. J Pers Med 2025 Sep 1;15(9):405. [doi: 10.3390/jpm15090405] [Medline: 41003108]
131. Caronni A, Sciumè L, Donzelli S, Zaina F, Negrini S. ISYQOL: a Rasch-consistent questionnaire for measuring health-related quality of life in adolescents with spinal deformities. Spine J 2017 Sep;17(9):1364-1372. [doi: 10.1016/j.spinee.2017.05.022] [Medline: 28529002]
132. Voinea GD, Postelnicu C, Butnariu S. Challenges involved in the design of an e-health application for a wearable scoliosis monitoring system. In: Stephanidis C, editor. HCI International 2016 – Posters' Extended Abstracts (HCI 2016): Springer, Cham; 2016:339-344. [doi: 10.1007/978-3-319-40542-1_56]
133. Lee KP, Wang Z, Zheng L, et al. Enhancing orthotic treatment for scoliosis: development of body pressure mapping knitwear with integrated FBG sensors. Sensors (Basel) 2025 Feb 20;25(5):1284. [doi: 10.3390/s25051284] [Medline: 40096079]
134. Borboni A, Zanoletti F, Faglia R. Development of an ultra-thin piezoresistive velostat-based wearable sensor for real-time pressure monitoring in adolescent idiopathic scoliosis bracing. In: Carbone G, Quaglia G, editors. Proceedings of I4SDG Workshop 2025 - IFToMM for Sustainable Development Goals: Springer, Cham; 2025:293-300. [doi: 10.1007/978-3-031-91151-4_33]
135. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34. [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]

Abbreviations

BAS: Body Appreciation Scale
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
EMG: electromyography
HRQoL: health-related quality of life
PICOS: Population, Intervention, Comparison, Outcome, and Study design
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-analyses Extension for Scoping Reviews
PROM: patient-reported outcome measure
RQ: research question
SOSORT: International Society on Scoliosis Orthopedic and Rehabilitation Treatment

XSL•FO
RenderX

Edited by L Buis; submitted 14.Dec.2025; peer-reviewed by C Cordani, T Babaee, YY Zou; revised version received 11.Feb.2026; accepted 09.Jun.2026; published 10.Aug.2026.

Please cite as:
Mathew A, Sykorova K, Pavel N, Johnsen MB, Pikkarainen M, Ali S, Kvammen MF, Gazerani P
Patient-Reported and Technology-Assisted Monitoring in Orthotic Management of Adolescents With Idiopathic Scoliosis: Scoping Review
JMIR Mhealth Uhealth 2026;14:e88880
URL: https://mhealth.jmir.org/2026/1/e88880
doi:10.2196/88880

© Anna Mathew, Kamila Sykorova, Nenad Pavel, Marianne Bakke Johnsen, Minna Pikkarainen, Sadeeq Ali, Mette Fløystad Kvammen, Parisa Gazerani. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Barriers and Enablers to Integrating Patient-Generated Health Data in Shared Decision-Making From Health Care Professional and Patient Perspectives: Scoping Review

Pavithren V S Pakianathan[1,2], BEng, MEng; Devender Kumar[3], PhD; Jayathissa Prabath[2,4], MBBS, MD, MSc; Rada Hussein[2], PhD; Josef Niebauer[2,5], Prof Dr, MBA, MD, PhD; Albrecht Schmidt[1], PhD; Jan Smeddinck[2], PhD

[1]Human-Centered Ubiquitous Media Group, Faculty of Mathematics, Informatics and Statistics, LMU Munich, Munich, Germany

[2]Ludwig Boltzmann Institute for Digital Health and Prevention, 22 Lindhofstrasse, Salzburg, Austria

[3]The Maersk Mc-Kinney Moller Institute, The Faculty of Engineering, University of Southern Denmark, Odense, Denmark

[4]University of Applied Sciences Technikum Wien, Vienna, Austria

[5]University Institute of Sports Medicine, Prevention and Rehabilitation, Paracelsus Medical University, Salzburg, Salzburg, Austria

Corresponding Author:
Pavithren V S Pakianathan, BEng, MEng
Human-Centered Ubiquitous Media Group, Faculty of Mathematics, Informatics and Statistics, LMU Munich, Munich, Germany

Abstract

Background: Advances in sensor technologies and increased adoption of wearables and smartphones by individuals have led to an abundance of patient-generated health data (PGHD). This data, when used effectively, could help to further augment the process of shared decision-making (SDM) to enable patient-centered care. However, the possible integration and usage of PGHD introduces complexities and challenges, which warrant considering both health care professional (HCP) and patient perspectives.

Objective: Summarize the relevant works from the past 10 years that reflect the perspectives of both HCPs and patients as key stakeholders on potential barriers and enablers to the integration of PGHD for SDM. We analyzed both perspectives to surface key challenges and opportunities with integrating PGHD throughout patient journeys, as well as clinical workflows.

Methods: Electronic searches were done using 3 databases: PubMed, ACM Digital Library, and IEEE Xplore for papers published between March 2013 and March 2023. Enablers and barriers mentioned by the stakeholders in included papers were extracted and analyzed using thematic analysis. An existing six-stage workflow model for integrating PGHD was used as a reference for deductive coding. Subsequently, considering barriers and enablers faced by both the HCPs and patients uncovered various tensions and alignments of perspectives, which could be addressed in future work and can inform concepts, designs, and development in PGHD for SDM.

Results: A total of 53 publications were included in the scoping review. Six main overarching themes for barriers and enablers were identified: (1) patient-provider relationship, (2) patient characteristics, (3) organizational factors, (4) medical ethics and law, (5) data-driven workflow, and (6) design and technology. The 6-stage workflow was further expanded based on the new findings to include 4 additional stages, which include contextual considerations outside of traditional clinical environments. In addition to partially corroborating previously established barriers in the 6-stage workflow model, several new barriers and enablers were identified throughout all stages. This model helps to further align the needs of HCPs and patients beyond the clinical setting and could benefit system designers who plan to integrate PGHD for SDM.

Conclusions: This scoping review demonstrates that there are several factors to consider for effectively integrating PGHD into health-related SDM. Notably, such factors extend outside the boundaries of traditional clinical settings. Although there is agreement between HCPs and patients on certain factors, there are also tensions to be addressed. Our augmented 10-stage workflow model offers system designers an overview of the challenges and enablers to consider while designing for PGHD integration in clinical workflows and patient journeys to improve SDM.

(JMIR Mhealth Uhealth 2026;14:e85197) doi:10.2196/85197

KEYWORDS

patient-generated health data; PGHD; scoping review; data integration; shared decision-making; digital health; digital biomarkers; personal informatics; technologies; digital health interventions; mHealth

XSL•FO
RenderX

Introduction

Overview

The challenges of aging populations and noncommunicable diseases across society have led to increased demands and stress on health care systems. Digital health technologies are seen as an enabler to solving such challenges. A review of telemedicine for chronic diseases revealed that wearable devices can help improve the management of chronic conditions by allowing for continuous monitoring of vital signs and reducing patient loads from hospitals [1], thereby improving the delivery of health care services. However, digital health technologies and consumer health care technologies are still facing hurdles in entering the integrated clinical market due to several factors such as infrastructure, technical training, and legal and ethics-related barriers [2]. Digital health technologies pave the path toward more personalized health care, and a key enabler to this end is the use of patient-generated health data (PGHD).

The rise of the Quantified Self movement led by people who track many kinds of data about themselves [3,4], followed by the mass market adoption of self-tracking technologies, coupled with advancements in mobile sensing technologies, has led to an increase in patients bringing their health data to clinical consultations via patient-initiated or clinician-initiated tracking [5,6]. PGHD refers to data created, recorded, or gathered by patients (or by family members or other caregivers) to help address one's health concerns [7]. Notably, PGHD excludes data from electronic health records (EHRs), records entered by health care professionals (HCPs), and patient-reported outcomes. The integration of PGHD has been known to bring about benefits such as (1) *better contextualization of symptoms* [8], (2) *better clinical decision-making* [9], and (3) *personalized and participatory care* [10,11].

Although PGHD has the potential to play a key role in the digitalization of health care, several challenges [12] lie in the path toward the adoption of PGHD as a foundation for enabling novel and additional insights and applications in clinical practice. Unlike traditional clinical data, which is typically produced through the actions of HCPs in tightly controlled clinical conditions, PGHD are usually gathered throughout situated daily living under strongly varying and individual circumstances. This poses challenges for the interpretation and reliability of PGHD with varying quality [13], but also opens opportunities in the form of "objective data windows" into daily living that would otherwise not be available. Akin to medical and health information being available through internet search [14], PGHD integration in patient journeys and clinical workflows is likely to impact medical examination and shared decision-making (SDM)—the collaborative process between an HCP and a patient to make health-related decisions [15,16] serving as a cornerstone of person-centered care.

We posit that the integration of PGHD in the health care context has resulted in a complex sociotechnical system where PGHD could traverse throughout the health data lifecycle [17] and be interacted with by multiple actors—humans and machines—resulting in the emerging field of human-data interaction [18]. Given the vital role PGHD plays in improving SDM [19] and empowering patients in proactively managing their health, it is essential to understand the needs and perspectives of patients and HCPs on how it can be integrated into care pathways.

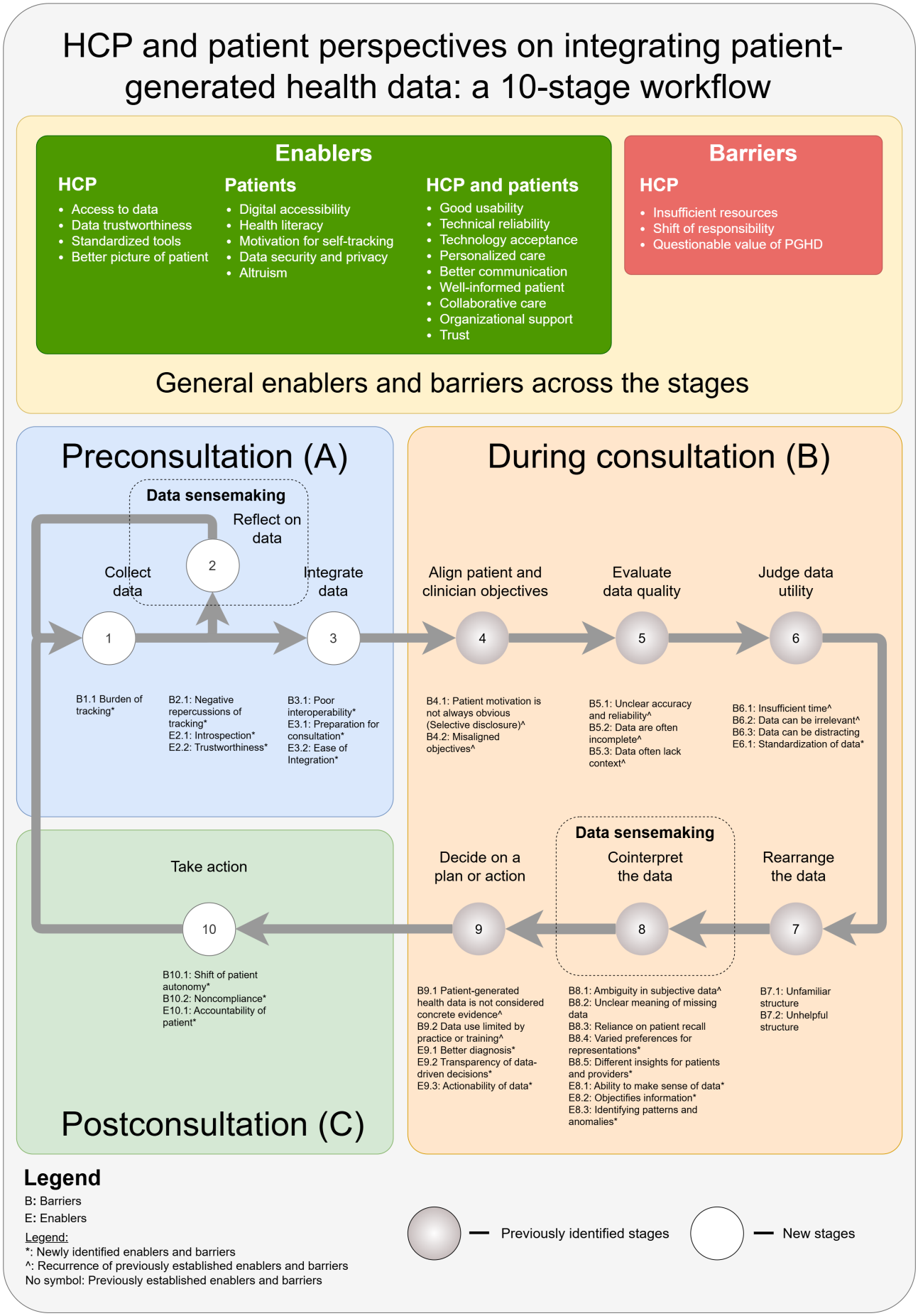
Over the past years, prior works have investigated PGHD from diverse perspectives—how individuals track data [20], how HCPs (can better) integrate data in clinical workflows [12], and how patients and HCPs collaborate with data [21]. However, there is a gap in holistically evaluating the needs and values of the key stakeholders—patients and HCPs—along patient journeys and HCPs' clinical workflows. Using an ecological lens to surface system-level interactions and reduce misunderstandings between stakeholders [6,21], our scoping review aims to map how patients and HCPs perceive the integration of PGHD within SDM processes across patient journeys and clinical workflows.

The main contributions of this scoping review paper are as follows: (1) a 10-stage workflow for aligning HCP and patient perspectives on the common barriers and enablers while integrating PGHD across patient journeys and clinical workflows, (2) mitigation strategies to overcome barriers across the patient journeys and clinical workflows, and (3) aligning areas of tension between HCP and patients across the data-enabled clinical workflows and patient data journeys and corresponding alignment strategies.

We use the findings of this scoping review to formulate a model—cf. 10-stage workflow model in Figure 1—which designers and health care technologists can use to analyze or pave data-enabled patient journeys and clinical workflows. By creating better data-enabled patient journeys, relationships between patients and HCPs can be improved, whereby patients could easily share information with HCPs and deliberate their preferences while participating in SDM [16].

XSL•FO
RenderX

Figure 1. The adapted 10-stage workflow for integrating PGHD. HCP: health care professional; PGHD: patient-generated health data.



XSL•FO
RenderX

Background

Overview

We drew on prior work on PGHD and patient-HCP collaboration with PGHD. This sets the context on how technology has influenced the relationship between patients and HCPs, resulting in both alignments and tensions on integrating PGHD across patient data journeys and clinical workflows.

PGHD

The Quantified Self movement, powered by the rise of technological advancements and increased accessibility to wearables, has led to increased patients bringing opportunistically captured PGHD to consultations [5,22]. *Quantifiers*—those who are part of the Quantified Self—are motivated to understand more about themselves through numbers and modify their lives accordingly [23]. They engage in such practices by using tracking technologies to self-collect personal data, including physiological, behavioral, or environmental information [24,25]. On the other hand, patients might engage in clinician-initiated tracking, which is purposefully capturing PGHD while receiving treatment or instructed by HCPs for long-term chronic care.

PGHD has also been enabled by remote monitoring technologies, which allow HCPs to issue technologies such as wearables and smartphone apps to patients who can take part in long-term treatment (eg, for postoperative care). The availability of PGHD during in-person consultations could facilitate more objectively grounded discussions between HCPs and patients. It could help to give HCPs a clearer picture of a patient's condition, improving the SDM process [7,19]. HCPs also believe that the integration of PGHD could improve symptom monitoring and assessment between visits, help personalize interventions and monitoring plans, aid with assessing clinical outcomes, promote patient self-management and behavior change support and prevention, as well as improve care delivery and quality assurance [22].

Motivated by the vision that PGHD are believed to help HCP "fill in the gaps" [26] and to unlock their potential, national frameworks have also been developed which guide their integration and governance. For example, the European Health Data Space [27], Digital Health Applications [28], and the UK's Personalized Health and Care 2020 framework [29] are frameworks or legislation that support the use, governance, interoperability, and integration of PGHD, fostering the uptake of digital health technologies by patients and HCPs alike. However, there are several technical, social, and organizational challenges to integrating PGHD in clinical practice [7,12].

From an HCP's perspective, PGHD can be challenging to use unless optimally prepared due to time constraints during consultations [12,30,31], the presence of irrelevant information or data, unfamiliarity with the structure of the collected data [32], and poor integration of PGHD into EHR systems [9], clinical workflows, and procedures. Furthermore, overheads due to administrative tasks related to data processing have been shown to create unnecessary workload and even burnout among health care providers [33].

From a patient perspective, there have been usability issues with the technologies, a shift in patient autonomy [31], increased occurrences of over-engagement with data which could in turn lead to a broad range of issues, including anxiety or depression and increased burden of tracking [9]. Moreover, patient expectations of HCPs to review their PGHD in its entirety during a time-limited consultation [34], emotional attachment and subjective view of their PGHD, and selective disclosure of PGHD could prevent doctors from having an unbiased view of patients' data [21].

Figueiredo and Chen [35] systematically explored the various dimensions of PGHD (eg, type, intended use, and duration) and highlighted the critical need to consider sociotechnical factors for its successful integration. This aligns with the broader shifts in health care, including the rise of information infrastructures in health care [36] and increased *data work* required of both patients and HCPs. The HCI community has also embarked on efforts to have a deeper understanding of this phenomenon [37,38]. PGHD integration should seek to balance the needs and interests of both patients and HCPs, as PGHD "negotiation" [21] between them may increase tensions and "asymmetries" [6] between them. Such tensions underscore the importance of designing systems that appropriately mediate the interaction between patients and HCPs, thereby enabling effective patient-centered care and SDM.

Designing for PGHD Integration and Patient-HCP Collaboration

Past works show that the generation of PGHD is not performed at a single point in time. Two seminal works evaluate this from a systematic perspective, though not specifically from the patient's viewpoint. Li et al [20] derived a stage-based model of personal informatics systems with five stages, namely (1) data preparation before tracking, (2) collection of information about oneself, (3) integration of information for sensemaking, (4) reflection on their personal information, and (5) taking action based on the information reflected upon. Primarily targeted toward the "quantifiers," it emphasizes the need for personal informatics systems to be designed in a "holistic manner across the stages." Epstein et al [39] subsequently adopted a *lived informatics* [40] perspective to characterize the integration of self-tracking into a quantifier's daily life. While these models primarily focus on personal use by "quantifiers," their stages and concepts can also be applied to patients who generate and share their data during clinical consultations. To systematically address the barriers HCPs face when integrating PGHD, West et al [12] developed a 6-stage workflow for integrating PGHD in clinical settings. The six stages are, namely, (1) aligning patient and HCP objectives, (2) evaluating data quality, (3) judging data quality, (4) rearranging the data, (5) interpreting the data, and (6) deciding on a plan or action. Overall, there needs to be alignment between HCPs and patients to pave the path toward patient-centered care.

Andersen et al [41] describe how patients and HCPs might have varying concerns in a clinical setting and emphasize that it is a challenge to align their concerns. They explain that for HCPs, concerns are about "professional issues to do with diagnosing and curing disease in accordance with accepted medical

XSL•FO
RenderX

standards." Whereas for a patient, concerns typically are about "personal experience and quality of life issues." They introduce a set of concepts for analyzing concerns that arise during the design of e-Health systems, those being (1) meaningful and making sense to both HCPs and patients, (2) actionable, wherein either HCPs or patients can act on it, and (3) feasible, wherein it is easy and convenient to do so within the organizational and social context [41]. Adopting the concepts by Andersen et al [41] for designing e-Health systems, we posit that the PGHD collected by patients must be (1) meaningful through structured scaffolding of how and what data is collected and presented; (2) actionable, where both HCPs and patients could craft and adopt appropriate (and possibly personalized) treatment plans; and (3) feasible where in-depth user-centered design methods could be used to understand how different hospital systems work, HCPs' practices, and the variances in sociocultural practices around patients' collection and sharing of PGHD among different geographic locations. To effectively integrate PGHD into patient data journeys, it is essential to adopt a systemic lens to this challenge and to address the needs of the adopters [42].

Research Gap and Motivation

Over the years, several publications have investigated the integration of PGHD into clinical practice from the patients' or the HCPs' perspectives [7,21,31,43-48]. To date, a summary of the enablers and barriers to using PGHD from both perspectives has not been published yet. In a study on patient-provider collaboration in the context of irritable bowel syndrome and weight management, Chung et al [21] emphasize that collaboration occurs in every stage of self-tracking and suggest that due to the social nature of patient-provider collaboration with data, it is necessary to investigate how HCPs and patients' roles shift across the stages. We aim to address this gap and align patient and HCP "concerns" [41] for better integration of PGHD in patient data journeys and clinical workflows.

Methods

Study Design

This scoping review is reported following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [47]. The scoping review approach was chosen over a systematic review process [49] to determine coverage of literature on the topic of PGHD integration in clinical pathways from the perspectives of both HCPs and patients, with an aim of aligning their needs. Prior reviews [12,25,50,51] around PGHD have focused primarily on a single stakeholder—HCPs. A formal review protocol was not registered or made publicly available.

We used the methodology formulated by Peters et al [52] to investigate the following research question: What barriers and enablers exist among key stakeholders—HCPs and patients—to integrate PGHD to inform SDM?

Search Strategy

Overview

After reviewing prior works in this area and having a consultation among the coauthors, we narrowed down to ACM Digital Library, PubMed, and IEEE databases, which cover research in both technical and medical fields. The search was conducted in PubMed, ACM Digital Library, and IEEE Xplore, chosen to capture literature spanning clinical, technical, and human-computer interaction domains. The search was conducted in March 2023 for papers published between March 2013 and March 2023. We chose 2013 as the starting year as it was a crucial point in the evolution of the field of digital health technologies driven by the democratization of smartphones and wearables [53]. Furthermore, in 2013, acknowledging this phenomenon, the FDA released its first clear framework on how digital health apps and wearables integrations would be regulated for personal and clinical use [54].

Several keywords were initially selected after considering highly visible related works in the area. Thereafter, the first and second authors discussed common themes to narrow down to the keywords in a scheme that can be summarized as follows:

1. patient-generated data *OR* patient-generated health data *OR* [...]
   AND
2. patient-reported outcomes *OR* self-tracking *OR* digital health
   AND
3. shared decision-making *OR* [...]
   AND
4. enablers *OR* barriers OR perspectives OR [...]

As the terminologies patient-generated data and PGHD have been used in prior work, both keywords were included in the search query. Other related keywords indicated as "[...] " and Boolean operators were included in the entire search query, which can be found in Multimedia Appendix 1.

Inclusion and Exclusion Criteria

The aim of applying inclusion criteria (ICs) and exclusion criteria (ECs) is to extract only publications relevant to the objective of this scoping review. We used the following set of ICs: (1) papers published in English in conference proceedings or a peer-reviewed journal, (2) papers related to noninvasive wearables that generated PGHD (eg, smartwatch, blood pressure monitor, etc) and patient self-reported data via journaling, etc, (3) studies which involved either HCPs or patients or both, (4) studies that explicitly mention SDM in person or remotely, mediated via technology, or have elements of SDM, and (5) the type of self-tracking could either be patient- or participant-initiated or clinician-initiated.

We used the following set of ECs: (1) review papers and study protocols, (2) papers referring solely to the personal health record of the patient, (3) papers with implants or wearables that require invasive surgery, and (4) gray literature and policy documents.

Further details about the IC and EC can be found in Multimedia Appendix 2.

XSL•FO
RenderX

Screening Process

To streamline the paper selection and review process, the results were exported to an AI-assisted collaborative review tool, Rayyan, which aids in duplicate detection and scoring suggestions. The AI system learns the patterns in the decision criteria and gives scoring suggestions after the author makes at least 50 inclusion or exclusion decisions. The first and second authors independently performed abstract screening and review of papers, and conflicts were resolved through discussions between the first and the second author after considering the research protocol and research question. The first author subsequently extracted the data from the final selection of papers. The detailed procedure is reported in the Results section.

Collating and Summarizing the Data

The extracted data were analyzed using a hybrid deductive-inductive thematic analysis guided by the 6-phase framework by Braun and Clarke [32] and the integrated approach to hybrid coding outline by Fereday and Muir-Cochrane [55] by a single researcher (see positionality statement in next subsection). The 6-stage workflow model initially proposed by West et al [12] was used as an a priori deductive framework. Initial data handling and preliminary inductive coding were performed by the first author using Excel (Microsoft Corp). To facilitate systematic analysis, the data and initial codes were then imported into MAXQDA (version 22). The first author conducted another round of familiarization to immerse in the data to refine the codes after consultation with the second author. Concurrently, open inductive coding was applied to capture barriers and enablers that fell outside the predefined framework. To ensure conceptual alignment, after reviewing the first 10 papers, the first author discussed code discrepancies with the second and last author. After coding for all the papers, the first, second, and last authors discussed the final codes and discussed the emerging themes. Clinician was used as shorthand for a wider understanding of HCP in this step. The detailed subthemes and corresponding categorization can be found in the codebook in Multimedia Appendix 3. We did not perform a formal critical appraisal of the papers, as the aim of this scoping review was to map the breadth of barriers and enablers reported across diverse study designs rather than to estimate any effect sizes.

Positionality

The lead researcher is a PhD candidate with a background in human-computer interaction and digital accessibility and has industry experience in digital transformation and building systems for Digital Health Interventions. He comes from a highly digitized nation in Southeast Asia with a robust health care system leading in digital health. His PhD research agenda focuses on PGHD integration in cardiac prevention pathways, and as such, he has vested interests in understanding the sociotechnical factors influencing its integration. To manage potential biases in single-coder analysis, he kept notes of his coding process and discussed emerging codes and themes with the second and last author. The authors have a background in the field of human-computer interaction, digital health, health informatics, cardiology, and digital health interventions for physical activity behavior change and human-computer interaction.

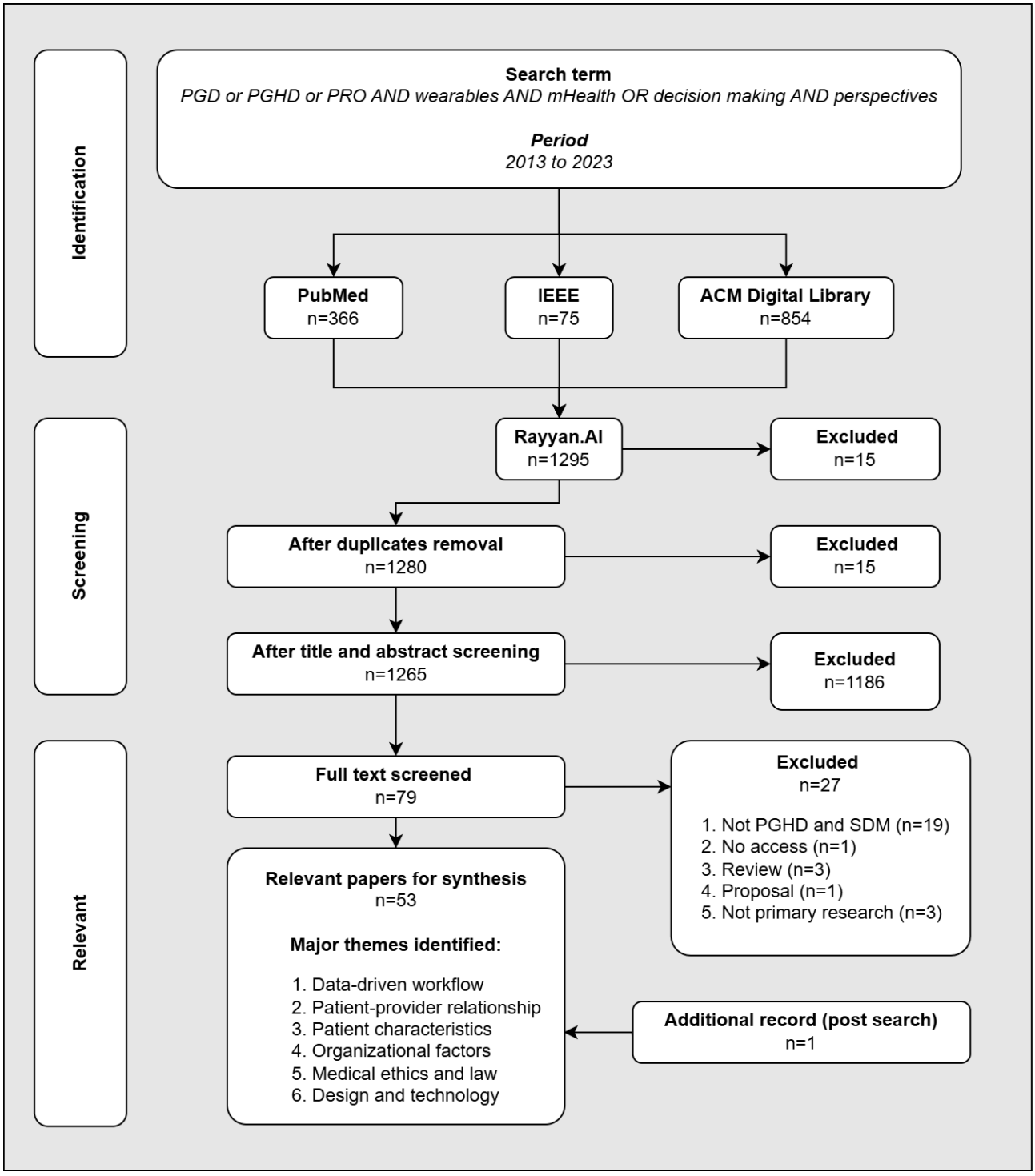
Results

Identified Papers

Our search in the ACM, PubMed, and IEEE digital libraries resulted in 1295 records. The distribution of the results was as follows: ACM 854, PubMed 366, and IEEE 75. After the automatic removal of duplicates by Rayyan and manual removal by the first author, 1280 papers remained. Next, the first and second authors read all titles and abstracts to filter by IC and EC, and 1186 papers were removed, resulting in 79 papers. After full-text screening, 27 papers were removed. Finally, 52 papers were selected for the full-text review by the first two authors. The first author subsequently performed the data extraction using a predefined Excel sheet that captured study characteristics and stakeholder-reported barriers and enablers. One paper was manually added as it was mentioned as an influential related work in four papers [34,44,47,56] published in subsequent years, resulting in 53 papers [6,7,9,12,13,21,30,31,34,43-48,56-93] (see Multimedia Appendix 4). A validation check-in was performed with the last author after reviewing 25 papers. The flow diagram for the review is provided in Figure 2.

XSL•FO
RenderX

Figure 2. Flow diagram showing data extraction process. mHealth: mobile health; PGD: patient-generated data; PGHD: patient-generated health data; PRO: patient-reported outcome; SDM: shared decision-making.



Characteristics of Included Papers

The works predominantly originated from English-speaking countries (Table 1). About 50% (n=33) of them were from the United States and more than 10% (n=7) from the United Kingdom. The final number is higher than in 53 studies, as some studies were conducted in more than one country.

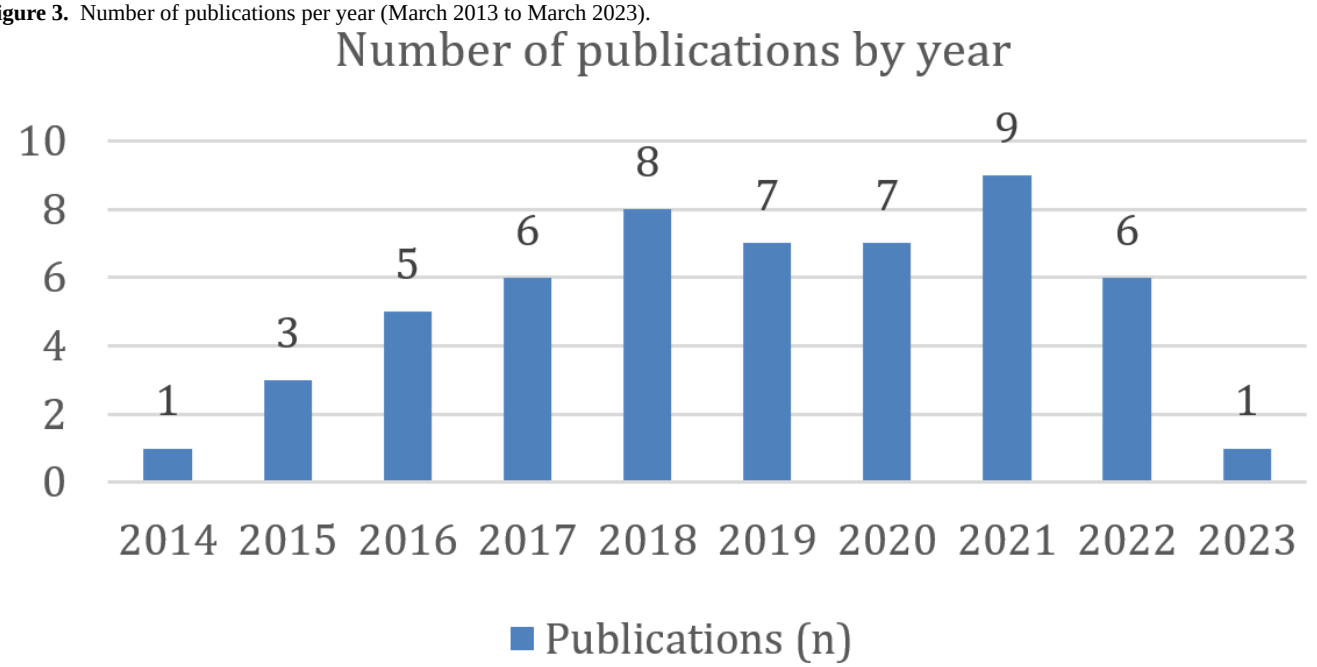
XSL•FO
RenderX

Table . Number of included studies by country of study setting.

Country/region	Values, n (%)
United States	33 (53.2)
United Kingdom	7 (11.3)
South Korea	3 (4.8)
Netherlands	3 (4.8)
Italy	2 (3.2)
India	2 (3.2)
Germany	2 (3.2)
Canada	2 (3.2)
Japan	1 (1.6)
France	1 (1.6)
Denmark	1 (1.6)
China	1 (1.6)
Brazil	1 (1.6)
Belgium	1 (1.6)
Australia	1 (1.6)
European Union	1 (1.6)
Total	62 (100)

The number of papers published annually (Figure 3) has steadily climbed and peaked in 2021. The drop in 2022 could be due to the COVID-19 pandemic, which resulted in a decline in in-person consultations. Furthermore, as the search was conducted in the first quarter of 2023, this resulted in a low count. Most of the papers were about noncommunicable diseases and chronic conditions. Only one paper included perspectives of patients with communicable diseases [91].

Figure 3. Number of publications per year (March 2013 to March 2023).



Factors Affecting the Integration of PGHD for SDM

In this section, we present the findings related to our guiding research question. The analysis identified several enablers and barriers related to the following six themes: (1) patient-provider collaboration with data, which refers to the interaction between HCPs and patients and several factors that affect this interaction, such as trust, expectations over data use, etc; (2) patient characteristics were mainly attributed to factors influencing patient behaviors and tracking habits or the effects of tracking; (3) organizational factors were related to processes within health care systems that could be improved or existing barriers preventing PGHD from integrating effectively into clinical

XSL•FO
RenderX

workflows; (4) medical ethics and law refer to ethical and legal considerations such as patient autonomy, the use of PGHD for insurance purposes, and the questionable value of PGHD for clinical decisions; (5) data-driven workflow-related factors are broadly associated with characteristics, processes, and tools related to data, such as integration, standardization, relevance, trustworthiness, etc; and finally, (6) design and technology refer to the technical implementation of tools that generate or use PGHD and the consideration of human factors in such tools.

The heatmap (Figure 4) shows the breakdown of themes by barriers and enablers and stakeholders based on the number of occurrences throughout the 53 papers selected for this review. Multiple coded segments within the same study were counted as a single occurrence. We also present a detailed heatmap showing the breakdown of codes by barriers and enablers raised by HCPs and patients in Figure 5. In Multimedia Appendix 5, we present an Excel sheet with detailed mapping of clinician or patient enablers or barriers across the 53 papers.

Figure 4. Heatmap showing the distribution of themes across stakeholder categories. Counts represent the number of studies in which a theme was identified at least once. A darker shade represents a higher count. CB: health care professional barrier; CE: health care professional enabler; PB: patient barrier; PE: patient enabler.

Theme	CE	PE	CB	PB
Data	22	13	31	17
Design and technology	3	9	8	13
Medical ethics/law	1	2	5	1
Organizational	18	14	27	6
Patient characteristics	7	20	18	17
Patient-provider relationship	31	24	13	8

XSL•FO
RenderX

Figure 5. Heatmap showing the distribution of high-level codes representing barriers and enablers across stakeholder categories. Counts represent the number of studies in which a code was identified at least once. A darker shade represents a higher count. Multiple coded segments within the same study were counted as a single occurrence. CB: health care professional barrier; CE: health care professional enabler; PB: patient barrier; PE: patient enabler; PGHD: patient-generated health data.

Theme	Codes	CE	PE	CB	PB
Data	Context	0	1	3	0
	Data driven workflow	11	5	16	3
	Health data literacy	2	0	7	6
	Insufficient data	0	0	2	3
	Relevance	3	1	6	5
	Security and privacy	0	1	6	6
	Sensemaking	13	8	6	4
	Trustworthiness	4	0	13	6
Design and technology	Acceptance of technology	0	0	1	1
	Accessibility	0	1	4	4
	Technical reliability	1	0	2	5
	Usability	2	9	2	5
Medical ethics/law	Insurance	1	0	0	0
	Patient autonomy	0	2	1	1
	Value of PGHD as evidence	0	0	4	0
Organizational	Collaboration and productivity	1	2	2	0
	Preparation for consultation	4	1	0	0
	Process, tools and policies	13	11	13	4
	Resource (time and workload)	2	1	21	4
Patient characteristics	Altruism	0	2	0	0
	Burden of tracking (workload)	1	1	4	8
	Habit of tracking	0	0	0	1
	Health literacy	2	1	2	1
	Introspectivity	2	15	0	1
	Mispresentation or hiding of data	1	0	5	2
	Negative repercussions (of tracking)	0	0	7	5
	Noncompliance	0	0	2	3
	Self-motivation	2	3	0	1
	Willingness to use PGHD and technology	0	1	2	0
Patient-provider relationship	Accountability and motivation of patient	5	3	0	3
	Better communication	14	11	1	0
	Better diagnosis	7	6	0	0
	Better picture of patient	15	8	3	1
	Collaborative care	4	10	1	0
	Expectations over PGHD use	2	4	7	3
	Objectify information	7	4	1	0
	Personalized care	13	11	1	0
	Shift of responsibility	0	2	4	0
	Transparency of data driven decisions	0	0	0	1
	Trust	1	5	4	2
	Well-informed patient	4	2	0	0

The most significant enabler for PGHD integration is how it could promote patient-provider collaboration with data, especially concerning improving communication, allowing for patient-centered care, and providing HCPs with a better picture of the patients. Aspirations followed this in terms of having mechanisms that enable effective sensemaking of health data and having data-driven clinical workflows.

Among barriers, data-driven workflow (eg, lack of standards, poor integration, and lack of access to data) has emerged as the most significant, followed by organizational factors (eg, unavailability of tools and lack of time) and patient characteristics (eg, selective disclosure of data and increased burden of tracking). In the subsequent sections, we will delve deeper into the specific factors.

XSL•FO
RenderX

Patient-Provider Collaboration With Data

Patient-provider collaboration with data is mainly seen as an enabler for the integration of PGHD. The key enablers include the ability to offer personalized care, facilitate communication between patients and HCPs, and improve diagnosis. For example, "both healthcare providers and patients use data to approach the unknown, whether they use it to identify the possible causes of infertility, to identify the fertile window, to decide on potential treatments, or to choose follow-up treatments" [56]. PGHD is also seen as an aid in building trust by objectifying subjective information: the physiotherapists found that the information provided by FitViz helped them in conducting "patient-driven" consultations [66].

While patients are often enthusiastic about sharing the data with HCPs, "patients also wanted to use their data to get recognition for their efforts and to show their doctor they take their health plan seriously" [21]. HCPs are concerned about a shift of responsibility as they might be more accountable for the data shared by the patients: "clinicians were concerned that by having continuous access to patient data, they would become responsible for monitoring that data for concerning changes that may be indicative of a mental health crisis, such as a manic episode or a suicide attempt" [65] and feel an obligation to interpret and explain the data: "simply" receiving PGHD around daily symptoms created a feeling of an 'obligation to act' for the clinicians" [71]. Moreover, with the capability of remote monitoring, patients "want to have the ability for their physician to conduct remote monitoring to verify the treatment plan is working and to pick up early warning signs of relapse or deterioration" [48].

Patient Characteristics and Perceptions

HCPs, in general, are worried about the (unintended) consequences of patients tracking and bringing their PGHD to consultations. They feel that an increased emotional attachment of patients to their PGHD could lead to depression and anxiety. For example, therapists highlight that "depression could be a potential side effect of tracking" [84] and patients could also "overreact to [negative] data" [43]. Both HCPs and patients have concerns about the effort required for tracking and feel that tracking could increase patients' burden; in a study by Ancker et al [91], a diabetes patient who had given up self-monitoring of blood glucose said, "it's too cumbersome for me."

In addition, patient motivations for sharing data with HCPs are not always clear due to *selective disclosure*, as the "patient makes the decision about what kind of data the clinician can see. [As a result] clinicians [were] uncomfortable" [65] with such data unless they had the entire dataset. HCPs are concerned about the "increased transparency via PGHD and how it could lead to noncompliance [by patients]" [31].

Although most feedback was negative, HCPs feel that having access to data could allow patients to better engage with their data through *introspection* and *motivate* them to self-treat their conditions. Patients feel more self-aware of their condition, and self-tracked data helps them see their habits [21]. Furthermore, patients also believed that sharing their data could benefit future research for treating patients with similar conditions: "they hoped their Fitbit data would benefit future veterans" [78].

Data-Driven Workflow

The data-driven workflow theme comprised several enablers and barriers. Within the theme, there are several factors considered as enablers, among which the generation of data, data actionability, and having access to standardized tools integrated well into the clinical workflow.

From the HCP's perspective, data sensemaking is a key enabler for unlocking the value of PGHD. For example, HCPs feel that "when patients struggled to identify triggers, visualizations of quantitative analyses could help patients and providers understand underlying nutrient-symptom relationships" [31]. Data sensemaking enabled patients to effectively use PGHD, facilitating better reflection of their health conditions and empowerment of self-care. For example, in a study of diet monitoring, "HCPs and participants were able to [make sense of their data to] identify eating patterns or potential triggers and used Foodprint as a structuring artifact to discuss actionable next steps" [34]. However, in a study with diabetes patients and providers, "commercially available visualizations that are built using clinical guidelines were useful for providers but not for patients" [83]. HCPs want standardized tools to interact with data and better understand the patient's condition. For instance, HCPs stress the importance of a "uniform clinician interface that integrates [PGHD] from various individual tools, supporting the findings in previous studies" [85] and take appropriate action, leading to better data actionability. Contextual information—how and when the data was collected—is also mentioned as a key to data actionability. Access to the data's context could allow for the discovery of "possible confounds including emotional and physical health, hydration, and exercise" [82]. Integration could enable HCPs to *access* and review data before consultations. In one of the studies, "geriatricians preferred to review the data before the consultation, and then implement a new part of their care plan based on the data" [70].

There are also several barriers identified. The introduction of new tools and technology meant that HCPs must have the necessary *literacy for* interacting with data. Some "[HCPs] doubt their ability to advise on tracking; many providers doubt patient ability to review tracking data" [30]. HCPs also highlight the need for better *data integration* with existing EHR and electronic medical records systems. For example, Zhu et al [81] found that as "commercial sleep tracking technology did not support data exporting or sharing, many clinicians continued to use paper sleep diary."

HCPs were also concerned about PGHD's *quantity*, *quality*, and *reliability*, which could, in turn, affect its *trustworthiness*. When "different patients use different apps, it [became] difficult for health care providers to interpret the data and assess the reliability of each app" [58]. Regarding the amount of data, some HCPs' feedback about receiving or reviewing data relevant to the consultation, given that unnecessary data could result in ineffective decision-making. From the patient's perspective, there is also a lack of awareness about what to track. As a result, "many wanted to track more data than may be feasible" and

XSL•FO
RenderX

wanted to have "smart defaults for how to track because they did not know what would reasonably balance burden and usefulness" [46].

Organizational Factors

Organizational factors mainly encompass factors about a health care organization such as a hospital. The primary barrier is HCPs' lack of time—"the cardiologists in our sample already encounter problems with data overload in their daily practice" [80] and the need for more time and effort to review patient data. HCPs share that "patient-provided data sources will add layers of data assessment to practice, and questioned whether this information may adversely affect efficient workflow" [87].

HCPs need more support, tools, and incentives from upper management to integrate PGHD into clinical care practices. HCPs feel that although activity monitoring tools could help to streamline means of collecting data, "current activity monitoring tools are insufficient to address program specific prehabilitation assessment needs; to exacerbate this challenge, stakeholder tailored tools do not currently exist" [67]. Regarding clinical workflow, HCPs feel that the ready availability of PGHD before consultations could be beneficial in allowing them to review patients' conduct to *prepare for consultations* beforehand, leading to better consultations.

> *Nearly all providers agreed that the best time to receive the patient-generated health report was immediately before a scheduled clinic visit.* [92]

This, in turn, leads to better consultations: "pre-visit notes helped health experts focus on participant goals and questions when reviewing food photos" [34].

Medical Ethics and Legal Aspects

Medical ethics and legal aspects refer to the responsibility of HCPs, ethics, reimbursement, and legal aspects of data. HCPs are worried about regulations such as HIPAA (Health Insurance Portability and Accountability Act) obstructing the introduction of PGHD into their care practices. Furthermore, autonomy is a factor for concern among HCPs and patients. For example, some "older adults perceived that monitoring PGHD increased the transparency of their (lack of) engagement in healthy behaviors" and "perceived [it] as a threat to autonomy" [31]. Finally, the unproven value of PGHD in health care—"lack of evidence for use" [9] of PGHD raises ethical concerns about how such data can be used for clinical decision-making, thereby limiting its adoption. An HCP felt frustrated rearranging the PGHD data and structuring it as evidence within their clinical framework [87]. Within the broader theme of ethics and legalities, one HCP mentions how PGHD could benefit insurance concerning a patient's recovery: "I think it's also good to show insurance kinda, hey, we are making changes, you know, getting them better in this aspect" [84].

Design and Technology

Both HCPs and patients highlight that well-implemented technologies with good usability can be a driver for adoption and sustained use. For example, among patients, learnability and ease of use—subqualities of usability—are favorable for patients in two studies: "GeniAuti made their recording process easier in everyday life" [60] and "ease-of-use of DiaFocus and found it easy to use and easy to learn to use" [57].

For HCPs, the usability of their tools is a key factor for adopting PGHD and missing features such as "inability to sort the data" [89] and added additional time. HCPs are also concerned about the reliability of the applications and the technologies that the patients adopted, especially when patients might use a myriad of them based on their preferences. On the other hand, "when hospitals offer the same app that has been clinically validated to all patients, health care providers can become, over time, knowledgeable about [it]" [58].

Overall, both stakeholders believe that technologies should be designed to be accessible so that population groups such as older adults can adopt these technologies.

> *Clinicians also alluded to old age as a potential challenge as the older generation is not as fluent with mHealth, wearable technology, and other devices that collect PGHD.* [31]

In a particular study, a participant's medical condition—rheumatoid arthritis—caused issues with their participation due to the tool's [lack of] accessibility [63].

PGHD Across a Patient Data Journey and HCP's Clinical Workflow: 10-Stage Workflow

Our analysis includes a wide range of barriers and enablers which can be arranged across a patient's journey and HCP's clinical workflow to provide logical flow. In doing so, we extend and augment the work by West et al [12]—a 6-stage workflow model aligning patient and HCP objectives—and the work by Li et al [20]—the stage-based model of personal informatics systems—to demonstrate how PGHD is integrated beyond clinical settings (Figure 1). Our updated workflow identifies four new stages encapsulated within the pre- and postconsultation phases in addition to the existing stages identified by West et al [12]—stage (1) align patient and HCP objectives, stage (2) evaluate data quality, stage (3) judge data quality, stage (4) rearrange the data, stage (5) cointerpret the data, and stage (6) decide on a plan or action. It also emphasizes the dyadic relationship between the HCPs and patients and the numerous factors influencing the integration of PGHD along the patient journey and HCP's clinical workflow. It starts from the data collection (stage 1) stage till a clinical action is taken (stage 10) and repeats (especially for patients with chronic or multiple conditions). The following subsections describe the newly added steps in further detail.

Collect Data (Stage 1)

Data collection is a crucial stage where patients initiate the tracking of PGHD. In some cases, they might be embarking on this activity on their own or through social influence; on the other hand, it could have been due to instructions from their HCP, for example, when treating medical conditions over the long term. However, it is also worth noting that the monitoring process requires additional effort—although there are methods for passive collection, for example, through wearable or stationary sensing devices, these do require considerable effort to set up and maintain. As such, both active (eg, through

XSL•FO
RenderX

occasional questionnaires or more frequent ecological momentary assessment) as well as passive sensing introduce an additional burden to patients—on top of managing their medical condition.

> *[Patients] mentioned that daily measurements were too burdensome or medicalizing, especially when they perceived their symptoms or blood pressure to be stable.* [69]

Reflect on Data (Stage 2)

In the reflection stage, patients can interact with the data they collected through the various sensemaking mechanisms available. This could be presented as data summaries, PDF reports, or graphs of different formats. However, at this stage, there are both facilitators and barriers. By accessing such data, patients could reflect on their physiological data such as stress levels, weight, etc, and information such as physical activity trends.

> *[P]atients looked to their data for self-awareness of their current lifestyle and described its value in terms of helping them see their habits.* [21]

This could allow them to be well-informed about themselves before consultations. However, if "veterans do not understand or trust the data, it could be difficult for them to be motivated to use it" [78]. There were also incidents in specific medical conditions—for example, in mental health—where having access to such data could reinforce negative views of their condition and agitate symptoms [72] or demotivate patients. Additionally, the constant recurrence of a condition—hence having increased tracking burden—also prevented some patients from self-monitoring: "many unengaged patients reported being easily discouraged by their [arterial fibrillation] AF recurrence, which they said caused them to self-monitor less" [77].

Integrate Data (Stage 3)

The integrate data stage entails the import of data into a clinical setting. Here, the patient attempts to "upload" their data into clinical systems, or the HCP could manually associate them with their EHR systems (a common need to store the foundations for decision-making). However, at this stage, HCPs faced trouble importing such data due to incompatibilities or insufficient interfaces to their systems. In a study, although "patients used mobile apps for their sleep diary, clinicians who received this data had to manually input each data point" [81] due to limitations in data sharing functionalities. There were issues about poor interoperability of self-tracking technologies—primarily due to the nonstandardization of devices or technologies—which prevented seamless integration into EHR systems. Notably, current-generation digital technologies for supporting sensemaking from PGHD cannot easily scale novel data types, handle incomplete or incongruent data, or integrate multimodal PGHD streams.

Take Action (10)

Finally, in the *taking action* stage, after a consultation with an HCP, the presence of data could allow patients to be more accountable for self-treating their conditions based on decisions made during the consultations. However, in some cases, there was a concern about the HCP's ability to monitor patients' conditions remotely or with more granular details; this could invade patients' privacy and threaten their *autonomy* [31]. In addition to this, although HCPs could give recommendations to patients on what tools to use to control their condition, they might not necessarily comply with it:

> *Providers sometimes recommended tracking tools to patients for clinical diagnosis and management. However, patients do not always follow those recommendations.* [13]

Tensions Between HCPs and Patients

Having mapped the barriers and enablers across the 10 stages, we now turn to the cross-cutting tensions that emerged between HCP and patient perspectives throughout the workflow. Excerpts from relevant papers are presented below each point:

Regarding tension 1, patients need clarification or are unaware of the *relevance of data* and need more *health literacy*. This results in (*time-limited*) HCPs facing an overload of (unnecessary) data.

> *[There is] a need to negotiate with patients when determining which data elements to collect, set patients' expectations for communication about PGHD, and let patients know that they would not be contacted if everything was normal.* [89]
>
> *Older adults and clinicians perceived that the lack of clinical knowledge by patients leads to a collection of irrelevant PGHD and decreases the usefulness of the information.* [31]

Regarding tension 2**,** patients, at times, have a subjective and emotional connection with their data and were hoping for detailed explanations of the data they brought for consultations. HCPs, however, intend to view the data objectively while making clinical decisions. This can result in mismatched expectations over collected data by patients and sometimes misunderstandings over how HCPs made clinical decisions using the data, which results in ineffective consultations.

> *Patients also expected providers to engage with the data and provide a personalized treatment plan.* [21]
>
> *The conflicting views often led to disagreement between the caregivers' and experts' interpretations of the children's challenging behaviors.* [60]

Regarding tension 3, patients feel that increased monitoring and visibility of PGHD could lead to disruption of their autonomy and loss of privacy; however, HCPs feel that having (access to) more (readily available) data could allow for better health support.

> *Participants were concerned about talking to clinicians they needed to become more familiar with and about the privacy of their PGD.* [68]

Regarding tension 4, HCPs desire a *complete picture of the patients (to facilitate decision-making*) but at the same time feel challenged by *increased workload* due to the increased data and insufficient time to handle them.

XSL•FO
RenderX

> *Provide the Context, Longer View, and Whole Picture.* [90]
>
> *If patients share the collected data with their doctor, the interpretation of the data and appropriate response becomes the doctor's response.* [80]

Regarding tension 5, patients have issues trusting the HCP's ability to interpret and review data. Furthermore, HCPs themselves had self-doubts about their capabilities.

> *These barriers include a lack of time to review detailed records, questions about providers' expertise, and skepticism about additional benefits of reviewing data.* [30]
>
> *The visualizations also caused some unease for [participants], who mentioned feeling "embarrassed" or "anxious" about giving the impression that they lacked necessary knowledge to explain the visualizations.* [82]

Regarding tension 6, patients and HCPs have varied preferences on which type of problem to focus on and how they wanted the data to be represented.

> *Patients and providers differ on the type of problem to focus on.* [83]
>
> *Data gives different insights to patients and providers.* [83]
>
> *Patients and providers use different representations of data to identify problems.* [83]

Regarding tension 7, ethical and legal challenges of data privacy, patient autonomy, and HCP accountability.

> *...the use of mHealth contributed to them being in control of their own health and to bring their own perspectives to the fore in consultations with health care professionals.* [69]
>
> *Older adults felt that monitoring PGHD increased the transparency of their (lack of) engagement in healthy behaviors.* [31]

Discussion

Summary of Results and Implications

Across the 53 reviewed papers, we identified 6 overarching themes influencing the integration of PGHD into patient journeys and clinical workflows to enable SDM: patient-provider collaboration, patient characteristics and perceptions, organizational factors, medical ethics and law, data-driven workflow factors, and design and technology. Furthermore, we surface 7 cross-cutting tensions faced by HCPs and patients. These findings are relevant for system designers and health care practitioners seeking to digitalize clinical pathways through the integration of PGHD.

Despite the potential of PGHD, there are still significant challenges. There is a lack of transparent processes, standardized tools, and communication about how data should be presented during consultations; individuals bringing their PGHD to consultations would currently most likely add additional workload for HCPs by providing irrelevant and nonstandard data and expecting interpretation. Furthermore, varying patient motivations and their selective data disclosure add layers of complexity to the relationship between HCPs and patients while integrating PGHD into clinical workflows.

Unpacking such challenges while making health and care increasingly patient-centric and assuring that patients feel heard and empowered is an essential goal in many national and international health care strategies (eg, other studies [94-97]). SDM is a key enabler of patient-centered care and helps with fostering adherence and self-determination as well as proactive approaches to health and care. In our 10-stage workflow model, we posit that stage 8 "cointerpretation of data"—or collaborative sensemaking between HCPs and patients—is the primary enabler for successful SDM when integrating PGHD. Prior work indicates that collaborative sensemaking of PGHD can improve patient-provider communication by facilitating deeper discussions regarding personal values and granting HCPs a more nuanced understanding of treatment progress [19,98]. As such, we argue that system designers should prioritize the development of interfaces that scaffold this collaborative sensemaking process to enable SDM.

Toward Enabling Integration of PGHD in Patient Journey and Clinical Workflows

Information infrastructures in health care are embedded within social arrangements and technologies and are spatiotemporal in nature [36]. HCI has a central role to play in investigating them and creating health care technologies that are fit for purpose [17].

In a clinical context, HCPs may already be facing increased risk of burnout due to poor technology integration [33], and given the potential benefits and burden PGHD brings, there is a need to prioritize how data-related tools can be better integrated. On a higher level, there needs to be organizational support in creating policies and processes to support the use of PGHD and creating data-driven workflows that seamlessly integrate such data into clinical practices. However, we identified challenges in a clinical setting involving how patients collect data and collaborate with HCPs in a clinical context.

Work on aligning and interactively navigating priorities between HCP and patient [12,21] appears clearly implicated. Tensions could arise from patient preferences or the subjectivity of data due to the emotional connectivity of the data conflicting with HCPs adopting a very objective view of data with their medical background—typically due to medical ethics. By juxtaposing the needs and priorities of both stakeholders, we discussed the alignments and tensions between them. We also identified design opportunities for system designers to consider addressing and balancing the stakeholders' priorities adequately by proactively building on enablers and avoiding or decisively tackling barriers.

Overall, the scoping review outcomes emphasize a need to adopt a systemic lens on the issue of integrating PGHD into clinical practice by considering its flow across the patient data journey and clinical workflows. Our research team, consisting of the first, second, and last authors, conducted a thorough review of the existing 6-stage workflow model [12] and the emerging codes to establish the new stages and the enablers and barriers

XSL•FO
RenderX

at each stage, and the potential tensions. This process resulted in the augmented 10-stage workflow model (Figure 1). This model (see simplified model in Multimedia Appendix 6) includes pre- and postconsultation phases, as well as corroborated formerly identified barriers and enablers and newly identified ones.

Addressing Barriers

To facilitate the practical application of our findings, we organize the barriers across the stages along with potential mitigation strategies in Table 2.

Table . Barriers and mitigation strategies.

Stages	Barriers	Mitigation strategies
Preconsultation		
Collect data (stage 1)	Burden of tracking	• Use passive tracking as much as possible and recommend only tracking that is necessary for consultations.
Reflect on data (stage 2)	Negative repercussions of tracking	• Educate patients on possible negative impacts of data and offer adequate resources to obtain support on data sensemaking.
Integrate data (stage 3)	Poor data integration	• Standardize interfaces and data export formats among commercial devices and applications for better integration into EHR[a] systems.
Post consultation		
Take action (stage 10)	Shift of patient autonomy and noncompliance	• Ensure that patient preferences on how their data is viewed and accessed are respected and communicated with them transparently. • Educate patients on the benefits of adherence.
All stages		
N/A[b]	Overburdened HCPs[c]	Across all stages: • Implement data sensemaking support in-line with clinical workflows. • Consider shifting to partially automated sensemaking steps that already occur before consultation
N/A	Lack of time for HCPs	• Introducing tools to improve the operational efficiency of clinical workflows
N/A	Questionable value of PGHD[d]	• Conduct empirical research on PGHDs and how they improve patient-centered care.

[a]EHR: electronic health record.

[b]N/A: not applicable.

[c]HCP: health care professional.

[d]PGHD: patient-generated health data.

Addressing Tensions

Apart from addressing key barriers, it is essential to address divergent perspectives resulting in tensions between the stakeholders. Given the dyadic relationship between patients and HCPs during SDM, it is essential to understand potential misalignments between both stakeholders, which results in tensions. In Table 3, we delineate these specific tensions alongside targeted alignment strategies designed to harmonize stakeholder expectations and foster collaborative care.

XSL•FO
RenderX

Table . Tensions and alignment strategies.

Number	Tension	Alignment strategies
1	Patients' lack of awareness about what data to bring vs time-starved HCPs[a]	Improve health (data) literacy of patients, align expectations, or establish standards over what data are needed for the consultation and for SDM[b], thereby engendering mutual trust.
2	Patients' subjectivity vs HCPs' objectivity	HCPs should set expectations regarding how PGHD[c] would be used for SDM and provide an explanation of how they influenced the SDM process.
3	HCPs' need for access to patient data vs patients' desire for privacy and autonomy	Patients and HCPs should set boundaries over how and when HCPs track PGHD to maintain patient autonomy. Appropriate privacy mechanisms could be designed for dynamic configuration of consent.
4	HCPs want to have a complete picture of patients vs limited consultation time	Equipping HCPs with the necessary tools such as clinical decision support and AI-assisted systems could help augment HCPs with PGHD interaction.
5	Patients' perception of lack of HCP data literacy affected trust	Improve HCPs' data literacy capabilities and assure patients about their abilities to make data-driven decisions.
6	Differing perspectives and approaches that patients and HCPs have when interacting with PGHD for diagnosis	Tools should promote mutual understanding between both patients and HCPs and support problem identification, preferably in an automated manner [83].
7	Ethical and legal need for addressing the dilemma of data privacy and patient autonomy vs HCP accountability	Foster cross-sectoral collaboration between patients, HCPs, and legislators to co-design the most appropriate governance frameworks for a given chronic disease context. This could include steps such as implementing dynamic consent and adapting policies such as HIPAA[d] in the United States and EHDS[e] to ensure that they do not diminish the quality of care [99]. Furthermore, alignment with medical ethics standards should incorporate national and contextual nuances, allowing for safeguards such as involvement of ethical committees for ethical oversight.

[a]HCP: health care professional.

[b]SDM: shared decision-making.

[c]PGHD: patient-generated health data.

[d]HIPAA: Health Insurance Portability and Accountability Act.

[e]EHDS: European Health Data Space.

Some of these alignment strategies could be considered as implications for design at an organizational or policy level based on the health care system. At an organizational level, system designers can adopt generative co-design activities with HCPs and patients, such as the approach by Höppchen et al [100] to design clinical pathways which systematically align the integration of PGHD.

Limitations

Our scoping review's findings were based on the papers derived from 3 databases, and other unpublished or nonpeer-reviewed works could be available in other repositories. They reflect studies primarily conducted in English-speaking, high-income countries, partly a result of limiting our search to English papers. As such, our review may underrepresent perspectives from low- and medium-income countries where infrastructural capabilities, sociocultural factors, and health care systems may vary significantly. We also acknowledge that no critical appraisal was conducted for individual studies as they are optional for scoping reviews and additionally due to the heterogeneity of the study designs included in our review. Furthermore, as mentioned in the positionality statement above, there could be researchers' subjectivity in the interpretation of the findings. As HCPs and patients unequally contributed to the findings, the perspectives should be explicitly assessed in empirical work and in a balanced manner. However, these preliminary findings give readers and designers a sense of factors to consider when integrating PGHD into digital health technologies and wider health care systems.

XSL•FO
RenderX

Conclusions

PGHD has immense potential to enhance patient-centered care and lead toward P4 (predictive, preventive, personalized, and participatory) medicine. However, its widespread integration in clinical settings still needs to be improved. Our research question aimed to paint a holistic and ecological understanding of the PGHD integration barriers and enablers perceived or experienced by patients and HCPs to better facilitate PGHD integration. By reviewing works from the past 10 years and incorporating both patient and HCP perspectives, we identified several enablers and barriers across 6 key themes. We build on the previous 6-stage model by West et al [12], adding 4 new stages beyond the clinical setting, including "collect data," "reflect on data," "integrate data," and "take action," and added the perspectives of both patients and HCPs to identify new barriers and enablers across the various stages of interaction with PGHD. Our 10-stage workflow model of aligning patient and HCP objectives in integrating PGHD implies several challenges to be tackled in a complex health data ecosystem [18] and health data lifecycle [17]. Furthermore, the model forms the foundation for deriving design implications around tensions and alignments between HCPs and patients along data-enabled patient journeys and clinical workflows. By addressing challenges introduced by PGHD, SDM could be more efficient for both parties, leading to better patient-centered care and life-accompanying personalized data-enabled health. Future work should investigate how our workflow model performs in real-world settings and expand evidence from underrepresented regions to improve the global applicability of PGHD integration among different health care systems.

Acknowledgments

We would like to thank all our colleagues who helped with proofreading this paper. The authors declare the use of generative AI (GenAI) in the research and writing process. According to the GAIDeT (2025; Generative Artificial Intelligence Delegation Taxonomy), the following tasks were delegated to GenAI tools under full human supervision: (1) proofreading and editing and (2) reformatting. The GenAI tool used was ChatGPT (version 4.5; OpenAI). Responsibility for the final paper lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the outcomes. Declaration submitted by the collective.

Funding

The authors declare that no financial support was received for the research, authorship, and/or publication of this paper.

Data Availability

All data are included in this paper and its supplementary materials.

Authors' Contributions

Conceptualization: PVSP, DK, JS
Methodology: PVSP, DK, JS
Analysis: PVSP, DK, JS
Validation: PVSP
Writing – original draft: PVSP
Writing – review and editing: All authors

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search query.
[DOCX File, 16 KB - mhealth_v14i1e85197_app1.docx]

Multimedia Appendix 2
Inclusion and exclusion criteria.
[DOCX File, 16 KB - mhealth_v14i1e85197_app2.docx]

Multimedia Appendix 3
Themes and definitions.
[DOCX File, 20 KB - mhealth_v14i1e85197_app3.docx]

Multimedia Appendix 4

XSL•FO
RenderX

Characteristics of included papers charted as a table. The study ID listed in this document is to be used in conjunction with the spreadsheet in Multimedia Appendix 5.
[DOCX File, 39 KB - mhealth_v14i1e85197_app4.docx]

Multimedia Appendix 5
An Excel sheet with detailed mapping of clinician or patient enablers or barriers across the 53 papers.
[XLSX File, 18 KB - mhealth_v14i1e85197_app5.xlsx]

Multimedia Appendix 6
Simplified 10-stage workflow model for integrating patient-generated health data.
[DOCX File, 155 KB - mhealth_v14i1e85197_app6.docx]

Checklist 1
PRISMA checklist.
[PDF File, 164 KB - mhealth_v14i1e85197_app7.pdf]

References

1. Peyroteo M, Ferreira IA, Elvas LB, Ferreira JC, Lapão LV. Remote monitoring systems for patients with chronic diseases in primary health care: systematic review. JMIR mHealth uHealth 2021 Dec 21;9(12):e28285. [doi: 10.2196/28285] [Medline: 34932000]
2. do Nascimento IJB, Abdulazeem H, Vasanthan LT, et al. Barriers and facilitators to utilizing digital health technologies by healthcare professionals. NPJ Digit Med 2023 Sep 18;6(1):161. [doi: 10.1038/s41746-023-00899-4] [Medline: 37723240]
3. Choe EK, Lee NB, Lee B, Pratt W, Kientz JA. Understanding quantified-selfers' practices in collecting and exploring personal data. Presented at: CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; Apr 26 to May 1, 2014. [doi: 10.1145/2556288.2557372]
4. Feng S, Mäntymäki M, Dhir A, Salmela H. How self-tracking and the quantified self promote health and well-being: systematic review. J Med Internet Res 2021 Sep 21;23(9):e25171. [doi: 10.2196/25171]
5. A'Court C, Jenkins W, Reidy C, Papoutsi C. Patient-initiated cardiovascular monitoring with commercially available devices: how useful is it in a cardiology outpatient setting? Mixed methods, observational study. BMC Cardiovasc Disord 2022 Sep 29;22(1):428. [doi: 10.1186/s12872-022-02860-x] [Medline: 36175861]
6. Oh CY, Luo Y, Jean B, Choe EK. Patients waiting for cues: information asymmetries and challenges in sharing patient-generated data in the clinic. Proc ACM Hum-Comput Interact 2022;6(CSCW1):1-23. [doi: 10.1145/3512954]
7. Zhu H, Colgan J, Reddy M, Choe EK. Sharing patient-generated data in clinical practices: an interview study. AMIA Annu Symp Proc 2016;2016:1303-1312. [Medline: 28269928]
8. Lordon RJ, Mikles SP, Kneale L, et al. How patient-generated health data and patient-reported outcomes affect patient-clinician relationships: a systematic review. Health Inf J 2020 Dec;26(4):2689-2706. [doi: 10.1177/1460458220928184] [Medline: 32567460]
9. Lavallee DC, Lee JR, Austin E, et al. mHealth and patient generated health data: stakeholder perspectives on opportunities and barriers for transforming healthcare. mHealth 2020;6:8. [doi: 10.21037/mhealth.2019.09.17] [Medline: 32190619]
10. Madanian S, Nakarada-Kordic I, Reay S, Chetty T. Patients' perspectives on digital health tools. PEC Innov 2023 Dec;2:100171. [doi: 10.1016/j.pecinn.2023.100171] [Medline: 37384154]
11. Boffetta P, Collatuzzo G. Application of P4 (predictive, preventive, personalized, participatory) approach to occupational medicine. Med Lav 2022 Feb 22;113(1):e2022009. [doi: 10.23749/mdl.v113i1.12622] [Medline: 35226650]
12. West P, Van Kleek M, Giordano R, Weal MJ, Shadbolt N. Common barriers to the use of patient-generated data across clinical settings. Presented at: CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems; Apr 21-26, 2018; Montreal, Quebec, Canada p. 1-13. [doi: 10.1145/3173574.3174058]
13. Abdolkhani R, Gray K, Borda A, DeSouza R, DeSouza R. Patient-generated health data management and quality challenges in remote patient monitoring. JAMIA Open 2019 Dec;2(4):471-478. [doi: 10.1093/jamiaopen/ooz036] [Medline: 32025644]
14. Lee K, Hoti K, Hughes JD, Emmerton LM. Consumer use of "Dr Google": a survey on health information-seeking behaviors and navigational needs. J Med Internet Res 2015 Dec 29;17(12):e288. [doi: 10.2196/jmir.4345] [Medline: 26715363]
15. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (Or it takes at least two to tango). Soc Sci Med 1997 Mar;44(5):681-692. [doi: 10.1016/S0277-9536(96)00221-3]
16. Elwyn G, Frosch D, Thomson R, et al. Shared decision making: a model for clinical practice. J Gen Intern Med 2012 Oct;27(10):1361-1367. [doi: 10.1007/s11606-012-2077-6] [Medline: 22618581]
17. Blandford A. HCI for health and wellbeing: challenges and opportunities. Int J Hum Comput Stud 2019 Nov;131:41-51. [doi: 10.1016/j.ijhcs.2019.06.007]
18. Mortier R, Haddadi H, Henderson T, McAuley D, Crowcroft J. Human-data interaction: the human face of the data-driven society. SSRN J 2014 Oct 1. [doi: 10.2139/ssrn.2508051]

XSL•FO
RenderX

19. Staehelin D, Dolata M, Stöckli L, Schwabe G. How patient-generated data enhance patient-provider communication in chronic care: field study in design science research. JMIR Med Inf 2024 Sep 10;12:e57406. [doi: 10.2196/57406] [Medline: 39255481]
20. Li I, Dey A, Forlizzi J. A stage-based model of personal informatics systems. Presented at: CHI ’10: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; Apr 10-15, 2010; Atlanta, GA p. 557-566. [doi: 10.1145/1753326.1753409]
21. Chung CF, Dew K, Cole A, et al. Boundary negotiating artifacts in personal informatics: patient-provider collaboration with patient-generated data. Presented at: 9th ACM Conference on Computer-Supported Cooperative Work (CSCW 2002); Feb 27 to Mar 2, 2016; San Francisco, CA p. 770-786. [doi: 10.1145/2818048.2819926]
22. Austin E, Lee JR, Amtmann D, et al. Use of patient-generated health data across healthcare settings: implications for health systems. JAMIA Open 2020 Apr 1;3(1):70-76. [doi: 10.1093/jamiaopen/ooz065]
23. Gimpel H, Nißen M, Görlitz R. Quantifying the quantified self: a study on the motivations of patients to track their own health. Presented at: ICIS 2013 Proceedings; Dec 15-18, 2013 URL: https://aisel.aisnet.org/icis2013/proceedings/HealthcareIS/3 [accessed 2026-07-30]
24. Shapiro M, Johnston D, Wald J, Mon D. Patient-generated health data. : RTI International; 2012 Apr URL: https://www.rti.org/publication/patient-generated-health-data-white-paper/fulltext.pdf [accessed 2026-07-30]
25. Guardado S, Karampela M, Isomursu M, Grundstrom C. Use of patient-generated health data from consumer-grade devices by health care professionals in the clinic: systematic review. J Med Internet Res 2024 May 31;26(1):e49320. [doi: 10.2196/49320]
26. Nittas V, Mütsch M, Ehrler F, Puhan MA. Electronic patient-generated health data to facilitate prevention and health promotion: a scoping review protocol. BMJ Open 2018 Aug 10;8(8):e021245. [doi: 10.1136/bmjopen-2017-021245] [Medline: 30099392]
27. European Health Data Space Regulation (EHDS). European Union. URL: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en [accessed 2026-07-30]
28. Digital health applications (DiGA). BfArM. URL: https://www.bfarm.de/EN/Medical-devices/Tasks/DiGA-and-DiPA/Digital-Health-Applications/_node.html [accessed 2026-07-30]
29. Personalised health and care 2020. GOV.UK. 2014 Nov 13. URL: https://www.gov.uk/government/publications/personalised-health-and-care-2020 [accessed 2026-07-30]
30. Chung CF, Cook J, Bales E, Zia J, Munson SA. More than telemonitoring: health provider use and nonuse of life-log data in irritable bowel syndrome and weight management. J Med Internet Res 2015 Aug 21;17(8):e203. [doi: 10.2196/jmir.4364] [Medline: 26297627]
31. Kim B, Ghasemi P, Stolee P, Lee J. Clinicians and older adults’ perceptions of the utility of patient-generated health data in caring for older adults: exploratory mixed methods study. JMIR Aging 2021 Nov 5;4(4):e29788. [doi: 10.2196/29788]
32. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
33. Ye J. The impact of electronic health record-integrated patient-generated health data on clinician burnout. J Am Med Inform Assoc 2021 Apr 23;28(5):1051-1056. [doi: 10.1093/jamia/ocab017] [Medline: 33822095]
34. Chung CF, Wang Q, Schroeder J, et al. Identifying and planning for individualized change: patient-provider collaboration using lightweight food diaries in healthy eating and irritable bowel syndrome. Proc ACM Interact Mobile Wearable Ubiquitous Technol 2019 Mar;3(1):7. [doi: 10.1145/3314394] [Medline: 31080941]
35. Figueiredo MC, Chen Y. Patient-generated health data: dimensions, challenges, and open questions. Found Trends Hum–Comput Interact 2020 Apr 22;13(3):165-297. [doi: 10.1561/1100000080]
36. Bjørn P, Kensing F. Special issue on information infrastructures for healthcare: the global and local relation. Int J Med Inf 2013 May;82(5):281-282. [doi: 10.1016/j.ijmedinf.2013.01.004] [Medline: 23422271]
37. Pine KH, Bossen C, Chen Y, et al. Data work in healthcare: challenges for patients, clinicians and administrators. Presented at: CSCW ’18 Companion: Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing; Nov 3-7, 2018; Jersey City, NJ p. 433-439. [doi: 10.1145/3272973.3273017]
38. Serban IB, Dritsa D, Jurado IC, et al. “I just see numbers, but how do you feel about your training?”: Clinicians’ data needs in telemonitoring for colorectal cancer surgery prehabilitation. Presented at: CSCW ’23 Companion: Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing; Oct 14-18, 2023. [doi: 10.1145/3584931.3607006]
39. Epstein DA, Ping A, Fogarty J, Munson SA. A lived informatics model of personal informatics. Presented at: UbiComp ’15: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 7-11, 2015; Osaka, Japan p. 731-742. [doi: 10.1145/2750858.2804250]
40. Rooksby J, Rost M, Morrison A, Chalmers M. Personal tracking as lived informatics. Presented at: CHI ’14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; Apr 26 to May 1, 2014. [doi: 10.1145/2556288.2557039]
41. Andersen TO, Bansler JP, Kensing F, et al. Aligning concerns in telecare: three concepts to guide the design of patient-centred e-Health. Comput Supported Coop Work 2018 Dec;27(3-6):1181-1214. [doi: 10.1007/s10606-018-9309-1]

XSL•FO
RenderX

42. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367. [doi: 10.2196/jmir.8775] [Medline: 29092808]
43. Bhat KS, Kumar N. Sociocultural dimensions of tracking health and taking care. Proc ACM Hum-Comput Interact 2020 Oct 14;4(CSCW2):1-24. [doi: 10.1145/3415200]
44. Caldeira C, Figueiredo MC, Dodakian L, de Souza CRB, Cramer SC, Chen Y. Towards supporting data-driven practices in stroke telerehabilitation technology. Proc ACM Hum-Comput Interact 2021 Apr 13;5(CSCW1):1-33. [doi: 10.1145/3449099]
45. Vaughn J, Kamkhoad D, Shaw RJ, Docherty SL, Subramaniam AP, Shah N. Seriously ill pediatric patient, parent, and clinician perspectives on visualizing symptom data. J Am Med Inf Assoc 2021 Jul 14;28(7):1518-1525. [doi: 10.1093/jamia/ocab037] [Medline: 33712836]
46. Piras EM, Miele F. Clinical self-tracking and monitoring technologies: negotiations in the ICT-mediated patient–provider relationship. Health Sociol Rev 2017 Jan 2;26(1):38-53. [doi: 10.1080/14461242.2016.1212316]
47. Schroeder J, Karkar R, Murinova N, Fogarty J, Munson SA, Munson SA. Examining opportunities for goal-directed self-tracking to support chronic condition management. Proc ACM Interact Mobile Wearable Ubiquitous Technol 2019 Dec;3(4):1-26. [doi: 10.1145/3369809] [Medline: 32656490]
48. West NEJ, Juneja M, Pinilla N, et al. Personalized vascular healthcare: insights from a large international survey. Eur Heart J Suppl 2022 Nov;24(Suppl H):H8-H17. [doi: 10.1093/eurheartjsupp/suac052] [Medline: 36382003]
49. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Dec;18(1):1-7. [doi: 10.1186/s12874-018-0611-x]
50. Nittas V, Lun P, Ehrler F, Puhan MA, Mütsch M. Electronic patient-generated health data to facilitate disease prevention and health promotion: scoping review. J Med Internet Res 2019 Oct 14;21(10):e13320. [doi: 10.2196/13320]
51. Philipp R, Barbara P, Gensichen J, Krcmar H. Insights on patient-generated health data in healthcare: a literature review. Presented at: PACIS 2022 Proceedings; Jul 5-9, 2022 URL: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1025&context=pacis2022 [accessed 2026-07-30]
52. Peters M, Tricco A, Munn Z, McInerney P, Khalil H, Godfrey C. Scoping reviews. In: Aromataris E, Jordan Z, Pilla B, Porritt K, Lockwood C, editors. JBI Manual for Evidence Synthesis: JBI; 2020. URL: https://jbi-global-wiki.refined.site/space/MANUAL/355862497/10.+Scoping+reviews [accessed 2026-08-05]
53. de-la-Fuente-Robles YM, Ricoy-Cano AJ, Albín-Rodríguez AP, López-Ruiz JL, Espinilla-Estévez M. Past, present and future of research on wearable technologies for healthcare: a bibliometric analysis using scopus. Sensors (Basel) 2022 Nov 8;22(22):8599. [doi: 10.3390/s22228599] [Medline: 36433195]
54. Policy for device software functions and mobile medical applications. : U.S. Food & Drug Administration; 2022 Sep URL: https://www.fda.gov/media/80958/download [accessed 2026-07-30]
55. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J Qual Methods 2006 Mar;5(1):80-92. [doi: 10.1177/160940690600500107]
56. Figueiredo MC, Su HI, Chen Y. Using data to approach the unknown: patients' and healthcare providers? Data practices in fertility challenges. Proc ACM Hum-Comput Interact 2021 Jan 5;4(CSCW3):1-35. [doi: 10.1145/3432926]
57. Bardram JE, Cramer-Petersen C, Maxhuni A, et al. DiaFocus: a personal health technology for adaptive assessment in long-term management of type 2 diabetes. ACM Trans Comput Healthcare 2023 Apr 30;4(2):1-43. [doi: 10.1145/3586579]
58. Wendrich K, Krabbenborg L. Digital self-monitoring of multiple sclerosis: interview study with dutch health care providers on the expected new configuration of roles and responsibilities. JMIR mHealth uHealth 2022;10(4):e30224. [doi: 10.2196/30224]
59. Seals A, Pilloni G, Kim J, et al. 'Are they doing better in the clinic or at home?': Understanding clinicians' needs when visualizing wearable sensor data used in remote gait assessments for people with multiple sclerosis. Presented at: CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems; Apr 30 to May 6, 2022; New Orleans, LA p. 1-16. [doi: 10.1145/3491102.3501989]
60. Jo E, Park S, Bang H, et al. GeniAuti: toward data-driven interventions to challenging behaviors of autistic children through caregivers' tracking. Proc ACM Hum-Comput Interact 2022 Mar 30;6(CSCW1):1-27. [doi: 10.1145/3512939]
61. Aarts T, Markopoulos P, Giling L, Vacaretu T, Pillen S. Snoozy: a chatbot-based sleep diary for children aged eight to twelve. Presented at: IDC '22: Proceedings of the 21st Annual ACM Interaction Design and Children Conference; Jun 27 to Jul 30, 2022. [doi: 10.1145/3501712.3529718]
62. White KM, Ivan A, Williams R, Galloway JB, Norton S, Matcham F. Remote measurement in rheumatoid arthritis: qualitative analysis of patient perspectives. JMIR Form Res 2021 Mar 9;5(3):e22473. [doi: 10.2196/22473] [Medline: 33687333]
63. Richter JG, Nannen C, Chehab G, et al. Mobile app-based documentation of patient-reported outcomes — 3-months results from a proof-of-concept study on modern rheumatology patient management. Arthritis Res Ther 2021 Apr 19;23(1):121. [doi: 10.1186/s13075-021-02500-3] [Medline: 33874994]
64. Rainey J, Verweij D, Dodds C, et al. Data contribution summaries for patient engagement in multi-device health monitoring research. Presented at: UbiComp/ISWC '21 Adjunct: Adjunct Proceedings of the 2021 ACM International Joint Conference

XSL•FO
RenderX

on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers; Sep 21-26, 2021. [doi: 10.1145/3460418.3479371]
65. Hoefer MJD, Van Kleunen L, Goodby C, Blackburn LB, Panati P, Voida S. The multiplicative patient and the clinical workflow: clinician perspectives on social interfaces for self-tracking and managing bipolar disorder. Presented at: DIS '21: Proceedings of the 2021 ACM Designing Interactive Systems Conference; Jun 28 to Jul 2, 2021. [doi: 10.1145/3461778.3461995]
66. Gupta A, Heng T, Shaw C, Gromala D, Leese J, Li L. Oh, I didn't do a good job: how objective data affects physiotherapist-patient conversations for arthritis patients. Presented at: PervasiveHealth '20: Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 18-20, 2020. [doi: 10.1145/3421937.3421991]
67. Zhu H, Moffa ZJ, Gui X, Carroll JM. Prehabilitation: care challenges and technological opportunities. Presented at: CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems; Apr 25-30, 2020. [doi: 10.1145/3313831.3376594]
68. Luo Y, Oh CY, Jean BS, Choe EK. Interrelationships between patients' data tracking practices, data sharing practices, and health literacy: onsite survey study. J Med Internet Res 2020 Dec 22;22(12):e18937. [doi: 10.2196/18937] [Medline: 33350960]
69. Jongsma KR, van den Heuvel JFM, Rake J, Bredenoord AL, Bekker MN. User experiences with and recommendations for mobile health technology for hypertensive disorders of pregnancy: mixed methods study. JMIR mHealth uHealth 2020 Aug 4;8(8):e17271. [doi: 10.2196/17271] [Medline: 32749225]
70. Alpert JM, Manini T, Roberts M, et al. Secondary care provider attitudes towards patient generated health data from smartwatches. NPJ Digit Med 2020;3(1):27. [doi: 10.1038/s41746-020-0236-4] [Medline: 32140569]
71. Tendedez H, Ferrario MA, McNaney R. The issue with that sort of data?": Clinicians' accountability concerns around COPD self-monitoring tools. Presented at: 2019 Computer Supported Cooperative Work and Social Computing Companion Publication (CSCW '19 Companion Proceedings); Nov 9-13, 2019. [doi: 10.1145/3311957.3359464]
72. Ng A, Kornfield R, Schueller SM, Zalta AK, Brennan M, Reddy M. Provider perspectives on integrating sensor-captured patient-generated data in mental health care. Proc ACM Hum-Comput Interact 2019 Nov 7;3(CSCW):1-25. [doi: 10.1145/3359217]
73. Iott B, Caverly T, Fishstrom A, King D, Meng G, Flynn A. Clinician perspectives on the user experience, configuration, and scope of use of a patient reported outcomes (PRO) dashboard. Presented at: PervasiveHealth'19: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 20-23, 2019; Trento, Italy p. 21-30. [doi: 10.1145/3329189.3329198]
74. Hue W, Jones S, Sengupta R. Exploring the future role of self-tracking data in the rheumatology clinic. Stud Health Technol Inf 2019;259(33–38):33-38. [Medline: 30923269]
75. Trauth E, Browning ER, Browning ER. Technologized talk: wearable technologies, patient agency, and medical communication in healthcare settings. Int J Sociotechnol Knowl Dev 2018 Jul 1;10(3):1-26. [doi: 10.4018/IJSKD.2018070101]
76. Tendedez H, McNaney R, Ferrario MA, Whittle J. Scoping the design space for data supported decision-making tools in respiratory care. Presented at: Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 21-24, 2018; New York, NY p. 217-226. [doi: 10.1145/3240925.3240983]
77. Reading M, Baik D, Beauchemin M, Hickey KT, Merrill JA. Factors influencing sustained engagement with ECG self-monitoring: perspectives from patients and health care providers. Appl Clin Inf 2018 Oct;9(4):772-781. [doi: 10.1055/s-0038-1672138] [Medline: 30304745]
78. Ng A, Reddy MC, Zalta AK, Schueller SM. Veterans' perspectives on Fitbit use in treatment for post-traumatic stress disorder: an interview study. JMIR Ment Health 2018 Jun 15;5(2):e10415. [doi: 10.2196/10415] [Medline: 29907556]
79. Mishra SR, Miller AD, Haldar S, et al. Supporting collaborative health tracking in the hospital. Presented at: CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems; Apr 21-26, 2018. [doi: 10.1145/3173574.3174224]
80. Gabriels K, Moerenhout T. Exploring entertainment medicine and professionalization of self-care: interview study among doctors on the potential effects of digital self-tracking. J Med Internet Res 2018 Jan 12;20(1):e10. [doi: 10.2196/jmir.8040] [Medline: 29330140]
81. Zhu H, Luo Y, Luo Y, Choe EK. Making space for the quality care: opportunities for technology in cognitive behavioral therapy for insomnia. Presented at: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems; May 6-11, 2017 p. 5773-5786. [doi: 10.1145/3025453.3025549]
82. Schroeder J, Hoffswell J, Chung CF, Fogarty J, Munson S, Zia J. Supporting patient-provider collaboration to identify individual triggers using food and symptom journals. Presented at: CSCW '17: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing; Feb 25 to Mar 1, 2017; Portland, OR p. 1726-1739. [doi: 10.1145/2998181.2998276]
83. Raj S, Newman MW, Lee JM, Ackerman MS. Understanding individual and collaborative problem-solving with patient-generated data. Proc ACM Hum-Comput Interact 2017 Dec 6;1(CSCW):1-18. [doi: 10.1145/3134723]

XSL•FO
RenderX

84. Malu M, Findlater L. Sharing automatically tracked activity data: implications for therapists and people with mobility impairments. Presented at: PervasiveHealth '17: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 23-26, 2017; Barcelona, Spain p. 136-145. [doi: 10.1145/3154862.3154864]
85. Kim Y, Heo E, Lee H, et al. Prescribing 10,000 steps like aspirin: designing a novel interface for data-driven medical consultations. Presented at: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems; May 6-11, 2017; Denver, CO p. 5787-5799. [doi: 10.1145/3025453.3025570]
86. Kelley C, Lee B, Wilcox L. Self-tracking for mental wellness: understanding expert perspectives and student experiences. Presented at: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems; May 6-11, 2017. [doi: 10.1145/3025453.3025750]
87. West P, Giordano R, Van Kleek M, Shadbolt N. The quantified patient in the doctor's office: challenges & opportunities. Presented at: CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems; May 7-12, 2016. [doi: 10.1145/2858036.2858445]
88. Kim Y, Ji S, Lee HJ, et al. "My doctor is keeping an eye on me!": exploring the clinical applicability of a mobile food logger. Presented at: CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose, CA p. 5620-5631. [doi: 10.1145/2858036.2858145]
89. Cohen DJ, Keller SR, Hayes GR, Dorr DA, Ash JS, Sittig DF. Integrating patient-generated health data into clinical care settings or clinical decision-making: lessons learned from Project HealthDesign. JMIR Hum Factors 2016 Oct 19;3(2):e26. [doi: 10.2196/humanfactors.5919] [Medline: 27760726]
90. Ryokai K, Michahelles F, Kritzler M, Syed S. Communicating and interpreting wearable sensor data with health coaches. Presented at: 9th International Conference on Pervasive Computing Technologies for Healthcare; May 20-23, 2015; Istanbul, Turkey p. 221-224. [doi: 10.4108/icst.pervasivehealth.2015.259055]
91. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. "You get reminded you're a sick person": personal data tracking and patients with multiple chronic conditions. J Med Internet Res 2015 Aug 19;17(8):e202. [doi: 10.2196/jmir.4209]
92. Nundy S, Lu CYE, Hogan P, Mishra A, Peek ME. Using patient-generated health data from mobile technologies for diabetes self-management support: provider perspectives from an academic medical center. J Diabetes Sci Technol 2014 Jan;8(1):74-82. [doi: 10.1177/1932296813511727] [Medline: 24876541]
93. Wu DTY, Xin C, Bindhu S, et al. Clinician perspectives and design implications in using patient-generated health data to improve mental health practices: mixed methods study. JMIR Form Res 2020 Aug 7;4(8):e18123. [doi: 10.2196/18123] [Medline: 32763884]
94. Boudioni M, McLaren S, Lister G. Patient empowerment: its implementation and systems within hospitals in England and Greece. Health Serv Manage Res 2018 Nov;31(4):180-194. [doi: 10.1177/0951484817752628] [Medline: 29376412]
95. WHO global strategy on people-centred and integrated health services: interim report. : World Health Organization; 2015 URL: https://www.afro.who.int/sites/default/files/2017-07/who-global-strategy-on-pcihs-main-document_final.pdf [accessed 2026-07-30]
96. Continuity and coordination of care: a practice brief to support implementation of the WHO framework on integrated people-centred health services. : World Health Organization; 2018 Nov 7 URL: https://iris.who.int/server/api/core/bitstreams/7c938c63-5570-4c0e-95f2-d310bb647704/content [accessed 2026-07-30]
97. Plans to ensure adequate and qualified workforce for community care services for ageing population. Ministry of Health Singapore. URL: https://www.moh.gov.sg/newsroom/plans-to-ensure-adequate-and-qualified-workforce-for-community-care-services-for-ageing-population/ [accessed 2026-08-04]
98. Dinkel A, Jahnen M. Patientenberichtete endpunkte – die bedeutung der subjektiven patientenperspektive für forschung und klinische versorgung [Article in German]. Urologie 2024 Sep;63(9):886-892. [doi: 10.1007/s00120-024-02405-4]
99. Position on the European Health Data Space. : Standing Committee of European Doctors (CPME); 2022 Oct 29 URL: https://www.cpme.eu/api/documents/adopted/2022/11/cpme.2022-065.FINAL.CPME.position.EHDS.pdf [accessed 2026-07-30]
100. Höppchen I, Kulnik ST, Meschtscherjakov A, et al. "Be with me and stay with me": insights from co-designing a digital companion to support patients transitioning from hospital to cardiac rehabilitation. Presented at: Proceedings of the 2024 ACM Designing Interactive Systems Conference; Jul 1-5, 2024. [doi: 10.1145/3643834.3661633]

Abbreviations

EC: exclusion criterion
EHR: electronic health record
HCP: health care professional
HIPAA : Health Insurance Portability and Accountability Act
IC: inclusion criterion
P4: predictive, preventive, personalized, and participatory
PGHD: patient-generated health data

XSL•FO
RenderX

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
SDM: shared decision-making

Edited by A Stone, L Buis; submitted 02.Oct.2025; peer-reviewed by M Ekwueme, O AKINLOLU, R Kurmashev, T Yusuff; revised version received 24.Jun.2026; accepted 24.Jun.2026; published 10.Aug.2026.

Please cite as:
V S Pakianathan P, Kumar D, Prabath J, Hussein R, Niebauer J, Schmidt A, Smeddinck J
Barriers and Enablers to Integrating Patient-Generated Health Data in Shared Decision-Making From Health Care Professional and Patient Perspectives: Scoping Review
JMIR Mhealth Uhealth 2026;14:e85197
URL: https://mhealth.jmir.org/2026/1/e85197
doi:10.2196/85197

© Pavithren V S Pakianathan, Devender Kumar, Jayathissa Prabath, Rada Hussein, Josef Niebauer, Albrecht Schmidt, Jan Smeddinck. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Patient-Facing AI-Enabled Digital Health Technologies and Quality of Life in Cancer: Systematic Review and Exploratory Meta-Analysis

Anica Ilic[1], PhD; Lene Kristine Juvet[2,3], PhD; Patrick Cairns[1], PhD; Kristen Elizabeth Thompson Thornton[1], PhD; Hanne Cathrine Lie[1], PhD

[1]Department of Behavioural Medicine, Institute of Basic Medical Sciences, Faculty of Medicine, University of Oslo, Postboks 1111 Blindern, Oslo, Norway

[2]Division of Infection Control, Norwegian Institute of Public Health, Oslo, Norway

[3]Department of Nursing and Health Sciences, Faculty of Health and Social Sciences, University of South-Eastern Norway, Borre, Norway

Corresponding Author:
Anica Ilic, PhD
Department of Behavioural Medicine, Institute of Basic Medical Sciences, Faculty of Medicine, University of Oslo, Postboks 1111 Blindern, Oslo, Norway

Abstract

Background: Cancer affects multiple physical, psychological, and social aspects of an individual's life. Cancer survivors frequently report unmet needs long after diagnosis and require ongoing support. AI is increasingly embedded in patient-facing digital health technologies (DHTs) in oncology, yet its impact on different domains of patients' and survivors' health-related quality of life (HRQOL) remains unclear.

Objective: This systematic review aims to (1) examine how AI has been integrated into patient-facing DHTs designed to support cancer survivors, (2) narratively synthesize the potential effects of these technologies on HRQOL and provide preliminary quantitative estimates through an exploratory meta-analysis, and (3) explore broader changes in additional patient-reported outcomes (PROs; secondary aim).

Methods: PubMed, PsycINFO, Embase, Scopus, CINAHL, and the Cochrane Library were searched for articles published between January 2020 and August 2025. Reference lists of included articles were hand-searched for additional eligible studies. Eligible studies enrolled cancer survivors of any age and disease stage, evaluated a patient-facing DHT with AI components, and assessed HRQOL. Nonoriginal research and non-English reports were excluded. Risk of bias was assessed in all controlled studies using RoB 2 (revised Cochrane risk of bias 2) or ROBINS-I V2 (Risk of Bias in Non-Randomized Studies—of Interventions, Version 2), according to study design. Data on HRQOL and other PROs were synthesized narratively, and exploratory random-effects meta-analyses were conducted for HRQOL domains.

Results: Eight reports from 7 studies from China and the United States (N=2867 participants) met the inclusion criteria, and 3 (n=292 participants) contributed to the exploratory meta-analysis. All studies included adults with various cancers at different stages and times since diagnosis. Most studies showed low risk of bias or some concerns (RoB 2), but one was evaluated as having a serious risk of bias (ROBINS-I V2). AI applications ranged from symptom monitoring to targeted education. The narrative synthesis suggested positive effects on selected HRQOL domains, particularly general health, with more pronounced effects in studies conducted in China. Exploratory meta-analyses demonstrated provisional moderate positive effects on global health (Hedges g=0.77, 95% CI 0.15 - 1.40) and social functioning (Hedges g=0.75, 95% CI 0.08 - 1.42), but no effects on physical functioning, role functioning, or emotional well-being. Other PROs indicated generally high user satisfaction and adherence, improved mental health outcomes, and reductions in physical symptoms. Only minor and mild adverse events were reported.

Conclusions: Current evidence, although limited, suggests that AI-enabled patient-facing DHTs may benefit survivors' HRQOL and other PROs, particularly in early survivorship. However, our findings are based on small and heterogeneous studies and should therefore be interpreted with caution. Robust trials with adequate sample sizes, longer follow-up, and appropriate control conditions, including DHTs without AI components, are needed to determine the specific contribution of AI.

Trial Registration: PROSPERO CRD420251021466; https://www.crd.york.ac.uk/PROSPERO/view/CRD420251021466

(JMIR Mhealth Uhealth 2026;14:e94793) doi:10.2196/94793

KEYWORDS

neoplasms; AI; digital health; quality of life; patient-reported outcome measures

XSL•FO
RenderX

Introduction

Cancer affects not only physical health but also the emotional, social, and psychological well-being of patients and their close networks across the cancer trajectory [1]. Cancer survivors often face symptoms, lifestyle adjustments, and report information needs that persist long after cancer treatment [2]. Although cancer survivors value continued personal follow-up care [2,3], it is rarely achievable in routine practice due to limited health care resources, competing priorities, geographic barriers, and the growing number of individuals living with and beyond cancer [4]. These challenges highlight the need for scalable solutions that can complement traditional survivorship care.

Patient-facing digital health technologies (DHTs), such as mobile apps and digital therapeutics, may help bridge these gaps by delivering accessible, personalized, and timely support to patients in diverse settings [5]. Unlike DHTs designed for health care professionals, patient-facing DHTs are directly operated by patients themselves, enabling self-monitoring, symptom reporting, psychoeducation, and behavioral interventions in daily life [5]. Incorporating AI components into DHTs might further enhance their potential impact [6,7]. For instance, machine learning (ML) and natural language processing (NLP) can facilitate the processing and interpretation of large volumes of heterogeneous data, including symptom reports, sensor data, and medical records [8]. These technologies can thereby enable patient-facing DHTs to provide tailored feedback, adapt digital interventions to individual needs, or identify moments when users are most receptive to receiving an intervention [8,9]. However, these potential benefits should be weighed against important concerns. Embedding AI into patient-facing DHTs may introduce additional risks to data privacy and exacerbate existing gaps in digital access [10], and consequently in access to health care. Furthermore, understanding how patients perceive and use these tools is necessary, as their experience will strongly influence engagement and uptake [11].

Although interest in patient-facing DHTs is increasing, few trials examine their effects from the patients' perspective [12]. Patient-reported outcomes (PROs) are critical indicators of how individuals perceive their health, functioning, and quality of life, as well as the impact of interventions on their daily lives [13]. PROs capture dimensions of health that cannot be observed or measured by clinicians alone and are particularly relevant in chronic and oncological conditions where maintaining quality of life is a primary treatment goal [14]. Health-related quality of life (HRQOL) represents a key multidimensional construct spanning physical, psychological, and social domains [15,16]. Various instruments exist to assess HRQOL, ranging from generic measures to cancer-specific scales, each assessing different facets of survivors' lived experiences [15,17]. However, it remains unclear whether AI-enabled DHTs adequately address the complex physical, emotional, and social needs of cancer survivors [18]. To address this gap, this systematic review aims to (1) examine how AI has been integrated into patient-facing DHTs designed to support cancer survivors, (2) narratively synthesize the potential effects of these technologies on HRQOL and provide preliminary quantitative estimates through an exploratory meta-analysis, and (3) explore broader changes in additional PROs (secondary aim).

Methods

This systematic review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines (Checklist 1) [19]. The protocol, including a detailed description of the search strategy, is registered in the PROSPERO systematic review registry (CRD420251021466).

Database Search

The search strategy consisted of 4 concept blocks (cancer, AI, DHT, and HRQOL) and combined free-text and thesaurus terms using AND/OR Boolean operators. Initial terms were identified by the first author (AI) through scoping searches and key articles in the field. The strategy was refined iteratively in consultation with a co-author (HCL) to capture relevant synonyms, acronyms, and spelling variants, and it was reviewed by an experienced librarian. Search strings were adapted for each database, and we focused on articles containing any of the terms in the title, abstract, and, where possible, the keywords. No further filters or language restrictions were applied at the search stage.

We conducted an initial search on May 20, 2025, in PubMed, PsycINFO, Embase, Scopus, CINAHL, and the Cochrane Library, covering records published from January 1, 2020, to the search date. The start date (January 1, 2020) was chosen to account for the rapid development of AI and to ensure the inclusion of the most recent evidence [20,21]. A second search was conducted in the same databases on August 26, 2025, adding 3 terms that had not been included previously. Therefore, we obtained a large number of duplicates. The reference lists of all included publications were screened to identify additional potentially relevant studies.

Eligibility Criteria

Reports were eligible if they (1) enrolled cancer survivors at any disease stage, including those undergoing active treatment or palliative care (in accordance with the definition of the National Cancer Institute) [22], (2) evaluated patient-facing DHTs that explicitly included AI components (AI-enabled DHTs), and (3) reported an assessment of HRQOL. For the purposes of this review, AI-enabled DHTs refer to patient-facing DHTs that incorporate an AI component, whether embedded within the platform or operating as an external module.

Nonoriginal reports, including other reviews, editorials, commentaries, conference abstracts, and study protocols, were excluded. Only studies that included a control group (eg, usual care and non–AI-enabled DHTs) were eligible for inclusion in the meta-analysis.

Although we initially considered including non-English articles relying on software-generated translations, these were ultimately excluded because we could not validate them with a native speaker and therefore could not avoid potential misinterpretation. However, these studies are cited in the results for completeness.

XSL•FO
RenderX

Study Selection

All identified references were screened using Covidence (Veritas Health Innovation Ltd) [23], and duplicates were automatically excluded. Two authors (AI and PC or KETT) independently screened all titles and abstracts against previously agreed-on inclusion and exclusion criteria. The same 2 authors screened the full texts of all articles previously included by at least one author to make a final decision on inclusion. Any disagreements were resolved through discussion or by consulting a third author. The study selection process is summarized in a PRISMA 2020 flow diagram.

Data Extraction

Data extraction was supported using SciSpace [24] and Perplexity [25]. All machine-extracted information was verified against the original manuscripts to ensure that no relevant information was misinterpreted or missed (AI) and validated by a second author (PC or KETT). Data were extracted from the published articles and any supplementary materials, and no individual participant data were requested. Where information was unclear or insufficiently detailed, we contacted the authors for clarification. However, no responses were received.

Collected data included (1) study information: title, authors, year, journal, country, and study design; (2) sample characteristics: sample size, sociodemographic information, and cancer-related characteristics for both intervention and control groups; (3) DHT details: type, purpose, AI components, input data sources, and output type; (4) HRQOL outcomes: baseline and postintervention measurements, instruments used, domains assessed, and main findings; and (5) additional PROs: instruments used and key results. When multiple articles reported on the same study, all relevant publications were considered to provide a comprehensive description of the methods and outcomes.

Risk of Bias

For randomized controlled trials (RCTs), the risk of bias was assessed using the revised Cochrane Risk of Bias 2 (RoB 2) tool [26], and for non-RCTs, the ROBINS-I V2 (Risk of Bias in Non-Randomized Studies—of Interventions, Version 2) [27]. Single-arm studies were not formally assessed for risk of bias and were reported narratively without inclusion in any quantitative synthesis.

Two authors (AI and LKJ) independently evaluated each study across all domains, with possible ratings of low risk, some concerns, or high risk (RoB 2), or low, moderate, serious, and critical concerns (ROBINS-I V2). The evaluations were then compared, and any discrepancies were resolved through discussion. When multiple articles reported on the same study, all relevant publications were considered to provide a comprehensive assessment.

Data Synthesis

For the purposes of this systematic review, studies were grouped as controlled (including randomized and nonrandomized trials) or single-arm studies. Controlled studies were eligible for quantitative synthesis (meta-analysis), whereas single-arm studies were only summarized narratively.

Narrative Synthesis

HRQOL outcomes were summarized using a narrative synthesis approach (AI) [28]. Given the small number of studies included, thematic clustering was not feasible. Instead, we presented results on a study-by-study basis, describing the reported direction and highlighting any patterns across interventions where possible. Similarly, we used a descriptive narrative approach to describe PROs beyond HRQOL.

Meta-Analysis

Because HRQOL instruments varied across studies, we extracted all domains reported at all time points and compared the underlying items to align conceptually similar domains across different instruments (Multimedia Appendix 1). We pooled outcomes only for domains that were conceptually comparable across instruments. For all instruments, higher scores indicated higher HRQOL. When studies reported multiple follow-up time points, we included the assessment closest to the end of the intervention to minimize heterogeneity in exposure duration.

For controlled studies, we conducted exploratory meta-analyses using random-effects models due to the clinical and methodological heterogeneity across digital health interventions and populations with cancer. For continuous HRQOL outcomes, provisional effect sizes were calculated as standardized mean differences (Hedges *g*) with 95% CIs. When authors reported SEs, we converted them to SDs to ensure comparability across studies. When the reported variability measure was not specified as SD or SE, it was assumed to represent SD, as SD is more commonly reported in HRQOL research. When studies reported both intention-to-treat and per-protocol analyses, we used data from the intention-to-treat analyses.

Statistical heterogeneity was assessed using the I^2 statistic. Sensitivity analysis was performed using a leave-one-out approach. Results of individual studies and pooled effect estimates were visually displayed using forest plots. Study characteristics and extracted outcome data were summarized using tables. All analyses were performed using the *meta* command suite in Stata (version 9.5; StataCorp LLC) [29].

Results

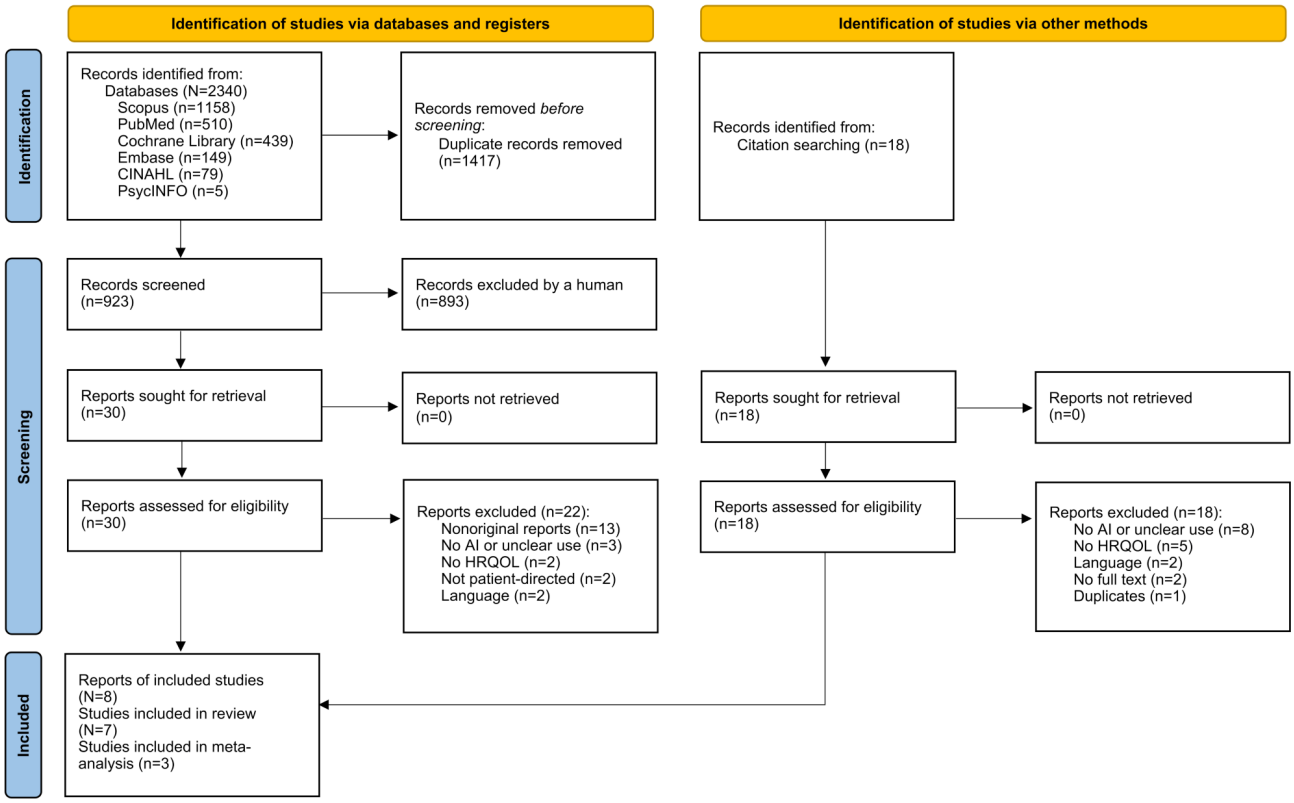
Study Selection

A comprehensive search of 6 scientific databases identified 2340 records (Figure 1). After removing duplicates (n=1417), 923 records were screened. Of these, 30 full-text articles were assessed for eligibility, and 8 publications from 7 studies (N=2867 participants) met the inclusion criteria and were included in the review.

Two potentially relevant records published in Chinese were excluded because reliable translation and validation were not feasible. One was a pilot study from an already included study [30], and the other reported data from a distinct study not otherwise included in this review [31]. Citation searching did not identify any additional eligible studies.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram. HRQOL: health-related quality of life. Source: [19].



Study Characteristics

The 7 included studies comprised 2 single-arm studies [32,33], 3 RCTs [34-37], 1 partial-crossover RCT [38], and 1 non-RCT [39] (Table 1).

The studies covered multiple cancer types and survivorship stages (Table 1). RCTs by Schmitz et al [38] in the United States and by Hu et al [34] and Jiang et al [35] in China focused exclusively on breast cancer survivors, while a US single-arm study by Horesh et al [33] included both breast and ovarian cancer survivors. Together, these studies targeted survivorship stages from active treatment through early survivorship to metastatic disease. The RCT by Li et al [37] examined early-stage postoperative lung cancer survivors in China, and the RCT by Kamdar et al [36] focused on patients with metastatic solid tumors receiving palliative care in the United States. The remaining studies included survivors of any cancer type in different contexts. The non-RCT by Zhang et al [39] focused on patients receiving chemotherapy in China, and the single-arm study by Buchan et al [32] included individuals at any stage of disease in the United States. Quality of life was assessed in all included studies as a primary or secondary outcome, and most studies used validated HRQOL instruments.

XSL•FO
RenderX

Table . Study characteristics[a].

Study	Study characteristics	Assessment details	Sample[b]	Cancer details	HRQOL[c] measure
Buchan et al [32], 2024	United States; single-arm cohort study	Within-group comparison; monthly assessments (T0, 49.1 mo; median follow-up 8.8 mo)	2423	Mixed cancers; any stage	Not specified
Horesh et al [33], 2022	United States; single-arm pre-post study	Within-group comparison; 2 assessments (T0, 24 d)	37	Breast and ovarian cancers; on active treatment, experiencing hot flashes	WHOQOL-BREF[d]
Hu et al [34], 2025; Jiang et al [35], 2024[j]	China; RCT[e]	General information online; 3 assessments (T0, 1 mo, 3 mo)	115 (intervention=60 and control=55)	Breast cancer; early survivorship	FACT-B[f]
Kamdar et al [36], 2024	United States; RCT	Usual care; 3 assessments (T0, 4 wk, 8 wk)	79 (intervention=39 and control=40)	Mixed (solid) cancers; metastatic, receiving palliative care	FACT-G[g]
Li et al [37], 2025	China; RCT	Usual care; 2 assessments (T0, 5 mo)	36 (intervention=20 and control=16)	Lung cancer (NSCLC)[h]; early survivorship, postoperative	EORTC QLQ-C30[i]
Schmitz et al [38], 2023[j]	United States; RCT, partial crossover	Delayed intervention; 3 assessments (T0, 3 mo, 6 mo[k])	33 (intervention=17 and control=16)	Breast cancer; metastatic	SF-36[l]
Zhang et al [39], 2025[j]	China, non-RCT	Non-RCT; usual care; 2 assessments (T0, 6 mo)	144 (intervention=72 and control=72)	Mixed cancers; receiving chemotherapy	EORTC QLQ-C30

[a]In total, 8 publications from 7 studies were included in this review.

[b]Sample sizes reported in this table reflect the number of participants included in the HRQOL analyses (including intervention and control groups). They may differ from the sample sizes reported elsewhere in the publications.

[c]HRQOL: health-related quality of life.

[d]WHOQOL-BREF: World Health Organization Quality of Life Scale–Brief version.

[e]RCT: randomized controlled trial.

[f]FACT-B: Functional Assessment of Cancer Therapy–Breast.

[g]FACT-G: Functional Assessment of Cancer Therapy–General.

[h]NSCLC: non–small cell lung cancer.

[i]EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire–Core 30.

[j]Included in the meta-analysis.

[k]Immediate group: intervention 0 - 6 months; delayed group: 3 - 6 months; control assessed 0 - 3 months prior to intervention.

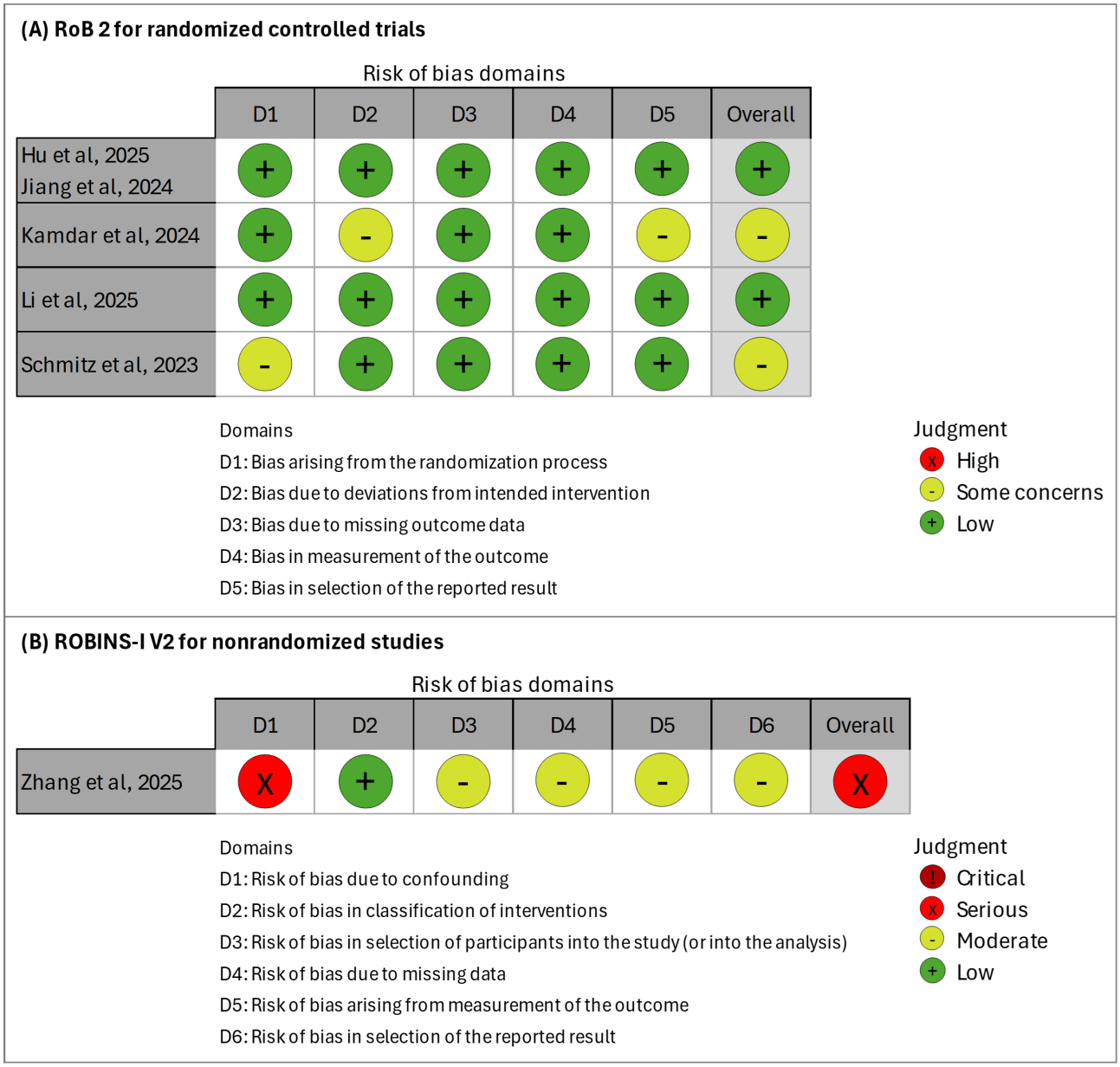
[l]SF-36: 36-item Short-Form Health Survey.

Risk of Bias

Of the 4 RCTs, 2 were rated as low risk of bias, and 2 studies were assessed as having some concerns. However, all concerns were observed in different domains and did not indicate a consistent source of bias across the evidence base. The non-RCT by Zhang et al [39] was found to be at serious risk of bias, primarily due to inadequate control of confounding factors in a time-separated convenience sample. There were additional concerns regarding the unclear management of exclusions and missing data. Furthermore, inconsistencies in the reporting of results suggest caution when interpreting the findings. No studies were evaluated as high (RoB 2) or critical (ROBINS-I V2) risk of bias. Therefore, none were excluded. Further details on the risk of bias evaluation are available in Figure 2.

XSL•FO
RenderX

Figure 2. Risk of bias assessment: (A) RoB 2 (Revised Cochrane risk of bias 2) for randomized controlled trials, (B) ROBINS-I V2 (Risk of Bias in Non-Randomized Studies—of Interventions, Version 2) for nonrandomized studies [34-39].



Aim 1: AI-Enabled DHTs for Patient Support in Cancer

Purpose

Most of the DHTs included in this review aimed to provide home-based support for monitoring and managing physical symptoms associated with cancer treatment, such as pain, nausea, fatigue, and sleep disturbances [33-36,38,39]. Three studies addressed psychosocial needs by offering supportive care [33-35,38]. Two studies specifically targeted cardio-oncology rehabilitation [37] or nutritional support for cancer survivors [32] (Table 2).

XSL•FO
RenderX

Table . Description of digital health technologies.

Study	Name	Purpose
Buchan et al [32], 2024	Ina	AI-driven virtual text message–based nutrition assistant providing evidence-based, personalized nutritional counseling, recipes, and symptom management support to improve dietary habits, symptom control, and HRQOL[a] for diverse cancer types.
Horesh et al [33], 2022	Bubble	AI-VR[b] intervention, delivered via mobile apps (Bubble and Luna) and VR goggles, with elements of cognitive behavioral therapy and mindfulness-based stress reduction, to reduce hot flash frequency and severity, alleviate psychological distress, improve sleep quality, and enhance HRQOL in women with breast and ovarian cancers. The intervention takes place in the "Frosty" virtual environment and is led by an avatar named Luna.
Hu et al [34], 2025; Jiang et al [35], 2024	AI-TA	AI-driven intelligent interactive mobile app designed to reduce physical, psychological, and social premature frailty, thereby improving HRQOL among young breast cancer survivors through personalized content delivery, symptom tracking, and 2-way web-based follow-up with health care professionals.
Kamdar et al [36], 2024	ePAL	Digital therapeutic mobile app designed to manage cancer-related pain in patients with advanced cancer through active pain monitoring, an AI algorithm that triages patient symptoms and identifies specific barriers to pain control, patient education modules, and real-time clinician outreach to address identified barriers to pain management.
Li et al [37], 2025	Recovery Plus Health (Recovery Plus USA Inc)	Cardio-oncology rehabilitation program delivered via the Recovery Plus mobile app using AI-driven exercise prescriptions with real-time heart rate monitoring to improve cardiopulmonary fitness, reduce cancer-related symptoms (fatigue, anxiety, and daytime dysfunction), and enhance HRQOL in early-stage lung cancer survivors.
Schmitz et al [38], 2023	Nurse AMIE	Conversational AI–based supportive care intervention delivered via Amazon Echo Show with Alexa to address symptoms (pain, fatigue, sleep disturbance, and distress), provide nutritional guidance, and improve HRQOL in patients with metastatic breast cancer.
Zhang et al [39], 2025	Not specified	Deep learning platform for symptom management designed to provide continuous, personalized symptom management for patients with cancer undergoing chemotherapy, aiming to reduce anxiety and depression and improve HRQOL by delivering real-time nursing support, symptom monitoring, and medical guidance outside hospital visits.

[a]HRQOL: health-related quality of life.

[b]VR: virtual reality.

Delivery Modality

Of the 4 interventions delivered via mobile applications, 2 were application-only [34-36], and 2 were delivered in combination with wearable devices such as virtual reality (VR) goggles [33] or heart rate bands [37]. The remaining interventions were delivered through web platforms or portals accessible on both computers and smartphones [39], via text messaging [32], or through a voice-based virtual assistant [38].

XSL•FO
RenderX

Six interventions included a human component alongside the DHTs. Three interventions encouraged participants who reported high symptoms to contact the health care team [32,37,38], and 2 automatically notified the health care team when any alarming symptoms were observed [36,38]. In 2 interventions, health care professionals supervised AI-driven recommendations and adjusted them when necessary [32,37]. Two DHTs also encouraged questions and peer group discussions [34,35,39].

AI Components

Across the interventions, AI was mainly used to personalize and optimize supportive care delivery. NLP was used for text analysis of patient input and for generating personalized responses, enabling the development of conversational interfaces [32,35,37]. ML and predictive analytics were used to predict patient needs, triage symptoms, and adapt interventions in real time [32,35,36,39]. Knowledge graphs and expert systems provided structured decision support, ensuring evidence-based recommendations [32,35,36]. Recommender systems combined keyword analysis and filtering to personalize the delivered content [33-35]. Other DHTs included AI-driven exercise prescription systems that adjusted training parameters based on wearable data [37], as well as speech recognition for voice-interactive virtual assistants [38].

Input Sources and Output Types

The included studies can be classified into 3 groups. First, several interventions in breast cancer used patient-reported symptoms, survey data, and basic clinical history to provide personalized education, symptom-management guidance, and supportive follow-up [33-35,38]. Second, studies focused on more intensive symptom management in people with advanced cancer or those receiving palliative care used repeated symptom reporting, clinical records, and real-time monitoring to trigger tailored coaching, clinician outreach, or escalation [36]. Third, studies centered on nutrition or broader supportive care used baseline medical and nutrition history, treatment details, and user input to deliver individualized advice, recipes, nursing guidance, and interactive messaging [32,37,39].

Aim 2: Effect on HRQOL

Multimedia Appendix 2 presents the magnitude and direction of changes for all presented outcomes.

Narrative Synthesis of HRQOL Findings

Across 5 controlled studies, the 3 studies conducted in China reported improvements in at least one HRQOL domain, while the 2 US studies found no significant changes. Li et al [37] found that the intervention group showed improved global health status compared with usual care at 5 months postintervention, although no between-group differences were observed across subdomains. Zhang et al [39] reported significant improvements across nearly all HRQOL domains in the intervention group, including overall health status, role functioning, cognitive functioning, emotional functioning, and social functioning. Physical functioning (sometimes referred to by the authors as “somatic function”) was the only domain that did not differ significantly between groups at 6 months postintervention [39]. Jiang et al [35] reported that although some HRQOL domains did not change significantly at 1-month postdiagnosis, all domains, including overall HRQOL, physical, social or family, emotional, and functional well-being, as well as breast cancer-specific concerns, showed significant between-group improvements by 3 months. Hu et al [34], reporting on the same study population, similarly found that total HRQOL improved significantly more in the intervention group than in the control group over this period. In the RCTs conducted by Kamdar et al [36] and Schmitz et al [38] in the United States, no significant effects of the intervention on HRQOL were observed.

The 2 single-arm studies conducted in the United States showed similar results. Buchan et al [32] reported that 81% of users experienced some improvement in their HRQOL after using the Ina virtual dietitian. Horesh et al [33] observed within-group improvements in physical and psychological domains 24 days after the intervention, with no changes in social or environmental domains.

Quantitative Synthesis of HRQOL Findings

From the 7 identified studies, 3 (n=292) contributed to the exploratory meta-analysis examining the effect of AI-enabled DHTs on HRQOL [35,38,39]. Studies were excluded from the meta-analysis for the following reasons: single-arm studies (n=2) [32,33], lack of quantitative HRQOL data (n=1) [36], reporting HRQOL change scores without corresponding pre- and postintervention values (n=1) [37], and multiple publications from the same study (n=1) [34]. For studies with incomplete HRQOL data, we contacted the authors but received no response. In cases of duplicate publications, the report providing the most comprehensive HRQOL data was used [35]. The 3 studies included in the meta-analysis were judged to have low [34,35], some concerns [38], and serious risk of bias [39], primarily due to potential confounding.

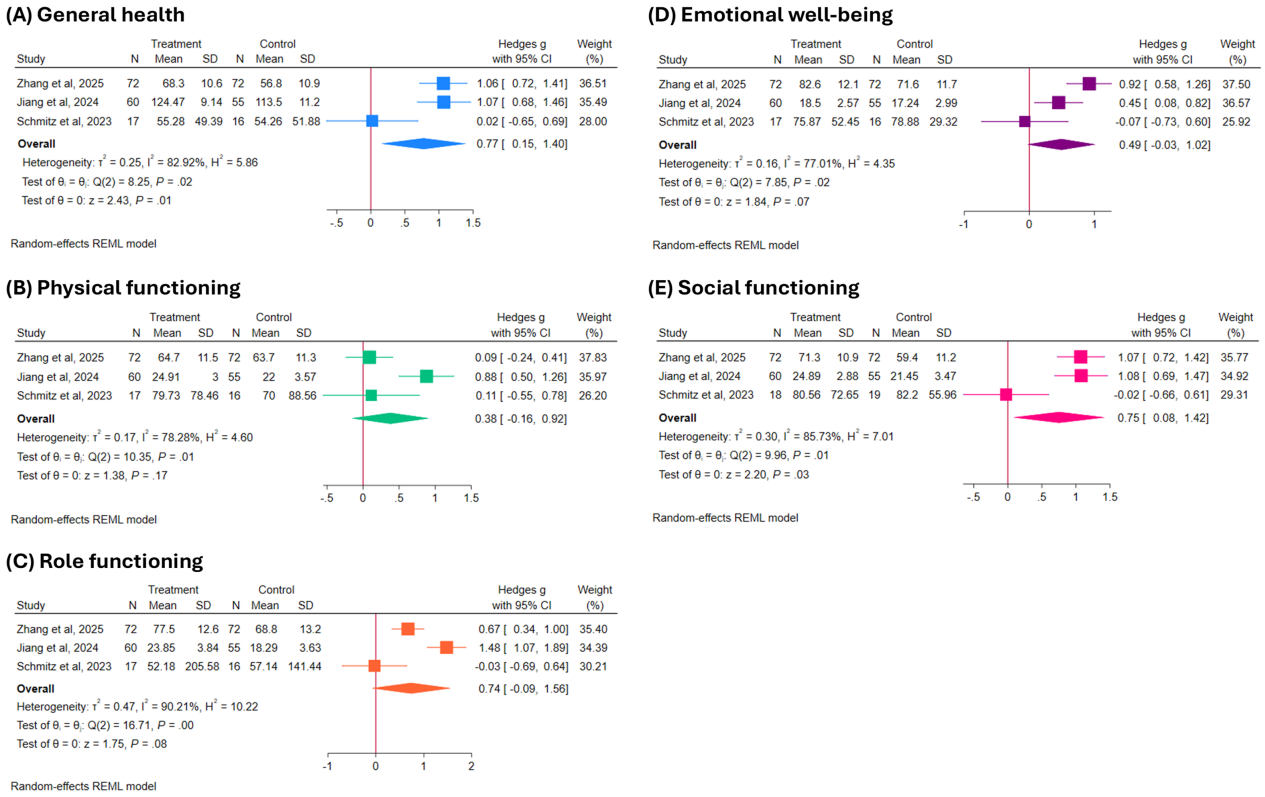
Five HRQOL domains were analyzed: general health, physical functioning, role functioning, emotional well-being, and social functioning. In the exploratory meta-analyses, statistically significant pooled effects were found for general health (Hedges g=0.77, 95% CI 0.15 to 1.40; P=.01; I^2=82.9%) and social functioning (Hedges g=0.75, 95% CI 0.08 to 1.42; P=.03; I^2=85.7%), both indicating moderate effect sizes. However, the high heterogeneity across these analyses warrants cautious interpretation. No statistically significant effects were found for physical functioning, role functioning, or emotional well-being (see Figure 3 for estimates).

Leave-one-out sensitivity analyses indicated that the pooled effect was sensitive to individual studies for 4 of the 5 HRQOL domains. Specifically, excluding the US study by Schmitz et al [38] increased effect sizes for general health, role functioning, emotional well-being, and social functioning, while physical functioning remained stable across all exclusion scenarios (full results are provided in Multimedia Appendix 3). Excluding the studies by Zhang et al [39] or Jiang et al [35] did not change the statistical significance of the pooled effects for any domain.

Due to the small number of included studies, other subgroup analyses, as well as formal tests for publication bias (Egger regression) and visual inspection of funnel plots, were not feasible [26].

XSL•FO
RenderX

Figure 3. Forest plots of pooled mean differences in various health-related quality of life domains: (A) general health, (B) physical functioning, (C) role functioning, (D) emotional well-being, and (E) social functioning [35,38,39]. REML: restricted maximum likelihood.



Aim 3: Changes in Other PROs

In addition to HRQOL, the included studies reported other PROs, including user satisfaction and adherence, mental health, physical symptoms, additional PROs, and adverse effects.

User Satisfaction and Adherence

Five studies reported outcomes related to satisfaction or acceptability [32,33,36,38,39]. Across these studies, users consistently rated the DHTs positively, with high scores for ease of use, clarity, and overall experience. Satisfaction levels generally exceeded 70%, and usability assessments indicated strong acceptability of the DHTs. Most users found the content helpful and the DHTs easy to navigate. However, in Zhang et al [39], a small proportion reported difficulties with navigation, internet access, or understanding specific content.

Adherence and engagement with the DHTs varied across studies. Overall, most users engaged with the DHTs at least occasionally, though the frequency and consistency of engagement differed by intervention type and activity. High adherence rates were reported in 2 interventions targeting early survivorship in China, with 88% to 97% of participants completing prescribed sessions or following recommended activities [34,37]. Engagement with DHTs targeting patients with advanced cancer in the United States declined gradually over time, with weekly use dropping from initially high rates to lower levels by later weeks [36,38]. Furthermore, specific features within single DHTs were used at varying rates. For example, most participants used nutritional tips, whereas fewer used the suggested recipes in the virtual dietitian intervention by Buchan et al [32].

Mental Health

Across the included studies, 4 of 6 interventions that reported mental health outcomes showed improvements. The US single-arm hot flash intervention study reported significant reductions in perceived stress, general psychological distress, and illness perceptions from pre- to postintervention [33]. Another study targeting women with breast cancer in China showed greater decreases in psychological frailty, psychological distress, symptom frequency, and global distress in the intervention group compared with controls on follow-up [34,35]. However, the US study by Schmitz et al [38] showed no significant between-group differences in distress among metastatic breast cancer survivors. Additional evidence indicated mixed effects for anxiety. Two controlled studies from China found decreases in anxiety. While Zhang et al [39] reported greater improvement in the control group than in the intervention group, Li et al [37] found the opposite. In addition, the RCT in US patients with metastatic cancer by Kamdar et al [36] found that anxiety worsened in the intervention group compared with the control group. Depression was assessed only in the study by Zhang et al [39] involving patients with mixed cancer undergoing chemotherapy in China, which found greater improvements in the intervention group than in the control group.

Physical Symptoms

Across the included studies, several DHTs showed improvements in physical symptoms. The virtual dietitian intervention by Buchan et al [32] reported high perceived usefulness for symptom management, with approximately one-third of users showing reductions in both symptom count

XSL•FO
RenderX

and total symptom severity over time. Significant reductions in hot flash interference and frequency, along with improvements in overall sleep quality, sleep onset latency, and subjective sleep quality were observed in a US hot flash intervention [33]. Similarly, a controlled study to manage premature frailty in young breast cancer survivors from China showed greater reductions in fatigue and physical symptoms, and an increase in physical activity in the intervention group, although group-by-time effects were not significant [34]. The RCT by Kamdar et al [36] found significant reductions in pain severity over 8 weeks, with a higher proportion of participants achieving substantial pain improvement compared with the control group. Additional benefits were reported in fatigue and daytime dysfunction, following the digital cardio-oncology rehabilitation program by Li et al [37]. One controlled study of US metastatic breast cancer survivors found no significant between-group differences in pain, sleep, and fatigue [38].

Other PROs

Other PROs were examined in the studies by Hu et al [34] and Jiang et al [35], which evaluated social support and self-efficacy following a digital humanistic program to manage premature frailty in young breast cancer survivors. Social support increased over time in both the intervention and control groups, with greater improvement observed in the intervention group [34,35]. In addition, Jiang et al [35] reported an increase in self-efficacy among participants receiving the intervention.

Adverse Effects

Across the included studies, adverse events were occasionally reported and generally mild. A small proportion of participants experienced cybersickness in the AI-VR intervention by Horesh et al [33] for the management of hot flashes. In the premature frailty trial conducted in China, adverse changes were uncommon, with only a few participants showing increases in symptom levels or declines in HRQOL [34]. A US intervention for pain management in patients with advanced cancer showed a small but statistically significant increase in anxiety symptoms in the intervention group, although values remained within the mild range and below clinically significant thresholds [36]. No adverse effects were identified in 2 studies [37,38], and the remaining manuscripts did not report adverse events [32,35,39].

Discussion

Principal Findings

This systematic review synthesized evidence on available patient-facing AI-enabled DHTs and their effect on HRQOL and broader PROs in cancer populations. Eight publications from 7 studies conducted in China and the United States evaluated a range of mobile- and web-based interventions, most of which aimed to support at-home symptom management. AI was used to personalize recommendations (eg, exercise, nutrition, and symptom triage), adapt delivery formats (eg, VR and voice interfaces), and tailor educational content. Overall, the effects on HRQOL were mixed. Results from the narrative synthesis showed that among the 5 controlled studies, all 3 RCTs conducted in China found significant improvements across multiple HRQOL domains, including global health status, role, cognitive, emotional, social, and functional well-being, while neither of the 2 US RCTs found significant between-group effects. The 2 single-arm US studies suggested within-group improvements, although the absence of control groups limits the interpretability of those findings. An exploratory meta-analysis of 5 HRQOL domains showed modest improvements only in general health and social functioning, but no effect on physical functioning, role functioning, and emotional well-being. However, these results should be interpreted with caution given the small number of contributing studies, limited sample sizes, heterogeneity, and the presence of risk-of-bias concerns in individual studies. Other PROs, including mental health, physical symptoms, self-efficacy, and social support, also showed mixed but generally encouraging results. Mental health outcomes were variable, with some DHTs reducing distress or anxiety and others showing no effect or small, but not clinically significant, increases in anxiety. Despite considerable heterogeneity in both interventions and populations, most DHTs were associated with improvements in key physical symptoms, including pain, fatigue, and sleep. User adoption and satisfaction were generally high, and only mild adverse effects were reported occasionally.

Several patterns might help explain the variation in effects across studies and guide the design of future AI-enabled DHTs for cancer survivorship care. Improvements in HRQOL and other PROs were more frequent in the trials conducted in China, particularly for general health, whereas effects were less consistent across specific HRQOL domains. Two trials conducted in the United States showed no benefits in HRQOL, while 2 observational studies from the United States lacked control groups. Methodological factors, including small samples, heterogeneous measures, and short follow-up periods, likely contributed to these differences. In digital oncology research, changes in HRQOL might require sustained use of DHTs over time, as physical and psychological symptoms develop gradually [40,41]. Trials with short follow-up periods may, therefore, have lacked sufficient time to detect change, even when adherence was high. Another explanation might be differences in perceptions of AI-enabled DHTs and in-person consultations across cultures [42]. These may influence how patients interpret symptom feedback, interact with automated recommendations, and integrate digital tools into their routines. Furthermore, cross-national research suggests that social desirability bias varies across cultures, with relatively higher levels observed in China [43]. This may have led to more favorable evaluations of AI-enabled DHTs, especially in clinician-administered surveys. However, these cultural explanations remain speculative and cannot be distinguished from other country-level differences.

Our findings should be interpreted in the context of heterogeneous definitions of cancer survivorship across studies, which translate into substantial differences in symptom burden and care needs. While some interventions targeted patients with metastatic or advanced disease receiving ongoing or palliative treatment, others focused on disease-free survivors after curative therapy. Consistent with this heterogeneity, trials reporting no improvements in HRQOL tended to include patients with metastatic disease. Previous studies have shown that, among

XSL•FO
RenderX

patients with high symptom burden in palliative care settings, supportive DHTs alone are often insufficient to improve overall HRQOL, although they may alleviate specific physical symptoms [44]. Our results have partially confirmed these findings. While AI-enabled tools may offer more personalized and proactive support than traditional DHTs, their benefits may be limited in advanced disease settings. This suggests that AI-enabled interventions should be integrated with comprehensive supportive care to improve HRQOL in advanced cancer.

Studies showing positive effects on HRQOL and PROs often included human involvement. This aligns with broader evidence that digital tools can initiate but rarely sustain long-term engagement among populations with complex needs without human involvement [45]. Hybrid models that combine AI-enabled monitoring and personalization with clinician or nurse follow-up leverage the strengths of both technology and human support. Thus, they can better translate digital engagement into behavior change and symptom improvement. Interventions that supported question-asking and peer discussion also reported positive effects on PROs. Features that enable communication with clinicians or connect patients with peers may increase self-efficacy, normalize experiences, and reduce isolation—mechanisms that are widely recognized in cancer survivorship research [46]. However, AI components themselves may introduce bias. For example, NLP or large language model–based systems that classify patient input or triage symptoms can introduce systematic measurement shifts, which could affect clinical decision support [47,48]. These findings highlight the value of designing comprehensive DHTs as platforms for holistic support.

User satisfaction and engagement were high across studies, indicating strong acceptability of DHTs containing AI components in cancer survivorship care. Notably, the effects observed on HRQOL and other PROs are similar to those reported for previous digital health interventions without AI [41,46]. However, because none of the studies included equivalent DHTs without AI components, we cannot determine whether these effects are attributable to the AI functionality, other DHT features, or nonspecific effects. Therefore, integrating AI might not significantly alter patients' experiences, either by enhancing outcomes or introducing new barriers. Instead, the effectiveness of DHTs on PROs seems to depend primarily on design quality and relevance of support rather than on whether the underlying system relies on AI. Future research should focus on larger, multisite, longitudinal trials, more diverse populations with cancer, the use of traditional DHTs without AI components for control purposes, and the identification of different engagement trajectories among users. In practice, AI-enabled DHTs may support HRQOL when paired with adequate human follow-up, peer-support features, and careful patient selection based on disease stage, symptom burden, and digital readiness.

Limitations

Our findings should be interpreted in light of several limitations. First, clinical heterogeneity limits the validity of pooled estimates. The included studies covered diverse types of cancer and stages of survivorship, making a single pooled estimate for cancer survivors problematic. Furthermore, the 3 different HRQOL instruments used across studies introduce measurement heterogeneity. Readers should, therefore, interpret findings from the exploratory meta-analyses with caution.

Second, the small number of eligible studies and participants limits the statistical power of the meta-analyses. While the narrative synthesis provides additional insights, the results from the meta-analyses should be considered exploratory.

Third, methodological limitations across studies reduce comparability and generalizability and may inflate observed effects. These limitations include small sample sizes, lack of control groups or randomization, short follow-up periods, and reporting limitations. None of the studies included equivalent DHTs without AI components, so it is not possible to determine whether the observed benefits are due to AI functionality, other DHT features, or nonspecific effects.

Fourth, geographic confounding limits generalizability. The limited geographical diversity (primarily studies from China vs US) prevents the isolation of intervention effects from country-level factors such as health care systems, cultural context, or implementation characteristics.

Fifth, how AI was defined and operationalized varied across studies, ranging from rule-based decision support to ML algorithms. This complicates direct comparisons across interventions and limits the interpretability of pooled findings. Furthermore, although the search strategy was intentionally broad and included multiple AI-related terms, studies that did not adequately describe the technical components of their DHTs may have been missed. Future reviews would benefit from applying a prespecified AI classification framework.

Finally, research on AI-enabled DHT is advancing rapidly. The numerous protocols identified during the screening process suggest that more publications are forthcoming. A living systematic review approach may, therefore, be valuable for maintaining an up-to-date evidence base.

Conclusions

This review provides a timely synthesis of evidence on AI-enabled DHTs for cancer care and establishes a conceptual and methodological foundation for future work. Although effects on HRQOL were mixed, some interventions showed benefits across HRQOL domains and other PROs. Importantly, as none of the studies used control groups with equivalent DHTs without AI components, the observed benefits cannot be attributed specifically to the AI components. Our findings are preliminary, and it is too early to conclude whether patient-facing AI-enabled DHTs improve cancer survivorship care. Future research should employ active control conditions to isolate the contribution of AI features specifically.

XSL•FO
RenderX

Acknowledgments

The authors acknowledge the use of generative AI tools, including ChatGPT, Perplexity (Claude, Sonnet 4.6), and Grammarly, for proofreading, editing, and summarizing text. All authors have reviewed the manuscript and take full responsibility for its content. The authors thank Marte Ødegaard for her assistance with the database search.

Funding

AI was supported by a Postdoc.Mobility fellowship from the Swiss National Science Foundation (SNSF; grant 225508). The SNSF had no role in the study design, data collection, analysis, interpretation, or manuscript preparation.

Data Availability

The data used in this systematic review were extracted from published studies. Data extraction tables can be provided by the corresponding author upon justified request.

Authors' Contributions

Conceptualization: AI
Data curation: AI
Formal analysis: AI, LKJ
Funding acquisition: AI
Investigation: AI, PC, KETT
Methodology: AI, LKJ
Project administration: AI
Supervision: HCL
Validation: LKJ, PC, KETT
Visualization: AI
Writing – original draft: AI
Writing – review & editing: AI, LKJ, PC, KETT, HCL

Conflicts of Interest

None declared.

Multimedia Appendix 1
Health-related quality of life domain comparison.
[DOCX File, 24 KB - mhealth_v14i1e94793_app1.docx]

Multimedia Appendix 2
Changes in patient-reported outcomes.
[DOCX File, 70 KB - mhealth_v14i1e94793_app2.docx]

Multimedia Appendix 3
Leave-one-out sensitivity analysis.
[DOCX File, 22 KB - mhealth_v14i1e94793_app3.docx]

Checklist 1
PRISMA 2020 checklist.
[PDF File, 162 KB - mhealth_v14i1e94793_app4.pdf]

References

1. Wang Y, Feng W. Cancer-related psychosocial challenges. Gen Psychiatr 2022;35(5):e100871. [doi: 10.1136/gpsych-2022-100871] [Medline: 36311374]
2. Evans Webb M, Murray E, Younger ZW, Goodfellow H, Ross J. The supportive care needs of cancer patients: a systematic review. J Cancer Educ 2021 Oct;36(5):899-908. [doi: 10.1007/s13187-020-01941-9] [Medline: 33492650]
3. Murchie P, Norwood PF, Pietrucin-Materek M, Porteous T, Hannaford PC, Ryan M. Determining cancer survivors' preferences to inform new models of follow-up care. Br J Cancer 2016 Dec 6;115(12):1495-1503. [doi: 10.1038/bjc.2016.352] [Medline: 27802453]

XSL•FO
RenderX

4. Love M, Debay M, Hudley AC, et al. Cancer survivors, oncology, and primary care perspectives on survivorship care: an integrative review. J Prim Care Community Health 2022;13:21501319221105248. [doi: 10.1177/21501319221105248] [Medline: 35678264]
5. Health Advances. Guidance to industry: classification of digital health technologies. : Digital Therapeutics Alliance; 2023 URL: https://dtxalliance.org/wp-content/uploads/2023/06/Guidance-to-Industry-Classification-of-Digital-Health-Technologies-2023Jun05.pdf [accessed 2026-06-04]
6. Silcox C, Zimlichmann E, Huber K, et al. The potential for artificial intelligence to transform healthcare: perspectives from international health leaders. NPJ Digit Med 2024 Apr 9;7(1):88. [doi: 10.1038/s41746-024-01097-6] [Medline: 38594477]
7. Fogel AL, Kvedar JC. Artificial intelligence powers digital medicine. NPJ Digit Med 2018;1(1):5. [doi: 10.1038/s41746-017-0012-2] [Medline: 31304291]
8. Graili P, Farhoudi B. The intersection of digital health and artificial intelligence: clearing the cloud of uncertainty. Digit Health 2025;11:20552076251315621. [doi: 10.1177/20552076251315621] [Medline: 39866886]
9. Keller R, Wangenheim F, Mair J, Kowatsch T. Chapter 6: receptivity to mobile health interventions. In: Jacobson N, Kowatsch T, Marsch L, editors. Digital Therapeutics for Mental Health and Addiction: Academic Press; 2023:65-77. [doi: 10.1016/B978-0-323-90045-4.00006-X]
10. Gerke S, Minssen T, Cohen G. Chapter 12: ethical and legal challenges of artificial intelligence–driven healthcare. In: Bohr A, Memarzadeh K, editors. Artificial Intelligence in Healthcare: Academic Press; 2020:295-336. [doi: 10.1016/B978-0-12-818438-7.00012-5]
11. Kay-Lambkin F, Heinsch M, Sampson D. Chapter 14: potential pitfalls and lessons learned. In: Jacobson N, Kowatsch T, Marsch L, editors. Digital Therapeutics for Mental Health and Addiction: Academic Press; 2023:177-189. [doi: 10.1016/B978-0-323-90045-4.00013-7]
12. Madanian S, Nakarada-Kordic I, Reay S, Chetty T. Patients' perspectives on digital health tools. PEC Innov 2023 Dec;2:100171. [doi: 10.1016/j.pecinn.2023.100171] [Medline: 37384154]
13. Definition of patient-reported outcome—NCI dictionary of cancer terms. National Cancer Institute. 2011. URL: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/patient-reported-outcome [accessed 2025-11-24]
14. Aiyegbusi OL, Nair D, Peipert JD, Schick-Makaroff K, Mucsi I. A narrative review of current evidence supporting the implementation of electronic patient-reported outcome measures in the management of chronic diseases. Ther Adv Chronic Dis 2021;12:20406223211015958. [doi: 10.1177/20406223211015958] [Medline: 34104376]
15. Hays RD, Reeve BB. Measurement and modeling of health-related quality of life. In: Heggenhougen HK, editor. International Encyclopedia of Public Health: Academic Press; 2008:241-252. [doi: 10.1016/B978-012373960-5.00336-1]
16. de Wit M, Hajos T. Health-related quality of life. In: Gellman MD, Turner JR, editors. Encyclopedia of Behavioral Medicine: Springer; 2013:929-931. [doi: 10.1007/978-1-4419-1005-9_753]
17. Kaplan RM, Hays RD. Health-related quality of life measurement in public health. Annu Rev Public Health 2022 Apr 5;43:355-373. [doi: 10.1146/annurev-publhealth-052120-012811] [Medline: 34882431]
18. Han R, Acosta JN, Shakeri Z, Ioannidis JPA, Topol EJ, Rajpurkar P. Randomised controlled trials evaluating artificial intelligence in clinical practice: a scoping review. Lancet Digit Health 2024 May;6(5):e367-e373. [doi: 10.1016/S2589-7500(24)00047-5] [Medline: 38670745]
19. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
20. Bohr A, Memarzadeh K. Chapter 2: the rise of artificial intelligence in healthcare applications. In: Bohr A, Memarzadeh K, editors. Artificial Intelligence in Healthcare: Academic Press; 2020:25-60. [doi: 10.1016/B978-0-12-818438-7.00002-2]
21. Pan LC, Wu XR, Lu Y, et al. Artificial intelligence empowered digital health technologies in cancer survivorship care: a scoping review. Asia Pac J Oncol Nurs 2022 Dec;9(12):100127. [doi: 10.1016/j.apjon.2022.100127] [Medline: 36176267]
22. Definitions. National Cancer Institute—Division of Cancer Control & Population Sciences. URL: https://cancercontrol.cancer.gov/ocs/definitions [accessed 2025-03-26]
23. Covidence. URL: https://www.covidence.org/ [accessed 2026-07-09]
24. SciSpace. URL: https://scispace.com/ [accessed 2026-07-09]
25. Perplexity AI. URL: https://www.perplexity.ai [accessed 2026-07-09]
26. Higgins JPT, Thomas J, Chandler J, et al. Cochrane Handbook for Systematic Reviews of Interventions Version 65: Cochrane; 2024. URL: https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current [accessed 2026-07-09]
27. ROBINS-i V2. Risk of Bias Tools. 2025. URL: https://www.riskofbias.info/welcome/robins-i-v2 [accessed 2026-07-09]
28. Campbell M, McKenzie JE, Sowden A, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ 2020 Jan 16;368:l6890. [doi: 10.1136/bmj.l6890] [Medline: 31948937]
29. Stata Statistical Software: Release 19. StataCorp. 2025. URL: https://www.stata.com/new-in-stata/ [accessed 2026-07-15]
30. Jiang L, Wang X, Xu J, Wu Y, Hu Y. Construction of an intelligent interactive nursing information support system and its application in young breast cancer survivors [Article in Chinese]. Chin J Nurs 2023;58(6):654-661 [FREE Full text]
31. Li M, Zhang H, Xia C, et al. Application practice of AI empowering post-discharge specialized disease management in postoperative rehabilitation of the lung cancer patients undergoing surgery [Article in Chinese]. Zhongguo Fei Ai Za Zhi 2025 Mar 20;28(3):176-182. [doi: 10.3779/j.issn.1009-3419.2025.102.11] [Medline: 40210477]

XSL•FO
RenderX

32. Buchan ML, Goel K, Schneider CK, Steullet V, Bratton S, Basch E. National implementation of an artificial intelligence–based virtual dietitian for patients with cancer. JCO Clin Cancer Inform 2024 Jun;8:e2400085. [doi: 10.1200/CCI.24.00085] [Medline: 38832697]
33. Horesh D, Kohavi S, Shilony-Nalaboff L, et al. Virtual reality combined with artificial intelligence (VR-AI) reduces hot flashes and improves psychological well-being in women with breast and ovarian cancer: a pilot study. Healthcare (Basel) 2022 Nov 11;10(11):2261. [doi: 10.3390/healthcare10112261] [Medline: 36421585]
34. Hu Y, Wiley J, Jiang L, et al. Digital humanistic program to manage premature frailty in young breast cancer survivors with gender perspective. NPJ Digit Med 2025 Jan 16;8(1):35. [doi: 10.1038/s41746-025-01439-y] [Medline: 39820349]
35. Jiang L, Xu J, Wu Y, Liu Y, Wang X, Hu Y. Effects of the "AI-TA" mobile app with intelligent design on psychological and related symptoms of young survivors of breast cancer: randomized controlled trial. JMIR Mhealth Uhealth 2024 Jun 4;12:e50783. [doi: 10.2196/50783] [Medline: 38833298]
36. Kamdar M, Jethwani K, Centi AJ, et al. A digital therapeutic application (ePAL) to manage pain in patients with advanced cancer: a randomized controlled trial. J Pain Symptom Manage 2024 Sep;68(3):261-271. [doi: 10.1016/j.jpainsymman.2024.05.033] [Medline: 38866116]
37. Li G, Zhou X, Deng J, et al. Digital therapeutics–based cardio-oncology rehabilitation for lung cancer survivors: randomized controlled trial. JMIR Mhealth Uhealth 2025 Feb 25;13:e60115. [doi: 10.2196/60115] [Medline: 39999435]
38. Schmitz KH, Kanski B, Gordon B, et al. Technology-based supportive care for metastatic breast cancer patients. Support Care Cancer 2023 Jun 20;31(7):401. [doi: 10.1007/s00520-023-07884-3] [Medline: 37338627]
39. Zhang J, Lv XN, Wang M, Zhang J, Qi F. Continuous nursing symptom management in cancer chemotherapy patients using deep learning. Sci Rep 2025;15(1):7990. [doi: 10.1038/s41598-025-92762-7] [Medline: 40055437]
40. Perry MB, Taylor S, Khatoon B, et al. Examining the effectiveness of electronic patient-reported outcomes in people with cancer: systematic review and meta-analysis. J Med Internet Res 2024 Jul 31;26:e49089. [doi: 10.2196/49089] [Medline: 39083791]
41. Buneviciene I, Mekary RA, Smith TR, Onnela JP, Bunevicius A. Can mHealth interventions improve quality of life of cancer patients? A systematic review and meta-analysis. Crit Rev Oncol Hematol 2021 Jan;157:103123. [doi: 10.1016/j.critrevonc.2020.103123] [Medline: 33190065]
42. Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. Lancet Digit Health 2021 Sep;3(9):e599-e611. [doi: 10.1016/S2589-7500(21)00132-1] [Medline: 34446266]
43. Bernardi RA. Associations between Hofstede's cultural constructs and social desirability response bias. J Bus Ethics 2006 Apr;65(1):43-53. [doi: 10.1007/s10551-005-5353-0]
44. Zhao B, Zhang S, Zhang T, Chen Y, Zhang C. Effect of applying digital health in palliative care for patients with advanced cancer: a meta-analysis and systematic review. Support Care Cancer 2024 Sep 17;32(10):664. [doi: 10.1007/s00520-024-08866-9] [Medline: 39297984]
45. Loughnane C, Laiti J, O'Donovan R, Dunne PJ. Systematic review exploring human, AI, and hybrid health coaching in digital health interventions: trends, engagement, and lifestyle outcomes. Front Digit Health 2025;7:1536416. [doi: 10.3389/fdgth.2025.1536416] [Medline: 40343215]
46. Bertrand A, Escot N, Girodet M, et al. Digital tools used to enhance the quality of life and well-being of adolescent and young adult cancer survivors: a systematic review of the literature. J Adolesc Young Adult Oncol 2026 Jun;15(3):165-170. [doi: 10.1177/21565333251377393] [Medline: 40938393]
47. Cross JL, Choma MA, Onofrey JA. Bias in medical AI: implications for clinical decision-making. PLOS Digit Health 2024 Nov;3(11):e0000651. [doi: 10.1371/journal.pdig.0000651] [Medline: 39509461]
48. Xu S, Wang Z, Wang H, Ding Z, Zou Y, Cao Y. LLM-based annotation and token-augmented modeling for emotional tone classification in online cancer peer-support posts. PLOS Digit Health 2026 May;5(5):e0001235. [doi: 10.1371/journal.pdig.0001235] [Medline: 42213728]

Abbreviations

DHT: digital health technology
HRQOL: health-related quality of life
ML: machine learning
NLP: natural language processing
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRO: patient-reported outcome
RCT: randomized controlled trial
RoB 2: revised Cochrane risk of bias 2
ROBINS-I V2: Risk of Bias in Non-Randomized Studies—of Interventions, Version 2
VR: virtual reality

XSL•FO
RenderX

Edited by Z Ni; submitted 06.Mar.2026; peer-reviewed by M Jani, Y Cao, Y Hu; revised version received 13.Jun.2026; accepted 16.Jun.2026; published 10.Aug.2026.

Please cite as:
Ilic A, Juvet LK, Cairns P, Thornton KET, Lie HC
Patient-Facing AI-Enabled Digital Health Technologies and Quality of Life in Cancer: Systematic Review and Exploratory Meta-Analysis
JMIR Mhealth Uhealth 2026;14:e94793
URL: https://mhealth.jmir.org/2026/1/e94793
doi:10.2196/94793

© Anica Ilic, Lene Kristine Juvet, Patrick Cairns, Kristen Elizabeth Thompson Thornton, Hanne Cathrine Lie. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Conceptualization From the Sensors to Suicide-Related Outcomes: Scoping Review Based on Layered Hierarchical Sensemaking Framework

Sohee Kim[1,2], BSN, RN; Jinyeong Kim[1], MSN, RN; Wai Tong Chien[3], BN (Hons), MPhil, PhD; Tzu Tsun Luk[4], RN, PhD; Yu Zhang[5,6], MSN, RN; Zhen Yang Abel Tan[4], BSN, RN; Eunju Park[1,2], MSN, RN; Heejung Kim[1,7,8,9], RN, PhD

[1]Department of Nursing, College of Nursing, Yonsei University, 50-1 Yonsei-ro, Seoul, Republic of Korea

[2]Brain Korea 21 FOUR Project, College of Nursing, Yonsei University, Seoul, Republic of Korea

[3]The Nethersole School of Nursing, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

[4]Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[5]School of Nursing, Peking University, Beijing, China

[6]Division of Sleep Medicine, Peking University People's Hospital, Beijing, China

[7]Yonsei Evidence Based Nursing Centre of Korea: A JBI Affiliated Group, Yonsei University, Seoul, Republic of Korea

[8]Mo-Im Kim Nursing Research Institute, Yonsei University, Seoul, Republic of Korea

[9]Institute for Innovation in Digital Healthcare, Yonsei University, Seoul, Republic of Korea

Corresponding Author:
Heejung Kim, RN, PhD
Department of Nursing, College of Nursing, Yonsei University, 50-1 Yonsei-ro, Seoul, Republic of Korea

Abstract

Background: Suicide is a leading cause of preventable mortality worldwide, with more than 700,000 deaths annually. Although suicidal ideation fluctuates rapidly, conventional risk assessments rely on retrospective self-report collected infrequently, and the detection of short-term suicide risk remains limited. Passive digital sensing using smartphones and wearable devices enables continuous monitoring of behavioral and physiological signals associated with suicide-related outcomes. However, current evidence remains fragmented, without a clear framework for translation into clinically interpretable risk indicators.

Objective: This scoping review synthesized and mapped passive digital markers associated with suicide-related outcomes via the layered hierarchical sensemaking framework (LHSF), which structures information from raw sensor data to high-level behavioral markers. We aimed to illustrate a clinically interpretable mapping of digital markers for suicide-specific digital phenotyping.

Methods: Following Arksey and O'Malley and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines, this scoping review was conducted using the population-concept-context framework (population: not restricted; concept: passively collected digital data from smartphones or wearable devices; and context: suicide-related outcomes). PubMed, CINAHL, PsycINFO, and IEEE Xplore were searched for studies published between 2015 and 2025. Studies were included if they (1) collected passive digital data from smartphones or wearable devices, and (2) measured suicide-related outcomes. Narrative mapping was conducted using LHSF to distinguish between low-level features (ie, measurable properties extracted from sensors) and high-level behavioral markers (ie, clinically meaningful constructs interpreted from low-level features).

Results: Of 626 studies identified, 14 (2.2%) met inclusion criteria. Six (42.9%) used predictive modeling, and 8 (57.1%) conducted correlational analyses. Among predictive studies (area under the curve [AUC]=0.56-0.89), a lower heart rate variability predicted an elevated suicide risk in 1 study (AUC=0.89). Of correlational studies, 7 (87.5%) of 8 reported at least one significant association between passive sensor data and suicide-related outcomes. Mapped to the LHSF, low-level features spanned 7 domains, linked to high-level markers, such as autonomic dysregulation, sleep disturbance, social withdrawal, smartphone use patterns, and suicide-related expression. Physiological indicators of autonomic regulation were associated with suicide-related outcomes in all 4 studies examining them and achieved the highest predictive performance (AUC=0.89). Smartphone use metrics were significantly associated in both studies, whereas linguistic (2/3 studies, 66.7%) and location-based features (2/2 studies, 100%) were associated with at least one outcome, with nonsignificant findings for some indicators or studies. Sleep parameters and movement intensity showed few significant associations.

XSL•FO
RenderX

Conclusions: Physiological indicators were associated with suicide-related outcomes across all relevant studies and showed the highest predictive performance (AUC=0.89), followed by smartphone-derived behavioral features. Linguistic and location-based features showed mixed associations, whereas sleep- and activity-related indicators showed few significant associations. Future research should prioritize multimodal data integration, algorithmic refinement, and external validation to strengthen clinical utility in digital suicide phenotyping based on the LHSF.

(JMIR Mhealth Uhealth 2026;14:e98616) doi:10.2196/98616

KEYWORDS

suicide; passive sensing; digital phenotyping; smartphones; wearable devices; scoping review; suicide-related outcomes

Introduction

Suicide is a leading cause of preventable mortality worldwide, with more than 700,000 individuals dying by suicide annually, accounting for approximately 1% of all deaths globally [1,2]. Recognizing its substantial public health burden, the United Nations has included a Sustainable Development Goals (SDG)–specific target (SDG 3.4.2) to reduce the national suicide mortality rate by one-third by 2030. Interviews and self-reported questionnaires are frequently used to evaluate suicide risk, such as suicidal ideation, plans, and past behaviors via standardized instruments [3,4]. However, these measures are typically administered at a single clinical encounter or at intervals and rely on individuals' recall of experiences over extended periods (eg, the past week or month) [5]. In practice, suicidal thoughts and related risk processes are highly dynamic, with meaningful fluctuations occurring over short timescales. Coppersmith et al [6] found that episodes of elevated suicidal ideation often lasted only 1 to 3 hours on average. An ecological momentary assessment (EMA) study in psychiatric inpatients with depression reported that momentary hopelessness and perceived burdensomeness significantly predicted suicidal ideation at the next assessment [7]. EMA studies indicate that only high-frequency, in situ assessments can capture short-term risk processes [8,9]. Consequently, conventional suicide risk assessments based on retrospective recall and infrequent administration often fail to detect brief but clinically critical risk elevations [8,9].

Researchers have increasingly used digital data as a complementary source of behavioral and psychological information to advance early identification and intervention. Recent advances in information and communication technologies have enabled passive digital sensing via smartphones and wearable devices, which has become a promising tool for continuous monitoring of mental health–related physiological and behavioral changes [10]. These devices capture real-world data on physiological reactions, physical functions, and daily activities, and enable assessments beyond clinical settings with ecological validity. This has fueled digital phenotyping research aiming to infer mental health states from patterns in digital traces [11]. Despite increasing digital sensing research on suicide risk, studies report heterogeneous digital indicators, including physiological (eg, heart rate variability [HRV]), behavioral (eg, physical movement), and smartphone-derived contextual features (eg, text language use).

However, this recent evidence remains inconsistent, and the sensors or digital features most robustly associated with suicidal thoughts and behaviors remain unclear for several reasons. First, the robustness of sensor signals is often insufficient; many passive sensing features, including physiological measures, lack stability across devices or populations and provide limited incremental value beyond traditional self-report measures [12]. Second, individual heterogeneity makes universal indicators difficult to define, as suicide risk is an idiosyncratic process rather than a monolithic state [13]. Although specific markers, such as active midnight smartphone use (11 PM to 1 AM), may predict next-day ideation, other indicators, such as decreased social frequency, remain ambiguous. These indicators may signal crisis for some but represent benign routine for others [14]. Such variability, often referred to as phenotypic equifinality, indicates that similar digital behaviors reflect different risk levels based on an individual's baseline, complicating standardized interpretation. Third, traditional gold-standard interviews often miss the proximal, transient triggers (eg, acute midnight loneliness or sudden sleep deprivation) that are visible only through "screenomics" data [15], creating a disconnect between real-time digital markers and traditional interviews. Consequently, clear guidance on the sensors or digital features that should be prioritized for reliable monitoring is lacking, limiting the translational potential of digital sensing in suicide or other clinical risk assessment.

Addressing these limitations requires a structured framework that can systematically translate passive sensor data into clinically meaningful suicide-related states. Mohr et al [16] proposed the layered hierarchical sensemaking framework (LHSF), which conceptualizes the transformation from low-level sensor data to high-level behavioral markers. This scoping review aimed to apply the LHSF to (1) systematically map how digital sensor data were transformed into low-level features, high-level behavioral markers, and suicide-related outcomes; and (2) propose a clinically interpretable structure for suicide-specific digital phenotyping to develop actionable digital markers.

Methods

Overview

This scoping review was conducted based on Arksey and O'Malley's methodological framework and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines (Checklist 1) [17,18]. The review was guided by the population-concept-context framework (population: not restricted; concept: passively collected digital data from smartphones or wearable devices; and context: suicide-related

XSL•FO
RenderX

outcomes). The study was conducted through collaboration among 3 research teams, with external validation performed by separate research teams. Fourteen studies were analyzed based on the LHSF using a structured data charting and synthesis approach [16], and Google Sheets was used as a data management tool. The protocol was registered in the Open Science Framework [19].

Search Strategy

Textbox 1 summarizes the query terms used in the search strategy. The search strategy was developed iteratively using the population-concept-context framework to ensure alignment with the review objectives. Four databases (PubMed, CINAHL, PsycINFO, and IEEE Xplore) were searched for studies published between January 1, 2015, and December 31, 2025. The search combined suicide-related terms with those related to digital sensing, as summarized in Textbox 1 (see full search strings in Multimedia Appendix 1). The final search was conducted in December 2025. Gray literature was excluded from the scope of this review. This boundary was established because the primary objective was to map granular digital signals onto the structured LHSF framework, which requires fully detailed and finalized descriptions of raw sensor processing, feature extraction, and statistical verification—details that are often omitted or partially reported in non–peer-reviewed literature. The reference list screening was conducted to identify additional relevant studies. A librarian at the affiliated medical library was consulted to review and refine the search strategy to enhance comprehensiveness and accuracy.

Textbox 1. Summary of the search terms.

Suicide-related terms include the following:

- Suicide
- Suicidal ideation
- Suicide attempt
- Suicide risk
- Suicidal behavior

Digital sensing terms include the following:

- Smartphone
- Wearable devices
- Passive sensing
- Digital marker
- Digital phenotype

Eligibility Criteria

We included studies that reported correlational or predictive analyses that explored digital markers with suicide-related outcomes. Studies that relied solely on active data collection methods or did not assess suicide-related outcomes were excluded. No language restrictions were applied. Textbox 2 summarizes the detailed inclusion and exclusion criteria. The publication period was restricted to studies from January 2015 onward, corresponding to the introduction of digital phenotyping in clinical research and wearable devices [20].

Textbox 2. Inclusion and exclusion criteria.

Inclusion criteria

- Full-text, peer-reviewed studies (including articles published online ahead of print)
- Studies that used passively collected digital data from smartphones or wearable devices
- Studies that measured suicide-related outcomes, including suicidal ideation, plans, attempts, behaviors, or suicide risk
- Studies that analyzed associations between passive digital data and suicide-related outcomes or developed predictive models to examine suicide-related outcomes

Exclusion criteria

- Studies that used only active data collection methods (eg, online surveys, self-report questionnaires, or ecological momentary assessment apps)
- Studies that did not assess suicide-related outcomes and only focused on general mental health indicators (eg, depression, stress, or nonsuicidal self-injury)
- Feasibility studies that did not examine associations or predictive relationships between passive digital data and suicide-related outcomes
- Reviews, protocols, commentaries, conference abstracts, or non–peer-reviewed studies

XSL•FO
RenderX

Study Selection

Two authors (SK and JK) independently screened the titles and abstracts of potentially eligible studies, followed by full-text review. If disagreements arose, they were resolved through discussion between 2 reviewers, and those not resolved were adjudicated by the principal investigator (HK).

Data Charting

Two authors (SK and JK) initially extracted the data, which were reviewed by the principal investigator (HK). Subsequently, 2 separate teams (WTC and YZ; and TTL and ZYAT) provided external validation and contributed to refining the analysis. Discrepancies were resolved through team discussion. Google Sheets was used to organize a structured data matrix for data charting. The data charting form was developed by the research team and refined through discussion prior to full data extraction. Study characteristics, device and sensor types, and low-level features and high-level behavioral markers were charted. Narrative mapping was conducted according to the LHSF [16]. This scoping review followed Joanna Briggs Institute recommendations for scoping reviews [21]. Consistent with these guidelines, methodological quality assessment was not performed.

Data Synthesis Using the LHSF

The LHSF was applied to organize data from raw sensor inputs to inferred behavioral patterns related to suicidality [16]. On the basis of the 4 hierarchical layers, the extracted features were grouped:

1. Sensor: hardware- or software-based sources within smartphones, wearables, or apps that passively collected environmental or behavioral data (eg, accelerometer, photoplethysmography, and phone usage logs);
2. Low-level features: quantitative indicators derived directly from raw sensor data (eg, step count and HRV);
3. High-level behavioral markers: psychosocial or behavioral patterns inferred from combinations of low-level features (eg, sleep disturbance and social withdrawal); and
4. Suicide-related outcomes: clinical indicators, such as suicidal ideation, suicide attempts, or suicide risk.

To synthesize findings across the 4 levels, we organized and categorized extracted features according to the LHSF structure. While the original framework was retained, specific digital features and behavioral markers were refined and expanded based on the included studies to reflect recent advances in digital phenotyping research. This approach enabled us to map contemporary empirical evidence onto the existing framework.

To ensure consistent operationalization across layers, explicit decision rules were applied during data charting. A feature was classified as a low-level indicator (layer 2) if it represented a directly quantifiable, sensor-derived measurement requiring no inferential interpretation. A construct was classified as a high-level behavioral marker (layer 3) if it represented a psychosocial or clinical construct that required aggregating or theoretically interpreting one or more low-level features. When the original study explicitly used a term corresponding to a layer 3 construct, it was retained as source defined. When no such label was present, 2 authors (SK and JK) independently derived a conceptually equivalent label based on the analytical context of the original study, indicated by superscript “a” in Table 1. Discrepancies in layer assignment or label derivation were resolved through discussion. A detailed overview of the operationalization criteria applied to each study is provided in Multimedia Appendix 2.

XSL•FO
RenderX

Table . Passive data sources and extracted features mapped to the layered hierarchical sensemaking framework (LHSF).

Study	Device	LHSF layer 1: sensor	LHSF layer 2: low-level features	LHSF layer 3: high-level behavioral markers	LHSF layer 4: suicide-related outcome (measures)
[22]	Smartphone	SMS text	Linguistic inquiry and word count–based word proportions (eg, first-person pronouns, emotion-related words) and communication activity (eg, message counts and sent-to-received message ratios)	Psychological and behavioral patterns[a] (self-focus, sentiment, and social engagement)	Suicide attempts: high risk; suicidal ideation: moderate risk; depressed and positive episodes: low/minimal risk (retrospective episode-based interview)
[23]	Smartphone	Screen time app	Smartphone screen time per app category	Smartphone use patterns[a]	Suicide risk: past suicide attempts, suicidal ideation frequency, suicide threat, and perceived likelihood of death by suicide (Suicidal Behaviors Questionnaire-revised)
[24]	Wearable (wristband)	Photoplethysmography sensor	High-frequency HRV[b] power	Reduced parasympathetic activity	Suicide severity: suicidal ideation and behavior (Columbia-Suicide Severity Rating Scale)
[25]	Smartphone, wearable (wristband)	Electrodermal activity sensor	Number of autonomic events	Autonomic arousal	Suicidal ideation (EMA[c])
[26]	Smartphone, wearable (wristband)	Not stated but sensors embedded in wristband	Sleep minutes and step counts	Sleep patterns and activity patterns	Suicidal ideation (Patient Health Questionnaire-item 9)
[27]	Smartphone	SMS text	Incidence of sleep-related text messages, the number of nightly texts sent, and the number of unique nightly hour bins	Sleep disturbance	Suicidal ideation; suicide attempts (clinical interview)
[28]	Smartphone, wearable (wristband)	Not stated but sensors embedded in wristband	Root-mean-square of successive differences, resting heart rate, total steps, and sleep duration	Distress, physical activity, and sleep disturbance	Next-day suicidal ideation (EMA using Columbia-Suicide Severity Rating Scale-based items)
[29]	Wearable (compression shirt)	Electrocardiogram and pressure sensors	HRV indices and respiratory rate	Autonomic imbalance	Current suicidal ideation; history of suicide attempts (Schedule for Affective Disorders and Schizophrenia-Lifetime version–based clinical interview)
[30]	Smartphone	Global positioning system, actigraphy (sensor was not stated); health data aggregation app; and phone usage logs (app usage logs)	Distance walked, time spent at home, steps taken, and app usage (binary)	Behavioral changes in daily activity profiles	1-week suicide-risk events (electronic health records)
[31]	Smartphone, wearable (wristband)	Three-axis accelerometer	Time in bed, sleep onset latency, sleep duration, wake after sleep onset, and sleep efficiency	Sleep disturbance	Next-day suicidal ideation (EMA)

XSL•FO
RenderX

Study	Device	LHSF layer 1: sensor	LHSF layer 2: low-level features	LHSF layer 3: high-level behavioral markers	LHSF layer 4: suicide-related outcome (measures)
[32]	Smartphone	GPS	Time at home, entropy (location variability), and distance traveled	Social connectedness, behavioral withdrawal, and anhedonia	Next-week suicidal events: attempts, psychiatric hospitalization, suicide-related emergency department visits (Self-injurious Thoughts and Behaviors Interview, weekly experience sampling method, Columbia-Suicide Severity Rating Scale); same-week/next-week suicidal ideation (weekly experience sampling method)
[33]	Wearable (wristband)	Photoplethysmography sensor, accelerometer, and gyroscope	Heart rate and HRV, movement index, wear-free time, step count, walking/running time, other exercise time, resting time, and total sleep time	Stress states (autonomic dysregulation), daily behavioral patterns[a]	Elevated suicide risk: suicidal ideation and behavior (Hamilton Depression Rating Scale-item 3)
[34]	Smartphone	Screen capture app	Dictionary-based word counts	Expression related to suicide risk[a]	Passive/active suicidal ideation, suicidal planning (EMA)
[35]	Smartphone	Screen capture app	Smartphone screen time (screenshot image counts, any use, and proportion of use)	Temporal patterns of smartphone use[a]	Suicide risk: suicidal ideation, suicidal planning, and suicidal desire (EMA)

[a]These categories represent high-level behavioral markers (LHSF layer 3) derived by the authors and were not explicitly defined as such in the original studies.

[b]HRV: heart rate variability.

[c]EMA: ecological momentary assessment.

Synthesis of Evidence Strength Across Feature Domains

To summarize the strength of evidence across feature domains, each study was classified according to the statistical results reported in the original article. A study was counted as showing a significant association for a given feature domain if at least one indicator in that domain was significantly associated with, or predictive of, at least one suicide-related outcome, such as suicidal ideation, planning, attempts, risk, or events. This classification was retained even when other indicators or outcomes within the same study were nonsignificant. Domain-level evidence was then described by the number of studies reporting significant associations, rather than by a single subjective judgment. Predictive performance was interpreted against a conventional discrimination threshold of an area under the curve (AUC) of 0.80 or above.

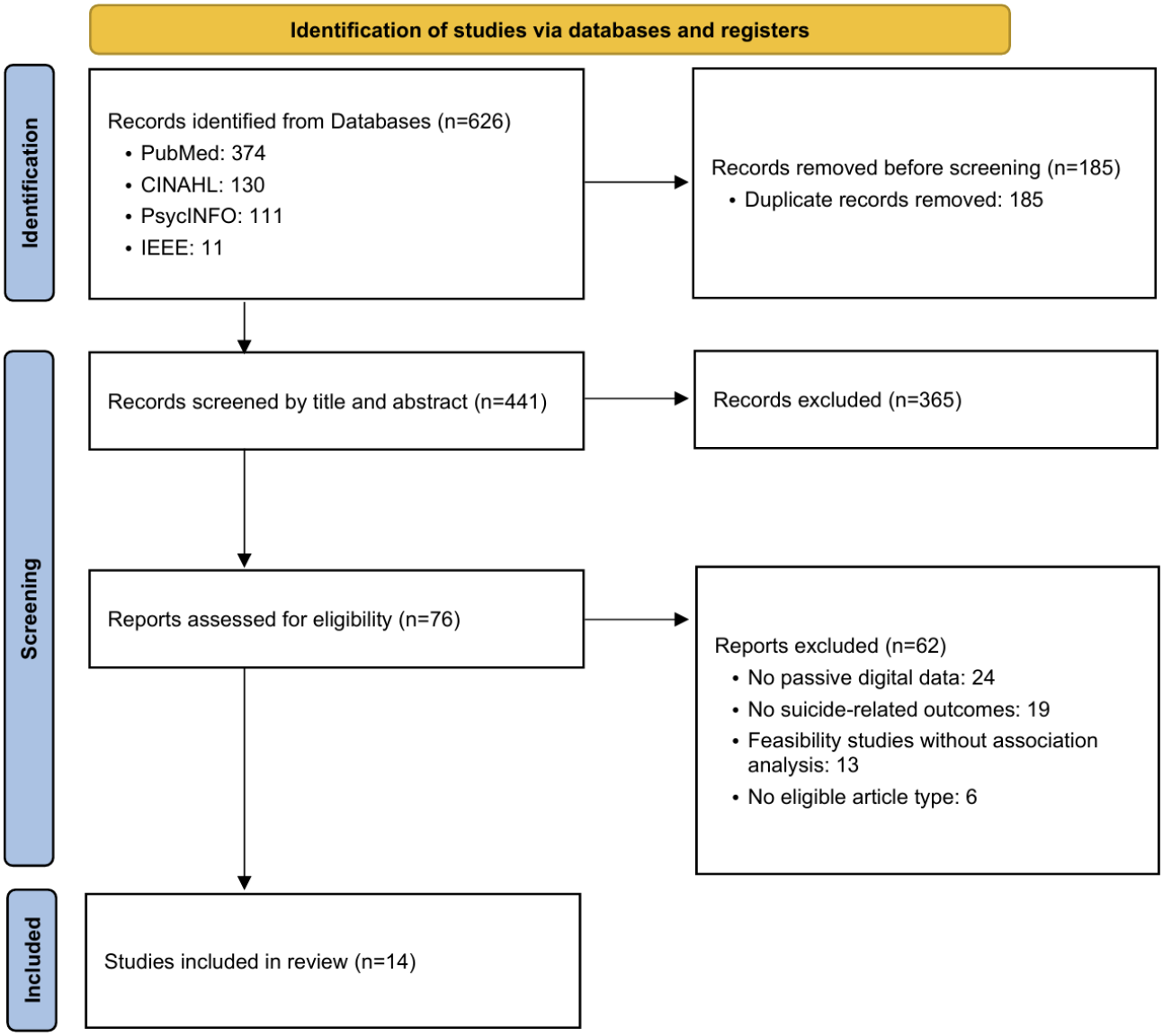
Results

Selection of Sources of Evidence

We conducted the final search in December 2025 and identified 626 studies across the 4 databases (PubMed: n=374; CINAHL: n=130; PsycINFO: n=111; and IEEE Xplore: n=11). After removing 185 duplicate records, 441 (70.4%) articles were screened based on their titles and abstracts. Of these, 365 (82.8%) records were excluded, and 76 (17.2%) full-text articles were assessed. Subsequently, 62 (81.6%) articles were excluded (Figure 1). Ultimately, 14 (18.4%) studies were included [22-35]. Reference list screening identified no additional eligible articles.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.



Characteristics of the Included Studies

Table 2 presents the characteristics of the included studies. Publication years of the 14 included studies ranged from 2020 to 2025, with most of the studies published in 2025 (n=5, 35.7%). Most studies were conducted in the United States (n=11, 78.6%). All included studies used observational designs, in which 12 (85.7%) were longitudinal studies, and 2 (14.3%) were cross-sectional. Furthermore, 11 (78.6%) studies were in prospective design, and 3 (21.4%) were retrospective ones. Regarding characteristics of participants, 3 (21.4%) studies included adolescents, and 11 (78.6%) studies included only adults. Most studies targeted high-risk populations (n=12, 85.7%), such as individuals who visited emergency departments for suicide-related reasons, psychiatric inpatients or outpatients, and those with a history of suicidal ideation or behavior. Duration of passive data collection varied across studies, ranging from 24 hours to 6 months.

XSL•FO
RenderX

Table . Characteristics of the included studies (N=14).

Author (year)	Country	Study design	Study participants			Passive data collection period
			Size, n	Types	Characteristics (female ratio/setting)	
Glenn et al (2020) [22]	United States	• Observational • Longitudinal • Quantitative • Retrospective • Pilot study	33	• High-risk adults who reported a past suicide attempt • Age: mean 20.4 (SD 2.4) y	• Female ratio: 84.8% • Community setting	2 wk
Coyne et al (2021) [23]	United States	• Observational • Cross-sectional[a] • Quantitative • Prospective (cohort)	281	• General adolescents and emerging adults • Age: mean 23.3 (SD 1.04) y	• Female ratio: 58% • Community setting	2 wk
Sheridan et al (2021) [24]	United States	• Observational • Longitudinal • Quantitative • Prospective	51	• High-risk adolescents • Age: mean 16.7 (SD 1.5; range 14-21) y	• Female ratio: 72.6% • Emergency department or psychiatric inpatient unit	7 d
Kleiman et al (2021) [25]	United States	• Observational • Longitudinal • Quantitative • Prospective (model fit improvement)	25	• High-risk adults • Age: mean 33.48 (SD 13.84, range 19-63) y	• Female ratio: 56% • Psychiatric inpatient unit	Inpatient stay +28 d after discharge
Horwitz et al (2022) [26]	United States	• Observational • Longitudinal • Quantitative • Prospective (cohort)	2881	• General adults (medical interns) • Age: mean 27.6 (SD 2.6) y	• Female ratio: 56.8% • Community setting	2 - 3 mo
Ladis et al (2023) [27]	United States	• Observational • Longitudinal • Quantitative • Retrospective • Pilot study	26	• High-risk young adults who reported past suicide attempts • Age: mean 20.8 (SD 2.6) y	• Female ratio: 84.6% • Community setting	2 wk
Czyz et al (2023) [28]	United States	• Observational • Longitudinal • Quantitative • Prospective	102	• High-risk young adults • Age: mean 20.9 (SD 2.1, range 18-25) y	• Female ratio: 81.4% • Emergency department	8 wk
Ortiz et al (2024) [29]	Canada	• Observational • Cross-sectional • Quantitative • Retrospective • Pilot study	53	• High-risk adults • Age: mean 44.7 (SD 13.1) y	• Female ratio: 66% • Psychiatric outpatient clinics	24 h

XSL•FO
RenderX

Author (year)	Country	Study design	Study participants			Passive data collection period
			Size, n	Types	Characteristics (female ratio/setting)	
Barrigon et al (2023) [30]	Spain	• Observational • Longitudinal • Quantitative • Prospective	225	• High-risk adults • Age: mean 43.24 (SD 14.13 y)	• Female ratio: 62.7% • Psychiatric outpatient clinic	6 mo
Mournet et al (2025) [31]	United States	• Observational • Longitudinal • Quantitative • Prospective • Case series	7	• High-risk adults • Age: mean 28 (SD 10.39, range 19-50) y	• Female ratio: 71.4% • Psychiatric outpatient program	6 wk
Auerbach et al (2025) [32]	United States	• Observational • Longitudinal • Quantitative • Prospective • Case series	186	• High-risk adolescents • Age: mean 16.43 (SD 1.68, range 13-18) y	• Female ratio: 79.6% • Psychiatric outpatient programs, emergency departments, or community-based recruitment	6 mo
Um et al (2025) [33]	Republic of Korea	• Observational • Longitudinal • Quantitative • Prospective	59 (patient group: n=39; healthy controls: n=20)	• High-risk adults and healthy controls • Age range 20-55 y	• Female ratio: 57.6% • Psychiatric outpatient clinic; healthy controls recruited via advertisement	2 mo
Ammerman et al[b] (2025) [34]	United States	• Observational • Longitudinal • Quantitative • Prospective	79	• High-risk adults who reported past-month active suicidal ideation or behaviors • Age: mean 35.15 (SD 11.07; range: 20 - 63) y	• Female ratio: 68.3% • Community setting	28 d
Jacobucci et al[b] (2025) [35]	United States	• Observational • Longitudinal • Quantitative • Prospective	79	• High-risk adults who reported past-month suicidal ideation or behaviors • Age: mean 35.15 (SD 11.07; range: 20 - 63) y	• Female ratio: 68.3% • Community setting	28 d

[a]Passive data were collected and analyzed cross-sectionally over a 2-week period within a larger longitudinal project.

[b]Studies 34 and 35 utilized the same dataset; however, they analyzed different independent variables (passive sensing-based text patterns vs. smartphone screen time).

Regarding demographic composition, most study samples were predominantly White or Caucasian, with White participants comprising 57% to 90% of samples across studies reporting race or ethnicity. Female participants comprised the majority

XSL•FO
RenderX

across all studies, with female ratios ranging from 56% to 84.8%. Although most studies recruited high-risk adult clinical populations, 3 (21.4%) studies included adolescents. Among studies reporting clinical diagnoses, depressive and anxiety disorders were the most represented conditions. Socioeconomic status was inconsistently reported; where reported, samples tended to include participants with different levels of college education or current employment. Detailed demographic and clinical sample characteristics are presented in Multimedia Appendix 3.

Passive Data Sources and Extracted Features Mapped to the LHSF

Overview

We mapped the passive data sources and features extracted from the 14 studies according to the 4-layer structure of the LHSF (Table 1) [16]. Results were presented from data sources to suicide-related outcomes. High-level behavioral markers (layer 3) were labeled using the terminology reported in the original studies. When not explicitly defined, conceptually similar descriptors derived by the authors were indicated with the superscript "a" (Table 1).

LHSF Layer 1: Sensors

Passive sensing devices were categorized as smartphones and wearable devices. All wearable devices were wristband based, except for 1 study that used a compression shirt equipped with sensors to measure the respiratory rate [29]. Smartphones were used either alone or in combination with wearable devices in studies that assessed suicide-related outcomes via EMA [25,28,31,34,35] or experience sampling method [32].

Wearable devices primarily collected physiological signals, including photoplethysmography [24,33], electrodermal activity [25], accelerometer and gyroscope data [33], electrocardiogram signals [29], and pressure sensor data [29], which were used to derive indicators, such as HRV and autonomic nervous system responses. In contrast, smartphones captured behavioral data, including location information via a global positioning system [30,32], accelerometer-based sleep metrics [31], screen time [23], screen capture data [34,35], phone usage logs [30], and SMS text data [22,27].

LHSF Layer 2: Low-Level Features

Low-level features were extracted from sensor data and categorized into physiological, behavioral, and digital indicators. Physiological indicators included heart rate and HRV indices, respiratory rate, and electrodermal activity [24,25,28,29,33]. Behavioral and physical activity indicators comprised step counts, distance traveled, exercise and rest duration, total sleep time, and mobility entropy [26,28,30-33]. Digital behavioral indicators included smartphone screen time and phone usage logs [23,30,35], dictionary-based word counts [34], number of nighttime text messages sent [27], and ratios of sent-to-received text messages [22]. Studies that analyzed text message content included dictionary-based linguistic features, such as emotion-related or sleep-related word usage [22,27].

LHSF Layer 3: High-Level Behavioral Markers

High-level behavioral markers were derived through an integration of low-level features. Studies (n=4, 28.6%) that leveraged physiological low-level features mapped these indicators to markers that reflected reduced parasympathetic activity or autonomic dysregulation, corresponding to altered autonomic nervous system regulation [24,25,29,33].

Studies (n=6, 42.9%) based on behavioral low-level features identified high-level markers that captured changes in overall daily activity profiles. These markers included sleep patterns or sleep disturbance, changes in activity patterns and activity levels, social connectedness, behavioral withdrawal, and anhedonia, which reflected behavioral and psychosocial characteristics [26,28,30-33].

Studies (n=4, 28.6%) that used smartphone-based digital behavioral features and reported high-level markers included smartphone use patterns or temporal changes in use, changes in daily activity profiles, and sleep disturbance [23,27,30,35]. In addition, 3 (21.4%) studies derived high-level markers from text-based analyses, which included expressions related to suicide risk, self-focus, sentiment, and social engagement, representing psychological and behavioral patterns [22,27,34].

LHSF Layer 4: Suicide-Related Outcomes

Suicide-related outcomes were assessed using clinical measures and real-time monitoring. Outcomes included suicidal ideation and planning, suicide attempts, severity, risk, and events.

Many studies reported suicide-related outcomes measured via EMA or experience sampling methods [25,28,31,32,34,35], including those that administered standardized clinical questionnaires via EMA [28]. In addition, several studies assessed suicide-related outcomes via structured clinical interviews or standardized questionnaire items [23,24,26,28,29,32,33]. The Columbia-Suicide Severity Rating Scale was the most frequently used clinical instrument [24,28,32]. One study identified suicide-related events from electronic health records [30].

Associations and Predictive Relationships Between Passive Digital Markers and Suicide-Related Outcomes

Overview

Table 3 presents the statistical analyses and main findings of the relationships between passively collected digital features and suicide-related outcomes across the 14 studies.

XSL•FO
RenderX

Table . Association and prediction between passive digital markers and suicide-related outcomes.

Study	Outcome type	Main statistical analysis methods	Predictive metric	Predictive performance	Main findings	Further suggestion
[22]	Association	Mixed effects models (generalized linear and linear mixed effects models)	Not applicable	Not applicable	• Steeper increases in anger (χ^2_3=7.83; *P*=.049) and steeper decreases in positive emotion (χ^2_3=41.67, *P*<.001) over time were associated with suicide attempt episodes (high risk) compared with other episodes (moderate risk, minimal or no risk). • Increases in self-focus did not uniquely distinguish suicide attempt episodes from all other episode types (*P*=.37 when compared to positive mood episodes).	• Prospective, frequently assessed designs • Objective classification of suicide-related events • Larger samples for adequate power • Development of machine learning models for real-time prediction of suicide risk • Advanced Natural Language Processing beyond dictionary-based methods
[23]	Association	Zero-inflated negative binomial regression and hierarchical regression analysis	Not applicable	Not applicable	• Higher use of entertainment apps in girls and reading apps in boys was associated with increased suicide risk (girls: β=.23 - .25; *P*<.001; boys: β=.28 - .29; *P*<.001). • Video game use was associated with higher suicide risk among boys experiencing cybervictimization (β=.27; *P*=.01).	• Replication in clinical high-risk samples • Longitudinal studies incorporating passive sensing from earlier waves • Repeated baseline and longitudinal assessment of suicide risk • Examination of contextual factors associated with media use

XSL•FO
RenderX

Study	Outcome type	Main statistical analysis methods	Predictive metric	Predictive performance	Main findings	Further suggestion
[24]	Association	Mixed effects models (random effect for patient)	Not applicable	Not applicable	• Increased high-frequency HRV[a] at night was significantly associated with decreased suicide severity (mean difference=11.89 ms/√Hz; *P*=.005).	• Validation in larger cohorts • Improvement of wearable data quality • Inclusion of additional HRV metrics
[25]	Prediction (model fit improvement)	Linear and generalized linear mixed effects models	Performance score	• Performance score improvement (daily level): mean 18.1% (range: 5.4% - 31.9%) • Performance score improvement (hourly level): mean 14.2% (2.7% - 22.9%)	• Electrodermal activity–based autonomic events improved model performance for predicting suicidal ideation severity (performance score improvement up to 31.90%) but worsened model fit for the presence of suicidal ideation.	• Replication in larger samples • Algorithm refinement to distinguish psychological distress from nonspecific physiological activation
[26]	Prediction	Three-step hierarchical logistic regression	AUC[b]	AUC: 0.74	• Passively collected sleep and physical activity features had no incremental predictive value for suicidal ideation and only negligible predictive value for depression (suicidal ideation: χ^2_4=0.7; *P*=.95; depression: χ^2_4=13.5, *P*=.009).	• Refinement of passive sensing features • Optimization of prediction windows and data processing strategies • Development of adaptive interventions based on early risk indicators
[27]	Association	Generalized linear mixed models, negative binomial models, and multilevel ordinal logistic regression	Not applicable	Not applicable		

XSL•FO
RenderX

Study	Outcome type	Main statistical analysis methods	Predictive metric	Predictive performance	Main findings	Further suggestion
					• Sleep-related linguistic features were not significantly associated with suicidal ideation or suicide attempt episodes (suicidal ideation: z=1.05; *P*=.29; and suicide attempt: z=1.88; *P*=.06), and nocturnal texting patterns were also not significantly associated with these outcomes (suicidal ideation: z=0.61; *P*=.54; and suicide attempt: z=0.83, *P*=.41).	• Integration of wearable sensors for objective sleep assessment • Refinement and validation of sleep-related linguistic dictionaries • Replication in larger and more diverse samples
[28]	Prediction	Multilevel machine learning (based classification and regression tree models)	AUC	• AUC (ecological momentary assessment only): 0.84 • AUC (passive only): 0.56 • AUC (ecological momentary assessment+ passive combined): 0.84	• Self-reported ecological momentary assessment features had excellent accuracy in predicting next-day suicidal ideation. • Passive sensor data alone had poor predictive performance and did not improve the model when combined.	• Prediction of suicide attempts beyond ideation • Replication in independent samples for external validation • Application of alternative machine learning approaches • Optimization of assessment frequency to balance burden and accuracy
[29]	Association	Logistic regression	Not applicable	Not applicable		• Longitudinal studies on diaphragmatic breathing • Mechanistic research on respiratory rate and hypocapnia

XSL•FO
RenderX

Study	Outcome type	Main statistical analysis methods	Predictive metric	Predictive performance	Main findings	Further suggestion
					• Suicide attempt history was associated with increased respiratory rate (U=14.1; P<.01) and higher low-frequency to high-frequency HRV ratio (*Wald* χ^2=4.1; P=.04). • Current suicidal ideation was associated only with a higher low-frequency to high-frequency HRV ratio (*Wald* χ^2=3.8; P<.05).	
[30]	Prediction	Machine learning (unsupervised model and Bayesian change-point detection)	AUC	AUC: 0.78	Behavioral changes detected from passive smartphone data predicted 1-wk suicide risk events, including suicide attempts and psychiatric emergency visits.	• Integrate passively collected smartphone data with actively collected self-report data, electronic health records, and clinical assessment data • Include a broader range of passively collected behavioral and environmental data to enhance short-term suicide risk prediction
[31]	Association	Individual linear regressions	Not applicable	Not applicable	Among actigraphy-derived sleep metrics (time in bed, sleep onset, sleep duration, sleep efficiency, and wake after sleep onset), only wake after sleep onset had a significant within-person association with next-day suicidal ideation in one of 7 participants (b=0.04; P=.02).	• Replicate in larger and more diverse samples • Examination of sleep regularity metrics • Individual-level analysis of actigraphy data
[32]	Prediction[c]		AUC	AUC: 0.64		

XSL•FO
RenderX

Study	Outcome type	Main statistical analysis methods	Predictive metric	Predictive performance	Main findings	Further suggestion
		Repeated-measures mixed effects logistic regression models			• Greater within-person increases in homestay were related to next-week suicidal events (aOR=1.99; *P*=.01). • Entropy and distance traveled were not (entropy: *P*=.14; distance: *P*=.13), and geolocation features were not related to next-week suicidal ideation (homestay: *P*=.90; entropy: *P*=.56; and distance: *P*=.38).	• Multimodal sensor and clinical data integration • External validation to improve generalizability
[33]	Prediction	Machine learning models	AUC	• AUC (single-step model): 0.88 • AUC (multi-step model): 0.89	HRV-related features and clinician-rated depression severity were key predictors of an elevated suicide risk (Hamilton Depression Rating Scale-item 3≥1) in patients with depressive disorders. Activity and sleep data had limited predictive value.	• Refining temporal resolution for real-time tracking • Expanding digital phenotypes (eg, geolocation, communication patterns) • Increasing sample size and diversity • Improving wearable adherence • External validation of machine learning models
[34]	Association	Zero-inflated negative binomial regression, Spearman correlation	Not applicable	Not applicable	Passively collected smartphone text patterns related to suicide risk were significantly associated with passive and active suicidal ideation (r =0.05 - 0.07, P<.05) and with suicidal planning based on within-person zero-inflated negative binomial effects (P<.05).	• Examine specific words driving divergent dictionary effects • Account for cooccurrence of suicidal ideation and planning • Apply flexible natural language processing or machine learning methods

XSL•FO
RenderX

Study	Outcome type	Main statistical analysis methods	Predictive metric	Predictive performance	Main findings	Further suggestion
[35]	Association	Bayesian zero-inflated negative binomial multilevel model	Not applicable	Not applicable	Within-person increases in smartphone screen time were associated with higher near-term (<3 h) suicidal ideation (passive: *b*=0.04, 95% CI 0.02 - 0.06; active: *b*=0.04, 95% CI 0.02 - 0.07) and suicidal planning (*b*=0.15, 95% CI 0.02 - 0.26).	• Examination of screen content and contextual factors • Refinement of temporal modeling • Validation in larger and more diverse samples

[a]HRV: heart rate variability.

[b]AUC: area under the curve.

[c]Ref 32 reported both associational (effect estimates) and predictive (AUC) findings from a single modeling approach and was classified as a predictive study for the analytic-approach count.

Statistical Analyses and Predictive Modeling

Across the 14 studies, 6 (42.9%) used predictive modeling approaches [25,26,28,30,32,33], and 8 (57.1%) conducted association or correlational analyses [22-24,27,29,31,34,35]. One predictive study [32] also reported association findings from the same modeling framework and was therefore included in the predictive group. Among studies that used predictive models, reported model performance varied, and AUC values ranged from 0.56 to 0.89.

Physiological Indicators

All 4 (28.6%) studies examining physiological digital features related to autonomic regulation reported at least one significant association with suicide-related outcomes [24,25,29,33]. Increased high-frequency HRV at night was significantly associated with decreased suicide severity [24]. Autonomic arousal events derived from electrodermal activity improved the model performance for predicting the severity of suicidal ideation; however, adding these features worsened model fit when the outcome was the presence of suicidal ideation [25]. Lower HRV-related features were identified as key predictors of an elevated suicide risk with excellent discrimination (AUC=0.89 for the multistep model) [33]. In contrast, physiological activity and sleep data had limited predictive value for suicide risk [33]. Individuals with a history of suicide attempts also exhibited increased respiratory rates and higher low-frequency to high-frequency HRV ratios compared with those without such history [29].

Mobility and Geolocation Indicators

Two (14.3%) studies examining mobility and geolocation-derived behavioral markers reported at least one significant association with suicide-related outcomes [30,32]. Behavioral changes derived from smartphone data predicted suicide risk events over 1 week with acceptable classification performance (AUC=0.78) [30]. In one predictive study that also examined within-person associations [32], greater increases in homestay were significantly associated with next-week suicidal events (adjusted odds ratio=1.99, *P*=.01); however, predictive accuracy based on the homestay model was modest (AUC=0.64). Other geolocation features, such as entropy and distance traveled, were not significant predictors of suicidal incidents. Additionally, no geographical information was associated with next-week suicidal ideation [32].

Smartphone Use Patterns

Two (14.3%) studies examining smartphone use patterns reported significant associations with suicide-related outcomes [23,35]. Higher entertainment and reading app use in girls and boys, respectively, were associated with increased suicide risk; additionally, video game use was associated with elevated risk among boys who had experienced cybervictimization [23]. Within-person increases in smartphone screen time were associated with higher levels of near-term (within 3 h prior to EMA assessment) suicidal ideation and planning [35].

Sleep Indicators

One case series study examined participant-specific associations between actigraphy-derived sleep features and next-day suicidal ideation using individual-level (within-person) regressions [31]. Assessed sleep metrics included time in bed, sleep onset, sleep duration, sleep efficiency, and wake after sleep onset. Among the 5 sleep metrics assessed, only wake after sleep onset was significantly associated with higher next-day suicidal ideation, and this association was observed in only 1 of the 7 participants. Sleep features examined in predictive contexts likewise showed no incremental predictive value [26,28].

Linguistic and Communication Markers

Three studies examined linguistic and communication-based markers derived from smartphone text data [22,27,34]. Of these, 2 reported at least one significant association [22,34], and 1 reported no significant association [27]. In a study by Ammerman et al [34], smartphone text patterns related to suicide risk were significantly associated with both passive and active suicidal ideation, as well as suicidal planning. In a retrospective episode-based analysis [22], trajectories of emotion-related language differed across suicide risk levels: before suicide attempts, anger increased and positive emotion decreased over

XSL•FO
RenderX

2 weeks, although self-focus did not consistently differentiate episode types. In contrast, sleep-related linguistic features and nocturnal texting patterns were not significantly associated with suicidal ideation or suicide attempts [27].

Nonsignificant or Limited Predictive Performance

Two (14.3%) studies reported nonsignificant or limited predictive performance for sleep- and activity-related passive features [26,28]. In predictive contexts, passively collected sleep and physical activity features had limited or no incremental predictive value for suicidal ideation when compared with, or added to, clinician-rated or self-reported measures [26,28]. In one study, passive sensor data alone demonstrated poor predictive performance for next-day suicidal ideation (AUC=0.56) and did not improve prediction when combined with EMA features [28].

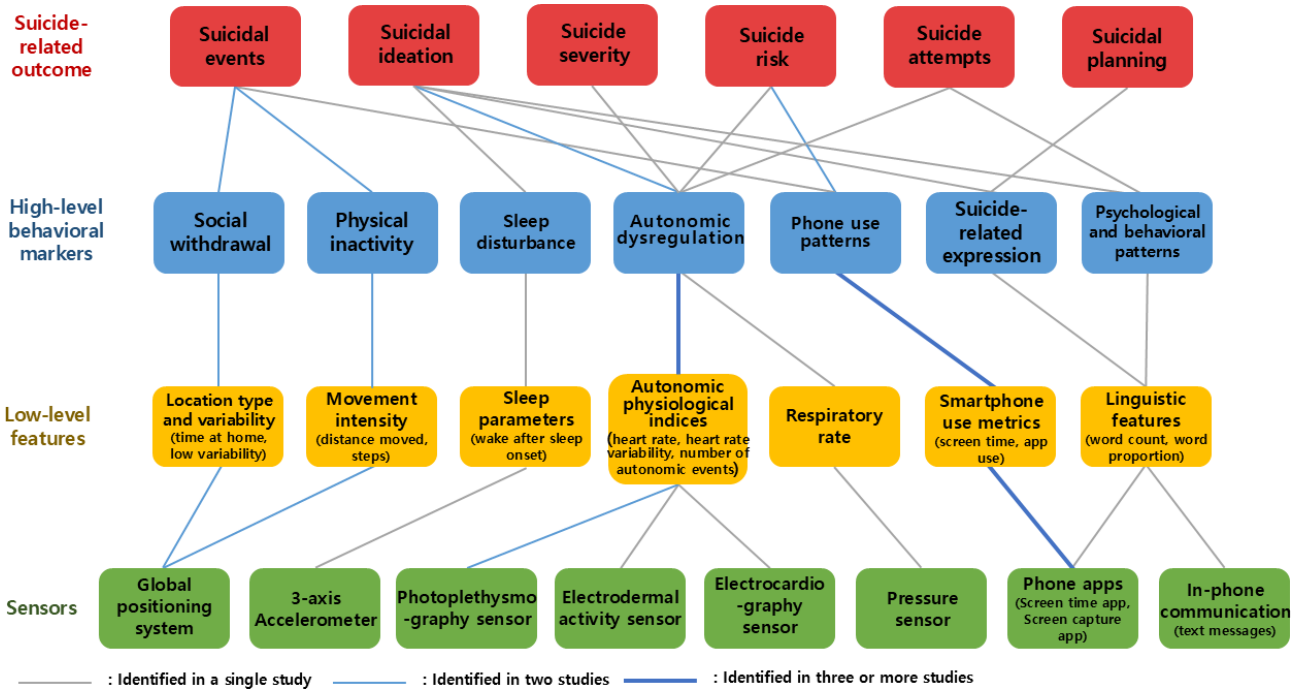
Future Research Directions Identified Across Studies

Across the included studies, several directions for future research were identified (Table 3). First, many studies emphasized the need for replication in larger and more diverse samples, along with the use of longitudinal designs, and more frequent assessments to improve the robustness and generalizability of findings [22-24,27,28,31,33,35]. Second, a consistent focus was placed on improving data quality and methodological approaches, including refinement of passive sensing features, optimization of temporal resolution, and the application of more advanced analytic techniques such as machine learning and natural language processing [22,25-28,33-35]. Third, these findings collectively highlight the critical need to bridge the gap between passively collected digital data and clinically interpretable indicators of suicide risk [30,32]. Although diverse data streams and analytic approaches are increasingly used, there remains a lack of structured approaches to translate low-level digital features into meaningful behavioral or clinical constructs. This underscores the importance of developing frameworks that can systematically organize and represent the relationships between raw data and clinically relevant suicide-related outcomes. Further details are provided in Table 3.

LHSF-Based Conceptual Synthesis of Passive Digital Markers and Suicide-Related Outcomes

Figure 2 presents the LHSF-based conceptual mapping synthesizing the reported associations between passive digital markers and suicide-related outcomes. Low-level features extracted from passive sensors consisted of 7 domains, including location type and variability, movement intensity, sleep parameters, autonomic physiological indices, respiratory rate, smartphone use metrics, and linguistic features.

Figure 2. Layered hierarchical sensemaking framework–based conceptual mapping of reported associations between passive digital data and suicide-related outcomes.



Across the 4 layers of the LHSF, physiological sensor-derived features were linked to autonomic dysregulation markers in all 4 relevant studies and showed associations with very diverse outcomes, including suicidal ideation, suicide severity, risk, and attempts. Behavioral and mobility-related features derived from the global positioning system and accelerometers mapped to high-level markers that captured social withdrawal, physical inactivity, and sleep disturbance. These markers showed associations with suicidal ideation and suicide-related events. Smartphone use metrics mapped to smartphone use patterns and were linked to suicide-related events and risk indicators, including suicidal ideation and planning. Linguistic features extracted from text messages mapped to high-level markers that reflected expression related to suicide risk and psychological and behavioral patterns. These markers were associated with suicidal ideation, planning, and suicide attempts. Overall, mapping revealed that passive digital markers were represented across multiple hierarchical levels of the LHSF and linked to diverse suicide-related outcomes.

XSL•FO
RenderX

Discussion

Main Findings

This scoping review proposed a suicide-specific adaptation of the LHSF to integrate fragmented evidence from digital sensing studies into clinically meaningful behavioral markers in suicide research. The adapted LHSF organized diverse digital features into higher-order constructs, such as autonomic dysregulation, sleep disturbance, and psychosocial dysfunction, and provided a conceptual bridge between technical sensor outputs and clinically interpretable suicide-related states.

Physiological Indicators

Autonomic indicators, particularly HRV-related features, were significantly associated with suicide-related outcomes in all 4 relevant studies and achieved the highest predictive performance among the included studies (AUC=0.89) [33]. Beyond prediction, HRV and autonomic measures were repeatedly associated across the broadest range of suicide-related outcomes in correlational studies, including suicide severity, current suicidal ideation, and attempt history [24,29]. However, they should be interpreted alongside other behavioral and contextual markers [24,36]. This pattern is consistent with prior evidence linking autonomic dysregulation to suicide risk [36,37]. Suicidality was associated with dysregulation of the autonomic nervous system, reflected in reduced parasympathetic activity and altered HRV [36]. Electrodermal activity improved model performance for predicting the severity of suicidal ideation; however, it worsened model fit when the outcome was the presence of suicidal ideation [25].

Passive physiological monitoring, including HRV assessment using wearable devices, might complement conventional suicide risk assessments, particularly for individuals who did not accurately disclose their suicidal ideation or behavior. Passively collected physiological data could be processed into HRV indices that reflect autonomic nervous system regulation and dysregulation [38,39]. These physiological changes could be clinically meaningful, as substantial alterations in HRV might be interpreted as indicators of psychological distress. However, as predictive performance varied across studies and outcomes, such measures should be appropriately considered as objective correlates associated with suicide-related risk rather than stand-alone clinical markers.

Mobility and Geolocation Indicators

Mobility- and geolocation-derived indicators suggest that changes in spatial behavior may reflect proximal signals of suicide-related risk, particularly when considered as within-person deviations rather than absolute behavioral levels. Increases in homestay and reductions in mobility were associated with subsequent suicidal events, indicating that disruptions in daily movement patterns might capture short-term risk dynamics [30,32]. These behavioral shifts might represent early manifestations of psychosocial withdrawal and loss of daily structure, both well-established precursors of suicidal crises, rather than reflect isolated symptoms. Therefore, abrupt deviations from an individual's routine can be interpreted as signals of functional deterioration difficult to capture using conventional self-reporting methods.

However, the predictive value of geolocation features differed across indicators; measures such as entropy and distance traveled were not significantly associated with suicide outcomes [32]. These findings could suggest that mobility-based markers might be sensitive to contextual and individual variability and thus should be interpreted in relation to personalized baseline patterns. From a clinical perspective, this supports an approach that prioritizes within-person changes over static thresholds, reflecting a broader shift toward dynamic and temporally sensitive models for suicide risk assessment [28].

Smartphone Use Patterns

Although smartphone use patterns captured dynamic behavioral changes associated with short-term suicide-related outcomes, their interpretation remained highly context dependent. When individuals spend more time on their smartphones, they are more likely to experience near-term suicidal ideation and planning, suggesting that heightened device engagement might reflect acute psychological states preceding suicide risk [35]. Research indicated significant associations between specific app usage patterns and suicide risk, with certain types of media use being linked to elevated risk under specific conditions [23]. In addition, behavioral changes derived from passively collected smartphone usage data demonstrated the ability to predict suicide-related events over short time frames, including suicide attempts and psychiatric emergency visits [30].

As a result, smartphone-derived behavioral data could serve as real-time proxies for the underlying cognitive and emotional processes. However, digital behaviors were inherently ambiguous and thus might reflect multiple psychological functions, including coping, avoidance, and social interaction. Hence, smartphone-based indicators should be interpreted with caution and combined with other data sources such as EMA or clinical assessments, to improve their reliability and clinical interpretability.

Sleep Indicators

Few sleep-related indicators were significantly associated with suicidal outcomes [31]. Only one actigraphy-derived metric, wake after sleep onset, was significantly associated with next-day suicidal ideation, and only in 1 of 7 participants at the individual level; other sleep parameters were not significantly associated [31]. A likely explanation is methodological rather than substantive. As actigraphy infers sleep from movement, it detects sleep with high sensitivity but poorly identifies wakefulness, so periods of quiet wakefulness are often misclassified as sleep, and metrics such as wake after sleep onset are often underestimated [40]. Self-reported sleep, in contrast, may capture the subjective distress more proximally linked to suicidal states, which could explain why passively sensed sleep metrics may show fewer significant associations. These findings suggest that the associations observed here may reflect current measurement constraints rather than the absence of a meaningful link. Methodological refinements, such as validating movement-based estimates against polysomnography and averaging across multiple nights to obtain more stable

XSL•FO
RenderX

parameter estimates [41], together with larger samples and more consistent measurement strategies, may strengthen future findings and clarify the role of sleep-related digital markers.

Linguistic and Communication Markers

Linguistic and communication-based indicators derived from smartphone data provided additional insights into suicide-related cognitive and emotional states, although findings varied across specific features. Text-based patterns that reflected suicide-related content were significantly associated with both passive and active suicidal ideation and planning [34]. Furthermore, longitudinal analyses demonstrated that temporal changes in emotion-related language, such as increases in anger and decreases in positive emotions, could differentiate suicide attempts from lower-risk states [22]. These findings suggest that language use may capture shifts in affective and cognitive processes preceding suicidal behavior.

However, not all linguistic features were significantly associated with suicide-related outcomes. Self-focus and sleep-related language patterns did not reliably distinguish between suicide risk states in some analyses [22,27]. Overall, linguistic markers may provide valuable indicators of internal psychological states. However, their clinical utility depends on integrating multiple features and considering variability across individuals and contexts.

Clinical Implications

These findings highlight the potential of passive digital markers for clinically meaningful suicide risk monitoring. In practice, the continuous monitoring of physiological, behavioral, and digital signals helps expand our assessment beyond static, retrospective viewpoints toward more dynamic and temporally sensitive models of suicide risk [12]. Approaches that prioritize within-person deviations from individual baseline patterns may enable earlier detection of subtle changes preceding suicidal crises, particularly among those who may not disclose suicidal ideation or behavior [14,34].

However, the predictive performance reported across the included studies requires careful interpretation. Among the predictive studies, AUC ranged from 0.56 to 0.89. In clinical screening contexts, AUC values of 0.80 or above are generally considered the minimum threshold for adequate discrimination [42]. By this standard, only one study achieved excellent performance (AUC=0.89) [33], whereas one study approached but did not reach this threshold (AUC=0.78) [30]. The remaining studies reported AUCs below this threshold (range 0.56-0.74) [26,28,32], and one predictive study reported only relative model performance improvements rather than an AUC [25]. These findings indicate that passive digital sensing for suicide risk prediction is still at an early stage and does not yet meet the performance levels required for clinical screening tools. Nevertheless, these AUC values should be interpreted in light of the inherent difficulty of predicting rare events such as suicidal behaviors and the reliance on passive data alone without integration with clinical assessments.

From a clinical perspective, these findings suggest that passive digital monitoring systems may function as continuous and low-burden safety nets and identify emerging risks in everyday contexts. In addition, integrating these digital markers into intervention frameworks, such as Just-In-Time Adaptive Interventions, may facilitate timely and context-sensitive responses, supporting early intervention before suicidal crises intensify [43]. These approaches may be particularly useful in community and outpatient settings, where continuous monitoring using only conventional assessment methods is unfeasible. Specifically, our findings suggest that physiological markers, such as HRV, which showed the strongest predictive performance and the broadest range of associations in our review, combined with behavioral and linguistic indicators, may provide a more context-sensitive approach to suicide risk assessment.

Study Limitations and Future Research

This review has several limitations. First, the relatively small number of included studies (n=14) limits the generalizability of our findings and the ability to establish a robust consensus on digital markers. Second, the lack of standardized data collection and extraction of suicide-related digital features across studies may have contributed to inconsistencies in the findings and limited comparability between studies. This issue is further compounded by simplified device classifications that do not account for variability across device brands, models, or operating systems (eg, iOS vs Android). Third, as a scoping review, this study did not formally assess the methodological quality of the included studies or perform a quantitative synthesis, such as a meta-analysis. Fourth, the exclusion of gray literature may have introduced publication bias. Although peer-reviewed publications were targeted to ensure complete technical descriptions for LHSF mapping, relevant negative or null results may exist outside these databases, potentially narrowing our synthesized evidence base. Fifth, most studies (78.6%) were conducted in the United States, and included samples were predominantly White or Caucasian samples and mostly female participants. This demographic and geographic homogeneity limits the generalizability of the present findings to culturally and demographically diverse populations. These limitations reflect the emerging nature of research in passive digital sensing for suicide-related outcomes and highlight the need for more rigorous and standardized methodological approaches. However, this review provides a structured framework for selecting digital markers for suicide prevention and offers a foundation for future research in digital suicide phenotyping.

Several directions for future research have emerged based on this review. First, future studies should adopt prospective longitudinal or real-time designs to examine multimodal integration, combining passive sensor data with EMA and electronic health record data to enhance clinical context and interpretability. Second, large-scale validation studies using advanced analytic approaches, such as machine learning and natural language processing, are needed, along with rigorous external validation, to improve the discrimination between general psychological distress and suicidal crises. Third, future research should move beyond correlational and predictive modeling and focus on intervention-focused research, such as clinical trials to develop and test actionable digital phenotypes within intervention frameworks, such as Just-In-Time Adaptive

XSL•FO
RenderX

Interventions frameworks. These approaches may enable real-time clinical alerts or timely responses when specific digital markers are detected. Fourth, future reviews in the field of digital suicide phenotyping should consider incorporating gray literature to mitigate potential publication bias and establish a more comprehensive evidence base. Fifth, future research should prioritize the inclusion of demographically and geographically diverse samples to strengthen the generalizability and clinical translatability of passive digital markers across cultural contexts.

Conclusions

This scoping review provides a suicide-specific adaptation of the LHSF to organize fragmented evidence on passive digital sensing into clinically interpretable behavioral markers. Physiological indicators related to autonomic regulation were associated with suicide-related outcomes across all relevant studies and showed the strongest predictive performance (AUC=0.89). Furthermore, smartphone-derived behavioral markers demonstrated additional potential for capturing proximal changes in suicidal states, whereas linguistic markers showed more variable findings. In contrast, sleep- and activity-related indicators showed few significant associations. This review clarifies how passive digital features can be translated into higher-order behavioral constructs and advances the conceptual foundation for digital suicide phenotyping. These findings highlight the potential of integrating multimodal digital data to enable continuous, real-time monitoring of suicide risk in natural settings. Future research should focus on improving the reliability, interpretability, and external validity of digital markers to develop clinically actionable systems for early detection and timely intervention in high-risk populations.

Acknowledgments

During the preparation of this manuscript, ChatGPT (OpenAI) and Claude (Anthropic) were used to refine and edit the author's English language. All suggestions were thoroughly reviewed and verified by the authors. The authors take full responsibility for the entire manuscript.

Funding

This research was supported by the National Research Foundation of Korea funded by the Ministry of Science and Information and Communication Technology (RS-2025-00563996). This research was supported by the Brain Korea 21 FOUR Project funded by the National Research Foundation of Korea, Yonsei University College of Nursing.

Data Availability

The data charting matrix generated during this scoping review is publicly available on the Open Science Framework [19].

Authors' Contributions

HK secured funding. HK, WTC, and TTL contributed to the conceptualization and methodology. SK and JK curated the data and conducted the formal analyses. SK was responsible for project administration. HK supervised the study. WTC, TTL, YZ, and ZYAT validated the findings. HK and SK prepared the visualizations. All authors contributed to drafting the original manuscript. HK, SK, WTC, and TTL critically reviewed and edited the manuscript. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search queries.
[DOCX File, 18 KB - mhealth_v14i1e98616_app1.docx]

Multimedia Appendix 2
Operationalization criteria for assigning features to LHSF layers 2 and 3.
[DOCX File, 23 KB - mhealth_v14i1e98616_app2.docx]

Multimedia Appendix 3
Demographic and clinical sample characteristics of the included studies.
[DOCX File, 24 KB - mhealth_v14i1e98616_app3.docx]

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 88 KB - mhealth_v14i1e98616_app4.docx]

XSL•FO
RenderX

References

1. GBD 2021 Suicide Collaborators. Global, regional, and national burden of suicide, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet Public Health 2025 Mar;10(3):e189-e202. [doi: 10.1016/S2468-2667(25)00006-4] [Medline: 39986290]
2. Naghavi M, Ong KL, Aali A, GBD 2021 Causes of Death Collaborators. Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2024 May 18;403(10440):2100-2132. [doi: 10.1016/S0140-6736(24)00367-2] [Medline: 38582094]
3. Arowosegbe A, Oyelade T. Suicide risk assessment and prevention tools in the UK: current landscape and future directions. Psychiatry International 2023;4(4):354-369. [doi: 10.3390/psychiatryint4040032]
4. Franklin JC, Ribeiro JD, Fox KR, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. Psychol Bull 2017 Feb;143(2):187-232. [doi: 10.1037/bul0000084] [Medline: 27841450]
5. Sarkhel S, Vijayakumar V, Vijayakumar L. Clinical practice guidelines for management of suicidal behaviour. Indian J Psychiatry 2023 Feb;65(2):124-130. [doi: 10.4103/indianjpsychiatry.indianjpsychiatry_497_22] [Medline: 37063624]
6. Coppersmith DDL, Ryan O, Fortgang RG, Millner AJ, Kleiman EM, Nock MK. Mapping the timescale of suicidal thinking. Proc Natl Acad Sci U S A 2023 Apr 25;120(17):e2215434120. [doi: 10.1073/pnas.2215434120] [Medline: 37071683]
7. Hallensleben N, Glaesmer H, Forkmann T, et al. Predicting suicidal ideation by interpersonal variables, hopelessness and depression in real-time. An ecological momentary assessment study in psychiatric inpatients with depression. Eur Psychiatry 2019 Feb;56:43-50. [doi: 10.1016/j.eurpsy.2018.11.003] [Medline: 30530103]
8. Gratch I, Choo TH, Galfalvy H, et al. Detecting suicidal thoughts: the power of ecological momentary assessment. Depress Anxiety 2021 Jan;38(1):8-16. [doi: 10.1002/da.23043] [Medline: 32442349]
9. Sedano-Capdevila A, Porras-Segovia A, Bello HJ, Baca-García E, Barrigon ML. Use of ecological momentary assessment to study suicidal thoughts and behavior: a systematic review. Curr Psychiatry Rep 2021 May 18;23(7):41. [doi: 10.1007/s11920-021-01255-7] [Medline: 34003405]
10. Shen S, Qi W, Zeng J, et al. Passive sensing for mental health monitoring using machine learning with wearables and smartphones: scoping review. J Med Internet Res 2025 Aug 14;27:e77066. [doi: 10.2196/77066] [Medline: 40811794]
11. Torous J, Kiang MV, Lorme J, Onnela JP. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR Ment Health 2016 May 5;3(2):e16. [doi: 10.2196/mental.5165] [Medline: 27150677]
12. Büscher R, Winkler T, Mocellin J, et al. A systematic review on passive sensing for the prediction of suicidal thoughts and behaviors. Npj Ment Health Res 2024 Sep 23;3(1):42. [doi: 10.1038/s44184-024-00089-4] [Medline: 39313519]
13. Coppersmith DDL, Kleiman EM, Millner AJ, et al. Heterogeneity in suicide risk: evidence from personalized dynamic models. Behav Res Ther 2024 Sep;180:104574. [doi: 10.1016/j.brat.2024.104574] [Medline: 38838615]
14. Jacobucci R, Jones SG, Blacutt M, Ammerman BA. Passive vs active nighttime smartphone use as markers of next-day suicide risk. JAMA Netw Open 2025 Nov 3;8(11):e2542675. [doi: 10.1001/jamanetworkopen.2025.42675] [Medline: 41217755]
15. Karas M, Huang D, Clement Z, et al. Smartphone screen time characteristics in people with suicidal thoughts: retrospective observational data analysis study. JMIR Mhealth Uhealth 2024 Oct 11;12:e57439. [doi: 10.2196/57439] [Medline: 39392706]
16. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 May 8;13:23-47. [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
17. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
18. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
19. Kim S, Kim J, Chien WT, et al. Development and Implementation of AI-Driven Individualized Program of Digital Suicide Prevention: Open Science Framework; 2026. URL: https://osf.io/vndkz [doi: 10.17605/OSF.IO/VNDKZ]
20. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. Nat Biotechnol 2015 May;33(5):462-463. [doi: 10.1038/nbt.3223] [Medline: 25965751]
21. Peters MDJ, Godfrey CM, McInerney P, Munn Z, Tricco AC, Khalil H. Chapter 11: scoping reviews. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis: JBI; 2020. [doi: 10.46658/JBIRM-20-01]
22. Glenn JJ, Nobles AL, Barnes LE, Teachman BA. Can text messages identify suicide risk in real time? A within-subjects pilot examination of temporally sensitive markers of suicide risk. Clin Psychol Sci 2020 Jul;8(4):704-722. [doi: 10.1177/2167702620906146] [Medline: 35692890]
23. Coyne SM, Hurst JL, Dyer WJ, et al. Suicide risk in emerging adulthood: associations with screen time over 10 years. J Youth Adolesc 2021 Dec;50(12):2324-2338. [doi: 10.1007/s10964-020-01389-6] [Medline: 33528704]
24. Sheridan DC, Baker S, Dehart R, et al. Heart rate variability and its ability to detect worsening suicidality in adolescents: a pilot trial of wearable technology. Psychiatry Investig 2021 Oct;18(10):928-935. [doi: 10.30773/pi.2021.0057] [Medline: 34555890]

XSL•FO
RenderX

25. Kleiman EM, Bentley KH, Maimone JS, et al. Can passive measurement of physiological distress help better predict suicidal thinking? Transl Psychiatry 2021 Dec 2;11(1):611. [doi: 10.1038/s41398-021-01730-y] [Medline: 34857731]
26. Horwitz A, Czyz E, Al-Dajani N, et al. Utilizing daily mood diaries and wearable sensor data to predict depression and suicidal ideation among medical interns. J Affect Disord 2022 Sep 15;313:1-7. [doi: 10.1016/j.jad.2022.06.064] [Medline: 35764227]
27. Ladis I, Valladares TL, Coppersmith DDL, et al. Inferring sleep disturbance from text messages of suicide attempt survivors: A pilot study. Suicide Life Threat Behav 2023 Feb;53(1):39-53. [doi: 10.1111/sltb.12920] [Medline: 36083138]
28. Czyz EK, King CA, Al-Dajani N, Zimmermann L, Hong V, Nahum-Shani I. Ecological momentary assessments and passive sensing in the prediction of short-term suicidal ideation in young adults. JAMA Netw Open 2023 Aug 1;6(8):e2328005. [doi: 10.1001/jamanetworkopen.2023.28005] [Medline: 37552477]
29. Ortiz A, Park Y, MacLean S, et al. A history of suicide attempt is associated with increased sympathetic activation in bipolar disorder. Can J Psychiatry 2024 Feb;69(2):126-137. [doi: 10.1177/07067437231194334] [Medline: 37583363]
30. Barrigon ML, Romero-Medrano L, Moreno-Muñoz P, et al. One-week suicide risk prediction using real-time smartphone monitoring: prospective cohort study. J Med Internet Res 2023 Sep 1;25:e43719. [doi: 10.2196/43719] [Medline: 37656498]
31. Mournet AM, Ruork AK, Kleiman EM, Rizvi SL. An actigraphy case series: sleep disturbance associations with suicidal ideation, self-harm and desire to live. Psychiatry Research Case Reports 2025 Jun;4(1):100244. [doi: 10.1016/j.psycr.2025.100244]
32. Auerbach RP, Bloom PA, Pagliaccio D, et al. Using smartphone GPS data to detect the risk of adolescent suicidal thoughts and behaviors. JAMA Netw Open 2025 Jan 2;8(1):e2456429. [doi: 10.1001/jamanetworkopen.2024.56429] [Medline: 39869336]
33. Um J, Park J, Lee DE, Ahn JE, Baek JH. Machine learning models to identify individuals with imminent suicide risk using a wearable device: a pilot study. Psychiatry Investig 2025 Feb;22(2):156-166. [doi: 10.30773/pi.2024.0257] [Medline: 40017279]
34. Ammerman BA, Kleiman EM, O'Brien C, et al. Smartphone-based text obtained via passive sensing as it relates to direct suicide risk assessment. Psychol Med 2025 May 9;55(e144):e144. [doi: 10.1017/S0033291725001199] [Medline: 40340954]
35. Jacobucci R, Blacutt M, Ram N, Ammerman BA. Smartphone screen time and suicide risk in daily life captured through high-resolution screenshot data. NPJ Digit Med 2025 May 29;8(1):321. [doi: 10.1038/s41746-025-01740-w] [Medline: 40442251]
36. Alacreu-Crespo A, Sebti E, Moret RM, Courtet P. From social stress and isolation to autonomic nervous system dysregulation in suicidal behavior. Curr Psychiatry Rep 2024 Jun;26(6):312-322. [doi: 10.1007/s11920-024-01503-6] [Medline: 38717659]
37. Bellato A, Admani MA, Deak C, et al. Autonomic dysregulation and self-injurious thoughts and behaviours in children and young people: A systematic review and meta-analysis. JCPP Adv 2023 Sep;3(3):e12148. [doi: 10.1002/jcv2.12148] [Medline: 37720589]
38. Hongn A, Bosch F, Prado LE, Ferrández JM, Bonomini MP. Wearable physiological signals under acute stress and exercise conditions. Sci Data 2025 Mar 28;12(1):520. [doi: 10.1038/s41597-025-04845-9] [Medline: 40155406]
39. Olivieri F, Biscetti L, Pimpini L, Pelliccioni G, Sabbatinelli J, Giunta S. Heart rate variability and autonomic nervous system imbalance: Potential biomarkers and detectable hallmarks of aging and inflammaging. Ageing Res Rev 2024 Nov;101:102521. [doi: 10.1016/j.arr.2024.102521] [Medline: 39341508]
40. Taibi DM, Landis CA, Vitiello MV. Concordance of polysomnographic and actigraphic measurement of sleep and wake in older women with insomnia. J Clin Sleep Med 2013 Mar 15;9(3):217-225. [doi: 10.5664/jcsm.2482] [Medline: 23493815]
41. Fekedulegn D, Andrew ME, Shi M, Violanti JM, Knox S, Innes KE. Actigraphy-based assessment of sleep parameters. Ann Work Expo Health 2020 Apr 30;64(4):350-367. [doi: 10.1093/annweh/wxaa007] [Medline: 32053169]
42. Hosmer DW Jr, Lemeshow S, Sturdivant RX. Applied Logistic Regression, 3rd edition: John Wiley & Sons; 2013. [doi: 10.1002/9781118548387]
43. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]

Abbreviations

AUC: area under the curve
EMA: ecological momentary assessment
HRV: heart rate variability
LHSF: Layered Hierarchical Sensemaking Framework
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
SDG: Sustainable Development Goals

XSL•FO
RenderX

Edited by A Stone; submitted 17.Apr.2026; peer-reviewed by NH Zainal; revised version received 12.Jun.2026; accepted 26.Jun.2026; published 13.Aug.2026.

Please cite as:
Kim S, Kim J, Chien WT, Luk TT, Zhang Y, Tan ZYA, Park E, Kim H
Conceptualization From the Sensors to Suicide-Related Outcomes: Scoping Review Based on Layered Hierarchical Sensemaking Framework
JMIR Mhealth Uhealth 2026;14:e98616
URL: https://mhealth.jmir.org/2026/1/e98616
doi:10.2196/98616

©Sohee Kim, Jinyeong Kim, Wai Tong Chien, Tzu Tsun Luk, Yu Zhang, Zhen Yang Abel Tan, Eunju Park, Heejung Kim. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 13.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobile Health in Home-Based Palliative Care: Realist Review of Contextual Factors and Mechanisms Influencing Outcomes

Nuhamin Tekle Gebre[1,2], MD; Oladayo Afolabi[1], PhD; Nahla Gafer[1], MD; Kennedy Bashan Nkhoma[1], PhD; Nicola Ayers[1], PhD; Charlotte Hanlon[3,4], MD, PhD; Richard Harding[1], PhD

[1]Cicely Saunders Institute, Florence Nightingale Faculty of Nursing, Midwifery and Palliative Care, King's College London, Bessemer Road, London, United Kingdom

[2]Global and Planetary Health Working Group, Institute of Medical Epidemiology, Biometrics, and Informatics, Center of Health Sciences, Medical Faculty, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany

[3]Division of Psychiatry, Institute for Neuroscience and Cardiovascular Research, University of Edinburgh, Edinburgh, United Kingdom

[4]Department of Psychiatry, School of Medicine, College of Health Science, Addis Ababa University, Addis Ababa, Ethiopia

Corresponding Author:
Nuhamin Tekle Gebre, MD
Cicely Saunders Institute, Florence Nightingale Faculty of Nursing, Midwifery and Palliative Care, King's College London, Bessemer Road, London, United Kingdom

Abstract

Background: The increasing prevalence of chronic illness presents a significant global health challenge due to growing end-of-life suffering. Palliative care is now an essential health service under Universal Health Coverage. Its integration into primary health care and use of mobile health (mHealth) have been recommended to improve access.

Objective: This study aimed to synthesize evidence on how mHealth interventions in home-based palliative care work, for whom, and in what context.

Methods: This realist review identified global evidence relevant to mHealth interventions for home-based palliative care through searches of 9 electronic databases (ie, MEDLINE, Embase, PsycINFO, Global Health, CINAHL, Web of Science, Cochrane Database of Systematic Reviews, Scopus, and Global Index Medicus), and forward citation tracking of included articles conducted in March 2026. The review followed the 5 key stages of a realist review: scoping the literature to assess what is important about the context of mHealth intervention in home-based palliative care and what mechanisms might be important in how such interventions result in their intended outcomes; articulating the underlying program theory and refining the review scope through consultation with international palliative care experts; conducting iterative searches for and appraisal of relevant evidence; extracting data; and narratively synthesizing the data, prioritized by relevance and rigor to generate conclusions and recommendations. The Framework of Complexity in Palliative Care Context, adapted from Bronfenbrenner's Ecological Systems Theory, was applied to examine the contextual influences and interactions among factors within the multilayered system. Context-mechanism-outcome configurations were developed and iteratively tested to refine the program theory for mHealth in home-based palliative care.

Results: A total of 4134 records were identified, of which 422 (10.21%) articles were retained for full-text screening, and 126 (3.05%) studies were included in the final synthesis. The contextual factors and mechanisms that positively influence the intended outcomes include (1) alleviating concerns and mitigating perceived threats about mHealth's suitability in palliative care through proper orientation for patients and carers, along with clear guidelines for health care professionals; (2) minimizing infrastructural and technological barriers through user-friendly designs and investment in sustainable models of mHealth for home-based palliative care; (3) engaging diverse stakeholders to ensure mHealth aligns with priority health care needs, user preferences, and the operations of implementing organizations and the wider health care system; (4) streamlining health management information systems through interoperable mHealth systems implemented at different levels of care; (5) enabling access to and regular communication with relevant health care teams; and (6) facilitating access to adequate information to empower users in palliative care services.

Conclusions: mHealth interventions can enhance home-based palliative care but must align with local contexts. It is recommended that mHealth interventions ensure safety and comfort, technology competence, tailored communication, empowerment of end users, family involvement, health care worker motivation, and system integration.

Trial Registration: PROSPERO CRD42022369443; https://www.crd.york.ac.uk/PROSPERO/view/CRD42022369443

(JMIR Mhealth Uhealth 2026;14:e80321) doi:10.2196/80321

XSL•FO
RenderX

KEYWORDS

community palliative care; home-based palliative care; digital health intervention; mobile health; mHealth intervention; telehealth

Introduction

The World Health Organization (WHO) advocates for digitally enabled care pathways to improve health care delivery through accurate data capture, enhanced communication, and care coordination [1]. Digital health interventions (DHIs) encompass a wide range of information and communications technologies that support health and health-related activities [2]. Mobile health (mHealth) is a subset of DHIs that uses mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices, in medical and public health practice [3,4]. In palliative care, this has been achieved through the use of patient-reported outcome measures and electronic health recordings [5,6]. During the COVID-19 pandemic, mHealth facilitated the remote monitoring and management of patients' symptoms, ensuring continuity of palliative care delivery [7,8].

mHealth technologies have demonstrated a positive impact on patient education, information sharing, communication, and clinical decision-making, as well as saving costs [9]. However, significant heterogeneity has been reported in the results of prior systematic reviews of mHealth effectiveness [3,10]. Challenges in the implementation of mHealth interventions include poor access to mobile devices and networks, usability and interoperability issues, and user concerns about data security and privacy [11,12]. Moreover, mHealth interventions are complex due to their multiple interacting components, the interplay of contextual factors, and the range of systems within which they are implemented [13]. To date, there is a paucity of research developing theoretical models of how these complex interventions improve health, psychosocial, and organizational outcomes or how contextual factors may influence their effectiveness [4]. Such evidence is needed to inform the development and evaluation of mHealth interventions for home-based palliative care and to ensure they can be successfully integrated into existing services.

Palliative care is an essential health service under Universal Health Coverage. However, of 56.8 million people who require the service every year, only an estimated 12% of them have access to it. Currently, most provision is in high-income countries, while 80% of the need is in low- and middle-income countries (LMICs) [14,15]. Palliative care aims to "improve the quality of life for patients and their families facing problems associated with life-threatening illness [16,17], through the prevention and relief of suffering by means of early identification and impeccable assessment and treatment of pain and other problems, physical, psychosocial and spiritual" [18-21].

Among the main strategies identified to enhance access to and improve coverage of palliative care are its integration into primary health care and the use of mHealth technology. Integrating health services within primary care is recognized as the most inclusive, equitable, and cost-effective strategy to achieving comprehensive health care coverage [22]. This approach enhances equitable access to services, especially for patients in remote or underserved areas, by enabling them to receive care in the comfort of their own homes. Furthermore, it ensures continuity of care, providing consistent and coordinated support throughout the illness trajectory, which contributes to better patient outcomes.

Leveraging technological interventions, such as mHealth into primary health care, enables timely and effective support for those in need, particularly patients in underserved regions [13,23,24]. The aim of this review was to determine how mHealth interventions improve both access to and the quality of home-based palliative care and to identify the contextual factors and mechanisms of action underlying their intended outcomes.

Methods

Design

In this realist review, we undertook a theory-informed synthesis of evidence on how, why, and under what circumstances mHealth works in home-based models of palliative care. The rationale for the design was that realist reviews provide an explanatory focus to reveal the mechanisms by which complex interventions work in relation to various contexts, allowing the development of a model that demonstrates how and why interventions work [25,26]. Realist methodology generates program theories using context-mechanism-outcome configurations (CMOCs), which are the main output of the review. Contexts (C) describe the circumstances in which the intervention is implemented, mechanisms (M) describe how the intervention works within specific contexts, and the interaction between these will lead to a particular outcome (O) [27,28].

This review adhered to the 5 main stages of a realist review, comprising scoping the literature, articulating underlying program theory and refining the scope of the review with stakeholders, searching for and appraising relevant evidence, extracting data, and synthesizing data to draw conclusions and recommendations [29]. The Realist and Meta-Narrative Evidence Synthesis: Evolving Standards were used to guide reporting [27,28]. The protocol for this review was registered on PROSPERO (registration CRD42022369443).

Procedure

Scoping the Literature to Develop Preliminary Hypotheses

First, a scoping search was undertaken to explore the context of palliative care needs at home, what is important about supporting such needs, how end users might engage with mHealth interventions, and the mechanisms underpinning home-based care. Nine electronic databases were systematically searched: MEDLINE, Embase, PsycINFO, and Global Health via Ovid Technologies, CINAHL via EBSCO Information Services, Web of Science via the Clarivate Web of Science platform, the Cochrane Database of Systematic Reviews via

XSL•FO
RenderX

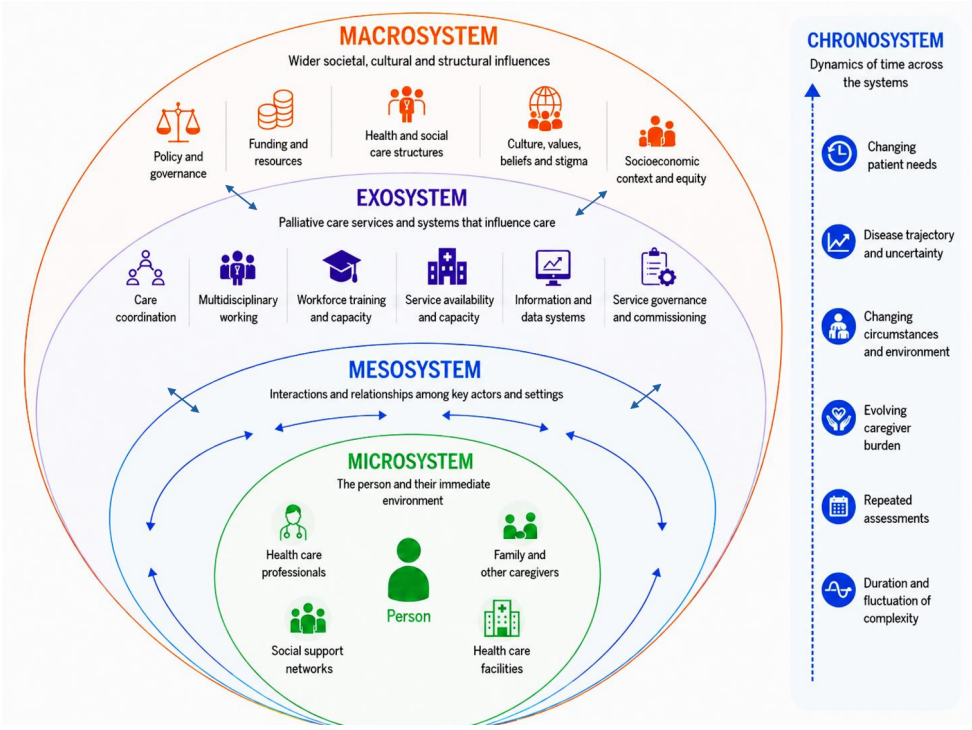
Cochrane Library, Scopus via the Elsevier Scopus, and Global Index Medicus via the WHO platform. The search applied a broad structure: [mHealth] AND [palliative care] AND [home-based care], combining subject headings and free-text searches of titles and abstracts. Search results were uploaded to an online data management software, Covidence [30]. The first author (NTG) screened all records for relevance to the topic of mHealth intervention in home-based palliative care, while the second and third reviewers (OA and NG) independently screened 50% of the identified articles, and any disagreement was resolved by team discussion. After title, abstract, and full-text screening, the authors familiarized themselves with the included studies and identified common themes. Preliminary *"if...then"* statements were developed to articulate how, why, and under what circumstances mHealth interventions influence home-based palliative care. These statements informed the initial context-mechanism-outcome hypotheses [29].

Articulating Underlying Program Theory and Refining the Scope of the Review With Stakeholders

The preliminary hypotheses generated from the scoping search were discussed with 14 global palliative care researchers and health care professionals through individual face-to-face and online meetings [31]. These discussions emphasized the importance of understanding the complex systems in which mHealth interventions are implemented, including their impact on technology engagement and subsequent outcomes.

To explore these contextual factors, we applied the Framework of Complexity in Palliative Care Contexts [32] adopted from Bronfenbrenner's Ecological Systems Theory [25,26]. This framework highlights interconnected environmental levels and provides a systemic perspective to analyze variations in context, multisectoral actors, and environments across different levels of the health care system (Figure 1). Insights from stakeholder consultations guided the refinement of the review scope and informed the subsequent iterative search strategy.

Figure 1. A framework of complexity in the palliative care context. Key: These levels are at the microsystem (person, needs, and characteristics), the chronosystem (dynamic influences of time), the mesosystem (interactions with family and health care professionals), the exosystem (palliative care services or systems), and the macrosystem (societal influences) (adapted from Pask et al [32], which is published under Creative Commons Attribution-NonCommercial 4.0 International License [33]).



Identify and Appraise Relevant Evidence

Following stakeholder consultation, an iterative evidence database search was conducted to identify studies relevant to the refined scope. The aim of this supplementary searching was to reach a point of theoretical saturation, where no new relevant evidence was generated for the particular theory building or testing within the synthesis [29].

The iterative search applied refined search terms informed by the scoping search and reassessment of the retained studies. In addition to the broader terms, specific functions of mHealth interventions commonly used in home settings, such as telehealth, telemedicine, and videoconferencing, were included. Searches were conducted in the same 9 electronic databases, with no restriction by year in March 2024 (full search strategy reproduced in Multimedia Appendix 1). The search results were uploaded and managed in Covidence. Only papers published in English were considered due to the practical constraints of engaging reviewers fluent in other languages.

In March 2026, the search was updated with forward citation searching of the final included studies using Google Scholar to identify more recent publications. Screening was conducted following predefined selection criteria; the first author (NTG) screened all titles and abstracts, while OA and NG independently

XSL•FO
RenderX

reviewed 50% of records. Discrepancies were resolved through team discussions.

Eligibility Criteria

The predefined selection criteria are described in Table 1.

Table . Summary of the eligibility criteria for all studies included in the review.

Eligibility domain	Inclusion criteria	Exclusion criteria
Participants or population	Patients, both adult and pediatric of all ages, with life-threatening conditions[a] or life-limiting illnesses[b], which are chronic conditions, which significantly impair quality of life and daily functioning, necessitating ongoing management to alleviate symptoms and provision of supportive care (eg, advanced dementia, progressive neurological diseases, and end organ failures [17]), or their informal caregivers (ie, *unpaid providers of physical, social, practical, and emotional support who may be blood or non–blood relatives* [34]), or formal (paid) health care professionals, community health workers, or health extension workers.	Patients solely treated within acute care or receiving only disease-modifying or curative treatment; and caregivers and health care providers with no palliative care roles
Intervention	mHealth[c] is a subset of DHIs[d] that uses mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices in medical and public health practice [1,2,35]. Studies were included if they featured the practice of palliative care supported by text message, voice or video call, or application-based or web-based technology (eg, through purposely designed apps or social media platforms) using mobile phones or tablets.	Interventions that do not use mHealth or with a limited mHealth application as part of a multifaceted intervention
Context	Home-based palliative care as a service delivered to seriously ill patients in the setting that a patient calls “home” (ie, a private residence) [36,37]. Studies were included irrespective of country or region to ensure representation of diverse sociocultural contexts.	Facility-based services (eg, primary health care facilities, hospitals), nursing homes, long-term care homes, and hospices (eg, outpatient, day care, and in patient)
Outcome(s)	All outcomes associated with mHealth-supported home-based palliative care; relating to patients, informal caregivers, or health systems.	No restriction

[a]Life-threatening conditions refer to conditions associated with a high risk of mortality.

[b]Life-limiting illnesses refer to chronic or progressive conditions that substantially impair quality of life and daily functioning, requiring ongoing symptom management and supportive care.

[c]mHealth: mobile health.

[d]DHI: digital health intervention.

Quality Assessment Based on Relevance and Rigor

Appraisal Process and Prioritization

The relevance of studies in this review was assessed by considering whether they contained sufficient detail and/or theoretical discussion to inform how and in what context mHealth interventions support home-based palliative care, that is, their contribution to generating program theory [29,38]. In accordance with previous research, the relevance of contributions was defined as high (well aligned to the review questions, with rich descriptions contributing to theory), medium (with some relevance to the review question and some insights to CMOCs), or low (meeting broad inclusion criteria and providing at least one insight) [29].

The Mixed Methods Appraisal Tool (MMAT) [39] was used to assess the methodological quality and rigor of the included studies, as it is useful to assess the quality of qualitative, quantitative, and mixed methods studies. The MMAT comprises 5 quality criteria, with 3 responses for each item (ie, positive [=1] and negative or unavailable information [both=0]). The 5 items are added with a score of 5 indicating high quality. The Joanna Briggs Institute quality assessment tool was used to assess the methodological rigor of the included systematic reviews, with scores allocated on a scale from 0 to 11. Using this tool, a score of 8 or more indicates high quality, a score of 6 to 8 indicates moderate quality, and a score less than 6 suggests low quality. Three authors (NTG, NG, and OA) independently assessed the methodological quality of the studies and met to discuss their appraisal. Differences were resolved through discussion with the entire research team [40].

As a realist review approach may use only relevant data from part of an article, no studies were excluded based on

XSL•FO
RenderX

methodological quality. However, studies with high relevance and rigor were given priority during theory development [27-29,31].

Data Extraction

Bespoke data extraction forms were developed to organize information obtained from the included studies. A careful review of the introduction, methods, results, and discussion sections assisted in identifying relevant information. Descriptive characteristics, such as author, year, study aims, design, population, and country, were recorded. Details were also extracted on the intervention components, contextual factors, implied or actual mechanisms attributed to the outcome of interest, and the intervention outcomes. Data extraction was undertaken by 2 of the authors (NTG and NG) working independently. Areas of discrepancy were resolved through discussion to reach consensus.

Data Synthesis

Evidence sources were synthesized narratively, prioritizing included studies by relevance and rigor [41] and shown as a set of CMOCs. Taking an iterative approach, the first author read and reread different sources of data and coded the evidence to identify prominent recurrent concepts concerning contexts, mechanisms, and outcomes [29]. Coding was done deductively to the concepts in the Framework of Complexity in Palliative Care Context and using the priority "if...then" statements, which informed the preliminary context-mechanism-outcome hypotheses. In this process, we drew logical conclusions from the observed patterns, particularly the contextual factors influencing mHealth interventions [32] and inductively (proposing insights from recurring observations), mainly regarding mechanisms, and retroductive coding (exploring new patterns using theory to propose causal pathways) [42]. This process allowed the development of mHealth program theory in palliative care and conceptualization of the causal pathways using contexts, mechanisms, and outcomes [43,44].

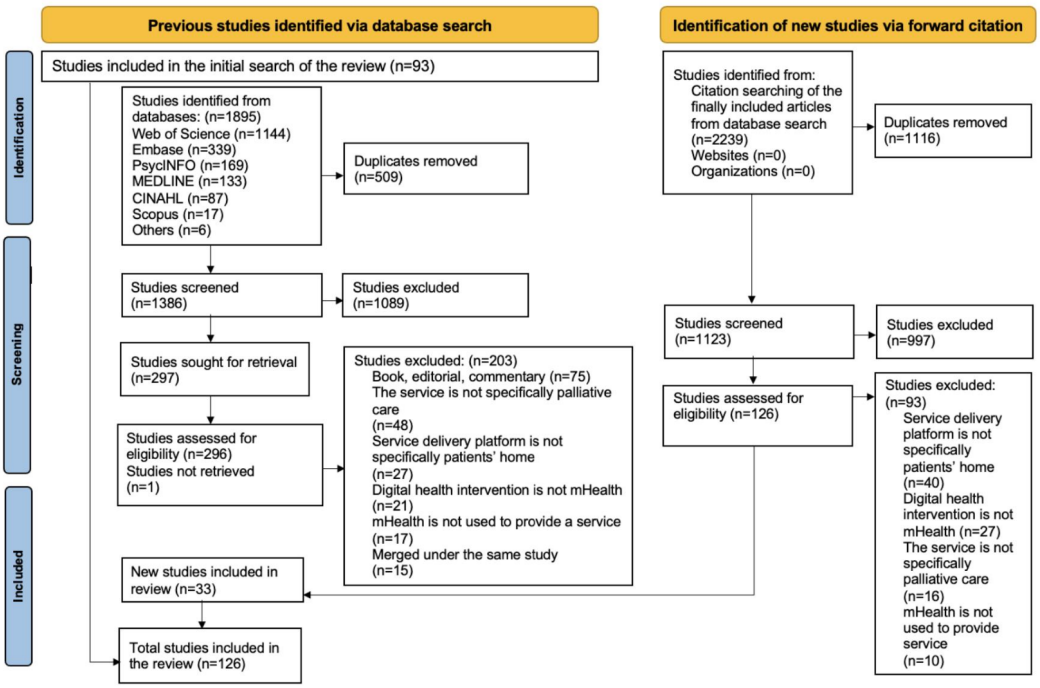
Results

Search Yield and Characteristics

Overview

In total, 4134 studies were identified via database searching (n=1895, 45.8%) and forward citation searching of the final included studies in March 2026 (n=2239, 54.2%). The selection process and included articles are illustrated in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram (Figure 2). Following the removal of duplicates and initial screening, 422 (10.21%) articles were retained for full-text screening, and 126 (3.05%) studies were included in the final synthesis, and of these, 20 (0.48%) articles from the initial scoping search contributed to the initial preliminary hypothesis generation. Details of the studies included in the final synthesis are presented in Table 2.

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of articles included in the review. mHealth: mobile health.



XSL•FO
RenderX

Table . Studies included in the review.

Author, year, country	Aim	Study design	Intervention
Abahussin et al, 2023, United Kingdom [45]	To design and develop an mHealth[a]-based pain recording system to support effective pain management in cancer.	User-centered design and usability testing	A mobile app and a web-based app for adult patients with cancer in home settings to record pain details. The data were accessed by health care professionals and researchers.
Allsop et al, 2019, 21 countries in Africa (see Multimedia Appendix 1) [35]	To assess the current use of mHealth in palliative care service delivery in the African region, identify barriers to mHealth use, and determine provider priorities for research development.	Mixed methods descriptive design	Mobile phone.
Al-Mondhiry et al, 2022, United States [46]	To use community-partnered participatory research to co-design and pretest a mobile app that focuses on palliative care priorities of clinicians and patients with advanced cancer.	Co-design workshops	Mobile app, that is, wellness journal to track patient-reported symptoms, goals, and medication use; information on self-management of symptoms; community resources; and patient and caregiver testimonial videos.
Ansari et al, 2022, United States [47]	To identify technology-based communication strategies to improve health outcomes in individuals with advanced cancer.	Integrative review	Mixed technologies, for example, health information technology or electronic health reporting, data sharing systems, and telehealth programs.
Aoki et al, 2006, Japan [48]	To evaluate the implementation of telepalliative care in a rural community for patients with terminal cancer who prefer home-based palliative care.	Mixed methods study	Telepalliative care using a Polycom view station with ISDN[b] line for communication, including telephone calls among health care providers and patients, and recording of vital signs.
Archer et al, 2021 [49]	To explore the impact of digital health interventions on the psychological outcomes of patients and families receiving pediatric palliative care.	Systematic review	Telehealth consultations via video link and telephone, telehealth intervention via Zoom videoconferencing or FaceTime, and the use of a bespoke website (My-Quality.net) to collect data on patient-reported outcomes.
Baldwin et al, 2012, United States [50]	To determine what organizational factors might account for differences in the use of telehospice.	Survey	Telehospice (ie, video conferencing) used by hospice agency staff and patients.
Balasubramanian et al, 2022, India [51]	To assess the level of satisfaction of patients receiving e-palliative homecare services and to evaluate the feasibility of e-palliative care in providing palliative care to patients with advanced cancer at their homes.	Prospective observational study	Implementation of e-palliative care through a technological interface, where a homecare team led by a palliative care–trained nurse visits patients with a camera-integrated laptop and e-palliative care software app installed, allowing remote communication with a palliative care physician in the hospital.
Bandini 2025, United States [52]	To examine current perspectives from interdisciplinary providers on the use of telehealth and in-person care for outpatient palliative care among underserved patients.	Qualitative study	Telehealth use among staff serving underserved population.

XSL•FO
RenderX

Author, year, country	Aim	Study design	Intervention
Basile et al, 2024 [53]	To map the literature on the use of technology by older adults with terminal illness receiving palliative care at home and to identify the technology systems in use, the impact of technology on communication between palliative care professionals and patients, and explore the strengths or weaknesses perceived by patients regarding the use of technology.	Scoping review	Tablets, smartphones, computers, videoconferencing devices, a pen with an embedded camera, and devices for remote symptom monitoring such as smartphones and wristbands.
Bauer et al, 2026, Denmark [54]	To develop program theory for a family-focused model of telepalliative care for adult patients and their families.	Qualitative study—workshops with key stakeholders, participant observations, and ethnographic interviews	Telepalliative consultation perceptions among patients with palliative care needs, family caregivers, SPCT[c] members, allied health care personnel, community care nurses, IT personnel, and health care managers from both community and tertiary care settings.
Bensink et al, 2004, Australia [55]	To design and test an internet-based videophone for pediatric palliative care services, focusing on assessing the feasibility and benefits of using the technology for telemedicine in home settings.	Feasibility study with qualitative exploration	Videophone: use of low-bandwidth, internet-based videophone for supporting pediatric palliative care in the home, with clinical consultations conducted twice each week.
Bensink et al, 2009, Australia [56]	To investigate the acceptability of providing videotelephone-based support to families receiving pediatric oncology-related palliative care and to explore the integration of videotelephone support from the time of diagnosis through outpatient care and support.	Randomized controlled trial	Access to a videotelephone service in addition to the usual 24-h on-call telephone support service, provided by specialist nurses for patient assessment, monitoring, family education, communication, and counseling.
Bethel et al, 2021, United States [57]	To evaluate technology system usability and Comprehensive Patient Assessment for Using Telehealth at Home implementation effectiveness in seriously ill older adults and their caregivers.	Qualitative study	Comprehensive Patient Assessment for Using Telehealth at Home intervention to meet remotely and assess patients in their own home.
Bhargava et al, 2021, Canada [58]	To develop and evaluate a remote symptom self-reporting app (ie, RELIEF) for community patients with palliative care needs, assessing its feasibility and acceptability for remote monitoring of patients through regular symptom self-reporting, with a focus on using technology to improve symptom management and care delivery.	Co-design and feasibility study	Patients self-reported symptoms daily using RELIEF and received over-the-phone assessments and interventions based on alerts, and some received support for expected death in the home.
Bonsignore et al, 2018, United States [59]	To describe and evaluate a telehealth palliative care program using the TapCloud app and videoconferencing in rural areas, focusing on feasibility, usability, symptom burden, and hospice transitions.	Mixed methods study	TapCloud app for remote patient monitoring, including symptom and medication management, messaging, and photo upload, as well as videoconferencing for real-time interactions.
Bradford et al, 2014, Australia [60]	To compare the costs of pediatric palliative care medical consultations conducted via the Home Telehealth Program with in-person consultations.	A cost minimization analysis	Palliative care consultations through the Home Telehealth Program at the Royal Children's Hospital in Brisbane, involving video consultations in the home.

XSL•FO
RenderX

Author, year, country	Aim	Study design	Intervention
Bradford et al, 2014, Australia [61]	To investigate the potential of telemedicine for the delivery of specialist support in the home or local community for children with life-limiting conditions requiring palliative care.	Qualitative study	Real-time audiovisual communication links between the specialist pediatric palliative care team and families caring for a child receiving palliative care at home, facilitated by personal computers, web cameras, and the internet.
Brännström et al, 2024, Sweden [62]	To describe health care professionals' experiences of video consultations in palliative care in community homecare and nursing homes in rural areas.	Qualitative study	Video consultations to access specialist palliative care service.
Burner-Fritsch et al, 2023, Germany [63]	To identify and explore challenges for the development of electronic patient-reported outcome measures as standardized assessment in specialized palliative home care, focusing on patient, health care professional, and organizational or structural levels.	Qualitative study	Patients used the electronic version of the Integrated Palliative Care Outcome Scale on their own web-enabled devices, completing it 1 to 2 times a week for at least 2 weeks.
Caetano et al, 2024, Portugal [64]	To explore whether caregivers feel prepared to provide informal palliative home care, their experiences, and the usefulness of telehealth in managing daily activities.	Qualitative study	Telehealth to support informal caregivers.
Calton et al, 2020, United States [65]	To characterize the experience of patients with serious illness and their caregivers receiving palliative care by telemedicine.	Mixed-methods telephone survey	Telemedicine acceptability and comfort among patients and caregivers.
Cameron 2021, United States [66]	To examine the comfort and emotional support that caregivers of home hospice patients derive from using electronic tablets for telehospice.	Mixed methods study	Caregivers received electronic tablets with the AVA[d] app for telehospice participation, were given a presurvey, shown how to use the tablets, completed surveys after setting up AVA in the home, and received a postsurvey 2 wk after the death of the loved one.
Carey et al, 2023, Ethiopia [67]	To co-design, develop, and evaluate a mobile phone–based remote monitoring system for use by palliative care patients in Ethiopia.	Co-design and qualitative study	Remote monitoring system where patients and families record their symptoms and access information to self-support their needs.
Christiansen et al, 2020, Sweden [68]	To describe perceptions of mHealth and its impact on HRQoL[e] among older adults with cognitive impairment.	Qualitative study applying phenomenography	mHealth intervention (SMART4MD) to improve the HRQoL of older adults with mild cognitive impairment and their informal caregivers.
Chua et al, 2022, United States [69]	To describe trends in in-person and video visit palliative care delivery challenges before and during the COVID-19 pandemic in the United States.	Quantitative nonrandomized	Early integrated telehealth palliative care (via video visit) or in-person palliative care, with at least monthly visits; initial visit in person within 4 wk of enrollment but could be conducted via video during the COVID-19 pandemic; subsequent follow-up visits every 4 wk, either via video or in person.

XSL•FO
RenderX

Author, year, country	Aim	Study design	Intervention
Collier et al, 2016, Australia [70]	To explore clinicians' perspectives and experiences of telehealth, perspectives about benefits and challenges for patients, families, and clinicians in using telehealth applications, and identified enablers and barriers for the use of telehealth in community palliative care.	Qualitative study	The intervention included self-report assessment tools, remote activity monitoring, video-based conferences, and virtual case conferences.
Collier et al, 2025, Australia and Aotearoa, New Zealand [71]	To explore the complexities of telehealth and palliative care with a focus on access and equity.	Qualitative study	Telehealth for extending reach of palliative care in underserved groups.
Cook et al, 2001, United States [72]	To analyze the successes and challenges of implementing telehospice care focusing on the role of senior executives' support, quality of care, and provider acceptance.	Observational study	Implementation of telehospice services using telemedicine technology, including training of potential telehospice users and providing care remotely via videophones over ordinary telephone lines to patients in rural and urban communities in Kansas and Michigan for at least a year.
Coyle et al 2002, United States [73]	To explore the effectiveness and benefits of using audiovisual communication as a complementary tool in caring for patients with complex palliative care needs who are isolated at home due to their medical conditions, aiming to enhance communication and support for both the patients and their caregivers.	Case report of a 3-mo trial	Trial of a commercially available audiovisual system, specifically a videophone (TeleEye 324 TM; LeadTek Research, Inc), placed in the patient's home using standard telephone wires for a 3-mo trial period.
Cruz et al 2023 [74]	To map the available knowledge regarding the use of apps for mobile devices to support adult patients in palliative care at home, specifically focusing on mHealth apps under study or development in the home care transition process for palliative care patients.	Systematic review	Participants received the mobile app intervention for monitoring symptoms, communication with health care professionals, and potentially receiving alerts and educational materials related to palliative care. The intervention was compared with controls in some studies, and outcomes were assessed in terms of changes in daily functioning, self-efficacy, and quality of life.
Davis et al, 2015, United States [75]	To describe the telecaring intervention and its outcomes, focusing on proactively identifying and addressing the needs of patients receiving hospice care at home and caregivers through a systematic phone-based care service.	A quality improvement initiative	TeleCaring consisted of daily proactive phone calls to patients and caregivers by specialists and nurses, made up to twice daily based on the patient or caregiver's preferred frequency and timing, providing up to 14 additional staff interactions each week. The calls aimed to address questions or concerns related to care and ensure supply needs were met.
Dickman Portz et al 2020, United States [11]	To explore provider perspectives regarding the utility of mHealth in palliative care.	Qualitative study	mHealth perception of use among providers from multiple disciplines working in palliative care settings.
Disalvo et al, 2021 [76]	To identify digital health technologies evaluated for supporting timely assessment and management of people with palliative care needs at home and their carers.	Meta-review	Remote patient monitoring interventions included monitoring of biometric data or symptoms, with alerts triggered to clinicians when patient scores were outside acceptable thresholds.

XSL•FO
RenderX

Author, year, country	Aim	Study design	Intervention
Dismore et al, 2025, United Kingdom [77]	To understand the experiences of patients, carers, and specialist palliative care professionals to receive or deliver video consultations in a rural setting.	Qualitative study	Video consultation perception of use to access specialist palliative care.
Dolan et al, 2021 [78]	To assess the safety and effectiveness of virtual care in end of life and palliative care.	Review	Virtual care, that is, video consultation, mobile apps, videos, websites, telephone support, email, and alert messages.
Domínguez et al, 2026, Ecuador [79]	To evaluate the technology usability and implementation feasibility of the nonintrusive devices for telemedicine system, a low-cost, smartwatch-based pain monitoring solution for palliative cancer care co-designed with health care staff from a cancer hospital in Ecuador.	Usability and feasibility study	Wearable-based pain monitoring system.
Doolittle et al, 2019, United States [80]	To explore the feasibility and effectiveness of using a TeleHospice approach in rural hospice care, focusing on enhancing patient- and family-centered, timely care.	Product design—community-engaged model	Use of low-cost, secure mobile tablets (iPads) for videoconferencing encounters in rural hospice care settings, involving an interprofessional team.
Evering et al, 2022, The Netherlands [81]	To explore the intention to use video communication by health care providers in interprofessional terminal care settings.	Cross-sectional survey	A video communication in terminal care.
Gonzalez, 2023, United States [82]	To evaluate the perception of usability of a tablet device among palliative care patients and informal caregivers.	Cross-sectional pilot study	A home health tablet device.
Gordon et al, 2022 [83]	To explore strategies for increasing access to palliative care among individuals living in remote or rural communities through the use of telehealth applications.	Rapid review	The interventions included videoconferencing, telephone-based interventions, web-based platforms, after-hours phone support service, remote patient monitoring and management, web-based videoconference platforms, and 24/7 phone consult services.
Guo et al, 2023, China [84]	To examine telehealth readiness and its related factors among Chinese palliative care specialist nurses.	Cross-sectional study	Telehealth (readiness assessed using the Chinese version of Telehealth Readiness Assessment Tools and the Innovative Self-Efficacy Scale).
Hancock et al, 2019 [85]	To describe the current use of telehealth in palliative care in the United Kingdom, evaluate telehealth initiatives against a digital service standard, and explore whether telehealth results in a reduction in emergency care access.	Systematic review	The intervention(s) was home telemonitoring using telephone or computer software to record clinical symptoms or signs from the patient's home. Patients were required to input specific data regarding symptoms and physical parameters, with some studies including telephone support.
Harding et al, 2021, India, Uganda, and Zimbabwe [9]	To design and evaluate a mobile phone app to improve communication between family caregivers, community caregivers, and palliative care teams, assess its acceptability and mechanisms of action, and propose refinements based on stakeholder feedback.	Mixed methods study	The intervention involved the use of a mobile phone app for family or community caregivers to report nonurgent patient and family outcomes on a dashboard accessible by the clinical palliative care team in 3 LMICs. The app aimed to facilitate patient assessment, improve staff awareness of patient concerns, enhance communication, and potentially improve patient outcomes.

XSL•FO
RenderX

Author, year, country	Aim	Study design	Intervention
Haroen et al, 2025 [86]	To investigate the benefits of integrating telehealth into community-based palliative care for people living with terminal illness and their caregivers.	Review	Telehealth, that is, telephone or video consultations and phone calls.
Haydon et al, 2021, Australia [87]	To explore the costs, service activity, and staff experiences resulting from the introduction of telehealth in a community palliative care service in Queensland, Australia.	Mixed methods study	In-person and telephone-based home visits and telehealth (video consultations).
Hebert et al, 2007, Canada [88]	To determine the potential for using videophones in palliative home care and to explore factors that could inform the eligibility criteria for video visits.	Retrospective chart review	Palliative home care visits that could have been carried out using videophones, telehospice care as a support for family caregivers, testing the usability of a videophone in palliative home care, and determining the potential for home telehealth in palliative care.
Hennemann-Krause et al, 2015, Brazil [89]	To examine telemedicine as a form of home and additional support for traditional outpatient care as a way to remotely monitor and manage the symptoms of patients with advanced cancer.	Prospective, longitudinal, qualitative, descriptive, and case series study	Web conferences to remotely monitor and manage the symptoms.
Hoek et al, 2017, The Netherlands [90]	To determine whether weekly teleconsultations from a hospital-based SPCT improved patient-experienced symptom burden compared to "care as usual" in home-dwelling patients with advanced cancer.	Randomized clinical trial	Weekly teleconsultations for a period of 13 wk in addition to their usual care. First, a teleconsultation device was installed at the patient's home. Patients who had not visited the SPCT before were evaluated at the outpatient clinic or during a home visit by one of the SPCT members.
Holland et al, 2014, United States [91]	To determine the resources needed to use the system and the quality of the audio and visual components to conduct virtual visits between a clinician at an academic center and community-dwelling adults living in rural locations.	Mixed-methods field study design	Virtual visits using a 3G-enabled Apple iPad, cellular phone data service, and a web-based video conference service.
Holmen et al, 2020 [92]	To identify and review the use of eHealth to communicate and support home-based pediatric palliative care.	Convergent, systematic mixed methods review	Interventions received by the study participants included videoconferences, the use of an online tool for monitoring quality of life, and the use of iPads for videoconferencing, among others.
Hutchinson et al, 2022, United States [93]	To explore the acceptability, feasibility, and emotional responsiveness of telepalliative care.	A mixed-methods formative study	Telemedicine perception of use, that is, acceptability, feasibility, and emotional responsiveness of telepalliative care.
Jess et al, 2019 [94]	To review and synthesize current evidence regarding the use of video consultations in both general and specialized palliative care to various patient groups.	Systematic integrative review	Use of video consultations in specialized palliative care settings
Jiang et al, 2023, Australia [95]	To assess the feasibility of a TH-SPC[f] model for an Australian rural setting, focusing on integrating telehealth technology to provide specialist palliative care to rural communities.	Prospective mixed methods pilot study	Video consultations with metropolitan-located specialist palliative care physicians alongside standard care.

XSL•FO
RenderX

Author, year, country	Aim	Study design	Intervention
Johnston, 2011 [96]	To explore the use of telehealth in relation to palliative care in the United Kingdom.	Review	Telehealth use for older people.
Johnston et al, 2012, United Kingdom [97]	To evaluate the current use of telehealth applications within palliative care across Scotland and assess how these applications are perceived by patients, carers, and professionals.	Qualitative evaluation design	Telehealth activity among the key stakeholders centered on the use of videoconferencing equipment for networking, consultations, and multidisciplinary meetings, and for education and training.
Johnston 2014 [98]	To scope information available from published and unpublished research on the current state of palliative home-based technology, practitioner-focused perspectives, patient-focused perspectives, quality of life, and the implications for clinical practice.	Review	Home-based technology perspectives of use among clinicians and patients and implications on practice.
Funderskov et al, 2019, Denmark [99]	To explore the advantages and disadvantages of using video consultations in SPC[g], focusing on the experiences of health care professionals and their perspectives on video consultations.	Qualitative study	Patients referred to specialized palliative care received an initial home visit by an SPC team head physician and an SPC team nurse. They were primarily in contact with the SPC team nurse by phone and had approximately one weekly video consultation using a tablet provided by the SPC team.
Kawashima et al, 2026 [100]	To evaluate the effectiveness of and implementation requirements for telehealth in palliative care patients with advanced cancer.	Review	Telehealth ie, videoconferencing, telephone, or text-based communication.
Keenan et al, 2021, United Kingdom [101]	To explore potential divergence and convergence in relation to health care professionals' and patients' acceptability of the use of telehealth within palliative care provision through the lens of Self-Determination Theory.	Qualitative study	Telehealth perceptions among health care professionals and patients.
Kidd et al, 2010 [102]	To review telehealth applications used in palliative care settings in the United Kingdom, identify who is using telehealth, for what purpose it is being used, and to explore if telehealth use is increasing within the clinical setting.	Systematic review	The interventions for health professionals, patients, carers, nurses, and GPs[h] include telehealth services such as out-of-hours telephone support, advice services, videoconferencing, and training and education for health care staff.
Kluger et al, 2024, United States [103]	To determine whether palliative care training for neurologists and remote access to a palliative care team can improve outcomes in patients with Parkinson disease and related disorders in community settings, focusing on patient quality of life, caregiver burden, and other patient-centered outcomes.	A pragmatic, stepped-wedge comparative effectiveness trial	Palliative care education for community neurologists and provision of team-based palliative care resources via telehealth, including consultations from an academic center's palliative care team.
Kwon et al, 2026, Republic of Korea [104]	To (1) develop the Hospice@Home system, a digital in-home hospice care solution; (2) explore preliminary indications of usability and feasibility among patients with terminal cancer and their caregivers; and (3) identify challenges for future implementation.	Development and usability pilot study	Digital in-home hospice care solution for patients with terminal cancer and their caregivers.

XSL•FO
RenderX

Author, year, country	Aim	Study design	Intervention
Middleton-Green et al, 2019, United Kingdom [105]	To evaluate the Gold Line telehealth initiative in providing support to individuals at home in the last year of life, focusing on patient and carer experiences and the extension of palliative care services to those not known to specialist palliative care.	Mixed methods evaluation	Intervention components: Gold Line service, training for health care teams, use of the Gold Standards Framework, access to electronic health records, recording call outcomes, and participation in interviews.
Lee et al, 2024, Ireland [106]	To understand patient experience.	Mixed quantitative and qualitative study	Video consultation—perception of use among patients.
Lind et al, 2008, Sweden [107]	To explore and describe palliative home care patients' experiences of assessing their pain using a pain diary together with digital pen and mobile Internet technology.	A case study with cross-case content analysis	A pain diary together with digital pen and mobile Internet technology. Included introduction of pain management clinical guidelines, pain treatment for patients with at least moderate pain, and follow-up of pain treatment using the IT system developed.
Lundereng et al, 2023 [108]	To systematically map published studies on HCPs'[i] experiences and perspectives on the use of telehealth in HBPC[j].	Scoping review	Video-based technology for teleconsultations among patients, families, and health care professionals, delivered by hospital-based HCPs, hospice workers, and home care professionals either independently or in collaboration with hospital-based staff.
Ma et al, 2025 [109]	To provide a comprehensive understanding of the roles of nurses and the multilevel facilitators and barriers to implementing nurse-delivered telehealth in home-based palliative care, which could inform future policy development, research, and clinical practice.	Review	Telehealth for home-based palliative care delivered by nurses.
Mathews et al, 2023 [110]	To describe the components of telehealth palliative care interventions for patients with advanced cancer before the COVID-19 pandemic, identify intervention components associated with improvements in outcomes, and evaluate the reporting of interventions.	Scoping review	Hybrid models, telephone-based interventions, videoconference-based interventions, asynchronous web-based intervention, mainly consisting of psychosocial content and delivered by nurses in the home setting. Participants included patients, family caregivers, and patient-caregiver dyads.
Maguire et al, 2025, United Kingdom [111]	To explore the usability, user experiences, and impact of the digital dyadic remote monitoring Care and Support System for Patients and Carers for patients in the last year of life, their informal carers, and health professionals involved in their care.	Mixed methods feasibility study	Digital dyadic remote monitoring Care and Support System for Patients and Carers for patients in the last year of life.
Maudlin et al, 2006, United States [112]	To determine if a telehealth model of care would benefit veteran patients with life-limiting illness by managing their physical, emotional, functional, and spiritual care needs during the last 2 y of their lives and fostering earlier enrollment into hospice through provider education on palliative care.	Pilot testing of the Advanced Illness/Palliative Care program	Technologies used included both video (telemonitors and videophones) and nonvideo (disease management and messaging) devices.

XSL•FO
RenderX

Author, year, country	Aim	Study design	Intervention
Mccall et al, 2008, United Kingdom [113]	To test the acceptability and usability of the ASyMSp[k] in monitoring and managing symptoms of patients receiving palliative care at home.	Explorative descriptive study	A mobile phone–based technology (ASyMSp) to monitor and manage symptoms reported by patients receiving palliative care at home, involving daily completion of an electronic questionnaire over a 30-d period.
Menon et al, 2015, United States [114]	To investigate the feasibility and describe patient outcomes of palliative care consultations via telemedicine in critically ill patients, assess the impact on decreasing the burden on family members, and provide care in alignment with patient wishes.	A retrospective review	Telemedicine palliative care conferences.
Miller et al, 2021 [115]	To explore the existing literature on pediatric TM-HBPC[l] for children with serious illness to identify research, determine the effectiveness, and highlight the need for more robust information on implementation, adaptation, and maintenance of TM-HBPC models.	Systematic scoping review	The intervention involved the delivery of home-based palliative care for children with serious illness via telemedicine, typically conducted by a hospital-based pediatric palliative care multidisciplinary team, with varying session durations and scheduling, and the use of different platforms for telemedicine consultations.
Morse et al, 2021, Tanzania [116]	To design and develop mPCL[m], a web and mobile app to support outpatient symptom assessment and care coordination and control, with a focus on pain.	Human-centered iterative design framework	Participants received the mPCL app for symptom assessment and care coordination, with some being provided loaned Android devices with the app preinstalled and personal smartphones installed the app. The intervention lasted for 4 mo.
Mukamal et al, 2025 [117]	To explore the experience of nononcologic older adult patients in palliative care with telemedicine	Review	Telemedicine to enhance palliative care for older adults.
Namasivayam et al, 2022 [118]	To review and map the available evidence on the use of telehealth in providing after-hours palliative care services in rural and remote Australia.	A scoping review	The intervention was the use of telehealth services provided by RNs[n], nurse specialists, GPs, specialist physicians, after-hours coordinators, and managers to enable patients to continue receiving after-hours palliative care at home in rural and remote Australia.
Narvaez et al, 2022 [119]	To map and evaluate existing research on the use of telemedicine and telehealth in palliative care patients during the COVID-19 outbreak.	Integrative review	Telehealth interventions including video conferencing, remote monitoring, and novel ways to improve care at home.
Nguyen et al, 2020, United States [120]	To compare a tech-supported model with video consultations to a standard model of home-based palliative care to determine non-inferiority in terms of interprofessional team coordination and patient outcomes.	Pragmatic, cluster-randomized non-inferiority trial	The intervention included nurse-led home visits coupled with remote physician consultation supervision via video for patients and caregivers in the tech-supported arm. Training on video visit procedures was provided to clinicians.
Oakley-Girvan et al, 2021, United States [121]	To develop a mobile app (TOGETHERCare) that enables cancer survivors' care partners to monitor the survivors' health and provide resources for the care partners.	Product design study with user involvement	Use of the TOGETHERCare mobile app by cancer care partners for a 28-day feasibility field test to remotely monitor cancer survivors' health and their own health, as well as to provide care partner resources.

XSL•FO
RenderX

Author, year, country	Aim	Study design	Intervention
Oelschlagel et al, 2021, Norway [122]	To explore municipal health care professionals' experiences with implementing an app for remote care in palliative home care for patients with cancer.	Qualitative, descriptive, exploratory design	Implementation of a technological solution named "remote home care" in palliative home care for patients with cancer, including providing a tablet device, installing digital medical devices.
Parker Oliver and Demiris, 2004, United States [123]	To assess the readiness of hospice organizations to accept technological innovation, specifically focusing on the current use of information and communication technology by hospice employees and their receptiveness to telemedicine or other advanced technology in patients' homes.	Cross-sectional survey study	Broad ICT[o] use—no specific intervention.
Oliver et al, 2012, United States [124]	To assess the state of evidence related to telehospice services and to highlight the need for strengthening the evidence base through various research methodologies.	Mixed methods study	Telephone advice lines, videophones, personal digital assistants, pen tablets, computers, comprehensive training for providers in the implementation of telehospice projects.
Olver et al, 2005, Australia [125]	To understand nurses' experience.	Qualitative study	Videophone communication among rural palliative care nurses and their more remote general nursing colleagues.
Osuji et al, 2020, United States [126]	To assess the usefulness and appropriateness of video visits in home-based palliative care from the perspective of physicians and nurses, anchored within the NASSS[p] framework.	Comparative effectiveness trial	Video visits were used primarily for follow-up visits as opposed to admission/start-of-care; intervention: video visits for follow-up care in home-based palliative care, facilitated by a nurse connecting patients with remote physicians.
Pare´ et al, 2009, Canada [127]	To evaluate the effects of a telehomecare intervention implemented in an oncology and palliative care unit in Quebec, Canada, focusing on nurses' satisfaction with the software app, perceived quality of care, and individual and group productivity.	Pre-post evaluation design	Implementation of the SyMO software with modules for care planning, developing treatment plan, nurse assignment, visit planning, route guidance, and documentation.
Parker et al, 2021, Australia [128]	To evaluate the usefulness of CarerHelp from the perspective of health professionals caring for individuals with life-limiting illnesses.	Quantitative study	A web-based online toolkit for carers.
Paul et al, 2019, Canada [129]	To gain a preliminary understanding of the experience of using mobile web-based videoconferencing for conducting in-home palliative care consults with older rural patients with life-limiting illness, assessing the clinical effectiveness, acceptability, service impact, and technical feasibility.	A descriptive, exploratory, proof-of-concept study	Palliative care consultations via mobile web-based videoconferencing with a palliative care physician consultant from their homes, with the presence of a palliative care clinical nurse specialist, and a home care nurse to manage the technology and support the patient. Participants also received training in the use of the technology.
Peñarrubia - San - Florencio et al, 2025, Spain [130]	To explore family perspectives and experiences with telehealth in PPC[q], focusing on their needs, perspectives, concerns, and hopes to refine digital care models.	Qualitative study	Telehealth to support families caring for children with palliative needs.

XSL•FO
RenderX

Author, year, country	Aim	Study design	Intervention
Phillips et al, 2008, Australia [131]	To evaluate the effectiveness of a generic after-hours telephone support service for palliative care patients and their caregivers in a rural setting in Australia, as well as to explore the application of a computer-based Palliative Care Clinical Information System to assist with the transfer of patient information.	Mixed methods study	After-Hours Telephone Support Service provided to palliative care patients and their caregivers.
Phongtankuel et al, 2018, United States [12]	To examine informal caregivers' receptivity and interest in using mHealth apps, along with identifying the app features caregivers perceived to be most useful in-home hospice care.	Qualitative study	mHealth apps with features such as video chat, medication/symptom tracking, and educational materials to improve communication and information exchange in home hospice care.
Pinto et al, 2017 [132]	To analyze the use of eHealth technologies and mobile apps in palliative care, focusing on the strengths, weaknesses, opportunities, and threats of these technologies.	Integrative literature review	eHealth technologies in palliative care used by adults, patients, nurses, physicians, family, carers, and older people.
Pinto et al, 2017, Portugal [133]	To introduce a web-based app for monitoring comfort in patients receiving palliative care.	Multiphase electronic app development process with pilot design	Monitoring comfort remotely using a web-based app, providing self-reported data on 11 comfort-related items on a Likert scale; using the app for 15 consecutive days, inputting data at least once a day; contacting the palliative care team in case of significant symptomatic distress; accessing the app through Android or Windows operating systems on smartphones, tablets, or computers; using the app to report symptoms and distress, with researchers monitoring the data.
Putranto and Rochmawati, 2020 [134]	To explore and synthesize evidence on cancer symptom management mobile apps for patients and their families, highlighting the role of nurses and the potential extension of mobile apps for managing other life-limiting illnesses.	Scoping review	Mobile apps for managing cancer symptoms at home. Nurses are assigned to observe and assess patient symptoms through the mobile app. The main function of the mobile app system is to record an assessment of patient symptoms, provide education and skills training, and offer video media adjusted to the reported symptoms.
Rana et al, 2025, United States [135]	To further explore adult patient and caregiver perspectives on telehealth, specifically the benefits of telehealth, trade-offs with in-person appointments, and the impact on patient care and family end-of-life preparations.	Qualitative study—single-arm prospective pilot study	Telehealth to support palliative care services for children and young adults with cancer.
Royackers et al, 2016, Canada [136]	To examine the experiences of informal caregivers who cared for a family member receiving palliative care as part of the eShift model of home care, focusing on their satisfaction with care delivery and ability to support their family members to die, at home.	Interpretive description methodology with semistructured interviews	eShift model of home-based palliative care using RNs, PSWs[r], and point-of-care technology with education provided to PSWs in palliative care.

XSL•FO
RenderX

Author, year, country	Aim	Study design	Intervention
Sadang et al, 2023, India [137]	To investigate barriers and facilitators to implementing Early Integrated Palliative Care via telehealth by surveying palliative care clinicians participating in a large-scale multisite palliative care trial comparing in-person and video visit-based services.	Multisite randomized trial	Early Integrated Palliative Care via secure video and/or in-person visits by advanced practice providers or board-certified physicians specializing in palliative care and/or oncology, at least monthly, beginning within 12 wk of diagnosis of advanced lung cancer.
Salem et al, 2020, Lebanon [138]	To assess caregiver and provider perceptions of the safety and efficacy of the Distance Support Program of a home-based palliative care provider.	Qualitative study	Distance Support Program including linking patients to a primary provider, face-to-face encounter, 24/7 telephone access, provider-initiated contact, and communication via telephone, messaging, or videophone.
Salimian et al 2019, Iran [139]	To develop the initial version of a mobile app for cancer palliative care for the purpose of improving the quality of life of cancer patients.	Mixed methods study	The mobile app "Ghasedak" for cancer palliative care, which includes functionalities such as user training on cancer and palliative care, clinician appointment reminders, personal health tracking, user guide, app settings, and patient notes.
Saysell and Routley, 2003, United Kingdom [140]	To address the national shortage of consultants in palliative medicine in the United Kingdom by implementing a telemedicine project to provide access to consultant advice and support for clinical nurse specialists in palliative care.	Mixed methods study	Weekly 2-h teleconference sessions between CNSs[s] and a consultant in palliative medicine to discuss patients with unresolved symptoms or significant palliative care issues.
Septian et al, 2025 [141]	To systematically synthesize the existing evidence from systematic reviews regarding: (1) the benefits of telehealth use on the quality of care and (2) the impact of telehealth on the quality of life of terminally ill patients.	Review	Telehealth to support terminally ill patients.
Schmitt et al, 2022, United States [142]	To describe the demographics of pediatric patients receiving telepalliative care, assess patient and family satisfaction with telepalliative care, and explore motives for using telepalliative services.	Retrospective EMR[t] chart review	Telepalliative visits.
Schoppee et al, 2020, United States [143]	To evaluate the acceptability of using the PAINReportIt completed on Wi-Fi-enabled tablets in the homes of patients with end-stage cancer and their caregivers for reporting pain related to end-stage cancer.	Pretest/posttest study	The PAINReportIt tool on a Wi-Fi-enabled tablet in patients' homes for reporting pain in real time to hospice nurses, with patients and caregivers completing the tool daily for a week.
Schuessler and Glarcher, 2024 [144]	To describe formal and informal caregivers' opportunities in telepalliative care and explore the associated ethical and safety challenges.	Integrative review of qualitative and quantitative studies.	The interventions included teleconsultation, supported by a step-by-step teleconsultation intervention protocol, telehealth options via telephone or videophone, web-based information, or smartphone apps for patient self-management, and the ENABLE CHF-PC[u] intervention involving face-to-face outpatient palliative care consultation and monthly telephone care coach sessions.

XSL•FO
RenderX

Author, year, country	Aim	Study design	Intervention
Slavin-Stewart et al, 2020, Canada [145]	To assess the feasibility of using FaceTime over a mobile cellular data network to provide improved access to specialist palliative care for patients in rural Nova Scotia, focusing on using low-cost technology to enhance access to palliative care services in underserviced areas.	Observational study	Participants received regular home-based visits from a palliative care nurse who used the FaceTime app to connect with the palliative care physician in Halifax. The nurse used an Apple iPad with the FaceTime app and a cellular data package during these visits.
Steindal et al, 2023 [146]	To critically appraise and synthesize evidence on patients' use of telehealth in home-based palliative care, focusing on the advantages and challenges experienced by patients.	Systematic mixed studies review with a convergent design.	Telehealth technology for home-based palliative care, including video consultations in synchronous mode for teleconsultations and remote monitoring in asynchronous mode for symptom management.
Stern et al, 2012, Canada [147]	To explore family caregivers and palliative cancer patients' patterns of use, perceptions of, and experiences with home telehealth in the context of palliative cancer care.	Mixed methods case study with a focus on qualitative methods	The intervention included specialist nurses available 24 h per day who communicated with patients and families using videophones, with optional remote monitoring of vital signs such as blood pressure, blood oxygen levels, and heart, lung, and abdominal sounds. Communication was done via telephone or videophone as required.
Sutherland et al, 2020 [148]	To evaluate the use of video consultations as an alternative to face-to-face visits in palliative care during the COVID-19 pandemic.	Systematic review	Video consultations using various technologies, cost reduction for patients, telehealth interventions (electronic patient records, telephone advice consultations, and telehealth monitoring), and home visits.
Tasneem et al, 2019, United States [149]	To assess patient perceptions on the impact of telemedicine video visits on care and to evaluate the ease of use and impact on health management of videoconference telemedicine for palliative care oncology patients in an urban setting.	Qualitative study	Telemedicine video visits conducted through web-based face-to-face videoconferences.
Tarbi et al, 2025, United States [150]	To explore if and how recommended in-person best practices for establishing human connection are adapted to the telehealth palliative care setting to enable discovery and description of practice innovations in this new care environment.	Qualitative study	Telepalliative consultations that support clinician connection-building communication practices in rural telepalliative care encounters.
Tieman et al, 2016, Australia [151]	To explore how telehealth can provide more regular engagement, continuous monitoring, and immediate access to resources for patients and families ultimately aiming to improve the efficiency and effectiveness of palliative care delivery in the community.	Prospective cohort study	Video-based conferences, virtual case conferences, self-report assessment tools, and remote activity monitoring.
Toenne et al, 2025, Germany [152]	To evaluate the acceptance and challenges of implementing teleconsultations among specialized outpatient pediatric palliative care health care professionals.	Survey	Telehealth.

XSL•FO
RenderX

Author, year, country	Aim	Study design	Intervention
Tong et al, 2022, United States [153]	To evaluate the experiences of remote volunteer palliative care consultants during the initial COVID-19 surge, examining motivations, emotional toll, lessons learned, and providing insights for future efforts in extending remote palliative care services.	Qualitative study	A palliative care virtual consultation program.
van Gurp et al, 2013, The Netherlands [154]	To describe elements of both the physical workplace and the cultural-social context of the palliative care practice, which are imperative for the use of teleconsultation technologies.	Qualitative study	Synchronous, audiovisual teleconsultation via webcam and iPad for patients with palliative care needs at home.
van Gurp et al, 2016, The Netherlands [155]	To investigate how teleconsultation supports the integration of primary care, specialist palliative care, and patient perspectives and services, as well as how patients and informal caregivers experience collaboration through teleconsultation.	Qualitative study	Teleconsultation between home-based patients with palliative care needs and a hospital-based SPCT conducted synchronously and audiovisual, including tripartite consultations between the patient, primary care physician, and SPCT.
Salvador Verges et al, 2022 [156]	To describe the current use of telemedicine in palliative care and assess stakeholders' views on the initiatives implemented worldwide regarding digital service standards.	Scoping review	Telemedicine for patient and caregiver support, professional education, and routine follow-up in palliative care.
Vincent et al, 2022, Canada [157]	To explore the experiences and perceptions of community palliative care providers, patients, and caregivers who delivered or received virtual palliative care as a component of home-based palliative care during the COVID-19 pandemic.	Qualitative study using phone and video-based semistructured interviews	Virtual palliative care as a component of home-based palliative care during the COVID-19 pandemic.
Washington et al, 2009, United States [158]	To measure telehospice acceptance in a national sample of hospice professionals from various disciplines and to explore the usefulness and ease of use of telehospice technology among different disciplinary backgrounds.	Cross-sectional study	The intervention was the use of videophone technology in the workplace for telehospice services, as demonstrated to participants through a web-based streaming video.
Weaver et al, 2021, United States [159]	To determine the acceptability of telehealth inclusion of a familiar pediatric palliative care provider during the first 2 home-based hospice visits to children, families, and adult-trained home hospice nurses in rural settings.	Case series	Virtual transition to home program with the inpatient pediatric palliative care physician joining the home hospice nurse's in-person visit as a telehealth presence using a FaceTime screen on day 1 and day 14 at home.
Whitten et al 2004, United States [160]	To investigate the use of telehospice services from patients' perspective, focusing on how telehospice is used, reasons for patient declination, and patient satisfaction and preferences regarding telehospice care.	Mixed methods	Telehospice services provided via videophones transmitting real-time video and audio images through the analog phone system, including nursing, social work, and spiritual care televisits.
Williams et al, 2024, United States [161]	To compare the provision of pediatric palliative care by delivery method.	A retrospective chart review	Telehealth for pediatric palliative care.
Wood et al, 2025 [162]	To synthesize how, why, and in which circumstances can videoconferencing be used to successfully meet the emotional support needs of adults receiving palliative care?	Realist review	Video consultations, ie, video calls among adult patients and the bereaved to support emotions by palliative care organizations.

XSL•FO
RenderX

Author, year, country	Aim	Study design	Intervention
Xu et al, 2023 [163]	To evaluate users' satisfaction with telehealth palliative care during COVID-19 and to identify facilitators and barriers to telehealth implementation in palliative care during COVID-19.	Systematic mixed studies review	Telehealth services for palliative care using digital communication and information technologies, such as computers, tablets, and mobile phone devices.
Zelko et al, 2022, Austria [164]	To test the effectiveness of a new tool for exchanging information between professionals in palliative care on the primary health care level, with a focus on improving interdisciplinary cooperation and communication.	Mixed methods study	A smartphone tool for exchanging information.
Zheng et al, 2016 [165]	To evaluate caregiver outcomes related to palliative telehealth interventions, with a focus on assessing the effectiveness of telehealth interventions for caregivers of patients in home-based palliative care.	Systematic review	Regular phones for counseling, internet-based interventions, and unspecified telehealth devices were used in the study to provide support and communication to caregivers in the palliative care setting.
Zimmermann et al, 2023, Germany [166]	To identify the needs and concerns of stakeholders regarding the introduction of telehealth in PPHC[v] and to define recommendations for telehealth in PPHC.	Mixed methods approach with a cross-sectional design	Telehealth interventions including video consultations, data transmission, access to patient records, symptom questionnaires, and communication support were provided to the study participants to improve the quality of care in pediatric palliative home care.

[a]mHealth: mobile health.

[b]ISDN: Integrated Service Digital Network.

[c]SPCT: specialist palliative care consultation team.

[d]AVA: Angela's Virtual Assist.

[e]HRQoL: health-related quality of life.

[f]TH-SPC: telehealth-assisted home-based specialist palliative care.

[g]SPC: specialized palliative care.

[h]GP: general practitioner.

[i]HCP: health care professional.

[j]HBPC: home-based palliative care.

[k]ASyMSp: Advanced Symptom Management System in Palliative Care.

[l]TM-HBPC: telemedicine home-based palliative care.

[m]mPCL: mobile-Palliative Care Link.

[n]RN: registered nurse.

[o]ICT: information and communication technology.

[p]NASSS: Non-adoption or Abandonment of technology by individuals and difficulties achieving Scale-up, Spread and Sustainability.

[q]PPC: pediatric palliative care.

[r]PSW: personal support worker.

[s]CNS: clinical nurse specialist.

[t]EMR: electronic medical record.

[u]CHF-PC: Concurrent Palliative Care Model for Older Adults with Heart Failure.

[v]PPHC: Paediatric Palliative Home Care.

Publication dates ranged from 2001 to 2026. Studies from high-income countries were conducted in the Americas (n=45, 35.71%), Europe (n=26, 20.63%), Australasia (n=12, 9.52%), and Asia (n=2, 1.58%). The LMICs represented were from Africa (n=4, 3.17%, representing 22 countries), the Middle East (n=2, 1.58%; ie, Iran and Lebanon), and Asia (n=2, 1.58%; ie, China and India) [9,35,116]. Studies featured qualitative design, including participatory co-design studies and case studies (n=48, 38.09%), mixed method studies (n=25, 19.84%), reviews (n=33, 26.19%), and quantitative studies, which included descriptive (n=14, 11.11%), randomized (n=4, 3.17%), and nonrandomized (n=2, 1.58%) trials.

The majority of the studies met most of the quality criteria (Multimedia Appendix 1). Regarding MMAT scores, the studies scored between 0 and 5 criteria, with a median score of 4.5 (IQR

XSL•FO
RenderX

3-5). For systematic reviews and meta-analyses, the median score was 9 (IQR 8-10) on the Joanna Briggs Institute quality assessment tool.

Focus of Included Studies

Overview

mHealth emerged as a key strategy for managing symptoms and improving palliative care delivery [52,62,64,70,77]. A recurring theme across studies was the use of technology to facilitate better communication [11,46,50,100,104,125] and care coordination among patients, caregivers, and health care providers [9,59,134]. Several studies concentrated on the psychological impact of digital interventions for both patients and families, highlighting their potential to alleviate distress and improve well-being [45,68,82,116,122,139,162]. End user experiences were critically assessed, with studies delving into patient and caregiver acceptance [93,98,101,106,109], satisfaction, and usability challenges [74,80,99,105,107,121]. Barriers to and facilitators of the wider adoption and integration of telehealth (ie, the use of electronic information and telecommunication technologies to support long-distance health care) into palliative care were also explored, identifying both technical and systemic influencers [41,127,130,132,135,141,150,152,161,167]. The applicability of telehealth in rural [62,117,168] or resource-constrained settings was reported with feasibility and potential to bridge gaps in health care delivery were examined [133,136,143]. Several studies further analyzed the resource utilization and cost-effectiveness of telehealth [49,145,151]. The role of mHealth in pediatric palliative care was also highlighted, as well as the rapid implementation and effectiveness of telehealth interventions during the COVID-19 pandemic [48,51,73,95,126,159].

Type of Technology Used

The studies assessed a broad range of electronic devices and described the features for the software used. mHealth devices used were mobile phones or mobile apps (n=23, 18.25%) [9,12,35,45,59,68,74,80,82,99,105,107,116,121,122,127,132-134,136,139,143,145,151], web-based platforms with videoconferencing features (n=15, 11.9%) [48,49,51,56,73,88,95,103,112,114,126,129,140,159,169], direct calls or SMS text messaging (n=4, 3.17%) [9,134,170], and videophones (n=9, 7.14%) [55,72,125,147,160]. The remaining studies included a mix of technologies (n=36, 28.57%) [47,56,63,83,86,87,91,92,94,102,108,110,115,118-120,123,124,137,144,146,148,149,153-155,163,165,166,171,172]. The specialized technologies in the studies were an embedded camera pen, remote symptom monitoring wristbands, and e-PC software with integrated camera laptop [51,79].

Purpose of the Interventions

mHealth was used for recording pain or other symptoms (n=9, 7.14%) [9,12,61,74,81,138,159,165,166] or remote monitoring (n=15, 11.9%) [11,47,56,63,67,78,86,89,102,111,119,124,126,129,139,147,170], support with medication [11,12], accessing information or serving as an educational resource [10,94,105,123,124,138,159,160,166], and care coordination [45,87,133,139]. Some of the mHealth apps also had built-in alert features [58,74,76]. A few studies engaged diverse stakeholder groups to help develop the interventions [70,81,84,126,127,137,139,140,153,158].

End Users of mHealth Interventions

The majority of the end users for mHealth interventions were health care professionals [35,72,76,81,84,108,122-124,127,140,164], adult patients [48,56,69,82,87,88,92,96,97,108,113,119-122,134,137,143,156,171], and informal caregivers [9,11,51,59,60,66,82,83,105,112,121-123,132,143,147,153,154,157,163,173]. Only 3 (2.38%) studies directly engaged children or adolescent patients [60,166,169], while other studies broadly explored the use of mHealth to support palliative care for the pediatric population [12,55,61,66,92,94,115,136,173].

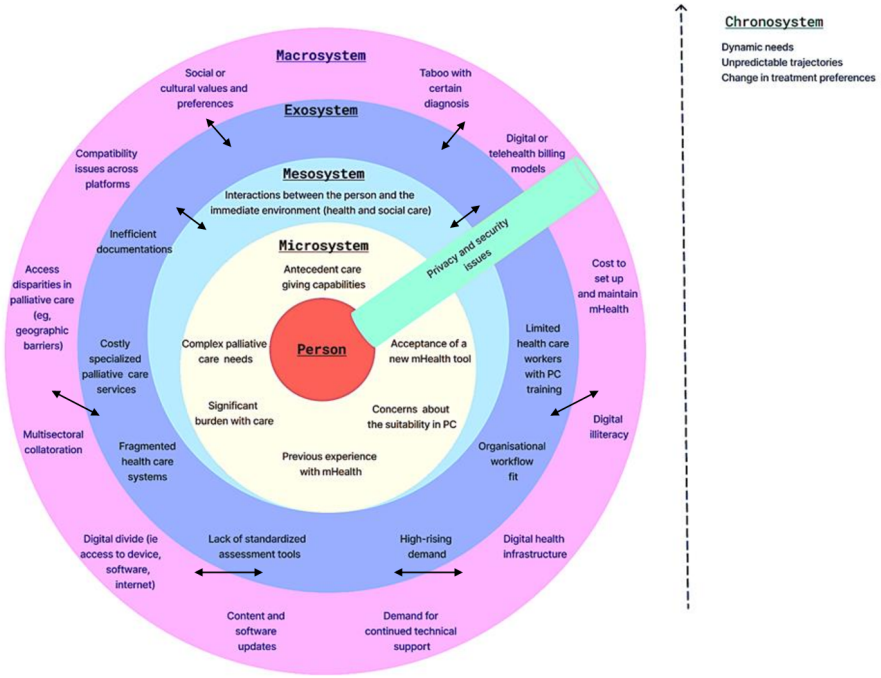
The Contextual Factors That Influence Mechanisms and Outcomes of mHealth

Overview

The contextual factors within which mHealth interventions were described using the Framework of Complexity in Palliative Care Context [32] (Figure 3). They encompassed a range of technological, organizational, economical, and social dimensions.

XSL•FO
RenderX

Figure 3. Contextual factors mapped to the domains of the framework of complexity in the palliative care (PC) context demonstrate the relationship between concepts mapped to the interconnected environmental layers of the ecological system model. Key: microsystem (person and immediate social environment and setting), chronosystem (influences of time), mesosystem (interactions between agents in the immediate environment), exosystem (factors within health services or systems), and macrosystem (factors beyond health care) [32]. mHealth: mobile health.



In the microsystem, important concepts were persons' or mHealth end users' needs (ie, complex health care and technology-related requirements) [45,63,142]. Individuals' characteristics were also key, which included limited knowledge about and the skills to support complex palliative care needs [92,97,123,156], the burden experienced by caregivers [9,12,121,128,142], and information gaps [88,134]. Concerning the acceptance of new mHealth tools, contextual concerns reported by end users encompassed their suitability in a palliative care context [95] and the perceived burden of monitoring symptoms for patients and family carers [48]. Furthermore, health care professionals had worries regarding reimbursement with the transition to virtual models of care [49,144]. In addition, having previous experience with similar DHIs and users' technical competency appeared to influence end users' confidence in using mHealth [57,76,90,92,99]. As depicted in the mesosystem, existing interpersonal relationships significantly influenced the utility of mHealth interventions. In fragmented health care systems, communication and information exchange challenges were particularly prominent, signifying the relevance of mHealth in such contexts [9,45,63,65,66,68,72,85,89,92,95,99,100,103,106,110,112,114,122,123,127,134,137,139,140,144,147,148,151,154,156,161,164,167,171,174]. Contextual factors that were identified in the chronosystem comprised the dynamic needs and fluctuating symptoms, which influenced patients' goals of care and treatment preferences [45,47,51,63,142] and thus impacted their willingness and ability to engage with mHealth technologies. Furthermore, end users, especially older adults, needed time to feel confident and become proficient with digital health tools [51,68].

In the exosystem, there were challenges in terms of a rising demand for palliative care [9,48,85,94,95,108,110,112,115,116,118-120,127,131,136], the lack of standardized assessment tools to appropriately identify patients who require palliative care [35,94,102,105,108,113,126,133,140,143], inefficient documentation and data sharing [87,127,147], and fragmented or siloed health care interventions [60,102,147]. Finally, contextual factors in the macrosystem encompassed disparities in health care access due to geographical barriers, lack of mHealth infrastructure, and inadequate technical support for end users [35,69,71,84,91,116,118,145,174]. Aligning technology with relevant social and cultural values [54,65,67] was reported as key to the successful adoption of mHealth [12,61,88,92,103,105,107,114,121,128,134]. The studies also highlighted the need for legal and organizational frameworks and regulations to ensure data privacy and security [51,94,110,115,119,120,124,126,132,148]. Multisectoral stakeholder engagement was important in the design and implementation stages as buy-in from each stakeholder influences intervention uptake and sustained use [49,99,103,134].

The CMOCs for the refined program theory are reported below and summarized in Table 3 and shown in Figure 4.

XSL•FO
RenderX

Table . List of context-mechanism-outcome configurations (CMOCs).

CMOCs	Description of CMOCs	Evidence summary (studies which contributed to CMOCs)
CMOC 1: Mitigating concerns and perceived threats	• 1.1 There are uncertainties about the suitability of mHealth[a] interventions in palliative care contexts in which patients need regular and holistic assessment and support (C), and where monitoring of worsening symptoms could be unsettling (C). Therefore, designing such tools tailored to end users' needs and capabilities, and providing appropriate orientation on the roles of end users, the frequency and timing of use, and what to expect from the interventions (IR) creates shared understanding (M) and comfort in using them (M) to potentially improve care delivery (O). • 1.2 The concerns about patient data security and privacy (C) necessitate that standard legal and ethical frameworks are established and adhered to (ie, to protect patient information) (IR), which creates feelings of security when recording and sharing data (M), thereby increasing the uptake of mHealth (O). • 1.3 The perceived threats to health care providers in terms of additional workload with less or no pay (C) demand the recognition of mHealth-supported home-based palliative care as standard payable services (eg, through the introduction of payment policies and billing models) (IR) and appropriate compensation for the health workforce (IR). This acknowledges the efforts and time health care workers spend, motivating them (M) to use mHealth and enabling better integration of mHealth with existing health care systems (O).	• [45,47,49,63,102,107,128,149,164] • [9,12,35,45,57,60,66,69,70,105,121,132,151] • [12,56,61,69,74,97,102,103,120,134,136,147,156,157]
CMOC 2: Minimizing infrastructural and technical barriers for user-friendly and sustained mHealth use	• The diverse infrastructural and technological experiences end users have (C) require training and education (IR) and maintaining low-cost, reliable infrastructure (IR) as well as inclusive and user-friendly design alongside offering technical support (IR). This improves confidence (M) and competency in using mHealth interventions (M), thus ensuring usability and wide-scale user penetration (O), which enables the standardization of health care delivery (O).	• [redacted]
CMOC 3: Engaging multiple stakeholders for better integration	• In contexts where patients, families, and care providers are required to navigate complex health care pathways (C), the development of mHealth interventions, which involves multiple stakeholders with diverse roles (IR), creates a shared understanding of the technical and system supports required (M), enhances collaborative relationships (M), and the integration of care pathways (M), which optimize resource utilization (O), potentially saving costs and time for both service users and health care systems (O).	• [redacted]

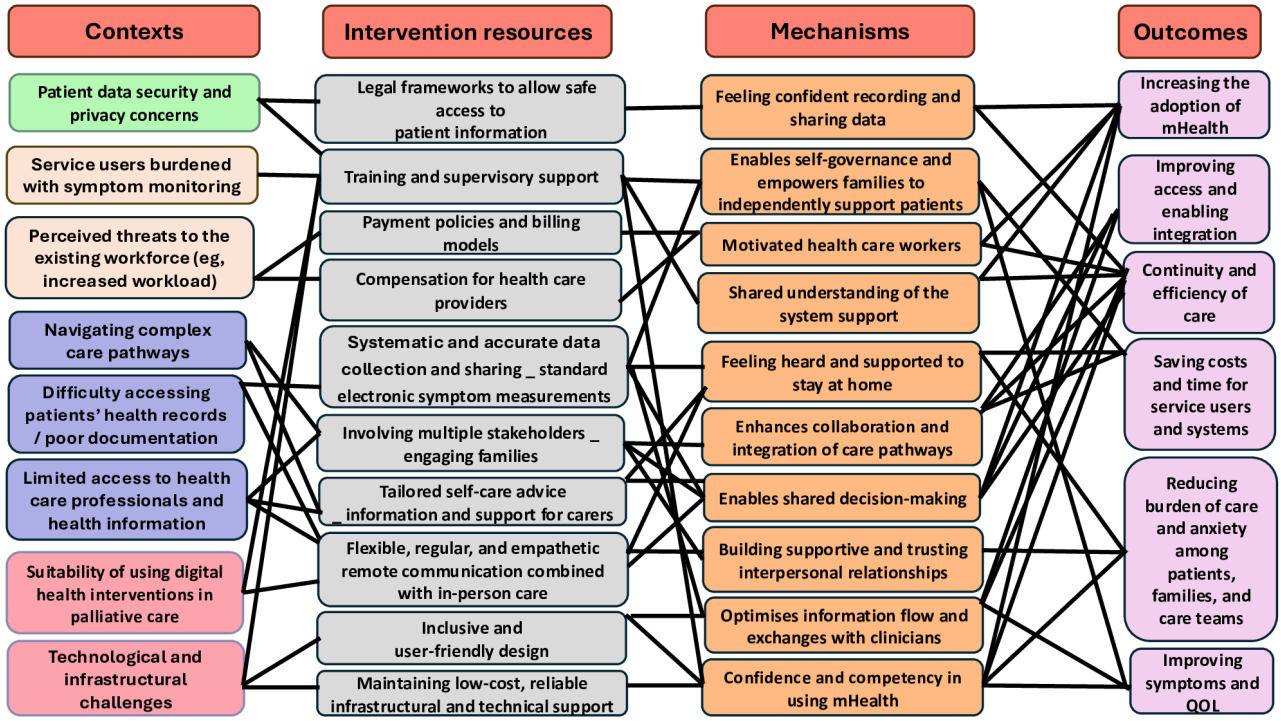
XSL•FO
RenderX

CMOCs	Description of CMOCs	Evidence summary (studies which contributed to CMOCs)
CMOC 4: Streamlined health management information system	• In home-based palliative care contexts where accessing patients' health records and documenting or sharing data is challenging (C), electronic data recording (IR) and standardized digital symptom measurement tools (IR) enable systematic and more accurate data capturing (M) as well as remote monitoring of symptoms (IR). This gives patients and caregivers a sense of comfort (M) and confidence knowing a health care team could be remotely accessible when needed (M), avoiding the inconvenience of unnecessary visits to health care facilities (O), and enabling preemptive management of problems, thus increasing the timeliness of the required support (O).	• [illegible]
CMOC 5: Regular communication and access to care	• 5.1 In the palliative care context where frequent communication with health care providers is necessary to support complex needs, often with unpredictable trajectories and limited access to health care professionals (C), mHealth provides a flexible and empathetic communication platform (IR), strengthening relationships between service users and service providers (M). This reduces the burden of care and anxiety among patients and families/informal caregivers (O), thus enhancing quality of life (O). • 5.2 In contexts where family caregivers are routinely closely involved in care provision (C), combining both in-person and virtual care (IR) and allowing families to join the conversation on equal terms, even remotely, (IR), enables shared decision-making (M) for better management of symptoms (O).	• [illegible] • [illegible]
CMOC 6: Access adequate information to support symptom	• In the palliative care context where adequate information about complex needs and options to support patients and caregivers is required (C), mHealth provides accessible self-care advice tailored to needs and the opportunity for knowledge exchange among health care workers (IR). This empowers end users (ie, patients, families, and junior health care professionals) to independently and effectively support symptoms (M), improving the efficiency of health care delivery (O) and quality of life (O).	• [illegible]

[a]mHealth: mobile health.

XSL•FO
RenderX

Figure 4. Conceptual model of context-mechanism-outcome configurations (CMOCs) illustrating how contextual factors within a multilevel system and mobile health (mHealth) intervention resources (eg, access to information or data) activate mechanisms (eg, improved capability, confidence, or strengthened interactions), leading to outcomes such as mHealth uptake, individual clinical outcomes, and health system–level impacts. The black lines indicate commonly reported pathways between context, intervention resources, mechanisms, and outcomes. Key: The first column (ie, context) is color coded according to the levels in the framework of complexity in palliative care context. Green: cross-cutting concept across levels in the system, beige: microsystem, blue: exosystem, pink: macrosystem. QOL: quality of life.



CMOC 1: Mitigating Concerns and Perceived Threats of mHealth

mHealth implementation may be resisted by various end users due to the concerns they may have about the intervention. One of the concerns is the suitability of using DHIs in palliative care for vulnerable patients and inadequately prepared informal caregivers who are already burdened with illness and caretaking responsibilities. This is mitigated by providing appropriate orientation on what to expect from DHIs, the role of end users, and when to use such tools. This in turn creates a shared understanding and tailored use directed to goals of care. Regarding apprehensions about patient data security and privacy, mHealth interventions that are designed adhering to international and local standard legal and ethical protocols and guidelines ensure feelings of security when recording and sharing data and safeguard patient information, which will encourage uptake and sustained use. Health care workers perceived they would experience job security threats and increased workload, for example, with parallel documentation. Furthermore, conventional health care systems were anticipated to take time to adopt mHealth enhanced models, with the need to develop payment protocols or billing systems and lease with the health workforce in the implementation. The shift toward mHealth creates concerns that existing interpersonal interactions would be disrupted. Thus, both service users and providers demanded that health care visits should be a combination of in-person and mHealth. Responding to such concerns will increase the acceptance of digital health technologies for improved care delivery.

CMOC 2: Minimizing Technical and Infrastructural Barriers for Sustained mHealth Use

Despite increasing access to mobile phones and increased routine use of mHealth, universal access to mobile devices has not been achieved. Reliability issues and inadequate technical support persist as key challenges in home-based palliative care. Sustained use depends on reliable infrastructure and technology tailored to local capacities. For mHealth to effectively support health care delivery, it should accommodate users' varying capabilities. National government and institutional support, through policies that promote mHealth adoption, is essential for successful implementation. Moreover, at the design phase, it is important to ensure that the technology is user-friendly and meets the needs of patients, families, and care providers. Health care workers who are confident in their technical skills are more likely to adopt these technologies. Yet, to be inclusive, all end users should receive orientation on the technical and functional aspects of such technologies for standardized use. This will improve self-efficacy and competency, thus ensuring usability with potential for wide-scale user penetration.

CMOC 3: Engaging Multiple Stakeholders for Better Integration

In contexts where both service users and care providers are required to navigate complex health care pathways, the development of clinically meaningful mHealth interventions must be designed with relevant multisectoral stakeholders from diverse disciplines. This approach will facilitate the consideration of unique perspectives and address the utility and usability needs and preferences of patients, family caregivers,

XSL•FO
RenderX

and health care workers. It is also important to assess the experiences and expectations of various end users, their competencies, and systems constraints to achieve optimal usage for intended outcomes. Furthermore, mHealth interventions must fulfill the standard requirements of health management information systems to ensure effective implementation and integration with routine care. The integration of mHealth in home-based palliative care presents numerous opportunities to enhance care delivery. For example, when mHealth interventions are seamlessly fitting into the standard workflow, health care workers' capacity to address patients' needs will increase, which improves the care experience for patients, optimizes resource use, and potentially saves costs and time for both service users and providers.

CMOC 4: Streamlined Health Management Information System

mHealth interventions represent efficient care delivery approaches that are well accepted by patients, families, and health care providers who value collaborative care planning and enhanced care coordination to improve patient outcomes. mHealth interventions designed to allow structured collection and reporting of symptoms, such as pain, using digitalized symptom reporting tools by patients or their carers enhance appropriate and timely management. Electronic health recording allows health care providers to closely monitor patients' conditions even remotely, triaging patients based on severity of symptoms for tailored response, which results in improved quality of care and health care workers' productivity. Primary care clinicians and generalist palliative care providers could document visits and share data for remote consultations with palliative care specialists, promoting knowledge exchange among health care teams. In addition, mHealth provides the opportunity for palliative care specialists to directly connect with service users (patients and family carers) through real-time audiovisual support. mHealth facilitates home-based care, improving comfort and convenience for patients and families while reducing risks linked to facility-based treatment.

CMOC 5: Improved Access to Care

In the management of chronic life-limiting illnesses, direct access to health care providers and regular symptom support are important. mHealth interventions designed to augment home-based palliative care play a pivotal role in facilitating access by surmounting geographical constraints, particularly for underserved populations and communities residing in remote areas. In palliative care context, such tools assist patients with an advanced stage of illness who have difficulties visiting health care facilities or those facing physical barriers to conventional health care services under unique circumstances, as observed during the COVID-19 pandemic. Both health care providers and recipients benefit from reduced burdens associated with logistical challenges and financial expenses associated with traveling to distant health care facilities. Moreover, mHealth fostered close communication between patients and providers and strengthened relationships. Family carers felt more supported, and their capacity to care for their loved ones improved, which reduced the anxiety associated with providing care for patients with serious illnesses. Considering the increasing demand for palliative care and the limited number of palliative care specialists, this is highly relevant to ensuring continuity of care and access to more equitable services.

CMOC 6: Access to Adequate Information to Manage Symptoms

The implementation of mHealth interventions serves as a means of disseminating information, thereby augmenting patient engagement in their health care journey. Patients are empowered as their comprehension of illness and its management deepens through personalized self-care advice to manage common symptoms. Empowered patients are more likely to engage in proactive healthy behaviors and adhere to medical advice, which improves symptoms and quality of life. Informal caregivers also benefit from applying mHealth tools in their caregiving roles to enhance their understanding and capabilities to support their loved ones. Health care professionals also use such interventions to access educational resources to assist them in care delivery.

Discussion

Principal Findings

This study presents a novel program theory to explore how, why, and under what circumstances mHealth interventions are designed and implemented to support better access to quality home-based palliative care. A program theory (Figure 4) is proposed on the causal explanations and key implementation strategies for successfully integrating mHealth interventions to support home-based palliative care. The important underlying mechanisms highlighted for mHealth intervention to support home-based service are the comfort patients and caregivers had staying at home having access to health care teams when needed; strengthened relationships between service users and providers through regular communication; empowerment of patients, families, and health care providers to independently and effectively manage symptoms through improved knowledge and skills; shared decision-making involving families; having a shared understanding of the technical and system support required; feeling confident and safe with using such technologies; motivation of health care workers through recognition and compensation of their time and effort; and enhanced collaboration and integration with existing health care systems.

Several systematic reviews on mHealth interventions have explored perceptions of end users on the advantages and challenges of their use for home-based palliative care. Our findings are supported by these reviews [3,108], which identified similar individual, interpersonal, organizational, and health systems factors influencing engagement with and effectiveness of mHealth. In palliative care, timely access to services impacts patient comfort and quality of life. mHealth has been identified as a health system–strengthening intervention [3,175]. The impact of mHealth has also been recognized in enhancing operations in pharmaceutical supply chain management systems [176] and in improving care coordination across levels of health care [177]. In studies from sub-Saharan Africa, such benefits were observed in populations living with HIV or AIDS and cancer [178,179]. Collectively, these reviews affirm that

XSL•FO
RenderX

mHealth not only enhances individual patient care but also strengthens the broader health care system, making it more accessible, efficient, and resilient. The role of informal caregivers in adapting DHIs in resource-limited settings has also been described previously [180]. Although continuous monitoring and immediate access to information and support can reduce anxiety in patients and their informal carers, this needs to be balanced with the burden of frequent assessment and reporting of symptoms by service users [181,182]. Furthermore, the combination of face-to-face alongside virtual contacts is encouraged [183]; building rapport in an initial in-person visit to develop better therapeutic relationships with subsequent contacts through technology. This prevents the care process from becoming depersonalized and mechanical, which contradicts the principles of person-centeredness in palliative care.

Strengths and Limitations

The evidence generated in this review drew from a broad global literature to determine how mHealth interventions improve access to and quality of home-based palliative care. It also incorporated perspectives from global and local stakeholders to refine the review focus and enhance the synthesis. The application of the framework of complexity in the palliative care context strengthened the comprehensive exploration of contextual influences of mHealth for home-based palliative care demonstrating interactions of multiple factors in the system. To allow us to generate findings and to provide comprehensive recommendations for home-based palliative care, we did not restrict mHealth use by type of user, the use case, conditions or diagnosis, or type of study. The emphasis of the synthesis was on underpinning mechanisms of action that influence outcomes, which have not been explored in detail in existing literature, nor explicitly stated in the included studies. Therefore, the hypotheses suggested in this study require further testing and refinement.

Contextual variations require intervention strategies that are tailored to the needs or preferences of end users. Most of the included studies in this review were notably from high-income countries; in 7 primary and multisite studies, only 24 countries were represented from LMICs. Therefore, application of these findings from this study needs careful consideration. Despite robust search strategies, only papers published in the English language were considered.

Recommendations

We have a number of recommendations from this review. To promote the effective integration of mHealth into home-based palliative care, it is essential to advocate for policies that support its adoption and seamless integration within health care systems. This can be achieved through the collaborative co-design of mHealth solutions with multisectoral stakeholders, ensuring that such tools address contextual barriers and align with existing health care infrastructure. The development of user-friendly mHealth interventions, tailored to the specific needs of end users, should be accompanied by comprehensive orientation and technical support to enhance usability and maximize benefits. Additionally, the involvement of informal caregivers, particularly in cultures that emphasize family engagement, is crucial for shared decision-making and the creation of personalized care plans that improve patient outcomes. Investing in reliable infrastructure and support systems is vital to minimize technical failures and ensure continuous care delivery. mHealth should also be leveraged to provide self-care guidance and equip informal caregivers with the necessary skills to support patients at home, thus alleviating pressure from the health care systems. In addition, health care providers must be adequately compensated for the additional time and effort required to operationalize mHealth interventions, which would encourage their active participation in enhancing patient care. Furthermore, the impact of mHealth interventions should be rigorously evaluated to provide evidence that informs policy decisions and promotes best practices that meet the needs of end users. Ongoing research is also necessary to explore the integration processes of mHealth interventions within health care systems, applying standardized frameworks to better understand factors influencing successful implementation.

Conclusions

To enhance usability and impact of mHealth interventions, it is important to adopt strategies that focus on the important mechanisms identified in this study. These mechanisms include patient and caregiver comfort with home-based care, strengthened provider-user relationships through regular communication, empowerment through better knowledge and skills, family-inclusive decision-making, shared understanding of the technical and system supports, confidence in using technology, motivation of health care workers through recognition and reimbursement, and improved integration with existing health care systems. Further study is needed to validate and evaluate these mechanisms in tailored interventions in LMICs.

Acknowledgments

A generative AI tool (ChatGPT, powered by OpenAI's GPT-4.1 model) was used solely for language editing, including grammar refinement and improvements in clarity and readability. The tool did not generate scientific content, influence the study design, data analysis, results, or interpretation. The authors take full responsibility for the accuracy and integrity of the manuscript. We thank Dr Munit Abdulreshid for assistance with table and image editing.

Funding

The authors declared no financial support was received for this work. NTG is a Doctoral researcher at King's College London, sponsored by the Africa International PGR Scholarship scheme. CH receives support from the National Institute for Health and

XSL•FO
RenderX

Care Research (NIHR) through the NIHR Global Health Research Group on Homelessness and Mental Health in Africa (NIHR134325), using UK aid from the UK Government. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. CH also receives support from Wellcome through grants 222154/Z20/Z and 223615/Z/21/Z. RH receives support from the NIHR through the NIHR Global Health Research Group on Global Health and Palliative Care (NIHR134440), using UK aid from the UK Government.

Data Availability

All data generated or analyzed during this study are included in this published article and Multimedia Appendix 1.

Authors' Contributions

Conceptualization: NTG (lead), CH (supporting), RH (supporting)
Formal analysis: NTG (lead), NG (supporting), OA (supporting)
Investigation: NTG (lead), NG (supporting), OA (supporting)
Methodology: NTG (lead), CH (supporting), RH (supporting), KN (supporting), NA (supporting)
Supervision: RH (lead), KN (supporting), CH (supporting), NA (supporting)
Validation: NG (supporting), OA (supporting)
Visualization: NTG (lead), NG (supporting), RH (supporting)
Writing – original draft: NTG (lead), CH (supporting), RH (supporting), KN (supporting), NA (supporting)
Writing – review & editing: NTG (lead), CH (supporting), RH (supporting), KN (supporting), NA (supporting), NG (supporting), OA (supporting).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials for the realist review, including the full database search strategy, quality appraisal summaries, and additional details on included studies.
[DOCX File, 278 KB - mhealth_v14i1e80321_app1.docx]

Checklist 1
PRISMA checklist.
[DOCX File, 278 KB - mhealth_v14i1e80321_app2.docx]

References

1. Mechael PN. The case for mHealth in developing countries. Innov Technol Gov Glob 2009 Jan 1;4(1):103-118. [doi: 10.1162/itgg.2009.4.1.103]
2. Murray E, Hekler EB, Andersson G, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851. [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
3. Finucane AM, O'Donnell H, Lugton J, Gibson-Watt T, Swenson C, Pagliari C. Digital health interventions in palliative care: a systematic meta-review. NPJ Digit Med 2021 Apr 6;4(1):64. [doi: 10.1038/s41746-021-00430-7] [Medline: 33824407]
4. Nwosu AC, McGlinchey T, Sanders J, et al. Identification of digital health priorities for palliative care research: modified Delphi study. JMIR Aging 2022 Mar 21;5(1):e32075. [doi: 10.2196/32075] [Medline: 35311674]
5. Consolo L, Colombo S, Basile I, et al. Barriers and facilitators of electronic patient-reported outcome measures (e-PROMs) for patients in home palliative cancer care: a qualitative study of healthcare professionals' perceptions. BMC Palliat Care 2023 Aug 4;22(1):111. [doi: 10.1186/s12904-023-01234-0] [Medline: 37542264]
6. Consolo L, Castellini G, Cilluffo S, Basile I, Lusignani M. Electronic patient-reported outcomes (e-PROMs) in palliative cancer care: a scoping review. J Patient Rep Outcomes 2022 Sep 23;6(1):102. [doi: 10.1186/s41687-022-00509-z] [Medline: 36138279]
7. Wu YR, Chou TJ, Wang YJ, et al. Smartphone-enabled, telehealth-based family conferences in palliative care during the COVID-19 pandemic: pilot observational study. JMIR Mhealth Uhealth 2020 Oct 28;8(10):e22069. [doi: 10.2196/22069] [Medline: 33021483]
8. Grewal US, Terauchi S, Beg MS. Telehealth and palliative care for patients with cancer: implications of the COVID-19 pandemic. JMIR Cancer 2020 Nov 24;6(2):e20288. [doi: 10.2196/20288] [Medline: 33049695]
9. Harding R, Carrasco JM, Serrano-Pons J, et al. Design and evaluation of a novel mobile phone application to improve palliative home-care in resource-limited settings. J Pain Symptom Manage 2021 Jul;62(1):1-9. [doi: 10.1016/j.jpainsymman.2020.09.045] [Medline: 33246073]

XSL•FO
RenderX

10. Allsop MJ, Powell RA, Namisango E. The state of mHealth development and use by palliative care services in sub-Saharan Africa: a systematic review of the literature. BMJ Support Palliat Care 2018 Jun;8(2):155-163. [doi: 10.1136/bmjspcare-2015-001034] [Medline: 27207721]
11. Dickman Portz J, Ford K, Bekelman DB, et al. "We're Taking Something So Human and Trying to Digitize": provider recommendations for mHealth in palliative care. J Palliat Med 2020 Feb;23(2):240-247. [doi: 10.1089/jpm.2019.0216] [Medline: 31526220]
12. Phongtankuel V, Shalev A, Adelman RD, et al. Mobile health technology is here—but are hospice informal caregivers receptive? Am J Hosp Palliat Care 2018 Dec;35(12):1547-1552. [doi: 10.1177/1049909118779018] [Medline: 29862851]
13. Global Diffusion of EHealth: Making Universal Health Coverage Achievable: Report of the Third Global Survey on EHealth: World Health Organization; 2017.
14. Knaul FM, Farmer PE, Krakauer EL, et al. Alleviating the access abyss in palliative care and pain relief-an imperative of universal health coverage: the Lancet Commission report. Lancet 2018 Apr 7;391(10128):1391-1454. [doi: 10.1016/S0140-6736(17)32513-8] [Medline: 29032993]
15. Connor SR, Morris C, Jaramillo E. Global Atlas of Palliative Care: World Health Organization (WHO); 2020.
16. Knaul FM, Bhadelia A, Rodriguez NM, Arreola-Ornelas H, Zimmermann C. The Lancet Commission on palliative care and pain relief—findings, recommendations, and future directions. Lancet Glob Health 2018 Mar;6:S5-S6. [doi: 10.1016/S2214-109X(18)30082-2]
17. Sleeman KE, de Brito M, Etkind S, et al. The escalating global burden of serious health-related suffering: projections to 2060 by world regions, age groups, and health conditions. Lancet Glob Health 2019 Jul;7(7):e883-e892. [doi: 10.1016/S2214-109X(19)30172-X] [Medline: 31129125]
18. Emanuel EJ, Fairclough DL, Slutsman J, Emanuel LL. Understanding economic and other burdens of terminal illness: the experience of patients and their caregivers. Ann Intern Med 2000 Mar 21;132(6):451-459. [doi: 10.7326/0003-4819-132-6-200003210-00005] [Medline: 10733444]
19. Reid EA, Kovalerchik O, Jubanyik K, Brown S, Hersey D, Grant L. Is palliative care cost-effective in low-income and middle-income countries? A mixed-methods systematic review. BMJ Support Palliat Care 2019 Jun;9(2):120-129. [doi: 10.1136/bmjspcare-2018-001499] [Medline: 30274970]
20. Radbruch L, De Lima L, Knaul F, et al. Redefining palliative care-a new consensus-based definition. J Pain Symptom Manage 2020 Oct;60(4):754-764. [doi: 10.1016/j.jpainsymman.2020.04.027] [Medline: 32387576]
21. Sepúlveda C, Marlin A, Yoshida T, Ullrich A. Palliative care: the World Health Organization's global perspective. J Pain Symptom Manage 2002 Aug;24(2):91-96. [doi: 10.1016/s0885-3924(02)00440-2] [Medline: 12231124]
22. Peeler A, Afolabi O, Adcock M, et al. Primary palliative care in low- and middle-income countries: a systematic review and thematic synthesis of the evidence for models and outcomes. Palliat Med 2024 Sep;38(8):776-789. [doi: 10.1177/02692163241248324] [Medline: 38693716]
23. Gomes B, Calanzani N, Curiale V, McCrone P, Higginson IJ, de Brito M. Effectiveness and cost-effectiveness of home palliative care services for adults with advanced illness and their caregivers. Cochrane Database Syst Rev 2013 Jun 6;2013(6):CD007760. [doi: 10.1002/14651858.CD007760.pub2] [Medline: 23744578]
24. New Horizons for Health through Mobile Technologies: World Health Organization; 2011. URL: http://whqlibdoc.who.int/publications/2011/9789241564250 [accessed 2026-06-24]
25. Bronfenbrenner U. Ecology of the family as a context for human development: research perspectives. In: Adolescents and Their Families: Routledge; 1999.
26. Bronfenbrenner U, Ceci SJ. Nature-nurture reconceptualized in developmental perspective: a bioecological model. Psychol Rev 1994 Oct;101(4):568-586. [doi: 10.1037/0033-295x.101.4.568] [Medline: 7984707]
27. Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R. RAMESES publication standards: realist syntheses. BMC Med 2013 Jan 29;11(1):21. [doi: 10.1186/1741-7015-11-21] [Medline: 23360677]
28. Greenhalgh T, Wong G, Jagosh J, et al. Protocol--the RAMESES II study: developing guidance and reporting standards for realist evaluation. BMJ Open 2015 Aug 3;5(8):e008567. [doi: 10.1136/bmjopen-2015-008567] [Medline: 26238395]
29. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist synthesis: an introduction. ESRC Research Methods Programme Working Paper Series.: ESRC Research Methods Programme University of Manchester; 2004 URL: https://www.researchgate.net/publication/228855827_Realist_Synthesis_An_Introduction [accessed 2026-06-24]
30. Covidence. URL: http://www.covidence.org [accessed 2026-06-25]
31. Abrams R, Park S, Wong G, et al. Lost in reviews: looking for the involvement of stakeholders, patients, public and other non-researcher contributors in realist reviews. Res Synth Methods 2021 Mar;12(2):239-247. [doi: 10.1002/jrsm.1459] [Medline: 32985074]
32. Pask S, Pinto C, Bristowe K, et al. A framework for complexity in palliative care: a qualitative study with patients, family carers and professionals. Palliat Med 2018 Jun;32(6):1078-1090. [doi: 10.1177/0269216318757622] [Medline: 29457743]
33. CC BY-NC 4.0: attribution-noncommercial 4.0 international. Creative Commons. URL: https://creativecommons.org/licenses/by-nc/4.0/ [accessed 2026-08-12]
34. Harding R. Informal caregivers in home palliative care. Prog Palliat Care 2013 Sep;21(4):229-231. [doi: 10.1179/1743291X13Y.0000000056]

XSL•FO
RenderX

35. Allsop MJ, Namisango E, Powell RA. A survey of mobile phone use in the provision of palliative care services in the African region and priorities for future development. J Telemed Telecare 2019 May;25(4):230-240. [doi: 10.1177/1357633X18767187] [Medline: 29635966]
36. Kahveci˙ K, Koç O, Aksakal H, Kahveci˙ K, Koç O, Aksakal H. Home-based palliative care. Bezmialem Science 2020 Jan 28;8(1):73-80. [doi: 10.14235/bas.galenos.2019.3208]
37. Roberts B, Robertson M, Ojukwu EI, Wu DS. Home based palliative care: known benefits and future directions. Curr Geri Rep 2021 Dec 1;10(4):141-147. [doi: 10.1007/s13670-021-00372-8]
38. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review--a new method of systematic review designed for complex policy interventions. J Health Serv Res Policy 2005 Jul;10 Suppl 1(1_suppl):21-34. [doi: 10.1258/1355819054308530] [Medline: 16053581]
39. Hong QN, Fàbregues S, Bartlett G, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. EFI 2018 Dec 18;34(4):285-291. [doi: 10.3233/EFI-180221]
40. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Based Healthc 2015 Sep;13(3):132-140. [doi: 10.1097/XEB.0000000000000055] [Medline: 26360830]
41. Rycroft-Malone J, McCormack B, Hutchinson AM, et al. Realist synthesis: illustrating the method for implementation research. Implement Sci 2012 Apr 19;7(1):33. [doi: 10.1186/1748-5908-7-33] [Medline: 22515663]
42. Siersbaek R, Ford JA, Burke S, Ní Cheallaigh C, Thomas S. Contexts and mechanisms that promote access to healthcare for populations experiencing homelessness: a realist review. BMJ Open 2021 Apr;11(4):e043091. [doi: 10.1136/bmjopen-2020-043091]
43. Anderson LM, Petticrew M, Rehfuess E, et al. Using logic models to capture complexity in systematic reviews. Res Synth Methods 2011 Mar;2(1):33-42. [doi: 10.1002/jrsm.32] [Medline: 26061598]
44. Greenhalgh T, Macfarlane F, Steed L, Walton R. What works for whom in pharmacist-led smoking cessation support: realist review. BMC Med 2016 Dec 16;14(1):209. [doi: 10.1186/s12916-016-0749-5] [Medline: 27978837]
45. Abahussin AA, West RM, Wong DC, Ziegler LE, Allsop MJ. Supporting pain self-management in patients with cancer: app development based on a theoretical and evidence-driven approach. JMIR Cancer 2023 Oct 9;9(1):e49471. [doi: 10.2196/49471] [Medline: 37812491]
46. Al-Mondhiry J, D'Ambruoso S, Pietras C, et al. Co-created mobile apps for palliative care using community-partnered participatory research: development and usability study. JMIR Form Res 2022 Jun 23;6(6):e33849. [doi: 10.2196/33849] [Medline: 35737441]
47. Ansari N, Wilson CM, Heneghan MB, Supiano K, Mooney K. How technology can improve communication and health outcomes in patients with advanced cancer: an integrative review. Support Care Cancer 2022 Aug;30(8):6525-6543. [doi: 10.1007/s00520-022-07037-y] [Medline: 35411467]
48. Aoki N, Ohta S, Yamamoto H, Kikuchi N, Dunn K. Triangulation analysis of tele-palliative care implementation in a rural community area in Japan. Telemed J E Health 2006 Dec;12(6):655-662. [doi: 10.1089/tmj.2006.12.655] [Medline: 17250487]
49. Archer S, Cheung NHY, Williams I, Darzi A. The impact of digital health interventions on the psychological outcomes of patients and families receiving paediatric palliative care: a systematic review and narrative synthesis. Palliat Med 2021 Dec;35(10):2017-2023. [doi: 10.1177/02692163211026523] [Medline: 34159851]
50. Baldwin PK, McAuley WJ, Van Nostrand J. Organizational factors associated with the use of telehospice. J Telemed Telecare 2012 Mar;18(2):90-93. [doi: 10.1258/jtt.2011.110803] [Medline: 22198959]
51. Balasubramanian S, Biji MS, Ranjith MK, Abhina SS. Patient satisfaction in home care services through e-Palliative care -an experience of tertiary cancer centre from Kerala. Indian J Palliat Care 2022;28(3):250-255. [doi: 10.25259/IJPC_36_2021] [Medline: 36072246]
52. Bandini JI, Tsuei J, Kavalieratos D, et al. Telehealth allows for flexibility and more equity": provider perspectives on telehealth for outpatient palliative care for underserved patients. J Palliat Med 2025 May;28(5):655-660. [doi: 10.1089/jpm.2024.0525] [Medline: 40009349]
53. Basile I, Consolo L, Colombo S, Rusconi D, Rampichini F, Lusignani M. Technology to support older adults in home palliative care: a scoping review. Am J Hosp Palliat Care 2024 Jun;41(6):673-690. [doi: 10.1177/10499091231189502]
54. Bauer EH, Schultz A, Smith AC, Brandt F, Bollig G, Dieperink KB. Using complex interventions to develop programme theory for a family‐focused model of telepalliative care. Scand J Caring Sci 2026 Mar;40(1):e70183. [doi: 10.1111/scs.70183] [Medline: 41556232]
55. Bensink M, Armfield N, Russell TG, Irving H, Wootton R. Paediatric palliative home care with internet-based video-phones: lessons learnt. J Telemed Telecare 2004;10 Suppl 1(1_suppl):10-13. [doi: 10.1258/1357633042614465] [Medline: 15603596]
56. Bensink ME, Armfield NR, Pinkerton R, et al. Using videotelephony to support paediatric oncology-related palliative care in the home: from abandoned RCT to acceptability study. Palliat Med 2009 Apr;23(3):228-237. [doi: 10.1177/0269216308100251] [Medline: 19073783]
57. Bethel C, Towers V, Crist JD, Silva GE, Shea K. A guide for intentional home telehealth assessment: patient and caregiver perceptions. Comput Inform Nurs 2021 May 28;39(12):943-947. [doi: 10.1097/CIN.0000000000000779] [Medline: 34050058]

XSL•FO
RenderX

58. Bhargava R, Keating B, Isenberg SR, Subramaniam S, Wegier P, Chasen M. RELIEF: A digital health tool for the remote self-reporting of symptoms in patients with cancer to address palliative care needs and minimize emergency department visits. Curr Oncol 2021 Oct 21;28(6):4273-4280. [doi: 10.3390/curroncol28060363] [Medline: 34898539]
59. Bonsignore L, Bloom N, Steinhauser K, et al. Evaluating the feasibility and acceptability of a telehealth program in a rural palliative care population: TapCloud for palliative care. J Pain Symptom Manage 2018 Jul;56(1):7-14. [doi: 10.1016/j.jpainsymman.2018.03.013] [Medline: 29551433]
60. Bradford NK, Armfield NR, Young J, Smith AC. Paediatric palliative care by video consultation at home: a cost minimisation analysis. BMC Health Serv Res 2014 Jul 28;14(1):328. [doi: 10.1186/1472-6963-14-328] [Medline: 25069399]
61. Bradford NK, Armfield NR, Young J, Herbert A, Mott C, Smith AC. Principles of a paediatric palliative care consultation can be achieved with home telemedicine. J Telemed Telecare 2014 Oct;20(7):360-364. [doi: 10.1177/1357633X14552370] [Medline: 25399995]
62. Brännström M, Philipsson J, Andersson S. Healthcare professionals' experiences of video consultations in palliative care in rural areas: an intervention study in community care. BMC Health Serv Res 2024 Jun 17;24(1):740. [doi: 10.1186/s12913-024-11196-5] [Medline: 38886714]
63. Burner-Fritsch I, Bolzani A, Hriskova K, Kolmhuber S, Bausewein C, Hodiamont F. Challenges developing an electronic patient-reported outcome measurement for palliative home care: a qualitative interview and focus group study. Palliat Med 2023 Feb;37(2):265-274. [doi: 10.1177/02692163221141487] [Medline: 36600402]
64. Caetano P, Querido A, Laranjeira C. Preparedness for caregiving role and telehealth use to provide informal palliative home care in Portugal: a qualitative study. Healthcare (Basel) 2024 Sep 25;12(19):1915. [doi: 10.3390/healthcare12191915] [Medline: 39408095]
65. Calton B, Shibley WP, Cohen E, et al. Patient and caregiver experience with outpatient palliative care telemedicine visits. Palliat Med Rep 2020 Dec 28;1(1):339-346. [doi: 10.1089/pmr.2020.0075] [Medline: 34223495]
66. Cameron P. Caregiver comfort with telehospice. Home Healthc Now 2021;39(1):32-38. [doi: 10.1097/NHH.0000000000000928] [Medline: 33417360]
67. Carey N, Abathun E, Maguire R, Wodaje Y, Royce C, Ayers N. Co-design and prototype development of the "Ayzot App": a mobile phone based remote monitoring system for palliative care. Palliat Med 2023 May;37(5):771-781. [doi: 10.1177/02692163231162408] [Medline: 37002562]
68. Christiansen L, Lindberg C, Sanmartin Berglund J, Anderberg P, Skär L. Using mobile health and the impact on health-related quality of life: perceptions of older adults with cognitive impairment. Int J Environ Res Public Health 2020 Apr 13;17(8):2650. [doi: 10.3390/ijerph17082650] [Medline: 32294911]
69. Chua IS, Olmsted M, Plotke R, et al. Video and in-person palliative care delivery challenges before and during the COVID-19 pandemic. J Pain Symptom Manage 2022 Dec;64(6):577-587. [doi: 10.1016/j.jpainsymman.2022.08.005] [Medline: 35985551]
70. Collier A, Morgan DD, Swetenham K, To THM, Currow DC, Tieman JJ. Implementation of a pilot telehealth programme in community palliative care: a qualitative study of clinicians' perspectives. Palliat Med 2016 Apr;30(4):409-417. [doi: 10.1177/0269216315600113] [Medline: 26290500]
71. Collier A, Appachoo S, Frey R, et al. Extending the reach of palliative care-a double-edged sword: a qualitative study of clinicians' experiences of delivering telehealth in Australia and Aotearoa New Zealand. BMC Palliat Care 2025 Nov 13;24(1):288. [doi: 10.1186/s12904-025-01932-x] [Medline: 41225429]
72. Cook DJ, Doolittle GC, Whitten PS. Administrator and provider perceptions of the factors relating to programme effectiveness in implementing telemedicine to provide end-of-life care. J Telemed Telecare 2001;7 Suppl 2:17-19. [doi: 10.1258/1357633011937335] [Medline: 11747648]
73. Coyle N, Khojainova N, Francavilla JM, Gonzales GR. Audio-visual communication and its use in palliative care. J Pain Symptom Manage 2002 Feb;23(2):171-175. [doi: 10.1016/s0885-3924(01)00402-x] [Medline: 11844640]
74. Cruz S, Fernandes C, Magalhães B. A scoping review of mobile apps for use with palliative patients in the context of home care. Int J Med Inform 2023 Sep;177:105166. [doi: 10.1016/j.ijmedinf.2023.105166] [Medline: 37527596]
75. Davis MS, Harrison KL, Rice JF, et al. A model for effective and efficient hospice care: proactive telephone-based enhancement of life through excellent caring,"TeleCaring" in advanced illness. J Pain Symptom Manage 2015 Sep;50(3):414-418. [doi: 10.1016/j.jpainsymman.2015.03.012] [Medline: 25896105]
76. Disalvo D, Agar M, Caplan G, et al. Virtual models of care for people with palliative care needs living in their own home: a systematic meta-review and narrative synthesis. Palliat Med 2021 Sep;35(8):1385-1406. [doi: 10.1177/02692163211024451] [Medline: 34169759]
77. Dismore L, Frew K, Wakefield D, Bryan C, Swainston K. Remote and rural communities face inequalities in access to specialist palliative care, could telemedicine enhance care? A qualitative study of patient, carer and healthcare professionals' experiences of video consultation. Palliat Care Soc Pract 2025 Sep;19:26323524251380632. [doi: 10.1177/26323524251380632] [Medline: 41049280]
78. Dolan H, Eggett C, Holliday L, Delves S, Parkes D, Sutherland K. Virtual care in end of life and palliative care: a rapid evidence check. J Telemed Telecare 2021 Dec;27(10):631-637. [doi: 10.1177/1357633X211046118] [Medline: 34726997]

XSL•FO
RenderX

79. Domínguez F, Heras J, Benjumea J, et al. Evaluating a wearable-based pain monitoring system in palliative cancer care: usability and feasibility study. JMIR Form Res 2026 Feb 6;10:e78098-e78098. [doi: 10.2196/78098] [Medline: 41650128]
80. Doolittle GC, Nelson EL, Spaulding AO, et al. TeleHospice: a community-engaged model for utilizing mobile tablets to enhance rural hospice care. Am J Hosp Palliat Care 2019 Sep;36(9):795-800. [doi: 10.1177/1049909119829458] [Medline: 30803247]
81. Evering RMH, Postel MG, van Os-Medendorp H, Bults M, den Ouden MEM. Intention of healthcare providers to use video-communication in terminal care: a cross-sectional study. BMC Palliat Care 2022 Nov 30;21(1):213. [doi: 10.1186/s12904-022-01100-5] [Medline: 36451219]
82. Gonzalez AF. Evaluating usability of a home health tablet device among adult palliative care patients. : Georgetown University; 2023 Mar 17 URL: https://repository.digital.georgetown.edu/handle/10822/1082679 [accessed 2026-06-24]
83. Gordon B, Mason B, Smith SLH. Leveraging trelehealth for delivery of palliative care to remote communities: a rapid review. J Palliat Care 2022 Apr;37(2):213-225. [doi: 10.1177/08258597211001184] [Medline: 33730904]
84. Guo J, Dai Y, Gong Y, Xu X, Chen Y. Exploring the telehealth readiness and its related factors among palliative care specialist nurses: a cross-sectional study in China. BMC Palliat Care 2023 Jun 28;22(1):82. [doi: 10.1186/s12904-023-01209-1] [Medline: 37370039]
85. Hancock S, Preston N, Jones H, Gadoud A. Telehealth in palliative care is being described but not evaluated: a systematic review. BMC Palliat Care 2019 Dec 13;18(1). [doi: 10.1186/s12904-019-0495-5] [Medline: 31835998]
86. Haroen H, Juniarti N, Sari CWM, et al. Integrating telehealth into community-based palliative care: a systematic review. J Multidiscip Healthc 2025;18:7339-7354. [doi: 10.2147/JMDH.S555226] [Medline: 41230403]
87. Haydon HM, Snoswell CL, Thomas EE, et al. Enhancing a community palliative care service with telehealth leads to efficiency gains and improves job satisfaction. J Telemed Telecare 2021 Dec;27(10):625-630. [doi: 10.1177/1357633X211048952] [Medline: 34726990]
88. Hebert MA, Paquin MJ, Whitten L, Cai P. Analysis of the suitability of “video-visits” for palliative home care: implications for practice. J Telemed Telecare 2007;13(2):74-78. [doi: 10.1258/135763307780096203] [Medline: 17359570]
89. Hennemann-Krause L, Lopes AJ, Araújo JA, Petersen EM, Nunes RA. The assessment of telemedicine to support outpatient palliative care in advanced cancer. Palliat Support Care 2015 Aug;13(4):1025-1030. [doi: 10.1017/S147895151400100X] [Medline: 25159308]
90. Hoek PD, Schers HJ, Bronkhorst EM, Vissers KCP, Hasselaar JGJ. The effect of weekly specialist palliative care teleconsultations in patients with advanced cancer –a randomized clinical trial. BMC Med 2017 Jun 19;15(1). [doi: 10.1186/s12916-017-0866-9] [Medline: 28625164]
91. Holland DE, Vanderboom CE, Ingram CJ, et al. The feasibility of using technology to enhance the transition of palliative care for rural patients. Comput Inform Nurs 2014 Jun;32(6):257-266. [doi: 10.1097/CIN.0000000000000066] [Medline: 24814998]
92. Holmen H, Riiser K, Winger A. Home-based pediatric palliative care and electronic health: systematic mixed methods review. J Med Internet Res 2020 Feb 28;22(2):e16248. [doi: 10.2196/16248] [Medline: 32130127]
93. Hutchinson RN, Anderson EC, Ruben MA, et al. A formative mixed-methods study of emotional responsiveness in telepalliative care. J Palliat Med 2022 Aug 1;25(8):1258-1267. [doi: 10.1089/jpm.2021.0589] [Medline: 35417249]
94. Jess M, Timm H, Dieperink KB. Video consultations in palliative care: a systematic integrative review. Palliat Med 2019 Sep;33(8):942-958. [doi: 10.1177/0269216319854938] [Medline: 31187688]
95. Jiang B, Bills M, Poon P. Integrated telehealth-assisted home-based specialist palliative care in rural Australia: a feasibility study. J Telemed Telecare 2023 Jan;29(1):50-57. [doi: 10.1177/1357633X20966466] [Medline: 33079611]
96. Johnston BM. 11 INVITED using technology in palliative care – a reality. Eur J Cancer 2011 Sep;47:S5. [doi: 10.1016/S0959-8049(11)70226-X]
97. Johnston B, Kidd L, Wengstrom Y, Kearney N. An evaluation of the use of telehealth within palliative care settings across Scotland. Palliat Med 2012 Mar;26(2):152-161. [doi: 10.1177/0269216311398698] [Medline: 21378067]
98. Johnston B. Palliative home-based technology from a practitioner’s perspective: benefits and disadvantages. Smart Homecare Technol Telehealth 2014:121-128. [doi: 10.2147/SHTT.S42687]
99. Funderskov KF, Boe Danbjørg D, Jess M, Munk L, Olsen Zwisler AD, Dieperink KB. Telemedicine in specialised palliative care: healthcare professionals’ and their perspectives on video consultations-a qualitative study. J Clin Nurs 2019 Nov;28(21-22):3966-3976. [doi: 10.1111/jocn.15004] [Medline: 31328336]
100. Kawashima A, Noguchi T, Furukawa T, Unesoko A, Togashi S, Sato K. Effectiveness of and implementation requirements for telehealth in palliative care patients with advanced cancer: a systematic review and meta-analysis. Palliat Med 2026 May;40(5):588-607. [doi: 10.1177/02692163251403395] [Medline: 41467572]
101. Keenan J, Rahman R, Hudson J. Exploring the acceptance of telehealth within palliative care: a self-determination theory perspective. Health Technol 2021 May;11(3):575-584. [doi: 10.1007/s12553-021-00535-9]
102. Kidd L, Cayless S, Johnston B, Wengstrom Y. Telehealth in palliative care in the UK: a review of the evidence. J Telemed Telecare 2010;16(7):394-402. [doi: 10.1258/jtt.2010.091108] [Medline: 20813893]

XSL•FO
RenderX

103. Kluger BM, Katz M, Galifianakis NB, et al. Patient and family outcomes of community neurologist palliative education and telehealth support in Parkinson disease. JAMA Neurol 2024 Jan 1;81(1):39-49. [doi: 10.1001/jamaneurol.2023.4260] [Medline: 37955923]
104. Kwon SH, Lee MA, Kim YJ, et al. Evaluation of Hospice@Home for home-based palliative care: development and usability pilot study. JMIR Form Res 2026 Jan 21;10(1):e79334. [doi: 10.2196/79334] [Medline: 41564291]
105. Middleton-Green L, Gadoud A, Norris B, et al. 'A Friend in the Corner': supporting people at home in the last year of life via telephone and video consultation—an evaluation. BMJ Support Palliat Care 2019 Dec;9(4):e26-e26. [doi: 10.1136/bmjspcare-2015-001016]
106. Lee J, Mc Carthy O, Ryan S, Kiely F. Video consultations in community palliative care–patient satisfaction: mixed methods study. BMJ Support Palliat Care 2024 Jan;13(e3):e1212-e1216. [doi: 10.1136/spcare-2022-004000]
107. Lind L, Karlsson D, Fridlund B. Patients' use of digital pens for pain assessment in advanced palliative home healthcare. Int J Med Inform 2008 Feb;77(2):129-136. [doi: 10.1016/j.ijmedinf.2007.01.013] [Medline: 17363323]
108. Lundereng ED, Nes AAG, Holmen H, et al. Health care professionals' experiences and perspectives on using telehealth for home-based palliative care: scoping review. J Med Internet Res 2023 Mar 29;25:e43429. [doi: 10.2196/43429] [Medline: 36989024]
109. Ma C, Fang Y, Zhang H, et al. Nurse-delivered telehealth in home-based palliative care: integrative systematic review. J Med Internet Res 2025 May 5;27(1):e73024. [doi: 10.2196/73024] [Medline: 40324776]
110. Mathews JJ, Chow R, Wennberg E, Lau J, Hannon B, Zimmermann C. Telehealth palliative care interventions for patients with advanced cancer: a scoping review. Support Care Cancer 2023 Jul 8;31(8):451. [doi: 10.1007/s00520-023-07907-z] [Medline: 37421447]
111. Maguire R, McCann L, Singleton C, et al. Exploring the impact of a remote monitoring system for palliative and End-of-Life CARE (CARE-PAC): mixed methods feasibility study. JMIR Form Res 2025 Sep 23;9(1):e69394. [doi: 10.2196/69394] [Medline: 40986856]
112. Maudlin J, Keene J, Kobb R. A road map for the last journey: home telehealth for holistic end-of-life care. Am J Hosp Palliat Care 2006;23(5):399-403. [doi: 10.1177/1049909106290807] [Medline: 17060308]
113. McCall K, Keen J, Farrer K, et al. Perceptions of the use of a remote monitoring system in patients receiving palliative care at home. Int J Palliat Nurs 2008 Sep;14(9):426-431. [doi: 10.12968/ijpn.2008.14.9.31121] [Medline: 19060793]
114. Menon PR, Stapleton RD, McVeigh U, Rabinowitz T. Telemedicine as a tool to provide family conferences and palliative care consultations in critically ill patients at rural health care institutions. Am J Hosp Palliat Care 2015 Jun;32(4):448-453. [doi: 10.1177/1049909114537110] [Medline: 24871344]
115. Miller KA, Baird J, Lira J, et al. The use of telemedicine for home-based palliative care for children with serious illness: a scoping review. J Pain Symptom Manage 2021 Sep;62(3):619-636. [doi: 10.1016/j.jpainsymman.2020.12.004] [Medline: 33348029]
116. Morse RS, Lambden K, Quinn E, et al. A mobile app to improve symptom control and information exchange among specialists and local health workers treating Tanzanian cancer patients: human-centered design approach. JMIR Cancer 2021 Mar 23;7(1):e24062. [doi: 10.2196/24062] [Medline: 33755022]
117. Mukamal RC, Dias LM, Sarti TD, Nunes R, Rego G. The role of telemedicine in enhancing palliative care for older adults: opportunities and challenges. BMC Geriatr 2025 Jul 2;25(1):474. [doi: 10.1186/s12877-025-06041-1] [Medline: 40604531]
118. Namasivayam P, Bui DT, Low C, et al. Use of telehealth in the provision of after-hours palliative care services in rural and remote Australia: a scoping review protocol. PLoS One 2022;17(1):e0261962. [doi: 10.1371/journal.pone.0261962] [Medline: 35025895]
119. Narvaez RA, Canaria A, Nifras SR, Listones N, Topacio R, Baua ME. The role of telehealth on quality of life of palliative care patients during the COVID-19 pandemic: an integrative review. Int J Palliat Nurs 2022 Dec 2;28(12):583-589. [doi: 10.12968/ijpn.2022.28.12.583] [Medline: 36520100]
120. Nguyen HQ, McMullen C, Haupt EC, et al. Findings and lessons learnt from early termination of a pragmatic comparative effectiveness trial of video consultations in home-based palliative care. BMJ Support Palliat Care 2022 Aug;12(e3):e432-e440. [doi: 10.1136/bmjspcare-2020-002553]
121. Oakley-Girvan I, Davis SW, Kurian A, et al. Development of a mobile health app (TOGETHERCare) to reduce cancer care partner burden: product design study. JMIR Form Res 2021 Aug 13;5(8):e22608. [doi: 10.2196/22608] [Medline: 34398787]
122. Oelschlägel L, Dihle A, Christensen VL, et al. Implementing welfare technology in palliative homecare for patients with cancer: a qualitative study of health-care professionals' experiences. BMC Palliat Care 2021 Sep 17;20(1):146. [doi: 10.1186/s12904-021-00844-w] [Medline: 34535125]
123. Parker Oliver DR, Demiris G. An assessment of the readiness of hospice organizations to accept technological innovation. J Telemed Telecare 2004 Jun 1;10(3):170-174. [doi: 10.1258/135763304323070832]
124. Oliver DP, Demiris G, Wittenberg-Lyles E, Washington K, Day T, Novak H. A systematic review of the evidence base for telehospice. Telemed J E Health 2012;18(1):38-47. [doi: 10.1089/tmj.2011.0061] [Medline: 22085114]
125. Olver I, Brooksbank M, Champion N, Keeley J. The use of videophones to enhance palliative care outreach nursing in remote areas. Prog Palliat Care 2005 Oct;13(5):263-267. [doi: 10.1179/096992605X57679]

XSL•FO
RenderX

126. Osuji TA, Macias M, McMullen C, et al. Clinician perspectives on implementing video visits in home-based palliative care. Palliat Med Rep 2020;1(1):221-226. [doi: 10.1089/pmr.2020.0074] [Medline: 34223480]
127. Paré G, Sicotte C, Chekli M, Jaana M, De Blois C, Bouchard M. A pre-post evaluation of a telehomecare program in oncology and palliative care. Telemed J E Health 2009 Mar;15(2):154-159. [doi: 10.1089/tmj.2008.0091] [Medline: 19292624]
128. Parker D, Hudson P, Tieman J, Thomas K, Saward D, Ivynian S. Evaluation of an online toolkit for carers of people with a life-limiting illness at the end-of-life: health professionals' perspectives. Aust J Prim Health 2021 Dec;27(6):473-478. [doi: 10.1071/PY21019] [Medline: 34802508]
129. Read Paul L, Salmon C, Sinnarajah A, Spice R. Web-based videoconferencing for rural palliative care consultation with elderly patients at home. Support Care Cancer 2019 Sep;27(9):3321-3330. [doi: 10.1007/s00520-018-4580-8] [Medline: 30613908]
130. Peñarrubia-San-Florencio L, Junior CG, Ricart S, et al. Family perspectives and experiences on implementing telehealth in pediatric palliative care: a qualitative approach. Eur J Pediatr 2025 Apr 8;184(5):287. [doi: 10.1007/s00431-025-06124-6] [Medline: 40199746]
131. Phillips JL, Davidson PM, Newton PJ, Digiacomo M. Supporting patients and their caregivers after-hours at the end of life: the role of telephone support. J Pain Symptom Manage 2008 Jul;36(1):11-21. [doi: 10.1016/j.jpainsymman.2007.08.017] [Medline: 18411012]
132. Pinto S, Caldeira S, Martins JC. E-Health in palliative care: review of literature, Google Play and App Store. Int J Palliat Nurs 2017 Aug 2;23(8):394-401. [doi: 10.12968/ijpn.2017.23.8.394] [Medline: 28854054]
133. Pinto S, Almeida F, Caldeira S, Martins JC. The Comfort app prototype: introducing a web-based application for monitoring comfort in palliative care. Int J Palliat Nurs 2017 Sep 2;23(9):420-431. [doi: 10.12968/ijpn.2017.23.9.420] [Medline: 28933997]
134. Putranto D, Rochmawati E. Mobile applications for managing symptoms of patients with cancer at home: a scoping review. Int J Nurs Pract 2020 Aug;26(4):1-14. [doi: 10.1111/ijn.12842] [Medline: 32347599]
135. Rana S, DeGroote NP, Harris E, et al. Interview results from a telehealth hospice intervention for children and young adults with cancer. J Pain Symptom Manage 2025 Apr;69(4):343-353. [doi: 10.1016/j.jpainsymman.2024.12.018] [Medline: 39746494]
136. Royackers A, Regan S, Donelle L. The eShift model of care: informal caregivers' experience of a new model of home-based palliative care. Prog Palliat Care 2016 Mar 3;24(2):84-92. [doi: 10.1179/1743291X15Y.0000000006]
137. Sadang KG, Centracchio JA, Turk Y, et al. Clinician perceptions of barriers and facilitators for delivering early integrated palliative care via telehealth. CANCERS (Basel) 2023 Nov 9;15(22):5340. [doi: 10.3390/cancers15225340] [Medline: 38001600]
138. Salem R, El Zakhem A, Gharamti A, Tfayli A, Osman H. Palliative care via telemedicine: a qualitative study of caregiver and provider perceptions. J Palliat Med 2020 Dec;23(12):1594-1598. [doi: 10.1089/jpm.2020.0002] [Medline: 32326816]
139. Salimian N, Ehteshami A, Ashouri-Talouki M. Developing Ghasedak: a mobile application to improve the quality of cancer palliative care. Acta Inform Med 2019 Mar;27(1):19-22. [doi: 10.5455/aim.2019.27.19-22] [Medline: 31213738]
140. Saysell E, Routley C. Telemedicine in community-based palliative care: evaluation of a videolink teleconference project. Int J Palliat Nurs 2003 Nov;9(11):489-495. [doi: 10.12968/ijpn.2003.9.11.11874] [Medline: 14676726]
141. Septian RB, Suryana R, Rukhama RA, Destyanto AA, Librianty R. The use of telehealth in home-based palliative care: an umbrella review of systematic reviews. Home Health Care Manag Pract 2025. [doi: 10.1177/10848223251359667]
142. Schmitt S, Ebby C, Doshi A, Bower K, Marc-Aurele K. Retrospective pediatric telepalliative care experience. J Palliat Med 2022 Feb;25(2):274-281. [doi: 10.1089/jpm.2021.0173] [Medline: 34550796]
143. Schoppee TM, Dyal BW, Scarton L, et al. Patients and caregivers rate the PAINReportIt wireless internet-enabled tablet as a method for reporting pain during end-of-life cancer care. Cancer Nurs 2020;43(5):419-424. [doi: 10.1097/NCC.0000000000000743] [Medline: 31517649]
144. Schuessler N, Glarcher M. Caregivers' perspectives on ethical challenges and patient safety in tele-palliative care. J Hosp Palliat Nurs 2024 Feb 1;26(1):E1-E12. [doi: 10.1097/NJH.0000000000000986] [Medline: 37747961]
145. Slavin-Stewart C, Phillips A, Horton R. A feasibility study of home-based palliative care telemedicine in rural Nova Scotia. J Palliat Med 2020 Apr;23(4):548-551. [doi: 10.1089/jpm.2019.0173] [Medline: 31532325]
146. Steindal SA, Nes AAG, Godskesen TE, et al. Advantages and challenges of using telehealth for home-based palliative care: systematic mixed studies review. J Med Internet Res 2023 Mar 13;25(100959882):e43684. [doi: 10.2196/43684] [Medline: 36912876]
147. Stern A, Valaitis R, Weir R, Jadad AR. Use of home telehealth in palliative cancer care: a case study. J Telemed Telecare 2012 Jul;18(5):297-300. [doi: 10.1258/jtt.2012.111201] [Medline: 22790013]
148. Sutherland AE, Stickland J, Wee B. Can video consultations replace face-to-face interviews? Palliative medicine and the COVID-19 pandemic: rapid review. BMJ Support Palliat Care 2020 Sep;10(3):271-275. [doi: 10.1136/bmjspcare-2020-002326] [Medline: 32457086]
149. Tasneem S, Kim A, Bagheri A, Lebret J. Telemedicine video visits for patients receiving palliative care: a qualitative study. Am J Hosp Palliat Care 2019 Sep;36(9):789-794. [doi: 10.1177/1049909119846843] [Medline: 31064195]

XSL•FO
RenderX

150. Tarbi EC, Ambrose N, Anderson EC, et al. It's hard to talk to a computer, I get it": an exploratory analysis of clinician connection-building communication practices in rural telepalliative care encounters. PEC Innov 2025 Jun;6:100377. [doi: 10.1016/j.pecinn.2025.100377] [Medline: 39974585]
151. Tieman JJ, Swetenham K, Morgan DD, To TH, Currow DC. Using telehealth to support end of life care in the community: a feasibility study. BMC Palliat Care 2016 Nov 17;15(1):94. [doi: 10.1186/s12904-016-0167-7] [Medline: 27855681]
152. Toenne R, Neuhaus U, Höcker M, et al. Exploring the viability of telehealth integration into specialised paediatric palliative care. Int J Palliat Nurs 2025 Feb 2;31(2):58-67. [doi: 10.12968/ijpn.2025.31.2.58] [Medline: 39969901]
153. Tong W, Kapur S, Fleet A, et al. Perspectives of remote volunteer palliative care consultants during COVID: a qualitative study. J Pain Symptom Manage 2022 Feb;63(2):321-329. [doi: 10.1016/j.jpainsymman.2021.07.028] [Medline: 34363952]
154. van Gurp J, van Selm M, van Leeuwen E, Hasselaar J. Transmural palliative care by means of teleconsultation: a window of opportunities and new restrictions. BMC Med Ethics 2013 Mar 7;14(1):12. [doi: 10.1186/1472-6939-14-12] [Medline: 23497307]
155. van Gurp J, van Selm M, van Leeuwen E, Vissers K, Hasselaar J. Teleconsultation for integrated palliative care at home: a qualitative study. Palliat Med 2016 Mar;30(3):257-269. [doi: 10.1177/0269216315598068] [Medline: 26269323]
156. Salvador Vergès À, Cusí Sánchez MV, Bossio Grigera P, et al. Determinants in stakeholder opinions about telemedicine in palliative care: a scoping review. Telemed J E Health 2022 Jul;28(7):932-941. [doi: 10.1089/tmj.2021.0441] [Medline: 34871034]
157. Vincent D, Peixoto C, Quinn KL, et al. Virtual home-based palliative care during COVID-19: a qualitative exploration of the patient, caregiver, and healthcare provider experience. Palliat Med 2022 Oct;36(9):1374-1388. [doi: 10.1177/02692163221116251] [Medline: 36071621]
158. Washington KT, Demiris G, Oliver DP, Day M. Telehospice acceptance among providers: a multidisciplinary comparison. Am J Hosp Palliat Care 2009 Dec;25(6):452-457. [doi: 10.1177/1049909108322289]
159. Weaver MS, Rosenberg AR, Fry A, Shostrom V, Wiener L. Impact of the coronavirus pandemic on pediatric palliative care team structures, services, and care delivery. J Palliat Med 2021 Aug;24(8):1213-1220. [doi: 10.1089/jpm.2020.0589] [Medline: 33350874]
160. Whitten P, Doolittle G, Mackert M. Telehospice in Michigan: use and patient acceptance. Am J Hosp Palliat Care 2004;21(3):191-195. [doi: 10.1177/104990910402100307] [Medline: 15188918]
161. Williams E, White C, Banerjee A, et al. Quality of outpatient pediatric palliative care telehealth: a retrospective chart review. J Pain Symptom Manage 2024 Jun;67(6):561-570. [doi: 10.1016/j.jpainsymman.2024.03.014] [Medline: 38514022]
162. Wood MJM, Nwosu AC, Dinham D, Seneque N, Walshe C. How, when and why is emotional support delivered using videoconferencing by adult palliative care services successful? A realist synthesis. Palliat Care 2025 Nov;19. [doi: 10.1177/26323524251363271]
163. Xu X, Ho MH, Lin CC. Telehealth in palliative care during the COVID‐19 pandemic: a systematic mixed studies review. Worldviews Evid Based Nurs 2023 Oct;20(5):476-491. [doi: 10.1111/wvn.12637] [Medline: 36942832]
164. Zelko E, Ramsak Pajk J, Škvarč NK. An innovative approach for improving information exchange between palliative care providers in slovenian primary health—a qualitative analysis of testing a new tool. Healthcare (Basel) 2022 Jan 22;10(2). [doi: 10.3390/healthcare10020216] [Medline: 35206829]
165. Zheng Y, Head BA, Schapmire TJ. A systematic review of telehealth in palliative care: caregiver outcomes. Telemed J E Health 2016 Apr;22(4):288-294. [doi: 10.1089/tmj.2015.0090] [Medline: 26360181]
166. Zimmermann J, Heilmann ML, Fisch-Jessen M, et al. Telehealth needs and concerns of stakeholders in pediatric palliative home care. Children (Basel) 2023 Jul 30;10(8):1315. [doi: 10.3390/children10081315] [Medline: 37628315]
167. Johnston B. UK telehealth initiatives in palliative care: a review. Int J Palliat Nurs 2011 Jun;17(6):301-308. [doi: 10.12968/ijpn.2011.17.6.301] [Medline: 21727889]
168. Tarbi EC, Durieux BN, Brain JM, et al. Measuring palliative care communication via telehealth: a pilot study. J Pain Symptom Manage 2023 Jul;66(1):e155-e161. [doi: 10.1016/j.jpainsymman.2023.04.001] [Medline: 37037343]
169. Weaver MS, Neumann ML, Navaneethan H, Robinson JE, Hinds PS. Human touch via touchscreen: rural nurses' experiential perspectives on telehealth use in pediatric hospice care. J Pain Symptom Manage 2020 Nov;60(5):1027-1033. [doi: 10.1016/j.jpainsymman.2020.06.003] [Medline: 32525081]
170. Doolittle GC, Whitten P, McCartney M, Cook D, Nazir N. An empirical chart analysis of the suitability of telemedicine for hospice visits. Telemed J E Health 2005 Feb;11(1):90-97. [doi: 10.1089/tmj.2005.11.90] [Medline: 15785226]
171. Steindal SA, Nes AAG, Godskesen TE, et al. Patients' experiences of telehealth in palliative home care: scoping review. J Med Internet Res 2020 May 5;22(5):e16218. [doi: 10.2196/16218] [Medline: 32369037]
172. Johnsen AT, Petersen MA, Sjøgren P, et al. Exploratory analyses of the Danish Palliative Care Trial (DanPaCT): a randomized trial of early specialized palliative care plus standard care versus standard care in advanced cancer patients. Support Care Cancer 2020 May;28(5):2145-2155. [doi: 10.1007/s00520-019-05021-7] [Medline: 31410598]
173. Philip J, Wawryk O, Pasanen L, Wong A, Schwetlik S, Collins A. Telehealth in outpatient delivery of palliative care: a prospective survey evaluation by patients and clinicians. Intern Med J 2022 Jul;52(7):1144-1153. [doi: 10.1111/imj.15721] [Medline: 35189018]

XSL•FO
RenderX

174. Paul LR, Walia R, Porquet-Seitz N, et al. Rural home care palliative videoconsultation project – phase 2. J Pain Symptom Manage 2018 Dec;56(6):e103-e104. [doi: 10.1016/j.jpainsymman.2018.10.349]
175. Capurro D, Ganzinger M, Pérez-Lu JE. Palliative care from a medical informatics perspective in Chile, Germany, and Peru. Stud Health Technol Inform 2013;192:1013. [Medline: 23920787]
176. Lind L, Karlsson D. A system for symptom assessment in advanced palliative home healthcare using digital pens. Med Inform Internet Med 2004 Jan;29(3-4):199-210. [doi: 10.1080/14639230400005966]
177. Allsop MJ, Kite S, McDermott S, Penn N, Millares-Martin P, Bennett MI. Electronic palliative care coordination systems: devising and testing a methodology for evaluating documentation. Palliat Med 2017 May;31(5):475-482. [doi: 10.1177/0269216316663881] [Medline: 27507636]
178. Mwase C, Nkhoma K, Allsop MJ. The role of digital health in palliative care for people living with HIV in sub-Saharan Africa: a systematic review. Digit Health 2022 Jan;8:20552076221133707. [doi: 10.1177/20552076221133707] [Medline: 36457812]
179. Schliemann D, Tan MM, Hoe WMK, et al. mHealth interventions to improve cancer screening and early detection: scoping review of reviews. J Med Internet Res 2022 Aug 15;24(8):e36316. [doi: 10.2196/36316] [Medline: 35969450]
180. Yang GM, Ke Y, Ng XH, Neo PSH, Cheung YB. Proactive symptom monitoring to initiate timely palliative care for patients with advanced cancer: a randomized controlled trial. Support Care Cancer 2025 Mar 5;33(3):249. [doi: 10.1007/s00520-025-09311-1] [Medline: 40038129]
181. Bienfait F, Petit M, Pardenaud R, Guineberteau C, Pignon A. Applying M-Health to palliative care: a systematic review on the use of M-Health in monitoring patients with chronic diseases and its transposition in palliative care. Am J Hosp Palliat Care 2020 Jul;37(7):549-564. [doi: 10.1177/1049909119885655] [Medline: 31773969]
182. Dilhani WNS, Mitchell S, Dale J, Toor K, Javaid M, MacArtney JI. A mixed-methods systematic review investigating the use of digital health interventions to provide palliative and end-of-life care for patients in low- and middle-income countries. Palliat Care Soc Pract 2024 Jan;18:26323524241236965. [doi: 10.1177/26323524241236965] [Medline: 38617095]
183. Widberg C, Wiklund B, Klarare A. Patients' experiences of eHealth in palliative care: an integrative review. BMC Palliat Care 2020 Oct 14;19(1):158. [doi: 10.1186/s12904-020-00667-1] [Medline: 33054746]

Abbreviations

CMOC: context-mechanism-outcome configuration
DHI: digital health intervention
LMIC: low- and middle-income country
mHealth: mobile health
MMAT: Mixed Methods Appraisal Tool
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
WHO: World Health Organization

Edited by L Buis; submitted 08.Jul.2025; peer-reviewed by K Zhang, N Roberts; revised version received 16.Mar.2026; accepted 14.Apr.2026; published 14.Aug.2026.

Please cite as:
Gebre NT, Afolabi O, Gafer N, Nkhoma KB, Ayers N, Hanlon C, Harding R
Mobile Health in Home-Based Palliative Care: Realist Review of Contextual Factors and Mechanisms Influencing Outcomes
JMIR Mhealth Uhealth 2026;14:e80321
URL: https://mhealth.jmir.org/2026/1/e80321
doi:10.2196/80321

© Nuhamin Tekle Gebre, Oladayo Afolabi, Nahla Gafer, Kennedy Bashan Nkhoma, Nicola Ayers, Charlotte Hanlon, Richard Harding. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Engagement of Users in Digital Health Applications: Scoping Review

Melek Aktas[1], PhD; Linda Cambon[2], PhD; Olivier Aromatario[1], PhD

[1]Mérisp Team Labelled League Against Cancer (CIC 1401), BPH (U1219), INSERM, University of Bordeaux, 146 Rue Léo Saignat, Bordeaux, France
[2]Prevention Unit, University Hospital of Bordeaux, Bordeaux, France

Corresponding Author:
Melek Aktas, PhD
Mérisp Team Labelled League Against Cancer (CIC 1401), BPH (U1219), INSERM, University of Bordeaux, 146 Rue Léo Saignat, Bordeaux, France

Abstract

Background: Mobile health (mHealth) uses mobile technology as a tool for prevention and health promotion. Research indicates that user engagement is crucial for effective mHealth interventions and improved health outcomes. However, many studies report low adoption rates, rapid decline after initial use, and a lack of acknowledgment of user implications in achieving outcomes. Thus, conceptualizing participation in mHealth is essential to identify key determinants for engaging users.

Objective: This scoping review aims to identify and characterize the attributes and definitions of user engagement in mHealth, examine engagement methods, and analyze barriers and facilitators influencing participation.

Methods: Following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines, Scopus, Web of Science, and PubMed databases were searched for publications between 2000 and 2025 with a 2-stage selection process.

Results: Out of 2489 articles identified, 1416 were screened, and 52 met the inclusion criteria. Half were recently published in the last 5 years (2020 - 2025). Existing literature focused on digital interventions for specific populations and health topics. Analysis revealed four main perspectives on engagement in mHealth: (1) usage metrics, (2) subjective user experiences, (3) a hybrid approach that combines both, and (4) a goal-oriented perspective (behavior change or health outcomes).

Conclusions: To understand the complexity and multifactorial nature of participation, it is relevant to conceptualize it as a dynamic mechanism enabling users to achieve their objectives. Both quantitative use and subjective user experience should be integrated to reach the optimal intervention dose. Recognizing users' evolving needs, uniqueness, and their socioenvironmental context interdependence, it is essential to involve users in all stages (design, implementation, and iterative evaluation of mHealth). Findings will inform an e-Delphi study to establish consensus on engagement criteria.

(JMIR Mhealth Uhealth 2026;14:e66002) doi:10.2196/66002

KEYWORDS

mHealth; user engagement; participation; digital health interventions; prevention; health promotion; scoping review; health app; behavior change; mobile health

Introduction

The World Health Organization's Global Observatory for eHealth defines mobile health (mHealth) as "medical and public health practices supported by mobile devices such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices" [1]. mHealth emerged as a promising approach to enhance health care, health promotion, and prevention delivery by leveraging mobile technologies such as smartphones, tablets, and wearable devices [2-5].

Community participation is a crucial element in health prevention and promotion, as highlighted by several important documents and reports. The Ottawa Charter for Health Promotion established community participation as one of the fundamental principles for health promotion [6], and the importance of community engagement in reducing health inequalities has been demonstrated [7]. Going forward, the World Health Organization's Global Strategy for Health for All by 2030 has also emphasized the importance of community participation in the design, implementation, and evaluation of public health programs [8]. These documents and reports have contributed to establishing the importance of community participation in health prevention and promotion, and encouraged governments and health organizations to partner with communities to design and implement effective health programs. Studies have also demonstrated a correlation between user engagement and the effectiveness of web-based behavior change interventions [9,10]. Evidence suggests that greater user

XSL•FO
RenderX

engagement is associated with better health outcomes in various mHealth interventions [11,12].

However, actual usage and sustained engagement with mHealth remain limited. According to data collected between 2010 and 2019 from 12,000 respondents, approximately 25% of downloaded apps were opened only once before being abandoned [13]. More recent findings based on a study of 2200 Android apps revealed that, on average, one out of every two apps is uninstalled within 30 days of initial download, with most of these occurring in the initial 24-hour period postinstallation, often due to unmet expectations or a disconnect between the app's marketing and the actual user expectation [14,15]. In addition to studying accessibility, it is therefore important to explore the actual usage of these tools [16].

Currently, there is no widely accepted definition of participation [17,18], and there is a variability [19,20] and a lack of clarity around the question of how to conceptualize engagement [21]. Sometimes described as a "suitcase word," the term engagement can be associated with "almost any action involving the population, to a greater or lesser extent." Participation can take on various forms and intensities, leading to different classifications in the literature. Comprehensive definitions of participation, its conceptual boundaries, and practical modalities are not always explicit [22]. To our knowledge, there is no consolidated consensus of the characteristics, methods, barriers, and facilitators related to user participation in this context.

Consequently, there is a need to better understand how this concept is mobilized in order to guide future evaluators and designers of digital interventions toward participatory, effective, and sustainable approaches [14].

For these reasons, given the wide and exploratory nature of the research questions and the diversity of the existing literature, a scoping review was conducted in order to identify existing literature on user participation in mHealth. The following research questions were formulated: How is user participation characterized in digital health interventions? What techniques and/or methods are implemented to involve users? What are the barriers and facilitators encountered in implementing user participation?

This paper aims to (1) identify the attributes and definitions of user engagement, (2) examine the methods used for engaging users in mHealth, and (3) identify barriers and facilitators encountered in implementing user participation.

Methods

Study Design

A scoping review was chosen as a systematic approach to map the evidence and identify key concepts and relevant theories on the subject. To guide this review, the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist for scoping reviews and tip sheets documents were used [23] (see Checklist 1).

Eligibility Criteria

Articles eligible for inclusion focused on mHealth interventions targeting health promotion and prevention. All reviews and primary articles were included. The articles had to be published in English or French between January 1, 2000, and April 14, 2025.

Exclusion criteria were studies focused on clinical interventions, disease management, and tool evaluations. Studies on health worker training were also excluded. Based on existing participation research in various other subject areas, interventions that consider engagement as simple service usage (such as logging in), information reception (where users only receive information), and consultation (where users express themselves using free text and this expression has no influence on the process) were excluded [24,25].

Information Sources

Literature searches were conducted in 3 databases: Scopus, Web of Science, and PubMed. These databases were chosen because of their focus on social behavior and medical sciences. Keywords were collected from articles of interest until saturation was reached. Then, keywords were refined considering the vocabulary of each database, adding additional keywords, using truncation to search by title, and sometimes by title and abstract. The search strategy was different for each database, adapted to the respective databases, and included the following terms divided into 3 categories relating to subject, medium, and field (see Multimedia Appendix 1).

Selection of Sources of Evidence

Search results were uploaded to Zotero to sort out duplicates. Then, the articles were transferred to Covidence, a web application for independent selection of articles by multiple researchers. A 2-stage selection process was established. First, titles and abstracts were screened by two authors (MA and OA) for inclusion or exclusion of articles. Then, discrepancies were solved through discussion until a consensus was reached. Finally, a full-text screening was carried out (MA).

Data Items and Synthesis of Results

The selected articles were analyzed using an evolving reading grid in Microsoft Excel, which includes a list of themes of interest based on the aim of the scoping review and research questions. These themes included (1) article presentation (goals, target population, etc); (2) attributes of participation (definitions, scales, etc); (3) methods used to engage users (theory, model, framework, etc); and (4) barriers, facilitators, and strategies to promote engagement.

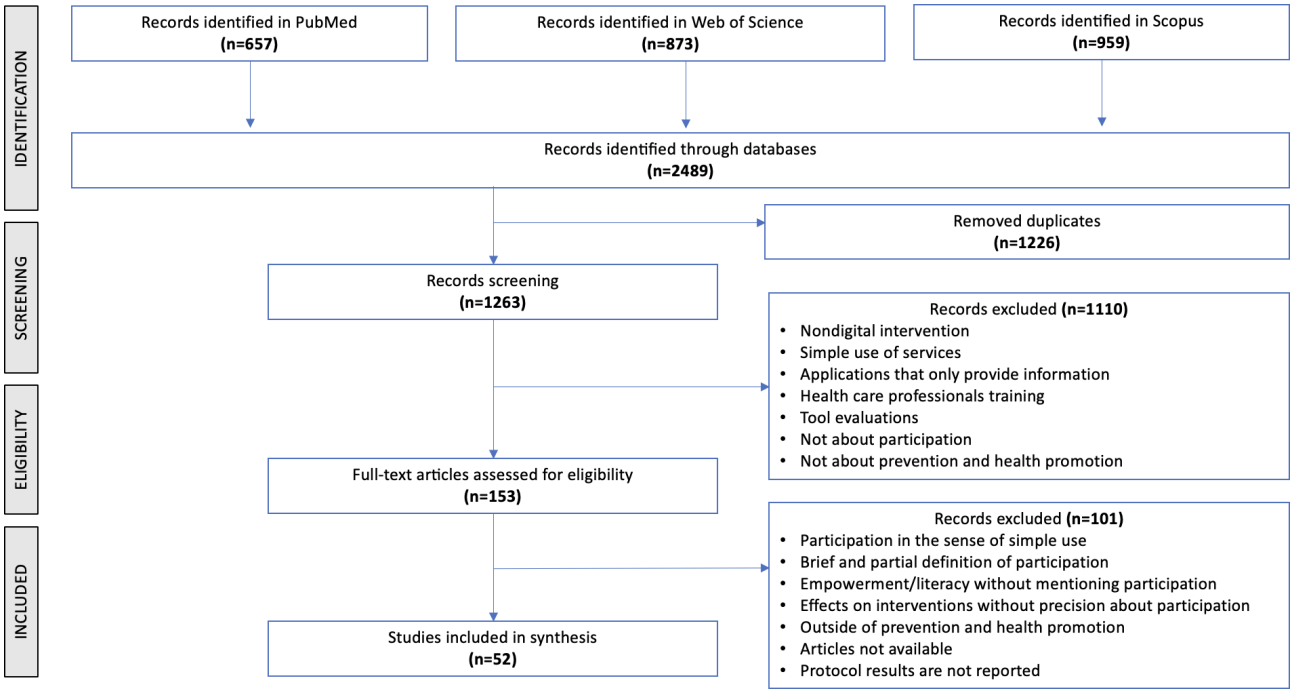
Results

Overview

A total of 2489 articles were identified, 1226 duplicates removed, resulting in 782 unique articles. After reviewing the titles and abstracts, 153 articles were selected to read in full. Following a review of the complete articles and application of exclusion criteria, 52 articles were finally selected for analysis. Our selection process is illustrated in a flowchart (Figure 1).

XSL•FO
RenderX

Figure 1. Flowchart of the article selection process.



Description

A total of 52 articles were analyzed in this study. The chronological distribution indicates that 25 (48%) articles were published in the last 5 years (2020 - 2025), which corresponds to nearly the same number as those published over the previous 20 years (2000 - 2020), with 27 (52%) articles. Of the articles reviewed, 39 focused on digital interventions for a specific target audience, with 14 of these specifically aimed at adolescents and young adults. The 16 articles that are not about a digital intervention for a target audience are thematic articles that do not address a specific digital health intervention. In terms of health topics, 40 articles focused on a specific health issue, with prevention and/or health promotion being the most common topic in 7 articles, followed by AIDS in 5 articles, obesity in 4 articles, and tobacco in 3 articles (see Multimedia Appendix 2) [10,14,17-21,26-70].

Depending on the article, some or all of the information regarding definitions and attributes of participation, scales and methods of assessment, techniques used to involve users, and facilitators and barriers to implementing participation were obtained (see Multimedia Appendix 3) [10,14,17-21,26-70].

The results are organized into the following main subcategories to guide interpretation: (1) engagement attributes and definitions, (2) scale and gradation of participation, (3) methods and techniques used to engage users in digital health interventions, (4) difficulties and barriers to engagement, and (5) facilitators and suggestions to support sustained participation.

Engagement Attributes

First, there are different attributes depending on the discipline: according to scientists, participation is adherence and satisfaction, while for computer scientists, it corresponds to increased attention and enjoyment [21]; in behavioral literature, participation is frequency or duration, while for the gamification industry, it is subjective experience with the service including affect, interest, attention, and flow [18,26].

One recurring way of conceptualizing engagement is the duration and frequency with which the participant uses the eHealth intervention [10,21,27]. In the literature, engagement is generally considered to be ease of use, measuring interactions with the features and functions of the digital solution [17,18]. Engagement with digital health interventions can be defined by using quantity (eg, total number of times the intervention was accessed), duration (eg, total duration of use), frequency (eg, patterns of use), and depth (eg, content consumed during the intervention) [28,29].

However, when it comes to the frequency or duration of time spent using a digital service, some studies suggest that low frequency of use can be just as effective as higher frequency [26], and an increase in usage time can signal low ease of use rather than attractive content [20]. In the context of mHealth, engagement behavior is frequently assessed through passive measures of application usage [30], which fail to provide a comprehensive understanding of engagement with a digital platform [19]. Additionally, there is ongoing discourse surrounding the definition of this “more than usage” [18]. A study indicates that it is important to focus on “effective engagement” rather than simply “increased engagement” [14]. In 3 articles, the authors differentiate between engagement (meaning interactions with the tool) and effective engagement (sufficient participation in the intervention to achieve the desired effects) [14,20,31]. To achieve these desired effects, there is a need to determine an optimal duration and frequency of engagement [20]. However, the concept of optimal engagement dose is not yet clear in the field of digital health interventions [14,18]. In addition, the concept of participation is often confused with compliance, which refers to an individual following an instruction and using the intervention as intended by the developers [20,21,32].

XSL•FO
RenderX

In addition, to characterize participation, some studies focus on the user's experience, but they do not define experience in the same way [21]. For some, user experience is about interaction [17,19] or the desire of a user to use the application for longer and at frequent intervals. User experience quality is characterized by attributes of challenge, positive affect, endurance, aesthetic and sensory appeal, attention, feedback, variety/novelty, interactivity, and perceived user control [33] and includes sensual (aesthetic or novel elements that promote attention and interest), emotional (positive or negative effect elicited), and spatiotemporal (perceptions and awareness of time and the physical environment that the user experiences during use) categories [10].

Although authors assert that the subjective user experience can provide information about the degree of immersion in an application, it does not generally explain the cognitive engagement in the behavior change process. A health application can be entertaining and have good usability, but it does not necessarily lead the user to reflect on their behavior or to take action toward achieving their goals [26].

In one study, to define participation, the aforementioned elements of usage and experience are combined. Engagement in digital interventions is defined through the measures of usage extent (eg, quantity, frequency, duration, and depth) and subjective experience (eg, attention, interest, and affect) [14,34,35]. In the field of digital health behavior change interventions, engagement should therefore integrate objective measures (defined based on usage habits measured using various tools to track the number of logins, time spent online, and amount of content used during the intervention period, and also physiological measures using wearable sensors such as heart rate and electrodermal activity, as well as eye-tracking to determine psychophysiological measures) and subjective measures (defined based on self-evaluation questionnaires measuring levels of engagement with digital games and the intervention) [14,20].

Finally, the literature includes some definitions focused on the objective to be achieved: health benefit [32,36], changes for benefit [37], or behavior change [38]. One source defines participation in eHealth and health behavior change as "the process of involving users in health content in a way that motivates and drives health behavior change," while specifying that this is influenced by a number of variables [14,37-39] and is a complex process because it is evolving [39]. Engagement itself is influenced by intervention characteristics such as content, delivery mode, and contextual characteristics such as physical environment and individual characteristics; there are also potential moderators and mediators of the engagement process [21]. This multidimensional aspect is also mentioned in several articles, where engagement is described as a multidimensional process that involves, among other things, cognitive, emotional, behavioral, and social dimensions [27,30,33,36,40]. One study examined 3 fundamental dimensions: behavior, cognition, and affect. At the behavioral level, the successful usage of a given system is characterized by its routine integration into the daily lives of its users, ease of use, and intuitive interface design that requires minimal effort, and dynamic adaptability to meet the needs of users in different contexts or periods. From a cognitive perspective, mHealth must function as an effective instrument to facilitate the achievement of objectives, while concurrently offering novel perspectives through the provision of pertinent, personalized information, thereby sustaining motivation over an extended period. Finally, on an affective level, a positive experience translates into the pleasure of making progress, general satisfaction in using the application, and a feeling of attachment to the intervention, to the point of feeling its absence when it is not available [18]. Thus, for evaluation, engagement is too complex to be evaluated by a single method, and a combination of several methods should be used [21,41]. The advent of novel technologies has precipitated the application of artificial intelligence, and more specifically, machine learning techniques, in the measurement of a diverse array of engagement. These phenomena encompass goal setting, learning, problem-solving, affect, and reflection, among others [40].

Scale and Gradation of Participation

The gradation of participation in articles is present in two different ways: either within the various stages of the process or through the establishment of a gradient that characterizes a level of engagement, ranging from low to high.

The conceptual model in Yardley et al [71] distinguishes between the micro and macro levels of engagement. *Micro-engagement* refers to the user's immediate interactions with the technology's features, including the extent of intervention use (eg, number of completed activities) and the user's experience (eg, level of interest and attention during activity completion, specifically linked to attention, interest, and affect). *Macro-engagement*, on the other hand, refers to how the user engages with the overall goal of behavior change. After a period of effective micro-engagement, the user may disengage from the platform while still being immersed in the behavior change process [10,18,21,28,34]. Another similar method is proposed by Cole-Lewis and colleagues, defining *Big E* as engagement in the targeted health behavior, and *little e* as engagement in the digital intervention for behavior change [17,40,42].

Regarding scales of participation ranging from least engaged to most engaged, scales are identified in some articles, but they are not specific to digital health interventions [43-48]. First, the most common participation scale is Arnstein's 8-rung ladder of citizen participation, ranging from nonparticipation in citizen power through the stages of manipulation, information, consultation, advocacy, partnership, delegation, control, and self-management [43]. Additionally, there are classifications in terms of usage quantity: an active user may post on a forum, for example, whereas a passive user may read an informative article [44]. Alternatively, a heavy user is one who uses the intervention several times a day or 1 - 2 days per week, and a light user is one who uses the intervention twice per month or once per month, or who has not used it in the past month [45]. There is also a 4-position psychological scale on a continuum ranging from minimum to maximum engagement [46], patient participation levels at 3 levels in the health care sector [32], a 4-stage scale representing the developmental nature of patient activation [47], and other specifically designed questionnaires

XSL•FO
RenderX

for a study [47,48]. Taki et al, on the other hand, suggest a mathematical indicator of engagement that uses 5 different indices to assess user engagement based on usage [20,72].

User typologies are also found. One study classifies users based on their alcohol consumption: "trackers" monitor and track their alcohol consumption, "cut-downers" intend to reduce their alcohol consumption (similar to trackers, they used not only the monitoring features of the application but also goal-setting and feedback functions), and "noncommitters" lack the motivation to use health applications [10]. Another study categorizes users into 3 groups based on their frequency of engagement in peer interactions: "conversation starters," who initiated the highest number of threads; "frequent posters," who posted the highest number of messages on forums; and "conversation attractors," whose messages received the most responses [19].

Methods/Techniques Used to Engage Users

Many authors emphasize the need for theory-based interventions in order to engage users [34]. A theory-based intervention is a systematic approach that uses theoretical models to understand how behaviors are influenced by factors such as beliefs, attitudes, and social norms. By applying these theories to the design of digital health interventions, designers can create programs that specifically target these factors, ensuring that they are based on solid data and effectively encourage user participation [28,46,49]. These theories can be specific to health behavior change models or specific to digital interventions [28]: derivatives of behavioral and neurocognitive theories and models such as the theory of planned behavior, social cognitive theory, self-regulation theory, and social learning theory or unified theory of acceptance and use of technology, which is a theoretical model to explain the adoption and use of technology by users [49,50].

Similarly, the use of appropriate behavior change techniques (BCTs) can also increase user engagement and help them maintain healthy behaviors in the long term. This can improve the overall user experience and make digital health interventions more effective. BCTs are the smallest active components of an intervention that are capable of changing behavior [31,34,51]. The 6 most common BCTs associated with engagement were social support, goal setting, feedback, prompts/cues, self-monitoring, and rewards [31,34]. However, the author claims that they depend on their design, dose, and duration, as well as their content, to optimize their effectiveness. The most advanced and complete BCT would be useless if it did not generate interest through an attractive design, or if it was not user-friendly and intuitive to use in terms of digital functionalities [33].

Behavioral incentive strategies, including action planning, prompts, and habit formation, have been demonstrated to promote sustained use. It is imperative that techniques such as goal setting with visualization and tracking of progress [52], self-monitoring, biofeedback, and behavioral feedback are prioritized for development. Furthermore, providing a range of rewards and incentives that align with distinct core values has been demonstrated to foster sustained engagement, given the recognition that user responses to incentives vary [51]. In addition, one of the most widespread methods for engaging users is user involvement in the intervention design, a process that takes different names such as user-centered design [37,42,53,54], cocreation [55-58], codevelopment [52], or participatory design [46,59]. It is important to involve end users from the beginning of the process to ensure that the intervention is relevant and accepted by users in order to trigger effective participation afterwards [45,60,61] and increase technology literacy and access [52]. Users were also invited to implement and/or evaluate the intervention [42,62,63]. Some authors mentioned that the process is spiral and works in a loop: ideas are generated by/with the community, implemented by the community, evaluated by the community, adjusted based on community feedback, and then implemented and evaluated [41,42,54,64].

Apart from being used for co-design, personalization is a technique used on its own to engage users [60,65]. Some articles mention the need to gather individuals' needs to tailor and personalize interventions in order to engage end users [14,27,30,46,60,66], for example, in adapting content by personalizing it according to variations in psychological, social, and behavioral profiles [28,50] or based on user typologies determined during application registration (eg, through a short questionnaire and targeting the application's content according to their typology) [10]. It is imperative to ensure that the data accurately reflects the individual and their situation in order to guarantee the credibility and accuracy of information and feedback [51,58]. The overall goal is to provide instant contextual support for targeted behaviors when the individual is most likely to be receptive. Just-in-time adaptive interventions could use sensory data, for example, a smartphone or smartwatch, and momentary information directly from participants [60].

Thus, autonomy is also one of the techniques used to engage users [30]. In a structured and self-directed behavior change intervention, all participants receive the same content, but each individual is encouraged to choose their own behavioral goals and activities [47,67]. In order to provide specific recommendations based on individuals' needs, personalized feedback can be used [27,28,48].

Regarding users, some authors also suggest using online peer communities to engage users as they provide resources for sharing experiences and achieving better outcomes. These digital health communities may include discussion forums, online support groups, and social networks for patients [27,30,32,41,43,50,65,68,69]. These digital health communities are particularly important for young people and marginalized populations as they provide an opportunity to address crucial behavior change issues [69] and allow social interaction, feeling supported, and understood by the community [19,28,66] to enhance participants' engagement in the program and foster their motivation [27], mentioned as a key to engaging users [26,30,33,39]. Regarding this, despite evidence of improved intervention engagement in physical interventions, motivational support such as motivational interviewing is rarely present in health apps for smartphones. One author indicates that the reason for the limited use of motivational interviewing may be its lack of a coherent theoretical framework and proposes that self-determination theory can be used as a theoretical basis

XSL•FO
RenderX

because it shares fundamental principles with motivational interviewing [26]. Self-determination theory, used by several authors, emphasizes autonomy, competence, and social relatedness to encourage user participation [26,41,65].

While motivational interviewing provides tools for self-reflection and satisfaction, gamification can provide experiences of autonomy, competence, and relatedness by adding fun and excitement to activities [14,26,28,43,55,59], and gamification can enhance hedonic and intrinsic motivation by addressing the psychological and emotional requirements of individuals through the use of game-like components [45,58]. Intrinsic motivation refers to performing an activity solely for pleasure, excitement, and interest. Authors assert that gamification of health applications is a promising approach to counter the often-decreasing long-term motivation of health application users [45].

Finally, emotion has been identified as a key mechanism of action that can engender user experience linked to engagement. Ensuring high user satisfaction with mHealth and "avoiding negative emotions" such as frustration, disappointment, or anxiety should result in good long-term engagement and an increase in the number of positive reviews [31,51].

Difficulties in Engaging Users

Authors encountered several barriers to user engagement in digital interventions. At the country level, in several countries and populations, technological adoption has been slow due to cost, infrastructure, design models, architecture, integration, usability, and the implementation of public policies. Adoption and deployment are also dependent on the training and education of users provided by authorities [32,51].

At the intervention level, difficulty in involving all populations in participatory design is encountered due to knowledge asymmetry [63]. Lack of user involvement in the design and a lack of desired features lead to reluctance to adopt the intervention [45,65,66]. In the long term, maintaining usage is also a recurring problem [14,45,64]: there is evidence indicating that people often download an app and never use it again, with a rapid decline in usage after the first download [73,74]. Research on the specific reasons for user abandonment and low engagement in online programs has identified a lack of motivation, an absence of goal tracking, and negative feedback when goals are not met [52].

Regarding the content of interventions, there is a lack of theory-based mHealth interventions, despite the fact that some behavior change theories have been well-validated and tested in evidence-based prevention, diagnostic, and care interventions [61]. As for interfaces, some studies emphasize that interfaces are non–user-friendly [29,43,50] and that it is challenging to compete with social media and other entertainment-based applications [28,60]. As for all other online tools, data security is also a recurring concern, particularly with respect to anonymity [43,50]. Once the results of an intervention are obtained, the literature mentions the difficulty of generalizing the results to other populations due to differences in context. What works in one context for a given subgroup of the population may be less effective or even harmful elsewhere for other subgroups [37].

At the user level, technical barriers and a lack of technological skills are cited as a hindrance to participation [42,43,66]. In addition, improving health literacy directly impacts the ability to act on health information and take greater control of health as individuals, families, and communities, and is considered an essential condition for patient participation [32,36,50,51]. Engagement also varies according to sociodemographic characteristics; young women and individuals with a higher level of education are more involved [50]. Additionally, lack of support [43,66] and lack of motivation were associated with low adoption of mHealth apps [45].

Facilitators and Suggestions to Engage Users

At the user level, engagement depends on several interdependent components such as social norms, positive and negative reinforcement factors, goal-setting, self-monitoring, self-evaluation [17,49], self-efficacy [70], social support [19,49,67], motivation [34,45], trust and adherence of the participants' family, and peer modeling [37,50], especially through social media and friends or any scientific authority [45,58]. In addition, the delivery of human support has been demonstrated to positively influence engagement with digital behavior change interventions and increase adherence with mHealth interventions. Support personnel like health coaches provided individualized and personalized support, and facilitated comprehension and application of BCTs, including goal setting and receiving feedback on behaviors that had been monitored [31].

The social facilitators associated with the intervention encompass the following factors: the establishment of positive relationships and effective communication, the development of trust in the intervention's leader, the usage of the program within a social context, and the cultivation of positive connections with fellow participants. The temporal and spatial flexibility facilitators of the intervention are characterized by their accessibility and adaptability to different environments (performed within the comfort and familiarity of the participant's home, thereby eliminating the need for travel) [70].

Regarding intervention features, gamification [19,26,33,45,51,59], accessibility and ease of use [14], interactivity [67], attractiveness [55,60], confidentiality and security of personal data [14], and adapting the material for potential users from socially or educationally disadvantaged backgrounds [37] all help users to engage and are avenues to explore to increase user engagement within digital health interventions. For instance, the configuration of push notifications could be adapted to align with the preferences of the end user, or alternatively, users have the option to disable them entirely. The hypothesis was that these customizations would provide a more pleasant user experience and facilitate the daily use of the mHealth apps [50,58].

Regarding the design of the interventions, results show that involving users at all stages [26,41], addressing basic psychological needs [26] including psychosocial determinants anchored in relevant behavioral theories, such as the social

XSL•FO
RenderX

ecological model, the COM-B (capability, opportunity, motivation—behavior) model, and social cognitive theory [17], using more systematic and progressive approaches for intervention development and evaluation [48], integrated skill acquisition, positive reinforcement, integration of cultural values and group norms [49] could be explored. One study suggests letting people address their own problems in their own way, but within a structured and supportive framework [67].

A conceptual synthesis of the determinants influencing user engagement in digital health interventions, as identified in this review, is presented in Multimedia Appendix 4.

Discussion

Principal Results

This paper identified the attributes and definitions of user engagement, examined the methods used for engaging users in mHealth, and identified barriers and facilitators encountered in implementing user participation. In the literature, a variety of information was found in the articles. Some articles contained information in all the categories (definition or attributes, scale or gradation of participation, methods, and facilitators or barriers of participation), while others provided detailed information only on one category. In 18 articles, the authors mention attributes or a definition concerning participation. Engagement in digital health interventions is described as multidimensional, encompassing both objective measures (eg, usage frequency, duration, and depth) and subjective experiences (eg, attention, interest, and affect), as well as the goal of behavior change. Methods used to involve users are the most frequently cited, with 38 articles. Main methods and techniques used are theory-based intervention design, BCTs, personalization, user-centered and participatory design, peer support, autonomy, gamification, and emotion management. The obstacles encountered in involving users are discussed in 19 articles. Difficulties in user engagement in digital health interventions arise from technical, contextual, and individual barriers, while facilitators include participatory approaches, personalization, gamification, social support, and human guidance to enhance motivation, adherence, and sustained use. Some articles deal with digital interventions in general, while others are based on one specific digital intervention. The literature on this topic is recent, with half of the articles selected dating from the last 5 years.

Combining Approaches to Define Participation

The characterization of participation in digital health interventions for health promotion and prevention is complex [39] and influenced by different factors [40]. This is consistent with the complexity of digital health interventions themselves, which often involve multiple components [14,21,27,30,37-39]. Addressing this complexity requires integration of multiple approaches. The first approach is a quantitative perspective, which conceptualizes participation through measurable indicators such as duration and frequency, focusing on the extent of tool usage [10,17,21,27,28]. The second is a user experience approach, which captures subjective dimensions including attention, interest, emotions, and perceptions [19,24,28,35]. Combining these approaches enables a more comprehensive understanding of effective user engagement with mHealth tools in achieving behavior change goals of users [32,36-38]. In this context, user participation in the intervention becomes a prerequisite for the involvement of the user with their behavior change objective [17,71]. This engagement is possible when users consume an optimal intervention dose [20]. Engagement can thus be conceptualized as a continuum, ranging from minimal involvement to an excessive, potentially counterproductive level. The aim is to identify and deliver the appropriate dose of engagement for each individual, thereby enabling the achievement of their behavior change goals.

Participation Is the Goal but How Can It Be Reached?

To achieve engagement, the literature mentions various methods and techniques in three levels: (1) user-related factors, (2) intrinsic characteristics of the intervention, and (3) implementation context. It is therefore essential to have a holistic vision to ensure the effective condition of participation.

User-Related Factors

Thus, mHealth interventions should enable individuals to enter into a process that allows them to make choices aimed at maintaining or improving their health status, as stated in the Ottawa Charter [6]. This means (1) being able to improve their individual ability to act through the use of effective BCTs such as Michie's [75,76] BCTs, (2) engaging in collective discussions that correspond to the needs and demands of users, (3) providing social support such as the collective discussion that founded in this review, (4) becoming aware of the effect of their environment on their behavior and being able to act on it [27,30,32,41,43,65,68,69], and (5) finding external human resources that improve the effectiveness of applications [77].

Intrinsic Characteristic of Intervention

To better understand users' needs and involve them in the development process, one of the most commonly used methods for engaging users is cocreation, which involves users in the creation phase of the tool. This cocreation is even more valuable because one of the reasons for the long-term decline in tool usage is the absence of functionalities that users need. However, it is not sufficient to simply include users in the development process, as the mHealth functionalities must be designed to evolve with users' needs and demands. The process is a spiral and works in a loop [41,42,54,64]. Users' needs vary significantly over time, and these changes are influenced by many factors, such as life events, health experiences, and interactions with the health care system. Therefore, it is advisable to monitor users' evolving needs over time to adapt the digital intervention accordingly and meet the changing needs of users [18,78]. So, engaging users in just one step is insufficient. Perhaps, it is necessary to rethink the design methods of applications that currently attempt to suit everyone, and which, as a result, are suited to no one, even if they are intended for a targeted population group. The design should be based on a user-centered approach that incorporates cocreation methods and iterative testing. It should also offer possibilities for adaptive personalization based on behavioral profiles.

Engagement techniques are grounded in robust theoretical models, such as the COM-B framework, and can be bolstered

XSL•FO
RenderX

by gamification strategies and interactive features that provide personalized feedback. The overall digital experience should combine an intuitive and ergonomic interface for ease of use with careful design for attractiveness.

Implementation Context

Data security is also a recurring concern in terms of the functionality of the intervention [43]. Furthermore, once the results of the intervention have been obtained, there may be difficulties in generalizing the results to other populations due to differences in context: what may work in a given context for a given subgroup of the population may be less effective or even harmful elsewhere for other subgroups [37]. In addition, engagement is frequently subjective, and the factors that motivate one user to repeatedly use an application may not be applicable to another user, or may not be applicable at a specific instance [35].

These 3 different levels interact dynamically to create a complex mechanism. For example, a poor user experience can reduce motivation and lead to the intervention being abandoned prematurely [79,80], whereas a favorable sociocultural context can reinforce long-term engagement [81].

Conclusions

The current state of knowledge allowed us to have a vision of the participation of users in digital interventions. However, the concept of participation, which is described as complex, evolving, and multifactorial, is a notion on which there is as yet no consensus. This paper offers insights into user participation in mHealth, as evidenced by the results of the review. Usage of the tool, combined with the user's experience of it, serves as an indicator of participation. This participation is a prerequisite for the user to receive the optimal dose of intervention content, thereby supporting the engagement of users with their behavior change objectives. To foster such engagement, a variety of methods can be mobilized. These methods fall into 3 key categories: user-related factors, intrinsic characteristics of the intervention, and the context of implementation. Such clarifications could then inform the methodology for evaluating and developing digital interventions. In order to address this gap, the results of this review are used in an e-Delphi study with the aim of establishing a consensus definition with experts. The results of this forthcoming study will be the subject of a subsequent publication.

Acknowledgments

This manuscript was prepared without the use of generative AI tools for ideation, writing, data analysis, interpretation, or figure creation. All content was produced and reviewed by the authors in accordance with JMIR Publications' transparency and author accountability policy. DeepL Write was used solely to assist in enhancing the clarity and quality of the manuscript.

Funding

MA's doctoral thesis was supported by the University Research Schools initiative founded through the PIA3 (project reference 17-EURE-0019) as part of the Graduate Program Digital Public Health (DPH) and the MedPrev Project of the Regional Union of Private Practice Physicians of Nouvelle-Aquitaine. This work was also supported by the Ligue Contre le Cancer through a research team label. The funders had no role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript.

Data Availability

All data generated or analyzed during this study are included in this published article and its multimedia appendices.

Authors' Contributions

Conceptualization: LC (lead), OA (equal), MA (supporting)
Data curation: OA (lead), MA (equal)
Formal analysis: MA (lead), OA (supporting)
Funding acquisition: LC (lead), OA (equal)
Investigation: MA (lead), OA (equal)
Methodology: LC (lead), OA (equal), MA (supporting)
Project administration: LC (lead), OA (equal)
Resources: LC (lead), OA (equal)
Supervision: LC (lead), OA (equal)
Validation: LC (lead), OA (equal)
Visualization: MA (lead), OA (supporting)
Writing – original draft: MA (lead), OA (supporting), LC (supporting)
Writing – review & editing: MA (lead), OA (supporting), LC (supporting)

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Search equations and keywords.
[PDF File, 91 KB - mhealth_v14i1e66002_app1.pdf]

Multimedia Appendix 2
General characteristics of articles.
[DOCX File, 30 KB - mhealth_v14i1e66002_app2.docx]

Multimedia Appendix 3
Information contained in the articles.
[DOCX File, 39 KB - mhealth_v14i1e66002_app3.docx]

Multimedia Appendix 4
Conceptual synthesis of the determinants influencing user engagement in digital health interventions.
[DOCX File, 150 KB - mhealth_v14i1e66002_app4.docx]

Checklist 1
PRISMA-ScR checklist.
[PDF File, 93 KB - mhealth_v14i1e66002_app5.pdf]

References

1. Tuberculosis: global task force on digital health. World Health Organization. 2015 Jun 3. URL: https://who-dev5.prgsdev.com/m/news-room/questions-and-answers/item/tuberculosis-global-task-force-on-digital-health [accessed 2026-04-25]
2. Lupton D. The use and value of digital media for information about pregnancy and early motherhood: a focus group study. BMC Pregnancy Childbirth 2016 Jul 19;16(1):171. [doi: 10.1186/s12884-016-0971-3] [Medline: 27435182]
3. Green Paper on mobile health ("mHealth"). European Commission. 2014 Apr 10. URL: https://digital-strategy.ec.europa.eu/en/library/green-paper-mobile-health-mhealth [accessed 2026-04-08]
4. Santé connectée: de la e-santé à la santé connectée [in French]. Conseil National de l'Ordre des Médecins. 2015 Jan. URL: https://www.conseil-national.medecin.fr/sites/default/files/2025-07/medecins-sante-connectee.pdf [accessed 2026-04-25]
5. Cambon L. Health smart devices and applications…towards a new model of prevention? Eur J Public Health 2017 Jun;27(3):390-391. [doi: 10.1093/eurpub/ckx019]
6. Ottawa charter for health promotion. World Health Organization. 1986. URL: https://www.who.int/teams/health-promotion/enhanced-wellbeing/first-global-conference [accessed 2026-04-08]
7. Closing the gap in a generation: health equity through action on the social determinants of health. World Health Organization. 2008 Aug 27. URL: https://www.who.int/publications/i/item/9789241563703 [accessed 2026-04-08]
8. Global strategy on human resources for health: workforce 2030. World Health Organization. 2016. URL: https://iris.who.int/items/d1ad8cde-1d9d-48d1-a1b0-42d33c4ce629 [accessed 2026-04-08]
9. Or CKL, Valdez RS, Casper GR, et al. Human factors and ergonomics in home care: current concerns and future considerations for health information technology. Work 2009;33(2):201-209. [doi: 10.3233/WOR-2009-0867] [Medline: 19713630]
10. Milward J, Deluca P, Drummond C, Kimergård A. Developing typologies of user engagement with the BRANCH alcohol-harm reduction smartphone app: qualitative study. JMIR Mhealth Uhealth 2018 Dec 13;6(12):e11692. [doi: 10.2196/11692] [Medline: 30545806]
11. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 6;20(6):e199. [doi: 10.2196/jmir.9275] [Medline: 29875089]
12. Kim J, Lim S, Min YH, et al. Depression screening using daily mental-health ratings from a smartphone application for breast cancer patients. J Med Internet Res 2016 Aug 4;18(8):e216. [doi: 10.2196/jmir.5598] [Medline: 27492880]
13. Percentage of mobile apps that have been used only once from 2010 to 2019. Statista. 2019 May. URL: https://www.statista.com/statistics/271628/percentage-of-apps-used-once-in-the-us/ [accessed 2026-04-08]
14. Saleem M, Kühne L, De Santis KK, Christianson L, Brand T, Busse H. Understanding engagement strategies in digital interventions for mental health promotion: scoping review. JMIR Ment Health 2021 Dec 20;8(12):e30000. [doi: 10.2196/30000] [Medline: 34931995]
15. Rosenfelder S. New data shows uninstalls remain a significant pain for apps. AppsFlyer. 2018 May 16. URL: https://www.appsflyer.com/blog/trends-insights/app-uninstall-trends/ [accessed 2026-04-25]
16. Weiss D, Rydland HT, Øversveen E, Jensen MR, Solhaug S, Krokstad S. Innovative technologies and social inequalities in health: a scoping review of the literature. PLoS One 2018;13(4):e0195447. [doi: 10.1371/journal.pone.0195447] [Medline: 29614114]

XSL•FO
RenderX

17. Cole-Lewis H, Ezeanochie N, Turgiss J. Understanding health behavior technology engagement: pathway to measuring digital behavior change interventions. JMIR Form Res 2019 Oct 10;3(4):e14052. [doi: 10.2196/14052] [Medline: 31603427]
18. Kelders SM, Kip H, Beerlage-de Jong N, Köhle N. What does it mean to be engaged with digital health interventions? A qualitative study into the experiences of engaged users and the views of professionals. Digit Health 2024;10:20552076241283530. [doi: 10.1177/20552076241283530] [Medline: 39376944]
19. Myneni S, Sridharan V, Cobb N, Cohen T. Content-sensitive characterization of peer interactions of highly engaged users in an online community for smoking cessation: mixed-methods approach for modeling user engagement in health promotion interventions. J Particip Med 2018 Jul 24;10(3):e9. [doi: 10.2196/jopm.9745] [Medline: 33052116]
20. Alshurafa N, Jain J, Alharbi R, Iakovlev G, Spring B, Pfammatter A. Is more always better? Discovering incentivized mHealth intervention engagement related to health behavior trends. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Dec;2(4):153. [doi: 10.1145/3287031] [Medline: 32318650]
21. Short CE, DeSmet A, Woods C, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292. [doi: 10.2196/jmir.9397] [Medline: 30446482]
22. Romero-Portier C, Darlington E. Comment promouvoir la participation dans les projets de promotion de la santé? Les facteurs clés selon les professionnels [How can participation in health promotion projects be encouraged? Professionals' views on the key factors at play]. Rev Epidemiol Sante Publique 2022 Aug;70(4):147-155. [doi: 10.1016/j.respe.2022.05.005] [Medline: 35715346]
23. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
24. Fung A, Wright EO. Deepening democracy: innovations in empowered participatory governance. Polit Soc 2001 Mar;29(1):5-41. [doi: 10.1177/0032329201029001002]
25. Hart RA. Children's participation: from tokenism to citizenship. UNICEF International Child Development Centre. 1992 Mar. URL: https://digitallibrary.un.org/record/227219 [accessed 2026-04-08]
26. Nurmi J, Knittle K, Ginchev T, et al. Engaging users in the behavior change process with digitalized motivational interviewing and gamification: development and feasibility testing of the Precious app. JMIR Mhealth Uhealth 2020 Jan 30;8(1):e12884. [doi: 10.2196/12884] [Medline: 32003750]
27. Nitsch M, Dimopoulos CN, Flaschberger E, et al. A guided online and mobile self-help program for individuals with eating disorders: an iterative engagement and usability study. J Med Internet Res 2016 Jan 11;18(1):e7. [doi: 10.2196/jmir.4972] [Medline: 26753539]
28. Hightow-Weidman LB, Horvath KJ, Scott H, Hill-Rorie J, Bauermeister JA. Engaging youth in mHealth: what works and how can we be sure? Mhealth 2021;7:23. [doi: 10.21037/mhealth-20-48] [Medline: 33898592]
29. Grieve N, Braaten K, MacPherson M, Liu S, Jung ME. Involving end users in the development and usability testing of a smartphone app designed for individuals with prediabetes: mixed-methods focus group study. JMIR Form Res 2025 Feb 11;9(1):e59386. [doi: 10.2196/59386] [Medline: 39935015]
30. Wagner B, Liu E, Shaw SD, et al. e-Wrapper: operationalizing engagement strategies in mHealth. In: UbiComp '17: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers: Association for Computing Machinery; 2017:790-798. [doi: 10.1145/3123024.312561]
31. Hawkes RE, Miles LM, Ainsworth B, Ross J, Meacock R, French DP. Engagement with a nationally-implemented digital behaviour change intervention: usage patterns over the 9-month duration of the National Health Service Digital Diabetes Prevention Programme. Internet Interv 2023 Sep;33:100647. [doi: 10.1016/j.invent.2023.100647] [Medline: 37502122]
32. Marin HF, Delaney C. Patient engagement and digital health communities. In: Global Health Informatics: How Information Technology Can Change Our Lives in a Globalized World: Academic Press; 2017:218-231. [doi: 10.1016/B978-0-12-804591-6.00011-2]
33. Fulton EA, Kwah KL, Wild S, Brown KE. Lost in translation: transforming behaviour change techniques into engaging digital content and design for the StopApp. Healthcare (Basel) 2018 Jul 6;6(3):75. [doi: 10.3390/healthcare6030075] [Medline: 29986396]
34. Milne-Ives M, Homer SR, Andrade J, Meinert E. Potential associations between behavior change techniques and engagement with mobile health apps: a systematic review. Front Psychol 2023;14:1227443. [doi: 10.3389/fpsyg.2023.1227443] [Medline: 37794916]
35. White BK, Burns SK, Giglia RC, Dhaliwal SS, Scott JA. Measuring user engagement with a socially connected, gamified health promotion mobile app. Int J Environ Res Public Health 2022 May 5;19(9):5626. [doi: 10.3390/ijerph19095626] [Medline: 35565015]
36. Musso M, Pinna R, Trombin M, et al. Social media to improve health promotion and health literacy for patients engagement. In: Zaramenskikh E, Fedorova A, editors. Digital Transformation and New Challenges: Springer; 2020:103-120. [doi: 10.1007/978-3-030-43993-4_10]
37. Rai T, Morton K, Roman C, et al. Optimizing a digital intervention for managing blood pressure in stroke patients using a diverse sample: integrating the person-based approach and patient and public involvement. Health Expect 2021 Apr;24(2):327-340. [doi: 10.1111/hex.13173] [Medline: 33316120]

XSL•FO
RenderX

38. Craig Lefebvre R, Tada Y, Hilfiker SW, Baur C. The assessment of user engagement with eHealth content: the eHealth engagement scale. J Comput Mediat Commun 2010;15(4):666-681. [doi: 10.1111/j.1083-6101.2009.01514.x]
39. Kim M, Yang J, Ahn WY, Choi HJ. Machine learning analysis to identify digital behavioral phenotypes for engagement and health outcome efficacy of an mHealth intervention for obesity: randomized controlled trial. J Med Internet Res 2021 Jun 24;23(6):e27218. [doi: 10.2196/27218] [Medline: 34184991]
40. Giovanelli A, Rowe J, Taylor M, et al. Supporting adolescent engagement with artificial intelligence-driven digital health behavior change interventions. J Med Internet Res 2023 May 24;25(1):e40306. [doi: 10.2196/40306] [Medline: 37223987]
41. Skinner H, Morrison M, Bercovitz K, et al. Using the internet to engage youth in health promotion. Promot Educ 1997 Dec;4(4):23-25. [doi: 10.1177/102538239700400411] [Medline: 9560853]
42. Lawrence K, Rodriguez DV, Feldthouse DM, et al. Effectiveness of an integrated engagement support system to facilitate patient use of digital diabetes prevention programs: protocol for a randomized controlled trial. JMIR Res Protoc 2021 Feb 9;10(2):e26750. [doi: 10.2196/26750] [Medline: 33560240]
43. Schroeer C, Voss S, Jung-Sievers C, Coenen M. Digital formats for community participation in health promotion and prevention activities: a scoping review. Front Public Health 2021;9:713159. [doi: 10.3389/fpubh.2021.713159] [Medline: 34869143]
44. Baltierra NB, Muessig KE, Pike EC, LeGrand S, Bull SS, Hightow-Weidman LB. More than just tracking time: complex measures of user engagement with an internet-based health promotion intervention. J Biomed Inform 2016 Feb;59:299-307. [doi: 10.1016/j.jbi.2015.12.015] [Medline: 26732997]
45. Mustafa AS, Ali N, Dhillon JS, Alkawsi G, Baashar Y. User engagement and abandonment of mHealth: a cross-sectional survey. Healthcare (Basel) 2022 Jan 24;10(2):221. [doi: 10.3390/healthcare10020221] [Medline: 35206837]
46. Graffigna G, Bosio C, Savarese M, Barello M, Barello S. "#I-Am-Engaged": conceptualization and first implementation of a multi-actor participatory, co-designed social media campaign to raise Italians citizens' engagement in preventing the spread of COVID-19 virus. Front Psychol 2020;11:567101. [doi: 10.3389/fpsyg.2020.567101] [Medline: 33250811]
47. Solomon M, Wagner SL, Goes J. Effects of a web-based intervention for adults with chronic conditions on patient activation: online randomized controlled trial. J Med Internet Res 2012 Feb 21;14(1):e32. [doi: 10.2196/jmir.1924] [Medline: 22353433]
48. Morrison L, Moss-Morris R, Michie S, Yardley L. Optimizing engagement with internet-based health behaviour change interventions: comparison of self-assessment with and without tailored feedback using a mixed methods approach. Br J Health Psychol 2014 Nov;19(4):839-855. [doi: 10.1111/bjhp.12083] [Medline: 24308806]
49. Kwan P, Sabado-Liwag M, Lee C, et al. Development of an online smoking cessation curriculum for Pacific Islanders: a community-based participatory research approach. Prog Community Health Partnersh 2017;11(3):263-274. [doi: 10.1353/cpr.2017.0031] [Medline: 29056618]
50. Ho TQA, Le LKD, Engel L, et al. Barriers to and facilitators of user engagement with web-based mental health interventions in young people: a systematic review. Eur Child Adolesc Psychiatry 2025 Jan;34(1):83-100. [doi: 10.1007/s00787-024-02386-x] [Medline: 38356043]
51. Mair JL, Hashim J, Thai L, et al. Understanding and overcoming barriers to digital health adoption: a patient and public involvement study. Transl Behav Med 2025 Jan 16;15(1):ibaf010. [doi: 10.1093/tbm/ibaf010] [Medline: 40167046]
52. Isakadze N, Marvel FA. What are the features promoting and deterring engagement in mobile health intervention use. JACC Adv 2023 Oct;2(8):100615. [doi: 10.1016/j.jacadv.2023.100615] [Medline: 38938370]
53. DeSmet A, Thompson D, Baranowski T, Palmeira A, Verloigne M, De Bourdeaudhuij I. Is participatory design associated with the effectiveness of serious digital games for healthy lifestyle promotion? A meta-analysis. J Med Internet Res 2016 Apr 29;18(4):e94. [doi: 10.2196/jmir.4444] [Medline: 27129447]
54. Mauka W, Mbotwa C, Moen K, et al. Development of a mobile health application for HIV prevention among at-risk populations in urban settings in East Africa: a participatory design approach. JMIR Form Res 2021;5(10). [doi: 10.2196/23204]
55. Pernencar C, Sousa P, Frontini R, et al. Planning a health promotion program: mobile app gamification as a tool to engage adolescents. Procedia Comput Sci 2018;138:113-118. [doi: 10.1016/j.procs.2018.10.016]
56. van Hierden Y, Dietrich T, Rundle-Thiele S. Designing an eHealth well-being program: a participatory design approach. Int J Environ Res Public Health 2021 Jul 6;18(14):7250. [doi: 10.3390/ijerph18147250] [Medline: 34299700]
57. Laidlaw R, Dixon D, Morse T, Beattie TK, Kumwenda S, Mpemberera G. Using participatory methods to design an mHealth intervention for a low income country, a case study in Chikwawa, Malawi. BMC Med Inform Decis Mak 2017 Jul 5;17(1):98. [doi: 10.1186/s12911-017-0485-6] [Medline: 28679428]
58. Dederichs M, Nitsch FJ, Apolinário-Hagen J. Piloting an innovative concept of e-Mental health and mHealth workshops with medical students using a participatory co-design approach and app prototyping: case study. JMIR Med Educ 2022 Jan 10;8(1):e32017. [doi: 10.2196/32017] [Medline: 35006085]
59. Silva CC, Silva Filho AG, Monteiro JS, et al. The development of digital games for food and nutrition education using popular education and participatory design. In: Wilton TL, editor. Rural Education: Teaching Strategies, Student Issues and Trends: Nova Science; 2018.
60. Partridge SR, Redfern J. Strategies to engage adolescents in digital health interventions for obesity prevention and management. Healthcare (Basel) 2018 Jun 21;6(3):70. [doi: 10.3390/healthcare6030070] [Medline: 29933550]

XSL•FO
RenderX

61. Toefy Y, Skinner D, Thomsen S. “Please don’t send us spam!” A participative, theory-based methodology for developing an mHealth intervention. JMIR Mhealth Uhealth 2016 Aug 17;4(3):e100. [doi: 10.2196/mhealth.6041] [Medline: 27535589]
62. Livingood WC, Monticalvo D, Bernhardt JM, et al. Engaging adolescents through participatory and qualitative research methods to develop a digital communication intervention to reduce adolescent obesity. Health Educ Behav 2017 Aug;44(4):570-580. [doi: 10.1177/1090198116677216] [Medline: 27811164]
63. Njie-Carr VPS, Jones-Parker H, Massey C, Baker D, Nganga-Good C. Leveraging community engagement to develop a mobile health application for older women with HIV infection. J Obstet Gynecol Neonatal Nurs 2018 Nov;47(6):833-843. [doi: 10.1016/j.jogn.2018.08.005] [Medline: 30278158]
64. Sucala M, Ezeanochie NP, Cole-Lewis H, Turgiss J. An iterative, interdisciplinary, collaborative framework for developing and evaluating digital behavior change interventions. Transl Behav Med 2020 Dec 31;10(6):1538-1548. [doi: 10.1093/tbm/ibz109] [Medline: 31328775]
65. van Bruinessen I, van’t Klooster JW, Boessen A, et al. Involving end-users in the development and implementation of a web-based, physical activity platform for elderly. In: IADIS International Conference e-Health 2017: International Association for Development of the Information Society; 2017.
66. Corbett T, Singh K, Payne L, et al. Understanding acceptability of and engagement with web-based interventions aiming to improve quality of life in cancer survivors: a synthesis of current research. Psychooncology 2018 Jan;27(1):22-33. [doi: 10.1002/pon.4566] [Medline: 29044832]
67. Chen Z, Koh PW, Ritter PL, Lorig K, Bantum EO, Saria S. Dissecting an online intervention for cancer survivors: four exploratory analyses of internet engagement and its effects on health status and health behaviors. Health Educ Behav 2015 Feb;42(1):32-45. [doi: 10.1177/1090198114550822] [Medline: 25288489]
68. Levine D, Madsen A, Wright E, Barar RE, Santelli J, Bull S. Formative research on MySpace: online methods to engage hard-to-reach populations. J Health Commun 2011 Apr;16(4):448-454. [doi: 10.1080/10810730.2010.546486] [Medline: 21391040]
69. Ronen K, Grant E, Copley C, Batista T, Guthrie BL. Peer group focused eHealth strategies to promote HIV prevention, testing, and care engagement. Curr HIV/AIDS Rep 2020 Oct;17(5):557-576. [doi: 10.1007/s11904-020-00527-w] [Medline: 32794071]
70. Turcotte S, Bouchard C, Rousseau J, et al. Factors influencing older adults’ participation in telehealth interventions for primary prevention and health promotion: a rapid review. Australas J Ageing 2024 Mar;43(1):11-30. [doi: 10.1111/ajag.13244] [Medline: 38014903]
71. Yardley L, Spring BJ, Riper H, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
72. Taki S, Lymer S, Russell CG, et al. Assessing user engagement of an mHealth intervention: development and implementation of the growing healthy app engagement index. JMIR Mhealth Uhealth 2017 Jun 29;5(6):e89. [doi: 10.2196/mhealth.7236] [Medline: 28663164]
73. Average iPhone app usage declines rapidly after first download. AppleInsider. 2009 Feb 19. URL: https://appleinsider.com/articles/09/02/19/iphone_app_usage_declining_rapidly_after_first_downloads [accessed 2026-04-08]
74. 21% of users abandon an app after one use. Upland Software. 2018. URL: https://uplandsoftware.com/localytics/resources/blog/21-percent-of-users-abandon-apps-after-one-use/ [accessed 2026-04-08]
75. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6(1):42. [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
76. Michie S, Richardson M, Johnston M, et al. The Behavior Change Technique Taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
77. Aromatario O, Van Hoye A, Vuillemin A, et al. How do mobile health applications support behaviour changes? A scoping review of mobile health applications relating to physical activity and eating behaviours. Public Health (Fairfax) 2019 Oct;175:8-18. [doi: 10.1016/j.puhe.2019.06.011] [Medline: 31374453]
78. Sahandi Far M, Eickhoff SB, Goni M, Dukart J. Exploring test-retest reliability and longitudinal stability of digital biomarkers for Parkinson disease in the m-Power data set: cohort study. J Med Internet Res 2021 Sep 13;23(9):e26608. [doi: 10.2196/26608] [Medline: 34515645]
79. Peters D, Calvo RA, Ryan RM. Designing for motivation, engagement and wellbeing in digital experience. Front Psychol 2018;9:797. [doi: 10.3389/fpsyg.2018.00797] [Medline: 29892246]
80. Bennett D, Mekler ED. Beyond intrinsic motivation: the role of autonomous motivation in user experience. ACM Trans Comput-Hum Interact 2024 Oct 31;31(5):1-41. [doi: 10.1145/3689044]
81. Spillers F, Asimakopoulos S. Does social user experience improve motivation for runners? A diary study comparing mobile health applications. In: Marcus A, editor. International Conference of Design, User Experience, and Usability: User Experience Design Practice: Springer; 2014:358-369. [doi: 10.1007/978-3-319-07638-6_35]

XSL•FO
RenderX

Abbreviations

BCT: behavior change technique
COM-B: capability, opportunity, motivation—behavior model
mHealth: mobile health
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by A Stone; submitted 03.Sep.2024; peer-reviewed by A Hashim, N Ribeiro; revised version received 04.Dec.2025; accepted 09.Dec.2025; published 15.May.2026.

Please cite as:
Aktas M, Cambon L, Aromatario O
Engagement of Users in Digital Health Applications: Scoping Review
JMIR Mhealth Uhealth 2026;14:e66002
URL: https://mhealth.jmir.org/2026/1/e66002
doi:10.2196/66002

© Melek Aktas, Linda Cambon, Olivier Aromatario. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 15.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Heart Rate Monitors for the Estimation of Physical Activity in Patients With Cardiovascular Disease: Systematic Review

Paulien Vermunicht[1,2], PhD; Christophe Buyck[1,2], MSc; Sebastiaan Naessens[2], MSc; Wendy Hens[2,3], PhD; Emeline Van Craenenbroeck[1,2], Prof Dr Med; Kris Laukens[4,5], Prof Dr; Lien Desteghe[1,2,6,7], Prof Dr; Hein Heidbuchel[1,2,6], Prof Dr Med

[1]Research Group Cardiovascular Diseases, University of Antwerp, Drie Eikenstraat 655, Antwerp, Belgium

[2]Department of Cardiology, Antwerp University Hospital, Drie Eikenstraat 655, Antwerp, Belgium

[3]Research Group Movant, University of Antwerp, Antwerp, Belgium

[4]Department of Computer Science, University of Antwerp, Antwerp, Belgium

[5]Biomedical Informatics Research Network Antwerp (Biomina), University of Antwerp, Antwerp, Belgium

[6]Faculty of Medicine and Life Sciences, Hasselt University, Hasselt, Belgium

[7]Center for Research and Innovation in Care, University of Antwerp, Antwerp, Belgium

Corresponding Author:
Paulien Vermunicht, PhD
Research Group Cardiovascular Diseases, University of Antwerp, Drie Eikenstraat 655, Antwerp, Belgium

Abstract

Background: Heart rate (HR) monitoring by wearable devices offers a physiological, personalized, and continuous method for assessing physical activity (PA) duration and intensity. However, methods to translate HR data into meaningful PA metrics are diverse and nonstandardized.

Objective: This study aims to provide an overview of how HR data are used to quantify PA behavior and estimate physiological outcomes in adult patients with cardiovascular disease (CVD).

Methods: A systematic search was performed in PubMed, Web of Science, and CENTRAL for studies published between 2014 and 2024. Eligible studies included adults with CVD or related risk factors wearing HR monitors to estimate PA. Data were synthesized narratively. The methodological quality of the included studies was evaluated using the Crowe Critical Appraisal Tool (CCAT; Michael Crowe).

Results: Twenty studies were included, spanning four HR-based PA estimation methods: (1) HR zone analysis (n=14), which assessed time spent in moderate-to-vigorous zones to evaluate guideline or training adherence; (2) physiological modeling (n=4), estimating outcomes such as energy expenditure (physical activity level) or cardiorespiratory fitness (maximal oxygen uptake); (3) change detection (n=1), using time-series and machine learning algorithms to quantify shifts in PA behavior; and (4) a derived personalized scoring system (n=1). While each approach demonstrated clinical promise of using HR data, external validation, and methodological transparency is often lacking.

Conclusions: HR-based PA estimation holds the promise of physiologically meaningful, personalized PA monitoring in CVD care. Modeling approaches and personalized scoring systems linking PA behavior to cardiovascular outcomes may provide highly needed clinical tools for PA management in patients. Research should prioritize algorithm transparency, clinical validation, and standardization.

(JMIR Mhealth Uhealth 2026;14:e79995) doi:10.2196/79995

KEYWORDS

exercise; heart rate; wearable electronic devices; fitness trackers; cardiac rehabilitation

Introduction

Engaging in sufficient physical activity (PA) is integral to the management of cardiovascular diseases (CVDs). The latest guidelines recommend ≥150 min/week of moderate-intensity PA, ≥75 min/week of vigorous-intensity PA, or an equivalent combination of both, typically expressed as the weekly accumulation of moderate-to-vigorous physical activity (MVPA), which can often be quantified using physiological indicators such as heart rate intensity zones [1]. PA is associated with reduction in cardiovascular and all-cause mortality, improvement in functional capacity and quality of life, and better control of key risk factors such as hypertension, dyslipidemia, and diabetes [1-4]. Despite the proven benefits of exercise-based cardiac rehabilitation (CR) in patients with CVD, with reductions of future cardiac events [5-7], its

XSL•FO
RenderX

long-term effectiveness is often limited by low adherence to recommended PA levels and inadequate monitoring of whether patients adhere to recommended PA levels, particularly when transitioning from supervised clinical settings (ie, phase 2 CR) to independent activities at home in phase 3 [8,9]. Nonadherence leads to a decline in cardiovascular function and undermines the potential long-term benefits of CR programs [10,11].

Continuous, accurate, and noninvasive monitoring of PA during patients' everyday activities could provide a solution to this challenge [12]. However, traditional PA monitoring methods such as self-reported questionnaires or activity logs are often subjective and prone to recall bias, and may not accurately reflect PA intensity [13,14]. Even objective methods like accelerometry primarily capture movement quantity (eg, steps and counts) and may approximate intensity (eg, via step cadence), but often do not adequately reflect physiological effort, making them less suited to fully assess PA intensity in clinical populations. In contrast, (cardiopulmonary) exercise testing or 6-minute walk tests offer objective insights but are limited to single time points and are impractical for frequent or real-world use [15,16].

Heart rate (HR) based PA estimation holds the promise of physiologically meaningful, personalized PA monitoring in CVD care. HR reflects the body's physiological response to activity intensity and is directly related to oxygen consumption and cardiac output [17]. Unlike movement-based metrics such as step count or distance walked, HR provides a direct measure of the body's metabolic demand and reaction to activity intensity, regardless of the type of exercise [18]. In line with clinical guidelines, PA intensity is often prescribed using target HR zones, especially in CR contexts [1,19]. This makes HR a more precise and individualized indicator of exertion and cardiorespiratory demand, and thus a uniquely valuable parameter for quantifying PA [19-21]. Advances in wearable technology have made near-continuous HR monitors, such as photoplethysmography-based wrist-worn devices and chest straps, accessible tools for continuous, noninvasive tracking of HR in everyday environments, provided they are properly validated.

However, translating HR data into meaningful, personalized metrics that guide patients and physicians toward adequate PA and optimal health outcomes remains a challenge. Although wearable devices can provide continuous HR measurements, there is a lack of standardized, transparent methods for interpreting this data in ways that are clinically actionable and tailored to patients' individual needs. While commercial devices such as those from Fitbit (Fitbit Inc) or Garmin (Garmin Ltd) often provide proprietary activity scores or readiness indicators, these metrics are typically based on "black box" algorithms that often rely more on accelerometer data than on HR data, and are rarely validated in clinical populations such as patients with CVD. Consequently, their clinical utility and safety in guiding PA remain uncertain.

The primary aim of this review is to explore how wearable HR monitors, including wrist-worn photoplethysmography sensors and electrocardiogram (ECG)-based chest straps across consumer-grade and research or medically validated devices, have been used to estimate and quantify PA with methodologies relevant to patients with CVD. Specifically, this review seeks to classify and describe the methodological approaches used to translate HR data into PA metrics and physiological estimates, including adherence rates and personalized scoring systems, across CR and broader CVD populations. In contrast to prior reviews that addressed wearable technologies in (tele) CR more broadly [22], this review focuses specifically on the methodological frameworks used to derive clinically meaningful PA indicators from HR data. By providing an overview of these methodologies, the review aims to offer insights into the availability and potential gaps in methods for HR-based PA monitoring in this population.

Methods

Overview

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 statement was followed in conducting this review (Checklist 1) [23]. This systematic review was not prospectively registered. A protocol was developed by the authors to guide the review process, but it was not published or made publicly available.

Search Strategy

A comprehensive search was conducted in PubMed, Web of Science, and the Cochrane Central Register of Controlled Trials (CENTRAL) on April 3, 2024 to identify relevant studies. The search strategy was developed based on the PICO (population, intervention, comparison, outcome) framework [24] and combined terms related to "CVD," "HR," "wearable HR monitors," and "PA" using Boolean operators. Search syntaxes were adapted for each database to account for differences in indexing and Boolean operator functionalities. Full search strings for each database are detailed in Table S1 in Multimedia Appendix 1. In addition to database searches, the reference lists of included articles were screened to identify additional relevant studies (backward citation search).

Eligibility Criteria

Studies were eligible for this review if they met the following inclusion criteria.

Population

Studies had to focus on adults aged 18 years or older with CVD. CVD was defined broadly and included individuals with coronary artery disease (CAD), myocardial infarction, acute coronary syndrome (ACS), heart failure with preserved or reduced ejection fraction (HFpEF/HFrEF), stroke, congenital heart disease, valvular diseases, and postrevascularization conditions (eg, percutaneous coronary intervention [PCI], coronary artery bypass grafting [CABG]). Studies conducted in patients with cardiovascular risk factors (eg, hypertension, diabetes, and metabolic syndrome) and in healthy populations were also considered if the authors explicitly stated that the HR-based PA estimation method described in the study was specifically intended or useful for application in cardiac populations.

XSL•FO
RenderX

HR-Based PA Estimation

Studies had to use wearable HR monitors (photoplethysmography-based wrist-worn devices or ECG-based chest straps) to estimate or quantify PA.

Study Types and Publication Criteria

Eligible studies included original research articles and study protocols. Protocols were included if they provided detailed methodological descriptions on how HR monitors or HR-based PA metrics would be used to assess PA, even if no results were reported. Including protocols allowed for the capture of methodological advancements relevant to the primary aim of the review. To ensure technological relevance, studies had to be published within the last 10 years (ie, between January 2014 and April 2024). Furthermore, only studies published in English with full-text availability were eligible.

Exclusion Criteria

Studies were excluded if HR monitors were solely used for exercise prescription or guidance, such as setting HR zones for training sessions, without quantifying or evaluating PA based on HR data. Studies were also excluded if they did not focus on a cardiac population or did not describe an HR-based PA estimation method applicable to cardiovascular health. Additionally, review articles, editorials, and conference abstracts were not eligible. Studies were also excluded if they were not published in English, were published more than 10 years ago, or lacked full-text availability.

Screening and Data Management

Duplicate records were identified and removed prior to screening, using a systematic deduplication method in EndNote, as described by Bramer et al [25] The screening process was conducted in multiple stages using Rayyan software (Rayyan Systems Inc.). Initially, 2 reviewers (PV and CB) independently screened article titles to exclude irrelevant studies. This was followed by a screening of abstracts. Articles deemed relevant based on abstract screening proceeded to full-text review, during which the reviewers assessed the complete articles to confirm their eligibility according to the predefined criteria. Studies where eligibility could not be determined from the abstract were included for full-text review. If discrepancies arose at any stage, they were resolved through discussion or, if necessary, a third reviewer (LD) was consulted to resolve them.

Data Extraction and Outcomes of Interest

A structured data extraction form was developed in Microsoft Excel and pilot-tested on 2 randomly selected studies to ensure consistent, clear, and comprehensive data collection. Two independent reviewers (PV and CB) extracted data from the included studies and any discrepancies were resolved through consensus discussions between the 2 reviewers (PV and CB); and if necessary, a third reviewer (LD) was consulted.

The primary outcome of interest in this review was the methodological approach used to translate wearable HR data into PA-related metrics or physiological estimates. Specifically, we extracted and classified how HR data were operationalized (eg, threshold-based HR zone analysis, regression-based energy expenditure models, cardiorespiratory fitness estimation, change detection algorithms, or personalized activity scores).

In addition, extracted data included study characteristics (eg, title, authors, publication year, country, funding, and design), participant demographics, HR monitor specifications, and details on the intervention or observation period (eg, training program and duration). We also extracted study-specific reported outcomes to provide contextual interpretation of each methodological approach. The reported outcome parameters varied across studies and included measurement validity indicators (eg, comparison with CPET or ECG, mean absolute [percentage] error, intraclass correlation coefficient, coefficient of determination), adherence rates, physiological outcomes (eg, maximal oxygen uptake [VO_2 max]), feasibility metrics, and safety events. Given the heterogeneity in devices, populations, and study designs, these outcomes were synthesized descriptively and were not pooled quantitatively.

Methodological Quality Assessment

The methodological quality of the included studies was evaluated using the Crowe Critical Appraisal Tool (CCAT; Michael Crowe) [26]. This tool was selected due to its flexibility and applicability across diverse study designs, making it particularly suitable for this review, which included randomized controlled trials, observational studies, feasibility studies, methodological studies, and protocol descriptions.

The CCAT assessed studies across 8 domains, including preliminaries, introduction, design, sampling, data collection, ethics, results, and discussion. Each domain is scored on a scale from 0 to 5, where 5 indicates that the criterion is fully met and 0 indicates that it is not addressed. Intermediate scores reflect partial fulfillment of the domain criteria according to the CCAT user guide [27]. Domain scores were assigned using the guidance provided in the CCAT user guide, where item descriptors support reviewer judgment when determining the final domain score [27]. The maximum total score is 40.

For protocol descriptions, only applicable domains of the CCAT were scored. Domains that were not applicable due to the absence of empirical results (eg, results and discussion) were marked as “not applicable”. The total score for protocols was recalculated as a percentage of the maximum applicable score (eg, if 6 domains were scored, the total score was calculated out of 30 instead of 40). This approach ensured that protocol descriptions were fairly assessed for their methodological relevance without being penalized for the absence of empirical findings.

Two independent reviewers (PV and CB) assessed methodological quality and resolved discrepancies through consensus discussion. CCAT scores did not influence study inclusion decisions and were used to contextualize interpretation rather than to weight or exclude findings within the narrative synthesis.

Although the CCAT is primarily designed as a methodological appraisal tool, several domains address aspects closely related to potential sources of bias (eg, study design, sampling strategy, data collection procedures, and discussion of bias and confounding). Therefore, CCAT scores provide an indirect

XSL•FO
RenderX

indication of methodological robustness and susceptibility to bias.

Data Synthesis

Given the heterogeneity of study designs and outcomes, a narrative synthesis was conducted. The primary aim of the synthesis was to identify and classify methodological approaches used to translate HR data into PA-related metrics.

During data extraction, the reported HR-based PA estimation methods from each study were examined and compared. Through an inductive process, recurring methodological approaches were identified and grouped into conceptual categories based on their underlying estimation strategy. This resulted in 4 distinct categories, explained in section "Heart Rate–Based Physical Activity Estimation Methods" of the Results. The categorization was discussed between the 2 reviewers (PV and CB) and finalized through consensus.

After categorization, studies within each methodological group were compared descriptively with respect to contextual characteristics such as HR monitor type, study population, and reported outcomes. Similarities and differences between studies were identified through iterative comparison across these dimensions. Methodological gaps were defined as areas where approaches lacked transparency, validation in cardiovascular populations, or consistent methodological reporting.

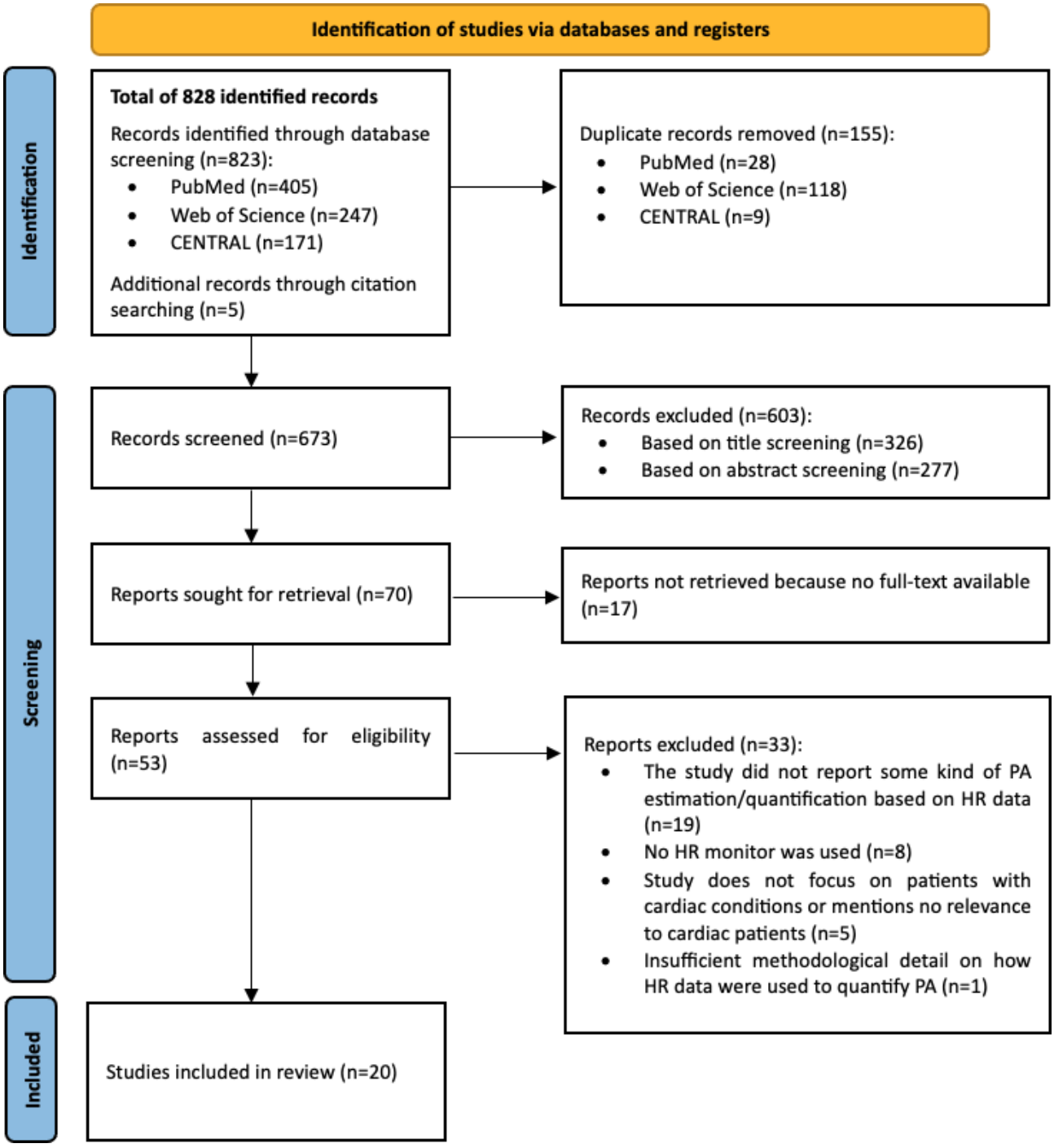
Results

Study Selection Process

The study selection process is summarized in a PRISMA 2020 flow diagram (Figure 1). The search for studies quantifying PA using HR monitors in patients with CVD yielded 828 results. After removal of duplicates, 673 publications were screened first on title and then on abstract. A total of 53 articles were screened on full text, 20 of which met the inclusion criteria.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the selection process. HR: heart rate; PA: physical activity.



General Study Characteristics

Of the 20 included studies, 15 [28-42] were research articles and 5 [43-47] were protocol papers (Table 1, left side). The study designs included 8 [28,29,38,43-47] randomized controlled trials (RCTs), 5 [30-33,42] nonrandomized interventional studies (4 feasibility studies and 1 pre-post study), and 7 [34-37,39-41] observational studies that monitored PA patterns using HR monitors without aiming to induce behavioral change. Studies were conducted across diverse geographical locations, primarily in Europe and the United Kingdom (n=12) [29,31,32,35,36,38-40,43,45-47], followed by North America (n=5) [28,34,37,41,44], Asia (n=2) [30,33], and Australia (n=1) [42]. Sample sizes varied widely, with feasibility studies including the fewest participants (range: 5 - 31), whereas other study designs included up to 300 participants.

XSL•FO
RenderX

Table . General study characteristics and population descriptions.

	General study characteristics						Study populations characteristics				
Study ID	First author (year)	Article type	Study design	Center type	Study country	Sample size (n)	Population	Gender distribution, n (%)	Age (years) [a]	BMI (kg/m²),[a]	Beta-Blocker use, n (%)
1	Alonso et al (2021) [28] *HEART Camp*	Research article	RCT[b]	Multicenter	United States	204	Patients with heart failure (HFrEF[c] and HFpEF[d])	• Female: 91 (45)[e] • Male: 113 (55)[e]	• HFpEF (intervention): 63.3 (9.4)[e] • HFpEF (control): 65.6 (9.3)[e] • HFrEF (intervention): 58.6 (13.3)[e] • HFrEF (control): 58.6 (10.1)[e]	• HFpEF (intervention): 35.6 (7.2)[e] • HFpEF (control): 36.4 (8.0)[e] • HFrEF (intervention): 34.7 (9.0)[e] • HFrEF (control): 33.9 (7.6)[e]	199 (98)
2	Crozier et al (2023) [43] *MOTIVATE-CR+*	Protocol paper	RCT (feasibility)	Multicenter	United Kingdom	72	Patients post myocardial infarction	—[f]	—	—	—
3	Mansfield et al (2017) [44] *PROPEL*	Protocol paper	RCT	Multicenter	Canada	192	Patients who have experienced stroke	—	—	—	—
4	Batalik et al (2020) [29]	Research article	RCT	Single-center	Czech Republic	51	Patients undergoing CR[g] (ACS[h], LVEF[i]>45%, and post heart revascularization)	• Female: 9 (18)[e] • Male: 42 (82)[e]	• Intervention: 56.5 (6.9)[e] • Control: 57.7 (7.6)[e]	• Intervention: 28.0 (3.5)[e] • Control: 28.3 (4.3)[e]	46 (90)
5	Chung et al (2019) [30]	Research article	Feasibility (interventional, non-RCT)	Single-center	South Korea	16	Healthy adults[j]	• Female: 11 (69)[e] • Male: 5 (31)[e]	• 26.3 (4.3)[e]	• Not disclosed	Not disclosed

XSL•FO
RenderX

	General study characteristics						Study populations characteristics				
Study ID	First author (year)	Article type	Study design	Center type	Study country	Sample size (n)	Population	Gender distribution, n (%)	Age (years) [a]	BMI (kg/m²),[a]	Beta-Blocker use, n (%)
6	Dobrican and Zampunieris (2016) [31]	Research article	Feasibility (interventional, non-RCT)	Single-center	Luxembourg	5	Healthy adults[j]	• Female: 0 (0)[e] • Male: 5 (100)[e]	• 21.0 - 48.0[k]	• Not disclosed	0 (0)
7	Dosbaba et al (2020) [45]	Protocol paper	RCT	Single-center	Czech Republic	76	Patients undergoing CR (ACS, LVEF>40%, post heart revascularization)	—	—	—	—
8	Quero et al (2017) [32]	Research article	Feasibility (interventional, non-RCT)	Single-center	Spain	Not explicitly disclosed[c]	Patients undergoing CR (CAD[l], low-risk)	• Not disclosed[m]	—	—	—
9	Suchy et al (2014) [46] *OptimEx-CLIN*	Protocol paper	RCT	Multi-center	Germany, Norway, Austria, Belgium	180	Patients with heart failure (HFpEF)	—	—	—	—

XSL•FO
RenderX

	General study characteristics						Study populations characteristics				
Study ID	First author (year)	Article type	Study design	Center type	Study country	Sample size (n)	Population	Gender distribution, n (%)	Age (years) [a]	BMI (kg/m²),[a]	Beta-Blocker use, n (%)
10	Xia et al (2023) [33]	Research article	Feasibility (interventional, non-RCT)	Single-center	China	31	Patients undergoing CR (CAD, ACS, and/or post heart revascularization)	• Female:1 (3)[e] • Male: 30 (97)[e]	• 56.2 (13.4)[e]	• 26.2 (2.9)[e]	20 (65)
11	Koffman et al (2023) [34]	Research article	Observational (prospective)	Single-center	United States	70	Patients who have experienced stroke	• Female: 32 (46)[e] • Male: 38 (54)[e]	• 61.0 (13.0)[e]	• Not disclosed	42 (60)
12	Schubert et al (2020) [35]	Research article	Observational (prospective)	Multi-center	United Kingdom, Germany	107 datasets from 84 patients	Valve disease patients (mitral or aortic)	• Female: 33 (39)[e] • Male: 51 (61)[e]	• 66.0 (IQR 56.0 - 75.0)[n]	• 26.8 (IQR 24.4 - 29.5)[n]	58 (54)
13	Sandberg et al (2016) [36]	Research article	Observational (prospective)	Single-center	Sweden	122	Adults with congenital heart disease	• Female: 48 (39)[e] • Male: 74 (61)[e]	• Simple lesions: 30.0 (IQR 20.7)[n] • Complex lesions: 133.7 (IQR 29.4)[n] • Controls: 31.4 (IQR 24.1)[n]	• Simple lesions: 25.1 (4.3)[e] • Complex lesions: 23.5 (4.0)[e] • Controls: 25.8 (5.3)[e]	18 (15)
14	Weeks et al (2018) [37]	Research article	Observational (prospective)	Single-center	USA	15	Patients undergoing inpatient rehabilitation (various conditions, including cardiac disorders and stroke)	• Female: 7 (47)[e] • Male: 8 (53)[e]	• 44.0 - 80.0[k]	• Not disclosed	4 (27)
15			RCT			300		—	—	—	—

XSL•FO
RenderX

	General study characteristics						Study populations characteristics				
Study ID	First author (year)	Article type	Study design	Center type	Study country	Sample size (n)	Population	Gender distribution, n (%)	Age (years) [a]	BMI (kg/m²),[a]	Beta-Blocker use, n (%)
	Brouwers et al (2017) [47] *Smart-Care-CAD*	Protocol paper		Single-center	Netherlands		Patients undergoing CR (CAD, ACS, and/or post heart revascularization)				

XSL•FO
RenderX

	General study characteristics						Study populations characteristics				
Study ID	First author (year)	Article type	Study design	Center type	Study country	Sample size (n)	Population	Gender distribution, n (%)	Age (years) [a]	BMI (kg/m²),[a]	Beta-Blocker use, n (%)
16	Kraal et al (2017) [38] *FIT@Home*	Research article	RCT	Single-center	Netherlands	90	Patients undergoing CR (ACS, and/or post heart revascularization)	• Female: 10 (11)[e] • Male: 80 (89)[e]	• 59.2 (8.5)[e]	• Intervention: 28.2 (3.9)[e] • Control: 27.8 (4.8)[e]	82 (91)
17	Kraal et al (2016) [39]	Research article	Observational (prospective)	Single-center	Netherlands	16	Patients undergoing CR (ACS, and/or post heart revascularization)	• Female: 0 (0)[e] • Male: 16 (100)[e]	• 55.8 (7.3)[e]	• Development: 28.6 (3.5)[e] • Validation: 30.7 (3.1)[e]	14 (88)
18	Rissanen et al (2022) [40] *Health-Beat*	Research article	Observational (prospective)	Single-center	Finland	74	Adults with cardiovascular risk factors	• Female: 56 (76)[e] • Male: 18 (24)[e]	• 55.8 (IQR 35.0 - 64.0)[n]	• 28.7 (IQR 21.9 - 39.9)[n]	0 (0)
19	Sprint et al (2017) [41]	Research article	Observational (prospective)	Single-center	United States	24	Patients undergoing inpatient rehabilitation (various conditions, including cardiac disorders and stroke)	• Female: 12 (50)[e] • Male: 12 (50)[e]	• Rehab group: 61.3 (12.7)[e] • Control group: 21.2 (2.2)[e]	• Not disclosed	4 (17)
20	Hannan et al (2021) [42]	Research article	Mixed methods pre-post study (interventional, non-RCT)	Single-center	Australia	18	Patients undergoing CR in maintenance phase (ACS, post heart revascularization and/or valvular surgery)	• Female: 3 (17)[e] • Male: 15 (83)[e]	• 56.0 (15.5)[e]	• 26.5 (4.4)[e]	3 (17)

[a]For each study, age/BMI was reported as originally provided by the authors. If age/BMI was available for the total study population, this value was reported. If only group-specific data were available, separate values are shown.

[b]RCT: randomized controlled trial.

XSL•FO
RenderX

[c]HFrEF: heart failure with a reduced ejection fraction.

[d]HFpEF: heart failure with a preserved ejection fraction.

[e]Mean (SD).

[f]Not applicable; applies to protocol papers where no results were reported.

[g]CR: cardiac rehabilitation.

[h]ACS: acute coronary syndrome (ie, myocardial infarction or unstable angina).

[i]LVEF: left ventricular ejection fraction.

[j]The study was included because of its focus on developing a system specifically for patients undergoing CR, despite the feasibility here being tested in healthy individuals.

[k]Range

[l]CAD: coronary artery disease.

[m]Case-based evaluation with a limited/nonspecified number of scenarios.

[n]Median (IQR)

Population Characteristics

An overview of the study populations is shown in Table 1 (right side). In terms of health characteristics, 8 studies [29,32,33,38,39,42,45,47] focused on patients eligible for CR following ACS, CAD, or revascularization procedures (eg, PCI or CABG). Other study populations included patients with heart failure (n=2) [28,46], post-stroke patients who have experienced a stroke (n=2) [34,44], individuals with cardiovascular risk factors (n=1) [40], congenital heart disease (n=1) [36], valve disease (n=1) [35], and patients postmyocardial infarction (n=1) [43]. Two studies [37,41] investigated inpatient rehabilitation patients with mixed conditions including cardiac disorders, while 2 feasibility studies [30,31] were conducted in healthy adults, assessing HR-based PA methods intended for populations with heart diseases.

CR studies had a predominantly male representation (82% - 100% [29,33,38,39,42]), whereas populations with heart failure (55% male [28]) and populations who have experienced strokes (54% male [34]) were more balanced. Age ranged from young healthy adults (mean 26.3, SD 4.3 years [30]) to older patients with heart failure (up to mean 65.6, SD 9.3 years [28]). Patients undergoing CR typically ranged between 55.8 (SD 7.3) years [39] and 59.2 (SD 8.5) years [38]. BMI values were generally between 26 and 31 kg/m² for CR populations [29,33,38,39,42], with the highest BMI observed in patients with heart failure (33.9 - 36.4 kg/m² [28]). Beta-blocker use was the highest among patients with heart failure (98% [28]) and ranged between 17% and 91% in CR populations [29,33,38,39,42].

HR Monitors

A total of 12 out of 20 studies [29-32,34,35,37,41-44,47] used a wrist-worn HR monitor (60%), 7 [33,36,38-40,45,46] used a chest-strap (35%), and 1 [28] did not specify the type (5%) (Table 2). Polar devices were the most commonly used (6 studies) [28,29,32,43,45,46], followed by Fitbit (4 studies) [34,37,41,44] and Garmin (2 studies) [38,39]. Other brands included Samsung, Philips, Firstbeat Technologies, and custom-designed wearables.

XSL•FO
RenderX

Table . Specifications of used heart rate monitors and heart rate-based physical activity estimation methods.

		HR monitor		HR[a]-based PA[b] estimation		Other PA estimation present?	
Study ID	First author (year)	Type	Brand/ Specification	Category (subcategory)	Details	Accelerometer-based	Questionnaire-based
1	Alonso et al. (2021) [28], HEART Camp	Unspecified	Polar (only the brand mentioned, not the type)	HR zone analysis (PA adherence)	• PA adherence is defined as: recording at least 120 minutes/week (ie, 80% of the recommended 150 minutes/week) of moderate intensity exercise (ie, 40% - 80% of HRR[c]) • HRR (HRmax – HRrest) determination: CPET[d]-based	No	Yes
2	Crozier et al. (2023) [43], MOTIVATE-CR+	Wrist-worn[e]	Polar Ignite 2[e]	HR zone analysis (PA adherence)	• PA adherence is defined as: recording at least 150 minutes/week of moderate intensity exercise (ie, 40% - 70% of HRR) • HRR (HRmax – HRrest) determination: not disclosed	Yes	Yes

XSL•FO
RenderX

		HR monitor		HR[a]-based PA[b] estimation		Other PA estimation present?	
Study ID	First author (year)	Type	Brand/ Specification	Category (subcategory)	Details	Accelerometer-based	Questionnaire-based
3	Mansfield et al. (2017) [44], PROPEL	Wrist-worn	Fitbit Charge	HR zone analysis (PA adherence)	• PA adherence is defined as: 1) recording at least 150 minutes/week of moderate intensity exercise (ie, 55% - 80% of HRmax) and/or (2) recording at least 150 "Fitbit active minutes"/week • HRmax determination: with age-dependent formula (specific formula not mentioned)	No	Yes
4	Batalik et al. (2020) [29]	Wrist-worn	Polar M430	HR zone analysis (training adherence)		No	No

XSL•FO
RenderX

Study ID	First author (year)	HR monitor		HR[a]-based PA[b] estimation		Other PA estimation present?	
		Type	Brand/ Specification	Category (subcategory)	Details	Accelerometer-based	Questionnaire-based

					• Adherence to a structured training program with predefined frequency, intensity, and duration (eg, 3 × 60-minutes/week at 70% - 80% of HRR):
					  • Intensity assessment: time spent in prescribed HR zone and average training intensity (HRR)
					  • Quantitative assessment: number of completed sessions
					• HRR (HRmax – HRrest) determination: CPET-based		

5	Chung et al. (2019) [30]	Wrist-worn	Custom-designed wearable device	HR zone analysis (training adherence)		No	No

XSL•FO
RenderX

Study ID	First author (year)	HR monitor		HR[a]-based PA[b] estimation		Other PA estimation present?	
		Type	Brand/ Specification	Category (subcategory)	Details	Accelerometer-based	Questionnaire-based

					• Adherence to a structured training program with predefined intensity (ie, 50% - 70% of HRmax):
					  • Intensity assessment: percentage of time within prescribed HR zone during exercise (referred to as 'target heart rate retention ratio')
					  • Quantitative assessment: number of completed sessions
					• HRmax determination: CPET-based		

6	Dobrican and Zampunieris (2016) [31]	Wrist-worn	Samsung Gear S2	HR zone analysis (training adherence)		No	No

XSL•FO
RenderX

Study ID	First author (year)	HR monitor		HR[a]-based PA[b] estimation		Other PA estimation present?	
		Type	Brand/ Specification	Category (subcategory)	Details	Accelerometer-based	Questionnaire-based

					• Adherence to a structured training program with predefined intensity (ie, 65% - 90% HRmax):
					  • Intensity assessment: time spent in prescribed HR zones
					  • Quantitative assessment: number of completed sessions
					• HRmax determination: with age-dependent formula (220 – age)		

7	Dosbaba et al. (2020) [45]	Chest-strap	Polar H10	HR zone analysis (training adherence)		No	No

XSL•FO
RenderX

Study ID	First author (year)	HR monitor		HR[a]-based PA[b] estimation		Other PA estimation present?	
		Type	Brand/ Specification	Category (subcategory)	Details	Accelerometer-based	Questionnaire-based

					• Adherence to a structured training program with predefined frequency, intensity, and duration (eg, 3×33-minutes/week at 85% - 95% of HRmax during intervals):
					  • Intensity assessment: time spent in prescribed HR zones
					  • Quantitative assessment: completion of ≥70% of prescribed sessions (regardless of HR)
					• HRmax determination: CPET-based		

8	Quero et al. (2017) [32]	Wrist-worn	Polar M600	HR zone analysis (training adherence)		No	No

XSL•FO
RenderX

Study ID	First author (year)	HR monitor		HR[a]-based PA[b] estimation		Other PA estimation present?	
		Type	Brand/ Specification	Category (subcategory)	Details	Accelerometer-based	Questionnaire-based

					• Adherence to a structured training program with predefined intensity and duration (ie, 15 - 40 min, target HR zones based on HRmax, HRrest, ventilatory thresholds):
					  • Intensity assessment: time spent in prescribed HR zones (classified via fuzzy logic algorithm)
					  • Quantitative assessment: number of completed sessions
					• Target HR zone determination: CPET-based		

9	Suchy et al. (2014) [46], OptimEx-CLIN	Chest-strap	Polar H7	HR zone analysis (training adherence)		Yes	No

XSL•FO
RenderX

Study ID	First author (year)	HR monitor		HR[a]-based PA[b] estimation		Other PA estimation present?	
		Type	Brand/ Specification	Category (subcategory)	Details	Accelerometer-based	Questionnaire-based

					• Adherence to a structured training program with predefined frequency, intensity, and duration (eg, 3 × 38-minutes/week at 90% - 95% of HRmax during intervals):
					  • Intensity assessment: time spent in prescribed HR zones
					  • Quantitative assessment: completion of ≥70% of prescribed sessions (regardless of HR)
					• HRmax determination: CPET-based		

10	Xia et al. (2023) [33]	Chest-strap	Recovery Plus Inc	HR zone analysis (training adherence)		No	No

XSL•FO
RenderX

Study ID	First author (year)	HR monitor		HR[a]-based PA[b] estimation		Other PA estimation present?	
		Type	Brand/ Specification	Category (subcategory)	Details	Accelerometer-based	Questionnaire-based

					• Adherence to a structured training program with predefined frequency, intensity, and duration (ie, 3 - 5 × 30-60-minutes/week, target HR=anaerobic threshold±5 - 10 bpm):
					  • Intensity assessment: time spent in prescribed HR zones
					  • Quantitative assessment: adherence defined as ≥10 minutes/day in target HR zone on ≥3 days/week
					• Anaerobic threshold determination: CPET-based		

11	Koffman et al. (2023) [34]	Wrist-worn	Fitbit Inspire 2	HR zone analysis (descriptive)		Yes	No

XSL•FO
RenderX

		HR monitor		HR[a]-based PA[b] estimation		Other PA estimation present?	
Study ID	First author (year)	Type	Brand/ Specification	Category (subcategory)	Details	Accelerometer-based	Questionnaire-based
					• Describe time in HR zone, without prescription of a structured training program: • Intensity assessment: time spent in PA categories determined based on high/low HR in combination with high/low step count • The threshold for high HR was HRrest + 20% HRR, as considered to be the threshold between sedentary and light activity intensity (additional: the high step count threshold was defined as the 25t percentile of the group steps per minute distribution among minutes with at least 1 step recorded (determined equally for all) • HRrest de-		

XSL•FO
RenderX

Study ID	First author (year)	HR monitor		HR[a]-based PA[b] estimation		Other PA estimation present?	
		Type	Brand/ Specification	Category (subcategory)	Details	Accelerometer-based	Questionnaire-based
					termination: mean HR between 2:00 AM and 4:00 AM • HRmax determination: with age-dependent formula (220 - age), also for beta-blocker users		
12	Schubert et al. (2020) [35]	Wrist-worn	Philips Health Watch (DL8791)	HR zone analysis (descriptive)	• Describe time in HR zone, without prescription of a structured training program: • Intensity assessment: time spent in moderate intensity zone (=50 - 88 bpm above HRrest) • HRrest determination: based on low-activity periods (ie, HR when maximum one step per minute and activity count accelerometer <475)	Yes	No
13	Sandberg et al. (2016) [36]	Chest-strap	Actiheart monitor (CamNTech Ltd, Cambridge, UK)	HR zone analysis (descriptive + PA adherence)		Yes	No

XSL•FO
RenderX

Study ID	First author (year)	HR monitor		HR[a]-based PA[b] estimation		Other PA estimation present?	
		Type	Brand/ Specification	Category (subcategory)	Details	Accelerometer-based	Questionnaire-based
					• Describe time in HR zone, without prescription of a structured training program: • Intensity assessment: time spent in moderate to vigorous zone, defined using two methods: (1) HR ≥1.75 * HRrest, and (2) HR ≥ HR value at 3 minutes into a step test (corresponding to 3 METs[f]) • PA adherence is defined as: recording at least 150 minutes/week of moderate to vigorous PA (HR ≥ HR value at 3 minutes into a step test), corresponding to		

XSL•FO
RenderX

		HR monitor		HR[a]-based PA[b] estimation		Other PA estimation present?	
Study ID	First author (year)	Type	Brand/ Specification	Category (subcategory)	Details	Accelerometer-based	Questionnaire-based
					• 21.4 minutes/day • HRrest determination: 30th lowest HR during monitoring period		
14	Weeks et al. (2018) [37]	Wrist-worn	Fitbit Charge	HR zone analysis (descriptive + PA adherence)		No	No

XSL•FO
RenderX

Study ID	First author (year)	HR monitor		HR[a]-based PA[b] estimation		Other PA estimation present?	
		Type	Brand/ Specification	Category (subcategory)	Details	Accelerometer-based	Questionnaire-based
					• Describe time in HR zone, without prescription of a structured training program: • Intensity assessment: time spent in target HR zone (ie, 55% - 80% of HRmax) and time spent in excessive HR zone (ie, >80% of HRmax) • PA adherence is defined as: recording at least 2×10-minutes in target HR zone (ie, 55% - 80% HRmax) on ≥3 days/week (American Heart Association / American Stroke Association (AHA/ASA[g]) guidelines) • HRmax determination: with age-dependent Tanaka formula (212 - 0.7× age), adjust-		

XSL•FO
RenderX

Study ID	First author (year)	HR monitor		HR[a]-based PA[b] estimation		Other PA estimation present?	
		Type	Brand/ Specification	Category (subcategory)	Details	Accelerometer-based	Questionnaire-based
					ed for beta-blocker users (Brawner formula: 164 - 0.7× age)		
15	Brouwers et al. (2017) [47], SmartCare-CAD	Wrist-worn	MIO Alpha	Physiological estimation using HR-based models (energy expenditure estimation – PAL)	Physical activity level (PAL) is calculated as total energy expenditure (TEE) divided by resting metabolic rate (RMR): • TEE[h] is estimated via multivariate regression using HR above rest (ie, HRnet), accelerometer output (activity counts per minute), weight, age, etc; exact coefficients are not reported • RMR[i] estimated using the Harris-Benedict equation (based on weight, height, age, and sex) • PAL thresholds: <3 (light), <6 (moderate), ≥6 (vigorous); typical daily PAL ranges from 1.2 (bed-bound) to >5 (elite athletes)	Yes	No
16	Kraal et al. (2017) [38], FIT@Home	Chest-strap	Garmin FR70			Yes	No

XSL•FO
RenderX

Study ID	First author (year)	HR monitor		HR[a]-based PA[b] estimation		Other PA estimation present?	
		Type	Brand/ Specification	Category (subcategory)	Details	Accelerometer-based	Questionnaire-based
				Physiological estimation using HR-based models (energy expenditure estimation – PAL)	Physical activity level (PAL) is calculated as total energy expenditure (TEE) divided by resting metabolic rate (RMR): • TEE is estimated via multivariate regression using HR above rest (ie, HRnet), accelerometer output (activity counts per min), weight, age, etc; exact coefficients are not reported • RMR estimated using the Harris-Benedict equation (based on weight, height, age, and sex) • PAL thresholds: <3 (light), <6 (moderate), ≥6 (vigorous); typical daily PAL ranges from 1.2 (bedbound) to >5 (elite athletes)		
17	Kraal et al. (2016) [39]	Chest-strap	Garmin (only the brand mentioned, not the type)	Physiological estimation using HR-based models (energy expenditure estimation – PAL)		Yes	No

XSL•FO
RenderX

		HR monitor		HR[a]-based PA[b] estimation		Other PA estimation present?	
Study ID	First author (year)	Type	Brand/ Specification	Category (subcategory)	Details	Accelerometer-based	Questionnaire-based
					Physical activity level (PAL) is calculated as total energy expenditure (TEE) divided by resting metabolic rate (RMR): • TEE is estimated via multivariate regression using HR above rest (ie, HRnet), accelerometer output (activity counts per min), weight, age, etc; exact coefficients are not reported • RMR estimated using the Harris-Benedict equation (based on weight, height, age, and sex) • PAL thresholds: <3 (light), <6 (moderate), ≥6 (vigorous); typical daily PAL ranges from 1.2 (bedbound) to >5 (elite athletes)		
18	Rissanen et al. (2022) [40], HealthBeat	Chest-strap	Bodyguard 2 (Firstbeat Technologies Oy)	Physiological estimation using HR-based models (VO_2max[j] estimation – Firstbeat)		Yes	No

XSL•FO
RenderX

		HR monitor		HR[a]-based PA[b] estimation		Other PA estimation present?	
Study ID	First author (year)	Type	Brand/ Specification	Category (subcategory)	Details	Accelerometer-based	Questionnaire-based
					• A proprietary algorithm, developed by Firstbeat Technologies Oy, estimates cardiorespiratory fitness levels (VO_2max in mL/kg/min) The full algorithm and exact details of the procedure are inaccessible, but the algorithm is based on integrating HR (to assess cardiac workload), heart rate variability (to reflect recovery and autonomic function), accelerometer data (to quantify motion and intensity), and individual characteristics (age, sex, weight, height)		

XSL•FO
RenderX

		HR monitor		HR[a]-based PA[b] estimation		Other PA estimation present?	
Study ID	First author (year)	Type	Brand/ Specification	Category (subcategory)	Details	Accelerometer-based	Questionnaire-based
19	Sprint et al. (2017) [41]	Wrist-worn	Fitbit Charge	PA change detection	Physical activity change detection (PACD) framework developed to detect changes in PA patterns using HR and step data: • Algorithms result in change score (CS[k]) quantifying the magnitude of detected PA changes between time windows • Strongest predictors of PA change included time in HR zone (ie, 55% - 80% of HRmax), total step count, and sedentary time reduction • HRmax determination: with age-dependent Tanaka formula (212 - 0.7× age), adjusted for beta-blocker users (Brawner formula: 164 - 0.7× age)	Yes	No
20	Hannan et al. (2021) [42]	Wrist-worn	Lynk2	Personalized activity intelligence (PAI) score calculation		No	No

XSL•FO
RenderX

		HR monitor		HR[a]-based PA[b] estimation		Other PA estimation present?	
Study ID	First author (year)	Type	Brand/ Specification	Category (subcategory)	Details	Accelerometer-based	Questionnaire-based
					Personal activity intelligence (PAI) is a HR–based metric calculated from exercise intensity and duration: • Based on a mathematical algorithm (available in supplementary material of other PAI-publication) • It uses HR responses (HRR calculations) over a 7-day rolling window; PAI scores are personalized using age, sex, HRrest, and HRmax • HRmax determination: with age-dependent formula (211 - 0.64× age), adjusted for beta-blocker or calcium channel blocker users (−15 bpm), or CPET-based when available • HRrest determination: not disclosed.		

[a]HR: heart rate.

[b]PA: physical activity.

[c]HRR: heart rate reserve, calculated as HRR = HRmax – HRrest.

[d]CPET: cardiopulmonary exercise testing.

[e]Only for intervention group – Polar Verity Sense arm strap for blinded measurements in the control group.

[f]MET: metabolic equivalent of task.

[g]AHA/ASA: American Heart Association / American Stroke Association.

XSL•FO
RenderX

[h]TEE: total energy expenditure.

[i]RMR: resting metabolic rate.

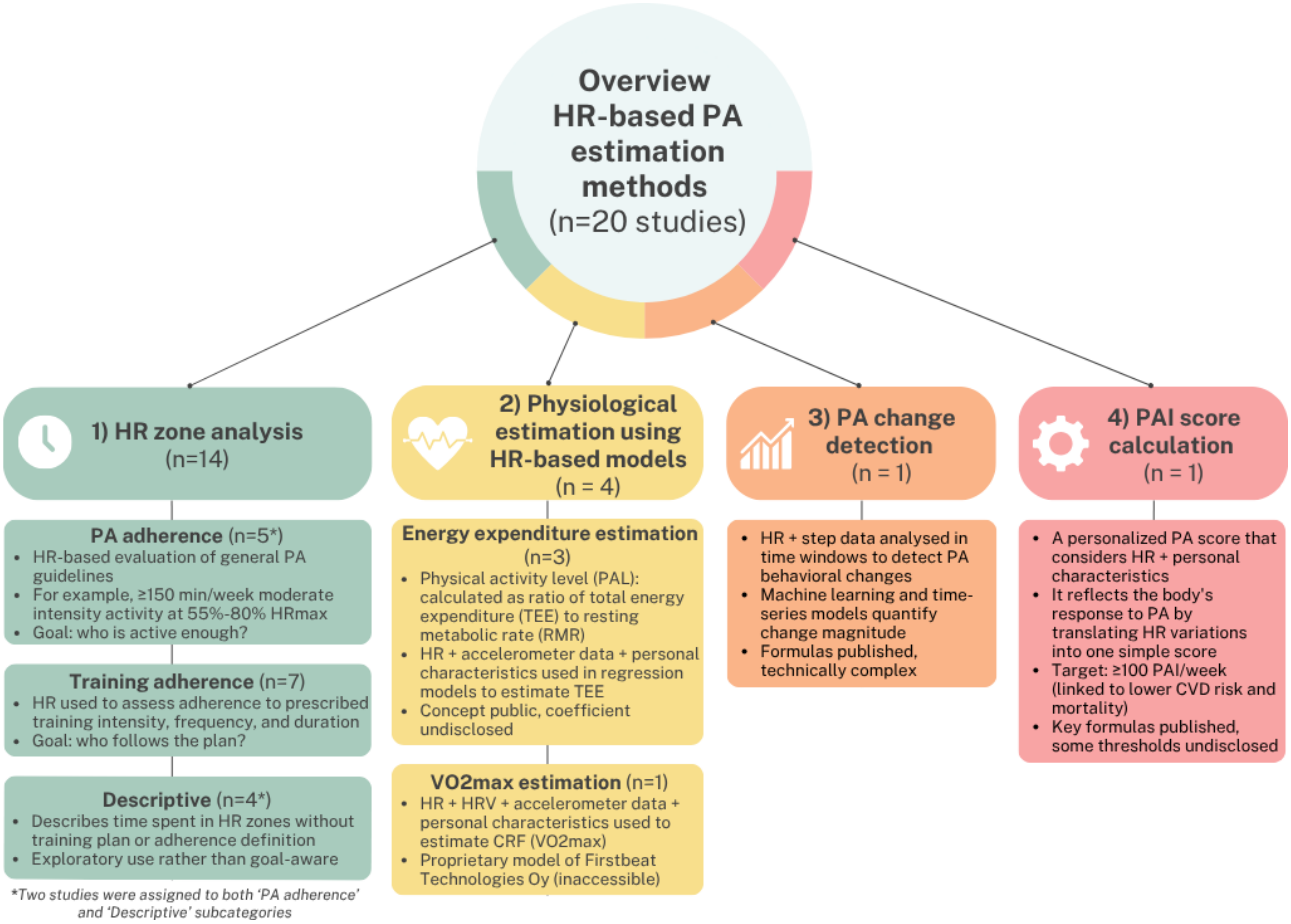
[j]VO_2max: maximal oxygen uptake.

[k]SC: change score.

HR-Based Physical Activity Estimation Methods

Four distinct approaches to HR-based PA estimation were identified: (1) HR zone analysis, (2) physiological estimation using HR-based models, (3) PA change detection, and (4) personalized activity scoring. These methods vary in complexity, ranging from simple threshold-based evaluations to algorithmic pattern detection. Each approach is summarized in Figure 2 and further elaborated in the subsections below, while detailed study specifications are presented in Table 2 (PA estimation approach) and in Table S2 in Multimedia Appendix 1.

Figure 2. Overview of heart rate-based physical activity estimation methods identified in this review. CRF: cardiorespiratory fitness; CVD: cardiovascular disease; HR: heart rate; HRmax: maximal heart rate; HRV: heart rate variability; PA: physical activity; PAI: personal activity intelligence; PAL: physical activity level; RMR: resting metabolic rate; TEE: total energy expenditure; VO_2max: maximal oxygen uptake.



HR Zone Analysis

Across 14 of the 20 included studies [28-37,43-46], the time spent in predefined HR zones was used to evaluate PA intensity and/or adherence, either in relation to general PA recommendations, structured training goals, or purely descriptively. These approaches relied on rule-based threshold classifications rather than analytical or statistical modeling, with HR values categorized into predefined HR-based intensity zones to quantify time spent at different activity levels.

HR zones were typically defined using heart rate reserve (HRR=HRmax [maximal heart rate]–HRrest) or percentages of HRmax. HRmax was determined using age-based formulas (n=4) [31,34,37,44], sometimes corrected for beta blocker use (n=1) [37], or measured via cardiopulmonary exercise testing (CPET; n=6) [28-30,32,45,46]. HRrest was estimated using self-established definitions such as the 30th lowest HR during monitoring, the average HR between 2:00 AM and 4:00 AM, or HR observed during low-activity periods (eg, HR when <1 step/minute combined with an accelerometer count <475). Most studies relied on transparent zone definitions, and most methods were reproducible. A subdivision was made for studies that applied HR zones as (1) behavioral benchmarks (eg, to assess adherence to PA guidelines), (2) criteria for training adherence, or (3) exploratory tools to describe PA intensity.

1. PA adherence: Around 5 [28,36,37,43,44] studies assessed whether participants met general PA recommendations based on HR-derived intensity thresholds. The internationally accepted target of ≥150 minutes per week of moderate-intensity activity was typically applied, or

XSL•FO
RenderX

translated into a daily equivalent (ie, ≥21.4 minutes/day). Moderate intensity was typically defined as 40% - 70% or 40% - 80% of HRR, or 55% - 80% of HRmax. One study defined the target threshold more pragmatically, using the HR at minute 3 of a step test as the intensity cut-off. In some cases, the HR-derived method of checking PA adherence was compared with assessments based on accelerometer-based indicators (n=2 [36,43]; eg, total accelerometer counts) and/or PA questionnaires (n=3 [28,43,44]; eg, report ≥150 minutes of moderate and/or vigorous intensity activity on a PA questionnaire, like the physical activity scale for individuals with physical disabilities [PASIPD]).

2. Training adherence: A total of 7 [29-33,45,46] studies evaluated adherence to a structured training program using HR to monitor both exercise intensity and completion. These programs defined frequency, duration, and intensity in advance (eg, 3×/week, 30 - 60 min/session, 70% - 80% HRR). Qualitative adherence was assessed based on intensity, including time spent in the prescribed HR zone, average session intensity (eg, 74.8% of HRR), or retention ratios (eg, percentage of total exercise time in the zone). Quantitative adherence referred to session completion, often defined as attending ≥70% of prescribed sessions, or meeting minimum thresholds such as ≥10 minutes/day in the target zone on ≥3 days/week.
3. Descriptive: In 4 [34-37] studies, HR zones were used without a behavioral target or structured plan. These studies described the time spent in various HR zones to characterize cardiovascular load or exercise intensity without defining adherence. The aim was typically exploratory, offering insight into physiological patterns of activity.

Feedback mechanisms: Among the 14 studies using HR zone analysis, feedback strategies varied in intensity and purpose. Four [30-33] training adherence studies delivered real-time feedback via wearable devices (eg, LED color changes or vibrations) when HR deviated from target zones. In at least 6 [28,29,43-46] studies, HR data were reviewed by health professionals via dashboards or apps (eg, Polar Flow; Polar Electro), with individual (telephone-based) or group-based counseling provided with a focus on knowledge, attitudes, self-efficacy, and self-management skills, and often informed by behavior change and motivational interviewing techniques. The remaining 4 [34-37] studies used HR monitoring purely for retrospective analysis, without participant-facing feedback or behavioral guidance.

Physiological Estimation Using HR-Based Models

A total of 4 studies [38-40,47] used HR data to estimate derived physiological PA-related outcomes rather than to assess PA behavior and adherence. These outcomes included energy expenditure, expressed as physical activity level (PAL), and cardiorespiratory fitness (CRF), typically estimated as VO_2max. In both cases, HR was used as a physiological input parameter, often in combination with accelerometer data and individual characteristics to inform regression or algorithmic models.

1. Energy expenditure–PAL: A total of 3 [38,39,47] studies estimated daily energy expenditure by calculating PAL, defined as the ratio of total energy expenditure (TEE) to resting metabolic rate. TEE was estimated using multivariate regression models that included HR above rest, accelerometer-derived activity counts per minute, and participant characteristics such as age and weight. The resting metabolic rate was estimated using the Harris-Benedict equation, based on sex, weight, height, and age. While none of the studies disclosed the exact regression coefficients or intercepts used to model TEE, the overall structure of the approach was well described and reproducible in principle. PAL was used as a continuous indicator of daily activity level, and typical values range from 1.2 (bed-bound) to above 5.0 (elite athletes). In these studies, no behavioral adherence thresholds or real-time feedback mechanisms were applied to PAL scores, though retrospective comparisons were made between groups (eg, cardiac telerehabilitation vs center-based programs).
2. VO_2max estimation–firstbeat: One study by Rissanen et al [40] estimated CRF by calculating VO_2max using a wearable device during a 30-minute self-paced walk. The algorithm integrated HR, heart rate variability, and movement data, alongside participant characteristics (eg, age, sex, height, weight, and HRmax), to model aerobic capacity. The HRmax determination was not disclosed. The estimation was validated against CPET-derived VO_2max values. The model showed strong agreement, with a mean absolute error of 3.1 mL/kg/min, mean absolute percentage error of 10.4%, and intraclass correlation coefficient of 0.88. The study acknowledged reduced accuracy in subgroups such as patients with type 2 diabetes. Developed by Firstbeat Technologies Oy, the algorithm is proprietary, with only conceptual elements publicly available. It remains inaccessible and is commercially protected, limiting reproducibility and external validation.

Physical Activity Change Detection

The study by Sprint et al [41] introduced the Physical Activity Change Detection (PACD) framework, an analytical method designed to identify significant changes in PA behavior over time using HR and step count data. PACD applied time-series analysis to detect meaningful fluctuations in PA intensity, frequency, and duration.

The dataset was segmented into consecutive time windows (eg, per week). Pairs of these windows were compared using three algorithmic methods: (1) small window permutation-based change detection (sw-PCAR), which uses Kullback-Leibler divergence to measure differences in activity patterns between two time windows, with significance assessed through a permutation test. (2) distance-based dissimilarity scorer (DDS), which computes the average Euclidean distance between summary statistics (eg, steps/day, time in HR zones), with distances above 0.25 interpreted as meaningful deviations. (3) Virtual classifier (VC), a machine learning method (decision trees) trained to distinguish activity patterns between windows, using classifier accuracy as a measure of behavioral change. Key predictors of PA change were fewer bouts in the target HR zone (55% - 80% HRmax), reduced sedentary time (<5 steps/minute), and lower daily step counts.

XSL•FO
RenderX

Each algorithm produced a change score reflecting the magnitude of PA change between time windows. In this study, PACD was applied retrospectively to characterize PA changes during inpatient rehabilitation, without being used for real-time clinical decision-making. Although no open-source code was provided, the authors detailed the full methodology, including formulas and statistical techniques, enabling potential reproduction of the analysis by researchers with appropriate technical expertise.

Personal Activity Intelligence (PAI) Score Calculation

One study by Hannan et al [42] used HR data to calculate the Personal Activity Intelligence (PAI) score, a metric that transforms HR data from a wearable device into a personalized, feedback score for long-term health maintenance. The PAI algorithm integrates HR intensity and duration over a rolling 7-day period and accounts for individual characteristics such as age, sex, HRrest, and HRmax [48]. The goal is to provide a single number reflecting whether an individual maintains sufficient cardiovascular stimulus to reduce morbidity and mortality risk [49]. A weekly score of ≥100 PAI is considered optimal for health benefits based on prior retrospective longitudinal cohort studies [50,51].

In the study of Hannan et al [42], PAI was used in a 2-phase design. Participants initially wore a HR monitor in a blinded phase (without receiving feedback), followed by an unblinded phase where they had access to their PAI score and were encouraged to increase it. HRmax, needed for the PAI calculation, was estimated using an age-dependent formula ($211 - 0.64 \times age$), with correction for beta- or calcium-channel blocker use (–15 bpm), or derived via CPET when available. The authors reported that 89% of participants improved their total PAI score following access to real-time feedback. The proportion of individuals achieving ≥50 PAI per week increased from 39% to 61% across phases [42].

Key components and equations behind the PAI algorithm, originally developed by the Cardiac Exercise Research Group (CERG) at the Norwegian University of Science and Technology (NTNU), have been published in supplementary material of earlier PAI-studies [18], although these were not eligible for inclusion in this review due to their retrospective, questionnaire-based methodology and noncardiac populations (see Discussion). The model is therefore largely reproducible for researchers with sufficient technical expertise, though some implementation details (eg, thresholds to begin earning PAI) remain undisclosed.

Methodological Quality Assessment

The CCAT quality ratings ranged from 50% to 87%, with a mean of 71% (Table 3). Four studies scored above 75%, indicating high methodological quality, while 2 [26,27] studies scored below 60%, reflecting some limitations. The remaining 14 studies fell within the moderate range of 60%-75%. Most studies performed well in the domains of introduction, data collection, and discussion, with the majority scoring between 4 and 5 in these areas and no study scoring below 3. Fifteen studies [28-30,32-37,39-43,47] reached a moderate score of 3 in the design domain. Sampling and ethics were the weakest domains; 5 studies [30-32,39,45] scored below 3 for sampling, and 3 studies [30-32] scored below 3 for ethics, highlighting issues with sample representativeness and ethical reporting.

XSL•FO
RenderX

Table . Quality rating using CCAT[a]. The methodological quality of the included studies was assessed using the CCAT [26,27], which evaluates eight domains: Preliminaries, Introduction, Design, Sampling, Data collection, Ethics, Results, and Discussion. Each domain is scored from 0 to 5, with a maximum total score of 40. For protocol descriptions, only applicable domains were scored. Domains not applicable due to the absence of empirical data (eg, Results and Discussion) were marked as not Applicable (N/A). The total score for protocols was recalculated as a percentage of the maximum applicable score (eg, if 6 domains were scored, the total score was calculated out of 30).

ID	Study	CCAT[a] rating per domain									
		1	2	3	4	5	6	7	8		
		Preliminaries (0 - 5)	Introduction (0 - 5)	Design (0 - 5)	Sampling (0 - 5)	Data collection (0 - 5)	Ethics (0 - 5)	Results (0 - 5)	Discussion (0 - 5)	Total score	Total percentage (%)
1	Alonso et al (2021) [28] *HEART Camp*	3	4	3	3	3	5	4	4	29/40	73
2	Crozier et al (2023) [43] *MOTIVATE-CR+*	4	4	3	3	3	4	—[b]	—	21/30	70
3	Mansfield et al (2017) [44] *PROPEL*	4	5	4	4	4	5	—	—	26/30	87
4	Batalik et al. (2020) [29]	4	4	3	3	4	4	3	4	29/40	73
5	Chung et al (2019) [30]	2	4	3	2	4	2	3	3	23/40	58
6	Dobrican and Zampunieris (2016) [31]	3	3	2	2	3	2	2	3	20/40	50
7	Dosbaba et al (2020) [45]	3	3	4	2	3	4	—	—	19/30	63
8	Quero et al (2017) [32]	3	4	3	2	4	2	3	3	24/40	60
9	Suchy et al (2014) [46] *OptimEx-CLIN*	4	4	4	4	3	4	—	—	23/30	77
10	Xia et al (2023) [33]	4	4	3	3	4	3	4	4	29/40	73
11	Koffman et al (2023) [34]	4	5	3	4	4	4	4	4	32/40	80

XSL•FO
RenderX

ID	Study	CCAT[a] rating per domain									
		1 Preliminaries (0 - 5)	2 Introduction (0 - 5)	3 Design (0 - 5)	4 Sampling (0 - 5)	5 Data collection (0 - 5)	6 Ethics (0 - 5)	7 Results (0 - 5)	8 Discussion (0 - 5)	Total score	Total percentage (%)
12	Schubert et al (2020) [35]	3	3	3	3	4	4	4	4	28/40	70
13	Sandberg et al (2016) [36]	3	4	3	4	4	4	4	3	29/40	73
14	Weeks et al (2018) [37]	4	4	3	3	4	4	3	4	29/40	73
15	Brouwers et al (2017) [47] *Smart-Care-CAD*	3	4	3	3	3	4	—	—	20/30	67
16	Kraal et al (2017) [38] *FIT@Home*	4	5	4	3	3	4	4	4	31/40	78
17	Kraal et al (2016) [39]	4	4	3	2	4	4	4	4	29/40	73
18	Rissanen et al. (2022) [40] *Health-Beat*	4	4	3	3	3	4	4	4	29/40	73
19	Sprint et al. (2017) [41]	4	4	3	3	4	4	3	4	29/40	73
20	Hannan et al. (2021) [42]	4	4	3	3	3	4	4	4	29/40	73

[a]CCAT: Crowe Critical Appraisal Tool.

[b]not applicable

The methodological quality assessment was used to contextualize the interpretation of findings but did not determine study inclusion. Importantly, studies with lower scores were distributed across the different methodological categories identified in the narrative synthesis, and the methodological approaches used to derive HR-based PA metrics were largely consistent across studies irrespective of CCAT score. Overall, the quality of the included studies was moderate to high.

Discussion

Principal Findings

This review provides the first structured overview of HR-based methodologies for PA estimation in patients with CVD, highlighting their clinical utility, strengths, and the implementation gaps. We identified 4 distinct estimation approaches: HR zone analysis, physiological estimation using HR-based models, PA change detection, and personal activity intelligence score calculation. Most studies (14/20) used HR zone thresholds to evaluate PA intensity or adherence, often within CR contexts. Other methods incorporated algorithmic

XSL•FO
RenderX

modeling or machine learning to derive metrics such as energy expenditure, CRF, or behavioral change. While these approaches varied in complexity and purpose, each demonstrated the potential of HR data to provide physiologically meaningful insight into daily PA patterns.

HR Zone Analysis and Guideline-Based Adherence

HR zone analysis emerged as the most commonly applied and straightforward method of HR-based PA estimation in this review. It is straightforward because it allows for a direct translation of international PA guidelines into quantifiable metrics, specifically, the time spent in moderate-to-vigorous intensity HR zones, and it aligns with the widely adopted the frequency, intensity, time, and type framework [1,4]. As such, it provides clinicians and researchers with an accessible method to both prescribe and evaluate PA intensity in line with global recommendations.

Both the World Health Organization (WHO) and the European Society of Cardiology (ESC) advise that adults should engage in at least 150 minutes of moderate or 75 minutes of vigorous intensity aerobic activity per week, or an equivalent combination of both, with intensity defined via several physiological indicators, including %VO_2max, %HRmax, %HRR, or ratings of perceived exertion [1,4]. Among the reviewed studies, %HRmax and %HRR were used most often to define intensity zones. ESC recommends that maximal HR values should ideally be based on exercise testing rather than estimated from age-predicted formulas, especially in clinical populations and in patients on rate-limiting medications such as beta-blockers [1]. Despite this, most studies in our review relied on simple prediction equations for HRmax, which may result in less precise zone definitions. Notably, none of the included studies explicitly reported whether vigorous-intensity minutes were counted double when evaluating weekly MVPA totals. While international guidelines do allow for this (eg, 75 min of vigorous activity or an equivalent combination with moderate activity) [1,4], the absence of clear reporting may limit interpretability and cross-study comparability. Including this distinction in future reporting could help ensure more accurate and standardized assessments of adherence.

HRrest, another component needed for calculating HRR, was often defined through study-specific approaches, such as averaging nighttime HR values or identifying low-HR periods during inactivity. While literature suggests that early morning supine HR, measured immediately after waking but before getting out of bed, may offer the most reliable estimate, there is no standardized or transparent method for deriving HRrest from wearable data at the time of this writing [52]. Most commercial devices provide HRrest estimates through proprietary algorithms, leaving researchers with limited insight into how these values are calculated.

In terms of adherence assessment, some studies focused on guideline-based targets (eg, ≥150 min/week of MVPA), while others evaluated compliance with structured training programs based on the frequency, intensity, time, and type principles. Importantly, HR monitoring enabled a distinction between session attendance (frequency) and session quality (intensity and duration), offering a more nuanced understanding of adherence by distinguishing between mere participation and actual achievement of intensity and duration targets [53].

Finally, several studies integrated HR zone data into feedback mechanisms, either in real time (eg, via smartwatches with LED zone indicators) or post-session through digital dashboards and clinician-guided counseling. Effective feedback has been shown to be a key driver of behavior change in PA interventions [54]. When combined with motivational strategies such as self-monitoring and goal setting, these feedback loops may support greater behavioral change and patient engagement toward sustained PA adherence [43,44,54].

HR Versus Accelerometer-Based PA Monitoring

Several studies in this review combined HR monitoring with accelerometer-based methods to quantify PA adherence and estimate physiological parameters such as energy expenditure and VO_2max. These multimodal approaches aimed to improve accuracy and/or cross-validate findings by combining the complementary strengths of each method [55]. HR monitoring provides a direct physiological measure of exercise intensity and cardiovascular load, making it a more personalized and clinically relevant indicator of effort than mechanical metrics such as step count or accelerometer counts. Moreover, the physical demand associated with a given accelerometer value can vary substantially between individuals depending on their fitness, age, or health status [56]. For example, walking at the same pace may represent light intensity for one person and moderate intensity for another [36].

Despite these limitations, accelerometer and step count data could play a valuable role in assessing sedentary time, an increasingly recognized factor in cardiovascular health [57]. Sedentary behavior, defined as waking behavior in a sitting, reclining, or lying posture with low energy expenditure, has been associated with cardiometabolic risk, even in individuals who meet weekly MVPA guidelines [58,59]. This risk profile may reflect distinct physiological pathways, underscoring the importance of addressing sedentary behavior alongside PA adherence [60]. In the PACD study included in this review, a reduction in sedentary time (defined as time with fewer than 5 steps per minute) emerged as a key predictor of meaningful PA behavior change [41]. When paired with HR data to reflect oxygen demand and PA intensity, step-based indicators of inactivity may offer a low-burden, scalable method for tracking both sedentary time and physiological load. However, it is important to assess PA and SB as independent behavioral domains, each with distinct health effects and potentially requiring their own intervention strategies [60].

From Activity Monitoring to Physiological Outcomes

Although most studies focused on measuring adherence to PA guidelines (eg, time in zones), a smaller subset advanced toward estimating physiological outcomes from HR data, such as energy expenditure (PAL) or cardiorespiratory fitness (VO_2max). This transition from behavioral metrics to physiological modeling represents a valuable next step, especially in clinical populations where improving aerobic capacity and reducing cardiovascular risk are primary goals.

XSL•FO
RenderX

For example, three studies estimated PAL, defined as the ratio of total energy expenditure to resting metabolic rate. These models typically combined HR above rest with accelerometer data and participant characteristics [38,39,47]. PAL is widely used in nutrition science to estimate daily energy needs, but in cardiovascular care its clinical application remains limited [61]. As a descriptive metric, it lacks behavioral thresholds, is rarely linked to feedback mechanisms, and often relies on regression models with undisclosed coefficients, limiting transparency and reproducibility in research and practice. A more advanced example is the proprietary Firstbeat algorithm, which estimated VO_2max using HR, HR variability, movement data, and participant characteristics during a 30-minute walk [40]. While generalizability to cardiac populations remains uncertain and requires further validation, the Firstbeat algorithm showed acceptable agreement with CPET-derived values in athletic cohorts or adults with cardiovascular risk factors and is now embedded in commercial wearables from Garmin, following their acquisition of Firstbeat Technologies Oy in 2020 [40,62,63]. While this integration supports scalability, the proprietary nature of the algorithm limits its adaptability and full validation in diverse patient groups.

Together, these examples illustrate a growing shift from pure activity tracking toward predictive modeling of physiological capacity. By combining HR data with individual and contextual factors, such models could move beyond surface-level monitoring to support personalized CR planning and long-term health surveillance. Such approaches may help clinicians and researchers estimate meaningful outcomes, such as changes in cardiorespiratory fitness, based on daily-life activity patterns, ultimately linking daily PA behavior to cardiovascular impact.

PAI: Validation, Challenges, and Future Directions

The PAI system stands out within this review as the only approach that translates continuous HR data into a personalized, interpretable single-number feedback metric intended to motivate behavioral change over time. PAI reflects accumulated cardiovascular effort across a rolling 7-day window. In the one included PAI study in this review, short-term improvements in PAI accumulation were observed after CR participants gained access to their PAI score via a smartphone application, although evidence remains exploratory [42].

As a caveat, the commonly cited ≥100 PAI/week threshold, proposed to reduce the risk of CVD and mortality, originates from retrospective analyses of large population studies (eg, the HUNT Study in Norway) [48,51,64-68]. During the screening process of this review, eight such validation studies were identified but excluded, as none used HR monitors to calculate PAI [48,51,64-68]. Instead, PAI scores were estimated post hoc from self-reported PA questionnaires, relying on participants' recall of activity frequency, intensity, and duration [48,51,64-68]. This introduces substantial subjectivity and deviates from the intended physiological, HR-based calculation of PAI, potentially reducing both accuracy and internal validity.

Moreover, 7 of the 8 validation studies were conducted in healthy adult populations, frequently excluding participants with cardiovascular conditions. The reported mean ages were relatively young, ranging from 42.6 [64] to 55.1 [68] years. One study included participants with self-reported angina pectoris, myocardial infarction, or stroke, with a higher mean [SD] age of 67.6 [10.3] years[66]. However, activity in that group was still assessed via questionnaires, and the population was not clinically verified. To date, no validation studies have assessed the accuracy or clinical utility of PAI when calculated from real HR data in older adults or patients with CVD, populations for whom accumulating ≥100 PAI points per week may be less achievable but important. Also, the PAI algorithm rewards sustained vigorous activity more than moderate or lower-intensity efforts, limiting its applicability for clinical populations where activities such as walking or cycling are more prevalent [64]. A summary table of the characteristics of the 8 excluded PAI validation studies is available in Table S3 in Multimedia Appendix 1.

Although the algorithm is largely reproducible based on published formulas, certain implementation details remain proprietary [18]. Researchers with technical expertise can replicate the model in principle, but public tools for doing so are limited. The MIA Health app, currently the only available platform for real-time PAI calculation, expanded its availability in January 2025 from a few Nordic countries to most of Europe, North America, Australia, New Zealand, Taiwan, and Brazil, potentially increasing both clinical and consumer access [69].

Recently, a follow-up metric known as the activity quotient (AQ) was introduced by the same research group as a refinement of the PAI. Some retrospective studies have validated AQ using the same questionnaire-based methodology as PAI and report stronger associations with VO_2max and mortality [70]. However, none of these studies used HR monitors, and prospective AQ studies with wearable devices are still forthcoming [69]. In parallel to these developments, our research group is developing the Antwerp Activity Index and smart Antwerp Activity Index models [71,72] (P Vermunicht, PhD, unpublished data, 2026). These tools aim to provide HR-based, personalized activity scores by integrating daily-life HR data with individual characteristics, and are designed to predict cardiorespiratory fitness changes over time. While preliminary development has been promising, full clinical validation and prospective outcome studies are currently underway. Such emerging models may further advance the translation of HR-derived activity monitoring into actionable, personalized health guidance in cardiac populations.

Together, these developments reflect a growing interest in HR-based PA quantification and behavioral feedback. However, their clinical relevance for cardiac and older adults will depend on prospective validation in target populations, improved algorithmic transparency, and appropriate feedback tailoring for diverse activity profiles.

Future Directions and Research Limitations

The reviewed studies illustrate a diverse range of HR-based methods, from guideline-driven time-in-zone metrics to more advanced models estimating physiological outcomes and personal activity scores. This variety reflects a growing interest in using HR data not only to monitor activity but also to estimate its cardiovascular implications. Future research should prioritize validation of these methods in real-world populations with CVD,

XSL•FO
RenderX

with attention to feasibility, clinical interpretability, and predictive value. Tools that go beyond simple time-in-zone tracking and instead estimate outcomes like energy expenditure or changes in cardiorespiratory fitness may offer greater clinical value. As illustrated by systems such as PAI or the Firstbeat VO_2max model, combining HR data with individual characteristics can generate personalized feedback with the potential to guide long-term health behavior. Future work should strive for a balance between algorithmic sophistication, transparency, and accessibility.

Although not a focus of this review, a key translational consideration for HR-based PA estimation is that the validity of any derived PA metric depends on the accuracy of the underlying HR data [73,74]. Although ECG-based chest straps provide highly accurate HR measurements, they are less practical for daily monitoring. Wrist-worn photoplethysmography devices are more convenient, but their accuracy can be affected by artifacts and they should be validated against an ECG before clinical use [75,76]. Recent work from our group has therefore focused on artifact detection and cleaning procedures that operate directly on device-generated photoplethysmography HR data from commercial wearables, aiming to remove unreliable data while preserving clinically meaningful activity-related HR fluctuations [71,76-78]. Such artifact removal strategies are highly relevant when implementing the HR-based PA quantification methods described in this review, particularly in real-world cardiac populations where pathophysiological factors such as peripheral edema, poor tissue perfusion, and a higher incidence of arrhythmias may affect HR reliability [76,79].

Beyond methodological development, successful implementation of HR-based PA estimation also depends on how these metrics are embedded within broader digital care strategies. Mobile health tools that support patient education, self-management, and structured follow-up may enhance engagement and facilitate behavior change, thereby increasing the clinical relevance of HR-derived activity metrics [80,81].

This review has several limitations. It was not prospectively registered in a public database (eg, PROSPERO [International Prospective Register of Systematic Reviews]), which may limit transparency regarding protocol deviations. Although a protocol was developed to guide the review process, it was not publicly available. As a narrative synthesis, it does not provide pooled effect sizes or meta-analytic comparisons. The included studies varied widely in design, sample size, and how HR-based metrics were operationalized. While this limits direct comparability, the primary aim of this review was to map the diversity of methodological approaches rather than pool results. In addition, studies using techniques such as PACD or personalized scoring systems remain scarce, restricting broader generalization. Continued development and validation of such tools, especially in older and cardiac populations, will be essential to support their integration into clinical practice. Finally, due to the rapid evolution of wearable technologies and increasing research interest in HR-based activity estimation, new studies may have been published after the final search date. For instance, an eligible outcome paper from the OptimEx-Clin trial was published shortly after the search period [82]. While this study was not included for methodological consistency, its design was captured through the previously published protocol paper, which is part of this review. Future updates of this review could consider incorporating such newly available outcome data.

82

Conclusions

This review provides the first structured synthesis of HR-based methodologies for PA estimation in patients with CVD. Most studies relied on HR zone analysis to evaluate adherence to international PA guidelines, while others explored more advanced modeling approaches to estimate physiological outcomes such as energy expenditure or cardiorespiratory fitness.

Our findings underscore that HR monitoring offers a physiologically meaningful and scalable way to estimate PA, especially in clinical populations where traditional step-based or questionnaire-based methods may fall short. Personalized HR-based feedback systems may offer intuitive activity targets for patients. However, variability in how HR metrics are defined, validated, interpreted, and applied, currently limits comparability and implementation. Future research should focus on validating HR-based methods in real-world populations with CVD, improving methodological transparency, and tailoring feedback to individual needs. Ensuring reliable HR measurements remains a prerequisite for meaningful HR-based PA estimation in clinical practice. As wearable technologies continue to advance, HR-derived metrics hold strong potential to support more personalized and effective PA guidance in both clinical care and daily life.

Funding

The authors declared no financial support was received for this work.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

Supporting information.
[PDF File, 381 KB - mhealth_v14i1e79995_app1.pdf]

Checklist 1
PRISMA 2020 checklist.
[PDF File, 156 KB - mhealth_v14i1e79995_app2.pdf]

References

1. Pelliccia A, Sharma S, Gati S, et al. 2020 ESC Guidelines on sports cardiology and exercise in patients with cardiovascular disease. Eur Heart J 2021 Jan 1;42(1):17-96. [doi: 10.1093/eurheartj/ehaa605] [Medline: 32860412]
2. Cleven L, Krell-Roesch J, Nigg CR, Woll A. The association between physical activity with incident obesity, coronary heart disease, diabetes and hypertension in adults: a systematic review of longitudinal studies published after 2012. BMC Public Health 2020 May 19;20(1):726. [doi: 10.1186/s12889-020-08715-4] [Medline: 32429951]
3. Dempsey PC, Rowlands AV, Strain T, et al. Physical activity volume, intensity, and incident cardiovascular disease. Eur Heart J 2022 Dec 7;43(46):4789-4800. [doi: 10.1093/eurheartj/ehac613] [Medline: 36302445]
4. Bull FC, Al-Ansari SS, Biddle S, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020 Dec;54(24):1451-1462. [doi: 10.1136/bjsports-2020-102955] [Medline: 33239350]
5. Huang R, Palmer SC, Cao Y, et al. Cardiac rehabilitation programs for chronic heart disease: a Bayesian network meta-analysis. Can J Cardiol 2021 Jan;37(1):162-171. [doi: 10.1016/j.cjca.2020.02.072] [Medline: 32485140]
6. Graham I, Atar D, Borch-Johnsen K, et al. European guidelines on cardiovascular disease prevention in clinical practice: executive summary: Fourth Joint Task Force of the European Society of Cardiology and Other Societies on cardiovascular disease prevention in clinical practice (constituted by representatives of nine societies and by invited experts). Eur Heart J 2007 Oct;28(19):2375-2414. [doi: 10.1093/eurheartj/ehm316] [Medline: 17726041]
7. Graham H, Prue-Owens K, Kirby J, Ramesh M. Systematic review of interventions designed to maintain or increase physical activity post-cardiac rehabilitation phase II. Rehabil Process Outcome 2020;9:1179572720941833. [doi: 10.1177/1179572720941833] [Medline: 34497468]
8. Giuliano C, Parmenter BJ, Baker MK, et al. Cardiac rehabilitation for patients with coronary artery disease: a practical guide to enhance patient outcomes through continuity of care. Clin Med Insights Cardiol 2017;11:1179546817710028. [doi: 10.1177/1179546817710028] [Medline: 28638244]
9. Bjarnason-Wehrens B, McGee H, Zwisler AD, et al. Cardiac rehabilitation in Europe: results from the European cardiac rehabilitation inventory survey. Eur J Cardiovasc Prev Rehabil 2010 Aug;17(4):410-418. [doi: 10.1097/HJR.0b013e328334f42d] [Medline: 20300001]
10. Hansen D, Dendale P, Raskin A, et al. Long-term effect of rehabilitation in coronary artery disease patients: randomized clinical trial of the impact of exercise volume. Clin Rehabil 2010 Apr;24(4):319-327. [doi: 10.1177/0269215509353262] [Medline: 20176771]
11. Moore SM, Charvat JM, Gordon NH, et al. Effects of a CHANGE intervention to increase exercise maintenance following cardiac events. Ann Behav Med 2006 Feb;31(1):53-62. [doi: 10.1207/s15324796abm3101_9] [Medline: 16472039]
12. Pepera G, Antoniou V, Su JJ, Lin R, Batalik L. Comprehensive and personalized approach is a critical area for developing remote cardiac rehabilitation programs. World J Clin Cases 2024 Apr 26;12(12):2009-2015. [doi: 10.12998/wjcc.v12.i12.2009] [Medline: 38680265]
13. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008 Nov 6;5:56. [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
14. Jacobs DR Jr, Ainsworth BE, Hartman TJ, Leon AS. A simultaneous evaluation of 10 commonly used physical activity questionnaires. Med Sci Sports Exerc 1993 Jan;25(1):81-91. [doi: 10.1249/00005768-199301000-00012] [Medline: 8423759]
15. Ross R, Blair SN, Arena R, et al. Importance of assessing cardiorespiratory fitness in clinical practice: a case for fitness as a clinical vital sign: a scientific statement from the American Heart Association. Circulation 2016 Dec 13;134(24):e653-e699. [doi: 10.1161/CIR.0000000000000461] [Medline: 27881567]
16. Holland AE, Spruit MA, Troosters T, et al. An official European Respiratory Society/American Thoracic Society technical standard: field walking tests in chronic respiratory disease. Eur Respir J 2014 Dec;44(6):1428-1446. [doi: 10.1183/09031936.00150314] [Medline: 25359355]
17. Fick A. By the measurement of blood quantum in the ventricles. Presented at: Proceedings of the Physico-Medical Society of Würzburg; 1870; Würzburg, Germany p. 16-17.
18. Nes BM, Gutvik CR, Lavie CJ, Nauman J, Wisløff U. Personalized Activity Intelligence (PAI) for prevention of cardiovascular disease and promotion of physical activity. Am J Med 2017 Mar;130(3):328-336. [doi: 10.1016/j.amjmed.2016.09.031] [Medline: 27984009]

XSL•FO
RenderX

19. Sartor F, Gelissen J, van Dinther R, Roovers D, Papini GB, Coppola G. Wrist-worn optical and chest strap heart rate comparison in a heterogeneous sample of healthy individuals and in coronary artery disease patients. BMC Sports Sci Med Rehabil 2018;10:10. [doi: 10.1186/s13102-018-0098-0] [Medline: 29881626]
20. Sartor F, Vernillo G, de Morree HM, et al. Estimation of maximal oxygen uptake via submaximal exercise testing in sports, clinical, and home settings. Sports Med 2013 Sep;43(9):865-873. [doi: 10.1007/s40279-013-0068-3] [Medline: 23821468]
21. Achten J, Jeukendrup AE. Heart rate monitoring: applications and limitations. Sports Med 2003;33(7):517-538. [doi: 10.2165/00007256-200333070-00004] [Medline: 12762827]
22. Jones AK, Yan CL, Rivera Rodriquez BP, Kaur S, Andrade-Bucknor S. Role of wearable devices in cardiac telerehabilitation: a scoping review. PLoS ONE 2023;18(5):e0285801. [doi: 10.1371/journal.pone.0285801] [Medline: 37256878]
23. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
24. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak 2007 Jun 15;7:16. [doi: 10.1186/1472-6947-7-16] [Medline: 17573961]
25. Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. De-duplication of database search results for systematic reviews in EndNote. J Med Libr Assoc 2016 Jul;104(3):240-243. [doi: 10.3163/1536-5050.104.3.014] [Medline: 27366130]
26. Crowe M, Sheppard L, Campbell A. Reliability analysis for a proposed critical appraisal tool demonstrated value for diverse research designs. J Clin Epidemiol 2012 Apr;65(4):375-383. [doi: 10.1016/j.jclinepi.2011.08.006] [Medline: 22078576]
27. Crowe M. Crowe Critical Appraisal Tool (CCAT) user guide version 14. JCU Library. 2013. URL: https://conchra.com.au/wp-content/uploads/2015/12/CCAT-user-guide-v1.4.pdf [accessed 2026-06-04]
28. Alonso WW, Kupzyk KA, Norman JF, et al. The HEART Camp exercise intervention improves exercise adherence, physical function, and patient-reported outcomes in adults with preserved ejection fraction HEART Failure. J Card Fail 2022 Mar;28(3):431-442. [doi: 10.1016/j.cardfail.2021.09.003] [Medline: 34534664]
29. Batalik L, Dosbaba F, Hartman M, Batalikova K, Spinar J. Benefits and effectiveness of using a wrist heart rate monitor as a telerehabilitation device in cardiac patients: a randomized controlled trial. Medicine (Baltimore) 2020 Mar;99(11):e19556. [doi: 10.1097/MD.0000000000019556] [Medline: 32176113]
30. Chung H, Lee H, Kim C, Hong S, Lee J. Patient-provider interaction system for efficient home-based cardiac rehabilitation exercise. IEEE ACCESS 2019;7:14611-14622. [doi: 10.1109/ACCESS.2019.2892717]
31. Dobrican RA, Zampunieris D. A proactive solution, using wearable and mobile applications, for closing the gap between the rehabilitation team and cardiac patients. Presented at: 2016 IEEE International Conference on Healthcare Informatics (ICHI); Oct 4-6, 2016; Chicago, IL p. 146-155. [doi: 10.1109/ICHI.2016.23]
32. Medina Quero J, Fernández Olmo MR, Peláez Aguilera MD, Espinilla Estévez M. Real-time monitoring in home-based cardiac rehabilitation using wrist-worn heart rate devices. SENSORS (Basel) 2017 Dec 12;17(12):2892. [doi: 10.3390/s17122892] [Medline: 29231887]
33. Xia K, Bai C, Ding R, et al. Feasibility of a smartphone app for prescribed exercise tutoring in patients with stable coronary heart disease. Digit Health 2023;9:20552076231197424. [doi: 10.1177/20552076231197424] [Medline: 37654723]
34. Koffman LJ, Crainiceanu CM, Roemmich RT, French MA. Identifying unique subgroups of individuals with stroke using heart rate and steps to characterize physical activity. J Am Heart Assoc 2023 Sep 19;12(18):e030577. [doi: 10.1161/JAHA.123.030577] [Medline: 37681556]
35. Schubert C, Archer G, Zelis JM, et al. Wearable devices can predict the outcome of standardized 6-minute walk tests in heart disease. NPJ Digit Med 2020;3:92. [doi: 10.1038/s41746-020-0299-2] [Medline: 32665977]
36. Sandberg C, Pomeroy J, Thilén U, Gradmark A, Wadell K, Johansson B. Habitual physical activity in adults with congenital heart disease compared with age- and sex-matched controls. Can J Cardiol 2016 Apr;32(4):547-553. [doi: 10.1016/j.cjca.2015.08.021] [Medline: 26777265]
37. Weeks DL, Sprint GL, Stilwill V, Meisen-Vehrs AL, Cook DJ. Implementing wearable sensors for continuous assessment of daytime heart rate response in inpatient rehabilitation. Telemed J E Health 2018 Apr 2;24(12):1014-1020. [doi: 10.1089/tmj.2017.0306] [Medline: 29608421]
38. Kraal JJ, Van den Akker-Van Marle ME, Abu-Hanna A, Stut W, Peek N, Kemps HM. Clinical and cost-effectiveness of home-based cardiac rehabilitation compared to conventional, centre-based cardiac rehabilitation: results of the FIT@Home study. Eur J Prev Cardiol 2017 Aug;24(12):1260-1273. [doi: 10.1177/2047487317710803] [Medline: 28534417]
39. Kraal JJ, Sartor F, Papini G, et al. Energy expenditure estimation in beta-blocker-medicated cardiac patients by combining heart rate and body movement data. Eur J Prev Cardiol 2016 Nov;23(16):1734-1742. [doi: 10.1177/2047487316667786] [Medline: 27625154]
40. Rissanen APE, Rottensteiner M, Kujala UM, Kurkela JLO, Wikgren J, Laukkanen JA. Cardiorespiratory fitness estimation based on heart rate and body acceleration in adults with cardiovascular risk factors: validation study. JMIR Cardio 2022 Oct 25;6(2):e35796. [doi: 10.2196/35796] [Medline: 36282560]
41. Sprint G, Cook D, Weeks D, Dahmen J, La Fleur A. Analyzing sensor-based time series data to track changes in physical activity during inpatient rehabilitation. Sensors (Basel) 2017 Sep 27;17(10):10. [doi: 10.3390/s17102219] [Medline: 28953257]

XSL•FO
RenderX

42. Hannan AL, Hing W, Coombes JS, et al. Effect of personal activity intelligence (PAI) monitoring in the maintenance phase of cardiac rehabilitation: a mixed methods evaluation. BMC Sports Sci Med Rehabil 2021 Oct 10;13(1):124. [doi: 10.1186/s13102-021-00350-9] [Medline: 34629086]
43. Crozier A, Cocks M, Hesketh K. 28 Mobile health biometrics to prescribe immediate remote physical activity for enhancing uptake to cardiac rehabilitation (motivate-CR+): protocol for a randomised controlled feasibility trial. In: BACPR Annual Conference Abstracts 2023:A14-A15. [doi: 10.1136/heartjnl-2023-BACPR.24]
44. Mansfield A, Brooks D, Tang A, et al. Promoting Optimal Physical Exercise for Life (PROPEL): aerobic exercise and self-management early after stroke to increase daily physical activity-study protocol for a stepped-wedge randomised trial. BMJ Open 2017 Jun 30;7(6):e015843. [doi: 10.1136/bmjopen-2017-015843] [Medline: 28667222]
45. Dosbaba F, Hartman M, Hnatiak J, Batalik L, Ludka O. Effect of home-based high-intensity interval training using telerehabilitation among coronary heart disease patients. Medicine (Baltimore) 2020 Nov 20;99(47):e23126. [doi: 10.1097/MD.0000000000023126] [Medline: 33217814]
46. Suchy C, Massen L, Rognmo Ø, et al. Optimising exercise training in prevention and treatment of diastolic heart failure (OptimEx-CLIN): rationale and design of a prospective, randomised, controlled trial. Eur J Prev Cardiolog 2014 Nov;21(2_suppl):18-25. [doi: 10.1177/2047487314552764]
47. Brouwers RWM, Kraal JJ, Traa SCJ, Spee RF, Oostveen L, Kemps HMC. Effects of cardiac telerehabilitation in patients with coronary artery disease using a personalised patient-centred web application: protocol for the SmartCare-CAD randomised controlled trial. BMC Cardiovasc Disord 2017 Jan 31;17(1):46. [doi: 10.1186/s12872-017-0477-6] [Medline: 28143388]
48. Kieffer SK, Croci I, Wisløff U, Nauman J. Temporal changes in a novel metric of physical activity tracking (personal activity intelligence) and mortality: the HUNT study, Norway. Prog Cardiovasc Dis 2019;62(2):186-192. [doi: 10.1016/j.pcad.2018.09.002] [Medline: 30227185]
49. Kieffer SK, Nauman J, Syverud K, et al. Association between personal activity intelligence (PAI) and body weight in a population free from cardiovascular disease - the HUNT study. Lancet Reg Health Eur 2021 Jun;5:100091. [doi: 10.1016/j.lanepe.2021.100091] [Medline: 34557819]
50. Nauman J, Sui X, Lavie CJ, et al. Personal activity intelligence and mortality - data from the aerobics center longitudinal study. Prog Cardiovasc Dis 2021;64:121-126. [doi: 10.1016/j.pcad.2020.05.005] [Medline: 32560967]
51. Nauman J, Nes BM, Zisko N, et al. Personal activity intelligence (PAI): a new standard in activity tracking for obtaining a healthy cardiorespiratory fitness level and low cardiovascular risk. Prog Cardiovasc Dis 2019;62(2):179-185. [doi: 10.1016/j.pcad.2019.02.006] [Medline: 30797801]
52. Quer G, Gouda P, Galarnyk M, Topol EJ, Steinhubl SR. Inter- and intraindividual variability in daily resting heart rate and its associations with age, sex, sleep, BMI, and time of year: retrospective, longitudinal cohort study of 92,457 adults. PLoS ONE 2020;15(2):e0227709. [doi: 10.1371/journal.pone.0227709] [Medline: 32023264]
53. Bannell DJ, France-Ratcliffe M, Buckley BJR, et al. Adherence to unsupervised exercise in sedentary individuals: a randomised feasibility trial of two mobile health interventions. Digit Health 2023;9:20552076231183552. [doi: 10.1177/20552076231183552] [Medline: 37426588]
54. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
55. Brage S, Ekelund U, Brage N, et al. Hierarchy of individual calibration levels for heart rate and accelerometry to measure physical activity. J Appl Physiol 2007 Aug;103(2):682-692. [doi: 10.1152/japplphysiol.00092.2006]
56. Schrack JA, Leroux A, Fleg JL, et al. Using heart rate and accelerometry to define quantity and intensity of physical activity in older adults. J Gerontol A Biol Sci Med Sci 2018 Apr 17;73(5):668-675. [doi: 10.1093/gerona/gly029] [Medline: 29509832]
57. Dempsey PC, Matthews CE, Dashti SG, et al. Sedentary behavior and chronic disease: mechanisms and future directions. J Phys Act Health 2020 Jan 1;17(1):52-61. [doi: 10.1123/jpah.2019-0377] [Medline: 31794961]
58. Farrahi V, Rostami M, Dumuid D, et al. Joint profiles of sedentary time and physical activity in adults and their associations with cardiometabolic health. Med Sci Sports Exerc 2022 Dec 1;54(12):2118-2128. [doi: 10.1249/MSS.0000000000003008] [Medline: 35881930]
59. Lavie CJ, Ozemek C, Carbone S, Katzmarzyk PT, Blair SN. Sedentary behavior, exercise, and cardiovascular health. Circ Res 2019 Mar;124(5):799-815. [doi: 10.1161/CIRCRESAHA.118.312669] [Medline: 30817262]
60. Ryan DJ, Stebbings GK, Onambele GL. The emergence of sedentary behaviour physiology and its effects on the cardiometabolic profile in young and older adults. Age (Dordr) 2015 Oct;37(5):89. [doi: 10.1007/s11357-015-9832-7] [Medline: 26315694]
61. Gerrior S, Juan W, Basiotis P. An easy approach to calculating estimated energy requirements. Prev Chronic Dis 2006 Oct;3(4):A129. [Medline: 16978504]
62. Firstbeat analytics: the strong partner that joined Garmin. Garmin. 2020. URL: https://www.garmin.com/en-SG/blog/firstbeat-analytics-the-strong-partner-that-joined-garmin [accessed 2025-04-23]
63. Gao WD, Nuuttila OP, Fang HB, Chen Q, Chen X. A new fitness test of estimating VO2max in well-trained rowing athletes. Front Physiol 2021;12:701541. [doi: 10.3389/fphys.2021.701541] [Medline: 34276423]

XSL•FO
RenderX

64. Nes BM, Gutvik CR, Lavie CJ, Nauman J, Wisløff U. Personalized activity intelligence (PAI) for prevention of cardiovascular disease and promotion of physical activity. Am J Med 2017 Mar;130(3):328-336. [doi: 10.1016/j.amjmed.2016.09.031] [Medline: 27984009]
65. Zisko N, Skjerve KN, Tari AR, et al. Personal activity intelligence (PAI), sedentary behavior and cardiovascular risk factor clustering - the HUNT study. Prog Cardiovasc Dis 2017;60(1):89-95. [doi: 10.1016/j.pcad.2017.02.007] [Medline: 28274818]
66. Kieffer SK, Zisko N, Coombes JS, Nauman J, Wisløff U. Personal activity intelligence and mortality in patients with cardiovascular disease: the HUNT study. Mayo Clin Proc 2018 Sep;93(9):1191-1201. [doi: 10.1016/j.mayocp.2018.03.029] [Medline: 30193673]
67. Hammer P, Tari AR, Franklin BA, Wen CP, Wisløff U, Nauman J. Personal activity intelligence and ischemic heart disease in a healthy population: China Kadoorie biobank study. J Clin Med 2022 Nov 4;11(21):6552. [doi: 10.3390/jcm11216552] [Medline: 36362780]
68. Tari AR, Selbæk G, Franklin BA, et al. Temporal changes in personal activity intelligence and the risk of incident dementia and dementia related mortality: a prospective cohort study (HUNT). EClinicalMedicine 2022 Oct;52:101607. [doi: 10.1016/j.eclinm.2022.101607] [Medline: 36034407]
69. Personal activity intelligence (PAI). CERG. 2015. URL: https://www.ntnu.edu/cerg/personal-activity-intelligence [accessed 2026-05-23]
70. Wai JPM, Wen CP, Tsai MK, et al. Association between activity quotient and cause-specific mortality - a prospective cohort study of 0.5 million participants in Asia. Prog Cardiovasc Dis 2025;89:53-60. [doi: 10.1016/j.pcad.2025.01.004] [Medline: 39798594]
71. Vermunicht P, Buyck C, Naessens S, et al. A novel machine learning procedure to detect and remove artefacts in heart rate data obtained from photoplethysmography wearables: a prospective cohort study. Digit Health 2026;12:20552076261426622. [doi: 10.1177/20552076261426622] [Medline: 41928833]
72. Buyck C, Vermunicht P, Piedrahita Giraldo JS, et al. Data driven physical activity quantification using chest strap-based heart rate measurements: the Antwerp Activity Index. Eur Heart J 2025 Nov 5;46(Supplement_1):46(Supplement_1. [doi: 10.1093/eurheartj/ehaf784.3883]
73. Fine J, Branan KL, Rodriguez AJ, et al. Sources of inaccuracy in photoplethysmography for continuous cardiovascular monitoring. Biosensors (Basel) 2021 Apr 16;11(4):126. [doi: 10.3390/bios11040126] [Medline: 33923469]
74. Charlton PH, Allen J, Bailón R, et al. The 2023 wearable photoplethysmography roadmap. Physiol Meas 2023 Nov 29;44(11):11. [doi: 10.1088/1361-6579/acead2] [Medline: 37494945]
75. Vermunicht P, Buyck C, Naessens S, et al. Optimization and pre-use suitability selection for wrist photoplethysmography-based heart rate monitoring in patients with cardiac disease. Eur Heart J Digit Health 2025 Sep;6(5):1024-1035. [doi: 10.1093/ehjdh/ztaf084] [Medline: 40984980]
76. Vermunicht P, Makayed K, Buyck C, et al. Continuous heart rate measurements in patients with cardiac disease: device comparison and development of a novel artefact removal procedure. Digit Health 2025;11:20552076251337598. [doi: 10.1177/20552076251337598] [Medline: 40547435]
77. Vermunicht P, Makayed K, Meysman P, et al. Validation of Polar H10 chest strap and Fitbit Inspire 2 tracker for measuring continuous heart rate in cardiac patients: impact of artefact removal algorithm. Europace 2023 May 24;25(Supplement_1):25. [doi: 10.1093/europace/euad122.550]
78. Vermunicht P, Buyck C, Makayed K, et al. Optimisation of artefact detection in photoplethysmography heart rate data: influence of different classifiers in machine learning models. Eur Heart J 2024 Oct 28;45(Supplement_1):45. [doi: 10.1093/eurheartj/ehae666.3533]
79. Ibrahim NS, Rampal S, Lee WL, Pek EW, Suhaimi A. Evaluation of wrist-worn photoplethysmography trackers with an electrocardiogram in patients with ischemic heart disease: a validation study. Cardiovasc Eng Technol 2024 Feb;15(1):12-21. [doi: 10.1007/s13239-023-00693-z] [Medline: 37973701]
80. Piepoli MF, Abreu A, Albus C, et al. Update on cardiovascular prevention in clinical practice: a position paper of the European Association of Preventive Cardiology of the European Society of Cardiology. Eur J Prev Cardiolog 2020 Jan;27(2):181-205. [doi: 10.1177/2047487319893035]
81. Batalik L, Su JJ, Yang C, Pepera G. Enhancing patient activation in cardiovascular health through personalized mobile health interventions and social support. Eur J Cardiovasc Nurs 2025 Mar 3;24(2):323-324. [doi: 10.1093/eurjcn/zvae161] [Medline: 39760154]
82. Mueller S, Kabelac M, Fegers-Wustrow I, et al. Comparison of exercise training modalities and change in peak oxygen consumption in heart failure with preserved ejection fraction: a secondary analysis of the OptimEx-Clin trial. Eur J Prev Cardiol 2025 Aug 25;32(11):926-936. [doi: 10.1093/eurjpc/zwae332]

Abbreviations

ACS: acute coronary syndrome
AQ: activity quotient
CABG: Coronary artery bypass grafting

XSL•FO
RenderX

CAD: coronary artery disease
CCAT: Crowe Critical Appraisal Tool
CENTRAL: Cochrane Central Register of Controlled Trials
CERG: Cardiac Exercise Research Group
CPET: cardiopulmonary exercise test
CR: cardiac rehabilitation
CRF: cardiorespiratory fitness
CVD: cardiovascular disease
ECG: electrocardiogram
ESC: European Society of Cardiology
HFpEF: heart failure with preserved ejection fraction
HFrEF: heart failure with reduced ejection fraction
HR: heart rate
HRR: heart rate reserve
MVPA: moderate-to-vigorous physical activity
NTNU: Norwegian University of Science and Technology
PA: physical activity
PACD: Physical Activity Change Detection
PAI: Personal Activity Intelligence
PAL: physical activity level
PCI: percutaneous coronary intervention
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
TEE: total energy expenditure
VO_2 max: maximal oxygen uptake
WHO: World Health Organization

Edited by M Dorsch; submitted 02.Jul.2025; peer-reviewed by A Ferariu, D Parmak, G Pepera, RB Lopez; revised version received 05.Mar.2026; accepted 22.Apr.2026; published 17.Jun.2026.

Please cite as:
Vermunicht P, Buyck C, Naessens S, Hens W, Van Craenenbroeck E, Laukens K, Desteghe L, Heidbuchel H
Heart Rate Monitors for the Estimation of Physical Activity in Patients With Cardiovascular Disease: Systematic Review
JMIR Mhealth Uhealth 2026;14:e79995
URL: https://mhealth.jmir.org/2026/1/e79995
doi:10.2196/79995

© Paulien Vermunicht, Christophe Buyck, Sebastiaan Naessens, Wendy Hens, Emeline Van Craenenbroeck, Kris Laukens, Lien Desteghe, Hein Heidbuchel. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 17.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

mHealth Technologies for the Care of Children With Congenital Heart Disease: Scoping Review

Mingzhu Wang[1*], MSN; Qiuyin Pan[1*], BSc; Shengrong Tan[1*], BSc; Yuanchun Kong[1], BSc; Zhengling Dai[1], BSc; Jiao Cai[1], BSc; Duoqin Bi[1], BSc; Jiali Zhou[1,2*], MSN

[1]Children's Heart Center, Guizhou Branch of Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine, Guizhou Hospital, Guiyang, Guizhou, China

[2]Shanghai Children's Medical Center Affiliated to Shanghai Jiaotong University School of Medicine, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Jiali Zhou, MSN
Children's Heart Center, Guizhou Branch of Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine, Guizhou Hospital, Guiyang, Guizhou, China

Abstract

Background: Mobile information technology (IT) is increasingly being used in the health care sector, and it can play a critical role in both the care of children with congenital heart disease (CHD) and the quality of life of their families.

Objective: This study aimed to conduct a scoping review of the application of mobile health (mHealth) technologies in the care of children with CHD. We summarized the forms of mHealth interventions and effects on CHD to provide a reference for future research in this field.

Methods: We searched PubMed; Embase; Web of Science; the Cochrane Library; CINAHL; China National Knowledge Infrastructure; Wanfang Data; the Chinese Biomedical Database; VIP Chinese Science and Technology Journal Database; National Guideline Clearinghouse of the United States; the website of the Registered Nurses' Association of Ontario, Canada; the Guidelines International Network; the American Heart Association; and the American Association of Cardiovascular and Pulmonary Rehabilitation. The search period was from the establishment of the databases to June 12, 2025. The retrieved literature was screened and analyzed.

Results: A total of 519 Chinese- and English-language articles were identified, with 44 (8.5%) studies meeting the inclusion criteria. The primary forms of mHealth interventions for patients with CHD included mobile apps, wearable devices, and remote monitoring equipment. The findings indicated that mHealth technologies could improve exercise capacity, nutritional status, psychological well-being, and quality of life in children with CHD.

Conclusions: The application of mHealth in the care of children with CHD is feasible and demonstrates positive effects. Future research should emphasize peer education and patient privacy protection while further exploring remote education and health management based on theoretical frameworks and intelligent ITs to enhance quality of life for both children with CHD and their parents.

(JMIR Mhealth Uhealth 2026;14:e83385) doi:10.2196/83385

KEYWORDS

congenital heart disease; mobile health; scoping review; pediatric; telemedicine

Introduction

Congenital heart disease (CHD) is a heterogeneous group of structural abnormalities in the cardiovascular system with an incidence rate of 7–9 per 1,000, ranking as the most common birth defect. Worldwide, over 1 million infants are born with CHD each year, and China reports 150,000 to 200,000 new cases annually [1-3]. CHD comprises a complex and diverse spectrum of conditions with severe consequences, imposing substantial economic burdens and psychological distress on families and society. As a major public health concern, CHD significantly impacts children's physical and mental health, as well as their quality of life.

Traditional face-to-face intervention models are often constrained by medical costs, travel distances, and time limitations, making it difficult to provide timely and effective personalized, precise, and continuous health care services. These constraints negatively impact patient health management. In contrast, mobile health (mHealth) apps transcend temporal and spatial barriers, ensuring both the continuity and accessibility of medical resources while motivating patients to actively engage in their own health management. Guidelines indicate

XSL•FO
RenderX

that remote comonitoring of cardiac patients' cardiopulmonary rehabilitation by both families and health care professionals is feasible, safe, and effective [4]. The use of mHealth technology for home monitoring of children with CHD can help identify clinical issues, prevent unnecessary emergency visits, ensure child safety, and positively impact caregivers' psychological well-being and health care use [5,6]. Although research on the application of mHealth technology in children with CHD has been conducted both domestically and internationally, it remains in the early stages. This study aimed to summarize the applications and effects of mHealth technology in the care of children with CHD, providing a reference to promote its further development in this field.

To address this gap in the literature, this study used a scoping review methodology to examine the application of mHealth technology in children with CHD. The specific objectives were to (1) systematically identify the application of forms of mobile IT in the nursing care of children with CHD; (2) summarize the effects of mobile IT on the exercise function, nutritional status, psychological state, and quality of life of these children; and (3) identify gaps in the existing research. This study aimed to highlight the limitations of the current literature and provide directions for future research on mHealth-supported self-management for children with CHD.

Methods

Literature Search Strategy

This study used a scoping review methodological framework to systematically map the current research landscape on mHealth management for children with CHD, identifying key concepts, evidence types, and knowledge gaps. Given the limited number of studies on the specific application of mHealth technologies in this pediatric population, this review aimed to delineate the overall research landscape rather than evaluate the effectiveness of specific interventions; therefore, a scoping review was chosen over a systematic review. This report adheres to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist. We systematically searched multiple databases from their inception to June 12, 2025, including international databases (PubMed, Embase, Web of Science, the Cochrane Library, the website of the Registered Nurse and CINAHL), Chinese databases (China National Knowledge Infrastructure, Wanfang Data, the Chinese Biomedical Database, and VIP Chinese Science and Technology Journal Database), and guideline repositories (Guidelines International Network, National Guideline Clearinghouse, National Institute for Health and Care Excellence, Scottish Intercollegiate Guidelines Network, the Canadian Medical Association clinical practice guideline, World Health Organization, American Heart Association, American Association of Cardiovascular and Pulmonary Rehabilitation, and Medlive). The search strategy for this study was structured around 3 core conceptual modules: the target population (children with CHD), the intervention measures (mHealth and digital technologies), and the research context (nutrition, exercise, management, monitoring, and education). In the target population module, we combined terms for "congenital heart disease." For the intervention measure module, keywords including "mobile health," "mHealth," "telemedicine," "smartphone," "mobile application," and "WeChat" were used. The modules were combined using the Boolean operator "AND."

This search strategy was first developed for the PubMed database and subsequently adapted for other databases according to their controlled vocabularies (eg, MeSH [Medical Subject Headings] terms) and syntax rules. Although the search strings varied across different databases, the core conceptual structure remained consistent to ensure the comprehensiveness and reproducibility of the search. The search strategy can be found in Multimedia Appendix 1.

Literature Inclusion and Exclusion Criteria

Inclusion criteria were (1) studies focusing on mHealth technologies for children with CHD, (2) a study population consisting of patients with CHD aged 18 years or younger, and (3) publications in Chinese or English. Exclusion criteria were (1) duplicate publications; (2) unavailable full-text articles; and (3) case reports, conference abstracts, or similar publication types.

Study Selection and Data Extraction Process

Literature screening and data extraction were independently conducted by 2 reviewers. First, the 2 reviewers independently performed title and abstract screening and full-text screening based on the inclusion and exclusion criteria; any disagreements were resolved through discussion or by consulting a third reviewer. Subsequently, the 2 reviewers independently extracted data using a predesigned structured data extraction form (Microsoft Excel format) and cross-checked the results. The extracted information included first author, publication year, country, study type, specific form of mHealth intervention (eg, apps, remote monitoring platforms, and social media), intervention content (eg, nutritional management, exercise rehabilitation, and parental psychological support), and main study findings. During the extraction process, any discrepancies were resolved through discussion to reach a consensus.

Results

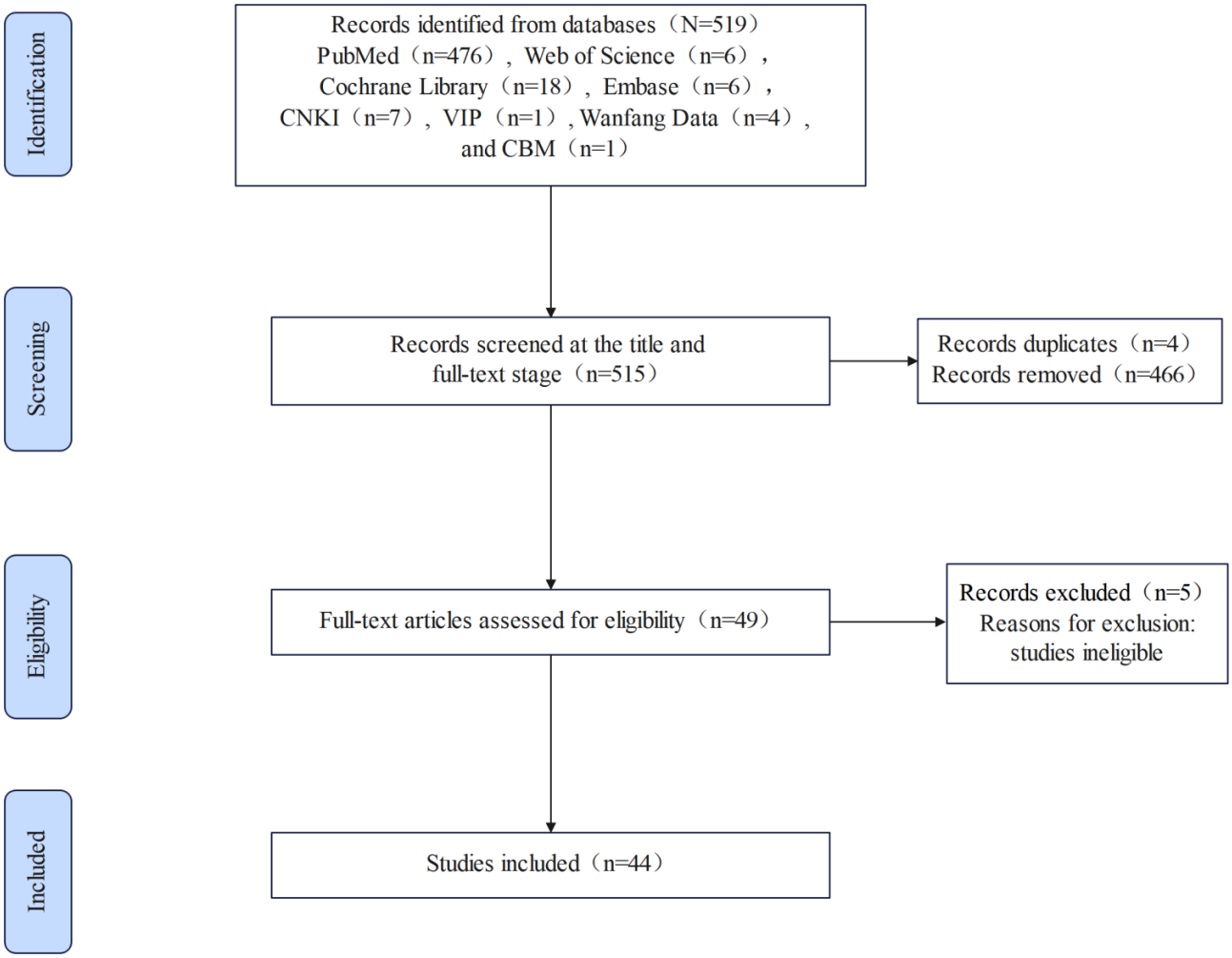
Literature Search Results

A preliminary search yielded a total of 519 relevant articles (n=476, 91.7% from PubMed; n=6, 1.2% from Web of Science; n=18, 3.5% from the Cochrane Library; n=6, 1.2% from Embase; n=7, 1.3% from China National Knowledge Infrastructure; n=1, 0.2% from VIP Chinese Science and Technology Journal Database; n=4, 0.8% from Wanfang Data; and n=1, 0.2% from the Chinese Biomedical Database). All articles were imported into a reference management software for deduplication. Of the 519 retrieved articles, after removing 4 (0.8%) duplicates, 515 (99.2%) remained for title and abstract screening. Following the review of titles and abstracts, of the 515 articles, 466 (90.5%) that were clearly irrelevant were excluded, leaving 49 (9.5%) for full-text review. During the full-text screening phase, of the 49 articles, 5 (10.2%) that did not meet the inclusion criteria were excluded, resulting in a final

XSL•FO
RenderX

inclusion of 44 (89.8%) articles, as shown in Figure 1 and Multimedia Appendix 2 [7-50].

Figure 1. Flowchart of the article screening process. CBM: Chinese Biomedical Database; CNKI: China National Knowledge Infrastructure.



Basic Characteristics of the Included Studies

Study Types and Country Distribution

Of the 44 included studies, 24 (54.5%) were intervention studies, including 9 (20.5%) randomized controlled trials and 6 (13.6%) observational studies. Geographically, 45.5% (n=20) of the studies were from China, mostly focusing on nursing interventions and health education using the WeChat platform; 20.5% (n=9) of the studies were from the United States; and 6.8% (n=3) of the studies were from Norway, emphasizing qualitative and survey research on parental acceptance of mobile apps and psychological support. The remaining studies were from India, the Netherlands, Italy, Japan, the United Kingdom, Germany, and Argentina.

Forms of Mobile IT Application

The forms of mobile IT used in the included studies comprised mobile apps for functions such as health education, symptom monitoring, and decision support; wearable devices for the collection of vital signs or physical activity data; and telemedicine platforms—including WeChat, video consultations, and telephone follow-ups—for postdischarge follow-up, home monitoring, or online guidance.

Overview of Intervention Content and Effects

The literature systematically demonstrated the broad application trend of mHealth technology in the management of children with CHD. In terms of intervention content, the studies focused on exercise intervention or cardiopulmonary rehabilitation, nutritional management, psychological support, and quality of life. Exercise interventions or cardiopulmonary rehabilitation facilitated the improvement of children's motor function through remote guidance or monitoring. Nutritional management encompassed breastfeeding guidance, removal of nasogastric tubes, and nutritional risk assessment. Psychological support primarily involved remote education and cognitive behavioral interventions to alleviate anxiety and depression in both children and their parents.

Discussion

Principal Findings

This study used a scoping review methodology to systematically examine the current status and effects of mobile IT in the nursing care of children with CHD. The findings indicate that mHealth technologies, primarily in the form of apps, wearable devices, and remote monitoring, demonstrated positive potential in improving exercise function, nutritional status, psychological

XSL•FO
RenderX

state, and quality of life in children with CHD. Additionally, these technologies significantly contributed to enhancing parental caregiving capacity and alleviating psychological burden. However, existing research still faces challenges such as high heterogeneity in interventions, lack of theoretical frameworks, privacy and security risks, and unequal accessibility of devices. These issues suggest that this field remains in the early stages of development, and future research requires further standardization and improvement.

Overview and Current Status of mHealth

mHealth uses mobile phones, apps, and monitoring devices to facilitate communication between patients and pediatric health care professionals, track health-related data, and enable access to pediatric health information. It incorporates functions such as tracking, feedback, and reminders, supporting disease screening, daily activity monitoring, medication adherence, and health education [51]. Unlike traditional telephone-based follow-up methods, mHealth has evolved toward greater specialization, systematization, and intelligence tailored to patients' disease characteristics and needs [52]. mHealth technology adherence refers to the extent to which health data transmitted by parents to health care providers align with recommended home-based symptom monitoring protocols. Parents with higher adherence to mHealth tools can more promptly identify symptoms in their children, leading to improved health outcomes [53]. Additionally, the mHealth management model offers a flexible and practical approach for remote postdischarge follow-up, diagnosis, and treatment in cardiac patients [54]. Currently, mHealth has become an indispensable medical tool; however, research on its application in CHD care remains limited. This review synthesized evidence on the use of mHealth in CHD management, aiming to provide insights for optimizing and expanding its adoption in pediatric CHD care.

Application Modalities of mHealth in Children With CHD

Mobile Apps

Currently, applications based on mobile phones, computers, and web platforms are widely used for vital sign monitoring, rehabilitation training, follow-up management, and health education in children with CHD. A study found that the implementation of an app integrating functionalities from prenatal diagnosis and neonatal screening to postoperative follow-up reduced the number of deaths among children with CHD from 598 to 353, representing a 41% decrease [7]. A survey also showed that, among 119 parents of children with CHD, 104 (87.4%) were willing to accept follow-up management based on WeChat [8]. Smartphone apps support remote monitoring of vital sign data, including oxygen saturation, blood pressure, temperature, and weight, and can assess medication adherence as well as the incidence and severity of symptoms. Relevant data can be uploaded to hospital information systems, facilitating timely awareness of the child's condition by medical staff. Furthermore, some apps incorporate integrated decision support systems to assist in formulating or adjusting treatment and nursing plans and provide health education materials in video and PDF formats along with online consultation features [9-11]. For instance, cardiac observation apps, through thoughtful design of content and functionality, effectively supplement traditional medical follow-up systems, helping parents gain more guidance and psychological support when facing their children's illness [12]. Carlson et al [13] developed a web-based automated follow-up system that automatically sends health status questionnaires to families at 30 days and 1 year after the child's discharge, aiding in the identification of high-risk patients requiring enhanced follow-up. The study also found a strong demand among families for knowledge related to heart disease and behavioral guidance. Future research could further explore the feasibility of conducting personalized remote rehabilitation, follow-up, and health education based on the severity of the child's condition and individual needs.

Wearable Devices

Current evidence indicates that wearable devices (eg, heart rate monitors, pedometers, and sensors) are primarily worn on the wrist, waist, or chest. These devices can continuously record vital sign parameters such as electrocardiography, heart rate, blood pressure, and oxygen saturation; some devices can also monitor weight changes [10,14]. However, existing studies are characterized by small sample sizes and short follow-up periods, making it difficult to assess long-term effects. Furthermore, the device models, monitored parameters, and intervention protocols vary considerably across studies, hindering cross-study comparability of results. Currently, wearables are predominantly used for monitoring, and there is a lack of integrated management models that combine them with early warning systems and remote interventions. Additionally, issues concerning the accuracy of device data, privacy and security, and children's compliance with wearing the devices require further investigation. Future research should focus on threshold setting for wearable device parameters and data storage functions. This would enable families to understand their children's health status in real time while also providing stored information to medical professionals. Such data would offer a basis for understanding children's condition at home and developing personalized intervention plans.

Remote Examination Devices

The application of remote examination devices in the diagnosis and treatment of children with CHD primarily encompasses modalities such as remote auscultation, remote consultation, remote diagnosis, and remote echocardiography. In the realm of remote auscultation, computer-aided auscultation systems, through heart sound acquisition, signal processing, and automated analysis algorithms, can effectively differentiate pathological murmurs, thereby reducing unnecessary referral burden [15]. Digital stethoscopes have also proven effective in identifying clinical issues, with no instances of missed diagnoses reported [16]. A remote diagnosis system in Japan has been successfully implemented for the prenatal diagnosis and pregnancy management of CHD in hospitals located in remote areas [17]. Regarding remote imaging diagnosis, tele-echocardiography has demonstrated accuracy, safety, and efficacy in screening children with CHD, potentially sparing families unnecessary interhospital transfer costs [18]. In terms

XSL•FO
RenderX

of teleconsultation, video-based teleconsultation saves time and medical costs for parents of children with CHD [19]. It also allows for modifications to the patient's treatment plan and facilitates communication regarding subsequent treatment strategies [20].

Despite the significant advantages demonstrated by remote examination devices, several challenges persist. Artificial intelligence–assisted auscultation requires expensive digital stethoscope equipment, and its widespread adoption and application may face economic and technological barriers. The operation of remote examination devices necessitates specialized training; the standardization of procedures and diagnostic consistency among primary health care providers require further evaluation. Furthermore, the lack of unified interoperability and data sharing standards among different devices limits the integrated application of telehealth systems. Future research should focus on reducing device costs, enhancing user-friendliness, and promoting the deployment of remote examination and diagnostic devices in primary care hospitals, thereby extending access to quality medical resources for patients in remote areas.

Potential of mHealth in Cardiac Rehabilitation

Physical Exercise

There is current evidence suggesting that mHealth technologies hold promise for enhancing the accessibility of cardiac exercise therapy and reducing implementation barriers [14]. However, findings regarding intervention effects remain inconsistent. Meyer et al [21] found that a web-based exercise intervention, while demonstrating the safety of 60 minutes of weekly physical activity, did not yield significant improvements in health-related physical fitness or quality of life among children with complex CHD. Conversely, other studies have indicated that wearable device–based cardiopulmonary rehabilitation can improve muscle endurance in children with heart disease [22]. Regarding study design, existing research on exercise interventions exhibits considerable heterogeneity. Variations in intervention duration, exercise modalities, and monitoring devices across studies render direct comparisons and synthesis of results difficult. Furthermore, most studies have small sample sizes and lack long-term follow-up data, precluding assessment of the long-term effects and safety of the interventions. The application of mHealth technologies in the field of exercise rehabilitation for children with CHD is still in an exploratory phase. Although existing evidence preliminarily confirms its safety and feasibility, its effectiveness remains unclear. Future research should conduct large-sample, multicenter randomized controlled trials that incorporate objective measures of physical capacity (eg, cardiopulmonary exercise testing) rather than relying solely on self-reports to validate the effectiveness of different exercise intervention protocols. Personalized remote exercise prescriptions could be developed tailored to children at different developmental stages—such as infants, school-aged children, and adolescents. Establishing a hospital-home remote supervision mechanism ensuring exercise safety under the joint monitoring of family members and medical staff is crucial. Additionally, exploring the integration of exercise monitoring data with early warning systems to enable real-time feedback and dynamic adjustments represents a key direction for future investigation.

Nutrition

Mobile ITs have played a positive role in nutritional risk identification and assessment, feeding guidance and nutritional improvement, and liberation from nasogastric tube feeding with associated home management for children with CHD. A telehealth home monitoring program facilitated the early identification of infants with CHD at risk of weight-for-age abnormalities, enabling timely customization of interventions by health care providers for this vulnerable population [23]. In another study, a nutritional risk assessment tool integrated into a mobile app automatically calculated and analyzed children's nutritional status upon data entry, providing decision support for clinicians to devise personalized nutritional interventions [24]. Multiple studies have demonstrated that telehealth education for parents of infants after CHD surgery—covering disease knowledge, postoperative care, home feeding, and complication management—can significantly enhance parental caregiving competence and improve the infants' postoperative nutritional status [25,26]. Notably, health education based on the WeChat platform has proven effective in improving breastfeeding efficiency and parental satisfaction for infants after CHD surgery [27]. Furthermore, remote nutritional management has been shown to promote neonatal growth and development while reducing the incidence of respiratory infections and hospital readmission rates [26]. Telehealth interventions grounded in multidisciplinary team collaboration combined with individualized patient assessment and effective communication with parents have facilitated the successful transition from nasogastric tube to oral feeding in children with CHD at home [28]. A WeChat-based discharge follow-up system also improved treatment adherence, enhanced cardiac function and nutritional status, and reduced the incidence of adverse events in these children [29].

Although existing studies generally support the positive effects of mHealth technologies on the nutritional management of children with CHD, several limitations remain. Some studies rely on parent-reported data, which may introduce recall bias, and the heterogeneity of nutritional outcome measures (eg, weight, length and height, BMI, feeding efficiency, and readmission rates) makes it difficult to conduct meta-analyses for integration. The nutritional assessment functions of mHealth tools are mostly limited to data calculation and currently lack deep integration with electronic health record systems or clinical decision support. Children with CHD are at higher nutritional risk than their healthy peers. Future research should establish a core outcome set, evaluate the differential effects of interventions across patients with varying nutritional risk stratification, focus on long-term nutritional follow-up in postoperative patients, explore deep integration of mHealth technologies with clinical nutritional guidelines, and enhance technology accessibility for low-income families.

Psychological Outcomes

Mobile ITs have also demonstrated utility in alleviating anxiety and depression and improving the psychological well-being of children with CHD, their parents, and adolescent patients. For

XSL•FO
RenderX

pediatric patients, these technologies have a positive impact on reducing anxiety and depression levels [30], with studies focusing on adolescents further confirming that mobile phone–based digital health interventions can enhance emotional well-being in this population [31]. Furthermore, Internet Plus–based health education, while enhancing disease-related knowledge among family members, can indirectly alleviate negative emotions in children and improve cardiac function while reducing complication rates [32]. At the parental level, multiple studies have shown that preoperative and postoperative health education and nursing guidance provided to parents of children with CHD via WeChat effectively ameliorates parental anxiety [33,34]. Nurse-guided mHealth interventions have also been associated with reduced anxiety and depression in parents following a prenatal diagnosis of critical CHD [35]. Additionally, telehealth services such as remote consultations and follow-up care enable timely assessment of the children's physical condition, reduce the need for hospital visits, and provide professional guidance for home care, thereby alleviating parental stress, anxiety, and depression and reducing caregiving burden [36-38].

However, existing research predominantly focuses on parental psychological states, with relatively fewer studies examining psychological outcomes in pediatric patients themselves, particularly young children. While interventions are diverse—including WeChat-based education, app-based interventions, and nurse-guided support—there is a lack of comparative studies evaluating the efficacy of different intervention modalities or assessing the long-term maintenance of psychological benefits. Furthermore, most studies use generic scales for psychological assessment, lacking condition-specific instruments validated for the pediatric CHD population. Children with CHD face unique psychosocial challenges during the transition to adulthood. Future research should prioritize the psychological health of patients during this critical transitional period, develop mHealth-based support programs for transition care, and explore integrated care models that combine psychological support for children with concurrent parental psychological interventions.

Quality of Life

There is current evidence suggesting that mHealth technology shows promising potential in improving quality of life for children with CHD and their parents, particularly demonstrating significant effects in enhancing treatment adherence, reducing complication rates, and alleviating parental caregiving burden. Studies have shown that telemedicine solutions have a positive impact on cardiovascular health [39]. Regarding the quality of life of pediatric patients, postdischarge follow-up management based on mHealth technology has demonstrated favorable intervention outcomes. Research indicates that using a WeChat-based follow-up nursing system can effectively improve out-of-hospital treatment adherence among children, reduce the incidence of adverse cardiac events, and enhance their quality of life [40]. An IT-supported transitional care model helped strengthen parental caregiving capabilities, improve nursing satisfaction, and reduce postoperative complications and readmission rates, thereby improving patient prognosis [41-43]. Remote health education targeting parents has also been associated with reduced complication rates and lower postdischarge loss to follow-up [33,44]. In terms of symptom monitoring, parents' use of mobile devices to upload daily symptom data facilitates timely identification of abnormalities by health care providers and enables prompt intervention, contributing to more positive health outcomes for children [53].

At the parental level, mHealth technologies have demonstrated multidimensional intervention value. Personalized decision support apps can enhance parents' understanding of common infant health conditions, assist them in completing health assessments and reports, and improve access to health care services [45]. Health education and psychological counseling delivered via the WeChat platform have proven effective in improving parental caregiving capacity, alleviating anxiety and grief, and reducing caregiving burden, thereby enhancing overall quality of life [46,47]. Individualized cardiac observation apps facilitate communication between parents and health care providers and improve readiness for hospital discharge [48]. Furthermore, WeChat-based guidance for preparing children for transthoracic echocardiography and for postoperative follow-up has been shown to alleviate parental anxiety and increase satisfaction with medication management [38,49]. Research focusing on mothers of children in intensive care units has similarly found that mHealth apps help reduce separation anxiety and improve the extent to which family needs are met [50].

Despite the generally positive findings, some studies have reported inconsistent effects of mHealth interventions. One telehealth home monitoring program, for instance, failed to significantly reduce psychological distress in parents of children with CHD [23]. These differences may be related to the complexity of the children's CHD and the technical usability of the mobile apps. Interventions targeting parental psychological support require further optimization. Current research exhibits considerable heterogeneity in intervention content, follow-up duration, and outcome measures, and no consensus has been reached regarding the effectiveness of interventions for parental stress. Therefore, current studies lack evidence on long-term quality of life improvements for both pediatric patients and their parents. Moreover, there is selection bias in the study designs—families willing to participate in mHealth research may inherently have higher health literacy and socioeconomic status, leading to overestimated effects [55,56]. Future research should incorporate intention-to-treat analyses and report dropout rates along with reasons for attrition. High-quality studies are still needed to identify optimal intervention strategies and the populations for which they are most appropriate.

Implementation Barriers: Health System–Level Challenges

Although the studies included in this review reported clinical effects of mHealth technologies, almost none discussed their implementation feasibility in real-world health systems. Below, we analyze barriers from 5 perspectives: cost, workflow integration, data security, equity, and recommendations for translation into practice.

XSL•FO
RenderX

Cost and Resource Allocation

Sustainable application of mHealth technologies requires ongoing hardware updates, software maintenance, and human resources. To date, no formal cost-effectiveness analysis of mHealth interventions for children with CHD has been conducted [57]. Moreover, the personnel costs associated with data monitoring are often concealed by research funding. Future studies should explicitly report direct and indirect costs as well as the incremental cost-effectiveness ratio.

Integration With Hospital Workflows

If an intervention cannot be integrated into existing workflows, its clinical efficacy loses practical relevance. Existing mHealth programs required nursing staff to periodically review data from apps but did not assess the additional workload. Furthermore, most apps were incompatible with hospital electronic medical record systems, necessitating manual data transcription and increasing the risk of errors and administrative burden [58]. Future research should prioritize platforms that seamlessly interface with electronic medical record systems and adopt automated alert algorithms to reduce manual review.

Data Security and Regulatory Compliance

mHealth technologies involve the transmission and storage of pediatric health information, and risks of privacy breaches during data collection, transmission, and storage are common. However, existing studies have inadequately described data encryption, anonymization, and user authorization mechanisms [59]. Future research should explicitly report data security measures in the “Methods” section and recommend developing open-source, locally compliant platforms.

Equity and Accessibility

The widespread use of mHealth technologies may exacerbate health inequities, with device costs and digital literacy being 2 core barriers for low-income families [60]. The initial investment in basic monitoring equipment (eg, smartphones and wearable sensors) and data plans may worsen the financial burden for some rural families, especially in the absence of insurance coverage or subsidies. Even when devices are available, insufficient digital literacy among caregivers (eg, inability to install apps or upload data) can compromise intervention effectiveness and adherence. Future research should report household income and baseline digital literacy and explore low-technology alternatives (eg, SMS text messages plus voice follow-up), device rental or subsidy programs, and community-based digital coaching as integrated support models.

Recommendations for Translation Into Practice

Future research, while reporting clinical outcomes, should focus on family acceptability, long-term intervention feasibility, nurse workload, and data security. Technologically, it is necessary to improve the accuracy of wearable devices; develop low-cost, high-precision solutions; and strengthen privacy protection and ethical compliance. Attention should be paid to digital accessibility for patients of low socioeconomic status and those in remote areas to narrow the digital divide. Research should be based on theoretical frameworks, include child-reported psychological outcomes and disease-specific measurement tools, and extend follow-up periods. Gamified interventions and qualitative research should be explored to gain an in-depth understanding of user needs.

Summary

mHealth technology holds positive potential in the nursing care of children with CHD, improving multiple health outcomes and reducing parental burden. Its mechanisms of action may be related to self-efficacy, adherence, and behavior change. Future research should conduct mechanism studies based on behavioral theory to provide a theoretical foundation for optimizing interventions.

Funding

This work was supported by two grants. The first grant is from the Guizhou Provincial Health Commission (Science and Technology Fund Project 2025, grant number zwkj2025-607). The second grant is from the Guizhou Provincial Department of Science and Technology under the Clinical Special Project (also referred to as the Guizhou Provincial Science and Technology Cooperation Achievement Project, General Program LC, where “LC” stands for “Linchuang” [meaning Clinical]), with grant number 161CHD (2026).

Authors' Contributions

Conceptualization: JZ
Data curation: MW, QP, ST
Formal analysis: MW, QP, ST
Writing – original draft: MW, ST, JZ, YK, ZD, JC, DB
Writing – review & editing: MW, QP, JZ

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy used in the PubMed database.
[XLSX File, 10 KB - mhealth_v14i1e83385_app1.xlsx]

XSL•FO
RenderX

Multimedia Appendix 2
Summary of included studies on digital health interventions for congenital heart disease.
[XLSX File, 120 KB - mhealth_v14i1e83385_app2.xlsx]

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 70 KB - mhealth_v14i1e83385_app3.docx]

References

1. Goldstein SA, Krasuski RA. Complex congenital heart disease in the adult. Annu Rev Med 2024 Jan 29;75:493-512. [doi: 10.1146/annurev-med-050922-052324] [Medline: 38285514]
2. Wei D, Zhang H, Li L, Wu T, Qin G, Luo Y. Feasibility analysis of simultaneous combined surgery for congenital heart disease complicated with other malformations. Chin J Thorac Cardiovasc Surg 2017;36(8):678-681 [FREE Full text]
3. Zimmerman MS, Smith AG, Sable CA. Global, regional, and national burden of congenital heart disease, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Child Adolesc Health 2020 Mar;4(3):185-200. [doi: 10.1016/S2352-4642(19)30402-X] [Medline: 31978374]
4. Valderrama P, Carugati R, Sardella A, et al. 2024 SIAC guidelines on cardiorespiratory rehabilitation in pediatric patients with congenital heart disease [Article in English, Spanish]. Rev Esp Cardiol (Engl Ed) 2024 Aug;77(8):680-689. [doi: 10.1016/j.rec.2024.05.001] [Medline: 38750931]
5. Stagg A, Giglia TM, Gardner MM, et al. Initial experience with telemedicine for interstage monitoring in infants with palliated congenital heart disease. Pediatr Cardiol 2023 Jan;44(1):196-203. [doi: 10.1007/s00246-022-02993-y] [Medline: 36050411]
6. Kauw D, Koole MA, van Dorth JR, et al. eHealth in patients with congenital heart disease: a review. Expert Rev Cardiovasc Ther 2018 Sep;16(9):627-634. [doi: 10.1080/14779072.2018.1508343] [Medline: 30079780]
7. Nair SM, Zheleva B, Dobrzycka A, Hesslein P, Sadanandan R, Kumar RK. A population health approach to address the burden of congenital heart disease in Kerala, India. Glob Heart 2021;16(1):71. [doi: 10.5334/gh.1034] [Medline: 34900562]
8. Lin WH, Chen YK, Lin SH, Cao H, Chen Q. Parents' understanding and attitudes toward the use of the WeChat platform for postoperative follow-up management of children with congenital heart disease. J Cardiothorac Surg 2023 Feb 8;18(1):66. [doi: 10.1186/s13019-023-02153-0] [Medline: 36755263]
9. Biersteker TE, Boogers MJ, de Lind van Wijngaarden RA, et al. Use of smart technology for the early diagnosis of complications after cardiac surgery: the Box 2.0 study protocol. JMIR Res Protoc 2020 Apr 21;9(4):e16326. [doi: 10.2196/16326] [Medline: 32314974]
10. Mastorci F, Lazzeri MF, Ait-Ali L, et al. Home-based intervention tool for cardiac telerehabilitation: protocol for a controlled trial. JMIR Res Protoc 2025 Jan 22;14:e47951. [doi: 10.2196/47951] [Medline: 39841521]
11. LeBlanc ME, Tallent S, Hornik CP, et al. Mobile health monitoring of children with CHDs. Cardiol Young 2024 Nov;34(11):2426-2431. [doi: 10.1017/S1047951124026222] [Medline: 39387239]
12. Hjorth-Johansen E, Børøsund E, Martinsen Østen I, Holmstrøm H, Moen A. Acceptability and initial adoption of the heart observation app for infants with congenital heart disease: qualitative study. JMIR Form Res 2023 Apr 5;7:e45920. [doi: 10.2196/45920] [Medline: 37018028]
13. Carlson L, O'Brien J, Gujral N, Chiang V, Del Nido P, Nathan M. Digital solution for follow-up in congenital cardiac surgery. Cardiol Young 2022 Jul;32(7):1032-1040. [doi: 10.1017/S1047951121003553] [Medline: 34497002]
14. White DA, Layton AM, Curran T, et al. eHealth technology in cardiac exercise therapeutics for pediatric patients with congenital and acquired heart conditions: a summary of evidence and future directions. Front Cardiovasc Med 2023;10:1155861. [doi: 10.3389/fcvm.2023.1155861] [Medline: 37332590]
15. Jani V, Danford DA, Thompson WR, Schuster A, Manlhiot C, Kutty S. The discerning ear: cardiac auscultation in the era of artificial intelligence and telemedicine. Eur Heart J Digit Health 2021;2(3):456-466. [doi: 10.1093/ehjdh/ztab059] [Medline: 36713594]
16. Stagg A, Giglia TM, Gardner MM, et al. Feasibility of digital stethoscopes in telecardiology visits for interstage monitoring in infants with palliated congenital heart disease. Pediatr Cardiol 2023 Dec;44(8):1702-1709. [doi: 10.1007/s00246-023-03198-7] [Medline: 37285041]
17. Mabuchi A, Waratani M, Tanaka Y, Mori T, Kitawaki J. Telediagnosis system for congenital heart disease in a Japanese prefecture. J Med Ultrason (2001) 2020 Jul;47(3):463-468. [doi: 10.1007/s10396-020-01020-y] [Medline: 32328807]
18. Makkar A, Milsten J, McCoy M, et al. Tele-echocardiography for congenital heart disease screening in a Level II Neonatal Intensive Care Unit with hybrid telemedicine system. Telemed J E Health 2021 Oct;27(10):1136-1142. [doi: 10.1089/tmj.2020.0440] [Medline: 33449839]
19. McCullough JE, Sinclair M, Gillender J, et al. Touching technology-parents' experiences of remote consultations for children with severe congenital cardiac conditions: quasi-experimental cohort study. JMIR Pediatr Parent 2024 Oct 22;7:e54598. [doi: 10.2196/54598] [Medline: 39437148]

XSL•FO
RenderX

20. Althabe M, Bove PM, Arancibia C, et al. Telemedicine in critical care: collaborative experience in postoperative pediatric cardiac care. Arch Cardiol Mex 2025;95(2):138-142. [doi: 10.24875/ACM.24000070] [Medline: 40445935]
21. Meyer M, Brudy L, Fuertes-Moure A, et al. E-health exercise intervention for pediatric patients with congenital heart disease: a randomized controlled trial. J Pediatr 2021 Jun;233:163-168. [doi: 10.1016/j.jpeds.2021.01.058] [Medline: 33516681]
22. Aronoff EB, Chin C, Opotowsky AR, et al. Facility-based and virtual cardiac rehabilitation in young patients with heart disease during the COVID-19 era. Pediatr Cardiol 2024 Oct;45(7):1533-1541. [doi: 10.1007/s00246-023-03202-0] [Medline: 37294337]
23. Medoff Cooper B, Marino BS, Fleck DA, et al. Telehealth home monitoring and postcardiac surgery for congenital heart disease. Pediatrics 2020 Sep;146(3):e20200531. [doi: 10.1542/peds.2020-0531] [Medline: 32817266]
24. Zhang QL, Lei YQ, Liu JF, Chen Q, Cao H. Telehealth education improves parental care ability and postoperative nutritional status of infants after CHD surgery: a prospective randomized controlled study. Paediatr Child Health 2022 Jun;27(3):154-159. [doi: 10.1093/pch/pxab094] [Medline: 35712039]
25. Zhang QL, Lin WH, Lin SH, Cao H, Chen Q. Using telemedicine strategy to implementing nutrition management for neonates after congenital heart disease surgery: a new nutrition management strategy. Front Pediatr 2022;10:918742. [doi: 10.3389/fped.2022.918742] [Medline: 35783326]
26. Horsley M, Hill GD, Kaskie S, Schnautz M, Brown J, Marcuccio E. Evaluation of an outpatient and telehealth initiative to reduce tube dependency in infants with complex congenital heart disease. Pediatr Cardiol 2022 Oct;43(7):1429-1437. [doi: 10.1007/s00246-022-02864-6] [Medline: 35333946]
27. Zhang Y, Lu L, Yang L, et al. Evaluation of a new digital pediatric malnutrition risk screening tool for hospitalized children with congenital heart disease. BMC Pediatr 2023 Mar 18;23(1):126. [doi: 10.1186/s12887-023-03899-1] [Medline: 36934232]
28. Yu XR, Xie WP, Liu JF, Dai WS, Cao H, Chen Q. Effect of WeChat follow-up education on breastfeeding efficiency of infants after congenital cardiac surgery and parents' satisfaction: a short-term follow-up study. J Paediatr Child Health 2021 Dec;57(12):1899-1904. [doi: 10.1111/jpc.15605] [Medline: 34085341]
29. Meng XM, Song P. Analysis of application effect of WeChat follow-up nursing in discharge follow-up care for children with congenital heart disease after closure surgery. Clin Res 2023;31(9):161-163 [FREE Full text]
30. Sutjonong T, Kalangi OA, Pulinggomang BB, Husada D. Impact of telemedicine on the mental status of parents of children with CHD: a meta-analysis. Pediatr Int 2023 Jan;65(1):e15457. [doi: 10.1111/ped.15457] [Medline: 36562667]
31. Willinger L, Oberhoffer-Fritz R, Ewert P, Müller J. Digital Health Nudging to increase physical activity in pediatric patients with congenital heart disease: a randomized controlled trial. Am Heart J 2023 Aug;262:1-9. [doi: 10.1016/j.ahj.2023.04.001] [Medline: 37030491]
32. Zhu WW, Jiang Y, Wang L. Application of "Internet + health education" in parental education of children with congenital heart disease and its effect on improving negative emotions of family members. Maternal and Child Health Care of China 2023;38(13):2471-2474 [FREE Full text]
33. Zhang QL, Xu N, Huang ST, Lin ZW, Cao H, Chen Q. WeChat-assisted health education and preoperative care improve the mental state of parents of children with ventricular septal defect. Psychol Health Med 2022 Apr;27(4):948-955. [doi: 10.1080/13548506.2021.1990360] [Medline: 34651528]
34. Jin L, Chen P, Fang X, Li C, Yang G, Liu A. Application of "Internet+" cognitive-behavioral intervention in caregivers of children with congenital heart disease undergoing elective surgery during the COVID-19 pandemic. Am J Transl Res 2023;15(4):2894-2902. [Medline: 37193177]
35. McKechnie AC, Elgersma KM, Ambrose MB, et al. Nurse-guided mobile health care program to reduce emotional distress experienced by parents of infants prenatally diagnosed with critical congenital heart disease: a pilot study. Prog Pediatr Cardiol 2024 Mar;72:101687. [doi: 10.1016/j.ppedcard.2023.101687] [Medline: 38130374]
36. Gillender J, McCullough JE, Browne R, Sinclair M, McCrossan B, Casey F. Experience of home monitoring of children with complex CHD during the COVID-19 pandemic: lessons learnt. Cardiol Young 2025 Feb;35(2):235-241. [doi: 10.1017/S1047951124026295] [Medline: 39865837]
37. Zhang QL, Huang ST, Xu N, Wang ZC, Cao H, Chen Q. Application of remote follow-up via the WeChat platform for patients who underwent congenital cardiac surgery during the COVID-19 epidemic. Braz J Cardiovasc Surg 2021 Aug 6;36(4):530-534. [doi: 10.21470/1678-9741-2020-0256] [Medline: 33355801]
38. Zhang QL, Liu JF, Xie WP, Cao H, Chen Q. The effect of WeChat on parental care burden, anxiety, and depression in children after CHD surgery during COVID-19 pandemic. Appl Clin Inform 2021 Aug;12(4):768-773. [doi: 10.1055/s-0041-1733850] [Medline: 34407558]
39. Schöneburg C, Willinger L, Uphoff I, Oberhoffer-Fritz R, Ewert P, Müller J. Interactive telehealth solutions for patients with congenital heart disease: a systematic review. CJC Pediatr Congenit Heart Dis 2025;4(2):69-80. [doi: 10.1016/j.cjcpc.2024.11.005] [Medline: 40470120]
40. Luo TN, Xie XR, Guo L. Application of WeChat follow-up nursing system in out-of-hospital care for discharged children with congenital heart disease after closure surgery. Chin Gen Pract Nurs 2021;19(1):95-97 [FREE Full text]

XSL•FO
RenderX

41. Zhang QL, Lin SH, Lin WH, Chen Q, Cao H. The effect of applying telehealth education to home care of infants after congenital heart disease surgery. Int J Qual Health Care 2023 Jan 14;35(1):mzac102. [doi: 10.1093/intqhc/mzac102] [Medline: 36537207]
42. Li GJ. Effect analysis of information-based methods combined with continuous nursing in out-of-hospital rehabilitation of children with complex congenital heart disease. Qinghai Med J 2024;54(2):36-39 [FREE Full text]
43. Chen L, Hu L, Zhou LH. Application of in-hospital comprehensive nursing combined with WeChat continuous intervention in children with critical congenital heart disease. J Qilu Nurs 2022;28(24):100-102 [FREE Full text]
44. Dun YT, Liu YC, Zhang LH. Evaluation on application effect of mobile internet management in continuous nursing of children after congenital heart disease surgery. Chin Nurs Manag 2019;19(7):968-972 [FREE Full text]
45. Hjorth-Johansen E, Børøsund E, Moen A, et al. Heart OBServation app: development of a decision support tool for parents of infants with severe cardiac disease. Cardiol Young 2023 Aug;33(8):1350-1358. [doi: 10.1017/S1047951122002438] [Medline: 35938297]
46. Zhang QL, Xu N, Huang ST, Chen Q, Cao H. WeChat-assisted preoperative health education reduces burden of care on parents of children with simple congenital heart disease: a prospective randomized controlled study. Braz J Cardiovasc Surg 2021 Oct 17;36(5):663-669. [doi: 10.21470/1678-9741-2020-0134] [Medline: 33355792]
47. Xie WP, Liu JF, Lei YQ, Cao H, Chen Q. Effects of WeChat follow-up management of infants who underwent ventricular septal defect repair on parents' disease knowledge and quality of life: a prospective randomized controlled study. J Card Surg 2021 Oct;36(10):3690-3697. [doi: 10.1111/jocs.15848] [Medline: 34318543]
48. Hjorth-Johansen E, Børøsund E, Moen A, et al. An individualised mobile app was beneficial for the mothers of infants with severe congenital heart defects. Acta Paediatr 2025 Jun;114(6):1238-1248. [doi: 10.1111/apa.17556] [Medline: 39707745]
49. Guo S, Lin WH, Lin SH, Zhang QL, Cao H, Chen Q. Using WeChat to guide preparation before transthoracic echocardiography reduces anxiety and improves satisfaction of parents of infants with congenital heart disease. J Cardiothorac Surg 2023 May 9;18(1):176. [doi: 10.1186/s13019-023-02225-1] [Medline: 37161515]
50. Liu Y, Chen X, Xiang Q, Wang H. Influence of application of ICU mobile app on demand satisfaction and separation anxiety of mothers of children with congenital heart disease. J Nurs Sci 2019;34(24):8-10 [FREE Full text]
51. Delva S, Waligora Mendez KJ, Cajita M, et al. Efficacy of mobile health for self-management of cardiometabolic risk factors: a theory-guided systematic review. J Cardiovasc Nurs 2021;36(1):34-55. [doi: 10.1097/JCN.0000000000000659] [Medline: 32040072]
52. Wang HH, Li YT, Huang J, Huang W, Wong MC. Advances and opportunities in the new digital era of telemedicine, e-health, artificial intelligence, and beyond. Hong Kong Med J 2023 Oct;29(5):380-382. [doi: 10.12809/hkmj235152] [Medline: 37880812]
53. Erickson LA, Emerson A, Russell CL. Parental mobile health adherence to symptom home monitoring for infants with congenital heart disease during the single ventricle interstage period: a concept analysis. J Spec Pediatr Nurs 2020 Oct;25(4):e12303. [doi: 10.1111/jspn.12303] [Medline: 32662243]
54. Wang L, Sun G, Tian J, et al. Application of mobile health management mode in remote management of patients with arrhythmia. J Nurs Sci 2017;32(3):78-80 [FREE Full text]
55. Baker-Smith CM, Sood E, Prospero C, Zadokar V, Srivastava S. Impact of social determinants and digital literacy on telehealth acceptance for pediatric cardiology care delivery during the early phase of the COVID-19 pandemic. J Pediatr 2021 Oct;237:115-124.e2. [doi: 10.1016/j.jpeds.2021.06.036] [Medline: 34174247]
56. Vaughan RM, Moore JA, Moreno JS, Dyer KJ, Oluyomi AO, Lopez KN. Remote care adoption in underserved congenital heart disease patients during the COVID-19 era. Pediatr Cardiol 2023 Feb;44(2):404-412. [doi: 10.1007/s00246-022-03042-4] [Medline: 36562781]
57. Sapanel Y, Tadeo X, Brenna CT, et al. Economic evaluation associated with clinical-grade mobile app-based digital therapeutic interventions: systematic review. J Med Internet Res 2023 Aug 1;25:e47094. [doi: 10.2196/47094] [Medline: 37526973]
58. Feldacker C, Usiri J, Kiruthu-Kamamia C, et al. Crossing the digital divide: the workload of manual data entry and integration between mobile health applications and eHealth infrastructure. Oxf Open Digit Health 2024;2(Suppl 2):ii9-ii17. [doi: 10.1093/oodh/oqae025] [Medline: 40191682]
59. Jiang J, Zheng Z. Medical information protection in internet hospital apps in China: scale development and content analysis. JMIR Mhealth Uhealth 2024 Jun 21;12:e55061. [doi: 10.2196/55061] [Medline: 38904994]
60. Ayemoba J, Schaller K, Nordgren R, Webster K, Romain C. Addressing disparities in telehealth access for children on the Southside of Chicago. Am J Surg 2025 Jul;245:116366. [doi: 10.1016/j.amjsurg.2025.116366] [Medline: 40344996]

Abbreviations

CHD: congenital heart disease
MeSH: Medical Subject Headings
mHealth: mobile health

XSL•FO
RenderX

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by M Dorsch; submitted 02.Sep.2025; peer-reviewed by N Barbazi, NR Lawrence; revised version received 19.Apr.2026; accepted 14.May.2026; published 17.Jul.2026.

Please cite as:
Wang M, Pan Q, Tan S, Kong Y, Dai Z, Cai J, Bi D, Zhou J
mHealth Technologies for the Care of Children With Congenital Heart Disease: Scoping Review
JMIR Mhealth Uhealth 2026;14:e83385
URL: https://mhealth.jmir.org/2026/1/e83385
doi:10.2196/83385

© Mingzhu Wang, Qiuyin Pan, Shengrong Tan, Yuanchun Kong, Zhengling Dai, Jiao Cai, Duoqin Bi, Jiali Zhou. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 17.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effects of Remote, Virtual, or Hybrid Cardiac Rehabilitation Supported by mHealth in Patients With Heart Failure: Systematic Review and Meta-Analysis

Kaidong Shao[1,2], MCM; Chunqiu Liu[1,2], MD; Tianshu Li[1,2], MCM; Huiyan Qu[1,2], MCM; Hua Zhou[1,2], MCM

[1]Institute of Cardiovascular Disease of Integrated Traditional Chinese and Western Medicine, Shuguang Hospital Affiliated to Shanghai University of Traditional Chinese Medicine, No. 528, Zhangheng Road, Pudong New Area, Shanghai, China

[2]Department of Cardiovascular Medicine, Shuguang Anhui Hospital Affiliated to Shanghai University of Traditional Chinese Medicine, Hefei, Anhui, China

Corresponding Author:
Hua Zhou, MCM
Institute of Cardiovascular Disease of Integrated Traditional Chinese and Western Medicine, Shuguang Hospital Affiliated to Shanghai University of Traditional Chinese Medicine, No. 528, Zhangheng Road, Pudong New Area, Shanghai, China

Abstract

Background: Structured exercise is a key component of cardiac rehabilitation (CR) for patients with heart failure (HF), but access to center-based cardiac rehabilitation (CBCR) is often limited. Mobile health (mHealth) platforms enable remote, virtual, or hybrid cardiac rehabilitation (RVH-CR) delivery.

Objective: This study aimed to evaluate the effectiveness and safety of structured, exercise-focused RVH-CR supported by mHealth compared with usual care or CBCR in patients with heart failure with reduced ejection fraction (HFrEF) or in HF populations predominantly comprising patients with HFrEF.

Methods: We searched PubMed, Web of Science, MEDLINE via Ovid, Cochrane CENTRAL, and CINAHL Complete from inception to April 27, 2026. Randomized controlled trials comparing mHealth-supported RVH-CR with usual care or CBCR were included. The primary outcome was exercise capacity, assessed by peak oxygen uptake (VO_2 peak) and 6-minute walk distance (6MWD). Secondary outcomes included health-related quality of life and safety. Data were pooled using random-effects meta-analysis stratified by comparator. Risk of bias was assessed with the Cochrane Risk of Bias Tool version 2, and evidence certainty was evaluated using GRADE (Grading of Recommendations Assessment, Development, and Evaluation).

Results: Eight randomized controlled trials with 1368 patients were included. In the CBCR comparison, mHealth-supported RVH-CR showed a statistically significant greater improvement in VO_2 peak than CBCR (mean difference [MD] 0.82, 95% CI 0.06-1.57; *P*=.03), although this finding was based on a limited number of trials. Compared with usual care, mHealth-supported RVH-CR was associated with improved 6MWD (MD 22.99, 95% CI 1.15-44.82; *P*=.04). Single-trial estimates suggested improvements in VO_2 peak (MD 2.50, 95% CI 0.88-4.12) and Minnesota Living with Heart Failure Questionnaire scores (standardized MD −0.57, 95% CI −0.98 to −0.17; *P*<.01) versus usual care. The certainty of evidence ranged from low to moderate. No intervention-related deaths or serious adverse events were reported, but sparse events and short follow-up limited conclusions regarding safety.

Conclusions: The effects of structured RVH-CR supported by mHealth differed according to comparator type, but the certainty of evidence ranged from low to moderate. Compared with usual care, mHealth-supported RVH-CR was associated with improved 6MWD. Compared with CBCR, mHealth-supported RVH-CR showed a significantly greater improvement in VO_2 peak in a limited number of trials, but superiority, equivalence, or noninferiority to CBCR cannot be concluded. Because usual care and CBCR are clinically distinct comparators, no single overall effect across comparator types should be inferred. Future studies should assess long-term outcomes and standardize structured exercise protocols across RVH-CR models.

Trial Registration: PROSPERO CRD420251162078; https://www.crd.york.ac.uk/PROSPERO/view/CRD420251162078

(JMIR Mhealth Uhealth 2026;14:e90422) doi:10.2196/90422

KEYWORDS

mobile health; mHealth; heart failure; meta-analysis; center-based cardiac rehabilitation; remote cardiac rehabilitation; virtual cardiac rehabilitation; hybrid cardiac rehabilitation

XSL•FO
RenderX

Introduction

Background

Heart failure (HF) is a complex clinical syndrome in which structural or functional cardiac abnormalities impair ventricular filling or ejection, giving rise to characteristic symptoms and signs [1]. HF remains a significant public health challenge globally [2,3]. Guidelines emphasize comprehensive disease management for patients with HF, including evidence-based pharmacotherapy, cardiac rehabilitation (CR), and structured follow-up [4,5]. Among HF phenotypes, heart failure with reduced ejection fraction (HFrEF) is clinically distinct and has been widely studied [6]. Multiple American Heart Association clinical practice guidelines endorse CR as an effective strategy for reducing mortality and improving quality of life across the spectrum of cardiovascular disease [7-9]. Core components of CR programs include patient assessment, individualized exercise training, nutritional counseling, psychological support, and health education, delivered by multidisciplinary teams comprising physicians, nurses, exercise physiologists, behavioral health specialists, dietitians, physical therapists, and respiratory therapists [10].

Exercise-based CR has been shown to improve functional capacity and exercise tolerance, reduce overall and HF-specific hospitalization rates, and enhance health-related quality of life (HRQoL) [11]. Despite these benefits, referral and participation rates for center-based cardiac rehabilitation (CBCR) are low globally, limited by transportation, social support, and socioeconomic factors [12,13].

To address these access barriers, CR delivery has expanded beyond traditional facility-based programs to include remote, virtual, and hybrid models that may be delivered in the home, community centers, fitness facilities, or other noncenter settings [9,14,15]. Contemporary consensus documents emphasize delivery mode rather than location: virtual CR refers to synchronous, real-time audiovisual delivery; remote CR refers to asynchronous delivery with clinician monitoring, feedback, or communication; and hybrid CR combines facility-based, virtual, and/or remote components [16,17]. Digital technologies are increasingly integrated into CR delivery, offering solutions to the logistical challenges inherent to CBCR while expanding the reach of structured rehabilitation care [10,17].

In recent years, the proliferation of smartphones and mobile apps has driven growing interest in mobile health (mHealth)-supported remote, virtual, and hybrid cardiac rehabilitation (RVH-CR) delivery [18]. mHealth refers to the use of mobile and wireless technologies to support the achievement of health objectives, encompassing smartphones, tablets, mobile apps, SMS, wearable activity trackers, biosensors, and connected digital platforms that enable bidirectional communication and real-time monitoring of health data between patients and clinicians [19,20].

Objectives

In recent clinical trials, mHealth-supported RVH-CR has been evaluated for its impact on patients with HF. However, there remains a lack of systematic reviews focusing specifically on exercise-based, structured RVH-CR supported by mHealth in patients with HFrEF or HFrEF-predominant HF populations. A meta-analysis conducted by Li et al [21] in 2025 compared the effects of mHealth home-based CR interventions versus usual care and CBCR across different cardiovascular conditions but lacked a specific focus on HF. Furthermore, the majority of systematic reviews [22-25] have focused on the impact of remote technologies on HF prognosis, with limited attention to exercise-based telerehabilitation modalities (such as videoconferencing platforms, computer-based systems, smartphone applications, or web-based home monitoring systems).

Throughout this review, we use “mHealth-supported RVH-CR” as the umbrella term for eligible interventions. “Remote,” “virtual,” and “hybrid” refer to CR delivery models, whereas “mHealth” refers to the portable, wireless, or connected digital technology layer supporting exercise prescription, monitoring, data transmission, clinician feedback, or supervision. “Home-based” is used only to describe the physical setting in which some intervention sessions occurred. “Telerehabilitation” is retained only when referring to terminology used by individual trials or previous reviews; otherwise, interventions are described using the standardized categories of RVH-CR. Given the emerging but still distinct literature on heart failure with preserved ejection fraction (HFpEF), the target population of this review was restricted to HFrEF or HFrEF-predominant HF populations to maintain clinical and pathophysiological coherence.

This systematic review aimed to identify and synthesize evidence from randomized clinical trials comparing the efficacy of mHealth-supported RVH-CR versus usual care or supervised CBCR in patients with HFrEF or HFrEF-predominant HF populations.

Methods

This systematic review and meta-analysis were developed and reported according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Checklist 1) [26]. The protocol was registered in PROSPERO (CRD420251162078).

Search Strategy

We searched 5 databases (PubMed, Web of Science, MEDLINE via the Ovid interface, Cochrane CENTRAL, and CINAHL Complete) from their inception to April 27, 2026. We also manually checked the reference lists of related studies and reviews. The search strategy combined MeSH terms and keywords adjusted for each database. The electronic search strategy was developed by HZ after discussions with the other authors (Multimedia Appendix 1). Key search terms covered the following 4 concepts: “heart failure,” “cardiac rehabilitation,” “mobile applications,” and “exercise.”

Selection Criteria

Studies were eligible if they met all of the following criteria: (1) the participants were adults (aged ≥18 y) diagnosed with HFrEF, or individuals with HF predominantly characterized by HFrEF. Trials were considered HFrEF-predominant if eligibility

XSL•FO
RenderX

criteria specified reduced left ventricular ejection fraction, or if the enrolled population with HF was predominantly composed of patients with reduced left ventricular ejection fraction, as reported by the trial authors; (2) interventions were not classified as mHealth-supported RVH-CR solely because they used videoconferencing, telephone support, SMS, or general telehealth communication. To be eligible, the intervention had to combine structured exercise-based CR with a patient-facing portable, wireless, or connected digital technology that supported exercise prescription, execution, physiologic or activity monitoring, patient-generated data transmission, clinician feedback, or supervision; (3) the control group consisted of either usual care (routine outpatient follow-up and general lifestyle advice without structured exercise or mHealth technology) or CBCR (supervised, facility-based exercise training and face-to-face follow-up); (4) the study design was a randomized controlled trial (RCT); and (5) at least one prespecified outcome was reported—exercise capacity (peak oxygen uptake [VO_2 peak] or 6-minute walk distance [6MWD]) or HRQoL.

Studies were excluded if (1) SMS, telephone, or videoconferencing was used solely for one-way reminders, general health education, or administrative communication, without structured exercise management, patient-generated data capture, or rehabilitation-specific feedback; (2) the full text was unavailable or the study was incomplete; or (3) the publication was not an original report (eg, review, conference abstract, or commentary).

Study Selection

All database search results were imported into EndNote X9 (Clarivate) for automated deduplication, followed by manual verification to ensure accuracy. Two reviewers (KS and CL) independently screened titles and abstracts against the eligibility criteria, followed by a full-text review of potentially eligible records. Discrepancies at any stage were resolved through discussion, and unresolved disagreements were adjudicated by a third reviewer (HZ).

Data Extraction

Two reviewers (KS and CL) independently extracted data using a standardized form developed in Microsoft Excel. Extracted items included study characteristics, participant demographics, intervention and control group details, and outcome measures. Discrepancies were resolved by consensus, with HZ serving as an arbiter when necessary. When data were missing or unclear, the corresponding authors were contacted by email to request the required information.

Risk of Bias and Quality of Evidence

Two reviewers (KS and TL) independently assessed the risk of bias (RoB) of each included RCT using the Cochrane Risk of Bias Tool version 2 (RoB 2), which evaluates 5 domains: the randomization process, deviations from intended interventions, missing outcome data, measurement of the outcome, and selection of the reported result. Each domain was rated as low risk, some concerns, or high risk, with an overall judgment derived according to RoB 2 guidance. Disagreements were resolved through discussion, with author HQ adjudicating unresolved cases.

The certainty of evidence for each outcome was independently assessed by 2 reviewers (KS and TL) using the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) framework through GRADEpro software (Evidence Prime). Evidence certainty was rated across 5 domains—RoB, inconsistency, indirectness, imprecision, and publication bias—and classified as high, moderate, low, or very low.

Statistical Analysis

The primary outcome was exercise capacity, operationalized as VO_2 peak and 6MWD. Secondary outcomes were HRQoL, assessed using the 36-Item Short Form Health Survey (SF-36) and the Minnesota Living with Heart Failure Questionnaire (MLHFQ). Where outcomes were amenable to quantitative synthesis, pooled effect estimates with 95% CIs were calculated using a random-effects model (DerSimonian-Laird method), which was prespecified to account for anticipated clinical and methodological heterogeneity across trials. Separate meta-analyses were conducted for 2 predefined comparisons: mHealth-supported RVH-CR versus usual care and mHealth-supported RVH-CR versus CBCR.

The meta-analyses were conducted using change-from-baseline values rather than postintervention values. For each outcome, pooled estimates represent between-group differences in change over the intervention period. As all outcomes were continuous, the mean difference (MD) was used when studies used identical measurement instruments, whereas the standardized mean difference (SMD) was calculated when different instruments or scale versions were used. Effect sizes were interpreted according to Cohen conventions: small (SMD 0.2), medium (SMD 0.5), and large (SMD 0.8). Statistical heterogeneity was quantified using the Cochran Q test and the I^2 statistic. Substantial heterogeneity was predefined as I^2>50%, Q test, P<.05, and clinically meaningful variability in the direction or magnitude of effects. Assessment of publication bias via funnel plot asymmetry and Egger regression test was not performed, as fewer than 10 studies were available for any single comparison—a threshold below which such tests lack adequate statistical power. All analyses were performed using Stata version 18.0 (StataCorp).

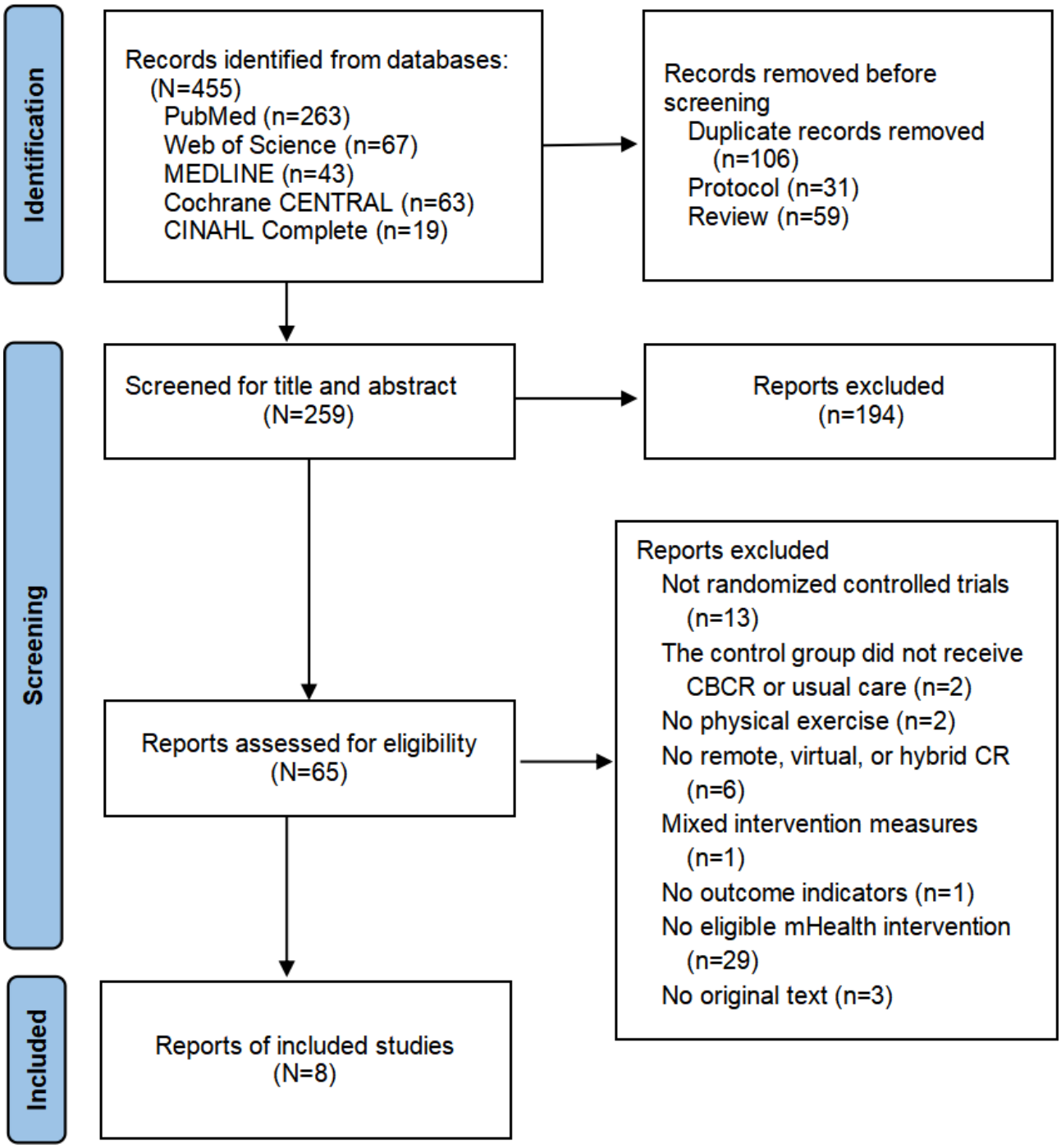
Results

Study Selection

As shown in Figure 1, 455 records were identified from 5 databases. After removing duplicates, protocols, and reviews, 259 records were screened by title and abstract. Of these, 194 were excluded, and 65 reports were assessed for eligibility.

XSL•FO
RenderX

Figure 1. Study selection. CBCR: center-based cardiac rehabilitation; CR: cardiac rehabilitation; mHealth: mobile health.



After full-text assessment, 57 reports were excluded because they were not RCTs, did not include an eligible control group, did not include a physical exercise or mHealth intervention, did not evaluate RVH-CR delivered outside traditional center-based settings, used mixed interventions, lacked relevant outcomes, or had no original text. Finally, 8 studies were included in the systematic review.

Study Characteristics

The 8 included studies [27-34] were published between 2010 and 2025. Five studies [30-34] were conducted in Europe, 2 [28,29] in Asia, and 1 [27] in Oceania. All studies were prospective, parallel-group RCTs that investigated mHealth-supported RVH-CR, with most patient exercise sessions conducted in the home setting. The included trials enrolled patients with HFrEF or HFrEF-predominant populations with HF. A condensed summary of the included studies is provided in Table 1, and detailed study characteristics are provided in Multimedia Appendix 2.

XSL•FO
RenderX

Table . Summary characteristics of the included randomized controlled trials.[a]

Study	Country	Total sample size, n (intervention:control)	Comparator	Eligibility rationale for mHealth-supported CR[b] classification
Hwang et al [27], 2017	Australia	53 (24:29)	CBCR[c] exercise plus education	Qualified as mHealth-supported[d] virtual CR because real-time video conferencing was combined with structured exercise supervision and patient-facing home monitoring of blood pressure and oxygen saturation rather than used only for remote communication.
Schmidt et al [33], 2025	Portugal	120 (75:45)	24-session supervised CBCR	Qualified because smartwatch-based exercise data upload and asynchronous clinician review supported structured home exercise prescription, monitoring, and feedback.
Piotrowicz et al [30], 2010	Poland	152 randomized; 131 included in final analysis (75:56)	Supervised center-based cycle-ergometer CR	Qualified because portable ECG[e] telemonitoring, mobile phone transmission, audiovisual prompts, and clinician contact were used to guide and supervise structured home exercise training.
Piotrowicz et al [31], 2021	Poland	850 randomized; QoL[f] analysis included 768 (377:391)	Usual care or observation with lifestyle and self-management advice	Qualified because hybrid CR incorporated connected ECG, blood pressure, weight, and device-related monitoring through a monitoring center to support structured CR delivery.
Nagatomi et al [28], 2022	Japan	30 (15:15)	Usual pharmacological and nonpharmacological care	Qualified because Fitbit-based activity monitoring, symptom/vital self-monitoring, and app/telephone feedback generated patient data used to support individualized exercise and clinical feedback.
Peng et al [29], 2018	China	98 (49:49)	Discharge education plus routine outpatient follow-up	Eligible because digital communication was integrated with structured exercise management, monitoring, and feedback rather than used only for education or reminders.
Piotrowicz et al [32], 2015	Poland	111 randomized; 107 included in final analysis (75:32)	Lifestyle and self-management advice; no supervised exercise	Eligible because portable connected monitoring was used to supervise and adjust structured Nordic walking training.

XSL•FO
RenderX

Study	Country	Total sample size, n (intervention:control)	Comparator	Eligibility rationale for mHealth-supported CR[b] classification
Lundgren et al [34], 2023	Norway	61 (31:30)	Usual care; guideline-based exercise encouraged	Eligible because synchronous CR was used to deliver structured exercise with active clinician supervision; however, it represents the broadest boundary of the mHealth-supported category and was retained because delivery depended on patient-facing connected digital technology.

[a]Values are mean or mean (SD) unless otherwise indicated. Randomized and final analysis samples are both reported when available. Detailed study characteristics are provided in Multimedia Appendix 2.

[b]CR: cardiac rehabilitation.

[c]CBCR: center-based cardiac rehabilitation.

[d]mHealth: mobile health.

[e]ECG: electrocardiogram.

[f]QoL: quality of life.

mHealth platforms in the included studies used broadband-connected laptops, tablet computers with videoconferencing systems, smartphones, smartwatches, and EHO (Pro Plus Company) devices [27-34]. Physiological monitoring was conducted using automated blood pressure monitors [27], electrocardiogram monitoring and transmission systems [30,31], wearable devices or smartwatches [28,33], and heart rate monitoring in a subset of participants [34]. Intervention durations ranged from 8 to 12 weeks. Exercise frequencies varied from daily to 5 times per week. The main type of exercise was aerobic exercise. Three studies also provided resistance training [28,29,31]. Two studies offered strength training [27,31]. Exercise prescriptions were generally individualized according to baseline functional capacity, symptoms, heart rate response, Borg perceived exertion score, or remotely transmitted physiological data. In most trials, physical therapists or multidisciplinary rehabilitation teams monitored progress and adjusted exercise intensity according to patient condition and training performance [27-29,33,34]. Psychological support was provided in 3 studies [30-32], nutritional counseling in 1 study [28], and lifestyle modification or self-management education was included in several studies [27,29,33,34].

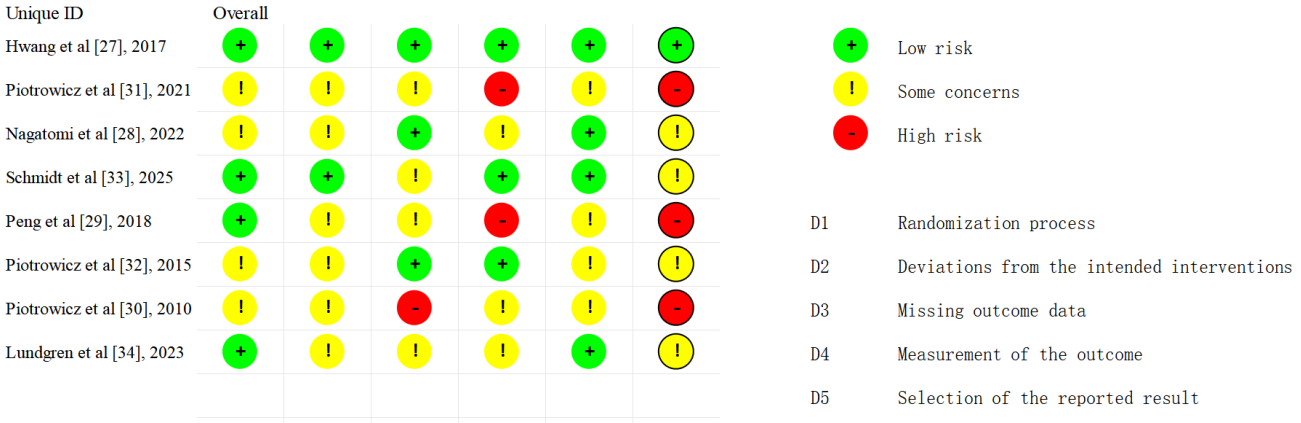
In 3 studies [27,30,33], the comparator was CBCR. In the remaining 5 studies [28,29,31,32,34], it was usual care. CBCR comprised supervised, facility-based exercise training and face-to-face educational or counseling sessions. Usual care consisted of routine outpatient follow-up, general recommendations for lifestyle modification, and self-management guidance, but excluded structured exercise training.

RoB

The included studies were assessed for RoB. One study [27] was considered to be at low risk because of concealed allocation, blinded outcome assessment, intention-to-treat analysis, trial registration, and minimal missing outcome data. Four studies [28,32-34] were rated as having some concerns, mainly due to the open-label nature of the interventions, incomplete reporting of allocation concealment or assessor blinding, baseline imbalance, or incomplete primary outcome data. Two studies [29,31] were judged to be at high risk primarily because their key outcomes were self-reported quality-of-life measures assessed in unblinded participants, whereas 1 study [30] was at high risk due to differential attrition and reliance on complete-case analysis. The most frequent sources of bias were lack of participant blinding, subjective outcome assessment, missing outcome data, and insufficient information on prespecified analysis plans. Therefore, pooled estimates for quality-of-life outcomes should be interpreted with caution, while findings based on objective exercise capacity outcomes, such as VO_2 peak and 6MWD, are likely to be less vulnerable to measurement bias but remain affected by attrition and incomplete blinding in selected trials. The RoB judgments are summarized in Figure 2. The specific details can be found in Multimedia Appendix 3.

XSL•FO
RenderX

Figure 2. Each risk of bias domain for each randomized controlled trial [27-34].



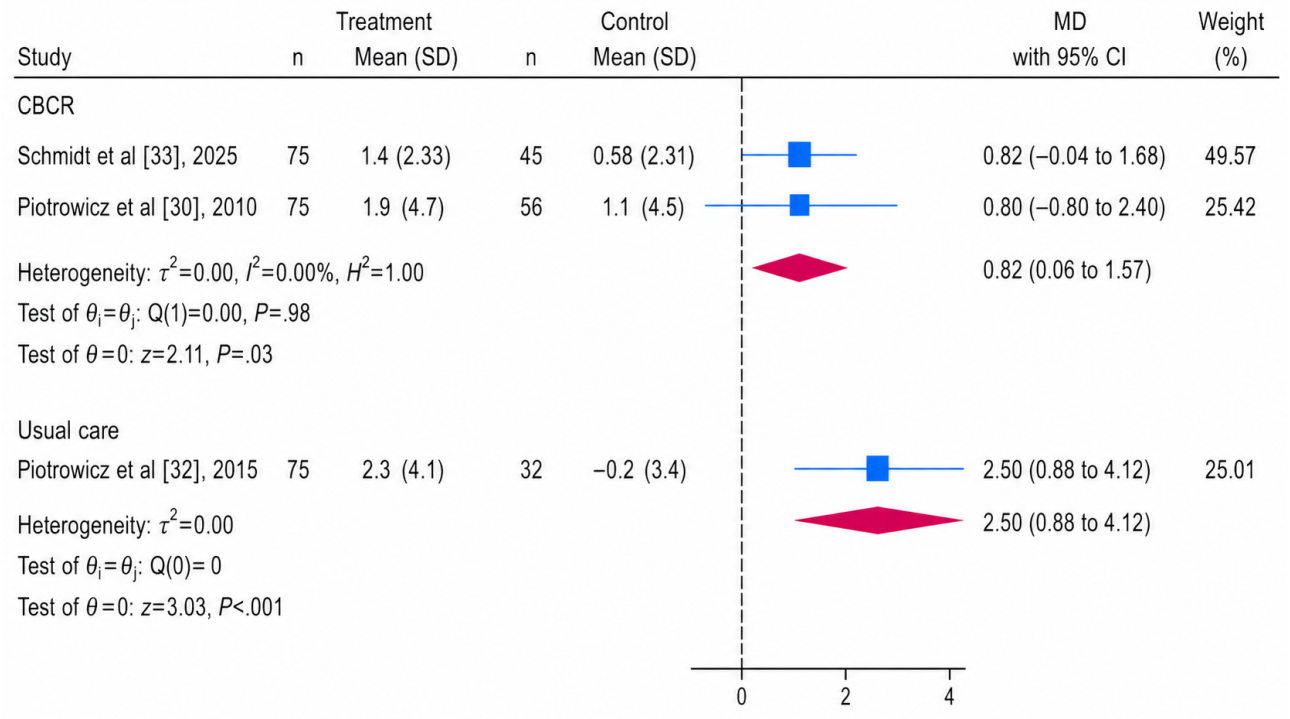
Outcome Measures

Exercise Capacity

Exercise capacity was assessed using VO_2 peak and 6MWD. Because the control groups differed clinically across studies, results were interpreted separately for comparisons with CBCR and usual care. The overall pooled estimates combining CBCR and usual care were not used as the primary basis for interpretation, because CBCR was an active exercise-based intervention, whereas usual care generally did not include structured exercise training. Therefore, we did not interpret any overall pooled effect combining usual care and CBCR, and the results below are presented separately for each comparator.

VO_2 peak was reported in 4 studies [30,32-34]. Three studies were included in the quantitative synthesis [30,32,33]. The study by Lundgren et al [34] was not included in the meta-analysis because postintervention VO_2 peak values were not reported, which precluded the calculation of change-from-baseline values and corresponding effect estimates. This study was, therefore, summarized narratively. In the CBCR subgroup, mHealth-supported RVH-CR showed a statistically significant greater improvement in VO_2 peak than CBCR (MD 0.82, 95% CI 0.06-1.57; P=.03; I^2=0%; Figure 3). One trial comparing mHealth-supported RVH-CR with usual care also showed a greater improvement in VO_2 peak in the intervention group (MD 2.50, 95% CI 0.88-4.12; Figure 3). Because this estimate was derived from a single trial, it should be interpreted as study-specific evidence rather than a pooled effect. It should not be considered evidence of established superiority over CBCR, particularly because the included trials were not designed to test superiority, equivalence, or noninferiority between delivery models.

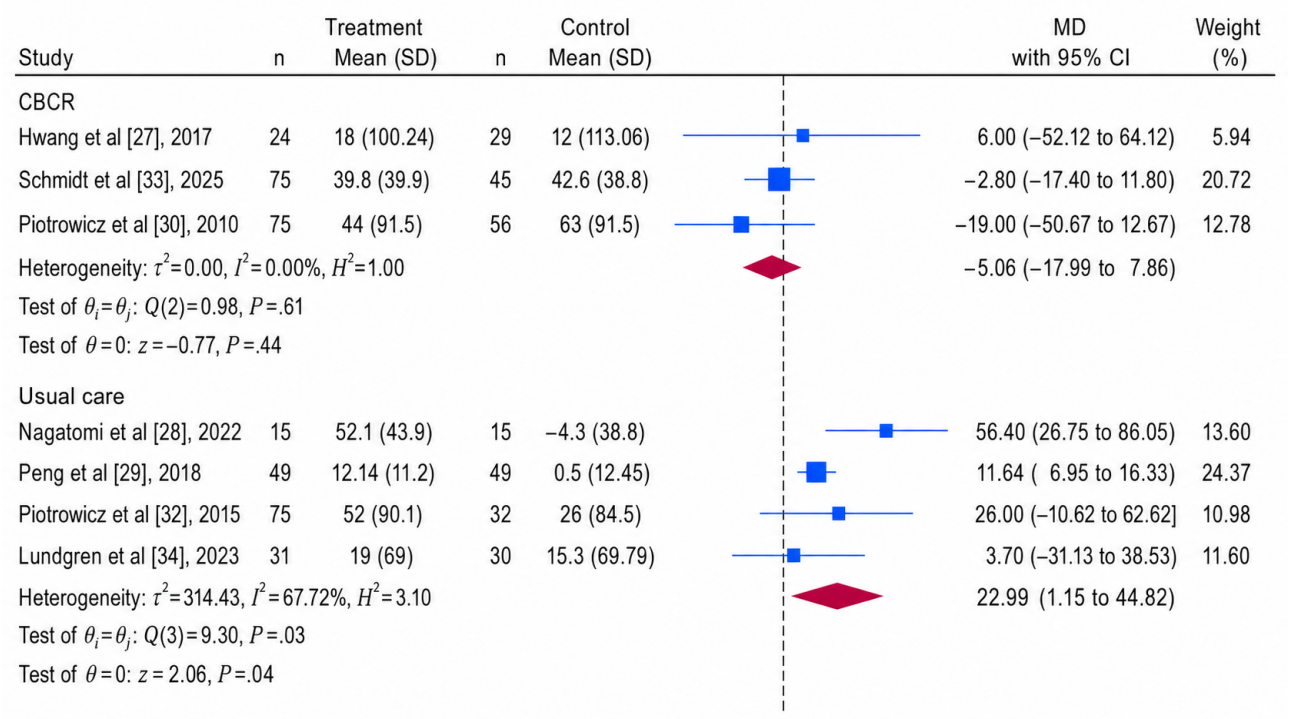
Figure 3. Forest plot showing the comparator-stratified effect of mobile health (mHealth)–supported RVH-CR on VO_2 peak [30,32,33]. CBCR: center-based cardiac rehabilitation; MD: mean difference; RVH-CR: remote, virtual, or hybrid cardiac rehabilitation; VO_2 peak: peak oxygen uptake.



XSL•FO
RenderX

Seven studies reported 6MWD and were included in the updated meta-analysis [27-30,32-34]. The results are shown in Figure 4. In the CBCR subgroup, mHealth-supported RVH-CR was not associated with a statistically significant difference in 6MWD compared with CBCR (MD −5.06, 95% CI −17.99 to 7.86; *P*=.44), with no observed heterogeneity across trials (I^2=0%) [27,30,33]. In the usual care subgroup, mHealth-supported RVH-CR was associated with a significant improvement in 6MWD compared with usual care (MD 22.99, 95% CI 1.15-44.82; *P*=.04) [28,29,32,34]. However, heterogeneity in this subgroup was moderate to substantial (I^2=67.72%), indicating meaningful between-study variability in effect estimates.

Figure 4. Forest plot showing the comparator-stratified effect of mobile health (mHealth)–supported RVH-CR on 6MWD [27-30,32-34]. 6MWD: 6-minute walk distance; CBCR: center-based cardiac rehabilitation; MD: mean difference; RVH-CR: remote, virtual, or hybrid cardiac rehabilitation.



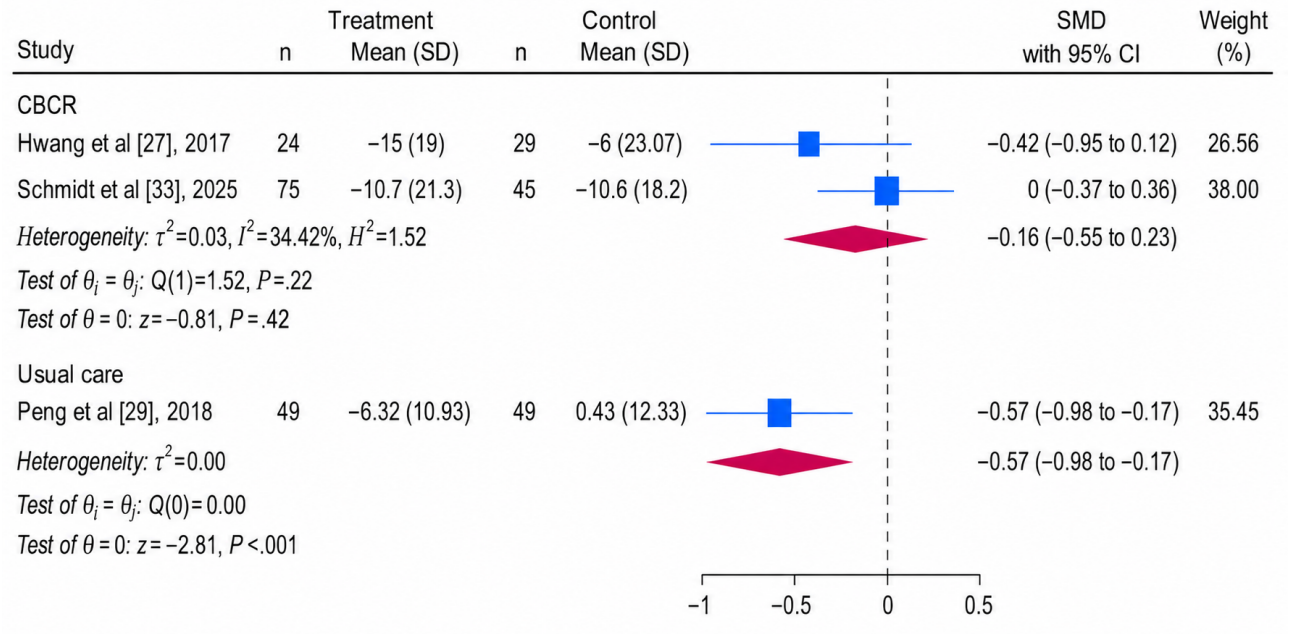
This heterogeneity indicates variability in treatment effects across trials and may reflect differences in patient populations, baseline functional capacity, intervention intensity, exercise frequency, remote monitoring strategies, adherence support, and the content of usual care. Therefore, the pooled 6MWD estimate should be interpreted as the average effect of a structured rehabilitation package supported by mHealth across heterogeneous intervention models, rather than as the isolated effect of mHealth technology itself.

HRQoL

HRQoL was assessed using the MLHFQ and SF-36 [27,29-33]. For all HRQoL analyses, effect directions were coded so that values favoring mHealth-supported RVH-CR indicated better HRQoL. Three studies assessed HRQoL using the MLHFQ [27,29,33]. Higher MLHFQ scores indicate worse HRQoL. The results are shown in Figure 5. In the CBCR subgroup, mHealth-supported RVH-CR showed no significant difference compared with CBCR (SMD −0.16, 95% CI −0.55 to 0.23; *P*=.42). One trial comparing mHealth-supported RVH-CR with usual care showed improved MLHFQ scores in the intervention group (SMD −0.57, 95% CI −0.98 to −0.17; *P*<.01). Because this usual care comparison was based on a single study, the result should be interpreted cautiously.

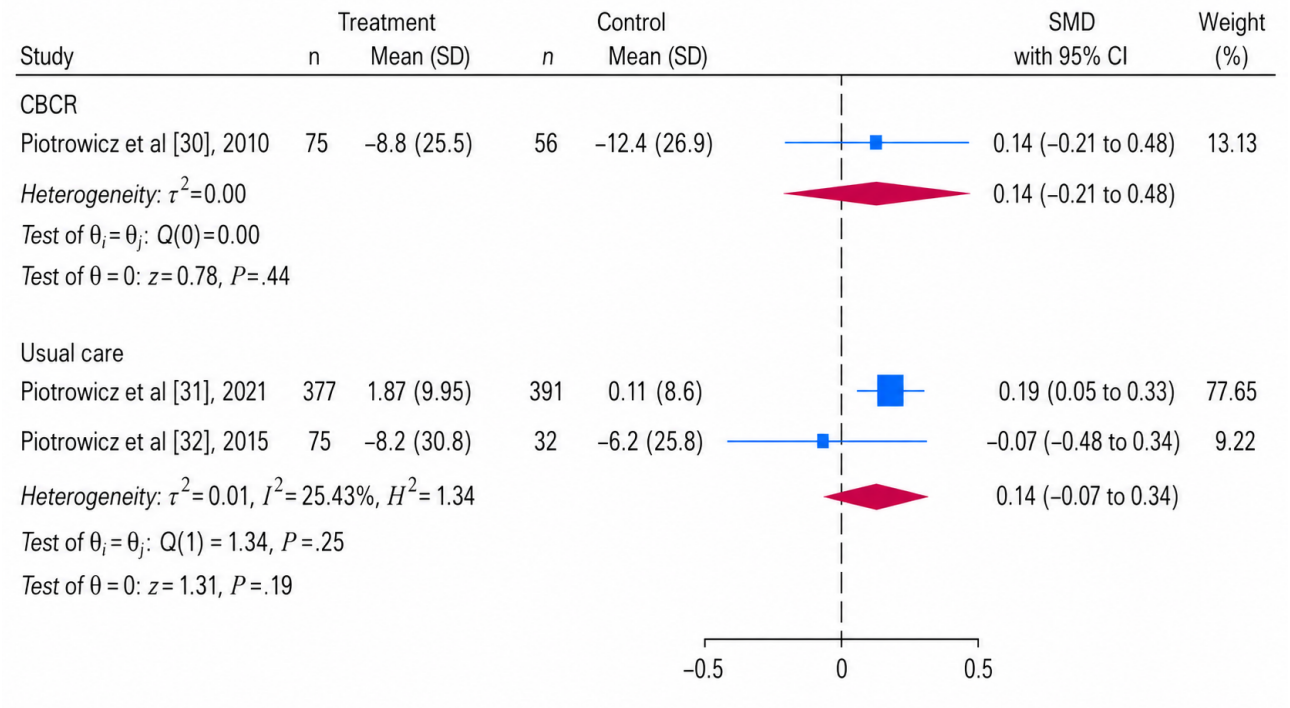
XSL•FO
RenderX

Figure 5. Forest plot showing the comparator-stratified effect of mobile health (mHealth)–supported RVH-CR on the Minnesota Living With Heart Failure Questionnaire [27,29,33]. CBCR: center-based cardiac rehabilitation; RVH-CR: remote, virtual, or hybrid cardiac rehabilitation; SMD: standardized mean difference.



Three studies assessed HRQoL using the SF-36 [30-32]. Higher SF-36 scores indicate better HRQoL. The results are shown in Figure 6. In the CBCR subgroup, 1 study found no significant difference between mHealth-supported RVH-CR and CBCR (SMD 0.14, 95% CI −0.21 to 0.48; *P*=.44) [30]. In the usual care subgroup, 2 studies showed no significant difference between mHealth-supported RVH-CR and usual care (SMD 0.14, 95% CI −0.07 to 0.34; *P*=.19) [31,32]. Therefore, the separated comparator analyses did not support a clear significant effect of mHealth-supported RVH-CR on SF-36 scores. These findings should be interpreted with caution because the SF-36 studies differed in sample size, intervention structure, and rehabilitation components.

Figure 6. Forest plot showing the comparator-stratified effect of mobile health (mHealth)–supported RVH-CR on SF-36 [30-32]. CBCR: center-based cardiac rehabilitation; RVH-CR: remote, virtual, or hybrid cardiac rehabilitation; SF-36: 36-Item Short Form Health Survey; SMD: standardized mean difference.



XSL•FO
RenderX

Mortality, Hospitalization, and Adverse Events

Mortality, hospitalization, and major adverse cardiovascular events were not consistently defined or reported across the included studies, and these outcomes were not suitable for meta-analysis. Available safety data were, therefore, summarized narratively.

Across the 8 included studies, no report explicitly attributed death to the mHealth-supported RVH-CR intervention. Hwang et al [27] reported no deaths, cardiac arrest, syncope, or falls during exercise sessions. Nagatomi et al [28] reported no serious adverse events such as death or cardiac arrest; 3 readmissions occurred in each group and were not considered causally related to the intervention. Peng et al [29] reported no significant complications or adverse outcomes during the program. One study [30] reported no deaths or HF-related hospitalizations attributable to the rehabilitation program, although 1 patient in the standard rehabilitation group died following elective orthotopic heart transplantation. In the TELEREH-HF (Telerehabilitation in Heart Failure Patients) quality-of-life analysis, 2 deaths occurred in the hybrid telerehabilitation group and 2 in the usual care group during the 9-week intervention or observation period [31]. One study [32] reported no deaths or hospitalizations in either group and no cardiovascular implantable electronic device interventions during Nordic walking training. Schmidt et al [33] reported 1 death in the home-based group, which was not related to the exercise training program, and no hospitalizations, musculoskeletal injuries, or other exercise-related adverse events during the 12-week study period. Lundgren et al [34] reported no exercise session-related adverse events during real-time supervised home-based telerehabilitation; 1 death occurred in the control group before the first follow-up.

Overall, no deaths or serious adverse events were explicitly attributed to mHealth-supported RVH-CR in the included trials. However, these findings should be interpreted cautiously because follow-up periods were short, event numbers were small, and clinical end points were inconsistently defined and reported. Therefore, the available evidence is insufficient to establish definitive conclusions regarding all-cause mortality, cardiovascular mortality, hospitalization, major adverse cardiovascular events, or overall intervention safety.

Certainty of Evidence

The certainty of evidence was assessed using the GRADE framework and is summarized in Table 2. Full domain-level judgments are provided in Multimedia Appendix 4. Because usual care and CBCR were clinically distinct comparators, certainty ratings were summarized separately by comparator type. Overall, the certainty was moderate for selected exercise capacity outcomes compared with CBCR and low for most usual-care and HRQoL comparisons. The most common reasons for downgrading were RoB, a small number of studies, imprecision, heterogeneity, and the subjective assessment of HRQoL outcomes.

XSL•FO
RenderX

Table . Summary of findings and certainty of evidence by outcome and comparator.

Outcome	Comparison	Participants, n	Effect estimate	Certainty	Main reasons for downgrading
VO_2 peak[a]	RVH-CR[b] versus usual care	107	MD[c] 2.50 mL/kg/min, 95% CI 0.88 to 4.12	Low	Single-study evidence, risk of bias, and imprecision due to small sample size
VO_2 peak	RVH-CR versus CBCR[d]	251	MD 0.82 mL/kg/min, 95% CI 0.06 to 1.57	Moderate	Limited number of trials and some risk-of-bias concerns
6MWD[e]	RVH-CR versus usual care	296	MD 22.99 m, 95% CI 1.15 to 44.82	Low	Risk-of-bias, heterogeneity, and imprecision
6MWD	RVH-CR versus CBCR	304	MD −5.06 m, 95% CI −17.99 to 7.86	Moderate	Some risk-of-bias concerns and limited number of trials
MLHFQ[f]	RVH-CR versus usual care	98	SMD[g] −0.57, 95% CI −0.98 to −0.17	Low	Single-study evidence, risk of bias, subjective outcome assessment, and imprecision
MLHFQ	RVH-CR versus CBCR	173	SMD −0.16, 95% CI −0.55 to 0.23	Moderate	Subjective outcome assessment and some risk-of-bias concerns
SF-36[h]	RVH-CR versus usual care	875	SMD 0.14, 95% CI −0.07 to 0.34	Low	Subjective outcome assessment and imprecision because the CI included the null effect
SF-36	RVH-CR versus CBCR	131	SMD 0.14, 95% CI −0.21 to 0.48	Low	Single-study evidence, risk of bias, subjective outcome assessment, and imprecision
Mortality and serious adverse events	All comparisons, narrative synthesis	1368	No intervention-related deaths or serious adverse events were reported	Low	Few events, short follow-up, and incomplete or inconsistent event reporting

[a]VO_2 peak: peak oxygen uptake.

[b]RVH-CR: remote, virtual, or hybrid cardiac rehabilitation.

[c]MD: mean difference.

[d]CBCR: center-based cardiac rehabilitation.

[e]6MWD: 6-minute walk distance.

[f]MLHFQ: Minnesota Living with Heart Failure Questionnaire.

[g]SMD: standardized mean difference.

[h]SF-36: 36-Item Short Form Health Survey.

Discussion

Principal Results

This review synthesized evidence from 8 RCTs, including 1368 patients with HFrEF or HFrEF-predominant populations. In comparator-specific analyses, mHealth-supported RVH-CR showed a statistically significant improvement in VO_2 peak compared with CBCR; however, this finding was based on a small number of trials and should be considered preliminary. Compared with usual care, mHealth-supported RVH-CR improved 6MWD, whereas observed changes in VO_2 peak and MLHFQ were based on single-trial estimates and should be interpreted cautiously. No significant differences were observed for 6MWD or MLHFQ in comparisons with CBCR, and comparator-specific analyses did not indicate a clear effect on SF-36 scores. Because the included trials were not designed as noninferiority studies, these results should not be taken as evidence of equivalence or noninferiority. Although selected functional and HRQoL outcomes improved, the certainty of evidence was low to moderate due to small sample sizes, heterogeneity, and reliance on self-reported HRQoL measures. Across the trials, no deaths or serious adverse events were directly attributed to structured RVH-CR supported by mHealth; however, the limited number of events and short follow-up prevent definitive conclusions regarding safety.

The clinical significance of the statistically significant findings should be interpreted cautiously. The improvement in 6MWD

XSL•FO
RenderX

versus usual care was approximately 23 m. In patients with chronic HF, a previous study estimated the minimal clinically important difference for the 6-minute walk test to be 30.1 m [35]. Therefore, the observed effect may be clinically relevant for some patients, although interpretation should consider baseline functional status, HF phenotype, comorbidities, and follow-up duration. The VO_2 peak difference versus CBCR was statistically significant but small in absolute magnitude (MD 0.82 mL/kg/min) and was based on a limited number of trials. Therefore, this finding should be viewed as preliminary and should not be interpreted as evidence of clinically meaningful superiority over CBCR.

mHealth-supported RVH-CR may provide an access-oriented delivery option for structured CR, particularly for patients with geographic, logistical, or resource constraints [36,37]. Remote monitoring, exercise prescription, and telecommunication support may help deliver individualized exercise programs and improve selected functional outcomes, although the evidence for HRQoL remains mixed and outcome-specific [38,39]. Clinicians should recognize that improvements versus usual care reflect both structured exercise and digital support. In this context, mHealth should be viewed as the enabling technology layer supporting CR delivery, whereas RVH-CR describes the care delivery models [16,17,40].

Comparison With Prior Literature

Previous systematic reviews and meta-analyses have studied telerehabilitation or eHealth interventions in patients with HF and other cardiac conditions. Most of these studies combined different types of interventions and patient populations. Li et al [21] evaluated phase II mHealth-supported CR across a broad cardiac population, including myocardial infarction, angina, revascularization, and HF, with interventions combining exercise, education, and remote monitoring. Rampengan et al [41] synthesized 37 RCTs evaluating eHealth interventions, including automated telemonitoring, human-led communication, and structured education, focusing on outcomes such as HF-related hospitalizations, mortality, quality of life, and HF-related knowledge. Although the review demonstrated a reduction in HF-related admissions, it did not specifically assess the effects of structured, exercise-based RVH-CR on functional capacity or exercise tolerance. Similarly, Gao et al [42] and Cavalheiro et al [43] assessed home-based or telerehabilitation programs broadly across cardiovascular populations, reporting improvements in functional capacity and quality of life, but without isolating structured exercise interventions in patients with HF. In contrast, this review focused specifically on patients with HFrEF or HFrEF-predominant HF populations, evaluating structured, exercise-based RVH-CR supported by mHealth. By restricting inclusion to interventions with a clearly defined exercise prescription and a prespecified HFrEF-predominant population, this review isolates the effects of structured exercise delivered outside traditional center-based settings while allowing direct comparison with CBCR.

The present review focused on HFrEF or HFrEF-predominant populations, whereas evidence for mHealth-supported CR in HFpEF remains limited and is beginning to emerge. For example, Subramanian et al [44] recently evaluated a 6-month mHealth-based CR program in older adults with HFpEF and reported improvements in quality of life but no significant improvement in 6MWD or VO_2 peak in the overall cohort.

Strengths and Limitations

This review has several limitations that should be considered when interpreting the findings. First, the small number of included studies (n=8) and the relatively modest sample sizes in individual trials limited statistical power. These constraints also precluded subgroup analyses and meta-regression to explore sources of heterogeneity or potential effect modifiers. This limitation contributed to GRADE downgrading of evidence certainty for most outcomes. Second, several key outcome comparisons, including VO_2 peak versus usual care and MLHFQ versus usual care, relied on single-trial estimates rather than pooled effects. This substantially limits the strength of inference. The statistically significant VO_2 peak finding in the CBCR comparison was based on a small number of trials. Therefore, this result should be interpreted as preliminary and hypothesis-generating rather than as definitive evidence of superiority over CBCR. Third, HRQoL outcomes were mostly assessed using self-reported questionnaires in unblinded participants. This introduces the risk of performance and detection bias and was a principal reason for GRADE downgrading of HRQoL outcomes. Fourth, publication bias could not be formally evaluated using funnel plots or the Egger regression test because fewer than 10 studies were available for any single comparison. Fifth, although all included interventions involved structured exercise-based RVH-CR delivered through mHealth platforms, there was considerable heterogeneity in the specific technologies, monitoring strategies, feedback mechanisms, and rehabilitation protocols used across studies. Sixth, the findings apply to HFrEF or HFrEF-predominant HF populations and should not be generalized to HFpEF, for which mHealth-supported CR evidence is still emerging and requires dedicated synthesis.

Despite these limitations, this review has notable strengths that enhance its clinical and methodological value. Unlike prior systematic reviews that examined telerehabilitation broadly across heterogeneous cardiac populations or combined multiple intervention modalities, this review specifically focused on structured, exercise-based RVH-CR supported by mHealth in patients with HFrEF or HFrEF-predominant populations with HF. This allowed for a more targeted and clinically meaningful synthesis. All included interventions incorporated a clearly defined exercise prescription with patient-generated data capture or clinician-mediated feedback, isolating the effects of structured exercise from general digital health support. RoB was assessed using the Cochrane RoB 2 tool, and evidence certainty was evaluated using the GRADE framework, ensuring methodological transparency and reproducibility. Analyses were stratified by comparator type (CBCR vs usual care) to avoid interpretive confounding arising from pooling structurally distinct control conditions.

Future Research Directions

Future research should prioritize larger, adequately powered RCTs with longer follow-up to assess the durability of

XSL•FO
RenderX

mHealth-supported RVH-CR effects on exercise capacity, quality of life, cardiovascular mortality, and hospitalization in patients with HF. Standardization of mHealth platforms, exercise prescription protocols, and monitoring intensity across trials is essential to reduce heterogeneity and enable more definitive evidence synthesis. Studies should systematically report adherence rates, patient engagement, and intervention fidelity. Greater attention to underrepresented populations, including older adults, women, and individuals from low-income and middle-income settings, is needed to establish generalizability and support equitable implementation of mHealth-supported CR. Future systematic reviews should specifically evaluate mHealth-supported RVH-CR in HFpEF populations.

Conclusions

The effects of structured RVH-CR supported by mHealth varied according to the comparator type in patients with HFrEF or HFrEF-predominant HF populations. Compared with usual care, mHealth-supported RVH-CR was associated with improved 6MWD, whereas single-trial estimates suggested improvements in VO_2 peak and MLHFQ scores. Compared with CBCR, mHealth-supported RVH-CR showed a statistically significant greater improvement in VO_2 peak in a limited number of trials, whereas no statistically significant differences were detected for 6MWD, MLHFQ, or SF-36. Because usual care and CBCR are clinically distinct comparators, these findings should not be interpreted as a single overall effect across all control conditions. No intervention-related deaths or serious adverse events were reported in the included trials, but its safety still needs to be studied. Equivalence or noninferiority to CBCR cannot be inferred from the available evidence. Future studies should evaluate long-term clinical outcomes and standardize intervention protocols.

Acknowledgments

During the preparation of this manuscript, the authors used AI-assisted tools, including ChatGPT (GPT-5.6 Thinking), for language editing and minor graphical and typographical refinements of figures. The AI tools were not used for generating research data, conducting statistical analyses, interpreting results, or formulating conclusions. All AI-assisted outputs were critically reviewed and revised by the authors, who assume full responsibility for the final content of the manuscript.

Funding

This research was funded by the Health Science Research Project of Anhui Province (2024Aa10040) and the Scientific Research Project of Anhui Education Department (2025AHGXZK20212).

Authors' Contributions

KS and HZ were responsible for conceptualization. KS conducted the database search. KS, CL, and TL carried out the paper screening, data extraction, and critical evaluation. KS and TL were responsible for data interpretation. HQ and HZ contributed to and provided suggestions for data interpretation. KS wrote the first draft. HQ and HZ reviewed and edited the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 12 KB - mhealth_v14i1e90422_app1.docx]

Multimedia Appendix 2
Characteristics of the included randomized controlled trials.
[DOCX File, 19 KB - mhealth_v14i1e90422_app2.docx]

Multimedia Appendix 3
Risk of bias tool version 2 summary table.
[DOCX File, 109 KB - mhealth_v14i1e90422_app3.docx]

Multimedia Appendix 4
GRADE (Grading of Recommendations Assessment, Development, and Evaluation) result table.
[DOCX File, 18 KB - mhealth_v14i1e90422_app4.docx]

Checklist 1
PRISMA checklist.

XSL•FO
RenderX

[DOCX File, 273 KB - mhealth_v14i1e90422_app5.docx]

References

1. Mosterd A, Hoes AW. Clinical epidemiology of heart failure. Heart 2007 Sep;93(9):1137-1146. [doi: 10.1136/hrt.2003.025270] [Medline: 17699180]
2. Ran J, Zhou P, Wang J, et al. Global, regional, and national burden of heart failure and its underlying causes, 1990-2021: results from the global burden of disease study 2021. Biomark Res 2025 Jan 23;13(1):16. [doi: 10.1186/s40364-025-00728-8] [Medline: 39849627]
3. Piepoli MF, Conraads V, Corrà U, et al. Exercise training in heart failure: from theory to practice. a consensus document of the Heart Failure Association and the European Association for Cardiovascular Prevention and Rehabilitation. Eur J Heart Fail 2011 Apr;13(4):347-357. [doi: 10.1093/eurjhf/hfr017] [Medline: 21436360]
4. Brown TM, Pack QR, Aberegg E, et al. Core components of cardiac rehabilitation programs: 2024 update: a scientific statement from the American Heart Association and the American Association of Cardiovascular and Pulmonary Rehabilitation. Circulation 2024 Oct 29;150(18):e328-e347. [doi: 10.1161/CIR.0000000000001289] [Medline: 39315436]
5. Kittleson MM. A clinician's guide to the 2022 ACC/AHA/HFSA guideline for the management of heart failure. J Card Fail 2022 May;28(5):831-834. [doi: 10.1016/j.cardfail.2022.03.346] [Medline: 35378258]
6. Murphy SP, Ibrahim NE, Januzzi JL Jr. Heart failure with reduced ejection fraction: a review. JAMA 2020 Aug 4;324(5):488-504. [doi: 10.1001/jama.2020.10262] [Medline: 32749493]
7. Taylor RS, Dalal HM, McDonagh STJ. The role of cardiac rehabilitation in improving cardiovascular outcomes. Nat Rev Cardiol 2022 Mar;19(3):180-194. [doi: 10.1038/s41569-021-00611-7] [Medline: 34531576]
8. Thompson A, Fleischmann KE, Smilowitz NR, et al. 2024 AHA/ACC/ACS/ASNC/HRS/SCA/SCCT/SCMR/SVM guideline for perioperative cardiovascular management for noncardiac surgery: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 2024 Nov 5;150(19):e351-e442. [doi: 10.1161/CIR.0000000000001285] [Medline: 39316661]
9. Thomas RJ, Beatty AL, Beckie TM, et al. Home-based cardiac rehabilitation: a scientific statement from the American Association of Cardiovascular and Pulmonary Rehabilitation, the American Heart Association, and the American College of Cardiology. J Am Coll Cardiol 2019 Jul 9;74(1):133-153. [doi: 10.1016/j.jacc.2019.03.008] [Medline: 31097258]
10. Golbus JR, Lopez-Jimenez F, Barac A, et al. Digital technologies in cardiac rehabilitation: a science advisory from the American Heart Association. Circulation 2023 Jul 4;148(1):95-107. [doi: 10.1161/CIR.0000000000001150] [Medline: 37272365]
11. Long L, Mordi IR, Bridges C, et al. Exercise-based cardiac rehabilitation for adults with heart failure. Cochrane Database Syst Rev 2019 Jan 29;1(1):CD003331. [doi: 10.1002/14651858.CD003331.pub5] [Medline: 30695817]
12. Mochari H, Lee JR, Kligfield P, Mosca L. Ethnic differences in barriers and referral to cardiac rehabilitation among women hospitalized with coronary heart disease. Prev Cardiol 2006;9(1):8-13. [doi: 10.1111/j.1520-037x.2005.3703.x] [Medline: 16407697]
13. Ruano-Ravina A, Pena-Gil C, Abu-Assi E, et al. Participation and adherence to cardiac rehabilitation programs. A systematic review. Int J Cardiol 2016 Nov 15;223:436-443. [doi: 10.1016/j.ijcard.2016.08.120] [Medline: 27557484]
14. Beatty AL, Beckie TM, Dodson J, et al. A new era in cardiac rehabilitation delivery: research gaps, questions, strategies, and priorities. Circulation 2023 Jan 17;147(3):254-266. [doi: 10.1161/CIRCULATIONAHA.122.061046] [Medline: 36649394]
15. Stefanakis M, Batalik L, Antoniou V, Pepera G. Safety of home-based cardiac rehabilitation: a systematic review. Heart Lung 2022;55:117-126. [doi: 10.1016/j.hrtlng.2022.04.016] [Medline: 35533492]
16. Hughes JW, Berry R, Brown TM, et al. Consensus statement on the virtual and remote delivery of cardiac and pulmonary rehabilitation and their components. J Cardiopulm Rehabil Prev 2025;45(6):387-396. [doi: 10.1097/HCR.0000000000001002] [Medline: 41171981]
17. Beatty AL, Brown TM, Corbett M, et al. Million Hearts cardiac rehabilitation think tank: accelerating new care models. Circ Cardiovasc Qual Outcomes 2021 Oct;14(10):e008215. [doi: 10.1161/CIRCOUTCOMES.121.008215] [Medline: 34587751]
18. Ozemek C, Berry R, Bonikowske AR, German C, Gavic AM. What has cardiac rehabilitation looked like in the COVID-19 pandemic: lessons learned for the future. Prog Cardiovasc Dis 2023;76(20-4):20-24. [doi: 10.1016/j.pcad.2023.01.001] [Medline: 36690287]
19. Wongvibulsin S, Habeos EE, Huynh PP, et al. Digital health interventions for cardiac rehabilitation: systematic literature review. J Med Internet Res 2021 Feb 8;23(2):e18773. [doi: 10.2196/18773] [Medline: 33555259]
20. Beatty AL, Fukuoka Y, Whooley MA. Using mobile technology for cardiac rehabilitation: a review and framework for development and evaluation. J Am Heart Assoc 2013 Nov 1;2(6):e000568. [doi: 10.1161/JAHA.113.000568] [Medline: 24185949]
21. Li L, Ringeval M, Wagner G, Paré G, Ozemek C, Kitsiou S. Effectiveness of home-based cardiac rehabilitation interventions delivered via mHealth technologies: a systematic review and meta-analysis. Lancet Digit Health 2025 Apr;7(4):e238-e254. [doi: 10.1016/j.landig.2025.01.011] [Medline: 40023729]

XSL•FO
RenderX

22. Molloy CD, Long L, Mordi IR, et al. Exercise-based cardiac rehabilitation for adults with heart failure—2023 Cochrane systematic review and meta-analysis. Eur J Heart Fail 2023 Dec;25(12):2263-2273. [doi: 10.1002/ejhf.3046] [Medline: 37850321]
23. Yamamoto S, Okamura M, Akashi YJ, et al. Impact of long-term exercise-based cardiac rehabilitation in patients with chronic heart failure—a systematic review and meta-analysis. Circ J 2024 Aug 23;88(9):1360-1371. [doi: 10.1253/circj.CJ-23-0820] [Medline: 38220206]
24. Hua C, Huang W, Chen Z, et al. Effects of exercise based cardiac rehabilitation delivery modes on chronic heart failure: a systematic review and network meta-analysis. Sci Rep 2024 Dec 28;14(1):31246. [doi: 10.1038/s41598-024-82608-z] [Medline: 39732888]
25. Taylor RS, Walker S, Smart NA, et al. Impact of exercise rehabilitation on exercise capacity and quality-of-life in heart failure: individual participant meta-analysis. J Am Coll Cardiol 2019 Apr 2;73(12):1430-1443. [doi: 10.1016/j.jacc.2018.12.072] [Medline: 30922474]
26. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100. [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
27. Hwang R, Bruning J, Morris NR, Mandrusiak A, Russell T. Home-based telerehabilitation is not inferior to a centre-based program in patients with chronic heart failure: a randomised trial. J Physiother 2017 Apr;63(2):101-107. [doi: 10.1016/j.jphys.2017.02.017] [Medline: 28336297]
28. Nagatomi Y, Ide T, Higuchi T, et al. Home-based cardiac rehabilitation using information and communication technology for heart failure patients with frailty. ESC Heart Fail 2022 Aug;9(4):2407-2418. [doi: 10.1002/ehf2.13934] [Medline: 35534907]
29. Peng X, Su Y, Hu Z, et al. Home-based telehealth exercise training program in Chinese patients with heart failure: a randomized controlled trial. Medicine (Baltimore) 2018 Aug;97(35):e12069. [doi: 10.1097/MD.0000000000012069] [Medline: 30170422]
30. Piotrowicz E, Baranowski R, Bilinska M, et al. A new model of home-based telemonitored cardiac rehabilitation in patients with heart failure: effectiveness, quality of life, and adherence. Eur J Heart Fail 2010 Feb;12(2):164-171. [doi: 10.1093/eurjhf/hfp181] [Medline: 20042423]
31. Piotrowicz E, Mierzyńska A, Banach M, et al. Quality of life in heart failure patients undergoing hybrid comprehensive telerehabilitation versus usual care—results of the Telerehabilitation in Heart Failure Patients (TELEREH-HF) randomized clinical trial. Arch Med Sci 2021;17(6):1599-1612. [doi: 10.5114/aoms.2020.98350] [Medline: 34900039]
32. Piotrowicz E, Zieliński T, Bodalski R, et al. Home-based telemonitored Nordic walking training is well accepted, safe, effective and has high adherence among heart failure patients, including those with cardiovascular implantable electronic devices: a randomised controlled study. Eur J Prev Cardiol 2015 Nov;22(11):1368-1377. [doi: 10.1177/2047487314551537] [Medline: 25261268]
33. Schmidt C, Magalhães S, Gois Basilio P, et al. Center- vs home-based cardiac rehabilitation in patients with heart failure: EXIT-HF randomized controlled trial. JACC Heart Fail 2025 May;13(5):695-706. [doi: 10.1016/j.jchf.2024.09.024] [Medline: 39895436]
34. Lundgren KM, Langlo KAR, Salvesen Ø, et al. Feasibility of telerehabilitation for heart failure patients inaccessible for outpatient rehabilitation. ESC Heart Fail 2023 Aug;10(4):2406-2417. [doi: 10.1002/ehf2.14405] [Medline: 37221704]
35. Shoemaker MJ, Curtis AB, Vangsnes E, Dickinson MG. Clinically meaningful change estimates for the six-minute walk test and daily activity in individuals with chronic heart failure. Cardiopulm Phys Ther J 2013 Sep;24(3):21-29. [Medline: 23997688]
36. Nso N, Nassar M, Mbome Y, et al. Comparative assessment of the long-term efficacy of home-based versus center-based cardiac rehabilitation. Cureus 2022 Mar;14(3):e23485. [doi: 10.7759/cureus.23485] [Medline: 35475109]
37. Alqahtani AS. Telehealth-based cardiac rehabilitation for heart failure: a systematic review of effectiveness, access, and patient-centred outcome. Medicina (Kaunas) 2025 Dec 23;62(1):25. [doi: 10.3390/medicina62010025] [Medline: 41597311]
38. Kikuchi A, Taniguchi T, Nakamoto K, et al. Feasibility of home-based cardiac rehabilitation using an integrated telerehabilitation platform in elderly patients with heart failure: a pilot study. J Cardiol 2021 Jul;78(1):66-71. [doi: 10.1016/j.jjcc.2021.01.010] [Medline: 33579602]
39. Bilbrey T, Martin J, Zhou W, et al. A dual-modality home-based cardiac rehabilitation program for adults with cardiovascular disease: single-arm remote clinical trial. JMIR mHealth uHealth 2024 Oct 1;12:e59098. [doi: 10.2196/59098] [Medline: 39150858]
40. Rechcińska A, Bralewska B, Mordaka M, Rechciński T. Digital technologies in cardiac rehabilitation for high-risk cardiovascular patients: a narrative review of mobile health, virtual reality, exergaming and virtual education. J Clin Med 2026 Feb 3;15(3):1193. [doi: 10.3390/jcm15031193] [Medline: 41682870]
41. Rampengan D, Surya S, Saisa D, et al. eHealth-supported self-management and telemonitoring in heart failure: a systematic review and meta-analysis of randomized controlled trials. Heart Int 2025;19(2):35-47. [doi: 10.17925/HI.2025.19.2.6] [Medline: 41377645]

XSL•FO
RenderX

42. Gao Y, Wang N, Zhang L, Liu N. Effectiveness of home-based cardiac telerehabilitation in patients with heart failure: a systematic review and meta-analysis of randomised controlled trials. J Clin Nurs 2023 Nov;32(21-22):7661-7676. [doi: 10.1111/jocn.16726] [Medline: 37062028]
43. Cavalheiro AH, Silva Cardoso J, Rocha A, Moreira E, Azevedo LF. Effectiveness of tele-rehabilitation programs in heart failure: a systematic review and meta-analysis. Health Serv Insights 2021;14:11786329211021668. [doi: 10.1177/11786329211021668] [Medline: 34188484]
44. Subramanian V, Betts T, Segar MW, et al. m-Health based cardiac rehabilitation program for management of older patients with HFpEF: a pilot RCT. JACC Heart Fail 2025 Apr;13(4):653-656. [doi: 10.1016/j.jchf.2024.10.024] [Medline: 39918533]

Abbreviations

6MWD: 6-minute walk distance
CBCR: center-based cardiac rehabilitation
CR: cardiac rehabilitation
GRADE: Grading of Recommendations Assessment, Development, and Evaluation
HF: heart failure
HFpEF: heart failure with preserved ejection fraction
HFrEF: heart failure with reduced ejection fraction
HRQoL: health-related quality of life
MD: mean difference
mHealth: mobile health
MLHFQ: Minnesota Living with Heart Failure Questionnaire
VO_2 peak: peak oxygen uptake
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RoB 2: risk of bias tool version 2
RVH-CR: remote, virtual, or hybrid cardiac rehabilitation
SF-36: 36-Item Short Form Health Survey
SMD: standardized mean difference
TELEREH-HF: Telerehabilitation in Heart Failure Patients

Edited by M Dorsch; submitted 29.Dec.2025; peer-reviewed by K Kitagaki, SC Surya, TAK Manne; revised version received 14.Jun.2026; accepted 18.Jun.2026; published 21.Jul.2026.

Please cite as:
Shao K, Liu C, Li T, Qu H, Zhou H
Effects of Remote, Virtual, or Hybrid Cardiac Rehabilitation Supported by mHealth in Patients With Heart Failure: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2026;14:e90422
URL: https://mhealth.jmir.org/2026/1/e90422
doi:10.2196/90422

© Kaidong Shao, Chunqiu Liu, Tianshu Li, Huiyan Qu, Hua Zhou. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 21.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mechanisms of Engagement With Mobile Health Apps for Adults With Long-Term Conditions: Overview of Systematic Reviews

Jeni Baykoca, MSc; Goretti Hurtado Barbeito, MSc; Christina Joanne Pearce, PhD; Madison Milne-Ives, PhD; Joanna Hudson, PhD; Sam Norton, PhD; Rona Moss-Morris, PhD

Department of Health Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, Guy's Hospital, Bermondsey Wing, 5th Floor, Health Psychology Section, London, England, United Kingdom

Corresponding Author:
Jeni Baykoca, MSc
Department of Health Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, Guy's Hospital, Bermondsey Wing, 5th Floor, Health Psychology Section, London, England, United Kingdom

Abstract

Background: Engagement is a necessary precondition for the effectiveness of mobile health (mHealth) apps for long-term physical health conditions (LTCs), particularly as health systems increasingly prioritize the deployment of scalable, self-guided digital interventions. Outside controlled research settings, where clinician involvement often drives engagement, little is known about whether, how, and why people engage with mHealth apps based on intrinsic motivation alone. Existing systematic reviews have cataloged behavioral engagement indicators but rarely assess the mechanisms underlying engagement.

Objective: This overview of systematic reviews aimed to (1) synthesize evidence on how engagement with mHealth apps for LTCs is defined, measured, and associated with health outcomes, (2) explore intrinsic and extrinsic motivational processes underlying engagement, and (3) provide practical guidance for developing scalable, user-centered digital health interventions that sustain sufficient engagement with minimal reliance on external drivers. Uniquely, we interpreted modifiable barriers and facilitators through a motivational lens that distinguishes extrinsic from intrinsic motives, mapping intrinsic motives onto autonomy, competence, and relatedness, as proposed by Self-Determination Theory (SDT).

Methods: Searches of MEDLINE, Web of Science, Epistemonikos, and gray literature (inception to June 9, 2025) identified systematic reviews reporting engagement indicators, engagement-outcome associations, or barriers and facilitators among adults with LTCs. Quantitative reviews were appraised using AMSTAR 2 (A Measurement Tool to Assess Systematic Reviews 2), and qualitative and mixed methods reviews were appraised using CASP (Critical Appraisal Skills Programme). A narrative synthesis was undertaken, and modifiable barriers and facilitators were independently mapped by 2 reviewers to extrinsic and intrinsic motivation and, for intrinsic factors, to the SDT constructs of autonomy, competence, and relatedness. Discrepancies were resolved through discussion with the wider research team.

Results: Nineteen reviews (12 quantitative, 5 mixed methods, and 2 qualitative) were included from 4684 records. Fourteen (74%) reviews did not define engagement, and the remaining 5 equated it with “usage” or “adherence,” precluding meta-analysis. Fourteen reviews reported microlevel behavioral indicators, but none captured macrolevel or effective engagement. Eight assessed engagement-outcome links; 7 reported positive associations, and 1 reported no effect. Seven reviews included nonmodifiable factors that influence engagement (eg, ethnicity), while 13 included modifiable factors. SDT mapping revealed that modifiable factors influencing autonomy (eg, personal relevance, flexibility), competence (eg, usability, technical support), and relatedness (eg, clinician endorsement, peer connection) underpin intrinsic engagement, whereas extrinsic barriers include restrictions to access (including cost).

Conclusions: Current evidence on engagement with mHealth apps remains conceptually inconsistent and methodologically fragmented, but motivational patterns are clear: engagement depends largely on intrinsic motives once external conditions are satisfied. Applying SDT provides the first mechanism-oriented explanation of how engagement operates, enabling practical recommendations for evaluating existing apps and designing future mHealth interventions that support autonomy, competence, and relatedness.

Trial Registration: PROSPERO CRD42024604784; https://tinyurl.com/57prxa3d

(JMIR Mhealth Uhealth 2026;14:e88382) doi:10.2196/88382

KEYWORDS

mHealth; mobile apps; chronic disease; motivation; digital health; user-centered design; health behavior; self-management; telemedicine

XSL•FO
RenderX

Introduction

Long-term physical health conditions (LTCs) represent a major and growing global challenge for health care systems, driven by their growing prevalence, increasing likelihood of multimorbidity, and rising costs of long-term care [1]. Digital health interventions offer scalable, flexible, and cost-effective potential to enhance LTC self-management and support patient-centered care, particularly in the context of limited health care resources and workforce constraints.

Reflecting this shift, digital health innovation has become a core policy priority. For example, the National Health Service 10 Year Health Plan for England emphasizes the role of digital technologies in empowering patients, improving access to care, and integrating physical and mental health services for people with LTCs [2]. Mobile health (mHealth) apps have garnered particular attention due to their accessibility, portability, and ability to deliver personalized interventions directly to users in real time [3-5]. With the increasing use of smartphones worldwide, mHealth apps can reach diverse populations across geographic and socioeconomic barriers [6]. mHealth apps also support multiple functions relevant to LTC self-management, including symptom monitoring, education, behavior change support, and communication with health care professionals. However, their effectiveness is partly constrained by engagement [7-9].

Engagement is necessary for effectiveness: an mHealth app that is not initiated, is insufficiently used, or is quickly abandoned cannot exert its intended effects [10]. Engagement with digital interventions is multidimensional, comprising behavioral indicators (eg, frequency and depth of use) and subjective indicators, including cognitive (perceived relevance and goal alignment) and affective (eg, interest, enjoyment, or emotional response) components [10,11]. Engagement also operates at multiple levels: the microlevel, capturing moment-to-moment interactions with the intervention (eg, logging in, completing a task), and the macrolevel (users' internalizing and integrating the intervention's goals into daily life) [11,12]. Importantly, effective engagement is not about engagement volume. Rather, it concerns the extent and type of engagement necessary to achieve desired outcomes, acknowledging that sustained use may not always be required if the user has internalized the intervention's purpose [11].

Although engagement has been described in detail using the aforementioned concepts [10-12], systematic reviews show inconsistent definitions and operationalization of engagement, impeding the synthesis of findings and making it difficult to understand if and why app use is sustained and health promoting [13,14]. As a result, there is limited consensus on how engagement should be conceptualized, how it is initiated and sustained over time, and, critically, how much engagement is "enough." This lack of clarity has contributed to widespread uncertainty about how to design mHealth apps that people will actually use, even when those apps are evidence-based and demonstrably effective under controlled conditions.

A key source of this confusion may be the limited attention paid to the motivational mechanisms underlying engagement. Engagement is often treated as being driven by features such as reminders or clinician oversight [8,11]. While such externally regulated strategies may increase short-term use, they are resource-intensive and difficult to scale, particularly in health care systems seeking to reduce clinician workload and maximize patient self-management [15]. In contrast, engagement is more likely to be sustained when individuals are internally motivated to use an app because they perceive it as personally valuable, relevant, and supportive of their goals [16]. Understanding intrinsic motives is therefore critical for realizing the scalability promise of mHealth app interventions for LTCs.

Self-Determination Theory (SDT) provides a well-established and empirically supported framework for understanding how and why people initiate and maintain behaviors [17]. SDT distinguishes between extrinsic motivation (behavior driven by external demands or rewards) and intrinsic motivation (behavior that is self-endorsed and aligned with personal values). According to SDT, intrinsic motivation is supported when three basic psychological needs—autonomy, competence, and relatedness—are satisfied. This framework is particularly relevant to mHealth apps, where achieving effective engagement often depends on users internalizing the value of app use rather than responding to extrinsic prompts alone [16,18]. Mapping app features and contextual factors onto SDT constructs offers a theoretically grounded, mechanistic understanding of how app design and features (eg, personalization, feedback, social features, and integration with care pathways) may support or undermine engagement.

Although numerous systematic reviews have reported engagement with mHealth apps for LTCs, few have synthesized findings within a coherent motivational framework [19,20]. Existing reviews frequently describe barriers and facilitators of engagement but stop short of explaining why these factors matter or how they relate to intrinsic vs extrinsic motives. This has restricted their ability to inform intervention design, evaluation, and policy, resulting in mixed findings and limited cumulative insight.

To address these gaps, this overview of systematic reviews aimed to (1) synthesize evidence on how engagement with mHealth apps for LTCs is defined, measured, and associated with health outcomes, (2) explore intrinsic and extrinsic motivational processes underlying engagement, and (3) provide practical guidance for developing scalable, user-centered digital health interventions that support effective engagement with minimal reliance on ongoing clinician input. By doing so, this overview integrates existing reviews to move beyond descriptive cataloging of "what" predicts engagement toward a theoretically grounded explanation of "how" and "why" commonly reported factors may influence it.

Methods

Ethical Considerations

This study did not involve human participants and therefore did not require ethical approval, in accordance with institutional and national guidelines.

XSL•FO
RenderX

Design

This overview of systematic reviews adhered to the PRIOR (Preferred Reporting Items for Overviews of Reviews) guidelines (Checklist 1) and was prospectively registered in the PROSPERO (International Prospective Register of Systematic Reviews; CRD42024604784) [21].

Deviations From Protocol

Three deviations were made to the registered protocol (PROSPERO CRD42024604784) prior to data extraction and synthesis to improve coherence and practical use:

1. Objective 2 in the PROSPERO registration (identify specific mHealth app-based intervention features that are associated with differing levels of engagement) was removed due to insufficient data in the included reviews.
2. A section on nonmodifiable factors influencing engagement (eg, age, sex, and ethnicity) was added to contextualize barriers and facilitators.
3. Modifiable barriers and facilitators of engagement were classified by motivational type (intrinsic or extrinsic), with intrinsically motivating indicators further mapped onto the basic psychological needs of SDT (autonomy, competence, and relatedness). This was done to propose plausible mechanisms through which app features may influence engagement.

Search Strategy

A comprehensive search was conducted in MEDLINE, Web of Science, Epistemonikos, and gray literature sources (Social Science Research Network, WorldCat, Health Management Information Consortium [Ovid]) from inception to June 9, 2025. Search terms combined keywords and MeSH related to "chronic conditions" and "mobile." No date or language limits were applied initially. The strategy was developed with domain experts and 2 senior librarians (Multimedia Appendix 1).

Eligibility Criteria

Systematic reviews were eligible if they

- included adults (≥18 y) with at least one LTC or persistent physical symptoms (eg, chronic pain or fatigue). For the purposes of this review, LTCs were defined as diagnosed physical health conditions that are persistent, require ongoing management, or have long-term consequences for health and functioning
- evaluated mHealth apps delivered via mobile phones or tablets
- reported behavioral engagement indicators, engagement-outcome associations, or barriers or facilitators to engagement

Reviews were excluded if they did not distinguish app-based interventions from other digital tools or focused solely on passive monitoring (eg, wearables). Reviews were therefore excluded when the population was defined solely by hypertension, obesity, isolated symptoms, exposures, risk factors, or health indicators, rather than by a diagnosed long-term physical condition within the scope of the review. This was based on World Health Organization (WHO) and Global Burden of Disease risk-factor frameworks reporting raised blood pressure (hypertension) and high BMI (overweight and obesity) as metabolic risk factors or health indicators rather than eligible long-term physical condition populations in their own right [22,23]. Search terms were intentionally broad to maximize sensitivity and capture related risk-factor literature; however, final inclusion was determined during screening based on how each review defined and reported the condition.

Study Selection

Search results were imported into EndNote X7 for deduplication [24]. Two reviewers independently screened titles and abstracts using Covidence, an online software tool to manage screening, extraction, and quality assessment for systematic reviews [25]. During screening, inclusion was limited to English, German, Spanish, or Turkish (the languages fluently spoken by the research team) to ensure accurate appraisal. This allowed us to note how many potentially relevant reviews were excluded for language while maintaining screening reliability. Full texts were retrieved for all records deemed potentially relevant by at least 1 reviewer. Ten percent of the full-text reviews were randomly selected for double screening to assess interrater reliability, yielding a Cohen κ of 0.62, which indicates substantial agreement between authors [26]. Discrepancies were resolved through discussion, and the remaining full-text screening was completed by the first author. Reference lists of included reviews were hand searched.

Data Extraction

Data extraction used a standardized form piloted by the research team. The first author extracted all data, with a random 10% independently checked by a second reviewer (GHB), yielding full agreement. The remaining data extraction was completed by the first author.

Quality Appraisal

Quantitative systematic reviews were appraised using the AMSTAR 2 (A Measurement Tool to Assess Systematic Reviews 2) tool and categorized as high, moderate, low, or critically low [27]. Qualitative and mixed methods reviews were appraised using the CASP (Critical Appraisal Skills Programme) qualitative checklist [28]. CASP does not provide a numeric scoring system and instead uses descriptive summaries with justifications. We considered the breadth of "yes" responses, with particular emphasis on core methodological areas, including clarity of aims, appropriateness of qualitative methodology and research design, adequacy of data collection, rigor of data analysis, clarity of findings, and the stated value of the research. We did not apply a prespecified numerical cutoff for most CASP items. Instead, reviews were judged as higher quality when "yes" responses were consistent across these core domains and when any limitations were minor or related to less applicable domains for systematic reviews, such as recruitment strategy. A random 10% were double checked, with full agreement. We considered alternative tools for quality appraisal, including the Joanna Briggs Institute (JBI) checklist for umbrella reviews. However, because our overview included qualitative and mixed methods reviews, we judged CASP to provide more explicit prompts for assessing the interpretive and methodological features of qualitative synthesis. CASP was therefore used as

XSL•FO
RenderX

a structured framework to appraise the qualitative synthesis components of the included reviews.

Data Synthesis

A narrative synthesis was undertaken. Reviews were grouped according to the engagement-related data they reported:

1. Behavioral indicators of engagement (amount, duration, depth, and frequency) [10,11]
2. Associations between engagement and health outcomes
3. Nonmodifiable barriers and facilitators of engagement (eg, age and ethnicity)
4. Modifiable barriers and facilitators of engagement (eg, digital literacy and app design)

Where definitions were absent, engagement-related terms (eg, usage, adherence, and compliance) were interpreted using the context from primary studies. For consistency in characterizing core mHealth app functionalities, we applied the WHO classification of digital health interventions [29].

Synthesis of modifiable barriers and facilitators offered insight into users' cognitive and affective responses to mHealth apps, reflecting how perceptions (eg, relevance and trust) and emotions (eg, interest and burden) influenced decisions to engage or disengage. This enabled exploration of subjective influences on engagement.

Modifiable barriers and facilitators of engagement were categorized as intrinsic or extrinsic motives. The intrinsic motives were then systematically mapped onto the 3 basic psychological needs proposed by SDT—autonomy, competence, and relatedness—using established definitions of SDT psychological needs [17].

- Autonomy: experiences of volition, choice, and personal relevance
- Competence: perceptions of efficacy, capability, and mastery
- Relatedness: feelings of connection, being cared for, and a sense of belonging with others

Mapping was completed independently by 2 reviewers (JB and MMI). Decisions about which barriers and facilitators aligned with each SDT need were informed by existing theoretical and empirical literature [30,31] and refined through iterative discussions within the research team.

The frequency with which themes were mapped to each psychological need was descriptively recorded across reviews and their included primary studies to enhance the transparency of the synthesis rather than to infer effect size or relative importance.

Although the Capability, Opportunity, Motivation-Behavior (COM-B) model was initially considered as an organizing framework [32], preliminary synthesis showed that most reviews described engagement processes in terms of the quality of motivation rather than broader behavioral determinants, such as physical capability or environmental opportunity. Common descriptors included perceived personal relevance and feelings of being supported, which align more closely with SDT's distinction between intrinsic and extrinsic motivation. Given that sustaining engagement with mHealth apps typically relies on intrinsic, self-endorsed motivation rather than externally regulated behavior [17], SDT provided a more coherent and mechanistic framework for interpreting engagement processes.

Primary study overlap across the included systematic reviews was assessed using a citation matrix. Rows represented unique primary studies, and columns represented included systematic reviews. Overlap was quantified descriptively by identifying primary studies that appeared in more than 1 review and by calculating the corrected covered area (CCA) [33]. As this overview of systematic reviews synthesized data at the systematic review level, primary study overlap was used to inform transparency and interpretation only; no primary study-level outcome data were extracted, reanalyzed, or pooled.

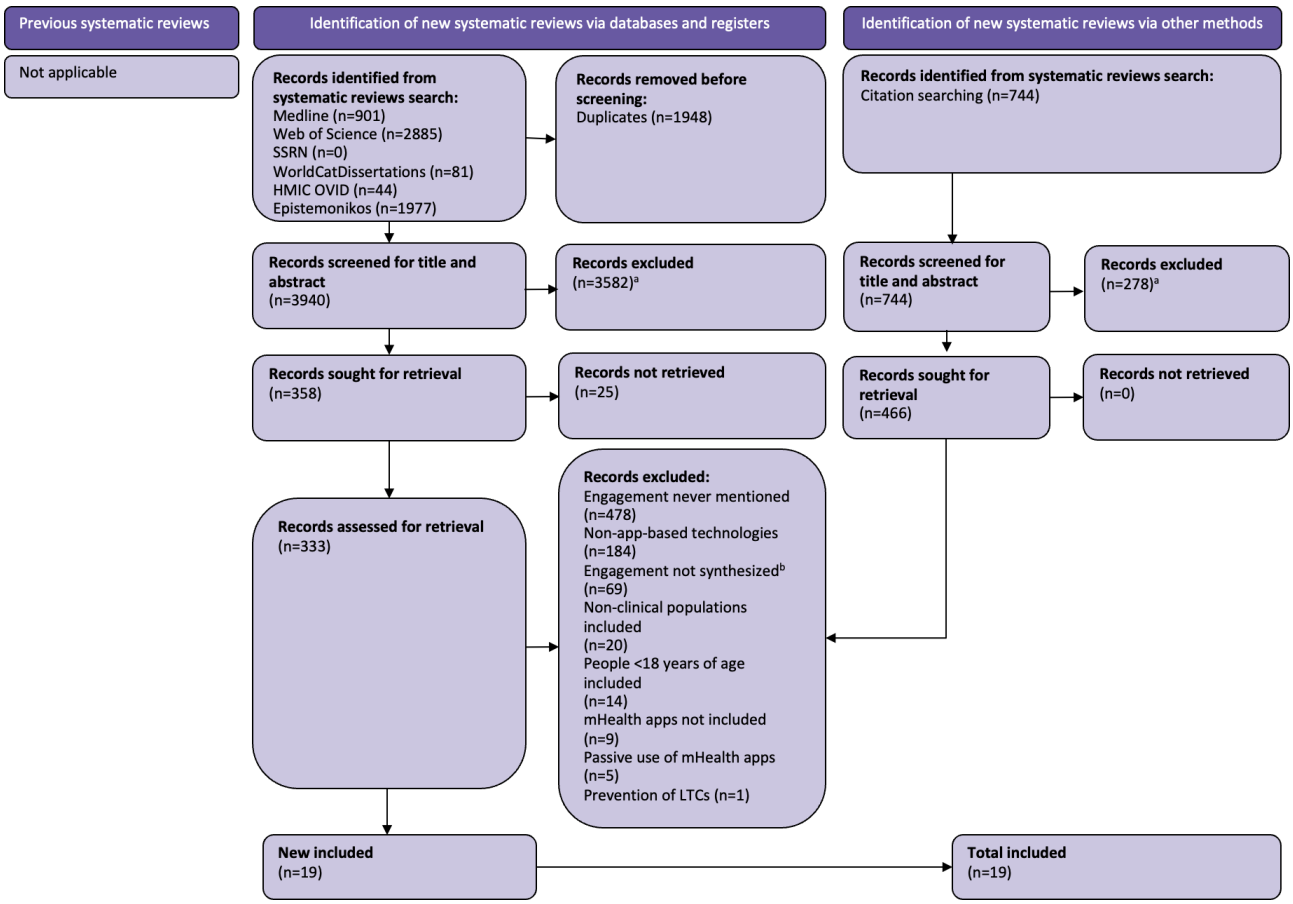
Results

Search Results

The search identified 5888 records from databases and registers and 744 through citation searching. After removing duplicates, 4684 records including 3940 (84.1%) from databases and registers and 744 (15.9%) from other methods. Overall, 799 full-text articles were assessed for eligibility by 2 independent reviewers, of which 333 (41.7%) were identified through databases and registers and 466 (58.3%) through other methods. The primary reason for exclusion (n=478, 59.9%) was the absence of any mention or reporting of engagement-related data. Ultimately, 19 systematic reviews were included in this overview of systematic reviews [34-52]. A PRIOR (Preferred Reporting Items for Overviews of Reviews) flow diagram is presented in Figure 1.

XSL•FO
RenderX

Figure 1. PRIOR (Preferred Reporting Items for Overviews of Reviews) flow diagram of the study selection process. [a]Reviews were of long-term risk factors (obesity, hypertension, etc) and/or transitory conditions (eg, gestational diabetes mellitus) (n=197) and/or searched app stores only (n=45). [b]The primary articles included measured engagement, but this was not synthesized in the review. SSRN: Social Science Research Network.



Characteristics of Included Reviews

The 19 systematic reviews comprised 12 (63%) quantitative, 5 (26%) mixed methods, and 2 (11%) qualitative syntheses published between 2015 and 2025. Of the 17 reviews with quantitative data, 3 (18%) conducted a meta-analysis. Of the 7 reviews with qualitative data, 1 (14%) undertook a formal meta-synthesis, while the remainder used thematic or narrative approaches. LTC populations included people living with cancer (n=6, 32%), diabetes (n=4, 21%), and chronic pain (n=2, 11%). Others focused on heart failure (n=1, 5%), stroke (n=1, 5%), chronic kidney disease (n=1, 5%), osteoarthritis (n=1, 5%), Parkinson disease (n=1, 5%), traumatic brain injury (n=1, 5%), and a mixed group of LTCs (n=1, 5%).

The number of primary studies per review ranged from 3 to 30 (total n=262), with sample sizes ranging from 1 to 3606 and mean participant ages typically between 50 and 63 years. Across the 262 primary studies, the three most common study designs were randomized controlled trials (RCTs; n=111, 42.4%), qualitative studies (n=62, 24%), and observational studies (n=33, 13%). Of the 195 primary studies that reported geographic origin, 83 (43%) were conducted in North America, 68 (35%) in Europe, 24 (12%) in Asia, 12 (6%) in Oceania, 4 (2%) in South America, and 4 (2%) in Africa. Table 1 summarizes these characteristics, and Multimedia Appendix 2 provides a more detailed table.

XSL•FO
RenderX

Table . Characteristics of included systematic reviews.[a]

Author, year	Systematic review characteristics			Primary studies' characteristics					
	Review type	Quality of reviews	Publication year range	Primary studies	Type of LTC[b]	Study designs for primary studies	Country of studies	Sample size, total (range)	Age (y), mean (range)
Alaslawi et al [36], 2022	Mixed methods	Moderate[c]	2015 - 2019	28	Diabetes	Qualitative (n=14), cross-sectional (n=12), cohort (n=1), mixed methods (n=1)	United States (n=10), Canada (n=3), United Kingdom (n=3), Australia (n=2), Saudi Arabia (n=2), Germany (n=2), Peru (n=1), Denmark (n=1), Rwanda (n=1), New Zealand (n=1), Norway (n=1), China (n=1)	Not reported	Not reported
Bezerra Giordan et al [45], 2022	Mixed methods	Moderate-to-high[c]	2012 - 2022	28	Heart failure	RCTs[d] (n=10), quasi-experimental (n=8), qualitative (n=10)	United States (n=15), Canada (n=4)	1397 (5-232)	Mean: 63.4
Campbell and Porter [39], 2015	Quantitative	Critically low[e]	2003 - 2013	5	Chronic kidney disease stages 3 - 5, including dialysis patients	RCTs (n=2), case studies/case reports (n=3)	United States (n=5)	60 (1 - 44)	Range: 50.3 - 70
de Melo Santana et al [40], 2023	Quantitative with meta-analysis	Low[e]	2018 - 2022	5	Chronic low back pain	RCTs (n=5)	Germany (n=2), Jordan (n=1), Denmark (n=1), Norway (n=1), India (n=1)	Total: 447	Range: 18 - 65
Diez Alvarez et al [47], 2024	Quantitative	Low[e]	2015 - 2020	3	Diabetes	Qualitative (n=1), interventional (n=1), observational (n=1)	Not specified	Range: 4 - 60	Not reported
Dunham et al [46], 2021	Mixed methods	Moderate[c]	2013 - 2020	10	Osteoarthritis with or without chronic pain	RCT (n=1), quasi-experimental (n=1), mixed methods (n=2), qualitative (n=6)	United States, United Kingdom, Australia	Range: 18 - 738	Not reported

XSL•FO
RenderX

Author, year	Systematic review characteristics			Primary studies' characteristics					
	Review type	Quality of reviews	Publication year range	Primary studies	Type of LTC[b]	Study designs for primary studies	Country of studies	Sample size, total (range)	Age (y), mean (range)
Frid et al [41], 2024	Quantitative	Moderate[e]	2015 - 2023	30	Breast cancer	RCT (n=25), Q-RCT (n=1), quasi-experimental (n=3), observational (n=1)	Not specified	3606 (35-490)	Not reported
He et al [42], 2022	Quantitative with meta-analysis	Critically low[e]	2011 - 2020	19	Type 2 diabetes	Multicenter RCT (n=9), others not specified	South Korea, Australia, China, Canada, Mexico, Netherlands, Norway, India, United States, Japan	2585 (54 - 247)	52.7 (31.7 - 68)
Hernandez Silva et al [43], 2019	Quantitative	Critically low[e]	2008 - 2017	7	Cancer (breast cancer [n=4], lung cancer [n=2], colorectal cancer [n=2], prostate cancer [n=1], lymphoma [n=1])	RCTs (n=2), quasi-randomized (n=1), non-randomized study with control group (n=1), single-arm studies (n=3)	United States (n=3), United Kingdom (n=2), Korea (n=1), Sweden (n=1)	Range: 16 - 356	58.5 (50.3 - 69)
Horn et al [51], 2025	Quantitative	Low[e]	2005 - 2023	11	Breast cancer	RCTs (n=11)	Netherlands (n=6), Japan (n=1), Turkey (n=1), Germany (n=1), United States (n=2)	Total: 2249	Range: 43.9 - 56.2
Lee et al [34], 2022	Quantitative	Low[e]	2013 - 2020	17	Parkinson disease	Observational (n=12), quasi-experimental (n=2), RCTs (n=3)	United States (n=7), England (n=2), Finland (n=2), Italy (n=2), Netherlands (n=2), United Kingdom (n=2), Australia (n=1), Belgium (n=1), Greece (n=1), Israel (n=1), Scotland (n=1)	Total: 1,246	63.02 (34-84)
MacLean et al [52], 2025	Quantitative	Low[e]	2016 - 2022	8	Chronic pain	RCTs (n=6), single-arm trials (n=2)	United States (n=4), Spain (n=1), Germany (n=1), Brazil (n=1), Australia (n=1)	Range: 20 - 206	50 (18–85)

XSL•FO
RenderX

Author, year	Systematic review characteristics			Primary studies' characteristics					
	Review type	Quality of reviews	Publication year range	Primary studies	Type of LTC[b]	Study designs for primary studies	Country of studies	Sample size, total (range)	Age (y), mean (range)
Magalhães et al [48], 2021	Quantitative	Critically low[e]	2007 - 2019	10	Cancer	Prospective intervention studies (n=6), randomized control studies (n=4) (pilot study [n=1], multicentric clinical trial [n=1])	United Kingdom, Switzerland, South Korea, China	Total: 616	Not reported
O'Neill et al [38], 2022	Qualitative	High[c]	2013 - 2019	14	Type 2 diabetes	Mixed methods (n=5), qual element of an RCT (n=3), 1 - 1 qualitative (n=5), focus group (n=1)	Norway (n=2), United States (n=5), Germany (n=1), Australia (n=3), United Kingdom (n=2), Canada (n=1)	Total: 248	Range: 24 - 80
Patail et al [50], 2025	Qualitative	High[c]	2018 - 2023	24	Diabetes	Face-to-face interviews (n=21), telephone interviews (n=2), electronic survey with open-ended questions (n=1)	Asia (n=4), Europe (n=5), United Kingdom (n=2), United States (n=8), Canada (n=1), Central Africa (n=1), multinational (Singapore and Germany) (n=1)	Not reported	Range: 23 - 81
Patterson et al [44], 2021	Mixed methods with meta-analysis	Moderate-to-high[c]	2015 - 2020	19	Coronary heart disease (n=10), hypertension (n=4), stroke (n=3), heart failure (n=1), peripheral artery disease (n=1)	RCT (n=10), non-RCT (n=3), cohort (n=6)	United States (n=10), Australia (n=1), Spain (n=2), Sweden (n=1), Norway (n=1), Israel (n=1), Scotland (n=1), Multicenter RCT (Spain, Germany, United Kingdom) (n=1), China (n=1), Germany (n=1)	Total: 1,543	59.7 (46.3 - 69)

XSL•FO
RenderX

Author, year	Systematic review characteristics			Primary studies' characteristics					
	Review type	Quality of reviews	Publication year range	Primary studies	Type of LTC[b]	Study designs for primary studies	Country of studies	Sample size, total (range)	Age (y), mean (range)
Rintala et al [35], 2023	Quantitative	Low[e]	2021 - 2022	11	Stroke survivors (chronic, subacute, and mixed stages)	Controlled clinical trials (n=7; RCT [n=5], non-RCT [n=2]), uncontrolled clinical trials (n=4)	Europe (n=5), United Kingdom (n=1), Spain (n=1), Netherlands (n=1), Israel (n=1), Asia (n=3), South Korea (n=1), Philippines (n=1), North America (n=1), United States (n=1), South America (n=1), Chile (n=1), Africa (n=1), Ghana (n=1)	Total: 264	Median: 59.3 (IQR: 55.3 - 61.0)
Vaezipour et al [37], 2019	Mixed methods	High[c]	2015 - 2017	4	Moderate-severe traumatic brain injury	Before-after design with preinterview, postinterview, and 2-month postintervention follow-up (n=1), participatory design approach for system development and presystem and postsystem evaluation (n=1), repeated-measures design (n=1), online survey (n=1)	Denmark (n=1), Canada (n=1), United States (n=1), Australia (n=1)	Total: 204	Range: 21 - 60
Whitehead and Seaton [49], 2016	Quantitative	Critically low[e]	2008 - 2014	9	Diabetes (n=5), chronic lung disease (n=3), cardiovascular disease (n=1)	RCTs (n=9)	Europe (n=3), Oceania (n=2), Asia (n=3), United States (n=1)	Range: 48 - 288	Range: 33.8 - 72.1

[a]If the systematic review included data from a population other than adults with long-term conditions, and/or interventions other than mobile health apps, only the data relevant to the overview were extracted.

[b]LTC: long-term condition.

[c]Quality of qualitative and mixed methods systematic reviews was rated using the CASP (Critical Appraisal Skills Programme) tool.

[d]RCT: randomized controlled trial.

[e]Quality of quantitative systematic reviews was rated using the AMSTAR 2 (A Measurement Tool to Assess Systematic Reviews 2) tool.

XSL•FO
RenderX

Quality Appraisal

Qualitative and mixed methods reviews were strong overall, with 3 rated "high quality" [37,38,50], 2 rated "moderate-to-high" [44,45], and 2 rated "moderate" [36,46], though reflexivity and analytic transparency were often limited (Table 1).

Quantitative reviews were generally low quality: 6 were rated "critically low" [39,40,42,43,48,49], 5 were rated "low" [34,47,51,52], and 1 was rated "moderate" [41]. None met the criteria for high quality, mainly due to missing protocol registration, limited risk-of-bias assessment, and insufficient reporting of heterogeneity or funding (Table 1). Multimedia Appendices 3 and 4 include the quality appraisal scores for AMSTAR 2 and CASP, respectively.

Overlap of Primary Studies

Across the 19 included systematic reviews, 262 unique primary studies were identified. Primary study overlap was slight [33]. Most primary studies appeared in only one review (n=243, 92.7%), while 19 studies appeared in more than one review (19/262, 7.3%). Of these overlapping studies, 14 appeared in 2 reviews, 4 appeared in 3 reviews, and 1 appeared in 4 reviews. The total number of primary study occurrences across reviews was 287. The CCA was 0.53%, indicating slight overlap. Given the low level of overlap and because synthesis was conducted at the systematic review level, all 19 reviews were retained in the narrative synthesis.

Description of mHealth App-Based Interventions

Across the 19 reviews, 98 distinct mHealth apps were reported. The most common app functionality, using the WHO classifications of digital health interventions, was symptom self-monitoring, which was present in 17 (90%) systematic reviews. Targeted delivery of health information, often as educational content tailored to users' LTCs, was identified in 16 (84%) reviews. Alerts and reminders were reported in 12 (63%) reviews. Active user data capture, primarily for symptom tracking and remote consultations with health care providers, appeared in 8 (42%) reviews. Peer group support features were reported in 5 (26%), while user-to-system feedback mechanisms were present in 2 (11%) reviews. Only 1 (5%) review included apps where users could access their personal health records. Full characteristics are provided in Multimedia Appendix 5.

Behavioral Engagement Indicators

For clarity, we describe behavioral indicators as microlevel usage metrics (eg, amount, duration, depth, and frequency of use) based on previous literature [10-12].

Of the 19 included systematic reviews, only 5 (26%) explicitly operationalized engagement a priori [36-38,46,50]. Fourteen (74%) reviews reported engagement metrics, but predominantly as microlevel indicators (eg, frequency of logins, duration of use, and number of tasks completed). Terms such as "usage," "adherence," and "engagement" were often used interchangeably, with no review operationalizing macrolevel engagement (eg, integration into daily life and alignment with goals). Definitions and thresholds varied widely, precluding meta-analysis. Multimedia Appendix 6 provides a detailed summary of the engagement indicators reported in each review.

Engagement-Health Outcome Associations

Eight of the 19 (42%) reviews assessed links between engagement and health outcomes. Seven of these reviews (moderate-to-high [n=1], moderate [n=1], low [n=4], and critically low [n=1]) reported positive associations between microlevel app use and outcomes such as improved symptom knowledge, health behavior change, or reduced distress [39,41,45,47,48,51,52]. One meta-analysis (critically low quality) found no significant association [40].

Importantly, no review linked macrolevel engagement with health outcomes or if engagement was sufficient to achieve intended health outcomes (ie, effective engagement) [11]. Qualitative reviews provided insight into subjective engagement indicators, such as users' sense of motivation or identification with the intervention, but no review integrated these with quantitative outcome data. Behavioral and subjective indicators of engagement were reported descriptively in separate strands in the included reviews, with no further analysis to link them.

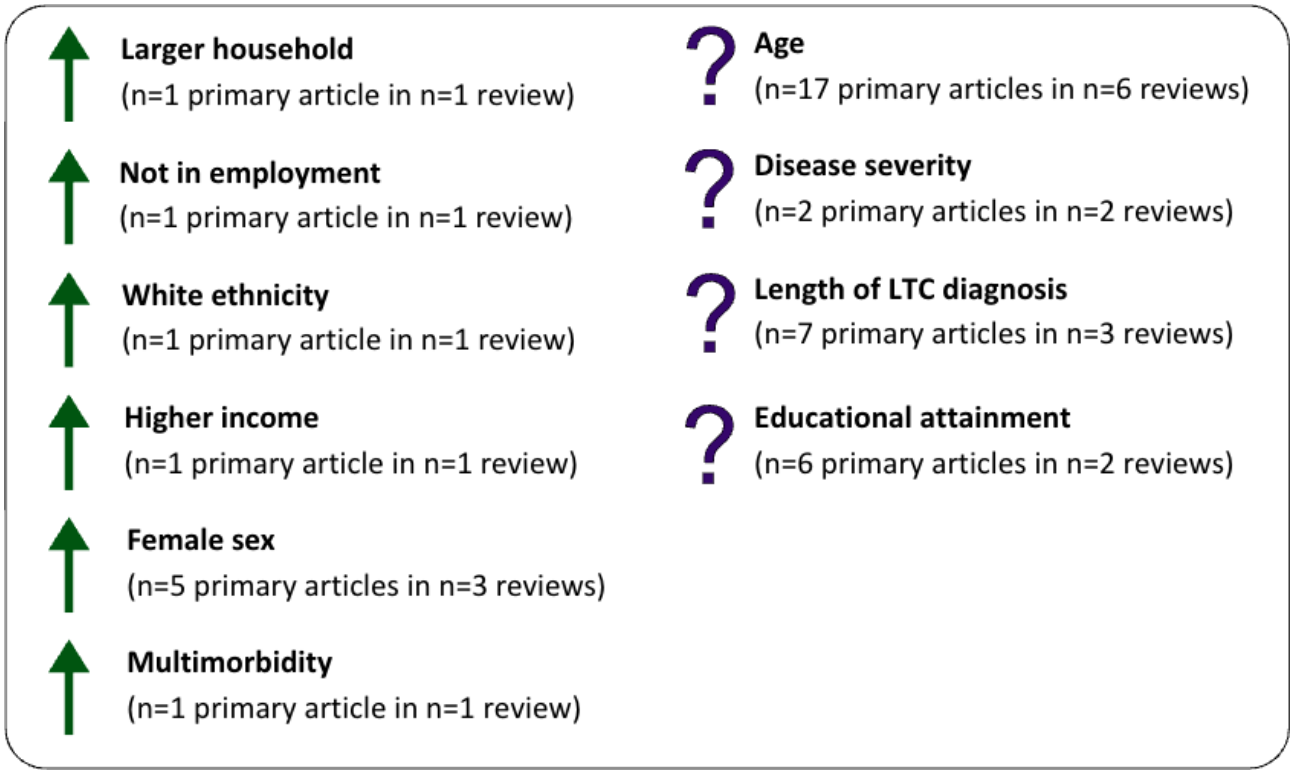
Nonmodifiable Barriers and Facilitators of Engagement

Seven of the 19 (37%) systematic reviews highlighted nonmodifiable user-specific barriers and facilitators of engagement [36,41,42,45,46,48,52] (Figure 2). A total of 57% (4/7) of reviews were quantitative (moderate [n=1], low [n=1], and critically low quality [n=2]), and 43% (3/7) of reviews were mixed methods (moderate-to-high [n=1] and moderate quality [n=2]).

Findings were mixed, as 2 reviews reported higher engagement among older adults (≥63 years in 1 review; age not specified in the other) [42,46], whereas another reported greater engagement among younger users [36]. Similarly, disease-related characteristics such as duration of diagnosis had mixed results, with both newly diagnosed individuals and those living with a condition for over 5 years demonstrating active engagement across different studies [36,42]. Findings related to disease severity or multimorbidity were also mixed, with one review reporting lower engagement [41] and another reporting higher engagement, potentially due to increased pain burden [40].

XSL•FO
RenderX

Figure 2. Nonmodifiable barriers and facilitators of engagement with mobile health apps reported in 37% (7/19) of systematic reviews. Upward arrows indicate higher engagement, and question marks indicate mixed findings. LTC: long-term condition.



Modifiable Barriers and Facilitators of Engagement

Thirteen of the 19 (68%) reviews reported modifiable barriers and facilitators of engagement with mHealth apps for LTCs [34,36-38,42-46,48-50,52]. Seven reviews were quantitative (moderate [n=1], low [n=2], and critically low quality [n=4]), 4 were mixed methods reviews (moderate-to-high [n=2] and moderate quality [n=2]), and 2 were qualitative reviews (high quality [n=2]).

Reviews were categorized into 9 overarching themes (Figure 3). Six themes mapped onto intrinsic motives: app design, app functionality, trust, time burden, baseline technological readiness, and baseline readiness for health behavior change; and 3 themes mapped onto extrinsic motives: awareness of the app, availability of the digital tool, and cost. The reported extrinsic motives function primarily as structural preconditions for engagement because they enable initial uptake but do not, by themselves, sustain engagement. A full thematic breakdown with illustrative quotations and frequency of reporting is provided in Multimedia Appendix 7.

For intrinsic motivation, barriers and facilitators supporting competence were the most frequently reported, appearing across 63% (12/19) of systematic reviews and 89.3% (234/262) of distinct primary studies. These included app design and functionality features that enabled users to feel capable of using the app effectively, such as ease of use, clear and accessible language, visual aids, real-time feedback, and symptom monitoring. Engagement was undermined when competence was challenged by technical problems, high data entry burden, information overload, unmet expectations, low digital literacy, or concerns that app use could exacerbate psychological distress related to the LTC.

Barriers and facilitators mapping onto autonomy were reported across 58% (11/19) of systematic reviews and 59.5% (156/262) of distinct primary studies. Autonomy-related barriers and facilitators reflected users' appraisal of relevance, control, and fit with daily life, including readiness for behavior change, perceived need for the app, ability to integrate app use into daily routines, goal-setting functionality, personalization, and privacy considerations.

Factors supporting relatedness were less frequently reported overall, with 53% (10/19) of systematic reviews and 40.5% (106/262) of primary studies, and focused on interpersonal connection and trust. Engagement facilitators included interaction with health care professionals and other users, endorsement by clinicians, secure data sharing with caregivers during emergencies, and integration with electronic medical records. Barriers to relatedness included lack of emotional connection to the app and stigmatizing or overly intrusive reminders.

The mapping of modifiable barriers and facilitators to intrinsic and extrinsic motivation informed the development of design recommendations for future evidence-based mHealth apps for people with LTCs (Table 2). Design implications were generated for all themes identified in the synthesis, irrespective of how frequently they were reported across primary studies. This approach reflects the conceptual aim of this overview of systematic reviews to understand mechanisms of engagement rather than to rank predictors by prevalence. Less frequently reported themes may still represent decisive barriers or facilitators for different user groups, LTCs, or contexts of use, and excluding them would risk overlooking design considerations essential for equitable, scalable, and effective engagement.

XSL•FO
RenderX

Figure 3. Motivational mechanisms of engagement with mobile health apps. EMR: electronic medical record; HCP: health care professional; mHealth: mobile health.

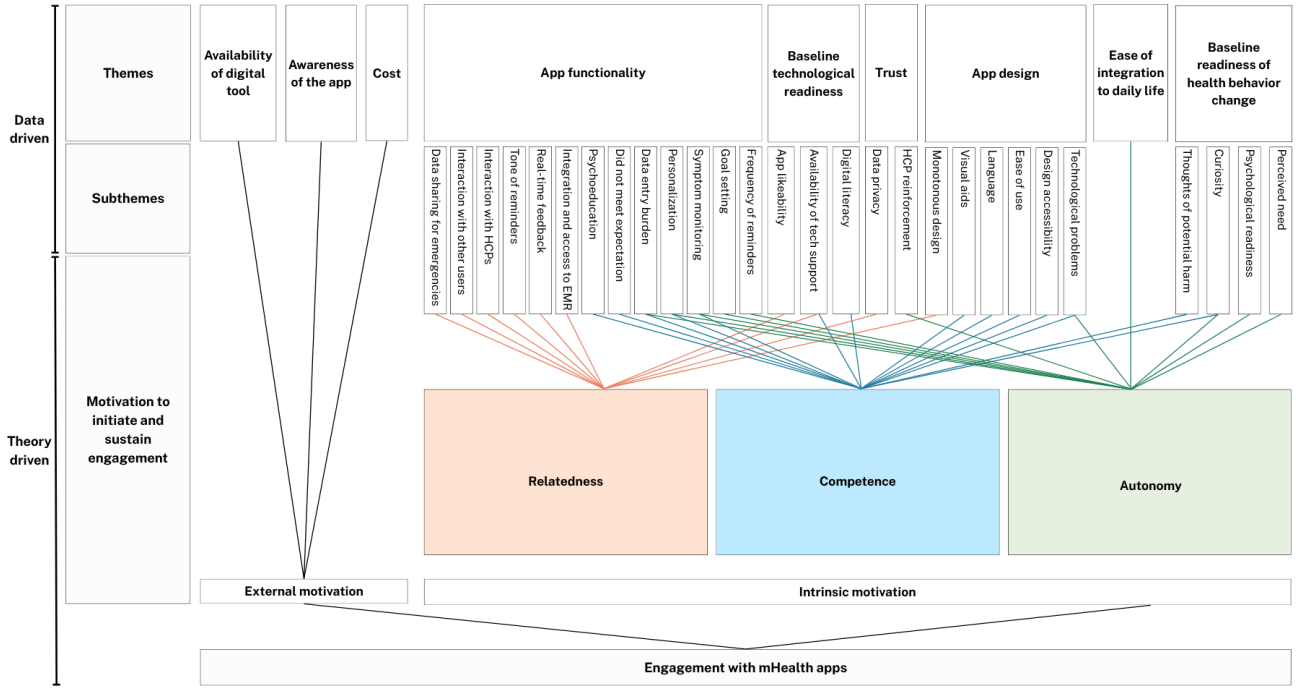


Table . Recommendations for mobile health app–based intervention development that support intrinsic and extrinsic motives.

Motivational category	Key engagement factors identified	Design implications for mHealth[a] apps
Intrinsic motivation		
Autonomy	• Limited perceived need for app use • Low psychological readiness for behavior change • Time burden and poor fit with daily routines • Perceived loss of control due to reminders or data entry demands • Privacy concerns	• Emphasize personal relevance during on-boarding • Use nonpressuring language • Reduce burden through automation • Enable flexible goal setting • Provide customizable reminders • Ensure transparency over data use • Offer tailored features that support user-led engagement
Competence	• Technical problems • Poor usability or accessibility • Information overload • Low digital literacy • Unmet expectations • Anxiety arising from symptom monitoring	• Use an accessible and intuitive design • Clear and plain-language communication • Visual information delivery • Emotionally supportive and timely feedback • Accurate expectation-setting • Ongoing technical support to accommodate varying literacy and skill levels
Relatedness	• Limited trust in app content • Absence of clinical endorsement • Lack of social or emotional connection • Stigmatizing or impersonal reminder tone	• Embed clinician endorsement • Enable optional peer interaction • Support secure data sharing with trusted others • Integrate with clinical systems • Use supportive, nonstigmatizing communication in reminders and pop-up notifications
Extrinsic motivation	• Lack of awareness of apps • Limited digital access or connectivity • Financial barriers	• Integrate app into routine care pathways • Design for low-connectivity environments (eg, offline functionality, lightweight data use) • Ensure core features are freely available and explore reimbursement, subsidy, or health system–supported funding models

[a]mHealth: mobile health.

XSL•FO
RenderX

Discussion

Principal Findings

To our knowledge, this overview of systematic reviews is the first to offer theory-informed, mechanism-oriented guidance for understanding and designing mHealth app–based interventions that support engagement among adults with LTCs. By mapping modifiable barriers and facilitators of engagement onto SDT, we move beyond descriptive lists of factors to explain why these influences matter and how they may support or undermine intrinsic motivation.

This approach aligns closely with the Medical Research Council framework for developing and evaluating complex interventions, which emphasizes the value of theory for specifying mechanisms of action, strengthening intervention design, and enhancing reproducibility. Applying SDT as an overarching motivational theory provides a coherent structure for identifying modifiable targets, linking app design and features to psychological processes that shape engagement, and treating engagement as an intentional, theorized component of an intervention rather than a by-product. The resulting recommendations (Table 2) therefore offer a workable, mechanism-based framework for integrating psychological needs into both the development and evaluation of mHealth interventions.

Across the 19 included systematic reviews, three consistent findings emerged: (1) conceptual ambiguity in engagement definitions, (2) limited and low-quality evidence on engagement-outcome associations, and (3) converging evidence on patterns in modifiable factors influencing engagement.

First, engagement was rarely defined explicitly. Reviews often used terms including “engagement,” “usage,” and “adherence” interchangeably and reported only microlevel behavioral indicators (eg, number of logins or tasks completed). No review operationalized macrolevel engagement (eg, sustained integration into daily life or alignment with goals). Consequently, the current evidence base largely reflects observed usage, limiting insight into whether engagement was clinically relevant to the person using the app.

Second, although 7 of the 8 reviews assessing engagement-outcome links reported positive associations, these findings were drawn from heterogeneous operationalizations of engagement and were often based on critically low-quality quantitative syntheses. Qualitative reviews contributed insight into subjective indicators of engagement, but no review integrated them with behavioral indicators. This methodological divide reflects broader measurement immaturity in engagement research rather than limitations of individual reviews.

Third, when modifiable barriers and facilitators to engagement were mapped to motivational processes, the vast majority were intrinsic—related to autonomy, competence, or relatedness—rather than extrinsic or structural. Most barriers and facilitators reflected whether users experienced the app as personally relevant, manageable, intuitive, trustworthy, and aligned with their own goals and capabilities. Only a small subset of factors was purely extrinsic (eg, awareness, connectivity, and cost), which is still prominent as this is the minimum needed to enable engagement. However, once extrinsic preconditions are met (eg, users can download the app), engagement largely depends on intrinsic motivational processes. Hence, mHealth apps that aim to replace or reduce clinician input must intentionally support these psychological needs.

SDT provided a clear mechanism through which to interpret these patterns, enabling us to identify specific motivational targets and derive design recommendations to guide the development of interventions that are more capable of sustaining engagement without ongoing clinician reinforcement.

Comparison With Prior Work

Our findings align with longstanding critiques that digital health research often emphasizes “how much” people engage through behavioral indicators while neglecting the quality of engagement and/or the “why” (ie, subjective [cognitive and affective] influences on engagement) [53-55]. Persistent heterogeneity in engagement terminology and measurement constrains the development of cumulative evidence and makes it difficult to establish whether disengagement reflects unmet needs of the user, symptom improvement, or success in achieving the intervention’s goal [10,56-58]. Without distinguishing between these possibilities, establishing cutoffs such as the “minimum effective dose” of engagement also remains elusive.

The positive engagement-outcome associations, such as improved symptom knowledge, health behavior change, or reduced distress, mirror earlier evidence, including Donkin et al [7] findings, but these associations remain difficult to interpret given inconsistent operationalizations and low review quality. Similarly, although persuasive design features (eg, tailoring and feedback) have been shown to improve engagement [8], none of the included reviews in this overview of systematic reviews analyzed user-level engagement or linked design features to specific mechanisms. Collectively, these findings highlight a need for standardized behavioral metrics and improved reporting of how specific app features are linked to engagement.

Implications

The implications include the evaluation of existing mHealth apps and the development of new mHealth app interventions.

Evaluation of Existing mHealth Apps

Definitions and measures of behavioral engagement indicators must be standardized. Future research should adopt a validated multidimensional operationalization of engagement that moves beyond single indicators such as logins. We recommend reporting the amount (eg, number of sessions), duration time spent (eg, minutes spent in sessions), depth (eg, number of tasks completed in the same session), and frequency (eg, number of logins) [10,11]. Complementary use of validated instruments such as the Twente Engagement With eHealth Technologies Scale (TWEETS) [59] can capture both behavioral and subjective dimensions of engagement, including interest, attention, and enjoyment. This combined approach is essential for distinguishing mere usage from outcome-relevant engagement.

XSL•FO
RenderX

Future Intervention Development

As outlined in Table 2, the motivational mechanisms identified, specifically those supporting autonomy, competence, and relatedness, translate into clear, actionable design recommendations. These recommendations connect psychological needs directly to design choices, supporting mechanism-informed intervention development.

Importantly, many autonomy- and competence-related barriers, such as digital exclusion, cost, and low digital literacy, are structural challenges requiring systemic responses. The features summarized in Table 2 include practical solutions such as offering offline or low-bandwidth app versions, SMS text message–based alternatives for users without smartphones, and device-loan schemes to expand access. To support competence among digitally excluded users, guided onboarding, clear tutorials, and accessible interface design (eg, large fonts, high-contrast modes, and multilingual options) are recommended. Moreover, partnerships with community health workers, primary care providers, and advocacy organizations (relatedness-supportive features) can improve trust and contextual relevance for the LTC populations that the mHealth apps aim to serve [60,61]. These measures directly address engagement gaps and help prevent mHealth innovations from exacerbating health inequalities [62-64]. Further work is needed to quantitatively confirm if some of these recommendations do, in fact, make a difference.

Strengths and Limitations

A key strength of this overview of systematic reviews is its theory-informed synthesis across quantitative, qualitative, and mixed methods reviews, enabling a multidimensional understanding of engagement with mHealth apps. Unlike prior syntheses focused narrowly on behavioral indicators, this review synthesized the barriers and facilitators of engagement within a motivational framework, which offers insight into subjective indicators. Applying SDT provided a coherent framework for mapping barriers and facilitators onto psychological needs and for generating design-oriented recommendations. Methodological rigor was supported through prospective PROSPERO (International Prospective Register of Systematic Reviews) registration and adherence to PRIOR guidelines.

Several limitations should be acknowledged. Review quality varied; quantitative reviews were often low quality, while qualitative and mixed methods reviews were stronger. In addition, the CASP qualitative checklist was developed for the appraisal of primary qualitative studies rather than review-level evidence syntheses. However, we considered it appropriate as a structured framework for assessing the qualitative and mixed methods components of included reviews because it provides methodological depth in domains such as reflexivity, rigor of analysis, credibility of findings, and the value of the research. This approach is also consistent with a previous overview of systematic reviews that used CASP as part of the quality appraisal of systematic reviews of qualitative studies [65]. Key populations, including older adults, people with low digital literacy, and those in rural or underserved settings, were underrepresented, limiting generalizability. Publication bias may also have excluded studies reporting no engagement-outcome associations. Applying SDT introduced some interpretive subjectivity, as most reviews were not theoretically framed; however, mappings were independently cross-checked to enhance consistency. These limitations reduce the certainty of the conclusions but highlight clear priorities for future research.

Conclusion

mHealth apps are a promising approach to supporting self-management of LTCs, but their effectiveness partly relies on users' engagement with the intervention and their ability to apply its therapeutic content in daily life. This overview of systematic reviews highlights conceptual and methodological gaps in engagement research and demonstrates the value of applying SDT to understand why engagement occurs and how it can be supported. The SDT-informed design guidance provides a foundation for developing more engaging, equitable, and user-centered mHealth app interventions. Embedding these principles in future evaluation and development can strengthen both user experience and clinical impact and contribute to more consistent, theory-driven progress in digital health.

The graphical abstract for the article is provided in Multimedia Appendix 8 .

Acknowledgments

The authors thank the senior librarians at King's College London for their assistance with the data search. No generative AI tools were used in writing the manuscript text.

Funding

This study was funded by the UK MS Society under award reference 180, and part funded by the National Institute of Health and Care Research (NIHR) Biomedical Research Centre (BRC): Maudsley.

Data Availability

All data extracted for this overview of systematic reviews are available within the article and its supplementary materials. No primary data were generated.

XSL•FO
RenderX

Authors' Contributions

Conceptualization: JB, JH, RM-M
Data curation: JB
Funding acquisition: JH, SN, RM-M
Investigation: JB, GHB, MM-I
Methodology: JB, SN
Supervision: CJP, JH, SN, RM-M
Validation: JB, GHB
Visualization: JB
Writing – original draft: JB
Writing – review & editing: GHB, CJP, MM-I, JH, SN, RM-M

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy for all databases and gray literature.
[DOCX File, 37 KB - mhealth_v14i1e88382_app1.docx]

Multimedia Appendix 2
Characteristics of included systematic reviews.
[DOCX File, 22 KB - mhealth_v14i1e88382_app2.docx]

Multimedia Appendix 3
Quality appraisal of quantitative systematic reviews (n=12) using the AMSTAR 2 tool.
[DOCX File, 43 KB - mhealth_v14i1e88382_app3.docx]

Multimedia Appendix 4
Quality appraisal of qualitative and mixed methods systematic reviews (n=7) using the CASP tool.
[DOCX File, 15 KB - mhealth_v14i1e88382_app4.docx]

Multimedia Appendix 5
Mobile health app–based intervention characteristics (n=19).
[DOCX File, 26 KB - mhealth_v14i1e88382_app5.docx]

Multimedia Appendix 6
Descriptive engagement indicators and their impact on health outcomes (n=19).
[DOCX File, 17 KB - mhealth_v14i1e88382_app6.docx]

Multimedia Appendix 7
Mapping modifiable barriers and facilitators of engagement to the Self-Determination Theory.
[DOCX File, 19 KB - mhealth_v14i1e88382_app7.docx]

Multimedia Appendix 8
Graphical abstract.
[PNG File, 254 KB - mhealth_v14i1e88382_app8.png]

Checklist 1
PRIOR checklist_Revision Round_20.06.2026.
[PDF File, 282 KB - mhealth_v14i1e88382_app9.pdf]

References

1. Ownby RL, Waldrop D, Davenport R, et al. A mobile app for chronic disease self-management for individuals with low health literacy: a multisite randomized controlled clinical trial. medRxiv. Preprint posted online on Apr 3, 2023. [doi: 10.1101/2023.04.01.23288020] [Medline: 37066256]

XSL•FO
RenderX

2. Fit for the future: 10 year health plan for England. : Department of Health and Social Care; 2025 URL: https://assets.publishing.service.gov.uk/media/6888a0b1a11f859994409147/fit-for-the-future-10-year-health-plan-for-england.pdf [accessed 2026-07-09]
3. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23. [doi: 10.2196/mhealth.8873] [Medline: 29343463]
4. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52. [doi: 10.2196/jmir.3951] [Medline: 25803266]
5. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform 2012 Feb;45(1):184-198. [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]
6. Ricciardi L, Mostashari F, Murphy J, Daniel JG, Siminerio EP. A national action plan to support consumer engagement via e-health. Health Aff (Millwood) 2013 Feb;32(2):376-384. [doi: 10.1377/hlthaff.2012.1216] [Medline: 23381531]
7. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 5;13(3):e52. [doi: 10.2196/jmir.1772] [Medline: 21821503]
8. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152. [doi: 10.2196/jmir.2104] [Medline: 23151820]
9. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11. [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
10. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
11. Yardley L, Spring BJ, Riper H, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
12. Villegas Mejía C, Remmerswaal D, Engels RCME, Ludden GDS, Boffo M. Macro-engagement in mHealth: exploring user engagement beyond the screen. Digit Health 2024;10:20552076231225591. [doi: 10.1177/20552076231225591] [Medline: 38269371]
13. Fawson S, Dibley L, Smith K, et al. Developing an online program for self-management of fatigue, pain, and urgency in inflammatory bowel disease: patients' needs and wants. Dig Dis Sci 2022 Jul;67(7):2813-2826. [doi: 10.1007/s10620-021-07109-9] [Medline: 34146202]
14. Carroll S, Moss-Morris R, Hulme K, Hudson J. Therapists' perceptions of barriers and facilitators to uptake and engagement with therapy in long-term conditions. Br J Health Psychol 2021 May;26(2):307-324. [doi: 10.1111/bjhp.12475] [Medline: 33043530]
15. Schueller SM, Torous J. Scaling evidence-based treatments through digital mental health. Am Psychol 2020 Nov;75(8):1093-1104. [doi: 10.1037/amp0000654] [Medline: 33252947]
16. Ng JYY, Ntoumanis N, Thøgersen-Ntoumani C, et al. Self-determination theory applied to health contexts: a meta-analysis. Perspect Psychol Sci 2012 Jul;7(4):325-340. [doi: 10.1177/1745691612447309] [Medline: 26168470]
17. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
18. Ntoumanis N, Ng JYY, Prestwich A, et al. A meta-analysis of self-determination theory-informed intervention studies in the health domain: effects on motivation, health behavior, physical, and psychological health. Health Psychol Rev 2021 Jun;15(2):214-244. [doi: 10.1080/17437199.2020.1718529] [Medline: 31983293]
19. Iribarren SJ, Akande TO, Kamp KJ, Barry D, Kader YG, Suelzer E. Effectiveness of mobile apps to promote health and manage disease: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2021 Jan 11;9(1):e21563. [doi: 10.2196/21563] [Medline: 33427672]
20. Eaton C, Vallejo N, McDonald X, et al. User engagement with mHealth interventions to promote treatment adherence and self-management in people with chronic health conditions: systematic review. J Med Internet Res 2024 Sep 24;26:e50508. [doi: 10.2196/50508] [Medline: 39316431]
21. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
22. Noncommunicable diseases. World Health Organization. 2025. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2026-07-07]
23. GBD 2021 Risk Factors Collaborators. Global burden and strength of evidence for 88 risk factors in 204 countries and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. The Lancet 2024 May;403(10440):2162-2203. [doi: 10.1016/S0140-6736(24)00933-4] [Medline: 38762324]
24. EndNote: citing the endnote program as a reference. Clarivate. 2013. URL: https://supportcenter.clarivate.com/s/article/Citing-the-EndNote-program-as-a-reference?language=en_US [accessed 2026-07-07]
25. Covidence. 2023. URL: https://www.covidence.org [accessed 2026-07-07]

XSL•FO
RenderX

26. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282. [doi: 10.11613/BM.2012.031] [Medline: 23092060]
27. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017 Sep 21;358:j4008. [doi: 10.1136/bmj.j4008] [Medline: 28935701]
28. Long HA, French DP, Brooks JM. Optimising the value of the Critical Appraisal Skills Programme (CASP) tool for quality appraisal in qualitative evidence synthesis. Res Methods Med Health Sci 2020 Sep;1(1):31-42. [doi: 10.1177/2632084320947559]
29. Classification of digital interventions, services and applications in health: a shared language to describe the uses of digital technology for health. : World Health Organization; 2023 URL: https://iris.who.int/server/api/core/bitstreams/bb20354e-ebee-44aa-8a58-7b1da85c6f95/content [accessed 2026-07-07]
30. Villalobos-Zúñiga G, Cherubini M. Apps that motivate: a taxonomy of app features based on self-determination theory. Int J Hum Comput Stud 2020 Aug;140:102449. [doi: 10.1016/j.ijhcs.2020.102449]
31. Jeno LM, Egelandsdal K, Grytnes JA. A qualitative investigation of psychological need-satisfying experiences of a mobile learning application: a Self-Determination Theory approach. Comput. Educ. Open. 2022 Dec;3:100108. [doi: 10.1016/j.caeo.2022.100108]
32. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6(1):42. [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
33. Hennessy EA, Johnson BT. Examining overlap of included studies in meta-reviews: guidance for using the corrected covered area index. Res Synth Methods 2020 Jan;11(1):134-145. [doi: 10.1002/jrsm.1390] [Medline: 31823513]
34. Lee J, Yeom I, Chung ML, Kim Y, Yoo S, Kim E. Use of mobile apps for self-care in people with Parkinson disease: systematic review. JMIR Mhealth Uhealth 2022 Jan 21;10(1):e33944. [doi: 10.2196/33944] [Medline: 35060910]
35. Rintala A, Kossi O, Bonnechère B, Evers L, Printemps E, Feys P. Mobile health applications for improving physical function, physical activity, and quality of life in stroke survivors: a systematic review. Disabil Rehabil 2023 Dec;45(24):4001-4015. [doi: 10.1080/09638288.2022.2140844] [Medline: 36325613]
36. Alaslawi H, Berrou I, Al Hamid A, Alhuwail D, Aslanpour Z. Diabetes self-management apps: systematic review of adoption determinants and future research agenda. JMIR Diabetes 2022 Jul 28;7(3):e28153. [doi: 10.2196/28153] [Medline: 35900826]
37. Vaezipour A, Whelan BM, Wall K, Theodoros D. Acceptance of rehabilitation technology in adults with moderate to severe traumatic brain injury, their caregivers, and healthcare professionals: a systematic review. J Head Trauma Rehabil 2019;34(4):E67-E82. [doi: 10.1097/HTR.0000000000000462] [Medline: 30608310]
38. O'Neill M, Houghton C, Crilly G, Dowling M. A qualitative evidence synthesis of users' experience of mobile health applications in the self-management of type 2 diabetes. Chronic Illn 2022 Mar;18(1):22-45. [doi: 10.1177/1742395320983877] [Medline: 33583207]
39. Campbell J, Porter J. Dietary mobile apps and their effect on nutritional indicators in chronic renal disease: a systematic review. Nephrology (Carlton) 2015 Oct;20(10):744-751. [doi: 10.1111/nep.12500] [Medline: 25959301]
40. de Melo Santana B, Raffin Moura J, Martins de Toledo A, et al. Efficacy of mHealth interventions for improving the pain and disability of individuals with chronic low back pain: systematic review and meta-analysis. JMIR Mhealth Uhealth 2023 Nov 2;11:e48204. [doi: 10.2196/48204] [Medline: 37962085]
41. Frid S, Amat-Fernández C, Fuentes-Expósito M, et al. Mapping the evidence on the impact of mHealth interventions on patient-reported outcomes in patients with breast cancer: a systematic review. JCO Clin Cancer Inform 2024 May;8:e2400014. [doi: 10.1200/CCI.24.00014] [Medline: 38710001]
42. He Q, Zhao X, Wang Y, Xie Q, Cheng L. Effectiveness of smartphone application-based self-management interventions in patients with type 2 diabetes: a systematic review and meta-analysis of randomized controlled trials. J Adv Nurs 2022 Feb;78(2):348-362. [doi: 10.1111/jan.14993] [Medline: 34324218]
43. Hernandez Silva E, Lawler S, Langbecker D. The effectiveness of mHealth for self-management in improving pain, psychological distress, fatigue, and sleep in cancer survivors: a systematic review. J Cancer Surviv 2019 Feb;13(1):97-107. [doi: 10.1007/s11764-018-0730-8] [Medline: 30635865]
44. Patterson K, Davey R, Keegan R, Freene N. Smartphone applications for physical activity and sedentary behaviour change in people with cardiovascular disease: a systematic review and meta-analysis. PLoS One 2021;16(10):e0258460. [doi: 10.1371/journal.pone.0258460] [Medline: 34634096]
45. Bezerra Giordan L, Tong HL, Atherton JJ, et al. The use of mobile apps for heart failure self-management: systematic review of experimental and qualitative studies. JMIR Cardio 2022 Mar 31;6(1):e33839. [doi: 10.2196/33839] [Medline: 35357311]
46. Dunham M, Bonacaro A, Schofield P, Bacon L, Spyridonis F, Mehrpouya H. Smartphone applications designed to improve older people's chronic pain management: an integrated systematic review. Geriatrics (Basel) 2021 Apr 8;6(2):40. [doi: 10.3390/geriatrics6020040] [Medline: 33917697]
47. Diez Alvarez S, Fellas A, Wynne K, et al. The role of smartwatch technology in the provision of care for type 1 or 2 diabetes mellitus or gestational diabetes: systematic review. JMIR Mhealth Uhealth 2024 Dec 3;12:e54826. [doi: 10.2196/54826] [Medline: 39625868]

XSL•FO
RenderX

48. Magalhães B, Fernandes C, Santos C, Martínez-Galiano JM. The use of mobile applications for managing care processes during chemotherapy treatments: a systematic review. Cancer Nurs 2021;44(6):E339-E360. [doi: 10.1097/NCC.0000000000000823] [Medline: 32384423]
49. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May 16;18(5):e97. [doi: 10.2196/jmir.4883] [Medline: 27185295]
50. Patail K, Pai HH, Greenfield G, et al. What are the perceptions and experiences of adults using mobile applications for self-management in diabetes? A systematic review. BMJ Open 2025 Jan 20;15(1):e086671. [doi: 10.1136/bmjopen-2024-086671] [Medline: 39833008]
51. Horn A, Jírů-Hillmann S, Widmann J, et al. Systematic review on the effectiveness of mobile health applications on mental health of breast cancer survivors. J Cancer Surviv 2025 Feb;19(1):1-17. [doi: 10.1007/s11764-023-01470-6] [Medline: 37906420]
52. MacLean RR, Shor R, Reilly ED, et al. Engagement in digital self-management interventions for chronic pain: a systematic review. Clin J Pain 2025 Jun 1;41(6):e1289. [doi: 10.1097/AJP.0000000000001289] [Medline: 40145149]
53. Bijkerk LE, Oenema A, Geschwind N, Spigt M. Measuring engagement with mental health and behavior change interventions: an integrative review of methods and instruments. Int J Behav Med 2023 Apr;30(2):155-166. [doi: 10.1007/s12529-022-10086-6] [Medline: 35578099]
54. Boucher EM, Raiker JS. Engagement and retention in digital mental health interventions: a narrative review. BMC Digit Health 2024;2(1):52. [doi: 10.1186/s44247-024-00105-9]
55. Elkes J, Cro S, Batchelor R, et al. User engagement in clinical trials of digital mental health interventions: a systematic review. BMC Med Res Methodol 2024 Aug 24;24(1):184. [doi: 10.1186/s12874-024-02308-0] [Medline: 39182064]
56. O'Brien HL, Toms EG. What is user engagement? A conceptual framework for defining user engagement with technology. J Am Soc Inf Sci 2008 Apr;59(6):938-955. [doi: 10.1002/asi.20801]
57. Torous J, Michalak EE, O'Brien HL. Digital health and engagement-looking behind the measures and methods. JAMA Netw Open 2020 Jul 1;3(7):e2010918. [doi: 10.1001/jamanetworkopen.2020.10918] [Medline: 32678446]
58. Sanatkar S, Baldwin PA, Huckvale K, et al. Using cluster analysis to explore engagement and e-attainment as emergent behavior in electronic mental health. J Med Internet Res 2019 Nov 28;21(11):e14728. [doi: 10.2196/14728] [Medline: 31778115]
59. Kelders SM, Kip H, Greeff J. Psychometric evaluation of the TWente Engagement with Ehealth Technologies Scale (TWEETS): evaluation study. J Med Internet Res 2020 Oct 9;22(10):e17757. [doi: 10.2196/17757] [Medline: 33021487]
60. Radu I, Scheermesser M, Spiess MR, Schulze C, Händler-Schuster D, Pehlke-Milde J. Digital health for migrants, ethnic and cultural minorities and the role of participatory development: a scoping review. Int J Environ Res Public Health 2023 Oct 23;20(20):6962. [doi: 10.3390/ijerph20206962] [Medline: 37887700]
61. Global strategy on digital health 2020-2025. : World Health Organization; 2021 URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2026-07-07]
62. Salemink K, Strijker D, Bosworth G. Rural development in the digital age: a systematic literature review on unequal ICT availability, adoption, and use in rural areas. J Rural Stud 2017 Aug;54:360-371. [doi: 10.1016/j.jrurstud.2015.09.001]
63. 2021 census: exploring the UK's digital divide. : Office for national statistics; 2021 URL: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/articles/exploringtheuksdigitaldivide/2019-03-04 [accessed 2026-07-07]
64. van Dijk J. The Digital Divide: Polity Press; 2020. URL: https://research.utwente.nl/en/publications/the-digital-divide-2/ [accessed 2026-07-07]
65. May CR, Cummings A, Myall M, et al. Experiences of long-term life-limiting conditions among patients and carers: what can we learn from a meta-review of systematic reviews of qualitative studies of chronic heart failure, chronic obstructive pulmonary disease and chronic kidney disease? BMJ Open 2016 Oct 5;6(10):e011694. [doi: 10.1136/bmjopen-2016-011694] [Medline: 27707824]

Abbreviations

AMSTAR 2: Assessing the Methodological Quality of Systematic Reviews 2
CASP: Critical Appraisal Skills Program
HMIC: Health Management Information Consortium (Ovid)
JBI: Joanna Briggs Institute
LTC: long-term physical health condition
mHealth: mobile health
PRIOR: Preferred Reporting Items for Overviews of Reviews
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
SDT: Self-Determination Theory

XSL•FO
RenderX

TWEETS: TWente Engagement with Ehealth Technologies Scale
WHO: World Health Organization

Edited by L Buis; submitted 16.Jan.2026; peer-reviewed by J Corser, P Misquitta; revised version received 24.Jun.2026; accepted 26.Jun.2026; published 24.Jul.2026.

Please cite as:
Baykoca J, Hurtado Barbeito G, Pearce CJ, Milne-Ives M, Hudson J, Norton S, Moss-Morris R
Mechanisms of Engagement With Mobile Health Apps for Adults With Long-Term Conditions: Overview of Systematic Reviews
JMIR Mhealth Uhealth 2026;14:e88382
URL: https://mhealth.jmir.org/2026/1/e88382
doi:10.2196/88382

© Jeni Baykoca, Goretti Hurtado Barbeito, Christina Joanne Pearce, Madison Milne-Ives, Joanna Hudson, Sam Norton, Rona Moss-Morris. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 24.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Perceptions of Adults Aged 50 Years and Older Regarding the Use of Wearable mHealth Technologies to Promote Physical Activity: Systematic Review and Meta-Ethnography

Beatriz Rodríguez-Martín[1,2,3*], RN, MSci, PhD, Prof Dr; Diego González-Arroyo[4*], RN; Susana Priego-Jiménez[2,5], MSci, PhD; María Lopez-González[2], MSci; María Medrano-Echeverría[6,7], PhD; Marta Carolina Ruiz-Grao[2,8], RN, MSci, PhD; Celia Álvarez-Bueno[2,9], RN, MSci, PhD, Prof Dr

[1]Department of Nursing, Faculty of Health Sciences, Physiotherapy and Occupational Therapy, University of Castilla-La Mancha, Talavera de la Reina, Spain

[2]ABC-Age Research Group, University of Castilla-La Mancha, Cuenca, Spain

[3]Network for Research on Chronicity, Primary Care, and Health Promotion (RICAPPS), Barcelona, Spain

[4]University Hospital Nuestra Señora del Prado, Health Service of Castilla-La Mancha, Talavera de la Reina, Spain

[5]University Hospital of Cuenca, Camino de El Terminillo s/n, Cuenca, Spain

[6]Department of Health Sciences, Institute for Innovation & Sustainable Food Chain Development, Public University of Navarra, Pamplona, Spain

[7]CIBER de Fisiopatología de la Obesidad y Nutrición (CIBEROBN), Instituto de Salud Carlos III, Madrid, Spain

[8]Department of Nursing, Nursing Faculty, Physiotherapy and Occupational Therapy, University of Castilla-La Mancha. Albacete, Albacete, Spain

[9]Universidad Politécnica y Artística del Paraguay, Asuncion, Paraguay

[*]these authors contributed equally

Corresponding Author:
Susana Priego-Jiménez, MSci, PhD
ABC-Age Research Group, University of Castilla-La Mancha, Cuenca, Spain

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2026/1/e111082

Abstract

Background: Despite advances in wearable mobile health (mHealth) technologies and their associated apps designed to promote physical activity, and the importance of adapting them to users, little is known about older adults' perceptions of these technologies.

Objective: This review aimed to synthesize and analyze qualitative evidence exploring the perceptions of adults aged 50 years and older regarding areas to improve, barriers to, and facilitators of wearable mHealth technologies (activity trackers and companion apps) to promote physical activity.

Methods: A qualitative systematic review and meta-ethnography was conducted. Comprehensive searches were performed across 8 databases (MEDLINE, Scopus, Web of Science, CINAHL, The Cochrane Library Plus, PsycINFO, ProQuest, and ÍnDICEs-CSIC) for articles published in English or Spanish between January 2013 and January 2024. The synthesis followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) and ENTREQ (Enhancing Transparency in Reporting the Synthesis of Qualitative Research) guidelines.

Results: Ten articles met the inclusion criteria and were synthesized using meta-ethnography. Three main themes emerged: (1) barriers to promoting physical activity caused by wearable mHealth technologies: personal barriers (physical aspects, perceptions about technology, and personal preferences), technological barriers (functionality, content, design, alarms, availability, and accessibility), and environmental barriers (season of the year); (2) personal facilitators (consideration that these apps improve health, perceptions about technology, and personal preferences), technological facilitators (functionality, content, and design), relational facilitators (technological and social support), environmental facilitators (seasonal variations), and health care professionals (support and monitoring by health care services); and (3) personal areas (perceptions about technology and personal preferences), technological areas (functionality, content, and design), and relational areas (technological support).

Conclusions: Although older adults acknowledge the potential of wearable mHealth technologies to promote physical activity, their effective engagement is hindered by distinct personal, technological, and environmental barriers. Bridging the digital divide requires designers to prioritize user-centered, age-friendly interfaces that are integrated with continuous support from health care

XSL•FO
RenderX

professionals. To promote genuine health equity, future research must rigorously report intersectional demographics to ensure that mHealth interventions mitigate, rather than inadvertently exacerbate, existing disparities.

(JMIR Mhealth Uhealth 2026;14:e67157) doi:10.2196/67157

KEYWORDS

eHealth; mHealth; mobile health; mobile app; app; application; smartphone; digital health; digital technology; digital intervention; electronic devices; meta-ethnography; older people; qualitative; physical activity; physical exercise; exercise; exercising; systematic review; review

Introduction

Population aging is a phenomenon of increasing magnitude worldwide. By 2025, approximately 1.2 billion people will be aged 60 years and older, reaching 2 billion by 2050 [1]. Additionally, according to the World Health Organization (WHO) forecasts, between 2000 and 2050, the percentage of people aged 60 years and older will double from 11% to 22% of the global population [2-5], surpassing the remaining population groups [6].

Population aging is associated with an increase in preventable chronic pathologies, such as rheumatism, arterial hypertension, heart disease, diabetes, chronic obstructive pulmonary disease, cerebrovascular disease, hearing loss, cataracts, depression, and dementia, as well as increasing disability, dependence, and the use of health system resources [7], which increases starting at age 55 years in men and from 60 years or more in women [7,8], accounting for 70% - 80% of health expenditures in Europe [9]. Older adults consume 50% of the time spent with primary care professionals, 70% of the time of geriatricians, and 62% of the time spent on pharmaceutical expenditure [7], which can be reduced by half if aging is accompanied by good health, reinforcing primary care and senior services [9].

Physical activity (PA), defined as any movement that involves energy consumption, has been shown to have physical and mental benefits for older adults, such as reducing cardiovascular disease, increasing quality of life, and maintaining independence [10-12]. The WHO recommends that adults older than 65 years of age should perform between 150 and 300 minutes of moderate-intensity aerobic PA or between 75 and 150 minutes of vigorous-intensity PA per week. In addition, performing 2 or more days of muscle-strengthening activities for all muscle groups and performing PA 3 or more days a week are recommended, prioritizing functional balance and moderate-intensity strength [13]. The benefits of these recommendations include decreasing the incidence of hypertension, cancer, and type 2 diabetes; improving mental health, cognitive health, and sleep; reducing cardiovascular mortality; preventing falls; and worsening bone and functional health [2,10,13,14]. Therefore, effective interventions aimed at promoting PA among older adults are necessary [14].

Information and communication technologies in health, which are being increasingly used by older people, have revolutionized how health services address the challenges of improving the quality, effectiveness, and safety of care [15,16]. Such technologies are great tools for promoting active aging. Prominent among these technologies is eHealth, defined by the WHO as "the cost-effective and safe use of information and communication technologies in support of health-related areas, including health services, health surveillance, health documentation and education, health knowledge and research" [17]. Within eHealth is mHealth, defined as medical and public health practices supported by mobile devices. In the context of PA promotion, mHealth rarely operates as a stand-alone software app; rather, it functions as a broader, integrated ecosystem that combines hardware (such as wearable activity trackers, smartbands, or smartwatches) and companion software apps [18]. The inclusion of these hybrid interventions is crucial, as their physical nature (hardware) and digital interface (software) jointly increase older adults' self-care, autonomy in decision-making, and self-efficacy in disease management. Furthermore, mHealth motivates older adults to engage in PA. At the systemic level, these technologies can improve care quality, reduce health care costs, and prevent unnecessary medical consultations. For the users, mHealth facilitates social interaction, digital inclusion, and intergenerational relationships, ultimately reducing loneliness and enhancing overall quality of life [19,20]. To this end, the characteristics of older adults, the ease and intuition of the use of devices and apps, their accuracy and veracity, and their needs, desires, and limitations should be considered [21-23]. Despite the benefits of these technologies for older people, factors such as anxiety and fear of technology, deficits in cognitive skills, health problems, lack of instructions and guidelines, and the cost of the technologies are barriers to their use [19]. Perceptions about the use of mHealth devices and apps to promote PA have been studied among parents and adolescents, which has revealed their satisfaction with these technologies and their effectiveness in promoting PA among children and adolescents [24]; however, few qualitative studies have analyzed their benefits among older adults, and few existing studies have focused on populations with specific pathologies, such as myeloid neoplasms [25] or breast cancer [26].

The success of such interventions depends on their ability to adapt to the needs of the user; thus, understanding how older adults can influence the use of mHealth devices and apps for the promotion of PA and analyzing their demands, contexts, and desires to develop more accessible, age-friendly, and acceptable technologies for this population group are essential [19].

This systematic review of qualitative evidence aimed to synthesize and analyze studies that explore the perceptions of adults aged 50 years and older regarding areas to improve, barriers, and facilitators of wearable mHealth technologies to promote PA. This information is key both for the tech developers

XSL•FO
RenderX

and for the health care professionals who recommend, train, and supervise users in their use.

Methods

Overview

A qualitative systematic review and meta-ethnography were carried out. Noblit and Hare [27] described this type of synthesis of qualitative research results, which is the most widely used qualitative synthesis approach in health care. The protocol of this review was registered in PROSPERO (CDR498374). This review was conducted following the recommendations of the Cochrane Qualitative and Implementation Methods Group Guidance [28], the ENTREQ (Enhancing Transparency in Reporting the Synthesis of Qualitative Research) statement, and the criteria of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) Statement [29] (Checklist 1).

Eligibility Criteria

The following inclusion and exclusion criteria were used to select the articles.

The inclusion criteria were as follows: (1) published in English and Spanish from January 2013 to January 2024 to include the latest publications on the subject, (2) used any type of qualitative approach or qualitative data collection technique, and (3) analyzed the use of mHealth ecosystems, specifically wearable PA trackers and their associated mobile apps, for the promotion of PA in healthy or chronically ill adults aged 50 years and older who were community-dwelling or institutionalized, functionally autonomous, and not classified as frail. The exclusion criteria were as follows: (1) studies that included participants with cognitive impairment or any mental pathology, (2) protocol studies or studies not completed, (3) mixed designs if the qualitative data were not analyzed in a disaggregated manner, and (4) studies that included the use of the wearable mHealth technologies exclusively for dietary recommendations.

Data Sources and Search Strategy

Two researchers independently conducted the search and selection of articles and resolved discrepancies through consensus, with a third researcher consulted when necessary, following the established criteria in the following databases: MEDLINE (PubMed), Scopus, Web of Science, ÍnDICEs-CSIC, ProQuest, PsycINFO, The Cochrane Library Plus, and CINAHL. In addition, a secondary search was conducted using the reference lists of the selected articles. In cases of disagreement, a third researcher approved the study. The search strategy included a combination of keywords adapted to each database. The complete search strategy is presented in Multimedia Appendix 1.

Study Selection and Data Abstraction

Article selection was performed independently by 2 researchers, with disagreements resolved by a third reviewer. First, the titles and abstracts of the articles were examined, and those that did not meet the inclusion criteria were discarded. The full texts of the selected articles were subsequently read, and the following information was extracted from each of them using an ad hoc Excel template: authors, year of publication, paradigmatic approach, objectives of the study, characteristics of the sample and sampling techniques, data collection techniques, main results, conclusions, and evaluation of methodological quality.

Assessment of Methodological Quality

The quality of the studies was assessed with the Joanna Briggs Institute Qualitative Assessment and Review Instrument (JBI-QARI) [30], which includes 10 items with 3 response options, “yes,” “no,” and “unclear,” to evaluate the methodological quality of the qualitative studies. Studies with poor methodological quality (fewer than 5 items) were rated as “excluded” and removed from the review.

Data Analysis and Synthesis

Following Noblit and Hare’s interpretive approach [27], the findings were synthesized through reciprocal translation, evolving into a line of argument that provides an original theoretical reconceptualization of mHealth apps to promote PA.

Data were systematically categorized using an abstraction form designed to distinguish between first-order constructs (participants’ quotations and metaphors), second-order constructs (interpretations based on patients’ quotations made by the original researchers), and third-order constructs (interpretations generated by our research team). By juxtaposing these layers, we moved beyond data aggregation toward a reciprocal translation of findings, synthesizing these accounts into a novel line of argument that reconceptualizes the phenomenon within the current literature (Multimedia Appendix 2).

The main themes, categories, subcategories, and codes were extracted. The codes were grouped by areas of similarity, resulting in the different themes, categories, and subcategories of facilitators, barriers, and areas for improvement for the promotion of PA among older adults. For the translation of the studies, the data were exported to a table containing the thematic tree, categories, subcategories, codes, and quotations to help us understand the relationships between the studies. No refutations or contradictions were found, as the data from all included studies were very similar. Finally, the researchers compared the translations to identify the lines of argument from which the synthesis would be developed.

Research Team and Reflexivity

The research team has previous experience in qualitative research, evidence synthesis, PA, and research involving older people. The team included registered nurses, physiotherapists, an anthropologist, and a graduate in PA. The findings were analyzed from an interprofessional perspective.

Results

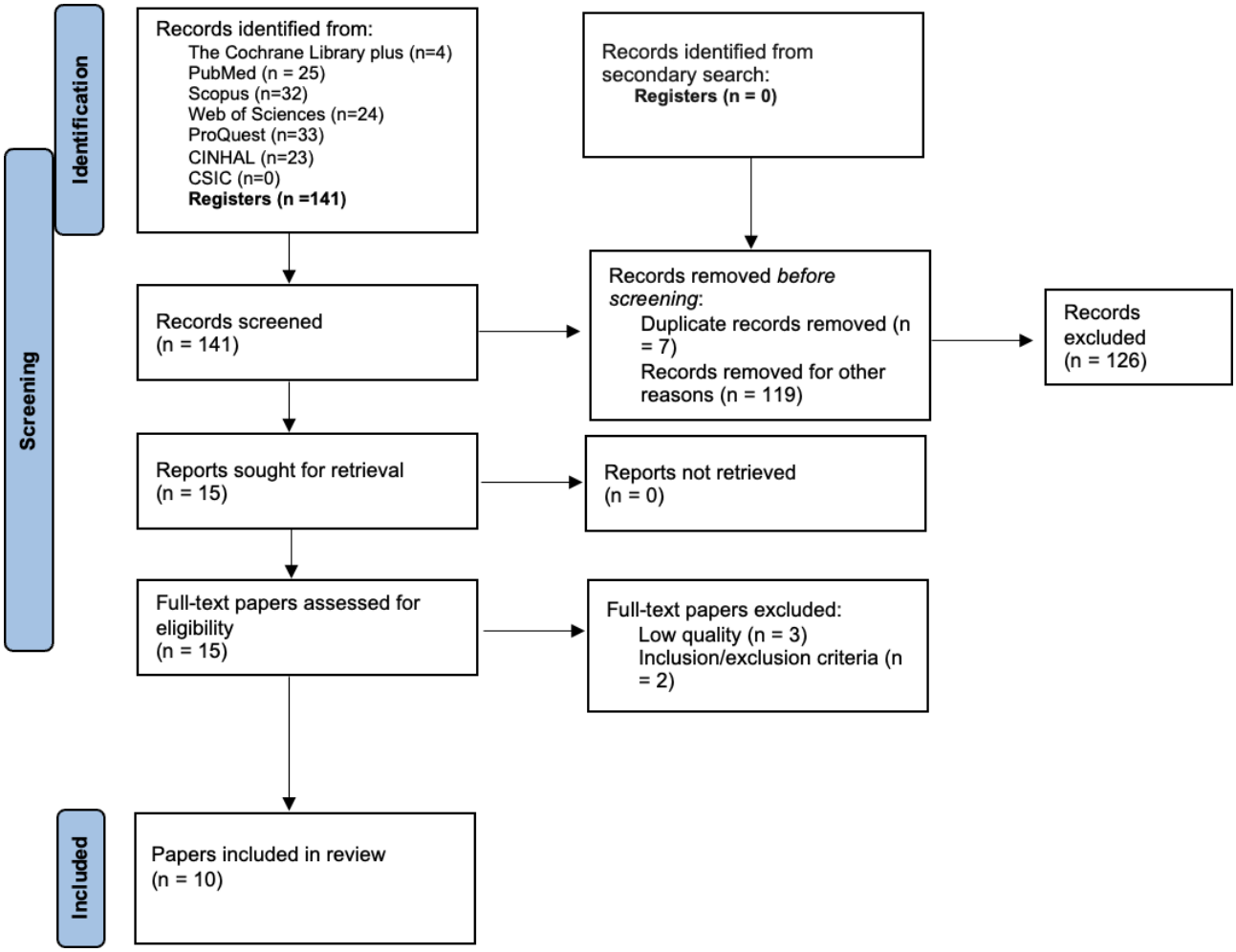
Study Selection and Characteristics

Following the database search, 141 references were identified, 7 of which were eliminated because they were duplicates across several databases. After the title and abstract were read, 15 articles met the inclusion criteria; 5 were excluded because of low methodological quality, specifically the use of the Joanna

XSL•FO
RenderX

Briggs Institute tool for qualitative studies. Finally, 10 articles were included in the narrative synthesis and meta-ethnography (Figure 1).

Figure 1. Flowchart of the search and selection process.



Synthesized Findings

The main characteristics of the studies analyzed are summarized in Table 1.

XSL•FO
RenderX

Table . Main characteristics of the studies analyzed.

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
Moore et al (2021) [31]	Qualitative systematic review, meta-ethnography type	To synthesize qualitative studies that analyzed the experience of older people who had used a PA[a] monitoring device for several days, and to determine which factors contributed to the acceptance and use of portable devices	• Twenty articles were included with a total of 349 participants aged 51 to 94 years who had used portable devices for several days and had analyzed their experiences • Intentional convenience sampling technique	Semistructured interviews and focus groups	• Three main themes explained older adults' experience with physical activity monitoring devices • Barriers: age-associated physiology and comorbidities. They were frustrated when the devices did not have the functions they wanted, were difficult to access, did not achieve their goal or capture all their activities, and had to rely on family or friends; when they were inaccurate, cumbersome, difficult to connect, incompatible with their phone, and complex instructions • Facilitators: continuous monitoring of the device and reaching their proposed goal. Supported by their family, friends, and health care personnel. That the devices were waterproof, small, comfortable,	• The factors that positively influence the use of PA devices and apps in older adults are personal motivation, ease of use, the objective set by the users, and considering their preferences. However, to overcome perceived barriers, such as usability or physiology-related barriers, older adults must be motivated and socially supported. In addition, device-related features (charging, instructions, and so on) should be improved.	10/10

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
					easy to put on and handle, safe, pleasant and easy to read display, long battery life, with a thin and flexible strap; that counted steps and had GPS, could measure health parameters, inform them of their progress and automatically synchronize with their other devices • Areas for improvement: clear and simple instructions, that the device was acceptable and reliable and that they did not have to charge it constantly		
Peng et al (2021) [32]	Phenomenology	To learn how users develop long-term habits following the use of PA monitoring devices (PA trackers)	• 20 adults over 65 years of age (11 women and 9 men) who had worn activity trackers for more than 6 months • Intentional convenience sampling technique	Two semistructured interviews were conducted with each participant			10/10

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
						• Contextual factors (such as the place where they place the tracker) and technological factors (such as set goals, reminders...) facilitate the adherence of older adults to use PA monitoring devices in the long term. • These user-identified factors can be used to design next-generation wearable activity trackers and facilitate behavior change in older adults.	

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
					• Analysis of the data obtained revealed the participants' perceptions of the use of physical activity trackers, classified into the following facilitators: • Facilitators: age-associated comorbidities. Setting a small goal and increasing it. Use of reminders and routines: they wore them first thing in the morning, charged them during the night, and always left them in the same place; they also used them in situations such as TV commercials, telephone conversations, or explored new routes to achieve the set goal. They were able to anticipate potential obstacles that might prevent them from using the activity tracker (eg, running out of battery). In addition,		

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
					they were prepared for dissatisfaction with not achieving the set goal and occasional lapses with a positive mindset and attitude		
Kononova et al (2019) [33]	Phenomenology	To determine the factors that favored or did not favor the use of activity trackers in people aged 65 years and older	• 48 people aged 65 years and older, of whom 17 had not used trackers, 9 had been using them in the short term, 11 had stopped using them, and 11 had used them in the long term • Intentional convenience sampling technique	Ten focus groups		• Improving the features and functions of mhealth[b] devices and apps perceived as barriers by older adults are essential aspects for their use. • However, being aware of the long-term benefits of app use, perceived social support, and motivation are facilitators that favor older adults' adherence to these devices.	10/10

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
					• Three themes explained participants' perceptions of the use of physical activity trackers that were categorized into the following barriers, facilitators, and areas for improvement: • Barriers: the devices did not have the function of establishing routines or an exercise schedule; inaccuracy in step counting, high price of the device, low interest in the device, and fear of sharing their data • Facilitators: that the device was attractive, with a comfortable bracelet and easy-to-read screen, waterproof, with long battery life; that other parameters such as sleep, heart rate, and so on, could be monitored; that the device allowed them to be more aware of their dai-		

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
					ly activity since its use reduced stress and promoted their independence; and to set goals, thus competing with family and friends • Areas for improvement: instructions for use and training on functionality		
Abouzahra and Ghasemaghaei (2020) [34]	Phenomenology	To examine the influence and effect of the use of PA monitoring devices in older adults	• 44 persons aged 65 years and older (20 men and 24 women) who had never used any PA monitoring device. • Intentional convenience sampling technique.	Two semistructured interviews with each participant.		• The use of PA monitoring devices in older adults is a complex process influenced by social (support from family and friends...), psychological (problem-solving skills), and technological (device complexity) factors. Taking these factors into account, strategies can be developed to improve the use of portable devices in this population and improve their health and quality of life.	10/10

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
					• The analysis of the data revealed 3 themes, classified into the following barriers, facilitators, and areas for improvement in the use of PA monitoring devices in older adults: • Barriers: inexperience in the use of technologies and the complexity of the device. The use of alarms and the high price of the device • Facilitators: alarms, previous experience with other technologies, and support from friends and family • Areas for improvement: monitoring other health parameters		
Ehn et al (2018) [35]	Phenomenology	To explore the experiences, requirements, and preferences of older adults for the use of PA monitoring devices		Individual semistructured interviews and focus groups.			10/10

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
			• Eight older adults aged 75 years and older; 6 had a mild walking disability and used wheeled walkers, and 2 participants walked without assistance. All knew how to use a smartphone, had not used an iPad, and 2 of them used computers. • Two physical activity recording wristbands were used. • Intentional convenience sampling technique.			• PA monitoring devices are an important tool to promote physical activity in older adults, but these should be easy to use and intuitive, with accurate activity log measurements, and display motivational messages.	

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
					• Three themes explained the participants' experiences using the activity monitoring devices that were categorized into barriers, facilitators, and areas for improvement: • Barriers: inaccurate measurements made by the devices. Many perceived the devices as fragile and therefore could break or damage them. They were afraid of being monitored and losing privacy, and some stopped using the device when it caused some kind of problem. Another barrier was that the device instructions were not in their native language. • Facilitators: achieving the proposed goal by making them more aware of how active they were even indoors. The use of reminders		

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
					and previous experience using other technologies made it easier for them to use the devices. They preferred the wristbands to be waterproof, flexible, small, comfortable, and easy to combine with clothing. • Areas for improvement: combining PA measurements with other health parameters and improvements in device battery life		
Fausset et al (2013) [36]	Phenomenology	To explore the use and attitudes of older adults toward 4 PA monitoring technologies and to learn how they integrate these devices into their daily lives	• Four married couples (6 perceived their health status as optimal and 2 as fair or poor) had a computer and internet connection and had no previous experience with any activity monitoring technology. • Intentional convenience sampling technique	Two semistructured interviews with each participant.		• Older adults perceive PA monitoring devices to be easy to use and useful as a key instrument for the promotion of healthy lifestyles. However, their high price, the inaccuracy of data measurement, and the loss of time in connecting and putting on the device are barriers to the use of these devices.	10/10

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
					• The most significant findings of the study were classified into barriers, facilitators, and areas for improvement. • Barriers: the inaccuracy in measuring the parameters, the high price, and the discomfort of wearing the bracelet at night • Facilitators: the enthusiasm and willingness shown by the participants to use the device and to be able to count the steps taken correctly. • Areas for improvement: long-term use, since half of the participants stated that they would not use the device continuously		
Mercer et al (2016) [37]	Phenomenology	To examine the usability and usefulness of activity monitoring devices in chronically ill older adults		Two focus groups were conducted.			10/10

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
			• 32 participants between 52 and 84 years of age with a chronic pathology. • Twelve participants did not have a cell phone, so they were provided with one. • The participants previously used a pedometer. Subsequently, they were assigned a device to use for at least 3 days. • Intentional convenience sampling technique			• For older adults with chronic diseases, wearable physical activity trackers are useful and acceptable. However, these users may need help setting up the device and learning how to interpret the data obtained.	

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
					• Three themes emerged in the study that described the acceptability of the devices by the older adults, which were classified into barriers, facilitators, and areas for improvement: • Barriers: participants were uncomfortable using the device because they thought they might break it. • Facilitators: that the devices were easy to use and made them more aware of their PA level, thanks to the alerts function, turning the set daily step goal into a game or competition with other users or with themselves. Finally, participants reflected that these devices could be useful in the health sector to improve their health and that they should be avail-		

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
					able in pharmacies and tax-free. • Areas for improvement: a simple manual of instructions, less technical terms, and clearer screens.		
Puri et al (2017) [38]	Mixed design, longitudinal quantitative study and qualitative study based on phenomenology	To evaluate the acceptability and use of activity monitoring devices in older adults considering the influencing factors identified in the literature and the technology acceptance model	• 20 adults older than 55 years of age, 12 women and 8 men. • Of the total, 18 used a computer daily, 14 owned a personal smartphone, and 17 had heard of activity monitoring devices. • Intentional convenience sampling technique	Four semistructured interviews were conducted in the qualitative part.		• Most older adults accept PA monitoring devices and perceive them as valuable to their lives. In addition, they are concerned about the features of the trackers (comfort, aesthetics, and price), whereas privacy is less of a concern for this group	10/10

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
					• Three themes explained older adults' perceptions of the acceptability of activity monitoring devices organized into the following barriers, facilitators, and areas for improvement. • Barriers: participants' lack of prior experience and seniors' sense of independence caused participants not to seek support from loved ones. • Facilitators: battery life, thin and flexible wristbands, sharing data with other users, and quick access to information managed by the device to understand the level and intensity of PA performed. In addition, the increased self-awareness of the device motivated them to make lifestyle changes and to inform themselves		

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
					about health-related topics. • Areas for improvement: the devices had to be easy to use, comfortable, and they were not afraid of breaking them		
Schlomann et al (2016) [39]	Phenomenology	To assess the experiences, opportunities, and barriers that older adults encounter in using PA trackers	• Six women aged 67-78 years who had no prior experience using wearable activity trackers but owned a smartphone. • Intentional convenience sampling technique	A focus group was conducted.		• Older adults are motivated to use PA monitoring devices. However, they perceive some problematic aspects that may cause some rejection in their use, such as the instructions provided, the handling, or the synchronization of these devices with the smartphone. • These aspects need to be considered for adequate use of PA monitoring in this age group.	10/10

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
					• Three themes assessed the barriers, facilitators, and areas for improvement perceived by the participants in the use of the activity monitoring devices. • Barriers: problems encountered when using the devices caused them anxiety, and they had to ask for help or call the researcher. In addition, they felt that the instructions provided were very brief and not in their native language. In addition, the devices were not waterproof, were incompatible or difficult to synchronize with some smartphones, there was inaccuracy in the step count, and they did not measure other clinical parameters (blood pressure, heart rate, etc.). • Facilitators: the devices		

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
					encouraged them to be more active and contributed to their personal well-being. Some participants changed their daily behavior due to the use of the device, comparing themselves with each other based on the steps they were able to take, or the information provided by their trackers. • Areas for improvement: participants did not perceive that the feedback provided by the devices forced them to be more active.		
Batsis et al (2016) [40]	Phenomenology	To evaluate the reliability and acceptability of a PA monitoring device (Fitbit) to promote behavior change in older adults with obesity residing in rural areas	• Eight older adults (4 women and 4 men) aged 65-80 years with BMI greater than 30 kg/m^2 and in need of weight loss. All participants had smartphones. • Intentional convenience sampling technique	A semistructured interview was conducted with each participant			10/10

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
						• Although the use of Fitbit is accepted by older adults with obesity, they perceive that they need advice from people (health care professionals, trainers, and so on) to teach them how to interpret the results obtained on the device and increase their confidence in the use of the device, thus bringing about a change in their behavior.	

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
					• The reliability and acceptability of the PA monitoring device in older adults in rural areas were evaluated and summarized in the following barriers and facilitators: • Barriers: it was observed that after the use of the PA monitoring devices, the total physical activity time of older adults decreased. • Facilitators: participants would recommend the use of the device as they were easy to use and found the feedback provided by the device useful. In addition, the social interaction was positively valued and motivated them to use the device, causing a positive behavior change. Finally, participants stated that the results obtained by the device should be explained by a health		

XSL•FO
RenderX

Author and year	Paradigmatic approach	Objectives of the study	Sample characteristics and sampling technique	Data collection technique	Main results	Conclusions	Methodological quality
					professional, trainer, and so on		

[a]PA: physical activity.

[b]mHealth: mobile health.

All the studies analyzed included adults older than 50 years, whose ages ranged broadly from 51 to 94 years. Most included mixed-gender samples, although one study exclusively analyzed the perspective of women [39]. With respect to comorbidities and functional status, several studies have moved beyond healthy populations to include individuals with chronic illnesses, older adults with obesity (BMI >30 kg/m^2), and participants with mild walking disabilities using wheeled walkers. Notably, formal categorizations of socioeconomic status (SES), culture, and ethnicity were absent from the primary studies' demographic reports. In terms of geographic scope, 5 studies were conducted in the United States [32-34,36,40], 2 in Canada [37,38], and the remaining in Sweden [35], Ireland [31], and Germany [39].

Concerning the objectives, 2 studies analyzed the barriers and facilitators perceived by older adults regarding the use of PA monitoring devices [33,39], another analyzed the experience of this population group using PA monitoring devices and their perceptions of factors influencing their long-term use [32], and the remaining analyzed the experiences, acceptability, effects, and preferences of older adults regarding the use of PA monitoring devices [34-38,40].

Most studies used the phenomenological approach [32-40], except for one study that used meta-ethnography [31]. The main data collection techniques used were focus groups [33,37], semistructured interviews [32,34,36,38,40], or a combination of both [31,35,39].

Description of the Main Findings

Seven of the analyzed studies reported the barriers, facilitators, and areas of improvement perceived by adults for the use of wearable mHealth technologies for PA promotion [31,33-35,37,38,41]; 1 synthesized the facilitators that promoted the use of wearable mHealth technologies among older adults [32], and 2 analyzed the facilitators and barriers that facilitated or hindered the use of these devices and apps [36,40].

Three major themes emerged from the narratives explaining older adults' perceptions of wearable mHealth technologies promoting PA: (1) barriers to their use of wearable mHealth technologies to promote PA: personal (physical and perceptions about technology and tastes and preferences), technological (functionality, content, design, alarms, availability, and accessibility), and environmental (seasonal variations); (2) facilitators of their use of wearable mHealth technologies to promote PA: personal (consideration that it improves health, perceptions about technology and tastes and preferences), technological (functionality, content, and design), relational (technological support and social support), environmental (seasonal variations), and health sector (support and monitoring of health services); and (3) main areas of improvement for the use of wearable mHealth technologies to promote PA: personal (perceptions about technology, tastes, and preferences), technological (functionality, content, and design), and relational (technological support). These themes are presented below with the corresponding categories, subcategories, and codes (Tables 2-4).

XSL•FO
RenderX

Table . Perceived barriers to the use of wearable mHealth[a] technologies for physical activity promotion by older adults (categories, subcategories, and codes).

Categories, subcategories, and codes	References
Personal barriers	
Physical aspects	
Decline in physical health	[31-33,35,37,39]
Physical limitations	[31,32,35]
Slow processing speed	[31,33,34]
Perceptions about technologies	
Lack of skill and insecurity in the use of technology	[31,33-36,38,39]
Low perception of technology use	[31,33-36,38,39]
Lack of previous experience	[31,33,34,38]
Negative feelings	[31,38-40]
Lack of interest and curiosity	[33,34,38]
Making mistakes by not asking for help from others around them	[31,38]
Personal preferences	
Change of routine	[31,33-35,38]
Decline in physical health	[35]
Technological barriers	
Functionality	
Cumbersome design	[31,33-35,37-39]
Technological problems with the devices	[31-35,38,39]
Lack of desired functions	[31,33-35,38,39]
Not age-friendly technology	[31,33-35,37,38]
Difficult to use	[31,33-35,38,39]
Unreliable	[31,33,35,36,39]
Complicated tablet or smartphone	[31,33-35]
Requires a smartphone	[31,34]
Not helpful in troubleshooting	[39]
Poor, absent, or negative feedback from the device	[31]
Content	
Not capturing all activities	[31,33,35]
Automatic targeting function	[31]
Design	
Large and rigid band	[31,33,35,38]
Frequent loading	[31,34,38]
Privacy	[33,35,38]
Fragile	[35,37,38]
Uncertainties about water or load damage	[31,35,39]
Difficult to put on	[31,35,36]
Big	[35,38]
Discomfort during nocturnal wear	[31,35]
Looks like a medical device (aesthetics)	[31]
Does not match clothing	[31]

XSL•FO
RenderX

Categories, subcategories, and codes	References
Complicated interaction when on the ankle	[31]
Alarms	
Frequent alarms or notifications	[31,34]
Availability	
High price	[33,34]
Accessibility	
Access to instructions, lack of training, and technicalities	[31,35,37-39]
Lack of practical training	[31,33]
Environmental barriers	
Seasonal variations	
Climatic conditions	[37,40]

[a]mHealth: mobile health.

XSL•FO
RenderX

Table . Perceived facilitators to the use of wearable mHealth[a] technologies for physical activity promotion by older adults (categories, subcategories, and codes).

Categories, subcategories, and codes	References
Personal facilitators	
Believe it improves health	
Improved physical activity	[33,35,39]
Stress reduction	[33,39]
Pain reduction	[33,39]
Improved health and fitness	[31-40]
Perceptions about technology	
Intrinsic motivation	[31-40]
Ability to use the device	[31,33-36,38]
Previous experience using technology	[31,33-35,38]
Confidence in the ability to remember how the technology works	[31,34,35,38]
Higher levels of actual physical activity than perceived by the technology	[31,33-40]
Personal preferences	
Positive attitude toward technology use	[31,34-40]
Achievement of personal goals	[31,32,35,37,39,40]
Perceived usefulness of technology	[31,33-35,38,39]
Emotional attachment to devices	[31,34,35,39]
Curiosity and interest in technology use	[31,35,38]
Including technology in users' daily routine	[32,38]
Technological facilitators	
Functionality	
Receive positive, real-time feedback from the device	[31-35,37-40]
Easy to use	[31-35,38,39]
Smart	[31,33,35,37,38]
Large, easy-to-read display	[31,33,37,38]
Automatic activity recording	[31,33,35,38]
Compatible with other daily life activities	[31,35,39]
Easy synchronization	[31,33]
Simple application	[31]
Large, easy-to-press buttons	[31]
Easy to see	[31]
Content	
Health-related features	[31,33-35,37,39]
Customizable notifications or alarms	[31-35,37]
Goal tracking	[31-33,35,38]
Step counting	[31,35,37,38]
Other device features (light and stopwatch)	[31,33,40]
Sleep tracking	[31,33,38]
Global Positioning System	[31]
Help section	[31]

XSL•FO
RenderX

Categories, subcategories, and codes	References
Design	
Comfortable location on the body and comfortable to wear	[31,37,38]
Small battery and longer battery life	[31,33,38]
Thin and flexible band	[31,33,38]
Pleasant or cool appearance	[31,33,38]
Water resistant	[31,33]
Device looks like a watch	[31,33]
Smaller design	[31,35]
Secure attachment	[31]
Relational facilitators	
Technological support	
Instructions and help to solve problems	[31]
Social support	
Support from family and friends	[31,33,34,38-40]
Peer support, interaction, and communication	[31,33,34,39,40]
Environmental facilitators	
Seasonal variations	
Weather conditions	[33,40]
Health care professionals	
Support and follow-up of health professionals	
Being a tool for health care professionals for follow-up and monitoring	[31,36-38,40]
Patient benefits due to the involvement of health care professionals	[31,36-38,40]
Being part of the treatment plan	[31,32,35,37,39]
Visualization of health information	[31,33-35]
Care team support	[31,32,37,40]
Sharing data with health care professionals	[33,35,38]

[a]mHealth: mobile health.

XSL•FO
RenderX

Table . Areas of improvement for the use of wearable mHealth[a] technologies for the promotion of physical activity by older adults (categories, subcategories, and codes).

Categories, subcategories, and codes	References
Personal areas	
Perceptions about technology	
Intrinsic motivation	[36]
Tastes and preferences	
Data visualization	[33-35,40]
Device feedback	[34,39]
Security	[37,38]
Personal goals	[34]
Attractiveness	[38]
Technological areas	
Functionality	
Age-friendly devices	[31,36,37,39]
Device reliability	[31,34]
Intelligibility	[31,38]
Synchronization with the smartphone	[33]
Large screen format and clear reading	[37]
Content	
Include health-related content	[34,35,39]
Monitoring of other health parameters	[34]
Usability	[38]
Design	
Instructions in native language	[35,39]
Comfortable devices	[31,38]
Sufficient battery life	[31,35]
Relational areas	
Technology support	
Simple, nontechnical instructions	[31,33,37]
Previous training	[31]

[a]mHealth: mobile health.

Barriers to the Use of Wearable mHealth Technologies for PA Promotion

Through the reciprocal translation of the primary literature, the synthesis demonstrates that mHealth ecosystem adoption among older adults is structurally hindered by an intersection of age-associated physiological decline and inadequate user-centered design. Personal barriers encompass sensory and physical deterioration, slower information processing speeds, and technology-induced anxiety, which frequently collide with an established reluctance to alter daily habits [31-35,37-40]. This vulnerability is severely compounded by technical deficits; users routinely experience frustration regarding cumbersome and fragile hardware, complex smartphone synchronization requirements, a lack of automated troubleshooting, and unreliable data visualization [31-39]. Furthermore, frequent notifications or alarms act as intrusive stressors [31,34], while high purchasing costs [33,34] and jargon-heavy instructions create severe access barriers [31,33,35,37-39]. These technical challenges are further compounded by environmental variables, as adverse seasonal weather (winter, rain, snow, and so on) restricts physical outdoor use [37,40]. These barriers are detailed in Table 2.

Facilitators for the Use of Wearable mHealth Technologies for PA Promotion

Conversely, a positive line of argument emerges when wearable technologies successfully bridge the gap between intrinsic health motivation and intuitive functionality and desire for functional independence [31-40]. At the individual level, older adults demonstrate high adherence when they perceive that tracking tools directly translate into stress reduction, pain management,

XSL•FO
RenderX

and measurable physical wellness [32-40]. From a technological standpoint, sustained engagement is facilitated by seamless, unobtrusive hardware configurations—specifically watch-like, waterproof devices featuring long battery life, flexible bands, and large, legible displays [31-35,37-40]. Software features that offer automated tracking, personalized goals, gamified self-competition, and real-time positive feedback act as powerful behavioral catalysts [31-35,37-40]. Critically, the synthesis highlights that hardware and software fail in isolation; relational and systemic scaffolding is paramount. Sustained use is driven by robust social networks (family and peers) [31,33,34,38-40] and, fundamentally, by the formal involvement of health care professionals who integrate these digital metrics into active clinical treatment plans and continuous medical monitoring [31-33,35-38,40]. Table 3 maps the comprehensive structure of these facilitators.

Areas of Improvement for the Use of Wearable mHealth Technologies for PA Promotion

To minimize early abandonment and maximize clinical utility, the synthesized data establishes explicit pathways for age-inclusive development. Participants universally demand a shift toward age-friendly engineering that prioritizes intuitive interfaces, native-language instruction manuals, enhanced data privacy, and the expansion of monitored metrics to include broader clinical parameters beyond basic pedometry [31,33-39]. Relatably, older adults highlight that closing the digital divide requires robust instructional scaffolding. Developers and health care networks must cooperate to provide accessible, nontechnical training and ongoing troubleshooting support to foster digital self-efficacy and prevent technology rejection [31,33,37]. These strategic directions for technological and clinical adaptation are synthesized in Table 4.

Results of the Analysis of the Quality of the Studies Analyzed

The methodological quality of the articles was analyzed with the JBI-QARI [30]. Table 5 shows the detailed results of the analysis of the methodological quality of the included studies after this tool was applied.

XSL•FO
RenderX

Table . Analysis of the quality of the studies analyzed with the Joanna Briggs Institute Qualitative Assessment and Review Instrument.

Author/s	Is there congruity between the stated philosophical perspective and the research methodology?	Is there congruity between the research methodology and the research question or objectives?	Is there congruity between the research methodology and the methods used to collect data?	Is there congruity between the research methodology and the representation and analysis of data?	Is there congruity between the research methodology and the interpretation of results?	Is there a statement locating the researcher culturally or theoretically?	Is the influence of the researcher on the research, and vice-versa, addressed?	Are participants, and their voices, adequately represented?	Is the research ethical according to current criteria or, for recent studies, and is there evidence of ethical approval by an appropriate body?	Do the conclusions drawn in the research report flow from the analysis, or interpretation, of the data?	Total
Moore et al 2021 [31]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10/10
Peng et al 2021 [32]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10/10
Kononova et al 2019 [33]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10/10
Abouzahra and Ghasemaghaei 2020 [34]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10/10
Ehn et al 2018 [35]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10/10
Fausset et al 2013 [36]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10/10
Mercer et al 2016 [37]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10/10
Puri et al 2017 [38]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10/10
Schlomann et al 2016 [39]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10/10
Batsis et al 2016 [40]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10/10

Discussion

Principal Results

The objective of this qualitative systematic review was to analyze the point of view of older adults on the aspects that can influence the use of wearable mHealth technologies for the promotion of PA and to analyze their demands, contexts, and desires to develop more accessible, age-friendly, and acceptable technologies for this population group [19].

People aged 50 years and older perceive personal barriers (physical aspects, perceptions about technologies, and tastes or preferences), technological barriers (functionality, content, design, alarms, availability, and accessibility), and environmental barriers (seasonal variations) that hinder the use of wearable mHealth technologies for the promotion of PA [31-38,40,41]. On the other hand, this population includes personal facilitators (which consider that it improves health, perceptions about technologies, and tastes or preferences), technological facilitators (functionality, content, and design), relational facilitators (technological support and social support), environmental facilitators (seasonal variations), and health professionals (support and follow-up by health professionals) that favor the use of wearable mHealth technologies [31-38,40,41]. Likewise, people more than 50 years of age perceive personal factors (perceptions about technologies and tastes or preferences), technological factors (functionality, content, and design), and relational factors (technological

XSL•FO
RenderX

support) as areas for improvement in mHealth devices and apps for the promotion of PA [31,33-38,40,41].

To illustrate the practical relevance and tangible impact of these technologies, specific mHealth interventions from synthesized studies demonstrate how targeted features can drive positive behavioral changes in vulnerable cohorts. For instance, the implementation of a commercial wearable device (Fitbit) among older adults with obesity residing in rural areas proved highly acceptable; it successfully motivated positive behavior change by providing useful real-time feedback and fostering social interaction [40]. Similarly, in a cohort of older adults living with chronic illnesses, wearable activity trackers functioned as powerful behavioral catalysts [37]. By using alert functions and setting daily step goals, these devices gamify PAs, significantly increasing users' awareness of their daily movement and transforming exercise into self-competition [37]. Furthermore, devices that allowed users to set small, progressive goals, supported by reminders and structured routines—such as wearing the device first thing in the morning—were instrumental in helping older adults develop long-term PA habits [32]. The increased self-awareness facilitated by these trackers also motivated older adults to make broader lifestyle changes and proactively seek health-related information [38]. These empirical examples underscore that when wearable mHealth technologies successfully integrate user-friendly feedback, gamification, and social support, they effectively translate theoretical facilitators into measurable health impacts for aging populations [32,37,38,40].

Comparison With Prior Work

In line with previous studies that have analyzed the use of wearable mHealth technologies among older adults [4,42], our results confirm that the following barriers hinder their use: physiological state, age, vision loss, hearing loss [31], high price [33,34], lack of interest in older adults, and inexperience and fear of using technologies and sharing data [33-35,37,38]. In addition, the results of this review coincide with those of previous studies showing the following facilitators for the use of devices and apps among older adults: previous experience with technology [32,34,38], being user-friendly [31,35], and favoring the independence and autonomy of older adults [33,35,38].

Another previous study performed on people with low back pain [43] confirmed, as in our review, that older adults must be supported by family members, friends, health professionals, and other users to be able to continuously use PA monitoring devices and apps [31,33,34,37,39,40]. In addition, older adults positively value the use of reminders and alarms [32-35,37] and receive feedback on whether they are using the device correctly [38,40]. Our results are in line with those of previous studies analyzing the use of mHealth ecosystems to promote PA in patients with myeloid neoplasms [25], revealing the need to establish specific objectives according to physiological conditions so that users can use PA monitoring devices appropriately [31,32,35].

In contrast to our findings [31,33,38,39], a study involving adolescent and young adult cancer survivors reported that PA monitoring devices were easy to use and synchronize, allowing for quick data retrieval [44]. Such ease of use is not reflected in the narratives of older adults in our review. Instead, older participants frequently expressed the belief that these technologies were not designed for them [31,35,39]. Moreover, in contrast to studies that claim that devices and apps are affordable and economically profitable [33,34], our results show that cost and access to mHealth technology are barriers for older adults [31,33-35,37,38,41]. These results highlight the need for mHealth technology to be friendlier to older adults, which has been termed age-friendly, and to be more flexible to potential users. One study reported that mHealth devices and apps should contain simple functions and simple navigation instructions to reduce the cognitive load of older adults, common and simple language, the option of zoom or large print text, and voice navigation (if they are visually impaired) [45]. These findings are consistent with the results of our studies [31-35,37-39].

For these reasons, governments, international organizations, health professionals, academic institutions, media, and the private sector must consider all these findings to eliminate the existing digital gap between older adults and the use of mHealth devices and apps. In this way, through learning and training, older adults can benefit from the use of wearable mHealth technologies by achieving better self-management of their health [46].

However, advocating for inclusive design is insufficient without addressing the economic realities of the digital health ecosystem. The technology sector generally recognizes the value of user-centered interfaces; rather, the primary barrier is economic. The high costs and lower profit margins associated with developing customized platforms often fail to support a compelling business case. To genuinely drive industry change, future research must actively develop this business case. By leveraging resources largely inaccessible to private developers—such as longitudinal electronic health records, clinical registries, and biobanks—the academic community must conduct robust Health Economics and Outcomes Research and budget impact analyses. Demonstrating a clear clinical and economic return on investment is essential for incentivizing the tech sector to prioritize and fund highly customizable, age-friendly mHealth technologies.

Limitations

This review included only articles in English and Spanish in the selected databases and excluded relevant articles on the phenomenon studied. In addition, no articles have studied the long-term use of wearable mHealth technologies among older adults, nor have the views of health professionals been considered. On the other hand, our study was based solely on the use of mHealth ecosystems that promote PA, and future studies should continue to investigate the repercussions of these devices and apps over time and from the point of view of professionals.

Another limitation of this meta-ethnography is the restricted demographic reporting in the primary literature. Although some studies detailed clinical parameters and geographic contexts, critical sociodemographic determinants such as SES, culture, and ethnicity were systematically omitted. This paucity of granular data precludes a robust intersectional analysis,

XSL•FO
RenderX

obscuring how overlapping social identities and structural inequalities influence mHealth engagement. Consequently, the current evidence may disproportionately reflect privileged cohorts and fail to capture the unique barriers faced by marginalized subpopulations. To advance health equity, future research must rigorously report diverse, intersectional demographic variables, ensuring that mHealth interventions mitigate rather than exacerbate the digital divide among aging populations.

Conclusions

This meta-ethnography demonstrates that older adults' engagement with wearable mHealth technologies for PA is not solely dictated by technological proficiency but by a complex interplay of personal, technological, and environmental factors. While usability constraints, age-related physical limitations, and privacy concerns significantly hinder adoption, older adults clearly recognize the potential of these digital health tools to enhance their well-being. Evidence from the synthesized cohorts highlights that when mHealth interventions use user-centered, age-friendly designs and integrate continuous support from health care professionals, they effectively drive sustained behavioral change and long-term adherence. However, a critical methodological gap persists in the current literature: the systemic omission of granular demographic data, specifically SES, cultural background, and ethnicity. This lack of intersectional reporting precludes a comprehensive understanding of how compounding structural inequalities shape digital health access. If mHealth is to promote true health equity, future research must transition from homogeneous sampling to rigorous, inclusive methodologies. Tech developers and clinicians must collaboratively address the multifaceted needs of diverse, marginalized older adult subpopulations. Only through this intersectional approach can we ensure that the proliferation of wearable mHealth technologies bridges, rather than inadvertently exacerbates, the existing digital divide.

Acknowledgments

No AI tools were used in the preparation of the manuscript.

Funding

This research has been funded by grant number 2025-GRIN-38519 from the University of Castilla-La Mancha and the European Regional Development Fund (FEDER).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 20 KB - mhealth_v14i1e67157_app1.docx]

Multimedia Appendix 2
Examples of first-order, second-order, and third-order constructs.
[DOCX File, 17 KB - mhealth_v14i1e67157_app2.docx]

Checklist 1
PRISMA 2020 checklist.
[DOCX File, 275 KB - mhealth_v14i1e67157_app3.docx]

References

1. World Health Organization, Noncommunicable Diseases and Mental Health Cluster, Department of Noncommunicable Disease Prevention and Health Promotion Ageing and Life Course. Active Ageing: A Policy Framework. Rev Esp Geriatría Gerontol 2002;37:74-105.
2. Núñez de Arenas-Arroyo S, Cavero-Redondo I, Alvarez-Bueno C, Sequí-Domínguez I, Reina-Gutiérrez S, Martínez-Vizcaíno V. Effect of eHealth to increase physical activity in healthy adults over 55 years: a systematic review and meta-analysis. Scand J Med Sci Sports 2021 Apr;31(4):776-789. [doi: 10.1111/sms.13903] [Medline: 33280182]
3. Buyl R, Beogo I, Fobelets M, et al. e-Health interventions for healthy aging: a systematic review. Syst Rev 2020 Jun 3;9(1):128. [doi: 10.1186/s13643-020-01385-8] [Medline: 32493515]
4. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health 2021 Aug 17;21(1):1556. [doi: 10.1186/s12889-021-11623-w] [Medline: 34399716]
5. Latorre Santos C. The aging of the population. Opportunities and challenges. Revista Ciencias de la Salud 2019;17(3):6-8. [doi: 10.12804/revistas.urosario.edu.co/revsalud/a.8347]

XSL•FO
RenderX

6. Pinilla Cárdenas MA, Ortiz Álvarez MA, Suárez-Escudero JC. Older adult: aging, disability, care, and day centers. A Review Revisión de tema Salud Uninorte 2022;37(2):488-505. [doi: 10.14482/sun.37.2.618.971]
7. Puig Vela JM, Adell Aparicio MC, Prat Marín A, Oromí Durich J. Population ageing as a health problem [Article in Spanish]. Med Integral 2000;36(5):190 [FREE Full text]
8. The 2018 aging report: economic & budgetary projections for the 28 EU member states (2016-2070). : European Union; 2018 URL: https://data.europa.eu/doi/10.2765/615631 [accessed 2026-06-29]
9. Envejecimiento y bienestar: una radiografía de las personas mayores. Oficina de Ciencia y Tecnología del Congreso de los Diputados (Oficina C). 2023. URL: https://oficinac.es/informes-c/envejecimiento [accessed 2026-06-29]
10. Kwan RYC, Salihu D, Lee PH, et al. The effect of e-health interventions promoting physical activity in older people: a systematic review and meta-analysis. Eur Rev Aging Phys Act 2020;17(1):7. [doi: 10.1186/s11556-020-00239-5] [Medline: 32336996]
11. Sun F, Norman IJ, While AE. Physical activity in older people: a systematic review. BMC Public Health 2013 May 6;13(1):449. [doi: 10.1186/1471-2458-13-449] [Medline: 23648225]
12. Hosseinpour M, Terlutter R. Your personal motivator is with you: a systematic review of mobile phone applications aiming at increasing physical activity. Sports Med 2019 Sep;49(9):1425-1447. [doi: 10.1007/s40279-019-01128-3] [Medline: 31144235]
13. WHO Guidelines on Physical Activity and Sedentary Behaviours: World Health Organization; 2021. URL: https://www.ncbi.nlm.nih.gov/books/NBK581973 [accessed 2026-06-29]
14. Muellmann S, Forberger S, Möllers T, Bröring E, Zeeb H, Pischke CR. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: a systematic review. Prev Med 2018 Mar;108:93-110. [doi: 10.1016/j.ypmed.2017.12.026] [Medline: 29289643]
15. Bernardo J, Apóstolo J, Loureiro R, et al. eHealth platforms to promote autonomous life and active aging: a scoping review. Int J Environ Res Public Health 2022 Nov 29;19(23):15940. [doi: 10.3390/ijerph192315940] [Medline: 36498018]
16. Robbins TD, Lim Choi Keung SN, Arvanitis TN. E-health for active ageing; a systematic review. Maturitas 2018 Aug;114:34-40. [doi: 10.1016/j.maturitas.2018.05.008] [Medline: 29907244]
17. Cortes Restrepo NV. Apps e-health en los procesos de enseñanza aprendizaje en enfermería. REDIIS Rev Investig E Innov En Salud 2020;3:100-111. [doi: 10.23850/rediis.v3i3.2980]
18. Bidargaddi N, Schrader G, Klasnja P, Licinio J, Murphy S. Designing m-Health interventions for precision mental health support. Transl Psychiatry 2020 Jul 7;10(1):222. [doi: 10.1038/s41398-020-00895-2] [Medline: 32636358]
19. Lorenzi LJ, Belo LF, Frohlich DM, Dourado VZ, Castro PC, Gomes GDO. Factors related to the adoption and adherence of physical activity mobile applications by older people: a scoping review protocol. BMJ Open 2021 Oct 8;11(10):e052414. [doi: 10.1136/bmjopen-2021-052414] [Medline: 34625417]
20. Kampmeijer R, Pavlova M, Tambor M, Golinowska S, Groot W. The use of e-health and m-health tools in health promotion and primary prevention among older adults: a systematic literature review. BMC Health Serv Res 2016 Sep 5;16 Suppl 5(Suppl 5):290. [doi: 10.1186/s12913-016-1522-3] [Medline: 27608677]
21. Padial Espinosa M, Pinzón Pulido S, Espinosa Almendro JM, Kalache A, Garrido Peña F. Longevity and digital revolution. The contribution of information and communication technologies to healthy aging. Gerokomos 2020;31(1):21-26. [doi: 10.4321/S1134-928X2020000100003]
22. Gustafson DH Sr, McTavish F, Gustafson DH Jr, et al. The effect of an information and communication technology (ICT) on older adults' quality of life: study protocol for a randomized control trial. Trials 2015 Apr 25;16(1):191. [doi: 10.1186/s13063-015-0713-2] [Medline: 25909465]
23. Kalem G, Turhan Ç. Mobile technology applications in the healthcare industry for disease management and wellness. Procedia Soc Behav Sci 2015 Jul;195:2014-2018. [doi: 10.1016/j.sbspro.2015.06.216]
24. Visier-Alfonso ME, Sánchez-López M, Rodríguez-Martín B, et al. Parents' perceptions of children's and adolescents' use of electronic devices to promote physical activity: systematic review of qualitative evidence. JMIR Mhealth Uhealth 2023 Jul 20;11:e44753. [doi: 10.2196/44753] [Medline: 37471127]
25. Loh KP, Sanapala C, Di Giovanni G, et al. Developing and adapting a mobile health exercise intervention for older patients with myeloid neoplasms: a qualitative study. J Geriatr Oncol 2021 Jul;12(6):909-914. [doi: 10.1016/j.jgo.2021.02.023] [Medline: 33676900]
26. Nguyen NH, Hadgraft NT, Moore MM, et al. A qualitative evaluation of breast cancer survivors' acceptance of and preferences for consumer wearable technology activity trackers. Support Care Cancer 2017 Nov;25(11):3375-3384. [doi: 10.1007/s00520-017-3756-y] [Medline: 28540402]
27. Noblit G, Hare R. Meta-ethnography. In: Synthesizing Qualitative Studies: Sage; 1988. [doi: 10.4135/9781412985000]
28. Kolaski K, Logan LR, Ioannidis JPA. Guidance to best tools and practices for systematic reviews. JBJS Rev 2023 Jun 1;11(6):00077. [doi: 10.2106/JBJS.RVW.23.00077] [Medline: 37285444]
29. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]

XSL•FO
RenderX

30. Lockwood C, Munn Z, Porritt K. Qualitative research synthesis: methodological guidance for systematic reviewers utilizing meta-aggregation. Int J Evid Based Healthc 2015 Sep;13(3):179-187. [doi: 10.1097/XEB.0000000000000062] [Medline: 26262565]
31. Moore K, O'Shea E, Kenny L, et al. Older adults' experiences with using wearable devices: qualitative systematic review and meta-synthesis. JMIR Mhealth Uhealth 2021 Jun 3;9(6):e23832. [doi: 10.2196/23832] [Medline: 34081020]
32. Peng W, Li L, Kononova A, Cotten S, Kamp K, Bowen M. Habit formation in wearable activity tracker use among older adults: qualitative study. JMIR Mhealth Uhealth 2021 Jan 19;9(1):e22488. [doi: 10.2196/22488] [Medline: 33464216]
33. Kononova A, Li L, Kamp K, et al. The use of wearable activity trackers among older adults: focus group study of tracker perceptions, motivators, and barriers in the maintenance stage of behavior change. JMIR Mhealth Uhealth 2019 Apr 5;7(4):e9832. [doi: 10.2196/mhealth.9832] [Medline: 30950807]
34. Abouzahra M, Ghasemaghaei M. The antecedents and results of seniors' use of activity tracking wearable devices. Health Policy Technol 2020 Jun;9(2):213-217. [doi: 10.1016/j.hlpt.2019.11.002]
35. Ehn M, Eriksson LC, Åkerberg N, Johansson AC. Activity monitors as support for older persons' physical activity in daily life: qualitative study of the users' experiences. JMIR Mhealth Uhealth 2018 Feb 1;6(2):e34. [doi: 10.2196/mhealth.8345] [Medline: 29391342]
36. Fausset CB, Mitzner TL, Price CE, Jones BD, Fain WB, Rogers WA. Older adults' use of and attitudes toward activity monitoring technologies. Proc Hum Factors Ergon Soc Annu Meet 2013 Sep;57(1):1683-1687. [doi: 10.1177/1541931213571374] [Medline: 31263349]
37. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: a mixed-methods evaluation. JMIR Mhealth Uhealth 2016 Jan 27;4(1):e7. [doi: 10.2196/mhealth.4225] [Medline: 26818775]
38. Puri A, Kim B, Nguyen O, Stolee P, Tung J, Lee J. User acceptance of wrist-worn activity trackers among community-dwelling older adults: mixed method study. JMIR Mhealth Uhealth 2017 Nov 15;5(11):e173. [doi: 10.2196/mhealth.8211] [Medline: 29141837]
39. Schlomann A, von Storch K, Rasche P, Rietz C. Means of motivation or of stress? The use of fitness trackers for self-monitoring by older adults. HBScience 2016 Aug;7(3):111-116. [doi: 10.1007/s16024-016-0275-6]
40. Batsis JA, Naslund JA, Gill LE, Masutani RK, Agarwal N, Bartels SJ. Use of a wearable activity device in rural older obese adults: a pilot study. Gerontol Geriatr Med 2016;2:2333721416678076. [doi: 10.1177/2333721416678076] [Medline: 28138502]
41. Schlomann A. A case study on older adults' long-term use of an activity tracker. Gerontechnology 2017 Aug 9;16(2):115-124. [doi: 10.4017/gt.2017.16.2.007.00]
42. Airola E. Learning and use of eHealth among older adults living at home in rural and nonrural settings: systematic review. J Med Internet Res 2021 Dec 2;23(12):e23804. [doi: 10.2196/23804] [Medline: 34860664]
43. Palazzo C, Klinger E, Dorner V, et al. Barriers to home-based exercise program adherence with chronic low back pain: patient expectations regarding new technologies. Ann Phys Rehabil Med 2016 Apr;59(2):107-113. [doi: 10.1016/j.rehab.2016.01.009] [Medline: 27050664]
44. Mendoza JA, Baker KS, Moreno MA, et al. A Fitbit and Facebook mHealth intervention for promoting physical activity among adolescent and young adult childhood cancer survivors: a pilot study. Pediatr Blood Cancer 2017 Dec;64(12). [doi: 10.1002/pbc.26660] [Medline: 28618158]
45. Ramdowar H, Khedo KK, Chooramun N. A comprehensive review of mobile user interfaces in mHealth applications for elderly and the related ageing barriers. Univ Access Inf Soc 2024 Nov;23(4):1613-1629. [doi: 10.1007/s10209-023-01011-z]
46. González Oñate C, Fanjul Peyró C. Aplicaciones móviles para personas mayores: un estudio sobre su estrategia actual [Article in Spanish]. Aula Abierta 2018;47(1):107. [doi: 10.17811/rifie.47.1.2018.107-112]

Abbreviations

ENTREQ: Enhancing Transparency in Reporting the Synthesis of Qualitative Research
JBI-QARI: Joanna Briggs Institute Qualitative Assessment and Review Instrument
mHealth: mobile health
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SES: socioeconomic status
WHO: World Health Organization

XSL•FO
RenderX

Edited by G Badicu; submitted 04.Oct.2024; peer-reviewed by B Wolff, KL Mauco, K Law, V Martinez-Vizcaino; revised version received 25.May.2026; accepted 27.May.2026; published 29.Jul.2026.

Please cite as:
Rodríguez-Martín B, González-Arroyo D, Priego-Jiménez S, Lopez-González M, Medrano-Echeverría M, Ruiz-Grao MC, Álvarez-Bueno C
Perceptions of Adults Aged 50 Years and Older Regarding the Use of Wearable mHealth Technologies to Promote Physical Activity: Systematic Review and Meta-Ethnography
JMIR Mhealth Uhealth 2026;14:e67157
URL: https://mhealth.jmir.org/2026/1/e67157
doi:10.2196/67157

© Beatriz Rodríguez-Martín, Diego González-Arroyo, Susana Priego-Jiménez, María Lopez-González, María Medrano-Echeverría, Marta Carolina Ruiz-Grao, Celia Álvarez-Bueno. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 29.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Patient Acceptance and Barriers to IoT Usage in Health Care: Systematic Literature Review

Zaenal Arifin, MTI; Putu Wuri Handayani, Prof Dr

Doktor Ilmu Komputer, Faculty of Computer Science, University of Indonesia, Fakultas Ilmu Komputer, Kampus UI Depok, Depok, Jawa Barat, Indonesia

Corresponding Author:
Zaenal Arifin, MTI
Doktor Ilmu Komputer, Faculty of Computer Science, University of Indonesia, Fakultas Ilmu Komputer, Kampus UI Depok, Depok, Jawa Barat, Indonesia

Abstract

Background: The Internet of Things (IoT) represents a transformative paradigm in health care service delivery, offering unprecedented potential to enhance quality of care, operational efficiency, and patient outcomes through interconnected devices and real-time data analytics. Despite rapid adoption, implementation depends on patient acceptance as end users. While literature extensively documents technical and institutional perspectives, a comprehensive understanding of patient acceptance factors remains fragmented, with high technology abandonment rates. Systematic synthesis of patient perspectives is critically needed to inform user-centered design, effective implementation strategies, and supportive policy frameworks.

Objective: This systematic literature review aims to identify and synthesize factors influencing patient acceptance of IoT technology in health care services, barriers hindering adoption, and effective strategies for enhancing acceptance.

Methods: Following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines, we systematically searched eight electronic databases (PubMed/MEDLINE, Scopus, IEEE Xplore, Web of Science, ScienceDirect, ACM Digital Library, ProQuest, and Google Scholar) for empirical studies published between January 2016 and December 2024. Inclusion criteria encompassed peer-reviewed empirical research examining patient perspectives on IoT technology in health care services, published in English or Indonesian. From 2537 initially identified papers, 62 studies met inclusion criteria after systematic screening and full-text evaluation. Quality assessment was conducted using the Mixed Methods Appraisal Tool.

Results: The 62 included studies represented diverse geographic contexts (Asia, Europe, North America, and the Middle East) and methodological approaches. Quality assessment revealed 45 (73%) studies of good-to-excellent quality. Perceived usefulness emerged as the strongest acceptance facilitator, identified in 55 of 62 (89%) studies, followed by perceived ease of use in 47 (76%) studies, and trust and security in 42 (68%) studies. Cost-effectiveness was identified in 32 (52%) studies as an important consideration. Primary barriers included data security concerns in 26 (42%) studies, privacy issues in 24 (39%) studies, lack of digital literacy in 22 (36%) studies, and resistance to change in 20 (32%) studies. Interoperability issues and high costs were identified in 19 (31%) studies and 18 (29%) studies, respectively. User-centered design was the most frequently recommended enhancement strategy in 20 (32%) studies, followed by user-friendly interface development in 19 (31%) studies, digital literacy programs in 18 (29%) studies, and health care professional involvement in 15 (24%) studies. Digital literacy functioned as a significant moderator, and trust served as both direct predictor and mediator.

Conclusions: Patient acceptance of IoT in health care represents a complex, multidimensional phenomenon requiring holistic approaches that integrate user-centered technology design, comprehensive digital literacy programs, trust-building mechanisms, and supportive policy frameworks. Successful implementation necessitates multilevel strategies addressing individual, organizational, and system factors simultaneously. With evidence-based, patient-centered approaches, IoT technology holds substantial potential to transform health care delivery into more proactive, personalized, and accessible services.

(JMIR Mhealth Uhealth 2026;14:e81260) doi:10.2196/81260

KEYWORDS

internet of things; healthcare IoT; technology acceptance; patient acceptance; patient-centric; systematic review; telemedicine; eHealth; mHealth; smart hospital; mobile health; mobile phone

XSL•FO
RenderX

Introduction

Background

The technological developments of the coming decade demonstrate fundamental transformations that will radically change how humans live, work, and interact. According to the McKinsey Global Institute (2024) report [1], waves of technological innovation represent not merely incremental changes, but a new paradigm that will integrate AI, quantum computing, and sustainable technologies into almost every aspect of human life. This condition is driven by multidisciplinary convergence between computer science, biotechnology, nanotechnology, and AI, opening new possibilities previously considered impossible. Sustainable technology has become a central paradigm in future innovation development, with renewable energy, green computing, and environmental solutions becoming the primary focus of global research. The Intergovernmental Panel on Climate Change (IPCC) report [2] confirms that technological innovation will be key in climate change mitigation, with projections of carbon emission reductions of up to 45% through green technology implementation by 2040. Based on the context of sustainable technology as the central paradigm of future innovation, the health care sector faces urgent real challenges that require resolution through modern technology implementation. One major problem currently faced by the health care sector is the increasing older adult population experiencing multiple chronic conditions (MCCs) such as hypertension, arthritis, heart disease, cancer, diabetes, and chronic kidney disease. This multimorbidity condition impacts increased mortality, disability, decreased physical function, and reduced patient quality of life [3]. Future technology trends addressing these issues include the internet of medical things (IoMT) for continuous monitoring [4] and expansion of telemedicine and telehealth, enabling remote consultations, reducing patient waiting times, and expanding access to specialists [5]. This digital transformation specifically targets reducing the older adult population health care costs while improving their outcomes and quality of life through personal and sustainable technology solutions.

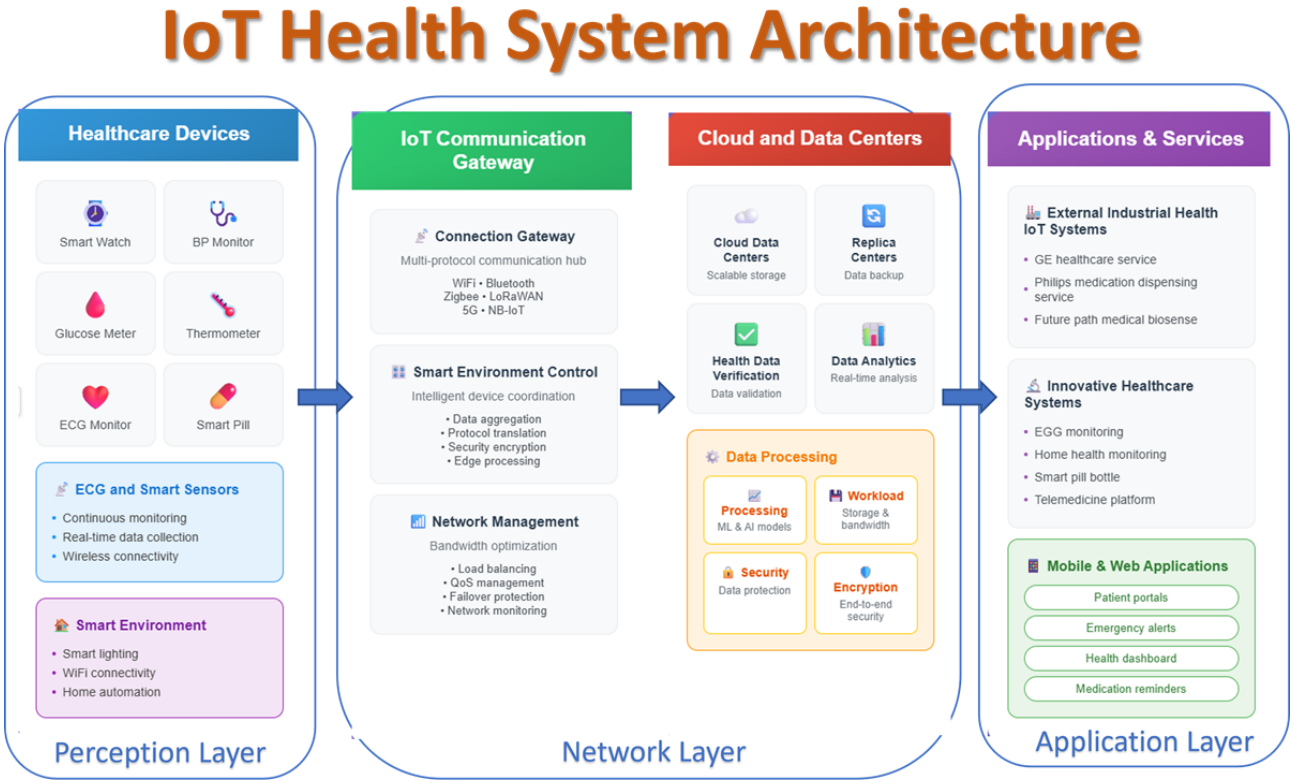
The rapid advancement of digital technology in recent decades has transformed various aspects of human life, including the health care sector. One of the most promising innovations in health care service transformation is the Internet of Things (IoT). IoT presents significant paradigm changes in health care service delivery, driven by the convergence of interconnected devices, data analytics, and communication networks. This technological synergy enables real-time patient health monitoring, remote chronic condition management, and improved efficiency in health care service operations. IoT infrastructure facilitates communication between various entities, including individuals and medical devices such as wireless sensors, thereby promoting increased access to quality and affordable health care services [6]. IoT implementation spans various health care service domains, offering innovative solutions for remote patient monitoring, medication adherence, and chronic disease management [7].

Although AI offers tremendous potential in the health care sector, the focus on IoT stems from its direct and practical applications in addressing current challenges in health care systems [8]. IoT provides basic infrastructure for collecting, transmitting, and analyzing patient data, building foundations for more sophisticated AI applications [9]. Integration of IoT devices, such as wearable sensors and remote monitoring systems, facilitates continuous data collection about physiological parameters, activity levels, and environmental conditions [10]. This real-time data flow enables proactive monitoring and early detection of health anomalies, reducing the need for frequent hospital visits and improving patient outcomes [10].

IoT refers to networks of physical devices connected through the Internet, enabling real-time data collection and exchange [11]. In academic contexts [11], IoT is described as a computing paradigm encompassing three main dimensions: communication, semantics, and interconnectivity. The communication dimension refers to devices' ability to exchange information through diverse network protocols. The semantic dimension focuses on interpretation and use of meaningful data, while interconnectivity refers to devices' ability to connect and interact in complex and dynamic ecosystems. In health care service contexts, IoT implementation (Figure 1) offers great potential for improving care quality, operational efficiency, and overall patient experience [12].

XSL•FO
RenderX

Figure 1. IoT health system architecture. BP: blood pressure; ECG: electrocardiogram; GE: General Electric; IoT: Internet of Things; QoS: quality of service.



The development of health care IoT market share shows very significant growth with latest data from Transforma Insights reporting that by the end of 2023 there were 16.1 billion active IoT devices worldwide, with growth projections reaching 399 billion devices in 2033 with a compound annual growth rate of 10%, while the total IoT market including connectivity modules and related core applications was valued at US $335 billion in 2023 and projected to increase to US $934 billion in 2033 [13]. Health care IoT adoption acceleration has been further accelerated by COVID-19 pandemic impacts, emphasizing urgent needs for remote health care solutions and digital technologies that can support health care service continuity in situations of physical and social restrictions [14].

Despite the transformative potential of IoT technology in health care services being extensively documented and discussed from technical, clinical, and institutional perspectives in various scientific literature, a comprehensive understanding of acceptance from patient perspectives as end users still shows significant limitations requiring deeper investigation [15]. This paradoxical phenomenon is reflected in high abandonment rates of digital health technologies, where various studies show that up to 50% of wearable health device users discontinue use within the first six months after initial adoption, indicating substantial gaps between the technical advancement of IoT devices and technology's ability to integrate seamlessly into users' daily lives and routines [16].

Health care IoT technology acceptance complexity is influenced by broad and multidimensional factor spectrums extending beyond mere technical considerations, including benefit perceptions related to health outcomes and quality of life improvements, ease of use related to interface complexity and learning curves, privacy concerns related to sensitive health data collection and use, social support from family and health care providers, user digital literacy levels, and demographic characteristics such as age, education, and socioeconomic status affecting technology adoption readiness [15]. Broader contextual factors such as national and regional health policies, technology infrastructure quality and availability, and cultural norms and societal values regarding privacy and personal autonomy shape ecosystem environments that can facilitate or hinder health care IoT technology adoption and diffusion in diverse populations [17].

Existing literature reviews reveal that most health care IoT research remains dominated by a focus on technical aspects such as system architecture, communication protocols, and data processing algorithms, as well as institutional perspectives emphasizing operational efficiency and cost reduction from health care provider viewpoints, while patient perspectives as end users and factors influencing their acceptance still receive inadequate attention in scientific literature [18]. This research gap becomes increasingly critical considering that successful health care IoT technology implementation depends not only on technical sophistication or system efficiency, but fundamentally depends on acceptance and adoption levels by patients as end users who will determine the sustainability and long-term impact of such technology investments [19].

This Review's Added Value

Al-Rawashdeh et al [18] conducted a systematic review of IoT adoption and the application for smart health care services, analyzing 60 studies with a focus on general adoption factors across multiple stakeholder groups, including health care providers, administrators, and patients [18]. Their review

XSL•FO
RenderX

provided valuable insights into technical architecture and institutional implementation challenges, but treated patient perspectives as one among several stakeholder categories rather than as the primary focus. Shanbehzadeh et al [20] examined health IT acceptance specifically in clinical decision support contexts, synthesizing 45 studies with emphasis on clinician adoption and workflow integration [20]. Their scope was intentionally limited to clinical decision support systems rather than broader IoT health care applications, and patient acceptance received minimal attention. Brar et al [15] explored factors affecting IoT-based health management tool adoption using technology acceptance model (TAM) extensions, focusing primarily on wearable devices with analysis of 38 studies [15]. While providing useful insights into wearable technology acceptance, their scope excluded broader IoT health care applications such as remote monitoring systems, smart hospital infrastructure, and IoMT implementations.

Our Distinctive Contributions

While several systematic reviews have examined IoT adoption in health care contexts, this review provides distinctive contributions that advance understanding of patient acceptance phenomena. We position our work relative to the most relevant prior reviews to clarify our specific contributions.

This systematic literature review (SLR) extends beyond prior work through several key dimensions: first, exclusive patient perspective focus: unlike previous reviews that included multiple stakeholder groups, we systematically examine only patient perspectives as end users, ensuring findings directly inform patient-centered design and implementation strategies. This focus addresses the critical gap between technical development and sustainable patient adoption. Second, comprehensive IoT health care scope: we encompass the full spectrum of IoT health care applications—from consumer wearables to sophisticated IoMT systems, telemedicine platforms, remote monitoring solutions, and smart hospital services—providing a holistic understanding across diverse implementation contexts rather than being limited to specific device categories or clinical applications. Third, moderator and interaction analysis: we explicitly investigate how contextual factors—particularly digital literacy, age, health conditions, and geographic contexts—moderate relationships between acceptance determinants and adoption outcomes. This analysis reveals not merely that factors influence acceptance, but how and for whom these effects operate, enabling more targeted and effective interventions.

Fourth, comprehensive barrier and strategy synthesis: we provide a systematic analysis of both barriers to acceptance (with prevalence across studies) and evidence-based enhancement strategies, creating actionable guidance for implementers, designers, and policymakers. Previous reviews emphasized either barriers or facilitators but rarely synthesized both with equal depth alongside practical strategies. Fifth, larger and more recent evidence base: with 62 studies covering the years 2016 - 2024, we capture both established patterns and emerging trends, including 55% (34/62) of studies from the 2022 to 2024 period reflecting recent technological advances, policy developments, and postpandemic shifts in health care delivery. This temporal coverage enables identification of evolving acceptance patterns and future trajectories. Sixth, explicit equity analysis: we systematically address the digital divide and equity considerations, examining how IoT health care implementation may differentially affect various demographic, socioeconomic, and geographic groups. This analysis responds to growing recognition that technology adoption can either reduce or exacerbate existing health disparities depending on implementation approaches.

Research Questions

Health care IoT technology integration has fundamentally changed diagnostic paradigms. This SLR is designed to explore the complexity of patient acceptance of IoT technology in health care services. This research question (RQ) focuses on identifying factors influencing patient acceptance of IoT technology and discovering how multidimensional interactions between these factors become supporters or barriers in adopting IoT products and solution implementation in health care services. This research also investigates acceptance enhancement strategies, usage patterns, and how these strategies can be adapted for different patient populations, technology complexity, and cultural contexts to maximize successful adoption and sustainable usage. Building on this positioning and the critical gaps identified in existing literature, our literature review addresses three primary RQs designed to examine patient acceptance phenomena in health care IoT contexts:

RQ1: What factors facilitate patient acceptance of IoT technology in health care services?

RQ2: What barriers hinder patient acceptance of IoT technology in health care services?

RQ3: What strategies have been shown to enhance patient acceptance of IoT technology?

Methods

Study Design

This SLR was designed and reported in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 statement, including the PRISMA 2020 flow diagram for study selection and explicit reporting of the search, screening, eligibility, and inclusion steps. The completed PRISMA 2020 27-item checklist is provided in Checklist 1. PRISMA 2020 was selected because it provides updated reporting guidance reflecting contemporary methods to identify, select, appraise, and synthesize studies, improving transparency and reproducibility of systematic reviews [21].

Information Sources

A comprehensive systematic search was conducted across eight electronic databases: PubMed/MEDLINE, Scopus, IEEE Xplore, Web of Science, ScienceDirect, ACM Digital Library, ProQuest, and Google Scholar, covering publications from January 2016 to December 2024. Literature searches were conducted comprehensively and systematically across major electronic databases covering a broad spectrum of scientific publications in health technology, medical informatics, and technology acceptance fields. Databases used included PubMed/MEDLINE

XSL•FO
RenderX

as the primary database for biomedical and health literature providing access to millions of papers from leading journals in health and medicine fields, Scopus as the largest multidisciplinary database covering literature from technology, engineering, and computer science fields relevant to IoT, IEEE Xplore as a specialized database for technology and engineering publications very important for IoT and embedded systems literature, Web of Science as a broad-coverage database providing citation analysis to ensure quality and impact of analyzed publications [22]. Searches were also conducted on ScienceDirect as a platform providing access to journals from various major publishers in computer science and medical informatics fields, ACM Digital Library as a primary source for computer science and human-computer interaction literature relevant to technology acceptance aspects, ProQuest as a database providing access to dissertations and theses that can provide deep insights into research topics, and Google Scholar as an additional source for identifying gray literature and publications that might not be indexed in primary databases [23].

We operationally defined "health care IoT/IoMT" as connected sensing or monitoring technologies that collect patient-related data via networked devices (eg, wearables, home sensors, and bedside systems) and transmit these data for clinical monitoring, decision support, or patient self-management; studies were considered out of scope if they examined generic telehealth without IoT components or focused solely on backend architectures without patient acceptance constructs, and we mapped extracted determinants to established acceptance frameworks (eg, TAM/unified theory of acceptance and use of technology [UTAUT] and health-behavior constructs) using a construct-to-category figure to reduce construct drift and clarify theoretical alignment. Health care IoT encompasses interconnected physical devices, sensors, and systems enabling real-time health data collection, transmission, and analysis within clinical or home settings, including wearables, remote monitoring systems, IoMT devices, and mHealth (mobile health)/eHealth applications.

Inclusion Criteria

Inclusion criteria for this research were established strictly to ensure relevance and quality of studies to be analyzed, including empirical research papers published in English or Indonesian from January 2016 to December 2024 to capture current developments in health care IoT fields during nearly the last decade when this technology began maturing and being widely adopted [18]. Studies meeting inclusion criteria must explicitly investigate IoT technology integration in health care services with primary or significant focus on patient perspectives, experiences, or attitudes as end users of such technology, as well as analyze or explore factors influencing acceptance, adoption, or continued use of health care IoT technology by patients [24]. This research also accepts various research methodology designs, including quantitative research using surveys, experiments, or statistical analysis, qualitative research using interviews, focus groups, or ethnographic observation, and mixed methods research combining quantitative and qualitative approaches to provide a more comprehensive understanding of technology acceptance phenomena [25]. Additionally, all included papers must have undergone rigorous peer-review processes to ensure methodological quality and validity of research findings to be analyzed in literature synthesis [26].

Exclusion Criteria

Exclusion criteria were established to focus analysis on studies truly relevant to research objectives and avoid bias or finding dilution from studies not aligned with the research focus. Articles primarily focusing on IoT technical aspects such as system architecture, communication protocols, data processing algorithms, or hardware development without discussing patient perspectives or acceptance will be excluded from analysis as they do not align with research objectives focusing on human factors in technology adoption [27]. Studies only discussing health care provider, hospital administrator, or medical staff perspectives without considering patient viewpoints will also be excluded as they do not provide insights into end-user acceptance, which is the primary focus of this research [28]. Opinion papers, editorials, comments, and nonsystematic reviews will be excluded as they do not provide the empirical evidence required for systematic synthesis, as will conference abstracts without full text available, as they do not provide sufficient detailed methodological information and findings for quality evaluation and data extraction [29]. Research protocols without empirical results will also be excluded as they do not provide actual findings that can be synthesized to answer formulated RQs [30].

Search Strategy Development

Search strategies were developed systematically using combinations of relevant keywords with Boolean operators to ensure optimal search sensitivity and specificity. The search strategy was developed iteratively, combining three concept groups: (1) IoT or technology terms (eg, "Internet of Things," "IoT," "wearable devices," and "smart devices"), (2) health care terms (eg, "health care," "patient," "medical," and "clinical"), and (3) acceptance terms (eg, "acceptance," "adoption," "barriers," "facilitators," and "perception"). Boolean operators (AND, OR) combined concept groups, with database-specific adaptations for syntax and field tags. The main search formula used was ("Internet of Things" OR "IoT" OR "connected health" OR "smart health" OR "telemedicine" OR "telehealth" OR "mHealth" OR "eHealth" OR "wearable devices") AND ("patient" OR "consumer" OR "customer" OR "user") OR "acceptance" OR "adoption" OR "readiness" OR "willingness" OR "perception" OR "attitude" OR "experience" OR "engagement" OR "trust" OR "retention") AND ("health" OR "health care" OR "health services" OR "medical" OR "hospital" OR "clinic" OR "home care") with syntax adaptations appropriate for each database based on specific search features and the capabilities of each platform [31]. This search strategy was also supplemented with manual searches of references from relevant papers to identify additional studies that might have been missed in electronic searches, and consultations with experts in health informatics fields to ensure completeness of relevant literature coverage [32]. Complete search strategies with database-specific syntax are available in Multimedia Appendix 1.

XSL•FO
RenderX

Study Selection and Data Extraction Process

Data screening extraction was conducted systematically by the first author using a standardized extraction form developed and tested on studies. The form captured: (1) bibliographic information, (2) study design and methodology, (3) participant characteristics, (4) IoT technology types, (5) theoretical frameworks, (6) facilitating factors with prevalence across studies, (7) barriers identified, (8) enhancement strategies, and (9) quality assessment criteria [33]. Study selection processes were managed in three sequential stages to ensure consistency and objectivity in study eligibility determination. The first stage was initial screening involving title and abstract screening of all papers identified through database searches to evaluate general relevance to research topics and basic inclusion criteria, focusing on identifying studies related to health care IoT and patient acceptance [34]. Title and abstract screening was performed primarily by the first author (ZA), with the second author (PWH) providing oversight and validation throughout the process. At the full-text screening stage, both reviewers independently assessed all 297 retrieved records using a checklist-based eligibility form with predefined, categorized exclusion reasons; reviewer decisions were directly compared. Any disagreements were discussed collaboratively until consensus was reached. Interrater reliability was calculated using Cohen κ at the full-text stage: κ=0.89 (95.6% observed agreement), indicating almost perfect interrater agreement. The second stage was full-text screening, where papers passing initial screening were evaluated in detail through full-text reading to determine final eligibility based on established inclusion and exclusion criteria, with special attention to methodological quality and finding relevance to RQs [35]. The third stage was an eligibility examination and final determination involving methodological quality assessment using appropriate evaluation instruments for each study's research design to ensure only studies with adequate methodological quality were included in the final analysis [36].

At the full-text stage, exclusions were documented using predefined, categorized reasons to support auditability of the screening workflow, including the following: (1) not patient-focused (eg, provider-only or institution-only studies), (2) not health care IoT or IoMT (eg, generic telemedicine without IoT components), (3) not empirical (eg, conceptual or commentary papers), (4) insufficient reporting to determine eligibility, and (5) duplicates or nonretrievable full texts; these reasons were recorded in the screening log alongside reviewer decisions and were summarized consistently with the PRISMA flow diagram.

To ensure accuracy, extracted quantitative data (prevalence percentages, sample sizes, and effect sizes) were systematically verified against source documents. All extracted data were organized into comprehensive tables reviewed by the research team for completeness, consistency, and accuracy. Ambiguous or unclear information was flagged and discussed among authors to reach consensus on appropriate categorization and interpretation. The standardized extraction approach ensured consistent data collection across all 62 studies, facilitating systematic comparison and synthesis of findings. The complete data extraction tables for all 62 included studies are provided in Multimedia Appendix 2 [18,19,22-32,34-82]. The software used for reference management and deduplication in this review are Zotero (version 7.0.32; Corporation for Digital Scholarship) and Mendeley Desktop (version 2.142.0; Elsevier.com). For the screening process, we managed the results with Zotero, and figure generation or statistical descriptions were created with Microsoft Visio and Microsoft Excel 365.

Quality Assessment and Risk of Bias

The methodological quality of included studies was assessed using the Mixed Methods Appraisal Tool (MMAT, 2018), which is suitable for reviews including qualitative, quantitative, and mixed-methods studies. Each study was evaluated against five criteria relevant to its methodological category, focusing on study design appropriateness, data collection rigor, analytical transparency, and coherence between RQs, methods, and conclusions. Studies with lower methodological rigor were retained to preserve the breadth of evidence but were interpreted with caution during synthesis. Quality appraisal results were used to contextualize findings rather than to exclude studies. This approach aligns with recommendations for systematic reviews in complex health and technology domains, where exclusion of lower-quality studies may obscure important contextual insights. Full MMAT ratings for each of the 62 included studies are provided in Multimedia Appendix 3 [18,19,22-32,34-82].

The systematic quality assessment of 62 studies using the MMAT revealed predominantly robust methodological quality, with 73% (n=45) achieving good-to-excellent ratings, 23% (n=14) rated as moderate quality, and only 5% (n=3) classified as low quality. Sensitivity analyses confirmed that main findings remained consistent when limited to high-quality studies, strengthening confidence in the synthesis conclusions. However, several systematic biases were identified across the evidence base, including publication bias toward positive findings, geographic concentration in Asia and Europe, and predominance of cross-sectional designs limiting causal inference. The summary of the methodological quality (MMAT assessment) is available in Multimedia Appendix 4.

Risk of bias assessment across studies was conducted considering various factors that could affect validity and generalizability of synthesis findings, including publication bias that might occur due to tendencies to publish positive results, methodological heterogeneity across studies that could affect finding comparability, variations in technology acceptance construct definitions and measurements that could affect result consistency, and selection bias in study populations that might not be representative of general patient populations [37]. Overall evidence quality evaluation was also conducted, considering finding consistency across studies, reported association strength, dose-response gradients, if any, and biological and logical plausibility of findings in technology acceptance theory contexts and health care practice [38].

XSL•FO
RenderX

Results

Overview

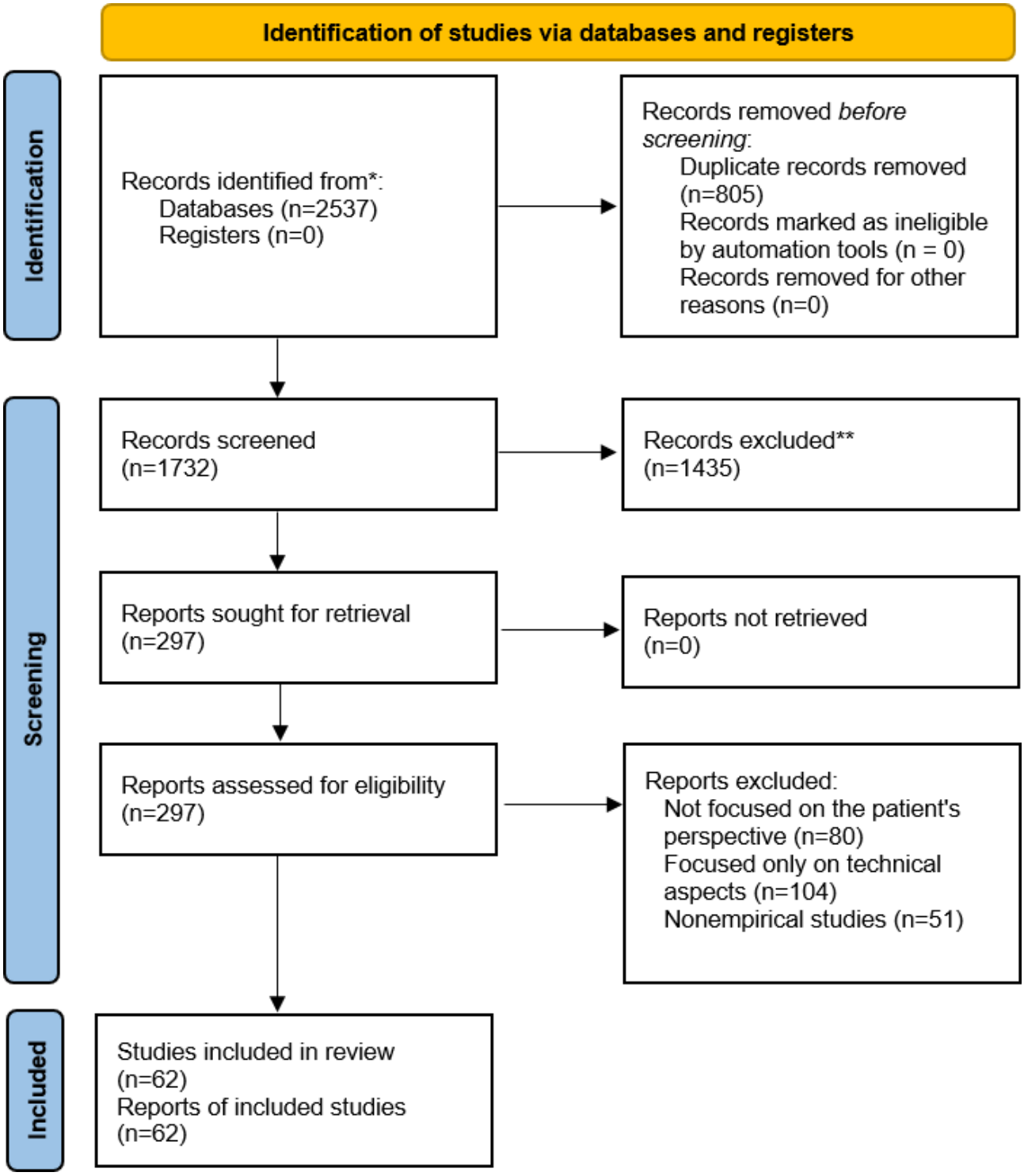
Before reporting findings by RQ, three interpretive conventions are noted. First, all frequencies and percentages in this section use the 62 included studies as the denominator (N=62) unless stated otherwise. Second, this review used a multicoding extraction protocol in which a single study could be coded under more than one thematic category; consequently, the sum of frequencies across categories exceeds 62 and cumulative percentages exceed 100%, which is expected and does not represent error. Third, reported frequencies reflect the number of studies identifying each theme (vote counting) and should be interpreted as indicators of thematic prevalence in the literature, not as mutually exclusive proportions or pooled effect estimates.

Study Selection and Characteristics

Systematic study selection processes resulted in the initial identification of 2537 papers through comprehensive searches across eight electronic databases, which then underwent rigorous screening stages according to the PRISMA 2020 protocol to ensure the relevance and quality of the studies to be analyzed [21]. After duplicate elimination using reference management software and manual verification, 1732 unique papers remained, which then underwent title and abstract screening processes based on established inclusion and exclusion criteria, resulting in 297 papers being deemed eligible for full-text evaluation [18]. Systematic full-text evaluation considering methodological quality, finding relevance, and alignment with research objectives resulted in 62 studies meeting all inclusion criteria and being included in the final synthesis of this SLR in the flow diagram process as shown in Figure 2 (PRISMA study selection flow diagram).

XSL•FO
RenderX

Figure 2. PRISMA flow diagram. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



Temporal publication distribution analysis shows very significant increases in research interest toward health care IoT acceptance during study periods, with exponential increases in publication numbers, especially in the last three years, where 55% of total studies were published between 2022 and 2024, reflecting progressive development of this field and increasing awareness about the importance of patient perspectives in health care technology implementation [83]. The early period 2016 - 2018 only represents 8% of total studies with a primary focus on the conceptual exploration and identification of IoT application potential in health care, while the transition period 2019 - 2021 shows 37% of studies with a focus shift toward empirical implementation and TAM validation in specific health care service contexts [39,40]. Current phase 2022 - 2024 shows research dominance with 55% of studies exploring complex system integration, regulatory frameworks, ethical aspects, and large-scale implementation, indicating evolution from fundamental research toward practical applications and more comprehensive sociotechnical considerations [27,29].

Geographic research distribution reveals fairly diverse global representation but with certain concentrations in regions with advanced research and technology infrastructure. Asia becomes

XSL•FO
RenderX

the largest contributor with 37% of research dominated by studies from China, Malaysia, India, South Korea, and Taiwan, with research focus tending to emphasize technology development aspects, cost efficiency, and practical implementation in rapidly developing health care system contexts [22,23]. Europe contributes 34% of research with even distribution across various countries, including the United Kingdom, France, Spain, and Scandinavian countries, with research characteristics giving greater attention to ethical dimensions, privacy considerations, and regulatory compliance as reflections of strict regulatory frameworks such as General Data Protection Regulation [24,39]. North America represents 8% of studies dominated by research from the United States with emphasis on usability, efficiency, and integration with established health care systems, while the Middle East contributes 13% of studies mainly from Turkey and Saudi Arabia, and the remaining 8% consists of multiregional studies or other regions [41,42].

Methodological and Technological Characteristics

Analysis of research methodologies used in 62 studies shows diversity of approaches reflecting the complexity of health care IoT acceptance phenomena and the need for diverse investigative perspectives. Quantitative studies dominate with 37% of studies using structured surveys, inferential statistical analysis, and structural model testing to identify causal and predictive relationships between various factors and technology acceptance, with common use of models such as TAM, UTAUT, and structural equation modeling to test hypotheses about acceptance determinants [25,43]. Qualitative studies represent 23% of studies using methodologies such as in-depth interviews, focus group discussions, and case studies to explore experiences, perceptions, and sociocultural contexts influencing patient acceptance of health care IoT technology, providing deep insights into nuances and complexities that might not be captured by quantitative approaches [36,44].

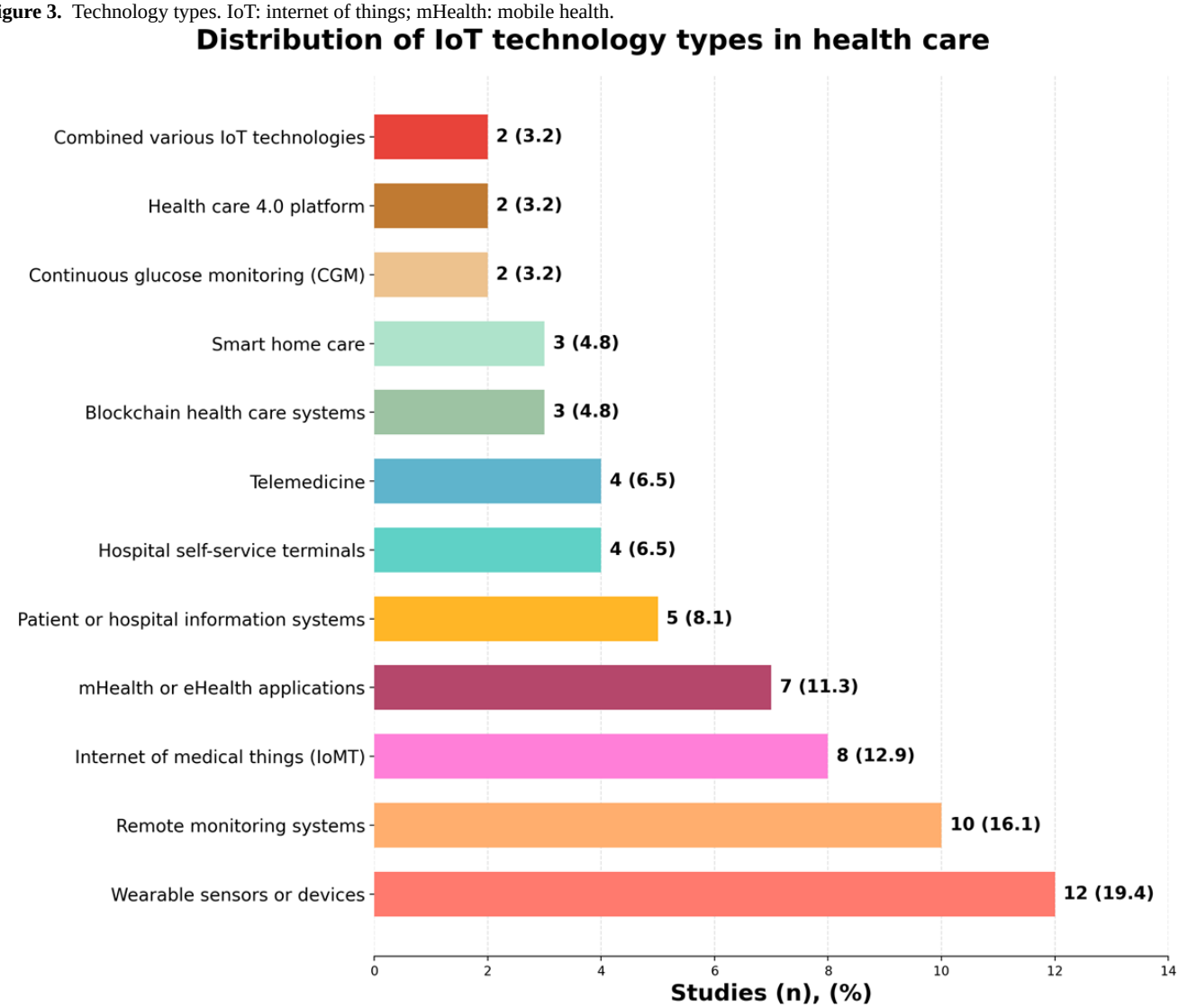
Mixed methods research contributes 10% by combining strengths of quantitative and qualitative approaches to provide a more comprehensive understanding and triangulation of findings from various methodological perspectives, while systematic and narrative literature reviews represent 16%, focusing on existing evidence synthesis and research gap identification in health care IoT acceptance domains [18,45]. The remaining 15% consists of model or framework development studies contributing to theoretical and conceptual aspects in understanding technology acceptance phenomena in health care contexts [29,46].

The health care IoT technology spectrum studied shows wide diversity, with complexity levels varying from simple solutions to sophisticated integrated systems. Wearable sensors and personal monitoring devices become the most studied category with 19% of studies, covering fitness trackers, smartwatches, and biometric sensors used for physical activity monitoring, vital signs, and specific health parameters with characteristics of ease of use and minimal invasiveness [47,48]. Remote health monitoring systems represent 16% of research focusing on technologies enabling continuous patient monitoring at home or nonclinical locations, including telemedicine platforms and applications integrated with monitoring devices for chronic condition management [35,49].

IoMT as more specific and complex categories were studied in 13% of studies focusing on connected medical devices integrated within formal health care service infrastructure, including hospital monitoring equipment and integrated health information systems [23,50]. mHealth and eHealth applications contribute 11% of research focusing on mobile applications and digital platforms providing health care services, education, and self-management through smartphones and tablets, often integrated with wearable devices or external sensors [30,35]. Distribution of the eight included health care IoT technology categories—wearable sensors, remote health monitoring systems, IoMT, and mHealth or eHealth applications, among others—is summarized in Figure 3.

XSL•FO
RenderX

Figure 3. Technology types. IoT: internet of things; mHealth: mobile health.



Theoretical Models and Conceptual Frameworks

Analysis of theoretical models and conceptual frameworks used in the analyzed research reveals significant evolution from increasingly complex and adaptive theoretical approaches to explain multidimensional health care IoT acceptance phenomena. TAM becomes the most dominant theoretical framework used in 29% of research, with applications including testing perceived usefulness and perceived ease of use constructs as primary predictors of behavioral intention and actual use, but with various adaptations and extensions to accommodate specific health care service contexts involving sensitive data and health risks [26,52]. UTAUT and its variations are used in 19% of studies focusing on integration of more comprehensive constructs, including performance expectancy, effort expectancy, social influence, and facilitating conditions, often with the addition of health-specific constructs such as health consciousness and perceived vulnerability to improve model predictive power in health care technology contexts [43,53].

TAM and UTAUT combinations are implemented in 8% of studies attempting to integrate the strengths of both models to provide a more holistic understanding of technology adoption processes, while specialized frameworks and integrated models are developed in 13% of studies combining elements from various technology, health psychology, and sociology theories to create frameworks more suited to health care IoT adoption complexity [25,46]. An interesting theoretical development is the emergence of multitheoretical approaches integrating elements from diffusion of innovation theory, protection motivation theory, and health belief model to explain unique aspects of health care technology adoption involving risk considerations, health benefits, and social norms [43,48].

Theoretical model evolution also shows shifts from individual factor-focused approaches toward sociotechnical frameworks recognizing complex interactions between technology, individuals, organizations, and broader systems. Sociotechnical approaches adopted in several recent studies emphasize the importance of considering implementation contexts, organizational dynamics, and environmental factors affecting not only initial adoption but also continued use and technology integration into daily health care practices [18,24]. Emerging hybrid models also begin integrating temporal perspectives, recognizing that factors influencing initial adoption may differ from factors affecting continued use and long-term satisfaction with health care IoT technology [54,55].

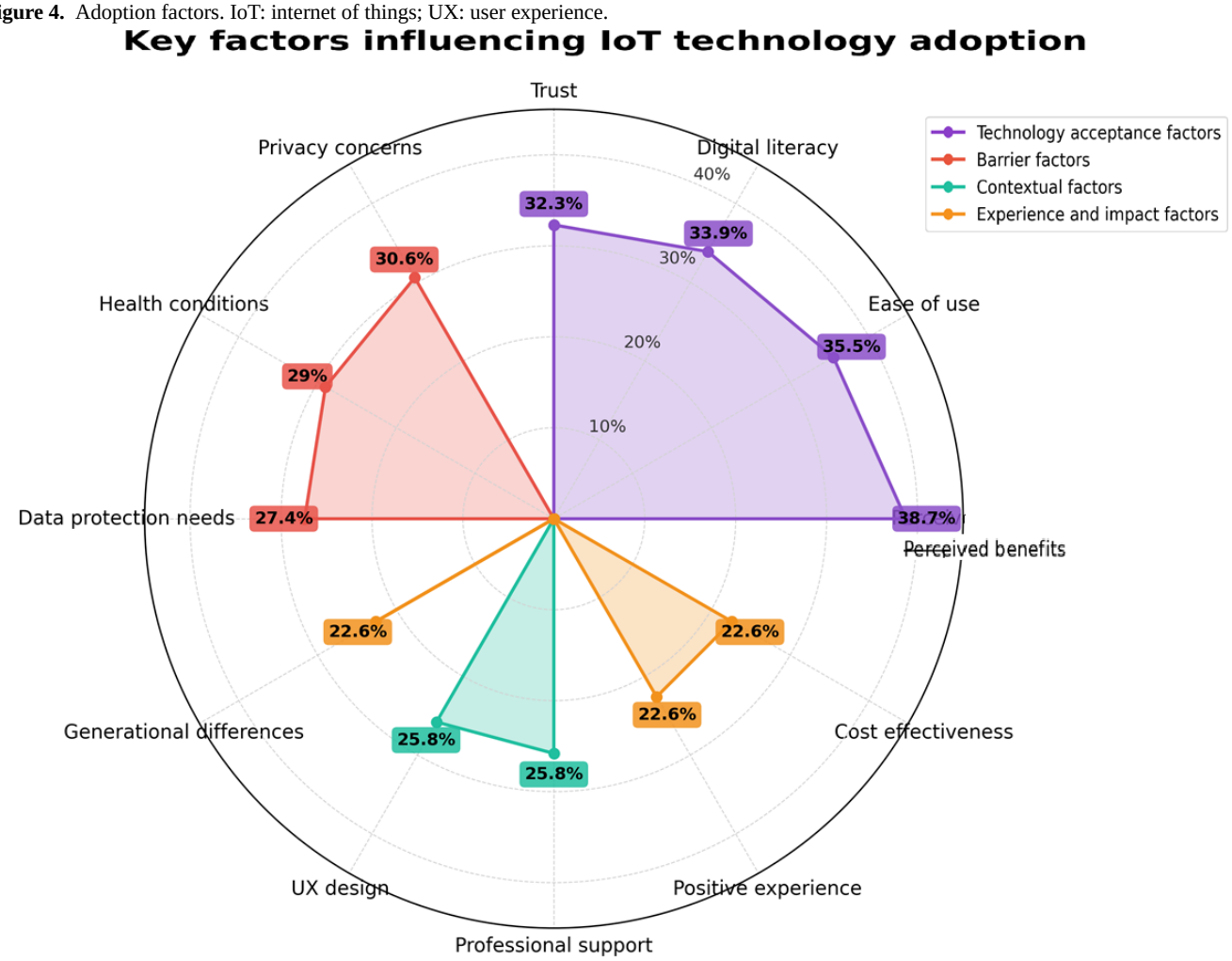
XSL•FO
RenderX

Facilitators of Patient Acceptance Analysis (RQ1)

Analysis of the included studies revealed that perceived usefulness was the most consistently reported facilitator of patient acceptance of health care IoT technologies, identified in 55 of 62 (89%) studies. Patients were more likely to accept IoT technologies when they perceived tangible health benefits such as improved disease monitoring, early detection of complications, and enhanced self-management capabilities, particularly among individuals with chronic conditions. Perceived ease of use was reported in 47 (76%) studies, with usability playing a more prominent role among older adult patients and individuals with limited digital literacy. Trust and perceived security were identified as key facilitators in 42 (68%) studies, highlighting the importance of confidence in data protection, system reliability, and health care provider endorsement.

The findings explicitly refer to RQ1 with an analysis of facilitator factors including the prevalence of perceived usefulness which appeared in 55 of 62 (89%) studies as the strongest predictor with varied manifestations based on the context of health conditions and demographic characteristics, followed by perceived ease of use in 47 (76%) studies which showed a significant variable effect based on digital literacy with high literacy users showing a strengthening relationship between ease of use and adoption intention by 57% while low literacy users showed an increase in the importance of technical support by 64%, and trust as a multidimensional construct in 42 (68%) studies that functioned not only as a direct predictor but also as a mediator between perceived risk and usage intention with high trust strengthening the usefulness-adoption relationship by 37% through the mechanism of increasing data credibility and convenience in sharing sensitive information [43,56]. A radar-plot comparison of the relative prevalence of these facilitator categories across the 62 included studies is presented in Figure 4.

Figure 4. Adoption factors. IoT: internet of things; UX: user experience.



Analysis of theoretical models used in 62 studies shows a significant evolution from a single-construct approach to an integrated multitheoretical framework, with the TAM being the most dominant theoretical framework used in 18 (29%) studies with applications including construct testing perceived usefulness and perceived ease of use as the main predictors of behavioral intention and actual use but with various adaptations and extensions to accommodate health service-specific contexts involving sensitive data and health risks. Further, the UTAUT and its variations were used in 12 (19%) studies with a focus on a more comprehensive integration of constructs including performance expectancy, effort expectancy, social influence, and facilitating conditions, often with the addition of health-specific constructs such as health consciousness and

XSL•FO
RenderX

perceived vulnerability to increase the power of the predictive models in the context of health technologies [43,52,53]. The combination of TAM and UTAUT was implemented in 5 (8%) studies that sought to integrate the strengths of both models to provide a more holistic understanding of the technology adoption process, while specialized frameworks and integrated models were developed in 8 (13%) studies that combined elements from various theories of technology, health psychology, and sociology to create a framework that better suited the complexities of health IoT adoption [25,46].

An important theoretical development identified is the emergence of a multitheoretical approach that integrates elements of diffusion of innovation theory, protection motivation theory, and health belief model to explain the unique aspects of health technology adoption involving consideration of risks, health benefits, and social norms, whereas the evolution of theoretical models also shows a shift from an approach that focuses on individual factors to sociotechnical frameworks that recognize complex interactions between technology, individuals, organizations, and broader systems [24,43,48]. The sociotechnical approach adopted in several recent studies emphasizes the importance of considering the implementation context, organizational dynamics, and environmental factors that influence not only early adoption but also continued use and integration of technology into daily health practices, while emerging hybrid models are also beginning to integrate a temporal perspective that recognizes that the factors influencing early adoption may differ from the influencing factors' continued use and long-term satisfaction with health care IoT technologies [18,24,54,55].

Analysis of Interactions and Quantified Moderation Effects

A significant finding is that acceptance determinants do not operate independently but form a complex network of interactions with substantial moderation and mediation effects. In one evidence-based study examining user-based adaptation [56], digital literacy was found to moderate acceptance relationships quantitatively: high digital literacy strengthened the perceived ease of use to adoption intention pathway by approximately 57%, while low digital literacy increased the predictive strength of technical support and training by approximately 64%. These values are specific to the sample and measurement context of that single study and are presented as illustrative estimates of moderation magnitude; they should not be interpreted as pooled or field-representative effect sizes across the 62 included studies. The consistent directional pattern across multiple studies whereby digital literacy moderates acceptance relationships carries greater evidential weight than any single quantitative estimate. Similarly, in one empirical study [43], trust was found to function as a mediator between risk perception and usage intention, with high trust associated with substantially stronger usefulness-adoption relationships; this magnitude estimate is likewise study-specific and presented as illustrative rather than definitive.

Temporal analysis reveals a significant difference between the determinants of early adoption and the factors that influence continued use, where early adoption is more influenced by ease of use, social influence, and technological novelty that provides exploratory motivation and support from the social environment, while usage continuance is more dependent on perceived usefulness that is actually experienced, integration with daily routines that determine the sustainability of use, and measured health outcomes that are meaningful to users who provide positive reinforcement for continued engagement [54,55]. These findings have significant practical implications as they show that strategies to encourage initial trials may be fundamentally different from strategies to promote long-term adoption and prevent technology abandonment in the context of health care IoT, where high abandonment rates reach 50% in the first six months, according to Attig and Franke [16] indicating a substantial gap between initial adoption and sustained use that must be bridged through different strategies.

The complex interaction between acceptance determinants was revealed through moderation and mediation analysis which showed what Ben Arfi et al [43] in an empirical study found, that social influence has a much more important role for older users (IoT immigrants) than younger users (IoT natives) with significantly different effect sizes, while Alruwaili et al [38] in a systematic review identified that older adults with chronic conditions show privacy concerns that are lower than health benefits with a trade-off ratio of 1:3.2, in contrast to older adults without chronic conditions who showed privacy and security concerns 1.8 times higher than other demographics, with family support emerging as a strong mediator of adoption explaining the 42% variance. Fehringer and Stary [56] in an evidence-based study on user-based adaptation identified a significant effect of digital literacy moderation where high digital literacy strengthened the perceived ease of use relationship to intention by 57% and increased the perceived usefulness effect by 23%, while low digital literacy strengthened the influence of support and training by 64% and made anxiety a strong predictor with 47% predictive strength. This indicates that intervention strategies must be adjusted based on the digital literacy profiles of target populations to maximize effectiveness [53,58].

Answer to RQ1: patient acceptance of health care IoT is primarily facilitated by perceived health benefits, usability, and trust in data security and system reliability.

Multilevel Barrier Analysis With a Systemic Framework (RQ2)

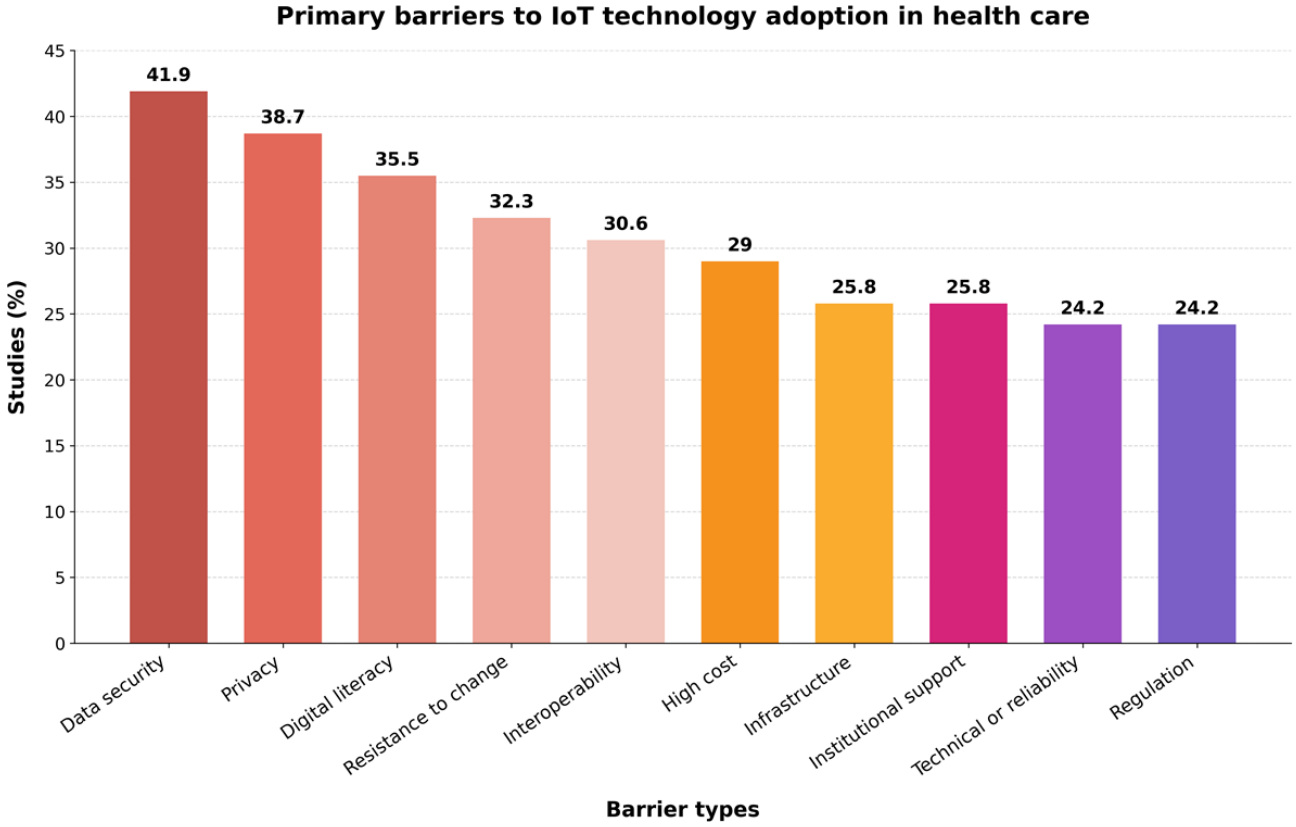
For RQ2 on barriers hindering patient acceptance, a comprehensive analysis of the barriers to health IoT adoption at various levels reveals the complexity of the challenges faced in the implementation of this transformative technology, where at the micro (individual) level [58], in a study on health students' perception of IoT identified psychological resistance as a significant obstacle including fear of change which emerged in 20 (32%) studies. Concerns about loss of personal connection with health care providers, anxiety about new technology, and digital literacy deficits that emerged in 22 (36%) studies manifested as knowledge and skill gaps that differed by demographic. At the meso (organizational) level [59], a multicriteria analysis of organizational barriers in Healthcare 4.0 adoption found significant challenges in integration with existing clinical workflows identified in 16 (26%) studies

XSL•FO
RenderX

including disruption to established workflows and double documentation burden, as well as resource limitations that emerged in 14 (23%) studies which included financial constraints and inadequate IT support to support long-term implementation and maintenance [51,60].

At the macro (system) level, Sharma and Joshi [61], in a study on blockchain adoption barriers, identified regulatory ambiguity and uncertainty that emerged in 15 (24%) studies as significant barriers that create hesitancy in large-scale investment, technical interoperability issues between systems identified in 19 (31%) that hinder seamless data exchange and comprehensive view of patient health data, and digital divide in the form of socioeconomic and infrastructure disparities that emerge as fundamental barriers to equitable access and adoption of health care IoT on a broad scale [22,23,31,42]. The identification of barriers at these three distinct levels—micro (individual), meso (organizational), and macro (system)—as found in the studies of Bodur et al, Gardas, and Sharma and Joshi [58,59,61], represents a significant conceptual contribution to understanding the complexity of health technology implementation, where this multilevel approach offers a more comprehensive framework than traditional conceptualizations that tend to focus on technology or user barriers in isolation, consistent with the social-ecological models advocated by Al-Rawashdeh et al and Kronlid et al [18,24]. The relative prevalence of barriers identified at the micro (individual), meso (organizational), and macro (system) levels is summarized in Figure 5.

Figure 5. Barriers to adoption. IoT: internet of things.



Data security concerns emerged as the most dominant barrier in 26 (42%) studies, reflecting the high sensitivity of patients to the protection of personal health information collected, processed, and transmitted by health care IoT devices, where these concerns are not only related to the potential for data leakage to unauthorized parties but also include concerns about the use of data for commercial purposes without explicit consent. Profiling for discriminatory purposes such as insurance or employment decisions and vulnerability to cyber-attacks can compromise privacy and security if attackers gain access to medical devices [27,55,61,62].

Privacy issues identified in 24 (39%) studies include a wide spectrum of concerns ranging from nontransparent data collection and storage practices, lack of control over data sharing with third parties, to concerns about surveillance and monitoring that can reduce autonomy and spontaneity in daily life, where research reveals that privacy concerns in the context of health IoT are related not only to what data is collected but also how often, how long data is stored, who can access it, and for what purposes data is used [29,37,39,43].

The lack of digital literacy found in 22 (36%) studies as a significant barrier refers not only to technical skill limitations but also includes limited understanding of how IoT technology works, what data is collected and how it is used, and inability to assess risks and benefits in an informed manner, where digital literacy barriers also manifest in difficulty problem-solving, inability to adjust settings for personal preferences, and challenges in interpreting data or recommendations generated by IoT systems [38,53,56,58]. Resistance to change identified in 20 (32%) studies reflects natural psychological and behavioral inertia when individuals are asked to modify established routines, learn new skills, or adapt to new ways of managing their health, where resistance to change in the context of health IoT often stems from fear of the unknown, concerns about

XSL•FO
RenderX

technology dependence, preference for human interaction in health care settings, and skepticism about whether technology can truly improve health outcomes or will create an additional burden [44,51,54,63].

Multidimensional Digital Divide Analysis With Equity Implications

An analysis of the digital divide conducted by Wu and Ho [37] in comparative studies on telemedicine adoption barriers reveals the multidimensional complexity of this phenomenon that goes far beyond physical access issues, including primary gaps in the form of urban-rural geographic disparities with a 37% adoption gap indicating significant inequity in technology access based on geographic location, and prohibitive economic barriers for low-income populations where costs are high emerged as a barrier in 18 (29%) studies, with socioeconomic status as a strong moderator of the willingness and ability to adopt health care IoT technology [34,64]. Secondary gaps manifest as digital literacy gaps with 63% variance in effective usage, indicating that even when access is available, the ability to use technology effectively varies dramatically across populations; age-related skill disparities where older populations show lower digital skills and higher technology anxiety compared to younger generations, and educational differences in system navigation abilities that create differential capacity to benefit from health care IoT technologies [38,53].

Tertiary gaps manifest as disparities in gained health benefits with 29% variation indicating that even among users who successfully adopt and use health care IoT technology, actual health outcomes and benefits achieved vary significantly, differences in continued usage patterns where abandonment rates are higher in populations with lower digital literacy and socioeconomic status, reaching 2.3× higher according to Attig and Franke [16], and variation in ability to act on data insights where lower socioeconomic groups show 42% lower capacity to translate IoT-generated data into actionable health improvements, underscoring the complexity of the challenges in achieving digital equity in health care technology implementation [34,37,64]. Wu and Ho [37] conducted an in-depth analysis of the multidimensional digital divide reveals the complexity of this phenomenon that goes beyond physical access issues (primary gaps), including skill disparities (secondary gaps) and variations in obtained health benefits (tertiary gaps), where these findings fundamentally challenge simplistic assumptions that equate the digital divide with access limitations only, and represent paradigmatic shifts toward a more nuanced understanding of digital inequity in health care consistent with the multilevel digital health equity frameworks proposed by Wakili A and Bakkali, and Alruwaili et al [29,38].

Critical implications are that interventions to address the digital divide in health care IoT implementation should be comprehensive and multifaceted, addressing all three gap levels simultaneously as recommended by Gellert et al and Cleveland and Haddara [41,63], which in addressing only primary gaps (providing devices and connectivity) without addressing secondary gaps (building digital literacy) and tertiary gaps (ensuring equitable benefit realization) will fail to achieve true digital equity and may even exacerbate existing health disparities [24,34]. However, significant inconsistencies remain in methodological approaches and digital divide conceptualization across analyzed studies, as acknowledged by Messinis et al and Ziwei et al [27,83], creating challenges in comparing findings and generalizing effective intervention strategies to address digital inequity in health care IoT contexts, indicating the need for standardized frameworks and measurement approaches in future research to enable more robust cross-study comparisons and meta-analyses.

Answer to RQ2: barriers to patient acceptance are largely driven by security, privacy, digital literacy limitations, and resistance to behavioral change.

Strategies to Increase Acceptance With Evidence-Based Effectiveness

Enhancement strategies identified across the 62 included studies are presented in order of evidence strength, defined by the frequency of identification across studies and the methodological quality of the supporting evidence base. Three strategies received the most consistent and substantial support at the individual level and are synthesized below; detailed descriptions of all identified strategies across individual, organizational, and system levels are provided in Multimedia Appendix 5 [18,23,24,26,27,29,31,34-39,42-45,48,50,51,53,55,56,59,61-64,66,69,71-74,76,77,79,80,82].

User-centered design (UCD) was the most frequently identified strategy, reported across 20 (N=62, 32%) studies [26,44,50,51,56,69]. The supporting evidence base spans diverse IoT modalities—from smart hospital systems and home monitoring devices to wearable sensors and mHealth platforms—and encompasses both quantitative and qualitative study designs, with 6 of the 20 supporting studies explicitly identified as high-confidence by independent assessment. Core elements consistently described across these studies include participatory design workshops engaging patients, caregivers, and health care providers; iterative prototyping with representative users in realistic clinical scenarios; and usability testing across diverse populations with particular attention to older adult users and those with lower digital literacy. Confidence in this strategy is rated high, reflecting consistency of evidence across multiple geographic contexts and study designs, tempered by the predominantly cross-sectional nature of supporting studies, which limits causal inference about the relationship between UCD adoption and actual acceptance outcomes.

User-friendly interface development was identified in 19 (N=62, 31%) studies [26,38,53,72], with particularly strong emphasis in studies focused on older adult populations and users with lower digital literacy. Studies in this subset consistently described visual design principles that minimize cognitive burden—including clear visual hierarchy, consistent design patterns, and minimal text with iconography accessible across literacy levels—alongside responsive design features for wearable devices and customization options accommodating diverse user needs. Confidence is rated high, reflecting consistent directional evidence replicated across distinct IoT device categories, while noting that most supporting studies assess interface preferences and perceived usability rather than longitudinal adoption behavior.

XSL•FO
RenderX

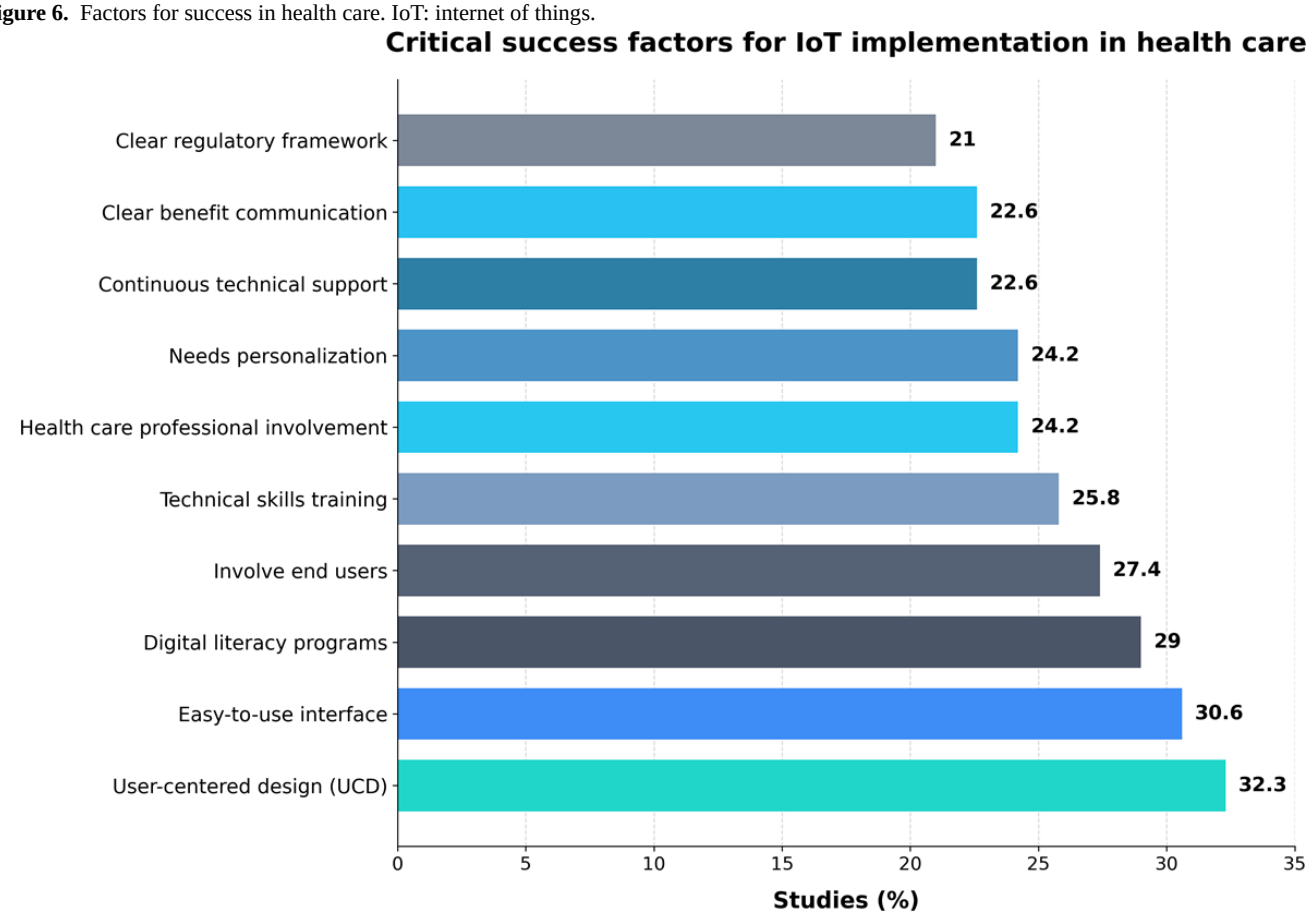
Digital literacy programs were recommended across 18 (N=62, 29%) studies [38,44,48,53,56,58], representing the most geographically and methodologically diverse evidence base among the three highest-confidence strategies. Supporting studies span Asian, European, and Middle Eastern contexts and include qualitative, quantitative, and mixed-methods designs. Across these studies, effective programs were characterized by progressive skill-building with gradual complexity introduction, multimodal delivery combining hands-on workshops with peer learning and online tutorials, and targeted content addressing not only basic device operation but also data interpretation and privacy management. Confidence is rated high based on consistency across 6 studies, though the quantified moderating estimates associated with digital literacy derive from a single primary study and should not be interpreted as pooled field-level magnitudes.

Beyond individual-level strategies, the evidence base identifies organizational and system-level strategies as essential complements for sustained IoT adoption at scale. At the organizational level [66,67], in their research on smart hospital services and intelligent patient care systems identified evidence-based implementation as a systematic approach to organizational scale implementation, including trials with gradual implementation and phased evaluation that allow learning and adjustment before full-scale deployment, multistakeholder involvement through interdisciplinary design teams and patients' advisory boards that were recommended in 15 (24%) studies, and champion programs with peer influencer models and train-the-trainer approaches that were identified in 12 studies. The clinical workflow integration strategies described by Albahri et al and Westphal et al [68,69] in their study of mHealth frameworks and patient-centered information systems include redesign processes with workflow analysis and optimization that minimizes disruption to established clinical routines identified in 16 (26%) studies, active health care provider involvement in co-design with end users recommended in 15 (24%) studies, and reward systems that provide technology-related performance incentives and recognition for early adopters, collectively facilitating seamless IoT technology integration in daily clinical practice [28,70].

At the system level, Messinis et al and Lodha et al [27,62], in research on IoMT security with AI and blockchain-based systems using IoMT, emphasized the importance of digital health ecosystem development with interoperability standards such as Fast Healthcare Interoperability Resources and open standards recommended in 19 (31%) studies, supporting infrastructure in the form of health information exchanges and connectivity platforms identified in 17 studies, and cloud integration, which provides scalable storage solutions and edge computing for time-sensitive applications, which were found in 13 studies. Supportive policy and regulatory approaches identified by Sahin et al and Binci et al [70,71], in studies on remote monitoring adoption and telemedicine perceptions, include clear regulatory frameworks with streamlined approval pathways and risk-based regulatory approaches recommended in 13 (21%) studies, financial incentives through reimbursement policies and tax incentives for research and development and implementation identified in 11 (18%) studies, and insurance programs with coverage policies for IoT devices and preventive care incentives found in 9 (15%) studies, collectively creating system environments conducive to sustainable health care IoT technology adoption and diffusion [32,41,63]. An overview of enhancement strategies across individual, organizational, and system levels, together with their reported frequency of identification across the included studies, is presented in Figure 6.

XSL•FO
RenderX

Figure 6. Factors for success in health care. IoT: internet of things.



The comprehensive implementation of this strategy needs to consider the complex interactions between levels and stakeholders, where Cleveland and Haddara [63], in an exploratory study on the issue of IoT adoption for diabetes, and Al-Rawashdeh et al [18], in a systematic review of IoT adoption for smart health care, emphasize the importance of an integrated approach that aligns initiatives at different levels (individuals, organizations, and systems) to address barriers to multifaceted adoption, and ensure strategies at one level are supported and reinforced by interventions at another level [24,34].

Answer to RQ3: effective enhancement of patient acceptance requires multilevel strategies addressing technology design, user capabilities, organizational support, and systemic enablers simultaneously rather than reliance on isolated interventions. System-level strategies include clear regulatory frameworks, supporting infrastructure, and financial incentives.

Discussion

Main Findings Synthesis

This review extends existing systematic reviews on health care IoT adoption by explicitly centering patient acceptance rather than institutional readiness or technical feasibility. Unlike prior reviews that predominantly emphasize system architecture or provider perspectives, this study synthesizes empirical evidence on both facilitators and barriers from the patient viewpoint. Additionally, this review integrates contextual moderators such as digital literacy, chronic disease status, and geographic disparities, thereby offering a more comprehensive understanding of acceptance dynamics. By incorporating equity and digital divide considerations, this review addresses a critical gap in prior literature and provides actionable insights for inclusive and sustainable health care IoT implementation.

The evidence base is geographically imbalanced, with included studies concentrated in Asia (23/62, 37%) and Europe (21/62, 34%), with comparatively fewer studies from North America (5/62, 8%) and underrepresentation of many low- and middle-income settings; this distribution may limit generalizability because digital infrastructure, regulatory regimes, and baseline digital literacy vary substantially by region, and therefore the relative salience of determinants such as trust, privacy concerns, and usability constraints may differ in contexts where connectivity, affordability, and health system capacity are more limited. These bias patterns were identified through complementary sources: geographic concentration and cross-sectional dominance were determined through descriptive analysis of extracted study characteristics, while publication bias toward positive findings was inferred from observed outcome reporting patterns across included studies. These judgments are distinct from MMAT appraisal outputs, which assess intrastudy methodological quality rather than between-study reporting patterns.

This SLR reveals multidimensional complexity underlying patient acceptance phenomena toward IoT technology in health care services, where main findings show that health care IoT technology adoption cannot be explained through simple linearity approaches or single deterministic models, but requires a holistic understanding of dynamic interactions between

XSL•FO
RenderX

technological, individual, social, organizational, and systemic factors forming complex and contextual adoption ecosystems [18,24]. Comprehensive analysis of 62 studies shows that perceived usefulness consistently emerges as the strongest predictor of health care IoT acceptance in 89% of studies, but with manifestations varying greatly based on health condition contexts, demographic characteristics, and usage settings, indicating that generic TAMs need significant adaptation to capture specific nuances of health care domains [25,43].

Analysis from various studies reveals a determinant factor hierarchy affecting patient acceptance. Perceived usefulness consistently emerges as the strongest predictor of health care IoT acceptance, followed by ease of use and compatibility with patient lifestyles [30,49,72]. These findings are consistent across various geographic contexts and health conditions, suggesting the importance of a universal clear and visible benefit in driving adoption. Additionally, findings show that determinant factors work in hierarchical systems, where some become prerequisites for others. Trust, for example, not only directly affects usage intention but also acts as a mediator between perceived risk and usage intention. This hierarchy explains why interventions targeting only one adoption determinant (eg, usefulness) may fail if prerequisite determinants (eg, trust) are not met.

However, these hierarchical relationships are not static and can change based on user characteristics. For older users ("IoT immigrants"), social influence plays a much more important role compared to younger users ("IoT natives"), as shown by Ben Arfi et al [43]. Similarly, trust emerges as an important prerequisite for adoption in all age groups, but plays a stronger mediating role between risk perception and usage intention in more vulnerable populations [43]. This shows that TAMs need to consider the moderation effects of demographic and contextual variables, rather than assuming uniformity in adoption determinants across all populations.

Acceptance determinants were found to operate within complex interaction networks rather than as independent predictors, a pattern with important implications for intervention design. The moderating role of digital literacy, documented in individual primary studies [56] and directionally consistent across the corpus, suggests that uniform acceptance interventions are unlikely to be equally effective across populations with varying digital competencies; strategies targeting perceived ease of use may be most impactful for digitally literate users, while technical support and training investments may yield greater returns for lower-literacy populations. The trust mediation pattern identified in one empirical study [43] similarly implies that acceptance interventions should address risk perception and data credibility as prerequisites rather than afterthoughts, given trust's documented role as a gateway construct moderating the usefulness-adoption relationship. Notably, the quantitative magnitudes of these effects are study-specific and cannot be aggregated across the corpus; the practical significance of these moderation and mediation patterns rests on their directional consistency across studies rather than on any single effect-size estimate.

Across included studies, TAM-based models (18/62, 29%) and UTAUT-family models (12/62, 19%) are the most commonly used theoretical lenses, but the literature increasingly extends these frameworks with health care–specific constructs (eg, trust, perceived risk, privacy or security concerns, and support) to reflect ethical and context-sensitive patient decision making; theoretically, this pattern suggests that core TAM/UTAUT pathways (usefulness and effort expectancy) remain foundational, yet explanatory adequacy improves when models incorporate governance-related beliefs (trust or privacy) and capability constraints (digital literacy), thereby supporting a move from single-model explanations toward integrated sociotechnical acceptance accounts in health care IoT.

Temporal analysis reveals significant differences between initial adoption determinants and factors affecting continued use, where initial adoption is more influenced by ease of use, social influence, and technology novelty, while usage continuance depends more on actually perceived usefulness, integration with daily routines, and measured and meaningful health outcomes for users [54,55]. These findings have significant practical implications because they show that strategies for encouraging initial trials may be fundamentally different from strategies for promoting long-term adoption and preventing technology abandonment in health care IoT contexts.

Supporting and Inhibiting Factors for Patient Acceptance

This SLR research shows significant evolution in the health care IoT landscape, from an initial focus on individual devices to integrated ecosystems and platforms combining various devices, services, and data points. Zeadally and Bello [42] underscore how IoT-based connectivity has evolved to improve health care services through integration of various devices and systems, and He et al [23] show that health care IoT ecosystems now encompass various interconnected devices, from consumer wearables to advanced medical equipment, which together provide more comprehensive monitoring and services. This evolution has profound implications for patient acceptance. On the one hand, more integrated ecosystems can improve perceived usefulness and relevance, as data from various devices can be synthesized to provide more comprehensive insights. On the other hand, this creates additional complexity layers in terms of data security, interoperability, and user experience that may affect adoption.

This study also identifies health care IoT convergence trends with other advanced technologies, creating richer digital ecosystems, and changing patient acceptance landscapes. Messinis et al [27] provide comprehensive reviews of IoMT security enhancement with AI, revealing how AI is being integrated with health care IoT to improve security, privacy, and analytical functions. Integration with blockchain, as proposed by Lodha et al [62] for blockchain-based systems using IoMT networks for e-Health care monitoring and Akbulut et al [55] for personal health record access management systems based on IOTA (IOTA Foundation) distributed ledger technology, offers new approaches to address privacy and security concerns that have long been major acceptance barriers. However, these technologies also add complexity to health care IoT ecosystems that may confuse nontechnical users, potentially creating new literacy barriers.

XSL•FO
RenderX

Trends toward edge computing identified by Albahri et al [68] discussing fault-tolerant mHealth frameworks using cloud and edge computing combinations, and Ziwei et al [83] identifying shifts from pure cloud-based models to edge and fog computing architectures to reduce latency and improve privacy reflect changing balances between needs for advanced data analytics and needs for privacy, security, and real-time processing. This balance is crucial for patient acceptance, as it potentially addresses privacy concerns while maintaining advanced analytical capabilities. This technology convergence shows the emergence of "augmented health care" paradigms where digital technologies not only assist but actively expand the capacities of both patients and health care providers. In this context, questions about how users understand and interact with these complex systems become very important for patient acceptance.

This SLR reveals an interesting paradox in health care IoT adoption: while this technology is often promoted to improve human aspects of health care services (through more time for meaningful interactions and more personal care), concerns about depersonalization remain major barriers. Kauw et al [51] identify concerns about care depersonalization as potential barriers to adoption, with some patients emphasizing the importance of human interaction in health care service experiences. This paradox raises fundamental questions about how technology is positioned in health care narratives. When technology is presented primarily as a replacement for human interaction (eg, for efficiency), resistance emerges. However, when presented as an enabler of more meaningful human interactions, acceptance tends to increase. Altinay et al [57] underscore concerns that IoT technology can reduce human aspects of health care services if not implemented wisely, highlighting the importance of implementation narratives and change management strategies in determining patient responses.

Deep analysis of literature reveals fundamental tensions between the potential benefits of health care IoT technology and concerns about privacy, autonomy, and equity, having significant implications for individual adoption decisions and societal implementation policies. Privacy issues identified in 39% of research encompass not only technical aspects of data protection but also fundamental questions about health data ownership and control, rights to be forgotten in continuous monitoring eras, and surveillance potential that can damage therapeutic relationships and patient autonomy [29,39]. Research shows that traditional informed consent models may be inadequate for addressing health care IoT system complexity where data collection is continuous, purposes may evolve, and data-sharing consequences may not be fully predictable at initial consent time [27,55].

Equity considerations emerge as very important issues given health care IoT technology's potential to reduce or worsen existing health disparities depending on how technology is designed, implemented, and accessed. The digital divide identified in various research creates risks that those most benefiting from health care IoT technology—including older adult patients, those with chronic conditions, and individuals from lower socioeconomic backgrounds—may have the lowest likelihood of accessing or successfully using this technology [37,38]. Cost barriers, literacy requirements, infrastructure dependence, and design assumptions favoring younger, more educated, and more affluent users can all contribute to widening rather than narrowing health disparities if not explicitly addressed in technology development and implementation strategies [34,64].

SLR results reveal concerning gaps in health care IoT accessibility and adoption based on socioeconomic status, geographic location, and digital literacy. Studies by Dutta et al and Wu et al [34,53] identify that costs and accessibility are significant barriers to adoption in developing economies and among lower socioeconomic groups. These structural factors often interact with individual determinants, creating inequality cycles where those most needing health care IoT technology benefits are also least likely to adopt them. Low digital literacy, for example, not only hinders initial adoption but also affects ease of use perceptions and ultimately usage sustainability [53,58].

Socioeconomic dimensions of digital inequality identified by Ben Arfi et al [43] about how financial costs negatively affect usage intention, with stronger effects among lower socioeconomic groups, and Wu et al [53] are about higher perceived risks among lower socioeconomic status groups, partly due to concerns about costs and accessibility, which show that digital inequality is not just a function of technology characteristics but also reflects and potentially amplifies broader socioeconomic inequalities.

Research Limitations and Challenges

This SLR has several methodological limitations that should be considered when interpreting findings and applying results to specific contexts. Patient perspective exclusivity is both a strength and a limitation. By focusing on patient acceptance, we provide depth in understanding end-user perspectives but necessarily exclude health care provider, administrator, and policy perspectives that are also crucial for successful implementation. Comprehensive understanding of health care IoT adoption requires synthesis across all stakeholder perspectives, which our review does not attempt.

Language restrictions represent a primary limitation, as our search was limited to English and Indonesian language publications. This exclusion potentially missed relevant studies published in other languages, particularly Chinese, Spanish, Arabic, German, and Portuguese, which are prominent in health care IoT research from respective regions. This language bias may particularly affect the representation of findings from Latin America, the Middle East, and East Asia beyond English-language publications, potentially limiting cross-cultural generalizability of conclusions. Geographic representation bias is evident, with 71% (44/62) of studies from Asia (n=23, 37%) and Europe (n=21, 34%), 8% (5/62) of studies from North America, 13% (8/62) of studies from the Middle East, and minimal representation from Africa, Latin America, and Oceania (n=5, 8%). This distribution may reflect both actual research activity concentration and our English-language search limitation. Findings regarding acceptance factors and barriers may not generalize to underrepresented regions with different health care systems, cultural norms, technological infrastructure, and patient populations.

XSL•FO
RenderX

Database coverage limitations exist despite searching eight major databases. Some relevant studies may have been missed due to database indexing variations, search term limitations, or publication in journals not indexed in the searched databases. Our search strategy, while comprehensive, relied on specific keyword combinations that may not have captured all relevant papers using alternative terminology for IoT technologies or acceptance constructs. Gray literature exclusion means we did not systematically search conference proceedings, dissertations, technical reports, or unpublished studies beyond what was captured in Google Scholar. This decision, made to focus on peer-reviewed evidence, may introduce publication bias as studies with positive or significant findings are more likely to be published in peer-reviewed journals than those with null or negative results. The true extent of failed implementations or low acceptance may be underrepresented in our synthesis.

There was limited subgroup analysis due to available data and reporting heterogeneity. While we identified important moderating factors (digital literacy, age, and health conditions), many included studies did not report sufficient data for systematic subgroup analyses. Our moderator findings therefore rely on studies that explicitly examined these interactions rather than systematic subgroup meta-analyses across all studies. Cross-sectional design dominance among included studies limits our ability to assess causality or track acceptance patterns over time. Most studies examined acceptance at single time points, meaning our understanding of how factors influence initial adoption may differ from factors affecting continued use, but evidence on this distinction is limited. Longitudinal research is critically needed to understand the temporal dynamics of IoT health care acceptance.

Future research needs to address these limitations through more diverse geographic representation, standardized measurement approaches, longitudinal study designs, inclusive recruitment strategies, and systematic attention to negative cases and implementation failures as important learning sources about barriers and facilitators for sustainable health care IoT adoption. Despite these limitations, we believe our systematic approach, comprehensive database searching, rigorous quality assessment, and transparent synthesis methods provide a reliable synthesis of current evidence on patient acceptance of IoT in health care. The consistency of main findings across studies with different designs, populations, and contexts strengthens confidence in core conclusions regarding the importance of perceived usefulness, trust, digital literacy, and the need for UCD approaches.

Practical Implications of Our Approach

This SLR provides significant contributions to diverse stakeholders in health care service ecosystems with comprehensive benefits. From theoretical perspectives, this SLR contributes to the body of knowledge about technology acceptance in health care service contexts by providing a comprehensive synthesis of factors affecting patient adoption, building and expanding existing theories.

For technology developers: implementation teams should prioritize UCD and onboarding that explicitly reduces perceived effort and cognitive burden, particularly for older adults and low-literacy users, because ease-of-use and support features repeatedly emerge as acceptance-enabling conditions in the included literature.

For health care organizations: implementation strategies must be informed by a comprehensive synthesis of facilitators, barriers, and proven enhancement approaches, enabling evidence-based deployment planning.

For policymakers: equity considerations and regulatory frameworks or guidelines need to be identified through systematic analysis of barriers and contextual variations across different populations and settings that facilitate adoption while ensuring patient safety. Policymakers should strengthen governance frameworks that increase patient trust—particularly regarding privacy, security, and accountability—because security and privacy concerns are among the most frequently reported acceptance barriers in this review.

For researchers: identification of methodological gaps, under-researched populations and contexts, and priority areas for future investigation to advance both theory and practice. Future studies should move beyond cross-sectional designs by using longitudinal and implementation-focused methods and by reporting standardized acceptance constructs and context variables to enable stronger synthesis and eventual meta-analytic work.

Future Research Directions

Based on gaps identified in current literature and emerging trends in health care IoT development, several important research priorities emerge that can advance theoretical understanding and practical implementation of patient-centered health care IoT solutions. Longitudinal research represents a primary priority given the current lack of understanding about how adoption patterns, user experiences, and health outcomes evolve over extended periods with health care IoT technology. Multiyear cohort studies tracking users from initial adoption to long-term use can provide important insights into factors driving continued engagement, technology abandonment patterns, evolution in user needs and preferences, and actual health usage and health care impacts of long-term IoT use [30,35]. Comparative research across various cultural, economic, and health care system contexts emerges as a second major priority, given the current geographic bias in the literature and the need to understand how contextual factors affect adoption patterns and optimal implementation strategies. Cross-cultural studies can illuminate how factors such as collectivism vs individualism, power distance, uncertainty avoidance, and other cultural dimensions affect attitudes toward health monitoring, data sharing, technology adoption, and relationships with health care providers in health care IoT system contexts [22,48]. Economic comparative research is also needed to understand how different health care financing systems, insurance models, and economic development levels affect access and sustainability of health care IoT interventions.

Implementation science research represents the third critical priority, focusing on understanding how evidence-based health care IoT interventions can be successfully translated into real-world practice settings with diverse stakeholders, competing

XSL•FO
RenderX

priorities, resource constraints, and complex organizational dynamics. Implementation research can answer questions about optimal training models for health care providers, effective strategies for patient orientation and support, sustainable financing models for long-term technology maintenance and updates, integration approaches minimizing workflow disruption, and scaling strategies maintaining fidelity while adapting to local contexts [60,69]. Methodological innovation represents the fourth priority area, with a need to develop new research approaches that can better capture the complexity of health care IoT adoption and use. This includes the development of standardized outcome measures for evaluating health care IoT interventions, the integration of objective usage data with subjective reports for comprehensive understanding of user experiences, participatory research methods involving patients as research partners in designing and evaluating health care IoT solutions, and mixed reality research approaches that can study health care IoT use in natural settings while maintaining experimental control [44,50]. Emerging technology research represents the fifth priority, given rapid innovation in fields such as AI integration, voice interfaces, augmented reality, blockchain applications, and edge computing that can significantly change health care IoT capability landscapes and user experiences. Research in these emerging areas needs to proactively address user acceptance factors, ethical implications, implementation challenges, and potential unintended consequences before technologies are widely deployed, rather than studying acceptance after technologies are already established in markets [27,62].

Conclusions

This SLR of 62 studies published between 2016 and 2024 provides comprehensive evidence that patient acceptance of IoT technology in health care represents a complex, multidimensional phenomenon requiring holistic approaches that extend beyond technical sophistication to encompass human, organizational, and systemic considerations. With appropriate attention to patient perspectives, UCD, digital equity, and multilevel implementation strategies, IoT technology holds substantial potential to transform health care delivery toward more proactive, personalized, accessible, and effective services. However, realizing this potential requires sustained commitment to placing patients at the center of technology design, implementation, and evaluation processes, ensuring technological advancement serves human flourishing rather than merely demonstrating technical capabilities. The evidence synthesized in this review provides a foundation for such patient-centered approaches, but continued research, evaluation, and adaptation will be essential as both technologies and health care contexts continue to evolve.

Acknowledgments

We want to convey our gratitude to Universitas Indonesia for the continuous support, particularly from the Directorate of Research and Community Engagement. Generative AI tools were not used to generate scientific content, literature searching, data interpretation, or analysis. Grammarly (Superhuman Platform Inc) was used to assist with English language editing, grammar checking, and improving readability, and all text and references were reviewed and verified by the authors to ensure accuracy and originality.

Funding

This research received no specific funding from any funding agency in public, commercial, or not-for-profit sectors.

Data Availability

All data supporting the findings of this systematic literature review are available within this paper and its supplementary materials. The complete PRISMA 2020 checklist, detailed search strategies for all eight databases, screening decisions log, data extraction forms, and quality assessment ratings are provided as Multimedia Appendices. This study's characteristics table for all 62 included studies is provided as Multimedia Appendix 6 [18,19,22-32,34-82]. Raw data from individual included studies can be accessed from the original publications as cited in our reference list. Additional materials, including search strategy documentation and analysis files, are available in the OSF (Open Science Framework) repository [84]. Requests for additional data or clarification should be directed to the corresponding author.

Authors' Contributions

Conceptualization: ZA
Data curation: ZA
Formal analysis: ZA
Investigation: ZA
Methodology: ZA, PWH
Project administration: ZA
Supervision: PWH
Validation: PWH
Visualization: ZA
Writing – original draft: ZA

XSL•FO
RenderX

Writing – review & editing: ZA, PWH

Conflicts of Interest

None declared.

Multimedia Appendix 1
Database search strategy for internet of things literature review.
[PDF File, 22 KB - mhealth_v14i1e81260_app1.pdf]

Multimedia Appendix 2
Internet of things systematic literature review data extraction table (version 1.2).
[DOCX File, 113 KB - mhealth_v14i1e81260_app2.docx]

Multimedia Appendix 3
Mixed Methods Appraisal Tool quality assessment for internet of things systematic literature review.
[XLSX File, 14 KB - mhealth_v14i1e81260_app3.xlsx]

Multimedia Appendix 4
Summary of methodological quality appraisal (Mixed Methods Appraisal Tool Assessment).
[PDF File, 30 KB - mhealth_v14i1e81260_app4.pdf]

Multimedia Appendix 5
Strategies for enhancing patient acceptance of internet of things in health care(version 1.2).
[DOCX File, 26 KB - mhealth_v14i1e81260_app5.docx]

Multimedia Appendix 6
Table study characteristics (version 1.2).
[XLSX File, 20 KB - mhealth_v14i1e81260_app6.xlsx]

Checklist 1
PRISMA 2020 27-Item Checklist V1.2.
[PDF File, 199 KB - mhealth_v14i1e81260_app7.pdf]

References

1. Yee L, Chui M, Roberts R, Issler M. McKinsey technology trends outlook 2024. McKinsey & Company. 2024 Jul 16. URL: https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/the-top-trends-in-tech-2024 [accessed 2026-07-22]
2. IPCC. URL: https://www.ipcc.ch/2024 [accessed 2026-07-22]
3. Khan MS, Shahid I, Siddiqi TJ, et al. Ten-year trends in enrollment of women and minorities in pivotal trials supporting recent US Food and Drug Administration approval of novel cardiometabolic drugs. J Am Heart Assoc 2020 Jun 2;9(11):e015594. [doi: 10.1161/JAHA.119.015594] [Medline: 32427023]
4. Khristich S, Nazarov V. Top 8 telemedicine trends and technologies in 2025. Tateeda Global. 2025 Jan 4. URL: https://tateeda.com/blog/future-of-telemedicine [accessed 2026-07-22]
5. Top healthcare technology trends in 2026. Acropolium. 2026 Jan 6. URL: https://acropolium.com/blog/top-healthcare-technology-trends [accessed 2026-07-22]
6. Turcu CE, Turcu CO. Internet of Things as key enabler for sustainable healthcare delivery. Procedia Soc Behav Sci 2013 Feb;73:251-256. [doi: 10.1016/j.sbspro.2013.02.049]
7. Agustini SA, Mulyatno IP, Amiruddin W. Perancangan sistem digitalisasi monitoring data RPM engine dan temperatur minyak lubrikan kapal berbasis internet of things (IoT) dengan metode simulasi prototype ESP32 [Article in Indonesian]. J Tek Perkapalan 2023 Dec;11(4) [FREE Full text]
8. Aeberhard JL, Radan AP, Soltani RA, et al. Introducing artificial intelligence in interpretation of foetal cardiotocography: medical dataset curation and preliminary coding-an interdisciplinary project. Methods Protoc 2024 Jan 4;7(1):5. [doi: 10.3390/mps7010005] [Medline: 38251198]
9. Al-Shargabi B, Abuarqoub S. IoT-enabled healthcare: benefits, issues and challenges . Presented at: ICFNDS '20; Nov 26-27, 2020; St Petersburg Russian Federation p. 1-5. [doi: 10.1145/3440749.3442596]
10. Nassra I, Capella JV. Data compression techniques in IoT-enabled wireless body sensor networks: a systematic literature review and research trends for QoS improvement. Internet Things 2023 Oct;23:100806. [doi: 10.1016/j.iot.2023.100806]

XSL•FO
RenderX

11. Atzori L, Iera A, Morabito G. The internet of things: a survey. Comput Netw 2010 Oct;54(15):2787-2805. [doi: 10.1016/j.comnet.2010.05.010]
12. Islam SMR, Kabir MH, Hossain M. The internet of things for health care: a comprehensive survey. IEEE Access 2015;3:678-708. [doi: 10.1109/ACCESS.2015.2437951]
13. Arnott M. Global IoT forecast report, 2023-2033. Transforma Insights. 2024 May 10. URL: https://transformainsights.com/research/reports/global-iot-forecast-report-2023-2033 [accessed 2026-07-22]
14. Singh RP, Javaid M, Haleem A, Suman R. Internet of things (IoT) applications to fight against COVID-19 pandemic. Diabetes Metab Syndr 2020;14(4):521-524. [doi: 10.1016/j.dsx.2020.04.041] [Medline: 32388333]
15. Brar PS, Shah B, Singh J, Ali F, Kwak D. Using modified technology acceptance model to evaluate the adoption of a proposed IoT-based indoor disaster management software tool by rescue workers. Sensors (Basel) 2022 Feb 26;22(5):1866. [doi: 10.3390/s22051866] [Medline: 35271012]
16. Attig C, Franke T. Why do people abandon activity trackers? The role of user diversity in discontinued use. Int J Hum Comput Interact 2023 May 9;39(8):1662-1674. [doi: 10.1080/10447318.2022.2067935]
17. Shi Y, Siddik AB, Masukujjaman M, Zheng G, Hamayun M, Ibrahim AM. The antecedents of willingness to adopt and pay for the IoT in the agricultural industry: an application of the UTAUT 2 theory. Sustainability 2022;14(11):6640. [doi: 10.3390/su14116640]
18. Al-Rawashdeh M, Keikhosrokiani P, Belaton B, Alawida M, Zwiri A. IoT adoption and application for smart healthcare: a systematic review. Sensors (Basel) 2022 Jul 19;22(14):5377. [doi: 10.3390/s22145377] [Medline: 35891056]
19. Ghiwaa T, Khan I, White M, Beloff N. Telemedicine adoption for healthcare delivery: a systematic review. Int J Adv Comput Sci Appl 2023 Nov;14(11):456-468. [doi: 10.14569/IJACSA.2023.01411125]
20. Shanbehzadeh M, Nopour R, Kazemi-Arpanahi H. Internet of Things (IoT) adoption model for early identification and monitoring of COVID-19 cases: a systematic review. Int J Prev Med 2022;13:112. [doi: 10.4103/ijpvm.IJPVM_667_20] [Medline: 36247189]
21. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
22. Zhang GW, Li B, Gu ZM, et al. In-depth examination of the functionality and performance of the internet hospital information platform: development and usability study. J Med Internet Res 2024 Nov 8;26:e54018. [doi: 10.2196/54018] [Medline: 39168813]
23. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan;25(1):30-36. [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
24. Kronlid C, Brantnell A, Elf M, Borg J, Palm K. Sociotechnical analysis of factors influencing IoT adoption in healthcare: a systematic review. Technol Soc 2024 Sep;78:102675. [doi: 10.1016/j.techsoc.2024.102675]
25. Yang Q, Al Mamun A, Hayat N, Jingzu G, Hoque ME, Salameh AA. Modeling the intention and adoption of wearable fitness devices: a study using SEM-PLS analysis. Front Public Health 2022 Jul 6;10:912303. [doi: 10.3389/fpubh.2022.918989]
26. Tu JC, Luo SC, Lee YL, Shih MF, Chiu SP. Exploring usability and patient attitude towards a smart hospital service with the technology acceptance model. Int J Environ Res Public Health 2022 May 16;19(10):6059. [doi: 10.3390/ijerph19106059] [Medline: 35627595]
27. Messinis S, Temenos N, Protonotarios NE, Rallis I, Kalogeras D, Doulamis N. Enhancing internet of medical things security with artificial intelligence: a comprehensive review. Comput Biol Med 2024 Mar;170:108036. [doi: 10.1016/j.compbiomed.2024.108036] [Medline: 38295478]
28. Alhasan A, Audah L, Ibrahim I, Al-Sharaa A, Al-Ogaili AS, M. Mohammed J. A case-study to examine doctors' intentions to use IoT healthcare devices in Iraq during COVID-19 pandemic. Int J Pervasive Comput 2022 Nov 25;18(5):527-547. [doi: 10.1108/IJPCC-10-2020-0175]
29. Wakili A, Bakkali S. Internet of things in healthcare: an adaptive ethical framework for IoT in digital health. Clin eHealth 2024 Dec;7:92-105. [doi: 10.1016/j.ceh.2024.07.001]
30. Fraterman I, Wollersheim BM, Tibollo V, et al. An eHealth app (CAPABLE) providing symptom monitoring, well-being interventions, and educational material for patients with melanoma treated with immune checkpoint inhibitors: protocol for an exploratory intervention trial. JMIR Res Protoc 2023 Oct 11;12:e49252. [doi: 10.2196/49252] [Medline: 37819691]
31. Calvillo-Arbizu J, Román-Martínez I, Reina-Tosina J. Internet of things in health: requirements, issues, and gaps. Comput Methods Programs Biomed 2021 Sep;208:106231. [doi: 10.1016/j.cmpb.2021.106231] [Medline: 34186337]
32. Rejeb A, Rejeb K, Treiblmaier H, et al. The internet of things (IoT) in healthcare: taking stock and moving forward. Internet Things 2023 Jul;22:100721. [doi: 10.1016/j.iot.2023.100721]
33. Methley AM, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. BMC Health Serv Res 2014 Nov 21;14:579. [doi: 10.1186/s12913-014-0579-0] [Medline: 25413154]
34. Dutta A, Kovid RK, Thatha M, Gupta J. Adoption of IoT-based healthcare devices: an empirical study of end consumers in an emerging economy. Paladyn J Behav Robotics 2023 Jul 29;14(1):20220121. [doi: 10.1515/pjbr-2022-0106]

XSL•FO
RenderX

35. Hossain MI, Yusof AF, Hussin ARC, lahad NA, Sadiq AS. Factors influencing adoption model of continuous glucose monitoring devices for internet of things healthcare. Internet Things 2021 Sep;15:100353. [doi: 10.1016/j.iot.2020.100353]
36. Hui CY, McKinstry B, Fulton O, Buchner M, Pinnock H. Patients' and clinicians' perceived trust in internet-of-things systems to support asthma self-management: qualitative interview study. JMIR mHealth uHealth 2021 Jul 16;9(7):e24127. [doi: 10.2196/24127] [Medline: 34269684]
37. Wu TC, Ho CTB. Barriers to telemedicine adoption during the COVID-19 pandemic in Taiwan: comparison of perceived risks by socioeconomic status correlates. Int J Environ Res Public Health 2023 Feb 16;20(4):3504. [doi: 10.3390/ijerph20043504] [Medline: 36834205]
38. Alruwaili MM, Shaban M, Ramadan OME. Digital health interventions for promoting healthy aging: a systematic review of adoption patterns, efficacy, and user experience. Sustainability 2023;15(23):16503. [doi: 10.3390/su152316503]
39. Mittelstadt B. Ethics of the health-related internet of things: a narrative review. Ethics Inf Technol 2017 Sep;19(3):157-175. [doi: 10.1007/s10676-017-9426-4]
40. Lu Y, Papagiannidis S, Alamanos E. Internet of things: a systematic review of the business literature from the user and organisational perspectives. Technol Forecast Soc Change 2018 Nov;136:285-297. [doi: 10.1016/j.techfore.2018.01.022]
41. Gellert GA, Rasławska-Socha J, Marcjasz N, et al. How virtual triage can improve patient experience and satisfaction: a narrative review and look forward. Telemed Rep 2023;4(1):292-306. [doi: 10.1089/tmr.2023.0037] [Medline: 37817871]
42. Zeadally S, Bello O. Harnessing the power of internet of things based connectivity to improve healthcare. Internet Things 2021 Jun;14:100074. [doi: 10.1016/j.iot.2019.100074]
43. Ben Arfi W, Ben Nasr I, Khvatova T, Ben Zaied Y. Understanding acceptance of eHealthcare by IoT natives and IoT immigrants: an integrated model of UTAUT, perceived risk, and financial cost. Technol Forecast Soc Change 2021 Feb;163:120437. [doi: 10.1016/j.techfore.2020.120437]
44. Grant S, Blom AW, Craddock I, Whitehouse M, Gooberman-Hill R. Home health monitoring around the time of surgery: qualitative study of patients' experiences before and after joint replacement. BMJ Open 2019 Dec 15;9(12):e032205. [doi: 10.1136/bmjopen-2019-032205] [Medline: 31843836]
45. AlQudah AA, Al-Emran M, Shaalan K. Technology acceptance in healthcare: a systematic review. Appl Sci 2021;11(22):10537. [doi: 10.3390/app112210537]
46. Al-Rawashdeh M, Keikhosrokiani P, Belaton B, Alawida M, Zwiri A. Effective factors for the adoption of IoT applications in nursing care: a theoretical framework for smart healthcare. J Build Eng 2024 Jul;89:109012. [doi: 10.1016/j.jobe.2024.109012]
47. Kos A, Umek A. Wearable sensor devices for prevention and rehabilitation in healthcare: swimming exercise with real-time therapist feedback. IEEE Internet Things J 2019;6(2):1331-1341. [doi: 10.1109/JIOT.2018.2850664]
48. Hayat N, Salameh AA, Malik HA, Yaacob MR. Exploring the adoption of wearable healthcare devices among the Pakistani adults with dual analysis techniques. Technol Soc 2022 Aug;70:102015. [doi: 10.1016/j.techsoc.2022.102015]
49. Bures M, Neumannova K, Blazek P, et al. A sensor network utilizing consumer wearables for telerehabilitation of post-acute COVID-19 patients. IEEE Internet Things J 2022 Dec;9(23):23795-23809. [doi: 10.1109/JIOT.2022.3188914] [Medline: 36514319]
50. Kotronis C, Routis I, Politi E, et al. Evaluating internet of medical things (IoMT)-based systems from a human-centric perspective. Internet Things 2019 Dec;8:100125. [doi: 10.1016/j.iot.2019.100125]
51. Kauw D, Huisma PR, Medlock SK, et al. Mobile health in cardiac patients: an overview on experiences and challenges of stakeholders involved in daily use and development. BMJ Innov 2020 Oct;6(4):184-191. [doi: 10.1136/bmjinnov-2019-000418]
52. Karahoca A, Karahoca D, Aksöz M. Examining intention to adopt to internet of things in healthcare technology products. Kybernetes 2018 Mar 16;47(4):742-770. [doi: 10.1108/K-02-2017-0045]
53. Wu Q, Huang L, Zong J. User interface characteristics influencing medical self-service terminals behavioral intention and acceptance by Chinese elderly: an empirical examination based on an extended UTAUT model. Sustainability 2023;15(19):14252. [doi: 10.3390/su151914252]
54. Du E, Quayle E, Macleod H. A qualitative study of patients' experiences and acceptance of computerised cognitive behavioural therapy in primary care, Scotland. tCBT 2021;14:e21. [doi: 10.1017/S1754470X21000210]
55. Akbulut S, Semantha FH, Azam S, et al. Designing a private and secure personal health records access management system: a solution based on IOTA distributed ledger technology. Sensors (Basel) 2023 May 29;23(11):5174. [doi: 10.3390/s23115174] [Medline: 37299900]
56. Fehringer H, Stary C. User-informed adaptation in IoT home healthcare: grounding development in empirical evidence. J Theor Appl Electron Commer Res 2023;18(4):1901-1925. [doi: 10.3390/jtaer18040096]
57. Altinay L, Alrawadieh Z, Tulucu F, Arici HE. The effect of hospitableness on positive emotions, experience, and well-being of hospital patients. Int J Hosp Manag 2023 Apr;110:103431. [doi: 10.1016/j.ijhm.2023.103431]
58. Bodur G, Gumus S, Gursoy NG. Perceptions of Turkish health professional students toward the effects of the internet of things (IOT) technology in the future. Nurse Educ Today 2019 Aug;79:98-104. [doi: 10.1016/j.nedt.2019.05.007] [Medline: 31112847]

XSL•FO
RenderX

59. Gardas BB. Organizational hindrances to Healthcare 4.0 adoption: a multicriteria decision analysis framework. J Multi-Criteria Decis Anal 2022;29(1-2):186-195. [doi: 10.1002/mcda.1766]
60. Fischer GS, da Rosa Righi R, de Oliveira Ramos G, Costa CD, Rodrigues J. ElHealth: using internet of things and data prediction for elastic management of human resources in smart hospitals. Eng Appl Artif Intell 2020 Jan;87:103285. [doi: 10.1016/j.engappai.2019.103285]
61. Sharma M, Joshi S. Barriers to blockchain adoption in health-care industry: an Indian perspective. J Glob Oper Strategic Sourc 2021 Jul 15;14(1):134-169. [doi: 10.1108/JGOSS-06-2020-0026]
62. Lodha L, Baghela VS, Bhuvana J, Bhatt R. A blockchain-based secured system using the internet of medical things (IOMT) network for e-healthcare monitoring. Meas Sens 2023 Dec;30:100904. [doi: 10.1016/j.measen.2023.100904]
63. Cleveland SM, Haddara M. Internet of things for diabetics: identifying adoption issues. Internet Things 2023 Jul;22:100798. [doi: 10.1016/j.iot.2023.100798]
64. Tortorella GL, Fogliatto FS, Espôsto KF, et al. Effects of contingencies on Healthcare 4.0 technologies adoption and barriers in emerging economies. Technol Forecast Soc Change 2020 Jul;156:120048. [doi: 10.1016/j.techfore.2020.120048]
65. Streeper NM, Dubnansky A, Sanders AB, Lehman K, Thomaz E, Conroy DE. Improving fluid intake behavior among patients with kidney stones: understanding patients' experiences and acceptability of digital health technology. Urology 2019 Nov;133:57-66. [doi: 10.1016/j.urology.2019.05.056] [Medline: 31374289]
66. Kwon H, An S, Lee HY, et al. Review of smart hospital services in real healthcare environments. Healthcare Inf Res 2022 Jan;28(1):3-15. [doi: 10.4258/hir.2022.28.1.3] [Medline: 35172086]
67. Wen MH, Bai D, Lin S, Chu CJ, Hsu YL. Implementation and experience of an innovative smart patient care system: a cross-sectional study. BMC Health Serv Res 2022 Jan 29;22(1):126. [doi: 10.1186/s12913-022-07511-7] [Medline: 35093036]
68. Albahri OS, Albahri AS, Zaidan AA, et al. Fault-tolerant mHealth framework in the context of IoT-based real-time wearable health data sensors. IEEE Access 2019;7:50052-50080. [doi: 10.1109/ACCESS.2019.2910411]
69. Westphal M, Yom-Tov GB, Parush A, et al. A patient-centered information system (myED) for emergency care journeys: design, development, and initial adoption. JMIR Form Res 2020 Feb 25;4(2):e16410. [doi: 10.2196/16410] [Medline: 32130144]
70. Sahin E, Kefeli U, Cabuk D, et al. Perceptions and acceptance of telemedicine among medical oncologists before and during the COVID-19 pandemic in Turkey. Support Care Cancer 2021 Dec;29(12):7497-7503. [doi: 10.1007/s00520-021-06290-x] [Medline: 34100135]
71. Binci D, Palozzi G, Scafarto F. Toward digital transformation in healthcare: a framework for remote monitoring adoption. TQM j 2022 Nov 29;34(6):1772-1799. [doi: 10.1108/TQM-04-2021-0109]
72. Cahill J, Portales R, McLoughin S, Nagan N, Henrichs B, Wetherall S. IoT/sensor-based infrastructures promoting a sense of home, independent living, comfort, and wellness. Sensors (Basel) 2019 Jan 24;19(3):485. [doi: 10.3390/s19030485] [Medline: 30682864]
73. Arfi WB, Nasr IB, Kondrateva G, Hikkerova L. The role of trust in intention to use the IoT in eHealth: application of the modified UTAUT in a consumer context. Technol Forecast Soc Change 2021 Jun;167:120688. [doi: 10.1016/j.techfore.2021.120688]
74. Hjelm K, Hedlund L. Internet-of-things (IoT) in healthcare and social services - experiences of a sensor system for notifications of deviant behaviours in the home from the users' perspective. Health Inf J 2022;28(1):14604582221075562. [doi: 10.1177/14604582221075562] [Medline: 35225069]
75. Lee JH, Park YR, Kweon S, Kim S, Ji W, Choi CM. A cardiopulmonary monitoring system for patient transport within hospitals using mobile internet of things technology: observational validation study. JMIR mHealth uHealth 2018 Nov 14;6(11):e12048. [doi: 10.2196/12048] [Medline: 30429115]
76. Martínez-Caro E, Cegarra-Navarro JG, García-Pérez A, Fait M. Healthcare service evolution towards the internet of things: an end-user perspective. Technol Forecast Soc Change 2018 Nov;136:268-276. [doi: 10.1016/j.techfore.2018.03.025]
77. Mittelstadt B. Designing the health-related internet of things: ethical principles and guidelines. Information 2017 Jul 1;8(3). [doi: 10.2139/ssrn.2943006]
78. Pulimamidi R. To enhance customer (or patient) experience based on IoT analytical study through technology (IT) transformation for E-healthcare. Meas: Sens 2024 Jun;33:101087. [doi: 10.1016/j.measen.2024.101087]
79. Ramdani B, Duan B, Berrou I. Exploring the determinants of mobile health adoption by hospitals in China: empirical study. JMIR Med Inf 2020 Jul 14;8(7):e14795. [doi: 10.2196/14795] [Medline: 32459630]
80. Ryu B, Kim S, Lee KH, Hwang H, Yoo S. Inpatient satisfaction and usage patterns of personalized smart bedside station system for patient-centered service at a tertiary university hospital. Int J Med Inf 2016 Nov;95:35-42. [doi: 10.1016/j.ijmedinf.2016.09.003] [Medline: 27697230]
81. Ziwei H, Dongni Z, Man Z, et al. The applications of internet of things in smart healthcare sectors: a bibliometric and deep study. Heliyon 2024 Feb 15;10(3):e25392. [doi: 10.1016/j.heliyon.2024.e25392] [Medline: 38356528]
82. Zobair KM, Houghton L, Tjondronegoro D, et al. Systematic review of internet of medical things for cardiovascular disease prevention among Australian first nations. Heliyon 2023 Nov;9(11):e22420. [doi: 10.1016/j.heliyon.2023.e22420] [Medline: 38074865]

XSL•FO
RenderX

83. Ziwei H, Dongni Z, Man Z, et al. Applications of the internet of things in smart healthcare sectors: a bibliometric and deep study. Heliyon 2024 Feb;10(3):e25025. [doi: 10.1016/j.heliyon.2024.e25025]
84. Arifin Z. IOT adoption in healthcare: a systematic review. OSF. URL: https://osf.io/vy67b/overview [accessed 2026-07-03]

Abbreviations

IoMT: internet of medical things
IoT: internet of things
mHealth: mobile health
MMAT: Mixed Methods Appraisal Tool
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RQ: research question
SLR: systematic literature review
TAM: technology acceptance model
UCD: user-centered design
UTAUT: unified theory of acceptance and use of technology

Edited by L Buis; submitted 30.Jul.2025; peer-reviewed by F Branco, LP Hung; revised version received 16.Jun.2026; accepted 16.Jun.2026; published 31.Jul.2026.

Please cite as:
Arifin Z, Handayani PW
Patient Acceptance and Barriers to IoT Usage in Health Care: Systematic Literature Review
JMIR Mhealth Uhealth 2026;14:e81260
URL: https://mhealth.jmir.org/2026/1/e81260
doi:10.2196/81260

© Zaenal Arifin, Putu Wuri Handayani. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 31.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Machine Learning Frameworks for Wearable-Based Stress Modeling in Naturalistic Settings: Scoping Review

Shifali Sharma, MS; Aswin Kumar Janakiraman, MS; Lujie Karen Chen, PhD

Department of Information Systems, University of Maryland, Baltimore County, 1000 Hilltop Cir, Baltimore, MD, United States

Corresponding Author:
Lujie Karen Chen, PhD
Department of Information Systems, University of Maryland, Baltimore County, 1000 Hilltop Cir, Baltimore, MD, United States

Abstract

Background: Stress, as commonly recognized, is an integral part of modern life and can significantly affect both mental and physical health. While substantial advancements have been made in measuring physical fitness through wearable devices, the detection and assessment of mental stress remain in their early stages.

Objective: The objective of this paper is to review recent studies of wearable-based stress detection in naturalistic settings, with a specific focus on characterizing machine learning frameworks inspired by the model card approach.

Methods: This review was conducted using the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist. A total of 353 articles were identified through searches in databases such as PubMed, MEDLINE, ScienceDirect, IEEE, ACM Digital Library, Web of Science, and Embase. Studies were considered eligible if they collected data from healthy adults in naturalistic settings using wearable devices and used machine learning models for stress detection.

Results: A total of 34 articles met the eligibility criteria, including 11 conference papers, 22 journal articles, and 1 preprint published between 2017 and 2024. From these studies, we analyzed key machine learning modeling decisions such as problem formulation, ground truth determination, and machine learning algorithms. Additionally, we examined the major contributions of each study, focusing on the challenges they addressed and the solutions they proposed. Based on these findings, we proposed a model card framework for reporting machine learning–based, wearable-based stress detection.

Conclusions: This scoping review highlights recent trends in machine learning models for stress detection and measurement using wearable signals. It underscores the need for improved standardization in reporting practices for datasets and key machine learning decisions, as well as the importance of addressing critical challenges associated with data collection in real-world settings. We hope this review will support and strengthen ongoing research efforts, promote knowledge sharing, and promote collaboration among researchers—ultimately advancing the field as a community.

(JMIR Mhealth Uhealth 2026;14:e76632) doi:10.2196/76632

KEYWORDS

mental health; stress detection; naturalistic setting; in the wild; field studies; wearables; physiology; machine learning models; scoping review

Introduction

Stress affects the well-being of many individuals in modern society [1,2]. From an evolutionary perspective, stress is a fight-or-flight response to perceived threat or danger, which prompts necessary survival actions to flee from dangerous situations [3]. However, sustained unmanaged stress may have long-term effects on both mental and physical health [4]. In extreme conditions, it may lead to serious adverse health outcomes such as depression, cardiovascular diseases [5-7], substance abuse, drugs and alcohol addiction, or self-harm behaviors such as suicide [8]. This underscores the importance of timely interventions or management of stress in daily life. Stress detection is fundamental to this type of support system, which permits continuous monitoring of stressful states in naturalistic settings, that is, everyday life. Given their nonintrusive and user-friendly design [9], wearables have been explored to monitor physical fitness, using metrics such as step counts or heart rates. In recent years, there have been emerging interests in using wearables to understand individuals' psychological fitness or well-being, in which stress detection and measurement are the major driving components.

Although stress detection from instruments such as wearables is not new, detecting stress in naturalistic settings poses significant new challenges. For studies conducted in controlled settings [10-12] in laboratory settings, researchers prescribe stress-inducing activities that participants engage with, such as the Trier Social Stress Test, the Stroop Color-Word Interference Test, the Montreal Imaging Stress Task, the Cold Pressor Test,

XSL•FO
RenderX

and the "Sing-a-Song" Stress Test [13]. Similarly, in studies outside of the laboratory, participants may participate in activities that are known to induce stress for some, such as hackathons, driving, or presentations [14]. In contrast, stress detection in naturalistic settings requires individuals to be involved in normal daily activities, often with significant uncertainty regarding the timing, duration, stressor, and stress responses. Moreover, there is additional complexity in dealing with motion artifacts of wearable signals resulting from ambulatory participants [13]. Both issues need to be appropriately addressed for reliable stress detection models.

The massive amounts of high-resolution data collected from wearables provide opportunities to leverage advanced analytics, machine learning (ML), and AI models for stress detection. In recent years, there has been a growing trend in stress detection research using ML as the primary modeling technique. This trend is supported by wearables data collected unobtrusively in naturalistic, real-world settings, driven in part by the availability of large-scale, open-source datasets that attempt to measure psychological constructs such as stress.

While several literature reviews have evaluated studies on wearable-based stress detection, most have focused on cataloging wearables, including their models, sensor types, and factors such as placement, cost, and usability [1,3,15-17].

Several prior reviews have examined aspects of ML-based stress modeling. For example, Namvari et al [16] focused on classification models, which discussed feature engineering approaches and summarized model performance. However, only 5 of the 23 studies included in their review were conducted exclusively in real-life or naturalistic settings. Similarly, the scoping review by Bolpagni et al [3] evaluated 56 studies, of which only 13 were conducted in real-world contexts. Their review primarily focused on preprocessing pipelines, feature extraction methods, and ML model types. Pinge et al [18] reviewed 39 studies in which stressors were predominantly derived from laboratory-induced or controlled stress-inducing stimuli, such as public speaking tasks or exposure to horror movies. Only a small subset of the reviewed studies involved those from free-living conditions. This review summarized preprocessing strategies, feature computation methods, ML techniques (including both classical and deep learning models), and commonly used performance metrics.

In summary, there are several notable gaps in the wearable-based ML-focused review for stress modeling. (1) None of these reviews focus exclusively on studies using data collected in real-world, naturalistic settings. (2) None of the reviews address methodological rigor explicitly, for example, by evaluating whether models use appropriate experimental setups to avoid data leakage [19-22]. This issue is particularly relevant but easy to overlook in wearable studies with repeated measures from the same participants, and failure to account for it can lead to overestimation of model performance. (3) None of the reviews focus on important problem formulation decisions, such as how to appropriately configure the training dataset so that input and output windows are correctly aligned so that it matches the modeling objectives, for example, whether the goal is to detect stress in the present moment (ie, nowcasting) or predict stress in the future (ie, forecasting). (4) There is no review attempting to explicitly tackle the standardization of reporting, for example, by adopting an existing ML model reporting framework such as model card [23]. This scoping review aims to address these critical gaps.

The main contributions of this paper are as follows:

1. Provided a focused review of recent studies on wearable-based stress detection exclusively in naturalistic settings using ML techniques.
2. Analyzed model performance with a focus on methodological rigor, explicitly examining potential threats to validity (eg, data leakage).
3. Analyzed key problem formulation decisions such as input and output window configuration, and distinctions between nowcasting and forecasting problems.
4. Proposed standardized report ML framework in wearable-based stress detection, inspired by the model card model reporting framework.

This review paper is organized as follows: The "Methods" section describes the process used to select papers for review, including eligibility criteria and the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram. The "Results" section presents the dataset features and ML frameworks used in the selected studies. Finally, the "Discussion" section analyzes the findings, identifies limitations, and provides recommendations for future research.

Methods

Overview

We adopt a scoping review approach to systematically identify gaps in the existing literature. Although stress detection has long been studied in laboratory settings, it remains challenging to translate these findings into real-life contexts. As such, our primary goal is to map current research on stress detection, identify existing gaps, and provide an overview of ML frameworks inspired by the concept of model cards. Specifically, this review examines how ML frameworks are used to detect stress in naturalistic settings using wearable sensors. The review is guided by the methodological framework proposed by Arksey and O'Malley [24] and adheres to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines (Checklist 1) [25].

Search Strategy

The primary search for this scoping review was conducted in April 2024 across 7 databases: PubMed, MEDLINE, ScienceDirect, IEEE Xplore, ACM Digital Library, Web of Science, and Embase. These databases were chosen for their extensive coverage of recent physiology research on stress detection. Our aim was to identify articles examining stress detection in naturalistic settings using biosignals and wearables. We restricted our search to English-language, peer-reviewed studies published between January 2017 and April 2024 that were available in full text. Initially, we tested various search strings—for instance, using "stress detection" alone yielded too many results, whereas "stress detection" AND ("naturalistic setting" OR "field study" OR "real life") yielded too few.

XSL•FO
RenderX

Ultimately, using more generic terms such as “stress detection,” “wearables,” and “physiology,” combined with AND operators, produced a sufficiently broad yet focused list of results. Research articles involving either primary or secondary data analyses (or both) were included. Primary data analysis refers to studies where data were collected by the authors, whereas secondary data analysis refers to studies using data collected by other researchers, which were either openly available or accessible upon request.

Eligibility Criteria

Studies were deemed eligible for inclusion if they used data collected in naturalistic (real-world or in-the-wild) settings. Here, naturalistic refers to situations in which stress is not experimentally induced or linked to known stress-inducing events (eg, taking a test), but instead reflects stress as it occurs in daily life. This criterion applied even if some studies use multiple datasets, including those from controlled or laboratory environments, provided that at least one dataset was collected in the wild. This decision was motivated by the fact that stress labeling mechanisms in laboratory studies differ fundamentally from those used in real-life or naturalistic contexts, where stress must be explicitly self-reported or otherwise captured from participants during their daily lives, which pose unique challenges. Given the study’s focus on acute stress, only research involving healthy populations in everyday contexts was considered. Consequently, the use of wearable devices was a prerequisite, as they facilitate continuous monitoring in daily life. Recognizing the surge in wearable technology adoption in recent years, the review encompassed studies published between 2017 and 2024. The specific eligibility criteria pertinent to this scoping review are detailed in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Data were collected in naturalistic settings where participants engaged in normal, everyday activities (including work).
- Healthy participants were involved. At least one type of data was collected using wearable devices to capture physiological and/or psychological signals.
- Machine learning models or other advanced data-driven techniques were used to detect stress.
- The study needs to be published in English.

Exclusion criteria

- Data were collected only in controlled settings (eg, in laboratory environments or when participants engaged in known stress-inducing activity such as a hackathon as described in the “Introduction”).
- The participants had been diagnosed with mental health disorders or other chronic conditions.
- No wearable devices were used in the data collection process.
- No machine learning or other advanced data-driven modeling techniques were used to detect stress.
- The study was not published in English.

Title and Abstract Screening

We began by applying basic eligibility filters (language and publication year) within the scientific databases themselves. After eliminating duplicate entries, the first author compiled all the retrieved records into a spreadsheet and conducted the initial title and abstract screening using preagreed eligibility criteria. The third author reviewed and verified the screened records and decisions regarding the inclusion and exclusion in the spreadsheet against the same criteria. During abstract screening, we identified publicly available, on-request datasets containing naturalistic wearable data for stress detection and reviewed their citations, adding any studies that used these datasets to our review. Any ambiguities or discrepancies were resolved through discussion until consensus was reached. Since formal full parallel double-screening was not performed, Cohen κ or percentage agreement was not calculated. Finally, the remaining full-text articles were reassessed for eligibility using the specified inclusion and exclusion criteria.

Full-Text Review, Data Extraction, and Analysis

Data extraction involved classifying studies into predefined categories to facilitate evidence synthesis, with a particular focus on the design of ML frameworks inspired by model cards. Extracted data elements included availability of datasets such as indicating whether datasets were open source, available upon request, or collected within the study design, demographic information such as participants' professions, total number of participants, and geographic location, details of wearable devices and biosignals used, duration of data collection, primary research focus such as problem addressed which are either originating from domain-centric or data-centric approaches, and ML models used. Additionally, we examined the formulation of problem statements, including the selection of input and output windows for nowcasting or forecasting tasks. We also assessed validation practices, with particular attention to measures implemented to prevent data leakage, such as appropriate data partitioning and cross-validation techniques.

Results

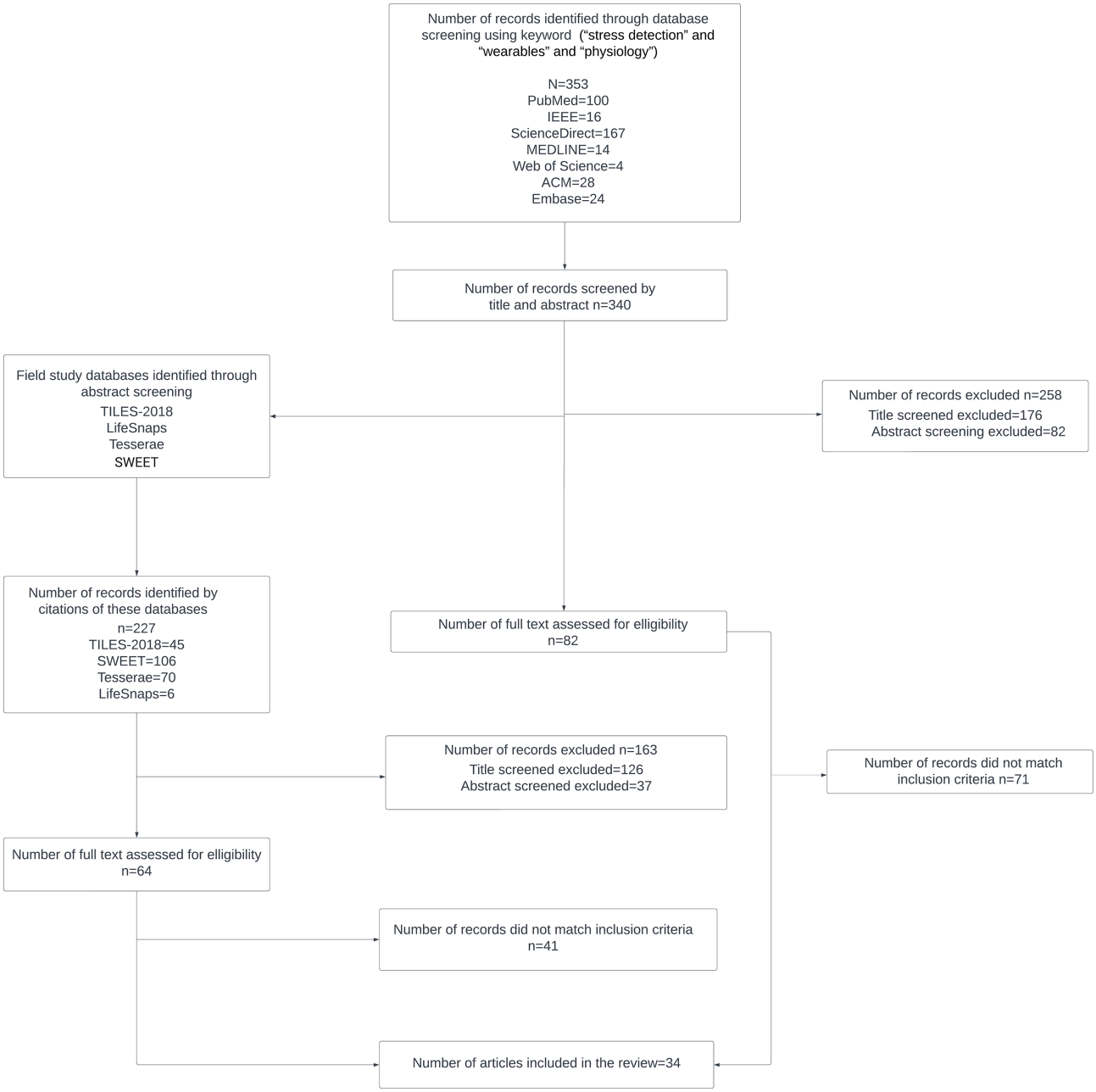
Literature Search

As shown in the flowchart in Figure 1, a total of 353 articles were identified through database searches: 100 from PubMed, 16 from IEEE, 167 from ScienceDirect, 14 from MEDLINE, 4

XSL•FO
RenderX

from Web of Science, 28 from ACM Digital Library, and 24 from Embase. During the title and abstract screening, 258 articles were excluded for not meeting the eligibility criteria.

Figure 1. Flowchart for paper screening.



During the abstract screening phase, we identified additional datasets that were collected using wearables to model wellness-related constructs in naturalistic settings, for example, TILES-2018 [26], Lifesnaps [27], Tesserae [28], and SWEET [29]. In total, we find 227 secondary analysis papers citing those datasets, with 45 of them using TILES-2018, 106 using SWEET, 70 using Tesserae, and 6 using Lifesnaps. After further title and abstract screening, 163 studies were excluded. This left a total of 146 studies considered for full-text review, with a mixture of primary and secondary analyses. Of these, 71 studies were excluded from the primary data analysis category, and 41 studies were excluded from the secondary data analysis category based on the eligibility criteria. In the end, 34 articles were selected for inclusion in this scoping review. For a detailed description of the search and selection strategy using PubMed, please refer to Multimedia Appendix 1.

Study Characteristics

Overview

Among the 34 selected papers, 11 studies were conference proceedings (32%), 22 were journal articles (65%), and 1 was a preprint (3%). The studies were published between 2017 and 2024. Of these, 24 out of 34 (71%) studies conducted secondary data analysis using datasets collected by other researchers—either open-source or available upon request. The remaining 10 out of 34 (29%) studies performed primary data analysis based on datasets collected by the authors specifically for the reported research. Fitbit was the most commonly used

XSL•FO
RenderX

wearable device, followed by Garmin, Empatica E4, and OMsignal garments. The duration of data collection across studies ranged from a single day to several months.

Dataset Characteristics

Although the dataset is not the primary focus of this review, we provide a brief overview of the dataset characteristics since it forms the foundation for the ML models. A comprehensive overview of the datasets used in the included studies is presented in Table 1. For detailed field descriptions, please refer to Multimedia Appendix 2. The frequently-used datasets as part of the secondary data analysis are TILES-2018, SWEET, LifeSnaps, and Tesserae. The populations represented include clinical providers, office and information workers, police officers, military personnel, university students, and remote workers, while a few studies did not specify the population. The number of participants in the datasets ranged from as few as 3 to as many as 1002. The duration of data collection varied, ranging from 1 day to 10 weeks.

XSL•FO
RenderX

Table . Dataset characteristics.

Dataset citation key	Dataset name	Participant pool	Number of participants	Data collection duration	Wearable device used
Mundnich et al (2020) [26]	TILES-2018	Clinical providers	212	10 weeks	Fitbit Charge 2 and OMsignal smart garment
Smets et al (2018) [29]	SWEET	Staff members in technology-oriented, banking, and public sector companies.	1002	5 days	imec's Chillband and chest patch
Yfantidou et al (2022) [27]	Lifesnaps	Participants were recruited through university mailing lists, and although not explicitly stated, they were likely university students from Greece, Cyprus, Italy, and Sweden.	71	4 months (2 waves, each lasting 2 months)	Fitbit Sense
Mattingly et al (2019) [28]	Tesserae	Information workers	757	56 days	Garmin Vivosmart 3, a waterproof wristwatch
Boateng and Kotz (2016) [30]	—[a]	Unknown	10	1 day	Zephyr chest strap and Amulet
Tervonen et al (2020) [31]	—	Office workers	74	4 weeks	Polar M600 smartwatch
Bavaresco et al (2020) [32]	—	Unknown	5	8 days	Polar H7 chest strap
Gjoreski et al (2017) [19]	—	Unknown	5	11 days, on average	Empatica E3 and E4 wrist devices
de Vries et al (2022) [33]	—	Dutch police officers	8	At least 15 weeks, up to 55 weeks.	Oura Ring (Generation 2)
Han et al (2020) [20]	—	Unknown	3	2 days	Empatica E4 wristband and Shimmer3 ECG chest strap
de Vries et al (2023) [34]	—	Dutch military	73	8 weeks	Garmin Tactix Charlie, smartwatch
Tump et al (2022) [35]	—	Remote-working employees of the global insurance company Cigna in 3 locations: the United States, the United Kingdom, and Hong Kong.	198	7 consecutive days	Garmin Vivosmart 4
Mishra et al (2020) [36]	—	University students	26	3 days	Polar H7 chest strap, Amulet wrist device, and a custom GSR[b] sensor
Schmidt et al (2019) [37]	—	University students	11	Approximately 16 days	Empatica E4 wristband

[a]Not available.

[b]GSR: Galvanic skin response.

Wearables Characteristic

The choice of wearables has a direct impact on the quality and diversity of data collected. A wide range of wearables has been used in these studies to unobtrusively collect data, with the most frequently used devices being the Garmin Vivosmart [22,35,38,39], Fitbit Charge 2 [21,40], OmSignal garment [41,42], Empatica E4 [20,37], Unihertz Jelly Pro smartphone [41,43], and the Imec Chillband [44]. Wearable sensors enable the capture of diverse data types across multiple modalities, including physiological data, for example, heart rate, temperature, and skin conductance [19,32,33,37,40,45-47]. Additional data types were also recorded, for example, phone usage, sleep duration, circadian rhythm, and movement

XSL•FO
RenderX

[21,35,40,48-50], audio signals [21,40,41,49], sociodemographic data [47,50], and environmental data (eg, temperature and humidity) [51]. These modalities provide options for studies to use either unimodal [20,30,36] or multimodal approaches [49,52,53].

ML Framework and Model Performances

In this section, we organize the analysis around key methodological decisions identified across the reviewed studies on wearable-based stress detection in naturalistic settings. At a high level, this task was framed as learning a mapping between the physiological signals collected from wearables and the psychological construct of stress. This was often formulated as a supervised learning framework (either a classification or regression problem), where the inputs are wearable sensing data and the outputs are stress labels.

Due to the nature of the data being collected, the reviewed studies varied across several methodological dimensions. We analyze the model frameworks along the following dimensions:

1. Ground truth stress labels: how the labels are collected and the types of labels collected.
2. Input and output windows: how to set up input-output pairs based on corresponding time windows and how the temporal relationships between input and output windows are informed by the problem being addressed—whether it involves in-the-moment stress detection (ie, nowcasting) or predicting future stress (ie, forecasting).
3. ML experiment design: how to appropriately set up ML experiments, for example, split training and testing data appropriately to avoid data leakage.
4. ML models: the primary ML algorithms used in the studies.
5. Challenges and solutions: the specific challenges that the reviewed papers aim to tackle and the main solutions proposed.
6. Analysis of model performance: analysis of model performance with respect to ML and specific solutions to address challenges, with caveats resulting from issues in reporting and methodological rigor.

Overview of Model Characteristics

Across the 34 included studies [19-22,30-59], the large majority adopted a global, population-level modeling approach, with 28 [19-22,30,32-40,42,44,46,48-50,52-59] out of 34 (82%) studies following this approach, while only 6 [31,41,43,45,47,51] out of 34 (18%) studies pursued personalization or built individualized models. With respect to model family, 21 [19,20,22,30-36,39,40,42,44,45,47,49,51,53,54,58] out of 34 (62%) studies used traditional ML, while 8 [21,37,41,43,46,48,56,57] out of 34 (24%) studies relied exclusively on deep learning. A further 4 [38,50,52,55] out of 34 (12%) studies combined traditional ML and deep learning, and 1 [59] out of 34 (3%) studies additionally incorporated a large language model (LLM), bringing the total share of studies that used deep learning in some capacity to 13 [21,37,38,41,43,46,48,50,52,55-57,59] out of 34 (38%) studies. Reporting of evaluation metrics was assessed among the 22 classification studies [19-21,30-32,35-38,42,44-49,52,55-58]. F_1-score was the most commonly reported metric, appearing in 19 [19,21,31,32,35-38,42,44-46,48,49,52,55-58] out of 22 (86%) classification studies, followed by accuracy in 12 [20,21,30-32,38,44,46,47,56-58] out of 22 (55%) studies, precision in 7 [19,32,36,38,52,57,58] out of 22 (32%) studies, and recall in 5 [32,36,38,57,58] out of 22 (23%) studies. Only 3 [42,52,55] out of 22 (14%) studies reported alternative discrimination or balance-aware metrics such as area under the receiver operating characteristic curve (AUC), balanced accuracy, or Matthews correlation coefficient (MCC), suggesting limited adoption of metrics that are robust to class imbalance. Among the 13 regression studies [22,32,34,38-41,43,50,51,53,54,59], all reported at least one quality measure such as R-squared or mean-squared error (MSE). Potential data leakage was identified in roughly half of the corpus, with 19 [30,32-35,38,39,41,42,44,45,47,49,53-56,58,59] out of 34 (56%) studies judged to have at least one source of potential leakage and the remaining 15 [19-22,31,36,37,40,43,46,48,50-52,57] out of 34 (44%) studies judged to be free of obvious leakage. Finally, in terms of temporal problem formulation, the field is dominated by nowcasting, with 27 [19-22,30-39,42-44,46,48,49,51-56,58] out of 34 (79%) studies framing their task as inferring a current or concurrent state; only 4 [40,47,50,57] out of 34 (12%) studies explicitly framed the task as forecasting a future state, and the formulation could not be unambiguously determined for 3 [41,45,59] out of 34 (9%) studies.

Details of the key components of the ML framework are presented in Tables 2 and 3, which summarize the reported best model performance; please refer to Multimedia Appendix 3 for additional details.

XSL•FO
RenderX

Table . Machine learning framework characteristics.

Paper citation key	Dataset citation key	Main problems addressed	Problem formulation	Input window	Output window	Model type	Sample size (N)	Possible data leakage
Boateng and Kotz 2016 [30]	Boateng and Kotz (2016) [30]	No specific problem was addressed	Nowcasting	1 minute or 5 minutes	15 minutes	Classification	10	Yes
Parousidou et al 2023 [45]	Yfantidou et al (2022) [27]	Personalization	Unclear	Unclear	Unclear	Classification	71	Yes
Tervonen et al 2020 [31]	Tervonen et al (2020) [31]	Personalization	Nowcasting	Current day	Current day	Classification	74	No
Bavaresco et al 2020 [32]	Bavaresco et al (2020) [32]	Context awareness	Nowcasting	20 minutes	20 minutes	Classification	5	Yes
Gjoreski et al 2017 [19]	Gjoreski et al (2017) [19]	Context awareness	Nowcasting	A range of windows from 10 minutes to 30 minutes	20 minutes	Classification	5	No
de Vries et al 2022 [33]	de Vries et al (2022) [33]	Temporal dependency	Nowcasting	Previous day	Current day	Regression (statistical)	8	Yes
Han et al 2020 [20]	Han et al (2020) [20]	No specific problem was addressed	Nowcasting	Current day	Current day	Classification	17	No
de Vries et al 2023 [34]	de Vries et al (2023) [34]	Temporal dependency	Nowcasting	Last night's sleep duration	at wake up	Regression (statistical)	73	Yes
Tump et al 2022 [35]	Tump et al (2022) [35]	Context awareness	Nowcasting	3 hours	3 hours	Classification	198	Yes
Mishra et al 2020 [36]	Mishra et al (2020) [36]	No specific problem was addressed	Nowcasting	1 minute	1 minute	Classification	27	No
Schmidt et al 2019 [37]	Schmidt et al (2019) [37]	No specific problem was addressed	Nowcasting	1 minute	1 minute	Classification	12	No
Jiang et al 2023 [54]	Mundnich et al (2020) [26]	Addressing limited supply of training data	Nowcasting	3 days centered on current day	Current day	Regression	212	Yes
Hadjiantonis et al 2020 [40]	Mundnich et al (2020) [26]	Temporal dependency	Forecasting	Previous day	Current day	Regression (statistical)	130	No
Gaballah et al 2021 [21]	Mundnich et al (2020) [26]	Temporal dependency	Nowcasting	Current shift	Current shift	Classification	212	No
Yu and Sano 2022 [48]	Mundnich et al (2020) [26]	Addressing limited supply of training data	Nowcasting	2.5 hours after the label was collected	Current shift	Classification	212	No
Burghardt et al 2021 [52]	Mundnich et al (2020) [26]	Temporal dependency	Nowcasting	3 days centered on the current day	Current day	Classification	212	No
Kao et al 2020 [51]	Mundnich et al (2020) [26]	Personalization	Nowcasting	3 days centered on the current day	Current day	Regression	212	No
Paromita et al 2023 [43]	Mundnich et al (2020) [26]	Personalization	Nowcasting	Current day	Current day	Regression	212	No
Pimentel et al 2021 [55]	Mundnich et al (2020) [26]	No specific problem was addressed	Nowcasting	Current day	Current day	Classification	212	Yes

XSL•FO
RenderX

Paper citation key	Dataset citation key	Main problems addressed	Problem formulation	Input window	Output window	Model type	Sample size (N)	Possible data leakage
Yang et al 2022 [46]	Mundnich et al (2020) [26]	Multiple modalities	Nowcasting	2 hours before the label was collected	Current shift	Classification	212	No
Zanna et al 2022 [56]	Mundnich et al (2020) [26]	Bias mitigation	Nowcasting	2 hours before the label was collected	Current shift	Classification	212	Yes
Tiwari and Falk 2021 [42]	Mundnich et al (2020) [26]	No specific problem was addressed	Nowcasting	Current day	Current day	Classification	212	Yes
Feng and Narayanan 2022 [49]	Mundnich et al (2020) [26]	Context awareness	Nowcasting	Current day	Current day	Classification (stable)	99	Yes
Feng et al 2021 [53]	Mundnich et al (2020) [26]	Context awareness	Nowcasting	Current day	Current day	Regression (stable, statistical)	113	Yes
Ravuri et al 2020 [41]	Mundnich et al (2020) [26]	Personalization	Unclear	Not given	Not given	Regression	154	Yes
Stojchevska et al 2022 [44]	Smets et al (2018) [29]	Context awareness	Nowcasting	1 hour	1 hour	Classification	1002	Yes
Booth et al 2022 [38]	Mattingly et al (2019) [28]	Temporal dependency	Nowcasting	Current day	Current day	Regression and classification	606	Yes
Martinez et al 2022 [39]	Mattingly et al (2019) [28]	No specific problem was addressed	Nowcasting	A range of windows from 5 minutes to 24 hours	Current day	Regression (statistical)	657	Yes
Robles-Granda et al 2021 [22]	Mattingly et al (2019) [28]	Temporal dependency	Nowcasting	Current day	Current day	Regression (stable)	757	No
Saylam and İncel 2023 [50]	Mattingly et al (2019) [28]	Temporal dependency	Forecasting	A range of windows from 1 day to 30 days before	A range of windows from 1 day to 7 days ahead	Regression	757	No
Li et al 2024 [57]	Mattingly et al (2019) [28]	Temporal dependency	Forecasting	A range of windows from 25 days to 35 days before	A range of windows from 12 to 19 days ahead	Classification	478	No
Saylam and Durmaz İncel 2022 [58]	Mattingly et al (2019) [28]	Context awareness	Nowcasting	Current day	Current day	Classification	757	Yes
Kim et al 2024 [59]	Yfantidou et al (2022) [27]	Use of LLM[a] in stress detection	Unclear	Unclear	Unclear	Regression	16	Yes
Paraschou et al 2023 [47]	Yfantidou et al (2022) [27]	No specific problem was addressed	Forecasting	Unclear	Unclear	Classification	71	Yes

[a]LLM: large language model.

XSL•FO
RenderX

Table . Machine learning model performance.

Paper citation key	Classification model					Regression model	Baseline model reported	Uncertainty measure reported
	Accuracy	F_1-score	Precision	Recall	Other measures	Quality measure		
Boateng and Kotz 2016 [30]	1.00	—[a]	—	—	—	N/A[b]	No	No
Parousidou et al 2023 [45]	—	0.66	—	—	—	N/A	Yes	No
Tervonen et al 2020 [31]	0.51	0.62	—	—	—	N/A	Yes	Yes
Bavaresco et al 2020 [32]	0.82	0.75	0.9	0.64	—	N/A	No	No
Gjoreski 2017 et al [19]	—	0.9	0.95	—	—	N/A	No	No
de Vries et al 2022 [33]	N/A	N/A	N/A	N/A	N/A	Adjusted R^2: 0.01 - 0.23	N/A	N/A
Han et al 2020 [20]	1.00	—	—	—	—	N/A	No	No
de Vries et al 2023 [34]	N/A	N/A	N/A	N/A	N/A	Marginal R^2=0.004	N/A	N/A
Tump et al 2022 [35]	—	0.47	—	—	—	N/A	Yes	Yes
Mishra et al 2020 [36]	—	0.7	0.57	0.91	—	N/A	No	No
Schmidt et al 2019 [37]	—	0.47	—	—	—	N/A	Yes	Yes
Jiang et al 2023 [54]	N/A	N/A	N/A	N/A	N/A	MSE[c]~1.00	Yes	Yes
Hadjiantonis et al 2020 [40]	N/A	N/A	N/A	N/A	N/A	r=0.24 (P<.01)	Yes	No
Gaballah et al 2021 [21]	0.66	0.64	—	—	—	N/A	No	No
Yu and Sano 2022 [48]	—	0.70	—	—	—	N/A	Yes	Yes
Burghardt et al 2021 [52]	—	0.23	0.16	—	AUC[d]=0.56	N/A	Yes	No
Kao et al 2020 [51]	N/A	N/A	N/A	N/A	N/A	RMSE[e]=0.89; r=0.42; R^2=0.15	Yes	No
Paromita et al 2023 [43]	N/A	N/A	N/A	N/A	N/A	One-tailed paired t test (P<.001)	No	Yes
Pimentel et al 2021 [55]	—	0.68	—	—	BACC[f]=0.65; MCC[g]=0.30	N/A	No	Yes
Yang et al 2022 [46]	0.58	0.72	—	—	—	N/A	No	No
Zanna et al 2022 [56]	0.43 - 0.54	0.30 - 0.42	—	—	—	N/A	No	No
Tiwari and Falk 2021 [42]		0.69	—	—	BACC=0.66; MCC=0.31	N/A	No	Yes

XSL•FO
RenderX

Paper citation key	Classification model					Regression model	Baseline model reported	Uncertainty measure reported
	Accuracy	F_1-score	Precision	Recall	Other measures	Quality measure		
Feng and Narayanan 2022 [49]	—	66	—	—	—	N/A	No	No
Feng et al 2021 [53]	N/A	N/A	N/A	N/A	N/A	Adjusted R^2: (PA[h] 0.152; NA[i] 0.034)	N/A	N/A
Ravuri et al 2020 [41]	N/A	N/A	N/A	N/A	N/A	ρ=0.08	No	Yes
Stojchevska et al 2022 [44]	0.42	41	—	—	—	N/A	No	Yes
Booth et al 2022 [38]	0.62	0.75	0.65	0.89	—	ρ=0.25	Yes	Yes
Martinez et al 2022 [39]	N/A	N/A	N/A	N/A	N/A	Marginal R^2: 0.022 (main); 0.032 (follow-up)	N/A	N/A
Robles-Granda et al 2021 [22]	N/A	N/A	N/A	N/A	N/A	SMAPE[j]=0.66	Yes	Yes
Saylam and İncel 2023 [50]	N/A	N/A	N/A	N/A	N/A	MAE[k]=0.47	No	No
Li et al 2024 [57]	0.74	72	0.73	0.71	—	N/A	No	No
Saylam and Durmaz İncel 2022 [58]	0.85	0.85	0.85	0.85	—	N/A	No	No
Kim et al 2024 [59]	N/A	N/A	N/A	N/A	N/A	MAE=0.32	No	Yes
Paraschou et al 2023 [47]	0.92	—	—	—	—	N/A	No	No

[a]Not available/not reported.

[b]N/A: not applicable.

[c]MSE: mean-squared error.

[d]AUC: area under the receiver operating characteristic curve.

[e]RMSE: root mean square error.

[f]BACC: balanced accuracy.

[g]MCC: Matthews correlation coefficient.

[h]PA: positive affect.

[i]NA: negative affect.

[j]SMAPE: symmetric mean absolute percentage error.

[k]MAE: mean absolute error.

Ground Truth Stress Labels

The goal of ML models is to predict or estimate psychological constructs, such as stress [43]. These constructs are often referred to as ground truth labels or targets within an ML framework. Since data are collected in naturalistic settings, aligning stress labels with physiological data from wearables poses a challenge, as there are no obvious ground truth labels, unlike in controlled environments where the stressor and stress-inducing periods are known with high levels of certainty. In naturalistic settings, stress labels are typically collected from participants through self-report mechanisms, such as ecological momentary assessment (EMA) [45]. EMA involves short surveys delivered through phone-based apps, a method used by most of the studies. In this section, we outline several key decisions involved in collecting ground truth labels using EMA.

XSL•FO
RenderX

EMA Prompts Frequency and Delivery Mechanisms

EMA was most frequently administered daily [48,51,54], capturing a range of psychological constructs. In fewer instances, more frequent assessments were conducted (eg, every 30 minutes or every 2 hours throughout the day) [30-32]. In all studies, EMA prompts were triggered by predefined schedules, with most being delivered via mobile apps, often through custom-developed apps [21,48,51]. A smaller number of studies used SMS messages to deliver EMA surveys [38,39,50].

Response Window

Since the data were collected in real-world contexts, participants were not expected to respond promptly to EMA probes as they may have been busy with other tasks. To control the delay in response and ensure the accurate timing of the label, studies may put an upper bound on the delay, that is, the probe will expire after a certain period of time [32] to ensure the relevance of labels.

Types of Stress Labels and Instruments Used

The most commonly collected psychology constructs in the studies were stress, which occurred in 26 [19-21,30-39,41-48,50,51,55,57-59] out of 34 (76%) studies, anxiety in 9 [22,39,43,50,51,53,55,56,59] out of 34 (26%) studies, positive or negative affect in 9 [22,39,41,43,49-51,53,54] out of 34 (26%) studies. Various instruments were used, including the Short State-Trait Anxiety Inventory (S-STAI) [47,59], which assessed both stress and anxiety, and the Perceived Stress Scale (PSS) [21,48]. Anxiety was also measured using the State-Trait Anxiety Inventory (STAI) [42,56]. The Positive and Negative Affect Schedule (PANAS) short scale [49,53] was used to assess both positive and negative effects.

Stress Measurement Scale

Additionally, different scales were used, such as the Likert scale [31,36,37], Numeric Rating Scale [33], and Visual Analog Scale [35], which is a gliding scale ranging from “not at all” to “extremely.” In a few studies, discrete ordered scales were used to categorize stress levels as low, medium, or high [30,58].

Preprocessing of Stress Labels

There were various preprocessing decisions researchers have made with regard to the stress labels in the reviewed studies. In most studies, the output targets in the ML model maintained the same granularity as the ground truth stress label. However, some studies applied downsampling techniques. For instance, when ground truth labels were collected every 30 minutes, they were aggregated into daily values [20,31] in modeling.

In classification studies, these outputs collected on a numeric scale were often categorized into bins or classes, such as low, medium, or high levels of stress or anxiety [19]. Most studies used binarized output targets for classification, while only a few applied multiclass classification. Due to the ordinal nature of the data, multiple categories were combined into fewer groups. For example, on a 5-point scale, ratings of 0 - 1 were classified as stress, while ratings of 2 - 4 were classified as no stress [19,32]. Another example involves merging the last 3 categories of a 5-point scale into a single “high stress” group, with the first 2 categories remaining as “low” and “medium stress” [44]. For nondiscrete values, such as on a 100-point scale, values below 50 were marked as low stress, and values above 50 were marked as high stress [35].

Two primary methods were used to define classification label thresholds: a personalized approach and a global approach. The personalized approach involved setting thresholds specific to individual participants, whereas the global approach applied the same thresholds across all participants. For personalized thresholds, various statistical methods were used, including majority voting, median calculation, and z-score analysis. The z-score method was the most commonly used [21,46,48,56], where z-scores were calculated for each individual to account for subjective variability. Data were then categorized into 2 classes; for example, negative classes included those with z-scores below zero, and vice versa. In the global approach, values were discretized using predefined scales or thresholds derived from the entire participant sample. For instance, some studies applied fixed global thresholds to classify stress and anxiety levels across all participants [42].

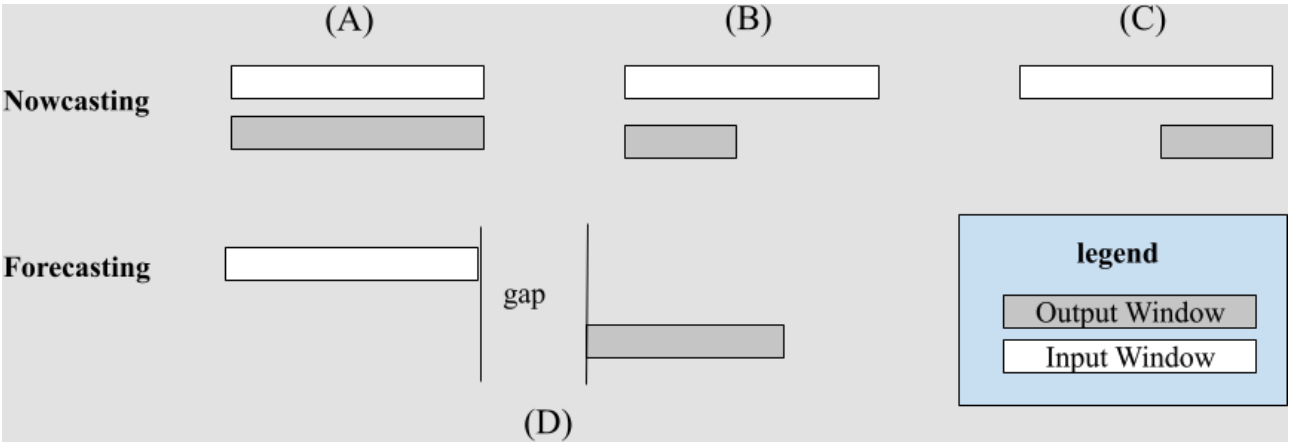
Input and Output Windows

As part of the modeling framework, we extracted the input and output windows reported in the reviewed studies. The input window refers to the time frame from which input features are extracted, typically derived from physiological and behavioral signals. On the other hand, the output window defines the period during which the model predicts target outcomes corresponding to the ground truth labels, as discussed in the “Ground Truth Stress Labels” section.

We use the terms nowcasting and forecasting to describe the relationship between input and output windows. Nowcasting (Figure 2A-C) refers to models that predict outcomes, such as stress, for the present or immediate future, resulting in a very short output window with no gap between input and output windows. In contrast, forecasting (Figure 2D) predicts stress over future time periods, with short-term forecasting focusing on the immediate next period and long-term forecasting predicting stress at a time point extended further into the future.

XSL•FO
RenderX

Figure 2. The relationship between input and output windows; nowcasting versus forecasting.



Several studies have used an exact overlap between input and output windows as shown in Figure 2A, often with a window size of 24 hours [31,51]. A smaller number of studies have used shorter input-output windows, such as every 20 minutes [32], hourly [44], or every 3 hours [35]. This exact overlap is a defining characteristic of nowcasting studies, as illustrated in Figure 1.

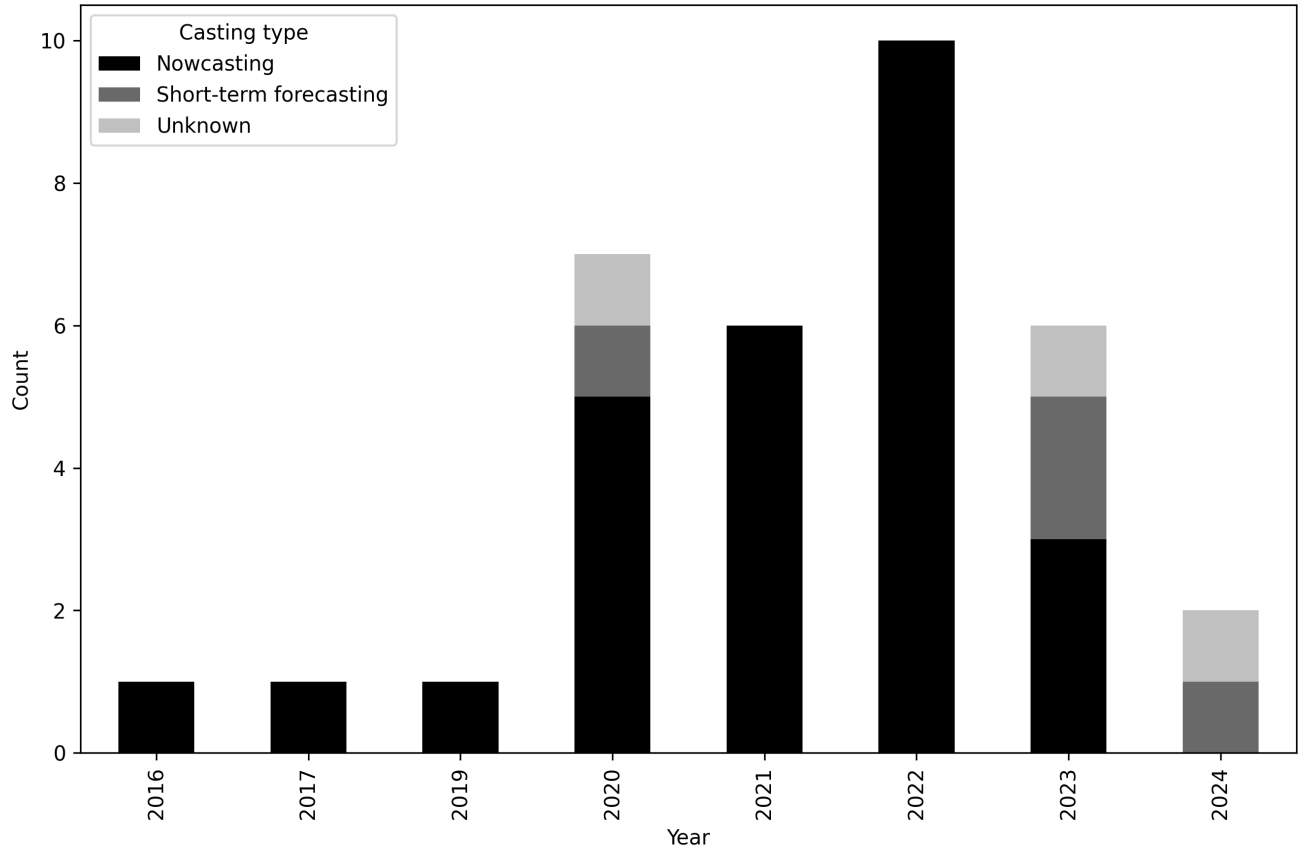
Additionally, many studies included partial overlap, where the input window is either longer or shorter than the output window, as shown in Figure 2B and C. For example, an input window might include data from the previous day, the current day, and the next day, while the output window focuses only on the current day [51,52] or both the previous and current days [54]. In other examples, input window is shorter than the output window; for example, the input window covers 1 minute and the output window spans 15 minutes [30], or where the input window is 2 hours paired with a 1-day output window [56], or an input window that selects an optimal range between 10 and 27.5 minutes while the output window remains fixed at 20 minutes [19].

In the forecasting studies identified, the input window is chosen to be immediately adjacent and trail the output window (Figure 2D). For example, the input window may consist of data from the previous day, with the output window set for the current day [40], in which case the gap is zero. In other instances, the input window might consist of data from the last 35, 30, or 25 days, while the output window could be set for 19, 17, or 12 days [57] into the future, starting from the prediction time, in which case the gap is larger than zero.

It has been observed from the selected studies that the majority of researchers have used a nowcasting relationship between the input and output windows, as illustrated in Figure 3.

XSL•FO
RenderX

Figure 3. Number of studies formulated as nowcasting versus forecasting problem, grouped by year.



Addressing Issues of Data Leakage

The reviewed studies differed in how their training, test, and validation sets were split. Since the wearable dataset is longitudinally collected, it is typical to see multiple input-output pairs extracted from the same participants. Data leakage can occur where data points in the training set and test set may originate from the same participant. This made participant-level splitting relevant to the assessment of potential data leakage and possible inflation of reported model performance.

In our review, we observed that 13 [30,32,35,36,41,42,45,47,54-56,58,59] out of 28 applicable (46%) studies [19-21,30-32,35-38,40-48,50-52,54-59] did not explicitly address potential data leakage issues as inferred from the description of the split of training, test, or validation set, while the rest addressed data leakage issues by using strategies such as leave-one-subject-out cross-validation.

ML Models

This section provides insights into the types of ML models used. The studies included in this review primarily focused on supervised learning paradigms, addressing both classification and regression tasks. A diverse range of models was used, spanning traditional ML frameworks to end-to-end deep learning architectures. For feature extraction, most studies relied on statistical summaries [30,31,36], while a smaller subset applied Fourier transformations, particularly for analyzing heart rate variability signals [19,20,42].

Traditional ML algorithms, valued for their simplicity and interpretability, such as support vector machines (SVMs), Naive Bayes, and k-nearest neighbors (KNN), were frequently used. Additionally, various ensemble methods, including bagging and boosting techniques, were widely applied [19,44,45,47,51]. Among these, random forest (RF) was the most commonly used in studies using traditional ML algorithms [19,36,38,44,45,49,50,52,54,58,59].

Table 4 summarizes the most frequently used traditional ML models across the studies.

XSL•FO
RenderX

Table . Counts of papers using specific traditional machine learning models.

Model	Count of papers using the model	Reference numbers
Random forest	11	Parousidou et al 2023 [45], Gjoreski et al 2017 [19], Mishra et al 2020 [36], Jiang et al 2023 [54], Burghardt et al 2021 [52], Feng and Narayanan 2022 [49], Booth et al 2022 [38], Robles-Granda et al 2021 [22], Saylam and İncel 2023 [50], Saylam and Durmaz İncel 2022 [58], Kim et al 2024 [59]
SVM[a]	9	Boateng and Kotz 2016 [30], Bavaresco et al 2020 [32], Gjoreski et al 2017 [19], Han et al 2020 [20], Mishra et al 2020 [36], Burghardt et al 2021 [52], Pimentel et al 2021 [55], Tiwari and Falk 2021 [42], Kim et al 2024 [59]
Boosting methods	7	Gjoreski et al 2017 [19], Stojchevska et al 2022 [44], Parousidou et al 2023 [45], Paraschou et al 2023 [47], Saylam and İncel 2023 [50], Kao et al 2020 [51], Jiang et al 2023 [54]
Naive Bayes	4	Parousidou et al 2023 [45], Bavaresco et al 2020 [32], Gjoreski et al 2017 [19], Han et al 2020 [20]
KNN[b]	4	Bavaresco et al 2020 [32], Gjoreski et al 2017 [19], Han et al 2020 [20], Jiang et al 2023 [54]
Logistic regression	4	Parousidou et al 2023 [45], Tump et al 2022 [35], Jiang et al 2023 [54], Burghardt et al 2021 [52]
Decision tree	3	Parousidou et al 2023 [45], Gjoreski et al 2017 [19], Robles-Granda et al 2021 [22]

[a]SVM: support vector machine.

[b]KNN: k-nearest neighbors.

In addition to traditional ML, the use of deep learning models has seen a significant rise. Models explored include recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, convolutional neural networks (CNNs), and LLMs. The choice of models is closely tied to the specific objectives and implications of each study, as discussed below.

Challenges and Solutions

The reviewed studies aimed to address specific challenges, each accompanied by proposed solutions. These challenges were either intrinsic to the data or inherent to the domain. Some common challenges and their corresponding solutions are outlined below

Personalization

Personalization was one of the challenges addressed in the reviewed studies, because of intrinsic between-person variability, as individuals may respond differently to the same stressors. In many of the selected studies, personalization was not considered, thus following a one-size-fits-all approach in modeling. However, a notable number of studies addressed personalization as part of the preprocessing phase. Two key personalization approaches were identified: user-based and group-based. While the user-based approach involves creating a unique model for each participant, the group-based approach groups participants based on selected attributes, such as demographics or gender.

Various clustering algorithms were used in the studies to create cohorts or groups, including k-means, spectral clustering, agglomerative clustering, mean shift, affinity propagation, and Gaussian mixture models (GMM) [51]. Some other studies also used additional techniques to enhance cluster quality, such as iterative clustering [41], which progressively refined participant groupings by repeating the clustering process multiple times. Another method listed in the reviewed studies was dimensionality reduction prior to clustering, which was achieved through principal component analysis (PCA) [43]. PCA simplified the feature space while preserving variance by projecting the data into a lower-dimensional space, whereas self-organizing maps (SOMs) [31] facilitated the organization of complex patterns within the data.

The clusters created during the preprocessing step were further trained using traditional ML techniques (eg, RF, decision trees, and linear regression) [45] as well as deep learning approaches, including metric learning with Siamese neural networks (SNNs) [43]. SNNs transform input data to ensure that similar data points are projected close together, while dissimilar points are spaced apart in the learned feature space. Additionally, feedforward neural networks (FFNs) [41] were used to fine-tune the generalized model by adjusting its weights through iterative training on group-specific data.

Clustering offers several benefits. For example, it may induce more cohesive groups as grouping users with similar characteristics helps the model detect common stress patterns within each group [45]. In another study, the user-based

XSL•FO
RenderX

collaborative filtering method is used by leveraging health information from other users. This type of model can fill gaps within groups and reduce the reliance on historical data from individual users [51].

Temporal Dependency

Data collected from wearable sensors inherently possess a time-dependent nature due to its intrinsic properties. Consequently, studies have explored methods to explicitly address these temporal dependencies to improve model accuracy, with the hope that those methods may reveal complex relationships that conventional ML models may overlook. Various models specifically designed to capture these temporal dependencies are listed below, along with a brief summary of their usage in the selected studies.

1. LSTM networks, designed to capture long-term dependencies in sequential data, were used as bidirectional LSTM (BI-LSTM) in one of the studies [21]. In that study, BI-LSTM captured both past and future values to predict current stress levels in a nowcasting setup. An hourly vector was processed in both the forward and backward directions to analyze data from start to finish and vice versa. The outputs from both directions were combined to generate the final predictions, offering a comprehensive understanding of the hourly temporal context. In another study, a lagged version of the data was explored to highlight trends over time, associating each data point with its historical values, allowing the model to learn from past trends [38].
2. Vector autoregression (VAR) [33] was used to predict current values based on past data through lag selection. Two approaches were taken: predicting sensor outcomes (eg, total sleep time and heart rate variability) using EMA variables such as stress, and predicting self-reported EMA variables based on sensor data. An impulse response function (IRF) was also applied to visualize how changes in predictor variables affect outcomes over time.
3. A hidden Markov model was used to capture how different physiological conditions evolve over time through hidden states. This study [52] aimed to detect both typical and atypical events by modeling physiological data, helping identify patterns that represent normal daily behaviors as well as deviations caused by stress and anxiety.
4. Higher-order networks [22] were used to model temporal dependencies using ensemble learning frameworks, integrating multiple ML models to predict physical, psychological, and job performance variables. These models leveraged time-dependent patterns in physiological signals by incorporating lagged data.
5. Conceptual frameworks such as dynamical systems theory [40] and chaos theory [57] have been applied to capture temporal relationships. One study used a dynamical systems model with linear regression to identify daily patterns of emotional self-regulation and stress spillover in health care professionals, effectively capturing recurring cycles. In another study that used chaos theory, phase space reconstruction was applied to transform low-dimensional data into high-dimensional data by combining stress levels from multiple days as output variables. This approach enabled the model to account for previous days' stress levels, ensuring that similar stress levels appeared closer in the reconstructed space. Additionally, integrating chaos theory with LSTM models that incorporated lagged values improved performance compared to using either chaos theory or LSTM models with lagged values independently.

Multiple Modalities

The feasibility of using all training modalities is often limited in real-world scenarios due to the restricted availability of sensors in wearables. To address this, one study focused on reducing the number of modalities during testing while incorporating more during training [46]. This was achieved using a technique called knowledge distillation, implemented as the "More to Less" framework. Knowledge distillation involves transferring knowledge from a stronger network to a weaker one by using classifier networks from different modalities. During training, the model used all available modalities, with each having its own classifier network. These classifiers exchanged knowledge through adaptive mechanisms defined in the "More to Less" framework, allowing each classifier to learn not only from its own modality but also from those performing better. As a result, testing was conducted with fewer modalities, as the network had already developed robust representations during training. This approach enhanced generalization by leveraging information from multiple modalities, even when fewer input modalities were available during testing.

Bias Mitigation

Bias in relation to demographic groupings was addressed in one reviewed study [56]. In that study, an LSTM model was used by incorporating multitask learning, which included both the protected attribute and the target variable. This approach aimed to mitigate bias related to the protected attribute. To achieve this, different loss weights were assigned. For example, if the target variable (eg, anxiety) is given a weight of 4.5 and the protected attribute (eg, gender) a weight of 0.5, the problem is treated as a multitask learning scenario. Each task has a separate loss function, ensuring that the nonprotected label, such as anxiety in this case, is prioritized.

Context Awareness

Contextual information is widely recognized as crucial in studies aimed at identifying the causes or triggers of stress, as learning context helps improve model accuracy. Various types of contextual data considered in these studies are outlined below:

1. One study compared 2 methods of context collection: ML-derived data, such as sleep and activity, versus self-reported data collected through EMA [44]. The results showed that ML-derived context significantly improved model performance over self-reported data.
2. Another study focused on key stressors among employees working from home, highlighting the role of environmental factors as contextual elements affecting stress and well-being [35]. This approach highlighted the importance of both personal and environmental context in remote work settings.

XSL•FO
RenderX

3. A separate set of studies integrated circadian cycles as a contextual factor for predicting wellness indicators—such as positive affect, negative affect, and life satisfaction—by analyzing data such as audio features, heart rate, sleep patterns, activity levels, and location [49,53].
4. Various models have been used in these studies, including ensemble methods such as gradient-boosted trees (eg, CatBoost), RF, logistic regression (LR), and linear regression.

Addressing Limited Supply of Training Data

Given the complexity of data collection in the naturalistic setting, there is a limited supply of high-quality datasets, along with challenges in capturing ground truth. To address those challenges, a study [54] explored zero-shot learning within a meta-learning framework. This framework incorporated various base learners, including RF, AdaBoost (AB), gradient boosting (GB), KNN, and LR. The approach enabled the model to learn from diverse tasks, allowing it to quickly adapt to new, unseen tasks with minimal additional training. By training on different tasks and populations, the model developed effective initialization weights, improving its ability to make accurate predictions on unseen data.

Use of LLM in Stress Detection

Recently, with advancements in AI, emerging technologies such as LLMs have been developed. These models are pretrained on vast amounts of text and can be applied to a wide range of tasks. One study in the review used LLMs along with wearable sensor data through structured prompts that incorporated user context, health knowledge, and temporal information [59]. The study focused on three approaches:

1. Zero-shot prompting: in this approach, the model received prompts without any prior examples. Models pretrained on task-specific data, such as medical or sensor data, performed better; for instance, Asclepius outperformed larger models.
2. Few-shot prompting: a small number of examples were included in the prompt to provide context for the model. This resulted in improved performance in larger models such as GPT-3.5 (OpenAI) and GPT-4 (OpenAI).
3. Fine-tuning: this method involves adjusting some or all parameters of a pretrained model using a dataset specific to the target task. HealAlpaca achieved the highest accuracy in this category.

Overall, few-shot prompting outperformed zero-shot prompting, while fine-tuning proved to be the most effective approach.

Analysis of Model Performance

In Table 3, we presented a detailed analysis of model performance and extracted key evaluation metrics, such as F_1-score, accuracy, precision, recall, and AUC for classification tasks, and MSE for regression tasks. We reported the best-performing model in each study. Where applicable, we also documented the corresponding model configurations and comparison baselines. It should be noted that data leakage considerations are only relevant for models designed to predict momentary stress labels. Six studies [22,33,34,39,49,53] fall outside this scope, including (1) statistical models (often linear or linear mixed-effects models) that fit the entire dataset to examine coefficient significance or overall goodness-of-fit; and (2) models predicting stable psychological constructs derived from one-time baseline surveys, where the unit of analysis is the individual rather than the event or time point. We identified a total of 13 studies [30,32,35,36,41,42,45,47,54-56,58,59] that had potential data leakage issues out of 28 applicable studies [19-21,30-32,35-38,40-48,50-52,54-59].

There are a few caveats in interpreting performance measures:

1. Substantial heterogeneity exists across datasets and problem formulations (eg, classification vs regression, input and output window configurations, nowcasting vs forecasting, and different discretization strategies for stress labels), making direct, apples-to-apples comparisons challenging.
2. Performance reporting is not standardized across studies. Although F_1-score is the most commonly reported metric for classification, several studies do not report it; in the case of class imbalance (which is often not reported), those metrics may not adequately reflect true model performance.
3. A notable proportion of studies (13 [30,32,35,36,41,42,45,47,54-56,58,59] out of 28 [46%] studies [19-21,30-32,35-38,40-48,50-52,54-59] among those applicable studies) has potential data leakage, most often due to inappropriate data-splitting strategies, which may inflate reported results. In fact, we note a few studies report unusually high performance (eg, with accuracy>85%), many of them with potential data leakage issues.
4. We note that a substantial proportion of studies (approximately 50%) do not report uncertainty quantification for performance metrics, such as CIs, SEs, SDs, or *P* values, making it difficult to determine whether observed performance differences are statistically meaningful or attributable to random variation.

There are several major findings:

1. Across studies, deep learning models achieved F_1-scores ranging from 0.44 to 0.72, whereas traditional ML models exhibited a wider performance range (F_1-score=0.47 - 0.90), making it difficult to draw a general conclusion about the absolute superiority of deep learning approaches. A rigorous comparison would require controlled, head-to-head experiments that hold other factors, such as data modality, problem formulation, or data split, constant. As 2 of those examples, Parousidou et al [45] show that traditional ML combined with personalization can outperform deep learning, while Pimentel et al [55] demonstrate that careful feature engineering enables traditional models to achieve superior performance relative to deep learning methods.
2. We note that performance also varied by modeling strategy: personalization-based approaches reported F_1-scores between 0.62 and 0.66; methods addressing limited data availability achieved F_1-scores around 0.70; context-aware models showed a wide range with F_1-score from 0.41 to 0.90; multimodal integration approaches reported an F_1-score of 0.72; and models accounting for temporal dependencies exhibited wide variability

XSL•FO
RenderX

(F_1-score=0.23 - 0.71). Finally, studies using LLMs reported accuracy values as high as 92%, although corresponding F_1-scores were not provided.

Ethical Considerations

Stress detection applications and associated wearable devices often involve the collection of fine-grained longitudinal data, as well as psychological profile information, either in raw form or inferred through computational models. Such data fall within the realm of sensitive private information and, if inadvertently disclosed, may pose significant risks to individuals. For example, employers could potentially misuse stress or negative affect profiles in human resource–related decision-making processes. In addition, weak data security policies may increase the likelihood of security breaches, which could enable the reidentification of individuals and lead to the leakage of personally identifiable information. Accordingly, research studies on stress detection, as well as the downstream applications built upon these systems, must adopt deliberate and robust measures to safeguard the entire data collection, storage, and processing pipeline. These measures are essential for providing a high level of assurance and protection to individuals, whether they participate as research subjects or engage with such systems as end users of downstream applications. Some emerging approaches may also be explored, such as federated learning, as suggested in one of the reviewed studies [59]. In this approach, raw data remain on the local device where they are recorded, thereby reducing the privacy risks associated with centralized cloud-based data storage and processing.

Equally important is the need for transparent and meaningful informed consent procedures. Data owners should be clearly informed about what types of data are being collected, how the data will be processed and analyzed, who will have access to the data, how long the data will be retained, and whether the data may be reused for secondary research purposes or commercial applications. Such transparency is critical for establishing trust and supporting individuals' autonomy in making informed decisions regarding participation. In addition, participants should be made aware of potential risks, including the possibility of reidentification or unintended disclosure of sensitive psychological or behavioral information. As noted, among the 14 datasets used across the reviewed studies, some based on primary data collection and others relying on secondary datasets, only 11 explicitly acknowledged obtaining proper institutional review board (IRB) approval, whereas the remaining 3 studies did not report such approval [19,20,32]. This observation highlights the continuing need for stronger ethical oversight and clearer reporting practices in stress detection research involving wearable and behavioral data.

Model Card for Wearable-Based Stress Modeling

In this section, we propose a model card framework tailored for wearable-based stress modeling, adapted from the general model card template introduced by prior work [23] in the ML community. The specific focus is on models that enable the momentary estimation of current stress or to forecast stress in the future. The original template includes sections such as model details, intended use, factors, metrics, evaluation data, training data, quantitative analysis, ethical considerations, and caveats and recommendations. Given the unique characteristics of wearable-based stress modeling, particularly the reliance on longitudinal, in-the-wild biosignal data collection and self-reported ground truth, we reorganize and customize the model card into two core components: (1) dataset and (2) modeling decisions. This structure foregrounds the data provenance and methodological choices that most strongly influence model validity, transparency, and reproducibility. A detailed model card for wearables, including its components, dimensions, and reporting requirements, is presented in Table 5 below.

XSL•FO
RenderX

Table . Proposed model card reporting template for wearable-based stress detection models.

Component and dimension	Reporting requirements
Modeling decisions	
Model metadata	• Model creators and affiliated organization • Model version, date, citation, and license • Point of contacts and link to code repository if applicable
Dataset	
Participant characteristics and context	• Participant demographics (eg, age, gender, and occupation) and their distributions • Sample size and duration and frequency of data collection • Geographic locations of participants • Naturalistic context: description of participants' typical daily life (eg, work schedules and routines) and common stressors relevant to the population under study
Wearable devices and signals	• Wearable devices used and placement • Biosignals captured (eg, heart rate, HRV[a], EDA[b], and accelerometry) and sampling frequency
Ground truth annotation	• Target variables (eg, perceived stress) and how they are captured • EMA[c] protocol details, including prompt frequency and scheduling strategy • Response window and compliance requirements • Survey instrument used • Descriptive statistics of target variables, including distributions across demographic subgroups, where applicable
Metadata	• Dataset versioning, documentation, point of contact, citations, licenses, and so on and any relevant contextual metadata (eg, time zones, device firmware, and protocol deviations)
Modeling decisions	
Data preparation	• Preprocessing steps, including handling of missing data, signal cleaning, normalization, multimodal data synchronization, and identification of subgroups
Problem formulation	• Task definition (eg, classification vs regression) • Temporal framing • Task specification • Nowcasting versus forecasting input and output window configurations
Experimental design	• Modeling setup and validation strategy • Data splitting procedures (training, validation, and test) • Assessment of potential data leakage and use mitigation strategies such as subject-level data-split
Modeling approach	• Provide detailed configuration of the model, including those do not perform well • Document hyperparameter tuning strategy and process • Report a variety of performance metrics, including uncertainty measures such as CIs, SE, or *P* value from statistical significance test • Report baseline model performance (including default model) for comparison. Decision thresholds (for classification models), if applicable
Dataset	
Ethical considerations	• Ethical approval (eg, IRB[d] and participant consent procedures) • Recruitment and sampling strategy, and discuss potential of sampling bias • Privacy protection strategy, eg, anonymization

XSL•FO
RenderX

Component and dimension	Reporting requirements
Modeling decisions	
Caveats and recommendations	• Justification of key modeling choices and the specific challenges they are intended to address • Discussion on factors influencing performance, including reporting of disaggregated performance across relevant subgroups where applicable • Discussion on sampling bias and generalization limitations, especially when transferring models across populations, contexts, or devices • Other known limitations, appropriate use cases, and guidance for interpretation

[a]HRV: heart rate variability.

[b]EDA: electrodermal activity.

[c]EMA: ecological momentary assessment.

[d]IRB: institutional review board.

Discussion

Principal Findings

Stress detection has become increasingly important due to the rising prevalence of stress in individuals' daily lives. Advances in wearable technologies have opened new opportunities for detecting stress in real-world settings. This scoping review aims to examine the existing literature on the use of ML applied to data collected from wearables in naturalistic environments for stress detection. In this section, we summarize the principal findings, the identified research gaps, and potential avenues for further investigation, while acknowledging the limitations of this review.

This review identified a lack of standardization in model reporting, which is one of the biggest barriers to advancing the field beyond existing research, and the generalizability of stress detection in naturalistic settings. This is due to substantial heterogeneity in how problem specifications are defined, with no standard format for model reporting, which limits benchmarking across current studies. The reviewed studies differed considerably in the wearables used, ground-truth labeling strategies, input-output window definitions, and ML approaches. We also identified methodological gaps, including potential data leakage, inconsistent reporting of performance metrics, limited handling of class imbalance, and relatively little explicit attention to motion artifacts. In addition, we noted that the reviewed studies addressed a range of challenges, though not uniformly, including personalization, temporal modeling, bias mitigation, modality-related concerns, challenges specific to in-the-wild settings, and the use of LLMs. These findings also motivated the proposed model card specification to improve comparability, transparency, and reproducibility across studies in the future.

Methodological Rigor

In our analysis of model frameworks, characteristics, and performance summarized in Tables 2 and 3, we observed that 13 [30,32,35,36,41,42,45,47,54-56,58,59] out of 28 (46%) applicable studies [19-21,30-32,35-38,40-48,50-52,54-59], specifically, those that developed ML models to predict momentary stress had potential data leakage issues. These issues primarily arise from ignoring the grouped structure of the data, where multiple observations or episodes are contributed by the same participant. When training, validation, and test splits are performed at the observation level rather than the participant level, data points from the same individual may appear in both the training and evaluation sets. This overlap can result in unintended information leakage from training to test data, thereby artificially inflating model performance. A commonly recommended practice to mitigate this issue is to perform data splitting at the subject level, ensuring that all observations from a given participant are assigned to a single split. Given that a substantial proportion of the reviewed studies exhibit this potential methodological flaw, the performance metrics reported in Table 3 should be interpreted with caution, as they may overestimate true generalization performance.

Another issue identified in our analysis of model performance is that a large proportion of studies do not explicitly report baseline model performance. This omission is particularly problematic in the presence of class imbalance, where commonly reported metrics such as accuracy may be misleading. For example, a naïve classifier that always predicts the majority class may achieve deceptively high accuracy, despite providing little meaningful predictive value. Moreover, performance metrics that are more robust to class imbalance, such as the AUC, are rarely reported. While some studies attempt to address this issue by using alternative measures, including balanced accuracy (BACC) or Matthews correlation coefficient (MCC), such practices remain the exception rather than the norm. In addition, we observe that a substantial proportion of studies do not report uncertainty quantification for performance metrics, such as CIs or SDs. The absence of these measures complicates performance comparison across models or approaches, as observed differences may simply reflect random variation rather than meaningful performance gains.

To address these methodological gaps, we explicitly incorporate these considerations into our proposed model card specification, with the goal of encouraging future research to adopt more rigorous experimental designs, performance measures, and reporting practices.

XSL•FO
RenderX

Standardization and Benchmarking

The complexity and richness of data collected from wearables and other devices in real-world settings create ample opportunities to explore a wide range of ML frameworks. While this flexibility allows researchers to pose a variety of research questions and explore various kinds of models, it also poses significant challenges when comparing the utility and performance of different models in tasks such as stress detection and estimation. As demonstrated in our analysis, the lack of standardization in reporting often makes it difficult to extract key specific details about ML modeling decisions, thus limiting the ability to compare model effectiveness across studies. This lack of standard and benchmarking is likely to slow progress in developing robust ML models for stress detection.

Future research could focus on the establishment of benchmark datasets and standardized tasks to enable meaningful comparisons. Additionally, the research community would benefit from adopting consistent reporting standards that ensure critical components of ML frameworks are clearly documented. For example, approaches such as model cards [23] promote transparency and reproducibility. There should also be a stronger emphasis on model quality control, including the adoption of best practices in experimental design to mitigate issues such as data leakage, which may contribute to the inflation of reported model performance. In this paper, we proposed a model card solution specifically tailored to modeling stress in the wild using wearable data, based on our analysis and understanding of the critical issues and related best practices related to modeling decisions, performance analysis, and reporting.

Longitudinal View of Stress: From Nowcasting to Forecasting

A distinguishing feature of real-life stress studies is the ability to examine stress over extended periods and explore its interactions with contextual factors. Despite these opportunities, we observed that most studies adopt a short-term approach to stress detection, where the input and output windows are closely aligned in time. Only a few recent studies have begun to explore forecasting methods that account for long-term temporal trends in psychological and physiological states.

Among the 4 forecasting-oriented papers, 2 adopt a conventional ML pipeline focused on next-day stress prediction. Paraschou et al [47] report an accuracy of 92.3% from a GB ML model, although the evaluation may be subject to data leakage concerns. Another study [50] systematically explores a wide range of input and output window configurations and finds that LSTM-based models generally achieve the strongest performance. In particular, the highest predictive accuracy is obtained when using data from the preceding 15 days (ie, input window=15 days) to predict next-day stress (ie, output window=1 day), although XGBoost (eXtreme Gradient Boosting) models show competitive performance under several configurations. In contrast, the remaining 2 papers adopt dynamic systems modeling approaches [40,57]. Hadjiantonis et al [40] fit a linear dynamical system model that characterizes next-day changes in emotional arousal as a linear function of the current day's perceived stress and emotional arousal, with model parameters estimated via linear regression. Li et al [57], however, model day-to-day perceived stress as a nonlinear dynamical system. The study follows a 2-step approach: first, chaos theory is used to identify the most reliable prediction horizon, which shows that an input window of 19 days maximizes predictability for forecasting stress over the subsequent 35 days. Using phase-space embedding and deep neural networks, the proposed method achieves accuracies of 74.4% for binary stress classification and 69.23% for 3-level stress classification under a person-wise 80/20 split.

We encourage future research to broaden the scope of stress modeling beyond momentary detection, incorporating longitudinal perspectives that capture the ebb and flow of stress over time. Such studies could also investigate the interplay between stress and mediating factors such as sleep, as well as coping strategies such as meditation, to provide a more comprehensive understanding of stress dynamics in daily life. Our review suggests that methodological innovations adopting multivariate and longitudinal perspectives are still emerging. Traditional ML pipelines provide a reasonable starting point by framing stress forecasting as a mapping between multimodal signal streams within predefined input windows and corresponding output windows. However, much of the existing work relies on black-box models that offer limited insight into the underlying mechanisms of stress-related dynamics. A complementary line of research draws on dynamical systems–inspired approaches, including both linear and nonlinear models. Within these frameworks, prediction targets often extend beyond a single future time point to a prediction horizon spanning multiple days or weeks. This shift enables more meaningful and theoretically grounded questions to be addressed, such as the evolution trajectory and long-term trends of stress dynamics over time. Chaos theory–based approaches, in particular, provide a useful lens for understanding the limits of predictability in forecasting future stress trajectories. Li et al [57] represent a promising first step in this direction, and future work could further advance this line of inquiry by incorporating richer, multivariate biosignals from wearable devices in both the input and output spaces.

For both ML-based and dynamical systems–based models, a key challenge in fitting highly complex models lies in data sparsity and limited transparency, as these approaches often require large amounts of high-quality longitudinal data while offering limited interpretability. One promising direction for addressing these challenges is the systematic integration of domain knowledge, such as insights from psychological theory, through principled frameworks such as neurosymbolic systems [60,61], which combine data-driven learning with theory-informed constraints and representations. Advancing such approaches will likely require close multidisciplinary collaboration between computing researchers and domain experts in fields such as psychology and health care.

Addressing Modeling Challenges in the Wild

Data collection in real-life settings, where participants are moving, poses significant challenges to the validity of physiological signals due to motion artifacts. Additionally, it remains unclear to what extent models trained on laboratory-controlled datasets generalize to real-world scenarios.

XSL•FO
RenderX

Our analysis found that only a few studies explicitly acknowledge and address these challenges. To address motion artifacts, a study [42] proposed subband HRV features in addition to traditional HRV benchmark features (time-domain and frequency-domain features). The proposed features were obtained by splitting the HRV tachogram into low-frequency (LF) and high-frequency (HF) bands and analyzing each band separately. From these bands, the authors computed nonlinear features, such as transfer entropy between LF and HF components, as well as spectral descriptor features, such as centroid, spread, skewness, kurtosis, crest, and spectral entropy. The study found that its Fuse-All model, which combined the traditional benchmark HRV features with the proposed subband features, gave better performance than the benchmark features alone. Another study [36] took a deliberate approach to compare the models developed in the lab versus those in the wild. They used a commercial off-the-shelf heart rate monitor in both laboratory and real-world settings on the same cohort of participants to ensure comparability. Stress labels collected in the wild were obtained through EMA-based self-reports. The results showed that the best-performing pipeline involved removing extreme values through trimming (ie, outlier removal), followed by data standardization using z-score normalization (trim_zscore). Furthermore, excluding periods of high physical activity further improved performance in real-world settings.

Model Transparency

In addition to evaluating model performance using common metrics such as accuracy, an equally important aspect of model quality is transparency. Transparent models not only allow for better diagnosis and refinement but also create opportunities for interdisciplinary collaboration with experts from fields such as psychology, who bring deep knowledge of stress and its mechanisms. However, many current modeling approaches rely on black-box methods—particularly those based on deep learning and, more recently, generative AI. While these models may improve performance in certain cases, they also make interpreting results more challenging [59]. Future research should explore modeling approaches that achieve a better balance between transparency and accuracy. By integrating domain knowledge from psychology and physiology, such models could not only predict stress but also explain why certain predictions are made—uncovering not just correlational relationships, but potentially causal ones that offer insights for designing effective stress interventions. Additionally, developing models and systems that facilitate clear communication of model decisions to end users could open a new line of inquiry focused on ensuring human-in-the-loop approaches in stress modeling as explored in the study by Paraschou [47].

LLMs in Wearable-Based Stress Detection

With the rise of LLMs, there is increasing interest in using them for mental health prediction tasks such as stress, anxiety, depression, and sleep disorders. One study included in this review demonstrated promising predictive performance using both larger LLMs, such as GPT-3.5 and GPT-4, and a smaller health-focused fine-tuned model, HealthAlpaca. The study suggested that performance depends not only on model size but also on prompting strategy, contextual information, and domain-specific fine-tuning, with smaller fine-tuned models performing comparably to or better than larger general-purpose LLMs. It also showed that the performance of larger models can be improved by providing examples through few-shot prompting or by adding contextual information to zero-shot prompts. However, several limitations remain. Wearable time-series data are high-dimensional, nonlinear, and continuous, making them more challenging for LLMs than conventional language inputs. In addition, the validity and interpretability of LLM-based predictions remain important concerns due to the lack of standardized evaluation benchmarks for nonlinguistic wearable data. These models also have substantial data and computational demands, as well as the risk of false-positive or hallucinated outputs in health-related settings. Future research should therefore examine explainability, validity, reliability, and whether LLM-based approaches provide meaningful advantages over more conventional models in wearable-based stress prediction.

Data Quality and Validity Considerations

Our review highlights several recurring data quality and validity challenges in wearable-based stress studies conducted in naturalistic settings. In particular, field-collected datasets are often shaped by study design constraints, resulting in nonuniform data distributions, limited sample sizes, short observation periods, and skewed participant characteristics (eg, gender, age, and other demographic attributes). These factors collectively constrain population representativeness and limit the external validity and generalizability of reported findings. In contrast, studies that rely on secondary or publicly available large datasets typically benefit from longer monitoring durations and more heterogeneous participant pools. While such datasets may offer improved statistical power and demographic coverage, they also introduce trade-offs, including reduced control over data collection protocols, labeling procedures, and contextual fidelity. Taken together, these differences underscore the importance of explicitly reporting dataset characteristics and carefully considering how data provenance and study design choices shape model validity and interpretability in real-world stress modeling. In our proposed model solution, we explicitly include sections on dataset characteristics to encourage robust study design, as well as transparency and standardization in reporting.

Limitations

Limitations of this scoping review include the challenge of ensuring that all relevant studies have been included and none have been overlooked due to language constraints, as well as the exclusion of unpublished or ongoing studies. Additionally, we focused on a limited range of databases, which restricts the scope of the studies retrieved. The search criteria and query string used may also limit the outcomes of the scoping review. Rather than implementing full parallel double-screening of all records, we adopted a variant of the screening approach in which one researcher conducted the initial screening, followed by a review by a second reviewer. As a result, interrater reliability metrics, such as Cohen κ or percentage agreement, were not calculated. The absence of formal double-screening and

XSL•FO
RenderX

corresponding reliability statistics may limit the reproducibility of the screening process.

Conclusion

Wearable devices offer valuable opportunities to measure and understand how stress manifests in real-world settings. ML and other advanced computational and statistical methods have increasingly been adopted to develop models that detect, characterize, and understand stress and related psychological constructs from physiological signals captured by wearables, as well as from self-reported data collected at increasingly fine temporal granularity using methods such as EMA. In this focused review, we examined recent studies that apply ML and advanced computational methods to model stress using data from wearable devices in the wild. This review provides an in-depth assessment of current modeling practices and reported performance in this domain and adds to existing reviews by focusing specifically on problem formulation, detailed analysis of model performance, and methodological rigor. In particular, we highlight key problem formulation decisions, such as input and output window configuration, and make the distinction between nowcasting and forecasting tasks, and conduct a detailed analysis of model performance across varied modeling approaches. Through this analysis, we observe a few general patterns of model performance as a function of modeling approaches. We also identify important methodological issues, such as potential data leakage, which may threaten the validity of reported results and inflate reported model performance. We also noted a lack of standardization in datasets, task definitions, and reporting practices, making cross-study comparisons challenging and potentially slowing progress in wearable-based stress detection research. In response to these findings, we propose a model card framework to guide modeling decisions, experimental design, and reporting practices, with the goal of promoting methodological rigor, enhancing transparency, and reproducibility. We hope this review serves as a roadmap for new researchers entering the field, while also offering a framework for researchers working on similar problems to share findings and collaboratively advance the field as a community.

Acknowledgments

Generative AI tools were not used to draft or write the manuscript. The authors reviewed and approved all content and remain fully responsible for the accuracy, originality, and integrity of the manuscript.

Funding

The authors declared no financial support was received for this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy and study selection for PubMed (PRISMA-ScR Item 8).
[DOCX File, 2105 KB - mhealth_v14i1e76632_app1.docx]

Multimedia Appendix 2
Dataset characteristics (XLSX file, 9 KB).
[XLSX File, 8 KB - mhealth_v14i1e76632_app2.xlsx]

Multimedia Appendix 3
Machine learning framework (XLSX file, 16 KB).
[XLSX File, 16 KB - mhealth_v14i1e76632_app3.xlsx]

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 47 KB - mhealth_v14i1e76632_app4.docx]

References

1. Razavi M, Ziyadidegan S, Sasangohar F. Machine learning techniques for prediction of stress-related mental disorders: a scoping review. Proc Hum Factors Ergon Soc Annu Meet 2022 Sep;66(1):300-304. [doi: 10.1177/1071181322661298]
2. Cohen S, Janicki‐deverts D. Who's stressed? Distributions of psychological stress in the United States in probability samples from 1983, 2006, and 2009. J Appl Soc Pyschol 2012 Jun;42(6):1320-1334. [doi: 10.1111/j.1559-1816.2012.00900.x]
3. Bolpagni M, Pardini S, Dianti M, Gabrielli S. Personalized stress detection using biosignals from wearables: a scoping review. Sensors (Basel) 2024 May 18;24(10):3221. [doi: 10.3390/s24103221] [Medline: 38794074]

XSL•FO
RenderX

4. Iqbal S, Howse J, Banu S, et al. The effects of stress on health. Psychiatr Ann 2024 Oct;54(10):e272-e276. [doi: 10.3928/00485713-20241004-02]
5. Dimsdale JE. Psychological stress and cardiovascular disease. J Am Coll Cardiol 2008 Apr 1;51(13):1237-1246. [doi: 10.1016/j.jacc.2007.12.024] [Medline: 18371552]
6. Genet JJ, Siemer M. Rumination moderates the effects of daily events on negative mood: results from a diary study. Emotion 2012 Dec;12(6):1329-1339. [doi: 10.1037/a0028070] [Medline: 22775130]
7. Wijsman J, Grundlehner B, Liu H, Hermens H, Penders J. Towards mental stress detection using wearable physiological sensors. Annu Int Conf IEEE Eng Med Biol Soc 2011;2011:1798-1801. [doi: 10.1109/IEMBS.2011.6090512] [Medline: 22254677]
8. Dobson R, Li LL, Garner K, Tane T, McCool J, Whittaker R. The use of sensors to detect anxiety for in-the-moment intervention: scoping review. JMIR Ment Health 2023 Feb 2;10(1):e42611. [doi: 10.2196/42611] [Medline: 36729590]
9. Scalise L, Cosoli G. Wearables for health and fitness: measurement characteristics and accuracy. Presented at: 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC); May 14-17, 2018; Houston, TX, USA p. 1-6. [doi: 10.1109/I2MTC.2018.8409635]
10. Bakker J, Pechenizkiy M, Sidorova N. What's your current stress level? detection of stress patterns from GSR sensor data. 2011 Presented at: 2011 IEEE 11th International Conference on Data Mining Workshops; Dec 11-14, 2011; Vancouver, Canada p. 573-580. [doi: 10.1109/ICDMW.2011.178]
11. Akmandor AO, Jha NK. Keep the stress away with SODA: stress detection and alleviation system. IEEE Trans Multi-Scale Comp Syst 2017;3(4):269-282. [doi: 10.1109/TMSCS.2017.2703613]
12. Sun FT, Kuo C, Cheng HT, Buthpitiya S, Collins P, Griss M. Activity-aware mental stress detection using physiological sensors. In: Mobile Computing, Applications, and Services: Springer; 2012:282-301. [doi: 10.1007/978-3-642-29336-8_16]
13. Can YS, Arnrich B, Ersoy C. Stress detection in daily life scenarios using smart phones and wearable sensors: a survey. J Biomed Inform 2019 Apr;92:103139. [doi: 10.1016/j.jbi.2019.103139] [Medline: 30825538]
14. Can YS, Chalabianloo N, Ekiz D, Ersoy C. Continuous stress detection using wearable sensors in real life: algorithmic programming contest case study. Sensors (Basel) 2019 Apr 18;19(8):1849. [doi: 10.3390/s19081849] [Medline: 31003456]
15. González Ramírez ML, García Vázquez JP, Rodríguez MD, Padilla-López LA, Galindo-Aldana GM, Cuevas-González D. Wearables for stress management: a scoping review. Healthcare (Basel) 2023 Aug 22;11(17):2369. [doi: 10.3390/healthcare11172369] [Medline: 37685403]
16. Namvari M, Lipoth J, Knight S, et al. Photoplethysmography enabled wearable devices and stress detection: a scoping review. J Pers Med 2022 Oct 31;12(11):1792. [doi: 10.3390/jpm12111792] [Medline: 36579537]
17. Long N, Lei Y, Peng L, Xu P, Mao P. A scoping review on monitoring mental health using smart wearable devices. Math Biosci Eng 2022 May 27;19(8):7899-7919. [doi: 10.3934/mbe.2022369] [Medline: 35801449]
18. Pinge A, Gad V, Jaisighani D, Ghosh S, Sen S. Detection and monitoring of stress using wearables: a systematic review. Front Comput Sci 2024;6:1478851. [doi: 10.3389/fcomp.2024.1478851]
19. Gjoreski M, Luštrek M, Gams M, Gjoreski H. Monitoring stress with a wrist device using context. J Biomed Inform 2017 Sep;73:159-170. [doi: 10.1016/j.jbi.2017.08.006] [Medline: 28803947]
20. Han HJ, Labbaf S, Borelli JL, Dutt N, Rahmani AM. Objective stress monitoring based on wearable sensors in everyday settings. J Med Eng Technol 2020 May;44(4):177-189. [doi: 10.1080/03091902.2020.1759707] [Medline: 32589065]
21. Gaballah A, Tiwari A, Narayanan S, Falk TH. Context-aware speech stress detection in hospital workers using bi-LSTM classifiers. Presented at: ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); Jun 6-12, 2021; Toronto, ON, Canada p. 8348-8352. [doi: 10.1109/ICASSP39728.2021.9414666]
22. Robles-Granda P, Lin S, Wu X, et al. Jointly predicting job performance, personality, cognitive ability, affect, and well-being. IEEE Comput Intell Mag 2021;16(2):46-61. [doi: 10.1109/MCI.2021.3061877]
23. Mitchell M, Wu S, Zaldivar A, et al. Model cards for model reporting. Presented at: FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency; Jan 29-31, 2019; Atlanta, GA, USA p. 220-229. [doi: 10.1145/3287560.3287596]
24. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
25. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
26. Mundnich K, Booth BM, L'Hommedieu M, et al. TILES-2018, a longitudinal physiologic and behavioral data set of hospital workers. Sci Data 2020 Oct 16;7(1):354. [doi: 10.1038/s41597-020-00655-3] [Medline: 33067468]
27. Yfantidou S, Karagianni C, Efstathiou S, et al. LifeSnaps, a 4-month multi-modal dataset capturing unobtrusive snapshots of our lives in the wild. Sci Data 2022 Oct 31;9(1):663. [doi: 10.1038/s41597-022-01764-x] [Medline: 36316345]
28. Mattingly SM, Gregg JM, Audia P, et al. The Tesserae project: large-scale, longitudinal, in situ multimodal sensing of information workers. Presented at: CHI EA '19: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland, UK p. 1-8. [doi: 10.1145/3290607.3299041]
29. Smets E, Rios Velazquez E, Schiavone G, et al. Large-scale wearable data reveal digital phenotypes for daily-life stress detection. NPJ Digit Med 2018;1(1):67. [doi: 10.1038/s41746-018-0074-9] [Medline: 31304344]

XSL•FO
RenderX

30. Boateng G, Kotz D. StressAware: an app for real-time stress monitoring on the amulet wearable platform. 2016 Presented at: 2016 IEEE MIT Undergraduate Research Technology Conference (URTC); Nov 4-6, 2016; Cambridge, MA p. 1-4. [doi: 10.1109/URTC.2016.8284068]
31. Tervonen J, Puttonen S, Sillanpää MJ, et al. Personalized mental stress detection with self-organizing map: from laboratory to the field. Comput Biol Med 2020 Sep;124:103935. [doi: 10.1016/j.compbiomed.2020.103935] [Medline: 32771674]
32. Bavaresco R, Barbosa J, Vianna H, Büttenbender P, Dias L. Design and evaluation of a context-aware model based on psychophysiology. Comput Methods Programs Biomed 2020 Jun;189:105299. [doi: 10.1016/j.cmpb.2019.105299] [Medline: 31935581]
33. de Vries HJ, Pennings HJM, van der Schans CP, Sanderman R, Oldenhuis HKE, Kamphuis W. Wearable-measured sleep and resting heart rate variability as an outcome of and predictor for subjective stress measures: a multiple N-of-1 observational study. Sensors (Basel) 2022 Dec 28;23(1):332. [doi: 10.3390/s23010332] [Medline: 36616929]
34. de Vries H, Oldenhuis H, van der Schans C, Sanderman R, Kamphuis W. Does wearable-measured heart rate variability during sleep predict perceived morning mental and physical fitness? Appl Psychophysiol Biofeedback 2023 Jun;48(2):247-257. [doi: 10.1007/s10484-022-09578-8] [Medline: 36622531]
35. Tump D, Narayan N, Verbiest V, et al. Stressors and destressors in working from home based on context and physiology from self-reports and smartwatch measurements: international observational study trial. JMIR Form Res 2022 Nov 10;6(11):e38562. [doi: 10.2196/38562] [Medline: 36265030]
36. Mishra V, Pope G, Lord S, et al. Continuous detection of physiological stress with commodity hardware. ACM Trans Comput Healthc 2020 Apr;1(2):1-30. [doi: 10.1145/3361562] [Medline: 32832933]
37. Schmidt P, Dürichen R, Reiss A, Laerhoven K, Plötz T. Multi-target affect detection in the wild: an exploratory study. ISWC '19: Proceedings of the 2019 ACM International Symposium on Wearable Computers 2019:211-219. [doi: 10.1145/3341163.3347741]
38. Booth BM, Vrzakova H, Mattingly SM, Martinez GJ, Faust L, D'Mello SK. Toward robust stress prediction in the age of wearables: modeling perceived stress in a longitudinal study with information workers. IEEE Trans Affective Comput 2022;13(4):2201-2217. [doi: 10.1109/TAFFC.2022.3188006]
39. Martinez GJ, Grover T, Mattingly SM, et al. Alignment between heart rate variability from fitness trackers and perceived stress: perspectives from a large-scale in situ longitudinal study of information workers. JMIR Hum Factors 2022 Aug 4;9(3):e33754. [doi: 10.2196/33754] [Medline: 35925662]
40. Hadjiantonis G, Paromita P, Mundnich K, et al. Dynamical systems modeling of day-to-day signal-based patterns of emotional self-regulation and stress spillover in highly-demanding health professions. Annu Int Conf IEEE Eng Med Biol Soc 2020 Jul;2020:284-287. [doi: 10.1109/EMBC44109.2020.9175604] [Medline: 33017984]
41. Ravuri V, Paromita P, Mundnich K, et al. Group-specific models of healthcare workers' well-being using iterative participant clustering. Presented at: 2020 Second International Conference on Transdisciplinary AI (TransAI); Sep 21-23, 2020; Irvine, CA, USA p. 115-118. [doi: 10.1109/TransAI49837.2020.00026]
42. Tiwari A, Falk TH. New measures of heart rate variability based on subband tachogram complexity and spectral characteristics for improved stress and anxiety monitoring in highly ecological settings. Front Signal Process 2021;1:737881. [doi: 10.3389/frsip.2021.737881]
43. Paromita P, Mundnich K, Nadarajan A, Booth BM, Narayanan SS, Chaspari T. Modeling inter-individual differences in ambulatory-based multimodal signals via metric learning: a case study of personalized well-being estimation of healthcare workers. Front Digit Health 2023;5:1195795. [doi: 10.3389/fdgth.2023.1195795] [Medline: 37363272]
44. Stojchevska M, Steenwinckel B, Van Der Donckt J, et al. Assessing the added value of context during stress detection from wearable data. BMC Med Inform Decis Mak 2022 Oct 15;22(1):268. [doi: 10.1186/s12911-022-02010-5] [Medline: 36243691]
45. Parousidou V, Yfantidou S, Karagianni C, Vakali A. Stress beats: a continuum of learning methods for personalized stress detection. Presented at: 2023 IEEE International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT); Oct 26-29, 2023; Venice, Italy p. 40-47. [doi: 10.1109/WI-IAT59888.2023.00012]
46. Yang H, Yu H, Sridhar K, Vaessen T, Myin-Germeys I, Sano A. More to less (M2L): enhanced health recognition in the wild with reduced modality of wearable sensors. Presented at: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); Jul 11-15, 2022; Glasgow, Scotland, United Kingdom p. 3253-3256. [doi: 10.1109/EMBC48229.2022.9871472]
47. Paraschou E, Yfantidou S, Vakali A. UnStressMe: explainable stress analytics and self-tracking data visualizations. Presented at: 2023 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops); Mar 13-17, 2023; Atlanta, GA, USA p. 340-342. [doi: 10.1109/PerComWorkshops56833.2023.10150274]
48. Yu H, Sano A. Semi-supervised learning and data augmentation in wearable-based momentary stress detection in the wild. Proc ACM Interact Mob Wearable Ubiquitous Technol 2023 Jun 12;7(2):1-23. [doi: 10.1145/3596246]
49. Feng T, Narayanan S. Exploring workplace behaviors through speaking patterns using large-scale multimodal wearable recordings: a study of healthcare providers. arXiv. Preprint posted online on Dec 18, 2022. [doi: 10.48550/arXiv.2212.09090]

XSL•FO
RenderX

50. Saylam B, İncel Ö. Quantifying digital biomarkers for well-being: stress, anxiety, positive and negative affect via wearable devices and their time-based predictions. Sensors (Basel) 2023 Nov 5;23(21):8987. [doi: 10.3390/s23218987] [Medline: 37960685]
51. Kao HT, Yan S, Hosseinmardi H, Narayanan S, Lerman K, Ferrara E. User-based collaborative filtering mobile health system. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020 Dec 17;4(4):1-17. [doi: 10.1145/3432703]
52. Burghardt K, Tavabi N, Ferrara E, Narayanan S, Lerman K. Having a bad day? detecting the impact of atypical events using wearable sensors. Presented at: Social, Cultural, and Behavioral Modeling: 14th International Conference, SBP-BRiMS 2021; Jul 6-9, 2021 p. 257-267. [doi: 10.1007/978-3-030-80387-2_25]
53. Feng T, Booth BM, Baldwin-Rodríguez B, Osorno F, Narayanan S. A multimodal analysis of physical activity, sleep, and work shift in nurses with wearable sensor data. Sci Rep 2021 Apr 22;11(1):8693. [doi: 10.1038/s41598-021-87029-w] [Medline: 33888731]
54. Jiang J, Lerman K, Ferrara E. Zero-shot meta-learning for small-scale data from human subjects. Presented at: 2023 IEEE 11th International Conference on Healthcare Informatics (ICHI); Jun 26-29, 2023; Houston, TX, USA p. 311-320. [doi: 10.1109/ICHI57859.2023.00049]
55. Pimentel A, Tiwari A, Narayanan S, Falk TH. Human mental state monitoring in the wild: are we better off with deeper neural networks or improved input features. 2021 Presented at: CMBEC 44; May 11, 2021 URL: https://proceedings.cmbes.ca/index.php/proceedings/article/view/941 [accessed 2026-07-02]
56. Zanna K, Sridhar K, Yu H, Sano A. Bias reducing multitask learning on mental health prediction. Presented at: 2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII); Oct 17-21, 2022; Nara, Japan p. 1-8. [doi: 10.1109/ACII55700.2022.9953850]
57. Li N, Zhang H, Feng L, Ding Y, Li H. Analyzing and identifying predictable time range for stress prediction based on chaos theory and deep learning. Health Inf Sci Syst 2024 Dec;12(1):16. [doi: 10.1007/s13755-024-00280-z] [Medline: 39185396]
58. Saylam B, Durmaz İncel Ö. Extracting digital biomarkers for unobtrusive stress state screening from multimodal wearable data. Presented at: Smart Technologies for Sustainable and Resilient Ecosystems Edge-IoT SmartGov 2022; Jun 18, 2023 p. 130-151. [doi: 10.1007/978-3-031-35982-8_10]
59. Kim Y, Xu X, McDuff D, Breazeal C, Park HW. Health-LLM: large language models for health prediction via wearable sensor data. arXiv. Preprint posted online on 2024. [doi: 10.48550/arXiv.2401.06866]
60. Garcez AD, Lamb LC. Neurosymbolic AI: the 3rd wave. Artif Intell Rev 2023 Nov;56(11):12387-12406. [doi: 10.1007/s10462-023-10448-w]
61. Monroe D. Neurosymbolic AI. Commun ACM 2022 Oct;65(10):11-13. [doi: 10.1145/3554918]

Abbreviations

AB: AdaBoost
AUC: area under the receiver operating characteristic curve
BACC: balanced accuracy
BI-LSTM: bidirectional long short-term memory network
CNN: convolutional neural network
EMA: ecological momentary assessment
FFN: feedforward neural network
GB: gradient boosting
GMM: Gaussian mixture model
HF: high frequency
IRB: institutional review board
IRF: impulse response function
KNN: k-nearest neighbors
LF: low frequency
LLM: large language model
LR: logistic regression
LSTM: long short-term memory
MCC: Matthews correlation coefficient
ML: machine learning
MSE: mean-squared error
PANAS: Positive and Negative Affect Schedule
PCA: principal component analysis
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

XSL•FO
RenderX

PSS: Perceived Stress Scale
RF: random forest
RNN: recurrent neural network
S-STAI: Short State-Trait Anxiety Inventory
SNN: Siamese neural network
SOM: self-organizing map
STAI: State-Trait Anxiety Inventory
STAI: State-Trait Anxiety Inventory
SVM: support vector machine
VAR: vector autoregression
XGBoost: eXtreme Gradient Boosting

Edited by R Chandrasekaran; submitted 29.Apr.2025; peer-reviewed by A Gupta, H Rafi, J Kallio, RT Potla, Y Zhang; revised version received 20.May.2026; accepted 22.May.2026; published 31.Jul.2026.

Please cite as:
Sharma S, Janakiraman AK, Chen LK
Machine Learning Frameworks for Wearable-Based Stress Modeling in Naturalistic Settings: Scoping Review
JMIR Mhealth Uhealth 2026;14:e76632
URL: https://mhealth.jmir.org/2026/1/e76632
doi:10.2196/76632

© Shifali Sharma, Aswin Kumar Janakiraman, Lujie Karen Chen. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 31.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Wearable Devices for Monitoring and Management of Comorbid Obstructive Sleep Apnea and Hypertension: Scoping Review

Hongyi Wang[1,2*], MMed; Anran Cheng[3*], MMed; Yusheng Li[1,2], MMed; Haining Zhang[1,2], PhD; Shun Fan[1,2], PhD; Rong Shen[4], BMed; Wei Zhang[1,2], MMed; Huanan Li[1,2], PhD; Jingui Wang[1,2], PhD

[1]Tuina Department, First Teaching Hospital of Tianjin University of Traditional Chinese Medicine, No. 88, Chang Ling Road, Li Qi Zhuang Jie, Xi Qing District, Tianjin, China

[2]National Clinical Research Center for Chinese Medicine, Tianjin, China

[3]Department of Sleep Medicine Center, Longhua Hospital Shanghai University of Traditional Chinese Medicine, Shanghai, China

[4]Anhui University of Traditional Chinese Medicine, Hefei, Anhui, China

[*]these authors contributed equally

Corresponding Author:
Jingui Wang, PhD
Tuina Department, First Teaching Hospital of Tianjin University of Traditional Chinese Medicine, No. 88, Chang Ling Road, Li Qi Zhuang Jie, Xi Qing District, Tianjin, China

Abstract

Background: As an established driver of hypertension, obstructive sleep apnea (OSA) generates a significant cardiovascular burden when both disorders are present. This intersection not only compromises nocturnal hemodynamics but also hampers long-term clinical management. Yet, current health care delivery rarely integrates the simultaneous and continuous tracking of these dual burdens. While wearable technology provides a noninvasive, pragmatic toolset for synchronized physiological monitoring, research remains largely siloed within single-disease frameworks. Consequently, clinical evidence supporting wearable applications specifically for comorbid populations remains sparse.

Objective: In this scoping review, we summarized the current applications of wearable technology for comorbid OSA and hypertension. The analysis primarily outlines device categories, monitored physiological indicators, prevalent clinical scenarios, and existing challenges.

Methods: The search for relevant literature spanned PubMed, Web of Science, Embase, and IEEE Xplore, covering the period from January 2015 to February 2026. Eligible studies included adults and used wearable or portable technologies to objectively track sleep- or respiratory-related indicators and cardiovascular/blood pressure metrics. Study selection, data charting, and evidence synthesis were conducted using a 2-reviewer process and descriptive and narrative approaches.

Results: Our initial search yielded 739 records. Following title and abstract screening, we reviewed 54 full texts, ultimately finalizing a cohort of 13 eligible studies. Published between 2015 and 2026 across 9 countries, these articles capture data from 5596 participants. Most were observational studies and device validation studies. Evaluated device types included wrist-worn devices, fingertip contact devices, patch and single-lead devices, and multiparameter portable monitoring systems. The clinical applications mainly focused on screening and risk stratification for comorbid OSA and hypertension, monitoring of abnormal nocturnal blood pressure and hemodynamic changes, and cardiovascular risk assessment and remote longitudinal management. However, research remains sparse, and different devices vary greatly in reference standards, diagnostic thresholds, and validation pathways; therefore, their clinical translational value still requires further validation.

Conclusions: Wearable devices may complement traditional assessment by providing continuous nocturnal and longitudinal data. However, at present, they are more suitable as auxiliary monitoring and risk warning tools in comorbidity management and cannot yet replace standard sleep studies or standard blood pressure monitoring. Future research should further shift from feasibility validation to large-sample, prospective, multicenter clinical studies in comorbid populations.

(JMIR Mhealth Uhealth 2026;14:e84506) doi:10.2196/84506

KEYWORDS

obstructive sleep apnea; hypertension; comorbidity; wearable devices; mHealth; continuous monitoring

XSL•FO
RenderX

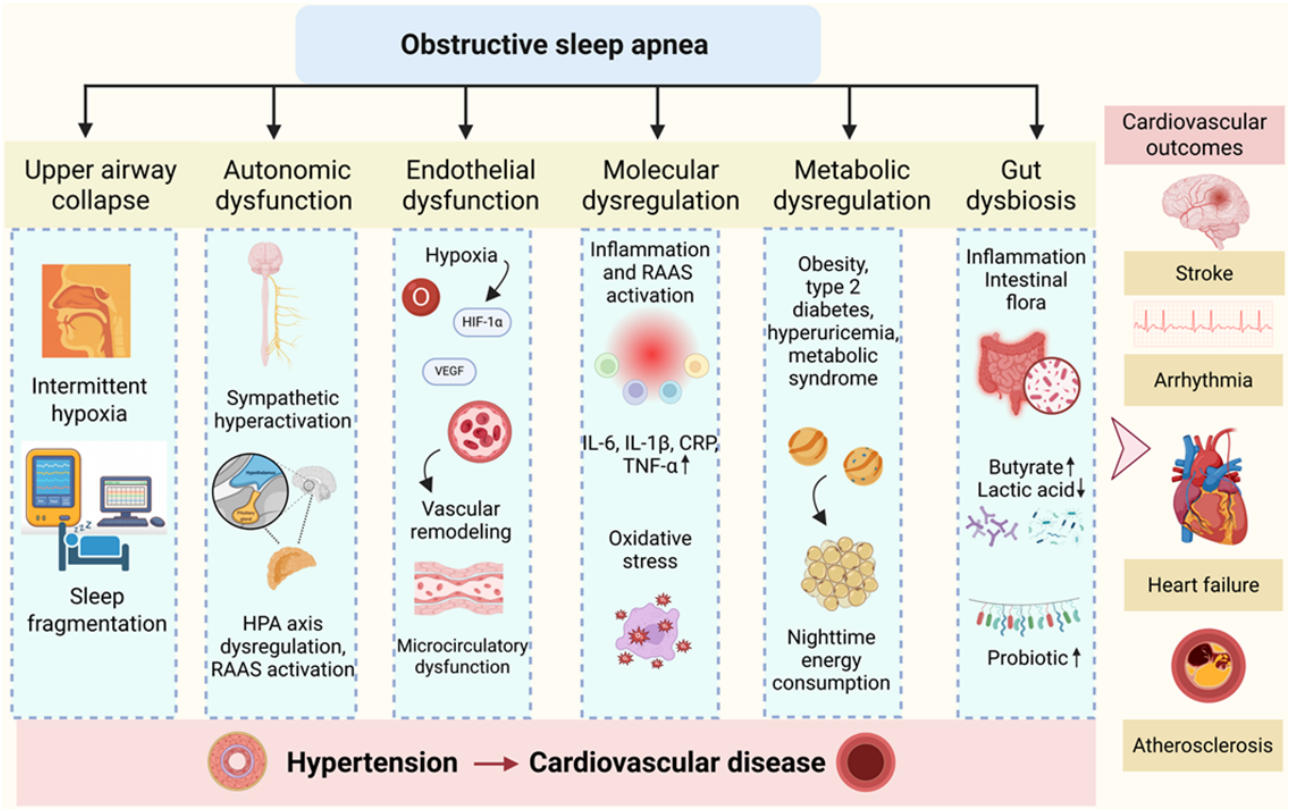
Introduction

Background

Obstructive sleep apnea (OSA) is one of the most common chronic respiratory diseases, affecting approximately 1 billion people worldwide, with the greatest management burden in China, the United States, Brazil, and India [1,2]. OSA has long been regarded as a disease that causes daytime sleepiness and reduced sleep quality. However, increasing evidence indicates that OSA is an independent risk factor for hypertension and a common cause of secondary hypertension [3]. The risk of incident hypertension is doubled in the population with OSA relative to healthy controls. OSA is also highly prevalent among patients with hypertension, particularly in more difficult-to-control cases [4,5]. These 2 conditions often coexist insidiously at night, and each apneic event is accompanied by marked hemodynamic fluctuations. This comorbid state is associated with a greater cardiovascular burden, abnormal nocturnal hemodynamics, and more complex long-term management, and it may also increase the lifetime risk of cardiovascular diseases such as coronary heart disease and stroke [6,7].

The link between OSA and hypertension may involve multiple interrelated mechanisms (Figure 1). The most central mechanism is recurrent nocturnal hypoxia-reoxygenation and hypercapnia. These nocturnal abnormalities can continuously activate the sympathetic nervous system, promote peripheral vasoconstriction, and increase blood pressure [8]. Neuroendocrine dysregulation may also be involved, as abnormal hypothalamic-pituitary-adrenal axis rhythms can lead to elevated cortisol levels and further increase blood pressure [9]. Vascular impairment is further characterized by endothelial dysfunction, which serves as a mechanistic link between the two disorders. Intermittent hypoxic events induce the expression of hypoxia-inducible factors, subsequently modulating downstream mediators such as vascular endothelial growth factor. These molecular events promote structural remodeling and compromise endothelial integrity, leading to an elevation in peripheral resistance [10]. Vascular alterations are exacerbated by inflammatory and oxidative stress pathways. Sleep fragmentation and intermittent hypoxia facilitate a systemic inflammatory response, which is characterized by elevated concentrations of interleukin-6, tumor necrosis factor-α, and C-reactive protein. These inflammatory mediators contribute to the progression of underlying vascular dysfunction [11,12]. Shifts in gut microbiota and metabolic processes can compromise intestinal epithelial integrity and elevate vascular tone [13]. Metabolic factors, including obesity, dyslipidemia, and insulin resistance, further contribute to cardiovascular impairment and reinforce the bidirectional relationship between OSA and hypertension [14,15]. Since these pathophysiological processes are predominantly nocturnal and dynamic, single-point outpatient measurements or assessments focused on a single disease often fail to represent the complete clinical profile of this comorbidity.

Figure 1. Proposed mechanisms linking obstructive sleep apnea and hypertension. CRP: C-reactive protein; HIF-1α: hypoxia-inducible factor 1 alpha; HPA: hypothalamic-pituitary-adrenal; IL-1β: interleukin 1 beta; IL-6: interleukin 6; RAAS: renin-angiotensin-aldosterone system; TNF-α: tumor necrosis factor alpha; VEGF: vascular endothelial growth factor.



At present, comorbid OSA and hypertension are still mostly managed separately across clinical specialties, with limited coordinated monitoring. Traditional data collection relies heavily on patient self-report or occasional in-hospital examinations

XSL•FO
RenderX

and cannot provide real-time monitoring [16]. The lack of key continuous nocturnal data makes it difficult to capture the dynamic relationship between OSA and hypertension. In response to these clinical needs, wearable devices have provided new technologies for the continuous monitoring and management of comorbid OSA and hypertension. These devices usually integrate miniaturized sensors into watches, wristbands, skin patches, rings, and glasses, and can continuously and noninvasively collect multidimensional physiological signals such as electrocardiography, heart rate, heart rate variability (HRV), blood pressure, peripheral oxygen saturation, and sleep [16]. Wearable devices are designed for long-term wear and can measure changes in indicators during both daytime and sleep, thereby filling the gap in continuous nocturnal data acquisition in traditional assessment and better reflecting the dynamic relationship between OSA-related events and blood pressure fluctuations [17,18]. Wearable devices are expected to provide more timely and objective information for the screening, monitoring, and long-term management of this comorbid condition.

In recent years, reviews on the application of wearable devices in sleep monitoring, blood pressure monitoring, and chronic disease management have continued to increase. Most existing reviews have focused on a single-disease context of OSA or hypertension, or have only addressed devices, algorithms, and sensing signals [19-22]. There remains a lack of reviews that integrate devices, indicators, validation pathways, and clinical applications within the context of comorbidity management, which limits the overall understanding of their value, applicability, and current limitations among researchers and clinicians.

Aims

This gap in the literature has hindered researchers' understanding of the technology. Therefore, we conducted a scoping review to evaluate the current applications of wearable devices in the screening and management of comorbid OSA and hypertension, analyze recent developments and existing challenges, and inform future research directions.

We specifically addressed four aims: (1) to categorize wearable technologies used for the screening and management of concurrent OSA and hypertension, (2) to delineate the physiological parameters these tools monitor to evaluate respiratory and cardiovascular metrics, (3) to describe the clinical settings for device implementation and their reported practical use, and (4) to evaluate prevailing technical and methodological barriers, specifically regarding reference standards and clinical translation.

Methods

Overview

A scoping review approach was adopted in this study to address the substantial heterogeneity in device types, monitored indicators, and clinical application scenarios across the research questions, as well as the evidence gaps in previous reviews. The Arksey and O'Malley framework and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) guidelines guided the methodological design and reporting of this scoping review. Related reporting checklists and structural details are available in Checklist 1 [23].

Inclusion and Exclusion Criteria

The inclusion criteria were as follows: (1) clinical studies involving adults (aged ≥18 years); (2) studies in which the included population was explicitly a comorbid OSA and hypertension population, or studies in which the population had a single disease but both sleep- or respiratory-related indicators and blood pressure/cardiovascular indicators were assessed to evaluate the relationship between the 2 conditions or their comorbidity-related clinical applications; (3) studies using noninvasive wearable or portable devices to obtain at least 1 type of objective physiological monitoring data; (4) original studies published in English; and (5) studies with accessible full text. The exclusion criteria were as follows: (1) reviews, systematic reviews, meta-analyses, study protocols, case reports, animal experiments, commentaries, conference abstracts, and articles without full text were excluded; (2) studies that focused only on OSA or only on hypertension without objectively assessing disease-related indicators of the other condition; (3) studies that used only traditional in-hospital or laboratory monitoring without wearable or portable monitoring data; and (4) studies that were not clinically or practically validated based on real human data were excluded.

Data Sources and Search Strategy

We comprehensively searched 4 databases, including PubMed, Web of Science, Embase, and IEEE Xplore. Original studies on wearable devices in the management of comorbid OSA and hypertension published between January 1, 2015, and February 28, 2026, were searched, and the final database search was completed on February 28, 2026. The search strategy included MeSH and free-text terms related to wearable devices and monitoring technologies (eg, "wearable electronics," "portable devices," "smartwatches," "actigraphs," "photoplethysmography," "remote monitoring," and "continuous monitoring"), OSA (eg, "obstructive sleep apnea," "sleep apnea," "sleep-disordered breathing," and "snoring"), and hypertension (eg, "hypertension" and "ambulatory blood pressure monitoring"); the full search strategy is provided in Multimedia Appendix 1. To ensure the comprehensiveness of the literature collection, we manually traced the reference lists of all finally included studies. All retrieved records were imported into EndNote 21 (Clarivate) for deduplication and screening, and duplicate records were manually checked.

Study Selection

Two researchers (HW and AC) independently conducted the initial screening of titles and abstracts to exclude records that did not meet the criteria. Following the preliminary screen, the remaining records underwent a complete full-text appraisal to confirm alignment with the predefined eligibility framework. Selection-related discrepancies were settled through consensus-based deliberation involving a third investigator (YL).

XSL•FO
RenderX

Data Extraction and Charting

Two authors (HW and AC) jointly developed a standardized data-extraction and charting form based on the aims of this review and the predefined research questions. Before full data extraction, the 2 authors piloted the form on 3 randomly selected included studies to calibrate the definitions of key variables and classification rules and to improve consistency between reviewers. Data charting was treated as an iterative process; during piloting and discussion, the categories and recording rules in the extraction form were refined and then applied consistently in the final form. Subsequently, both authors independently extracted and charted data from all included studies. Disagreements were first resolved through discussion and consensus between the 2 authors; if consensus could not be reached, a third researcher (YL) was consulted.

We extracted study-level information across 5 domains: (1) study characteristics, including author, publication year, country, study design, sample size, participant characteristics, and study setting; (2) device characteristics, including wearable device type, sensing modality, and continuous monitoring capability; (3) physiological monitoring parameters, including sleep- or respiratory-related indicators and cardiovascular or hemodynamic indicators; (4) methodological characteristics, including reference standards, diagnostic or stratification thresholds, and validation design; and (5) clinical characteristics, including clinical application scenarios and study limitations. For studies involving multiple monitoring targets or application scenarios, the research team determined the primary focus according to the study objective, main outcome, and emphasis of reported findings, while also recording secondary or supportive features. The charted data were used to develop Table 1, Figure 2 [24-36], and Multimedia Appendices 2-4.

XSL•FO
RenderX

Table . Summary of study characteristics. Categories are not mutually exclusive; individual studies may contribute to multiple rows. For studies involving multiple monitoring targets or clinical applications, classification was based on the primary study objective, main outcome, and emphasis of reported findings. Respiratory event frequency metrics included AHI[a], REI[b], ODI[c], or algorithm-estimated equivalents. Nocturnal BP[d] variables included mean BP, peak BP, morning BP surge, BP variability, beat-to-beat BP, and cuffless BP estimates. Reference standards and assessment pathways were recorded as reported in each study.

Category	Studies, n (%)	References
Study design		
Observational studies	8 (62)	[24-31]
Device validation studies	4 (31)	[32-35]
Feasibility studies	1 (8)	[36]
Study setting/data source		
Sleep laboratory or hospital-based studies	9 (69)	[25-28,30,32-35]
Home or remote wearable monitoring studies	3 (23)	[24,29,36]
Community-based or real-world screening studies	3 (23)	[24,29,31]
Population/clinical context		
Suspected OSA[e]/screening populations	9 (69)	[24,27-29,32-36]
Confirmed OSA	3 (23)	[27,28,33]
CPAP-treated[f] OSA	1 (8)	[36]
Primary hypertension	1 (8)	[25]
Older adults with hypertension	1 (8)	[31]
Obese adults at risk of abnormal nocturnal BP	1 (8)	[26]
Device type		
Wrist-worn devices	3 (23)	[24,29,36]
Fingertip contact devices	4 (31)	[25,26,30,34]
Patch and single-lead devices	2 (15)	[28,31]
Multiparameter portable monitoring systems	4 (31)	[27,32,33,35]
Sensor technologies		
PAT-based[g] wearable studies	2 (15)	[25,26]
PPG/pulse[h] oximetry–based wearable studies	8 (62)	[24,27,30,32-36]
Single-lead ECG–based[i] screening studies	1 (8)	[31]
PTT-derived[j] cuffless BP monitoring studies	2 (15)	[32,35]
Pulse wave/vascular analysis studies	2 (15)	[30,34]
Acoustic/tracheal sound wearable studies	1 (8)	[28]
Consumer sleep/activity tracker studies	1 (8)	[29]
Sleep and respiratory variables		
Sleep duration or sleep-variability metrics	3 (23)	[26,29,36]
Respiratory event frequency metrics (AHI/REI/ODI or equivalents)	12 (92)	[24-28,30-36]
Nocturnal hypoxia metrics (eg, SpO_2[k] nadir, T<90, ODT[l])	9 (69)	[24-28,32-35]
Event- or stage-specific analyses	4 (31)	[27,28,33,35]
Cardiovascular variables		
Nocturnal BP level, surge, or variability	7 (54)	[25-27,32,33,35,36]
Beat-to-beat or continuous BP monitoring	3 (23)	[27,32,36]

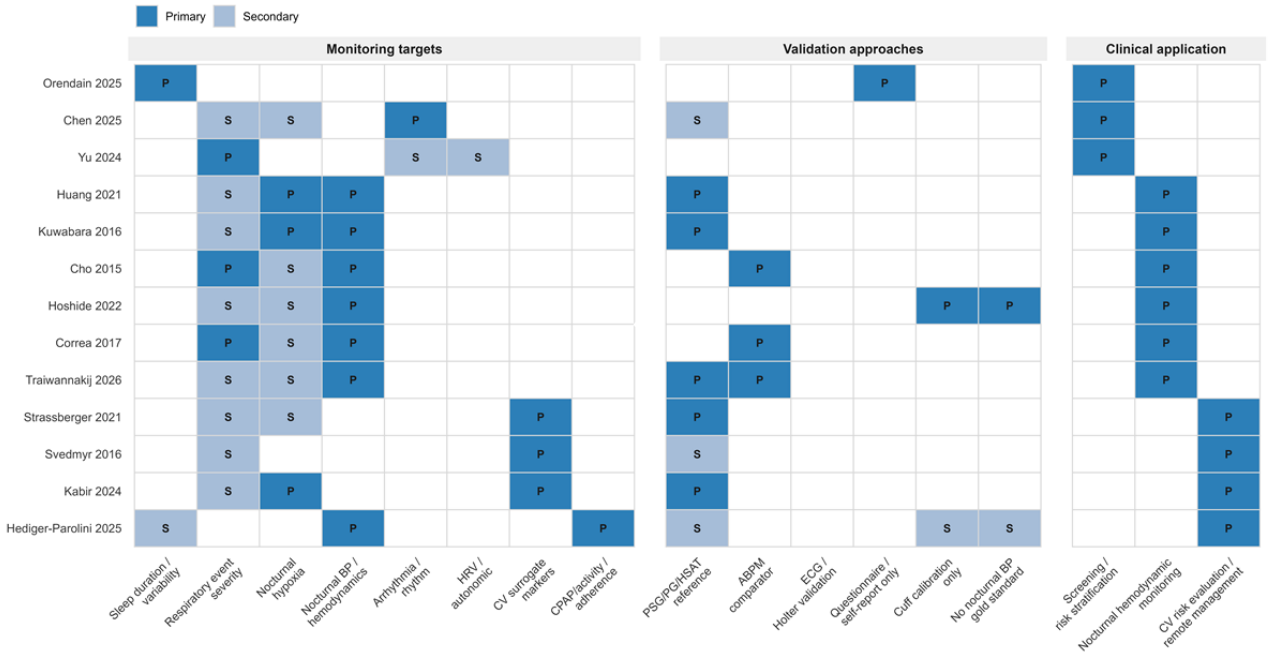
XSL•FO
RenderX

Category	Studies, n (%)	References
Arrhythmia or rhythm abnormality monitoring	2 (15)	[24,31]
HRV[m] or autonomic markers	2 (15)	[30,31]
Pulse wave or vascular stiffness markers	2 (15)	[30,34]
Reference standard/assessment pathway		
PSG[n], polygraphy, or HSAT[o]	8 (62)	[24,27,28,32-36]
ABPM[p]	3 (23)	[25,26,35]
ECG/Holter	1 (8)	[31]
Questionnaire or self-report	1 (8)	[29]
Office cuff calibration without an external nocturnal BP gold standard	2 (15)	[32,36]
No external nocturnal BP gold standard despite BP-focused monitoring	3 (23)	[27,32,33]
Clinical application		
OSA comorbidity screening and risk stratification	3 (23)	[24,29,31]
Nocturnal BP abnormality and hemodynamic monitoring	6 (46)	[25-27,32,33,35]
CV[q] risk evaluation and remote longitudinal management	4 (31)	[28,30,34,36]

[a]AHI: apnea-hypopnea index.
[b]REI: respiratory event index.
[c]ODI: oxygen desaturation index.
[d]BP: blood pressure.
[e]OSA: obstructive sleep apnea.
[f]CPAP: continuous positive airway pressure.
[g]PAT: peripheral arterial tonometry.
[h]PPG: photoplethysmography.
[i]ECG: electrocardiography.
[j]PTT: pulse transit time.
[k]SpO_2: peripheral oxygen saturation.
[l]ODT: oxygen desaturation time.
[m]HRV: heart rate variability.
[n]PSG: polysomnography.
[o]HSAT: home sleep apnea test.
[p]ABPM: ambulatory blood pressure monitoring.
[q]CV: cardiovascular.

XSL•FO
RenderX

Figure 2. Study-by-dimension matrix of monitoring targets, validation approaches, and clinical applications across the included studies [24-36]. P indicates a primary focus, and S indicates a secondary or supportive feature. Blank cells indicate that the dimension was not central to the study. ABPM: ambulatory blood pressure monitoring; BP: blood pressure; CPAP: continuous positive airway pressure; CV: cardiovascular; ECG: electrocardiography; HRV: heart rate variability; HSAT: home sleep apnea test; PG: polygraphy; PSG: polysomnography.



Synthesis of Results

We synthesized the charted data using descriptive analysis and narrative synthesis. Using the charted data, we grouped similar findings into common themes. Information related to wearable device applications, sensing modalities, monitored indicators, reference standards, validation pathways, and clinical uses was described and organized. We did not prespecify the number of thematic categories or restrict the number of categories to which each study could contribute, in order to capture the diversity of the current evidence in terms of device types, monitoring targets, validation approaches, and clinical application scenarios. For each study, we reported the device characteristics, monitored variables, diagnostic or stratification thresholds, reference standards, and main findings as closely as possible to the original report, including information provided in supplementary materials where available. Similar content was then grouped into common themes without changing the meaning of the original studies.

Because the included studies were markedly heterogeneous in terms of wearable device types, sensing technologies, outcome assessment indicators, and measurement settings, we did not attempt to quantitatively pool the data or conduct a meta-analysis. Instead, the key study information, relevant findings, and methodological characteristics reported in each study were summarized and recorded. This approach was consistent with the purpose of a scoping review, which was to summarize study characteristics, identify evidence gaps, and clarify directions for future research rather than to estimate pooled effect sizes. Therefore, no formal risk-of-bias assessment or methodological quality appraisal was conducted for the included studies; relevant methodological and practical limitations were summarized in the Results and Discussion sections.

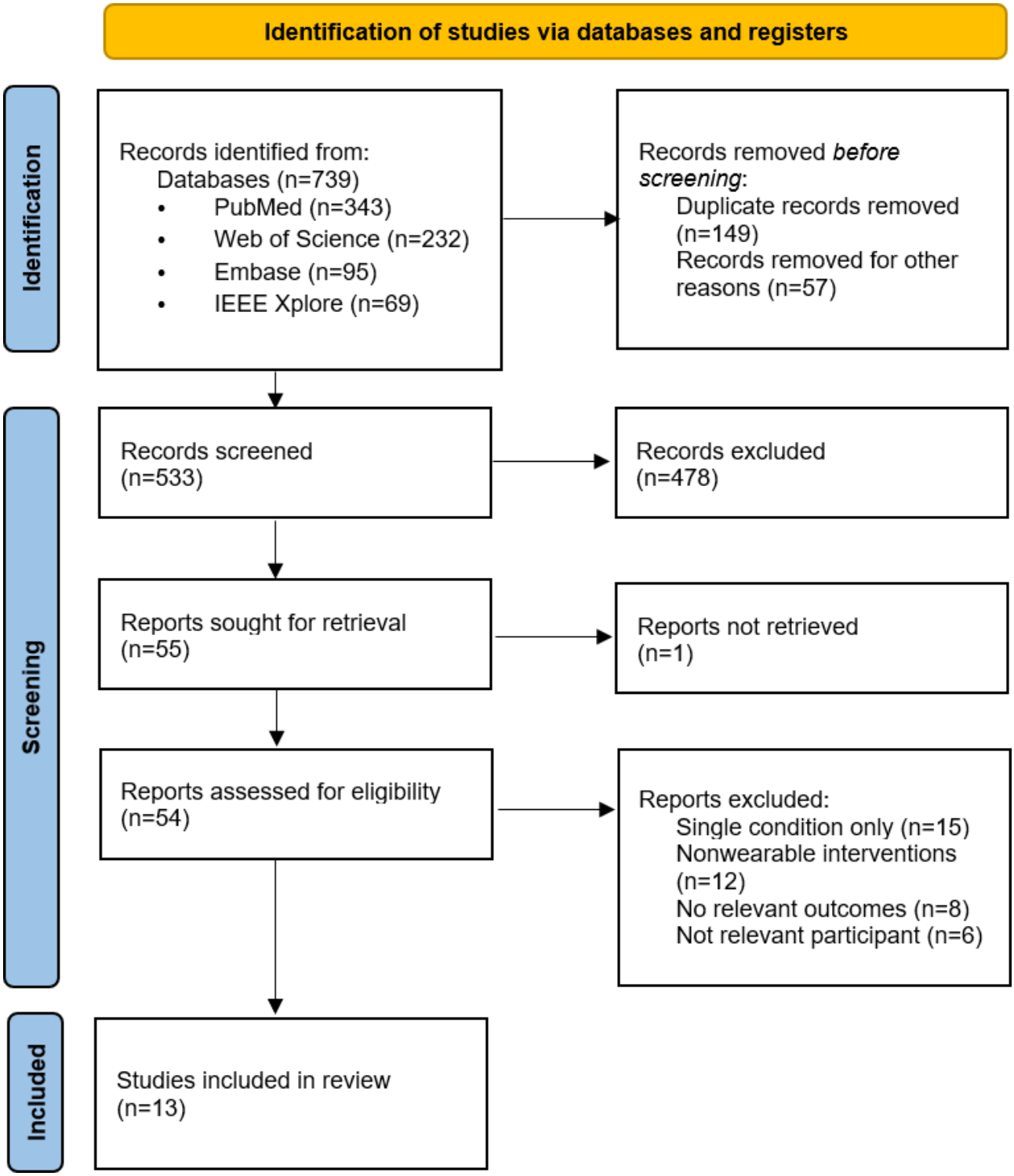
Results

Study Selection

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram (Figure 3) outlines the study selection process. A total of 739 studies were identified from the 4 databases. After removal of duplicates and records excluded for other reasons, 533 studies entered the title and abstract screening stage, and 478 articles were excluded after screening. Full texts of 55 articles were sought and assessed, of which 1 could not be retrieved. After full-text screening of the remaining 54 articles, 41 studies were excluded. Finally, 13 studies were considered eligible.

XSL•FO
RenderX

Figure 3. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of study selection.



Characteristics of Included Studies

The 13 studies included in this scoping review were published between 2015 and 2026 and were conducted in 9 countries or regions, namely China, Japan, South Korea, Thailand, the United States, Canada, Brazil, Switzerland, and Germany/Sweden, involving a total of 5596 participants. Among them, 1 real-world study included initial screening data from more than 1 million users [24]. The included literature comprised 8 (62%) observational studies [24-31], 4 (31%) device validation studies [32-35], and 1 (8%) feasibility study [36]. The study populations were markedly heterogeneous and covered multiple clinical scenarios in the management of comorbid OSA and hypertension. Participants included patients with OSA at different stages, ranging from suspected screening and confirmed diagnosis to treatment with continuous positive airway pressure (CPAP), as well as patients with primary hypertension, older adults with hypertension, and obese adults at risk of abnormal nocturnal blood pressure. A summary of the characteristics of the included studies is shown in Table 1, and detailed information on the included studies is provided in Multimedia Appendix 2.

XSL•FO
RenderX

Wearable Device Types

The 13 included studies used various types of wearable devices, showing the characteristics of portability and continuous monitoring; detailed device information is provided in Multimedia Appendix 3. These included wrist-worn devices (n=3, 23%) [24,29,36], fingertip contact devices (n=4, 31%) [25,26,30,34], patch and single-lead devices (n=2, 15%) [28,31], and multiparameter portable monitoring systems (n=4, 31%) [27,32,33,35]. Wrist-worn devices mainly included smartwatches, wristbands, and activity trackers, and because of their greater convenience and practicality, they were mainly used for community screening, daily monitoring, and remote follow-up. Multiple studies used fingertip contact devices to collect nocturnal respiratory events and pulse wave signals in order to extract key characteristic parameters such as oxygen desaturation index (ODI), HRV, pulse propagation time (PPT), and pulse transit time (PTT). Patch and single-lead devices were used for the noninvasive continuous acquisition of acoustic or electrocardiographic signals. Some studies also applied multiparameter portable systems that combined sleep respiratory monitoring with continuous ambulatory blood pressure monitoring (ABPM) to synchronously assess nocturnal respiratory events and their related hemodynamic changes.

Sensor Technologies

The included studies used multiple sensing technologies for the noninvasive acquisition of sleep and cardiovascular parameters. Optical sensing was the most commonly used technology, and photoplethysmography and pulse oximetry were the most widely applied, mainly for recording hypoxic events, hemodynamic fluctuations, and assessing nocturnal cardiovascular risk [24,27,30,34,36]. Two studies further combined optical and electrical signal sensing and extracted PTT by acquiring electrocardiography and fingertip photoplethysmography signals to achieve cuffless blood pressure estimation, thereby extending monitoring from OSA to the preliminary assessment of nocturnal hypertension or masked hypertension [32,35]. Some studies also used peripheral arterial tonometry (PAT), acoustic sensing, and actigraphy-related sensing technologies. Cho et al [25] and Correa et al [26] used PAT-related signals to assess changes in peripheral vascular tone and sympathetic nervous activity associated with sleep-related respiratory events [25,26]. Accelerometers were commonly used for activity state identification, sleep staging, and variability analysis [29,36]. Kabir et al [28] collected tracheal breath sounds and heart sound signals using miniature microphones and accelerometers and analyzed them synchronously with polysomnography.

Monitored Parameters

Based on the multimodal sensing technologies described above, current evidence suggests that some wearable devices can simultaneously capture sleep-related respiratory parameters and hemodynamic parameters. In the sleep-respiratory domain, the devices were mainly used to assess OSA severity and nocturnal hypoxic burden. The monitored indicators included the apnea-hypopnea index (AHI) or respiratory event index (REI), ODI, minimum or mean peripheral oxygen saturation (SpO_2), sleep time with oxygen saturation below 90%, or other indicators of hypoxic exposure [24-27,30-33,35]. Some studies obtained sleep rhythm indicators through long-term monitoring, including total sleep time, sleep efficiency, and sleep variability [26,29,30]. Other studies focused on respiratory signal features such as nasal airflow, snoring, and thoracoabdominal respiratory effort [28,32]. In comparison, cardiovascular and hemodynamic parameters may better reflect the comorbidity characteristics of OSA and hypertension. Wearable devices were able not only to derive conventional indicators such as resting heart rate and HRV [29,31] but also to obtain continuous beat-to-beat blood pressure, PTT-based nocturnal ambulatory blood pressure [27,32,35], systolic blood pressure peaks triggered by hypoxic events [33], PAT amplitude attenuation caused by sympathetic activation [25,26], and PPT, which is related to vascular stiffness or nocturnal cardiovascular burden [30]. In addition, some studies also recorded changes in heart sound intensity and nocturnal abnormal arrhythmic events [24,28,31]. Overall, there was marked heterogeneity across the included studies in terms of reference standards, assessment pathways, and diagnostic or stratification thresholds, including different assessment methods such as polysomnography, home sleep apnea testing (HSAT), ABPM, and questionnaire tools. Blood pressure thresholds are reported as used in the original studies for diagnostic or stratification purposes; treatment targets were not synthesized because most included studies focused on screening and monitoring rather than treatment-goal attainment. Therefore, the reported device performance should be interpreted objectively in the context of specific study settings (Multimedia Appendix 4).

Clinical Applications

Overview

Evidence suggests that wearable technology for comorbid OSA and hypertension management largely focuses on three functional areas: initial comorbidity screening and risk stratification in community and home settings, continuous monitoring of abnormal nocturnal hemodynamics and changes in respiratory events, and multimodal cardiovascular risk assessment and exploration of remote management. Figure 2 summarizes the monitoring targets, validation approaches, and major clinical application scenarios across the included studies.

Screening and Risk Stratification for Comorbid OSA and Hypertension

Three studies involved the application of wearable devices in the screening and risk stratification of comorbid OSA and hypertension [24,29,31]. Existing evidence suggests that wearable devices have been increasingly applied to the early identification and risk stratification of this comorbidity. In a large real-world screening study, wrist-worn smart devices were used for initial screening in 1,056,494 users, identifying 19,563 individuals at high risk for OSA, and 1054 patients with OSA were confirmed among those who subsequently underwent HSAT or polysomnography evaluation; the devices could also simultaneously indicate the risk of arrhythmia [24]. A study including 2573 older adults with hypertension in primary care settings found that continuous monitoring based on wearable electrocardiography and the autocorrelated wave detection with adaptive threshold (ACAT) algorithm could be used for community-based initial screening of OSA and could help

XSL•FO
RenderX

improve the identification of individuals at high risk for moderate to severe OSA [31]. Another study analyzed long-term sleep data collected by wearable devices together with hypertension information and found that increased sleep variability was associated with higher OSA risk and increased hypertension risk, suggesting that wearable devices may be used for the preliminary identification of comorbidity risk [29]. The application of wearable devices is no longer limited to the detection of OSA alone but has extended to the preliminary identification of comorbidity risk. However, their role still lies in early identification and referral stratification, and they cannot replace standard sleep monitoring or the diagnostic pathway for OSA.

Monitoring of Abnormal Nocturnal Blood Pressure and Hemodynamic Changes

A total of 6 studies involved the monitoring of abnormal nocturnal blood pressure and respiratory event–related hemodynamic changes [25-27,32,33,35]. Wearable devices can not only monitor nocturnal blood pressure but also record the dynamic relationship between OSA-related hypoxic events and abnormal blood pressure fluctuations. Multiple studies found that the more severe the hypoxic events, the more pronounced the increase in nocturnal systolic blood pressure and blood pressure variability, suggesting that OSA-related hemodynamic changes are often event-triggered [27,33]. In a multicenter study of 330 patients with suspected sleep-disordered breathing, beat-to-beat PTT-estimated blood pressure was compared with fixed-interval PTT-estimated blood pressure. The results showed that the 2 methods were generally similar in mean blood pressure and blood pressure variability measures, but beat-to-beat monitoring was more sensitive in capturing maximum and minimum values [32]. Another study, using 24-hour ABPM as the reference, found that the PTT estimation method based on combined electrocardiography and photoplethysmography signals was only weakly correlated with nocturnal ABPM and tended to overestimate nocturnal blood pressure in patients with moderate to severe OSA, indicating that a gap in agreement with the gold-standard ABPM still remains [35].

Cardiovascular Risk Assessment and Remote Longitudinal Management

A total of 4 studies involved wearable devices in cardiovascular risk assessment and remote longitudinal management in this comorbidity [28,30,34,36]. In the assessment of hypertension and cardiovascular risk, 2 studies involving populations with suspected OSA showed that wearable devices could provide warning indicators such as the cardiac risk index, pulse wave attenuation index, and PPT. The studies reported that the cardiac risk index may be superior to conventional OSA severity indicators in identifying individuals at high cardiovascular risk, and that the derivation of these parameters may help identify high-risk cardiovascular phenotypes of OSA associated with increased arterial stiffness [30,34]. Another study found that heart sound intensity increased significantly after the termination of respiratory events and was closely associated with the degree of hypoxia and the concomitant fluctuations in blood pressure and heart rate, suggesting that patch-based devices may complement the assessment of the comorbid cardiovascular burden related to OSA from the perspective of cardiac mechanical load [28]. In terms of remote longitudinal management, a feasibility study including only 10 patients with OSA receiving CPAP treatment found that home monitoring using a multimodal wearable system could be used for the continuous observation of sleep, activity, and blood pressure changes and may support the remote management of intervention adherence and blood pressure fluctuations [36].

Technical Limitations and Challenges

Technical and translational obstacles persist in the application of wearables for managing comorbid OSA and hypertension. A total of 5 (38%) studies lacked adequate reference standard validation or used different assessment pathways, which limited the comparability of results across different devices and algorithms [25,29-32]. Selection biases attributable to voluntary enrollment, specific referral cohorts, and single-center designs were identified across multiple studies, restricting the generalizability of observations across diverse demographics and clinical outcomes. Methodological inconsistencies were observed in specific reports, such as the reliance on self-reported hypertension status (1/13, 8%) rather than clinical diagnostic verification [29]. Another study involved cardiovascular risk assessment derived from questionnaires and telephone-based follow-ups, a process susceptible to subjective bias compared to objective clinical measurements [24]. These observations support the use of wearables for risk stratification rather than as direct diagnostic substitutes. Data from Traiwannakij et al [35] indicated weak concordance between PTT-derived cuffless blood pressure monitoring and gold-standard ABPM; specifically, a systematic overestimation of nocturnal values was observed among patients with moderate to severe OSA. In addition, in remote or multimodal monitoring settings, obstacles such as reduced quality of nocturnal blood pressure data, difficulty in device calibration, and reliance on manual processing for cross-device data synchronization also remain, all of which limit the clinical application of wearable devices [36].

Discussion

Principal Findings

To our knowledge, this scoping review is the first review to systematically summarize the evidence on the application of wearable devices in the screening, monitoring, and management of OSA and hypertension from the perspective of this comorbidity. A total of 13 studies were included, involving 4 types of wearable devices, namely wrist-worn devices, fingertip contact devices, patch and single-lead devices, and multiparameter portable monitoring systems, covering a range of application scenarios from clinical validation to real home settings. Previous reviews have explored the feasibility of wearable devices in the monitoring of OSA or hypertension alone. From the perspective of comorbidity, this review further integrated the existing evidence on sleep-related respiratory signals and nocturnal hemodynamic information in the assessment of this comorbidity. The findings showed that wearable devices can simultaneously obtain sleep-related respiratory parameters and cardiovascular-related indicators,

XSL•FO
RenderX

and have shown preliminary application value in the identification of populations at high risk for OSA, the preliminary identification of abnormal nocturnal blood pressure, cardiovascular risk stratification, and remote follow-up. However, the current evidence remains limited and heterogeneous, particularly in terms of reference standards, diagnostic thresholds, and validation pathways. The following sections further discuss the clinical significance, technical boundaries, and future directions of current sleep and blood pressure monitoring.

Value of Sleep and Blood Pressure Monitoring From a Comorbidity Perspective

The clinical evaluation of OSA has long centered on AHI and the hypoxia and sleep architecture indicators obtained from conventional sleep monitoring. However, from the perspective of comorbid OSA and hypertension, sleep monitoring alone is insufficient to fully reflect the true clinical significance of OSA. AHI and REI mainly reflect the frequency of respiratory events and are less able to capture information more closely related to cardiovascular risk, such as nocturnal hypoxic burden, autonomic nervous responses, and abnormal circadian blood pressure patterns [6,8]. Comorbidity management does not simply mean adding more monitoring indicators, but rather incorporating the nocturnal cardiovascular burden underlying respiratory events into the overall assessment. Therefore, the focus of monitoring in comorbidity management should not be limited to AHI and ODI, but should also include indicators that better reflect the characteristics of OSA-related hypertension, such as hypoxic burden, respiratory effort burden, elevated nocturnal blood pressure, reduced nocturnal blood pressure dipping, abnormal morning blood pressure surge, and blood pressure variability. Traditional assessment mainly relies on patient recall or single in-hospital examinations and, in particular, lacks nocturnal data, making it difficult to effectively capture the dynamic relationship between OSA and hypertension [16]. The 24-hour ambulatory blood pressure and nocturnal blood pressure are closely associated with cardiovascular risk, and home morning blood pressure is closely associated with prognosis in patients with cardiovascular disease [37]. Traditional assessment may miss high-risk phenotypes such as masked hypertension, nondipper/riser blood pressure patterns, and morning blood pressure surge [5,38]. Reliance on a single sleep monitoring and routine outpatient assessment is therefore insufficient to fully identify the true risk of OSA-related hypertension. At the same time, OSA-related nocturnal hemodynamic changes may also manifest as autonomic dysfunction and increased blood pressure variability, suggesting that comorbidity management should further incorporate these related indicators [39]. Taken together, the assessment of comorbid OSA and hypertension should not remain limited to single examinations, but should further shift toward long-term dynamic monitoring.

The most direct clinical value of wearable devices lies in supporting long-term, longitudinal, and individualized monitoring in the management of comorbid OSA and hypertension. ABPM can monitor day-night blood pressure variability and detect masked hypertension, but it cannot assess daily and long-term blood pressure variability, including night-to-night variability; home blood pressure monitoring (HBPM) is suitable for tracking daily and long-term blood pressure trends, but measurements are mostly concentrated at fixed time points, such as in the morning and at night [40]. Single-night polysomnography is difficult to use to reflect OSA severity and its night-to-night variability, whereas ABPM and HBPM can provide blood pressure information, each still has limitations in long-term continuous dynamic tracking. Existing studies have shown through multinight sleep monitoring and repeated blood pressure measurements that the greater the night-to-night variability of OSA, the higher the risk of uncontrolled hypertension, and this association is independent of mean AHI [41]. Continuous monitoring may help more accurately identify hypertension risk and nocturnal vascular burden. For patients with OSA receiving CPAP treatment, wearable devices may support long-term adherence tracking and assist in evaluating posttreatment blood pressure improvement and cardiovascular benefit [36]. Consistent CPAP use correlates with lower rates of recurrent cardiovascular episodes, serving as an integral element of secondary prevention for the population with OSA [42]. For individuals presenting with both resistant hypertension and OSA, CPAP therapy facilitates reductions in 24-hour mean and nocturnal blood pressure and is more likely to correct nondipper/riser blood pressure patterns [43]. Therefore, the future value of wearable devices lies not only in improving the convenience of screening but also in providing continuous data support for treatment adherence tracking, blood pressure response assessment, and long-term cardiovascular risk stratification.

Clinical Limitations and Technical Boundaries of Wearable Monitoring

At present, the value of wearable technologies is reflected more in screening, dynamic monitoring, and long-term follow-up rather than in replacing standard diagnosis, and polysomnography remains the gold standard for OSA. The included studies found that wearable devices are feasible for identifying high-risk populations in large-scale real-world studies and primary care screening studies and are more suitable for initial screening and risk stratification [24,31]. Consistent with previous studies, most wearable devices lack complete electroencephalography channels, making full sleep staging difficult and limiting the accurate identification of microarousal-related hypopnea events, respiratory effort, and more complex respiratory events [44,45]. Signal fidelity frequently degrades due to postural shifts, ambient noise, and motion artifacts [28]. Proprietary algorithms differ across manufacturers, complicating their generalization to patients with complex comorbidities [31,45].

Continuous home-based blood pressure tracking via cuffless systems does not currently serve as a full replacement for standard ABPM or HBPM protocols [35]. Persistent technical limitations such as signal instability and mandatory calibration requirements, alongside sensitivity to postural changes and localized skin perfusion variations, affect the precision of these measurements [46,47]. The primary technical challenge involves the consistent recording of 24-hour blood pressure fluctuations rather than the accuracy of isolated measurement approximations [48]. Although certain cuffless tools demonstrate acceptable

XSL•FO
RenderX

performance during the day, their nocturnal estimates frequently lack precision and may fail to document transient pressure surges or circadian irregularities with sufficient reliability [27,48]. Inconsistencies in validation frameworks and algorithmic logic among manufacturers prevent data uniformity. Large-scale clinical datasets focused on blood pressure fluctuations specific to nocturnal OSA events also remain scarce. Therefore, cuffless blood pressure monitoring is currently better positioned as a tool for trend observation and risk warning than as a substitute for ABPM or HBPM.

Multimodal Sensing and Missing-Modality Handling

Recent studies have shown that intelligent OSA identification technologies are gradually shifting from single-signal analysis to multimodal sensing fusion. By integrating visual signals, snoring audio, body movement information, and multiple electrophysiological signals, related models have shown greater accuracy than single-modality approaches in OSA identification and stratification [49]. To address sensor detachment, motion artifacts, and modality missingness during long-term wearable monitoring, some studies have also begun to explore missing-modality reconstruction strategies to improve model robustness under incomplete input conditions. For example, the MIMAR-OSA model was able to maintain relatively high monitoring performance under partial signal loss through multimodal data integration and missing-modality reconstruction [50]. However, these advances are currently still limited to the automatic identification, severity stratification, and sleep staging of OSA itself, and are mainly focused on improvements in signal processing, feature extraction, and algorithm robustness. They have not yet been extended to the identification of abnormal nocturnal blood pressure or the long-term management of comorbid OSA and hypertension. Therefore, such studies may be more suitable as supporting technologies for future digital management of this comorbidity, but prospective clinical validation in the setting of comorbid OSA and hypertension is still lacking.

Exploration of a Sleep and Blood Pressure Closed-Loop Management Pathway

Existing studies have mainly involved digital care, wearable device monitoring, risk alert platforms, and long-term management, and a mature “sleep and blood pressure” closed-loop management model has not yet been established for comorbid OSA and hypertension [51,52]. Based on the included evidence, such a framework should at least involve 4 key components. At the input level, key information reflecting the comorbid state should be continuously acquired, including multinight sleep-related respiratory events, hypoxic burden, nocturnal blood pressure or blood pressure rhythm changes, HRV, and treatment adherence. At the output level, a more reasonable goal is not to directly generate an automated diagnosis, but to identify risk signals relevant to comorbidity management, such as suspected OSA-related nocturnal hypertension, persistent abnormal hypoxic burden, reduced CPAP adherence, or disordered nocturnal blood pressure rhythms. These risk alerts should primarily prompt clinicians to arrange polysomnography, supplement ABPM, or intensify remote follow-up, rather than delegating diagnostic or therapeutic decisions to autonomous algorithms. The safety boundaries of sleep and blood pressure management need to be clearly defined. Data outputs from wearable platforms operate as auxiliary indicators that do not replace gold-standard monitoring or bypass professional clinical appraisal. The use of wearables alongside CPAP remote monitoring is a feasible approach for home observation [36]. Future sleep and blood pressure closed-loop pathways are envisioned as auxiliary frameworks rather than fully autonomous diagnostic systems.

Barriers to Clinical Implementation

Practical obstacles persist in the clinical application of wearables for patients with concurrent OSA and hypertension. The existence of continuous data streams does not automatically translate into improved management efficiency. Clinical staff still have limited training in devices and related algorithms, and the interpretation of device data and verification of abnormal alerts may create an additional workload. In addition, there is still a lack of standardized clinical pathways, evaluation frameworks, and reimbursement systems applicable to the management of comorbid OSA and hypertension, making these technologies difficult to implement in routine clinical practice. Moreover, patient adherence, motion artifacts, and reduced nocturnal signal quality may further affect their application in clinical settings.

Therefore, in future clinical translation, it may be necessary to gradually establish a minimum necessary dataset for the management of comorbid OSA and hypertension, including core physiological indicators, device algorithm metadata, signal quality, and adherence information, in order to reduce the problem that outputs from different platforms are difficult to compare or cannot be incorporated into the clinical interpretation workflow [53,54]. On this basis, core indicators and quality control standards should also be standardized to improve the transferability and traceability of data across different systems. Responsibility for data quality should be further clarified. For example, synchronization, quality labeling, and abnormal alert generation could be completed at the device or platform level, while the clinical team would review the alert results, thereby reducing the data overload and additional interpretation burden caused by continuous monitoring [52,55].

Methodological Limitations

This scoping review has several methodological limitations. Only English-language studies were included, and gray literature and preprints were not included, which may have introduced language bias and publication bias. Second, because the current evidence on comorbid OSA and hypertension remains limited, this review included some studies conducted in single-disease populations that simultaneously assessed sleep- or respiratory-related indicators and blood pressure/cardiovascular indicators. Although this strategy expanded the scope of evidence covered, it also increased the heterogeneity across study populations, application scenarios, and outcome indicators. In addition, the number of included studies was limited, and some studies had small sample sizes, which restricted the robustness of the evidence. There was substantial heterogeneity across the included studies in terms of device types, reference standards, assessment pathways, and outcome definitions, and

XSL•FO
RenderX

no meta-analysis was performed in this review. Overall, the conclusions of this review are more suitable for informing subsequent validation studies and real-world research, and they should still be interpreted with caution in specific clinical implementations.

Future Directions

Based on the evidence gaps identified in this scoping review, future research should further shift from feasibility validation to clinically translational research in comorbidity management. Existing evidence mainly consists of observational studies and device validation studies and has focused primarily on whether devices can identify the risk of OSA and hypertension or record nocturnal physiological changes, with insufficient attention to the populations, indicators, and outcomes that are truly critical in comorbidity management. Future studies should more strongly emphasize large-sample, prospective, and multicenter research in clearly defined comorbid populations and should validate key nocturnal and long-term outcome indicators against reference standards such as polysomnography and ABPM. At the same time, the research focus should move beyond consistency with standard testing toward the effects of wearable devices on clinical outcomes such as blood pressure control, CPAP adherence, monitoring of nocturnal blood pressure patterns, and patient experience. Recent studies have also suggested that bedtime and circadian rhythm characteristics may be important additional factors for future comorbidity risk assessment [56]. Future comorbidity management should not be limited to the monitoring of apneic events alone, but should also incorporate sleep behavior, circadian rhythm characteristics, and hemodynamic changes into a comprehensive assessment. Finally, it will also be necessary to gradually establish core indicators, validation pathways, and reporting standards applicable to the management of comorbid OSA and hypertension, so that sleep and blood pressure data collected by wearable devices can better serve clinical care and be progressively integrated into standardized management pathways.

Conclusions

Wearable devices support the screening, nocturnal monitoring, and remote follow-up of comorbid OSA and hypertension. Specifically, they provide the continuous nocturnal data and longitudinal follow-up metrics that are often absent in conventional diagnostic workflows. Their clinical value is gradually extending from the identification of single sleep events to the continuous observation of comorbidity-related nocturnal cardiovascular burden. However, the existing evidence still mainly comes from observational studies and device validation studies, and there is marked heterogeneity across devices in terms of reference standards, diagnostic thresholds, and validation pathways. At the current stage, wearable devices are more suitable as auxiliary monitoring and risk warning tools in comorbidity management and cannot replace standard sleep testing or standard blood pressure monitoring.

Acknowledgments

The authors gratefully acknowledge the support of the Tianjin Science and Technology Program. During the preparation of this work, the authors used ChatGPT in order to improve language. The authors reviewed, revised, and verified all AI-assisted content and take full responsibility for the final manuscript. HL and JW are co-corresponding authors for this paper.

Funding

This work was supported by the Tianjin Science and Technology Program Funding Grant (24ZYJDSY00290) and the National Administration of Traditional Chinese Medicine's "Hundred-Thousand-Ten Thousand Talents" Inheritance and Innovation Project (Qihuang Project, 885022).

Data Availability

No new datasets were generated in this scoping review. All data analyzed were derived from published articles and are presented in the manuscript and multimedia appendices.

Authors' Contributions

Conceptualization: JW, HL, HW
Methodology: SF, HZ
Validation: YL, RS
Visualization: HW, AC
Writing – original draft: HW, AC
Writing – review and editing: HL, WZ
Supervision: JW, HL
Project administration: HL
Funding acquisition: HL
All authors reviewed the final manuscript and approved the submission.
HW and AC contributed equally to this work.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full search strategies for all databases.
[DOCX File, 22 KB - mhealth_v14i1e84506_app1.docx]

Multimedia Appendix 2
Characteristics of included studies.
[DOCX File, 35 KB - mhealth_v14i1e84506_app2.docx]

Multimedia Appendix 3
Device characteristics and key findings.
[DOCX File, 35 KB - mhealth_v14i1e84506_app3.docx]

Multimedia Appendix 4
Diagnostic or stratification thresholds.
[DOCX File, 32 KB - mhealth_v14i1e84506_app4.docx]

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 86 KB - mhealth_v14i1e84506_app5.docx]

References

1. Lévy P, Kohler M, McNicholas WT, et al. Obstructive sleep apnoea syndrome. Nat Rev Dis Primers 2015;1(15015):15015. [doi: 10.1038/nrdp.2015.15] [Medline: 27188535]
2. Benjafield AV, Ayas NT, Eastwood PR, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. Lancet Respir Med 2019;7(8):687-698. [doi: 10.1016/S2213-2600(19)30198-5] [Medline: 31300334]
3. Tokunou T, Ando SI. Recent advances in the management of secondary hypertension-obstructive sleep apnea. Hypertens Res 2020;43(12):1338-1343. [doi: 10.1038/s41440-020-0494-1] [Medline: 32546778]
4. Dominguez JE, Grotegut CA, Wright MC, Habib AS. Obstructive sleep apnea among gravidas with chronic hypertension compared to matched controls: a prospective cohort study. Anesth Analg 2023;136(2):205-214. [doi: 10.1213/ANE.0000000000006223] [Medline: 36355613]
5. Shiina K. Obstructive sleep apnea -related hypertension: a review of the literature and clinical management strategy. Hypertens Res 2024;47(11):3085-3098. [doi: 10.1038/s41440-024-01852-y] [Medline: 39210083]
6. Redline S, Azarbarzin A, Peker Y. Obstructive sleep apnoea heterogeneity and cardiovascular disease. Nat Rev Cardiol 2023;20(8):560-573. [doi: 10.1038/s41569-023-00846-6] [Medline: 36899115]
7. Ettehad D, Emdin CA, Kiran A, et al. Blood pressure lowering for prevention of cardiovascular disease and death: a systematic review and meta-analysis. Lancet 2016;387(10022):957-967. [doi: 10.1016/S0140-6736(15)01225-8] [Medline: 26724178]
8. Martinot JB, Le-Dong NN, Malhotra A, Pépin JL. Respiratory effort during sleep and prevalent hypertension in obstructive sleep apnoea. Eur Respir J 2023;61(3):2201486. [doi: 10.1183/13993003.01486-2022] [Medline: 36455958]
9. Kritikou I, Basta M, Vgontzas AN, et al. Sleep apnoea and the hypothalamic-pituitary-adrenal axis in men and women: effects of continuous positive airway pressure. Eur Respir J 2016;47(2):531-540. [doi: 10.1183/13993003.00319-2015] [Medline: 26541531]
10. Cowie MR, Linz D, Redline S, Somers VK, Simonds AK. Sleep disordered breathing and cardiovascular disease: JACC state-of-the-art review. J Am Coll Cardiol 2021;78(6):608-624. [doi: 10.1016/j.jacc.2021.05.048] [Medline: 34353537]
11. Imani MM, Sadeghi M, Khazaie H, Emami M, Sadeghi Bahmani D, Brand S. Evaluation of serum and plasma interleukin-6 levels in obstructive sleep apnea syndrome: a meta-analysis and meta-regression. Front Immunol 2020;11:1343. [doi: 10.3389/fimmu.2020.01343] [Medline: 32793188]
12. Popadic V, Brajkovic M, Klasnja S, et al. Correlation ofdyslipidemia and inflammation with obstructive sleep apnea severity. Front Pharmacol 2022;13:897279. [doi: 10.3389/fphar.2022.897279] [Medline: 35694268]
13. Huang CG, Lin WN, Hsin LJ, et al. Exploring the interplay of gut microbiota and systemic inflammation in pediatric obstructive sleep apnea syndrome and its impact on blood pressure status: a cross-sectional study. Int J Mol Sci 2024;25(24):13344. [doi: 10.3390/ijms252413344] [Medline: 39769109]

XSL•FO
RenderX

14. Kurnool S, McCowen KC, Bernstein NA, Malhotra A. Sleep apnea, obesity, and diabetes-an intertwined trio. Curr Diab Rep 2023;23(7):165-171. [doi: 10.1007/s11892-023-01510-6] [Medline: 37148488]
15. Giampá SQC, Lorenzi-Filho G, Drager LF. Obstructive sleep apnea and metabolic syndrome. Obesity (Silver Spring) 2023;31(4):900-911. [doi: 10.1002/oby.23679] [Medline: 36863747]
16. Williams GJ, Al-Baraikan A, Rademakers FE, et al. Wearable technology and the cardiovascular system: the future of patient assessment. Lancet Digit Health 2023;5(7):e467-e476. [doi: 10.1016/S2589-7500(23)00087-0] [Medline: 37391266]
17. Song Y, Min J, Gao W. Wearable and implantable electronics: moving toward precision therapy. ACS Nano 2019;13(11):12280-12286. [doi: 10.1021/acsnano.9b08323] [Medline: 31725255]
18. Alipour P, El-Aghil M, Foo A, Azizi Z. Leveraging mobile health and wearable technologies for the prevention and management of atherosclerotic cardiovascular disease. Curr Atheroscler Rep 2025;27(1):31. [doi: 10.1007/s11883-024-01272-w] [Medline: 39932603]
19. Sinou N, Sinou N, Koutroulakis S, Filippou D. The role of wearable devices in blood pressure monitoring and hypertension management: a systematic review. Cureus 2024;16(12):e75050. [doi: 10.7759/cureus.75050] [Medline: 39749057]
20. Hasanain M, Akram A, Kabir A, et al. Diagnostic value of wearable and contactless technologies for detecting obstructive sleep apnea: a systematic review and meta-analysis. Sleep Breath 2025;29(6):368. [doi: 10.1007/s11325-025-03544-4] [Medline: 41310031]
21. Flood JM, Cistulli PA, Tong BK, Mohammadieh A, de Chazal P, Stewart GM. Current and novel device-led approaches for targeted obstructive sleep apnea screening, diagnosis, treatment, and long-term management. Expert Rev Med Devices 2026;23(3):303-319. [doi: 10.1080/17434440.2026.2625362] [Medline: 41611622]
22. Osa-Sanchez A, Ramos-Martinez-de-Soria J, Mendez-Zorrilla A, Ruiz IO, Garcia-Zapirain B. Wearable sensors and artificial intelligence for sleep apnea detection: a systematic review. J Med Syst 2025;49(1):66. [doi: 10.1007/s10916-025-02199-8] [Medline: 40387964]
23. Westphaln KK, Regoeczi W, Masotya M, et al. From Arksey and O'Malley and beyond: customizations to enhance a team-based, mixed approach to scoping review methodology. MethodsX 2021;8:101375. [doi: 10.1016/j.mex.2021.101375] [Medline: 34430271]
24. Chen Y, Zhang H, Li J, Xu P, Guo Y, Xie L. Screening OSA in Chinese smart device consumers: a real-world arrhythmia-related study. Nat Sci Sleep 2025;17:663-676. [doi: 10.2147/NSS.S509097] [Medline: 40297265]
25. Cho JS, Ihm SH, Kim CJ, et al. Obstructive sleep apnea using Watch-PAT 200 is independently associated with an increase in morning blood pressure surge in never-treated hypertensive patients. J Clin Hypertens (Greenwich) 2015;17(9):675-681. [doi: 10.1111/jch.12581] [Medline: 26033308]
26. Correa CM, Gismondi RA, Cunha AR, Neves MF, Oigman W. Twenty-four hour blood pressure in obese patients with moderate-to-severe obstructive sleep apnea. Arq Bras Cardiol 2017;109(4):313-320. [doi: 10.5935/abc.20170130] [Medline: 28876375]
27. Huang K, Zhou Y, Huang Z, et al. Associations between nocturnal continuous blood pressure fluctuations and the characteristics of oxygen desaturation in patients with obstructive sleep apnea: a pilot study. Sleep Med 2021;84:1-7. [doi: 10.1016/j.sleep.2021.05.005] [Medline: 34090008]
28. Kabir MM, Assadi A, Saha S, et al. Unveiling the impact of respiratory event-related hypoxia on heart sound intensity during sleep using novel wearable technology. Nat Sci Sleep 2024;16(1623-36):1623-1636. [doi: 10.2147/NSS.S480687] [Medline: 39430234]
29. Orendain N, Bagsic SS, Ramos E, Pandit J, Owens RL, Jaiswal S. Digital biometrics in predicting risk for obstructive sleep apnea and hypertension: decentralized, prospective cohort study. J Med Internet Res 2025;27:e77641. [doi: 10.2196/77641] [Medline: 41340392]
30. Svedmyr S, Zou D, Sommermeyer D, et al. Vascular stiffness determined from a nocturnal digital pulse wave signal: association with sleep, sleep-disordered breathing, and hypertension. J Hypertens 2016;34(12):2427-2433. [doi: 10.1097/HJH.0000000000001111] [Medline: 27607460]
31. Yu X, Guan L, Su P, et al. Study on OSA screening and influencing factors in community-based elderly hypertensive patients based on single-lead wearable ECG devices. Sleep Breath 2024;28(6):2445-2456. [doi: 10.1007/s11325-024-03136-8] [Medline: 39207664]
32. Hoshide S, Yoshihisa A, Tsuchida F, et al. Pulse transit time-estimated blood pressure: a comparison of beat-to-beat and intermittent measurement. Hypertens Res 2022;45(6):1001-1007. [doi: 10.1038/s41440-022-00899-z]
33. Kuwabara M, Hamasaki H, Tomitani N, Shiga T, Kario K. Novel triggered nocturnal blood pressure monitoring for sleep apnea syndrome: distribution and reproducibility of hypoxia-triggered nocturnal blood pressure measurements. J Clin Hypertens (Greenwich) 2017;19(1):30-37. [doi: 10.1111/jch.12878] [Medline: 27411291]
34. Strassberger C, Zou D, Penzel T, et al. Beyond the AHI-pulse wave analysis during sleep for recognition of cardiovascular risk in sleep apnea patients. J Sleep Res 2021;30(6):e13364. [doi: 10.1111/jsr.13364] [Medline: 34032334]
35. Traiwannakij S, Tripipitsiriwat A, Kunanon S, et al. Comparative assessment of pulse transit time-derived blood pressure and ambulatory blood pressure monitoring in patients with obstructive sleep apnea. Sleep Breath 2026;30(1):38. [doi: 10.1007/s11325-026-03601-6] [Medline: 41670937]

XSL•FO
RenderX

36. Hediger-Parolini A, Charbonnier F, Grandin S, et al. Remote monitoring of patients with obstructive sleep apnea syndrome treated with CPAP: a feasibility study. Front Sleep 2025;4:1601389. [doi: 10.3389/frsle.2025.1601389] [Medline: 41425197]
37. Yang WY, Melgarejo JD, Thijs L, et al. Association of office and ambulatory blood pressure with mortality and cardiovascular outcomes. JAMA 2019;322(5):409-420. [doi: 10.1001/jama.2019.9811] [Medline: 31386134]
38. Crinion SJ, Kleinerova J, Kent B, et al. Non-dipping nocturnal blood pressure correlates with obstructive sleep apnoea severity in normotensive subjects and may reverse with therapy. ERJ Open Res 2021;7(3):00338-2021. [doi: 10.1183/23120541.00338-2021] [Medline: 34409095]
39. Tobushi T, Floras JS. Sleep apnea, autonomic disturbances, and blood pressure variability. Hypertension 2024;81(9):1837-1844. [doi: 10.1161/HYPERTENSIONAHA.124.20433] [Medline: 38957967]
40. Tomitani N, Hoshide S, Kario K. Novel blood pressure monitoring methods: perspectives for achieving “perfect 24-h blood pressure management”. Hypertens Res 2023;46(8):2051-2053. [doi: 10.1038/s41440-023-01329-4] [Medline: 37280261]
41. Lechat B, Loffler KA, Reynolds AC, et al. High night-to-night variability in sleep apnea severity is associated with uncontrolled hypertension. NPJ Digit Med 2023;6(1):57. [doi: 10.1038/s41746-023-00801-2] [Medline: 36991115]
42. Sánchez-de-la-Torre M, Gracia-Lavedan E, Benitez ID, et al. Adherence to CPAP treatment and the risk of recurrent cardiovascular events: a meta-analysis. JAMA 2023;330(13):1255-1265. [doi: 10.1001/jama.2023.17465] [Medline: 37787793]
43. Martínez-García MA, Capote F, Campos-Rodríguez F, et al. Effect of CPAP on blood pressure in patients with obstructive sleep apnea and resistant hypertension: the HIPARCO randomized clinical trial. JAMA 2013;310(22):2407-2415. [doi: 10.1001/jama.2013.281250] [Medline: 24327037]
44. Chiang AA, Jerkins E, Holfinger S, et al. OSA diagnosis goes wearable: are the latest devices ready to shine? J Clin Sleep Med 2024;20(11):1823-1838. [doi: 10.5664/jcsm.11290] [Medline: 39132687]
45. Zhou G, Zhao W, Zhang Y, et al. Comparison of OPPO Watch sleep analyzer and polysomnography for obstructive sleep apnea screening. Nat Sci Sleep 2024;16(125-41):125-141. [doi: 10.2147/NSS.S438065] [Medline: 38348055]
46. Walzel S, Sebestova H, Rafl-Huttova V, Rozanek M, Rafl J. Long-term accuracy and stability of blood pressure measurements from a smartwatch: prospective validation study. Digit Health 2026;12:20552076261415923. [doi: 10.1177/20552076261415923] [Medline: 41602947]
47. Lunardi M, Muhammad F, Shahzad A, et al. Performance of wearable watch-type home blood pressure measurement devices in a real-world clinical sample. Clin Res Cardiol 2024;113(10):1393-1404. [doi: 10.1007/s00392-023-02353-7] [Medline: 38112747]
48. Schutte AE. Wearable cuffless blood pressure tracking: when will they be good enough? J Hum Hypertens 2024;38(9):669-672. [doi: 10.1038/s41371-024-00932-3] [Medline: 38997475]
49. Qiu X, Wang C, Li B, et al. An audio-semantic multimodal model for automatic obstructive sleep apnea-hypopnea syndrome classification via multi-feature analysis of snoring sounds. Front Neurosci 2024;18:1336307. [doi: 10.3389/fnins.2024.1336307] [Medline: 38800571]
50. Qiu X, Wei Y, Tan X, et al. MIMAR-OSA: enhancing obstructive sleep apnea diagnosis through multimodal data integration and missing modality reconstruction. Pattern Recognit 2026;169:111917. [doi: 10.1016/j.patcog.2025.111917]
51. Haverinen J, Harju T, Mikkonen H, Liljamo P, Turpeinen M, Reponen J. Digital care pathway for patients with sleep apnea in specialized care: mixed methods study. JMIR Hum Factors 2024;11:e47809. [doi: 10.2196/47809] [Medline: 38386368]
52. Verbraecken J, Amodio E, Basoglu OK, et al. European Respiratory Society statement on advanced telemedicine for obstructive sleep apnoea (e-sleep). Eur Respir J 2025;66(5):2500557. [doi: 10.1183/13993003.00557-2025] [Medline: 40907996]
53. Baumert M, Cowie MR, Redline S, et al. Sleep characterization with smart wearable devices: a call for standardization and consensus recommendations. Sleep 2022;45(12):zsac183. [doi: 10.1093/sleep/zsac183] [Medline: 35913733]
54. Tabari P, Costagliola G, De Rosa M, Boeker M. State-of-the-art Fast Healthcare Interoperability Resources (FHIR)-based data model and structure implementations: systematic scoping review. JMIR Med Inform 2024;12:e58445. [doi: 10.2196/58445] [Medline: 39316433]
55. Claggett J, Petter S, Joshi A, Ponzio T, Kirkendall E. An infrastructure framework for remote patient monitoring interventions and research. J Med Internet Res 2024;26:e51234. [doi: 10.2196/51234] [Medline: 38815263]
56. Zhou R, Shan G, Hu Y, et al. Association of bedtime with co-morbid hypertension in OSA: a community-based study. Sleep Breath 2026;30(2):96. [doi: 10.1007/s11325-026-03579-1] [Medline: 41854932]

Abbreviations

ABPM: ambulatory blood pressure monitoring
AHI: apnea-hypopnea index
CPAP: continuous positive airway pressure
HBPM: home blood pressure monitoring
HRV: heart rate variability
HSAT: home sleep apnea test

XSL•FO
RenderX

ODI: oxygen desaturation index
OSA: obstructive sleep apnea
PAT: peripheral arterial tonometry
PPT: pulse propagation time
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
PTT: pulse transit time
REI: respiratory event index
SpO_2: peripheral oxygen saturation

Edited by A Stone, L Buis; submitted 20.Sep.2025; peer-reviewed by M Lamanna, X Tan; revised version received 27.Jun.2026; accepted 29.Jun.2026; published 31.Jul.2026.

Please cite as:
Wang H, Cheng A, Li Y, Zhang H, Fan S, Shen R, Zhang W, Li H, Wang J
Wearable Devices for Monitoring and Management of Comorbid Obstructive Sleep Apnea and Hypertension: Scoping Review
JMIR Mhealth Uhealth 2026;14:e84506
URL: https://mhealth.jmir.org/2026/1/e84506
doi:10.2196/84506

© Hongyi Wang, Anran Cheng, Yusheng Li, Haining Zhang, Shun Fan, Rong Shen, Wei Zhang, Huanan Li, Jingui Wang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 31.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effectiveness of Mobile Health Interventions in Pediatric Cancer: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Huilu Yao[1,2], BSc; Yiting Wen[3], BSc; Hongxiu Wang[4], BSc; Ying Xiao[3], BSc; Meiling Deng[5], BSc; Wei Yang[6], BSc; Yuqin Li[7], BSc; Xiaorong Mao[2], PhD

[1]College of Nursing, Chengdu University of Traditional Chinese Medicine, Chengdu, China

[2]Department of Nursing, Sichuan Provincial People's Hospital, School of Medicine, University of Electronic Science and Technology of China, No. 32, West Section 2, 1st Ring Road, Qingyang District, Chengdu, Sichuan Province, China

[3]School of Medicine, University of Electronic Science and Technology of China, Chengdu, China

[4]College of Nursing, North Sichuan Medical College, Nanchong, China

[5]Department of Neurosurgery, Sichuan Provincial People's Hospital, School of Medicine, University of Electronic Science and Technology of China, Chengdu, China

[6]Department of Thoracic Surgery, Sichuan Provincial People's Hospital, School of Medicine, University of Electronic Science and Technology of China, Chengdu, China

[7]Department of Cancer Center, Sichuan Provincial People's Hospital, School of Medicine, University of Electronic Science and Technology of China, Chengdu, China

Corresponding Author:
Xiaorong Mao, PhD
Department of Nursing, Sichuan Provincial People's Hospital, School of Medicine, University of Electronic Science and Technology of China, No. 32, West Section 2, 1st Ring Road, Qingyang District, Chengdu, Sichuan Province, China

Abstract

Background: Cancer poses a significant threat to children's health, and mobile health (mHealth) is emerging as a key tool for remote disease management, health education, and follow-up. However, evidence of its effectiveness remains limited.

Objective: This study aimed to summarize the effects of mHealth interventions for pediatric cancer compared with usual care, providing evidence-based support for optimizing intervention models and improving patient outcomes.

Methods: A systematic search of 14 databases identified randomized controlled trials (RCTs) on mHealth apps for pediatric patients with cancer from inception to August 1, 2025. Two reviewers independently screened studies, extracted data, assessed bias risk, and graded evidence quality. The meta-analysis was conducted using RevMan 5.4 and Stata 15.

Results: A total of 24 RCTs involving 2645 patients were included. This review found that mHealth interventions significantly reduced infection rates (odds ratio [OR] 0.25, 95% CI 0.10-0.60; *P*=.002) and the overall incidence of peripherally inserted central catheter (PICC) complications (OR 0.16, 95% CI 0.10-0.24; *P*<.001), while improving quality of life (standardized mean difference [SMD] 1.34, 95% CI 0.13-2.55; *P*=.03), self-management ability (SMD 6.39, 95% CI 1.26-11.53; *P*=.01), and treatment adherence (OR 2.83, 95% CI 1.41-5.66; *P*=.003). However, mHealth interventions had no significant effect on PICC catheter displacement (OR 0.44, 95% CI 0.15-1.29; *P*=.13) or health knowledge (SMD 4.44, 95% CI −2.40 to 11.29; *P*=.20). Further high-quality studies are needed to verify their impact in these areas. The intervention components covered 9 behavior change techniques: goals and planning, feedback and monitoring, social support, shaping knowledge, repetition and substitution, reward and threat, comparison of outcomes, natural consequences, and regulation.

Conclusions: This systematic review and meta-analysis synthesized evidence from RCTs. The findings support the use of mHealth to reduce infections and PICC-related complications among pediatric patients with cancer while improving quality of life, self-management capabilities, and treatment adherence. These results underscore the importance of incorporating mHealth strategies into pediatric cancer care and guide the development and enhancement of future mHealth interventions.

Trial Registration: PROSPERO CRD420251108938; https://www.crd.york.ac.uk/PROSPERO/view/CRD420251108938

(JMIR Mhealth Uhealth 2026;14:e86836) doi:10.2196/86836

KEYWORDS

pediatric cancer; mobile health; adherence; quality of life; meta-analysis

XSL•FO
RenderX

Introduction

Cancer remains a major threat to the health and survival of children [1]. While advances in medical technology have improved survival rates [2], children living with cancer continue to face considerable challenges. First, the disease itself compromises immune function and causes neutropenia, increasing susceptibility to bacterial, fungal, and viral infections [3], which is one of the leading causes of mortality in this population [4]. Second, following diagnosis, pediatric patients with cancer typically undergo intensive treatments. Chemotherapy, used in over 95% of pediatric cancer cases [5], is associated with adverse effects such as pain, vomiting, and fatigue, which substantially diminish quality of life (QoL) [6]. Additionally, pediatric patients typically require the placement and maintenance of a peripherally inserted central catheter (PICC) during chemotherapy, yet PICC-related complications—including bloodstream infections, venous thrombosis, and mechanical issues—remain frequent. These complications can delay antitumor therapy, prolong hospitalization, and increase mortality risk [7]. The complexity of disease management, coupled with limited access to routine care information [8] and frequent care transitions between hospital and home [9], places a heavy burden on pediatric patients and their families. Therefore, timely and effective health education and self-management support are crucial.

In this context, mobile health (mHealth) has emerged as a key tool in health care interventions due to its convenience, efficiency, accessibility, and low cost. mHealth has been shown to overcome barriers related to limited human resources and distance [10]. The World Health Organization defines mHealth as the use of mobile information technology devices, such as mobile phones, tablets, wearable devices, and wireless pedometers, in medical and public health practices [11]. It enables remote services, including symptom monitoring, health education, vital sign tracking, alerts, and medication guidance [12,13].

Previous reviews have examined mHealth apps in pediatric cancer care. Ramsey et al [14] found that mHealth can reduce pain and improve treatment adherence, though findings regarding health-related QoL were inconsistent. González-Díaz et al [15] observed that mobile apps effectively reduced the incidence and severity of symptoms, such as pain and nausea, with high usability and acceptance among patients and caregivers. Similarly, Upreti et al [16] noted that pain monitoring apps reduced pain intensity and decreased moderate-to-severe pain episodes, demonstrating good usability and satisfaction. Zhu et al [17] concluded that mHealth enables multidimensional pain assessment (intensity, frequency, location, and associated symptoms) with a favorable user experience. Delemere et al [18] found that incorporating high-frequency behavior change techniques (BCTs), such as feedback, monitoring, and social support, in pediatric cancer-related mHealth apps enhances treatment adherence and caregiving coping abilities. Mehdizadeh et al [19] further observed that smartphone apps improve symptom reporting adherence, health care communication, and medication compliance, with good user acceptance.

However, existing reviews have several limitations. First, most analyses included primarily pilot or small-sample exploratory studies, with no comprehensive synthesis of randomized controlled trials (RCTs) on mHealth apps for pediatric patients with cancer. Second, outcomes have often focused on mHealth device usability and pain management, lacking thorough evidence on clinical effectiveness in areas such as infection prevention and PICC care [18]. Third, some reviews combined children and adolescents in their analyses, resulting in high sample heterogeneity. These reviews also failed to distinguish between mHealth users (children or caregivers) and treatment stages (treatment or recovery phases) [20], making it difficult to rule out confounding factors influencing outcomes [15]. This limits the specificity and applicability of the evidence. Fourth, prior reviews have largely been restricted to English-language databases and evidence from high-income countries, offering limited insight into mHealth effectiveness in low-income and middle-income regions [15,19]. Furthermore, many existing mHealth platforms lack clinical and evidence-based input in their design, underscoring the need for high-quality evidence to guide development [21].

To address these limitations, this study conducted a systematic search of Chinese and English databases, aiming to synthesize the effects of mHealth interventions in children aged 0‐18 years undergoing cancer treatment by using meta-analyses based on RCTs. The primary outcomes are infection incidence, QoL, and PICC-related complications. Secondary outcomes include health knowledge, self-management ability, and treatment adherence. The BCT taxonomy was used to categorize intervention components across studies, thereby providing more practice-oriented evidence for mHealth apps in pediatric cancer care.

Methods

Ethical Considerations

This study did not involve the collection of primary data from human participants. All data were derived from previously published sources and are appropriately cited. Therefore, ethical approval was not required.

Study Registration

This review was registered with PROSPERO (registration number: CRD420251108938). During the review process, this study strictly adhered to the registered protocol. In the initial search strategy, broad telehealth-related terms were applied to comprehensively assess the multidimensional effects of telehealth interventions on pediatric patients with cancer. However, during the systematic search and screening process, it was observed that most (over 90%) of the studies meeting the original inclusion criteria used mHealth-based interventions—such as mobile apps, web-based platforms, or smart devices—rather than traditional telehealth approaches (eg, telephone follow-ups or video consultations) [22,23]. Meanwhile, a sufficient number of RCTs were identified. To enhance internal homogeneity and ensure that the review provides precise, high-quality evidence with clear practical implications, the scope of this review was refined to focus exclusively on RCTs of mHealth interventions. The registration

XSL•FO
RenderX

record has been updated accordingly, and the systematic review and meta-analysis were conducted in strict accordance with the most recent registration record [24].

Search Strategies

The systematic review and meta-analysis were conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 statement [25] (Checklist 1). A comprehensive search was conducted across PubMed, Web of Science, Cochrane Library, Embase, CINAHL, Scopus, ScienceDirect, ProQuest, PsycINFO, OVID, Chinese National Knowledge Infrastructure, WanFang, Weip Database, and China Biology Medicine database, with each searched from inception to August 1, 2025. The search terms combined both Medical Subject Headings (MeSH) terms and free-text keywords: (cancer or oncology or tumor or neoplasm or leukemia) and (child or pediatric) and (mHealth or internet-based intervention or software application). The detailed search strategy is provided in Multimedia Appendix 1.

Study Selection and Data Extraction

Two researchers (HY and YX) first removed duplicates using EndNote X21 (Clarivate Analytics) and then screened the studies in 2 steps: first by title and abstract, and then by full text. Any discrepancies that arose during the screening process were resolved through discussion between the 2 researchers, or, if necessary, by consulting a third researcher (XM). The literature screening followed the participants, intervention, comparison, outcomes, and study design [26]. Inclusion criteria were as follows:

1. Participants: children aged 0 - 18 years diagnosed with cancer
2. Intervention: mobile-based or wireless-based tools such as apps, websites, wearables, or social media
3. Comparison: usual care (paper-based education or face-to-face monitoring)
4. Outcomes: infection incidence, QoL, PICC-related complications, health knowledge, self-management ability, and treatment adherence
5. Study design: RCTs

Exclusion criteria included (1) reviews, abstracts, theses, systematic reviews, meta-analyses, and case reports; (2) unavailable full texts; (3) interventions targeting only parents; (4) publications not in Chinese or English; and (5) interventions limited to calls or videos.

Two researchers (HW and YW) independently extracted the following data using a standardized form, including the first author, publication year, country, sample size, age, sex, cancer type, intervention duration and setting, mHealth platform, BCT clusters, intervention control groups, and outcome measurement tools.

Quality Assessment

Two researchers (HY and YW) independently assessed the methodological quality of the included studies using the revised Cochrane Risk of Bias 2 tool, evaluating selection, performance, attrition, detection, and reporting biases, each rated as low, some concerns, or high [27]. The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) framework was applied to appraise the quality of evidence using GRADEpro. Results were categorized into 4 levels (high, moderate, low, and very low) according to predefined criteria. The assessment incorporated study design, risk of bias, inconsistency in population, heterogeneity of findings, statistical precision of effect estimates, and publication bias [28]. Discrepancies were resolved through discussion or, if needed, consultation with a third researcher (XM).

Statistical Analysis

This study adhered to the Cochrane Handbook [29] for data synthesis and analysis, using RevMan 5.4 (Nordic Cochrane Center) and Stata 15.0 (StataCorp) for statistical procedures. Effect measures included standardized mean differences (SMD, 95% CI) and odds ratios (OR, 95% CI). Heterogeneity was assessed using I^2 and P values; when I^2 >50% and P<.1 (indicating substantial heterogeneity), a random-effects model was used; otherwise, a fixed-effects model was used. Sensitivity analyses were conducted using a stepwise exclusion, while subgroup analyses and meta-regression were used to explore heterogeneity. Publication bias was assessed with funnel plots and Egger's test. A P<.05 was considered statistically significant.

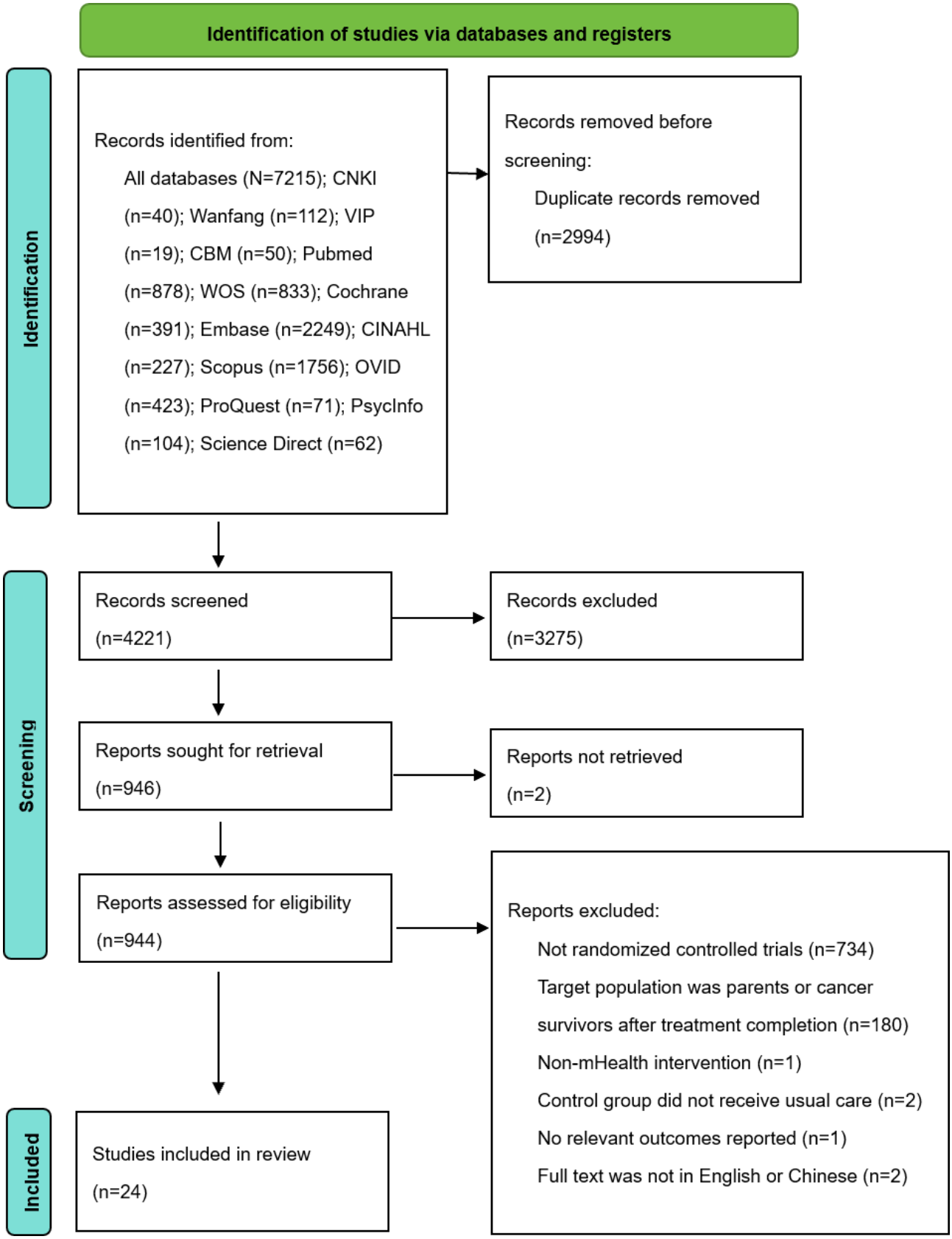
Results

Search Results

A total of 7215 studies were identified, with 24 meeting the inclusion criteria for the systematic review and meta-analysis (Figure 1).

XSL•FO
RenderX

Figure 1. Study flowchart. mHealth: mobile health.



Descriptions of the Included Studies

This study included 24 RCTs published between 2017 and 2025, the main characteristics of which are summarized in Table 1. Most of the studies (19) were conducted in China [30-48], while one each was conducted in America [49], Canada [50], Iran [51], the Netherlands [52], and Turkey [53]. These studies involved a total of 2645 children with cancer, with sample sizes ranging from 18 to 444 participants. The cancer types included hematologic malignancies, brain tumors, bone tumors,

XSL•FO
RenderX

neuroblastoma, sarcomas, and other solid tumors. Participants were aged 0 - 18 years. The intervention durations varied from 7 days to 2 years. The intervention settings included home-based (9 studies [30-33,35,37,41,51,52]), hospital-based (3 studies [34,50,53]), and combined home and hospital (12 studies [36,38-40,42-49]).

XSL•FO
RenderX

Table . Characteristics of the included studies.

Author, year, country	Sample size (T[a],C[b]), n	Age[c] (y)	Male/ female (T,C)	Cancer type	Duration (wk)	Setting	mHealth platform	BCT[d] clusters	Intervention control groups	Outcome measurement tool
Zhong et al [30], 2025, China	70, 71	T=4.73 (2.59); C=5.07 (3.08)	35/35, 45/26	Leukemia	12	Home	WeChat	Goals and planning[e], feedback and monitoring[f], social support[g], natural consequences[h], reward and threat[i]	Usual care[j]	SEAMS[k]; missed or incorrect medication rate; unplanned hospitalization rate; platform usage evaluation
Simon et al [52], 2024, Netherlands	79, 79	T=7.5 (5.1); C=7.5 (5.4)	41/38, 38/41	Hematologic malignancies, neurological tumor, solid tumors	4	Home	KLIK Pijnmonitor software	Feedback and monitoring, shaping knowledge[l], natural consequences, social support	Usual care	NRS-11[m]; BPI-SF[n]; ET[o]; app acceptance; pain incidence
Shahri et al [51], 2024, Iran	33, 33	T=15.55 (1.85); C=15.21 (1.79)	18/15, 12/21	Leukemia, lymphoma, glioma, osteosarcoma	4	Home	Website + SMS text messaging	Shaping knowledge, social support, natural consequences	Usual care	Gastrointestinal symptom incidence; gastrointestinal symptom self-management score
Zhao et al [31], 2023, China	40, 40	T=8.25 (2.02); C=8.30 (1.68)	20/21, 25/14	Leukemia	4	Home	WeChat	Goals and planning, feedback and monitoring, shaping knowledge, social support, comparison of outcomes[p], natural consequences	Usual care	PICC[q]-related complications; catheter care status
Lv et al [32], 2023, China	50, 49	T=8.45 (2.52); C=8.77 (2.64)	29/21, 26/23	Brain tumors	12	Home	WeChat	Social support, shaping knowledge, feedback and monitoring, reward and threat	Usual care	KAP[r] of catheter care; PICC-related complications; PedsQL 4.0[s]

XSL•FO
RenderX

Author, year, country	Sample size (T[a],C[b]), n	Age[c] (y)	Male/ female (T,C)	Cancer type	Duration (wk)	Setting	mHealth platform	BCT[d] clusters	Intervention control groups	Outcome measurement tool
Hu et al [33], 2022, China	45, 44	T=7.32 (2.02); C=7.49 (2.01)	24/21, 24/20	Leukemia	12	Home	WeChat	Social support, shaping knowledge, reward and threat, feedback and monitoring	Usual care	SCAS-C[t]; ESCA[u]; PICC-related complications
Semerci et al [53], 2022, Turkey	26, 31	T and C: 8 - 18	15/11, 17/14	Bone tumors, leukemia, lymphoma, brain tumors	1	Hospital	"5inD" software	Repetition and substitution[v], feedback and monitoring	Usual care	ARINVc[w]; ARINVp[x]
Breakey et al [50], 2022, Canada	39, 42	T and C: 15.2 (1.7)	26/10, 17/23	Leukemia, solid tumors, lymphoma, brain tumors	12	Hospital	Website	Shaping knowledge, social support, feedback and monitoring, goals and planning	Usual care	Satisfaction questionnaire ; REDCap[y] logs; website engagement; AWC[z] and caregiver self-reports
Ren et al [34], 2021, China	118, 118	T and C: 0 - 18	33/26, 37/22	Brain tumors	≥3	Hospital	Medical software + WeChat + Hospital system	Shaping knowledge, goals and planning, social support, feedback and monitoring, reward and threat, natural consequences	Usual care	PPUS-CC[aa]; RHDS[ab]
Xie [35], 2021, China	30, 30	T=5.94 (1.81); C=5.96 (1.76)	20/10, 18/12	Wilms tumor	8	Home	WeChat	Shaping knowledge, social support, goals and planning, feedback and monitoring	Usual care	SF-36[ac]; PSQI[ad]; Self-developed Treatment Adherence/ Satisfaction Scale
Jiang [36], 2021, China	40, 40	T=3.47 (1.27); C=3.52 (1.64)	23/17, 21/19	Leukemia	144	Hospital + Home	WeChat	Shaping knowledge, social support, goals and planning, repetition and substitution, regulation[ae], natural consequences	Usual care	Self-developed Resistance/ Satisfaction Scale; CMFS[af]

XSL•FO
RenderX

Author, year, country	Sample size (T[a],C[b]), n	Age[c] (y)	Male/ female (T,C)	Cancer type	Duration (wk)	Setting	mHealth platform	BCT[d] clusters	Intervention control groups	Outcome measurement tool
Ye et al [37], 2021, China	9, 9	T=6.79 (1.15); C=6.83 (1.13)	5/4, 6/3	Leukemia	40	Home	WeChat	Shaping knowledge, social support	Usual care	Adherence to prescribed treatment
Li et al [38], 2020, China	30, 30	T=4.30 (1.75); C=4.28 (1.86)	23/17, 24/16	Malignant tumor	24	Hospital + Home	WeChat	Shaping knowledge, social support, goals and planning, feedback and monitoring	Usual care	PedsQL 4.0; PedsQL 3.0; nursing satisfaction questionnaire
Song et al [40], 2020, China	36, 36	T=4.39 (2.14); C=4.38 (2.13)	20/16, 21/15	Leukemia	__[ag]	Hospital + Home	WeChat＋ Telephone	Shaping knowledge, social support, goals and planning, natural consequences	Usual care	PICC-related complications
Wang et al [39], 2020, China	30, 30	T and C: 0 - 9	17/13, 18/12	Leukemia	48	Hospital + Home	WeChat	Shaping knowledge, social support, goals and planning, feedback and monitoring, comparison of outcomes, natural consequences	Usual care	Complication incidence; treatment adherence questionnaire
Yao et al [41], 2020, China	65, 65	T=11.40 (4.65); C=10.37 (3.97)	33/29, 35/25	Leukemia	12	Home	Medical software	Shaping knowledge, feedback and monitoring, social support, natural consequences	Usual care	PICC-related complications rate; catheter care omission rate; catheter maintenance time and cost; PICC-specific nursing service satisfaction scale; Children's PICC Self-Management Ability Scale

XSL•FO
RenderX

Author, year, country	Sample size (T[a],C[b]), n	Age[c] (y)	Male/ female (T,C)	Cancer type	Duration (wk)	Setting	mHealth platform	BCT[d] clusters	Intervention control groups	Outcome measurement tool
Bhatia et al [49], 2020, United States	230, 214	T=8.6 (5.6 - 14.3); C=7.5 (5.3 - 14.0)	154/76, 148/66	Leukemia	20	Hospital + Home	Website + SMS text messaging + MEMS	Shaping knowledge, social support, repetition and substitution, feedback and monitoring, natural consequences	Usual care	MEMS[ah] usage records to calculate adherence and safety
Huang et al [42], 2019, China	17, 15	T=6.5 (3.5); C=6.3 (3.7)	10/7, 7/8	Leukemia	144	Hospital + Home	WeChat	Social support, goals and planning, feedback and monitoring	Usual care	PedsQL 4.0; caregiver self-management ability score
Zhong [43], 2018, China	36, 36	T=6.3 (3.1); C=6.8 (2.6)	18/18, 20/16	Leukemia	3	Hospital + Home	WeChat	Shaping knowledge, social support, goals and planning, feedback and monitoring, natural consequences, social support, regulation	Usual care	PICC Knowledge Assessment Questionnaire; PICC Self-Management Ability Scale; PICC Patient Satisfaction Questionnaire; PICC-related complications rate
Wen [44], 2018, China	52, 50	T=8.8 (1.5); C=9.3 (1.8)	29/23, 26/24	Leukemia	12	Hospital + Home	WeChat	Shaping knowledge, social support, regulation, natural consequences	Usual care	HAD[ai]; PedsQL™ 4.0; time of PICC catheterization and the incidence of complications
Lu et al [45], 2018, China	23, 23	T=5.1 (1.43); C=5.9 (1.85)	12/11, 13/10	Wilms tumor	96	Hospital + Home	WeChat	Shaping knowledge, feedback and monitoring, goals and planning, natural consequences	Usual care	Adherence to prescribed treatment; Quality of Life Scale

XSL•FO
RenderX

Author, year, country	Sample size (T[a],C[b]), n	Age[c] (y)	Male/ female (T,C)	Cancer type	Duration (wk)	Setting	mHealth platform	BCT[d] clusters	Intervention control groups	Outcome measurement tool
Ding et al [46], 2017, China	70, 70	—	—	Wilms tumor, neuroblastoma, hepatoblastoma, osteosarcoma	4	Hospital + Home	WeChat	Shaping knowledge, feedback and monitoring, social support, goals and planning, natural consequences	Usual care	PICC Knowledge Questionnaire; catheter care adherence; self-developed patient satisfaction questionnaire; PICC-related complications rate
Qin et al [47], 2017, China	70, 70	T=5.36 (1.58); C=5.62 (1.60)	41/29, 36/34	Lymphoma, rhabdomyosarcoma, neuroblastoma, Wilms tumor	48	Hospital + Home	WeChat	Shaping knowledge, social support, feedback and monitoring, comparison of outcomes, natural consequences	Usual care	PICC-related complications rate

XSL•FO
RenderX

Author, year, country	Sample size (T[a],C[b]), n	Age[c] (y)	Male/ female (T,C)	Cancer type	Duration (wk)	Setting	mHealth platform	BCT[d] clusters	Intervention control groups	Outcome measurement tool
Xiang et al [48], 2017, China	56, 56	T=5.3 (3.1); C=5.6 (3.5)	60, 52	Leukemia	—	Hospital + Home	WeChat + Telephone	Social support, feedback and monitoring, shaping knowledge, natural consequences	Usual care	PICC-related complications rate; self-developed patient satisfaction questionnaire

[a]T: experimental group.

[b]C: control group.

[c]Age is given as mean (SD), range values, and median (IQR).

[d]BCT: behavior change technique.

[e]*Goals and planning:* Personalized planning of medication, exercise, follow-up visits, disease management, and personal development.

[f]*Feedback and monitoring:* Recording medication information, disease management, assessments, and test results with data visualization, generating corresponding guidance and recommendations; providing regular reminders, alerts for critical values or risky behaviors; and enabling video consultations or scheduling in-person follow-ups.

[g]*Social support:* Establishing 2-way communication channels with other families and medical staff or offering assistance and reimbursement consultations.

[h]*Natural consequences:* Enabling individuals to understand or experience the natural consequences of their behaviors.

[i]*Reward and threat:* Using virtual tasks or gamified elements to assess the mastery of self-care knowledge and providing point-based rewards or badge recognition.

[j]*Usual care:* During hospitalization, treatments are administered according to medical orders, with face-to-face monitoring of vital signs and rehabilitation care, while maintaining a comfortable ward environment. Before discharge, guidance is provided via verbal instructions, printed materials, or health education videos, covering medication, diet, and catheter care. At home, care and record-keeping are performed by family members, with medical staff monitoring the condition and addressing questions through regular phone calls or outpatient follow-ups. This approach does not involve mobile apps or online platform support.

[k]SEAMS: Self-Efficacy for Appropriate Medication Use Scale.

[l]*Shaping knowledge:* Providing staged information delivery, search, and storage functions related to disease, symptoms, catheter care, and medication management.

[m]NRS-11: Numerical Rating Scale–11.

[n]BPI-SF: Brief Pain Inventory–short form.

[o]ET: emotion thermometer.

[p]*Comparison of outcomes:* Promoting belief enhancement and behavior change through social comparison and modeling by sharing successful cases, organizing parent communication for mutual learning, and demonstrating practical experiences and caregiving skills.

[q]PICC: peripherally inserted central catheter.

[r]KAP: knowledge, attitude, and practice.

[s]PedsQL 4.0: Pediatric Quality of Life Inventory 4.0.

[t]SCAS-C: Spence Children's Anxiety Scale—child version.

[u]ESCA: Exercise of Self-Care Agency Scale.

[v]*Repetition and substitution:* Conducting online cognitive-behavioral interventions, professional psychological counseling, or other complementary medical interventions.

[w]ARINVc: Children's Version of the Adapted Rhodes Index of Nausea and Vomiting.

[x]ARINVp: Parent Version of the Adapted Rhodes Index of Nausea and Vomiting.

[y]REDCap: Research Electronic Data Capture.

[z]AWC: adolescents with cancer.

[aa]PPUS-CC: Parent Perception of Uncertainty Scale–Child Chronic Illness version.

[ab]RHDS: Readiness for Hospital Discharge Scale.

[ac]SF-36: 36-Item Short Form Survey.

[ad]PSQI: Pittsburgh Sleep Quality Index.

[ae]*Regulation:* Helping individuals reduce negative emotions, relieve tension and stress, and maintain a more stable and positive psychological and physiological state through interventions such as stress management and music therapy.

[af]CMFS: Children's Medical Fear Scale.

[ag]Not available.

[ah]MEMS: Medication Event Monitoring System.

[ai]HAD: Hospital Anxiety and Depression Scale.

XSL•FO
RenderX

Regarding intervention platforms used for the experimental groups, 15 studies [30-33,35-39,42-47] used WeChat (a widely used mobile-based social media and communication app in China that supports messaging, group communication, and content dissemination[54]), 1 used a website [50], 3 used professional medical apps [41,52,53], and 5 used mixed platforms [34,40,48,49,51]. All interventions were conducted under the guidance of medical staff. Figure 2 shows the BCTs used in each study. A total of 13 studies involved goals and planning [30,31,34-36,38-40,42,43,45,46,50]; 19 studies included feedback and monitoring [30-35,38,39,41-43,45-50,52,53]; 21 provided social support [30-44,46-48,50-52]; 21 included shaping knowledge; 3 involved repetition and substitution [36,49,53]; 4 included reward and threat [30,32-34]; 3 conducted comparison of outcomes [31,39,47]; 16 incorporated natural consequences [30,31,34,36,39-41,43-49,51,52]; 2 involved regulation [36,44]. A total of 3 reported security measures such as password login and patient-specific data verification [41,50,51]. The control groups received usual care.

Figure 2. The behavior change techniques used in each study [30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51-53].

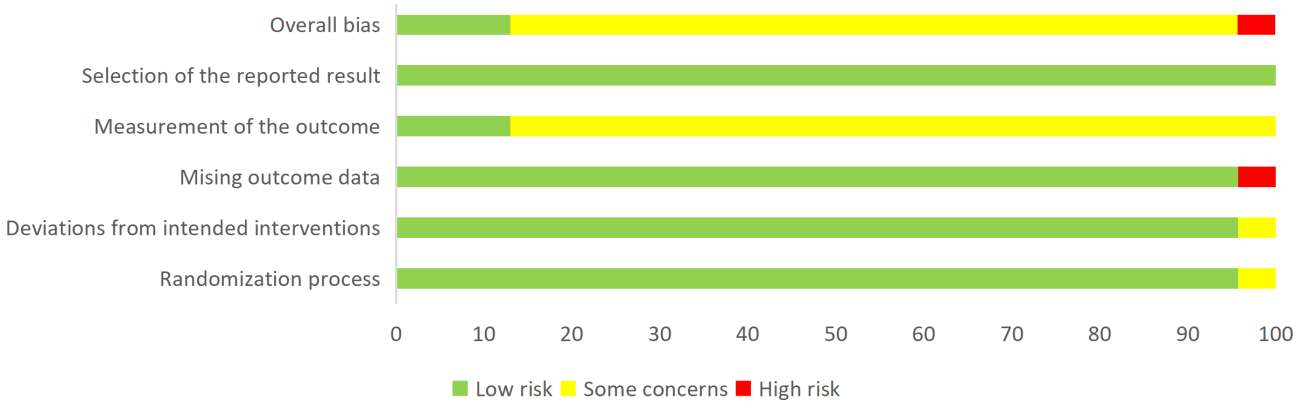


Quality Assessment Results

Among the 24 included studies (Figures 3 and 4), 3 were rated as low risk of bias [32,41,49]. One study had a high risk due to unaddressed missing data [50]; the remaining 20 had some concerns (1 study [52] due to potential bias in the intervention, 1 study [37] for not reporting allocation concealment, and the other 18 due to subjective outcome assessments potentially influenced by positive psychological suggestion).

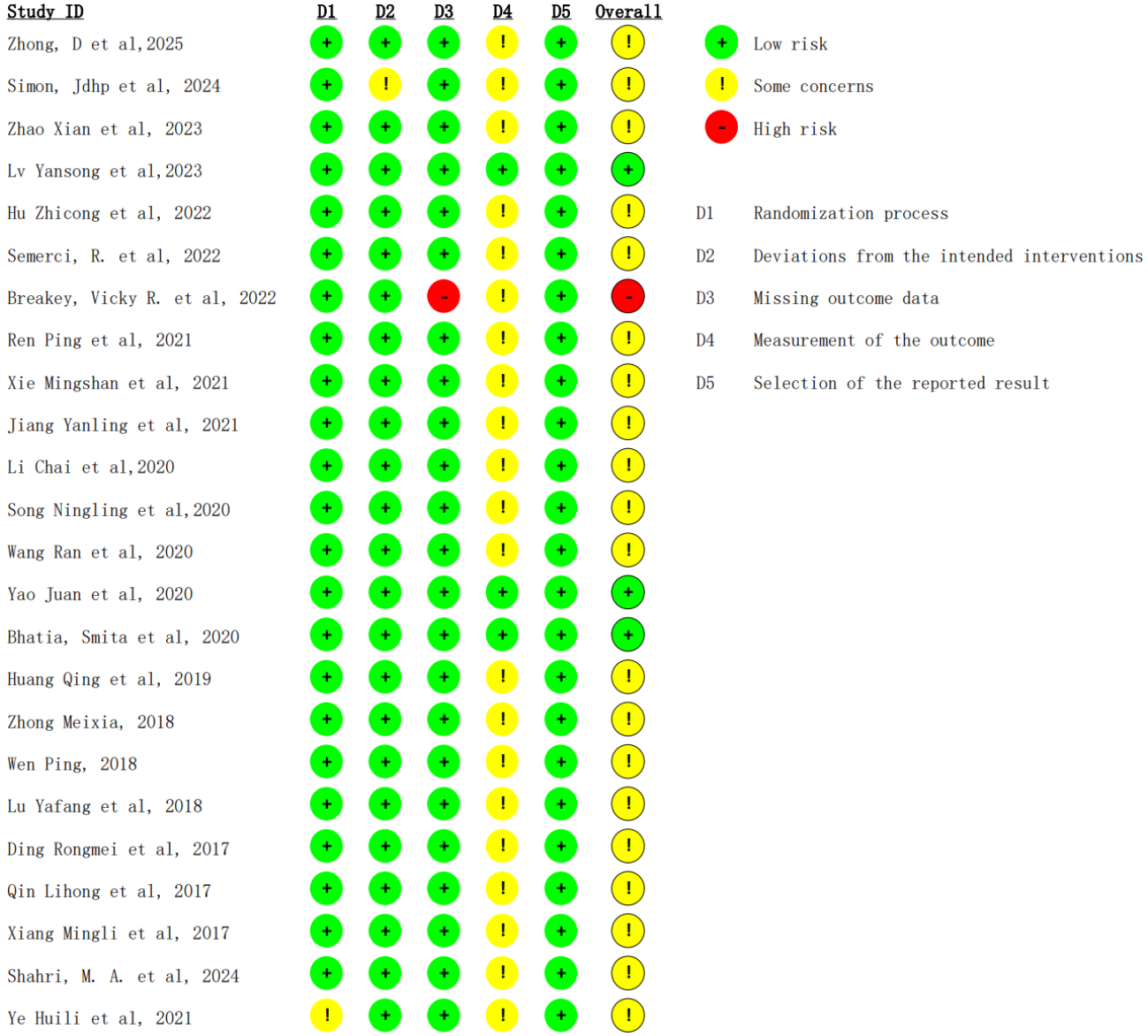
XSL•FO
RenderX

Figure 3. Risk of bias graph.



GRADE evidence quality ratings are presented in Multimedia Appendix 2. Evidence for infection incidence, overall incidence of PICC complications, PICC phlebitis, PICC puncture site bleeding, PICC puncture site infection, PICC occlusion, PICC catheter dislodgement, and PICC catheter displacement was rated as moderate. In contrast, evidence quality for QoL, PICC thrombogenesis, health knowledge, children's self-management ability, and treatment adherence was rated as very low.

Figure 4. Risk of bias summary [30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51-53].



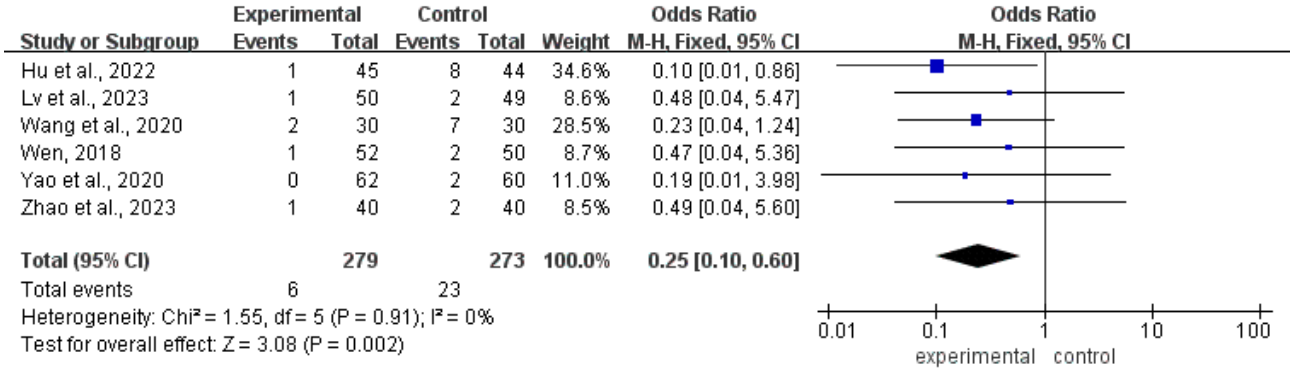
XSL•FO
RenderX

Meta-Analysis Results

Infection Incidence

Six [31-33,39,41,44] of the 24 studies investigated the effectiveness of mHealth-based interventions in controlling infection rates among pediatric patients with cancer. The infections reported in the included studies primarily involved skin, bloodstream, catheter-related, oral, and perianal infections. The studies demonstrated homogeneity (P=.91; I^2=0%), thus enabling analysis using a fixed-effects model. The results showed that mHealth-based interventions could significantly reduce the incidence of infections (OR 0.25, 95% CI 0.10-0.60; P=.002, Figure 5).

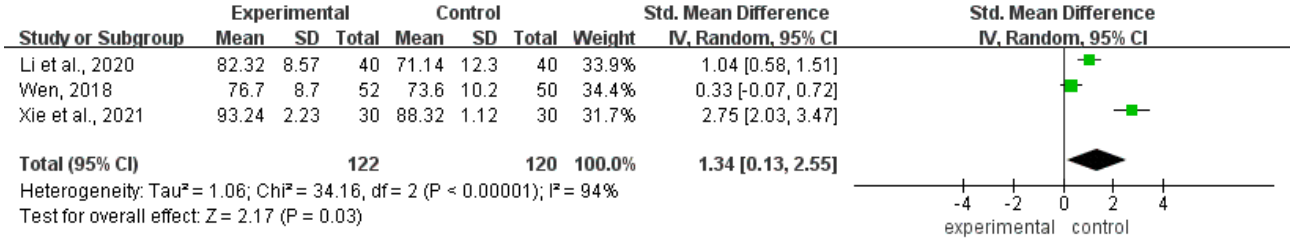
Figure 5. Forest plot of the effect of mHealth on infection incidence [31,32,33,39,41,44].



Quality of Life

Of the 24 studies included in the analysis, three [35,38,44] reported total QoL scale scores for pediatric patients with cancer who received mHealth-based interventions. Due to high heterogeneity across the studies (P<0.001; I^2=94%), a random-effects model was used for the analysis. Meta-analysis results showed that mHealth-based interventions were more effective than control in improving the QoL of children with cancer (SMD =1.34, 95% CI 0.13-2.55; P=.03, Figure 6).

Figure 6. Forest plot of the effect of mobile health on quality of life [35,38,44].

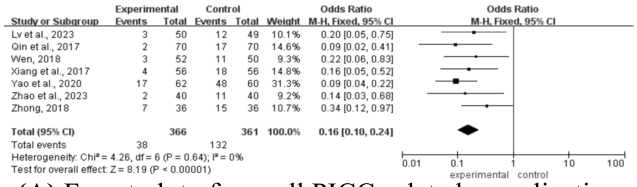


PICC-Related Complications

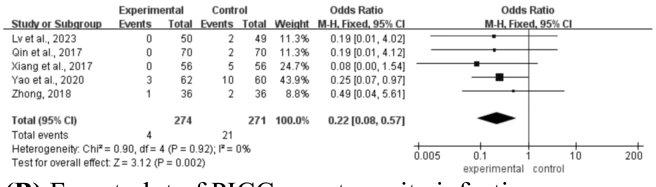
Among the 24 studies, 10 [31-33,40,41,43,44,46-48] reported the incidence of PICC-related complications in mHealth interventions, including the overall incidence of PICC complications (n=7) [31,32,41,43,44,47,48], puncture site infection (n=5) [32,41,43,47,48], phlebitis (n=9) [31-33,41,43,44,46-48], thrombogenesis (n=3) [33,43,47], puncture site bleeding (n=5) [32,41,43,44,47], catheter occlusion (n=8) [32,33,40,41,43,44,46,48], catheter dislodgement (n=9) [31-33,40,41,43,44,46,48], and catheter displacement (n=3) [43,46,47]. Heterogeneity tests revealed no significant heterogeneity among studies for any outcome measure (P>.1; I^2=0%), thus allowing analysis using a fixed-effect model. Meta-analysis results showed that in the mHealth intervention groups, the incidence of overall complications (OR 0.16, 95% CI 0.10-0.24; P<.001, Figure 7A), puncture site infection (OR 0.22, 95% CI 0.08-0.57; P=.002, Figure 7B), phlebitis (OR 0.30, 95% CI 0.16-0.58; P=.0003, Figure 7C), thrombogenesis (OR 0.15, 95% CI 0.03-0.82; P=.03, Figure 7D), puncture site bleeding (OR 0.28, 95% CI 0.12-0.61; P=.001, Figure 7E), catheter occlusion (OR 0.33, 95% CI 0.16-0.65; P=.002, Figure 7F) and catheter dislodgement (OR 0.29, 95% CI 0.16-0.54; P<.001, Figure 7G) were all lower than in the control groups. However, there was no significant difference in the incidence of catheter displacement (OR 0.44, 95% CI 0.15-1.29; P=.13, Figure 7H).

XSL•FO
RenderX

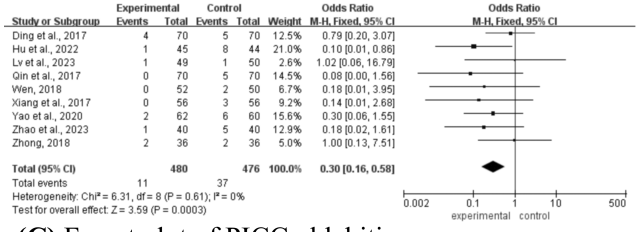
Figure 7. Forest plot of the effect of mHealth on (A) overall PICC-related complications [31,32,41,43,44,47,48], (B) PICC puncture site infection [32,41,43,47,48], (C) PICC phlebitis [31,32,33,41,43,44,46,47,48], (D) PICC thrombogenesis [33,43,47], (E) PICC puncture site bleeding [32,41,43,44,47], (F) PICC catheter occlusion [32,33,40,41,43,44,46,48], (G) PICC catheter dislodgement [31,32,33,40,41,43,44,46,48], and (H) PICC catheter displacement [43,46,47]. PICC: peripherally inserted central catheters.



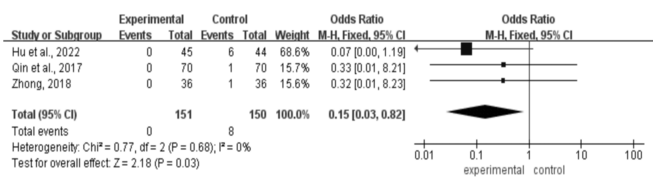
(A) Forest plot of overall PICC-related complications



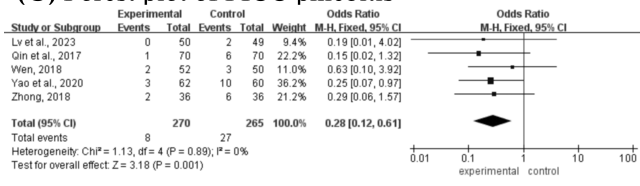
(B) Forest plot of PICC puncture site infection



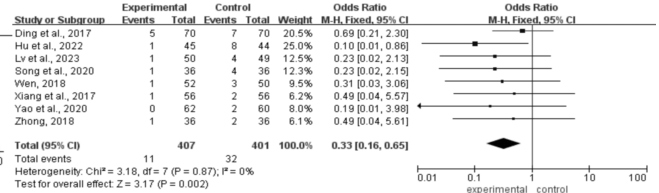
(C) Forest plot of PICC phlebitis



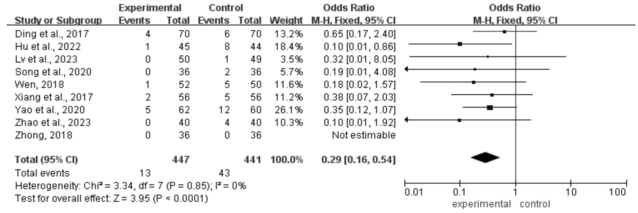
(D) Forest plot of PICC thrombogenesis



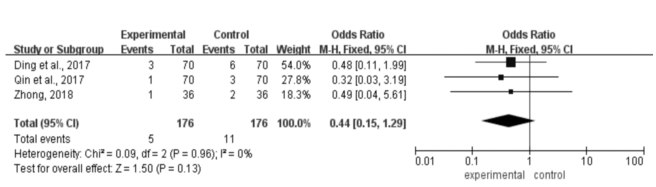
(E) Forest plot of PICC puncture site bleeding



(F) Forest plot of PICC catheter occlusion



(G) Forest plot of PICC catheter dislodgement

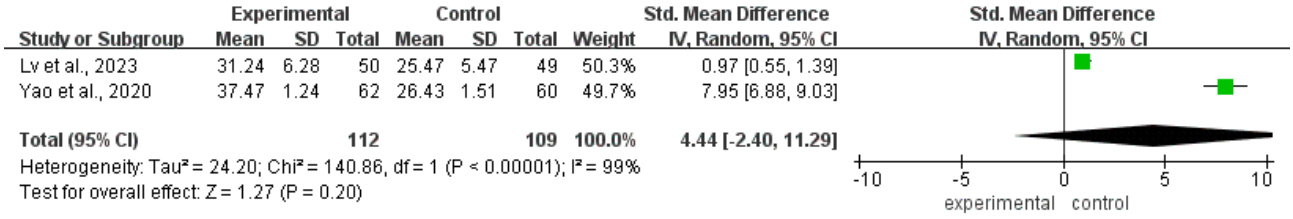


(H) Forest plot of PICC catheter displacement

Health Knowledge

Among the 24 studies, 2 [32,41] reported the effects of mHealth-based interventions on health knowledge. Owing to substantial heterogeneity (P<.001; I^2=99%), a random-effects model was used. There were no significant differences in health knowledge compared with the control group (SMD 4.44, 95% CI −2.40 to 11.29; P=.20, Figure 8).

Figure 8. Forest plot of the effect of mobile health on health knowledge [32,41].

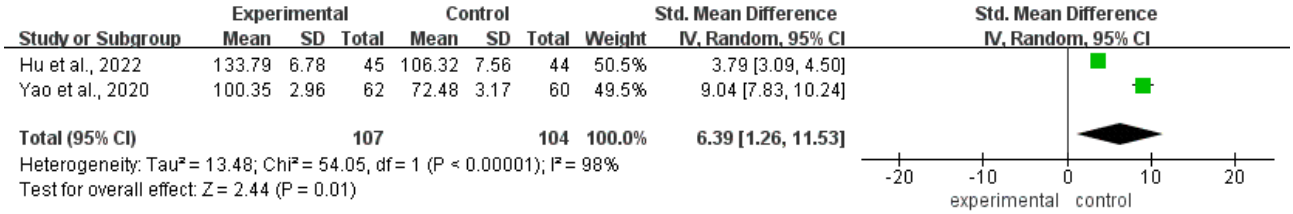


Self-Management Ability

Among the 24 studies, 2 [33,41] reported the effects of mHealth-based interventions on self-management abilities in pediatric patients with cancer compared to usual care. Due to high heterogeneity across studies (P<.001; I^2=98%), a random-effects model was used for analysis. The results of the meta-analysis indicated that mHealth-based interventions were more effective than control in improving children's self-management abilities (SMD 6.39, 95% CI 1.26-11.53; P=.01, Figure 9).

XSL•FO
RenderX

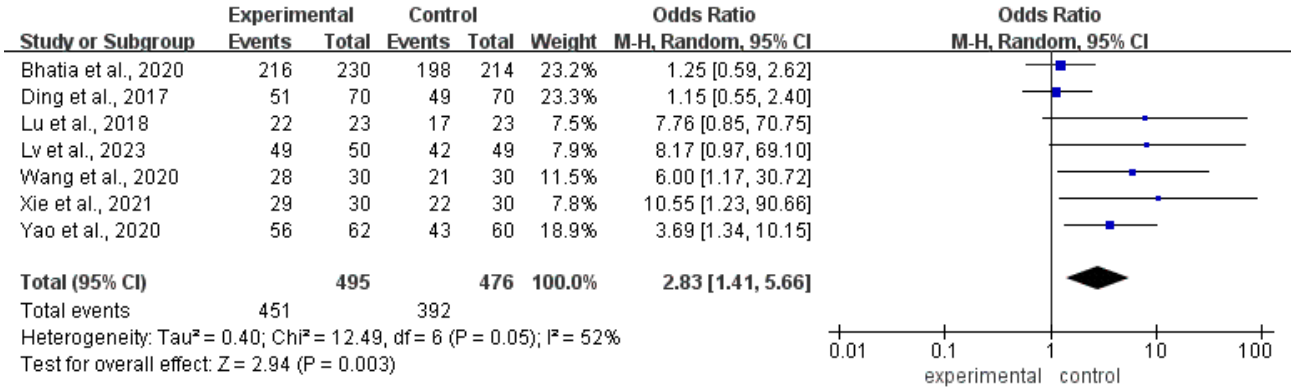
Figure 9. Forest plot of the effect of mobile health on self-management ability [33,41].



Treatment Adherence

Of the 24 studies, 10 [32,35-37,39,41,42,45,46,49] assessed the effectiveness of mHealth-based interventions compared to usual care in improving treatment adherence among pediatric patients with cancer. Two studies [36,42] were not pooled because of differences in scoring directions, and 1 study [37] was not entered into the quantitative synthesis owing to internally inconsistent and unverifiable adherence data; the remaining 7 studies reporting dichotomous adherence outcomes were included in the meta-analysis. Heterogeneity testing revealed significant heterogeneity among the studies (P=.05; I^2=52%), which prompted the use of a random-effects model. The results indicated that children in the mHealth group had higher treatment adherence than those in the control group (OR 2.83, 95% CI 1.41-5.66; P=.003, Figure 10).

Figure 10. Forest plot of the effect of mobile health on treatment adherence [32,35,39,41,45,46,49].



Sensitivity Analyses

For most outcomes, excluding any single study did not alter the effect sizes of mHealth-based interventions, indicating stable results (see Multimedia Appendix 3). Exceptions were observed for QoL, PICC complication incidence, and treatment adherence. For quality of life, excluding either study by Li et al [38] or Xie [35] led to significant changes in the overall effect size (SMD 1.52, 95% CI −0.86 to 3.90; P=.21; SMD 0.67, 95% CI −0.03 to 1.38; P=.06), indicating instability. Exclusion study by Hu et al [33] (OR 0.33, 95% CI 0.03-3.19; P=.34), substantially altered the effect estimate for thrombogenesis. In contrast, heterogeneity in treatment adherence decreased markedly after excluding the study by Ding et al [46] (P=.11; I^2=45%), while the pooled effect size remained stable (OR 3.72, 95% CI 1.68-8.22; P<.001). Two outcome indicators (health knowledge and self-management ability) were not subjected to sensitivity analysis as they were only involved in 2 studies.

Publication Bias

The assessment of publication bias in the effects of mHealth–based interventions on infection incidence, QoL, PICC-related complications, health knowledge, self-management ability, and treatment adherence in pediatric patients with cancer was facilitated by the use of funnel plots. The majority of the funnel plots demonstrated symmetrical patterns in the outcomes, with the exception of those pertaining to health knowledge and self-management ability (Multimedia Appendix 4). In addition, the Egger test was performed to evaluate the incidence of infection (t=0.56; P=.61), the QoL (t=12.12; P=.052), PICC-related complications (t=0.42; P=.69), PICC puncture site infection (t=−0.66; P=.55), PICC phlebitis (t=−1.59; P=.16), PICC thrombogenesis (t=20.91; P=.03), PICC puncture site bleeding (t=−0.39; P=.72), PICC catheter occlusion (t=−2.35; P=.07), PICC catheter dislodgement (t=−2.09; P=.08), and PICC catheter displacement (t=−0.59; P=.66), and treatment adherence (t=5.37; P=.003). The results indicated the absence of substantial publication bias, with the exception of the PICC thrombogenesis, health knowledge, self-management ability, and treatment adherence.

Subgroup Analysis and Meta-Regression

Subgroup analyses and meta-regression were not conducted because fewer than 10 studies were available for each outcome. This decision aligns with the Cochrane Handbook [55] guidelines to prevent false-positive results, insufficient statistical power, and unstable results.

Discussion

Principal Findings

This review aims to evaluate whether mHealth-based interventions are more effective than usual care for pediatric patients with cancer. The quality of the evidence for each

XSL•FO
RenderX

outcome measure in the included studies was graded using GRADEpro, with the following results:

1. Moderate certainty evidence: mHealth-based interventions significantly reduce the incidence of infections and PICC complications (including overall complications, puncture site infection, phlebitis, puncture site bleeding, catheter occlusion, and catheter dislodgement) compared to usual care in pediatric patients with cancer, with no significant effect observed for PICC catheter displacement.
2. Very low certainty evidence: mHealth-based interventions outperform usual care in improving QoL, self-management ability, treatment adherence, and reducing the PICC thrombogenesis rate among pediatric patients with cancer, but do not significantly enhance their health knowledge.

Effect of mHealth Interventions on Infection Rates

This is the first review to summarize the effect of mHealth on infection rates in pediatric patients with cancer, demonstrating a significant protective benefit. The interventions used 8 BCTs to establish effective bidirectional communication between clinicians and patients, offering practical insights for clinical implementation. Notably, the majority of interventions used WeChat, a social platform that dominates the Chinese digital landscape with over 1 billion monthly active users and preinstallation on more than 90% of smartphones across all age demographics. Its comprehensive functionality encompasses instant messaging (text, voice, and video messages), free audio and video calls, group chats (eg, interactions between health care professionals and patients or among patients), and private messaging for individualized consultation. In health care settings, WeChat is commonly used through official accounts (for health education dissemination), mini-programs (for data recording or questionnaire administration), and multimedia transmission (eg, image or video reports), offering advantages of low cost and wide reach [54]. Nevertheless, this app presents notable limitations. Interactions frequently depend on medical staff availability, potentially resulting in delayed responses [56]. In addition, current health education materials often require manual compilation and uploading, reducing efficiency and increasing clinician workload—potentially compromising care quality and safety. There is a clear need for more automated, dynamic content delivery systems.

Effect of mHealth Interventions on QoL

The results of a meta-analysis indicate that mHealth-based interventions significantly improved the QoL for pediatric patients with cancer compared to usual care. These findings address previous gaps in quantitative measurement [15] and reconcile inconsistencies attributed to confounding factors in interventions and populations [14]. The included mHealth interventions incorporated 6 BCTs, supporting psychological, social, and symptom-related needs, thereby reducing risks of psychological distress, infections, and PICC complications. However, high heterogeneity and sensitivity analyses indicate instability in these results. All studies on this outcome measure used WeChat as the platform, differing only in sample size and measurement tools. This suggests that the instability and heterogeneity of these intervention outcomes may be influenced more by methodological factors than by mHealth. In addition, the observed large effect size of QoL may be partly attributable to the elevated effect size in Xie's [35] study, which assessed outcomes at 2 months post intervention—a shorter timeframe than other studies (3 mo, 6 mo)—resulting in stronger short-term effects. Furthermore, the employment of subjective assessment scales in all studies may have resulted in an overestimation of results. Therefore, while mHealth interventions show potential for improving the QoL in pediatric patients with cancer, their long-term effects require further confirmation through more rigorous evidence-based studies before integration into routine clinical care.

Effect of mHealth Interventions on Incidence of PICC-Related Complications

This study is the first to summarize the effectiveness of mHealth-based interventions in reducing PICC-related complications within the pediatric oncology field. Results indicate that mHealth management significantly lowers overall complication rates compared to usual care, likely by mitigating recall bias and misunderstandings associated with traditional education methods (eg, demonstrations or printed manuals) [57]. The mHealth intervention in this outcome involved 8 BCTs, offering real-time monitoring and correction, scheduling and reminders, online assessment, and follow-up. Despite these advantages, the nonsignificant effect of mHealth on reducing catheter displacement may be attributed to children's growth and frequent postural changes, which increase the risk of mechanical traction [58]. Hu et al [33] implemented a reward and threat system (weekly quizzes with prizes for successful venous catheterization) that generated a more pronounced intervention advantage in the intervention group compared to the control group. In contrast, the remaining 2 studies [43,47] used health education manuals in their control groups, potentially reducing the distinction from the experimental group. This suggests that certain functions of mHealth may not necessarily offer a significant advantage over health education manuals for specific outcomes. Overall, mHealth appears effective for reducing PICC complications, though efficacy may vary by complication type and specific intervention components.

Effect of mHealth Interventions on Treatment Adherence

Similar to previous reviews [15], treatment adherence was significantly better in the mHealth group than in the usual care group. The mHealth intervention in this study integrated 9 BCTs. These strategies collectively address core aspects of treatment adherence, such as timely medication administration, maintaining appropriate lifestyle habits, and attending regular follow-up appointments and assessments [59]. Additionally, dedicated online accounts facilitate peer interaction, reducing stigma and enhancing adherence [60]. However, these results exhibit some heterogeneity and extreme effect sizes. The observed heterogeneity may be partly attributable to the 3-arm randomized controlled design used by Ding et al [46]. The extreme effect sizes may be related to the smaller sample sizes. Notably, while mHealth interventions significantly improve adherence among pediatric patients with cancer, participant attrition remains a significant issue. This may stem from limited digital literacy among children and their parents, which can

XSL•FO
RenderX

hinder the effective use of mobile devices and negatively impact adherence [61]. Therefore, clinical practice should establish timely feedback and reminder mechanisms, which may be key factors in maintaining long-term engagement among pediatric patients.

Effect of mHealth Interventions on Health Knowledge and Self-Management Ability

This review found that mHealth-based interventions did not significantly improve health knowledge scores, possibly due to a lack of age-appropriate educational content. The 4 BCT clusters used may not suffice for children's cognitive levels [62]. Future interventions should incorporate age-tailored content, interactive designs, and parental involvement to enhance learning outcomes. In contrast, mHealth interventions demonstrated favorable effects on improving children's self-management ability. This result pertains to 5 BCT clusters. Benefits likely stem from flexible, real-time education and feedback [63], empowering both children and caregivers [62]. The high effect sizes may reflect the use of subjective scales and potential Hawthorne effects.

However, Yao et al [41] used a professional medical platform for intervention, yielding effect sizes higher than those reported in studies using WeChat platforms [32,33] for both outcome measures. This discrepancy may explain the higher heterogeneity and wider confidence intervals observed, warranting cautious interpretation of the findings. Notably, Yao et al [41] was also the only study to conduct a cost-related analysis, reporting that the cost of hospital visits for PICC maintenance in the intervention group 117.50 CNY (US $17.07; IQR 105.00-132.50) was significantly lower than that in the control group 321 CNY (US $46.66; IQR 261.50-374.25), limiting the generalizability of the evidence. Therefore, balancing patient economic benefits with clinician workload remains a key challenge for health care administrators.

Strengths and Limitations

This review has several strengths. First, it provides a systematic review and meta-analysis of mHealth effectiveness in pediatric cancer, moving beyond usability to report quantitative clinical outcomes—including infection rates and PICC complications—for the first time. It also offers evidence on QoL, health knowledge, self-management, and treatment adherence, establishing a foundation for future research and practice. Second, by including only RCTs—the gold standard in clinical evidence—and restricting participants to pediatric patients undergoing treatment (excluding parents or survivors), it minimizes confounding and enhances translational potential. Third, it identifies and categorizes BCTs across studies, standardizing intervention components and improving comparability and replicability for future design. Fourth, the inclusion of Chinese databases supplements evidence from developing countries, addressing a gap in prior research.

Limitations should also be acknowledged. First, most included studies lacked the implementation of blinding, resulting in a current lack of high-certainty evidence. This may be related to the inherent difficulty of setting up blinding for mHealth interventions. Second, 79% of the included studies were conducted in China, with interventions predominantly delivered via the WeChat platform. However, the usage context of WeChat differs from that of mHealth tools commonly used in other countries, such as standalone mHealth apps or web-based platforms. The integration of daily life functions with health care services within WeChat may result in higher user engagement and retention compared with independent medical apps, potentially representing a key factor underlying the observed intervention effectiveness. Consequently, this WeChat-based model, characterized by high user engagement, may be difficult to directly replicate in other countries or regions with different health care systems and mHealth infrastructure, potentially limiting the generalizability of the findings to other linguistic and cultural contexts. Third, as a nonmedical social platform, WeChat has limitations in clinical applications. Specifically, it shows deficiencies in data encryption, regulatory compliance, liability definition, and ethical oversight during use. Simultaneously, medical staff may experience vigilance fatigue and blurred professional boundaries during specialized communications on such platforms due to unclear responsibility frameworks [64]. Moreover, it also lacks integration with clinical systems (eg, electronic health records), structured data support, and advanced algorithmic feedback [65]. Fourth, the limited number of included studies precluded subgroup analysis or meta-regression, restricting deeper exploration of heterogeneity sources, including the influence of BCT number on outcomes. Fifth, only Chinese and English articles were included due to the authors' reading limitations, which may have resulted in the omission of high-quality articles in other languages. Finally, although the initial protocol registration planned to include a range of telehealth interventions and study designs, the scope of the review was refined during the implementation phase to focus specifically on RCTs of mHealth-based interventions. This refinement improved conceptual clarity and internal validity, while also strengthening the clinical and practical relevance of the findings. However, this focused scope may limit the applicability of the conclusions to other forms of telehealth (such as telephone-based consultations or video-enabled remote care) and to nonrandomized study designs that more closely reflect real-world practice (eg, stepped-wedge or cluster-based designs).

Prospects

The findings of this study suggest that the extant evidence remains inadequate, and subsequent research may investigate the following domains.

First, future platforms could integrate ecological momentary assessment and smart biosensors to enable dynamic, predictive monitoring. Establishing automated, risk-stratified feedback mechanisms and exploring the incorporation of large language models could enhance efficiency through automated content updates and personalized information delivery. This would allow clinicians to focus more on data review and clinical decision-making [66]. Second, given the advantage of WeChat highlighted in this review, future developments should prioritize embedding mHealth modules into widely used ecosystem apps to minimize user burden, rather than developing isolated apps. Third, research quality should be enhanced. Adopting internationally recognized core outcome sets would facilitate

XSL•FO
RenderX

cross-study comparisons [67]. Large-scale, multicenter RCTs should be conducted with extended follow-up periods to observe the temporal dynamics of intervention effects and long-term clinical outcomes. Fourth, equity and safety should be prioritized. Preuse assessments of digital health literacy and preferences among patients, parents, and health care providers should be conducted, offering tailored training to ensure resource equity during mHealth implementation. Platform users should be clearly defined to mitigate the adverse effects of inappropriate mobile device use on children's vision, sleep, and other areas [68]. Finally, implementation research should be strengthened. Clear cost-benefit analyses should be conducted; barriers and facilitators to mHealth adoption should be evaluated, such as health care institutions' feasibility regarding equipment investment, technical maintenance, staffing costs, and time allocation; and how mHealth interventions integrate with existing care processes and electronic information systems should be examined to avoid additional workload, duplicate documentation, or process conflicts, thereby enhancing technology usability and sustainability.

Conclusion

This systematic review and meta-analysis synthesizes evidence from RCTs on mHealth interventions for children with cancer. The findings suggest that mHealth holds significant value in pediatric oncology care, demonstrating greater effectiveness than usual care in reducing the incidence of infection and various PICC-related complications, including overall complications, site infection, phlebitis, thrombosis, bleeding, occlusion, and dislodgement. mHealth interventions also significantly improved patients' QoL, self-management ability, and treatment adherence. However, current evidence does not show a significant advantage for mHealth in reducing PICC catheter displacement or enhancing health knowledge. Large effect sizes for some outcomes may be related to subjective measurement tools and small sample sizes. Limitations such as lack of blinding and considerable heterogeneity reduce the certainty of evidence and may lead to overestimation of effects; therefore, results should be interpreted cautiously. Despite these limitations, existing evidence supports the potential benefits and applicability of mHealth in pediatric cancer care. Future research should conduct higher-quality, multinational RCTs and use methods such as network meta-analysis to identify the most effective BCTs within mHealth platforms. Concurrently, integrating implementation science and cost-effectiveness analyses will be crucial, with a focus on aligning mHealth technologies with existing clinical workflows to improve their feasibility, scalability, and sustainable impact.

Acknowledgments

The authors only used generative artificial intelligence tools to polish the English language during the refinement of the initial draft to enhance the readability of the manuscript. Artificial intelligence was not involved in the generation of scientific content, such as research design, data analysis, result interpretation, or chart generation. All references were independently retrieved, added, and verified by the authors themselves. The authors carefully reviewed and revised all text assisted by artificial intelligence and take full responsibility for the content of the manuscript.

Funding

This study was funded by Chengdu Municipal Health Commission Scientific Research Project—professional academic or number: 2024062; Sichuan Province science and Technology Department key research and development project or number: 23ZDYF2075; Chengdu Science and Technology Bureau Science and technology innovation research and development project or number: 2022-YF05-01621-SN.

Data Availability

This study did not create or analyze new data, so data sharing is not applicable.

Authors' Contributions

Conceptualization: HY, XM
Data curation: HY, YW
Formal analysis: HY, HW, MD
Funding acquisition: XM
Methodology: HY, YX
Project administration: HY, YW
Software: HY, YL
Supervision: XM
Validation: HY
Visualization: HY, WY
Writing - original draft: HY, YW, HW
Writing - review & editing: HY, YW, YX, HW, WY, MD, XM

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 24 KB - mhealth_v14i1e86836_app1.docx]

Multimedia Appendix 2
GRADE assessment results.
[DOCX File, 25 KB - mhealth_v14i1e86836_app2.docx]

Multimedia Appendix 3
Results of sensitivity analyses.
[DOCX File, 382 KB - mhealth_v14i1e86836_app3.docx]

Multimedia Appendix 4
Funnel plots.
[DOCX File, 769 KB - mhealth_v14i1e86836_app4.docx]

Checklist 1
PRISMA checklist.
[DOCX File, 29 KB - mhealth_v14i1e86836_app5.docx]

References

1. Kappelmann L, Götte M, Krombholz A, Hüter J, Fischer B. Factors that influence physical activity behavior in children and adolescents during and after cancer treatment: a qualitative systematic review of the literature. Pediatr Exerc Sci 2024 May 1;36(2):106-114. [doi: 10.1123/pes.2022-0111] [Medline: 37890837]
2. Gu Y, Yu J, Li L, et al. Parents' perceptions of physical activity for their children with cancer: a qualitative meta-synthesis. Front Pediatr 2025;13:1402516. [doi: 10.3389/fped.2025.1402516] [Medline: 40224387]
3. Lehrnbecher T, Laws HJ. Infectious complications in pediatric cancer patients [Article in German]. Klin Padiatr 2005 Nov;217 Suppl 1:S3-S8. [doi: 10.1055/s-2005-872498] [Medline: 16288349]
4. Villeneuve S, Aftandilian C. Neutropenia and infection prophylaxis in childhood cancer. Curr Oncol Rep 2022 Jun;24(6):671-686. [doi: 10.1007/s11912-022-01192-5] [Medline: 35230594]
5. Katabalo DM, Abraham M, Kidenya BR, Liwa A, Schroeder K. Prescription patterns of supportive care medications among children receiving chemotherapy treatments at a major referral hospital in Tanzania: where are we in managing chemotherapy-induced toxicities? Front Oncol 2025;15:1444565. [doi: 10.3389/fonc.2025.1444565] [Medline: 40071090]
6. Fedhila F, Hannachi MW, Jbebli E, et al. Impact of music therapy on quality of life in children with cancer. Children (Basel) 2023 Aug 31;10(9):1486. [doi: 10.3390/children10091486] [Medline: 37761447]
7. Nunn JL, Takashima MD, Wray-Jones EM, Soosay Raj TA, Hanna DMT, Ullman AJ. Central venous access device adverse events in pediatric patients with cancer: a systematic review and meta-analysis. Support Care Cancer 2024 Sep 16;32(10):662. [doi: 10.1007/s00520-024-08853-0] [Medline: 39283363]
8. Ilic A, Sievers Y, Roser K, Scheinemann K, Michel G. The information needs of relatives of childhood cancer patients and survivors: a systematic review of qualitative evidence. Patient Educ Couns 2023 Sep;114:107840. [doi: 10.1016/j.pec.2023.107840] [Medline: 37321115]
9. Walling EB, Fiala M, Connolly A, Drevenak A, Gehlert S. Challenges associated with living remotely from a pediatric cancer center: a qualitative study. J Oncol Pract 2019 Mar;15(3):e219-e229. [doi: 10.1200/JOP.18.00115] [Medline: 30702962]
10. Jacob SA, Daas R, Feliciano A, LaMotte JE, Carroll AE. Caregiver experiences with accessing sickle cell care and the use of telemedicine. BMC Health Serv Res 2022 Feb 22;22(1):239. [doi: 10.1186/s12913-022-07627-w] [Medline: 35193570]
11. mHealth: new horizons for health through mobile technologies. : World Health Organization; 2011 URL: https://iris.who.int/server/api/core/bitstreams/ad1b13c0-7c82-47b4-8dd5-f0a26c3a3cc3/content [accessed 2026-02-24]
12. van der Kamp MR, Hengeveld VS, Brusse-Keizer MGJ, Thio BJ, Tabak M. eHealth technologies for monitoring pediatric asthma at home: scoping review. J Med Internet Res 2023 Jul 21;25:e45896. [doi: 10.2196/45896] [Medline: 37477966]
13. Oates GR, Mims C, Geurs R, et al. Mobile health platform for self-management of pediatric cystic fibrosis: impact on patient-centered care outcomes. J Cyst Fibros 2023 Sep;22(5):823-829. [doi: 10.1016/j.jcf.2023.04.009] [Medline: 37085386]

XSL•FO
RenderX

14. Ramsey WA, Heidelberg RE, Gilbert AM, Heneghan MB, Badawy SM, Alberts NM. eHealth and mHealth interventions in pediatric cancer: a systematic review of interventions across the cancer continuum. Psychooncology 2020 Jan;29(1):17-37. [doi: 10.1002/pon.5280] [Medline: 31692183]
15. González-Díaz A, Pérez-Ardanaz B, Suleiman-Martos N, Gómez-Urquiza JL, Canals Garzón C, Gómez-Salgado J. The role of mobile applications in enhancing the health-related quality of life of children with cancer: a systematic review and meta-analysis. Children (Basel) 2025 Jul 14;12(7):927. [doi: 10.3390/children12070927] [Medline: 40723120]
16. Upreti A, Dahal S, Thapa S. Use of mobile health applications in pain management in paediatric cancer care: a descriptive literature review [Bachelor's thesis]. : Metropolia University of Applied Sciences; 2024 URL: https://www.theseus.fi/bitstream/handle/10024/872697/Upreti_Dahal_Thapa.pdf?sequence=2 [accessed 2026-02-23]
17. Zhu J, Fang J, Wang L, Chen W. Application of mobile health care in pain management of children with malignant tumors: a scoping review [Article in Chinese]. Chin Nurs Manag 2024;24(8):1261-1265. [doi: 10.3969/j.issn.1672-1756.2024.08.026]
18. Delemere E, Maguire R. Behaviour change techniques within smartphone application connected health interventions for paediatric cancer: a review. Presented at: 2023 IEEE International Symposium on Technology and Society (ISTAS); Sep 13-15, 2023. [doi: 10.1109/ISTAS57930.2023.10306058]
19. Mehdizadeh H, Asadi F, Mehrvar A, Nazemi E, Emami H. Smartphone apps to help children and adolescents with cancer and their families: a scoping review. Acta Oncol 2019 Jul;58(7):1003-1014. [doi: 10.1080/0284186X.2019.1588474] [Medline: 30915872]
20. Cheng L, Duan M, Mao X, Ge Y, Wang Y, Huang H. The effect of digital health technologies on managing symptoms across pediatric cancer continuum: a systematic review. Int J Nurs Sci 2020;8(1):22-29. [doi: 10.1016/j.ijnss.2020.10.002] [Medline: 33575441]
21. Skeens MA, Jackson DI, Sutherland-Foggio MS, Sezgin E. mHealth apps in the digital marketplace for pediatric patients with cancer: systematic search and analysis. JMIR Pediatr Parent 2024 Oct 1;7:e58101. [doi: 10.2196/58101] [Medline: 39352720]
22. Herold F, Theobald P, Gronwald T, Rapp MA, Müller NG. Going digital - a commentary on the terminology used at the intersection of physical activity and digital health. Eur Rev Aging Phys Act 2022 Jul 16;19(1):17. [doi: 10.1186/s11556-022-00296-y] [Medline: 35840899]
23. Istepanian RSH. Mobile health (m-Health) in retrospect: the known unknowns. Int J Environ Res Public Health 2022 Mar 22;19(7):3747. [doi: 10.3390/ijerph19073747] [Medline: 35409431]
24. Effectiveness of mobile health (mHealth) interventions in pediatric cancer: a systematic review and meta-analysis of randomized controlled trials. PROSPERO. URL: https://www.crd.york.ac.uk/PROSPERO/view/CRD420251108938 [accessed 2026-03-24]
25. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
26. Amir-Behghadami M, Janati A. Population, Intervention, Comparison, Outcomes and Study (PICOS) design as a framework to formulate eligibility criteria in systematic reviews. Emerg Med J 2020 Jun;37(6):387. [doi: 10.1136/emermed-2020-209567] [Medline: 32253195]
27. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
28. Schünemann H, Brożek J, Guyatt G, Oxman A. GRADE handbook. : The GRADE Working Group; 2013 URL: https://www.rama.mahidol.ac.th/ceb/sites/default/files/public/pdf/journal_club/2017/GRADE%20handbook.pdf [accessed 2026-02-23]
29. Higgins JPT, Thomas J, Chandler J, et al. Cochrane Handbook for Systematic Reviews of Interventions, 2nd edition: John Wiley & Sons; 2019. URL: https://training.cochrane.org/handbook [accessed 2026-02-23]
30. Zhong D, Liu Y, Mo L, Zhang L, Shen X, Xiang Q. Establishment and application of an intelligent management platform for home medication for children with leukemia. PLoS One 2025;20(4):e0320790. [doi: 10.1371/journal.pone.0320790] [Medline: 40168348]
31. Zhao X, Xia C, Wang Y. Application effects of WeChat platform-based continuous nursing on children with leukemia undergoing peripherally inserted central catheter catheterization [Article in Chinese]. China Mod Med 2023;30(11):188-192 [FREE Full text]
32. Lv Y, Tian H, Yang Y, Li Y, Xue J. Effect of information-based continuous health guidance on catheter maintenance compliance and complications in school-age children with malignant tumors undergoing PICC catheterization [Article in Chinese]. J Clin Pathol Res 2023;43(4):788-795. [doi: 10.11817/j.issn.2095-6959.2023.222029]
33. Hu Z, Chang X, Jia X. The application value of video-based continuity management services in alleviating the condition of children with leukemia undergoing deep venous catheterization during the induction phase [Article in Chinese]. Chin Gen Pract Nurs 2022;20(24):3377-3379. [doi: 10.12104/j.issn.1674-4748.2022.24.016]
34. Ren P, Gu Y, Li H, Ma L. The effect of mobile platforms on health education for children with brain tumors [Article in Chinese]. Chin Gen Pract Nurs 2021;19(11):1462-1467. [doi: 10.12104/j.issn.1674-4748.2021.11.006]
35. Xie M. Analysis of the impact of continued care based on WeChat platform on postoperative quality of life of children with Wilms tumor [Article in Chinese]. Chin Foreign Med Treat 2021;40(33):179-183 [FREE Full text]

XSL•FO
RenderX

36. Jiang Y. Application value analysis of knowledge attitude practice intervention combined with music therapy in the nursing of children with leukemia based on WeChat platform [Article in Chinese]. Psychologies Magazine 2021:111-112. [doi: 10.19738/j.cnki.psy.2021.05.054]
37. Ye H, Yu W. Application of nurse-patient WeChat group in continuing nursing care of children with acute leukemia [Article in Chinese]. World Latest Med Inf 2021;21(82):331-332. [doi: 10.3969/j.issn.1671-3141.2021.82.161]
38. Li C, Wei P, Pan J, Wang Z. Application of continuous nursing based on WeChat platform in life of children with malignant tumor [Article in Chinese]. Chin J Sch Doct 2020;34(7) [FREE Full text]
39. Wang R, Yang X. The application of WeChat groups for nurses and patients in the continuity care of children with acute leukemia [Article in Chinese]. Diet Health 2020;7(7):164 [FREE Full text]
40. Song N. Analysis of the application of continuing care in PICC catheterization for children with leukemia [Article in Chinese]. J Contemp Clin Med 2020;33(6). [doi: 10.3969/j.issn.2095-9559.2020.06.026]
41. Yao J, Mo L, Guo X, Deng Z, Liu K, Li X. Application of internet plus multiple linkage on home-based continuous nursing care for PICC in children with malignant tumor [Article in Chinese]. Chin J Nurs 2020;55(5):756-761. [doi: 10.3761/j.issn.0254-1769.2020.05.025]
42. Huang Q, Ling Z. Application analysis of outpatient management of childhood leukemia based on WeChat platform [Article in Chinese]. China Health Care Nutr 2019;29(31):383 [FREE Full text]
43. Zhong M, Wang L. The application of telephone follow-up and WeChat follow-up continuous care in acute leukemia children discharged with PICC catheters [Article in Chinese]. Chin Gen Pract Nurs 2018;16(14):1774-1777. [doi: 10.12104/j.issn.1674-4748.2018.14.043]
44. Wen P, Bai Y, Luo M, Hu Y, Lei X. Effects of continual nursing on quality of life and parental care ability for acute leukemia children with PICC in remission induction stage after discharge [Article in Chinese]. Mod Clin Nurs 2018;17(3):38-43. [doi: 10.3969/j.issn.1671-8283.2018.03.09]
45. Lu Y, Yao K, Zhang C. Effect of continuation nursing based on WeChat platform on postoperative quality of life in children with nephroblastoma [Article in Chinese]. Chronic Pathematol J 2018;19(S1):11-14. [doi: 10.16440/j.cnki.1674-8166.2018.s1.005]
46. Ding R, Fan L, Wang P. The Application of the combination of WeChat platform and nursing path in pediatric patients with tumors and PICC tube [Article in Chinese]. China Digit Med 2017;12(9):98-100. [doi: 10.3969/j.issn.1673-7571.2017.09.034]
47. Qin L, Ren A. The application effect of WeChat platform in the continuous care of tumor patients with PICC [Article in Chinese]. Chin J Woman Child Health Res 2017;28(2):308-312 [FREE Full text]
48. Xiang M, Yan M. The application of extended care services in the maintenance of peripherally inserted central catheters during the intermission of chemotherapy in children with leukemia [Article in Chinese]. J Basic Clin Oncol 2017;30(3):265-267. [doi: 10.3969/j.issn.1673-5412.2017.03.030]
49. Bhatia S, Hageman L, Chen Y, et al. Effect of a daily text messaging and directly supervised therapy intervention on oral mercaptopurine adherence in children with acute lymphoblastic leukemia: a randomized clinical trial. JAMA Netw Open 2020 Aug 3;3(8):e2014205. [doi: 10.1001/jamanetworkopen.2020.14205] [Medline: 32852553]
50. Breakey VR, Gupta A, Johnston DL, et al. A pilot randomized control trial of teens taking charge: a web-based self-management program for adolescents with cancer. J Pediatr Hematol Oncol Nurs 2022;39(6):366-378. [doi: 10.1177/27527530211068778] [Medline: 35759365]
51. Shahri MA, Farahani AS, Rassouli M, Khabazkhoob M, Aghbelagh SM. The impact of telenursing on the self-management of gastrointestinal symptoms in adolescent cancer patients receiving chemotherapy. Cancer Nurs 2026;49(2):121-128. [doi: 10.1097/NCC.0000000000001447] [Medline: 39680047]
52. Simon J, Schepers SA, van Gorp M, et al. Pain monitoring app leads to less pain in children with cancer at home: results of a randomized controlled trial. Cancer 2024 Jul 1;130(13):2339-2350. [doi: 10.1002/cncr.35100] [Medline: 37947136]
53. Semerci R, Akgün Kostak M, Taşkın C. The effect of using an interactive mobile application for the management of chemotherapy-induced nausea and vomiting in children: randomized controlled study. Eur J Oncol Nurs 2022 Jun;58:102121. [doi: 10.1016/j.ejon.2022.102121] [Medline: 35306247]
54. Chen X, Zhou X, Li H, Li J, Jiang H. The value of WeChat application in chronic diseases management in China. Comput Methods Programs Biomed 2020 Nov;196:105710. [doi: 10.1016/j.cmpb.2020.105710] [Medline: 32858284]
55. Geissbühler M, Hincapié CA, Aghlmandi S, Zwahlen M, Jüni P, da Costa BR. Most published meta-regression analyses based on aggregate data suffer from methodological pitfalls: a meta-epidemiological study. BMC Med Res Methodol 2021 Jun 15;21(1):123. [doi: 10.1186/s12874-021-01310-0] [Medline: 34130658]
56. Leahy AB, Schwartz LA, Li Y, et al. Electronic symptom monitoring in pediatric patients hospitalized for chemotherapy. Cancer 2021 Aug 15;127(16):2980-2989. [doi: 10.1002/cncr.33617] [Medline: 33945640]
57. Skinner TC, Barnard K, Cradock S, Parkin T. Patient and professional accuracy of recalled treatment decisions in out-patient consultations. Diabet Med 2007 May;24(5):557-560. [doi: 10.1111/j.1464-5491.2007.02129.x] [Medline: 17367303]
58. Connolly B, Amaral J, Walsh S, Temple M, Chait P, Stephens D. Influence of arm movement on central tip location of peripherally inserted central catheters (PICCs). Pediatr Radiol 2006 Aug;36(8):845-850. [doi: 10.1007/s00247-006-0172-8] [Medline: 16758187]

XSL•FO
RenderX

59. Hall AE, Paul C, Bryant J, et al. To adhere or not to adhere: rates and reasons of medication adherence in hematological cancer patients. Crit Rev Oncol Hematol 2016 Jan;97:247-262. [doi: 10.1016/j.critrevonc.2015.08.025] [Medline: 26412718]
60. Hasei J, Ishida H, Katayama H, et al. Utilizing the metaverse to provide innovative psychosocial support for pediatric, adolescent, and young adult patients with rare cancer. Cancers (Basel) 2024 Jul 23;16(15):2617. [doi: 10.3390/cancers16152617] [Medline: 39123345]
61. Su J, Wang J, Sun X. The relationship between health literacy and online health information search among caregivers of pediatric oncology children. Chin Gen Pract 2025;28(10):1207-1212. [doi: 10.12114/j.issn.1007-9572.2023.0814]
62. Mehdizadeh H, Asadi F, Emami H, Mehrvar A, Nazemi E. An mHealth self-management system for support children with acute lymphocytic leukemia and their caregivers: qualitative co-design study. JMIR Form Res 2022 Apr 15;6(4):e36721. [doi: 10.2196/36721] [Medline: 35228195]
63. Novrianda D, Herini ES, Haryanti F, Supriyadi E, Lazuardi L. Chemo assist for children mobile health application to manage chemotherapy-related symptoms in acute leukemia in Indonesia: a user-centered design approach. BMC Pediatr 2023 May 30;23(1):274. [doi: 10.1186/s12887-023-04076-0] [Medline: 37254039]
64. Jeyaraman M, Ramasubramanian S, Kumar S, et al. Multifaceted role of social media in healthcare: opportunities, challenges, and the need for quality control. Cureus 2023 May;15(5):e39111. [doi: 10.7759/cureus.39111] [Medline: 37332420]
65. Shu T, Xu F, Li H, Zhao W. Investigation of patients' access to EHR data via smart apps in Chinese Hospitals. BMC Med Inform Decis Mak 2021 Jul 30;21(Suppl 2):53. [doi: 10.1186/s12911-021-01425-w] [Medline: 34330258]
66. Pool J, Indulska M, Sadiq S. Large language models and generative AI in telehealth: a responsible use lens. J Am Med Inform Assoc 2024 Sep 1;31(9):2125-2136. [doi: 10.1093/jamia/ocae035] [Medline: 38441296]
67. Lei R, Shen Q, Yang B, et al. Core outcome sets in child health: a systematic review. JAMA Pediatr 2022 Nov 1;176(11):1131-1141. [doi: 10.1001/jamapediatrics.2022.3181] [Medline: 36094597]
68. Meyers N, Glick AF, Mendelsohn AL, et al. Parents' use of technologies for health management: a health literacy perspective. Acad Pediatr 2020;20(1):23-30. [doi: 10.1016/j.acap.2019.01.008] [Medline: 30862511]

Abbreviations

BCT : behavior change technique
GRADE: Grading of Recommendations Assessment, Development, and Evaluation
IQR: Interquartile Range
MeSH: Medical Subject Headings
mHealth : mobile health
OR: odds ratios
PICC: peripherally inserted central catheter
PRISMA : Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QoL: quality of life
RCT: randomized controlled trial
SMD: standardized mean difference

Edited by JN Siebert; submitted 31.Oct.2025; peer-reviewed by S Zheng, Z Zhao; accepted 17.Feb.2026; published 22.Apr.2026.

Please cite as:
Yao H, Wen Y, Wang H, Xiao Y, Deng M, Yang W, Li Y, Mao X
Effectiveness of Mobile Health Interventions in Pediatric Cancer: Systematic Review and Meta-Analysis of Randomized Controlled Trials
JMIR Mhealth Uhealth 2026;14:e86836
URL: https://mhealth.jmir.org/2026/1/e86836
doi:10.2196/86836

© Huilu Yao, Yiting Wen, Hongxiu Wang, Ying Xiao, Meiling Deng, Wei Yang, Yuqin Li, Xiaorong Mao. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 22.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effectiveness of mHealth-Based Nutritional Interventions on Iron Status of Pregnant Women: Systematic Review of Randomized Controlled Trials

Saba Abraham Belay[1,2], MPH; Afework Mulugeta Bezabih[3], PhD; Wim Van Petegem[4], Prof Dr; Christophe Matthys[1,5,6], Prof Dr

[1]Department of Chronic Diseases and Metabolism, Clinical and Experimental Endocrinology Unit, KU Leuven, Herestraat 49 ON1 Bus 902, Leuven, Belgium

[2]Continuing Professional Development Center, College of Health Sciences, Mekelle University, Mekelle, Ethiopia

[3]Department of Nutrition and Dietetics, School of Public Health, College of Health Sciences, Mekelle University, Mekelle, Ethiopia

[4]Engineering Technology Education Research (ETHER) Unit, Group T Leuven Campus, Faculty of Engineering Technology, KU Leuven, Leuven, Belgium

[5]Department of Endocrinology, University Hospitals Leuven, KU Leuven, Leuven, Belgium

[6]Division of Human Nutrition, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

Corresponding Author:
Christophe Matthys, Prof Dr
Department of Chronic Diseases and Metabolism, Clinical and Experimental Endocrinology Unit, KU Leuven, Herestraat 49 ON1 Bus 902, Leuven, Belgium

Abstract

Background: Anemia is a global health concern. It is disproportionately prevalent among pregnant women in low-resource regions, where iron deficiency is the leading cause. Given the multifactorial nature of anemia, a range of nutritional interventions is recommended. However, effective implementation is often hindered by limited health care access, poor adherence to supplementation, and gaps in nutrition knowledge and counseling. To address these challenges and optimize hemoglobin (Hb) levels among pregnant women, mobile health (mHealth)−based nutritional interventions offer a promising alternative.

Objective: The aim of the study is to review available evidence on the effectiveness of mHealth-based nutritional interventions on iron status (Hb and/or serum ferritin concentration) among pregnant women.

Methods: Searches were conducted in Embase, CINAHL, Cochrane Library, PubMed, Web of Science, and Scopus, and supplemented by snowballing to identify additional relevant studies from citation lists. The key search strings comprised 4 concepts: “mobile health,” “nutritional intervention,” “Hb, anemia or iron deficiency anemia,” and “pregnant women.” Predefined inclusion and exclusion criteria were applied during screening. The methodological quality of included studies was assessed using the Risk of Bias 2 tool. The primary end point was the change in mean Hb concentration or serum ferritin level. Effect sizes (ESs) were calculated as standardized mean differences, including Cohen *d* and Hedges *g*.

Results: Of the 14,284 studies identified, only 11 randomized controlled trials were included. These studies used various modes of delivery, including mobile phone calls (n=1), SMS text messaging (n=3), and mobile apps (n=4), with some using more than 2 modes (n=3). The effect of mHealth-based nutritional interventions on iron status varied significantly. In total, 4 studies demonstrated a large ES (>0.8), with 3 relying on WhatsApp Messenger as an mHealth delivery mode. Approximately 82% (9/11) of the included studies reported a positive effect (*P* values ranging from <.001 to .047) of the intervention on Hb level, whereas 2 studies reported no statistically significant association (*P*=.33 and *P*=.35, respectively). Notably, interventions with the largest ES achieved clinically significant improvements in Hb concentration, with within- and between-group differences exceeding 1 g/dL. However, including behavioral change theories and nutrition-sensitive components was not consistently associated with larger ESs. Due to high heterogeneity (I^2>95%), attributed to variations in mHealth delivery modes, functions, and interactive features across the included studies, meta-analysis could not be performed.

Conclusions: This review demonstrates that mHealth-supported nutritional interventions effectively optimize Hb concentration in pregnant women. While SMS text messaging was less effective in improving Hb concentration, combining it with another mHealth delivery mode, such as phone calls, improved intervention effectiveness. However, the variability in mHealth delivery modes, functions, and interactive features underscores the need for tailored strategies that account for context-specific challenges, digital literacy, and access to technology to enhance effectiveness.

Trial Registration: PROSPERO CRD42025627769; https://www.crd.york.ac.uk/PROSPERO/view/CRD42025627769

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2026;14:e81001) doi:10.2196/81001

KEYWORDS

mHealth; mobile health; anemia; pregnancy; systematic review; randomized controlled trials

Introduction

Anemia during pregnancy, defined by the World Health Organization as a hemoglobin (Hb) concentration below 11 g/dL, remains a critical public health problem [1,2]. Globally, anemia affects nearly 36.8% of pregnant women, with the prevalence disproportionately higher in low-resource settings [2-4]. The etiology of anemia is multifactorial, involving nutritional deficiencies, mainly caused by deficiencies in hematopoietic nutrients such as vitamins A, B_2, B_6, B_{12}, C, D, E, and folate, as well as minerals like iron, copper, and zinc; infections; chronic disease; inflammation; and genetic Hb disorders [5-7]. Iron deficiency anemia alone accounts for over half of cases, driven by increased gestational iron demands, poor dietary intake, and dietary habits such as inadequate consumption of heme iron sources (animal-based foods) and high intake of iron inhibitors, including polyphenols (tea and coffee) and phytates (whole grains) [4-7].

Maternal anemia is associated with adverse perinatal outcomes, including stillbirth, preterm birth, low birth weight, intrauterine growth retardation, and high neonatal mortality rates [8-10]. Furthermore, anemia during pregnancy is linked to neurodevelopmental disorders, attention-deficit/hyperactivity disorder, and intellectual disability in children [11]. In mothers, anemia raises the risk of pre-eclampsia, postpartum depression, and maternal mortality [3].

Recognizing the consequences of anemia, countries have been implementing both nutrition-specific and nutrition-sensitive solutions aimed at preventing and managing the condition. The nutrition-specific interventions focus on the most immediate causes of anemia, particularly inadequate intake of iron and other hematopoietic nutrients. Strategies include adhering to iron and folic acid supplementation, promoting dietary diversity, implementing iron fortification programs, and providing nutritional education and counseling. Conversely, nutrition-sensitive interventions tackle the underlying causes of anemia by addressing food security, ensuring access to adequate health services, and maintaining a safe and hygienic environment [5,12,13].

The effectiveness of iron and folic acid supplementation and dietary intake is closely associated with the dietary behavior of individuals. According to Marijn Stok et al [14], dietary behavior relates to all phenomena related to food choices, eating habits, and dietary intake patterns. Various studies have highlighted the pivotal role of nutritional education and counseling in influencing dietary behaviors, enhancing nutritional knowledge and attitudes, and promoting anemia prevention and management strategies [3,15,16]. A meta-analysis focused on the impact of nutritional education and counseling on iron status of pregnant women in low- and middle-income countries (LMICs) found that women who received nutritional education and counseling experienced an average increase of 0.88 g/dL (95% CI 0.63-1.13 g/dL; $P<.001$) in Hb concentration and a 34% lower risk of anemia compared to the control group [16].

Despite the potential of nutritional education and counseling in improving health and dietary behavior, a review targeting 4 countries (2 in Africa and 2 in Asia) found that pregnant women were not reached frequently enough, were not counseled adequately, and consequently lacked motivation, nutritional knowledge, and optimal dietary practice [17]. These gaps highlight the need for nutritional education and counseling to be supported by innovative solutions, such as mobile health (mHealth), to effectively influence dietary behavior through improving knowledge, attitudes, and dietary practices [5,18].

mHealth, defined as the use of mobile technologies in health care, has emerged as a promising tool for improving the accessibility, affordability, and effectiveness of nutrition care services in both high-income countries and LMICs [18,19]. Furthermore, nowadays, the delivery of health behavior change interventions increasingly relies on mHealth technologies due to their widespread penetration across socioeconomic levels [20]. For example, an umbrella review of 47 meta-analyses, comprising 507 randomized controlled trials (RCTs), documented an impact of mHealth on dietary behavior. According to this review, mHealth interventions have small-to-moderate improvements across various dietary outcomes [21].

However, while several reviews have explored the role of mHealth interventions in improving maternal and child health outcomes, there remains a lack of systematic reviews specifically focusing on mHealth-based nutritional interventions and their impact on the iron status of pregnant women. To our knowledge, no review has identified, appraised, and synthesized the existing evidence focused on nutritional interventions supported by mHealth in optimizing the iron status of pregnant women.

Therefore, this systematic review aims to evaluate the effectiveness of mHealth-based nutritional interventions on the iron status of pregnant women through a synthesis of evidence from existing RCTs. Additionally, this review examined the effects of mHealth-based interventions on secondary outcomes, including adherence to iron and folic acid supplementation, dietary intake, gestational weight gain, and nutritional knowledge, attitudes, and practices, as well as patient-centered implementation outcomes reported in the included studies. The findings are intended to inform and guide the development of future mHealth interventions with relevant evidence to address anemia during pregnancy.

Methods

Overview

The review protocol was registered in PROSPERO (International Prospective Register of Systematic Reviews; CRD42025627769) on January 17, 2025, with prespecified primary and additional

XSL•FO
RenderX

secondary outcomes (adherence to iron and folic acid supplementation, dietary intake, gestational weight gain, and nutritional knowledge and practice). Attitudes and implementation outcomes were introduced later as additional secondary outcomes not included in the original registration. The paper was structured following the updated guidelines for PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Checklist 1) statement and methodological considerations of conducting a systematic review of RCTs [22].

Search Strategy

The searches were conducted in CINAHL (via EBSCOhost), Embase, PubMed (including MEDLINE, via NCBI), CENTRAL (via Cochrane Library), Scopus, and Web of Science Core Collection using a predeveloped search strategy (Multimedia Appendix 1). A combination of indexed terms, database-specific keywords, and MeSH terms was used to improve the search. The search query retrieved all studies that included key terms such as mHealth, nutritional interventions, Hb, anemia or iron deficiency anemia, and pregnant women in either the title, abstract, or keywords. Besides, references and citation lists of all included studies and relevant systematic reviews that met our eligibility criteria were further screened through snowballing.

Eligibility Criteria

The population, intervention, comparison, and outcome scheme was used in this review to formulate the review question and define the eligibility criteria. In addition, the study design, setting, publication language, and period were prespecified as part of the inclusion and exclusion criteria (Multimedia Appendix 2). We included individual and cluster randomized controlled trials (cRCTs), as these study designs are considered the gold standard for clinical research, generating robust and reliable conclusions. This is due to the random allocation of participants to competing interventions and the analytic approaches that support causal inference [23,24]. In contrast, we excluded nonrandomized studies of interventions, including quasi-RCTs, non-RCTs, cohort studies, and case-control studies, as these designs present methodological challenges. Effect estimates obtained from nonrandomized studies of interventions are more subject to additional sources of bias, such as confounding [23,25].

Trials fulfilling the inclusion criteria and published between January 1, 2003, and February 28, 2025 (the date of the final search) were included. All pregnant women were included, regardless of gestational age, Hb level, age, or country of residence. We included studies reporting mean Hb concentration as part of their study outcomes. Thus, we described iron status in terms of mean Hb concentration (expressed in g/dL) or iron deficiency anemia, defined as Hb <11 g/dL or serum ferritin <12 µg/L) [1].

Nutritional interventions delivered either as standalone interventions (using only mHealth components) or as part of a comprehensive intervention package (mHealth combined with other non-mHealth components) were included in the review. We considered various mobile technology tools that facilitate communication in remote areas, including phone calls, video calls, SMS text messages, multimedia messaging services, mobile apps, and social media platforms (eg, WhatsApp, Facebook, Telegram, and TikTok). We also included studies comparing mHealth-based nutritional interventions with standard antenatal care, described as “usual care,” “standard care,” “routine care,” or “standard antenatal care.”

Non–peer-reviewed papers were excluded due to concerns about scientific consistency, methodological compliance, and reliability, which are better ensured in peer-reviewed publications. Nutrition interventions that were supported or delivered by any means other than mHealth technologies were also excluded. Additionally, we excluded studies published before 2003, as the widespread adoption of mHealth-based interventions and the emergence of related publications occurred after this period [26]. This criterion ensures that included studies reflect contemporary mHealth interventions using mobile technologies.

Study Selection and Data Extraction

The retrieved records were imported into EndNote (version 21; Clarivate) software for reference management and duplicate removal. After duplicates were manually verified and removed by 1 author (SAB), the remaining papers were transferred to Rayyan (Qatar Computing Research Institute) for systematic screening.

Two authors (SAB and CM) independently screened the titles and abstracts and categorized papers as relevant (met the eligibility criteria), irrelevant (did not meet the eligibility criteria), and uncertain (inconclusive information on eligibility criteria) following the removal of duplicates. Both authors then reviewed and assessed “potentially relevant” and “uncertain” papers in full text based on the eligibility criteria. Any discrepancies arising during the selection process were resolved through discussion and, when necessary, the involvement of a third reviewer (AMB).

Based on the guidelines in the Cochrane Handbook for Systematic Reviews of Interventions, a comprehensive data extraction form was developed by 1 author (SAB) and subsequently reviewed and refined by 3 authors (CM, WVP, and AMB). The form was piloted on at least 2 included studies to ensure reliability and reproducibility.

The following information was extracted from the included studies:

- Study identifiers: author details and publication details.
- Participant characteristics: age (in years), gestational age (in weeks), and anemia status (Hb concentration in g/dL).
- Study methodology: study setting, sample size and sampling procedures, study design, eligibility criteria, processes for randomization, allocation and blinding, and statistical analysis details.
- Intervention description: characteristics and components of the intervention, intervention delivery methods, duration of follow-up, and details of the comparison intervention.
- Intervention outcomes: attrition rate, continuous data (mean and SD) of Hb concentration, and any reported secondary outcomes.

XSL•FO
RenderX

Following the development of the extraction form, data were extracted by 1 author (SAB) and independently verified by 2 authors (CM and AMB). Discrepancies were resolved through consensus or, if necessary, by consulting a fourth reviewer (WVP). All relevant information was obtained from the full texts of the included studies. Corresponding authors were contacted when clarification was required or when data were missing.

Evaluation of the Methodological Quality of the Studies

Two authors (SAB and CM) independently assessed the methodological quality of the included studies using the revised Risk of Bias 2 tool [27]. In accordance with Risk of Bias 2 guidance, the risk of bias was assessed at the outcome level rather than the study level. Specifically, we evaluated the risk of bias for the results related to the primary outcome of interest.

The tool uses five domains to incorporate all types of bias currently considered to affect the results of RCTs. The domains include (1) risk of bias arising from the randomization process, (2) risk of bias due to deviations from the intended interventions, (3) risk of bias due to missing outcome data, (4) risk of bias in measurement of the outcome, and (5) risk of bias in selection of the reported result, followed by the assessment of the overall risk of bias. On this basis, studies were categorized as having a “high” risk of bias (high risk in at least 1 domain or some concerns across multiple domains), “some concerns” (some concerns in at least 1 domain without any domain rated as high risk), or a “low” risk of bias (low risk across all domains). Any disagreements during the quality assessment were resolved through discussion, involving third and fourth reviewers (AMB and WVP) as needed. Additionally, an independent reviewer (Kidu Gidey) cross-checked the risk-of-bias judgments within domains.

Dealing With Missing Data

For missing data in the included studies (eg, SDs or means), we first attempted to contact the corresponding authors via email to obtain the required information. If data could not be retrieved, we followed guidance from the Cochrane Handbook to derive missing statistics from other reported measures using the generic inverse variance method [23]. For instance, missing SDs were calculated from CIs for means, SEs, *t* statistics, or *P* values when available. In addition, when data were symmetrically distributed, reported medians were considered as a reasonable substitute for means. Studies with missing data were not directly excluded; instead, they were considered in sensitivity analyses to assess the robustness of the findings. We excluded studies with missing data that were irretrievable and judged to be at high risk of bias from effect size (ES) estimation.

Data Synthesis

The characteristics of mHealth interventions are described in terms of mHealth function, interaction feature, and delivery mode.

mHealth Function

We categorized mHealth functions in each study following the classification proposed by Knop et al [28], which outlines 12 mHealth functions. Among these, we focused on the “client education and behavior change communication” function [28]. To systematically analyze the behavior change components within this function, we mapped each intervention to the corresponding behavior change techniques (BCTs) derived from the BCT Taxonomy v1, which includes 93 hierarchically clustered techniques [29]. This approach enables integration of the mHealth function classification with specific and theory-based BCTs.

mHealth Interaction Features

In addition to describing the mHealth feature, it is crucial to examine the level of interaction between the patient or participants and these features. As reported by Knop et al [28], interactions are categorized as unidirectional (1-way communication), bidirectional (2-way communication), and multidirectional (complex interactions involving multiple stakeholders). Following the description by Donevant et al [30], which aligns with existing terminology, we described these interactions as passive and interactive features. The passive feature does not require any additional response or action from the patient within the mHealth app, including 1-way text messaging and reminders. In contrast, interactive features enable patients to respond or engage on the matter, including interactive prompts and 2-way communication via texting, email, phone calls, and other mobile apps [30].

Measures of Mean Hb Concentration (g/dL)

The primary end point of this review was the change in mean Hb concentration (g/dL). For each included study, we calculated the mean change in Hb concentration (g/dL) from baseline. We assessed within-group differences (pre- vs postintervention) and between-group differences based on postintervention mean Hb concentrations (intervention vs control).

Measures of Statistical Significance (P Value)

For the evaluation of effectiveness, an intervention was classified as effective if the mHealth-based nutritional intervention resulted in statistically significant ($P<.05$) outcomes, as compared to the control group. The intervention was classified as ineffective if there was no statistical difference between the intervention and control groups.

Measures of Intervention Effect

The standardized mean difference, a commonly used ES measure in RCTs, including 6 cRCTs and 5 individual RCTs, was calculated. For consistency across study designs, we estimated the standardized mean difference using both Cohen *d* and Hedges *g*. Cohen *d* is a standard metric for ES estimation, while Hedges *g* is similar but incorporates a correction factor to reduce bias in small sample sizes [31,32].

The ES (Cohen *d*) for continuous outcomes was calculated using the formula: $d=(M1-M2)SDp$, where M_1 and M_2 represent the mean Hb concentrations of the intervention and control groups, respectively, and SD_p is the pooled SD of the 2 groups [31]. Group-level summary statistics, which had already accounted for the study design in the original analysis, were used to estimate the ES in cRCTs. Consequently, no additional clustering adjustments were made.

XSL•FO
RenderX

We summarized the intervention's effect on iron status separately for each study. The magnitude of Hedges *g* is interpreted using Cohen convention, where an ES of <0.20 is small, 0.50 to 0.80 is medium, and scores >0.80 are large [31]. Comprehensive Meta-Analysis software was used to calculate the ES.

Heterogeneity

Heterogeneity was quantified using the chi-square and the I^2 statistic. These measures assess the variability in the intervention effects across studies. Following Cochrane guidelines, a chi-square greater than the *df* and a small *P* value (eg, *P*<.05) indicate evidence of heterogeneity of intervention effects (variation in effect estimates beyond chance) [23]. I^2 values were interpreted as: 0% to 40%: might not be important, 30% to 60%: moderate heterogeneity, 50% to 90%: substantial heterogeneity, and 75% to 100%: considerable heterogeneity [23]. Quantitative synthesis was considered only when heterogeneity was low to moderate. With substantial or considerable heterogeneity, results were synthesized narratively.

Secondary Outcome

Based on a review of the existing literature, we defined the following secondary outcomes. Adherence to iron and folic acid supplementation was defined as the consumption of at least 4 iron-folic acid tablets per week for the recommended period [33]. Dietary intake was defined as food and nutrient consumption at an individual, household, or population level over a period [34,35]. Gestational weight gain was characterized as the recommended range of weight a pregnant woman should gain during pregnancy to optimize maternal and child health outcomes [36]. Nutritional knowledge was defined as the knowledge of nutrition, including the ability to recall nutrition and diet-related terminology [37]. Attitude was described as an individual's feeding or eating behavior influenced by feelings, motivations, perceptions, and thoughts [37]. Practice was operationalized as an individual's actions that could affect his or her nutrition, such as eating, feeding, cooking, and selecting foods [37].

In addition, the following implementation outcomes were defined based on the framework proposed by Proctor et al [38]. Appropriateness was defined as the perceived relevance or compatibility of the intervention for the target groups to address the issue or problem. Acceptability or satisfaction was defined as the perception among the target groups that a given practice or intervention is agreeable, palatable, or satisfactory. Implementation was defined as the cost impact of an implementation effort. Feasibility was defined as the extent to which the newly developed intervention can be successfully used or carried out within a given setting. Adoption or uptake was defined as the intention, initial decision, or action to try or use the intervention. Fidelity was defined as the degree to which an intervention was implemented as intended by its developers. Sustainability was defined as the extent to which a newly implemented intervention is maintained or institutionalized within a service setting over time.

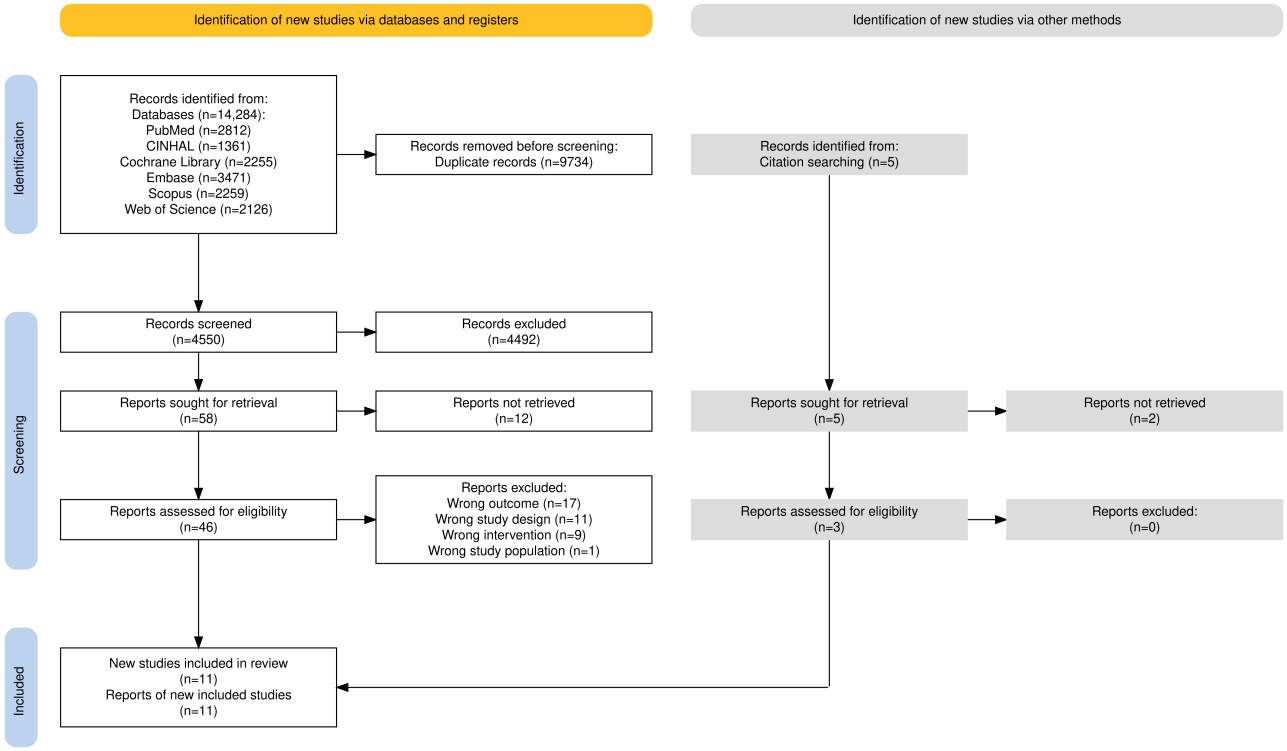
Results

Search Results

As shown in Figure 1 [39], a total of 14,284 studies were extracted from 6 databases, and 5 additional papers through citation searching. We screened titles and abstracts of 4550 papers. Of these, we assessed the full text of 46 papers against the eligibility criteria. Following the full-text assessment, 11 studies published between 2018 and 2023 were included in this review. A total of 38 records were excluded, with detailed reasons for exclusion provided in the PRISMA diagram.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of study selection process.



Study Characteristics

Table 1 summarizes characteristics across 11 included studies. Among these, 6 studies [40-45] were individual RCTs conducted in health facility settings. Of the 5 cRCT studies, only 1 study [46] was conducted at the community level, while the remaining 4 [15,47-49] were conducted in health facilities. In terms of geographical distribution, the majority (7 studies) [15,41,43-46,48] were conducted in low- and lower-middle-income countries, while the remaining 4 studies [40,42,47,49] were conducted in high- and upper-middle-income countries, according to the World Bank classification [50]. Only 2 continents were represented in this systematic review, with Asia contributing 9 (81.8%) studies, and the remaining 2 (18.2%) studies were conducted in Africa. The sample size in the included studies ranged from 59 in Thailand [42] to 413 participants in Nepal [46], with a total of 2191 (mean 199.2, SD 104.1) study participants.

XSL•FO
RenderX

Table . The characteristics of included studies.

Study and publication year	Study location	Study design	Sample size	Inclusion criteria			Exclusion criteria
				Age (years)	Gestational weeks	Hb[a] level (g/dL)	Pregnant women with the following conditions were excluded from the study
Elsharkawy et al (2022) [40]	Saudi Arabia	RCT[b]	• n=196 • Intervention: 98 • Control: 98	18 - 45	14 - 16	<11	Unable to read or write, had multiple fetuses, hereditary anemia, or chronic illnesses, had a Hb level of <7 g/dL, or had been on iron and folic acid supplementation for more than 1 week.
Sontakke et al (2022) [41]	India	RCT	• n=240 • Intervention: 120 • Control: 120	—[c]	13 - 28	8 - 11	With hemoglobinopathies or proven worm infestations.
Singh et al (2020) [46]	Nepal	cRCT[d]	• n=413 • Intervention: 199 • Control: 214	15 - 45	13 - 28	—	—
Abd Rahman et al (2022) [47]	Malaysia	cRCT	• n=104 • Intervention: 50 • Control: 54	20 - 40	13 - 24	7 - 10.9	With underlying chronic diseases, symptoms of anemia, or plans for abortion in the current pregnancy.
Xuto et al (2022) [42]	Thailand	RCT	• n=59 • Intervention: 30 • Control: 29	18 - 40	11-13	—	With a twin pregnancy and unwilling to participate due to threatened abortion or termination of pregnancy.
Sharma et al (2022) [44]	India	RCT	• n=143 • Intervention: 71 • Control: 72	—	<13	—	—
Washington et al (2023) [48]	Liberia	cRCT	• n=150 • Intervention: 75 • Control: 75	—	—	—	With chronic illnesses.
Wakwoya et al (2023) [15]	Ethiopia	cRCT	• n=326 • Intervention:163 • Control: 163	—	<16	—	Had chronic illnesses, refused to provide verbal consent, or intended to leave the study area.

XSL•FO
RenderX

Study and publication year	Study location	Study design	Sample size	Inclusion criteria			Exclusion criteria
				Age (years)	Gestational weeks	Hb[a] level (g/dL)	Pregnant women with the following conditions were excluded from the study
Ahmad et al (2022) [49]	Indonesia	cRCT	• n=110 • Intervention: 55 • Control: 55	—	>28	—	—
Sharma et al (2023) [43]	India	RCT	• n=250 • Intervention: 127 • Control: 123	—	14 - 24	<11	Had a history of obstetric, medical, or psychological problems.
Abujilban et al (2019) [45]	Jordan	RCT	• n=200 • Intervention: 100 • Control: 100	18 - 45	13 - 28	<10.5	Illiterate or had a history of obstetric, medical, or psychological problems.

[a]Hb: hemoglobin.

[b]RCT: randomized controlled trial.

[c]No report.

[d]cRCT: cluster randomized controlled trial.

The age range of participants at baseline was reported in 5 studies [40,42,45-47], ranging from 15 to 45 years. Except for 1 study [48], all studies provided information on the gestation age (in weeks) of the participants. Additionally, 5 studies [40,41,43,45,47] selected the participants based on their Hb levels (Hb <11 g/dL), while the remaining 6 studies did not consider the anemia status of the participants as a selection criterion.

Description of the mHealth Interventions

Descriptions of the mHealth and nutritional interventions from the reviewed studies are summarized in Tables 2 and 3. Among the included studies, 5 studies [40,43,45,47,49] used WhatsApp Messenger to deliver text, audio, and video-based educational messages to pregnant women in the intervention group. Different modes of delivery, including mobile phone calls, SMS text messaging, and mobile apps, were reported. Some studies used more than 1 mode of delivery [43,48,49]. Approximately 91.9% of the included studies used client education and behavior change communication to convey educational information to the target groups.

XSL•FO
RenderX

Table . Summary of mobile health (mHealth) characteristics[a].

Study	Delivery mode	mHealth function	Interaction feature	Intervention (mHealth group) received	Follow-up period (weeks)
Elsharkawy et al [40]	WhatsApp texts	• Education • Reminder[b] • Feedback	Interactive	One educational message, four medication reminders, and more than three participant feedback	12
Sontakke et al [41]	Phone calls	• Reminder • Feedback	Interactive	Biweekly phone calls and reminder notifications	12
Singh et al [46]	SMS or texting	• Education	Passive	One message every 2 weeks (4 - 6 months), weekly thereafter until childbirth	12
Abd Rahman et al [47]	WhatsApp video	• Education	Passive	Daily 3 - to 5-minute educational video (6 days), followed by weekly reminders (weeks 2 - 5)	12
Xuto et al [42]	SMS or texting	• Education	Passive	Two text messages per week (13 - 40 gestational weeks)	28
Sharma et al [44]	Mobile app	• Education • Reminder • Feedback	Passive	Nutritional advice, test tracking, data input, graphical or textual visualization of data, and medication reminder	Until delivery
Washington et al [48]	SMS and phone calls	• Education	Passive	Biweekly “antenatal care telereminder” texts and calls	12
Wakwoya et al [15]	SMS or texting	• Education	Passive	Weekly serial SMS messages	12
Ahmad et al [49]	Phone calls and WhatsApp text	• Education • Feedback	Interactive	Educational flyers via WhatsApp and telephonic nutrition education	12
Sharma et al [43]	SMS or texting, phone calls, and WhatsApp audio	• Education • Reminder • Feedback	Interactive	Four SMS, one WhatsApp audio message, and six weekly calls	4
Abujilban et al [45]	WhatsApp video	• Education	Passive	15-minute educational video	12

[a]All participants in the control groups received routine antenatal care.

[b]Reminder: medication reminder.

XSL•FO
RenderX

Table . Summary of mobile health (mHealth) components, nutritional interventions, and behavior change techniques (BCTs) or behavior change theory applied.

Study	mHealth component	Other components	Nutrition interventions or topics covered by mHealth		BCT taxonomy applied	Behavior change theory
			Nutrition interventions (sensitive)	Nutrition interventions (specific)		
Elsharkawy et al [40]	Educational texts	PowerPoint slides, guidelines, brochures	IFAS[a], dietary advice, and anemia prevention and management	—[b]	Knowledge shaping, prompts and cues, feedback and monitoring, social support	—
Sontakke et al [41]	Phone call reminders	—	IFAS	—	Prompts and cues, goal setting, feedback and monitoring, social support	—
Singh et al [46]	Educational texts	HCP[c] training	General nutrition advice	Health care service use	Knowledge shaping	—
Abd Rahman et al [47]	Educational videos and reminders	—	IFAS, dietary advice, and anemia prevention and management	—	Knowledge shaping, natural consequences	HBM[d]
Xuto et al [42]	Educational texts	—	Dietary advice	—	Knowledge shaping	—
Sharma et al [44]	Apps-based content	—	IFAS, dietary advice, and malaria prevention	Reproductive health care	Knowledge shaping, feedback and monitoring, prompts and cues, social support	—
Washington et al [48]	Educational texts and calls	HCPs or guideline manual	IFAS, dietary advice, malaria prevention, and intestinal parasite control	Hygiene and health care service use	Knowledge shaping, prompts and cues, social support	—
Wakwoya et al [15]	Educational texts	Face-to-face nutritional counseling and brochures	IFAS, dietary advice, use of iodized salt, malaria, and intestinal parasite prevention and control	Hygiene and health care service use	Knowledge shaping	HBM
Ahmad et al [49]	Educational calls and texts	Nutrition booklet and food monitoring card	IFAS and dietary advice	Hygiene	Knowledge shaping, prompts and cues, feedback and monitoring, social support	TPB[e]
Sharma et al [43]	Reminder calls, texts, and audio messages	—	IFAS	—	Feedback and monitoring, prompts and cues	—
Abujilban et al [45]	Educational videos	—	Iron supplementation and dietary management	—	Knowledge shaping	—

[a]IFAS: iron and folic acid supplementation.

[b]No report.

[c]HCP: health care provider.

[d]HBM: health belief model.

[e]TPB: theory of planned behavior.

Regarding mHealth interaction features, 7 studies [15,42,44-48] incorporated a passive feature (unidirectional communication approach), using push technology to deliver educational content or reminders to the target groups. In contrast, 5 of the studies [40,41,43,44,49] used an interactive feature (bidirectional communication; Table 2).

The follow-up duration across studies ranged from 4 weeks [43] to 28 weeks [42], with a mean of 12.8 (SD 5.9) weeks estimated from 10 studies. One study [44] did not provide a clear

XSL•FO
RenderX

intervention duration. Approximately 82% of studies implemented an intervention follow-up period of 12 weeks.

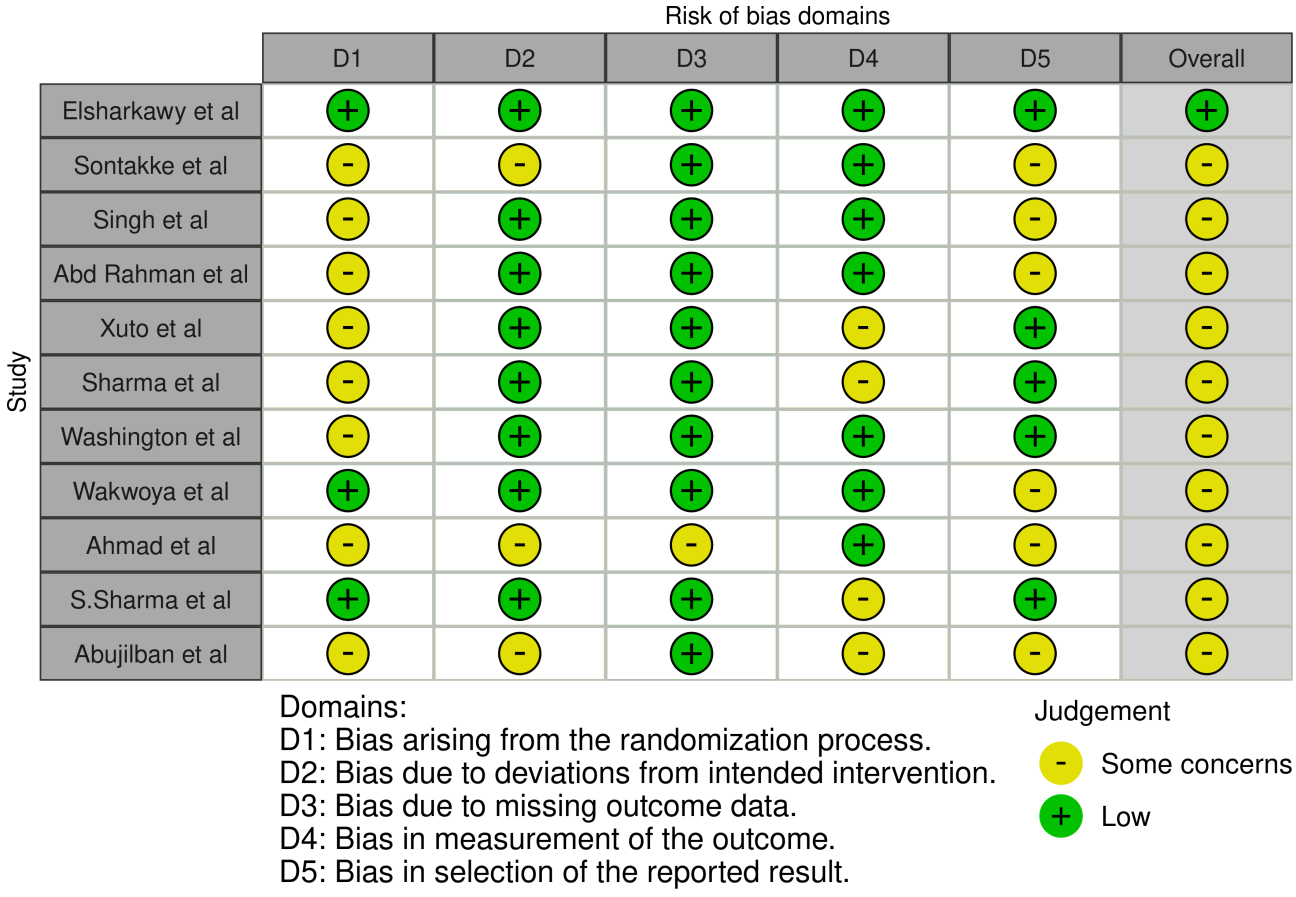
This review also identified a range of BCTs across the included studies, as presented in Table 3. Approximately 54.5% of the studies [40,41,43,44,48,49] applied 2 or more BCTs, with the most frequently used techniques including shaping knowledge, prompts and cues, and feedback and monitoring. Additionally, only 3 studies [15,47,49] incorporated a behavior change theory, such as the health belief model and the theory of planned behavior.

In total, 5 studies [15,40,46,48,49] supplemented their mHealth interventions with additional strategies such as face-to-face presentations, nutritional counseling, guideline and brochure distribution, and health care provider capacity-building training sessions to enhance intervention effectiveness. All studies incorporated nutrition-specific topics, conveying advice on key areas such as iron and folic acid supplementation, dietary intake, and strategies for preventing and managing anemia, malaria, and intestinal parasites. However, nutrition-sensitive topics related to health care service use, hygiene, and reproductive health were covered in only 5 studies [15,44,46,48,49].

Risk of Bias in Included Studies

The methodological quality of the studies varied noticeably. A total of 90.9% of studies were judged to have some concern regarding the risk of bias, with only 1 study classified as having a low risk of bias. Approximately 8 (72.7%) of the studies did not report the method used for allocation sequence concealment, and more than half (54.5%) demonstrated a bias in the selection of reported results. Blinding practice differed across the included studies. In total, 5 studies [15,40,42,47,48] were able to blind participants. The risk of bias for each included study is described in Figure 2 [51], and detailed descriptions of the risk of bias judgments are provided in Multimedia Appendix 3 [15,40-49].

Figure 2. Risk of bias plot for included randomized controlled trials [15,40-49].

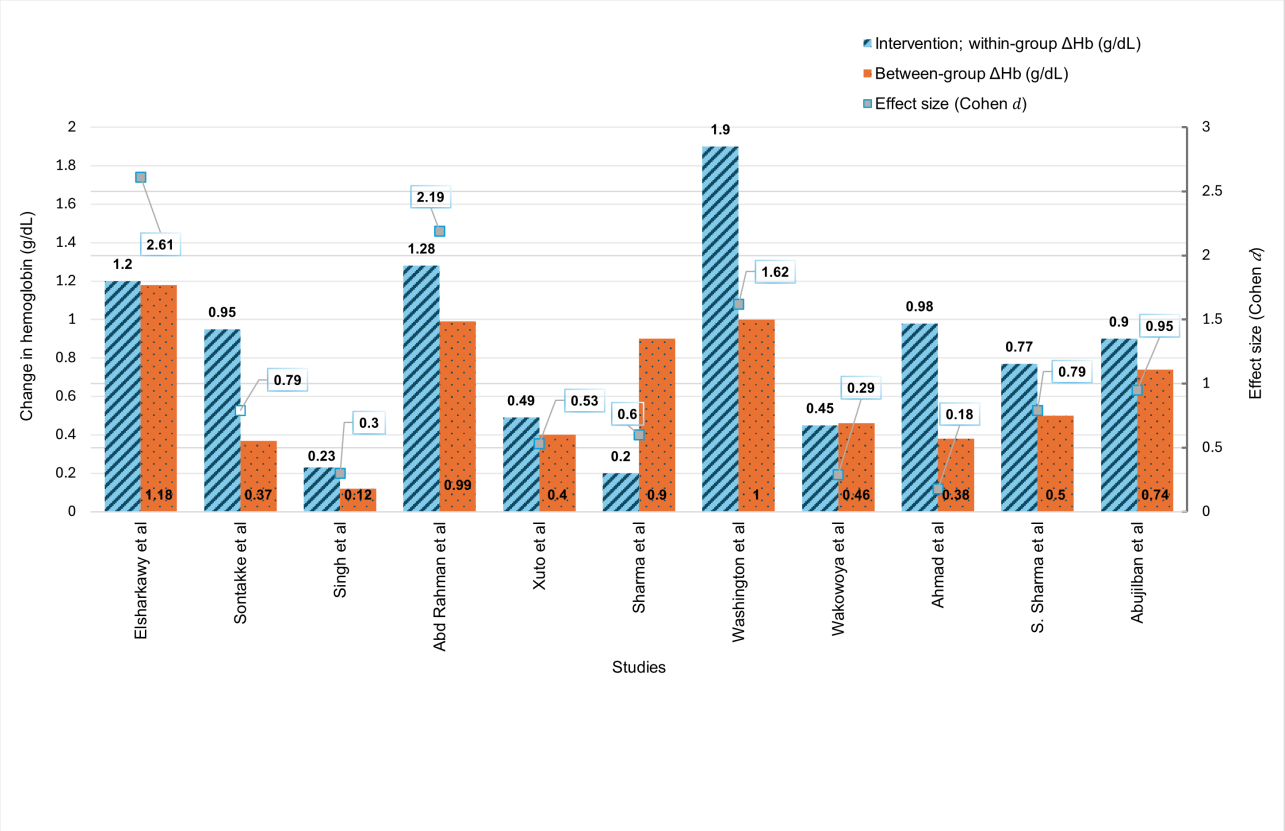


Effects of mHealth-Based Nutritional Interventions

Of the 11 included studies, 10 reported participants' Hb concentration (g/dL) using mean and SD to describe iron status, while 1 study [43] used the median and IQR to report pre- and postintervention changes in Hb concentration. As Figure 3 shows, the highest mean Hb concentration difference within the intervention group (1.9 g/dL) was reported by Washington et al [48], while the smallest difference (0.2 g/dL) was observed in the study of Sharma et al [44]. The smallest between-group difference (intervention vs control) was 0.19 g/dL [49], whereas the largest was 1.18 g/dL [40]. Notably, interventions with the largest ES (eg, ES=2.61, 95% CI 2.23-3.00; ES=2.19, 95% CI 1.70-2.68; ES=1.62, 95% CI 1.25-1.99) achieved clinically meaningful improvements in Hb concentration. These increases exceeded the widely recognized 1 g/dL threshold, indicative of a positive therapeutic response to anemia in pregnant women, and are associated with significant improvements in maternal health outcomes [52,53].

XSL•FO
RenderX

Figure 3. Mean hemoglobin change (ΔHb) within and between groups (g/dL) and corresponding effect sizes (Cohen *d*) across included studies [15,40-49].



Almost 82% (9/11) of studies showed a positive association between intervention and Hb concentration. However, 2 studies [42,49] with small ES (Cohen *d*=0.5) found no association between the mHealth intervention and the Hb concentration in the intervention group (*P*=.33 and *P*=.35, respectively; Table 4).

XSL•FO
RenderX

Table . Summary of study effect size for hemoglobin level (g/dL).

Studies	Mode of delivery	Control			Intervention				Effect size[a]	
		Sample size	Preintervention	Postintervention	Sample size	Preintervention	Postintervention		Cohen *d* with 95% CI	Hedges *g*
		n (%)	Mean (SD)	Mean (SD)	n (%)	Mean (SD)	Mean (SD)	*P* value		
Elsharkawy et al [40]	WTM[b]	98 (9)	9.99 (0.32)	10.01 (0.37)	98 (8.8)	9.96 (0.30)	11.16 (0.50)	.001	2.61 (2.23 to 3.00)	2.60
Sontakke et al [41]	MPC[c]	120 (11.1)	9.48 (0.68)	10.06 (0.70)	120 (10.8)	9.74 (0.68)	10.69 (0.89)	.001	0.79 (0.52 to 1.05)	0.78
Singh et al [46]	TM[d]	199 (18.4)	11.00 (1.15)	11.11 (1.10)	214 (19.3)	11.2 (1.08)	11.43 (1.00)	.02	0.30 (0.11 to 0.50)	0.30
Abd Rahman et al [47]	WVM[e]	50 (4.6)	10.12 (0.66)	10.41 (0.44)	54 (4.9)	10.20 (0.51)	11.48 (0.53)	.001	2.19 (1.70 to 2.68)	2.17
Xuto et al [42]	TM	29 (2.7)	11.66 (1.01)	11.75 (1.37)	30 (2.7)	11.95 (0.68)	12.44 (1.25)	.33[f]	0.53 (0.01 to 1.05)	0.52
Sharma et al [44]	Mapp[g]	72 (6.6)	11.60 (1.30)	10.90 (1.40)	71 (6.4)	11.30 (1.50)	11.50 (1.40)	.03	0.43 (0.10 to 0.76)	0.43
Washington et al [48]	TM and PC[h]	75 (6.9)	10.80 (0.92)	11.70 (0.65)	75 (6.8)	10.90 (0.94)	12.80 (0.71)	.047	1.62 (1.25 to 1.99)	1.61
Wakwoya et al [15]	TM	163 (15)	12.20 (1.44)	12.19 (1.16)	163 (14.7)	12.08 (1.15)	12.53 (1.18)	.01	0.29 (0.07 to 0.51)	0.29
Ahmad et al [49]	PC and TM	55 (5.1)	11.35 (1.25)	11.95 (1.16)	55 (5)	11.16 (0.90)	12.14 (0.96)	.35[f]	0.18 (−0.20 to 0.55)	0.18
Sharma et al [43]	TM, PC, and WAM[i]	123 (11.3)	9.93 (0.75)	10.20 (0.60)	127 (11.5)	9.93 (1.05)	10.70 (0.67)	.001	0.79 (0.53 to 1.04)	0.78
Abujilban et al [45]	WVM	100 (9.2)	9.55 (0.86)	9.71 (1.08)	100 (9)	9.66 (0.70)	10.56 (0.96)	.01	0.83 (0.54 to 1.12)	0.83

[a]Effect size (Cohen *d*): 0-0.1: no effect; 0.2-0.4: small effect; 0.5-0.7: intermediate effect; 0. 8->1: large effect.

[b]WTM: WhatsApp text message.

[c]MPC: mobile phone call.

[d]TM: text message.

[e]WVM: WhatsApp video message.

[f]Nonsignificant *P* value ($P \geq .05$).

[g]Mapp: mobile app.

[h]PC: phone call.

[i]WAM: WhatsApp audio message.

The ES varied among studies, ranging from ES=2.61 (95% CI 2.23 - 3.00) [40] to ES=0.18 (95% CI −0.20 to 0.55) [49]. Among the studies with a small ES (Cohen *d*=0.5), the majority used SMS text messages as their mode of delivery [15,46]. Studies with the highest ESs (ES=2.61, 95% CI 2.23 - 3.00 and ES=2.20, 95% CI 1.70-2.68) used WhatsApp as an mHealth delivery mode, incorporating multiple mHealth functions [40,47]. In total, 3 studies used more than 1 mode of delivery, with ES varying across studies. A detailed description of the intervention's effect on participants' iron status is presented in Table 4.

Secondary Outcomes

As presented in Table 5, 10 studies [15,40-48] reported at least 1 of the secondary outcomes relevant to this review. Among these 10 studies, 5 showed a statistically significant effect ($P<.05$) of mHealth intervention on adherence to iron and folic acid supplements in the intervention group, with adherence ranging from 63.8% [43] to 96% [48] in the intervention group.

XSL•FO
RenderX

Table . Summary of secondary outcomes.

Outcome categories and studies	Summary of findings
Adherence to IFAS[a]	
Elsharkawy et al [40]	Higher adherence to IFAS in the intervention group (90.8%) versus the control group (66.4%), *P*<.001
Sharma et al [43]	Higher adherence to IFAS in the intervention group (63.8%) versus the control group (39%), *P*=.001
Abujilban et al [45]	Higher mean IFAS adherence in the intervention group (mean 14.13, SD 2.68) versus the control group (mean 11.45, SD 3.02), *P*<.01
Washington et al [48]	Higher adherence to IFAS in the intervention group (96%) versus the control group (84%), *P*=.02
Gestational weight gain	
Singh et al [46]	Higher mean weight in the intervention group (mean 52.36, SD 6.28 kg) versus the control group (mean 50.94, SD 5.90 kg), *P*=.02
Wakwoya et al [15]	Slightly higher weight gain in the intervention group (9.69 kg) versus the control group (7.75 kg), but not statistically significant, *P*<.27
Xuto et al [42]	Higher mean weight gain in the intervention group (mean 12.77, SD 4.92 kg) versus the control group (mean 11.98, SD 6.42 kg), but not statistically significant, *P*=.60
Dietary intake	
Abd Rahman et al [47]	Higher mean dietary intake in the intervention group (mean 20.55, SD 5.12) versus the control group (mean 15.98, SD 3.68), *P*<.001
Wakwoya et al [15]	Intervention group consumed more: dark green vegetables (mean 17.96, SD 4.23 vs mean 10.37, SD 2.27; *P*<.001), vitamin C-rich foods (mean 5.42, SD 2.01 vs mean 4.20, SD 2.26; *P*=.005), eggs (mean 4.02, SD 1.04 vs mean 3.35, SD 1.10; *P*<.01), and meat, liver, and fish (mean 22.66, SD 4.77 vs mean 18.06, SD 3.36; *P*<.001) than the control group
Washington et al [48]	Intervention group consumed more fruits (91% vs 72%; aOR[b] 3.77, 95% CI 1.49-9.54; *P*=.003), vegetables (96% vs 87%; aOR 3.69, 95% CI 0.97-14.00; *P*=.04), and fish, meat, or poultry (93% vs 83%; aOR 2.93, 95% CI 0.99-8.70; *P*=.04)
Nutritional knowledge	
Elsharkawy et al [40]	Higher mean knowledge score in intervention group (mean 54.33, SD 10.92) versus the control group (mean 24.79, SD 10.19), *P*<.001
Abd Rahman et al [47]	Higher mean knowledge score in intervention group (mean 26.41, SD 2.17) versus the control group (19.10, SD 3.44), *P*<.001
Abujilban et al [45]	Higher mean knowledge score in intervention group (mean 35.33, SD 9.55) versus the control group (mean 18.74, SD 9.55), *P*<.01
Attitude and practice	
—[c]	No included studies reported on this outcome
Implementation outcomes	
Sharma et al [44]	About 92.5% of pregnant women using the SwasthGarbh app expressed complete satisfaction

[a]IFAS: iron and folic acid supplement.

[b]aOR: adjusted odds ratio.

[c]Not available.

Among studies that evaluated the effect of mHealth intervention on gestational weight, only 1 study [46] reported a statistically significant association (*P*=.02). However, according to Wakwoya et al [15] and Xuto et al [42], although weight gain was slightly higher among pregnant women in the intervention group compared to the control group, the difference was not statistically significant (*P*=.27 and *P*=.60, respectively). Only 1 study [44] measured satisfaction levels among pregnant women adhering to the mHealth intervention. However, no studies provided data on maternal nutrition attitudes and practices or other patient-centered implementation outcomes, such as appropriateness, affordability, feasibility, acceptability,

XSL•FO
RenderX

adoption, fidelity, sustainability, or the cost of mHealth intervention.

Heterogeneity

Considerable heterogeneity was observed across studies, likely attributable to variance in participant characteristics (maternal age, gestational age, and Hb status), intervention features (mHealth delivery mode, mHealth interaction, and function), and socioeconomic and geographical contexts. Although a random-effects meta-analysis was conducted, heterogeneity was extreme (I^2=95%; χ^2_{10}=196.7; P<.01), and the wide prediction interval (−0.92 to 2.82) restricted the interpretability of pooled effect estimates. Additionally, sensitivity analyses were performed to assess the influence of outlining studies, including those with large ESs and those reporting nonstatistically significant results. However, exclusion of these did not reduce heterogeneity, which remained high (I^2>88%), nor did it substantially change the direction or statistical significance of the pooled effect. Therefore, we provided a narrative synthesis following guidance from the Cochrane Handbook for summarizing findings when meta-analysis was not feasible [23]

Discussion

Principal Findings

To the best of our knowledge, this is the first systematic review comprehensively assessing the effectiveness of mHealth-based nutritional interventions on maternal iron status. In this systematic review, 9 of 11 studies revealed a positive effect of mHealth-based nutritional intervention on Hb concentration during pregnancy. However, the ES varied from large ES (Cohen d>0.8) to small ES (Cohen d<0.5). Among the studies with a large ES, 3 used WhatsApp as an mHealth mode of delivery. These studies were conducted in high and upper-middle-income countries, where smartphone availability and access to data bundles are high, enabling mHealth interventions via WhatsApp platforms. In contrast, in LMICs, mHealth interventions mainly rely on SMS text messaging or telephone calls, as these delivery modes are more accessible through basic mobile devices [54,55]. The limited smartphone ownership, poor internet connectivity, and lower digital literacy in LMICs may contribute to reduced engagement with more interactive mHealth platforms.

According to a systematic review by Kante and Målqvist [56], the effectiveness of SMS-based interventions in LMICs varied across studies and settings, largely due to the differences in intervention type, content, frequency, and implementation approach. Some studies reported positive effects of SMS text messaging–based interventions on increasing the use of maternal and child health care services and improving adherence to iron supplementation among pregnant women [57,58]. These studies also identified SMS text messaging as a low-cost, easily personalized intervention that can be sent directly to target groups [57,58]. However, other studies found no significant effect of SMS text messaging on maternal health outcomes, including iron status during pregnancy [59].

Consistent with these findings, in our review, SMS text messaging was the most common mHealth delivery mode, used in 5 studies [15,42,43,46,48] conducted in LMICs. Despite its applicability, texting was associated with lower effectiveness than phone calls or WhatsApp-based interventions. This reduced effectiveness may be due to the limited educational content that can be conveyed via SMS text messaging and the lack of multimedia support, which could aid in information retention. Notably, the findings from this review suggest that combining SMS text messaging with other mHealth delivery modes, such as phone calls, significantly enhances intervention effectiveness. For instance, studies by Washington et al [48] and Sharma et al [43] reported ES of 1.61 (95% CI 1.25-1.99) and 0.79 (95% CI 0.53-1.04), respectively, when SMS text messaging was combined with phone calls.

Moreover, a study suggested that combining SMS text messaging with an interactive feature could improve the effectiveness of mHealth interventions [59]. Similarly, the highest ES (2.61, 95% CI 2.23-3.00) in this review was observed in a study that used text-based education via WhatsApp, supplemented by interactive features or 2-way communication [40].

In total, 3 studies [40,47,48] demonstrating the highest ES also showed clinically significant improvements in mean Hb concentration, with differences greater than 1 g/dL within- and between-group comparisons. These studies conveyed nutritional advice through different delivery modes, in which 2 studies [40,48] reported higher adherence to iron and folic acid supplements, and the third study [47] found higher dietary iron intake among intervention groups adhering to mHealth-supported nutrition interventions. Furthermore, the mean Hb concentrations observed in these 3 studies were notably higher than the pooled results from 2 meta-analyses, which found mean Hb differences of 0.46 and 0.89 g/dL in pregnant women taking iron with or without folic acid, respectively [60,61]. Similarly, a review by Engidaw et al [16] reported a mean Hb difference of 0.88 g/dL based on pooled results from 39 studies, of which only 8 used an mHealth-based approach. While it is acknowledged that comparing individual study results with pooled meta-analysis data has limitations, the higher mean Hb concentration difference observed in this review may be attributed to the nutritional interventions supported by mHealth technologies. This technological support may have improved participants' adherence to the interventions, thereby enhancing the intervention outcomes.

All 11 studies included in this review focused on nutrition-specific interventions, which address the most proximal causes of anemia, such as dietary intake and iron and folic acid supplementation. Nonetheless, 5 studies [15,44,46,48,49] incorporated nutrition-sensitive interventions, which address underlying determinants of anemia, such as maternal health care access and hygiene practices. Notably, this review found that studies incorporating nutrition-sensitive components did not demonstrate higher ES compared to those focusing solely on nutrition-specific interventions. This may be due to 2 key factors. First, the complex and multifactorial etiology of anemia may hinder the direct effects of nutrition-sensitive interventions, especially within short study time frames [6]. Second, mHealth

XSL•FO
RenderX

platforms may not be optimally suited for delivering certain nutrition-sensitive interventions, such as maternal reproductive health services that require in-depth counseling, physical demonstrations by skilled health care providers, or access to physical resources [62].

Additionally, the meta-review by Moorthy et al [13] synthesized data from 118 systematic reviews to assess the impact of nutritional interventions on Hb concentration and anemia outcomes in specific populations, including pregnant women. This comprehensive analysis highlighted the effectiveness of key nutrition-specific interventions, such as iron and folic acid supplementation and intermittent preventive treatment in pregnancy for malaria, in optimizing Hb concentration among pregnant women. However, only 5 of the 118 systematic reviews focused on the impact of nutrition-sensitive interventions on Hb concentration and anemia prevalence, indicating a knowledge gap in the evidence [13]. This paucity of data constrains our ability to draw robust conclusions about the effectiveness of nutrition-sensitive interventions in enhancing maternal iron status.

The effectiveness of mHealth-based nutritional interventions may also rely on the integration of BCTs, which refer to the specific techniques used in the intervention to promote behavior change [63,64]. However, despite their broad use in health promotion, only 3 (27.3%) studies in this review explicitly used existing behavior change theories such as the health belief model or the theory of planned behavior. Notably, one study incorporating the health belief model reported a small ES (ES=0.29, 95% CI 0.07-0.51), while another study with the largest ES (ES=2.61, 95% CI 2.23-3.00) did not apply any behavior change theories [15,40]. This discrepancy suggests that the effectiveness of mHealth interventions may not rely solely on the inclusion of a single behavior change theory or BCTs but rather on the integration of diverse theories or techniques alongside optimal mHealth delivery modes [64,65]. These findings align with systematic reviews, emphasizing that combining multiple BCTs, such as feedback and monitoring, prompts and cues, personalization, and goals and planning, enhances intervention outcomes [66,67].

Limitations of the Review

A key limitation of this review is the inability to include a meta-analysis due to substantial heterogeneity across studies, including variations in mHealth delivery modes, intervention content, and types of nutritional interventions. As a result, a formal statistical assessment of publication bias (eg, funnel plot analysis) could not be conducted. Nevertheless, studies with null or nonsignificant findings may be underrepresented due to publication bias.

This review was restricted to peer-reviewed studies published in English, potentially leading to language and publication bias by excluding relevant evidence from the gray literature. Additionally, interrater reliability statistics (eg, Cohen κ) were not calculated, as data extraction was conducted by only 1 reviewer (SAB).

Furthermore, many of the included studies (90.9%) were judged to have some concerns regarding the risk of bias, potentially limiting the quality of evidence. Finally, this review did not focus on mHealth interventions designed for health care providers, which may have influenced maternal health outcomes indirectly by enhancing maternal nutrition care and adherence to antenatal guidelines.

Conclusions

This systematic review demonstrates that mHealth-supported nutritional interventions can effectively optimize Hb concentrations in pregnant women. Interventions using WhatsApp showed the most significant impact, potentially due to their capacity to deliver multimedia-rich content, thereby facilitating better information retention. Conversely, interventions relying solely on SMS text messaging were less effective; however, combining SMS text messaging with other delivery modes, such as phone calls, improved overall effectiveness.

The integration of mHealth interventions into maternal health care for anemia prevention and management is both feasible and supported by evidence. Nevertheless, the variability in mHealth delivery modes, functions, and interactive features underscores the need for tailored strategies that consider context-specific challenges, digital literacy levels, and access to technology to increase effectiveness.

Implications for Practice and Future Research

The current evidence is predominantly generated from studies conducted in a limited number of countries, which constrains the generalizability of findings. Future research should focus on evaluating the effectiveness of mHealth-based nutritional interventions on maternal iron status across diverse geographical and socioeconomic contexts to strengthen the applicability of the conclusions. Additionally, the insights from this review can guide future researchers in understanding the impact of various mHealth delivery modes, functions, and interactive features on the iron status of pregnant women. Such understanding will assist mHealth intervention designers and implementation researchers in adopting and developing context-specific approaches and identifying implementation strategies to improve effectiveness.

Given the importance of patient-centered care, mHealth interventions must be designed with a personalized approach. Thus, future researchers should also emphasize evaluating implementation outcomes to ensure the effectiveness and sustainability of mHealth interventions in real-world settings.

Acknowledgments

The authors wish to thank Thomas Vandendriessche, Norin Hamouda, Krizia Tuand, Chayenne Van Meel, and Jaana Van Overwalle, the reference librarians of the KU Leuven Libraries—2Bergen (Leuven, Belgium), for their help in conducting the systematic literature search.

XSL•FO
RenderX

Funding

SAB received a PhD scholarship from Vlaamse Interuniversitaire Raad—Universitaire Ontwikkelingssamenwerking through the Global Minds Program, funded by the Belgian Directorate General for Development Cooperation and Humanitarian Aid. The funder was not involved in the conceptualization, analysis, interpretation, or writing of this manuscript.

Data Availability

All data generated or analyzed during this review are included within this paper and its supplementary information files.

Authors' Contributions

SAB and CM conceptualized the review, developed the initial search strategy, and conducted the study selection, data extraction, quality assessment, and data analysis. AMB and WVP contributed to the review and interpretation of the findings. SAB drafted the initial manuscript with input from CM. All authors critically revised the manuscript for important intellectual content and approved the final version. All authors had full access to all data in the study and take responsibility for the integrity and accuracy of the data and analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy used for PubMed (including MEDLINE via NCBI), Embase, Web of Science Core Collection, Scopus, CENTRAL (via Cochrane Library), and CINAHL (via EBSCOhost), combining terms related to mobile health, nutritional interventions, iron status, and pregnant women.
[DOCX File, 18 KB - mhealth_v14i1e81001_app1.docx]

Multimedia Appendix 2
Eligibility criteria used to define the research question.
[DOCX File, 18 KB - mhealth_v14i1e81001_app2.docx]

Multimedia Appendix 3
Risk of bias assessment for randomized controlled trials.
[DOCX File, 88 KB - mhealth_v14i1e81001_app3.docx]

Checklist 1
PRISMA checklist.
[DOCX File, 274 KB - mhealth_v14i1e81001_app4.docx]

References

1. Haemoglobin concentration for the diagnosis of anemia and assessment of severity. World Health Organization. 2011 May 31. URL: https://www.who.int/publications/i/item/WHO-NMH-NHD-MNM-11.1 [accessed 2024-12-05]
2. Karami M, Chaleshgar M, Salari N, Akbari H, Mohammadi M. Global prevalence of anemia in pregnant women: a comprehensive systematic review and meta-analysis. Matern Child Health J 2022 Jul;26(7):1473-1487. [doi: 10.1007/s10995-022-03450-1] [Medline: 35608810]
3. Skolmowska D, Głąbska D, Kołota A, Guzek D. Effectiveness of dietary interventions in prevention and treatment of iron-deficiency anemia in pregnant women: a systematic review of randomized controlled trials. Nutrients 2022 Jul 23;14(15):3023. [doi: 10.3390/nu14153023] [Medline: 35893877]
4. Davidson EM, Scoullar MJL, Peach E, et al. Quantifying differences in iron deficiency-attributable anemia during pregnancy and postpartum. Cell Rep Med 2023 Jul 18;4(7):101097. [doi: 10.1016/j.xcrm.2023.101097] [Medline: 37413986]
5. da Silva Lopes K, Yamaji N, Rahman MO, et al. Nutrition-specific interventions for preventing and controlling anaemia throughout the life cycle: an overview of systematic reviews. Cochrane Database Syst Rev 2021 Sep 26;9(9):CD013092. [doi: 10.1002/14651858.CD013092.pub2] [Medline: 34564844]
6. Global anaemia reduction efforts among women of reproductive age: impact, achievement of targets and the way forward for optimizing efforts. World Health Organization. 2020. URL: https://iris.who.int/handle/10665/336559 [accessed 2026-03-13]
7. Chaparro CM, Suchdev PS. Anemia epidemiology, pathophysiology, and etiology in low- and middle-income countries. Ann N Y Acad Sci 2019 Aug;1450(1):15-31. [doi: 10.1111/nyas.14092] [Medline: 31008520]

XSL•FO
RenderX

8. Geta TG, Gebremedhin S, Omigbodun AO. Prevalence and predictors of anemia among pregnant women in Ethiopia: systematic review and meta-analysis. PLoS One 2022;17(7):e0267005. [doi: 10.1371/journal.pone.0267005] [Medline: 35895619]
9. Kassa GM, Muche AA, Berhe AK, Fekadu GA. Prevalence and determinants of anemia among pregnant women in Ethiopia; a systematic review and meta-analysis. BMC Hematol 2017;17(17):17. [doi: 10.1186/s12878-017-0090-z] [Medline: 29075500]
10. Owais A, Merritt C, Lee C, Bhutta ZA. Anemia among women of reproductive age: an overview of global burden, trends, determinants, and drivers of progress in low- and middle-income countries. Nutrients 2021 Aug 10;13(8):2745. [doi: 10.3390/nu13082745] [Medline: 34444903]
11. Nyarko SH, Greenberg LT, Saade GR, et al. Association between iron deficiency anemia and severe maternal morbidity: a retrospective cohort study. Ann Epidemiol 2024 Dec;100(10-5):10-15. [doi: 10.1016/j.annepidem.2024.10.006] [Medline: 39442771]
12. Nutritional anaemias: tools for effective prevention and control. World Health Organization. 2017. URL: https://www.who.int/publications/i/item/9789241513067 [accessed 2026-03-13]
13. Moorthy D, Merrill R, Namaste S, Iannotti L. The impact of nutrition-specific and nutrition-sensitive interventions on hemoglobin concentrations and anemia: a meta-review of systematic reviews. Adv Nutr 2020 Nov 16;11(6):1631-1645. [doi: 10.1093/advances/nmaa070] [Medline: 32845972]
14. Marijn Stok F, Renner B, Allan J, et al. Dietary behavior: an interdisciplinary conceptual analysis and taxonomy. Front Psychol 2018;9:1689. [doi: 10.3389/fpsyg.2018.01689] [Medline: 30298030]
15. Wakwoya EB, Belachew T, Girma T. Effect of intensive nutrition education and counseling on hemoglobin level of pregnant women in East Shoa zone, Ethiopia: randomized controlled trial. BMC Pregnancy Childbirth 2023 Sep 19;23(1):676. [doi: 10.1186/s12884-023-05992-w] [Medline: 37726668]
16. Engidaw MT, Lee P, Fekadu G, Mondal P, Ahmed F. Effect of nutrition education during pregnancy on iron-folic acid supplementation compliance and anemia in low- and middle-income countries: a systematic review and meta-analysis. Nutr Rev 2025 Jul 1;83(7):e1472-e1487. [doi: 10.1093/nutrit/nuae170] [Medline: 39545365]
17. Sanghvi T, Nguyen PH, Tharaney M, et al. Gaps in the implementation and uptake of maternal nutrition interventions in antenatal care services in Bangladesh, Burkina Faso, Ethiopia and India. Matern Child Nutr 2022 Apr;18(2):e13293. [doi: 10.1111/mcn.13293] [Medline: 34816602]
18. Gayesa RT, Ngai FW, Xie YJ. The effects of mHealth interventions on improving institutional delivery and uptake of postnatal care services in low-and lower-middle-income countries: a systematic review and meta-analysis. BMC Health Serv Res 2023 Jun 9;23(1):611. [doi: 10.1186/s12913-023-09581-7] [Medline: 37296420]
19. Mokaya M, Kyallo F, Vangoitsenhoven R, Matthys C. Clinical and patient-centered implementation outcomes of mHealth interventions for type 2 diabetes in low-and-middle income countries: a systematic review. Int J Behav Nutr Phys Act 2022 Jan 6;19(1):1. [doi: 10.1186/s12966-021-01238-0] [Medline: 34991593]
20. McGovern L, O'Toole L, Houshialsadat Z, O'Reilly SL. Women's perspectives on mHealth behavior change interventions for the management of overweight, obesity, or gestational diabetes: a qualitative meta-synthesis. Obes Rev 2024 Aug;25(8):e13761. [doi: 10.1111/obr.13761] [Medline: 38733067]
21. Singh B, Ahmed M, Staiano AE, et al. A systematic umbrella review and meta-meta-analysis of eHealth and mHealth interventions for improving lifestyle behaviours. NPJ Digit Med 2024 Jul 5;7(1):179. [doi: 10.1038/s41746-024-01172-y] [Medline: 38969775]
22. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
23. Higgins JT, Chandler J, Cumpston M, Li T, Page MJ. In: Welch VA, editor. Cochrane Handbook for Systematic Reviews of Interventions Version 65: Cochrane; 2024. URL: https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current [accessed 2026-04-06]
24. Hariton E, Locascio JJ. Randomised controlled trials—the gold standard for effectiveness research: study design: randomised controlled trials. BJOG 2018 Dec;125(13):1716. [doi: 10.1111/1471-0528.15199] [Medline: 29916205]
25. Yao M, Wang Y, Busse JW, et al. Evaluating the impact of including non-randomised studies of interventions in meta-analysis of randomised controlled trials: a protocol for a meta-epidemiological study. BMJ Open 2023 Jul 26;13(7):e073232. [doi: 10.1136/bmjopen-2023-073232] [Medline: 37495391]
26. Alanzi T, Rehman SU, Khan MA, Istepanian RSH. The evolution and mapping trends of mobile health (m-Health): a bibliometric analysis (1997-2023). Mhealth 2024;10(23):23. [doi: 10.21037/mhealth-23-20] [Medline: 39114458]
27. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
28. Knop MR, Nagashima-Hayashi M, Lin R, et al. Impact of mHealth interventions on maternal, newborn, and child health from conception to 24 months postpartum in low- and middle-income countries: a systematic review. BMC Med 2024 May 15;22(1):196. [doi: 10.1186/s12916-024-03417-9] [Medline: 38750486]

XSL•FO
RenderX

29. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
30. Donevant SB, Estrada RD, Culley JM, Habing B, Adams SA. Exploring app features with outcomes in mHealth studies involving chronic respiratory diseases, diabetes, and hypertension: a targeted exploration of the literature. J Am Med Inform Assoc 2018 Oct 1;25(10):1407-1418. [doi: 10.1093/jamia/ocy104] [Medline: 30137383]
31. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition: Routledge; 2013.
32. Durlak JA. How to select, calculate, and interpret effect sizes. J Pediatr Psychol 2009 Oct;34(9):917-928. [doi: 10.1093/jpepsy/jsp004] [Medline: 19223279]
33. Sendeku FW, Azeze GG, Fenta SL. Adherence to iron-folic acid supplementation among pregnant women in Ethiopia: a systematic review and meta-analysis. BMC Pregnancy Childbirth 2020 Mar 4;20(1):138. [doi: 10.1186/s12884-020-2835-0] [Medline: 32131751]
34. Dietary assessment: a resource guide to method selection and application in low resource settings. Food and Agriculture Organization of the United Nations. 2018. URL: https://openknowledge.fao.org/server/api/core/bitstreams/3dc75cfc-9128-4f29-9d76-8d1f792078f0/content [accessed 2026-03-13]
35. O'Hara C, Gibney ER. Dietary intake assessment using a novel, generic meal-based recall and a 24-hour recall: comparison study. J Med Internet Res 2024 Feb 14;26:e48817. [doi: 10.2196/48817] [Medline: 38354039]
36. Truong YN, Yee LM, Caughey AB, Cheng YW. Weight gain in pregnancy: does the Institute of Medicine have it right? Am J Obstet Gynecol 2015 Mar;212(3):362. [doi: 10.1016/j.ajog.2015.01.027] [Medline: 25725659]
37. Gezimu W, Bekele F, Habte G. Pregnant mothers' knowledge, attitude, practice and its predictors towards nutrition in public hospitals of Southern Ethiopia: a multicenter cross-sectional study. SAGE Open Med 2022;10:20503121221085843. [doi: 10.1177/20503121221085843] [Medline: 35356812]
38. Proctor E, Silmere H, Raghavan R, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76. [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
39. Haddaway NR, Page MJ, Pritchard CC, McGuinness LA. PRISMA2020: an R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. Campbell Syst Rev 2022 Jun;18(2):e1230. [doi: 10.1002/cl2.1230] [Medline: 36911350]
40. Elsharkawy NB, Abdelaziz EM, Ouda MM, Oraby FA. Effectiveness of health information package program on knowledge and compliance among pregnant women with anemia: a randomized controlled trial. Int J Environ Res Public Health 2022 Feb 26;19(5):2724. [doi: 10.3390/ijerph19052724] [Medline: 35270420]
41. Sontakke P, Dwidmuthe KS, Kawathalkar A, Bhalerao A. Effect of mobile phone call reminders with standard therapy versus standard therapy alone on compliance with iron supplementation in antenatal women with iron deficiency anemia: a randomized controlled trial. Cureus 2022 Sep;14(9):e29501. [doi: 10.7759/cureus.29501] [Medline: 36299926]
42. Xuto P, Toyohiko K, Prasitwattanaseree P, Sriarporn P. Effect of receiving text messages on health care behavior and state anxiety of Thai pregnant women: a randomized controlled trial. Int J Community Based Nurs Midwifery 2022 Jan;10(1):18-29. [doi: 10.30476/IJCBNM.2021.89364.1604] [Medline: 35005038]
43. Sharma S, Smitha MV, Balakrishnan D. Telephonic intervention to combat non-adherence to oral iron-folic acid supplementation in pregnancy: a randomized controlled trial. Eur J Obstet Gynecol Reprod Biol X 2023 Dec;20:100235. [doi: 10.1016/j.eurox.2023.100235] [Medline: 37736306]
44. Sharma S, Soni S, Kaushik S, et al. SwasthGarbh: a smartphone app for improving the quality of antenatal care and ameliorating maternal-fetal health. IEEE J Biomed Health Inform 2023 Jun;27(6):2729-2738. [doi: 10.1109/JBHI.2022.3211426] [Medline: 36191117]
45. Abujilban S, Hatamleh R, Al-Shuqerat S. The impact of a planned health educational program on the compliance and knowledge of Jordanian pregnant women with anemia. Women Health 2019 Aug;59(7):748-759. [doi: 10.1080/03630242.2018.1549644] [Medline: 30596538]
46. Singh JK, Acharya D, Paudel R, et al. Effects of female community health volunteer capacity building and text messaging intervention on gestational weight gain and hemoglobin change among pregnant women in Southern Nepal: a cluster randomized controlled trial. Front Public Health 2020;8(312):312. [doi: 10.3389/fpubh.2020.00312] [Medline: 32766199]
47. Abd Rahman R, Idris IB, Md Isa Z, Abd Rahman R. The effectiveness of a theory-based intervention program for pregnant women with anemia: a randomized control trial. PLoS One 2022;17(12):e0278192. [doi: 10.1371/journal.pone.0278192] [Medline: 36473006]
48. Washington K, Leila SA, Corazon VCB, Angelina RB, Mark DA. Does mobile phone tele-reminder increased iron status and diet quality of pregnant women availing antenatal care services at Phebe and Charles B. Dunbar Hospitals Bong County, Liberia A cluster randomized control trial. Am J Med Sci Innov 2023;2(2):153-163. [doi: 10.54536/ajmsi.v2i2.1590]
49. Ahmad A, Wagustina S, Dwiriani CM, et al. The efficacy of nutrition education on anemia and upper arm circumference among pregnant women in Aceh Besar District of Indonesia during the COVID-19 pandemic. Jgizipangan 2022 Mar 27;17(1):27-36. [doi: 10.25182/jgp.2022.17.1.27-36]

XSL•FO
RenderX

50. Metreau E, Young KE, Eapen SG. World bank country classifications by income level for 2024-2025. World Bank. 2024. URL: https://blogs.worldbank.org/en/opendata/world-bank-country-classifications-by-income-level-for-2024-2025 [accessed 2026-03-31]
51. McGuinness LA, Higgins JPT. Risk-of-bias VISualization (robvis): an R package and Shiny web app for visualizing risk-of-bias assessments. Res Synth Methods 2021 Jan;12(1):55-61. [doi: 10.1002/jrsm.1411] [Medline: 32336025]
52. Ubom AE, Begum F, Ramasauskaite D, et al. FIGO good practice recommendations on anemia in pregnancy, to reduce the incidence and impact of postpartum hemorrhage (PPH). Int J Gynaecol Obstet 2025 Dec;171(3):993-1007. [doi: 10.1002/ijgo.70529] [Medline: 41031541]
53. Tandon R, Jain A, Malhotra P. Management of iron deficiency anemia in pregnancy in India. Indian J Hematol Blood Transfus 2018 Apr;34(2):204-215. [doi: 10.1007/s12288-018-0949-6] [Medline: 29622861]
54. Materia FT, Smyth JM, Puoane T, et al. Implementing text-messaging to support and enhance delivery of health behavior change interventions in low- to middle-income countries: case study of the Lifestyle Africa intervention. BMC Public Health 2023 Aug 10;23(1):1526. [doi: 10.1186/s12889-023-16388-y] [Medline: 37563595]
55. Demena BA, Artavia-Mora L, Ouedraogo D, Thiombiano BA, Wagner N. A systematic review of mobile phone interventions (SMS/IVR/Calls) to improve adherence and retention to antiretroviral treatment in low-and middle-income countries. AIDS Patient Care STDS 2020 Feb;34(2):59-71. [doi: 10.1089/apc.2019.0181] [Medline: 32049555]
56. Kante M, Målqvist M. Effectiveness of SMS-based interventions in enhancing antenatal care in developing countries: a systematic review. BMJ Open 2025 Feb 25;15(2):e089671. [doi: 10.1136/bmjopen-2024-089671] [Medline: 40000082]
57. Lee SH, Nurmatov UB, Nwaru BI, Mukherjee M, Grant L, Pagliari C. Effectiveness of mHealth interventions for maternal, newborn and child health in low- and middle-income countries: systematic review and meta-analysis. J Glob Health 2016 Jun;6(1):010401. [doi: 10.7189/jogh.06.010401] [Medline: 26649177]
58. Dewi A, Wirasto A, Dewi TS. The use of mobile-phone applications to improve the compliance of pregnant women to take iron tablets: a systematic review. Research Square. Preprint posted online on Jun 7, 2023. [doi: 10.21203/rs.3.rs-3023111/v1]
59. Sondaal SFV, Browne JL, Amoakoh-Coleman M, et al. Assessing the effect of mHealth interventions in improving maternal and neonatal care in low- and middle-income countries: a systematic review. PLoS One 2016;11(5):e0154664. [doi: 10.1371/journal.pone.0154664] [Medline: 27144393]
60. Haider BA, Olofin I, Wang M, et al. Anaemia, prenatal iron use, and risk of adverse pregnancy outcomes: systematic review and meta-analysis. BMJ 2013 Jun 21;346:f3443. [doi: 10.1136/bmj.f3443] [Medline: 23794316]
61. Peña-Rosas JP, De-Regil LM, Garcia-Casal MN, Dowswell T. Daily oral iron supplementation during pregnancy. Cochrane Database Syst Rev 2015 Jul 22;2015(7):CD004736. [doi: 10.1002/14651858.CD004736.pub5] [Medline: 26198451]
62. Labrique AB, Wadhwani C, Williams KA, et al. Best practices in scaling digital health in low and middle income countries. Global Health 2018 Nov 3;14(1):103. [doi: 10.1186/s12992-018-0424-z] [Medline: 30390686]
63. Watson D, Mushamiri P, Beeri P, et al. Behaviour change interventions improve maternal and child nutrition in sub-Saharan Africa: a systematic review. PLOS Glob Public Health 2023;3(3):e0000401. [doi: 10.1371/journal.pgph.0000401] [Medline: 36996036]
64. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4. [doi: 10.2196/jmir.1376] [Medline: 20164043]
65. Cho YM, Lee S, Islam SMS, Kim SY. Theories applied to m-Health interventions for behavior change in low- and middle-income countries: a systematic review. Telemed J E Health 2018 Oct;24(10):727-741. [doi: 10.1089/tmj.2017.0249] [Medline: 29437546]
66. Aguiar M, Trujillo M, Chaves D, Álvarez R, Epelde G. mHealth apps using behavior change techniques to self-report data: systematic review. JMIR Mhealth Uhealth 2022 Sep 9;10(9):e33247. [doi: 10.2196/33247] [Medline: 36083606]
67. Dugas M, Gao GG, Agarwal R. Unpacking mHealth interventions: a systematic review of behavior change techniques used in randomized controlled trials assessing mHealth effectiveness. Digit Health 2020;6:2055207620905411. [doi: 10.1177/2055207620905411] [Medline: 32128233]

Abbreviations

BCT: behavior change technique
cRCT: cluster randomized controlled trial
ES: effect size
Hb: hemoglobin
LMIC: low- and middle-income countries
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial

XSL•FO
RenderX

Edited by M Waring; submitted 21.Jul.2025; peer-reviewed by B Wogayehu, ME Heidari; revised version received 16.Feb.2026; accepted 17.Feb.2026; published 09.Apr.2026.

Please cite as:
Belay SA, Bezabih AM, Petegem WV, Matthys C
Effectiveness of mHealth-Based Nutritional Interventions on Iron Status of Pregnant Women: Systematic Review of Randomized Controlled Trials
JMIR Mhealth Uhealth 2026;14:e81001
URL: https://mhealth.jmir.org/2026/1/e81001
doi:10.2196/81001

© Saba Abraham Belay, Afework Mulugeta Bezabih, Wim Van Petegem, Christophe Matthys. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 9.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

mHealth-Based Gamification Interventions to Promote Health Among Older Adults: Scoping Review

Xi Chen[1*], PhD; Yishu Zhu[1*], MS; Dian Jiang[1], PhD; Jundan Huang[1], PhD; Qi Xie[1], PhD; Yishi Chen[1], MS; Jing Chang[1], MS; Weiping Huang[2], MS; Hongting Ning[3], PhD; Hui Feng[1], PhD

[1]Xiangya School of Nursing, Central South University, No. 172, Tongzipo Road, Yuelu District, Changsha, China

[2]Teaching and Research Section of Clinical Nursing, Xiangya Hospital of Central South University, Changsha, China

[3]School of Nursing, Hunan University of Chinese Medicine, Changsha, China

[*]these authors contributed equally

Corresponding Author:
Hui Feng, PhD
Xiangya School of Nursing, Central South University, No. 172, Tongzipo Road, Yuelu District, Changsha, China

Abstract

Background: Healthy aging has emerged as a global priority. However, older adults' participation in health promotion programs remains low, and traditional health promotion models have achieved limited success in fostering sustained engagement among this population. Mobile health (mHealth)–based gamification interventions offer a promising way to address these challenges. However, no published reviews support or oppose the use of mHealth-based gamification interventions as health promotion strategies in older adults.

Objective: The study aimed to identify mHealth interventions using gamification to promote health among older adults.

Methods: Our scoping review was conducted following the Joanna Briggs Institute recommendations for scoping reviews and Arksey and O'Malley's framework. The process followed PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines and PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Literature Search Extension) checklist. A comprehensive literature search was conducted across 8 databases: PubMed, Scopus, Web of Science, Embase, Cochrane Library, CINAHL, PsycARTICLES, and IEEE Xplore Digital Library, from their inception to December 10, 2025. Two reviewers independently screened titles, abstracts, and full texts via Rayyan, with disagreements resolved by a third reviewer.

Results: This scoping review identified 11 studies. Only 1 article was published before 2022. The interventions were found to improve enjoyment and motivation (n=5), cognitive function (n=3), physical activity (n=2), and digital literacy (n=2). Individual studies also reported improvements in mental health (n=1) and adherence (n=1), a reduction in suicidal ideation (n=1), improvements in physical function (n=1), the promotion of social engagement (n=1), and the identification of mild cognitive impairment (n=1). Game elements used were ranked by frequency as progress, challenges, goals, levels, reward, sensation, storytelling or narration, leaderboard, surprise, and avatar. No research was found to use the game element of "social sharing." mHealth types included augmented and virtual reality–based training systems, wearable devices, mobile phones, tablets, and Windows platforms and devices. Notably, only 4 studies applied theoretical frameworks, and 3 omitted the concrete approach to gamification.

Conclusions: As the first scoping review to identify and map mHealth-based gamification interventions for older adults, this study highlights their potential as an innovative approach to health promotion. By systematically synthesizing evidence regarding intervention designs, gamification strategies, and preliminary health outcomes, it establishes a foundation for future inquiry. However, this review is limited by the small number of included studies, precluding broad generalizations. Future research should assess long-term impacts, integrate theoretical frameworks, establish reporting guidelines, design personalized social-interactive interventions, and expand to broader health domains. Ultimately, these insights provide targeted guidance for developing age-appropriate digital health solutions, contributing to the realization of active aging.

(JMIR Mhealth Uhealth 2026;14:e82368) doi:10.2196/82368

KEYWORDS

mHealth; gamification; health promotion; older adults; health behavior; mobile health

XSL•FO
RenderX

Introduction

Globally, the proportion of older people continues to rise. By 2050, the number of people over the age of 65 years is expected to reach approximately one-sixth of the total population [1]. The aging process often leads to cognitive and physical changes in older adults that, if not addressed, can lead to a variety of health complications, including dementia [2]. Maintaining the health of older people is not only related to the quality of individual life but also the core issue of coping with the social and economic pressure of aging.

Healthy aging has also become a global priority, with the World Health Organization identifying the promotion of older people's health as one of its sustainable development goals [3]. However, the number of older people participating in health promotion programs is small [4,5]. In addition, traditional health promotion models have had limited success in engaging older persons in sustained engagement. Therefore, there is an urgent need for innovative interventions to improve the health behaviors of older persons.

Mobile health (mHealth) is showing its unique potential as an effective way to address these challenges. As a new and rapidly growing field, mHealth enables the implementation of predefined health promotion interventions and practices through mobile devices and their applications [6]. mHealth is an ideal choice for health promotion because of its significant benefits: continuous availability of services, wider access, equity in access to services, personalization of content, lower cost, and increased service capability and efficiency [7]. In recent years, with the increasing penetration of smartphones and the overall increase in digital literacy among the public, public and private health care systems have made significant progress in using mobile apps to address public health challenges [7]. In this context, mHealth is an innovative tool that can greatly broaden access to health promotion services for older people and flexibly adapt to a wide range of health-related needs and contexts.

There is a growing interest in integrating game-like elements into mobile apps, known as gamification, to increase user engagement and enable personalization [8,9]. This review uses a definition provided by Deterding et al [10-12]. They define gamification as the use of game elements (such as goal-setting, progress bars, rewards, feedback, badges, points, competition, cooperation, leaderboards, etc) rather than mature games in nonplayable services and applications to enhance user experience and engagement [10-12]. While game elements are not always clearly distinguishable, they refer to the typical features of a game [10]. To be clear, this review focuses on gamification, not serious games, because serious games provide a pure gaming experience by creating complete and immersive games. In contrast to serious games, this article focuses not on the full game but on the fun elements.

Systems that incorporate gamification often use incentives to promote the adoption of mHealth technologies and healthy behaviors, including immediate success feedback (rewards), continuous progress feedback, and goal setting. For example, users receive badges as rewards when they win, and they can share their achievements on social networks [13]. In rehabilitation and behavioral sciences, gamification has been integrated into mHealth, virtual reality (VR) technology, and robotics and has been effective in motivating patients with cardiovascular disease [14-16], diabetes [17,18], and stroke [19-21] to participate more actively in long-term health management. Multiple studies have shown that mobile app–based gamification interventions can improve patients' adherence to medical protocols and effectively manage their health conditions [22-24]. In addition, a literature review of empirical studies on gamification further confirms the effectiveness of gamification in improving user motivation and engagement [25], which is the core goal of many established behavior change strategies [26-30]. Taken together, gamification seems to be an effective strategy for promoting health [31].

Older people are becoming a growing segment of technology users with unique needs and challenges [32]. Furthermore, in the field of health care, there is a lack of understanding of gamification as a design strategy, and its application is often restricted to the use of video games [33] or blindly incorporating game elements without discrimination to motivate certain behaviors (such as physical activity) [34]. In addition, the concept of gamification is relatively new [10], and research on mHealth-based gamification interventions for the age group 60 years and older is still in the early stages of exploration. Therefore, for older people, more in-depth investigations and customized reviews are necessary. Unfortunately, in existing literature databases, we were unable to search for an English-language scoping review that explored the current use of mHealth-based gamification interventions to promote health in older adults.

To address these gaps, this study adopted a scoping review methodology to identify mHealth interventions using gamification to promote health among older adults. To provide a comprehensive overview, this review specifically analyzes (1) intervention characteristics and the health promotion topics addressed; (2) gamification characteristics; (3) the needs and experiences of older adults regarding these interventions; and (4) current knowledge gaps to provide recommendations for future research.

Methods

Overview

This study was structured according to the framework proposed by Arksey and O'Malley [35], with our methodology being guided by the steps described in the Joanna Briggs Institute manual for key phases, including study selection; data mapping; and the collation, synthesis, and reporting of results [36]. The reporting of this scoping review aligned with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist (Checklist 1) and PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Literature Search Extension) checklist (Checklist 2) [37]. To ensure methodological rigor, inclusion and exclusion criteria were predefined prior to initiating the review process.

XSL•FO
RenderX

Search Strategy

All authors discussed and decided to include the following 8 databases: PubMed, Scopus, Web of Science, Embase, Cochrane Library, CINAHL, PsycARTICLES, and IEEE Xplore Digital Library. The IEEE Xplore Digital Library is a more technology-oriented database that allows us to search for articles in the fields of informatics and engineering. A comprehensive literature search was conducted across all databases, covering the period from their inception to an initial search date of February 18, 2025. This search was subsequently updated by rerunning the same strategy on December 10, 2025, to identify any newly published relevant studies. The language of publication was limited to English to ensure an accurate understanding of the research.

The search strategy included terms across three categories: (1) mHealth (eg, telemedicine OR telehealth OR e-health OR m-Health OR eHealth OR mHealth OR "mobile applications" OR applications OR application OR app OR apps OR online OR mobile OR internet OR "web based" OR Smartphone OR "phone, smart" OR "smart phones" OR smartphones OR "smart phone" OR "phones, smart" OR "cell phone" OR iphone OR android OR iOS OR "Wearable Electronic Devices" OR website OR digital* OR system* OR electronic* OR technolog* OR device OR framework* OR "social support" OR Facebook OR "networks, social"), (2) gamification (eg, Gamification OR "game-based learning" OR gaming OR gamif* OR "game element*" OR game* OR "game mechanic*" OR gameful* OR "game design element*" OR "game-design element*" OR "game interface element*" OR "game feature" OR "game-like element*" OR "videogame element*"), and (3) older adults (eg, "older adult*" OR "older people" OR elderly OR "older population" OR "senior citizen*").

The search terms were derived from Medical Subject Headings (MeSH) and free-text words. These keywords were then combined using Boolean operators (AND, OR) to ensure a comprehensive search. All search algorithms were modified and refined for each database. For a detailed search strategy, see Multimedia Appendix 1. A supplementary manual search was conducted by reviewing the reference lists of all included articles. Additionally, ResearchGate was used as a targeted supplementary tool for two specific purposes: (1) to retrieve the full texts of potentially eligible peer-reviewed articles that were otherwise inaccessible, including directly requesting full texts from the authors; and (2) to review the publication profiles of key authors from the included studies to identify any other relevant peer-reviewed journal publications meeting our inclusion criteria. The inclusion and exclusion criteria are mentioned in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- This review primarily includes interventions based on "gamification" rather than "serious games," both of which are defined in the introduction. We explicitly did not exclude studies simply due to the absence of a clear definition of "gamification" or "serious games" within the text. Instead, we require that gamification be explicitly stated or at least 1 game element used in the article
- Peer-reviewed English-language journal publications with no date limit until the final search date
- Original empirical research, including qualitative and quantitative research (must be experimental research)
- Gamification is delivered through mHealth devices such as computers, smartphones, and wearables
- The purpose of gamification is to promote health in older adults, including but not limited to physical activity, cognitive function, mental health, nutritional intake, and social interaction
- The mean age of the participants included in the article is 60 years or older

Exclusion criteria

- Excluding reviews, protocols, design documents, letters, magazines, dissertations, abstracts, reviews, posters, preprints, conference proceedings, conference abstracts, nonprimary data sources, books, and book chapters
- Excludes serious or mature games (eg, video games, immersive virtual reality games, and augmented reality exergames)
- If the entire article cannot be accessed or retrieved, the study will also be excluded
- Exclude studies that only mention gamification superficially, for example, as a buzzword or as a suggestion for future research, without it being a core component of the study's intervention or analysis

Study Selection

We first used the software EndNote X21 (Clarivate) to automatically identify and remove duplicates. Then, we imported the deduplicated documents from EndNote X21 into Rayyan software and continued to perform manual deduplication in Rayyan software. Prior to the formal review, a calibration exercise was conducted on a random sample of one-fifth of the titles and abstracts. Two reviewers (XC and YZ) independently screened this sample to ensure consistency. Following a precedent set in similar reviews, an interrater agreement of nine-tenths was achieved before proceeding with the screening of the remaining articles [38]. The complete screening was performed independently by the 2 reviewers using Rayyan software. Initially, the 2 authors (XC and YZ) screened all study titles and abstracts for eligibility based on inclusion and exclusion criteria. Next, the 2 authors (XC and YZ) examined all full-text articles to determine whether they included the use of mHealth-based gamification interventions to promote health in older adults. Finally, a supplementary manual search was

XSL•FO
RenderX

performed by looking at the list of references for the included articles and using ResearchGate. To ensure the uniformity of the screening, the authors conducted several training sessions in Rayyan with the co-reviewers. After each stage, the 2 authors (XC and YZ) cross-checked their results and discussed any differences. Any disagreements were resolved by a third researcher (HF).

Data Extraction

Two researchers (XC and YZ) independently used the data extraction table of Excel and extracted and encoded study characteristics (first author name and year of publication, country, study design, and study objectives), participant characteristics (mean age, percentage of female participants, and target population), intervention characteristics (sample size, study setting, type of mHealth used, the theory used, and duration), gamification characteristics (definition of gamification, gamification elements, reasons for using gamification elements, and the concrete approach to gamification), and the needs and experiences of older adults with mHealth-based gamification interventions, key measurements, and study results. We extracted game elements based on the classification in the research by Aschentrup et al [31], including reward, sensation, progress, challenges, surprise, storytelling or narration, social sharing, level, leaderboard, goals, and avatar.

The extracted data were cross-checked between authors (XC and YZ), and any discrepancies within data were discussed. Differences were resolved through consultation with the third author (HF). The first author (XC) synthesized the extracted data for analysis. All data were presented using the data charts suggested by Arksey and O'Malley [35].

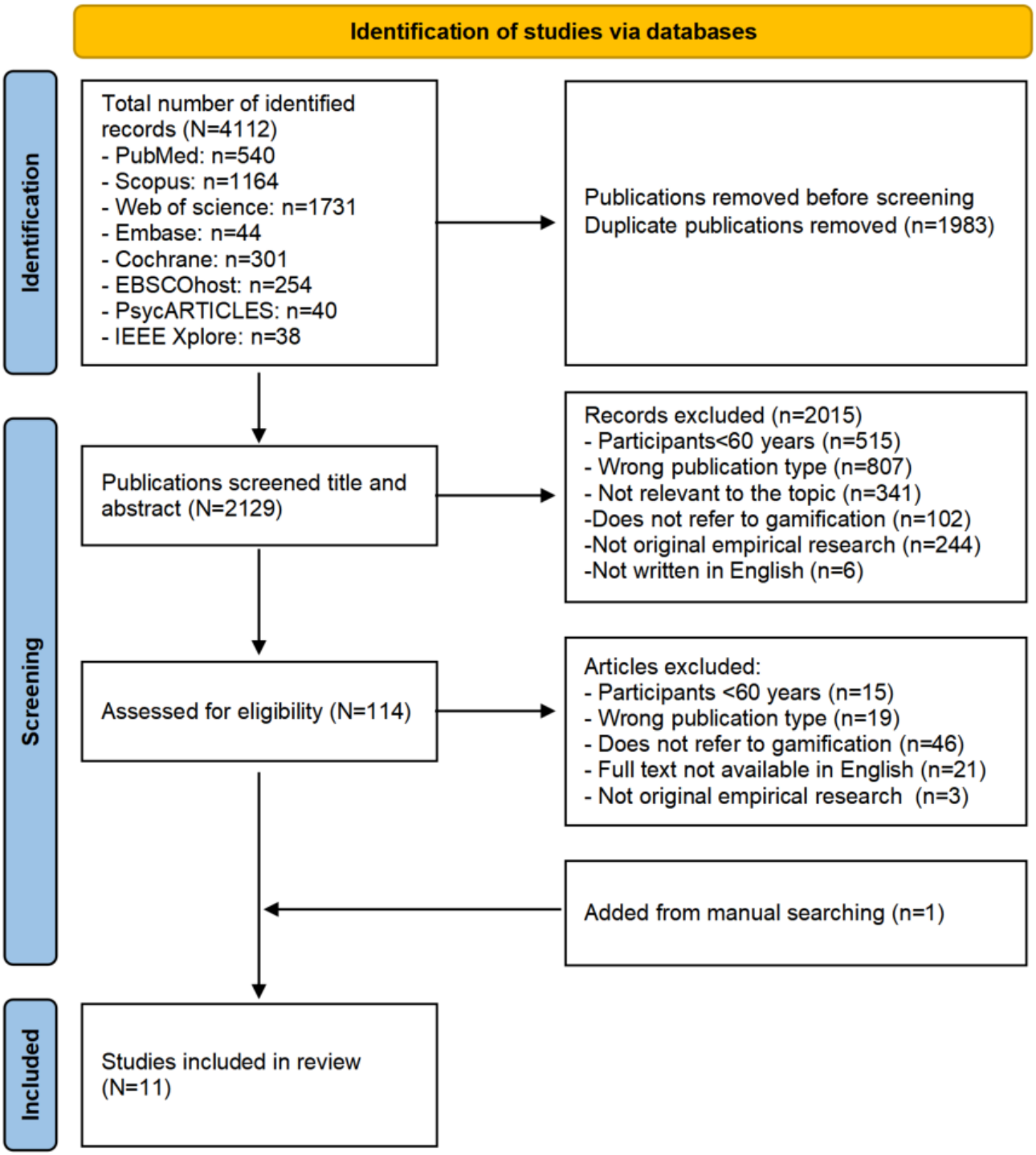
Results

Study Selection

A total of 4112 records were found in our database. We first used the software EndNote X21 to automatically identify and delete duplicates, with 1928 duplicates removed. We then imported the deduplicated documents from EndNote X21 into Rayyan software, and continued to perform manual deduplication, which excluded 55 out of 2184 publications. The authors (XC and YZ) then used Rayyan software to screen all study titles and abstracts of these publications, achieving an interrater agreement rate of 0.978. The screening of titles and abstracts resulted in the exclusion of 2015 studies. The full texts of the remaining articles were then assessed for eligibility, resulting in the inclusion of 10 studies. An additional study was identified through a manual search of reference lists and using ResearchGate, bringing the final number of included studies to 11. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram (Figure 1) provides more details on the screening process and reasons for exclusion.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of article selection.



Study Characteristics

Table S1 in Multimedia Appendix 1 summarizes the research characteristics of the 11 articles included in this study. These studies were published between 2015 and 2025, with only 1 out of 8 articles published before 2022. These studies were distributed in the following 7 countries: China (3/11), South Korea (2/11), Switzerland (1/11), Spain (1/11), the United States (1/11), Japan (1/11), Thailand (1/11), and Malaysia (1/11). The research design included the following 3 types: 4 randomized controlled trials (RCTs), 4 quasi-experiments, 2 mixed methods studies, and 1 qualitative study. The research objectives of the included studies focused on developing mHealth-based gamification interventions for the older population and exploring their acceptability, feasibility, usability, safety, and effectiveness among older adults.

Participant Characteristics

All included studies provided detailed reports on the female proportion, which was relatively high, ranging from 0.476 to 1.000. The 11 included studies recruited diverse populations of older adults. Three studies focused on participants with specific cognitive conditions, including those with early symptoms of memory disorders [39], individuals at risk for Alzheimer disease and related dementias [40], and older adults with no diagnosed cognitive impairments [41]. Another 2 studies targeted

XSL•FO
RenderX

participants with specific physical or mental health statuses: high-risk, postdischarge patients with chronic pulmonary disease [42], and older women experiencing suicidal ideation [43]. A further study recruited hospital inpatients without specifying their condition [44]. The remaining 5 studies did not explicitly specify the health status of their participants.

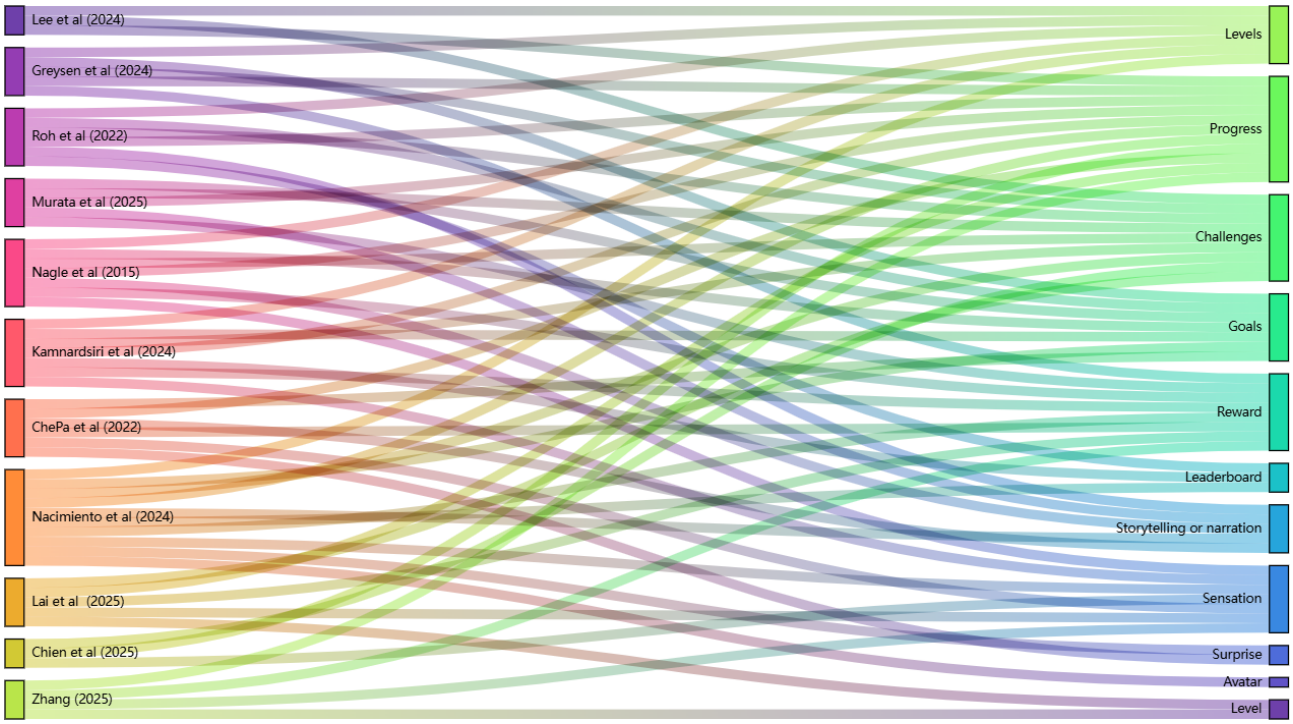
Intervention Characteristics

A total of 545 participants were included in this review, with sample sizes ranging from 12 to 138. The study setting covered communities (n=3), hospitals (n=3), homes (n=2), care centers (n=1), and universities (n=1). One study did not explicitly specify the study setting. The mHealth types comprised augmented and VR-based training systems, wearable devices, mobile phones, tablets, and Windows platforms and devices. Some studies used a single device, while others used a combination of technologies. Furthermore, only 4 studies used theory. These included the application of behavioral economic principles [40], gamification theory [45], technology acceptance model [41], and human-centered design framework [42]. Among all the studies, only 4 studies provided the duration of a single intervention, ranging from 5 minutes to 60 minutes. Three studies provided the intervention frequency, all of which were 3 times a week. Seven studies provided the total duration of the intervention, ranging from 3 to 14 weeks. For detailed information, please refer to Table S1 in Multimedia Appendix 1.

Gamification Characteristics

Of the 11 included studies, only 5 provided a definition of gamification and were consistent with this review. The number of game elements used in these studies ranged from 1 to 11. Multimedia Appendix 2 illustrates the frequency of each game element across the included studies. Interestingly, no research was found to use the game element of "social sharing." Figure 2 presents a Sankey diagram illustrating the specific many-to-many relationships between each study and the elements it used. Overall, each included study integrated at least 3 game elements, with the study by Nacimiento-García et al [2] using the highest number of game elements (n=10). The primary reasons cited for incorporating game elements were to promote engagement and motivation (n=7), enhance adherence (n=3), and increase enjoyment (n=3). Other reasons for using game elements are detailed in Table S2 in Multimedia Appendix 1. Additionally, 7 studies described "the concrete approach to gamification," indicating that reporting on this aspect requires strengthening in future research.

Figure 2. Sankey diagram illustrating the distribution and flow of game elements across the included studies. The left column represents the individual studies, while the right column lists the identified game elements. Each colored flow connects a study to the specific game elements it used [2,39-48].



Needs and Experiences

Three included studies reported on user needs and experiences through qualitative methods. Positive user experiences, reported across the studies, included high acceptability, satisfaction, emotional satisfaction, enhanced enjoyment, motivation to continue use, a refreshing experience, perceived cognitive benefits, and improvements in physical activity, memory, and sleep quality, alongside practical design aspects such as clear instructions, good design, suitability for personal preferences, and accessibility for play at any time [2,48]. The interventions' social and collaborative aspects were also described as valuable for fostering a sense of community [2].

Reported challenges and negative experiences included technical issues such as system freezes, difficulty with game mechanics, relatively small character and object sizes, and misalignment between in-game and actual movements, as well as experiences of stress and initial physical fatigue [2,48]. Furthermore, user needs and preferences were found to differ by gender, with men

XSL•FO
RenderX

valuing cognitive challenges, while women emphasized practical memory support [41].

Health Promotion Domains and Recommendations

Overview

The 11 included studies reported several positive outcomes. The interventions were found to improve enjoyment and motivation (n=5), cognitive function (n=3), physical activity (n=2), and digital literacy (n=2). Individual studies also reported improvements in mental health (n=1) and adherence (n=1), a reduction in suicidal ideation (n=1), improvements in physical function (n=1), the promotion of social engagement (n=1), and the identification of mild cognitive impairment (n=1).

Improve Physical Function

One RCT by Lee et al [46] compared the efficacy of an 8-week mHealth-based gamification intervention against conventional exercise in 40 community-dwelling older adults (aged 60 - 85 years). The study found significant improvements in physical function in both groups, concluding that the gamified intervention produced comparable exercise effects to conventional exercise. Notably, the study examined mHealth-based gamification interventions as an integrated whole rather than evaluating individual game elements. Future research is needed to identify specific game elements most effective in enhancing older adults' physical function.

Improve Adherence

In a 3-week home-based trial with 21 healthy older adults, Nagler et al [47] investigated the impact of user control over game design elements (difficulty adjustment, reward selection, and visual theme customization) on adherence to a memory training game. The study compared a "user-control" mode to an "automatic" mode and found that the user-control group demonstrated significantly higher play frequency, total engagement duration, and memory task accuracy. These findings suggest that user-driven gamification enhances adherence and performance. However, the study combined 3 game elements, preventing identification of individual effects, and its small sample of healthy, motivated participants limited generalizability, while the nonrandomized design hindered causal inference. Additionally, the research did not specify the concrete approach to gamification, leaving unclear how exactly game mechanics were integrated into the mHealth intervention.

Promote Physical Activity

Greysen et al [40] conducted a 12-week RCT with 94 older adults at risk for Alzheimer's disease. They found that an mHealth-based gamification intervention with a support partner led to a significant increase in physical activity (1699 steps/day; $P<.001$) compared to controls (step goals with daily feedback), a gain that was partially sustained at a 3-week follow-up. However, the authors noted several limitations, including the short intervention and follow-up periods and the exclusion of nine participants post randomization, which could introduce bias.

The RCT by Roh et al [45] targeted older adults to evaluate a gamification-based smartphone app for promoting brain health. Over 8 weeks, 40 older adults were randomized to the app intervention or control group, with outcomes measured via physical activity questionnaires, cognitive scores, and the Situational Motivation Scale. The intervention group showed significant improvements in moderate-intensity physical activity ($P=.01$) and an 8-week physical activity difference ($P=.04$) versus controls. However, the study faced challenges, including a small sample size and a short intervention period, which may limit generalizability to long-term behavior change in older adults.

Identify Mild Cognitive Impairment

Murata et al [44] developed and tested a gamified mobile app, based on the N-back task, to screen for mild cognitive impairment in 138 older adults. The study found significant correlations between gameplay performance metrics (eg, swipe time, accuracy) and cognitive function (Mini-Mental State Examination scores). The authors concluded that their gamified app could identify mild cognitive impairment with an accuracy comparable to standard dementia screening tests. However, the study did not address participants' past smart device experience, which may confound motor-based metrics. Additionally, reliance on device-specific factors (eg, swipe time) raises cross-device reproducibility concerns.

Improve Cognitive Function and Promote Digital Literacy and Social Engagement

Nacimiento-García et al [2] developed a tailored mHealth-based gamification intervention, with games targeting memory, attention, and problem-solving. A pilot study found the intervention effectively promoted cognitive function, digital literacy, and social engagement. However, the authors noted that advanced tools like VR glasses posed initial usability barriers for some older users, and they recommended that future research should expand participant diversity and conduct longitudinal controlled studies to assess long-term impacts.

ChePa et al [39] used an mHealth-based gamification intervention for older individuals with memory disorders and found that it led to significant cognitive improvement, enhanced memory recall, and improved digital literacy. A key limitation noted was the lack of a specific description of how the gamification was implemented.

In a qualitative study by Lai et al [41], participants reported perceived cognitive benefits from using a Cantonese-language gamified mobile app. However, the reported cognitive benefits were based on self-reported experiences only; no objective cognitive or usage data were collected, and the small sample size limits the generalizability of findings across different demographic subgroups.

Improve Enjoyment and Motivation

Nacimiento-García et al [2] found that gamified experiences were enjoyable in mHealth-based interventions. Roh et al [45] demonstrated that such interventions significantly enhanced intrinsic motivation ($P=.01$), which subsequently promoted physical activity ($P=.01$). Kamnardsiri et al [48] reported comparable benefits in a home-based, low-intensity, gamification-based interactive physical-cognitive training pilot; participants showed a significant increase in enjoyment ($P=.001$)

XSL•FO
RenderX

over 4 weeks. However, this study focused on prototype usability and did not evaluate clinical physical or cognitive outcomes, nor did it specify the concrete approach to gamification.

Lai et al [41] reported that participants experienced emotional satisfaction and expressed motivation to continue using a gamified application; however, the study did not assess enjoyment or motivation using quantitative measures. Given that the study was exploratory and based on voluntary participation, potential self-selection bias should be taken into account when interpreting the results.

Chien et al [42] reported that participants in a home-based pulmonary telerehabilitation program using the Intelligent Pulmonary Rehabilitation Exercise System experienced increased engagement, motivation, and enjoyment through personalized feedback, collaborative goal-setting, and gamified exercise modules. However, the study used a single-group design with a modest sample size, evaluations were primarily conducted during hospitalization, and the absence of a control group or comparison with other treatment modalities limits generalizability.

Improve Mental Health and Reduce Suicidal Ideation

Zhang [43] conducted an RCT involving 120 older women with suicidal ideation to evaluate a gamified group-based VR physical activity intervention. The study concluded that this approach serves as a promising strategy to enhance psychological well-being and mitigate suicide risk. However, the intervention lacked a guiding theoretical framework. Critically, the study design did not allow for the identification of the specific effects of individual game elements. Furthermore, the reliance on indirect indicators rather than direct suicide assessment, combined with the absence of blinding and longitudinal follow-up, presents methodological limitations.

Discussion

Summary of Key Findings

This is the first scoping review to examine mHealth-based gamification interventions for health promotion among older adults. This scoping review identified 11 studies. We provide a comprehensive overview of mHealth-based gamification interventions for older people's health promotion in countries like China, Korea, and the United States. Only 1 article was published before 2022. These studies demonstrated the interventions' effectiveness in improving physical and cognitive function, adherence, enjoyment, motivation, and mental health; promoting physical activity, digital literacy, and social engagement; identifying mild cognitive impairment; and reducing suicidal ideation.

Game elements used were ranked by frequency as progress, challenges, goals, levels, rewards, sensation, storytelling or narration, leaderboard, surprise, and avatar. However, due to insufficient information in these studies, it remains unclear whether intervention outcomes should be attributed to single game elements or synergistic effects of multiple elements. mHealth types included augmented and VR-based training systems, wearable devices, mobile phones, tablets, and Windows platforms and devices. Notably, only 4 studies applied theoretical frameworks (behavioral economic principles and gamification theory, technology acceptance model, and human-centered design framework), and 3 omitted the concrete approach to gamification. Given these findings, there is substantial potential for wider adoption of mHealth-based gamification interventions for clinicians and patients in future health promotion among older adults.

Intervention Characteristics

First, all included studies in this research had relatively short intervention periods, with the longest lasting only 14 weeks [39], and 1 study even involved a single intervention lasting just 5 to 10 minutes [44]. This finding aligns with Aschentrup et al [31], where only 2 studies had a 12-week intervention period. In reality, the long-term effects of interventions are one of the most critical aspects for evaluating intervention efficacy. Additionally, most studies did not report specific details of intervention dosage, particularly intervention frequency and the duration of single sessions. Therefore, future research should explore long-term effects and report intervention dosage details to facilitate study reproducibility.

Second, the mHealth types involved in the studies included augmented and VR-based training systems, wearable devices, mobile phones, tablets, and Windows platforms and devices, with mobile phones and tablets being the most commonly used. This contradicts the findings of Xu et al [49], whose study indicated that most research used wearable devices. A possible reason for this discrepancy is that their study focused solely on physical activity in participants, and wearable devices can provide real-time feedback on daily steps and energy expenditure through specially designed algorithms [50,51]. Thus, when using mHealth-based gamification interventions to promote physical activity among older adults in the future, we recommend prioritizing wearable devices as the mHealth type.

Finally, in terms of theoretical applications, only 4 studies applied theoretical frameworks: 1 used behavioral economic principles to select gamification elements [40], 1 used gamification theory to guide intervention design [45], 1 used the technology acceptance model to explore older adults' experiences with the gamified mobile app, and 1 adopted a human-centered design framework to guide the development of an mHealth-based gamified intervention system. This finding differs from previous research. For example, Montagni et al [52] noted that almost all included studies were grounded in robust theories, which were proven effective in promoting understanding and evaluating information related to vaccines and behavioral change end points. Literature reviews also show that theory-driven interventions are more effective than non–theory-based ones [53]. Therefore, we advocate that future designs of mHealth-based gamification interventions for older people's health promotion should be closely integrated with theoretical frameworks.

Gamification Characteristics

First, in this study, the most frequently used game element was progress, followed by challenges and goals. The most commonly used game elements align with a previous study [49]. Notably,

XSL•FO
RenderX

goals serve as a key technique for behavioral change [54]. When combined with progress and feedback, they can significantly boost intrinsic motivation [55]. The only game element unused in all included studies was social sharing, which contradicts prior findings. For example, Xu et al [49] showed that when using mHealth-based gamification interventions for physical activity, the second most common game element was social interaction, which not only enhanced users' enjoyment but also promoted motivation for physical activity through social incentives. The discrepancy might be attributed to differing inclusion criteria for gamification elements across studies [56,57]. Additionally, Mahmoudi et al [38] emphasized the importance of considering user preferences and personalized game elements, which was not reflected in our study. Thus, developers and researchers are encouraged to integrate social interaction, user preferences, and personalized gamification elements into interventions to enhance user engagement and, in turn, their effectiveness in health-related outcomes.

Second, gamification is one of the most critical components of mHealth-based gamification interventions. However, it is worth noting that among the 11 included studies, 3 failed to describe how gamification was specifically implemented in mHealth. This could undermine the reproducibility and comparability of research findings, thereby hindering the accurate evaluation of intervention efficacy and the formulation of evidence-based guidelines. Such oversight also aligns with previous review studies [31,38,58]. Thus, future research is recommended to establish corresponding reporting guidelines, emphasizing the need to strengthen this aspect of reporting for field advancement.

Finally, it is noteworthy that each study explained the reasons for using game elements, primarily, including promoting engagement and motivation, enhancing adherence, increasing enjoyment, and improving intervention efficacy. These findings are consistent with previous studies. For example, Mahmoudi et al [38] noted that most studies included in their research aimed to use game elements to enhance the adherence and motivation of children with disabilities. Additionally, a study has found that introducing gamified applications can improve intervention effectiveness by encouraging users to use them continuously and frequently [59].

Needs and Experiences

Only 3 studies collected older users' needs and experiences. The positive feedback can be categorized into 4 main aspects: first, the attributes of interventions, covering practical design aspects such as user-friendly design, personalization, and accessibility for play at any time; second, user engagement and subjective experience, which included high acceptability, satisfaction, emotional satisfaction, enhanced enjoyment, motivation to continue use, and a refreshing experience; third, health and functional benefits, such as perceived cognitive benefits and improvements in physical activity, memory, and sleep quality; and fourth, social and cultural elements, where the interventions' social and collaborative aspects were described as valuable for fostering a sense of community. Possible explanations include that game elements enhance enjoyment and self-efficacy among participants. Additionally, integrating immersive and gamified elements provides dual benefits of cognitive and emotional stimulation, enhancing the overall user experience [2]. These results are also supported by previous studies [60,61].

Key issues included usability and technical flaws, including system freezes, difficult game mechanics, small character and object sizes, and movement misalignment, alongside reports of stress and initial physical fatigue. Usability is a critical aspect of improving acceptability and intervention adherence [48]. A prior study has shown that user-friendly interfaces, in particular, are a key factor in promoting positive acceptance among older users unfamiliar with new technologies [62]. Thus, future research should prioritize investigating needs and experiences, as this is crucial for designing mHealth-based gamification interventions. And more attention should be paid to user-centered design principles to ensure high acceptability and usability.

Health Promotion Domains and Recommendations

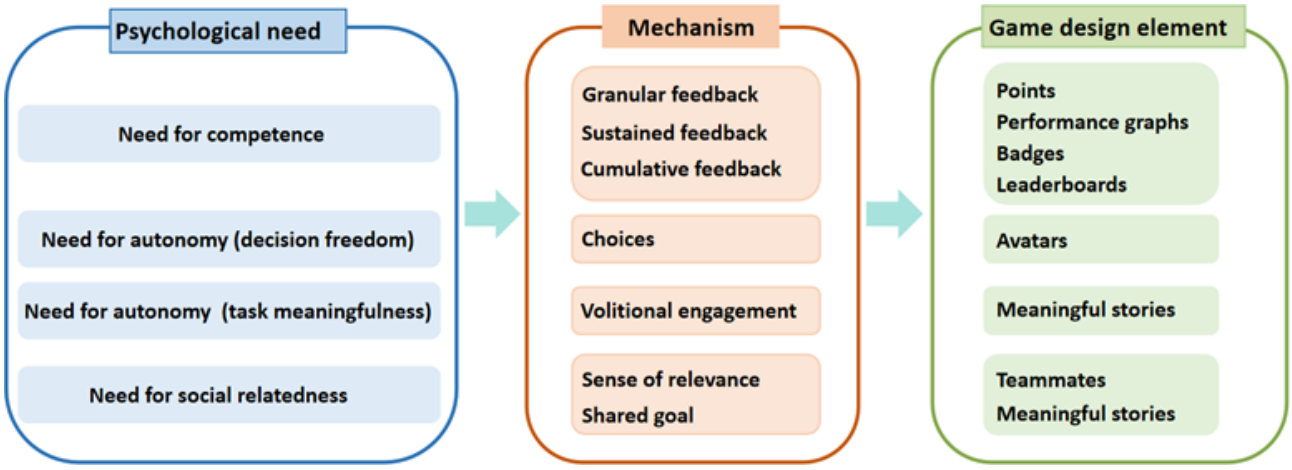
Our study demonstrates that mHealth-based gamification interventions effectively improve physical and cognitive function, enhance adherence, promote physical activity, identify mild cognitive impairment, promote digital literacy and social engagement, improve mental health, reduce suicidal ideation, and enhance enjoyment and motivation among older adults.

These findings are consistent with previous studies. A study by Luo et al [63] showed that mHealth-based gamification interventions effectively improve pre-exposure prophylaxis adherence among men who have sex with men. Xu et al [49] found that mHealth-based gamification interventions may increase participation in physical activity. Huang et al [64] noted that using gamification to promote self-management among patients with chronic diseases has been widely recognized, including promoting physical exercise and rehabilitation training, enhancing motivation for symptom management, providing psychological support, improving cognitive function, and improving medication adherence. Additionally, research has indicated that when gamification is combined with emerging technologies, the findings may be more promising [65]. This could be attributed to digital tools being unconstrained by time and space [66], and gamification elements potentially enhancing their appeal and acceptability. Thus, future intervention designs should consider the integration of new technologies and gamification.

The motivational effects of game elements can also be explained theoretically. Sailer et al [67] constructed a relationship between various gamification elements (such as points, leaderboards, and badges) and Ryan and Deci's self-determination theory [68] (Figure 3). It is a theory of motivation that identifies 3 universal basic psychological needs that determine human behavior: competence, autonomy, and social inclusion [68]. If 1 or more of these needs are met (eg, through gamification elements), there is a positive impact on human behavior and the factors that influence it [67].

XSL•FO
RenderX

Figure 3. Psychological needs with matching game design elements.



Beyond the topics identified in our review, prior research has demonstrated the potential of mHealth-based gamification interventions across other health domains. For example, Montagni et al [52] showed that digital gamification enhances vaccination knowledge and ultimately expands vaccination coverage. Another study highlighted its potential in promoting positive oral health behaviors [69]. To guide this future work, key methodological findings should also be considered. Xu et al [49] found that combining multiple game elements in mHealth-based gamification interventions is more effective than single elements, and theory-guided interventions outperform those without theoretical grounding. Therefore, we recommend that future empirical research should not only explore broader health domains but also (1) systematically compare the effects of single versus combined game elements and (2) ensure that the design and implementation of interventions are grounded in established theoretical frameworks.

Practical and Research Implications

Our study suggests that mHealth-based gamification interventions are a promising approach for enhancing older adults' health. Although this review identifies potential benefits, given the limited number of included studies, these findings should be interpreted with a cautious and critical perspective. We propose several recommendations for future research and practice:

First, to expand the innovative capacity outlined above, determining long-term impacts is crucial. Future studies should standardize research settings, increase sample sizes, extend intervention and follow-up periods, and conduct high-quality RCTs. They should also report intervention dosage details to facilitate study reproducibility.

Additionally, most existing studies lack theoretical frameworks. Thus, we advocate for integrating theoretical frameworks into future designs of mHealth-based gamification interventions for older health promotion, as theory-guided interventions may be more effective than nontheoretical ones.

Furthermore, future research is needed to explore whether intervention outcomes can be attributed to single game elements or synergistic effects of multiple elements.

Moreover, future research should establish corresponding reporting guidelines, particularly focusing on how gamification is specifically implemented in mHealth, thereby advancing the field and enhancing the reproducibility of studies.

Additionally, developers and researchers are encouraged to integrate social interaction, user preferences, and personalized gamification elements to create tailored, older-friendly interventions, thereby enhancing user engagement and health-related effectiveness.

Subsequently, future studies should prioritize investigating users' needs and experiences, which are critical factors in designing mHealth-based gamification interventions.

Finally, empirical research for older adults should also expand mHealth-based interventions to broader health domains, such as exploring their impacts on the mental health of older populations.

Limitations

We acknowledge several limitations in this review. First and foremost is the limited number of included studies, many of which were pilot studies or used small sample sizes. Consequently, the current evidence primarily supports the feasibility and preliminary potential of mHealth-based gamification rather than definitive broad effectiveness. Although we searched 8 databases to ensure comprehensiveness, the application of gamification in geriatric health remains a relatively emerging field.

Second, a major limitation of this scoping review is the overreliance on published peer-reviewed research, which may introduce publication bias and limit the comprehensiveness of the mapping. Relying solely on published works explicitly excludes gray literature such as technical reports, institutional documents, theses, dissertations, and government publications. These materials often provide valuable insights into early-stage developments and practical implementations. Although these contents may not have entered the formal academic literature, their contribution to this field is crucial. The absence of these perspectives may lead to an incomplete understanding of the overall landscape and real-world applications. While the focus of this review was deliberately on published and accessible peer-reviewed empirical evidence to ensure a high baseline of

XSL•FO
RenderX

methodological rigor, future reviews should incorporate gray literature to capture a broader range of evidence, particularly for this emerging field.

Third, the protocol for this scoping review was not prospectively registered, which may diminish transparency and increase the potential for unintentional bias in our review process. To mitigate this, all authors critically reflected on the different steps. Fourth, restricting the search to English-language literature may have excluded relevant studies from non–English-speaking countries, potentially affecting the comprehensiveness and generalizability of our findings. Fifth, this review focused on describing effects and did not conduct a meta-analysis. Sixth, all included studies were conducted in high- or middle-income countries, limiting the generalizability of findings to low-income contexts. Furthermore, restrictions due to the specific definition of gamification adopted (ie, excluding serious games) may have limited the scope of relevant interventions included in this review. Finally, given the rapidly evolving digital landscape, the temporal limitation of the search period necessitates cautious interpretation as new research continues to emerge.

Conclusion

As the first review to systematically scope this specific area, our study identifies mHealth-based gamification as a novel and viable strategy for fostering healthy aging. Although the findings are preliminary and require cautious interpretation due to the limited number of studies, the synthesized evidence highlights the capacity of these interventions to enhance enjoyment, motivation, cognitive function, physical activity, digital literacy, and other aspects among older adults. By bridging the gap between geriatric health needs and emerging digital technologies, this review provides a much-needed baseline for the field. However, to transform these emerging interventions into scalable real-world solutions, future research must prioritize rigorous, theory-driven, and long-term evaluations, alongside standardized reporting and personalized, socially interactive designs. Ultimately, this work lays the groundwork for developing more effective, age-tailored digital health tools that can tangibly improve the quality of life for the aging population.

Acknowledgments

The authors declare the use of generative artificial intelligence (GAI) in the research and writing process. According to the GAIDeT (Generative AI Delegation Taxonomy; 2025), the following tasks were delegated to GAI tools under full human supervision: proofreading and editing. The GAI tool used was Gemini 3 Pro Preview. Responsibility for the final manuscript lies entirely with the authors. GAI tools are not listed as authors and do not bear responsibility for the final outcomes. Artificial intelligence (Gemini 3 Pro Preview) was used for minor language refinement and editing, specifically to improve grammar, spelling, verb tense, sentence flow, and English idiomatic expression. All outputs were reviewed and modified by the authors, and the authors take full responsibility for the final content. This declaration was submitted by the authors.

Funding

This work was supported by the Natural Science Foundation Project of Hunan Province (grant 2025JJ90277 and 2025JJ60537), Key Research and Development Program of Hunan Province (grant 2025JK2118), and Health Research Project of Hunan Provincial Health Commission (grant 20255135).

Authors' Contributions

Conceptualization: DJ, HN, QX, XC, YC, YZ
Data curation: JH, XC, YZ
Formal analysis: JH, XC, YZ
Funding acquisition: HF, HN
Investigation: XC, YZ
Methodology: DJ, JH, QX, XC, YZ
Project administration: XC, YZ
Resources: HF, HN
Supervision: JH, QX, XC, YZ
Validation: DJ, QX, XC, YZ
Visualization: DJ, JC, WH, XC, YC, YZ
Writing – original draft: DJ, JH, XC, YZ
Writing – review & editing: HN, HF, JC, WH, XC, YC

HN and HF are the corresponding authors of this study.
HN can be reached via email at (Hongting4960@hnucm.edu.cn).

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Complete database search strategies, summary of intervention characteristics, and overview of gamification characteristics of included studies.
[DOCX File, 44 KB - mhealth_v14i1e82368_app1.docx]

Multimedia Appendix 2
Frequency distribution of game elements across the 11 included studies.
[PNG File, 52 KB - mhealth_v14i1e82368_app2.png]

Checklist 1
PRISMA-ScR checklist.
[PDF File, 115 KB - mhealth_v14i1e82368_app3.pdf]

Checklist 2
PRISMA-S checklist.
[PDF File, 86 KB - mhealth_v14i1e82368_app4.pdf]

References

1. Padeiro M, Santana P, Grant M. Global aging and health determinants in a changing world. In: Aging: From Fundamental Biology to Societal Impact: Academic Press; 2023:3-30. [doi: 10.1016/B978-0-12-823761-8.00021-5]
2. Nacimiento-García E, González-González CS, Colombo-Ruano L, Gutiérrez-Vela FL. Gamification and immersive experiences: a gamified approach for promoting active aging. Appl Sci 2024;14(23):10880. [doi: 10.3390/app142310880]
3. Bending the trends to promote health and well-being: a strategic foresight on the future of health promotion. : World Health Organization; 2022 URL: https://iris.who.int/server/api/core/bitstreams/9e4c885e-409a-4be4-bc2b-979bb5f91607/content [accessed 2026-04-07]
4. Dogra S, Dunstan DW, Sugiyama T, Stathi A, Gardiner PA, Owen N. Active aging and public health: evidence, implications, and opportunities. Annu Rev Public Health 2022 Apr 5;43(1):439-459. [doi: 10.1146/annurev-publhealth-052620-091107] [Medline: 34910580]
5. Patel MS, Cacchione J, Yehia BR. 4 ways to remove “sludge” from health care processes. Harvard Business Review. 2022 Mar 11. URL: https://hbr.org/2022/03/4-ways-to-remove-sludge-from-health-care-processes [accessed 2024-02-29]
6. Price M, Yuen EK, Goetter EM, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014 Sep;21(5):427-436. [doi: 10.1002/cpp.1855] [Medline: 23918764]
7. Olff M. Mobile mental health: a challenging research agenda. Eur J Psychotraumatol 2015;6(1):27882. [doi: 10.3402/ejpt.v6.27882] [Medline: 25994025]
8. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Jul;71:31-48. [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]
9. King D, Greaves F, Exeter C, Darzi A. “Gamification”: influencing health behaviours with games. J R Soc Med 2013 Mar;106(3):76-78. [doi: 10.1177/0141076813480996] [Medline: 23481424]
10. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining “gamification”. Presented at: Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments; Sep 28-30, 2011. [doi: 10.1145/2181037.2181040]
11. Marston HR, Hall AK. Gamification: applications for health promotion and health information technology engagement. In: Handbook of research on holistic perspectives in gamification for clinical practice: IGI Global; 2016:78-104. [doi: 10.4018/978-1-4666-9522-1.ch005]
12. Huotari K, Hamari J. Defining gamification: a service marketing perspective. Presented at: 16th International Academic MindTrek Conference; Oct 3-5, 2012 URL: https://research.aalto.fi/en/publications/defining-gamification-a-service-marketing-perspective/ [accessed 2026-04-07] [doi: 10.1145/2393132.2393137]
13. Blohm I, Leimeister JM. Gamification. Wirtschaftsinf 2013 Aug;55(4):275-278. [doi: 10.1007/s11576-013-0368-0]
14. Sawa R, Saitoh M, Morisawa T, et al. The potential application of commercially available active video games to cardiac rehabilitation: scoping review. JMIR Serious Games 2022 Mar 18;10(1):e31974. [doi: 10.2196/31974] [Medline: 35302503]
15. Cechetti NP, Bellei EA, Biduski D, Rodriguez JPM, Roman MK, De Marchi ACB. Developing and implementing a gamification method to improve user engagement: a case study with an m-health application for hypertension monitoring. Telematics Inf 2019 Aug;41:126-138. [doi: 10.1016/j.tele.2019.04.007]
16. Davis AJ, Parker HM, Gallagher R. Gamified applications for secondary prevention in patients with high cardiovascular disease risk: a systematic review of effectiveness and acceptability. J Clin Nurs 2021 Oct;30(19-20):3001-3010. [doi: 10.1111/jocn.15808] [Medline: 33872436]
17. Höchsmann C, Infanger D, Klenk C, Königstein K, Walz SP, Schmidt-Trucksäss A. Effectiveness of a behavior change technique-based smartphone game to improve intrinsic motivation and physical activity adherence in patients with type 2

XSL•FO
RenderX

diabetes: randomized controlled trial. JMIR Serious Games 2019 Feb 13;7(1):e11444. [doi: 10.2196/11444] [Medline: 30758293]
18. Theng YL, Lee JWY, Patinadan PV, Foo SSB. The use of videogames, gamification, and virtual environments in the self-management of diabetes: a systematic review of evidence. Games Health J 2015 Oct;4(5):352-361. [doi: 10.1089/g4h.2014.0114] [Medline: 26287926]
19. Chen Y, Abel KT, Janecek JT, Chen Y, Zheng K, Cramer SC. Home-based technologies for stroke rehabilitation: a systematic review. Int J Med Inform 2019 Mar;123:11-22. [doi: 10.1016/j.ijmedinf.2018.12.001] [Medline: 30654899]
20. Hashim SHM, Ismail M, Manaf H, Hanapiah FA. Development of dual cognitive task virtual reality game addressing stroke rehabilitation. Presented at: ICVARS '19: Proceedings of the 2019 3rd International Conference on Virtual and Augmented Reality Simulations; Feb 23-25, 2019. [doi: 10.1145/3332305.3332312]
21. Ozgur AG, Khodr H, Bruno B, et al. Detecting compensatory motions and providing informative feedback during a tangible robot assisted game for post-stroke rehabilitation. Presented at: 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN); Aug 8-21, 2021. [doi: 10.1109/RO-MAN50785.2021.9515325]
22. Allam A, Kostova Z, Nakamoto K, Schulz PJ. The effect of social support features and gamification on a web-based intervention for rheumatoid arthritis patients: randomized controlled trial. J Med Internet Res 2015 Jan 9;17(1):e14. [doi: 10.2196/jmir.3510] [Medline: 25574939]
23. AlMarshedi A, Wills GB, Ranchhod A. The Wheel of Sukr: a framework for gamifying diabetes self-management in Saudi Arabia. Procedia Comput Sci 2015;63:475-480. [doi: 10.1016/j.procs.2015.08.370]
24. Dennis TA, O'Toole L. Mental health on the go: effects of a gamified attention bias modification mobile application in trait anxious adults. Clin Psychol Sci 2014 Sep 1;2(5):576-590. [doi: 10.1177/2167702614522228] [Medline: 26029490]
25. Hamari J, Koivisto J, Sarsa H. Does gamification work? -- a literature review of empirical studies on gamification. Presented at: 2014 47th Hawaii International Conference on System Sciences (HICSS); Jan 6-9, 2014. [doi: 10.1109/HICSS.2014.377]
26. Johnson D, Deterding S, Kuhn KA, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016 Nov;6:89-106. [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
27. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010;31(1):399-418. [doi: 10.1146/annurev.publhealth.012809.103604] [Medline: 20070207]
28. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
29. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011;6(1):42. [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
30. Reeve J. Understanding Motivation and Emotion, 6th edition: John Wiley & Sons; 2014.
31. Aschentrup L, Steimer PA, Dadaczynski K, Mc Call T, Fischer F, Wrona KJ. Effectiveness of gamified digital interventions in mental health prevention and health promotion among adults: a scoping review. BMC Public Health 2024 Jan 2;24(1):69. [doi: 10.1186/s12889-023-17517-3] [Medline: 38167010]
32. Hill R, Betts LR, Gardner SE. Older adults' experiences and perceptions of digital technology: (dis)empowerment, wellbeing, and inclusion. Comput Hum Behav 2015 Jul;48:415-423. [doi: 10.1016/j.chb.2015.01.062]
33. Warsinsky S, Schmidt-Kraepelin M, Rank S, Thiebes S, Sunyaev A. Conceptual ambiguity surrounding gamification and serious games in health care: literature review and development of Game-Based Intervention Reporting Guidelines (GAMING). J Med Internet Res 2021 Sep 10;23(9):e30390. [doi: 10.2196/30390] [Medline: 34505840]
34. Al-Rayes S, Al Yaqoub FA, Alfayez A, et al. Gaming elements, applications, and challenges of gamification in healthcare. Inf Med Unlocked 2022;31:100974. [doi: 10.1016/j.imu.2022.100974]
35. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
36. Peters M, Godfrey CM, McInerney, Khalil H, Parker D, Baldini Soares C. Chapter 11: scoping reviews. In: JBI Manual for Evidence Synthesis: Joanna Briggs Institute; 2020:467-473. [doi: 10.46658/JBIRM-20-01]
37. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev 2021 Jan 26;10(1):39. [doi: 10.1186/s13643-020-01542-z] [Medline: 33499930]
38. Mahmoudi E, Yejong Yoo P, Chandra A, et al. Gamification in mobile apps for children with disabilities: scoping review. JMIR Serious Games 2024 Sep 6;12(1):e49029. [doi: 10.2196/49029] [Medline: 39240675]
39. ChePa N, Sie-Yi LL, Yusof N, Wan Yahaya WAJ, Ishak R. Impacts of game-based psychotherapy intervention to elderly with memory disorder. Entertain Comput 2023 Jan;44:100532. [doi: 10.1016/j.entcom.2022.100532]
40. Greysen SR, Oon AL, Harkins K, et al. Effect of gamification with a support partner to increase physical activity in older adults at risk for Alzheimer's disease: the STEP 4Life randomized clinical trial. Alzheimers Dement 2024 Aug;20(8):5450-5459. [doi: 10.1002/alz.14058] [Medline: 38962958]
41. Lai FHY, Yip BCB, Hai EYK, Darling C, Man DWK. Acceptability and feasibility in using gamified mobile application for working memory training in older adults. Health Soc Care Community 2025 Jan;2025(1):9925547. [doi: 10.1155/hsc/9925547]

XSL•FO
RenderX

42. Chien SY, Hu HC, Tseng W. Design of an AI-driven home-based pulmonary telerehabilitation system to enhance patient engagement. Digit Health 2025;11:20552076251393295. [doi: 10.1177/20552076251393295] [Medline: 41229937]
43. Zhang J. Preventing suicide among older women in China: the integration of gamified group-based exercise in a virtual reality environment. Front Psychol 2025;16:1626476. [doi: 10.3389/fpsyg.2025.1626476] [Medline: 40989272]
44. Murata N, Nishii S, Usuha R, et al. A gamified N-back app for identifying mild-cognitive impairment in older adults. JMA J 2025 Jan 15;8(1):174-182. [doi: 10.31662/jmaj.2024-0217] [Medline: 39926062]
45. Roh HW, Ryu H, Jeong S, et al. The effectiveness of a motivational enhancement smartphone application promoting lifestyle improvement for brain health: a randomized controlled trial. PLoS One 2022;17(6):e0267806. [doi: 10.1371/journal.pone.0267806] [Medline: 35771740]
46. Lee EL, Ko MH, Shin MJ, et al. The effect of convergence gamification training in community-dwelling older people: a multicenter, randomized controlled trial. J Am Med Dir Assoc 2022 Mar;23(3):373-378. [doi: 10.1016/j.jamda.2021.05.041] [Medline: 34216552]
47. Nagle A, Riener R, Wolf P. High user control in game design elements increases compliance and in-game performance in a memory training game. Front Psychol 2015;6:1774. [doi: 10.3389/fpsyg.2015.01774] [Medline: 26635681]
48. Kamnardsiri T, Kumfu S, Munkhetvit P, Boripuntakul S, Sungkarat S. Home-based, low-intensity, gamification-based, interactive physical-cognitive training for older adults using the ADDIE model: design, development, and evaluation of user experience. JMIR Serious Games 2024 Oct 29;12:e59141. [doi: 10.2196/59141] [Medline: 39470391]
49. Xu L, Shi H, Shen M, et al. The effects of mHealth-based gamification interventions on participation in physical activity: systematic review. JMIR mHealth uHealth 2022 Feb 3;10(2):e27794. [doi: 10.2196/27794] [Medline: 35113034]
50. Preusse KC, Mitzner TL, Fausset CB, Rogers WA. Older adults' acceptance of activity trackers. J Appl Gerontol 2017 Feb;36(2):127-155. [doi: 10.1177/0733464815624151] [Medline: 26753803]
51. Brickwood KJ, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR mHealth uHealth 2019 Apr 12;7(4):e11819. [doi: 10.2196/11819] [Medline: 30977740]
52. Montagni I, Mabchour I, Tzourio C. Digital gamification to enhance vaccine knowledge and uptake: scoping review. JMIR Serious Games 2020 May 18;8(2):e16983. [doi: 10.2196/16983] [Medline: 32348271]
53. Moore G, Cambon L, Michie S, et al. Population health intervention research: the place of theories. Trials 2019 Jun 11;20(1):285. [doi: 10.1186/s13063-019-3383-7] [Medline: 31186053]
54. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013 Dec 12;1(1):e3. [doi: 10.2196/games.3139] [Medline: 25658754]
55. Hicks K, Gerling K, Richardson G, Pike T, Burman O, Dickinson P. Understanding the effects of gamification and juiciness on players. Presented at: 2019 IEEE Conference on Games (CoG); Aug 20-23, 2019. [doi: 10.1109/CIG.2019.8848105]
56. Cheng VWS, Davenport T, Johnson D, Vella K, Hickie IB. Gamification in apps and technologies for improving mental health and well-being: systematic review. JMIR Ment Health 2019 Jun 26;6(6):e13717. [doi: 10.2196/13717] [Medline: 31244479]
57. Camargo MC, Barros RM, Brancher JD, Barros VTO, Santana M. Designing gamified interventions for autism spectrum disorder: a systematic review. In: van der Spek E, Göbel S, Do EYL, Clua E, Hauge JB, editors. Entertainment Computing and Serious Games: First IFIP TC 14 Joint International Conference, ICEC-JCSG 2019, Arequipa, Peru, November 11–15, 2019, Proceedings: Springer; 2019:341-352. [doi: 10.1007/978-3-030-34644-7_28]
58. Tran S, Smith L, El-Den S, Carter S. The use of gamification and incentives in mobile health apps to improve medication adherence: scoping review. JMIR mHealth uHealth 2022 Feb 21;10(2):e30671. [doi: 10.2196/30671] [Medline: 35188475]
59. Schoeppe S, Alley S, Van Lippevelde W, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 7;13(1):127. [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
60. Bonnechère B, Langley C, Sahakian BJ. The use of commercial computerised cognitive games in older adults: a meta-analysis. Sci Rep 2020 Sep 17;10(1):15276. [doi: 10.1038/s41598-020-72281-3] [Medline: 32943742]
61. Alnajjar F, Khalid S, Vogan AA, Shimoda S, Nouchi R, Kawashima R. Emerging cognitive intervention technologies to meet the needs of an aging population: a systematic review. Front Aging Neurosci 2019;11:291. [doi: 10.3389/fnagi.2019.00291] [Medline: 31798439]
62. Lu MH, Lin W, Yueh HP. Development and evaluation of a cognitive training game for older people: a design-based approach. Front Psychol 2017;8:1837. [doi: 10.3389/fpsyg.2017.01837] [Medline: 29089914]
63. Luo Q, Zhang Y, Wang W, Cui T, Li T. mHealth-based gamification interventions among men who have sex with men in the HIV prevention and care continuum: systematic review and meta-analysis. JMIR mHealth uHealth 2024 Apr 15;12(1):e49509. [doi: 10.2196/49509] [Medline: 38623733]
64. Huang X, Xiang X, Liu Y, Wang Z, Jiang Z, Huang L. The use of gamification in the self-management of patients with chronic diseases: scoping review. JMIR Serious Games 2023 Dec 22;11:e39019. [doi: 10.2196/39019] [Medline: 38133907]
65. Tozzi AE, Gesualdo F, D'Ambrosio A, Pandolfi E, Agricola E, Lopalco P. Can digital tools be used for improving immunization programs? Front Public Health 2016;4:36. [doi: 10.3389/fpubh.2016.00036] [Medline: 27014673]

XSL•FO
RenderX

66. Atkinson KM, Wilson K, Murphy MSQ, et al. Effectiveness of digital technologies at improving vaccine uptake and series completion - a systematic review and meta-analysis of randomized controlled trials. Vaccine 2019 May 21;37(23):3050-3060. [doi: 10.1016/j.vaccine.2019.03.063] [Medline: 31036457]
67. Sailer M, Hense JU, Mayr SK, Mandl H. How gamification motivates: an experimental study of the effects of specific game design elements on psychological need satisfaction. Comput Hum Behav 2017 Apr;69:371-380. [doi: 10.1016/j.chb.2016.12.033]
68. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
69. Moreira R, Silveira A, Sequeira T, et al. Gamification and oral health in children and adolescents: scoping review. Interact J Med Res 2024 Apr 4;13(1):e35132. [doi: 10.2196/35132] [Medline: 38573750]

Abbreviations

MeSH: Medical Subject Headings
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Literature Search Extension
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
RCT: randomized controlled trial
VR: virtual reality

Edited by L Buis; submitted 13.Aug.2025; peer-reviewed by H Söbke, LF Belo; revised version received 23.Mar.2026; accepted 23.Mar.2026; published 04.May.2026.

Please cite as:
Chen X, Zhu Y, Jiang D, Huang J, Xie Q, Chen Y, Chang J, Huang W, Ning H, Feng H
mHealth-Based Gamification Interventions to Promote Health Among Older Adults: Scoping Review
JMIR Mhealth Uhealth 2026;14:e82368
URL: https://mhealth.jmir.org/2026/1/e82368
doi:10.2196/82368

© Xi Chen, Yishu Zhu, Dian Jiang, Jundan Huang, Qi Xie, Yishi Chen, Jing Chang, Weiping Huang, Hongting Ning, Hui Feng. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 4.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effectiveness of Text Messaging Interventions on BMI Among Adults With Prediabetes: Systematic Review and Meta-Analysis

Arkers Kwan Ching Wong[1,2], PhD; Shun Yan Choi[1], BSc; Tiffany Tsz Yan Cheng[1], BSc; Chelsea Cheuk Yiu Cheng[1], BSc; Ivy Yuk Man Kwong[1], BSc; Cheuk Yin Fong[1], BSc; Chi Sum Wong[1], BSc; Ka Lo Chan[1], BSc

[1]School of Nursing, Hong Kong Polytechnic University, GH 502, 1 Cheong Wan Road, Hung Hom, China (Hong Kong)

[2]Joint Research Centre for Primary Health Care, Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

Corresponding Author:
Arkers Kwan Ching Wong, PhD
School of Nursing, Hong Kong Polytechnic University, GH 502, 1 Cheong Wan Road, Hung Hom, China (Hong Kong)

Abstract

Background: Prediabetes is a growing global health concern. Lifestyle modification is the cornerstone of management, yet scalable delivery strategies are needed. SMS text messaging is a promising mobile health approach for behavior change, but its effectiveness for metabolic outcomes in prediabetes remains uncertain.

Objective: This study aims to evaluate the effectiveness of text message–delivered lifestyle programs for BMI among adults with prediabetes and, secondarily, for weight, waist circumference, hemoglobin A_{1c} (HbA_{1c}), total cholesterol, and diabetes incidence.

Methods: We conducted a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)-guided systematic review and meta-analysis of randomized controlled trials (searching the PubMed, MEDLINE, CINAHL, Embase, Web of Science, and the Cochrane Library databases; from April 2005 to March 2025). Three reviewers independently screened studies, extracted data, and assessed risk of bias (using the Cochrane Risk of Bias tool), and random-effects meta-analyses were performed.

Results: In total, 7 randomized controlled trials (n=4632) met the inclusion criteria. SMS text messaging did not significantly affect BMI compared with standard care (mean difference [MD] −0.17 kg/m², 95% CI −0.85 to 0.25; I^2=32%). Secondary outcomes were similarly nonsignificant: weight (MD −0.46 kg, 95% CI −1.74 to 0.83; I^2=27%); waist circumference (MD −0.36 cm, 95% CI −1.09 to 0.36; I^2=13%); HbA_{1c} (MD −0.05%, 95% CI −0.17 to 0.07; I^2=89%); total cholesterol (standardized MD −0.00, 95% CI −0.06 to 0.06; I^2=0%); and diabetes incidence (odds ratio 0.84, 95% CI 0.63-1.12, I^2=51%; n=3515 across 3 trials). Estimates were small and, for several outcomes, notably HbA_{1c} and diabetes incidence, were imprecise and heterogeneous, indicating substantial between-study variability.

Conclusions: In adults with prediabetes, lifestyle programs incorporating SMS text messaging showed no meaningful effect on BMI or other metabolic outcomes; pooled estimates were near null with considerable uncertainty and heterogeneity for some end points. Variability in intervention design, dose, and tailoring likely contributed to these results. Future trials should extend follow-up, report body composition outcomes in addition to BMI, and test tailored, interactive, closed-loop SMS text messaging strategies with adequate power to resolve heterogeneous effects.

Trial Registration: PROSPERO CRD420251055972; https://www.crd.york.ac.uk/PROSPERO/view/CRD420251055972

(JMIR Mhealth Uhealth 2026;14:e78521) doi:10.2196/78521

KEYWORDS

prediabetes; text messages; meta-analysis; telehealth; mobile health; Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PRISMA

Introduction

Prediabetes affects approximately 374 million adults aged 20 to 79 years worldwide and is projected to increase to 730 million by 2045 [1,2]. It is defined by impaired fasting glucose or impaired glucose tolerance, with blood glucose levels elevated but below the threshold for a type 2 diabetes diagnosis [3,4]. Prediabetes is associated with increased risks of cardiovascular disease, microvascular complications, and all-cause mortality, and approximately 5% to 10% of individuals with prediabetes progress to type 2 diabetes annually [1]. Due to its asymptomatic nature, prediabetes often remains undiagnosed until significant disease onset, underscoring the need for early preventive strategies.

Lifestyle modification is the first-line approach for prediabetes management, demonstrating efficacy in delaying or preventing

XSL•FO
RenderX

type 2 diabetes [5]. Compared with pharmacological interventions, lifestyle programs are preferred by individuals and focus on dietary changes, increased physical activity, and behavioral support [6]. Adopting healthy eating patterns, such as the Mediterranean diet, improves glycemic control [7], while enhanced physical activity improves insulin sensitivity and reduces diabetes risk [8]. Behavioral strategies, including goal setting and self-monitoring, are critical for sustaining lifestyle changes [9]. Given that an elevated BMI is a strong risk factor for diabetes progression, targeting BMI through lifestyle interventions is central to prediabetes prevention efforts [10].

Mobile health (mHealth) technologies have emerged as scalable tools for delivering lifestyle interventions. Numerous studies have demonstrated that mHealth interventions, including mobile apps, wearable devices, and SMS text messaging, can effectively support behavior change and reduce the incidence of type 2 diabetes [11-14]. Among these mHealth interventions, SMS text messaging stands out for its simplicity. SMS text messaging, in particular, is widely accessible, cost-effective, and convenient, leveraging ubiquitous mobile phone use to promote health behaviors [15-17]. Messages can be retained, accessed flexibly, and tailored to individual needs, which may enhance engagement and adherence [11-14].

SMS text messaging interventions, guided by behavior change theories such as self-determination theory, offer motivational support and personalized feedback that facilitate sustainable behavior change [18-20]. Compared with other mHealth modalities, SMS text messaging imposes minimal user burden and provides an efficient, scalable approach to lifestyle modification [21]. These advantages are particularly relevant for individuals with prediabetes, who often face barriers to program participation, including socioeconomic constraints, time limitations, and low motivation [16,22].

Despite growing interest in SMS text messaging interventions for lifestyle modification, evidence specific to adults with prediabetes remains limited. Prior reviews have predominantly focused on populations with diagnosed diabetes or obesity, which may differ in baseline awareness, motivation, and response to interventions. As prediabetes continues to rise globally, understanding the effectiveness of SMS text messaging–based interventions in this specific population is critical.

This systematic review and meta-analysis aimed to evaluate the effectiveness of lifestyle modification programs using SMS text messaging on BMI reduction among adults with prediabetes, compared with standard care. Secondary outcomes included changes in weight, waist circumference, hemoglobin A_{1c} (HbA_{1c}), total cholesterol, and incidence of diabetes. By synthesizing available evidence, this study addresses the lack of targeted evidence for digital interventions in prediabetes management, a distinct group often underrepresented in diabetes prevention research. The findings are expected to directly inform the development of mHealth strategies tailored to the population with prediabetes, optimizing early intervention efforts. Furthermore, by identifying the role of SMS text messaging in BMI reduction and metabolic improvements, this review provides a foundation for implementing scalable, low-cost digital solutions to curb diabetes progression at both clinical and public health levels.

Methods

Overview

This systematic review and meta-analysis was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines (Checklist 1). It was registered in PROSPERO with the number CRD420251055972. A protocol for this systematic review was not prepared.

Search Strategy

A comprehensive literature search was performed by 3 independent reviewers (TTYC, CYF, and AKCW) across 6 electronic databases: PubMed, MEDLINE, CINAHL, Embase, Web of Science, and the Cochrane Library. Studies published between April 1, 2005, and March 31, 2025, were included. The reference lists of included studies were also hand-searched independently by the 3 reviewers (TTYC, CYF, and AKCW) to identify additional eligible studies. Any disagreements among the reviewers were resolved through discussion.

Only peer-reviewed articles were considered; gray literature and review articles were excluded to ensure the inclusion of high-quality evidence. Search strategies were developed using key terms including “prediabetes” and “text messaging” and were expanded using Medical Subject Headings (MeSH) terms where appropriate. Detailed search strategies for each database are provided in Multimedia Appendix 1.

Study Selection

Studies were eligible for inclusion if they met the following criteria: randomized controlled trials (RCTs), participants aged ≥18 years diagnosed with prediabetes, interventions consisting of lifestyle modification programs (ie, diet control, exercise, or behavioral therapy), and use of text messages—defined as SMS, multimedia messaging service, or instant messaging via mobile apps—as the primary intervention delivery channel.

The exclusion criteria were as follows: participants diagnosed with type 2 diabetes mellitus (DM) or gestational diabetes, participants taking medications known to affect glucose tolerance, and studies without complete analytical outcome data.

Outcome

The primary outcome of interest was change in BMI. Secondary outcomes included changes in body weight, waist circumference, HbA_{1c}, total cholesterol, and incidence of DM.

Quality Assessment and Risk of Bias

The risk of bias for the included studies was assessed using the Risk of Bias tool (Cochrane Collaboration). This tool evaluates the following five domains: (1) bias arising from the randomization process, (2) bias due to deviations from intended interventions, (3) bias due to missing outcome data, (4) bias in the measurement of outcomes, and (5) bias in the selection of reported results. Any disagreements among the 3 reviewers were resolved through discussion. As there were fewer than 10

XSL•FO
RenderX

included studies, a funnel plot to assess publication bias was not constructed.

Certainty of Evidence (Grading of Recommendations, Assessment, Development, and Evaluation)

Beyond study-level risk of bias, we appraised the outcome-level certainty of evidence using the GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach for each meta-analyzed outcome (BMI, body weight, waist circumference, HbA_{1c}, total cholesterol, and diabetes incidence). Because all included studies were RCTs, each outcome began at high certainty and could be rated down for risk of bias, inconsistency, indirectness, imprecision, and publication bias. The domains were assessed as follows:

- Risk of bias; considered the risk of bias domain ratings contributing to each outcome
- Inconsistency; considered heterogeneity (I^2 and between-study variability) and overlap of study CIs
- Indirectness; assessed population, intervention, comparator, and outcome alignment with the review question
- Imprecision; considered 95% CIs (including whether they crossed the line of no effect) and optimal information size
- Publication bias; formal tests were not performed given <10 studies; we qualitatively considered small-study effects and selective reporting.

We present GRADE ratings and explicit downgrading rationales in a summary of findings table (Table S1 in Multimedia Appendix 2) and detailed footnotes (Multimedia Appendix 2). No outcomes were upgraded.

Data Extraction

Data from the included studies were retrieved and independently entered into a Microsoft Excel spreadsheet by 3 independent reviewers (TTYC, CYF, and AKCW). Additional variables—including author, study location, study population, study duration, intervention and control details, data collection time points, outcome measures, and results—were also extracted from the selected studies (Multimedia Appendix 3).

Data Synthesis and Analysis

Meta-analyses were conducted using Review Manager (version 5.4.1; Cochrane Collaboration). Outcomes were analyzed at the longest reported follow-up time point and illustrated by forest plots. For continuous variables, mean differences (MDs) with 95% CIs were calculated under a random-effects model. The meta-analysis was conducted using the postintervention values reported by the individual studies. For dichotomous variables, odds ratios and 95% CIs were computed using the inverse variance method.

Statistical heterogeneity was assessed using the chi-square test, the I^2 statistic, and associated P values ($P<.05$ was considered statistically significant). I^2 values were interpreted as follows: unimportant (0% - 40%), moderate (30% - 60%), substantial (50% - 90%), or considerable (75% - 100%) heterogeneity. Given the limited number of included trials (<10), subgroup analyses, meta-regression, and sensitivity analyses were not conducted.

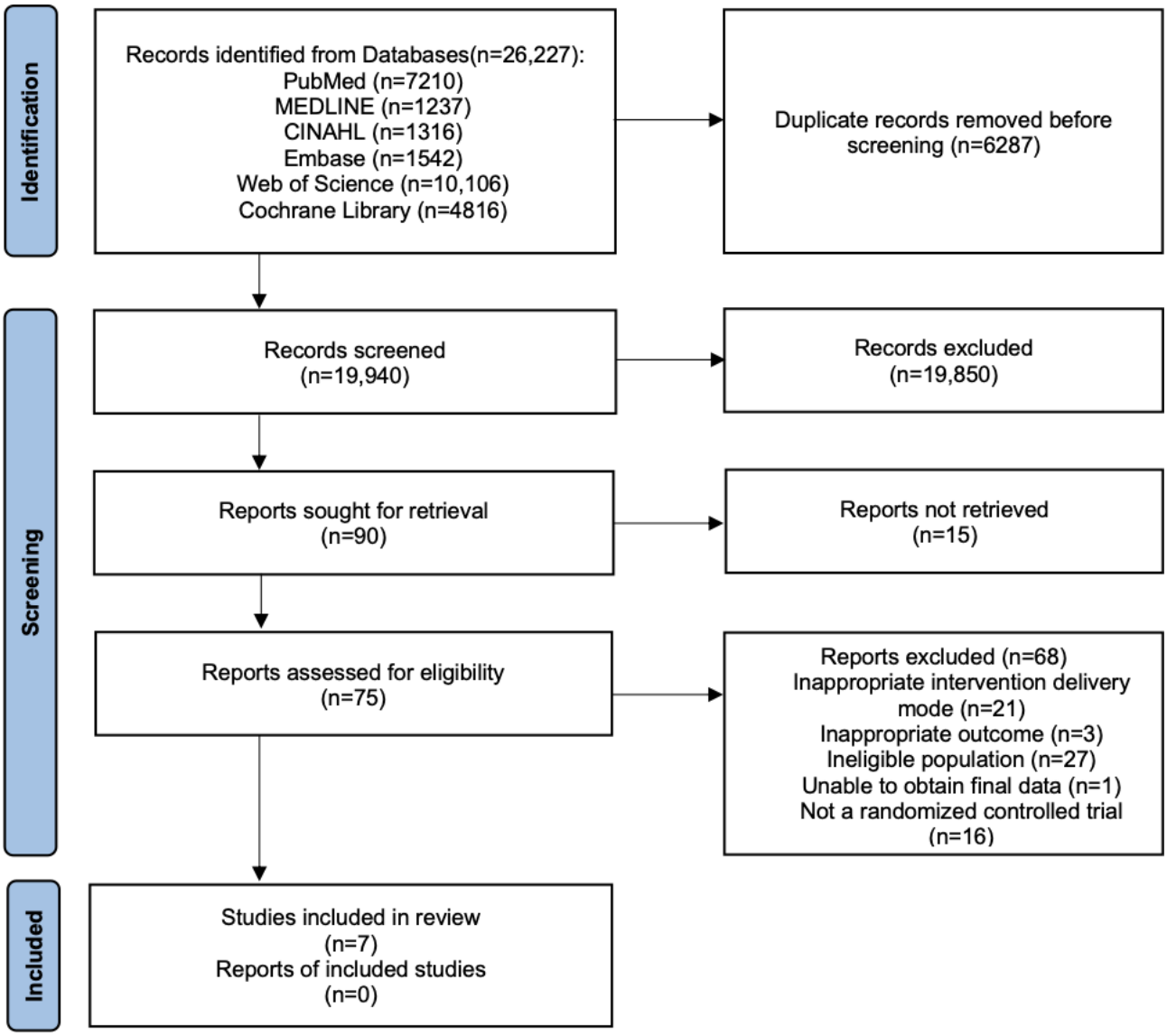
Results

Study Selection

A total of 19,940 publications were identified after duplicate removal. Title and abstract screening excluded 19,850 (99.5%) publications. Of the remaining 90 publications, 75 (83.3%) were available for full-text review and assessed for eligibility. In total, 68 (90.7%) publications were excluded, primarily due to inappropriate intervention delivery modes, ineligible populations, or study designs other than RCTs. Only 7 (9.4%) publications met the eligibility criteria and were included in this review [16,23-28] (Figure 1).

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flowchart.



Risk of Bias Assessment

Risk of bias among the 7 included RCTs, assessed using the Cochrane Risk of Bias tool, ranged from “some concerns” to “high risk.” All included studies demonstrated low risk in the domain “bias in selection of the reported result.” However, 3 of 7 included studies were rated as “some concerns” for the domains of bias arising from the randomization process, missing outcome data, and measurement of outcomes. All included studies were rated as “some concerns” for bias due to deviations from intended interventions, likely reflecting the nonblinded nature of the interventions. A summary of the risk of bias assessments is presented in Table 1.

XSL•FO
RenderX

Table . Risk of bias of the included studies.

Study	Randomization process	Deviation from intended intervention	Missing outcome data	Measurement of outcomes	Selection of the reported result	Overall risk of bias
Al-Hamdan et al [26]	Some concerns	Some concerns	Low	Low	Low	Some concerns
Bootwong and Intarut [16]	Some concerns	Some concerns	Low	Low	Low	Some concerns
Chung et al [27]	Low	Some concerns	Some concerns	Some concerns	Low	High
Khunti et al [28]	Low	Some concerns	Some concerns	Some concerns	Low	High
Nanditha et al [23]	Low	Some concerns	Some concerns	Low	Low	Some concerns
Ramachandran et al [24]	Low	Some concerns	Low	Some concerns	Low	Some concerns
Wong et al [25]	Some concerns	Some concerns	Some concerns	Low	Low	High

Certainty of Evidence (GRADE) Summary

Outcome-level certainty ratings and rationales are reported in Table S1 (summary of findings) in Multimedia Appendix 2. In brief, certainty was most commonly rated down for inconsistency (between-study heterogeneity) and imprecision (wide CIs and limited sample size per outcome). Publication bias could not be formally assessed due to fewer than 10 studies, and no outcomes were upgraded.

Characteristics of Participants and Studies

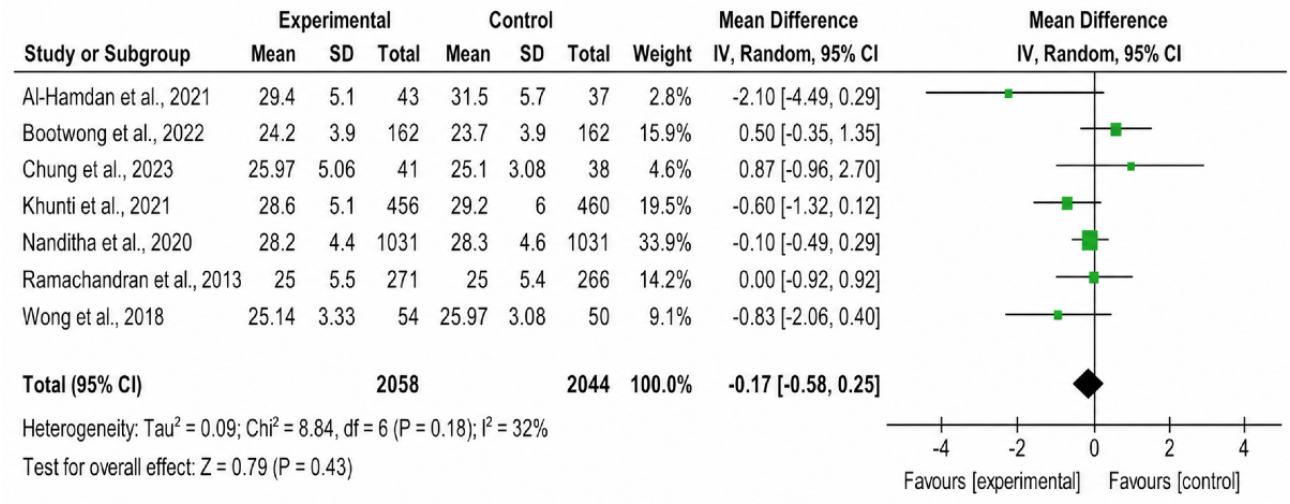
The 7 included RCTs enrolled a total of 4632 participants diagnosed with prediabetes. All participants owned smartphones and were free of major comorbidities. The mean age of participants ranged from 43.5 to 54.2 (SD 5.6) years. Of 4632 participants, 2843 (61.4%) were male.

Among the participants, 2056 (44.4%) individuals received lifestyle modification programs incorporating SMS text messaging, while 2576 (55.6%) individuals received comparable programs without SMS text messaging support. Intervention durations ranged from 12 to 48 months, with 24 months (n=2) and 48 months (n=2) being the most common durations. Delivery formats varied: 4 studies delivered regular text messages promoting lifestyle modification [16,23-25], 2 studies combined mHealth apps with SMS text messaging support [26,27], and 1 study integrated face-to-face sessions with text message follow-ups [28]. For the control group, usual care included related education (n=4), paper-based health information materials (n=2), and physical activity management (n=1). Characteristics of all included studies are summarized in Multimedia Appendix 3.

Quantitative Synthesis

The forest plots of both primary and secondary outcomes are shown in Figures 2-7.

Figure 2. Forest plot of the effect of SMS text messaging interventions on BMI compared with standard care in adults with prediabetes: 1 primary outcome; outcome 1.1 BMI [16,23-28].



XSL•FO
RenderX

Figure 3. Forest plot of the effect of SMS text messaging interventions on hemoglobin A_{1c} compared with standard care in adults with prediabetes [23,26-28].

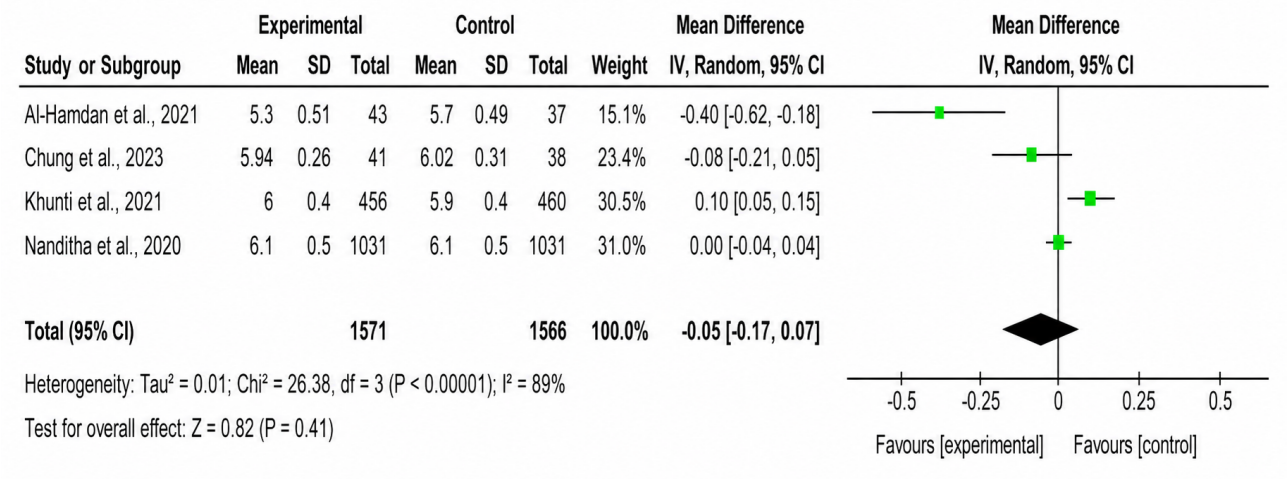


Figure 4. Forest plot of the effect of SMS text messaging interventions on body weight compared with standard care in adults with prediabetes [16,23-26,28].

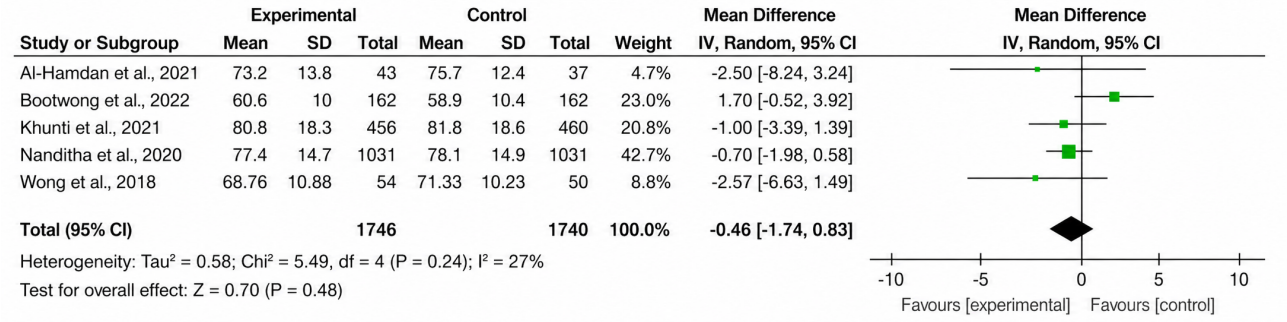
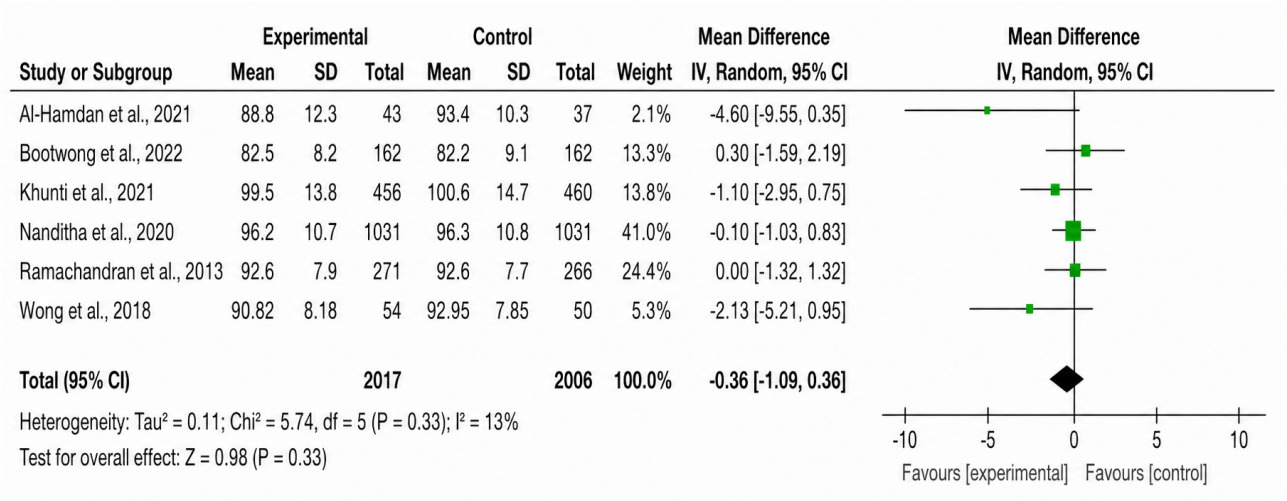


Figure 5. Forest plot of the effect of SMS text messaging interventions on waist circumference [16,23-26,28].



XSL•FO
RenderX

Figure 6. Forest plot of the effect of SMS text messaging interventions on total cholesterol [23-26,28].

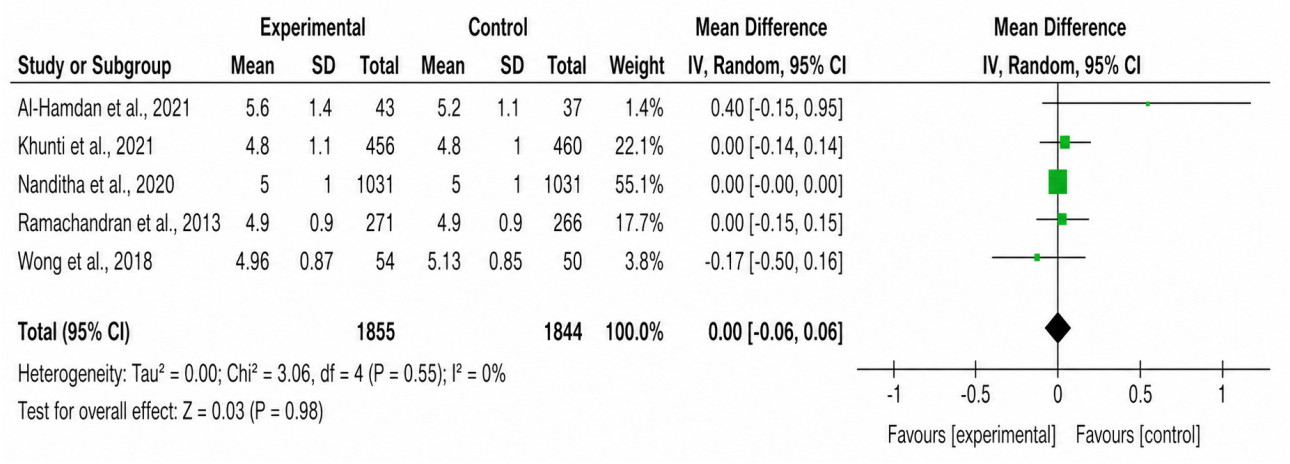
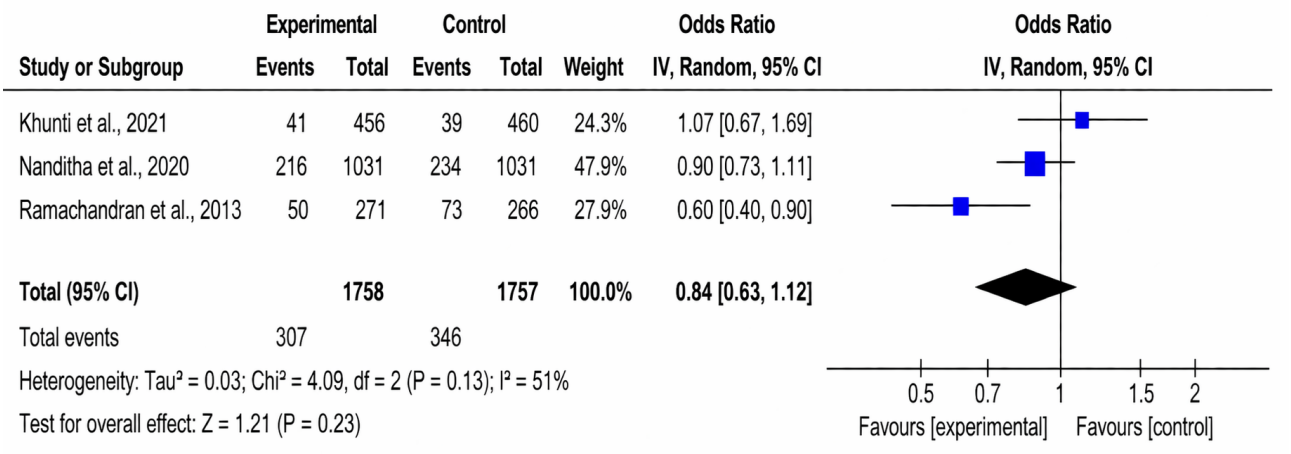


Figure 7. Forest plot of the effect of text messaging interventions on diabetes incidence [23,24,28].



Primary Outcome: BMI

All 7 studies assessed BMI (4102/4634, 88.5% participants). The pooled MD in BMI change was not statistically significant between intervention and control groups (MD –0.17, 95% CI –0.85 to 0.25; *P*=.43), with low heterogeneity observed (χ^2_6=8.8; I^2=32%; *P*=.18).

Secondary Outcomes

Weight

Of the 7 studies, 5 (71.4%) reported weight outcomes. The pooled MD was –0.46 (95% CI –1.74 to 0.83; *P*=.48), indicating no significant difference in weight change between groups. Low heterogeneity was observed (χ^2_5=5.49; I^2=27%; *P*=.24).

Waist Circumference

In total, 6 (85.7%) studies assessed waist circumference. The pooled MD was –0.36 (95% CI –1.09 to 0.36; *P*=.33), with low heterogeneity (χ^2_5=5.74, I^2=13%; *P*<.001).

HbA_{1c}

Overall, 4 studies assessed HbA_{1c} levels. Only 1 study reported a significant improvement favoring the SMS text messaging group. The pooled MD was –0.05 (95% CI –0.17 to 0.07; *P*=.41), with high heterogeneity across studies (χ^2_3=3.06, I^2=89%; *P*=.49).

Total Cholesterol

A meta-analysis of 5 studies showed no significant difference in total cholesterol levels between the intervention and control groups (standardized MD –0.00, 95% CI –0.06 to 0.06; *P*=.98). No heterogeneity was observed (χ^2_4=18.2, I^2=0%; *P*=.55).

DM Incidence

DM incidence was reported in 3 studies (3515/653, 18.6% participants). Moderate heterogeneity was noted (χ^2_2=4.09, I^2=51%; *P*=.13). Among intervention participants, 307 (17.5%) of 1758 developed diabetes, compared with 346 (19.7%) of 1757 in the control group. The odds ratio for DM incidence was 0.84 (95% CI 0.63 to 1.12; *P*=.23), indicating no statistically significant difference between groups.

Discussion

Principal Findings

This systematic review and meta-analysis synthesized evidence on SMS text message–delivered lifestyle programs for adults with prediabetes, focusing on changes in BMI, weight, waist circumference, HbA_{1c}, total cholesterol, and diabetes incidence.

XSL•FO
RenderX

Overall, the pooled estimates did not demonstrate statistically significant improvements in these metabolic outcomes compared with standard care, although several individual trials showed favorable trends.

mHealth technologies can extend the reach of prevention programs by lowering barriers related to access, workforce capacity, and cost. SMS text messaging is especially attractive because it is simple, low-cost, and highly scalable [18,29]. For adults with prediabetes—a group often characterized by low disease awareness and variable motivation—SMS text messaging may provide timely prompts for early lifestyle change. However, in contrast to prior meta-analyses reporting benefits of SMS text messaging for smoking cessation, weight loss, cardiovascular risk reduction, or glycemic control in type 2 diabetes [30-37], our results showed null effects on anthropometry and cardiometabolic markers in prediabetes. These differences likely reflect population context and mechanism: approaches successful in patients with established chronic disease may not translate directly to prevention in largely asymptomatic prediabetes.

Several design features may explain the limited impact observed. First, interventions varied widely in duration, message frequency, delivery channel, and content design. Frequencies ranged from multiple messages per week to biweekly contact; the literature is mixed on whether a higher cadence improves outcomes or induces message fatigue, suggesting an outcome- and population-specific optimum. Second, most programs used standardized, nontailored content. Evidence from digital behavior change research indicates that dynamic tailoring—linking message content, timing, and cadence to a person's current behavior, preferences, and psychosocial state—improves engagement and maintenance [38]. Third, the predominant communication style was unidirectional. Bidirectional designs that solicit brief replies, troubleshoot barriers, and escalate support after nonresponse may enhance accountability and adherence but were rarely implemented. Finally, several trials combined SMS text messaging with additional components (eg, app access, in-person education, or follow-up calls), which improved ecological validity but reduced the ability to attribute effects specifically to SMS text messaging alone and added heterogeneity in behavioral dose and support intensity [39-41].

Interpreting Null BMI Effects Through Body Composition and Fat Distribution

BMI is an imperfect proxy for metabolic risk because it conflates fat and lean mass and cannot distinguish visceral adipose tissue (VAT) from subcutaneous fat or capture ectopic depots (hepatic or intramuscular fat). Small but meaningful reductions in VAT—or preservation or increases in skeletal muscle mass with modest fat loss—can yield little or no net change in BMI while still improving insulin sensitivity, lipid handling, and inflammatory tone. Early behavior change can also redistribute fat (from visceral to subcutaneous compartments) or prevent sarcopenic weight loss, effects that BMI cannot detect.

In our synthesis, waist circumference similarly showed a precise null effect with relatively narrow CIs; however, centimeter-level changes over short follow-up periods may be below detection thresholds, and meaningful body composition adaptations often require ≥6 to 12 months to manifest. Moreover, adherence decay and metabolic adaptation (reductions in energy expenditure during weight loss) can flatten BMI trajectories when interventions are light touch and largely unidirectional, as seen in most included programs. Therefore, future trials should include body composition outcomes (eg, dual-energy X-ray absorptiometry–derived percent fat and lean mass, VAT area by imaging, and multifrequency bioimpedance) and waist-to-height ratio alongside the BMI.

Regarding publication bias, the small number of studies per outcome precluded formal small-study tests. Notably, several pooled estimates centered closely on the null value with relatively narrow CIs (eg, total cholesterol), suggesting that, at the tested dose and design—generic, 1-way prompts over limited durations—SMS text messaging alone is unlikely to generate the sustained energy deficit and resistance-type stimulus required to materially shift BMI or central adiposity in prediabetes. Thus, the persistence of null results appears plausible even if publication bias might ordinarily favor positive findings.

Clinical and Public Health Implications

SMS text messaging appears best positioned as a reach-and-reinforcement layer rather than a stand-alone metabolic intervention. In clinical pathways, brief messaging can reinforce clinic-initiated goals, prompt self-monitoring (weight, steps, and dietary logs), and support attendance at higher-intensity services (group classes, dietitian visits, or structured exercise). At the service level, programs should incorporate closed-loop features (eg, automated outreach after nonresponse, brief telecoaching, or referral triggers when weight plateaus) and integrate device-captured data to enable tailoring.

Commissioners planning population strategies should pair SMS text messaging with an effective dose (≥12 - 24 weeks), link to evidence-based nutrition and progressive resistance training supports, and evaluate outcomes beyond BMI, including waist-to-height ratio, body fat percentage, VAT surrogates, and glycemic variability. Reporting of engagement analytics (delivery, open, reply, and attrition rates) and behavioral mediators (diet quality, step counts, and resistance training adherence) is essential to clarify mechanisms and optimize deployment.

Limitations

Several limitations of this study should be acknowledged. Moderate to high heterogeneity was identified in outcomes such as HbA_{1c} levels and diabetes incidence, primarily due to the limited number of available studies (n=3 - 4) for these outcomes. Variability in study designs—including differences in intervention duration, SMS text messaging content, and follow-up periods—also posed challenges in synthesizing and comparing results across studies. Due to the small number of studies for certain outcomes, conducting meta-regression analyses and subgroup analyses to adjust for these sources of heterogeneity was not feasible, potentially impacting the reliability of pooled estimates. Assessment of funnel plot asymmetry for publication bias was also precluded due to the limited number of studies.

XSL•FO
RenderX

Furthermore, this meta-analysis did not exclude studies that integrated additional mHealth components, such as mobile apps, alongside SMS text messaging. The inclusion of multicomponent interventions may have introduced bias, limiting the ability to isolate the independent effects of SMS text messaging on clinical outcomes in prediabetes. This review is also limited by its focus on BMI, which may require a longer follow-up duration to detect meaningful changes. Other measures of body composition, such as fat distribution or muscle mass, were not assessed.

Conclusions

In summary, this systematic review found that lifestyle modification programs using SMS text messaging did not produce significant improvements in BMI among adults with prediabetes compared with standard care. The limited effectiveness observed may reflect variations in intervention design, including content, frequency, and delivery mode, highlighting the need for more optimized, theory-driven messaging strategies.

Future research should focus on developing standardized, tailored, and interactive SMS text messaging interventions with larger sample sizes and longer follow-up periods. Isolating SMS text messaging as a primary intervention component will also be critical to better determine its role in prediabetes management and diabetes prevention efforts.

Funding

The authors declare that no financial support was received for this study.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: AKCW
Formal analysis: AKCW
Investigation: CCYC, IYMK, CYF, CSW, KLC
Methodology: AKCW, CCYC, IYMK, CYF, CSW, KLC
Project administration: AKCW
Supervision: AKCW
Writing—original draft: SYC, TTYC
Writing—review and editing: AKCW, SYC

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies.
[PDF File, 114 KB - mhealth_v14i1e78521_app1.pdf]

Multimedia Appendix 2
Certainty of evidence (GRADE) summary.
[DOCX File, 29 KB - mhealth_v14i1e78521_app2.docx]

Multimedia Appendix 3
Data extraction table.
[PDF File, 214 KB - mhealth_v14i1e78521_app3.pdf]

Checklist 1
PRISMA checklist.
[DOCX File, 32 KB - mhealth_v14i1e78521_app4.docx]

References

1. Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Res Clin Pract 2019 Nov;157:107843. [doi: 10.1016/j.diabres.2019.107843] [Medline: 31518657]
2. Magliano D, Boyko EJ. IDF Diabetes Atlas, 10th edition: International Diabetes Federation; 2021.

XSL•FO
RenderX

3. Buysschaert M, Medina JL, Buysschaert B, Bergman M. Definitions (and current controversies) of diabetes and prediabetes. Curr Diabetes Rev 2016;12(1):8-13. [doi: 10.2174/1573399811666150122150233] [Medline: 25612821]
4. Prediabetes – your chance to prevent type 2 diabetes. Center for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/diabetes/prevention-type-2/prediabetes-prevent-type-2.html [accessed 2026-04-12]
5. Echouffo-Tcheugui JB, Perreault L, Ji L, Dagogo-Jack S. Diagnosis and management of prediabetes: a review. JAMA 2023 Apr 11;329(14):1206-1216. [doi: 10.1001/jama.2023.4063] [Medline: 37039787]
6. Galaviz KI, Weber MB, Suvada KBS, et al. Interventions for reversing prediabetes: a systematic review and meta-analysis. Am J Prev Med 2022 Apr;62(4):614-625. [doi: 10.1016/j.amepre.2021.10.020] [Medline: 35151523]
7. Lambrinou E, Hansen TB, Beulens JW. Lifestyle factors, self-management and patient empowerment in diabetes care. Eur J Prev Cardiol 2019 Dec;26(2_suppl):55-63. [doi: 10.1177/2047487319885455] [Medline: 31766913]
8. Colberg SR, Sigal RJ, Fernhall B, et al. Exercise and type 2 diabetes: the American College of Sports Medicine and the American Diabetes Association: joint position statement. Diabetes Care 2010 Dec;33(12):e147-e167. [doi: 10.2337/dc10-9990] [Medline: 21115758]
9. Artinian NT, Fletcher GF, Mozaffarian D, et al. Interventions to promote physical activity and dietary lifestyle changes for cardiovascular risk factor reduction in adults: a scientific statement from the American Heart Association. Circulation 2010 Jul 27;122(4):406-441. [doi: 10.1161/CIR.0b013e3181e8edf1] [Medline: 20625115]
10. Saha S, Muradi V, Preethi BL, Kalra P. Assessment of body mass index in prediabetics. Int J Clin Exper Physiol 2021;8(4):155-157. [doi: 10.5530/ijcep.2021.8.4.39]
11. Fischer HH, Durfee MJ, Raghunath SG, Ritchie ND. Short message service text message support for weight loss in patients with prediabetes: pragmatic trial. JMIR Diabetes 2019 Apr 15;4(2):e12985. [doi: 10.2196/12985] [Medline: 30985289]
12. Graham SA, Pitter V, Hori JH, Stein N, Branch OH. Weight loss in a digital app-based diabetes prevention program powered by artificial intelligence. Digit Health 2022;8:20552076221130619. [doi: 10.1177/20552076221130619] [Medline: 36238752]
13. Lim SL, Ong KW, Johal J, et al. A smartphone app-based lifestyle change program for prediabetes (D'LITE study) in a multiethnic Asian population: a randomized controlled trial. Front Nutr 2022;8:780567. [doi: 10.3389/fnut.2021.780567] [Medline: 35141265]
14. Wang Y, Xue H, Huang Y, Huang L, Zhang D. A systematic review of application and effectiveness of mHealth interventions for obesity and diabetes treatment and self-management. Adv Nutr 2017 May;8(3):449-462. [doi: 10.3945/an.116.014100] [Medline: 28507010]
15. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017;12(2):e0170581. [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
16. Bootwong P, Intarut N. The effects of text messages for promoting physical activities in prediabetes: a randomized controlled trial. Telemed J E Health 2022 Jun;28(6):896-903. [doi: 10.1089/tmj.2021.0303] [Medline: 34619066]
17. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415. [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
18. Head KJ, Noar SM, Iannarino NT, Grant Harrington N. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
19. Lev Arey D, Blatt A, Gutman T. A self-determination theory and acceptance and commitment therapy-based intervention aimed at increasing adherence to physical activity. Front Psychol 2022;13:935702. [doi: 10.3389/fpsyg.2022.935702] [Medline: 36051214]
20. Bol N, Smit ES, Lustria ML. Tailored health communication: opportunities and challenges in the digital era. Digit Health 2020;6:2055207620958913. [doi: 10.1177/2055207620958913] [Medline: 33029355]
21. Lauffenburger JC, Barlev RA, Sears ES, et al. Preferences for mHealth technology and text messaging communication in patients with type 2 diabetes: qualitative interview study. J Med Internet Res 2021 Jun 11;23(6):e25958. [doi: 10.2196/25958] [Medline: 34114964]
22. Yoon S, Wee S, Loh DH, Bee YM, Thumboo J. Facilitators and barriers to uptake of community-based diabetes prevention program among multi-ethnic Asian patients with prediabetes. Front Endocrinol (Lausanne) 2022;13:816385. [doi: 10.3389/fendo.2022.816385] [Medline: 35295990]
23. Nanditha A, Thomson H, Susairaj P, et al. A pragmatic and scalable strategy using mobile technology to promote sustained lifestyle changes to prevent type 2 diabetes in India and the UK: a randomised controlled trial. Diabetologia 2020 Mar;63(3):486-496. [doi: 10.1007/s00125-019-05061-y] [Medline: 31919539]
24. Ramachandran A, Snehalatha C, Ram J, et al. Effectiveness of mobile phone messaging in prevention of type 2 diabetes by lifestyle modification in men in India: a prospective, parallel-group, randomised controlled trial. Lancet Diabetes Endocrinol 2013 Nov;1(3):191-198. [doi: 10.1016/S2213-8587(13)70067-6] [Medline: 24622367]
25. Wong CKH, Siu SC, Wong KW, Yu EYT, Lam CLK. Five-year effectiveness of short messaging service (SMS) for pre-diabetes. BMC Res Notes 2018 Oct 10;11(1):709. [doi: 10.1186/s13104-018-3810-y] [Medline: 30309382]
26. Al-Hamdan R, Avery A, Al-Disi D, Sabico S, Al-Daghri NM, McCullough F. Efficacy of lifestyle intervention program for Arab women with prediabetes using social media as an alternative platform of delivery. J Diabetes Investig 2021 Oct;12(10):1872-1880. [doi: 10.1111/jdi.13531] [Medline: 33638248]

XSL•FO
RenderX

27. Chung HW, Tai CJ, Chang P, Su WL, Chien LY. The effectiveness of a traditional Chinese medicine-based mobile health app for individuals with prediabetes: randomized controlled trial. JMIR Mhealth Uhealth 2023 Jun 20;11:e41099. [doi: 10.2196/41099] [Medline: 37338977]
28. Khunti K, Griffin S, Brennan A, et al. Promoting physical activity in a multi-ethnic population at high risk of diabetes: the 48-month PROPELS randomised controlled trial. BMC Med 2021 Jun 3;19(1):130. [doi: 10.1186/s12916-021-01997-4] [Medline: 34078362]
29. Armanasco AA, Miller YD, Fjeldsoe BS, Marshall AL. Preventive health behavior change text message interventions: a meta-analysis. Am J Prev Med 2017 Mar;52(3):391-402. [doi: 10.1016/j.amepre.2016.10.042] [Medline: 28073656]
30. Owei I, Umekwe N, Ceesay F, Dagogo-Jack S. Awareness of prediabetes status and subsequent health behavior, body weight, and blood glucose levels. J Am Board Fam Med 2019;32(1):20-27. [doi: 10.3122/jabfm.2019.01.180242] [Medline: 30610138]
31. Tam HL, Wong EM, Cheung K, Chung SF. Effectiveness of text messaging interventions on blood pressure control among patients with hypertension: systematic review of randomized controlled trials. JMIR Mhealth Uhealth 2021 Sep 22;9(9):e24527. [doi: 10.2196/24527] [Medline: 34550078]
32. Emeran A, Burrows R, Loyson J, Behardien MR, Wiemers L, Lambert E. The effect of text message-based mHealth interventions on physical activity and weight loss: a systematic review and meta-analysis. Am J Lifestyle Med 2024 Aug 15:15598276241268324. [doi: 10.1177/15598276241268324] [Medline: 39554952]
33. Orr JA, King RJ. Mobile phone SMS messages can enhance healthy behaviour: a meta-analysis of randomised controlled trials. Health Psychol Rev 2015;9(4):397-416. [doi: 10.1080/17437199.2015.1022847] [Medline: 25739668]
34. Scott-Sheldon LA, Lantini R, Jennings EG, et al. Text messaging-based interventions for smoking cessation: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016 May 20;4(2):e49. [doi: 10.2196/mhealth.5436] [Medline: 27207211]
35. Skinner R, Gonet V, Currie S, Hoddinott P, Dombrowski SU. A systematic review with meta-analyses of text message-delivered behaviour change interventions for weight loss and weight loss maintenance. Obes Rev 2020 Jun;21(6):e12999. [doi: 10.1111/obr.12999] [Medline: 32043809]
36. Shariful Islam SM, Farmer AJ, Bobrow K, et al. Mobile phone text-messaging interventions aimed to prevent cardiovascular diseases (Text2PreventCVD): systematic review and individual patient data meta-analysis. Open Heart 2019;6(2):e001017. [doi: 10.1136/openhrt-2019-001017] [Medline: 31673381]
37. Haider R, Sudini L, Chow CK, Cheung NW. Mobile phone text messaging in improving glycaemic control for patients with type 2 diabetes mellitus: a systematic review and meta-analysis. Diabetes Res Clin Pract 2019 Apr;150:27-37. [doi: 10.1016/j.diabres.2019.02.022] [Medline: 30822496]
38. Sahin C, Courtney KL, Naylor PJ, E Rhodes R. Tailored mobile text messaging interventions targeting type 2 diabetes self-management: a systematic review and a meta-analysis. Digit Health 2019;5:2055207619845279. [doi: 10.1177/2055207619845279] [Medline: 31041110]
39. Wong AK, Wong FK, Bayuo J, Chow KK, Wong SM, Lee AY. A randomized controlled trial of an mHealth application with nursing interaction to promote quality of life among community-dwelling older adults. Front Psychiatry 2022;13:978416. [doi: 10.3389/fpsyt.2022.978416] [Medline: 36329920]
40. Arambepola C, Ricci-Cabello I, Manikavasagam P, Roberts N, French DP, Farmer A. The impact of automated brief messages promoting lifestyle changes delivered via mobile devices to people with type 2 diabetes: a systematic literature review and meta-analysis of controlled trials. J Med Internet Res 2016 Apr 19;18(4):e86. [doi: 10.2196/jmir.5425] [Medline: 27095386]
41. Wong AKC, Bayuo J, Wong FKY, et al. Experiences of receiving an mHealth application with proactive nursing support among community-dwelling older adults: a mixed-methods study. BMC Nurs 2024 Apr 7;23(1):232. [doi: 10.1186/s12912-024-01909-w] [Medline: 38584270]

Abbreviations

DM: diabetes mellitus
GRADE: Grading of Recommendations Assessment, Development and Evaluation
HbA_{1c}: hemoglobin A_{1c}
MD: mean difference
MeSH: Medical Subject Headings
mHealth: mobile health
RCT: randomized controlled trial
VAT: visceral adipose tissue

XSL•FO
RenderX

Edited by L Buis; submitted 04.Jun.2025; peer-reviewed by C Chiaranai, LD Nys; revised version received 04.Nov.2025; accepted 19.Mar.2026; published 30.Apr.2026.

Please cite as:
Wong AKC, Choi SY, Cheng TTY, Cheng CCY, Kwong IYM, Fong CY, Wong CS, Chan KL
Effectiveness of Text Messaging Interventions on BMI Among Adults With Prediabetes: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2026;14:e78521
URL: https://mhealth.jmir.org/2026/1/e78521
doi:10.2196/78521

© Arkers Kwan Ching Wong, Shun Yan Choi, Tiffany Tsz Yan Cheng, Chelsea Cheuk Yiu Cheng, Ivy Yuk Man Kwong, Cheuk Yin Fong, Chi Sum Wong, Ka Lo Chan. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 30.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluating Effectiveness of mHealth Interventions for Rehabilitation in Patients With Head and Neck Cancer: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Yang Xiao[1,2], MSN; Qiuyu Zhou[1,2], MSN; Feiruo Hong[1,2], PhD; Yuwei Lin[1,2], MSN; Enhong Li[2], MSN; Xuefen Yu[2], PhD

[1]School of Medicine, Zhejiang University, Hangzhou, Zhejiang, China

[2]Stomatology Hospital, School of Stomatology, Zhejiang University School of Medicine, Zhejiang Provincial Clinical Research Center for Oral Diseases, Zhejiang Key Laboratory of Oral Biomedical, Hangzhou, Zhejiang, China

Corresponding Author:
Xuefen Yu, PhD
Stomatology Hospital, School of Stomatology, Zhejiang University School of Medicine, Zhejiang Provincial Clinical Research Center for Oral Diseases, Zhejiang Key Laboratory of Oral Biomedical, Hangzhou, Zhejiang, China

Abstract

Background: Patients with head and neck cancer (HNC) frequently experience functional impairments and psychological distress following surgery or radiotherapy. While mobile health (mHealth) interventions are increasingly integrated into clinical care to support patient self-management and home-based recovery, evidence of their effectiveness in HNC remains inconsistent.

Objective: This study aims to evaluate the effectiveness of mHealth interventions on quality of life (QoL), psychological symptoms, physical symptoms, and functional recovery among HNC survivors.

Methods: This systematic review and meta-analysis followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. Overall, 12 electronic databases were searched for randomized controlled trials published up to December 1, 2025. Two reviewers independently performed study selection, data extraction, and risk of bias assessment using the Cochrane Risk of Bias 2.0 tool. Certainty of evidence was evaluated using the GRADE (Grading of Recommendations, Assessment, Development, and Evaluation) approach. Pooled effects were calculated as standardized mean differences (SMDs) with 95% CIs. The primary outcome was QoL; secondary outcomes included anxiety, depression, fatigue, pain, swallowing function, and maximal interincisal opening (MIO).

Results: A total of 26 randomized controlled trials involving 2385 participants were included. mHealth interventions significantly improved QoL (SMD 0.64, 95% CI 0.41-0.88; *P*<.001), anxiety (SMD −0.75, 95% CI −1.42 to −0.08; *P*=.03), depression (SMD −0.89, 95% CI −1.37 to −0.40; *P*<.001), and fatigue (SMD −0.85, 95% CI −1.19 to −0.51; *P*<.001). Pain showed a small reduction (SMD −0.37, 95% CI −0.49 to −0.24; *P*<.001). However, the improvement in swallowing function reached only borderline significance (SMD 0.66, 95% CI 0.28-1.04; *P*=.04), suggesting that current evidence for this outcome is fragile. No significant effect was found for MIO (*P*=.68). Subgroup analysis revealed that interventions featuring home practice support, self-monitoring, and shorter durations (<3 months) yielded stronger clinical effects. The overall certainty of evidence ranged from low to very low.

Conclusions: mHealth interventions effectively enhance QoL and alleviate psychosocial distress in patients. However, evidence for improving swallowing function and MIO remains insufficient. Future research should prioritize standardized protocols and high-quality trials to validate long-term clinical value.

Trial Registration: PROSPERO CRD42024559825; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024559825

(JMIR Mhealth Uhealth 2026;14:e78109) doi:10.2196/78109

KEYWORDS

mobile health; head and neck cancer; rehabilitation; quality of life; depression; pain; systematic review; meta-analysis

Introduction

Head and neck cancer (HNC) is an umbrella term for tumors that arise in the oral cavity, oropharynx, larynx, hypopharynx, and paranasal sinuses [1]. As the seventh most common cancer globally, HNC accounted for approximately 4.7% of new cases and 4.9% of cancer-related deaths in 2022 [1-3]. Treatment for HNC typically involves surgery, radiotherapy, and systemic therapies such as chemotherapy, targeted therapy, or immunotherapy [4]. Despite advancements in these modalities, patients frequently experience severe treatment-related adverse

XSL•FO
RenderX

effects. Physical impairments, including dysphagia, trismus, fatigue, and pain, frequently co-occur with psychological distress such as anxiety and depression, severely compromising quality of life (QoL) [5,6]. These symptoms can persist long-term, representing a significant burden to comprehensive rehabilitation. Studies have shown that HNC survivors report lower QoL than survivors of other cancer types [7,8].

In oncology, "rehabilitation" refers to interventions that aim to restore or maintain physical function, manage treatment-related symptoms, and improve psychological well-being during the recovery phase [9]. This stage is essential for enhancing the long-term QoL for survivors. Traditional rehabilitation models mainly rely on face-to-face consultations, which face challenges such as uneven resource distribution, limited accessibility, and low patient adherence [9,10]. Geographical barriers often prevent patients in remote areas from receiving timely, high-quality care, leaving adverse reactions poorly managed. Furthermore, time constraints in clinical settings make it difficult for physicians to capture a comprehensive record of patients' daily symptoms. After discharge, the lack of continuous and personalized support often leads to delayed symptom management and increased complications.

Mobile health (mHealth) offers a promising solution to chronic disease management and cancer rehabilitation [11,12]. It encompasses both hardware, such as smartphones, tablets, and wearable activity tracking or sensor devices, and software, including digital health platforms and mobile apps (eg, WeChat mini-programs) [13]. By using wireless connectivity, these tools support real-time symptom monitoring, personalized rehabilitation exercises, and enhanced information accessibility [14,15]. mHealth technologies streamline patient-provider communication, promote home-based self-management, and enhance overall health outcomes. Furthermore, these applications enable patients to track their health status in real time while fostering social and psychological support through connectivity with peers and health care professionals [13].

Although mHealth shows potential, existing evidence remains inconsistent [13,16-24]. Meta-analyses by Buneviciene et al [17] and Zhang et al [24] support the effectiveness of mHealth in improving QoL and alleviating anxiety or depression in general cancer populations. However, these findings primarily derive from mixed-cancer populations. Some reviews indicate that current mHealth interventions for HNC often involve small samples and nonrandomized designs. For instance, Calver et al [18] found insufficient data to confirm the effectiveness of psychological interventions on QoL. Additionally, high heterogeneity in intervention formats and durations persists, with a notable lack of high-quality randomized controlled trials (RCTs) targeting specific outcomes such as swallowing function [13].

Therefore, based on these inconsistent findings and the recent increase in original research, this meta-analysis aims to assess the effectiveness of mHealth interventions in improving QoL, anxiety, depression, fatigue, pain, swallowing function, and maximal interincisal opening (MIO) in patients with HNC. Furthermore, we explored the effectiveness of different interventions, such as intervention delivery format and intervention durations. This study will provide evidence for the future development and clinical implementation of mHealth-based rehabilitation for patients with HNC.

Methods

Ethical Considerations

This article is a systematic review and meta-analysis. It does not contain any studies with human participants or animals performed by any of the authors.

Design

The review protocol was registered with PROSPERO (CRD42024559825). We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) reporting guidelines (Checklist 1) [25].

Data Sources and Search Strategies

We searched 12 electronic databases from their inception to December 1, 2025. These included PubMed, Embase, Cochrane Central Register of Controlled Trials, Web of Science, PsycInfo, CINAHL, Scopus, Google Scholar, China National Knowledge Infrastructure, Chinese Biomedical Literature Database, Weipu (VIP), and Wanfang Data. Search terms included Medical Subject Headings and keywords such as "head and neck cancer," "mHealth," "digital health," and "RCT." We also manually searched the reference lists of relevant articles and screened gray literature, including ClinicalTrials.gov and the World Health Organization International Clinical Trials Registry Platform. Detailed search strategies for each database are provided in Multimedia Appendix 1.

Eligibility Criteria

We used the PICOS (participants, intervention, comparison, outcome, and study design) framework [26] to define our inclusion criteria: (1) participants: patients diagnosed with HNC of any age or gender after surgery or radiotherapy; (2) interventions: mHealth-based tools, such as smartphone apps, WeChat mini-program, wearable devices, or digital health platforms; (3) comparison: usual care, no intervention, or non-mHealth interventions; (4) outcomes: reported at least one targeted outcome, including pain, anxiety, depression, fatigue, QoL, swallowing function, MIO, adherence, self-efficacy, or perceived stress (mean and SD); (5) study design: RCTs; and (6) language: published in English or Chinese.

The exclusion criteria were as follows: (1) studies focused only on diagnosis, screening, or application development without evaluating clinical or patient-reported outcomes; (2) study types such as case reports, reviews, conference abstracts, or qualitative studies; (3) non-mHealth interventions (eg, telephone-only or face-to-face only); and (4) studies with incomplete data or results that could not be extracted.

To accurately evaluate the effectiveness of these interventions, this study adopts the following definition of key concepts. QoL refers to a multidimensional construct covering physical, psychological, and social functions [27]. Swallowing function is a core survival indicator that is often affected by chemoradiotherapy [28]. MIO serves as the key parameter for

XSL•FO
RenderX

evaluating trismus after treatment [29]. Regarding mHealth functions, self-monitoring refers to patients actively recording their own symptoms [30]. Home practice support involves providing tailored rehabilitation guidance at home through digital media [31]. Telemedical support emphasizes bidirectional interaction or real-time feedback between patients and health care providers rather than the use of a simple tool [32].

Study Selection

We used EndNote X21 (Clarivate) software to manage records and deduplication. Two reviewers (YX and QZ) independently screened titles and abstracts. Then, they evaluated the full text of potentially eligible studies. Any disagreements were resolved by a third reviewer (XY). We used Cohen κ to measure the agreement between reviewers, following standard categories: ≤0 (poor), 0.01 - 0.20 (slight), 0.21 - 0.40 (fair), 0.41 - 0.60 (moderate), 0.61 - 0.80 (substantial), and 0.81 - 1.00 (almost perfect agreement) [33].

Data Extraction

Two independent authors (YX and QZ) extracted data using a standardized Microsoft Excel spreadsheet. Extracted information included first author, year, country, setting, patient characteristics, sample size, mean age, intervention and control details, duration, follow-up, outcomes, and assessment tools. For studies with multiple control groups, we selected the most appropriate comparison, such as usual care. Any discrepancies were resolved through discussion with a third author (XY).

Risk of Bias Assessment and Quality of Evidence

Two researchers (YX and QZ) independently assessed the risk of bias in the included RCTs using the Cochrane Risk of Bias 2.0 (RoB 2.0) tool [34]. Evaluations covered 5 domains: the randomization process, deviations from intended interventions, missing outcome data, measurement of the outcome, and selection of the reported results. Each domain was rated as “low risk,” “some concerns,” or “high risk.”

The certainty of the evidence was evaluated using the GRADE (Grading of Recommendations, Assessment, Development, and Evaluation) approach [35]. Evidence from RCTs was initially classified as “high” certainty and downgraded to “moderate,” “low,” or “very low” based on 5 factors: risk of bias, inconsistency, indirectness, imprecision, and publication bias. Conversely, evidence could be upgraded for a large effect size or dose-response gradient. For both RoB 2.0 and GRADE assessments, the κ coefficient was calculated to determine interrater agreement. Any disagreements were resolved through consensus or by consulting a third author (XY).

Statistical Analysis

Statistical analyses were performed using Review Manager (version 5.4; The Cochrane Collaboration), Stata (version 18.0; StataCorp), and R software (version 4.4.1; R Foundation for Statistical Computing). Continuous outcomes were synthesized using standardized mean differences (SMDs) with 95% CIs. For instruments where scoring directions differed, scores were inverted to ensure a uniform orientation for each outcome. Heterogeneity between studies was assessed using the Cochrane Q test (significance level at $P<.10$) and the I^2 statistic. An I^2 value >50% indicated significant heterogeneity. A fixed-effects model was used when heterogeneity was low; otherwise, a random-effects model was applied [36]. Statistical significance was set at $P<.05$. The magnitude of SMD was classified as small (<0.4), medium (0.4 - 0.7), or large (>0.7) [37]. Additionally, 95% prediction intervals (PIs) were calculated for outcomes with significant heterogeneity [38]. These intervals estimate the range of treatment effects expected in future clinical settings, offering a more robust interpretation of the observed variance. Subgroup and sensitivity analyses were performed to explore sources of heterogeneity. Publication bias was evaluated using funnel plots and the Egger regression test [39].

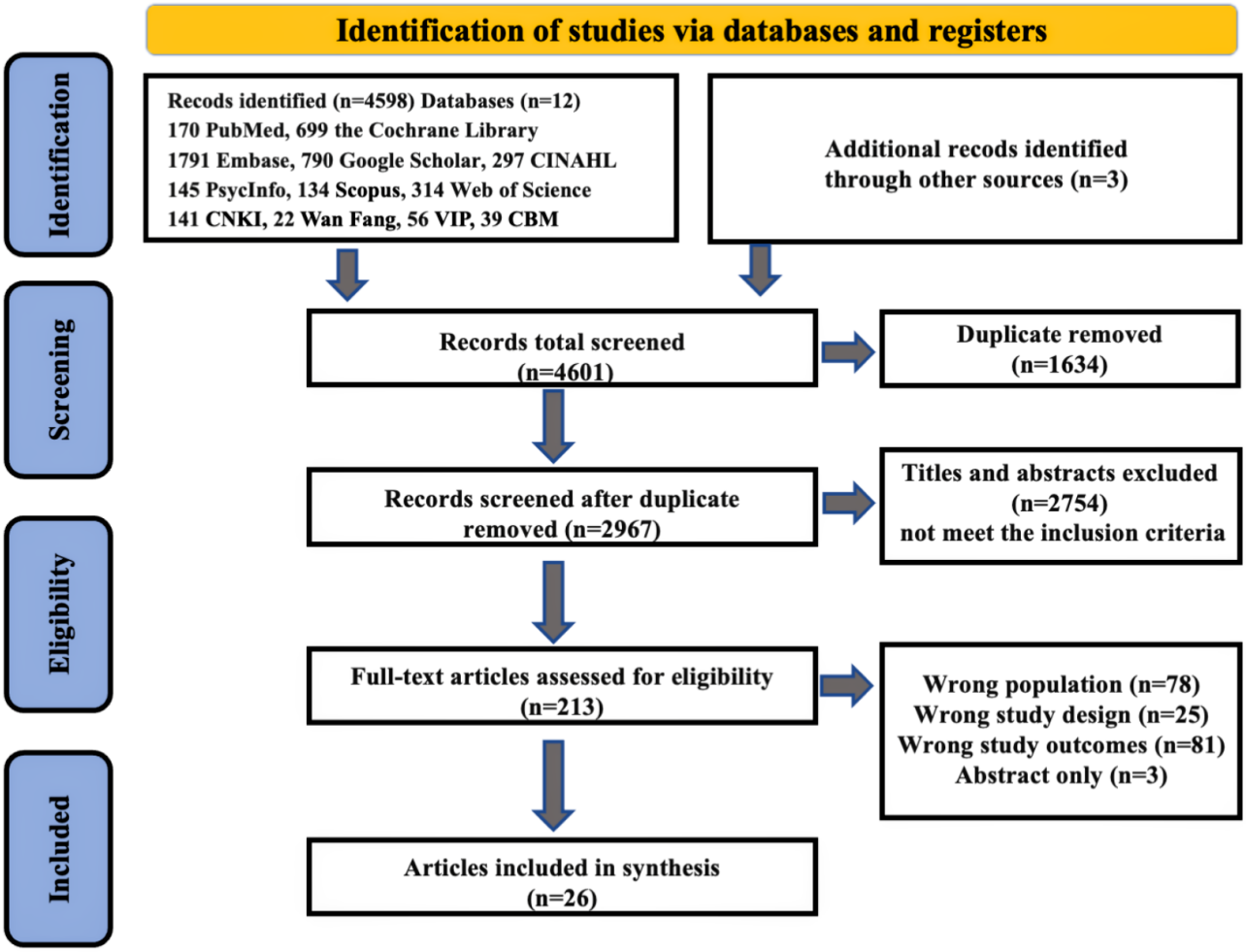
Results

Search Outcomes

The literature search yielded 4601 records, of which 1634 were removed as duplicates. The remaining 2967 publications were evaluated at the title/abstract phase using predetermined inclusion/exclusion criteria, resulting in 2754 exclusions. Of 213 articles advancing to full-text assessment, 187 were excluded for reasons detailed in the PRISMA flowchart (Figure 1). Ultimately, 26 RCTs were included in the systematic review and meta-analysis [40-65]. The interrater reliability for study selection was substantial at the title and abstract screening stage (Cohen κ=0.85) and reached excellent agreement during the full-text review (κ=0.92).

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of studies included in the review.



Study Characteristics

A total of 2385 participants were included across the 26 studies. The key characteristics of these studies are summarized in Table 1. The included trials were conducted in various countries, including China, Belgium, Canada, Korea, Germany, the Netherlands, Australia, and the United States, with publication dates ranging from 2015 to 2025. Regarding cancer subtypes, nasopharyngeal carcinoma was the most frequently studied (n=6), followed by oropharyngeal (n=3), laryngeal (n=3), oral (n=3), and thyroid cancer (n=2).

XSL•FO
RenderX

Table . General characteristics of the included studies.

Study	Country	Design	Population	Sample size, n (intervention/control)	Intervention	Control group	Intervention duration	Outcomes and outcome measures
Baudelet et al 2023 [40]	Belgium	RCT[a]	Oropharyngeal cancer	148 (49/49)	App-supported prophylactic swallowing exercises (home practice support)	Paper group/therapist group	1 month	Adherence
Cen and Yang 2019 [41]	China	RCT	Nasopharyngeal carcinoma	140 (74/75)	Health management app (follow-up reminders)	Routine health education	2 months	EORTC QLQ-H&N35[b]: fatigue, pain, swallowing function
Chen et al 2021 [42]	China	RCT	Laryngeal cancer	100 (50/50)	Mobile education app (home practice support, follow-up reminders)	Usual care	1 month	Anxiety (SAS[c]), depression (SDS[d]), quality of life (self-developed questionnaire)
Dai et al 2024 [43]	China	RCT	HNC[e]	44 (22/22)	Health enjoy system (home practice support, follow-up reminders)	Usual care	3 months	MIO[f] (incisor measuring ruler), self-efficacy (SUPPH[g]), treatment adherence
Di et al 2017 [45]	China	RCT	Nasopharyngeal carcinoma	65 (32/33)	Multimodal smartphone app (self-monitoring and reporting, follow-up reminders)	Usual care	6 months	Quality of life, fatigue, pain (EORTC QLQ-C30[h])
Di and Li 2018 [44]	China	RCT	Nasopharyngeal carcinoma	132 (65/67)	Medical smartphone app (self-monitoring and reporting, telemedical support, follow-up reminders)	Usual care	6 months	Quality of life (EORTC QLQ-C30)
Ducharme et al 2024 [46]	Canada	RCT	HNC	39 (24/15)	PTSD[i] Coach game app (self-monitoring and reporting, home practice support)	Usual care	6 months	Quality of life (FACT-HN[j]), anxiety, depression (HADS)[k]
Kim and Hwang 2025 [47]	Korea	RCT	Thyroid cancer	62 (31/31)	TOP note smartwatch app (self-monitoring and reporting, home practice support, follow-up reminders)	Usual care	3 months	Perceived stress (Korea Perceived Stress Scale, NHANES version)

XSL•FO
RenderX

Study	Country	Design	Population	Sample size, n (intervention/control)	Intervention	Control group	Intervention duration	Outcomes and outcome measures
Lee et al 2023 [48]	Korea	RCT	HNC	40 (27/13)	ePRO-CTCAE app (self-monitoring and reporting, telemedical support)	Diary to record symptoms	2 months	Quality of life (EORTC QLQ-C30)
Liao et al 2020 [50]	China	RCT	Nasopharyngeal carcinoma	80 (40/40)	Home-based digital care platform (home practice support, follow-up reminders)	Usual care	6 months	EORTC QLQ-C30: pain, swallowing function
Liao et al 2022 [49]	China	RCT	Nasopharyngeal carcinoma	116 (58/58)	WeChat mini-program (home practice support, follow-up reminders)	Usual care	6 months	Fatigue (CRF[l] scale), pain, swallowing function (EORTC QLQ-H&N35)
Lyu et al 2016 [51]	China	RCT	HNC	108 (57/51)	WeChat (telemedical support, follow-up reminders)	Telephone follow - up	6 months	Patient satisfaction
Pang et al 2023 [52]	China	RCT	Oral and maxillofacial tumor	64 (33/31)	WeChat app (follow-up reminders)	Usual care	6 months	Mastication function (Modified Sato questionnaire), swallowing function (MDADI[m])
Peng 2019 [53]	China	RCT	Oral cancer	140 (70/70)	WeChat-based interactive education platform (telemedical support, follow-up reminders)	Routine health education	6 months	UW-QOL[n]: quality of life, pain, swallowing function, anxiety; Kessler Psychological Distress Scale: depression
Pfeifer et al 2015 [54]	USA	RCT	HNC	80 (45/35)	Telemedicine for symptom management (self-monitoring and reporting, telemedical support)	Usual care	6 weeks	Quality of life (FACT-HN)
Rades et al 2022 [55]	Germany	RCT	HNC	53 (25/28)	Reminder App of skin and oral care (follow-up reminders)	Usual care	3 months	Adherence, oral mucositis

XSL•FO
RenderX

Study	Country	Design	Population	Sample size, n (intervention/control)	Intervention	Control group	Intervention duration	Outcomes and outcome measures
Sprave et al 2023 [56]	Germany	RCT	HNC	100 (50/50)	ePRO Monitoring System (self-monitoring and reporting, telemedical support)	Usual care	3 months	Quality of life (FACT-HN), fatigue, pain (EORTC QLQ-C30), swallowing function (ESAS[o]), anxiety, depression (HADS)
Starmer et al 2023 [57]	USA	RCT	HNC	91 (44/47)	HNC Virtual Coach with adherence and swallowing support (home practice support, self-monitoring and reporting, follow-up reminders)	Paper exercise logs	3 months	Swallowing function (MDADI), quality of life (FACT-HN)
Sun et al 2024 [58]	China	RCT	Thyroid cancer	104 (62/42)	Multitheory model-based mHealth via WeChat (follow-up reminders)	Usual care	3 months	Anxiety (PHQ-4[p]), depression (PHQ-4), quality of life (EQ-5D VAS[q])
Van Den Brink et al 2007 [59]	Netherlands	RCT	HNC	184 (35/128)	Laptop-based eHealth support system (self-monitoring and reporting, telemedical support)	Usual care	6 weeks	Quality of life (FACT-HN)
van der Hout et al 2020 [60]	Netherlands	RCT	HNC	99 (54/45)	Oncokompas App with HRQOL[r] and symptom monitoring (self-monitoring and reporting, follow-up reminders)	Usual care	6 months	Pain, swallowing function (EORTC QLQ-H&N35)
Wall et al 2020 [61]	Australia	RCT	Oropharyngeal cancer	79 (26/27)	SwallowIT App with remote swallowing support (home practice support, follow-up reminders)	Usual care	3 months	Swallowing function (MDADI), anxiety (HADS-A), depression (HADS-D)
Wang et al 2019 [62]	China	RCT	Oral cancer	60 (30/30)	Telephone-based (telemedical support, follow-up reminders)	Usual care	3 months	Mandibular function impairment (MFIQ[s])

XSL•FO
RenderX

Study	Country	Design	Population	Sample size, n (intervention/control)	Intervention	Control group	Intervention duration	Outcomes and outcome measures
Wang et al 2025 [63]	China	RCT	Laryngeal cancer	80 (40/40)	Internet+ extended care model (home practice support, telemedical support, follow-up reminders)	Usual care	6 months	Swallowing function, quality of life, patient satisfaction (SSA[t])
Wu et al 2025 [64]	China	RCT	HNC	108 (54/54)	Health Belief Model–based Opencare App (home practice support, follow-up reminders)	Usual care	3 months	MIO, treatment adherence, self-efficacy (HBMQ[u]), fatigue (HBMQ), pain (BPI[v]), quality of life (AQoL-6D[w])
Xu et al 2019 [65]	China	RCT	Laryngeal cancer	60 (30/30)	Almond Doctor App (self-monitoring and reporting, home practice support, follow-up reminders)	Usual care	3 months	Anxiety (SAS), depression (SDS), quality of life (QLICP-HN[x])

[a]RCT: randomized controlled trial.

[b]EORTC QLQ-H&N35: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Head and Neck 35.

[c]SAS: Self-Rating Anxiety Scale.

[d]SDS: Self-Rating Depression Scale.

[e]HNC: Head and neck cancer.

[f]MIO: Maximal interincisal opening.

[g]SUPPH: Strategies Used by People to Promote Health.

[h]EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30.

[i]PTSD: Posttraumatic stress disorder.

[j]FACT-HN: Functional Assessment of Cancer Therapy – Head and Neck.

[k]HADS: Hospital Anxiety and Depression Scale.

[l]CRF: Cancer-Related Fatigue scale.

[m]MDADI: M. D. Anderson Dysphagia Inventory.

[n]UW-QOL: University of Washington Quality of Life questionnaire.

[o]ESAS: Edmonton Symptom Assessment System.

[p]PHQ-4: Patient Health Questionnaire-4.

[q]EQ-5D VAS: EuroQol Five-Dimensional Visual Analogue Scale.

[r]HRQOL: Health-Related Quality of Life.

[s]MFIQ: Mandibular Function Impairment Questionnaire.

[t]SSA: Swallowing Satisfaction Assessment.

[u]HBMQ: Health Belief Model Questionnaire.

[v]BPI: Brief Pain Inventory.

[w]AQoL-6D: Assessment of Quality of Life 6D.

[x]QLICP-HN: Quality of Life Instruments for Cancer Patients – Head and Neck.

Characteristics of Interventions and Outcome Measures

The mHealth interventions exhibited significant diversity in design and functionality. The most prevalent delivery mode was smartphone apps (n=14), which typically integrated features such as symptom monitoring, educational modules, clinical reminders, and remote follow-up (Multimedia Appendix 1). A notable subset of studies (n=6), primarily conducted in China, used WeChat-based platforms, including mini-programs and

XSL•FO
RenderX

interactive education systems. Other delivery formats included smartwatch-based apps, web-based platforms, and multicomponent telemedicine models. Functional components often focused on symptom self-monitoring, home-based exercise guidance, and bidirectional clinician-patient communication. Several interventions incorporated theoretical frameworks, such as the Health Belief Model, while others used gamification or virtual coaching to improve patient engagement. Intervention durations ranged from 4 weeks to 12 months.

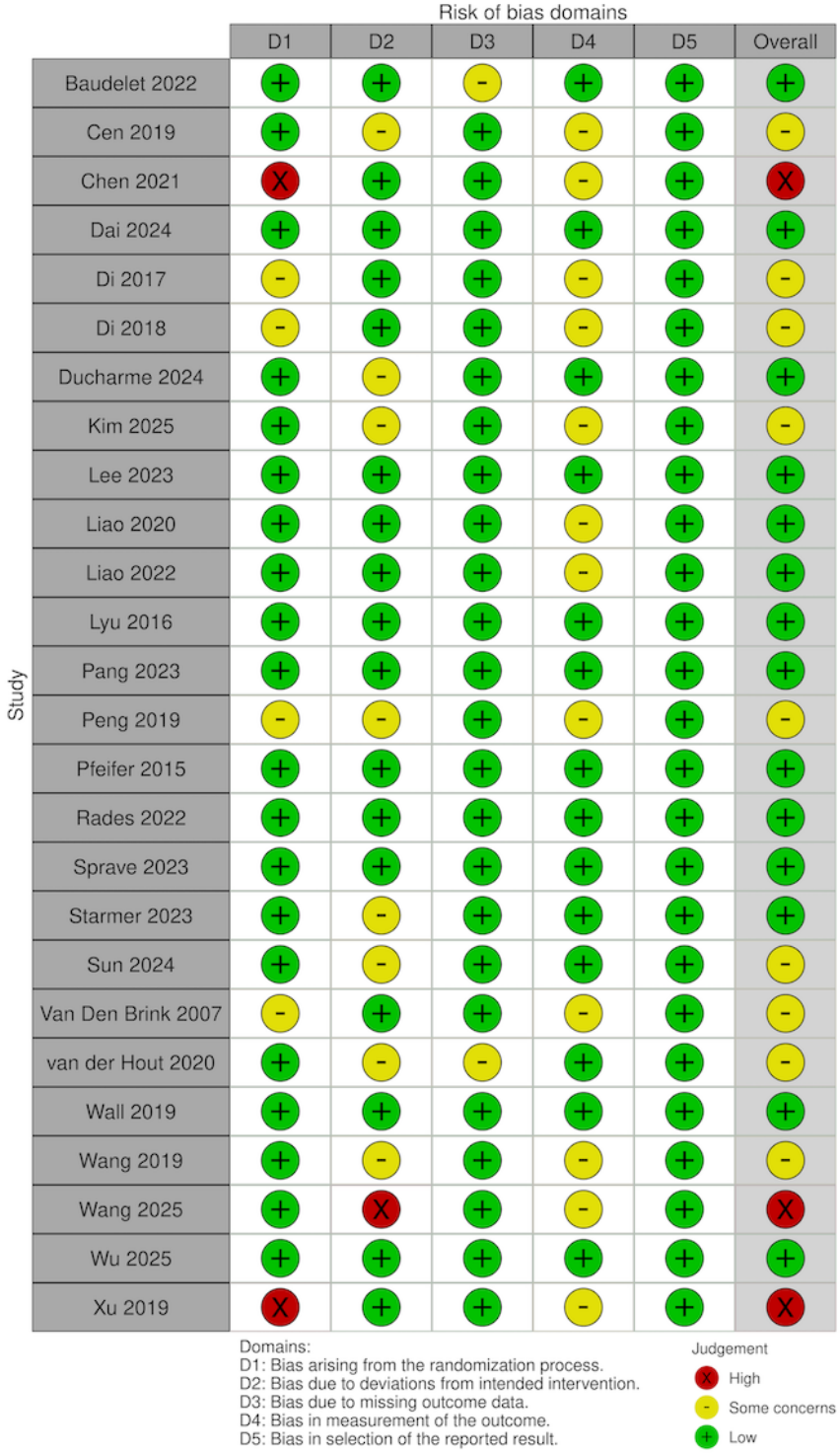
All assessment tools used in the included studies were validated instruments. QoL was most frequently measured using the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30) and its head and neck-specific module (EORTC QLQ-H&N35; 10 studies), followed by the University of Washington Quality of Life questionnaire (n=3) and the Functional Assessment of Cancer Therapy–Head and Neck (n=2). For psychological outcomes, the Hospital Anxiety and Depression Scale was the most common (n=6), followed by the Self-Rating Anxiety/Depression Scale (n=3) and Patient Health Questionnaire-9/Generalized Anxiety Disorder-7 (n=4). Fatigue and pain were primarily assessed through subscales of the EORTC tools. Swallowing function was evaluated using the M. D. Anderson Dysphagia Inventory (n=3) or EORTC QLQ-H&N35 (n=2). Additional outcomes included MIO and self-efficacy (using the Strategies Used by People to Promote Health or the Health Belief Model Questionnaire, and the Mandibular Function Impairment Questionnaire).

Risk of Bias

Methodological quality was evaluated by two reviewers independently using the RoB 2.0 tool. The interrater consistency was high, with κ coefficients ranging from 0.82 to 1.00. Detailed results are presented in Figure 2. While all studies were RCTs, 8 did not specify the method of allocation concealment, resulting in “some concerns” in the randomization process. Due to the nature of mHealth interventions, blinding of participants and personnel was largely unfeasible; thus, most studies were rated as “some concerns” regarding deviations from intended interventions. Three studies were classified as “high risk” for attrition bias due to substantial missing data or high dropout rates. Regarding outcome measurement, only 6 studies reported the use of blinded assessors. Overall, 14 studies were rated as having a “low risk” of bias [40,43,46,48-52,54-57,61,64], 9 as “some concerns” [41,44,45,47,53,58-60,62],” and 3 as “high risk” [42,63,65].” Despite the risk of attrition in certain trials—often due to disease progression or treatment-related mortality—these studies were retained in the analysis to ensure a comprehensive synthesis of the current evidence. GRADE assessment of the 7 outcomes indicated that the overall quality of evidence rating ranged from low to very low (Multimedia Appendix 1). This was primarily due to the high risk of bias in most studies, substantial heterogeneity, and imprecision in certain results.

XSL•FO
RenderX

Figure 2. Risk of bias assessment [40-65].



Qualitative Synthesis of Findings

The included studies indicate that mHealth interventions for patients with HNC primarily focus on 4 domains: health education, self-monitoring, psychological support, and remote guidance. Most platforms adopted a multicomponent approach to facilitate self-management. For instance, Ducharme et al [46] used the "PTSD Coach" app, which integrated psychoeducation with coping tools for self-assessment and social support. Kim et al [47] developed "TOP note," a theory-based application that combined goal setting with smartwatch-linked feedback to manage stress and sleep among thyroid cancer survivors. The degree of clinical interaction varied across platforms; for example, Baudelet et al [40] found that while app-based exercise programs provided clear digital instructions and reminders, the absence of real-time clinical feedback potentially attenuated therapeutic efficacy compared to traditional therapist-led modalities.

XSL•FO
RenderX

Regarding psychological outcomes, mHealth interventions showed significant potential. Ducharme et al [46] reported improved anxiety, depression, and QoL scores at the 3-month follow-up. Kim et al [47] observed significant reductions in perceived stress and improvements in both subjective and objective sleep quality. However, the evidence for physical rehabilitation was more heterogeneous. While mHealth tools were generally well-accepted, patient adherence remained a critical challenge, particularly during intensive treatment phases. Baudelet et al [40] noted that adherence to app-based swallowing exercises declined over time and was lower than that of face-to-face therapy.

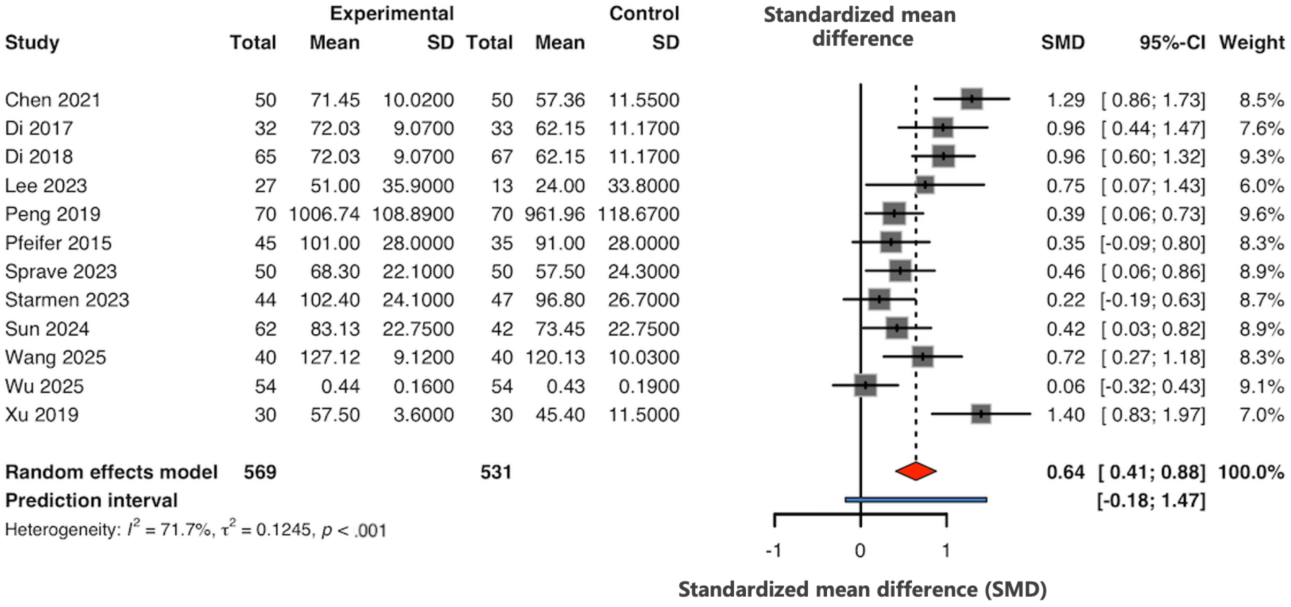
Other studies emphasized different strengths of mHealth. Lyu et al [51] used WeChat for postdischarge follow-up and found it improved patient satisfaction, reduced anxiety, and supported continuity of care. Van Den Brink et al [59] used an early web-based system that helped detect serious complications early and gave patients a strong sense of security. Rades et al [55] tested a simple reminder app for skin and mouth care during radiotherapy; while it did not reach statistical significance, it showed a trend toward fewer severe side effects. Additionally, Wall et al [61] showed that the digital swallowing therapy program "SwallowIT" was clinically equivalent to traditional in-person therapy, suggesting it serves as a viable and practical alternative for HNC rehabilitation.

Meta-Analysis Results

Quality of Life

A total of 12 (46%) studies [42,44,45,48,53,54,56-58,63-65] (46%) provided data on QOL. The pooled results indicated that mHealth interventions had a significant positive impact on the QOL of patients with HNC compared to routine care (SMD 0.64, 95% CI 0.41-0.88; I^2=72%) (Figure 3). However, the 95% PI ranged from −0.18 to 1.47, suggesting that the impact of mHealth on QOL may vary across different clinical settings.

Figure 3. Meta-analysis on quality of life [42,44,45,48,53,54,56-58,63-65].

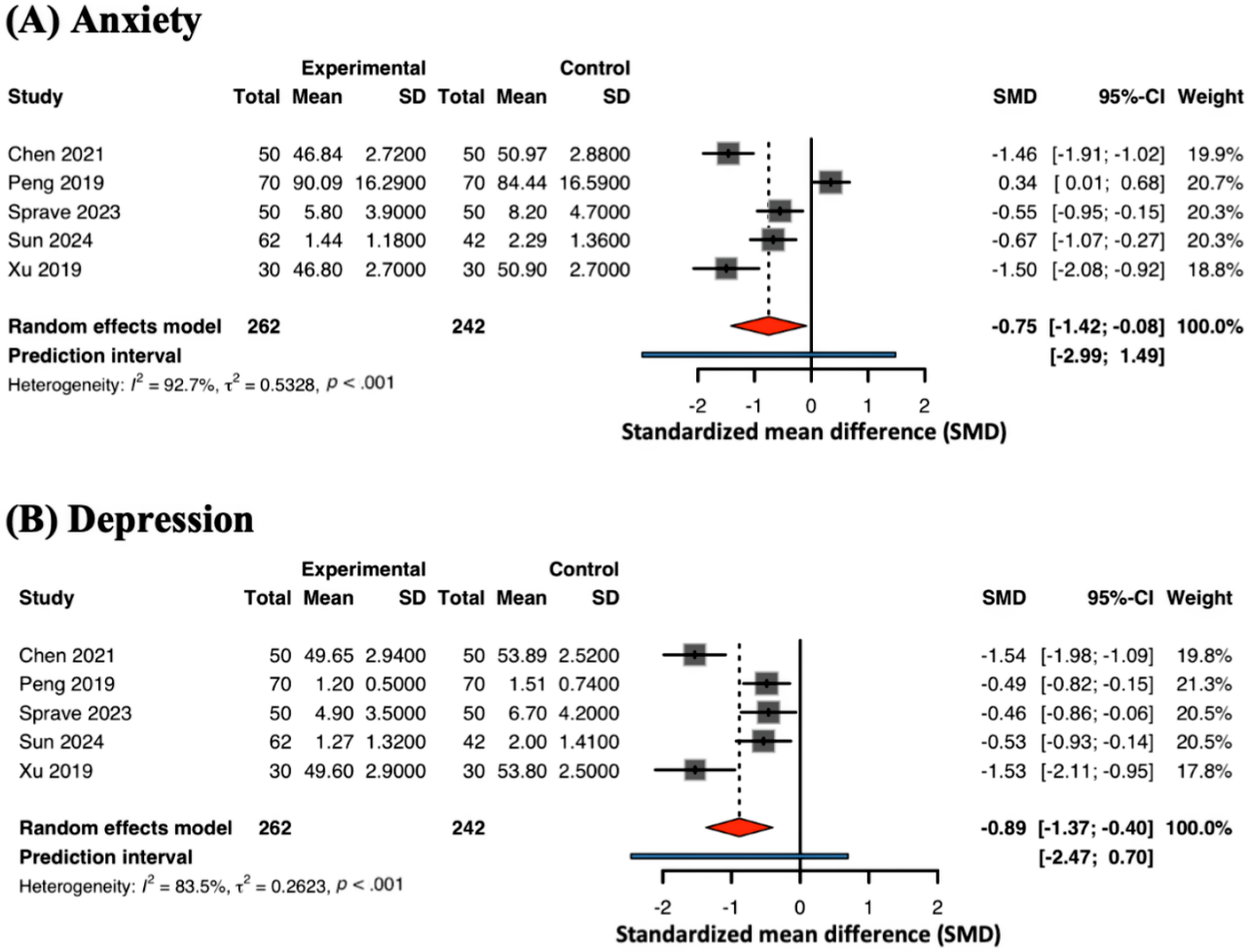


Anxiety and Depression

Five studies [42,53,56,58,65] (19%) evaluated the effect of mHealth on psychological outcomes (Figure 4). The meta-analysis demonstrated that mHealth interventions significantly alleviated anxiety symptoms (SMD −0.75, 95% CI −1.42 to −0.08; I^2=93%). Due to extreme heterogeneity, the 95% PI was wide and encompassed the null value (−2.99 to 1.49), indicating that the intervention's effect on anxiety remains highly variable across different populations. Similarly, 5 studies reported depression outcomes, showing a significant reduction in depressive symptoms (SMD −0.89, 95% CI −1.37 to −0.40; I^2=84%), with a 95% PI of −2.47 to 0.70. These findings suggest that digital platforms are effective on average, though the wide PIs necessitate caution in generalizing these results.

XSL•FO
RenderX

Figure 4. Meta-analysis on (A) anxiety and (B) depression [42,53,56,58,65].



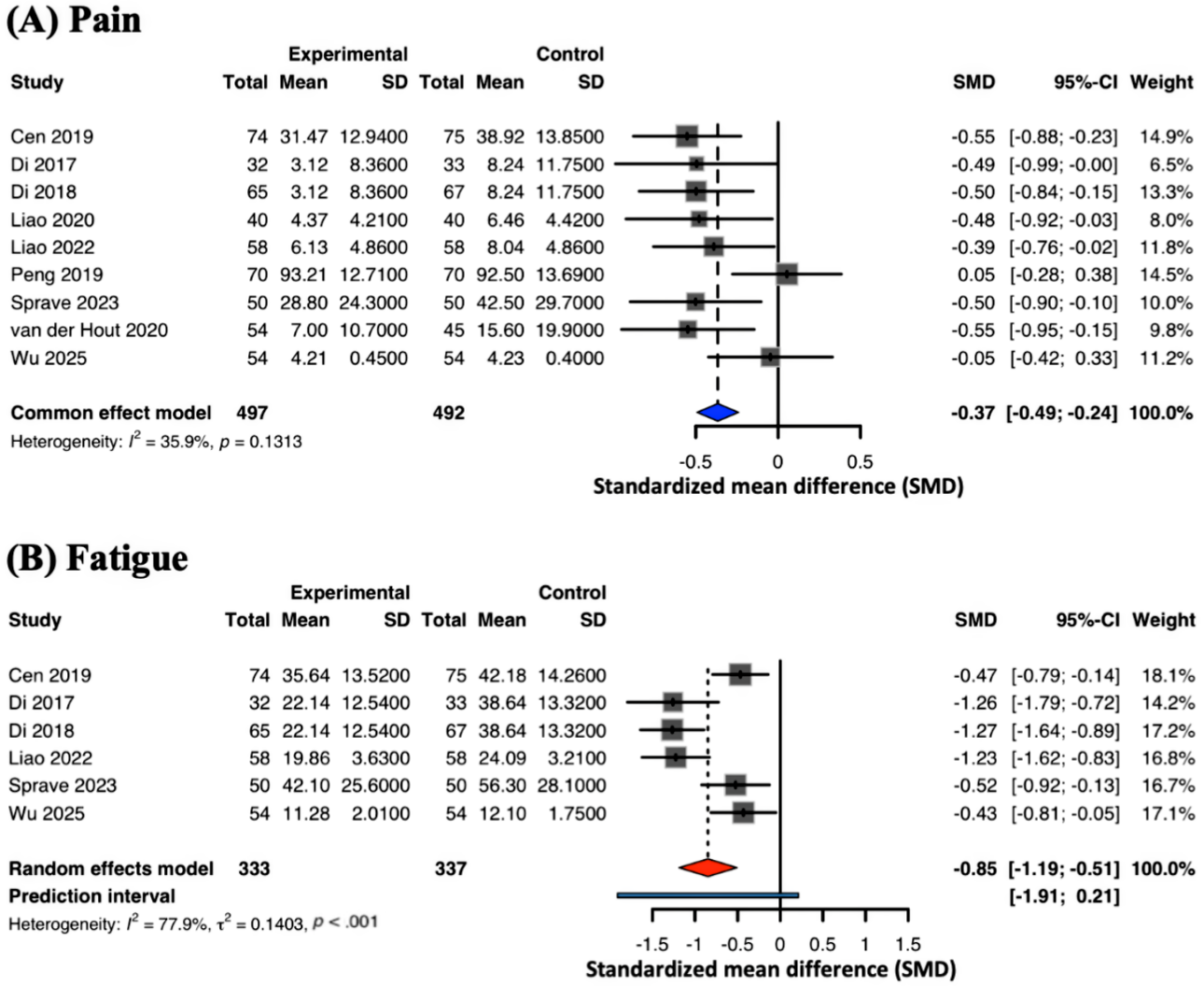
Fatigue and Pain

Fatigue was evaluated in 6 studies [41,44,45,49,56,64] (23%), with the pooled estimate showing a significant beneficial effect of mHealth interventions (SMD −0.85, 95% CI −1.19 to −0.51; I^2=78%) (Figure 5). The 95% PI was −1.91 to 0.21, reflecting considerable variation in expected fatigue reduction.

Nine studies [41,44,45,49,50,53,56,60,64] (35%) assessed pain intensity, and the meta-analysis revealed a small but statistically significant reduction in pain scores (SMD −0.37, 95% CI −0.49 to −0.24; I^2=36%) (Figure 5). The lower heterogeneity in this outcome supports a more consistent effect across studies.

XSL•FO
RenderX

Figure 5. Meta-analysis on (A) pain and (B) fatigue [41,44,45,49,50,53,56,60,64].

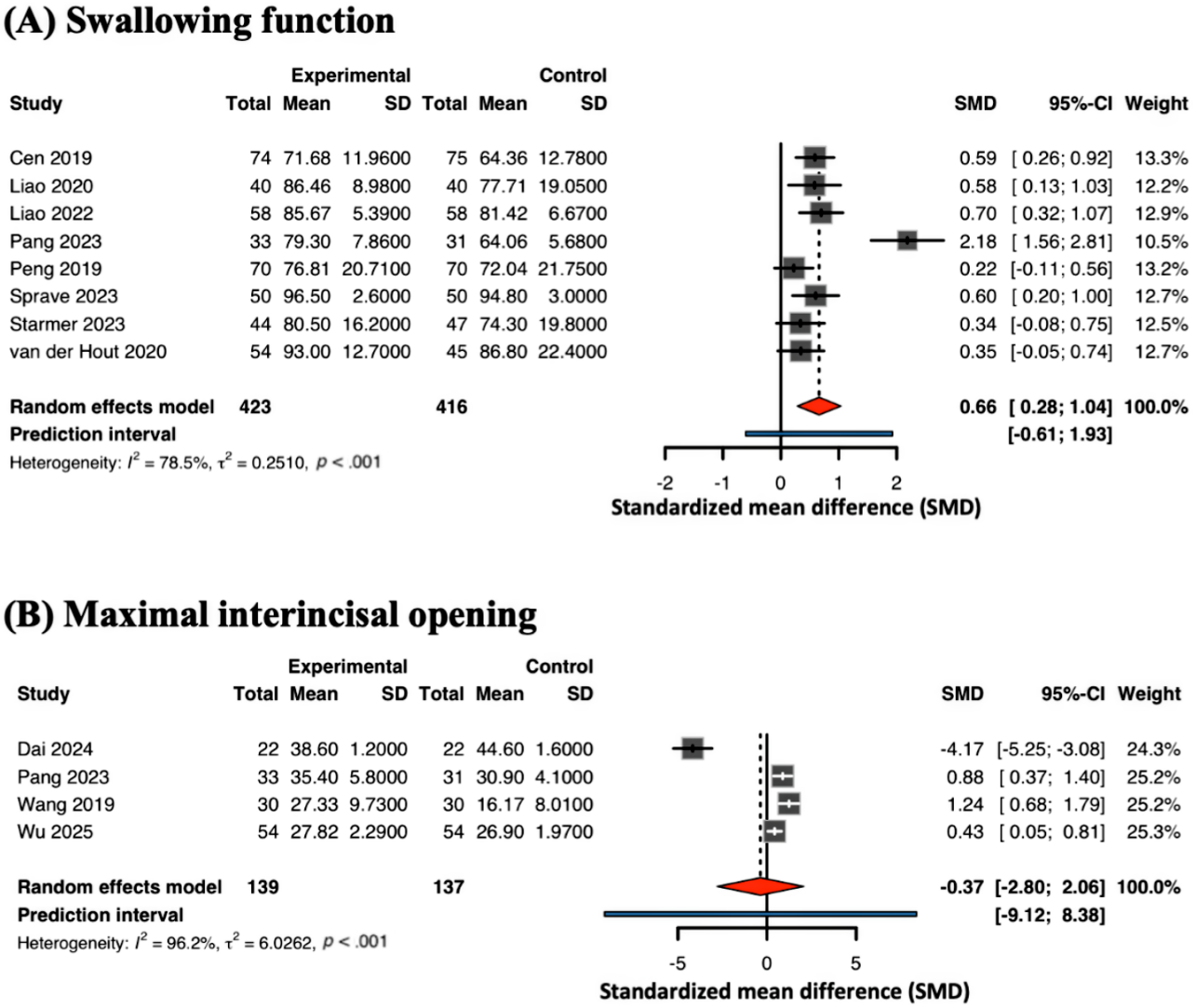


Swallowing Function

Eight studies [41,49,50,52,53,56,57,60] (31%) reported on swallowing function. The improvement in swallowing function showed borderline significance (SMD 0.66, 95% CI 0.28-1.04; I^2=91%) (Figure 6), suggesting that the current evidence is fragile and should be interpreted with caution. The 95% PI ranged from −0.61 to 1.93, which indicates that the “average” effect may not be clinically representative for all future patients, and the intervention’s efficacy remains variable.

XSL•FO
RenderX

Figure 6. Meta-analysis on (A) swallowing function and (B) maximal interincisal opening [41,43,49,50,52,53,56,57,60,62,64].



Maximal Interincisal Opening

Four studies [43,52,62,64] (15%) examined MIO as an outcome. The meta-analysis found no statistically significant difference between mHealth and control groups (SMD −0.37, 95% CI −2.80 to 2.06; *P*=.68; I^2=96%) (Figure 6). The wide confidence interval and very high heterogeneity indicate inconsistency across studies, likely due to differences in intervention intensity, follow-up duration, and baseline function. Pooled effect sizes and 95% CIs for all examined outcomes are summarized in Multimedia Appendix 1.

Subgroup Analysis

As shown in Table 2, subgroup analyses were conducted across outcomes, including anxiety, depression, fatigue, pain, swallow function, and QOL. We stratified the data by assessment scale, intervention format, and intervention duration. Additionally, we examined the impact of specific components, including telemedical support, self-monitoring, and home practice support. Detailed results for each outcome are presented in Multimedia Appendix 1.

Home practice support significantly reduced symptoms of anxiety (SMD −1.48, 95% CI −1.83 to −1.12; *P*<.001; I^2=0%) and depression (SMD −1.53, 95% CI −1.88 to −1.18; *P*<.001; I^2=0%), while simultaneously enhancing QoL (SMD 0.72, 95% CI 0.19-1.25; *P*=.008; I^2=86%). Self-monitoring was strongly associated with better swallowing function (SMD 0.50, 95% CI 0.31-0.70; *P*<.001; I^2=0%). The therapeutic impact was found to be moderated by intervention duration. Short-term interventions (<3 months) showed significant effects on both QoL (SMD 0.68, 95% CI 0.38-0.98; *P*<.001; I^2=70%) and swallowing function (SMD 0.59, 95% CI 0.34-0.85; *P*<.001; I^2=0%). Long-term interventions (≥3 months) remained effective for QoL (SMD 0.74, 95% CI 0.44-1.03; *P*<.001; I^2=52%). In contrast, while the pooled effect size for swallowing function appeared positive, it failed to reach statistical significance (SMD 0.42, 95% CI –0.18 to 1.02; *P*=.17; I^2=93%). However, interventions without telemedical support showed stronger effects across all outcomes compared to those with such support (*P*>.05).

XSL•FO
RenderX

Table . Subgroup analyses of mobile health intervention efficacy.

Subgroup variable	Studies, n	SMD[a] (95% CI)	*P* value	I^2 (%)
Anxiety				
Home practice support	2	–1.48 (–1.83 to –1.12)	<.001	0
Self-monitoring	2	–1.00 (–1.93 to –0.07)	.03	86
Telemedical support	3	–0.55 (–1.53 to 0.44)	<.001	85
Without telemedical support	2	–1.06 (–1.84 to –0.29)	.03	85
Depression				
Home practice support	2	–1.53 (–1.88 to –1.18)	<.001	0
Telemedical support	2	–0.48 (–0.77 to –0.19)	.003	83
Without telemedical support	3	–1.18 (–1.48 to –0.89)	<.001	83
Swallowing function				
Self-monitoring	4	0.50 (0.31 to 0.70)	<.001	0
Duration: short-term (≤3 months)	2	0.59 (0.34 to 0.85)	<.001	0
Duration: long-term (>3 months)	7	0.42 (–0.18 to 1.02)	.17	93
Scale: EORTC[b]-based	5	0.57 (0.40 to 0.74)	<.001	0
Quality of life				
Home practice support	5	0.72 (0.19 to 1.25)	<.001	86
Duration: short-term (≤3 months)	7	0.68 (0.38 to 0.98)	<.001	70
Duration: long-term (>3 months)	4	0.74 (0.44 to 1.03)	<.001	52
Pain				
Self-monitoring	4	–0.48 (–0.66 to –0.31)	<.001	65
Scale: EORTC-based	6	–0.50 (–0.64 to –0.35)	<.001	0
With telemedical support	4	–0.34 (–0.52 to –0.16)	<.001	62
Without telemedical support	4	–0.39 (–0.56 to –0.22)	<.001	10

[a]SMD: Standardized mean difference.

[b]EORTC: European Organisation for Research and Treatment of Cancer.

Sensitivity Analysis

The sensitivity analyses for the meta-analysis results of anxiety, depression, fatigue, QoL, and pain showed that the pooled estimates remained robust when any single study was excluded. However, swallowing function reached only borderline significance in the random-effects model and was not robust in sensitivity analysis. The results of MIO revealed that after excluding the study by Dai et al [43], the overall pooled result changed from nonstatistically significant (SMD −0.37, 95% CI −2.80 to 2.06; *P*=.68; I^2=96%) to statistically significant (SMD 0.81, 95% CI 0.34-1.28; *P*<.001; I^2=66%). Detailed sensitivity analysis results are presented in Multimedia Appendix 1.

Publication Bias

Publication bias was assessed using funnel plots and Egger tests including QoL, pain, and swallowing function. For other indicators, the number of included studies was insufficient to perform these tests. Visual inspection of the funnel plots showed general symmetry, and Egger tests revealed no evidence of significant publication bias for QoL (*P*=.59), pain (*P*=.78), or swallowing function (*P*=.91) (Multimedia Appendix 1).

Discussion

Principal Findings

HNC survivors often suffer from persistent functional deficits and psychological distress. Compared with other cancers, they require specialized and continuous rehabilitation after discharge.

XSL•FO
RenderX

With the rapid advancement of digital health technologies, mHealth has been an increasingly used approach to support self-management and improve health outcomes in cancer care. This systematic review and meta-analysis provides evidence-based data regarding the efficacy of mHealth in enhancing QoL and alleviating symptoms of anxiety, depression, fatigue, pain, and swallowing function among patients with HNC. Furthermore, we analyzed the influence of various intervention formats and durations on these outcomes, providing evidence for the development of optimized digital health protocols.

Our findings indicate that mHealth interventions significantly enhance QoL and alleviate psychological distress, such as anxiety and depression, aligning with results from previous systematic reviews [17,18,24]. These improvements are primarily attributed to real-time psychological support, peer interaction, and continuous symptom tracking, which help reduce uncertainty about the illness and lower the psychological burden during rehabilitation. Evidence suggests that the mitigation of anxiety and depression is concurrently associated with improved QoL [66]. Despite variations in intervention formats, durations, and target populations, sensitivity analyses confirmed the robustness of these positive psychosocial outcomes. Sun et al [58] used an intervention based on the multitheory model and found significant improvements in anxiety, fear of cancer recurrence, and overall QoL, but no meaningful changes in physical function or somatic symptoms. These results indicate that the primary benefit of such interventions lies in addressing the psychological and emotional challenges faced by cancer survivors. Furthermore, subgroup analyses revealed that “home practice support” is associated with robust improvements in anxiety, depression, and QoL. This indicates that integrating psychological support with specific rehabilitation tasks can enhance self-efficacy and behavioral adherence [58]. The sustained improvement in QoL observed in long-term interventions (≥3 months) further highlights the importance of mHealth in maintaining long-term well-being.

Studies report that pain and fatigue are the most prominent physical symptoms among patients with HNC, particularly in postoperative populations. Our findings indicate that mHealth interventions lead to moderate improvements in both symptoms; specifically, the alleviation of fatigue was relatively pronounced, whereas the reduction in pain was more limited. These findings align with previous systematic reviews [67]. For instance, Hernandez Silva et al [67] reported that self-management-based mHealth interventions effectively improved fatigue in cancer survivors but failed to reach statistical significance for pain symptoms. Among the studies included in this analysis, interventions using self-management applications or WeChat-based platforms showed superior effects in mitigating fatigue and pain [49,56,64]. This efficacy may be attributed to the integration of personalized reminders, symptom tracking, and real-time feedback, which collectively enhance patient self-management capabilities and treatment adherence.

Dysphagia and trismus are common complications following HNC treatment [68]. These impairments can lead to malnutrition, dehydration, weight loss, chronic aspiration, and aspiration pneumonia, significantly affecting patients’ QoL and long-term prognosis. Although this review did not observe an overall significant improvement in MIO through mHealth, the results indicated that self-monitoring significantly enhanced swallowing outcomes. Encouraging patients to actively record and monitor their swallowing status may enhance rehabilitation awareness and behavioral adherence [22]. In addition, the effects on swallowing function were more pronounced in short-term interventions (<3 months), highlighting the importance of early intervention, which aligns with the findings of a recent systematic review [24]. Current evidence on MIO is highly inconsistent and sensitive to individual study inclusion, warranting cautious interpretation. Future research should prioritize large-scale, high-quality trials focused on dysphagia and trismus. Moreover, the long-term benefits of mHealth on these functional outcomes require further validation through rigorous longitudinal studies.

Interestingly, the subgroup analysis showed that interventions without telemedical support appeared to be more effective. This finding is counterintuitive as it challenges the assumption that “human-in-the-loop” interventions are inherently more effective than fully automated systems. For example, some “telemedical” support was limited to low-intensity and noninteractive notifications, which might increase a patient’s cognitive load without providing meaningful clinical feedback. In contrast, simple tools like WeChat may be more successful because they are easier to use in daily life. However, a note of caution is necessary when interpreting this finding. This result may be a statistical artifact specific to the limited number of studies in these subgroups rather than a general rule that “less is more” in mHealth. The better results seen in automated systems likely reflect the high quality of their content and fewer technical difficulties, rather than a disadvantage of human interaction. Future research should focus on identifying which types of human involvement truly contribute to clinical value in HNC rehabilitation.

The extreme heterogeneity observed in outcomes such as anxiety and swallowing function has critical clinical implications. While the pooled SMDs suggest an overall benefit, the 95% PIs for these outcomes encompass the null value. This indicates that the “average” effect may not be clinically representative of every setting, and the results should be interpreted with extreme caution. The observed variance primarily stems from diverse intervention designs (eg, WeChat vs specialized apps), patient population variability (eg, tumor sites and treatment modalities), and inconsistent assessment tools. These findings underscore the need for standardized intervention protocols to better validate the long-term clinical and economic value of mHealth in HNC rehabilitation.

Implications for Clinical Practice

Our findings provide clear guidance for mHealth apps in HNC rehabilitation. Health care professionals should use mHealth to enhance symptom management and QoL, particularly during the first 3 months posttreatment. Effective mHealth tool design should prioritize “home practice support” and “self-monitoring” functions. Using concise mini-programs or lightweight apps is generally more effective and less burdensome than complex

XSL•FO
RenderX

teleconsultation systems. Furthermore, interventions should be based on established behavior change theories and tailored to specific tumor sites and treatment modalities. An ideal care model involves a "hospital-community-home" integrated system, where mHealth platforms facilitate daily tracking and structured training, supported by periodic multidisciplinary feedback. Beyond automated education, platforms should incorporate features for private inquiries and personalized responses. Finally, addressing barriers for older patients and those with low digital literacy through age-appropriate design is essential to ensure equity. Future research requires high-quality, large-scale RCTs to validate functional outcomes, along with systematic evaluations of long-term efficacy and health economic benefits across diverse HNC subtypes and cultural backgrounds.

Strengths and Limitations

This study has several strengths. First, our comprehensive search across multiple databases ensured extensive literature coverage. Second, we exclusively included RCTs, and most of the selected trials had a low risk of bias. In addition, we evaluated both psychosocial and physiological outcomes by detailed subgroup analyses.

Despite these strengths, several limitations must be acknowledged. First, the generalizability of our findings is limited by geographic and platform concentration. Nearly half of the included studies rely on 2 specific platforms: WeChat in China and OncoKompas in the Netherlands. These tools depend on localized digital infrastructures that may not be available elsewhere. Second, most evidence comes from high- or middle-income settings with high smartphone penetration. Consequently, our results may not apply to patients in low-resource settings, rural areas, or low- and middle-income countries. Finally, substantial heterogeneity was observed, and the overall quality of evidence for several outcomes remains low to very low.

Conclusions

mHealth interventions are a promising adjunct for improving QoL and alleviating psychological distress, fatigue, and pain in patients with HNC. Home practice support and self-monitoring are pivotal components for driving efficacy. Future research should prioritize high-quality, large-scale RCTs to further validate the long-term effectiveness and standardized implementation of these digital interventions.

Acknowledgments

The authors would like to express their gratitude to all the researchers whose studies were included in this meta-analysis.

Funding

This work was supported by the Medical Science and Technology Project of Zhejiang Province (grant 2023KY131).

Data Availability

The data that support the findings of this study are available from the original articles included in this systematic review, which are publicly available from the respective journals and databases. Further details can be obtained from the corresponding author upon reasonable request.

Authors' Contributions

YX conceptualized and designed the study, developed the methodology, performed data analysis, validation, and investigation, curated the data, created visualizations, and wrote the initial draft. QZ contributed to data curation and methodology. YL was involved in validation. EL conducted a formal analysis. FH assisted with data curation. XY conceptualized the study and provided supervision. All authors reviewed, edited, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies, GRADE (Grading of Recommendations, Assessment, Development, and Evaluation) evidence profiles, intervention characteristics, and supplementary statistical analyses, including subgroup, sensitivity, and publication bias analyses.
[DOCX File, 12115 KB - mhealth_v14i1e78109_app1.docx]

Checklist 1
PRISMA checklist.
[DOCX File, 22 KB - mhealth_v14i1e78109_app2.docx]

References

XSL•FO
RenderX

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024;74(3):229-263. [doi: 10.3322/caac.21834] [Medline: 38572751]
2. Sun H, Yu M, An Z, et al. Global burden of head and neck cancer: epidemiological transitions, inequities, and projections to 2050. Front Oncol 2025;15:1665019. [doi: 10.3389/fonc.2025.1665019] [Medline: 41079071]
3. Zhou T, Huang W, Wang X, et al. Global burden of head and neck cancers from 1990 to 2019. iScience 2024 Mar 15;27(3):109282. [doi: 10.1016/j.isci.2024.109282] [Medline: 38455975]
4. Amaral MN, Faísca P, Ferreira HA, Gaspar MM, Reis CP. Current insights and progress in the clinical management of head and neck cancer. Cancers (Basel) 2022 Dec 10;14(24):6079. [doi: 10.3390/cancers14246079] [Medline: 36551565]
5. de Vette SPM, van Rijn-Dekker MI, Van den Bosch L, et al. Evaluation of a comprehensive set of normal tissue complication probability models for patients with head and neck cancer in an international cohort. Oral Oncol 2025 Apr;163:107224. [doi: 10.1016/j.oraloncology.2025.107224] [Medline: 40023984]
6. Taylor KJ, Amdal CD, Bjordal K, et al. Long-term health-related quality of life in head and neck cancer survivors: a large multinational study. Int J Cancer 2024 May 15;154(10):1772-1785. [doi: 10.1002/ijc.34861] [Medline: 38312044]
7. Andreassen R, Jönsson B, Hadler-Olsen E. Oral health related quality of life in long-term survivors of head and neck cancer compared to a general population from the seventh Tromsø study. BMC Oral Health 2022 Mar 30;22(1):100. [doi: 10.1186/s12903-022-02140-2] [Medline: 35354441]
8. van Nieuwenhuizen AJ, Buffart LM, Brug J, Leemans CR, Verdonck-de Leeuw IM. The association between health related quality of life and survival in patients with head and neck cancer: a systematic review. Oral Oncol 2015 Jan;51(1):1-11. [doi: 10.1016/j.oraloncology.2014.09.002] [Medline: 25262163]
9. Goncalves Leite Rocco P, Reategui-Rivera CM, Finkelstein J. Telemedicine applications for cancer rehabilitation: scoping review. JMIR Cancer 2024 Aug 21;10:e56969. [doi: 10.2196/56969] [Medline: 39079103]
10. da Silva HEC, Santos GNM, Ferreira Leite A, et al. The feasibility of telehealth in the monitoring of head and neck cancer patients: a systematic review on remote technology, user adherence, user satisfaction, and quality of life. Support Care Cancer 2022 Oct;30(10):8391-8404. [doi: 10.1007/s00520-022-07109-z] [Medline: 35524146]
11. Nasi G, Cucciniello M, Guerrazzi C. The performance of mHealth in cancer supportive care: a research agenda. J Med Internet Res 2015 Feb 13;17(1):e9. [doi: 10.2196/jmir.3764] [Medline: 25720295]
12. Moreno-Ligero M, Moral-Munoz JA, Salazar A, Failde I. mHealth intervention for improving pain, quality of life, and functional disability in patients with chronic pain: systematic review. JMIR Mhealth Uhealth 2023 Feb 2;11:e40844. [doi: 10.2196/40844] [Medline: 36729570]
13. Li Y, Chen W, Liang Y, Yang L, Hou L. Evaluation of mobile health technology interventions for the postdischarge management of patients with head and neck cancer: scoping review. JMIR Mhealth Uhealth 2023 Oct 23;11:e49051. [doi: 10.2196/49051] [Medline: 37870887]
14. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3(1):4. [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
15. Deniz-Garcia A, Fabelo H, Rodriguez-Almeida AJ, et al. Quality, usability, and effectiveness of mHealth apps and the role of artificial intelligence: current scenario and challenges. J Med Internet Res 2023 May 4;25:e44030. [doi: 10.2196/44030] [Medline: 37140973]
16. Ahn JH, Kim M. Effects of mobile app-based interventions on quality of life in patients with breast cancer: a systematic review and meta-analysis. Gen Hosp Psychiatry 2025;97:107-117. [doi: 10.1016/j.genhosppsych.2025.10.002] [Medline: 41075577]
17. Buneviciene I, Mekary RA, Smith TR, Onnela JP, Bunevicius A. Can mHealth interventions improve quality of life of cancer patients? A systematic review and meta-analysis. Crit Rev Oncol Hematol 2021 Jan;157:103123. [doi: 10.1016/j.critrevonc.2020.103123] [Medline: 33190065]
18. Calver L, Tickle A, Moghaddam N, Biswas S. The effect of psychological interventions on quality of life in patients with head and neck cancer: a systematic review and meta-analysis. Eur J Cancer Care (Engl) 2018 Jan;27(1):e12789. [doi: 10.1111/ecc.12789] [Medline: 29094780]
19. Nezenega ZS, Aychiluhm SB, Ross AG, et al. Systematic review and meta-analysis of digital interventions for mental health in cancer patients and survivors. NPJ Digit Med 2025 Dec 27;9(1):83. [doi: 10.1038/s41746-025-02266-x] [Medline: 41455832]
20. Qin M, Chen B, Sun S, Liu X. Effect of mobile phone app-based interventions on quality of life and psychological symptoms among adult cancer survivors: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022 Dec 19;24(12):e39799. [doi: 10.2196/39799] [Medline: 36534460]
21. Su Y, Zhang S, Duan L, Hu X. The effectiveness of telemedicine on distress, physical function and self-efficacy in patients with cancer: a meta-analysis of randomised controlled trials. J Clin Nurs 2026 Mar;35(3):1142-1151. [doi: 10.1111/jocn.70089] [Medline: 40910721]
22. Yang W, Du Y, Chen M, et al. Effectiveness of home-based telerehabilitation interventions for dysphagia in patients with head and neck cancer: systematic review. J Med Internet Res 2023 Sep 8;25:e47324. [doi: 10.2196/47324] [Medline: 37682589]

XSL•FO
RenderX

23. Zhang C, Hong Y, Chen Y, Arbing R, Chen WT, Huang F. Effectiveness of nurse-led mHealth interventions on symptom outcomes in adult patients with cancer: a systematic review and meta-analysis. BMC Nurs 2025 Oct 31;24(1):1356. [doi: 10.1186/s12912-025-03981-2] [Medline: 41174780]
24. Zhang X, Sun S, Jiangenuer L, et al. Effect of mobile health (mHealth) on improving anxiety, depression and quality of life in cancer patients: a systematic review and meta -analysis. J Affect Disord 2025 Apr 1;374:11-25. [doi: 10.1016/j.jad.2025.01.016] [Medline: 39793615]
25. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
26. Methley AM, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. BMC Health Serv Res 2014 Nov 21;14(1):579. [doi: 10.1186/s12913-014-0579-0] [Medline: 25413154]
27. Hinkle J, Wiener L, Altschuler A, et al. What constitutes quality of life? Perspectives of adolescents and young adults with advanced cancer. J Natl Compr Canc Netw 2023 Dec;21(12):1243-1250. [doi: 10.6004/jnccn.2023.7067] [Medline: 38081124]
28. Ihara Y, Kato H, Tashimo Y, et al. Changes in oral function, swallowing function, and quality of life in patients with head and neck cancer: a prospective cohort study. BMC Oral Health 2022 Jul 17;22(1):293. [doi: 10.1186/s12903-022-02329-5] [Medline: 35843950]
29. McMillan H, Barbon CEA, Cardoso R, et al. Manual therapy for patients with radiation-associated trismus after head and neck cancer. JAMA Otolaryngol Head Neck Surg 2022 May 1;148(5):418-425. [doi: 10.1001/jamaoto.2022.0082] [Medline: 35297966]
30. Ester M, Wagoner CW, Dreger J, et al. Effectiveness of a self-monitoring app in supporting physical activity maintenance among rural Canadians with cancer after an exercise oncology program: cluster randomized controlled trial. JMIR Cancer 2023 Sep 7;9(1):e47187. [doi: 10.2196/47187] [Medline: 37676714]
31. Parsons CE, Madsen MA, Jensen KL, et al. Smartphone monitoring of participants' engagement with home practice during mindfulness-based stress reduction: observational study. JMIR Ment Health 2020 Jan 14;7(1):e14467. [doi: 10.2196/14467] [Medline: 31934873]
32. Tian L, Wen Y, Li J, Guan J, Li T, Fan J. Effects of remote web-based interventions on the physiological and psychological states of patients with cancer: systematic review with meta-analysis. JMIR Mhealth Uhealth 2025 Jun 12;13:e71196. [doi: 10.2196/71196] [Medline: 40505121]
33. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282. [Medline: 23092060]
34. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928. [doi: 10.1136/bmj.d5928] [Medline: 22008217]
35. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008 Apr 26;336(7650):924-926. [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]
36. Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 2008 Mar 15;336(7644):601-605. [doi: 10.1136/bmj.39465.451748.AD] [Medline: 18316340]
37. Cohen J. Statistical Power Analysis for the Behavioral Sciences: Routledge; 2013.
38. IntHout J, Ioannidis JPA, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. BMJ Open 2016 Jul 12;6(7):e010247. [doi: 10.1136/bmjopen-2015-010247] [Medline: 27406637]
39. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634. [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
40. Baudelet M, Duprez F, Van den Steen L, et al. Increasing adherence to prophylactic swallowing exercises during head and neck radiotherapy: the multicenter, randomized controlled PRESTO-trial. Dysphagia 2023 Jun;38(3):886-895. [doi: 10.1007/s00455-022-10513-6] [Medline: 36121560]
41. Cen J, Yang L. Application effect of health management APP on patients with nasopharyngeal carcinoma. Chin Nurs Manag 2019;19(11):1695-1701 [FREE Full text]
42. Chen S, Ma H, Wang Q. Application of mobile education APP in postoperative home care of patients with laryngeal cancer surgery [in Chinese]. Nurs Res 2021;35(12):2245-2247. [doi: 10.12102/j.issn.1009-6493.2021.12.036]
43. Dai YM, Virtanen H, Zhu Y, Wan H. Effectiveness of mHealth intervention for trismus exercise in patients with head and neck cancer undergoing proton and heavy ion therapy: a randomized control trial. Support Care Cancer 2024 Jun 29;32(7):470. [doi: 10.1007/s00520-024-08679-w] [Medline: 38951291]
44. Di R, Li G. Use of a smartphone medical app improves complications and quality of life in patients with nasopharyngeal carcinoma who underwent radiotherapy and chemotherapy. Med Sci Monit 2018 Sep 4;24:6151-6156. [doi: 10.12659/MSM.908146] [Medline: 30178780]
45. Di RQ, Li GW, Zhao YL, Li XD. Effect of smart-phone application on complications of chemoradiotherapy and quality of life in patients with nasopharyngeal carcinoma [in Chinese]. J Clin Otorhinolaryngol Head Neck Surg 2017;31(3):215-218. [doi: 10.13201/j.issn.1001-1781.2017.03.012]

XSL•FO
RenderX

46. Ducharme L, Lo C, Hier M, et al. PTSD Coach as an early mobile intervention to improve cancer-related anxiety and psychosocial oncology uptake in patients newly diagnosed with head and neck cancer: pilot randomized controlled trial. Pilot Feasibility Stud 2024 Dec 21;10(1):153. [doi: 10.1186/s40814-024-01556-7] [Medline: 39709459]
47. Kim M, Hwang SY. Development and effects of a smartphone application to improve self-management in workers who underwent thyroid cancer surgery. Cancer Nurs 2025 Jan 10. [doi: 10.1097/NCC.0000000000001459] [Medline: 39792358]
48. Lee M, Kang D, Kang E, et al. Efficacy of the PRO-CTCAE mobile application for improving patient participation in symptom management during cancer treatment: a randomized controlled trial. Support Care Cancer 2023 May 6;31(6):321. [doi: 10.1007/s00520-023-07779-3] [Medline: 37148373]
49. Liao T, Qiu L, Zhu J, Li J, Zhang Y, Yang L. A mHealth-based nursing model for assessing the health outcomes of the discharged patients with nasopharyngeal carcinoma: a pilot RCT. BMC Nurs 2022 Aug 1;21(1):210. [doi: 10.1186/s12912-022-00993-0] [Medline: 35915490]
50. Liao T, Yang L, Wang R, et al. Construction and application of home-based continuous care platform for patients with nasopharyngeal carcinoma [in Chinese]. Chin J Nurs 2020;55(9):1399-1405. [doi: 10.3761/j.issn.0254-1769.2020.09.023]
51. Lyu KX, Zhao J, Wang B, et al. Smartphone application WeChat for clinical follow-up of discharged patients with head and neck tumors. Chin Med J 2016;129(23):2816-2823. [doi: 10.4103/0366-6999.194635]
52. Pang P, Lin S, Chen H, Shan Y, Sun C. Improving masticatory and swallowing ability of postoperative oral and maxillofacial tumor patients by telerehabilitation—a randomized controlled trial. Clin Rehabil 2023 Sep;37(9):1178-1188. [doi: 10.1177/02692155231166326]
53. Peng C. Application of network interactive education in nursing care of patients with oral cancer [in Chinese]. Chin Nurs Res 2019;33(3):376-380. [doi: 10.12102/j.issn.1009-6493.2019.03.035]
54. Pfeifer MP, Keeney C, Bumpous J, et al. Impact of a telehealth intervention on quality of life and symptom distress in patients with head and neck cancer. J Community Support Oncol 2015 Jan;13(1):14-21. [doi: 10.12788/jcso.0101] [Medline: 25839061]
55. Rades D, Zwaan I, Cacicedo J, et al. Impact of a mobile application (reminder app) on acute toxicity during radiotherapy of head-and-neck cancer – results of a randomized phase III trial (RAREST-02). BMC Cancer 2022;22:989. [doi: 10.1186/s12885-022-10088-3]
56. Sprave T, Pfaffenlehner M, Stoian R, et al. App-controlled treatment monitoring and support for patients with head and neck cancer undergoing radiotherapy: results from a prospective randomized controlled trial. J Med Internet Res 2023 Oct 19;25:e46189. [doi: 10.2196/46189] [Medline: 37856185]
57. Starmer HM, Klein D, Montgomery A, et al. Head and neck virtual coach: a randomized control trial of mobile health as an adjunct to swallowing therapy during head and neck radiation. Dysphagia 2023 Jun;38(3):847-855. [doi: 10.1007/s00455-022-10506-5] [Medline: 35960394]
58. Sun X, Jiang Y, Wang J, et al. Effects of a mobile health intervention based on a multitheoretical model of health behavior change on anxiety and depression, fear of cancer progression, and quality of life in patients with differentiated thyroid cancer: a randomized controlled trial. BMC Med 2024 Oct 15;22(1):466. [doi: 10.1186/s12916-024-03652-0] [Medline: 39407174]
59. van den Brink JL, Moorman PW, de Boer MF, et al. Impact on quality of life of a telemedicine system supporting head and neck cancer patients: a controlled trial during the postoperative period at home. J Am Med Inform Assoc 2007 Mar 1;14(2):198-205. [doi: 10.1197/jamia.M2199]
60. van der Hout A, van Uden-Kraan CF, Holtmaat K, et al. Role of eHealth application Oncokompas in supporting self-management of symptoms and health-related quality of life in cancer survivors: a randomised, controlled trial. Lancet Oncol 2020 Jan;21(1):80-94. [doi: 10.1016/S1470-2045(19)30675-8] [Medline: 31838009]
61. Wall LR, Ward EC, Cartmill B, et al. Prophylactic swallowing therapy for patients with head and neck cancer: a three‐arm randomized parallel‐group trial. Head Neck 2020 May;42(5):873-885. [doi: 10.1002/hed.26060] [Medline: 31903689]
62. Wang TJ, Su JH, Leung KW, Liang SY, Wu SF, Wang HM. Effects of a mouth-opening intervention with remote support on adherence, the maximum interincisal opening, and mandibular function of postoperative oral cancer patients: a randomized clinical trial. Eur J Oncol Nurs 2019 Jun;40:111-119. [doi: 10.1016/j.ejon.2019.04.001] [Medline: 31229200]
63. Wang Y, Xiong L, Wu C, Li S, Zhang J, Chen H, et al. Application study of “internet+” extended care based on IMB model in swallowing rehabilitation training of postoperative patients with laryngeal cancer [in Chinese]. Lab Med Clin 2025;22(16):2271-2275. [doi: 10.3969/j.issn.1672-9455.2025.16.021]
64. Wu T, Han T, Feng Z, et al. Effectiveness and determinants of a digital health application-delivered intervention to promote rehabilitation exercises in non-adherent head and neck cancer patients: a clinical trial. Eur J Oncol Nurs 2025 Dec;79:103000. [doi: 10.1016/j.ejon.2025.103000] [Medline: 41086689]
65. Xu X, Sun N, Nie P. Effect of continued nursing based on APP platform on treatment compliance and quality of life in patients discharged after surgery for laryngeal cancer [in Chinese]. Chin J Clin Oncol Rehabil 2019;26(9):1140-1143. [doi: 10.13455/j.cnki.cjcor.2019.09.32]
66. Henry M, Chen LM, Ducharme L, et al. Genetic risk for depression and quality of life in patients with head and neck cancer. JAMA Otolaryngol Head Neck Surg 2024 Jul 1;150(7):598-606. [doi: 10.1001/jamaoto.2024.0376] [Medline: 38814668]

XSL•FO
RenderX

67. Hernandez Silva E, Lawler S, Langbecker D. The effectiveness of mHealth for self-management in improving pain, psychological distress, fatigue, and sleep in cancer survivors: a systematic review. J Cancer Surviv 2019 Feb;13(1):97-107. [doi: 10.1007/s11764-018-0730-8] [Medline: 30635865]
68. Parsons A, Dewan K. Dysphagia and dysphonia after head and neck cancer. Oral Dis 2025 Sep;31(9):2753-2760. [doi: 10.1111/odi.15152] [Medline: 39370743]

Abbreviations

EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30
EORTC QLQ-H&N35: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Head and Neck 35
GRADE: Grading of Recommendations, Assessment, Development, and Evaluation
HNC: head and neck cancer
mHealth: mobile health
MIO: maximal interincisal opening
PI: prediction interval
PICOS: participants, intervention, comparison, outcome, and study design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QoL: quality of life
RCT: randomized controlled trial
RoB 2.0: Cochrane Risk of Bias 2.0
SMD: standardized mean difference

Edited by R Chandrasekaran; submitted 27.May.2025; peer-reviewed by H D'Angelo, L Hou, X Bi, Z Zandesh; accepted 20.Feb.2026; published 30.Apr.2026.

Please cite as:
Xiao Y, Zhou Q, Hong F, Lin Y, Li E, Yu X
Evaluating Effectiveness of mHealth Interventions for Rehabilitation in Patients With Head and Neck Cancer: Systematic Review and Meta-Analysis of Randomized Controlled Trials
JMIR Mhealth Uhealth 2026;14:e78109
URL: https://mhealth.jmir.org/2026/1/e78109
doi:10.2196/78109

© Yang Xiao, Qiuyu Zhou, Feiruo Hong, Yuwei Lin, Enhong Li, Xuefen Yu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 30.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Wearable Devices for Remote Monitoring of Chronic Diseases: Systematic Review

Masresha Derese Tegegne[1], MPH; Sharareh Rostam Niakan Kalhori[1,2], PhD; Paulo Haas[1], MSc; Viktor MG Sobotta[1], MSc; Joana Warnecke[1], PhD; Thomas M Deserno[1], PhD

[1]Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Mühlenpfordtstraße 23, Braunschweig, Germany

[2]Department of Health Information Management and Medical Informatics, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran

Corresponding Author:
Masresha Derese Tegegne, MPH
Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Mühlenpfordtstraße 23, Braunschweig, Germany

Abstract

Background: Wearable devices enable the remote collection of health parameters, supporting the outpatient care plans recommended by the World Health Organization to manage chronic diseases. While disease-specific monitoring is accurate, a comprehensive analysis of wearables across various chronic diseases helps to standardize remote patient monitoring systems.

Objective: This review aimed to identify wearables for remote monitoring of chronic diseases, focusing on (1) wearable devices, (2) sensor types, (3) health parameters, (4) body locations, and (5) medical applications.

Methods: We developed a search strategy and conducted searches across three databases: PubMed, Web of Science, and Scopus. After reviewing 1160 articles, we selected 61 that addressed cardiovascular, cancer, neurological, metabolic, respiratory, and other diseases. We created a data analysis method based on our 5 objectives to organize the articles for a comprehensive analysis.

Results: From the 61 articles, 39 (64%) used wearable bands such as smartwatches, wristbands, armbands, and straps to monitor chronic diseases. Wearable devices commonly included various sensor types, such as accelerometers (n=39, 64%), photoplethysmographic sensors (n=18, 30%), biopotential meters (n=17, 28%), pressure meters (n=11, 18%), and thermometers (n=9, 15%). These sensors collected diverse health parameters, including acceleration (n=39, 64%), heart rate (n=24, 39%), body temperature (n=9, 15%), blood pressure (n=8, 13%), and peripheral oxygen saturation (n=7, 11%). Common sensor body locations were the wrist, followed by the upper arm and the chest. The medical applications of wearable devices were neurological (n=21, 34%) and cardiovascular diseases (n=15, 25%). Additionally, researchers applied wearable devices for wellness and lifestyle monitoring (n=39, 64%), mainly for activity (n=39, 100%) and sleep (n=10, 26%).

Conclusions: This review underscores that wearable devices primarily function as bands, commonly worn on the wrist, to monitor chronic diseases. These devices collect data on acceleration, heart rate, body temperature, blood pressure, and peripheral oxygen saturation, with a focus on neurological and cardiovascular diseases. Our findings provide a foundational road map for designing generalized remote patient monitoring systems to manage multimorbidity and support standardized terminology for interoperability across digital health systems. To translate this into practice, we recommend that future research prioritize pragmatic clinical trials with medically certified devices.

Trial Registration: PROSPERO CRD42023460873; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023460873

(JMIR Mhealth Uhealth 2026;14:e74071) doi:10.2196/74071

KEYWORDS

wearable; sensor; health parameter; vital sign; body location; medical application; standardized terminology

Introduction

Background

According to the World Health Organization (WHO), chronic diseases are long-lasting, noninfectious, and progressively worse over time [1]. These include cardiovascular diseases, neurological disorders, cancer, respiratory diseases, and metabolic disorders, which collectively account for 74% of annual global mortality and make them the leading cause of death worldwide [1,2]. The burden of chronic diseases extends beyond health, severely straining health care resources. The Centers for Disease Control and Prevention reports that 90% of total health care expenditures in the United States address individuals with chronic diseases [3]. This expenditure

XSL•FO
RenderX

encompasses ongoing treatments, regular medical consultations, and other health care–related costs [4].

Owing to their long-lasting nature and the ongoing commitment to curative health services, health care professionals mostly manage chronic diseases on an outpatient basis [5]. However, preventable factors related to these chronic diseases lead to sudden death [6]. The WHO prioritizes individuals with chronic diseases and underscores the need to develop efficient outpatient treatment strategies [7]. Health care providers achieve this by remotely monitoring health parameters [8,9]. Compared to inpatient health care delivery, remote patient monitoring (RPM) systems have become a promising option for managing chronic diseases [10,11]. RPM systems support early detection by real-time monitoring of health and well-being [12] using wearable devices that integrate 1 or more sensors to collect relevant health parameters remotely [13-15]. This reduces clinic visits and conserves time and health care resources [16].

Wearable-based RPM systems rely on custom-built sensors to gather and analyze biometric and physiological data [17]. Contemporary devices such as smartwatches and fitness trackers have several sensors [18] and often pair with mobile apps. Integrating these wearable devices with cloud platforms facilitates efficient data storage and easy access for health care providers. Examples of cloud platforms include Apple HealthKit [19], Google Fit [20], Microsoft Azure Health Data Services [21], Amazon AWS HealthLake [22], and Biofourmis [23]. Such technologies offer substantial potential for clinical trials, fostering research advancements and improving patient care.

Current articles on wearable devices highlight their potential in managing chronic diseases. Various sensors, such as accelerometers, gyroscopes, magnetometers, biopotential meters, photoplethysmographic (PPG) sensors, and thermometers, monitor health parameters remotely. These include activity, electrocardiography (ECG), electroencephalography, electromyography, heartbeats, heart rate, sleep patterns, and body temperature [24-31].

Currently, wearable devices play a key role in monitoring specific chronic diseases, enhancing the effectiveness of RPM by focusing on disease-specific health parameters. For instance, continuous glucose monitors track glucose levels in diabetes management [32], biopotential meters monitor heart rhythms in various conditions [33], and inertial measurement units (IMU) evaluate body movements for diagnosing stroke and neurodegenerative diseases [34,35].

Previous review articles on wearable devices for remote monitoring of chronic diseases often focus on a single condition, a specific population, or a limited clinical setting [36,37]. This narrow focus creates a knowledge gap in understanding their applicability across a broader range of chronic diseases. The increasing variability of wearable devices and stand-alone RPM systems presents substantial challenges for interoperability with digital health systems [38,39], including electronic health record systems [40]. These challenges become even more pressing in the context of rising rates of multimorbidity [41], where patients require integrated monitoring solutions that can simultaneously track multiple conditions. This systematic review addresses these gaps by applying a five-category standardized terminology framework to analyze wearable devices used for remote monitoring of chronic diseases.

Objectives

This systematic review aimed to explore wearable devices and sensor types for remote monitoring of chronic diseases. Specifically, we answered the following research questions:

1. Which wearable devices are used to monitor chronic diseases?
2. Which sensor types are integrated into these wearable devices?
3. Which health parameters do these sensors collect?
4. Where are these sensors located on the human body?
5. Which medical applications are supported by the wearable devices?

Methods

Overview

This systematic review's design and reporting follow the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Checklist 1) [42]. We registered the protocol in PROSPERO (CRD42023460873). Although we specified Quality Assessment Tool for Diverse Designs for quality appraisal in the registered protocol, we used the Quality Assessment with Diverse Studies (QuADS) critical appraisal tool instead, as it was better aligned with the design and reporting characteristics of the included articles.

Information Source and Search Strategy

We developed a search strategy and sought full-text articles published in English between January 2019 and December 2023 across three databases: PubMed, Web of Science, and Scopus. Our search strategy incorporated three terms: wearable device, remote monitoring, and chronic disease, as well as relevant synonyms. We used the AND operator between the keywords and the OR operator within the synonyms (Multimedia Appendix 1).

Eligibility Criteria

To ensure a concentrated analysis of wearable devices for remote monitoring of chronic diseases, we established the following eligibility criteria.

- Inclusion criteria:
  - Articles that used wearable devices for the remote monitoring of chronic diseases
  - Articles that used sensors for the remote collection of health parameters
  - Original research articles published in peer-reviewed journals
- Exclusion criteria:
  - Articles that used unobtrusive, implantable, or nonwearable devices or sensor types
  - Articles that did not involve remote monitoring or data collection
  - Nonjournal publications
  - Review articles

XSL•FO
RenderX

Data Management and Extraction

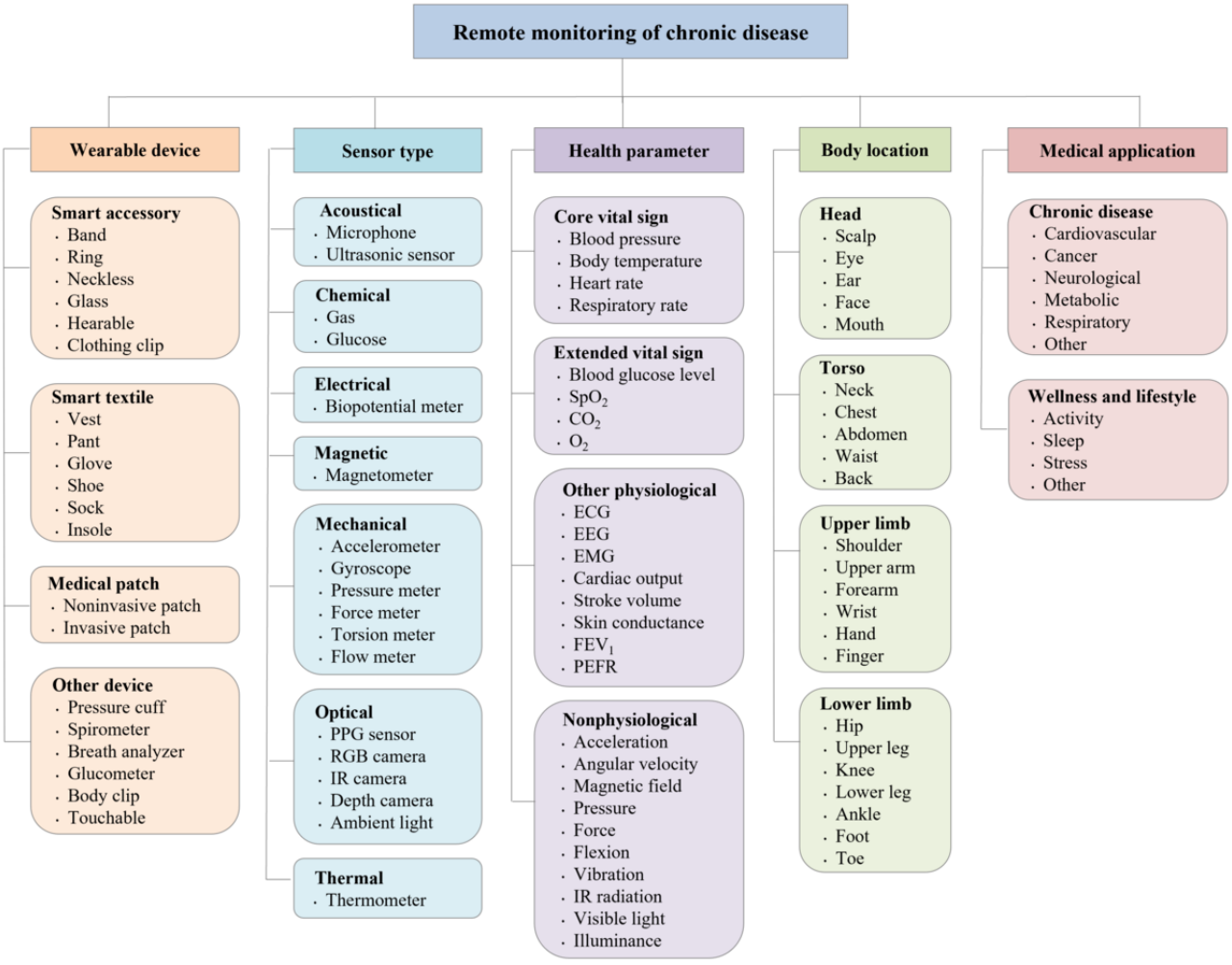
We conducted data extraction through regular team communication while reviewing the articles' titles, abstracts, and full texts. We engaged in thorough interranker discussions to resolve discrepant findings. For data extraction, we used a well-organized spreadsheet (Microsoft Excel) with the following parameters: author names, year of publication, country, study design, study phase, device name, device status, device type, sensor types, health parameters, body locations, and medical applications (Multimedia Appendix 2).

Standardized Terminology

Overview

We developed a standardized terminology based on existing evidence supplemented by insights from our research team. Our terminology encompasses five predefined categories: wearable device, sensor type, health parameter, body location, and medical application (Figure 1). We systematically organized the articles by these categories to facilitate a thorough analysis of the available evidence. The following sections detail the rationale and evidence.

Figure 1. Our standardized terminology framework outlines 5 categories used to guide the systematic analysis of articles on remote monitoring of chronic diseases. CO_2: carbon dioxide; ECG: electrocardiography; EEG: electroencephalography; EMG: electromyography; FEV_1: forced expiratory volume in the first second; IR: infrared; PEFR: peak expiratory flow rate; PPG: photoplethysmographic; O_2: oxygen; RGB: red, green, blue; SpO_2: peripheral oxygen saturation.



Wearable Device

We defined a wearable device as an Internet of Things–enabled electronic device that users wear externally on the body to monitor and collect health parameters, either continuously or sporadically, and to transmit data remotely for health monitoring purposes [43]. Our definition encompasses noninvasive and minimally invasive devices that users wear externally and that do not require surgical implantation [44]. We explicitly excluded fully implantable devices, environmental sensors, and purely handheld devices not designed for body attachment [45]. We categorized wearable devices into four types:

- Smart accessory: a stand-alone electronic device that users wear continuously on the body, including bands (smartwatches, wristbands, armbands, and straps), rings, necklaces, glasses, hearables, and clothing clips
- Smart textile: a device that integrates into everyday clothing items such as vests, pants, gloves, shoes, socks, and insoles
- Medical patch: a device with adhesive skin contact or a minimally invasive needle
- Other device: a device that users wear sporadically for specific measurements, such as pressure cuffs, spirometers,

XSL•FO
RenderX

breath analyzers, glucometers, body clips, and touchable devices, which can remotely transmit data

Sensor Type

Sensor types refer to the categories of sensing technologies that wearable devices use for remote monitoring of chronic diseases, grouped by the physical or chemical properties they measure (ie, acoustical, chemical, electrical, magnetic, mechanical, optical, and thermal) [46].

Health Parameter

Health parameters are measurable indicators that provide insights into an individual's health status, including core vital signs, extended vital signs, other physiological data, and nonphysiological data [14].

Body Location

Body location specifies the outer body regions for sensor placement to collect health parameters. On the basis of the classification proposed by Kim et al [47], we grouped these regions into four categories: head, torso, upper limb, and lower limb.

Medical Application

Medical applications refer to the specific uses of wearable devices and sensors designed to monitor, manage, or support health conditions. The primary application of wearable devices focuses on chronic disease management while optionally offering wellness and lifestyle tracking, which can supplement disease monitoring. After identifying more than 10 distinct diseases, we categorized them into five groups based on their physiological systems: cardiovascular (eg, heart diseases and hypertension), cancer, neurological (eg, stroke, epilepsy, neurodegenerative disorders, and peripheral neuropathy), metabolic (eg, diabetes and obesity), respiratory (eg, asthma, chronic obstructive pulmonary disease), and other diseases [48,49]. This allows us to focus on broader applications of wearable devices across various diseases rather than focusing solely on specific conditions.

Quality Appraisal

Our review encompassed various study designs, including user-centered design, observational, experimental, and mixed methods articles. To assess the quality of these diverse articles, we used the QuADS critical appraisal tool [50]. The QuADS includes 13 criteria rated on a scale ranging from 0 to 3 (0=not at all, 1=very slightly, 2=moderately, and 3=complete), with a total quality score ranging from 0 to 39 (Multimedia Appendix 3) [51].

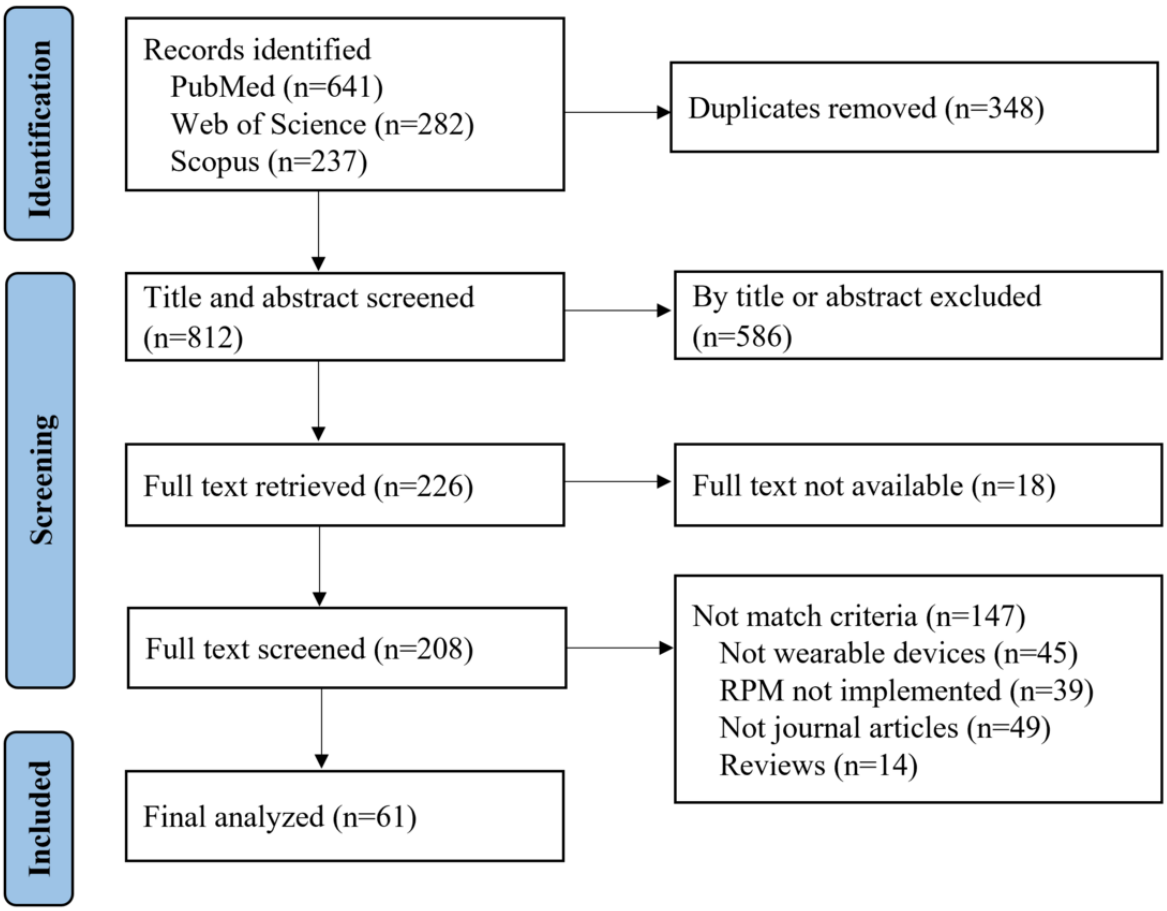
Results

Identified Articles

We identified 1160 articles from the 3 electronic databases. After removing duplicates, we considered 812 articles for title and abstract screening and excluded 586. Of the 226 articles screened, 18 were unavailable in full text. We thoroughly assessed 208 full-text articles for eligibility and excluded 147 that did not meet the inclusion and exclusion criteria. Finally, we included 61 articles (Figure 2).

XSL•FO
RenderX

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart describing the article selection process for this review. RPM: remote patient monitoring.



Characteristics of Included Articles

The included articles span Europe, North America, and the Asia-Pacific region. A total of 29 (48%) articles were conducted in Europe, with Italy contributing the most [25,31,52-55], followed by the United Kingdom [54,56-59]. North America ranks second, with nearly all articles conducted in the United States [26-28,32,34,60-73], except one article conducted in Canada [74].

Researchers used various research methodologies and study designs. Observational designs were the most prevalent (29/61, 48%), followed by 13 (21%) articles using user-centered design paired with experimental validation. Nine (15%) articles relied solely on experimental methods, and 7 (11%) articles focused exclusively on user-centered design. Only 3 (5%) articles adopted a mixed methods approach. In line with these designs, most articles (52/61, 85%) aimed to develop RPM systems and conduct feasibility and usability testing. In contrast, only 9 (15%) articles focused on late-stage validation and clinical efficacy evaluation.

We identified 74 instances of wearable devices used across the included articles. Of these, 28 (38%) instances involved dedicated medical devices approved by regulatory bodies, such as the US Food and Drug Administration (FDA) or the European Conformity Medical Device Regulation (CE-MDR). Most of these (24/28, 86%) collected vital signs and other physiological parameters. Twenty-six (35%) instances were consumer-grade wellness devices lacking medical certification, nearly all (23/26, 88%) designed as bands. Fourteen (19%) instances represented the early-stage development of research prototype devices. Additionally, 6 (8%) used consumer-grade devices with certified medical functions (FDA/CE-MDR/HIPAA [Health Insurance Portability and Accountability Act] approved) [29,34,65,70,75,76] (Table 1).

XSL•FO
RenderX

Table . Summary of the characteristics of the included articles (N=61).

Characteristics and category	Articles, n (%)[a]	Reference
Location		
Europe	29 (48)	[24,25,29,31,33,35,52-59,75-89]
North America	20 (33)	[26-28,32,34,60-74]
Asia-Pacific	12 (20)	[90-101]
Study design[b]		
User-centered design	7 (11)	[27,55,71,77,85,96,97]
Observational	29 (48)	[24,28,29,32,33,35,53,57,58,61-64,67,69,70,72-74,76,79,81-83,86,90,94,95,99]
Experimental	9 (15)	[34,54,66,68,75,78,89,93,100]
User-centered design with experimental validation	13 (21)	[25,31,52,56,59,60,65,84,87,88,91,92,98]
Mixed methods	3 (5)	[26,80,101]
Study phase[c]		
Technology development and piloting	21 (34)	[26,27,31,52,55,56,59-61,65,71,77,84,85,87,88,91,92,96-98]
Feasibility and usability testing	31 (51)	[24,25,28,32,33,35,54,57,58,63,64,66-70,72-74,76,78-82,90,93,95,99-101]
Clinical validation	5 (8)	[29,62,75,86,94]
Clinical outcome evaluation	4 (7)	[34,53,83,89]
Device status[d]		
Dedicated medical device (FDA[e]/CE-MDR[f] approved)	28 (38)	[28,32,53,55,57,58,60-62,64,66,68,71,73,79-83,86,88,93,100]
Consumer-grade device with certified medical function (FDA/CE-MDR/HIPAA[g] approved)	6 (8)	[29,34,65,70,75,76]
Consumer-grade wellness device	26 (35)	[24,26,29,35,54,57,61,63-65,67,69,72,78,80,84,87,89-91,95,96,99]
Research prototype device	14 (19)	[24,25,27,52,54,56,58,59,77,80,85,92,98,101]

[a]Percentages may not total 100% due to rounding.

[b]The study design describes the methodological approach used in each article.

[c]The study phase indicates the stage of development or evaluation addressed.

[d]The regulatory status reflects the device's classification at the time of the study, not its current status. Some articles used multiple devices; therefore, device status is reported based on 74 wearable device instances.

[e]FDA: US Food and Drug Administration.

[f]CE-MDR: the European Conformity Medical Device Regulation.

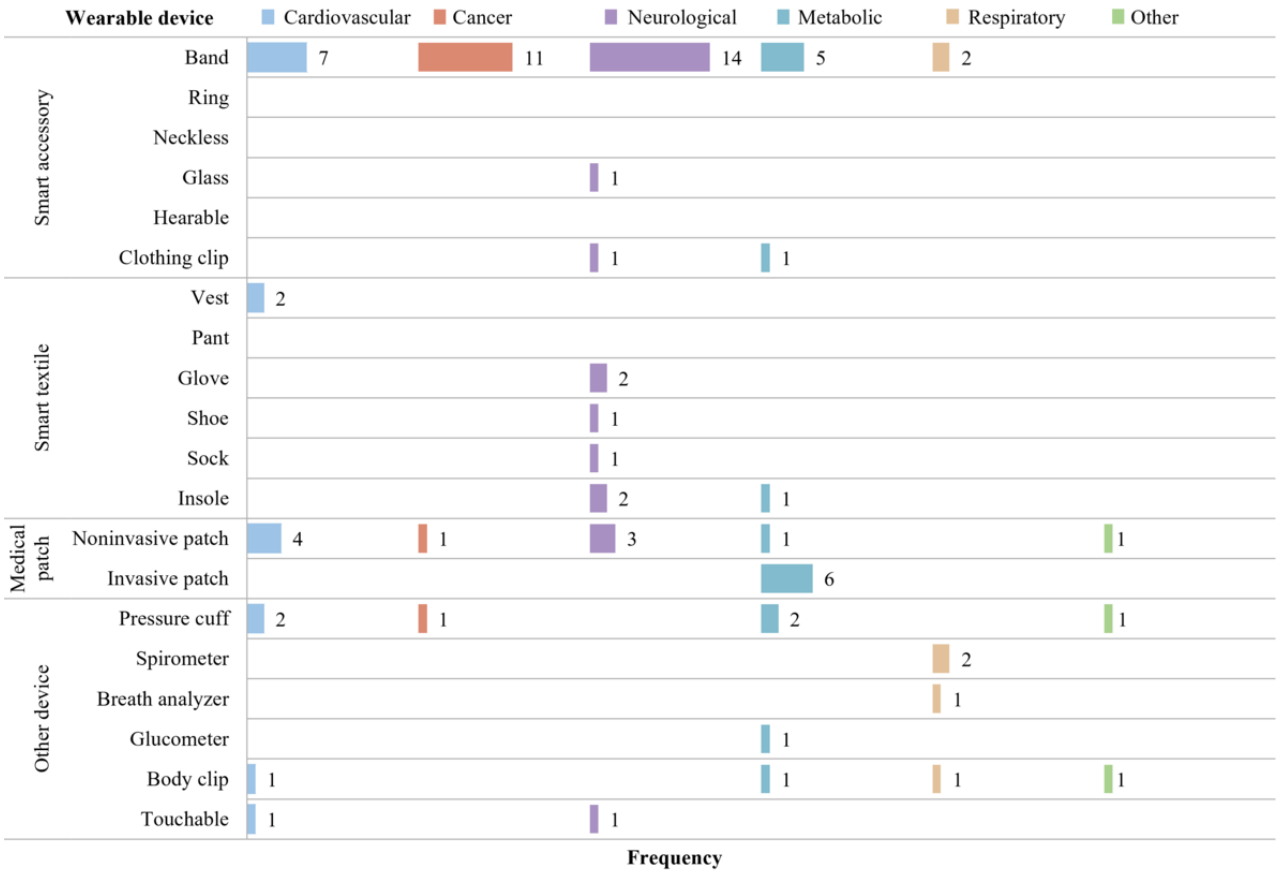
[g]HIPAA: Health Insurance Portability and Accountability Act.

Wearable Device

We identified 16 distinct wearable devices for remote monitoring of chronic diseases (Figure 3). Bands, noninvasive patches, and pressure cuffs emerged as the most frequently used devices across various chronic diseases, appearing in 39 (64%), 10 (16%), and 6 (10%) articles, respectively. Neurological disease monitoring used 9 devices, with bands being the most common [24,25,31,35,54,56,65,72,73,87,89,91,98,101]. Researchers used 8 distinct wearable devices for metabolic diseases, with invasive patches [32,53,55,60,80,83] being exclusive to this category and the most frequently used, followed by bands [29,55,80,95,96]. Cardiovascular monitoring involved 6 wearable devices, primarily bands [67,69,75,76,84,93,94] and noninvasive patches [33,62,71,85]. For cancer monitoring, the articles almost exclusively reported on bands [34,58,63,64,78,81,82,90,97,99,100].

XSL•FO
RenderX

Figure 3. Occurrence of wearable devices by categories summarized in Figure 1. Rows represent the number of articles (N=61) that use each wearable device; colored columns (disease categories) may exceed N because some articles report on multiple devices.



Sensor Type

The wearable devices feature 14 sensor types. Accelerometers, PPG sensors, biopotential meters, pressure meters, and thermometers are most frequently used to monitor chronic conditions, appearing in 39 (64%), 18 (30%), 17 (28%), 11 (18%), and 9 (15%) articles, respectively. The wearable devices typically monitor neurological diseases using a diverse array of 10 sensors, preferring accelerometers [24,25,27,31,35,54,56,65,68,72,73,87,89,91,98,101], gyroscopes [24,25,27,31,35,54,56,65,68,73,91,98,101], and magnetometers [24,25,31,56,65,73,98] being the most common. Researchers used force meters [54,56], torsion meters [27,98], and wearable cameras [70], exclusively for neurological diseases.

We identified sensors such as biopotential meters [33,62,71,75,77,79,85,86], accelerometers [62,67,71,79,84,86,93], PPG sensors [67,69,74,84,93], pressure meters [74,76,77,94], and thermometers [62,71,93] as key sensors for the remote monitoring of cardiovascular diseases, ranked by their usage frequency. Similarly, we identified that these sensors monitor cancer, although their frequency varies. Accelerometers [34,58,63,64,81,82,90,99,100], PPG sensors [78,81,82,97,99], biopotential meters [64,81,82], thermometers [81,82,97], and pressure meters [64] are the most common. Additionally, our analysis showed that the articles used glucose sensors [32,53,55,60,80,83,96] exclusively and frequently for monitoring metabolic diseases (Figure 4).

XSL•FO
RenderX

Figure 4. Occurrence of sensor types by categories summarized in Figure 1. Rows represent the number of articles (N=61) that use each sensor; colored columns (disease categories) may exceed N because some articles report on multiple sensors. IR: infrared; PPG: photoplethysmographic; RGB: red, green, blue.

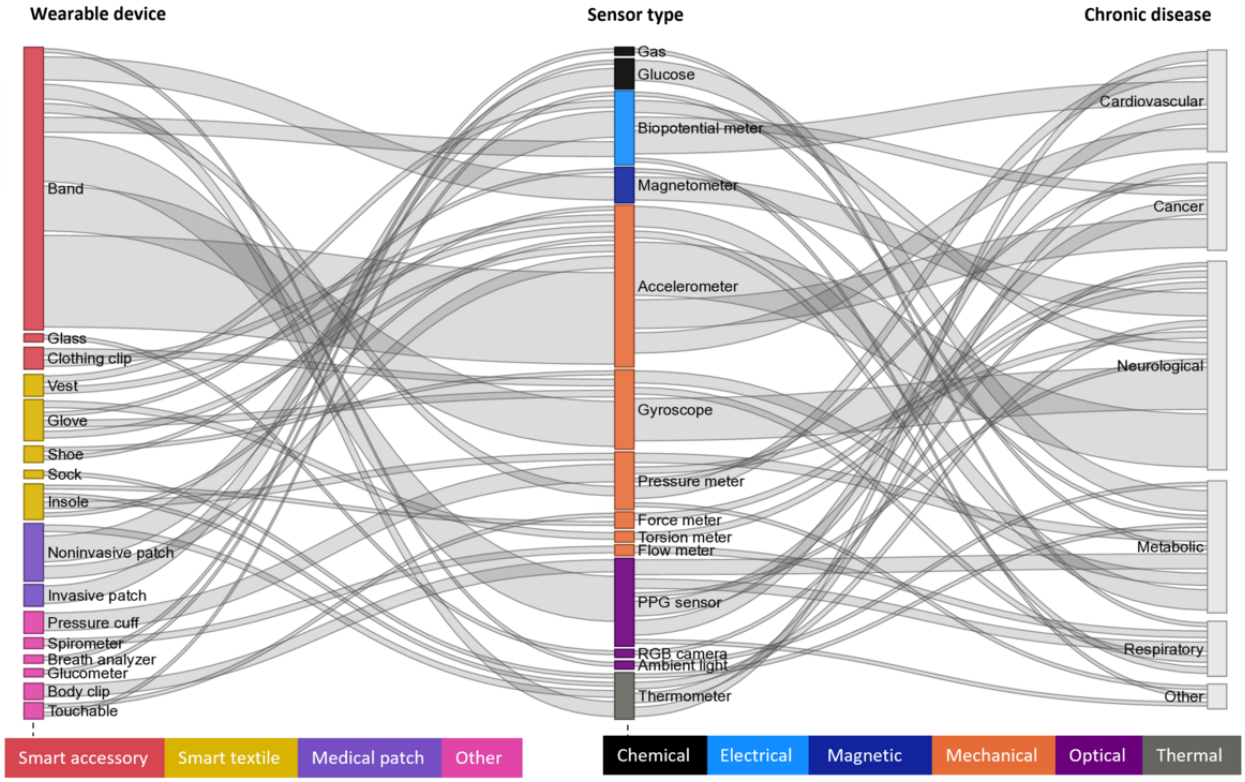


Wearable Device, Sensor Type, and Chronic Disease Mapping

A Sankey diagram visually illustrates the relationships among wearable devices, sensor types, and the corresponding chronic diseases they monitor (Figure 5). Bands have emerged as the most widely used wearable devices, integrating 8 sensor types. They primarily incorporate accelerometers [24,25,31,35,54,56,65,72,73,87,89,91,98,101], gyroscopes [24,25,31,35,54,56,65,73,91,98,101], and magnetometers [24,25,31,56,65,73,98] to facilitate remote monitoring of neurological diseases. They also often feature PPG sensors [29,54,55,57,61,67,69,78,80-82,84,93,96,97,99] and biopotential meters [64,75] to track heart rate and ECG. Moreover, bands are increasingly equipped with other sensor types such as electrodermal activity (EDA) biopotential meters [54,81,82], skin thermometers [54,81,82,93,97], oscillometric blood pressure meters [76,94], barometric altimeters [80], and ambient light sensors [80], enabling remote monitoring of skin conductance, body temperature, blood pressure, atmospheric pressure, and illumination, respectively.

XSL•FO
RenderX

Figure 5. A Sankey diagram illustrating the mapping between wearable devices, sensor types, and the chronic diseases targeted for remote monitoring. Node and link widths reflect match frequency, and node colors correspond to wearable device and sensor types. PPG: photoplethysmographic; RGB: red, green, blue.



Noninvasive medical patches incorporate accelerometers [58,62,71,80] and thermometers [62,71], in addition to common types of biopotential meters such as ECG [33,62,71,80,88], electroencephalography [66,92], electromyography [26], and impedance [85]. Notably, glucose sensors are the only sensor type integrated into minimally invasive devices [32,53,55,60,80,83]. Furthermore, traditional pressure cuffs, such as the blood pressure monitors iHealth (iHealth Labs Inc) [88,96] and Withings (Withings) [29], now transmit their data remotely. Additionally, pulse oximeters such as iHealth [96] and Onyx (Nonin Medical Inc) [88] enhance the functionality of traditional body clips by enabling the remote transmission of peripheral oxygen saturation (SpO_2) and heart rate data.

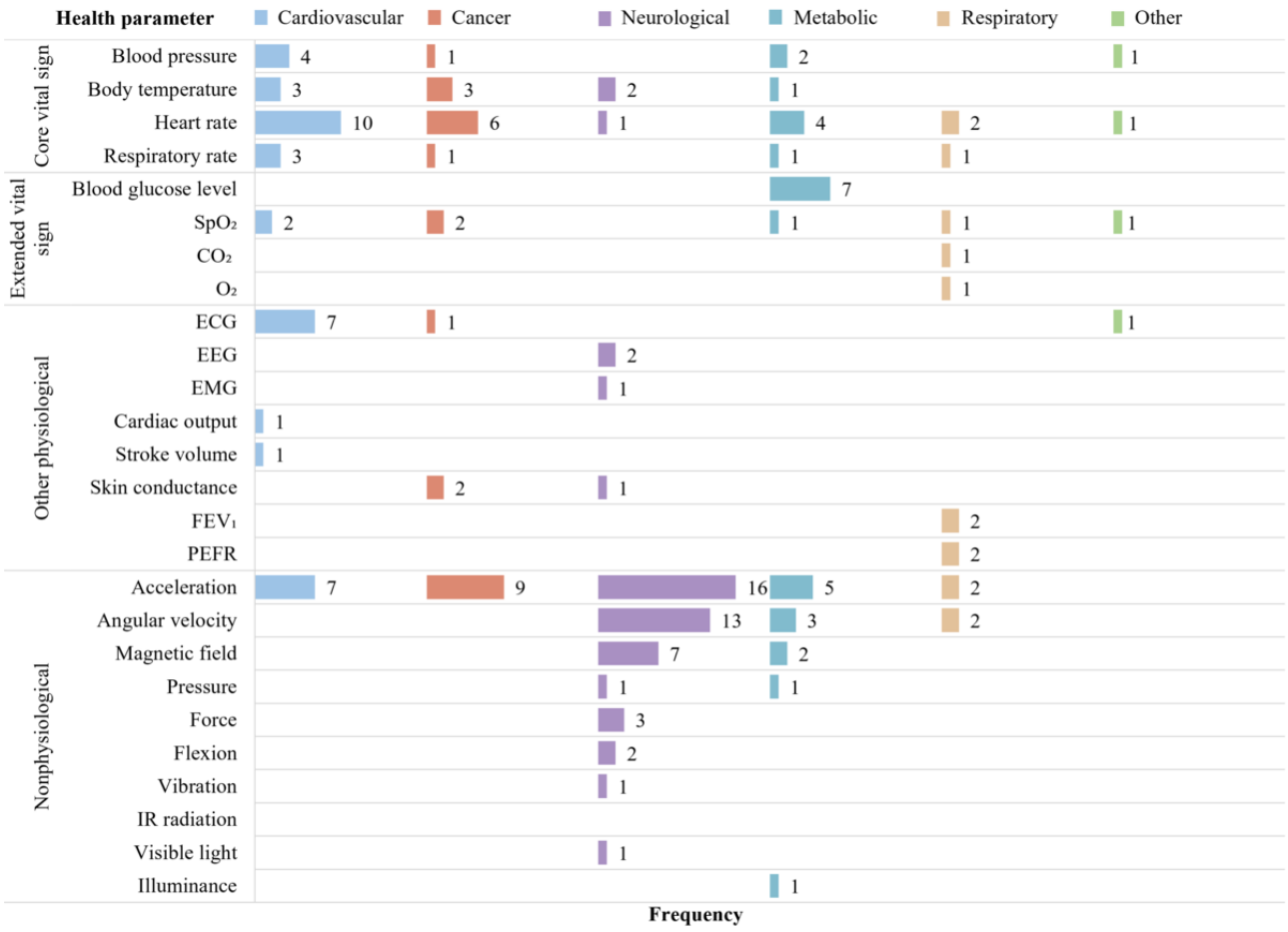
Health Parameter

Sensors gather a total of 25 health parameters for remote monitoring of chronic diseases (Figure 6). Acceleration is a key health parameter for monitoring various chronic diseases, as indicated by 39 (64%) articles. Specifically, it is the most commonly collected health parameter in the remote monitoring of neurological diseases [24,25,27,31,35,54,56,65,68,72,73,87,89,91,98,101] and cancer [34,58,63,64,81,82,90,99,100].

The articles consistently reported on monitoring heart rate, body temperature, blood pressure, and SpO_2, as reported in 24 (39%), 9 (15%), 8 (13%), and 7 (11%) articles, respectively. These core and extended vital signs are essential health parameters for remote monitoring of cardiovascular, cancer, and metabolic diseases. On the other hand, articles primarily monitor neurological diseases using nonphysiological parameters. Furthermore, metabolic monitoring focuses on core and extended vital signs and nonphysiological parameters (Figure 6).

XSL•FO
RenderX

Figure 6. Occurrence of health parameters by categories summarized in Figure 1. Rows represent the number of articles (N=61) that collect each health parameter; colored columns (disease categories) may exceed N because some articles collect multiple health parameters. CO_2: carbon dioxide; ECG: electrocardiography; EEG: electroencephalography; EMG: electromyography; FEV_1: forced expiratory volume in the first second; IR: infrared; PEFR: peak expiratory flow rate; PPG: photoplethysmographic; O_2: oxygen; RGB: red, green, blue; SpO_2: peripheral oxygen saturation.

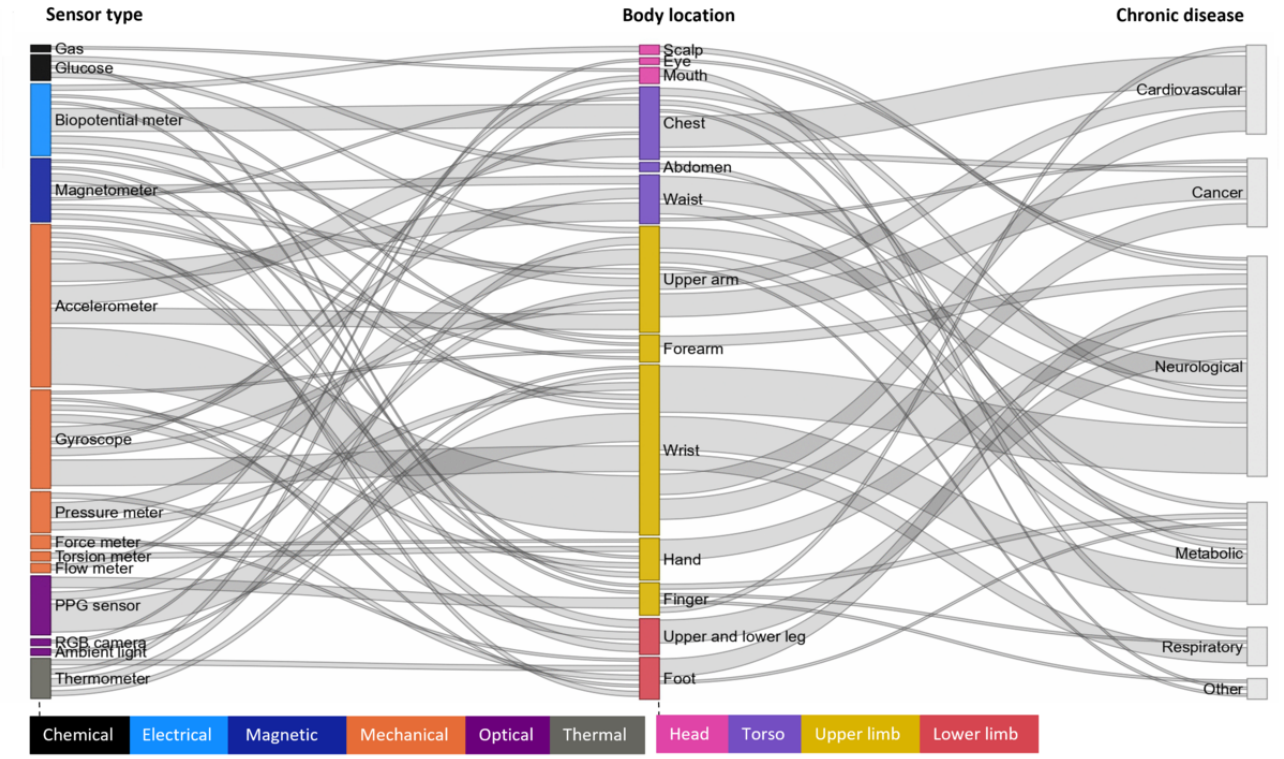


Body Location

The wrist is the most frequent body location for sensor placement, comprising 8 different sensor types. A Sankey diagram illustrates the linkage between accelerometers, PPG sensors, and gyroscopes to the wrist, highlighting that these sensor types are mostly positioned on the wrist (Figure 7). The upper arm is the second-most frequent body location, comprising 8 sensor types, including pressure meters [29,64,74,77,88,96] and accelerometers [35,56,81,82,93,98]. The chest is the third-most common body location, accommodating 5 sensor types. Researchers commonly place biopotential meters [33,62,71,79,80,88] here to monitor cardiovascular diseases. Additionally, the waist is an ideal location for IMU sensors, including accelerometers [64,65,68,72,73,80,95], gyroscopes [65,68,73,80], and magnetometers [65,73,95] for monitoring neurological, metabolic, and cancer diseases. Furthermore, researchers place various sensors on the hand, forearm, upper, and lower legs to monitor neurological diseases [26,27,31,56,73,98].

XSL•FO
RenderX

Figure 7. A Sankey diagram illustrating body locations for sensor placement in the remote monitoring of various chronic diseases. Node and link widths reflect match frequency, and node colors correspond to sensor type and body locations. PPG: photoplethysmographic; RGB: red, green, blue.

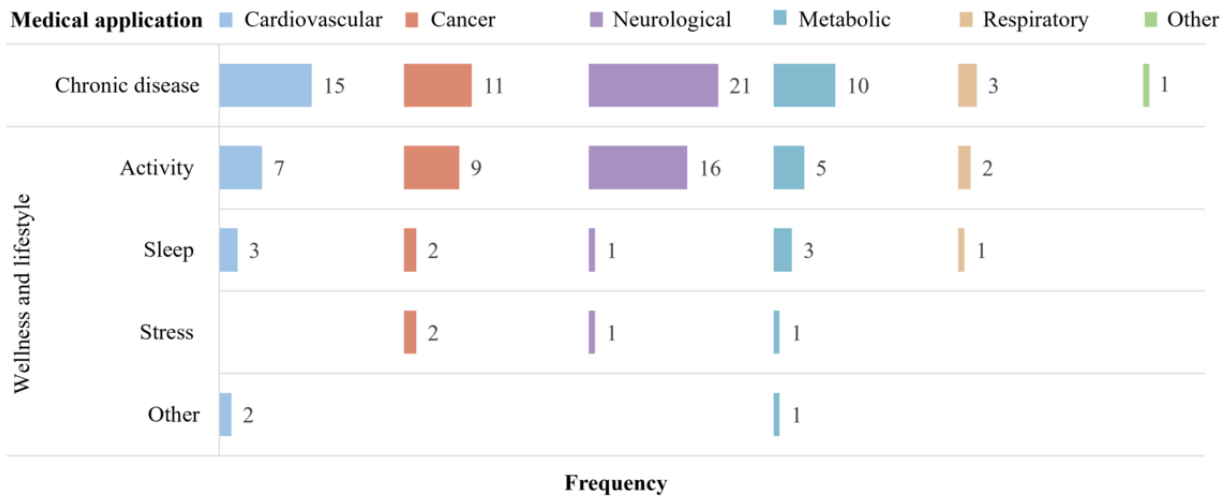


Medical Application

The articles primarily highlighted the use of wearable devices for monitoring specific chronic diseases as their primary medical application, with the option to include wellness and lifestyle monitoring. Among chronic diseases, neurological diseases are the primary focus, accounting for more than one-third (21/61, 34%) of the articles. Cardiovascular diseases follow, with 15 (25%) articles, while wearable devices monitored cancer, metabolic diseases, and respiratory diseases in 11 (18%), 10 (16%), and 3 (5%) articles, respectively (Figure 8).

Furthermore, more than half (39/61, 64%) of the articles targeted wellness and lifestyle applications alongside disease-specific metrics. Activity tracking is a fundamental wellness and lifestyle metric collected across 39 (100%) articles. Notably, activity remains essential for monitoring neurological diseases. Furthermore, sleep monitoring is another application widely reported in 10 (26%) articles for various chronic conditions. For cancer, neurological, and metabolic diseases, the articles described remote stress-level monitoring (Figure 8).

Figure 8. Frequency of key medical applications by categories summarized in Figure 1. Rows represent the number of articles (N=61 for chronic disease monitoring and N=39 for wellness and lifestyle applications); colored columns may exceed N because some articles report on multiple wellness or lifestyle applications.



XSL•FO
RenderX

Quality Assessment of Included Articles

More than half of the included articles fully met 11 of the 13 QuADS criteria, receiving the top rating (Complete) on the QuADS scale (Table 2). Specifically, 56 (92%) and 54 (89%) articles clearly described the data collection procedure and stated the research aim, respectively. Fifty-two (85%) articles provided detailed descriptions of the research setting and population, used an appropriate format and content for the data collection tool, and applied proper methods of analysis to address the research aim. In contrast, only 15 (25%) articles thoroughly addressed stakeholder involvement in the design and conduct of the article, while 11 (18%) articles did not report stakeholder involvement at all (Multimedia Appendix 3).

Table . Quality assessment scores for the 13 Quality Assessment with Diverse Studies (QuADS) criteria [50].

QuADS criteria	Rating scale[a], n (%)[b]			
	Not at all	Very slightly	Moderately	Complete
Theoretical or conceptual underpinning to the research	0 (0)	1 (2)	25 (41)	35 (57)
Statement of research aim/s	0 (0)	1 (2)	6 (10)	54 (89)
Clear description of research setting and target population	0 (0)	2 (3)	7 (11)	52 (85)
Study design appropriate to address the stated research aim/s	0 (0)	0 (0)	17 (28)	44 (72)
Appropriate sampling to address the research aim/s	2 (3)	5 (8)	31 (51)	23 (38)
Rationale for choice of data collection tool/s	3 (5)	5 (8)	16 (26)	37 (61)
Format and content of data collection tool appropriate to research aim/s	0 (0)	0 (0)	9 (15)	52 (85)
Description of data collection procedure	0 (0)	0 (0)	5 (8)	56 (92)
Recruitment data provided	4 (7)	5 (8)	14 (23)	38 (62)
Justification for analytic method selected	7 (11)	5 (8)	18 (30)	31 (51)
Method of analysis appropriate to answer the research aim/s	2 (3)	0 (0)	7 (11)	52 (85)
Evidence that stakeholders were considered in design/conduct	11 (18)	12 (20)	23 (38)	15 (25)
Strengths and limitations critically discussed	5 (8)	0 (0)	23 (38)	33 (54)

[a]Each cell indicates the number of articles (N=61) that met the corresponding QuADS criterion at the given rating scale.

[b]Percentages may not total 100% due to rounding.

Discussion

Principal Findings and Interpretation

We identified the wearable devices, sensor types, health parameters, body locations, and medical applications for remote monitoring of various chronic diseases. We found a clear trend toward wrist-worn bands primarily tracking acceleration, followed by core and extended vital signs. These devices mainly monitor neurological and cardiovascular diseases.

Wearable devices predominantly come in the form of bands, a trend driven by consumer electronics giants such as Fitbit [34,55,67,72,90], Apple [55,69,84,99], Garmin [57,80], Huawei [94,96], Samsung [99], Withings [75], Omron HeartGuide [76], and Misfit [63]. These brands offer a diverse range of wearable devices, including smartwatches, wristbands, armbands, and straps. However, most consumer-grade wellness devices lack medical certification, and 23 of 26 are bands. This finding urges future researchers and clinicians to look beyond wearability and user-friendliness and to consider regulatory status. The emergence of consumer-grade wellness devices with certified medical functions (FDA/CE-MDR approval), although still a minority, represents a significant development for RPM systems by addressing the need for clinical validation.

A wide range of wearable devices integrated accelerometers to track acceleration, which emerged as the most frequently monitored and clinically relevant health parameter across

XSL•FO
RenderX

multiple chronic diseases. This is because acceleration offers valuable insights for designing activity plans that support wellness and slow disease progression [102]. Our findings further highlight its particular prevalence in the remote monitoring of neurological disorders and cancer. This aligns with previous research, demonstrating that tracking acceleration enables researchers to gain a deeper understanding of movement patterns, serving as key indicators of Parkinson disease [103-105], and to assess motor function and recovery following a stroke [106,107]. Acceleration is also a key health parameter for monitoring physical fitness levels and evaluating treatment effectiveness in patients with cancer [108].

Vital sign monitoring remains a cornerstone for the remote management of chronic diseases [109], as it provides critical insights into a patient's overall physiological status [110] and serves as an early warning signal for detecting and preventing patient deterioration [111]. Our findings support this principle in the context of wearable devices, revealing that articles frequently use PPG sensors, biopotential meters, pressure meters, and thermometers to track core and extended vital signs, including heart rate, body temperature, blood pressure, and SpO_2.

The wrist is the dominant body location for sensor placement, primarily due to ongoing technological innovations in bands, which improve comfort and usability [112] while enabling the integration of key sensor types, such as accelerometers and PPG sensors [113,114]. When data quality is essential, researchers often choose the upper arm to minimize motion artifacts that corrupt health parameters, compared to the more dynamic wrist and hand [115]. Sensor placement is specific to the health parameter. The chest remains the major site for placing biopotential meters (ECG tracking), which reflects clinical practice [116]. Conversely, PPG sensors do not perform well when placed on the chest, likely due to factors such as skin type and hair density [117-120]. To quantify motor symptoms in the limbs, researchers monitoring neurological diseases often place IMUs on the extremities (eg, the upper arm, forearm, hand, and leg) [121-123].

Wearable devices primarily target neurological and cardiovascular diseases. A recent report from the Institute for Health Metrics and Evaluation supports our findings, identifying neurological diseases as the leading global cause of disease burden and disability, affecting 3.4 billion people, and surpassing the impact of cardiovascular diseases [124,125]. The effectiveness of wearable-based RPM systems for early prediction and prevention is mainly due to their ability to capture quantifiable biomarkers, such as gait analysis, tremors, speech patterns, and cognitive functions. Alongside this, researchers continue to use wearable devices to monitor cardiovascular diseases, which are the leading cause of death globally [126].

Beyond chronic disease monitoring, there is a significant emphasis on wellness and lifestyle monitoring. Approximately two-thirds of the articles track physical activities, such as counting steps and determining movement. Additionally, movement-related parameters, such as tremors, motor symptom assessment, and muscle vibration, are essential for diagnosing neurological diseases. This aligns with the WHO report, which states that promoting healthy behaviors or responding to warning signs can prevent 80% of chronic diseases [127]. Furthermore, our results align with previous research, demonstrating that regular physical activity is essential for managing chronic diseases [128,129], preventing the onset of new diseases [130-132], reducing medication needs [133,134], and enhancing quality of life [135-138]. This emphasizes the importance of setting physical activity plans as a strategy to prevent the progression of chronic diseases.

Another application of this lifestyle approach is the growing focus on remote sleep monitoring. Evidence from the Centers for Disease Control and Prevention [139] and the Population Reference Bureau [140] supports our findings, highlighting the strong association between poor sleep and various chronic diseases, including cardiovascular and metabolic disorders. As a result, we emphasize the importance of monitoring and addressing sleep problems to prevent deterioration in health and promote healthy behaviors in patients with chronic conditions.

Comparison With Prior Work

Several systematic reviews examined the use of wearable devices for remote monitoring of chronic diseases. Some reviews evaluated the impact of wearable-based RPM systems on patient outcomes [36,41,141]. However, these articles do not specify the wearable devices, sensor types, body locations, or health parameters. Other reviews analyze medical applications of RPM systems [142,143], but they do not provide cross-disease mapping of wearable devices, sensor types, body locations, or monitored health parameters. Some reviews focus on specific applications, such as primary health care [37], rehabilitation [14], or physical activity monitoring [144], but do not compare wearable devices or parameters across health conditions.

In contrast, our systematic review provides a framework for wearable devices, sensor types, body locations, health parameters, and medical applications. We visualize relationships among wearable devices, sensor types, and chronic diseases, as well as between sensor types, body locations, and chronic diseases. Building on the recommendation from Cajamarca et al [145], we argue that managing multiple chronic conditions represents a forward-looking approach to address the global rise in multimorbidity. For instance, data from Watson et al [146] show that 51.4% of US adults live with multiple chronic conditions. By identifying shared components across chronic conditions, our findings provide foundational evidence for designing generalized wearable-based RPM systems. Such multifunctional systems cost-effectively monitor multiple chronic conditions simultaneously, moving beyond disease-specific solutions [147].

Martins et al [148] highlighted the need for common standards to address ongoing interoperability challenges across diverse RPM systems. Identifying shared components across pilot projects, commercial products, and CE/FDA-approved wearable devices represents a critical first step toward the semantic harmonization of clinical data across chronic conditions [149]. Our work fosters a deeper understanding of the increasing diversity of RPM systems, supports the assessment of Fast Healthcare Interoperability Resources requirements, encourages performing concept mapping, and promotes the development

XSL•FO
RenderX

of extensions to medical terminologies, including International Classification of Diseases, 11th Revision [150], Systematized Nomenclature of Medicine—Clinical Terms [151], and Logical Observation Identifiers Names and Codes [152].

Limitations and Future Research Directions

This review has several limitations. First, we limited the scope to wearable devices used for the remote monitoring of chronic diseases, excluding unobtrusive or implantable sensors, as well as articles that do not implement RPM systems. While this allows for a more targeted analysis, it may reduce the generalizability of the results to broader sensor technologies. Second, most of the included articles (52/61, 85%) focused on developing RPM systems and conducting early-stage feasibility and usability assessments. Although the field demonstrates strong innovation, it lacks substantial late-stage validation and evaluation of clinical impact on patient outcomes. Consequently, our findings are primarily based on reported device usage frequencies in early-stage and feasibility studies, rather than on demonstrated clinical outcomes. Therefore, readers should not interpret the frequency of a device's appearance in this review as evidence of its clinical effectiveness. Third, we did not evaluate technical aspects, such as long-term device reliability, low power consumption, or the impact of motion artifacts on sensor accuracy. While these factors significantly affect the long-term monitoring of chronic diseases, they fall outside the predefined objectives of this review and require further investigation in future research.

Given this predominance of early-stage feasibility articles, we strongly recommend a shift in research priorities. Future work should prioritize late-stage clinical trials that use medically certified wearable devices. Such articles are essential for rigorously evaluating the clinical efficacy, data privacy, and system maturity, and the impact of motion artifacts, thereby moving RPM systems beyond proof-of-concept toward evidence-based implementation.

Conclusions

This review confirms that the most frequent devices for remote monitoring of chronic diseases are bands, with the wrist being the preferred body location, followed by the upper arm and the chest. The most common sensor types are accelerometers, PPG sensors, and biopotential meters, which primarily collect health parameters such as acceleration, heart rate, and body temperature. Key applications of wearable devices include monitoring neurological and cardiovascular diseases, tracking activity, and assessing sleep quality. Overall, these findings highlight the need for a foundational road map for designing generalized RPM systems that can manage multimorbidity and support standardized terminology to enhance interoperability across digital health systems.

Acknowledgments

No generative artificial intelligence (AI) was used to generate scientific content, images, or any other portion of the manuscript. AI-assisted tools were used only to check grammar, spelling, and language clarity.

Funding

The authors thank TU Braunschweig for providing open-access publication funds.

Data Availability

All data are included in the manuscript and its supplementary materials.

Authors' Contributions

MDT and TMD conceptualized the work and designed the terminology. MDT and VMGS handled the article selection. All authors discussed and approved the extracted data. MDT designed the original draft. TMD, PH, VMGS, SRNK, and JW reviewed and edited the manuscript. MDT, JW, PH, and VMGS conducted the quality assessment of the included articles. MDT and TMD worked on the visualizations and finalized the manuscript. TMD supervised the work. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy used for identifying articles on wearable devices for remote monitoring of chronic diseases.
[DOCX File, 16 KB - mhealth_v14i1e74071_app1.docx]

Multimedia Appendix 2
Complete dataset extracted from the included articles.
[XLSX File, 22 KB - mhealth_v14i1e74071_app2.xlsx]

Multimedia Appendix 3

XSL•FO
RenderX

Quality appraisal of included articles using the 13 Quality Assessment With Diverse Studies criteria.
[DOCX File, 38 KB - mhealth_v14i1e74071_app3.docx]

Checklist 1
PRISMA checklist.
[DOCX File, 281 KB - mhealth_v14i1e74071_app4.docx]

References

1. Noncommunicable diseases. : World Health Organization, Regional Office for Europe; 2011 URL: https://iris.who.int/server/api/core/bitstreams/ac06aa4f-119b-45ed-bfd6-b54fdbdc4f21/content [accessed 2026-01-06]
2. National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP). Centers for Disease Control and Prevention (CDC). 2023. URL: https://www.cdc.gov/nccdphp/index.html [accessed 2026-01-06]
3. Fast facts: health and economic costs of chronic conditions. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/chronic-disease/data-research/facts-stats/index.html [accessed 2026-01-06]
4. Hoffman C, Rice D, Sung HY. Persons with chronic conditions. Their prevalence and costs. JAMA 1996 Nov 13;276(18):1473-1479. [Medline: 8903258]
5. Caton CLM. The new chronic patient and the system of community care. Hosp Community Psychiatry 1981 Jul;32(7):475-478. [doi: 10.1176/ps.32.7.475] [Medline: 7239471]
6. Kirby SE, Dennis SM, Jayasinghe UW, Harris MF. Patient related factors in frequent readmissions: the influence of condition, access to services and patient choice. BMC Health Serv Res 2010 Jul 21;10(1):216. [doi: 10.1186/1472-6963-10-216] [Medline: 20663141]
7. Morabia A, Abel T. The WHO report “Preventing chronic diseases: a vital investment” and us. Soz Praventivmed 2006;51(2):74. [doi: 10.1007/s00038-005-0015-7] [Medline: 18027782]
8. Yetisen AK, Martinez-Hurtado JL, Ünal B, Khademhosseini A, Butt H. Wearables in medicine. Adv Mater 2018 Jun 11;30(33):e1706910. [doi: 10.1002/adma.201706910] [Medline: 29893068]
9. Yilmaz T, Foster R, Hao Y. Detecting vital signs with wearable wireless sensors. Sensors (Basel) 2010;10(12):10837-10862. [doi: 10.3390/s101210837] [Medline: 22163501]
10. Kofoed S, Breen S, Gough K, Aranda S. Benefits of remote real-time side-effect monitoring systems for patients receiving cancer treatment. Oncol Rev 2012 Mar 5;6(1):e7. [doi: 10.4081/oncol.2012.e7] [Medline: 25992209]
11. Hicks LL, Fleming DA, Desaulnier A. The application of remote monitoring to improve health outcomes to a rural area. Telemed J E Health 2009 Sep;15(7):664-671. [doi: 10.1089/tmj.2009.0009] [Medline: 19694598]
12. Deen MJ. Information and communications technologies for elderly ubiquitous healthcare in a smart home. Pers Ubiquit Comput 2015 Jul;19(3-4):573-599. [doi: 10.1007/s00779-015-0856-x]
13. Baig MM, GholamHosseini H, Moqeem AA, Mirza F, Lindén M. A systematic review of wearable patient monitoring systems - current challenges and opportunities for clinical adoption. J Med Syst 2017 Jul;41(7):115. [doi: 10.1007/s10916-017-0760-1] [Medline: 28631139]
14. Patel S, Park H, Bonato P, Chan L, Rodgers M. A review of wearable sensors and systems with application in rehabilitation. J Neuroeng Rehabil 2012 Apr 20;9(1):21. [doi: 10.1186/1743-0003-9-21] [Medline: 22520559]
15. Guo Y, Liu X, Peng S, et al. A review of wearable and unobtrusive sensing technologies for chronic disease management. Comput Biol Med 2021 Feb;129:104163. [doi: 10.1016/j.compbiomed.2020.104163] [Medline: 33348217]
16. Fortune E, Crenshaw JR, Sosnoff JJ. Editorial: wearable sensors for remote health monitoring and intelligent disease management. Front Sports Act Living 2021;3:788165. [doi: 10.3389/fspor.2021.788165] [Medline: 34927069]
17. Liao Y, Thompson C, Peterson S, Mandrola J, Beg MS. The future of wearable technologies and remote monitoring in health care. Am Soc Clin Oncol Educ Book 2019 Jan;39(39):115-121. [doi: 10.1200/EDBK_238919] [Medline: 31099626]
18. Apple Health. Apple. URL: https://www.apple.com/health/ [accessed 2026-01-06]
19. HealthKit. Apple Developer. 2024. URL: https://developer.apple.com/design/human-interface-guidelines/healthkit [accessed 2026-01-06]
20. Google Fit. Google. 2024. URL: https://www.google.com/fit/ [accessed 2026-01-06]
21. Azure Health Data Services. Microsoft Azure. 2024. URL: https://azure.microsoft.com/en-us/products/health-data-services [accessed 2026-01-06]
22. AWS HealthLake. Amazon Web Services. 2024. URL: https://aws.amazon.com/de/healthlake/ [accessed 2026-01-06]
23. Biofourmis. 2024. URL: https://www.biofourmis.com/ [accessed 2026-01-06]
24. Ruokolainen J, Haladijan J, Juutinen M, et al. Mobilemicroservices architecture for remote monitoring of patients: a feasibility study. Stud Health Technol Inform 2022 Jun 6;290:200-204. [doi: 10.3233/SHTI220061] [Medline: 35673000]
25. Zedda A, Gusai E, Caruso M, et al. DoMoMEA: a home-based telerehabilitation system for stroke patients. Annu Int Conf IEEE Eng Med Biol Soc 2020 Jul;2020:5773-5776. [doi: 10.1109/EMBC44109.2020.9175742] [Medline: 33019286]

XSL•FO
RenderX

26. Marin-Pardo O, Phanord C, Donnelly MR, Laine CM, Liew SL. Development of a low-cost, modular muscle-computer interface for at-home telerehabilitation for chronic stroke. Sensors (Basel) 2021 Mar 5;21(5):1806. [doi: 10.3390/s21051806] [Medline: 33807691]
27. Ravichandran V, Sadhu S, Convey D, et al. iTex gloves: design and in-home evaluation of an e-textile glove system for tele-assessment of Parkinson's disease. Sensors (Basel) 2023 Mar 7;23(6):2877. [doi: 10.3390/s23062877] [Medline: 36991587]
28. Scholten HJ, Shih CD, Ma R, Malhotra K, Reyzelman AM. Utilization of a smart sock for the remote monitoring of patients with peripheral neuropathy: cross-sectional study of a real-world registry. JMIR Form Res 2022 Mar 1;6(3):e32934. [doi: 10.2196/32934] [Medline: 35230248]
29. Knijff JM, Houdijk E, van der Kaay DCM, et al. Objective home-monitoring of physical activity, cardiovascular parameters, and sleep in pediatric obesity. Digit Biomark 2022;6(1):19-29. [doi: 10.1159/000522185] [Medline: 36148249]
30. Majumder S, Mondal T, Deen MJ. Wearable sensors for remote health monitoring. Sensors (Basel) 2017 Jan 12;17(1):130. [doi: 10.3390/s17010130] [Medline: 28085085]
31. Albani G, Ferraris C, Nerino R, et al. An integrated multisensor approach for the remote monitoring of Parkinson's disease. Sensors (Basel) 2019 Nov 2;19(21):4764. [doi: 10.3390/s19214764] [Medline: 31684020]
32. Bergenstal RM, Layne JE, Zisser H, et al. Remote application and use of real-time continuous glucose monitoring by adults with type 2 diabetes in a virtual diabetes clinic. Diabetes Technol Ther 2021 Feb;23(2):128-132. [doi: 10.1089/dia.2020.0396] [Medline: 33026839]
33. Santala OE, Lipponen JA, Jäntti H, et al. Continuous mHealth patch monitoring for the algorithm-based detection of atrial fibrillation: feasibility and diagnostic accuracy study. JMIR Cardio 2022 Jun 21;6(1):e31230. [doi: 10.2196/31230] [Medline: 35727618]
34. Gao Z, Ryu S, Zhou W, et al. Effects of personalized exercise prescriptions and social media delivered through mobile health on cancer survivors' physical activity and quality of life. J Sport Health Sci 2023 Nov;12(6):705-714. [doi: 10.1016/j.jshs.2023.07.002] [Medline: 37467931]
35. Blanc M, Roy AL, Fraudet B, et al. Evaluation of a digitally guided self-rehabilitation device coupled with telerehabilitation monitoring in patients with Parkinson disease (TELEP@RK): open, prospective observational study. JMIR Serious Games 2022 Feb 7;10(1):e24946. [doi: 10.2196/24946] [Medline: 35129449]
36. Kamei T, Kanamori T, Yamamoto Y, Edirippulige S. The use of wearable devices in chronic disease management to enhance adherence and improve telehealth outcomes: a systematic review and meta-analysis. J Telemed Telecare 2022 Jun;28(5):342-359. [doi: 10.1177/1357633X20937573] [Medline: 32819184]
37. Peyroteo M, Ferreira IA, Elvas LB, Ferreira JC, Lapão LV. Remote monitoring systems for patients with chronic diseases in primary health care: systematic review. JMIR Mhealth Uhealth 2021 Dec 21;9(12):e28285. [doi: 10.2196/28285] [Medline: 34932000]
38. Bent B, Sim I, Dunn JP. Digital medicine community perspectives and challenges: survey study. JMIR Mhealth Uhealth 2021 Feb 3;9(2):e24570. [doi: 10.2196/24570] [Medline: 33533721]
39. Canali S, Schiaffonati V, Aliverti A. Challenges and recommendations for wearable devices in digital health: data quality, interoperability, health equity, fairness. PLOS Digit Health 2022 Oct;1(10):e0000104. [doi: 10.1371/journal.pdig.0000104] [Medline: 36812619]
40. Dinh-Le C, Chuang R, Chokshi S, Mann D. Wearable health technology and electronic health record integration: scoping review and future directions. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e12861. [doi: 10.2196/12861] [Medline: 31512582]
41. Mattison G, Canfell O, Forrester D, et al. The influence of wearables on health care outcomes in chronic disease: systematic review. J Med Internet Res 2022 Jul 1;24(7):e36690. [doi: 10.2196/36690] [Medline: 35776492]
42. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. PLOS Med 2009 Jul 21;6(7):e1000097. [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
43. Seo YW, La Marca V, Tandon A, Chiao JC, Drummond CK. Exploring the design for wearability of wearable devices: a scoping review. Computers 2024;13(12):326. [doi: 10.3390/computers13120326]
44. Iqbal SMA, Mahgoub I, Du E, Leavitt MA, Asghar W. Advances in healthcare wearable devices. NPJ Flex Electron 2021;5(1):9. [doi: 10.1038/s41528-021-00107-x]
45. Abhinav V, Basu P, Verma SS, et al. Advancements in wearable and implantable BioMEMS devices: transforming healthcare through technology. Micromachines (Basel) 2025 Apr 28;16(5):522. [doi: 10.3390/mi16050522] [Medline: 40428648]
46. White RM. A sensor classification scheme. IEEE Trans Ultrason Ferroelectr Freq Control 1987;34(2):124-126. [doi: 10.1109/t-uffc.1987.26922] [Medline: 18290099]
47. Kim N, Ock YS, Kim M. A classification scheme of wearable devices for the identification of strategic market segment. ICIC express Lett Part B Appl 2017;8(2):269-275 [FREE Full text]
48. Dunn J, Runge R, Snyder M. Wearables and the medical revolution. Per Med 2018 Sep;15(5):429-448. [doi: 10.2217/pme-2018-0044] [Medline: 30259801]
49. Kristoffersson A, Lindén M. A systematic review on the use of wearable body sensors for health monitoring: a qualitative synthesis. Sensors (Basel) 2020 Mar 9;20(5):1502. [doi: 10.3390/s20051502] [Medline: 32182907]

XSL•FO
RenderX

50. Harrison R, Jones B, Gardner P, Lawton R. Quality assessment with diverse studies (QuADS): an appraisal tool for methodological and reporting quality in systematic reviews of mixed- or multi-method studies. BMC Health Serv Res 2021 Feb 15;21(1):144. [doi: 10.1186/s12913-021-06122-y] [Medline: 33588842]
51. Miller L, Alele FO, Emeto TI, Franklin RC. Epidemiology, risk factors and measures for preventing drowning in Africa: a systematic review. Medicina (Kaunas) 2019 Sep 25;55(10):637. [doi: 10.3390/medicina55100637] [Medline: 31557943]
52. Radogna AV, Siciliano PA, Sabina S, Sabato E, Capone S. A low-cost breath analyzer module in domiciliary non-invasive mechanical ventilation for remote COPD patient monitoring. Sensors (Basel) 2020 Jan 24;20(3):653. [doi: 10.3390/s20030653] [Medline: 31991608]
53. Boscari F, Ferretto S, Uliana A, Avogaro A, Bruttomesso D. Efficacy of telemedicine for persons with type 1 diabetes during Covid19 lockdown. Nutr Diabetes 2021 Jan 5;11(1):1. [doi: 10.1038/s41387-020-00147-8] [Medline: 33414391]
54. Gatsios D, Antonini A, Gentile G, et al. Feasibility and utility of mHealth for the remote monitoring of Parkinson disease: ancillary study of the PD_manager randomized controlled trial. JMIR Mhealth Uhealth 2020 Jun 29;8(6):e16414. [doi: 10.2196/16414] [Medline: 32442154]
55. Cappon G, Cossu L, Boscari F, Bruttomesso D, Sparacino G, Facchinetti A. An integrated mobile platform for automated data collection and real-time patient monitoring in diabetes clinical trials. J Diabetes Sci Technol 2022 Nov;16(6):1555-1559. [doi: 10.1177/19322968211024620] [Medline: 34218721]
56. Huo W, Angeles P, Tai YF, et al. A heterogeneous sensing suite for multisymptom quantification of Parkinson's disease. IEEE Trans Neural Syst Rehabil Eng 2020 Jun;28(6):1397-1406. [doi: 10.1109/TNSRE.2020.2978197] [Medline: 32305925]
57. Althobiani MA, Ranjan Y, Jacob J, et al. Evaluating a remote monitoring program for respiratory diseases: prospective observational study. JMIR Form Res 2023 Nov 24;7:e51507. [doi: 10.2196/51507] [Medline: 37999935]
58. Lévi F, Komarzynski S, Huang Q, et al. Tele-monitoring of cancer patients' rhythms during daily life identifies actionable determinants of circadian and sleep disruption. Cancers (Basel) 2020 Jul 17;12(7):1938. [doi: 10.3390/cancers12071938] [Medline: 32708950]
59. Beach C, Cooper G, Weightman A, Hodson-Tole EF, Reeves ND, Casson AJ. Monitoring of dynamic plantar foot temperatures in diabetes with personalised 3D-printed wearables. Sensors (Basel) 2021 Mar 2;21(5):1717. [doi: 10.3390/s21051717] [Medline: 33801346]
60. Forlenza GP, Ekhlaspour L, Breton M, et al. Successful at-home use of the tandem control-IQ artificial pancreas system in young children during a randomized controlled trial. Diabetes Technol Ther 2019 Apr;21(4):159-169. [doi: 10.1089/dia.2019.0011] [Medline: 30888835]
61. Bui AAT, Hosseini A, Rocchio R, et al. Biomedical REAl-Time Health Evaluation (BREATHE): toward an mHealth informatics platform. JAMIA Open 2020 Jul;3(2):190-200. [doi: 10.1093/jamiaopen/ooaa011] [Medline: 32734159]
62. Stehlik J, Schmalfuss C, Bozkurt B, et al. Continuous wearable monitoring analytics predict heart failure hospitalization: the LINK-HF multicenter study. Circ Heart Fail 2020 Mar;13(3):e006513. [doi: 10.1161/CIRCHEARTFAILURE.119.006513] [Medline: 32093506]
63. Ghods A, Shahrokni A, Ghasemzadeh H, Cook D. Remote monitoring of the performance status and burden of symptoms of patients with gastrointestinal cancer via a consumer-based activity tracker: quantitative cohort study. JMIR Cancer 2021 Nov 26;7(4):e22931. [doi: 10.2196/22931] [Medline: 34842527]
64. Peterson SK, Basen-Engquist K, Demark-Wahnefried W, et al. Feasibility of mobile and sensor technology for remote monitoring in cancer care and prevention. AMIA Annu Symp Proc 2021;2021:979-988. [Medline: 35308916]
65. Lee BC, An J, Kim J, Lai EC. Performing dynamic weight-shifting balance exercises with a smartphone-based wearable telerehabilitation system for home use by individuals with Parkinson's disease: a proof-of-concept study. IEEE Trans Neural Syst Rehabil Eng 2023;31:456-463. [doi: 10.1109/TNSRE.2022.3226368] [Medline: 36455080]
66. Frankel MA, Lehmkuhle MJ, Watson M, et al. Electrographic seizure monitoring with a novel, wireless, single-channel EEG sensor. Clin Neurophysiol Pract 2021;6:172-178. [doi: 10.1016/j.cnp.2021.04.003] [Medline: 34189361]
67. Sohn A, Speier W, Lan E, et al. Integrating remote monitoring into heart failure patients' care regimen: a pilot study. PLoS ONE 2020;15(11):e0242210. [doi: 10.1371/journal.pone.0242210] [Medline: 33211733]
68. Darcy B, Rashford L, Shultz ST, et al. Gait device treatment using telehealth for individuals with stroke during the COVID-19 pandemic: nonrandomized pilot feasibility study. JMIR Form Res 2023 May 19;7:e43008. [doi: 10.2196/43008] [Medline: 37204830]
69. Weng D, Ding J, Sharma A, et al. Heart rate trajectories in patients recovering from acute myocardial infarction: a longitudinal analysis of Apple Watch heart rate recordings. Cardiovasc Digit Health J 2021 Oct;2(5):270-281. [doi: 10.1016/j.cvdhj.2021.05.003] [Medline: 35265918]
70. Noorian AR, Bahr Hosseini M, Avila G, et al. Use of wearable technology in remote evaluation of acute stroke patients: feasibility and reliability of a Google glass-based device. J Stroke Cerebrovasc Dis 2019 Oct;28(10):104258. [doi: 10.1016/j.jstrokecerebrovasdis.2019.06.016] [Medline: 31296476]
71. Chi WN, Reamer C, Gordon R, et al. Continuous remote patient monitoring: evaluation of the heart failure cascade soft launch. Appl Clin Inform 2021 Oct;12(5):1161-1173. [doi: 10.1055/s-0041-1740480] [Medline: 34965606]

XSL•FO
RenderX

72. Waddell KJ, Patel MS, Wilkinson JR, et al. Deploying digital health technologies for remote physical activity monitoring of rural populations with chronic neurologic disease. Arch Rehabil Res Clin Transl 2023 Mar;5(1):100250. [doi: 10.1016/j.arrct.2022.100250] [Medline: 36968173]
73. Sharma M, Mishra RK, Hall AJ, et al. Remote at-home wearable-based gait assessments in Progressive Supranuclear Palsy compared to Parkinson's Disease. BMC Neurol 2023 Dec 11;23(1):434. [doi: 10.1186/s12883-023-03466-2] [Medline: 38082255]
74. Ho K, Novak Lauscher H, Cordeiro J, et al. Testing the feasibility of sensor-based home health monitoring (TEC4Home) to support the convalescence of patients with heart failure: pre-post study. JMIR Form Res 2021 Jun 3;5(6):e24509. [doi: 10.2196/24509] [Medline: 34081015]
75. Campo D, Elie V, de Gallard T, et al. Atrial fibrillation detection with an analog smartwatch: prospective clinical study and algorithm validation. JMIR Form Res 2022 Nov 4;6(11):e37280. [doi: 10.2196/37280] [Medline: 35481559]
76. Vaseekaran M, Kaese S, Görlich D, Wiemer M, Samol A. WATCH-BPM-comparison of a WATCH-type blood pressure monitor with a conventional ambulatory blood pressure monitor and auscultatory sphygmomanometry. Sensors (Basel) 2023 Oct 31;23(21):8877. [doi: 10.3390/s23218877] [Medline: 37960576]
77. Bleda AL, Melgarejo-Meseguer FM, Gimeno-Blanes FJ, et al. Enabling heart self-monitoring for all and for AAL-portable device within a complete telemedicine system. Sensors (Basel) 2019 Sep 14;19(18):3969. [doi: 10.3390/s19183969] [Medline: 31540042]
78. Filakova K, Janikova A, Felsoci M, et al. Home-based cardio-oncology rehabilitation using a telerehabilitation platform in hematological cancer survivors: a feasibility study. BMC Sports Sci Med Rehabil 2023 Mar 23;15(1):38. [doi: 10.1186/s13102-023-00650-2] [Medline: 36959613]
79. Blockhaus C, List S, Waibler HP, et al. Wearable cardioverter-defibrillator used as a telemonitoring system in a real-life heart failure unit setting. J Clin Med 2021 Nov 22;10(22):5435. [doi: 10.3390/jcm10225435] [Medline: 34830724]
80. Kytö M, Koivusalo S, Tuomonen H, et al. Supporting the management of gestational diabetes mellitus with comprehensive self-tracking: mixed methods study of wearable sensors. JMIR Diabetes 2023 Oct 31;8:e43979. [doi: 10.2196/43979] [Medline: 37906216]
81. Pavic M, Klaas V, Theile G, et al. Mobile health technologies for continuous monitoring of cancer patients in palliative care aiming to predict health status deterioration: a feasibility study. J Palliat Med 2020 May;23(5):678-685. [doi: 10.1089/jpm.2019.0342] [Medline: 31873052]
82. Pavic M, Klaas V, Theile G, Kraft J, Tröster G, Guckenberger M. Feasibility and usability aspects of continuous remote monitoring of health status in palliative cancer patients using wearables. Oncology (Williston Park, NY) 2020;98(6):386-395. [doi: 10.1159/000501433] [Medline: 31336377]
83. Notemi LM, Amoura L, Fall Mostaine F, et al. Long-term efficacy of sensor-augmented pump therapy (Minimed 640G system) combined with a telemedicine follow-up in patients with type 1 diabetes: a real life study. J Clin Transl Endocrinol 2022 Dec;30:100306. [doi: 10.1016/j.jcte.2022.100306] [Medline: 36238800]
84. Werhahn SM, Dathe H, Rottmann T, et al. Designing meaningful outcome parameters using mobile technology: a new mobile application for telemonitoring of patients with heart failure. ESC Heart Fail 2019 Jun;6(3):516-525. [doi: 10.1002/ehf2.12425] [Medline: 30868756]
85. Ausín JL, Ramos J, Lorido A, Molina P, Duque-Carrillo JF. Wearable and noninvasive device for integral congestive heart failure management in the IoMT paradigm. Sensors (Basel) 2023 Aug 9;23(16):7055. [doi: 10.3390/s23167055] [Medline: 37631594]
86. Kovacs B, Müller F, Niederseer D, et al. Wearable cardioverter-defibrillator-measured step count for the surveillance of physical fitness during cardiac rehabilitation. Sensors (Basel) 2021 Oct 25;21(21):7054. [doi: 10.3390/s21217054] [Medline: 34770361]
87. Sigcha L, Polvorinos-Fernández C, Costa N, et al. Monipar: movement data collection tool to monitor motor symptoms in Parkinson's disease using smartwatches and smartphones. Front Neurol 2023;14:1326640. [doi: 10.3389/fneur.2023.1326640] [Medline: 38148984]
88. Morales-Botello ML, Gachet D, de Buenaga M, Aparicio F, Busto MJ, Ascanio JR. Chronic patient remote monitoring through the application of big data and internet of things. Health Informatics J 2021;27(3):14604582211030956. [doi: 10.1177/14604582211030956] [Medline: 34256646]
89. Lipsmeier F, Taylor KI, Postuma RB, et al. Reliability and validity of the Roche PD mobile application for remote monitoring of early Parkinson's disease. Sci Rep 2022 Jul 15;12(1):12081. [doi: 10.1038/s41598-022-15874-4] [Medline: 35840753]
90. Chung IY, Jung M, Lee SB, et al. An assessment of physical activity data collected via a smartphone app and a smart band in breast cancer survivors: observational study. J Med Internet Res 2019 Sep 6;21(9):13463. [doi: 10.2196/13463] [Medline: 31493319]
91. Chae SH, Kim Y, Lee KS, Park HS. Development and clinical evaluation of a web-based upper limb home rehabilitation system using a smartwatch and machine learning model for chronic stroke survivors: prospective comparative study. JMIR Mhealth Uhealth 2020 Jul 9;8(7):e17216. [doi: 10.2196/17216] [Medline: 32480361]
92. Shivaraja TR, Remli R, Kamal N, Wan Zaidi WA, Chellappan K. Assessment of a 16-channel ambulatory dry electrode EEG for remote monitoring. Sensors (Basel) 2023 Mar 31;23(7):3654. [doi: 10.3390/s23073654] [Medline: 37050713]

XSL•FO
RenderX

93. Wong CK, Un KC, Zhou M, et al. Daily ambulatory remote monitoring system for drug escalation in chronic heart failure with reduced ejection fraction: pilot phase of DAVID-HF study. Eur Heart J Digit Health 2022;3(2):284-295. [doi: 10.1093/ehjdh/ztac024] [Medline: 36713022]
94. Lee WL, Danaee M, Abdullah A, Wong LP. Is the blood pressure-enabled smartwatch ready to drive precision medicine? Supporting findings from a validation study. Cardiol Res 2023 Dec;14(6):437-445. [doi: 10.14740/cr1569] [Medline: 38187511]
95. Lee M, Shin J. Change in waist circumference with continuous use of a smart belt: an observational study. JMIR Mhealth Uhealth 2019 May 2;7(5):e10737. [doi: 10.2196/10737] [Medline: 31045500]
96. Ramesh J, Aburukba R, Sagahyroon A. A remote healthcare monitoring framework for diabetes prediction using machine learning. Healthc Technol Lett 2021 Jun;8(3):45-57. [doi: 10.1049/htl2.12010] [Medline: 34035925]
97. Ayyoubzadeh SM, Baniasadi T, Shirkhoda M, et al. Remote monitoring of colorectal cancer survivors using a smartphone app and internet of things-based device: development and usability study. JMIR Cancer 2023 Feb 15;9:e42250. [doi: 10.2196/42250] [Medline: 36790851]
98. Guo L, Wang J, Wu Q, et al. Clinical study of a wearable remote rehabilitation training system for patients with stroke: randomized controlled pilot trial. JMIR Mhealth Uhealth 2023 Feb 23;11:e40416. [doi: 10.2196/40416] [Medline: 36821348]
99. Wu JM, Ho TW, Chang YT, et al. Wearable-based mobile health app in gastric cancer patients for postoperative physical activity monitoring: focus group study. JMIR Mhealth Uhealth 2019 Apr 23;7(4):e11989. [doi: 10.2196/11989] [Medline: 31012858]
100. Ha L, Wakefield CE, Mizrahi D, et al. A digital educational intervention with wearable activity trackers to support health behaviors among childhood cancer survivors: pilot feasibility and acceptability study. JMIR Cancer 2022 Aug 17;8(3):e38367. [doi: 10.2196/38367] [Medline: 35976683]
101. Toh SFM, Gonzalez PC, Fong KNK. Usability of a wearable device for home-based upper limb telerehabilitation in persons with stroke: a mixed-methods study. Digit Health 2023;9:20552076231153737. [doi: 10.1177/20552076231153737] [Medline: 36776407]
102. Pate RR, Pratt M, Blair SN, Haskell WL, Macera CA, Bouchard C, et al. Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. JAMA 1995;273(5):402-407. [doi: 10.1001/jama.1995.03520290054029] [Medline: 7823386]
103. Thorp JE, Adamczyk PG, Ploeg HL, Pickett KA. Monitoring motor symptoms during activities of daily living in individuals with Parkinson's disease. Front Neurol 2018;9:1036. [doi: 10.3389/fneur.2018.01036] [Medline: 30619024]
104. Das S, Trutoiu L, Murai A, et al. Quantitative measurement of motor symptoms in Parkinson's disease: a study with full-body motion capture data. Annu Int Conf IEEE Eng Med Biol Soc 2011;2011:6789-6792. [doi: 10.1109/IEMBS.2011.6091674] [Medline: 22255897]
105. Ossig C, Antonini A, Buhmann C, et al. Wearable sensor-based objective assessment of motor symptoms in Parkinson's disease. J Neural Transm (Vienna) 2016 Jan;123(1):57-64. [doi: 10.1007/s00702-015-1439-8] [Medline: 26253901]
106. Gubbi J, Rao AS, Fang K, Yan B, Palaniswami M. Motor recovery monitoring using acceleration measurements in post acute stroke patients. Biomed Eng Online 2013 Apr 16;12:33. [doi: 10.1186/1475-925X-12-33] [Medline: 23590690]
107. Balasubramanian S, Colombo R, Sterpi I, Sanguineti V, Burdet E. Robotic assessment of upper limb motor function after stroke. Am J Phys Med Rehabil 2012 Nov;91(11 Suppl 3):S255-S269. [doi: 10.1097/PHM.0b013e31826bcdc1] [Medline: 23080041]
108. Dadhania S, Williams M. Wearable accelerometers in cancer patients. In: Lim CP, Chen YW, Vaidya A, Mahorkar C, Jain LC, editors. Handbook of Artificial Intelligence in Healthcare: Vol 2: Practicalities and Prospects: Springer International Publishing; 2022:109-147. [doi: 10.1007/978-3-030-83620-7_5]
109. Harries AD, Zachariah R, Kapur A, Jahn A, Enarson DA. The vital signs of chronic disease management. Trans R Soc Trop Med Hyg 2009 Jun;103(6):537-540. [doi: 10.1016/j.trstmh.2008.12.008] [Medline: 19155031]
110. Mohammed KI, Zaidan AA, Zaidan BB, et al. Real-time remote-health monitoring systems: a review on patients prioritisation for multiple-chronic diseases, taxonomy analysis, concerns and solution procedure. J Med Syst 2019 Jun 11;43(7):223. [doi: 10.1007/s10916-019-1362-x] [Medline: 31187288]
111. Ko HYK, Tripathi NK, Mozumder C, Muengtaweepongsa S, Pal I. Real-time remote patient monitoring and alarming system for noncommunicable lifestyle diseases. Int J Telemed Appl 2023;2023:9965226. [doi: 10.1155/2023/9965226] [Medline: 38020047]
112. Davoudi A, Mardini MT, Nelson D, et al. The effect of sensor placement and number on physical activity recognition and energy expenditure estimation in older adults: validation study. JMIR Mhealth Uhealth 2021 May 3;9(5):e23681. [doi: 10.2196/23681] [Medline: 33938809]
113. Schroeder NL, Romine WL, Kemp SE. A scoping review of wrist-worn wearables in education. Comput Educ Open 2023 Dec;5:100154. [doi: 10.1016/j.caeo.2023.100154]
114. Henriksen A, Haugen Mikalsen M, Woldaregay AZ, et al. Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. J Med Internet Res 2018 Mar 22;20(3):e110. [doi: 10.2196/jmir.9157] [Medline: 29567635]

XSL•FO
RenderX

115. Maeda Y, Sekine M, Tamura T. Relationship between measurement site and motion artifacts in wearable reflected photoplethysmography. J Med Syst 2011 Oct;35(5):969-976. [doi: 10.1007/s10916-010-9505-0] [Medline: 20703691]
116. Stracina T, Ronzhina M, Redina R, Novakova M. Golden standard or obsolete method? Review of ECG applications in clinical and experimental context. Front Physiol 2022;13:867033. [doi: 10.3389/fphys.2022.867033] [Medline: 35547589]
117. Hochstadt A, Havakuk O, Chorin E, et al. Continuous heart rhythm monitoring using mobile photoplethysmography in ambulatory patients. J Electrocardiol 2020;60:138-141. [doi: 10.1016/j.jelectrocard.2020.04.017] [Medline: 32361522]
118. Sañudo B, De Hoyo M, Muñoz-López A, Perry J, Abt G. Pilot study assessing the influence of skin type on the heart rate measurements obtained by photoplethysmography with the Apple watch. J Med Syst 2019 May 22;43(7):195. [doi: 10.1007/s10916-019-1325-2] [Medline: 31119387]
119. Boonya-Ananta T, Rodriguez AJ, Ajmal A, et al. Synthetic photoplethysmography (PPG) of the radial artery through parallelized Monte Carlo and its correlation to body mass index (BMI). Sci Rep 2021 Jan 28;11(1):2570. [doi: 10.1038/s41598-021-82124-4] [Medline: 33510428]
120. Fallow BA, Tarumi T, Tanaka H. Influence of skin type and wavelength on light wave reflectance. J Clin Monit Comput 2013 Jun;27(3):313-317. [doi: 10.1007/s10877-013-9436-7] [Medline: 23397431]
121. Halloran S, Tang L, Guan Y, Shi JQ, Eyre J. Remote monitoring of stroke patients' rehabilitation using wearable accelerometers. Presented at: Proceedings of the 2019 ACM International Symposium on Wearable Computers; Sep 9-13, 2019; London, United Kingdom p. 72-77. [doi: 10.1145/3341163.3347731]
122. Zampogna A, Mileti I, Palermo E, et al. Fifteen years of wireless sensors for balance assessment in neurological disorders. Sensors (Basel) 2020 Jun 7;20(11):3247. [doi: 10.3390/s20113247] [Medline: 32517315]
123. Cicirelli G, Impedovo D, Dentamaro V, Marani R, Pirlo G, D'Orazio TR. Human gait analysis in neurodegenerative diseases: a review. IEEE J Biomed Health Inform 2022 Jan;26(1):229-242. [doi: 10.1109/JBHI.2021.3092875] [Medline: 34181559]
124. Steinmetz JD, Seeher KM, Schiess N, GBD 2021 Nervous System Disorders Collaborators, et al. Global, regional, and national burden of disorders affecting the nervous system, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet Neurol 2024 Apr;23(4):344-381. [doi: 10.1016/S1474-4422(24)00038-3] [Medline: 38493795]
125. The Lancet Neurology: neurological conditions now leading cause of ill health and disability globally, affecting 34 billion people worldwide. Institute for Health Metrics and Evaluation. 2024 Mar 14. URL: https://www.healthdata.org/news-events/newsroom/news-releases/lancet-neurology-neurological-conditions-now-leading-cause-ill [accessed 2026-01-06]
126. New study reveals latest data on global burden of cardiovascular disease. Institute for Health Metrics and Evaluation. 2023 Dec 11. URL: https://www.healthdata.org/news-events/newsroom/news-releases/new-study-reveals-latest-data-global-burden-cardiovascular [accessed 2026-01-06]
127. WHO action framework for the prevention and control of chronic diseases. : World Health Organization; 2006 URL: https://www.afro.who.int/sites/default/files/2017-06/Action%20Framework%20for%20the%20Prevention%20and%20Control%20of%20Chronic%20Disease.pdf [accessed 2026-01-06]
128. Phillips SM, Cadmus-Bertram L, Rosenberg D, Buman MP, Lynch BM. Wearable technology and physical activity in chronic disease: opportunities and challenges. Am J Prev Med 2018 Jan;54(1):144-150. [doi: 10.1016/j.amepre.2017.08.015] [Medline: 29122356]
129. Chiauzzi E, Rodarte C, DasMahapatra P. Patient-centered activity monitoring in the self-management of chronic health conditions. BMC Med 2015 Apr 9;13(1):77. [doi: 10.1186/s12916-015-0319-2] [Medline: 25889598]
130. Adami PE, Negro A, Lala N, Martelletti P. The role of physical activity in the prevention and treatment of chronic diseases. Clin Ter 2010;161(6):537-541. [Medline: 21181084]
131. Anderson E, Durstine JL. Physical activity, exercise, and chronic diseases: a brief review. Sports Med Health Sci 2019 Dec;1(1):3-10. [doi: 10.1016/j.smhs.2019.08.006] [Medline: 35782456]
132. Kruk J. Physical activity in the prevention of the most frequent chronic diseases: an analysis of the recent evidence. Asian Pac J Cancer Prev 2007;8(3):325-338. [Medline: 18159963]
133. Kesaniemi YK, Danforth E Jr, Jensen MD, Kopelman PG, Lefèbvre P, Reeder BA. Dose-response issues concerning physical activity and health: an evidence-based symposium. Med Sci Sports Exerc 2001 Jun;33(6 Suppl):S351-S358. [doi: 10.1097/00005768-200106001-00003] [Medline: 11427759]
134. Börjesson M, Onerup A, Lundqvist S, Dahlöf B. Physical activity and exercise lower blood pressure in individuals with hypertension: narrative review of 27 RCTs. Br J Sports Med 2016 Mar;50(6):356-361. [doi: 10.1136/bjsports-2015-095786] [Medline: 26787705]
135. Painter P, Krasnoff J, Paul SM, Ascher NL. Physical activity and health-related quality of life in liver transplant recipients. Liver Transpl 2001 Mar;7(3):213-219. [doi: 10.1053/jlts.2001.22184] [Medline: 11244162]
136. Acree LS, Longfors J, Fjeldstad AS, et al. Physical activity is related to quality of life in older adults. Health Qual Life Outcomes 2006 Jun 30;4(1):37. [doi: 10.1186/1477-7525-4-37] [Medline: 16813655]
137. Stroud NM, Minahan CL. The impact of regular physical activity on fatigue, depression and quality of life in persons with multiple sclerosis. Health Qual Life Outcomes 2009 Jul 20;7:68. [doi: 10.1186/1477-7525-7-68] [Medline: 19619337]

XSL•FO
RenderX

138. Bogdanovic G, Stojanovich L, Djokovic A, Stanisavljevic N. Physical activity program is helpful for improving quality of life in patients with systemic lupus erythematosus. Tohoku J Exp Med 2015 Nov;237(3):193-199. [doi: 10.1620/tjem.237.193] [Medline: 26490344]
139. About sleep and your heart health. Centers for Disease Control and Prevention. 2024 May 15. URL: https://www.cdc.gov/heart-disease/about/sleep-and-heart-health.html [accessed 2026-01-06]
140. New evidence on sleep's role in aging and chronic disease. : Population Reference Bureau; 2018 URL: https://www.prb.org/resources/new-evidence-on-sleeps-role-in-aging-and-chronic-disease/ [accessed 2026-01-06]
141. Tan SY, Sumner J, Wang Y, Wenjun Yip A. A systematic review of the impacts of remote patient monitoring (RPM) interventions on safety, adherence, quality-of-life and cost-related outcomes. NPJ Digit Med 2024 Jul 18;7(1):192. [doi: 10.1038/s41746-024-01182-w] [Medline: 39025937]
142. Jafleh EA, Alnaqbi FA, Almaeeni HA, Faqeeh S, Alzaabi MA, Al Zaman K. The role of wearable devices in chronic disease monitoring and patient care: a comprehensive review. Cureus 2024 Sep;16(9):e68921. [doi: 10.7759/cureus.68921] [Medline: 39381470]
143. Malasinghe LP, Ramzan N, Dahal K. Remote patient monitoring: a comprehensive study. J Ambient Intell Human Comput 2019 Jan;10(1):57-76. [doi: 10.1007/s12652-017-0598-x]
144. Kristoffersson A, Lindén M. A systematic review of wearable sensors for monitoring physical activity. Sensors (Basel) 2022 Jan 12;22(2):573. [doi: 10.3390/s22020573] [Medline: 35062531]
145. Cajamarca G, Proust V, Herskovic V, Cádiz RF, Verdezoto N, Fernández FJ. Technologies for managing the health of older adults with multiple chronic conditions: a systematic literature review. Healthcare (Basel) 2023 Nov 3;11(21):2897. [doi: 10.3390/healthcare11212897] [Medline: 37958041]
146. Watson KB, Wiltz JL, Nhim K, Kaufmann RB, Thomas CW, Greenlund KJ. Trends in multiple chronic conditions among US adults, by life stage, Behavioral Risk Factor Surveillance System, 2013-2023. Prev Chronic Dis 2025 Apr 17;22:E15. [doi: 10.5888/pcd22.240539] [Medline: 40245168]
147. De Guzman KR, Snoswell CL, Taylor ML, Gray LC, Caffery LJ. Economic evaluations of remote patient monitoring for chronic disease: a systematic review. Value Health 2022 Jun;25(6):897-913. [doi: 10.1016/j.jval.2021.12.001] [Medline: 35667780]
148. Martins D, Lewerenz S, Carmo A, Martins H. Interoperability of telemonitoring data in digital health solutions: a scoping review. Front Digit Health 2025;7:1502260. [doi: 10.3389/fdgth.2025.1502260] [Medline: 40330872]
149. Palojoki S, Lehtonen L, Vuokko R. Semantic interoperability of electronic health records: systematic review of alternative approaches for enhancing patient information availability. JMIR Med Inform 2024 Apr 25;12:e53535. [doi: 10.2196/53535] [Medline: 38686541]
150. International Classification of Diseases 11th Revision (ICD-11). World Health Organization. 2022. URL: https://icd.who.int/en/ [accessed 2026-01-06]
151. SNOMED International. URL: https://www.snomed.org/ [accessed 2026-01-06]
152. LONIC. URL: https://loinc.org/ [accessed 2026-01-06]

Abbreviations

CE-MDR: European Conformity Medical Device Regulation
ECG: electrocardiography
FDA: US Food and Drug Administration
HIPAA: Health Insurance Portability and Accountability Act
IMU: inertial measurement unit
PPG: photoplethysmographic
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QuADS: Quality Assessment With Diverse Studies
RPM: remote patient monitoring
SpO_2: peripheral oxygen saturation
WHO: World Health Organization

XSL•FO
RenderX

Edited by A Stone; submitted 17.Mar.2025; peer-reviewed by AI Faisal, A Kumar, W Jiang; revised version received 25.Nov.2025; accepted 25.Nov.2025; published 11.Feb.2026.

Please cite as:
Tegegne MD, Niakan Kalhori SR, Haas P, Sobotta VMG, Warnecke J, Deserno TM
Wearable Devices for Remote Monitoring of Chronic Diseases: Systematic Review
JMIR Mhealth Uhealth 2026;14:e74071
URL: https://mhealth.jmir.org/2026/1/e74071
doi:10.2196/74071

© Masresha Derese Tegegne, Sharareh Rostam Niakan Kalhori, Paulo Haas, Viktor MG Sobotta, Joana Warnecke, Thomas M Deserno. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Video-Based Motion Capture Smartphone Apps for Testing Human Motor Performance Skills: Scoping Review

Clara Sophia Zoeller[1*], MSc; Claudia Niessner[1*], PhD; Manuel Fleps[1], MSc; Thorsten Klein[2], MSc; Anke Hanssen-Doose[2], PhD; Alexander Burchartz[1], PhD; Alexander Woll[1], Prof Dr; Thorsten Stein[1], Prof Dr

[1]Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Engler-Bunte-Ring 15, Karlsruhe, Germany

[2]Institute of Movement and Sport, Karlsruhe University of Education, Karlsruhe, Germany

[*]these authors contributed equally

Corresponding Author:
Clara Sophia Zoeller, MSc
Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Engler-Bunte-Ring 15, Karlsruhe, Germany

Abstract

Background: Good motor performance skills (MPS) are relevant in all stages of life. Higher MPS are associated with enhanced cognitive abilities and physical and mental health. The assessment of MPS is important to identify deficits in MPS at an early stage and to implement interventions to address these deficits. One method to assess MPS is through marker-based motion capture in a laboratory setting with multiple cameras. However, this approach is expensive and time-consuming, making it impractical, for example, in large-scale studies for MPS assessment. Recent advancements (eg, artificial intelligence) in technology (eg, smartphone cameras) have opened up innovative solutions for various challenges (eg, testing large sample sizes). A potential solution is using video-based smartphone apps to assess MPS through markerless motion capture with a single camera.

Objective: The objectives of this scoping review were to summarize existing smartphone apps designed to digitally assess MPS through motion capture, identify the target population of the apps, determine whether the apps have been validated, and summarize the specific MPS that were assessed.

Methods: The scoping review was conducted in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analysis extension for Scoping Reviews) guidelines. The search was conducted in March 2024 using PubMed, Scopus, SPORTDiscus, Web of Science, Education Resources Information Centre, and SAGE Publications. All included studies investigated video-based motion capture smartphone apps to assess MPS.

Results: A total of 10 studies met the inclusion criteria. Seven different smartphone apps were used within the studies, 6 of which have already been validated. The MPS assessed through the apps were gait, breaststroke, running, countermovement jump, and shoulder mobility, and 1 study assessed a functional movement test battery. The studied populations were healthy adults, older adults, athletes, or individuals with neurological illnesses.

Conclusions: The assessment of MPS through smartphone apps represents a promising tool, which can be used in a variety of fields, such as health and performance monitoring, coaching, and scientific research. In the future, more studies should focus on developing new smartphone apps to assess different MPS and validate these apps.

(JMIR Mhealth Uhealth 2026;14:e65474) doi:10.2196/65474

KEYWORDS

smartphone app; motion capture; motor performance skills; motor assessment; automated video assessment; AI-based motion capture; kinematics; artificial intelligence; mobile health; scoping review; physical health; mental health; neurological illnesses; healthy adults; older; athletes

Introduction

The comprehensive development of motor performance skills (MPS) is critical for healthy human development [1] and remains important throughout all stages of life [2]. Overall, studies have shown that higher MPS are associated with increased physical activity and physical fitness at all ages [3-5], which have both physical and mental health benefits (eg, reduced depressive symptoms) [3,6]. Furthermore, there is an inverse relationship between MPS levels and body weight [5,7,8].

Given the multitude of terms used in motor performance research, it is important to select a definition of the term “motor performance skills” that will be used in this study. In their scoping review, Sortwell et al [9] established a general definition of the MPS term based on existing descriptions in the literature. MPS encompass a range of movements, including those used for space coverage (eg, walking), object manipulation (eg,

XSL•FO
RenderX

throwing), and movements that are needed for overcoming obstacles in vertical, diagonal, or horizontal directions (eg, climbing and jumping). Additionally, they include different kinds of pulling, pushing, and holding. These MPS serve as the foundation for specialized movement and the acquisition of sport-specific competencies [9]. The term MPS is often used interchangeably with the terms "motor skills" and "fundamental movement skills" [9,10].

The use of motor performance test batteries to assess motor performance allows for a precise examination of the current state and development of MPS. The results of these test batteries can be used to target interventions aimed at reducing deficits and improving performance [1,11,12]. However, assessing MPS using test profiles only provides limited or mainly subjective information regarding execution errors or the quality of the movement performed. A quantitative biomechanical measurement of kinematic variables collected via motion tracking systems could serve as an alternative or complementary approach [13,14].

The current gold standard for biomechanical movement assessment is using marker-based 3D motion tracking systems in a laboratory environment [14,15]. However, this method is time-consuming, expensive, and complex to implement, limiting its application in large-scale population studies [15-17]. To address these limitations, some studies have validated markerless 2D motion capture systems by comparing them to the gold standard. For example, a recent study showed a very high agreement between the gold standard and a single smartphone camera in a countermovement jump analysis, where videos were evaluated by Sbsq-pose (Subsequent GmbH), an artificial intelligence (AI)–based motion tracking system using deep learning to predict the 3D positions of 24 skeletal key points from 2D video recordings [18]. Indeed, the use of AI has gained increasing attention in biomechanics [19]. Today, there are pose estimation models that can be used to analyze video recordings of human movement [19]. These pose estimation models are created by deep learning algorithms, which are capable of identifying anatomical landmarks from digital videos [13]. To ensure that the assigned anatomical landmarks are correctly recognized, these human body models are first trained on large datasets [13]. Human annotators manually label the positions of anatomical landmarks in the training images [13,20]. Subsequently, the AI algorithm is able to automatically track human movements within recorded videos [13]. The trained pose estimation models also enable motion capture to be integrated into various everyday apps, such as smartphone apps [13,19]. There are already several smartphone apps that provide biomechanical monitoring, such as gait analysis, which provides information on spatiotemporal parameters (eg, cadence and foot strike), joint kinematics (eg, knee flexion angle) but not joint kinetics [21]. These apps enable automatic kinematic detection of whole-body and partial-body movements, including motion analysis that can be used in the field with minimal effort and even with large sample sizes. The advantage of using smartphone apps to collect MPS is that it is a simple and cost-effective way to perform biomechanical studies on a large population without the need for a lot of equipment or a laboratory environment, as the majority of individuals possess a smartphone [13]. In 2022, 6.4 billion individuals worldwide owned a smartphone, with an estimated 6.9 billion in 2024 and an estimated total of 7.7 billion by 2028 [22]. The widespread ownership of smartphones enables testing of large numbers of individuals, which is not feasible with a laboratory-based approach. The use of smartphone apps may be a particularly beneficial approach for population-based fitness studies, such as the MoMo study [23,24] in Germany or as those using the ALPHA test battery, which includes highly reliable and valid measures of cardiorespiratory fitness, muscular strength, and anthropometry [23-25]. For instance, a large-scale fitness study in European youth successfully collected data from nearly 8 million test results across 34 countries, encompassing a broad age range and using representative national data [25]. By using smartphone apps, a similar number of participants could be assessed on their fitness levels more efficiently. This approach minimizes the need for extensive travel and logistical resources, thereby enabling greater flexibility in test scheduling. By allowing participants to complete assessments remotely from their homes, it eliminates the requirement for in-person appointments or visits to testing facilities, thus streamlining the data collection process and enhancing accessibility for a broader population. However, it is imperative to acknowledge the limitations inherent to such apps, which often exhibit performance deficiencies. A number of factors may contribute to inaccurate movement detection by smartphone cameras. These include, for example, a low sampling rate, the presence of human-like objects in the room, and inaccurate camera positioning. These issues can result in the inability to precisely detect anatomical orientation points, which, in turn, can lead to measurement errors [13]. Furthermore, it should be noted that certain apps are designed to operate within a cloud computing environment, necessitating a stable internet connection for optimal functionality. This aspect should be considered when using such s.

To date, there is one review that gives an overview of video-based motion capture smartphone apps for MPS testing, but it only focuses on healthy adults [26]. Currently, no reviews provide a general overview for all population types.

This gap in the literature hinders the ability of researchers and practitioners to fully understand the capabilities and limitations of these technologies. Therefore, we conducted a scoping review to address this gap by providing a comprehensive overview of the current state of research on existing smartphone apps designed to assess the biomechanics of MPS through markerless motion capture. This review aimed to highlight the advancements, identify potential areas for improvement, and guide future research and development in this field.

We specifically addressed four aims: (1) to summarize existing smartphone apps designed to digitally assess MPS through motion capture, (2) to elucidate the specific MPS that these apps assess, (3) to determine whether these apps have been validated, and (4) to identify the target population of each app.

Methods

This scoping review was performed in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic

XSL•FO
RenderX

Reviews and Meta-Analysis extension for Scoping Reviews) guidelines (Checklist 1) [27]. A study protocol was designed in advance and then registered on OSF [28].

Eligibility Criteria

Inclusion Criteria

The studies included in the review had to perform motor performance tests that were assessed through video-based motion capture integrated into smartphone apps. The papers were included if they were written in English, published up to February 2024, involved human participants, and assessed motor performance through an AI-based video analysis (without human supervision).

Exclusion Criteria

All articles that used the smartphone's built-in sensors or accelerometers for the assessment of MPS were excluded from the analysis. We also excluded dissertations and conference papers. Meta-analyses, scoping reviews, and systematic reviews were not included in this scoping review but were considered separately in this paper as part of the basis for discussion and can be found in the "Meta-Analyses and Reviews" section.

Search Strategy

To identify potentially relevant articles, the search was conducted on March 5, 2024, in six electronic databases: PubMed, Scopus, SPORTDiscus, Web of Science, Education Resources Information Centre, and SAGE Publications. Before beginning the search, 2 reviewers (CSZ and CN) screened various publications and reviews in the research area, discussed the results, and determined the search term. The search terms included topics related to (1) MPS, (2) motion capture, and (3) smartphone apps, which were combined by Boolean operators. The search terms are shown in the Multimedia Appendix 1. The search was performed on the title and abstract of the papers, and the study type (validation study) was screened manually. There were no preset restrictions on the target population, so all ages were included. Before performing the search, the search strategy was reviewed by an expert (Dr Janis Fiedler) in mobile health technologies. Gray literature was not considered in this scoping review.

Selection of Sources of Evidence

To ensure a consistent and transparent selection process, a standardized screening form was developed by 2 reviewers (CSZ and CN) based on the central research objective: to provide a comprehensive overview of smartphone apps that assess MPS through markerless motion capture. The form was piloted using a small set of studies to calibrate inclusion decisions and refine key criteria such as app type, motion capture method, and degree of automation.

Study Selection

The search results were exported into the citation program Zotero (Corporation for Digital Scholarship). After removing duplicates, the titles and abstracts were screened by one reviewer (CSZ) using the open-source machine learning software "ASReview" (ASReview LAB developers) developed at Utrecht University [29]. To start training the model, 2 studies were initially selected for inclusion, and 2 studies were selected for exclusion [29]. Subsequently, the software presented 1 paper at a time, and the reviewer determined if it should be included or excluded based on its title and abstract. The title and abstract screenings were stopped after 50 consecutive articles were excluded by the reviewer [30,31]. Afterward, the second reviewer (CN) screened the articles excluded by the first reviewer to ensure that no relevant articles were falsely excluded. Following the title and abstract screening, the software provided a list of the included articles, which were then screened manually. If the full text was not available, authors were contacted and asked to provide it. The full-text screening was performed independently by 2 reviewers (CSZ and MF).

Data Charting and Extraction

To chart the data, 2 reviewers created a data charting form in Microsoft Excel to determine which variables should be extracted. Data extraction was then performed by 1 reviewer (CSZ) and verified by another author (TK). During the screening process, we found that information on the validation of the apps was more relevant than initially anticipated. Therefore, a fourth research objective addressing this aspect was added. This refers to whether the app was compared with another measurement instrument, such as a marker-based motion tracking system, regardless of whether the app was validated in the included study or in a previous study. The variables listed in the data chart included information about the authors (first author and year of publication), study design, study aims, population (sample size, age, and gender), name of the smartphone app, assessed MPS, if the app has been validated, and the main findings of the studies. The studies were grouped according to the smartphone app and the MPS studied.

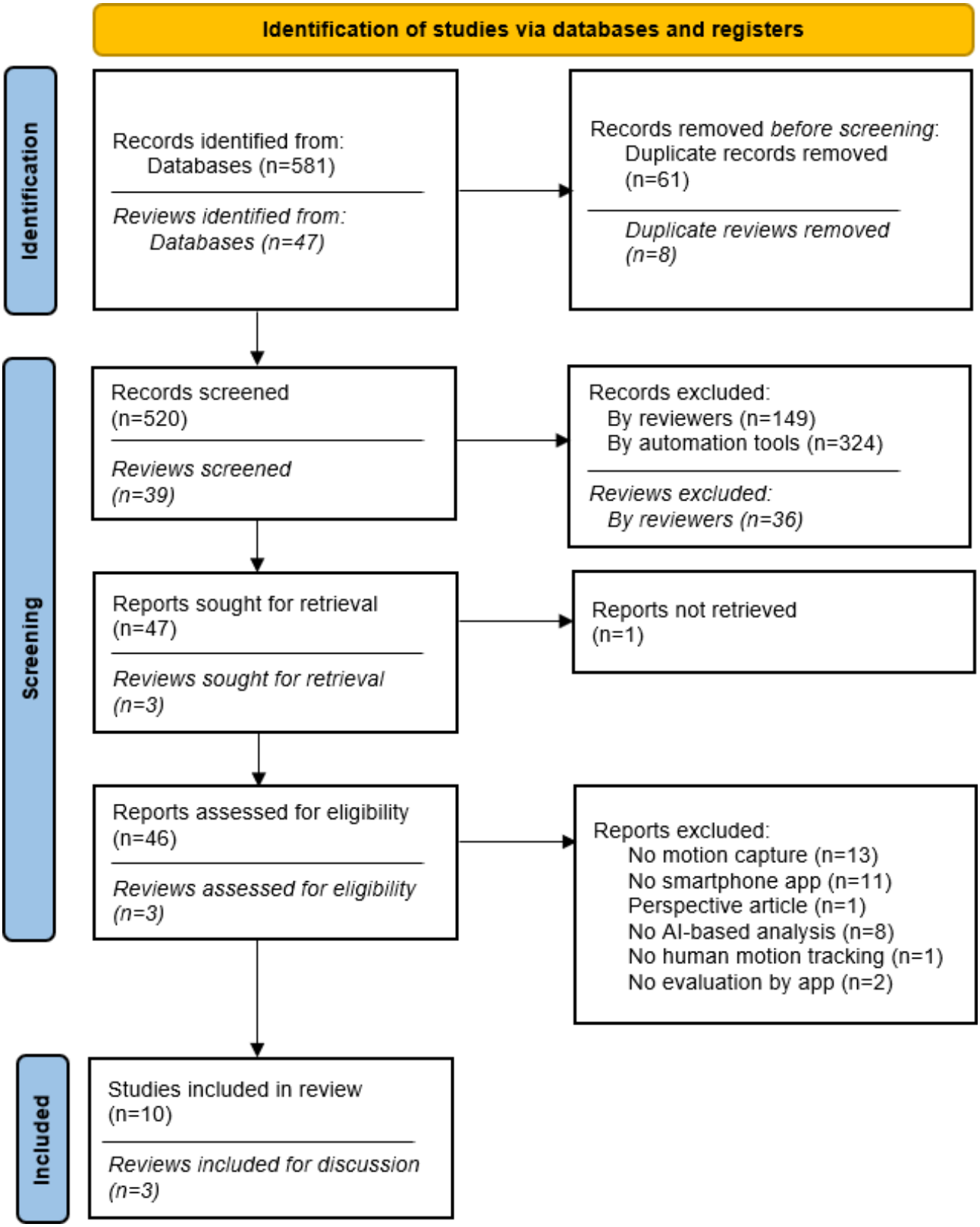
Results

Study Selection

A total of 581 articles were identified by searching the electronic databases. After removing duplicates, 520 (89.4 %) articles were included in the title and abstract screening in ASReview. The number of articles was subsequently reduced to 47 (8.1 %), with 1 article being excluded due to unavailability. Therefore, 46 (7.9 %) articles were included in the full-text screening. On the basis of the full-text screening, a total of 10 (1.7 %) articles were included in this scoping review. A detailed overview of the selection process is shown in Figure 1.

XSL•FO
RenderX

Figure 1. Flow diagram of search strategy. Italicized words refer to the selection of separately considered meta-analyses and reviews. AI: artificial intelligence.



Study Characteristics

The included articles were published between 2021 and 2024, with the most articles published in 2022 (n=3) and 2023 (n=4). Four of the studies were conducted in Japan [32-35]. The remaining studies were conducted in Australia [36], Germany [37], Spain [38], the United Kingdom [21], and the United States [39]. One study did not indicate the country in which it was conducted [40]. Research designs of the studies were mainly validation studies (n=5) [21,32,36,37,39], 1 proof-of-concept study [38], 2 cross-sectional studies [34,40], and 2 prospective, observational studies [33,35] (Table 1).

XSL•FO
RenderX

Table . Study characteristics.

Study	Study type	Participants: sample size; sex (male or female); age (y), mean (SD)	Smartphone app; pose estimation algorithm	MPS[a]	Validated
Aoyagi et al [32]	Validation study	• N/A[b]	TDPT-GT[c]; modified Resnet34 image classification model	Gait	Yes[d]
Azhand et al [37]	Validation study	• N=44 healthy older adults; N/A; 73.9 (6.0)	N/A; series of algorithms: Convolutional Pose Machine 25 body joint model, VNect model, and GASTNet 3D uplifting model	Gait	Yes
Balsalobre-Fernández [38]	Proof-of-concept case study and validation study	• N=1 healthy physically active man; n=1 male and n=0 female; 36	My Jump Lab; Apple Vision framework	CMJ[e]	Yes
Fanton et al [39]	Validation study	• N=150 (n=113 healthy adults, n=17 athletes, and n=20 MI[f] adults); n=56 males and n=94 females; 18-85	Amazon Halo Movement; N/A	Single leg stance, forward lunge, overhead squat, feet together squat, overhead reach	Yes
Feng et al [40]	Cross-sectional study	• N=8 professional swimmers; n=8 males and n=0 females; 21.8 (2.6)	N/A; series of algorithms: Gaussian filtering, Canny operator, double threshold method, and Hough transformation	Breaststroke	No
Iseki et al [35]	Prospective and observational study	• N=23 patients with PD[g]; n=13 males and n=10 females; 70.1 (6.0) • N=23 patients with iNPH[h]; n=16 males and n=7 females; 77.0 (6.4) • N=92 healthy adults; n=36 males and n=56 females; 72.3 (6.3)	TDPT-GT; Modified Resnet34 image classification model	Gait	Yes[d]
Iseki et al [34]	Cross-sectional study	• N=114 pathological patients (n=48 iNPH, n=21 PD, and n=55 other NMD[i]); n=52 males and n=62 females; 74.5 (7.8) • N=160 healthy adults; n=91 males and n=69 females; 72.9 (11.1)	TDPT-GT; Modified Resnet34 image classification model	Gait	Yes[d]

XSL•FO
RenderX

Study	Study type	Participants: sample size; sex (male or female); age (y), mean (SD)	Smartphone app; pose estimation algorithm	MPS[a]	Validated
van den Hoorn et al [36]	Validation study	• N=20 healthy adults; n=10 males and n=10 females; 36 (13)	mymobility App; Apple Vision framework	Shoulder abduction, adduction, flexion, and extension	Yes
Yamada et al [33]	Prospective and observational study	• N=15 healthy adults; n=9 males and n=6 females; 39.1 (20.1) • N=92 older adults; n=36 males and n=56 females; 73.0 (6.3) • N=47 patients with iNPH; n=32 males and n=15 females; 77.3 (6.3)	TDPT-GT; Modified Resnet34 image classification model	Gait	Yes[d]
Young et al [21]	Validation study	• N=31 healthy, experienced runners; n=20 males and n=11 females; 34.5 (9.7)	N/A; BlazePose	Running	Yes

[a]MPS: motor performance skills.

[b]N/A: not applicable.

[c]TDPT-GT: Three-Dimensional Pose Tracker for Gait Test.

[d]Preliminary validation indicates an initial comparison with another.

[e]CMJ: countermovement jump.

[f]MI: movement impaired.

[g]PD: Parkinson disease.

[h] iNPH: idiopathic normal pressure.

[i]NMD: neuromuscular diseases.

MPS and Smartphone Apps

The following results are presented in accordance with the initial 3 aims of this review and are summarized descriptively to provide an overview of the identified studies (Table 1, Multimedia Appendix 2).

Gait

In the studies included in this review, gait was the most commonly assessed MPS (n=5). Four of the studies used the Three-Dimensional Pose Tracker for Gait Test (TDPT-GT) smartphone app to assess this MPS [32-35]. Aoyagi et al [32] developed the TDPT-GT app to quantitatively assess pathological gait patterns and performed a preliminary validation against a Vicon 3D motion tracking system. The app uses a modified ResNet34 image classification model to detect the coordinates of 24 anatomical key points and calculate the 3D angles of the lumbar, bilateral hip, neck, knee, and ankle joints. To conduct the test, the subjects were instructed to walk in a circle with a diameter of 1 m, both clockwise and counterclockwise [32-35]. The app is capable of reconstructing 3D full-body movements from 2D motion capture data [32].

The other 3 studies examined the gait patterns of idiopathic normal pressure hydrocephalus (iNPH), Parkinson disease, and other neuromuscular diseases, as well as older adults and healthy adults [33-35]. Two studies by Iseki et al [34,35] compared pathological gait patterns with healthy gait patterns.

The final study included in the review used the TDPT-GT app to determine distinct indices for three pathological gait patterns: shuffling gait, wide-based gait, and short-stepped gait [33].

The fifth study developed a new smartphone app for gait assessment and examined the validity and repeatability of the app (the app name was not mentioned) [37]. To test the validity of the smartphone app, a comparison with the GAITRite System was performed. The app includes a series of pose estimation algorithms starting with the Convolutional Pose Machine 25 body joint model, followed by the VNect model and the GASTNet 3D uplifting model, to extract four gait parameters: gait speed, step length, step time, and cadence.

XSL•FO
RenderX

Running

One study developed an unnamed smartphone app that uses the deep learning pose estimation tool BlazePose (MediaPipe framework, Google LLC) to assess running gait in experienced runners [21]. To validate the app, the researchers simultaneously recorded the participants running on a treadmill with an iPhone 13 smartphone camera and a Vicon 3D motion tracking system [21].

Countermovement Jump

The study of Balsalobre-Fernández [38] used the Apple Vision framework (Apple Inc.), integrated into the My Jump Lab app, to assess countermovement jump (CMJ) height. The app was validated by comparing the results from My Jump Lab to those recorded simultaneously using a force plate.

Shoulder Mobility

One study performed shoulder range of motion tests by using the mymobility app, which uses the Apple Vision framework (Apple Inc.) for pose estimation [36]. For validation, participants performed shoulder abduction, adduction, flexion, and extension while being filmed with an iPhone 13s camera, with a sample rate of 30 frames per second, and a marker-based Vicon motion tracking system for comparison.

Functional Movement Test Battery

The study of Fanton et al [39] validated the Amazon Halo Movement app (Amazon Ltd) by comparing it to validated functional movement tests (FMTs) and sensor-collected kinematic variables. The FMT battery of the smartphone app was performed by participants wearing a full-body suit equipped with motion sensors. The app includes 5 FMTs, which are used to assess posture, stability, and mobility (single leg balance, forward lunge, overhead reach, feet together squat, and overhead squat) [39].

Breaststroke

The study of Feng et al [40] used a nonvalidated smartphone app to assess breaststroke technique. The app records the swimmer's movement in the water through a smartphone camera and calculates the angles between the thigh and trunk, thigh and calf, and the trunk and the horizontal plane. The video processing integrated into the app starts with Gaussian filtering, followed by the calculation of gradient amplitude and direction with a Canny operator, double threshold method for edge closure, and, finally, the detection of a straight line by Hough transformation [40].

Population

Most studies used the apps to assess the MPS of healthy adults (n=6) [33,35,36,38,39], 3 studies included athletes (general athletes [39], swimmers [40], and runners [21]), 2 studies included older persons [33,37], and 4 studies included clinical populations, that is, Parkinson disease [34,35], iNPH [33-35], movement impairment [39], and other neuromuscular diseases [34]. One of the studies did not provide any information about their study population [32] (Multimedia Appendix 3).

Synthesis of the Findings

The included studies demonstrate a growing interest in smartphone-based digital assessment tools for MPS, with gait analysis being the most frequently targeted skill, especially in clinical contexts such as Parkinson disease and iNPH (Multimedia Appendix 4). Despite the variety of apps, most tools have been validated through intraclass correlation coefficient or gold standard comparisons such as marker-based 3D motion capture. Deep learning algorithms, particularly modified ResNet34, dominate technical implementation. However, there are clear research gaps: few studies address upper body skills, and most studies focus on adults in controlled settings rather than broader, general-use or preventive contexts. These insights highlight future needs for diversified and validated apps, including those for different target groups (eg, children), and increased attention to real-world usability and accessibility (Table 2).

XSL•FO
RenderX

Table . Overview table of key findings.

Category and subcategory	Studies, n (%)	Validated, n (%)	Example applications or notes	Studies included
Motor performance skill				
Gait	5 (50)	5 (50)	TDPT-GT[a], CPM[b]+VNect	Aoyagi et al [32]; Azhand et al [37]; Iseki et al [34,35]; Yamada et al [33]
Running	1 (10)	1 (10)	BlazePose	Young et al [21]
CMJ[c]	1 (10)	1 (10)	My Jump Lab	Balsalobre-Fernández [38]
Shoulder mobility	1 (10)	1 (10)	mymobility	van den Hoorn et al [36]
Functional movement battery	1 (10)	1 (10)	Amazon Halo Movement	Fanton et al [39]
Swimming (breaststroke)	1 (10)	—[d]	Series of algorithms: Gaussian filtering, Canny operator, double threshold method, Hough transformation	Feng et al [40]
Target population				
Healthy adults	6 (60)	—	Most common group	Balsalobre-Fernández [38]; van den Hoorn et al [36]; Young et al [21]; Azhand et al [37]; Iseki et al [34,35]; Yamada et al [33]
Athletes	3 (30)	—	Swimmers, runners, general athletes	Young et al [21]; Feng et al [40]; Fanton et al [39]
Older adults	2 (20)	—	Often in gait-related studies	Azhand et al [37]; Yamada et al [33]
Clinical populations	4 (40)	—	PD[e], iNPH[f], MI[g], NMD[h]	Iseki et al [34,35]; Yamada et al [33]; Fanton et al [39]
Validation methods				
ICC[i]	6 (60)	—	My Jump Lab, BlazePose, CPM-based apps	Azhand et al [37]; Balsalobre-Fernández [38]; van den Hoorn et al [36]; Young et al [21]; Fanton et al [39]
AUC[j]	3 (30)	—	TDPT-GT for neurological assessment	Iseki et al [34,35]; Yamada et al [33]
Motion capture comparison	4 (40)	—	Vicon system	Balsalobre-Fernández [38]; van den Hoorn et al [36]; Young et al [21]; Azhand et al [37]
No formal validation	1 (10)	—	Swimming app	Feng et al [40]
Algorithms and frameworks				
Modified ResNet34	4 (40)	—	Clinical gait tracking (TDPT-GT)	Aoyagi et al [32]; Iseki et al [34,35]; Yamada et al [33]
Apple Vision Framework	2 (20)	—	My Jump Lab, mymobility	Balsalobre-Fernández [38]; van den Hoorn et al [36]
BlazePose	1 (10)	—	Running gait analysis	Young et al [21]
CPM+VNect+GASTNet	1 (10)	—	Multistep pose extraction for gait	Azhand et al [37]
Algorithm series	1 (10)	—	Hough/Canny in swimming	Feng et al [40]

[a]TDPT-GT: Three-Dimensional Pose Tracker for Gait Test.

[b]CPM: convolutional pose machine.

[c]CMJ: countermovement jump.

XSL•FO
RenderX

[d]Not applicable.
[e]PD: Parkinson disease.
[f]iNPH: idiopathic normal pressure hydrocephalus.
[g]MI: movement impairment.
[h]NMD: neuromuscular disease.
[i]ICC: intraclass correlation.
[j]AUC: area under the curve.

Meta-Analyses and Reviews

Selection of Meta-Analyses and Reviews

As meta-analyses, scoping reviews, and systematic reviews were not included in the scoping review, they were excluded from the screening process conducted by ASReview and were screened manually (CSZ and CN). A detailed overview of the screening is also found in Figure 1, shown in italics. To gain a more comprehensive understanding of the research field and to provide a basis for discussion, the inclusion criteria were considered more generously here. As a result, reviews were included that consider pose estimation methods without apps or apps that require a human operator to assess MPS.

Findings From Meta-Analyses and Reviews

We did not find any meta-analyses, reviews, or similar studies analyzing video-based motion capture smartphone apps. Using our search terms, we identified 47 meta-analyses and reviews, of which 3 have been included. However, one of the reviews included studies that examined apps that require manual operation, while the others included studies on pose estimation but without the use of smartphone apps. The 3 reviews present the respective findings relevant to specific subareas of video-based motion capture smartphone apps, which justified their inclusion in the evaluation.

The systematic review by Silva et al [41] focused on smartphone apps for capturing strength, power, change of direction, and velocity assessment. The included studies analyzed a total of 11 different smartphone apps, but 10 of them required a human operator to manually perform the assessments. Only the included study by Balsalobre-Fernández et al [42] used a video-based motion capture app (My Lift app). However, the app focused on tracking the barbell used during the snatch motion, rather than on the human movement itself. Nonetheless, this review provides valuable insights for an overview of the research and supports our aim of identifying existing apps, even if most of them require manual operation.

Furthermore, the review by Stenum et al [13] provides a general overview of the research on AI-based human pose estimation and its application in monitoring MPS, but without the use of smartphone apps. This review presents the extensive range of applications for these estimation tools, which are used in monitoring human development, in the motor assessment of neurological diseases, and in performance optimization. The advantages of these tools include a cost-effective and widely accessible alternative to motion capture systems in the laboratory [13].

In light of the growing use of pose estimation software in AI-based smartphone apps, the review by Stenum et al [13] was included to provide a more comprehensive understanding of the field.

Another review examined various technologies for markerless motion capture in combination with pose estimation in clinical surveys [43]. The review included a total of 65 studies, which, in turn, identified 9 different technical devices for markerless motion capture. The 3 most commonly used technologies were the Microsoft Kinect system, cameras such as GoPro, and smartphone video recordings. The review offers a comprehensive overview of the current state of research in this field and highlights the potential of video-based motion capture as a screening tool [43].

Discussion

Principal Findings

Video-based motion capture smartphone apps enable automatic kinematic detection of whole-body and partial-body movements, including motion analysis that can be used in the field with minimal effort and even with large sample sizes. To provide an overview of the current state of research on such smartphone apps, the scoping review had four aims: (1) to summarize existing smartphone apps designed to digitally assess MPS through motion capture, (2) to elucidate the specific MPS that these apps assess, (3) to determine whether these apps have been validated, and (4) to identify the target population of each app.

The findings of this scoping review indicate that there is a variety of smartphone apps that enable the recording and evaluation of MPS through video-based motion capture. Here, “gait” was the most frequently assessed MPS in the included studies, which was mainly recorded using the TDPT-GT app [32-35]. A total of 7 different smartphone apps were used in the aforementioned studies. Six of these apps have already been validated by comparing them with the corresponding gold standard methods. The smartphone apps identified here only include healthy adults, older adults, and individuals with neurological diseases as the target population. Children and adolescents have not yet been the focus of such smartphone apps, but it would be important to develop apps for this target group, for example, to identify motor development disorders in this target group at an early stage to provide appropriate interventions.

As demonstrated by the publication years of the articles, the research field of automatically assessing MPS with the use of video-based smartphone apps is a relatively novel and expanding area. The first publications in this field appeared in 2021 [37].

There are several advantages to assessing MPS through video-based motion capture smartphone apps. One advantage

XSL•FO
RenderX

is that it represents a more cost-effective alternative to laboratory-based systems, such as marker-based 3D motion capture or direct observation by trained professionals and offers the advantage that it can be performed in any location [14,16,17]. Furthermore, the pose estimation software integrated into the app does not require the attachment of markers on the participant, making it less time-consuming than marker-based motion tracking systems [14,17]. Additionally, some groups of people, such as older adults or those with disabilities, may be unable to tolerate the lengthy process of marker attachment and subsequent measurement [35].

These apps offer versatile use in coaching, health and performance monitoring, and scientific research. In scientific research, they can be used to perform testing outside of the laboratory setting and when no expensive motion capture system is available [13]. In the field of health monitoring, these applications would enable the testing of patients at home, as exemplified by the TDPT-GT app [32-35] included in this review. The advantage of testing at home is that patients are not required to visit clinics, physicians' offices, or physiotherapists as frequently [44]. Furthermore, these smartphone apps enable patients to be screened more frequently, which, on the one hand, relieves the burden on physicians and, on the other hand, allows the identification of developmental disorders or other impairments at an earlier stage, thus enabling timely intervention [13]. Additionally, it enables patients to be screened in their everyday environment, which is of great relevance depending on the clinical picture [44]. However, it must be considered that there is a nonnegligible challenge in the use of smartphone apps in epidemiologic health studies, which is the consideration of data sensitivity and privacy issues [45]. These applications often collect and process sensitive health data, which must be protected to ensure participants' privacy. Robust data encryption, secure data storage, and strict access controls are essential to safeguard this information. Furthermore, clear communication with participants about data usage, consent, and their rights is imperative to maintain trust and compliance with data protection regulations. Ensuring these measures are in place will facilitate the ethical and effective use of smartphone apps in health research [45].

In addition to the frequent screening of patients with impairments, the applications can also be used to monitor the athletic performance, as demonstrated in the study of Feng et al [40]. The advantage of the used pose estimation software integrated into the apps is that it enables real-time motion tracking and therefore can provide real-time feedback [13]. This is a valuable addition to the feedback provided by the coach, and it allows for a more frequent performance monitoring, as it enables monitoring during training sessions when the coach is not present. The potential for real-time motion tracking and feedback represents a valuable addition to the field of health monitoring. To ensure these apps deliver reliable and valid results, further development and validation against gold standard methods are essential [13].

Another advantage of using smartphone apps that have integrated AI for pose estimation is that they are less prone to inaccuracies than apps that require an operator to estimate, for example, the start and stop frames of a movement and then set these events manually [41,44].

As the technological revolution progresses, the potential of smartphone apps in scientific research, health monitoring, and athletic performance becomes increasingly evident.

Limitations

Despite the benefits of pose estimation, the following limitations must be considered when using these algorithms. For example, wearing loose clothing, such as dresses, can cause anatomical landmarks to be misidentified or obscured by other parts of the body. Environmental factors, such as lighting conditions, background clutter, or camera positioning, can also significantly affect the accuracy of pose estimation [46]. Variability in these external factors may introduce inconsistencies in results, especially in field-based assessments where standardization is difficult to maintain. It should also be noted that cameras with low sampling rates may not be able to correctly capture fast movements or that the training data used to train the algorithms may be very different from the movements to be analyzed and therefore be prone to error [13]. Another salient problem concerning the training data is that the annotators may lack anatomical knowledge, which can compromise their ability to accurately identify anatomical landmarks. Open-source datasets frequently undergo labeling through the utilization of crowdsourcing methodologies. This method is prone to interindividual variability, as the annotation process is often subjective and may differ between individuals [47,48]. Furthermore, most current pose estimation algorithms are trained on 2D video data, which may introduce inaccuracies when estimating joint angles and body segment orientations in 3D space. While some approaches attempt to reconstruct 3D poses, these are often limited in precision compared to marker-based systems, particularly in complex, multiplanar movements or occluded positions [46]. For these reasons, it is not yet possible to completely replace marker-based 3D motion capture in the laboratory with smartphone apps, but they already enable valid testing and thus facilitate the testing of large sample sizes.

As this scoping review did not aim to evaluate the quality or magnitude of validation outcomes, no conclusions regarding the strength of validity across applications can be drawn.

From a methodological perspective, scoping reviews inherently carry limitations, such as potential selection bias and the lack of critical appraisal. Although we aimed to follow rigorous inclusion criteria, the decision to include or exclude studies may have introduced subjective bias. Furthermore, due to the rapidly evolving nature of pose estimation technologies, it is possible that some recently published or ongoing work was not captured at the time of our review.

Looking forward, we recognize that including articles written in multiple languages could enhance the comprehensiveness of future reviews. By broadening the scope to include non-English studies, we can ensure a more inclusive and thorough understanding of the available research on this topic.

XSL•FO
RenderX

Conclusions

To our knowledge, this is the first scoping review to provide a comprehensive overview of smartphone apps that assess MPS using video-based markerless motion capture. Our findings highlight a growing interest in this cutting-edge field, with a variety of apps already available. These apps demonstrate significant potential across multiple domains, including enhanced monitoring, accurate kinematic measurements, and real-time feedback.

Looking ahead, there is a clear need for the development of more smartphone apps that assess MPS through video-based motion capture. It is crucial that these apps undergo rigorous validation against gold standard systems, such as marker-based 3D motion tracking systems [14,15,49]. Moreover, it is essential to ensure these apps are tested across diverse populations, including children, adolescents, adults, older adults, and individuals with disabilities.

Future research should also explore the integration of these apps with other health monitoring technologies, such as wearable sensors and telehealth platforms, to create comprehensive health management systems. Additionally, leveraging these apps in remote and underserved areas could offer valuable health insights and improve access to medical care. As AI and machine learning continue to advance, the accuracy and functionality of these apps will be further enhanced, solidifying their roles as effective tools for health and performance monitoring.

In summary, video-based markerless motion capture apps on smartphones hold transformative potential for health and performance monitoring. Continued innovation and validation in this field could lead to more accessible, accurate, and comprehensive monitoring systems, thereby facilitating the work of scientists and the development of interventions.

Acknowledgments

The authors would like to thank Dr Janis Fiedler for his assistance in reviewing and refining our search strategy. During the preparation of this work, the authors used DeepL Write to improve the language and readability of the paper. After using this tool or service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

This work was developed in the research project ”COMO-Study“ (2023-2026). The national study on the impact of the COVID-19 pandemic on the physical and mental health and health behaviors of children and adolescents against the background of socioecological contexts in Germany is funded by the German Federal Ministry of Education and Research (funding reference number: 01UP2222).

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

Conceptualization: CN, CSZ
Screening: CSZ, CN, MF
Data extraction: CSZ, MF, TK
writing – original draft: CSZ, CN
writing – review and editing: CSZ, CN, TS, AH-D, AB
Supervision: TS, AW
All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search term.
[DOCX File, 14 KB - mhealth_v14i1e65474_app1.docx]

Multimedia Appendix 2
Key findings of the included studies.
[DOCX File, 16 KB - mhealth_v14i1e65474_app2.docx]

Multimedia Appendix 3
Heatmap of the motor performance skills (MPS) per target population and validated MPS.
[PNG File, 56 KB - mhealth_v14i1e65474_app3.png]

XSL•FO
RenderX

Multimedia Appendix 4
Heatmap of the algorithms and frameworks used in smartphone motor performance skill applications.
[PNG File, 35 KB - mhealth_v14i1e65474_app4.png]

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 17 KB - mhealth_v14i1e65474_app5.docx]

References

1. Barnett LM, Stodden D, Cohen KE, et al. Fundamental movement skills: an important focus. J Teach Phys Educ 2016;35(3):219-225. [doi: 10.1123/jtpe.2014-0209]
2. Voelcker-Rehage C. Motor-skill learning in older adults—a review of studies on age-related differences. Eur Rev Aging Phys Act 2008 Apr;5(1):5-16. [doi: 10.1007/s11556-008-0030-9]
3. Stodden D, Langendorfer S, Roberton MA. The association between motor skill competence and physical fitness in young adults. Res Q Exerc Sport 2009 Jun;80(2):223-229. [doi: 10.1080/02701367.2009.10599556] [Medline: 19650387]
4. Robinson LE, Stodden DF, Barnett LM, et al. Motor competence and its effect on positive developmental trajectories of health. Sports Med 2015 Sep;45(9):1273-1284. [doi: 10.1007/s40279-015-0351-6] [Medline: 26201678]
5. Lubans DR, Morgan PJ, Cliff DP, Barnett LM, Okely AD. Fundamental movement skills in children and adolescents. Sports Med 2010 Dec;40(12):1019-1035. [doi: 10.2165/11536850-000000000-00000] [Medline: 21058749]
6. Global status report on physical activity. : World Health Organization; 2022 URL: https://iris.who.int/server/api/core/bitstreams/8804f1b0-dbae-4e58-a251-36fd14dc7e02/content [accessed 2026-01-29]
7. Cattuzzo MT, Dos Santos Henrique R, Ré AHN, et al. Motor competence and health related physical fitness in youth: a systematic review. J Sci Med Sport 2016 Feb;19(2):123-129. [doi: 10.1016/j.jsams.2014.12.004] [Medline: 25554655]
8. Hulteen RM, Morgan PJ, Barnett LM, Stodden DF, Lubans DR. Development of foundational movement skills: a conceptual model for physical activity across the lifespan. Sports Med 2018 Jul;48(7):1533-1540. [doi: 10.1007/s40279-018-0892-6] [Medline: 29524160]
9. Sortwell A, Behringer M, Granacher U, et al. Advancing sports science and physical education research through a shared understanding of the term motor performance skills: a scoping review with content analysis. Int J Kinesiol Sports Sci 2022;10(3):18-27. [doi: 10.7575/aiac.ijkss.v.10n.3p.18]
10. Sortwell A, Newton M, Marinho DA, Knijnik J, Ramirez-Campillo R. Potential role of plyometric training in the development of motor performance skills: a narrative review. Kinesiol Rev 2022;11(2):158-170. [doi: 10.1123/kr.2021-0006]
11. Ulrich D. Test of Gross Motor Development: Examiner's Manual, 2nd edition: Pro-Ed; 2000. URL: https://openlibrary.org/books/OL3966132M/Test_of_Gross_Motor_Development
12. Payne VG, Isaacs LD. Human Motor Development: A Lifespan Approach: Routledge; 2017. URL: https://www.routledge.com/Human-Motor-Development-A-Lifespan-Approach/Payne-Isaacs/p/book/9781032697130 [accessed 2026-01-29]
13. Stenum J, Cherry-Allen KM, Pyles CO, Reetzke RD, Vignos MF, Roemmich RT. Applications of pose estimation in human health and performance across the lifespan. Sensors (Basel) 2021 Nov 3;21(21):7315. [doi: 10.3390/s21217315] [Medline: 34770620]
14. Mündermann L, Corazza S, Andriacchi TP. The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications. J Neuroeng Rehabil 2006 Mar 15;3(1):6. [doi: 10.1186/1743-0003-3-6] [Medline: 16539701]
15. Reinking MF, Dugan L, Ripple N, et al. Reliability of two-dimensional video-based running gait analysis. Int J Sports Phys Ther 2018 Jun;13(3):453-461. [Medline: 30038831]
16. Mauntel TC, Cameron KL, Pietrosimone B, Marshall SW, Hackney AC, Padua DA. Validation of a commercially available markerless motion-capture system for trunk and lower extremity kinematics during a jump-landing assessment. J Athl Train 2021 Feb 1;56(2):177-190. [doi: 10.4085/1062-6050-0023.20] [Medline: 33480993]
17. Eltoukhy M, Kuenze C, Oh J, Wooten S, Signorile J. Kinect-based assessment of lower limb kinematics and dynamic postural control during the star excursion balance test. Gait Posture 2017 Oct;58:421-427. [doi: 10.1016/j.gaitpost.2017.09.010] [Medline: 28910654]
18. Barzyk P, Zimmermann P, Stein M, Keim D, Gruber M. AI-smartphone markerless motion capturing of hip, knee, and ankle joint kinematics during countermovement jumps. Eur J Sport Sci 2024 Oct;24(10):1452-1462. [doi: 10.1002/ejsc.12186] [Medline: 39205332]
19. Ngali MZ, An WC. Artificial intelligence in biomechanics pose estimation and the applications in human-related engineering. 2023 Presented at: Proceedings of the 1st International Conference on Frontier of Digital Technology Towards a Sustainable Society; Jan 26-27, 2022. [doi: 10.1063/5.0125285]

XSL•FO
RenderX

20. Andriluka M, Pishchulin L, Gehler P, Schiele B. 2D human pose estimation: new benchmark and state of the art analysis. Presented at: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Jun 23-28, 2014; Columbus, OH, USA p. 3686-3693. [doi: 10.1109/CVPR.2014.471]
21. Young F, Mason R, Morris R, Stuart S, Godfrey A. Internet-of-things-enabled markerless running gait assessment from a single smartphone camera. Sensors (Basel) 2023 Jan 7;23(2):696. [doi: 10.3390/s23020696] [Medline: 36679494]
22. Number of smartphone mobile network subscriptions worldwide from 2016 to 2025, with forecasts from 2025 to 2028. Statista. 2024. URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2024-05-07]
23. Woll A, Kurth BM, Opper E, Worth A, Bös K. The "Motorik-Modul" (MoMo): physical fitness and physical activity in German children and adolescents. Eur J Pediatr 2011 Sep;170(9):1129-1142. [doi: 10.1007/s00431-010-1391-4] [Medline: 21318230]
24. Woll A, Klos L, Burchartz A, et al. Cohort profile update: the Motorik-Modul (MoMo) longitudinal study-physical fitness and physical activity as determinants of health development in German children and adolescents. Int J Epidemiol 2021 May 17;50(2):393-394. [doi: 10.1093/ije/dyaa281] [Medline: 33709121]
25. Ortega FB, Leskošek B, Blagus R, et al. European fitness landscape for children and adolescents: updated reference values, fitness maps and country rankings based on nearly 8 million test results from 34 countries gathered by the FitBack network. Br J Sports Med 2023 Mar;57(5):299-310. [doi: 10.1136/bjsports-2022-106176] [Medline: 36623866]
26. El-Rajab I, Klotzbier TJ, Korbus H, Schott N. Camera-based mobile applications for movement screening in healthy adults: a systematic review. Front Sports Act Living 2025;7:1531050. [doi: 10.3389/fspor.2025.1531050] [Medline: 40416048]
27. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
28. Zöller CS, Niessner C, Burchartz A. Video-based smartphone applications for motor performance testing: a scoping review - study protocol. 2024. URL: https://osf.io/pj42u/overview [accessed 2026-02-12]
29. van de Schoot R, de Bruin J, Schram R, et al. An open source machine learning framework for efficient and transparent systematic reviews. Nat Mach Intell 2021;3(2):125-133. [doi: 10.1038/s42256-020-00287-7]
30. Quan Y, Tytko T, Hui B. Utilizing ASReview in screening primary studies for meta-research in SLA: a step-by-step tutorial. Research Methods in Applied Linguistics 2024 Apr;3(1):100101. [doi: 10.1016/j.rmal.2024.100101]
31. Zhu EM, Buljac-Samardžić M, Ahaus K, Sevdalis N, Huijsman R. Implementation and dissemination of home and community-based interventions for informal caregivers of people living with dementia: a systematic scoping review protocol. BMJ Open 2022 Jan 31;12(1):e052324. [doi: 10.1136/bmjopen-2021-052324] [Medline: 35105624]
32. Aoyagi Y, Yamada S, Ueda S, et al. Development of smartphone application for markerless three-dimensional motion capture based on deep learning model. Sensors (Basel) 2022 Jul 14;22(14):5282. [doi: 10.3390/s22145282] [Medline: 35890959]
33. Yamada S, Aoyagi Y, Iseki C, et al. Quantitative gait feature assessment on two-dimensional body axis projection planes converted from three-dimensional coordinates estimated with a deep learning smartphone app. Sensors (Basel) 2023 Jan 5;23(2):617. [doi: 10.3390/s23020617] [Medline: 36679412]
34. Iseki C, Suzuki S, Fukami T, et al. Fluctuations in upper and lower body movement during walking in normal pressure hydrocephalus and Parkinson's disease assessed by motion capture with a smartphone application, TDPT-GT. Sensors (Basel) 2023 Nov 18;23(22):9263. [doi: 10.3390/s23229263] [Medline: 38005649]
35. Iseki C, Hayasaka T, Yanagawa H, et al. Artificial intelligence distinguishes pathological gait: the analysis of markerless motion capture gait data acquired by an iOS application (TDPT-GT). Sensors(Basel) 2023 Jul 7;23(13):6217. [doi: 10.3390/s23136217] [Medline: 37448065]
36. van den Hoorn W, Lavaill M, Cutbush K, Gupta A, Kerr G. Comparison of shoulder range of motion quantified with mobile phone video-based skeletal tracking and 3D motion capture-preliminary study. Sensors (Basel) 2024 Jan 15;24(2):534. [doi: 10.3390/s24020534] [Medline: 38257626]
37. Azhand A, Rabe S, Müller S, Sattler I, Heimann-Steinert A. Algorithm based on one monocular video delivers highly valid and reliable gait parameters. Sci Rep 2021 Jul 7;11(1):14065. [doi: 10.1038/s41598-021-93530-z] [Medline: 34234255]
38. Balsalobre-Fernández C. Real time estimation of vertical jump height with a markerless motion capture smartphone app: a proof-of-concept case study. Proc Inst Mech Eng P J Sports Eng Technol 2024:17543371241227817. [doi: 10.1177/17543371241227817]
39. Fanton M, Harari Y, Giffhorn M, et al. Validation of Amazon Halo Movement: a smartphone camera-based assessment of movement health. NPJ Digit Med 2022 Sep 6;5(1):134. [doi: 10.1038/s41746-022-00684-9] [Medline: 36065060]
40. Feng Z, Ren A, Chen Z, et al. Portable and feasible device to evaluate the technique of swimming practitioners. IEEE Access 2022;10:88093-88104. [doi: 10.1109/ACCESS.2022.3200743]
41. Silva R, Rico-González M, Lima R, Akyildiz Z, Pino-Ortega J, Clemente FM. Validity and reliability of mobile applications for assessing strength, power, velocity, and change-of-direction: a systematic review. Sensors (Basel) 2021 Apr 8;21(8):2623. [doi: 10.3390/s21082623] [Medline: 33918020]

XSL•FO
RenderX

42. Balsalobre-Fernández C, Geiser G, Krzyszkowski J, Kipp K. Validity and reliability of a computer-vision-based smartphone app for measuring barbell trajectory during the snatch. J Sports Sci 2020 Mar;38(6):710-716. [doi: 10.1080/02640414.2020.1729453] [Medline: 32079484]
43. Lam WWT, Tang YM, Fong KNK. A systematic review of the applications of markerless motion capture (MMC) technology for clinical measurement in rehabilitation. J Neuroeng Rehabil 2023 May 2;20(1):57. [doi: 10.1186/s12984-023-01186-9] [Medline: 37131238]
44. Ferrer-Mallol E, Matthews C, Stoodley M, et al. Patient-led development of digital endpoints and the use of computer vision analysis in assessment of motor function in rare diseases. Front Pharmacol 2022;13:916714. [doi: 10.3389/fphar.2022.916714] [Medline: 36172196]
45. The ongoing journey to commitment and transformation: digital health in the WHO european region. : World Health Organization, Regional Office for Europe; 2023 URL: https://cdn.who.int/media/docs/librariesprovider2/data-and-evidence/english-ddh-260823_7amcet.pdf?sfvrsn=4c674522_2&download=true [accessed 2026-01-29]
46. Wade L, Needham L, McGuigan P, Bilzon J. Applications and limitations of current markerless motion capture methods for clinical gait biomechanics. PeerJ 2022;10:e12995. [doi: 10.7717/peerj.12995] [Medline: 35237469]
47. Cronin NJ. Using deep neural networks for kinematic analysis: challenges and opportunities. J Biomech 2021 Jun 23;123:110460. [doi: 10.1016/j.jbiomech.2021.110460] [Medline: 34029787]
48. Needham L, Evans M, Cosker DP, et al. The accuracy of several pose estimation methods for 3D joint centre localisation. Sci Rep 2021 Oct 19;11(1):20673. [doi: 10.1038/s41598-021-00212-x] [Medline: 34667207]
49. Albert JA, Owolabi V, Gebel A, Brahms CM, Granacher U, Arnrich B. Evaluation of the pose tracking performance of the Azure Kinect and Kinect v2 for gait analysis in comparison with a gold standard: a pilot study. Sensors (Basel) 2020 Sep 8;20(18):5104. [doi: 10.3390/s20185104] [Medline: 32911651]

Abbreviations

AI: artificial intelligence
FMT: functional movement test
iNPH: idiopathic normal pressure hydrocephalus
MPS: motor performance skills
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analysis extension for Scoping Reviews
TDPT-GT: Three-Dimensional Pose Tracker for Gait Test

Edited by L Buis; submitted 16.Aug.2024; peer-reviewed by B Liew, J Simmich; accepted 08.Jan.2026; published 19.Feb.2026.

Please cite as:
Zoeller CS, Niessner C, Fleps M, Klein T, Hanssen-Doose A, Burchartz A, Woll A, Stein T
Video-Based Motion Capture Smartphone Apps for Testing Human Motor Performance Skills: Scoping Review
JMIR Mhealth Uhealth 2026;14:e65474
URL: https://mhealth.jmir.org/2026/1/e65474
doi:10.2196/65474

© Clara Sophia Zoeller, Claudia Niessner, Manuel Fleps, Thorsten Klein, Anke Hanssen-Doose, Alexander Burchartz, Alexander Woll, Thorsten Stein. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 19.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Digital Health Interventions in Children and Adolescents With Type 1 Diabetes Mellitus and Their Impact on Clinical and Behavioral Outcomes: Scoping Review

Eirena Beh[1,2], MBBS; Jiaying Lin[1,2], MBBS; Geraldyn Leong[1,2], MBBS; Pei Yi Loh[3]; Ee Jin Poh[3]; Daniel Chan[2,3,4,5], MBBS

[1]Department of Paediatric Medicine, KK Women's and Children's Hospital, 100 Bukit Timah Road, Singapore, Singapore

[2]Paediatrics Academic Clinical Programme, KK Women's and Children's Hospital, Singapore, Singapore

[3]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[4]Endocrinology Service, KK Women's and Children's Hospital, Singapore, Singapore

[5]Duke-NUS Medical School, Singapore, Singapore

Corresponding Author:
Eirena Beh, MBBS
Department of Paediatric Medicine, KK Women's and Children's Hospital, 100 Bukit Timah Road, Singapore, Singapore

Abstract

Background: Effective self-management of type 1 diabetes mellitus (T1DM) by children and adolescents remains challenging despite advances in insulin delivery and glucose monitoring technologies. Mobile health (mHealth) interventions have emerged as promising tools to support pediatric diabetes care. However, their clinical impact and the behavioral mechanisms through which they operate—particularly those grounded in social cognitive theory (SCT)—are not well established.

Objective: This scoping review assesses mHealth apps focused on the management of T1DM in the pediatric population and looks into the underlying behavioral frameworks in accordance with SCT.

Methods: We conducted a scoping review of 5 databases (PubMed, Cochrane Library, EMBASE, CINAHL, and Scopus) for English-language studies published between January 2000 and July 2024. Eligible studies evaluated mHealth apps for children and adolescents with T1DM (≤18 years) and reported outcomes including glycemic control, self-efficacy, adherence, self-management, or quality of life. Data were extracted and synthesized according to clinical outcomes and the presence of SCT constructs, namely self-efficacy, behavioral capability, expectations, reinforcements, and reciprocal determinism.

Results: Of 5607 studies screened, 12 met the inclusion criteria. These comprised 4 randomized controlled trials, 4 pilot studies, 2 pre-post intervention studies, 1 retrospective cohort study, and 1 double crossover trial. App features included glucose logging, insulin tracking, bolus calculators, reminders, gamification, and structured educational content. Hemoglobin A_{1c} (HbA_{1c}) outcomes were reported by 9 studies; 4 demonstrated statistically significant improvements, while the others reported stability or no change. Several studies also reported improvements in treatment adherence and perceived self-efficacy. Of the 12 studies, 11 incorporated at least 1 SCT construct, with most integrating behavioral capability and self-efficacy as core components. Interventions using multiple SCT constructs showed greater promise for supporting sustained behavior change.

Conclusions: mHealth apps for pediatric T1DM are complex behavioral interventions that often leverage key principles of SCT to promote effective self-management. Although the evidence supports modest benefits in glycemic control and behavioral outcomes, heterogeneity in study design and outcome measurement limits broader generalizability. Future research should prioritize the development and evaluation of SCT-informed digital interventions with standardized outcome frameworks to improve pediatric diabetes care.

(JMIR Mhealth Uhealth 2026;14:e79338) doi:10.2196/79338

KEYWORDS

mhealth; mobile health; digital health; eHealth; pediatrics; diabetes; diabetes mellitus; type 1 diabetes; pediatric diabetes

Introduction

Type 1 diabetes mellitus (T1DM) is a chronic autoimmune condition characterized by the destruction of insulin-producing pancreatic β-cells, leading to insulin deficiency and hyperglycemia. T1DM is among the 5 most common chronic illnesses in children attending school [1]. Findings observed from the SEARCH for Diabetes in Youth study [2] noted the prevalence of T1DM in youth aged 0 years to 19 years increased by 21.1% between 2001 and 2009, with an annual incidence

XSL•FO
RenderX

growth rate of 1.4% reported between 2002 and 2012. This increase was noted across various demographic and ethnic groups.

T1DM management carries a substantial burden—including clinical, cognitive, and psychosocial—for both children and their caregivers, requiring daily insulin administration; frequent blood glucose monitoring; precise dietary management; and ongoing adherence to a structured, health-oriented lifestyle. Advancements in technologies in the treatment of T1DM such as insulin pumps, continuous glucose monitoring (CGM) systems, and intermittently scanned CGM systems have significantly improved patients' metabolic control and, in turn, have helped to mitigate long-term diabetic complications [3].

Despite advances in insulin delivery systems and glucose monitoring technologies, achieving optimal glycemic control continues to rely heavily on effective self-management. Effective management requires strict and sustained discipline in regular glucose monitoring, carbohydrate counting, and adherence to an insulin regimen. This remains a long-term challenge, particularly among children and adolescents as they navigate emerging autonomy in the self-management of their diabetes [4]. With the advent of digital health, patients and health care providers can now leverage digital tools such as video conferencing, mobile health (mHealth) apps, and remote monitoring systems to improve accessibility to consistent health care, revolutionizing the management of T1DM. Recognizing its potential, the American Diabetes Association advocates that telehealth should complement in-person visits to achieve optimal glycemic management [5]. As such, exploring effective and innovative strategies that enhance self-management skills in the pediatric population with T1DM becomes critical. With the rapid advancement of mHealth technologies, mHealth apps have emerged as promising tools to support disease management by children and adolescents with T1DM. These digital interventions integrate features such as real-time reminders, interactive education, gamification, and data sharing to promote treatment adherence and self-management. Although several studies suggest that mHealth apps may improve glycemic control and psychosocial outcomes in pediatric T1DM, the extent and consistency of their clinical and behavioral impact remain unclear.

Social cognitive theory (SCT) [6,7] is a widely applied behavioral framework that explains how individuals acquire and maintain health behaviors through the dynamic interaction of personal factors, behavioral patterns, and environmental influences—a concept known as reciprocal determinism. Key constructs include self-efficacy, behavioral capability, expectations, and reinforcement, all of which are critical to sustaining self-management in chronic conditions such as T1DM. This scoping review aimed to synthesize the current evidence on mHealth interventions for pediatric T1DM, with a focus on their effects on key outcomes including hemoglobin A_{1c} (HbA_{1c}), self-efficacy, treatment adherence, and quality of life. In doing so, we also examined the extent to which SCT constructs are embedded within these interventions to better understand how digital tools are designed to promote behavior change and support sustainable improvements in pediatric diabetes care.

Methods

We conducted our scoping review with reference to the updated Joanna Briggs Institute guidance for scoping reviews. This was also aligned with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines for reporting scoping reviews [8]. The review protocol was preregistered on the Open Science Framework.

Data Sources and Search Strategy

We conducted a comprehensive literature search in 5 databases including PubMed, Cochrane Library, EMBASE, CINAHL, and Scopus. A broad search strategy was used and included the following terms as MeSH, EMBASE Subject Headings terms, major headings, or title or abstract keywords: ("diabetes mellitus") AND ("pediatrics" OR "child" OR "adolescent" OR "infant" OR "minor" OR "puberty" OR "schools")) AND ("mobile applications" OR "telemedicine" OR "telehealth" OR "mobile health").

Search results were limited to studies published in English between January 2000 and July 2024. This time frame ensures the inclusion of studies reflecting contemporary technological capabilities and behavioral frameworks relevant to mHealth interventions. The studies were screened and selected by 3 independent reviewers (EB, JL, and GL) and adjudicated by an independent fourth reviewer (DC).

Screening and Data Extraction

We used the web-based software Covidence [9] for title and abstract screening, full-text review, and subsequent data extraction. We extracted the following information about each study: author, year of publication, study design, sample size, type of mHealth intervention, and main outcome findings.

Eligibility Criteria

We included specific study designs such as randomized controlled trials, observational studies, and prospective interventional studies published between January 2000 and July 2024. We focused on studies reporting on the impact of mHealth apps on children and adolescents with T1DM. The definition for a pediatric population in this review was ≤18 years old.

We looked at important outcomes related to T1DM including HbA_{1c}, self-efficacy, adherence to therapy, T1DM self-management skills, and quality of life. T1DM self-management included carbohydrate counting and insulin dosing, treatment of hyperglycemia, and prevention and treatment of hypoglycemia. Quality of life measurements were assessed objectively using validated questionnaires, including items assessing satisfaction with lifestyle, impact on life and relationships, and future concerns.

We excluded feasibility trials, study protocols, design studies, commentaries, case reports, and studies that were not published in English.

XSL•FO
RenderX

Data Synthesis and Analysis

We conducted a thematic analysis using a deductive, semi-iterative approach guided by predefined constructs from SCT. Interventions were categorized and mapped to the 5 core SCT constructs: reciprocal determinism, behavioral capability, reinforcement, expectations, and self-efficacy. This analysis was performed by our team of reviewers independently (EB, JL, GL, PYL, and EJP), with disagreements first discussed and resolved among reviewers followed by further referral for adjudication by DC. We also assessed the reported impact of each intervention on the clinical and behavioral management of pediatric T1DM.

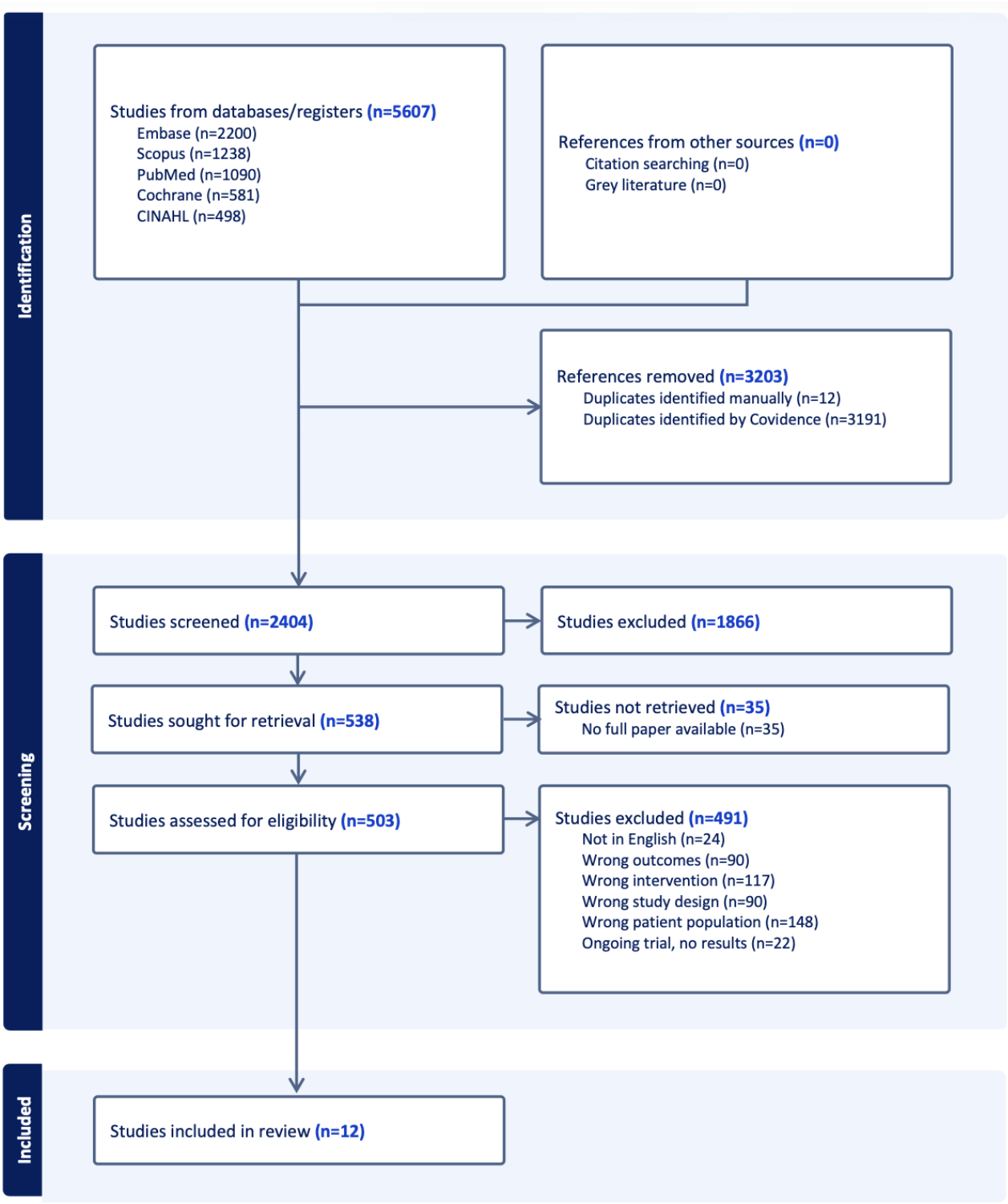
Results

Overview

Our database search, conducted in July 2024, yielded a total of 5607 studies. After removal of duplicates, there were 2404 studies. Through screening of title and abstracts, 1866 papers were excluded. A further 491 studies were excluded after full-text review. The screening and selection process is illustrated in the PRISMA flowchart (Figure 1).

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of study selection for the scoping review on mobile health (mHealth) interventions for pediatric type 1 diabetes mellitus.



We identified a total of 12 suitable studies. Of the 12 papers, 4 were randomized controlled trials, 4 were pilot studies, 2 were pre-post intervention studies, 1 was a retrospective cohort study, and 1 was a randomized double crossover study. Table 1 gives an overview of the included studies. Different diabetes mHealth apps were evaluated across the studies. App features included glucose level recording, insulin dose recording, feedback on glucose measurements, physical activity promotion, daily reminders, educational features, and reward systems. Some apps also combined multiple features on the platform.

XSL•FO
RenderX

Table . Study characteristics.

Domain	Author (year)	Title	Study design	Participants, n	Age (years), range	Outcome measures	App features
B[a], R[b], E[c], S[d]	Cafazzo et al (2012) [10]	Design of an mHealth[e] app for the self-management of adolescent type 1 diabetes: a pilot study	Pilot study	20	12 - 16	Frequency of daily blood glucose readings, HbA_{1c}[f], self-care inventory survey, and Diabetes Family Responsibility Questionnaire	Gamification incentives and blood glucose trend analysis
B, E, S	Nilsson (2016) [11]	How young people can learn about newly diagnosed type 1 diabetes	Prospective pilot study	15	6.5 - 16	Qualitative interview regarding knowledge and learning	App with educational content about injection technique, hyperglycemia, and hypoglycemia; diabetes game; an anatomy app; and a carbohydrate counter provided during the hospital stay
B, R, S	Goyal et al (2017) [12]	A mobile app for the self-management of type 1 diabetes among adolescents: a randomized controlled trial	RCT[g]	92	11 - 16	HbA_{1c}, hypoglycemic event, self-monitoring of blood glucose, and self-initiated adjustments to regimen	Out-of-range blood glucose trend alerts, coaching around out-of-range trend causes, and fixed point-based incentive system
B, S	Clifton et al (2018) [13]	A tablet-based educational tool: toward more comprehensive pediatric patient education	Prospective pilot study	106	6 - 17	Perceived diabetes knowledge	Education regarding insulin regime, doses, treatment of hyperglycemia, prevention and treatment of hypoglycemia, and to teach patients and caregivers
B, R, E, S	Klee et al (2018) [14]	An intervention by a patient-designed do-it-yourself mobile device app reduces HbA_{1c} in children and adolescents with type 1 diabetes: a randomized double-crossover Study	Randomized double crossover	55	10 - 18	HbA_{1c}, QoL[h], and hypoglycemia	Logging blood glucose readings, carbohydrate/sugar intake, and insulin dosing and regular, remote feedback from the diabetes care team

XSL•FO
RenderX

Domain	Author (year)	Title	Study design	Participants, n	Age (years), range	Outcome measures	App features
B, R, E, S	Pramanik et al (2019) [15]	Smartphone app as motivational intervention to improve glycemic control in adolescents with type 1 diabetes	Pre-post intervention study (single group)	34	11 - 18	HbA_{1c}	Daily reminders for insulin, meals, and exercise; thrice-daily reminders (morning, evening, bedtime) on weekdays; and an additional midday reminder on weekends for insulin administration, meal timing, and exercise
R	Hilliard et al (2020) [16]	Type 1 Doing Well: pilot feasibility and acceptability study of a strengths-based mHealth app for parents of adolescents with type 1 diabetes	RCT	80	12 - 17	HbA_{1c} (%) and qualitative interviews	Praise messages and tracking diabetes-related behaviors with a weekly summary report
B, S	Alfonsi et al (2020) [17]	Carbohydrate counting app using image recognition for youth with type 1 diabetes: pilot randomized control trial	RCT	46	8.5 - 17	HbA_{1c}, carbohydrate counting accuracy, counting errors, QoL for youth, self-care, and patient or parent responsibility	Food identification and carbohydrate counting
B, R, E, S	Holtz et al (2021) [18]	An mHealth-based intervention for adolescents with type 1 diabetes and their parents: pilot feasibility and efficacy single-arm study	Prospective pilot study	30 (adolescent-parent pairs)	10 - 15	Diabetes care adherence, QoL, family conflict, and HbA_{1c}	Parent-adolescent communication
B, R, S	Kime et al (2023) [19]	Evaluation of the DigiBete App, a self-management app for type 1 diabetes: experiences of young people, families, and healthcare professionals	Pre-post intervention study (single group)	1343	5 - 16	Qualitative interviews	Educational advice and resources such as guidance, quizzes, and educational and instructional videos; store insulin ratios/doses; and communication with the diabetes team via the app
B, S	Bender et al (2024) [20]	Impact of Sten-O Starter on glycemic management in children and adolescents with type 1 diabetes in the north region of Denmark	Retrospective cohort study	181	≤18[i]	HbA_{1c}, time in range, time above range, and time below range	Educational app (each topic with animation, quizzes, and games) and instruction videos

XSL•FO
RenderX

Domain	Author (year)	Title	Study design	Participants, n	Age (years), range	Outcome measures	App features
B, R, E, S	Holtz et al (2024) [21]	The effect of an mHealth intervention for adolescents with type 1 diabetes and their parents	RCT	33 (adolescent-parent pairs)	10 - 15	HbA_{1c}, self-management, Diabetes Behavior Rating scale, and QoL via Impact on Family Scale	Parent-child communication

[a]B: behavioral.

[b]R: reinforcement.

[c]E: expectations.

[d]S: self-efficacy.

[e]mHealth: mobile health.

[f]HbA_{1c}: hemoglobin A_{1c}.

[g]RCT: randomized controlled trial.

[h]QoL quality of life.

[i]Inclusion criterion for the study.

The results for each of the included outcomes are described in the following sections.

HbA_{1c}

HbA_{1c} values as an outcome measure were reported in 9 studies [10,12,14-18,20,21]. In these studies, there were 355 people in the intervention groups and 208 people in the control groups. Significant improvement in HbA_{1c} values (reduced HbA_{1c} values) were reported within the intervention groups in 4 studies (Alfonsi et al [17]: −0.35%, *P*=.03; Klee et al [14]: –0.33%, *P*=.048; Pramanik et al [15]: −0.91%, *P*=.004; Goyal et al [12] −0.58%, *P*=.02). Significant differences were not found in 3 studies [10,16,20]. HbA_{1c} stability in the intervention group (HbA_{1c}: −0.22%, *P*=.26) compared with worsening in the control group (HbA_{1c}:+0.52%, *P*=.02) was reported in 1 study [21].

Holtz et al [18] did not find a significant change in HbA_{1c} levels between the groups; however, a higher use of the mobile app correlated with a greater improvement in HbA_{1c} levels (*P*<.005).

Hypoglycemia

The occurrence of hypoglycemic events as an outcome measure was reported in 3 studies [12,20,21].

Bender et al [20] found a significant difference in the reduction of the time below range between the intervention and control groups at both 6 months (1.2% vs 2.6%, *P*=.02) and 12 months (1% vs 2%, *P*=.05) posttreatment. Klee et al [14] and Goyal et al [12] did not find any significant differences in hypoglycemic events.

Self-Efficacy

Self-efficacy as an outcome measure was reported in 3 studies [13,16,17].

Alfonsi et al [17] reported increased carbohydrate counting accuracy (*P*=.008) and reduced frequency of individual counting errors >10 g (*P*=.047).

Hilliard et al [16] used the Diabetes Strengths and Resilience Measure for Adolescents to assess how often patients engaged in T1DM-related behaviors, but found no significant differences between the groups.

Through a multiple-choice survey designed with a 5-point Likert-type scale, Clifton et al [13] assessed the change in caregivers' knowledge level for T1DM in general, hypoglycemia, hyperglycemia, and diabetic ketoacidosis. They found that 57% (36/63) of the intervention group perceived an increase in diabetes knowledge compared with 39% (16/41) of the control group, though the difference did not reach statistical significance (*P*=.059).

Treatment Adherence

The impact on treatment adherence was evaluated in 4 studies [10,12,18,21].

Holtz et al [18] and Hilliard et al [16] used the Diabetes Behavior Rating Scale, showing significant improvements in diabetes behavior 12 weeks postintervention (*P*=.02 in both studies).

Goyal et al [12] used the Self Care Inventory questionnaire to measure adherence to treatment recommendations but did not find any significant changes in scores. Similarly, Cafazzo et al [10] did not find any significant changes in Self Care Inventory scores but did report a 50% (2.4 to 3.6 per day) increase in the daily average frequency of blood glucose measurements (*P*=.006).

Quality of Life

Quality of life was measured through various questionnaires in 7 studies [10,12,16-19,21]. Alfonsi et al [17], Cafazzo et al [10], and Goyal et al [12] used the Diabetes Quality of Life for Youth questionnaire. Hilliard et al [16] used the Monitoring Individual Needs in Diabetes Youth Questionnaire. Klee et al [14] used the Quality of Life for Youth questionnaire, and Holtz et al [21] used the PedsQl questionnaire. None of the studies found significant differences between the intervention and control groups.

Holtz et al [18] reported significant improvement in quality of life measured via the PedsQoL generic scale (before study:

XSL•FO
RenderX

mean 4.02, SD 0.84; after study: mean 4.27, SD 0.73; t_{24}=2.48, P=.01, d=0.32).

Family Quality of Life and Family Conflict

The impact on quality of life in the family was evaluated in 5 studies [10,12,16,18,21].

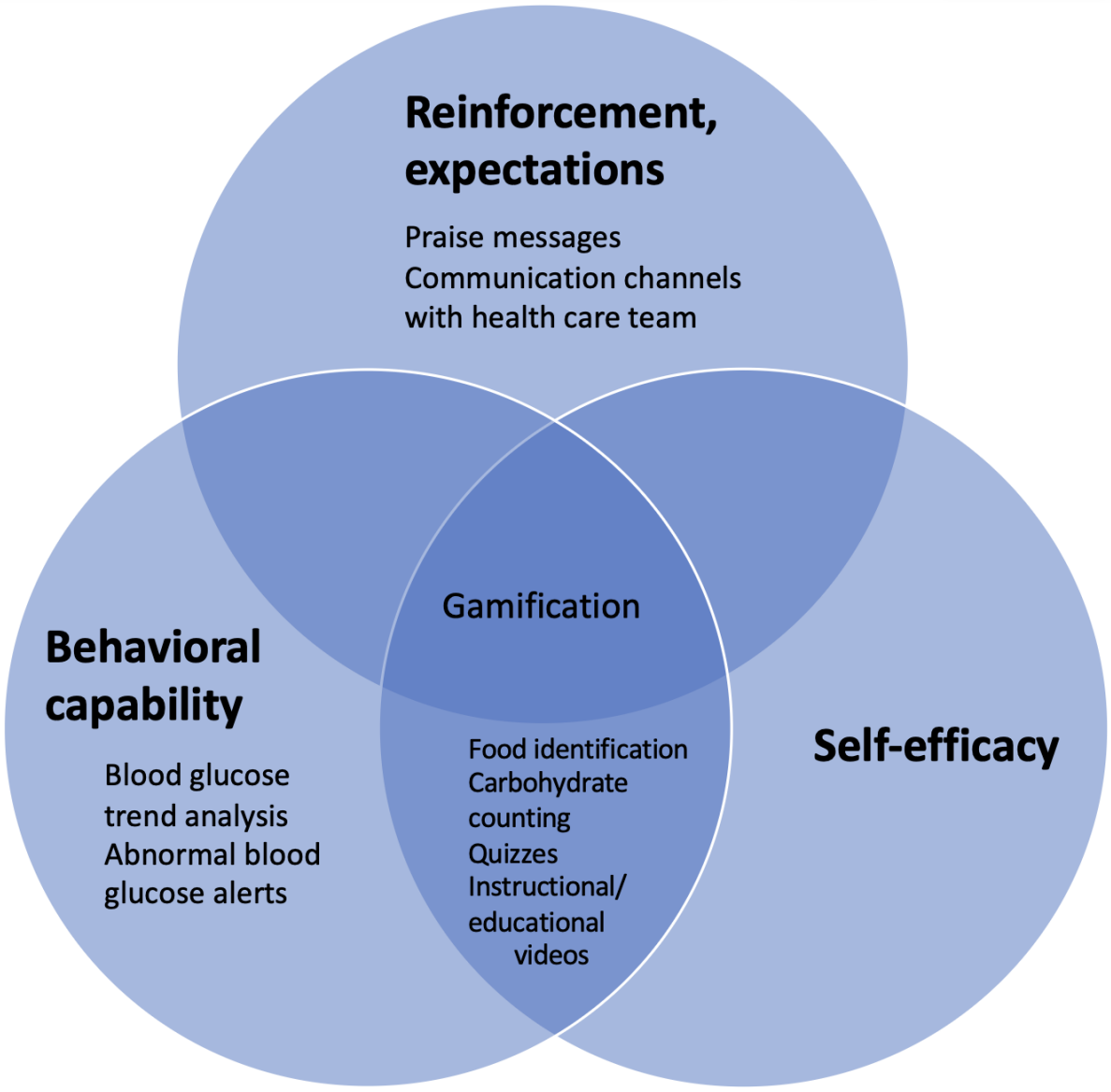
Holtz et al [21] used the Impact on Family Scale in their study, showing that parents' perceptions of quality of family life improved significantly for the intervention group (P=.02).

Hilliard et al [16] used the Diabetes Family Impact Scale, Holtz et al [18] used the Revised Diabetes Family Conflict scale, and Cafazzo et al [10] and Goyal et al [12] both used the Diabetes Family Responsibility Questionnaire to examine family conflict and parent-adolescent interaction around diabetes. However, no significant differences between intervention and control groups were found.

SCT Constructs Across the mHealth Apps

All included mHealth apps incorporated at least one SCT construct, with two-thirds of studies using at least 3 constructs within the same app. Among the individual SCT constructs, behavioral capability and self-efficacy were the most frequently used, appearing in 11 of the 12 included mHealth apps. Many of the apps had various tools: DigiBete [19] had the ability to store insulin ratios/doses and pump settings and guidance videos on the administration of injections, while WebDia [14] had a bolus calculator function with a section centered on carbohydrate estimation with pictorial assistance. In the iSpy study [17], the use of a mobile app designed to assist youth with T1DM with counting carbohydrates made food identification through images possible. Figure 2 illustrates the key features of the mHealth apps and their associated SCT construct(s).

Figure 2. Mapping of mobile health (mHealth) app features to core constructs of social cognitive theory in pediatric type 1 diabetes management.



Gamified features were used within the mHealth apps in 5 studies [10,11,18,20,21]. The Bant app [12] in particular was designed with a rewards algorithm that allocated game-like experience points for adherence to blood glucose testing with bonus points and level-ups awarded accordingly. A total of 161 rewards (an average of 8 rewards each) were distributed to

XSL•FO
RenderX

participants based on their frequency of blood glucose measurement, with 50% (10/20) of participants collecting more than 10 awards. The use of gamification and rewards encouraged adherence and reinforced positive health behavior.

The majority of the studies provided educational modules to participants, thereby enhancing their behavioral capability. Alfonsi et al [17] and Klee et al [14] focused on nutrition with food identification and carbohydrate counting, while other studies [11-13,19] included analysis of blood glucose trends and awareness of hypoglycemia and hyperglycemia. Some studies also included educational materials regarding insulin storage, calculation, and administration. The DigiBete [19] and Sten-O Starter [20] apps also made use of quizzes to help users consolidate their learning and understanding of their own condition, with awards for completion that allowed for positive reinforcement.

Intervention Engagement and User Analytics

Hilliard et al [16] reported parental use of the app for an average of 106.1 (SD 37.1) days over the 3-month study period, with parents logging in 1 or more times a day on 80% of the days. Alfonsi et al [17] reported 43% (9/21) of iSpy participants were still engaged (medium to high users) at the end of the study. Both noted high app acceptability ratings among their participants pertaining to usefulness, ease of use, and learning on the app.

Kime et al [19] reported an average of 4.4 videos viewed per app user. Uniquely, their app allowed for communication and interactions with health care professionals. It was noted that more than one-half of health care professionals surveyed (99/178, 55.6%) reported that they frequently or very frequently sent out information and news to support their patients and families via the clinic portal in the app.

Holtz et al [18] reported an average number of 63 (SD 27.7) days of app use, with 84% (21/25) of the participants having use in the medium or high use categories over the 90-day study period. Linear regression analysis demonstrated a significant effect between the overall use of the app and improvement in HbA_{1c} levels ($F_{1,20}$=9.74, P<.005; R^2=0.33).

Klee et al [14] shared that 80% (27/33) of participants reported using Webdia daily. However, the subgroup analysis showed that frequency of app use did not influence the effect of intervention, although overall results showed a significant difference in the decrease in HbA_{1c} between intervention and control groups.

Goyal et al [12] reported 35% (16/46) of the participants were moderately or highly engaged users (uploaded self-monitored blood glucose [SMBG] concentrations ≥3 days a week) over the 12-month period. Exploratory regression analysis demonstrated a statistically significant association between increased SMBG and improved HbA_{1c} in the intervention group. For a subgroup of users with SMBG≥5 times daily, there was a significant improvement in HbA_{1c} of 0.58% (P=.02) compared with the parallel control subgroup who did not have a significant HbA_{1c} change.

Holtz et al [21] reported an average of 23.3 (SD 16.44) days of parental app use and 57.72 (SD 25.4) days of adolescent app use over the 12-week study period. However, they did not identify any correlation between adolescents' use days and HbA_{1c} levels.

Of the studies, 4 [11,13,15,20] did not report any usage data, as these were either interventions implemented within the health care setting or they were not included as a studied outcome.

Discussion

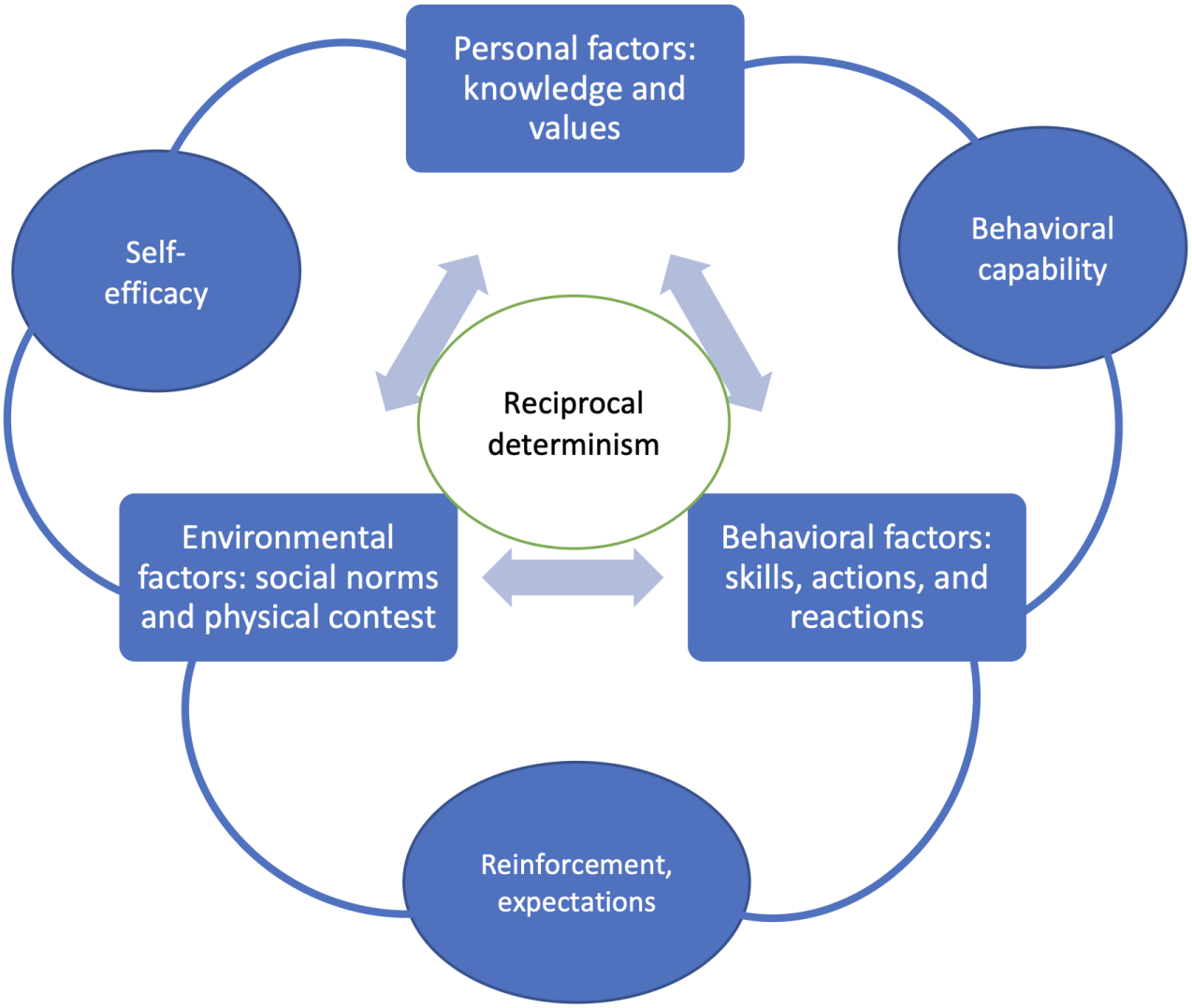
Principal Findings

We performed a scoping review to examine mHealth apps focused on the management of T1DM in the pediatric population, looking into their effects on clinical and patient-related outcomes, as well as examining the underlying behavioral frameworks informed by SCT. To our knowledge, no review has been conducted in the past decade to specifically examine pediatric digital health interventions for T1DM, with the most recent being published by Blake et al [22] in 2015. Although recent meta-analyses and umbrella reviews have evaluated mHealth interventions for diabetes [23,24], these reviews have primarily addressed adult populations or mixed-age cohorts, with limited focus on pediatric-specific outcomes. However, children and adolescents with T1DM face distinct developmental and psychosocial challenges that may influence both the design and effectiveness of digital interventions. Although Trnka et al [25] focused on categorizing coaching-based mHealth technologies for pediatric T1DM, our review builds on this by systematically evaluating a broader range of interventions, with specific attention to clinical outcomes such as HbA_{1c} and hypoglycemia, as well as behavioral outcomes including treatment adherence and self-efficacy. Furthermore, we uniquely analyzed the theoretical foundations of these interventions, drawing on constructs from SCT to elucidate the behavioral mechanisms underlying their impact.

The results of our study suggest that mHealth apps for children and adolescents with T1DM are complex interventions grounded in multiple constructs of SCT. A widely used framework in behavioral interventions [26], SCT [27] explains how behaviors are learned within a social context (Figure 3). This process is grounded in the interplay of personal factors, environmental influences, and behavior itself—a concept known as reciprocal determinism. Key constructs of SCT include (1) self-efficacy (belief in one's ability to succeed), (2) behavioral capability (knowledge and skills to perform a behavior), (3) expectations (anticipated outcomes), and (4) reinforcements (responses that influence behavior recurrence) [6].

XSL•FO
RenderX

Figure 3. Triadic reciprocal causation and key elements of social cognitive theory.



For reciprocal determinism, SCT proposes that behavior is influenced by 3 interlocking factors: personal, behavioral, and environmental factors [28]. Behavior is influenced by a continuous interaction between personal factors, environmental factors, and behavior itself.

Behavioral capability refers to an individual's ability to perform a behavior, thus relying on their knowledge and skills [28]. The individual then learns from the consequences of their behavior.

Internal and external responses to a person's behavior then positively or negatively reinforce the behavior. Expectations refer to the anticipated consequences one has of their behavior.

Self-efficacy refers to the level of a person's confidence in their ability to successfully perform a behavior [29]. It is influenced by an individual's own capabilities and environmental factors.

Impact of mHealth on Clinical Outcomes in Pediatric T1DM

Glucose Control

mHealth apps for children and adolescents with T1DM could help improve glycemic control to varying degrees. Among the included studies, some mHealth apps were reported to significantly reduce HbA_{1c} over the course of exposure to the intervention, while others kept HbA_{1c} levels stable as opposed to control groups. This is consistent with observations in previous evidence, with umbrella reviews by Wang et al [24] and Kitsiou et al [23] showing pooled favorable effects of mHealth interventions on HbA_{1c}, at least in the short term. This is a notable observation, as there are unique challenges for maintaining glycemic control in the pediatric population, ranging from hormonal fluctuations, conflicting academic commitments, and struggles with the growing responsibility of diabetes management to the impact of family and peer pressure [30]. Given the pervasiveness of smartphone use even among children and youths today [31], mHealth apps could potentially help by improving self-management ability and health literacy of children and adolescents, which tend to be poorer than that in adults [32], allowing for better treatment adherence and glycemic control. However, one should approach the matter with cautious optimism, as the effectiveness of mHealth apps remains heterogeneous across different studies and its sustained benefits in the long term remain unknown.

XSL•FO
RenderX

Educational Value and Treatment Adherence

mHealth apps can be of educational value by providing structured education that is easily accessible to patients and may facilitate improvement in communication between health care providers and patients, to align treatment steps and goals [33]. Many studies highlighted the usefulness of digital health interventions for improving patients' and their families' knowledge of their disease. From the DigiBete study [19], children and families with T1DM felt that the app was helpful for navigating the large volume of information available on T1DM and allowed them to retrieve information on T1DM readily. Of health care professionals, 83.7% (149/178) believed that the DigiBete app helped patients manage their T1DM better. Our findings corroborate with a previous cross-sectional study that found that educating children with T1DM and their parents regarding major complications of T1DM increased overall adherence to treatment and decreased their HbA_{1c} levels [34].

Family Interaction and Quality of Life

Studies have found that mHealth apps have the potential to improve family interactions as well as patients' and family members' perception of their quality of life. The MyT1DMHero app [18] linked adolescents with T1DM to their parents through two separate interfaces, allowing parents to establish a blood glucose testing schedule on the app so that the child could follow the schedule accordingly and update their blood glucose readings. Parents and children could send each other preprogrammed messages within the app to facilitate positive communication as well. The Type 1 Doing Well app [16], aimed at promoting supportive family diabetes management, allows parents to record positive T1DM-related actions that their child engaged in every day; ratings would be summarized and presented as a strengths summary for each family. There were multiple positive comments from families who felt that the app allowed them to generate a calmer atmosphere at home and helped adolescents feel that their efforts were recognized. Many of these apps, such as Webdia [14] and iSpy [17], received good feedback but did not show statistically significant differences in the quality of life from baseline and at postintervention. T1DM management can be particularly challenging for families, as it requires teamwork between parents and child [35] and conflicts can be common [36,37]. Ineffective communication can lead to youths feeling unsupported. mHealth technology offers a promising avenue to make behavioral interventions for family diabetes management more accessible [38,39]. The extent to which family dynamics and glycemic management are correlated is currently unknown, but the accumulated knowledge within a family has also been shown to have positive effects on glycemic management [40]. These studies show that, although mHealth apps are promising avenues to better improve family interactions and quality of life, more robust evidence is still required.

SCT and its Application to mHealth

The majority of studies used at least 3 SCT constructs within the same app. This underscores the complex interplay between these elements, as they do not function in isolation but could reinforce one another in the design of mHealth apps and driving behavioral change. For instance, building behavioral capability such as equipping children with essential self-care skills could enhance self-efficacy [41], which in turn can strengthen expectations of improved health outcomes [42]. Similarly, reinforcements, such as gamified incentives, have been shown to sustain engagement while bolstering productivity and self-efficacy in pediatric T1DM management through tangible progress [43]. These combinations of constructs align with current taxonomies of the ideal app for people with diabetes [44], including tips and support, rewards, and reminders. Consequently, the dynamic interaction between these constructs also mirror reciprocal determinism, where personal confidence, environmental support (eg, caregiver involvement, app-based reminders), and self-management behaviors continuously influence one another [45].

Among individual SCT constructs, behavioral capability and self-efficacy were the most frequently used. Their prevalence reflects the emphasis of mHealth interventions on fostering self-management, which is in keeping with existing knowledge on the clinical efficacy of self-monitoring as a behavioral change technique in the management of diabetes [46]. This focus on self-management is particularly crucial in the management of diabetes in pediatric populations, where competency in self-care skills (from diet, exercise, and treatment adherence,to monitoring) from early in the life course could go a long way in supporting healthy growth and prevent the development of potentially debilitating complications [47,48]. mHealth interventions could help equip children with the necessary skills and reinforce long-term self-care behaviors while instilling confidence in their ability to manage their condition [49,50]. The spotlight cast on patient self-management also aligns with a broader shift toward patient empowerment, balancing autonomy with structured caregiver and health care provider support.

In this review, studies that used gamification and improved T1DM self-management demonstrated significant differences in clinical outcomes. The use of graphics, aesthetically colorful images, and avatars are important elements in mHealth apps that can make the apps more appealing to use and better engage children and adolescents. Gamification leverages SCT by incorporating its core constructs. For example, completing levels in a game and earning rewards build mastery, provide motivation, and boost self-efficacy. Gamification also works through positive reinforcement using points or social comparisons where users are rewarded for their performance and the relevance of their achievements are recognized on the gaming platform [51].

Intervention Engagement

A key functionality of mHealth in clinical practice is improving clinical outcomes through behavior change and enhancement of patient adherence and treatment compliance. More frequent engagement is generally associated with improved treatment adherence or self-management [33].

Participant engagement and retention continue to be fundamental challenges for mHealth app studies [52]. Similarly, 2 of the included studies [17,18] reported initial high levels of usage followed by decreasing levels over time. This could be attributed to the "novelty effect" [53], whereby participants are more

XSL•FO
RenderX

motivated to good performance when an experience is new to them. Additional feedback given for the mHealth apps included the lack of personalization [16], need for repeated manual entries [12,14], and requiring a secondary mobile or supplementary device [12].

Several studies [12,14,18] that reported data on user engagement noted medium to high usage levels; however, this did not necessarily translate to improved clinical outcomes. To examine the reasons for this gap, we considered research design and participant factors. Regarding the participant factors, the quantity of use might not reflect the quality of use. A user's degree of engagement is influenced by the depth of the individual's investment in the interaction with the digital tool, and this investment may be defined temporally, affectively, and/or cognitively [54]. Hence, current engagement measures are not able to fully detail user characteristics and their interactions with the app platform. For research design factors, there is a lack of standardized classification for user engagement. Studies tended to focus on a limited number of indicators of engagement, leading to obscuration of more complex patterns and posing a challenge for the generalizability of findings [55]. To address these factors, Shani et al [55] proposed an organizing framework to guide the measurement, reporting, and interpretation of engagement in digital intervention research. They suggested the conceptualization of digital interventions as a collection of stimuli (such as notifications and reminders) and tasks (such as opening the mobile app and practicing a relaxation technique) and considering user engagement with digital interventions as a continuous process. Moving forward, the adoption of such a standardized framework would be useful for measuring, reporting, and interpreting user engagement with digital interventions.

mHealth in the Changing Landscape of T1DM Management

There have been transformative advances in technologies to support T1DM management including CGM devices, insulin pumps, and automated insulin delivery systems. There have been significant improvements in both glycemic control and quality of life measures for those who have integrated the use of these technologies into care of their diabetes [56].

Fang et al [57] reported a substantial increase in the use of CGM and insulin pumps among youths. However, the lowest rates of diabetes technology uptake and poorest glycemic control were observed among Hispanic, non-Hispanic Black, and state-insured youths and adults, and these disparities have persisted or even widened over time. Racial, ethnic, and socioeconomic disparities in access to diabetes technology have been identified as common barriers to uptake of diabetes technology in several studies [58-60]. This has also been compounded by individual-level barriers such as physical interferences from the device and body image concerns [61]. Studies [62,63] examining the psychosocial barriers to the uptake of diabetes technology have also highlighted varied user experiences with technical failures, fear of device dependence, and uncertainty of device accuracy. Diabetes care then defers back to SMBG and multiple daily insulin injections. Although CGM and insulin pump technologies are increasingly regarded as the standard of care in contemporary T1DM management, digital health interventions remain as both a valuable and accessible adjunct for individuals who either choose to delay the adoption of these technologies or simply lack access due to psychosocial or socioeconomic barriers. Such tools can still potentially support essential self-management behaviors including carbohydrate counting, insulin logging, and structured education. They could also play a complementary role in addressing educational, behavioral, and psychosocial needs that may not be the primary objective of these technologies.

Proprietary CGM manufacturers feature mobile apps that provide real-time glucose monitoring, trend alerts, and data sharing. Nonetheless, third-party apps such as mySugr and Glucose Buddy remain highly rated apps [64,65] on various mobile operating systems. Zivkovic et al [66] investigated the impact of transitioning from SMBG to CGM alongside the use of a mHealth app, mySugr, on users' glycemic control. They found that this transition significantly improved glycemic control in terms of glucose variability and mean glucose levels. There remains an extensive selection of third-party apps catering to the various needs of those with T1DM, with features supporting data tracking, data analysis, food insights, and motivational support either in the form of online communities or chatbots. mHealth apps offer the potential of an integrated platform for seamless pairing with CGM sensors and devices while providing opportunity for real-time feedback, timely intervention, communication support (with family, health care providers, user community), and education about T1DM.

Strengths and Limitations

This scoping review possesses several methodological and conceptual strengths. First, it represents the most up-to-date synthesis specifically focused on mHealth interventions for pediatric T1DM, addressing a critical evidence gap in the digital health literature. Unlike prior reviews that aggregated data across broader or mixed populations, this review targets children and adolescents—a group with unique behavioral, cognitive, and developmental needs that warrant tailored intervention strategies. The next notable strength lies in our use of deductive thematic analysis guided by an underpinning conceptual framework based on SCT. Intervention components were systematically mapped to key SCT constructs—self-efficacy, behavioral capability, reinforcement, expectations, and reciprocal determinism—allowing for an in-depth exploration of how behavioral change mechanisms are embedded in mHealth design. This theory-driven approach moves beyond descriptive synthesis to offer explanatory insight into how and why mHealth tools may support effective self-management in pediatric T1DM. Finally, by evaluating both clinical outcomes (such as HbA_{1c} and hypoglycemia) and behavioral outcomes (such as treatment adherence, self-efficacy, and quality of life), this review provides a comprehensive and translational understanding of the impact of mHealth interventions. The findings offer valuable guidance for the development and refinement of future digital health strategies that are both evidence-based and behaviorally grounded.

However, we do acknowledge that several limitations should be considered when interpreting the findings of this review.

XSL•FO
RenderX

First, the overall number of eligible studies was limited, and many were studies with small sample sizes and short follow-up durations. This limits the statistical power and generalizability of the reported outcomes, particularly with respect to long-term clinical effectiveness. Next, due to the limited availability of high-quality randomized controlled trials, we included a range of nonrandomized study designs. Although this approach aligns with scoping review methodology and allows for broader coverage of real-world evidence, it introduces inherent risks of bias and limits the ability to draw causal inferences regarding intervention effectiveness. Third, there was substantial heterogeneity in outcome reporting across studies. Clinical and behavioral endpoints—such as HbA_{1c}, adherence, and quality of life—were assessed using a variety of instruments, often with differing definitions, scales, or time points. This lack of standardization hinders direct comparisons across studies and precludes meta-analytic synthesis. Another notable limitation was the lack of inclusion of youths with T1DM as members of our research team to co-design the methodology of our work. Their lived experiences would also have provided deeper insights into contextualizing our findings and direct applicability of digital health. Using the key findings from this paper, our study team aims to involve children and young persons living with diabetes in subsequent qualitative and mixed methods studies examining the role of digital health interventions in both their lives as well as their families.

Conclusion

In conclusion, mHealth apps for children and adolescents with T1DM are complex behavioral interventions that often incorporate multiple constructs from SCT. Current evidence suggests that these interventions may contribute to improved glycemic control, with several studies reporting stability or reductions in HbA_{1c} relative to control groups. Although the evidence base is heterogeneous, mHealth interventions show potential benefits across a range of behavioral and psychosocial outcomes, including self-efficacy, treatment adherence, quality of life, hypoglycemia management, and family dynamics. To strengthen the evidence and guide future implementation, further research is needed, using standardized outcome measures and robust methodological designs to better understand how theory-informed digital tools influence self-management and holistic well-being in pediatric T1DM.

Conflicts of Interest

None declared.

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File, 86 KB - mhealth_v14i1e79338_app1.docx]

References

1. Miller GF, Coffield E, Leroy Z, Wallin R. Prevalence and costs of five chronic conditions in children. J Sch Nurs 2016 Oct;32(5):357-364. [doi: 10.1177/1059840516641190] [Medline: 27044668]
2. Chiang JL, Maahs DM, Garvey KC, et al. Type 1 diabetes in children and adolescents: a position statement by the American Diabetes Association. Diabetes Care 2018 Sep;41(9):2026-2044. [doi: 10.2337/dci18-0023] [Medline: 30093549]
3. Diabetes Control, Complications Trial Research Group, Nathan DM, Genuth S. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med 1993 Sep 30;329(14):977-986. [doi: 10.1056/NEJM199309303291401]
4. Clements MA, Foster NC, Maahs DM, et al. Hemoglobin A1c (HbA1c) changes over time among adolescent and young adult participants in the T1D exchange clinic registry. Pediatr Diabetes 2016 Aug;17(5):327-336. [doi: 10.1111/pedi.12295] [Medline: 26153338]
5. ElSayed NA, Aleppo G, Bannuru RR, et al. 1. Improving care and promoting health in populations: Standards of Care in Diabetes—2024. Diabetes Care 2024 Jan 1;47(Supplement_1):S11-S19. [doi: 10.2337/dc24-S001]
6. Nabavi RT, Bijandi MS. Bandura's social learning theory & social cognitive learning theory. ResearchGate. 2012. URL: https://www.researchgate.net/publication/267750204_Bandura [accessed 2026-02-22]
7. Schunk DH, DiBenedetto MK. Motivation and social cognitive theory. Contemp Educ Psychol 2020 Jan;60:101832. [doi: 10.1016/j.cedpsych.2019.101832]
8. Peters MDJ, Marnie C, Tricco AC, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020 Oct;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
9. Covidence. URL: https://www.covidence.org [accessed 2026-02-22]
10. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012 May 8;14(3):e70. [doi: 10.2196/jmir.2058] [Medline: 22564332]
11. Nilsson S. How young people can learn about newly diagnosed type 1 diabetes. Nurs Child Young People 2016 Apr;28(3):22-26. [doi: 10.7748/ncyp.28.3.22.s21] [Medline: 27063051]
12. Goyal S, Nunn CA, Rotondi M, et al. A mobile app for the self-management of type 1 diabetes among adolescents: a randomized controlled trial. JMIR Mhealth Uhealth 2017 Jun 19;5(6):e82. [doi: 10.2196/mhealth.7336] [Medline: 28630037]

XSL•FO
RenderX

13. Clifton DC, Benjamin RW, Brown AR, Ostrovsky DA, Narayan AP. A tablet-based educational tool: toward more comprehensive pediatric patient education. Clin Pediatr (Phila) 2018 Sep;57(10):1176-1182. [doi: 10.1177/0009922818766621] [Medline: 29575940]
14. Klee P, Bussien C, Castellsague M, et al. An intervention by a patient-designed do-it-yourself mobile device app reduces HbA1c in children and adolescents with type 1 diabetes: a randomized double-crossover study. Diabetes Technol Ther 2018 Dec;20(12):797-805. [doi: 10.1089/dia.2018.0255] [Medline: 30403495]
15. Pramanik BK, Angelin JJ, Mathai VJ, Mathai S, Korula S, Simon A. Smartphone app as motivational intervention to improve glycemic control in adolescents with type 1 diabetes. Indian J Pediatr 2019 Dec;86(12):1118-1123. [doi: 10.1007/s12098-019-03035-x] [Medline: 31353430]
16. Hilliard ME, Cao VT, Eshtehardi SS, et al. Type 1 Doing Well: pilot feasibility and acceptability study of a strengths-based mHealth app for parents of adolescents with type 1 diabetes. Diabetes Technol Ther 2020 Nov 1;22(11):835-845. [doi: 10.1089/dia.2020.0048]
17. Alfonsi JE, Choi EEY, Arshad T, et al. Carbohydrate counting app using image recognition for youth with type 1 diabetes: pilot randomized control trial. JMIR Mhealth Uhealth 2020 Oct 28;8(10):e22074. [doi: 10.2196/22074] [Medline: 33112249]
18. Holtz B, Mitchell KM, Holmstrom AJ, et al. An mHealth-based intervention for adolescents with type 1 diabetes and their parents: pilot feasibility and efficacy single-arm study. JMIR Mhealth Uhealth 2021 Sep 14;9(9):e23916. [doi: 10.2196/23916] [Medline: 34519670]
19. Kime N, Zwolinsky S, Pringle A. Evaluation of the DigiBete app, a self-management app for type 1 diabetes: experiences of young people, families, and healthcare professionals. Children (Basel) 2023 Dec 16;10(12):1933. [doi: 10.3390/children10121933] [Medline: 38136135]
20. Bender C, Jensen MH, Skindbjerg SB, et al. Impact of Sten-O Starter on glycemic management in children and adolescents with type 1 diabetes in the north region of Denmark. Diabetes Epidemiology and Management 2024 Apr;14:100200. [doi: 10.1016/j.deman.2024.100200]
21. Holtz BE, Mitchell KM, Holmstrom AJ, et al. The effect of an mHealth intervention for adolescents with type 1 diabetes and their parents. J Telemed Telecare 2024 Aug;30(7):1155-1162. [doi: 10.1177/1357633X221125835] [Medline: 36177538]
22. Blake H, Roberts A, Stanulewicz N. Telemedicine and mHealth interventions for children and young people with type one diabetes (T1DM). International Journal of Diabetes and Endocrinology ;1(1):100-104 [FREE Full text]
23. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: an overview of systematic reviews. PLoS ONE 2017;12(3):e0173160. [doi: 10.1371/journal.pone.0173160] [Medline: 28249025]
24. Wang Y, Min J, Khuri J, et al. Effectiveness of mobile health interventions on diabetes and obesity treatment and management: systematic review of systematic reviews. JMIR Mhealth Uhealth 2020 Apr 28;8(4):e15400. [doi: 10.2196/15400] [Medline: 32343253]
25. Trnka P, Aldaghi T, Muzik J. Categorization of mHealth coaching technologies for children or adolescents with type 1 diabetes: systematic review. JMIR Pediatr Parent 2024 Oct 10;7:e50370. [doi: 10.2196/50370] [Medline: 39388720]
26. Ng CY, Thomas-Uribe M, Yang YA, et al. Theory-based health behavior interventions for pediatric chronic disease management: a systematic review. JAMA Pediatr 2018 Dec 1;172(12):1177-1186. [doi: 10.1001/jamapediatrics.2018.3039] [Medline: 30357260]
27. Bandura A. Social foundations of thought and action. In: Marks DF, editor. The Health Psychology Reader: SAGE Publications Ltd; 2002:94-106. [doi: 10.4135/9781446221129.n6]
28. Nickerson C. Albert Bandura's social cognitive theory. Simply Psychology. 2025 Mar 31. URL: https://www.simplypsychology.org/social-cognitive-theory.html [accessed 2026-02-23]
29. Health behavior and health education, part three, chapter eight: key constructs. University of Pennsylvania. URL: https://www.med.upenn.edu/hbhe4/part3-ch8-key-constructs.shtml [accessed 2026-02-23]
30. Gerber BS, Solomon MC, Shaffer TL, Quinn MT, Lipton RB. Evaluation of an internet diabetes self-management training program for adolescents and young adults. Diabetes Technol Ther 2007 Feb;9(1):60-67. [doi: 10.1089/dia.2006.0058] [Medline: 17316099]
31. Ko Y, Park S. Impacts of problematic smartphone use on children: perspectives from main caregivers. Arch Psychiatr Nurs 2023 Oct;46:59-64. [doi: 10.1016/j.apnu.2023.08.007] [Medline: 37813505]
32. Guo J, Whittemore R, Grey M, Wang J, Zhou ZG, He GP. Diabetes self-management, depressive symptoms, quality of life and metabolic control in youth with type 1 diabetes in China. J Clin Nurs 2013 Jan;22(1-2):69-79. [doi: 10.1111/j.1365-2702.2012.04299.x] [Medline: 23106340]
33. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3(1):4. [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
34. Wysocki T, Lochrie A, Antal H, Buckloh LM. Youth and parent knowledge and communication about major complications of type 1 diabetes: associations with diabetes outcomes. Diabetes Care 2011 Aug;34(8):1701-1705. [doi: 10.2337/dc11-0577] [Medline: 21659641]
35. Wysocki T, Nansel TR, Holmbeck GN, et al. Collaborative involvement of primary and secondary caregivers: associations with youths' diabetes outcomes. J Pediatr Psychol 2009 Sep;34(8):869-881. [doi: 10.1093/jpepsy/jsn136] [Medline: 19112077]

XSL•FO
RenderX

36. Jaser SS. Family interaction in pediatric diabetes. Curr Diab Rep 2011 Dec;11(6):480-485. [doi: 10.1007/s11892-011-0222-y] [Medline: 21853415]
37. Spencer J, Cooper H, Milton B. Qualitative studies of type 1 diabetes in adolescence: a systematic literature review. Pediatr Diabetes 2010 Aug;11(5):364-375. [doi: 10.1111/j.1399-5448.2009.00603.x] [Medline: 19895566]
38. Schueller SM, Muñoz RF, Mohr DC. Realizing the potential of behavioral intervention technologies. Curr Dir Psychol Sci 2013 Dec;22(6):478-483. [doi: 10.1177/0963721413495872]
39. Mulvaney SA, Ritterband LM, Bosslet L. Mobile intervention design in diabetes: review and recommendations. Curr Diab Rep 2011 Dec;11(6):486-493. [doi: 10.1007/s11892-011-0230-y] [Medline: 21960031]
40. Wennick A, Hallström I. Swedish families' lived experience when a child is first diagnosed as having insulin-dependent diabetes mellitus: an ongoing learning process. J Fam Nurs 2006 Nov;12(4):368-389. [doi: 10.1177/1074840706296724] [Medline: 17099116]
41. Cockcroft EJ, Clarke R, Dias RP, et al. Effectiveness of educational and psychoeducational self-management interventions in children and adolescents with type 1 diabetes: a systematic review and meta-analysis. Pediatr Diabetes 2024;2024:2921845. [doi: 10.1155/2024/2921845] [Medline: 40302946]
42. Tabernero C, Rebollo-Román Á, Villaécija-Rodríguez J, Luque B. Charting a path to health: the empowering influence of self-efficacy for the self-management of type 1 diabetes in children and adolescents. Int J Psychol 2024 Dec;59(6):1121-1132. [doi: 10.1002/ijop.13240] [Medline: 39238255]
43. Rewolinski JA, Kelemen A, Liang Y. Type I diabetes self-management with game-based interventions for pediatric and adolescent patients. Comput Inform Nurs 2020 Jun 25;39(2):78-88. [doi: 10.1097/CIN.0000000000000646] [Medline: 32590404]
44. Martinez-Millana A, Jarones E, Fernandez-Llatas C, Hartvigsen G, Traver V. App features for type 1 diabetes support and patient empowerment: systematic literature review and benchmark comparison. JMIR Mhealth Uhealth 2018 Nov 21;6(11):e12237. [doi: 10.2196/12237] [Medline: 30463839]
45. Bandura A. The self system in reciprocal determinism. American Psychologist 1978;33(4):344-358. [doi: 10.1037/0003-066X.33.4.344]
46. Tarricone R, Petracca F, Svae L, Cucciniello M, Ciani O. Which behaviour change techniques work best for diabetes self-management mobile apps? Results from a systematic review and meta-analysis of randomised controlled trials. EBioMedicine 2024 May;103:105091. [doi: 10.1016/j.ebiom.2024.105091] [Medline: 38579364]
47. Ispriantari A, Agustina R, Konlan KD, Lee H. Family-centered interventions for children and adolescents with type 1 diabetes mellitus: an integrative review. Child Health Nurs Res 2023 Jan;29(1):7-23. [doi: 10.4094/chnr.2023.29.1.7] [Medline: 36760109]
48. Nukpezah RN, Salisu WJ, Doat AR. Self-care practices among children and adolescents with diabetes: an integrative review. Asian Journal of Research in Nursing and Health 2021;4(1):140-157 [FREE Full text]
49. Berndt RD, Takenga C, Preik P, et al. Impact of information technology on the therapy of type-1 diabetes: a case study of children and adolescents in Germany. J Pers Med 2014 Apr 16;4(2):200-217. [doi: 10.3390/jpm4020200] [Medline: 25563223]
50. Liu K, Xie Z, Or CK. Effectiveness of mobile app-assisted self-care interventions for improving patient outcomes in type 2 diabetes and/or hypertension: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2020 Aug 4;8(8):e15779. [doi: 10.2196/15779] [Medline: 32459654]
51. Krath J, Schürmann L, von Korflesch HFO. Revealing the theoretical basis of gamification: a systematic review and analysis of theory in research on gamification, serious games and game-based learning. Comput Human Behav 2021 Dec;125:106963. [doi: 10.1016/j.chb.2021.106963]
52. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022 Apr 26;24(4):e35120. [doi: 10.2196/35120] [Medline: 35471414]
53. Marek MW, Vivian Wu WC. Motivational affordances of the novelty effect in TELL. 2021 Presented at: 2021 International Conference on Advanced Learning Technologies (ICALT); Tartu, Estonia p. 268-269. [doi: 10.1109/ICALT52272.2021.00086]
54. Torous J, Michalak EE, O'Brien HL. Digital health and engagement-looking behind the measures and methods. JAMA Netw Open 2020 Jul 1;3(7):e2010918. [doi: 10.1001/jamanetworkopen.2020.10918] [Medline: 32678446]
55. Nahum-Shani I, Yoon C. Towards the science of engagement with digital interventions. Curr Dir Psychol Sci 2024 Aug;33(4):239-246. [doi: 10.1177/09637214241254328] [Medline: 39552747]
56. Lundgrin EL, Kelly CA, Bellini N, Lewis C, Rafi E, Hatipoglu B. Diabetes technology trends: a review of the latest innovations. The Journal of Clinical Endocrinology & Metabolism 2025 Feb 25;110(Supplement_2):S165-S174. [doi: 10.1210/clinem/dgaf034]
57. Fang M, Xu Y, Ballew SH, et al. Trends and disparities in technology use and glycemic control in type 1 diabetes. JAMA Netw Open 2025 Aug 1;8(8):e2526353. [doi: 10.1001/jamanetworkopen.2025.26353] [Medline: 40788645]
58. Tsai D, Flores Garcia J, Fogel JL, Wee CP, Reid MW, Raymond JK. Diabetes technology experiences among Latinx and non-Latinx youth with type 1 diabetes. J Diabetes Sci Technol 2022 Jul;16(4):834-843. [doi: 10.1177/19322968211029260] [Medline: 34225480]

XSL•FO
RenderX

59. Pauley ME, Berget C, Messer LH, Forlenza GP. Barriers to uptake of insulin technologies and novel solutions. Med Devices (Auckl) 2021;14:339-354. [doi: 10.2147/MDER.S312858] [Medline: 34803408]
60. Conway RB, Snell-Bergeon J, Honda-Kohmo K, et al. Disparities in diabetes technology uptake in youth and young adults with type 1 diabetes: a global perspective. J Endocr Soc 2024 Nov 26;9(1):bvae210. [doi: 10.1210/jendso/bvae210] [Medline: 39703363]
61. Kubiak T, Priesterroth L, Barnard-Kelly KD. Psychosocial aspects of diabetes technology. Diabet Med 2020 Mar;37(3):448-454. [doi: 10.1111/dme.14234] [Medline: 31943354]
62. Brew-Sam N, Chhabra M, Parkinson A, et al. Diabetes technology experiences of young people living with type 1 diabetes mellitus and their parents: hybrid theoretical foundation guided analysis. medRxiv. Preprint posted online on Nov, 2021. [doi: 10.1101/2021.11.08.21265793]
63. Vloemans AF, van Beers CAJ, de Wit M, et al. Keeping safe. Continuous glucose monitoring (CGM) in persons with type 1 diabetes and impaired awareness of hypoglycaemia: a qualitative study. Diabet Med 2017 Oct;34(10):1470-1476. [doi: 10.1111/dme.13429] [Medline: 28731509]
64. Veazie S, Winchell K, Gilbert J, et al. Mobile applications for self-management of diabetes, technical brief, no. 31. : Agency for Healthcare Research and Quality (US); 2018 URL: https://pubmed.ncbi.nlm.nih.gov/30088878/ [Medline: 30088878]
65. Fu HNC, Rizvi RF, Wyman JF, Adam TJ. Usability evaluation of four top-rated commercially available diabetes apps for adults with type 2 diabetes. Comput Inform Nurs 2020 Jun;38(6):274-280. [doi: 10.1097/CIN.0000000000000596] [Medline: 31904594]
66. Zivkovic J, Mitter M, Theodorou D, Kober J, Mueller-Hoffmann W, Mikulski H. Transitioning from self-monitoring of blood glucose to continuous glucose monitoring in combination with a mHealth app improves glycemic control in people with type 1 and type 2 diabetes. Diabetes Technol Ther 2025 Jan 1;27(1):10-18. [doi: 10.1089/dia.2024.0169]

Abbreviations

CGM: continuous glucose monitoring
HbA_{1c}: hemoglobin A_{1c}
mHealth: mobile health
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
SCT: social cognitive theory
SMBG: self-monitored blood glucose
T1DM: type 1 diabetes mellitus

Edited by M DeJonckheere; submitted 20.Jun.2025; peer-reviewed by H Witteman, N Juniarti, X Rong; accepted 05.Feb.2026; published 02.Mar.2026.

Please cite as:
Beh E, Lin J, Leong G, Loh PY, Poh EJ, Chan D
Digital Health Interventions in Children and Adolescents With Type 1 Diabetes Mellitus and Their Impact on Clinical and Behavioral Outcomes: Scoping Review
JMIR Mhealth Uhealth 2026;14:e79338
URL: https://mhealth.jmir.org/2026/1/e79338
doi:10.2196/79338

© Eirena Beh, Jiaying Lin, Geraldyn Leong, Pei Yi Loh, Ee Jin Poh, Daniel Chan. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 2.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Examining the Use of Consumer Wearable Devices and Digital Tools for Stress Measurement in College Students: Scoping Review of Methods

Aarti Sathyanarayana[1,2*], PhD; Ohida Binte Amin[1,2*], BSc; Jennie An[1,2], MS; Jukka Pekka Onnela[3], PhD

[1]Bouve College of Health Sciences, Northeastern University, 360 Huntington Avenue, Boston, MA, United States

[2]Khoury College of Computer Science, Northeastern University, Boston, MA, United States

[3]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Aarti Sathyanarayana, PhD
Bouve College of Health Sciences, Northeastern University, 360 Huntington Avenue, Boston, MA, United States

Abstract

Background: College-aged students face persistent academic and social stress that adversely affects their mental and physical health. Digital phenotyping with wearable devices enables real-time stress monitoring from continuous physiological signals, supporting just-in-time therapeutic interventions to improve student well-being. Despite rapid advances in wearables and analytical methods, it remains unclear which devices, physiological signals, and machine learning or deep learning approaches are most commonly used for stress detection in this population.

Objective: This study aimed to systematically review the literature to identify best practices and emerging trends in stress measurement using wearable technology and digital tools among college-aged students. We sought to evaluate commonalities in sensor types, datasets, and machine learning approaches used for stress detection.

Methods: A systematic search was conducted across medical and computer science databases, including Embase, PubMed, IEEE Xplore, and ACM Digital Library, for studies published between January 2020 and December 2025. Studies were included if they examined psychological stress detection using wearable or digital tools among college-aged students and were excluded if they focused on nonpsychological stress, were reviews or prototypes without a defined study population, or lacked clear population information. Two reviewers independently screened studies and extracted data on the wearable sensors, physiological signals, datasets, and modeling approaches to summarize trends in stress prediction.

Results: A total of 134 studies met the inclusion criteria and were included in the review from the original 792 papers. Electrodermal activity was the most frequently used physiological signal, appearing in 57.5% (n=77) of studies, and wrist-worn wearable devices were the predominant sensing modality. Among studies that compared algorithms, support vector machines were identified as the most commonly applied and best-performing model in 33.3% (n=45) of cases. Overall, 62.8% (n=84) of included studies relied on preexisting datasets, and approximately 80% (n=67) of those used the Wearable Stress and Affect Detection dataset, which contains only 15 participants. Demographic reporting was inconsistent, as 27.6% (n=37) of studies did not report sex distribution, and only 4 studies justified the sample size. The use of temporal modeling algorithms was limited, despite their importance for capturing the dynamic, time-varying nature of stress. This review highlights persistent gaps and underscores the need for more diverse datasets and advanced modeling approaches to improve stress detection accuracy.

Conclusions: Our review innovatively synthesizes wearable-based stress detection research focused on college-aged students. Unlike prior reviews that aggregate heterogeneous populations or focus primarily on algorithmic performance, this review focused on wearable sensors, physiological signals, modeling approaches, and methodological quality to identify persistent gaps limiting real-world deployment. These findings inform the development of more generalizable monitoring systems to support early mental health intervention in students.

(JMIR Mhealth Uhealth 2026;14:e64144) doi:10.2196/64144

KEYWORDS

digital phenotyping; wearable technology; stress detection; machine learning; college students

XSL•FO
RenderX

Introduction

With the widespread adoption of wearable devices, numerous stress monitoring frameworks have been designed specifically for undergraduate students [1-3], given their heightened susceptibility to psychological stress. This need is underscored by findings that over 80% of undergraduate students report experiencing significant stress related to their academic life [4]. University life can be particularly overwhelming, as many students experience independent living for the first time while navigating self-care and decision-making [5]. While positive stress can sometimes enhance academic performance, persistent and long-lasting chronic stress can negatively impact both mental and physical health [6]. By proactively managing stress, individuals can mitigate the risk of stress-related health issues, including cardiovascular problems, gastrointestinal issues, mental health disorders, substance abuse, and chronic diseases such as diabetes or hypertension [7]. Stress also significantly disrupts sleep [8], social interactions [9], and academic performance [10], contributing to insomnia [11], anxiety [12], and a weakened immune system [13]. Digital phenotyping of stress, leveraging wearable and mobile technologies, enables just-in-time stress management solutions that help prevent chronic stress from compromising long-term health.

In recent years, the use of consumer wearables to monitor physical activity [14] and other lifestyle traits [15] has become more prevalent. For example, many commercial consumer wearables are being used to keep track of and improve upon fitness regimens [16]. With this increased availability of wearables comes the possibility for real-time health management using these commercial devices that are more convenient and lightweight [17]. The use of wearables to passively monitor physiological signals and the subsequent analysis using various machine learning and deep learning models brings enormous benefits for health management [18]. By passively tracking heart rate (HR) or heart rate variability (HRV), skin temperature, electrodermal activity (EDA), electroencephalogram, electrocardiogram (ECG), acceleration, and other physiological variables, smartphones and wearable sensors can provide features related to signs indicative of poor mental health [19]. Stress is reflected in the body with increased EDA or HR, reflecting the autonomic nervous system and hypothalamic-pituitary-adrenal axis activity [20]. Many studies have tracked these biosignals with commercial digital tools to build models to measure stress [21]. In this review, we examine the trends in the current use of these digital tools to measure stress.

Stress assessment using wearable and digital technologies has been conducted across both controlled laboratory experiments and real-world, free-living conditions. In laboratory settings, studies commonly use well-established stress elicitation tasks [22] with resting periods used as baselines. Commonly used tasks include the Trier Social Stress Test (TSST), mental arithmetic tasks [23] (eg, the Montreal Imaging Stress Task [24]), the Stroop color-word test, public speaking, startle response tests, cold pressor tests, and stress-inducing video stimuli [25]. Across these studies, researchers used varying combinations of physiological signals and derived diverse feature sets following preprocessing steps such as artifact removal, signal normalization, and feature selection [26]. In contrast, stress monitoring in free-living environments relies on self-reported stress measures alongside passive and unobtrusive sensing approaches that capture daily physiological and behavioral patterns using wearable devices and smartphones [27]. These approaches vary widely in sensor availability, feature extraction methods, and contextual information, leading to substantial heterogeneity in how stress is represented and quantified across wearables and digital tools.

Alongside variability in study design, stress capture methods, and physiological sensing, approaches for stress prediction differ markedly across studies. Both traditional machine learning [28,29] and deep learning [30,31] models have been applied to physiological time-series data to identify stress episodes and enable just-in-time interventions. However, it remains unclear which modeling paradigms are most appropriate for different physiological signals and smartphone-derived active and passive sensing data, how model architectures should be designed to capture temporal stress dynamics, and whether increased model complexity consistently yields performance gains. These methodological challenges hinder the translation of wearable-based stress detection systems into practical tools for continuous monitoring and personalized support in college-aged populations, underscoring the need for systematic evidence synthesis and clearer methodological pathways for future research.

This review aims to identify trends in current research and highlight areas for improvement that future researchers should focus on. There is a need to understand which algorithms perform best, which wearables are most used, and which signals are most informative. The topic of this review is identifying moments of high stress using digital tools and ubiquitous data in college-aged students. We examine both machine learning and deep learning advancements in the field, as well as comparisons of methods, where a scoping review is the most appropriate synthesis method to address the stated objectives. Our population of interest includes college students aged 18 - 24 years. Publication dates of interest include conference and journal papers published between 2020 and 2025, as we focus on advancements in the field, including newer wearable devices and algorithms. We are also narrowing our focus to college students, as university is a particularly stressful place where their health and lifestyle habits are likely to fluctuate [32]. Academic stress is directly linked to health crises such as anxiety and depression, indicating an opportunity to monitor stress and prevent health from deteriorating [33]. In this scoping review, we summarize the wearables used, signals measured, and algorithms performed to measure stress. We then discuss trends in data and practices across papers. We conduct a quality assessment of all included studies. We also provide an overview of the results and a discussion of limitations and future possibilities for stress measurement. As a result, this scoping review aims to synthesize recent research on wearable or digital tool–based stress detection among college-aged students by summarizing the sensing technologies used, the physiological and behavioral signals measured, the machine learning and deep learning models applied, and key methodological practices, to

XSL•FO
RenderX

identify current trends, limitations, and directions for future research.

Methods

Overview

We conducted a scoping review to characterize current research on stress detection using wearable and digital tools among college-aged students. This review synthesizes studies published between January 2020 and December 2025 to summarize commonly used wearable devices, physiological signals, datasets, and machine learning or deep learning approaches for identifying high-stress moments. By organizing existing methods and conducting a quality assessment, this review provides an overview of methodological practices and highlights areas for future research in wearable-based stress measurement. This scoping review adhered to the methodological framework proposed by Arksey and O'Malley [34], which includes identifying the research question, identifying relevant studies, study selection, charting the data, and collating, summarizing, and reporting the results. Finally, this scoping review was conducted and reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines to ensure transparency and reproducibility [35].

Protocol and Registration

No formal review protocol was registered for this scoping review, as the objective was to map the scope and characteristics of existing evidence in stress prediction research using wearable technology.

Eligibility Criteria

We defined eligibility criteria to ensure that only relevant and methodologically appropriate studies were included in this review. Studies were included if they measured or classified psychological stress using physiological signals from a tool, wearable, or sensor. Only experimental or observational studies published in English were considered. The target population was college students aged 18 - 24 years. Studies that partially included this age range were eligible if they explicitly mentioned students as a distinct group or if the mean age, along with the SD, fell within the target population. Studies were excluded if they focused on nonpsychological stress (eg, mechanical stress), were review papers, extended abstracts, or prototype descriptions without a defined study population. Papers without clear population details or those identifying participants solely by employment (eg, "office workers" or "hospitalized patients") were also excluded.

Information Sources

We searched IEEE Xplore, ACM Digital Library, PubMed, and Embase for conference and journal papers covering studies published between January 2020 and December 2025, a time frame selected to capture recent developments in wearable sensing technologies and stress detection methodologies.

Search

We used a combination of terms related to the key concepts of psychological stress, wearable devices, and sensors (full search per database is provided in Multimedia Appendix 1). We extracted each database searched and the platform used, including IEEE Xplore, ACM Digital Library, PubMed, and Embase, in accordance with PRISMA-S (PRISMA literature search extension) [36], and all databases were searched independently rather than through a multidatabase platform. No multidatabase searching or study registry searching was conducted. No additional online resources (eg, tables of contents, print conference proceedings, and websites) were browsed. No additional search methods were used, including citation searching, contacting authors or experts, or setting up citation alerts. The full search strategies for each database are provided in Multimedia Appendix 1, including the specification that no filters or limits other than language (English) and publication date (January 2020 to December 2025) were applied. Search strategies were developed with input from 2 academic librarians; however, search strategies from prior reviews were not reused, and no formal peer review of the search strategy was conducted. No additional methods were used to update the search. Therefore, searches were limited to studies published in English within the specified date range. No restrictions were applied based on study design. All retrieved records were initially screened. Following screening, records were imported into Rayyan (Rayyan Systems Inc) [37], where duplicate entries were identified and removed. The deduplicated set of records was then used for abstract and full-text screening.

Selection of Sources of Evidence

Two independent reviewers screened all records using a 2-stage selection process. Studies were checked for eligibility by 2 reviewers independently screening titles and abstracts. This first round of filtering focused on relevance. Abstracts were also screened for population. Some papers did not mention population in the abstract and were thus moved to full-text screening. This resulted in 261 papers for full-text screening. During this second round of filtering, studies were also checked for eligibility by 2 researchers independently reviewing the full text. Disagreements at any stage of eligibility and filtering were resolved by the 2 reviewers discussing their reasons for either inclusion, exclusion, or neither. Full agreement was reached for abstract and full-text screening, leading to the final inclusion of 134 papers.

Data Charting

A standardized data-charting form was jointly developed by 2 reviewers to identify and extract relevant information aligned with the review objectives. The form was pilot-tested on a subset of included studies and refined iteratively to ensure completeness and consistency. Two reviewers independently charted data from all eligible studies, compared their entries, and resolved discrepancies through discussion. All data were extracted directly from the published papers, and no additional information was sought from study authors.

XSL•FO
RenderX

Data Items

To extract consistent information from each paper, we conducted systematic data extraction as outlined in Tables 1-3. Extracted variables included study details (title, authors, publication date, study purpose, and data collection duration), sample characteristics (age, sex, sample size, and demographic information), sensor type, and all available feature categories used in the study (sleep, physiological signals, calorie intake or expenditure, phone use, activity, location, and survey or EMA data). For studies conducting algorithm comparisons, we additionally extracted the types of signals analyzed, devices used, algorithms tested, performance measures, best-performing algorithm, validation strategy, and outcome measures.

XSL•FO
RenderX

Table . Summary characteristics of 134 included studies.

Study	Sample (n)	Sex	Age (years), mean (SD)	Sleep	Physiological signals	Calorie intake or expenditure	Phone use	Activity	Location	Survey	Total feature types
Bellante et al [38]	15	3 females and 12 males	27.5 (2.4)		✔						1
Faro and Giordano [39]	__[a]	—	College students		✔			✔	✔		3
Faro et al [40]	31	—	College students		✔						1
Iranfar et al [41]	95	95 males	20.43 (2.17)		✔						1
Mohammadi et al [42]	18	5 females and 13 males	27.5 (2.4)		✔						1
Mustafa et al [43]	15	3 females and 12 males	27.5 (2.4)		✔						1
Arsalan and Majid [44]	40	20 females and 20 males	24.86 (6.69)		✔						1
Li and Sano [45]	239	—	College students		✔			✔			2
Can et al [27]	14	5 females and 9 males	23.5 (N/A)[a]		✔						1
Cheadle et al [46]	100	61 females and 39 males	20.4 (N/A)		✔						1
Chen et al [47]	30	20 females and 10 males	23 (NA)							✔	1
Gupta et al [48]	15	3 females and 12 males	27.5 (2.4)		✔			✔			2
Panganiban and de Leon [49]	36	—	21.5 (N/A)		✔						1
Gasparini et al [50]	36	14 females and 22 males	24.7 (3.3)		✔						1
Azgomi et al [51]	20	—	College students		✔			✔			2
Yu and Sano [31]	243	—	College students		✔			✔			2
Han et al [52]	17	4 females and 13 males	24 (N/A)		✔						1
Wu et al [53]	264	113 females and 151 males	22.8 (N/A)		✔			✔			2
Jelsma et al [54]	100	—	College students		✔						1

XSL•FO
RenderX

Study	Sample (n)	Sex	Age (years), mean (SD)	Sleep	Physiological signals	Calorie intake or expenditure	Phone use	Activity	Location	Survey	Total feature types
Lai et al [55]	15	3 females and 12 males	27.5 (2.4)		✔						1
Liakopoulos et al [56]	Multiple datasets	Multiple datasets	Multiple datasets		✔						1
Li and Sano [57]	239	—	College students		✔			✔			2
Hssayeni and Ghoraani [58]	15	3 females and 12 males	27.5 (2.4)		✔			✔			2
Gil-Martin et al [59]	15	3 females and 12 males	27.5 (2.4)		✔			✔			2
Han et al [60]	20	—	College students		✔						1
Mishra et al [61]	27	15 females and 12 males	23 (3.24)		✔			✔		✔	3
Mishra et al [26]	90	—	Graduate and undergraduate students		✔						1
Momeni et al [62]	60	60 males	20.43 (2.17)		✔						1
Rashid et al [63]	15	3 females and 12 males	27.5 (2.4)		✔						1
Bobade and Vani [18]	15	3 females and 12 males	27.5 (2.4)		✔			✔			2
Yannam et al [64]	70	—	Undergraduate	✔	✔		✔	✔	✔		5
Pakhomov et al [65]	18	14 females and 4 males	20.1 (2.01)		✔			✔			2
Holder et al [66]	11	10 females and 1 male	27.5 (2.4)		✔			✔			2
Elzeiny and Qaraqe [67]	22	5 females and 17 males	27.5 (2.4)		✔						1
Heo et al [68]	15	3 females and 12 males	27.5 (2.4)		✔						1
Kar et al [69]	15	3 females and 12 males	27.5 (2.4)		✔			✔			2
Prashant et al [70]	15	3 females and 12 males	27.5 (2.4)		✔						1

XSL•FO
RenderX

Study	Sample (n)	Sex	Age (years), mean (SD)	Sleep	Physiological signals	Calorie intake or expenditure	Phone use	Activity	Location	Survey	Total feature types
Samyoun et al [71]	15	3 females and 12 males	27.5 (2.4)		✔						1
Silva et al [72]	82	63 females and 19 males	22.13 (5.55)	✔	✔	✔					3
Islam et al [73]	20	7 females, 12 males, and 1 nonbinary	22 (N/A)	✔	✔			✔	✔		4
Vidal et al [32]	49	25 females and 24 males	18.1 (N/A)	✔						✔	2
Wu et al [74]	169	81 females and 88 males	22.8 (6.2)		✔						1
Mitro et al [75]	30	22 males and 8 females	27.5 (2.4)		✔						1
Zhu et al [28]	112	—	—		✔						3
Tutunji et al [76]	84	32 males and 52 females	College students		✔					✔	5
Lange et al [77]	15	12 males	27.5 (2.4)		✔						4
Abdul et al [78]	20	—	—		✔						2
Almadhor et al [79]	15	12 males and 3 females	27.5 (2.4)		✔						6
Vos e al [29]	136	—	—		✔						13
Mai and Chung [80]	15	—	30 (7)		✔					✔	2
Sepanloo et al [81]	12	—	29.6 (10.1)		✔						3
Gedam et al [2]	200	128 male and 72 female	23 (N/A)		✔						3
Darwish et al [82]	1017	496 males and 454 females	27.5 (2.4)		✔						3
Lim et al [83]	5	4 males and 1 female	—		✔						2
Bloomfield et al [3]	525	144 males and 381 females	22 (N/A)	✔	✔					✔	6
Nazeer et al [84]	15	12 males and 3 females	27.5 (2.4)		✔						6

XSL•FO
RenderX

Study	Sample (n)	Sex	Age (years), mean (SD)	Sleep	Physiological signals	Calorie intake or expenditure	Phone use	Activity	Location	Survey	Total feature types
Almadhor et al [85]	15	12 males and 3 females	27.5 (2.4)		✔						6
Stržinar et al [86]	15	12 males and 3 females	27.5 (2.4)		✔						1
Chen and Lee [30]	30	6 males and 24 females	20.4 (N/A)		✔						3
Feng et al [87]	15	12 males and 3 females	27.5 (2.4)		✔						6
Xuanzhi et al [88]	15+	—	—		✔						2
Vidal et al [89]	55	—	18.5 (N/A)	✔	✔					✔	2
Fauzi et al [90]	15	12 males and 3 females	27.5 (2.4)		✔						4
Tazarv et al [91]	20	13 males and 7 females	25 (N/A)		✔						4
Alfredo et al [92]	35	—	—		✔						4
Su et al [93]	18403	8565 males and 9838 females	118.5 (N/A)							✔	1
Wang et al [94]	15	12 males and 3 females	27.5 (2.4)		✔						1
Can and André [95]	14	9 males and 5 females	23 (N/A)		✔						3
Prajod et al [96]	135	—	—		✔						4
Ganesan et al [97]	15	12 males and 3 females	27.5 (2.4)		✔						7
Sun et al [98]	21	—	23 (2.91)		✔					✔	2
Neigel et al [99]	103	91 males and 12 females	21.8 (1.9)	✔	✔			✔			4
Pogliaghi et al [100]	15	12 males and 3 females	27.5 (2.4)		✔						2
Jaiswal et al [101]	64	—	—		✔						1
Rashid et al [102]	15	12 males and 3 females	27.5 (2.4)		✔						7

XSL•FO
RenderX

Study	Sample (n)	Sex	Age (years), mean (SD)	Sleep	Physiological signals	Calorie intake or expenditure	Phone use	Activity	Location	Survey	Total feature types
Narwat et al [103]	15	12 males and 3 females	27.5 (2.4)		✔						3
Kafková et al [104]	15+	—	—		✔						2
Lopez et al [105]	166	—	21 (N/A)	✔	✔	✔		✔			5
Wilfred et al [106]	25	—	—		✔						2
Jaiswal et al [107]	60	—	—		✔						1
Gaitan-Padilla et al [108]	12	5 males and 7 females	—		✔						2
Gupta et al [109]	15	12 males and 3 females	27.5 (2.4)		✔						3
Beierle and Pryss [110]	15	12 males and 3 females	27.5 (2.4)		✔						4
Masrur et al [111]	15+	—	College students		✔						1
Sakanti et al [112]	15	12 males and 3 females	27.5 (2.4)		✔						6
Shedage et al [113]	15	12 males and 3 females	27.5 (2.4)		✔						7
Gaitan-Padilla et al [114]	5	4 males and 1 female	22.6 (0.55)		✔						2
Tanwar et al [115]	15	12 males and 3 females	27.5 (2.4)		✔						6
Gullapalli et al [116]	18	—	20 (N/A)		✔						1
Sadruddin et al [117]	15	12 males and 3 females	27.5 (2.4)		✔						6
Jahanjoo et al [118]	15	12 males and 3 females	27.5 (2.4)		✔						1
Parousidou et al [119]	15	12 males and 3 females	27.5 (2.4)		✔						6
Karpagam et al [120]	15	12 males and 3 females	27.5 (2.4)		✔						3
Sethia et al [121]	36	32 males and 4 females	21 (N/A)		✔						4

XSL•FO
RenderX

Study	Sample (n)	Sex	Age (years), mean (SD)	Sleep	Physiological signals	Calorie intake or expenditure	Phone use	Activity	Location	Survey	Total feature types
Hasanpoor et al [122]	15	12 males and 3 females	27.5 (2.4)		✔						1
Benita et al [123]	15	12 males and 3 females	27.5 (2.4)		✔						1
Hsu [124]	10	—	College students		✔						1
Carmisciano et al [125]	15	12 males and 3 females	27.5 (2.4)		✔						2
Warrier et al [126]	15	12 males and 3 females	27.5 (2.4)		✔						5
Calbert and Tonekaboni [127]	5	2 males and 3 females	College students		✔						4
Hoang et al [1]	15	12 males and 3 females	27.5 (2.4)		✔						6
Kumar et al [128]	15	12 males and 3 females	27.5 (2.4)		✔						6
Hasanpoor et al [129]	15	12 males and 3 females	27.5 (2.4)		✔						1
Le et al [130]	10	—	College students		✔						3
Fernandez et al [131]	30	15 males and 15 females	28 (N/A)		✔						1
Tanwar et al [132]	15	12 males and 3 females	27.5 (2.4)		✔						3
Huang et al [133]	15	12 males and 3 females	27.5 (2.4)		✔						1
Oh et al [134]	15	12 males and 3 females	27.5 (2.4)		✔						6
Thapa et al [135]	15	12 males and 3 females	27.5 (2.4)		✔						6
Abdelfattah et al [136]	15	12 males and 3 females	27.5 (2.4)		✔						6
Tsiampa et al [137]		—	College students		✔						1
Fazeli et al [138]	14	—	College students		✔	✔		✔			8

XSL•FO
RenderX

Study	Sample (n)	Sex	Age (years), mean (SD)	Sleep	Physiological signals	Calorie intake or expenditure	Phone use	Activity	Location	Survey	Total feature types
Subathra and Malarvizhi [139]	15	12 males and 3 females	27.5 (2.4)		✔						2
Shikha et al [140]	36	—	20 (N/A)		✔						3
Andreas et al [141]	15	12 males and 3 females	27.5 (2.4)		✔						6
Lee et al [21]	15	12 males and 3 females	27.5 (2.4)		✔						6
Kasnesis et al [142]	15	12 males and 3 females	27.5 (2.4)		✔						6
Ciharova et al [143]	42	13 males and 29 females	20.79 (N/A)		✔						2
Darwish et al [144]	15	12 males and 3 females	27.5 (2.4)		✔						3
Nuamah [145]	32	—	25.2 (2.3)		✔						2
Saylam and İncel [19]	700	—	College students	✔	✔			✔			4
Sa-nguannarm et al [146]	15	12 males and 3 females	27.5 (2.4)		✔						6
Nelson et al [147]	103	—	College students		✔		✔			✔	3
Dahal et al [148]	15	12 males and 3 females	27.5 (2.4)		✔						1
Aqajari et al [149]	11	4 males and 7 females	22.91 (5.05)		✔						1
Jiao et al [150]	32	14 males and 18 females	22.69 (3.73)		✔						1
Yuting and Rashid [33]	502	476 males and 26 females	College students	✔				✔		✔	3
Lotfi et al [151]	168	168 females	122.5 (N/A)					✔			3
Belwafi et al [23]	36	8 males and 28 females	21 (N/A)		✔						1
Patanè et al [152]	16	—	College students				✔	✔			3
Subathra et al [153]	46	40 males and 6 females	22 (N/A)		✔						2

XSL•FO
RenderX

Study	Sample (n)	Sex	Age (years), mean (SD)	Sleep	Physiological signals	Calorie intake or expenditure	Phone use	Activity	Location	Survey	Total feature types
Li et al [25]	177	89 males and 88 females	20.37 (2.97)		✔					✔	3
Van der Mee et al [154]	95	15 males and 80 females	20 (N/A)		✔					✔	2
Rosenbach et al [24]	60	20 males and 40 females	27.5 (5.6)		✔						3

[a]Not available.

XSL•FO
RenderX

Table . Details for studies conducting algorithm comparisons.

Study	Device used	Physiological or nonphysiological signals	Algorithm	Performance measure	Best performing algorithm	Validation
Bellante et al [38]	Wrist and chest devices	BVP[a], EDA[b], and ESP[c]	DT[d], bagging DT, RF[e], Extra Trees, AdaBoost[f] DT, SVM[g], KNN[h], LR[i], and LDA[j]	Accuracy and F_1-score	SVM	Leave-one-out cross-validation (LOOCV)
Iranfar et al [41]	Biopac BioNomadix System	EDA, RESP[k], ECG[l], and PPG[m]	LDA, SVM, RF, XGBoost[n], Isolation forest, and Bayesian ridge algorithm	Accuracy	XGBoost	Group k-fold cross-validation (k=10)
Mohammadi et al [42]	__[o]	ECG and EDA	KNN, DT, RF, SVM, and FCM[p]	Accuracy, sensitivity, and specificity	KNN	Train and test split
Mustafa et al [43]	SA9309M, AD8232, and MAX30205	HR[q], SC[r], and TEMP[s]	ANN[t], KNN, DT, and SVM	Accuracy	DT	Train and test split
Arsalan and Majid [44]	MUSE EEG[u], Shimmer GSR[v], and PPG optical pulse clip	EEG, GSR, and PPG	KNN, DT, RF, MLP[w], and SVM	Accuracy and F_1-score	SVM	LOOCV
Can et al [27]	Smartwatch and Empatica E4	EDA and HR	MLP, RF (n=100), KNN (n=3), SVM, and LR	Accuracy	RF and SVM	10-fold CV[x]
Panganiban and de Leon [49]	Smartphone and CorSense	PRV[y] from PPG	KNN, NN[z], SVM, RF, and AdaBoost	Accuracy	RF	Stratified k-fold CV
Gasparini et al [50]	Shimmer3 GSR	BVP	SVM linear kernel and CNN[aa]	Accuracy, precision, recall, and F_1-score	CNN	Train and test split
Yu and Sano [31]	Wrist device and Android phone data	ACC[ab], SC, and TEMP	LSTM[ac], combination of LSTM and CNN	MAE[ad] and statistical analyses	LSTM	5-fold CV
Han et al [52]	Shimmer3 ECG, Shimmer 3 GSR+, and Empatica E4	ECG, PPG, and GSR	KNN (k=1, 3, 5, 7, and 9), SVM, and Naïve Bayes classifier	Accuracy	SVM	10-fold CV
Liakopoulos et al [56]	Body sensors, wrist, and chest devices	ECG, EDA, and HR	CNN, SVM, KNN, RF, and NN	Accuracy and F_1-score	SVM	10-fold and LOSO[af] CV
Hssayeni and Ghoraani [58]	Wrist and chest devices	RESP, ECG, EMA[ag], EDA, TEMP, and ACC	Gradient tree boosting and CNN	MAE and r	CNN	LOOCV
Mishra et al [61]	Polar H7, Amulet wrist, and custom-made GSR sensor	HR, activity data, EMA prompts, and GSR	SVM and RF	Accuracy and F_1-score	SVM	LOOCV
Mishra et al [26]	Polar H10, Polar H7, and Empatica E4	HR and EDA	SVM and RF	Precision, recall, and F_1-score	SVM with HR, RF for HR and EDA	LOOCV
Bobade and Vani [18]	Wrist and chest devices	ACC, ECG, BVP, TEMP, RESP, EMG[ah], and EDA	KNN, LDA, RF, DT, AdaBoost, Kernel SVM, and ANN	Accuracy	ANN	LOOCV

XSL•FO
RenderX

Study	Device used	Physiological or nonphysiological signals	Algorithm	Performance measure	Best performing algorithm	Validation
Elzeiny and Qaraqe [67]	PPG sensor and Empatica E4	IBI[ai] and BVP	CNN, RF, Extra Trees, extremely randomized trees, and SVM	Accuracy	CNN and Extra Trees	CNN: 5-fold cross validation and ML:[aj] 10-fold cross validation
Prashant et al [70]	Wrist and chest devices	ECG	LDA, RF (100 base estimators), SVM (Gaussian kernel), and ANN	Accuracy	RF	Train and test split
Silva et al [72]	Microsoft Smartband 2	HR, SC, TEMP, calorie intake and expenditure, and sleep patterns	Logistic regression, NN, Naïve Bayes, SVM, RF, and KNN	Sensitivity and specificity	NN	Train and test split
Islam et al [73]	Fitbit Charge 2 and Android	HR, sleep, step count, GPS location, sound intensity, and light data	LR, KNN, SVM, and NN	Accuracy	SVM	10-fold CV
Zhu et al [28]	Empatica E4, Affectiva Q Curve, and Shimmer3	EDA, PPG, and ECG	SVM, RF, KNN, Naïve Bayes, and LR	Accuracy, recall, precision, and F_1-score	SVM	LOSO and 10-fold CV
Sepanloo et al [81]	Empatica E4 and Zephyr BioHarness 3 chest straps	HR, EDA, and TEMP	RF, gradient boosting classifier, and stacking models	Accuracy, precision, recall, F_1-score, and support	Stacking models	Stratified 5-fold CV
Gedam et al [2]	Empatica E4 and RespiBAN	ECG, GSR, and TEMP	KNN, SVM, DT, RF, AdaBoost, XGBoost[n], and gradient boosting	Accuracy, precision, recall, F_1-score, and AUC[ak]	XGBoost	Train and test split and 10-fold CV
Alfredo et al [92]	Empatica E4	TEMP, EDA, BVP, and salivary cortisol	SVM, AdaBoost, RF, LDA, and KNN	Accuracy	RF and KNN	Train and test split
Su et al [93]	—	Self-reports (PSQI[al], DASS-21[am], CD-RISC[an], and IPAQ)[ao]	RF LR, SVM, and FNN[ap]	Accuracy, specificity, and F_1-score	RF	Train and test split
Wang et al [94]	Empatica E4 and RespiBAN	HRV[ae]	SVM and KNN	Accuracy, F_1-score, recall, and precision	SVM	10-fold CV
Prajod et al [96]	RespiBAN, Empatica E4, TMSI Mobi, IOM biofeedback device, and Actiwave Cardio Monitor	ECG, EDA, BVP, and TEMP	RF, SVM, and MLP	F_1-score and accuracy	RF	LOSO
Narwat et al [103]	RespiBAN	EDA, ECG, and TEMP	CNN, KNN, and XGBoost	Accuracy, precision, recall, F_1-score, and support	CNN	—
Sadruddin et al [117]	Empatica E4 and RespiBAN	ECG, EDA, EMG, ACC, TEMP, and RESP	DT, XGBoost, LR, and LDA	Accuracy	XGBoost	10-fold CV
Jahanjoo et al [118]	Empatica E4 and RespiBAN	PPG	KNN, LDA, SVM, DT, RF, and AdaBoost	Accuracy	SVM	CV
Karpagam et al [120]	Empatica E4	ACC, EDA, and TEMP	RF and LR	Accuracy	RF	10-fold CV

XSL•FO
RenderX

Study	Device used	Physiological or nonphysiological signals	Algorithm	Performance measure	Best performing algorithm	Validation
Hsu [124]	Empatica E4	EDA	LDA, SVM, and KNN	Precision, recall, F_1-score, and accuracy	SVM	Train and test split
Calbert and Tonekaboni [127]	Hexoskin vests and Actigraph watches	HR, RESP, breathing volume, and movement	RF, KNN, XGBoost, and NN	Accuracy	RF	LOSO
Le et al [130]	Empatica E4	HR, EDA, and TEMP	SVM and KNN	F_1-score and accuracy	KNN	10-fold CV
Fernandez et al [131]	EEG Enobio device and the BIOPAC MP36	EEG	LightGBM[aq], CNN, KNN, and SVM	Accuracy	LightGBM	Train and test split and 5-fold CV
Shikha et al [140]	Empatica E4	EDA, PPG, and ACC	Gradient Boosting, SVM, KNN, RF, and EBM[ar]	Accuracy	Gradient boosting	—

XSL•FO
RenderX

Study	Device used	Physiological or nonphysiological signals	Algorithm	Performance measure	Best performing algorithm	Validation
Aqajari et al [149]	Samsung Galaxy Gear Sport watches	PPG	KNN, RF, and XGBoost	F_1-score	RF	5-fold CV

[a]BVP: blood volume pulse.
[b]EDA: electrodermal activity.
[c]ESP: echo squeezing protocol.
[d]DT: decision tree.
[e]RF: random forest.
[f]AdaBoost: adaptive boosting.
[g]SVM: support vector machine.
[h]KNN: k-nearest neighbor.
[i]LR: logistic regression.
[j]LDA: linear discriminant analysis.
[k]RESP: response.
[l]ECG: electrocardiogram.
[m]PPG: photoplethysmography.
[n]XGBoost: extreme gradient boosting.
[o]Not available.
[p]FCM: fuzzy c-means.
[q]HR: heart rate.
[r]SC: skin conductance.
[s]TEMP: temperature.
[t]ANN: artificial neural network.
[u]EEG: electroencephalogram.
[v]GSR: galvanic skin response.
[w]MLP: multilayer perceptron.
[x]CV: cross-validation.
[y]PRV: pulse rate variability.
[z]NN: neural network.
[aa]CNN: convolutional neural network.
[ab]ACC: accelerometer.
[ac]LSTM: long short-term memory.
[ad]MAE: mean absolute error.
[ae]HRV: heart rate variability.
[af]LOSO: leave-one-subject-out.
[ag]EMA: ecological momentary assessment.
[ah]EMG: electromyography.
[ai]IBI: interbeat interval.
[aj]ML: machine learning
[ak]AUC: area under the receiver operating characteristic curve.
[al]PSQI: Pittsburgh Sleep Quality Index.
[am]DASS-21: Depression Anxiety Stress Scales–21.
[an]CD-RISC: Connor–Davidson Resilience Scale.
[ao]IPAQ: International Physical Activity Questionnaire.
[ap]FNN: feedforward neural network.
[aq]LightGBM: light gradient boosting machine.
[ar]EBM: explainable boosting machine.

XSL•FO
RenderX

Table . Details for studies testing or comparing their own framework or conducting statistical analyses.

Study	Device used	Features used	Algorithm analysis	Performance measure	Results	Validation
Faro and Giordano [39]	ECG[a] wearable and wearable body sensor network	HR[b], activity, time, and location	ANN[c] and SOM[d] for proposed framework	Classification tool	Model successful	Train/test split
Faro et al [40]	ECG wearable and wearable body sensor network	HR	SOFM[e]	—[f]	Defined as accurate enough	Train/test split
Li and Sano [45]	Wrist	SC[g], TEMP[h], and ACC[i]	L2 and 1-norm regularized multitask least squares regression	Mean squared error and MAE[j]	Early fusion better	Train/test split
Cheadle et al [46]	SAM[k] activity wearable, EDA[l] sensor, and Empatica E4	EDA[l]	Linear regression	Statistical correlation	Support prior findings that perceived microaggressive discrimination increases negative emotion	—
Chen et al [47]	Personalized system and surveys	Survey questions	Proposed framework	MAE	—	—
Gupta et al [48]	RespiBAN and Empatica E4	ECG, EMG[m], TEMP, RESP[n], BVP[o], EDA, and ACC	CNN[p] and k-medoid clustering	Accuracy and execution time	Success	4-fold CV[q]
Azgomi et al [51]	Affectiva Q Curve and Nonin Wireless WristOx2 oximeter	SC, TEMP, ACC, HR, and blood oxygenation	Bayesian filtering with an expectation maximization (EM)	*t* test comparison	Success	—
Wu et al [53]	Wrist and smartphone	EDA, PPG[r], TEMP, and ACC	Proposed framework and SVM[s]	Accuracy	Framework proposed	5-fold CV
Jelsma et al [54]	Wrist-worn EDA sensor, Empatica E4, and smartphone	EDA	Econometric fixed-effects with robust SE regression approach	Statistical analyses	—	—
Lai et al [55]	Wearable body sensor network	TEMP and EDA	Proposed framework with Res-TCN[t] classifier	Accuracy	High accuracy	LOOCV[u]
Li and Sano [57]	Wrist	TEMP, SC, and ACC	MTL[v] linear regression model and k-means clustering for the proposed framework	MSE[w] and MAE	The framework can extract features better than feature crafting or static autoencoders, and temporal features demonstrated significantly higher precision than static and crafted features.	4-fold CV
Gil-Martin et al [59]	RespiBAN and Empatica E4	ACC, TEMP, RESP, ECG, EMG, EDA, and BVP	CNN	Accuracy and F_1	—	LOOCV
Han et al [60]	Wrist	EDA, TEMP, ACC, HR, and blood oxygenation	Adversarial networks and transfer learning	Accuracy	Disentangled adversarial transfer learning framework	LOOCV
Momeni et al [62]	Biopac system	ECG, RESP, PPG, and EDA	XGBoost[x] algorithm	Accuracy and F_1	—	Group Shuffle Split CV with 10 iterations.
Rashid et al [63]	Wrist-based PPG sensor	BVP	CNN	Accuracy and F_1	Success	LOOCV

XSL•FO
RenderX

Study	Device used	Features used	Algorithm analysis	Performance measure	Results	Validation
Yannam et al [64]	Smartphones (Android) and fitness trackers (eg, OnePlus Band)	User screen time, devices around user, mobile and application usage stats, mobile interaction, location data, HR, sleep data, and step counts	Proposed framework	—	—	—
Pakhomov et al [65]	Fitbit	HR and activity	*t* test, significance levels, and Spearman rank test	—	—	—
Holder et al [66]	Empatica E4	ACC, BVP, EDA, and TEMP	KNN[y], DT[z], and CNN	Accuracy and F_1	Single modality showed promise	LOOCV
Heo et al [68]	PPG sensor	HR	DT, RF[aa], Adaboosting[ab], 9-NN[ac], LDA[ad], SVM, gradient-boosting, and the proposed framework OMDP[ae]	Accuracy and F_1	OMDP	LOOCV
Kar et al [69]	Wrist and chest	ACC, EDA, and TEMP	Binary classifier based on GRU[af] and RNN[ag]	Precision, recall, F_1, and accuracy	Support the use of a modest set of signals that are easily collected on wearables.	
Samyoun et al [71]	Smart wrist devices	ECG, EDA, EMG, TEMP, and RESP	RF, Extra Trees (EXT), DT, LDA, LR[ah], and MLP[ai]	Accuracy and F_1	Chest better than wrist sensors, and a combination of both is better than just chest.	LOOCV
Vidal Bustamante et al [32]	Wearables, wristband actigraphy data, and smartphone-based self-report surveys.	Self-report surveys on physical health, daily consumption habits, positive and negative affect, studying behaviors, stress levels and sources, sociability and support, and actigraphy	Linear modeling and clustering	BIC[aj]	—	—
Wu et al [74]	Empatica E4	EDA, BVP, and HR	K-means model with 2 clusters	Silhouette score	Comparable to state-of-the-art unsupervised methods.	—
Tutunji et al [76]	Empatica E4	HR, SC, ST[ak], ACC, and surveys	Linear mixed-effects models, paired sample *t* test, and RF	Error rate	Individualized models combined EMA[al] with physiology performed best, while group-based models performed worse.	LOSO[am] and LOBO[an]
Abdul Kader et al [78]	Empatica E4	ACC, BVP, TEMP, EDA, HR, and HRV[ao]	DNN[ap]	Accuracy, precision, recall, F_1-score, and AUROC[aq]	Privacy-preserving stress detection system using federated learning, providing privacy to the patient's data.	CV

XSL•FO
RenderX

Study	Device used	Features used	Algorithm analysis	Performance measure	Results	Validation
Vos et al [29]	Empatica E4, Mobi, and RespiBAN	EDA, HRV, ECG, ACC, EDA, ST, HR, SPO2[ar], ACC, BVP, IBI[as], EMG, and RESP	RF, SVM, ANN, and XGBoost	Accuracy, precision, recall, and F_1-score	An ensemble ML[at] model trained on a synthesized multidataset to improve the generalization of prediction.	LOSO
Darwish et al [82]	Fitbit Sense 2, Flowtime, Movesense, Prana, and Sentio Solutions Feel Terapeutics	ECG, EDA, and RESP	RF, XGBoost, KNN, LR, DT, AdaBoost, Extra Trees, Bagging classifier, LDA, and QDA[au]	Accuracy, precision, recall, and F_1-score	Validated multimodal wearable data in controlled (WESAD)[av] and real-life (SWEET)[aw] datasets for binary and 5-class stress detection.	CV
Bloomfield et al [3]	Oura Ring	Sleep, surveys, ACC, HR, HRV, and RESP	Mixed-effects regression models	Coefficient and *P* value	Used sleep estimates from wearables in the prediction of perceived stress.	—
Nazeer et al [84]	Customized proposed STRESSCARE and stress detection sensor	ECG, EDA, BVP, EMG, TEMP, and sweat	XGBoost, DT, RF, and SVM	Accuracy and F_1-score	Wrist-worn sensors (2-class and 3-class) prediction model performed worse than chest sensors (2-class).	Exploring various combinations of input sensor data.
Xuanzhi et al [88]	Empatica E4 and RespiBAN	EDA and HRV	Attention mechanism-based XLNet model, BrainNet, Xception, EfficientNetB4, VGG19, ResNet-50, MobileNet, and InceptionV3	Accuracy, recall, precision, and F_1-score	Proposed attention mechanism-based XLNet model for continuous stress monitoring.	Train/test split and CV
Vidal et al [89]	Actigraphy	Sleep duration and self-reports on stress and sleep	Individual-level linear model with a Bayesian framework	Bayesian metrics (pd, UIs, ROPE, ESS, and R-hat)	Negative associations between sleep duration and perceived stress in participants.	Stable estimates of lead-lag associations.
Tazarv et al [91]	Samsung Galaxy Gear Sport	PPG, ACC, GYR[ax], and atmospheric pressure	SVM, XGBoost, and RF with a context-aware Deep Q-Network (DQN)	Recall	A model with a context-aware active learning strategy for fine-grained, personalized stress detection worked with fewer queries.	LOSO
Ganesan et al [97]	Empatica E4	ACC, PPG, ECG, EMG, EDA, RESP, and TEMP	DNN and 1D-CNN	ROC-AUC[ay], F_1-score, accuracy, latency, and memory	An optimized, cost-effective, real-time, and energy-efficient DNN model demonstrated superior performance.	—

XSL•FO
RenderX

Study	Device used	Features used	Algorithm analysis	Performance measure	Results	Validation
Neigel et al [99]	Oura Ring	HR, HRV, activity, and sleep	Mixed effects model	*P* value and regression coefficients	Heightened waking HR and max waking HR, alongside sleep HR, sleep HRV, activity patterns, and sleep phases, during periods coinciding with significant academic and societal events.	—
Pogliaghi et al [100]	Empatica E4	EDA and BVP	RF, XGBoost, and MTL	F_1-score and accuracy	The proposed MTL model improved compared to single-task models.	LOSO
Lopez et al [105]	Fitbits	Calories burned, HR, sleep, steps, and distance	AdaBoost	F_1-score	Aggregation levels of 4 and 12 hours performed best with the calories and sleep modalities outperforming other modalities.	LOSO
Wilfred et al [106]	Wyoware devices	EMG and GSR[az]	Transfer learning model networks with CNN compared with SVM, DNN, LSTM[ba], and CNN + LSTM	Accuracy, precision, recall, and F_1-score	The proposed stress detection tool, equipped with an IoT[bb] system and VR[bc], worked best.	—
Gaitán-Padilla et al [108]	customized wearable polymeric optical fiber sensor, fiber Bragg grating, and ECG sensor	Pulse and RESP	Bagged DT, KNN, DT, and SVM	Accuracy, precision, recall, and F_1-score	Used a low-cost wearable polymeric optical fiber sensor to classify stress.	Comparison
Gupta et al [109]	Empatica E4 and RespiBAN	ECG, PPG, and GSR	RF, SVM, LDA, KNN, NN, and DT	Accuracy, sensitivity, specificity, precision, F_1-score, Matthew's correlation coefficient, and Cohen kappa	Wrist-worn sensors performed less than chest-worn sensors.	LOOCV
Sakanti et al [112]	RespiBAN	ACC, ECG, EDA, EMG, TEMP, and RESP	Extreme gradient boosting	Accuracy	Evaluated extreme gradient boosting in stress classification with high accuracy.	—
Shedage et al [113]	Empatica E4 and RespiBAN	BVP, ECG, EDA, EMG, RESP, TEMP, and ACC	LR, DT, RF, and SEL[bd]	Accuracy	SEL worked for a generalized, personalized model. SEL: LR, DT, and RF as base model and RF as meta model.	—
Tanwar et al [115]	Empatica E4 and RespiBAN	ECG, EDA, EMG, ACC, TEMP, and RESP	XGBoost, LGBoost[be], and CatBoost[bf]	Accuracy	Evaluated the effectiveness of data fusion methods, an accuracy increases with increase in modalities, and 5 modalities had best performance.	Train/test split

XSL•FO
RenderX

Study	Device used	Features used	Algorithm analysis	Performance measure	Results	Validation
Gullapalli et al [116]	PPG sensors in consumer-grade earbud devices	HRV	RF	Accuracy, specificity, and sensitivity	Compared stress detection with the most prominent HRV library HeartPy.	—
Parousidou et al [119]	Empatica E4 and RespiBAN	ECG, EDA, EMG, ACC, TEMP, and RESP	LDA, log reg, DT, NB[bg], RF, GB, user-based splitting, single-attribute splitting, multiattribute splitting, single task learning, and MTL.	F_1-score	Personalized approach performed better in lab settings and worse in the wild, outperforming one-size-fits-all.	—
Sethia et al [121]	Empatica E4	IBI from HRV, BVP, EDA, and TEMP	GB, RF, DT, SVM, KNN, and XGBoost	Accuracy	EDA + BVP + HRV performed well with GB for 2-level and 3-level stress classification, with HRV and EDA being the most important features.	—
Benita et al [123]	Empatica E4	PPG	CNN	Accuracy	Developed a stress detection system investigating CNN.	Train/test split
Carmisciano et al [125]	Empatica E4 and RespiBAN	EDA and HR	FDA[bh], RF, and LM[bi]	Partial R-squared	FDA models generally fit better than LM and RF.	—
Warrier et al [126]	RespiBAN	ECG, EDA, EMG, RR, TEMP, and ACC	DNN and federated learning	Accuracy	Federated learning–based stress detection method, focused on privacy protection with high accuracy.	Train/test split
Hoang et al [1]	Empatica E4 and RespiBAN	ECG, EDA, EMG, ACC, TEMP, and RESP	XGBoost	F_1-score, precision, and recall	Personalization performed better	Train/test split
Hasanpoor et al [129]	Empatica E4	PPG	CNN	Accuracy	Optimized model of reduced size and space addressing resource constraints.	Train/test split
Tanwar et al [132]	Empatica E4 and RespiBAN	ECG, EMG, and RESP	A hybrid deep learning network consisting of long short-term memory and gated recurrent unit (LSTM-GRU) with an attention layer	Accuracy	Proposed well-performing personalized stress detection.	—
Huang et al [133]	RespiBAN	ECG	A hybrid model combining CNN and SVM	Accuracy	A hybrid model combining a CNN and SVM performed with high accuracy.	Train/test split

XSL•FO
RenderX

Study	Device used	Features used	Algorithm analysis	Performance measure	Results	Validation
Oh et al [134]	RespiBAN	ACC, ECG, EDA, EMG, TEMP, and RESP	Three CNN-based classifiers and an ensemble attention module	Accuracy	An ensemble-based stress detection model that used multimodal features and metadata to capture personalized patterns.	Train/test split
Thapa et al [135]	Empatica E4 and RespiBAN	ECG, EDA, EMG, ACC, TEMP, and RESP	Conducted experiments using 4 state-of-the-art LLMs:[bj] GPT (4 and 3.5-Turbo), Llama2, BioMistralDARE, and Gemini-Pro.	Accuracy and MAE	For LLMs, parameter size did not correlate with accuracy; smaller models such as GPT-3.5-Turbo performed comparably to larger ones like GPT-4, though these models overall performed worse.	—
Tsiampa et al [137]	Empatica E4	EDA	Statistical correlation analyses	Correlation	A relationship exists between EDA and stress levels related to social media content, with a strong correlation.	—
Fazeli et al [138]	Garmin vivoactive 4S	HR, HRV, number of floors climbed, BMR[bk] kilocalories, distance traveled, activity levels, SPO2, and RESP	RNN, LSTM, and MLP	Accuracy	Proposed a multimodal semisupervised framework for tracking physiological precursors of the stress response; Late-fusion + Supervised Training + Contrastive Regularization performed best.	—
Subathra and Malarvizhi [139]	Empatica E4	EDA and HR	K-means and agglomerative clustering	Silhouette score	Agglomerative clustering obtained in the proposed method outperformed.	—
Andreas et al [141]	Empatica E4 and RespiBAN	ECG, EDA, EMG, ACC, TEMP, and RESP	CNNs in conjunction with transfer learning	Accuracy	Proposed method's effectiveness outperformed state-of-the-art classification techniques in the field using transfer learning.	—
Lee et al [21]	Empatica E4 and RespiBAN	ECG, EDA, EMG, ACC, TEMP, and RESP	DNN augmented with attention mechanisms	Accuracy	Enhanced DNN capabilities by integrating both raw signals and human-engineered features altogether.	LOSO
Kasnesis et al [142]	Empatica E4 and RespiBAN	ECG, EDA, EMG, ACC, TEMP, and RESP features extracted by a temporal CNN.	TranSenseFuser is comprised of temporal convolutions followed by feature-level or sequence-level multihead attention.	Accuracy and F_1-score	Model performed well for stress prediction.	LOSO

XSL•FO
RenderX

Study	Device used	Features used	Algorithm analysis	Performance measure	Results	Validation
Ciharova et al [143]	VU-AMS	ECG and EDA	Bayesian ridge regression	Accuracy, F_1-score, and r2	Performance ranged from acceptable to good, but only for the presentation stressor, best algorithm performance was a weak relationship between the detected and observed score	LOSO
Darwish et al [144]	RespiBAN	ECG, EDA, and RESP	RF, XGBoost, KNN, LR, DT, AB[bl], ET, BAG[bm], QDA, LDA, and ensemble models using majority voting and weighted averaging.	Accuracy	Binary stress classification performed better than five-class classification	K-fold CV
Nuamah [145]	Empatica E4 and Tobii Pro Glasses 2	Vagally mediated heart variability measures (vmHRV) and task-evoked pupillary response (TEPR)	Mixed-effects modeling	r2	vmHRV measures and TEPR are sensitive enough to quantify psychophysiological responses to recurrent task-induced stress	—
Saylam and İncel [19]	Fitbit	Step counts, active minutes, HR, and sleep metrics	RF, XGBoost, LSTM, and regression	MAE	With MTL, RF had the lowest error while looking back 7 and 15 days	—
Sa-nguannarm et al [146]	Empatica E4 and RespiBAN	ECG, EDA, EMG, ACC, TEMP, and RESP	Bi-LSTM	Accuracy and F_1-score	The human lifelong monitoring model Bi-LSTM for stress behavior recognition performed well.	Train/test split
Nelson et al [147]	Smartphone	PPG	Mixed-effects modeling	r2	Smartphone-based PPG significantly covaries with self-reported stress and anxiety.	—
Dahal et al [148]	RespiBAN	HRV	RF	Accuracy	Identified person-specific stress events with an accuracy higher than 99% after a global training framework.	15-fold CV
Jiao et al [150]	PL3516 Powerlab 16/35 with TN1012/ST Pulse Transducer	PRV[bn]	SVM model with linear and radial basis function kernel	Accuracy	Developed a pulse rate variability detection model with RFE[bo] feature selection.	5-fold CV
Belwafi et al [23]	EEG[bp] sensor	EEG	Statistical thresholding mechanism on EEG bands	Accuracy, precision, recall, and F_1-score	Proposed statistical thresholding mechanism on EEG bands approach achieved an average accuracy of 88.89%.	—

XSL•FO
RenderX

Study	Device used	Features used	Algorithm analysis	Performance measure	Results	Validation
Patane et al [152]	Smartphone	Phone call duration, conversation, physical activity, app usage, and academic deadlines	RNN, Bi-LSTM, transformer with prompt tuning	MAE and MSE	Personalized mental well-being monitoring with RNN, Bi-LSTM, and transformer with prompt tuning, where prompt-based adaptation achieved lower prediction error.	Train/validation/test split a 70%-10%-20% ratio.
Subathra et al [153]	Custom-built wrist device	HR and EDA	Bi-LSTM	Accuracy and F_1-score	Developed a wearable band, in Bi-LSTM, got F1-score of 99.38% and 98.88% in multiple datasets.	Train/validation/test split a 70%-10%-20% ratio.
Li et al [25]	PPG sensor	DASS-21[bq] stress score, PRV, and dPPG	1DCNN-Bi-LSTM, cross-attention, and XGBoost	MAE and RMSE[br]	Analysis found fusion of PRV and dPPG signals yielded best detection performance.	5-fold CV
Van der Mee et al [154]	Garmin smartwatch	Garmin HRV-derived stress score and mood EMAs.	Firstbeat analytic algorithms, mixed-effects regression, logistic multilevel models, and ANOVA	AUC and statistical significance	Analysis found Garmin Stress Score was associated with high- and moderate-intensity positive mood; it was not associated with states of high arousal negative mood.	Statistical association analysis

XSL•FO
RenderX

Study	Device used	Features used	Algorithm analysis	Performance measure	Results	Validation
Rosenbach et al [24]	Garmin Vivosmart 4 and Polar H10 chest strap	Garmin stress score, HRV, and HR	Linear mixed effect model	Statistical significance	Analysis found HR showed the strongest association with self‐reported stress, while the Garmin stress score demonstrated only marginal predictive value.	Statistical association analysis

[a]ECG: electrocardiogram.
[b]HR: heart rate.
[c]ANN: artificial neural network.
[d]SOM: self-organizing map.
[e]SOFM: self-organizing feature map.
[f]Not available.
[g]SC: skin conductance.
[h]TEMP: temperature.
[i]ACC: accelerometer.
[j]MAE: mean absolute error.
[k]SAM: Self-Assessment Manikin.
[l]EDA: electrodermal activity.
[m]EMG: electromyography.
[n]RESP: response.
[o]BVP: blood volume pulse.
[p]CNN: convolutional neural network.
[q]CV: cross-validation.
[r]PPG: photoplethysmography.
[s]SVM: support vector machine.
[t]Res-TCN: residual temporal convolutional network.
[u]LOOCV: leave-one-out cross-validation.
[v]MTL: multitask learning.
[w]MSE: mean squared error.
[x]XGBoost: extreme gradient boosting.
[y]KNN: k-nearest neighbor.
[z]DT: decision tree.
[aa]RF: random forest.
[ab]Ada-boosting: adaptive boosting.
[ac]NN: neural network.
[ad]LDA: linear discriminant analysis.
[ae]OMDP: optimized model decision process.
[af]GRU: gated recurrent unit.
[ag]RNN: recurrent neural network.
[ah]LR: logistic regression.
[ai]MLP: multilayer perceptron.
[aj]BIC: Bayesian information criterion.
[ak]ST: skin temperature.
[al]EMA: ecological momentary assessment.
[am]LOSO: leave-one-subject-out.
[an]LOBO: leave-one-batch-out.
[ao]HRV: heart rate variability.
[ap]DNN: deep neural network.
[aq]AUROC: area under the receiver operating characteristic curve.
[ar]SPO2: peripheral capillary oxygen saturation.

XSL•FO
RenderX

[as]IBI: interbeat interval.
[at]ML: machine learning.
[au]QDA: quadratic discriminant analysis.
[av]WESAD: Wearable Stress and Affect Detection.
[aw]SWEET: Stress in the Wild and Everyday Environment.
[ax]GYR: gyroscope.
[ay]ROC-AUC: receiver operating characteristic–area under the curve.
[az]GSR: galvanic skin response.
[ba]LSTM: long short-term memory.
[bb]IoT: internet of things.
[bc]VR: virtual reality.
[bd]SEL: stacked ensemble learning.
[be]LGBoost: Light Gradient Boosting Machine.
[bf]CatBoost: categorical boosting.
[bg]NB: naive Bayes.
[bh]FDA: functional data analysis.
[bi]LM: linear model.
[bj]LLM: large language model.
[bk]BMR: basal metabolic rate.
[bl]AB: adaptive boosting.
[bm]BAG: bootstrap aggregating.
[bn]PRV: pulse rate variability.
[bo]RFE: recursive feature elimination.
[bp]EEG: electroencephalogram.
[bq]DASS-21: Depression Anxiety Stress Scale–21 item.
[br]RMSE: root mean squared error.

Critical Appraisal of Individual Sources of Evidence

Although critical appraisal is not required for scoping reviews, we conducted an assessment of study quality to better contextualize the strengths and limitations of the included evidence. To address the quality of each paper, we scored every paper across 4 categories on a scale from 0 to 2 as described in Multimedia Appendix 2 and shown in Multimedia Appendix 3. Given the diverse study designs among the extracted papers, we adopted a methodology similar to that used by De Angel et al [155]. This approach integrates the AXIS appraisal tool [156] for cross-sectional studies with the Newcastle-Ottawa Scale [157] for longitudinal studies. Papers were assessed using a 3-point scoring system: 2 points for fully meeting the criteria, 1 point for partial fulfillment, and 0 points for nonfulfillment.

Effect measures extracted from the included studies consisted of accuracy, F_1-score, sensitivity, specificity, precision, recall, and other performance metrics reported for stress detection. These measures were used to compare model performance across studies. For population characteristics, the mean age and corresponding SDs were extracted whenever available.

Synthesis of Results

Due to differences in study designs, methodologies, and outcome reporting, results were synthesized descriptively. Key study characteristics, signals measured, algorithms used, and sensor types were organized into structured tables to enable comparison across studies. Frequencies of the most commonly measured signals, best-performing algorithms, and most-used sensors were calculated and visualized using bar plots. Missing summary statistics were extracted as reported, with no additional transformations applied. No meta-analysis, subgroup analysis, or meta-regression was conducted; instead, the synthesis focused on identifying overarching trends across the included studies. Because the focus of this review was to characterize stress detection methods used among college-aged populations, we extracted data elements that were directly relevant to the review objectives, including participant characteristics, sensor types, physiological signals, analytical methods, and model performance outcomes. Broader intervention-related data items (eg, intervention protocols, adverse event reporting, and clinical outcome metrics) did not apply to the observational and experimental studies included in this review. Therefore, the extraction approach was intentionally streamlined to ensure consistency, interpretability, and comparability across heterogeneous study designs. In addition, we developed an evidence gap map to conceptually organize and summarize the literature across study conditions, methodological enablers, analytical approaches, barriers, and outcomes, highlighting recurring patterns as well as persistent gaps, following a prior standardized method [158].

Ethical Considerations

This study is a systematic review of previously published literature and did not involve the collection of primary data from human participants. No new data were generated, and no individuals were directly recruited, observed, or intervened upon as part of this research. Accordingly, a formal review by an Institutional Review Board or Research Ethics Board was not sought. This determination is consistent with standard guidance that systematic reviews relying exclusively on publicly available,

XSL•FO
RenderX

previously published data do not constitute human participant research requiring ethics board oversight.

All included studies were previously published in peer-reviewed journals and were assumed to have undergone appropriate ethical review by their respective authors and institutions before publication. No personally identifiable information was accessed, extracted, or reported at any stage of this review. The conduct of this review adhered to the ethical principles outlined in the World Medical Association Declaration of Helsinki and complied with applicable institutional, regional, and international standards for research integrity.

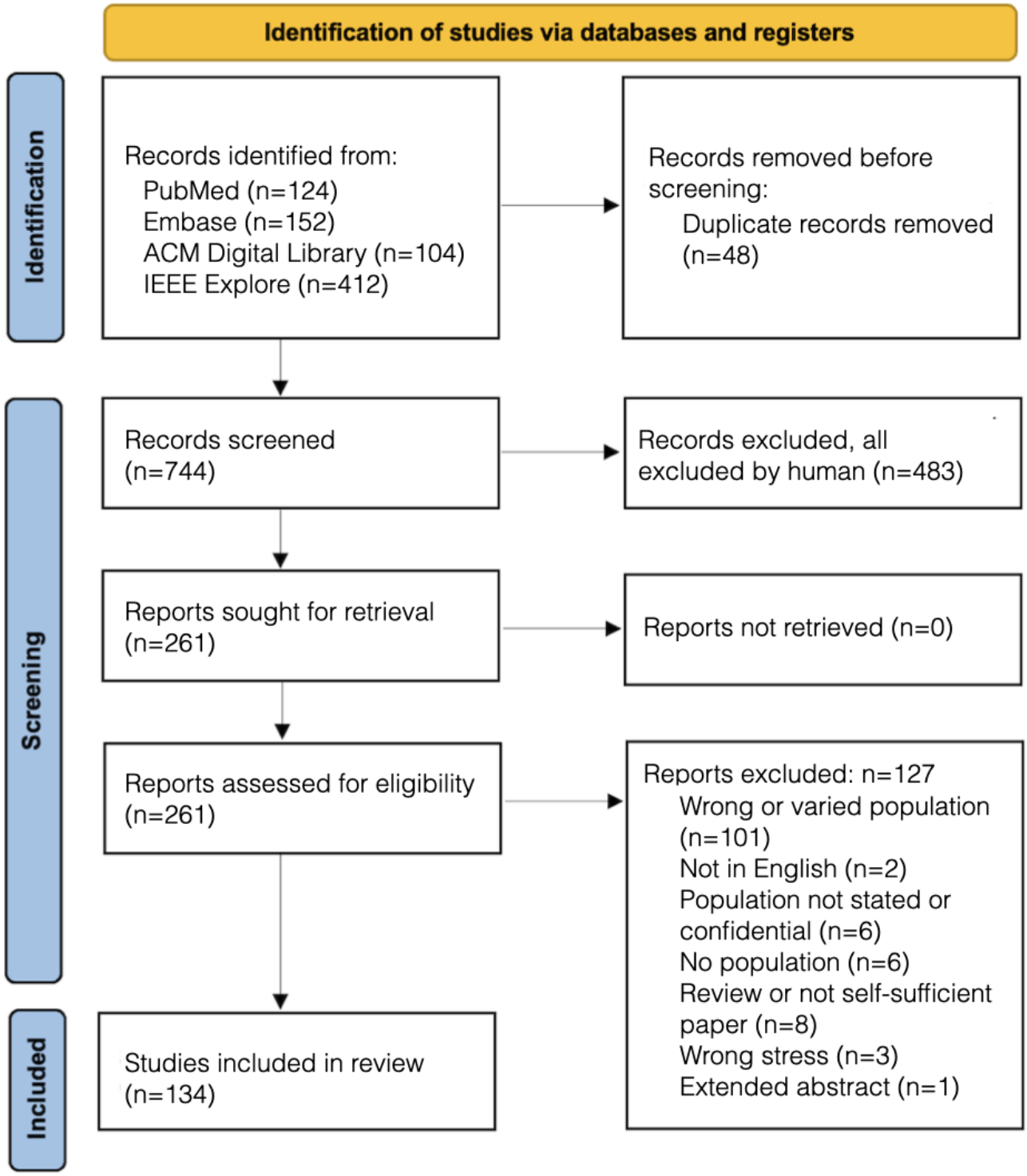
Results

Selection of Sources of Evidence

Records were screened from IEEE Xplore, ACM Digital Library, Embase, and PubMed, with most records coming from technical journals. A total of 134 studies were included in the review out of the original 792 records, as illustrated in Figure 1 and Multimedia Appendix 4. Forty-eight records were removed after deduplication. Of the remaining records, 483 were excluded after 744 abstracts were screened for relevance. In total, 127 records were excluded after 261 full texts were screened for relevance and correct population. Summary characteristics of the final 134 included studies are provided in Figure 1 and Table 1.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for study selection from medical and computer science databases.



Demographic and Geographic Characteristics of Included Studies

Our population of interest included college-aged students aged 18 - 24 years. In terms of sex demographics, about 72.4% (97/134) of studies specified the number of participants who were male, female, or nonbinary. Among the selected studies, 2 of 134 (1.5%) studies [39,137] failed to mention a sample size. Across the studies that reported sex distribution, most had a higher proportion of male participants than female, indicating a demographic imbalance that may limit the generalizability of findings. In terms of racial demographics, in 42 papers published from 2020 to 2022, about 9.5% (n=13) of papers included race distribution across their sample population, and 21 (50%) studies included other relevant health information, including preexisting conditions, mental health, and underlying illnesses. From papers published from 2020 to 2025, 26 (19.4%) studies were conducted in Europe, 14 (10.4%) studies were conducted in Asia, 2 (1.5%) studies were conducted in the Middle East, 21 (15.7%) studies were conducted in the United States, 3 (2.2%) studies were conducted in South America, and other studies did not explicitly mention where they were conducted. The higher

XSL•FO
RenderX

number of studies conducted in Europe and the United States compared to Asia and other regions suggests regional variations in digital health adoption, research funding, and accessibility of wearable technologies. These differences may influence trends in stress detection research, highlighting the need for region-specific digital health strategies to address varying technological infrastructures, health care priorities, and user needs.

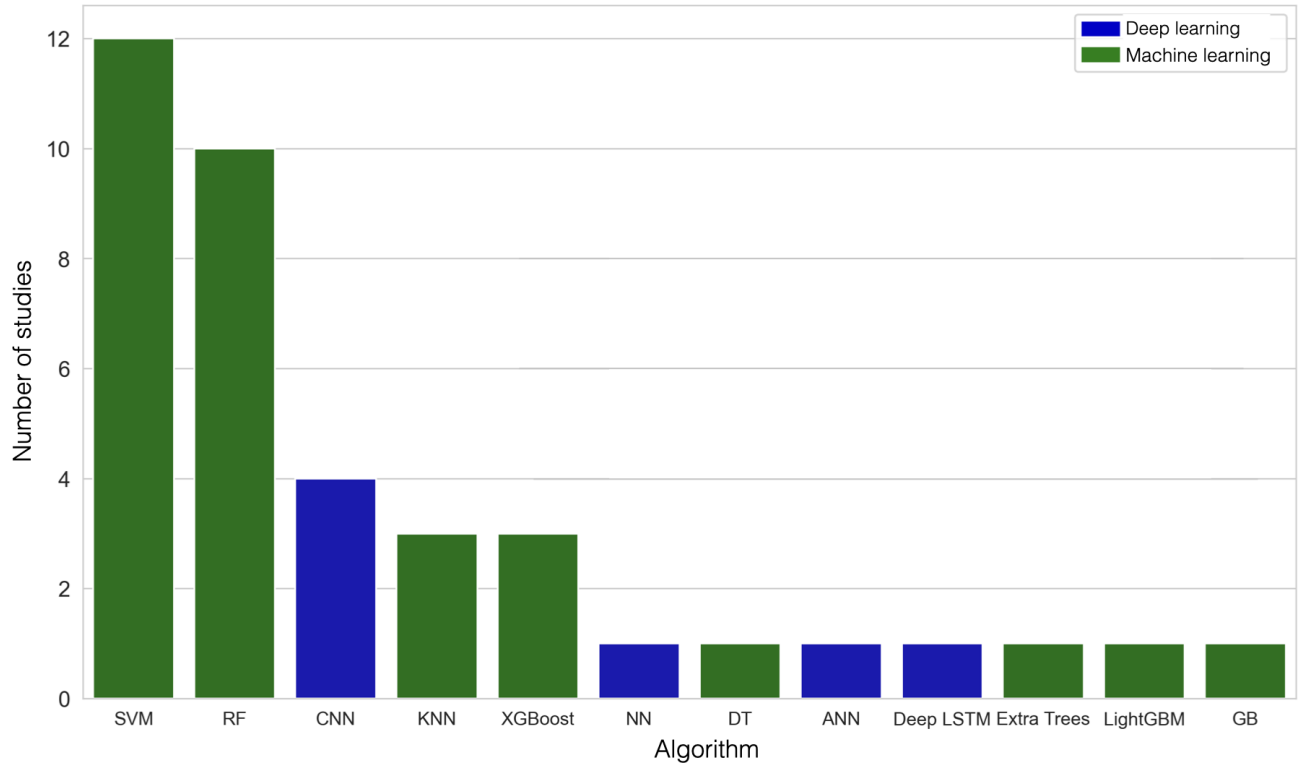
Study Design and Data Collection Characteristics

More than half (62.8%, n=134) of the studies used preexisting datasets to implement their method of stress measurement. The rest of the studies were experimental in nature and carried out "de novo" data collection. Seventeen studies [32,45,46,54,57,61,65,72,73] were longitudinal in nature, published from 2020 to 2022, and 10 studies [3,76,82,89,91,95,99,105,152,154] were longitudinal in nature, published from 2023 to 2025, meaning data were collected for the same study population over a period rather than collected at one time point cross-sectionally. These longitudinal data consist of repeated observations at the individual level rather than data collected at multiple time points across different populations. Individual-level effects are confounded with cohort effects in cross-sectional studies, so being able to isolate and study the effect of time as a repeated measure is critical. Of these longitudinal studies published from 2020 to 2022, 2 were clear in addressing how they handled missing data. These studies either imputed missing values with each person's channel-wise mean values of the day, where days with >25% sensor data missing were discarded [45], or removed missing data [57]. It is difficult to collect comprehensive, complete data from sensors longitudinally, where data are not always complete for each participant. About 6.6% (n=9) of studies included a recruitment method for participants. Two studies used volunteers, and 1 study invited participants to participate.

Approaches in Stress Detection Research

The extracted studies were classified into 3 primary methodological categories: algorithm comparisons (shown in Table 2), the development of custom stress measurement frameworks, and statistical analyses (illustrated in Table 3). Studies focusing on algorithm comparison primarily used 2 approaches: machine learning models, such as support vector machines (SVMs), random forest (RF), k-nearest neighbors, and extreme gradient boosting (XGBoost), which used handcrafted features for stress detection, or deep learning methods, such as convolutional neural networks (CNNs), to automatically extract relevant features [159]. Among the studies reviewed, SVM demonstrated the highest performance, with 33.3% (n=45) of papers identifying it as the best-performing algorithm, as illustrated in Figure 2. In comparison, 11.1% (n=15) of the studies reported CNN as the best-performing model [50,58,67,103]. One study evaluated 3 boosting algorithms—XGBoost, Light Gradient Boosting Machine, and CatBoost—tree-based ensemble methods that iteratively improve weak learners to enhance classification, evaluating the effectiveness of data fusion methods [115].

Figure 2. Best-performing algorithms across 36 studies comparing established methods. ANN: artificial neural network, CNN: convolutional neural network, DT: decision trees, Extra Trees: extremely randomized trees, GB: gradient boosting, KNN: k-nearest neighbor, LightGBM: light gradient boosting machine, LSTM: deep long short-term memory, NN: neural network, RF: random forest, SVM: support vector machine, XGBoost: extreme gradient boosting.



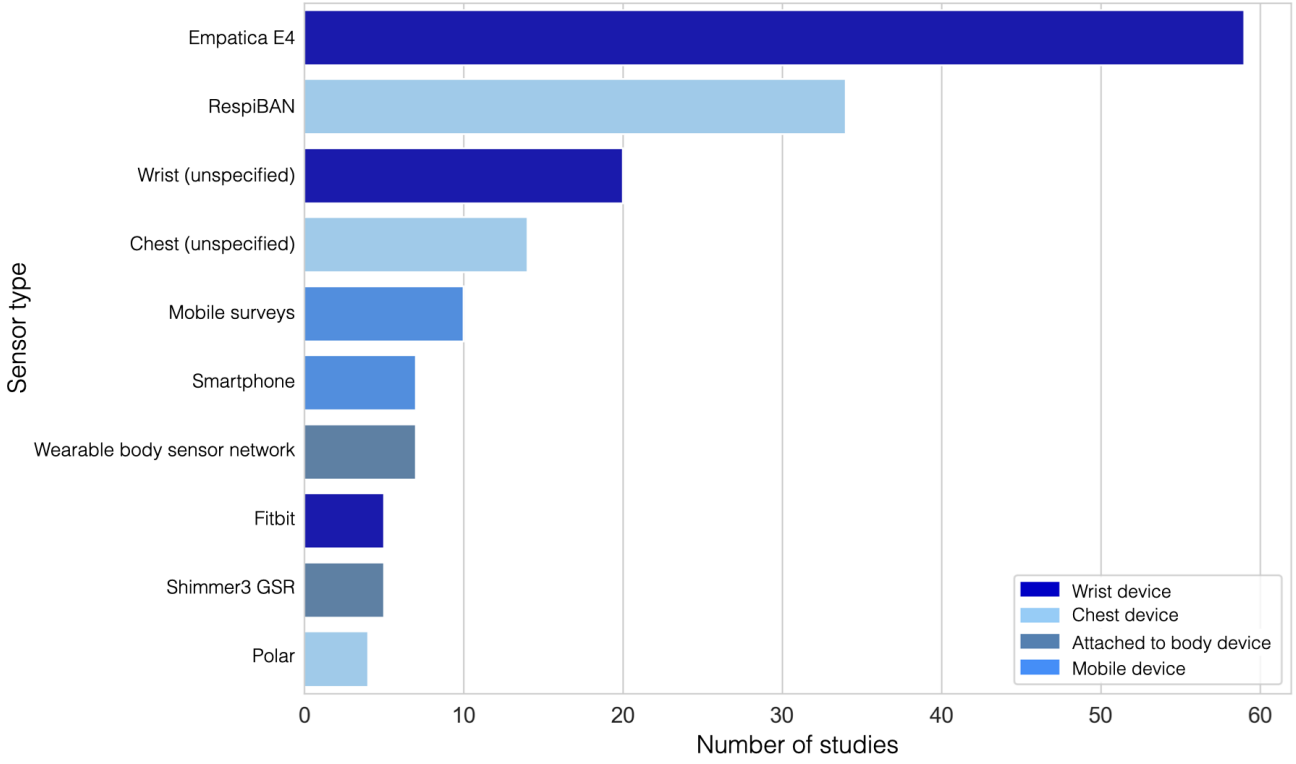
One paper [31] focused on comparing long short-term memory (LSTM) and a combination of LSTM and CNN. This study found LSTM alone to perform better. Two studies in Table 3 that focused on a single framework supported the use of a single

XSL•FO
RenderX

modality or a modest set of signals [66,69]. Studies that focused on the comparison of chest and wearable devices found chest devices to perform better [84,109], but chest devices in combination with wrist devices performed the best [71]. Most of these studies focused on time-agnostic algorithms, as shown in Table 2. We also found studies using wrist wearables (eg, Empatica, Microsoft Smartband 2, Fitbit Charge 2, and Samsung Galaxy Gear Sport Watches) and chest-worn devices, with core physiological signals such as EDA, galvanic skin response, HR, photoplethysmography, HRV, respiration, or temperature, evaluated using k-fold cross-validation, leave-one-out cross-validation, or leave-one-subject-out evaluation, and reported performance metrics such as F_1-score, accuracy, precision, and recall. In the "best" column, classic machine learning models were most often SVM, followed by RF, while deep learning wins were fewer (occasional CNN, deep neural network, and a single LSTM). Few studies in Table 2 incorporated nonphysiological or contextual signals [61,72,73]. Recent studies examining the association between sleep and stress have leveraged data from the Oura Ring [3,99]. Two recent studies using Garmin smartwatch–derived stress scores found significant associations with high- and moderate-intensity positive mood in 1 study [154], while another reported a stronger association of HR with self‐reported stress, and the Garmin stress score demonstrated marginal predictive value [24].

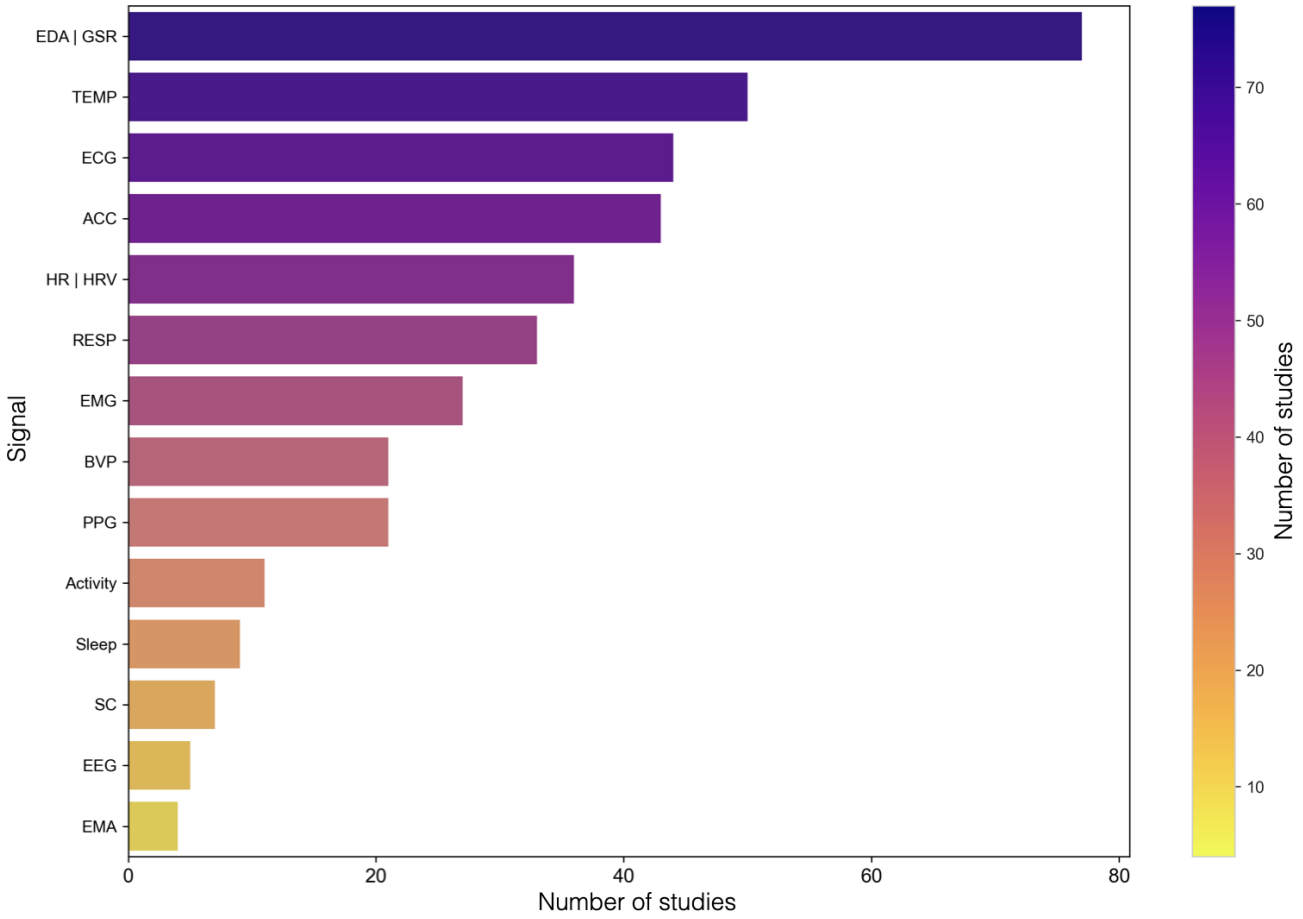
Studies mainly aggregated stress on a binary or 3-tier scale, meaning participants were either identified as stressed or not stressed, as opposed to being measured on a continuous scale. Here, a continuous scale captures stress fluctuation over time rather than binary or categorical labels. Sensors or tools used to measure physiological signals included various wrist, chest, and full-body sensors alongside mobile surveys. Figure 3 details the various devices used and shows that wrist sensors, in general, were the most widely used sensor type. About 72.4% (n=97) of the studies used well-validated stress tests or tasks for their models, such as the TSST [160], mental arithmetic tests, video stimuli, the Stroop color word test, startle response tests, cold-pressor tests, or public speaking, to reliably trigger stress responses while incorporating restful periods as a baseline [22]. About 8.3% (n=11) of the studies used self-reported SMS text messaging surveys in their supervised machine learning models. The various physiological features and signals measured are illustrated in Figure 4. The most common signal was EDA, appearing in 57.5% (n=77) of studies. Figure 4 shows the top signals measured per study, including instances where papers used multiple signals together.

Figure 3. Top 10 sensors used across all 134 studies.



XSL•FO
RenderX

Figure 4. Distribution of top physiological signals used in reviewed studies, including ecological momentary assessment (EMA) as a self-report measure. Many studies used multiple signals, which are counted in the bar plot. ACC: acceleration, BVP: blood volume pulse, ECG: electrocardiography, EDA: electrodermal activity, EEG: electroencephalogram, EMG: electromyography, GSR: galvanic skin response, HR: heart rate, HRV: heart rate variability, PPG: photoplethysmography, RESP: respiration, SC: skin conductance, TEMP: temperature.



Most Commonly Used Wearable Stress and Affect Detection Datasets in Stress Detection

Of the 62.8% of studies that used some preexisting datasets, around 80% (n=67) used the Wearable Stress and Affect Detection (WESAD) dataset, for instance, including papers published from 2020 to 2022 [18,38,42,43,48,55,56,58,59,63,66-71] or a few published from 2023 to 2025 [2,28,75,79,86,94,103,113,122,141]. This dataset was commonly referenced in papers included in this review. This dataset is publicly available and is a widely used dataset for stress and affect detection [161]. The mean age of participants is 27.5 years with a SD of 2.4 years. The sample included 3 females and 12 males for a total of 15 participants. Heavy smokers and pregnant women were excluded, and the participants were composed of graduate students. The signals collected include physiological and motion data from chest-worn and wrist-worn devices. Measurements include blood volume pulse, ECG, EDA, electromyography, respiration, body temperature, and 3-axis acceleration. The protocol used elicits 3 emotional states: baseline, stress, and amusement, followed by a meditation phase. Benchmarks for comparison used the well-studied stress induction method, the TSST, with 0.93 accuracy and 0.91 F_1-score for distinguishing stress, using a linear discriminant analysis classifier, using only chest-based physiological signals.

Although many papers used this same dataset, they experimented with different physiological signals as well as motion data when extracting features for modeling. Modeling and validation methods also varied. The algorithms with the best performance when applied to the WESAD dataset included SVM, RF, XGBoost, k-nearest neighbor, decision tree, deep neural network, self-supervised learning, artificial neural networks, large language models, and CNN. In addition to WESAD, recently published papers used other datasets, including SWELL [29], AffectiveROAD [81], VerBIO [96], S-TEST, or DS-3 [101].

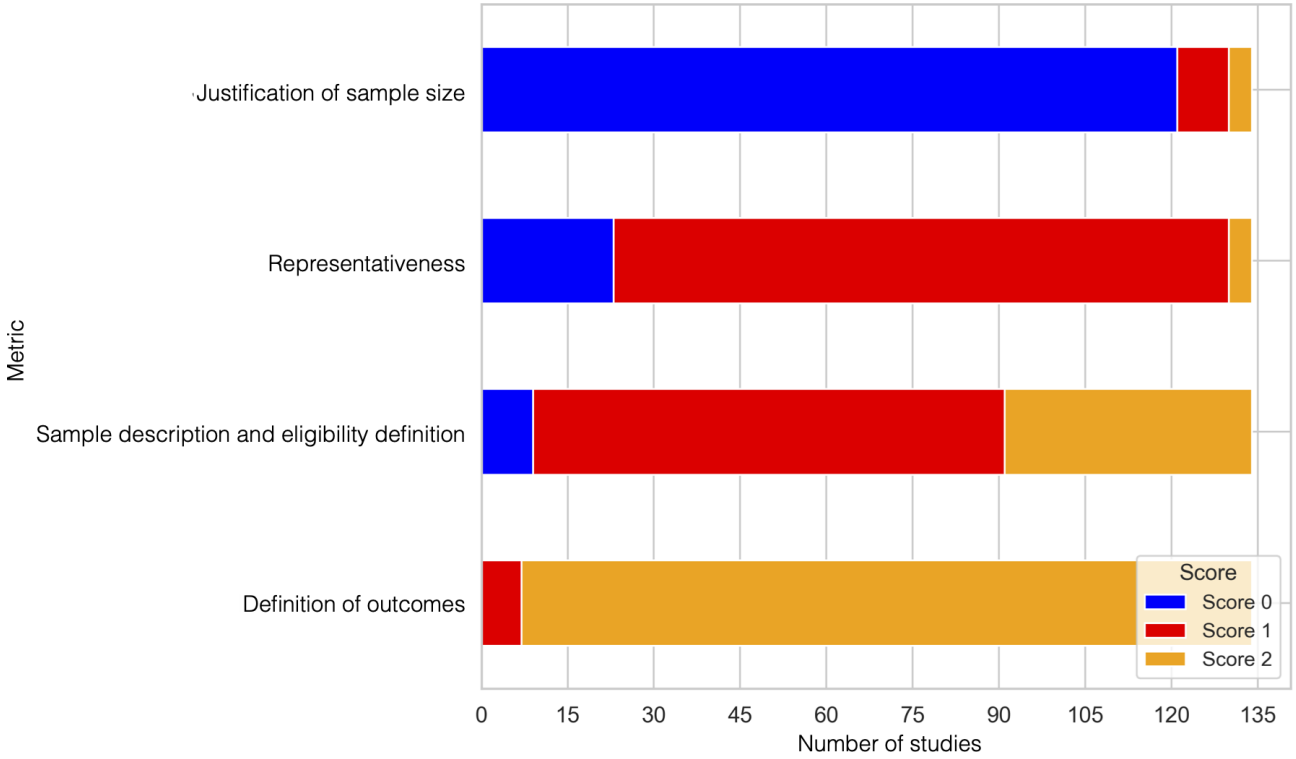
Quality Assessment of Included Studies

Figure 5 shows a breakdown of quality score assessments for all extracted papers, broken down into 4 categories. Papers were scored 0, 1, or 2 for each category. An explanation of each category's scoring is provided in Multimedia Appendix 2, and the individual score breakdown by category for each paper is provided in Multimedia Appendix 3. In general, outcomes and sample descriptions were clearly stated, with most papers having a quality score of 2. However, representativeness and justification of sample size were areas in which many papers did not perform as well. Representativeness was cited as a common issue across many papers, as samples were limited due to recruitment processes for participants or the data that were available. The samples were also limited by age due to the demographic of interest in this review. Around 27.6% (n=37)

XSL•FO
RenderX

of papers failed to give sex demographic information. Most papers analyzed used experimental data from other sources or open-source, publicly accessible datasets such as the WESAD dataset, which did not justify the chosen sample size. From papers published from 2020 to 2022, only 2% of papers failed to give sample size information; however, sample size justification was rarely given, although the papers that did address this issue cited their voluntary recruitment process as a limitation. Almost none of the studies analyzed did a power analysis to determine sample size before running the stress studies, which is a major shortcoming. Across recent papers published from 2023 to 2025, almost all clearly defined outcomes and described their samples, but very few addressed representativeness, and only 3 papers [24,143,154] justified their sample size published, highlighting a major gap in methodological rigor.

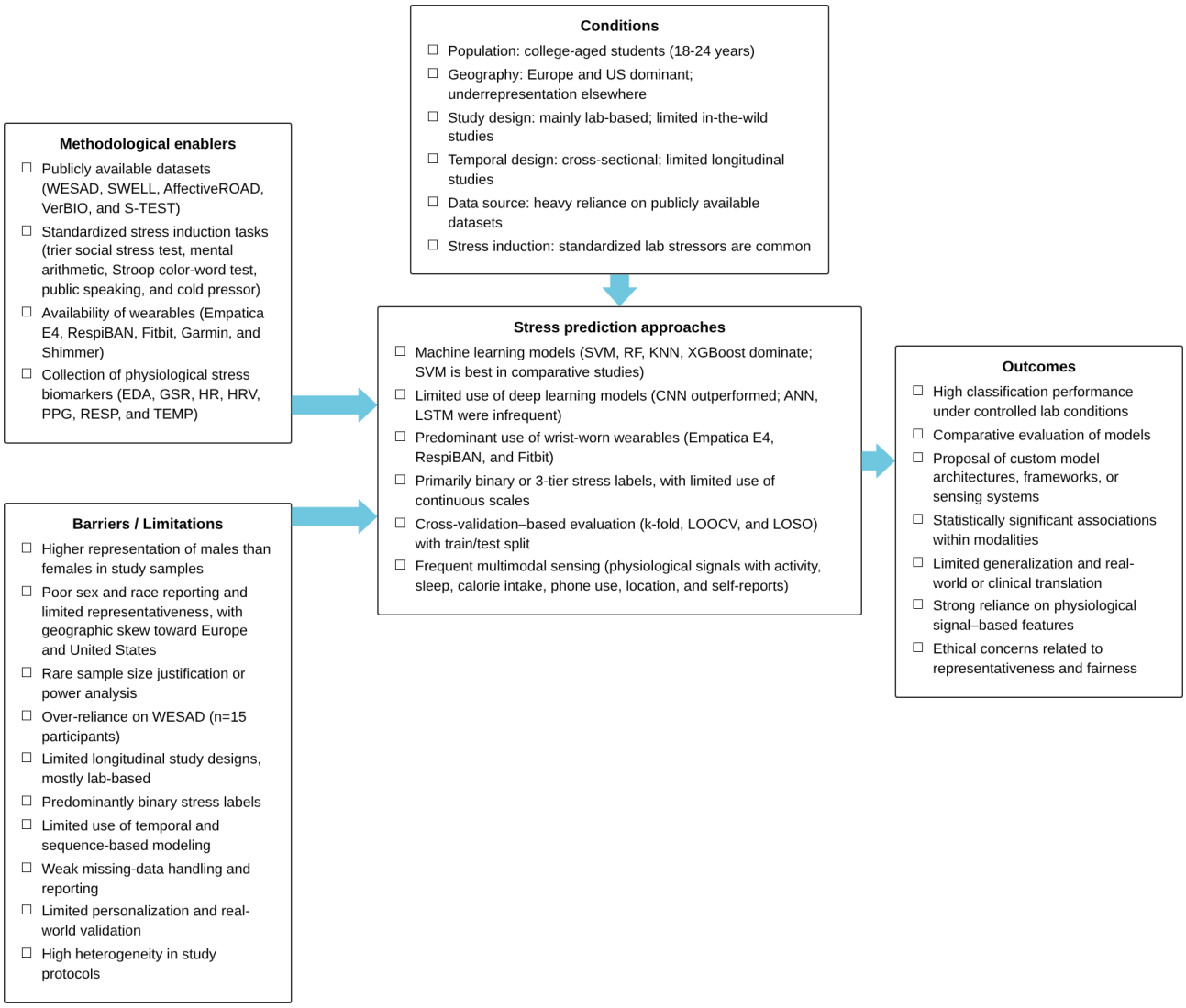
Figure 5. Quality of the literature in each domain. The figure shows the scoring across papers in each category from 0 to 2, with 0 indicating not fulfilled, 1 indicating partially fulfilled, and 2 indicating fulfilled.



Finally, these findings point to substantial heterogeneity and a meaningful risk of bias across the included studies. The wide variation in sample sizes, inconsistent reporting of demographic characteristics, limited disclosure of health information, and strong geographic skew toward Europe and the United States contribute to structural differences that complicate direct comparison of results. This heterogeneity is further shaped by the heavy reliance on the WESAD dataset, a publicly available dataset with only 15 predominantly male participants, with a mean age of 27.5 years, which results in many studies concluding a small and demographically narrow sample. Such repeated use of a single dataset increases the likelihood that reported model performance reflects the characteristics of WESAD participants rather than capturing variability among college-aged students. Accordingly, the synthesized findings should be interpreted with caution, acknowledging that both heterogeneity in study design and risk of bias in sampling and reporting may influence observed performance patterns and limit the extent to which results can be generalized. Using a relational synthesis approach, Figure 6 presents an evidence gap map that synthesizes methodological enablers, study conditions, stress prediction approaches, barriers, and outcomes observed across the included studies. The map illustrates a research landscape shaped by publicly available datasets, standardized in-laboratory stress protocols, and widespread use of wrist-worn physiological sensors. At the same time, it highlights recurring constraints including a predominance of laboratory-based study designs, heavy reliance on publicly available datasets, and limited demographic representativeness. While many studies report strong classification performance using classical machine learning models under controlled conditions, comparatively fewer examine temporal stress dynamics, personalization, or real-world deployment.

XSL•FO
RenderX

Figure 6. Gap map summarizing methodological enablers, study conditions, modeling approaches, barriers, and outcomes in wearable-based stress prediction studies among college students. ACC: acceleration, BVP: blood volume pulse, ECG: electrocardiography, EDA: electrodermal activity, EEG: electroencephalogram, EMA: ecological momentary assessment, EMG: electromyography, GSR: galvanic skin response, HR: heart rate, HRV: heart rate variability, PPG: photoplethysmography, RESP: respiration, SC: skin conductance, TEMP: temperature.



Discussion

Overview

In this scoping review, we examined how stress is measured among college-aged students using wearable technologies and machine learning methods between 2020 and 2025, to identify commonly used wearables, the most informative physiological signals, and the best-performing algorithms. Across the literature, we found that SVMs among traditional machine learning models and CNNs among deep learning models were the strongest performers for stress classification. Wrist-worn devices were the predominant sensor platform, and EDA was the most frequently measured and most informative signal. However, most studies relied on small, homogeneous samples, frequently used controlled laboratory datasets such as WESAD, and commonly used binary (stressed vs not stressed) labeling approaches, raising concerns about representativeness and ecological validity. Our quality assessment further revealed inconsistent demographic reporting, insufficient justification of sample sizes, limited attention to social determinants of stress, and substantial variation in how psychological stress was defined, elicited, and validated across studies.

Modeling Approaches for Stress Prediction

Regarding stress prediction model performance, the strong performance of SVMs can be attributed to their robustness in handling high-dimensional physiological data [33,144], their ability to generalize well by maximizing the margin between classes, and their effectiveness in small and imbalanced datasets, which are common in stress detection studies [162]. Additionally, the flexibility of SVM in using different kernel functions [163] allows them to model complex, nonlinear relationships in physiological signals without requiring deep feature extraction. These advantages likely contribute to their superior performance compared with other traditional machine learning models in stress classification. However, SVMs are computationally expensive and may not be practical for real-time applications [164]. More efficient and scalable approaches are needed to enhance practicality in the field. Deep learning models, particularly CNNs, outperformed traditional machine learning approaches in comparative analyses [82,85]. Although

XSL•FO
RenderX

CNNs capture spatial patterns in temporal data, they do not have memory in their architecture, reducing their effectiveness on longitudinal temporal data [165], indicating a need for algorithms that explicitly model temporal patterns, such as RNNs [74]. One study focusing on the comparison of various machine learning and deep learning methods attempted to use a version of an RNN in the form of an LSTM. This paper reported the greatest performance with LSTM alone, as opposed to a combination of LSTM and CNN, indicating some value in noting and using temporal patterns. In addition, emerging evaluations of large language models for stress prediction [135] did not perform well and suggest that parameter count does not consistently correlate with performance. For example, GPT-3.5-Turbo performed comparably to GPT-4 on WESAD [109]. These findings indicate that identifying key biomarkers is essential for improving model efficiency [115]. From 2023 to 2025, published literature emphasized personalization and multitask learning to enhance stress-prediction performance and generalizability [70,79,98,107,112,127]. In addition, 1 study explored stress detection in a virtual reality environment integrated with an Internet of Things system, demonstrating the potential of immersive technologies for stress monitoring [85].

Wearable Technologies and Physiological Signals

Wrist wearables were most commonly considered [166] as they seem less encumbering than full body or chest wearables [22] while attaining better measurement of physiological signals than surveys or smartphones. Other wearable sensors used across studies included chest wearables, full body sensors, or some combination of chest and wrist wearable signals. We saw that EDA was the most frequently measured signal across papers and is important in stress detection [167], since it provides valuable information about a person's sympathetic nervous system activity, which is closely linked to emotional responses, including stress. Most papers used multiple signals in their model building, with EDA most commonly contributing to a more accurate model. For instance, building a stress detection model incorporating both HR and EDA [22,26,81] data might allow for a more comprehensive, accurate, and context-aware assessment of stress and other emotional responses. Ensuring the reliability and reproducibility of physiological measurements is crucial for real-world stress detection [26]. Variability in sensor accuracy, signal quality, and environmental factors can impact consistency [22]. Validating models across diverse settings improves generalizability and practical applicability [168].

Conceptualizing and Measuring Psychological Stress

We saw that most studies used a binary model of stress in which an individual is identified as either stressed or not stressed. A few studies extended beyond binary classification by using multiclass stress prediction (eg, 3-class [62] or 5-class [59,125] models), which allows a somewhat finer-grained view but still treats stress as discrete states. There is a need for a model more in line with how human stress manifests, such as a continuous scale [26,169]. For example, an individual might feel mildly stressed, which is worth noting and which cannot be captured on a binary scale of stress [150]. On a binary scale, mild stress may be interpreted as either diminished or heightened stress. A continuous scale for stress monitoring is valuable for capturing individual differences and for understanding the dynamic nature of stress [150].

We found a lack of detailed explanations on how psychological stress was identified. Accurately distinguishing psychological stress from other physiological responses is complex, as HR alone is insufficient for stress detection [154]. For example, HR alone cannot reliably indicate stress, as an elevated HR may result from various factors [170], such as jogging or facing an unprepared mathematics test. A stress detection model based solely on HR data could misclassify natural variations in HR, such as those caused by excitement or physical activity during social events, as stress, leading to inaccurate assessments [169]. One critical detail to note in studies of stress is the differentiation between physiological and mental stress. This distinction is complicated for wearable devices [154]. To accommodate this, studies need to look at a participant's resting data while they are confirmed to be stressed, as well as their accelerometer data, if necessary, to check movement patterns, and consider these factors while detecting significant stress moments [169]. One's activity must be noted to clearly identify psychological stress. Many studies used some well-validated stress tasks to account for this concern, but could benefit from clearer explanations of how their stress tasks accommodate this issue. These stress tasks mostly used tests such as mental arithmetic, Stroop test, public speaking, or cold-pressor tests, with participants putting their hands in ice water, to benchmark stress [22,26]. By contrast, other datasets (eg, "A Wearable Exam Stress Dataset for Predicting Cognitive Performance in Real-World Settings" [124]) inferred stress levels indirectly from examination grades, raising concerns about the accuracy of stress labeling. Studies that did not incorporate a stress task often used self-report surveys to monitor whether someone is stressed [168,171]. Self-report measures often face challenges with accuracy and completeness [172]. While frequent and timely survey prompts can improve accuracy, they do not fully address issues of completeness. Additionally, repeated survey checks may increase participant burden, potentially leading to survey fatigue and lower response rates [173]. There is also a need for better transparency regarding the wording of questions and the frequency of surveys to ensure consistency and minimize bias [174].

Concerns Related to Study Design and Reporting

When analyzing the quality of research, we saw a need for larger sample sizes [175]. Larger sample sizes help reduce bias, provide a better representation of the target population, and lower the impact of outlier participants [176]. We observed that many studies relied on the WESAD dataset [177], a widely used dataset for stress and affect detection. However, WESAD includes only 15 participants, making it a limited representation of the college student population. Additionally, since WESAD data were collected in a controlled laboratory setting, they do not reflect real-world ("in the wild") stress detection, where external factors and daily life variability play a significant role [171,178,179]. In fact, 1 study that used WESAD achieved strong performance under laboratory conditions but failed to generalize effectively in real-world settings [119], further underscoring the limitations of laboratory-based datasets.

XSL•FO
RenderX

Many studies did not report racial or ethnic demographics or have a representative sample regarding sex. This was a commonly identified issue within papers, as many samples relied on volunteers. Many papers also failed to report on other demographics of their samples besides sex or ethnicity, such as populations for exclusion. This includes excluding populations taking certain medications, populations with certain mental health histories, populations engaging in drug use, or pregnant populations. Knowing the populations for exclusion is crucial for replicability and transparency, as well as for bias detection and interpretation of results [180-182]. Although our population of interest was students, there is a need for more varied student demographics in samples regarding sex, race, and ethnicity, capturing different social determinants [183]. Given that stress is influenced by various social determinants [184,185], future studies should incorporate factors such as socioeconomic status, neighborhood context, physical environment, racial minority representation, and health-lifestyle interactions [186]. Including these elements would provide a more comprehensive understanding of stress in college students. One paper mentioned that its sample may not be representative because participants were recruited from an elite, private university [32]. Along these lines, there is a need for better justification of sample selection as well as sample size. Finally, missing data present a significant challenge in stress studies, affecting both comparability across studies and the reliability of findings [187]. The way missing data is handled, whether through imputation, exclusion, or other techniques, can influence study outcomes and lead to biased conclusions [188]. There is a need for more complete data and more detailed descriptions of how missing data were handled, particularly in longitudinal studies [189].

Relationship to Prior Reviews and Contribution of This Work

Prior literature reviews have explored various aspects of stress detection using wearable technology and machine learning. A meta-analysis examined the effectiveness of wearable AI in diagnosing and predicting stress among students, while emphasizing the need for real-world validation and improvements [190]. Another review categorized stress detection approaches based on different wearable sensor types and environments such as driving, studying, and working [191]. A separate study systematically assessed biosignal responses to psychological stress, analyzing electroencephalogram, ECG, EDA, HRV, respiration, and temperature to evaluate their reliability and consistency [192]. A prior review also examined machine learning techniques used in stress monitoring research, focusing on model generalization when training on public datasets [20]. Another review focused on wearable technologies and smart devices for detecting depression, anxiety, and stress, discussing physiological markers such as HRV, EDA, and electroencephalogram, along with their market availability [193]. Finally, a review analyzed physiological parameters such as HR, temperature, humidity, blood pressure, and speech, exploring various stress detection sensors and machine learning-based classification techniques [194]. Our scoping review extends this literature by specifically focusing on stress measurement in college-aged students, reviewing recent papers published from January 2020 to December 2025, analyzing common datasets, sensor types, and the best-performing machine learning algorithms used in research. We also evaluate weaknesses in current methodologies through a quality assessment while identifying best practices in study design, feature selection, sensor use, and algorithmic approaches.

Taken together, the findings of this scoping review highlight that progress in wearable-based stress detection for college-aged students [3,32,46,73] is constrained primarily by methodological and conceptual design choices rather than sensor availability for digital phenotyping of stress [195] or algorithmic capacity [18,28,30]. While multimodal physiological sensing, particularly EDA combined with cardiac measures, shows consistent promise [22,26], the field remains highly reliant on small, controlled datasets such as WESAD [177] and binary stress formulations that fail to capture the continuous [26,169], context-dependent nature of stress in students' daily lives [171]. Advancing this area will require a shift toward larger [175], more diverse cohorts that reflect different social determinants of health [186], and real-world datasets that support generalizable human behavior modeling [168,196]; along with transparent reporting of participant characteristics, exclusion criteria, and missing data handling [189]; and modeling approaches that explicitly account for temporal patterns [95], personalization [1], and contextual information from students' behavioral patterns [152]. These improvements are not only methodological but also ethical, and without representative samples and robust validation in real-world settings, stress detection systems might risk reinforcing bias [197] and producing misleading inferences when deployed in student populations [183]. By synthesizing recent evidence and identifying persistent gaps, this review provides a foundation for designing more reliable, interpretable, and equitable stress monitoring systems that can support just-in-time interventions and inform institutional strategies to improve student mental health [5].

Limitations

Our focused and systematic approach targeting stress in college students in recent years allows for a more detailed analysis. Recency allows for analysis of the most up-to-date and commonly used sensors as well as the newest algorithms. By systematically categorizing the approach taken by each study, along with the devices used and signals measured, we can synthesize the information, establish trends, and make conclusions about best-performing methods and practices. Many studies relied on commonly used datasets, such as WESAD. Using the same dataset across different research projects enables benchmarking, allowing for direct comparison of methodologies and an understanding of why results may vary across approaches. A common challenge in the reviewed papers was the inclusion of multiple populations or datasets within a single study. While our primary focus was on college students, some papers analyzed mixed populations or multiple datasets. However, as long as college students were included, these studies were still considered in our review. Many papers also used overlapping datasets such as the WESAD dataset, although different papers used different parts of the dataset along with different models. This may lead to some redundancy in findings. The commonly used dataset, WESAD, with only 15 participants, had limited sample sizes, introducing potential bias and reducing

XSL•FO
RenderX

the likelihood of capturing a truly representative population. Additionally, only studies published in English were included, as this was the language accessible to our reviewers, which may have led to the exclusion of relevant research.

Conclusions

This scoping review provides a focused synthesis of wearable- and digital tool–based stress detection research specifically among college-aged students, a population often overlooked or aggregated with broader adult samples in prior reviews. Current research highlights the need for larger and more diverse samples to improve representativeness, as many studies rely on a limited number of existing datasets, potentially leading to overlapping findings. Greater diversity in sex and ethnic demographics, along with clearer justification of sample sizes and improved demographic reporting, is essential for understanding population-level stress patterns. Methodologically, most studies conceptualized stress as a binary state (stressed vs not stressed), failing to capture variations in intensity, such as mild or moderate stress that can be chronic and clinically meaningful. Few studies used algorithms such as RNNs, which can capture temporal patterns, despite the importance of tracking stress progression over time. Greater emphasis on time-dependent modeling could enhance the understanding of how stress evolves. Many studies failed to clearly distinguish between psychological stress and physiological stress responses, despite the critical need for distinct measurement approaches. More precise definitions and methodologies are necessary to differentiate between these 2 aspects of stress effectively. In real-world settings, these limitations constrain the generalizability and clinical usefulness of stress detection systems.

To strengthen the credibility and generalizability of future research, studies should provide clear justifications for their sample sizes and, where possible, aim to recruit larger cohorts that reduce bias and improve statistical reliability. The field would also benefit from the development and use of more varied datasets, which can limit overlap across studies and reduce potential sources of bias. Increasing diversity in participant recruitment is essential; researchers should ensure representation across race, sex, socioeconomic status, and environmental contexts, as well as variation in behavioral and lifestyle factors such as sleep duration and efficiency, physical activity, phone usage, social media engagement, and mobility patterns. Detailed demographic reporting should accompany all studies to enhance transparency and enable meaningful comparisons across research efforts. Future analytical approaches should incorporate algorithms capable of capturing temporal patterns to model fluctuations in stress over time. Rather than relying solely on binary stress categorizations, researchers should develop models that characterize stress as a dynamic and progressive state, allowing for the detection of mild, moderate, and chronic stress levels. Clear explanations of baseline stress measurements are also needed to ensure that resting conditions are consistently defined and comparable across studies. Finally, stress prediction models should increasingly focus on personalization while maintaining robust privacy protections for participants.

Acknowledgments

We would like to thank librarian Alissa Cilfone and Lauri Fennell for their consultation regarding database search strategies and the development of search terms. We used a generative artificial intelligence (AI) tool (ChatGPT-5.2; OpenAI) to polish the initial draft of the manuscript and Microsoft 365 Word built-in tools for spell and grammar checks, solely for language refinement, proofreading, summarization, and reformatting to improve the clarity and readability of the manuscript. No generative AI tools were used to generate any scientific content, figures, results, analyses, or interpretations. All citations were identified, verified, and added manually by the authors, and no AI-generated references were used.

Funding

This study represents independent research funded by Northeastern University’s Project-Based Exploration for the Advancement of Knowledge (PEAK) Experience #2: The Base Camp Award and Northeastern University’s FY23 Transforming Interdisciplinary Experiential Research (Tier) 1 Seed Grant: assessing the scalability and feasibility of digitally phenotyping stress.

Authors' Contributions

AS, OBA, and JA contributed to the literature search and data extraction. AS, OBA, and JO contributed to data analysis and interpretation. All authors contributed to writing the manuscript, and all authors approved the manuscript. All authors guaranteed the integrity of the work. AS and OBA contributed equally to this work and are co-first authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms and phrases.
[DOCX File, 38 KB - mhealth_v14i1e64144_app1.docx]

Multimedia Appendix 2

XSL•FO
RenderX

Quality assessment scoring details.
[DOCX File, 38 KB - mhealth_v14i1e64144_app2.docx]

Multimedia Appendix 3
Quality scores by paper.
[DOCX File, 60 KB - mhealth_v14i1e64144_app3.docx]

Multimedia Appendix 4
Study key and publication information.
[DOCX File, 168 KB - mhealth_v14i1e64144_app4.docx]

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 250 KB - mhealth_v14i1e64144_app5.docx]

References

1. Hoang TH, Dang TK, Trang NTH. Personalized stress detection for university students using wearable devices. Presented at: 2025 19th International Conference on Ubiquitous Information Management and Communication (IMCOM); Jan 3-5, 2025; Bangkok, Thailand p. 1-7. [doi: 10.1109/IMCOM64595.2025.10857507]
2. Gedam S, Dutta S, Jha R. Analyzing mental stress in Indian students through advanced machine learning and wearable technologies. Sci Rep 2025 Jul 1;15(1):20610. [doi: 10.1038/s41598-025-06918-6] [Medline: 40595085]
3. Bloomfield LSP, Fudolig MI, Kim J, et al. Predicting stress in first-year college students using sleep data from wearable devices. PLOS Digit Health 2024 Apr;3(4):e0000473. [doi: 10.1371/journal.pdig.0000473] [Medline: 38602898]
4. Substance Abuse In College Students: Statistics & Rehab Treatment. American Addiction Centers. 2024. URL: https://americanaddictioncenters.org/blog/college-coping-mechanisms [accessed 2023-06-29]
5. Regehr C, Glancy D, Pitts A. Interventions to reduce stress in university students: a review and meta-analysis. J Affect Disord 2013 May 15;148(1):1-11. [doi: 10.1016/j.jad.2012.11.026] [Medline: 23246209]
6. Schmidt MV, Sterlemann V, Müller MB. Chronic stress and individual vulnerability. Ann N Y Acad Sci 2008 Dec;1148(1):174-183. [doi: 10.1196/annals.1410.017] [Medline: 19120107]
7. Can YS, Arnrich B, Ersoy C. Stress detection in daily life scenarios using smart phones and wearable sensors: a survey. J Biomed Inform 2019 Apr;92:103139. [doi: 10.1016/j.jbi.2019.103139] [Medline: 30825538]
8. Lo Martire V, Caruso D, Palagini L, Zoccoli G, Bastianini S. Stress & sleep: a relationship lasting a lifetime. Neurosci Biobehav Rev 2020 Oct;117:65-77. [doi: 10.1016/j.neubiorev.2019.08.024] [Medline: 31491473]
9. Avitsur R, Powell N, Padgett DA, Sheridan JF. Social interactions, stress, and immunity. Immunol Allergy Clin North Am 2009 May;29(2):285-293. [doi: 10.1016/j.iac.2009.02.006] [Medline: 19389582]
10. Buddhiprabha DDP, Shabbeer A, Veena N, Shailaja S. Stress and academic performance. Int J Indian Psychol 2016;3(3):71-82. [doi: 10.25215/0303.068]
11. Birch JN, Vanderheyden WM. The molecular relationship between stress and insomnia. Adv Biol (Weinh) 2022 Nov;6(11):e2101203. [doi: 10.1002/adbi.202101203] [Medline: 35822937]
12. Robinson L. Stress and anxiety. Nurs Clin North Am 1990 Dec;25(4):935-943. [doi: 10.1016/S0029-6465(22)02991-7] [Medline: 2235645]
13. Dhabhar FS. Effects of stress on immune function: the good, the bad, and the beautiful. Immunol Res 2014 May;58(2-3):193-210. [doi: 10.1007/s12026-014-8517-0] [Medline: 24798553]
14. Strath SJ, Rowley TW. Wearables for promoting physical activity. Clin Chem 2018 Jan;64(1):53-63. [doi: 10.1373/clinchem.2017.272369] [Medline: 29118062]
15. Spil T, Sunyaev A, Thiebes S, Van Baalen R. The adoption of wearables for a healthy lifestyle: can gamification help? 2017 Jan 4 Presented at: 50th Annual Hawaii International Conference on System Sciences (HICSS-50); Jan 4, 2017. [doi: 10.24251/HICSS.2017.437]
16. Passos J, Lopes SI, Clemente FM, et al. Wearables and internet of things (IoT) technologies for fitness assessment: a systematic review. Sensors (Basel) 2021 Aug 11;21(16):5418. [doi: 10.3390/s21165418] [Medline: 34450860]
17. Kaewkannate K, Kim S. A comparison of wearable fitness devices. BMC Public Health 2016 May 24;16(1):433. [doi: 10.1186/s12889-016-3059-0] [Medline: 27220855]
18. Bobade P, Vani M. Stress detection with machine learning and deep learning using multimodal physiological data. Presented at: 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA); Jul 15-17, 2020; Coimbatore, India p. 51-57. [doi: 10.1109/ICIRCA48905.2020.9183244]
19. Saylam B, İncel Ö. Multitask learning for mental health: depression, anxiety, stress (DAS) using wearables. Diagnostics (Basel) 2024 Feb 26;14(5):501. [doi: 10.3390/diagnostics14050501] [Medline: 38472973]

XSL•FO
RenderX

20. Vos G, Trinh K, Sarnyai Z, Rahimi Azghadi M. Generalizable machine learning for stress monitoring from wearable devices: a systematic literature review. Int J Med Inform 2023 May;173:105026. [doi: 10.1016/j.ijmedinf.2023.105026] [Medline: 36893657]
21. Lee H, Chang J, Jaewon K, Han B, Park SM. Developing an explainable deep neural network for stress detection using biosignals and human-engineered features. SSRN. Preprint posted online on Aug 5, 2024. [doi: 10.2139/ssrn.4881618]
22. Amin OB, Mishra V, Tapera TM, Volpe R, Sathyanarayana A. Extending stress detection reproducibility to consumer wearable sensors. arXiv. Preprint posted online on May 9, 2025. [doi: 10.48550/arXiv.2505.05694]
23. Belwafi K, Alsuwaidi A, Mejri S, Djemal R. Brain-inspired signal processing for detecting stress during mental arithmetic tasks. Brain Inf 2025 Dec;12(1):34. [doi: 10.1186/s40708-025-00281-y]
24. Rosenbach H, Itzkovitch A, Gidron Y, Schonberg T. Assessing stress level scores against wearables-driven physiological measurements. Stress Health 2025 Dec;41(6):e70125. [doi: 10.1002/smi.70125] [Medline: 41292097]
25. Li M, Li J, Chen Y, Hu B. Stress severity detection in college students using emotional pulse signals and deep learning. IEEE Trans Affective Comput 2025 Jul;16(3):1942-1954. [doi: 10.1109/TAFFC.2025.3547753]
26. Mishra V, Sen S, Chen G, et al. Evaluating the reproducibility of physiological stress detection models. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020 Dec;4(4):1-29. [doi: 10.1145/3432220] [Medline: 36189150]
27. Can YS, Gokay D, Kılıç DR, Ekiz D, Chalabianloo N, Ersoy C. How laboratory experiments can be exploited for monitoring stress in the wild: a bridge between laboratory and daily life. Sensors (Basel) 2020 Feb 4;20(3):838. [doi: 10.3390/s20030838] [Medline: 32033238]
28. Zhu L, Spachos P, Ng PC, et al. Stress detection through wrist-based electrodermal activity monitoring and machine learning. IEEE J Biomed Health Inform 2023 May;27(5):2155-2165. [doi: 10.1109/JBHI.2023.3239305] [Medline: 37022004]
29. Vos G, Trinh K, Sarnyai Z, Rahimi Azghadi M. Ensemble machine learning model trained on a new synthesized dataset generalizes well for stress prediction using wearable devices. J Biomed Inform 2023 Dec;148:104556. [doi: 10.1016/j.jbi.2023.104556]
30. Chen Q, Lee BG. Deep learning models for stress analysis in university students: a Sudoku-based study. Sensors (Basel) 2023 Jul 2;23(13):6099. [doi: 10.3390/s23136099] [Medline: 37447948]
31. Yu H, Sano A. Passive sensor data based future mood, health, and stress prediction: user adaptation using deep learning. Presented at: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) in conjunction with the 43rd Annual Conference of the Canadian Medical and Biological Engineering Society; Jul 20-24, 2020; Montreal, Canada p. 5884-5887 URL: https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=9167168 [accessed 2026-03-19] [doi: 10.1109/EMBC44109.2020.9176242]
32. Vidal Bustamante CM, Coombs G 3rd, Rahimi-Eichi H, et al. Fluctuations in behavior and affect in college students measured using deep phenotyping. Sci Rep 2022 Feb 4;12(1):1932. [doi: 10.1038/s41598-022-05331-7] [Medline: 35121741]
33. Yuting L, Rashid RABA. Beyond the books: how sleep, school belonging, and physical activity affect the mental health of students under academic stress. Acta Psychol (Amst) 2025 Aug;258:105213. [doi: 10.1016/j.actpsy.2025.105213] [Medline: 40609190]
34. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
35. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
36. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. Syst Rev 2021 Jan 26;10(1):39. [doi: 10.1186/s13643-020-01542-z] [Medline: 33499930]
37. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5(1):210. [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
38. Bellante A, Bergamasco L, Bogdanovic A, et al. EMoCy: towards physiological signals-based stress detection. Presented at: 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI); Jul 27-30, 2021; Athens, Greece p. 1-4. [doi: 10.1109/BHI50953.2021.9508611]
39. Faro A, Giordano D. Prognostics and management of mental stress by aiot monitoring and schlegel diagrams. Presented at: 2021 IEEE International Smart Cities Conference (ISC2); Sep 7-10, 2021; Manchester, United Kingdom p. 1-7. [doi: 10.1109/ISC253183.2021.9562844]
40. Faro A, Giordano D, Venticinque M. Finding the proper mental stress model depending on context using edge devices and machine learning. Presented at: 2020 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS); Jan 27-28, 2021; Bali, Indonesia p. 161-166 URL: https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=9359628 [accessed 2026-03-19] [doi: 10.1109/IoTaIS50849.2021.9359701]
41. Iranfar A, Arza A, Atienza D. ReLearn: a robust machine learning framework in presence of missing data for multimodal stress detection from physiological signals. Annu Int Conf IEEE Eng Med Biol Soc 2021 Nov:535-541. [doi: 10.1109/EMBC46164.2021.9630040] [Medline: 34891350]
42. Mohammadi A, Fakharzadeh M, Baraeinejad B. An integrated human stress detection sensor using supervised algorithms. IEEE Sensors J 2022;22(8):8216-8223. [doi: 10.1109/JSEN.2022.3157795]

XSL•FO
RenderX

43. Mustafa A, Alahmed M, Alhammadi A, Soudan B. Stress detector system using iot and artificial intelligence. Presented at: 2020 Advances in Science and Engineering Technology International Conferences (ASET); Feb 4 to Apr 9, 2020; Dubai, United Arab Emirates p. 1-6. [doi: 10.1109/ASET48392.2020.9118345]
44. Arsalan A, Majid M. Human stress classification during public speaking using physiological signals. Comput Biol Med 2021 Jun;133:104377. [doi: 10.1016/j.compbiomed.2021.104377] [Medline: 33866254]
45. Li B, Sano A. Early versus late modality fusion of deep wearable sensor features for personalized prediction of tomorrow's mood, health, and stress. Presented at: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) in conjunction with the 43rd Annual Conference of the Canadian Medical and Biological Engineering Society; Jul 20-24, 2020; Montreal, Canada p. 5896-5899. [doi: 10.1109/EMBC44109.2020.9175463]
46. Cheadle JE, Goosby BJ, Jochman JC, Tomaso CC, Kozikowski Yancey CB, Nelson TD. Race and ethnic variation in college students' allostatic regulation of racism-related stress. Proc Natl Acad Sci U S A 2020 Dec 8;117(49):31053-31062. [doi: 10.1073/pnas.1922025117] [Medline: 33229568]
47. Chen M, Xiao W, Li M, Hao Y, Hu L, Tao G. A multi-feature and time-aware-based stress evaluation mechanism for mental status adjustment. ACM Trans Multimedia Comput Commun Appl 2022 Feb 28;18(1s):1-18. [doi: 10.1145/3462763]
48. Gupta D, Bhatia MPS, Kumar A. Resolving data overload and latency issues in multivariate time-series IoMT data for mental health monitoring. IEEE Sensors J 2021 Nov 15;21(22):25421-25428. [doi: 10.1109/JSEN.2021.3095853]
49. Panganiban FC, de Leon FA. Stress detection using smartphone extracted photoplethysmography. Presented at: 2021 IEEE Region 10 Symposium (TENSYMP); Aug 23-25, 2021; Jeju, Republic of Korea p. 1-7. [doi: 10.1109/TENSYMP52854.2021.9550905]
50. Gasparini F, Grossi A, Bandini S. A deep learning approach to recognize cognitive load using PPG signals. 2021 Jun 29 Presented at: PETRA '21: Proceedings of the 14th PErvasive Technologies Related to Assistive Environments Conference; Jun 29, 2021; Corfu, Greece p. 489-495 URL: https://dl.acm.org/doi/proceedings/10.1145/3453892 [accessed 2026-03-19] [doi: 10.1145/3453892.3461625]
51. Azgomi HF, Cajigas I, Faghih RT. Closed-loop cognitive stress regulation using fuzzy control in wearable-machine interface architectures. IEEE Access 2021;9:106202-106219. [doi: 10.1109/ACCESS.2021.3099027]
52. Han HJ, Labbaf S, Borelli JL, Dutt N, Rahmani AM. Objective stress monitoring based on wearable sensors in everyday settings. J Med Eng Technol 2020 May 18;44(4):177-189. [doi: 10.1080/03091902.2020.1759707]
53. Wu J, Zhang Y, Zhao X. Stress detection using wearable devices based on transfer learning. Presented at: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); Dec 9-12, 2021; Houston, TX p. 3122-3128. [doi: 10.1109/BIBM52615.2021.9669904] [Medline: 36704639]
54. Jelsma EB, Goosby BJ, Cheadle JE. Do trait psychological characteristics moderate sympathetic arousal to racial discrimination exposure in a natural setting? Psychophysiology 2021 Apr;58(4):e13763. [doi: 10.1111/psyp.13763] [Medline: 33462861]
55. Lai K, Yanushkevich SN, Shmerko VP. Intelligent stress monitoring assistant for first responders. IEEE Access 2021;9:25314-25329. [doi: 10.1109/ACCESS.2021.3057578]
56. Liakopoulos L, Stagakis N, Zacharaki EI, Moustakas K. CNN-based stress and emotion recognition in ambulatory settings. Presented at: 2021 12th International Conference on Information, Intelligence, Systems & Applications (IISA); Jul 12-14, 2021; Chania Crete, Greece p. 1-8. [doi: 10.1109/IISA52424.2021.9555508]
57. Li B, Sano A. Extraction and interpretation of deep autoencoder-based temporal features from wearables for forecasting personalized mood, health, and stress. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020 Jun 15;4(2):1-26. [doi: 10.1145/3397318]
58. Hssayeni MD, Ghoraani B. Multi-modal physiological data fusion for affect estimation using deep learning. IEEE Access 2021;9:21642-21652. [doi: 10.1109/ACCESS.2021.3055933]
59. Gil-Martin M, San-Segundo R, Mateos A, Ferreiros-Lopez J. Human stress detection with wearable sensors using convolutional neural networks. IEEE Aerosp Electron Syst Mag 2022 Jan 1;37(1):60-70. [doi: 10.1109/MAES.2021.3115198]
60. Han M, Ozdenizci O, Wang Y, Koike-Akino T, Erdogmus D. Disentangled adversarial transfer learning for physiological biosignals. Presented at: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) in conjunction with the 43rd Annual Conference of the Canadian Medical and Biological Engineering Society; Jul 20-24, 2020; Montreal, Canada p. 422-425 URL: https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=9167168 [accessed 2026-03-19] [doi: 10.1109/EMBC44109.2020.9175233]
61. Mishra V, Pope G, Lord S, et al. Continuous detection of physiological stress with commodity hardware. ACM Trans Comput Healthcare 2020 Apr 30;1(2):1-30. [doi: 10.1145/3361562]
62. Momeni N, Valdes AA, Rodrigues J, Sandi C, Atienza D. CAFS: cost-aware features selection method for multimodal stress monitoring on wearable devices. IEEE Trans Biomed Eng 2022 Mar;69(3):1072-1084. [doi: 10.1109/TBME.2021.3113593] [Medline: 34543185]
63. Rashid N, Chen L, Dautta M, Jimenez A, Tseng P, Al Faruque MA. Feature augmented hybrid CNN for stress recognition using wrist-based photoplethysmography sensor. 2021 Aug 2 Presented at: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); Nov 1-5, 2021. [doi: 10.1109/EMBC46164.2021.9630576]

XSL•FO
RenderX

64. Yannam PKR, Venkatesh V, Gupta M. Research study and system design for evaluating student stress in indian academic setting. Presented at: 2022 14th International Conference on COMmunication Systems & NETworkS (COMSNETS); Jan 4-8, 2022; Bangalore, India p. 54-59. [doi: 10.1109/COMSNETS53615.2022.9668379]
65. Pakhomov SVS, Thuras PD, Finzel R, Eppel J, Kotlyar M. Using consumer-wearable technology for remote assessment of physiological response to stress in the naturalistic environment. In: Cabiati M, editor. PLoS ONE 2020;15(3):e0229942. [doi: 10.1371/journal.pone.0229942] [Medline: 32210441]
66. Holder R, Sah RK, Cleveland M, Ghasemzadeh H. Comparing the predictability of sensor modalities to detect stress from wearable sensor data. Presented at: 2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC); Jan 8-11, 2022; Las Vegas, NV p. 557-562. [doi: 10.1109/CCNC49033.2022.9700682]
67. Elzeiny S, Qaraqe M. Automatic and intelligent stressor identification based on photoplethysmography analysis. IEEE Access 2021;9:68498-68510. [doi: 10.1109/ACCESS.2021.3077358]
68. Heo S, Kwon S, Lee J. Stress detection with single PPG sensor by orchestrating multiple denoising and peak-detecting methods. IEEE Access 2021;9:47777-47785. [doi: 10.1109/ACCESS.2021.3060441]
69. Kar SP, Kumar Rout N, Joshi J. Assessment of mental stress from limited features based on GRU-RNN. Presented at: 2021 IEEE 2nd International Conference on Applied Electromagnetics, Signal Processing, & Communication (AESPC); Nov 26-28, 2021; Bhubaneswar, India p. 1-4. [doi: 10.1109/AESPC52704.2021.9708506]
70. Prashant Bhanushali S, Sadasivuni S, Banerjee I, Sanyal A. Digital machine learning circuit for real-time stress detection from wearable ECG sensor. Presented at: 2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS); Aug 19-20, 2020; Springfield, MA p. 978-981. [doi: 10.1109/MWSCAS48704.2020.9184466]
71. Samyoun S, Sayeed Mondol A, Stankovic JA. Stress detection via sensor translation. Presented at: 2020 16th International Conference on Distributed Computing in Sensor Systems (DCOSS); May 25-27, 2020; Marina del Rey, CA p. 19-26 URL: https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=9178819 [accessed 2026-03-19] [doi: 10.1109/DCOSS49796.2020.00017]
72. Silva E, Aguiar J, Reis LP, Sá JOE, Gonçalves J, Carvalho V. Stress among Portuguese medical students: the EuStress solution. J Med Syst 2020 Jan 2;44(2):45. [doi: 10.1007/s10916-019-1520-1] [Medline: 31897774]
73. Islam TZ, Wu Liang P, Sweeney F, et al. College life is hard! - shedding light on stress prediction for autistic college students using data-driven analysis. Presented at: 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC); Jul 12-16, 2021; Madrid, Spain p. 428-437. [doi: 10.1109/COMPSAC51774.2021.00066]
74. Wu Y, Daoudi M, Amad A, Sparrow L, D'Hondt F. Unsupervised learning method for exploring students' mental stress in medical simulation training. 2020 Oct 25 Presented at: ICMI '20 Companion: Companion Publication of the 2020 International Conference on Multimodal Interaction; Oct 25, 2020; Virtual Event, The Netherlands p. 165-170 URL: https://dl.acm.org/doi/proceedings/10.1145/3395035 [accessed 2026-03-19] [doi: 10.1145/3395035.3425191]
75. Mitro N, Argyri K, Pavlopoulos L, et al. AI-enabled smart wristband providing real-time vital signs and stress monitoring. Sensors (Basel) 2023 Mar 4;23(5):2821. [doi: 10.3390/s23052821] [Medline: 36905025]
76. Tutunji R, Kogias N, Kapteijns B, et al. Detecting prolonged stress in real life using wearable biosensors and ecological momentary assessments: naturalistic experimental study. J Med Internet Res 2023 Oct 19;25:e39995. [doi: 10.2196/39995] [Medline: 37856180]
77. Lange L, Wenzlitschke N, Rahm E. Generating synthetic health sensor data for privacy-preserving wearable stress detection. Sensors (Basel) 2024 May 11;24(10):3052. [doi: 10.3390/s24103052] [Medline: 38793906]
78. Abdul Kader L, Al-Shargie F, Tariq U, Al-Nashash H. One-channel wearable mental stress state monitoring system. Sensors (Basel) 2024 Aug 20;24(16):5373. [doi: 10.3390/s24165373] [Medline: 39205067]
79. Almadhor A, Sampedro GA, Abisado M, et al. Wrist-based electrodermal activity monitoring for stress detection using federated learning. Sensors (Basel) 2023 Apr 14;23(8):3984. [doi: 10.3390/s23083984] [Medline: 37112323]
80. Mai ND, Chung WY. On-chip mental stress detection: integrating a wearable behind-the-ear EEG device with embedded tiny neural network. IEEE J Biomed Health Inform 2025 Mar;29(3):1872-1885. [doi: 10.1109/JBHI.2024.3519600] [Medline: 40030726]
81. Sepanloo K, Shevelev D, Son YJ, Aras S, Hinton JE. Assessing physiological stress responses in student nurses using mixed reality training. Sensors (Basel) 2025 May 20;25(10):3222. [doi: 10.3390/s25103222] [Medline: 40432013]
82. Darwish BA, Rehman SU, Sadek I, Salem NM, Kareem G, Mahmoud LN. From lab to real-life: a three-stage validation of wearable technology for stress monitoring. MethodsX 2025 Jun;14:103205. [doi: 10.1016/j.mex.2025.103205] [Medline: 39996105]
83. Lim KYT, Nguyen Thien MT, Nguyen Duc MA, Posada-Quintero HF. Application of DIY electrodermal activity wristband in detecting stress and affective responses of students. Bioengineering (Basel) 2024 Mar 20;11(3):291. [doi: 10.3390/bioengineering11030291] [Medline: 38534565]
84. Nazeer M, Salagrama S, Kumar P, et al. Improved method for stress detection using bio-sensor technology and machine learning algorithms. MethodsX 2024 Jun;12:102581. [doi: 10.1016/j.mex.2024.102581] [Medline: 38322136]
85. Almadhor A, Sampedro GA, Abisado M, Abbas S. Efficient feature-selection-based stacking model for stress detection based on chest electrodermal activity. Sensors (Basel) 2023 Jul 25;23(15):6664. [doi: 10.3390/s23156664] [Medline: 37571448]

XSL•FO
RenderX

86. Stržinar Ž, Sanchis A, Ledezma A, Sipele O, Pregelj B, Škrjanc I. Stress detection using frequency spectrum analysis of wrist-measured electrodermal activity. Sensors (Basel) 2023 Jan 14;23(2):963. [doi: 10.3390/s23020963] [Medline: 36679760]
87. Feng M, Fang T, He C, Li M, Liu J. Affect and stress detection based on feature fusion of LSTM and 1DCNN. Comput Methods Biomech Biomed Engin 2024;27(4):512-520. [doi: 10.1080/10255842.2023.2188988] [Medline: 36919485]
88. Xuanzhi L, Hakeem A, Mohaisen L, et al. BrainNet: an automated approach for brain stress prediction utilizing electrodermal activity signal with XLNet model. Front Comput Neurosci 2024;18:1482994. [doi: 10.3389/fncom.2024.1482994] [Medline: 39512386]
89. Vidal Bustamante CM, Coombs Iii G, Rahimi-Eichi H, et al. Precision assessment of real-world associations between stress and sleep duration using actigraphy data collected continuously for an academic year: individual-level modeling study. JMIR Form Res 2024 Apr 30;8:e53441. [doi: 10.2196/53441] [Medline: 38687600]
90. Fauzi MA, Yang B, Yeng P. Improving stress detection using weighted score-level fusion of multiple sensor. 2022 Nov 22 Presented at: SIET '22: Proceedings of the 7th International Conference on Sustainable Information Engineering and Technology; Jan 13, 2023; Malang, Indonesia p. 65-71 URL: https://dl.acm.org/doi/proceedings/10.1145/3568231 [accessed 2026-03-19] [doi: 10.1145/3568231.3568242]
91. Tazarv A, Labbaf S, Rahmani A, Dutt N, Levorato M. Active reinforcement learning for personalized stress monitoring in everyday settings. 2023 Jun 21 Presented at: CHASE '23: Proceedings of the 8th ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies; Jan 22, 2024; Orlando, FL p. 44-55 URL: https://dl.acm.org/doi/proceedings/10.1145/3580252 [accessed 2026-03-19] [doi: 10.1145/3580252.3586979]
92. Alfredo RD, Nie L, Kennedy P, et al. "That student should be a lion tamer!" stressviz: designing a stress analytics dashboard for teachers. 2023 Mar 13 Presented at: LAK2023: LAK23: 13th International Learning Analytics and Knowledge Conference; Mar 13, 2023; Arlington, TX p. 57-67 URL: https://dl.acm.org/doi/proceedings/10.1145/3576050 [accessed 2026-03-19] [doi: 10.1145/3576050.3576058]
93. Su Y, Ge L, Wei G. Random forest model predicts stress level in a sample of 18,403 college students. 2024 Jun 21 Presented at: CAIBDA '24: Proceedings of the 2024 4th International Conference on Artificial Intelligence, Big Data and Algorithms; Oct 24, 2024; Zhengzhou, China p. 588-593 URL: https://dl.acm.org/doi/proceedings/10.1145/3690407 [accessed 2026-03-19] [doi: 10.1145/3690407.3690507]
94. Wang L, Hao J, Zhou TH, Song F. ECG stress detection model based on heart rate variability feature extraction. 2023 Jun 17 Presented at: HP3C '23: Proceedings of the 2023 7th International Conference on High Performance Compilation, Computing and Communications; Nov 16, 2023; Jinan, China p. 184-188 URL: https://dl.acm.org/doi/proceedings/10.1145/3606043 [accessed 2026-03-19] [doi: 10.1145/3606043.3606069]
95. Can YS, André E. Performance exploration of RNN variants for recognizing daily life stress levels by using multimodal physiological signals. 2023 Oct 9 Presented at: ICMI '23: Proceedings of the 25th International Conference on Multimodal Interaction; Oct 9, 2023; Paris, France p. 481-487 URL: https://dl.acm.org/doi/proceedings/10.1145/3577190 [accessed 2026-03-19] [doi: 10.1145/3577190.3614159]
96. Prajod P, Mahesh B, André E. Stressor type matters! --- exploring factors influencing cross-dataset generalizability of physiological stress detection. 2024 Nov 4 Presented at: ICMI '24: Proceedings of the 26th International Conference on Multimodal Interaction; Nov 4, 2024; San Jose, Costa Rica p. 508-517 URL: https://dl.acm.org/doi/proceedings/10.1145/3678957 [accessed 2026-03-19] [doi: 10.1145/3678957.3685738]
97. Ganesan P, Thota YR, Shehata H, Nikoubin T. TinyML based stress detection utilizing PPG signals: a lightweight approach for smart wearable devices. 2025 Jun 30 Presented at: Proceedings of the Great Lakes Symposium on VLSI 2025; Jun 30, 2025; New Orleans, LA p. 941-946 URL: https://dl.acm.org/doi/proceedings/10.1145/3716368 [accessed 2026-03-19] [doi: 10.1145/3716368.3735274]
98. Sun X, Zhao L, Gao R, Wang X. Stress recognition based on the markov transition field of electrodermal activity. 2025 Jan 10 Presented at: BIC '25: Proceedings of the 2025 5th International Conference on Bioinformatics and Intelligent Computing; Jan 10, 2025; Shenyang, China p. 467-472 URL: https://dl.acm.org/doi/proceedings/10.1145/3724979 [accessed 2026-03-19] [doi: 10.1145/3724979.3725051]
99. Neigel P, Vargo A, Tag B, Kise K. Using wearables to unobtrusively identify periods of stress in a real university environment. 2024 Oct 5 Presented at: ISWC '24: Proceedings of the 2024 ACM International Symposium on Wearable Computers; Oct 5, 2024; Melbourne, Australia p. 17-24 URL: https://dl.acm.org/doi/proceedings/10.1145/3675095 [accessed 2026-03-19] [doi: 10.1145/3675095.3676608]
100. Pogliaghi A, Di Lascio E, Gashi S, Piciucco E, Santini S, Gjoreski M. Multi-task learning for stress recognition. 2022 Sep 11 Presented at: Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022; Sep 11, 2022 URL: https://dl.acm.org/doi/proceedings/10.1145/3544793 [accessed 2026-03-19] [doi: 10.1145/3544793.3563404]
101. Jaiswal D, Chatterjee D, B s M, Ramakrishnan RK, Pal A. GSR based generic stress prediction system. Presented at: Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing; Oct 8, 2023. [doi: 10.1145/3594739.3610734]

XSL•FO
RenderX

102. Rashid N, Mortlock T, Faruque MAA. Stress detection using context-aware sensor fusion from wearable devices. IEEE Internet Things J 2023 Aug 15;10(16):14114-14127. [doi: 10.1109/JIOT.2023.3265768]
103. Narwat N, Kumar H, Jadon JS, Singh A. Multi-sensory stress detection system. Presented at: 2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence); Jan 18-19, 2024; Noida, India p. 685-689. [doi: 10.1109/Confluence60223.2024.10463214]
104. Kafková J, Pirník R, Janota A, Kuchár P. Stress classification utilising AI studio. Presented at: 2025 26th International Carpathian Control Conference (ICCC); May 19-21, 2025; Starý Smokovec, High Tatras, Slovakia p. 1-5. [doi: 10.1109/ICCC65605.2025.11022862]
105. Lopez R, Shrestha A, Hickey K, et al. Screening students for stress using fitbit data. Presented at: 2024 IEEE International Conference on Big Data (BigData); Dec 15-18, 2024; Washington, DC p. 3931-3934. [doi: 10.1109/BigData62323.2024.10825089]
106. Wilfred JJ, B P, Nirosha R. Real-time stress detection and management using iot sensors and virtual reality technology. Presented at: 2025 8th International Conference on Trends in Electronics and Informatics (ICOEI); Apr 24-25, 2025. [doi: 10.1109/ICOEI65986.2025.11013460]
107. Jaiswal D, Mukhopadhyay S, Sharma V. TinyStressNet: on-device stress assessment with wearable sensors on edge devices. Presented at: 2024 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops); Mar 11-15, 2024; Biarritz, France p. 166-171. [doi: 10.1109/PerComWorkshops59983.2024.10502631]
108. Gaitán-Padilla M, Múnera M, José Pontes M, Eduardo Vieira Segatto M, Cifuentes CA, Diaz CAR. Development of a polymeric optical fiber sensor for stress estimation: a comparative analysis between physiological sensors. IEEE Sensors J 2024 Oct 15;24(20):32140-32149. [doi: 10.1109/JSEN.2024.3435399]
109. Gupta R, Bhongade A, Gandhi TK. Multimodal wearable sensors-based stress and affective states prediction model. Presented at: 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS); Mar 17-18, 2023; Coimbatore, India p. 30-35. [doi: 10.1109/ICACCS57279.2023.10112973]
110. Beierle F, Pryss R. Automating the development of stress detection systems. Presented at: 2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE); Jul 24-27, 2023; Las Vegas, NV p. 2694-2696. [doi: 10.1109/CSCE60160.2023.00432]
111. Masrur N, Halder N, Rashid S, Setu JH, Islam A, Ahmed T. Performance analysis of ensemble and DNN models for decoding mental stress utilizing ECG-based wearable data fusion. Presented at: 2024 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom); Jun 24-27, 2024; Tbilisi, Georgia p. 276-279. [doi: 10.1109/BlackSeaCom61746.2024.10646297]
112. Sakanti MM, Siniaev V, Amaris A, Luo WJ, Kuncoro CBD. Psychological stress classification using extreme gradient boosting algorithm. Presented at: 2024 15th International Conference on Information and Communication Technology Convergence (ICTC); Oct 16-18, 2024; Jeju Island, Republic of Korea p. 946-950. [doi: 10.1109/ICTC62082.2024.10827020]
113. Shedage PS, Pouriyeh S, Parizi RM, Han M, Sannino G, Dehbozorgi N. Stress detection using multimodal physiological signals with machine learning from wearable devices. Presented at: 2024 IEEE Symposium on Computers and Communications (ISCC); Jun 26-29, 2024; Paris, France p. 1-6. [doi: 10.1109/ISCC61673.2024.10733703]
114. Gaitán-Padilla M, Múnera M, Cifuentes CA, Monteiro ME, Pontes MJ, Diaz CAR. Stress classification using a low-cost optical fiber physiological sensor: a preliminary study. Presented at: 2023 SBMO/IEEE MTT-S International Microwave and Optoelectronics Conference (IMOC); Nov 5-9, 2023. [doi: 10.1109/IMOC57131.2023.10379658]
115. Tanwar R, Singh G, Pal PK. FuSeR: fusion of wearables data for stress recognition using explainable artificial intelligence models. Presented at: 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT); Jul 6-8, 2023; Delhi, India p. 1-6. [doi: 10.1109/ICCCNT56998.2023.10307589]
116. Gullapalli BT, Nathan V, Rahman MM, Kuang J, Gao JA. A framework for extracting heart rate variability features from earbud-PPG for stress detection. Annu Int Conf IEEE Eng Med Biol Soc 2024 Jul;2024:1-5. [doi: 10.1109/EMBC53108.2024.10782088] [Medline: 40039377]
117. Sadruddin S, Khairnar VD, Vora DR. Machine learning based assessment of mental stress using wearable sensors. Presented at: 2024 11th International Conference on Computing for Sustainable Global Development (INDIACom); Feb 28 to Mar 1, 2024; New Delhi, India p. 351-355. [doi: 10.23919/INDIACom61295.2024.10498597]
118. Jahanjoo A, TaheriNejad N, Aminifar A. High-accuracy stress detection using wrist-worn PPG sensors. Presented at: 2024 IEEE International Symposium on Circuits and Systems (ISCAS); Jul 2, 2024; Singapore, Singapore p. 1-5. [doi: 10.1109/ISCAS58744.2024.10558012]
119. Parousidou V, Yfantidou S, Karagianni C, Vakali A. Stress beats: a continuum of learning methods for personalized stress detection. Presented at: 2023 IEEE International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT); Oct 26-29, 2023; Venice, Italy p. 40-47. [doi: 10.1109/WI-IAT59888.2023.00012]
120. Karpagam GR, Vardhan V M H, K K K, P P, Ramesh P, Sathyendira B S. Physiological data-based stress detection: from wrist sensors to cloud computing and user feedback integration. Presented at: 2024 International Conference on Smart Systems for Electrical, Electronics, Communication and Computer Engineering (ICSSEECC); Jun 28-29, 2024; Coimbatore, India p. 386-391. [doi: 10.1109/ICSSEECC61126.2024.10649521]

XSL•FO
RenderX

121. Shikha S, Sethia D, Indu S. Optimization of wearable biosensor data for stress classification using machine learning and explainable AI. IEEE Access 2024;12:169310-169327 [FREE Full text] [doi: 10.1109/ACCESS.2024.3463742]
122. Hasanpoor Y, Tarvirdizadeh B, Alipour K, Ghamari M. Wavelet-based analysis of photoplethysmogram for stress detection using convolutional neural networks. Presented at: 2023 11th RSI International Conference on Robotics and Mechatronics (ICRoM); Dec 19-21, 2023; Tehran, Islamic Republic of Iran p. 501-506. [doi: 10.1109/ICRoM60803.2023.10412512]
123. Benita DS, Ebenezer AS, Susmitha L, Subathra MSP, Priya SJ. Stress detection using cnn on the wesad dataset. Presented at: 2024 International Conference on Emerging Systems and Intelligent Computing (ESIC); Feb 9-10, 2024; Bhubaneswar, India p. 308-313. [doi: 10.1109/ESIC60604.2024.10481604]
124. Hsu A. Quantifying exam stress progressions using electrodermal activity and machine learning. Presented at: 2023 IEEE 23rd International Conference on Bioinformatics and Bioengineering (BIBE); Dec 4-6, 2023; Dayton, OH p. 434-438. [doi: 10.1109/BIBE60311.2023.00077]
125. Carmisciano L, Boschi T, Chiaromonte F, Delmastro F, Vandin A. Investigating functional data analysis for wearable physiological sensor data in stress evaluation. Presented at: 2024 IEEE Symposium on Computers and Communications (ISCC); Jun 26-29, 2024; Paris, France p. 1-6. [doi: 10.1109/ISCC61673.2024.10733576]
126. Warrier LC, Ragesh GK, Ram Samarth BB, Gurumurthy K. Privacy-preserved stress detection from wearables using federated learning. Presented at: 2024 IEEE 5th India Council International Subsections Conference (INDISCON); Aug 22-24, 2024; Chandigarh, India p. 1-6. [doi: 10.1109/INDISCON62179.2024.10744249]
127. Calbert L, Tonekaboni NH. Temporal dynamics of classroom stress: insights from wearable sensors and machine learning. Presented at: 2024 International Conference on Machine Learning and Applications (ICMLA); Dec 18-20, 2024; Miami, FL p. 377-384. [doi: 10.1109/ICMLA61862.2024.00057]
128. Kumar S, Raj Chauhan A, Kumar A, Yang G. Resp-BoostNet: mental stress detection from biomarkers measurable by smartwatches using boosting neural network technique. IEEE Access 2024;12:149861-149874. [doi: 10.1109/ACCESS.2024.3461588]
129. Hasanpoor Y, Rostami A, Tarvirdizadeh B, Alipour K, Ghamari M. Real-time stress detection via photoplethysmogram signals: implementation of a combined continuous wavelet transform and convolutional neural network on resource-constrained microcontrollers. Presented at: 2024 32nd International Conference on Electrical Engineering (ICEE); May 14-16, 2024. [doi: 10.1109/ICEE63041.2024.10668302]
130. Le Tran Thuan T, Nguyen PK, Gia QN, Tran AT, Le QK. Machine learning algorithms for stress level analysis based on skin surface temperature and skin conductance. Presented at: 2024 IEEE 6th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS); Jun 14-16, 2024. [doi: 10.1109/ECBIOS61468.2024.10885479]
131. Fernandez J, Martínez R, Innocenti B, López B. Contribution of EEG signals for students' stress detection. IEEE Trans Affective Comput 2025;16(2):1235-1246. [doi: 10.1109/TAFFC.2024.3503995]
132. Tanwar R, Pal PK, Singh G. Wearables based personalised stress recognition using signal processing and hybrid deep learning model. Presented at: 2024 International Conference on Computer, Electronics, Electrical Engineering & their Applications (IC2E3); Jun 6-7, 2024; Srinagar Garhwal, India p. 1-6. [doi: 10.1109/IC2E362166.2024.10827149]
133. Huang M, Yang H, Sun N, et al. Study of a hybrid CNN-SVM model for stress detection with automated heart rate variability feature extraction method. Presented at: 2024 3rd International Conference on Health Big Data and Intelligent Healthcare (ICHIH); Dec 13-15, 2024; Zhuhai, China p. 316-319. [doi: 10.1109/ICHIH63459.2024.11064910]
134. Oh K, Choi JK, Park H, Lee S. Personalized ensemble based stress detection using wearable sensor data. Presented at: 2025 27th International Conference on Advanced Communications Technology (ICACT); Feb 16-19, 2025; Pyeong Chang, Korea, Republic of p. 470-475. [doi: 10.23919/ICACT63878.2025.10936652]
135. Thapa B, Rivas M, Griffith H, Rathore H. StressLLM: large language models for stress prediction via wearable sensor data. Presented at: 2025 IEEE International Conference on Consumer Electronics (ICCE); Jan 11-14, 2025; Las Vegas, NV p. 1-6. [doi: 10.1109/ICCE63647.2025.10929774]
136. Abdelfattah E, Joshi S, Tiwari S. Machine and deep learning models for stress detection using multimodal physiological data. IEEE Access 2025;13:4597-4608. [doi: 10.1109/ACCESS.2024.3525459]
137. Tsiampa K, Zhu L, Spachos P, Plagianakos VP. Investigating feasibility of stress detection from social media content through wearables. Presented at: GLOBECOM 2023 - 2023 IEEE Global Communications Conference; Dec 4-8, 2023; Kuala Lumpur, Malaysia p. 1173-1178. [doi: 10.1109/GLOBECOM54140.2023.10437938]
138. Fazeli S, Levine L, Beikzadeh M, et al. A self-supervised framework for improved data-driven monitoring of stress via multi-modal passive sensing. Presented at: 2023 IEEE International Conference on Digital Health (ICDH); Jul 2-8, 2023; Chicago, IL p. 177-183. [doi: 10.1109/ICDH60066.2023.00033]
139. Subathra P, Malarvizhi S. Autoencoder-based human stress detection system using biological signals. Presented at: 2024 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI); Apr 17-18, 2024; Chennai, India p. 1-7. [doi: 10.1109/RAEEUCCI61380.2024.10547833]
140. Shikha S, Sethia D, Indu S. CorLMI-fsa: an efficient feature selection approach for stress classification using physiological signals. Presented at: 2025 Fifth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT); Jan 9-10, 2025; Bhilai, India p. 1-7. [doi: 10.1109/ICAECT63952.2025.10958862]

XSL•FO
RenderX

141. Andreas A, Mavromoustakis CX, Song H, Batalla JM. Optimisation of CNN through transferable online knowledge for stress and sentiment classification. IEEE Trans Consumer Electron 2024;70(1):3088-3097. [doi: 10.1109/TCE.2023.3319111]
142. Kasnesis P, Chatzigeorgiou C, Feidakis M, Gutiérrez Á, Patrikakis CZ. TranSenseFusers: a temporal CNN-transformer neural network family for explainable PPG-based stress detection. Biomed Signal Process Control 2025 Apr;102:107248. [doi: 10.1016/j.bspc.2024.107248]
143. Ciharova M, Amarti K, van Breda W, et al. Machine-learning detection of stress severity expressed on a continuous scale using acoustic, verbal, visual, and physiological data: lessons learned. Front Psychiatry 2025;16:1548287. [doi: 10.3389/fpsyt.2025.1548287] [Medline: 40585547]
144. Darwish BA, Salem NM, Kareem G, Mahmoud LN, Sadek I. Evaluating the potential of wearable technology in early stress detection: a multimodal approach. medRxiv. Preprint posted online on Jul 21, 2024. [doi: 10.1101/2024.07.19.24310732]
145. Nuamah J. Effect of recurrent task-induced acute stress on task performance, vagally mediated heart rate variability, and task-evoked pupil response. Int J Psychophysiol 2024 Apr;198:112325. [doi: 10.1016/j.ijpsycho.2024.112325] [Medline: 38447701]
146. Sa-nguannarm P, Elbasani E, Kim JD. Human activity recognition for analyzing stress behavior based on Bi-LSTM. THC 2023 Sep 15;31(5):1997-2007. [doi: 10.3233/THC-235002]
147. Nelson BW, Harvie HMK, Jain B, Knight EL, Roos LE, Giuliano RJ. Smartphone photoplethysmography pulse rate covaries with stress and anxiety during a digital acute social stressor. Psychosom Med 2023 Sep 1;85(7):577-584. [doi: 10.1097/PSY.0000000000001178] [Medline: 37409791]
148. Dahal K, Bogue-Jimenez B, Doblas A. Global stress detection framework combining a reduced set of HRV features and random forest model. Sensors (Basel) 2023 May 31;23(11):5220. [doi: 10.3390/s23115220] [Medline: 37299947]
149. Aqajari SAH, Labbaf S, Tran PH, et al. Context-aware stress monitoring using wearable and mobile technologies in everyday settings. arXiv. Preprint posted online on Dec 14, 2023. [doi: 10.1101/2023.04.20.23288181]
150. Jiao Y, Wang X, Liu C, et al. Feasibility study for detection of mental stress and depression using pulse rate variability metrics via various durations. Biomed Signal Process Control 2023 Jan;79:104145. [doi: 10.1016/j.bspc.2022.104145]
151. Lotfi F, Lotfi A, Lotfi M, Bjelica A, Bogdanović Z. Enhancing smart healthcare with female students' stress and anxiety detection using machine learning. Psychol Health Med 2025 Aug 9;30(7):1465-1484. [doi: 10.1080/13548506.2025.2484698]
152. Patanè G, Sorrenti A, Bellitto G, Palazzo S. Continual learning strategies for personalized mental well-being monitoring from mobile sensing data. 2025 Oct 27 Presented at: PILM '25: Proceedings of the International Workshop on Personalized Incremental Learning in Medicine; Oct 27, 2025; Dublin, Ireland p. 9-17. [doi: 10.1145/3746259.3760432]
153. Subathra P, Malarvizhi S, Ferents Koni Jiavana K, Patil S. A wearable electronic band for stress understanding using machine learning. IEEE Sensors J 2025 Oct 15;25(20):38639-38648. [doi: 10.1109/JSEN.2025.3590380]
154. van der Mee DJ, Koyuncu Z, Lemmers-Jansen ILJ. Are you stressed or just excited? What the Garmin Stress Score can say about your mood. Journal of Affective Disorders Reports 2025 Jul;21:100974. [doi: 10.1016/j.jadr.2025.100974]
155. De Angel V, Lewis S, White K, et al. Digital health tools for the passive monitoring of depression: a systematic review of methods. NPJ Digit Med 2022 Jan 11;5(1):3. [doi: 10.1038/s41746-021-00548-8] [Medline: 35017634]
156. Downes MJ, Brennan ML, Williams HC, Dean RS. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). BMJ Open 2016 Dec 8;6(12):e011458. [doi: 10.1136/bmjopen-2016-011458] [Medline: 27932337]
157. Wells G, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta- analysis. URL: https://www.researchgate.net/publication/261773681_The_Newcastle-Ottawa_Scale_NOS_for_Assessing_the_Quality_of_Non-Randomized_Studies_in_Meta-Analysis [accessed 2026-02-13]
158. Gagliardi AR, Berta W, Kothari A, Boyko J, Urquhart R. Integrated knowledge translation (IKT) in health care: a scoping review. Implementation Sci 2015 Dec;11(1):38. [doi: 10.1186/s13012-016-0399-1]
159. Shaheen F, Verma B, Asafuddoula M. Impact of automatic feature extraction in deep learning architecture. Presented at: 2016 International Conference on Digital Image Computing; Nov 30 to Dec 26, 2016; Gold Coast, Australia p. 1-8. [doi: 10.1109/DICTA.2016.7797053]
160. Allen AP, Kennedy PJ, Dockray S, Cryan JF, Dinan TG, Clarke G. The Trier Social Stress Test: principles and practice. Neurobiol Stress 2017 Feb;6:113-126. [doi: 10.1016/j.ynstr.2016.11.001] [Medline: 28229114]
161. WESAD (wearable stress and affect detection). Kaggle. URL: https://ubicomp.eti.uni-siegen.de/home/datasets/icmi18/ [accessed 2023-06-29]
162. Zhang P, Jung G, Alikhanov J, Ahmed U, Lee U. A reproducible stress prediction pipeline with mobile sensor data. Proc ACM Interact Mob Wearable Ubiquitous Technol 2024 Aug 22;8(3):1-35. [doi: 10.1145/3678578] [Medline: 40093941]
163. Patle A, Chouhan DS. SVM kernel functions for classification. Presented at: 2013 International Conference on Advances in Technology and Engineering (ICATE 2013); Jan 23-25, 2013; Mumbai p. 1-9. [doi: 10.1109/ICAdTE.2013.6524743]
164. Li YF, Kwok J, Zhou ZH. Cost-sensitive semi-supervised support vector machine. AAAI 2010 Jul 3;24(1):500-505. [doi: 10.1609/aaai.v24i1.7661]
165. Ayeni JA, Department of Computer Sciences, Ajayi Crowther University, Oyo, Nigeria. Convolutional neural network (CNN): the architecture and applications. Appl J Phys Sci 2022 Dec 30;4(4):42-50. [doi: 10.31248/AJPS2022.085]

XSL•FO
RenderX

166. de Arriba-Pérez F, Santos-Gago JM, Caeiro-Rodríguez M, Ramos-Merino M. Study of stress detection and proposal of stress-related features using commercial-off-the-shelf wrist wearables. J Ambient Intell Human Comput 2019 Dec;10(12):4925-4945. [doi: 10.1007/s12652-019-01188-3]
167. Setz C, Arnrich B, Schumm J, La Marca R, Troster G, Ehlert U. Discriminating stress from cognitive load using a wearable EDA device. IEEE Trans Inform Technol Biomed 2009;14(2):410-417. [doi: 10.1109/TITB.2009.2036164]
168. Xu X, Liu X, Zhang H, et al. GLOBEM: cross-dataset generalization of longitudinal human behavior modeling. Proc ACM Interact Mob Wearable Ubiquitous Technol 2022 Jan 11;6(4):1-34. [doi: 10.1145/3569485]
169. Sarker H, Tyburski M, Rahman MM, et al. Finding significant stress episodes in a discontinuous time series of rapidly varying mobile sensor data. 2016 May 7 Presented at: CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems; May 7, 2016; San Jose, CA p. 4489-4501 URL: https://dl.acm.org/doi/proceedings/10.1145/2858036 [accessed 2026-03-19] [doi: 10.1145/2858036.2858218]
170. Perini R, Veicsteinas A. Heart rate variability and autonomic activity at rest and during exercise in various physiological conditions. Eur J Appl Physiol 2003 Oct;90(3-4):317-325. [doi: 10.1007/s00421-003-0953-9] [Medline: 13680241]
171. Mishra V, Hao T, Sun S, et al. Investigating the role of context in perceived stress detection in the wild. 2018 Oct 8 Presented at: UbiComp '18: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers; Oct 8, 2018; Singapore p. 1708-1716 URL: https://dl.acm.org/doi/proceedings/10.1145/3267305 [accessed 2026-03-19] [doi: 10.1145/3267305.3267537]
172. Möller A, Kranz M, Schmid B, Roalter L, Diewald S. Investigating self-reporting behavior in long-term studies. 2013 Apr 27 Presented at: CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; Apr 27, 2013; Paris, France p. 2931-2940 URL: https://dl.acm.org/doi/proceedings/10.1145/2470654 [accessed 2026-03-19] [doi: 10.1145/2470654.2481406]
173. Fass-Holmes B. Survey fatigue--what is its role in undergraduates' survey participation and response rates? J Interdiscip Stud Educ 2022 [FREE Full text]
174. Wen CKF, Schneider S, Stone AA, Spruijt-Metz D. Compliance with mobile ecological momentary assessment protocols in children and adolescents: a systematic review and meta-analysis. J Med Internet Res 2017 Apr 26;19(4):e132. [doi: 10.2196/jmir.6641] [Medline: 28446418]
175. Riley RD, Ensor J, Snell KIE, et al. Importance of sample size on the quality and utility of AI-based prediction models for healthcare. Lancet Digit Health 2025 Jun;7(6):100857. [doi: 10.1016/j.landig.2025.01.013] [Medline: 40461350]
176. Kaplan RM, Chambers DA, Glasgow RE. Big data and large sample size: a cautionary note on the potential for bias. Clinical Translational Sci 2014 Aug;7(4):342-346 [FREE Full text] [doi: 10.1111/cts.12178]
177. Schmidt P, Reiss A, Duerichen R, Marberger C, Van Laerhoven K. Introducing WESAD, a multimodal dataset for wearable stress and affect detection. 2018 Oct 2 Presented at: ICMI '18: Proceedings of the 20th ACM International Conference on Multimodal Interaction; Oct 2, 2018; Boulder, CO p. 400-408 URL: https://dl.acm.org/doi/proceedings/10.1145/3242969 [accessed 2026-03-19] [doi: 10.1145/3242969.3242985]
178. Xu X, Chikersal P, Doryab A, et al. Leveraging routine behavior and contextually-filtered features for depression detection among college students. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Sep 9;3(3):1-33. [doi: 10.1145/3351274]
179. Xu X, Chikersal P, Dutcher JM, et al. Leveraging collaborative-filtering for personalized behavior modeling: a case study of depression detection among college students. Proc ACM Interact Mob Wearable Ubiquitous Technol 2021 Mar 19;5(1):1-27. [doi: 10.1145/3448107]
180. Salmasi V, Lii TR, Humphreys K, Reddy V, Mackey SC. A literature review of the impact of exclusion criteria on generalizability of clinical trial findings to patients with chronic pain. PR9 2022;7(6):e1050. [doi: 10.1097/PR9.0000000000001050]
181. Humphreys K. A review of the impact of exclusion criteria on the generalizability of schizophrenia treatment research. Clin Schizophr Relat Psychoses 2017;11(1):49-57. [doi: 10.3371/1935-1232-11.1.49] [Medline: 28548580]
182. Wong JJ, Jones N, Timko C, Humphreys K. Exclusion criteria and generalizability in bipolar disorder treatment trials. Contemp Clin Trials Commun 2018 Mar;9:130-134. [doi: 10.1016/j.conctc.2018.01.009]
183. Alegría M, NeMoyer A, Falgàs Bagué I, Wang Y, Alvarez K. Social determinants of mental health: where we are and where we need to go. Curr Psychiatry Rep 2018 Sep 17;20(11):95. [doi: 10.1007/s11920-018-0969-9] [Medline: 30221308]
184. McEwen BS, Gianaros PJ. Central role of the brain in stress and adaptation: links to socioeconomic status, health, and disease. Ann N Y Acad Sci 2010 Feb;1186(1):190-222. [doi: 10.1111/j.1749-6632.2009.05331.x] [Medline: 20201874]
185. Jackson RW, Treiber FA, Turner JR, Davis H, Strong WB. Effects of race, sex, and socioeconomic status upon cardiovascular stress responsivity and recovery in youth. Int J Psychophysiol 1999 Jan;31(2):111-119. [doi: 10.1016/S0167-8760(98)00044-0]
186. Braveman P, Egerter S, Williams DR. The social determinants of health: coming of age. Annu Rev Public Health 2011;32(1):381-398. [doi: 10.1146/annurev-publhealth-031210-101218] [Medline: 21091195]
187. Chien WS, Lee CC. Understanding missing data bias in longitudinal mental stress detection. Presented at: 2024 IEEE 20th International Conference on Body Sensor Networks (BSN); Oct 15-17, 2024; Chicago, IL p. 1-4. [doi: 10.1109/BSN63547.2024.10780500]
188. McCombe N, Liu S, Ding X, et al. Practical strategies for extreme missing data imputation in dementia diagnosis. IEEE J Biomed Health Inform 2021;26(2):818-827. [doi: 10.1109/JBHI.2021.3098511]

XSL•FO
RenderX

189. Ibrahim JG, Molenberghs G. Missing data methods in longitudinal studies: a review. TEST (Madr) 2009 May;18(1):1-43. [doi: 10.1007/s11749-009-0138-x]
190. Abd-Alrazaq A, Alajlani M, Ahmad R, et al. The performance of wearable AI in detecting stress among students: systematic review and meta-analysis. J Med Internet Res 2024 Jan 31;26:e52622. [doi: 10.2196/52622] [Medline: 38294846]
191. Gedam S, Paul S. A review on mental stress detection using wearable sensors and machine learning techniques. IEEE Access 2021;9:84045-84066. [doi: 10.1109/ACCESS.2021.3085502]
192. Giannakakis G, Grigoriadis D, Giannakaki K, Simantiraki O, Roniotis A, Tsiknakis M. Review on psychological stress detection using biosignals. IEEE Trans Affective Comput 2022 Jan 1;13(1):440-460. [doi: 10.1109/TAFFC.2019.2927337]
193. Hickey BA, Chalmers T, Newton P, et al. Smart devices and wearable technologies to detect and monitor mental health conditions and stress: a systematic review. Sensors (Basel) 2021 May 16;21(10):3461. [doi: 10.3390/s21103461] [Medline: 34065620]
194. Shanmugasundaram G, Yazhini S, Hemapratha E, Nithya S. A comprehensive review on stress detection techniques. Presented at: 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN); Mar 29-30, 2019; Pondicherry, India p. 1-6. [doi: 10.1109/ICSCAN.2019.8878795]
195. Onnela JP. Opportunities and challenges in the collection and analysis of digital phenotyping data. Neuropsychopharmacology 2021 Jan;46(1):45-54. [doi: 10.1038/s41386-020-0771-3] [Medline: 32679583]
196. Xu X, Zhang H, Sefidgar Y, et al. GLOBEM dataset: multi-year datasets for longitudinal human behavior modeling generalization. arXiv. Preprint posted online on Nov 4, 2023 URL: http://arxiv.org/abs/2211.02733 [accessed 2024-10-03]
197. Gjoreski M, Gjoreski H, Luštrek M, Gams M. Continuous stress detection using a wrist device: in laboratory and real life. 2016 Presented at: UbiComp '16: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct; Sep 12-16, 2016; Heidelberg, Germany p. 1185-1193. [doi: 10.1145/2968219.2968306]

Abbreviations

CNN: convolutional neural network
ECG: electrocardiogram
EDA: electrodermal activity
HR: heart rate
HRV: heart rate variability
LSTM: long short-term memory
PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
RF: random forest
SVM: support vector machine
TSST: Trier Social Stress Test
WESAD: Wearable Stress and Affect Detection
XGBoost: extreme gradient boosting

Edited by S Brini; submitted 09.Jul.2024; peer-reviewed by M Matabuena, RK Nath; revised version received 05.Jan.2026; accepted 06.Jan.2026; published 30.Mar.2026.

Please cite as:
Sathyanarayana A, Amin OB, An J, Onnela JP
Examining the Use of Consumer Wearable Devices and Digital Tools for Stress Measurement in College Students: Scoping Review of Methods
JMIR Mhealth Uhealth 2026;14:e64144
URL: https://mhealth.jmir.org/2026/1/e64144
doi:10.2196/64144

© Aarti Sathyanarayana, Ohida Binte Amin, Jennie An, Jukka Pekka Onnela. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 30.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mental Health Apps Implemented in the Workplace: Scoping Review of Trends and Gaps in Evaluation Research

Sheila Addanki[1], MSc, PhD; Luciana Macedo[2], PT, PhD; Joy MacDermid[3], BSc, BScPT, MSc, PhD; Sandra Moll[2], MSc(OT), PhD

[1]University of British Columbia, Vancouver, BC, Canada

[2]School of Rehabilitation Science, Faculty of Health Sciences, McMaster University, 1400 Main Street West, Hamilton, ON, Canada

[3]Western University, London, ON, Canada

Corresponding Author:
Sandra Moll, MSc(OT), PhD
School of Rehabilitation Science, Faculty of Health Sciences, McMaster University, 1400 Main Street West, Hamilton, ON, Canada

Abstract

Background: Technology-based solutions to support the mental health needs of workers are on the rise, as evidenced by the growing body of research related to e–mental health apps implemented with workers or within the context of workplaces. This expanding landscape of evidence related to mental health apps underscores the necessity of summarizing and consolidating the different ways in which studies are evaluating real-world technology-based interventions in a complex setting such as a workplace.

Objective: The aim of this scoping review is to summarize the growing body of evidence evaluating mental health apps with workers. Specific research questions include "What types of mental health apps are studied?" "With whom are they being evaluated?" and "What outcomes are being tracked with what tools?"

Methods: The scoping review followed 5 stages: specifying the research question, identifying relevant literature, selecting studies, extracting data, and synthesizing the findings. The search strategy was applied across 6 databases (PsycINFO, Embase, MEDLINE, Cochrane Database, OVID Healthstar, and OVID Emcare) to identify relevant studies from January 2000 to August 2023.

Results: From the 288 abstracts screened, 54 studies met the inclusion criteria for this review. Over two-thirds of the studies were randomized controlled trials. The studies included 44 different apps, comprising 23 structured self-guided apps, 15 unstructured self-guided apps, and 6 apps designed as adjuncts to other interventions. Evaluation approaches included examining user engagement and utilization, investigating the impact on users, and evaluating the implementation process. Most studies evaluated the impact on individual mental health–related outcomes as well as workplace-related outcomes.

Conclusions: This scoping review provides a comprehensive overview of the ways in which studies are currently evaluating workplace mental health apps. The review highlights key trends and gaps in the existing research, noting that most studies focus on the effects of mental health apps on individual users, while only a limited number of studies explore how implementing such dynamic interventions within complex real-world settings (eg, workplaces) may influence their overall effectiveness. Future research should explore the implementation process to identify factors that promote and/or hinder the optimal use and impact of mental health apps for workers.

Trial Registration: OSF Registries osf.io/hxndj; https://doi.org/10.17605/OSF.IO/HXNDJ

(JMIR Mhealth Uhealth 2026;14:e57046) doi:10.2196/57046

KEYWORDS

mobile apps; digital health; mobile health; mHealth; mental health apps; workplace; workers; scoping review; user engagement; smartphones; eHealth; technology-based interventions

Introduction

Mental health issues have a significant human and financial cost for society and workplaces. Poor mental health and loss of productivity are estimated to cost the global economy approximately US $1 trillion each year [1], and this figure is projected to rise, given the COVID-19 pandemic's impact on workplace mental health. Rapid transitions and changes in workplace demands linked to the pandemic have led to an increase in reports of perceived stress, depression, and anxiety among working adults, particularly for working mothers, people of minority ethnicities, and those working in frontline positions [2-4]. Limited access to traditional mental health services, along with the longstanding stigma associated with seeking support,

XSL•FO
RenderX

has led to a growing need to expand access to a range of mental health supports to address the unmet needs of the workforce and reduce the burden of mental health issues.

In recent years, there has been a growing interest in digital health technologies, including web-based and smartphone-based interventions as tools to address mental health problems. One of the main advantages of digital mental health interventions is their ability to overcome traditional barriers to accessing mental health care by providing on-demand access to quality mental health information and support [5,6]. For example, many people are reluctant to seek help due to the stigma associated with mental illness. Digital interventions, such as mental health apps, offer a private and anonymous way to access support without the fear of being judged. Digital interventions can also help overcome geographical barriers by providing access for people who live in remote or rural areas, where mental health services may be limited or nonexistent [5]. In addition, digital interventions can be accessed at any time and from anywhere, which can help overcome the time constraints that many working adults face when trying to access traditional mental health services. Mental health apps can be used across the full spectrum of mental health care, from prevention to management. For example, there are apps that help people assess and manage stress and anxiety, apps that provide support for people with depression, and apps that help people improve their well-being.

There is a growing body of research that supports the effectiveness of mental health apps in both the general and clinical populations [4,7]. A review of app-based mental health interventions found them to be effective in reducing symptoms of depression and anxiety [4,7], as well as improving mental health literacy and well-being [8] in both clinical and non-clinical populations [4,7-9]. However, a review of publicly available popular mental health apps found that only 2 out of the 6 mental health apps identified were empirically evaluated [10], underscoring the gap between mental health apps available publicly and those that are evaluated in empirical studies. There is an urgent need for high-quality studies in this field to explore mental health apps that are evidence-based, empirically supported, and scalable to broader users.

There is a growing interest in the use of mental health apps across different settings such as clinical, postsecondary, and community settings, and more recently in workplaces. Several reviews provide evidence to support that digital mental health interventions implemented in workplaces have positive effects on improving general mental health and workplace-related outcomes [11-13]. Studies have explored the effects of these interventions in improving well-being, symptoms of stress, and several workplace-related outcomes, such as work effectiveness or productivity and job satisfaction [11-13]. However, the majority of the studies in these reviews consisted of web-based interventions, with only a small percentage of them exploring mental health apps.

As the availability and use of mental health apps in workplace settings become more prevalent, there is a growing need to identify and synthesize the existing evidence related to their systematic evaluation with workers. In this paper, we aim to conduct a scoping review to synthesize and critically reflect on the body of evidence in this emerging field. Specifically, we will summarize and identify trends and gaps in the evaluation of mental health apps in workplace settings.

Methods

Study Design

We followed the 5 steps of scoping review methodology [14]: specifying the research question, identifying relevant literature, selecting studies, extracting data, and synthesizing the findings. The protocol for this review is registered on the Open Science Framework. We also adhered to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines (Checklist 1) to ensure consistent and transparent reporting [15]. To examine and summarize the breadth of evidence regarding the evaluation of mental health apps for working adults, we developed the following research questions based on the population, intervention, comparison, and outcome (PICO) framework:

1. How are mental health apps being evaluated with workers and/or in the context of a workplace?
2. What types of mental health apps are studied?
3. With whom are the applications being evaluated (workers or workplace sectors)?
4. What key variables are being assessed, and what tools are being used to track progress and impact?

Inclusion Criteria

Studies were included in this review if they met the following inclusion criteria: (1) the primary study intervention was a mobile mental health app that was tested by employees; (2) the study evaluated the effectiveness, implementation, impact, or use of the mental health app; (3) study participants included workers currently employed in any field and/or implemented in the context of a workplace; (4) the study included any quantitative or qualitative evaluation design; (5) the study was published in a peer-reviewed journal; and (6) the study was available in English. Studies were excluded if: (1) the intervention in the study was primarily a web-based or internet-based intervention (ie, not a mobile app) and (2) the study design only focused on app development (not evaluation) of a mental health app for workers. Conference proceedings and gray literature were not included, as the key focus of this review was to present a summary of published peer-reviewed studies that systematically evaluated mental health apps.

Search Strategy

We searched the following electronic databases: PsycINFO, Embase, MEDLINE, Cochrane Database, OVID Healthstar, and OVID Emcare for peer-reviewed studies from January 2000 to August 2023. A comprehensive search strategy, including search terms, subject headings, and syntax requirements for the databases, was adapted from a previous systematic review of digital mental health interventions for the workplace [11] and was modified in consultation with a health science librarian. Search terms included a combination of constructs such as “Mental Health,” “Workplace,” and “Mobile Application” (Multimedia Appendix 1). The search was conducted from May 2023 to August 2023.

XSL•FO
RenderX

Study Screening

Data screening and extraction were carried out by 2 independent reviewers (SA and ML) using Covidence systematic review software (Veritas Health Innovation), an online tool that supports screening and data extraction for reviews, to facilitate a systematic screening process. All studies were screened independently by 2 reviewers, and any discrepancies in study selection were discussed in a review meeting. A third reviewer, SM, was consulted to resolve any disagreements between the 2 primary reviewers and to improve clarity in decision-making when necessary.

Data Extraction and Analysis

Data extraction was an iterative process designed to develop and refine a data charting sheet in Microsoft Excel to ensure all relevant data were captured comprehensively. The fields of data extracted included study characteristics such as the year, country of publication, study design, workplace context or industry where the study was conducted, intervention characteristics, timeframe, and evaluation approaches including both assessment tools and outcomes explored in the study. The initial extraction chart was piloted with 5 studies and was later revised to capture additional information relevant to our research questions. Analysis of these data included a summary of the study characteristics, types of app interventions, study populations, and approaches to evaluation (variables and evaluation tools). Descriptive summaries of the trends in each category will be outlined.

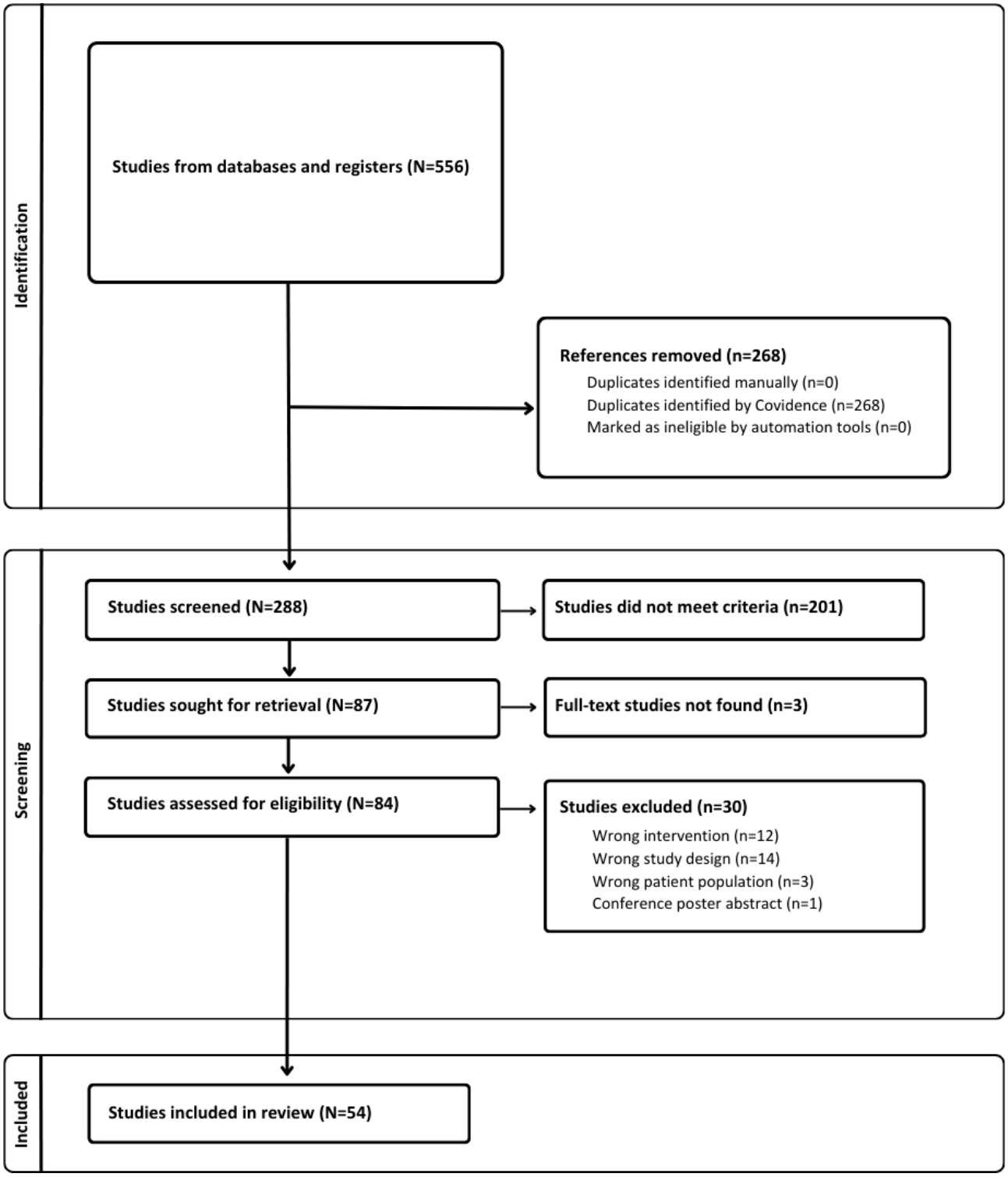
Results

Overview

We searched 6 electronic databases, identifying 288 studies after the removal of duplicates. After screening titles and abstracts, 85 studies were deemed eligible for full-text screening. Subsequently, 54 studies met our inclusion criteria for being included in this scoping review (Figure 1). Data from these studies were extracted and charted to generate a comprehensive summary of the breadth and trends in the body of evidence of mental health app evaluation in the workplace. Of these 54 studies, 11 were protocols for studies in progress; hence, no results were available for these studies at the time of the review. Upon careful consideration of the protocols and clinical trial registrations, we were able to gather relevant information related to the key outcomes being evaluated and the processes by which the studies were being conducted, which was deemed sufficient to answer our main research question: “How are mental health apps being evaluated with workers and/or in the context of the workplace?” As a result, we decided to report these ongoing studies in our review.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the study screening process.



Study Characteristics

Among the 54 studies included, approximately two-thirds were randomized controlled trials (RCTs; protocols or trial registrations n=11 [16-26]; completed RCTs n=21 [27-48]; pilot RCTs n=3 [49-51]). A smaller subset of studies adopted nonrandomized methods such as pre-post nonexperimental designs (n=13) [52-63], with even fewer studies using mixed methods (n=5) [64-67] and qualitative designs (n=1) [68]. Over two-thirds of the studies (n=37) were published after the onset of the COVID-19 pandemic, reflecting accelerated remote work and rapid digital transformation in workplaces, with over 26% (n=14) of the studies on this topic published in the year 2022 alone. Table 1 outlines a summary of the key characteristics of all included studies.

XSL•FO
RenderX

Table . Summary of study characteristics that are leading to significant shortages.

Characteristics	Studies per characteristic (n)
Manuscript availability	• Full text (43) • Ongoing (registered) clinical trials or protocols (11)
Study research design	• Protocols (11) • Pilot RCT[a] (3) • RCT (22) • Cross-sectional pre- or postdesign (12) • Mixed method studies (5) • Qualitative (1)
Year of publication	• 2010 (1) • 2013 (1) • 2014 (1) • 2016 (2) • 2018 (8) • 2019 (5) • 2020 (9) • 2021 (7) • 2022 (13) • 2023 (7)
Location of study	• United States (13) • Australia (12) • United Kingdom (4) • South Korea (5) • Finland (3) • Japan (2) • Brazil (1) • Europe (2) • Italy (1) • Sweden (1) • Germany (1) • Spain (1) • Canada (1) • Malaysia (1) • Netherlands (1) • Slovenia (1) • Taiwan (1) • Vietnam (1) • England (2)
Organizational sector	• Across multiple industries (7) • Health care and social services (23) • Information and technology (4) • Education (2) • Agriculture (1) • Construction (1) • Apprentice (1) • Large corporations or consumer-based Industry (5) • Unspecified (10)
Focus of intervention	• Structured self-guided (23) • Unstructured self-guided (15) • Adjunct to other interventions (6)

[a]RCT: randomized controlled trial.

Types of Mental Health App Interventions

A total of 44 unique apps were evaluated across the 54 studies included in this review. There was considerable variability across these apps in terms of their therapeutic approach, goals, functionalities, target population, and intervention duration. Over half of the apps (23/44, 52%) included a programmatic structure with predetermined guidelines or prescribed instructions on how to use them, while 34% (15/44) were open-ended and user driven without a fixed guideline for use. Based on these distinctions, we broadly classified these apps as: (1) structured self-guided, (2) unstructured self-guided, and (3) an adjunct to other interventions. See Table 2 for an

XSL•FO
RenderX

overview of the identified apps categorized by type, including proposed purpose and target population noted in the study.

XSL•FO
RenderX

Table . Overview of apps by type and study target population.

Name of app	Overview	Target population
Unstructured self-guided		
e-pD-Work intervention [16]	Self-guided modules to promote physical exercise, improve sleep, expand relationships, solve problems, improve communication, assertiveness, decision-making, and manage work stress.	General practitioners across multiple primary care centers
The Healthy Minds Program [28]	Meditation app focused on awareness, connection, insight, and a model of well-being.	Employees from district school system
Stress management app (no official name found) [31]	Mental health information and mindfulness strategies (eg, music meditation and breathing).	Nurses employed at college hospitals
Lift Intervention [44]	Self-directed mindfulness program with tips and contents for integrating mindfulness into daily life.	Nurses from COVID-19 units at a single hospital
InMind [48]	Mindfulness training, including measurement of heart rate variability with recommended content.	Employees in 3 organizations: legal, construction, and a PR[a] firm
24alife app [49]	Customized activities for relaxation, exercise training, nutritional advice, and physiological responses to stress based on initial assessment of lifestyle habits	Employees from an Information and Communication Technologies company
Jibun kiroku (a self-record app) [52]	CBT[b]-based intervention focuses on self-monitoring and awareness and recording of negative thoughts, daily activities, and daily mood	Employees experiencing mild to moderate psychological distress
The Wellness Hub app [55]	Mental health–related content and resources, and links to local resources.	Employees from a large health system
Listen Leon [56]	Encourages employees to send anonymous strengths-oriented descriptive feedback to colleagues	Employees from 5 industries: retail, consulting, banking, human resources, and education
Stress management app (no official name found) [57]	Mental health and lifestyle assessment, daily lifestyle management, peer support, tools for searching nearby mental health professionals, and relaxation training	Employees from 3 national hospitals and health promotion centers
Healthy Outcomes at Work app [63,64]	Psychoeducation tailored to the needs of the organization (info re: well-being events, access to vocational rehabilitation assistant, option to provide feedback (via app) to management	Child and Family social workers and hospital staff
The Mood Map [65]	Enables daily mood monitoring (happiness, sadness, anxiety, and anger) at regular intervals. Therapeutic strategies include cognitive reappraisal exercises, breathing, and relaxation techniques.	Employees at a large corporation with high stress levels
Oiva (app available in Finnish only) [66]	Promotes stress management and well-being, based on ACT[c], with short skills building sessions (visual or audio).	Employees from 2 information communication and technology organizations
Brightr app [69]	Monitoring of behavior, sleep, physical activity, nutrition, and shift work by providing tailored feedback based on responses to short in-app questionnaires and sleep and physical activity tracking.	Employees from an information and technology company
Structured self-guided		
Unmind App [18,41]	Six to eight 10 - 20 minute self-guided sessions based on principles of CBT and ACT to manage and improve mental health and well-being. Users had 3 weeks to complete their allocated intervention and were sent a weekly intervention reminder message.	Full-time or part-time employees from a research recruitment platform

XSL•FO
RenderX

Name of app	Overview	Target population
SOLAR[d] [20]	Eight modules over 5 weeks focused on behavioral skill development or psychoeducation, delivered in animations, videos, activities, and notifications, that participants can self-complete at their own pace. Developed as an extension of in-person training.	Firefighters experiencing distress (scored between 7 and 18 on the 6-item Kessler Psychological Distress Scale
WorkingWell [22]	Designed for individuals with severe mental health issues to cope on the job. Users are asked to select up to three goals to work towards and on which to reflect in the coming week. Daily reminders of goals upon app opening and new goals can be chosen for each week.	Employees with severe mental illness using supported employment services
EMPOWER [23]	7-week access to intervention: (1) screening tools for stressful psychosocial working conditions, mental distress, and work function; (2) psychoeducational CBT modules to promote mental health and work functioning.	Employees from SME[e] and public agencies
WEDiary[f] [24]	Two-week intervention that facilitates daily self-monitoring of goals and achievements at work.	Employees across different organizations
Holidaily [25]	Daily prompts based on behavioral activation strategies designed to promote recovery behavior and mental detachment from work to extend the beneficial effects of vacation into their daily work life.	Employees who are awaiting to go on vacation
Headspace [26,27,40,68]	Daily guided 10-minute mindfulness practices or techniques including breathing, imagination, and body scan exercises focusing on different bodily sensations.	Staff from 2 emergency departments (NHS[g]), health care workers, and university employees
BetterLife [29]	Guided, 10 weeks (50 min/wk) self-help therapy informed by principles of CBT and problem-solving therapy for addressing stress, depression, anxiety, and sleep disorders.	Employees with elevated perceived stress (score of 14+ on the Perceived Stress Scale)
Well-being mobile app (official name not found) [30]	A 8-week program (4 lessons/wk) with content focused on techniques to reduce stress and promote well-being (breathing techniques, relaxation techniques, and guided meditation).	Female employees at a large private tertiary care hospital
The MoodHacker [32]	A 6-week intervention with sequenced content and prompts to track mood and activities daily. App tracks mood and positive activity planning, promotes cognitive restructuring, mindful self-awareness, gratitude expression, and identifying and using strengths.	Employees from several institutions recruited in partnership with EAP[h]
Smartphone resiliency training (no official name found) [33]	A 6-week intervention for monitoring and awareness for sleep and mood, happiness and positivity, energy, and focus, and productivity. Users were prompted to select a goal (one of the topics) to work on for the next few weeks.	Employees from a large research hospital and medical center
ABC Stress Management app [36]	Six-week (1 module/wk) intervention with two types of stress management apps—Program A, a free-choice multimodule program and Program B, a fixed-sequential order iCBT[i] multimodule program.	Employees from a large public hospital (Vietnam)
ACT-based app [37]	Six weekly modules based on ACT principles to manage stress. Reflections accessible by a therapist who could send encouraging personal messages to the user.	Middle managers from private sector company

XSL•FO
RenderX

Name of app	Overview	Target population
HeadGear [17,38]	30-day therapeutic intervention (1 challenge/day). Content is based on behavioral activation and mindfulness-based therapies. Features include mood monitoring and access to a tech helpline.	Employees from male-dominated industry partner organizations: agriculture, freight, and mining
Anchored app (adapted from HeadGear app) [45,51]	30-day intervention with daily tasks or "challenge" based on evidence-based techniques (behavioral activation, mindfulness, and coping skills training) delivered using videos, exercise, and goal setting to target stress.	Adults experiencing workplace stress from any industry
HeadGear Apprentice app [67]	30-day therapeutic intervention (1 challenge/day). Modifications to HeadGear include gamification, changes to navigation, and customized content for apprentice support.	Trainees enrolled in an apprenticeship program
Kelaa Mental Resilience app [39]	Four weeks of psychoeducational modules designed to reduce stress and promote well-being specifically in the workplace context by providing options for self-monitoring of behaviors, cognitions, and emotions.	Employees from public and private sectors from 3 countries
Foundations app [43]	Daily active use encouraged during first 2 weeks. User chooses one of 6 focus areas during onboarding (relaxation, sleep, anxious thoughts, feeling down, self-esteem, and stress), then relevant programs and activities are presented based on CBT, mindfulness-based CBT, relaxation techniques, and positive psychology.	Health care workers across health system
Mobile-based Stress Management Intervention [46]	Recommended use of 10 minutes twice a week for 6 weeks. Combination of 3 modules that include interventions for relaxation, self-management, and counseling.	White collar workers with elevated symptoms of perceived stress (Perceived Stress Scale-10≥22)
Calm app [47]	Recommend 10 minutes per day over 8 weeks. Incorporates mindfulness, breathing techniques, and body scans.	Employees of a large consumer electronics retailer
P4Well [50]	Three mobile apps: (1) Nokia Wellness Diary, (2) Firstbeat Mobile Coach (fitness training), and (3) SelfRelax (personalized 4-wk relaxation program).	Male working adults (25 - 45 y) with exhaustion, stress symptoms, or sleeping problems
Naluri app [53]	16-week EAP offering personalized programming to promote behavior change (educational modules, health journal, and a habit tracker) and access to a team of professional health coaches via text-based messages and video calls.	Employees from an unspecified company in Malaysia
Shift app [54,60]	30-day intervention based on content and features of HeadGear app, modified to meet the needs of physician population.	Junior medical officers from 2 hospital sites
VA[j] Mindfulness Coach App [61]	12 audio-guided mindfulness exercises on a range of techniques (eg, body or breath awareness, compassion, and mindful eating).	Nurses from psychiatric facility reporting burnout
Smartphone-based meditation app (No official name found) [62]	Ten days of 10-minute audio-recorded mindfulness sessions.	Firefighters from a large metropolitan city
Adjunct to other interventions		
MATESmobile program [21]	Supplements in-person training with a focus on reinforcing face-to-face training messages and enabling links to mental health support as needed.	Construction workers
Spire Stone wearable device and mobile app [34]	Brief modules for mindfulness training with monitoring of respiratory patterns and real-time biofeedback of physiological stress.	Employees from a large technology corporation across seven cities.

XSL•FO
RenderX

Name of app	Overview	Target population
Dayzz [42]	Personalized sleep training app that is an extension to an online SHAW[k] program.	Employees at a large health care organization.
Brain-sensing eyeglasses and a corresponding mobile app [58]	Eyeglasses collected EEG[l] activity during mindfulness, transferring the data to the app. App provided real-time acoustic feedback from data to promote relaxation.	Professionals with top management duties at a public service company
Smartphone-delivered biofeedback training (no official name found) [35]	Self-managed protocol with videos containing guided meditation practices and real-time biofeedback.	Psychiatric ward nurses in 3 hospitals who experienced workplace violence or abuse

[a]PR: public relations.

[b]CBT: cognitive behavioral therapy.

[c]ACT: acceptance and commitment therapy.

[d]SOLAR: The Skills for Life Adjustment and Resilience.

[e]SME: small and medium enterprises.

[f]WEDiary: work engagement diary.

[g]NHS: National Health Service.

[h]EAP: employee assistance program.

[i]iCBT: internet-based cognitive behavioral therapy.

[j]VA: Veterans Affairs.

[k]SHAW: Sleep Health and Wellness.

[l]EEG: electroencephalogram.

Structured self-guided apps (n=23 apps) had a structured pathway in the form of programs or modules that users needed to complete, often in the form of modules, lessons, exercises, or activities. For instance, the HeadGear app (cited in 8 studies) is a 30-day intervention informed by mindfulness-based techniques that include daily challenges users are instructed to complete [17,38,45,51,54,60,67]. Headspace included daily guided 10-minute mindfulness practices or techniques [26,27,40,68]. Several studies noted that the app content was informed by evidence-based therapies, including behavioral activation and mindfulness [24,38,48,53,62,70], acceptance and commitment therapy [37,42,43,68], and principles of cognitive behavioral therapy [32,59], but varied widely in terms of the extent to which they adhered to evidence-based guidelines.

Unstructured self-guided apps (n=15) were more open-ended and provided users with the choice to select and engage with intervention components on demand. For instance, the Oiva app [66] consisted of information sessions on stress management and well-being based on the principles of acceptance and commitment therapy but had no prescribed guidance for engaging with the app. Many apps included meditation or relaxation exercises [28,31,35,44,48,49], self-monitoring tools (eg, mood, physical activity, and stress) [52,57,65,69], and/or links to resources [55,57].

Finally, there were 5 apps that were designed to be implemented in conjunction with other interventions. Three were used to supplement in-person training [21,34,42], and three included wearable devices linked to the mental health app [34,35,58]. One study, for example, combined a brief in-person mindfulness training followed by monitoring of respiratory patterns through a wearable device, with real-time biofeedback received through the app [34].

In terms of intervention duration, most studies (n=38) implemented a fixed intervention period, typically ranging from 2 to 8 weeks, with 4 weeks being the most common. Outcomes were generally assessed at baseline and at the end of the intervention, although some studies included longer follow-up periods, with the longest extending to 12 months [38].

Study Populations

Participant demographics varied significantly across the studies. Some studies targeted the implementation of the intervention with specific gender groups, among individuals with symptoms of mental health issues and/or in specific workplace sectors. For instance, one study targeted male participants experiencing mild to moderate symptoms of stress and/or depression to assess the feasibility of a multicomponent, technology-based (P4Well) intervention [50]. Two other studies specifically recruited individuals from male-dominated industries to evaluate the effectiveness of a therapeutic app (HeadGear) [38,45]. One study focused on evaluating the effectiveness of a well-being and stress management app for working women [30].

Study participants worked in a range of industries, including health and social services (n=23), which largely aimed at implementing therapeutic apps, large corporations or consumer-based industries (n=5), information and technology (n=5), education (n=2), agriculture (n=1), apprenticeships (n=1), and construction (n=1). A subset of studies implemented their interventions across multiple industries (n=7), and a few studies did not explicitly specify the contextual setting (n=10) in which the research was being conducted.

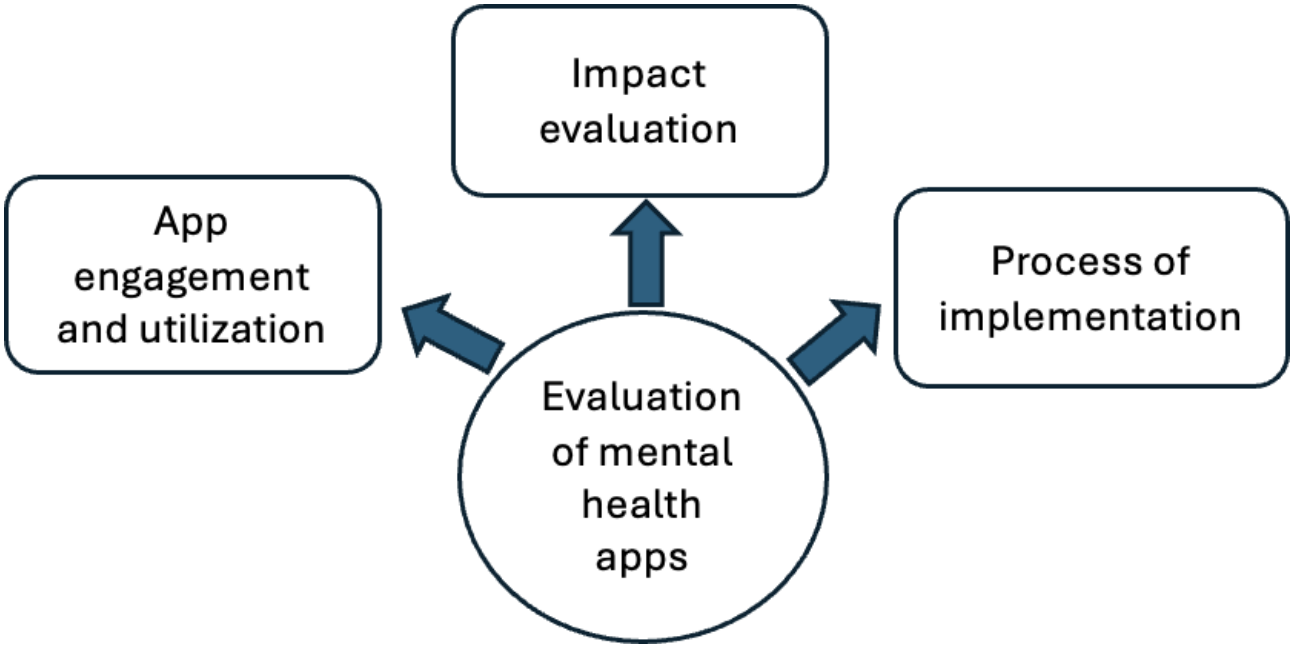
Approach to Evaluation

We identified three primary ways (Figure 2) in which studies evaluated mental health apps with workers: (1) app engagement and utilization; (2) impact evaluation (mental distress, mental

XSL•FO
RenderX

health, and workplace outcomes); and (3) implementation process evaluation. The variables and tools within each approach are summarized in Multimedia Appendix 2.

Figure 2. Three primary approaches to evaluating mental health apps in the workplace.



App Engagement and Utilization

Most of the studies reported on users' engagement with the intervention by tracking app usage patterns. App use ranged from simply tracking the number of downloads or logins to tracking the number of times specific features were used, completion of specific tasks, and tracking the amount of time spent in the app. In the structured interventions, app usage included tracking adherence to the prescribed protocols. The data was typically gathered through built-in analytics that highlighted aggregate usage patterns of the app, although some studies gathered subjective reports of app usage through survey or interview questions. There were at least a dozen studies that did not report app usage data.

User feedback was another common way of tracking engagement, noted in 22 (41%) of the studies. Typically, the follow-up evaluation included exploring perceptions of the app's key features, such as usability, acceptability, quality, motivation to use the app, and overall satisfaction, and/or providing an opportunity to suggest improvements. Standardized measures, such as the Mobile App Rating Scale [70], were used in 5 studies [18,25,41,51,67]; the Systems Usability Scale [71] was used in 2 studies [22,32]; and the eHealth Impact Questionnaire was used in 1 study [16]. Other studies used more informal feedback approaches.

Impact Evaluation

Most of the research on the impact of mental health apps for workers focuses on multiple primary and secondary outcomes, ranging from mental distress to mental health and well-being, as well as specific work-related outcomes for individual app users. Table 3 highlights the most common outcomes that were evaluated, along with an overview of the tools that were used to track impact.

XSL•FO
RenderX

Table . Impact outcomes and measurement tools across studies.

Outcomes and variables	Most frequent measures used
Mental distress	
Symptoms of depression [16-20,23,25,29,31-35,38,40-42,44-47,50-55,57,60,62,67]	• PHQ-9[a] and PHQ-8 (n=16) • DASS-21[b] (n=4) • Center for Epidemiologic Studies Depression Scale (n=3) • Patient Reported Outcomes Measuring Scale (n=3) • Beck Depression Inventory (n=2)
Symptoms of anxiety [16-18,20,23,28,29,31,32,34,38,41,43,45,46,49,51,52,54,55,60,62,65,67]	• GAD-7[c] and GAD-2 (n=16) • STAI[d] (n=2)
Psychological distress [21,37,43,50,52]	• General Health Questionnaire (n=2) • Kessler Psychological Distress Scale (n=2) • General Symptom Index
Perceived stress [25-31,34,37,39,45,46,48,49,57,58,62,63,68]	• PSS[e] (n=16) • Biological markers of stress (n=4)
Psychological trauma [20,55]	• PCL-5[f]
Alcohol use [51,52,55,57]	• AUDIT-C[g] (n=2) • The Daily Drinking Questionnaire (n=1)
Sleep [25,26,33,39,42,47,49]	• Pittsburgh Sleep Quality Index (n=2) • Insomnia Severity Index (n=2) • Fatigue inventory • Sleeping troubles from COPSOQ-II[h]
Mental health	
Well-being [17,18,23,27,28,30,31,33,38-41,43,45,48,51,52,55,56,63,66-68]	• WHO[i]-5 Well-Being Index (n=12) • Warwick-Edinburgh Mental Wellbeing Scale (n=7)
Quality of life [16,23,29,34,50]	• Visual Analog Scale • Euro Quality of Life Questionnaire • World Health Organization Quality of Life • CDC[j] health-related quality of life metrics
Resilience [17,35,38,39,43,45,47,51,55,57]	• Connor-Davidson Resilience Scale (n=3) • Brief Resilience Scale (n=4) • Resilience Scale (n=2) • Korean Resilience Quotient
Self-compassion [28,33,40]	• Self-Compassion Scale (n=2)
Life satisfaction [28,49,50,66]	• 5-item SWLS[k] (n=2) • Meaning in Life Questionnaire–Presence Subscale
Mindfulness [26-28,40,66]	• Five Facets Mindfulness Questionnaire (n=3) • Mindful Attention Awareness Scale
Social connections [28,57]	• Social Readjustment Rating Scale • Social connection (NIH[l] Toolbox Loneliness Questionnaire)
Work-related outcomes	
Work stress [26,31,35,51,57]	• SISQ[m] • Korean Occupational Stress Scale (n=2) • Siegrist Job Strain Scale • Occupational Stress Indicator-2

XSL•FO
RenderX

Outcomes and variables	Most frequent measures used
Burnout [25-27,33,40,44,45,49,50,61,62,68]	• Maslach Burnout Inventory (n=8) • Bergen Burnout Inventory (n=3)
Work engagement [22,24,26,29,36,46,56,63,66]	• Utrecht Work Engagement Scale (9 items; n=7) • Job motivation (job match survey and Multidimensional Work Motivation Scale; n=2)
Job satisfaction [16,22,63]	• Self-report (n=3)
Work performance or productivity [17,23,32,38,39,41-43,45,47,51,56,60,63,67]	• HPQ[n] (n=6) • The Work Productivity and Activity Impairment Questionnaire (n=3) • WLQ[o] • Stanford Presenteeism Scale • Productivity cost (count of absenteeism and presenteeism; n=2)
Working conditions [16,21,31,39,56,63,64]	• MSIT[p] (n=2) • Job Content Questionnaire • Korean-Emotional Labor scale • Psychosocial safety climate • Strengths Use and Deficit Correction scale
Other outcomes	
Mental health knowledge [21,32]	• Author developed 14 multiple-choice scale • GAT[q] suicide awareness questionnaire
Help-seeking or health care utilization [21,23,42,43,47,54,60,72]	• General Help Seeking Questionnaire (n=2) • Intentions of help seeking • Self-report of medical visits or help accessed (n=3)
Healthy behaviors or physical function [23,49,57]	• IPAQ[r] • Fitness test (strength and flexibility) • Exercise and diet patterns
Cognitive function [28,50,58]	• Neurocognitive efficiency tests • Psychological flexibility (Acceptance and Action Questionnaire-2) • Perseverative thinking and Drexel Defusion Scale

[a]PHQ: Patient Health Questionnaire.
[b]DASS-21: Depression, Anxiety, and Stress Scale–21.
[c]GAD: Generalized Anxiety Disorder.
[d]STAI: State-Trait Anxiety Inventory.
[e]PSS: Perceived Stress Scale.
[f]PCL-5: Posttraumatic Stress Disorder Checklist for *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*.
[g]AUDIT-C: Alcohol Use Disorders Identification Test–Concise.
[h]COPSOQ-II: Copenhagen Psychosocial Questionnaire-II.
[i]WHO: World Health Organization.
[j]CDC: Centers for Disease Control and Prevention.
[k]SWLS: Satisfaction with Life Scale.
[l]NIH: National Institutes of Health.
[m]SISQ: Single-Item Stress Question.
[n]HPQ: Health and Work Performance Questionnaire.
[o]WLQ: Work Limitations Questionnaire.
[p]MSIT: Management Standards Indicator Tool.
[q]GAT: General Awareness Training.
[r]IPAQ: International Physical Activity Questionnaire.

Impact on mental distress was the most common outcome, evaluated in 42 of the 54 (78%) studies (Table 3). Symptoms of depression were evaluated in 30 (56%) of the studies, often using standardized tools such as the Patient Health Questionnaire-9 [73]. Symptoms of anxiety were evaluated in 24 (44%) of the studies, many using the Generalized Anxiety

XSL•FO
RenderX

Disorder (GAD-2 or GAD-7) [74]. Perceived stress was another common outcome measure used in 18 of the 54 (33%) studies, with tools such as the Perceived Stress Scale [75]. It is interesting to note that four of the studies incorporated biological indicators of stress [49,58,62]. Other measures of change in mental distress included tools to track symptoms of trauma, alcohol use, and sleep or insomnia.

Impact on mental health and well-being was another common outcome, evaluated across 34 (63%) studies. The primary outcome was the assessment of impact on well-being (n=24, 44%), using standardized tools such as the World Health Organization–5 (WHO-5) Well-Being Index [76] or the 14-item Warwick-Edinburgh Wellbeing Scale [77]. Other well-being outcomes included resilience (n=10), life satisfaction (n=7), quality of life (n=5), mindfulness (n=5), self-compassion (n=3), and social connection (n=2). These measures are congruent with many of the tools that focused on mindfulness and psychoeducation for coping with stress.

Impact on work-related outcomes was examined in 37 (68%) studies. Common work-related outcomes included work productivity and performance, burnout, job satisfaction, work engagement, and subjective perceptions of work stress and working conditions. Work performance outcomes, noted in at least 15 (30%) studies [17,23,32,38,39,41-43,45,47,51,56,60,63,67], included standardized tools such as the Work Limitations Questionnaire, the WHO Health and Work Performance Questionnaire, and the Work Productivity and Activity Impairment Questionnaire. Many of these tools included presenteeism or absenteeism measures, but several studies evaluated these separately [23,40,43,45,63,67], with some noting the productivity cost for organizations [23,47,67]. Burnout was another relatively common outcome noted in 12 (22%) of the studies [25-27,33,40,44,45,49,50,61,62,68], and evaluated using tools such as the Maslach Burnout Inventory or Bergen Burnout Inventory. Positive measures of job satisfaction were noted in four studies, often tracked through self-report [16,19,22,63]. Work engagement was another outcome noted in 9 (17%) studies [22,24,26,29,36,46,56,63,66], typically measured by tools such as the Utrecht Work Engagement Scale. In addition to work outcomes, 14 (26%) studies evaluated whether using a mental health app could change users' perceptions of their working conditions including perceived work stress [25,26,31,35,51,57], job demands or strain [16,31,36], psychosocial climate [21,23], and perceived social support, including support from leaders and co-workers [39,43,63,64].

Other outcomes used to track impact included changes in mental health knowledge [21,32], cognitive function [28,50,58], physical function [49], or engagement in healthy behaviors related to diet and exercise [23,57]. Help-seeking or health care utilization was another outcome noted in 7 (13%) studies [21,23,42,43,47,54,60], particularly with apps that promoted connection to mental health resources. It is interesting to note that very few studies evaluated behavior change.

Process of Implementation

While most studies evaluated the overall effectiveness of the technology on individual or work-related outcomes, only a limited number of studies (n=5) focused on evaluating how the process of implementation affected these outcomes. Protocol studies included in this review proposed examining the relationship between the extent of app usage [19] in terms of its influence on the overall effectiveness of the intervention and understanding strategies needed to initiate and promote the use of the app [22]. Implementation outcomes in 3 other studies involved the evaluation of individual, contextual, and implementation-related barriers and facilitators that influenced the uptake, use, and overall impact of the mental health intervention [66,68,69]. Process measures were incorporated in order to generate recommendations for optimizing the utilization and effectiveness of the mental health technology.

Discussion

Principal Findings

This scoping review identified 54 studies (43 completed RCTs and 11 protocols) evaluating mobile health (mHealth) applications in workplace contexts. The main trends in the research were (1) research on workplace mental health apps has expanded rapidly in recent years, particularly following the pandemic; (2) the majority of studies were conducted in health and social services, with few in male-dominated industries; (3) about half of the 44 distinct apps adopted a structured, self-guided approach with modules or activities for the user to complete, another third were unstructured apps to be used as needed, and 11% of the apps were designed to be used in conjunction with a wearable sensor or web-based program; and (4) commonly evaluated outcomes included mental distress (including perceived stress, anxiety, and depression), mental wellness (well-being and resilience), and workplace-related outcomes (burnout, workplace engagement, work productivity, and performance), with limited studies on the process of implementation. Together, these findings highlight important gaps in both the design and evaluation of workplace mental health apps, underscoring the need for greater use of implementation science frameworks to guide future research.

Study Characteristics

The results of this review profiled 43 completed studies and 11 study protocols that evaluate the impact of mHealth interventions on a working population. A previous systematic review, published in 2019, on e–mental health interventions in the workplace identified 22 studies, but only 3 studies included apps as a primary intervention [78]. The pandemic and post-pandemic era have precipitated a rapid increase in app-related research with workers (41 studies in the past 5 years). This may reflect a shift to mHealth approaches to support the growing mental health needs of workers. The number of completed randomized clinical trials (24/54, 44%) is promising, reflecting high-quality research designs to evaluate impact. Furthermore, this review included 11 trial registrations or published protocols, reflecting that the research is ongoing. The rise of digital technology and associated research on mental health apps for workers speaks to the importance of summarizing current evaluation practices and gaps in the evidence.

XSL•FO
RenderX

Types of Mental Health Applications

The review also highlighted diversity in the nature of mHealth interventions, with 44 different types of unique applications across the mental health spectrum, from promoting wellness to managing illness. In fact, several studies included wearable devices linked to the app, which may reflect a new technology trend. As noted in other review studies of mobile apps in the workplace, many were based on principles of cognitive behavioral therapy or mindfulness, combining information sharing with self-monitoring and self-management activities [11-13,79]. It is interesting to note, however, that even in the structured apps, the mechanisms of action were not consistently explained. There were a few studies that explicitly cited theories such as behavioral activation and cognitive behavioral approaches, but many of the app tools and techniques were not necessarily embedded within a psychotherapeutic protocol. This is a significant gap since explicit links between app design and expected outcomes are important aspects of evidence-based interventions [80].

Over half of the studies featured structured, self-guided apps with specific modules or activities that users were directed to complete, often over a 2 to 8-week period. Outcomes related to engagement, including completion of the activities, were more amenable to tracking "success" than apps that were less structured and available to use "as needed." Given high attrition rates in app use, recommendations for optimizing engagement noted in the literature include providing "challenges" within the technology over short periods to build habits, providing regular "push notifications," and connecting with an online coach or therapist [81]. Additional research is needed, however, to systematically evaluate the effect of these implementation strategies, and longitudinal research is needed to track the impact of these strategies over time [79,81]. Only a few apps (Headspace, Headgear, or Shift) had more than 2 studies examining their effectiveness, reflecting the need for more robust evaluation.

Study Participants

Although studies in this review were conducted across a range of sectors, the highest proportion (23/54, 42%) was in health and social services. Given the mental health crisis noted in health and social services that is leading to significant shortages in staffing [44,68], these studies indicate promising efforts to evaluate innovative strategies to support the mental health of this group of workers. Studies in other sectors were more limited, although there were several in service-based sectors (education and information technology), with comparatively fewer studies in male-dominated industries such as construction. Unfortunately, 10 of the studies did not specify the workplace context where the app was implemented, making it difficult to identify relevant recommendations for specific workplace sectors. Given the importance of the workplace context, future comparative research is needed to explore how mobile apps function differently across industry sectors and organizations. Furthermore, many studies included in this review lacked sufficient reporting details regarding the intervention, research setting, and evaluation plan, making it challenging for future reproducibility, comparability, and dissemination. There are guidelines, such as the 16-item mERA (mHealth Evidence Reporting and Assessment) checklist, that can be used to improve the reporting quality of interventions and support replicability of research [82]. This checklist provides comprehensive recommendations for study authors on adequately describing the content, context, and technology infrastructure to implement the mobile health technology with a specific population group [82].

Key Evaluation Variables and Tools

The review highlights 3 different categories of evaluations conducted by studies exploring mental health apps. User engagement and feedback were important outcomes tracked across many studies, either objectively through app metrics or subjectively through a recall of usage patterns. Feedback about usability, acceptability, and relevance was tracked qualitatively but also through standardized measures such as the Mobile Application Rating Scale [70] and the System Usability Scale [71]. Tracking user perceptions, engagement, and interaction with mental health apps is critical in establishing a relationship between usage (dose) and desired outcomes (response) [79]. Individual studies varied in their operational definitions of these constructs along with how they were measured, ranging from objective measures of usage (frequency and duration of use) that are passively collected from within the app to subjective measures of user ratings of usefulness and satisfaction, making it difficult to conduct cross-study comparisons.

Currently, there is no consensus on what constitutes optimal engagement for a mental health app to produce desired effects [83,84]. One study exploring different types of app use, features, and their effect on outcomes found that the right combination of different features (ie, learning, goal setting, and self-tracking) and their use at a moderate level could lead to desired outcomes, thereby refuting the common belief that higher usage always leads to better outcomes [85]. Collectively, each of the identified measures could be valuable in determining whether users who reported higher levels of usage, found the app useful, and were satisfied were more likely to engage with the intervention optimally, subsequently leading to significant improvements in outcomes. However, only 1 study protocol proposed assessing the link between the dose (the extent to which an app needs to be used) and the response (a significant improvement in outcomes) [19].

In addition to assessing usage and engagement, user experience and feedback were typically assessed only in feasibility studies. Since apps require iterative updates and maintenance, assessing user experience and feedback over time is important when attempting to deploy them in a large-scale, real-world setting [84].

In terms of impact, it is interesting to note that there was a range of approaches to evaluation, ranging from impact on reducing mental distress (78%) to improving mental well-being (63%) or work-related outcomes (68%). Since all of the studies were conducted in the context of the workplace, these workplace-related outcomes may be important for organizational decision-makers who are considering the value of investing in mHealth technology. Although there were 37 studies that considered workplace outcomes, there was a wide range of tools

XSL•FO
RenderX

to evaluate these outcomes, ranging from downstream measures of productivity and burnout to more upstream measures of work engagement, perceived working conditions, and job satisfaction. Although there is a growing body of evidence indicating that digital mental health interventions implemented in workplaces can improve work effectiveness and productivity [13], tracking these outcomes can be complex, considering that studies need to speculate when changes in these workplace performance outcomes will take place and embed appropriate follow-up time points to measure them [86]. Study designs need to clearly outline the components of the intervention and the mechanisms of change to define what constitutes proximal (short-term or immediate) outcomes and distal (long-term) outcomes and determine appropriate time points for the measurement of these outcomes.

Even when apps are designed to be evidence-based, this does not guarantee that the expected outcomes will occur. Evaluating a complex mental health intervention in the context of the workplace requires a multilevel approach, considering the implementation process as well as its impact [81,86]. While several studies in this review acknowledged the complexities of implementing a digital intervention in a workplace setting, very few systematically evaluated the process of implementation, and there appears to be a gap in the application of existing theories to systematically guide this inquiry.

Several implementation science theories and frameworks can be useful to guide further research in understanding and unpacking the complexities of implementing technology across different contexts and achieving large-scale adoption and sustained effects. For example, the Consolidated Framework for Implementation Research (CFIR) provides a summary of evidence-based constructs across 5 domains: the intervention, the individual user, the inner context, the outer context, and the process of implementation [87]. These constructs can help identify and explain the facilitators and barriers to the uptake and impact of the intervention [87]. Another similar framework, called the nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) framework [88], highlights six key domains that can be used to guide researchers in evaluating the success of an mHealth intervention implemented in an organizational context. It considers the characteristics of (1) the condition being addressed (eg, mental health), (2) the technology-based intervention, (3) the adopters of the technology, (4) the value proposition for the users, (5) the organization where the technology is being implemented, and (6) the wider system or society. The NASSS framework also reflects on the complex and dynamic interaction between the 6 domains over time, which can help answer questions related to users' engagement with apps over time [88]. Systematic evaluation of the implementation process can facilitate exploration of key facilitators and barriers that influence the uptake and effectiveness of the intervention. While traditional health care interventions typically emphasize evaluation of their effects on key outcomes, mental health app–based interventions necessitate the examination of the different ways in which users engage with the intervention as well as exploring other implementation-related and contextual factors that influence the uptake and overall impact [79-81,83,86].

Limitations

To our knowledge, this review is one of the first to summarize current research regarding approaches to evaluating mental health apps specifically in the context of the workplace. While a broad objective was established at the onset of the review in order to sufficiently capture all relevant studies, we restricted our search to studies that were published in English to ensure feasibility in terms of timelines and resources. Another limitation was the focus only on descriptions of the evaluation approaches outlined in the studies. There was no formal assessment of the study quality or synthesis of study findings. Future systematic reviews are needed to critically evaluate the current research evidence to identify the impact of mental health apps, and to profile what works for whom and in what context. This is particularly important to assist organizational leaders, policymakers, and stakeholders in deciding how to invest in and optimize the use of mental health apps with frontline workers.

Recommendations

Results from this study informed the following recommendations for future research regarding app-based mental health interventions in the workplace. First, future research should improve transparency and consistency in the reporting of studies by incorporating guidelines such as the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) [89]. Second, evaluation studies need to identify and measure appropriate outcomes by examining the congruence between intervention goals and outcomes when exploring the effectiveness of mental health apps. Third, there is a need for longitudinal studies to assess the long-term effects of mental health apps, particularly when evaluating their impact on workplace outcomes that may require a longer duration to detect meaningful changes. Fourth, long-term effects of mental health apps, particularly when evaluating their impact on workplace outcomes that may require longer duration to detect meaningful changes. Fifth, there is a need to investigate strategies for integrating mental health apps into existing workplace settings by considering organizational factors and fostering an implementation science lens. Applying implementation theory can generate new insights about what works for whom and in what context. Sixth, there is a need to investigate strategies for integrating mental health apps into existing workplace settings by considering organizational factors and fostering an implementation science lens. Applying implementation theory can generate new insights about what works for whom and in what context. Finally, a future systematic review is needed to synthesize and appraise current evidence regarding the effectiveness of mental health apps for workers.

Conclusions

In conclusion, the past decade has seen tremendous advancements in the field of mHealth innovation, including its application in the context of the workplace. Traditional experimental approaches to evaluating mHealth interventions, however, may not be sufficient to capture the complexities of the facilitators and barriers that shape the adoption of the intervention and the sustainability of its effects within diverse

XSL•FO
RenderX

workplaces. This methodological gap calls for research in this field to embrace new implementation science methodologies to truly capture the nuances of technology adoption and reliably measure its effectiveness over time.

Acknowledgments

The authors thank Meghan Lees, who assisted as a second reviewer in the initial screening process, and Melissa Parker and Ante Cuvalo, who provided assistance with revisions to the final paper.

Funding

The authors would like to acknowledge the support for the article processing fee by the Public Health Agency of Canada.

Data Availability

Data tables generated and analyzed during this study are included in this published study and in multimedia appendices.

Authors' Contributions

SA led the scoping review, including conceptualization, data collection, analysis, and drafting of initial and revised versions of the paper. SM provided supervision throughout the process, from conceptualization to analysis, and actively supported the writing of the initial and revised drafts of the paper. LM and JM provided input into the analysis of the study findings and contributed to the review and editing of the final paper. Generative artificial intelligence was not used in any portion of the manuscript writing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 20 KB - mhealth_v14i1e57046_app1.docx]

Multimedia Appendix 2
Study summary table.
[DOCX File, 78 KB - mhealth_v14i1e57046_app2.docx]

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 111 KB - mhealth_v14i1e57046_app3.docx]

References

1. The Lancet Global Health. Mental health matters. Lancet Glob Health 2020 Nov;8(11):e1352. [doi: 10.1016/S2214-109X(20)30432-0] [Medline: 33069297]
2. Brooks SK, Webster RK, Smith LE, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920. [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
3. Tulk C, Bartram M, Leslie K, Atanackovic J, Chamberland-Rowe C, Bourgeault IL. The impact of COVID-19 on the mental health and substance use health (MHSUH) workforce in Canada: a mixed methods study. Hum Resour Health 2023 Feb 8;21(1):9. [doi: 10.1186/s12960-023-00797-6] [Medline: 36755246]
4. Rauschenberg C, Schick A, Hirjak D, et al. Evidence synthesis of digital interventions to mitigate the negative impact of the COVID-19 pandemic on public mental health: rapid meta-review. J Med Internet Res 2021 Mar 10;23(3):e23365. [doi: 10.2196/23365] [Medline: 33606657]
5. Olff M. Mobile mental health: a challenging research agenda. Eur J Psychotraumatol 2015;6(1):27882. [doi: 10.3402/ejpt.v6.27882] [Medline: 25994025]
6. Watts SE, Andrews G. Internet access is not restricted globally to high income countries: So why are evidenced based prevention and treatment programs for mental disorders so rare? Asian J Psychiatr 2014 Aug;10:71-74. [doi: 10.1016/j.ajp.2014.06.007] [Medline: 25042956]
7. Firth J, Torous J, Nicholas J, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298. [doi: 10.1002/wps.20472] [Medline: 28941113]

XSL•FO
RenderX

8. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 1;3(1):e7. [doi: 10.2196/mental.4984] [Medline: 26932350]
9. Seegan PL, Miller MJ, Heliste JL, Fathi L, McGuire JF. Efficacy of stand-alone digital mental health applications for anxiety and depression: a meta-analysis of randomized controlled trials. J Psychiatr Res 2023 Aug;164:171-183. [doi: 10.1016/j.jpsychires.2023.06.019] [Medline: 37352813]
10. Wasil AR, Palermo EH, Lorenzo-Luaces L, DeRubeis RJ. Is there an app for that? A review of popular apps for depression, anxiety, and well-being. Cogn Behav Pract 2022 Nov;29(4):883-901. [doi: 10.1016/j.cbpra.2021.07.001]
11. Stratton E, Jones N, Peters SE, Torous J, Glozier N. Digital mHealth interventions for employees: systematic review and meta-analysis of their effects on workplace outcomes. J Occup Environ Med 2021 Aug 1;63(8):e512-e525. [doi: 10.1097/JOM.0000000000002267] [Medline: 34050095]
12. Howarth A, Quesada J, Silva J, Judycki S, Mills PR. The impact of digital health interventions on health-related outcomes in the workplace: a systematic review. Digit Health 2018;4:2055207618770861. [doi: 10.1177/2055207618770861] [Medline: 29942631]
13. Carolan S, Harris PR, Cavanagh K. Improving employee well-being and effectiveness: systematic review and meta-analysis of web-based psychological interventions delivered in the workplace. J Med Internet Res 2017 Jul 26;19(7):e271. [doi: 10.2196/jmir.7583] [Medline: 28747293]
14. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
15. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
16. Bellón JA, Rodríguez-Morejón A, Conejo-Cerón S, et al. A personalized intervention to prevent depression in primary care based on risk predictive algorithms and decision support systems: protocol of the e-predictD study. Front Psychiatry 2023;14:1163800. [doi: 10.3389/fpsyt.2023.1163800] [Medline: 37333911]
17. Deady M, Johnston DA, Glozier N, et al. A smartphone application for treating depressive symptoms: study protocol for a randomised controlled trial. BMC Psychiatry 2018 Jun 1;18(1):166. [doi: 10.1186/s12888-018-1752-5] [Medline: 29859060]
18. Economides M, Male R, Bolton H, Cavanagh K. Feasibility and preliminary efficacy of app-based audio tools to improve sleep health in working adults experiencing poor sleep: a multi-arm randomized pilot trial. Sleep 2023 Jul 11;46(7):zsad053. [doi: 10.1093/sleep/zsad053] [Medline: 36919571]
19. Kosenkranius MK, Rink FA, de Bloom J, van den Heuvel M. The design and development of a hybrid off-job crafting intervention to enhance needs satisfaction, well-being and performance: a study protocol for a randomized controlled trial. BMC Public Health 2020 Jan 28;20(1):115. [doi: 10.1186/s12889-020-8224-9] [Medline: 31992266]
20. Metcalf O, Gibson K, Fredrickson J, Finlayson-Short L, Varker T, O'Donnell M. Design, development and randomised controlled trial protocol of a smartphone-delivered version of "SOLAR" for emergency service workers to manage stress and trauma. BMJ Open 2023 Feb 10;13(2):e062710. [doi: 10.1136/bmjopen-2022-062710] [Medline: 36764725]
21. Milner A, King TL, Scovelle AJ, et al. A blended face-to-face and smartphone intervention for suicide prevention in the construction industry: protocol for a randomized controlled trial with MATES in Construction. BMC Psychiatry 2019 May 14;19(1):146. [doi: 10.1186/s12888-019-2142-3] [Medline: 31088405]
22. Nicholson J, Wright SM, Carlisle AM. Pre-post, mixed-methods feasibility study of the WorkingWell mobile support tool for individuals with serious mental illness in the USA: a pilot study protocol. BMJ Open 2018 Feb 6;8(2):e019936. [doi: 10.1136/bmjopen-2017-019936] [Medline: 29437756]
23. Olaya B, Van der Feltz-Cornelis CM, Hakkaart-van Roijen L, et al. Study protocol of EMPOWER: a cluster randomized trial of a multimodal eHealth intervention for promoting mental health in the workplace following a stepped wedge trial design. Digit Health 2022;8:20552076221131145. [doi: 10.1177/20552076221131145] [Medline: 36276189]
24. Tokita M, Kobayashi S, Miyanaka D, et al. Effects of a smartphone-based positive reflection diary on work engagement among Japanese workers: randomized controlled trial. JMIR Mhealth Uhealth 2025 Sep 25;13:e55664. [doi: 10.2196/55664] [Medline: 40997313]
25. Smyth A, de Bloom J, Syrek C, et al. Efficacy of a smartphone-based intervention—"Holidaily"—promoting recovery behaviour in workers after a vacation: study protocol for a randomised controlled trial. BMC Public Health 2020 Aug 26;20(1):1286. [doi: 10.1186/s12889-020-09354-5] [Medline: 32843021]
26. Rich RM, Ogden J, Morison L. A randomized controlled trial of an app-delivered mindfulness program among university employees: effects on stress and work-related outcomes. Int J Workplace Health Manag 2021 Feb 22;14(2):201-216. [doi: 10.1108/IJWHM-04-2020-0046]
27. Xu HG, Eley R, Kynoch K, Tuckett A. Effects of mobile mindfulness on emergency department work stress: a randomised controlled trial. Emerg Med Australas 2022 Apr;34(2):176-185. [doi: 10.1111/1742-6723.13836] [Medline: 34378320]
28. Hirshberg MJ, Frye C, Dahl CJ, et al. A randomized controlled trial of a smartphone-based well-being training in public school system employees during the COVID-19 pandemic. J Educ Psychol 2022 Nov;114(8):1895-1911. [doi: 10.1037/edu0000739] [Medline: 36387982]

XSL•FO
RenderX

29. Hwang H, Kim SM, Netterstrøm B, Han DH. The efficacy of a smartphone-based app on stress reduction: randomized controlled trial. J Med Internet Res 2022 Feb 15;24(2):e28703. [doi: 10.2196/28703] [Medline: 35166687]
30. Coelhoso CC, Tobo PR, Lacerda SS, et al. A new mental health mobile app for well-being and stress reduction in working women: randomized controlled trial. J Med Internet Res 2019 Nov 7;21(11):e14269. [doi: 10.2196/14269] [Medline: 31697244]
31. Hwang WJ, Jo HH. Evaluation of the effectiveness of mobile app–based stress-management program: a randomized controlled trial. Int J Environ Res Public Health 2019 Nov 3;16(21):4270. [doi: 10.3390/ijerph16214270] [Medline: 31684169]
32. Birney AJ, Gunn R, Russell JK, Ary DV. MoodHacker mobile web app with email for adults to self-manage mild-to-moderate depression: randomized controlled trial. JMIR Mhealth Uhealth 2016 Jan 26;4(1):e8. [doi: 10.2196/mhealth.4231] [Medline: 26813737]
33. Mistretta EG, Davis MC, Temkit M, Lorenz C, Darby B, Stonnington CM. Resilience training for work-related stress among health care workers: results of a randomized clinical trial comparing in-person and smartphone-delivered interventions. J Occup Environ Med 2018 Jun;60(6):559-568. [doi: 10.1097/JOM.0000000000001285] [Medline: 29370014]
34. Smith EN, Santoro E, Moraveji N, Susi M, Crum AJ. Integrating wearables in stress management interventions: promising evidence from a randomized trial. Int J Stress Manag 2020;27(2):172-182. [doi: 10.1037/str0000137]
35. Hsieh HF, Huang IC, Liu Y, Chen WL, Lee YW, Hsu HT. The effects of biofeedback training and smartphone-delivered biofeedback training on resilience, occupational stress, and depressive symptoms among abused psychiatric nurses. Int J Environ Res Public Health 2020 Apr 22;17(8):2905. [doi: 10.3390/ijerph17082905] [Medline: 32331460]
36. Sasaki N, Imamura K, Tran TTT, et al. Effects of smartphone-based stress management on improving work engagement among nurses in Vietnam: secondary analysis of a three-arm randomized controlled trial. J Med Internet Res 2021 Feb 23;23(2):e20445. [doi: 10.2196/20445] [Medline: 33620328]
37. Ly KH, Asplund K, Andersson G. Stress management for middle managers via an acceptance and commitment-based smartphone application: a randomized controlled trial. Internet Interv 2014 Jul;1(3):95-101. [doi: 10.1016/j.invent.2014.06.003]
38. Deady M, Glozier N, Calvo R, et al. Preventing depression using a smartphone app: a randomized controlled trial. Psychol Med 2022 Feb;52(3):457-466. [doi: 10.1017/S0033291720002081] [Medline: 32624013]
39. Weber S, Lorenz C, Hemmings N. Improving stress and positive mental health at work via an app-based intervention: a large-scale multi-center randomized control trial. Front Psychol 2019;10:2745. [doi: 10.3389/fpsyg.2019.02745] [Medline: 31866915]
40. Taylor H, Cavanagh K, Field AP, Strauss C. Health care workers' need for Headspace: findings from a multisite definitive randomized controlled trial of an unguided digital mindfulness-based self-help app to reduce healthcare worker stress. JMIR Mhealth Uhealth 2022 Aug 25;10(8):e31744. [doi: 10.2196/31744] [Medline: 36006668]
41. Taylor CB, Fitzsimmons-Craft EE, Graham AK. Digital technology can revolutionize mental health services delivery: the COVID-19 crisis as a catalyst for change. Int J Eat Disord 2020 Jul;53(7):1155-1157. [doi: 10.1002/eat.23300] [Medline: 32449523]
42. Robbins R, Weaver MD, Quan SF, et al. Evaluating the impact of a sleep health education and a personalised smartphone application on sleep, productivity and healthcare utilisation among employees: results of a randomised clinical trial. BMJ Open 2022 Sep 14;12(9):e062121. [doi: 10.1136/bmjopen-2022-062121] [Medline: 36104122]
43. Gnanapragasam SN, Tinch-Taylor R, Scott HR, et al. Multicentre, England-wide randomised controlled trial of the "Foundations" smartphone application in improving mental health and well-being in a healthcare worker population. Br J Psychiatry 2023 Feb;222(2):58-66. [doi: 10.1192/bjp.2022.103] [Medline: 36040419]
44. Pratt EH, Hall L, Jennings C, et al. Mobile mindfulness for psychological distress and burnout among frontline COVID-19 nurses: a pilot randomized trial. Ann Am Thorac Soc 2023 Oct;20(10):1475-1482. [doi: 10.1513/AnnalsATS.202301-025OC] [Medline: 37289650]
45. Deady M, Collins DAJ, Lavender I, et al. Selective prevention of depression in workers using a smartphone app: randomized controlled trial. J Med Internet Res 2023 Aug 24;25:e45963. [doi: 10.2196/45963] [Medline: 37616040]
46. Lee YJ. Effects of a mobile health intervention on activities of stress self-management for workers. Work 2023;75(1):233-241. [doi: 10.3233/WOR-211406] [Medline: 36591675]
47. Huberty JL, Espel-Huynh HM, Neher TL, Puzia ME. Testing the pragmatic effectiveness of a consumer-based mindfulness mobile app in the workplace: randomized controlled trial. JMIR Mhealth Uhealth 2022 Sep 28;10(9):e38903. [doi: 10.2196/38903] [Medline: 36169991]
48. Yoon SI, Lee SI, Suh HW, Chung SY, Kim JW. Effects of mobile mindfulness training on mental health of employees: A CONSORT-compliant pilot randomized controlled trial. Medicine (Baltimore) 2022 Sep 2;101(35):e30260. [doi: 10.1097/MD.0000000000030260] [Medline: 36107583]
49. Jukic T, Ihan A, Strojnik V, Stubljar D, Starc A. The effect of active occupational stress management on psychosocial and physiological wellbeing: a pilot study. BMC Med Inform Decis Mak 2020 Dec 3;20(1):321. [doi: 10.1186/s12911-020-01347-z] [Medline: 33272279]

XSL•FO
RenderX

50. Lappalainen P, Kaipainen K, Lappalainen R, et al. Feasibility of a personal health technology-based psychological intervention for men with stress and mood problems: randomized controlled pilot trial. JMIR Res Protoc 2013 Jan 9;2(1):e1. [doi: 10.2196/resprot.2389] [Medline: 23611946]
51. Collins DAJ, Harvey SB, Lavender I, Glozier N, Christensen H, Deady M. A pilot evaluation of a smartphone application for workplace depression. Int J Environ Res Public Health 2020 Sep 16;17(18):6753. [doi: 10.3390/ijerph17186753] [Medline: 32947994]
52. Hamamura T, Suganuma S, Ueda M, Mearns J, Shimoyama H. Standalone effects of a cognitive behavioral intervention using a mobile phone app on psychological distress and alcohol consumption among Japanese workers: pilot nonrandomized controlled trial. JMIR Ment Health 2018 Mar 22;5(1):e24. [doi: 10.2196/mental.8984] [Medline: 29567634]
53. Jesuthasan J, Low M, Ong T. The impact of personalized human support on engagement with behavioral intervention technologies for employee mental health: an exploratory retrospective study. Front Digit Health 2022;4:846375. [doi: 10.3389/fdgth.2022.846375] [Medline: 35574254]
54. Counson I, Bartholomew A, Crawford J, et al. Development of the Shift smartphone app to support the emotional well-being of junior physicians: design of a prototype and results of usability and acceptability testing. JMIR Form Res 2021 Dec 2;5(12):e26370. [doi: 10.2196/26370] [Medline: 34860662]
55. Golden EA, Zweig M, Danieletto M, et al. A resilience-building app to support the mental health of health care workers in the COVID-19 era: design process, distribution, and evaluation. JMIR Form Res 2021 May 5;5(5):e26590. [doi: 10.2196/26590] [Medline: 33872189]
56. Gradito Dubord MA, Forest J, Balčiūnaitė LM, Rauen E, Jungert T. The power of strength-oriented feedback enlightened by self-determination theory: a positive technology-based intervention. J Happiness Stud 2022 Aug;23(6):2827-2848. [doi: 10.1007/s10902-022-00524-3]
57. Baek JH, Kim JH, Oh S, Kim JY, Baik S. Smart stress care: usability, feasibility and preliminary efficacy of fully automated stress management application for employees. Psychiatry Investig 2018 Oct;15(10):991-999. [doi: 10.30773/pi.2018.08.10] [Medline: 30301304]
58. Crivelli D, Fronda G, Venturella I, Balconi M. Stress and neurocognitive efficiency in managerial contexts: a study on technology-mediated mindfulness practice. Int J Workplace Health Manag 2019;12:42-56. [doi: 10.1108/IJWHM-07-2018-0095]
59. Deady M, Johnston D, Milne D, et al. Preliminary effectiveness of a smartphone app to reduce depressive symptoms in the workplace: feasibility and acceptability study. JMIR Mhealth Uhealth 2018 Dec 4;6(12):e11661. [doi: 10.2196/11661] [Medline: 30514694]
60. Sanatkar S, Counson I, Mackinnon A, Bartholomew A, Glozier N, Harvey S. Preliminary investigation of Shift, a novel smartphone app to support junior doctors' mental health and well-being: examination of symptom progression, usability, and acceptability after 1 month of use. J Med Internet Res 2022 Sep 21;24(9):e38497. [doi: 10.2196/38497] [Medline: 36129745]
61. Gbeddy GA. Using mindfulness-based practice to reduce work-related stress and burnout among psychiatric nurses [Dissertation]. : The University of North Carolina at Chapel Hill; 2021. [doi: 10.17615/9znk-jr21]
62. Pace TWW, Zeiders KH, Cook SH, et al. Feasibility, acceptability, and preliminary efficacy of an app-based meditation intervention to decrease firefighter psychological distress and burnout: a one-group pilot study. JMIR Form Res 2022 Jun 8;6(6):e34951. [doi: 10.2196/34951] [Medline: 35675115]
63. Ravalier JM. Co-design, delivery, and evaluation of wellbeing initiatives for NHS staff: the HOW (healthier outcomes at work) NHS project. Int J Environ Res Public Health 2022 Apr 12;19(8):4646. [doi: 10.3390/ijerph19084646] [Medline: 35457512]
64. Ravalier JM, Wainwright E, Smyth N, Clabburn O, Wegrzynek P, Loon M. Co-creating and evaluating an app-based well-being intervention: the HOW (Healthier Outcomes at Work) Social Work Project. Int J Environ Res Public Health 2020 Nov 24;17(23):8730. [doi: 10.3390/ijerph17238730] [Medline: 33255460]
65. Morris ME, Kathawala Q, Leen TK, et al. Mobile therapy: case study evaluations of a cell phone application for emotional self-awareness. J Med Internet Res 2010 Apr 30;12(2):e10. [doi: 10.2196/jmir.1371] [Medline: 20439251]
66. Muuraiskangas S, Harjumaa M, Kaipainen K, Ermes M. Process and effects evaluation of a digital mental health intervention targeted at improving occupational well-being: lessons from an intervention study with failed adoption. JMIR Ment Health 2016 May 11;3(2):e13. [doi: 10.2196/mental.4465] [Medline: 27170553]
67. Deady M, Glozier N, Collins D, et al. The utility of a mental health app in apprentice workers: a pilot study. Front Public Health 2020;8:389. [doi: 10.3389/fpubh.2020.00389] [Medline: 33014953]
68. Xu HG, Tuckett A, Kynoch K, Eley R. A mobile mindfulness intervention for emergency department staff to improve stress and wellbeing: a qualitative study. Int Emerg Nurs 2021 Sep;58:101039. [doi: 10.1016/j.ienj.2021.101039] [Medline: 34333332]
69. de Korte EM, Wiezer N, Janssen JH, Vink P, Kraaij W. Evaluating an mHealth app for health and well-being at work: mixed-method qualitative study. JMIR Mhealth Uhealth 2018 Mar 28;6(3):e72. [doi: 10.2196/mhealth.6335] [Medline: 29592846]

XSL•FO
RenderX

70. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849] [Medline: 27287964]
71. Grier RA, Bangor A, Kortum P, Peres SC. The system usability scale: beyond standard usability testing. Proc Hum Factors Ergon Soc Annu Meet 2013;57:187-191. [doi: 10.1177/1541931213571042]
72. Wilson CJ, Deane FP, Ciarrochi J, Rickwood D. General Help Seeking Questionnaire (GHSQ). APA PsycTests. 2005. URL: https://doi.org/10.1037/t42876-000 [accessed 2026-03-18]
73. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatr Ann 2002 Sep;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
74. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
75. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
76. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176. [doi: 10.1159/000376585] [Medline: 25831962]
77. Tennant R, Hiller L, Fishwick R, et al. The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Nov 27;5:63. [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
78. Phillips EA, Gordeev VS, Schreyögg J. Effectiveness of occupational e-mental health interventions: a systematic review and meta-analysis of randomized controlled trials. Scand J Work Environ Health 2019 Nov 1;45(6):560-576. [doi: 10.5271/sjweh.3839] [Medline: 31184758]
79. Cameron G, Mulvenna M, Ennis E, et al. Effectiveness of digital mental health interventions in the workplace: umbrella review of systematic reviews. JMIR Ment Health 2025 Jan 24;12:e67785. [doi: 10.2196/67785] [Medline: 39854722]
80. Iribarren SJ, Akande TO, Kamp KJ, Barry D, Kader YG, Suelzer E. Effectiveness of mobile apps to promote health and manage disease: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2021 Jan 11;9(1):e21563. [doi: 10.2196/21563] [Medline: 33427672]
81. Bernard RM, Toppo C, Raggi A, et al. Strategies for implementing occupational eMental health interventions: scoping review. J Med Internet Res 2022 Jun 1;24(6):e34479. [doi: 10.2196/34479] [Medline: 35648457]
82. Agarwal S, LeFevre AE, Lee J, et al. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
83. Yardley L, Spring BJ, Riper H, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015]
84. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
85. Zhang R, Nicholas J, Knapp AA, et al. Clinically meaningful use of mental health apps and its effects on depression: mixed methods study. J Med Internet Res 2019 Dec 20;21(12):e15644. [doi: 10.2196/15644] [Medline: 31859682]
86. Tsantila F, Coppens E, De Witte H, et al. Developing a framework for evaluation: a theory of change for complex workplace mental health interventions. BMC Public Health 2023 Jun 17;23(1):1171. [doi: 10.1186/s12889-023-16092-x] [Medline: 37330460]
87. Damschroder LJ, Reardon CM, Opra Widerquist MA, Lowery J. Conceptualizing outcomes for use with the consolidated framework for implementation research (CFIR): the CFIR outcomes addendum. Implement Sci 2022 Jan 22;17(1):7. [doi: 10.1186/s13012-021-01181-5] [Medline: 35065675]
88. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367. [doi: 10.2196/jmir.8775] [Medline: 29092808]
89. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: Improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
mERA: mHealth Evidence Reporting and Assessment
mHealth: mobile health
NASSS: nonadoption, abandonment, scale-up, spread, and sustainability
PICO: population, intervention, comparison, and outcome

XSL•FO
RenderX

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
RCT: randomized controlled trial
WHO: World Health Organization

Edited by L Buis; submitted 03.Feb.2024; peer-reviewed by A Hassan, A Lim, A Jain; revised version received 10.Nov.2025; accepted 02.Feb.2026; published 31.Mar.2026.

Please cite as:
Addanki S, Macedo L, MacDermid J, Moll S
Mental Health Apps Implemented in the Workplace: Scoping Review of Trends and Gaps in Evaluation Research
JMIR Mhealth Uhealth 2026;14:e57046
URL: https://mhealth.jmir.org/2026/1/e57046
doi:10.2196/57046

© Sheila Addanki, Luciana Macedo, Joy MacDermid, Sandra Moll. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 31.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploiting Unsupervised Free-Living Data for Cardiorespiratory Fitness Estimation: Systematic Review and Meta-Analysis

Alexios Dosis[1,2], MSc, MD; Aron Berger Syversen[3], BSc, MSc; Mikolaj R Kowal[1,2], MBChB, PgDIP; Daniel Grant[4], BSc; Jim Tiernan[2], MBBS, PhD; David Wong[5], MEng, DPhil; David G Jayne[1,2], BSc, MBBCHIR, MD

[1]Leeds Institute of Medical Research, University of Leeds, Clinical Sciences Building, St James's University Hospital, Beckett Street, Leeds, United Kingdom

[2]Abdominal Medicine and Surgery, Leeds Teaching Hospitals, Leeds, United Kingdom

[3]School of Computing, University of Leeds, Leeds, England, United Kingdom

[4]Cardiorespiratory Department, Leeds Teaching Hospitals, Leeds, United Kingdom

[5]5 Leeds Institute of Health Sciences, University of Leeds, Leeds, England, United Kingdom

Corresponding Author:
Alexios Dosis, MSc, MD
Leeds Institute of Medical Research, University of Leeds, Clinical Sciences Building, St James's University Hospital, Beckett Street, Leeds, United Kingdom

Abstract

Background: Current methods of cardiorespiratory fitness (CRF) assessment may discriminate against frail individuals who are challenged to perform a maximal cardiopulmonary exercise test. CRF estimations from free-living wearable data, captured over extended time periods, may offer a more representative assessment and increase usability in clinical settings.

Objective: This study aimed to review current evidence behind this novel concept and evaluate the performance and quality of models developed to estimate CRF from free-living, unsupervised data.

Methods: Following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, we systematically searched 4 databases (MEDLINE, Embase, Scopus, and arXiv) for studies reporting the development of models to estimate CRF from continuous free-living wearable data. Studies conducted entirely under controlled laboratory conditions were excluded. Performance metrics were combined in a meta-correlation analysis using a random-effects model and Fisher Z transformation.

Results: Of 1848 papers screened, 18 met the eligibility criteria, with a total of 31,072 participants. The weighted mean age was 46.9 (SD 1.46) years. Multiple computational techniques were used, with 8 studies employing more advanced machine learning models. The meta-correlation analysis revealed a pooled overall estimate of 0.83 with a 95% CI 0.77 - 0.88. The I^2 test indicated high heterogeneity at 97%. Risk of bias assessment found most concerns in the data analysis domain, with studies often lacking clarity around the data handling process.

Conclusions: A promising preliminary agreement between CRF predictions and measured values was noted. However, no definite conclusions can be drawn for clinical implementation due to high heterogeneity among the included studies and lack of external validation. Nonetheless, continuous data streams appear to be a valuable resource that could lead to a step change in how we measure and monitor CRF.

Trial Registration: PROSPERO CRD42024593878; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024593878

(JMIR Mhealth Uhealth 2026;14:e69996) doi:10.2196/69996

KEYWORDS

wearables; cardiorespiratory fitness; free-living data; machine learning; perioperative medicine

Introduction

Cardiorespiratory fitness (CRF) is regarded as a key element of anesthetic preassessment and preoperative decision-making, reflecting an individual's aerobic capacity and ability to withstand and recover from surgery. The most widely recognized measure of CRF is maximal oxygen uptake (VO_2max), a strong indicator of the cardiorespiratory system's ability to capture, transport, and use oxygen during exercise. VO_2max is inversely related to all-cause mortality and linked to several other health outcomes, such as cardiovascular disease (CVD), dementia, and depression [1-3]. In surgery, VO_2max serves as a prognostic marker and is currently the gold standard predictor of early postoperative cardiorespiratory morbidity [4,5].

XSL•FO
RenderX

Cardiopulmonary exercise testing (CPET) is a maximal dynamic test used to assess CRF and global exercise response. CPET is typically performed on a cycle ergometer or a treadmill under conditions of graduated physiological stress, involving computerized gas-exchange analysis of breath-by-breath ventilation. The test is routinely used in cardiorespiratory medicine as a diagnostic tool to distinguish between ventilatory and cardiac exercise intolerance [6]. In the preoperative setting, it is usually used as a risk assessment tool for major surgery to aid clinical decision-making regarding suitability for surgery and to guide perioperative management [7,8].

Despite its proven ability for risk stratification, there remain some drawbacks to this method. VO_2max measurements through maximal exercise can be strenuous and challenging for older or frail adults and those with musculoskeletal conditions who may be limited by pain rather than exertion [9]. Performance-reducing factors, such as peripheral arterial disease, osteoarthritis, and poor effort, have also been associated with inaccurate measurements, which may impact clinical decision-making [10]. In addition, high costs, the requirement for highly trained staff to undertake the test, and reduced hospital availability render regular VO_2max monitoring impractical.

To overcome these limitations, several VO_2max prediction models have been developed: nonexercise models are usually derived from lifestyle and anthropometric data, while submaximal exercise tests rely on prespecified protocols that involve heart rate (HR) monitoring at certain speeds, such as the 20-meter shuttle test or the modified shuttle walking test [9,11,12]. These methods offer an alternative; yet, they are not widely used in routine clinical practice due to some inherent limitations. Submaximal tests rely on the assumption that mechanical efficiency is the same for everyone, often leading to inaccurate VO_2 max estimations, and self-reported physical activity measures are subject to social desirability and recall bias [13]. Equally, lack of protocol standardization has raised concerns about the validity and reliability of submaximal tests [14]. Nonetheless, it seems a great limitation of the above CPET alternatives is their inability to capture and assess unstructured and incidental ambulatory activity accurately [15].

This gap has encouraged the exploration of wearable devices that can collect a substantial amount of information about an individual's activities in daily life, regardless of frequency, duration, or intensity [16]. Wearable technology has experienced a remarkable uptake over the last years, with more users appreciating the potential benefits for health and fitness tracking [17]. Commercially available wearables already offer VO_2max estimations; however, their algorithms are primarily based on short periods of structured exercise data, and their resulting VO_2max estimations have a large degree of error at the individual level [18]. Continuous monitoring of unstructured physical activity, however, shows promise in a variety of settings, enabling constant tracking of physiological data in an unobtrusive manner [16]. Physiological signals captured over longer periods may be more representative of CRF for certain populations. In view of the great potential of these devices, we aimed to explore whether CRF can be accurately predicted leveraging wearable data from unsupervised free-living conditions, outside the controlled laboratory environment. In this paper, we systematically review the research methodology behind the proposed models, the associated challenges and limitations, and discuss the feasibility of applying this concept for CRF estimation in health care settings.

Methods

Search Strategy and Study Selection Process

This study was registered with the international database for systematic reviews PROSPERO (CRD42024593878). We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement and recommendations for systematic reviews [19]. Relevant studies were located from a systematic electronic search of 4 databases (MEDLINE, Embase, Scopus, and arXiv), and the last search was performed on July 27, 2024. The full search strategy and key terms are available in Multimedia Appendix 1.

We used online systematic review software to blind reviewers and screen titles and abstracts after removing duplicates [20]. Conflicts were resolved through direct discussion between the 2 reviewers (AD and ABS) after unblinding. The full papers of potentially eligible studies were scrutinized against the eligibility criteria. Citation chaining of references was also completed by AD. Any additional studies identified were subsequently reviewed and assessed for inclusion by a third investigator (MRK; Textbox 1).

XSL•FO
RenderX

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria:

- All papers reporting the development of a prediction model to estimate maximal oxygen uptake from longitudinal free-living wearable data.
- "Free-living" is defined as data collected in unsupervised, uncontrolled, real-world settings.
- Mixed designs with some simulated activities permitted, provided that at least some unsupervised activity was captured and analyzed.
- Human studies published in English.
- All wearable devices are eligible, including accelerometers, electrocardiogram (ECG) biosensors, commercial smartwatches, and optical heart rate sensors (photoplethysmogram).
- No restriction applied to the clinical setting studied.

Exclusion criteria:

- Studies in which the authors focused solely on physical activity and energy expenditure estimation.
- Monitoring occurring exclusively under controlled laboratory conditions, with no free-living activity studied.
- Wearable data including only exercise activity.
- Studies in which the authors did not report a prediction model but only correlations of wearable metrics with measures of cardiorespiratory fitness.
- Systematic reviews, literature reviews, surveys, conference proceedings, or meeting proceedings.
- Studies focusing on adolescents and young children.

We considered all papers reporting the development of a prediction model to estimate VO_2max from longitudinal free-living wearable data. For this review, "free-living" was defined as data collected in unsupervised, uncontrolled, real-world settings. Mixed designs with some simulated activities were also permitted, provided that at least some unsupervised activity was being captured and analyzed by the authors. A limit was set to human studies published in English. All wearable devices were eligible, including accelerometers, electrocardiogram (ECG) biosensors, commercial smartwatches, and optical HR sensors (photoplethysmogram). No restriction was applied to the clinical setting studied.

Studies were disqualified if (1) the authors focused solely on physical activity and energy expenditure estimation; (2) monitoring occurred exclusively under controlled conditions in a laboratory setting, and no free-living activity was studied; (3) wearable data included only exercise activity; (4) the authors did not report a prediction model, but only correlations of various wearable metrics with measures of CRF; (5) they were systematic or literature reviews, surveys, conference, or meeting proceedings; and (6) studies focusing on adolescents and young children.

Data Extraction and Model Performance Assessment

To ensure consistency, a standardized form was piloted and modified until consensus was reached between 2 authors and the senior investigator for the data extraction tool. Two reviewers (AD and ABS) retrieved all relevant data independently, and a third author (MRK) verified the accuracy of the records, cross-referencing with sources to resolve discrepancies. For each study, the following items were extracted: study details, demographics, setting and sample size, wearable device used for monitoring, and the baseline method used to obtain the ground truth (control). We recorded features derived from wearable data and the preprocessing techniques researchers used for feature extraction. We extracted details on the various machine learning (ML) models that were used, as well as prediction accuracy metrics and the validation process reported.

Quality Assessment

Qualitative appraisal of each included study was independently performed by 2 authors using a modified version of the Prediction model Risk Of Bias Assessment Tool (PROBAST) following the updated TRIPOD-AI (Transparent Reporting of a multivariable or ML prediction model for Individual Prognosis Or Diagnosis–artificial intelligence) guidance [21,22]. In case of disagreements, the opinion of the third author was sought.

Studies received a score of "low," "unclear," or "high" risk of bias on five major domains: (1) predictor choice and definition; (2) participant selection, including source and study setting; (3) outcome measurement; and (4) analysis and methodological quality of the proposed model. Overall judgment was rated as unclear if at least 1 domain was regarded as unclear, and similarly as high if any domain was rated as high. Risk-of-bias plots were created for quality assessment using the "robvis" software package in R (R Foundation for Statistical Computing).

Data Analysis

We calculated and reported descriptive statistics to outline each study characteristic. Summary measures are reported as means or medians, including measures of dispersion such as SDs. Key metrics of model accuracy were identified and combined for quantitative analysis.

Frequently reported metrics included the Pearson correlation coefficient (r) and the R^2 values. Other metrics such as the standard error of estimate (SEE) and the root-mean-square error were also reported but less frequently. Although not technically an accuracy metric, where available, the r coefficient was used to indicate how well the model's predictions aligned with the CPET values.

XSL•FO
RenderX

A meta-analysis of correlation estimates was undertaken to integrate measures of performance across the included studies and provide a more objective and systematic assessment. In this instance, reported correlation coefficients (r) were used as the primary effect size, as they represented the most consistently reported metric in this review. When only R^2 values were available, we converted them by taking the square root to obtain the corresponding r values between the predicted and actual VO_2max values and assessing the direction of association in each study. RStudio (version 2024.04.2+764; R Foundation for Statistical Computing) was used for all statistical analyses [23]. Two packages, "metafor" and "robumeta," were installed to perform the meta-analysis. Fisher Z transformation was applied to convert correlation estimates to a more normally distributed metric and obtain standard effect sizes. A restricted maximum likelihood estimation method was used for a standard random-effects model to conduct the meta-analysis [24]. The random-effects model assigns less study weight to larger studies with less variance [24]. Results of the meta-correlation analysis were presented visually using a forest plot. Subgroup analysis was also performed, comparing regression-based models with more advanced ML methodologies.

We assessed heterogeneity using the I^2 and τ^2 statistics. The I^2 statistic quantifies the proportion of total variation in effect sizes that was due to heterogeneity rather than chance. We considered I^2 values of 25%, 50%, and 75% to represent low, moderate, and high heterogeneity, respectively [25]. The τ^2 represents another method to assess the between-study variance, focusing on the absolute variability of true effect size, with higher values indicating greater heterogeneity [26]. The Egger test was used to assess the likelihood of publication bias, which was also presented visually with a funnel plot.

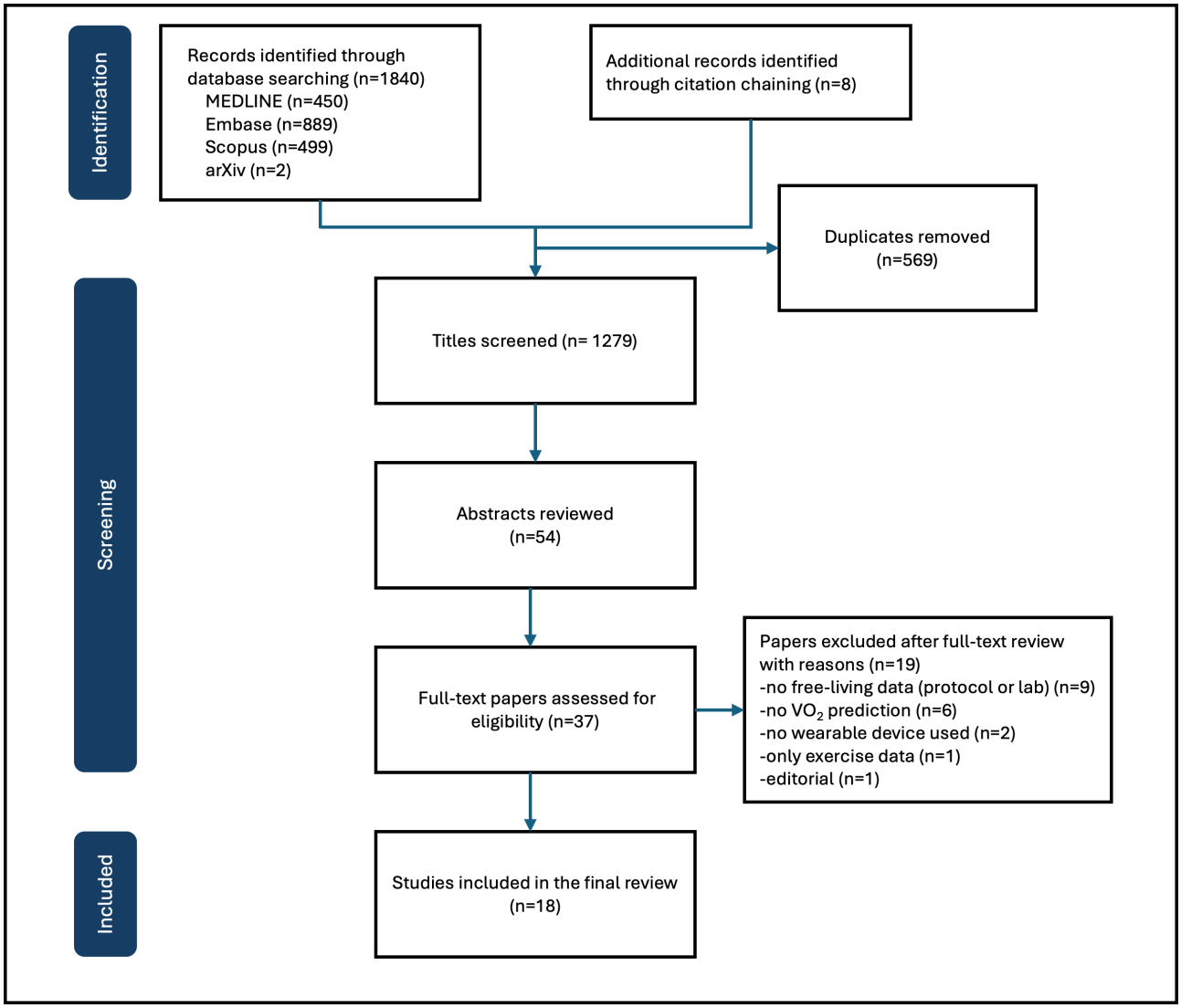
Results

Overview

The combined literature search generated 1848 papers, with 1279 remaining after deduplication. The PRISMA flowchart (Figure 1) shows the paper selection process. Following title and abstract screening, 37 papers qualified for full-text review. Eighteen studies were accepted in the final set with a total sample size of 31,072 participants. Sample sizes varied greatly across studies and ranged from 13 to 12,425. Only 4 studies had a sample size of more than 1000 patients, indicating that most research in this field is based on a relatively small number of participants [27-30].

XSL•FO
RenderX

Figure 1. Flowchart of study selection following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidance. VO_2: maximal oxygen uptake.



The characteristics of each included study are shown in Table 1 [2,4,5,7,11,12,16,27-37]. Participant-level weighted mean age was 46.9 (SD 1.46) years, with a male participant distribution of 48.9%. Notably, a few studies had exclusively male or female participants, while others had more balanced samples. Most studies included data from volunteers recruited in a prospective manner (n=13, 72%). Smaller studies focused primarily on healthy participants (n=13), in contrast to the larger cohorts (n=5) that used data from population studies involving hundreds of patients (the Fenland and Framingham studies) [38,39]. Overall, there were only 2 research trials that targeted patients scheduled for preoperative assessment [4,7].

XSL•FO
RenderX

Table . Summary of study characteristics.

Authors	Year	Sample size (M[a]), n (%)	Age (years), mean (range)	Sensors and modality used	Wearable monitoring (days)	Participants	Control	Reference VO_2, mean (SD)
Plasqui and Westerterp [2]	2005	25 (40)	28 (18-50)	Tracmor elastic belt triaxial accelerometer and Polar (S610i) HR[b] monitor wristwatch	7 (daytime only)	Healthy volunteers	Maximal GXT[c] cycle	M=49.5±10.2; F[d]=40.7±8.4
Plasqui and Westerterp [33]	2006	26 (53.8)	29 (18-50)	Tracmor elastic belt triaxial accelerometer and Polar (S610i) HR monitor wristwatch	7 (daytime only)	Healthy volunteers	Maximal GXT cycle	44.6±10.5
Cao [34]	2009	189 (0)	49.6 (20-69)	Kenz Lifecorder uniaxial accelerometer	7 (daytime only)	Healthy volunteers	Maximal GXT cycle	31.4±7.4
Cao et al [35]	2010	148 (0)	47 (20-69)	Kenz Lifecorder uniaxial accelerometer and triaxial accelerometer	7 (daytime only)	Healthy volunteers	Maximal GXT cycle	30.8±5.9
Novoa et al [4]	2011	38 (79)	62.8 (38-80)	OMROM Walking Style ProW accelerometer (two modes-aerobic mode after 10 min of walking at 60 steps per min)	7-41 (daytime only)	Patients scheduled for lung resection	Maximal GXT cycle and Arterial gas	20.3±4.6
Altini et al [11]	2016	46 (45)	24.7 (NR[e])	ECG[h] Necklace (one lead ECG) and ADXL330 triaxial accelerometer and Mobile phone for GPS coordinates	14 (during laboratory protocols and free-living)	NR	Maximal GXT cycle	44±9.8
Altini et al [5]	2016	51 (47)	NR[e]	ECG Necklace (one lead ECG) and ADXL330 triaxial accelerometer	14 (during laboratory protocols and free-living)	Healthy volunteers	Maximal GXT cycle	NR[e]
Beltrame et al [37]	2017	13 (100)	26 (NR)	Hexoskin smartshirt (hip accelerometer, three lead ECG and respiration bands)	4 (free-living 9 AM to 5 PM) and simulated ADLs[f]	Healthy volunteers	Simulated ADLs and Pseudorandom Ternary sequence.	NR

XSL•FO
RenderX

Authors	Year	Sample size (M[a]), n (%)	Age (years), mean (range)	Sensors and modality used	Wearable monitoring (days)	Participants	Control	Reference VO_2, mean (SD)
Ahn et al [31]	2017	24 (100)	27.5 (NR)	Shimmer ECG sensor (18), for measurement of 2-lead ECG and triaxial accelerometer	4 (daytime only)	Healthy volunteers	Maximal GXT treadmill	48.5±5.3
Kwon et al [12]	2019	240 (47)	42 (20-65)	Fitbit (Fitbit Charge; Fitbit) (triaxial accelerometer, PPG)[g]	3 (daytime only)	Healthy volunteers	Maximal GXT treadmill	36.25
Bonomi et al [16]	2020	40 (48)	25 (18-55)	Chest-belt HR ECG monitor (RS800CX, Polar, Wrist activity monitor (Tracmor)	5 (daytime) and simulated ADLs	Healthy volunteers	Maximal GXT cycle	M=45.7±6.1; F=40±6.6
Jones et al [7]	2021	49 (65)	65 (NR)	Garmin Vivosmart HR+ activity tracker wristwatch (PPG, accelerometer, GPS)	7 continuous	Patients for preoperative assessment	Maximal GXT cycle	18.2±4.5
Spathis et al [30]	2021	2100 (46)	48.7 (35-65)	Combined HR and uniaxial movement sensor (Actiheart) and wrist triaxial accelerometer	6 continuous	Population-based cohort study (Fenland)	Submaximal GXT treadmill	NR
Wu et al [28]	2022	12,425 (47) 181	47.7 (35-65)	Combined HR and uniaxial movement sensor (Actiheart) and wrist triaxial accelerometer	6 continuous	Fenland study population-based cohort and UK Biobank	Submaximal GXT treadmill Maximal GXT test	NR; 32.95
Spathis et al [27]	2022	11,059 (47)	47.7 (35-65)	Combined HR and uniaxial movement sensor (Actiheart) and wrist triaxial accelerometer	6 continuous	Fenland study population-based cohort	Submaximal GXT treadmill	M=41.95±4.61; F=37.44±4.73
Frade et al [32]	2022	43 (74.4)	37.5 (19-72)	Hexoskin smartshirt (hip accelerometer, three lead ECG and respiration bands)	7 (daytime only)	Volunteers (chronic disease allowed)	Maximal GXT cycle	32.09
Neshitov et al [29]	2023	3894 (67)	42 (20-65)	Apple Watch and Garmin watch (PPG, triaxial accelerometer, GPS)	mean 287, SD 149	Healthy volunteers -consented to Welltory app	Estimated by smartwatch device	36.16±6.66

XSL•FO
RenderX

Authors	Year	Sample size (M[a]), n (%)	Age (years), mean (range)	Sensors and modality used	Wearable monitoring (days)	Participants	Control	Reference VO_2, mean (SD)
Zhang et al [36]	2024	662 (41)	53 (NR)	Apple Watch (PPG, triaxial accelerometer, GPS)	mean 128 (daytime only)	Framingham study cohort	Maximal GXT cycle	M 27±7; F 22±6

[a]M: male.

[b]HR: heart rate.

[c]GXT: graded exercise test.

[d]F: female.

[e]NR: not reported.

[f]ADL: activities of daily living.

[g]PPG: photoplethysmogram.

[h]ECG: electrocardiogram.

A wide variety of wearable devices were used, including triaxial accelerometers, ECG sensors, and smartwatches with optical photoplethysmogram sensors. All studies used accelerometers for motion tracking, while HR monitoring was performed either with ECG (n=9) or optical sensors (n=6). The monitoring durations also differed, with most studies tracking participants over a few days, usually 3 - 7, while others extended up to nearly a year [29]. Measurement of VO_2max as the ground truth was typically obtained through a maximal CPET test, though in 4 trials, a submaximal treadmill test was used. In 1 study, reference VO_2max was not directly measured but estimated using the proprietary algorithm of a smartwatch device [29]. Due to disparities in participant populations, the reference mean VO_2max varied significantly between patient-centered studies (weighted mean VO_2max 19.2 mL/kg/min, SD 1.48) and healthy volunteer studies (40.2 mL/kg/min, SD 6.58; z test, P=.03).

Features

Feature extraction is a crucial first step in signal processing, transforming complex raw data points into meaningful numerical features that can be interpreted and processed in a model [40]. In high-volume continuous wearable data, feature extraction can also help to reduce dimensionality without losing important information. This process makes data handling easier and speeds computation by focusing only on the most relevant aspects of the data [41]. We observed, however, that in 6 studies, researchers adopted features as reported from the internal proprietary algorithm of the manufacturer without further analysis of the raw data. Table 2 summarizes the analytical methods used in the included studies, while Table 3 provides an overview of each feature type with examples in the studies used.

XSL•FO
RenderX

Table . Summary of analytical methods, models, and wearable features used in the included studies.

Study, year	Wearable features	Model used	Preprocessing techniques and analysis	Model performance	Validation	Principal finding
Plasqui and Westerterp [2], 2005	[a]HR/ACM[b] index (ACM: activity counts per minute)	Multiple linear regression	Minute averages for HR and activity counts averaged over the 7 days, ratio of HR/ACM, and missing data removed	SEE[c]=12.4%; r=0.87	NR[d]	Fitness index HR/ACM significantly related to VO_2max[e] corrected for body composition and age
Plasqui and Westerterp [33], 2006	HR/ACM index	Multiple linear regression 2nd equation tested	Same as above and groups combined and sorted for activity counts	r=0.86; SEE=10.7%; Bland-Altman systematic error=5.6%	Cross-validation	A second equation for the fitness index HR/ACM had to be tested to predict VO_2[f]
Cao et al [34], 2009	Daily SC[g]	Hierarchical linear regression	Steps per day provided and handling of missing data not reported	SEE=10.9%; r=0.81	Split test	SC was a significant contributor to the prediction of the measured VO_2max
Cao et al [35], 2010	MVPA[h], VPA[i], and SC	Hierarchical linear regression	As provided (minutes spent in MVPA and VPA) and handling of missing data not reported.	SEE=9.66%; r=0.863	Cross-validation, Subgroup analysis	VPA significantly increased the explained variance in VO_2max, adjusted for age
Novoa et al [4], 2011	Daily SC, Aerobic SC, time spent in aerobic activity, and daily distance in km	Linear regression	As provided and handling of missing data not reported	R^2=0.93	Bootstrapping (1000 iterations)	VO_2max can be significantly predicted by the mean daily walked distance
Altini et al [11], 2016	HR at different walking speeds, stay regions, and activity composites (ie, HR at relative time spent in each activity)	Nonnested hierarchical Bayesian regression, SVM[j] classifier, and HMM[k] for transitions between activities	LDA[l] activity is classified into 6 clusters, accelerometer data was bandpassed between 0.1 and 10 Hz to isolate dynamic components, HR was extracted from R-R intervals and averaged over 15 seconds, and missing data not analyzed	R^2=0.76; RMSE[m]=249.4; SEE=5.79%	Leave-one-participant-out cross-validation	Contextualizing HR by means of activity and speed improved correlation between free-living HR and CRF[n]
Altini et al [5], 1985	HR/min while lying down and while walking at 3.5 and 5.5 km/h	Multiple Linear regression, SVM classifier	HR was extracted from R-R intervals and averaged over 15 seconds windows and unusable ECG[o] was discarded. Acceleration signal was segmented in nonoverlapping intervals of 5s	R^2=0.78; RMSE=284.7	Leave-one-participant-out cross-validation	Submaximal context-specific HR can be used to estimate VO_2max

XSL•FO
RenderX

Study, year	Wearable features	Model used	Preprocessing techniques and analysis	Model performance	Validation	Principal finding
Beltrame et al [37], 2018	Means of HR, VE[p], BF[q], hip acceleration, and SC in the 2 conditions ("active-inactive")	Random forest	HR was averaged every 16 beats, VE and BF average of the last 7 respiration cycles, all features were time-aligned, low-pass filtered at 0.01 Hz. Fast Fourier transformation and frequency domain analysis for hip acceleration, and when hip acceleration was >0.05 g, data were labeled as "active"; otherwise, they were "inactive	r=0.88	Leave-one-participant-out cross-validation	Predicted oxygen uptake data during ADLs[r] were strongly correlated with the temporal characteristics of the VO_2 during a controlled protocol
Ahn et al [31], 2017	aEE[s] (nonlinear model derived from ACM horizontal and ACM vertical signals and HR per minute	Linear regression between HR and aEE	The tri-axial acceleration was band-pass filtered (0.25 to 7 Hz). ECG data were band-pass filtered (5 to 20 Hz), the R-R intervals were averaged for 1 min and converted to a HR, and used only increasing HR periods and excluded data with inaccurate ECG	R^2=0.74; r=0.87; SEE=11.85%	Split test	aEE can be used to estimate VO_2max during daily activities
Kwon et al [12], 2019	HR and daily PA in terms of METs, used slope of HR and PA	Linear regression	Moving average filter was applied, data points at which both HR and physical activity data increased were selected	R^2=0.651; SEE=3.518; 9.6%	PRESS[t] for cross-validation	VO_2max can be estimated using novel features, aEE and the slope between physical activity and HR
Bonomi et al [16], 2020	Acceleration and HR- fitness index named TEE-pulse[u]	Stepwise linear regression	Motion intensity was defined as activity counts per minute, acceleration signal was processed in overlapping windows of 60 seconds, and activity is grouped in (sedentary, or other) based on a set of counts thresholds	RMSE=367 or 12.4%; SEE=13.09%; r=0.89; MAE=10.2%	Leave-one-participant-out cross-validation	The daily average TEE-pulse was highly correlated to the mean TEE-pulse measured in the laboratory without the need for specific exercise protocol
Jones et al [7], 2021	Floors climbed, total number of steps and total distance, average HR and resting HR	Linear regression	Features were used as provided by the device and averaged across the 7-day wear period. Self-reported METs[v] from questionnaires	AIC[w]=181.62; R^2=0.74; r=0.86; AUC[x]=0.93	NR	Using all the wearable variables together in linear regression gave a stronger correlation between the measured CPET[y] values, specifically for peak VO_2

XSL•FO
RenderX

Study, year	Wearable features	Model used	Preprocessing techniques and analysis	Model performance	Validation	Principal finding
Spathis et al [30], 2021	HR per minute, acceleration (magnitude calculated through ENMO[z]), resting HR	Step2Heart Deep neural network, (CNN[aa] learn spatial and RNN[ab] temporal features)	Noisy heart data removed with a Gaussian process robust regression, participants with less than 72 hours of wear were removed, and accelerometry and ECG signals were summarized to a common time resolution of one observation per 15 seconds	AUC=0.70; RMSE=9.54 (HR forecasting only)	Split test	A general-purpose self-supervised feature extractor for wearable data was developed. HR forecasting transfer learning of learned physiological representations
Wu et al [28], 2022	HR and movement 26 features combining HR, movement data, and time-series metadata	UDAMA[ac]	Nonwear periods were removed (periods of nonphysical HR and no movement), downsampled the sampling rate to 15 minutes and used the first 600 timesteps, and pretrained on noisy data and used adversarial training on the BBVS dataset	R^2=0.392; r=0.665; MSE[ad]=30.79; MAE[ae]=4.44	Split test and 3-fold cross validation	A novel model proposal to leverage noisy data from source domain (wearable dataset) to improve modeling for accurate fitness estimation at scale
Spathis et al [27], 2022	48 features: Raw acceleration derived through ENMO, HR, HRV[af], MVPA for each feature mean, minimum, maximum, SD, and the slope of a linear regression fit	Deep neural network- adaptive representation learning	Nonwear periods removed, movement intensities were converted into standard METs, principal component analysis for noise reduction, tSNE[ag], a nonlinear dimension-reduction technique was applied	R^2=0.658; RMSE=2.956; r=0.82; RMSE=8.998 (Biobank only)	Split test and External validation in UK biobank cohort (181 patients)- Maximal GXT[ah] testing	A deep learning framework for predicting CRF was developed, combining learned features from HR and accelerometer free-living data without context awareness
Frade et al [32], 2023	Mean HR and BF, minute ventilation, tidal volume, mean hip acceleration, and mean SC	SVM (support vector regression formulation)	Abnormal HR and BR[ai] were excluded with a preprocessing algorithm (not mentioned), and all raw data were averaged	r=0.804; MAE=3.84	k-fold cross-validation	Hemodynamic domain presented statistically higher importance to predict the VO_2max compared with activity and Pulmonary domains

XSL•FO
RenderX

Study, year	Wearable features	Model used	Preprocessing techniques and analysis	Model performance	Validation	Principal finding
Neshitov et al [29], 2023	HR and SC/min, HR/cadence ratio, daily MET, and HR response to cadence increase	Quantile regression (for each quantile a gradient boosting model was trained)	The HR stream was resampled to 1 measurement per minute and averaged over consecutive 1-minute intervals, cadence is the number of steps made during the same 1-minute interval, continuous ranked probability score used for hyperparameter tuning, and model trained on estimated VO_2 from wearable device	Test set: ECE[aj]=0.032; IQR=3.948; MedPE[ak]=0.01; and Direct VO_2 dataset: ECE=0.084; IQR=4.705; MedPE=0.35	Split test for training External validation (10 healthy volunteers) Maximal GXT treadmill	Anthropometric characteristics were the most influential feature, followed by cadence to HR ratio. The proposed model provides a point estimation and a probabilistic prediction of VO_2; to estimate the prediction's uncertainty

XSL•FO
RenderX

Study, year	Wearable features	Model used	Preprocessing techniques and analysis	Model performance	Validation	Principal finding
Zhang et al [36], 2024	Daily SC, mean HR	Multivariable linear regression, sensitivity analysis for age, BMI, gender	Defined HR as nonactive if recording interval was >1 minute, hourly steps <30, inferred motion context status for HR measures that lacked motion, and excluded days with <5 hours of wearing time	R^2=0.07 - 0.12	NR	Every 1.3 mL/kg/min higher peak VO_2 corresponded to a 2.4-bpm lower nonactive HR. Physical activity with 1.3 mL/kg/min higher peak VO_2 was associated with nearly 1000 more daily steps

[a]HR: heart rate.
[b]ACM: activity counts per minute.
[c]SEE: standard error of estimate.
[d]NR: not reported.
[e]VO_2max: maximal oxygen uptake.
[f]VO_2: measured oxygen uptake.
[g]SC: step count.
[h]MVPA: moderate to vigorous physical activity.
[i]VPA: vigorous physical activity.
[j]SVM: support vector machine.
[k]HMM: hidden Markov model.
[l]LDA: latent Dirichlet allocation.
[m]RMSE: root-mean-square error.
[n]CRF: cardiorespiratory fitness.
[o]ECG: electrocardiogram.
[p]VE: minute ventilation.
[q]BF: breathing frequency.
[r]ADL: activities of daily living.
[s]aEE: activity energy expenditure.
[t]PRESS: predicted residual error sum of squares.
[u]TEE: total energy expenditure.
[v]MET: metabolic equivalent task.
[w]AIC: Akaike Information Criteria.
[x]AUC: area under the receiver operating characteristic curve.
[y]CPET: cardiopulmonary exercise testing.
[z]ENMO: Euclidean norm minus one.
[aa]CNN: convolutional neural network.
[ab]RNN: recurrent neural network.
[ac]UDAMA: unsupervised domain adaptation via multidiscriminator adversarial training framework.
[ad]MSE: mean squared error.
[ae]MAE: mean absolute error.
[af]HRV: heart rate variability.
[ag]tSNE: t-distributed stochastic neighbor embedding.
[ah]GXT: graded exercise test.
[ai]BR: breathing rate.
[aj]ECE: expected calibration error.
[ak]MedPE: median prediction error.

XSL•FO
RenderX

Table . Wearable features grouped by type, with examples used in each of the included studies.

Feature type	Studies	Examples
Motion	• Cao et al [34] and Cao et al [35] • Novoa et al [4] • Beltrame et al [37] • Jones et al [7] • Spathis et al [27] and Spathis et al [30] • Frade et al [32] • Zhang et al [36]	• Daily step count • Moderate-to-vigorous physical activity • Vigorous physical activity daily distance, mean hip acceleration, and acceleration magnitude
Cardiac	• Altini et al [11] • Beltrame et al [37] • Ahn et al [31] • Jones et al [7] • Spathis et al [30] • Frade et al [32] • Zhang et al [36]	• Average HR[a] and resting HR • HR/min • Heart rate variability measures
Contextualized HR	• Plasqui and Westerterp [2] and Plasqui and Westerterp [33] • Altini et al [5] and Altini et al [11] • Bonomi et al [16] • Kwon et al [12] • Wu et al [28] • Spathis et al [27]	• HR/ACM[b] ratio index • HR response to cadence increase activity composites (HR at relative time spent in an activity) • HR at different walking speeds • slope of HR and physical activity • HR/cadence ratio
Other	• Beltrame et al [37] • Ahn et al [31] • Kwon et al [12] • Frade et al [32] • Neshitov et al [29]	• V_E[c], BF[d], and tidal volume • aEE[e] • TEE[f] • MET[g]

[a]HR: heart rate.

[b]ACM: activity counts per minute.

[c]VE: minute ventilation.

[d]BF: breathing frequency.

[e]aEE: activity energy expenditure.

[f]TEE: total energy expenditure.

[g]MET: metabolic equivalent.

Motion Features

Activity features from the accelerometer data included mainly daily step count (SC), distance covered, time spent in anaerobic or sedentary activity (stay regions), and acceleration or walking speed (Table 3). From these, the most frequently reported feature was the daily SC, which in most instances was precalculated from the accelerometer's proprietary algorithm. We found that some researchers reported steps as an average across several days, removing the temporal context which might be useful in analyzing trends [2,4]. Intensity of movement and walking speed was described as time spent in sedentary or vigorous activity, which can be calculated from the acceleration as activity counts per minute [16]. One study used simulated daily activities to correlate aerobic dynamics of variable intensity [37]. Finally, distance covered was either extracted directly from the device or computed as the total number of steps taken in a day multiplied by the stride length of the participant [4].

Cardiac Features

Features extracted from the ECG and optical sensors included average HR, resting HR, HR per minute, mean HR, and ΔHR (difference between current and previous HR value to capture the magnitude of changes in cardiac activity). ECG signals are generally complex; they are subject to motion artifact, creating noise that affects quality [42]. Various filtering methods were applied to reduce noise. Researchers commonly use a band-pass filter between 5 and 10 Hz to remove artifacts and enhance the detection of the heartbeat from the R-R intervals. The R-R intervals were usually averaged over set-time windows, discarding any inaccurate values [11,31]. Beltrame et al [37] averaged HR every 16 beats, passing HR data through a low-pass filter at 0.01 Hz, removing high frequencies.

Contextualized HR

The combination of HR and activity data, often referred to as contextualized HR, was reported in more recent studies. This contextual dimension of wearable signals can be vital in understanding the physiological cardiac response to exercise. Such features comprised HR at variable-intensity walking speeds, HR and SC per minute, HR/cadence ratio, and HR response to cadence increase (Table 2) [5,27,29]. Kwon et al [12] examined the slope between the concurrent increase in physical activity and HR, while others studied time-series

XSL•FO
RenderX

metadata of HR and movement signals, mining numerous features [27,28]. Advanced signal processing methods, such as principal component analysis and fast Fourier transform, were implemented to tackle noisy heart data, but this was not standardized across the included studies. Authors frequently used resampling techniques (standardizing time intervals between data points) to align HR with movement data, helping to contextualize HR within the corresponding physical activity [30].

Other Features

Less common features included energy expenditure estimates in terms of metabolic equivalents, which were calculated in 3 studies based on daily physical activity and proprietary algorithms [12,16,29]. On 1 occasion, breathing frequency and minute ventilation were extracted as the average of the last 7 respiration cycles, based on respiration bands integrated into the wearable device used for monitoring [37].

Models

Multiple modeling techniques were used among the included studies, with more advanced ML models such as support vector machines (SVMs) and deep learning gaining interest over regression in recent studies. This trend follows the usage of preprocessing techniques such as fast Fourier transform and frequency domain analysis and represents an effort to mine the raw data and uncover hidden patterns. A detailed breakdown of modeling and preprocessing techniques is provided in Table 2.

Eleven studies used linear models, which allow interpretation of outcomes through reporting of coefficients that give direct insights into how each predictor influences the outcome measured [2,4,7,11,12,16,31,33-36]. The earliest study that examined whether the ratio of HR to activity counts could predict VO_2max was from Plasqui and Westerterp [2]. Two studies used correlation analysis to identify the strongest predictors of VO_2max and built on this using linear regression [4,7]. The rest of the studies outlined combinations of modern ML techniques. Three studies reported SVMs, with Frade et al [32] presenting support vector regression, which is an SVM formulation for regression problems. SVMs are supervised models that identify an optimal decision boundary (hyperplane) to separate data points into distinct classes with the aim of maximizing the margin between observed and predicted values.

Beltrame et al [37] were the only group to consider a random forest model. Random forests aggregate several decision trees together as a group but introduce randomness to prevent overfitting [37]. Another study trained several gradient boosting models and then fitted a quantile regression algorithm to predict the distribution of VO_2max with CIs [29]. In boosting, models are trained sequentially, building on the errors or residuals of the previous model to improve their prediction accuracy.

Finally, 3 studies [27,28,30] leveraged large-scale free-living datasets to predict CRF with variations of neural networks and deep learning. Wu et al [28] introduced a novel 2-stage approach building an adversarial training framework based on unsupervised domain adaptation. The proposed model was pretrained with noisy health-related labels in a fully supervised setting to improve its performance on high-quality, gold-standard data. Coarse- and fine-grained discriminators were used to better handle the distribution shifts between source (silver-standard) and target (gold-standard) datasets. Spathis et al [27] applied principal component analysis to denoise the raw data and developed deep neural network models able to capture nonlinear relationships between numerous wearable features and VO_2 max.

Length of Available Data Required for Prediction

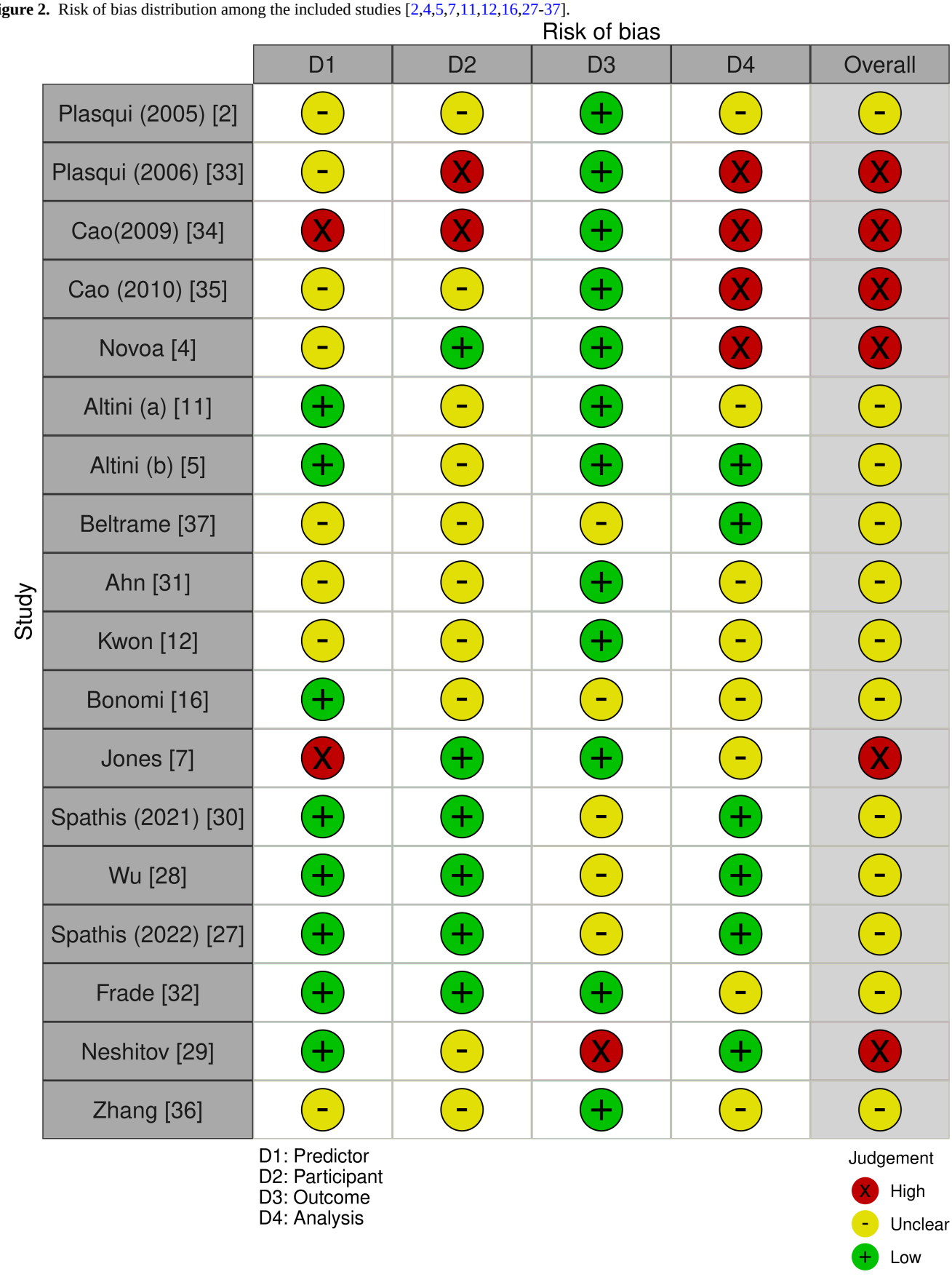
Some studies examined the minimum length of free-living wearable data that would be required to reach reliable conclusions regarding VO_2max estimations, but no agreement was observed. The most thorough assessment was provided by Neshitov et al [29], who tested the degree of certainty for 5 different models using various amounts of available data. They advocated that a minimum of 200 minutes is required for an error estimation range of 4.5 mL/kg/min, but more than 1000 minutes is needed to improve this to under 4 mL/kg/min. Ahn et al [31] plotted the correlation coefficient values with the included measurement time and found no drastic improvements between VO_2 max and estimated values past the 900-minute mark, which yielded an r value of 0.81. On the contrary, other researchers reported that even 10 minutes per day of good-quality data might be sufficient to predict VO_2max. Beltrame et al [37] determined the 10-minute window from the frequency domain analysis as the ideal size for data extraction based on iterative testing of different window lengths (200-1000 seconds, incrementing by 100 seconds). A window length of 600 seconds (10 minutes) was found to provide the best balance between maximizing frequency resolution and ensuring enough reliable samples for analysis across participants. Altini [5,11] proposed this on a theoretical basis, as many submaximal protocols are of a 10-minute duration (eg, 6-minute walking test). Other studies did not account for a minimum length but excluded patients with <72 hours of data from the analysis [27]. Ultimately, this large disparity in the length of data required for feature engineering reflects the exploratory nature of some of the included studies.

Quality of Studies

Risk of bias distribution for each domain is provided in Figure 2 [2,4,5,7,11,12,16,27-37]. Critical appraisal of the included papers showed that only 1 study was classified as “low risk of bias” for all domains. CRF was aptly measured as the outcome of interest in 12 (67%) studies using data from maximal exercise tests (also provided in Multimedia Appendix 2). Appropriate selection of participants and predictors was documented in 6 and 8 studies, respectively. Higher degrees of bias were observed in the analysis domain with robust reporting of analytical methods noted only in 6 (33%) studies. Handling of missing data was not reported adequately in 5 studies [4,5,11,34,35], while others excluded data from analysis arbitrarily, without fully justifying their decisions [12].

XSL•FO
RenderX

Figure 2. Risk of bias distribution among the included studies [2,4,5,7,11,12,16,27-37].



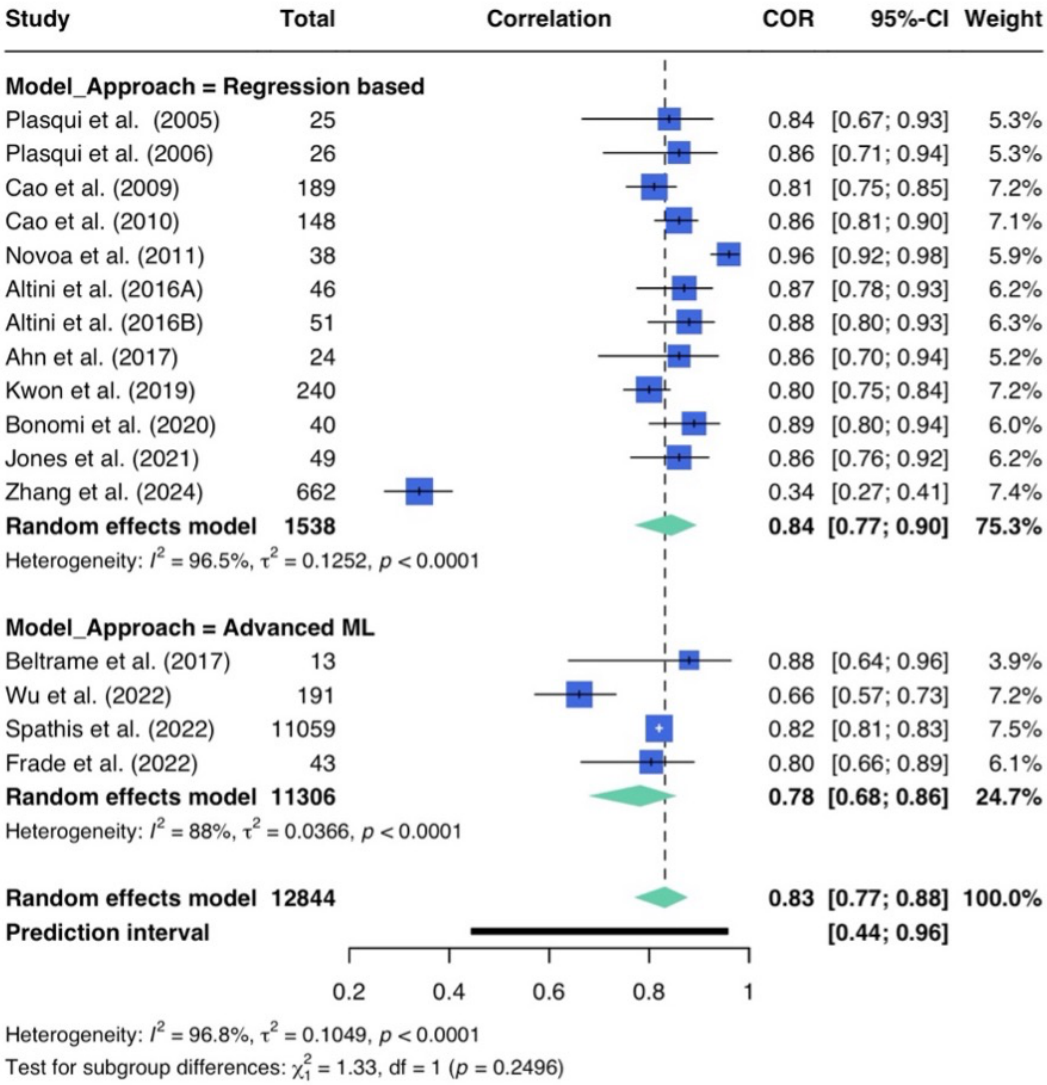
Most studies (n=13) reported internal validation methods for their predictive model, such as split-test or leave-one-participant-out cross-validation (Table 2). Model validation was not considered in 3 studies, and only 2 papers tested their algorithm externally on unseen data [27,28].

XSL•FO
RenderX

Model Performance

Various model performance metrics were reported (Table 2). The weighted average SEE was 9.03%, indicating that models overall predict VO_2max with an error of approximately 9%. In total, 16 studies were included in the meta-correlation analysis, which is provided in Figure 3 [2,4,5,7,11,12,16,27,28,31-37]. The pooled overall estimate of r=0.83 with a 95% CI of 0.77-0.88 from the random-effects model indicates a positive agreement between predicted and observed VO_2 max values. Heterogeneity among the included studies was high, with I^2=97% and a Q test of statistical significance (P<.01). Furthermore, moderate variance was observed (τ^2=0.1049), suggesting underlying differences in how well VO_2 max is predicted across studies. Subgroup analysis comparing regression-based methods with more advanced ML methodologies favored regression, but the difference was not statistically significant (P<.24; Figure 3 [2,4,5,7,11,12,16,27,28,31-37]).

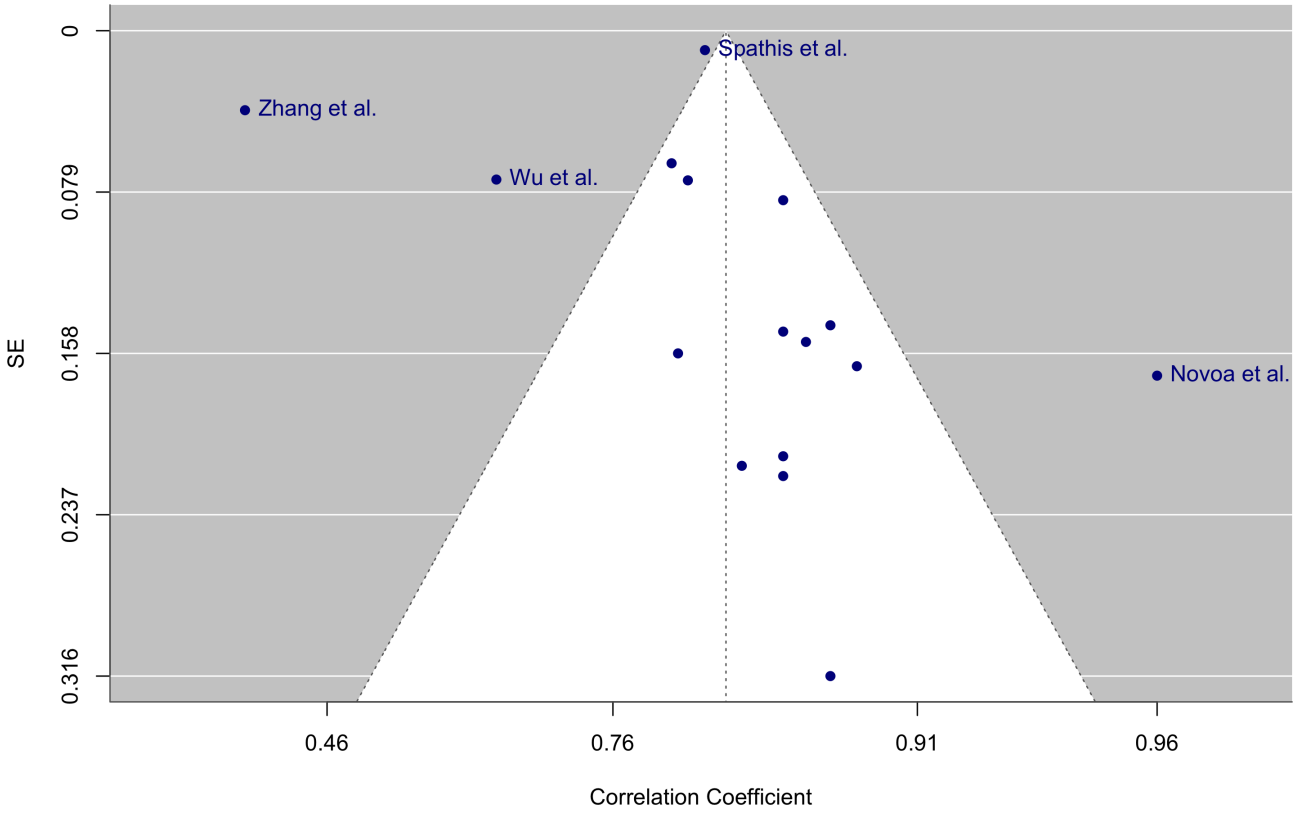
Figure 3. Forest plot of the meta-correlation analysis between maximal oxygen uptake estimates and reference values. Random-effects study weighting was calculated as the inverse sum of the in-study variance and the between-study variance (τ^2). Subgroup analysis comparing modeling approaches is also presented [2,4,5,7,11,12,16,27,28,31-37].



In studies reporting high correlations, there are several common features. While most research indicated the use of validation methods with unseen data (train-test split and cross-validation) to test model performance, several studies reported the R^2 values from the linear regression model [4,7]. Another common factor among the highest-performing models was the incorporation of data collected from laboratory protocols, either to contextualize or interpret free-living data, into feature extraction and modeling processes [5,11,16]. In addition, the funnel plot provided in Figure 4 [4,27,28,36] revealed asymmetry with 4 outliers, which was also confirmed with an Egger test (z=2.29; P=.02). This suggests the presence of small-study effects and publication bias. Notably, the overoptimistic, smaller-sized study by Novoa et al [4] is likely overrepresented in the pooled estimate, and results need to be interpreted with caution.

XSL•FO
RenderX

Figure 4. A funnel plot illustrating the distribution of study effect sizes to assess potential publication bias. Four studies fall outside the expected range, suggesting potential publication bias [4,27,28,36].



Discussion

Principal Findings

This systematic review identified research using real-world, unsupervised wearable data to develop predictive models for CRF estimation, focusing on VO_2 max as the measure of interest. Our study adds to the literature as the first to appraise evidence in this field and showcase the ability of advanced ML algorithms to harness the power of unstructured physical activity outside controlled laboratory settings. The included meta-correlation analysis revealed a pooled overall estimate of 0.83 with a 95% CI of 0.77-0.88, and a mean SEE of 9.06%, demonstrating a promising overall agreement between predicted VO_2max and ground truth. Authors experimented with a range of sensor modalities and various population groups, but models were predominantly designed based on small-sized, healthy volunteer data. Several features were extracted from the free-living information, including SC and distance covered, resting and mean HR, and cardiac response to cadence increase. Quality control of the eligible studies showed that authors were consistent in predictor and outcome reporting, but analytical methods were often ambiguous and included some arbitrary decisions regarding data manipulation.

Advantages

The concept of CRF estimation based on free-living activity holds considerable potential, and results from this study suggest that this could be a pragmatic alternative to CPET. Leveraging longitudinal wearable data can aid preoperative risk assessment for frail patients or those with musculoskeletal conditions that underperform during CPET (indicated usually by a respiratory exchange ratio of <1.10) [8]. Researchers have argued that physiological signals captured over longer time periods may even be more representative of cardiac health in these populations [5,37] compared to a snapshot laboratory measurement. In addition, at-home monitoring offers a convenient and unintrusive assessment without the need for specific protocols, improving patient experience and reducing the psychological stress related to the hospital environment [43]. Considering decreasing costs and increasing accessibility [44], continuous monitoring could represent a complementary, more cost-effective, efficient method that can be scaled to accommodate all patients [7]. Serial measurements of VO_2max can not only help patients track progress and meet targets set during prehabilitation and rehabilitation programs but also guide clinician decision-making [45].

CRF is a well-established marker of CVD and all-cause mortality [3]. Considering that low levels of CRF may precede the clinical detection of CVD, early recognition and intervention are of patient benefit [37]. Wearable-driven evaluation of the aerobic response during unsupervised activities of daily living holds prognostic value in tracking changes in fitness over time, as demonstrated by Spathis et al [27]. Arguably, models that predict future CRF levels could help identify early-stage CVD before general symptom manifestation [37]. Finally, scrutinizing free-living data with advanced ML presents a rare opportunity to study patient behavior and activity habits, shedding light on individual CRF levels [37]. But above all, tailored interventions can be implemented promptly to improve patient fitness and outcomes [46].

XSL•FO
RenderX

Contextualizing HR

The advent of wearables is undoubtedly transforming the landscape of health monitoring, providing clinical teams with a substantial amount of user-generated time-series data [27]. Although some earlier published studies used aggregates over several days as features (average steps or HR data), potentially losing information on trends and variability [2,33-36], most researchers in this review worked on extracting features from the raw signals, with seven studies aligning cardiac and activity points to contextualize HR and gain insights into the participants' physiological response to workload. Based on this principle, Plasqui and Westerterp [2] were the first to publish a fitness index as a ratio between acceleration and HR. Studies of more advanced modeling included other multimodal features such as HR at certain speeds or activities and the HR response to acceleration and recovery [5,29,37]. As physical activity encompasses both body movement and an associated cardiovascular response, leveraging these signals concurrently allows for a better evaluation of the temporal dynamics of CRF and enhances the understanding of the individual's physiology [27,37,47]. It should be emphasized, though, that contextualizing HR in unlabelled data requires navigating many intricacies, as investigators need to account for external factors that can influence HR, such as emotional stress, illness, heat, or medications that can potentially lead to invalid results.

In addition, a key observation arising from studying the multimodal features is the inverse relationship between VO_2max and HR at a given physical activity, which conforms to what is seen during the submaximal tests [5,29,33]. This observation reinforces the concept of estimating fitness from free-living activity, as even in the absence of controlled settings, behaviors approximating submaximal laboratory conditions will spontaneously occur [47]. Neshitov et al [29] demonstrated this inverse relationship, with the slope of the HR-over-cadence regression line being lower for participants with high than those with low VO_2max, and it was mainly noticeable between the 60 - 100 steps per minute exercise effort. Interestingly, Bonomi et al [16] highlighted the need for activity-specific prediction equations, showcasing models that combined energy expenditure and HR based on different activity types. Ultimately, tailoring predictive models to account for specific activity patterns and physiological responses enhances the accuracy of CRF predictions.

Challenges

Despite their potential, free-living data present some intrinsic statistical and computational challenges [48]. Using wearables in out-of-hospital, free-living settings often results in lower-quality, noisy data that require heavy filtering and preprocessing to become usable. Vigorous human motion can disturb on-body sensors and easily corrupt cardiac and accelerometer signals [40]. Aside from noise, missing data can also prevent meaningful features from being extracted. As reported in the "Results" section, preprocessing techniques are an essential step in the data-mining process to ensure that only reliable data points contribute to predictions. Another challenge lies in the precise physical activity detection in unlabeled data. Owing to the diverse nature of daily living, activity patterns overlap, and assumptions are occasionally made on how certain patterns in the data correlate with physical activities [5,49]. Furthermore, the abundance of sedentary data often leads to a data imbalance bias when low-intensity activities are overrepresented, leading to inaccurate estimations [5,16]. Consequently, due to the novelty of the task and the challenges outlined, there is currently no consensus on a specific approach and model that is most suited for free-living data.

Regression and ML Approaches

Interestingly, in the subgroup analysis, regression-based models appeared to perform slightly better than ML approaches. Much research examining the 2 approaches has repeatedly demonstrated comparable performance between regression-based and ML approaches, but this is not universal [38,39]. Nonetheless, this finding from our review warrants attention, as it likely stems from issues such as reporting bias and overfitting rather than genuine superiority. Results are influenced by the notable disparity in sample sizes, with 1 study driving the pooled estimate in ML modeling studies [27]. Finally, the limited external validation in ML suggests that the robustness and interpretability of simpler models can, in some cases, outweigh the complexity when appropriate validation has not been considered.

Limitations

This systematic review showed that free-living data can be valuable for CRF prediction and may prove a useful alternative in a variety of clinical settings. However, some limitations merit attention. First, although we observed promising preliminary agreement between VO_2max estimates and predictions, we need to acknowledge that, although we chose the correlation coefficient as the primary effect size for the meta-correlation analysis for its availability, it is not an accuracy metric and therefore does not imply that predictions are close in absolute terms. Further, as in certain instances, conversion of the R^2 was applied, this may have artificially inflated perceived predictive ability and, as such, influenced the overall result. Error-based metrics such as root-mean-square error or SEE would better capture accuracy, which is particularly relevant in clinical settings, but these measures were not consistently reported.

Second, there was significant heterogeneity and variance among the included studies, which varied in trial design, sample size, wearable device used, and modeling approach. Understandably, this limits the generalizability of our results, and the pooled estimate needs to be interpreted cautiously. However, in contrast to the synthesis of randomized trials, heterogeneity is frequently noted in meta-analyses of predictive modeling studies, mainly due to the disparity of eligible study designs or models [50].

Some studies used resting HR as a feature in their models, limiting monitoring to daytime periods only, excluding nocturnal HR data [31,32,36,37]. Research, however, demonstrates that using nighttime data can yield closer estimations to true testing values when it comes to resting HR [51]. As such, using daytime-derived resting HR may have implications for model performance and potentially lead to erroneous results. Godkin et al [51] underline the lack of standardization and considerable shortcomings among the criteria and methods used to estimate

XSL•FO
RenderX

resting HR. Since daytime HR can be affected by numerous behavioral, psychological, and environmental factors, we advocate for continuous monitoring of free-living data that captures both activity and rest phases for a more stable profile of HR distributions.

Considerations should also be made when interpreting the outputs of the presented models, as the papers reporting the highest correlations between predicted and actual measurements share several common features. Notably, several studies presented outputs from regression models without using unseen data for validation. Directly comparing these outputs against models validated on unseen data likely overestimates the model's true predictive ability. In such cases, high performance may reflect only how well the data fit the model, not its ability to generalize. Therefore, the lack of external validation may contribute to inflation in the aggregated meta-correlation analysis. Under such conditions, a single pooled correlation does not necessarily reflect a uniform level of accuracy and should certainly not be viewed as evidence of clinical readiness. Instead, it demonstrates the potential for future research that a strong association may be achieved.

Another limitation concerning the applicability and validation of the reported models is the selection bias, as most trials recruited young, healthy volunteers, making models less applicable to patient populations. We found that models predict VO_2max with an average 9%, which, arguably, may be clinically relevant in borderline cases—for instance, during preoperative risk assessment of frail individuals. Consequently, no direct conclusions can be drawn for clinical decision-making, and future research should focus more on medical settings to assess the effect of patient-specific factors, such as regular medication and comorbidities, in model training. Interestingly, although there were several different devices used in the included trials, no study explored the practicalities of monitoring patients remotely to collect the free-living data, which could explain the quality issues of noise and missingness that these datasets exhibit.

Implications for Real-World Use

The absence of a shared methodological framework across studies remains a major barrier to translation. As stated previously in the "Limitations" subheading, heterogeneity in this review was high, with studies using distinct device types, signal processing strategies, and validation tactics, often with insufficient external testing. Consequently, this limits confidence in generalizability and reproducibility, making direct comparisons difficult. From a clinical perspective, a model optimized for one setting may fail to transfer effectively in different patient groups, sensors, or wear patterns [52].

For clinical adoption to become feasible, several credibility issues need to be addressed. Pitfalls such as model overfitting, lack of standardized analytical pipelines, and limited evidence that performance is stable under real-world conditions (device updates, medication effects on HR) present significant challenges. Similar challenges have been identified in studies of prediction modeling for CVDs, highlighting the need for independent tools to assess replicability and external validation [52]. Therefore, until uniform data handling and transparent external validation become routine, results from research in this field should be considered promising but not ready for clinical application.

Future Considerations

Despite the challenges and limitations identified in this review, several models reported here should not be overlooked in this expanding research field. Future research should aim to streamline the deployment of wearable devices in out-of-hospital settings and educate and support patients and clinical teams. Furthermore, given the increasing influence of the wearable industry in health care, it is essential for such predictive models to undergo rigorous validation before being fully integrated into clinical practice. Establishing consensus on feature extraction, validation, and reporting guided by frameworks such as TRIPOD-AI and recent calls for transparency in wearable research (INTERLIVE) are recommended for future research to yield reproducible and clinically useful results [21,53].

For public health systems, regulatory frameworks regarding the digital storage, privacy, and security of the vast amount of patient-generated data should be considered early. With the myriad of wearables available, it is important that feasibility work is undertaken to set standards for a reliable and accurate technology, helping avoid a repetitive cycle of temporary models being developed that cannot be extrapolated to clinical contexts or used in clinical practice. Finally, a cost-effective analysis will determine the viability of these remote monitoring systems, ensuring they offer a sustainable solution for patients and health care systems.

Conclusion

This work explores a novel concept for CRF estimation from unsupervised free-living patient data. Contrary to the current gold-standard CPET, which is a snapshot of the individual's functional capacity, wearable health monitoring in free-living conditions generates rich datasets that can be exploited to train models for fitness estimation. Several models are discussed in this paper, with studies applying ML to mine raw data and enhance accuracy.

The combined results from this review show promise, with good preliminary agreement between predictions and measured values. However, no firm conclusions can be drawn for clinical implementation due to the heterogeneity of the studies and the lack of external validation. Nonetheless, continuous data streams appear to be a valuable resource for ML methods to shed light on human behavior and health, leading to a step change in how we measure and monitor CRF, ultimately aiming to improve health outcomes.

XSL•FO
RenderX

Acknowledgments

This study was supported with in-kind assistance from the National Institute for Health and Care Research's HealthTech Research Centre in Accelerated Surgical Care (HRC-ASC) and Leeds BRC.

Funding

AD is funded through a Leeds Hospitals Charity Research Fellowship (A2001580 – Making Surgery Safer for Patients With Bowel Cancer) and a Bowel Research UK Grant (BRUK_SG_24012). ABS is funded through UK Research and Innovation. This research was funded by the Engineering and Physical Sciences Research Council (grant no EP/S024336/1).

Data Availability

This is a systematic review, and data were extracted from the included studies and presented in the tables above. No additional data are applicable in this instance.

Authors' Contributions

AD conceptualized the study and conducted the literature search. Data extraction was performed by AD, ABS, and MRK. AD and DW carried out the analysis. The original draft was prepared by AD, with review and editing contributed by ABS, DW, DG, JT, and DGJ. Supervision was provided by JT and DGJ.

Conflicts of Interest

None declared.

Multimedia Appendix 1 [DOCX File, 17 KB - mhealth_v14i1e69996_app1.docx]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File, 276 KB - mhealth_v14i1e69996_app2.docx]

References

1. Mandsager K, Harb S, Cremer P, Phelan D, Nissen SE, Jaber W. Association of cardiorespiratory fitness with long-term mortality among adults undergoing exercise treadmill testing. JAMA Netw Open 2018 Oct 5;1(6):e183605. [doi: 10.1001/jamanetworkopen.2018.3605]
2. Plasqui G, Westerterp KR. Accelerometry and heart rate as a measure of physical fitness: proof of concept. Med Sci Sports Exerc 2005 May;37(5):872-876. [doi: 10.1249/01.mss.0000161805.61893.c0] [Medline: 15870644]
3. Lang JJ, Prince SA, Merucci K, et al. Cardiorespiratory fitness is a strong and consistent predictor of morbidity and mortality among adults: an overview of meta-analyses representing over 20.9 million observations from 199 unique cohort studies. Br J Sports Med 2024 May;58(10):556-566. [doi: 10.1136/bjsports-2023-107849]
4. Novoa NM, Varela G, Jiménez MF, Ramos J. Value of the average basal daily walked distance measured using a pedometer to predict maximum oxygen consumption per minute in patients undergoing lung resection. Eur J Cardiothorac Surg 2011 May;39(5):756-762. [doi: 10.1016/j.ejcts.2010.08.025] [Medline: 21146419]
5. Altini M, Casale P, Penders J, Ten Velde G, Plasqui G, Amft O. Cardiorespiratory fitness estimation using wearable sensors: Laboratory and free-living analysis of context-specific submaximal heart rates. J Appl Physiol (1985) 2016 May 1;120(9):1082-1096. [doi: 10.1152/japplphysiol.00519.2015] [Medline: 26940653]
6. Laveneziana P, Di Paolo M, Palange P. The clinical value of cardiopulmonary exercise testing in the modern era. Eur Respir Rev 2021 Mar 31;30(159):200187. [doi: 10.1183/16000617.0187-2020]
7. Jones L, Tan L, Carey-Jones S, et al. Can wearable technology be used to approximate cardiopulmonary exercise testing metrics? Perioper Med 2021 Dec;10(1). [doi: 10.1186/s13741-021-00180-w]
8. Pritchard A, Burns P, Correia J, et al. ARTP statement on cardiopulmonary exercise testing 2021. BMJ Open Resp Res 2021 Nov;8(1):e001121. [doi: 10.1136/bmjresp-2021-001121]
9. Noonan V, Dean E. Submaximal Exercise Testing: Clinical Application and Interpretation. Phys Ther 2000 Aug 1;80(8):782-807. [doi: 10.1093/ptj/80.8.782]
10. Glaab T, Taube C. Practical guide to cardiopulmonary exercise testing in adults. Respir Res 2022 Jan 12;23(1):9. [doi: 10.1186/s12931-021-01895-6] [Medline: 35022059]
11. Altini M, Casale P, Penders J, Amft O. Cardiorespiratory fitness estimation in free-living using wearable sensors. Artif Intell Med 2016 Mar;68:37-46. [doi: 10.1016/j.artmed.2016.02.002] [Medline: 26948954]
12. Kwon SB, Ahn JW, Lee SM, et al. Estimating maximal oxygen uptake from daily activity data measured by a watch-type fitness tracker: Cross-sectional study. JMIR Mhealth Uhealth 2019;7(6):e13327. [doi: 10.2196/13327]

XSL•FO
RenderX

13. Gulati M, McBride PE. Functional capacity and cardiovascular assessment: Submaximal exercise testing and hidden candidates for pharmacologic stress. Am J Cardiol 2005 Oct;96(8):11-19. [doi: 10.1016/j.amjcard.2005.06.016]
14. Bennett H, Parfitt G, Davison K, Eston R. Validity of submaximal step tests to estimate maximal oxygen uptake in healthy adults. Sports Med 2016 May;46(5):737-750. [doi: 10.1007/s40279-015-0445-1]
15. Tudor-Locke CE, Myers AM. Challenges and opportunities for measuring physical activity in sedentary adults. Sports Med 2001;31(2):91-100. [doi: 10.2165/00007256-200131020-00002]
16. Bonomi AG, Ten Hoor GA, de Morree HM, Plasqui G, Sartor F. Cardiorespiratory fitness estimation from heart rate and body movement in daily life. J Appl Physiol (1985) 2020 Mar 1;128(3):493-500. [doi: 10.1152/japplphysiol.00631.2019] [Medline: 31999530]
17. Ferreira JJ, Fernandes CI, Rammal HG, Veiga PM. Wearable technology and consumer interaction: A systematic review and research agenda. Comput Human Behav 2021 May;118:106710. [doi: 10.1016/j.chb.2021.106710]
18. Molina-Garcia P, Notbohm HL, Schumann M, et al. Validity of estimating the maximal oxygen consumption by consumer wearables: A systematic review with meta-analysis and expert statement of the INTERLIVE network. Sports Med 2022 Jul;52(7):1577-1597. [doi: 10.1007/s40279-021-01639-y] [Medline: 35072942]
19. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. PLoS Med 2021 Mar;18(3):e1003583. [doi: 10.1371/journal.pmed.1003583] [Medline: 33780438]
20. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5(1):210. [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
21. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024 Apr 16;385:e078378. [doi: 10.1136/bmj-2023-078378] [Medline: 38626948]
22. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med 2019 Jan 1;170(1):51-58. [doi: 10.7326/M18-1376]
23. R: The R Project for Statistical Computing. R Core Team. 2021. URL: https://www.r-project.org/ [accessed 2024-09-09]
24. Quintana DS. From pre-registration to publication: a non-technical primer for conducting a meta-analysis to synthesize correlational data. Front Psychol 2015;6. [doi: 10.3389/FPSYG.2015.01549/BIBTEX]
25. Daraj LR, AlGhareeb M, Almutawa YM, Trabelsi K, Jahrami H. Systematic review and meta-analysis of the correlation coefficients between nomophobia and anxiety, smartphone addiction, and insomnia symptoms. Healthcare (Basel) 2023 Jul 19;11(14):2066. [doi: 10.3390/healthcare11142066] [Medline: 37510507]
26. Bakbergenuly I, Hoaglin DC, Kulinskaya E. Methods for estimating between-study variance and overall effect in meta-analysis of odds ratios. Res Synth Methods 2020 May;11(3):426-442. [doi: 10.1002/jrsm.1404] [Medline: 32112619]
27. Spathis D, Perez-Pozuelo I, Gonzales TI, et al. Longitudinal cardio-respiratory fitness prediction through wearables in free-living environments. NPJ Digit Med 2022 Dec 1;5(1). [doi: 10.1038/s41746-022-00719-1]
28. Wu Y, Spathis D, Jia H, et al. Turning silver into gold: domain adaptation with noisy labels for wearable cardio-respiratory fitness prediction. . Preprint posted online on Nov 28, 2022. [doi: 10.48550/ARXIV.2211.10475]
29. Neshitov A, Tyapochkin K, Kovaleva M, et al. Estimation of cardiorespiratory fitness using heart rate and step count data. Sci Rep 2023 Sep 22;13(1):15808. [doi: 10.1038/s41598-023-43024-x] [Medline: 37737296]
30. Spathis D, Perez-Pozuelo I, Brage S, Wareham NJ, Mascolo C. Self-supervised transfer learning of physiological representations from free-living wearable data. Presented at: ACM CHIL '21: ACM Conference on Health, Inference, and Learning; Apr 8-10, 2021. [doi: 10.1145/3450439.3451863]
31. Ahn JW, Hwang SH, Yoon C, Lee J, Kim HC, Yoon HJ. Unobtrusive estimation of cardiorespiratory fitness with daily activity in healthy young men. J Korean Med Sci 2017;32(12):1947. [doi: 10.3346/jkms.2017.32.12.1947]
32. Frade MCM, Beltrame T, Gois MDO, et al. Toward characterizing cardiovascular fitness using machine learning based on unobtrusive data. PLOS ONE 2023 Mar 1;18(3):e0282398. [doi: 10.1371/journal.pone.0282398]
33. Plasqui G, Westerterp KR. Accelerometry and heart rate as a measure of physical fitness: cross-validation. Med Sci Sports Exerc 2006 Aug;38(8):1510-1514. [doi: 10.1249/01.mss.0000228942.55152.84] [Medline: 16888467]
34. Cao ZB, Miyatake N, Higuchi M, Ishikawa-Takata K, Miyachi M, Tabata I. Prediction of VO2max with daily step counts for Japanese adult women. Eur J Appl Physiol 2009 Jan;105(2):289-296. [doi: 10.1007/s00421-008-0902-8] [Medline: 18985375]
35. Cao ZB, Miyatake N, Higuchi M, Miyachi M, Ishikawa-Takata K, Tabata I. Predicting VO2max with an objectively measured physical activity in Japanese women. Med Sci Sports Exerc 2010 Jan;42(1):179-186. [doi: 10.1249/MSS.0b013e3181af238d] [Medline: 20010115]
36. Zhang Y, Wang X, Pathiravasan CH, et al. Association of smartwatch-based heart rate and physical activity with cardiorespiratory fitness measures in the community: Cohort study. J Med Internet Res 2024 Jun 13;26(1):e56676. [doi: 10.2196/56676] [Medline: 38870519]
37. Beltrame T, Amelard R, Wong A, Hughson RL. Extracting aerobic system dynamics during unsupervised activities of daily living using wearable sensor machine learning models. J Appl Physiol 2018 Feb 1;124(2):473-481. [doi: 10.1152/japplphysiol.00299.2017]

XSL•FO
RenderX

38. Lindsay T, Westgate K, Wijndaele K, et al. Descriptive epidemiology of physical activity energy expenditure in UK adults (The Fenland study). Int J Behav Nutr Phys Act 2019 Dec 9;16(1):126. [doi: 10.1186/s12966-019-0882-6] [Medline: 31818302]
39. Kannel WB, Kannel C, Paffenbarger RS, Cupples LA. Heart rate and cardiovascular mortality: The Framingham study. Am Heart J 1987 Jun;113(6):1489-1494. [doi: 10.1016/0002-8703(87)90666-1]
40. Syversen A, Dosis A, Jayne D, Zhang Z. Wearable sensors as a preoperative assessment tool: A review. Sensors (Basel) 2024 Jan 12;24(2):482. [doi: 10.3390/s24020482] [Medline: 38257579]
41. Khalid S, Khalil T, Nasreen S. A survey of feature selection and feature extraction techniques in machine learning. 2014 Presented at: 2014 Science and Information Conference (SAI); Aug 27-29, 2014. [doi: 10.1109/SAI.2014.6918213]
42. Kang S, Paul A, Jeon G. Reduction of mixed noise from wearable sensors in human-motion estimation. Computers & Electrical Engineering 2017 Jul;61:287-296. [doi: 10.1016/j.compeleceng.2017.05.030]
43. Kumar A, Dubey P, Ranjan A. Assessment of anxiety in surgical patients: An observational study. Anesth Essays Res 2019;13(3):503. [doi: 10.4103/aer.AER_59_19]
44. Mück JE, Ünal B, Butt H, Yetisen AK. Market and patent analyses of wearables in medicine. Trends Biotechnol 2019 Jun;37(6):563-566. [doi: 10.1016/j.tibtech.2019.02.001]
45. Waterland JL, Ismail H, Granger CL, et al. Prehabilitation in high-risk patients scheduled for major abdominal cancer surgery: a feasibility study. Perioper Med 2022 Aug 23;11(1). [doi: 10.1186/s13741-022-00263-2]
46. Franklin BA, Eijsvogels TMH, Pandey A, Quindry J, Toth PP. Physical activity, cardiorespiratory fitness, and cardiovascular health: A clinical practice statement of the American Society for Preventive Cardiology Part II: Physical activity, cardiorespiratory fitness, minimum and goal intensities for exercise training, prescriptive methods, and special patient populations. American Journal of Preventive Cardiology 2022 Dec;12:100425. [doi: 10.1016/j.ajpc.2022.100425]
47. Hallgrímsson HT, Jankovic F, Althoff T, Foschini L. Learning individualized cardiovascular responses from large-scale wearable sensors data. . Preprint posted online on Dec 4, 2018 URL: http://arxiv.org/abs/1812.01696 [accessed 2024-06-04]
48. Ferguson M, Shulman M. Cardiopulmonary exercise testing and other tests of functional capacity. Curr Anesthesiol Rep 2022;12(1):26-33. [doi: 10.1007/s40140-021-00499-6] [Medline: 34840532]
49. Wang Z, Zhang Q, Lan K, et al. Enhancing instantaneous oxygen uptake estimation by non-linear model using cardio-pulmonary physiological and motion signals. Front Physiol 2022;13:Wang. [doi: 10.3389/fphys.2022.897412]
50. Ahmed I, Debray TP, Moons KG, Riley RD. Developing and validating risk prediction models in an individual participant data meta-analysis. BMC Med Res Methodol 2014 Dec;14(1). [doi: 10.1186/1471-2288-14-3]
51. Godkin FE, Van Ooteghem K, Beyer KB, et al. Measuring resting heart rate during daily life using wearable technology: Examining the impact of behavioral context and methodological criteria. Digit Health 2025 May;11. [doi: 10.1177/20552076251367506]
52. Cai YQ, Gong DX, Tang LY, et al. Pitfalls in developing machine learning models for predicting cardiovascular diseases: Challenge and solutions. J Med Internet Res 2024;26:e47645. [doi: 10.2196/47645]
53. Schumann M, Feuerbacher JF, Heinrich L, et al. Using free-living heart rate data as an objective method to assess physical activity: A scoping review and recommendations by the INTERLIVE-network targeting consumer wearables. Sports Med 2025 Feb;55(2):275-300. [doi: 10.1007/s40279-024-02159-1]

Abbreviations

CPET: cardiopulmonary exercise testing
CRF: cardiorespiratory fitness
CVD: cardiovascular disease
ECG: electrocardiogram
HR: heart rate
ML: machine learning
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROBAST: Prediction Model Study Risk of Bias Assessment Tool
SC: step count
SEE: standard error of estimate
SVM: support vector machine
TRIPOD-AI: Transparent Reporting of a multivariable or machine learning prediction model for Individual Prognosis Or Diagnosis–artificial intelligence
VO_2 max: maximal oxygen uptake

XSL•FO
RenderX

Edited by L Buis; submitted 13.Dec.2024; peer-reviewed by A Martinko, B Xiang, S Tedesco, S Jiang; revised version received 26.Oct.2025; accepted 17.Dec.2025; published 27.Jan.2026.

Please cite as:
Dosis A, Syversen AB, Kowal MR, Grant D, Tiernan J, Wong D, Jayne DG
Exploiting Unsupervised Free-Living Data for Cardiorespiratory Fitness Estimation: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2026;14:e69996
URL: https://mhealth.jmir.org/2026/1/e69996
doi:10.2196/69996

© Alexios Dosis, Aron Berger Syversen, Mikolaj R Kowal, Daniel Grant, Jim Tiernan, David Wong, David G Jayne. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 27.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Digital Health Interventions to Support Chronic Disease Management: Systematic Scoping Review

Abdullah Al Mahmud[1], PhD; Shane Joachim[2], MSc; Prem Prakash Jayaraman[3], PhD; Caitlin Learmonth[4], MSc; Shivani Tyagi[5], PhD; Abdur Rahim Mohammad Forkan[2], PhD; Muhammad Shuakat[6], PhD; Nilmini Wickramasinghe[7], PhD; Jack Wheeler[8], MSc; Stephanie Best[9], PhD; Alison Trainer[8], PhD

[1]Centre for Design Innovation, Department of Architectural and Industrial Design, Swinburne University of Technology, John St., Hawthorn, Melbourne, Australia

[2]Department of Computing Technologies, Swinburne University of Technology, Melbourne, Australia

[3]School of Science, Computing and Engineering Technologies, Factory of the Future and Digital Innovation Lab, Swinburne University of Technology, Melbourne, Australia

[4]Swinburne University of Technology, Melbourne, Australia

[5]Department of Communication Design, School of Design and Architecture, Swinburne University of Technology, Melbourne, Australia

[6]La Trobe University, Melbourne, Australia

[7]School of Computing, Engineering & Mathematical Sciences, La Trobe University, Melbourne, Australia

[8]Parkville Familial Cancer Centre, Peter MacCallum Cancer Centre, Melbourne, Australia

[9]School of Health Sciences, University of Melbourne, Melbourne, Australia

Corresponding Author:
Abdullah Al Mahmud, PhD
Centre for Design Innovation, Department of Architectural and Industrial Design, Swinburne University of Technology, John St., Hawthorn, Melbourne, Australia

Abstract

Background: Health interventions delivered by digital platforms are gaining popularity and are evolving to address the needs of patients with chronic diseases. The heterogeneity of chronic diseases requires that digital health platforms vary in their approaches to chronic disease management.

Objective: This review aimed to explore the characteristics of digital health platforms and the corresponding digital interventions developed to support patients with chronic diseases. This includes those platforms' design, development, and the metrics by which any incremental benefits they provide are assessed.

Methods: We searched electronic databases including Scopus, Web of Science, PsycINFO, IEEE Xplore, MEDLINE, and Embase. Relevant articles published from January 2013 to November 2024 were extracted. Extracted data were then synthesized using qualitative content analysis and presented in narrative form with relevant tables.

Results: In total, we identified 69 digital health platforms supporting the management of 20 chronic diseases. Most platforms were mobile apps (n=22) or a combination of web and mobile apps (n=15). Most of the platforms (n=44) were tailored to support self-management of chronic diseases. These platforms also provided a web-based portal where health care providers could review and manage the information recorded by patients. In 77% (53/69) of the studies, patients reported that the digital interventions delivered by the platform improved their quality of life, their health, and their ability to self-manage their chronic diseases. In addition, health care providers reported positive outcomes, including improved clinical utility and patient communication. While short-term health outcomes of the digital health interventions were largely positive, long-term health outcomes remain unknown. This was because most of the studies were short-term pilots and often formative in nature (n=42). Many had limited sample sizes, limited participant uptake of the digital platforms, and technical issues. In many cases, further personalization of platforms was required to meet patients' self-management needs.

Conclusions: Digital health interventions can be beneficial in the management of chronic disease. The adoption of digital interventions in combination with regular clinical care can improve health outcomes, support self-management, and enhance communication between patients and health care providers. However, long-term user engagement is the major barrier to their long-term success. High dropout rates, often resulting from a lack of motivation or technical issues, testify to the need for adaptive, low-burden interventions that function seamlessly in users' daily lives. Adopting user-centered and co-design approaches that engage both clinicians and patients in designing digital health platforms may enhance the usability and uptake of such platforms.

(JMIR Mhealth Uhealth 2026;14:e63742) doi:10.2196/63742

XSL•FO
RenderX

KEYWORDS

mHealth; chronic disease; digital technology; co-design; user-centered design; user centered; chronic disease management; support; scoping review; design; development; digital health platform; qualitative content analysis; web-based; self-management; quality of life; clinical utility; patient communication

Introduction

Background

Chronic diseases are defined as long-lasting conditions (ie, 1 y or more) and require ongoing care and account for 41 million (74%) deaths worldwide [1-3]. They can cause disability that affects the quality of life and reduces life expectancy. Patients with chronic diseases may face several challenges when managing their condition, such as (1) conflicting knowledge about the disease or how to manage it, (2) access to care, and (3) communication with health care providers [4]. Chronic diseases pose a significant burden on health care systems, families, and caregivers [5]. Therefore, the prevention and management of chronic diseases has become a global priority, as the prevalence of those diseases can undermine social and economic development.

Digital health interventions use technology to deliver health care services or treatments and facilitate knowledge exchange [6]. These interventions are designed to enhance the quality of patient care by capturing and conveying information in a digital format. Digital health interventions may involve electronic medical records (EMRs), mobile apps or web applications, and wearable sensors such as Fitbit (Google). Technologies such as digital health platforms are gaining increased use in managing chronic diseases [7]. The proliferation of mobile apps and the ubiquitous nature of information technology have fueled the development of platforms that support the management of chronic disease [8,9]. The digital health technologies discussed in this study include mobile apps, web applications, electronic health records (EHRs), EMRs, wearable devices, and telehealth services [10]. We have used digital platforms as an umbrella term throughout the paper to denote these technologies.

With the rapid growth of digital platforms for chronic disease management, a systematic synthesis of these platforms is needed to inform effective and efficient care. Recent studies have investigated the framework for managing chronic diseases [11] and the potential of technology adoption [12]. To our knowledge, no studies have examined how digital platforms support the management of chronic diseases. Due to the heterogeneity of digital platforms, it is vital to investigate the types of digital platforms that are available and the usability and acceptability of these platforms. In addition, it is important to investigate the processes that led to their design and development, as well as the metrics used to assess their benefits. In this study, digital platforms are characterized as assortments of web-based and mobile applications and related technologies that are used to deliver health care services [13].

Aims

This review aims to explore the characteristics of digital health platforms and corresponding digital health interventions that support patients with chronic disease. This exploration will include those technologies' design, their development processes, and the metrics by which their incremental and long-term benefits have been assessed.

Methods

Overview

This review adopts the scoping review methodology proposed by Arksey and O'Malley [14] because we are interested in identifying and mapping emerging evidence [15]. PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) checklist [16] was used throughout the review to ensure adherence (Checklist 1). However, there are some differences between the registered protocol and this paper regarding database searching. Furthermore, 2 additional databases (Medline and Embase) have been searched, and the search period has been extended to November 2024 for all the databases.

Step 1: Identifying the Research Questions

First, what are the characteristics of digital platforms that support the management of chronic diseases, including self-management and provider-led management?

Second, what principles and theoretical frameworks have been used to design or co-design these platforms?

Third, how were these platforms evaluated for clinical utility?

Finally, what is the effectiveness of those platforms?

Step 2: Search Strategy

The search terms used for the literature search are "chronic disease" OR "chronic illness," OR "long-term conditions," OR "chronic conditions" AND "Digital" OR "mHealth" OR "App" AND "management."

We did not include the exact term "self-management" in our search, nor the standard indexing terms that databases use for that idea (eg, PubMed's MeSH heading Self Care and Embase's Emtree term self-management). To reduce the chance of missing papers, we implemented citation chasing, that is, for each included study, we checked its reference list and looked up newer papers that cite it.

Initially, databases such as Scopus, Web of Science, PsycINFO, and IEEE Xplore were searched from January 2013 to 30 November 2022. Later, an additional search was conducted on these databases from November 2022 to November 2024. Furthermore, 2 additional databases, MEDLINE and Embase, were searched from January 2013 to November 2024 (Refer to Multimedia Appendix 1 for the search outcomes). Articles retrieved were imported into Covidence software (Veritas Health Innovation Ltd) [17], and duplicated items were automatically removed.

XSL•FO
RenderX

Step 3: Study Selection

The following inclusion and exclusion criteria were used.

Inclusion criteria

First, the study must be a peer-reviewed journal article and present primary data. Second, it should be published within the last 12 years (January 2013 to November 2024). We chose 2013 as the starting year to reflect a critical turning point in the evolution of digital health technologies. Around this time, the widespread use of smartphones and mobile apps, breakthroughs in wearable sensors, and more prevalent usage of EHRs began to transform chronic disease management [18,19]. Third, it should be available in English. Finally, it should involve digital applications in the context of chronic disease management.

Exclusion criteria

The study was excluded if (1) it was a review article or opinion piece, or (2) hypothetical use of digital technology was found in it.

Initially, titles and abstracts were reviewed against the selection criteria and were marked as "include," "exclude," or "uncertain." Two reviewers (AAM and MS) conducted the screening independently, and regular discussions with the research team were undertaken to resolve any discrepancies and to fine-tune selection criteria. This screening and discussion process continued until we reached a consensus. Subsequently, for the included studies, a full-text review was carried out (AAM and MS) against the selection criteria following the same screening procedure.

Step 4: Data Extraction and Charting the Data

Two authors (AAM and MS) developed a data charting form to identify relevant information to extract from the included studies. Using this form, the following data were extracted: study citation, publication type, authors, study location, study year, user acceptability of the digital interventions or digital platforms, and outcome of the study (quantitative results, qualitative themes, recommendations, key learnings, and limitations). AAM and MS charted the extracted information.

Step 5: Collating, Summarizing, and Reporting the Results

First, the extracted data were analyzed using descriptive statistics (eg, frequencies). This provided numerical summaries of (1) digital platforms and their characteristics, (2) chronic diseases, (3) platform design principles, and (4) outcomes. These details were presented using tables, charts, and graphs, followed by a summary. Second, 2 authors (AAM and CL) independently analyzed the extracted data thematically to identify themes. Results from the 2 reviewers' thematic analysis were combined to select the final collection of themes.

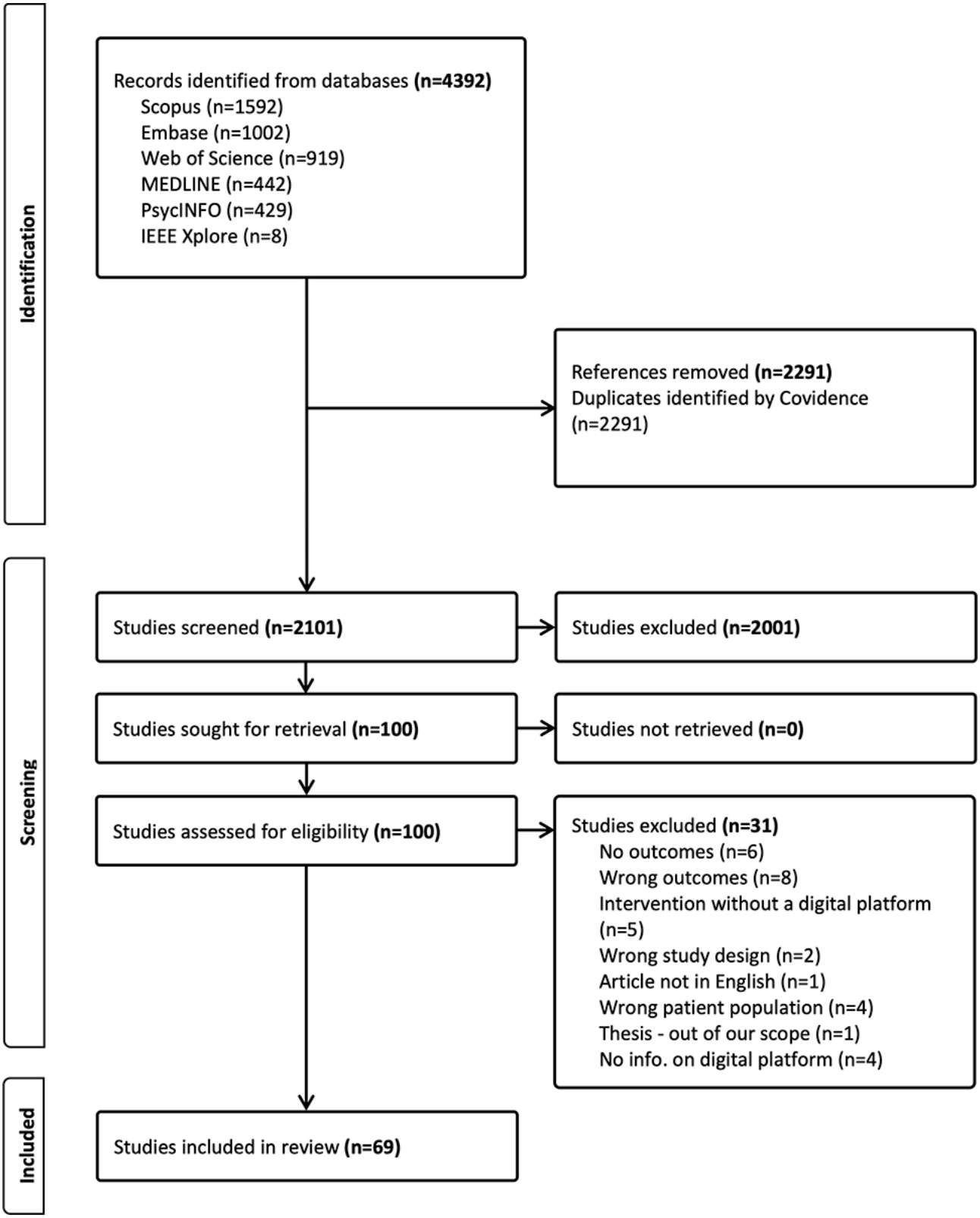
Results

Overview

A total of 4392 studies were identified from the 4 databases. Of these, 2291 were duplicates, leaving 2101 to be screened. In total, 2001 studies were excluded during the title and abstract screening process, and 100 were assessed for eligibility. Of these, 69 studies [20-88] met our eligibility criteria and were included for review. Refer to Figure 1 for the article selection process.

XSL•FO
RenderX

Figure 1. PRISMA diagram showing the article selection process.



Characteristics of the Included Studies

In total, 69 studies were conducted in 26 countries. Of the total, 15 studies (22%) were conducted in the United States [22,25-27,30,31,37,38,49,50,52,55,61,66,85]. Canada had 5 studies (7%) [38,49,50,76,84], and China had 10 studies (15%) [24,32,41,43,58,63,65,69,79,81], while 4 studies (5%) [28,59,67,83] were conducted in Spain. The most common chronic disease reported was type 2 diabetes (n=17, 20%), followed by heart failure (n=11, 13%), chronic obstructive pulmonary disease (COPD; n=9, 11%), and hypertension (n=9, 11%) (Multimedia Appendix 2). In total, 27 studies (39%) [21,22,24,27-29,31,33,34,37,41,42,45,49,50,52,54,55,59,60-68,69] are reported as randomized controlled trials (RCTs), and the

XSL•FO
RenderX

remaining 42 studies (61%) are formative studies [20,23,25,26,30,32,35,36,38-40,43,44,46-48,51,53,56-58,61,62,70-88]. Table 1 summarizes the aims of the digital platforms, the platform types, and the study settings.

Table . Summary of the platforms.

Platform characteristics	Total, n (%)
Aim of the digital platform (n=77)	
Self-management[a]	44 (57)
Behavior changes	17 (22)
Communication with health care providers	16 (20)
Types of digital platforms (n=69)	
Web-based application	15 (22)
Mobile app	33 (48)
Combination of web-based application and mobile app (ie, multimodal)	14 (20)
Other[b] (SMS text messaging)	7 (10)
Wearable device (in combination with web app, smartphone app, or both)	16 (23)
Study setting (n=78)	
Hospital or primary care setting	34 (44)
Home	36 (46)
Online Community	8 (10)

[a]Includes self-monitoring of symptoms, medication, physical activity, etc.

[b]One intervention used telehealth and SMS text messaging.

Characteristics of the Study Participants

All the included studies had adult participants (aged >18 y). Participants from various age groups were included in the studies (30 - 35 y: n=2, 40 - 45 y: n=1, 45 - 50 y: n=1, 60 y or older: n=19, and 75 - 80 y: n=1). Several studies (n=15) did not specify the mean age or age range. Most of the studies are dominated by female participants, especially studies related to the management of chronic diseases such as asthma (eg, 80 women vs 26 men [20]), hypertension (eg, 49/67, 73% female [21]), diabetes, and rheumatoid arthritis. Some studies target specific gender groups to address health disparities or disease prevalence, such as Black women with hypertension [22]. Most chronic conditions are higher in females, which may be a reason for overrepresentation. Some studies, like those dealing with heart failure (eg, N=25, 100% male participants [23]), atrial fibrillation (eg, 59/96, 61.5% male [24]), and other cardiovascular disorders (eg, 48/79, 61% male [54]), might have a greater proportion of males. In some instances, these conditions may be more prevalent in men within specific age groups. Most male-dominated studies occur in areas such as cardiovascular health, where men are more prone to certain conditions.

In total, 6 studies [40,54,57,58,71,89] mentioned the literacy levels of participants in the participant eligibility criteria. Only 1 study [26] mentioned digital literacy (ie, technical and cognitive abilities to use information and communication technologies [89]), defined as "acceptable literacy level to read and write with a smartphone," in the inclusion criteria. None of the studies measured the participants' digital literacy level as part of the intervention. Most studies (n=19; 61%) had predominantly male participants (66%). Where literacies were mentioned, the most reported literacy was linguistic (English language literacy, n=2 out of 6), followed by health literacy [90], defined as the capability to process and understand health information (n=2 out of 6).

Characteristics of the Digital Interventions

The digital interventions identified in the literature targeted a range of chronic diseases and thus had a variety of operational aims. Multimedia Appendix 3 provides a summary of the digital intervention strategies used in the studies. The definition of each digital intervention is extracted from the National Institute for Health and Care Excellence (NICE) Framework [91]. An analysis of the digital intervention strategies based on the NICE Evidence Standards Framework for digital health [1] found self-management to be the most prevalent strategy, appearing in 68 studies (99%) [20-46,49-86,88]. Collaborative care (ie, self-management plus provider-led management) followed with 43 occurrences (39%) [20-25,27,28,32-37,39-42,44,51-53,56,57,60-63,65,67,69,70,72,74,77-83,85,87], while information and education were present in 27 studies (39%) [20-24,26,33,37,40,45,46,57,59,63,64,66,72-79,81,84,85]. Personal health record systems were identified in 25 papers (62%), digital therapeutics in 15 papers (22%) [22,23,25,28,35,44,45,51,59,60,63,66,72,73,77], clinical decision support systems in 8 papers (12%) [26,28,44,48,59,65,70,72], and active monitoring in 1 paper (2%) [30].

XSL•FO
RenderX

Self-management frequently co-occurs with collaborative care (47 times), reflecting strong integration between patient-driven health management interventions and the need for clinical support (ie, provider-led management). Information and education (31 times) and personal health record systems (25 times) also frequently co-occur with self-management, underscoring the importance of providing users with relevant knowledge and real-time data to make informed health decisions. This, in turn, enhances self-care practices and promotes better health outcomes. Additionally, digital therapeutics (15 times) and clinical decision support systems (8 times) demonstrate strong associations, emphasizing their role in remote treatment guidance and data-driven clinical decision-making. These findings highlight how personalized digital interventions, combined with real-time insights, can enhance both patient engagement and clinical oversight.

Features of the Digital Platforms and Behavior Change Techniques Offered

Overview

The most common features by far are self-monitoring and tracking (65/69 studies, 94%), showing the trend of users' empowerment for active management (Table 2) of chronic diseases with support from health care providers. Medication reminders (59/69 studies, 86%) and behavioral support features (48/69 sources, 70%) are essential in helping patients stay committed to their treatment plans, with alerts (59/69 studies, 86%) and gamification using points, badges, or levels (8/69 studies, 12%) playing key roles in engagement. Most of the platforms are also targeted at improving user awareness and engagement by educating the users about better health practices (55/69 sources, 80%) and communicating with health care providers (45/69 studies, 72%). Such features ensure that patients are supported not only through technology but also by and through human contact.

XSL•FO
RenderX

Table . Summary of the most common features across the 69 studies, including their frequencies.

Category and feature	Frequency
Mobile app (n=32)	
Self-monitoring	15
Patient education	10
Reminders	18
Data tracking	14
Communication (between patient and health care providers)	12
Personalized feedback	8
Medication adherence	7
Activity tracking	9
Health reporting (eg, symptom, weight, and blood pressure)	6
Push notifications	7
Web-based application (n=15)	
Health data management	8
Patient education	5
Health reporting	6
Care team communication	4
Goal setting	3
Progress tracking	4
Data visualization	2
Document management (eg, laboratory results and prescriptions)	2
Personalized feedback	1
Combination of web-based application and mobile app (multimodal; n=16)	
Health data integration	7
Communication (between patient and health care providers)	5
Health tracking	6
Goal setting	5
Personalized feedback	4
Reminders and notifications	5
Symptom tracking	4
Educational content	3
Activity tracking	3
Wearable device (in combination with web application or mobile app; n=16)	
Physical activity tracking	9
Health data measurement (eg, heart rate, blood pressure, oxygen saturation, and SpO2[a])	8
Real-time monitoring	7
Data syncing (with mobile app or cloud)	6
Reminder and alert functionality	5
Health reporting	4
Medication reminders	3
Data visualization	3

XSL•FO
RenderX

Category and feature	Frequency
Other (SMS text messaging; n=7)	
Reminders (medication, appointments, etc)	7
Behavioral triggers	5
Education (health tips and guidance)	6
Symptom reporting	4
Daily check-ins or reports	3

[a]SpO2: peripheral oxygen saturation.

Table 2 provides a summary of the features and their frequencies by category. Routine input of health data was facilitated through surveys and questionnaires, using free-text or drop-down menu functionalities. Some apps facilitated image uploads, allowing patients to supply photos of wounds, rashes, or other relevant aspects of their conditions to be assessed by health care providers [27,28]. To support the self-management and tracking of health data, many platforms use tools for self-measurement and reporting of key health indicators, such as blood pressure (BP), glucose levels, and heart rate. Furthermore, 2 platforms embedded these functionalities in motion sensors [29,30]. However, the majority of platforms incorporating self-reporting functionalities did so through Bluetooth-enabled technology, such as smart watches, BP monitors, and scales [31-36]. These either fed data directly to the platforms or provided data for participants to input manually. Refer to Table 3 for the overview of the wearable devices used in the included studies.

XSL•FO
RenderX

Table . Wearable devices used in the included studies (n=16).

Wearable device	Type of wearable	Purpose	Study
Fitbit	Activity tracker or fitness band	General health monitoring, physical activity, and heart rate tracking	Oh et al [34]
Apple Watch	Smartwatch	ECG[a], heart rate monitoring, activity tracking, and general health monitoring	Guo et al [32]
Omron BP Cuff	Blood pressure monitor	Blood pressure tracking and monitoring for hypertension	Evans et al [30]
Dexcom CGM	Continuous glucose monitor	Continuous glucose monitoring for diabetes	Schnall et al [37]
iHealth Pulse Oximeter	Pulse oximeter	Monitoring oxygen levels (SpO2[b]) for respiratory conditions	Guo et al [32]
AliveCor KardiaMobile	ECG monitor	ECG readings for detecting arrhythmias	Gray et al [38]
Masimo Pulse Oximeter	Pulse oximeter	Monitoring oxygen saturation levels (SpO2)	Kryger et al [27]
Apple Watch	Smartwatch	Heart rate monitoring, ECG, and wellness tracking	Burda et al [39]
Fitbit	Activity tracker or fitness band	Step counting, heart rate monitoring, and sleep tracking	Bailey et al [40]
FreeStyle Libre	Continuous glucose monitor	Continuous glucose monitoring for diabetes management	Dorsch et al [31]
Google Glass	Smart glasses	Augmented reality for hands-free monitoring during medical procedures	Zhang et al [41]
Quell	Pain management wearable	Nerve stimulation for chronic pain relief	Cormican and Dowling [88]
Oura Ring	Wearable sleep tracker	Sleep quality tracking and recovery monitoring	Poppe et al [42]
Masimo	Pulse oximeter	Continuous SpO2 monitoring for respiratory health	Ji et al [43]
WHOOP Strap	Wearable fitness tracker	Sleep and recovery tracking for physical performance enhancement	Doyle et al [44]
iHealth Pulse Oximeter	Pulse oximeter	Monitoring oxygen levels in the blood	Ji et al [43]

[a]ECG: electrocardiogram.

[b]SpO2: peripheral oxygen saturation.

Only 5 studies [29,44-47] explicitly mentioned the use of artificial intelligence (AI) in the digital platforms. In these, AI was used to aggregate and analyze patient data across multimodal platforms. This was achieved through machine learning algorithms or recommender systems on platforms that use conversation agents. For example, the Snapcare app (Snapcare Technologies Pvt Ltd) gathered daily activity data over 12 weeks to address chronic back pain, including walking distance and workouts [29]. Notifications were sent to patients based on app usage and physical activity data collected by built-in phone sensors. This information was automatically transferred to a secure server, where machine learning algorithms examined the daily data on physical activity and produced suggestions for the session the following day. Similarly, the ProACT digital platform helps older patients manage their multimorbidity, including diabetes, congestive heart failure, COPD, and chronic heart disease [44]. Health information, including BP, heart rate, blood glucose, pulse oximetry, weight, activity, and sleep readings, is collected using off-the-shelf technologies. ProACT's AI algorithms can gain knowledge from the data to provide more accurate personalized recommendations and highlight a condition that needs attention. Likewise, using an embodied conversational agent named Laura, My Diabetes Coach (The University of Melbourne) provides patients with individualized support, monitoring, and motivational coaching [45]. The algorithms and conversational scripts that direct each person's progress were developed using behavior change theories, and they can accommodate recommendations by a general practitioner. Natural language processing and automated speech recognition are used to enhance the capability of a voice-enabled chatbot in user interactions in French [46]. The Medly voice app (University

XSL•FO
RenderX

Health Network) for heart failure management leverages various AI technologies, including machine learning, natural language processing, and automated speech recognition, which process the speech of users and formulate responses [47].

In total, 14 of the platforms included in the review were multimodal. For example, a platform might function as a combination of web-based and smartphone-based applications or work in combination with a wearable Bluetooth device, such as a smartwatch or BP monitor. Of the total, 8 interventions included some form of wearable technology or measurement device. These ranged from smartwatches and activity trackers [30-32,40,44] to BP monitors [33,34,36,44], glucose monitors [34,44], scales [31], and sleep monitoring devices [44]. Additionally, 15 interventions [26-28,35,37,39,44,47,48,50,53,58,61,62,71] were designed to work in conjunction with in-person consultations. These included functionalities that were integrated with routine clinical visits to monitor appointments, provide technical support, and review progress. Digital platforms in 10 studies supported multimorbidity [27,34,35,38,44,48-52].

Behavior change was broadly reported to be a multistep process requiring numerous complementary features. Tailoring features to the specific needs of participants based on both demographics and disease type was shown to be important in achieving behavior change [32,48,53]. Although not all of the studies explicitly used theoretical frameworks for behavioral change, many of the platforms are underpinned by theories intended to enable behavior change. Theories applied include social cognitive theory, which is detailed in the work of Dale et al [54], and the transtheoretical model, which is elaborated upon by Salari et al [53]. Furthermore, Sittig et al [55] explore the Fogg Behavior Model, and both cognitive theory and self-efficacy theory are examined by Park et al [60].

A range of methods was used to deliver tailored behavior change features to facilitate user engagement and improved health outcomes. Reminders, goal-setting, and motivational messages were widely used in 5 studies [38,39,44,56,88]. Those features aimed to motivate patients to engage with their health interventions and help them stay on track with their treatment plans. Behavioral trigger messaging was used to promote engagement, motivation, and self-belief. It was also used to provide reinforcement to participants based on behavior change theories [55]. Similarly, theory-based approaches specific to tailored SMS text messaging were used to improve personalization and increase the acceptability of the health interventions [32,53].

Gamification and learning material were incorporated in some studies for the optimization of motivation and to increase users' knowledge of managing their conditions. Gamification, either through rewards or challenges, stimulated usage of the app, while educational information [26,33,40,57] provided valuable information to allow patients to better understand their conditions and make informed decisions on their treatment. Furthermore, 5 platforms used gamification techniques to enhance user experience while embedding learning principles [29,42,45,58,59]. In-app conversation agents supported by AI [45] were used in a platform to change behavior too.

Furthermore, self-monitoring, tailored feedback, and social support were key features of some platforms [26,35,40,60]. These features fostered engagement and behavior change. Self-monitoring permitted users to monitor their health measurements (ie, BP or blood glucose), while personalized feedback assisted in matching interventions to individual needs and encouraging ownership of one's health. Social support, either through peer interaction or direct contact with health care providers, also facilitated long-term engagement by providing users with emotional encouragement and accountability.

Chronic Disease Management: Patient Self-Management and Provider-Led Care

In analyzing digital platform features, we distinguish 2 complementary approaches: patient self-management and provider-led management. Provider-led management refers to the coordinated, proactive activities delivered by health care services and clinicians to support people living with chronic conditions, typically including comprehensive care planning, risk stratification, guideline-concordant treatment, continuity, between-visit care coordination, and use of clinical information systems for monitoring and quality improvement across settings and providers [92]. In contrast, patient-led self-management encompasses the day-to-day work undertaken by people living with chronic conditions, including symptom monitoring, medication adherence, lifestyle and behavioral changes, decision-making, help-seeking, and the usage of tools and supports, including digital platforms [93].

The vast majority of digital platforms (n=36) were designed to be used at home by patients, and this was usually done in conjunction with their ongoing health care plans. To support at-home interventions, some platforms incorporated automated SMS text messaging, including conversational agents and trigger SMS text messaging [45,55]. The key focus of platforms that were exclusively home-based was self-management through input and self-monitoring of health data. Of the 36 interventions, 13 designed to be used in the home were also implemented in primary care settings, including hospitals. The digital platforms, which contained a patient-facing mobile health (mHealth) app, primarily focused on self-management of the symptoms of a chronic condition. The data collected by an mHealth app is usually passed to a practitioner-facing portal, where the practitioner can view patient data and adjust care plans. In some cases, these digital platforms supported bidirectional communication, allowing the practitioner to directly communicate with patients through the platform [32,35,45,53,60,61]. A few platforms also supported the booking of appointments with relevant health care providers [33,48,49] or sent reminders about upcoming appointments [43].

Because a high number of digital platforms are tailored to support self-management (n=44), the majority of platforms were primarily managed by the patients themselves. However, where the digital platforms were multimodal, consisting of a smartphone app and a web-based application, the smartphone apps tended to be managed by the patients, while the web-based applications were used by health care providers to review and manage patient information through an online portal.

XSL•FO
RenderX

Principles Used in the Design and Development of the Digital Platforms

In total, 25 studies (36%) used some form of co-design, consultative, or user-centered approach to developing the digital platforms. Furthermore, 11 studies focused on platforms that either had already existed or were adapted from existing models. Co-design approaches varied and included expert consultation with health care providers and user-centered iterative approaches, which sought feedback from patients throughout an intervention period.

Most of the platforms were developed by using surveys or focus groups to gather information and determine requirements. Often, these surveys and focus groups lacked the full participation of end users. In some studies, the iterative development approach also included gaining input from caregivers, patients, and related stakeholders. For example, the iMHere 2.0 [35] system was iteratively designed, developed, and evaluated with patients involved at all stages. Other studies that discussed the development of digital platforms did not follow or report on the iterative development process. One platform [62] was not accepted as it failed to meet the needs of the patients. Studies adopting an iterative approach tended to make changes to the digital intervention throughout the testing period in response to feedback, whereas those using professional consultations at the beginning and end of the process did not make changes during the testing period. Platforms primarily seeking to influence behavior change tended to be developed using more theoretically based approaches, such as behavior change models and social cognitive theory [54,55].

Outcomes of Studies Assessing the Digital Platforms

Outcomes of the digital platform assessments reported in the studies are grouped into 2 categories: outcomes of randomized controlled studies and other formative studies.

Outcomes of RCTs

The majority of the included studies demonstrated significant benefits and health outcomes, such as symptom reduction, better disease control, and improved quality of life. A smaller proportion of studies showed no significant effect on the primary outcomes, with improvements in only secondary measures that did not significantly impact the main health indicators. Some trials did not have the expected results, sometimes due to high dropout rates, lack of long-term effects, or problems with patient engagement (Table 4).

XSL•FO
RenderX

Table . Summary of the randomized controlled trial studies (n=27).

Study	Favorable outcomes	No effects	Name of the platform and target disease
Li et al [63]	Significant improvement in disease control at 6 months (*P*=.001).	At 12 months, no significant difference between groups (*P*=.90).	A smart system of disease management; rheumatoid arthritis
Goulding et al [64]	Decreased relapse risk in the low-risk group (*P*=.02), improved depressive symptoms (*P*=.02), and relational quality of life (*P*=.02).	There was no significant improvement in relapse risk for the high-risk group (*P*=.62).	A smartphone-based self-management intervention (LiveWell) app and website; bipolar disorder
Goodman and Locke [21]	Significant improvements in blood pressure were found, with systolic BP[a] decreasing from 140 to 134 mm Hg (*P*=.001) and diastolic BP decreasing from 78 to 74 mm Hg (*P*=.007)	No effect on the usual care group for HbA1c[b] (*P*=.19).	A mobile phone-delivered diabetes intervention; diabetes
Zhang et al [65]	Significant improvements in HbA1c (*P*<.05), diastolic BP (*P*<.05), and fasting plasma glucose (*P*<.05).	No effect on BMI.	A digital health technology to provide shared decision-making–informed dietary intervention; diabetes
Buis et al [66]	Both groups showed significant reductions in systolic BP (*P*<.001).	No significant differences between groups for BP or other outcomes (*P*=.99).	MI-BP, a culturally tailored multibehavior mobile health intervention; hypertension
Zhang et al [41]	Significant improvement in HbA1c (*P*<.001) and ABC[c] control rate (*P*=.025). ABC goals are HbA1c <7%, systolic BP/diastolic BP <140/80 mm Hg, and LDL-C[d] <2.6 mmol/L.	No changes in LDL-C or blood pressure (*P*=.95).	SMARTDiabetes app; diabetes
Tabernero et al [67]	Significant improvement in emotional well-being and self-efficacy for chronic disease management (*P*<.05).	No effect on anxiety and depression levels.	Psychological interventions delivered via mHealth[e] technology; chronic cardiac diseases
Hartch et al [52]	Significant improvements in medication adherence (Cohen *d*=−0.52, *P*=.014) and medication self-efficacy (Cohen *d*=0.43, *P*=.035).	No significant effect on medication knowledge or social support (*P*=.15).	Medisafe app; hypertension, diabetes, and asthma
Xu et al [24]	Significant improvement in anticoagulation knowledge, medication compliance, and patient satisfaction (*P*<.05).	No effect on bleeding or thrombotic events.	Alfalfa app; atrial fibrillation
Abel et al [22]	Significant reduction in systolic BP (*P*<.001), weight (*P*<.001), and physical activity (*P*=.018).	No significant difference in BP control between the treatment and control groups (*P*=.99). No sustained effect on BP control after 6 months.	Chronic disease self-management program; hypertension
Oh et al [34]	Significant reduction in body fat mass (*P*=.04) and HbA1c (*P*=.03) in the integrative mHealth group.	No significant changes in body weight, BMI, blood pressure, or HbA1c.	Integrative mHealth platform; hypertension and diabetes
Sittig et al [55]	Statistically significant improvements in self-efficacy (*P*=.008) and exercise (*P*=.01) in high and mid-users.	No significant differences across groups in overall health measures (ANOVA).	“capABILITY” app; diabetes
Patnaik et al [68]	Significant improvements in weight, BMI, waist circumference, and systolic blood pressure (*P*<.05).	No significant difference in metabolic equivalent (MET) levels (*P*=.54).	A mobile interactive platform: an Android-based application; diabetes
Jia et al [69]	Significant improvement in HbA1c (*P*<.001) and ABC control (*P*=.025).	No effect on hypoglycemia or weight gain (*P*=.95).	Graded the ROADMAP app and a website; diabetes

XSL•FO
RenderX

Study	Favorable outcomes	No effects	Name of the platform and target disease
Lear et al [49]	Fewer hospitalizations and in-hospital days (*P*<.05).	No significant reduction in hospitalization rates (*P*=.12).	An internet-based self-management and symptom monitoring program for diabetes, heart failure, ischemic heart disease, chronic kidney disease, or chronic obstructive pulmonary disease
Gray et al [50]	Participants set meaningful self-management goals.	No significant improvement in quality of life (*P*=.24).	Electronic patient-reported outcome mobile app and portal system; multiple chronic diseases such as arthritis, asthma, and hypertension
Kryger et al [27]	Significant reduction in urinary tract infections (*P*=.03).	No significant change in psychosocial outcomes.	iMHere mHealth system; spinal cord injury
Gong et al [45]	Significant improvement in quality of life (*P*=.04).	No significant change in HbA1c (*P*=.83).	My Diabetes Coach program, an app-based interactive embodied conversational agent; diabetes
Chhabra et al [29]	Significant improvement in disability (*P*<.001).	No significant change in pain.	Snapcare app; low back pain
Puig et al [28]	High patient satisfaction (85.2% find it useful, 91.4% would recommend it).	No significant change in quality of life or clinical outcomes.	+Approp app; HIV
Dale et al [54]	Significant improvements in medication adherence (*P*=.004).	No long-term effects on lifestyle changes (*P*=.13).	An mHealth-delivered comprehensive cardiac rehabilitation program called Text4Heart for coronary heart disease
Dorsch et al [31]	Improvement in Minnesota Living with Heart Failure Questionnaire at 6 weeks (*P*=.04).	No effect on self-reported HF management (*P*=.78).	ManageHF4Life app; heart failure
Velardo et al [33]	High compliance with self-monitoring (96% of symptom diaries completed).	No significant impact on patient outcomes or disease progression.	"Self-management and support programme (EDGE);" COPD[f]
Poppe et al [42]	Improvement in physical activity in some groups (*P*<.05).	Limited improvements in sitting time and moderate physical activity (*P*=.09 to *P*=.07).	MyPlan 2.0 comprises a website and an optional mobile app for diabetes
Schnall et al [37]	Significant improvement in 5 symptoms (anxiety, depression, neuropathy, fever or chills, and weight loss) (*P*<.05).	No effect on other symptoms.	mobile video information provider app; HIV
Morcillo-Muñoz et al [59]	Improvement in catastrophizing, rumination, and quality of life (*P*<.05).	No effect on magnification or satisfaction with health.	NO+Dolor (NO+ Pain) app; chronic pain
Park et al [60]	Improved self-care behavior (*P*=.01) and physical activity.	No improvement in self-efficacy for managing dyspnea. The number of steps per day did not significantly differ at 6 months.	Smartphone app-based self-management program; COPD

[a]BP: blood pressure.

[b]HbA1c: hemoglobin A1c.

[c]ABC: the "ABC" goals for type 2 diabetes management and stands for A1C (a measure of blood sugar), blood pressure, and cholesterol (specifically low-density lipoprotein cholesterol).

[d]LDL-C: low-density lipoprotein cholesterol.

[e]mHealth: mobile health.

[f]COPD: chronic obstructive pulmonary disease.

mHealth technologies are emerging as a promising solution in managing chronic conditions, offering patients a convenient and efficient way to monitor and improve their health. These randomized controlled studies (n=26) that reported favorable outcomes indicate that the impact of the mHealth intervention is evident in chronic disease management [21,22,24,27-29,31,33,34,37,41,42,45,49,50,52,54,55,59,63-69]. The clinical improvements were the more consistent outcome

XSL•FO
RenderX

across many of the included studies. Out of 26 RCTs, 30.8% (n=8) reported favorable improvements in key clinical health metrics, including BP [22,65,66,68], hemoglobin A1c (HbA1c) levels [21,34,41,65,69], weight [22,68], and blood glucose levels [65]. For example, the SMARTDiabetes trial [41] demonstrated better glycemic control in the intervention arm, improving HbA1c, BP, and low-density lipoprotein cholesterol control. These improvements were particularly notable in chronic diseases like diabetes and hypertension, suggesting that mHealth interventions can be effective in helping patients manage these conditions over time.

Beyond symptoms and clinical improvements, many studies found that these mHealth apps could create better self-management; 12 out of 26 studies (46%) demonstrated that patients became more involved in managing their health by tracking symptoms, medication adherence, and healthier behavior changes, such as increased physical activity [22,24,31,33,37,41,42,45,52,55,67,68]. For example, in the bipolar disorder study [64], participants using the smartphone-based self-management intervention showed reduced depressive symptoms and improved relational quality of life. This underlines the fact that mHealth tools are instrumental in improving clinical outcomes and empowering patients to take greater control of their conditions.

A common theme from these studies is increased patient engagement, which is favorable in 14 (54%) of these studies [22,24,31,33,37,41,42,45,52,54,55,63,67,68]. Most participants shared that engaging in mobile apps to track symptoms and remind themselves about their medication and educational content increased patients' interest in their health. In the study conducted on rheumatoid arthritis, it was observed that patients who were exposed to the mobile app had better control of their disease, while the smart system of disease management group [63] had a higher rate of patients with controlled disease than the control group, at 71% versus 64.5%.

Besides engagement, user satisfaction also appeared consistently high across the positive-outcome studies. Specifically, participants expressed appreciation for the ease of use in navigating the interfaces, personalized feedback, and easy access to health care information in studies by Xu et al [24] and Puig et al [28]. For example, in the study on the Alfalfa App [24], there was improved medication adherence among patients, $P<.001$, and a very high satisfaction regarding the app's utility in managing anticoagulation therapy.

Another important insight from the positive studies is the reduction in disease symptoms, reported in 5 of the 26 RCTs (19.2%) [27,31,37,63,64]. Indeed, many studies reported significant improvements in specific symptoms such as pain, depression, anxiety, and fatigue. This was especially true in conditions like HIV, where the mobile video information provider app [37] helped alleviate symptoms, such as neuropathy, anxiety, and depression, while also increasing medication adherence.

mHealth interventions also helped most patients improve their overall quality of life. Many studies have shown evidence for the above fact as the common resultant factor in the case of chronic diseases. Statistical improvements in scores over health-related quality of life were noticed to be significantly higher among app users than control subjects in the "My Diabetes Coach" study [45]; thus, providing evidence that such tools are effective in managing not only the clinical symptom improvement but also in emotional function and life satisfaction improvement.

In addition to clinical outcomes and user satisfaction, these mHealth interventions brought positive behavioral changes. Many studies reported that participants became more physically active, followed exercise routines more consistently, and had healthier dietary habits (11/26, 42%). For example, in a diabetes study [68], systolic BP, body fat, and BMI decreased significantly ($P<.001$) among the intervention group, thus indicating the effectiveness of mobile apps in bringing about healthier lifestyle changes.

Finally, 1 study [59] examined the cost-effectiveness of mHealth interventions. It noted that these tools are effective and economically viable, presenting affordable solutions for managing chronic diseases, especially in resource-constrained settings. For example, the chronic pain therapy study [59] established that app-based mobile treatments for pain management were effective and cost-effective, incorporating them into existing treatment plans.

Outcomes of the Formative Studies

The following sections present the findings of the formative studies (n=42) [20,23,25,26,30,32,35,36,38-40,43,44,46-48,51,53,56-58,61,62,70-88]. The most frequent outcomes are patient engagement and satisfaction, but also clinical improvements and self-management behavior are the focus of many studies. There are some usability issues and challenges reported in the studies (Table 5).

XSL•FO
RenderX

Table . Summary of the outcomes and challenges mentioned in the formative studies (n=42).

Outcomes (favorable, no effect, and challenges)	Explanation	Frequency
Favorable outcomes		
Patient engagement and satisfaction (eg, ease of use and positive feedback) [30,32,38-40,44,46,51,58,61,70-78,88]	Most of the studies reported a high level of patient engagement and satisfaction, naming ease of use and personalized feedback as major advantages. Many studies have shown that user-friendly digital health tools improved adherence to health regimens and helped patients manage their diseases more effectively.	22
Clinical improvements (eg, pain reduction, improved blood pressure, and LDL-C[a] levels) [25,32,39,40,43,71,79,80]	These studies reported various clinical improvements, including pain reduction, blood pressure control, improvement in LDL-C level, and overall disease management. Key results included highlighting digital health tools' potential to enhance self-management, support patients to achieve clinical goals, and improve overall health.	8
Self-management behavior (eg, adherence to medication and lifestyle changes) [32,38-40,43,44,46,51,53,56,61,70-73,75,78,81,82]	Overall, digital health tools across these studies helped improve self-management behaviors in medication adherence, lifestyle changes, and engagement in physical activity. The most significant improvements were seen in the management of chronic diseases like COPD[b], diabetes, hypertension, and heart failure.	19
Remote monitoring and resource usage (eg, reduction in hospital visits and health care costs) [25,32,43,51,71,83]	These studies highlight how remote monitoring systems can avert hospital admissions and reduce health costs by helping patients manage their conditions at home and, in real-time, provide the clinician with timely interventions. These findings were seen in heart failure, asthma, COPD management, hypertension, and ankylosing spondylitis, showing how digital health tools could improve clinical outcomes and optimize the use of resources within health care systems.	6
No effect		
No effect on disease management (eg, no improvement in symptom control or disease management) [30,38,48,70,79,84]	Some studies did not find any significant improvement in disease management, especially when the digital health tools did not sufficiently help engage patients or when the patients had barriers such as a lack of interest or limited use of digital health tools. In some cases, the personalization or customization of the tools was insufficient, leading to low effectiveness in managing the conditions.	6
Challenges encountered		
Usability issues (eg, technical issues and user disengagement) [32,35,38,39,61,72,88]	The most common usability issues reported across studies included technical problems, such as device inaccuracies, data syncing issues, and interface complexity. User disengagement was also another common challenge in many instances due to a lack of motivation, the tediousness of the process, and issues relating to poor integration into existing healthcare workflows.	7

XSL•FO
RenderX

Outcomes (favorable, no effect, and challenges)	Explanation	Frequency
Challenges in provider integration (eg, issues with workflows and data sharing) [32,38,43,44,61,71,77,78,81,83]	Provider integration challenges were highly reported in many studies, especially on integrating digital health tools with clinical workflows and the sharing of data between patients and providers. In most cases, the difficulty in adopting digital health tools in routine clinical care was cited as a barrier to clinical decision-making, with issues such as data synchronization and interoperability, assuring that providers can use the data collected remotely efficiently.	10
Privacy and data security concerns (eg, data interoperability and concerns about privacy) [32,38,43,44,61,71,76,81,85]	Privacy and data security issues were consistently identified in the reviews, particularly regarding transmission, storage, and interoperability with other healthcare information systems. There were issues of patient consent, data sharing, and following regulatory policies such as the Health Insurance Portability and Accountability Act (HIPAA). A concern was raised about protecting sensitive health information from unauthorized access.	9
Lack of participation (eg, limited use due to lack of time, motivation, or technical issues) [32,38,43,44,46,51,56,61,70-75,78,81,82]	Several of these studies repeatedly mentioned problems of non-participation for which technical issues, such as malfunction of a device and connectivity problems, together with a lack of motivation, were major reasons for dropouts and inconsistent use of digital health tools. Time constraints were also a significant factor in disengagement, as the patients struggled to integrate such tools into daily life. Personalization and support appeared very pivotal for long-term engagement.	17

[a]LDL-C: low-density lipoprotein cholesterol.

[b]COPD: chronic obstructive pulmonary disease.

One of the significant trends observed across the formative studies is the potential of digital health tools to improve the management of chronic conditions such as COPD, hypertension, diabetes, and heart failure. A significant percentage of studies (17/42, 40%) reported favorable clinical and self-management outcomes, including improved BP control, weight management, and self-efficacy. For example, HbA1c was reduced by approximately 0.79% [57], and patients suffering from COPD on the Wellinks mHealth platform showed improved symptoms and quality of life [71]. Similarly, 1 study [36] demonstrated that digital management tools for hypertension reduced systolic BP/diastolic BP by 14/5 mm Hg. These results suggest that digital tools can offer tangible improvements in managing chronic diseases, particularly when integrated with traditional care methods.

The second trend to emerge from these studies is the centrality of user-centered design in the overall success of digital health interventions. A substantial proportion of studies (13/42, 31%) indicated that designing the digital tool for patient preferences and needs enhances engagement and satisfaction. For example, the iMHere 2.0 system, which offers personalized app modules to support various self-management tasks, was praised for its customizability and ability to keep patients engaged [61]. Similarly, the Wellinks mHealth platform for COPD was well-received due to its ease of use and support in daily disease management [71]. The above findings point out the importance of developing digital tools that are not only functional but can also be tailored according to the needs of the patients in improving usability and increasing engagement.

Whereas the initial engagement and clinical outcomes from the studies were generally good, the longer-term health outcomes tend to be more mixed. The main challenges with the long-term maintenance of digital interventions were mentioned in 12% (5/42) of the selected studies, where initially engaged patients stopped using the tools due to various barriers, such as motivational issues, technical problems, or difficulties in maintaining regular use. For example, some patients in the studies of COPD management dropped off after an initial burst of engagement due to difficulties in integrating the technology into their daily routines [71]. This points to the need for continuous engagement strategies and more user-friendly designs to maintain patient involvement over the long term [26].

The studies also indicate an increased awareness that, in treating chronic conditions of a complex nature, it is more often than not challenging to rely on one-size-fits-all approaches. Some studies with generalized tools showed positive outcomes; others (5/42, 12%) indicated that tools must be customized to meet individual patient needs. For instance, a digital health tool for diabetes showed promising results in improving medication adherence but struggled with user engagement in the long term, particularly among patients who required more personalized support [39]. Many studies emphasized the need for adaptive technologies that can adjust to the changing needs of patients

XSL•FO
RenderX

and those that can integrate seamlessly into existing health care systems.

Despite promising results, the need for further research and development is a constant note in several studies, as shown by 9.52% (4/42). These studies have shown that while the performance of digital tools has a promising side, there are serious gaps in personalization, scalability, and integration into health systems [60,62]. Issues of provider workflow problems [38], data interoperability [38], and assurance about patient privacy concerns were considered the most important to resolve to make them more acceptable.

Discussion

Principal Findings

This review highlights the growing potential of digital platforms in enabling both self-management (patient-facing monitoring, decision support, and behavior change) and provider-led management (remote monitoring dashboards, care coordination, and clinical decision support). The digital platforms were primarily designed for use at home and complement patients' routine health care practices seamlessly, giving major importance to self-monitoring, personalization, and motivational aspects such as rewards. While self-management is valued, our findings suggest that platforms without embedded communication facilities with health care professionals or social support may limit user interaction and effectiveness. The fact that collaborative care and self-management co-occurred in the included studies underscores the value of hybrid interventions that combine patient autonomy with professional oversight.

A significant proportion of the studies reported the usage of co-design or user-centered design approaches as the best practice when developing health interventions. The lack of methodological details and theoretical underpinnings in some studies hinders clear conclusions regarding their effectiveness. While the review found a range of features, there was limited evidence on which features are most effective at facilitating long-term engagement. Most studies were feasibility or pilot studies with brief follow-up periods and small or nonrandomized samples. As such, large-scale assessments of the effectiveness of platforms, particularly for long-term health outcomes, are lacking. Platforms addressing single chronic diseases were more likely to exhibit improved outcomes, whereas those addressing multimorbidity encountered more usability and implementation issues, underscoring the need for more personalized designs to meet complex health needs.

One of the significant barriers to the long-term adoption of digital health solutions was user disengagement. Several longitudinal studies reported high dropout rates due to a lack of motivation or technical issues, highlighting the importance of adaptive and personalized engagement strategies that are unobtrusive and seamlessly integrate digital health platforms into everyday life.

In the following sections, we elaborate on the principal findings and compare and contrast them with relevant literature.

Characteristics of Digital Platforms Supporting the Management of Chronic Diseases

Our findings show that most of the interventions focused on self-management of chronic diseases. We also found that the vast majority of interventions were designed to be used at home by participants, usually in conjunction with their ongoing health care plans. Self-management is critical in managing chronic disease [94,95], and new digital platforms heavily target this aim. According to the literature, chronic disease can be managed well by balancing traditional medical care with self-management [96]. Effective self-management requires optimal communication with health care teams [97]. Although self-management requires support from health care providers and caregivers, this review discovers that not all identified platforms offer such options. Similarly, it is necessary that digital platforms provide options for social support to maintain long-term engagement in self-managing disease [98]. Only a few platforms in our review had options for social support. Our findings suggest that features like self-tracking, customization, and rewards support users' engagement with digital platforms [12]. Although we have identified a range of features in the studied platforms, there is still a lack of evidence in the included studies as to which features are best for supporting long-term engagement. Therefore, more research is needed to investigate which features of digital platforms will best support long-term user engagement and motivation.

The findings from applying the NICE Evidence Standards Framework emphasize the necessity of a multifaceted and integrated approach to delivering effective digital health interventions. The significance of self-management reflects the growing emphasis on patient empowerment, enabling individuals to actively monitor, track, and manage their health. However, self-management alone is not always sufficient, as structured support systems improve engagement and adherence. The frequent co-occurrence of self-management with collaborative care (47 times) highlights the effectiveness of a hybrid model that combines patient autonomy with a form of clinical oversight. This integration ensures that while patients take a leading role in their health management, they are still supported by health care professionals who provide guidance and add to the rigor of the digital intervention. However, as noted earlier, some platforms lack direct communication with health care providers, which could reflect as a limiting factor in their effectiveness as a digital intervention. This comprehensive analysis highlights the interdependencies among digital intervention strategies, advocating for a cohesive, patient-centered approach. Self-management is most effective when complemented by collaborative care, education, and evidence-based tools, ensuring engagement, clinical effectiveness, and long-term sustainability in chronic disease management.

Principles and Frameworks Used in Designing and Developing the Platforms

We have observed that more than half of the studies reported some form of co-design, consultative, or user-centered approach to the development of digital platforms. This is a positive indication that health care interventions are co-developed to

XSL•FO
RenderX

meet the needs of the stakeholders, as suggested in the literature [99,100]. Overall, adopting a co-design approach might have contributed to positive outcomes. However, due to the lack of relevant data in those studies, we could not determine if there was any relationship between the co-design of the digital platforms and their effectiveness in managing chronic disease. A lack of theoretical basis in some of the included studies limits their reliability. Therefore, it is suggested that the future development of digital platforms needs solid theoretical support, and such support may well improve effectiveness and user engagement.

A range of behavior-change techniques was used in the included studies, generally informed by behavior-change models and social cognitive theory [101]. However, future studies should focus on measuring the effects of these design principles, determining the extent to which they contribute to the efficacy and continued use of the platforms. In addition, we identified 5 platforms that used gamification techniques to enhance user experience and embed learning principles. This suggests that platform designers considered the notion that gamification improves health behaviors, as reported in the literature [102].

Effectiveness and Efficacy of the Platforms

Most of the included studies were pilot in nature, focusing either on design, development, usability, uptake, or clinical utility. In many of these studies, not all features of the platforms were tested for effectiveness. One key reason for not testing all the features of a digital platform during the trial period (for example, study by Doyle et al [44]) was that the development was based on the initial success of the trial, and this was the focus of the majority of the included studies. This suggests that future work should identify those studies where the full potential of digital platforms is evaluated, so that more complete conclusions can be drawn about the platforms' effectiveness.

It is important to note that the majority of studies reporting improved health outcomes focused on a single or localized chronic disease. Conversely, those reporting no effects on health outcomes largely focused on multiple chronic diseases. This suggests that focused digital platforms have a higher likelihood of positive health outcomes than those targeting multiple chronic health issues. Further research is needed to investigate how digital platforms could successfully be designed and evaluated to manage multimorbidity (ie, the presence of 2 or more long-term health conditions). Furthermore, it is worth noting that the testing duration of the digital platforms was generally short in most studies. In some cases, the formal sample size was not calculated since the studies were feasibility or pilot studies. This highlights the importance of meticulously evaluating digital platforms with appropriate sample sizes in future studies to ensure the validity and reliability of the research.

Uptake of the Digital Health Interventions

A few of the included studies, which were conducted over a longer period (ie, 6 months or more), explicitly mentioned dropout rates and the causes of disengagement, including lack of motivation, technical issues, or health problems. For example, a study [86] aimed to understand the effects of long-term (eg, 12 months) smartphone-based self-monitoring in patients with lipid metabolism disorders reported that 43% (43/100) of patients never started using the app due to a lack of time, health problems, lack of motivation, and technical problems. Dropout due to technical issues (eg, poor wireless connection) was also reported in another study, where an mHealth system was developed for managing chronic conditions [61], where 1 patient dropped out of the study after the initial engagement. Another included study [39] mentioned that approximately 80% (400/500) of users used Mobiab (ForaCare Suisse AG) for managing diabetes for less than 1 week. Such a dropout may imply that the daily use requirements of the app were challenging for individuals to maintain. User fatigue is a likely issue where daily data entry or constant interaction with digital tools is required. As digital health interventions often require continuous self-monitoring and engagement—whether through tracking symptoms, inputting data, or responding to feedback—users may experience burnout, leading to disengagement. A high dropout rate (95/162, 58.6%) was mentioned in a study [66] where the effect of an mHealth self-monitoring intervention among black individuals with uncontrolled hypertension was tested for 1 year. While digital health tools show promise in improving short-term health outcomes, dropout rates, user fatigue, and sustainability remain substantial challenges. For digital health interventions to be successful in the long run, continuous engagement strategies and adaptive features must be prioritized to ensure users remain motivated and that interventions can be integrated smoothly into everyday life. However, further investigation is required to understand the dropout and disengagement with digital health interventions.

Digital literacies play a key role in the uptake of digital health interventions [103]. Researchers found that even though patients with low literacy may have access to technology, they may not be able to use it without any help [104]. Some studies were conducted with adults aged between early (20 - 39 y) [27,35,43,61] and middle adulthood (40 - 59 y) [30,34,40,55]. Chronic diseases are more common among older adults, who may have poor digital literacy and difficulty adopting new information technologies. In the included studies, no information was collected regarding the digital literacy levels of the participants. However, some of the included studies had information about income and education level that are linked to the participants' digital skills and health literacy. Low digital literacy may be a barrier to the adoption and engagement of digital platforms. Another limitation of the included studies is that not all studies reported parameters such as participants' skills, experience, or level of education, and these might have contributed to the infrequent use of the platform [28] or withdrawal from the study [42]. Some of the studies excluded patients who were unwilling to participate or could not meet the study requirements. Such nonparticipants may provide insights for the design of more user-centered platforms. Therefore, future studies should account for the varying digital literacies of different cohorts of platform users, as this may impact the overall feasibility of digital health interventions.

Participant confidentiality, including data security, was not widely discussed in the papers included in this review. Only 1 paper provided a detailed overview of the data security measures

XSL•FO
RenderX

of their platform [39]. Researchers have reported that health and fitness apps often violate users' privacy by not following existing guidelines and regulations [105,106]. Therefore, it is of the utmost importance to protect the health data that digital platforms gather. Ensuring the privacy and security of health data may contribute to the long-term uptake of digital health interventions.

Implications and Real-World Adoption

While most studies focus on the efficacy of digital health interventions in pilot settings, large-scale integration within already prevailing health care systems was not reported in any studies. However, some studies have promising steps toward integration within the current health care system. As an example, a mobile app called Mobiab [39] for diabetes management integrated with health care systems was partially successful, particularly where patient data could be automatically transmitted to clinicians through platforms like mobile apps. The study reported that the use of different software was an additional complication for the clinicians, as they already used some commercial software. Also, the researchers did not implement data into hospital information systems due to the lack of communication interface specifications. A hypertension management platform was developed that can be integrated into EHRs to facilitate real-time patient data analysis and effective decision-making [85]. However, the efficacy of the system is yet to be tested. Another application called "electronic patient-reported outcome mobile app and portal system" was developed for people with complex care needs. During the trial, the electronic patient-reported outcome system was not interfaced with other existing technology systems, that is, EMRs or other available platforms, but the system was designed in such a way that interoperability could be a possibility [50].

None of the studies explicitly discusses interoperability, clinician workload, or regulatory constraints, although these could hinder the implementation of digital health solutions in real-world health care settings [107,108]. These barriers are implicit in the included studies and require investigation in future studies. For example, one of our included studies [39] refers to the long-term involvement of clinicians in managing diabetes but does not discuss the potential impact it would have on clinician workload. Clinicians would be required to review the data generated by the app, resulting in an increased workload if the app is not providing actionable insights or is not well-integrated into clinical workflows. Research shows that the lack of seamless integration of digital health platforms into health care systems poses obstacles to broader adoption and implementation [109,110] and may hinder sustainability and scalability. In addition, regulatory restrictions on patient data protection (eg, General Data Protection Regulation and Health Insurance Portability and Accountability Act) are not discussed in the included studies, but these would be necessary to make the platform health care regulation compliant and legally viable for real-world adoption and implementation. Such concerns were echoed in a recent study, where authors highlighted the need for transparent data government policies to be implemented in order to meet regulatory requirements and address security concerns [107].

Many of the papers included in this review suggested that further research into the interventions was needed. Reasons for this included limited sample sizes, limited participant uptake, technical issues, and the need for further personalization of the platforms. Furthermore, while short-term health outcomes of the digital interventions were largely positive, long-term outcomes remain generally unknown. Thus, it was frequently suggested that strategies to maintain long-term use were required, along with further analysis of platform use. Additionally, it was noted in some instances that future studies should account for the varying digital literacies of different cohorts of users, as this may impact the overall feasibility of digital health interventions. Finally, where studies focused on a single or localized chronic health issue, it was broadly concluded that platform design could easily be replicated to address other health issues. However, interventions that sought to address multiple chronic diseases had a higher incidence of technical issues or problems with usability and feasibility. This could be because those studies have methodological challenges, such as higher sample sizes and complex sampling frames to measure the required outcomes. However, such challenges were not reported in those studies, and this suggests that further research is needed to enable digital health interventions to effectively address multiple chronic health conditions.

Limitations

There are several limitations to our study. Our objective was to explore the broader perspective of chronic diseases rather than focusing on individual chronic diseases. Therefore, the search strategy used was general terms related to chronic disease management rather than specific conditions such as diabetes, hypertension, cardiovascular disease, or COPD. While this approach allowed us to draw on a wide range of interventions, it may have inadvertently omitted condition-specific intervention studies using narrowly defined keywords. This could have affected the completeness of the review in 2 ways. First, we might have missed highly specialized interventions tailored to the unique management needs of particular chronic diseases. Second, excluding disease-specific search terms could have led to underrepresenting certain populations or technologies in specific disease domains. Nonetheless, studies in this review constitute an important and representative sample of the current digital intervention landscape for chronic disease management. These studies encompass a diverse range of technologies, user groups, and interventions to identify trends, design issues, and gaps in digital interventions for chronic disease management. We acknowledge that a more targeted search strategy, perhaps in a future scoping or systematic review, would offer more information about condition-specific digital health innovations and their impact.

Our search strategy did not include the exact keyword "self-management" or corresponding controlled-vocabulary terms. Although the broader "management" concept and citation chasing were used to capture patient-led self-management studies indirectly, records that exclusively use "self-management" terminology may have been missed. A future update should incorporate a self-management term cluster (eg, "self-management," "self care," and "self-monitor") and mapped controlled vocabulary to improve sensitivity.

XSL•FO
RenderX

In this review, we only included articles that were published in English. For the included studies, we have primarily reported the results qualitatively. Where available, we reported the frequency of outcomes but were unable to capture the effect size due to the variability of the studies. Furthermore, not all studies measured the impact of digital platforms; several of them instead measured the usability and acceptability of those platforms. Therefore, we could not compare all the outcomes, and in some cases, the outcomes were inconclusive due to the preliminary nature of the studies. Finally, the digital platforms included in the studies were designed for diverse users with varying degrees of digital literacy. However, we could not analyze how the digital literacies of the participants contributed to outcomes because such data were not reported in the studies.

Conclusions

This study provides a comprehensive overview of digital platforms for managing chronic diseases, delineating features for self-management versus provider-led management. Overall, the vast majority of papers in this review concluded that digital health interventions can be beneficial in managing chronic health issues. They also indicated that the adoption of such methods in combination with regular clinical care has the potential to improve health outcomes, support self-management, and support communication between patients and health care providers. However, challenges remain in long-term engagement, overcoming technological barriers, and integrating these tools into existing workflows in health care. The effectiveness and acceptance of digital health interventions vary based on patient characteristics, such as age, health literacy, and the capacity for intervention tailoring. Success will be contingent on interventions that can fulfill specific patient needs through user-centered, tailored engagement while being effortless to use and integrated seamlessly within the health care ecosystem. These tools, therefore, require further research for their full development so that they are adaptable, scalable, and meet the diverse needs of patients with chronic conditions. More research is needed to further develop these tools for wider acceptance and improving engagement.

Funding

This work was supported jointly by Peter MacCallum Cancer CentreCenter, Swinburne University of Technology, and the Digital Health CRC Limited (“DHCRC”). DHCRC is funded under the Commonwealth’s Cooperative Research Centres (CRC) Program.

Authors' Contributions

Conceptualization: AT, AAM, SB
Formal analysis: AAM, CL, SJ, MS, ST, ARMF
Funding acquisition: AT, PPJ
Investigation: AAM, CL, SJ
Methodology: AAM, AT, NW, SB
Validation: AAM, SJ, CL
Visualization: AAM, CL, SJ
Writing – original draft: AAM, CL, SJ, PPJ
Writing – review and editing: All authors
All authors reviewed the final manuscript and approved the submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Database search outcome.
[DOCX File, 20 KB - mhealth_v14i1e63742_app1.docx]

Multimedia Appendix 2
Chronic diseases reported in the included studies (n=83).
[PNG File, 77 KB - mhealth_v14i1e63742_app2.png]

Multimedia Appendix 3
Digital intervention strategies, features of the digital platforms, and co-occurrence matrix of the digital intervention strategies.
[DOCX File, 50 KB - mhealth_v14i1e63742_app3.docx]

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 86 KB - mhealth_v14i1e63742_app4.docx]

XSL•FO
RenderX

References

1. Chronic disease. US Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/chronic-disease/about/index.html#:~:text=Chronic%20diseases%20are%20defined%20broadly,of%20daily%20living%20or%20both [accessed 2024-11-11]
2. Noncommunicable diseases. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2023-11-13]
3. The top 10 causes of death. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death [accessed 2023-11-14]
4. Liddy C, Blazkho V, Mill K. Challenges of self-management when living with multiple chronic conditions: systematic review of the qualitative literature. Can Fam Physician 2014 Dec;60(12):1123-1133. [Medline: 25642490]
5. Hajat C, Stein E. The global burden of multiple chronic conditions: a narrative review. Prev Med Rep 2018 Dec;12:284-293. [doi: 10.1016/j.pmedr.2018.10.008] [Medline: 30406006]
6. Murray E, Hekler EB, Andersson G, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851. [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
7. Samal L, Fu HN, Camara DS, Wang J, Bierman AS, Dorr DA. Health information technology to improve care for people with multiple chronic conditions. Health Serv Res 2021 Oct;56 Suppl 1(Suppl 1):1006-1036. [doi: 10.1111/1475-6773.13860] [Medline: 34363220]
8. Anekwe TD, Rahkovsky I. Self-management: a comprehensive approach to management of chronic conditions. Am J Public Health 2018 Dec;108(S6):S430-S436. [doi: 10.2105/AJPH.2014.302041r]
9. Sheng Y, Doyle J, Bond R, Jaiswal R, Gavin S, Dinsmore J. Home-based digital health technologies for older adults to self-manage multiple chronic conditions: a data-informed analysis of user engagement from a longitudinal trial. Digit Health 2022;8:20552076221125957. [doi: 10.1177/20552076221125957] [Medline: 36171962]
10. What is digital health technology and what can it do for me? National Institute of Health and Care Research. 2022. URL: https://evidence.nihr.ac.uk/collection/what-is-digital-health-technology/ [accessed 2025-12-14]
11. Bashi N, Fatehi F, Mosadeghi-Nik M, Askari MS, Karunanithi M. Digital health interventions for chronic diseases: a scoping review of evaluation frameworks. BMJ Health Care Inform 2020 Mar;27(1):e100066. [doi: 10.1136/bmjhci-2019-100066] [Medline: 32156751]
12. Nittas V, Zecca C, Kamm CP, Kuhle J, Chan A, von Wyl V. Digital health for chronic disease management: an exploratory method to investigating technology adoption potential. PLoS ONE 2023;18(4):e0284477. [doi: 10.1371/journal.pone.0284477] [Medline: 37053272]
13. Digital health platform handbook: building a digital information infrastructure (infostructure) for health. World Health Organization. 2020. URL: https://www.who.int/publications/i/item/9789240013728 [accessed 2025-12-14]
14. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
15. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143. [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
16. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
17. Covidence. URL: https://www.covidence.org/ [accessed 2025-12-05]
18. Gagnon MP, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
19. Lupton D. Health promotion in the digital era: a critical commentary. Health Promot Int 2015 Mar;30(1):174-183. [doi: 10.1093/heapro/dau091] [Medline: 25320120]
20. Genberg EM, Viitanen HT, Mäkelä MJ, Kautiainen HJ, Kauppi PM. Impact of a digital web-based asthma platform, a real-life study. BMC Pulm Med 2023 May 12;23(1):165. [doi: 10.1186/s12890-023-02467-8] [Medline: 37173716]
21. Goodman KS, Locke E. Health impact of a mobile-delivered diabetes intervention to control blood pressure in older adults. AJPM Focus 2024 Aug;3(4):100244. [doi: 10.1016/j.focus.2024.100244] [Medline: 39034932]
22. Abel WM, Efird JT, Crane PB, Ferdinand KC, Foy CG, DeHaven MJ. Use of coaching and technology to improve blood pressure control in Black women with hypertension: pilot randomized controlled trial study. J of Clinical Hypertension 2023 Jan;25(1):95-105. [doi: 10.1111/jch.14617]
23. Chrysohoou C, Tsamadias V, Kariori M, et al. Evaluation of the usability of the digital platform navigator KardioUp for the journey of patients with chronic heart failure. Hellenic J Cardiol 2024;75:32-40. [doi: 10.1016/j.hjc.2023.06.001] [Medline: 37295668]
24. Xu W, Huang X, Lin Q, et al. Application of Alfalfa app in the management of oral anticoagulation in patients with atrial fibrillation: a multicenter randomized controlled trial. BMC Med Inform Decis Mak 2024 Oct 9;24(1):294. [doi: 10.1186/s12911-024-02701-1] [Medline: 39385171]
25. Reid D, Mehta J, Anis K, Mehta S. Impact of remote patient monitoring platform on patients with moderate to severe persistent asthma: observational study. JMIR Form Res 2023 Dec 28;7:e51065. [doi: 10.2196/51065] [Medline: 38153783]

XSL•FO
RenderX

26. Alharbey R, Chatterjee S. An mHealth assistive system "MyLung" to empower patients with chronic obstructive pulmonary disease: design science research. JMIR Form Res 2019 Mar 19;3(1):e12489. [doi: 10.2196/12489] [Medline: 30888329]
27. Kryger MA, Crytzer TM, Fairman A, et al. The effect of the interactive mobile health and rehabilitation system on health and psychosocial outcomes in spinal cord injury: randomized controlled trial. J Med Internet Res 2019 Aug 28;21(8):e14305. [doi: 10.2196/14305] [Medline: 31464189]
28. Puig J, Echeverría P, Lluch T, et al. A specific mobile health application for older HIV-infected patients: usability and patient's satisfaction. Telemed J E Health 2021 Apr;27(4):432-440. [doi: 10.1089/tmj.2020.0098] [Medline: 32667858]
29. Chhabra HS, Sharma S, Verma S. Smartphone app in self-management of chronic low back pain: a randomized controlled trial. Eur Spine J 2018 Nov;27(11):2862-2874. [doi: 10.1007/s00586-018-5788-5] [Medline: 30324496]
30. Evans J, Papadopoulos A, Silvers CT, et al. Remote health monitoring for older adults and those with heart failure: adherence and system usability. Telemed J E Health 2016 Jun;22(6):480-488. [doi: 10.1089/tmj.2015.0140] [Medline: 26540369]
31. Dorsch MP, Farris KB, Rowell BE, Hummel SL, Koelling TM. The effects of the ManageHF4Life mobile app on patients with chronic heart failure: randomized controlled trial. JMIR Mhealth Uhealth 2021 Dec 7;9(12):e26185. [doi: 10.2196/26185] [Medline: 34878990]
32. Guo X, Gu X, Jiang J, et al. A hospital-community-family-based telehealth program for patients with chronic heart failure: single-arm, prospective feasibility study. JMIR Mhealth Uhealth 2019 Dec 13;7(12):e13229. [doi: 10.2196/13229] [Medline: 31833835]
33. Velardo C, Shah SA, Gibson O, et al. Digital health system for personalised COPD long-term management. BMC Med Inform Decis Mak 2017 Feb 20;17(1):19. [doi: 10.1186/s12911-017-0414-8] [Medline: 28219430]
34. Oh SW, Kim KK, Kim SS, Park SK, Park S. Effect of an integrative mobile health intervention in patients with hypertension and diabetes: crossover study. JMIR Mhealth Uhealth 2022 Jan 11;10(1):e27192. [doi: 10.2196/27192] [Medline: 35014961]
35. Setiawan IMA, Zhou L, Alfikri Z, et al. An adaptive mobile health system to support self-management for persons with chronic conditions and disabilities: usability and feasibility studies. JMIR Form Res 2019 Apr 25;3(2):e12982. [doi: 10.2196/12982] [Medline: 31021324]
36. Milani RV, Lavie CJ, Bober RM, Milani AR, Ventura HO. Improving hypertension control and patient engagement using digital tools. Am J Med 2017 Jan;130(1):14-20. [doi: 10.1016/j.amjmed.2016.07.029] [Medline: 27591179]
37. Schnall R, Cho H, Mangone A, Pichon A, Jia H. Mobile health technology for improving symptom management in low income persons living with HIV. AIDS Behav 2018 Oct;22(10):3373-3383. [doi: 10.1007/s10461-017-2014-0] [Medline: 29299790]
38. Gill A, Khan AI, Hans PK, Kuluski K, Cott C. The electronic patient reported outcome tool: testing usability and feasibility of a mobile app and portal to support care for patients with complex chronic disease and disability in primary care settings. JMIR Mhealth Uhealth 2016 Jun 2;4(2):e58. [doi: 10.2196/mhealth.5331] [Medline: 27256035]
39. Burda V, Mráz M, Schneider J, Novák D. Managing diabetes using Mobiab: long-term case study of the impact of a mobile app on self-management. JMIR Diabetes 2022 Apr 20;7(2):e36675. [doi: 10.2196/36675] [Medline: 35442201]
40. Bailey JF, Agarwal V, Zheng P, et al. Digital care for chronic musculoskeletal pain: 10,000 participant longitudinal cohort study. J Med Internet Res 2020 May 11;22(5):e18250. [doi: 10.2196/18250] [Medline: 32208358]
41. Zhang P, Tao X, Ma Y, et al. Improving the management of type 2 diabetes in China using a multifaceted digital health intervention in primary health care: the SMARTDiabetes cluster randomised controlled trial. Lancet Reg Health West Pac 2024 Aug;49:101130. [doi: 10.1016/j.lanwpc.2024.101130] [Medline: 39056088]
42. Poppe L, De Bourdeaudhuij I, Verloigne M, et al. Efficacy of a self-regulation-based electronic and mobile health intervention targeting an active lifestyle in adults having type 2 diabetes and in adults aged 50 years or older: two randomized controlled trials. J Med Internet Res 2019 Aug 2;21(8):e13363. [doi: 10.2196/13363] [Medline: 31376274]
43. Ji X, Wang Y, Ma Y, et al. Improvement of disease management and cost effectiveness in Chinese patients with ankylosing spondylitis using a smart-phone management system: a prospective cohort study. Biomed Res Int 2019;2019:2171475. [doi: 10.1155/2019/2171475] [Medline: 30931322]
44. Doyle J, Murphy E, Gavin S, et al. A digital platform to support self-management of multiple chronic conditions (ProACT): findings in relation to engagement during a one-year proof-of-concept trial. J Med Internet Res 2021 Dec 15;23(12):e22672. [doi: 10.2196/22672] [Medline: 34914612]
45. Gong E, Baptista S, Russell A, et al. My Diabetes Coach, a mobile app-based interactive conversational agent to support type 2 diabetes self-management: randomized effectiveness-implementation trial. J Med Internet Res 2020 Nov 5;22(11):e20322. [doi: 10.2196/20322] [Medline: 33151154]
46. Babington-Ashaye A, de Moerloose P, Diop S, Geissbuhler A. Design, development and usability of an educational AI chatbot for people with haemophilia in Senegal. Haemophilia 2023 Jul;29(4):1063-1073. [doi: 10.1111/hae.14815] [Medline: 37347648]
47. Barbaric A, Munteanu C, Ross H, Cafazzo JA. Design of a patient voice app experience for heart failure management: usability study. JMIR Form Res 2022 Dec 6;6(12):e41628. [doi: 10.2196/41628] [Medline: 36472895]
48. Breckner A, Litke N, Göbl L, et al. Effects and processes of an mHealth intervention for the management of chronic diseases: prospective observational study. JMIR Form Res 2022 Aug 25;6(8):e34786. [doi: 10.2196/34786] [Medline: 36006666]

XSL•FO
RenderX

49. Lear SA, Norena M, Banner D, et al. Assessment of an interactive digital health-based self-management program to reduce hospitalizations among patients with multiple chronic diseases: a randomized clinical trial. JAMA Netw Open 2021 Dec 1;4(12):e2140591. [doi: 10.1001/jamanetworkopen.2021.40591] [Medline: 34962560]
50. Steele Gray C, Chau E, Tahsin F, et al. Assessing the implementation and effectiveness of the electronic patient-reported outcome tool for older adults with complex care needs: mixed methods study. J Med Internet Res 2021 Dec 2;23(12):e29071. [doi: 10.2196/29071] [Medline: 34860675]
51. Rossetto F, Borgnis F, Isernia S, et al. System integrated digital empowering and telerehabilitation to promote patient activation and well-being in chronic disabilities: a usability and acceptability study. Front Public Health 2023;11:1154481. [doi: 10.3389/fpubh.2023.1154481] [Medline: 37250091]
52. Hartch CE, Dietrich MS, Lancaster BJ, Stolldorf DP, Mulvaney SA. Effects of a medication adherence app among medically underserved adults with chronic illness: a randomized controlled trial. J Behav Med 2024 Jun;47(3):389-404. [doi: 10.1007/s10865-023-00446-2] [Medline: 38127174]
53. Salari R, R Niakan Kalhori S, GhaziSaeedi M, Jeddi M, Nazari M, Fatehi F. Mobile-based and cloud-based system for self-management of people with type 2 diabetes: development and usability evaluation. J Med Internet Res 2021 Jun 2;23(6):e18167. [doi: 10.2196/18167] [Medline: 34076579]
54. Pfaeffli Dale L, Whittaker R, Jiang Y, Stewart R, Rolleston A, Maddison R. Text message and internet support for coronary heart disease self-management: results from the Text4Heart randomized controlled trial. J Med Internet Res 2015 Oct 21;17(10):e237. [doi: 10.2196/jmir.4944] [Medline: 26490012]
55. Sittig S, Wang J, Iyengar S, Myneni S, Franklin A. Incorporating behavioral trigger messages into a mobile health app for chronic disease management: randomized clinical feasibility trial in diabetes. JMIR Mhealth Uhealth 2020 Mar 16;8(3):e15927. [doi: 10.2196/15927] [Medline: 32175908]
56. Nabovati E, Rangraz Jeddi F, Tabatabaeizadeh SM, Hamidi R, Sharif R. Design, development, and usability evaluation of a smartphone-based application for nutrition management in patients with type II diabetes. J Diabetes Metab Disord 2023;22(1):315-323. [doi: 10.1007/s40200-022-01140-x]
57. NOVIANI L, DIANTINI A, SUBARNAS A. Implementation of mobile application for diabetes self-management on self efficacy and clinical outcome in Indonesia. International Journal of Pharmaceutical Research (09752366) 2020 [FREE Full text]
58. Duan H, Wang Z, Ji Y, et al. Using goal-directed design to create a mobile health app to improve patient compliance with hypertension self-management: development and deployment. JMIR Mhealth Uhealth 2020 Feb 25;8(2):e14466. [doi: 10.2196/14466] [Medline: 32130161]
59. Morcillo-Muñoz Y, Sánchez-Guarnido AJ, Calzón-Fernández S, Baena-Parejo I. Multimodal chronic pain therapy for adults via smartphone: randomized controlled clinical trial. J Med Internet Res 2022 May 11;24(5):e36114. [doi: 10.2196/36114] [Medline: 35373776]
60. Park SK, Bang CH, Lee SH. Evaluating the effect of a smartphone app-based self-management program for people with COPD: a randomized controlled trial. Appl Nurs Res 2020 Apr;52:151231. [doi: 10.1016/j.apnr.2020.151231] [Medline: 31955942]
61. Parmanto B, Pramana G, Yu DX, Fairman AD, Dicianno BE, McCue MP. iMHere: a novel mHealth system for supporting self-care in management of complex and chronic conditions. JMIR Mhealth Uhealth 2013 Jul 11;1(2):e10. [doi: 10.2196/mhealth.2391] [Medline: 25100682]
62. Brandl LC, Liebram C, Schramm W, Pobiruchin M. A German smartphone-based self-management tool for psoriasis: community-driven development and evaluation of quality-of-life effects. JMIR Form Res 2022 Jul 7;6(7):e32593. [doi: 10.2196/32593] [Medline: 35797109]
63. Li C, Huang J, Wu H, et al. Management of rheumatoid arthritis with a digital health application: a multicenter, pragmatic randomized clinical trial. JAMA Netw Open 2023 Apr 3;6(4):e238343. [doi: 10.1001/jamanetworkopen.2023.8343] [Medline: 37058302]
64. Goulding EH, Dopke CA, Rossom R, Jonathan G, Mohr D, Kwasny MJ. Effects of a smartphone-based self-management intervention for individuals with bipolar disorder on relapse, symptom burden, and quality of life: a randomized clinical trial. JAMA Psychiatry 2023 Feb 1;80(2):109-118. [doi: 10.1001/jamapsychiatry.2022.4304] [Medline: 36542401]
65. Zhang W, Zhu X, Que X, Zhang X. The effects of shared decision-making informed dietary intervention based on digital health technology in older adults with type 2 diabetes mellitus: a randomized controlled trial. Digit Health 2024;10:20552076241272514. [doi: 10.1177/20552076241272514] [Medline: 39403713]
66. Buis LR, Kim J, Sen A, et al. The effect of an mHealth self-monitoring intervention (MI-BP) on blood pressure among black individuals with uncontrolled hypertension: randomized controlled trial. JMIR Mhealth Uhealth 2024 Jun 28;12:e57863. [doi: 10.2196/57863] [Medline: 38941601]
67. Tabernero C, Gutiérrez-Domingo T, Steca P, et al. Effectiveness of mindfulness and positive strengthening mHealth interventions for the promotion of subjective emotional wellbeing and management of self-efficacy for chronic cardiac diseases. J Pers Med 2022 Nov 25;12(12):1953. [doi: 10.3390/jpm12121953] [Medline: 36556174]

XSL•FO
RenderX

68. Patnaik L, Panigrahi SK, Sahoo AK, Mishra D, Muduli AK, Beura S. Effectiveness of mobile application for promotion of physical activity among newly diagnosed patients of type II diabetes - a randomized controlled trial. Int J Prev Med 2022;13(1):54. [doi: 10.4103/ijpvm.IJPVM_92_20] [Medline: 35706879]
69. Jia W, Zhang P, Zhu D, et al. Evaluation of an mHealth-enabled hierarchical diabetes management intervention in primary care in China (ROADMAP): a cluster randomized trial. PLoS Med 2021 Sep;18(9):e1003754. [doi: 10.1371/journal.pmed.1003754] [Medline: 34547030]
70. Burka D, Gupta R, Moran AE, et al. Keep it simple: designing a user-centred digital information system to support chronic disease management in low/middle-income countries. BMJ Health Care Inform 2023 Jan;30(1):e100641. [doi: 10.1136/bmjhci-2022-100641] [Medline: 36639189]
71. Gelbman BD, Reed CR. An integrated, multimodal, digital health solution for chronic obstructive pulmonary disease: prospective observational pilot study. JMIR Form Res 2022 Mar 17;6(3):e34758. [doi: 10.2196/34758] [Medline: 35142291]
72. Rusch A, Carley I, Badola P, et al. Digital mental health interventions for chronic serious mental illness: findings from a qualitative study on usability and scale-up of the life goals app for bipolar disorder. Front Digit Health 2022;4:1033618. [doi: 10.3389/fdgth.2022.1033618] [Medline: 36479190]
73. Hietbrink EAG, Middelweerd A, van Empelen P, et al. A digital lifestyle coach (E-Supporter 1.0) to support people with type 2 diabetes: participatory development study. JMIR Hum Factors 2023 Jan 12;10(1):e40017. [doi: 10.2196/40017] [Medline: 36633898]
74. O'Neill K, Gormley C, Kelly MG, et al. Service development project to pilot a digital technology innovation for video direct observation of therapy in adult patients with asthma. BMJ Open Qual 2024 Jul 15;13(3):e002626. [doi: 10.1136/bmjoq-2023-002626] [Medline: 39009461]
75. Salim H, Cheong AT, Sharif-Ghazali S, et al. A self-management app to improve asthma control in adults with limited health literacy: a mixed-method feasibility study. BMC Med Inform Decis Mak 2023 Sep 27;23(1):194. [doi: 10.1186/s12911-023-02300-6] [Medline: 37759184]
76. Barbaric A, Munteanu C, Ross H, Cafazzo JA. A voice app design for heart failure self-management: proof-of-concept implementation study. JMIR Form Res 2022 Dec 21;6(12):e40021. [doi: 10.2196/40021] [Medline: 36542435]
77. Díaz-Mohedo E, Carrillo-León AL, Calvache-Mateo A, Ptak M, Romero-Franco N, Carlos-Fernández J. App-Mohedo®: a mobile app for the management of chronic pelvic pain. A design and development study. Int J Med Inform 2024 Jun;186:105410. [doi: 10.1016/j.ijmedinf.2024.105410] [Medline: 38507980]
78. Kim DY, Kwon H, Nam KW, Lee Y, Kwon HM, Chung YS. Remote management of poststroke patients with a smartphone-based management system integrated in clinical care: prospective, nonrandomized, interventional study. J Med Internet Res 2020 Feb 27;22(2):e15377. [doi: 10.2196/15377] [Medline: 32130140]
79. Xie X, Wang X, Li A, et al. A study of the effectiveness of mobile health application in a self-management intervention for kidney transplant patients. Iran J Kidney Dis 2023 Sep;17(5):263-270. [doi: 10.52547/ijkd.7693] [Medline: 37838936]
80. Holtz BE, Mitchell KM, Holmstrom AJ, et al. The effect of an mHealth intervention for adolescents with type 1 diabetes and their parents. J Telemed Telecare 2024 Aug;30(7):1155-1162. [doi: 10.1177/1357633X221125835] [Medline: 36177538]
81. Xing F, Guo Y, Xia N, et al. Mobile app-assisted family physician program for improving blood pressure outcome in hypertensive patients. BMC Prim Care 2023 Jan 10;24(1):8. [doi: 10.1186/s12875-023-01965-2] [Medline: 36627556]
82. Miki T, Yamada J, Ishida S, Sakui D, Kanai M, Hagiwara Y. Exploring the feasibility and initial impact of an mHealth-based disease management program for chronic ischemic heart disease: formative study. JMIR Form Res 2024 Aug 22;8:e56380. [doi: 10.2196/56380] [Medline: 39173150]
83. Naranjo-Rojas A, Perula-de Torres L, Cruz-Mosquera FE, Molina-Recio G. Usability of a mobile application for the clinical follow-up of patients with chronic obstructive pulmonary disease and home oxygen therapy. Int J Med Inform 2023 Jul;175:105089. [doi: 10.1016/j.ijmedinf.2023.105089] [Medline: 37172506]
84. Lalloo C, Hundert A, Harris L, et al. Capturing daily disease experiences of adolescents with chronic pain: mHealth-mediated symptom tracking. JMIR Mhealth Uhealth 2019 Jan 17;7(1):e11838. [doi: 10.2196/11838] [Medline: 30664472]
85. Funes Hernandez M, Babakhanian M, Chen TP, et al. Design and implementation of an electronic health record-integrated hypertension management application. J Am Heart Assoc 2024 Jan 16;13(2):e030884. [doi: 10.1161/JAHA.123.030884] [Medline: 38226516]
86. Steinert A, Eicher C, Haesner M, Steinhagen-Thiessen E. Effects of a long-term smartphone-based self-monitoring intervention in patients with lipid metabolism disorders. Assist Technol 2020;32(2):109-116. [doi: 10.1080/10400435.2018.1493710] [Medline: 29944463]
87. Fee C, Fuller J, Guss CE, et al. A digital platform to support HIV case management for youth and young adults: mixed methods feasibility study. JMIR Form Res ;6(11):e39357. [doi: 10.2196/39357]
88. Cormican O, Dowling M. Evaluating a digital self-management tool for people living with multiple myeloma. Cancer Nursing Practice 2020;20(1):36-42 [FREE Full text]
89. Digital Literacy. 2024. URL: https://literacy.ala.org/digital-literacy [accessed 2024-11-05]
90. Health literacy. World Health Organization. 2024. URL: https://www.who.int/teams/health-promotion/enhanced-wellbeing/ninth-global-conference/health-literacy [accessed 2024-11-05]

XSL•FO
RenderX

91. Evidence standards framework for digital health technologies. National Institute for Health and Care Excellence. 2019. URL: https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf [accessed 2025-12-14]
92. National framework for the integrated prevention and management of chronic disease in Ireland 2020–2025. : Health Service Executive; 2020 URL: https://www.hse.ie/eng/about/who/cspd/icp/chronic-disease/documents/national-framework-integrated-care.pdf [accessed 2025-12-14]
93. Van de Velde D, De Zutter F, Satink T, et al. Delineating the concept of self-management in chronic conditions: a concept analysis. BMJ Open 2019 Jul 16;9(7):e027775. [doi: 10.1136/bmjopen-2018-027775] [Medline: 31315862]
94. Modi AC, Pai AL, Hommel KA, et al. Pediatric self-management: a framework for research, practice, and policy. Pediatrics 2012 Feb;129(2):e473-e485. [doi: 10.1542/peds.2011-1635] [Medline: 22218838]
95. Grande SW, Longacre MR, Palmblad K, et al. Empowering young people living with juvenile idiopathic arthritis to better communicate with families and care teams: content analysis of semistructured interviews. JMIR Mhealth Uhealth 2019 Feb 22;7(2):e10401. [doi: 10.2196/10401] [Medline: 30794202]
96. Lorig K. Self-management of chronic illness: a model for the future. Generations: Journal of the American Society on Aging 1993;17(3):11-14 [FREE Full text]
97. Grady PA, Gough LL. Self-management: a comprehensive approach to management of chronic conditions. Am J Public Health 2014 Aug;104(8):e25-e31. [doi: 10.2105/AJPH.2014.302041] [Medline: 24922170]
98. Bellin MH, Dosa N, Zabel TA, Aparicio E, Dicianno BE, Osteen P. Self-management, satisfaction with family functioning, and the course of psychological symptoms in emerging adults with spina bifida. J Pediatr Psychol 2013;38(1):50-62. [doi: 10.1093/jpepsy/jss095] [Medline: 22976508]
99. Lupton D. The digitally engaged patient: self-monitoring and self-care in the digital health era. Soc Theory Health 2013 Aug;11(3):256-270. [doi: 10.1057/sth.2013.10]
100. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30. [doi: 10.2196/jmir.4055] [Medline: 25639757]
101. Schunk DH, Ellen LU. Social cognitive theory and motivation. In: The Oxford Handbook of Human Motivation: Oxford University Press; 2012. [doi: 10.1093/oxfordhb/9780195399820.013.0002]
102. Al-Rayes S, Al Yaqoub FA, Alfayez A, et al. Gaming elements, applications, and challenges of gamification in healthcare. Informatics in Medicine Unlocked 2022;31:100974. [doi: 10.1016/j.imu.2022.100974]
103. Kemp E, Trigg J, Beatty L, et al. Health literacy, digital health literacy and the implementation of digital health technologies in cancer care: the need for a strategic approach. Health Prom J of Aust 2021 Feb;32(S1):104-114 [FREE Full text] [doi: 10.1002/hpja.387]
104. Vollbrecht H, Arora V, Otero S, Carey K, Meltzer D, Press VG. Evaluating the need to address digital literacy among hospitalized patients: cross-sectional observational study. J Med Internet Res 2020 Jun 4;22(6):e17519. [doi: 10.2196/17519] [Medline: 32496196]
105. Papageorgiou A, Strigkos M, Politou E, Alepis E, Solanas A, Patsakis C. Security and privacy analysis of mobile health applications: the alarming atate of practice. IEEE Access 2018;6:9390-9403. [doi: 10.1109/ACCESS.2018.2799522]
106. Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the far side of mobile health: information security and privacy of mobile health apps on iOS and Android. JMIR Mhealth Uhealth 2015 Jan 19;3(1):e8. [doi: 10.2196/mhealth.3672] [Medline: 25599627]
107. Alzghaibi H. Healthcare practitioners' perceptions of mHealth application barriers: challenges to adoption and strategies for enhancing digital health integration. Healthcare (Basel) 2025 Feb 25;13(5):494. [doi: 10.3390/healthcare13050494] [Medline: 40077056]
108. Leigh S, Ashall-Payne L, Andrews T. Barriers and facilitators to the adoption of mobile health among health care professionals from the United Kingdom: discrete choice experiment. JMIR Mhealth Uhealth 2020 Jul 6;8(7):e17704. [doi: 10.2196/17704] [Medline: 32628118]
109. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. JMIR Mhealth Uhealth 2019 Apr 16;7(4):e11223. [doi: 10.2196/11223] [Medline: 30990458]
110. Giebel GD, Abels C, Plescher F, et al. Problems and barriers related to the use of mHealth apps from the perspective of patients: focus group and interview study. J Med Internet Res 2024 Apr 23;26:e49982. [doi: 10.2196/49982] [Medline: 38652508]

Abbreviations

AI: artificial intelligence
BP: blood pressure
COPD: chronic obstructive pulmonary disease
EHR: electronic health record
EMR: electronic medical record

XSL•FO
RenderX

HbA1c: hemoglobin A1c
mHealth: mobile health
NICE: National Institute for Health and Care Excellence
PRISMA-ScR: Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews
RCT: randomized controlled trial

Edited by L Buis; submitted 06.Jul.2024; peer-reviewed by A Dima, B Seljelid, N Nøhr; revised version received 06.Sep.2025; accepted 01.Oct.2025; published 14.Jan.2026.

Please cite as:
Al Mahmud A, Joachim S, Jayaraman PP, Learmonth C, Tyagi S, Forkan ARM, Shuakat M, Wickramasinghe N, Wheeler J, Best S, Trainer A
Digital Health Interventions to Support Chronic Disease Management: Systematic Scoping Review
JMIR Mhealth Uhealth 2026;14:e63742
URL: https://mhealth.jmir.org/2026/1/e63742
doi:10.2196/63742

© Abdullah Al Mahmud, Shane Joachim, Prem Prakash Jayaraman, Caitlin Learmonth, Shivani Tyagi, Abdur Rahim Mohammad Forkan, Muhammad Shuakat, Nilmini Wickramasinghe, Jack Wheeler, Stephanie Best, Alison Trainer. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobile Apps to Improve Health Parameters in Healthy Adults: Systematic Review

Gaia Leuzzi[1,2], PhD; Mirko Job[1], PhD; Carola Cosentino[3], PhD; Riccardo Testa[4], MSc; Aldo Scafoglieri[2], PhD; Marco Testa[1], PhD

[1]Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genoa, Campus of Savona, Savona, Italy

[2]Department of Physical Education and Rehabilitation, Experimental Anatomy Research Group (EXAN), Vrije Universiteit of Brussel (VUB), Brussels, Belgium

[3]Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genoa, Genova, Italy

[4]Department of Electrical, Electronics, Telecommunication Engineering and Naval Architecture, University of Genoa, Genova, Italy

Corresponding Author:
Marco Testa, PhD
Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genoa, Campus of Savona, Savona, Italy

Abstract

Background: Recently, mobile health and mobile apps have been proposed as a potential tool to improve different outcomes (eg, daily steps, blood glucose) in both people with and without chronic conditions. In particular, healthy people could benefit from these tools by improving health variables and for prevention. Previous evidence investigated different types of health interventions adopting apps in various settings and populations, but evidence of their effectiveness is still unclear.

Objective: The aim was to assess the effectiveness of mobile apps in improving health variables (eg, daily steps, maximal aerobic capacity) in healthy adults, involving an intervention regarding physical activity, diet, or their combination thereof. Evidence would suggest if apps could be effectively adopted in health interventions aiming toward prevention.

Methods: A systematic review was performed using Medline via PubMed, Cochrane Library—CENTRAL, and Embase. Only randomized controlled trials comparing the same intervention provided with and without a mobile app or a treatment and a mobile app compared with the treatment only were included in this systematic review. The Risk of Bias tool 2.0 was used to assess the risk of bias, and the GRADE (Grading of Recommendations, Assessment, Development and Evaluation) was adopted for rating the certainty of evidence.

Results: Considering studies up to June 2025, only 2 studies were included in the review of mobile apps for physical activity, and none were included for mobile apps for diet and none for mobile apps for physical activity and diet combined. The quality of evidence of the 2 studies included was low due to a high risk of bias, several missing data, and deviation from the original interventions, suggesting a scarce rigor in the methodology adopted. Therefore, mobile apps' effectiveness in improving diet, physical activity, or their combination cannot be assessed.

Conclusions: Despite the widespread use of mobile apps for health and the large number of relative publications, the results of this systematic review did not allow us to ascertain the effectiveness of mobile apps for health, but they provided fundamental insights for future research. Hence, it is not possible to state if apps for health might be used as supporting tools for health interventions aiming toward prevention and health improvements in healthy people. There is an urgent need to develop stronger evidence of apps' effectiveness in addressing different populations and types of interventions for different health domains.

Trial Registration: PROSPERO CRD42023485803; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023485803

(JMIR Mhealth Uhealth 2026;14:e66881) doi:10.2196/66881

KEYWORDS

active aging; diet; exercise; healthy ageing; health apps; mobile apps; physical activity

Introduction

Promoting active aging [1] requires new strategies to reach different populations in a feasible and effective way to help people improve their health status [2,3]. For this reason, the use of technology interventions is getting paramount attention to help people improve their health variables [4], identified as both clinical (eg, blood pressure, weight) and nonclinical outcomes (eg, daily activity or sleep) [5,6]. The attention to the use of technology-supported health interventions is due to the ease of

XSL•FO
RenderX

the use of mobile devices, their portability, their quality-price ratio [7], and the quantity of information they can provide with good data storage and live data analysis [8].

The World Health Organization describes this type of app as mobile health app (mHealth), defining it as "medical and public health practices supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices" [9]. Wireless devices can be fitness trackers, smartwatches, and smartphones, which makes it easy to collect several types of data (ie, number of daily steps, macronutrients, sleep, and stress level) from different health-related spheres, such as physical activity and diet [10,11], automatically (ie, using a wearable device [12]), through user's action or both. In particular, to do so, smartphones allow for downloading different types of apps, including health ones. These apps can be focused on just 1 health-related sphere [13,14] or a combination thereof [15].

Due to their easy use and versatility, people with and without diseases can benefit from mHealth apps to prevent or treat different conditions [16,17]. In line with that, different studies have been carried out in the last few years to test the effectiveness of apps for health on people with [18] and without diseases [19]. Unfortunately, the quality of the studies is low, and the results are often controversial [20], thus not allowing for definitive results on this topic. Moreover, to understand the effect of apps for health, one should compare the effectiveness of an intervention provided with and without the adoption of an app, while most of these studies adopt apps as baseline treatment with different interventions as adjunctive therapy, therefore testing the effectiveness of the adjunctive therapy rather than the app one [21,22]. Furthermore, mobile apps were chosen as mHealth to be investigated, as they are one of the most adopted technologies worldwide in health contexts [23], and those targeting fitness, nutrition, and healthy living are widely diffused [24].

Hence, the main purpose of this systematic review is to analyze the effectiveness of mobile apps in improving healthy adults' (ie, >18 years old) health variables, analyzing only randomized controlled trials (RCTs) that compare the same intervention with and without this technology in physical activity, diet, and a combination thereof.

Methods

The protocol of this systematic review was created and submitted to PROSPERO [25] (CRD42023485803). Furthermore, PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Checklist 1) [26] guidelines and PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Literature Search Extension; Checklist 2) [27] were followed to report this review.

Deviations from the Protocol

The protocol initially restricted inclusion to English-language studies. During screening, it became clear that this would underrepresent available evidence and introduce potential language bias. We therefore expanded the criteria to include studies in other languages when reliable translation or accurate data extraction was possible. This deviation was made to enhance the review's completeness and international representativeness, while maintaining all other methodological criteria.

Study Objective

The main objective of this systematic review was to analyze the effectiveness of mobile apps for physical activity and diet to improve healthy adults' health variables. The research question was as follows: are mobile apps effective in improving health-related variables in healthy adults? To address this topic, three different options were investigated: (1) mobile apps for physical activity, (2) mobile apps for diet, and (3) mobile apps for physical activity and diet. The main outcomes listed hereafter correspond to each of the abovementioned points: (1) physical activity variables (eg, daily steps, moderate-to-vigorous physical activity), (2) diet variables (eg, weight, BMI), and (3) physical activity and diet variables as mentioned above.

Eligibility Criteria

For this systematic review, studies were considered eligible if they were published RCTs. No limitations on publication time were set, and RCTs published online until June 3, 2025, were included. Systematic reviews, reviews, meta-analyses, single-case studies, case series, observational studies, books, documents, guidelines, reports, and conference abstracts were excluded. Gray literature, systematic reviews, and meta-analyses were consulted, but not considered eligible, to find useful studies.

We included all those studies that involved healthy adult participants (>18 years old), with no cognitive impairments, musculoskeletal or neuromotor diseases, chronic conditions (eg, diabetes, hypertension), obese (ie, BMI≥30), or pregnant women. Interventions were considered eligible if a mobile app was used as an intervention to improve variables related to physical activity, diet, or their combination or only as a supportive technology to a specific intervention. No limitations on time or sessions of interventions were set. Mobile apps, including automatic or self-reported data collection and the use of a wearable device, were taken into account. Conversely, we excluded interventions with non-standalone apps, using high-cost sensors, exergames, nonmobile monitoring systems, robotics systems, or just clinician's telemedicine. Moreover, studies that did not precisely describe the health conditions of their population were excluded.

Search Strategy

Three scientific databases were sought for the study research: Medline via PubMed, Cochrane Library—CENTRAL, and Embase. They were chosen as they are reported as mandatory by the *Cochrane Handbook for Systematic Reviews of Interventions* [28].

The literature search was performed on the databases up to June 3, 2025, and the results were later merged into a single file to be subsequently uploaded onto Covidence [29], where the automatic duplicate detection was conducted. Specific search strings were created for the 3 databases, mixing Boolean

XSL•FO
RenderX

operators (ie, AND, OR), MeSH terms, and keywords. The research strategy is reported below.

- Mobile apps and physical activity and RCT: ((mHealth) OR (m-health) OR ("mobile health") OR ("mobile application") OR ("mobile app") OR ("smartphone application") OR ("smartphone app") OR (apps) OR (smartphone) OR ("Mobile Applications"[Mesh]) OR ("Smartphone"[Mesh])) AND ((fitness) OR ("physical exercise") OR ("physical fitness") OR ("fitness behavior") OR ("Physical Fitness"[Mesh]) OR ("Exercise"[Mesh]) OR (pedometer) OR (steps) OR (exercise) OR ("training exercise") OR ("heart rate variability") OR ("Heart rate") OR ("Heart Rate"[Mesh])) AND ((single blind) OR (double blind) OR (trial) OR (random*) OR (randomized) OR (randomized controlled))
- Mobile apps and diet and RCT: ((mHealth) OR (m-health) OR ("mobile health") OR ("mobile application") OR ("mobile app") OR ("smartphone application") OR ("smartphone app") OR (apps) OR (smartphone) OR ("Mobile Applications"[Mesh]) OR ("Smartphone"[Mesh])) AND ((diet) OR ("calorie counter") OR ("calorie counting") OR (calorie) OR ("calorie intake") OR (diet) OR ("Diet"[Mesh]) OR (dieting) OR ("weight loss") OR ("weight loss"[Mesh]) OR ("Weight Reduction Programs"[Mesh])) AND ((single blind) OR (double blind) OR (trial) OR (random*) OR (randomized) OR (randomized controlled))

Selection Process

Regarding the first research question, 2 researchers (GL and MJ) manually and independently screened titles and abstracts of the retrieved papers and evaluated them against the inclusion criteria. At the same time, for the second research question, 2 researchers (GL and RT) followed the same procedure. For the third research question, studies were identified among the papers selected by the above-described screening. The eligibility of the studies was then agreed upon through a consensus meeting between the 2 authors of each review and, in case of disagreement, a third researcher (CC) was consulted to reach a final decision. Afterward, the full texts of the selected papers were further screened against the inclusion criteria following the same process.

Data Collection

Two researchers (GL and MJ and GL and RT) proceeded blindly and independently to extract specific data from each study such as authors, year of publication, country, intervention setting, study design, total number of participants, number of participants for each experimental group, mean age of the participants and standard deviation (if available), number of female and male participants, type and timing of intervention sessions for both experimental groups, number and timing of follow-ups, outcomes, key conclusions, and eventually even a researcher's comment on each study. Moreover, all data available in each study were extracted and reported, such as mean, median, IQR, SD, number of follow-ups, and data registered at each follow-up. In case of missing data, authors were contacted.

Data Items

The most relevant characteristics of the selected studies are summarized in Table 1. According to our research questions, the outcomes of this systematic review are grouped as follows: (1) physical activity outcomes, (2) diet outcomes, and (3) physical activity and diet outcomes. No limits were identified for the reporting of any outcome. In case of missing data, authors were contacted.

Table . Study characteristics.

Study characteristics	Zongpa et al [30] (2020)	Muntaner-Mas et al [31] (2021)
Total number of participants	47	66
App name	Take a Walk	Vidahora
Type of intervention	App+diet indications	App
Type of control	No app+diet indications	No app
Primary outcome	VO_2 max[a] + HRV[b] + FBG[c] + adherence	Weigh + waist and hip circumference + 20-m shuttle run test + handgrip + standing long jump test + 4 × 10 m shuttle run test + sit and reach + IFIS[d]
Follow-ups	Week 4	Week 9
Intervention sessions and duration	6 reminders/day, 1 for every working hour	Free

[a]VO_2 max: maximal aerobic capacity.

[b]HRV: heart rate variability.

[c]FBG: fasting blood glucose.

[d]IFIS: International Fitness Scale.

Risk of Bias Assessment

The risk of bias (RoB) assessment was performed, independently and blindly, by 2 researchers for each study (GL and MJ or GL and RT, respectively) following the Revised Cochrane Risk of Bias tool 2.0 (RoB 2.0) [32] for RCTs or the Rob 2 CRT for cluster-randomized controlled trials [33]. This tool aims at assessing the RoB specifically for 5 domains: "Risk of bias arising from the randomization process," "Risk of bias due to

XSL•FO
RenderX

deviations from the intended interventions," "Risk of bias due to missing outcome data," "Risk of bias in measurement of outcome," and "Risk of bias in selection of the reported results." Consequently, an overall RoB for the study is provided. Domains and studies can be classified at low, moderate, or high RoB. The tool also allows one to indicate "no information" as an answer for each item of every domain and, in this case, it would often be considered at high RoB. A third researcher (CC) was contacted in case of disagreement to reach a consensus.

Statistical Analysis

Data from each study were extracted and reported, and a descriptive statistic was performed. For intergroup comparisons, the mean, SD, and/or mean differences for pre- and posttreatment conditions were reported. Additionally, the 1- or 2-tailed *t* tests for normally distributed data and the Mann-Whitney *U* test for nonnormally distributed data were also reported if performed in the studies. All statistical analyses were performed using the Jamovi statistical software [34].

Quality of Evidence

To perform the quality of evidence assessment, the GRADE (Grading of Recommendations, Assessment, Development and Evaluation) [35] approach was used via the GRADEpro GTD tool. This tool helps assess both the certainty of evidence and the strength of recommendations. The evaluation process took into account 5 different domains: risk of bias, imprecision (eg, sample size, confidence intervals), inconsistency (eg, heterogeneity), indirectness (eg, eligibility criteria against actual studies included), and publication bias (eg, bias in results publication).

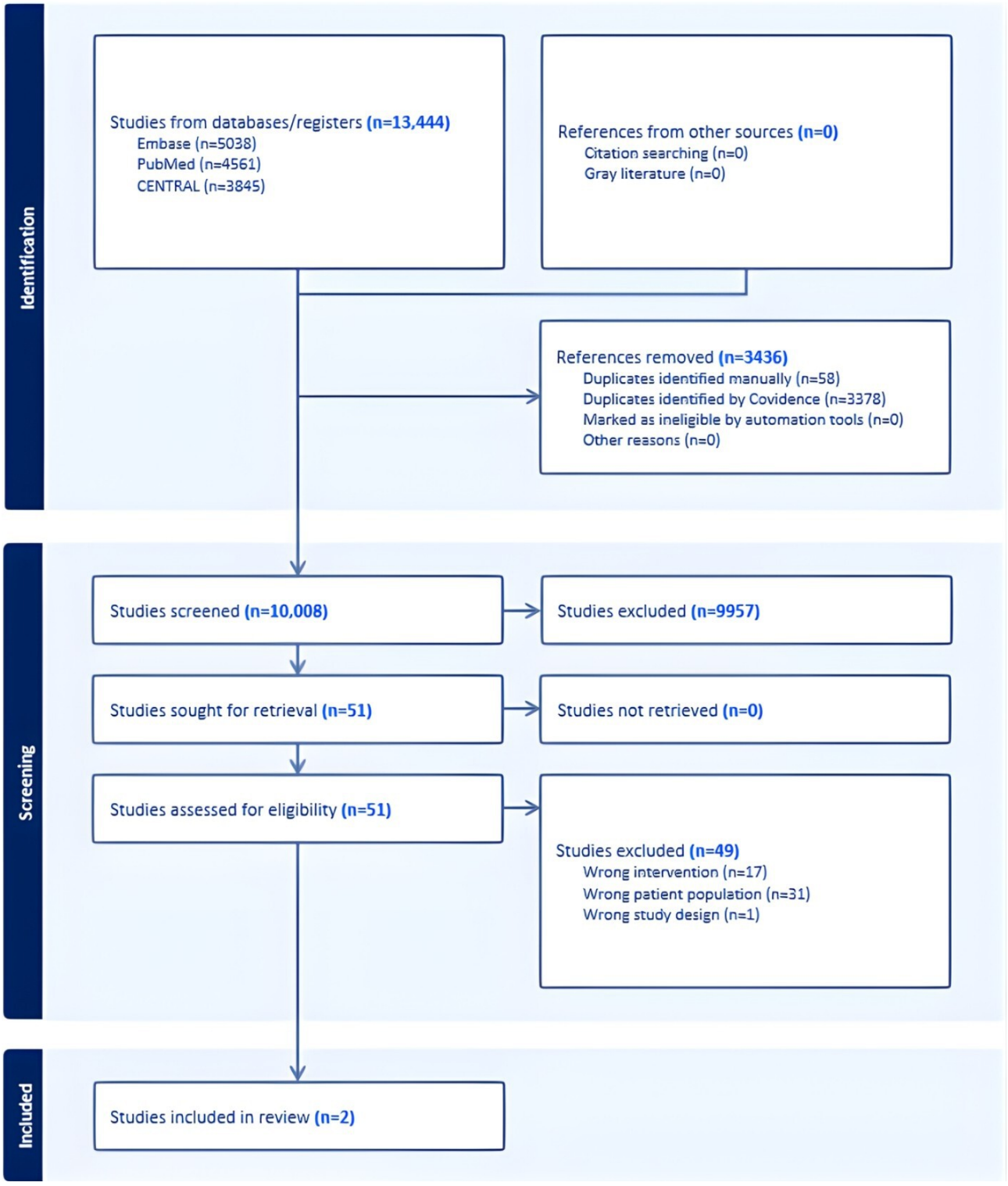
Results

Study Selection

The literary search process for the first review identified a total of 13,444 studies. Duplicate removal eliminated 3436 studies, leaving a total of 10,008 studies to screen. After applying the inclusion and exclusion criteria to titles and abstracts, 51 studies were left [19,30,31,36-82]. Full-text studies were read independently by 2 researchers (GL and MJ), and in due course, another 49 papers were excluded [19,36-82], resulting in the final inclusion of 2 studies for further analysis [30,31]. The complete research process is graphically displayed in Figure 1, and the reasons for exclusions are reported in Multimedia Appendix 1. Multimedia Appendix 2 reports the complete research processes of the other 2 research questions.

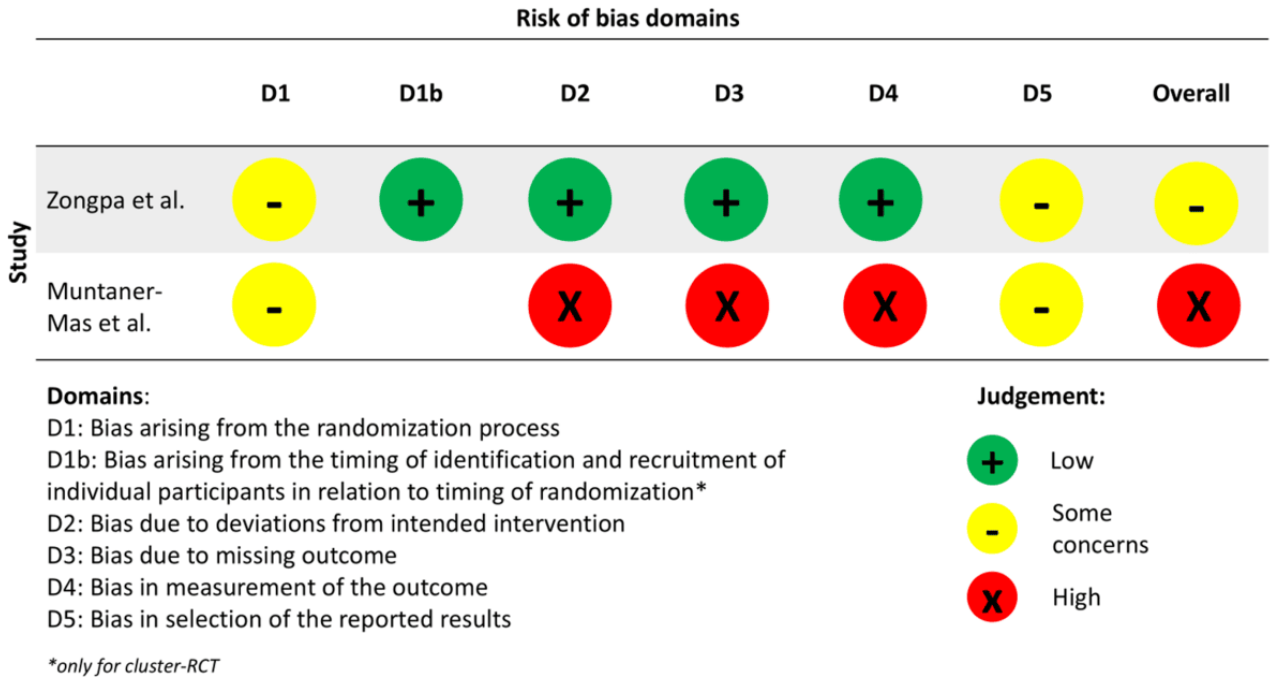
XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for mobile apps for physical activity.



XSL•FO
RenderX

Figure 2. Risk of bias (RoB) assessment for randomized controlled trial (RCT) and cluster-RCT [30,31].



Regarding the second and third questions (ie, mobile app for diet, and physical activity and diet), unfortunately, no studies were included. The selection processes are shown in Multimedia Appendix 2.

Study Characteristics

Among the 2 studies included in the final analysis, the first was a cluster RCT published in 2020 [30], and the second one was an RCT published in 2021 [31]. Both had only 1 intervention group and 1 control group [30,31]. The time of the intervention ranged from 4 to 9 weeks. The countries of study development were India [30] and Spain [31]. The studies' characteristics are indicated in Table 1. The first study [30] involved 47 participants, and the intervention consisted of giving a smartphone app (Take a Walk) to remind them to perform a few minutes of walking at regular intervals. The intervention group received "walk breaks" reminders each hour of work for a total of 6 reminders a day. The intervention lasted 4 weeks. Both intervention and control groups received an indication to follow a standard diet of 2300 Kcal/day. Physical activity–related variables were evaluated using different tests or variables, such as the fasting blood glucose, VO_2 max (maximal aerobic capacity), heart-rate variability, and adherence to walking breaks.

In the second study [31], involving 66 participants, the intervention group was provided with an app (Vidahora) to be used for 9 weeks. Participants were free to use the app whenever they wanted and were only advised to perform at least 10 minutes of physical activity 3 times a day. The second study assessed physical activity–related variables using several performance tests, anthropometric tests, and a questionnaire. Anthropometric tests were weight, hips, and waist circumferences. Performance tests were 20-m shuttle run test, handgrip test, standing long jump test, 4×10 m shuttle run test, and sit and reach test. The questionnaire used was the International Fitness Scale [83].

The characteristics of both studies are reported in Table 1, while the studies' outcomes are reported in Table 2.

XSL•FO
RenderX

Table . Physical activity-related outcomes.

Authors (app used) and test	Baseline		Post-intervention	
	Intervention group	Control group	Intervention group	Control group
Zongpa et al [30] (Take a Walk)				
VO_2 max[a] (mL/kg/min), median (IQR)	45.3 (39.0-52.3)	36.0 (36.0-41.2)	47.6 (39.6-55.9)	37.5 (35-40.3)
HRV[b]				
Time domain, median (IQR)				
SDNN[c] interval (ms)	52.7 (51.4-53.4)	49.1 (44.8-52.1)	54.5 (52.5-60.2)	49.4 (48.1-50.6)
RMSSD[d] interval (ms)	52.4 (48.2-54.2)	54.2 (52.4-55.8)	58.1 (57.4-58.6)	55.1 (53.2-55.7)
NN50[e] (beats)	129.4 (129.1-129.7)	129.7 (128.9-129.2)	130.3 (129.6-132.1)	126.2 (123.8-127.9)
pNN50[f] (%)	37.4 (37.3-37.6)	37.2 (37.0-37.2)	39.0 (38.2-39.2)	35.3 (34.2-37.1)
Frequency domain, median (IQR)				
VLF[g] (ms^2/Hz)	103.00 (101.0-105.0)	103.0 (90.0-103.0)	107.0 (102.0-112.0)	90.1 (84.7-103.1)
LF[h] (ms^2/Hz)	981.0 (972.0-988.2)	984.3 (983.6-985.7)	986.5 (978.0-996.0)	976.0 (962.0-984.0)
HF[i] (ms^2/Hz)	970.0 (958.0-984.0)	981.5 (965.0-984.0)	986.0 (973.0-994.5)	975.0 (962.0-982.0)
LF/HF[j] (%)	1.02 (0.98 - 1.04)	1.0 (0.96 - 1.04)	1.0 (0.9 - 1.0)	1.0 (0.98 - 1.04)
Nonlinear index, median (IQR)				
SD1[k] (ms)	37.2 (36.1-38.3)	22.1 (22.0-25.1)	41.1 (38.2-44.1)	35.8 (34.5-37.7)
SD2[k] (ms)	55.0 (51.8-59.7)	33.5 (30.7-33.5)	88.6 (72.0-88.6)	59.4 (57.6-61.1)
SD1/SD2 (%)	1.5 (1.2-1.8)	1.5 (1.2-1.5)	2.1 (1.9-2.2)	1.7 (1.3-2.0)
FBG[l] (mmol/dL), median (IQR)	89.0 (78.0-93.2)	87.0 (81.0-88.0)	83.0 (72.0-86.0)	87.0 (81.0-92.1)
Adherence	NA[m]	NA	NA	NA
Muntaner-Mas et al [31] (Vidahora)				
Weight (kg), mean (SD)	65.1 (12.1)	65.1 (13.7)	65.7 (12.1)	65.5 (13.8)
Waist circumference (cm), mean (SD)	77.1 (9.9)	81.0 (11.8)	76.0 (11.7)	80.2 (11.2)
Hip circumference (cm), mean (SD)	96 (8.8)	96.3 (9.0)	93.3 (9.1)	95.7 (10.2)
20-m shuttle run (laps), mean (SD)	6.6 (3.2)	5.8 (3.4)	7.7 (2.8)	5.7 (3.3)
Handgrip strength (kg), mean (SD)	30.8 (8.0)	28.9 (7.9)	32.1 (9.0)	28.2 (9.1)
Standing broad jump (cm), mean (SD)	155.4 (35.2)	146.0 (31.0)	169.7 (35.9)	150.5 (29.4)
4×10 m shuttle run (sec), mean (SD)	11.4 (1.3)	11.7 (1.4)	11.3 (1.3)	12.0 (1.5)
Sit-and-reach (cm), mean (SD)	19.9 (8.9)	20.2 (8.9)	21.6 (9.7)	21.7 (8.3)
General physical fitness, mean (SD)	3.2 (0.7)	3.3 (0.9)	3.6 (0.6)	3.2 (0.9)
Cardiorespiratory fitness, mean (SD)	2.8 (1.0)	2.8 (1.1)	3.1 (1.0)	2.7 (1.1)

XSL•FO
RenderX

Authors (app used) and test	Baseline		Post-intervention	
	Intervention group	Control group	Intervention group	Control group
Muscular fitness, mean (SD)	3.2 (0.7)	0.1 (1.1)	3.4 (0.7)	3.2 (1.0)
Speed-agility, mean (SD)	3.3 (0.7)	3.3 (1.0)	3.5 (0.7)	3.3 (1.0)
Flexibility, mean (SD)	2.8 (0.9)	2.8 (1.2)	3.1 (1.2)	2.9 (1.1)

[a]VO_2 max: maximal aerobic capacity.

[b]HRV: heart rate variability.

[c]SDNN interval: standard deviation of NN intervals.

[d]RMSSD: root mean square of successive RR interval differences.

[e]NN50: successive RR intervals that differ by more than 50 ms.

[f]pNN50: percentage of successive RR intervals that differ by more than 50 ms.

[g]VLF: very low frequency of power.

[h]LF: absolute power of the low-frequency band (0.04-0.15 Hz).

[i]HF: absolute power of the high-frequency band (0.15-0.4 Hz).

[j]LF/HF: ratio of LF-HF power.

[k]SD1 and SD2: Poincaré plots perpendicular to line of identity.

[l]FBG: fasting blood glucose.

[m]NA: not available but requested to authors.

App's Characteristics

Information about the Take a Walk app was limited; it was described as a simple Java–based Android app that allowed participants to set personalized reminders for walking, including customizable times and data. Participants were only required to manually set when to receive the reminders to walk.

The Vidahora app was made of 4 different sections: the first section dedicated to a quiz about healthy habits, the second section dedicated to the challenges for improving different physical activities' components via suggested video exercises (eg, strength, aerobic exercise, yoga), the third section hosted an artificial intelligence–assisted chatbot that could ask the participant about progress in a friendly way, and the last section was for setting the user data (eg, username, personal data). Badges for achievements were also present, as well as individual and community challenges with daily or weekly aims. Participants in the intervention groups were invited to use the app as they wished, with the only suggestion of recording at least 3 sessions of a minimum of 10 minutes per week of physical education.

Risk of Bias in Studies

The RoB assessment for the included RCT studies is graphically reported in Figure 2 using the Robvis tool [84]. The first study presents some concerns in the overall RoB, due to the randomization process since it is not clear how the experimenters performed it. Moreover, the study reports the registration of a protocol with a registration number, but in the mentioned database, it is not possible to find the protocol. In this case, it cannot be excluded that the results were not analyzed by a prespecified analysis plan that was finalized before unblinded outcome data were available for analysis. Indeed, the authors reported having several missing data but without any reasonable explanation.

The same 2 domains (D1 and D5) influence the overall RoB of the second study as well. The second study has the overall RoB indicated as "high risk." Domains 2, 3, and 4 are at high RoB, and those are influencing the overall RoB. The corresponding authors of the studies were contacted, but no answer was ever received.

GRADE Assessment

The assessment of the quality of evidence adopting the GRADE approach could not be performed due to the high heterogeneity of the outcomes considered in the studies included. First, the 2 studies did not consider the same outcome. Moreover, even involving the same sphere of interest (ie, physical activity) did not consider the same outcomes. Specifically, the first study [30] mainly considered physiological outcomes (ie, VO_2 max, blood glucose), while the second study [31] investigated performance outcomes (eg, 20-m shuttle run test, handgrip). Therefore, it is not possible to assess the quality of this evidence.

Discussion

Principal Findings

The included studies were overall characterized by a high RoB due to many missing values, high dropout, small sample sizes, and poor data reporting. In particular, the randomization processes were evaluated with the RoB 2.0 tool with "some concerns" as this tool requires this scoring if the paper does not describe the randomization procedures adopted in detail but just mention their adoption. Additionally, for 1 study, a protocol was not available. Moreover, the RoB of the second study was influenced by the decision of not including in the data analysis the participants who had missing data in 1 of the evaluations or were outliers. Thus, it might not be excluded that the results were influenced.

XSL•FO
RenderX

Furthermore, their study designs were significantly different, and it was not possible to make a direct comparison between their results. Our findings highlight the urgent need for standardized outcome measures to enable the generation of stronger, comparable evidence in this field. For each health domain examined, a validated set of standardized outcomes should be developed, allowing for their consistent use across diverse study designs and settings, including RCTs. Such standardization would facilitate more accurate assessments of mobile app effectiveness on health outcomes in healthy adults. Due to the heterogeneity of outcome measures among the included studies, it remains difficult to draw definitive conclusions regarding the efficacy of health-related mobile apps in this population. Therefore, future studies should aim to include larger sample sizes to enhance statistical power and improve the reliability of findings. Additionally, greater participant numbers may also help mitigate issues related to dropout and incomplete data during interventions. Moreover, intervention times should also be standardized to be able to compare results from different studies. Another problem that emerges from our results is the lack of standardized apps or guidelines to develop them for different health domains, and this might be seen from the different outcomes considered in each study included in our review. A standardized version of the health app could allow for having a set of common health data across different apps, with the possibility of adding other health variables specific to each app based on its characteristics and aims. In this way, studies could compare the use of different health apps for the same domain and consider a minimum set of common health variables.

Other systematic reviews were carried out on mHealth in the last years [85-89] and were also characterized by a very limited number of selected studies with high heterogeneity, and therefore they could not assess mHealth effectiveness. Indeed, many of the studies they included presented mixed results of the delivered interventions and the way of delivering them [64,87,90,91]. Hence, those studies did not assess the effectiveness of the same intervention delivered with and without mHealth, as we conversely did in our work. Considering the available literature, it is fundamental to emphasize the need to evaluate the effectiveness of specific mHealth interventions. However, the intrinsic variability in the designs and the scarce quality of the currently available studies do not allow us to state if mobile apps can be effective to improve health variables.

To overcome this problem, we decided to include in our work the studies that compared the same intervention provided with and without the mobile app, and additionally also the studies considering the same treatment provided via mobile app against the treatment. By applying these severe criteria, many studies were excluded from this systematic review for improper control, leaving only 2 studies to analyze. Consistent with our findings, other works reported the need for more studies with clearer designs to test the effectiveness of mHealth technology in different settings [92-95].

Studies Included

Digging into the included studies, Zongpa et al reported that physical activity–related variables improved over a 4-week period, specifically VO_2 max, heart rate variability, and fasting blood glucose. It should be noted that dropouts were 11.32%, and VO_2 max improvements may be questioned because the validity of a submaximal test in healthy people is questionable. Moreover, the VO_2 max improvement reported by this study was only 1.33%, while the minimal clinically important difference for VO_2 max should be higher than 6% [96]. Finally, the results should be taken carefully, as the results included in the analysis considered only an intervention adherence of at least 70%, thus imposing a possible bias in the selection of the results.

In the second study, Muntaner-Mas et al reported that many physical activity–related variables evaluated in the study improved, and the authors decided to split them into 3 categories: fatness indicators, physical fitness components, and self-reported fitness.

Starting from fatness indicators (ie, weight, waist, and hip circumference), no changes were obtained that could be attributable to the app. Physical fitness components (ie, 20-m shuttle run laps, handgrip strength, standing broad jump, 4×10 m shuttle run, and sit-and-reach) were improved, but even though few changes are indicated as statistically significant, the actual improvements are minimal and might not even be clinically relevant [97]. Second, improvements of a few units of centimeters or seconds obtained in 9 weeks and from a healthy and young population could be considered scarce. The category “self-reported physical fitness” explored 5 domains, and all the components were evaluated via Likert-type questions and reported the results obtained from the International Fitness Scale. The results improved for the intervention group and decreased for the control group, but considering the possible bias emerging from the self-evaluation, improvements should be carefully addressed since changes were minimal. Even in this category, the results on mobile app effectiveness could not be considered conclusive.

Despite the study considering the mobile app effective in improving physical activity–related variables, there were many missing data (ie, about 30%), and the sample size is limited. Summarizing, the effectiveness of the Vidahora app cannot be assessed. Although the selected studies highlighted that diet and physical activity levels can be improved by mHealth apps, their weak methodological design raises some concerns about their conclusions.

Limitation

A limitation of this work must be acknowledged: this systematic review included only studies involving healthy participants. Therefore, it is not possible to report anything about people with chronic conditions or pathologies. Additionally, a librarian was not consulted to develop the research strings, as people with expertise in conducting systematic reviews, and in their methodology, were consulted.

Conclusions

Despite the studies we included seeming to support the effectiveness of mobile apps to improve physical activity, diet-related variables, or their combination in healthy adults,

XSL•FO
RenderX

their poor methodological quality as well as the high variability in literature does not allow any definitive conclusion on this topic. Besides, the long-term effects of mobile apps interventions on different outcomes are scarce [92,95], and further research is needed. In addition, the interventions' (eg, activity, diet) effectiveness should be tested a priori and then provided via mHealth. Some urgent needs emerge from the literature analyzed and from this study. Specifically, for future studies, there is a need for high-quality RCT designs with large sample sizes to better assess the possible effects of health apps and the generalizability of results. Moreover, there is a need for clearer and consistent methodology that could provide stronger evidence of effectiveness, more transparent reporting of results that would prevent any bias and would additionally allow for acknowledging what is not working with apps for health and why, and addressing healthy people to test the mobile apps' effectiveness in preventing diseases and improving health conditions. Furthermore, standardized outcomes for each health domain of interest (eg, physical activity, diet) should be adopted, allowing for comparing the results of different studies and populations. Additionally, different studies could include, in the same health domain of interest, the chosen standardized outcomes as well as new ones, to try to expand possible results. Moreover, mHealth should be tested and validated by both patients and users before using them to deliver an intervention. Further research should test mobile apps as a tool supporting preventive approaches for health and well-being in young people and healthy participants as well. Finally, clear guidelines should be created on how to build up different types of mHealth, specifically for mobile apps, to standardize this process among health apps and to further try to ensure better use of this technology in the active aging and well-being fields.

Acknowledgments

We would like to sincerely express our gratitude to Dr Simone Battista for his support in carrying out this work. In all parts of this manuscript, generative artificial intelligence was used solely to ensure clarity and correct syntax. It was not used to create or contribute any insights into the research itself or any of its components.

Funding

No funding was received for this study.

Data Availability

All the relevant details and information are presented in the manuscript. Nonetheless, the data are available on reasonable request from the authors.

Authors' Contributions

GL, CC, MJ, and MT performed study conceptualization. GL and CC performed methodology. GL, RT, and MJ conducted the analysis, investigation, and data extraction. GL wrote the original draft preparation. GL, MJ, CC, RT, AS, and MT reviewed and edited it. All authors read and approved the final manuscript.

Multimedia Appendix 1
Reasons for exclusions (mobile apps for physical activity).
[DOCX File, 19 KB - mhealth_v14i1e66881_app1.docx]

Multimedia Appendix 2
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagrams.
[DOCX File, 54 KB - mhealth_v14i1e66881_app2.docx]

Checklist 1
PRISMA checklist.
[DOCX File, 22 KB - mhealth_v14i1e66881_app3.docx]

Checklist 2
PRISMA-S checklist.
[DOCX File, 30 KB - mhealth_v14i1e66881_app4.docx]

References

1. Walker A. A strategy for active ageing. Int Soc Secur Rev 2002 Jan;55(1):121-139. [doi: 10.1111/1468-246X.00118]
2. Cai T, Verze P, Bjerklund Johansen TE. The quality of life definition: where are we going? Uro 2021 Feb;1(1):14-22. [doi: 10.3390/uro1010003]

XSL•FO
RenderX

3. Noble N, Paul C, Turon H, Oldmeadow C. Which modifiable health risk behaviours are related? A systematic review of the clustering of Smoking, Nutrition, Alcohol and Physical activity ('SNAP') health risk factors. Prev Med 2015 Dec;81:16-41. [doi: 10.1016/j.ypmed.2015.07.003] [Medline: 26190368]
4. Lee M, Lee H, Kim Y, et al. Mobile app-based health promotion programs: a systematic review of the literature. Int J Environ Res Public Health 2018 Dec 13;15(12):2838. [doi: 10.3390/ijerph15122838] [Medline: 30551555]
5. Leuzzi G, Recenti F, Giardulli B, Scafoglieri A, Testa M. Exploring digital health: a qualitative study on adults' experiences with health apps and wearables. Int J Qual Stud Health Well-being 2025 Dec 31;20(1). [doi: 10.1080/17482631.2024.2447096]
6. Measures of health status, quality of life and health care. Health Knowledge. URL: https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1c-health-care-evaluation-health-care-assessment/measures-health-status [accessed 2024-03-19]
7. Randazzo V, Ferretti J, Pasero E. Anytime ECG monitoring through the use of a low-cost, user-friendly, wearable device. Sensors (Basel) 2021 Sep 9;21(18):6036. [doi: 10.3390/s21186036] [Medline: 34577247]
8. Al-Rawashdeh M, Keikhosrokiani P, Belaton B, Alawida M, Zwiri A. IoT adoption and application for smart healthcare: a systematic review. Sensors (Basel) 2022 Jul 19;22(14):5377. [doi: 10.3390/s22145377] [Medline: 35891056]
9. Park YT. Emerging new era of mobile health technologies. Healthc Inform Res 2016 Oct;22(4):253-254. [doi: 10.4258/hir.2016.22.4.253] [Medline: 27895955]
10. Thornton L, Gardner LA, Osman B, et al. A multiple health behavior change, self-monitoring mobile app for adolescents: development and usability study of the Health4Life app. JMIR Form Res 2021 Apr 12;5(4):e25513. [doi: 10.2196/25513] [Medline: 33843590]
11. Leuzzi G, Job M, Scafoglieri A, Testa M. Smartphone Apps and Wearables for Health Parameters in Young Adulthood: Cross-Sectional Study. JMIR Hum Factors ;12:e64629-e64629. [doi: 10.2196/64629]
12. Smuck M, Odonkor CA, Wilt JK, Schmidt N, Swiernik MA. The emerging clinical role of wearables: factors for successful implementation in healthcare. NPJ Digit Med 2021 Mar 10;4(1):45. [doi: 10.1038/s41746-021-00418-3] [Medline: 33692479]
13. Slazus C, Ebrahim Z, Koen N. Mobile health apps: an assessment of needs, perceptions, usability, and efficacy in changing dietary choices. Nutrition 2022 Sep;101:111690. [doi: 10.1016/j.nut.2022.111690] [Medline: 35660502]
14. Nardi W, Roy A, Dunsiger S, Brewer J. Analyzing the impact of mobile app engagement on mental health outcomes: secondary analysis of the unwinding anxiety program. J Med Internet Res 2022 Aug 15;24(8):e33696. [doi: 10.2196/33696] [Medline: 35969440]
15. Ruf A, Koch ED, Ebner-Priemer U, Knopf M, Reif A, Matura S. Studying microtemporal, within-person processes of diet, physical activity, and related factors using the APPetite-mobile-app: feasibility, usability, and validation study. J Med Internet Res 2021 Jul 5;23(7):e25850. [doi: 10.2196/25850] [Medline: 34342268]
16. Xu H, Long H. The effect of smartphone app-based interventions for patients with hypertension: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Oct 19;8(10):e21759. [doi: 10.2196/21759] [Medline: 33074161]
17. Sarker MHR, Moriyama M, Rashid HU, et al. Chronic kidney disease awareness campaign and mobile health education to improve knowledge, quality of life, and motivation for a healthy lifestyle among patients with chronic kidney disease in Bangladesh: randomized controlled trial. J Med Internet Res 2022 Aug 11;24(8):e37314. [doi: 10.2196/37314] [Medline: 35969429]
18. Bylappa BK, Kamath DY, Josephine IS, et al. Usability and feasibility assessment of a smartphone application (Suhriday) for heart failure self-care remote monitoring in an Indian tertiary health care setting: a pilot mixed-methods study. BMJ Open 2022 Aug 24;12(8):e056962. [doi: 10.1136/bmjopen-2021-056962] [Medline: 36002201]
19. Al-Nawaiseh HK, McIntosh WA, McKyer LJ. An-m-health intervention using smartphone app to improve physical activity in college students: a randomized controlled trial. Int J Environ Res Public Health 2022 Jun 13;19(12):7228. [doi: 10.3390/ijerph19127228] [Medline: 35742477]
20. Edwards D, Williams J, Carrier J, Davies J. Technologies used to facilitate remote rehabilitation of adults with deconditioning, musculoskeletal conditions, stroke, or traumatic brain injury: an umbrella review. JBI Evid Synth 2022 Aug 1;20(8):1927-1968. [doi: 10.11124/JBIES-21-00241] [Medline: 35971198]
21. Ek A, Alexandrou C, Söderström E, et al. Effectiveness of a 3-month mobile phone-based behavior change program on active transportation and physical activity in adults: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jun 8;8(6):e18531. [doi: 10.2196/18531] [Medline: 32510462]
22. ANZCTR - Registration. Australian New Zealand Clinical Trials Registry (ANZCTR). URL: https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=377945&showOriginal=true&isReview=true [accessed 2022-10-21]
23. Khan R, Khan S, Almohaimeed HM, Almars AI, Pari B. Utilization, challenges, and training needs of digital health technologies: perspectives from healthcare professionals. Int J Med Inform 2025 May;197:105833. [doi: 10.1016/j.ijmedinf.2025.105833] [Medline: 39954392]
24. Wang Y, Wu T, Chen Z. Active usage of mobile health applications: cross-sectional study. J Med Internet Res 2021 Dec 22;23(12):e25330. [doi: 10.2196/25330] [Medline: 34941545]
25. Schiavo JH. PROSPERO: an international register of systematic review protocols. Med Ref Serv Q 2019;38(2):171-180. [doi: 10.1080/02763869.2019.1588072] [Medline: 31173570]

XSL•FO
RenderX

26. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
27. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. Syst Rev 2021 Jan 26;10(1):39. [doi: 10.1186/s13643-020-01542-z] [Medline: 33499930]
28. Higgins JPT, Thomas J, Chandler J, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 63 (Updated February 2022): Cochrane; 2022. URL: https://www.training.cochrane.org/handbook/archive/v6.3 [accessed 2026-01-02]
29. Covidence. URL: https://www.covidence.org/ [accessed 2026-01-02]
30. Zongpa TC, Chandrasekaran B, Arumugam A. Effectiveness of a smartphone directed physical activity program on cardiometabolic disease risk in desk based office employees. a pragmatic, two arm, parallel, cluster randomised trial. Muscles Ligaments Tendons J 2020;10(4):713-723. [doi: 10.32098/mltj.04.2020.19]
31. Muntaner-Mas A, Sanchez-Azanza VA, Ortega FB, et al. The effects of a physical activity intervention based on a fatness and fitness smartphone app for University students. Health Informatics J 2021;27(1):1460458220987275. [doi: 10.1177/1460458220987275] [Medline: 33446036]
32. RoB 2: a revised Cochrane risk-of-bias tool for randomized trials. Cochraine Bias. URL: https://methods.cochrane.org/bias/resources/rob-2-revised-cochrane-risk-bias-tool-randomized-trials [accessed 2022-08-30]
33. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
34. Jamovi (version 2.3.28). jamovi. URL: https://www.jamovi.org [accessed 2026-01-07]
35. GRADE approach. Cochrane. URL: https://training.cochrane.org/grade-approach [accessed 2022-09-26]
36. Wilson D, Driller MW, Johnston B, Gill ND. A contactless app-based intervention to improve health behaviors in airline pilots: a randomized trial. Am J Prev Med 2023 May;64(5):666-676. [doi: 10.1016/j.amepre.2022.12.011] [Medline: 36641335]
37. Fanning J, Roberts S, Hillman CH, Mullen SP, Ritterband L, McAuley E. A smartphone "app"-delivered randomized factorial trial targeting physical activity in adults. J Behav Med 2017 Oct;40(5):712-729. [doi: 10.1007/s10865-017-9838-y] [Medline: 28255750]
38. Edney SM, Olds TS, Ryan JC, et al. A social networking and gamified app to increase physical activity: cluster RCT. Am J Prev Med 2020 Feb;58(2):e51-e62. [doi: 10.1016/j.amepre.2019.09.009] [Medline: 31959326]
39. Tulasiram B, Chandrasekaran B. Are smartphones better in guiding physical activity among sedentary young adults? A randomised controlled trial. Muscles Ligaments Tendons J 2021;11(1):77. [doi: 10.32098/mltj.01.2021.10]
40. Mamede A, Noordzij G, Jongerling J, Snijders M, Schop-Etman A, Denktas S. Combining web-based gamification and physical nudges with an app (MoveMore) to promote walking breaks and reduce sedentary behavior of office workers: field study. J Med Internet Res 2021 Apr 12;23(4):e19875. [doi: 10.2196/19875] [Medline: 33843593]
41. Balk-Møller NC, Poulsen SK, Larsen TM. Effect of a nine-month web- and app-based workplace intervention to promote healthy lifestyle and weight loss for employees in the social welfare and health care sector: a randomized controlled trial. J Med Internet Res 2017 Apr 10;19(4):e108. [doi: 10.2196/jmir.6196] [Medline: 28396303]
42. Ek A, Alexandrou C, Delisle Nyström C, et al. The Smart City Active Mobile Phone Intervention (SCAMPI) study to promote physical activity through active transportation in healthy adults: a study protocol for a randomised controlled trial. BMC Public Health 2018 Jul 16;18(1):880. [doi: 10.1186/s12889-018-5658-4] [Medline: 30012116]
43. Glynn LG, Hayes PS, Casey M, et al. Effectiveness of a smartphone application to promote physical activity in primary care: the SMART MOVE randomised controlled trial. Br J Gen Pract 2014 Jul;64(624):e384-e391. [doi: 10.3399/bjgp14X680461] [Medline: 24982490]
44. Cai X, Qiu S, Luo D, et al. Effects of peer support and mobile application-based walking programme on physical activity and physical function in rural older adults: a cluster randomized controlled trial. Eur Geriatr Med 2022 Oct;13(5):1187-1195. [doi: 10.1007/s41999-022-00682-w] [Medline: 36001254]
45. King AC, Hekler EB, Grieco LA, et al. Effects of three motivationally targeted mobile device applications on initial physical activity and sedentary behavior change in midlife and older adults: a randomized trial. PLoS ONE 2016;11(6):e0156370. [doi: 10.1371/journal.pone.0156370] [Medline: 27352250]
46. Murawski B, Plotnikoff RC, Rayward AT, et al. Efficacy of an m-health physical activity and sleep health intervention for adults: a randomized waitlist-controlled trial. Am J Prev Med 2019 Oct;57(4):503-514. [doi: 10.1016/j.amepre.2019.05.009] [Medline: 31542128]
47. van Drongelen A, Boot CR, Hlobil H, Twisk JW, Smid T, van der Beek AJ. Evaluation of an mHealth intervention aiming to improve health-related behavior and sleep and reduce fatigue among airline pilots. Scand J Work Environ Health 2014 Nov;40(6):557-568. [doi: 10.5271/sjweh.3447] [Medline: 25121620]
48. Zhou M, Fukuoka Y, Mintz Y, et al. Evaluating machine learning-based automated personalized daily step goals delivered through a mobile phone app: randomized controlled trial. JMIR Mhealth Uhealth 2018 Jan 25;6(1):e28. [doi: 10.2196/mhealth.9117] [Medline: 29371177]

XSL•FO
RenderX

49. Baumann H, Heuel L, Bischoff LL, Wollesen B. Efficacy of individualized sensory-based mhealth interventions to improve distress coping in healthcare professionals: a multi-arm parallel-group randomized controlled trial. Sensors (Basel) 2023 Feb 19;23(4):2322. [doi: 10.3390/s23042322] [Medline: 36850920]
50. Arrogi A, Bogaerts A, Seghers J, et al. Evaluation of stAPP: a smartphone-based intervention to reduce prolonged sitting among Belgian adults. Health Promot Int 2019 Feb 1;34(1):16-27. [doi: 10.1093/heapro/dax046] [Medline: 28973149]
51. Gabbiadini A, Greitemeyer T. Fitness mobile apps positively affect attitudes, perceived behavioral control and physical activities. J Sports Med Phys Fitness 2019 Mar;59(3):407-414. [doi: 10.23736/S0022-4707.18.08260-9] [Medline: 29619794]
52. Murawski B, Plotnikoff RC, Lubans DR, et al. Examining mediators of intervention efficacy in a randomised controlled m-health trial to improve physical activity and sleep health in adults. Psychol Health 2020 Nov 1;35(11):1346-1367. [doi: 10.1080/08870446.2020.1756288]
53. Saran T, Pedrycz A, Mucha D, Mucha D. Follow-up monitoring of physical activity after rehabilitation by means of a mobile application: effectiveness of measurements in different age groups. Adv Clin Exp Med 2018 Aug;27(8):1037-1044. [doi: 10.17219/acem/69131] [Medline: 29962117]
54. Gremaud AL, Carr LJ, Simmering JE, et al. Gamifying accelerometer use increases physical activity levels of sedentary office workers. J Am Heart Assoc 2018 Jul 2;7(13):e007735. [doi: 10.1161/JAHA.117.007735] [Medline: 29967221]
55. Skvortsova A, Rodrigues TC, de Buisonjé D, et al. Increasing the effectiveness of a physical activity smartphone intervention with positive suggestions: randomized controlled trial. J Med Internet Res 2022 Mar 1;24(3):e32130. [doi: 10.17605/OSF.IO/CWJES]
56. Yoshimura E, Tajiri E, Michiwaki R, Matsumoto N, Hatamoto Y, Tanaka S. Long-term effects of the use of a step count-specific smartphone app on physical activity and weight loss: randomized controlled clinical trial. JMIR Mhealth Uhealth 2022 Oct 24;10(10):e35628. [doi: 10.2196/35628] [Medline: 36279159]
57. Martin SS, Feldman DI, Blumenthal RS, et al. mActive: a randomized clinical trial of an automated mhealth intervention for physical activity promotion. J Am Heart Assoc 2015 Nov 9;4(11):e002239. [doi: 10.1161/JAHA.115.002239] [Medline: 26553211]
58. Kim Y, Lee H, Chung ML. Living labs for a mobile app-based health program: effectiveness of a 24-week walking intervention for cardiovascular disease risk reduction among female Korean-Chinese migrant workers: a randomized controlled trial. Arch Public Health 2022 Aug 4;80(1):181. [doi: 10.1186/s13690-022-00941-z] [Medline: 35927769]
59. Plotnikoff RC, Jansson AK, Duncan MJ, et al. mHealth to support outdoor gym resistance training: the ecofit effectiveness RCT. Am J Prev Med 2023 Jun;64(6):853-864. [doi: 10.1016/j.amepre.2023.01.031] [Medline: 36804197]
60. Fukuoka Y, Haskell W, Lin F, Vittinghoff E. Short- and long-term effects of a mobile phone app in conjunction with brief in-person counseling on physical activity among physically inactive women: the mPED randomized clinical trial. JAMA Netw Open 2019 May 3;2(5):e194281. [doi: 10.1001/jamanetworkopen.2019.4281] [Medline: 31125101]
61. Spring B, Schneider K, McFadden HG, et al. Multiple behavior changes in diet and activity: a randomized controlled trial using mobile technology. Arch Intern Med 2012 May 28;172(10):789-796. [doi: 10.1001/archinternmed.2012.1044] [Medline: 22636824]
62. Shcherbina A, Hershman SG, Lazzeroni L, et al. The effect of digital physical activity interventions on daily step count: a randomised controlled crossover substudy of the MyHeart Counts Cardiovascular Health Study. Lancet Digit Health 2019 Nov;1(7):e344-e352. [doi: 10.1016/S2589-7500(19)30129-3] [Medline: 33323209]
63. Mönninghoff A, Fuchs K, Wu J, Albert J, Mayer S. The effect of a future-self avatar mobile health intervention (FutureMe) on physical activity and food purchases: randomized controlled trial. J Med Internet Res 2022 Jul 7;24(7):e32487. [doi: 10.2196/32487] [Medline: 35797104]
64. Gür F, Gür GC, Ayan V. The effect of the ERVE smartphone app on physical activity, quality of life, self-efficacy, and exercise motivation for inactive people: a randomized controlled trial. Eur J Integr Med 2020 Oct;39:101198. [doi: 10.1016/j.eujim.2020.101198]
65. Schroé H, Van Dyck D, De Paepe A, et al. Which behaviour change techniques are effective to promote physical activity and reduce sedentary behaviour in adults: a factorial randomized trial of an e- and m-health intervention. Int J Behav Nutr Phys Act 2020 Oct 7;17(1):127. [doi: 10.1186/s12966-020-01001-x] [Medline: 33028335]
66. Safran Naimark J, Madar Z, Shahar DR. The impact of a Web-based app (eBalance) in promoting healthy lifestyles: randomized controlled trial. J Med Internet Res 2015 Mar 2;17(3):e56. [doi: 10.2196/jmir.3682] [Medline: 25732936]
67. Damschroder LJ, Buis LR, McCant FA, et al. Effect of adding telephone-based brief coaching to an mHealth app (Stay Strong) for promoting physical activity among veterans: randomized controlled trial. J Med Internet Res 2020 Aug 4;22(8):e19216. [doi: 10.2196/19216] [Medline: 32687474]
68. MacPherson MM, Merry KJ, Locke SR, Jung ME. Effects of mobile health prompts on self-monitoring and exercise behaviors following a diabetes prevention program: secondary analysis from a randomized controlled trial. JMIR Mhealth Uhealth 2019 Sep 5;7(9):e12956. [doi: 10.2196/12956] [Medline: 31489842]
69. Olson R, Wipfli B, Hanson GC, et al. Weight loss maintenance among truck drivers in the SHIFT randomised controlled trial, USA. Occup Environ Med 2025 Jul 9;82(4):168-175. [doi: 10.1136/oemed-2024-109903] [Medline: 40355262]

XSL•FO
RenderX

70. Simmering JE, Polgreen LA, Francis SL, Strom AJ, Segre AM, Polgreen PM. Using a Fitbit-based walking game to improve physical activity among U.S. veterans. Mil Med 2025 Jan 16;190(1-2):194-201. [doi: 10.1093/milmed/usae280] [Medline: 38829720]
71. Mateo-Orcajada A, Vaquero-Cristóbal R, Mota J, Abenza-Cano L. Training and detraining effects of a physical activity program implemented through mobile applications in adolescents. ClinicalTrials.gov. 2023. URL: https://clinicaltrials.gov/study/NCT06164041 [accessed 2026-01-02]
72. Ryan DJ, Ross MH, Simmich J, et al. TRACK & ACT: a pragmatic randomised controlled trial exploring the comparative effectiveness of pedometers and activity trackers for changing physical activity and sedentary behaviour in inactive individuals. J Act Sedent Sleep Behav 2023 May 1;2(1):12. [doi: 10.1186/s44167-023-00018-4] [Medline: 40217539]
73. Collombon EHGM, Bolman CAW, de Bruijn GJ, Peels DA, Verboon P, Lechner L. The efficacy of online physical activity interventions with added mobile elements within adults aged 50 years and over: randomized controlled trial. Appl Psychol Health Well Being 2024 Nov;16(4):1921-1943. [doi: 10.1111/aphw.12568] [Medline: 38925643]
74. Lugade V, Torbitt M, O'Brien SR, Silsupadol P. Smartphone- and paper-based delivery of balance intervention for older adults are equally effective, enjoyable, and of high fidelity: a randomized controlled trial. Sensors (Basel) 2023 Aug 27;23(17):7451. [doi: 10.3390/s23177451] [Medline: 37687907]
75. Javed A, Kim DS, Hershman SG, et al. Personalized digital behaviour interventions increase short-term physical activity: a randomized control crossover trial substudy of the MyHeart Counts Cardiovascular Health Study. Eur Heart J Digit Health 2023 Oct;4(5):411-419. [doi: 10.1093/ehjdh/ztad047] [Medline: 37794870]
76. Ghazala S, Veluswamy SK, Ravindra S, Arena R, Myers J. Efficacy of mHealth-based workplace health promotion strategy in improving cardiorespiratory fitness in a healthcare setting: a randomized controlled study. J Occup Environ Med 2024 Dec 1;66(12):1083-1090. [doi: 10.1097/JOM.0000000000003229] [Medline: 39322289]
77. Zhang N, Zhou M, Li M, Ma G. Effects of smartphone-based remote interventions on dietary intake, physical activity, weight control, and related health benefits among the older population with overweight and obesity in China: randomized controlled trial. J Med Internet Res 2023 Apr 28;25:e41926. [doi: 10.2196/41926] [Medline: 37115608]
78. Sanchez-Trigo H, Maher C, Godino JG, Sañudo B. Effects of an mHealth physical activity intervention to prevent osteoporosis in premenopausal women. A randomized controlled trial. J Sci Med Sport 2023 Oct;26(10):545-552. [doi: 10.1016/j.jsams.2023.09.004] [Medline: 37739855]
79. Pettersson B, Lundin-Olsson L, Skelton DA, et al. Effectiveness of the safe step digital exercise program to prevent falls in older community-dwelling adults: randomized controlled trial. J Med Internet Res 2025 Mar 31;27:e67539. [doi: 10.2196/67539] [Medline: 40163860]
80. Chandrasekaran B, Rao CR, Pesola AJ, Arumugam A. Effectiveness of technology-assisted and self-directed interventions to sit less and move more among Indian desk-based office workers: a three-arm cluster randomised controlled trial (SMART-STEP trial). Appl Ergon 2025 Sep;127:104528. [doi: 10.1016/j.apergo.2025.104528] [Medline: 40199231]
81. Pomkai N, Potharin D, Widyastari DA, et al. Effectiveness of an mHealth application for physical activity promotion among Thai older adults: a randomized controlled trial. Inquiry 2024;61:469580241309869. [doi: 10.1177/00469580241309869] [Medline: 39718167]
82. Gómez-Cuesta N, Mateo-Orcajada A, Meroño L, Abenza-Cano L, Vaquero-Cristóbal R. A mobile app-based intervention improves anthropometry, body composition and fitness, regardless of previous active-inactive status: a randomized controlled trial. Front Public Health 2024;12:1380621. [doi: 10.3389/fpubh.2024.1380621] [Medline: 39193194]
83. Ortega FB, Ruiz JR, España-Romero V, et al. The International Fitness Scale (IFIS): usefulness of self-reported fitness in youth. Int J Epidemiol 2011 Jun;40(3):701-711. [doi: 10.1093/ije/dyr039] [Medline: 21441238]
84. McGuinness LA, Higgins JPT. Risk-of-bias VISualization (robvis): an R package and Shiny web app for visualizing risk-of-bias assessments. Res Synth Methods 2021 Jan;12(1):55-61. [doi: 10.1002/jrsm.1411] [Medline: 32336025]
85. Beatty AL, Fukuoka Y, Whooley MA. Using mobile technology for cardiac rehabilitation: a review and framework for development and evaluation. J Am Heart Assoc 2013 Nov 1;2(6):e000568. [doi: 10.1161/JAHA.113.000568] [Medline: 24185949]
86. Car J, Gurol-Urganci I, de Jongh T, Vodopivec-Jamsek V, Atun R. Mobile phone messaging reminders for attendance at healthcare appointments. Cochrane Database Syst Rev 2012 Jul 11(7):CD007458. [doi: 10.1002/14651858.CD007458.pub2] [Medline: 22786507]
87. de Jongh T, Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst Rev 2012 Dec 12;12(12):CD007459. [doi: 10.1002/14651858.CD007459.pub2] [Medline: 23235644]
88. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52. [doi: 10.2196/jmir.3951] [Medline: 25803266]
89. Bloomfield GS, Vedanthan R, Vasudevan L, Kithei A, Were M, Velazquez EJ. Mobile health for non-communicable diseases in Sub-Saharan Africa: a systematic review of the literature and strategic framework for research. Global Health 2014 Jun 13;10(1):49. [doi: 10.1186/1744-8603-10-49] [Medline: 24927745]

XSL•FO
RenderX

90. Worringham C, Rojek A, Stewart I. Development and feasibility of a smartphone, ECG and GPS based system for remotely monitoring exercise in cardiac rehabilitation. PLoS ONE 2011 Feb 9;6(2):e14669. [doi: 10.1371/journal.pone.0014669] [Medline: 21347403]
91. Fukui K, Suzuki Y, Kaneda K, et al. Do “Stay-at-Home Exercise” videos induce behavioral changes in college students? A randomized controlled trial. Sustainability 2021 Nov 1;13(21):11600. [doi: 10.3390/su132111600]
92. Aranda-Jan CB, Mohutsiwa-Dibe N, Loukanova S. Systematic review on what works, what does not work and why of implementation of mobile health (mHealth) projects in Africa. BMC Public Health 2014 Feb 21;14(1):188. [doi: 10.1186/1471-2458-14-188] [Medline: 24555733]
93. Bacigalupo R, Cudd P, Littlewood C, Bissell P, Hawley MS, Buckley Woods H. Interventions employing mobile technology for overweight and obesity: an early systematic review of randomized controlled trials. Obes Rev 2013 Apr;14(4):279-291. [doi: 10.1111/obr.12006] [Medline: 23167478]
94. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]
95. Free C, Phillips G, Watson L, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013;10(1):e1001363. [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
96. Banks AZ, Mentz RJ, Stebbins A, et al. Response to exercise training and outcomes in patients with heart failure and diabetes mellitus: insights from the HF-ACTION trial. J Card Fail 2016 Jul;22(7):485-491. [doi: 10.1016/j.cardfail.2015.12.007] [Medline: 26687984]
97. Bohannon RW. Minimal clinically important difference for grip strength: a systematic review. J Phys Ther Sci 2019 Jan;31(1):75-78. [doi: 10.1589/jpts.31.75] [Medline: 30774209]

Abbreviations

GRADE: Grading of Recommendations, Assessment, Development and Evaluation
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses
PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Literature Search Extension
RCT: randomized controlled trial
RoB: risk of bias
VO_2 max: maximal oxygen consumption

Edited by A Stone; submitted 25.Sep.2024; peer-reviewed by S Werts-Pelter, WJ He; revised version received 17.Nov.2025; accepted 17.Nov.2025; published 16.Jan.2026.

Please cite as:
Leuzzi G, Job M, Cosentino C, Testa R, Scafoglieri A, Testa M
Mobile Apps to Improve Health Parameters in Healthy Adults: Systematic Review
JMIR Mhealth Uhealth 2026;14:e66881
URL: https://mhealth.jmir.org/2026/1/e66881
doi:10.2196/66881

© Gaia Leuzzi, Mirko Job, Carola Cosentino, Riccardo Testa, Aldo Scafoglieri, Marco Testa. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 16.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Advancements in Wearable Sensor Technologies for Health Monitoring in Terms of Clinical Applications, Rehabilitation, and Disease Risk Assessment: Systematic Review

Bonsang Gu[1], MS; Hyeon Su Kim[1,2], BEng; HyunBin Kim[2], BEng; Jun-Il Yoo[3], MD, PhD

[1]Department of Biomedical Research Institute, Inha University Hospital, Incheon, Republic of Korea

[2]Program in Biomedical Science and Engineering, Inha University, Incheon, Republic of Korea

[3]Department of Orthopedic Surgery, Inha University Hospital, Inha University College of Medicine, Incheon, Incheon, Republic of Korea

Corresponding Author:
Jun-Il Yoo, MD, PhD
Department of Orthopedic Surgery
Inha University Hospital, Inha University College of Medicine
27, Inhang-ro, Jung-gu
Incheon, Incheon
Republic of Korea
Phone: 82 10 3242 4980
Email: furim@daum.net

Abstract

Background: Wearable sensor technologies such as inertial measurement units, smartwatches, and multisensor systems have emerged as valuable tools in clinical and real-world health monitoring. These devices enable continuous, noninvasive tracking of gait, mobility, and functional health across diverse populations. However, challenges remain in sensor placement standardization, data processing consistency, and real-world validation.

Objective: This systematic review aimed to evaluate recent literature on the clinical and research applications of wearable sensors. Specifically, it investigated how these technologies are used to assess mobility, predict disease risk, and support rehabilitation. It also identified limitations and proposed future research directions.

Methods: This review was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. We searched the PubMed, Scopus, and Web of Science databases up to March 9, 2025. Inclusion criteria focused on studies using wearable sensors in clinical or real-world environments. A total of 30 eligible studies were identified for qualitative synthesis. Data extracted included study design, population characteristics, sensor type and placement, machine learning algorithms, and clinical outcomes.

Results: Of the included studies, 43% (13/30) were observational, 27% (8/30) were experimental, and 10% (3/30) were randomized controlled trials. Inertial measurement unit–based sensors were used in 67% (20/30) of the studies, with wrist-worn devices being the most common (13/20, 65%). Machine learning techniques were frequently applied, with random forest (6/30, 20%) and deep learning (5/30, 17%) models predominating. Clinical applications spanned Parkinson disease, stroke, multiple sclerosis, and frailty, with several studies (4/30, 13%) reporting high predictive accuracy for fall risk and mobility decline (area under the receiver operating characteristic curve up to 0.97).

Conclusions: Wearable sensors show strong potential for mobility monitoring, disease risk assessment, and rehabilitation tracking in clinical and real-world settings. However, challenges remain in standardizing sensor protocols and data analysis. Future research should focus on large-scale, longitudinal studies; harmonized machine learning pipelines; and integration with cloud-based health systems to improve scalability and clinical translation.

(JMIR Mhealth Uhealth 2026;14:e76084) doi:10.2196/76084

KEYWORDS

wearable sensors; health monitoring; systematic review; gait analysis; rehabilitation; mobile health; mHealth

XSL•FO
RenderX

Introduction

Throughout this paper, we use standardized terminology to ensure clarity and consistency. The term "wearable sensor" refers to any body-worn device capable of measuring physiological or biomechanical parameters, including inertial measurement unit (IMU)–based sensors, smartwatches, smart insoles, and multisensor systems. "IMU" refers specifically to devices incorporating accelerometers, gyroscopes, and magnetometers for motion tracking. When referring to specific device types (eg, smartwatches and smart insoles), we use the precise terminology to distinguish their unique features and applications.

Wearable sensors have gained significant attention in clinical research and health care for their ability to provide continuous, real-world assessments of mobility and physiological health. These devices, including IMU-based sensors, smartwatches, and multisensor systems, have transformed traditional gait and activity monitoring by enabling remote, noninvasive tracking of movement patterns and health status [1]. The integration of advanced analytics, particularly machine learning (ML), has further enhanced their diagnostic and predictive capabilities, positioning wearable sensors as key tools in digital health and precision medicine [2].

Gait analysis and mobility tracking have been central to wearable sensor applications, particularly in neurological, musculoskeletal, and age-related conditions. In Parkinson disease (PD), wearable sensors have been used to detect subtle changes in gait speed, stance and swing phase durations, and postural instability, aiding in early disease detection and progression monitoring. In stroke rehabilitation, these sensors enable remote motor recovery assessment and provide continuous mobility data outside traditional clinical settings [3]. Wearable sensors also demonstrate high efficacy in frailty assessment and fall risk prediction, offering objective, real-time alternatives to conventional tools such as the Performance-Oriented Mobility Assessment (POMA) and Timed Up and Go (TUG) tests.

Despite their growing clinical adoption, several challenges hinder the widespread implementation of wearable sensor technology. Variability in sensor placement, study methodologies, and data processing techniques limits cross-study comparability and reproducibility [3]. Additionally, while controlled laboratory studies have validated their accuracy, real-world validation remains insufficient, necessitating further large-scale, longitudinal studies to assess their usability and reliability across diverse populations [4]. Furthermore, standardization of ML frameworks and data interpretation methodologies is essential to ensure consistent clinical application [5].

This systematic review aimed to provide a comprehensive evaluation of wearable sensor research, analyzing their clinical applications, technological advancements, and methodological challenges. By synthesizing evidence from recent studies, we highlight key trends in wearable sensor use, discuss their implications for health care, and propose future directions to enhance their impact in mobility monitoring and rehabilitation.

Methods

Study Design

The protocol of this review was not registered in PROSPERO due to its exploratory nature and inclusion of emerging sensor studies. However, the review process followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidance to ensure methodological transparency. The study selection process followed PRISMA 2020 guidelines, including identification, screening, eligibility assessment, and final inclusion. This review focused on studies published in the last 10 years that investigated the applications and effectiveness of wearable sensors, including smartwatches, in remote health monitoring, rehabilitation, and disease assessment. Full-text articles were included to ensure a comprehensive analysis. We aimed to synthesize evidence on clinical and research applications of wearable sensors, particularly for gait analysis, fall risk assessment, and disease monitoring. Given study heterogeneity, we categorized and synthesized the findings narratively, emphasizing disease-specific insights and sensor use trends.

Search Strategy

A comprehensive database search was conducted across PubMed, Scopus, and Web of Science.

The search strategy combined terms related to wearable technologies, inertial sensors, digital biomarkers, and rehabilitation. Representative Boolean operators were used as follows:

> *("smartwatch" OR "smart watch" OR "wearable sensor" OR "wearable sensor") AND ("accelerometer" OR "acceleration sensor" OR "inertial sensor" OR "IMU") AND ("remote monitoring" OR "digital biomarkers" OR "telemedicine" OR "wearable health tracking") AND (aging OR older adults OR elderly OR Parkinson OR stroke OR "gait disorders" OR "neurological disorders" OR "movement disorders" OR "fall risk" OR rehabilitation OR "functional mobility" OR sarcopenia OR osteoarthritis OR dementia) NOT review.*

The initial search yielded 4226 records. Of these 4226 records, after removing duplicates and studies unrelated to clinical applications (n=3664, 86.7%), 562 (13.3%) remained for screening. Of these 562 studies, those focusing solely on technical performance comparisons (n=501, 89.1%) were excluded, leaving 61 (10.9%) for eligibility assessment. Of these 61 studies, an additional 31 (51%) were excluded due to limited relevance to disease-related applications, resulting in 30 (49%) studies included in the final review. The search was finalized on March 9, 2025. In total, 30 studies met the inclusion criteria and were included in the final synthesis. Additional references cited throughout the manuscript (n=43) were used for background, context, and methodological justification.

XSL•FO
RenderX

Study Selection Process

A single researcher conducted study selection and data extraction following a predefined protocol to minimize bias. The eligibility criteria were clearly defined and consistently applied. Studies published from March 9, 2015, to March 9, 2025, were considered, reflecting a 10-year search window. Eligible study designs included randomized controlled trials (RCTs), observational studies, and experimental validation studies conducted in either clinical or real-world settings. Both research-grade and commercial wearable sensors were included provided that they reported measurable health or functional outcomes. Only English-language, peer-reviewed articles were included, and conference abstracts, reviews, and purely technical feasibility reports without human participants were excluded. Quality appraisal using the Newcastle-Ottawa Scale (NOS), Joanna Briggs Institute (JBI) appraisal tools, or version 2 of the Cochrane risk-of-bias tool for randomized trials (RoB 2) was conducted to describe study rigor but did not influence inclusion decisions.

Any uncertainties during the selection process were resolved by re-evaluating studies against the predefined inclusion criteria. Data extraction was conducted manually using a standardized form. No independent reviewers cross-checked the extracted data, which is acknowledged as a limitation. Missing or unclear data were clarified when possible, by contacting the corresponding authors. No automation tools were used for data collection.

Participant Selection in the Included Studies

The included studies targeted diverse populations, including healthy adults; older individuals; and patients diagnosed with neurological disorders (eg, stroke and PD), musculoskeletal disorders (eg, sarcopenia and osteoarthritis), or metabolic conditions (eg, diabetes). Participants could walk independently and provided informed consent. We excluded studies lacking clear participant definitions or standard gait analysis metrics to maintain consistency. Pediatric studies were excluded except for those including toddler cohorts (aged <3 years) and specifically designed for developmental gait analysis. To ensure data quality, studies were required to report a minimum wear time of 30 minutes of valid sensor data per session.

Wearable Sensor Technology

The wearable sensors reviewed featured advanced components such as high-precision accelerometers, gyroscopes, and pressure sensors. These sensors accurately captured key gait and mobility parameters: step length, stance and swing phase durations, plantar pressure distribution, and center of pressure. Smartwatches were primarily used for activity tracking and remote monitoring, whereas wearable sensors and foot-mounted sensors specialized in gait and postural assessments. The wearable sensors integrated seamlessly into daily life, ensuring high usability and real-world applicability.

Data Acquisition and Analysis

Data were collected using a variety of wearable sensor systems, including IMUs, smart insoles, smartwatches, and pressure-sensing devices. Sensor placement varied by study objective and included the wrist, waist, ankle, thigh, lumbar spine, and foot. The IMU sensors incorporated accelerometers, gyroscopes, and magnetometers with sampling rates ranging from 10 to 1149 Hz depending on the device and measurement context. Pressure-sensitive insoles provided additional biomechanical insights through plantar pressure distribution and force-time characteristics.

Wireless data transmission via Bluetooth or cloud platforms enabled real-time monitoring and digital biomarker extraction. Embedded preprocessing algorithms were applied to reduce noise, improve signal quality, and enhance feature extraction accuracy. Studies used various ML techniques, including random forest (6/30, 20%), deep learning (5/30, 17%), elastic net regression (4/30, 13%), and principal component analysis (PCA; 2/30, 7%), for pattern recognition, mobility classification, and disease risk prediction. All reported quantitative values (eg, area under the receiver operating characteristic curve [AUROC], accuracy, and improvement rate) were extracted from individual studies and are presented descriptively, not as pooled estimates.

The methodological quality of the included studies was systematically evaluated using appropriate assessment tools based on the study design. The NOS was applied to prospective cohort studies, whereas the JBI critical appraisal checklist was used for observational, cross-sectional, and experimental studies. For RCTs, the RoB 2 tool was used to ensure a robust evaluation of study quality. The results of the quality assessment guided the interpretation of the reliability and clinical applicability of the findings. Studies were categorized as low (JBI or NOS score of ≥8), moderate (JBI or NOS score of 6-7), or high (JBI or NOS score of ≤5) risk of bias according to established thresholds.

Ethical Considerations

All studies included adhered to ethical standards for human research, following the Declaration of Helsinki. As this study is a systematic review of published literature and did not involve human participants, interventions, or identifiable private data, ethics approval was not required. Data privacy and participant well-being were prioritized across the studies.

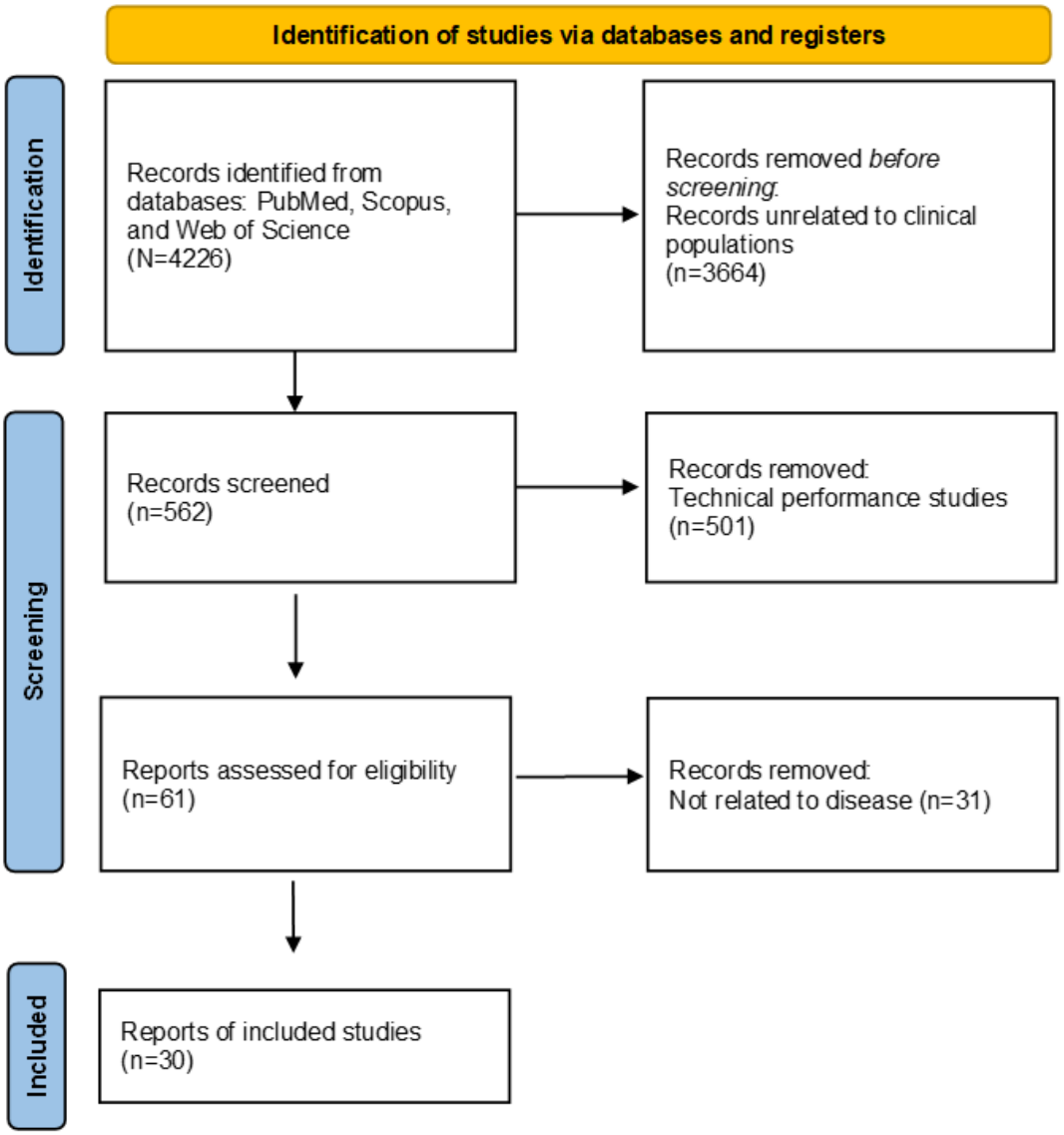
Results

Study Design and Population Characteristics

A total of 4226 records were identified through database searching. After removing duplicates and screening titles and abstracts, 562 articles remained for full-text assessment. Of these, 30 studies met the inclusion criteria and were included in the final review. The full screening process is summarized in Figure 1 (PRISMA 2020 flow diagram). Among the analyzed studies, the most frequently used research design was prospective observational studies, accounting for 43% (13/30) of the total. Experimental studies comprised 27% (8/30), whereas RCTs were limited to 10% (3/30). Cross-sectional studies and cohort studies accounted for 13% (4/30) and 7% (2/30), respectively. Observational studies were most common due to their feasibility, whereas experimental studies validated sensor-based assessments. RCTs were scarce, indicating limited rigorous intervention-based evaluations. Cohort studies were included in long-term monitoring applications.

XSL•FO
RenderX

Figure 1. PRISMA flow diagram illustrating the screening process of papers for study selection.



The studies focused on various populations, with healthy adults being the most common participant group, as summarized in Table 1. Studies involving neurodegenerative diseases such as PD, stroke, Huntington disease, and multiple sclerosis (MS) were also prevalent. Healthy adults were often included for sensor validation and reliability assessment, whereas studies on neurodegenerative diseases primarily aimed at mobility and functional monitoring. Research on frailty in older adults (4/30, 13%) focused on mobility assessment, balance, and fall risk analysis. The reviewed studies covered a wide range of age groups and health conditions. Studies involving healthy adults typically included participants from early adulthood to middle age, with some extending to adolescent and pediatric populations. Research on neurodegenerative diseases such as PD and Huntington disease focused primarily on older adults, whereas cardiovascular disease– and frailty-related studies involved older participants as well. Certain studies (12/30, 40%) targeted specific conditions such as osteoarthritis, rheumatoid arthritis, MS, cystic fibrosis, stroke, and spinal cord injury, highlighting the diverse application of wearable sensors, particularly smartwatches, in different clinical populations. The distribution of male and female participants varied across studies, with some (1/30, 3%) focusing specifically on age-based differences in sensor performance and health monitoring.

XSL•FO
RenderX

Table 1. Summary of study populations in the reviewed studies.

Study	Participant age (y)	Number of male/female participants	Health condition
Bolam et al [6], 2021	Mean 66.8 (SD 7.0)	6/8	Knee osteoarthritis
Angelucci and Aliverti [7], 2023	Mean 26.8 (SD NR[a]; range 23-54)	9/11	Healthy adults
Greene et al [8], 2021	PD[b] 1 (clinical cohort): mean 67.3 (SD 7.1); PD 2 (exercise cohort): mean 64.9 (SD 7.3)	27/17	PD
Gordon et al [9], 2019	Mean 51 (SD 12)	9/8	HD[c]
Presley et al [10], 2023	Range 13-24	6/12	Healthy adults and adolescents
De Cannière et al [11], 2020	Mean 63 (SD 1)	65/24	CVD[d] and heart failure
Nunes et al [12], 2024	HD: mean 51.9 (SD 11); pre-HD stage: mean 36.5 (SD 13.1); control: mean 58.9 (SD 12.2)	8/8 (HD group)	HD, pre-HD stage, and controls
Mahadevan et al [13], 2020	Control: mean 43.9 (SD 10); PD: mean 68.1 (SD 8.1)	Control: 27/33; PD: 23/12	PD
Seo et al [14], 2024	Mean 61 (SD 12)	12/7	Stroke (upper-limb hemiparesis)
Odhiambo et al [15], 2023	Mean 27.25 (SD NR; range 20-56)	16/12	Healthy adults
Hwang and Effenberg [16], 2021	Mean 29.8 (SD 6.8)	6/6	Healthy adults
Wu et al [17], 2021	Mean 48 (range 26-66)	2/2	Healthy adults
John and Soangra [18], 2022	Control: mean 74 (SD 8.7); Stroke: mean 69 (SD 8.4)	Control: 6/8; Stroke: 8/6	Stroke vs healthy older adults
Meyer et al [19], 2022	Mean 51 (SD 7)	5/16	Multiple sclerosis
Toumieux et al [20], 2015	Not specified	Not specified	Stroke
Elstub et al [21], 2022	Not specified	6/3	Healthy runners
Perraudin et al [22], 2018	RA[e]: mean 50.7 (SD 11.4); PA[f]: mean 47.5 (SD 15.5); OA[g]: mean 60.7 (SD 4.5); healthy: mean 48 (SD 13.6)	RA: 18 female; PA: 2 female; OA: 10 female; healthy: 15 female	RA, PA, and OA
Giggins et al [23], 2025	Mean 77.5 (SD 8.4)	12/39	Frailty levels
Savi et al [24], 2020	Mean 37.5 (SD 11.5)	6/18	Cystic fibrosis
Haghi et al [25], 2023	Mean 26.7 (SD NR)	22/8	Healthy adults
DasMahapatra et al [26], 2018	Mean 52 (SD 10.6)	28/86	Multiple sclerosis
Sun et al [27], 2019	OA: mean 78.2 (SD 6.1); YA[h]: mean 24.4 (SD 3.9)	OA: 4/6; YA: 6/6	Healthy adults
Ramezani et al [28], 2019	Community: mean 82.16 (SD 9.55); hospital: mean 84.22 (SD 13.87)	Community: 41/104; hospital: 5/4	Older adults with multiple chronic conditions
Liew et al [29], 2024	Mean 20.8 (SD 1.6)	9/9	Healthy adults
Kwon et al [30], 2019	13-35 months (1-year-olds: 11; 2-year-olds: 11)	10/12	Healthy toddlers
Martin et al [31], 2015	Mean 58 (SD 8; range 18-69)	26/22	CVD prevention and obesity
Hup et al [32], 2024	Various	Not specified	OHCA[i] and SCA[j]
Browne et al [33], 2020	Mean 13.3 (SD 2.3)	9/11	Pediatric obesity
Burns et al [34], 2020	≥18	Not specified	Rotator cuff and shoulder pain
Bailey et al [35], 2024	Mean 51 (SD 9; range 28-63)	11/4	Spinal cord injury and wheelchair users

[a]NR: not reported.

[b]PD: Parkinson disease.

XSL•FO
RenderX

[c]HD: Huntington disease.
[d]CVD: cardiovascular disease.
[e]RA: rheumatoid arthritis.
[f]PA: psoriatic arthritis.
[g]OA: old adult.
[h]YA: young adult.
[i]OHCA: out-of-hospital cardiac arrest.
[j]SCA: sudden cardiac arrest.

Sensor Use and Data Analysis in Wearable Research

Studies relied heavily on IMU sensors (20/30, 67%) and smartwatches (8/30, 27%). Shoe-mounted sensors and multisensor systems incorporating electrocardiograms were less frequently used. Wrist-worn sensors (13/30, 43%) were the most common due to ease of wear and practical data collection. Smartwatches, as a subset of wrist-worn devices, were frequently used for continuous activity tracking and health monitoring. Additionally, ankle and thigh placements (7/30, 23%) were primarily used for gait analysis, whereas foot and insole sensors (2/30, 7%) were implemented for more specialized balance and gait assessments.

The predominant activity type studied was gait analysis, which appeared in 60% (18/30) of the studies, followed by activities of daily living (12/30, 40%), balance assessments (8/30, 27%), energy expenditure evaluations (6/30, 20%), and rehabilitation exercises (6/30, 20%). Gait analysis was especially relevant in research focused on neurodegenerative diseases and mobility impairments, where smartwatches were often used for free-living gait monitoring. Studies on activities of daily living leveraged smartwatches for continuous data collection in real-world settings. Balance assessments were primarily conducted for frailty and fall risk evaluations, with some smartwatch-based apps integrating accelerometry for postural control analysis.

Data processing in wearable sensor research used a range of ML techniques. Random forest was the most commonly applied method (6/30, 20%), followed by deep learning models (5/30, 17%), elastic net regression and support vector machines (SVMs; 4/30, 13%), and PCA (2/30, 7%). Random forest was frequently used in gait analysis and activity recognition, whereas deep learning models were applied for long-term movement pattern analysis, particularly in smartwatch-based apps. Elastic net and SVM were commonly used for classification tasks, and PCA was used for dimensionality reduction, optimizing the performance of wearable sensor data processing.

Clinical Applications of Wearable Sensors

Overview

The diverse clinical applications of wearable sensors, categorized into 5 main areas—healthy individuals, age-related conditions, neurological conditions, musculoskeletal disorders, and metabolic conditions—are summarized in Table 2.

XSL•FO
RenderX

Table 2. Summary of clinical results for wearable sensors across populations.

Category and condition	Key findings	Clinical significance and utility
Rehabilitation assessment and functional recovery		
Knee arthroplasty (TKA[a] and UKA[b])	Bone stimulus: +52%; impact load: +371%; OKS[c]: +52%; EQ-5D: +32%	IMU[d]-based wearables support accurate postoperative monitoring and personalized recovery for knee arthroplasty.
CR[e]	6MWD[f] prediction error: 42.8 m; R^2=0.661	CR progress can be reliably tracked remotely, improving long-term care.
Stroke—upper-limb rehabilitation	Movement quality classification accuracy: 92%; F_1-score: 0.95	IMUs enable remote monitoring of home-based rehabilitation, enhancing adherence and personalization.
Rotator cuff injury rehabilitation	Exercise classification accuracy: 99.99%	Smartwatches improve rehabilitation compliance and exercise tracking at home.
Lumbar mobility assessment	Significant ROM[g] differences between wrist and lumbar sensors (up to 8, *P*=.003)	Wrist-worn sensors allow for remote lumbar mobility assessment for rehabilitation use.
Disease state prediction and risk assessment		
PD[h]—tremor and fall risk	Tremor detection accuracy: 83%; sensitivity: 86%; specificity: 86%; fall risk prediction RMSE[i]: 0.42	Wearables objectively monitor PD symptoms and fall risk for early intervention.
HD[j]	Sensitivity: 85%; specificity: 72%; accuracy: 81%; AUROC[k]: 0.82	HD motor decline can be tracked remotely for personalized care.
Frailty assessment in older adults	QTUG[l] accuracy: 75.8%; ScanWatch-enhanced model: 79.3%	Wearables detect frailty early, enabling preventive intervention in older adults.
OHCA[m] detection	Optimized balance between sensitivity and specificity	OHCA can be detected in real time via wearables for rapid emergency response.
Activity and behavior tracking		
Arthritis—pain and function monitoring	5-STS[n] performance significantly correlated with morning pain scores (*P*<.05)	IMUs enable noninvasive, remote tracking of arthritis pain and function.
CF[o]—activity monitoring	Fitbit and iOS smartphone showed strong agreement with SWA[p]	Consumer wearables offer scalable, affordable physical activity monitoring.
Consumer vs research-grade wearables (activity monitoring)	Z-Track sedentary behavior detection: AUC[q]=0.95; MVPA[r] detection: AUC=0.93	Fitbit-level devices reliably track sedentary behavior and activity levels.
Medication adherence monitoring	Medication intake detection accuracy: 93.6%; sensitivity: 92%	Wearables automate medication tracking, improving adherence in chronic care.
Gait analysis and balance assessment		
Gait symmetry analysis (head-worn sensor)	Gait event detection accuracy: 99.35%	Head-worn sensors support gait symmetry analysis for neurological rehabilitation.
Balance assessment (smartwatch based)	Strong correlation between smartwatch and research-grade sensors (*r*=0.861-0.970)	Smartwatches enable balance monitoring at home to prevent falls.
MS[s]—balance and mobility	AUROC=0.97	Wearables track long-term mobility and balance in MS, supporting personalized rehabilitation.

[a]TKA: total knee arthroplasty.

[b]UKA: unicompartmental knee arthroplasty.

[c]OKS: Oxford knee score.

[d]IMU: inertial measurement unit.

[e]CR: cardiac rehabilitation.

[f]6MWD: 6-minute walk distance.

[g]ROM: range of motion.

[h]PD: Parkinson disease.

[i]RMSE: root mean square error.

[j]HD: Huntington disease.

[k]AUROC: area under the receiver operating characteristic curve.

[l]QTUG: quantitative timed up and go.

XSL•FO
RenderX

[m]OHCA: out-of-hospital cardiac arrest.

[n]5-STS: 5-time sit-to-stand assessment.

[o]CF: cystic fibrosis.

[p]SWA: sensewear armband.

[q]AUC: area under the curve.

[r]MVPA: moderate to vigorous physical activity.

[s]MS: multiple sclerosis.

Rehabilitation Assessment and Functional Recovery

Wearable sensors were used to analyze gait metrics in both young and older adults. Studies on young adults focused on plantar pressure distribution, step length, swing time, and ground reaction force, achieving high accuracy in real-time gait analysis, such as 95% using the FreeWalker system with a 1000-Hz sampling frequency. Advanced ML techniques further enhanced center of pressure prediction accuracy by over 30%. Among older adults, wearable sensors were effective in assessing fall risk and mobility. Improvements were observed in swing time (+6.45%) and slip and trip classification accuracy, which exceeded 98% (P<.05).

Disease State Prediction and Risk Assessment

Studies addressed frailty and fall history using wearable sensors to measure load distribution, gait phases, and stance and swing time. Load distribution assessments demonstrated high reliability, with intraclass correlation coefficient values reaching 0.91 and strong correlations for the left (r=0.7171) and right (r=0.7502) foot. Fall risk indexes provided significant predictive accuracy, with AUROC values of 0.919 (P<.05), making them comparable to traditional tools such as the POMA and TUG tests. These findings emphasize the potential of wearable sensors for early identification of frailty and fall risk in older adults.

Activity and Behavior Tracking

Wearable sensors were used to evaluate gait characteristics in individuals with stroke, MS, and PD. Among stroke survivors, significant reductions in gait speed and step length were observed compared to controls, with strong correlations between Fugl-Meyer Assessment lower-limb scores and stance time differences (R^2=0.71). In MS, high agreement was reported between the FeetMe and GAITRite systems (intraclass correlation coefficient>0.8), validating the utility of wearable sensors for mobility monitoring. For individuals with PD, significant differences were detected in gait speed, stride length, and swing and stance time compared to healthy controls (P<.05), demonstrating the role of wearable sensors in tracking disease progression.

Gait Analysis and Balance Assessment

Wearable sensors were effective in managing diabetes and other metabolic disorders. For diabetes, total contact casts reduced forefoot contact area by 5% and peak pressure by 8% (P<.05), effectively offloading pressure and reducing the risk of complications. These devices provide actionable data that support better management of metabolic health and reduce disease-related complications.

Quality Assessment Results

The quality assessment results are summarized in Table 3, with the studies rated using the JBI critical appraisal tools scoring between 5 and 8 out of 10, NOS-rated cohort studies scoring between 6 and 7 out of 9, and RoB 2–rated RCTs scoring 8 out of 10. Of the 30 included studies, 6 (20%) were rated as low risk, and 24 (80%) were rated as having a moderate risk of bias. Studies investigating wearable sensor–based mobility assessments, gait analysis, and rehabilitation applications showed high feasibility and reliability, particularly those incorporating real-time monitoring and signal processing techniques. However, several studies (26/30, 87%) exhibited limitations such as small sample sizes, lack of validation in real-world settings, and limited applicability to diverse patient populations. Additionally, some studies (12/30, 40%) faced technical challenges, including sensor displacement errors, signal-to-noise ratio issues, and data synchronization difficulties.

XSL•FO
RenderX

Table 3. Quality assessment summary of the reviewed studies.

Study	Study design	Quality score	Risk-of-bias category
Bolam et al [6], 2021	Prospective cohort study	7/9 (NOS[a])	Moderate
Angelucci and Aliverti [7], 2023	Experimental study	6/10 (JBI[b] tool)	Moderate
Greene et al [8], 2021	Experimental study	7/10 (JBI tool)	Moderate
Gordon et al [9], 2019	Observational study	6/10 (JBI tool)	Moderate
Presley et al [10], 2023	Experimental study	8/10 (JBI tool)	Low
De Cannière et al [11], 2020	Prospective cohort study	7/9 (NOS)	Moderate
Nunes et al [12], 2024	Observational study	6/10 (JBI tool)	Moderate
Mahadevan et al [13], 2020	Observational study	8/10 (JBI tool)	Low
Seo et al [14], 2024	Observational study	7/10 (JBI tool)	Moderate
Odhiambo et al [15], 2023	Experimental study	8/10 (JBI tool)	Low
Hwang and Effenberg [16], 2021	Observational study	7/10 (JBI tool)	Moderate
Wu et al [17], 2021	Observational study	7/10 (JBI tool)	Moderate
John and Soangra [18], 2022	Observational study	6/10 (JBI tool)	Moderate
Meyer et al [19], 2022	Observational study	7/10 (JBI tool)	Moderate
Toumieux et al [20], 2015	Experimental study	5/10 (JBI tool)	Moderate
Elstub et al [21], 2022	Experimental study	7/10 (JBI tool)	Moderate
Perraudin et al [22], 2018	Observational study	7/10 (JBI tool)	Low
Giggins et al [23], 2025	Cross-sectional study	6/10 (JBI tool)	Moderate
Savi et al [24], 2020	Cross-sectional study	7/10 (JBI tool)	Moderate
Haghi et al [25], 2023	Experimental study	8/10 (JBI tool)	Moderate
DasMahapatra et al [26], 2018	Observational study	6/10 (JBI tool)	Low
Sun et al [27], 2019	Experimental study	7/10 (JBI tool)	Moderate
Ramezani et al [28], 2019	Pilot study	6/10 (JBI tool)	Moderate
Liew et al [29], 2024	Cross-sectional study	6/10 (JBI tool)	Moderate
Kwon et al [30], 2019	Observational study	7/10 (JBI tool)	Moderate
Martin et al [31], 2015	RCT[c]	8/10 (RoB 2[d])	Low
Hup et al [32], 2024	Clinical study	7/10 (JBI tool)	Moderate
Browne et al [33], 2020	RCT	8/10 (RoB 2)	Moderate
Burns et al [34], 2020	Prospective cohort study	7/9 (NOS)	Moderate
Bailey et al [35], 2024	Cross-sectional study	7/10 (JBI tool)	Moderate

[a]NOS: Newcastle-Ottawa Scale.

[b]JBI: Joanna Briggs Institute.

[c]RCT: randomized controlled trial.

[d]RoB 2: version 2 of the Cochrane risk-of-bias tool for randomized trials.

Discussion

Expanding the Role of Wearable Sensors in Health Monitoring

Wearable sensor technology, including IMU-based smartwatches, smart insoles, and multisensor systems, has significantly transformed health monitoring, rehabilitation tracking, and disease risk assessment [36]. These devices enable continuous, real-world tracking of mobility and functional health, addressing key limitations of traditional clinical assessments [37]. The reviewed studies highlight these devices' diverse applications in neurological, musculoskeletal, cardiovascular, and metabolic conditions, supporting early disease detection, remote therapy adherence, and precision rehabilitation [38].

Observational studies accounted for 43% (13/30) of the reviewed studies, reflecting the feasibility of longitudinal monitoring, whereas experimental studies made up 27% (8/30), playing a crucial role in validating sensor-based assessments. However,

XSL•FO
RenderX

the limited number of RCTs, representing only 10% (3/10) of the studies, underscores the need for rigorous intervention-based research to establish causal relationships between wearable sensor use and patient outcomes. The study populations were diverse, with healthy individuals comprising 33.3% of the participants, often included for sensor validation and reliability testing. Clinical populations, including individuals with PD, stroke, frailty, and cardiovascular conditions, were the primary focus of applied research, demonstrating the potential for wearable sensors to support patient management in real-world health care settings.

Evolution of Wearable Sensor Applications

In our review, 30% (9/30) of the included studies conducted validation in healthy adults, indicating that early wearable-sensor research primarily focused on device feasibility and performance testing before expanding into clinical populations. These devices have revolutionized mobility monitoring, particularly in neurodegenerative conditions such as PD, MS, and stroke, where continuous tracking of gait parameters enables early detection of motor impairments and disease progression. They also play a significant role in frailty assessment (4/30, 13%) and fall risk prediction. Smart insoles demonstrate high predictive accuracy (AUROC=0.919; P<.05) as noninvasive, real-world mobility assessment tools.

Technological Integration and Advances in Data Processing

The studies primarily used IMU-based systems (20/30, 67%) and smartwatches (8/30, 27%). Wrist-worn sensors were the most common, representing 43% (13/30) of the devices used, as they offer practicality, ease of wear, and convenience for everyday use. Ankle- and thigh-mounted sensors accounted for 23% (7/30) of applications and were primarily used for gait and posture assessments, whereas multisensor systems integrating electrocardiograms and pressure sensors provided additional biomechanical and cardiovascular insights, although they were less frequently studied.

Advances in ML have significantly enhanced data interpretation and predictive capabilities in wearable sensor applications [2]. Random forest models, applied in 20% (6/30) of the studies, were widely used for gait classification and activity recognition, whereas deep learning techniques were applied in 17% (5/30) of the studies and demonstrated high accuracy in long-term movement analysis. Elastic net regression and SVMs were used in 13% (4/30) of cases for classification tasks, whereas PCA was used in 7% (2/30) of the studies to reduce dimensionality and optimize data processing. However, variability in feature extraction methods remains a challenge. Standardized approaches are needed to improve reproducibility and clinical translation.

Clinical Applications of Wearable Sensors

Wearable sensors demonstrated strong feasibility across multiple health care applications, including rehabilitation monitoring, disease risk assessment, activity tracking, and gait analysis. For rehabilitation assessment, wearable sensors improved postsurgical monitoring in patients who underwent knee arthroplasty, showing 52% better bone stimulus and 371% better impact load tracking. Wearable sensors for cardiac rehabilitation demonstrated reliable 6-minute walk distance prediction, with an error of 42.8 m and an R^2 value of 0.661, facilitating remote patient monitoring. In stroke rehabilitation, IMU-based movement quality assessments achieved 92% accuracy (F_1-score=0.95), supporting their use for personalized therapy and remote monitoring.

Recent studies have also extended the application of IMU-based wearable sensors to shoulder rehabilitation. Tranquilli et al [39] demonstrated that a single IMU could simultaneously capture joint mobility and muscle strength dynamics during postinjury recovery. Ajčević et al [40] applied IMU sensors to quantify shoulder kinematics and evaluate therapeutic response in adhesive capsulitis, whereas Parel et al [41] introduced a kinematic biofeedback program integrating inertial sensors for patients after rotator cuff repair. These studies highlight the versatility of IMU technology for upper-limb functional assessment and real-time feedback during rehabilitation.

Wearable sensors also played a key role in disease prediction and risk assessment. In PD monitoring, wearable technology achieved an accuracy of 83% in tremor detection and both a sensitivity and specificity of 86% in fall risk prediction, supporting the feasibility of early intervention strategies. Fall risk assessments using wearable sensors reached an AUROC value of 0.919, demonstrating their ability to provide noninvasive, real-world alternatives to clinical assessments such as the TUG and POMA tests.

Activity and behavior tracking applications showed high accuracy, particularly in arthritis-related pain and function monitoring, where significant correlations were observed between morning pain scores and 5-time sit-to-stand performance, with P values of less than .05. Consumer wearables such as Fitbit and iOS-integrated smartwatches achieved a strong agreement with research-grade sensors, with an AUROC of 0.93, demonstrating their feasibility for large-scale, real-world activity tracking. In medication adherence monitoring, smartwatch-based tracking achieved an accuracy of 93.6%, highlighting its potential for improving adherence in chronic disease management.

Gait and balance assessments using wearable sensors provided highly accurate insights into functional mobility. Head-worn IMU–based gait symmetry analysis reached an accuracy of 99.35%, indicating its effectiveness in neuromuscular rehabilitation and postural correction. Wearable sensor–based assessments of balance and mobility for patients with MS achieved an AUROC of 0.97, reinforcing their potential to support personalized rehabilitation planning and disease progression monitoring.

The included studies encompassed diverse populations, including healthy adults, neurological patients (eg, PD), individuals with musculoskeletal disorders, and pediatric or rehabilitation cohorts. This diversity introduces biomechanical and physiological variability in gait patterns, movement strategies, and sensor placement feasibility. Differences in muscle coordination, assistive device use, and experimental environments further contribute to heterogeneity. Given these

XSL•FO
RenderX

variations, direct quantitative comparisons between studies were avoided. Instead, a narrative synthesis was used to identify overarching technological and methodological trends across populations. This approach emphasizes generalizable insights—such as the importance of standardized placement, calibration, and cross-population validation—while acknowledging disease-specific distinctions in biomechanics and sensor performance.

Many of the included studies (19/30, 63%) used ML algorithms such as random forest, deep learning, elastic net regression, and PCA for signal interpretation and disease classification. However, reporting transparency and methodological rigor varied substantially. Several studies (26/30, 87%) were limited by small sample sizes and internal validation only, increasing the risk of model overfitting. In addition, few studies (4/30, 13%) provided sufficient details regarding cross-validation protocols, feature selection strategies, or hyperparameter optimization.

Adherence to standardized ML reporting frameworks—such as the Transparent Reporting of a Multivariable Model for Individual Prognosis or Diagnosis–Artificial Intelligence and Prediction of Model Risk of Bias Assessment Tool–Artificial Intelligence—was rarely observed, which may affect reproducibility and generalizability.

Future research should emphasize external validation, open-source code sharing, and adherence to established ML reporting standards to ensure reliability and transparency in sensor-based clinical modeling.

Challenges in Wearable Sensor Research

Despite the promising applications of wearable sensors, several challenges remain that must be addressed to ensure widespread clinical adoption and real-world impact.

First, small sample sizes (26/30, 87% of the studies) and limited real-world validation (12/30, 40% of the studies) reduced finding generalizability. Short study durations (8/30, 27%) also hindered long-term effectiveness assessment. Beyond these methodological limitations, the scope of this review was restricted to English-language, peer-reviewed publications, excluding gray literature such as conference abstracts and theses. This language restriction and publication bias may have favored studies reporting positive or statistically significant outcomes, potentially overestimating the clinical impact of wearable sensor technologies. Furthermore, although some studies (3/30, 10%) discussed the potential cost-effectiveness of sensor-based systems, no direct economic evaluations were identified, limiting the ability to substantiate financial feasibility claims.

Technical challenges also persist. Variability in signal-to-noise ratios, sensor displacement errors, and inconsistencies in data collection protocols underscore the need for improved hardware design and standardized preprocessing algorithms. Differences in feature extraction and model architectures limit cross-study comparisons and reproducibility.

Finally, the field urgently requires greater standardization. Variability in sensor placement, protocols, and data interpretation hinders reproducibility and large-scale comparison. Establishing consensus-driven guidelines for wearable sensor research—including standardized task protocols, data reporting frameworks, and model transparency criteria—will be essential to enable scalability, reproducibility, and eventual clinical translation.

Future Directions

To fully realize the potential of wearable sensors in health care, future research should focus on several key areas. Expanding RCTs is essential to establish causal relationships between wearable sensor use and health outcomes beyond feasibility studies. Standardized data analysis frameworks will improve comparability and reproducibility, enabling integration into multicenter trials and large-scale studies. Long-term, multicenter studies will enhance real-world validation and assess sensor accuracy, usability, and adoption across health care settings.

Integration with cloud-based platforms and telemedicine will enhance scalability and enable real-time remote monitoring across diverse populations [42]. Cost-effectiveness analyses will determine financial feasibility and accessibility, supporting broader health care adoption and effective use in resource-limited settings [43].

Conclusions

This systematic review highlights the growing clinical relevance of wearable sensors for rehabilitation monitoring, disease risk assessment, and personalized health care. IMU-based smartwatches, multisensor systems, and gait-monitoring devices demonstrate high accuracy in mobility assessment, fall risk prediction, and chronic disease management for digital health and precision medicine.

Despite their utility, the following challenges remain: small sample sizes, real-world validation gaps, and inconsistent ML methodologies. Future research should standardize protocols, expand clinical trials, and integrate sensors into telemedicine and cloud-based analytics platforms.

Overcoming these challenges will enable wearable sensors to revolutionize health care through real-time, noninvasive monitoring that bridges traditional clinical assessments and continuous real-world health tracking.

Acknowledgments

This research was supported by a grant from the Korean Advanced Research Projects Agency for Health initiative through the Korea Health Industry Development Institute funded by the Ministry of Health and Welfare, Republic of Korea (grant RS-2024-00507256). This research was also supported by a grant from the Patient-Centered Clinical Research Coordinating Center funded by the Ministry of Health and Welfare, Republic of Korea (grant RS-2025-02213601).

XSL•FO
RenderX

Data Availability

This systematic review paper does not include original data or materials as it primarily synthesizes and analyzes existing studies. Therefore, there are no specific datasets or materials available for sharing.

Authors' Contributions

BG contributed to the conception and design of the study, data analysis, interpretation of the results, and drafting of the manuscript. Yoo JI supervised the study and served as the corresponding author. All other authors critically reviewed the manuscript and approved the final version for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA 2020 checklist.
[PDF File (Adobe PDF File), 103 KB - mhealth_v14i1e76084_app1.pdf]

References

1. Wang X, Yu H, Kold S, Rahbek O, Bai S. Wearable sensors for activity monitoring and motion control: a review. Biomimetic Intell Robot 2023 Mar;3(1):100089. [doi: 10.1016/j.birob.2023.100089]
2. Wei S, Wu Z. The application of wearable sensors and machine learning algorithms in rehabilitation training: a systematic review. Sensors (Basel) 2023 Sep 05;23(18):7667 [FREE Full text] [doi: 10.3390/s23187667] [Medline: 37765724]
3. Prisco G, Pirozzi MA, Santone A, Esposito F, Cesarelli M, Amato F, et al. Validity of wearable inertial sensors for gait analysis: a systematic review. Diagnostics (Basel) 2024 Dec 27;15(1):36 [FREE Full text] [doi: 10.3390/diagnostics15010036] [Medline: 39795564]
4. Liu L, Pu Y, Fan J, Yan Y, Liu W, Luo K, et al. Wearable sensors, data processing, and artificial intelligence in pregnancy monitoring: a review. Sensors (Basel) 2024 Oct 04;24(19):6426 [FREE Full text] [doi: 10.3390/s24196426] [Medline: 39409471]
5. Ortiz BL, Gupta V, Kumar R, Jalin A, Cao X, Ziegenbein C, et al. Data preprocessing techniques for AI and machine learning readiness: scoping review of wearable sensor data in cancer care. JMIR Mhealth Uhealth 2024 Sep 27;12:e59587 [FREE Full text] [doi: 10.2196/59587] [Medline: 38626290]
6. Bolam SM, Batinica B, Yeung TC, Weaver S, Cantamessa A, Vanderboor TC, et al. Remote patient monitoring with wearable sensors following knee arthroplasty. Sensors (Basel) 2021 Jul 29;21(15):5143 [FREE Full text] [doi: 10.3390/s21155143] [Medline: 34372377]
7. Angelucci A, Aliverti A. An IMU-based wearable system for respiratory rate estimation in static and dynamic conditions. Cardiovasc Eng Technol 2023 Jun;14(3):351-363 [FREE Full text] [doi: 10.1007/s13239-023-00657-3] [Medline: 36849621]
8. Greene BR, Premoli I, McManus K, McGrath D, Caulfield B. Predicting fall counts using wearable sensors: a novel digital biomarker for Parkinson's disease. Sensors (Basel) 2021 Dec 22;22(1):54 [FREE Full text] [doi: 10.3390/s22010054] [Medline: 35009599]
9. Gordon MF, Grachev ID, Mazeh I, Dolan Y, Reilmann R, Loupe PS, et al. Quantification of motor function in huntington disease patients using wearable sensor devices. Digit Biomark 2019 Sep 06;3(3):103-115 [FREE Full text] [doi: 10.1159/000502136] [Medline: 32095771]
10. Presley BM, Sklar JC, Hazelwood SJ, Berg-Johansen B, Klisch SM. Balance assessment using a smartwatch inertial measurement unit with principal component analysis for anatomical calibration. Sensors (Basel) 2023 May 09;23(10):4585 [FREE Full text] [doi: 10.3390/s23104585] [Medline: 37430500]
11. De Cannière H, Corradi F, Smeets CJ, Schoutteten M, Varon C, Van Hoof C, et al. Wearable monitoring and interpretable machine learning can objectively track progression in patients during cardiac rehabilitation. Sensors (Basel) 2020 Jun 26;20(12):3601 [FREE Full text] [doi: 10.3390/s20123601] [Medline: 32604829]
12. Nunes AS, Potter İ, Mishra RK, Casado J, Dana N, Geronimo A, et al. Using wearable sensors and machine learning to assess upper limb function in Huntington's disease. Research Square :1-26 Preprint posted online on June 3, 2024 [FREE Full text] [doi: 10.21203/rs.3.rs-4355136/v1]
13. Mahadevan N, Demanuele C, Zhang H, Volfson D, Ho B, Erb MK, et al. Development of digital biomarkers for resting tremor and bradykinesia using a wrist-worn wearable device. NPJ Digit Med 2020 Jan 15;3:5 [FREE Full text] [doi: 10.1038/s41746-019-0217-7] [Medline: 31970290]
14. Seo NJ, Coupland K, Finetto C, Scronce G. Wearable sensor to monitor quality of upper limb task practice for stroke survivors at home. Sensors (Basel) 2024 Jan 16;24(2):554 [FREE Full text] [doi: 10.3390/s24020554] [Medline: 38257646]

XSL•FO
RenderX

15. Odhiambo CO, Ablonczy L, Wright PJ, Corbett CF, Reichardt S, Valafar H. Detecting medication-taking gestures using machine learning and accelerometer data collected via smartwatch technology: instrument validation study. JMIR Hum Factors 2023 May 04;10:e42714 [FREE Full text] [doi: 10.2196/42714] [Medline: 37140971]
16. Hwang TH, Effenberg AO. Head trajectory diagrams for gait symmetry analysis using a single head-worn IMU. Sensors (Basel) 2021 Oct 05;21(19):6621 [FREE Full text] [doi: 10.3390/s21196621] [Medline: 34640945]
17. Wu J, Kuruvithadam K, Schaer A, Stoneham R, Chatzipirpiridis G, Easthope CA, et al. An intelligent in-shoe system for gait monitoring and analysis with optimized sampling and real-time visualization capabilities. Sensors (Basel) 2021 Apr 19;21(8):2869 [FREE Full text] [doi: 10.3390/s21082869] [Medline: 33921846]
18. John J, Soangra R. Visualization-driven time-series extraction from wearable systems can facilitate differentiation of passive ADL characteristics among stroke and healthy older adults. Sensors (Basel) 2022 Jan 13;22(2):598 [FREE Full text] [doi: 10.3390/s22020598] [Medline: 35062557]
19. Meyer BM, Depetrillo P, Franco J, Donahue N, Fox SR, O'Leary A, et al. How much data is enough? A reliable methodology to examine long-term wearable data acquisition in gait and postural sway. Sensors (Basel) 2022 Sep 15;22(18):6982 [FREE Full text] [doi: 10.3390/s22186982] [Medline: 36146348]
20. Toumieux P, Chevalier L, Sahuguède S, Julien-Vergonjanne A. Optical wireless connected objects for healthcare. Healthc Technol Lett 2015 Sep 24;2(5):118-122 [FREE Full text] [doi: 10.1049/htl.2015.0028] [Medline: 26609417]
21. Elstub LJ, Nurse CA, Grohowski LM, Volgyesi P, Wolf DN, Zelik KE. Tibial bone forces can be monitored using shoe-worn wearable sensors during running. J Sports Sci 2022 Aug;40(15):1741-1749 [FREE Full text] [doi: 10.1080/02640414.2022.2107816] [Medline: 35938189]
22. Perraudin CG, Illiano VP, Calvo F, O'Hare E, Donnelly SC, Mullan RH, et al. Observational study of a wearable sensor and smartphone application supporting unsupervised exercises to assess pain and stiffness. Digit Biomark 2018 Oct 23;2(3):106-125 [FREE Full text] [doi: 10.1159/000493277] [Medline: 32095762]
23. Giggins OM, Vavasour G, Doyle J. Unsupervised assessment of frailty status using wearable sensors: a feasibility study among community-dwelling older adults. Adv Rehabil Sci Pract 2025 Feb 15;14:27536351241311845. [doi: 10.1177/27536351241311845] [Medline: 39958411]
24. Savi D, Graziano L, Giordani B, Schiavetto S, Vito CD, Migliara G, et al. New strategies of physical activity assessment in cystic fibrosis: a pilot study. BMC Pulm Med 2020 Oct 30;20(1):285 [FREE Full text] [doi: 10.1186/s12890-020-01313-5] [Medline: 33126875]
25. Haghi M, Ershadi A, Deserno TM. Recognizing human activity of daily living using a flexible wearable for 3D spine pose tracking. Sensors (Basel) 2023 Feb 12;23(4):2066 [FREE Full text] [doi: 10.3390/s23042066] [Medline: 36850664]
26. DasMahapatra P, Chiauzzi E, Bhalerao R, Rhodes J. Free-living physical activity monitoring in adult US patients with multiple sclerosis using a consumer wearable device. Digit Biomark 2018 Apr 13;2(1):47-63 [FREE Full text] [doi: 10.1159/000488040] [Medline: 32095756]
27. Sun R, Aldunate RG, Sosnoff JJ. The validity of a mixed reality-based automated functional mobility assessment. Sensors (Basel) 2019 May 11;19(9):2183 [FREE Full text] [doi: 10.3390/s19092183] [Medline: 31083514]
28. Ramezani R, Zhang W, Xie Z, Shen J, Elashoff D, Roberts P, et al. A combination of indoor localization and wearable sensor-based physical activity recognition to assess older patients undergoing subacute rehabilitation: baseline study results. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e14090 [FREE Full text] [doi: 10.2196/14090] [Medline: 31293244]
29. Liew BX, Crisafulli O, Evans DW. Quantifying lumbar sagittal plane kinematics using a wrist-worn inertial measurement unit. Front Sports Act Living 2024 May 14;6:1381020 [FREE Full text] [doi: 10.3389/fspor.2024.1381020] [Medline: 38807615]
30. Kwon S, Honegger K, Mason M. Daily physical activity among toddlers: hip and wrist accelerometer assessments. Int J Environ Res Public Health 2019 Nov 01;16(21):4244 [FREE Full text] [doi: 10.3390/ijerph16214244] [Medline: 31683776]
31. Martin SS, Feldman DI, Blumenthal RS, Jones SR, Post WS, McKibben RA, et al. mActive: a randomized clinical trial of an automated mhealth intervention for physical activity promotion. J Am Heart Assoc 2015 Nov 09;4(11):e002239 [FREE Full text] [doi: 10.1161/JAHA.115.002239] [Medline: 26553211]
32. Hup RG, Linssen EC, Eversdijk M, Verbruggen B, Bak MA, Habibovic M, et al. Rationale and design of the BECA project: smartwatch-based activation of the chain of survival for out-of-hospital cardiac arrest. Resusc Plus 2024 Feb 09;17:100576 [FREE Full text] [doi: 10.1016/j.resplu.2024.100576] [Medline: 38370313]
33. Browne S, Kechadi MT, O'Donnell S, Dow M, Tully L, Doyle G, et al. Mobile health apps in pediatric obesity treatment: process outcomes from a feasibility study of a multicomponent intervention. JMIR Mhealth Uhealth 2020 Jul 08;8(7):e16925 [FREE Full text] [doi: 10.2196/16925] [Medline: 32673267]
34. Burns D, Razmjou H, Shaw J, Richards R, McLachlin S, Hardisty M, et al. Adherence tracking with smart watches for shoulder physiotherapy in rotator cuff pathology: protocol for a longitudinal cohort study. JMIR Res Protoc 2020 Jul 05;9(7):e17841 [FREE Full text] [doi: 10.2196/17841] [Medline: 32623366]
35. Bailey DP, Ahmed I, Cooper DL, Finlay KA, Froome HM, Nightingale TE, et al. Validity of a wrist-worn consumer-grade wearable for estimating energy expenditure, sedentary behaviour, and physical activity in manual wheelchair users with spinal cord injury. Disabil Rehabil Assist Technol 2025 Apr;20(3):708-714 [FREE Full text] [doi: 10.1080/17483107.2024.2405895] [Medline: 39301994]

XSL•FO
RenderX

36. Porciuncula F, Roto AV, Kumar D, Davis I, Roy S, Walsh CJ, et al. Wearable movement sensors for rehabilitation: a focused review of technological and clinical advances. PM R 2018 Sep;10(9 Suppl 2):S220-S232 [FREE Full text] [doi: 10.1016/j.pmrj.2018.06.013] [Medline: 30269807]
37. Merritt MW. Health researchers' ancillary care obligations in low-resource settings: how can we tell what is morally required? Kennedy Inst Ethics J 2011 Dec;21(4):311-347 [FREE Full text] [doi: 10.1353/ken.2011.0019] [Medline: 22187929]
38. Tao Q, Liu S, Zhang J, Jiang J, Jin Z, Huang Y, et al. Clinical applications of smart wearable sensors. iScience 2023 Jul 27;26(9):107485 [FREE Full text] [doi: 10.1016/j.isci.2023.107485] [Medline: 37636055]
39. Tranquilli C, Bernetti A, Picerno P. Ambulatory joint mobility and muscle strength assessment during rehabilitation using a single wearable inertial sensor. Medicina dello Sport 2013;66(4):583-597 [FREE Full text]
40. Ajčević M, Deodato M, Murena L, Miladinovic A, Mezzarobba S, Accardo A. Assessment of mobility deficit and treatment efficacy in adhesive capsulitis by measurement of kinematic parameters using IMU sensors. In: Proceedings of the IEEE International Symposium on Medical Measurements and Applications. 2020 Jul 10 Presented at: MeMeA '20; June 1-July 1, 2020; Bari, Italy URL: https://ieeexplore.ieee.org/document/9137157 [doi: 10.1109/MeMeA49120.2020.9137157]
41. Parel I, Candoli V, Filippi MV, Padolino A, Merolla G, Sanniti S, et al. Shoulder rehabilitation exercises with kinematic biofeedback after arthroscopic rotator cuff repair: protocol for a new integrated rehabilitation program. JMIR Res Protoc 2023 Mar 22;12:e35757 [FREE Full text] [doi: 10.2196/35757] [Medline: 36947146]
42. Junaid SB, Imam AA, Balogun AO, De Silva LC, Surakat YA, Kumar G, et al. Recent advancements in emerging technologies for healthcare management systems: a survey. Healthcare (Basel) 2022 Oct 03;10(10):e39 [FREE Full text] [doi: 10.3390/healthcare10101940] [Medline: 36292387]
43. De Sario Velasquez GD, Borna S, Maniaci MJ, Coffey JD, Haider CR, Demaerschalk BM, et al. Economic Perspective of the Use of Wearables in Health Care: A Systematic Review. Mayo Clin Proc Digit Health 2024 Sep;2(3):299-317 [FREE Full text] [doi: 10.1016/j.mcpdig.2024.05.003] [Medline: 40206120]

Abbreviations

AUROC: area under the receiver operating characteristic curve
IMU: inertial measurement unit
JBI: Joanna Briggs Institute
ML: machine learning
MS: multiple sclerosis
NOS: Newcastle-Ottawa Scale
PCA: principal component analysis
PD: Parkinson disease
POMA: Performance-Oriented Mobility Assessment
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RoB 2: version 2 of the Cochrane risk-of-bias tool for randomized trials
SVM: support vector machine
TUG: Timed Up and Go

Edited by M Giurgiu; submitted 17.Apr.2025; peer-reviewed by A Uchenna, P Picerno; comments to author 26.Jul.2025; accepted 17.Nov.2025; published 09.Jan.2026.

Please cite as:
Gu B, Kim HS, Kim H, Yoo JI
Advancements in Wearable Sensor Technologies for Health Monitoring in Terms of Clinical Applications, Rehabilitation, and Disease Risk Assessment: Systematic Review
JMIR Mhealth Uhealth 2026;14:e76084
URL: https://mhealth.jmir.org/2026/1/e76084
doi:10.2196/76084
PMID:

©Bonsang Gu, Hyeon Su Kim, HyunBin Kim, Jun-Il Yoo. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 09.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete

XSL•FO
RenderX

bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Evaluating the Effectiveness of Digital Interventions for Stress Management in Pregnant Women: Systematic Review and Meta-Analysis

Jeung-Im Kim[1], PhD; Joo Yun Lee[2], PhD; So-Hee Park[3], MSN

[1]School of Nursing, College of Medicine, Soonchunhynag University, Asan-si, Republic of Korea

[2]College of Nursing, Gachon University, Incheon, Republic of Korea

[3]School of Nursing, College of Medicine, Soonchunhyang University, Asan-si, Republic of Korea

Corresponding Author:
Joo Yun Lee, PhD
College of Nursing, Gachon University
191, Hambangmoe-ro, Yeonsu-gu
Incheon, 21936
Republic of Korea
Phone: 82 10 9169 0130
Email: jooyun@gachon.ac.kr

Abstract

Background: Psychological stress during pregnancy is common and has been associated with adverse maternal and neonatal outcomes. Digital health interventions (DHIs) have emerged as a scalable approach to support stress management during pregnancy, yet evidence remains fragmented, and prior reviews have largely focused on broad perinatal mental health outcomes or delivery platforms rather than stress-specific effects and targeted intervention components.

Objective: This systematic review and meta-analysis aimed to evaluate the effectiveness of DHIs specifically designed to reduce stress during pregnancy and to examine how intervention strategies and delivery methods are associated with stress outcomes.

Methods: We conducted a systematic review and meta-analysis following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines. Randomized controlled trials and quasi-experimental studies involving pregnant women were eligible if they evaluated any digitally delivered intervention—such as mobile apps, web-based programs, or telemedicine—intended to reduce stress, and reported validated stress outcomes. We searched CINAHL, the Cochrane Library, Embase, and PubMed from database inception through November 2025. Risk of bias was assessed using the Cochrane risk of bias 2 tool for randomized trials and the risk of bias in nonrandomized studies of interventions tool for nonrandomized studies. Where appropriate, effect sizes were pooled using random-effects meta-analysis with the Hartung–Knapp–Sidik–Jonkman method and reported as standardized mean differences.

Results: A total of 19 studies were included. Overall, DHIs were associated with a significant reduction in stress compared with control conditions (standardized mean difference –0.45, 95% CI –0.59 to –0.32; 95% prediction interval –0.78 to –0.13), with low to moderate heterogeneity. Strategy-based subgroup analyses indicated that mindfulness- and education-focused interventions showed favorable effects, but formal tests for between-subgroup differences were not statistically significant. Evidence certainty was rated as moderate, primarily due to risk-of-bias concerns in some trials.

Conclusions: This review provides stress-focused evidence that DHIs can support stress reduction during pregnancy and extends existing literature by systematically disaggregating interventions according to delivery methods, functional features, and content strategies. This study offers a component-oriented synthesis that informs the design and selection of digital stress-management interventions for pregnant women. In real-world antenatal care, these tools may complement clinician-delivered services by expanding access to low-intensity, scalable support, particularly when interventions integrate skills-based content with supportive digital functions. Future research should directly compare single versus combined strategies and evaluate implementation across diverse populations and care settings.

(JMIR Mhealth Uhealth 2026;14:e66267) doi:10.2196/66267

XSL•FO
RenderX

KEYWORDS

pregnancy; stress management; digital health; mHealth; telemedicine; systematic review; meta-analysis

Introduction

Overview

Pregnancy represents a major life transition that is frequently accompanied by heightened psychological stress, including pregnancy-specific stress related to concerns about maternal health, fetal well-being, and childbirth. Recent evidence indicates that pregnancy-specific stress is prevalent even among low-risk populations and tends to increase with perceived pregnancy risk and advancing gestational age [1]. Elevated stress during pregnancy has been consistently associated with adverse maternal mental health outcomes as well as negative neonatal and child health outcomes, underscoring the importance of timely and effective stress management during the antenatal period [2-4].

A range of nonpharmacological interventions has been developed to address stress during pregnancy, and their effectiveness has been demonstrated across multiple studies. These include psychoeducational programs, structured psychotherapies such as cognitive behavioral therapy and interpersonal psychotherapy, and mind–body approaches including mindfulness and yoga [5-8]. Traditionally, such interventions have been delivered through face-to-face sessions. While effective, in-person delivery can be constrained by barriers related to access, cost, time, geographic distance, and the availability of trained professionals—limitations that may be particularly salient during pregnancy.

Digital health interventions (DHIs) have emerged as a scalable and low-threshold approach for delivering stress-management strategies, with growing evidence supporting their effectiveness in the general population. Meta-analytic evidence indicates that app-based stress-management interventions yield small but statistically significant improvements across self-reported, physiological, and neuroendocrine stress outcomes [9], and similar modest reductions in perceived stress have been reported in randomized trials of mental health smartphone apps [10]. These findings suggest that digitally delivered interventions can effectively support stress reduction when designed and implemented appropriately.

However, evidence specific to pregnancy remains more limited and less conclusive. Much of the existing literature on DHIs during pregnancy has focused on broader perinatal mental health outcomes—particularly depression and anxiety—rather than stress as a primary target. For example, reviews of digital mindfulness and nurse-led eHealth interventions have reported consistent benefits for depressive and anxiety symptoms but mixed or inconclusive effects on stress [11,12], while telemedicine-based psychological interventions often combine antenatal and postpartum populations and rarely prioritize stress outcomes [13]. An umbrella review of DHIs for perinatal women suggested overall benefits for psychological outcomes, including stress; however, its findings were pooled across heterogeneous populations, intervention types, outcomes (stress, depression, and anxiety), and perinatal stages, with only a small subset of reviews specifically addressing stress during pregnancy [14]. Collectively, these syntheses highlight that stress-focused evidence in pregnant populations remains fragmented and underdeveloped.

At the same time, DHIs offer distinct methodological and practical advantages that warrant more nuanced evaluation. Digital interventions are inherently multicomponent, comprising combinations of delivery methods (eg, mobile apps, web-based platforms, and telemedicine), functional features (eg, self-monitoring, automated feedback, and reminders), and intervention contents (eg, psychoeducation, mindfulness, relaxation, and cognitive behavioral skills). Behavioral science frameworks emphasize that technology functions primarily as a delivery vehicle rather than the therapeutic agent itself [15], and empirical assessments of stress-management apps similarly distinguish delivery modalities, functional features, and therapeutic content as separable but interacting components [16]. Despite this, prior reviews of digital mental health interventions during pregnancy have rarely synthesized evidence across these dimensions, limiting understanding of which combinations of delivery methods, functions, and contents are most relevant for stress reduction in pregnant women.

Rationale

Taken together, the existing literature highlights a persistent gap in evidence regarding digital interventions specifically designed to reduce stress during pregnancy. Although DHIs for perinatal mental health have received growing attention, most prior reviews have focused primarily on depression and anxiety or have pooled multiple psychological outcomes, with stress often insufficiently examined as a primary outcome. As a result, the effectiveness of digital interventions targeting stress during pregnancy remains unclear.

Moreover, previous syntheses have largely categorized digital interventions by delivery platform, providing limited insight into how specific intervention components contribute to stress reduction. DHIs are inherently multicomponent, integrating delivery methods, functional features, and therapeutic content; however, the relative importance of these components for managing pregnancy-related stress has not been systematically examined.

To address these gaps, the present systematic review and meta-analysis provide an updated and focused synthesis of digital stress-management interventions during pregnancy. By centering stress reduction as the primary outcome and examining intervention effectiveness through the lens of delivery methods and intervention content, this review aims to clarify how digital interventions can be optimally designed and implemented to support stress management in real-world antenatal care settings.

Objectives

Following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidance [17] and PRISMA-S [18], this systematic review aimed to (1) identify and describe digital stress-management interventions for

XSL•FO
RenderX

pregnant women (Population), including delivery platforms and key components (Intervention); (2) evaluate the effectiveness of these interventions compared with usual care, wait-list, or active comparators (Comparator) on validated stress outcomes (Outcome); and (3) summarize intervention strategies and delivery methods associated with engagement and stress reduction.

Methods

Overview

We conducted a systematic review and meta-analysis to evaluate digitally delivered interventions for reducing stress in pregnant women. The review adhered to the PRISMA guidelines [17] and PRISMA-S (PRISMA literature search extension) [18] (Multimedia Appendix 1).

Eligibility Criteria

Eligibility criteria were predefined using the PICOS (population, intervention, comparison, outcome, and study design) framework. Participants were pregnant women. Interventions included any digitally delivered program primarily aimed at reducing or managing stress during pregnancy (antenatal period), delivered via mobile apps, web-based platforms, telemedicine, or other technology-enabled modalities. Studies in which the intervention was initiated postpartum were excluded. Comparators included usual care, waitlist/no intervention, or nondigital interventions. Outcomes included stress assessed using validated instruments; when the Depression Anxiety Stress Scales (DASS) was used, only the stress subscale was extracted. Study designs included RCTs and quasi-experimental/nonrandomized intervention studies with extractable pre-post outcome data.

Exclusion criteria included (1) postpartum-only interventions (interventions initiated after delivery), (2) interventions not primarily targeting stress management (eg, programs targeting depression, weight control, smoking cessation, or other conditions in which stress was only a secondary outcome without a stress-focused rationale), (3) insufficient outcome data to estimate effects (no extractable pre-post data or between-group comparisons), and (4) conference abstracts or trial registry entries without sufficient methodological detail or outcome data for extraction and risk-of-bias assessment.

Information Sources

The initial systematic search was conducted across 4 databases (CINAHL, Cochrane Library, Embase, and PubMed) from database inception to September 2023. In response to reviewer comments and to ensure the currency of the evidence base, the search was updated by re-running the same search strategy in all databases for records published from September 2023 through November 30, 2025. The updated search was executed on December 10, 2025, by JYL and SHP, and all newly retrieved records were screened and incorporated into the final synthesis as appropriate. Searches were conducted with the assistance of 2 medical librarians, and the final search strategies for each database are reported in Multimedia Appendix 2. Each database was searched independently using its native platform (PubMed, Embase.com, CINAHL via EBSCOhost, and the Cochrane Library). Databases were not searched simultaneously on a single platform. In addition to database searching, reference lists of included studies were manually screened to identify potentially relevant articles using citation searching methods. We did not contact study authors, experts, or intervention developers to obtain additional data, nor were other information sources such as trial registries or gray literature repositories searched.

Search Strategy

Electronic search strategies were developed in consultation with 2 medical librarians, who proposed candidate keywords and controlled vocabulary; the research team reviewed and refined the strategies, and final searches were executed by the study team with librarian support. All search strategies were newly developed for this review and were not adapted from or reused based on search strategies reported in previous literature reviews. The strategy combined controlled vocabulary (MeSH [Medical Subject Headings], Emtree, and CINAHL Subject Headings) and keywords related to pregnancy, digital health, and stress. Pregnancy-related terms included “pregnan*,” “pregnancy,” “antenatal,” and “prenatal” (and related indexing terms). Digital health terms included “digital health,” “mHealth,” “eHealth,” “telemedicine,” “mobile app,” and “smartphone.” Stress-related terms included “stress” and relevant intervention terms (eg, relaxation, mindfulness, cognitive behavioral). Animal studies were excluded, and results were limited to English-language publications. Full search strategies for each database are provided in Multimedia Appendix 2.

Selection Process

All records were imported into EndNote and deduplicated. Three reviewers (JIK, JYL, and SHP) independently screened titles and abstracts to identify potentially eligible studies. Full texts were retrieved for records deemed relevant or uncertain; when full-text articles were difficult to obtain, a medical librarian assisted with document retrieval. Two reviewers (JIK and SHP) independently assessed full-text eligibility and cross-checked decisions, with disagreements resolved through discussion (and adjudication by a third reviewer when necessary). For the updated search, a second screening pass was completed on December 12, 2025, and all newly retrieved articles were screened and assessed for eligibility using the same procedures as the initial review, with eligible studies incorporated into the final synthesis.

Data Collection Process and Data Items

Data was extracted using a piloted structured form capturing intervention purpose, participant characteristics, timing and duration, intervention strategies/components and digital functions, delivery mode, comparator, sample size, stress measures, and outcome values. When outcomes were measured repeatedly, the assessment closest to intervention completion was extracted; follow-up outcomes were extracted separately where available. Two reviewers independently entered data into an Excel file and verified consistency. Discrepancies were resolved by consensus among 3 reviewers to ensure accuracy and consistent terminology.

XSL•FO
RenderX

Study Risk of Bias Assessment

Risk of bias was assessed using RoB 2 (a revised Cochrane risk-of-bias tool for randomized trials) for randomized trials [19] and ROBINS-I (risk of bias in nonrandomized studies of interventions) for quasi-experimental/nonrandomized studies [20]. Two reviewers (JIK and SHP) independently rated each study, and disagreements were resolved through discussion with adjudication by a third reviewer [JYL] when needed. RoB 2 judgments were made across 5 domains and summarized as low risk, some concerns, or high risk. ROBINS-I judgments were made across 7 domains and summarized as low, moderate, serious, or critical risk of bias. Risk-of-bias assessments informed sensitivity analyses and the certainty of evidence (GRADE [Grading of Recommendations, Assessment, Development and Evaluation]).

Effect Measures

The primary outcome was the change in stress from baseline to postintervention (closest to intervention completion). When DASS was used, only the stress subscale was extracted. Continuous outcomes were synthesized as standardized mean differences (SMDs; Hedges *g*) and reported with 95% CIs [21], with negative values indicating lower stress in the intervention group. When studies did not report means and SDs directly, effect estimates were derived from available statistics (eg, *F* values, odds ratios, chi-square values) using standard conversion methods.

Synthesis Methods

Study characteristics (population, intervention components and digital functions, delivery mode, stress measures, and effects) were summarized narratively. For quantitative synthesis, meta-analyses were conducted using random-effects models to reflect expected between-study variability in true effects. Analyses were performed in in R (version 4.5.2; R Foundation for Statistical Computing, Vienna, Austria) using the Hartung–Knapp–Sidik–Jonkman method [22]. Heterogeneity was quantified using τ^2 and I^2, with the Cochran Q test reported where appropriate. To distinguish the average pooled effect from the expected distribution of effects across settings, we reported 95% prediction intervals for the main meta-analyses (but not for subgroup analyses). CIs describe uncertainty around the pooled mean effect, whereas prediction intervals indicate the plausible range of effects in a new setting. Prespecified subgroup analyses (eg, intervention strategies/components and delivery modes) and, when sufficient studies were available, meta-regression were used to explore potential sources of heterogeneity. Sensitivity analyses excluded studies at high or serious risk of bias and tested alternative assumptions for imputed parameters where applicable. Small-study effects were assessed using funnel plots and Egger's regression test when ≥10 studies were available for a given main analysis.

Certainty Assessment

Two reviewers (JIK and SHP) independently assessed certainty of evidence using GRADE, with disagreements resolved through consensus and third-reviewer adjudication. Certainty was evaluated across risk of bias, inconsistency, indirectness, imprecision, and publication bias. Downgrading decisions were documented in the Summary of Findings table and were informed by RoB 2/ROBINS-I judgments, heterogeneity (including the extent to which prediction intervals indicated variability across settings), imprecision (width of CIs and information size), and evidence of small-study effects (funnel/Egger).

Results

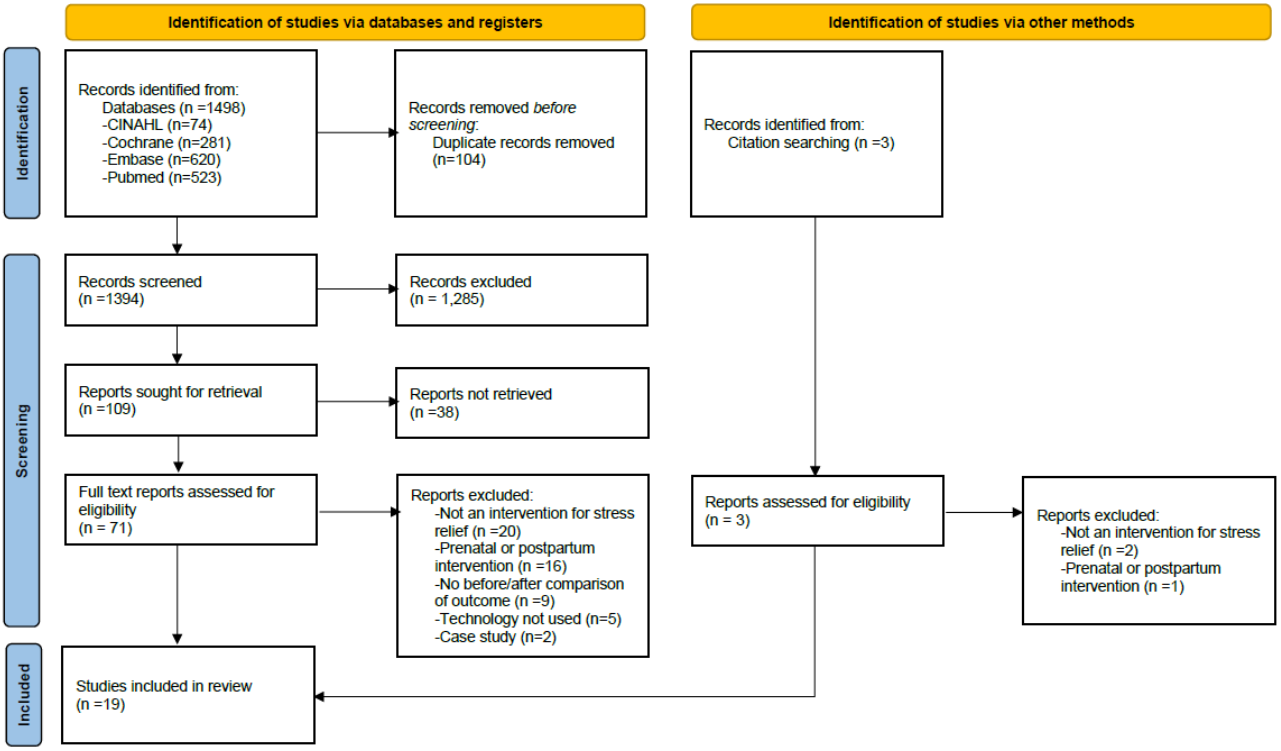
Study Selection

The study selection process is summarized in Figure 1. Studies were excluded if they targeted health conditions other than stress (eg, depression or anxiety, weight control, or smoking cessation) without a stress-focused intervention rationale, or if stress was measured only as a secondary outcome. Studies describing intervention development without reporting evaluable effectiveness outcomes were also excluded. Citation searching identified additional records, which were assessed at the full-text level but ultimately did not meet the inclusion criteria. Reasons for exclusion at each stage were documented and are presented in the PRISMA flow diagram.

Following the updated search and eligibility assessment of newly retrieved records, 4 additional studies were included. In total, 19 studies (15 from the initial search and 4 from the updated search) were included in the final review.

XSL•FO
RenderX

Figure 1. Study selection process based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram.



Study Characteristics

Included studies were published between 2016 and 2025 [23-41]. Among them, 7 studies [24,26,28,31,33,34,38] were conducted in the United States (37%), followed by studies from Iran (n=2) [29,37], China (n=2) [25,34], Korea (n=2) [24,32], and other countries. Overall, there were 12 RCTs (63%) and 7 quasi-experimental studies (37%; Table 1).

XSL•FO
RenderX

Table 1. Characteristics of the included studies.

Author	Country (year)	Target	Gestational age at intervention	Intervention period	Delivery mode	Intervention contents	Intervention details
Randomized controlled study							
Balderas-Díaz et al [23]	Spain (2022)	At risk of having SGA[a] fetuses	Second trimester and part of the third trimester	16 weeks	Mobile	Education and information	• Learn about specific topics, carry out the proposed activities to strengthen training • Four components: Medical advice, health care, communication with the fetus for stimulation purposes, and emotional management
Lee et al [24]	Korea (2023)	Working pregnant women	From ≤34 weeks	4 weeks	Mobile	Education and information with peer support	• Four components: Eight education sessions, a health log, a diary, and an anonymous discussion board • Each session lasts approximately 15 minutes • Four targeted practices: Sleep and rest, eating, physical activity, and stress management
Sun et al [25]	China (2021)	At risk of perinatal depression (with an EPDS[b] score >9 or a PHQ-9[c] score >4)	12-20 weeks	8 weeks	Mobile	Mindfulness	• Formal mindfulness training: Body scan, mindful breathing, mindful stretching, and mindful meditation lasting 15-25 minutes per day • Informal training: Pausing in the midst of daily life, mindful eating, mindful walking, and 3-minute breathing practices
Smith et al [26]	United States (2021)	Obstetric patients of outpatient clinic during the COVID-19 pandemic	14-34 weeks	30 days	Mobile	Mindfulness	• Mindfulness meditation, sleep stories, and nature sounds • Encourage use of 70 minutes per week, preferably 10 minutes per day
Krusche et al [27]	United Kingdom (2018)	Pregnant women recruited online	12-34 weeks	4 weeks	Website	Mindfulness	• Formal and informal meditation practices such as body scan, mindful movement, breathing space, and mindful eating • Online sessions, assignments, and emails
Dennis-Tiwary et al [28]	United States (2017)	Pregnant women recruited from a large urban hospital	19-29 weeks	4 weeks	Mobile	Game	• Gamified attention bias modification training (ABMT), incorporating video game-like features such as animated characters and sound effects • 10 rounds each day (≤10 minutes of play) for 4 days/week: 160 rounds total for the duration of the study
Kia [29]	Iran (2023)	No stressful event other than COVID-19 disease in the preceding 6 months	12 or more weeks	3 weeks	Mobile	Education and information	• Educational theme: (1) COVID-19 disease explanation, personal hygiene, and transmission modes, (2) COVID-19 in pregnancy and childbirth, (3) COVID-19 and breastfeeding, (4) COVID-19 and infants • Five sessions of 30 minutes, 2 sessions per week

XSL•FO
RenderX

Author	Country (year)	Target	Gestational age at intervention	Intervention period	Delivery mode	Intervention contents	Intervention details
Mauriello et al [30]	United States (2016)	Pregnant women recruited from 6 locations of 3 federally funded health centers	From <19 weeks	24 weeks	Mobile	Education and information	• Stage-matched and tailored guidance, 3 interactive sessions focused on 2 priority health behavior risks (smoking, stress management, fruits and vegetables), individually tailored and stage-matched change strategies • Feedback messages within stage-matched activities, including tools such as calculators, quizzes, action plans, support messages, and recipe ideas • Three antenatal sessions are spaced approximately 12 weeks apart, with 2 postpartum assessment-only sessions at 1 month and 4 months
Chua et al [31]	Singapore (2024)	Heterosexual couples recruited from a public tertiary hospital	From >24 weeks	From 24 weeks to 1 month postpartum	Mobile	Education and information and support	• Six components: (1) Education center (multimedia resources, chatbot), (2) Ask an expert (Experienced nurses/midwives will respond to queries within 24 h), (3) Smile center (mood rating, guided mindfulness), (4) Positivity space (support forum), (5) Gamification features (virtual badges, rewards), (6) Helpline
Park et al [32]	Korea (2025)	Pregnant women recruited from an obstetrics and gynecology center	1-32 weeks	8 weeks	Mobile	Mindfulness	• Four sections: (1) breathing mindfulness meditation, (2) body scanning, (3) emotional awareness, (4) self-kindness mindfulness • Each session was divided into 2 subsessions—instruction and practice sessions. • Instructed to practice each session at least twice
Tandon et al [33]	United States (2025)	Pregnant women recruited from 6 university affiliated prenatal care clinics	From <22 weeks	maximum of 14 weeks	Mobile, in-person	Education and Mindfulness	• Twelve-session manualized intervention: CBT[d] content related to behavioral activation, identification, and reframing unhelpful thought patterns, and promotion of positive interactions • Includes various mindfulness practices • Received just-in-time (JIT) messages were sent within 24 h of an elevated stress reading
Tian et al [34]	China (2025)	Couples expecting their first child recruited from outpatient centers	12-20 weeks	6 weeks	Mobile	Mindfulness	• Six one-week modules, each with a thematic session and 6 audio-guided home practice sessions • Mindfulness exercises that involved couple collaboration and interaction
Quasi-experimental study							

XSL•FO
RenderX

Author	Country (year)	Target	Gestational age at intervention	Intervention period	Delivery mode	Intervention contents	Intervention details
Porter et al [35]	United States (2022)	Pregnant women recruited from a university hospital	<15 weeks to 28 weeks	13 weeks	Mobile	Mindfulness	• Meditations tailored to the trimester and specific physical and emotional states of participants • Recommend daily usage of 10-20 minutes
Jallo et al [36]	United States (2017)	Hospitalized with preterm labor	22+0-33+6 weeks	8 days	Mobile	Education and relaxation	• Educational overview (stress, stress response, and impact on health) • Four guided imagery audio files, including relaxation, focused breathing, positive affirmations, and multiple multisensory images, each lasting 15-20 minutes long • A stress self-assessment scale
Hashemzahi et al [37]	Iran (2022)	With perceived stress (PSS[e] score ≥21.8) and moderate to severe anxiety (CDAS[f] score ≥17)	20-28 weeks	2 weeks	Telemedicine (WhatsApp)	Education and support	• Educational content covered topics such as familiarity with COVID-19, its effect on pregnancy, prevention and self-care during pregnancy, and guidelines for care and prevention during childbirth, postpartum, and breastfeeding • The content was shared with the group every other day over 2 weeks, comprising 6 sessions via WhatsApp. • The researcher followed up to ensure that participants watched the educational videos and addressed any questions or concerns raised by participants
Tsai et al [38]	Taiwan (2018)	Low-risk pregnant women in the outpatient department of a medical center	From 16-24 to 36-38 weeks	12-22 weeks	Website (accessible via smartphone and computer)	Education and information	• Four modules: Maternity health records, antenatal health education, self-management journals, newborn birth records • Participants could store their own information and records • Automatic pop-up windows provide antenatal health information according to a woman's gestational age
Buultjens et al [39]	Australia (2023)	Low-risk pregnant women at one community health site	From 28-30 weeks to 36-38 weeks	6-10 weeks (Antenatal)	Telemedicine (video conference)	Education and support	• One-to-one pregnancy clinical care, integrating standard pregnancy health assessments with structured online small-group interdisciplinary education and peer support, thus incorporating broader psychosocial aspects • Four group antenatal and 4 group postnatal education sessions
Kubo et al [40]	United States (2021)	With moderate-to-moderately-severe depressive symptoms (PHQ-8[g] score 10-19)	From <28 weeks	6 weeks	Mobile	Mindfulness	• Basic mindfulness condition- or situation-specific courses • Headspace courses including breathing exercises, body scan, noting, and visualization • Additional short (1-2 min) lecture videos designed to increase the understanding of mindfulness and encourage its integration into daily life • Participants asked to use the app for 10-20 min a day over 6 weeks

XSL•FO
RenderX

Author	Country (year)	Target	Gestational age at intervention	Intervention period	Delivery mode	Intervention contents	Intervention details
Barber and Masters-Awatere [41]	New Zealand (2022)	Pregnant women recruited from midwifery clinics, through antenatal educators, and via social media	<24 weeks	12 weeks	Mobile	Education and relaxation	• Four types of modules: (1) interactive self-assessments with associated feedback, (2) activities for relaxation, stress management and planning for parenthood, (3) information about psychological and social changes in pregnancy, and (4) discussions with a partner or support person to build social support and interactive reflection and problem-solving

[a]SGA: small for gestational age.

[b]EPDS: Edinburgh Postnatal Depression Scale.

[c]PHQ: patient health questionnaire.

[d]CBT: cognitive behavioral therapy.

[e]PSS: Perceived Stress Scale.

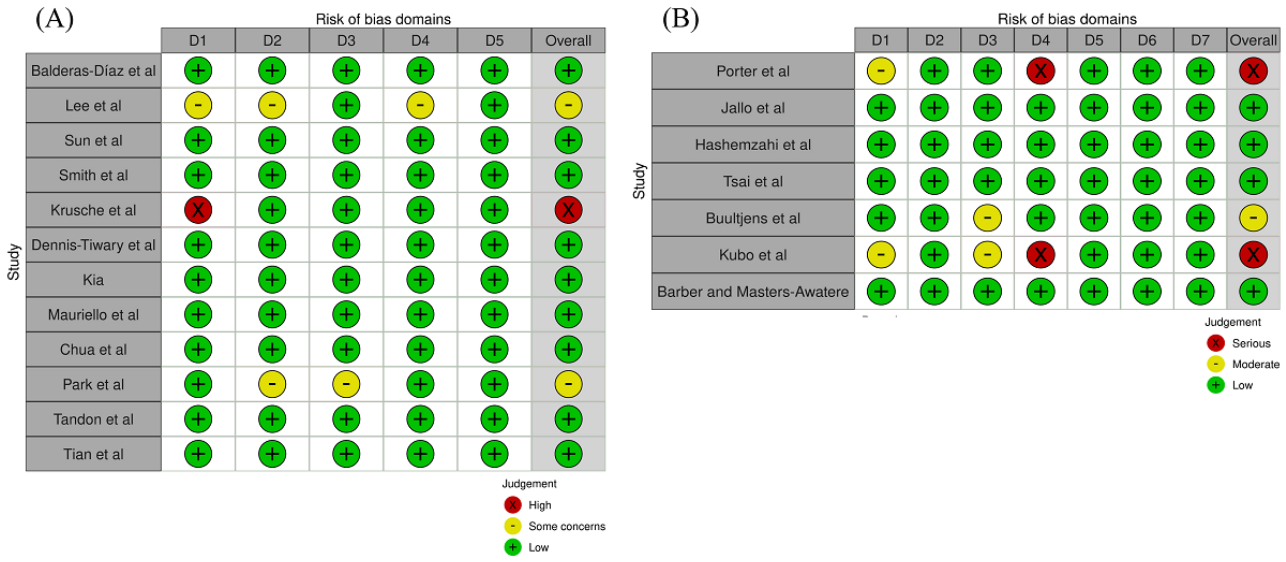
[f]CDAS: Corona Disease Anxiety Scale.

[g]PHQ-8: Patient Health Questionnaire-8.

Risk of Bias in Studies

The results of the risk of bias assessment are presented in Figure 2 [23-41]. Among the 12 RCTs assessed with RoB 2, one study was judged at high risk of bias overall, driven by concerns in the randomization process (Domain 1). Two additional RCTs were rated as having some concerns overall, reflecting incomplete reporting or concerns in specific domains, while the remaining RCTs were judged at low risk of bias across domains.

Figure 2. Risk of bias assessments. (A) Risk of bias for randomized controlled trials assessed using the revised Cochrane risk-of-bias tool. Domains: D1 (randomization process), D2 (deviations from intended interventions), D3 (missing outcome data), D4 (measurement of the outcome), and D5 (selection of the reported result). (B) Risk of bias for nonrandomized studies assessed using the risk of bias in nonrandomized studies of interventions. Domains: D1 (confounding), D2 (selection of participants), D3 (classification of interventions), D4 (deviations from intended interventions), D5 (missing data), D6 (measurement of outcomes), and D7 (selection of the reported result) [23-41].



Among the 7 nonrandomized studies assessed with ROBINS-I, 2 studies were judged at serious risk of bias overall, primarily due to deviations from intended interventions (Domain 4). One study was judged at moderate risk of bias, and the remaining studies were judged at low risk across ROBINS-I domains. No study was judged at critical risk of bias.

Results of Individual Studies

Study Population

All studies targeted pregnant women, with 12 studies focusing on low-risk pregnancies (63%), while the remaining studies targeted pregnant women with specific issues or risk factors. Two studies focused on women at risk of perinatal depression, while one study each targeted women with perceived stress,

XSL•FO
RenderX

stress induced by the COVID-19 pandemic, small for gestational age fetuses, and those hospitalized due to preterm labor. Additionally, one study focused on stress management in working pregnant women. Among the 12 studies, 9 studies in the low-risk group primarily started in the second trimester and often continued into the third trimester. Interventions targeting pregnant women with depression or stress and those carrying small for gestational age fetuses started in the second trimester. Interventions for preterm labor were initiated between 22 and 33 weeks of gestation. Meanwhile, interventions for working pregnant women could be initiated at any time before 34 weeks of gestation.

Intervention Strategies

When categorizing the intervention strategies, 4 studies provided only education and information (21%), while 7 studies incorporated additional strategies (37%). Specifically, 2 studies included professional support, and 2 studies incorporated peer support, which we classified as the strategy of "seeking support." Three studies included relaxation content, such as guided imagery audio files or activities for relaxation, such as watching humorous YouTube videos. One of these studies also included peer support. Furthermore, 7 studies primarily focused on mindfulness, and one study offered an intervention in the form of a game (Table 2).

Table 2. Intervention strategies and functions included in the stress management intervention.

Category and details	Study
Intervention strategies	
Education and information	
Education, medical advice: text, video files, automatic pop-up, or videoconference	Barber and Masters-Awatere [41], Balderas-Díaz et al [23], Lee et al [24], Kia [29], Jallo et al [36], Hashemzahi et al [37], Tsai et al [38], Buultjens et al [39], Chua et al [31], Tandon et al [33]
Seeking support	
Anonymous discussion board	Lee et al [24], Chua et al [31]
Online group peer support	Buultjens et al [39]
Conversation with the partner, other supporters	Barber and Masters-Awatere [41]
Question and answer from the researcher	Hashemzahi et al [37]
Relaxation	
Guided imagery audio files	Barber and Masters-Awatere [41], Jallo et al [36]
Activities for relaxation	Barber and Masters-Awatere [41]
Mindfulness	
Formal training: body scan, mindful breathing, mindful stretching, mindful meditation, and visualization: text, video files, and audio files	Sun et al [25], Smith et al [26], Krusche et al [27], Porter et al [35], Kubo et al [40], Park et al [32], Tandon et al [33], Tian et al [34]
Informal training: mindful eating, mindful walking	Sun et al [25], Krusche et al [27]
Game	
Attention bias modification training	Dennis-Tiwary et al [28]
Functions	
Monitoring (recording)	
Health log, diary	Lee et al [24], Tsai et al [38], Tandon et al [33]
Mindfulness journal	Sun et al [25]
Self-assessment scale	Barber and Masters-Awatere [41], Jallo et al [36]
Automated feedback	
Tailored guidance, feedback messages: programmed messages and chatbot	Barber and Masters-Awatere [41], Mauriello et al [30], Chua et al [31], Tandon et al [33]
Reminder	
Notifications and reminders to prompt use	Barber and Masters-Awatere [41], Sun (WeChat) et al [25], Krusche (e-mail) et al [27], Kubo et al [40], Tian et al [34]

In the process of content extraction, some elements were classified as functions according to the categorization by Paganini et al [16]. For instance, monitoring (recording) included studies where participants regularly assessed stress or fetal development through health logs, diaries, or self-assessment tools. Automated feedback involved the delivery of tailored guidance and feedback messages through programmed communication. There was an intervention incorporated with

XSL•FO
RenderX

interactive chatbot functions that responded to users' questions in real time, providing guidance related to stress management and pregnancy. Additionally, 4 studies included a reminder function, which prompted participants to re-engage with the digital intervention if they had not participated for a certain period.

Intervention Delivery Mode

The digital interventions, which were the focus of the studies, were delivered via mobile app in 15 (79%) studies, via the web in 2 (10.5%) studies, and via telemedicine in 2 studies, using video conferencing and WhatsApp Messenger.

Measure of Stress

Stress was measured using various tools. Ten (53%) studies used the Perceived Stress Scale, making it the most frequently used measure. Four (21%) studies used DASS, with one of these also measuring salivary cortisol. Other measures used in the studies included the COVID-19 stress score, prenatal distress questionnaire, Pregnancy Stress Rating Scale-36, Visual Analog Stress Scale, and the Stage of Change in Stress Management, each being used in one study.

Intervention Outcomes

Among the studies, significant changes in intervention outcomes were observed in 10 cases, while in 8 studies, stress levels decreased but were not statistically significant (Table 3). In the study that used a game for stress reduction [28], the DASS results did not exhibit a significant difference, whereas the salivary cortisol results did.

XSL•FO
RenderX

Table 3. Comparisons and outcomes in the included studies.

Author	Intervention group		Control group		Outcomes
	Provided content	Number of participants	Provided content	Number of participants	
Randomized controlled trials					

Balderas-Díaz et al [23]	VivEmbarazo app	15 couples	Routine perinatal care (not mHealth[a] system)	24 couples	• Significant difference in:
					  • Maternal stress (PSS[b]) ↓
					  • Baby weight at birth ↑
					  • Gestational age at birth ↑
					  • Preterm ↓
					• Nonsignificant difference
					  • Maternal depression (EPDS[c])

Lee et al [24]	Self-care for Pregnant Women at Work (SPWW) app	60	Application that only had surveys	56	• At week 4[d], significant difference in:
					  • Pregnancy stress (PDQ[e]) ↓
					  • Pregnancy hassles ↓
					  • Health practice in pregnancy ↑
					• Nonsignificant difference in:
					  • Work stress
					  • Fear of childbirth

Sun et al [25]	Spirits Healing app	74	WeChat health consultations to control attention	84	• At week 8[d], significant difference in:
					  • Depression (EPDS) ↓
					  • Anxiety symptoms (GAD-7[f]) ↓
					  • Position effect ↑
					• Nonsignificant difference in:
					  • Perceived stress (PSS)
					  • Negation effect
					  • Sleep-related problems
					  • Fatigue
					  • Prospective memory
					  • Fear of childbirth

Smith et al [26]	Calm app	33	Routine perinatal care	27	• At day 30[d], significant difference in:
					  • Stress (PSS) ↓
					• Nonsignificant difference in:
					  • Depression (HADS[g])
					  • Anxiety (HADS)
					  • Sleep disturbance

Krusche et al [27]	Website: Be Mindful Online	22 out of 107 respondents who completed the initial survey	Routine perinatal care	50 out of 78 respondents who completed the initial survey	• At day 45[d], significant difference in:
					  • Stress (PSS) ↓
					  • Depression (EPDS) ↓
					  • Pregnancy distress ↓
					• Nonsignificant difference in:
					  • Anxiety (GAD-7)

Dennis-Tiwary et al [28]	ABMT app	15	App with placebo mode	14	• Significant difference in:
					  • Stress (salivary cortisol) ↓
					• Nonsignificant difference in:
					  • Stress (DASS[h])
					  • Anxiety (DASS and HAM-A[i])
					  • Depression (DASS)

XSL•FO
RenderX

Author	Intervention group		Control group		Outcomes
	Provided content	Number of participants	Provided content	Number of participants	
Kia [29]	Mobile-based health educational intervention	40	A PDF file of the educational content	40	• Significant difference in: COVID-19 stress score (CSS-18[j]) ↓

Mauriello et al [30]	iPad- Healthy Pregnancy: Step by Step	169	Brochures named March of Dimes on the target behaviors	166	• At the third trimesterd, nonsignificant difference in:
					  • Stress management
					  • Fruit and vegetable consumption

Chua et al [31]	Parentbot- a digital health care assistant	59 couples	Routine perinatal care	59 couples	• At postpartum months 1 and 3[d], significant difference in
					  • Anxiety (State–Trait Anxiety) ↓
					• Nonsignificant difference in:
					• Stress (PSS)
					  • Depression (EPDS)
					  • Support
					  • Parent-child bonding

Park et al [32]	Mindfulness-based mobile intervention	66	Routine perinatal care, wait-list	67	• At 4 weeks[d], significant difference in:
					  • Anxiety (DASS-21) ↓
					  • Emotional well-being ↑
					  • Maternal-fetal attachment ↑
					• Nonsignificant difference in:
					  • Stress (DASS-21)
					  • Depression (DASS-21)
					  • Postnatal depression (EPDS)

Tandon et al [33]	Personalized stress management	49	Routine perinatal care	51	• At week 1 and postpartum months 1 and 3[d], significant difference in:
					  • Depression (PROMIS[k]) ↓
					  • Perceived stress (PSS-10) ↓
					  • Behavioral activation ↑
					  • Decentering ↑
					  • Mood regulations
					• Nonsignificant difference in:
					  • Anxiety (STAI[l])
					  • Social support

Tian et al [34]	WeChat mini-program (mobile digital platform)	80 couples	Routine perinatal care	80 couples	• At 2 weeks and postpartum week 6[d], significant difference in:
					  • Maternal perceived stress (PSS-10) ↓
					  • Maternal depression (EPDS) ↓
					  • Paternal depression (EPDS) ↓
					  • Mindfulness (FFMQ[m]) ↑
					  • Infant Neuropsychological Development ↑
					• Nonsignificant difference in:
					  • Anxiety (GAD-7)
					  • Symptoms of sleep problems
					  • Fatigue

XSL•FO
RenderX

Author	Intervention group		Control group		Outcomes
	Provided content	Number of participants	Provided content	Number of participants	
Porter et al [35]	Expectful mindfulness app	12 out of 21 participants who completed at least one meditation	Routine perinatal care	247	• At 28 weeks[d], significant difference in stress (PSS) ↓

Jallo et al [36]	Stress coping intervention app	5 out of 15 participants who completed baseline measure	—[n]	—	• Significant change in:
					  • Maternal stress (VASS[o]) ↓
					• Nonsignificant change in:
					  • Perceived stress (PSS)
					  • Stress coping (CSES[p])

Hashemzahi et al [37]	Telemedicine (WhatsApp Messenger)	50	Routine perinatal care	50	• Significant difference in:
					  • Perceived stress (PSS) ↓
					  • Corona disease anxiety (CDAS[q]) ↓

Tsai et al [38]	Web-based antenatal care system	68	Routine perinatal care: face-to-face individual consulting	67	• Significant difference in:
					  • Pregnancy-related stress (PSRS-36[r]) ↓
					  • Self-efficacy (GSE[s]) ↑

Buultjens et al [39]	Perinatal care, education and support (PECS) intervention	40	Routine perinatal care	21	• At 36-38 weeks[d], significant difference in:
					  • Depression (DASS and EPDS)
					• Nonsignificant difference in
					  • Stress and anxiety (DASS)

Kubo et al [40]	Headspace mindfulness app	20	—	—	• Significant change in:
					  • Perceived stress (PSS-10) ↓
					  • Depression (PHQ-8[t]) ↓
					  • Sleep disturbance (PSQI[u]) ↓
					  • Mindfulness (FFMQ) ↑

Barber and Masters-Awatere [41]	Positively Pregnant (PP) mobile app	42	—	—	• Significant change in:
					  • Stress (DASS) ↓
					• Nonsignificant change in:
					  • Depression and anxiety (DASS)
					  • Depression (APDS)

[a]mHealth: mobile health

[b]PSS: Perceived Stress Scale.

[c]EPDS: Edinburgh Postnatal Depression Scale.

[d]In the case of repeated measures, the closest measured values at the end of the intervention were compared.

[e]PDQ: prenatal distress questionnaire.

[f]GAD-7: Generalized Anxiety Disorder.

[g]HADS: Hospital Anxiety and Depression Scale.

[h]DASS: Depression Anxiety Stress Scale.

[i]HAM-A: Hamilton Anxiety Scale.

[j]CSS-18: COVID-19 stress score.

[k]PROMIS: Patient-Reported Outcomes Measurement Information System Depression Scale.

[l]STAI: State-Trait Anxiety Inventory Scale.

[m]FFMQ: Five Facet Mindfulness Questionnaire.

XSL•FO
RenderX

[n]Not applicable.
[o]VASS: Visual Analog Stress Scale.
[p]CSES: Coping Self-Efficacy Scale.
[q]CDAS: Corona Disease Anxiety Scale.
[r]PSRS-36: Pregnancy Stress Rating Scale-36.
[s]GSE: General Self-Efficacy Scale.
[t]PHQ: patient health questionnaire.
[u]PSQI: Pittsburgh Sleep Quality Index.

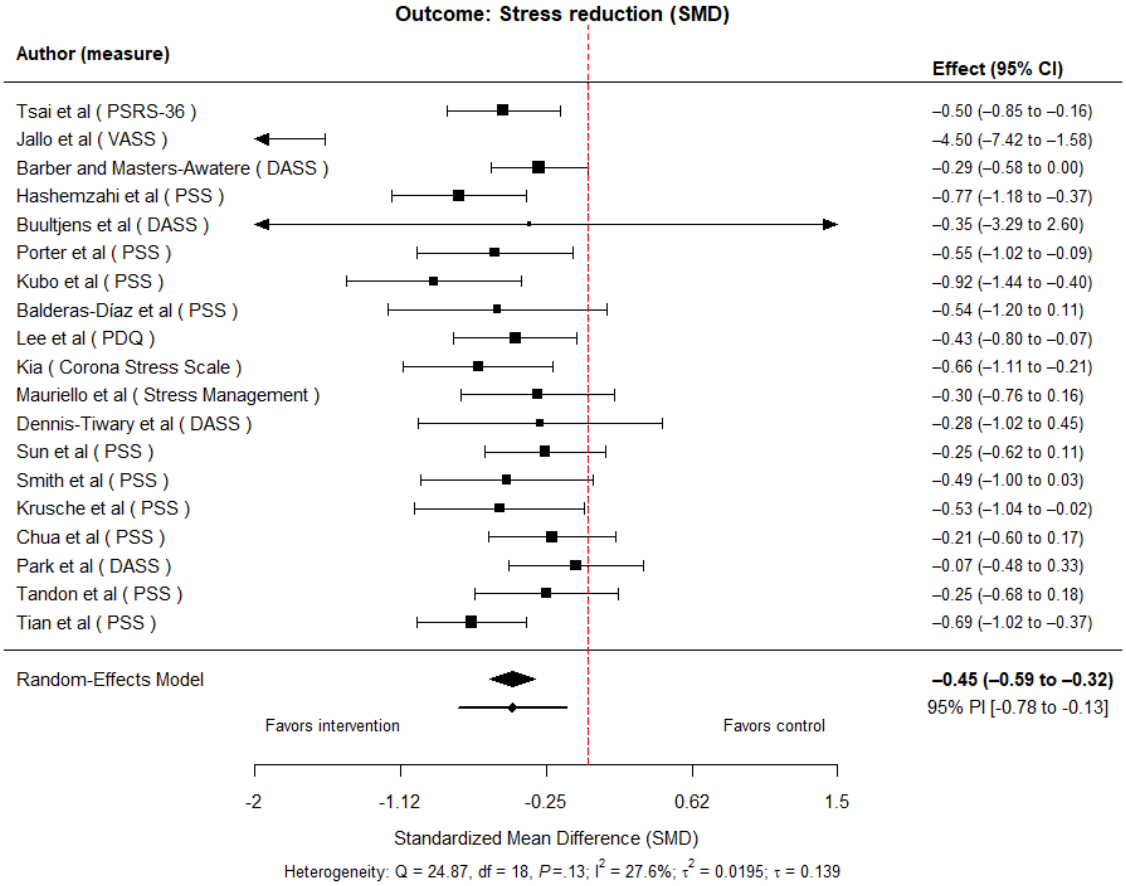
Results of Syntheses

The meta-analysis results of these studies were as follows.

Overall

Across 19 studies, effect sizes were calculated and synthesized. A random-effects meta-analysis using the Hartung–Knapp–Sidik–Jonkman method yielded a pooled average effect of SMD –0.45 (95% CI –0.59 to –0.32), indicating a statistically significant reduction in stress (t_{18}= –6.97, P<.001). Between-study heterogeneity was low to moderate (τ^2=0.0195, τ=0.1395, I^2=27.62%, Q (18)=24.87, P=.13). The 95% PI ranged from –0.78 to –0.13, indicating that the true effects in new settings are expected to remain beneficial overall, although the magnitude of benefit may vary across implementations (Figure 3 [23-41]).

Figure 3. Forest plot of the overall effect of digital interventions on stress reduction [23-41]. DASS: Depression Anxiety Stress Scale; PDQ: prenatal distress questionnaire; PSRS-36: Pregnancy Stress Rating Scale-36; PSS: Perceived Stress Scale; VASS: Visual Analog Stress Scale.



By Strategies

When the intervention effects were categorized by strategies, the pooled estimates were directionally beneficial across subgroups, but statistical certainty varied:

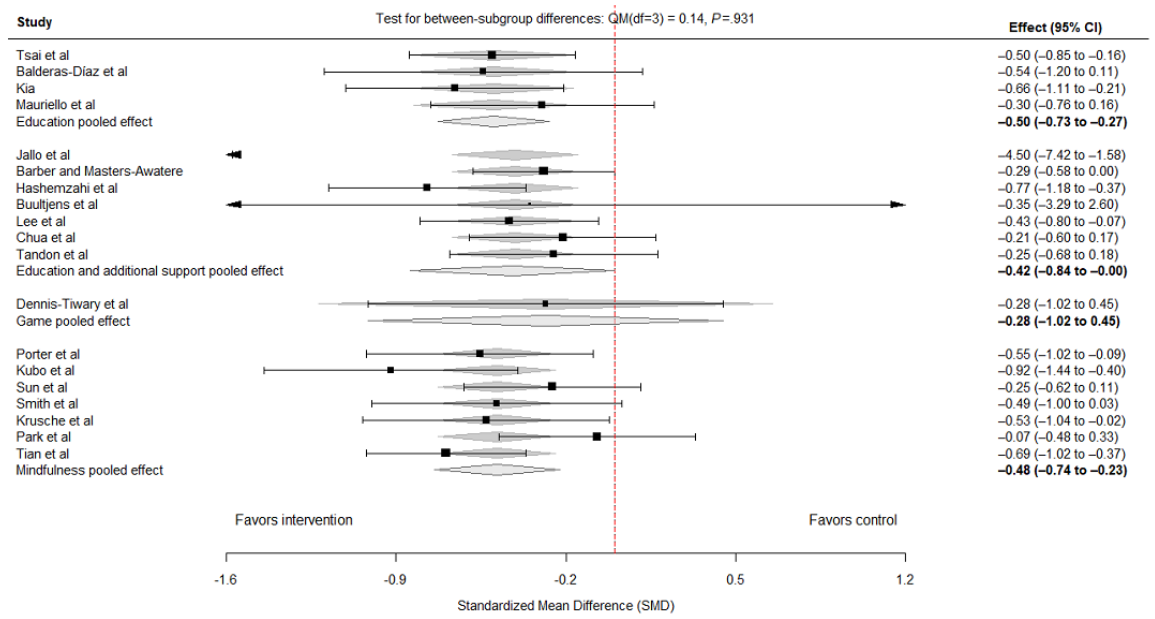
- Education only (k=4): SMD –0.50, 95% CI (–0.73 to –0.27), I^2=0%, τ^2=0
- Education with additional support (k=7): SMD –0.42, 95% CI (–0.84 to 0.00),I^2=52.8%, τ^2=0.059
- Game-based interventions (k=1): SMD –0.28, 95% CI (–1.02 to 0.45),I^2=0%, τ^2=0

XSL•FO
RenderX

- Mindfulness (k=7): SMD –0.48, 95% CI (–0.74 to –0.23),I^2=39.4%, τ^2=0.031

These findings suggest that while several strategies (education-only and mindfulness) show statistically significant average effects, estimates for strategies supported by fewer studies (game-based) or with greater heterogeneity (education with additional support) are less precise and should be interpreted cautiously (Figure 4 [23-41]). A test for between-subgroup differences based on meta-regression showed no statistically significant differences in effect sizes across intervention strategies (Q statistic for moderators: QM_3=0.14; P=.93).

Figure 4. Forest plot for the effects of the intervention by strategies [23-41].

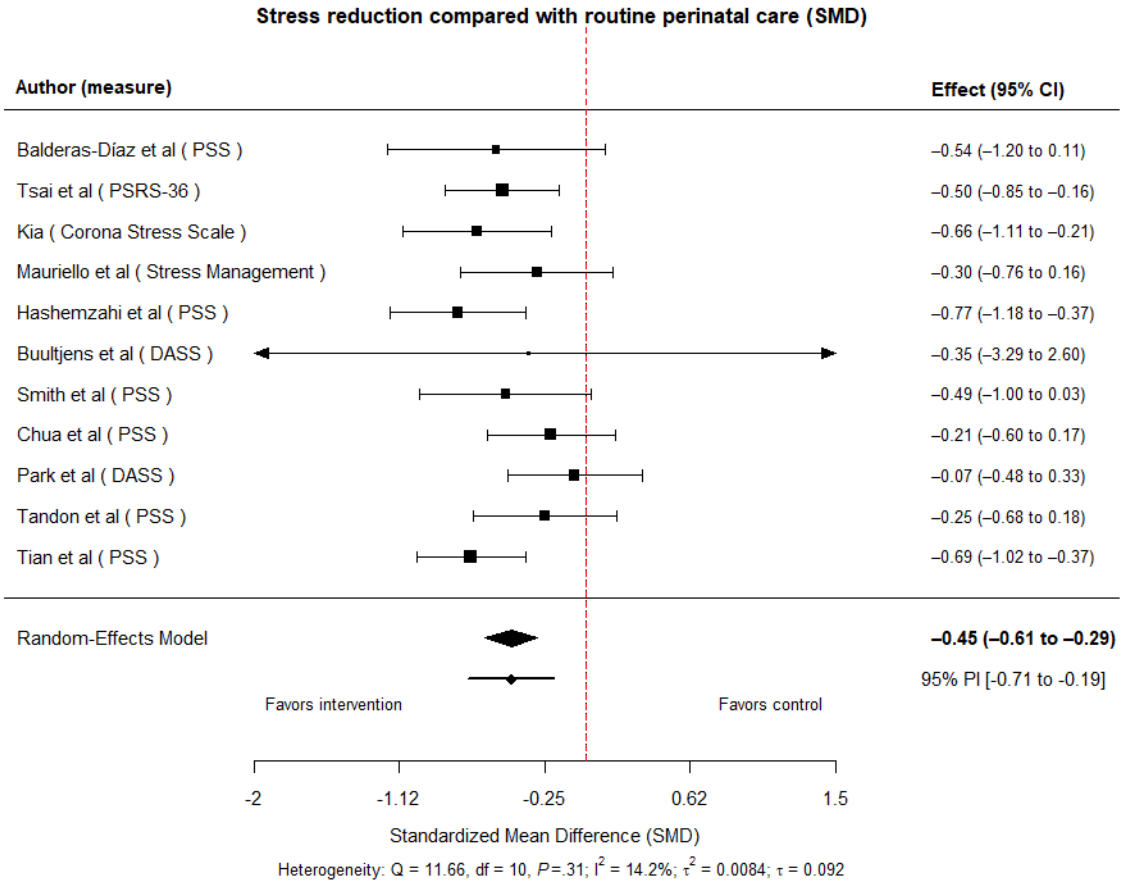


Comparison With Routine Antenatal Care

Studies comparing DHIs with routine antenatal care/usual care showed a significant reduction in stress (pooled average effect SMD −0.45, 95% CI −0.61 to −0.29, t_{10}=−6.27, P<.001). Heterogeneity was low (Q (10)=11.66, P=.31; I^2=14.24%, τ^2=0.0084; Figure 5 [23,26,29-34,37-39]).

XSL•FO
RenderX

Figure 5. Forest plot for the effects compared with routine perinatal care [23,26,29-34,37-39].



Certainty of Evidence

Twelve of the 19 included effect sizes were derived from RCTs. Using the GRADE, the certainty of evidence for stress reduction was rated as moderate, downgraded by one level for risk of bias (1/12 RCTs rated overall high risk, and 2/12 rated some concerns in RoB 2), with no serious concerns identified for inconsistency, indirectness, imprecision, or publication bias.

The certainty of evidence was rated as moderate, downgraded by one level for risk of bias. The pooled effect estimate for RCTs was SMD −0.39 (95% CI −0.52 to −0.26), reflecting a moderate level of certainty in the stress-reducing effect of digital interventions (Multimedia Appendix 3).

Discussion

Principal Findings

Guided by our objectives to characterize antenatal DHIs for stress management, evaluate their effectiveness, and summarize intervention strategies associated with engagement, this systematic review and meta-analysis found that digitally delivered stress-management interventions are generally effective in reducing stress among pregnant women compared with routine care or control conditions. These findings support the growing role of DHIs as a feasible and scalable approach for addressing pregnancy-related stress. Consistent with an umbrella review of digital interventions for perinatal psychological outcomes, digital approaches appear capable of reducing stress symptoms overall, although prior syntheses often pooled stress with other mental health outcomes [14]. They are partially aligned with a recent systematic review and meta-analysis of digitally delivered mindfulness interventions in pregnant women, which reported clear benefits for depression and anxiety and more variable effects for stress [11]. Our findings extend this evidence by focusing specifically on stress as a primary outcome during pregnancy.

When interventions were classified by dominant strategy, mindfulness-based approaches tended to show the most consistent signals of stress reduction. This pattern aligns with earlier reviews of perinatal mindfulness interventions and with emerging trials of digital mindfulness programs tailored to pregnancy [13,42]. Recent randomized trials suggest that mobile-delivered mindfulness programs can reduce depression, underscoring the potential value of skills-based, self-directed content when engagement is maintained [43]. Education-focused programs were also beneficial, particularly when paired with supportive elements such as coaching, peer interaction, or structured follow-up, suggesting that informational content alone may be insufficient for sustained stress reduction. In contrast, gamified interventions were evaluated in only one

XSL•FO
RenderX

study and showed mixed results, with improvements detected in a biological stress indicator but not in self-reported stress, underscoring the need for further research using multiple outcome measures. Importantly, although subgroup-specific point estimates varied across strategies, the formal test for between-subgroup differences was not statistically significant. This indicates insufficient meta-analytic evidence to conclude that any single strategy is superior for stress reduction during pregnancy. Given the limited number of studies within some subgroups and the resulting low statistical power to detect moderator effects, these subgroup patterns should be interpreted cautiously and viewed as exploratory rather than confirmatory.

Beyond intervention strategy, how these programs were delivered also appeared to shape feasibility and engagement. Across studies, most interventions were delivered via mobile apps, with fewer web-based platforms and telemedicine approaches, reflecting contemporary trends in digital health delivery for pregnant populations. The predominance of mobile app–based interventions observed in this review mirrors widespread smartphone use among pregnant women and supports the feasibility of delivering brief, scalable stress-management content within routine antenatal care [16,44]. The increasing dominance of mobile-delivered interventions in more recent studies highlights a shift toward app-based platforms that can leverage a wider range of interactive functions—such as continuous monitoring, automated feedback, and reminders—which may enhance engagement and responsiveness compared with more static web-based programs. While earlier meta-analyses of internet-delivered psychological interventions primarily emphasized benefits for depression and anxiety outcomes [45], our findings suggest that web-based interventions can also contribute to stress reduction during pregnancy, albeit based on a smaller number of trials.

Beyond delivery mode alone, consistent patterns emerged in how interventions were constructed. Most DHIs combined mobile or web-based delivery with self-guided use, supported by functional features such as self-monitoring, automated feedback, reminders, or asynchronous communication, and delivered therapeutic content including psychoeducation, mindfulness practices, relaxation techniques, or cognitive behavioral therapy–based skills. Multi-component designs were the norm rather than the exception, indicating that digital stress-management interventions during pregnancy rarely rely on a single active ingredient. Compared with earlier reviews that grouped interventions primarily by delivery methods or broad perinatal mental health outcomes [13,46], the present review adopts a strategy-oriented perspective that provides a more clinically meaningful framework for understanding how digital interventions may be designed to reduce stress during pregnancy. These findings highlight the need for future research that directly compares single-strategy interventions (such as mindfulness alone or psychoeducation alone) with multi-component interventions, in order to clarify whether additive or synergistic effects contribute to stress reduction in this population.

Interpretation of these findings should consider both between-study variability and study quality. Although between-study variability did not indicate extreme inconsistency, the prediction interval suggests that beneficial effects are likely across future settings, albeit with meaningful variation in magnitude. Reporting prediction intervals alongside CIs helps distinguish the average effect from the distribution of effects expected in real-world implementations [19]. Subgroup findings should therefore be interpreted in the context of overall consistency rather than as definitive evidence of differential effectiveness across strategies [47]. Finally, assessments of small-study effects should be interpreted cautiously, as funnel plot asymmetry may reflect mechanisms beyond publication bias, including true heterogeneity or methodological differences across studies [48].

Clinical Implications

These results suggest that DHIs can serve as a practical adjunct to routine antenatal care by extending access to evidence-based stress-management skills, particularly for women who face barriers to in-person services. Interventions that combine structured mindfulness practice or targeted education with functional support (eg, reminders, progress feedback, brief coaching, or moderated peer support) may be especially useful for maintaining adherence and reinforcing skills [16]. Implementation may be facilitated by integrating these tools into prenatal education pathways and ensuring appropriate guidance for women with elevated distress who may need stepped-up care. Recent trials suggest that mobile-delivered mindfulness may improve depressive symptoms during pregnancy, but sustained engagement and equitable access remain key implementation challenges [43]. From a design perspective, the most promising programs provided clear goals, brief and repeatable programs, and opportunities for personalization. Future DHIs could also explore safe integration of passive sensing or wearable-enabled feedback to tailor content in real time, while ensuring data privacy and minimizing user burden [47,48]. Recent trials also highlight the practical importance of reporting engagement metrics (eg, module completion, frequency of practice, and prompt response) so that implementation decisions can be based not only on efficacy but also on real-world use patterns [43].

Limitations

Several limitations should be noted. The number of trials within some strategy subgroups was small, limiting precision and the ability to detect differential effects. Our restriction to English-language publications and the use of only 4 databases may have missed relevant studies, and most included trials were conducted in high-income countries, which may limit generalizability to low- and middle-income contexts. Outcomes were primarily self-reported, raising the possibility of reporting bias. In addition, incomplete reporting of intervention details and adherence in some trials constrained the interpretation of which components drove benefit. Because the review was not prospectively registered, transparency may be lower than in preregistered reviews; however, we sought to mitigate this limitation by adhering to PRISMA guidance and applying the GRADE approach to characterize certainty of evidence. Moreover, given the rapid evolution of digital perinatal interventions, studies published after our search window may

XSL•FO
RenderX

influence future pooled estimates; ongoing evidence surveillance or living-review approaches may therefore be warranted [14,43].

Conclusions

Overall, this systematic review and meta-analysis provide novel evidence that digitally delivered stress-management interventions can meaningfully reduce stress during pregnancy, highlighting the potential of digital health approaches as a complementary component of antenatal care. Unlike previous reviews that primarily categorized interventions by delivery platform or aggregated stress with other psychological outcomes, this study offers an innovative strategy-based synthesis that disentangles how specific intervention contents and supportive features contribute to stress reduction in pregnant women. By distinguishing intervention strategies rather than technologies alone, the findings advance understanding of what works within digital stress-management programs during pregnancy.

Across studies, the most consistent benefits were observed for mindfulness-based approaches and for educational interventions combined with supportive features such as coaching or peer interaction, underscoring the importance of active skill-building and engagement rather than information provision alone. Importantly, while digital tools are not intended to replace clinician-delivered care, the results indicate that they can extend access to evidence-based coping strategies and provide scalable, low-intensity support that fits within real-world antenatal workflows. As mobile app–based interventions continue to expand, their ability to integrate multiple functions—such as self-monitoring, feedback, and reminders—positions them as particularly feasible and adaptable tools for routine maternity care.

By clarifying the effectiveness of digital stress-management interventions specifically during pregnancy and by framing interventions according to their strategic components, this review contributes actionable insights for clinicians, researchers, and health system planners. The findings support the integration of digital stress-management programs as an adjunct to standard antenatal services, with the potential to improve reach, equity, and continuity of psychosocial support for pregnant women in diverse care settings.

Acknowledgments

We are grateful to Sangkeun Hyon, Dong-Hwan Shin, and Saeam Kim, medical librarians at the Soonchunhyang University Seoul Hospital, for their contributions to database searches and the collection of full text articles for this study. Generative artificial intelligence was used to support English translation and language editing. The authors carefully reviewed and revised all artificial intelligence–assisted text and remain fully responsible for the content of the manuscript.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (number 2023R1A2C1003355 and RS-2023-00239536) as well as by Soonchunhyang University.

Authors' Contributions

JIK contributed to the conceptualization, data curation, and formal analysis, and was involved in writing the original draft, reviewing and editing the manuscript, and providing supervision. JYL contributed to conceptualization and methodology, conducted data curation and formal analysis, and participated in writing the original draft and reviewing and editing the manuscript. SHP contributed to data curation and formal analysis and participated in writing the original draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA 2020 Checklist and PRISMA-S Checklist.
[DOCX File , 40 KB - mhealth_v14i1e66267_app1.docx]

Multimedia Appendix 2
Detailed Search Strategies across Databases.
[DOCX File , 16 KB - mhealth_v14i1e66267_app2.docx]

Multimedia Appendix 3
Effect Estimates from Randomized Controlled Trials with GRADE Quality Assessment.
[DOCX File , 25 KB - mhealth_v14i1e66267_app3.docx]

XSL•FO
RenderX

References

1. Maharramova A, Gur EY. The pregnant women's perception of risks and pregnancy stress levels: a cross-sectional study from Turkey. Rev Assoc Med Bras (1992) 2024;70(6):e20231270 [FREE Full text] [doi: 10.1590/1806-9282.20231270] [Medline: 39045948]
2. Williamson SP, Moffitt RL, Broadbent J, Neumann DL, Hamblin PS. Coping, wellbeing, and psychopathology during high-risk pregnancy: a systematic review. Midwifery 2023;116:103556. [doi: 10.1016/j.midw.2022.103556] [Medline: 36427386]
3. Caparros-Gonzalez RA, Torre-Luque ADL, Romero-Gonzalez B, Quesada-Soto JM, Alderdice F, Peralta-Ramírez MI. Stress during pregnancy and the development of diseases in the offspring: a systematic-review and meta-analysis. Midwifery 2021;97:102939. [doi: 10.1016/j.midw.2021.102939] [Medline: 33647755]
4. Cho H, Kim J. Moderating effect of general self-efficacy on the relationship between pregnancy stress, daily hassles stress, and preterm birth risk in women experiencing preterm labor: a cross-sectional study. J Korean Acad Nurs 2024;54(3):329-339. [doi: 10.4040/jkan.24008] [Medline: 39248420]
5. Missler M, Donker T, Beijers R, Ciharova M, Moyse C, de Vries R, et al. Universal prevention of distress aimed at pregnant women: a systematic review and meta-analysis of psychological interventions. BMC Pregnancy Childbirth 2021;21(1):276 [FREE Full text] [doi: 10.1186/s12884-021-03752-2] [Medline: 33794828]
6. Nasrollahi M, Ghazanfar Pour M, Ahmadi A, Mirzaee M, Alidousti K. Effectiveness of mindfulness-based stress reduction on depression, anxiety, and stress of women with the early loss of pregnancy in southeast Iran: a randomized control trial. Reprod Health 2022 Dec 29;19(1):233 [FREE Full text] [doi: 10.1186/s12978-022-01543-2] [Medline: 36581926]
7. Corrigan L, Moran P, McGrath N, Eustace-Cook J, Daly D. The characteristics and effectiveness of pregnancy yoga interventions: a systematic review and meta-analysis. BMC Pregnancy Childbirth 2022;22(1):250 [FREE Full text] [doi: 10.1186/s12884-022-04474-9] [Medline: 35337282]
8. Huo L. Effects of prenatal psychotherapies and psychosocial interventions on depressive symptoms, anxious symptoms and stress: a systematic review and network meta-analysis. Front Psychiatry 2025;16:1624924 (forthcoming) [FREE Full text]
9. Sîrbu V, David OA. Efficacy of app-based mobile health interventions for stress management: a systematic review and meta-analysis of self-reported, physiological, and neuroendocrine stress-related outcomes. Clin Psychol Rev 2024;114:102515 [FREE Full text] [doi: 10.1016/j.cpr.2024.102515] [Medline: 39522422]
10. Linardon J, Firth J, Torous J, Messer M, Fuller-Tyszkiewicz M. Efficacy of mental health smartphone apps on stress levels: a meta-analysis of randomised controlled trials. Health Psychol Rev 2024;18(4):839-852 [FREE Full text] [doi: 10.1080/17437199.2024.2379784] [Medline: 39041586]
11. Mefrouche ML, Siegmann E, Böhme S, Berking M, Kornhuber J. The effect of digital mindfulness interventions on depressive, anxiety, and stress symptoms in pregnant women: a systematic review and meta-analysis. Eur J Investig Health Psychol Educ 2023;13(9):1694-1706 [FREE Full text] [doi: 10.3390/ejihpe13090122] [Medline: 37754461]
12. Han Y, Tian Q, Xu M, Zhao W, Wang Z, Zhang W. Effects of nurse-led e-health interventions on the health-related outcomes of pregnant women: a systematic review. J Clin Nurs 2025;34(1):88-107. [doi: 10.1111/jocn.17560] [Medline: 39568149]
13. Stentzel U, Grabe HJ, Schmidt S, Tomczyk S, van den Berg N, Beyer A. Mental health-related telemedicine interventions for pregnant women and new mothers: a systematic literature review. BMC Psychiatry 2023;23(1):292 [FREE Full text] [doi: 10.1186/s12888-023-04790-0] [Medline: 37118689]
14. Lau Y, Chew HSJ, Ang WHD, Ang WW, Yeo CY, Lim GZQ, et al. Effects of digital health interventions on the psychological outcomes of perinatal women: umbrella review of systematic reviews and meta-analyses. Health Psychol Rev 2024;18(2):229-254 [FREE Full text] [doi: 10.1080/17437199.2023.2185654] [Medline: 36919443]
15. Mohr DC, Schueller SM, Riley WT, Brown CH, Cuijpers P, Duan N, et al. Trials of intervention principles: evaluation methods for evolving behavioral intervention technologies. J Med Internet Res 2015;17(7):e166 [FREE Full text] [doi: 10.2196/jmir.4391] [Medline: 26155878]
16. Paganini S, Meier E, Terhorst Y, Wurst R, Hohberg V, Schultchen D, et al. Stress management apps: systematic search and multidimensional assessment of quality and characteristics. JMIR Mhealth Uhealth 2023;11:e42415 [FREE Full text] [doi: 10.2196/42415] [Medline: 37642999]
17. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
18. Rethlefsen ML, Kirtley S, Waffenschmidt S, Patricia Ayala A, Moher D. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev 2021;10(1). [doi: 10.31219/osf.io/sfc38]
19. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
20. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]

XSL•FO
RenderX

21. Borenstein M. How to understand and report heterogeneity in a meta-analysis: the difference between I-squared and prediction intervals. Integr Med Res 2023;12(4):101014 [FREE Full text] [doi: 10.1016/j.imr.2023.101014] [Medline: 38938910]
22. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard Dersimonian-Laird method. BMC Med Res Methodol 2014;14:25 [FREE Full text] [doi: 10.1186/1471-2288-14-25] [Medline: 24548571]
23. Balderas-Díaz S, Rodríguez-Fórtiz MJ, Garrido JL, Bellido-González M, Guerrero-Contreras G. A psycho-educational intervention programme for parents with SGA foetuses supported by an adaptive mHealth system: design, proof of concept and usability assessment. BMC Med Inform Decis Mak 2022;22(Suppl 4):291 [FREE Full text] [doi: 10.1186/s12911-022-02036-9] [Medline: 36357878]
24. Lee Y, Kim S, Choi S. Effectiveness of mobile-based intervention self-care for pregnant women at work: a randomized controlled trial. J Occup Health 2023;65(1):e12402 [FREE Full text] [doi: 10.1002/1348-9585.12402] [Medline: 37144249]
25. Sun Y, Li Y, Wang J, Chen Q, Bazzano AN, Cao F. Effectiveness of smartphone-based mindfulness training on maternal perinatal depression: randomized controlled trial. J Med Internet Res 2021;23(1):e23410 [FREE Full text] [doi: 10.2196/23410] [Medline: 33502326]
26. Smith R, Mahnert ND, Foote J, Saunders KT, Mourad J, Huberty J. Mindfulness effects in obstetric and gynecology patients during the coronavirus disease 2019 (COVID-19) pandemic: a randomized controlled trial. Obstet Gynecol 2021;137(6):1032-1040 [FREE Full text] [doi: 10.1097/AOG.0000000000004316] [Medline: 33957663]
27. Krusche A, Dymond M, Murphy SE, Crane C. Mindfulness for pregnancy: a randomised controlled study of online mindfulness during pregnancy. Midwifery 2018;65:51-57. [doi: 10.1016/j.midw.2018.07.005] [Medline: 30099285]
28. Dennis-Tiwary TA, Denefrio S, Gelber S. Salutary effects of an attention bias modification mobile application on biobehavioral measures of stress and anxiety during pregnancy. Biol Psychol 2017;127:148-156 [FREE Full text] [doi: 10.1016/j.biopsycho.2017.05.003] [Medline: 28478138]
29. Kia ZR. The effect of educational intervention with mobile health technology on COVID-19 induced stress among pregnant women: a randomized controlled trial. J Midwifery Reprod Health 2023;11(3):3782-3793 [FREE Full text] [doi: 10.22038/JMRH.2022.66480.1940]
30. Mauriello LM, Van Marter DF, Umanzor CD, Castle PH, de Aguiar EL. Using mHealth to deliver behavior change interventions within prenatal care at community health centers. Am J Health Promot 2016;30(7):554-562. [doi: 10.4278/ajhp.140530-QUAN-248] [Medline: 26305603]
31. Chua JYX, Choolani M, Chee CYI, Yi H, Chan YH, Lalor JG, et al. The effectiveness of parentbot - a digital healthcare assistant - on parenting outcomes: a randomized controlled trial. Int J Nurs Stud 2024;160:104906. [doi: 10.1016/j.ijnurstu.2024.104906] [Medline: 39305680]
32. Park S, Cho HY, Park JY, Chung K, Jhung K. Development and evaluation of a mindfulness-based mobile intervention for perinatal mental health: randomized controlled trial. J Med Internet Res 2025;27:e56601 [FREE Full text] [doi: 10.2196/56601] [Medline: 39823585]
33. Tandon SD, Moskowitz JT, Edwards RC, Zhang Y, Giase G, Sinche B, et al. Effects of a personalized stress management intervention on maternal mental health: a randomized clinical trial. Arch Womens Ment Health 2025;28(6):1585-1595. [doi: 10.1007/s00737-025-01619-5] [Medline: 40960524]
34. Tian Y, Ma R, Cui N, Wang J, Huang Y, Guo K, et al. Effects of digital mindfulness training for couples on psychological distress and infant neuropsychological development: randomized controlled trial. J Med Internet Res 2025;27:e77260 [FREE Full text] [doi: 10.2196/77260] [Medline: 41271207]
35. Porter AC, Hunter S, Noonan K, Hoffman MC. A mindfulness application for reducing prenatal stress. J Midwifery Womens Health 2022;67(4):442-447 [FREE Full text] [doi: 10.1111/jmwh.13359] [Medline: 35403807]
36. Jallo N, Thacker LR, Menzies V, Stojanovic P, Svikis DS. A stress coping app for hospitalized pregnant women at risk for preterm birth. MCN Am J Matern Child Nurs 2017;42(5):257-262. [doi: 10.1097/NMC.0000000000000355] [Medline: 28817447]
37. Hashemzahi M, Khayat S, Khazaeian S. Effect of COVID-19 self-care training via telemedicine on perceived stress and corona disease anxiety in pregnant women: a quasi-experimental study. J Midwifery Reprod Health 2022;10(1):3066-3074 [FREE Full text] [doi: 10.22038/jmrh.2021.60589.1728]
38. Tsai Y, Hsu Y, Hou T, Chang C. Effects of a web-based antenatal care system on maternal stress and self-efficacy during pregnancy: a study in Taiwan. J Midwifery Womens Health 2018;63(2):205-213. [doi: 10.1111/jmwh.12685] [Medline: 29533525]
39. Buultjens M, Gill J, Fielding J, Lambert KA, Vondeling K, Mastwyk SE, et al. Maternity care during a pandemic: can a hybrid telehealth model comprising group interdisciplinary education support maternal psychological health? Women Birth 2023;36(3):305-313 [FREE Full text] [doi: 10.1016/j.wombi.2022.09.007] [Medline: 36184532]
40. Kubo A, Aghaee S, Kurtovich EM, Nkemere L, Quesenberry CP, McGinnis MK, et al. mHealth mindfulness intervention for women with moderate-to-moderately-severe antenatal depressive symptoms: a pilot study within an integrated health care system. Mindfulness (N Y) 2021;12(6):1387-1397 [FREE Full text] [doi: 10.1007/s12671-021-01606-8] [Medline: 33723491]

XSL•FO
RenderX

41. Barber CC, Masters-Awatere B. Positively pregnant: development and piloting of a mobile app for social and emotional well-being in pregnancy. Appl Psychol Health Well Being 2022;14(4):1255-1272. [doi: 10.1111/aphw.12333] [Medline: 34959260]
42. Lucena L, Frange C, Pinto ACA, Andersen ML, Tufik S, Hachul H. Mindfulness interventions during pregnancy: a narrative review. J Integr Med 2020;18(6):470-477. [doi: 10.1016/j.joim.2020.07.007] [Medline: 32798196]
43. Kim M, Park JY, Park S, Chung K, Cho HY, Do G, et al. Effectiveness of a mindfulness-based mobile intervention for improving perinatal mental health and reducing depression during pregnancy: randomized controlled trial. J Med Internet Res 2025;27:e75630 [FREE Full text] [doi: 10.2196/75630] [Medline: 41037777]
44. Nabovati E, Farzandipour M, Vahedpoor Z, Akbari H, Anvari S, Sharif R, et al. Pregnant women's use and attitude toward mobile phone features for self-management. BMC Med Inform Decis Mak 2023;23(1):77 [FREE Full text] [doi: 10.1186/s12911-023-02172-w] [Medline: 37101302]
45. Neo HS, Tan JH, Ang WHD, Lau Y. Internet-delivered psychological interventions for reducing depressive, anxiety symptoms and fear of childbirth in pregnant women: a meta-analysis and meta-regression. J Psychosom Res 2022;157:110790. [doi: 10.1016/j.jpsychores.2022.110790] [Medline: 35367919]
46. Nair U, Armfield NR, Chatfield MD, Edirippulige S. The effectiveness of telemedicine interventions to address maternal depression: a systematic review and meta-analysis. J Telemed Telecare 2018;24(10):639-650. [doi: 10.1177/1357633X18794332] [Medline: 30343660]
47. Borenstein M, Higgins JPT. Meta-analysis and subgroups. Prev Sci 2013;14(2):134-143. [doi: 10.1007/s11121-013-0377-7] [Medline: 23479191]
48. Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011;343:d4002. [doi: 10.1136/bmj.d4002] [Medline: 21784880]

Abbreviations

DASS: Depression Anxiety Stress Scale
DHI: digital health intervention
GRADE: Grading of Recommendations, Assessment, Development and Evaluation
MeSH: Medical Subject Headings
PICOS: population, intervention, comparison, outcome, and study design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension
RoB 2: revised Cochrane risk-of-bias tool
ROBINS-I: risk of bias in nonrandomized studies of interventions
SMD: standardized mean difference

Edited by S Brini; submitted 09.Sep.2024; peer-reviewed by E Motrico, A Finlay; comments to author 12.Mar.2025; revised version received 28.Dec.2025; accepted 28.Dec.2025; published 26.Jan.2026.

Please cite as:
Kim JI, Lee JY, Park SH
Evaluating the Effectiveness of Digital Interventions for Stress Management in Pregnant Women: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2026;14:e66267
URL: https://mhealth.jmir.org/2026/1/e66267
doi:10.2196/66267
PMID:

©Jeung-Im Kim, Joo Yun Lee, So-Hee Park. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 26.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Calorie-Counting Apps for Monitoring and Managing Calorie Intake in Adults Living With Weight-Related Chronic Diseases: Decade-Long Scoping Review (2013-2024)

Kaylee Dugas[1,2*], MD; Marie-Andrée Giroux[1,2*], MD, PhD; Abdelatif Guerroudj[1,2*], MD; Jazna Leger[1,2*], MD; Asal Rouhafzay[1,3*], PhD; Ghazal Rouhafzay[4], PhD; Jalila Jbilou[1,2,3], MD, PhD

[1]Centre de Formation Médicale du Nouveau-Brunswick, Moncton, NB, Canada

[2]Université de Sherbrooke, Sherbrooke, QC, Canada

[3]School of Psychology, Université de Moncton, Moncton, NB, Canada

[4]Department of Computer Science, Université de Moncton, Moncton, NB, Canada

[*]these authors contributed equally

Corresponding Author:
Jalila Jbilou, MD, PhD
School of Psychology
Université de Moncton
18, avenue Antonine-Maillet Moncton
Moncton, NB, E1A 3E9
Canada
Phone: 1 5068584000 ext 4931
Email: jalila.jbilou@umoncton.ca

Abstract

Background: Overweight and obesity, as defined by the World Health Organization, correspond to BMI values of 25-29.9 kg/m² for overweight and ≥30 kg/m² for obesity. Both conditions remain major public health challenges worldwide due to their strong link with type 2 diabetes, cardiovascular disease, and hypertension, which place a heavy clinical and economic burden on health care systems. In Canada, obesity rates are notably high, with vulnerable populations disproportionately affected due to socioeconomic barriers, limited access to preventive care, and higher comorbidity rates. Calorie-counting Mobile health (mHealth) apps support dietary self-monitoring and weight control; however, varied designs and evidence complicate assessment of feasibility and effectiveness.

Objective: This study aimed to systematically evaluate the structure and content of 46 calorie-counting apps, identify factors related to their acceptability and feasibility among adults living with obesity or weight-related chronic diseases, and formulate evidence-based recommendations for app developers, clinicians, and researchers.

Methods: We conducted a scoping review of papers on calorie-counting apps published between January 2013 and March 2024. We identified 771 records and applied the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) eligibility criteria. Data on app functions, features, user engagement, and acceptability and feasibility among adults with overweight or related chronic conditions were synthesized to generate practical recommendations for designing and clinically implementing calorie-counting apps.

Results: A total of 68 studies met the inclusion criteria. Randomized controlled trials (23/68, 34%) and cohort studies (16/68, 24%) were the most common designs. Most studies targeted adults with overweight or obesity (53/68, 78%), while diabetes and hypertension were less frequently represented. In total, 46 distinct calorie-counting apps were identified, with MyFitnessPal and Lose It! being the most frequently studied. Nearly all apps (45/46, 98%) offered calorie logging, often through manual entry supported by food databases, and about half included goal-setting features. The most cited acceptability factors were personalization, automation, user-friendly design, and data sharing with health care professionals; barriers included technical issues, limited food databases, and manual entry. Adherence declined over time. For example, self-monitoring with MyFitnessPal decreased from 5.4 to 1.4 days per week from weeks 4 to 12, while use of Lose It! dropped to 4 days per week by the end of 12 weeks. Twelve recommendations were developed to enhance the feasibility and acceptability of calorie-counting apps for people living with weight-related chronic diseases.

XSL•FO
RenderX

Conclusions: Calorie-counting apps hold potential as tools for supporting individuals living with obesity and weight-related chronic diseases. To improve clinical usability, app developers should enhance engagement via personalization and automation, ensuring food database comprehensiveness, and minimizing tracking effort. Further research should validate effectiveness and strategies for sustaining adherence, thereby informing development of user-friendly mHealth interventions.

(JMIR Mhealth Uhealth 2026;14:e64139) doi:10.2196/64139

KEYWORDS

calorie counting; clinical feasibility; mHealth; obesity management; weight-related chronic diseases

Introduction

One of the major challenges that health care systems around the world are currently facing is the prevalence of weight-related chronic diseases [1]. According to the World Health Organization, an estimated 2.5 billion adults aged 18 years and older were living with overweight, and 890 million were living with obesity in 2022. This corresponds to 43% and 16% of the adult population being overweight or obese, respectively. In Canada, the prevalence of obesity is even higher, with just over 26% of Canadians having a BMI of 30 or above in 2016 [2]. These Canadian statistics vary greatly within the country, with Atlantic provinces showing the highest figures. Notably, New Brunswick currently has the highest rates of obesity at 34.1%. While there can be many contributing factors, low income (42% of households earning less than CAD $60,000 [US $43,842] a year), low education rates (40% of the population having not completed postsecondary studies), and poor health behaviors account for the most significantly lacking determinants of health [3]. These figures are worrisome not only because of their drastic increase within the past 30 years [2], but also due to the well-established link between obesity and several health conditions, notably type 2 diabetes, cardiovascular diseases, and hypertension [4-6]. These weight-related chronic illnesses all lead to significant costs for society [7].

Tackling the epidemic of obesity and weight-related chronic diseases can be complex, especially when considering the impact of socioeconomic factors linked to obesity [2]. In general, the therapeutic approach includes interventions aiming to create a negative calorie balance, which can notably be achieved by dietary self-monitoring [8]. Current dietary self-monitoring methods include the traditional paper food diaries, wearable devices, websites, personal digital assistants, and mobile phone apps [9]. With the current smartphone ownership rates at 90% in certain countries, such as the United States [10], and the COVID-19 pandemic, which encouraged the use of telemedicine [11], mobile health (mHealth) apps in the form of calorie-counting apps show a promising field for calorie intake control. Such apps represent novel approaches that should be included in dieticians' and health care professionals' toolkit to help patients manage weight and weight-related chronic illnesses [12].

The increase in popularity of calorie-counting apps has prompted many studies to attempt to establish their efficacy in achieving weight loss [13,14]. A systematic review and meta-analysis of 12 randomized controlled trials demonstrated that mobile phone apps help achieve significant weight loss and BMI reduction in comparison with control groups, thus demonstrating the potential for mobile phone apps in weight management [15]. Calorie-counting apps are very heterogeneous in terms of functionalities, features, and objectives, making it difficult to compare them and provide evidence [16]. This, combined with the fact that calorie-counting apps often see a drop in adherence after only 3 to 5 weeks of use in most cases [17], contributes to difficulties with establishing clinical feasibility. Therefore, before implementing these apps as weight management options in a clinical setting, there is a clear need for studies to analyze key design elements and to identify the factors increasing the acceptability and feasibility of these calorie-counting apps. Some studies have already begun tackling this topic, but have either not evaluated calorie-counting apps specifically or have only studied such questions in healthy adult populations, and not in patients living with weight-related chronic illnesses.

As illustrated, it is urgent to establish the state of knowledge on existing apps and thereby identify knowledge gaps for future direction of research on calorie-counting apps in an adult population living with obesity and weight-related chronic disease. Accordingly, this paper explores the structure and content of calorie-counting apps and examines the factors influencing their acceptability and feasibility among adults living with obesity or weight-related chronic disease. The findings also inform evidence-based recommendations for app developers, clinicians, and researchers.

Methods

Overview

Scoping reviews are conducted to map existing theoretical and empirical research on a given topic. The main aims of a scoping review are to identify gaps in knowledge and theories and propose directions for future research. Thus, considering the aims of our study, a scoping review is the most appropriate and structured approach to map the literature. This scoping review was guided by the methodological guidance for scoping reviews as initially proposed by Arksey and O'Malley [18] and updated by Levac et al [19]. The manuscript was drafted according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist [20]. The study protocol was not registered and was not published.

Search Strategy

The search strategy was built through team discussion. The research team consisted of 4 medical students and 1 postdoctoral fellow in data science, working under the supervision of a public health physician holding a PhD in community health, with recognized expertise in health care management and

XSL•FO
RenderX

implementation science. Literature search was conducted on May 17, 2024, in PubMed for papers published from January 2013 to March 2024. A period of 10 years was considered sufficient, given the large number of studies published and the rapid advancement of knowledge on the topic. Our search strategy included the following five main concepts: (1) calorie-counting apps, (2) weight-related chronic diseases, (3) calorie intake or weight, (4) acceptability and feasibility, and (5) structure, features, and functionalities. For each concept, we used a list of keywords presented in Textbox 1. These terms were either formatted as free-text words or as controlled vocabulary (Medical Subject Headings; MeSH) in PubMed and IEEE. Asterisks were used at the root of certain words when relevant to extend the scope of the Boolean search. The search was restricted to titles and abstracts. Selected papers were handled by the bibliographic reference manager, Zotero [21], then exported in Microsoft Excel form for paper screening and data extraction.

XSL•FO
RenderX

Textbox 1. List of keywords used in PubMed and IEEE.

Calorie-counting apps

- App to count calories
- Apps for tracking calories
- Apps for tracking food
- Calorie counting app
- Calorie smart phone app
- Calorie tracking app
- Food logging app
- Diet app
- Diet tracker
- Diet tracking app
- Diet-tracking app
- Dietary mobile app
- Food imaging app
- Food monitoring app
- Food intake monitoring app
- Food intake tracking app
- Food picture app
- Food recognition
- Food recognition app
- Food balance estimation
- Food scan
- Food-scan
- Foodscan
- Nutrition app
- Nutrition tracking app
- Weight app
- Weight management app

Weight-related chronic diseases

- Cardiovascular disease
- Cardiovascular Diseases (MeSH)
- Chronic disease
- Chronic Disease (MeSH)
- Chronic health condition
- Chronic illness
- Diabetes
- Type 2 diabetes
- Diabetic patients
- Diabetes Mellitus (MeSH)
- Heart disease
- Hypertension
- Hypertension (MeSH)

XSL•FO
RenderX

- Long-term health condition
- Metabolic Syndrome (MeSH)
- Obese
- Obesity
- Obesity (MeSH)
- Overweight (MeSH)

Calorie intake or weight

- Body Mass Index (MeSH)
- Body Weight Changes (MeSH)
- Diet control
- Diet, Reducing (MeSH)
- Dieting
- Weight control
- Weight Gain (MeSH)
- Weight loss
- Weight Loss (MeSH)
- Weight management
- Weight monitor
- Weight reduction

Acceptability and feasibility

- Acceptability
- Feasibility
- System Usability Scale
- Usability
- Usefulness

Structure, features, and functionalities

- Content
- Design
- Feature
- Function
- Structure

To identify factors related to acceptability, each publication was independently reviewed by 4 authors to extract narrative descriptions of users' and health care professionals' perceptions of the apps, including aspects such as usability, usefulness, satisfaction, and perceived barriers. Discrepancies between reviewers were resolved through discussion with the principal investigator.

Study Selection

The initial search strategy led to 771 records. After removing duplicates, 639 studies were retrieved for screening. We applied a preset list of inclusion and exclusion criteria to identify relevant papers (Table 1). A total of 4 team members discussed the screening strategy and completed a calibration using 4.5% (29/639) papers to ensure interrater reliability. Disagreements regarding inclusion were resolved through team discussion, with the principal investigator serving as referee. Following the PRISMA flowchart (see Results section), we retained 68 papers for data extraction. The principal investigator then reviewed the full list of papers for data synthesis.

XSL•FO
RenderX

Table 1. Criteria for inclusion and exclusion of papers.

Criteria	Inclusion	Exclusion
Population	• Participants living with weight-related chronic disease (including overweight or obesity, prediabetes or diabetes, prehypertension or hypertension, cardiovascular disease, and metabolic syndrome) • Adult population (aged 18 years and older) • Human participants	• Papers targeting participants with mental illnesses or eating disorders • Papers targeting pediatric participants exclusively
Intervention	• Use of a calorie-counting app • English or French app	• Use of health apps that do not include calorie-counting features (eg, exercise tracker, heart rate tracker) • Intervention limited to using text messages only • Use of app for data collection only (eg, phone-based questionnaires)
Comparison	• Not restricted	—[a]
Outcome	• Not restricted	—
Study design	• Quantitative design • Qualitative design	• Systematic review, narrative review, scoping review, and case reports
Type of documents	• Peer-reviewed papers	• Conference abstracts, conference proceedings, book chapters, letters to editors, opinion papers, governmental reports, theses, policy reports, and papers unavailable through institutional access
Dates	• Published between January 1, 2013, and March 31, 2024	—
Language	• Full text available in English or French	• Abstract in English or French, but full text in another language • All other languages

[a]Not applicable.

Although specific keywords related to the outcome of interest in our study were used in the search strategy, the outcomes did not represent the inclusion criteria that needed to be present in papers. In fact, the final set of papers needed to include papers describing the app structure and functionalities. It was not an inclusion criterion whether these papers aimed to address the effects of calorie-counting apps on weight management, monitoring, and control, or on healthy eating, or to evaluate the acceptability and feasibility of these apps.

Data Extraction

Data were extracted using a standardized form developed in Excel and approved by all authors. The variables extracted are described in Table S1 in Multimedia Appendix 1. We systematically collected information on the following:

- Study characteristics: geographic location, research protocol, and type of weight-related chronic disease in the study population.
- App characteristics: names and objectives of the calorie-counting apps, logging-in features and functionalities, calorie-counting functionalities, goal-setting functionalities, user interaction features, and aspects related to validity, accuracy, and reliability.
- Determinants of acceptability: user and health care professional perspectives, including information on user-friendliness, usability, usefulness, satisfaction, and barriers limiting acceptability.
- Determinants of feasibility: user and health care professional perspectives, including the lack of certain functionalities, accuracy of calorie-counting features, socioeconomic barriers, and data privacy concerns.

Data on adherence were also extracted inductively during the data extraction process, as adherence emerged as a recurring theme across the included studies.

To ensure intercoder reliability, daily team meetings were organized, and a calibration exercise consisting of data extraction from 5 papers (7%) was performed. The principal investigator acted as referee in cases of disagreement.

Notably, information on intervention duration was extracted when available from the included studies; however, this variable was not reported in all sources.

Data Analysis and Synthesis

After extracting information from the included papers, we performed descriptive statistics (eg, count and proportions) to present an overview of the main characteristics of the retrieved studies (including geographic location, type of design, and weight-related chronic diseases). The results were summarized using a qualitative approach, built upon thematic content analysis, to synthesize, organize, and conceptualize findings. Whenever a dimension or an aspect could not be sorted into an existing category of concepts, we invited a meeting to discuss the possibilities of classification or the creation of a new

XSL•FO
RenderX

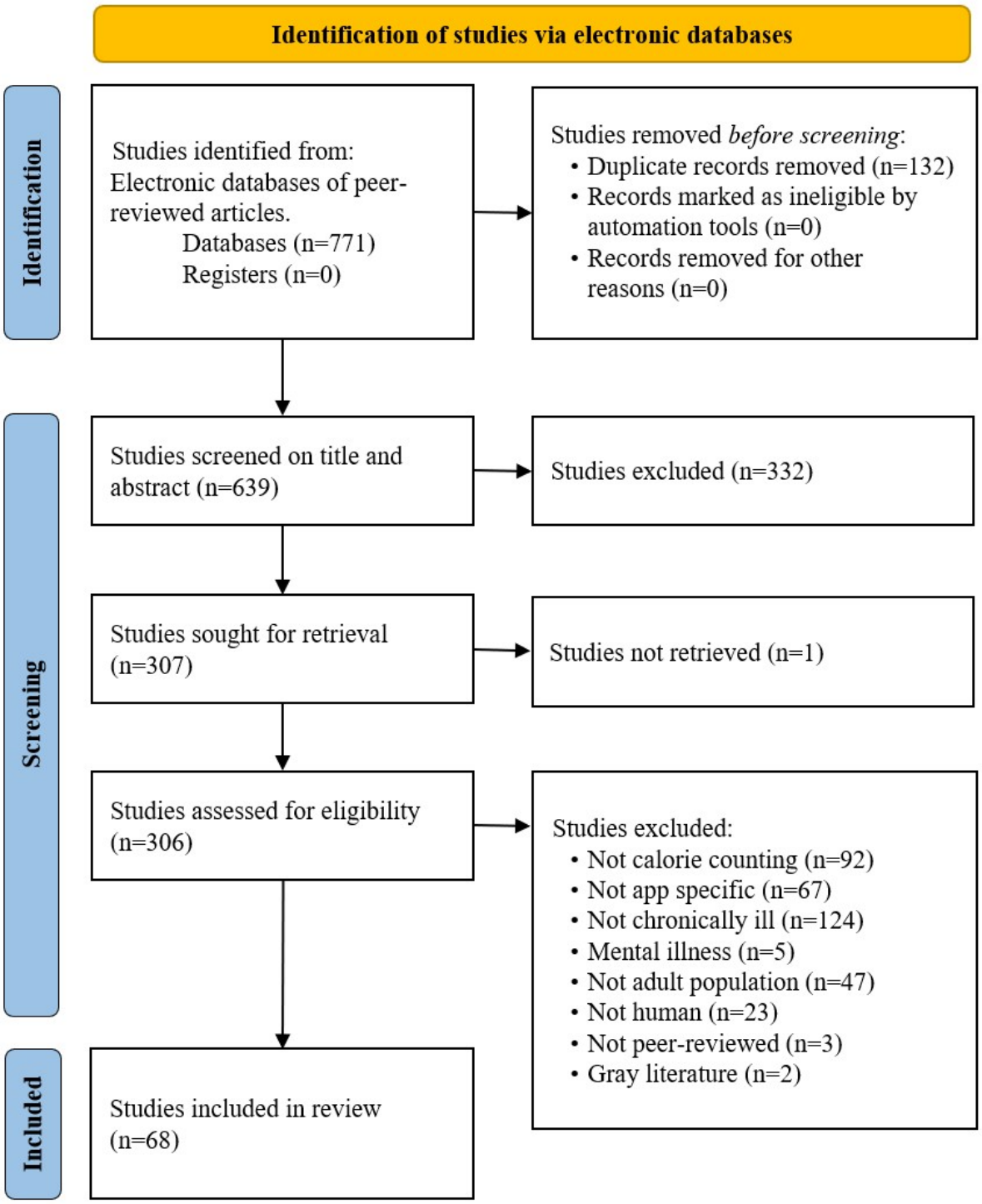
conceptual category. The final conceptual categorization first included features and functionalities of logging in, calorie-counting, goal setting, app interactions with users, other key features, and validity. It then included factors contributing to acceptability, feasibility, and adherence (see full list in results).

Results

Description of Included Papers

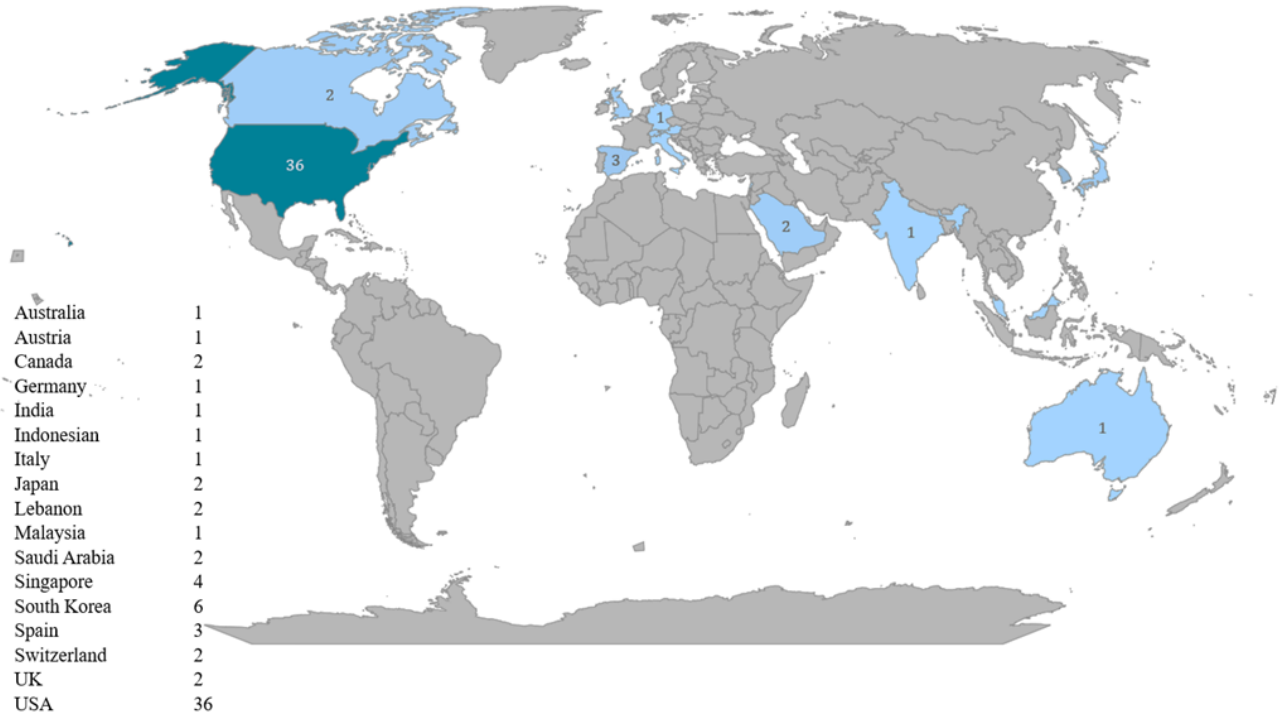
Of the 68 retrieved papers [13-17,22-85], (Figure 1), most studies were conducted in the United States (36/68, 52.9%), followed by South Korea (6/68, 8.8%) and Singapore (4/68, 5.9%; Figure 2).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. Note that excluded papers can fall under multiple exclusion criteria.



XSL•FO
RenderX

Figure 2. Distribution of included articles by geographic location.



Randomized controlled trial (23/68, 34%) and cohort study (16/68, 24%) represented the main research designs used, followed by cross-sectional study (9/68, 13%), mixed methods design (5/68, 7%), qualitative study (5/68, 7%), and quasi-experimental study (10/68, 15%; Table 2).

Table 2. Distribution of included papers by study design.

Study design	Papers, n (%)
Randomized controlled trial	23 (34)
Cohort study	16 (24)
Cross-sectional study	9 (13)
Mixed method design (qualitative-quantitative)	5 (7)
Qualitative study	5 (7)
Quasi-experimental study	10 (15)

Among the 68 retrieved papers [13-17,22-85], 53 (78%) studies included participants living with obesity [13-17,22-31,36,37,40,42-44,49-53]. Participants with prediabetes or diabetes (type I or II) and prehypertension or hypertension were included in 18 of 68 (26.5%) [27-30,36,37,40,42-44,50,51,53,55,58,59,75,78] and 10 of 68 (14.7%) papers [28,30,40,43,52,64-66,69,81], respectively. Multimorbidity, or the presence of more than one weight-related chronic disease within a study population, was identified in 18 of 68 (26.5%) papers [22,27-30,36,37,40,42-44,50-53,59,69,81]. For the other weight-related chronic diseases included in this scoping review, cardiovascular disease was only found in 1 (1.5%) paper [81], and metabolic syndrome was not identified in any paper.

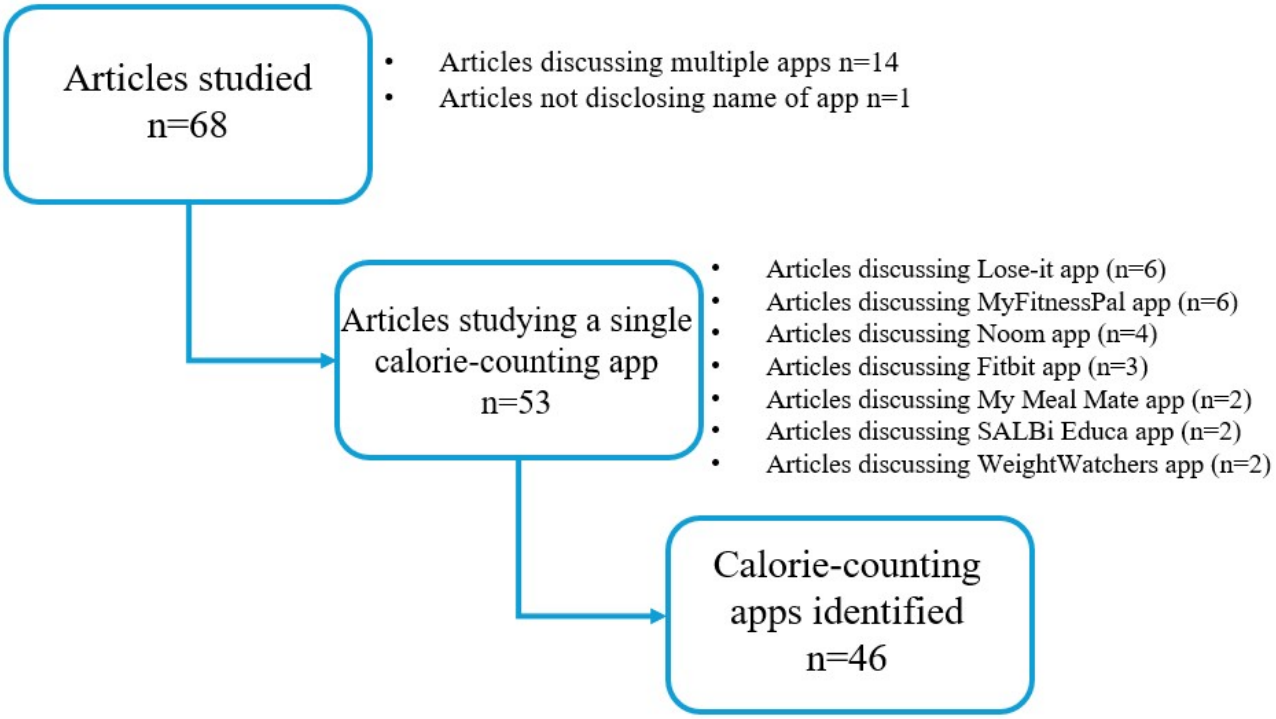
Out of the papers reviewed, only 19 of 68 (27.9%) provided detailed or brief descriptions of the theoretical or conceptual foundations of calorie-counting apps [22,26,29,32-36,39,41,45-48,50,54-56,72]. Among these, the social cognitive theory appeared in 8 of 68 (11.8%) studies [22,27-29,37,43,72,82]. Self-management education and chronic disease self-management frameworks were each used in 4 of 68 (5.9%) studies [25,27,51,82]. Additionally, the self-regulation theory and behavioral theory were each referenced in 3 of 68 (4.4%) papers. The goal setting theory was mentioned in 2 of 68 (2.9%) papers [68,69], while the information–motivation–behavioral skills model and behavioral economics theory were each used in a single paper.

Among the 68 papers identified through our search, 14 (20.6%) examined more than one calorie-counting app [23,24,26,28,49,50,60,65,66,73,77,78,80,85]. Within this subset, 7 papers did not provide a detailed evaluation or noted commonalities and differences among the apps [23,24,26,50,65,73,80], while the other 7 conducted detailed comparisons and evaluations [28,49,60,66,77,78,85]. One paper (1.5%) did not disclose the name of the app studied [69]. The remaining 53 papers (78%) focused on evaluating or discussing a single app [13-17,22,25,27,29-48,51-59,61-64,67-72,74-76,79,81-84]. In sum of all papers, MyFitnessPal was the most frequently studied app (20/68, 29.4%)

XSL•FO
RenderX

[15,16,23,24,28,46,48,49,52,54,57,60,63,67,76,79-81,84,85], followed by Lose It! (14/68, 20.6%) [15,16,23,24,31,44,49,54,61,62,68,71,75,80], the Fitbit app (9/68, 13.2%) [16,31,37,40,67,75,77,80,84], and Noom (9/68, 13.2%) [16,28,31,52,67-70,73]. The process used to extract individual apps is illustrated in Figure 3, while Table 3 presents the 46 apps along with their available establishment year, country, and a brief description.

Figure 3. Flowchart illustrating process of identifying calorie counting apps through selected articles.



XSL•FO
RenderX

Table 3. Description of the 46 single apps included in the retrieved studies.

App name: study	Established year	Country	Short description
Akser Waznk; Alturki and Gay [22]	__[a]	Saudi Arabia	The app was developed in Saudi Arabia and assists users with obesity in adopting a healthy lifestyle through motivation and dietary tracking.
Argus; Ferrara et al [23]	2012	United States	The app tracks activity, dietary intake, and sleep to support weight management.
Bitesnap; Gioia et al [24]	2017	United States	The app helps users track food intake and manage their diet through a visual approach.
Cronometer; Gioia et al [24]	2005	Canada	The app provides detailed nutrition tracking and calorie-counting to support weight management.
DialBetics Lite; Kondo et al [25]	—	Japan	The app records diet, physical exercise, and diabetes self-management data.
DietLens; Tahir and Loo [26]	2018	Singapore	The app uses AI[b] to identify foods and estimate nutritional composition from images.
Doctor Diary; Kim et al [27]	—	India	The app tracks calories and diabetes-related data and enables monitoring by health care professionals.
Easy Diet Diary; Chen and Allman-Farinelli [28]	2012	Australia	The Australian app provides calorie-counting and dietary tracking with physical activity monitoring.
EatsUp; Agustina et al [29]	—	Indonesian	The Indonesian app records dietary data and recommends menus based on caloric targets.
ENGAGED; Pellegrini et al [30]	2012	United States	The app was developed as part of the ENGAGED trial, allowing for dietary self-monitoring and calorie counting, as well as physical activity tracking.
FatSecret; Ferrera et al [23]	2007	Australia	A free food diary app that allows users to log the amounts and types of foods and beverages consumed, providing nutritional information, including calorie estimates.
FDDB extender app [86]	—	Germany	The app assists users in tracking calories for weight reduction or maintenance.
Fitbit; Burke et al [31]	2007	United States	The app tracks calories, diet, and other health-related metrics.
Food Tracker; Tahir and Loo [26]	2019	Canada	The app logs food intake and monitors nutritional consumption for health management.
FoodCam; Raju and Sazonov [32]	2015	United Kingdom	The app uses image recognition to identify foods and provide nutritional information.
FoodLog; Lemacks et al and Aizawa et al [33,34]	2013	Japan	The app tracks meals and monitors nutritional intake.
Fooducate; Chaudhry [35]	2010	Israel	The app helps users make healthier food choices by providing nutritional insights.
GoCarb; Anthimopoulos et al [36]	2016	Switzerland	The app estimates the carbohydrate content of meals from food images for diabetes management.
Good Measures; Olfert et al [37]	—	United States	The app is designed to support healthy food choices and behavior change related to eating and exercise via food entry and nutritional balance evaluation.
iDAT [38]	—	Singapore	The app is designed for the Singapore population. Functions as a calorie counter, helping users to balance calories consumed with calories burned daily.
Im2Calories; Myers et al [39]	2015	Japan	The app uses deep learning to estimate calorie content from food photographs.
January AI; Zahedani et al [40]	2023	United States	The app integrates CGM[c] and HR[d] data with user-entered diet and activity data.
Keenoa; Moyen et al [41]	2020	Canada	The app uses image-assisted food diaries with AI-based food recognition.
KIT-Nutrition app; Schusterbauer et al [42]	—	United States	The app combines dietary self-monitoring with diabetes management features.

XSL•FO
RenderX

App name: study	Established year	Country	Short description
LIBIT; Oh et al [43]	—	Korea	The app records diet and exercise, connects to home devices, and provides feedback.
Lifesum; Ferrara et al [23]	2013	Switzerland	The app combines calorie counting with healthy lifestyle recommendations.
Lose It!; Burke et al [44]	2008	United States	The app allows users to record daily food intake and physical activity for weight loss.
Menu-Match; Beijbom et al [45]	2015	United States	The app helps users with dietary restrictions find suitable restaurant meals.
Mobile food record (mFR); Ahmad et al [46]	2021	United Kingdom	The app captures before-and-after meal photos to assess dietary intake and patterns.
MyDietCam, Tahir and Loo [26] and Chui et al [47]	2020	Malaysia	The app analyzes meal photos to provide nutritional information and personalized diet advice.
MyFitnessPal; Tosi et al [48]	2005	United States	The app tracks foods and beverages to provide nutritional information and calorie analysis.
MyMacros+; Gioia et al [24]	2014	United States	The app tracks macronutrients for weight management.
MyMealMate; Carter et al [49]	2012	Australia	The app facilitates weight loss with an electronic food diary for dietary self-monitoring.
MyNetDiary; Fu et al [50]	2007	United States	The app provides personalized weight loss advice and tracks food intake.
MyPlate by Livestrong; Ferrara et al [23]	2004	United States	The app tracks calories and nutrients and supports diabetic users.
nBuddy Diabetes (Nutrition Buddy Diabetes); Lim et al [51]	2017	Singapore	The app supports diabetes self-management through meal logging and activity tracking.
Noom; Jin et al [52]	2008	United States	The app combines calorie tracking with behavioral coaching for weight management.
Nutritionix; Kay et al [53]	—	United States	The app enables self-monitoring of dietary intake for daily nutrition tracking.
PlateMate; Zhou et al [54]	2011	United States	The app estimates the nutritional content of foods from photos using crowdsourcing.
SAlBi (Salud, Alimentación, Bienestar, and educación); Gonzalez-Ramirez et al [55]	2022	Spain	The app provides self-monitoring and tailored dietary advice based on the Mediterranean diet.
Snap-n-Eat; Zhang et al [56]	2015	United States	The app automatically recognizes food and estimates calorie and nutrient content.
SparkPeople; Bardus et al [57]	2001	United States	The app tracks diet, exercise, and weight changes over time.
T1DEXI; Riddell et al [58]	2020	United States	The app tracks food intake and exercise for diabetes management.
Unnamed prototype app	—	United States	The prototype app provides real-time feedback on diet quality and heart disease risk.
WeightWatchers (WW); O'Neil [59]	1963	United States	The app uses a point system to promote healthy eating and weight loss.
Yazio; Puigdomènech et al [60]	2013	Germany	The app assists with calorie counting and diet planning for weight loss.

[a]Not available.

[b]AI: artificial intelligence.

[c]CGM: continuous glucose monitoring.

[d]HR: heart rate.

Synthesis of Results

The data extracted from the 68 papers were categorized into four dimensions: (1) features and functionalities of calorie-counting apps, (2) acceptability, (3) feasibility, and (4) adherence. These dimensions are presented below in more detail.

XSL•FO
RenderX

Features and Functionalities

Logging-In

When beginning to use an app, it may ask the user to provide personal information, such as an email, to allow the user to sign up and create an account to use the app. This allows for information on the mobile device to be saved as the user uses the app, linking the user's information to their account. The review of papers underlined two methods for the storage of information on an app: the creation of a new account on the app's software itself or signing in by linking to an already existing social media account. Of the 46 total apps discussed, 16 (35%) were said to have a login feature. One additional paper discussing this feature was asked not to specify the app portrayed. Doctor Diary [27], Akser Waznk [22], and Lose It! [61] were apps that were mentioned to use email addresses as a means of signing up to the platform. Akser Waznk was also described as offering the option of linking to a Facebook, Instagram, or X (formerly known as Twitter) account for users to sign up to the app [22].

Additionally, papers also described the process of entering personal and health information in order to accurately customize goals and recommendations. Akser Waznk [22], Lose It! [62], and Noom [52] were reported to ask users to enter their gender, age, height, and weight at the time of sign-up. Additionally, Akser Waznk had an optional step of adding health information, such as the user's medications and chronic diseases [22]. Doctor Diary [27], MyFitnessPal [63], and FDDB extender [64] had mentions of a login feature requiring user information, but no further description was provided. The 18 other apps had no mention of the logging-in process, according to this study, although it is possible that this feature is present in these apps and has yet to be described.

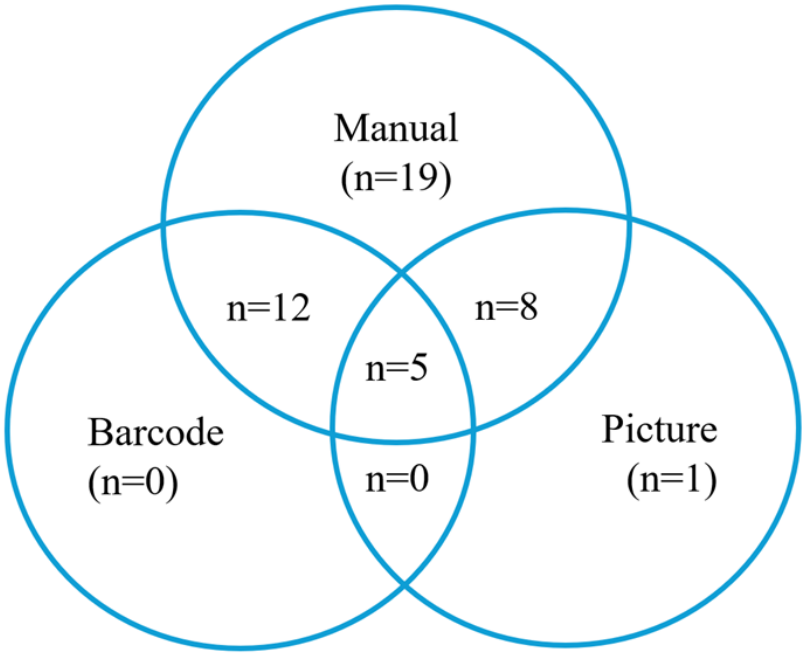
Calorie-Counting Features

When determining the acceptability of an app, the way that the process of calorie counting is done is crucial [65]. On that same note, an app that allowed for fast analysis of calorie content (meaning that the user needed to spend as little time as possible to log their calorie intake, as well as the ease of use of the said feature) was generally viewed as favorable [66]. Among the 68 papers, 17 (25%) included more than one app or did not mention the availability of calorie-counting functionality. The remaining 51 papers described 3 main methods of calorie counting: manual entry, picture-based entry, and barcode scanning.

Manual entry was generally described as the ability of the user to enter food products to get an estimation of total calories (ie, diary data entry). A total of 44 out of the 46 (95.6%) apps used calorie manual entry. Although the general concept of manual entry is well described, there is one significant difference when it comes to the functionality of calorie estimation: the integration of a food database. These databases, which could include up to 6 million foods within the app [67], provided a base of caloric information to be retrieved when users logged their consumption and removed the need for users to enter the calorie counting themselves. The databases could also be tailored to match certain population demographics (17/46, 37% apps had country-specific food databases). Once a food item was found, the user quantified the amount consumed. Of the 44 manual entry method apps, 26 (56.5%) had the use of a food database mentioned within their respective paper. Finally, it is worth noting that LIBIT had a voice recognition feature for logging meals [43].

Picture-based entry was described as the process of taking a picture of a food before consumption to get a calorie count estimation by the app. This was usually done through an integrated software that recognizes food items [62]. A total of 14 (30.4%) of the 46 apps mentioned this method of calorie counting (Figure 4).

Figure 4. Venn diagram shows overlap of calorie-counting methods used by apps.



XSL•FO
RenderX

The last method of calorie counting was through barcode scanning. This method relied on scanning barcodes of store-bought food items, which allowed the app to identify food items, and the user then logged the quantity consumed to get a calorie intake estimate [64]. A total of 17 (36.9%) of the 46 apps were mentioned to have this feature.

It is important to mention that certain apps include multiple methods for calorie counting. A total of 17 apps had both manual entry and barcode entry, while 5 apps had a combination of all 3 methods. These relationships are portrayed in Figure 4.

Goal Setting

Another feature that was often described within the calorie-counting apps designed for individuals living with chronic health conditions was the ability for users to set goals within the app. Indeed, among the 46 individual apps discussed in papers, 24 (52.2%) had a form of goal setting. While the specific type of goal could vary in composition, the most commonly described is one that calculates a calorie budget or allotment via the input of a weight loss goal determined by the user, as well as the input of certain characteristics such as current weight, height, and physical activity level (usually done when going through the account setup). One method by which this caloric budget calculation was made was through the Harris-Benedict equation [29]. Goal setting in the form of a weight loss target and a caloric budget was clearly described by Wharton et al [68] and was found to be the most common type of goal setting since it is used by 17 of the 24 apps (70.8%) that have a goal setting functionality.

Of the remaining 7 apps, 6 had goal settings mentioned without specifying the way it was integrated, and 1 was mentioned to have goal setting in the form of nutritional balance, which consisted of encouraging users to preset a goal of servings per food category, such as fruits and vegetables [69].

App Interaction With User

Some mobile apps could interact with the user to encourage use of the app and its functionalities (ie, tracking). These interactions could be done while the user was actively using the app (synchronous) or when the app was not in current use (asynchronous), such as providing reminders to the user. Of the 46 total apps, 30 (65.2%) mentioned in the papers interacting with the user by using reminders or by providing intake recommendations and feedback. Notably, MyFitnessPal [67], Akser Waznk [22], Noom [70], EatsUp [29], nBuddy Diabetes [51], Lose It! [44], January AI [40], Keenoa [41], WeightWatchers (WW) [59], FatSecret, Cronometer [24], Yazio [60], MyPlate by Livestrong, Lifesum, SparkPeople, Argus, and MyMealMate [49] used reminder messages to log intake or to increase adherence to recommendations. Eisenhauer et al [71] found that text messages, such as Lose It!, would improve self-monitoring in users at 3 and 6 months. While the presence of reminder functionalities was extracted, the frequency or timing of these reminders was not consistently reported across studies and, therefore, was not included in the data synthesis.

Interactions also consisted of providing feedback and health information to the users. Doctor Diary [27], MyFitnessPal [67], DialBetics Lite [25], SAlBi educa [55,72], Nutritionix [53], January AI [40], LIBIT [43], EatsUp [29], MyMealMate [49], and Good Measures [37] provided feedback on nutrition. This consisted of proposition suggestions on energy intake and food recommendations. Gonzalez-Ramirez et al [55] stated that users found the tailored feedback messages to be one of the most useful features used in the SAlBi educa app and that 86.1% of users believed this feedback would have a positive effect on user diet, while Kay et al [53] stated that only 55% of users believed this feedback was useful in the Nutritionix app. Nutrition feedback was also conducted in different manners other than live in-app messages, such as providing health-related papers in Noom [73] or educational videos on nBuddy Diabetes [74]. Some apps also displayed the logging of calories in visual feedback for the users, which allowed a more comprehensive way of showcasing their data. Easy Diet Diary [28], MyFitnessPal [67], Lose It! [75], Fitbit [31], ENGAGED [30], SAlBi educa [55], and MyMealMate [49] were apps that mentioned presenting the user's intake data in graphs or charts, according to the select papers.

Other Functionalities

Most of the apps studied in the selected papers had multiple functions other than calorie counting, for example, weight or blood glucose monitoring. Other key features described in the literature were related to the ways users can log health information. These key features included remote patient monitoring by health care providers, graphical or visual data, and sharing progress with other users. Some apps included information on recipes and color schemes, as well as the possibility to upgrade to a premium version (with more features and functionalities). Out of the 46 total apps, 34 (74%) were mentioned to have some of these features. Doctor Diary [27], EatsUp [29], LIBIT [43], Good Measures [37], MyFitnessPal [76], Lose It! [71], SAlBi educa [55], January AI [40], MyMacros+, Lifesum, SparkPeople, Argus, and WeightWatchers (WW) [13] had features other than calorie counting for users to track, such as weight tracking or glucose monitoring, according to the literature retrieved.

Sharing the user's progress with health care providers could be done in real-time tracking or later, providing insight into the user's tracking habits. A total of 15 (32.6%) out of the 46 total apps included this feature. Doctor Diary [27], Noom [70], EatsUp [29], Easy Diet Diary [28], nBuddy Diabetes [51], FDDB extender [64], LIBIT [43], Nutrinaut [42], Good Measures [37], Cronometer, MyNetDiary, WeightWatchers, Lifesum, MyPlate by Livestrong, and SparkPeople were described as apps that allowed health care providers to access the user's tracking data. Where specified, health care providers involved were physicians and registered dieticians.

An additional key feature discussed for 15 (32.6%) of the 46 apps was the ability for users to have real-time chats with other users or to share their progress on the app for others to see. Lose It! offered discussion boards that would have benefited the users more if they were done in a synchronous manner, which would have increased the feeling of support, according to participants [71]. Akser Wankz and WeightWatchers also provided features for users to chat with others about their progress [22]. Other apps like iDAT, Yazio, Lifesum, SparkPeople, Argus,

XSL•FO
RenderX

WeightWatchers, and Fooducate allowed users to share their progress directly to a social media platform, Facebook [78]. MyFitnessPal's networking feature was assessed, and 80% of participants reported having "no friends" on the app, which highlighted the minimal use of this feature [79].

Color schemes were also discussed in select papers, providing an insight into the importance of aesthetics in applications. Out of the 46 total apps, 13 (28.3%) had discussions on the aesthetics used. Both EatsUp [29] and Akser Waznk [22] were discussed in terms of their use of color schemes. The use of colors allowed for an increase in appeal and encouraged the use of apps. According to Alshathri et al [80], MyFitnessPal, Fitbit, and Lose It! were also some of the highest-ranked apps in engagement and aesthetics.

Some apps also had the option of premium access, allowing the user to pay for more advanced features not included in the free version of the app, such as Noom [73], MyFitnessPal [81], and Lose It! [62]. In general, 13 (28.3%) of the 24 apps mentioned have a premium option, which offers exclusive functions with enhanced support and patient monitoring. Although premium features were available, most participants expressed a preference for free apps, with cost emerging as one of the primary factors influencing their app selection [64].

Overall, free, of charge, and self-explanatory apps were most preferred by users. Sharing their progress with others or through a platform were features deemed to be less important by users [66].

Validity, Accuracy, and Reliability

Calorie-counting apps that were available to the public used different methods to log calories and could provide recommendations to users for healthier options. For reliability, these apps could rely on evidence-based theories, use known food databases, or be developed by a team of health care professionals. Of the total 46 apps, 22 (47.8%) were mentioned in papers to use at least one of these methods. EatsUp [29], Nutritionix [53], Kit-Nutrition [42], and SAlBi educa [55] were apps mentioned in the retrieved literature to have been developed or maintained by health care professionals, such as dieticians or physicians.

As discussed previously, food databases were sometimes used by the apps to correctly estimate calorie counting for the user. This thereby provided reliable, accurate information to the users. Easy Diet Diary was described to use the Australian Food and Nutrient Database 2011-2012 [28] while FDDB extender used the FoodData Central of the US Department of Agriculture and the Souci-Fachmann-Kraut databases [64]. Additionally, SAlBi educa used the BEDCA database for reference during food logging [55]. January AI was also said to use known databases, but this was not further detailed in the chosen literature [40]. On the other hand, Nutritionix itself was highlighted as the largest verified database for nutrition information by [53].

In terms of evidence-based theories, these apps were mainly developed to manage chronic diseases, such as obesity or diabetes. They could thereby rely on chronic disease management or behavioral theories to develop an effective calorie-counting app, like the social cognitive theory. Akser Waznk [22], EatsUp [29], MyMealMate [82], SAlBi educa [72], and Good Measures [37] were apps that were described to have relied on evidence-based theories on nutrition or chronic disease management. Lee et al [73] rated Noom the highest score of information accuracy, but the paper did not provide additional information on the theories used to develop the app. Keenoa was evaluated to have moderate to strong relative validity compared to an Automated Self-Administered Dietary Assessment Tool, but more validity establishment was needed [41]. Table 4 summarizes the main identified features and functionalities, along with the number of papers where they were presented and a short description.

XSL•FO
RenderX

Table 4. Description of the identified features and functionalities.

Feature	Apps with the feature mentioned (N=46), n (%)	Description
Logging in	• 16 (35)	Logging in to the app allows users to save information. The sign-up process may require personal or health information.
Calorie counting	• 45 (98)	There are 3 general ways by which apps count calories: manual entry, barcode scanning, and picture-based entry. Almost all apps have some form of manual entry, which is facilitated by the integration of a food database in most cases. A total of 5 apps offer a combination of methods.
Goal setting	• 24 (52)	Goal setting is generally available in the form of a weight loss goal, which allows the app to calculate a caloric budget. One app offers the setting of nutritional balance goals.
App interaction with the user	• 30 (65)	Interactions consisted of reminders to increase app use or the provision of feedback and recommendations on nutrition.
Other	• Other functions of the app: 34 (74) • Monitoring by health care providers • 15 (33) • Premium subscription: 13 (28) • Data sharing: 13 (28) • Visual aesthetics: 13 (28)	Most apps have a combination of calorie counting and other functionalities, including information exchange with health care professionals or other patients, data visualization, and an upgrade to a premium version.
Validity, accuracy, and reliability	• 22 (48)	To be reliable, some apps were developed according to evidence-based theories, such as behavioral theories. Other ways of ensuring reliability include health care professionals in the development of apps or using known food databases for accurate nutritional information.

Acceptability

From the scoping review database, we extracted the factors that favor or limit the acceptability of calorie-counting apps from the perspective of users and of health care professionals. When identifying factors contributing to acceptability, we used the following definition as provided by Garizábalo-Dávila et al [83]: “Acceptability is the perception of patients and health professionals versus determining whether the intervention is appropriate to address the problem, in a reasonable, adequate, and convenient manner for its application in daily life.” We summarized below the key findings according to the various structural, cultural, socioeconomic, behavioral, quality standards, and ethical factors identified.

Structural Factors

Various characteristics of the apps’ format, features, functionalities, and content influenced their acceptability from the users’ perspective (Textbox 2). The scoping review showed that features or functionalities personalized according to users’ preferences and characteristics increased the acceptability of calorie-counting apps. For instance, users were very satisfied with the possibility to choose color schema and app themes through a variety of options [22]. Users also liked the feature of allowing them to highlight the nutrients of most importance to them in the app when tracking their dietary intake [28]. The personalization of reminders or feedback generated by the app improved acceptability. Receiving app recommendations based on previous successes or failures to meet dietary goals [65] or getting a personalized progress report each week [84] are examples of functionalities that increased users’ satisfaction.

XSL•FO
RenderX

Textbox 2. Facilitators of acceptability of calorie-counting apps.

Format (structural)

- Esthetically pleasing design
- Streamlined interface
- Ease of carrying a smartphone
- Paperless food record

Features (structural)

- Personalized features
- Self-monitoring features
- User-friendly screen features
- Food image
- Technical support from the app provider (data derived from both users and health care professionals)

Functionalities (structural)

- Augmented reality
- Automatic calorie or nutrient estimation
- Automatic food recording
- Integration with other apps or platforms
- Sharing information with a health care professional, family, or on social networks
- Ability to monitor goals
- Useful and personalized reminders or feedback

Content (structural)

- Availability of varied food items
- Appropriate and clear terminology
- Appropriate translation
- No need for training
- Availability of support or training

Cultural

- Health awareness (data reported exclusively by health care professionals)

Socioeconomic

- e-literacy
- App free of charge
- Gender
- Rurality

Behavioral

- Sense of accountability
- Motivation
- Performance expectancy
- Habit
- Anxiety risk

Quality standards

- Validation

XSL•FO
RenderX

- Certification

In Textbox 2, all items represent data reported by participants, unless indicated.

Automated functionalities and the possibility to share app data through well-integrated platforms increased app usability or usefulness and increased users' acceptability (Textbox 2). For instance, users enjoyed an app including automatic calorie or nutrient estimation [66] and automatic food recording through barcode readers, meal image analysis, or artificial intelligence-based food detection [65,66,69]. On the contrary, manual entry of food consumption represented a major deterrent to using a calorie-counting app [85]. Users were satisfied with the possibility to share data recorded on the app with health care professionals, family members, or on social media. Because saving and sharing dietary record history was not available on all calorie-counting apps, the absence of this functionality was also a factor hampering acceptability. Users also noted that the functionalities of some apps were limited by bugs, crashes, and insufficient updates, and these factors all reduced acceptability.

The self-monitoring features were considered by users to be useful components of an app. Being able to daily track food intake [84], to easily log dietary intake with copy-paste functionalities [28], or to easily input beverage count [29] was appreciated. Users were also motivated by the ability to monitor goals or to fill out an action plan weekly [84]. Being able to visualize a weight chart showing the initial, current, and goal weight on a progress screen on some apps represented a great source of satisfaction for users [22]. These self-monitoring features, being available as paperless food records on an easy-to-carry device such as a smartphone, were also noted by users as a strength in terms of usability [29].

In terms of interface and content, user-friendliness was highlighted as an important characteristic of a calorie-counting app. Users' satisfaction increased with the presence of self-explanatory features that did not require training [66] or when users could easily find all subscreens on the main screen [22]. With respect to recording food consumption, the availability of a variety of food items represented a clear strength for an app [37]. This was especially true when app designers chose to use food images over text as an option to select dietary items [29]. On the contrary, the acceptability of an app decreased when users were unable to find specific food items on the app, either because they were unavailable or difficult to find. More specifically, many papers highlighted that an app's usefulness decreased when lacking multiethnic, local, and home-cooked foods as well as take-out meals and restaurant options [28,37,65,79,81]. In addition, confusion on the app with respect to portion size was another barrier to acceptability. Finally, users noted the importance of clear terminology [72] and appropriate translation [81] in calorie-counting apps. Nutritional messages that were easily understood by users with a lower level of education were also identified as a positive factor of acceptability [72]. Finally, users' satisfaction was higher when the app design was aesthetically pleasing, for instance, when it used attractive color schemes, an appropriate font size, and a streamlined interface [79].

Other Factors

Among cultural factors, health awareness was recognized by both users [37,74,78,81] and health care professionals [22] as a factor that increased the acceptability of an app. In parallel, some users identified that they were not comfortable sharing their electronic dietary record with a health care professional, and this limited the usefulness of the app [28].

The scoping review identified many socioeconomic factors contributing to the acceptability of calorie-counting apps. The most frequently cited factor was e-literacy. In fact, some apps were considered user-friendly, but only for users familiar with smartphone apps [28]. One study identified that this could be especially true for older patients [75]. The influence of the level of e-literacy on the acceptability of calorie-counting apps was not restricted to users, as one study reported that health care professionals with low levels of e-literacy could also lack the skills to efficiently use or counsel patients using calorie-counting apps [28]. Financial aspects were also addressed as factors influencing acceptability. Free apps had a higher level of acceptability [66], but not being able to afford a smartphone was a major barrier to app usability [37]. In addition, the usefulness of certain recommendations generated by apps varied with the type of work schedules. For instance, shift workers reported that, depending on their shift schedule during a given week, they were not always able to meet fixed weekly goals included in recommendations automatically generated by the app [65]. Sex affected adherence to app use, as females were more consistent users compared to males [78]. Finally, participants living in rural areas compared to those in urban areas shared more favorable views of calorie-counting apps [37].

Behavioral characteristics of users could affect the acceptability of calorie-counting apps. For users who mentioned the importance of being kept accountable, the usefulness of using a calorie-counting app increased when combined with the oversight from a dietician [28]. In 2024, Chew et al [74] showed that the intention to use AI-assisted weight management apps is influenced by factors such as age, anxiety risk, and the desire to maintain a healthy diet. These factors together explain a significant portion of why people might choose to use these apps. However, self-regulation, depression risk, BMI, and waist circumference do not seem to affect the decision to use AI-assisted weight management apps.

In terms of quality standards, validation and certification of an app represented criteria for selecting a nutrition and diet app by users [66]. Although automated functionalities in an app increased its acceptability level as addressed above, the lack of accuracy of automated functionalities led to user frustration [69]. Finally, for dieticians, a key barrier to using calorie-counting apps was the lack of confidence associated with the reliability of sharing patients' personal information [28].

In Textbox 3, all items represent data reported by participants, unless indicated.

XSL•FO
RenderX

Textbox 3. Barriers to the acceptability of calorie-counting apps.

Functionalities (structural)

- Saving and sharing history and favorite not supported
- Bugs, crashes, or server connectivity
- Insufficient app updates
- Manual entry of food consumption
- Impossibility to edit and delete entered values
- Reliability of data sharing unknown (data reported exclusively by health care professionals)

Content (structural)

- Demotivating reminders
- Unavailability or difficulty in finding food items
- Confusion with portion sizes

Cultural

- Discomfort with sharing app data with a health care professional

Socioeconomic

- Internet unavailability
- Low level of e-literacy (data derived from both users and health care professionals)
- Older age
- Smartphone availability
- Work schedule

Quality standards

- Automated techniques not accurate
- App not validated (data derived from both users and health care professionals)
- App not certified (data derived from both users and health care professionals)
- Confidentiality of data sharing is unknown (data reported exclusively by health care professionals)

Feasibility

Barriers that affected the feasibility of the apps in a clinical context were classified into 5 categories: structural, cultural, socioeconomic, behavioral, and accuracy (Textbox 4).

XSL•FO
RenderX

Textbox 4. Barriers to feasibility from the patient (or health care professional) perspective.

Structural

- App crashes
- Server connectivity
- Technical problems
- Calorie logging difficulties
- Limited smartphone platform
- No barcode scanning option
- Time consuming

Cultural

- Limited food database

Socioeconomic

- Web version of the mobile app
- No compatible smartphone

Behavioral

- Food recommendation
- Portion estimation by the user

Accuracy

- App underestimating calories
- App overestimating calories

Structural

It was not unusual for the users to experience app crashes, technical problems, and difficulty with server or internet connectivity [42,79,85].

Some apps did not contain the barcode scanning feature or needed the internet for the entry of food items, which then caused loss of motivation [22]. Logging food and beverage intake in an app could be effortful and time-consuming, taking an average of 15 to 20 minutes per day, and behavioral fatigue sometimes resulted [42,87]. On a similar note, one study found that dieticians thought the Easy Diet Diary was time-consuming to monitor and made it difficult to review patient records in a timely manner. The issues created a loss in motivation to keep using the app [28].

Socioeconomic

Some participants did not have a smartphone, or the smartphone was not compatible with the app [22,42]. Lose It! had a smartphone and web-based version, although the web version could be limiting since it did not contain the same features. For example, the photo feature of the app was not available through the web version [62].

Behavioral and Cultural

Users did not always accurately estimate food portions and nutritional contents [69]. Food databases were available to help users, but the databases could be quite large for the calorie-counting apps, and could be missing certain food products or recipes, especially when it came to more regional food items and homemade meals [41,42,64,72]. Capacity and accuracy of the search engine and the variety of meals and food terminology are key aspects to consider when food databases are created.

Accuracy

In a prospective controlled trial that assessed the quality and effectiveness of popular calorie-counting apps in weight management and behavior change, calorie and activity recommendations were compared with standards, and over 65% of apps over- or underestimated calorie intake [85]. Recent evidence further emphasizes that incorporating structured behavior change strategies within mobile app interventions can significantly enhance their effectiveness in improving weight-related outcomes [88].

Adherence

All calorie-counting apps saw a decline in adherence over time, although both the definition and level of adherence differed from one study to another. Definitions ranged from logging total daily dietary intake to simply using the app at least once a day, with some studies requiring a minimum number of calories to be recorded. Despite varying definitions of adherence, studies consistently showed that adherence declined over time, with several reporting a sharp drop, in some cases to nearly zero, after the intervention period. Adherence levels were typically measured by collecting the app data, which tracked how frequently and how much information participants entered.

XSL•FO
RenderX

However, recent findings suggest that strong early app engagement may predict better long-term adherence and greater weight loss outcomes [89].

Notably, some apps had good adherence in the first few months before decreasing. One study reported that participants self-monitored with MyFitnessPal an average of 5.4 days/week during the first 4 weeks, declining to 1.4 days/week during weeks 5-12, and to 0 days/week after the 12-week intervention [67]. Another study in Doctor Diary found that 14/32 participants (43.8%) failed to record three daily meals throughout the 8-week period, meaning only 18/32 participants (56.2%) were considered adherent based on the study's criteria [21].

Comparingly, a study on "Lose it!" found that in the first 6 weeks, participants used the app daily but decreased to 4 days per week at the end of the 12-week study [75]. Multiple studies showed that Lose It! had high adherence compared to a food journal group or an electronic diary group [31,61,68]. In one study that evaluated the adherence of MyMealMate, calorie counting declined over time so that by 6 months, 7 of the participants (16%) in the smartphone group had managed to record their dietary intake every day, but no participants in the food journal and web group had achieved this [49]. The following section outlines 4 factors found to enhance adherence, also summarized in Textbox 5.

Textbox 5. Factors that facilitate adherence are divided into 4 categories: structural, user-related, financial, and environmental.

Structural

- Photo feature
- Paired with a structured program
- Paired with a Facebook group
- Simplified monitoring

User-related

- Health awareness
- High motivation by the participant
- Higher education

Financial

- Monetary incentive
- Access to the premium version

Environmental

- Internet access

Socioeconomic Factors

In a study comparing premium versions of apps to free versions, they found that premium versions had a higher adherence rate, defined as logging of 800 daily calories or more [71]. Although another study with MyFitnessPal Premium also confirmed that the adherence to logging all meals 4 or more days a week still decreased over time, from 39% to 8% to 0% at week 1, 12, and 24, respectively [81].

One study examined the effectiveness of a monetary incentive with a deposit system with the goal of motivating participants using Noom. With this in place, more than half of the participants completed the 16-week program by logging three or more times a day [73].

User-Related Factors

Adherence to daily app use was found to be associated with good comprehension of the impact associated with their health condition and the desire for weight loss [85]. Similarly, low motivation for habit change was associated with low daily adherence [78,85].

In terms of specific user demographics, a study on Good Measures noted that adherence to daily meal-logging in the app did not vary between peri-urban and rural sites [37]. In another study, participants with a higher level of education had a higher general adherence to the app [53].

Structural Factors

Some features were found to have increased participants' adherence to the app. For example, a retrospective cohort study evaluating Lose It! showed that the group who used the photo feature by taking a snapshot of their food items and selecting the right items from a list of suggestions to display calories had 6.1 more total logged days than those who did not use this feature [62]. This remained significant after adjusting for user demographics such as age, gender, and BMI.

One way to increase adherence was to pair the app with a structured program or a Facebook group [31,81], but another documented way is to simplify calorie counting. For example, monitoring only the high-calorie foods consumed that have limited nutritional value. In a study comparing detailed counting of calories on Fitbit to simplified calorie counting, the group with the simplified option had better adherence throughout the

XSL•FO
RenderX

study, with a median day of self-monitoring of 97% compared to 49% for the detailed group [84].

Environmental Factors

The need for an internet connection to manually enter food intake in the app was negatively associated with daily food logging adherence [85].

Discussion

Recommendations

Based on the results of our scoping review, we provided a list of recommendations to improve acceptability and feasibility associated with the use of calorie-counting apps to monitor and manage weight in adult populations living with obesity and weight-related chronic disease. We present below our recommendations for developers, clinicians, and researchers (Table 5), based on potential improvements associated with behavioral, structural, socioeconomic, and research and development components of the use of calorie-counting apps. These recommendations are a result of collecting information on the functionalities, features, and user engagement metrics that factored into the acceptability and feasibility of the apps.

Table 5. Recommendations to improve the acceptability and feasibility of implementing the use of calorie-counting apps in clinical practice in populations living with weight-related chronic disease.

Category and subcategory	Recommendations
Behavioral	
Motivation	• Ensure a thoughtful frequency of reminders, feedback, and progress reports to help users stay on track without inducing fatigue or discouragement • Allow users to opt out of notifications to avoid discouragement in certain users • Include game features in apps, deliver app-based interventions in a context of group competition, and/or provide financial incentives associated with app use to increase motivation
Guidance	• Provide more guidance on how to make healthy food choices and how to deal with diet slips to help users progress toward their goals
Structural	
App function	• Ensure regular app updates to stop the frequent app crashes and decrease frustration from app users • Allow offline functioning to improve consistency of calorie logging • Offer users simplified calorie logging options, such as only keeping track of high-calorie and low-nutrition food, to reduce the workload associated with calorie tracking
Content	• Include a variety of multiethnic, local, and home-cooked foods as well as take-out meals and restaurant options for users to choose from when logging dietary intake • Include features adapted for users with low vision, such as simplified terminology and screen organization, as well as clearly defined tabs • Include a tutorial with the app to make sure users understand all functionalities
Socioeconomic	
Costs	• Offer a free or very affordable app when implementing an app used in clinical interventions to avoid reduced accessibility for populations with low income
Research and development	
Engagement	• Involve future users and health care professionals in all stages of app development and clinical interventions, as well as in associated research projects, to foster co-construction of technological and clinical approaches

Behavioral Recommendations

Motivation

To develop a calorie-counting app that is well-suited for clinical implementation, it is important to consider some factors that keep users engaged. Motivation is one of the major factors fostering long-term app use and thus promoting adherence. Our scoping review showed that reminders, feedback, encouragement, and progress reports were well received by users (see the Acceptability section). For the best results, they should be thoughtfully spaced out so that they help users to stay on track without being too frequent to avoid fatigue or discouragement. This is especially true for the progress reports, where too frequent reports can lead to stagnant results and loss of motivation. In addition, allowing users to opt out of notifications to avoid feeling discouraged represents another associated recommendation. To increase motivation to use calorie-counting apps, other behavioral recommendations have been suggested, such as including game features in apps, delivering app-based interventions in a context of group competition, and providing financial incentives associated with app use [79].

XSL•FO
RenderX

Guidance

Our scoping review suggested that greater user understanding of calorie intake and its health implications may be linked to better engagement with calorie-counting apps [85]. Accordingly, providing structured guidance on healthy food choices and strategies to address dietary lapses could support user adherence and long-term behavior change.

Structural Recommendations

App Function

We formulated various recommendations related to the app function. First, ensuring regular app updates to stop the frequent app crashes would decrease frustration from app users. In addition, having an offline option would enable users to be more consistent by tracking calories when they do not have internet access. To reduce the workload and time spent tracking calories, we recommend offering users simplified calorie-counting options, such as only keeping track of high-calorie and low-nutrition food, for instance. Another recommendation to improve function while reducing workload for users would be to rely on artificial intelligence to perform automatic estimation of the calorie content of food items shown in pictures. This could significantly simplify the task of calorie logging for users and hence increase motivation and adherence.

Content

The content of calorie-counting apps could also be improved. To reduce workload for users and increase app usefulness, we recommend including a variety of multiethnic, local, and home-cooked foods, as well as take-out meals and restaurant options for users to choose from when logging into dietary intake. Moreover, app content should reflect diverse cultural and racial backgrounds to ensure inclusivity and relevance across populations with different dietary habits and nutritional needs. In addition, to tailor the app for users with low vision, we recommend the use of simplified terminology and screen organization with clearly defined tabs that can be easily seen. Finally, including in the app content a tutorial for users to understand all the functions the apps have to offer also represents a key recommended improvement for calorie-counting apps.

Socioeconomic Recommendations

Our scoping review identified that the cost of a calorie-counting app could represent a barrier limiting its acceptability and, consequently, the adherence to the app. We recommend offering a free or very affordable app to improve accessibility for users with lower income, thus making it easily recommendable by health care professionals to all patients.

Research and Development Recommendation

Successful implementation of calorie-counting apps in clinical interventions targeting patients living with weight-related chronic diseases relies on reaching high enough acceptability, feasibility, and adherence levels for both users and health care professionals. For that purpose, we provide a general recommendation to foster engagement of these stakeholders in all stages required to develop an app, but also to implement app use in clinical practice. Co-construction of the technological and clinical approaches together with users and health care professionals would be instrumental to strengthening acceptability, feasibility, and adherence.

Limitations

Given our search strategy, which required papers to include populations living with chronic conditions, studies that only described apps without targeting these populations were not necessarily included. This may have limited the depth of our features and functionalities of calorie-counting apps analysis. Even when combining app descriptions from the included studies, it is possible that some features were missed if the primary objective of certain papers was not to describe the app in detail. We acknowledge that this may have led to the omission of specific app features. Nevertheless, we believe that our study still achieved its objective of providing an overview of the most common design features of calorie-counting apps for adults living with chronic conditions. Another limitation of this review is the lack of consistent reporting on reminder frequency within the primary studies, which prevented analysis of how reminder timing or intensity might influence user engagement.

Implications for Future Research

Building an app with these recommendations in mind will be a crucial step toward developing an application that meets the needs of adults living with chronic conditions in clinical settings. Another essential step in the development of these applications is the need for future research. As is the case with any intervention destined for a clinical setting, the validity of these apps, as well as an evaluation of their effectiveness, should be documented through future research. Many of these applications have a goal of promoting weight loss and healthy eating behaviors, and the effectiveness with which they can achieve those goals should be studied. As was done throughout the development of many studied apps in the papers selected by this study, future research should also evaluate the acceptability and the feasibility of the app throughout every stage of development to allow for necessary adjustments to be made. These future research initiatives should involve users and health care professionals in every stage, from app design to implementation in clinical practice. This will allow them to better meet the needs of the various stakeholders through fostering their engagement.

Throughout this paper, it was also possible to highlight several elements of acceptability and feasibility for calorie-counting apps, with more general information for app structure and content. While this was partly due to the nature of our scoping review, which was to provide an overview, we still noticed that the literature was somewhat limited when it came to detailed explanations of features and functionalities, especially for commercially available apps. Therefore, future research exploring specific features and functionalities of apps and their relationship with acceptance among adults living with chronic conditions is needed.

Conclusion

Calorie-counting apps can support management of obesity and related chronic diseases provided they reduce logging friction, personalize guidance, integrate with clinical workflows, and

XSL•FO
RenderX

leverage comprehensive, culturally inclusive food databases. Persistent barriers, technical instability, incomplete databases, and the time cost of manual entry contribute to erosion of adherence over time. Future development should emphasize automation (barcode- and image-based capture, offline use), privacy-preserving clinician data sharing, and simplified monitoring paradigms that maintain engagement. Rigorous evaluations are needed to verify automated calorie estimation, benchmark database quality, and assess equity and accessibility.

Acknowledgments

We would like to thank Marc Harper, bibliometric expert at Université de Moncton, for his valuable contributions to the development of the search strategy, including the formulation of keyword chains and Boolean syntax, conducting comprehensive searches across electronic databases, eliminating duplicates, performing initial data cleaning, and organizing the literature retrieved in Zotero.

Data Availability

All data files are available from the corresponding author upon reasonable request.

Funding

This study was supported through grant programs from the Université de Moncton and the Centre de formation médicale du Nouveau-Brunswick. This study was supported by a postdoctoral grant from the Université de Moncton (PLEIADE grant) awarded to AR under the supervision of JJ and GR. JJ also received funding from ResearchNB (grant PM_2025_017) to conduct this study and cover publication fees. Additional support for data analysis was provided through a funding award from the McKenna Institute on behalf of Shoppers Drug Mart. We extend our sincere thanks to the Centre de formation médicale du Nouveau-Brunswick for their administrative assistance.

Authors' Contributions

Conceptualization: JJ
Methodology: JJ
Validation: JJ
Data extraction: KD, MAG, AG, JL, AR
Data synthesis: KD, MAG, AG, JL, AR
Writing – original draft: KD, MAG, AG, JL, AR
Writing – review & editing: KD, AR, GR, JJ
Supervision: JJ
Project administration: GR, JJ

Conflicts of Interest

None declared.

Multimedia Appendix 1
Metadata of the extraction form used to gather the data from the 68 articles included in the scoping review.
[DOCX File , 22 KB - mhealth_v14i1e64139_app1.docx]

Multimedia Appendix 2
PRISMA-ScR checklist.
[PDF File (Adobe PDF File), 511 KB - mhealth_v14i1e64139_app2.pdf]

References

1. NCD Risk Factor Collaboration (NCD-RisC). Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19·2 million participants. Lancet 2016;387(10026):1377-1396 [FREE Full text] [doi: 10.1016/S0140-6736(16)30054-X] [Medline: 27115820]
2. Twells LK, Janssen I, Kuk JL. Canadian adult obesity clinical practice guidelines. Obesity Canada. 2020. URL: https://obesitycanada.ca/guidelines/epidemiology [accessed 2025-12-04]
3. Population health. New Brunswick Health Council. 2024. URL: https://nbhc.ca/data/browse/population-health [accessed 2025-12-04]
4. Jiang S, Lu W, Zong X, Ruan H, Liu Y. Obesity and hypertension. Exp Ther Med 2016;12(4):2395-2399 [FREE Full text] [doi: 10.3892/etm.2016.3667] [Medline: 27703502]

XSL•FO
RenderX

5. Koliaki C, Liatis S, Kokkinos A. Obesity and cardiovascular disease: revisiting an old relationship. Metabolism 2019;92:98-107. [doi: 10.1016/j.metabol.2018.10.011] [Medline: 30399375]
6. Verma S, Hussain ME. Obesity and diabetes: an update. Diabetes Metab Syndr 2017;11(1):73-79. [doi: 10.1016/j.dsx.2016.06.017] [Medline: 27353549]
7. Obesity and overweight. World Health Organization. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2025-12-04]
8. Wyatt HR. Update on treatment strategies for obesity. J Clin Endocrinol Metab 2013;98(4):1299-1306 [FREE Full text] [doi: 10.1210/jc.2012-3115] [Medline: 23443815]
9. Raber M, Liao Y, Rara A, Schembre SM, Krause KJ, Strong L, et al. A systematic review of the use of dietary self-monitoring in behavioural weight loss interventions: delivery, intensity and effectiveness. Public Health Nutr 2021;24(17):5885-5913 [FREE Full text] [doi: 10.1017/S136898002100358X] [Medline: 34412727]
10. Mobile fact sheet. Pew Research Center. 2024. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2025-12-04]
11. Gareev I, Gallyametdinov A, Beylerli O, Valitov E, Alyshov A, Pavlov V, et al. The opportunities and challenges of telemedicine during COVID-19 pandemic. Front Biosci (Elite Ed) 2021;13(2):291-298 [FREE Full text] [doi: 10.52586/E885] [Medline: 34937315]
12. Lieffers JRL, Hanning RM. Dietary assessment and self-monitoring with nutrition applications for mobile devices. Can J Diet Pract Res 2012;73(3):e253-e260. [doi: 10.3148/73.3.2012.e253] [Medline: 22958633]
13. Pagoto S, Xu R, Bullard T, Foster GD, Bannor R, Arcangel K, et al. An evaluation of a personalized multicomponent commercial digital weight management program: single-arm behavioral trial. J Med Internet Res 2023;25:e44955 [FREE Full text] [doi: 10.2196/44955] [Medline: 37642986]
14. Valinskas S, Nakrys M, Aleknavičius K, Jonusas J, Lileikienė A. User engagement and weight loss facilitated by a mobile app: retrospective review of medical records. JMIR Form Res 2023;7:e42266 [FREE Full text] [doi: 10.2196/42266] [Medline: 36692936]
15. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 2015;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
16. Metzendorf M, Wieland LS, Richter B. Mobile health (m-health) smartphone interventions for adolescents and adults with overweight or obesity. Cochrane Database Syst Rev 2024;2(2):CD013591. [doi: 10.1002/14651858.CD013591.pub2] [Medline: 38375882]
17. Yu Z, Sealey-Potts C, Rodriguez J. Dietary self-monitoring in weight management: current evidence on efficacy and adherence. J Acad Nutr Diet 2015;115(12):1931-1938. [doi: 10.1016/j.jand.2015.04.005] [Medline: 26028176]
18. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
19. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
20. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
21. Your personal research assistant. Zotero. URL: https://www.zotero.org/ [accessed 2025-09-12]
22. Alturki R, Gay V. The development of an arabic weight-loss app akser waznk: qualitative results. JMIR Form Res 2019;3(1):e11785 [FREE Full text] [doi: 10.2196/11785] [Medline: 30869639]
23. Ferrara G, Kim J, Lin S, Hua J, Seto E. A focused review of smartphone diet-tracking apps: usability, functionality, coherence with behavior change theory, and comparative validity of nutrient intake and energy estimates. JMIR Mhealth Uhealth 2019;7(5):e9232 [FREE Full text] [doi: 10.2196/mhealth.9232] [Medline: 31102369]
24. Gioia S, Vlasac IM, Babazadeh D, Fryou NL, Do E, Love J, et al. Mobile apps for dietary and food timing assessment: evaluation for use in clinical research. JMIR Form Res 2023;7:e35858 [FREE Full text] [doi: 10.2196/35858] [Medline: 37327038]
25. Kondo M, Okitsu T, Waki K, Yamauchi T, Nangaku M, Ohe K. Effect of information and communication technology-based self-management system dialbeticslite on treating abdominal obesity in the specific health guidance in Japan: randomized controlled trial. JMIR Form Res 2022;6(3):e33852 [FREE Full text] [doi: 10.2196/33852] [Medline: 35323122]
26. Tahir GA, Loo CK. A comprehensive survey of image-based food recognition and volume estimation methods for dietary assessment. Healthcare (Basel) 2021;9(12):1676. [doi: 10.3390/healthcare9121676] [Medline: 34946400]
27. Kim Y, Lee H, Seo JM. Integrated diabetes self-management program using smartphone application: a randomized controlled trial. West J Nurs Res 2022;44(4):383-394. [doi: 10.1177/0193945921994912] [Medline: 33655794]
28. Chen J, Allman-Farinelli M. Impact of training and integration of apps into dietetic practice on dietitians' self-efficacy with using mobile health apps and patient satisfaction. JMIR Mhealth Uhealth 2019;7(3):e12349. [doi: 10.2196/12349] [Medline: 30829577]

XSL•FO
RenderX

29. Agustina R, Febriyanti E, Putri M, Martineta M, Hardiany NS, Mustikawati DE, et al. Development and preliminary validity of an Indonesian mobile application for a balanced and sustainable diet for obesity management. BMC Public Health 2022;22(1):1221 [FREE Full text] [doi: 10.1186/s12889-022-13579-x] [Medline: 35725407]
30. Pellegrini CA, Conroy DE, Phillips SM, Pfammatter AF, McFadden H, Spring B. Daily and seasonal influences on dietary self-monitoring using a smartphone application. J Nutr Educ Behav 2018;50(1):56-61.e1 [FREE Full text] [doi: 10.1016/j.jneb.2016.12.004] [Medline: 29325663]
31. Burke LE, Sereika SM, Bizhanova Z, Parmanto B, Kariuki J, Cheng J, et al. The effect of tailored, daily, smartphone feedback to lifestyle self-monitoring on weight loss at 12 months: the smarter randomized clinical trial. J Med Internet Res 2022;24(7):e38243 [FREE Full text] [doi: 10.2196/38243] [Medline: 35787516]
32. Raju VB, Sazonov E. FOODCAM: A novel structured light-stereo imaging system for food portion size estimation. Sensors (Basel) 2022;22(9):3300 [FREE Full text] [doi: 10.3390/s22093300] [Medline: 35590990]
33. Lemacks JL, Adams K, Lovetere A. Dietary intake reporting accuracy of the bridge2u mobile application food log compared to control meal and dietary recall methods. Nutrients 2019;11(1):199 [FREE Full text] [doi: 10.3390/nu11010199] [Medline: 30669430]
34. Aizawa K, Maruyama Y, Li H, Morikawa C. Food balance estimation by using personal dietary tendencies in a multimedia food log. IEEE Trans. Multimedia 2013;15(8):2176-2185. [doi: 10.1109/TMM.2013.2271474]
35. Chaudhry BM. Food for thought. Mhealth 2019;5:20 [FREE Full text] [doi: 10.21037/mhealth.2019.06.02] [Medline: 31463306]
36. Anthimopoulos M, Dehais J, Shevchik S, Ransford BH, Duke D, Diem P, et al. Computer vision-based carbohydrate estimation for type 1 patients with diabetes using smartphones. J Diabetes Sci Technol 2015;9(3):507-515 [FREE Full text] [doi: 10.1177/1932296815580159] [Medline: 25883163]
37. Olfert MD, Barr ML, Hagedorn RL, Long DM, Haggerty TS, Weimer M, et al. Feasibility of a mHealth approach to nutrition counseling in an appalachian state. J Pers Med 2019;9(4):50 [FREE Full text] [doi: 10.3390/jpm9040050] [Medline: 31757057]
38. iDAT app (version 2.2.0). Health Promotion Board. 2015. URL: https://hpb.gov.sg/newsroom/article/decommissioning-of-idat-app [accessed 2024-06-25]
39. Myers A, Johnston N, Rathod V, Korattikara A, Gorban A, Silberman N, et al. Im2Calories: towards an automated mobile vision food diary. 2015 Presented at: IEEE International Conference on Computer Vision (ICCV); December 07-13, 2015; Santiago, Chile p. 1233-1241. [doi: 10.1109/iccv.2015.146]
40. Zahedani AD, McLaughlin T, Veluvali A, Aghaeepour N, Hosseinian A, Agarwal S, et al. Digital health application integrating wearable data and behavioral patterns improves metabolic health. NPJ Digit Med 2023;6(1):216 [FREE Full text] [doi: 10.1038/s41746-023-00956-y] [Medline: 38001287]
41. Moyen A, Rappaport AI, Fleurent-Grégoire C, Tessier A, Brazeau A, Chevalier S. Relative validation of an artificial intelligence-enhanced, image-assisted mobile app for dietary assessment in adults: randomized crossover study. J Med Internet Res 2022;24(11):e40449 [FREE Full text] [doi: 10.2196/40449] [Medline: 36409539]
42. Schusterbauer V, Feitek D, Kastner P, Toplak H. Two-stage evaluation of a telehealth nutrition management service in support of diabesity therapy. Stud Health Technol Inform 2018;248:314-321. [Medline: 29726453]
43. Oh SW, Kim K, Kim SS, Park SK, Park S. Effect of an integrative mobile health intervention in patients with hypertension and diabetes: crossover study. JMIR Mhealth Uhealth 2022;10(1):e27192 [FREE Full text] [doi: 10.2196/27192] [Medline: 35014961]
44. Burke LE, Zheng Y, Ma Q, Mancino J, Loar I, Music E, et al. The SMARTER pilot study: testing feasibility of real-time feedback for dietary self-monitoring. Prev Med Rep 2017;6:278-285 [FREE Full text] [doi: 10.1016/j.pmedr.2017.03.017] [Medline: 28409090]
45. Beijbom O, Joshi N, Morris D, Saponas S, Khullar S. Menu-menu-match: restaurant-specific food logging from images. : IEEE; 2015 Presented at: IEEE Winter Conference on Applications of Computer Vision; January 05-09, 2015; Waikoloa, HI, USA p. 844-851. [doi: 10.1109/wacv.2015.117]
46. Ahmad Z, Kerr DA, Bosch M, Boushey CJ, Delp EJ, Khanna N, et al. A mobile food record for integrated dietary assessment. MADiMa16 (2016) 2016;2016:53-62 [FREE Full text] [doi: 10.1145/2986035.2986038] [Medline: 28691119]
47. Chui T, Tan J, Li Y, Raynor HA. Validating an automated image identification process of a passive image-assisted dietary assessment method: proof of concept. Public Health Nutr 2020;23(15):2700-2710 [FREE Full text] [doi: 10.1017/S1368980020000816] [Medline: 32517834]
48. Tosi M, Radice D, Carioni G, Vecchiati T, Fiori F, Parpinel M, et al. Accuracy of applications to monitor food intake: evaluation by comparison with 3-d food diary. Nutrition 2021;84:111018. [doi: 10.1016/j.nut.2020.111018] [Medline: 33046348]
49. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013;15(4):e32 [FREE Full text] [doi: 10.2196/jmir.2283] [Medline: 23587561]

XSL•FO
RenderX

50. Fu HNC, Rizvi RF, Wyman JF, Adam TJ. Usability evaluation of four top-rated commercially available diabetes apps for adults with type 2 diabetes. Comput Inform Nurs 2020;38(6):274-280 [FREE Full text] [doi: 10.1097/CIN.0000000000000596] [Medline: 31904594]
51. Lim SL, Tay MHJ, Ong KW, Johal J, Yap QV, Chan YH, et al. Association between mobile health app engagement and weight loss and glycemic control in adults with type 2 diabetes and prediabetes (D'LITE Study): prospective cohort study. JMIR Diabetes 2022;7(3):e35039 [FREE Full text] [doi: 10.2196/35039] [Medline: 36178718]
52. Jin T, Kang G, Song S, Lee H, Chen Y, Kim S, et al. The effects of dietary self-monitoring intervention on anthropometric and metabolic changes via a mobile application or paper-based diary: a randomized trial. Nutr Res Pract 2023;17(6):1238-1254 [FREE Full text] [doi: 10.4162/nrp.2023.17.6.1238] [Medline: 38053827]
53. Kay MC, Miller HN, Askew S, Spaulding EM, Chisholm M, Christy J, et al. Patterns of engagement with an application-based dietary self-monitoring tool within a randomized controlled feasibility trial. AJPM Focus 2022;1(2):100037 [FREE Full text] [doi: 10.1016/j.focus.2022.100037] [Medline: 37791242]
54. Zhou J, Bell D, Nusrat S, Hingle M, Surdeanu M, Kobourov S. Calorie estimation from pictures of food: crowdsourcing study. Interact J Med Res 2018;7(2):e17 [FREE Full text] [doi: 10.2196/ijmr.9359] [Medline: 30401671]
55. Gonzalez-Ramirez M, Sanchez-Carrera R, Cejudo-Lopez A, Lozano-Navarrete M, Salamero Sánchez-Gabriel E, Torres-Bengoa MA, et al. Short-term pilot study to evaluate the impact of salbi educa nutrition app in macronutrients intake and adherence to the mediterranean diet: randomized controlled trial. Nutrients 2022;14(10):2061 [FREE Full text] [doi: 10.3390/nu14102061] [Medline: 35631202]
56. Zhang W, Yu Q, Siddiquie B, Divakaran A, Sawhney H. "Snap-n-Eat": food recognition and nutrition estimation on a smartphone. J Diabetes Sci Technol 2015;9(3):525-533 [FREE Full text] [doi: 10.1177/1932296815582222] [Medline: 25901024]
57. Bardus M, Ali A, Demachkieh F, Hamadeh G. Assessing the quality of mobile phone apps for weight management: user-centered study with employees from a lebanese university. JMIR Mhealth Uhealth 2019;7(1):e9836 [FREE Full text] [doi: 10.2196/mhealth.9836] [Medline: 30672742]
58. Riddell M, Li Z, Gal R, Calhoun P, Jacobs P, Clements M, T1DEXI Study Group. Examining the acute glycemic effects of different types of structured exercise sessions in type 1 diabetes in a real-world setting: the type 1 diabetes and exercise initiative (T1DEXI). Diabetes Care 2023;46(4):704-713 [FREE Full text] [doi: 10.2337/dc22-1721] [Medline: 36795053]
59. O'Neil PM, Miller-Kovach K, Tuerk PW, Becker LE, Wadden TA, Fujioka K, et al. Randomized controlled trial of a nationally available weight control program tailored for adults with type 2 diabetes. Obesity (Silver Spring) 2016;24(11):2269-2277 [FREE Full text] [doi: 10.1002/oby.21616] [Medline: 27804264]
60. Puigdomènech E, Robles N, Balfegó M, Cuatrecasas G, Zamora A, Saigí-Rubió F, et al. Codesign and feasibility testing of a tool to evaluate overweight and obesity apps. Int J Environ Res Public Health 2022;19(9):5387 [FREE Full text] [doi: 10.3390/ijerph19095387] [Medline: 35564781]
61. Ifejika NL, Bhadane M, Cai CC, Noser EA, Grotta JC, Savitz SI. Use of a smartphone-based mobile app for weight management in obese minority stroke survivors: pilot randomized controlled trial with open blinded end point. JMIR Mhealth Uhealth 2020;8(4):e17816. [doi: 10.2196/17816] [Medline: 32319963]
62. Ben Neriah D, Geliebter A. Weight loss following use of a smartphone food photo feature: retrospective cohort study. JMIR Mhealth Uhealth 2019;7(6):e11917 [FREE Full text] [doi: 10.2196/11917] [Medline: 31199300]
63. Hahn SL, Kaciroti N, Eisenberg D, Weeks HM, Bauer KW, Sonneville KR. Introducing dietary self-monitoring to undergraduate women via a calorie counting app has no effect on mental health or health behaviors: results from a randomized controlled trial. J Acad Nutr Diet 2021;121(12):2377-2388 [FREE Full text] [doi: 10.1016/j.jand.2021.06.311] [Medline: 34427188]
64. Baum Martinez I, Peters B, Schwarz J, Schuppelius B, Steckhan N, Koppold-Liebscher DA, et al. Validation of a smartphone application for the assessment of dietary compliance in an intermittent fasting trial. Nutrients 2022;14(18):3697 [FREE Full text] [doi: 10.3390/nu14183697] [Medline: 36145073]
65. Chew H, Lim S, Kim G, Kayambu G, So B, Shabbir A, et al. Essential elements of weight loss apps for a multi-ethnic population with high BMI: a qualitative study with practical recommendations. Transl Behav Med 2023;13(3):140-148. [doi: 10.1093/tbm/ibac090] [Medline: 36689306]
66. Vasiloglou MF, Christodoulidis S, Reber E, Stathopoulou T, Lu Y, Stanga Z, et al. Perspectives and preferences of adult smartphone users regarding nutrition and diet apps: web-based survey study. JMIR Mhealth Uhealth 2021;9(7):e27885 [FREE Full text] [doi: 10.2196/27885] [Medline: 34328425]
67. Patel ML, Hopkins CM, Brooks TL, Bennett GG. Comparing self-monitoring strategies for weight loss in a smartphone app: randomized controlled trial. JMIR Mhealth Uhealth 2019;7(2):e12209 [FREE Full text] [doi: 10.2196/12209] [Medline: 30816851]
68. Wharton CM, Johnston CS, Cunningham BK, Sterner D. Dietary self-monitoring, but not dietary quality, improves with use of smartphone app technology in an 8-week weight loss trial. J Nutr Educ Behav 2014;46(5):440-444. [doi: 10.1016/j.jneb.2014.04.291] [Medline: 25220777]

XSL•FO
RenderX

69. Kwon BC, VanDam C, Chiuve SE, Choi HW, Entler P, Tan P, et al. Improving heart disease risk through quality-focused diet logging: pre-post study of a diet quality tracking app. JMIR Mhealth Uhealth 2020;8(12):e21733 [FREE Full text] [doi: 10.2196/21733] [Medline: 33355537]
70. Kim HH, Kim Y, Michaelides A, Park YR. Weight loss trajectories and related factors in a 16-week mobile obesity intervention program: retrospective observational study. J Med Internet Res 2022;24(4):e29380 [FREE Full text] [doi: 10.2196/29380] [Medline: 35436211]
71. Eisenhauer CM, Brito F, Kupzyk K, Yoder A, Almeida F, Beller RJ, et al. Mobile health assisted self-monitoring is acceptable for supporting weight loss in rural men: a pragmatic randomized controlled feasibility trial. BMC Public Health 2021;21(1):1568 [FREE Full text] [doi: 10.1186/s12889-021-11618-7] [Medline: 34407782]
72. Gonzalez-Ramirez M, Cejudo-Lopez A, Lozano-Navarrete M, Salamero Sánchez-Gabriel E, Torres-Bengoa MA, Segura-Balbuena M, et al. SAlBi educa (Tailored Nutrition App for Improving Dietary Habits): initial evaluation of usability. Front Nutr 2022;9:782430 [FREE Full text] [doi: 10.3389/fnut.2022.782430] [Medline: 35520281]
73. Lee J, Bae S, Park D, Kim Y, Park J. The effectiveness of a monetary reimbursement model for weight reduction via a smartphone application: a preliminary retrospective study. Sci Rep 2020;10(1):15714 [FREE Full text] [doi: 10.1038/s41598-020-72908-5] [Medline: 32973278]
74. Chew HSJ, Achananuparp P, Dalakoti M, Chew NWS, Chin YH, Gao Y, et al. Public acceptance of using artificial intelligence-assisted weight management apps in high-income southeast Asian adults with overweight and obesity: a cross-sectional study. Front Nutr 2024;11:1287156 [FREE Full text] [doi: 10.3389/fnut.2024.1287156] [Medline: 38385011]
75. Zheng Y, Weinger K, Greenberg J, Burke LE, Sereika SM, Patience N, et al. Actual use of multiple health monitors among older adults with diabetes: pilot study. JMIR Aging 2020;3(1):e15995 [FREE Full text] [doi: 10.2196/15995] [Medline: 32202506]
76. Pagoto S, Tulu B, Waring ME, Goetz J, Bibeau J, Divito J, et al. Slip buddy app for weight management: randomized feasibility trial of a dietary lapse tracking app. JMIR Mhealth Uhealth 2021;9(4):e24249 [FREE Full text] [doi: 10.2196/24249] [Medline: 33792547]
77. Pagoto S, Xu R, Bullard T, Foster GD, Bannor R, Arcangel K, et al. An evaluation of a personalized multicomponent commercial digital weight management program: single-arm behavioral trial. J Med Internet Res 2023;25:e44955 [FREE Full text] [doi: 10.2196/44955] [Medline: 37642986]
78. Goh G, Tan NC, Malhotra R, Padmanabhan U, Barbier S, Allen JC, et al. Short-term trajectories of use of a caloric-monitoring mobile phone app among patients with type 2 diabetes mellitus in a primary care setting. J Med Internet Res 2015;17(2):e33 [FREE Full text] [doi: 10.2196/jmir.3938] [Medline: 25648130]
79. Laing BY, Mangione CM, Tseng C, Leng M, Vaisberg E, Mahida M, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. Ann Intern Med 2014;161(10 Suppl):S5-S12 [FREE Full text] [doi: 10.7326/M13-3005] [Medline: 25402403]
80. Alshathri DM, Alhumaimeedy AS, Al-Hudhud G, Alsaleh A, Al-Musharaf S, Aljuraiban GS. Weight management apps in Saudi Arabia: evaluation of features and quality. JMIR Mhealth Uhealth 2020;8(10):e19844 [FREE Full text] [doi: 10.2196/19844] [Medline: 33104013]
81. Rowland S, Ramos AK, Trinidad N, Quintero S, Johnson Beller R, Struwe L, et al. Feasibility, usability and acceptability of a mhealth intervention to reduce cardiovascular risk in rural hispanic adults: descriptive study. JMIR Form Res 2022;6(12):e40379 [FREE Full text] [doi: 10.2196/40379] [Medline: 36563025]
82. Carter MC, Burley VJ, Cade JE. Weight loss associated with different patterns of self-monitoring using the mobile phone app my meal mate. JMIR Mhealth Uhealth 2017;5(2):e8 [FREE Full text] [doi: 10.2196/mhealth.4520] [Medline: 28153814]
83. Garizábalo-Dávila CM, Rodríguez-Acelas AL, Cañon-Montañez W. Usefulness of acceptability and feasibility assessment in studies of nursing interventions. Invest Educ Enferm 2023;41(1):e2 [FREE Full text] [doi: 10.17533/udea.iee.v41n1e02] [Medline: 37071859]
84. Patel ML, Cleare AE, Smith CM, Rosas LG, King AC. Detailed versus simplified dietary self-monitoring in a digital weight loss intervention among racial and ethnic minority adults: fully remote, randomized pilot study. JMIR Form Res 2022;6(12):e42191 [FREE Full text] [doi: 10.2196/42191] [Medline: 36512404]
85. Banerjee P, Mendu VVR, Korrapati D, Gavaravarapu SM. Calorie counting smart phone apps: effectiveness in nutritional awareness, lifestyle modification and weight management among young Indian adults. Health Informatics J 2020;26(2):816-828 [FREE Full text] [doi: 10.1177/1460458219852531] [Medline: 31195885]
86. Extender app (Version 5.7.1. Food Database GmbH (FDDB). 2024. URL: https://fddb.info/db/en/index.html [accessed 2024-06-25]
87. Butryn ML, Godfrey KM, Martinelli MK, Roberts SR, Forman EM, Zhang F. Digital self-monitoring: does adherence or association with outcomes differ by self-monitoring target? Obes Sci Pract 2020;6(2):126-133 [FREE Full text] [doi: 10.1002/osp4.391] [Medline: 32313670]
88. Li S, Zhou Y, Tang Y, Ma H, Zhang Y, Wang A, et al. Behavior change resources used in mobile app-based interventions addressing weight, behavioral, and metabolic outcomes in adults with overweight and obesity: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2025;13:e63313 [FREE Full text] [doi: 10.2196/63313] [Medline: 40829125]

XSL•FO
RenderX

89. Lehmann M, Jones L, Schirmann F. App engagement as a predictor of weight loss in blended-care interventions: retrospective observational study using large-scale real-world data. J Med Internet Res 2024;26:e45469 [FREE Full text] [doi: 10.2196/45469] [Medline: 38848556]

Abbreviations

mHealth: mobile health
PRISMA-ScR: Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews

Edited by G Badicu; submitted 12.Jul.2024; peer-reviewed by S Rowland, S Sulaiman, K Ufholz; comments to author 29.Jun.2025; revised version received 30.Oct.2025; accepted 07.Nov.2025; published 01.Apr.2026.

Please cite as:
Dugas K, Giroux MA, Guerroudj A, Leger J, Rouhafzay A, Rouhafzay G, Jbilou J
Calorie-Counting Apps for Monitoring and Managing Calorie Intake in Adults Living With Weight-Related Chronic Diseases: Decade-Long Scoping Review (2013-2024)
JMIR Mhealth Uhealth 2026;14:e64139
URL: https://mhealth.jmir.org/2026/1/e64139
doi:10.2196/64139
PMID:41329042

©Kaylee Dugas, Marie-Andrée Giroux, Abdelatif Guerroudj, Jazna Leger, Asal Rouhafzay, Ghazal Rouhafzay, Jalila Jbilou. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 01.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Key Features of Engagement Strategies in Nutrition Apps for Adults: Scoping Review

Maria F Vasiloglou[1], MSc, PhD; Zoë van der Heijden[2], MSc; Eric Antoine Scuccimarra[1], MSc; Alberto Conde Freniche[1], MSc, M Vet Med; Desiree A Lucassen[2], MSc, PhD; Nienke de Vlieger[3], MSc, PhD; Frédéric Ronga[4], MSc, PhD; Elske Brouwer-Brolsma[2], MSc, PhD; Tamara Bucher[3,5], MSc, PhD

[1]Nestlé Institute of Health Sciences, Nestlé Research, Lausanne, Switzerland

[2]Division of Human Nutrition and Health, Wageningen University & Research, Wageningen, The Netherlands

[3]School of Health Sciences, Faculty of Health and Medicine, University of Newcastle, Callaghan, NSW, Australia

[4]Strategic Research Operations Unit, Nestlé Research, Lausanne, Switzerland

[5]Division of Nutrition and Dietetics, Bern University of Applied Sciences, Bern, Switzerland

Corresponding Author:
Maria F Vasiloglou, MSc, PhD
Nestlé Institute of Health Sciences, Nestlé Research
Rte du Jorat 57
Lausanne, 1000
Switzerland
Phone: 41 0762438288
Email: maria.vasiloglou@rd.nestle.com

Abstract

Background: Nutrition apps offer scalable opportunities to support dietary behavior change and prevent chronic diseases. Their success depends on sustained user engagement, which is essential yet challenging to achieve and, consequently impacts the long-term effectiveness of these digital tools. Engagement strategies have been widely explored in digital health, but a comprehensive synthesis focusing on nutrition apps for adults is lacking.

Objective: This scoping review aimed to map the current engagement approaches and metrics implemented in nutrition apps targeting adults and to identify how user engagement is defined across studies.

Methods: We conducted a search of the PubMed, Scopus, Cochrane, and Web of Science databases for relevant studies published from January 1, 2013, to June 30, 2024. The inclusion criteria included original adult interventional or observational studies that evaluated nutrition apps and reported user-engagement strategies or metrics. Two reviewers independently screened records in Covidence, with discrepancies resolved by a third reviewer. Data were charted across study characteristics, engagement strategies, and engagement metrics and then synthesized narratively.

Results: A total of 59 studies that used apps to improve dietary behaviors were included in our analysis, including randomized controlled trials, observational trials, and mixed methods studies. Most of these apps were designed for adults who were overweight and obese. The studies were primarily conducted in North America and Europe and were randomized controlled trials or nonrandomized intervention studies, with varying durations and sample sizes. Engagement strategies varied widely, and engagement was typically measured by frequency of specific function use and frequency of app use, followed by retention rate. The most common engagement strategies reported in studies were push notifications (n=29, 49%), behavioral theory integration (n=24, 41%), personalization and customization (n=19, 32%), and goal-setting features (n=18, 31%). Only 31% (n=18) of studies provided an explicit definition of "user engagement," and definitions were highly heterogeneous. Engagement measurement was dominated by quantitative system-recorded metrics, including time and frequency of using specific functions (n=38, 64%), app use frequency (n=34, 58%), and retention (n=17, 29%). Few studies assessed qualitative or long-term engagement dimensions, and long-duration studies rarely integrated adaptive or contextualized engagement mechanisms. Research apps more frequently used theory-driven strategies compared with commercial apps, which tended to emphasize streamlined user experience.

Conclusions: Although several engagement strategies are commonly used, their implementation is inconsistent and often lacks grounding in conceptual frameworks. Research in the future needs to prioritize the use of common definitions for user engagement and measurement criteria while implementing user-centered design methods and using multiple research approaches to study the

XSL•FO
RenderX

complex patterns of user engagement. The evidence base for engagement strategies needs strengthening because it will support the development of sustainable nutrition mobile health interventions.

(JMIR Mhealth Uhealth 2026;14:e82276) doi:10.2196/82276

KEYWORDS

mobile health; mHealth; nutrition apps; user engagement; digital health; behavior change; scoping review; mobile apps; dietary behavior

Introduction

Background

Noncommunicable diseases, including cardiovascular disease, cancer, respiratory conditions, and diabetes, are the leading causes of death globally, accounting for more than 70% of deaths worldwide [1]. Overweight and obesity significantly increase the risk of these noncommunicable diseases, impacting morbidity and mortality rates [1]. The global obesity epidemic is worsening, with significant implications for physical and economic health [2].

Digital health tools for diet monitoring, planning, and precision nutrition have the potential to provide scalable tools to support health care professionals in managing nutrition-related diseases and obesity [3]. Despite the potential benefits, the integration of digital health tools into large, complex health systems remains limited. Challenges such as the need for user-centered, context-dependent, and customizable designs hinder widespread adoption, particularly in low-income countries with limited financial resources [4].

Mobile health (mHealth) apps, specifically those targeting nutrition, have proliferated over the past decade, aiding users in tracking food intake, providing personalized dietary advice, and promoting healthier lifestyles. There were more than 330,000 digital health apps available, with a substantial proportion focusing on nutrition [5]. These apps are increasingly being recognized for their potential to improve poor dietary habits and diet-related chronic diseases. The scalability and wide reach of mHealth apps, combined with their capacity to deliver personalized dietary guidance at a low cost, position them as a valuable approach to promote healthier diets and thereby support public health efforts to reduce the burden of chronic disease [6]. Two primary categories can be identified based on the developers' affiliations and goals: academic apps are created by nutrition or dietetics experts to offer a dependable, scientifically validated tool, primarily for research purposes. Conversely, consumer apps are usually developed by private companies specializing in digital development, aimed at the general public, with a primary focus on commercial objectives [7].

Although there is modest evidence that nutrition apps can be effective in promoting healthier dietary behaviors [6], a significant challenge lies in maintaining long-term user engagement with the app ("Little e"), which is necessary to ensure sufficient exposure to relevant behavior change techniques [8] and ultimately engagement in the behavior change process and outcomes ("Big E") [9].

Although there is modest evidence that nutrition apps can be effective in promoting healthier dietary behaviors [6], a significant challenge lies in maintaining long-term user engagement, which is necessary to ensure sufficient exposure to relevant behavior change techniques [8]. User engagement is a multifaceted concept and includes both the extent of app use, such as frequency, duration, and depth of access, as well as the quality of the user experience, including attention, interest, and emotional involvement. Despite the growing popularity of mHealth apps, many nutrition apps struggle to retain user engagement over time [10-12]. For example, users of the commercial weight loss app Lose It! engaged for an average of only 29 days [13]. Similarly, in a short-term intervention, participants recorded vegetable intake on just 11 out of 28 days, with engagement declining as the study progressed [14]. Low engagement is a common challenge in app-based interventions, with approximately 71% of users discontinuing within the first 90 days. Since this significantly undermines effectiveness [15-17], it is crucial to understand the factors that support sustained user engagement [18].

In an effort to understand user engagement in mHealth apps, several systematic reviews have been conducted to synthesize the different approaches in digital health interventions, particularly in mental health [19-23], physical activity, and chronic disease prevention and management [18,24,25]. In addition to these focused reviews, broader systematic reviews have also been performed, encompassing digital health interventions through multiple modes of delivery beyond mHealth apps, combining a variety of lifestyle behaviors [26,27]. Furthermore, specific reviews have examined digital health interventions for population groups, such as children [28] or older adults [29]. These studies provide an overview of useful engagement strategies in mHealth more broadly and for specific applications (ie, physical activity, mental health, and chronic disease).

To address the challenge of sustaining engagement in nutrition apps, various strategies have been implemented to enhance motivation and user experience, including gamification, personalization [30], and the integration of social features [31]. However, to the best of our knowledge, no review has yet provided a comprehensive overview of the full range of strategies used to promote user engagement in nutrition apps for the general adult population.

Objectives

This scoping review aims to provide an overview of the current landscape of predefined strategies implemented in apps used for nutrition interventions to enhance user engagement, identify gaps in the literature, and guide future research and app

XSL•FO
RenderX

development. Another goal is to map and synthesize the definitions of user engagement used in the studies, creating a comprehensive and unified definition.

Methods

Overview

We chose a scoping review to explore how user engagement is defined, measured, and promoted in nutrition apps due to the broad and heterogeneous nature of the literature across disciplines such as nutrition, behavioral science, and digital health. This approach enables a comprehensive overview of existing evidence without limiting study design or outcomes. This review is guided by the methodological framework originally proposed by Arksey and O'Malley [32], which outlines 5 key stages: identifying the research question, identifying relevant studies, study selection, charting the data, and summarizing and reporting the results. This scoping review was conducted in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist, which can be found in Multimedia Appendix 1.

Search Strategy and Selection Criteria

We searched 4 databases, namely PubMed, Scopus, Cochrane, and Web of Science, for relevant studies published between January 1, 2013, and June 30, 2024. The detailed search strategy can be found in Multimedia Appendix 1.

Eligibility Criteria

Our inclusion and exclusion criteria followed the population, intervention, comparison, outcome, and study design framework (Multimedia Appendix 1). This scoping review included studies that examined engagement strategies within nutrition apps used by adults aged above 18 years, across a range of health conditions. The review focused on interventions that incorporate elements such as gamification, personalized content, social integration, and behavior change techniques aimed at enhancing user engagement. Excluded were studies involving children and adolescents below 18 years, and those focusing on non–nutrition-related apps (eg, fitness-only or medication adherence apps without a primary nutrition component). Furthermore, studies that assessed engagement without reporting specific metrics such as daily active users, time spent in the app, or adherence rates were excluded. Only original interventional and observational studies were considered, excluding any type of review (eg, umbrella reviews, scoping reviews, or narrative reviews); conference abstracts; case reports; expert opinions without empirical data; and editorials, as well as quality and content analyses of apps found in app stores.

Screening and Data Extraction

Covidence software was used to manage the screening flow, beginning with the removal of duplicate records, followed by title and abstract screening conducted by pairs of reviewers (MFV, EAS, EB-B, TB, ACF, FR, SL, NvD, and ZvdH). All potentially eligible abstracts underwent full-text assessment by pairs of reviewers (MFV, EAS, SL, ACF, EB-B, FR, ZvdH, NdV, and DAL). Data extraction of eligible papers was performed by a single investigator and then verified by a second investigator (MFV or ZvdH). Any discrepancies were resolved through discussion with a third reviewer (EB-B).

An Excel (Microsoft Corp) data extraction spreadsheet was developed to systematically compile the following study-level information: author, year, aim, study design, location, total number of participants, female or male, intervention group, control group, duration, age, BMI, condition or disease, type of app, commercial or research, app features, intervention description, incentive, user-engagement strategies, description of those strategies, engagement definition, app evaluation criteria, specification of metrics, description of metrics, relationship between user engagement with smartphone apps and main outcome, app evaluation results, and strengths and limitations. Extracted data on engagement strategies were classified into the following categories: push notifications, personalization, goal setting, game-like features, multimedia formats, monitoring (real time or passive), peer-to-peer communication, access to professionals, content reliability, quality of feedback, behavioral theory, or none or insufficiently specified. User metrics (ie, engagement metric or indicators as specified in the protocol) were categorized into average session length, daily app use, drop-offs, most visited or active screens, overall time used, retention rate, time and frequency of using specific functions, frequency of app use, and others.

The effect on health outcomes were extracted from the included studies and were first categorized as follows: (1) *positive outcomes*—significant improvements ($P<.05$) or positive trends ($P<.10$), (2) *negative outcomes*—significant deterioration ($P<.05$) or negative trends ($P<.10$), (3) *neutral or no outcomes*—no significant changes ($P\geq.10$), or (4) not reported—no or insufficient data on health outcomes. Subsequently, the number of studies reporting positive, neutral, or neutral or no outcomes was counted for each category of user metrics to describe how frequently each metric appeared in studies with different types of outcomes. This counting procedure was purely descriptive and was not intended to infer causality, comparative effectiveness, or the predictive value of specific engagement metrics.

Finally, the duration of the studies was defined as short when it lasted less than a month, medium when it lasted 1 to 3 months, and long when it lasted more than 3 months. These classifications were developed specifically for this study by the authors.

Results

In the following section, the outcomes from the screening and analysis process are presented, including the characteristics of the included studies, the engagement strategies, and the metrics identified.

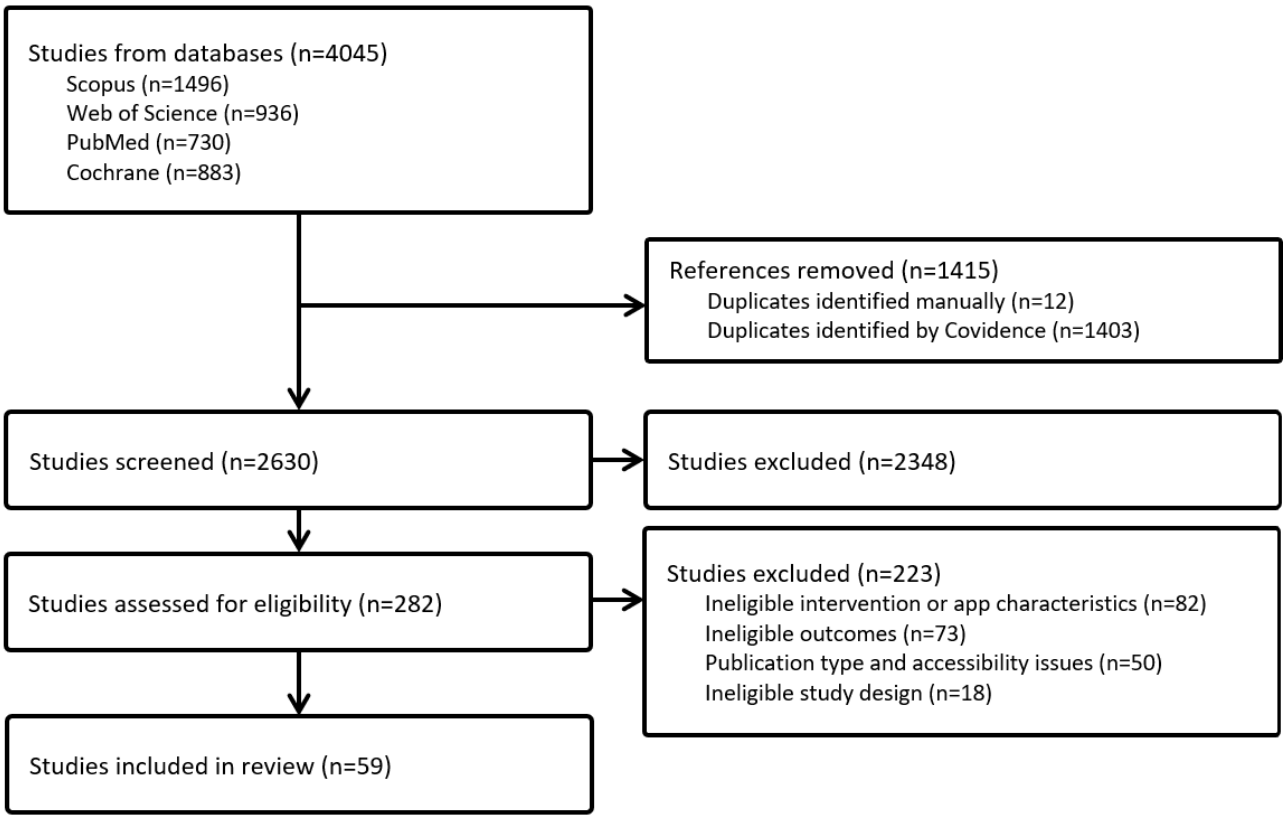
Search Results

Our searches retrieved 4045 records, and after the exclusion of duplications, 2630 (65%) records underwent abstract screening, of which 282 (7%) underwent full-text review (Figure 1). On the basis of the established inclusion and exclusion criteria, 59

XSL•FO
RenderX

studies were included in our scoping review (Multimedia Appendix 1).

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) flowchart.



Characteristics of Included Studies

Population

The scoping review included a total of 59 studies examining nutrition apps. The most common study design was a randomized controlled trial (RCT), accounting for 41% (24/59) of the studies. The remaining study designs were nonrandomized intervention studies (11/59, 19%), mixed methods studies (10/59, 17%), observational studies (8/59, 14%), and secondary analyses (6/59, 9%). Most studies (26/59, 44%) were conducted in North America, followed by Europe (14/59, 24%), Australia (9/59, 15%), Asia (7/59, 12%), and New Zealand (2/59, 3%); 1 (N=59, 2%) study did not specify a location. Sample sizes varied across studies from 12 to 1,011,008. Most of the participants (50/59, 85%) were adults (aged 18-55 years). In 24% (14/59) of the studies, women made up 75% to 90% of the participants, and in 39% (23/59) of the studies, women constituted 50% to 75% of the study population.

The duration of studies also varied widely. Most studies (22/59, 37%) had a medium duration of 1 to 3 months. A smaller proportion (10/59, 17%) had a short duration of less than 1 month, including 2 (N=59, 3%) studies that lasted 1 week or less. Studies with a long duration of more than 3 months accounted for 32% (19/59), spanning from over 3 months to over a year. A notable portion did not specify the study duration or left it unreported (8/59, 14%). Most studies (24/59, 41%) focused on weight management. Diabetes (type 1, type 2, and gestational) was the focus in 10% (6/59) of the studies, while pregnancy was the target in 7% (4/59) of the studies. Table 1 provides detailed information about the studies and their populations.

In our analysis, the distribution between commercial and research-based nutrition apps was assessed. Research apps constituted the majority, comprising 75% (44/59) of the studies included in our review. In contrast, commercial apps were represented in 25% (15/59) of the studies. Detailed characteristics of the included studies can be found in Multimedia Appendix 1.

XSL•FO
RenderX

Table 1. Basic characteristics of the study populations (N=59).

Characteristic	Studies, n (%)
Participants	
≤20	6 (10)
20-50	9 (15)
50-100	14 (24)
100-200	14 (24)
200-500	7 (12)
500-1000	3 (5)
>1000	6 (10)
Duration	
≤1 month	10 (17)
1 month-3 months	22 (37)
≥3 months	19 (32)
Not specified or not applicable	8 (14)
Mean age group (y)	
18-65	53 (90)
≥65	3 (5)
Not reported	3 (5)
Mean BMI (kg/m^2)	
Underweight (<18.5)	0 (0)
Normal (18.5-24.9)	4 (7)
Overweight (25.0-29.9)	14 (24)
Obesity 1 (30.0-34.9)	19 (32)
Obesity 2 (35.0-39.9)	6 (10)
Not reported	20 (34)
Condition or disease	
Overweight or obesity	24 (41)
Diabetes (type 1, type 2, or gestational)	6 (10)
Pregnancy	4 (7)
Mental health conditions	3 (5)
Older adults	2 (3)
Suboptimal lifestyle habits	2 (3)
Others	7 (14)
Not reported	12 (20)

Features

A variety of user-engagement strategies were used across the included nutrition apps (refer to Figure S1 in Multimedia Appendix 1). The most frequently applied strategies were push notifications or prompts (29/59, 49%), behavioral theory (24/59, 41%), personalization and customization (19/59, 32%), and goal setting (18/59, 31%). Less commonly reported were features such as peer-to-peer communication (9/59, 15%), game-like features (8/59, 14%), and multimedia formats (5/59, 8%). A total of 4 (N=59, 7%) studies did not report on user-engagement strategies applied or described them insufficiently.

To better understand how these strategies were operationalized, data were extracted on the specific app features used. The most common feature was self-monitoring or tracking (44/59, 75%), followed by prompts or notifications (25/59, 41%). Personalized feedback (22/59, 37%), goal-setting tools (15/59, 27%), and outcome tracking or progress summaries (13/59, 22%) were also frequently included. Less commonly used features included multimedia materials and social features (both 11/59, 19%),

XSL•FO
RenderX

data sharing or professional access (9/59, 15%), and more advanced technologies such as artificial intelligence chatbots or image recognition (2/59, 3%). A small number of studies reported features such as rewards (4/59, 7%), reflective logging tools (7/59, 12%), or photo-based modules (6/59, 10%).

User-Engagement Metrics Included

To evaluate how users interacted with these features, most of the studies (38/59, 64%) included metrics of functions related to time and frequency of specific functions. Many of them also included frequency of app use (34/59, 58%). Moreover, some of the studies (17/59, 29%) included retention rate and others included the number of drop-offs (15/59, 25%). Fewer studies (n=11, 19%) included metrics such as the overall time that an app was used, the daily app use time (9/59, 15%), the most visited or the most active screens (8/59, 14%), and the average session length (3/59, 5%; Figure S2 in Multimedia Appendix 1).

Table 2 presents the health outcomes reported in studies that included a given engagement metric. For each metric, we examined only the subset of studies that reported that specific metric and assessed whether they described improvements, no changes, or insufficient detail regarding effects on health outcomes. Most metrics were included in studies reporting positive effects. There were no studies that reported a negative association of a metric with health outcomes or components.

Table 2. Frequency of engagement metrics applied, used, and corresponding health outcomes reported in studies that included the specific metric (N=59).

Engagement metric	Improvements or positive changes, n	No changes, n	Not reported or insufficient detail, n	Studies, n (%)
Time or frequency of using specific functions	14	5	19	38 (64)
Frequency of app use	12	6	16	34 (58)
Retention	6	2	9	17 (29)
Drop-offs	2	3	10	15 (25)
Overall time used	4	3	4	11 (19)
Retention	3	0	7	10 (17)
Daily app use	4	1	4	9 (15)
Most visited or active screens	2	1	5	8 (14)

Commercial Versus Research Apps

A total of 15 (N=59, 25%) commercial apps and 44 (75%) research apps were analyzed to compare user-engagement strategies. Push notifications or prompts were used in 33% (5/15) of commercial apps and in 55% (24/44) of research apps, suggesting that research apps may more frequently incorporate reminder-based engagement. Personalization and customization features were equally present in commercial and research apps, each used by 33% (5/15) and 32% (14/44) of the apps, respectively. Goal setting was more common in research apps (15/44, 34%) compared to commercial apps (3/15, 20%). Game-like features were used by 13% (2/15) of commercial apps and 14% (6/44) of research apps.

App Evaluation Results Among Apps That Included Evaluation

In our review, 32 (54%) of the 59 included studies reported various dimensions of app evaluation. The most frequently assessed aspect was user satisfaction, noted in 50% (16/32) of the studies that included evaluations. This was followed by ease of use, reported in 25% (8/32) of the studies. App functionalities were evaluated in 28% (9/32) of the studies. Aesthetics, intention to use, and attitude toward the app were each evaluated in 9 (n=32, 28%) studies. Attitude toward one's own health was assessed in only 3% (1/32) of the studies. Suggestions for improvement were noted in 19% (6/32) of the studies. Additionally, 19% (6/32) of the studies did not report specific categories of user feedback.

App Evaluation Criteria

Feasibility was the most frequently assessed criterion, reported in 13 of 59 (22%) studies. This was closely followed by user satisfaction, which was evaluated in 12 of 59 (20%) studies. Usability and acceptability were each assessed in 10 of 59 (17%) studies. Finally, 26 of 59 (44%) studies did not report any specific evaluation criteria.

Key Engagement Strategies Among Adult Populations With Specific Needs

Key approaches varied across the 30 studies analyzing engagement strategies in apps targeting specific populations of adults. Among the 24 studies on apps for individuals with overweight and obesity, push notifications and prompts (n=12, 50%), behavioral theory (n=12, 50%), goal setting (n=8, 33%), peer-to-peer communication (n=5, 21%), and access to professionals (n=5, 21%) appeared as key engagement strategies. Engagement tracking focused on assessing the time and frequency of using specific functions (15/24, 63%), app use frequency (15/24, 63%), and the retention rate (11/24, 46%). Similarly, the 6 studies on apps for individuals living with diabetes prioritized behavioral theory (n=5, 83%), with half (n=3, 50%) of them incorporating push notifications, professional support, and real-time monitoring features, while

XSL•FO
RenderX

goal setting (n=1, 17%) and passive data monitoring (n=2, 33%) were less common.

Variation in Engagement Strategies Across Study Durations

In studies with different durations, distinct patterns emerged. Of the 10 (17%) short-duration studies (<1 month), 8 (80%) studies used push notifications or prompts extensively. The 22 (N=59, 37%) medium-duration studies (1-3 months) highlighted the importance of behavioral theory, with 11 (50%) of them incorporating it. The 19 (N=59, 32%) long-duration studies (>3 months) also favored behavioral theory and push notifications equally, each used in 42% (8/19) of the cases.

Short-duration studies prioritized tracking the time and frequency of using specific functions (7/10, 70%) and the frequency of app use (6/10, 60%). In medium-duration studies, 73% (16/22) focused on the time and frequency of using specific functions, and 55% (12/22) focused on the frequency of app use. Long-duration studies showed a significant measurement of time and frequency of using specific functions (9/19, 47%) and retention rates (6/19, 32%).

The presence of defined engagement strategies was examined across 59 studies. Of these, 18 (31%) studies provided an explicit definition of user engagement, while 9 (15%) studies presented an implicit conceptual definition. However, most of the studies (32/59, 55%) did not define user engagement at all. Among the 27 studies that contained either explicit or implicit definitions, various dimensions were highlighted. The most referenced criterion was use patterns, mentioned in 21 (n=27, 78%) studies. Feature interaction was the second most common, appearing in 8 (n=27, 30%) studies, followed by behavioral indicators in 6 (n=27, 22%) studies and user feedback and responses in 5 (n=27, 19%) studies. Outcome-based definitions, linking engagement to specific user achievements or health outcomes, were noted in 3 (n=27, 11%) studies. The least common approach was game-like features, mentioned in just 1 (n=27, 4%) study.

To summarize how user engagement was operationalized across the included studies, we derived a working description based on those studies that provided an explicit definition of user engagement (18/59, 31%). This description is intended as a synthesis of observed operationalizations rather than a novel or normative definition:

> *User engagement is operationalized as the extent, frequency, and persistence of users' interactions with the app over time. It's typically quantified using system-recorded metrics, such as the number of active or logged-in days, total time spent in the app, frequency of accessing app modules, and the number or intensity of recorded actions (e.g., logging meals or viewing content). In some cases, these indicators are combined into composite scores, points totals, or weekly usage summaries, while other studies operationalize express engagement as the proportion of days with recorded activity. A small number of studies additionally incorporate the user's reflections on factors that support or hinder continued app usage.*

Discussion

Principal Findings

This scoping review provides a comprehensive overview of 59 mHealth nutrition studies, focusing on user-engagement strategies and metrics. Most apps targeted adults with overweight and/or obesity and were tested in RCTs and nonrandomized intervention studies. The included study samples comprised a predominantly female population. Study durations and sample sizes varied. Engagement was typically measured by the time and frequency of using specific functions and overall app use frequency. Push notifications and behavioral theory elements were the most frequently reported engagement strategies.

Push notifications and behavioral theory (elements) were commonly used to enhance engagement, aligning with findings from other digital health interventions, including mental and social well-being [33], diabetes management [34], weight loss interventions [35], or physical activity promotion [36]. While their widespread use suggests versatility across health contexts and adult populations, poorly timed or excessive push notifications can lead to user fatigue or disengagement, emphasizing the need to optimize the implementation of the engagement strategies. Notably, user engagement is known to be influenced by individual factors, including an individual's digital literacy and technology acceptance [37,38], health and demographic characteristics [39], or behavioral and psychological factors [40]. In addition, contextual factors also affect user engagement. For instance, sending notifications at specific times during weekends may yield better engagement rates [41,42]. Therefore, a standardized approach may not engage all users effectively, highlighting the need for adaptive frameworks that tailor engagement strategies to individual user characteristics and contextual factors.

Looking into the differences between commercial and research apps, the latter often used structured engagement tools such as push notifications and goal setting, likely reflecting a focus on behavioral theory and evidence-based practices. In contrast, commercial apps tended to emphasize user experience and simplicity since they are designed for broader appeal, with minimal features to reduce user burden. Both app types similarly adopted personalization and game-like elements, suggesting their broad value. For future studies and development, commercial apps could adopt more evidence-based strategies, while research apps might benefit from commercial design strengths.

The adoption of evidence-based strategies in mHealth apps is essential to ensure their effectiveness, equity, and sustainability in real-world settings. Evidence-based approaches integrate the best available scientific evidence with professional expertise and contextual considerations, thereby reducing reliance on ad hoc or intuition-driven decision-making [43]. In this review, engagement was primarily assessed using quantitative metrics, often referred to as "Little e" engagement, such as time spent on specific functions and overall app use, aligning with existing literature [25]. While standardized quantitative metrics aid comparability and benchmarking, they provide only a limited

XSL•FO
RenderX

view on user engagement, which also involves emotional, cognitive, and behavioral dimensions [44]. Most studies focused on short-term metrics, limiting insights into sustained engagement. Therefore, longitudinal assessments are needed to better understand long-term user interaction, while qualitative measures, such as interviews and in-app feedback, could provide deeper insights into user experience motivations, satisfaction, perceived usability, and barriers, and for explaining why certain engagement strategies succeed or fail [44-46]. Therefore, a mixed methods approach may better capture engagement dynamics and may help optimize future mHealth nutrition tools to better meet user needs and preferences. This variation between research and commercial apps underscores the need for standardized guidelines to ensure consistent quality and effectiveness in behavior change apps, an aim already addressed by several ongoing initiatives [47].

This review highlights gaps in how user engagement is conceptualized, defined, and measured. Only 31% (n=18) of studies explicitly defined engagement, limiting comparability and generalizability, an issue noted in multiple previous reviews on the topic [48,49]. Despite existing conceptual frameworks [44,45], their adoption remains limited, leading to high heterogeneity in engagement metrics [50]. This inconsistency hinders the understanding of real-world challenges, such as low uptake of nutrition apps, and impedes the development of standardized measures. These measures are crucial for comparability, establishing benchmarks, and deriving robust conclusions on effective engagement strategies. Future research should prioritize implementing established frameworks and developing standardized metrics to enhance cross-study synthesis and support evidence-based best practices for mHealth nutrition tools.

Limitations

Despite the broad scope of this review, several limitations must be acknowledged. The diversity of the included studies, both in design and in the involved engagement strategies, made direct comparisons of effectiveness challenging. Additionally, the findings highlight the need for further investigation into how engagement strategies can be tailored to individual user characteristics and contextual factors. The included studies use a wide range of methodologies, including RCTs, pilot RCTs, secondary analyses of RCTs, observational cohort studies, and mixed methods research, such as feasibility and acceptability studies. There is also significant variability in sample sizes, ranging from as few as 12 participants in the study by Robinson et al [51] to more than a million (1,011,008) in the study by Serrano et al [13]. Moreover, while this review analyses engagement among users who accessed the included interventions, it is crucial to acknowledge that nonuse may reflect structural and contextual barriers rather than a lack of interest. Several barriers, including limited access to digital infrastructure, low digital literacy, and socioeconomic constraints, may prevent certain populations from engaging with nutrition apps. The latter could affect both the extent and quality of recorded engagement, potentially leading to underestimation or misrepresentation of true user behavior. Thus, it is important to note that our findings represent the experiences of the included active users.

A significant limitation observed across the reviewed studies is the lack of clear, standardized definitions for essential terms such as “engagement,” which is often undefined, as well as broad concepts such as “acceptability” and “feasibility.” Furthermore, few studies include reflective discussions on how to improve user engagement, despite its critical role in the success of interventions [44]. The absence of standardized reporting metrics further complicates the synthesis of findings, limiting the ability to draw meaningful comparisons across studies. It should be noted that the counting of studies reporting positive, negative, or neutral findings across engagement-metric categories was purely descriptive. This approach does not allow any causal, comparative, or predictive inferences regarding the effectiveness of specific engagement metrics, and the observed distribution reflects reporting patterns in the included studies rather than true differences in impact.

Finally, another limitation of this study was the uneven distribution between commercial and research-based nutrition apps across the reviewed articles, with the latter constituting a majority (n=44, 75%). Whereas commercial apps typically prioritize consumer design principles and advanced user interface aesthetics, research apps may focus more on evidence - based content, and, in some cases, compensate users to sustain use. Therefore, the insights gained from this review can be further explored in future studies that incorporate a broader range of commercial apps.

Future Directions

Future research should focus on more rigorous evaluations, potentially through RCTs or long-term observational studies, to establish clearer insights into the efficacy of these strategies.

Co-design approaches in the development of nutrition and health apps have been shown to enhance user engagement and improve intervention effectiveness. Many studies have used participatory methods, involving stakeholders in the design process to ensure that the app aligns with user needs and preferences. Some commonly used co-design techniques include participatory research approaches, stakeholder engagement, and iterative design workshops. According to Davis et al [52] and Jackson et al [53], the benefits of co-design include higher user satisfaction, improved app usability, and a greater likelihood of meeting users’ needs, particularly by addressing cultural considerations. While co-design in mHealth is still an emerging field, it has demonstrated effectiveness across various health domains, including disease management and health promotion. Further research is needed to fully implement co-design processes and evaluate their impact on nutrition interventions.

Although many studies report use patterns, such as log in frequency and session duration, few provide insights into how users engage with specific behavioral components of nutrition apps. This lack of detail makes it difficult to identify which features drive engagement and behavior change. Additionally, engagement metrics vary widely across studies, making comparisons and synthesis challenging. For example, “active use” may be defined as a single session in one study but as daily interactions in another, complicating efforts to draw meaningful conclusions.

XSL•FO
RenderX

Moreover, most studies rely on static, one-size-fits-all engagement metrics that do not account for changes in user behavior over time. Without adaptive metrics, where use patterns dynamically inform engagement strategies, apps risk missing opportunities for personalization and sustained engagement. This static approach limits the potential for iterative improvements and real-time feedback loops that could optimize user experience and long-term outcomes. A recent methodological framework has been developed to analyze and model the dynamic processes involved in participant engagement with mHealth apps and identify factors that foster or undermine engagement in an individual's respective context [54,55]. However, the uptake of this type of approach is still very limited. Future research on user engagement may find this type of approach particularly valuable to understand how different contextual factors impact engagement as a dynamic process.

Conclusions

This review provides a broad overview of the various strategies used to enhance user engagement in nutrition apps. mHealth nutrition apps predominantly target adults with overweight or obesity and are often evaluated through controlled trials with female-majority samples. Despite variability in study design, engagement is commonly tracked via use patterns, with push notifications and behavioral theory–based strategies emerging as key tools to enhance user involvement. By understanding and implementing these strategies, we can design more engaging and effective digital tools to promote healthy dietary behaviors and ultimately contribute to improved public health outcomes.

Funding

The authors declared that no financial support was received for this work.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

MFV, TB, and EB-B designed, conceptualized, and defined the methods in this study. All the authors contributed to the screening of the studies. MFV drafted the manuscript, and all the authors reviewed the final manuscript. All authors contributed to screening abstracts and full texts. MFV and ZvdH conducted the data extraction.

Conflicts of Interest

At the time of publication, MFV, ACF, EAS, and FR were employees of Nestlé Research, which is part of the Société des Produits Nestlé S.A.

Multimedia Appendix 1
Search strategy, PICOS, user-engagement definitions, list of included studies, basic study characteristics, and characteristics of the apps.
[DOCX File , 226 KB - mhealth_v14i1e82276_app1.docx]

Multimedia Appendix 2
Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist.
[DOCX File , 87 KB - mhealth_v14i1e82276_app2.docx]

References

1. Noncommunicable diseases. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2025-04-03]
2. Gregg EW, Shaw JE. Global health effects of overweight and obesity. N Engl J Med 2017 Jul 06;377(1):80-81 [FREE Full text] [doi: 10.1056/NEJMe1706095] [Medline: 28604226]
3. Abeltino A, Riente A, Bianchetti G, Serantoni C, De Spirito M, Capezzone S, et al. Digital applications for diet monitoring, planning, and precision nutrition for citizens and professionals: a state of the art. Nutr Rev 2025 Feb 01;83(2):e574-e601. [doi: 10.1093/nutrit/nuae035] [Medline: 38722240]
4. Ruyobeza B, Grobbelaar SS, Botha A. Hurdles to developing and scaling remote patients' health management tools and systems: a scoping review. Syst Rev 2022 Aug 30;11(1):179 [FREE Full text] [doi: 10.1186/s13643-022-02033-z] [Medline: 36042505]
5. Digital health trends 2024. IQVIA. 2024 Dec 12. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports-and-publications/reports/digital-health-trends-2024 [accessed 2026-04-21]

XSL•FO
RenderX

6. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
7. Khazen W, Jeanne JF, Demaretz L, Schäfer F, Fagherazzi G. Rethinking the use of mobile apps for dietary assessment in medical research. J Med Internet Res 2020 Jun 18;22(6):e15619 [FREE Full text] [doi: 10.2196/15619] [Medline: 32554383]
8. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
9. Cole-Lewis H, Ezeanochie N, Turgiss J. Understanding health behavior technology engagement: pathway to measuring digital behavior change interventions. JMIR Form Res 2019 Oct 10;3(4):e14052 [FREE Full text] [doi: 10.2196/14052] [Medline: 31603427]
10. Zarnowiecki D, Mauch CE, Middleton G, Matwiejczyk L, Watson WL, Dibbs J, et al. A systematic evaluation of digital nutrition promotion websites and apps for supporting parents to influence children's nutrition. Int J Behav Nutr Phys Act 2020 Feb 10;17(1):17 [FREE Full text] [doi: 10.1186/s12966-020-0915-1] [Medline: 32041640]
11. Kohl LF, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
12. Grady A, Pearson N, Lamont H, Leigh L, Wolfenden L, Barnes C, et al. The effectiveness of strategies to improve user engagement with digital health interventions targeting nutrition, physical activity, and overweight and obesity: systematic review and meta-analysis. J Med Internet Res 2023 Dec 19;25:e47987 [FREE Full text] [doi: 10.2196/47987] [Medline: 38113062]
13. Serrano KJ, Coa KI, Yu M, Wolff-Hughes DL, Atienza AA. Characterizing user engagement with health app data: a data mining approach. Transl Behav Med 2017 Jun;7(2):277-285 [FREE Full text] [doi: 10.1007/s13142-017-0508-y] [Medline: 28616846]
14. Nour M, Chen J, Allman-Farinelli M. Young adults' engagement with a self-monitoring app for vegetable intake and the impact of social media and gamification: feasibility study. JMIR Form Res 2019 May 10;3(2):e13324 [FREE Full text] [doi: 10.2196/13324] [Medline: 31094322]
15. Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014 Feb 14;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
16. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256 [FREE Full text] [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
17. Cavallo DN, Tate DF, Ward DS, DeVellis RF, Thayer LM, Ammerman AS. Social support for physical activity-role of Facebook with and without structured intervention. Transl Behav Med 2014 Dec;4(4):346-354 [FREE Full text] [doi: 10.1007/s13142-014-0269-9] [Medline: 25584083]
18. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]
19. Gan DZ, McGillivray L, Larsen ME, Christensen H, Torok M. Technology-supported strategies for promoting user engagement with digital mental health interventions: a systematic review. Digit Health 2022 Jun 01;8:20552076221098268 [FREE Full text] [doi: 10.1177/20552076221098268] [Medline: 35677785]
20. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016 Jan 08;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
21. Elkes J, Cro S, Batchelor R, O'Connor S, Yu LM, Bell L, et al. User engagement in clinical trials of digital mental health interventions: a systematic review. BMC Med Res Methodol 2024 Aug 24;24(1):184 [FREE Full text] [doi: 10.1186/s12874-024-02308-0] [Medline: 39182064]
22. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
23. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
24. Abqari U, van 't Noordende AT, Richardus JH, Isfandiari MA, Korfage IJ. Strategies to promote the use of online health applications for early detection and raising awareness of chronic diseases among members of the general public: a systematic literature review. Int J Med Inform 2022 Jun;162:104737 [FREE Full text] [doi: 10.1016/j.ijmedinf.2022.104737] [Medline: 35358894]

XSL•FO
RenderX

25. Eaton C, Vallejo N, McDonald X, Wu J, Rodríguez R, Muthusamy N, et al. User engagement with mHealth interventions to promote treatment adherence and self-management in people with chronic health conditions: systematic review. J Med Internet Res 2024 Oct 24;26:e50508 [FREE Full text] [doi: 10.2196/50508] [Medline: 39316431]
26. Mclaughlin M, Delaney T, Hall A, Byaruhanga J, Mackie P, Grady A, et al. Associations between digital health intervention engagement, physical activity, and sedentary behavior: systematic review and meta-analysis. J Med Internet Res 2021 Feb 19;23(2):e23180 [FREE Full text] [doi: 10.2196/23180] [Medline: 33605897]
27. Aguiar M, Trujillo M, Chaves D, Álvarez R, Epelde G. mHealth apps using behavior change techniques to self-report data: systematic review. JMIR Mhealth Uhealth 2022 Sep 09;10(9):e33247 [FREE Full text] [doi: 10.2196/33247] [Medline: 36083606]
28. Schoeppe S, Alley S, Rebar AL, Hayman M, Bray NA, Van Lippevelde W, et al. Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: a review of quality, features and behaviour change techniques. Int J Behav Nutr Phys Act 2017 Jun 24;14(1):83 [FREE Full text] [doi: 10.1186/s12966-017-0538-3] [Medline: 28646889]
29. van Acker J, Maenhout L, Compernolle S. Older adults' user engagement with mobile health: a systematic review of qualitative and mixed-methods studies. Innov Aging 2023;7(2):igad007 [FREE Full text] [doi: 10.1093/geroni/igad007] [Medline: 37007638]
30. Zenun Franco R, Fallaize R, Weech M, Hwang F, Lovegrove JA. Effectiveness of web-based personalized nutrition advice for adults using the eNutri web app: evidence from the EatWellUK randomized controlled trial. J Med Internet Res 2022 Apr 25;24(4):e29088 [FREE Full text] [doi: 10.2196/29088] [Medline: 35468093]
31. Lister C, West JH, Cannon B, Sax T, Brodegard D. Just a fad? Gamification in health and fitness apps. JMIR Serious Games 2014 Aug 04;2(2):e9 [FREE Full text] [doi: 10.2196/games.3413] [Medline: 25654660]
32. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
33. Pretolesi D, Motnikar L, Till B, Uhl J. Notifying users: customisation preferences for notifications in health and well-being applications. In: Meschtscherjakov A, Midden C, Ham J, editors. Persuasive Technology. Cham, Switzerland: Springer; 2023.
34. Narindrarangkura P, Ye Q, Boren SA, Khan U, Simoes EJ, Kim MS. Analysis of healthy coping feedback messages from diabetes mobile apps: validation against an evidence-based framework. J Diabetes Sci Technol 2023 Jan;17(1):152-162 [FREE Full text] [doi: 10.1177/19322968211043534] [Medline: 34530644]
35. Hales S, Turner-McGrievy G, Fahim A, Freix A, Wilcox S, Davis RE, et al. A mixed-methods approach to the development, refinement, and pilot testing of social networks for improving healthy behaviors. JMIR Hum Factors 2016 Feb 12;3(1):e8 [FREE Full text] [doi: 10.2196/humanfactors.4512] [Medline: 27026535]
36. Schwarz A, Winkens LH, de Vet E, Ossendrijver D, Bouwsema K, Simons M. Design features associated with engagement in mobile health physical activity interventions among youth: systematic review of qualitative and quantitative studies. JMIR Mhealth Uhealth 2023 Mar 06;11:e40898 [FREE Full text] [doi: 10.2196/40898] [Medline: 36877551]
37. Nunes A, Limpo T, Castro SL. Acceptance of mobile health applications: examining key determinants and moderators. Front Psychol 2019 Dec 10;10:2791 [FREE Full text] [doi: 10.3389/fpsyg.2019.02791] [Medline: 31920836]
38. Fan S, Jain RC, Kankanhalli MS. A comprehensive picture of factors affecting user willingness to use mobile health applications. ACM Trans Comput Healthcare 2024 Jan 13;5(1):1-31. [doi: 10.1145/3626962]
39. Simões de Almeida R, Marques A. User engagement in mobile apps for people with schizophrenia: a scoping review. Front Digit Health 2022;4:1023592 [FREE Full text] [doi: 10.3389/fdgth.2022.1023592] [Medline: 36703941]
40. Steinberg JR, Yeh C, Jackson J, Saber R, Niznik CM, Leziak K, et al. Optimizing engagement in an mHealth intervention for diabetes support during pregnancy: the role of baseline patient health and behavioral characteristics. J Diabetes Sci Technol 2022 Nov;16(6):1466-1472 [FREE Full text] [doi: 10.1177/19322968211035441] [Medline: 34423677]
41. Bidargaddi N, Almirall D, Murphy S, Nahum-Shani I, Kovalcik M, Pituch T, et al. To prompt or not to prompt? A microrandomized trial of time-varying push notifications to increase proximal engagement with a mobile health app. JMIR Mhealth Uhealth 2018 Nov 29;6(11):e10123 [FREE Full text] [doi: 10.2196/10123] [Medline: 30497999]
42. Bidargaddi N, Pituch T, Maaieh H, Short C, Strecher V. Predicting which type of push notification content motivates users to engage in a self-monitoring app. Prev Med Rep 2018 Sep;11:267-273 [FREE Full text] [doi: 10.1016/j.pmedr.2018.07.004] [Medline: 30109172]
43. Brownson RC, Fielding JE, Maylahn CM. Evidence-based public health: a fundamental concept for public health practice. Annu Rev Public Health 2009 Apr 01;30:175-201. [doi: 10.1146/annurev.publhealth.031308.100134] [Medline: 19296775]
44. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
45. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]

XSL•FO
RenderX

46. Abusamhadana GA, Elias NF, Mukhtar M, Mokhtar UA. Items for measuring user engagement success in information systems development. In: Proceedings of the 2019 International Conference on Electrical Engineering and Informatics. 2019 Presented at: ICEEI 2019; July 9-10, 2019; Bandung, Indonesia. [doi: 10.1109/iceei47359.2019.8988816]
47. Australian mHealth apps assessment framework: options report. Australasian Institute of Digital Health. 2024. URL: https://digitalhealth.org.au/wp-content/uploads/2024/06/AIDH_mHealth-Framework-Options.pdf [accessed 2026-04-21]
48. Flaherty SJ, McCarthy M, Collins AM, McCafferty C, McAuliffe FM. Exploring engagement with health apps: the emerging importance of situational involvement and individual characteristics. Eur J Mark 2020 Dec;55(13):122-147. [doi: 10.1108/ejm-06-2019-0531]
49. König LM, Attig C, Franke T, Renner B. Barriers to and facilitators for using nutrition apps: systematic review and conceptual framework. JMIR Mhealth Uhealth 2021 Apr 01;9(6):e20037 [FREE Full text] [doi: 10.2196/20037] [Medline: 34254938]
50. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. Psychiatr Serv 2019 Jul 01;70(7):538-544 [FREE Full text] [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]
51. Robinson E, Higgs S, Daley AJ, Jolly K, Lycett D, Lewis A, et al. Development and feasibility testing of a smart phone based attentive eating intervention. BMC Public Health 2013 Jul 09;13:639 [FREE Full text] [doi: 10.1186/1471-2458-13-639] [Medline: 23837771]
52. Davis SR, Peters D, Calvo RA, Sawyer SM, Foster JM, Smith L. "Kiss myAsthma": using a participatory design approach to develop a self-management app with young people with asthma. J Asthma 2018 Sep;55(9):1018-1027 [FREE Full text] [doi: 10.1080/02770903.2017.1388391] [Medline: 29064746]
53. Jackson DN, Sehgal N, Baur C. Benefits of mHealth co-design for African American and Hispanic adults: multi-method participatory research for a health information app. JMIR Form Res 2022 Mar 09;6(3):e26764 [FREE Full text] [doi: 10.2196/26764] [Medline: 35262496]
54. De La Torre SA, El Mistiri M, Hekler E, Klasnja P, Marlin B, Pavel M, et al. Modeling engagement with a digital behavior change intervention (HeartSteps II): an exploratory system identification approach. J Biomed Inform 2024 Oct;158:104721 [FREE Full text] [doi: 10.1016/j.jbi.2024.104721] [Medline: 39265816]
55. Hekler EB, Rivera DE, Martin CA, Phatak SS, Freigoun MT, Korinek E, et al. Tutorial for using control systems engineering to optimize adaptive mobile health interventions. J Med Internet Res 2018 Jun 28;20(6):e214 [FREE Full text] [doi: 10.2196/jmir.8622] [Medline: 29954725]

Abbreviations

mHealth: mobile health
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RCT: randomized controlled trial

Edited by L Buis; submitted 28.Aug.2025; peer-reviewed by M Aktas; comments to author 15.Dec.2025; revised version received 24.Mar.2026; accepted 25.Mar.2026; published 12.May.2026.

Please cite as:
Vasiloglou MF, van der Heijden Z, Scuccimarra EA, Conde Freniche A, Lucassen DA, de Vlieger N, Ronga F, Brouwer-Brolsma E, Bucher T
Key Features of Engagement Strategies in Nutrition Apps for Adults: Scoping Review
JMIR Mhealth Uhealth 2026;14:e82276
URL: https://mhealth.jmir.org/2026/1/e82276
doi:10.2196/82276
PMID:

©Maria F Vasiloglou, Zoë van der Heijden, Eric Antoine Scuccimarra, Alberto Conde Freniche, Desiree A Lucassen, Nienke de Vlieger, Frédéric Ronga, Elske Brouwer-Brolsma, Tamara Bucher. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 12.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Cost-Effectiveness of Virtual Emergency Care Models: Systematic Review

Ravi Shankar[1], PhD; Linda Wang[2], PhD; Soon Hoe Ho[2], MPH; Mei Fong Liew[3,4,5], MBBS; Satya Pavan Kumar Gollamudi[5,6], MBBS; Tze Chin Wong[7], BEng; Serene Wong[3,4,5], MBBS

[1]Clinical Research & Innovation Office, Tan Tock Seng Hospital, National Healthcare Group, Singapore, Singapore

[2]Fast and Chronic Programmes, Alexandra Hospital, Singapore, Singapore

[3]Division of Respiratory and Critical Care Medicine, Alexandra Hospital, National University Health System, Singapore, Singapore

[4]Division of Respiratory and Critical Care Medicine, National University Hospital, National University Health System, Singapore, Singapore

[5]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[6]Division of Advanced Internal Medicine, Alexandra Hospital, University Health System, Singapore, Singapore

[7]Ambulatory Operations, Alexandra Hospital, Singapore, Singapore

Corresponding Author:
Ravi Shankar, PhD
Clinical Research & Innovation Office, Tan Tock Seng Hospital, National Healthcare Group
11 Jalan Tan Tock Seng
Singapore, 308433
Singapore
Phone: 65 6256 6011
Email: ravisr.srivastava@gmail.com

Abstract

Background: Virtual care technologies have rapidly expanded in emergency medicine, particularly following the COVID-19 pandemic. However, comprehensive economic evaluations of their cost-effectiveness remain fragmented across different clinical applications and health care settings, creating uncertainty for policymakers and health care administrators considering implementation.

Objective: This study aimed to systematically review and synthesize evidence on the cost-effectiveness of virtual emergency care models compared to traditional in-person emergency care across diverse clinical conditions, populations, and health care settings.

Methods: We conducted a systematic review following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, searching 8 electronic databases (PubMed, Embase, Scopus, Web of Science, CINAHL, Cochrane Library, MEDLINE, and PsycINFO) from inception to February 2025. We included full economic evaluations comparing virtual emergency care interventions with usual care. Two reviewers independently screened studies, extracted data, and assessed quality using the Drummond checklist and Consensus Health Economic Criteria (CHEC) list. Evidence certainty was evaluated using Grading of Recommendations Assessment, Development, and Evaluation (GRADE) methodology. Given heterogeneity in interventions and methods, we conducted a narrative synthesis by virtual care modality and clinical application.

Results: From 5817 identified references, 13 studies met inclusion criteria, representing diverse virtual care modalities across 6 countries (United States, Australia, Italy, Canada, Haiti, and Belgium). All included studies reported favorable economic outcomes for virtual emergency care. Video consultation was the most common modality (11/13 studies), achieving 31% to 73% reduction in patient transfers and cost savings of US $73 (AUD $105) to US $5118 per encounter. A total of 6 (46%) studies found virtual care to be dominant (both less costly and more effective). Incremental cost-effectiveness ratios ranged from US $1273 (€990) to US $108,363 per quality-adjusted life year, with most below accepted willingness-to-pay thresholds. Transfer avoidance was the primary economic driver, particularly in rural settings. Quality assessment revealed high methodological rigor (mean Drummond score 92.3%, SD 6.0%; mean CHEC score 95%, SD 4.2%). Using GRADE, evidence certainty was rated high for cost-effectiveness, moderate for transfer reduction and quality of life improvements, and low for emergency department length of stay and mortality benefits.

Conclusions: Virtual emergency care demonstrates strong and consistent cost-effectiveness across diverse clinical conditions, populations, and health care settings. The evidence particularly supports implementation for stroke care, pediatric emergencies,

XSL•FO
RenderX

and rural/remote populations where transfer avoidance drives substantial economic benefits. All evaluated modalities achieved favorable economic outcomes, suggesting technology should match context rather than maximize sophistication. These findings provide robust economic justification for expanding virtual emergency care access and removing regulatory barriers. As health care systems face mounting pressures from aging populations, workforce shortages, and budget constraints, virtual emergency care offers a proven strategy for improving access and quality while reducing costs.

Trial Registration: PROSPERO CRD42025648218; https://www.crd.york.ac.uk/PROSPERO/view/CRD42025648218

(JMIR Mhealth Uhealth 2026;14:e82143) doi:10.2196/82143

KEYWORDS

cost-effectiveness; economic evaluation; emergency care; telemedicine; virtual care

Introduction

The Challenge of Emergency Department Pressures

Emergency departments (EDs) worldwide face unprecedented challenges, including overcrowding, workforce shortages, and geographic disparities in access to specialist care [1,2]. These pressures have intensified following the COVID-19 pandemic, which accelerated the adoption of virtual care technologies while highlighting systemic vulnerabilities in emergency care delivery [3]. These issues not only hinder operational efficiency but also lead to delays in care, reduced patient and staff satisfaction, and increased adverse clinical outcomes [4,5].

Virtual Emergency Care Models

For the purposes of this review, we define "virtual emergency care" as any technology-mediated health care interaction that replaces or augments in-person emergency care delivery, encompassing synchronous telemedicine (real-time video or audio consultations), asynchronous telemedicine (store-and-forward transmission of clinical data), remote monitoring, and digital triage tools. We use "virtual emergency care" as the overarching term, with specific modalities (eg, telestroke, telepsychiatry, and telephone consultation) specified where relevant.

Virtual emergency care models have emerged as a promising approach for addressing these challenges by potentially expanding access, improving efficiency, and containing costs [4-6]. However, these technologies also raise concerns, including the risk of diagnostic error without physical examination, potential mistriage, technology dependence and failure modes, patient safety considerations, and the possibility of exacerbating health disparities for populations with limited digital access or literacy [6-8]. Patient-to-provider telemedicine enables individuals to engage with emergency clinicians via synchronous (eg, video) or asynchronous (eg, text) methods for clinical assessment without the need for physical ED visits [9,10]. Provider-to-provider teleconsultations, such as telestroke or teletrauma, extend specialist expertise to individuals in underserved rural and remote areas where access to specialized emergency medical services is often limited or nonexistent [11]. Digital triage tools include symptom checkers, decision support algorithms, and chatbot-based platforms that guide disposition and urgency based on self-reported symptoms [12]. Remote monitoring systems, such as wearables or sensor-based technologies, enable continuous physiologic surveillance and timely escalation for conditions such as sepsis or acute decompensated heart failure [13,14].

Economic Implications of Virtual Emergency Care

The economic implications of virtual emergency care are particularly significant given that emergency services account for a substantial proportion of health care expenditures. In the United States alone, ED visits generate over US $150 billion in annual costs, with rural and underserved populations facing additional burdens from unnecessary transfers and delayed care [15]. While the clinical benefits of virtual emergency care have been increasingly documented, the economic case for implementation remains a critical consideration for health care decision-makers. Economic evaluation provides a framework for comparing the costs and outcomes of virtual care models against traditional approaches, informing resource allocation decisions in resource-constrained health care systems. While virtual care technologies have demonstrated clinical efficacy in various emergency applications, their economic value proposition remains incompletely understood, with cost-effectiveness evidence scattered across different clinical domains and implementation contexts.

Gaps in Existing Evidence

Previous systematic reviews have predominantly emphasized clinical effectiveness, patient satisfaction, or implementation feasibility [16,17]. Some other reviews have examined telemedicine applications in specific emergency conditions or settings [7,9,18-20], but none have comprehensively synthesized the economic evidence across the full spectrum of virtual emergency care modalities. This gap is problematic for health care decision-makers who must justify infrastructure investments and policy changes based on robust economic evidence. Furthermore, the heterogeneity of virtual care interventions, from simple telephone consultations to sophisticated hub-and-spoke telestroke networks, necessitates a nuanced analysis of which modalities provide optimal value in different contexts.

The urgency of understanding virtual emergency care economics has intensified as health care systems globally face budget constraints while demanding improved access and quality. Policymakers require evidence not only of clinical effectiveness but also of economic sustainability and return on investment (ROI). Additionally, the rapid evolution of virtual care technologies and payment models necessitates a current synthesis of available evidence to guide implementation decisions.

XSL•FO
RenderX

Study Aim

This systematic review addresses these knowledge gaps by comprehensively evaluating the cost-effectiveness of virtual emergency care models compared to traditional in-person care. We examine economic outcomes across different virtual care modalities, clinical applications, and implementation settings while identifying factors that influence cost-effectiveness. Our analysis aims to provide actionable evidence for health care administrators, policymakers, and clinicians considering virtual emergency care implementation.

Specifically, we sought to answer the following research question: "In patients requiring emergency medical care (population), is virtual care technology including telemedicine, remote monitoring, and digital triage (intervention) cost-effective compared to traditional in-person emergency care (comparator), as measured by incremental cost-effectiveness ratios (ICERs), cost savings, and quality-adjusted life years (QALYs; outcomes), across various health care settings and time horizons (setting/time)?"

Methods

Study Design and Protocol

We conducted this systematic review following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Multimedia Appendix 1) [21]. The review protocol was registered prospectively in PROSPERO (International Prospective Register of Systematic Reviews; registration number CRD42025648218). No ethical approval was required as this study involved analysis of published literature only.

Search Strategy

A comprehensive search strategy was developed to identify all relevant economic evaluations of virtual emergency care. We searched 8 electronic databases from inception to February 2025: PubMed (n=2084), Embase (n=1827), Scopus (n=916), Web of Science (n=645), CINAHL (n=199), The Cochrane Library (n=82), MEDLINE (n=30), and PsycINFO (n=3).

The search strategy combined 3 concept blocks using Boolean operators: (1) virtual care terms ("telemedicine," "telehealth," "virtual care," "remote consultation," "digital health," "eHealth," and "mHealth"); (2) emergency care terms ("emergency," "urgent care," "acute care," "emergency department," "emergency room," "trauma," and "critical care"); and (3) economic evaluation terms ("cost-effectiveness," "cost-benefit," "cost-utility," "economic evaluation," "cost analysis," "ROI," "ICER," "QALY," and "cost-minimization"). The search strategy was refined iteratively with input from an information specialist. While we primarily used free-text terms to maximize sensitivity across databases with varying indexing systems, we acknowledge that the absence of MeSH terms for economic evaluation concepts (eg, "cost-benefit analysis" [MeSH]) and emergency-specific controlled vocabulary may have reduced search precision. During the screening process, one eligible study was initially misclassified during automated deduplication in Covidence due to similar authorship and title to another included study [22,23]. This study was subsequently identified through manual verification, retrieved, and included in the final analysis. Our broad digital health terms were designed to capture the rapidly evolving terminology in this field, though this approach may have introduced some irrelevant results during initial screening. We recognize that certain virtual care terms (eg, "teleconsultation," "teleurgency," and "store-and-forward") may not have been captured by our search terms. No language or date restrictions were applied. The full search strategy for each database is available in Multimedia Appendix 2.

Additionally, we performed citation searching of included studies (n=14) and searched gray literature sources (n=17) to identify unpublished economic evaluations, government reports, and health technology assessments.

Eligibility Criteria

Studies were included if they met all predefined eligibility criteria. Eligible studies focused on patients of any age presenting to EDs or requiring emergency medical care. The intervention of interest was any form of virtual care technology applied to emergency care delivery, encompassing synchronous or asynchronous telemedicine, remote monitoring, digital triage tools, or mobile health apps. Prehospital emergency care interventions, including ambulance-based telemedicine, were eligible for inclusion provided they met all other criteria for full economic evaluation. Stand-alone digital triage tools without clinician involvement (eg, automated symptom checkers used independently by patients without subsequent clinician assessment) were excluded, as these represent a distinct category of intervention with different economic evaluation considerations. This distinction was made a priori to maintain a focus on clinician-mediated virtual care models. Studies were required to include a comparator group involving traditional in-person emergency care or usual care practices. Only full economic evaluations were considered, specifically those that incorporated both cost and outcome data through cost-effectiveness, cost-utility, cost-benefit, or cost-minimization analyses. Relevant outcomes included economic measures such as cost-effectiveness ratios and other forms of comparative economic analysis.

Studies were excluded if they reported only costs or only outcomes without integrating both into an economic evaluation. Abstracts without accessible full texts, even after attempts to contact authors, were not considered. Additional exclusions applied to studies with unsuitable designs such as reviews, editorials, commentaries, or protocols; studies focusing on nonemergency settings or conditions (such as routine outpatient care or scheduled procedures); those lacking a clear economic evaluation framework; and those not involving direct patient care, such as research limited to provider education or administrative systems.

Study Selection

All identified references were imported into Covidence systematic review software for deduplication and screening. Two reviewers independently screened the titles and abstracts based on predefined eligibility criteria. Full texts of studies deemed potentially eligible were then retrieved and assessed independently by the same reviewers. Any disagreements were

XSL•FO
RenderX

resolved through discussion and, if needed, with input from a third reviewer. Interrater agreement was found to be strong across both stages of screening. The entire screening process, including reasons for exclusion, was documented in accordance with PRISMA guidelines. During quality checks, 1 study that had been inadvertently excluded during automated deduplication was identified. This study, which shared the same lead author and similar title as another included study, was manually retrieved, assessed for eligibility, and subsequently included in the final synthesis.

Data Extraction

A standardized data extraction form was developed and pilot-tested to ensure consistency, with initial testing conducted on a small sample of studies. Two reviewers independently extracted data from all included studies across several key domains. Study characteristics were recorded, including the first author, publication year, country, study design, and funding source. Population details such as sample size, age, sex distribution, emergency conditions, triage acuity levels, and comorbidities were documented. Information on intervention characteristics included the type of virtual care model, technology platform used, provider types involved, frequency and duration of interventions, and how the virtual model related to in-person care. Comparator details focused on usual care practices and the associated providers and resources. Economic evaluation methods were captured, including the type of analysis, analytic perspective, time horizon, discount rate, currency, and price year. Outcome measures covered effectiveness indicators such as QALYs, disability-adjusted life years (DALYs), and clinical outcomes, as well as cost-related metrics including direct, indirect, and societal costs, and ICERs. Data on the types of sensitivity analyses conducted and key results were also extracted. Finally, implementation factors, including any barriers and facilitators identified, were noted. Any discrepancies in data extraction were resolved through discussion and careful reexamination of the original source materials.

Quality Assessment

Methodological quality of included studies was assessed using 2 validated checklists: the Drummond checklist for economic evaluations [24] and the Consensus Health Economic Criteria (CHEC) list [25]. The combined use of both checklists is recommended by the Cochrane Handbook for Systematic Reviews of Interventions for appraising the methodological quality of full economic evaluations [26], as each tool addresses complementary quality domains. The Drummond checklist (10 items) focuses on the completeness and rigor of reporting in economic evaluations, including research question clarity, study design appropriateness, cost identification, incremental analysis, and uncertainty assessment [27]. The CHEC list (19 items) emphasizes methodological quality and reporting transparency, covering domains such as appropriate study population, time horizon, discount rate, outcomes measurement, and generalizability discussion [25]. This dual-checklist approach has been used in previous systematic reviews of economic evaluations to provide a more comprehensive quality assessment, as noted by Watts and Li [28] in their meta-review of checklist use in 346 systematic reviews of economic evaluations. Where the 2 tools assessed overlapping domains (eg, research question clarity and cost identification), we prioritized the more detailed tool's assessment for that domain.

Two reviewers independently rated each study, with disagreements resolved through consensus. We acknowledge that presenting quality scores as percentages has limitations, as these checklists were not originally designed to produce aggregate numerical scores, and individual checklist items may vary in importance [29]. The percentage scores should therefore be interpreted as broad indicators of methodological quality rather than precise quantitative measures. Item-level quality assessment results are presented in Multimedia Appendix 3 to allow readers to evaluate specific methodological strengths and weaknesses across studies.

Evidence Synthesis

We adapted the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach to assess the certainty of evidence for key outcomes [30], recognizing that GRADE was originally developed for clinical interventions rather than economic evaluations. Our adaptation followed the framework described by Brunetti et al [31], which provides guidance on considering resource use and rating the quality of economic evidence within the GRADE system. We additionally drew on the conceptual GRADE approach for assessing the certainty of modeled evidence outlined by Brozek et al [32], which confirmed that the standard GRADE domains apply to evidence from modeling studies, including those used in health economic evaluations.

Specifically, we assessed risk of bias (considering study design, model assumptions, and potential conflicts of interest), inconsistency (direction and magnitude of findings across studies), indirectness (applicability of study populations and settings to the review question), imprecision (confidence in effect estimates), and publication bias. For upgrading considerations, we evaluated the magnitude of effects and the consistency of the direction of findings across studies. The reference to dose-response relationships that appeared in the original protocol has been removed, as this criterion is not directly applicable to economic evaluations in this context [31]. Given the heterogeneity in interventions, populations, and economic evaluation methods, we conducted a narrative synthesis organized by virtual care modality and clinical application.

Subgroup and Sensitivity Analyses

We planned a priori subgroup analyses to explore potential sources of heterogeneity across included studies. Subgroup analyses examined variations by clinical condition, virtual care modality, geographic setting, patient age group, and economic perspective. These subgroup analyses are distinct from the sensitivity analyses conducted within individual included studies, which tested the robustness of each study's findings to variations in key parameters and assumptions. We report the results of within-study sensitivity analyses as part of our data extraction and discuss their implications for the certainty of findings.

XSL•FO
RenderX

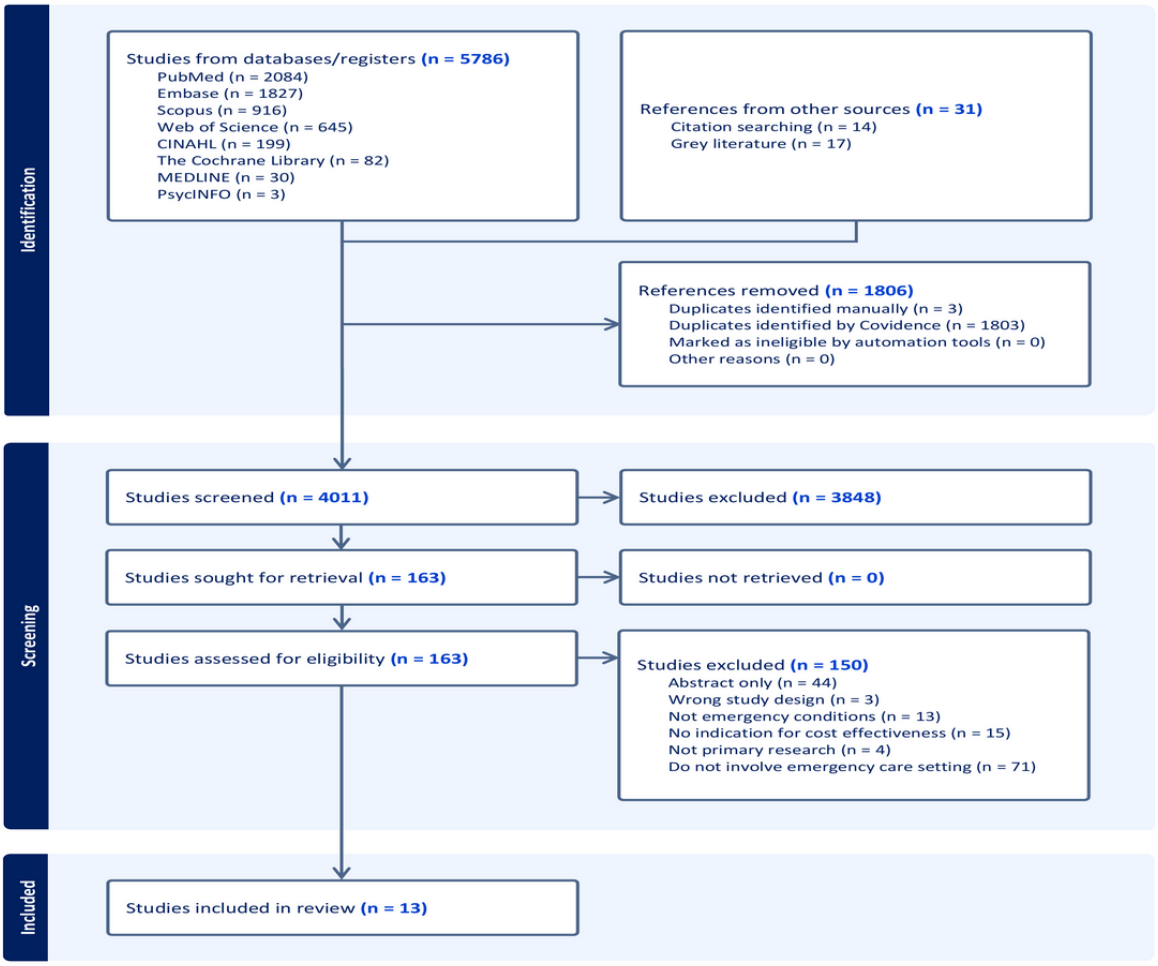
Costs were reported in the original currencies and price years as presented in the included studies. We did not convert costs to a common currency or adjust for inflation, as the primary objective was to assess the direction and magnitude of cost-effectiveness findings within each study's context rather than to pool absolute cost figures. This approach is consistent with narrative synthesis methodology for heterogeneous economic evaluations. However, we acknowledge that this limits direct cross-study cost comparisons. Where available, we report the original publication year for each study to facilitate reader interpretation.

Results

Study Selection

The systematic search identified 5786 references from electronic databases. After removal of 1806 duplicates (3 identified manually and 1803 by Covidence), 3980 unique database references remained. Combined with 31 references from citation searching (n=14) and gray literature (n=17), a total of 4011 references were screened at the title and abstract level. Title and abstract screening excluded 3848 studies, leaving 163 for full-text review. Exclusion categories at the full-text stage were abstract only without accessible full text (n=44), unsuitable study design (n=3), focused on nonemergency clinical conditions such as routine outpatient management (n=13), no formal cost-effectiveness analysis (n=15), not primary research (n=4), and did not involve an emergency care delivery setting despite addressing acute conditions (n=71). The distinction between "nonemergency conditions" and "did not involve emergency care settings" reflects studies where the clinical condition was nonacute (eg, chronic disease management) vs studies where the condition was acute but the care setting was not an ED or emergency care pathway (eg, postdischarge follow-up and scheduled urgent care). A total of 13 studies met all inclusion criteria and were included in the review (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.



Study Characteristics

The 13 included studies were published between 2011 and 2024, with over half (n=7, 54%) published after 2015, reflecting recent growth in virtual emergency care evaluation. Studies originated from 6 countries: the United States (n=6), Australia (n=2), Italy (n=2), Canada (n=1), Haiti (n=1), and Belgium (n=1), representing diverse health care systems and economic contexts (Table 1).

XSL•FO
RenderX

Table 1. Study characteristics of the included studies.

Study (first author, year)	Country	Sample size	Population	Emergency conditions	Virtual care model	Technology	Provider type	Comparator
Brunetti, 2014 [33]	Italy	109,750 ECGs[a]	Patients with suspected acute cardiac disease	Acute coronary syndrome, arrhythmia	Prehospital telemedicine ECG triage	12-lead ECG via mobile phone	Remote cardiologist (24/7)	Conventional ED[b] triage
Demaerschalk, 2013 [34]	United States	1112 (hypothetical)	Patients with acute ischemic stroke	Acute ischemic stroke	Hub-and-spoke telestroke network	Videoconferencing	Vascular neurologists (24/7)	No telestroke network
Espinoza, 2017 [35]	Belgium	2282 (registry), 1000 (modeled)	Patients with suspected stroke	Acute stroke (all types)	In-ambulance telemedicine	Live AV[c] + vital signs	Neurologist + EMS[d]	Standard stroke care
Flaherty, 2022 [36]	Haiti	10,585 (projected)	Children 0-9 years	Diarrheal disease, respiratory infections, skin infections	Nighttime telemedicine with medication delivery	Telephone call center	Nurses with physician oversight	Hospital emergency care
Kadel, 2018 [37]	Italy	1676	Neurosurgical emergencies	TBI[e], spinal injuries, ICH[f], brain cancer	Televideoconsultation	Multispecialist video system	Consultant neurosurgeon	Transfer to the referral hospital
Kim, 2022 [38]	Australia	1024	Patients with stroke <4.5 h onset	Stroke	ED-based telestroke	Mobile computer/camera	Physician	No virtual consultation
Nelson, 2011 [23]	United States	8 spoke hospitals	Acute ischemic stroke	Acute ischemic stroke	ED-based telestroke	2-way audiovisual	Stroke specialists	Usual care without telestroke
Nelson 2016 [22]	United States	864 (98 pre, 766 post)	Acute ischemic stroke	Acute ischemic stroke	Hub-spoke telestroke (POTN[g])	2-way AV videoconferencing	Stroke specialists	Nontelestroke care
Pace, 2023 [39]	Canada	108	Pediatric emergencies <18 years	Critical pediatric emergencies (CTAS[h]-1/2)	Pediatric teleresuscitation	Videoconferencing	Pediatric specialists	Standard ED care
Thomas, 2018 [40]	United States	494	Pediatric psychiatric emergencies	Suicide intent, self-harm, depression/anxiety	ED-based telepsychiatry	Videoconferencing	Psychiatrists, social workers	Ambulance transfer
Von Huben, 2024 [41]	Australia	49,228 ED presentations	Any age, rural ED patients	General ED presentations	Virtual Rural Generalist Service (75% virtual)	HD[i] videoconferencing	Rural generalist doctors	Local/locum doctors
Whetten, 2018 [42]	United States	777 consultations	Neuroemergent conditions	AIS[j], mTBI[k], neuroemergent disorders	ED-based telemedicine consultation	2-way AV with image transfer	Neurologists/neurosurgeons	Standard rural care
Yang, 2015 [43]	United States	135	Pediatric critical care <18 years	Asthma, bronchiolitis, pneumonia	ED-based pediatric telemedicine	Live interactive AV	Pediatric critical care physicians	Telephone consultation

[a]ECG: electrocardiogram.

[b]ED: emergency department.

[c]AV: audiovisual.

[d]EMS: emergency medical services.

[e]TBI: traumatic brain injury.

[f]ICH: intracranial hemorrhage.

[g]POTN: Providence Oregon Telestroke Network.

[h]CTAS: Canadian Triage and Acuity Scale.

[i]HD: high definition.

[j]AIS: acute ischemic stroke.

XSL•FO
RenderX

[k]mTBI: mild traumatic brain injury.

Study designs varied, with model-based economic evaluations being most common (n=7, 54%) [22,23,34,35,39,42,43], followed by observational cost-effectiveness analyses (n=4, 31%) [8,36,37,40] and 2 (15%) within-trial analyses [22,38]. Sample sizes ranged from 108 to 109,750 patients, with a median of 1068 (IQR 706-4358) patients (Table 1). The largest study [8] evaluated a prehospital electrocardiogram triage program in Italy serving a population of 4 million.

Clinical conditions addressed included stroke/cerebrovascular emergencies (n=7, 54%), pediatric emergencies (n=4, 31%), general neuroemergencies (n=2, 15%), psychiatric emergencies (n=1, 8%), and cardiac emergencies (n=1, 8%). One study [41] examined general ED presentations without condition-specific focus (Table 1).

Economic Outcomes

Economic evaluation methods varied across studies, with cost-effectiveness analysis being most common (n=8), followed by cost-utility analysis (n=3), cost-minimization (n=1), and cost-consequence analysis (n=1). A total of 8 (62%) studies adopted a health care system perspective, while 4 (31%) used a broader societal perspective incorporating patient and family costs (Table 2).

XSL•FO
RenderX

Table 2. Economic evaluation results of virtual emergency care models.[a]

Study (first author, year)	Type of analysis	Perspective	Time horizon	Effectiveness measure	Costs (intervention vs comparator)	ICER[b]	Key findings	Sensitivity analysis
Brunetti, 2014 [33]	Cost-minimization/cost-utility	Regional health system	1 year	Lives saved (STEMI[c]); QALYs[d]	€16.70 vs €24.80-55.20 per ECG[e] (costs in 2012 €)	€1927/QALY (€990 to –€2508 after correction)	Cost-saving; lower cost per ECG/Consultation	Range analysis (savings of €8.10-€38.40 per ECG/consultation)
Demaerschalk, 2013 [34]	Cost-utility	Societal	Lifetime	QALYs	Intervention cost US $1436 lower compared to standard care over lifetime (costs in 2011 US dollars)	Dominant (saves US $1436, gains 0.02 QALYs)	Cost-saving and more effective	1-way, 2-way sensitivity analyses
Espinoza, 2017 [35]	Cost-utility	Health care payer	Lifetime	QALYs	Savings of US $30 compared to standard care (costs in 2014 US dollars)	Dominant (saves US $30, gains 0.00456 QALYs)	Cost-effective from 6-min time gain; dominant from 12 min	1-way, 2-way, probabilistic (5000 iterations)
Flaherty, 2022 [36]	Cost-effectiveness	Societal	1 year	DALYs[f] averted	US $317,898/year vs US $89,392/year (costs in average US dollars from 2019 July-2021 June)	US $1288/DALY averted	Cost-effective (below 3 × GDP threshold)	Univariate sensitivity analysis
Kadel, 2018 [37]	Cost-effectiveness	Health service	4 years	Transfers avoided	€2326 vs €4173 per patient (costs in 2009 €)	€365/transfer avoided (dominant)	73% transfer reduction; €3,095,863 total savings	None performed
Kim, 2022 [38]	Cost-effectiveness	Societal	1 year	QALYs	AUD $82,259 vs AUD $82,449 at 12 months (costs in 2018 Australian dollars)	AUD $0-$5000/QALY (50.6% iterations)	Cost-effective; small cost savings with QALY gains	Deterministic and probabilistic (1000 iterations)
Nelson, 2011 [23]	Cost-effectiveness	Societal	90 days and lifetime	QALYs	Infrastructure + care costs 90-day: US $14, 274 vs US $13, 872 lifetime: US $133,527 vs US $130, 343 (costs in 2008 US dollars)	90-day: US $108,363/QALY; lifetime: US $2449/QALY	Cost-effective over lifetime	1-way and Monte Carlo (10,000 iterations)

XSL•FO
RenderX

Study (first author, year)	Type of analysis	Perspective	Time horizon	Effectiveness measure	Costs (intervention vs comparator)	ICER[b]	Key findings	Sensitivity analysis
Nelson, 2016 [22]	Cost-effectiveness	Spoke and hub	Inpatient stay	QALYs	Spoke Δ: US $51-$1956; hub Δ: US $863-$2767	Spoke: US $1322-$50,687/QALY; Hub: US $22,363-$71,703/QALY	Cost-effective; best for severe strokes; tPA[g] rate tripled	1-way by NIHSS[h] and cost-sharing; probabilistic (10,000)
Pace, 2023 [39]	Cost-effectiveness	Societal	3 years	YLL[i] averted	CAD $2032.73 vs CAD $317.45 (costs in 2022 Canadian dollars)	CAD $64.61/YLL averted	Cost-effective; reduced mortality	Deterministic (±20%) and probabilistic (10,000 runs)
Thomas, 2018 [40]	Cost-efficiency	Hospital system	Single encounter	ED[j] LOS[k], disposition rates	US $3493 vs US $8611 per patient (median charges; costs in 2015 US dollars)	Not calculated	Lower charges, shorter LOS (5.5 vs 8.3 hours)	None reported
Von Huben, 2024 [41]	Cost-consequence	Health care funder	1 year	Multiple outcomes	AUD $134 vs AUD $240 per ED encounter (costs in 2022 AUD)	Dominant (less costly, more effective)	AUD $105 savings per ED encounter; improved outcomes	1-way and scenario analyses
Whetten, 2018 [42]	Cost-effectiveness	Health care payer	90 days	QALYs	US $4241 savings per patient (costs in 2015 US dollars)	Dominant (saves US $21,205/QALY)	Cost-saving; 0.20 QALY gain per patient	1-way, tornado diagram, Monte Carlo (10,000)
Yang, 2015 [43]	Cost-effectiveness and ROI[l]	Health care payer	1 year	Transfers avoided	US $3641 vs US $8303 per child/ED/year (costs in 2013 US dollars)	Dominant (saves US $4662 per child/ED/year; ROI: 1.28	31% transfer reduction; US $46,620 annual savings per ED	1-way, 2-way, probabilistic Monte Carlo (5000 iterations)

[a]Costs reported in non-US currencies in this table were converted to US dollars using the average annual exchange rate for the currency and cost year reported in each study: €1=US $1.39 (2009); €1=US $1.29 (2012); AUD $1=US $0.75 (2018); AUD $1=US $0.69 (2022); CAD $1=US $0.77 (2022). The same rates are applied throughout the manuscript.

[b]ICER: incremental cost-effectiveness ratio.

[c]STEMI: ST-elevation myocardial infarction.

[d]QALY: quality-adjusted life year.

[e]ECG: electrocardiogram.

[f]DALY: disability-adjusted life year.

[g]tPA: tissue plasminogen activator.

[h]NIHSS: National Institutes of Health Stroke Scale.

[i]YLL: years of life lost.

[j]ED: emergency department.

[k]LOS: length of stay.

[l]ROI: return on investment.

Cost-Effectiveness Results

Virtual emergency care was associated with favorable economic outcomes across all included studies, though the strength and mechanisms of cost-effectiveness varied by clinical domain. Stroke-related applications (n=6) demonstrated the most consistent evidence of dominance, driven by increased thrombolysis rates and transfer avoidance. Pediatric applications (n=4) achieved cost-effectiveness primarily through reduced transfers and averted hospitalizations, while psychiatric and

XSL•FO
RenderX

general emergency applications were each supported by single studies, limiting the certainty of subgroup-specific conclusions. A total of 6 (46%) studies found virtual care to be dominant, that is, both less costly and more effective than usual care. These included telestroke networks [34,42], pediatric telemedicine [43], neuroemergency consultation [37], and rural generalist services [41].

Among studies reporting ICERs, the time horizon significantly influenced results: Nelson et al [23] reported an ICER of US $108,363/QALY at 90 days, which improved to US $2449/QALY over a lifetime horizon. Nelson et al [22] demonstrated that from the spoke hospital perspective, ICERs ranged substantially depending on the proportion of implementation costs borne by the spoke facility: US $1322/QALY when the spoke bore 0% of implementation costs, US $25,991/QALY at 50%, and US$50,687/QALY at 100%. From the hub perspective, the overall ICER ranged from US $22,363/QALY to US $71,703/QALY. Cost-effectiveness was most pronounced for patients with severe stroke (National Institutes of Health Stroke Scale [NIHSS] score ≥15), with spoke ICERs ranging from US $7794/QALY to US $40,071/QALY, while mild strokes (NIHSS <5) were not cost-effective from the spoke perspective when spokes bore ≥50% of implementation costs. Most reported ICERs fell below commonly accepted willingness-to-pay thresholds of US $50,000-$100,000 per QALY [44].

Cost savings varied by setting and intervention type. Per-encounter savings ranged from US $73 (AUD $105) [41] to US $5118 [40]. The primary driver of cost savings was a reduction in patient transfers, which decreased by 31%-73% across studies [37,43]. Transfer avoidance alone generated savings of US $3729 per patient in [42]. The prehospital model demonstrated different economics: Espinoza et al [35] found in-ambulance telemedicine to be dominant with time gains of 12 minutes or more, saving US $30 per patient while gaining 0.00456 QALYs. Notably, this intervention became cost-effective with time gains as small as 6 minutes, demonstrating that time gains influenced cost-effectiveness thresholds.

Clinical Effectiveness

Virtual care interventions consistently improved clinical outcomes alongside economic benefits. QALY gains ranged from 0.02 [34] to 0.20 [42] per patient (over a lifetime), with stroke and neuroemergency applications showing the largest improvements. Nelson et al [22] reported an incremental effectiveness of 0.039 QALYs per patient for telestroke compared to nontelestroke assisted care, with telestroke implementation leading to a more than threefold increase in tissue plasminogen activator (tPA) treatment rates (from 9.2% to 31.7%). This treatment effect was most dramatic in moderate-severity strokes (NIHSS 5-14), where a nearly seven-fold increase in tPA treatment was observed (6.9% to 47.7%). The Haiti pediatric service [36] averted 199.76 DALYs annually at a cost of US $1288 per DALY averted, which is well below the country's GDP-based cost-effectiveness threshold.

ED length of stay decreased by 2.8 hours with virtual consultations in Thomas et al [40]. This study documented a median ED length of stay of 5.5 hours for telepsychiatry vs 8.3 hours for usual care (P<.001). Hospital admission rates fell significantly: Thomas et al [40] reported a reduction from 56% to 42% for pediatric psychiatric emergencies. Importantly, these efficiency gains occurred without compromising quality or safety, with patient satisfaction exceeding 90% in all studies reporting this outcome.

Quality Assessment

Methodological quality was independently assessed using the Drummond checklist (10 items focused on analytical rigor) and the CHEC list (19 items emphasizing reporting standards). The included studies demonstrated consistently high quality, with mean scores of 9.29 (SD 0.60) out of 10 criteria met for the Drummond checklist and 18.04 (SD 0.79) out of 19 criteria met for the CHEC list. While we present these aggregate figures for descriptive purposes, we note that these checklists were designed as qualitative assessment tools rather than quantitative scoring instruments, and individual items are not equally weighted. Item-level quality assessment results are presented in Multimedia Appendix 3 to allow readers to evaluate specific methodological strengths and weaknesses. The term "overall mean quality score" has been removed to avoid implying a validated composite metric. All 13 (100%) studies achieved an "Excellent" quality rating (≥85%), with Kim et al [38] scoring the highest at 100% (Table 3). All studies clearly stated research questions, appropriately selected the type of economic evaluation, comprehensively identified costs, and explicitly reported perspectives and time horizons. Sensitivity analyses were conducted in the majority of studies (12/13, 92%), often using multiple approaches. The most common limitations were inadequate discussion of generalizability (observed in 5/12, 42% of studies) and incomplete conflict of interest reporting (7/12, 58%). Notably, no study scored below 83% on either checklist, indicating that the economic evaluations of virtual emergency care consistently meet high methodological standards. This uniform methodological rigor, paired with consistently favorable economic findings.

Using GRADE methodology, we assessed the certainty of evidence as high for cost-effectiveness outcomes based on 10 studies with consistent positive findings. Evidence certainty was moderate for transfer rate reduction and quality of life improvements, and low for ED length of stay and mortality benefits due to study design limitations and imprecision (Table 4).

The evidence generally supports virtual emergency care implementation, with 6/10 economic evaluations demonstrating dominance (cost-saving with better outcomes). However, one study showed unfavorable short-term cost-effectiveness, and effectiveness measures varied across studies. Large reductions were observed primarily in transfer rates (31%-90%), with cost impacts varying by setting and condition.

XSL•FO
RenderX

Table 3. Quality assessment of included studies using Drummond and CHEC (Consensus Health Economic Criteria) checklists.

Study (first author, year)	Condition	Economic perspective	Drummond score (out of 10)	CHEC score (out of 19)	Overall quality
Brunetti, 2014 [33]	Acute coronary syndrome	Health care system	8.0	16	Good (82.10%)
Demaerschalk, 2013 [34]	Acute ischemic stroke	Societal	10.0	18.5	Excellent (98.69%)
Espinoza, 2017 [35]	Stroke	Health care payer	10.0	17.5	Excellent (96.05%)
Flaherty, 2022 [36]	Pediatric emergencies	Societal	9.5	18.5	Excellent (96.18%)
Kadel, 2018 [37]	Neurosurgical emergencies	Health service	9.0	18.5	Excellent (93.68%)
Kim, 2022 [38]	Stroke	Societal	10.0	19.0	Excellent (100.00%)
Nelson, 2011 [23]	Acute ischemic stroke	Societal	10.0	18.5	Excellent (98.69%)
Nelson, 2016 [22]	Acute ischemic stroke	Spoke and hub perspectives	9.0	18.0	Excellent (92.37%)
Pace, 2023 [39]	Pediatric resuscitation	Societal	9.0	19.0	Excellent (95.00%)
Thomas, 2018 [40]	Behavioral health emergencies (pediatric)	Hospital system	9.0	17.5	Excellent (91.05%)
Von Huben, 2024 [41]	Mixed ED[a] presentations	Health care funder	9.0	18.5	Excellent (93.68%)
Whetten, 2018 [42]	Stroke	Health care payer	9.0	18.0	Excellent (92.37%)
Yang, 2015 [43]	Pediatric conditions	Health care payer	9.0	18.0	Excellent (92.37%)

[a]ED: emergency department.

XSL•FO
RenderX

Table 4. GRADE (Grading of Recommendations Assessment, Development, and Evaluation) evidence summary table: overall evidence assessment by outcome.[a]

Outcome	Studies, n	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Virtual care	Control	Effect estimate	Certainty	Importance
Cost-effectiveness (ICER[b]/QALY[c])	10 [23,33-39,42,43]	6 RCTs[d]/quasi-experimental, 4 modeling studies	Serious (–1)[e]	Not serious	Not serious	Not serious	Large effect (+1), dose response (+1)	671/1924	855/2107	7/10 dominant, 3/10 cost-effective ICERs: US $1273 (€990)-US $108,363/QALY	⊗ high	Critical
Transfer rates	7 [23,34-37,42,43]	4 observational, 3 modeling	Serious (–1)[f]	Not serious	Not serious	Not serious	Large effect (+1), consistent direction	423/1389	756/1389	31-85% reduction RR[g] 0.15-0.77	⊗ moderate	Critical
ED length of stay	4 [38-41]	3 observational, 1 RCT	Serious (–1)[h]	Not serious	Not serious	Serious (–1)[i]	Consistent direction	829/1846	1017/2039	2.8-3.2 hours reduction MD[j] -2.95 hours	⊗ low	Important
Quality of life (QALYs gained)	6 [23,33-35,38,42]	3 RCTs, 3 modeling	Serious (–1)[k]	Not serious	Not serious	Not serious	All positive effects	712/1547	919/1730	0.02-0.20 QALYs gained MD +0.11 QALYs	⊗ moderate	Critical
Mortality/lives saved	3 [33,39,41]	1 observational, 2 modeling	Serious (–1)[l]	Not serious	Serious (–1)[m]	Not serious	Strong association	185/654	231/687	Lives saved for STEMI[n], 26.55 YLL[o] averted OR[p] 0.73 (0.52-0.95)	⊗ low	Critical
Patient satisfaction	3 [33,39,41]	3 observational	Serious (–1)[q]	Not serious	Not serious	Serious (–1)[r]	Large effect	NR	NR	>90% approval rates	⊗ low	Important

[a]Costs reported in non-US currencies in this table were converted to US dollars using the average annual exchange rate for the currency and cost year reported in each study: €1=US $1.39 (2009); €1=US $1.29 (2012); AUD $1=US $0.75 (2018); AUD $1=US $0.69 (2022); CAD $1=US $0.77 (2022). The same rates are applied throughout the manuscript.

[b]ICER: incremental cost-effectiveness ratio.

[c]QALY: quality-adjusted life year.

[d]RCT: randomized controlled trial.

[e]Most studies nonrandomized with potential selection bias.

[f]Observational studies with potential confounding.

[g]RR: risk ratio.

[h]Limited randomized data, measurement inconsistencies.

[i]Wide CIs, small sample sizes.

[j]MD: mean difference.

[k]Mix of study designs with methodological limitations.

[l]Limited studies, potential reporting bias.

[m]Indirect mortality measures in some studies.

[n]STEMI: ST-elevation myocardial infarction.

[o]YLL: years of life lost.

[p]OR: odds ratio.

[q]Subjective measures, potential response bias.

[r]Limited sample sizes, measurement variability.

XSL•FO
RenderX

Subgroup Analyses

By Clinical Condition

Stroke care showed the most robust evidence, with 5/6 studies demonstrating dominance [22,23,34,37,42]. The consistent mechanism was increased thrombolysis rates through rapid specialist consultation, generating both clinical benefits and cost savings through avoided transfers and improved outcomes. Nelson et al [22] extended this evidence by demonstrating that the cost-effectiveness of telestroke varied significantly by stroke severity: severe strokes (NIHSS ≥15) showed the most favorable ICERs (US $7794/QALY to US $40,071/QALY from the spoke perspective), while mild strokes (NIHSS <5) were dominated when spokes bore ≥50% of implementation costs. This severity-stratified analysis, the first of its kind in the telestroke cost-effectiveness literature, provides critical nuance for implementation planning. Demaerschalk et al [34] documented US $1436 lower costs and 0.02 QALY gains over patient lifetimes in their hub-and-spoke telestroke network evaluation. The prehospital stroke model [35] achieved cost-effectiveness through a different mechanism, through time savings enabling earlier treatment rather than transfer avoidance. Their model showed that every minute saved translated into measurable economic and clinical benefits.

Pediatric emergency applications [36,39,43] achieved favorable economics through different mechanisms. Critical care consultations avoided unnecessary transfers (31% reduction) [43], while the Haiti model prevented hospitalizations through early intervention. Yang et al [43] reported ROI ranging from 1.28 to 1.96 across pediatric studies, with annual cost savings of US $46,620 per ED treating 10 children with telemedicine.

By Geographic Setting

Rural and remote implementations [22,23,34,36,41-43] consistently showed superior cost-effectiveness compared to urban applications, driven primarily by high transfer costs and distances. Rural settings achieved 6 of 7 dominant outcomes, with savings amplified by avoided air medical transports and family travel costs. Whetten et al [42] demonstrated that transfer rates from rural hospitals to urban medical facilities decreased from 85% to 5% (no tPA) and from 90% to 23% (tPA) with their Access to Critical Cerebral Emergency Support Services (ACCESS) telestroke program. Nelson et al [22] provided additional evidence from the Pacific Northwest, where 17 spoke facilities in the Providence Oregon Telestroke Network (POTN) served rural communities. Their analysis uniquely demonstrated how cost-sharing arrangements between hub and spoke facilities influence cost-effectiveness, with the spoke perspective yielding ICERs as low as US $1322/QALY when implementation costs were fully subsidized by the hub.

Urban network implementations [23,34] achieved cost-effectiveness through operational efficiency and care standardization rather than transfer reduction. The Italian regional networks [8,37] demonstrated how high-volume centralized systems can achieve economies of scale.

By Technology Type

Video consultation technology's [22,23,34,35,37-43] versatility across clinical conditions and ability to support visual assessment appeared critical for both clinical and economic success (Table 5). These implementations typically featured 2-way, high-definition audiovisual systems connecting EDs to remote specialists. Hub-and-spoke models were common, with specialist coverage ratios of 1:7-8 peripheral sites as documented by Demaerschalk et al [34] and Nelson et al [23]. Nelson et al [22] used a similar hub-and-spoke model with 2 hub facilities (1 primary stroke center and 1 comprehensive stroke center) supporting 17 spoke facilities, demonstrating scalability of this approach.

Other virtual modalities included mobile electrocardiogram transmission [8], telephone consultation with medication delivery [36], and a hybrid model combining 75% virtual with 25% in-person coverage [41]. Technology sophistication varied considerably, from basic telephone systems in resource-limited Haiti to advanced platforms with peripheral examination devices and rapid imaging transfer capabilities. Lower-technology solutions proved viable in resource-constrained settings. The Haiti telephone-based model [36] achieved cost-effectiveness despite, or perhaps because of, its simplicity, suggesting technology should match context rather than maximize sophistication. The in-ambulance telemedicine model [35] represents an intermediate technology approach, combining real-time consultation with automated data transmission. This prehospital intervention uniquely demonstrated that time gains as small as 6 minutes could achieve cost-effectiveness, with dominance achieved at 12 minutes.

XSL•FO
RenderX

Table 5. Virtual care modality comparison—types, cost-effectiveness, and implementation contexts.[a]

Virtual care modality	Technology platform	Clinical applications	Cost-effectiveness findings	Implementation setting	Key success factors	Challenges/limitations
Real-time video consultation	• 2-way HD[b] videoconferencing • Mobile carts with cameras • Peripheral examination devices	• Stroke [23,34,35,42] • Pediatric emergencies [39,40,43] • Neuroemergencies [37,38] • Psychiatric emergencies [40]	• 3 of 4 studies indicate the dominant cost-saving of the intervention • CAD $64.61 per YLL[c] averted; Saves US $4662 per child/ED[d]/year • € 365 per transfer avoided; • 0-5000 AUD/QALY[e] • Lower charges; shorter LOS[f] for the intervention	• Rural EDs to urban specialists • Hub-and-spoke networks • 24/7 availability	• Specialist expertise access • Rapid decision-making • Visual assessment capability • High clinician satisfaction	• Technology infrastructure costs • Internet connectivity in rural areas • Staff training requirements
Mobile ECG[g] transmission	• 12-lead ECG recorder • Mobile phone transmission • Central hub interpretation	• Acute coronary syndrome • Cardiac arrhythmias • Prehospital triage [33]	• €990-€2508/QALY • €16.70 vs €24.80-€55.20/ECG • Lives saved for STEMI[h] patients	• Prehospital EMS[i] units • Regional coverage (4M population) • 109,750 ECGs/year	• Rapid cardiac diagnosis • Avoided unnecessary transports • 24/7 cardiologist availability	• Limited to cardiac conditions • Requires EMS training • Mobile network dependency
Telephone consultation	• Call center platform • Twilio intake system • Beacon driver dispatch	• Pediatric preemergencies • Nighttime service • Medication delivery [36]	• US $1288/DALY[j] averted • Cost-effective below 3× GDP • 199.76 DALYs averted/year	• Haiti (resource-limited) • Nighttime-only service • Home medication delivery	• Addresses access barriers • Low technology requirements • Integrated medication delivery	• Limited to pre-emergency cases • No visual assessment • Requires delivery infrastructure
Hybrid virtual-physical model	• HD videoconferencing • Wireless telehealth carts • 75% virtual, 25% in-person	• General ED presentations • Rural hospital coverage predominantly lower acuity [41]	• Dominant strategy • AUD $105 savings/ED encounter • Improved quality metrics	• 29 rural hospitals • 24/7 coverage • Virtual rural generalists	• Workforce sustainability • Reduced locum dependency • Maintained care continuity	• Requires some physical presence • Complex scheduling • Multiple site coordination

[a]Costs reported in non-US currencies in this table were converted to US dollars using the average annual exchange rate for the currency and cost year reported in each study: €1=US $1.39 (2009); €1=US $1.29 (2012); AUD $1=US $0.75 (2018); AUD $1=US $0.69 (2022); CAD $1=US $0.77 (2022). The same rates are applied throughout the manuscript.

[b]HD: high definition.

[c]YLL: years of life lost.

[d]ED: emergency department.

[e]QALY: quality-adjusted life year.

[f]LOS: length of stay.

XSL•FO
RenderX

[g]ECG: electrocardiogram.

[h]STEMI: ST-elevation myocardial infarction.

[i]EMS: emergency medical services.

[j]DALY: disability-adjusted life year.

By Provider Type

Provider types delivering virtual care included neurologists and stroke specialists [23,34,35,37,38,42], pediatric critical care physicians and psychiatrists [36,39,40,43], rural generalist physicians [41], cardiologists [8], and nurses with physician oversight [36]. Most programs operated 24/7, though the Haiti service functioned only during nighttime hours when traditional care was unavailable. A unique prehospital model was evaluated by Espinoza et al [35], featuring in-ambulance telemedicine with live bidirectional audio-video communication, automated vital signs transmission, and 24/7 neurologist consultation. This system served 150 patients per ambulance annually, with 39% of eligible patients receiving telemedicine consultation.

Sensitivity analyses were conducted in 12/13 (92%) studies. The most common approaches were 1-way deterministic sensitivity analysis (n=9), probabilistic sensitivity analysis using Monte Carlo simulation (n=7, with iterations ranging from 1000 to 10,000), 2-way sensitivity analysis (n=4), and scenario analysis (n=2). Kadel et al [37] were the only study that did not perform a sensitivity analysis. Key findings from sensitivity analyses included results were robust to parameter variation in 10/12 studies; Nelson et al [23] found that the ICER was most sensitive to the time horizon assumption, shifting from US $108,363/QALY at 90 days to US $2449/QALY over a lifetime; Nelson et al [22] demonstrated through probabilistic sensitivity analysis (10,000 iterations) that telestroke was the most cost-effective strategy in at least half of simulation runs when willingness-to-pay thresholds exceeded US $5000/QALY (spoke bearing 0% costs), US $30,000/QALY (50% costs), or US $52,000/QALY (100% costs); Espinoza et al [35] identified the minimum time gain threshold (6 minutes) needed for cost-effectiveness; and Yang et al [43] reported that ROI estimates remained favorable (range 1.28-1.96) across probabilistic iterations. Key uncertainty drivers across studies included time horizon, transfer rates, intervention use rates, cost-sharing arrangements, and discount rates. No study identified scenarios under which virtual care became definitively cost-ineffective in their primary sensitivity analyses, though shorter time horizons and lower use rates reduced the magnitude of economic benefits.

Cost Components Analysis

Direct medical costs consistently favored virtual care, with per-consultation costs ranging from €16.70 (US $21) for electrocardiogram telemedicine [8] to US $600 per consultation for neurological telemedicine [42]. While initial infrastructure investment was substantial, break-even typically occurred within 1-3 years based on operational savings. The Victorian Stroke Telemedicine program showed cost savings of US $142 (AUD $190) per patient (control: US $61,631 [AUD $82,449]; intervention: US $61,489 [AUD $82,259]) at 12 months [38]. Direct medical costs consistently favored virtual care across multiple studies. Brunetti et al [8] reported telemedicine electrocardiogram consultation costs of €16.70 (US $21) per encounter compared to €24.80-€55.20 (US $32-71) for conventional ED visits in Italy. Similarly, Von Huben et al [41] documented cost savings of US $73 (AUD $105) per ED encounter for their Virtual Rural Generalist Service (US $93 vs US $167 [AUD $134 vs AUD $240] for usual care). Nelson et al [22] provided detailed cost breakdowns from POTN: spoke facility fixed costs included monthly camera fees (US $1980/month), 1-time equipment setup (US $11,000), and yearly training (US $768), while hub facility costs included 1-time equipment (US $26,300), annual wireless connectivity (US $4500), a full-time telestroke coordinator (US $89,679), and yearly training (US $900). These real-world cost data, derived from encounter-level financial records rather than literature estimates, represent a significant contribution to understanding the true infrastructure investment required for telestroke implementation.

Infrastructure investment varied considerably by program scope and technology. While initial setup costs could be substantial, rapid ROI was consistently demonstrated. Yang et al [43] calculated an ROI of 1.28 in their base case, rising to 1.96 in probabilistic analyses for pediatric telemedicine consultations.

Transfer and transportation represented the largest cost differential, particularly in rural settings. Virtual care programs achieved dramatic reductions in transfer rates—from 85%-90% to 5%-23% in patients with stroke [42] and a 73% reduction in neurosurgical transfers [37]. Nelson et al [22] reported a slight decline in transfer probability for tPA-treated patients (55.6% to 51.9%) but an increase among untreated patients (11.2% to 16.8%), the latter likely reflecting improved identification of patients requiring higher-level care. Beyond direct ambulance costs, virtual care eliminated family travel costs and productivity losses. The Canadian pediatric teleresuscitation program documented mean costs of US $1563 (CAD $2032.73) for telemedicine cases vs US $244 (CAD $317.45) for controls [39]. Transfer and transportation costs represented the largest area of savings. Whetten et al [42] reported cost savings of US $3729 per patient, primarily from reduced unnecessary transfers (85% to 5% for non-tPA patients and 90% to 23% for tPA patients). Kadel et al [37] documented a 73% reduction in transfer rates for neurosurgical emergencies, saving 139,916 km in travel distance and €3,095,863 (US $4,303,250) over 4 years. Beyond direct ambulance costs, virtual care eliminated associated family expenses and productivity losses.

Staffing efficiencies emerged as a significant economic driver. Hub-and-spoke models achieved specialist coverage ratios of 1:7-8 sites compared to traditional 1:1 models [23,34]. The POTN model [22] demonstrated a similar approach with 2 hub facilities supporting 17 spoke facilities, with an average of 14.4 patients with acute ischemic stroke per spoke per year. Von Huben et al [41] documented substantial reductions in expensive locum physician shifts (from 1456 to 609 days), with the Virtual Rural Generalist Service achieving a price per activity unit of

XSL•FO
RenderX

AUD $1047 compared to AUD $1753 for usual care. Staffing efficiencies emerged as a critical economic driver. They reported a 60% reduction in expensive locum physician shifts from 1456 to 609 days annually, with the price per activity unit lower for virtual care (AUD $1047) compared to usual care (AUD $1753). Hub-and-spoke models achieved efficient specialist coverage, as demonstrated by Demaerschalk et al [34] with 1 hub hospital supporting 7 spoke hospitals.

The Haiti pediatric telemedicine program demonstrated unique cost-effectiveness at US $1288 per DALY averted, well below the World Health Organization cost-effectiveness threshold of 3 times per capita GDP (US $3531) [36]. This program included medication delivery, representing an innovative model combining consultation and pharmaceutical services.

Implementation Factors

Studies consistently identified high initial technology costs as the primary implementation barrier, though rapid ROI mitigated this concern. Yang et al [43] demonstrated an ROI of 1.28 in base case analysis, with pediatric telemedicine generating annual cost-savings of US $46,620 per ED treating 10 children. Rural connectivity challenges were addressed through tiered technology approaches—the Haiti study successfully operated using basic telephone systems [36], while other programs required high-definition videoconferencing [41]. Technology infrastructure costs emerged as the primary implementation barrier across studies. However, flexibility in technology requirements facilitated adoption. Flaherty et al [36] successfully implemented a nighttime pediatric telemedicine service in Haiti using basic telephone technology, while Von Huben et al [41] used high-definition videoconferencing in Australian rural hospitals.

Facilitators of successful implementation included strong clinical champions, executive support, and phased rollout strategies targeting high-need populations. Programs focusing on time-critical conditions like stroke demonstrated early wins that facilitated expansion. The Victorian Stroke Telemedicine program strategically excluded 4 large hospitals with existing stroke protocols, focusing implementation where impact would be greatest [38]. Programs targeting high-need populations demonstrated early wins—stroke programs consistently showed increased tPA administration rates and improved outcomes [23,34].

Integration with existing workflows proved critical for success. Kadel et al [37] demonstrated efficiency with 75% of teleconsultations completed within 15 minutes and 90% within 30 minutes. The ACCESS program achieved consultation completion rates with neurologists available for rural EDs, resulting in cost savings of US $4241 per patient and QALY gains of 0.20 per patient [42]. Clinical workflow integration proved critical for success. Kadel et al [37] reported that 75% of teleneurosurgery consultations were completed within 15 minutes and 90% within 30 minutes, demonstrating efficiency that encouraged provider adoption. Similarly, Brunetti et al [8] achieved rapid electrocardiogram interpretation with remote cardiologists available 24/7.

Volume and case mix significantly influenced economic outcomes. The Brunetti et al study processed 109,750 electrocardiograms from 274,198 total emergency calls in 2012, achieving economies of scale with a cost per electrocardiogram/consultation of €16.70 (US $21) [8]. Programs with lower volumes, such as the Belgian in-ambulance telemedicine study with 39% use rates, still achieved cost-effectiveness when time gains exceeded 6-12 minutes [35].

Provider acceptance was enhanced by demonstrable benefits to work-life balance and patient outcomes. Thomas et al [40] reported reduced ED length of stay (5.5 vs 8.3 hours) and lower charges (US $3493 vs US $8611) for telepsychiatry compared to usual care. Pace et al [39] documented an ICER of US $50 (CAD $64.61) per year of life lost averted for pediatric teleresuscitation, demonstrating clear value.

Funding models and reimbursement structures varied by country and health care system. In the United States, several studies noted consultation fees ranging from US $600 per consult [42] to operational costs of US $222,800 annually for telepsychiatry services [40]. The dominant finding across studies was that virtual emergency care programs generated net savings despite initial investments.

Discussion

Principal Findings

This systematic review found that virtual emergency care was associated with favorable economic outcomes across diverse clinical applications and health care settings. While these findings suggest economic viability, the predominance of model-based evaluations (7/13, 54% of included studies) and the absence of empirical negative findings warrant cautious interpretation. Model-based analyses rely on assumptions about clinical pathways, costs, and transition probabilities that may not fully capture real-world complexities. Multiple mechanisms underlie these favorable outcomes, including avoiding patient transfers and improving operational efficiency. Our results present a supportive economic case beyond direct cost savings, highlighting workforce sustainability benefits as well. For instance, one study documented a 60% reduction in the use of costly locum physicians [41].

It is important to recognize that the clinical heterogeneity across included studies spanning telestroke, pediatric telemedicine, psychiatric emergencies, cardiac triage, and rural generalist services limits the generalizability of aggregate conclusions. Each clinical domain involves distinct patient populations, care pathways, cost structures, and outcome measures (QALYs, DALYs, transfers avoided, and length of stay). While the overall direction of findings was consistently positive, the evidence base is strongest for stroke care (supported by multiple studies with concordant findings) and more preliminary for psychiatric and general emergency applications (each supported by single studies).

Comparison With Prior Work

Our findings extend earlier domain-specific reviews in several important ways. Previous systematic reviews of telestroke cost-effectiveness [7] have demonstrated favorable outcomes

XSL•FO
RenderX

in stroke-specific applications, while reviews of telemedicine in emergency settings [7,17,18] have primarily focused on clinical effectiveness rather than economic outcomes. Our review is the first to synthesize economic evidence across the full spectrum of virtual emergency care modalities, demonstrating that the favorable economics observed in telestroke extend to pediatric, psychiatric, and general emergency applications. This broader scope strengthens the evidence base for policy decisions that span multiple clinical domains.

Unlike the mixed results reported for telemedicine cost-effectiveness in nonemergency settings by Whitten et al [45], our findings were uniformly positive—a discrepancy that may reflect emergency care's unique characteristics (high baseline costs, time-sensitive decisions, and clear outcome metrics) rather than a true absence of unfavorable results. Furthermore, the success of simpler technologies in resource-constrained settings challenges the assumption that high technological sophistication is necessary for cost-effectiveness. For example, a telephone-based consultation model in Haiti was found to be cost-effective [36], suggesting that matching the technology to the context can be more important than using the most advanced tools. Our review also highlights the prehospital telemedicine model as an important middle ground: ambulance-based systems can achieve cost-effectiveness without a full hospital infrastructure.

Moreover, by including full economic evaluations rather than cost-only studies, our review provides more robust evidence than prior reviews that examined only cost data without health outcome integration.

Interpretation of Cost-Effectiveness Findings

The cost-effectiveness of virtual emergency care was driven by distinct mechanisms across clinical domains. In stroke care, the primary value proposition was enabling time-sensitive thrombolysis through rapid specialist consultation, where even modest improvements in treatment times translated into substantial QALY gains and reduced long-term disability costs. The strong lifetime cost-effectiveness (ICER of US $2449/QALY) despite unfavorable short-term economics (US $108,363/QALY at 90 days) underscores the importance of capturing downstream benefits in economic evaluations of acute interventions.

In pediatric applications, cost-effectiveness was primarily driven by transfer avoidance and reduced unnecessary hospitalizations, reflecting the high baseline costs of pediatric emergency transfers and the ability of specialist teleconsultation to safely manage patients locally. The variability in ICERs across studies—ranging from US $1273 (€990)/QALY to US $108,363/QALY—reflects differences in time horizons, health care system costs, comparator practices, and the specific clinical conditions studied. These contextual factors mean that the transferability of specific ICER values to new settings requires careful consideration of local cost structures, existing care pathways, and patient populations.

Policy and Clinical Implications

The evidence supports consideration of expanding virtual emergency care programs, particularly in settings where the economic case is strongest, while acknowledging that implementation decisions should account for local contextual factors not fully captured in economic models. The rapid uptake of virtual care during temporary COVID-19 policy relaxations demonstrated the feasibility of swift implementation. Health care administrators should prioritize high-impact virtual applications (eg, telestroke and pediatric critical care) in underserved areas to maximize early ROI and build organizational support; documented break-even periods of 12-24 months can reassure budget-conscious decision-makers.

Moreover, the fact that all economic perspectives examined showed positive results further strengthens the investment case: even analyses using a narrow health care system perspective demonstrated cost savings, while broader societal perspective analyses revealed added benefits such as reduced costs for families and productivity gains. For clinicians, it is significant that the economic benefits of virtual emergency care were accompanied by quality improvements (maintained or improved patient outcomes, >90% patient satisfaction, and reduced ED length of stay). This finding suggests that virtual care can enhance care quality rather than compromise it, while also reducing clinicians' workload through improved patient flow.

Methodological Considerations

Although the included studies used diverse economic evaluation methods reflecting real-world variability, this heterogeneity complicates direct comparisons and underscores the need for more standardized reporting of cost-effectiveness outcomes (eg, consistent use of ICER calculations, uniform costing methods, and common time horizons). The development of a core set of economic outcome measures for virtual emergency care could significantly improve comparability across studies.

While 54% (7/13) of the studies were model-based analyses, thereby raising some questions about generalizability, the consistency of findings between these models and the observational studies is reassuring. However, the reliance on modeling introduces inherent uncertainty, as results depend on assumptions about clinical pathways, transition probabilities, and cost structures that may not hold across all settings.

Overall, the quality of the economic evidence was generally high: on average, studies met over 90% of criteria on validated quality assessment checklists, though we note that these percentage scores should be interpreted as descriptive indicators rather than precise quality metrics, as these checklists were not designed for numerical aggregation. However, future evaluations should strengthen their reporting of generalizability and better disclose any conflicts of interest. Encouragingly, 92% (12/13) of the studies included sensitivity analyses, indicating that uncertainty was appropriately handled—which is an important practice for decision-maker confidence.

Sensitivity Analysis and Uncertainty

While 92% (12/13) of included studies conducted sensitivity analyses, several important sources of uncertainty merit

XSL•FO
RenderX

discussion. Key parameters driving variation in cost-effectiveness estimates included the time horizon of analysis, with shorter horizons (eg, 90 days) producing less favorable ICERs than lifetime models; patient volume and use rates, which influenced the spread of fixed infrastructure costs; transfer rates and distances, which were the dominant cost drivers in rural settings; and discount rates applied to future health benefits.

Notably, no study systematically explored scenarios in which virtual care would not be cost-effective, such as settings with very low patient volume, high existing specialist availability, or unreliable technology infrastructure. Future evaluations should explicitly model conditions under which virtual emergency care may not represent good value for money, to provide decision-makers with a more complete picture of when implementation is and is not warranted.

Equity and Global Generalizability

The evidence base is dominated by high-income countries (the United States, Australia, Canada, Italy, and Belgium), with only one study conducted in a low-income setting (Haiti). This geographic concentration significantly limits the generalizability of findings to low- and middle-income countries (LMIC), where health care infrastructure, digital connectivity, workforce composition, and cost structures differ substantially. While the Haiti study demonstrated that simpler technologies can achieve cost-effectiveness in resource-constrained environments, this single study cannot address the diverse challenges faced across LMIC. Claims about global scalability should therefore be interpreted with caution.

Future research should prioritize economic evaluations in LMIC, where the potential impact of virtual emergency care may be greatest but where implementation barriers—including limited broadband access, electricity reliability, and trained workforce availability—are also most pronounced. Equity considerations, including whether virtual emergency care reduces or exacerbates disparities related to digital literacy, language, age, or socioeconomic status, also require dedicated investigation.

Strengths and Limitations

This review implemented several measures to minimize bias. We conducted a comprehensive search across multiple databases, including gray literature and non-English studies, and applied rigorous quality assessments using validated tools alongside the GRADE approach. Furthermore, by focusing exclusively on full economic evaluations, we ensured that our conclusions reflect true value generation rather than cost shifting between settings.

Despite these strengths, the study has several limitations. The interventions evaluated were heterogeneous, which precluded a meaningful meta-analysis. In addition, most included studies adopted a health care system (rather than societal) perspective, potentially underestimating the full benefits of virtual care.

Publication bias represents a significant concern in this review. Despite our comprehensive search across 8 databases and gray literature sources, all 13 included studies reported favorable economic outcomes for virtual emergency care. The complete absence of negative or neutral findings is noteworthy and may reflect selective publication of studies with positive results, selective reporting of favorable outcomes within studies, or genuine cost-effectiveness of virtual emergency care across all evaluated contexts. The possibility that unfavorable economic evaluations remain unpublished or were not identified by our search cannot be excluded. This limitation should temper the certainty with which the overall positive conclusions are interpreted.

Several important cost dimensions are underrepresented in the included studies. Ongoing technology maintenance and upgrade costs, which can be substantial as hardware ages and software requires updates, were rarely incorporated. Staff training costs—both initial and recurrent—were inconsistently reported, yet represent a significant and recurring investment. Workflow integration costs, including the time and resources required to redesign clinical processes, develop protocols, and manage change, were largely absent from economic evaluations. Additionally, the included studies did not systematically account for potential downstream costs such as follow-up care resulting from remote diagnostic limitations, liability and malpractice considerations, or the costs of addressing technology failures and downtime.

The long-term sustainability of cost savings remains uncertain, as initial efficiencies may diminish as programs scale, technology costs evolve, or reimbursement structures change. While early evidence indicates that the benefits are sustained or even amplified over time, longer-term studies will be important to confirm this durability.

While this review adhered to PRISMA 2020 guidelines, certain recommended items—including formal assessment of reporting bias across studies and provision of underlying extracted data—were not fully addressed. The complete dataset of extracted study characteristics and economic outcomes is available from the corresponding author upon request.

Future Research Directions

Looking ahead, further research is needed in several key areas. First, economic evaluation methods should be standardized to facilitate cross-study comparisons and enable future meta-analyses. Second, researchers should evaluate emerging virtual care modalities such as AI-driven triage algorithms and integrated digital emergency systems to determine which features truly drive cost-effectiveness and which simply add expense. Third, dedicated studies should assess the equity impacts of virtual emergency care to determine whether it reduces or exacerbates health care disparities, with particular attention to the digital divide. Fourth, implementation science approaches are needed to optimize how virtual emergency care is delivered. Comparative studies of different care models, training methods, and change management strategies could offer practical guidance for health systems, given that current variation in program success suggests significant room for improvement. Finally, long-term evaluations are crucial to assess how the benefits of virtual emergency care evolve over time, how quality of care is maintained as programs scale up, and how these models adapt to changing technology landscapes.

XSL•FO
RenderX

Future evaluations should also adopt longer time horizons and incorporate comprehensive implementation cost frameworks—including technology maintenance, staff training, workflow redesign, and opportunity costs—to provide a more complete economic picture. Studies should explicitly model conditions under which virtual emergency care may not be cost-effective to help decision-makers identify appropriate implementation contexts.

Despite these research needs, the existing evidence supports broader implementation of virtual emergency care, particularly for high-need populations and time-sensitive emergency conditions where its benefits are most pronounced.

Conclusion

This systematic review found consistent evidence of favorable economic outcomes for virtual emergency care across diverse clinical conditions, settings, and modalities. The evidence is strongest for stroke care and rural/remote implementations, where transfer avoidance drives substantial savings. While all included studies reported cost-effectiveness or cost savings, the predominance of model-based analyses, the absence of evidence from LMIC, and the possibility of publication bias warrant cautious interpretation. These findings support the continued expansion of virtual emergency care, particularly for high-need populations and time-sensitive conditions, while highlighting the need for standardized economic evaluation methods, long-term sustainability assessments, and equity-focused research to guide future implementation decisions.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files. The complete search strategy, data extraction forms, and quality assessment ratings are available in Multimedia Appendices 2 and 3. The dataset of extracted study characteristics and economic outcomes is available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: RS, SW
Methodology: RS
Formal analysis: RS
Screening: LW, SHH
Data extraction: LW, SHH
Data curation: RS
Writing – original draft: RS
Writing – review and editing: RS, LW, SHH, MFL, SPKG, TCW, SW
Supervision: MFL, SPKG, SW
Project administration: RS

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[PDF File (Adobe PDF File), 161 KB - mhealth_v14i1e82143_app1.pdf]

Multimedia Appendix 2
Full electronic search strategies for all eight databases.
[DOCX File , 19 KB - mhealth_v14i1e82143_app2.docx]

Multimedia Appendix 3
Item-level methodological quality assessment of the included studies using the Drummond and CHEC checklists.
[DOCX File , 23 KB - mhealth_v14i1e82143_app3.docx]

References

1. Di Somma S, Paladino L, Vaughan L, Lalle I, Magrini L, Magnanti M. Overcrowding in emergency department: an international issue. Intern Emerg Med 2015;10(2):171-175. [doi: 10.1007/s11739-014-1154-8] [Medline: 25446540]

XSL•FO
RenderX

2. Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L. Emergency department crowding: a systematic review of causes, consequences and solutions. PLoS One 2018;13(8):e0203316 [FREE Full text] [doi: 10.1371/journal.pone.0203316] [Medline: 30161242]
3. Baum A, Kaboli PJ, Schwartz MD. Reduced in-person and increased telehealth outpatient visits during the COVID-19 pandemic. Ann Intern Med 2021;174(1):129-131 [FREE Full text] [doi: 10.7326/M20-3026] [Medline: 32776780]
4. Hsia RY, Sarkar N, Shen Y. Impact of ambulance diversion: Black patients with acute myocardial infarction had higher mortality than Whites. Health Aff (Millwood) 2017;36(6):1070-1077 [FREE Full text] [doi: 10.1377/hlthaff.2016.0925] [Medline: 28583966]
5. Sun BC, Hsia RY, Weiss RE, Zingmond D, Liang L, Han W, et al. Effect of emergency department crowding on outcomes of admitted patients. Ann Emerg Med 2013;61(6):605-611.e6 [FREE Full text] [doi: 10.1016/j.annemergmed.2012.10.026] [Medline: 23218508]
6. Natafgi N, Shane DM, Ullrich F, MacKinney AC, Bell A, Ward MM. Using tele-emergency to avoid patient transfers in rural emergency departments: an assessment of costs and benefits. J Telemed Telecare 2018;24(3):193-201. [doi: 10.1177/1357633X17696585] [Medline: 29278984]
7. Baratloo A, Rahimpour L, Abushouk AI, Safari S, Lee CW, Abdalvand A. Effects of telestroke on thrombolysis times and outcomes: a meta-analysis. Prehosp Emerg Care 2018;22(4):472-484. [doi: 10.1080/10903127.2017.1408728] [Medline: 29345529]
8. Brunetti ND, Scalvini S, Acquistapace F, Parati G, Volterrani M, Fedele F, et al. Telemedicine for cardiovascular disease continuum: a position paper from the Italian Society of Cardiology Working Group on Telecardiology and Informatics. Int J Cardiol 2015;184:452-458. [doi: 10.1016/j.ijcard.2015.02.023] [Medline: 25755064]
9. Ward MM, Jaana M, Natafgi N. Systematic review of telemedicine applications in emergency rooms. Int J Med Inform 2015;84(9):601-616. [doi: 10.1016/j.ijmedinf.2015.05.009] [Medline: 26072326]
10. Zachrison KS, Boggs KM, M Hayden E, Espinola JA, Camargo CA. A national survey of telemedicine use by US emergency departments. J Telemed Telecare 2020;26(5):278-284. [doi: 10.1177/1357633X18816112] [Medline: 30558518]
11. Dorsey ER, Topol EJ. Telemedicine 2020 and the next decade. Lancet 2020;395(10227):859. [doi: 10.1016/S0140-6736(20)30424-4] [Medline: 32171399]
12. Wiedermann CJ, Mahlknecht A, Piccoliori G, Engl A. Redesigning primary care: the emergence of artificial-intelligence-driven symptom diagnostic tools. J Pers Med 2023;13(9):1379 [FREE Full text] [doi: 10.3390/jpm13091379] [Medline: 37763147]
13. Schwamm LH, Rosenthal ES, Hirshberg A, Schaefer PW, Little EA, Kvedar JC, et al. Virtual telestroke support for the emergency department evaluation of acute stroke. Acad Emerg Med 2004;11(11):1193-1197 [FREE Full text] [doi: 10.1197/j.aem.2004.08.014] [Medline: 15528584]
14. Bashiri M, Greenfield LJ, Oliveto A. Telemedicine interest for routine follow-up care among neurology patients in Arkansas. Telemed J E Health 2016;22(6):514-518. [doi: 10.1089/tmj.2015.0112] [Medline: 26684500]
15. Ashwood JS, Mehrotra A, Cowling D, Uscher-Pines L. Direct-to-consumer telehealth may increase access to care but does not decrease spending. Health Aff (Millwood) 2017;36(3):485-491. [doi: 10.1377/hlthaff.2016.1130] [Medline: 28264950]
16. Zachrison KS, Boggs KM, Hayden EM, Espinola JA, Camargo CA. Understanding barriers to telemedicine implementation in rural emergency departments. Ann Emerg Med 2020;75(3):392-399. [doi: 10.1016/j.annemergmed.2019.06.026] [Medline: 31474481]
17. Shigekawa E, Fix M, Corbett G, Roby DH, Coffman J. The current state of telehealth evidence: a rapid review. Health Aff (Millwood) 2018;37(12):1975-1982. [doi: 10.1377/hlthaff.2018.05132] [Medline: 30633674]
18. Langabeer JR, Gonzalez M, Alqusairi D, Champagne-Langabeer T, Jackson A, Mikhail J, et al. Telehealth-enabled emergency medical services program reduces ambulance transport to urban emergency departments. West J Emerg Med 2016;17(6):713-720 [FREE Full text] [doi: 10.5811/westjem.2016.8.30660] [Medline: 27833678]
19. Ahmed AA, Mojiri M, Daghriri AA, Hakami OA, Alruwaili RF, Khan RA, et al. The role of telemedicine in emergency department triage and patient care: a systematic review. Cureus 2024;16(12):e75505. [doi: 10.7759/cureus.75505] [Medline: 39803013]
20. Sharifi Kia A, Rafizadeh M, Shahmoradi L. Telemedicine in the emergency department: an overview of systematic reviews. Z Gesundh Wiss 2022:1-15 [FREE Full text] [doi: 10.1007/s10389-021-01684-x] [Medline: 35103232]
21. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
22. Nelson RE, Okon N, Lesko AC, Majersik JJ, Bhatt A, Baraban E. The cost-effectiveness of telestroke in the Pacific Northwest region of the USA. J Telemed Telecare 2016;22(7):413-421. [doi: 10.1177/1357633X15613920] [Medline: 26541170]
23. Nelson RE, Saltzman GM, Skalabrin EJ, Demaerschalk BM, Majersik JJ. The cost-effectiveness of telestroke in the treatment of acute ischemic stroke. Neurology 2011;77(17):1590-1598 [FREE Full text] [doi: 10.1212/WNL.0b013e318234332d] [Medline: 21917781]
24. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. United Kingdom: Oxford University Press; 2015.

XSL•FO
RenderX

25. Evers S, Goossens M, de Vet H, van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: consensus on health economic criteria. Int J Technol Assess Health Care 2005;21(2):240-245 [FREE Full text] [Medline: 15921065]
26. Shemilt I, Mugford M, Byford S, Drummond M, Eisenstein E, Knapp M. Incorporating economics evidence. In: Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series. London: Cochrane; 2008:449-479.
27. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. BMJ 1996;313(7052):275-283 [FREE Full text] [doi: 10.1136/bmj.313.7052.275] [Medline: 8704542]
28. Watts RD, Li IW. Use of checklists in reviews of health economic evaluations, 2010 to 2018. Value Health 2019;22(3):377-382 [FREE Full text] [doi: 10.1016/j.jval.2018.10.006] [Medline: 30832977]
29. Frederix GWJ. Check your checklist: the danger of over- and underestimating the quality of economic evaluations. Pharmacoecon Open 2019;3(4):433-435 [FREE Full text] [doi: 10.1007/s41669-019-0118-3] [Medline: 30680675]
30. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336(7650):924-926 [FREE Full text] [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]
31. Brunetti M, Shemilt I, Pregno S, Vale L, Oxman AD, Lord J, et al. GRADE guidelines: 10. Considering resource use and rating the quality of economic evidence. J Clin Epidemiol 2013;66(2):140-150. [doi: 10.1016/j.jclinepi.2012.04.012] [Medline: 22863410]
32. Brozek JL, Canelo-Aybar C, Akl EA, Bowen JM, Bucher J, Chiu WA, GRADE Working Group. GRADE guidelines 30: the GRADE approach to assessing the certainty of modeled evidence-an overview in the context of health decision-making. J Clin Epidemiol 2021;129:138-150 [FREE Full text] [doi: 10.1016/j.jclinepi.2020.09.018] [Medline: 32980429]
33. Brunetti ND, De Gennaro L, Correale M, Santoro F, Caldarola P, Gaglione A, et al. Pre-hospital electrocardiogram triage with telemedicine near halves time to treatment in STEMI: a meta-analysis and meta-regression analysis of non-randomized studies. Int J Cardiol 2017;232:5-11. [doi: 10.1016/j.ijcard.2017.01.055] [Medline: 28089154]
34. Demaerschalk BM, Switzer JA, Xie J, Fan L, Villa KF, Wu EQ. Cost utility of hub-and-spoke telestroke networks from societal perspective. Am J Manag Care 2013;19(12):976-985 [FREE Full text] [Medline: 24512034]
35. Espinoza AV, Devos S, van Hooff R, Fobelets M, Dupont A, Moens M, et al. Time gain needed for in-ambulance telemedicine: cost-utility model. JMIR Mhealth Uhealth 2017;5(11):e175 [FREE Full text] [doi: 10.2196/mhealth.8288] [Medline: 29175808]
36. Flaherty KE, Klarman MB, Cajusma Y, Schon J, Exantus L, Beau de Rochars VM, et al. A nighttime telemedicine and medication delivery service to avert pediatric emergencies in Haiti: an exploratory cost-effectiveness analysis. Am J Trop Med Hyg 2022;106(4):1063-1071 [FREE Full text] [doi: 10.4269/ajtmh.21-1068] [Medline: 35189597]
37. Kadel R, Evans-Lacko S, Tramarin A, Stopazzolo G. Cost-effectiveness of tele-video-consultation for the neuro-surgical emergency management at the general hospitals in Italy. Front Neurosci 2018;12:908 [FREE Full text] [doi: 10.3389/fnins.2018.00908] [Medline: 30564091]
38. Kim J, Tan E, Gao L, Moodie M, Dewey H, Bagot K, et al. Cost-effectiveness of the Victorian Stroke Telemedicine Program. Aust Health Rev 2022;46(3):294-301. [doi: 10.1071/AH21377] [Medline: 35589669]
39. Pace A, Faught BE, Law M, Mateus L, Roy M, Sulowski C, et al. Economic evaluation of tele-resuscitation intervention on emergency department pediatric visits in the Niagara Region, Canada a pilot study. Front Health Serv 2023;3:1105635 [FREE Full text] [doi: 10.3389/frhs.2023.1105635] [Medline: 37342797]
40. Thomas JF, Novins DK, Hosokawa PW, Olson CA, Hunter D, Brent AS, et al. The use of telepsychiatry to provide cost-efficient care during pediatric mental health emergencies. Psychiatr Serv 2018;69(2):161-168. [doi: 10.1176/appi.ps.201700140] [Medline: 29032703]
41. Von Huben A, Thompson AE, Wilson A, Luscombe GM, Haigh A, Howard K, et al. An economic evaluation of the virtual rural generalist service versus usual care in Western NSW Local Health District. Med J Aust 2024;221 Suppl 11:S28-S36. [doi: 10.5694/mja2.52530] [Medline: 39647925]
42. Whetten J, van der Goes DN, Tran H, Moffett M, Semper C, Yonas H. Cost-effectiveness of Access to Critical Cerebral Emergency Support Services (ACCESS): a neuro-emergent telemedicine consultation program. J Med Econ 2018;21(4):398-405 [FREE Full text] [doi: 10.1080/13696998.2018.1426591] [Medline: 29316820]
43. Yang NH, Dharmar M, Yoo B, Leigh JP, Kuppermann N, Romano PS, et al. Economic evaluation of pediatric telemedicine consultations to rural emergency departments. Med Decis Making 2015;35(6):773-783 [FREE Full text] [doi: 10.1177/0272989X15584916] [Medline: 25952744]
44. Neumann PJ, Cohen JT, Weinstein MC. Updating cost-effectiveness--the curious resilience of the $50,000-per-QALY threshold. N Engl J Med 2014;371(9):796-797. [doi: 10.1056/NEJMp1405158] [Medline: 25162885]
45. Whitten PS, Mair FS, Haycox A, May CR, Williams TL, Hellmich S. Systematic review of cost effectiveness studies of telemedicine interventions. BMJ 2002;324(7351):1434-1437 [FREE Full text] [doi: 10.1136/bmj.324.7351.1434] [Medline: 12065269]

XSL•FO
RenderX

Abbreviations

ACCESS: Access to Critical Cerebral Emergency Support Services
CHEC: Consensus Health Economic Criteria
DALY: disability-adjusted life year
ED: emergency department
GRADE: Grading of Recommendations Assessment, Development, and Evaluation
ICER: incremental cost-effectiveness ratio
LMIC: low- and middle-income countries
NIHSS: National Institutes of Health Stroke Scale
POTN: Providence Oregon Telestroke Network
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
QALY: quality-adjusted life year
ROI: return on investment
tPA: tissue plasminogen activator

Edited by L Buis; submitted 09.Aug.2025; peer-reviewed by P Usuemerai, MA Rezaei, J Roca; comments to author 10.Feb.2026; revised version received 10.Jun.2026; accepted 10.Jun.2026; published 05.Aug.2026.

Please cite as:
Shankar R, Wang L, Ho SH, Liew MF, Kumar Gollamudi SP, Wong TC, Wong S
Cost-Effectiveness of Virtual Emergency Care Models: Systematic Review
JMIR Mhealth Uhealth 2026;14:e82143
URL: https://mhealth.jmir.org/2026/1/e82143
doi:10.2196/82143
PMID:

©Ravi Shankar, Linda Wang, Soon Hoe Ho, Mei Fong Liew, Satya Pavan Kumar Gollamudi, Tze Chin Wong, Serene Wong. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 05.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Application of Mobile Health Technologies to Prevent Unintentional Childhood Injuries: Scoping Review

Huan Wang[1*], MS; Fengjiao Pu[1*], MS; Dongjun Jing[1*], MS; Hangcheng Liu[2], MNS; Miao Li[3], MNS; Xixi Li[1], PhD

[1]Department of Nursing, Mianyang Central Hospital, School of Medicine, University of Electronic Science and Technology of China, Mianyang, Sichuan, China

[2]Department of Obstetrics and Gynecology, Mianyang Central Hospital, School of Medicine, University of Electronic Science and Technology of China, Mianyang, Sichuan, China

[3]School of Nursing, North Sichuan Medical College, Nanchong, Sichuan, China

[*]these authors contributed equally

Corresponding Author:
Xixi Li, PhD
Department of Nursing
Mianyang Central Hospital, School of Medicine, University of Electronic Science and Technology of China
12 Changjia Lane, Jingzhong Street
Fucheng District
Mianyang, Sichuan, 621000
China
Phone: 86 18281690555
Email: 68218001@qq.com

Abstract

Background: Unintentional childhood injuries represent a major public health issue affecting the lives and health of children worldwide, with the risk and injury type dynamically changing with age and developmental stage. To effectively prevent unintentional injuries, intervention measures must be adjusted based on actual circumstances and kept up to date in real time. Mobile health (mHealth) technologies can meet this need, but their specific functions and intervention methods remain unclear at present.

Objective: This study aimed to conduct a scoping review of research on the application of mHealth technologies in the prevention and emergency management of unintentional injuries among children, with a focus on identifying the types of mHealth technologies used, intervention contents, and outcome measures, thereby providing a reference for related studies.

Methods: We systematically searched 5 databases—PubMed, Web of Science, Embase, Cochrane Library, and CINAHL—from their inception to December 31, 2025. Data extraction and synthesis were performed on the included studies.

Results: Of 1085 articles, 15 (1.3%) met the inclusion criteria, and 1 additional reference was added during full-text review. Thus, 16 studies were included. The included studies comprised 9 (56.2%) randomized controlled trials, 2 (12.5%) quasi-experimental studies, 3 (18.8%) mixed methods studies, 1 (6.2%) qualitative study, and 1 (6.2%) descriptive study. mHealth technologies included apps, WeChat, mobile-based e-learning programs, and virtual reality combined with wearable devices. Their functions included health education, interactive communication, monitoring and reminders, behavioral training, questionnaires and record-keeping, and emergency response. Outcome measures included the incidence of unintentional injuries among children; knowledge, attitudes, and behaviors regarding injuries; other psychological and social indicators; and usability of mHealth technology.

Conclusions: mHealth technologies have shown some progress in reducing unintentional injuries among children. Content based on this technology has demonstrated positive effects in reducing injury incidence, enhancing caregivers' awareness, and improving related attitudes and behavioral capabilities. However, only 2 mobile apps are currently available to the public. Future research should optimize the functional design of mHealth, strengthen sustainability, explore long-term implementation methods, and conduct large-scale, long-term studies to further evaluate its impact on the incidence of unintentional injuries among children.

Trial Registration: Open Science Framework 8QPR6; https://osf.io/8qpr6

(JMIR Mhealth Uhealth 2026;14:e94038) doi:10.2196/94038

XSL•FO
RenderX

KEYWORDS

child; mobile health; scoping review; technologies; unintentional injuries

Introduction

Background

Unintentional injuries are bodily harm caused by sudden events, including accidental falls, drowning, burns, suffocation, and poisoning [1]. It is estimated that unintentional injuries are the leading cause of death among children aged 1 to 19 years globally and the fifth leading cause of death among infants aged <1 year [2,3]. The 2019 Global Burden of Disease Study reveals that 205,000 children and adolescents aged 5 to 24 years died from unintentional injuries, with children aged 5 to 9 years facing the highest risk of death. According to China's Under-5 Child Mortality Surveillance System reports, between 2010 and 2020, a total of 7925 children died from unintentional injuries, and the proportion of deaths among injuries rose from 15.2% in 2010 to 23.8% in 2020 [4]. It not only poses a severe threat to children's lives and health but also imposes heavy economic burdens and psychological stress on families and society. The occurrence of unintentional injuries involves multiple factors, including the child's age and developmental stage, the guardians' educational and supervisory levels, and the safety conditions of the surrounding environment [5-7]. Among all the risk factors, age and developmental stage are important predictors, indicating that injury types and risk factors vary dynamically with the child's age and developmental stage. For example, young children are more prone to unintentional injuries, and infants aged <1 year are commonly more prone to choking or suffocation; while children over 1 year of age are more likely to experience accidental falls or burns [6]. It is this characteristic that requires interventions for unintentional injuries to be timely adjusted and updated to the child's age and developmental stage.

The World Health Organization published the World Report on Child Injury Prevention (2008), which outlines key interventions across multiple areas, including legislation, product safety, environmental modification, education, and first aid [8]. However, the strategies, both domestically and internationally, primarily rely on offline approaches that have limited coverage, target smaller populations, and yield short-term effects, wasting time and money and lacking sustainability. With the rapid development of IT, mobile health (mHealth) offers a new perspective and platform for preventing and controlling unintentional childhood injuries. mHealth refers to the use of smartphones, wearable devices, mobile apps, and other technologies to deliver health promotion, disease prevention, and public health management services to the public [9]. This technology offers convenience, real-time access, personalization, and intelligence, enabling dynamic updates to intervention content and long-term tracking, thereby addressing the shortcomings of traditional approaches and aligning with the requirements of unintentional injuries [10]. Currently, mHealth technologies are widely used in areas such as chronic disease management, rehabilitation, cancer care, and perinatal care, and show great promise [11-14].

In the field of child injury prevention, some studies have also attempted to develop mHealth interventions to prevent unintentional injuries in children. For example, Feng et al [15] used WeChat to regularly send information on the prevention of unintentional injuries and emergency strategies to parents. Cooray et al [16] used a mobile app to help parents of children aged <1 year improve their ability to prevent infant falls. Wong et al [17] used serious games to help children learn traffic safety signs and understand potential safety hazards in different areas. Because researchers use different mHealth technologies, the target populations, functional focuses, and intervention outcomes vary. A review is needed to clarify the format, advantages, effectiveness, and potential limitations of each technology. Currently, there are few systematic reviews in this field. Schulze et al [18] conducted a review of mobile apps for unintentional injuries in children aged <7 years, detailing the user groups, specific needs, and evaluation metrics of these apps. However, this study included only 5 papers published before 2021, resulting in limited evidence that does not comprehensively reflect the full scope of the field. Shetty et al [19] reviewed studies on interventions for unintentional injuries in children from 2013 to 2023, including 9 eHealth-based studies; however, they similarly failed to provide detailed descriptions of the technologies' features. Overall, there remains a lack of systematic reviews that comprehensively describe the current status of mHealth technology applications, the types of technologies, and the effectiveness of interventions for preventing unintentional injuries among children.

Objectives

This study aims to provide a systematic review of the current status of mHealth technology applications in interventions for unintentional injuries in children. The primary objectives are to (1) summarize the types of mHealth technologies used for unintentional injuries in children through a systematic review of the existing literature and (2) examine how these technologies are used in interventions for unintentional injuries in children, what the outcome measures of these interventions are, and how effective these interventions are.

Methods

The study followed the Joanna Briggs Institute Manual for Scoping Reviews Framework and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews; Multimedia Appendix 1) [20]. The protocol was registered in Open Science Framework.

Defining Research Questions

The review was guided by the following research questions:

1. What mHealth technologies are used in interventions for unintentional injuries among children? Which technologies are currently available?
2. What are the target populations and intervention content of studies on mHealth technology-based interventions?
3. What are the outcome measures for mHealth technologies?

XSL•FO
RenderX

4. How effective are mHealth technologies in preventing unintentional injuries among children?

Definition of mHealth Technology

The concept of mHealth originates from eHealth, which refers to the use of information and communication technology to support health care services and information dissemination [21]. With the widespread adoption of smartphones and tablets, eHealth services—which previously relied on fixed devices or locations—are gradually shifting toward mobile devices as their primary platform. This study follows the World Health Organization's definition, which defines mHealth as health care and public health practices supported by mobile devices [9]. Whether a technology falls under the category of mHealth is primarily determined by whether it is accessed via mobile devices such as smartphones and tablets. Accordingly, in addition to common apps and wearable devices, this study includes within the scope of mHealth any technology accessed or operated through mobile portable devices, such as internet platforms, serious games, and virtual reality (VR) technology.

Search Strategy

A systematic search was conducted across 5 databases: PubMed, Web of Science, Embase, Cochrane Library, and CINAHL. A combination of subject headings and free-text terms was used, with additional reference tracing. The search covers each database from its inception to December 31, 2025. Search terms included "children," "child," "infant," "child, preschool," "accidental injuries," "unintentional injuries," "accidental," "injuries," "mobile applications," "app," "mobile phone," "social media," "digital health," "digital media," "telemedicine," "internet," "text messaging," "virtual," "digital technology," "artificial intelligence," "Web," "WeChat," "online social network," "message," and "smartphone." The search terms and strategy used are presented in Multimedia Appendix 2.

Literature Inclusion and Exclusion Criteria

Literature inclusion criteria were established based on population, concept, and context principles: (1) the study population comprised children, child caregivers, or child educators; (2) the study concept involved mHealth technology research; (3) the context pertained to studies related to unintentional injuries among children. The study types included original research such as randomized controlled trials, quasi-experimental studies, mixed method studies, and observational studies. Exclusion criteria were as follows: (1) duplicate publications; (2) if there are multiple papers reporting on the same study, only the one reporting the final results will be included; (3) study protocols; (4) publications for which the full text is unavailable; and (5) publications not in Chinese or English.

Literature Screening and Data Extraction

Literature screening was conducted using NoteExpress software (version 4.1.0.10030; Beijing Aegean Software), with 2 researchers independently performing initial and duplicate screening. Disagreements were resolved through discussion with a third researcher. Extracted data included authors, publication year, country, study type, study subjects, sample size, app type, study duration, specific implementation methods, control group measures, and outcome measures.

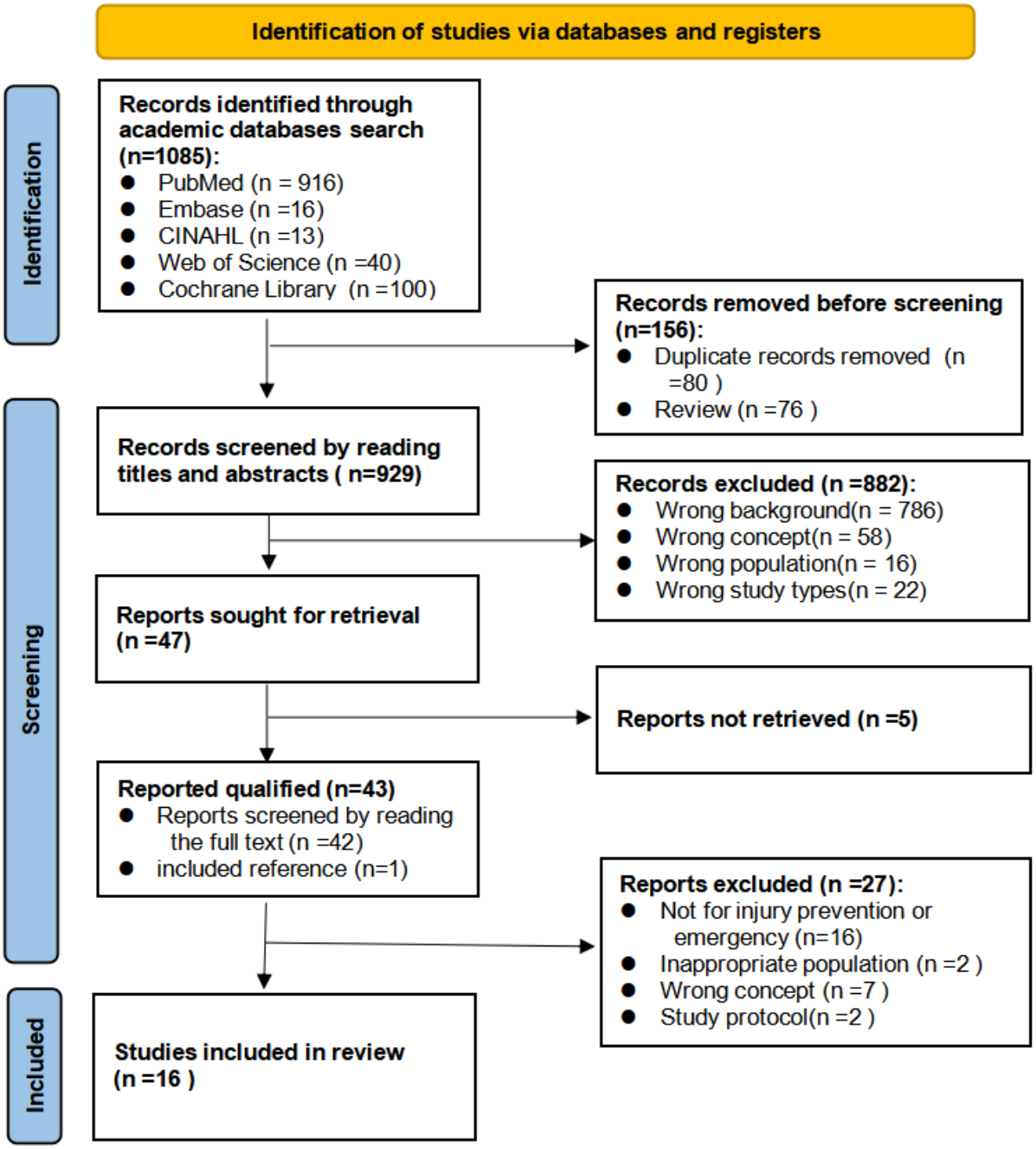
Results

Literature Search Results

The initial search yielded 1085 articles. After removing duplicates and reviews, 929 articles remained. Following a review of titles and abstracts, 882 articles were excluded. After reviewing the full texts, 25 articles were excluded for irrelevant research topics, 2 were excluded as protocols, and 5 could not be retrieved. There were 15 remaining articles, and 1 article was included through reference tracing. Finally, a total of 16 articles were ultimately included. The literature screening process is illustrated in Figure 1.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) flow diagram.



Basic Characteristics of Included Literature

All included studies were in English. The collection comprised 9 randomized controlled trials [15,17,22-28], 2 quasi-experimental studies [29,30], 3 mixed method studies [16,31,32], 1 qualitative study [33], and 1 descriptive study [34]. Geographically, 6 studies were from China [15,17,25,26,28,30], 3 from the United States [23,24,32], 1 from South Korea [22,29,34], 2 from Australia [16,27], and 1 each from the United Kingdom [33] and Germany [31]. Specific characteristics are shown in Multimedia Appendix 3.

Types of mHealth Technologies

Mobile Apps

Twelve studies [16,17,23-25,27-29,31-34] used apps as their delivery platform. Among these, 6 apps [16,17,23,25,27,28,33,34] were specifically developed by the research teams to meet their own research needs. Ning et al [25] and He et al [28] used the same app (Bao Hu San, Protective Umbrella), which was developed by Ning et al [25] for injury prevention education among urban child caregivers; He et al [28] adapted and improved it based on that study to address rural conditions. The game app Safe City, developed by Wong et al [17], is designed to help children develop proper traffic safety behaviors. This app is currently available for download via a web browser, but users must provide a mobile phone

XSL•FO
RenderX

number to register. The apps developed by Kang et al [34] and Jones et al [33] provide education on the prevention of unintentional injuries to schoolteachers and mothers, respectively; however, all of them are currently in the development and usability testing phases. A total of 4 apps [24,27,29,31,32] have been developed by relevant organizations in advance. The app used by McKenzie et al [24] and Roberts et al [32] (Make Safe Happen) was developed by the US Centers for Injury Prevention and Control and is specifically designed for parents of children aged 0 to 12 years. Its features include age-appropriate safety information, room safety checklists, links to purchase safety products, the ability to create shopping lists and set calendar reminders, and the option to add a poison control hotline to phone contacts with a single tap. However, this app is no longer being updated and is unavailable for download. The app used in the study by Kim and Cho [29] was developed by the South Korean Ministry of Education to help children handle emergencies they might encounter at school or in daily life; the download link was provided in the literature but is currently unavailable. Müller et al [31] conducted an evaluation of the Vergiftungsunfalle bei Kindern ("child poisoning accidents") app, developed by the German Federal Institute for Risk Assessment in 2013. This app is designed for caregivers and primarily provides information on toxic substances, prevention advice, and emergency call functions. This app is currently available for download on the Google Play Store and the App Store. The Cool Runnings app, used in the study by Burgess et al [27], was developed by a research team at the University of Queensland's Center for Pediatric Burns and Trauma to raise mothers' awareness of scald injuries from hot beverages and improve their emergency response capabilities; however, the app is no longer being updated and is unavailable for download.

WeChat

Two studies [15,26] used WeChat to provide relevant resources for child caregivers. Feng et al [15] established WeChat groups and an official account to provide caregivers with resources on injury prevention; the WeChat official account registered for this study is currently searchable but is no longer being updated. In the other study, Yan et al [26] provided on-site training to child caregivers at schools and subsequently used WeChat groups to regularly distribute learning materials on child restraint systems.

Mobile-Based e-Learning Program

Choi and Ahn [22] developed an e-learning program using Storyline 360 e-learning software and generated a course link. The course includes videos, audio, and slides, and parents can access the link via computers, tablets, and smartphones to study at their convenience. In addition, the course supports repeat viewing and progress tracking, allowing caregivers to review the material whenever needed. This program is not available to the public.

Mobile-based VR technology and wearable devices: Schwebel et al [30] developed a VR pedestrian safety training system. This system uses a smartphone as both the platform for running VR content and the display device; the smartphone is inserted into a Viewmaster VR stereoscopic display equipped with goggles and a joystick to run a virtual crosswalk environment. Children undergo 12 days of repeated crossing training in the classroom. The system provides immediate feedback based on their crossing performance and automatically records behavioral data such as collisions and unsafe crossings. This technology is currently still in the experimental stage and has not yet been made available to the public.

Study Duration and Follow-Up Periods

The duration of the included studies varied. Specifically, for studies assessing mHealth interventions, the intervention periods ranged from 1 week to 12 months. In contrast, among studies adopting observational or cross-sectional designs, only 2 reported the study duration: 7-10 days [32] and 3 months [16], respectively.

Target Population

Eleven studies [15,16,22,24-28,31-33] focused on caregivers of children, covering those caring for children aged 0 to 12 years; 4 studies [17,23,29,30] targeted children directly, with participants ranging in age from 5 to 13 years; and 1 study [34] was conducted with schoolteachers.

Injury Type Addressed

Nine studies [15,22,24,25,28,29,32-34] addressed unintentional injuries as a whole, including common types such as choking, aspiration, falls, poisoning, drowning, and burns. The remaining 7 studies focused on specific types of injuries, including 2 on traffic accidents [17,26], 1 on burns and scalds [27], 1 on accidental falls among infants [16], 1 on accidental poisoning [31], and 1 on the prevention of bicycle injuries and animal bites [23].

Theoretical Framework

The included studies drew upon behavior change theory [16,23], Orem's self-care theory [29], and the Haddon model [15] to guide the design of learning content; Choi and Ahn [22] designed their learning program based on the analysis, design, development, implementation, and evaluation (ADDIE) model; Wong et al [17] developed a game app by integrating social cognitive theory with an experiential game model; Ning et al [25] designed an app by combining behavior change theory, the Haddon model, and the rational analysis framework for mobile education; Müller et al [31] evaluated the app's perceived usefulness, perceived ease of use, trust, and intention to use based on the technology acceptance model.

Intervention Content

Health Education

A total of 15 studies [15-17,22-29,31-34] used mHealth technologies to deliver health education on unintentional injuries to children and caregivers via text, images, audio, video, and games. The interventions included raising awareness of injuries, changing attitudes or beliefs regarding injury supervision, providing home environment checklists and hazard-identification tools, disseminating knowledge and skills on prevention, and teaching emergency response methods following an incident. The curriculum designed by Choi and Ahn [22] covered 6 scenarios: the bedroom, living room, bathroom, kitchen,

XSL•FO
RenderX

outdoors, and inside a vehicle. Each scenario included prevention knowledge for injury types such as choking, aspiration, falls, poisoning, physical injuries, drowning, and burns. The program also provided information on safety products but did not include purchase links. Ning et al [25] and He et al [28] used the same app to provide education on unintentional injuries to caregivers of urban and rural children, respectively. The Make Safe Happen program used by McKenzie et al [24] and Roberts et al [32] categorizes children by age into 5 groups: 0 to 11 months, 12 to 23 months, 2 to 4 years, 5 to 9 years, and 10 to 12 years. It develops personalized safety information tailored to the types of injuries most common in each age group and provides home environment checklists to help caregivers identify potential hazards. Feng et al [15] conducted a study that used WeChat groups and official accounts to regularly send weekly updates to parents regarding the prevention of common childhood accidents, aiming to enhance their awareness of such injuries. Wong et al [17], Burgess et al [27], and Dixon et al [23] incorporated game-based simulations of relevant scenarios to improve caregivers' or children's awareness of and ability to prevent unintentional injuries. Furthermore, Cooray et al [16] focused on the differences in fall risks between infants and children over 1 year of age. They provided caregivers with customized safety information centered on four themes—safe feeding, furniture use, infant product safety, and the environment—demonstrating a personalized educational strategy tailored to population characteristics, which further highlights the refinement and personalization of health education.

Interactive Communication

Five studies [15,16,23,25,28] provided communication support channels via mHealth technology for caregivers of children. Through these platforms, caregivers can raise questions encountered while implementing accident prevention measures in user forums, discussion areas, or expert consultation channels, and receive insights from experienced caregivers or targeted guidance and answers from health care professionals and researchers. Ning et al [25] focused on facilitating mutual communication among users and between users and researchers. Building on this, He et al [28] optimized expert consultation and customer service functions, added online forums and themed activities, and strengthened interactions among researchers, experts, and caregivers. Feng et al [15] collected caregivers' questions via WeChat groups, with community doctors providing responses within 48 hours to ensure timely guidance. Cooray et al [16] focused on collaborative communication, with researchers and caregivers jointly developing emergency response plans to enable rapid and effective action in the event of an infant fall. Dixon et al [23] encouraged increased communication between children and their caregivers regarding the prevention of bicycle injuries and pet bites.

Behavioral Training

Five studies [17,23,26,27,30] used mHealth technologies—such as WeChat, smartphone-based VR systems, and gaming apps—to deliver injury prevention training to children and caregivers. The content included safe road-crossing skills for children, proper child safety seat installation methods for caregivers, and the identification and management of household safety hazards. Schwebel et al [30] used a smartphone-based VR system to train children's road-crossing behavior, enabling their performance to meet adult standards. Wong et al [17] used a game to simulate road-crossing scenarios and train children's safety behaviors; to prevent game addiction, the study implemented an energy-point mechanism that required a 2-hour wait before restarting once energy was depleted. Dixon et al [23] combined an app with a game to simulate street environments, indoor and outdoor residential settings, animal shelters, and parks, requiring children to learn and practice safe cycling skills and how to interact with animals while responding to various risk scenarios. Burgess et al [27] used gamification techniques to help caregivers master measures to prevent accidental poisoning and burns in children.

Monitoring and Reminders

Six studies [16,24,27,29,32,34] used mobile apps or WeChat to monitor and provide reminders about the safety behaviors or learning progress of children and their caregivers. McKenzie et al [24] and Roberts et al [32] guided caregivers to set up reminders in the app for testing and replacing safety equipment and monitored their implementation. Kim and Cho [29] established a classroom log system within the app, requiring students to record their completion of the previous day's learning tasks daily. Cooray et al [16] used task tracking lists and progress dashboards to show parents the completion status of safety tasks and sent regular SMS text message reminders prompting users to read relevant articles. Burgess et al [27] required caregivers to upload photos of home equipment between push notifications to assess the safety status of the equipment, and to complete pop-up quizzes to reinforce injury prevention knowledge. Kang et al [34] designed a child safety checklist for teachers to use for self-assessment of safety behaviors, with the system providing feedback through colors or markers.

Other Intervention Components

In terms of data collection and user settings, 2 studies [25,28] used mobile platforms to collect data on childhood accidents, survey responses, and user feedback online. Among these, Bao Hu San [25,28] also allowed users to customize the app interface to suit their personal preferences, thereby enhancing user satisfaction. Regarding emergency features, Make Safe Happen [24,32] included a poison control hotline that could be added to a user's phone contacts with a single tap; Jones et al [33] provided emergency phone numbers for caregivers to look up; and Müller et al [31] included a 1-tap button to call a poison control center and provided guidelines for pediatric care to help caregivers seek immediate medical assistance.

Outcome Measures and App Effectiveness

Overview

The studies included in this review used various methods to evaluate the effectiveness of mHealth technology interventions, including researcher-designed questionnaires, participant self-reports, and individual or focus group interviews. Consequently, there were significant variations in the naming and measurement of outcome measures. For ease of presentation, this study categorized these measures into 5 groups.

XSL•FO
RenderX

Incidence of unintentional injuries among children

Three studies [15,25,28] measured the incidence of unintentional injuries in children as the primary outcome, using caregivers' self-reported data for statistical analysis. Among these, He et al [28] and Feng et al [15] both found that the incidence of unintentional injuries in the intervention group was lower than that in the control group after the intervention was completed; however, the incidence of falls in the intervention group led by He et al [28] was slightly higher than that in the control group (adjusted risk ratio 1.17, 95% CI 1.06-1.30); whereas in the 6-month follow-up period reported by Ning et al [25], although the incidence in the intervention group decreased from 8.76% to 8.11%, the difference was not statistically significant (P=.59).

Knowledge, Attitudes, and Behaviors

Knowledge Level

Eight studies [15,17,22-24,26,27,29] measured knowledge level as one of the outcome indicators, primarily including awareness of unintentional injuries [22], knowledge of prevention and safety [17,23,24], knowledge of child restraint system use [26], and first aid knowledge in the event of an accident [15,27,29]. All were assessed using questionnaires designed by the researchers. Among these, the studies by Choi and Ahn [22], Feng et al [15], and Wong et al [17] did not observe any significant statistical differences.

Attitudes

Three studies [25,26,28] used attitude levels as outcome measures, including caregivers' attitudes toward accident prevention [25,28] and the use of child restraint systems [26]. Among these 3 studies, only the study by Ning et al [25] did not show statistically significant differences.

Behavior

Eleven studies [15-17,22-26,28,30,32] used behavioral competence of children and their caregivers as one of the outcome measures, including injury prevention and safety behaviors [17,22-25,32], caregiver supervision behaviors [15,28], traffic safety behaviors [30], use of child restraint systems [26], and behaviors with the potential to reduce the risk of infant falls [16]. Regarding the measurement of this outcome indicator, with the exception of Wong et al [17], who observed no significant differences, all other studies reported statistically significant differences.

Psychosocial Outcomes

Wong et al [17] used the Strengths and Difficulties Questionnaire to assess psychosocial problems in children. While the study did not observe statistically significant changes in these outcomes following the intervention, an analysis of the dose-response relationship revealed that cumulative game scores were associated with improvements in internalizing problems at the 3-month mark of the intervention, suggesting that serious games can enhance children's psychosocial functioning. Feng et al [15] assessed improvements in caregivers' beliefs regarding the prevention of unintentional injuries following the intervention. This indicator was divided into 3 components based on relevant literature: injury attribution beliefs, preventability beliefs, and responsibility beliefs. The results showed that mHealth technology can improve caregivers' beliefs regarding the preventability of unintentional injuries (β=.344, 95% CI 0.152-0.537; P<.001). Schwebel et al [30] measured children's self-efficacy when crossing the street using both questionnaires and behavioral observations. The study showed that the proportion of children able to safely cross busy streets increased from 3.6% to 19.6% (odds ratio [OR] 4.7; P=.002).

Other Indicators

Other indicators included home environment safety (β=3.38, 95% CI 2.51-4.16; P<.001) [28], the use rate of child restraint systems (OR 0.29, 95% CI 0.11-0.79; P=.016) [26], and caregivers' awareness of injury prevention (which increased significantly from 63% before the intervention to 81%, P<.01) [32]. Additionally, 2 studies [16,32] identified barriers and facilitators to caregivers' adoption of unintentional injury prevention measures through focus group interviews. Both studies indicated that simplifying tasks and increasing app reminders and feedback features could promote the implementation of injury prevention measures; conversely, incomplete functional content, impractical recommendations, time constraints, high costs, cumbersome tasks, and a lack of family support hindered their implementation.

Usability of mHealth Technologies

Eleven studies [15-17,22-26,28,30,32] reported metrics related to mHealth technologies, primarily including user satisfaction [22,26,34], user engagement [16,23,27,32,33], user experience [30-33], user acceptance [16,23], the educational effectiveness of smartphones (frequency, practicality, interest, and ease of understanding) [29], economic losses resulting from childhood accidents [25], and the economic costs of the apps [15,25], as well as the identification of facilitators and barriers affecting user acceptance through targeted individual interviews [31]. The specific results are presented in Multimedia Appendix 3.

Discussion

Overview of mHealth Technologies for Unintentional Childhood Injuries

The results of this study indicate that mHealth technologies currently used in the field of unintentional childhood injuries include mobile apps, WeChat, mobile-based e-learning programs, VR technology, and wearable devices. Apps and WeChat are primarily targeted at child caregivers, as their busy daily schedules make it difficult to set aside time for in-person training. Apps and WeChat platforms offer convenience and quick access to health information, helping caregivers learn in spare moments or obtain necessary information promptly when needed. Among the mobile apps included in this study, with the exception of the 3 apps cited by Wong et al [17], Dixon et al [23], and Kim and Cho [29], all others were designed for child caregivers, with one specifically targeting teachers. In terms of functional design, health education or regular learning content is primarily delivered through audio, video, text, images, and gamified formats; user forums or comment sections facilitate interaction between users and researchers or health care professionals; check-in features and progress bars are used to monitor users' learning progress and adherence to safety

XSL•FO
RenderX

measures; additionally, some apps include emergency contact information or links to purchase safety products. The results of this study indicate that participants demonstrated high levels of engagement, acceptance, satisfaction, and positive user experience with the apps they used. For children, games and VR technology are currently widely adopted. This study included 2 gaming apps, one VR app combined with wearable devices, and one health education-focused app. Key features included simulations of real-world environments such as streets, residential areas, dog shelters, and parks, designed to help children develop appropriate safety behaviors and prevent traffic accidents and pet bites. The health education-focused app provides children with emergency response knowledge for various situations, covering 8 categories of emergency scenarios: school safety, off-campus safety, responding to violent incidents, traffic accidents, accidental poisoning, basic first aid skills, natural disasters, and evacuation drills—totaling 48 specific emergencies. Each scenario includes a brief description and response recommendations, with some supplemented by instructional videos. Compared with traditional lecture-style education, children demonstrate greater acceptance of this engaging, interactive learning method.

Positive Effects and Development Requirements

This study describes the effectiveness of mHealth-based interventions in preventing unintentional injuries among children. The results indicate that such interventions have a positive impact on reducing the incidence of unintentional injuries among children, enhancing the safety knowledge of children and caregivers, and improving their safety attitudes and behaviors. However, among the mHealth technologies currently used for interventions targeting unintentional injuries in children, only the Vergiftungsunfälle bei Kindern and Safe City apps are available for download and use via app stores or web browsers. The remaining technologies are either not publicly released or have ceased to be updated and maintained. The development and long-term maintenance of mHealth apps require significant investments of time, effort, and funding, while the actual social and health benefits are relatively limited, making it difficult for many projects to sustain operations [35]. Furthermore, mHealth technologies can overcome spatial and temporal constraints and address the shortfall of caregivers who are unable to participate in in-person training due to time constraints through online interventions. The online intervention model itself may experience low user adherence, with some users potentially losing motivation to continue participating after initial use [36]. Although this study did not assess user adherence, based on the dropout rates reported by Feng et al (24.38%) [15], Ning et al (32.19%) [25], and Burgess et al (51%) [27], long-term user adherence may be relatively low, which could be one of the reasons why mHealth programs fail to operate sustainably over the long term. The World Health Organization advocates that countries incorporate mHealth into their national public health strategies [37]. The “Action Plan for Injury Prevention and Control (2026-2030)” issued by China’s National Center for Disease Control and Prevention also emphasizes exploring and promoting appropriate technologies and strengthening education on child injury prevention and control skills, providing clear policy guidance for mHealth interventions. Therefore, it is recommended that future research leverage national public health programs to conduct long-term, large-scale studies on mHealth technologies. Relying on the public health system to overcome cost and adherence barriers can promote the sustainable application of mHealth interventions.

Outcome Measures and Assessment Tools

Regarding the evaluation of the effectiveness of mHealth technologies, most studies have focused on caregivers’ knowledge, attitudes, and behaviors as primary outcome measures, while only 3 studies have primarily examined the incidence of unintentional injuries in children. The occurrence of unintentional injuries in children is influenced by multiple factors, including the home environment, the level of supervision, and the child’s developmental stage. To detect statistically significant differences, large sample sizes and long-term follow-up may be required, resulting in high research costs and significant implementation challenges [38,39]. Most of the studies included in this review had sample sizes ranging from 100 to 200 participants, with intervention durations of approximately 1 week to 3 months, which may not meet the basic requirements for conducting effective statistical tests for this outcome. Furthermore, regarding the selection of assessment tools, the studies included in this review all used self-developed or adapted questionnaires. Since the data rely on self-reports from caregivers, there is potential for bias and subjectivity, which may lead to a lack of comparability across study results [35,40]. Regarding the assessment of mHealth technology usability metrics, the literature included in this study generally reported favorable evaluations of these metrics by children and their caregivers across multiple dimensions, including user acceptance, engagement, user experience, and satisfaction. However, these evaluations were also largely based on assessment tools and criteria developed by the researchers themselves, lacking uniformity and comparability. Only the studies by Kang et al [34] and Müller et al [31] conducted standardized assessments of app usability and user satisfaction using the validated User Interaction Satisfaction Scale and the System Usability Scale, respectively. Currently, the academic community has developed a variety of assessment tools with good reliability and validity for the systematic evaluation of the usability of mHealth technologies. Therefore, it is recommended that future studies adopt these standardized, validated assessment questionnaires wherever possible to enhance the scientific rigor and comparability of research findings [41].

Limitations

This study is a scoping review, so the quality of the included literature was not assessed. Additionally, the literature search was limited to English-language publications, which may have led to incomplete coverage of studies from certain regions or specific populations. Finally, this study focused on research that has been officially published and contains complete methodological information; gray literature, which often exists in the form of preliminary abstracts or unpublished data, usually does not meet the inclusion criteria for this review, so it was not systematically searched.

XSL•FO
RenderX

Conclusions

This study provides a scoping review of the current use of mHealth technologies for the prevention and control of unintentional injuries among children. The findings indicate that mHealth technologies play a positive role in improving the knowledge, attitudes, and behaviors of children and their caregivers regarding unintentional injuries. However, only 2 mobile apps are currently available to the public. It is recommended that future efforts continue to optimize the functional design of mHealth, strengthen the technology's sustainability, explore methods that can be implemented over the long term, and conduct large-scale, long-term studies to further evaluate the impact of mHealth on the incidence of unintentional injuries among children.

Funding

This study is jointly supported by 3 research projects, 2 of which are funded by the Mianyang Municipal Federation of Social Sciences: the Special Research Project on Vocational Education, named Job Content Analysis of Infection Prevention and Control Nurses in Northwestern Sichuan and Development of a Competency-Based Training Program (grant MY2025ZX686), and the General Project, named A Survey on the Current Status of Family First-Aid Competence for Infant Unintentional Injuries in Mianyang and a Progressive Intervention Strategy Based on Social Cognitive Theory (grant MY2026YB151). The third project is funded by the Chuanbei Health Humanities Research Program under the Nanchong Municipal Social Science Planning Project (grant NC26CB16).

Authors' Contributions

Conceptualization: HW
Methodology: HW, XL, HL, ML
Formal analysis: HW, HL, ML
Data curation: FP, DJ
Supervision: XL
Writing—original draft: HW
Writing—review and editing: XL, HL, ML
All authors have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR checklist.
[PDF File (Adobe PDF File), 155 KB - mhealth_v14i1e94038_app1.pdf]

Multimedia Appendix 2
Search strategies.
[DOCX File , 15 KB - mhealth_v14i1e94038_app2.docx]

Multimedia Appendix 3
Basic characteristics of included studies.
[DOCX File , 33 KB - mhealth_v14i1e94038_app3.docx]

References

1. International Statistical Classification of Diseases and Related Health Problems 10th Revision. World Health Organization. 2019. URL: https://icd.who.int/browse10/2019/en [accessed 2026-07-21]
2. GBD 2021 Causes of Death Collaborators. Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2024;403(10440):2100-2132 [FREE Full text] [doi: 10.1016/S0140-6736(24)00367-2] [Medline: 38582094]
3. DeGeorge KC, Neltner CE, Neltner BT. Prevention of unintentional childhood injury. Am Fam Physician 2020;102(7):411-417 [FREE Full text] [Medline: 32996759]
4. Yu X, Wang Y, He C, Kang L, Miao L, Wu Y, et al. The trend of unintentional injury-related mortality among children aged under-five years in China, 2010-2020: a retrospective analysis from a national surveillance system. BMC Public Health 2023;23(1):673 [FREE Full text] [doi: 10.1186/s12889-023-15546-6] [Medline: 37041562]

XSL•FO
RenderX

5. Foettinger L, Doerwald F, Kalbitz A, Bammann K. Risk factors and parental risk perception of unintentional home injuries of children under 6 years in Germany: a secondary data analysis. J Safety Res 2022;81:326-332. [doi: 10.1016/j.jsr.2022.03.009] [Medline: 35589303]
6. Ji J, Yang H, Zeng C, Chen O, Lu Q. Bridging the gap: parental supervision as a mediator between home environment and unintentional injuries in children under 3 years. Int J Public Health 2025;70:1606726 [FREE Full text] [doi: 10.3389/ijph.2025.1606726] [Medline: 40051470]
7. Xie Y, Yu X, Wu X, Zhang W, Feng Z, Xiao F, et al. Association between the guardians' educational levels and unintentional injuries in children aged 6-18 in Shenzhen, China. BMC Public Health 2024;24(1):2344 [FREE Full text] [doi: 10.1186/s12889-024-19748-4] [Medline: 39210286]
8. World report on child injury prevention. World Health Organization. URL: https://www.who.int/publications/i/item/9789241563574 [accessed 2026-07-21]
9. mHealth: new horizons for health through mobile technologies. World Health Organization. URL: https://www.afro.who.int/publications/mhealth-new-horizons-health-through-mobile-technologie [accessed 2026-02-10]
10. Deniz-Garcia A, Fabelo H, Rodriguez-Almeida AJ, Zamora-Zamorano G, Castro-Fernandez M, Alberiche Ruano MDP, WARIFA Consortium. Quality, usability, and effectiveness of mHealth apps and the role of artificial intelligence: current scenario and challenges. J Med Internet Res 2023;25:e44030 [FREE Full text] [doi: 10.2196/44030] [Medline: 37140973]
11. Ameyaw EK, Amoah PA, Ezezika O. Effectiveness of mHealth apps for maternal health care delivery: systematic review of systematic reviews. J Med Internet Res 2024;26:e49510 [FREE Full text] [doi: 10.2196/49510] [Medline: 38810250]
12. Chen X, Liu F, Lin S, Yu L, Lin R. Effects of virtual reality rehabilitation training on cognitive function and activities of daily living of patients with poststroke cognitive impairment: a systematic review and meta-analysis. Arch Phys Med Rehabil 2022;103(7):1422-1435. [doi: 10.1016/j.apmr.2022.03.012] [Medline: 35417757]
13. Park SH, Lee H. Comparing the effects of home visits and telenursing on blood glucose control: a systematic review of randomized controlled trials. Int J Nurs Stud 2023;148:104607. [doi: 10.1016/j.ijnurstu.2023.104607] [Medline: 37839308]
14. Wu W, Graziano T, Salner A, Chen M, Judge MP, Cong X, et al. Acceptability, effectiveness, and roles of mHealth applications in supporting cancer pain self-management: integrative review. JMIR Mhealth Uhealth 2024;12:e53652 [FREE Full text] [doi: 10.2196/53652] [Medline: 39024567]
15. Feng Y, Ma X, Zhang Q, Jiang R, Lu J, Chen K, et al. Effectiveness of WeChat-group-based parental health education in preventing unintentional injuries among children aged 0-3: randomized controlled trial in Shanghai. BMC Public Health 2022;22(1):2086 [FREE Full text] [doi: 10.1186/s12889-022-14462-5] [Medline: 36380326]
16. Cooray N, Ho C, Bestman A, Adams S, Nassar N, Keay L, et al. Exploring the potential of a behavior theory-informed digital intervention for infant fall prevention: mixed methods longitudinal study. JMIR Pediatr Parent 2024;7:e47361 [FREE Full text] [doi: 10.2196/47361] [Medline: 38170580]
17. Wong RS, Tung KTS, Ho FKW, Wong WHS, Chow CB, Chan KL, et al. Effect of a mobile game-based intervention to enhance child safety: randomized controlled trial. J Med Internet Res 2024;26:e51908 [FREE Full text] [doi: 10.2196/51908] [Medline: 38354042]
18. Schulze A, Lindemann AK, Brand F, Geppert J, Menning A, Stehr P, et al. Interact J Med Res 2023;12:e45258 [FREE Full text] [doi: 10.2196/45258] [Medline: 37672312]
19. Shetty S, Nayak BS, George A, Shetty A, Guddattu V. Evidence of interventions for the prevention of unintentional injuries: scoping review. JMIR Pediatr Parent 2025;8:e67877 [FREE Full text] [doi: 10.2196/67877] [Medline: 40294437]
20. Peters MDJ, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Implement 2021;19(1):3-10. [doi: 10.1097/XEB.0000000000000277] [Medline: 33570328]
21. Global Observatory for eHealth. World Health Organization. URL: https://www.who.int/observatories/global-observatory-for-ehealth [accessed 2026-06-18]
22. Choi Y, Ahn HY. Developing and evaluating a mobile-based parental education program for preventing unintentional injuries in early childhood: a randomized controlled trial. Asian Nurs Res (Korean Soc Nurs Sci) 2021;15(5):329-336 [FREE Full text] [doi: 10.1016/j.anr.2021.12.001] [Medline: 34915186]
23. Dixon CA, Ammerman RT, Johnson BL, Lampe C, Hart KW, Lindsell CJ, et al. A randomized controlled field trial of iBsafe-a novel child safety game app. Mhealth 2019;5:3 [FREE Full text] [doi: 10.21037/mhealth.2019.01.02] [Medline: 30842951]
24. McKenzie LB, Roberts KJ, McAdams RJ, Abdel-Rasoul M, Kristel O, Szymanski A, et al. Efficacy of a mobile technology-based intervention for increasing parents' safety knowledge and actions: a randomized controlled trial. Inj Epidemiol 2021;8(1):56 [FREE Full text] [doi: 10.1186/s40621-021-00350-w] [Medline: 34593040]
25. Ning P, Cheng P, Schwebel DC, Yang Y, Yu R, Deng J, et al. An app-based intervention for caregivers to prevent unintentional injury among preschoolers: cluster randomized controlled trial. JMIR Mhealth Uhealth 2019;7(8):e13519 [FREE Full text] [doi: 10.2196/13519] [Medline: 31400105]
26. Yan S, Yang J, Fu J, Ding K, Ye W, Chen X, et al. Assessing an app-based child restraint system use intervention in China: an RCT. Am J Prev Med 2020;59(3):e141-e147. [doi: 10.1016/j.amepre.2020.02.003] [Medline: 32334955]

XSL•FO
RenderX

27. Burgess J, Watt K, Kimble RM, Cameron CM. Combining technology and research to prevent scald injuries (the Cool Runnings Intervention): randomized controlled trial. J Med Internet Res 2018;20(10):e10361 [FREE Full text] [doi: 10.2196/10361] [Medline: 30305263]
28. He J, Zhao M, Li J, Schwebel DC, Fu Y, Wang W, et al. App-based prevention of preschooler unintentional injury in rural China: a cluster randomized controlled trial. NPJ Digit Med 2025;8(1):760 [FREE Full text] [doi: 10.1038/s41746-025-02134-8] [Medline: 41398035]
29. Kim SJ, Cho H. The effect of smartphone-delivered emergency preparedness education on coping knowledge among fifth- and sixth-grade elementary schoolchildren in South Korea. J Sch Nurs 2017;33(6):434-445. [doi: 10.1177/1059840516680267] [Medline: 27884947]
30. Schwebel DC, Wu Y, Li P, Severson J, He Y, Xiang H, et al. Featured article: evaluating smartphone-based virtual reality to improve Chinese schoolchildren's pedestrian safety: a nonrandomized trial. J Pediatr Psychol 2018;43(5):473-484 [FREE Full text] [doi: 10.1093/jpepsy/jsx147] [Medline: 29216384]
31. Müller P, Schulze A, Geppert J, Menning A, Brand F, Stehr P, et al. Poisoning accidents in young children-theory-based evaluation of an mHealth app. Digit Health 2025;11:20552076251362753 [FREE Full text] [doi: 10.1177/20552076251362753] [Medline: 40777834]
32. Roberts KJ, McAdams RJ, Kristel OV, Szymanski AM, McKenzie LB. Qualitative and quantitative evaluation of the Make Safe Happen app: mobile technology-based safety behavior change intervention for parents. JMIR Pediatr Parent 2019;2(1):e12022 [FREE Full text] [doi: 10.2196/12022] [Medline: 31518322]
33. Jones F, Whitehouse A, Dopson A, Palaghias N, Aldiss S, Gibson F, et al. Reducing unintentional injuries in under fives: development and testing of a mobile phone app. Child Care Health Dev 2020;46(2):203-212. [doi: 10.1111/cch.12729] [Medline: 31782175]
34. Kang KA, Kim SJ, Kang SR, Lee SH, Kim YY, Ellis KW. Development and preliminary testing of a smartphone-based injury-prevention application (S-IPA) for teachers at child-care centers in South Korea. J Community Health Nurs 2017;34(3):147-159. [doi: 10.1080/07370016.2017.1340767] [Medline: 28767294]
35. Subramaniam A, Hensley E, Stojancic R, Vaughn J, Shah N. Careful considerations for mHealth app development: lessons learned from QuestExplore. Mhealth 2022;8:24 [FREE Full text] [doi: 10.21037/mhealth-21-51] [Medline: 35928515]
36. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
37. Global Strategy on Digital Health 2020-2025. World Health Organization. URL: https://www.who.int/publications/i/item/9789240020924 [accessed 2026-07-21]
38. Gallagher L, Breslin G, Leavey G, Curran E, Rosato M. Determinants of unintentional injuries in preschool age children in high-income countries: a systematic review. Child Care Health Dev 2024;50(1):e13161. [doi: 10.1111/cch.13161] [Medline: 37555597]
39. Reading R, Harvey I, McLean M. Cluster randomised trials in maternal and child health: implications for power and sample size. Arch Dis Child 2000;82(1):79-83 [FREE Full text] [doi: 10.1136/adc.82.1.79] [Medline: 10630921]
40. Ghebreab L, Kool B, Lee A, Morton S. Comparing primary caregivers' reported injury data with routinely recorded injury data to assess predictors of childhood injury. BMC Med Res Methodol 2023;23(1):91 [FREE Full text] [doi: 10.1186/s12874-023-01900-0] [Medline: 37041484]
41. Hajesmaeel-Gohari S, Khordastan F, Fatehi F, Samzadeh H, Bahaadinbeigy K. The most used questionnaires for evaluating satisfaction, usability, acceptance, and quality outcomes of mobile health. BMC Med Inform Decis Mak 2022;22(1):22 [FREE Full text] [doi: 10.1186/s12911-022-01764-2] [Medline: 35081953]

Abbreviations

ADDIE: Analysis, Design, Development, Implementation, and Evaluation
mHealth: mobile health
OR: odds ratio
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
VR: virtual reality

XSL•FO
RenderX

Edited by JN Siebert; submitted 24.Feb.2026; peer-reviewed by P Stehr, C Morgan, A Hamdan; comments to author 13.Apr.2026; accepted 06.Jul.2026; published 14.Aug.2026.

Please cite as:
Wang H, Pu F, Jing D, Liu H, Li M, Li X
Application of Mobile Health Technologies to Prevent Unintentional Childhood Injuries: Scoping Review
JMIR Mhealth Uhealth 2026;14:e94038
URL: https://mhealth.jmir.org/2026/1/e94038
doi:10.2196/94038
PMID:

©Huan Wang, Fengjiao Pu, Dongjun Jing, Hangcheng Liu, Miao Li, Xixi Li. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Mechanical Contact Conditions in Wearable Reflectance Photoplethysmography: Scoping Review

Chenxi Yang[1,2], PhD; Jiahang Xie[1], BEng; Zifei He[1], BEng; Jianqing Li[1], PhD; Chengyu Liu[1,2], PhD

[1]School of Instrument Science and Engineering, Southeast University, Nanjing, Jiangsu, China

[2]State Key Laboratory of Digital Medical Engineering, Southeast University, Nanjing, Jiangsu, China

Corresponding Author:
Chenxi Yang, PhD
School of Instrument Science and Engineering
Southeast University
Number 2 Sipailou
Nanjing, Jiangsu, 210009
China
Phone: 86 15370025236
Email: chenxiyang@seu.edu.cn

Abstract

Background: Cardiovascular diseases remain a major global health burden, highlighting the need for long-term physiological monitoring. Photoplethysmography (PPG) is widely used in wearable devices for noninvasive monitoring of heart rate (HR), rhythm, and oxygen saturation in mobile health (mHealth) apps. However, the reliability of wearable reflectance PPG depends on sensing conditions, including sensor-skin contact force and pressure.

Objective: This scoping review maps how contact force and contact pressure have been defined, controlled, measured, represented, and reported in wearable or wearable-relevant reflectance PPG studies; characterizes the reported signal-, waveform-, feature-, and task-level responses under different contact conditions; and identifies methodological and evidence gaps.

Methods: A comprehensive literature search was conducted in PubMed, IEEE Xplore, Scopus (Elsevier), and Web of Science Core Collection (Clarivate) from database inception to June 1, 2026. Studies were eligible if they addressed contact force or contact pressure in wearable or wearable-relevant reflectance PPG and reported measurement approaches, measurement sites and device configurations, force or pressure representation, or PPG responses across signal quality, waveform and feature characteristics, and downstream physiological estimation. Database filters were applied, where available, to restrict results to English-language publications. Search results were imported into EndNote for deduplication. Database searches were supplemented by reference-list screening. After screening, 53 reports were sought for retrieval; 1 was not retrieved, 52 were assessed at full text, and 21 studies met the inclusion criteria.

Results: The 21 included studies showed substantial heterogeneity, with sample sizes ranging from single-participant experiments to a wrist PPG dataset including 1142 participants. Most human studies recruited healthy volunteers, whereas some used public datasets or tissue-vessel phantoms, and 1 combined theoretical modeling with human-participant validation. The mapped evidence identified contact force and contact pressure as important measurement conditions in wearable reflectance PPG. Across the included studies, different contact conditions were associated with changes in alternating current/direct current components, amplitude- and morphology-related features, derivative-based indices, wavelength-dependent responses, and fiducial-point detection. Several studies also examined downstream physiological tasks, including HR, oxygen saturation, pulse transit or arrival time, blood pressure–related estimates, and HR variability.

Conclusions: The evidence mapped in this scoping review supports a multilevel conceptual pathway in wearable reflectance PPG, in which mechanical conditions at the sensor-skin interface are associated with changes in PPG signal characteristics, derived features, and, in a smaller body of studies, downstream physiological estimation. Evidence remains limited by small samples, short-term controlled protocols, and inconsistent reporting of mechanical parameters, including units, contact area, and probe geometry. Insufficient population diversity and limited free-living validation further restrict generalizability. Future research should standardize reporting of contact conditions and device geometry, incorporate real-world validation, and develop force-aware signal-quality assessment algorithms, adaptive attachment designs, and context-aware models to improve mHealth and cardiovascular monitoring.

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2026;14:e99333) doi:10.2196/99333

KEYWORDS

contact force; contact pressure; wearable photoplethysmography; reflectance photoplethysmography; signal quality; waveform morphology; artifact susceptibility; temporal stability; physiological parameter estimation

Introduction

Background

Cardiovascular diseases remain a major cause of morbidity, mortality, and health care burden worldwide [1,2]. As many cardiovascular conditions may remain asymptomatic in their early stages, timely risk identification and intervention are critical for improving outcomes and preventing disease progression [3,4]. However, current screening and diagnostic approaches still largely depend on professional medical settings and centralized equipment, limiting their accessibility for long-term follow-up and routine management. Therefore, scalable and low-cost digital health technologies for continuous monitoring are increasingly important in cardiovascular disease prevention and management [5]. In this context, photoplethysmography (PPG) has become one of the most widely used sensing modalities in digital health and mobile health (mHealth) apps because of its noninvasive nature, convenience, and suitability for long-term use.

By optically detecting pulsatile blood volume changes, PPG can support the assessment of heart rate (HR), oxygen saturation, blood perfusion, and certain vascular function–related indices [6]. With advances in sensors, wearable devices, and artificial intelligence [7], PPG has expanded from traditional pulse and oxygen saturation monitoring to broader applications such as HR monitoring [8], blood pressure (BP) assessment, oxygen saturation estimation [9], and cardiovascular health monitoring. With the rapid development of wearable technologies, PPG has been widely integrated into smartwatches, wristbands, patches, and other devices for continuous and noninvasive health monitoring [10-12]. In homes, communities, and other nonclinical settings, these wearable PPG systems offer practical advantages for remote monitoring and proactive health management, highlighting their strong potential for large-scale cardiovascular health assessment [5].

At the same time, the effectiveness of PPG in wearable applications is fundamentally constrained by signal quality, which in turn is highly dependent on specific sensing conditions, including sensor geometry, the skin-sensor interface, ambient light, contact force, and measurement-site characteristics [13,14]. Previous studies have identified contact force and contact pressure as relevant interface conditions associated with differences in sensor-skin coupling, local vascular loading, waveform quality, and downstream interpretability [13,15]. However, contact force has often been treated as a secondary issue. In practice, sensor-skin pressure is not constant and changes with wearing conditions. Too little contact force weakens mechanical coupling, facilitates relative motion, and increases susceptibility to ambient light and random noise. Excessive contact force can compress superficial vessels and capillary beds, alter local microcirculation, suppress pulsatile blood-volume changes, and distort the waveform [16]. Across analytical levels, contact conditions have been reported in relation to differences in alternating current/direct current (AC/DC) components, waveform morphology, signal-quality indices, and feature stability [16,17]. In wearable settings, these differences have also been examined alongside posture, motion, tissue deformation, strap fit, and downstream tasks, including peripheral oxygen saturation (SpO_2) estimation [15,18,19], heart-rate monitoring, heart rate variability (HRV) analysis, pulse arrival time (PAT)/pulse transit time (PTT) extraction, blood pressure modeling, and other PPG-driven applications [15,20,21].

Related Reviews

To position this scoping review within the existing literature, we reviewed representative PPG-related reviews published over the past decade. Existing reviews have addressed PPG from several perspectives, including its physiological basis, waveform characteristics, signal processing methods, and applications in HR, BP, oxygen saturation, and arrhythmia detection [7-10,22-29]. Other reviews have focused on wearable PPG devices, summarizing sensor structures, measurement sites, device forms, system architectures, and application scenarios [30-34]. In addition, some reviews have discussed PPG signal analysis from the perspectives of signal quality, artifact sources, dataset resources, and algorithmic tools [7,29,35-37]. Overall, these reviews have mainly approached PPG from the perspectives of signal technology, wearable implementation, signal quality, or downstream physiological applications. Existing reviews have acknowledged contact pressure, wearing tightness, and sensor-skin interface stability as factors related to PPG measurement quality [29-32], but these issues have generally been treated as secondary acquisition considerations within broader discussions of wearable devices, signal quality, or physiological applications. As a result, limited attention has been given to how contact force and contact pressure are defined, measured, controlled, represented, and reported across wearable reflectance PPG studies. This scoping review addresses this gap by using contact force and contact pressure as the central organizing concepts and mapping measurement and control approaches, anatomical sites, device configurations, reporting practices, and corresponding PPG responses across signal quality, waveform characteristics, derived features, and downstream physiological estimation. It further examines the distribution of evidence across populations, experimental contexts, and protocol durations to clarify the methodological structure of the field and identify areas in which evidence remains limited.

Objectives

This scoping review aimed to map and characterize the existing literature on contact force and contact pressure in wearable reflectance PPG. The review was guided by the following research questions: (1) How have contact force and contact

XSL•FO
RenderX

pressure been defined, applied, controlled, measured, estimated, and reported in wearable reflectance PPG studies? (2) Which signal-level, waveform-level, feature-level, signal-quality, and task-level outcomes have been investigated under different contact conditions? (3) How is the available evidence distributed across measurement sites, sensor configurations, study populations, experimental settings, and outcome categories? In addition, this review aimed to identify current evidence gaps and methodological challenges to inform future force-aware wearable PPG acquisition, sensor-skin interface design, algorithm development, and standardization in wearable cardiovascular monitoring.

Methods

Protocol and Registration

No review protocol was registered or publicly published for this scoping review. The eligibility criteria, information sources, search strategy, study-selection procedures, and initial data-charting framework were established before full-text screening. The data-charting categories were subsequently refined iteratively during data extraction as familiarity with the terminology, methodological diversity, and outcome categories of the included studies increased. The final review methods are reported in this section and Multimedia Appendices 1-3.

Study Design

This study used a scoping review methodology to address the breadth and heterogeneity of the available evidence on contact force and contact pressure in wearable reflectance PPG. The review mapped how these contact conditions were defined, measured, and reported, together with the associated PPG outcomes and evidence gaps. The review was conducted following the framework proposed by Arksey and O'Malley [38] and was reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist (Multimedia Appendix 1) [39]. In addition, the literature search was reported according to the PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension; Multimedia Appendix 2) [40] checklist to strengthen the transparency and reproducibility of the search process.

Eligibility Criteria

Inclusion Criteria

Peer-reviewed journal articles and conference proceedings published in English were considered for inclusion. Eligible studies were required to examine or report contact force, contact pressure, or related sensor-skin interface conditions in wearable or wearable-relevant reflectance PPG and to provide relevant evidence on methodological characteristics, including measurement sites and device configurations, and PPG outcomes across signal-quality, waveform, feature, artifact-susceptibility, and downstream-task domains. Specifically, studies were required to meet all 3 core criteria: (1) the study must involve contact force, contact pressure, or sensor-skin interface conditions (including direct measurement, controlled application, or qualitative description); (2) the study must use wearable or wearable-relevant reflectance-mode PPG systems; and (3) the study must report at least one PPG-related methodological characteristic or outcome that was observed under, compared across, or interpreted in relation to the contact condition. In this review, contact force refers to the total mechanical load applied at the sensor-skin interface, whereas contact pressure refers to the distribution of that load over the effective contact area. Terms and units were retained as reported by the source studies and were not treated as interchangeable unless sufficient contact-area or geometric information was available.

Exclusion Criteria

Non–peer-reviewed publications, non-English articles, and studies without accessible full text were excluded. Studies were further excluded if they met 1 or more of the following criteria: (1) did not address contact force, contact pressure, or sensor-skin interface–related factors in the context of PPG signal acquisition or analysis; (2) focused solely on transmission-mode PPG systems without clear relevance to wearable or wearable-relevant applications; or (3) were review articles, editorials, commentaries, conference abstracts, or other nonoriginal research publications.

Information Sources and Search

To identify relevant studies, a comprehensive literature search strategy was designed by the review authors and further refined through iterative discussion. The search strategy was developed with reference to PRISMA-S reporting principles and methodological recommendations from Cochrane and Joanna Briggs Institute, emphasizing the combined use of controlled vocabulary terms, free-text terms, field modifiers, and a broad range of synonyms. The search strategy was built around 3 concepts: PPG and wearable reflectance PPG; contact force/contact pressure and the sensor-skin interface; and signal quality, waveform morphology, feature extraction, and downstream physiological estimation. The strategy combined controlled vocabulary terms, where available, with free-text keywords and was first developed for PubMed before being translated to the other databases with syntax-specific adaptations. The search was conducted in PubMed, IEEE Xplore, Scopus (Elsevier), and Web of Science Core Collection (Clarivate), accessed through the university library, covering records from database inception to June 1, 2026. Search terms included concepts related to PPG, wearable-device applications, contact conditions, and relevant signal- and task-level outcomes. Database filters were applied, where available, to restrict results to English-language publications. Review articles, editorials, commentaries, conference abstracts, and other nonoriginal publications were excluded during screening because this scoping review aimed to map original research evidence. The complete database-specific search strategies are reported in Multimedia Appendix 3. All search results were exported as RIS files and imported into EndNote (Clarivate Analytics) for reference management and deduplication. Duplicate records were first identified using EndNote's automated duplicate-detection function and then checked manually. The database searches were supplemented by backward citation searching of the reference lists of included studies. Potentially relevant reports were cross-checked against the original database

XSL•FO
RenderX

exports using DOI and title information. No formal peer review of the search strategy by an information specialist or librarian was performed; however, the search strategy was iteratively reviewed by the author team and checked against PRISMA-S reporting items.

Selection of Sources of Evidence

All retrieved records were imported into EndNote for reference management, and duplicate records were removed before screening. Title and abstract screening was independently conducted by 2 reviewers (JX and ZH) to identify potentially relevant studies. Each record was assessed by both reviewers, and studies deemed relevant were subsequently subjected to full-text review. Full-text articles were independently evaluated by the same 2 reviewers according to the predefined inclusion and exclusion criteria. Any disagreements arising during the title and abstract screening or full-text review stages were resolved through discussion; if consensus could not be reached, a third reviewer (CY) was consulted for final adjudication. The reference-list entries identified through citation searching were screened by the same reviewers, and potentially relevant reports were assessed using the same inclusion and exclusion criteria applied to reports identified through the database searches. Reports were assigned to the citation-searching pathway only after confirming that they were not represented in the original database exports. Given the scoping nature of this review, the selection process aimed to map the available evidence broadly and transparently according to the predefined eligibility criteria.

Data Charting Process and Data Items

Relevant data were extracted from the included studies using a standardized data-charting approach. Extracted items covered study and publication characteristics; participant or experimental-platform characteristics; measurement sites and device configurations; contact-condition terminology, measurement or control approaches, and force or pressure representation; experimental protocols; key methodological reporting details; and reported PPG outcomes and main findings across signal and signal-quality, waveform, feature, artifact-susceptibility, and downstream-task domains. The data-charting process was iterative, allowing refinement of categories as familiarity with the literature increased. Information unavailable in the source articles was recorded as not reported.

Critical Appraisal of Individual Sources of Evidence

A formal critical appraisal of individual sources of evidence was not conducted because this scoping review aimed to map the breadth, characteristics, and methodological heterogeneity of the available evidence rather than to assess the certainty of the evidence or generate pooled effect estimates.

Synthesis of Results

To address the review questions, the charted data were first summarized using descriptive numerical analysis. Frequencies and percentages were calculated for key study, methodological, and outcome characteristics, with categories treated as nonmutually exclusive where appropriate. The individual sources of evidence were subsequently analyzed and synthesized narratively to compare methodological approaches, study contexts, and reported PPG outcomes across contact conditions. Finally, cross-study synthesis was conducted using evidence mapping, study-level linkage analysis, and study-level methodological and analytical coverage mapping to characterize the distribution, connections, and breadth of the available evidence. Owing to substantial heterogeneity in study designs, contact-condition definitions, measurement protocols, and outcome measures, a meta-analysis was not conducted.

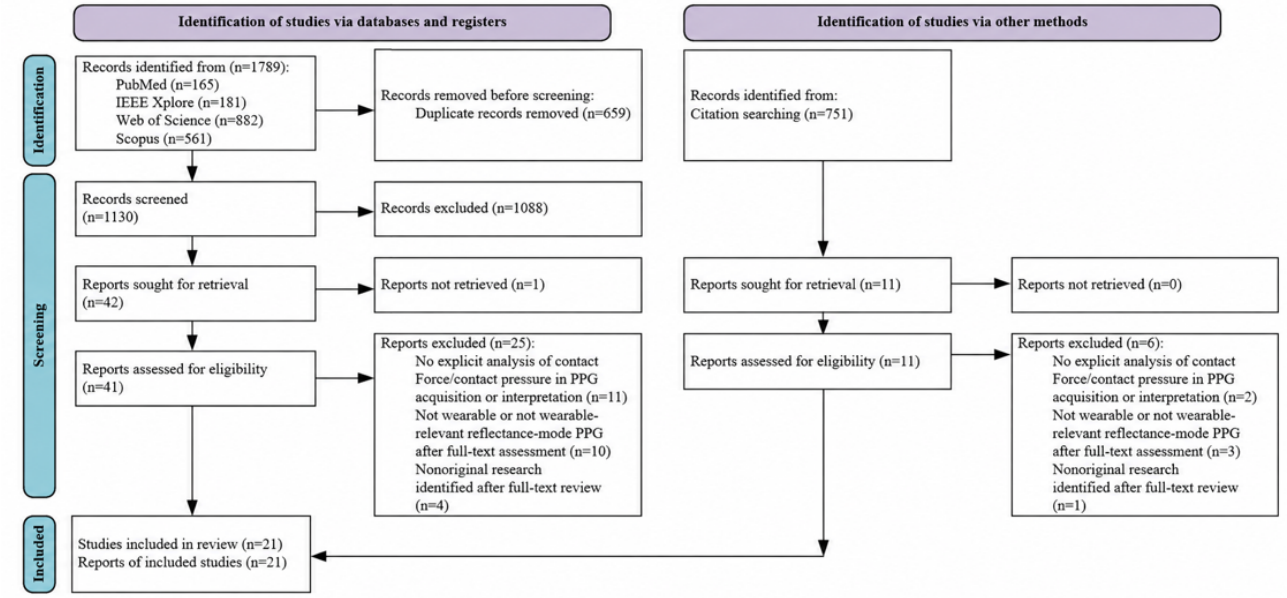
Results

Selection of Sources of Evidence

Figure 1 presents the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram illustrating the identification, screening, eligibility assessment, and final inclusion of studies. Database searches identified 1789 records, of which 1130 remained after deduplication and were screened; 16 reports were included through the database pathway. In addition, 751 reference-list entries were screened through citation searching, resulting in 5 additional included reports. Overall, 21 studies were included in the review.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. PPG: photoplethysmography.



Characteristics of Sources of Evidence

Characteristics of Publication and Baseline

Multimedia Appendix 4 summarizes the baseline characteristics of the 21 included studies, including publication year, publication type, country or region, and primary study focus [15,16,18,20,21,41-56]. The included studies were published between 2006 and 2026. Of these 21 studies, 15 (71%) [15,16,20,21,42-52] were published from 2020 onward, showing that most of the included evidence was published in recent years; besides, 16 (76%) [15,16,21,41-44,46-51,53-55] were journal articles and 5 (23.8%) [18,20,45,52,56] were peer-reviewed conference papers. In terms of country or region, the included studies were mainly from the United States (6/21, 29%) [18,42,47,48,52,56], China, including Hong Kong (3/21, 14%) [15,43,54], the United Kingdom (2/21, 10%) [16,46], Finland (2/21, 10%) [20,51], and Italy (2/21, 10%) [21,49], with 1 study each from South Korea [41], Slovakia [44], Canada [45], Spain [50], Latvia [53], and Israel [55]. Country or region was summarized according to the reported study location when available; when the study location was not explicitly stated, it was assigned according to the first or corresponding author's institutional affiliation.

Characteristics of Study Designs and Research Focus

The study-level characteristics of all included sources are summarized in Multimedia Appendix 4, with additional methodological and outcome details provided in Tables 1-3 and Multimedia Appendix 5. The included studies contributed to 3 broad and nonmutually exclusive research domains: contact-condition measurement, control, sensing, or estimation; directly measured signal-, waveform-, or feature-level responses; and directly evaluated downstream wearable monitoring tasks. Of the 21 included studies, 4 (19%) [41,43-45] had a primary methodological focus, 16 (76%) [15,16,18,20,41-45,47,49-51,53,55,56] reported at least 1 directly measured signal-, waveform-, or feature-level response, and 10 (48%) [15,18,21,42,47,49,51,52,54,56] directly evaluated at least 1 downstream task under different contact conditions. As these domains were nonmutually exclusive, individual studies could contribute to more than 1 domain.

XSL•FO
RenderX

Table 1. Summary of contact-force control and sensing methods.

Study	Study population and sample size	Measurement site/experimental condition	Method type	How force is controlled	How force is sensed	Force/pressure reporting and convertibility	Key hardware/implementation	Main information provided
Sim et al [41]	• Healthy human-participant validation • Main posture-change experiment: 1 healthy participant • Additional contact-force–dependency experiment: 4 participants. • Male/female distribution: not reported	Radial artery at the wrist	Active closed-loop control	Probe displacement is actively adjusted with a thermo-pneumatic actuator to maintain constant contact force in real time.	A FlexiForce A201 flexible force sensor measures force through force-dependent conductance changes.	• Contact force • Unit: Newton; target force: 0.6 N; measurable range: 0.2-4.0 N • Effective contact area: not reported • Contact-force–related signal • Unit/range: not reported	• Wrist platform comprising a force regulator, main body, force sensor, and reflectance PPG[a] probe.	Demonstrates a closed-loop contact-force regulation strategy for improving wrist PPG stability under posture-related contact changes.
Liu et al [43]	• Sample size and participant health status: not reported	Wrist; elastic wristband positioned over the radial and ulnar artery regions.	Multipoint synchronous monitoring	More of a force-sensing/regulation platform than a strict actuator-based closed loop; emphasis is on synchronous spatial monitoring.	Four ceramic piezoelectric sensors track changes in mechanical stress/contact force.	• Geometry/contact area: wrist-contact PPG and piezoelectric sensors; sensor diameter reported as 12 mm, but the effective skin-contact area for pressure conversion was not reported. • Contact force; authors used the term contact pressure force. • Unit/range: Newton; 5 force levels	• Wearable platform integrating 9 PPG sensors and 4 piezoelectric sensors. • The PPG sensors form a 3 × 3 array embedded in an elastic wristband, enabling simultaneous acquisition of 9 PPG channels and 4 contact force channels.	Provides a multipoint view of how local contact-force variation and spatial sensor position affect wrist PPG waveform behavior.
Přibil et al [44]	• 6 healthy volunteers: 4 males and 2 females • Mean age 56 (SD 8) years • All healthy volunteers	Left and right index fingers; five force levels	Direct contact force	No active control is emphasized; the main goal is direct quantification of probe-skin force.	A force-sensitive resistor directly measures probe-skin contact force.	• Male: 0.147-1.078 N; female: 0.147-0.753 N • Geometry/area: The force-sensitive resistor active sensing region had a diameter of 4 mm, but the effective skin-contact area was not reported; pressure in mm Hg (millimeters of mercury)/kPa (kilopascals) was not reported.	• PPG probe integrated with humidity, temperature, and force-sensitive resistor sensing.	Provides direct sensor-level documentation of finger PPG changes under different applied contact-pressure-force levels.

XSL•FO
RenderX

Study	Study population and sample size	Measurement site/experimental condition	Method type	How force is controlled	How force is sensed	Force/pressure reporting and convertibility	Key hardware/implementation	Main information provided
Fortin et al [45]	• 1 participant: male (28 years old), right-handed, pale White skin tone; self-administered protocol because of COVID-19 self-isolation restrictions.	Wrist and index fingertip	Indirect force estimation	No active force-control device is added; force is changed experimentally using a clamp and adjustment screw.	A miniature load cell provides the ground-truth force during development; the long-term aim is to estimate force from raw PPG alone.	• Contact force • Unit/range: Newton; 0-2.6 N in 0.2 N increments • Effective contact area: not reported	• 3D-printed clamp, 500 g load cell, HX711 module, and Arduino. • Force sampled from 0 to 2.6 N in 0.2 N steps. • 55 features extracted from raw PPG and modeled with a bagged-tree approach.	Shows the feasibility of estimating contact force from raw PPG features without adding a dedicated force sensor to the final wearable system.

[a]PPG: photoplethysmography.

Table 2. Measurement sites and wearable configurations in representative studies.

Study	Study design	Sample size and participant characteristics	Main measurement site	Wearable configuration/device form	Reference/auxiliary channel	Main findings
Charlton et al [46]	Measurements in different postures and sensor heights to examine effects of posture and sensor position relative to the heart on wrist PPG[a] signal quality.	1142 participants from the Aurora-BP dataset with varying ages and health statuses.	Wrist	Wrist-worn reflectance PPG	Posture and sensor height considered	Wrist PPG signal quality was strongly shaped by body posture and sensor height relative to the heart, indicating that wrist-site performance depends on acquisition geometry rather than anatomical site alone.
Ho et al [47]	A wrist-finger dual-channel dataset study for contact-pressure effects on wrist PPG morphology.	27 healthy young participants (16 males and 11 females)	Wrist	Force-adjustable wrist module	High-perfusion fingertip reference	Wrist posture changed sensor-skin contact pressure and produced different wrist PPG morphologies; the wrist setup supports analysis of wrist-site distortion in smartwatch-like monitoring
Dresher and Mendelson [18]	A forehead reflectance pulse-oximetry walking study under different contact pressures; involved an oxygen-saturation monitoring context	10 healthy volunteers; male/female distribution: not reported	Forehead above the eye	Elastic-headband–based forehead reflectance pulse oximeter with custom housing	Fingertip reference	The forehead above the eye was evaluated as a motion-tolerant reflectance pulse-oximetry site, but its performance depended on headband fixation and appropriate sensor-skin pressure during walking.
Chan et al [48]	A wearable reflectance pulse-oximetry study at the sternum; involved a continuous SpO_2[b] monitoring context	Total sample size, sex distribution, and participant characteristics: not reported	Mid-sternum	Chest/central reflectance PPG	Continuous SpO_2 task	The mid-sternum was assessed as a central wearable reflectance PPG site using a chest patch; this configuration enabled continuous SpO_2 monitoring but required attention to weak local perfusion and respiratory artifacts.

[a]PPG: photoplethysmography.

[b]SpO_2: peripheral oxygen saturation.

XSL•FO
RenderX

Table 3. Static and dynamic study protocols in representative contact-force research.

Study	Study design/protocol characteristics	Sample size and participant characteristics/platform	Measurement site	Protocol design	Force/pressure reporting	Main observation and finding
Scardulla et al [49]	A dynamic in vivo step-exercise wrist PPG[a] study evaluating how preset wristband contact pressure affects PPG-derived heart-rate accuracy under different stepping intensities.	17 participants (12 males and 5 females), with a mean age of 36 (SD 11) years; health status was not explicitly reported.	Wrist reflectance PPG with electrocardiogram chest-strap reference.	• Participants stepped on a 22.5 cm platform at 90, 120, and 140 bpm for 60 seconds, preceded by 10 seconds of rest. • Three contact pressures were randomized across 3 activity intensities, yielding 9 tests per participant.	• Contact pressure • Unit/range: mm Hg (millimeters of mercury); 12, 33, and 54 mm Hg. • Geometry/contact area: contact area = 473 mm2	Dynamic step-exercise protocol: PPG-derived heart rate was compared with electrocardiogram-derived heart rate under 3 contact pressures and 3 stepping intensities. Contact pressure affected heart rate agreement during movement, showing that pressure optimization remains important in dynamic wrist PPG recordings.
Sirkiä et al [20]	A quasi-static in vivo finger multiwavelength PPG pressure-ramp study examining how controlled external finger pressure affects wavelength-dependent PPG waveforms, vascular occlusion/opening behavior, and pulse-foot timing.	3 volunteers (1 female); health status and age were not reported	Index finger of the right hand; 5-wavelength multiwavelength PPG with controlled external pressure.	• External pressure was continuously increased by lowering a bar with a stepper motor until the maximum allowed pressure was above systolic blood pressure and then released. • Multiwavelength PPG and pressure signals were recorded synchronously, with 3 measurements per volunteer and reference blood pressure after each measurement.	• External pressure/contact pressure. • Unit/range: mm Hg; pressure was increased above systolic blood pressure and then released. • Geometry/contact area: multiwavelength PPG sensor printed circuit board diameter of approximately 9.6 mm; effective skin-contact area not reported.	Quasi-static finger pressure-ramp protocol: multiwavelength PPG was observed during external pressure increase and release. Pressure produced wavelength-dependent blockage/opening behavior and pulse-timing changes, indicating that external pressure alters vascular responses differently across optical wavelengths.
Scardulla et al [21]	A standardized dynamic in vivo treadmill wrist PPG study evaluating heart-rate accuracy, precision, and uncertainty under different strap pressures and exercise speeds.	25 healthy volunteers; male/female distribution was not reported.	Wrist PPG prototype with electrocardiogram chest-strap reference.	• For each pressure condition, blood pressure was measured first. • Participants then performed treadmill trials at 3, 6, and 8 km/h for 90 seconds each with 60-second rest intervals, followed by a 5-minute seated PPG recording and another blood pressure measurement. • The full sequence was repeated at 20, 60, and 75 mm Hg.	• Contact pressure. • Unit/range: mm Hg; 20, 60, and 75 mm Hg. • Effective contact area: not reported.	Dynamic treadmill protocol: heart rate accuracy, precision, and uncertainty were evaluated under 3 strap pressures and 3 treadmill speeds. Intermediate pressure, especially around 60 mm Hg, generally provided better PPG-heart rate performance than lower or higher pressures.

XSL•FO
RenderX

Study	Study design/protocol characteristics	Sample size and participant characteristics/platform	Measurement site	Protocol design	Force/pressure reporting	Main observation and finding
May et al [16]	A static in vitro tissue-vessel phantom study examining how progressively increased sensor contact pressure affects reflectance PPG signal quality and morphological features under simulated blood pressure states.	In vitro tissue-vessel phantom; no human participants; sex distribution not applicable.	Above the embedded vessel in the phantom.	• Under pulsatile flow fixed at 60 bpm, 4 hemodynamic states were simulated: hypotension, normotension, stage 1 hypertension, and stage 2 hypertension. • A linear actuator progressively increased sensor pressure until complete vessel occlusion.	• Contact pressure. • Unit/range: mm Hg; continuously increased until complete vessel occlusion; optimal range was reported as 35.1-48.1 mm Hg. • Effective contact area: not reported.	Static phantom protocol: signal-to-noise ratio and 17 morphological features were observed as contact pressure increased until vessel occlusion. PPG signal quality and morphology changed nonlinearly with pressure, with an optimal pressure range before complete occlusion.
Ho et al [47]	A static/quasi-static seated wrist-finger dual-channel PPG dataset study examining how wrist contact-pressure variation affects wrist PPG morphology, with fingertip PPG recorded simultaneously as a reference.	27 healthy young participants (16 males and 11 females); mean age 24.3 (SD 2.74) years	Wrist PPG with fingertip PPG reference	• Each participant completed 6 sessions. • Each session included blood pressure measurement, contact-pressure adjustment, and 4 minutes of seated recording, with a 2-minute rest between sessions. • Wrist PPG, fingertip PPG, load cell, electrocardiogram, blood pressure, and SpO_2[2] were synchronized.	• Contact pressure. • Unit/range: load-cell reading; gradually increased across 6 sessions. • Geometry/contact area: load cell placed between the wrist PPG sensor and movable jaw; effective skin-contact area was not reported.	Static/quasi-static seated wrist-finger protocol: wrist PPG morphology was observed across 6 contact-pressure sessions with fingertip PPG as a reference. Wrist contact-pressure variation systematically changed wrist PPG morphology, supporting dual-site recordings to characterize pressure-related wrist-site distortion.

[a]PPG: photoplethysmography.

[b]SpO_2: peripheral oxygen saturation.

Characteristics of Participants and Data Sources

The included sources of evidence varied substantially in terms of participant characteristics and data sources. Most human-participant studies involved healthy volunteers [18,20,21,41,42,44,45,47,49-53,56], whereas some studies used larger datasets or study populations with broader participant characteristics [46,52]. Sample sizes ranged from single-case feasibility studies [41,45] to large-scale wrist PPG datasets [46]. Several studies did not fully report participant age, sex distribution, or health status [18,43,48,50,55], limiting characterization of population diversity across the evidence base. In addition to the in vivo human studies, the evidence base included an in vitro tissue-vessel phantom study [16] and a study combining theoretical modeling with human-participant experimental validation [54].

Characteristics of Measurement Sites and Wearable Reflectance PPG Configurations

Measurement-site categories were nonmutually exclusive because some studies included more than 1 PPG acquisition site or used a secondary PPG site as a reference channel. Of the 21 included studies, 10 (48%) [15,21,41,43,45-47,49,52,55] involved wrist-based configurations, 9 (43%) [20,42,44,45,47,50,51,54,55] involved finger-based measurements, and 3 (14%) [18,55,56] involved forehead-based configurations. One study (5%) [48] examined a sternum-based configuration, 1 (5%) [53] examined other anatomical sites, and 1 (5%) [16] used an in vitro tissue-vessel phantom rather than a human anatomical measurement site. As several studies involved more than 1 measurement site, the summed percentages exceeded 100%. Primary device or experimental configurations were coded into mutually exclusive categories according to the principal PPG setup used in each study. Of the 21 included studies, 7 (33%) [15,21,41,43,46,49,52] used single-site

XSL•FO
RenderX

wrist-based systems, 6 (29%) [20,42,44,50,51,54] used single-site finger-based systems, and 3 (14%) [45,47,55] used multisite configurations involving more than 1 PPG acquisition site; 2 (10%) [18,56] studies used forehead-based reflectance pulse-oximetry systems, 1 (5%) [48] used a sternum-based reflectance PPG system, 1 (5%) [53] used a probe at other anatomical sites, and 1 (5%) [16] used an in vitro tissue-vessel phantom rather than an in vivo wearable configuration. Several secondary design features were identified across these primary categories. Three studies (14%) [15,41,47] used force-adjustable or force-regulated platforms, 3 (14%) [43,47,55] used multichannel or multisite PPG acquisition, and 3 (14%) [20,51,55] used multiwavelength reflectance PPG configurations. These secondary categories were nonmutually exclusive because individual systems could incorporate more than 1 design feature. Detailed study-level configurations, fixation strategies, optical settings, and reference measurements are presented in Table 2.

Characteristics of Contact-Force and Contact-Pressure Reporting

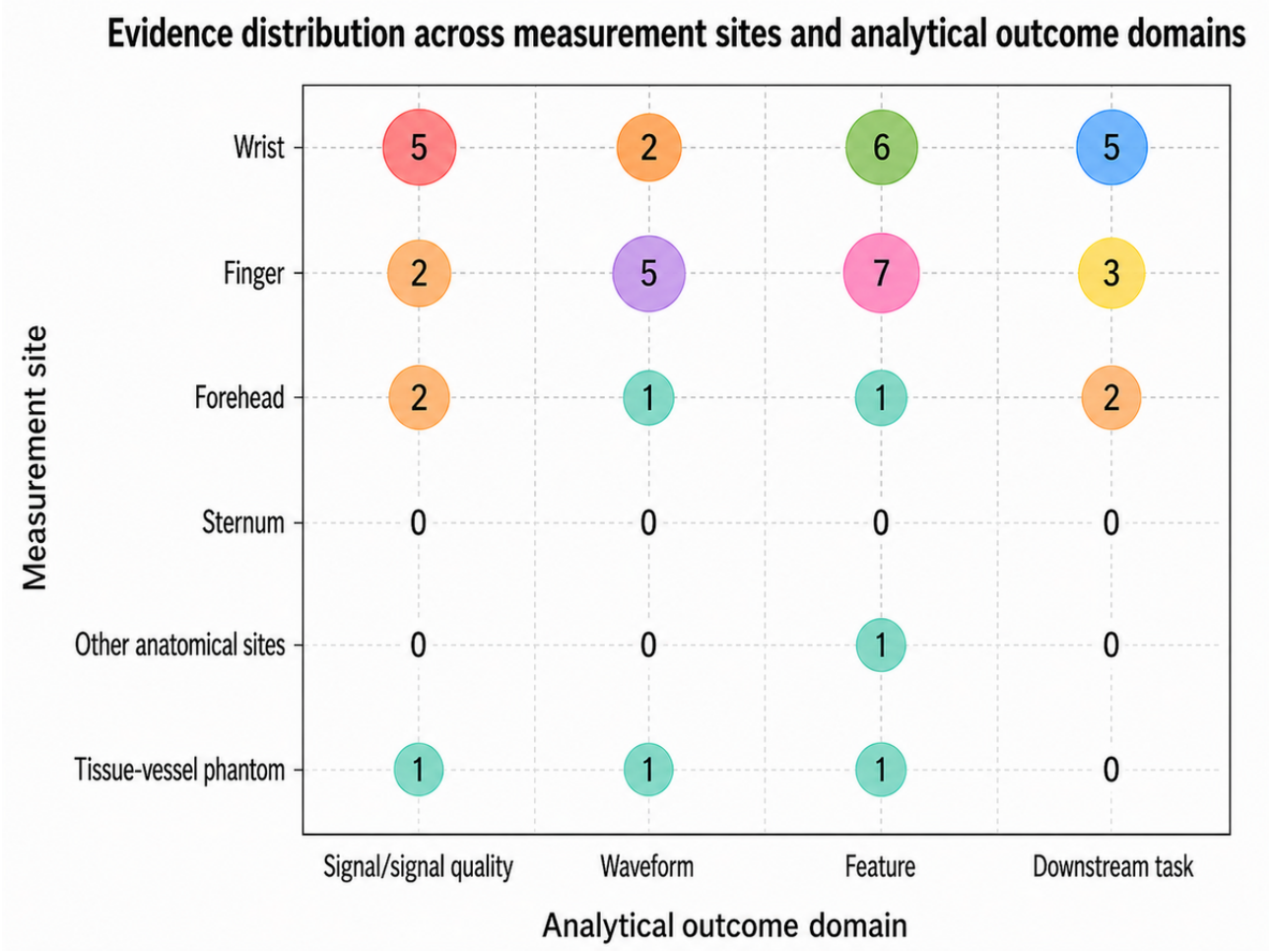
The representation of contact conditions differed substantially across the included studies. Based on the principal mechanical quantity or descriptor reported in each study, the 21 studies were classified into 4 mutually exclusive categories. As many as 5 (24%) studies [15,41,44,45,54] used force-based representations. Of these, all 5 [15,41,44,45,54] explicitly reported contact force in Newtons. Among them, Teng and Zhang [54] examined discrete applied contact-force levels through theoretical modeling and validation involving human participants. Ten studies (48%) [16,18,20,21,42,49,51,53,55,56] used pressure-based representations reported in physical pressure units, including mm Hg (millimeters of mercury), kPa (kilopascal), or torr. These representations included externally applied or cuff-derived pressure values [20,42,51], preset wristband or strap-pressure levels [21,49], and locally applied probe-skin, sensor-skin, or device-housing pressure [16,18,53,55,56]. Four studies (19%) [43,47,50,52] used nonstandardized sensor-output, normalized, indirect, or categorical representations. These included contact-force–related piezoelectric sensor output without a reported physical unit [43], sequential load-cell readings without conversion to a standardized force or pressure value [47], normalized force divided into predefined ranges [50], and pressure-related waveform-morphology labels [52]. Two studies (10%) [46,48] described sensor-skin contact qualitatively or contextually without directly quantifying the applied force or pressure. One study [49] explicitly reported the sensor-skin contact area. Four studies [20,43,44,47] provided partial sensor, probe, printed circuit board, or device-geometry information but did not report the effective skin-contact area. For the remaining studies, contact-area or probe-geometry information was insufficient or unavailable. Sensor or device dimensions were not assumed to represent effective skin-contact area unless this relationship was explicitly established. Consequently, reported force values could not always be converted into contact pressure or directly compared with pressure values across devices. Although contact force and contact pressure are mechanically related, they were not treated as interchangeable when sufficient contact-area information was unavailable.

Characteristics of Multilevel PPG Responses Under Different Contact Force/Pressure

The reported outcomes spanned signal or signal-quality, waveform, feature, and downstream-task domains. Sixteen studies [15,16,18,20,41-45,47,49-51,53,55,56] contributed direct signal-, waveform-, or feature-level evidence, whereas 10 studies [15,18,21,42,47,49,51,52,54,56] evaluated at least 1 downstream task under different contact conditions. These categories were nonmutually exclusive. Detailed study-level findings are presented below and synthesized across measurement sites, protocol contexts, and analytical domains in Figures 2-4 and Table 4. The studies considered in the empirical linkages in Figure 3 are [15,18,42,47,49,51,52,56].

XSL•FO
RenderX

Figure 2. Evidence distribution across measurement sites and analytical outcome domains. Bubble size represents the number of unique studies contributing direct empirical response evidence to each site-outcome combination. Studies could contribute to more than 1 cell. Empty cells indicate that no direct empirical evidence of a response to the contact condition was mapped to that combination.



Results of Individual Sources of Evidence

Overview

To avoid duplicating the detailed study-level tables, the following narrative highlights key findings from individual sources.

Contact-Force Control and Sensing Methods

As outlined in Table 1, the included studies can be broadly categorized into 4 methodological groups according to their approaches to contact force and contact pressure in wearable reflectance PPG: active closed-loop force control, direct contact force/pressure measurement, multipoint contact-force monitoring, and indirect contact-force estimation based on PPG signals. These approaches reflect 4 distinct methodological priorities, namely, stabilizing contact conditions, quantifying contact state, characterizing spatial variations in contact, and reducing reliance on additional hardware. Sim et al [41] represented the first direction by developing an active closed-loop force-control platform, through which the analysis could be performed under constant contact conditions, thereby emphasizing signal stability and experimental repeatability. By contrast, Liu et al [43] extended the problem from single-point force regulation to multipoint synchronous sensing, enabling the investigation of relationships between spatial variations in contact state and PPG waveform behavior. Přibil et al [44] further highlighted the direct measurement and quantitative characterization of probe-skin contact force by integrating force-sensitive sensing into the optical probe, which improved the physical interpretability of experimental conditions and facilitated comparison across studies. Fortin et al [45], in turn, explored a different direction by attempting to estimate contact force indirectly from raw PPG signals, thereby reducing dependence on dedicated force-sensing hardware.

Measurement Sites and Wearable Configurations

As shown in Table 2, from the perspective of measurement site and device configuration, current reflectance PPG research has moved beyond the simple question of where a sensor should be placed. Instead, it increasingly treats measurement site, wearing form, reference channel, and target task as a coupled system-level design problem. Charlton et al [46] reported that wrist PPG signal quality varied across body postures and sensor heights relative to the heart. Ho et al [47] further showed that this systemic issue also involves contact-force control and reference-channel configuration; by combining a force-adjustable wrist module with a high-perfusion fingertip reference, they more clearly revealed the coupling among measurement site, wearing structure, and reference design. The study by Dresher and Mendelson [18] likewise indicated that the effectiveness of the forehead as a measurement site cannot

XSL•FO
RenderX

be understood independently of the specific wearing configuration. Using a forehead reflectance pulse oximetry device consisting of an elastic headband and a custom housing, they systematically evaluated the performance of PPG signals measured from the supraorbital forehead region under different contact pressures and showed that forehead signal quality and measurement reliability during motion are shaped not only by the measurement site itself but also by the combined effects of fixation structure and contact-pressure conditions. Chan et al [48] extended the measurement site from the periphery to the mid-sternum, further indicating that although more central reflectance PPG holds promise for continuous oxygen saturation monitoring, its performance remains constrained by multiple factors, including local perfusion, optical path design, and wearing strategy.

Static and Dynamic Study Protocols

As summarized in Table 3, existing protocols can be broadly categorized into static/quasi-static designs and dynamic in vivo designs. Static and quasi-static protocols primarily examined contact-condition–related changes in waveform morphology, signal-to-noise ratio (SNR), and derived features while minimizing motion-related variation. Dynamic protocols examined whether similar response patterns were observed during stepping, walking, running, or other wearable-use conditions. In the dynamic category, Scardulla et al [49] investigated wrist PPG during step exercise at 3 preset contact pressures and showed that contact pressure had a greater impact on heart-rate measurement quality than exercise intensity itself. Scardulla et al [21] later extended this line of work using a more standardized treadmill protocol across 20, 60, and 75 mm Hg and found that 60 mm Hg offered the best overall balance of accuracy and precision, although the optimal pressure still varied across individuals. By contrast, quasi-static and static studies have been more focused on mechanism. Sirkiä et al [20], using a controlled pressure ramp on the fingertip with multiwavelength PPG, showed that externally applied pressure changes not only oscillation amplitudes but also pressure-dependent occlusion behavior and interchannel timing differences, indicating depth-dependent vascular responses. May et al [16], in an in vitro tissue-vessel phantom study with continuously increasing pressure under multiple simulated blood pressure states, demonstrated the existence of an optimal contact-pressure range and further showed that temporal features are relatively robust, whereas amplitude- and geometry-related features are substantially more pressure-sensitive. Ho et al [47] further contributed a quasi-static wrist-finger dual-channel dataset collected under sedentary conditions, in which wrist pressure was gradually adjusted while a fingertip channel was maintained as a high-quality reference; their results showed systematic pressure-dependent transitions in wrist PPG morphology and corresponding degradation in downstream HR and HRV estimation under suboptimal pressure.

Contact-Condition–Related Findings in Reflectance PPG

Signal and Signal-Quality Responses

As outlined in Table S1 in Multimedia Appendix 5, the included signal-quality studies reported contact-condition–related differences across several PPG dimensions, including signal amplitude, perfusion-related measures, noise, baseline stability, motion artifacts, and morphology-based quality indices. Lambert Cause et al [50] evaluated normalized fingertip force across 5 bins using skewness, kurtosis, relative power, entropy, zero-crossing rate, and perfusion index and found that different signal-quality indices favored different force ranges. May et al [16] demonstrated in vitro that the SNR first increased and then decreased as contact pressure rose, with a clear optimum before full occlusion. He et al [15] quantified the force dependence of high-frequency noise, baseline drift, and motion artifacts, showing that higher force could reduce some motion-related degradation, although not uniformly across all indices. Charlton et al [46] further emphasized that contact pressure should be interpreted jointly with posture, sensor height, and skin contact state when signal quality is assessed at the wrist.

Waveform and Feature Responses

As outlined in Table S2 in Multimedia Appendix 5, the included studies reported contact-condition–related differences not only in signal amplitude but also in waveform morphology and derived feature behavior. May et al [16] showed that PPG amplitude and SNR followed an inverted-U pattern as pressure increased. Ho et al [47] revealed that low-force wrist PPG often appeared single-peaked or noise-dominated, whereas moderate force revealed a more canonical 2-peak morphology with a clearer systolic peak, dicrotic notch, and diastolic peak; at still higher force, the waveform could collapse back toward a distorted single-peak shape. Sirkiä et al [20,51] further demonstrated that contact pressure affects both AC and DC components and that wavelength channels respond differently. Grabovskis et al [53] investigated the effect of probe contact pressure on PPG-based conduit artery stiffness assessment and showed that variable probe pressure affected the AC PPG second-derivative peak ratio b/a, a morphology-derived parameter associated with arterial stiffness. Fine and Kaminsky [55] showed that, in reflection pulse oximetry, gamma, a dual-wavelength PPG-derived ratio parameter used in arterial oxygen saturation calculation, increased with increasing applied pressure.

Findings on Artifact Susceptibility and Temporal Stability

As outlined in Table S3 in Multimedia Appendix 5, several included studies examined contact conditions in relation to artifact susceptibility and the temporal stability of PPG-derived measurements. At the wavelength level, Sirkiä et al [20,51] found that shorter wavelengths were more pressure-sensitive than red and infrared channels, with earlier onset of occlusion-like behavior and waveform distortion as pressure increased. At the feature level, Sim et al [41] showed that amplitude-dominated indices were more force-sensitive than interval-dominated indices. At the task level, Scardulla et al [49] found that excessively low pressure promoted poor contact and motion artifacts during exercise, whereas excessively high pressure could reduce precision by compressing local microcirculation.

Findings for SpO2 and Other Downstream Estimation Tasks

As outlined in Table S4 in Multimedia Appendix 5, a smaller body of evidence examined contact force and contact pressure in relation to downstream physiological estimation tasks beyond

XSL•FO
RenderX

signal- and waveform-level outcomes. For SpO_2 estimation, He et al [15], in a study of wrist reflectance PPG-based oxygen saturation prediction, showed that explicitly incorporating force-related features into an SpO_2 prediction model improved estimation accuracy, reducing the mean absolute error to 0.8811%. For BP estimation, Chandrasekhar et al [42], in a study using PPG-derived PAT for potential cuffless BP measurement, found that contact-pressure–induced changes in waveform morphology and fiducial-point location shifted PAT_{foot} and PAT_{peak} by 22 (SD 2) ms and 40 (SD 7) ms, respectively. Here, PAT_{foot} refers to PAT estimated from the foot of the PPG waveform, whereas PAT_{peak} refers to PAT estimated from the systolic peak. These timing shifts corresponded to approximately 11 and 20 mm Hg of BP estimation error, respectively, indicating that pressure-induced waveform distortion can directly propagate into BP estimation. For heart-rate estimation, Dresher and Mendelson [56] designed a reflectance pulse oximeter housing with contact-pressure measurement and showed that excessive sensor-skin pressure can impair local blood flow, attenuate or even eliminate PPG signals, and thereby affect pulse-oximetry measurement accuracy. Their dual-housing design improved HR accuracy under large contact-pressure variations. Their findings linked contact-pressure variation with differences in pulse-oximetry and heart-rate performance. Scardulla et al [21,49], in studies evaluating PPG signal performance under different contact-pressure conditions, showed that heart-rate estimation is strongly pressure-dependent, with intermediate pressures generally providing the best group-level performance, whereas both loose and excessively tight configurations degrade accuracy through different mechanisms. Hu et al [52], in wrist-PPG–based HRV estimation under sedentary conditions, demonstrated that even subtle waveform morphology distortions can degrade HRV estimation performance. They showed that morphology-aware modeling reduced the gap between wrist PPG and reference measurements, indicating that pressure-related morphology changes may impair HRV estimation by disturbing beat-to-beat timing and waveform-feature extraction.

Nonlinear Force Dependence and Moderate-Force Behavior

Several included studies reported nonlinear or task-dependent response patterns under their respective experimental conditions. As outlined in Table S5 in Multimedia Appendix 5, for example, May et al [16] found that both waveform amplitude and SNR first increased and then decreased as pressure rose. Lambert Cause et al [50] reported that the perfusion index, which is strongly amplitude-dependent, was also highest in lower-to-moderate force ranges before declining at higher force. Their signal-quality indices likewise showed that both low and high extremes were suboptimal, whereas moderate force better preserved waveform detail. Sirkiä et al [20,51] further showed that SpO_2 estimates remained relatively stable at low pressure but began to rise as pressure approached diastolic and mean arterial pressure, indicating that pressure-induced changes in AC/DC relationships and feature points can bias downstream metrics. Teng and Zhang [54] investigated the effect of sensor contact force on PPG-derived PTT using theoretical modeling and validation involving human participants. By considering the nonlinear pressure-volume relationship of the finger arterial wall, they showed that PTT increased with contact force under positive transmural pressure, reached a maximum near zero transmural pressure, and then remained nearly constant under negative transmural pressure. This finding indicates that sensor contact force can nonlinearly alter PPG-derived timing parameters through changes in local transmural pressure and arterial wall mechanics. Across these studies, nonlinear or inverted-U-like response patterns were reported under the respective experimental conditions. Intermediate contact levels were frequently associated with better signal quality, waveform clarity, or task performance than the lowest or highest tested levels. Importantly, the location and shape of this optimum are not fixed but vary with the physiological target, wavelength, measurement site, and feature type.

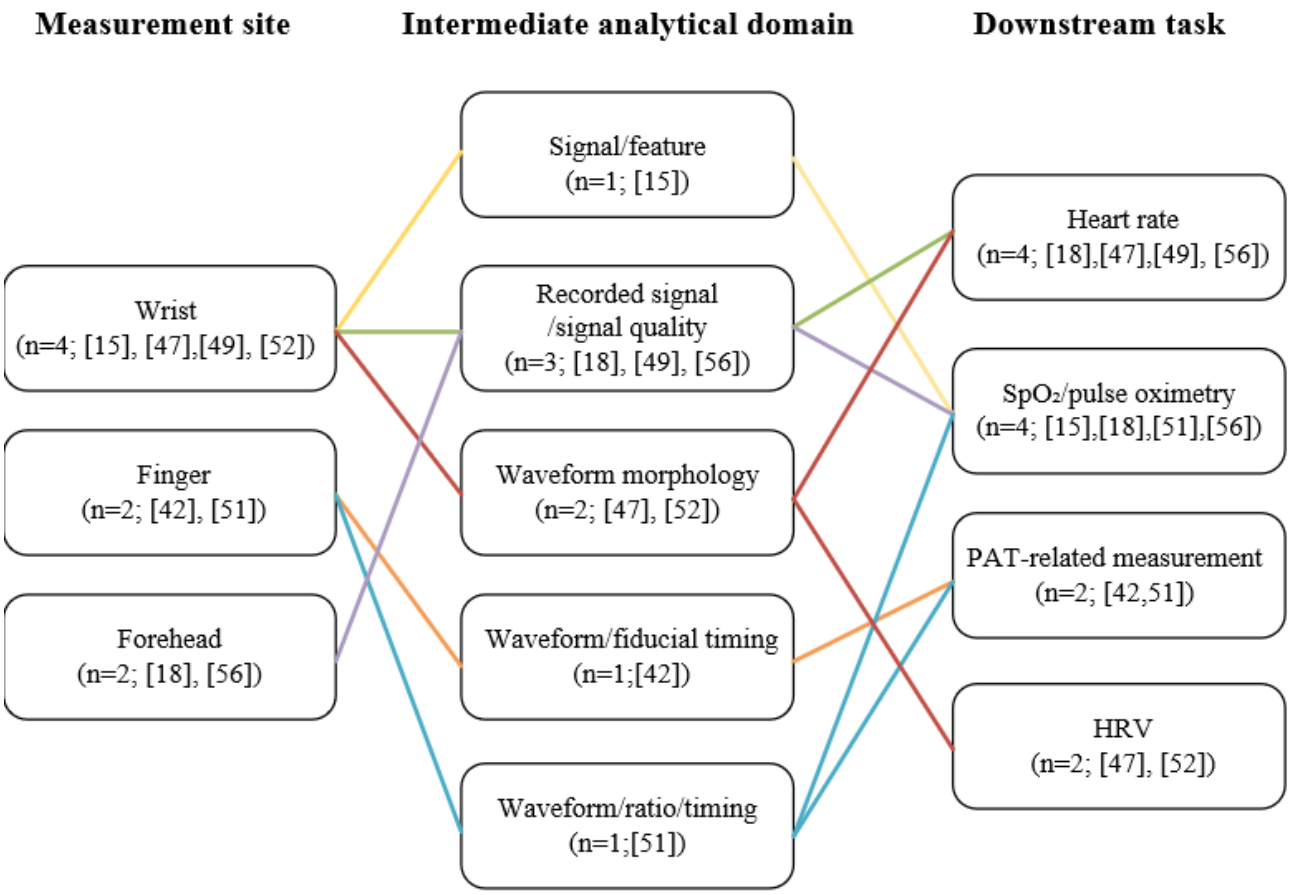
Synthesis of Results

Evidence Distribution Across Measurement Sites and Analytical Outcome Domains

Figure 2 maps direct empirical response evidence across measurement sites and analytical outcome domains. At the wrist, direct empirical evidence included 5 signal- or signal-quality-level studies [15,21,41,43,49], 2 waveform-level studies [43,47], 6 feature-level studies [15,43,45,47,52,55], and 5 downstream-task studies evaluating SpO_2, HR, or HRV [15,21,47,49,52]. At the finger, 2 studies reported signal-level responses [44,50], 5 examined AC/DC behavior, pulse shape, or waveform fiducial points [20,42,44,51,55], 7 evaluated timing-, morphology-, perfusion-, or ratio-based features [20,42,44,45,50,51,55], and 3 assessed SpO_2, PAT, or PTT as downstream tasks [42,51,54]. At the forehead, 2 studies reported signal- or signal-quality-level responses [18,56], 1 examined waveform-level responses [56], 1 evaluated feature-level responses [55], and 2 assessed HR or pulse-oximetry performance [18,56]. Evidence from other sites was limited: 1 study [53] at other anatomical sites reported morphology- and derivative-based feature responses, whereas the tissue-vessel phantom study [16] contributed signal-, waveform-, and feature-level evidence across progressively increased pressure conditions. No direct empirical contact-condition response was mapped for the sternum study [48]. As individual studies could contribute to more than 1 measurement-site or analytical-domain category, the categories were nonmutually exclusive. The figure summarizes the distribution of evidence but does not determine whether intermediate PPG responses and downstream tasks were connected within the same study; these complete within-study linkages are presented in Figure 3.

XSL•FO
RenderX

Figure 3. Complete within-study empirical linkages from measurement sites through intermediate photoplethysmography domains to downstream tasks. Each colored pathway represents a single study; branches indicate that a study contributed evidence linking the same pathway to multiple downstream tasks. HRV: heart rate variability; PAT: pulse arrival time; SpO_2: peripheral oxygen saturation.



Evidence Linkages Across Measurement Sites, Intermediate Analytical Domains, and Downstream Tasks

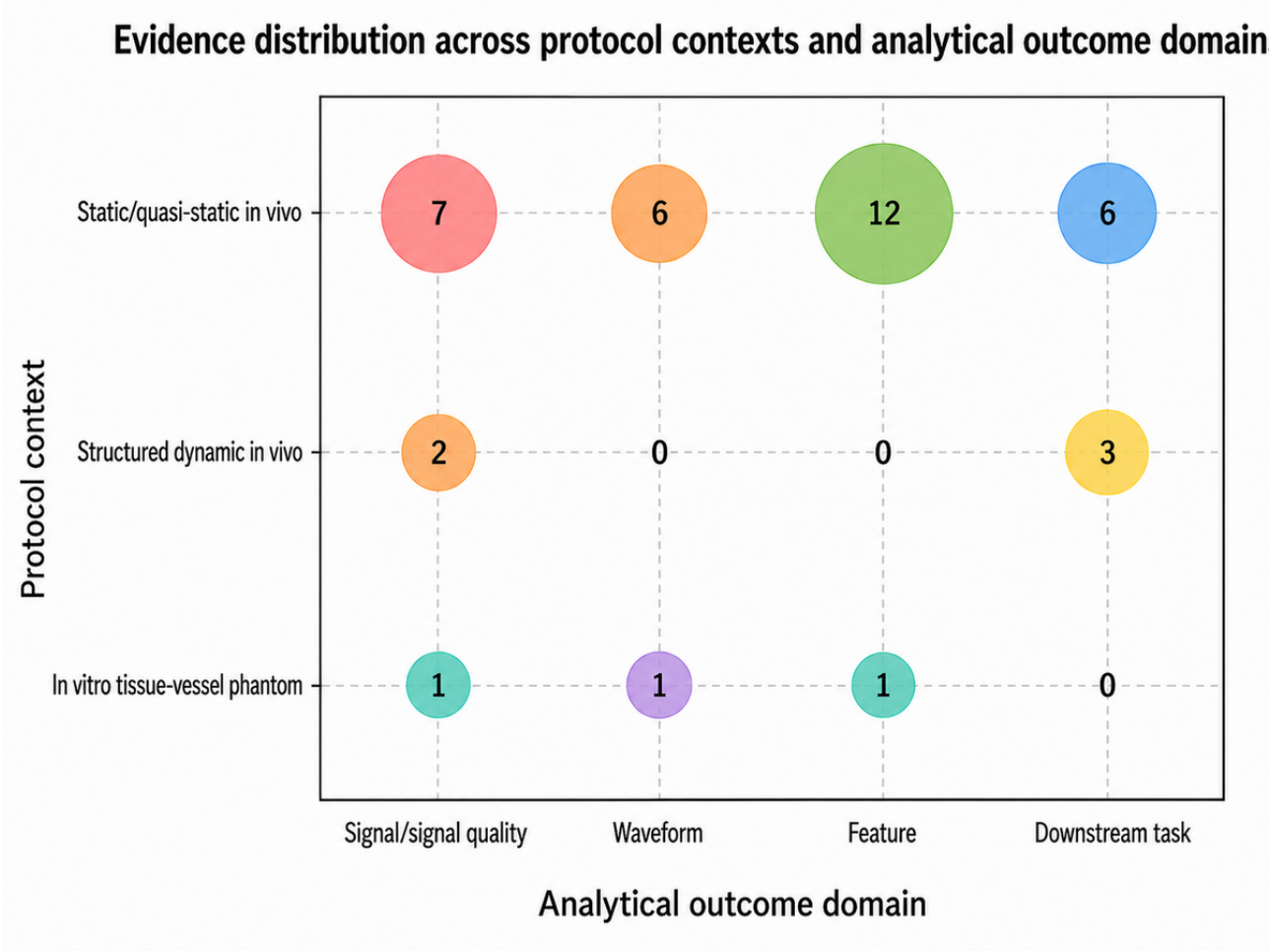
At the wrist, He et al [15] linked contact-force–responsive signal-quality indices, including high-frequency noise, baseline drift, and motion artifact, together with contact-force information, to PPG-based SpO_2 prediction. Ho et al [47] linked pressure-related changes in waveform morphology, including the systolic peak, dicrotic notch, and diastolic peak, to HR and HRV estimation. Scardulla et al [49] linked contact stability and motion-artifact susceptibility during exercise to PPG-derived HR evaluation. Hu et al [52] used raw wrist PPG waveforms and pressure-related morphology labels to estimate NN intervals and beat count for HRV computation. At the finger, Chandrasekhar et al [42] linked changes in waveform-foot and systolic-peak timing to PAT_{foot} and PAT_{peak}. Sirkiä et al [51] linked AC/DC components, wavelength-dependent waveform responses, amplitude- and perfusion-related characteristics, and pulse-foot and pulse-peak timing to SpO_2, PAT_{foot}, and PAT_{peak}. At the forehead, Dresher and Mendelson [18] linked contact-pressure–related PPG signal quality and measurement reliability during walking to HR and SpO_2 performance. In a related housing study, Dresher and Mendelson [56] linked pressure-related differences in forehead PPG amplitude and signal quality to SpO_2 and HR measurement performance. Overall, the complete within-study linkages involved signal-quality indices and contact-force information for SpO_2 prediction [15], waveform morphology for HR and HRV estimation [47,52], dynamic signal stability for HR evaluation [49], fiducial-point timing for PAT measurement [42], AC/DC and wavelength-dependent responses for SpO_2 and PAT calculation [51], and forehead PPG signal quality and amplitude for HR and pulse-oximetry performance [18,56].

Evidence Distribution Across Protocol Contexts and Outcome Domains

Under static or quasi-static in vivo conditions, direct empirical evidence included 7 studies [15,41,43-45,50,56] reporting signal- or signal-quality-level responses, 6 [20,42,44,47,51,55] examining waveform-level changes, 12 [15,20,42-45,47,50-53,55] evaluating feature-level responses, and 7 [15,42,47,51,52,54,56] assessing downstream tasks. Under structured dynamic conditions, 2 studies [18,49] reported signal- or signal-quality-level responses, whereas 3 [18,21,49] evaluated downstream tasks, primarily HR-related performance. In the in vitro tissue-vessel phantom protocol, 1 study [16] provided signal-, waveform-, and feature-level evidence across controlled pressure conditions. Overall, Figure 4 maps directly measured or calculated PPG responses across principal protocol contexts and analytical outcome domains. The mapped distribution indicates that waveform- and feature-level evidence was concentrated in static or quasi-static in vivo protocols, whereas structured dynamic protocols mainly addressed signal behavior and downstream HR performance.

XSL•FO
RenderX

Figure 4. Distribution of direct empirical evidence across principal protocol contexts and analytical outcome domains. Bubble area represents the number of unique studies contributing evidence to each protocol-outcome combination. Each study was assigned to 1 principal protocol category but could contribute to multiple outcome domains. Empty cells indicate that no direct empirical evidence was mapped to that combination.



Study-Level Coverage Across Methodological and Analytical Domains

Table 4 provides a study-level overview of the methodological and analytical domains covered by the 21 included studies. Measurement site and principal protocol context were identifiable for all studies. Participant characteristics and contact-condition representation were more consistently documented than effective contact area or probe geometry. Sensor, printed circuit board, load-cell, or device dimensions without sufficient information on the effective skin-contact interface were classified as partial rather than sufficient reporting. Analytical coverage varied across studies, with some studies addressing multiple signal-, waveform-, feature-, and downstream-task domains. Overall, the matrix highlights variation in methodological reporting and analytical breadth rather than differences in study quality.

XSL•FO
RenderX

Table 4. Overview of the methodological and analytical domains covered by the 21 included studies.

Study	Study context		Methodological reporting			Analytical coverage			
	Site	Protocol	Participants	Contact-condition representation	Geometry/contact area	Signal/signal quality	Waveform	Feature	Downstream task
He et al [15]	Wrist	Static/quasi-static in vivo	Sufficient	Sufficient	Insufficient or unavailable	Assessed	Not assessed	Assessed	Assessed
May et al [16]	Tissue-vessel phantom	In vitro tissue-vessel phantom	N/A[a]	Sufficient	Insufficient or unavailable	Assessed	Assessed	Assessed	Not assessed
Sim et al [41]	Wrist	Static/quasi-static in vivo	Sufficient	Sufficient	Insufficient or unavailable	Assessed	Assessed	Not assessed	Not assessed
Chandrasekhar et al [42]	Finger	Static/quasi-static in vivo	Sufficient	Sufficient	Insufficient or unavailable	Not assessed	Assessed	Assessed	Assessed
Dresher et al [18]	Forehead	Structured dynamic in vivo	Insufficient or unavailable	Sufficient	Insufficient or unavailable	Assessed	Not assessed	Not assessed	Assessed
Sirkiä et al [20]	Finger	Static/quasi-static in vivo	Sufficient	Sufficient	Partial or indirect	Not assessed	Assessed	Assessed	Not assessed
Scardulla et al [21]	Wrist	Structured dynamic in vivo	Sufficient	Sufficient	Insufficient or unavailable	Assessed	Not assessed	Not assessed	Assessed
Liu et al [43]	Wrist	Static/quasi-static in vivo	Insufficient or unavailable	Partial or indirect	Partial or indirect	Assessed	Assessed	Assessed	Not assessed
Přibil et al [44]	Finger	Static/quasi-static in vivo	Sufficient	Sufficient	Partial or indirect	Assessed	Not assessed	Assessed	Not assessed
Fortin et al [45]	Multiple sites	Static/quasi-static in vivo	Sufficient	Sufficient	Insufficient or unavailable	Assessed	Not assessed	Assessed	Not assessed
Charlton et al [46]	Wrist	Observational	Sufficient	Insufficient or unavailable	Insufficient or unavailable	Assessed	Not assessed	Not assessed	Not assessed
Ho et al [47]	Multiple sites	Static/quasi-static in vivo	Sufficient	Partial or indirect	Partial or indirect	Not assessed	Assessed	Assessed	Assessed
Chan et al [48]	Sternum	Observational	Insufficient or unavailable	Insufficient or unavailable	Insufficient or unavailable	Not assessed	Not assessed	Not assessed	Assessed
Scardulla et al [49]	Wrist	Structured dynamic in vivo	Sufficient	Sufficient	Sufficient	Assessed	Not assessed	Not assessed	Assessed
Lambert Cause et al [50]	Finger	Static/quasi-static in vivo	Insufficient or unavailable	Partial or indirect	Insufficient or unavailable	Assessed	Not assessed	Assessed	Not assessed
Sirkiä et al [51]	Finger	Static/quasi-static in vivo	Sufficient	Sufficient	Insufficient or unavailable	Assessed	Assessed	Assessed	Assessed
Hu et al [52]	Wrist	Static/quasi-static in vivo	Sufficient	Partial or indirect	Insufficient or unavailable	Not assessed	Assessed	Assessed	Assessed
Grabovskis et al [53]	Other anatomical site	Static/quasi-static in vivo	Sufficient	Sufficient	Insufficient or unavailable	Not assessed	Assessed	Assessed	Assessed
Teng and Zhang [54]	Finger	Static/quasi-static protocol with theoretical modeling and validation of human participants	Sufficient	Sufficient	Insufficient or unavailable	Not assessed	Not assessed	Not assessed	Assessed

XSL•FO
RenderX

Study	Study context		Methodological reporting			Analytical coverage			
	Site	Protocol	Partici-pants	Contact-condition representa-tion	Geometry/contact area	Signal/sig-nal quality	Waveform	Feature	Down-stream task
Fine and Kamin-sky [55]	Multiple sites	Static/quasi-static in vivo	Insufficient or unavail-able	Sufficient	Insufficient or un-available	Assessed	Assessed	As-sessed	Assessed
Dresher and Mendel-son [56]	Forehead	Static/quasi-static in vivo	Sufficient	Sufficient	Insufficient or un-available	Assessed	Assessed	Not as-sessed	Assessed

[a]N/A: not applicable.

Discussion

Principal Findings

This scoping review identified 21 studies that examined how contact force or contact pressure was controlled, measured, estimated, or characterized in wearable or wearable-relevant reflectance PPG, together with the reported signal-, waveform-, feature-, and task-level outcomes under different contact conditions [15,16,18,20,21,41-56]. Previous reviews have synthesized the physiological basis of PPG, waveform analysis, signal-processing methods, and downstream applications such as HR, BP, oxygen saturation, and arrhythmia detection [7-10,22-29]; wearable-device architectures and configurations [30-34]; and signal-quality challenges, datasets, and analytical tools [7,29,35-37]. Within these broader reviews, contact pressure, wearing tightness, and sensor-skin interface stability have generally been discussed as acquisition or signal-quality considerations rather than as the central framework for organizing the evidence [29-32]. Building on this literature, our review places contact force and contact pressure at the center of the synthesis, distinguishes these 2 mechanical quantities, and integrates evidence on their measurement, control, representation, and reporting with outcomes across multiple levels of PPG analysis. It also maps the distribution of evidence across anatomical sites, device configurations, populations, experimental settings, and protocol durations.

The mapped evidence indicates that contact force and contact pressure are important measurement conditions associated with variability in wearable reflectance PPG observations and should not be treated solely as minor wearing-related factors. Across the included studies, different contact conditions were reported in relation to differences in signal amplitude, signal quality, motion susceptibility, waveform morphology, feature stability, and downstream physiological estimation [15,16,18,20,21,41,42,46-56]. However, these responses were not uniform across feature categories or experimental settings. Amplitude-, morphology-, and derivative-based parameters generally varied more across contact conditions than some timing-related features, suggesting that feature stability may depend on the type of parameter examined [16,41,53]. The reviewed studies further indicated that contact-condition–related responses varied across measurement sites, wavelengths, fixation strategies, postures, and motion conditions, rather than following a single relationship that could be generalized across devices or settings [18,20,21,46-48,51]. Their practical relevance also appeared to differ by task: applications that rely on waveform morphology, amplitude relationships, or precise fiducial-point localization may require more task-specific evaluation than applications based primarily on preserved periodicity [15,20,21,42,49,51,52]. Compared with previous broad reviews of wearable PPG, this review therefore provides a focused methodological and conceptual map of the sensor-skin mechanical interface. It also identifies incomplete mechanical reporting, limited real-world validation, restricted population and measurement-site diversity, and incomplete linkage between intermediate signal characteristics and downstream estimation as important priorities for future research.

Limitations

Inconsistent Reporting of Contact Force and Contact Pressure

Despite growing interest, the mapped evidence remains methodologically heterogeneous [15,16,18,20,21,41-56]. A major limitation is the lack of a standardized framework for representing, measuring, and reporting contact force and contact pressure [15,16,18,20,21,41-56]. Importantly, contact force and contact pressure are mechanically related but not interchangeable: contact force describes the total load applied to the sensor-skin interface, whereas contact pressure depends on how that load is distributed over the effective contact area [15,16,20,21,41-47,50-56]. Across studies, contact conditions were described using different units or concepts, including Newtons, mm Hg, normalized force units, cuff pressure, preset strap levels, inferred tightness, or qualitative descriptions of skin contact [15,16,18,20,21,41-56]. This heterogeneity limits direct comparison and the development of quantitative design guidance [15,16,20,21,41-47,49-56]. It also limits the transferability of terms such as “moderate force” or “optimal pressure” across wearable configurations [16,20,21,47,49-51,54]. In particular, when studies report only force without specifying the effective contact area, probe geometry, or active sensing area, the corresponding contact pressure cannot be reliably derived [15,41,43-45,47,50,52]. Even when similar terms are used, the underlying physical meaning may differ because reported pressure can depend on probe geometry, active sensing area, tissue compliance, or the way force is converted into pressure [16,20,21,47,49-51,53,54].

XSL•FO
RenderX

A more standardized reporting convention would therefore be an important step toward improving comparability and reproducibility [57,58]. Future studies should report the effective sensor-skin contact area or probe geometry when reporting contact force, or directly report contact pressure in standardized units such as mm Hg or kPa [57,58]. This would improve reproducibility, cross-study comparability, and practical translation for wearable PPG device design.

Limited Real-World and Longitudinal Validation

A major evidence gap is the limited validation of contact conditions during prolonged and unsupervised wearable use. Most included studies used small samples, short recordings, highly structured protocols, static or quasi-static tasks, or predefined pressure conditions [16,18,20,21,41,42,44,45,47,49-51,53-56]. These designs are informative for characterizing responses under controlled conditions, but they provide limited evidence about the variability encountered during everyday wear. During free-living use, the sensor-skin interface may vary with posture transitions, arm movement, sweating, changes in skin hydration, tissue deformation, strap displacement, repeated device removal and reattachment, and prolonged wear [14,18,21,46,48,49]. Accordingly, the available laboratory evidence may not capture the full range or temporal instability of contact conditions in natural settings [14,18,21,46,49]. Current evidence does not establish whether a contact condition measured at the start of a recording remains stable over several hours, whether users reproduce similar conditions after repeated donning, or whether contact-related variability is consistent across days. Short supervised walking, stepping, or treadmill protocols should therefore be distinguished from longitudinal or free-living validation. Future studies should jointly document contact force or pressure, motion, posture, wear duration, and contextual factors rather than evaluating contact conditions only under isolated, idealized protocols [14,21,46,47,49,58]. Recent benchmarking efforts, including the QUMPHY D4 report [58], also indicate the value of standardized benchmark datasets in addition to algorithm development. For wearable reflectance PPG, future public datasets should document acquisition context more systematically, including site, device configuration, fixation strategy, wear duration, and other factors that may be associated with signal quality and estimation performance [46,47,58]. Repeated-measures studies should report within-session drift, between-day reproducibility, reattachment procedures, strap or fastening settings, and the frequency of contact-related signal-quality failures.

Gap Between Signal-Level Changes and Task-Level Estimation Bias

As shown in Figure 3, complete within-study pathways connecting measurement site, an intermediate signal-, waveform-, or feature-level domain, and a downstream task were identified in 8 studies [15,18,42,47,49,51,52,56]. The remaining studies did not report a complete site-intermediate-domain-downstream-task pathway [16,20,21,41,43-46,48,50,53-55]. This distinction is important because signal or feature differences observed under varying contact conditions do not necessarily correspond to clinically or algorithmically meaningful estimation differences. Their practical relevance is likely to depend on the target task, the signal properties used by the algorithm, and the selected reference method. Conversely, task-level differences should not be attributed solely to contact conditions without also considering motion, wavelength, device configuration, reference-measurement uncertainty, and participant characteristics. Future studies should therefore evaluate contact conditions, intermediate signal or feature behavior, and task-level estimation error within the same experimental protocol and against an appropriate reference standard. Stratified analyses across contact conditions, together with sensitivity or feature-ablation analyses, may help identify which signal properties are most closely associated with task-specific performance differences [15,21,42,49,52,57,58].

Limited Population and Measurement-Site Diversity

A further evidence gap concerns population and measurement-site diversity. The evidence was concentrated at the wrist and finger: 10 [15,21,41,43,45-47,49,52,55] of the 21 included studies involved wrist-based configurations, and 9 [20,42,44,45,47,50,51,54,55] involved finger measurements, whereas only 3 [18,55,56] examined the forehead, 1 [48] examined the sternum, and 1 [53] examined another anatomical site; these categories were nonmutually exclusive. Given differences in anatomical structure, device geometry, fixation strategy, and optical measurement configuration across sites, numerical pressure ranges reported at the wrist or finger should not be assumed to apply directly to the forehead, sternum, ear, or other anatomical locations [18,46-48,55,59,60]. Most human studies recruited healthy volunteers [18,20,21,41,42,44,45,47,49-54,56], leaving limited direct evidence for older adults, children, patients with cardiovascular or peripheral vascular conditions, and individuals with edema or impaired perfusion. Previous studies and systematic reviews have reported differences in pulse-oximetry accuracy across skin-pigmentation groups, particularly under lower oxygen-saturation conditions [61-64]. However, the included contact-condition literature did not adequately characterize whether skin pigmentation modifies the relationships between contact conditions and wearable reflectance PPG outcomes. More broadly, the reported specificity of reflectance PPG and SpO_2 estimation to participant and measurement-site characteristics suggests that population and site characteristics should be considered when interpreting contact-condition–related findings [59,60]. Future studies should report participant age, sex distribution, health status, the method used to characterize skin tone or pigmentation, and relevant vascular or tissue characteristics more consistently. Harmonized multisite protocols are needed to determine whether reported contact-condition responses are site-specific or reproducible across anatomical locations. Findings derived mainly from healthy adults and common peripheral sites should therefore be generalized cautiously to broader populations and other anatomical locations [46,59-64].

XSL•FO
RenderX

Scope Boundaries for Transmission-Mode and Hybrid PPG Systems

The focus on wearable or wearable-relevant reflectance PPG represents a deliberate scope boundary of this review. Accordingly, the mapped findings should be applied cautiously to transmission-mode and hybrid reflection-transmission PPG systems. Transmission-mode PPG remains important in wearable pulse oximetry, particularly in finger-clip, ring-type, earlobe, and in-ear devices, in which the light source and photodetector are positioned on opposite sides of relatively thin tissue [30,65-67]. These configurations may provide strong pulsatile signals for SpO_2 monitoring, but they involve different anatomical constraints and fixation mechanisms, including clips, rings, and in-ear fittings. Differences in optical path, contact area, tissue compression, and fixation mean that reported contact-condition responses may not be directly comparable across optical modes. Numerical force or pressure ranges and response patterns mapped in reflectance studies should therefore not be transferred directly to transmission-mode or hybrid devices. Several methodological principles remain relevant across optical modes, including distinguishing force from pressure and documenting geometry, effective contact area, fixation, and calibration. For hybrid systems, the reflectance and transmission interfaces should be described separately rather than summarized using a single wearing-tightness descriptor. For each sensing interface, studies should report the measurement site, optical mode, wavelength, emitter-detector geometry, fixation method, effective contact area, and contact force or pressure. Direct comparative studies, and potentially a separate evidence synthesis focused on transmission-mode and hybrid systems, are needed before cross-mode recommendations can be established.

Methodological Limitations and Scope of the Review

This review also has methodological limitations related to its scope and review process. Although the review followed a scoping review framework and was reported with reference to PRISMA-ScR and PRISMA-S [38-40], the protocol was not formally registered or publicly available. The search strategy was developed by the review authors and refined iteratively, but it did not undergo formal peer review by an information specialist or librarian. In addition, only English-language peer-reviewed journal articles and conference proceedings with accessible full texts were included, whereas non-English studies, non-peer-reviewed sources, conference abstracts, preprints, technical reports, theses, and other gray literature were excluded. Although 4 major databases were searched, publication bias and incomplete database coverage cannot be fully ruled out. Finally, because this was a scoping review, no formal risk-of-bias assessment, critical appraisal, or quantitative meta-analysis was performed; therefore, the findings should be interpreted as a structured mapping and qualitative synthesis of the available evidence rather than as a quantitative estimate of effect size or certainty of evidence.

Conclusions

Wearable reflectance PPG has become an important sensing technology for mHealth. As wearable PPG moves from controlled settings to real-world use, measurement quality varies with anatomical site, skin and tissue characteristics, motion, and the sensor-skin interface. This scoping review identified a multilevel relationship between contact conditions at the sensor-skin interface and wearable PPG outcomes. Differences in contact conditions were associated with changes in signal characteristics and derived features and, in a smaller body of studies, with differences in downstream physiological estimates such as HR, oxygen saturation, PTT or PAT, blood pressure–related indicators, and HRV. Therefore, contact force should not be treated as a minor wearing-related factor because contact-condition–related differences may extend from PPG signal acquisition to derived features and, in some studies, to downstream physiological estimation. Current evidence remains limited by short-term controlled study designs, inconsistent reporting of force or pressure conditions, insufficient description of contact area and probe geometry, and limited validation in long-term, unsupervised, or free-living settings. Only a few studies have directly linked contact-condition–related signal or feature differences with the robustness of downstream estimation models. Future research should therefore adopt standardized reporting of contact force or pressure, probe geometry, contact area, body site, wavelength, and fixation strategy. Long-term real-world validation, force-aware signal-quality assessment, robustness testing, adaptive attachment design, and context-aware modeling will be essential for improving the reliability and comparability of wearable PPG systems.

Acknowledgments

The authors thank all individuals who provided feedback and support during the development of this review. The authors also declare the use of generative AI (GAI) in the research and writing process. According to the GAIDeT (Generative AI Delegation Taxonomy) (2025) framework, the following tasks were delegated to GAI tools under full human supervision: text generation, proofreading and editing, translation, and reformatting. The GAI tool used was: ChatGPT 5.5/5.6 (OpenAI). Responsibility for the final manuscript lies entirely with the authors. GAI tools are not listed as authors and do not bear responsibility for the final outcomes.

Data Availability

All data extracted and charted in this scoping review are provided in Multimedia Appendices 4 and 5. No additional datasets were generated.

XSL•FO
RenderX

Funding

This work was supported in part by the National Natural Science Foundation of China (grant 62471132) and the National Key Research and Development Program of China (grant 2023YFC3603604).

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist and explanation.
[DOCX File , 64 KB - mhealth_v14i1e99333_app1.docx]

Multimedia Appendix 2
PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension) checklist.
[DOCX File , 15 KB - mhealth_v14i1e99333_app2.docx]

Multimedia Appendix 3
Search strategy and associated keywords.
[DOCX File , 40 KB - mhealth_v14i1e99333_app3.docx]

Multimedia Appendix 4
Baseline characteristics of the included studies with extracted contact-force/contact-pressure information.
[DOCX File , 41 KB - mhealth_v14i1e99333_app4.docx]

Multimedia Appendix 5
Supporting evidence tables for contact-condition–related findings in wearable reflectance photoplethysmography.
[DOCX File , 48 KB - mhealth_v14i1e99333_app5.docx]

References

1. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1736-1788 [FREE Full text] [doi: 10.1016/S0140-6736(18)32203-7] [Medline: 30496103]
2. GBD 2017 DALYsHALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1859-1922 [FREE Full text] [doi: 10.1016/S0140-6736(18)32335-3] [Medline: 30415748]
3. van Daalen KR, Zhang D, Kaptoge S, Paige E, Di Angelantonio E, Pennells L. Risk estimation for the primary prevention of cardiovascular disease: considerations for appropriate risk prediction model selection. The Lancet Global Health 2024 Aug;12(8):e1343-e1358. [doi: 10.1016/s2214-109x(24)00210-9]
4. Wong M, Dai Y, Ge J. Pan-vascular disease: what we have done in the past and what we can do in the future? Cardiol Plus 2024;9(1):1-5 [FREE Full text] [doi: 10.1097/CP9.0000000000000078] [Medline: 38584611]
5. Nie G, Zhao Q, Tang G, Li Y, Hong S. Artificial intelligence-derived photoplethysmography age as a digital biomarker for cardiovascular health. Commun Med (Lond) 2025 Nov 19;5(1):481 [FREE Full text] [doi: 10.1038/s43856-025-01188-9] [Medline: 41258400]
6. Gu X, Tang W, Han J, Sangha V, Liu F, Gowda SN, et al. Cardiac health assessment across scenarios and devices using a multimodal foundation model pretrained on data from 1.7 million individuals. Nat Mach Intell 2026 Feb 24;8(2):220-233. [doi: 10.1038/s42256-026-01180-5] [Medline: 41757248]
7. Nie G, Zhu J, Tang G, Zhang D, Geng S, Zhao Q, et al. A scoping review of deep learning methods for photoplethysmography data. Health Data Sci 2026 May 15:e1 (forthcoming) [FREE Full text] [doi: 10.34133/hds.0474]
8. Pankaj, Kumar A, Komaragiri R, Kumar M. A review on computation methods used in photoplethysmography signal analysis for heart rate estimation. Arch Computat Methods Eng 2021 May 10;29(2):921-940. [doi: 10.1007/s11831-021-09597-4]
9. Park J, Seok HS, Kim S, Shin H. Photoplethysmogram analysis and applications: an integrative review. Front Physiol 2021 Mar 1;12:808451 [FREE Full text] [doi: 10.3389/fphys.2021.808451] [Medline: 35300400]
10. Pereira T, Tran N, Gadhoumi K, Pelter MM, Do DH, Lee RJ, et al. Photoplethysmography based atrial fibrillation detection: a review. NPJ Digit Med 2020;3:3 [FREE Full text] [doi: 10.1038/s41746-019-0207-9] [Medline: 31934647]

XSL•FO
RenderX

11. Williams GJ, Al-Baraikan A, Rademakers FE, Ciravegna F, van de Vosse FN, Lawrie A, et al. Wearable technology and the cardiovascular system: the future of patient assessment. The Lancet Digital Health 2023 Jul;5(7):e467-e476. [doi: 10.1016/s2589-7500(23)00087-0]
12. Krittanawong C, Rogers AJ, Johnson KW, Wang Z, Turakhia MP, Halperin JL, et al. Integration of novel monitoring devices with machine learning technology for scalable cardiovascular management. Nat Rev Cardiol 2021 Feb 09;18(2):75-91 [FREE Full text] [doi: 10.1038/s41569-020-00445-9] [Medline: 33037325]
13. Cox J, Butlin M, Akeila E, Bentley G, Avolio A, Qasem A. Impact of finger contact force on aortic waveform parameters derived from finger photoplethysmography using a transfer function. New York, NY: IEEE; 2024 Presented at: 2024 IEEE SENSORS; October 20-23, 2024; Kobe, Japan p. 1-4. [doi: 10.1109/sensors60989.2024.10785128]
14. Orini M, Guvensen G, Jamieson A, Chaturvedi N, Hughes A. Movement, sweating, and contact pressure as sources of heart rate inaccuracy in wearable devices. 2022 Presented at: Computing in Cardiology 2022 (CinC 2022); September 4-7, 2022; Tampere, Finland p. 1-4. [doi: 10.22489/CinC.2022.237]
15. He Z, Cai Z, Sun J, Li J, Liu C, Yang C. A preliminary quantitative analysis of contact force effects on peripheral oxygen saturation prediction from reflectance photoplethysmography. IEEE Sensors J 2026:1-1. [doi: 10.1109/jsen.2026.3671734]
16. May JM, Mejía-Mejía E, Nomoni M, Budidha K, Choi C, Kyriacou PA. Effects of contact pressure in reflectance photoplethysmography in an in vitro tissue-vessel phantom. Sensors (Basel) 2021 Dec 16;21(24):8421 [FREE Full text] [doi: 10.3390/s21248421] [Medline: 34960512]
17. Wang L, Liu S, Li G, Zhu R. Interface sensors with skin piezo-thermic transduction enable motion artifact removal for wearable physiological monitoring. Biosens Bioelectron 2021 Sep 15;188:113325. [doi: 10.1016/j.bios.2021.113325] [Medline: 34030098]
18. Dresher R, Mendelson Y. Reflectance forehead pulse oximetry: effects of contact pressure during walking. Conf Proc IEEE Eng Med Biol Soc 2006;2006:3529-3532. [doi: 10.1109/IEMBS.2006.260136] [Medline: 17946185]
19. Liu C, Correia R, Ballaji H, Korposh S, Hayes-Gill B, Morgan S. Optical fibre-based pulse oximetry sensor with contact force detection. Sensors (Basel) 2018 Oct 26;18(11):3632 [FREE Full text] [doi: 10.3390/s18113632] [Medline: 30373119]
20. Sirkiä J, Panula T, Kaisti M. The effects of external pressure on multi-wavelength photoplethysmography signals. 2021 Presented at: 2021 Computing in Cardiology (CinC); September 12-15, 2021; Brno, Czech Republic p. 1-4. [doi: 10.23919/CinC53138.2021.9662771]
21. Scardulla F, Cosoli G, Gnoffo C, Antognoli L, Bongiorno F, Diana G, et al. Experimental analysis on the effect of contact pressure and activity level as influencing factors in PPG sensor performance. Sensors (Basel) 2025 Jul 18;25(14):4477 [FREE Full text] [doi: 10.3390/s25144477] [Medline: 40732604]
22. Ryals S, Chiang A, Schutte-Rodin S, Chandrakantan A, Verma N, Holfinger S, et al. Photoplethysmography-new applications for an old technology: a sleep technology review. J Clin Sleep Med 2023 Jan 01;19(1):189-195 [FREE Full text] [doi: 10.5664/jcsm.10300] [Medline: 36123954]
23. Sun Y, Thakor N. Photoplethysmography revisited: from contact to noncontact, from point to imaging. IEEE Trans Biomed Eng 2016 Mar;63(3):463-477. [doi: 10.1109/tbme.2015.2476337]
24. Elgendi M, Fletcher R, Liang Y, Howard N, Lovell NH, Abbott D, et al. The use of photoplethysmography for assessing hypertension. NPJ Digit Med 2019 Jun 26;2(1):60 [FREE Full text] [doi: 10.1038/s41746-019-0136-7] [Medline: 31388564]
25. Chen G, Zou L, Ji Z. A review: blood pressure monitoring based on PPG and circadian rhythm. APL Bioeng 2024 Sep;8(3):031501 [FREE Full text] [doi: 10.1063/5.0206980] [Medline: 39049850]
26. Faris Ali N, Al-Nashash H, Elfadel IM, Atef M. A review of photoplethysmography-based blood pressure monitoring: from cloud-based machine learning to TinyML edge deployment. IEEE Access 2025;13:204210-204240. [doi: 10.1109/access.2025.3639253]
27. El-Hajj C, Kyriacou P. A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure. Biomedical Signal Processing and Control 2020 Apr;58:101870. [doi: 10.1016/j.bspc.2020.101870]
28. Almarshad MA, Islam MS, Al-Ahmadi S, BaHammam AS. Diagnostic features and potential applications of PPG signal in healthcare: a systematic review. Healthcare (Basel) 2022 Mar 16;10(3):547 [FREE Full text] [doi: 10.3390/healthcare10030547] [Medline: 35327025]
29. Charlton PH, Pilt K, Kyriacou PA. Establishing best practices in photoplethysmography signal acquisition and processing. Physiol Meas 2022 May 25;43(5):050301 [FREE Full text] [doi: 10.1088/1361-6579/ac6cc4] [Medline: 35508148]
30. Castaneda D, Esparza A, Ghamari M, Soltanpur C, Nazeran H. A review on wearable photoplethysmography sensors and their potential future applications in health care. Int J Biosens Bioelectron 2018;4(4):195-202 [FREE Full text] [doi: 10.15406/ijbsbe.2018.04.00125] [Medline: 30906922]
31. Kim KB, Baek HJ. Photoplethysmography in wearable devices: a comprehensive review of technological advances, current challenges, and future directions. Electronics 2023 Jul 03;12(13):2923. [doi: 10.3390/electronics12132923]
32. Knight S, Lipoth J, Namvari M, Gu C, Hedayati M, Syed-Abdul S, et al. The accuracy of wearable photoplethysmography sensors for telehealth monitoring: a scoping review. Telemed J E Health 2023 Jun 01;29(6):813-828. [doi: 10.1089/tmj.2022.0182] [Medline: 36288566]

XSL•FO
RenderX

33. Namvari M, Lipoth J, Knight S, Jamali AA, Hedayati M, Spiteri RJ, et al. Photoplethysmography enabled wearable devices and stress detection: a scoping review. J Pers Med 2022 Oct 31;12(11):1792 [FREE Full text] [doi: 10.3390/jpm12111792] [Medline: 36579537]
34. Alshehhi A, H.Abbasi Q, Ghannam R. Photoplethysmography for measuring cognitive load in XR environments: a systematic review. Advanced Sensor Research 2026 Mar 24;5(3):e00142. [doi: 10.1002/adsr.202500142]
35. Fahim MMH. Systematic review of online photoplethysmography (PPG) tool kits. International Journal of Innovative Science and Research Technology (IJISRT) 2025 Dec 18;10(12):883. [doi: 10.38124/ijisrt/25dec587]
36. Penmetcha V, Rambabu L, Smith BG, Mantle O, Edmiston T, Hobbs L, et al. Evaluating diversity in open photoplethysmography datasets: protocol for a systematic review. JMIR Res Protoc 2025 Oct 01;14:e73040-e73040 [FREE Full text] [doi: 10.2196/73040] [Medline: 41032881]
37. Charlton P. Photoplethysmography datasets. GitHub. 2022 Aug 2. URL: https://peterhcharlton.github.io/post/ppg_datasets/ [accessed 2026-05-28]
38. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
39. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
40. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, PRISMA-S Group. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. Syst Rev 2021 Jan 26;10(1):39 [FREE Full text] [doi: 10.1186/s13643-020-01542-z] [Medline: 33499930]
41. Sim J, Ahn B, Doh I. A contact-force regulated photoplethysmography (PPG) platform. AIP Adv 2018;8(4):045210. [doi: 10.1063/1.5020914]
42. Chandrasekhar A, Yavarimanesh M, Natarajan K, Hahn J, Mukkamala R. PPG sensor contact pressure should be taken into account for cuff-less blood pressure measurement. IEEE Trans Biomed Eng 2020 Nov;67(11):3134-3140. [doi: 10.1109/tbme.2020.2976989]
43. Liu Z, Xiang T, Zhou R, Ji N, Zhang Y. A multi-channel photoplethysmography array with contact-force regulation for tonoarteriographic imaging. Conn Health Telemed 2024 Feb 19;3(1):300001. [doi: 10.20517/chatmed.2023.11]
44. Přibil J, Přibilová A, Dermek T. Wearable PPG optical sensor enabling contact pressure and skin temperature measurement. J Electr Eng 2025;76(2):134-146 [FREE Full text] [doi: 10.2478/jee-2025-0014]
45. Fortin P, Blum J, Weill-Duflos A. Contact force estimation from raw photoplethysmogram signal. 2020 Presented at: 2020 IEEE SENSORS; October 25-28, 2020; Virtual p. 1-4. [doi: 10.1109/SENSORS47125.2020.9278658]
46. Charlton PH, Marozas V, Mejía-Mejía E, Kyriacou PA, Mant J. Determinants of photoplethysmography signal quality at the wrist. PLOS Digit Health 2025 Jun 27;4(6):e0000585 [FREE Full text] [doi: 10.1371/journal.pdig.0000585] [Medline: 40577339]
47. Ho MY, Pham HM, Saeed A, Ma D. WF-PPG: a wrist-finger dual-channel dataset for studying the impact of contact pressure on PPG morphology. Sci Data 2025 Feb 03;12(1):200 [FREE Full text] [doi: 10.1038/s41597-025-04453-7] [Medline: 39900957]
48. Chan M, Ganti V, Heller J, Abdallah C, Etemadi M, Inan O. Enabling continuous wearable reflectance pulse oximetry at the sternum. Biosensors (Basel) 2021 Dec 17;11(12):521 [FREE Full text] [doi: 10.3390/bios11120521] [Medline: 34940278]
49. Scardulla F, D'Acquisto L, Colombarini R, Hu S, Pasta S, Bellavia D. A study on the effect of contact pressure during physical activity on photoplethysmographic heart rate measurements. Sensors (Basel) 2020 Sep 05;20(18):5052 [FREE Full text] [doi: 10.3390/s20185052] [Medline: 32899540]
50. Lambert Cause J, Solé Morillo ?, García-Naranjo JC, Stiens J, da Silva B. The impact of contact force on signal quality indices in photoplethysmography measurements. Applied Sciences 2024 Jun 29;14(13):5704. [doi: 10.3390/app14135704]
51. Sirkiä J, Panula T, Kaisti M. Investigating the impact of contact pressure on photoplethysmograms. Biomedical Engineering Advances 2024 Jun;7:100123. [doi: 10.1016/j.bea.2024.100123]
52. Hu C, Pham H, Ma D. Morphology-aware HRV estimation from wrist PPG in sedentary scenarios. 2025 Presented at: UbiComp '25: The 2025 ACM International Joint Conference on Pervasive and Ubiquitous Computing/ISWC ACM International Symposium on Wearable Computers; October 12-13, 2025; Espoo, Finland p. 745-750. [doi: 10.1145/3714394.3756171]
53. Grabovskis A, Marcinkevics Z, Rubins U, Kviesis-Kipge E. Effect of probe contact pressure on the photoplethysmographic assessment of conduit artery stiffness. J Biomed Opt 2013 Feb;18(2):27004. [doi: 10.1117/1.JBO.18.2.027004] [Medline: 23377011]
54. Teng XF, Zhang YT. Theoretical study on the effect of sensor contact force on pulse transit time. IEEE Trans Biomed Eng 2007 Aug;54(8):1490-1498. [doi: 10.1109/TBME.2007.900815] [Medline: 17694870]
55. Fine I, Kaminsky A. Possible error in reflection pulse oximeter readings as a result of applied pressure. J Healthc Eng 2019;2019:7293813 [FREE Full text] [doi: 10.1155/2019/7293813] [Medline: 31781359]

XSL•FO
RenderX

56. Dresher R, Mendelson Y. A new reflectance pulse oximeter housing to reduce contact pressure effects. 2006 Presented at: IEEE 32nd Annual Northeast Bioengineering Conference; April 1-2, 2006; Easton, PA p. 35-36. [doi: 10.1109/NEBC.2006.1629746]
57. Elgendi M, Haugg F, Fletcher RR, Allen J, Shin H, Alian A, et al. Recommendations for evaluating photoplethysmography-based algorithms for blood pressure assessment. Commun Med (Lond) 2024 Jul 12;4(1):140. [doi: 10.1038/s43856-024-00555-2] [Medline: 38997447]
58. Hackstein U, Alastruey J, Aston P, Bench C, Charlton P, Coquelin L. Benchmark problems and benchmark datasets for the evaluation of machine and deep learning methods on photoplethysmography signals: the D4 report from the QUMPHY project. arXiv. Preprint published online April 1, 2026. URL: https://arxiv.org/abs/2604.01398 [accessed 2026-08-17]
59. Mohamed Tanveejul APS, Alex A, Preejith SP. A study on the subject and location specificity in reflectance based SpO2 estimation using R-value based calibration curve. 2020 Presented at: 2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA); June 1 to July 1, 2020; Bari, Italy p. 1-6. [doi: 10.1109/MeMeA49120.2020.9137227]
60. Berwal D, Kuruba A, Shaikh AM, Udupa A, Baghini MS. SpO2 measurement: non-idealities and ways to improve estimation accuracy in wearable pulse oximeters. IEEE Sensors J 2022 Jun 15;22(12):11653-11664. [doi: 10.1109/jsen.2022.3170069]
61. Bickler P, Feiner J, Severinghaus J. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. Anesthesiology 2005 Apr;102(4):715-719. [doi: 10.1097/00000542-200504000-00004] [Medline: 15791098]
62. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial bias in pulse oximetry measurement. N Engl J Med 2020 Dec 17;383(25):2477-2478. [doi: 10.1056/nejmc2029240]
63. Shi C, Goodall M, Dumville J, Hill J, Norman G, Hamer O, et al. The accuracy of pulse oximetry in measuring oxygen saturation by levels of skin pigmentation: a systematic review and meta-analysis. BMC Med 2022 Aug 16;20(1):267 [FREE Full text] [doi: 10.1186/s12916-022-02452-8] [Medline: 35971142]
64. Martin D, Johns C, Sorrell L, Healy E, Phull M, Olusanya S, et al. Effect of skin tone on the accuracy of the estimation of arterial oxygen saturation by pulse oximetry: a systematic review. Br J Anaesth 2024 May;132(5):945-956 [FREE Full text] [doi: 10.1016/j.bja.2024.01.023] [Medline: 38368234]
65. Tamura T, Maeda Y, Sekine M, Yoshida M. Wearable photoplethysmographic sensors—past and present. Electronics 2014 Apr 23;3(2):282-302. [doi: 10.3390/electronics3020282]
66. Charlton PH, Allen J, Bailón R, Baker S, Behar JA, Chen F, et al. The 2023 wearable photoplethysmography roadmap. Physiol Meas 2023 Nov 29;44(11):111001 [FREE Full text] [doi: 10.1088/1361-6579/acead2] [Medline: 37494945]
67. Huang C, Chan M, Chen C, Lin B. Novel wearable and wireless ring-type pulse oximeter with multi-detectors. Sensors (Basel) 2014 Sep 19;14(9):17586-17599 [FREE Full text] [doi: 10.3390/s140917586] [Medline: 25244586]

Abbreviations

AC: alternating current
BP: blood pressure
DC: direct current
HR: heart rate
HRV: heart rate variability
kPa: kilopascal
mHealth: mobile health
mm Hg: millimeters of mercury
PAT: pulse arrival time
PPG: photoplethysmography
PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
PTT: pulse transit time
SNR: signal-to-noise ratio
SpO2: peripheral oxygen saturation

XSL•FO
RenderX

Edited by S Brini; submitted 24.Apr.2026; peer-reviewed by CT Hsiao, M Chee; comments to author 25.May.2026; revised version received 14.Aug.2026; accepted 14.Aug.2026; published 24.Aug.2026.

Please cite as:
Yang C, Xie J, He Z, Li J, Liu C
Mechanical Contact Conditions in Wearable Reflectance Photoplethysmography: Scoping Review
JMIR Mhealth Uhealth 2026;14:e99333
URL: https://mhealth.jmir.org/2026/1/e99333
doi:10.2196/99333
PMID:

©Chenxi Yang, Jiahang Xie, Zifei He, Jianqing Li, Chengyu Liu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 24.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Continuous Temperature Monitoring via Wearable Devices for Fever or Infection Management in Acute Care Hospitals: Scoping Review of Clinical Implementation and Data Analytics

Yuwei Liu[1,2]; Qimeng Zhao[2]; Ka Li[1,3,4]; Dawn Dowding[2]

[1]West China School of Nursing / West China Hospital, Sichuan University, Chengdu, Sichuan, China

[2]Division of Nursing, Midwifery and Social Work, School of Health Sciences, University of Manchester, Manchester, United Kingdom

[3]Nursing Key Laboratory of Sichuan Province, Chengdu, Sichuan, China

[4]Medicine and Engineering Interdisciplinary Research Laboratory of Nursing & Materials, West China Hospital, Sichuan University, Chengdu, Sichuan, China

Corresponding Author:
Ka Li
West China School of Nursing / West China Hospital
Sichuan University
No.37 Guoxue Lane
Chengdu, Sichuan, 610041
China
Phone: 86 18980601488
Email: lika127@126.com

Abstract

Background: Infections are a major health concern in hospitalized patients. Fever is an early sign of infection, making temperature monitoring essential for infection surveillance. Wearable devices are increasingly being explored for continuous temperature monitoring in acute care hospitals, but how temperature data from wearables are monitored, presented, and used to support fever or infection management in clinical practice remains poorly understood.

Objective: This study aims to map existing evidence on the use of temperature data from wearables for fever or infection management in acute care hospitals, focusing on wearables' characteristics; data transmission, storage, and presentation strategies; data preprocessing and analytic approaches; and the maturity of wearables' clinical integration.

Methods: We searched MEDLINE, Embase, Web of Science, CINAHL, and IEEE Xplore for publications from January 2013 to July 2026 using the population, concept, context (PCC) framework. Primary research studies using wearables to monitor patients' body temperature in acute care hospitals for fever or infection management, published in English or Chinese, were eligible. Temperature data analyses were classified according to the descriptive, diagnostic, predictive, and prescriptive analytics framework. The clinical integration of wearables was assessed using a tailored maturity framework modified from the World Health Organization's stages of maturity for digital health interventions.

Results: We included 29 publications from 26 studies, published between 2018 and 2026. Eighteen wearable devices were identified, monitoring temperature at the axilla, chest, wrist, or upper arm. Continuous data streams were predominantly transmitted in real time, while 5 studies used non–real-time batch uploads, periodic synchronization, or device-memory downloads. Only 7 studies presented wearable temperature data in clinical settings, displaying data on mobile devices and/or centralized monitoring stations; only 1 study reported enabling alerts through electronic health records (EHRs). Invalid sensor data filtering methods varied across studies, such as physiological thresholds, firmware quality scores, and statistical outlier detection. Temperature data were most frequently analyzed by descriptive analytics (n=19) to depict the frequency, timing, and duration of fever episodes, followed by diagnostic analytics (n=6) to identify risk factors or distinguish causes of fever, and predictive analytics (n=9) to forecast impending fever or infections. None of the included studies used prescriptive analytics. For clinical integration, most studies were at the clinical validation stage (n=9) or clinical research stage (n=17), corresponding to an early stage of maturity; 3 studies reached the routine clinical practice stage, and none progressed to multicenter implementation or full integration stages.

Conclusions: While promising, the clinical integration of wearables remains at an early stage of maturity. Critical gaps exist in translating temperature data into actionable clinical insights, with limited data presentation, poor EHR interoperability, and

XSL•FO
RenderX

underdeveloped analytic approaches. Future research should prioritize standardized data cleaning frameworks, workflow integration, and clinical interpretation to facilitate the active use of wearable data in clinical decision-making.

Trial Registration: OSF Registries osf.io/v6sp8; https://osf.io/v6sp8

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2025-103630

(JMIR Mhealth Uhealth 2026;14:e89773) doi:10.2196/89773

KEYWORDS

body temperature; fever; infections; mHealth; mobile health; scoping review; wearable device

Introduction

Infections are a major health burden in hospitalized patients, particularly those acquired within health care settings, posing a significant threat to global patient safety [1]. An estimated 5.4% to 7.8% of patients in acute care hospitals across Europe [2] and 12% to 18% in the Eastern Mediterranean Region [3] are affected by health care–associated infections. Timely recognition and earlier treatment of infections are crucial for reducing morbidity and mortality [4]. As one of the earliest and most apparent symptoms, fever is a vital diagnostic and monitoring indicator that reflects the clinical course of many infectious diseases [5]. Therefore, temperature monitoring serves as a fundamental component of infection surveillance and management, supporting clinical decisions on whether to initiate a fever workup or prescribe antibiotics.

Traditional temperature measurement mainly relies on intermittent manual recording by nurses. However, the intervals between observations cannot fully capture temperature changes and may miss febrile events or delay the detection of infections [6]. In recent years, wearable technologies have enabled continuous temperature monitoring in acute care hospitals [7]. Wearable devices are electronic sensors designed to be directly worn on the body to perceive and record various physiological parameters [8], encompassing a range of sensor-based devices such as patches, rings, wristbands, and smartwatches. Featuring wireless connectivity, real-time data collection, and remote monitoring capabilities, wearables have attracted growing interest for clinical applications. Recent studies have explored the accuracy and feasibility of wearables for temperature monitoring in diverse clinical settings, such as surgical wards, emergency departments, oncology wards, and intensive care units (ICUs) [9].

The high-frequency temperature data generated by wearables contain rich physiological information, which holds the potential to transform infection care by providing data-driven insights into inflammatory responses and infection trajectories [10]. With advanced analytical approaches, these data can be used to derive temperature features that capture temporal dynamics such as trends, patterns, and circadian rhythms. These features could serve as digital biomarkers to support early detection of febrile events, personalized alerts, and timely interventions [11]. For instance, a fever detection algorithm using a chest patch captured fever episodes 4.3 hours earlier than clinical assessments in patients at high risk of febrile neutropenia [12]. Vargas et al [13] demonstrated that entropy analysis of continuous body temperature could distinguish bacterial infections from other causes of fever, potentially enabling earlier antibiotic cessation. More broadly, wearable data are increasingly integrated into AI predictive models of infectious disease within real-time surveillance and clinical decision support ecosystems [14].

However, translating this potential into practice remains hindered by considerable hurdles when deploying wearables within the complex, high-noise environments of acute care hospitals. Wearable data collected in such environments typically contain various anomalies, such as missing values, outliers, motion artifacts, random noise, recording errors, and redundant data [15]. Without effective filtering, these anomalies may trigger false alerts and contribute to alert fatigue. Although wearables technically streamline routine temperature measurement and reduce time and effort [16], they may counterproductively increase nurses' workload without proper clinical integration [17]. Moreover, while continuous monitoring captures subtle physiological fluctuations, difficulties in interpreting the sheer volume of temperature data often prevent these data from being translated into actionable clinical insights [18], resulting in a paradoxical situation where cognitive overload and data waste coexist. Given these limitations, wearables may simply be used as tools for passive data collection rather than active clinical decision support [19], with their potential remaining largely underexplored.

To better integrate wearables into clinical practice, there is a need to incorporate wearable data into existing clinical workflows and facilitate data interpretation to support clinical decision-making. However, most current studies remain focused on evaluating the accuracy and feasibility of wearables in clinical care [20]. While existing literature reviews have summarized available wearables and their clinical validation findings [9,21], there is a lack of synthesized evidence on how temperature data from wearables are processed, presented, and interpreted in clinical practice. For the next step, it is essential to shift the research focus from device performance assessment toward proactive data usage in clinical contexts. Therefore, to bridge the gap between technical feasibility and clinical utility, this study aims to review evidence on the use of wearables for continuous temperature monitoring in acute care hospitals to support fever or infection management, focusing on strategies for data transmission, presentation, and use, and the overall maturity of these wearables' clinical integration.

XSL•FO
RenderX

Methods

Overview

We chose a scoping review because it provides the flexibility to synthesize heterogeneous evidence across different wearable device types, clinical contexts, and data analytical methods in this rapidly evolving field. We conducted the review following the Joanna Briggs Institute (JBI) manual for scoping reviews [22] and reported it in alignment with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [23]. The protocol was registered on the Open Science Framework (OSF) [24] and published [25] ahead of this review. The PRISMA-ScR checklist is provided in Multimedia Appendix 1.

Research Question

Our review addressed the following questions:

- Question 1: what types of wearable devices have been used in acute care hospitals to monitor body temperature for fever or infection management?
- Question 2: how are the temperature data from wearable devices monitored, transmitted, and presented in acute care hospitals?
- Question 3: what analytical approaches have been used to process temperature data from wearable devices to support the management of fever or infections?
- Question 4: to what extent have these wearable devices been integrated into clinical practice?

Selection of Studies

We adopted the Population, Concept, Context (PCC) framework to guide the development of search strategies and eligibility criteria.

- Population: hospitalized patients who are confirmed, suspected, or at risk of developing fever or infections.
- Concept: wearable devices that are capable of wireless, continuous, noninvasive body temperature monitoring by being worn on the external body surface [8]; fever was defined as a pathological elevation in the body's set-point temperature [26,27]; infections refer to the invasion and multiplication of microorganisms, including bacteria, viruses, fungi, or parasites, in body tissues. We define continuous monitoring as automated time-series data acquisition at a predefined sampling frequency in real time or near real time (eg, every 1 minute or 5 minutes) [28], without manual initiation of each measurement.
- Context: acute care hospitals, which provide short-term, hospital-based care for patients with acute illness or injury. All inpatient wards and emergency departments were eligible.

The MEDLINE, Embase, Web of Science, CINAHL, and IEEE Xplore databases were searched from January 2013 to July 2026, as bibliometric analysis indicates a rapid increase in research on wearable technologies for health monitoring since 2013 [29,30]. The search was initially conducted in February 2025 and updated in July 2026 (both by reviewer YL). Search terms were organized into categories, including "wearable*," "body temperature," "fever," "infecti*," "patient*," and "hospital." Specific brands of wearables were identified from previous reviews and added to search strings [7,9]. Gray literature, including conference proceedings, preprints, and dissertations, was also searched. The detailed search strategies for databases and websites for gray literature are provided in Multimedia Appendix 2. The full search strategy was reviewed by a professional medical librarian using the PRESS (Peer Review of Electronic Search Strategies) checklist [31].

We uploaded the search results to Rayyan [32] to remove duplicates and conduct online screening. During the pilot screening, we compiled a preliminary list of potential temperature monitoring devices, discussed which qualified as wearables, and reached consensus on the eligible boundaries before formal screening. Then, a 2-step process, including title and abstract screening and full-text screening, was used. One reviewer (YL) screened all literature, with 20% randomly selected and independently screened by a second reviewer (QZ) to assess screening consistency. The interrater agreement using Cohen κ coefficients was 0.778 and 0.746 for each screening step, indicating substantial agreement. Disagreements were resolved through discussion with a third reviewer (DD). For multiple records derived from the same study, such as conference abstracts and research protocols, the formally published article was retained. The reference lists and citation lists of eligible articles and relevant literature reviews were manually searched to identify additional publications.

Studies were included if they used wearables to monitor patients' body temperature in acute care hospitals for the purpose of fever or infection management, including but not limited to detection, diagnosis, prediction, treatment, or prognostic assessment. Primary research studies published in English or Chinese were eligible regardless of study design or publication type. Studies were excluded if they focused on monitoring local tissue temperature (eg, wound or foot), reported the technical design or development of wearables, assessed device accuracy exclusively without clinical interpretation of temperature data, or focused on the prevalence of epidemic infectious diseases rather than clinical care.

Data Extraction and Analysis

Data were extracted from the main text and supplementary materials using a predefined data extraction form (Multimedia Appendix 3), including bibliographic and contextual attributes, wearable device characteristics, and strategies for data transmission, storage, and presentation. We also extracted temperature data analysis strategies, including data preprocessing, derived temperature features, and analytic methods. We classified temperature data analyses according to the descriptive, diagnostic, predictive, and prescriptive analytics framework [33,34], and further assessed the maturity stage of wearables' integration in clinical practice with reference to the World Health Organization's (WHO) manual, "*Monitoring and Evaluating Digital Health Interventions*" [35]. We also identified the role of nurses by reviewing the main text, author lists, and affiliations. We did not assess the methodological quality of the included publications in an attempt to map all

XSL•FO
RenderX

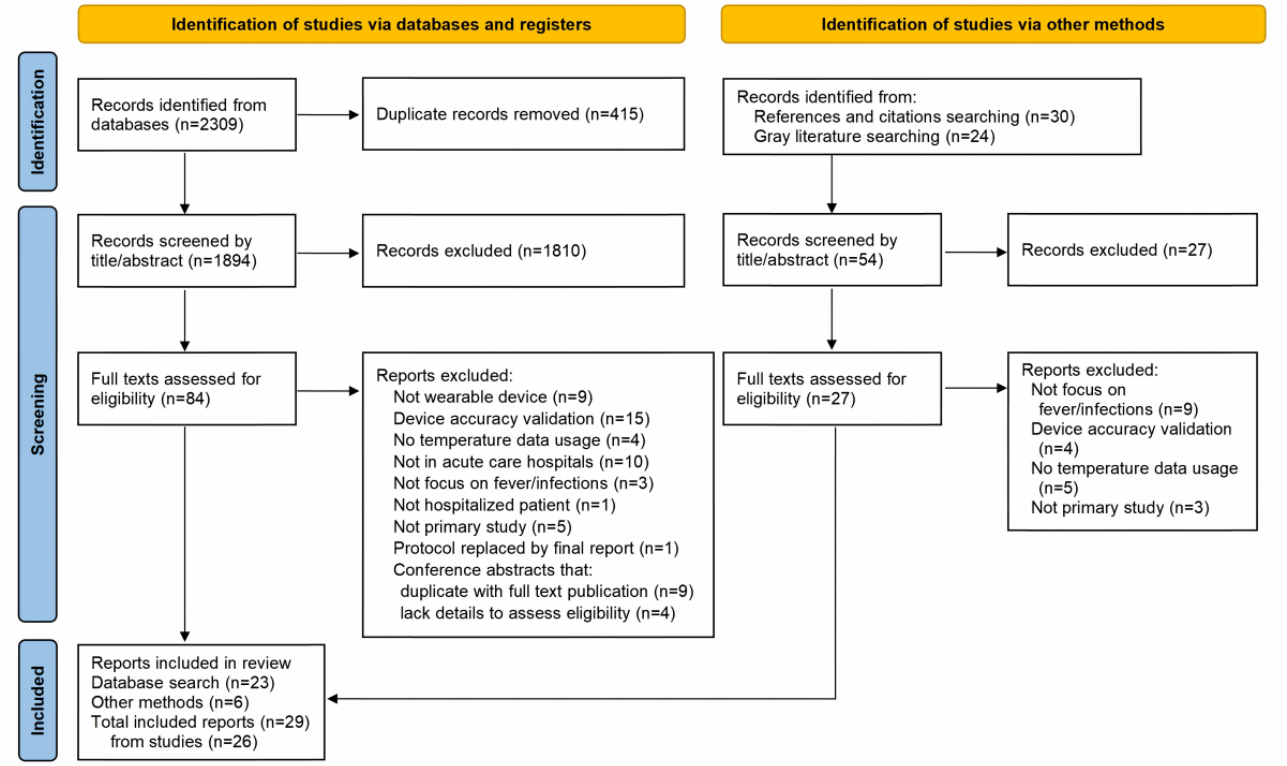
available evidence, and quality appraisal is not mandatory in a scoping review according to JBI methodology [36]. Data extraction was performed by one reviewer (YL). A second reviewer (QZ) independently extracted data from a random 20% of the included publications to evaluate the reliability of the data extraction process. Discrepancies were discussed to reach a resolution, and a third researcher (DD) was consulted for unresolved discrepancies. A descriptive synthesis of the findings was presented in tables and figures.

Results

Study Selection

A total of 2309 records were identified through database searches. After duplicate removal (n=415), 1948 records (1894 from database searches and 54 from manual identification) were screened by title and abstract. Subsequently, 111 records (84 from database search screening and 27 from other methods) underwent full-text screening. We included 29 publications (from 26 studies) in this review. Figure 1 presents the study selection process. Three publications [6,37,38] reported data derived from the same patient cohort in one study, while 2 publications [39,40] analyzed data from another study.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.



Study Characteristics

The 29 included publications (from 26 studies) comprised 25 journal articles (including 1 research letter with original data) and 4 conference abstracts, published in English (28/29, 97%) and Chinese (1/29, 3%) between 2018 and 2026. The included studies were conducted in Asia (9/26, 35%), Europe (8/26, 31%), North America (6/26, 23%), and Oceania (3/26, 12%). The majority were prospective observational studies (20/26, 77%), while 3 were retrospective observational studies, 1 was a randomized controlled trial, and 2 observational studies did not specify whether they were prospective or retrospective. Most studies were single-center (24/26, 92%), with only 2 studies being multicenter. Study populations were adult patients (17/26, 65%), pediatric patients (5/26, 19%), women in labor or pregnancy (3/26, 12%), and older adult patients (1/26, 4%). Clinical settings comprised oncology wards (9/26, 35%), surgical wards (6/26, 23%), medical wards (6/26, 23%), prenatal or delivery rooms (3/26, 12%), and ICUs (2/26, 8%). The primary events of interest were febrile events (15/26, 58%; including fever, febrile neutropenia, cytokine release syndrome [CRS], and intrapartum fever) and infections (11/26, 42%; including sepsis, pneumonia, intraamniotic infection, postoperative infectious complications, and COVID-19).

Wearable Device Characteristics and Monitoring Parameters

A total of 18 wearable devices were identified (Table 1), including 4 chest patches, 6 axillary patches, 2 smartwatches, 2 wristbands, 3 upper arm or leg bands, and 1 neck patch. Most devices (11/18, 61%) monitored multiple physiological parameters, while 7 devices were dedicated temperature sensors. Seven devices measured axillary temperature, while 11 devices measured skin temperature at various sites, including the chest (n=3), wrist (n=4), upper arm or leg (n=3), and neck (n=1). Five devices provided estimated core temperature (VitalPatch, CORE, Everion, Moni-Patch, and iThermonitor), with measurements derived from axillary or skin temperature readings. Axillary

XSL•FO
RenderX

patches were the most frequently used device, with the highest number of studies reported in oncology wards (Figure 2).

Among the 18 identified wearables, 78% (14/18) were medical-grade and 11% (2/18) were consumer-grade, while the grade was not specified for 2 devices. Ten wearables were approved by the US Food and Drug Administration (FDA), with the majority (8/10) approved as class II devices for use by health care professionals on general patients in health care settings, while 2 were approved for home use or ovulation tracking (Multimedia Appendix 3). However, reporting on the regulatory status and clinical validation of wearables showed gaps. Among the included publications, only 41% (12/29) reported the device grade, and just 35% (10/29) confirmed FDA approval or Conformité Européenne (CE) marking. Only 35% (10/29) of publications validated the wearables by comparing the temperature data with standard clinical measurements as part of their results, and 17% (5/29) cited previously published validation findings. One article noted that the wearable device used in their study had not yet been clinically validated [41], and 2 articles mentioned manufacturer preclinical validation [42,43]. The remaining publications (11/29, 38%) did not mention clinical validation.

XSL•FO
RenderX

Table 1. Characteristics of wearables identified from the included studies.

Wearable sensor; grade; regulatory approval	Physical location	Physical metrics	Monitoring frequency	Battery life
TempTraq TT-100; medical; FDA[a]/CE[b]	Axilla	Axillary temperature	10 seconds[c]; 2 minutes[d]	1 day[e]
iThermonitor; medical; FDA	Axilla	Axillary temperature (estimate core temperature)	4 seconds[d]	30 days
Lifetemp; medical[e]; CE[e]	Axilla	Axillary temperature, heart rate, and respiratory rate intervals	1 minute[d]	5 days[e]
SteadyTemp; medical[e]; FDA/CE[e]	Axilla	Axillary temperature	5 minutes[d]	10 days[e]
thynC MT100D; medical[e]; MFDS[e,f]	Axilla	Axillary temperature	1 minute[d]	Not reported
FeverScout; medical[e]; FDA/CE[e]	Axilla	Axillary temperature	15 seconds[c]; 2 minutes[d]	Rechargeable[e]
SensiumVitals; medical; CE	Chest patch with an axillary sensor	Axillary temperature, heart rate, and respiratory rate	2 minutes[d]	5 days
VitalPatch; medical; FDA/CE[e]	Chest	Skin temperature, Core temperature, heart rate, heart rate variability, and respiratory rate	1 minute[d]	5 days[e]
Biobeat BB-613WP; medical; FDA	Chest	Skin temperature, heart rate, blood pressure, respiratory rate, cardiac output, cardiac index, and SVR[g]	15 minutes[d]	6 days
Biosensor Voyage; not reported; not reported	Chest	Skin temperature, heart rate, and respiratory rate	30 seconds[d]	3 days
HEARThermo wristband; not reported; not reported	Wrist	Skin temperature and heart rate	10 seconds[d]	Not reported
Ava bracelet; medical[e]; FDA[e]	Wrist	Skin temperature, pulse rate, heart rate variability, and respiratory rate	10 seconds[d]	Rechargeable
HUAWEI WATCH 3; consumer; N/A[h]	Wrist	Skin temperature, heart rate, respiratory rate, SpO_2[i], and cough sound	25 Hz[c]	Not reported
Empatica E4 Watch; medical; FDA[e]	Wrist	Skin temperature, pulse rate, respiratory rate, electrodermal activity, SpO_2, and activity	4 Hz[c]	3 days (rechargeable)
Everion; medical; FDA/CE	Upper arm	Skin temperature, core temperature, heart rate, heart rate variability, respiratory rate, electrodermal activity, SpO_2, blood pressure wave, and activity	1 second[c]; 1 minute[d]	Rechargeable
CORE; consumer; N/A	Upper arm or chest	Core temperature	1 second[d]	Rechargeable
Gen2; medical; FDA	Upper arm	Skin temperature, respiratory rate, pulse rate, SpO_2, motion, and blood pressure	2 seconds[d]	Rechargeable[e]
Moni-Patch; medical[e]; PMDA[e,j]	Neck	Core temperature	Not reported	Not reported

[a]FDA: US Food and Drug Administration.

[b]CE: Conformité Européenne.

[c]Indicates sampling frequency.

[d]Indicates recording frequency.

[e]Information not reported in the included studies was obtained from official regulatory databases/websites and manufacturers' websites.

[f]MFDS: Ministry of Food and Drug Safety, Republic of Korea.

[g]SVR: systemic vascular resistance.

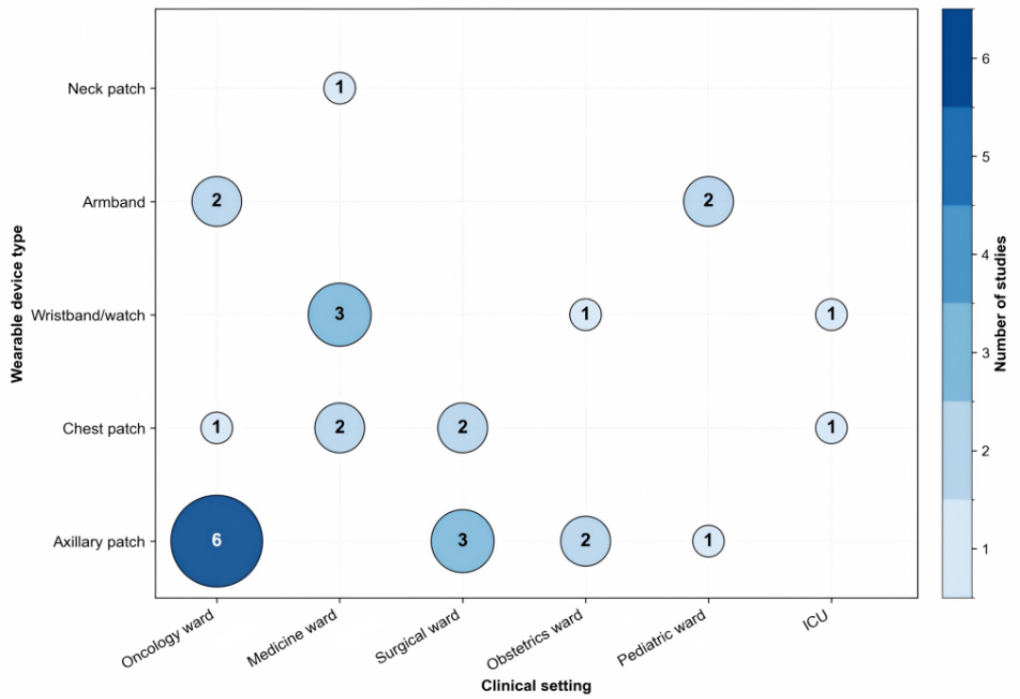
[h]N/A: not applicable.

[i]SpO_2: peripheral oxygen saturation.

[j]PMDA: Pharmaceuticals and Medical Devices Agency, Japan.

XSL•FO
RenderX

Figure 2. Number of included studies by clinical setting and wearable device type. Bubble size and color intensity represent study count for each combination. Studies using 2 wearable devices were counted under both types. ICU: intensive care unit.



There was variation in monitoring frequency and duration across studies. Body temperature was monitored at varying intervals, including every 1, 2, 4, 10, or 30 seconds; every 1, 2, 5, or 15 minutes; or at sampling frequencies of 4 Hz or 25 Hz. The median duration of temperature monitoring per patient ranged from 3 hours to 20 days across studies. Battery life across the included wearables ranged from 1 to 30 days. Holt et al [44] reported that, despite a theoretical maximum battery life of 3 days, the device recorded data only up to 44 hours in practice. To meet the requirements of clinical monitoring durations, different strategies were used depending on the device, such as daily charging [41,45,46], battery replacement [47], or regular device change every 24 hours [39,42,48], 3 days [49,50], or 6 days [51].

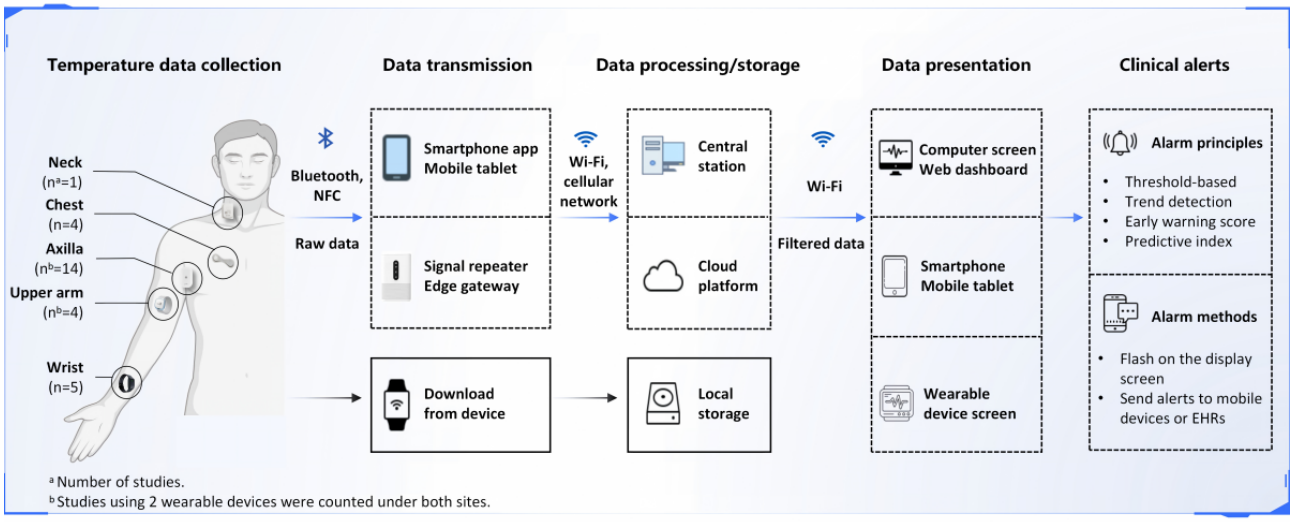
Temperature Data Transmission, Storage, and Presentation

Data Transmission

Figure 3 illustrates the data transmission strategies. Wearable data were predominantly transmitted through a wireless pipeline (20/26, 77%), initially from sensors via Bluetooth to nearby mobile devices (smartphones, tablets, or iPads), engineered bridges, signal repeaters, or transportable gateways. Subsequently, 16 studies described the further transmission of data from mobile devices or bridges to cloud service platforms (n=11) or central computers in the wards (n=5). These data were typically transmitted in real time or as continuous streams. Data latency was quantitatively reported in only 1 feasibility study [46], where data arrived at the dashboard with a median delay of 5 minutes for Everion and 2 minutes for CORE. After addressing transmission bottlenecks caused by 2 wearables competing on a single gateway, the median delay for both devices decreased to 0 minutes (range 0-2 minutes). Kim et al [52] used a dual-gateway wireless architecture that allowed the patch to communicate simultaneously with 2 nearby gateways, thereby ensuring uninterrupted transmission with minimal data latency or signal loss.

XSL•FO
RenderX

Figure 3. Data transmission, storage, and presentation in clinical practice.



In contrast, 19% (5/26) of studies used non–real-time data transmission methods, which included a batch upload strategy with 10-minute intervals [48], synchronizing temperature data each morning via a smartphone app following overnight monitoring [43], reading out patch data via near-field communication (NFC) every 24 to 72 hours using a smartphone app [53], or downloading data from the devices' internal memory after regular removal [44,45]. The remaining 1 study did not specify data transmission strategies or frequency.

Data Storage and Protection

Four studies reported that data could be stored temporarily on the device for 18 hours [54], 90 hours [45], or up to 6 days [55] with time stamps [47]. This feature supported local data retention during periods without network coverage, with buffered data automatically synchronizing upon connection restoration. Eight of the studies [41,46,50,54,56-59] reported that wearable data were stored on cloud platforms, with 2 studies [41,46] irreversibly deleting data from the cloud dashboard after they were transferred to a local research database. Four studies [47,52,55,60] stored wearable data on local central workstations or hospital computer systems. One study [45] reported that data were stored on the device, downloaded via a stationary Bluetooth interface for local storage, without using a web application due to regulatory requirements. Reporting on data security remains poor: only 38% (10/26) of studies described privacy protection strategies, including access restrictions [17,50,55,56], data anonymization [39], encryption for data in transit and data stored [41,46], local data storage [45], and the use of a secure wireless network and secure communication protocols [52]. Among these, 2 studies [45,49] reported alignment with the European Union (EU) General Data Protection Regulations.

Data Presentation in Clinical Practice

Only 27% (7/26) of studies presented temperature data from wearables in clinical settings, displaying data on mobile devices (n=5) and/or centralized monitoring interfaces (n=6). Among these, 4 studies [17,51,54,61] used multiple interfaces, displaying data both on mobile devices carried by nurses and on ward computer screens [17,61] or web-browser–based dashboards enabling remote access [51,54]. In contrast, 3 studies used a single interface, visualizing temperature curves either on a central computer screen in the nurse station [47,60] or on a paired bedside tablet [62]. Two studies reported that the wearable systems could send digital alerts to health care professionals via smartphones [17,61] or electronic health records (EHRs) [17]. Apart from the single study allowing alerts through EHRs [17], no other studies reported integration with EHR systems.

Conversely, 42% (11/26) of studies did not present wearable data in clinical settings, using the devices solely for research data collection. Among these, 9 studies explicitly stated that wearable data were blinded to clinical staff [41,45,48,50-53,58,59] to avoid influencing clinical decisions [46], while 2 studies [44,45] directly downloaded data from the device without dashboards. In 1 study [49], although heart rate, respiratory rate, and peripheral oxygen saturation (SpO_2) were available in real time for caregivers to support care, skin temperature was not directly available because its clinical relevance was considered uncertain. The remaining studies (8/26, 30.8%) lacked sufficient information to determine whether data were displayed or provided to clinical staff.

The Use of Temperature Data

Data Quality Assessment and Preprocessing

The reporting of missing sensor data varied across the studies. Only 62% (16/26) of studies reported the reasons and proportions of missing temperature data, while 2 studies indicated that no missing data were observed [47,60]. In the remaining 31% (8/26) of studies, missing data were not mentioned. The reported causes for missing data were grouped into 4 main categories: patient factors (eg, discomfort, nonadherence to the wearing protocol, and improper handling), device factors (eg, device malfunction and battery issues), wearing quality (eg, device detachment, poor skin contact, and sweating), and signal factors (eg, synchronization failure, motion artifacts, and patients moving out of signal coverage). The metrics used to quantify missing data also varied. Five studies described the hours with low-quality or missing data (ranging from 3.6% to 11.7%), 5 studies reported the hours with valid

XSL•FO
RenderX

data (ranging from 67% to 99%), while 6 studies reported the cases with missing values (ranging from 2% to 31.4%).

Several methodologies to detect invalid data were identified from the included studies, including heuristic filtering based on physiological plausibility thresholds, firmware-integrated quality indexing, and statistical outlier detection frameworks. Six studies used thresholds to filter out implausible values (eg, below 35.5 °C or 35 °C for axillary temperature [39,40,50,52] or 31 °C for skin temperature [44]), with one of them involving patients with terminal illness using a lower threshold of 25 °C for skin temperature [49]. Four studies [41,45,46,51] reported that wearables' firmware generated a quality score for temperature data per minute, though the use of the quality score differed. In the study by Garbern et al [54], a quality score was calculated based on signal quality each minute; windows with quality scores below 75% were labeled as low quality and excluded from analysis. In contrast, Jacobsen et al [45] did not use the quality score to select data; instead, they retained only periods with at least 3000 data points per hour for analysis. Eisenkraft et al [51] picked outliers with a difference from the mean exceeding the Q3 + 3 × IQR threshold. Additionally, the initial 5 minutes of data were often excluded due to device warm-up and stabilization periods [49].

Strategies for handling invalid or missing sensor data were heterogeneous across studies, spanning data filtering, smoothing, imputation, and native algorithmic handling. Some studies excluded entire cases [43,44,50,52,53,56] when extensive missingness occurred (eg, patch fell off over 6 hours), or targeted removal of data points identified as anomalies [17,44,50,51,53,54]. For smoothing, van Goor et al [49] computed 15-minute median values to reduce the impact of short-lasting outliers and missing data. For imputation, Rajeeve et al [59] applied forward-fill methods to address missing sensor data, whereas Debnath et al [40] retained the dropped data without imputation, considering the imputation process may introduce potential bias. Finally, 2 studies embedded missing-data handling directly within their modeling pipeline. Chen et al [56] used extreme gradient boosting (XGBoost), which handles missing values during training by automatically learning the optimal split direction; Jacobsen et al [45] used end-to-end deep learning algorithms that could handle artifacts and data gaps without extra preprocessing.

Temperature Features Derived From Wearable Data

Three major categories of temperature features were derived from raw continuous readings, including basic descriptive statistics, threshold-based features, and temporal features. Basic descriptive statistics were the most commonly used features, including the mean (n=14), median (n=3), maximum body temperature (n=5), SD (n=5), variance, and coefficient of variation (n=1). These features were used to characterize the temperature dataset and to compare temperature measurements across different monitoring methods, patient cohorts, and anatomical sites. For example, Liu et al [47] compared the peak temperature monitored by continuous monitoring and intermittent monitoring, finding that continuous monitoring captured 0.29 °C higher peak values.

The second major category was threshold-based temperature features, which were identified in 21 publications. These studies used various predefined cutoffs (eg, 37.2 °C, 37.8 °C, 38 °C, and 38.5 °C) to identify fever, subsequently using continuous temperature data to quantify the features of fever episodes, including frequencies, the cumulative duration [62] or percentage of time spent above the threshold [54], and severity (quantified by calculating the area under the time-temperature curve [62]). Notably, Rajeeve et al [59] reported that individualized axillary temperature thresholds (baseline mean + 2 SD) combined with a fixed threshold (36.4 °C) enabled earlier detection of temperature elevation before clinical recognition of CRS episodes than a simpler fixed 38.0 °C threshold did. Four studies [17,51,54,61] relied on thresholds set by early warning score (EWS) systems to trigger alerts when body temperature deviated from the preset physiological range. Eight studies focused on the timing of fever detection [6,39,42,47,50,52,58,60]. For instance, Sampson et al [42] compared the differences in the timing of fever detection, showing that continuous temperature monitoring detected fever events 12 hours earlier than standard monitoring. Wang et al [60] used 24-hour continuous recordings and a brute-force strategy to simulate different intermittent monitoring schedules, identifying the combination with the highest fever detection rate.

Ten publications further incorporated temporal features of body temperature, including trends, trajectory patterns, and circadian rhythmicity. Five studies explored visual trends and patterns: 3 studies observed rising temperature trends associated with infections [6,51,62]; 2 studies tried visual inspection of temperature curves in the 48 hours preceding fever episodes, although they did not yield clinically significant predictive information [41,46]. A more computational pattern recognition technique was used by Ren et al [37], who used dynamic time-warping to characterize temporal trajectory patterns to differentiate between fever etiologies. Another 3 studies focused on circadian rhythmicity, either by modeling the rhythm with cosinor models to detect a loss of circadian variation in skin temperature among patients who were critically ill with COVID-19 [49], or by using residual deviations from the patient's individual circadian rhythm pattern to predict febrile adverse events [6,45]. Two studies developed machine learning models using temperature data as inputs, with time-series features generated by aggregating temperature data into rolling windows of 6 to 14 hours in one study [59], while the other used a fixed 12-hour window [63].

Classification of Temperature Data Analytics

An established data analytics framework categorizes 4 analytic approaches by the questions they address: "What happened?" (descriptive), "Why did this happen?" (diagnostic), "What might happen in the future?" (predictive), and "What should be done next?" (prescriptive) [34]. Using this framework, we classified the temperature data analyses in the included publications as descriptive (used in 19/29, 65% of studies), diagnostic (used in 6/29, 21% of studies), and predictive (used in 9/29, 31% of studies); none used prescriptive analytics (Table 2). Five publications used multiple analytic approaches.

XSL•FO
RenderX

Table 2. Classification of data analytics for wearable-generated temperature data.

Types of analytics and description of the included studies	Temperature features	Analysis techniques
Descriptive analytics (n=19)		
Compared differences in temperatures recorded across monitoring methods, patient cohorts, or anatomical sites [43,44,47]	Mean, SD, median, and maximum	Student *t* test and ANOVA
Counted the numbers and frequencies of febrile events and compared the time difference between fever episodes detected by continuous and intermittent monitoring [17,38,39,42,47,48,50,52,58,61]	Timing of fever	Student *t* test
Developed a heuristic scoring system to quantify the clinical relevance of fever detection, incorporating fever duration (short spike <60 minutes; prolonged episode ≥60 minutes) and clinical actions triggered [53]	Thresholds and fever duration	Descriptive comparison
Hourly points from the continuous temperature data were used to simulate 1 to 24 intermittent daily measurements to identify the optimal timing and frequency achieving the highest fever detection rate [60]	Counts and frequencies	Brute-force strategy
Compared the performance of fixed, individualized (baseline mean + 2SD), and combined thresholds for axillary temperature in detecting cytokine release syndrome [59]	Thresholds	Grid search
Detected abnormal values or temperature deviations from preset thresholds to trigger alerts [17,61]	Thresholds	Described the counts and frequencies
Visualized patterns of temperature, such as rising trends or fluctuations in temperature curves [41,46,51]	Visual patterns	Line plots
Described the disappearance of the circadian rhythm of skin temperature in patients with critical illness, in comparison with recovered patients [49]	Circadian rhythmicity	Cosinor models
Diagnostic analytics (n=6)		
Classified fevers of infectious or non-infectious causes using temperature trajectory patterns within 4 hours around fever start time [37]	Trajectory patterns	Dynamic time warping and k-means clustering
Quantified fever intensity and evaluated its association with epidural analgesia duration and dosage to identify risk factors for fever [62]	Area under time-temperature curve	Logistic regression and ROC[a] curve analysis
Examined the correlation between skin temperature and oral/ear canal temperature and identified a diagnostic cut-off value of skin temperature for detecting fever or infections [44,48,57]	Correlations	ROC curve analysis
Analyzed the correlation between standard deviation and peak temperature to reveal an increased temperature variability before and after the spike [39]	Correlations	Pearson correlation and linear regression
Predictive analytics (n=9)		
Predicted SCC[b] by inputting temperature data into a deep neural network to calculate an SCC score reflecting dissimilarity from normal patterns [45]	Cosine similarity	Deep neural network
Used aggregated temperature features (median, variance, coefficient of variation) as inputs for pulmonary infection prediction models [56]	Median, variance, coefficient of variation	XGBoost[c]
Extracted residual deviations of temperature from the patient's individual circadian rhythm pattern as early signs of febrile adverse events [6]	Circadian residuals	Circadian modeling and genetic algorithm
Used continuous temperature data as inputs for machine learning models to predict the risk of advanced sepsis [54], cytokine release syndrome [59], or COVID-19 deterioration within the next 24 hours [63]	Mean, SD, extreme value, and percentage of time spent over thresholds	Machine learning models
Used skin temperature features (eg, mean and maximum) within various time windows before fever onset as early predictive signals for impending fever episodes [40,55,57]	Correlations	Logistic regression and generalized estimating equation

[a]ROC: receiver operating characteristic.

[b]SCC: serious clinical complications.

[c]XGBoost: extreme gradient boosting.

Descriptive analytics were used in 18 publications. Descriptive analytics describe the temperature data using basic descriptive statistics such as means, frequency, and variation, and present the data through histograms, box plots, scatter plots, and line charts. These studies reported that continuous temperature monitoring via wearables enabled earlier fever detection [17,38,39,42,47,48,50,58,61], captured higher temperatures [43,44,47], detected more febrile events [38,42], and yielded

XSL•FO
RenderX

higher fever detection rates [52,60] than conventional intermittent monitoring, suggesting the potential of wearables to improve fever or infection detection.

Diagnostic analytics were used in 6 publications. Diagnostic analytics involve comparing coexisting trends, uncovering correlations between variables, and determining causal relationships where possible to identify the risk factors or causes of fever. Pearson correlation, linear regression, and logistic regression were used to explore associations between temperature variables [39,57] and identify risk factors for fever [62]; receiver operating characteristic (ROC) analysis was used to derive diagnostic cutoff values for fever based on wearable skin temperature [44,48]; and k-means clustering was used to classify fevers of infectious or noninfectious causes [37]. These studies highlighted the potential of wearable data to support clinical diagnosis related to fever or infections.

Predictive analytics were identified in 9 publications. By using historical data and statistical modeling, predictive analytics identify patterns and predictors of future events and forecast the probability of fever or infections. Methods included logistic regression, machine learning models, and circadian modeling, with temperature features serving as predictive inputs. Six studies developed predictive models to estimate the risk of pulmonary infection [56], febrile adverse events [6,45], CRS [59], deterioration [63], and advanced sepsis [54]. Another 3 studies [40,55,57] focused on identifying skin temperature features within various time windows before fever onset as early predictive signals for impending fever episodes. These findings suggested the potential of wearables to enhance personalized fever or infection risk prediction.

Prescriptive analytics uses algorithms with “if” and “else” rules to analyze data and recommend optimal next steps or actions. In health care, this approach can be operationalized through clinical decision support systems that provide therapeutic recommendations, for example, whether medication or physical examinations are required. However, none of the included studies used prescriptive analytics.

Maturity Stages of Wearable Device Integration in Clinical Practice

Table 3 maps the original WHO stage of maturity for digital health interventions and our modified maturity framework tailored to wearables, outlining 6 phases in a continuum. As specified in the eligibility criteria, studies focused on prototype design, preclinical testing, or only on validating the accuracy of wearables were excluded. Among the 29 included publications, most were classified as being at the clinical validation stage (stage 3a: 9/29, 31%) or clinical research stage (stage 3b: 17/29, 59%). The primary objectives of these studies were to evaluate the performance of wearables for monitoring body temperature in clinical settings or to collect research data, corresponding to the pilot stage in the WHO framework. In 3 studies, wearables were applied in routine clinical practice (stage 4: 3/29, 10%) as a supplement to standard temperature monitoring by nurses, aligning with the demonstration stage. None of the studies in this review had progressed to multicenter implementation or full integration.

XSL•FO
RenderX

Table 3. Maturity stages of wearable devices' integration in clinical practice, modified from WHO[a] "Monitoring and Evaluating Digital Health Interventions."

WHO stage of maturity and modified stage of maturity	Publications	Focus on wearable devices at each stage
Early		
Preprototype		
Stage 1: proof of concept	Not included	Hypothesis building, clinical needs/context assessment, preliminary solution modeling, and initial technical feasibility testing
Prototype		
Stage 2: preclinical test	Not included	Technical design, functionality assessment, and testing of technical stability (eg, device compatibility, data transmission range, and system robustness) in simulated or controlled environments
Pilot		
Stage 3a: clinical validation	n=9	Validation of accuracy, precision, and feasibility in real clinical settings, typically via field tests and clinical trials using gold standards for comparison [39,41,42,46,47,50,52,58,61]
Stage 3b: clinical research	n=17	Wearables are used exclusively for data collection in research, with the data collected not being provided to support clinical practice [6,17,37,38,40,43-45,48,53-57,59,62,63]
Mid		
Demonstration		
Stage 4: routine clinical application	n=3	Wearables are integrated into routine clinical workflows as a supplement to standard care; wearable system operating independently from the hospital's EHR[b] [49,51,60]
Advanced		
Scale-up		
Stage 5: multicenter implementation	n=0	Wearables are adopted as part of routine clinical practice across multiple centers or scaled up to subnational, national, or population levels.
Integrated and sustained program		
Stage 6: full integration	n=0	Full integration into clinical practice, with EHR interoperability, use in decision support systems, and supportive policies and financing enabling broader health system incorporation.

[a]WHO: World Health Organization.

[b]EHR: electronic health record.

Nurses' Roles

The findings show that nurses were engaged in both research and clinical practice related to wearable devices. As researchers (identified in 7 studies), nurses acted as principal investigators who led and conducted the research [47,48,50,55,57,61] or served as clinical research nurses responsible for patient recruitment, device maintenance, and data collection [54,61]. In clinical practice, 12 studies described nurses as direct users of wearable devices for temperature monitoring, evaluating temperature data generated by the devices, and responding to device alerts.

Discussion

Principal Findings

This scoping review assessed 29 publications from 26 studies using wearables to monitor body temperature in acute care hospitals for fever or infection management. We identified a variety of wearables used across diverse patient populations and clinical settings since 2018, indicating growing interest in this emerging technology despite its recent introduction to clinical practice. By mapping the current evidence on wearable temperature data pipelines from sampling through transmission and presentation to clinical use, we also identified overall limited clinical integration and critical gaps at each stage. These findings may serve as a practical reference for the design and implementation of wearable-based monitoring systems within clinical workflows and highlight barriers to be addressed to move toward more sustainable clinical adoption.

Different Types of Wearables and Clinical Utility of Their Temperature Data

This review identified a preference for medical-grade devices in clinical settings, while consumer-grade devices predominated in home-based remote monitoring in oncology [20] or cardiovascular care [64]. This preference reflects the need for accurate temperature data in clinical practice, which influences

XSL•FO
RenderX

the selection of wearables and how the data can be used. Wearables generally provide temperature readings by monitoring thermal activity at the skin surface, which is susceptible to ambient conditions and skin perfusion [65]. In order to meet clinical and regulatory standards, medical-grade wearables adopted several strategies to deliver accurate temperature data, including monitoring at more reliable anatomical sites such as the axilla, which closely approximates core temperature when properly acquired [66], and developing algorithms to estimate core temperature from raw sensor signals. Accordingly, the average readings of estimated core temperatures and axillary temperatures were around 36.3 °C to 37.2 °C [40,62], which were mainly considered in studies that presented wearable data in clinical practice.

Unsurprisingly, skin temperatures directly provided by wearables at wearing sites, such as the wrist or chest, remain more variable in terms of validity and clinical utility. The mean skin temperatures were reported around 33.9 °C to 34.8 °C in reviewed studies [43,44], showing significant deviation from commonly used clinical values, particularly when determining fever. Therefore, several studies withheld skin temperature from clinical staff or discarded data from analyses [67-69], highlighting a cautious stance toward its clinical relevance. Some studies limited data collection periods, monitoring wrist skin temperature only during sleep to reduce daily activity influences [43]. However, studies have embraced more innovative strategies to actively use skin temperature, such as identifying elevated values as a predictive signal for fever [55], detecting abnormal circadian rhythms in patients with critical illness [49], and directly using its cutoff values to predict infection [44]. These methods view skin temperature as an independent output pathway of the thermoregulatory system, offering a potentially clinically useful route for its application. Notably, 1 study reported that both axillary and skin temperature emerged as reliable biomarkers, yet prioritized wearable-derived skin temperature over axillary measurements for its higher sampling frequency, continuous availability, and ease of future deployment [59]. Therefore, for clinical interpretation of wearable temperature data, health care providers should not be limited by absolute value comparability to traditional measurements but integrate trend-based and pattern-recognition approaches to expand clinical utility of wearable data across diverse monitoring contexts.

Filtering Anomalies Before Using Temperature Data From Wearables

Although medical-grade wearables exhibit accuracy, prolonged usage in real-world clinical environments still presents significant data quality challenges arising from patient nonadherence, poor skin contact, device malfunction, sensor drift, and signal artifacts. Efforts have been made by manufacturers to assign quality scores to each temperature reading via wearables' firmware [41,45,46,51]. However, the reliability of quality scores has not been adequately validated, and their calculations often lack transparency due to proprietary algorithms and patent protection. As indicated by Koenig et al [41], the wearables' quality assessment was generally plausible, but it occasionally misclassified vital signs by assigning good quality to incorrectly measured data and poor quality to plausible extreme values. Therefore, studies in this review differed in their use of quality scores. We highly recommend that future studies validate quality scores before deciding whether to incorporate them.

Identifying anomalies should be guided by both data characteristics and clinical domain knowledge, as anomalies may still fall within the physiological boundaries yet represent clinically significant events [15]. The most commonly used method in the reviewed studies relied on thresholds to identify invalid temperature data, which has significant limitations. Removing axillary temperature readings below 35.5 °C may overlook inadvertent hypothermia in patients with critical illness [70] or patients who had undergone surgery [66]. Moreover, recent studies suggest that human basal body temperature has declined over the past century [71]. A systematic review calculated the normal range for axillary temperature to be 35.01 °C to 36.93 °C [72], indicating that a fixed 35.5 °C may not be able to account for individual variations in baseline temperatures. Additionally, the high variability in skin temperature makes it more challenging to determine reasonable thresholds for outlier removal. Instead, specialized artifact detection algorithms have been developed in a broader context. For example, Bardia et al [73] developed algorithms to identify adjacent temperature readings with slopes exceeding 0.08 or an absolute change greater than 0.25 °C, which effectively filtered intraoperative temperature artifacts. Future efforts should develop and incorporate such standardized data cleaning frameworks based on specific clinical contexts and professional knowledge, enabling personalized anomaly detection while preserving clinically relevant events.

After identifying anomalies in temperature data, handling incomplete datasets remains challenging. Most reviewed studies simply excluded cases with high missingness or deleted anomalous window segments. While retaining only valid data is feasible for basic statistical comparisons, removing data disrupts time-series signal continuity and limits specific feature extractions that require continuous data as inputs, such as entropy metrics for nonlinear dynamics [13]. On the other hand, the decision to impute missing data also requires careful evaluation. A study comparing imputation techniques for wireless continuous vital signs found that linear interpolation reduced feature bias (eg, 2-hour window mean and slope) relative to no imputation, whereas other techniques (eg, last observation carried forward and spline interpolation) increased bias, with imputation errors growing significantly with gap length [74]. Some advanced strategies have emerged as alternatives to imputation. For instance, dimensionality and size reduction based on intrinsic and target similarity can be used to eliminate features and samples that are highly sensitive to missing values [75], a principle that may extend to physiological time-series modeling. In addition, algorithms with built-in tolerance to missing data may improve model robustness and generalizability. Overall, we recommend selecting the appropriate preprocessing approach based on the specific downstream task, following careful assessment of potential bias.

XSL•FO
RenderX

Advancing Temporal Pattern Mining and Interpretation for Wearables Data

Our review found that descriptive analytics remain the predominant approach for analyzing wearable-generated temperature data, with visual inspection of the trends and spikes in temperature curves being a common approach. This approach is intuitive, convenient, and easy to interpret in clinical practice, without requiring additional computational resources. However, it may not allow for detecting a specific pattern or predictive signals preceding infections [41]. In contrast, time series analysis and machine learning models for diagnostic or predictive applications were used in only one-third of the publications reviewed, indicating underutilization of these advanced analytical capabilities.

To leverage the hidden temporal patterns in high-frequency temperature data, machine learning and time series models represent an important methodological evolution. These methods extract temperature features that indicate circadian rhythm variations and pattern similarities, capturing subtle yet crucial indicators of infection that are not discernible through direct visual inspection. Beyond the temperature features identified in this review, studies have examined temperature dynamics in a broader context. For example, Papaioannou et al [76] assessed temperature complexity using wavelet transformation and multiscale entropy to classify patients with systemic inflammatory response syndrome, sepsis, and septic shock. Bhavani et al [77] identified 4 temperature trajectories from hourly recorded EHR data, revealing sepsis subphenotypes with distinct biomarker profiles. These advances highlight the need for further integration of advanced data analytics for temperature temporal pattern mining and knowledge discovery. These algorithms offer promising solutions to tackle data overload challenges by filtering meaningful and actionable information from massive wearable data; however, they still require further integration into computerized decision support systems to enable intelligent analysis and transform data into simpler and more usable forms.

Integration of Wearable Devices Into Clinical Practice

We found that several wearables have been used for temperature monitoring as a supplement to standard nursing measurements, but their integration into routine clinical practice remains limited. The wearable systems were deployed independently, with only 1 study enabling alerts through EHRs [17]. This may stem from health systems' requirement to establish a dedicated secure network for wearables, separated from the main network to prioritize data privacy [78]. However, this approach creates data silos. Although temperature is a key indicator for infection detection, final clinical diagnosis or treatment decisions require synthesis of wearable data with laboratory findings, clinical assessments, and patient history. Advancing system interoperability is essential to support comprehensive clinical decision-making rather than switching between systems to access fragmented data. Potential solutions include establishing standardized APIs through third-party integration platforms and developing plug-and-play interoperability standards [78]. Additionally, to streamline nursing workflow, more efforts are still needed to achieve automatic uploading of wearable data to nursing records, rather than requiring manual transcription of data from one screen to another. However, as previously discussed, such system integration must incorporate automatic artifact filtering algorithms to prevent recording erroneous data without professional oversight.

To better support clinical practice using wearables, we identified several gaps requiring further attention regarding data access interfaces, data latency, and data security. First, some studies deployed central monitoring systems at the nursing station, which enable simultaneous multipatient monitoring but prevent immediate access to real-time readings when assessing patients at bedside. A hybrid approach combining central and mobile interfaces would optimize both oversight and point-of-care accessibility. Second, some studies used batch uploads at fixed intervals, which conserve energy and extend battery life but may delay clinical reactions to patient deterioration. Transmission frequency must align with clinical context to ensure real-time data streaming, with transparent reporting enabling proper data latency assessment. Third, data security is a growing concern, yet only 10 studies reported specific protection strategies, with 50% (5/10) conducted in Europe under strict EU privacy regulations [20]. There remains a need for stronger adherence to safety and privacy regulations globally.

Overall, the clinical integration of wearables for fever and infection management remains predominantly at an early stage of maturity, aligning with existing evidence from digital health interventions for postoperative monitoring [79] and oncology care [20]. Although the scope of regulatory approval limited some devices' clinical application, the reasons more frequently reported for not providing data to support clinical decision-making were insufficient data accuracy and limited real-time availability [53]. We also noted that certain devices (eg, SensiumVitals, Everion, and Gen2, listed in Multimedia Appendix 3) are not intended for use in critical care patients, potentially associated with reduced measurement reliability under unstable hemodynamic conditions. This explains their limited use in ICUs (Figure 2) and underscores the need for caution when using wearables for patients who are clinically deteriorating in general wards and avoiding sole reliance on wearable data for clinical decisions. Beyond technical limitations, there are systemic barriers to overcome before wearables can become widely accepted with full clinical integration, such as data ownership and privacy, data standards and interoperability, and reimbursement policies [80]. Additionally, the geographic concentration of reviewed studies in well-resourced areas highlights potential equity challenges, as disparities in device access, digital literacy, and infrastructure may limit scaled implementation and exacerbate health inequalities. Nevertheless, the emergence of wearables in clinical settings shows a developmental trajectory outlined in the maturity framework, which delineates the accumulation of evidence throughout the device lifecycle from initial prototype to full clinical integration.

Future Directions

With the rapid iteration of wearable technology, novel devices are increasingly emerging. While existing evidence mainly focused on lightweight rigid wearables, we identified some

XSL•FO
RenderX

textile-based flexible wearables for remote monitoring of infectious diseases in the literature search. However, these textile-based devices are still in the proof-of-concept [81], technical testing [82], and animal experimental stages [83], possibly reflecting the current technological bottlenecks in achieving clinical accuracy and limited maturity in regulatory approval and clinical validation [84]. This development reflects an evolution in wearable technologies toward softer, lighter, and safer sensing materials. In future studies, our maturity framework could serve as a tool to support health care providers in evaluating novel wearables' readiness for clinical deployment and selecting suitable devices. Noncontact vital sign monitoring technologies, such as camera-based or robotic sensing systems, represent another emerging direction in continuous physiological surveillance in hospital environments [85] and may offer supplementary options for body temperature monitoring in special populations, such as neonates with fragile skin intolerant to adhesive sensors [86], although these technologies fall outside the scope of this review.

As wearables become increasingly used in clinical practice, nurses need to be better prepared for this transformation. Nurses are not only frontline users but also key researchers in this field, with nearly one-third of the studies reviewed being nurse-led. In future clinical deployment, large-scale data streams from wearables will require a robust infrastructure and storage capacity to prevent data loss or overwriting. These data could be incorporated into the clinical research data warehouse as a valuable source. Moreover, to harness the value of big data and AI, it is important for nurses to develop digital literacy and data science competencies to integrate nursing expertise into data interpretation. This will empower nurses to play an active role in wearable-driven health care transformation, rather than being passive end users of monitoring technology.

The observed lack of transparency highlights the need for reporting guidelines specific to clinical research using wearables. Our findings indicate low reporting rates on fundamental aspects, such as device grade (12/29, 41%) and regulatory status (10/29, 34%). While wearable data quality is crucial for appropriate clinical interpretation [87], many publications did not report data accuracy (11/29, 38%) and missing data (8/26, 31%). To improve rigor, the device grade, regulatory status, and clinical validation should be standard reporting elements. For feasibility and pilot studies, data transmission, latency, and loss rates could be taken into assessment and reporting. In predictive modeling research, reporting missing sensor data and any imputation methods is crucial. Data privacy and clinical accessibility could also be incorporated into reporting guidelines.

Limitations of This Review

This review has several limitations. First, we focused specifically on the use of wearables for monitoring body temperature in fever or infection management. The assessment of maturity is limited to this specific area, and maturity stages may differ in other fields. Second, we did not assess the quality or risk of bias of the included studies, which means our findings are based on all available evidence, regardless of its methodological rigor. Third, we only included studies published in English or Chinese, which may have excluded evidence published in other languages. Additionally, only a random 20% of the records were screened and extracted independently by 2 reviewers. While this approach is methodologically acceptable [88], it remains a potential source of bias. Furthermore, we identified the roles of nurses from the main texts, author, and affiliation information. However, confirming the presence of nurses among the authors was not always possible due to the lack of explicitly stated nursing qualifications.

Conclusions

This scoping review summarized ways of using temperature data from wearables across diverse clinical contexts, showing promise in optimizing fever or infection management in acute care hospitals, particularly when integrated with advanced data analytic approaches. However, the overall integration of wearables into clinical practice remains at an early stage of maturity. Key issues include data quality challenges, clinical interpretation of skin surface temperature, and interoperability between wearable systems and EHRs. Future efforts should prioritize standardized data cleaning frameworks and EHR interoperability, and facilitate clinical interpretation of wearable data to promote its active use in decision-making. These efforts will help to unlock the full potential of wearables in driving health care transformation.

Acknowledgments

The authors thank librarian Ms Claire Hodkinson at The University of Manchester Library for reviewing the search strategy. We also thank Mr Xinyu Dong for the valuable suggestions during the stakeholder consultation.

During the preparation of this work, the authors used ChatGPT 4.0 (OpenAI) to proofread the manuscript for grammatical errors. After using this tool and service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

XSL•FO
RenderX

Funding

This work was funded by the Sichuan Provincial Science and Technology Program (2024NSFSC1688, 2023YFS0066) and the China Scholarship Council (CSC) to YL. The funders had no role in the study design, data collection and analysis, preparation of the manuscript, or decision to publish.

Authors' Contributions

Conceptualization: YL (lead), DD (supporting)
Data curation: YL (lead), QZ (supporting)
Formal analysis: YL (lead), QZ (supporting), DD (supporting)
Funding acquisition: YL
Investigation: YL (lead), QZ (supporting)
Methodology: YL (lead), DD (supporting)
Supervision: DD (lead), KL (supporting)
Validation: YL (lead), QZ (supporting), DD (supporting)
Visualization: YL
Writing—original draft: YL
Writing—review and editing: YL (lead), QZ (supporting), KL (supporting), DD (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR checklist.
[DOCX File , 319 KB - mhealth_v14i1e89773_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 156 KB - mhealth_v14i1e89773_app2.docx]

Multimedia Appendix 3
Data extraction form.
[XLSX File (Microsoft Excel File), 4100 KB - mhealth_v14i1e89773_app3.xlsx]

References

1. Global report on infection prevention and control. 1st edition. World Health Organization. 2022. URL: https://www.who.int/publications/i/item/9789240051164 [accessed 2025-09-30]
2. Suetens C, Latour K, Kärki T, Ricchizzi E, Kinross P, Moro ML, Healthcare-Associated Infections Prevalence Study Group. Prevalence of healthcare-associated infections, estimated incidence and composite antimicrobial resistance index in acute care hospitals and long-term care facilities: results from two European point prevalence surveys, 2016 to 2017. Euro Surveill 2018;23(46):1800516 [FREE Full text] [doi: 10.2807/1560-7917.ES.2018.23.46.1800516] [Medline: 30458912]
3. Ghashghaee A, Benis MR, Aryankhesal A, Tanha K, Hosseinifard H, Janani L, et al. The prevalence of hospital-acquired infections in the EMRO: a systematic review and meta-analysis from 2000 to 2018. SSRN Electron J 2019:1-14. [doi: 10.2139/ssrn.3339500]
4. Leung LY, Huang HL, Hung KK, Leung CY, Lam CC, Lo RS, et al. Door-to-antibiotic time and mortality in patients with sepsis: systematic review and meta-analysis. Eur J Intern Med 2024;129:48-61. [doi: 10.1016/j.ejim.2024.06.015] [Medline: 39034174]
5. Mackowiak PA, Chervenak FA, Grünebaum A. Defining fever. Open Forum Infect Dis 2021;8(6):ofab161 [FREE Full text] [doi: 10.1093/ofid/ofab161] [Medline: 34476283]
6. Flora C, Tyler J, Mayer C, Warner DE, Khan SN, Gupta V, et al. High-frequency temperature monitoring for early detection of febrile adverse events in patients with cancer. Cancer Cell 2021;39(9):1167-1168 [FREE Full text] [doi: 10.1016/j.ccell.2021.07.019] [Medline: 34388378]
7. Wells CI, Xu W, Penfold JA, Keane C, Gharibans AA, Bissett IP, et al. Wearable devices to monitor recovery after abdominal surgery: scoping review. BJS Open 2022;6(2):zrac031 [FREE Full text] [doi: 10.1093/bjsopen/zrac031] [Medline: 35388891]
8. Lu L, Zhang J, Xie Y, Gao F, Xu S, Wu X, et al. Wearable health devices in health care: narrative systematic review. JMIR Mhealth Uhealth 2020;8(11):e18907 [FREE Full text] [doi: 10.2196/18907] [Medline: 33164904]

XSL•FO
RenderX

9. Leenen JPL, Leerentveld C, van Dijk JD, van Westreenen HL, Schoonhoven L, Patijn GA. Current evidence for continuous vital signs monitoring by wearable wireless devices in hospitalized adults: systematic review. J Med Internet Res 2020;22(6):e18636 [FREE Full text] [doi: 10.2196/18636] [Medline: 32469323]
10. Papaioannou VE, Sertaridou EN, Chouvarda IG, Kolios GC, Pneumatikos IN. Determining rhythmicity and determinism of temperature curves in septic and non-septic critically ill patients through chronobiological and recurrence quantification analysis: a pilot study. Intensive Care Med Exp 2019;7(1):53 [FREE Full text] [doi: 10.1186/s40635-019-0267-9] [Medline: 31486940]
11. Yehya N, Fitzgerald JC, Hayes K, Zhang D, Bush J, Koterba N, et al. Temperature trajectory sub-phenotypes and the immuno-inflammatory response in pediatric sepsis. Shock 2022;57(5):645-651 [FREE Full text] [doi: 10.1097/SHK.0000000000001906] [Medline: 35066512]
12. Verma N, Haji-Abolhassani I, Ganesh S, Vera-Aguilera J, Paludo J, Heitz R, et al. A Novel Wearable Device for Continuous Temperature Monitoring & Fever Detection. IEEE J Transl Eng Health Med 2021;9:2700407. [doi: 10.1109/JTEHM.2021.3098127] [Medline: 34765323]
13. Vargas B, Cuesta-Frau D, González-López P, Fernández-Cotarelo MJ, Vázquez-Gómez Ó, Colás A, et al. Discriminating bacterial infection from other causes of fever using body temperature entropy analysis. Entropy (Basel) 2022;24(4):510 [FREE Full text] [doi: 10.3390/e24040510] [Medline: 35455174]
14. Odone A, Barbati C, Amadasi S, Schultz T, Resnik DB. Artificial intelligence and infectious diseases: an evidence-driven conceptual framework for research, public health, and clinical practice. Lancet Infect Dis 2026;26(3):e152-e167. [doi: 10.1016/S1473-3099(25)00412-8] [Medline: 40972628]
15. Ortiz BL, Gupta V, Kumar R, Jalin A, Cao X, Ziegenbein C, et al. Data preprocessing techniques for AI and machine learning readiness: scoping review of wearable sensor data in cancer care. JMIR Mhealth Uhealth 2024;12:e59587 [FREE Full text] [doi: 10.2196/59587] [Medline: 38626290]
16. Leenen JPL, Rasing HJM, Kalkman CJ, Schoonhoven L, Patijn GA. Process evaluation of a wireless wearable continuous vital signs monitoring intervention in 2 general hospital wards: mixed methods study. JMIR Nurs 2023;6:e44061 [FREE Full text] [doi: 10.2196/44061] [Medline: 37140977]
17. Joshi M, Ashrafian H, Arora S, Sharabiani M, McAndrew K, Khan SN, et al. A pilot study to investigate real-time digital alerting from wearable sensors in surgical patients. Pilot Feasibility Stud 2022;8(1):140 [FREE Full text] [doi: 10.1186/s40814-022-01084-2] [Medline: 35794669]
18. Khanna AK, Flick M, Saugel B. Continuous vital sign monitoring of patients recovering from surgery on general wards: a narrative review. Br J Anaesth 2025;134(2):501-509 [FREE Full text] [doi: 10.1016/j.bja.2024.10.045] [Medline: 39779421]
19. Guardado S, Karampela M, Isomursu M, Grundstrom C. Use of patient-generated health data from consumer-grade devices by health care professionals in the clinic: systematic review. J Med Internet Res 2024;26:e49320 [FREE Full text] [doi: 10.2196/49320] [Medline: 38820580]
20. Cloß K, Verket M, Müller-Wieland D, Marx N, Schuett K, Jost E, et al. Application of wearables for remote monitoring of oncology patients: a scoping review. Digit Health 2024;10:20552076241233998 [FREE Full text] [doi: 10.1177/20552076241233998] [Medline: 38481796]
21. Amin T, Mobbs RJ, Mostafa N, Sy LW, Choy WJ. Wearable devices for patient monitoring in the early postoperative period: a literature review. Mhealth 2021;7:50 [FREE Full text] [doi: 10.21037/mhealth-20-131] [Medline: 34345627]
22. Peters MD, Godfrey C, McInerney P, Munn Z, Tricco A, Khalil H. Scoping reviews. In: Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z, editors. JBI Manual for Evidence Synthesis. Adelaide, Australia: JBI; 2024.
23. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
24. Liu Y, Zhao Q, Li K, Dowding D. The use of continuous temperature data from wearable devices to identify fever or infections in acute hospital settings: a scoping review. Open Science Framework. 2025. URL: https://osf.io/v6sp8/overview [accessed 2025-12-18]
25. Liu Y, Zhao Q, Li K, Dowding D. Use of continuous temperature data from wearable devices for fever or infection management in acute hospital settings: a scoping review protocol. BMJ Open 2025;15(12):e103630 [FREE Full text] [doi: 10.1136/bmjopen-2025-103630] [Medline: 41469073]
26. Thompson HJ. Fever: a concept analysis. J Adv Nurs 2005;51(5):484-492 [FREE Full text] [doi: 10.1111/j.1365-2648.2005.03520.x] [Medline: 16098165]
27. Purssell E. Fever in children--a concept analysis. J Clin Nurs 2014;23(23-24):3575-3582. [doi: 10.1111/jocn.12347] [Medline: 24103112]
28. Brasier N, Wang J, Dincer C, Güder F, Schauwecker I, Schaffarczyk D, et al. Next-generation digital biomarkers: continuous molecular health monitoring using wearable devices. Trends Biotechnol 2024;42(3):255-257. [doi: 10.1016/j.tibtech.2023.12.001] [Medline: 38233230]
29. Kageyama I, Kurata K, Miyashita S, Lim Y, Sengoku S, Kodama K. A bibliometric analysis of wearable device research trends 2001-2022—a study on the reversal of number of publications and research trends in China and the USA. Int J Environ Res Public Health 2022;19(24):16427 [FREE Full text] [doi: 10.3390/ijerph192416427] [Medline: 36554307]

XSL•FO
RenderX

30. Zhang N, Peng Y, Guo Q. Visual analysis of research trends and hotspots in wearable electronic devices in the medical field: a bibliometric study. Digit Health 2024;10:20552076241305233 [FREE Full text] [doi: 10.1177/20552076241305233] [Medline: 39679002]
31. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS peer review of electronic search strategies: 2015 guideline statement. J Clin Epidemiol 2016;75:40-46 [FREE Full text] [doi: 10.1016/j.jclinepi.2016.01.021] [Medline: 27005575]
32. AI-powered systematic review management platform. Rayyan. URL: https://www.rayyan.ai/ [accessed 2025-12-18]
33. Paganelli AI, Mondéjar AG, da Silva AC, Silva-Calpa G, Teixeira MF, Carvalho F, et al. Real-time data analysis in health monitoring systems: a comprehensive systematic literature review. J Biomed Inform 2022;127:104009 [FREE Full text] [doi: 10.1016/j.jbi.2022.104009] [Medline: 35196579]
34. Parrales-Bravo F, Cevallos-Torres L, Vasquez-Cevallos L, Caicedo-Quiroz R, Tolozano-Benites R, Gómez-Rodríguez V. A review of the use of data analytics to address preeclampsia in Ecuador between 2020 and 2024. Diagnostics (Basel) 2025;15(8):978 [FREE Full text] [doi: 10.3390/diagnostics15080978] [Medline: 40310379]
35. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. World Health Organization. 2016. URL: https://iris.who.int/handle/10665/252183 [accessed 2025-02-28]
36. Peters MDJ, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
37. Ren X, Mayhew K, Rozwadowski M, Khan S, Rao A, Kumar R, et al. Distinguishing infection- vs non-infection-related febrile neutropenia in HCT patients by machine learning analysis of continuous body temperature data collected using a wearable sensor. Transplant Cell Ther 2024;30(2):S119-S120. [doi: 10.1016/j.jtct.2024.01.006]
38. Khan SN, Tewari M, Lee S, Flora C, Madhukar R, Mayhew K, et al. High-frequency temperature monitoring detects clinically actionable fevers otherwise missed by standard-of-care monitoring during treatment of hematologic malignancies. Blood 2024;144(1):7500. [doi: 10.1182/blood-2024-194678]
39. Koppel R, Debnath S, Zanos TP, Saadi N, Potak D, Meirowitz N, et al. Efficacy of continuous monitoring of maternal temperature during labor using wireless axillary sensors. J Clin Monit Comput 2022;36(1):103-107. [doi: 10.1007/s10877-020-00625-5] [Medline: 33394269]
40. Debnath S, Koppel R, Saadi N, Potak D, Weinberger B, Zanos TP. Prediction of intrapartum fever using continuously monitored vital signs and heart rate variability. Digit Health 2023;9:20552076231187594 [FREE Full text] [doi: 10.1177/20552076231187594] [Medline: 37448783]
41. Koenig C, Ammann RA, Kuehni CE, Roessler J, Brack E. Continuous recording of vital signs with a wearable device in pediatric patients undergoing chemotherapy for cancer-an operational feasibility study. Support Care Cancer 2021;29(9):5283-5292 [FREE Full text] [doi: 10.1007/s00520-021-06099-8] [Medline: 33655413]
42. Sampson M, Hickey V, Huber J, Alonso PB, Davies SM, Dandoy CE. Feasibility of continuous temperature monitoring in pediatric immunocompromised patients: a pilot study. Pediatr Blood Cancer 2019;66(6):e27723. [doi: 10.1002/pbc.27723] [Medline: 30884117]
43. Brun R, Girsberger J, Rothenbühler M, Argyle C, Hutmacher J, Haslinger C, et al. Wearable sensors for prediction of intraamniotic infection in women with preterm premature rupture of membranes: a prospective proof of principle study. Arch Gynecol Obstet 2023;308(5):1447-1456 [FREE Full text] [doi: 10.1007/s00404-022-06753-4] [Medline: 36098832]
44. Holt SG, Yo JH, Karschimkus C, Volpato F, Christov S, Smith ER, et al. Monitoring skin temperature at the wrist in hospitalised patients may assist in the detection of infection. Intern Med J 2020;50(6):685-690 [FREE Full text] [doi: 10.1111/imj.14748] [Medline: 31908128]
45. Jacobsen M, Gholamipoor R, Dembek TA, Rottmann P, Verket M, Brandts J, et al. Wearable based monitoring and self-supervised contrastive learning detect clinical complications during treatment of Hematologic malignancies. NPJ Digit Med 2023;6(1):105 [FREE Full text] [doi: 10.1038/s41746-023-00847-2] [Medline: 37268734]
46. Koenig C, Ammann RA, Schneider C, Wyss J, Roessler J, Brack E. Continuous timely monitoring of core temperature with two wearable devices in pediatric patients undergoing chemotherapy for cancer - a comparison study. Support Care Cancer 2024;32(3):188 [FREE Full text] [doi: 10.1007/s00520-024-08366-w] [Medline: 38400942]
47. Liu Y, Liu C, Gao M, Wang Y, Bai Y, Xu R, et al. Evaluation of a wearable wireless device with artificial intelligence, iThermonitor WT705, for continuous temperature monitoring for patients in surgical wards: a prospective comparative study. BMJ Open 2020;10(11):e039474 [FREE Full text] [doi: 10.1136/bmjopen-2020-039474] [Medline: 33208327]
48. Dambrosio N, Porter M, Bauer E, Levitan N, Liedtke D, De LM, et al. Identifying neutropenic fever earlier: an application of a skin patch for continuous temperature monitoring. Blood 2018;132(1):4718. [doi: 10.1182/blood-2018-99-114181]
49. van Goor HMR, van Loon K, Breteler MJM, Kalkman CJ, Kaasjager KAH. Circadian patterns of heart rate, respiratory rate and skin temperature in hospitalized COVID-19 patients. PLoS One 2022;17(7):e0268065 [FREE Full text] [doi: 10.1371/journal.pone.0268065] [Medline: 35797369]
50. Liu S, Nataraja RM, Medhurst A, Pacilli M. Continuous wireless temperature monitoring using TempTraq® in children. Pediatr Res 2026;99(7):2680-2686. [doi: 10.1038/s41390-025-04553-5] [Medline: 41233627]

XSL•FO
RenderX

51. Eisenkraft A, Maor Y, Constantini K, Goldstein N, Nachman D, Levy R, et al. Continuous remote patient monitoring shows early cardiovascular changes in COVID-19 patients. J Clin Med 2021;10(18):4218 [FREE Full text] [doi: 10.3390/jcm10184218] [Medline: 34575328]
52. Kim YK, Ahn SY, Lee SJ, Song CB, Hong Y, Kim R, et al. Novel wearable-based real-time temperature monitoring in hospitals for febrile adverse events in patients with cancer: a prospective feasibility study. Sensors (Basel) 2025;25(23):7111 [FREE Full text] [doi: 10.3390/s25237111] [Medline: 41374486]
53. Schäfer L, Dickel F, Strohmayer K, Koele W, Leber B, Sucher R, et al. Early detection of postoperative infections using continuous temperature monitoring: a prospective clinical trial. J Clin Monit Comput 2026;40(3):721-733. [doi: 10.1007/s10877-025-01383-y] [Medline: 41348412]
54. Garbern SC, Mamun GMS, Shaima SN, Hakim N, Wegerich S, Alla S, et al. A novel digital health approach to improving global pediatric sepsis care in Bangladesh using wearable technology and machine learning. PLOS Digit Health 2024;3(10):e0000634 [FREE Full text] [doi: 10.1371/journal.pdig.0000634] [Medline: 39475844]
55. Chung YT, Yeh CY, Chen CC, Lai CH, Lin YH, Lin CY, et al. Early changes in skin surface temperature predict body temperature increases in patients with fever: a pilot study. Intensive Crit Care Nurs 2024;83:103652. [doi: 10.1016/j.iccn.2024.103652] [Medline: 38402650]
56. Chen Y, She D, Guo Y, Chen W, Li J, Li D, et al. Smartwatch-based algorithm for early detection of pulmonary infection: validation and performance evaluation. Digit Health 2024;10:20552076241290684. [doi: 10.1177/20552076241290684] [Medline: 39465220]
57. Chen YJ, Chung YT, Chen CC, Chen YC, Liang SH, Wang JL, et al. [Detection of fever using continuous skin temperature monitoring in hospitalized older adults]. Hu Li Za Zhi 2023;70(1):60-69. [doi: 10.6224/JN.202302_70(1).09] [Medline: 36647311]
58. Fyfe RC, Larsen M, Blud D, Weinkove R. Continuous temperature monitoring for the detection of fever in haemato-oncology patients: a pilot study of two wearable devices (THERMAL). Intern Med J 2025;55(11):1928-1933. [doi: 10.1111/imj.70199] [Medline: 41001994]
59. Rajeeve S, Wilkes M, Zahradka N, Tomalin L, Quidwai M, Pan D, et al. Detection of cytokine release syndrome using wearables and cytokine profiling following CAR-T therapy for myeloma. JCI Insight 2026;11(12):e203988. [doi: 10.1172/jci.insight.203988] [Medline: 42084945]
60. Wang S, Ji G, Feng X, Huang L, Luo J, Yu P, et al. Temperature measurement timings and the fever detection rate after gastrointestinal surgery: retrospective cross-sectional study. Interact J Med Res 2024;13:e50585 [FREE Full text] [doi: 10.2196/50585] [Medline: 39383527]
61. Downey CL, Croft J, Ainsworth G, Buckley H, Shinkins B, Randell R, et al. Trial of remote continuous versus intermittent NEWS monitoring after major surgery (TRaCINg): a feasibility randomised controlled trial. Pilot Feasibility Stud 2020;6(1):183 [FREE Full text] [doi: 10.1186/s40814-020-00709-8] [Medline: 33292669]
62. Zhao B, Li B, Wang Q, Jia J, Song X. Time- and dose-dependent correlations between patient-controlled epidural analgesia and intrapartum maternal fever. BMC Anesthesiol 2021;21(1):31 [FREE Full text] [doi: 10.1186/s12871-021-01249-1] [Medline: 33514322]
63. Miyagawa K, Matsumoto S, Sasaki M, Osaki T, Furukubo M, Tsuchimoto H, et al. Machine learning-based prediction model of COVID-19 deterioration using wearable core temperature and oxygen saturation sensor data. Circulation 2022;146(Suppl_1):A9842. [doi: 10.1161/circ.146.suppl_1.9842]
64. Hughes A, Shandhi MMH, Master H, Dunn J, Brittain E. Wearable devices in cardiovascular medicine. Circ Res 2023;132(5):652-670 [FREE Full text] [doi: 10.1161/CIRCRESAHA.122.322389] [Medline: 36862812]
65. Böttcher S, Vieluf S, Bruno E, Joseph B, Epitashvili N, Biondi A, et al. Data quality evaluation in wearable monitoring. Sci Rep 2022;12(1):21412 [FREE Full text] [doi: 10.1038/s41598-022-25949-x] [Medline: 36496546]
66. Sessler DI. Perioperative temperature monitoring. Anesthesiology 2021;134(1):111-118. [doi: 10.1097/ALN.0000000000003481] [Medline: 32773677]
67. Choi A, Chung K, Chung SP, Lee K, Hyun H, Kim JH. Advantage of vital sign monitoring using a wireless wearable device for predicting septic shock in febrile patients in the emergency department: a machine learning-based analysis. Sensors (Basel) 2022;22(18):7054 [FREE Full text] [doi: 10.3390/s22187054] [Medline: 36146403]
68. Garbern SC, Mbanjumucyo G, Umuhoza C, Sharma VK, Mackey J, Tang O, et al. Validation of a wearable biosensor device for vital sign monitoring in septic emergency department patients in Rwanda. Digit Health 2019;5:2055207619879349 [FREE Full text] [doi: 10.1177/2055207619879349] [Medline: 31632685]
69. Weenk M, Koeneman M, van de Belt TH, Engelen LJLPG, van Goor H, Bredie SJH. Wireless and continuous monitoring of vital signs in patients at the general ward. Resuscitation 2019;136:47-53. [doi: 10.1016/j.resuscitation.2019.01.017] [Medline: 30685546]
70. Thomas-Rüddel DO, Hoffmann P, Schwarzkopf D, Scheer C, Bach F, Komann M, MEDUSA study group. Fever and hypothermia represent two populations of sepsis patients and are associated with outside temperature. Crit Care 2021;25(1):368 [FREE Full text] [doi: 10.1186/s13054-021-03776-2] [Medline: 34674733]
71. Protsiv M, Ley C, Lankester J, Hastie T, Parsonnet J. Decreasing human body temperature in the United States since the industrial revolution. Elife 2020;9:e49555 [FREE Full text] [doi: 10.7554/eLife.49555] [Medline: 31908267]

XSL•FO
RenderX

72. Geneva II, Cuzzo B, Fazili T, Javaid W. Normal body temperature: a systematic review. Open Forum Infect Dis 2019;6(4):ofz032 [FREE Full text] [doi: 10.1093/ofid/ofz032] [Medline: 30976605]
73. Bardia A, Deshpande R, Michel G, Yanez D, Dai F, Pace NL, et al. Demonstration and performance evaluation of two novel algorithms for removing artifacts from automated intraoperative temperature data sets: multicenter, observational, retrospective study. JMIR Perioper Med 2022;5(1):e37174 [FREE Full text] [doi: 10.2196/37174] [Medline: 36197702]
74. van Rossum MC, da Silva PMA, Wang Y, Kouwenhoven EA, Hermens HJ. Missing data imputation techniques for wireless continuous vital signs monitoring. J Clin Monit Comput 2023;37(5):1387-1400 [FREE Full text] [doi: 10.1007/s10877-023-00975-w] [Medline: 36729298]
75. Tutsoy O, Sumbul HE. A novel deep machine learning algorithm with dimensionality and size reduction approaches for feature elimination: thyroid cancer diagnoses with randomly missing data. Brief Bioinform 2024;25(4):bbae344 [FREE Full text] [doi: 10.1093/bib/bbae344] [Medline: 39007597]
76. Papaioannou VE, Chouvarda IG, Maglaveras NK, Pneumatikos IA. Temperature variability analysis using wavelets and multiscale entropy in patients with systemic inflammatory response syndrome, sepsis, and septic shock. Crit Care 2012;16(2):R51 [FREE Full text] [doi: 10.1186/cc11255] [Medline: 22424316]
77. Bhavani SV, Spicer A, Sinha P, Malik A, Lopez-Espina C, Schmalz L, et al. Distinct immune profiles and clinical outcomes in sepsis subphenotypes based on temperature trajectories. Intensive Care Med 2024;50(12):2094-2104. [doi: 10.1007/s00134-024-07669-0] [Medline: 39382693]
78. Dinh-Le C, Chuang R, Chokshi S, Mann D. Wearable health technology and electronic health record integration: scoping review and future directions. JMIR Mhealth Uhealth 2019;7(9):e12861 [FREE Full text] [doi: 10.2196/12861] [Medline: 31512582]
79. McLean KA, Knight SR, Diehl TM, Varghese C, Ng N, Potter MA, et al. Readiness for implementation of novel digital health interventions for postoperative monitoring: a systematic review and clinical innovation network analysis. Lancet Digit Health 2023;5(5):e295-e315 [FREE Full text] [doi: 10.1016/S2589-7500(23)00026-2] [Medline: 37100544]
80. Ginsburg GS, Picard RW, Friend SH. Key issues as wearable digital health technologies enter clinical care. N Engl J Med 2024;390(12):1118-1127. [doi: 10.1056/NEJMra2307160] [Medline: 38507754]
81. Vaghasiya JV, Mayorga-Martinez CC, Pumera M. Telemedicine platform for health assessment remotely by an integrated nanoarchitectonics FePS/rGO and TiC-based wearable device. Npj Flex Electron 2022;6(1):73 [FREE Full text] [doi: 10.1038/s41528-022-00208-1] [Medline: 35990769]
82. Raposo A, Marques L, Correia R, Melo F, Valente J, Pereira T, et al. e-CoVig: a novel mHealth system for remote monitoring of symptoms in COVID-19. Sensors (Basel) 2021;21(10):3397 [FREE Full text] [doi: 10.3390/s21103397] [Medline: 34068131]
83. Rajkumar SK, Samhitha PSA, Meenakshi T. Innovative flexible body sensor for in vitro vital signs monitoring. 2023 Presented at: 2023 International Conference on Next Generation Electronics (NEleX); December 14-16, 2023; Vellore, India p. 1-6. [doi: 10.1109/nelex59773.2023.10421589]
84. Azeem M, Shahid M, Masin I, Petru M. Design and development of textile-based wearable sensors for real-time biomedical monitoring; a review. J Text Inst 2024;116(1):80-95. [doi: 10.1080/00405000.2024.2318500]
85. Gongor F, Tutsoy O. Doctor robots: design and implementation of a heart rate estimation algorithm. Int J of Soc Robotics 2022;14(6):1435-1461. [doi: 10.1007/s12369-022-00888-9]
86. Voss F, Lyra S, Blase D, Leonhardt S, Lüken M. A setup for camera-based detection of simulated pathological states using a neonatal phantom. Sensors (Basel) 2022;22(3):957 [FREE Full text] [doi: 10.3390/s22030957] [Medline: 35161702]
87. van Melzen R, Haveman ME, Schuurmann RCL, Struys MMRF, de Vries JPM. Implementing wearable sensors for clinical application at a surgical ward: points to consider before starting. Sensors (Basel) 2023;23(15):6736 [FREE Full text] [doi: 10.3390/s23156736] [Medline: 37571519]
88. Mak S, Thomas A. Steps for conducting a scoping review. J Grad Med Educ 2022;14(5):565-567 [FREE Full text] [doi: 10.4300/JGME-D-22-00621.1] [Medline: 36274762]

Abbreviations

CE: Conformité Européenne
CRS: cytokine release syndrome
EHR: electronic health record
EU: European Union
EWS: early warning score
FDA: US Food and Drug Administration
ICU: intensive care unit
JBI: Joanna Briggs Institute
NFC: near-field communication
OSF: Open Science Framework
PCC: population, concept, context

XSL•FO
RenderX

PRESS: Peer Review of Electronic Search Strategies
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
ROC: receiver operating characteristic
SpO2: peripheral oxygen saturation
WHO: World Health Organization
XGBoost: extreme gradient boosting

Edited by L Buis; submitted 21.Jan.2026; peer-reviewed by P Sharma, CK Drummond, J McClay, O Tutsoy; comments to author 24.Jun.2026; revised version received 31.Jul.2026; accepted 20.Aug.2026; published 11.Sep.2026.

Please cite as:
Liu Y, Zhao Q, Li K, Dowding D
Continuous Temperature Monitoring via Wearable Devices for Fever or Infection Management in Acute Care Hospitals: Scoping Review of Clinical Implementation and Data Analytics
JMIR Mhealth Uhealth 2026;14:e89773
URL: https://mhealth.jmir.org/2026/1/e89773
doi:10.2196/89773
PMID:

©Yuwei Liu, Qimeng Zhao, Ka Li, Dawn Dowding. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Effectiveness of Wrist-Worn Wearable–Guided Cardiac Rehabilitation Delivered Through Fully Remote or Hybrid Programs in Patients With Cardiovascular Disease: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Jun Hwan Choi[1], MD; So Young Lee[1], MD, PhD; Kyung Cheon Seo[2], MD; Sang Yoon Lee[3], MD, PhD; Hyo Kyung Lee[4], PhD; Dongbin Heo[1], MD; Hyun Ryu[2], MD; Jong Hoon Kim[4], BS; Jeong Eul Kwon[4], BS; Bo Ryun Kim[2,5*], MD, PhD; Michael Y Lee[5*], MD, MHA, CPE

[1]Department of Rehabilitation Medicine, Jeju National University Hospital, Jeju National University, Jeju, Republic of Korea

[2]Department of Physical Medicine and Rehabilitation, Korea University Anam Hospital, Korea University College of Medicine, Seoul, Republic of Korea

[3]Department of Rehabilitation Medicine, Seoul National University Hospital, Seoul National University, Seoul, Republic of Korea

[4]School of Industrial Management Engineering, Korea University College of Engineering, Seoul, Republic of Korea

[5]H Ben Taub Department of Physical Medicine and Rehabilitation, Baylor College of Medicine, Houston, TX, United States

[*]these authors contributed equally

Corresponding Author:
Bo Ryun Kim, MD, PhD
Department of Physical Medicine and Rehabilitation
Korea University Anam Hospital
Korea University College of Medicine
73 Goryeodae-ro
Seongbuk-gu
Seoul, 02841
Republic of Korea
Phone: 82 2 920 6471
Fax: 82 2 929 9951
Email: brkim08@korea.ac.kr

Abstract

Background: Cardiovascular disease (CVD) remains the leading cause of death worldwide, yet participation in cardiac rehabilitation (CR) remains suboptimal. Wrist-worn wearables may support fully remote or hybrid CR through activity monitoring and individualized feedback, but the effectiveness and consistency of benefits remain uncertain.

Objective: This study aimed to evaluate the effectiveness of wrist-worn wearable–guided CR delivered through fully remote or hybrid programs initiated during phase II or III CR in patients with CVD.

Methods: A total of 8 databases were searched for randomized controlled trials published from January 1, 2015, through August 7, 2026. Eligible trials enrolled adults with CVD and integrated a wrist-worn wearable into fully remote or hybrid CR. Studies focused primarily on heart failure or using wearables only for outcome measurement were excluded. Primary outcomes were peak oxygen consumption (VO_2), 6-minute walk distance (6MWD), and daily steps. Random-effects meta-analyses used mean differences (MDs), except for quality of life, which used standardized MDs. Between-study variance was estimated using the DerSimonian-Laird method, pooled-effect CIs used the Knapp-Hartung adjustment, and 95% prediction intervals (PIs) used the Nagashima-Noma-Furukawa parametric bootstrap method. Risk of bias was assessed with Cochrane Risk of Bias 2. Subgroup and univariable meta-regression analyses explored study and intervention characteristics.

Results: A total of 18 trials involving 1993 randomized participants, reported in 20 publications, were included; 17 contributed to meta-analysis. Pooled estimates favored wearable-guided CR for peak VO_2 (10 studies; 591 participants; MD 1.69, 95% CI 0.46-2.92, 95% PI –2.21 to 5.39 mL/kg/min), 6MWD (9 studies; 1094 participants; MD 31.03, 95% CI 7.66-53.92, 95% PI –62.11 to 123.18 m), and daily steps (8 studies; 474 participants; MD 1062.8, 95% CI 383.5-1825.0, 95% PI –503.0 to 2860.0 steps/day).

XSL•FO
RenderX

All 3 PIs included the null, indicating uncertainty about consistency across future comparable settings. No clear pooled effects were observed for anaerobic threshold, BMI, waist circumference, grip strength, or quality of life. Subgroup differences were not statistically significant. In univariable meta-regression, a higher proportion of female participants was associated with greater peak VO_2 improvement, although this aggregate-level finding was exploratory.

Conclusions: Wrist-worn wearable–guided CR delivered through fully remote or hybrid programs may improve average exercise capacity and physical activity in patients with CVD. However, PIs included the null, indicating uncertainty about consistency across future settings. These programs may extend activity monitoring, individualized feedback, and behavioral support beyond center-based CR, but their multicomponent nature prevents attributing effects to wearables alone. Larger, rigorously designed trials with standardized protocols and outcomes should identify the patients and program components most likely to benefit. Future research should also evaluate long-term adherence, safety, equity, and implementation within routine multidisciplinary CR across diverse health systems.

(JMIR Mhealth Uhealth 2026;14:e101306) doi:10.2196/101306

KEYWORDS

cardiac rehabilitation; cardiovascular diseases; digital health; randomized controlled trials as topic; wearable electronic devices

Introduction

Cardiovascular disease (CVD) remains the leading cause of death worldwide. Annual deaths are projected to rise from 20.5 million in 2025 to 35.6 million in 2050, with global prevalence increasing by approximately 90% [1]. Its economic burden is also substantial; in the United States, annual cardiovascular health care costs are projected to reach US $1.49 trillion by 2050, with an additional US $361 billion in productivity losses [2]. Long-term risk reduction and functional recovery therefore remain major public health priorities.

Cardiac rehabilitation (CR) is a core component of secondary prevention, integrating individualized exercise, risk-factor management, education, and psychosocial support [3]. It reduces cardiovascular mortality and readmissions and improves exercise capacity and quality of life (QoL) [4,5], yet participation and completion remain suboptimal [6,7]. Travel, limited program availability, work and caregiving responsibilities, socioeconomic constraints, and physical limitations restrict engagement [8-10]. Home-based, community-based, remote, hybrid, and digitally supported models have therefore been developed to extend monitoring and behavioral support beyond centers [11-15].

Wearable devices are body-worn sensors that collect and transmit physiological or behavioral data, including chest-strap monitors, biosensors, smart garments, smartwatches, and activity trackers [15]. Previous reviews examined wearable activity monitoring during CR maintenance, heterogeneous wearable-assisted home CR, broader eHealth interventions, or mixed devices including pedometers and accelerometers [16-19]. They did not isolate actively guided wrist-worn programs initiated during phase II or III CR. Wrist-worn devices are clinically relevant because they are accessible, scalable, and can continuously monitor heart rate (HR) and activity with low user burden. Several recent randomized controlled trials (RCTs) also warrant updated synthesis.

Individual RCTs of wrist-worn wearable–guided CR have been small and inconsistent: Zhang et al [20] and Zheng et al [21] reported significant peak oxygen consumption (VO_2) gains, whereas Yagi et al [22] and Mitropoulos et al [23] did not. Effects on exercise capacity, physical activity, and related outcomes therefore remain uncertain.

This systematic review and meta-analysis evaluated wrist-worn wearable–guided CR delivered through fully remote or hybrid programs initiated during phase II or III CR in patients with CVD. We assessed exercise capacity, physical activity, and related outcomes and explored heterogeneity through subgroup and meta-regression analyses of participant, intervention, and comparator characteristics.

Methods

This systematic review and meta-analysis of RCTs followed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 [24]. The completed expanded checklist is provided in Multimedia Appendix 1.

Eligibility Criteria

Eligibility was defined using population, intervention, comparator, outcomes, and study design framework. We included adults with CVD receiving wrist-worn wearable–guided CR delivered through fully remote or hybrid programs initiated during phase II or III. Phase denoted when the intervention began, not eligibility of all delivery settings. Both phases were included to span structured postacute rehabilitation through longer-term maintenance; phase was extracted as a prespecified subgroup variable. Intervention duration was unrestricted.

Eligible participants included patients with coronary artery disease, acute coronary syndrome, ischemic heart disease, and those who had undergone percutaneous coronary intervention (PCI) or cardiac surgery, such as coronary artery bypass grafting, valve surgery, or aortic surgery, or transcatheter valve intervention. Studies primarily focused on patients with heart failure were excluded; however, studies enrolling mixed populations of patients with cardiovascular disease remained eligible when patients with heart failure constituted only a subset of the overall population.

Eligible interventions used a wrist-worn smartwatch or smartband as a core component of structured remote or hybrid CR to guide exercise prescription, monitor physical activity,

XSL•FO
RenderX

provide individualized feedback, or manage adherence. Conventional center-based CR alone and wearable use solely for passive monitoring or outcome assessment were ineligible. Comparators included usual care, unsupported home exercise, passive or minimally supported monitoring, and structured center-based CR.

Eligible studies were full-text RCTs published in English or Korean from January 1, 2015, through August 7, 2026. Conference abstracts, study protocols, review articles, nonrandomized or observational studies, and secondary or overlapping reports without independent eligible outcome data were excluded. Gray literature, including conference abstracts, dissertations, and unpublished materials, was not included because the review was restricted to full-text, peer-reviewed RCTs providing sufficient methodological information for eligibility and risk-of-bias assessment.

Information Sources

We searched PubMed, Web of Science, CENTRAL, Embase, Scopus, CINAHL, KoreaMed, and KMbase. The original search (November 14, 2025) was updated on August 7, 2026, covering January 1, 2015, through August 7, 2026. Reference lists of included reports and relevant reviews were examined. Trial registries, regulatory databases, and gray literature were not searched.

Search Strategy

Following PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Literature Search Extension) [25], database-specific strategies combined controlled vocabulary and free-text terms for CVD or cardiac procedures, wrist-worn wearables or remote monitoring, CR or related delivery models, and RCTs. Syntax, field tags, proximity operators, and filters were adapted by database. The August 7, 2026, update used the same concepts and updated date limits; overlapping records were deduplicated when day-specific restrictions were unavailable.

Complete original and updated strategies, platforms, dates, limits, filters, and yields are provided in Multimedia Appendix 2. No automated search tools were used; the strategies did not undergo peer review using the Peer Review of Electronic Search Strategies, and authors or manufacturers were not contacted.

Selection Process

Records were exported to EndNote (version 21; Clarivate) and deduplicated. Two reviewers (BRK and JHC) independently screened titles and abstracts; reports considered potentially eligible by either were retrieved. Four reviewers (BRK, JHC, S Young Lee, and KCS) independently assessed and cross-checked full texts, recorded exclusion reasons, and resolved disagreements by consensus. No automation or crowdsourcing was used, and no translation was required.

Data Collection Process

Three reviewers (BRK, JHC, and S Young Lee) extracted data using a standardized Excel form; 4 reviewers cross-checked entries and resolved discrepancies by consensus. Extracted variables covered study design; participant characteristics; CR phase and delivery model; intervention, wearable, feedback, exercise, and comparator characteristics; outcomes; and assessment timing.

Companion reports were linked as one study, with outcome data taken from the relevant report. Each cohort contributed only 1 estimate per outcome, preventing double counting. No automated extraction tools were used.

Data Items

Primary outcomes were peak VO_2, 6-minute walk distance (6MWD), and daily steps. Peak VO_2 was the highest directly measured oxygen uptake during maximal or symptom-limited cardiopulmonary exercise testing; relative values (mL/kg/min) were preferred, and equivalent VO_2 max measures were included. The 6MWD represented submaximal functional capacity, and daily steps represented habitual ambulatory activity rather than functional independence in activities of daily living.

Secondary outcomes included anaerobic threshold (AT), BMI, waist circumference, grip strength, and QoL. Peak VO_2 and AT were expressed in mL/kg/min, 6MWD in meters, daily steps in steps/day, BMI in kg/m^2, waist circumference in centimeters, and grip strength in kilograms. QoL was assessed using validated instruments, including the 36-Item Short Form Health Survey (SF-36), 12-Item Short Form Health Survey, Kansas City Cardiomyopathy Questionnaire, World Health Organization Quality of Life-Brief version, and EQ-5D. Additional outcomes reported by the included studies were recorded descriptively but were not included in the predefined quantitative syntheses.

We also extracted study setting and design; diagnosis; randomized, analyzed, and outcome-specific sample sizes; dropout; age; sex; intervention duration and dose; follow-up; and comparator components. Combined exercise included aerobic and resistance training; comparators were categorized as minimal/usual care, passive or less intensive wearable monitoring, or active structured CR.

Information that was missing or unclear in the published reports was recorded as not reported and was not imputed unless it could be directly calculated from the available data. When only the proportion of male participants was reported, the proportion of female participants was calculated as 100% minus the reported male proportion. Randomized, analyzed, and outcome-specific sample sizes were extracted separately when available; the sample size used in each meta-analysis was determined according to the population contributing to the corresponding outcome estimate, as described in the Synthesis Methods section.

Study Risk of Bias Assessment

Risk of bias was assessed at the outcome level using the revised Cochrane Risk of Bias 2 (RoB 2) tools in accordance with the Cochrane Handbook for Systematic Reviews of Interventions [26,27]. Four reviewers (BRK, JHC, S Young Lee, and KCS) independently assessed the studies assigned to them, with each outcome-specific assessment performed by at least 2 reviewers. The reviewers subsequently discussed their assessments, and disagreements were resolved by consensus.

XSL•FO
RenderX

The standard RoB 2 tool was used for parallel RCTs, the cluster-specific tool for the cluster RCT, and the crossover tool for the crossover RCT. Design-specific judgments, including recruitment timing and period or carryover effects, were mapped to harmonized D1-D5 labels for Multimedia Appendix 3.

Domain and overall judgments followed RoB 2 algorithms: low risk required all domains to be low; some concerns required at least one concern without a high-risk domain; and high risk reflected at least one high-risk domain or multiple concerns substantially lowering confidence. Plots used R (version 4.5.2; R Foundation for Statistical Computing) [28]; no automation was used.

Effect Measures

All outcomes were continuous. Mean differences (MDs) with 95% CIs were used when outcomes were reported using a common measurement scale. Accordingly, MDs were calculated for peak VO_2 and AT in mL/kg/min, 6MWD in meters, daily steps in steps/day, BMI in kg/m^2, waist circumference in centimeters, and grip strength in kilograms. For QoL, which was assessed using different validated instruments across studies, standardized mean differences (SMDs; Hedges *g*) with 95% CIs were calculated. Before calculating SMDs, the direction of each QoL scale was harmonized so that effect estimates had a consistent interpretation.

Positive MDs represented effects favoring the intervention for peak VO_2, AT, 6MWD, daily steps, and grip strength, whereas negative MDs represented effects favoring the intervention for BMI and waist circumference. For QoL, the direction of effect was defined consistently after harmonization of the scales. No prespecified thresholds were used to categorize pooled effects as small, moderate, or large.

Synthesis Methods

Outcome-specific meta-analyses included studies with clinically comparable measures and sufficient data to calculate an effect and variance. Otherwise, studies remained in the review and risk-of-bias assessment but not the synthesis. Postintervention values were preferred; change scores were used when unavailable. Units were converted as needed.

For studies reporting outcomes at multiple follow-up time points, only 1 assessment per outcome was included to avoid repeated inclusion of the same participants. The assessment conducted at or closest to the completion of the intervention was preferentially selected. When an assessment at intervention completion was unavailable, the earliest eligible postintervention assessment was used. For each outcome, the sample size corresponding to the reported summary statistics was used. The randomized sample size was used when results were reported according to the intention-to-treat principle, whereas the outcome-specific analyzed sample size was used when analyses were based on participants with available outcome data. Missing outcome data were not imputed unless the required statistics could be derived directly from information reported in the original article using established meta-analytic methods.

Multiple reports originating from the same trial cohort were linked and treated as companion publications. When companion reports provided data for different outcomes, data were extracted from the report relevant to each outcome. Each trial cohort contributed only one effect estimate to any given outcome-specific meta-analysis. For multiarm trials, only the comparison relevant to the eligibility criteria was included, and no shared intervention or comparator group was entered more than once in the same pooled analysis. Thus, no participant was double-counted within any outcome-specific meta-analysis.

MDs with 95% CIs were calculated for peak VO_2, 6MWD, daily steps, AT, BMI, waist circumference, and grip strength because these outcomes were reported using consistent measurement units. For QoL, SMDs using Hedges *g* with 95% CIs were calculated because different validated instruments were used across studies. Quantitative synthesis was restricted to validated overall or composite QoL scores with sufficient numerical data to calculate an effect estimate and variance. An unvalidated sum of SF-36 domain scores was not considered an eligible overall QoL score. Where necessary, score directions were harmonized so that higher values consistently represented better QoL.

Pooled effect estimates were calculated using inverse-variance random-effects models because clinical and methodological heterogeneity was anticipated across study populations and interventions. Between-study variance (τ^2) was estimated using the DerSimonian-Laird method, and 95% CIs for pooled effects were calculated using the Knapp-Hartung adjustment. Statistical heterogeneity was described using Cochran Q, τ^2, and I^2. I^2 was not interpreted using fixed low, moderate, or high categories because it represents the relative proportion of observed variability attributable to between-study heterogeneity rather than its absolute magnitude or clinical importance and can be imprecise when few studies are available. Heterogeneity was therefore interpreted jointly using τ^2, 95% prediction intervals, and clinical and methodological differences among studies [29,30].

For the main meta-analyses, 95% prediction intervals were calculated using the Nagashima-Noma-Furukawa parametric bootstrap method [31] to estimate the range in which the effect of a future comparable study might be expected. Prediction intervals were not calculated for subgroup analyses because of the small number of studies within individual subgroups and the resulting instability of subgroup-specific estimates. A reliable 95% prediction interval was not reported for QoL because marked between-study variability remained after exclusion of the study with an invalid SF-36 total score. Forest plots presented study-specific estimates, Knapp-Hartung-adjusted pooled estimates, and clearly labeled normal-normal framework (NNF) prediction intervals.

Subgroup and meta-regression analyses were planned to investigate potential sources of heterogeneity and patient- and intervention-level factors associated with intervention effectiveness. The specific outcomes eligible for these analyses were determined after data extraction based on the number of contributing studies and the availability of relevant subgroup and study-level covariate data, rather than on the magnitude or statistical significance of the pooled effects.

XSL•FO
RenderX

Subgroup analyses were conducted for peak VO_2 and 6MWD, for which sufficient studies were available, according to the CR phase at intervention initiation, provision of real-time feedback, type of wearable device, type of exercise, and comparator characteristics. Subgroup analyses were performed using restricted maximum likelihood estimation with the Knapp-Hartung adjustment. Differences between subgroups were evaluated using formal tests for subgroup interaction. The comparator-based subgroup analysis was added during the peer-review process and was therefore considered post hoc. Given the limited number of studies within individual subgroups, the findings were interpreted cautiously. A diagnosis-based subgroup analysis (eg, post-PCI, post–coronary artery bypass grafting, or mixed populations of patients with cardiovascular disease) was not performed because several studies enrolled heterogeneous populations of patients with cardiovascular disease and too few studies were available within specific diagnostic categories to support reliable subgroup estimates.

Univariable random-effects meta-regression analyses were conducted for peak VO_2 and 6MWD, for which sufficient study-level covariate data were available. These analyses were performed using restricted maximum likelihood estimation with the Knapp-Hartung adjustment. Mean participant age, proportion of female participants, and intervention duration were examined as potential moderators. Multivariable meta-regression was not performed because of the limited number of contributing studies. Meta-regression coefficients with 95% CIs and 2-sided *P* values were reported. Given the limited statistical power and use of aggregate study-level data, the meta-regression findings were interpreted as hypothesis-generating.

Sensitivity analyses were performed by excluding the cluster-randomized trial from the affected outcome-specific meta-analyses. A leave-one-out sensitivity analysis was additionally conducted for 6MWD by sequentially excluding each study and recalculating the pooled effect estimate and heterogeneity statistics. All analyses were performed using R version 4.5.2 (R Foundation for Statistical Computing) [28]. Statistical significance was defined as a 2-sided *P* value of <.05.

Reporting Bias Assessment

Small-study effects were assessed for outcome-specific meta-analyses including at least 10 studies through visual inspection of funnel plots and the Egger regression test. Accordingly, these assessments were performed for peak VO_2, which included 10 studies. Formal tests for funnel plot asymmetry were not performed for the remaining outcomes because fewer than 10 studies were available, resulting in limited statistical power and unreliable assessments of asymmetry. Funnel plot asymmetry was interpreted as potential evidence of small-study effects rather than definitive evidence of publication bias [26].

Certainty Assessment

The certainty of evidence was assessed for all predefined primary and secondary outcomes using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach [32]. The assessed outcomes included peak VO_2, 6MWD, daily steps, AT, BMI, waist circumference, grip strength, and QoL. Four reviewers (BRK, JHC, S Young Lee, and KCS) participated in the assessment. Outcomes were independently evaluated in reviewer pairs using a structured Microsoft Excel form, and disagreements were resolved through discussion and consensus among all reviewers. Final judgments were summarized using GRADEpro guideline development tool [33].

The certainty of evidence for each outcome was evaluated across 5 domains: risk of bias, inconsistency, indirectness, imprecision, and publication bias. Because the evidence was derived from RCTs, the initial certainty was rated as high and was downgraded by 1 or 2 levels when serious or very serious concerns were identified within a domain. Downgrading decisions considered the outcome-level RoB 2 assessments, clinical and statistical heterogeneity, consistency in the direction and magnitude of effects, width of the 95% CIs, amount of available information, applicability to the review question, and evidence of possible publication bias or small-study effects. Prediction intervals were considered supplementary information when evaluating inconsistency but were not used as an independent criterion for downgrading. The overall certainty of evidence was categorized as high, moderate, low, or very low.

Results

Study Selection

The original database search identified 1428 records, and the updated search conducted on August 7, 2026, identified 552 records, resulting in a total of 1980 records identified through electronic databases. Three additional reports were identified through other sources across the original and updated searches. After removal of 633 duplicate database records, 1347 database records underwent title and abstract screening, of which 1249 were excluded.

The remaining 98 database reports were sought for retrieval and successfully retrieved for eligibility assessment. The 3 reports identified through other sources were also retrieved and assessed. Thus, a total of 101 reports were assessed for eligibility, and 81 were excluded for the following reasons: not RCTs (n=9), not meeting the eligibility criteria for study participants (n=7), not meeting the intervention criteria (n=14), protocol publications (n=30), overlapping or salami publications without distinct eligible data (n=7), or abstract-only reports without sufficient full-text data (n=14).

The remaining 20 publications, representing 18 unique trials, were included in the systematic review [20-23,34-49]. The updated database search resulted in the inclusion of Nishio et al [38], while the supplementary manual search resulted in the inclusion of Claes et al [39]. Of the 20 included reports, 19 reports representing 18 unique trials contributed outcomes eligible for the RoB 2 assessment [20-23,34-47,49], while 18 reports representing 17 unique studies contributed data to the quantitative meta-analysis [20-23,34-47].

Lahtio et al [35] and Hakala et al [45] were identified as companion publications arising from the same cluster-randomized trial (ISRCTN61225589). Both reports were

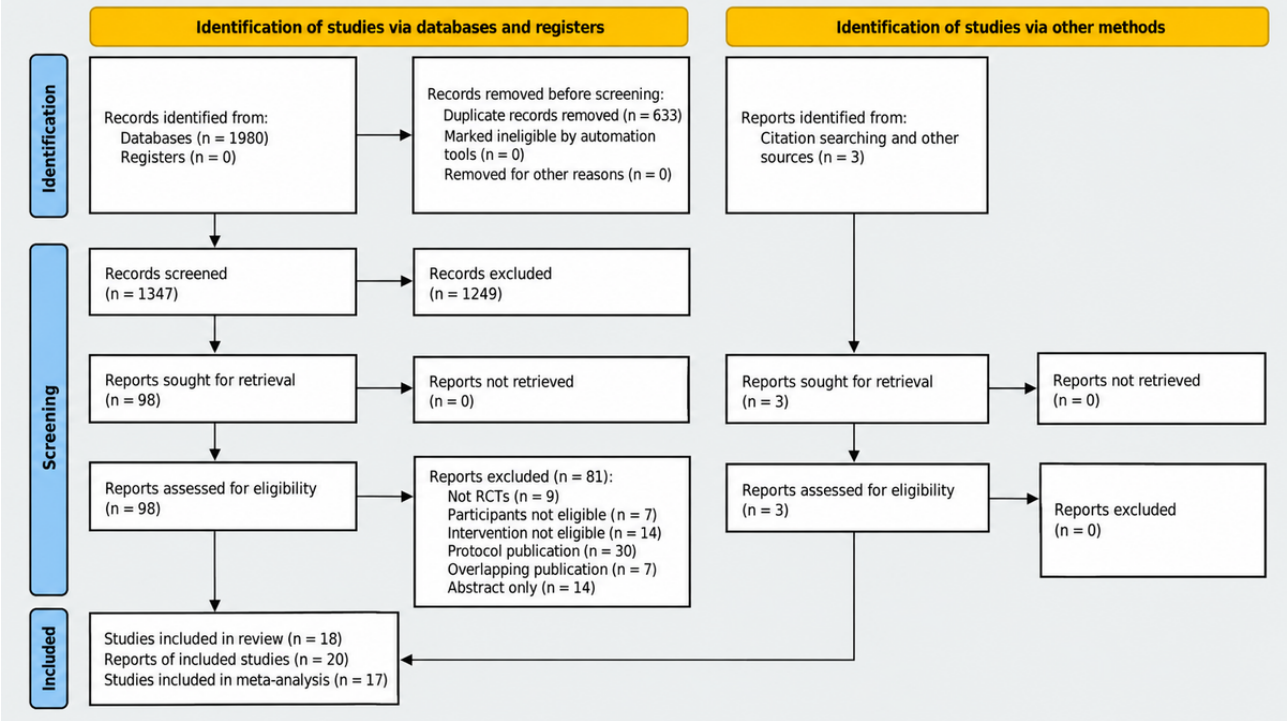
XSL•FO
RenderX

retained because they provided distinct eligible outcomes: Lahtio et al [35] contributed data on 6MWD, BMI, and waist circumference, whereas Hakala et al [45] contributed data on daily steps. The reports were linked and treated as arising from a single underlying study.

Mitropoulos et al [23] and Ntovoli et al [48] were also identified as companion publications arising from the same trial cohort (NCT06071273). Ntovoli et al [48] did not report any of the outcomes predefined for this review and was therefore not included in the outcome-level RoB 2 assessment or quantitative meta-analysis. However, it was retained as a companion report in the systematic review and linked to Mitropoulos et al [23]. Accordingly, the cohort was counted as a single unique study.

One included study [49] did not contribute to the quantitative meta-analysis because it reported quality-adjusted life-years, a composite outcome incorporating both quality and duration of life that was not directly comparable with the conventional QoL outcomes reported by the other studies. Each unique study contributed only one effect estimate to any given outcome-specific meta-analysis; therefore, no participant was double-counted within the same pooled analysis. The study-selection process is presented in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram of study selection. The PRISMA diagram shows identification, screening, eligibility, and inclusion. The review contained 20 publications from 18 trials; 17 trials contributed to the meta-analysis. RCT: randomized controlled trial.



Study Characteristics

Table 1 summarizes the characteristics of the 18 unique trials reported in 20 publications representing 18 unique trials, published between 2018 and 2025 [20-23,34-49]. At the unique-trial level, 6 studies were conducted in the United States [36,37,40,43,44,46], 2 each in China [20,21], the Czech Republic [34,41], and Japan [22,38], and 1 each in Greece [23,48], Finland [35,45], Uzbekistan [47], the Netherlands [49], and Belgium [42]. One multicenter study was conducted across Belgium and Ireland [39]. Companion publications from Greece [23,48] and Finland [35,45] were linked at the trial level, and each pair was counted only once in the country distribution.

XSL•FO
RenderX

Table 1. Characteristics of 20 publications representing 18 unique RCTs conducted across multiple countries and published between 2018 and 2025, evaluating wrist-worn wearable–guided CR delivered through fully remote or hybrid programs in adults with CVD, with interventions initiated during phase II or phase III CR.

Study, country	Study design	Population (intervention vs control)[a]	Intervention characteristics	Comparator type and components	Outcomes assessed[b]	Outcome assessment/follow-up
Zhang et al [20], (2025) China	Single-center RCT[c]	CHD[d] (angina, MI[e], PCI[f]); n=35 vs 35; dropout: 3 vs 5 (8/70, 11.4%); age: mean 62.67 (SD 4.89) years vs mean 57.2 (SD 13) years; male sex: 56.2% vs 63.3%	Timing: at discharge; delivery: home-based CR[g] with wearable+app+ICT[h]; components: real-time feedback, remote supervision, PA[i] and hemodynamic monitoring, medication reminders, educational content; device: smartwatch (Le-W20 Heart Health Watch); real-time feedback: yes; exercise: aerobic; dose: not reported; duration: 3 months	Usual care (standard HBCR[j] without technological assistance)	Peak VO_2[k], AT[l], QoL[m]	3 months
Zheng et al [21], (2024) China	Single-center RCT	CHD (PCI); n=55 vs 55; dropout: 2 vs 2 (4/110, 3.6%); age: mean 63.49 (SD 9.53) vs mean 63.04 (SD 10.44) years; male sex: 60% vs 57%	Timing: post discharge; delivery: home-based CR with wearable+app+ICT; components: remote supervision, PA and HR[n] monitoring; device: smartband (model not reported); real-time feedback: no; exercise: aerobic; dose: 3-5 days/week, session duration not reported; duration: 3 months	Usual care (discharge instructions and a health guide handbook)	6MWD[o], peak VO_2, AT, QoL	3 months
Yagi et al [22], (2025) Japan	Single-center RCT	CVD[p] (angina, MI, CHF[q]); n=23 vs 21; dropout: 4 vs 1 (5/44, 11.4%); age: 73 vs 76 years; male sex: 68% vs 95%	Timing: after outpatient center-based CR; delivery: home-based CR with wearable; components: PA and HR self-monitoring; device: smartwatch (Fitbit Versa 3); real-time feedback: no; exercise: aerobic, resistance; dose: 30 min/session, 5 days/week; duration: 3 months	Usual care (supervised hospital-based CR without wearable support)	Peak VO_2, 6MWD, AT, grip strength	3 months
Mitropoulos et al [23], (2024) Greece	Single-center RCT	CAD[r] (MI); n=18 vs 18; dropout: 3 vs 3 (6/36, 16.7%); age: mean 54.1 (SD 8.5) vs mean 53.5 (SD 6.6) years; male sex: 73% vs 73%	Timing: less than 4 weeks after MI; delivery: home-based CR with wearable+app+ICT; components: real-time feedback, remote supervision, PA and hemodynamic monitoring; device: smartwatch (ScanWatch); real-time feedback: yes; exercise: aerobic, resistance, balance, flexibility; dose: 60 min/session, 3 days/week; duration: 6 months	Community-based CR (supervised gym-based exercise without hemodynamic monitoring)	Peak VO_2, daily steps	6 months
Wohlfahrt et al [34], (2024)[s] Czech Republic	Single-center, crossover RCT	MI; n=32 vs 32; dropout: 1 vs 2 (3/64, 4.7%); age: mean 51.6 (SD 10.2) vs mean 51.2 (SD 10.5) years; male sex: 90% vs 90%	Timing: 1 month after MI; delivery: home-based CR with wearable+app+ICT; components: remote supervision, PA monitoring; device: smartwatch (model not reported); real-time feedback: no; exercise: aerobic; dose: individualized daily step count target (derived from 6MWT[t]: 0.9 × 5 × 6MWT steps); duration: 3 months per period	Usual care (physical activity advice without wearable support)	Peak VO_2, 6MWD, QoL	3 months for the meta-analysis (first period before crossover); total trial duration, 6 months
Lahtio et al [35], (2023)[u] Finland	Single-center, cluster RCT	CHD (angina, MI, PCI, CABG[v]); n=29 (3 clusters) vs 30 (3 clusters); dropout: 3 vs 3 (6/59, 10.2%); age: mean 59.7 (SD 6.0) vs mean 59.2 (SD 6.1) years; male sex: 77% vs 86%	Timing: PCI or CABG 3-12 months prior to the study; delivery: home-based CR with wearable+app+ICT; components: remote supervision, PA and HR monitoring, educational content; device: smartband (Fitbit Charge); real-time feedback: no; exercise: aerobic, resistance, flexibility; dose: not reported; duration: 12 months	Usual care (conventional CR without remote technology)	6MWD, BMI, waist circumference, QoL	12 months
Beckie et al [36], (2023) United States	Multicenter RCT	CHD; n=23 vs 24; dropout: 1 vs 4 (5/47, 10.6%); age: mean 62.7 (SD 7.3) vs mean 59.7 (SD 10.5) years; male sex: 0% vs 0%	Timing: CHD with or without PCI within the prior 5 years; delivery: home-based CR with wearable+app+ICT; components: remote supervision, PA monitoring, educational content; device: smartwatch (Android model not reported); real-time feedback: no; exercise: aerobic, balance, flexibility; dose: not reported; duration: 3 months	Usual care (CR workbook/DVD with weekly telephone support)	6MWD, BMI, waist circumference, QoL	3 months

XSL•FO
RenderX

Study, country	Study design	Population (intervention vs control)[a]	Intervention characteristics	Comparator type and components	Outcomes assessed[b]	Outcome assessment/follow-up
Lindman et al [37], (2021) United States	Multicenter RCT	TAVR[w]; n=25 vs 25; dropout: 0 (0%); age: mean 76 (SD 7) vs mean 76 (SD 9) years; male sex: 64% vs 68%	Timing: 1 month post-TAVR; delivery: home-based CR with wearable+app+ICT; components: remote supervision, PA and HR monitoring; device: smartband (Fitbit Alta HR); real-time feedback: no; exercise: aerobic, resistance; dose: not reported; duration: 6 weeks	Usual care (activity monitor without feedback or exercise instructions)	Daily steps, 6MWD, grip strength, QoL	6 weeks
Chokshi et al [40], (2018) United States	Multicenter RCT	CHD (angina, MI); n=50 vs 55; dropout: 1 vs 0 (1/105, 0.9%); age: mean 60 (SD 9.5) vs mean 59.1 (SD 11.5) years; male sex: 72.0% vs 67.3%	Timing: recent-onset ACS[x] or coronary catheterization; delivery: home-based CR with wearable+app+ICT; components: remote supervision, PA monitoring; device: smartband (Misfit Shine); real-time feedback: no; exercise: aerobic; dose: not reported; duration: 16 weeks	Usual care (wearable step monitoring without personalized goals or feedback)	Daily steps	4 months
Batalik et al [41], (2020) Czech Republic	Single-center RCT	CVD (angina, MI, PCI, CABG); n=28 vs 28; dropout: 3 vs 2 (5/56, 8.9%); age: mean 56.5 (SD 6.9) vs mean 57.7 (SD 7.6) years; male sex: 80% vs 85%	Timing: CVD in the last 6 months; delivery: home-based CR with wearable+app+ICT; components: remote supervision, PA and HR monitoring; device: smartband (Polar M430); real-time feedback: no; exercise: aerobic; dose: 60 min/session, 3 days/week; duration: 3 months	Center-based CR (supervised outpatient exercise)	Peak VO_2, QoL	3 months
Avila et al [42], (2020)[y] Belgium	Single-center RCT	CAD; n=30 vs 30 vs 30; dropout: 4 vs 1 vs 5 (10/90, 11.1%); age: mean 62.2 (SD 7.1) vs mean 62.0 (SD 7.4) vs mean 63.7 (SD 7.4) years; male sex: 88% vs 90% vs 92%	Timing: after discharge from phase II center-based CR; delivery: home-based CR with wearable+app+ICT; components: remote supervision, PA and HR monitoring; device: smartwatch (Garmin Forerunner), sensewear mini armband (BodyMedia); real-time feedback: no; exercise: aerobic; dose: 30 min/session, 5 days/week; duration: 3 months	Center-based CR (supervised exercise); usual care (physical activity advice only)	Peak VO_2, daily steps, BMI, waist circumference, grip strength, QoL	12 months
Park et al [43], (2021) United States	Multicenter RCT	CVD; n=32 vs 28; dropout: 6 vs 3 (9/60, 15.0%); age: mean 66.7 (SD 8.6) vs mean 66.8 (SD 8.7) years; male sex: 73% vs 80%	Timing: after discharge from phase II center-based CR; delivery: home-based CR with wearable+app+ICT; components: remote supervision, PA and hemodynamic self-monitoring, educational content; device: smartband (Fitbit Charge); real-time feedback: no; exercise: aerobic; dose: not reported; duration: 2 months	Usual care (pedometer, step diary, and telephone support)	Daily steps, 6MWD	2 months
Duscha et al [44], (2018) United States	Single-center RCT	CVD (angina, MI, PCI, CABG, valve repair, CHF [4/25, 16%]); n=21 vs 11; dropout: 5 vs 2 (7/32, 21.9%); age: mean 59.9 (SD 8.1) vs mean 66.5 (SD 7.2) years; male sex: 81.2% vs 66.7%	Timing: after discharge from phase II center-based CR; delivery: home-based CR with wearable+app+ICT; components: remote supervision, PA monitoring; device: smartband (Fitbit Charge); real-time feedback: no; exercise: aerobic; dose: not reported; duration: 3 months	Usual care (physician-directed physical activity advice)	Peak VO_2, daily steps	3 months
Hakala et al [45], (2021)[u] Finland	Single-center, cluster RCT	CHD (angina, MI, PCI, CABG); n=29 vs 30 (same cohort as Lahtio et al[35]); dropout: 6/59 (10.2%); age: mean 59.7 (SD 6.0) vs mean 59.2 (SD 6.1) years; male sex: 76% vs 87%	Timing: CHD who attended the center-based CR program; delivery: home-based CR with wearable+app+ICT; components: remote supervision, PA and HR monitoring, educational content; device: wrist-worn smartband (Fitbit Charge), hip-worn accelerometer (Fitbit Zip); real-time feedback: no; exercise: aerobic; dose: not reported; duration: 12 months	Usual care (conventional CR without remote technology)	Daily steps	12 months

XSL•FO
RenderX

Study, country	Study design	Population (intervention vs control)[a]	Intervention characteristics	Comparator type and components	Outcomes assessed[b]	Outcome assessment/follow-up
Dodson et al [46], (2025) United States	Multicenter RCT	CHD (MI, PCI, CABG); n=298 vs 102; dropout: 27 vs 17 (44/400, 11.0%); age: 71.0 vs 71.0 years; male sex: 72.5% vs 73.5%	Timing: hospitalized for AMI[z], PCI, or CABG within prior 4 weeks; delivery: home-based CR with wearable+app+ICT; components: remote supervision, PA and HR monitoring; device: smartband (Fitbit Inspire); real-time feedback: no; exercise: aerobic; dose: 30 min/session, 5 days/week; duration: 3 months	Usual care (traditional ambulatory CR permitted)	6MWD, QoL	3 months
Madina et al [47], (2025) Uzbekistan	Single-center RCT	MI; n=150 vs 150; dropout: 0 (0%); age: mean 57.8 (SD 8.9) vs mean 59.1 (SD 9.5) years; male sex: 67.3% vs 62.6%	Timing: MI within the 3 months prior to study; delivery: home-based CR with wearable+app+ICT; components: remote supervision, PA and HR monitoring, early warning sign reporting, medication reminders, dietary advice; device: smartband (model not reported); real-time feedback: no; exercise: aerobic; dose: not reported; duration: 6 months	Standard CR (exercise program and monthly physiotherapist sessions without digital monitoring)	6MWD, QoL	6 months
Ntovoli et al [48], (2025)[aa] Greece	Single-center RCT	CAD (MI); same trial as Mitropoulos et al [23]; age: 18-45 years (26.7%), 46-55 years (20.1%), 56-65 years (40%), >65 years (13.3%); male sex: 80%	Timing: less than 4 weeks after MI; delivery: home-based CR with wearable+app+ICT; components: real-time feedback, remote supervision, PA and hemodynamic monitoring; device: smartwatch (ScanWatch); real-time feedback: yes; exercise: aerobic, resistance, balance, flexibility; dose: 60 min/session, 3 days/week; duration: 6 months	Community-based CR (supervised gym-based exercise)	Perceived well-being (PERMA profiler), CR adherence, percentage of peak HR, RPE[ab]	6 months
Brouwers et al [49], (2021) Netherlands	Multicenter RCT	CAD (ACS, PCI); n=153 vs 147; dropout: 4 vs 12 lost to FU[ac] (16/300, 5.3%); age: mean 60.8 (SD 9.3) vs mean 60.5 (SD 9.8) years; male sex: 87.6% vs 89.8%	Timing: at phase II of outpatient CR; delivery: home-based CR with wearable+app+ICT; components: remote supervision, PA and HR monitoring; device: smartwatch (Mio Alpha), hip-worn accelerometer (ActiGraph wGT3x-BT); real-time feedback: no; exercise: aerobic, functional, resistance; dose: not reported; duration: 3 months CTR[ad]+9 months relapse prevention	Center-based CR (supervised group-based exercise)	QoL	12 months
Nishio et al [38], (2025) Japan	Single-center RCT	CAD (PCI, CABG); n=25 vs 25; dropout: 1 vs 0 lost to FU (1/50, 2.0%); age: mean 63.8 (SD 6.4) vs mean 62.6 (SD 7.9) years; male sex: 95.8% vs 92%	Timing: outpatient phase III CR candidates, after baseline CPET[ae]; delivery: home-based CR with wearable+app+ICT; components: real-time PA and HR monitoring, remote supervision, weekly individualized text-message coaching, monthly videoconferencing; device: smartwatch (Fitbit Sense); real-time feedback: yes; exercise: aerobic; dose: ≥30 min/session, 5 days/week; duration: 3 months	Wearable-supported CPET-based home exercise without online coaching	Peak VO_2, AT, daily steps, BMI	3 months

XSL•FO
RenderX

Study, country	Study design	Population (intervention vs control)[a]	Intervention characteristics	Comparator type and components	Outcomes assessed[b]	Outcome assessment/follow-up
Claes et al [39], (2020) Belgium	Multicenter RCT	CVD (PCI/CABG/valve repair); n=60 vs 60; dropout: 7 vs 13 lost to FU (20/120, 16.7%); age: mean 61.7 (SD 14.5) vs mean 59.6 (SD 13.2) years; male sex: 81.7% vs 81.7%	Timing: initiated during the final 4 weeks of center-based phase II CR and continued into phase III; delivery: home-based CR with wearable+app+ICT; components: remote supervision, PA and HR monitoring, telecoaching, social networking, automated supportive text messages or emails; device: smartband (Microsoft Band 2); real-time feedback: no; exercise: aerobic; dose: 150 min/week moderate-intensity PA; duration: 6 months	Usual care (physical activity advice without structured CR)	Peak VO_2, daily steps, grip strength, BMI, QoL	6 months

[a]Unless otherwise indicated, values are presented as intervention vs control. For studies with more than 2 arms, values are presented in the order of the study arms.

[b]Measurement methods and units: peak VO_2 and VO_2 max were measured using cardiopulmonary exercise testing (CPET) and expressed in mL/kg/min. AT was determined during CPET and expressed in mL/kg/min. The 6-minute walk distance was measured using the 6-minute walk test and expressed in meters. Physical activity was assessed using the wearable device or accelerometer specified in each study and reported as steps/day, minutes/day, or minutes/week. Heart rate was expressed as beats/min; blood pressure, as mmHg; body weight, as kg; BMI, as kg/m2; waist, hip, and calf circumferences, as cm; grip strength, as kg; left ventricular ejection fraction, as a percentage; and glycated hemoglobin, as a percentage. Quality of life and other patient-reported outcomes were assessed using the validated instruments specified in the original studies.

[c]RCT: randomized controlled trial.

[d]CHD: coronary heart disease.

[e]MI: myocardial infarction.

[f]PCI: percutaneous coronary intervention.

[g]CR: cardiac rehabilitation.

[h]ICT: information and communication technology.

[i]PA: physical activity.

[j]HBCR: home-based cardiac rehabilitation.

[k]VO_2: oxygen consumption.

[l]AT: anaerobic threshold.

[m]QoL: quality of life.

[n]HR: heart rate.

[o]6MWD: 6-minute walk distance.

[p]CVD: cardiovascular disease.

[q]CHF: congestive heart failure.

[r]CAD: coronary artery disease.

[s]Wohlfahrt et al [34] used a crossover design. Data from the first 3-month period before crossover were used for the meta-analysis to avoid potential carryover effects. The 64 participants were randomized to 2 treatment sequences (32 vs 32).

[t]6MWT: 6-minute walk test.

[u]Lahtio et al [35] and Hakala et al [45] are companion publications arising from the same cluster-randomized trial (ISRCTN61225589). They reported different eligible outcomes and were treated as one unique underlying trial for study-level counting, including the country distribution. No participant was counted more than once within any outcome-specific meta-analysis.

[v]CABG: coronary artery bypass grafting.

[w]TAVR: transcatheter aortic valve replacement.

[x]ACS: acute coronary syndrome.

[y]Avila et al [42] used a 3-arm design. For each outcome-specific meta-analysis, only the eligible wearable-supported intervention arm and the corresponding usual-care comparator were included; no comparator group was entered more than once.

[z]AMI: acute myocardial infarction.

[aa]Ntovoli et al [48] and Mitropoulos et al [23] are companion publications arising from the same randomized trial (NCT06071273). Ntovoli et al did not report any of the outcomes predefined for this review and therefore did not contribute to the outcome-level Risk of Bias 2 assessment or quantitative meta-analysis. The reports were linked and treated as one unique underlying trial for study-level counting, including the country distribution.

[ab]RPE: rating of perceived exertion.

[ac]FU: follow-up.

[ad]CTR: cardiac telerehabilitation.

[ae]CPET: cardiopulmonary exercise testing.

Lahtio et al [35] and Hakala et al [45] were companion publications arising from the same cluster-randomized trial (ISRCTN61225589), while Mitropoulos et al [23] and Ntovoli et al [48] were companion publications arising from the same

XSL•FO
RenderX

randomized trial (NCT06071273). Each companion-publication pair was linked and treated as representing one unique underlying trial for study-level counting, including the country distribution.

Of the 18 unique trials, 10 were single-center RCTs [20-23,34,38,41,42,44,47], 7 were multicenter RCTs [36,37,39,40,43,46,49], and 1 was a cluster RCT reported in 2 companion publications [35,45]. Wohlfahrt et al [34] used a crossover RCT design, whereas the remaining individually randomized studies used parallel-group designs.

All eligible intervention groups implemented remote or hybrid CR incorporating a wrist-worn wearable device as a core intervention component [20-23,34-49]. The wearable technologies included smartbands [21,35,37,39-47] and smartwatches [20,22,23,34,36,38,48,49]. A small number of studies incorporated additional sensors for intervention delivery or outcome assessment [39,45,49]. Nishio et al [38] used a Fitbit Sense integrated with a real-time monitoring platform and compared additional online coaching with wearable-supported CR without coaching. Claes et al [39] evaluated the PATHway technology-enabled home-based CR platform, which incorporated a Microsoft Band 2, remotely delivered exercise sessions, physical activity monitoring, automated feedback, education, and behavioral support.

The comparator interventions included usual care, home-based exercise without wearable-guided support, passive or minimally supported wearable monitoring, and structured center- or community-based CR [20-23,34-49]. Thus, the presence of a wearable device was not necessarily exclusive to the intervention group. Rather, the eligible intervention groups actively integrated wearable-derived data into CR delivery through monitoring, feedback, coaching, exercise guidance, or adherence management.

In terms of intervention components, nearly all studies included remote supervision and physical activity monitoring [20,21,23,34-49]. Many studies also incorporated HR or hemodynamic monitoring [20-23,35,37-39,41-43,45-49], whereas real-time or individualized feedback was provided in a subset of studies [20,23,38,39,48]. Additional components included educational or behavioral-support content [20,35,36,39,43,45,47] and warning systems for abnormal physiological signals [20,47].

The exercise programs were predominantly aerobic, including the intervention evaluated by Nishio et al [38], although several studies implemented combined training incorporating aerobic and resistance exercise [22,23,35,37,39,48,49]. Some interventions additionally included balance or flexibility exercises [23,35,36,39,48]. When reported, the prescribed exercise dose generally ranged from 30 to 60 minutes per session, performed 3-5 times per week [22,23,34,38,41,42,46,48].

Intervention duration ranged from 6 weeks to 12 months, with most interventions lasting approximately 3-6 months [20-23,34,36,38-42,44,46-49]. Follow-up assessments were generally aligned with intervention completion, although some studies reported extended follow-up to 12 months [42,46,49]. When multiple assessment time points were available, the assessment conducted at or closest to intervention completion was preferentially selected for quantitative synthesis. When no assessment was reported at intervention completion, the earliest eligible postintervention assessment was used. In Avila et al [42], the relevant outcomes were available at the 12-month follow-up following a 3-month intervention.

Overall, the 18 unique trials randomized 1993 participants with CVD receiving wrist-worn wearable–guided CR through fully remote or hybrid programs initiated during phase II or phase III CR. Of the 20 included publications, 18 publications representing 17 unique trials contributed data to at least 1 quantitative meta-analysis [20-23,34-47]. Because the number of participants contributing data differed across outcomes and assessment time points, outcome-specific sample sizes are reported with each meta-analysis. Across the included studies, mean participant age ranged from approximately 51 to 76 years, and the proportion of male participants varied substantially, ranging from 0% to approximately 96%; 1 study enrolled only female participants [36]. Individual study sample sizes ranged from 25 to 400 participants. Most participants had coronary artery disease or had undergone PCI or cardiac surgery.

Risk of Bias in Studies

Among the 20 publications representing 18 unique trials, 19 reports were included in the outcome-level RoB assessment. Ntovoli et al [48], a companion publication from the same trial as Mitropoulos et al [23], was not assessed because it did not report any of the predefined outcomes of interest. RoB was assessed separately for each outcome using the standard or design-specific Cochrane RoB 2 tool, as appropriate.

For peak VO_2, 4 of the 10 contributing studies were rated as having low overall RoB and 6 as having some concerns [20-23,34,38,39,41,42,44], with no studies rated as having high RoB. For 6MWD, 8 of the 9 contributing studies were rated as having some concerns and 1 as having high RoB [21,22,34-37,43,46,47]. All studies contributing to BMI [35,36,38,39,42] and waist circumference [35,36,42] were rated as having some concerns. For AT, 2 of the 4 contributing studies were rated as having low RoB and 2 as having some concerns [20-22,38]. For grip strength, 3 of the 4 contributing studies were rated as having some concerns and 1 as having high RoB [22,37,39,42].

In contrast, substantial RoB concerns were observed for daily steps and QoL. For daily steps, 7 of the 9 studies were rated as having high overall RoB [37,38,40,42-45], primarily because of bias due to deviations from intended interventions. For QoL, 10 of the 13 studies were rated as having high overall RoB [20,34-36,38,39,42,46,47,49], primarily because of bias in outcome measurement. The outcome-specific RoB assessments are summarized in Multimedia Appendix 3.

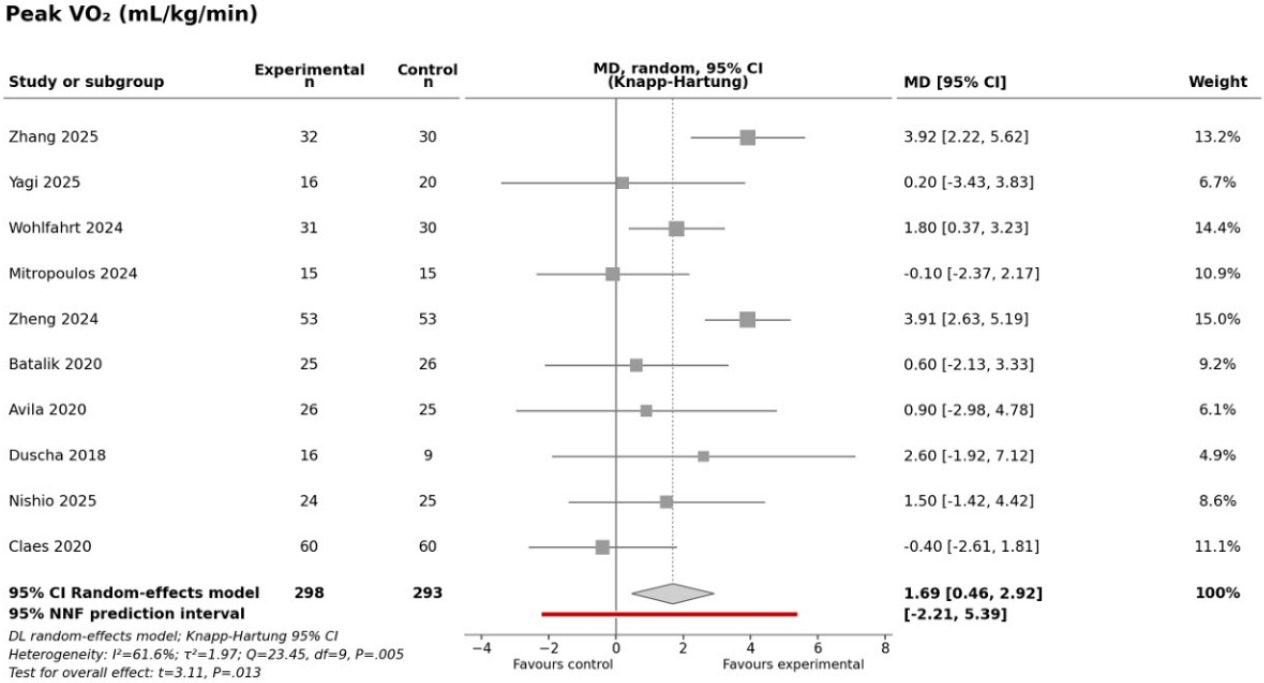
Results of Individual Studies

Outcome-specific summary statistics and effect estimates for each included study are presented in the forest plots (Figures 2-9) and the corresponding supplementary tables (Multimedia Appendix 4). For continuous outcomes, individual study effects are reported as MDs with 95% CIs for peak VO_2, 6MWD, daily

XSL•FO
RenderX

steps, AT, BMI, waist circumference, and grip strength. For QoL, individual study effects are reported as SMDs using Hedges *g* with 95% CIs because different validated instruments were used across studies. The number of participants contributing data to each outcome-specific effect estimate is also presented. Individual study findings varied in magnitude and, for several outcomes, in direction, as reflected in the forest plots and the between-study heterogeneity reported in the Results of Syntheses section.

Figure 2. Random-effects meta-analysis of peak oxygen consumption (VO2). Study and pooled random-effects mean differences (MDs) are shown with 95% CIs. DerSimonian-Laird (DL) estimated between-study variance. Pooled CIs used the Knapp-Hartung adjustment. The labeled normal-normal framework (NNF) 95% prediction interval represents a future comparable study.



Results of Syntheses

Overview

For each outcome, the pooled estimate and 95% CI represent the estimated average effect across the included studies, whereas the 95% prediction interval describes the expected distribution of true effects across comparable clinical settings while accounting for between-study heterogeneity.

Primary Outcomes

Peak VO2

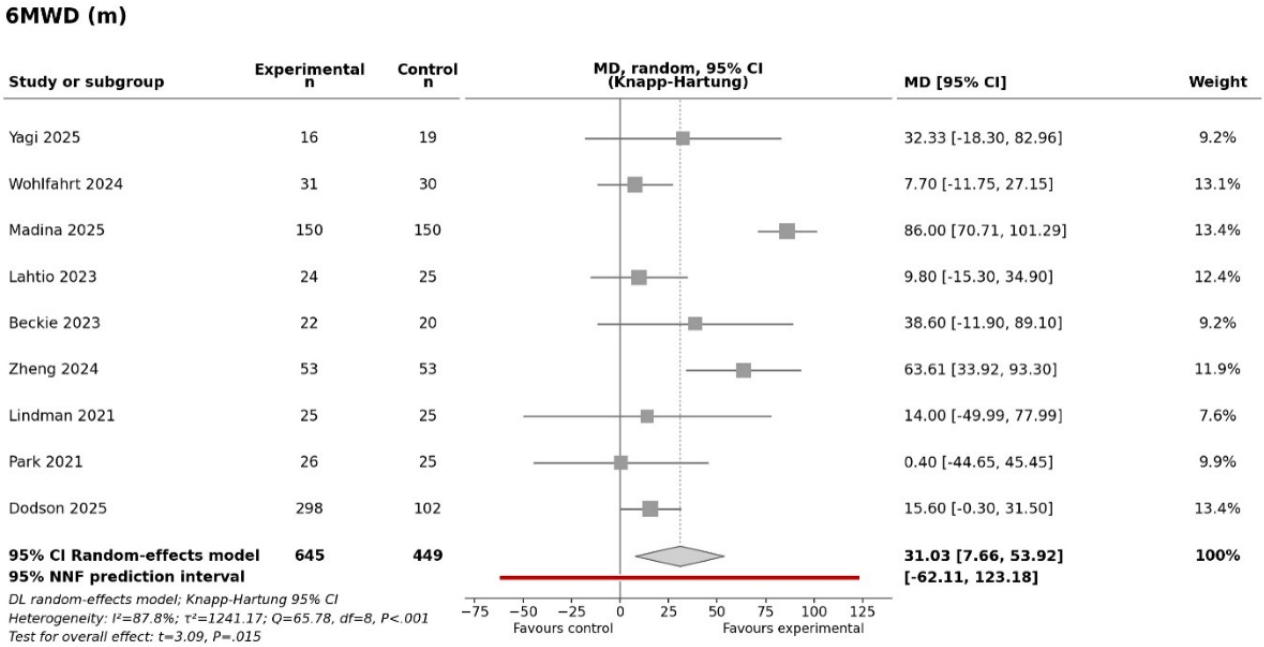
A total of 10 studies involving 591 participants were included in the meta-analysis of peak VO_2 [20-23,34,38,39,41,42,44]. On average, wrist-worn wearable–guided CR delivered through fully remote or hybrid programs was associated with greater peak VO_2 than the control interventions (MD 1.69, 95% CI 0.46-2.92 mL/kg/min; *P*=.01). Between-study variability was reflected by τ^2=1.97; I^2=61.6%; and Cochran Q=23.45 (*P*=.005). The NNF 95% prediction interval ranged from –2.21 to 5.39 mL/kg/min and included the null. Thus, although the estimated average effect favored the intervention, a benefit may not be observed consistently across future comparable settings (Figure 2).

6MWD

A total of 9 studies involving 1094 participants were included in the meta-analysis of 6MWD [21,22,34-37,43,46,47]. The pooled estimate indicated greater average 6MWD in the intervention groups than in the control groups (MD 31.03, 95% CI 7.66-53.92 m; *P*=.02). Between-study variability was reflected by τ^2=1241.17; I^2=87.8%; and Cochran Q=65.78 (*P*<.001). The NNF 95% prediction interval was wide and included the null (–62.11 to 123.18 m), indicating uncertainty in the effect expected in a future comparable setting despite the favorable average effect (Figure 3).

XSL•FO
RenderX

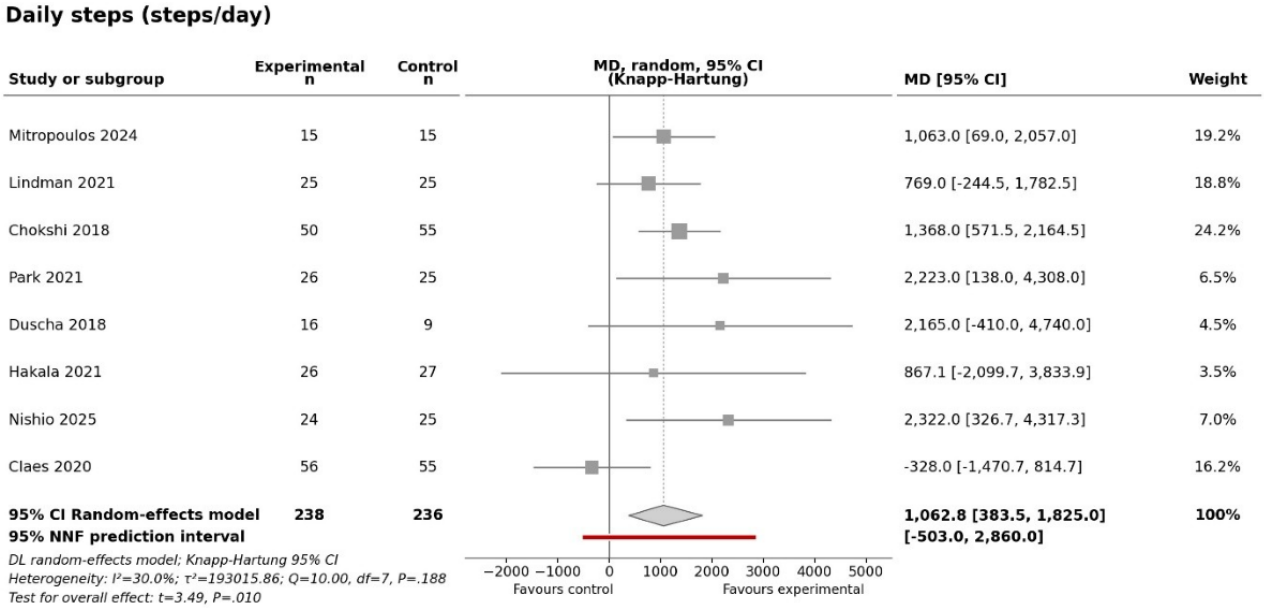
Figure 3. Random-effects meta-analysis of 6-minute walk distance (6MWD). Study and pooled random-effects mean differences (MDs) for 6MWD are shown with 95% CIs. DerSimonian-Laird (DL) estimated variance, pooled CIs used Knapp-Hartung adjustment, and the labeled normal-normal framework (NNF) 95% prediction interval represents a future comparable study.



Daily Steps

A total of 9 studies reported daily steps. Avila et al [42] was excluded from quantitative synthesis because the report did not provide sufficient numerical information to calculate a daily-step effect estimate and its variance. The remaining 8 studies involving 474 participants were included [23,37-40,43-45]. On average, the intervention groups accumulated more daily steps than the control groups (MD 1062.8, 95% CI 383.5-1825.0 steps/day; *P*=.01). Between-study variability was reflected by τ^2=193,015.86; I^2=30.0%; and Cochran Q=10.00 (*P*=.19). The NNF 95% prediction interval included the null (–503.0 to 2860.0 steps/day), suggesting that a benefit may not be observed consistently across future comparable settings (Figure 4).

Figure 4. Random-effects meta-analysis of daily steps. Study and pooled random-effects mean differences (MDs) in daily steps are shown with 95% CIs. DerSimonian-Laird (DL) estimated variance, pooled CIs used the Knapp-Hartung adjustment, and the labeled normal-normal framework (NNF) 95% prediction interval represents a future comparable study.



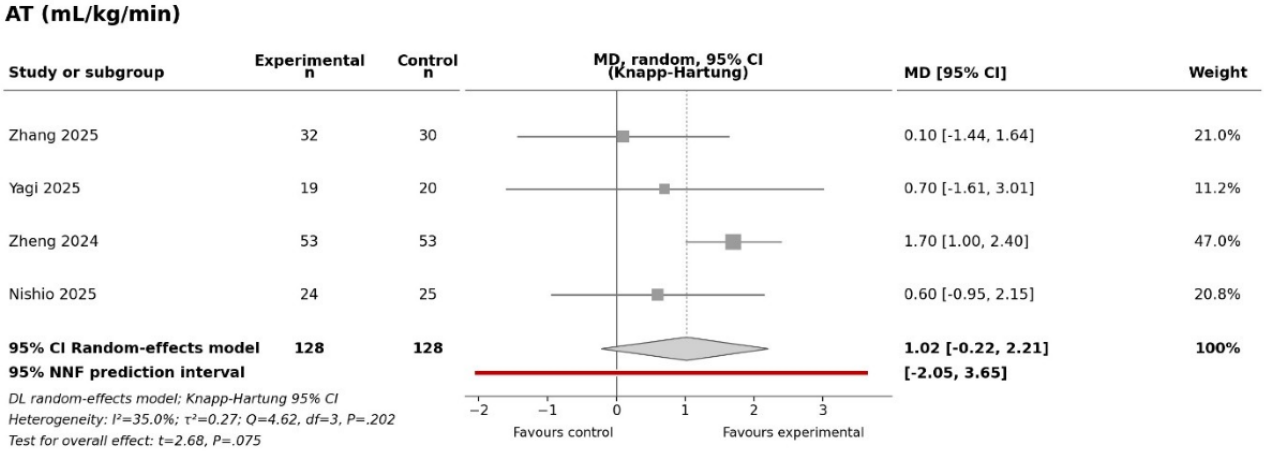
XSL•FO
RenderX

Secondary Outcomes

AT

A total of 4 studies involving 256 participants were included in the meta-analysis of AT [20-22,38]. The pooled estimate favored the intervention but was imprecise and included the null after Knapp-Hartung adjustment (MD 1.02, 95% CI –0.22 to 2.21 mL/kg/min; *P*=.08). Between-study variability was reflected by τ^2=0.27; I^2=35.0%; and Cochran Q=4.62 (*P*=.20). The NNF 95% prediction interval also included the null (–2.05 to 3.65 mL/kg/min), providing no clear evidence of a consistent benefit (Figure 5).

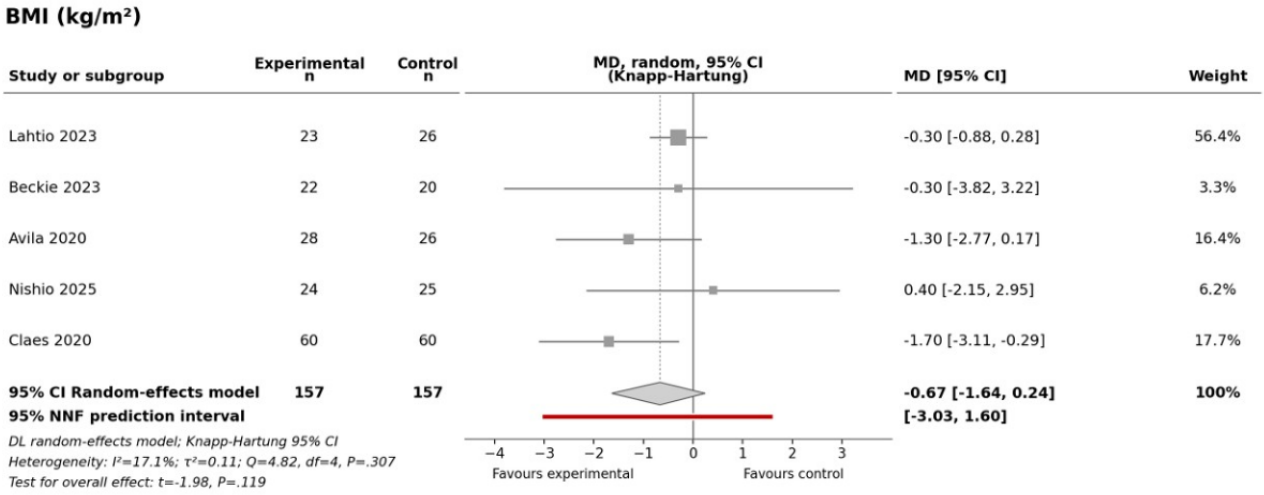
Figure 5. Random-effects meta-analysis of anaerobic threshold (AT). Study and pooled random-effects mean differences (MDs) in AT are shown with 95% CIs. DerSimonian-Laird (DL) estimated variance, pooled CIs used the Knapp-Hartung adjustment, and the labeled normal-normal framework (NNF) 95% prediction interval represents a future comparable study.



BMI

A total of 5 studies involving 314 participants were included in the meta-analysis of BMI [35,36,38,39,42]. The pooled estimate suggested a possible reduction but was imprecise and included the null after Knapp-Hartung adjustment (MD –0.67, 95% CI –1.64 to 0.24 kg/m^2; *P*=.12). Between-study variability was reflected by τ^2=0.11; I^2=17.1%; and Cochran Q=4.82 (*P*=.31). The NNF 95% prediction interval included the null (–3.03 to 1.60 kg/m^2), providing no clear evidence of a consistent reduction (Figure 6).

Figure 6. Random-effects meta-analysis of BMI. Study and pooled random-effects mean differences (MDs) in BMI are shown with 95% CIs. DerSimonian-Laird (DL) estimated variance, pooled CIs used Knapp-Hartung adjustment, and the labeled normal-normal framework (NNF) 95% prediction interval represents a future comparable study.

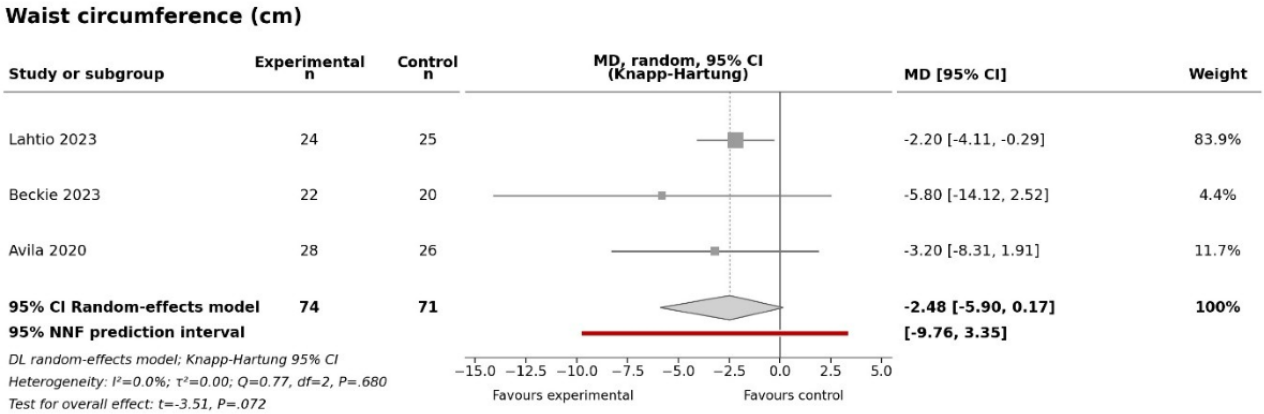


Waist Circumference

A total of 3 studies involving 145 participants were included in the meta-analysis of waist circumference [35,36,42]. The pooled estimate suggested a possible reduction but was imprecise and included the null after Knapp-Hartung adjustment (MD –2.48, 95% CI –5.90 to 0.17 cm; *P*=.07). The estimated between-study variance was zero (τ^2=0; I^2=0%; Cochran Q=0.77; *P*=.68), but this estimate was based on only 3 studies and should not be interpreted as evidence of identical effects. The NNF 95% prediction interval included the null (–9.76 to 3.35 cm), providing no clear evidence of a consistent reduction (Figure 7).

XSL•FO
RenderX

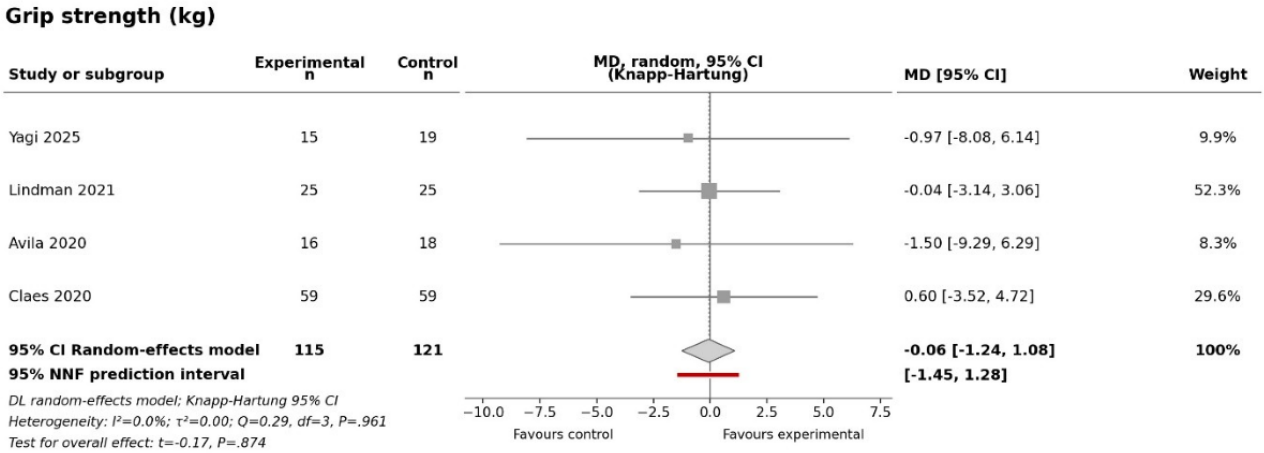
Figure 7. Random-effects meta-analysis of waist circumference. Study and pooled random-effects mean differences (MDs) in waist circumference are shown with 95% CIs. DerSimonian-Laird (DL) estimated variance, pooled CIs used the Knapp-Hartung adjustment, and the labeled normal-normal framework (NNF) 95% prediction interval represents a future comparable study.



Grip Strength

A total of 4 studies involving 236 participants were included in the meta-analysis of grip strength [22,37,39,42]. The pooled analysis showed no evidence of an average between-group difference (MD –0.06, 95% CI –1.24 to 1.08 kg; *P*=.87). The estimated between-study variance was zero (τ^2=0; I^2=0%; Cochran Q=0.29; *P*=.96), although only 4 studies contributed. The NNF 95% prediction interval also included the null (–1.45 to 1.28 kg), providing no evidence of a consistent benefit (Figure 8).

Figure 8. Random-effects meta-analysis of grip strength. Study and pooled random-effects mean differences (MDs) in grip strength are shown with 95% CIs. DerSimonian-Laird (DL) estimated variance. Pooled CIs used Knapp-Hartung adjustment. The labeled normal-normal framework (NNF) 95% prediction interval represents a future comparable study.

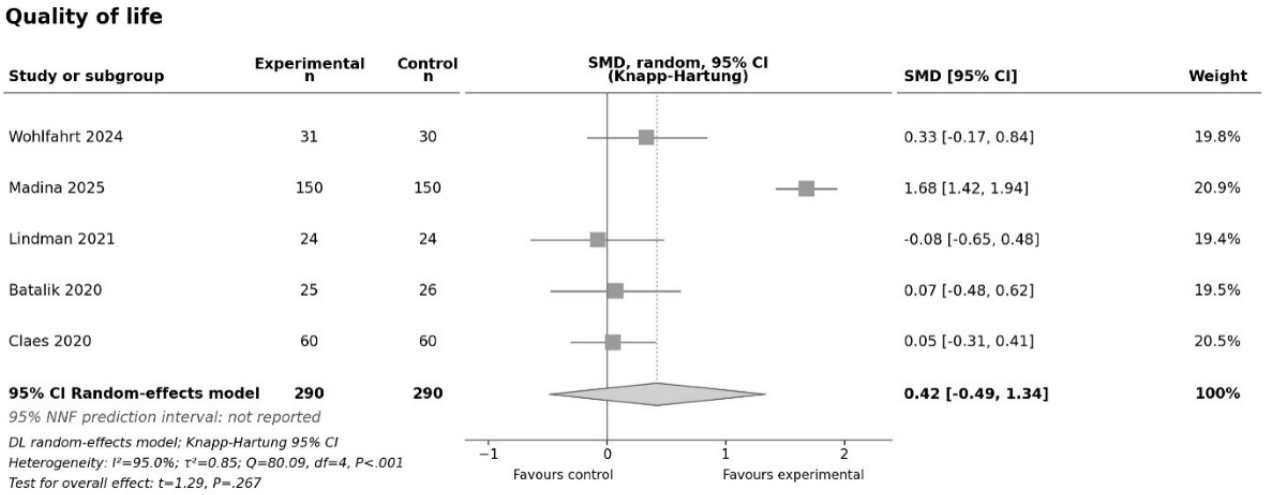


QoL

Although 13 studies reported QoL outcomes, only 5 studies involving 580 participants reported a valid overall or composite QoL score with sufficient numerical data for quantitative synthesis [34,37,39,41,47]. Zhang et al [20] was excluded because the reported SF-36 total score was not a validated summary score and was inconsistent with the stated 0-100 scoring range. The exploratory synthesis showed no clear average between-group difference (SMD 0.42, 95% CI –0.49 to 1.34; *P*=.27). Marked between-study variability remained (τ^2=0.85; I^2=95.0%; Cochran Q=80.09; *P*<.001), and a reliable NNF 95% prediction interval was not reported. Given the clinical and statistical variability across instruments and outcome domains, the pooled estimate was interpreted cautiously, and findings were considered primarily descriptive (Figure 9).

XSL•FO
RenderX

Figure 9. Random-effects meta-analysis of QoL. Study and pooled random-effects Hedges g standardized mean differences (SMDs) are shown with 95% CIs, and pooled CIs used Knapp-Hartung adjustment. Zhang et al [20] was excluded because its 36-Item Short Form Health Survey total was not validated. No reliable normal-normal framework (NNF) prediction interval was reported because variability remained marked. DL: DerSimonian-Laird.



Subgroup Analyses

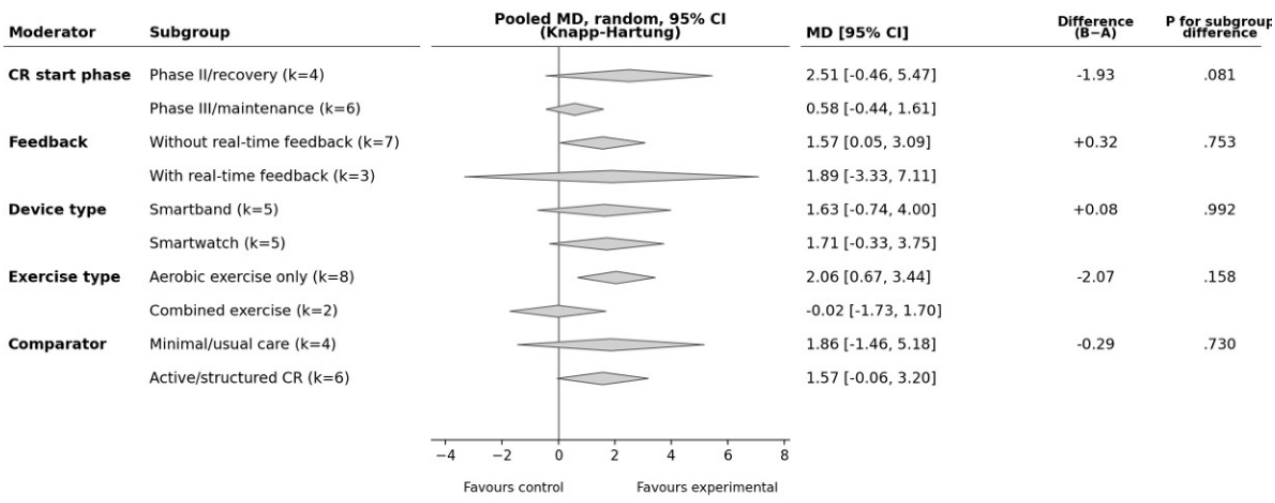
Peak VO2

Subgroup analyses for peak VO_2 were conducted according to the CR phase at intervention initiation, provision of real-time feedback, wearable-device type, exercise type, and comparator characteristics. With restricted maximum likelihood estimation and Knapp-Hartung-adjusted subgroup CIs, the average effect was numerically larger in studies in which CR was initiated during phase II (MD 2.51, 95% CI –0.46 to 5.47 mL/kg/min; 4 studies) than in those initiated during phase III (MD 0.58, 95% CI –0.44 to 1.61 mL/kg/min; 6 studies). However, the formal Knapp-Hartung interaction test did not demonstrate a statistically significant difference between the subgroups (interaction: *P*=.08).

The average effect was also numerically larger in studies using aerobic exercise alone (MD 2.06, 95% CI 0.67-3.44 mL/kg/min; 8 studies) than in studies using combined aerobic and resistance exercise (MD –0.02, 95% CI –1.73 to 1.70 mL/kg/min; 2 studies); however, the Knapp-Hartung subgroup difference was not statistically significant (interaction: *P*=.16).

No statistically significant subgroup differences were identified according to provision of real-time feedback (interaction: *P*=.75), wearable-device type (interaction: *P*=.99), or comparator characteristics (interaction: *P*=.73). Overall, none of the formal interaction tests reached statistical significance (Figure 10). The numerically larger effects observed with phase II initiation and aerobic-only exercise should therefore be considered hypothesis-generating, particularly given the small number of studies within some subgroups.

Figure 10. Subgroup analyses of peak oxygen consumption (VO2). Random-effects subgroup estimates used restricted maximum likelihood and Knapp-Hartung 95% CIs. No interaction was statistically significant (cardiac rehabilitation [CR] phase: P=.08; exercise type: P=.16). MD: mean difference.

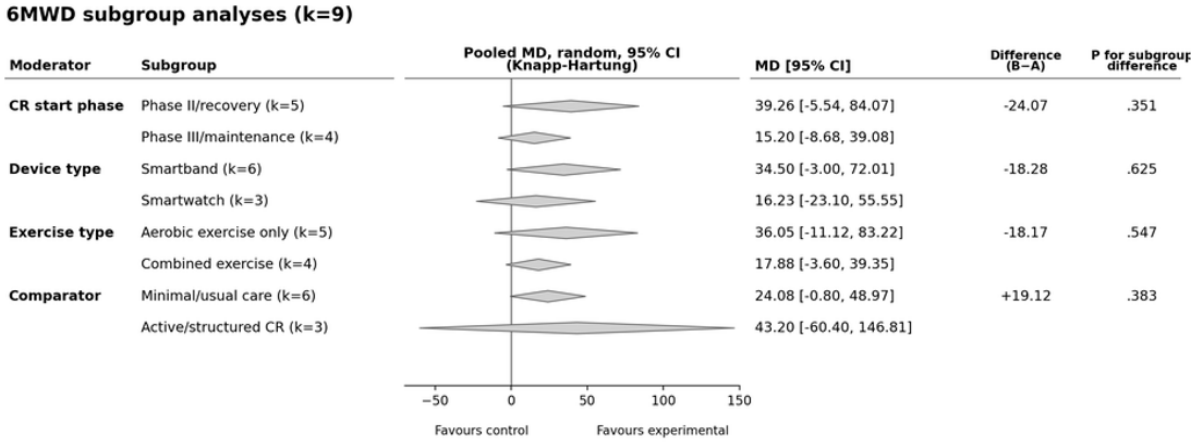


XSL•FO
RenderX

6MWD

Subgroup analyses for 6MWD were conducted according to the CR phase at intervention initiation, wearable-device type, exercise type, and comparator characteristics. With restricted maximum likelihood estimation and Knapp-Hartung-adjusted subgroup CIs, the pooled MD was 39.26 (95% CI –5.54 to 84.07 m; 5 studies) for interventions initiated during phase II and 15.20 (95% CI –8.68 to 39.08 m; 4 studies) for interventions initiated during phase III (interaction: *P*=.35). The pooled MD was 34.50 (95% CI –3.00 to 72.01 m; 6 studies) for smartbands and 16.23 (95% CI –23.10 to 55.56 m; 3 studies) for smartwatches (interaction: *P*=.63). For aerobic exercise alone and combined exercise, the pooled MDs were 36.05 (95% CI –11.12 to 83.22 m; 5 studies) and 17.88 (95% CI –3.60 to 39.35 m; 4 studies), respectively (interaction: *P*=.55). For minimal or usual-care and active or structured CR comparators, the pooled MDs were 24.08 (95% CI –0.80 to 48.97 m; 6 studies) and 43.20 (95% CI –60.40 to 146.81 m; 3 studies), respectively (interaction: *P*=.38). None of the interaction tests were statistically significant (Figure 11).

Figure 11. Subgroup analyses of 6-minute walk distance (6MWD). Random-effects subgroup estimates for 6MWD used restricted maximum likelihood and Knapp-Hartung 95% CIs. No interaction was statistically significant. CR: cardiac rehabilitation; MD: mean difference.



Given the limited number of studies within several subgroups, all subgroup findings were interpreted cautiously. The comparator-based subgroup analyses were conducted post hoc in response to peer review. Prediction intervals were not calculated for subgroup analyses because subgroup-specific estimates were based on a limited number of studies. Study-level results for the peak VO_2 and 6MWD subgroup analyses are provided in Figures S1-S9 in Multimedia Appendix 4.

Meta-Regression Analyses

Peak VO2

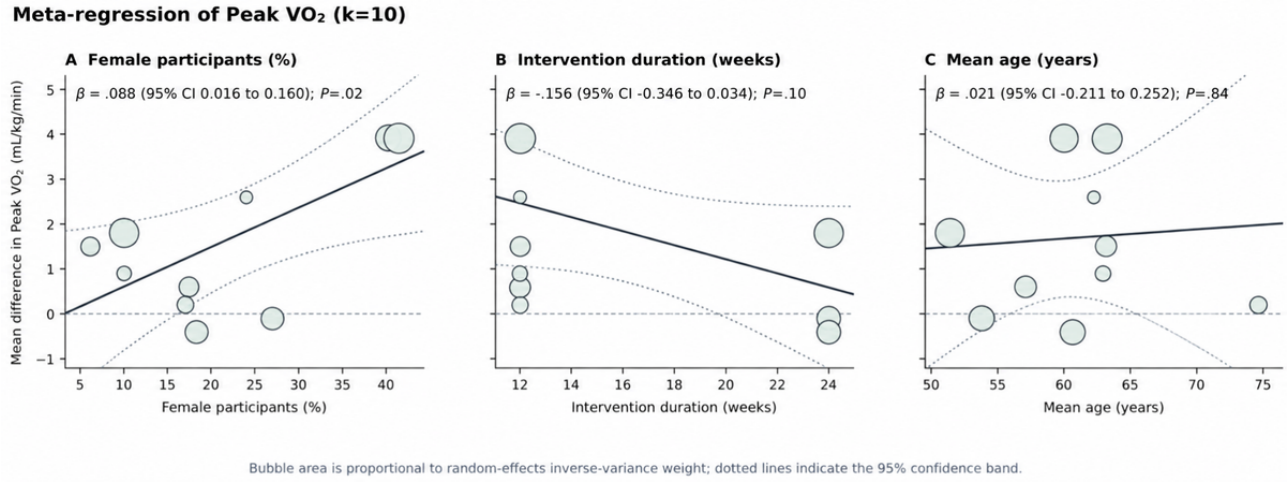
Univariable random-effects meta-regression analyses were conducted across the 10 studies contributing to the peak VO_2 meta-analysis to examine associations with the proportion of female participants, intervention duration, and mean participant age.

A higher proportion of female participants was associated with a larger intervention effect on peak VO_2 (β=.088, 95% CI 0.016-0.160 mL/kg/min per 1-percentage-point increase in the proportion of female participants; *P*=.02). Thus, at the study level, a 10-percentage-point increase in the proportion of female participants corresponded to an estimated 0.88-mL/kg/min larger MD in peak VO_2. This finding should be interpreted cautiously because the analysis was based on aggregate study-level data and does not establish an individual-level sex-specific treatment effect.

Intervention duration showed a negative association with the intervention effect, but the association did not reach the prespecified 2-sided significance threshold (β=–.156, 95% CI –0.346 to 0.034 mL/kg/min per additional week; *P*=.10). Mean participant age was not associated with the intervention effect (β=.021, 95% CI –0.211 to 0.252 mL/kg/min per additional year; *P*=.84). The peak VO_2 meta-regression results are presented in Figure 12.

XSL•FO
RenderX

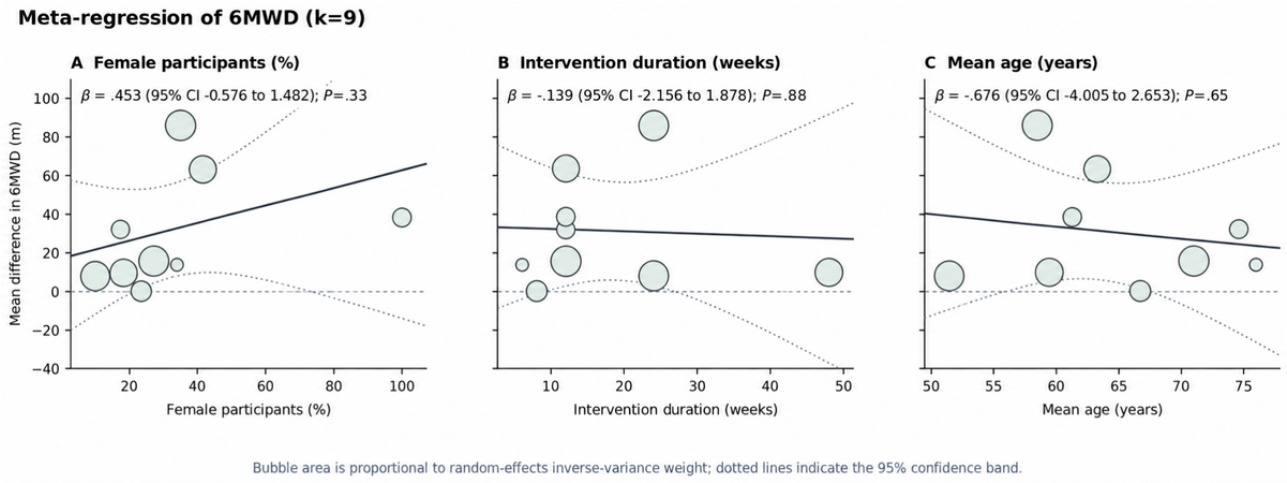
Figure 12. Univariable meta-regression analyses of peak oxygen consumption (VO2). Bubble plots show random-effects meta-regression of peak VO2 mean difference against female proportion, intervention duration, and mean age across 10 studies; area reflects inverse-variance weight; and dotted lines show 95% confidence bands.



6MWD

Univariable random-effects meta-regression analyses were conducted across the 9 studies contributing to the 6MWD meta-analysis. No statistically significant associations were identified for any of the examined study-level moderators. The proportion of female participants was not significantly associated with the intervention effect on 6MWD (β=.453, 95% CI –0.576 to 1.482 m per 1-percentage-point increase; *P*=.33). Intervention duration was also not associated with the effect estimate (β=–.139, 95% CI –2.156 to 1.878 m per additional week; *P*=.88), nor was mean participant age (β=–.676, 95% CI –4.005 to 2.653 m per additional year; *P*=.65). The 6MWD meta-regression results are presented in Figure 13.

Figure 13. Univariable meta-regression analyses of 6-minute walk distance (6MWD). Bubble plots show random-effects meta-regression of 6MWD mean difference against female proportion, intervention duration, and mean age across 9 studies; area reflects inverse-variance weight; and dotted lines show 95% confidence bands.



Given the small number of contributing studies, limited statistical power, and use of aggregate study-level covariates, all meta-regression findings were considered exploratory and hypothesis-generating. In particular, the association between the proportion of female participants and the peak VO_2 effect should not be interpreted as evidence that individual female participants experienced greater benefits.

Sensitivity Analyses

Exclusion of the Cluster-Randomized Trial

Sensitivity analyses excluding the cluster-randomized trial were conducted for 6MWD, daily steps, BMI, and waist circumference, which were the outcomes to which the trial contributed.

For 6MWD, exclusion of the cluster-randomized trial did not materially change the pooled average effect (MD 33.95, 95% CI 7.60-60.30 m; *P*=.02). Between-study variability remained (I^2=88.4%; τ^2=1343.55), and the NNF 95% prediction interval included the null (–67.52 to 135.99 m).

For daily steps, the pooled average effect remained statistically significant after exclusion of the cluster-randomized trial (MD 1084.5, 95% CI 294.6-1961.1 steps/day; *P*=.02). Between-study variability was reflected by I^2=39.9% and τ^2=265,305.58, and the NNF 95% prediction interval included the null (–704.8 to 3077.3 steps/day).

For BMI, exclusion of the cluster-randomized trial resulted in a larger estimated reduction, but the Knapp-Hartung-adjusted

XSL•FO
RenderX

CI included the null (MD –1.18, 95% CI –2.48 to 0.35 kg/m^2; P=.08; I^2=0%; τ^2=0). The NNF 95% prediction interval also included the null (–3.68 to 1.71 kg/m^2).

For waist circumference, exclusion of the cluster-randomized trial moved the point estimate farther from the null (from –2.48 to –3.91 cm), but precision decreased because only 2 studies remained (MD –3.91, 95% CI –8.27 to 0.44 cm; P=.08; I^2=0%). An NNF prediction interval was not reported because only 2 studies remained, below the number for which the method has demonstrated adequate performance. Thus, the estimated effect became larger in magnitude, but the evidence became less precise and remained sensitive to a single-study exclusion.

Leave-One-Out Analysis for 6MWD

A leave-one-out sensitivity analysis was conducted for 6MWD by sequentially excluding each of the 9 contributing studies. The pooled MDs ranged from 20.95 to 34.44 m, remained positive, and retained 95% CIs that excluded the null in every analysis. The corresponding NNF prediction intervals nevertheless included the null throughout.

Exclusion of Madina et al [47] was accompanied by a decrease in I^2 from 87.8% to 42.5%, while the pooled MD remained statistically significant (MD 20.95, 95% CI 3.57-38.36 m). This finding indicates that the relative inconsistency among study results was lower after exclusion of this study; however, the change in I^2 was not interpreted as an absolute reduction in between-study heterogeneity. Exclusion of the other individual studies did not materially reduce I^2, which ranged from 87.2% to 89.4%; however, these relative inconsistency estimates were interpreted alongside $\tau^2$and the prediction intervals rather than as absolute measures of heterogeneity. The complete leave-one-out analysis is presented in Figure S14 in Multimedia Appendix 4.

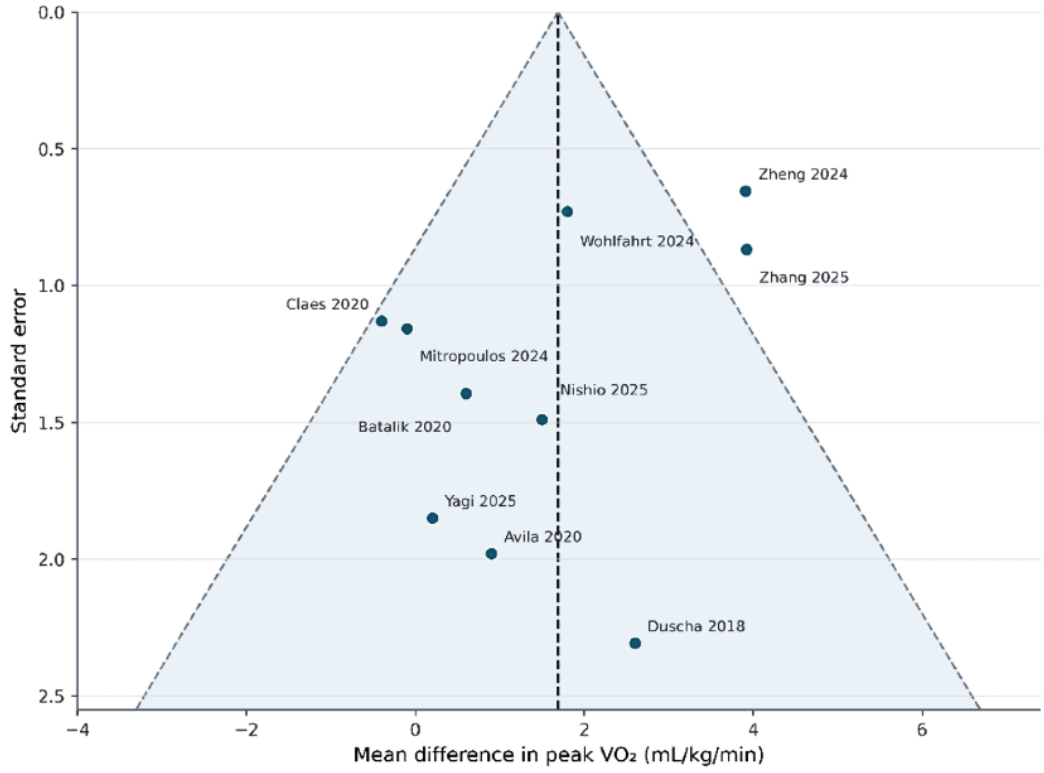
Overall, the sensitivity analyses preserved the direction of the pooled effects for 6MWD, daily steps, and BMI. Knapp-Hartung-adjusted pooled effects remained statistically significant for 6MWD and daily steps after exclusion of the cluster-randomized trial, whereas the BMI estimate included the null. The waist-circumference estimate increased in magnitude but lost precision after exclusion of the cluster-randomized trial. Prediction intervals continued to include the null where estimable.

Reporting Biases

Potential small-study effects were assessed for peak VO_2, for which 10 studies were available. Visual inspection of the funnel plot suggested some asymmetry. Egger regression test yielded an intercept of –2.24 (P=.10), which did not reach the prespecified 2-sided significance threshold of P<.05. Together with the visual asymmetry, this suggested possible small-study effects, although the evidence was inconclusive. This finding should be interpreted cautiously because only 10 studies were available and funnel plot asymmetry may result from between-study heterogeneity, methodological differences, or other factors rather than publication bias alone (Figure 14).

Formal tests for funnel plot asymmetry were not performed for the remaining outcomes because fewer than 10 studies were included in each meta-analysis, resulting in limited statistical power and potentially unreliable assessments.

Figure 14. Funnel plot for the meta-analysis of peak oxygen consumption (VO2). The funnel plot shows peak VO2 mean differences against SEs for 10 randomized controlled trials. The vertical line marks the pooled mean difference, and diagonals mark the expected 95% limits. Mild visual asymmetry was not supported by Egger regression (intercept –2.24, SE 1.19; P=.10) and is not direct evidence of publication bias.



XSL•FO
RenderX

Certainty of Evidence

GRADE certainty ranged from very low to moderate. Peak VO_2 was moderate, downgraded for inconsistency because effects varied and the NNF 95% prediction interval included the null. 6MWD was moderate, downgraded for inconsistency because effects varied materially and the wide NNF interval included the null. Daily steps were low, downgraded 2 levels for very serious risk of bias, mainly from deviations from intended interventions.

AT and BMI were moderate, each downgraded for imprecision. Waist circumference was low, downgraded for risk of bias and imprecision; only 145 participants contributed, the Knapp-Hartung CI and NNF interval included the null, and cluster-RCT exclusion left 2 studies with less precision. Grip strength was low because of risk of bias and imprecision. QoL was very low because of very serious risk of bias, very serious inconsistency, and imprecision. Table S1 in Multimedia Appendix 4 provides detailed judgments.

Discussion

Principal Findings

This systematic review and meta-analysis found favorable average effects of wrist-worn wearable–guided CR delivered through fully remote or hybrid programs on peak VO_2, 6MWD, and daily steps in patients with CVD, with interventions initiated during phase II or phase III CR. No clear pooled effects were found for AT, BMI, waist circumference, grip strength, or QoL after applying a consistent Knapp-Hartung framework. The prediction intervals for all 3 primary outcomes included the null, indicating uncertainty about whether the favorable average effects would be reproduced in a future comparable setting. Certainty varied across outcomes, and findings for daily steps and QoL were further limited by risk-of-bias concerns. Subgroup and meta-regression findings remain exploratory.

The observed average improvements in peak VO_2 and 6MWD may be clinically relevant because both outcomes are associated with cardiovascular prognosis and functional independence. Peak VO_2 is a strong predictor of cardiovascular and all-cause mortality, and modest improvements in cardiorespiratory fitness have been associated with reduced mortality risk [50,51]. The pooled 6MWD estimate was also comparable to commonly reported thresholds for clinically meaningful change in CR populations [52]. These findings are consistent with CR guidelines that identify cardiorespiratory fitness and functional capacity as central treatment targets [3,8]. Nevertheless, the prediction intervals for both outcomes included the null. The pooled estimates should therefore be interpreted as favorable average effects, not effects expected to occur uniformly across future fully remote or hybrid programs.

Subgroup analyses showed numerically larger peak VO_2 improvements when wrist-worn wearable–guided CR was initiated during phase II rather than phase III and when interventions emphasized aerobic exercise rather than combined aerobic and resistance exercise. However, neither formal interaction test reached statistical significance. These findings should therefore be considered hypothesis-generating rather than evidence that CR phase or exercise type modifies intervention effectiveness. The larger numerical effect during phase II may reflect greater potential for physiological recovery during early outpatient rehabilitation, whereas phase III focuses primarily on the maintenance of previously achieved gains [8,53]. Similarly, aerobic training directly targets the central and peripheral determinants of maximal oxygen uptake [54], although resistance exercise remains important for muscular strength and overall function [55,56]. The small number of studies in some subgroups precludes recommending one CR phase or exercise modality over another.

No statistically significant subgroup differences were identified according to wearable-device type, real-time feedback, or comparator characteristics. Most devices shared core functions such as step counting and HR monitoring, which may have limited the ability to detect device-specific differences [57]. Moreover, the effectiveness of feedback may depend on its frequency, personalization, and integration into exercise prescription rather than simply on whether feedback is available [58]. Control interventions also ranged from usual or minimally supported care to structured center-based, community-based, or wearable-supported programs. Consequently, the absence of statistically significant subgroup interactions should not be interpreted as demonstrating equivalence among devices, feedback approaches, or comparator conditions.

Meta-regression associated higher study-level female proportion with greater peak VO_2 improvement, while age and intervention duration were not significant moderators. With few studies and aggregate covariates, these exploratory results are susceptible to ecological bias and do not establish greater individual benefit in women or absence of age- or duration-related effects. Remote or hybrid CR may nevertheless improve access for women facing logistical, caregiving, social, or referral barriers [59], a consideration distinct from the observed treatment-effect association.

The pooled analysis indicated a favorable average effect on daily steps with wrist-worn wearable–guided CR; however, confidence is limited because most contributing studies had high overall risk of bias and evidence certainty was low. The prediction interval included the null, indicating that benefit may not be observed consistently across comparable fully remote or hybrid settings. Wearable- and eHealth-supported rehabilitation may facilitate self-monitoring, goal setting, and feedback [58,60]. However, greater step counts do not necessarily indicate reduced sedentary time, higher activity intensity, or sustained long-term behavioral change because these dimensions were not assessed consistently.

The pooled estimates for AT, BMI, and waist circumference were imprecise and included the null after Knapp-Hartung adjustment. AT may respond differently from peak VO_2 depending on baseline fitness, exercise intensity, and measurement methods [61-63]. The anthropometric point estimates may reflect increased physical activity, as aerobic exercise and sustained activity have been associated with reduced central adiposity [64,65], but most included programs did not provide intensive dietary or weight-management

XSL•FO
RenderX

components. For waist circumference, excluding the cluster-randomized trial moved the point estimate farther from the null but reduced precision because only 2 studies remained. These secondary findings should therefore be considered uncertain and should not be attributed specifically to wearable use.

No clear improvement was observed in grip strength, which may reflect the limited emphasis on progressive resistance training in many wrist-worn wearable–guided CR programs. Step-based activity monitoring alone does not provide the training stimulus required to increase muscular strength [55,56]. The available evidence did not establish a clear effect on QoL. The exploratory pooled estimate was highly uncertain because most contributing studies had high overall risk of bias, between-study heterogeneity was extreme, and the certainty of evidence was rated as very low. Therefore, the nonsignificant pooled estimate should not be interpreted as evidence of no effect. Differences in generic and disease-specific instruments and the broader psychosocial and clinical determinants of QoL may also have contributed to the observed variability [66].

Previous reviews of wearable- and eHealth-supported cardiovascular rehabilitation generally found gains in physical activity or exercise capacity [9,10,16-19]. Alam et al [9] reported that quantitative syntheses tended to favor consumer wearables for steps and moderate to vigorous activity, and Heimer et al [10] found that eHealth-supported phase III CR improved activity and capacity with a small QoL benefit. Our QoL estimate was more uncertain, likely reflecting differences in populations, phases, program components, comparators, follow-up, and instruments. Telerehabilitation can extend CR beyond centers [67], but multicomponent benefits cannot be attributed to the wearable alone.

From a clinical perspective, wrist-worn wearable–guided CR may help extend activity monitoring and behavioral support beyond center-based sessions. This may be particularly relevant because conventional CR participation remains limited by travel distance, transportation, program availability, work and caregiving responsibilities, and physical limitations [8,68,69]. Wrist-worn devices have relatively low user burden and can be incorporated into daily life [70,71]. Their most appropriate role may therefore be to complement structured multidisciplinary CR by supporting individualized goals, activity monitoring, and communication between patients and rehabilitation teams rather than to replace supervised exercise prescription or comprehensive secondary prevention.

This review focused on a clinically relevant and scalable category of wearable technology and included only RCTs. Absolute effect measures enhanced clinical interpretability, and prediction intervals distinguished uncertainty surrounding average effects from the expected variation across comparable settings. Subgroup, meta-regression, and sensitivity analyses were used to explore potential sources of heterogeneity, and companion publications were linked to prevent double counting.

Limitations include material variation in effects, especially for 6MWD and QoL, arising from populations, CR phase, duration, prescriptions, wearable functions, comparators, and timing. I^2 was therefore interpreted with τ^2, prediction intervals, and clinical diversity. Few studies supported several outcomes and subgroups; aggregate meta-regression had limited power and could not establish patient-level modification. Risk-of-bias concerns included deviations, missing data, and outcome measurement. Heterogeneous multicomponent interventions prevented isolating wearable effects, and limited postintervention follow-up precluded conclusions about durability or cardiovascular events. The review lacked prospective registration and a published protocol; some exploratory-analysis eligibility was refined after extraction, and the comparator subgroup was post hoc. Restriction to peer-reviewed English- or Korean-language full texts and exclusion of gray literature may have increased dissemination bias; funnel plots and Egger regression could assess, but not establish, publication bias.

Conclusions

Wrist-worn wearable–guided CR delivered through fully remote or hybrid programs may improve average exercise capacity and physical activity in patients with CVD during phase II or III CR. Favorable pooled effects were observed for peak VO_2, 6MWD, and daily steps; however, the prediction intervals included the null, indicating that these benefits may not be consistently reproduced across future settings. No clear effects were observed for AT, BMI, waist circumference, grip strength, or QoL. When integrated into structured multidisciplinary CR, wrist-worn wearables may provide a scalable means of extending activity monitoring, individualized feedback, and behavioral support beyond supervised clinical encounters. Nevertheless, the limited evidence base, risk-of-bias concerns, intervention heterogeneity, and multicomponent nature of the programs warrant cautious interpretation. Larger, rigorous trials should determine which patients and intervention components derive the greatest benefit and how wearable-derived data can support adaptive, personalized CR.

Acknowledgments

The authors declare the use of generative AI during manuscript revision. ChatGPT (OpenAI) was used under full human supervision to assist with language editing, organization of responses to reviewers, consistency checking across textual and graphical reporting, and code-assisted preparation of revised figures. The tool was not used to make decisions regarding study eligibility, risk-of-bias assessment, Grading of Recommendations Assessment, Development and Evaluation judgments, or the final interpretation of the findings. All AI-assisted outputs, statistical inputs and calculations, text, and figures were critically reviewed and verified by the authors. The authors made all final scientific and editorial decisions and assume full responsibility for the accuracy and integrity of the manuscript. The generative AI tool is not listed as an author and bears no responsibility for the final work.

XSL•FO
RenderX

Funding

This work was supported by National Research Foundation of Korea grants funded by the Korean government (Ministry of Science and Information and Communication Technology; RS-2024-00336696 and RS-2026-25469859). The funder had no involvement in the study design, data collection, analysis, interpretation of the results, or preparation of the manuscript.

Authors' Contributions

Conceptualization: JHC, BRK, ML, S Yoon Lee, HKL
Methodology: JHC, BRK, ML, S Young Lee, KCS, S Yoon Lee, HKL, HR, DH
Formal analysis: JHC, BRK, ML, KCS, HR, DH, JHK, JEK
Investigation: Not reported
Data curation: JHC, BRK, ML, S Young Lee, KCS, HR, DH, JHK, JEK
Software: JHK, JEK
Project administration: BRK, ML
Supervision: S Yoon Lee, HKL
Funding acquisition: BRK
Writing—original draft: JHC
Writing—review & editing: BRK, ML, S Young Lee

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA 2020 expanded checklist.
[DOCX File , 50 KB - mhealth_v14i1e101306_app1.docx]

Multimedia Appendix 2
Complete search strategies.
[DOCX File , 41 KB - mhealth_v14i1e101306_app2.docx]

Multimedia Appendix 3
Summary of outcome-specific Risk of Bias assessments.
[DOCX File , 540 KB - mhealth_v14i1e101306_app3.docx]

Multimedia Appendix 4
Supplementary figures and tables.
[DOCX File , 3929 KB - mhealth_v14i1e101306_app4.docx]

References

1. Chong B, Jayabaskaran J, Jauhari SM, Chan SP, Goh R, Kueh MTW, et al. Global burden of cardiovascular diseases: projections from 2025 to 2050. Eur J Prev Cardiol 2025;32(11):1001-1015 [FREE Full text] [doi: 10.1093/eurjpc/zwae281] [Medline: 39270739]
2. Kazi DS, Elkind MS, Deutsch A, Dowd WN, Heidenreich P, Khavjou O, American Heart Association. Forecasting the economic burden of cardiovascular disease and stroke in the United States through 2050: a presidential advisory from the American Heart Association. Circulation 2024;150(4):e89-e101 [FREE Full text] [doi: 10.1161/CIR.0000000000001258] [Medline: 38832515]
3. Bäck M, Wilhelm M, Marcin T, Machado FVC, Sandberg A, Buccheri S, ESC Guidelines Advisory Group. 2026 ESC guidelines on cardiac rehabilitation. Eur J Prev Cardiol 2026:zwag417 [FREE Full text] [doi: 10.1093/eurjpc/zwag417] [Medline: 42661424]
4. Dibben GO, Faulkner J, Oldridge N, Rees K, Thompson DR, Zwisler A, et al. Exercise-based cardiac rehabilitation for coronary heart disease: a meta-analysis. Eur Heart J 2023;44(6):452-469 [FREE Full text] [doi: 10.1093/eurheartj/ehac747] [Medline: 36746187]
5. Lawler PR, Filion KB, Eisenberg MJ. Efficacy of exercise-based cardiac rehabilitation post-myocardial infarction: a systematic review and meta-analysis of randomized controlled trials. Am Heart J 2011;162(4):571-584.e2. [doi: 10.1016/j.ahj.2011.07.017] [Medline: 21982647]
6. Suaya JA, Shepard DS, Normand ST, Ades PA, Prottas J, Stason WB. Use of cardiac rehabilitation by Medicare beneficiaries after myocardial infarction or coronary bypass surgery. Circulation 2007;116(15):1653-1662. [doi: 10.1161/CIRCULATIONAHA.107.701466] [Medline: 17893274]

XSL•FO
RenderX

7. Ritchey MD, Maresh S, McNeely J, Shaffer T, Jackson SL, Keteyian SJ, et al. Tracking cardiac rehabilitation participation and completion among Medicare beneficiaries to inform the efforts of a national initiative. Circ Cardiovasc Qual Outcomes 2020;13(1):e005902 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.119.005902] [Medline: 31931615]
8. Brown TM, Pack QR, Aberegg E, Brewer LC, Ford YR, Forman DE, American Heart Association Exercise, Cardiac RehabilitationSecondary Prevention Committee of the Council on Clinical Cardiology; Council on CardiovascularStroke Nursing; Council on LifestyleCardiometabolic Health;Council on Quality of CareOutcomes Research. Core components of cardiac rehabilitation programs: 2024 update: a scientific statement from the American Heart Association and the American Association of Cardiovascular and Pulmonary Rehabilitation. Circulation 2024;150(18):e328-e347 [FREE Full text] [doi: 10.1161/CIR.0000000000001289] [Medline: 39315436]
9. Alam S, Zhang M, Harris K, Fletcher LM, Reneker JC. The impact of consumer wearable devices on physical activity and adherence to physical activity in patients with cardiovascular disease: a systematic review of systematic reviews and meta-analyses. Telemed J E Health 2023;29(7):986-1000. [doi: 10.1089/tmj.2022.0280] [Medline: 36445750]
10. Heimer M, Schmitz S, Teschler M, Schäfer H, Douma E, Habibovic M, et al. eHealth for maintenance cardiovascular rehabilitation: a systematic review and meta-analysis. Eur J Prev Cardiol 2023;30(15):1634-1651 [FREE Full text] [doi: 10.1093/eurjpc/zwad145] [Medline: 37154363]
11. Taylor RS, Dalal HM, McDonagh STJ. The role of cardiac rehabilitation in improving cardiovascular outcomes. Nat Rev Cardiol 2022;19(3):180-194 [FREE Full text] [doi: 10.1038/s41569-021-00611-7] [Medline: 34531576]
12. Thomas RJ, Beatty AL, Beckie TM, Brewer LC, Brown TM, Forman DE, et al. Home-based cardiac rehabilitation: a scientific statement from the American Association of Cardiovascular and Pulmonary Rehabilitation, the American Heart Association, and the American College of Cardiology. Circulation 2019;140(1):e69-e89 [FREE Full text] [doi: 10.1161/CIR.0000000000000663] [Medline: 31082266]
13. Rawstorn JC, Gant N, Direito A, Beckmann C, Maddison R. Telehealth exercise-based cardiac rehabilitation: a systematic review and meta-analysis. Heart 2016;102(15):1183-1192. [doi: 10.1136/heartjnl-2015-308966] [Medline: 26936337]
14. Widmer RJ, Collins NM, Collins CS, West CP, Lerman LO, Lerman A. Digital health interventions for the prevention of cardiovascular disease: a systematic review and meta-analysis. Mayo Clin Proc 2015;90(4):469-480 [FREE Full text] [doi: 10.1016/j.mayocp.2014.12.026] [Medline: 25841251]
15. Dunn J, Runge R, Snyder M. Wearables and the medical revolution. Per Med 2018;15(5):429-448 [FREE Full text] [doi: 10.2217/pme-2018-0044] [Medline: 30259801]
16. Hannan AL, Harders MP, Hing W, Climstein M, Coombes JS, Furness J. Impact of wearable physical activity monitoring devices with exercise prescription or advice in the maintenance phase of cardiac rehabilitation: systematic review and meta-analysis. BMC Sports Sci Med Rehabil 2019;11:14 [FREE Full text] [doi: 10.1186/s13102-019-0126-8] [Medline: 31384474]
17. Antoniou V, Davos CH, Kapreli E, Batalik L, Panagiotakos DB, Pepera G. Effectiveness of home-based cardiac rehabilitation, using wearable sensors, as a multicomponent, cutting-edge intervention: a systematic review and meta-analysis. J Clin Med 2022;11(13):3772 [FREE Full text] [doi: 10.3390/jcm11133772] [Medline: 35807055]
18. Yu T, Xu H, Sui X, Zhang X, Pang Y, Yu T, et al. Effectiveness of eHealth interventions on moderate-to-vigorous intensity physical activity among patients in cardiac rehabilitation: systematic review and meta-analysis. J Med Internet Res 2023;25:e42845 [FREE Full text] [doi: 10.2196/42845] [Medline: 36989017]
19. Maximidou T, Mons U. Benefits of wearable-based cardiac rehabilitation interventions in secondary prevention of coronary artery disease - a systematic review and meta-analysis. Am J Prev Cardiol 2025;23:101015 [FREE Full text] [doi: 10.1016/j.ajpc.2025.101015] [Medline: 40585341]
20. Zhang S, Wang Y, Wu J, Ma C, Meng X. Effectiveness of smartwatch device on adherence to home-based cardiac rehabilitation in patients with coronary heart disease: randomized controlled trial. JMIR Mhealth Uhealth 2025;13:e70848 [FREE Full text] [doi: 10.2196/70848] [Medline: 40966536]
21. Zheng Y, Guo J, Tian Y, Qin S, Liu X. Effect of home-based cardiac telerehabilitation in patients after percutaneous coronary intervention: a randomized controlled trial. Comput Inform Nurs 2024;42(12):898-904. [doi: 10.1097/CIN.0000000000001167] [Medline: 39087895]
22. Yagi E, Matsumura K, Uchigashima Y, Shiroyama J, Hase M, Nanba T, et al. Integration of wearables into a cardiac rehabilitation program and its impact on physical activity and exercise capacity in older patients with cardiovascular disease. Adv Rehabil Sci Pract 2025;14:27536351251343538. [doi: 10.1177/27536351251343538] [Medline: 40453481]
23. Mitropoulos A, Anifanti M, Koukouvou G, Ntovoli A, Alexandris K, Kouidi E. Exploring the effects of real-time online cardiac telerehabilitation using wearable devices compared to gym-based cardiac exercise in people with a recent myocardial infarction: a randomised controlled trial. Front Cardiovasc Med 2024;11:1410616 [FREE Full text] [doi: 10.3389/fcvm.2024.1410616] [Medline: 38903965]
24. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
25. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, PRISMA-S Group. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev 2021;10(1):39 [FREE Full text] [doi: 10.1186/s13643-020-01542-z] [Medline: 33499930]

XSL•FO
RenderX

26. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
27. Schünemann HJ, Higgins JPT, Vist GE, Glasziou P, Akl EA, Skoetz N. Chapter 14: completing "summary of findings" tables and grading the certainty of the evidence. In: Cochrane Handbook for Systematic Reviews of Interventions. Chichester: John Wiley & Sons; 2019:375-402.
28. R: a language and environment for statistical computing. The R Core Team. URL: http://www.R-project.org [accessed 2026-09-22]
29. Brini S, Leung TI. Value and credibility of meta-analysis: tutorial on enhancing methodological rigor and AI-powered efficiency. J Med Internet Res 2026;28:e92132 [FREE Full text] [doi: 10.2196/92132] [Medline: 42390911]
30. Higgins JPT, López-López JA. Reflections on the I-squared index for measuring inconsistency in meta-analysis. Res Synth Methods 2026;17(3):389-402. [doi: 10.1017/rsm.2025.10062] [Medline: 41992898]
31. Nagashima K, Noma H, Furukawa T. Prediction intervals for random-effects meta-analysis: a confidence distribution approach. Stat Methods Med Res 2019;28(6):1689-1702. [doi: 10.1177/0962280218773520] [Medline: 29745296]
32. GRADE Working Group. GRADE Handbook for Grading Quality of Evidence and Strength of Recommendations. Lyon, France: The GRADE Working Group; 2013.
33. GRADEpro GDT. GRADEpro Guideline Development Tool Software. Krakow, Poland: McMaster University and Evidence Prime; 2021.
34. Wohlfahrt P, Jenča D, Melenovský V, Mrázková J, Šramko M, Kotrč M, et al. Remote, smart device-based cardiac rehabilitation after myocardial infarction: a pilot, randomized cross-over SmartRehab study. Mayo Clin Proc Digit Health 2024;2(3):352-360 [FREE Full text] [doi: 10.1016/j.mcpdig.2024.06.001] [Medline: 40206122]
35. Lahtio H, Heinonen A, Paajanen T, Sjögren T. The added value of remote technology in cardiac rehabilitation on physical function, anthropometrics, and quality of life: cluster randomized controlled trial. J Med Internet Res 2023;25:e42455 [FREE Full text] [doi: 10.2196/42455] [Medline: 37043264]
36. Beckie TM, Sengupta A, Dey AK, Dutta K, Ji M, Chellappan S. A mobile health behavior change intervention for women with coronary heart disease: a randomized controlled pilot study. J Cardiopulm Rehabil Prev 2024;44(1):40-48. [doi: 10.1097/HCR.0000000000000804] [Medline: 37285601]
37. Lindman BR, Gillam LD, Coylewright M, Welt FGP, Elmariah S, Smith SA, et al. Effect of a pragmatic home-based mobile health exercise intervention after transcatheter aortic valve replacement: a randomized pilot trial. Eur Heart J Digit Health 2021;2(1):90-103 [FREE Full text] [doi: 10.1093/ehjdh/ztab007] [Medline: 34048509]
38. Nishio R, Dohi T, Yokoyama M, Nakade T, Takahashi N, Chikata Y, et al. Wearable devices in remote cardiac rehabilitation with and without weekly online coaching for patients with coronary artery disease: randomized controlled trial. JMIR Mhealth Uhealth 2025;13:e63797 [FREE Full text] [doi: 10.2196/63797] [Medline: 40395066]
39. Claes J, Cornelissen V, McDermott C, Moyna N, Pattyn N, Cornelis N, et al. Feasibility, acceptability, and clinical effectiveness of a technology-enabled cardiac rehabilitation platform (Physical Activity Toward Health-I): randomized controlled trial. J Med Internet Res 2020;22(2):e14221 [FREE Full text] [doi: 10.2196/14221] [Medline: 32014842]
40. Chokshi NP, Adusumalli S, Small DS, Morris A, Feingold J, Ha YP, et al. Loss-framed financial incentives and personalized goal-setting to increase physical activity among ischemic heart disease patients using wearable devices: the ACTIVE REWARD randomized trial. J Am Heart Assoc 2018;7(12):e009173 [FREE Full text] [doi: 10.1161/JAHA.118.009173] [Medline: 29899015]
41. Batalik L, Dosbaba F, Hartman M, Batalikova K, Spinar J. Benefits and effectiveness of using a wrist heart rate monitor as a telerehabilitation device in cardiac patients: a randomized controlled trial. Medicine (Baltimore) 2020;99(11):e19556 [FREE Full text] [doi: 10.1097/MD.0000000000019556] [Medline: 32176113]
42. Avila A, Claes J, Buys R, Azzawi M, Vanhees L, Cornelissen V. Home-based exercise with telemonitoring guidance in patients with coronary artery disease: does it improve long-term physical fitness? Eur J Prev Cardiol 2020;27(4):367-377 [FREE Full text] [doi: 10.1177/2047487319892201] [Medline: 31787026]
43. Park LG, Elnaggar A, Lee SJ, Merek S, Hoffmann TJ, Von Oppenfeld J, et al. Mobile health intervention promoting physical activity in adults post cardiac rehabilitation: pilot randomized controlled trial. JMIR Form Res 2021;5(4):e20468 [FREE Full text] [doi: 10.2196/20468] [Medline: 33861204]
44. Duscha BD, Piner LW, Patel MP, Craig KP, Brady M, McGarrah RW, et al. Effects of a 12-week mHealth program on peak VO and physical activity patterns after completing cardiac rehabilitation: a randomized controlled trial. Am Heart J 2018;199:105-114. [doi: 10.1016/j.ahj.2018.02.001] [Medline: 29754647]
45. Hakala S, Kivistö H, Paajanen T, Kankainen A, Anttila M, Heinonen A, et al. Effectiveness of distance technology in promoting physical activity in cardiovascular disease rehabilitation: cluster randomized controlled trial, a pilot study. JMIR Rehabil Assist Technol 2021;8(2):e20299 [FREE Full text] [doi: 10.2196/20299] [Medline: 34142970]
46. Dodson JA, Adhikari S, Schoenthaler A, Hochman JS, Sweeney G, George B, et al. Rehabilitation at home using mobile health for older adults hospitalized for ischemic heart disease: the RESILIENT Randomized clinical trial. JAMA Netw Open 2025;8(1):e2453499 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.53499] [Medline: 39775808]
47. Iqbal SMA, Mahgoub I, Du E, Leavitt MA, Asghar W. Advances in healthcare wearable devices. NPJ Flex Electron 2021;5(1):504-510. [doi: 10.1038/s41528-021-00107-x]

XSL•FO
RenderX

48. Ntovoli A, Mitropoulos A, Anifanti M, Koukouvou G, Kouidi E, Alexandris K. Can online exercise using wearable devices improve perceived well-being? A study among patients with coronary artery disease. Sensors (Basel) 2025;25(3):698 [FREE Full text] [doi: 10.3390/s25030698] [Medline: 39943336]
49. Brouwers RWM, van der Poort EKJ, Kemps HMC, van den Akker-van Marle ME, Kraal JJ. Cost-effectiveness of cardiac telerehabilitation with relapse prevention for the treatment of patients with coronary artery disease in the Netherlands. JAMA Netw Open 2021;4(12):e2136652 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.36652] [Medline: 34854907]
50. Ross R, Blair SN, Arena R, Church TS, Després JP, Franklin BA, American Heart Association Physical Activity Committee of the Council on Lifestyle Cardiometabolic Health, Council on Clinical Cardiology, Council on Epidemiology Prevention, Council on Cardiovascular Stroke Nursing, Council on Functional Genomics Translational Biology, Stroke Council. Importance of assessing cardiorespiratory fitness in clinical practice: a case for fitness as a clinical vital sign: a scientific statement from the American Heart Association. Circulation 2016;134(24):e653-e699. [doi: 10.1161/CIR.0000000000000461] [Medline: 27881567]
51. Kodama S, Saito K, Tanaka S, Maki M, Yachi Y, Asumi M, et al. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis. JAMA 2009;301(19):2024-2035. [doi: 10.1001/jama.2009.681] [Medline: 19454641]
52. Gremeaux V, Troisgros O, Benaïm S, Hannequin A, Laurent Y, Casillas J, et al. Determining the minimal clinically important difference for the six-minute walk test and the 200-meter fast-walk test during cardiac rehabilitation program in coronary artery disease patients after acute coronary syndrome. Arch Phys Med Rehabil 2011;92(4):611-619 [FREE Full text] [doi: 10.1016/j.apmr.2010.11.023] [Medline: 21440707]
53. Ades PA. Cardiac rehabilitation and secondary prevention of coronary heart disease. N Engl J Med 2001;345(12):892-902. [doi: 10.1056/NEJMra001529] [Medline: 11565523]
54. Montero D, Diaz-Cañestro C, Lundby C. Endurance training and VO2max: role of maximal cardiac output and oxygen extraction. Med Sci Sports Exerc 2015;47(10):2024-2033. [doi: 10.1249/MSS.0000000000000640] [Medline: 25680086]
55. Bassett DR, Howley ET. Limiting factors for maximum oxygen uptake and determinants of endurance performance. Med Sci Sports Exerc 2000;32(1):70-84. [doi: 10.1097/00005768-200001000-00012] [Medline: 10647532]
56. Kirkman DL, Lee D, Carbone S. Resistance exercise for cardiac rehabilitation. Prog Cardiovasc Dis 2022;70:66-72 [FREE Full text] [doi: 10.1016/j.pcad.2022.01.004] [Medline: 35122871]
57. Iqbal SMA, Mahgoub I, Du E, Leavitt MA, Asghar W. Advances in healthcare wearable devices. NPJ Flex Electron 2021;5(1):9. [doi: 10.1038/s41528-021-00107-x]
58. Brickwood KJ, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019;7(4):e11819 [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]
59. Resurrección DM, Motrico E, Rigabert A, Rubio-Valera M, Conejo-Cerón S, Pastor L, et al. Barriers for nonparticipation and dropout of women in cardiac rehabilitation programs: a systematic review. J Womens Health (Larchmt) 2017;26(8):849-859. [doi: 10.1089/jwh.2016.6249] [Medline: 28388314]
60. Jakicic JM, Davis KK, Rogers RJ, King WC, Marcus MD, Helsel D, et al. Effect of wearable technology combined with a lifestyle intervention on long-term weight loss: the IDEA randomized clinical trial. JAMA 2016;316(11):1161-1171 [FREE Full text] [doi: 10.1001/jama.2016.12858] [Medline: 27654602]
61. Balady GJ, Arena R, Sietsema K, Myers J, Coke L, Fletcher GF, American Heart Association Exercise, Cardiac Rehabilitation,Prevention Committee of the Council on Clinical Cardiology, Council on EpidemiologyPrevention, Council on Peripheral Vascular Disease, Interdisciplinary Council on Quality of CareOutcomes Research. Clinician's guide to cardiopulmonary exercise testing in adults: a scientific statement from the American Heart Association. Circulation 2010;122(2):191-225. [doi: 10.1161/CIR.0b013e3181e52e69] [Medline: 20585013]
62. Jones AM, Carter H. The effect of endurance training on parameters of aerobic fitness. Sports Med 2000;29(6):373-386. [doi: 10.2165/00007256-200029060-00001] [Medline: 10870864]
63. Bensimhon DR, Leifer ES, Ellis SJ, Fleg JL, Keteyian SJ, Piña IL, HF-ACTION Trial Investigators. Reproducibility of peak oxygen uptake and other cardiopulmonary exercise testing parameters in patients with heart failure (from the heart failure and a controlled trial investigating outcomes of exercise traiNing). Am J Cardiol 2008;102(6):712-717 [FREE Full text] [doi: 10.1016/j.amjcard.2008.04.047] [Medline: 18773994]
64. Armstrong A, Jungbluth Rodriguez K, Sabag A, Mavros Y, Parker HM, Keating SE, et al. Effect of aerobic exercise on waist circumference in adults with overweight or obesity: a systematic review and meta-analysis. Obes Rev 2022;23(8):e13446 [FREE Full text] [doi: 10.1111/obr.13446] [Medline: 35383401]
65. Waller K, Kaprio J, Kujala UM. Associations between long-term physical activity, waist circumference and weight gain: a 30-year longitudinal twin study. Int J Obes (Lond) 2008;32(2):353-361. [doi: 10.1038/sj.ijo.0803692] [Medline: 17653065]
66. Anderson L, Thompson DR, Oldridge N, Zwisler A, Rees K, Martin N, et al. Exercise-based cardiac rehabilitation for coronary heart disease. Cochrane Database Syst Rev 2016;2016(1):CD001800 [FREE Full text] [doi: 10.1002/14651858.CD001800.pub3] [Medline: 26730878]

XSL•FO
RenderX

67. Frederix I, Solmi F, Piepoli M, Dendale P. Cardiac telerehabilitation: a novel cost-efficient care delivery strategy that can induce long-term health benefits. Eur J Prev Cardiol 2017;24(16):1708-1717 [FREE Full text] [doi: 10.1177/2047487317732274] [Medline: 28925749]
68. Turk-Adawi K, Supervia M, Lopez-Jimenez F, Pesah E, Ding R, Britto RR, et al. Cardiac rehabilitation availability and density around the globe. EClinicalMedicine 2019;13:31-45 [FREE Full text] [doi: 10.1016/j.eclinm.2019.06.007] [Medline: 31517261]
69. Ruano-Ravina A, Pena-Gil C, Abu-Assi E, Raposeiras S, van 't Hof A, Meindersma E, et al. Participation and adherence to cardiac rehabilitation programs. A systematic review. Int J Cardiol 2016;223:436-443. [doi: 10.1016/j.ijcard.2016.08.120] [Medline: 27557484]
70. Schweizer T, Gilgen-Ammann R. Wrist-worn and arm-worn wearables for monitoring heart rate during sedentary and light-to-vigorous physical activities: device validation study. JMIR Cardio 2025;9:e67110 [FREE Full text] [doi: 10.2196/67110] [Medline: 40116771]
71. Müller AM, Wang NX, Yao J, Tan CS, Low ICC, Lim N, et al. Heart rate measures from wrist-worn activity trackers in a laboratory and free-living setting: validation study. JMIR Mhealth Uhealth 2019;7(10):e14120 [FREE Full text] [doi: 10.2196/14120] [Medline: 31579026]

Abbreviations

6MWD: 6-minute walk distance
AT: anaerobic threshold
CR: cardiac rehabilitation
CVD: cardiovascular disease
GRADE: Grading of Recommendations Assessment, Development and Evaluation
HR: heart rate
MD: mean difference
PCI: percutaneous coronary intervention
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Literature
PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Literature Search Extension
QoL: quality of life
RCT: randomized controlled trial
RoB 2: Cochrane Risk of Bias 2
SF-36: 36-Item Short Form Health Survey
SMD: standardized mean difference
VO2: oxygen consumption

Edited by S Brini; submitted 29.May.2026; peer-reviewed by P Liu, B Schmitz; comments to author 24.Jul.2026; revised version received 20.Sep.2026; accepted 21.Sep.2026; published 27.Oct.2026.

Please cite as:
Choi JH, Lee SY, Seo KC, Lee SY, Lee HK, Heo D, Ryu H, Kim JH, Kwon JE, Kim BR, Lee MY
Effectiveness of Wrist-Worn Wearable–Guided Cardiac Rehabilitation Delivered Through Fully Remote or Hybrid Programs in Patients With Cardiovascular Disease: Systematic Review and Meta-Analysis of Randomized Controlled Trials
JMIR Mhealth Uhealth 2026;14:e101306
URL: https://mhealth.jmir.org/2026/1/e101306
doi:10.2196/101306
PMID:

©Jun Hwan Choi, So Young Lee, Kyung Cheon Seo, Sang Yoon Lee, Hyo Kyung Lee, Dongbin Heo, Hyun Ryu, Jong Hoon Kim, Jeong Eul Kwon, Bo Ryun Kim, Michael Y Lee. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 27.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Measuring Perceived Stress in Healthy Adults in Daily Life: Scoping Review of Smartphone-Based Self-Assessment

Toon Colman[1], MSc; Tom Van Daele[1,2], PhD; Bert Bonroy[3], PhD; Jean-Marie Aerts[4], PhD; Glen Debard[3], PhD; Jasper Gielen[4], PhD; Romy Sels[3], MSc; Catharina Nina Van Oost[4], MSc; Maria Faurholt-Jepsen[5,6], MD, DMSc; Nele A J De Witte[1], PhD

[1]Psychology & Technology, Centre of Expertise Care and Well-being, Thomas More University of Applied Sciences, Antwerp, Belgium

[2]Centre for Technological Innovation, Mental Health and Education, Queen's University Belfast, Belfast, Northern Ireland, United Kingdom

[3]Mobilab & Care, Centre of Expertise Care and Well-being, Thomas More University of Applied Sciences, Geel, Belgium

[4]Department of Biosystems, Research group M3-BIORES, KU Leuven, Leuven, Belgium

[5]Copenhagen Affective Disorder Research Center (CADIC), Psychiatric Center Copenhagen, Frederiksberg, Denmark

[6]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

Corresponding Author:
Toon Colman, MSc
Psychology & Technology, Centre of Expertise Care and Well-being
Thomas More University of Applied Sciences
Molenstraat 8
Antwerp, 2018
Belgium
Phone: 32 32410824
Email: toon.colman@thomasmore.be

Abstract

Background: Smartphone-based self-assessment methods, such as ecological momentary assessment (EMA), are increasingly used to measure perceived stress in research and emerging clinical applications. However, the proliferation of studies and methods has made it difficult for researchers and practitioners to maintain an overview and make informed methodological choices. An updated synthesis of smartphone-based self-report methods is therefore needed.

Objective: This scoping review aims to systematically map smartphone-based methods for assessing self-reported perceived stress in healthy adults during daily life, to describe measurement characteristics and methodological choices, and to narratively examine the extent to which validity and reliability evidence are reported.

Methods: Following PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines, we searched a total of 7 databases: Web of Science, Embase, MEDLINE, PubMed, IEEE Xplore, ACM, and PsycInfo. Eligible studies assessed general perceived stress in healthy adults using smartphone-based self-report measures in daily life. Screening was conducted independently by at least 2 reviewers at the abstract and full-text levels. Methodological characteristics of stress measurement were extracted and summarized descriptively.

Results: From 8777 deduplicated records, 108 studies met all eligibility criteria. Publication frequency increased notably over the review period. Most studies used observational designs, with EMA-style prompting schedules. Stress was typically assessed using very brief measures: most instruments (84/114, 74%) consisted of a single item, most commonly answered using a Likert-type scale or a visual analog scale. Measurement frequency ranged from once per week to multiple prompts per day, with intensive approaches (≥3 prompts per day) being most common. Despite the prevalence of these instruments, only a minority of studies reported validity or reliability measures. When provided, evidence typically consisted of correlations with established stress questionnaires or affective measures. A small subset conducted elaborate psychometric analyses or validation of instruments.

Conclusions: Smartphone-based stress assessment is widespread but methodologically heterogeneous. This review provides an updated methodological overview to support researchers in selecting or developing robust smartphone-based measures of perceived stress. Only a small minority of studies reported formal psychometric evaluations, highlighting the need for more systematic validation of smartphone-based stress assessment.

Trial Registration: OSF Registries 7UM6B; https://osf.io/7um6b

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2026;14:e94204) doi:10.2196/94204

KEYWORDS

smartphone; ecological momentary assessment; psychological stress; self-report; mobile apps; surveys and questionnaires

Introduction

Background

Stress is a multifaceted psychological and physiological response that plays a central role in mental and physical health. According to Cohen et al [1], stress can be described as the process in which "environmental demands tax or exceed the adaptive capacity of an organism, resulting in psychological and biological changes that may place persons at risk for disease." Chronic stress, without sufficient recovery, is indeed associated with numerous adverse outcomes, including infectious disease [2], cardiovascular disease [3], and depression [4], as well as substantial economic costs [5]. Hence, improving stress monitoring and interventions can have a substantial personal and societal impact.

Being a complex response, stress involves changes in autonomic nervous system activity, behavior, and subjective mood. Physiologically, it is associated with decreases in heart rate variability (HRV) [6], increases in electrodermal activity [7], and the release of stress-related hormones such as cortisol [8]. Equally important, however, is the psychological response: the subjective experience of feeling stressed, occurring when individuals perceive that demands exceed their resources [9]. Physiological and psychological stress measures are related, but they capture distinct processes on different timescales and, therefore, their correlation is often modest [10]. Physiological responses are also not stress-specific and are influenced by contextual variables, such as physical movement or high-arousal positive emotions [11]. Therefore, the accurate measurement of self-reported perceived stress (or subjective stress) remains crucial.

Several questionnaires assess perceived stress, including the Perceived Stress Scale (PSS) [12], the stress subscale of the Depression Anxiety Stress Scales (DASS) [13], and the Kessler Psychological Distress Scale (K10) [14]. Although these retrospective measures are widely used and psychometrically supported [15-17], they are limited by recall bias and low temporal resolution: they cannot capture rapid fluctuations in stress throughout daily life [18]. Ecological momentary assessment (EMA), also termed the experience sampling method (ESM), addresses these limitations by collecting repeated real-time reports of thoughts, emotions, and behaviors in natural environments, enhancing ecological validity and reducing recall-related bias [18]. With the widespread use of smartphones, EMA has become more feasible and scalable, expanding the possibilities for high-resolution monitoring of perceived stress [19-22]. Smartphone-based EMA methods have also been used to measure stress in patient populations, demonstrating the opportunities in clinical contexts [23]. Examples of populations that have been studied include patients with HIV [24] or tinnitus [25].

EMA and similar methods allow researchers and practitioners to collect self-reported stress data in real time, but several methodological challenges remain. EMA protocols rely on brief assessments that are administered repeatedly, often multiple times per day. To minimize participant burden and increase compliance, single-item measures or abbreviated versions of existing questionnaires are often used [20]. While the validity and reliability of brief EMA measures are generally supported, results vary across participants, topics, and studies [26,27]. In addition, when traditional (retrospective) questionnaires are shortened and adapted to a momentary assessment context, they may not automatically retain their psychometric properties [28]. The validity and reliability of EMA stress measures are a multifaceted problem. First, validity can be examined by studying whether the EMA measure correlates with a validated questionnaire, such as the PSS (convergent validity), and by investigating whether it correlates as theoretically expected with distinct but related constructs, such as positive and negative affect (criterion and discriminant validity) [28,29]. Second, the reliability of EMA measures can be investigated using multilevel approaches, as there is variability not only between participants but also within participants (due to repeated measurements) [28,29].

A systematic review by Þórarinsdóttir et al [20] showed that by 2016, smartphone-based self-assessment of stress had already proliferated, driven by widespread smartphone availability and growing interest in personalized health monitoring. Their review revealed heterogeneity in methodological choices such as scale type, number of items, and sampling frequency, and highlighted that many studies relied on unvalidated brief instruments. Consequently, to date, important questions remain regarding the validity and reliability of these measures.

Study Objectives

It is likely that over the past decade, since the review by Þórarinsdóttir et al [20], the scientific literature on smartphone-based self-assessment of stress has grown even more steeply, and more validated approaches have been developed. The lack of a recent comprehensive overview of smartphone-based methods for assessing perceived stress poses challenges for researchers seeking to make informed methodological decisions.

Thus, the present scoping review systematically maps the methods used to measure self-reported perceived stress via smartphones in healthy adults during daily life. Here, we focus on general perceived (or subjective) stress, as opposed to stress associated with specific events or domains, such as stress related to traumatic life events. Because the field is characterized by considerable methodological diversity in instrument content, response formats, sampling schedules, and implementation strategies, a systematic review is currently not feasible. Instead, a scoping review is the most appropriate way to map the existing evidence base. We aim to update the findings of Þórarinsdóttir

XSL•FO
RenderX

et al [20] by synthesizing the methodological characteristics of recent approaches and exploring the extent to which their validity and reliability have been studied.

Specifically, this review addressed the following research questions:

1. Which smartphone-based self-report methods have been used to assess perceived stress in healthy adults during daily life since 2016?
2. What are the methodological characteristics of smartphone-based stress assessment, including study characteristics (eg, study duration, sample size, population, and assessment context) and measurement characteristics (eg, item phrasing, number of items, response format, measurement frequency, and prompting method)?
3. What evidence regarding the validity and reliability of these instruments has been reported?

Methods

Review Framework and Protocol Registration

This scoping review was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 statement [30] and the PRISMA extension for Scoping Reviews (PRISMA-ScR) [31]. The tip sheets accompanying the PRISMA-ScR statement were also consulted. The PRISMA-ScR checklist can be found in Multimedia Appendix 1 [31]. The protocol was developed following the methodological guidelines of the Joanna Briggs Institute Scoping Review Methodology Group [32] and was preregistered on OSF Registries [33]. Compared with the preregistered protocol, only 2 minor deviations occurred. First, the ACM Guide to Computing Literature was searched instead of the ACM Full Text Collection. Because the ACM Guide to Computing Literature includes all records from the ACM Full Text Collection as well as additional records, this modification resulted in a more comprehensive search. Second, although participant age was extracted from included studies as planned, age data were not synthesized using descriptive statistics. Age was reported inconsistently across studies, with some reporting means and SDs, others reporting medians, and others reporting age ranges only. Because these data could not be summarized in a consistent manner across studies, age was not included in the final quantitative summary.

Eligibility Criteria

Following the population, concept, and context framework of the Joanna Briggs Institute [34], studies were eligible when they targeted healthy adults (population), assessed general perceived stress (concept), and used a smartphone-based self-assessment method administered during daily life (context). The population of healthy adults was defined as individuals aged 18 years or older without diagnosed physical or mental health conditions and without indications of unhealthy living circumstances (eg, homelessness), as outlined in the World Health Organization Housing and Health Guidelines [35]. To update the earlier review by Þórarinsdóttir et al [20], only records published between January 1, 2016, and the search date of June 14, 2024, were considered.

Studies were excluded if they met any of the following criteria: (1) wrong publication type (eg, nonempirical report such as a protocol or not written in English), (2) wrong population (eg, minors, unhealthy adults, or unhealthy housing conditions), (3) no measurement of general perceived stress levels in daily life (eg, stress not quantified, not general but specific stress, or not in daily life), (4) not a smartphone-based assessment (eg, stress not measured on a smartphone), (5) duplicate data (eg, in multiple publications reporting the same dataset, only the most complete version was retained, ie, the study with the most participants or the longest duration if sample size was equal), and (6) missing data on one or more of the main variables, preventing extraction of core characteristics (eg, stress item phrasing or sampling frequency).

Information Sources and Search Strategy

We searched generalist and domain-specific databases in psychology and technology. Searches were conducted in Web of Science Core Collection (via the Clarivate Web of Science platform), Embase and MEDLINE (via Embase), PubMed (via PubMed), PsycInfo (via Ovid), IEEE Xplore Digital Library (via IEEE Xplore), and the ACM Guide to Computing Literature (via ACM Digital Library). Because PubMed also includes all MEDLINE-indexed records, there was overlap between the searches conducted through Embase and PubMed, which was addressed during the deduplication process. However, PubMed also includes records that are not (yet) indexed in MEDLINE, thereby increasing the sensitivity of the search strategy.

No gray literature (eg, master's theses) was included because the review aimed to identify smartphone-based stress assessment methods that were described in sufficient methodological detail to allow extraction of key measurement characteristics and any available evidence regarding validity or reliability. No supplementary reference list searching, citation tracking, hand-searching, or expert consultation was performed.

The general search query was *(Stress* OR psychological stress OR emotional stress) AND (smartphone* OR "cell phone*" OR cellphone* OR "cellular phone*" OR mobile phone* OR iPhone* OR iOS OR Android OR mobile application* OR mobile app OR mobile apps OR ecological momentary assessment* OR experience sampling method* OR ESM OR EMA)*. Some databases required adaptation of the query to accommodate their specific search fields and operators. Prior to finalizing the strategy, we piloted the search string in all databases and refined it iteratively to maximize sensitivity (inclusion of all relevant studies) while maintaining sufficient specificity (limiting irrelevant results).

As an example, the full search strategy for PubMed was to enter the following search query in the basic search field: *(Stress*[Title/Abstract] OR "psychological stress"[Title/Abstract:~5] OR "emotional stress"[Title/Abstract:~5]) AND (smartphone*[Title/Abstract] OR "cell phone"[Title/Abstract:~5] OR "cell phones"[Title/Abstract:~5] OR cellphone*[Title/Abstract] OR "cellular phone"[Title/Abstract:~5] OR "cellular phones"[Title/Abstract:~5] OR "mobile phone"[Title/Abstract: ~5] OR "mobile phones"[Title/Abstract:~5] OR iPhone*[Title/Abstract] OR iOS[Title/Abstract] OR Android[Title/Abstract] OR "mobile*

XSL•FO
RenderX

application"[Title/Abstract:~5] OR "mobile applications"[Title/Abstract: ~5] OR "mobile app"[Title/Abstract:~5] OR "mobile apps"[Title/Abstract:~5] OR "ecological momentary assessment"[Title/Abstract] OR "experience sampling method*"[Title/Abstract] OR ESM[Title/Abstract] OR EMA[Title/Abstract])*. After conducting the search, the publication date filter was set from January 1, 2016, to June 14, 2024. All other database-specific search queries are available in the preregistered protocol [33].

Screening and Selection Process

For the screening and selection of studies, all retrieved records were imported into Rayyan (Rayyan Systems, Inc) [36] for deduplication and screening. After removing duplicate entries, the remaining unique records were first screened at the abstract level using the predefined inclusion and exclusion criteria. Abstract screening was conducted independently by 2 reviewers, who were blinded to each other's decisions. Records deemed potentially eligible proceeded to full-text screening, which was carried out using the same blinded dual-reviewer procedure. After both the abstract and full-text stages, any conflicts between reviewers were resolved through discussion until consensus was reached. Reviewers had an academic and professional background in psychology.

Data Extraction

Overview

The primary aim of the extraction process was to capture the methodological characteristics of smartphone-based stress measurement. Data extraction was first piloted on a subset of 10 studies by 1 reviewer, after which the variables to be extracted were discussed with a second reviewer and refined accordingly. After finalizing the extraction approach, 1 reviewer extracted data from all included studies, after which a second independent reviewer checked the dataset. Any discrepancies or irregularities identified during this check were resolved through discussion.

Study Characteristics

To describe the characteristics of studies on smartphone-based self-assessment of stress, several variables were extracted. The main variables were (1) study population (eg, university students or employees), (2) assessment context (eg, daily life or workplace), and (3) study duration. Studies were required to report these variables to be eligible for inclusion. Additional variables that were extracted but were less essential were (1) study location, (2) study design (eg, observational cohort or interventional), and (3) number of participants. Participant age was also extracted but not retained in the final quantitative summary because of inconsistencies in reporting formats across studies.

Characteristics of Stress Self-Assessment Methods

To characterize the measurement characteristics of smartphone-based self-assessment of stress, several variables were extracted. The main variables were (1) the number of items measuring stress, (2) the stress items, (3) the method for self-assessment (eg, Likert scale or binary yes-or-no response), (4) the number of times per day that stress was measured, and (5) mobile modality and prompting method (eg, measurement via text message or app notification). Studies were required to report these variables to be eligible for inclusion. An additional variable extracted was the smartphone operating system.

Evidence on Validity and Reliability

For each included study, we documented whether evidence regarding the validity or reliability of the stress measure was reported. When available, information was extracted on the type of evidence provided, the procedures used to establish validity or reliability, and any reported psychometric indexes.

Analysis Plan

The extracted data were summarized using descriptive statistics to provide an overview of study characteristics and characteristics of smartphone-based self-assessment methods. In line with scoping review methodology, neither a formal quality assessment nor a meta-analysis was undertaken. The evidence on validity and reliability was not formally appraised but instead synthesized narratively, due to the exploratory nature of the review and the limited reporting of psychometric properties. Given the wide range of reporting practices, from no evidence to a comprehensive psychometric evaluation, it was not feasible to quantify validity and reliability evidence systematically. Therefore, we provide a narrative overview of the strategies used to establish validity and reliability in the included studies and highlight illustrative examples.

All data processing and analysis were conducted using R (version 4.3.0; R Foundation for Statistical Computing) [37], along with the packages *readxl* [38], *writexl* [39], *officer* [40], and *flextable* [41]. To enhance comparability across studies, some variables were harmonized by reducing the level of detail when appropriate, such as by binning numeric variables into categories. Study duration was converted to an estimated number of days when reported in alternative units (eg, week, month, and year), using standard conversions (7, 30, and 365 days).

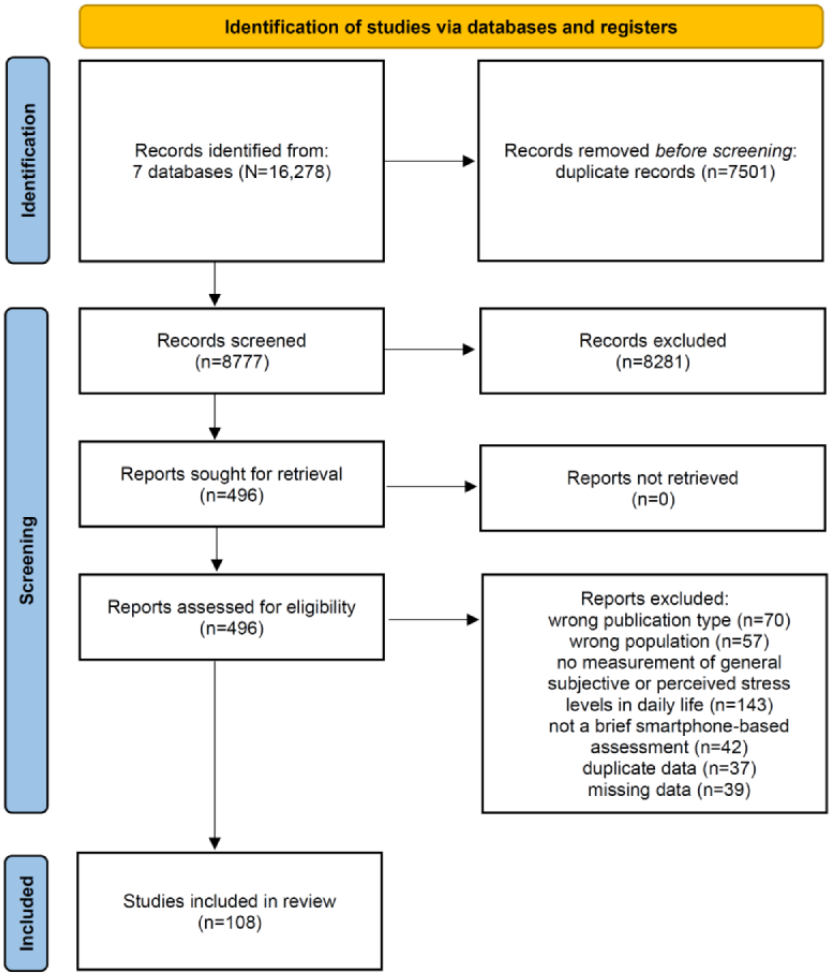
Results

Overview

After deduplication, 8777 records were screened. Of these, 496 progressed to full-text review, and 108 studies met all eligibility criteria and were included in the final synthesis. A PRISMA flow diagram of the screening and selection process is presented in Figure 1. The data table presenting all included studies can be found in Table S1 in Multimedia Appendix 2 [28,29,42-147]. The same table is included in Excel format with sorting and filtering functionalities in Multimedia Appendix 3 [28,29,42-147], to aid the closer investigation of individual studies.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram presenting the flow of the screening and selection process.



Study Characteristics

The 108 included studies were published between 2016 and mid-2024, with yearly publication counts of 5 (2016), 5 (2017), 2 (2018), 11 (2019), 13 (2020), 21 (2021), 17 (2022), 25 (2023), and 9 (2024). Most studies appeared in peer-reviewed journals (105/108, 97%), with a small number from conference proceedings (3/108, 3%). Geographically, the research was conducted primarily in North America (n=53) and Europe (n=38), followed by Asia (n=8), Oceania (n=3), the Middle East (n=1), South America (n=1), and multiregional studies (n=4).

Regarding study design, most publications used an observational cohort design (87/108, 81%), including longitudinal and EMA-based approaches, as well as studies that analyzed only one arm or group of an intervention. Other studies focused specifically on data collection for the development or training of machine learning models (n=7). Interventional studies were less common and included randomized (n=8), nonrandomized (n=1), and within-subject designs (n=3). Finally, 2 publications used a (multiple) n-of-1 methodology.

Although all studies involved healthy adults in daily life, the samples varied: 37 studies recruited university students, and 19 focused on working adults (of which 16 also investigated stress in the workplace). Others targeted specific subgroups, such as parents (n=8), older adults (n=2), romantic couples (n=2), or specialized populations (n=9), including pregnant women, smokers, drinkers, diet-interested individuals, meditators, and Native Americans residing on a reservation. Finally, 11 studies explicitly mentioned the COVID-19 pandemic as a context. Sample sizes varied widely, ranging from 1 to 21,923 participants, with a median value of 97 (IQR 50.25-182.50). Study duration (ie, the duration of smartphone-based stress assessment) ranged from 1 to 365 days with a median value of 14 (IQR 7-21.5).

Characteristics of Stress Self-Assessment Methods

Across all 108 studies, 114 instruments measuring perceived stress were identified. A total of 6 studies used 2 different instruments within the same protocol (eg, daily single-item EMA plus a weekly short scale). Most instruments (84/114, 74%) relied on a single-item measure. A total of 23 instruments used short scales of 2 to 4 items, and 7 used scales of 5 items or more. Note that only items specifically measuring stress were counted; often, other constructs were measured alongside stress.

The phrasing of the items (ie, the questions for the self-assessment) was grouped into 3 categories: (1) 60 out of 114 instruments measured stress in a momentary manner (eg,

XSL•FO
RenderX

"How stressed do you feel right now?" or "At the moment, I feel stressed"), (2) 30 instruments relied on a period-based assessment (eg, "How stressful has your day been so far?" or "How stressed were you in the last 30 minutes?"), and (3) 17 instruments relied on an appraisal-based approach in the style of typical stress questionnaires (eg, "I feel that I cannot cope with the things I have to do" or "I am on top of things"). Seven instruments combined multiple of these approaches within a single scale. Regarding the response format, Likert scales (75/114, 66%) and visual analog scales (VAS; 33/114, 29%) were most common. Less frequently used formats included pictorial scales (n=2) and binary yes-or-no responses, either alone (n=1) or combined with a Likert scale or a VAS (n=3).

Across the 108 included studies, measurement frequency varied. Most studies used an intensive approach of 3 or more measurements or a variable number of measurements each day (73/108, 68%), which is typical of EMA methodology. A considerable number of studies measured stress once (n=25) or twice (n=5) per day. Finally, 5 studies used a nonintensive approach of fewer than 1 measurement per day (eg, once per week).

Most studies (83/108, 77%) administered stress assessments using a mobile app with push notifications. Another 22 studies used SMS or messaging apps to deliver survey links to an online platform (eg, Qualtrics), sometimes automated using services such as SurveySignal. One study used the smartphone timer app as a reminder, and 2 studies delivered and received assessment items directly via text message. This made it technically possible to include participants with older cell phones instead of smartphones. Some studies specified whether they supported Android (Google) devices (n=27), iOS (Apple) devices (n=2), or both (n=23).

Evidence on Validity and Reliability

Overall, relatively few studies reported robust evidence on the validity or reliability of their smartphone-based stress measures. The available evidence ranged from indirect indications of validity, through correlational assessments with external measures, to more formal psychometric evaluations.

Several studies used items adapted from existing stress questionnaires, providing indirect evidence of validity insofar as the original instruments had undergone prior validation. Most commonly, researchers adapted items from the PSS [12] (eg, [28,42-52]). Other studies drew on items from the DASS [13] (eg, [53]), the Positive and Negative Affect Schedule (PANAS) [148] (eg, [52,54]), the Psychological Stress Measure-9 (PSM-9) [149] (eg, [55,56]), or the Stress and Coping Inventory (SCI) [150] (eg, [57]). Several studies also adopted a single-item stress question developed by the American Psychological Association (APA) [151] (eg, [58,59]). However, none of these instruments were originally designed or validated to capture momentary stress. Consequently, researchers modified the time frames and, in some cases, the response formats, to approximate momentary assessment, possibly raising questions about the validity of these adapted measures for real-time stress evaluation.

A subset of studies provided evidence for convergent, criterion, or discriminant validity, typically by correlating momentary stress ratings with established questionnaires or physiological markers. This is an important step toward establishing the validity of the measurement approach. For example, Gidlow et al [60], Feneberg et al [61], and Jones et al [62] reported significant associations with the PSS, and daSilva et al [63] found similar results for their pictorial stress scale. Other instruments that were used included the PSM-9 (eg, [64]), the Achenbach Adult Self-Report (ASR) [152] (eg, [65]), and the Subcomponents of Affect Scale (SAS) [153] (eg, [66]). Þórarinsdóttir et al [67], in a study explicitly designed to assess validity, reported correlations between smartphone-based stress assessments and the PSS, Eysenck Personality Questionnaire Neuroticism (EPQ-N) [154], and the Functioning Assessment Short Test (FAST) [155]. Other studies similarly used a battery of affective or stress-related instruments (eg, [28,29,68-72]).

Studies examining associations between smartphone-based stress measures and physiological indicators produced mixed findings. For example, Gidlow et al [60] and Feneberg et al [61] did not observe significant associations with hair cortisol. Giessing et al [73] similarly reported no associations with salivary cortisol or α-amylase. In contrast, Martinez et al [74] found a small but significant association between self-reported stress and HRV, and de Vries et al [75] reported comparable findings for some participants in a multiple n-of-1 study. Gordon and Mendes [76] also found some evidence of an association with heart rate and blood pressure reactivity.

Finally, a small number of studies conducted psychometric evaluations. At the basic level, this included estimating indexes of internal consistency or reliability. For multi-item measures, reliability was estimated using Cronbach α (eg, [56]) or multilevel McDonald ω to estimate between- and within-person reliability (eg, [43,51]). Within-person reliability was also estimated using variance decomposition techniques (eg, [52]). For single-item measures, researchers sometimes calculated intraclass correlations (ICCs) to quantify the proportion of variance attributable to between-person (as opposed to within-person) differences (eg, [72,77,78]). Lower between-person ICC values suggest that moment-to-moment, within-person changes account for most of the variability, which might be preferable in EMA studies aiming to capture momentary variations in stress.

At the most advanced level, a small number of studies performed comprehensive factor-analytic assessments to examine the underlying structure of the measurement instrument. Murray et al [28] validated an EMA-adapted version of the PSS (PSS-EMA), demonstrating reliability at both within- and between-person levels and confirming the scale's factorial structure using multilevel confirmatory factor analysis. They further established convergent and criterion validity through associations with related constructs, such as stress, anxiety, depression, and negative affect. Vikoler et al [29] validated the Di-Eu-Stress State (DESS) scale, a dedicated EMA measure designed to capture both stress (distress) and eustress. Their evaluation likewise included establishing within- and between-person reliability, multilevel confirmatory factor analysis, and associations with a range of constructs, such as stress, anxiety, depressiveness, and affect.

XSL•FO
RenderX

Discussion

Main Findings

This scoping review systematically mapped 108 studies that used smartphone-based methods to assess self-reported perceived stress in healthy adults during daily life. Findings are consistent with the observations made in the earlier review by Þórarinsdóttir et al [20], particularly regarding the heterogeneity of measurement strategies, scale types, number of items, and sampling frequency. In recent years, however, both the quantity and scope of research have expanded considerably. The present review identified 108 studies, compared with the previous set of 35 studies. Most studies used observational designs with EMA-style prompting schedules, often involving 3 or more assessments per day. Across studies, stress was predominantly measured using very brief instruments, with 74% (84/114) relying on single-item measures and most using Likert scales or VAS. Single-item measures have also proven dominant in EMA research in other domains such as pain, momentary depression, and happiness [19,156-158]

Despite the widespread use of these approaches, evidence on validity and reliability was still relatively limited. Some studies adapted items from established stress or affect scales, such as the PSS or PANAS, or reported correlations with validated stress measures and related constructs. Some studies examined reliability or ICC, and the present review did retrieve a small number of recent studies on validated EMA measures, including multilevel psychometric evaluations, such as PSS-EMA [28] and a newly developed EMA stress scale [29]. These efforts signal an emerging methodological maturation, although comprehensive validation remains the exception rather than the norm.

Implications for Selecting and Evaluating Instruments

Based on the present scoping review and existing guidelines on self-assessments in daily life, several pragmatic recommendations can be made for researchers who are entering the field and selecting or developing smartphone-based stress instruments.

First, several recommendations have been proposed for EMA research in general and apply to EMA stress research as well. These include alignment between the measurement approach and the research question [159]. Specific study aims will be paramount in guiding protocol development, including timing and sampling schemes for EMA research, such as event-contingent, time-contingent, or hybrid designs [159]. When applied to the stress field, momentary items (eg, “right now”) might be preferable for capturing rapid stress fluctuations or responses to specific events, as required in high-resolution EMA designs. Period-based items (eg, “in the last 30 minutes” or “today”; also referred to as a coverage model of EMA) may be appropriate for studies focusing on larger time frames but might introduce greater recall demands or bias [160]. Appraisal-based items (eg, “I feel like I cannot cope with the things I have to do”) can strictly speaking be either momentary or period based and are mainly characterized by their alignment with traditional appraisal-based questionnaires such as the PSS, often assessing multiple dimensions of stress. App-based data collection with either a custom-developed mobile app or a readily available app, preferably on both Android and iOS, provides optimal flexibility [19].

Second, in addition to study alignment, measurement properties are relevant to consider. Single-item instruments reduce burden and facilitate intensive sampling but may limit psychometric precision. Researchers should prioritize instruments with demonstrated psychometric quality or include convergent measures (standardized stress questionnaires and affective scales) to evaluate measurement validity in their sample. The current review has shown that some single-item measures do provide satisfactory correlations with extended instruments. In addition, reliable multi-item EMA instruments were identified, such as the PSS-EMA [28] or the DESS scale [29]. When aligned with the research design, such instruments may be preferable. When appropriate, physiological indicators such as HRV (eg, measured using a wearable) can also offer complementary information to enrich the dataset with data measured at a different temporal resolution [19,21,159]. Mobile apps can enable integration with smartphone sensors and wearables but may require technical expertise, impose operating system constraints, and increase participant onboarding demands. SMS- or messaging app–based assessments enhance inclusivity, especially for participants using older devices, but offer less flexibility and integration with contextual data.

Opportunities for Improving the EMA Tradition in Stress Research

More training on the design and implementation of EMA approaches is warranted. When assessing studies, not all methodological choices appear carefully considered or prepared. Within and outside the field of stress research, many EMA measures have used novel questions without substantial prior psychometric evaluation, which implies potential problems in terms of replication and validity of the findings [156,159,160]. In addition to training on the approach itself, providing device training has been recommended to assess comprehension and promote adoption [156,160]. In line with this, transparent reporting of implementation choices supports reproducibility and informs future study design and interpretation of findings. Several reporting checklists have already been developed in related fields but could be applied to daily-life stress monitoring [19,161-163].

Because reliability is sample dependent, documenting reliability coefficients is relevant [160]. Reliability can be quantified through Cronbach α or multilevel McDonald ω for multi-item scales. The amount of variance attributed to within- and between-person differences can be quantified using ICC for single-item measures. Factor-analytic approaches provide even stronger evidence of construct validity. Alternative approaches that have been proposed to support item validity are real-time cognitive interviewing, planned missingness designs, and tests of context-dependent differential item functioning and item parameter drift [160].

Progress can also be made by further integrating smartphone-based self-report methods with physiological, behavioral, and contextual data. There is evidence for the complementary relevance of stress physiology, although

XSL•FO
RenderX

temporal resolution is key and daily-life monitoring is challenging [8]. Wearable devices increasingly enable continuous monitoring of physiological parameters, and combining these multimodal data streams with EMA-based stress ratings offers opportunities for digital phenotyping and machine learning–based stress detection [164]. However, the validity of prediction models depends fundamentally on the quality of the EMA stress measures used as ground truth, underscoring the need for robust psychometric evaluation.

Limitations and Future Directions

Several limitations must be acknowledged. First, this review focused on general perceived stress in healthy adults, excluding large bodies of work on domain-specific stress (eg, trauma-related stress [165] or caregiver stress [166]), as well as studies involving clinical populations, minors, or unhealthy housing conditions. The review by Yang et al [23] supported the feasibility of stress monitoring in an adult patient sample but noted challenges related to the complexity of EMA data and completion rates. The approaches documented in the current review are nevertheless relevant to the broader domain of smartphone-based stress assessment and may provide additional insights into existing methodology. Second, due to the methodological diversity of included studies, we did not conduct a formal risk of bias assessment or quantitative synthesis of validity evidence. Although frameworks such as the Patient-Reported Outcomes Measurement Information System (PROMIS) [167] could facilitate structured comparisons of psychometric quality, their application requires a narrower, more homogeneous subset of studies. Another methodological limitation is that no interrater reliability coefficient was calculated during the screening of studies. Third, because of the high number of publications and to facilitate comparison, studies with missing data on key methodological variables (eg, stress item phrasing or measurement frequency) were excluded. Although this improved the interpretability of the descriptive synthesis, it may have introduced selection bias if reporting completeness was related to certain study characteristics. Finally, many studies in this review recruited university students or white-collar workers in high-income countries. Future work should include broader demographic and cultural contexts to enhance generalizability.

Conclusions

Smartphone-based self-assessment of perceived stress in daily life has become widespread, with notable growth over the past decade. The field of research is characterized by substantial study heterogeneity, with the use of single-item EMA measures and intensive sampling schedules nevertheless being most common. Some studies provide promising evidence for the validity and reliability of smartphone-based stress instruments, and a small number conduct rigorous multilevel psychometric evaluations. Nevertheless, comprehensive validation remains relatively rare.

Together, these findings indicate that smartphone-based stress assessment has become widely used, but studies are methodologically diverse and often rely on instruments whose psychometric properties have not been adequately established. Researchers should carefully align instrument selection with their study objectives and, whenever possible, prioritize measures with demonstrated validity and reliability or report psychometric indicators such as ICCs or multilevel reliability estimates. By synthesizing existing methods and the available validity and reliability evidence, this review supports more informed instrument selection and the development of robust, evidence-based strategies for mobile stress assessment. Future work would benefit from more systematic psychometric evaluation and transparent reporting of measurement characteristics to strengthen comparability across studies.

Funding

This study received internal funding from KU Leuven through the StressTRACS research project (C3/23/043). The funder had no involvement in the study design, data analysis, interpretation of findings, or preparation of the manuscript.

Data Availability

The dataset generated and analyzed in this scoping review is available in Multimedia Appendices 2 and 3.

Generative AI Use

The authors declare the use of generative AI in the research and writing process. In line with the Generative AI Delegation Taxonomy [168], the following tasks were delegated to generative AI tools under full human supervision: proofreading and editing (ie, improving sentence construction in sections). The generative AI tool used was Copilot GPT-5.2 (Microsoft). Responsibility for the final manuscript lies entirely with the authors. Generative AI tools are not listed as authors and do not bear responsibility for the final outcomes.

Authors' Contributions

Conceptualization: NAJDW (lead), TC (equal), TVD (supporting), BB (supporting)
Data curation: TC
Formal analysis: TC
Funding acquisition: BB (lead), JMA (supporting), NAJDW (supporting)
Investigation: TC (lead), NAJDW (supporting), TVD (supporting)

XSL•FO
RenderX

Methodology: TC (lead), NAJDW (supporting), TVD (supporting)
Project administration: BB
Supervision: NAJDW
Visualization: TC (lead), NAJDW (supporting)
Writing–original draft: TC (lead), NAJDW (supporting)
Writing–review & editing: TC (lead), NAJDW (equal), TVD (supporting), RS (supporting), BB (supporting), GD (supporting), MF-J (supporting), JG (supporting), CNVO (supporting), JMA (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR checklist.
[DOCX File , 80 KB - mhealth_v14i1e94204_app1.docx]

Multimedia Appendix 2
Data table of all included studies (Microsoft Word).
[DOCX File , 69 KB - mhealth_v14i1e94204_app2.docx]

Multimedia Appendix 3
Data table of all included studies (Microsoft Excel).
[XLSX File (Microsoft Excel File), 41 KB - mhealth_v14i1e94204_app3.xlsx]

References

1. Cohen S, Kessler RC, Gordon LU. Strategies for measuring stress in studies of psychiatric and physical disorders. In: Measuring Stress: A Guide for Health and Social Scientists. Oxford, UK: Oxford University Press; 1995.
2. O'Connor DB, Thayer JF, Vedhara K. Stress and health: a review of psychobiological processes. Annu Rev Psychol 2021 Jan 04;72:663-688. [doi: 10.1146/annurev-psych-062520-122331] [Medline: 32886587]
3. Sara JD, Toya T, Ahmad A, Clark MM, Gilliam WP, Lerman LO, et al. Mental stress and its effects on vascular health. Mayo Clin Proc 2022 May;97(5):951-990 [FREE Full text] [doi: 10.1016/j.mayocp.2022.02.004] [Medline: 35512885]
4. Madsen IE, Nyberg ST, Magnusson Hanson LL, Ferrie JE, Ahola K, Alfredsson L, et al. Job strain as a risk factor for clinical depression: systematic review and meta-analysis with additional individual participant data. Psychol Med 2017 Jun;47(8):1342-1356 [FREE Full text] [doi: 10.1017/S003329171600355X] [Medline: 28122650]
5. Martinez MF, O'Shea KJ, Kern MC, Chin KL, Dinh JV, Bartsch SM, et al. The health and economic burden of employee burnout to U.S. employers. Am J Prev Med 2025 Apr;68(4):645-655. [doi: 10.1016/j.amepre.2025.01.011] [Medline: 40019422]
6. Immanuel S, Teferra MN, Baumert M, Bidargaddi N. Heart rate variability for evaluating psychological stress changes in healthy adults: a scoping review. Neuropsychobiology 2023;82(4):187-202 [FREE Full text] [doi: 10.1159/000530376] [Medline: 37290411]
7. Klimek A, Mannheim I, Schouten G, Wouters EJ, Peeters MW. Wearables measuring electrodermal activity to assess perceived stress in care: a scoping review. Acta Neuropsychiatr 2023 Mar 24;37:e19. [doi: 10.1017/neu.2023.19] [Medline: 36960675]
8. Weber J, Angerer P, Apolinário-Hagen J. Physiological reactions to acute stressors and subjective stress during daily life: a systematic review on ecological momentary assessment (EMA) studies. PLoS One 2022 Jul 27;17(7):e0271996 [FREE Full text] [doi: 10.1371/journal.pone.0271996] [Medline: 35895674]
9. Lazarus RS, Folkman S. Transactional theory and research on emotions and coping. Eur J Pers 1987 Sep;1(3):141-169. [doi: 10.1002/per.2410010304]
10. Vaessen T, Rintala A, Otsabryk N, Viechtbauer W, Wampers M, Claes S, et al. The association between self-reported stress and cardiovascular measures in daily life: a systematic review. PLoS One 2021 Nov 19;16(11):e0259557 [FREE Full text] [doi: 10.1371/journal.pone.0259557] [Medline: 34797835]
11. Siepe BS, Tutunji R, Rieble CL, Proppert RK, Fried EI. Associations between ecological momentary assessment and passive sensor data in a large student sample. J Psychopathol Clin Sci 2025 Nov;134(8):912-925. [doi: 10.1037/abn0001013] [Medline: 40742761]
12. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
13. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-U]

XSL•FO
RenderX

14. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SL, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002 Sep 26;32(6):959-976. [doi: 10.1017/S0033291702006074]
15. Adu P, Popoola T, Iqbal N, Medvedev ON, Simpson CR. Validating the Depression Anxiety Stress Scales (DASS-21) across Germany, Ghana, India, and New Zealand using Rasch methodology. J Affect Disord 2025 Aug 15;383:363-373 [FREE Full text] [doi: 10.1016/j.jad.2025.04.099] [Medline: 40294823]
16. Wojujutari AK, Idemudia ES. Consistency as the currency in psychological measures: a reliability generalization meta-analysis of Kessler Psychological Distress Scale (K-10 and K-6). Depress Anxiety 2024 Oct 16;2024:3801950. [doi: 10.1155/2024/3801950] [Medline: 40226709]
17. Yılmaz Koğar E, Koğar H. A systematic review and meta-analytic confirmatory factor analysis of the Perceived Stress Scale (PSS-10 and PSS-14). Stress Health 2024 Feb;40(1):e3285. [doi: 10.1002/smi.3285] [Medline: 37341705]
18. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
19. de Vries LP, Baselmans BM, Bartels M. Smartphone-based ecological momentary assessment of well-being: a systematic review and recommendations for future studies. J Happiness Stud 2021;22(5):2361-2408 [FREE Full text] [doi: 10.1007/s10902-020-00324-7] [Medline: 34720691]
20. Þórarinsdóttir H, Kessing LV, Faurholt-Jepsen M. Smartphone-based self-assessment of stress in healthy adult individuals: a systematic review. J Med Internet Res 2017 Feb 13;19(2):e41 [FREE Full text] [doi: 10.2196/jmir.6397] [Medline: 28193600]
21. Wrzus C, Neubauer AB. Ecological momentary assessment: a meta-analysis on designs, samples, and compliance across research fields. Assessment 2023 Apr;30(3):825-846 [FREE Full text] [doi: 10.1177/10731911211067538] [Medline: 35016567]
22. Yin H, Zhu H, Gu J, Qin H, Ding W, Guo N, et al. Mobile-based ecological momentary assessment and intervention: bibliometric analysis. Front Psychiatry 2024 Feb 26;15:1300739 [FREE Full text] [doi: 10.3389/fpsyt.2024.1300739] [Medline: 38469030]
23. Yang YS, Ryu GW, Choi M. Methodological strategies for ecological momentary assessment to evaluate mood and stress in adult patients using mobile phones: systematic review. JMIR Mhealth Uhealth 2019 Apr 01;7(4):e11215 [FREE Full text] [doi: 10.2196/11215] [Medline: 30932866]
24. Cook PF, Schmiege SJ, Bradley-Springer L, Starr W, Carrington JM. Motivation as a mechanism for daily experiences' effects on HIV medication adherence. J Assoc Nurses AIDS Care 2018;29(3):383-393 [FREE Full text] [doi: 10.1016/j.jana.2017.09.003] [Medline: 28967490]
25. Wilson MB, Kallogjeri D, Joplin CN, Gorman MD, Krings JG, Lenze EJ, et al. Ecological momentary assessment of tinnitus using smartphone technology: a pilot study. Otolaryngol Head Neck Surg 2015 May;152(5):897-903 [FREE Full text] [doi: 10.1177/0194599815569692] [Medline: 25676150]
26. Stinson L, Liu Y, Dallery J. Ecological momentary assessment: a systematic review of validity research. Perspect Behav Sci 2022 May 06;45(2):469-493 [FREE Full text] [doi: 10.1007/s40614-022-00339-w] [Medline: 35719870]
27. Song J, Howe E, Oltmanns JR, Fisher AJ. Examining the concurrent and predictive validity of single items in ecological momentary assessments. Assessment 2023 Jul;30(5):1662-1671 [FREE Full text] [doi: 10.1177/10731911221113563] [Medline: 36004406]
28. Murray AL, Xiao Z, Zhu X, Speyer LG, Yang Y, Brown RH, et al. Psychometric evaluation of an adapted version of the perceived stress scale for ecological momentary assessment research. Stress Health 2023 Oct;39(4):841-853. [doi: 10.1002/smi.3229] [Medline: 36697362]
29. Vikoler T, Dániel K, Traut-Mattausch E. The Di-Eu-Stress State Scale (DESS Scale): development and validation of a scale measuring state distress and eustress. Eur J Psychol Assess 2024 May 29;41(6):486-505. [doi: 10.1027/1015-5759/a000829]
30. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
31. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
32. Peters MD, Godfrey C, McInerney P, Khalil H, Larsen P, Marnie C, et al. Best practice guidance and reporting items for the development of scoping review protocols. JBI Evid Synth 2022 Apr 01;20(4):953-968. [doi: 10.11124/JBIES-21-00242] [Medline: 35102103]
33. Colman T, De Witte NA, Van Daele T. Scoping review protocol: measuring self-reported stress in healthy adults in daily life - a scoping review of smartphone-based assessment methods. Open Science Framework. 2025. URL: https://doi.org/10.17605/OSF.IO/7UM6B [accessed 2026-02-25]
34. Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z. JBI manual for evidence synthesis. JBI. 2024. URL: https://doi.org/10.46658/JBIMES-24-01 [accessed 2026-02-25]

XSL•FO
RenderX

35. WHO housing and health guidelines. World Health Organization. 2018. URL: https://www.who.int/publications/i/item/9789241550376 [accessed 2026-02-25]
36. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
37. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2023. URL: https://www.R-project.org/ [accessed 2026-02-25]
38. Wickham H, Bryan J. readxl: Read Excel Files. Tidyverse. 2026. URL: https://readxl.tidyverse.org [accessed 2026-02-25]
39. Ooms J, Denney B. writexl: Export Data Frames to Excel 'xlsx' Format. rOpenSci. 2026. URL: https://ropensci.r-universe.dev/writexl [accessed 2026-02-25]
40. Gohel D, Moog S, Heckmann M. officer: manipulation of Microsoft Word and PowerPoint documents. ArData. 2026. URL: https://ardata-fr.github.io/officeverse/ [accessed 2026-02-25]
41. Gohel D, Skintzos P. flextable: functions for tabular reporting. ArData. 2026. URL: https://ardata-fr.github.io/flextable-book/ [accessed 2026-02-25]
42. Aulbach MB, van Alebeek H, Jones CM, Blechert J. Why we don't eat as intended: moderators of the short-term intention-behaviour relation in food intake. Br J Health Psychol 2024 Sep;29(3):576-588. [doi: 10.1111/bjhp.12714] [Medline: 38290795]
43. Dunton GF, Ke W, Dzubur E, O'Connor SG, Lopez NV, Margolin G. Within-subject effects of stress on weight-related parenting practices in mothers: an ecological momentary assessment study. Ann Behav Med 2019 Mar 28;53(5):415-425 [FREE Full text] [doi: 10.1093/abm/kay053] [Medline: 31222228]
44. Fitz N, Kushlev K, Jagannathan R, Lewis T, Paliwal D, Ariely D. Batching smartphone notifications can improve well-being. Comput Hum Behav 2019 Dec;101:84-94. [doi: 10.1016/j.chb.2019.07.016]
45. Kaden SJ, Dalton ED. Momentary fluctuations in emotional intelligence and stress predict changes in disordered eating. J Am Coll Health 2024 Nov;72(8):2668-2675. [doi: 10.1080/07448481.2022.2122724] [Medline: 36170498]
46. Kaiser T, Butter B, Arzt S, Pannicke B, Reichenberger J, Ginzinger S, et al. Time-lagged prediction of food craving with qualitative distinct predictor types: an application of BISCWIT. Front Digit Health 2021 Sep 20;3:1-8. [doi: 10.3389/fdgth.2021.694233]
47. King ZD, Moskowitz J, Egilmez B, Zhang S, Zhang L, Bass M, et al. micro-stress EMA: a passive sensing framework for detecting in-the-wild stress in pregnant mothers. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Sep;3(3):1-22 [FREE Full text] [doi: 10.1145/3351249] [Medline: 32432212]
48. McLeod K, Spachos P, Plataniotis KN. Smartphone-based wellness assessment using mobile environmental sensors. IEEE Syst J 2021 Jun;15(2):1989-1999. [doi: 10.1109/JSYST.2020.3004558]
49. Morin L, Laurin JC, Doucerain M, Grégoire S. Dyadic association between new parents' mindfulness and relationship satisfaction: mediating role of perceived stress. J Soc Pers Relat 2023 May;40(5):1448-1470 [FREE Full text] [doi: 10.1177/02654075231156404] [Medline: 37255719]
50. Pryss R, John D, Schlee W, Schlotz W, Schobel J, Kraft R, et al. Exploring the time trend of stress levels while using the crowdsensing mobile health platform, TrackYourStress, and the influence of perceived stress reactivity: ecological momentary assessment pilot study. JMIR Mhealth Uhealth 2019 Oct 30;7(10):e13978 [FREE Full text] [doi: 10.2196/13978] [Medline: 31670692]
51. Ruf A, Neubauer AB, Koch ED, Ebner-Priemer U, Reif A, Matura S. Individual differences in the dietary response to stress in ecological momentary assessment: does the individual-difference model need expansion? Appl Psychol Health Well Being 2023 May;15(2):629-649. [doi: 10.1111/aphw.12400] [Medline: 36082743]
52. Schultchen D, Reichenberger J, Mittl T, Weh TR, Smyth JM, Blechert J, et al. Bidirectional relationship of stress and affect with physical activity and healthy eating. Br J Health Psychol 2019 May;24(2):315-333 [FREE Full text] [doi: 10.1111/bjhp.12355] [Medline: 30672069]
53. John-Henderson NA, Oosterhoff B, Kampf TD, Hall B, Johnson LR, Laframboise ME, et al. Historical loss: implications for health of American Indians in the Blackfeet community. Ann Behav Med 2022 Feb 11;56(2):193-204 [FREE Full text] [doi: 10.1093/abm/kaab032] [Medline: 33969868]
54. Pannicke B, Kaiser T, Reichenberger J, Blechert J. Networks of stress, affect and eating behaviour: anticipated stress coping predicts goal-congruent eating in young adults. Int J Behav Nutr Phys Act 2021 Jan 09;18(1):9 [FREE Full text] [doi: 10.1186/s12966-020-01066-8] [Medline: 33422046]
55. Grégoire S, Chénier C, Doucerain M, Lachance L, Shankland R. Ecological momentary assessment of stress, well-being, and psychological flexibility among college and university students during acceptance and commitment therapy. Can J Behav Sci 2020;52(3):231-243. [doi: 10.1037/cbs0000175]
56. Grégoire S, Doucerain M, Morin L, Finkelstein-Fox L. The relationship between value-based actions, psychological distress and well-being: a multilevel diary study. J Context Behav Sci 2021 Apr;20:79-88. [doi: 10.1016/j.jcbs.2021.03.006]
57. Zander-Schellenberg T, Collins IM, Miché M, Guttmann C, Lieb R, Wahl K. Does laughing have a stress-buffering effect in daily life? An intensive longitudinal study. PLoS One 2020 Jul 09;15(7):e0235851 [FREE Full text] [doi: 10.1371/journal.pone.0235851] [Medline: 32645063]

XSL•FO
RenderX

58. Freisthler B, Gruenewald PJ, Tebben E, Shockley McCarthy K, Price Wolf J. Understanding at-the-moment stress for parents during COVID-19 stay-at-home restrictions. Soc Sci Med 2021 Jun;279:114025 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114025] [Medline: 34004571]
59. Gallagher J, Carr LJ. Leisure but not occupational physical activity and sedentary behavior associated with better health. J Occup Environ Med 2021 Nov 01;63(11):e774-e782. [doi: 10.1097/JOM.0000000000002365] [Medline: 34456325]
60. Gidlow CJ, Randall J, Gillman J, Silk S, Jones MV. Hair cortisol and self-reported stress in healthy, working adults. Psychoneuroendocrinology 2016 Jan;63:163-169. [doi: 10.1016/j.psyneuen.2015.09.022] [Medline: 26447679]
61. Feneberg AC, Forbes PA, Piperno G, Pronizius E, Stijovic A, Skoluda N, et al. Diurnal dynamics of stress and mood during COVID-19 lockdown: a large multinational ecological momentary assessment study. Proc Biol Sci 2022 May 25;289(1975):20212480 [FREE Full text] [doi: 10.1098/rspb.2021.2480] [Medline: 35611528]
62. Jones M, Taylor A, Liao Y, Intille SS, Dunton GF. Real-time subjective assessment of psychological stress: associations with objectively-measured physical activity levels. Psychol Sport Exerc 2017 Jul;31:79-87 [FREE Full text] [doi: 10.1016/j.psychsport.2017.03.013] [Medline: 29151810]
63. daSilva AW, Huckins JF, Wang W, Wang R, Campbell AT, Meyer ML. Daily perceived stress predicts less next day social interaction: evidence from a naturalistic mobile sensing study. Emotion 2021 Dec;21(8):1760-1770. [doi: 10.1037/emo0000994] [Medline: 35041440]
64. Swendeman D, Comulada WS, Koussa M, Worthman CM, Estrin D, Rotheram-Borus MJ, et al. Longitudinal validity and reliability of brief smartphone self-monitoring of diet, stress, and physical activity in a diverse sample of mothers. JMIR Mhealth Uhealth 2018 Sep 21;6(9):e176 [FREE Full text] [doi: 10.2196/mhealth.9378] [Medline: 30249576]
65. Milne A, Novelli A, Watson C, Yousef H, Holterman LA, Rosen L, et al. The potential of real-time behavior and well-being assessments to improve medical student wellness. Acad Med 2024 Jun 01;99(6):608-612 [FREE Full text] [doi: 10.1097/ACM.0000000000005642] [Medline: 38266202]
66. Jenkins BN, Martin LT, Helen Lee HY, Hunter JF, Acevedo AM, Pressman SD. Affect variability and cortisol in context: the moderating roles of mean affect and stress. Psychoneuroendocrinology 2024 Aug;166:107064 [FREE Full text] [doi: 10.1016/j.psyneuen.2024.107064] [Medline: 38713929]
67. Þórarinsdóttir H, Faurholt-Jepsen M, Ullum H, Frost M, Bardram JE, Kessing LV. The validity of daily self-assessed perceived stress measured using smartphones in healthy individuals: cohort study. JMIR Mhealth Uhealth 2019 Aug 19;7(8):e13418 [FREE Full text] [doi: 10.2196/13418] [Medline: 31429413]
68. Christmann CA, Zolynski G, Hoffmann A, Bleser G. Towards more interactive stress-related self-monitoring tools to improve quality of life. In: Bagnara S, Tartaglia R, Albolino S, Alexander T, Fujita Y, editors. Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018). Cham, Switzerland: Springer; 2018.
69. Komarov O, Ko LW, Jung TP. Associations among emotional state, sleep quality, and resting-state EEG spectra: a longitudinal study in graduate students. IEEE Trans Neural Syst Rehabil Eng 2020 Apr;28(4):795-804. [doi: 10.1109/TNSRE.2020.2972812] [Medline: 32070988]
70. Melcher J, Lavoie J, Hays R, D'Mello R, Rauseo-Ricupero N, Camacho E, et al. Digital phenotyping of student mental health during COVID-19: an observational study of 100 college students. J Am Coll Health 2023 Apr;71(3):736-748. [doi: 10.1080/07448481.2021.1905650] [Medline: 33769927]
71. Tung NY, Yap Y, Bei B, Luecken LJ, Wiley JF. Resilience to early family risk moderates stress-affect associations: a 14-day ecological momentary assessment study. J Affect Disord 2022 Aug 15;311:134-142. [doi: 10.1016/j.jad.2022.05.075] [Medline: 35597463]
72. Wenze SJ, Gaugler TL, Sheets ES, DeCicco JM. Momentary experiential avoidance: within-person correlates, antecedents, and consequences and between-person moderators. Behav Res Ther 2018 Aug;107:42-52. [doi: 10.1016/j.brat.2018.05.011] [Medline: 29859916]
73. Giessing L, Oudejans RR, Hutter V, Plessner H, Strahler J, Frenkel MO. Acute and chronic stress in daily police service: a three-week N-of-1 study. Psychoneuroendocrinology 2020 Dec;122:104865. [doi: 10.1016/j.psyneuen.2020.104865] [Medline: 32961407]
74. Martinez GJ, Grover T, Mattingly SM, Mark G, D'Mello S, Aledavood T, et al. Alignment between heart rate variability from fitness trackers and perceived stress: perspectives from a large-scale in situ longitudinal study of information workers. JMIR Hum Factors 2022 Aug 04;9(3):e33754 [FREE Full text] [doi: 10.2196/33754] [Medline: 35925662]
75. de Vries HJ, Pennings HJ, van der Schans CP, Sanderman R, Oldenhuis HK, Kamphuis W. Wearable-measured sleep and resting heart rate variability as an outcome of and predictor for subjective stress measures: a multiple N-of-1 observational study. Sensors (Basel) 2022 Dec 28;23(1):332 [FREE Full text] [doi: 10.3390/s23010332] [Medline: 36616929]
76. Gordon AM, Mendes WB. A large-scale study of stress, emotions, and blood pressure in daily life using a digital platform. Proc Natl Acad Sci U S A 2021 Aug 03;118(31):e2105573118 [FREE Full text] [doi: 10.1073/pnas.2105573118] [Medline: 34326265]
77. Ewert C, Hoffmann CF, Schröder-Abé M. Stress processing mediates the link between momentary self-compassion and affective well-being. Mindfulness 2022 Aug 09;13:2269-2281. [doi: 10.1007/s12671-022-01954-z]

XSL•FO
RenderX

78. Schilling R, Herrmann C, Ludyga S, Colledge F, Brand S, Pühse U, et al. Does cardiorespiratory fitness buffer stress reactivity and stress recovery in police officers? A real-life study. Front Psychiatry 2020 Jun 23;11:594 [FREE Full text] [doi: 10.3389/fpsyt.2020.00594] [Medline: 32670116]
79. Aalbers G, Hendrickson AT, Vanden Abeele MM, Keijsers L. Smartphone-tracked digital markers of momentary subjective stress in college students: idiographic machine learning analysis. JMIR Mhealth Uhealth 2023 Mar 23;11:e37469. [doi: 10.2196/37469] [Medline: 36951924]
80. Aguilar-Raab C, Stoffel M, Hernández C, Rahn S, Moessner M, Steinhilber B, et al. Effects of a mindfulness-based intervention on mindfulness, stress, salivary alpha-amylase and cortisol in everyday life. Psychophysiology 2021 Dec;58(12):e13937. [doi: 10.1111/psyp.13937] [Medline: 34525214]
81. Bavaresco R, Barbosa J, Vianna H, Büttenbender P, Dias L. Design and evaluation of a context-aware model based on psychophysiology. Comput Methods Programs Biomed 2020 Jun;189:105299 [FREE Full text] [doi: 10.1016/j.cmpb.2019.105299] [Medline: 31935581]
82. Berrocal A, Concepcion W, De Dominicis S, Wac K. Complementing human behavior assessment by leveraging personal ubiquitous devices and social links: an evaluation of the peer-ceived momentary assessment method. JMIR Mhealth Uhealth 2020 Aug 07;8(8):e15947. [doi: 10.2196/15947] [Medline: 32763876]
83. Boemo T, Martín-Garcia O, Pacheco-Romero AM, Blanco I, Lafit G, Myin-Germeys I, et al. Not just emotion regulation, but cognition: an experience sampling study testing the relations of ecological interpretation biases and use of emotion regulation strategies with momentary affective states during daily life functioning. Behav Res Ther 2024 Jun;177:104550 [FREE Full text] [doi: 10.1016/j.brat.2024.104550] [Medline: 38688821]
84. Brandt MJ, Crawford JT, Van Tongeren DR. Worldview conflict in daily life. Soc Psychol Pers Sci 2019;10(1):35-43. [doi: 10.1177/1948550617733517]
85. Brophy-Herb HE, Brincks A, Cook JL, Stacks A, Vallotton CD, Frosch C, et al. Stress intensity and exhaustion among infant and toddler teachers: descriptive analysis and associations with sources of stress and coping strategy use. Early Educ Dev 2023;34(7):1545-1564 [FREE Full text] [doi: 10.1080/10409289.2022.2151399] [Medline: 37849911]
86. Brophy-Herb HE, Stacks AM, Frosch C, Brincks AM, Cook JL, Vallotton CD, et al. The effects of a relationship-focused professional development intervention on infant and toddler teachers' mindfulness-based strategies for coping. Early Child Educ J 2024 Feb;52(2):243-257. [doi: 10.1007/s10643-022-01416-0] [Medline: 39092205]
87. Burg MM, Schwartz JE, Kronish IM, Diaz KM, Alcantara C, Duer-Hefele J, et al. Does stress result in you exercising less? Or does exercising result in you being less stressed? Or is it both? Testing the bi-directional stress-exercise association at the group and person (N of 1) level. Ann Behav Med 2017 Dec;51(6):799-809 [FREE Full text] [doi: 10.1007/s12160-017-9902-4] [Medline: 28290065]
88. Vidal Bustamante CM, Coombs G, Rahimi-Eichi H, Mair P, Onnela JP, Baker JT, et al. Fluctuations in behavior and affect in college students measured using deep phenotyping. Sci Rep 2022 Feb 04;12(1):1932 [FREE Full text] [doi: 10.1038/s41598-022-05331-7] [Medline: 35121741]
89. Chen-Sankey JC, Choi K, Kirchner TR, Feldman RH, Butler J, Mead EL. Flavored cigar smoking among African American young adult dual users: an ecological momentary assessment. Drug Alcohol Depend 2019 Mar 01;196:79-85 [FREE Full text] [doi: 10.1016/j.drugalcdep.2018.12.020] [Medline: 30754022]
90. Chin B, Slutsky J, Raye J, Creswell JD. Mindfulness training reduces stress at work: a randomized controlled trial. Mindfulness (N Y) 2019 Apr;10(4):627-638 [FREE Full text] [doi: 10.1007/s12671-018-1022-0] [Medline: 30923585]
91. Courtney JB, Lipsey T, Braun B, Henry K, Nelson TL, Li K. Using ecological momentary assessment to examine the effects of duty status on acute stress and tiredness in firefighters: a pilot study. J Occup Environ Med 2020 Oct;62(10):859-870. [doi: 10.1097/JOM.0000000000001988] [Medline: 32769799]
92. Dancík D, Kasanova Z, Hajdúk M, Heretik A. Attachment, stress and emotions in daily life: an experience sampling study. Stud Psychol 2021;63(4):323-336. [doi: 10.31577/sp.2021.04.830]
93. DaSilva AW, Huckins JF, Wang R, Wang W, Wagner DD, Campbell AT. Correlates of stress in the college environment uncovered by the application of penalized generalized estimating equations to mobile sensing data. JMIR Mhealth Uhealth 2019 Mar 19;7(3):e12084 [FREE Full text] [doi: 10.2196/12084] [Medline: 30888327]
94. de Vries H, Kamphuis W, Oldenhuis H, van der Schans C, Sanderman R. Moderation of the stressor-strain process in interns by heart rate variability measured with a wearable and smartphone app: within-subject design using continuous monitoring. JMIR Cardio 2021 Oct 04;5(2):e28731 [FREE Full text] [doi: 10.2196/28731] [Medline: 34319877]
95. Dekker CA, Baumgartner SE. Is life brighter when your phone is not? The efficacy of a grayscale smartphone intervention addressing digital well-being. Mob Media Commun 2023 Nov 09;12(3):688-708. [doi: 10.1177/20501579231212062]
96. Duong JB, Carta KE, Walters SN, Simo Fiallo N, Benamu DI, Jumonville G, et al. Relationship closeness as a protective factor against the sensitizing effect of adversity history. J Soc Pers Relatsh 2023 Jul 06;40(11):3752-3780. [doi: 10.1177/02654075231185043]
97. Ellison WD, Trahan AC, Pinzon JC, Gillespie ME, Simmons LM, King KY. For whom, and for what, is experience sampling more accurate than retrospective report? Pers Individ Differ 2020;163:110071. [doi: 10.1016/j.paid.2020.110071]
98. Engelen L, Chau JY, Burks-Young S, Bauman A. Application of ecological momentary assessment in workplace health evaluation. Health Promot J Austr 2016 Feb;27(3):259-263. [doi: 10.1071/HE16043] [Medline: 27596817]

XSL•FO
RenderX

99. Faulkner G, Weatherson KA, Duncan MJ, Wunderlich KB, Puterman E. Exploring work-time affective states through ecological momentary assessment in an office-based intervention to reduce occupational sitting. J Phys Act Health 2023 Jun 01;20(6):566-570. [doi: 10.1123/jpah.2022-0495] [Medline: 37037458]
100. Felt JM, Russell MA, Johnson JA, Ruiz JM, Uchino BN, Allison M, et al. Within-person associations of optimistic and pessimistic expectations with momentary stress, affect, and ambulatory blood pressure. Anxiety Stress Coping 2023 Sep;36(5):636-648 [FREE Full text] [doi: 10.1080/10615806.2022.2142574] [Medline: 36371799]
101. Fritz H, Tarraf W, Saleh DJ, Cutchin MP. Using a smartphone-based ecological momentary assessment protocol with community dwelling older African Americans. J Gerontol B Psychol Sci Soc Sci 2017 Sep 01;72(5):876-887 [FREE Full text] [doi: 10.1093/geronb/gbw166] [Medline: 28057696]
102. Gerstberger L, Blanke ES, Keller J, Brose A. Stress buffering after physical activity engagement: an experience sampling study. Br J Health Psychol 2023 Sep;28(3):876-892. [doi: 10.1111/bjhp.12659] [Medline: 37037566]
103. Gignac F, Righi V, Toran R, Paz Errandonea L, Ortiz R, Mijling B, et al. Short-term NO2 exposure and cognitive and mental health: a panel study based on a citizen science project in Barcelona, Spain. Environ Int 2022 Jun;164:107284 [FREE Full text] [doi: 10.1016/j.envint.2022.107284] [Medline: 35576732]
104. Goldman A, York Cornwell E. Stand by me: social ties and health in real-time. Socius 2023;9 [FREE Full text] [doi: 10.1177/23780231231171112] [Medline: 37822581]
105. Goodday SM, Karlin E, Alfarano A, Brooks A, Chapman C, Desille R, Stress And Recovery Participants, et al. An alternative to the light touch digital health remote study: the stress and recovery in frontline COVID-19 health care workers study. JMIR Form Res 2021 Dec 10;5(12):e32165 [FREE Full text] [doi: 10.2196/32165] [Medline: 34726607]
106. Goodson PN, Lopez RB, Denny BT. Perceived stress moderates emotion regulation success in real-world contexts: an ecologically-valid multilevel investigation. Anxiety Stress Coping 2024 Jul;37(4):501-514. [doi: 10.1080/10615806.2023.2278057] [Medline: 37929316]
107. Gregersen T, Mercer S, MacIntyre P, Talbot K, Banga C. Understanding language teacher wellbeing: an ESM study of daily stressors and uplifts. Lang Teach Res 2023;27(4):862-883 [FREE Full text] [doi: 10.1177/13621688820965897]
108. Hernandez J, McDuff D, Infante C, Maes P, Quigley K, Picard R. Wearable ESM: differences in the experience sampling method across wearable devices. In: MobileHCI '16: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services. New York, NY: Association for Computing Machinery; 2016:195-205.
109. Hojjaty A, Zawadzki MJ. Examining how morning stress forecasts relate to subsequent ecological momentary assessments of stress and coping. Stress Health 2024 Jun;40(3):e3344. [doi: 10.1002/smi.3344] [Medline: 37924512]
110. Huckins JF, DaSilva AW, Hedlund EL, Murphy EI, Rogers C, Wang W, et al. Causal factors of anxiety and depression in college students: longitudinal ecological momentary assessment and causal analysis using Peter and Clark Momentary Conditional Independence. JMIR Ment Health 2020 Jun 10;7(6):e16684 [FREE Full text] [doi: 10.2196/16684] [Medline: 32519971]
111. Jakobsson P, Fagerlund A, Laakso M, Londen M. The emotional well-being of teachers and school personnel: an experience sampling study. Int J Emot Educ 2023;15(2):54-70. [doi: 10.56300/SEDD8938]
112. Kollárik M, Heinzel CV, Miché M, Lieb R, Wahl K. Exam-related unwanted intrusive thoughts and related neutralizing behaviors: analogues to obsessions and compulsions. PLoS One 2022;17(7):e0270692 [FREE Full text] [doi: 10.1371/journal.pone.0270692] [Medline: 35789213]
113. Kou L, Tao Y, Kwan MP, Chai Y. Understanding the relationships among individual-based momentary measured noise, perceived noise, and psychological stress: a geographic ecological momentary assessment (GEMA) approach. Health Place 2020 Jul;64:102285. [doi: 10.1016/j.healthplace.2020.102285] [Medline: 32819555]
114. Lawrence K, Bucci S, Berry K, Carter LA, Taylor P. Fluctuations in proximity seeking and paranoia. Br J Clin Psychol 2022 Nov;61(4):1019-1037 [FREE Full text] [doi: 10.1111/bjc.12372] [Medline: 35570710]
115. Li CT, Cao J, Li TM. Eustress or distress: an empirical study of perceived stress in everyday college life. In: UbiComp '16: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. New York, NY: Association for Computing Machinery; 2016:1209-1217.
116. Loth KA, Ji Z, Wolfson J, Fisher J, Berge J, Neumark-Sztainer D. Momentary predictors of a broad range of food parenting practices within a population-based sample of parents of preschool-aged children. Front Public Health 2022;10:944734 [FREE Full text] [doi: 10.3389/fpubh.2022.944734] [Medline: 36726615]
117. Maher JP, Batts M, Rebar AL, Mead MP, Zaplatosch ME, Hevel DJ, et al. Bidirectional relations between daily stress and sleep among Black emerging adults. Ann Behav Med 2022 Nov 05;56(11):1089-1100. [doi: 10.1093/abm/kaac015] [Medline: 35986720]
118. Massar SA, Ong JL, Lau T, Ng BK, Chan LF, Koek D, et al. Working-from-home persistently influences sleep and physical activity 2 years after the Covid-19 pandemic onset: a longitudinal sleep tracker and electronic diary-based study. Front Psychol 2023;14:1145893 [FREE Full text] [doi: 10.3389/fpsyg.2023.1145893] [Medline: 37213365]
119. Masucci MD, Martin V, Le TP, Cohen AS. Negative schizotypy attenuates the effect of momentary stress on social dysfunction related to COVID-19 social distancing. Schizophr Res 2022 May;243:24-31 [FREE Full text] [doi: 10.1016/j.schres.2022.02.028] [Medline: 35231831]

XSL•FO
RenderX

120. Maxhuni A, Hernandez-Leal P, Morales E, Sucar L, Osmani V, Mayora O. Unobtrusive stress assessment using smartphones. IEEE Trans Mobile Comput 2021 Jun 1;20(6):2313-2325. [doi: 10.1109/TMC.2020.2974834]
121. McLeod S, Berry K, Taylor P, Wearden A. Romantic attachment and support preferences in new mothers: the moderating role of stress. J Soc Pers Relatsh 2021;38(5):1535-1552. [doi: 10.1177/0265407521990432]
122. Mehrotra A, Tsapeli F, Hendley R, Musolesi M. MyTraces: investigating correlation and causation between users' emotional states and mobile phone interaction. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017;1(3). [doi: 10.1145/3130948]
123. Messner EM, Sariyska R, Mayer B, Montag C, Kannen C, Schwerdtfeger A, et al. Insights—future implications of passive smartphone sensing in the therapeutic context. Verhaltenstherapie 2022;32(Suppl 1):86-95. [doi: 10.1159/000501951]
124. Müller SR, Peters H, Matz SC, Wang W, Harari GM. Investigating the relationships between mobility behaviours and indicators of subjective well–being using smartphone–based experience sampling and GPS tracking. Eur J Pers 2020 Sep 01;34(5):714-732. [doi: 10.1002/per.2262]
125. Oved S, Mofaz M, Lan A, Einat H, Kronfeld-Schor N, Yamin D, et al. Differential effects of COVID-19 lockdowns on well-being: interaction between age, gender and chronotype. J R Soc Interface 2021 Jun;18(179):20210078 [FREE Full text] [doi: 10.1098/rsif.2021.0078] [Medline: 34062107]
126. Pacanowski CR, Dominick G, Crosby RD, Engel SG, Cao L, Linde JA. Daily self-weighing compared with an active control causes greater negative affective lability in emerging adult women: a randomized trial. Appl Psychol Health Well Being 2023 Nov;15(4):1695-1713. [doi: 10.1111/aphw.12463] [Medline: 37339756]
127. Park SH, Petrunoff NA, Wang NX, van Dam RM, Sia A, Tan CS, et al. Daily park use, physical activity, and psychological stress: a study using smartphone-based ecological momentary assessment amongst a multi-ethnic Asian cohort. Ment Health Phys Act 2022 Mar;22:100440. [doi: 10.1016/j.mhpa.2022.100440]
128. Peterson-Sockwell H, Rejeski WJ, Fanning J, Porges SW, Heilman KJ, Laurienti PJ, et al. Differential momentary reports of stress and affect associated with alcohol consumption in middle-aged versus younger adults. Subst Use Misuse 2023;58(5):666-675 [FREE Full text] [doi: 10.1080/10826084.2023.2177967] [Medline: 36852426]
129. Righetti F, Gere J, Hofmann W, Visserman ML, Van Lange PA. The burden of empathy: partners' responses to divergence of interests in daily life. Emotion 2016 Aug;16(5):684-690. [doi: 10.1037/emo0000163] [Medline: 26950367]
130. Schneider E, Hopf D, Eckstein M, Scheele D, Aguilar-Raab C, Herpertz SC, et al. Stress during the COVID-19 pandemic moderates pain perception and momentary oxytocin levels. J Clin Med 2023 Mar 16;12(6):2333 [FREE Full text] [doi: 10.3390/jcm12062333] [Medline: 36983333]
131. Shahane AD, Godfrey DA, Denny BT. Predicting real-world emotion and health from spontaneously assessed linguistic distancing using novel scalable technology. Emotion 2023 Oct;23(7):2002-2012. [doi: 10.1037/emo0001211] [Medline: 36808975]
132. Sheng R, Yang X, Zhou Y, Liu X, Xu W. COVID-19 and eating problems in daily life: the mediating roles of stress, negative affect and posttraumatic stress symptoms. Psychol Rep 2023 Feb;126(1):34-51 [FREE Full text] [doi: 10.1177/00332941211045315] [Medline: 34784823]
133. Shevchenko Y, Huber N, Reips UD. Psychological well-being during the COVID-19 pandemic: combining a web survey with experience sampling methodology. PLoS One 2023;18(3):e0282649 [FREE Full text] [doi: 10.1371/journal.pone.0282649] [Medline: 36961837]
134. Stieger S, Schmid I, Altenburger P, Lewetz D. Short-, medium-, and long-term impact of watching humorous video clips on stress and well-being: an experience sampling method-based field experiment. Comput Human Behav Rep 2023 May;10:100270. [doi: 10.1016/j.chbr.2023.100270]
135. Stoffel M, Abbruzzese E, Rahn S, Bossmann U, Moessner M, Ditzen B. Covariation of psychobiological stress regulation with valence and quantity of social interactions in everyday life: disentangling intra- and interindividual sources of variation. J Neural Transm (Vienna) 2021 Sep;128(9):1381-1395 [FREE Full text] [doi: 10.1007/s00702-021-02359-3] [Medline: 34181094]
136. Tate AD, Trofholz A, Youngblood A, Goldschmidt AB, Berge JM. Association between parental resource depletion and parent use of specific food parenting practices: an ecological momentary assessment study. Appetite 2024 Aug 01;199:107368. [doi: 10.1016/j.appet.2024.107368] [Medline: 38643902]
137. Tomko RL, Saladin ME, McClure EA, Squeglia LM, Carpenter MJ, Tiffany ST, et al. Alcohol consumption as a predictor of reactivity to smoking and stress cues presented in the natural environment of smokers. Psychopharmacology (Berl) 2017 Feb;234(3):427-435 [FREE Full text] [doi: 10.1007/s00213-016-4472-x] [Medline: 27838746]
138. Tschacher W, Lienhard N. Mindfulness is linked with affectivity in daily life: an experience-sampling study with meditators. Mindfulness 2021 Apr 01;12(6):1459-1472. [doi: 10.1007/s12671-021-01615-7]
139. Tump D, Narayan N, Verbiest V, Hermsen S, Goris A, Chiu CD, et al. Stressors and destressors in working from home based on context and physiology from self-reports and smartwatch measurements: international observational study trial. JMIR Form Res 2022 Nov 10;6(11):e38562 [FREE Full text] [doi: 10.2196/38562] [Medline: 36265030]
140. Wen H, Sobolev M, Vitale R, Kizer J, Pollak JP, Muench F, et al. mPulse mobile sensing model for passive detection of impulsive behavior: exploratory prediction study. JMIR Ment Health 2021 Jan 27;8(1):e25019 [FREE Full text] [doi: 10.2196/25019] [Medline: 33502330]

XSL•FO
RenderX

141. Wenze SJ, Battle CL, Huntley ED, Gaugler TL, Kats D. Ecological momentary assessment of postpartum outcomes in mothers of multiples: lower maternal-infant bonding, higher stress, and more disrupted sleep. Arch Womens Ment Health 2023 Jun;26(3):361-378 [FREE Full text] [doi: 10.1007/s00737-023-01317-0] [Medline: 37118548]
142. Wilson-Mendenhall CD, Condon P, Hasenkamp W, MacKerron G, Quigley KS, Barrett LF. The Qwantify app dataset: a remote experience sampling study of desire, emotion, and well-being. Front Psychol 2022;13:1054292 [FREE Full text] [doi: 10.3389/fpsyg.2022.1054292] [Medline: 36524170]
143. Wooldridge JS, Soriano E, Filip TF, Moore RC, Eyler LT, Herbert MS. Compassion dynamics in medical students: an ecological momentary assessment study. J Clin Psychol Med Settings 2025 Mar;32(1):51-59. [doi: 10.1007/s10880-024-10003-x] [Medline: 38402300]
144. Xu W, Ding X, Zhuang Y, Yuan G, An Y, Shi Z, et al. Perceived haze, stress, and negative emotions: an ecological momentary assessment study of the affective responses to haze. J Health Psychol 2020 Mar;25(4):450-458. [doi: 10.1177/1359105317717600] [Medline: 28810492]
145. Yap Y, Bei B, Wiley JF. Daily coping moderates the relations between stress and actigraphic sleep: a daily intensive longitudinal study with ecological momentary assessments. Sleep Med 2021 Dec;88:231-240. [doi: 10.1016/j.sleep.2021.10.012] [Medline: 34798439]
146. Yung ST, Chen Y, Zawadzki MJ. Loneliness and psychological distress in everyday life among Latinx college students. J Am Coll Health 2023 Jul;71(5):1407-1416. [doi: 10.1080/07448481.2021.1927051] [Medline: 34242538]
147. Zawadzki MJ, Hussain M, Kho C. Comparing multidimensional facets of stress with social, emotional, and physical well-being using ecological momentary assessment among a Hispanic sample. Stress Health 2022 Apr;38(2):375-387. [doi: 10.1002/smi.3098] [Medline: 34494721]
148. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037/0022-3514.54.6.1063]
149. Lemyre L, Lalande-Markon MP. Psychological Stress Measure (PSM-9): integration of an evidence-based approach to assessment, monitoring, and evaluation of stress in physical therapy practice. Physiother Theory Pract 2009 Jul;25(5-6):453-462. [doi: 10.1080/09593980902886321] [Medline: 19842868]
150. Satow L. SCI - stress- und coping-inventar. Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID). 2012. URL: https://www.psycharchives.org/en/item/8478339e-2308-476e-86d8-3ca4383baec8 [accessed 2026-02-25]
151. Are teens adopting adults' stress habits? American Psychological Association. 2014. URL: https://www.apa.org/news/press/releases/stress/2013/teen-stress [accessed 2026-02-25]
152. Rescorla LA, Achenbach TM. The Achenbach System of Empirically Based Assessment (ASEBA) for ages 18 to 90 years. In: Maruish ME, editor. The Use of Psychological Testing for Treatment Planning and Outcomes Assessment: Instruments for Adults. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 2004:115-152.
153. Jenkins BN, Cross MP, Donaldson CD, Pressman SD, Fortier MA, Kain ZN, et al. The Subcomponents of Affect Scale (SAS): validating a widely used affect scale. Psychol Health 2023;38(8):1032-1055 [FREE Full text] [doi: 10.1080/08870446.2021.2000612] [Medline: 34846253]
154. Eysenck HJ, Eysenck SBG. Eysenck Personality Questionnaire-Revised (EPQ-R). APA PsycTests. 1993. URL: https://psycnet.apa.org/doi/10.1037/t05461-000 [accessed 2026-02-25]
155. Rosa AR, Sánchez-Moreno J, Martínez-Aran A, Salamero M, Torrent C, Reinares M, et al. Validity and reliability of the Functioning Assessment Short Test (FAST) in bipolar disorder. Clin Pract Epidemiol Ment Health 2007 Jun 07;3:5 [FREE Full text] [doi: 10.1186/1745-0179-3-5] [Medline: 17555558]
156. Kim H, Kim S, Kong SS, Jeong YR, Kim H, Kim N. Possible application of ecological momentary assessment to older adults' daily depressive mood: integrative literature review. JMIR Ment Health 2020 Jun 02;7(6):e13247 [FREE Full text] [doi: 10.2196/13247] [Medline: 32484442]
157. May M, Junghaenel DU, Ono M, Stone AA, Schneider S. Ecological momentary assessment methodology in chronic pain research: a systematic review. J Pain 2018 Jul;19(7):699-716 [FREE Full text] [doi: 10.1016/j.jpain.2018.01.006] [Medline: 29371113]
158. Tynan M, Virzi N, Wooldridge JS, Morse JL, Herbert MS. Examining the association between objective physical activity and momentary pain: a systematic review of studies using ambulatory assessment. J Pain 2024 Apr;25(4):862-874 [FREE Full text] [doi: 10.1016/j.jpain.2023.10.021] [Medline: 37914094]
159. Myin-Germeys I, Kasanova Z, Vaessen T, Vachon H, Kirtley O, Viechtbauer W, et al. Experience sampling methodology in mental health research: new insights and technical developments. World Psychiatry 2018 Jun;17(2):123-132 [FREE Full text] [doi: 10.1002/wps.20513] [Medline: 29856567]
160. Stone AA, Schneider S, Smyth JM. Evaluation of pressing issues in ecological momentary assessment. Annu Rev Clin Psychol 2023 May 09;19:107-131 [FREE Full text] [doi: 10.1146/annurev-clinpsy-080921-083128] [Medline: 36475718]
161. Dao KP, De Cocker K, Tong HL, Kocaballi AB, Chow C, Laranjo L. Smartphone-delivered ecological momentary interventions based on ecological momentary assessments to promote health behaviors: systematic review and adapted checklist for reporting ecological momentary assessment and intervention studies. JMIR Mhealth Uhealth 2021 Nov 19;9(11):e22890 [FREE Full text] [doi: 10.2196/22890] [Medline: 34806995]

XSL•FO
RenderX

162. Liao Y, Skelton K, Dunton G, Bruening M. A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: an adapted STROBE Checklist for Reporting EMA Studies (CREMAS). J Med Internet Res 2016 Jun 21;18(6):e151 [FREE Full text] [doi: 10.2196/jmir.4954] [Medline: 27328833]
163. van Roekel E, Keijsers L, Chung JM. A review of current ambulatory assessment studies in adolescent samples and practical recommendations. J Res Adolesc 2019 Sep;29(3):560-577 [FREE Full text] [doi: 10.1111/jora.12471] [Medline: 31573762]
164. Smets E, Rios Velazquez E, Schiavone G, Chakroun I, D'Hondt E, De Raedt W, et al. Large-scale wearable data reveal digital phenotypes for daily-life stress detection. NPJ Digit Med 2018 Dec 12;1:67 [FREE Full text] [doi: 10.1038/s41746-018-0074-9] [Medline: 31304344]
165. Weiss DS. The impact of event scale: revised. In: Wilson JP, Tang CK, editors. Cross-Cultural Assessment of Psychological Trauma and PTSD. Boston, MA: Springer; 2007.
166. Robinson BC. Validation of a Caregiver Strain Index. J Gerontol 1983 May;38(3):344-348. [doi: 10.1093/geronj/38.3.344] [Medline: 6841931]
167. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Med Care 2007 May;45(5 Suppl 1):S3-11 [FREE Full text] [doi: 10.1097/01.mlr.0000258615.42478.55] [Medline: 17443116]
168. Suchikova Y, Tsybuliak N, Teixeira da Silva JA, Nazarovets S. GAIDeT (Generative AI Delegation Taxonomy): A taxonomy for humans to delegate tasks to generative artificial intelligence in scientific research and publishing. Account Res 2026 Apr;33(3):2544331. [doi: 10.1080/08989621.2025.2544331] [Medline: 40781729]

Abbreviations

DASS: Depression Anxiety Stress Scales
DESS: Di-Eu-Stress State
EMA: ecological momentary assessment
EPQ-N: Eysenck Personality Questionnaire Neuroticism
ESM: experience sampling method
HRV: heart rate variability
ICC: intraclass correlation coefficient
K10: Kessler Psychological Distress Scale
PANAS: Positive and Negative Affect Schedule
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSM-9: Psychological Stress Measure-9
PSS: Perceived Stress Scale
VAS: visual analog scale

Edited by L Buis; submitted 26.Feb.2026; peer-reviewed by M Val, D Zavala; comments to author 16.Jun.2026; revised version received 05.Oct.2026; accepted 07.Oct.2026; published 05.Nov.2026.

Please cite as:
Colman T, Van Daele T, Bonroy B, Aerts JM, Debard G, Gielen J, Sels R, Van Oost CN, Faurholt-Jepsen M, De Witte NAJ
Measuring Perceived Stress in Healthy Adults in Daily Life: Scoping Review of Smartphone-Based Self-Assessment
JMIR Mhealth Uhealth 2026;14:e94204
URL: https://mhealth.jmir.org/2026/1/e94204
doi:10.2196/94204
PMID:

©Toon Colman, Tom Van Daele, Bert Bonroy, Jean-Marie Aerts, Glen Debard, Jasper Gielen, Romy Sels, Catharina Nina Van Oost, Maria Faurholt-Jepsen, Nele A J De Witte. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 05.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Daily Negative Affect and Reaction Time Inconsistency in Emerging Adults: Ecological Momentary Assessment Study

Lauren Alexandra Rutter[1], PhD; Pei-Ying Chen[2], PhD; Prabhvir Lakhan[1], BS; Laura Thi Germine[3], PhD; Zoe Hawks[4], PhD

[1]Department of Psychological and Brain Sciences, Indiana University, 1101 E 10th Street, Bloomington, IN, United States

[2]Digital Scholarship & Sciences, University of Nebraska-Lincoln Libraries, Lincoln, NE, United States

[3]Division of Brain and Cognitive Health Technology, Department of Neurolgy, Albert Einstein College of Medicine, Bronx, NY, United States

[4]Brooklyn Health, Brooklyn, NY, United States

Corresponding Author:
Lauren Alexandra Rutter, PhD
Department of Psychological and Brain Sciences, Indiana University, 1101 E 10th Street, Bloomington, IN, United States

Abstract

Background: Anxiety and mood disorders, characterized by elevated negative affect (NA) and cognitive impairments, are highly prevalent among college students. Within-person (WP) NA variability, which captures moment-to-moment fluctuations in NA, provides unique insights into emotional processes that are not reflected in mean NA levels. Cognitive variability, particularly reaction time (RT) inconsistency, is increasingly recognized as a sensitive marker of cognitive health and functional integrity. Although prior research links NA to cognitive variability, the short-term dynamics of these associations in naturalistic settings remain understudied. College students provide an ideal population for examining these dynamics using ecological momentary assessment (EMA).

Objective: This study investigated the association between WP NA and RT inconsistency, hypothesizing that higher WP fluctuations in NA would predict increased RT inconsistency. We also examined the moderating roles of practice effects and covariates, including neuroticism, insomnia, and sex.

Methods: Using EMA, 99 university students completed morning and evening assessments over 14 days, including a cognitive task measuring RT inconsistency (standard deviation in trial-level RT) and self-reported NA. Multilevel modeling was used to assess WP fluctuations in NA and their impact on RT inconsistency, accounting for time (session number), between-person differences in NA, and covariates such as sleep problems, neuroticism, age, sex, and use of a touch device.

Results: WP fluctuations in NA significantly predicted increased RT inconsistency (exp(β)=1.022, 95% CI 1.008 - 1.037; *P*=.007), supporting the hypothesis that NA variability disrupts cognitive performance. Male students exhibited lower RT inconsistency than female students, with a small effect size (exp(β)=0.824, 95% CI 0.694 - 0.977; *P*=.049). Finally, EMA sessions were inversely associated with RT inconsistency, with a stronger effect up to session 3 (exp(β)=0.930, 95% CI 0.879 - 0.985; *P*=.03) than after session 3 (exp(β)=0.986, 95% CI=0.979 - 0.992; *P*<.001), indicating practice effects.

Conclusions: Momentary fluctuations in NA influence cognitive variability, particularly in the early stages of repeated cognitive tasks, underscoring the role of emotional processes in cognitive performance. Practice effects and individual differences, such as sex and insomnia, influence these associations. These findings highlight the use of EMA for understanding cognitive-affective processes and suggest potential intervention targets, such as addressing NA, to improve cognitive functioning in emotionally vulnerable populations like college students.

(JMIR Mhealth Uhealth 2026;14:e64397) doi:10.2196/64397

KEYWORDS

ecological momentary assessment; negative affect; cognition; emotions; questionnaire; neuroticism; anxiety; depression; insomnia; stress

Introduction

Anxiety and mood disorders, also known as emotional disorders, are the most common mental disorders and leading contributors to the global burden of disability [1,2]. Emotional disorders are often chronic and highly comorbid due to their shared features, including a high level of negative affect (NA) and strong reactivity to NA, also known as neuroticism [3]. Elevated NA is typically accompanied by both affective dysregulation and other cognitive impairments such as lower executive functioning, reduced attention capacity and processing speed, and impaired short-term memory, all of which are associated

XSL•FO
RenderX

with the severity of emotional disorders and the likelihood of relapse [4,5]. NA has been extensively linked to emotional disorders, yet its momentary fluctuations and their cognitive implications remain underexplored. This study bridges this gap by using ecological momentary assessment (EMA) to examine the real-time interplay between NA and cognitive variability.

Recent research using EMA has examined within-person (WP) variation in NA. Broadly, this body of work demonstrates that dynamic fluctuations in NA are associated with increased psychiatric symptoms [6-9]. Emotional inertia, or the tendency for emotional states to persist over time, represents another critical dimension of affective dynamics [10]. Prolonged negative emotional states may indicate difficulties in emotional recovery, contrasting with high emotional instability, which reflects rapid fluctuations in affect. Both extremes of these dynamics are linked to the risk of mood disorders [11]. EMA allows researchers to measure both affect inertia and affect fluctuations in real-time, offering a nuanced perspective on these emotional processes and their cognitive implications. This study focuses on momentary WP fluctuations in NA, which have been consistently linked to psychiatric symptoms. In prior EMA research, in a study of 40 individuals with depression and 40 healthy controls, participants with depression exhibited more variable NA [8]. In another EMA study (N=365), participants with a current depression and/or anxiety disorder had higher NA fluctuations, suggesting sensitivity to internal and external sources of stress coupled with poor emotion regulation skills [11]. Based on a meta-analysis of 79 studies on emotion dynamics and psychological well-being, low psychological well-being tends to occur when there are more variable and unstable emotions, especially negative emotions [7].

NA is consistently linked to cognition, but the mechanisms accounting for these relationships remain undercharacterized and poorly understood [12]. One theory is that the associations between NA and cognitive ability may reflect the underlying integrity of the nervous system and the efficiency of information processing [12]. Efficient information processing refers to the ability to quickly and accurately process stimuli and integrate them with appropriate responses. Furthermore, efficient information processing could explain why better cognitive ability is related to better physical and mental health [12]. However, the naturalistic, short-term (eg, day-to-day) time course of WP associations between NA and cognition remains unclear. This study integrates Barlow et al's [3] framework on neuroticism and affective variability and Jokela's [12] theory on neural efficiency to establish a unified understanding of how fluctuations in NA influence cognitive performance. Reaction time (RT) inconsistency serves as a sensitive indicator of these cognitive-affective dynamics, reflecting momentary fluctuations in attention and processing speed. By using EMA, we aim to elucidate the real-world implications of these theoretical constructs.

WP variation in cognition has been discussed in the literature in several ways, including (1) variability across tasks measured at a single time point, known as dispersion; (2) variability across trials within a single task, known as inconsistency; and (3) variability over time across repeated task administrations, referred to as fluctuations [13]. Cognitive dispersion and inconsistency have been shown to increase with age and may serve as important indicators of cognitive functioning across the lifespan [14-16]. In comparison to the number of studies on variation in cognition in older adults, much less is known about physical and mental health-related correlates of WP variation in cognition in emerging adults. Our study targets WP variation in cognition in college-age students using RT-based measurement, specifically, standard deviation in trial-level RTs. Standard deviation in RT, which we refer to as "RT inconsistency," is increasingly accruing evidence as an indicator of cognitive health, which also includes domains of executive functioning, attention, and memory [17]. RT inconsistency has potential for furthering our understanding of dynamic processes in health, cognition, and affect [18,19]. As a WP metric that can be readily assessed using EMA, RT inconsistency may yield novel insights in emerging adults who are experiencing stress and NA in their daily lives. Moreover, RT inconsistency, rather than mean RT, can represent a dimension of cognitive performance that is a dispositional characteristic reflecting underlying changes in psychological processes or neurological functioning [19]. However, while RT inconsistency is a valuable indicator of cognitive health, it is also important to study other factors, such as sleep, fatigue, and other learned behaviors. Our study controls for baseline insomnia and neuroticism in assessing the dynamic relationship between RT inconsistency and NA variability.

College students face significant academic, social, and personal stressors, including high levels of academic pressure, the transition to independent living, and the challenge of navigating new social environments [20]. These stressors contribute to emotional dysregulation, particularly fluctuations in NA, which can impair cognitive functioning such as attention and memory [21]. Furthermore, college students are a population with high rates of emotional disorders and treatment-seeking behaviors [20], making them an ideal group for studying the dynamic interplay between NA and cognition. These developmental transitions, along with the increased prevalence of anxiety and mood disorders, underscore the importance of understanding how fluctuations in affect influence cognitive performance during this critical period of young adulthood [22]. In a study of older and young adults who reported on daily stress and completed daily tasks of cognitive performance, WP variability in stress predicted cognitive fluctuations, evidenced by WP RT variability on a 2-back working memory task in both groups [23]. This is consistent with the attention-depletion hypothesis, which posits that stress-related cognitive interference disrupts attentional resources, weakening attention-dependent cognitive processing [23,24].

While the impact of stress on college students, the impact of NA on higher-level cognition [25,26], and the associations among stress, NA, and cognition in individuals with emotional disorders (see review by Pechtel and Pizzagalli [27]) have been well-established, there remains a surprising lack of published research using EMA to understand dynamic, short-term, WP connections among these factors. A better understanding of the mechanisms that connect momentary changes in affect and momentary changes in cognition is now possible through EMA. Specifically, a better understanding of RT inconsistency and its

XSL•FO
RenderX

relationship with NA may advance our understanding of psychological functioning and cognitive health [13,17]. Understanding the relationships between cognition, NA, and emotional disorder symptoms in college-age students and emerging adults could aid in providing targeted directions for treatments of individuals vulnerable to stress, NA, and disrupted cognitive performance.

In summary, symptoms of emotional distress are associated with elevated NA [3,28], greater NA variability and instability [29-31], and poorer cognitive performance [32]. However, NA and cognition are not typically measured together in high-frequency study designs, which obscures insights into the momentary effects of NA on cognition. This study addressed this gap by examining WP associations between NA and RT inconsistency in university students. Participants completed daily EMAs measuring NA and cognition. For each EMA, RT inconsistency was operationalized as the standard deviation in trial-level RTs [15,33,34]. RT inconsistency was chosen as the primary measure due to its sensitivity to moment-to-moment fluctuations in cognitive processing. Compared to other measures, such as working memory or attention switching, RT inconsistency captures variability in attentional control and neural efficiency [12], making it particularly relevant for studying the cognitive impact of NA.

Using a 14-day EMA design, we examined WP associations between NA and RT inconsistency in 99 university students. We hypothesized that momentary WP increases in NA would be associated with greater RT inconsistency at subsequent assessments. As a secondary hypothesis, we examined whether individual differences in sex, neuroticism, and insomnia severity were associated with RT inconsistency. Examining these processes in daily life may clarify mechanisms linking emotional distress to functional impairment.

Methods

Participants

Participants were sampled from (1) a psychology student subject pool at a large Midwestern university (n=78) and (2) a paid study in which individuals were recruited through university online classified advertisements and flyers posted around campus (n=72). Inclusion criteria included being a university student aged 18 years or older with an email address, access to the internet, and a device to complete study assessments. Exclusion criteria included individuals younger than 18 years.

Ethical Considerations

This study was reviewed and approved by the Indiana University Institutional Review Board as human subjects research (protocol numbers 2011722518 and 12537). All study procedures complied with institutional and federal guidelines for the ethical conduct of research involving human participants. Informed consent was obtained electronically from all participants prior to study enrollment. Participants were informed that participation was voluntary and that they could withdraw from the study at any time without penalty. Informed consent covered both the baseline assessment and the subsequent 14-day EMA protocol. To protect the privacy and confidentiality of participants, all data were deidentified. Identifying data (email addresses used for study communication) were stored separately from survey and cognitive task data on secure servers accessible only to the research team via 2-factor authentication. Participants recruited through the psychology subject pool received course credit for participation. Participants recruited through paid advertisements were compensated up to US $60 for completing at least 75% (18/24) of EMA assessments. No additional incentives were provided.

Procedure

The 14-day EMA protocol was chosen to capture meaningful WP variability while maintaining participant retention, as longer durations may lead to increased dropout rates and data loss [35]. The frequency of daily assessments (2 - 13 d) was selected to balance participant burden with the need to capture fluctuations in NA and cognitive performance. This schedule ensures adequate data collection while minimizing participant fatigue, which could compromise the quality of responses. The morning and evening sessions are consistent with prior EMA studies [23], and they allow us to assess affective and cognitive fluctuations at multiple time points.

Participants completed all study assessments on their personal devices. They were prompted to complete EMAs twice daily via email. We assessed for technical issues after each EMA and explored device compatibility in our prior work [36]. All assessments for the study were completed via Qualtrics and TestMyBrain.org. TestMyBrain.org is a citizen science research initiative that allows people to participate in studies in exchange for individualized feedback on their performance or self-reported characteristics [36-38]. Data from TestMyBrain.org have been shown to be of similar high quality to data gathered in a laboratory setting [37]. On day 1 (baseline), participants were emailed at 7 AM to complete their baseline assessment battery, and they were sent a reminder email at 5 PM to complete their day 1 battery by midnight. They were advised to complete the assessment battery when they had approximately 45 minutes of uninterrupted time, specifically, “in 1 sitting, without distraction.” On the following days, participants received daily links to their email addresses to complete study assessments in the morning and evening each day. Morning study reminders were sent at 7 AM and evening study reminders were sent at 4 PM. Participants were reminded to complete EMAs on days 2 to 13 when they had 5 minutes of uninterrupted time. On day 14 (termination), participants were sent links to complete their assessment battery at 7 AM, with a reminder at 5 PM to complete the assessments before midnight. They were advised to complete the battery when they had approximately 35 minutes of uninterrupted time. When participants missed more than 2 days in a row of tests, they were reminded by email to complete as many assessments as possible. Participants were compensated, either by course credit or electronic gift cards, approximately 2 days following the study termination.

Measures

This study is part of a larger study measuring cognition and emotions over time. Relevant assessments for this study are included in Textbox 1.

XSL•FO
RenderX

Textbox 1. Schedule of ecological momentary assessments.

Ecological momentary assessment days 1 - 14

- Day 1 (baseline)
  - Demographic questionnaire
  - Depression, Anxiety, and Stress (DASS)
  - Generalized Anxiety Disorder-7 (GAD-7)
  - Positive and Negative Affect Schedule—Short Form (PANAS-SF)
  - Multidimensional Emotional Disorder Inventory (MEDI)
  - Insomnia Severity Index (ISI)
  - Choice Reaction Time (Choice RT)
- Days 2 - 13
  - Morning (7 AM-12 PM)
    - PANAS-SF
    - Choice RT
  - Evening (4 PM-12 AM)
    - PANAS-SF
- Day 14
  - DASS
  - GAD-7
  - PANAS-SF
  - MEDI
  - Choice RT

Positive Affect and Negative Affect Schedule—Short Form (PANAS-SF)

The PANAS-SF is a widely used and extensively validated 20-item self-report scale assessing positive affect (PA) and NA [39]. Items are rated on a 5-point Likert scale ranging from 1 ("Very slightly or not at all") to 5 ("Extremely"), yielding scores from 10 to 50 for PA and 10 to 50 for NA. Higher scores indicate higher levels of affect for both subscales. The PANAS sampling schedule is described in Table 1. On day 1, participants were instructed to complete the PANAS based on how they felt in the past week. On days 2 to 14, participants were instructed to complete the PANAS based on how they felt at the time of EMA (ie, "right now"). In our sample, internal consistency was acceptable (Cronbach α=0.75).

XSL•FO
RenderX

Table . Sample characteristics of participants who completed any ecological momentary assessments after baseline (day 1).

Variable	More than 50% (n=99)	Less than 50% (n=17)	*P* value[a]
Sample, n (%)			.10
Student	37 (37)	10 (59)	
Paid	62 (63)	7 (41)	
Age (y), mean (SD; range)	22.9 (7; 18.0-56.0)	21.9 (7.3; 18.0-47.0)	.07
Sex, n (%)	74 (75)	9 (53)	.08
Neuroticism[b], mean (SD; range)	16 (8; 1-33)	15 (9; 0-29)	.6
DASS-Depression[c], mean (SD; range)	9 (9; 0-38)	10 (11; 0-40)	.9
DASS-Anxiety[d], mean (SD; range)	8 (7; 0-38)	6 (8; 0-30)	.2
GAD-7[e] total, mean (SD; range)	6.2 (4.9; 0.0-21.0)	5.5 (4.7; 0.0-19.0)	.6

[a]Pearson chi-square test, Wilcoxon rank-sum test, and Fisher exact test.

[b]Multidimensional Emotional Disorder Inventory Neuroticism scale.

[c]DASS-Depression: Depression Anxiety and Stress Scale—Depression subscale.

[d]DASS-Anxiety: Depression Anxiety and Stress Scale—Anxiety subscale.

[e]GAD-7: Generalized Anxiety Disorder-7.

Insomnia Severity Index

The Insomnia Severity Index is a 7-item self-report scale validated for measuring the insomnia symptom severity over the previous 2 weeks [40-42]. Items are rated on a 5-point Likert scale ranging from 0 (“none”) to 4 (“very severe”), yielding total scores of 0 to 28. Total scores are interpreted such that 0 to 7 indicates the absence of insomnia, 8 to 14 indicates subthreshold insomnia, 15 to 21 indicates moderate insomnia, and 22 to 28 indicates severe insomnia. In the sample, internal consistency was good (Cronbach α=0.88).

Depression Anxiety and Stress Scales

The Depression Anxiety and Stress Scales (DASS) is a 21-item measure rated on a 4-point Likert scale from 0 (“does not apply to me at all”) to 3 (“applied to me very much, or most of the time”) [43]. It assesses levels of depression, generalized anxiety, and general NA or stress. Based on factor analytic work, the DASS demonstrates excellent psychometric properties [42,43]. In our sample, internal consistency was good (Cronbach α=0.89).

Multidimensional Emotional Disorder Inventory

The Multidimensional Emotional Disorder Inventory (MEDI) assesses transdiagnostic dimensions of emotional disorder psychopathology [44]. It consists of 49 items that are rated on a 9-point Likert scale from 0 (“Not characteristic of me/does not apply to me”) to 8 (“Extremely characteristic of me/applies to me very much”). The initial validation of the MEDI used factor analysis to determine the factor structure of 9 empirically supported transdiagnostic dimensions proposed by Rosellini and Brown [44]. These 9 dimensions include neurotic temperament, positive temperament, depression, autonomic arousal, somatic anxiety, social anxiety, intrusive cognitions, traumatic re-experiencing, and avoidance. The MEDI showed strong evidence of reliability and validity in its initial psychometric evaluation in clinical and outpatient samples [43]. We used the MEDI neuroticism subscale as our measure of neuroticism, with higher scores indicating a higher degree of neuroticism. In our sample, internal consistency was excellent (Cronbach α=0.94).

TestMyBrain Choice RT

TestMyBrain choice RT measures psychomotor response speed, response selection, and cognitive inhibition [45,46]. It was adapted for brief digital administration from Maljkovic and Nakayama [47] and demonstrates high between-person (BP) and WP reliability in EMA [48]. Each administration lasted approximately 2.5 minutes and included 4 practice trials followed by 24 task trials. On each trial, participants are presented with 3 vertically aligned squares (blue or yellow) containing an arrow pointing either left or right. One of the 3 squares is always of a different color from the other 2, and participants are asked to choose the direction of the arrow in the differently colored square as quickly as possible by clicking or tapping their touchscreen. Participants had 5000 ms to respond to each trial, with an intertrial interval of 700 to 1500 ms. RT inconsistency was computed for each EMA as the standard deviation of trial-level RTs.

Data Cleaning

EMA data cleaning aimed to maximize the signal-to-noise ratio by excluding data that were strongly suggestive of careless responding. From our full sample of 150 individuals, we first excluded participants who failed to complete EMAs after the baseline assessment (day 1). The remaining sample included 116 participants. Second, we excluded participants who failed to complete the morning EMA, including both choice RT and PANAS, in at least 50% (7/14) of assessments. Finally, we excluded EMA data when choice RT accuracy was less than 50% (ie, chance).

XSL•FO
RenderX

Missing data primarily arose from missed EMA assessments. Rather than imputing missing values, we used listwise deletion at the observation level within mixed-effects models, such that only available observations contributed to each model. In this study, missingness reflected missed EMA prompts rather than partially completed surveys, meaning that entire measurement occasions were absent rather than individual items within completed assessments. For this reason, we prioritized analyzing observed EMA data using mixed-effects models rather than imputing unobserved moments. Sensitivity analyses comparing models applying stricter data-cleaning criteria showed consistent results, suggesting minimal bias introduced by this criterion. After these cleaning procedures, 99 participants were included in the analyses.

Data Analyses

Prior to analysis, we conducted a power analysis to evaluate the sample size needed to detect WP associations between NA and RT inconsistency in multilevel models. Based on prior similar EMA studies, we anticipated effects in the small-to-moderate to moderate range [7,23]. Assuming a medium effect size (f^2=0.15), a 2-sided significance α level of .05, and a desired power of 0.80, a sample of approximately 80 to 90 participants with repeated observations was estimated to be sufficient to detect the primary WP effect. The final analytic sample included 99 participants contributing 1177 EMA observations, which provided adequate power to detect medium-sized WP effects. However, given early attrition and noncompliance, the study may be underpowered to detect small effects. Accordingly, nonsignificant findings should be interpreted cautiously.

Data were analyzed in RStudio Version 2023.06.0+421 using *tidyverse* for data wrangling and visualization (version 1.3.0) [49] and *lme4* for multilevel modeling (version 1.1 - 26) [50]. First, we standardized DASS and MEDI neuroticism for ease of interpretation. Next, general linear mixed models (ie, multilevel models where sessions were modeled as nested within persons) were used to examine the impact of WP fluctuations in NA on RT inconsistency, accounting for time (ie, session) in the study, BP differences in NA, sex, age, neuroticism, and insomnia.

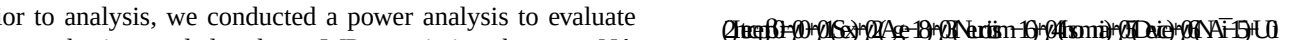
Predictors were entered in stages to quantify incremental variance explained, and intercepts were allowed to vary for each participant (ie, as random effects). In model 0, we ran an empty means, random-intercept-only model to examine the overall pattern of individual differences in RT inconsistency. In model 1, a 2-piece discontinuous model of change by describing the change before and after EMA session 3 as 2 separate linear slopes was fitted to capture the apparent breakpoint in session 3 (Multimedia Appendix 1). In model 2, NA was added at levels 1 and 2. In model 3, level 1 and cross-level interactions were added between EMA session and NA. In model 4, sex, age, neuroticism, insomnia, and response device were added at level 2. These covariates were included to account for potential differences in cognitive performance related to maturation, sex, and the device used.

Comparison of model fit across models, including an alternative model describing change as a continuous slope, suggests that model 4 fit the data best based on the likelihood ratio test. The diagnosis of model assumptions (linearity, constant and normal distribution of variance, and independence of residuals across levels) for model 4, along with a check of collinearity and autocorrelation as well as model fit comparisons, is provided in Multimedia Appendix 1. Model 4 equations are provided below:

Level 1 (observation level)



Level 2 (person level)



(3)Slope 1: $\beta 1i=\gamma 10+\gamma 11(\overline{NA}i-15)$

(4)Slope 2: $\beta 2i=\gamma 20+\gamma 21(\overline{NA}i-15)$

(5)WP NA:$\beta 3i=\gamma 30$

(6)WP NA by Slope 1:$\beta 4i=\gamma 40$

(7)WP NA by Slope 2:$\beta 5i=\gamma 50$

Results

Of the 150 recruited participants, 22.7% did not complete any EMAs and are considered early dropouts (n=34). Early dropout was defined as completing a portion of the day 1 baseline measures but not beginning EMAs over the next 13 days. The majority of early dropouts were participants from the psychology study pool.

We compared the study pool and paid study samples before and after excluding early dropouts. Before excluding early dropouts, compliance and age differed between the samples (Table 2). The study pool sample tended to be younger and less compliant. As shown in Table 2, 40% (31/78) of the student sample dropped out after day 1, compared to only 4.2% (3/72) of the paid sample. After excluding early dropouts, age no longer differed between the samples (Table 1). Sex, levels of neuroticism, depression, and anxiety did not differ between the samples (Tables 1 and 2). This indicates that once participants began completing the study EMAs, they were not significantly more likely to be compliant if they were older or experiencing less depression and anxiety. The largest dropout occurred before beginning the EMA. Participants in the paid study were more likely to complete more than 50% of EMAs than those in the student sample. Morning assessments were chosen for data analyses to minimize potential confounds introduced by daily fatigue cycles or stress accumulation. Despite a dropout rate of 22.7% (34/150), comparisons between completers and noncompleters showed no significant differences in baseline characteristics, reducing concerns about systematic bias.

XSL•FO
RenderX

Table . Sample characteristics of participants who completed the baseline assessment (day 1; N=150).

Variable	Student (n=78)	Paid (n=72)	*P* value[a]
Group, n (%)			<.001
Dropout	31 (40)	3 (4.2)	
<50	10 (13)	7 (9.7)	
>50	37 (47)	62 (86)	
Age (y; n=148), mean (SD; range)	20.0 (3.7; 18.0-47.0)	24.4 (7.9; 19.0-56.0)	<.001
Sex (n=148), n (%)	60 (77)	50 (71)	.4
Neuroticism[b] (n=148), mean (SD; range)	16 (8; 0-34)	16 (9; 0-33)	.9
DASS-Depression[c] (n=148), mean (SD; range)	10 (10; 0-42)	9 (9; 0-40)	.9
DASS-Anxiety[d] (n=148), mean (SD; range)	8 (8: 0-38)	7 (7; 0-34)	.5
GAD-7[e] total (n=148), mean (SD; range)	6.6 (5.3; 0.0-21.0)	6.1 (4.4; 0.0-19.0)	.8

[a]Pearson chi-square test and Wilcoxon rank-sum test.

[b]Multidimensional Emotional Disorder Inventory Neuroticism scale.

[c]DASS-Depression: Depression Anxiety and Stress Scale—Depression subscale.

[d]DASS-Anxiety: Depression Anxiety and Stress Scale—Anxiety subscale.

[e]GAD-7: Generalized Anxiety Disorder-7.

Next, we examined WP associations between NA and RT inconsistency. Results for models 1 to 4 are shown in Table 3. The intraclass correlation coefficient for RT inconsistency was 0.45 from model 0 (results not shown), indicating that 45% of its variance was due to BP mean differences. RT inconsistency decreased by 10.3% between sessions 1 and 2, but only by 1.3% after session 3 (model 1). Model 2 introduced both BP and WP NA as the main effects, while model 3 included their interaction with both slopes. Including interaction terms improved the model fit ($\Delta\chi^2_4$=11.58; *P*=.02).

We focus our interpretation on model 4, which examines whether RT inconsistency is predicted by WP fluctuations in NA, controlling for sex, age, insomnia level, and the response device used. Consistent with hypotheses, WP fluctuations in NA predicted RT inconsistency. Specifically, RT inconsistency increased by 2.2% when NA was higher than usual. Having a higher-than-usual NA during the first 2 sessions seems to attenuate its effect by 1.1%, but not so afterward. In addition, we observed significant main effects of our control variables: RT inconsistency was nearly 25% lower among individuals with moderate-to-severe insomnia, but using touch devices increased RT inconsistency by 11.8%. RT inconsistency was also 17% higher for women compared to men (*P*=.049).

XSL•FO
RenderX

Table . Results of general mixed-effects models predicting reaction time (RT) inconsistency[a].

Predictors	Model 1[b], RT inconsistency[c]			Model 2[d], RT inconsistency			Model 3[e], RT inconsistency			Model 4[f], RT inconsistency		
	Estimates	95% CI	*P* value	Estimates	95% CI	*P* value	Estimates	95% CI	*P* value	Estimates	95% CI	*P* value
(Intercept)	205.253	186.422 - 225.985	<.001	203.399	183.909 - 224.954	<.001	184.921	164.387 - 208.020	<.001	202.457	173.153 - 236.721	<.001
Sessions 1 - 3	0.892	0.855 - 0.931	<.001	0.897	0.856 - 0.939	<.001	0.944	0.893 - 0.999	.08	0.930	0.879 - 0.985	.03
Sessions 3 - 14	0.987	0.981 - 0.994	<.001	0.987	0.981 - 0.994	<.001	0.986	0.979 - 0.993	<.001	0.986	0.979 - 0.992	<.001
morningNAbp[g]	—[i]	—	—	0.993	0.972 - 1.013	.58	0.989	0.961 - 1.017	.49	1.002	0.971 - 1.033	.91
morningNAwp[j]	—	—	—	1.002	0.996 - 1.007	.58	1.023	1.008 - 1.037	.005	1.022	1.008 - 1.037	.007
Sessions 1 - 3 × morningNAbp	—	—	—	—	—	—	1.005	0.993 - 1.018	.49	1.005	0.993 - 1.018	.56
Sessions 3 - 14 × morningNAbp	—	—	—	—	—	—	0.999	0.997 - 1.001	.49	0.999	0.997 - 1.001	.56
Sessions 1 - 3 × morningNAwp	—	—	—	—	—	—	0.989	0.980 - 0.999	.07	0.989	0.979 - 0.998	.047
Sessions 3 - 14 × morningNAwp	—	—	—	—	—	—	0.999	0.997 - 1.001	.53	1.000	0.998 - 1.002	.74
Male	—	—	—	—	—	—	—	—	—	0.824	0.694 - 0.977	.049
Age[k]	—	—	—	—	—	—	—	—	—	1.004	0.994 - 1.014	.56
Neuroticism[l]	—	—	—	—	—	—	—	—	—	1.003	0.993 - 1.013	.66
Mild insomnia	—	—	—	—	—	—	—	—	—	0.912	0.786 - 1.059	.38
Moderate-to-severe insomnia	—	—	—	—	—	—	—	—	—	0.757	0.622 - 0.921	.02
Touch device	—	—	—	—	—	—	—	—	—	1.118	1.045 - 1.195	.006

[a]Estimates are exponentially transformed, and *P* values are adjusted for false discovery rate.

[b]Random effects: residual variance (σ^2): 0.14; random-intercept variance ($\tau 00$): 0.12 id (participant-level grouping variable for the random effect); intraclass correlation: 0.45; N: 99 id; observations: 1177; marginal R2/conditional R2: 0.037/0.467.

[c]RT inconsistency: standard deviation in reaction time (log-transformed).

[d]Random effects: σ^2: 0.14; $\tau 00$: 0.12 id; intraclass correlation: 0.45; N: 99 id; observations: 1177; marginal R2/conditional R2: 0.039/0.469.

XSL•FO
RenderX

[e]Random effects: σ^2: 0.14; τ00: 0.11 id; intraclass correlation: 0.45; N: 99 id; observations: 1177; marginal R2/conditional R2: 0.044/0.471.

[f]Random effects: σ^2: 0.14; τ00: 0.10 id; intraclass correlation: 0.42; N: 99 id; observations: 1177; marginal R2/conditional R2: 0.115/0.483.

[g]NA: negative affect.

[h]BP: between-person centered at 15).

[i]Not applicable.

[j]WP: within-person.

[k]Age: centered at 18.

[l]Neuroticism: Multidimensional Emotional Disorder Inventory Neuroticism subscale centered at 16.

Discussion

Principal Findings

Attention dysregulation is a cardinal symptom of emotional disorders [51-53], and prior work demonstrates that trial-level RT variability reflects attentional control processes and underlying information processing efficiency [12,23,24,34]. Using a 14-day EMA with repeated cognitive (choice RT) and affective (PANAS) measures, the current study examined whether WP fluctuations in NA (deviations from an individual's mean NA) were associated with RT inconsistency in daily life. This approach allowed for an examination of the dynamic associations between cognition and affect as they unfold in naturalistic settings.

Consistent with our primary hypothesis, momentary WP fluctuations in NA predicted greater RT inconsistency, even after accounting for BP differences in NA, neuroticism, insomnia, age, sex, and device type. Additionally, individual differences in sex and insomnia severity were associated with RT inconsistency, partially supporting our secondary hypothesis. Together, these findings suggest that acute increases in NA relative to an individual's typical affect may disrupt attention and information processing above and beyond sleep problems or individual differences in emotional distress. These results extend prior work linking depression, anxiety [54,55], and stress [56-58] to cognitive complaints in young adults by showing that NA changes influence cognition at the momentary, WP level. Notably, the association between NA fluctuations and RT inconsistency was strongest during the early EMA sessions, suggesting that momentary affective states may be particularly disruptive to cognitive performance before participants become familiar and comfortable with a task.

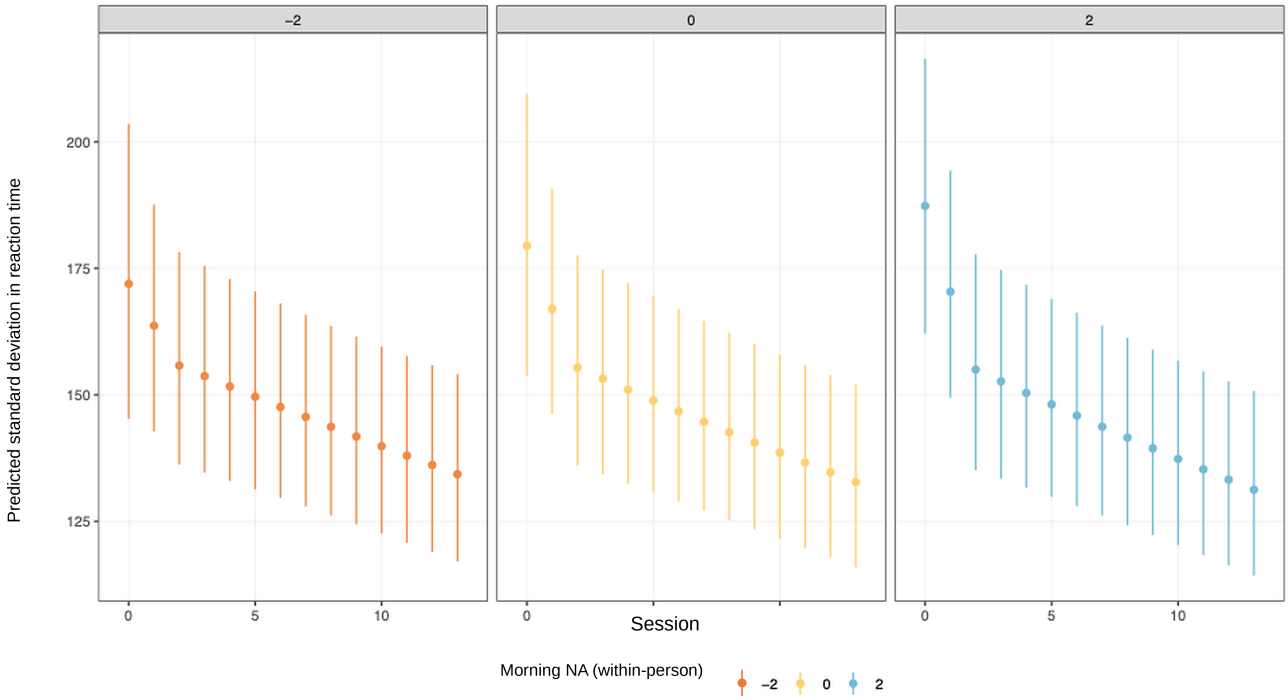
Potential Mechanisms and Interpretation

There are several potential mechanisms that account for the association between momentary NA fluctuations and RT inconsistency. Much like anxiety's effects on performance [59], it is possible that the effects of NA on cognition depend on perceptions of task effort and difficulty, which may be influenced by practice. Early in the study, elevated NA fluctuations may more strongly tax attentional resources, whereas repeated exposure to the same tasks may reduce cognitive load. This is speculative and requires further testing.

An alternative explanation involves speed-accuracy tradeoffs. For example, at the beginning of the study, participants experiencing low and high NA may similarly prioritize accuracy. At the end of the study, participants experiencing high NA may be more likely to prioritize speed by rapidly tapping the screen, thereby reducing RT inconsistency [60-63]. Indeed, the interaction between earlier sessions and NA, while statistically significant, highlights the dynamic nature of cognitive performance over time. An interaction plot (Figure 1) illustrates how RT inconsistency changes across sessions at varying levels of NA, providing insights into the moderating role of practice. Future research is required to replicate these findings, test competing explanations for the interaction between NA fluctuations and practice, and examine generalizability in relation to other cognitive domains. Additionally, future work could incorporate multiple equivalent cognitive tasks to reduce task-specific learning effects and allow for a more accurate assessment of cognitive variability.

XSL•FO
RenderX

Figure 1. Predicted values of standard deviation reaction time (centered) based on within-person NA and session. NA: negative affect.



Strengths and Limitations

There are many strengths of the study and EMA approaches in general. EMA offers the opportunity to measure daily cognitive functioning in naturalistic environments, which can enhance our understanding of contemporaneous, WP relationships between cognitive and affective processes in clinical and community samples. In clinical assessments of cognition, there may be nuances that emerge outside the context of formal assessment, prior to the onset of a disorder. To be most useful, assessments of cognition should be easy to obtain in real-world settings at multiple points across time, but few validated tests exist. Our study used brief, repeated cognitive tasks that have been validated by our team [37,45,48,64-66] and are reliably administered on any mobile device (laptop, desktop, iPad, tablet, iPhone, and Android). Moreover, our approach used assessment at multiple timepoints, repeatedly, which allowed for measurement of intraindividual RT inconsistency across several timepoints, which is increasingly recognized as a warning sign of future cognitive decline and poorer prognosis. This approach allows for the detection of very subtle cognitive changes associated with depression and anxiety, as well as with aging-related changes in cognitive functioning.

Despite the many strengths of this study, several limitations should be considered. First, our sample was predominantly female and White, which limits generalizability. Future research should aim for greater diversity to validate these results in broader populations, including diverse age, gender, and socioeconomic groups. Our study also consisted of college-age, emerging adults [22]. The high dropout rate and reliance on psychology students as participants limit the generalizability of our findings. While dropout was partially mitigated by comparing baseline characteristics, the homogeneity of the sample restricts applicability to broader populations and reduces the sample size for analyses. In our study, paid participants showed lower dropout rates than psychology pool students. Participant fatigue and lack of strong incentives may have influenced results. Missing EMA assessments may introduce bias if missingness is systematically related to participants' affect or cognitive performance. We did not formally test whether missingness was completely at random, which limits our ability to rule out systematic patterns of missingness. However, our sample was primarily impacted by noncompliance in the EMA, which could be addressed by more frequent and personalized reminder emails or a more thorough introduction to the study during the consent process. Participants were instructed to complete the assessments when they were not distracted, but we cannot confirm that occurred, and participants completed assessments at inconsistent timings based on their schedules. In future work, text messages may prompt more immediate responses than email reminders.

As with other EMA, a lack of control over the participants' environment, device type, and network speed may have impacted our results [36,67]. A lack of experimental control is both a benefit and a limitation of EMA. Unfortunately, the study cannot determine if participants were actually trying their best or performing for course credit in an apathetic manner, and rely on data cleaning to ensure for data quality. Conducting EMA in real-world settings introduces variability due to environmental distractions (eg, noise and multitasking), device differences, and participant engagement. These factors, while enhancing ecological validity, may obscure the true relationship between NA and RT inconsistency. Future studies should consider standardized device protocols and environmental monitoring, as well as more self-reported focus checks.

Another important limitation of our design is that our choice RT task was repeated each day from days 2 to 13. Although our ongoing work addresses this, in this study, we did not alter stimuli in each EMA, which challenges efforts to disentangle

XSL•FO
RenderX

the effects of task-specific learning (ie, practice effects) from the effects of memory for task stimuli and associated responses. Practice effects may also arise from factors that we did not control for, such as participant fatigue and task repetition. Future studies should administer multiple psychometrically equivalent cognitive tasks during EMA to mitigate the effects of memory on performance [68]. Isolating the effects of learning on cognitive performance holds future promise to inform digital risk phenotyping. For example, emerging evidence in older adults suggests that individual differences in learning rate and peak (asymptotic) cognitive performance are related to biological age and clinical impairment [69].

As a final limitation of our study, focusing solely on RT inconsistency as a measure of cognitive performance variability may be too narrow. Cognitive variability can be captured by various other metrics, and additional cognitive tasks can provide a more comprehensive assessment of cognitive functioning over time. We chose RT inconsistency because it is rapid and inexpensive [34] compared to a full cognitive battery. Given the push toward more individualized care through the use of basic and clinical science [70], RT inconsistency may be a particularly useful metric, with the potential to change the way we view the connection between cognition and psychiatric symptoms.

Implications and Future Directions

In sum, momentary WP fluctuations in NA across 14 days predicted RT inconsistency, but the magnitude of this effect changed over the course of the study. These findings support the notion that NA is a form of cognitive load that disrupts cognitive performance, and even subtle fluctuations in NA are a form of stress that impacts cognitive functioning. These findings have implications for the mental health and academic performance of college students, given that internalizing conditions are on the rise and college students are increasingly seeking treatment for emotional disorders and attentional complaints. Perhaps interventions to target NA (enhancing PA, reducing NA, or directly targeting fluctuations in NA) could improve cognitive functioning in college students. For example, mindfulness-based stress approaches [71] and cognitive behavioral therapy have shown promise in reducing affective instability and improving cognitive function [72]. These interventions could be tailored to college students and emerging adults, a population particularly vulnerable to emotional and cognitive challenges, to enhance resilience and cognitive performance. By linking these findings to practical strategies, our long-term goal is to provide actionable insights for researchers and clinicians.

Additional research is necessary to examine the strength of these WP associations in relation to psychological symptoms, with potential implications for digital risk screening and preventative (eg, cognitive behavioral) intervention. Indeed, changes in affect and cognition frequently precede psychological symptoms [73,74], suggesting that momentary cognitive monitoring using scalable technologies may support early risk assessment. This study's focus on momentary NA fluctuations complements prior research on emotional inertia [10], which has been associated with prolonged emotional disturbances and reduced cognitive resilience. Future research should examine how inertia and NA fluctuations may interact to influence cognitive performance, using EMA to capture their dynamic interplay in real-world settings.

Finally, EMA can help identify patterns of emotional responses that may signal the need for early interventions aimed at improving emotional regulation. While we focused on WP differences in NA variability, BP effects may modulate outcomes. Additional work examining how stable traits, such as neuroticism, insomnia, and sex, interact with WP fluctuations in NA to influence cognitive variability would deepen our understanding of these dynamic relationships. Future directions include the integration of passive data to better understand cognition, stress, and sleep, as well as testing in clinical samples and aging samples to understand connections between depression, cognitive impairment, and cognitive decline. Moreover, longer (eg, longitudinal burst) time series data are recommended to enable personalized risk modeling.

Conclusions

This study provides preliminary evidence that momentary WP fluctuations in NA are associated with RT inconsistency in daily life, although the magnitude of this association changed over the course of the study. Our findings support the value of EMA and brief cognitive tasks to capture dynamic links between affect and cognitive functioning in real-world contexts. Future work using longer monitoring periods and passive data may further clarify how daily fluctuations in affect contribute to cognitive performance variability and psychiatric risk.

Acknowledgments

During the revision of this manuscript, the authors used generative artificial intelligence (AI) tools (ChatGPT) for editorial assistance and organization of revision materials. Generative AI was not used to generate data, conduct statistical analyses, or produce study results. The authors reviewed and edited all AI-assisted content and take full responsibility for the content of the manuscript.

Funding

This work is supported by the National Institute on Aging (K01 AG086598 to LAR).

XSL•FO
RenderX

Data Availability

Analysis code is available on OSF [75]. The participant-level data analyzed in this study are not publicly available because they contain intensive longitudinal ecological momentary assessment and cognitive task data. Deidentified data may be available from the corresponding author upon reasonable request, subject to institutional approvals and data-sharing agreements.

Authors' Contributions

Conceptualization: LAR, LTG, ZH
Data curation: PL
Formal analysis: LAR, PYC
Funding acquisition: LAR
Investigation: LAR, PL
Methodology: LAR, LTG, ZH
Project administration: LAR, PL
Resources: LTG
Software: PYC
Supervision: LAR, LTG
Visualization: PYC
Writing—original draft: LAR
Writing—review and editing: LAR, PYC, PL, LTG, ZH

Conflicts of Interest

ZH has received salary support and stock options from Brooklyn Health, as well as consulting fees from Blueprint Health. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Supplemental materials.
[DOCX File, 738 KB - mhealth_v14i1e64397_app1.docx]

References

1. Greenberg PE, Fournier AA, Sisitsky T, Pike CT, Kessler RC. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). J Clin Psychiatry 2015 Feb;76(2):155-162. [doi: 10.4088/JCP.14m09298] [Medline: 25742202]
2. Murray CJL, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2197-2223. [doi: 10.1016/S0140-6736(12)61689-4] [Medline: 23245608]
3. Barlow DH, Ellard KK, Sauer-Zavala S, Bullis JR, Carl JR. The origins of neuroticism. Perspect Psychol Sci 2014 Sep;9(5):481-496. [doi: 10.1177/1745691614544528] [Medline: 26186755]
4. Crow AJD. Associations between neuroticism and executive function outcomes: response inhibition and sustained attention on a continuous performance test. Percept Mot Skills 2019 Aug;126(4):623-638. [doi: 10.1177/0031512519848221] [Medline: 31146642]
5. Forgas JP. Affect and cognition. Perspect Psychol Sci 2008 Mar;3(2):94-101. [doi: 10.1111/j.1745-6916.2008.00067.x] [Medline: 26158876]
6. Dejonckheere E, Mestdagh M, Houben M, et al. The bipolarity of affect and depressive symptoms. J Pers Soc Psychol 2018 Feb;114(2):323-341. [doi: 10.1037/pspp0000186] [Medline: 29389216]
7. Houben M, Van Den Noortgate W, Kuppens P. The relation between short-term emotion dynamics and psychological well-being: a meta-analysis. Psychol Bull 2015 Jul;141(4):901-930. [doi: 10.1037/a0038822] [Medline: 25822133]
8. Nelson J, Klumparendt A, Doebler P, Ehring T. Everyday emotional dynamics in major depression. Emotion 2020 Mar;20(2):179-191. [doi: 10.1037/emo0000541] [Medline: 30589297]
9. Dhillon S, Videla-Nash G, Foussias G, Segal ZV, Zakzanis KK. On the nature of objective and perceived cognitive impairments in depressive symptoms and real-world functioning in young adults. Psychiatry Res 2020 May;287:112932. [doi: 10.1016/j.psychres.2020.112932] [Medline: 32272334]
10. Waugh CE, Kuppens P. Affect Dynamics: Springer International Publishing; 2021. [doi: 10.1007/978-3-030-82965-0]
11. Schoevers RA, van Borkulo CD, Lamers F, et al. Affect fluctuations examined with ecological momentary assessment in patients with current or remitted depression and anxiety disorders. Psychol Med 2021 Aug;51(11):1906-1915. [doi: 10.1017/S0033291720000689] [Medline: 32234092]
12. Jokela M. Why is cognitive ability associated with psychological distress and wellbeing? Exploring psychological, biological, and social mechanisms. Pers Individ Dif 2022 Jul;192:111592. [doi: 10.1016/j.paid.2022.111592]

XSL•FO
RenderX

13. Stawski RS, MacDonald SWS, Brewster PWH, Munoz E, Cerino ES, Halliday DWR. A comprehensive comparison of quantifications of intraindividual variability in response times: a measurement burst approach. J Gerontol B Psychol Sci Soc Sci 2019 Feb 15;74(3):397-408. [doi: 10.1093/geronb/gbx115] [Medline: 29029201]
14. Bielak AAM, Cherbuin N, Bunce D, Anstey KJ. Intraindividual variability is a fundamental phenomenon of aging: evidence from an 8-year longitudinal study across young, middle, and older adulthood. Dev Psychol 2014 Jan;50(1):143-151. [doi: 10.1037/a0032650] [Medline: 23586940]
15. Hultsch DF, MacDonald SWS, Dixon RA. Variability in reaction time performance of younger and older adults. J Gerontol B Psychol Sci Soc Sci 2002 Mar;57(2):101-115. [doi: 10.1093/geronb/57.2.p101] [Medline: 11867658]
16. Kochan NA, Bunce D, Pont S, Crawford JD, Brodaty H, Sachdev PS. Reaction time measures predict incident dementia in community-living older adults: the Sydney Memory and Ageing Study. Am J Geriatr Psychiatry 2016 Mar;24(3):221-231. [doi: 10.1016/j.jagp.2015.12.005] [Medline: 26905045]
17. MacDonald SWS, Stawski RS. Intraindividual Variability—An Indicator of Vulnerability or Resilience in Adult Development and Aging? In: Diehl M, Hooker K, Sliwinski MJ, editors. Handbook of Intraindividual Variability across the Life Span: Routledge; 2015:231-257.
18. Diehl M, Hooker K, Sliwinski MJ. Handbook of Intraindividual Variability Across the Life Span: Routledge; 2015. [doi: 10.4324/9780203113066]
19. MacDonald SWS, Stawski RS. Longitudinal changes in response time mean and inconsistency exhibit predictive dissociations for risk of cognitive impairment. Neuropsychology 2020 Mar;34(3):264-275. [doi: 10.1037/neu0000608] [Medline: 31789567]
20. Hunt J, Eisenberg D. Mental health problems and help-seeking behavior among college students. J Adolesc Health 2010 Jan;46(1):3-10. [doi: 10.1016/j.jadohealth.2009.08.008] [Medline: 20123251]
21. Polanczyk GV, Willcutt EG, Salum GA, Kieling C, Rohde LA. ADHD prevalence estimates across three decades: an updated systematic review and meta-regression analysis. Int J Epidemiol 2014 Apr;43(2):434-442. [doi: 10.1093/ije/dyt261] [Medline: 24464188]
22. Arnett JJ, Žukauskienė R, Sugimura K. The new life stage of emerging adulthood at ages 18-29 years: implications for mental health. Lancet Psychiatry 2014 Dec;1(7):569-576. [doi: 10.1016/S2215-0366(14)00080-7] [Medline: 26361316]
23. Sliwinski MJ, Smyth JM, Hofer SM, Stawski RS. Intraindividual coupling of daily stress and cognition. Psychol Aging 2006 Sep;21(3):545-557. [doi: 10.1037/0882-7974.21.3.545] [Medline: 16953716]
24. Kahneman D, Tversky A. On the psychology of prediction. Psychol Rev 1973;80(4):237-251. [doi: 10.1037/h0034747]
25. Blanchette I, Richards A. The influence of affect on higher level cognition: a review of research on interpretation, judgement, decision making and reasoning. Cogn Emot 2010 Jun;24(4):561-595. [doi: 10.1080/02699930903132496]
26. Shackman AJ, Salomons TV, Slagter HA, Fox AS, Winter JJ, Davidson RJ. The integration of negative affect, pain and cognitive control in the cingulate cortex. Nat Rev Neurosci 2011 Mar;12(3):154-167. [doi: 10.1038/nrn2994] [Medline: 21331082]
27. Pechtel P, Pizzagalli DA. Effects of early life stress on cognitive and affective function: an integrated review of human literature. Psychopharmacology (Berl) 2011 Mar;214(1):55-70. [doi: 10.1007/s00213-010-2009-2] [Medline: 20865251]
28. Sauer-Zavala S, Boswell JF, Gallagher MW, Bentley KH, Ametaj A, Barlow DH. The role of negative affectivity and negative reactivity to emotions in predicting outcomes in the unified protocol for the transdiagnostic treatment of emotional disorders. Behav Res Ther 2012 Sep;50(9):551-557. [doi: 10.1016/j.brat.2012.05.005] [Medline: 22738907]
29. Rutter LA, Ten Thij M, Lorenzo-Luaces L, Valdez D, Bollen J. Negative affect variability differs between anxiety and depression on social media. PLoS One 2024;19(2):e0272107. [doi: 10.1371/journal.pone.0272107] [Medline: 38381769]
30. Sperry SH, Walsh MA, Kwapil TR. Emotion dynamics concurrently and prospectively predict mood psychopathology. J Affect Disord 2020 Jan 15;261:67-75. [doi: 10.1016/j.jad.2019.09.076] [Medline: 31600589]
31. Wichers M, Peeters F, Geschwind N, et al. Unveiling patterns of affective responses in daily life may improve outcome prediction in depression: a momentary assessment study. J Affect Disord 2010 Jul;124(1-2):191-195. [doi: 10.1016/j.jad.2009.11.010] [Medline: 20004977]
32. Zainal NH, Newman MG. Larger increase in trait negative affect is associated with greater future cognitive decline and vice versa across 23 years. Depress Anxiety 2021 Feb;38(2):146-160. [doi: 10.1002/da.23093] [Medline: 32840954]
33. MacDonald SWS, Li SC, Bäckman L. Neural underpinnings of within-person variability in cognitive functioning. Psychol Aging 2009 Dec;24(4):792-808. [doi: 10.1037/a0017798] [Medline: 20025396]
34. Rutter LA, Vahia IV, Passell E, Forester BP, Germine L. The role of intraindividual cognitive variability in posttraumatic stress syndromes and cognitive aging: a literature search and proposed research agenda. Int Psychogeriatr 2021 Jul;33(7):677-687. [doi: 10.1017/S1041610220000228] [Medline: 32172714]
35. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4(1):1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
36. Passell E, Strong RW, Rutter LA, et al. Cognitive test scores vary with choice of personal digital device. Behav Res Methods 2021 Dec;53(6):2544-2557. [doi: 10.3758/s13428-021-01597-3] [Medline: 33954913]

XSL•FO
RenderX

37. Germine L, Nakayama K, Duchaine BC, Chabris CF, Chatterjee G, Wilmer JB. Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. Psychon Bull Rev 2012 Oct;19(5):847-857. [doi: 10.3758/s13423-012-0296-9] [Medline: 22829343]
38. Rutter LA, Lind C, Howard J, Lakhan P, Germine L. Posttraumatic stress symptom severity is associated with impaired processing of emotional faces in a large international sample. J Trauma Stress 2022 Aug;35(4):1263-1272. [doi: 10.1002/jts.22834] [Medline: 35366020]
39. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988 Jun;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
40. Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001 Jul;2(4):297-307. [doi: 10.1016/s1389-9457(00)00065-4] [Medline: 11438246]
41. Blais FC, Gendron L, Mimeault V, Morin CM. Evaluation de l'insomnie: validation de trois questionnaires [Article in French]. Encephale 1997;23(6):447-453. [Medline: 9488928]
42. Morin CM, Belleville G, Bélanger L, Ivers H. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep 2011 May 1;34(5):601-608. [doi: 10.1093/sleep/34.5.601] [Medline: 21532953]
43. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]
44. Rosellini AJ, Brown TA. The Multidimensional Emotional Disorder Inventory (MEDI): assessing transdiagnostic dimensions to validate a profile approach to emotional disorder classification. Psychol Assess 2019 Jan;31(1):59-72. [doi: 10.1037/pas0000649] [Medline: 30160498]
45. Germine LT, Joormann J, Passell E, et al. Neurocognition after motor vehicle collision and adverse post-traumatic neuropsychiatric sequelae within 8 weeks: initial findings from the AURORA study. J Affect Disord 2022 Feb 1;298(Pt B):57-67. [doi: 10.1016/j.jad.2021.10.104] [Medline: 34800569]
46. Rutter LA, Vahia IV, Forester BP, Ressler KJ, Germine L. Heterogeneous indicators of cognitive performance and performance variability across the lifespan. Front Aging Neurosci 2020;12:62. [doi: 10.3389/fnagi.2020.00062] [Medline: 32210793]
47. Maljkovic V, Nakayama K. Priming of pop-out: I. Role of features. Mem Cognit 1994 Nov;22(6):657-672. [doi: 10.3758/bf03209251] [Medline: 7808275]
48. Singh S, Strong R, Xu I, et al. Ecological momentary assessment of cognition in clinical and community samples: reliability and validity study. J Med Internet Res 2023 Jun 2;25:e45028. [doi: 10.2196/45028] [Medline: 37266996]
49. Wickham H, Averick M, Bryan J, et al. Welcome to the Tidyverse. J Open Source Softw 2019;4(43):1686. [doi: 10.21105/joss.01686]
50. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
51. Dalgleish T, Watts FN. Biases of attention and memory in disorders of anxiety and depression. Clin Psychol Rev 1990 Jan;10(5):589-604. [doi: 10.1016/0272-7358(90)90098-U]
52. Kircanski K, Joormann J, Gotlib IH. Cognitive aspects of depression. Wiley Interdiscip Rev Cogn Sci 2012;3(3):301-313. [doi: 10.1002/wcs.1177] [Medline: 23240069]
53. Reinholdt-Dunne ML, Mogg K, Bradley BP. Attention control: relationships between self-report and behavioural measures, and symptoms of anxiety and depression. Cogn Emot 2013;27(3):430-440. [doi: 10.1080/02699931.2012.715081] [Medline: 22917576]
54. Bierman EJM, Comijs HC, Jonker C, Beekman ATF. Symptoms of anxiety and depression in the course of cognitive decline. Dement Geriatr Cogn Disord 2007;24(3):213-219. [doi: 10.1159/000107083] [Medline: 17690554]
55. Suddell S, Mahedy L, Skirrow C, Penton-Voak IS, Munafò MR, Wootton RE. Cognitive functioning in anxiety and depression: results from the ALSPAC cohort. R Soc Open Sci 2023 Aug;10(8):221161. [doi: 10.1098/rsos.221161] [Medline: 37564071]
56. Boals A, Banks JB. Effects of traumatic stress and perceived stress on everyday cognitive functioning. Cogn Emot 2012;26(7):1335-1343. [doi: 10.1080/02699931.2011.651100] [Medline: 22963576]
57. Klein K, Barnes D. The relationship of life stress to problem solving: task complexity and individual differences. Soc Cogn 1994 Sep;12(3):187-204. [doi: 10.1521/soco.1994.12.3.187]
58. Stawski RS, Sliwinski MJ, Smyth JM. Stress-related cognitive interference predicts cognitive function in old age. Psychol Aging 2006 Sep;21(3):535-544. [doi: 10.1037/0882-7974.21.3.535] [Medline: 16953715]
59. Yerkes RM, Dodson JD. The relation of strength of stimulus to rapidity of habit ‐ formation. J Comp Neurol Psychol 1908 Nov;18(5):459-482. [doi: 10.1002/cne.920180503]
60. Baird BM, Le K, Lucas RE. On the nature of intraindividual personality variability: reliability, validity, and associations with well-being. J Pers Soc Psychol 2006 Mar;90(3):512-527. [doi: 10.1037/0022-3514.90.3.512] [Medline: 16594835]
61. Eid M, Diener E. Intraindividual variability in affect: reliability, validity, and personality correlates. J Pers Soc Psychol 1999;76(4):662-676. [doi: 10.1037/0022-3514.76.4.662]

XSL•FO
RenderX

62. Hallion LS, Kusmierski SN, Caulfield MK. Worry alters speed-accuracy tradeoffs but does not impair sustained attention. Behav Res Ther 2020 May;128:103597. [doi: 10.1016/j.brat.2020.103597] [Medline: 32217356]
63. Mestdagh M, Pe M, Pestman W, Verdonck S, Kuppens P, Tuerlinckx F. Sidelining the mean: the relative variability index as a generic mean-corrected variability measure for bounded variables. Psychol Methods 2018 Dec;23(4):690-707. [doi: 10.1037/met0000153] [Medline: 29648843]
64. Germine LT, Duchaine B, Nakayama K. Where cognitive development and aging meet: face learning ability peaks after age 30. Cognition 2011 Feb;118(2):201-210. [doi: 10.1016/j.cognition.2010.11.002] [Medline: 21130422]
65. Germine L, Reinecke K, Chaytor NS. Digital neuropsychology: challenges and opportunities at the intersection of science and software. Clin Neuropsychol 2019 Feb;33(2):271-286. [doi: 10.1080/13854046.2018.1535662] [Medline: 30614374]
66. Hartshorne JK, Germine LT. When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. Psychol Sci 2015 Apr;26(4):433-443. [doi: 10.1177/0956797614567339] [Medline: 25770099]
67. Passell E, Dillon DG, Baker JT, et al. Digital Cognitive Assessment: Results from the TestMyBrain NIMH Research Domain Criteria (RDoC) Field Test Battery Report. PsyArXiv. Preprint posted online on Jan 26, 2019. [doi: 10.31234/osf.io/dcszr]
68. Hawks ZW, Strong R, Jung L, et al. Accurate prediction of momentary cognition from intensive longitudinal data. Biol Psychiatry Cogn Neurosci Neuroimaging 2023 Aug;8(8):841-851. [doi: 10.1016/j.bpsc.2022.12.002] [Medline: 36922302]
69. Oravecz Z, Harrington KD, Hakun JG, et al. Accounting for retest effects in cognitive testing with the Bayesian double exponential model via intensive measurement burst designs. Front Aging Neurosci 2022;14:897343. [doi: 10.3389/fnagi.2022.897343] [Medline: 36225891]
70. Stuss DT. Individual differences in cognitive neuropsychology. Pers Individ Dif 2017 Nov;118:4-6. [doi: 10.1016/j.paid.2016.10.037]
71. Guendelman S, Medeiros S, Rampes H. Mindfulness and emotion regulation: insights from neurobiological, psychological, and clinical studies. Front Psychol 2017;8:220. [doi: 10.3389/fpsyg.2017.00220] [Medline: 28321194]
72. Hofmann SG, Asnaani A, Vonk IJJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: a review of meta-analyses. Cognit Ther Res 2012 Oct 1;36(5):427-440. [doi: 10.1007/s10608-012-9476-1] [Medline: 23459093]
73. Curtiss JE, Mischoulon D, Fisher LB, et al. Rising early warning signals in affect associated with future changes in depression: a dynamical systems approach. Psychol Med 2023 May;53(7):3124-3132. [doi: 10.1017/S0033291721005183] [Medline: 34937601]
74. Wichers M, Groot PC, Psychosystems, ESM Group, EWS Group. Critical slowing down as a personalized early warning signal for depression. Psychother Psychosom 2016;85(2):114-116. [doi: 10.1159/000441458] [Medline: 26821231]
75. OSF. URL: https://osf.io/yzdqw/ [accessed 2026-05-29]

Abbreviations

BP: between-person
DASS: Depression Anxiety and Stress Scales
EMA: ecological momentary assessment
MEDI: Multidimensional Emotional Disorder Inventory
NA: negative affect
PA: positive affect
PANAS-SF: Positive and Negative Affect Schedule—Short Form
RT: reaction time
WP: within-person

Edited by A Stone, L Buis; submitted 16.Jul.2024; peer-reviewed by N Nazari, S Lau; revised version received 08.May.2026; accepted 11.May.2026; published 09.Jun.2026.

Please cite as:
Rutter LA, Chen PY, Lakhan P, Germine LT, Hawks Z
Daily Negative Affect and Reaction Time Inconsistency in Emerging Adults: Ecological Momentary Assessment Study
JMIR Mhealth Uhealth 2026;14:e64397
URL: https://mhealth.jmir.org/2026/1/e64397
doi:10.2196/64397

© Lauren Alexandra Rutter, Pei-Ying Chen, Prabhvir Lakhan, Laura Thi Germine, Zoe Hawks. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 9.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited.

XSL•FO
RenderX

The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Objective Smartphone Screen Time Monitoring Among Caregivers of Preschool-Aged Children: Reliability Study of Monitoring Duration

Joshua Culverhouse[1], PhD; Taylor Adair[1], BS; Meghan Restino[1], BS; Heidi M Weeks[2], PhD; Jenny S Radesky[3], MD; R Glenn Weaver[1], PhD; Sarah Burkart[1], PhD; Hannah Parker[1], PhD; Olivia L Finnegan[1], PhD; Anthony J Holmes[1], PhD; Keagan Kiely[1], PhD; Rahul Ghosal[4], PhD; Srihari Nelakuditi[5], PhD; Michael W Beets[1], PhD; Bridget Armstrong[1], PhD

[1]Department of Exercise Science, University of South Carolina, PHRC - Room 140, 921 Assembly St, Columbia, SC, United States

[2]University of Michigan School of Public Health, Ann Arbor, MI, United States

[3]University of Michigan Medical School, Ann Arbor, MI, United States

[4]Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC, United States

[5]Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, United States

Corresponding Author:
Joshua Culverhouse, PhD
Department of Exercise Science, University of South Carolina, PHRC - Room 140, 921 Assembly St, Columbia, SC, United States

Abstract

Background: Reliable measurement of smartphone screen time is important for research examining associations between digital media use and health outcomes. Although objective monitoring tools reduce bias associated with self-report, the number of monitoring days required to reliably estimate typical smartphone use remains unclear. This question is particularly important for study design because longer monitoring periods increase participant burden and the likelihood of missing data.

Objective: This study aimed to determine the minimum number of valid monitoring days within a 7-day measurement period required to reliably estimate average daily smartphone screen time during the same 7-day period among primary caregivers of preschool-aged children using 2 objective methods.

Methods: This secondary analysis included caregivers participating in 2 observational studies conducted in the Eastern United States. Android users installed Chronicle, a passive sensing application that records raw Android UsageEvents data, which were preprocessed into smartphone session events (pickups where the device was unlocked). iOS users submitted daily screenshots of Apple Screen Time summaries. Reliability analyses included 100 Android Chronicle participants and 129 iOS Screenshot participants with at least 7 consecutive complete days of data. The primary outcome was day-level total smartphone screen time. For Android Chronicle, this was defined as the summed duration of device-unlock sessions across each calendar day, whereas for iOS Screenshot, this was defined as the daily total reported by Apple Screen Time. One-way random-effects intraclass correlation coefficients (ICC [1,1]) with 95% bootstrap CIs (500 resamples) were calculated separately for Android and iOS data. The Spearman-Brown prophecy formula was used to estimate the number of monitoring days required to achieve target reliability thresholds (ICC≥0.70, 0.80, and 0.90). Sensitivity analyses using the combined durations of sessions and glances produced similar conservative monitoring-day recommendations.

Results: Single-day ICCs were 0.69 (95% CI 0.63-0.76) for Android Chronicle and 0.71 (95% CI 0.65-0.77) for iOS Screenshot data. Point-estimate projections indicated that 2 monitoring days were sufficient to achieve an ICC≥0.80 for both data sources. Using the more conservative criterion that the lower bound of the projected 95% CI exceeded 0.80, 3 monitoring days were required for both Android and iOS data. Screen time was significantly lower on weekends than on weekdays, suggesting that monitoring periods should include at least 1 weekend day.

Conclusions: Among caregivers of preschool-aged children, relatively short objective monitoring periods may provide reliable estimates of average daily smartphone screen time within a 7-day monitoring period. Two monitoring days were sufficient under point-estimate projections, whereas 3 days were required using a conservative lower-bound criterion. These findings support the feasibility of brief smartphone monitoring protocols for estimating the 7-day average duration, while emphasizing the importance of including weekend representation and avoiding the extension of these estimates to longer-term habitual smartphone behavior or screen time metrics beyond duration without further evidence.

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2026;14:e82846) doi:10.2196/82846

KEYWORDS

screen time; passive sensing; smartphone; sampling duration; reliability

Introduction

Background

Smartphone ownership is nearly universal in the United States, with more than 90% of adults owning a smartphone [1]. Alongside rising usage, growing evidence links smartphone use to a range of health and psychosocial outcomes [2-4]. However, progress in this area depends on the quality of the measures used to quantify smartphone behavior. In digital media research, screen time and related exposure variables are commonly used as predictors or outcomes, but many studies have relied on retrospective self-report measures [5-7].

Limitations of Self-Reported Smartphone Use

Self-reported media use is vulnerable to recall error, social desirability bias, and difficulty estimating frequent behaviors embedded in daily routines [8,9]. These limitations may be especially relevant for smartphone use, which often occurs through short, repeated interactions across varied contexts and alongside other activities [10]. A systematic review and meta-analysis found that self-reported digital media use correlated only moderately with logged use and was rarely an accurate reflection of objectively recorded behavior [7]. This suggests that self-reported and logged measures may not be interchangeable and supports the use of objective measurement when estimating smartphone exposure.

Objective Smartphone Measurement

Objective smartphone measures generally fall into several related approaches. On Android devices, timestamped operating-system event logs can be processed to construct smartphone use behaviors such as sessions, glances, and app use episodes [11]. Research tools such as Chronicle and the Effortless Assessment Research System (EARS) use these types of passive sensing data to provide detailed measures of device or app interaction [12-14]. Other intensive approaches, such as Screenomics, capture repeated screenshots to characterize digital context and content, offering rich behavioral detail but with greater privacy, storage, and processing demands [15]. In contrast, iOS research often relies on less granular approaches or participant-submitted screenshots of Apple Screen Time summaries because Apple restricts third-party access to detailed device activity data. Although these objective approaches reduce reliance on retrospective self-report, they still require decisions about the number of monitoring days needed to obtain reliable estimates.

Within-Person Variability and Monitoring Duration

Smartphone use also varies meaningfully within individuals across days and situations. Recent work distinguishing person-specific from situation-specific variation in media use found that a large share of media use variance occurs within persons across situations, underscoring the need for repeated measurement designs rather than single time point estimates [16]. At the same time, smartphone behavior may include habitual or automatic interactions triggered by contextual cues, making it difficult for participants to accurately recall or summarize their use [17]. Together, these features create a practical measurement challenge: researchers need objective monitoring protocols that are long enough to capture typical use, but not so long that they increase participant burden or missing data.

Prior studies have addressed similar questions using correlation-based approaches, examining the association between short-term and longer-term monitoring windows [18,19]. For instance, Wilcockson et al [18] found that 5 days of monitoring were sufficient to characterize weekly smartphone duration and 2 days were sufficient for checking behavior (pickups lasting <15 s). Although informative, correlations primarily capture rank-order consistency across individuals and do not directly estimate measurement reliability, absolute agreement, or the contribution of within-person variability.

Reliability-Based Study Planning

An alternative approach uses intraclass correlation coefficients (ICCs) and the Spearman-Brown prophecy formula [20,21]. ICC-based approaches are useful for planning measurement protocols because they quantify the proportion of observed variance attributable to stable between-person differences, provide an estimate of measurement error in the exposure, and can be paired with the Spearman-Brown formula to estimate how many repeated days are needed to achieve a prespecified reliability threshold. These methods are widely used in physical activity and other behavioral monitoring research to determine the minimum monitoring periods required for stable estimates of behavior [22-26], but have not been widely applied to objective smartphone screen time measurement.

Study Aim

In this study, we used the ICC and Spearman-Brown methods to estimate the minimum number of valid monitoring days required, within a 7-day monitoring period, to reliably estimate average daily smartphone screen time duration among primary caregivers of young children. Separate analyses were conducted for 2 objectively collected data sources: (1) passive sensing via the Chronicle app on Android devices and (2) screenshots of Apple Screen Time on iOS devices. Because these data sources reflect different platform-specific measurement approaches and were collected from distinct participant groups, the aim was not to test equivalence or directly compare Android and iOS estimates. Rather, our aim was to provide data-source–specific reliability estimates and practical guidance for researchers studying smartphone behavior in this population, where some days of monitoring may be missing, invalid, or incomplete.

XSL•FO
RenderX

Methods

Participants and Study Design

This secondary analysis used data from 2 observational studies involving caregivers of preschool-aged children: a pilot feasibility study [27] and an ongoing longitudinal cohort study [28] (R01HD112311). Both studies recruited nonrandom samples via social media across the eastern United States and used identical smartphone screen time assessment protocols. Inclusion criteria were (1) primary caregiver of a child aged 3 to 5 years, (2) ownership of a smartphone, and (3) ability to read and speak English. Remote data collection included surveys, ecological momentary assessments, and passive smartphone sensing.

Pilot study data were collected from September 2020 to February 2022 using a 30-day monitoring protocol, whereas cohort study data were collected from September 2024 to May 2025 using a 14-day monitoring protocol. Android Chronicle data were available from both studies, whereas iOS Screenshot data were collected only in the cohort study.

Ethical Considerations

Both parent studies were approved by the University of South Carolina Institutional Review Board (IRB) in August 2020 (Pro00092634) and December 2023 (Pro00133829). All participants provided informed consent before participation, and the IRB approvals covered secondary analyses of the collected data. Participant data were deidentified before analysis and stored on secure, access-restricted university servers. Smartphone data were processed using participant identifiers rather than names or directly identifying information. Participants received up to US $440 per assessment period for completing the broader study protocol, including parent surveys, daily diaries, ecological momentary assessments, and other remote data collection procedures. Smartphone screenshot submissions were completed as part of the compensated daily diary protocol (US $5 per completion). Chronicle data collection was not directly compensated.

Smartphone Screen Time Measurement

Android Chronicle

Android users installed Chronicle, a passive sensing application that records timestamped device interaction events [12]. Chronicle accesses raw Android event-log data through the Android UsageEvents framework, which captures application transitions, screen interactions, and device state events [13]. Data are uploaded at regular intervals to a secure server when the device is connected to Wi-Fi.

For the current analysis, raw Android event logs were used to derive smartphone interaction measures based on an adapted implementation of the framework proposed by Parry and Torth [11] for constructing meaningful smartphone use behaviors from Android event streams. This approach constructed smartphone sessions (device-unlock episodes), glances (brief screen activations without unlocking), and foreground app use episodes from timestamped event sequences (Figure 1).

Figure 1. Example preprocessed Chronicle output.

participant_id	interaction_type	application_label	app_package_name	event_timestamp	start_timestamp	stop_timestamp	duration_secs	timezone	processing_date	processing_version
example_data	Glance	Android system	android	2025-01-14 18:40:15.426-05:0	1/14/25 6:40:15 PM	1/14/25 6:40:25 PM	10.4	America/New_York	8/10/2026	v1.0
example_data	Session	Android system	android	2025-01-14 18:40:26.865-05:0	1/14/25 6:40:26 PM	1/14/25 6:41:33 PM	67.1	America/New_York	8/10/2026	v1.0
example_data	App Usage	minimalist phone	com.qqlabs.minimalistlauncher	2025-01-14 18:40:27.072-05:0	1/14/25 6:40:27 PM	1/14/25 6:40:28 PM	1.3	America/New_York	8/10/2026	v1.0
example_data	Unknown importance: 11	Calendar	com.google.android.calendar	2025-01-14 18:40:27.545-05:0	1/14/25 6:40:27 PM	NA	NA	America/New_York	8/10/2026	v1.0
example_data	Unknown importance: 11	Gmail	com.google.android.gm	2025-01-14 18:40:27.546-05:0	1/14/25 6:40:27 PM	NA	NA	America/New_York	8/10/2026	v1.0
example_data	App Usage	Spotify	com.spotify.music	2025-01-14 18:40:28.398-05:0	1/14/25 6:40:28 PM	1/14/25 6:41:01 PM	33	America/New_York	8/10/2026	v1.0
example_data	Unknown importance: 11	Android Auto	com.google.android.projection.gearhead	2025-01-14 18:40:29.332-05:0	1/14/25 6:40:29 PM	NA	NA	America/New_York	8/10/2026	v1.0
example_data	App Usage	Spotify	com.spotify.music	2025-01-14 18:41:01.411-05:0	1/14/25 6:41:01 PM	1/14/25 6:41:33 PM	32.5	America/New_York	8/10/2026	v1.0
example_data	Session	Android system	android	2025-01-14 18:47:37.986-05:0	1/14/25 6:47:37 PM	1/14/25 6:52:03 PM	265	America/New_York	8/10/2026	v1.0
example_data	App Usage	Spotify	com.spotify.music	2025-01-14 18:47:38.154-05:0	1/14/25 6:47:38 PM	1/14/25 6:47:45 PM	6.9	America/New_York	8/10/2026	v1.0
example_data	App Usage	Spotify	com.spotify.music	2025-01-14 18:47:45.046-05:0	1/14/25 6:47:45 PM	1/14/25 6:48:09 PM	24.8	America/New_York	8/10/2026	v1.0
example_data	App Usage	minimalist phone	com.qqlabs.minimalistlauncher	2025-01-14 18:48:09.843-05:0	1/14/25 6:48:09 PM	1/14/25 6:48:24 PM	15.2	America/New_York	8/10/2026	v1.0
example_data	Unknown importance: 11	Gboard	com.google.android.inputmethod.latin	2025-01-14 18:48:10.954-05:0	1/14/25 6:48:10 PM	NA	NA	America/New_York	8/10/2026	v1.0
example_data	App Usage	Chrome	com.android.chrome	2025-01-14 18:48:25.082-05:0	1/14/25 6:48:25 PM	1/14/25 6:51:59 PM	214.5	America/New_York	8/10/2026	v1.0
example_data	Unknown importance: 11	Calendar	com.google.android.calendar	2025-01-14 18:50:27.621-05:0	1/14/25 6:50:27 PM	NA	NA	America/New_York	8/10/2026	v1.0
example_data	Unknown importance: 11	Gmail	com.google.android.gm	2025-01-14 18:50:27.624-05:0	1/14/25 6:50:27 PM	NA	NA	America/New_York	8/10/2026	v1.0
example_data	Unknown importance: 11	Android Auto	com.google.android.projection.gearhead	2025-01-14 18:50:29.340-05:0	1/14/25 6:50:29 PM	NA	NA	America/New_York	8/10/2026	v1.0
example_data	App Usage	Moto App Launcher	com.motorola.launcher3	2025-01-14 18:51:59.573-05:0	1/14/25 6:51:59 PM	1/14/25 6:51:59 PM	0.2	America/New_York	8/10/2026	v1.0
example_data	App Usage	Spotify	com.spotify.music	2025-01-14 18:51:59.743-05:0	1/14/25 6:51:59 PM	1/14/25 6:52:03 PM	3.3	America/New_York	8/10/2026	v1.0
example_data	Glance	Android system	android	2025-01-14 18:52:06.649-05:0	1/14/25 6:52:06 PM	1/14/25 6:52:20 PM	13.8	America/New_York	8/10/2026	v1.0
example_data	Session	Android system	android	2025-01-14 18:52:30.871-05:0	1/14/25 6:52:30 PM	1/14/25 6:53:10 PM	40.1	America/New_York	8/10/2026	v1.0
example_data	App Usage	Spotify	com.spotify.music	2025-01-14 18:52:31.189-05:0	1/14/25 6:52:31 PM	1/14/25 6:52:36 PM	5.6	America/New_York	8/10/2026	v1.0
example_data	Unknown importance: 11	Calendar	com.google.android.calendar	2025-01-14 18:52:32.145-05:0	1/14/25 6:52:32 PM	NA	NA	America/New_York	8/10/2026	v1.0
example_data	Unknown importance: 11	Gmail	com.google.android.gm	2025-01-14 18:52:32.153-05:0	1/14/25 6:52:32 PM	NA	NA	America/New_York	8/10/2026	v1.0
example_data	App Usage	Spotify	com.spotify.music	2025-01-14 18:52:36.838-05:0	1/14/25 6:52:36 PM	1/14/25 6:53:10 PM	34.1	America/New_York	8/10/2026	v1.0
example_data	Unknown importance: 11	Gboard	com.google.android.inputmethod.latin	2025-01-14 18:58:11.026-05:0	1/14/25 6:58:11 PM	NA	NA	America/New_York	8/10/2026	v1.0
example_data	Glance	Android system	android	2025-01-14 19:01:09.125-05:0	1/14/25 7:01:09 PM	1/14/25 7:01:20 PM	11.5	America/New_York	8/10/2026	v1.0
example_data	Session	Android system	android	2025-01-14 19:01:29.665-05:0	1/14/25 7:01:29 PM	1/14/25 7:01:47 PM	17.9	America/New_York	8/10/2026	v1.0
example_data	App Usage	Spotify	com.spotify.music	2025-01-14 19:01:29.878-05:0	1/14/25 7:01:29 PM	1/14/25 7:01:31 PM	2	America/New_York	8/10/2026	v1.0
example_data	App Usage	minimalist phone	com.qqlabs.minimalistlauncher	2025-01-14 19:01:31.895-05:0	1/14/25 7:01:31 PM	1/14/25 7:01:33 PM	1.6	America/New_York	8/10/2026	v1.0
example_data	App Usage	Moto App Launcher	com.motorola.launcher3	2025-01-14 19:01:33.452-05:0	1/14/25 7:01:33 PM	1/14/25 7:01:33 PM	0	America/New_York	8/10/2026	v1.0
example_data	App Usage	minimalist phone	com.qqlabs.minimalistlauncher	2025-01-14 19:01:33.480-05:0	1/14/25 7:01:33 PM	1/14/25 7:01:36 PM	2.7	America/New_York	8/10/2026	v1.0
example_data	Unknown importance: 11	Gboard	com.google.android.inputmethod.latin	2025-01-14 19:01:34.605-05:0	1/14/25 7:01:34 PM	NA	NA	America/New_York	8/10/2026	v1.0
example_data	App Usage	WhatsApp	com.whatsapp	2025-01-14 19:01:36.250-05:0	1/14/25 7:01:36 PM	1/14/25 7:01:36 PM	0.7	America/New_York	8/10/2026	v1.0
example_data	App Usage	WhatsApp	com.whatsapp	2025-01-14 19:01:36.971-05:0	1/14/25 7:01:36 PM	1/14/25 7:01:38 PM	1.9	America/New_York	8/10/2026	v1.0

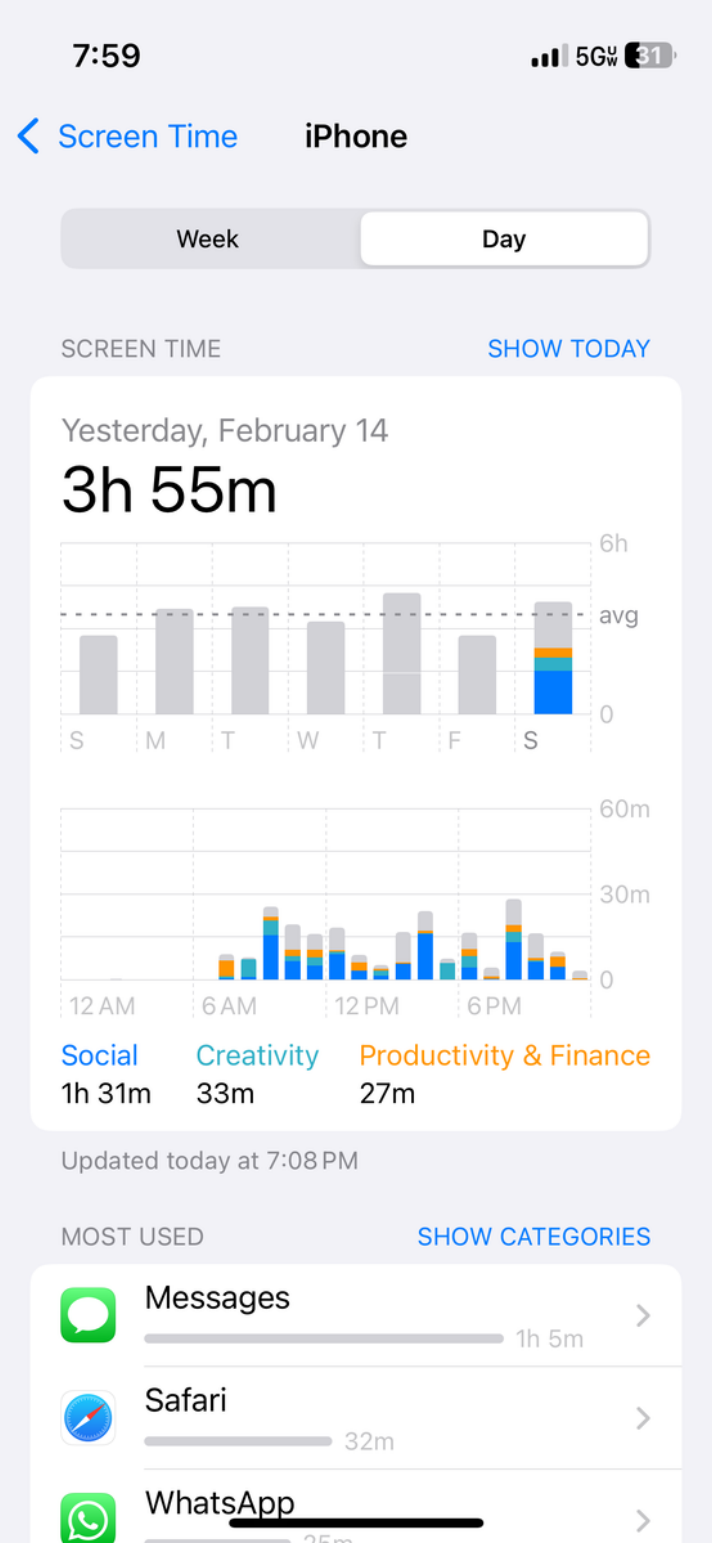
The primary Android outcome was total daily smartphone session duration, aggregated by each calendar day. In a sensitivity analysis, we also examined combined session and glance duration. Conservative monitoring-day recommendations were unchanged; therefore, the primary analyses presented focus on session-derived estimates.

iOS Screenshot

Comparable passive sensing data were not available on iOS devices because Apple restricts third-party access to detailed app activity data. Therefore, iOS users submitted daily screenshots of their previous day's built-in Apple Screen Time summary via a daily diary survey link texted to their smartphones (Figure 2). Surveys were administered through Qualtrics and automatically distributed using an internal messaging system (WearData).

XSL•FO
RenderX

Figure 2. Example iOS Screenshot (Apple Screen Time).



Screenshots were manually reviewed for correct date, device, and completeness. They were excluded if the displayed date did not match the target reporting day (submitted day minus 1), if the image was a duplicate submission for an already processed day, if the image was unclear or unreadable, or if the hourly bar chart was incomplete or could not be reliably extracted. When multiple valid screenshots were submitted for the same participant-day, only one screenshot was retained. These checks were completed before aggregation to daily totals. We then applied a custom open-source Python script [29] to extract per-hour screen time values from the screenshot bar chart. Hourly values were summed to generate total daily screen time. To validate this approach, we compared the extracted totals against the daily totals displayed on a random sample of 100 screenshots. The mean difference was −2.7±11.7 minutes per day (relative difference of −0.01%±0.04%), with a Lin concordance coefficient of 0.998, demonstrating excellent agreement. This validation confirmed the accuracy of extracting

XSL•FO
RenderX

and summing hourly values from submitted screenshots relative to the daily total displayed by Apple Screen Time. However, it should not be interpreted as criterion validation of Apple Screen Time against raw device logs because Apple's underlying screen time definitions and algorithms are proprietary.

Data Processing and Cleaning

Android Chronicle data were processed using a reproducible preprocessing and cleaning pipeline. Raw Android event logs were first transformed into structured smartphone interaction events, including sessions, glances, and foreground app use episodes [30]. Although this study used session-derived duration measures as the primary outcome, app use episodes were retained during preprocessing and cleaning to support transparent reconstruction of smartphone interaction events and future analyses. Cleaning procedures then standardized timestamps and time zones, collapsed adjacent split app use segments, truncated predefined low-engagement or background applications, identified implausibly long events, and flagged incomplete or potentially unreliable days [31].

To reduce inflation of duration estimates, predefined low-engagement applications (eg, launcher, clock, screensaver, or other background-system applications) were truncated to a maximum duration of 10 minutes rather than excluded entirely. The truncation threshold was chosen in line with prior work using Chronicle data [13]. A list of these apps is available with the cleaning script linked above. Events with implausibly long durations (>3 h) were reviewed using prespecified rules based on Google Play Store app categories (GenreID), with prolonged durations considered plausible for "video players," "entertainment," "social," and "games" apps. All events lasting for more than 6 hours were considered implausible and truncated to 10 minutes. Prolonged inactivity gaps (>12 h without recorded Android events or device shutdown markers) were treated as likely periods of missing data. Days containing such gaps, as well as partial start and end days, were excluded from analysis.

To summarize the application of these rules, we examined the exclusion of partial or incomplete days across all available Chronicle records processed before analytic window selection. Across these records, 10% of participant-days were excluded because of partial coverage or prolonged inactivity gaps. To summarize the impact of long-session truncation on the data used in the primary reliability analyses, we examined the final 7-day analytic windows. Within these analytic windows, 9 of 64,644 pickup/session events (0.01%) exceeded 6 hours and were truncated. These rules were applied systematically using prespecified criteria to reduce implausible inflation of screen time estimates. We did not interpret uncleaned values as valid alternative estimates because they included incomplete days and anomalously long events that were considered likely data artifacts.

Screenshot-based iOS data required only verification of complete 24-hour coverage and of valid screenshot formatting before aggregation into daily totals. For the current analysis, we included only participants with at least 7 consecutive complete days of objective smartphone use data after cleaning. For participants with more than 7 consecutive complete days available, we selected the first 7 days from the first qualifying consecutive sequence after cleaning. This rule was applied consistently across participants and data sources to define a reproducible 7-day analytic window.

Statistical Analysis

Overview

We fitted a linear mixed-effects model (lme4) with data source (iOS vs Android), day type (weekend vs weekday), and their interaction as fixed effects, and a random intercept for participant to account for repeated daily observations within individuals. Fixed-effect estimates (*B*), standard errors (SEs), 95% CIs, and *P* values are reported.

We conducted separate analyses for Android Chronicle and iOS Screenshot data because these data sources were derived from different platform-specific measurement approaches. The aim was not to directly compare Android and iOS screen time estimates but to provide data-source–specific reliability estimates for 2 objective approaches to measuring daily smartphone use duration.

The primary outcome for both data sources was day-level total smartphone screen time in hours. In the Android Chronicle data, this was defined as the summed duration of device-unlock sessions for each calendar day. In the iOS Screenshot data, this was defined as the daily total reported by Apple Screen Time and extracted from participant-submitted screenshots. The following approaches were used, as described in the following sections.

ICC

A 1-way random-effects ICC for single measures with absolute agreement (ICC [1,1]) was calculated to quantify the proportion of variance attributable to stable between-person differences in daily screen time duration [20]. ICC (1,1) was selected because the aim was to estimate the reliability of a single day of monitoring sampled from a broader universe of possible days, rather than agreement between a fixed set of raters or measurement occasions. This provides a direct estimate of test-retest reliability for a single day of monitoring. ICCs were computed using the *irr::icc*() function in R (R Foundation for Statistical Computing) [32]. To quantify uncertainty around the single-day ICC estimate, we generated a 95% CI using nonparametric bootstrapping (resampling participants with replacement; 500 replicates). For each bootstrap replicate, the single-day ICC estimate was entered into the Spearman-Brown prophecy formula to generate projected ICC_k values and corresponding estimates of the number of monitoring days required to achieve target reliability thresholds, thereby propagating uncertainty from the single-day ICC through all projected reliability estimates.

Spearman-Brown Prophecy Formula

To estimate how reliability improves with additional days, we calculated the projected expected ICC for k=2 to 7 days using the following formula:

$$ICCk = k \cdot ICC1 / (1 + (k-1) \cdot ICC1)$$

where ICC_1 is the reliability from a single day of monitoring, and ICC_k is the projected reliability for k days [21,33]. Then,

XSL•FO
RenderX

for each bootstrap draw of ICC_1, we applied the following formula to estimate the number of monitoring days needed to reach thresholds of ICC=0.70, 0.80, and 0.90:

$$k* = ICCtarget(1-ICC1)\ ICC1(1-ICCtarget)$$

Here, ICC_{target} is the desired reliability level (0.70, 0.80, or 0.90), and k^* is the projected number of days needed to achieve that level. We summarized the resulting k^* distribution as the median (95% CI) and showed the full ICC vs days curves (with 95% bootstrap CI). For interpretation, we focus on the point at which the lower bound of the 95% CI of the projected ICC curve first reaches ≥0.80, a commonly accepted threshold for adequate reliability in behavioral research [34]. Projected monitoring durations were rounded up to the nearest whole day for practical interpretation.

The Spearman-Brown projections assume that repeated monitoring days function as exchangeable indicators of the same underlying average daily behavior. This assumption may be imperfect for smartphone screen time because use can vary systematically by day type. Therefore, the projected monitoring durations were interpreted as estimates of the number of valid days needed to estimate the average daily smartphone duration within the observed 7-day monitoring period. They were not interpreted as evidence that the same number of days would reliably estimate longer-term habitual smartphone use across weeks, seasons, or other contexts.

To assess potential differences between included and excluded participants, we compared available demographic characteristics by inclusion status within each data source. Because only 7 Android Chronicle participants were excluded, Android comparisons were interpreted descriptively. For iOS Screenshot participants, age was compared using a 2-tailed Welch *t* test, and categorical characteristics were compared using Fisher exact test. All statistical analyses were conducted in R (version 4.4.0), and statistical significance was set at $P<.05$.

Reporting Guidelines

This manuscript was reviewed using the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist for observational studies. The completed checklist is provided in Checklist 1.

Results

Participant Characteristics

Of the 316 participants with objective smartphone use data available for screening, 229 (72.5%) were included in the final analytic sample (Table 1). For Android (Chronicle), 100 of 107 participants (93.5%) had at least 7 consecutive valid days after cleaning. For iOS Screenshot, 129 of 209 participants (61.7%) had at least 7 consecutive valid screenshots. Because only 7 Android Chronicle participants were excluded, included vs excluded comparisons were interpreted descriptively. Among iOS Screenshot participants, included and excluded participants did not differ significantly by age, sex, ethnicity, employment, or income (all $P>.05$), but differed by race ($P=.001$) and education ($P=.002$). Included iOS participants were more likely to be White and to have a college degree or higher than excluded iOS participants.

Table . Participant inclusion by smartphone data source.

Data source	Available for screening, n	Included, n (%)	Excluded, n (%)
Android Chronicle	107	100 (93.5)	7 (6.5)
iOS Screenshot	209	129 (61.7)	80 (38.3)
Total	316	229 (72.5)	87 (27.5)

For the final analytic sample, the mean age was 36.4 years (SD 5.7), and the sample was predominantly female primary caregivers (218/229, 95.2%), which is typical of child and family-based research [35]. Because data sources were platform-specific, Android and iOS users represent distinct subsamples, with some demographic differences (Table 2). In particular, iOS Screenshot participants were more likely to have a college degree or higher, be employed, and report household income of US $100,000 or more. These differences should be considered when interpreting any descriptive comparisons between platform and data-source groups. However, these differences were not adjusted for, as the focus of this analysis was on within-sample measurement reliability rather than between-group comparisons.

XSL•FO
RenderX

Table . Participant demographics by smartphone screen time data source.

Characteristics	Android Chronicle (n=100)	iOS Screenshot (n=129)
Age (y), mean (SD)	36.6 (6.8)	36.3 (4.7)
Sex, n (%)		
Female	94 (94)	124 (96.1)
Race, n (%)		
Black	19 (19)	15 (11.6)
White	74 (74)	104 (80.6)
Asian, more than 1 race, or other	7 (7)	10 (7.8)
Hispanic ethnicity (yes), n (%)	5 (5)	8 (6.2)
Education[a], n (%)		
High school or less	11 (11)	3 (2.3)
Some college or 2-year degree	39 (39)	20 (15.5)
College graduate or higher	50 (50)	106 (82.2)
Employment[a], n (%)		
Employed	54 (54)	98 (76.0)
Student	28 (28)	19 (14.7)
Unemployed	11 (11)	5 (3.9)
Other	7 (7)	7 (5.4)
Income[a], n (%)		
<US $20,000	10 (10)	2 (1.6)
US $20,000-$49,000	23 (23)	13 (10.1)
US $50,000-$99,000	39 (39)	39 (30.2)
≥US $100,000	28 (28)	75 (58.1)

[a]Android Chronicle and iOS Screenshot participants differed on some demographic characteristics. In particular, iOS Screenshot participants were more likely to have a college degree or higher, be employed, and report household income ≥US $100,000.

Data Source and Day Type Differences

A linear mixed-effects model with a random intercept for participant, including both data source and day type as fixed effects, showed no significant difference in mean daily screen time between iOS and Android users (*B*=15.7 min, SE=21.4, 95% CI –26.2 to 57.5; *P*=.46). However, screen time was significantly lower on weekends than on weekdays (*B*=–16.1 min, SE=5.6, 95% CI –27.0 to –5.2; *P*=.004). The interaction between data source and day type was not significant (*B*=6.5 min, SE=11.3; *P*=.56), indicating that the weekend-weekday difference did not vary by data source. Reliability analyses were conducted across all 7 monitoring days rather than separately by day type because the aim was to estimate the reliability of a typical week of smartphone use. Mean daily screen time by day of the week is shown in Multimedia Appendix 1.

Reliability and Monitoring Duration

Table 3 presents the reliability results for daily smartphone screen time. For Android Chronicle data, the single-day ICC was 0.69 (95% CI 0.63-0.76). For iOS Screenshot data, it was 0.71 (95% CI 0.65-0.77). For interpretability, projected monitoring durations were rounded up to the nearest whole day when deriving practical recommendations. Using Spearman-Brown projections based on the point estimates, 2 monitoring days were sufficient for both Android Chronicle and iOS Screenshot data to achieve projected ICC≥0.80 (Figure 3). Using the more conservative lower-bound approach, 3 days were required for both data sources. Sensitivity analyses using combined session and glance duration produced a single-day ICC of 0.64 (95% CI 0.57-0.72). The projected monitoring durations were 1.3 (95% CI 1.0-1.7) days, 2.2 (95% CI 1.7-2.9) days, and 5.0 (95% CI 3.9-6.6) days for ICC thresholds of 0.70, 0.80, and 0.90, respectively. After rounding, both the point-estimate and conservative recommendations were 3 days to achieve ICC≥0.80.

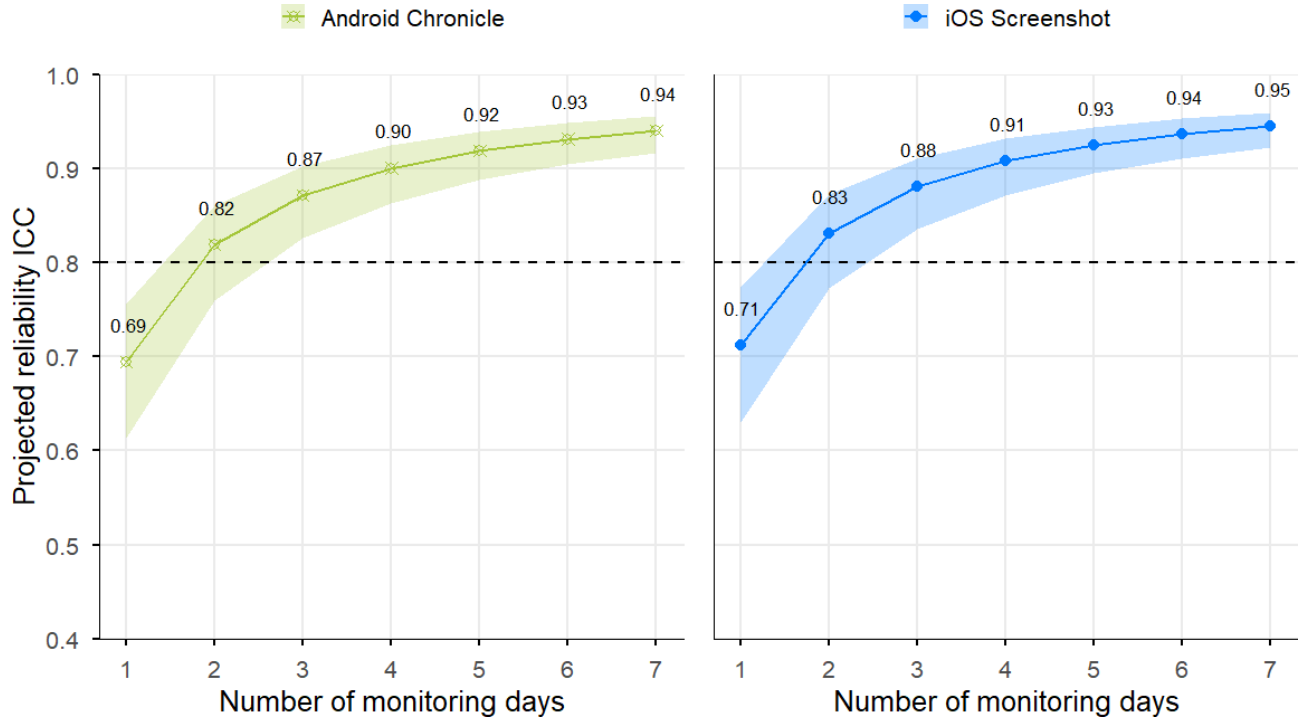
XSL•FO
RenderX

Table . Single-day intraclass correlation coefficients (ICCs; 95% bootstrapped CI), and projected number of days needed for target reliabilities of 0.7, 0.8, and 0.9, by data source.

Data source	Single-day ICC (95% CI)	Projected number of days (95% CI)			Between-person variance	Within-person variance	Practical recommendation for ICC≥0.80[a]
		ICC=0.7	ICC=0.8	ICC=0.9			
Android Chronicle (n=100)	0.69 (0.63-0.76)	1.0 (0.8-1.3)	1.8 (1.4-2.3)	4.0 (3.1-5.1)	26,788	11,804	Point estimate: 2 days; conservative: 3 days
iOS Screenshot (n=129)	0.71 (0.65-0.77)	0.9 (0.7-1.3)	1.6 (1.2-2.2)	3.6 (2.8-4.8)	22,265	9018	Point estimate: 2 days; conservative: 3 days

[a]Projected monitoring durations were rounded up to the nearest whole day. Conservative recommendations were based on the lower bound of the 95% CI for projected ICC estimates.

Figure 3. Projected reliability of average daily smartphone screen time by number of monitoring days, estimated using the Spearman-Brown prophecy formula. Points show projected intraclass correlation coefficients (ICC [1,1]) estimates for 1 to 7 monitoring days, shaded bands show 95% CIs, and the dashed horizontal line indicates ICC=0.80.



Discussion

Principal Findings

This study examined the minimum number of monitoring days required to reliably estimate average daily smartphone screen time among primary caregivers of preschool-aged children using objective Android and iOS monitoring approaches. Daily screen time duration estimates demonstrated high reliability even with a small number of days. When using the point estimate, projections indicated that 2 monitoring days were sufficient for both Android Chronicle and iOS Screenshot data to achieve ICC≥0.80. When using the more conservative lower-bound criterion of the 95% CIs, 3 days were required for both data sources. These estimates should be interpreted as the number of valid days needed to estimate average daily smartphone duration within a 7-day monitoring period. Although the weekday-weekend difference was modest in absolute terms, screen time differed systematically by day type. Therefore, shortened protocols should include weekend representation. For a 3-day protocol, this would most pragmatically involve at least 2 weekdays and 1 weekend day.

Our findings align with prior research in other forms of objective behavioral monitoring, such as physical activity, where 3 to 7 days of accelerometer wear are typically recommended to obtain reliable estimates of movement behaviors [22-26]. Interestingly, smartphone screen time appears to require even fewer days, potentially reflecting a degree of stability of daily phone use, at least among primary caregivers.

Previous studies that estimated the number of days needed for reliable smartphone measurement relied on correlation-based analyses [18,19]. For instance, Wilcockson et al [18,19] found that 5 days were sufficient to characterize weekly usage using Pearson correlations. Although these studies provide valuable

XSL•FO
RenderX

early insights, correlation-based methods primarily reflect rank-order consistency and do not directly quantify measurement reliability. In other words, correlations measure whether individuals maintain the same relative position over time, but do not quantify measurement reliability in a statistical sense. In contrast, our approach uses ICCs, which directly estimate the proportion of total variance attributable to stable between-person differences in average daily screen time over a week [20]. Combined with the Spearman-Brown formula, this approach allows us to estimate how many days of monitoring are needed to achieve a desired level of reliability (eg, ICC≥0.80) while accounting for both within-person and between-person variability in screen time [34]. Although our results are similar to earlier findings, they provide a methodologically rigorous basis for these recommendations by directly estimating measurement reliability. This strengthens confidence in the use of short monitoring periods for future study design.

Objective methods for monitoring smartphone use have progressed rapidly, with built-in summaries (eg, iOS Screenshot) and passive sensing tools (eg, Android Chronicle) each presenting distinct advantages and trade-offs [36]. Built-in summaries, such as Apple's Screen Time, offer built-in, user-friendly reports of device and app activity, but rely on proprietary algorithms that define and categorize usage and can be updated without warning, limiting insight into what actually counts as screen time. Because Apple Screen Time algorithms are proprietary and Android Chronicle estimates were derived from reconstructed session events, it is difficult to determine whether observed differences in screen time estimates between data sources reflect true behavioral differences, sample differences, or platform-specific measurement differences. However, the purpose of this study was not to establish equivalence between Android and iOS estimates. Rather, we conducted separate analyses to provide reliability estimates for 2 distinct objective data sources commonly available to researchers studying smartphone use.

Passive sensing platforms such as Chronicle, Qustodio, and EARS can capture more granular event-level data and enable analyses beyond aggregate daily or hourly summaries [14,37]. However, these approaches often require complex processing pipelines and vary in validation, cost, and ease of implementation. A strength of the present study is the use of documented and reproducible processing procedures for both passive sensing and screenshot-based data. Future work should prioritize head-to-head comparisons of smartphone monitoring tools, validation of algorithmic definitions, and open standards for deriving digital behavior metrics.

Limitations

Several limitations should be noted when interpreting these findings. First, the sample consisted of primary caregivers of young children, the majority of whom were female and already participating in observational studies about family screen time. This raises questions about selection bias, observer effects, and generalizability to other populations. Caregivers of preschool-aged children may also have more constrained and predictable daily routines than adolescents, older adults, or adults without young children [38]. In addition, because inclusion required 7 consecutive valid monitoring days, the analytic sample may overrepresent participants with greater adherence or more complete data capture. This concern may be particularly relevant for iOS Screenshot data, where included participants differed from excluded participants in race and education.

Second, Android Chronicle and iOS Screenshot data came from different platform and data source groups, and iOS Screenshot data were collected only in the cohort study. These groups differed in several demographic characteristics, including education, income, and employment. Because the study was designed to estimate reliability separately within each data source rather than to compare platforms directly, these differences were not adjusted for statistically. However, they limit interpretation of any descriptive Android-iOS differences in screen time estimates.

Third, a single 7-day monitoring window may not capture longer-term habitual smartphone behavior, as routines and occasional deviations may vary across weeks or seasons. For participants with more than 7 consecutive complete days available, the analytic window was defined as the first 7 days from the first qualifying consecutive sequence after cleaning; therefore, we cannot rule out the possibility that smartphone behavior differed during other portions of the monitoring period. This limitation in the monitoring window is important for interpreting the Spearman-Brown projections. The Spearman-Brown approach assumes that repeated monitoring days are exchangeable indicators of the same underlying behavior, but this assumption may be imperfect when behavior varies systematically by day type or over longer time frames. Recent work in physical activity measurement has shown that Spearman-Brown projections from short observation windows can overestimate longer-term aggregate reliability, and has proposed directly estimating reliability using repeated observation windows [39]. The present study was not designed to support this type of repeated-window analysis. Therefore, our findings should be interpreted as estimates of the number of valid days needed to estimate average daily smartphone duration within a 7-day monitoring period, rather than the number of days needed to characterize longer-term habitual smartphone use.

Scope of Inference

The present findings apply specifically to total daily smartphone use duration measured over a 7-day monitoring window. Although total duration is a common and practical exposure metric, it does not capture other potentially important dimensions of smartphone behavior, including frequency of pickups, timing of use, fragmentation or clustering of use across the day, duration of individual sessions, app-specific use, or the content and context of interactions [40-42]. Reliability estimates for total daily duration should, therefore, not be assumed to generalize to these more granular behavioral metrics. Future work should examine how many monitoring days are needed to reliably estimate pattern-based and content-specific smartphone exposures, such as social media duration, messaging

XSL•FO
RenderX

frequency, nighttime use, checking behavior, or fragmented vs sustained use.

Conclusions

In conclusion, this study suggests that relatively few valid days of objective smartphone monitoring may be sufficient to estimate average daily smartphone screen time within a 7-day monitoring period among caregivers of preschool-aged children. These findings support the feasibility of brief monitoring protocols when some days are missing, invalid, or incomplete, helping to balance measurement reliability with participant burden and logistical constraints. However, these estimates should not be interpreted as establishing the number of days needed to measure longer-term habitual smartphone use or more granular smartphone behaviors beyond daily duration. Future studies with longer repeated monitoring windows should examine whether short-term reliability projections align with reliability across distinct weeks or longer time frames and across additional metrics.

Acknowledgments

The authors thank the participating families and study staff who contributed to data collection and management. The generative AI tool ChatGPT (OpenAI) was used for proofreading. All substantive decisions, analyses, interpretation, and final manuscript content were reviewed and approved by the authors, who take full responsibility for the accuracy and integrity of the work.

Funding

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award P20GM130420 and by the Eunice Kennedy Shriver National Institute of Child Health and Human Development under award R01HD112311. The funders had no role in the study design, data collection, analysis, interpretation, manuscript preparation, or decision to submit the manuscript for publication.

Data Availability

Analysis scripts used to calculate intraclass correlation coefficients, bootstrap confidence intervals, Spearman-Brown projections, and sensitivity analyses are publicly available on GitHub [43]. The participant-level analytic dataset is not publicly available because data collection and primary outcome analyses for the parent study are ongoing. An anonymized, aggregated dataset sufficient to reproduce the analyses will be made available from the corresponding author upon reasonable request and will be deposited publicly following completion of parent study data collection and primary outcome analyses.

Authors' Contributions

Conceptualization: JC, BA
Data curation: TA, MR
Formal analysis: JC
Investigation: JSR, HMW, SB, HP, OLF, AJH, KK
Methodology: RG, SN, JC
Project administration: TA, MR
Supervision: RGW, BA, MWB
Writing – original draft: JC
Writing – review & editing: JC, BA, TA, MR, RG, SN, JSR, HMW, SB, HP, OLF, AJH, KK, RGW, MWB
All authors approved the final manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Multimedia Appendix 1
Mean (SD) of daily smartphone screen time in minutes.
[PNG File, 20 KB - mhealth_v14i1e82846_app1.png]

Checklist 1
STROBE checklist.
[DOCX File, 38 KB - mhealth_v14i1e82846_app2.docx]

References

1. Mobile fact sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2025-08-17]

XSL•FO
RenderX

2. Ratan ZA, Parrish AM, Zaman SB, Alotaibi MS, Hosseinzadeh H. Smartphone addiction and associated health outcomes in adult populations: a systematic review. Int J Environ Res Public Health 2021 Nov 22;18(22):12257. [doi: 10.3390/ijerph182212257] [Medline: 34832011]
3. Wacks Y, Weinstein AM. Excessive smartphone use is associated with health problems in adolescents and young adults. Front Psychiatry 2021;12:669042. [doi: 10.3389/fpsyt.2021.669042] [Medline: 34140904]
4. Adamczewska-Chmiel K, Dudzic K, Chmiela T, Gorzkowska A. Smartphones, the epidemic of the 21st century: a possible source of addictions and neuropsychiatric consequences. Int J Environ Res Public Health 2022 Apr 23;19(9):5152. [doi: 10.3390/ijerph19095152] [Medline: 35564547]
5. Beynon A, Hendry D, Lund Rasmussen C, et al. Measurement method options to investigate digital screen technology use by children and adolescents: a narrative review. Children (Basel) 2024 Jun 21;11(7):754. [doi: 10.3390/children11070754] [Medline: 39062204]
6. Kaye LK, Orben A, Ellis DA, Hunter SC, Houghton S. The conceptual and methodological mayhem of “screen time”. Int J Environ Res Public Health 2020 May 22;17(10):3661. [doi: 10.3390/ijerph17103661] [Medline: 32456054]
7. Parry DA, Davidson BI, Sewall CJR, Fisher JT, Mieczkowski H, Quintana DS. A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. Nat Hum Behav 2021 Nov;5(11):1535-1547. [doi: 10.1038/s41562-021-01117-5] [Medline: 34002052]
8. Molaib KM, Sun X, Ram N, Reeves B, Robinson TN. Agreement between self-reported and objectively measured smartphone use among adolescents and adults. Comput Hum Behav Rep 2025 Mar;17:100569. [doi: 10.1016/j.chbr.2024.100569]
9. Júdice PB, Sousa-Sá E, Palmeira AL. Discrepancies between self-reported and objectively measured smartphone screen time: before and during lockdown. J Prev (2022) 2023 Jun;44(3):291-307. [doi: 10.1007/s10935-023-00724-4] [Medline: 36692818]
10. Griffioen N, Scholten H, Lichtwarck-Aschoff A, van Rooij M, Granic I. Everyone does it—differently: a window into emerging adults’ smartphone use. Humanit Soc Sci Commun 2021;8:177. [doi: 10.1057/s41599-021-00863-1]
11. Parry D, Torth R. Extracting meaningful measures of smartphone usage from android event log data: a methodological primer. Comput Commun Res 2025 Oct 31;7(1):1. [doi: 10.5117/CCR2025.1.8.PARR]
12. Barr R, Kirkorian H, Radesky J, et al. Beyond screen time: a synergistic approach to a more comprehensive assessment of family media exposure during early childhood. Front Psychol 2020;11:1283. [doi: 10.3389/fpsyg.2020.01283] [Medline: 32754078]
13. Radesky JS, Weeks HM, Ball R, et al. Young children’s use of smartphones and tablets. Pediatrics 2020 Jul;146(1):e20193518. [doi: 10.1542/peds.2019-3518] [Medline: 32482771]
14. Lind MN, Kahn LE, Crowley R, Reed W, Wicks G, Allen NB. Reintroducing the Effortless Assessment Research System (EARS). JMIR Ment Health 2023 Apr 26;10:e38920. [doi: 10.2196/38920] [Medline: 37099361]
15. Reeves B, Ram N, Robinson TN, et al. Screenomics: a framework to capture and analyze personal life experiences and the ways that technology shapes them. Hum Comput Interact 2021;36(2):150-201. [doi: 10.1080/07370024.2019.1578652] [Medline: 33867652]
16. Schnauber-Stockmann A, Scharkow M, Karnowski V, Naab TK, Schlütz DM, Pressmann P. Distinguishing person-specific from situation-specific variation in media use: a meta-analysis. Communic Res 2024;52(2):143-172. [doi: 10.1177/00936502241262664]
17. Bayer JB, LaRose R. Technology habits: progress, problems, and prospects. In: Verplanken B, editor. The Psychology of Habit: Theory, Mechanisms: Springer; 2018:111-130. [doi: 10.1007/978-3-319-97529-0_7]
18. Wilcockson TDW, Ellis DA, Shaw H. Determining typical smartphone usage: what data do we need? Cyberpsychol Behav Soc Netw 2018 Jun;21(6):395-398. [doi: 10.1089/cyber.2017.0652] [Medline: 29781716]
19. Pan YC, Lin HH, Chiu YC, Lin SH, Lin YH. Temporal stability of smartphone use data: determining fundamental time unit and independent cycle. JMIR mHealth uHealth 2019 Mar 26;7(3):e12171. [doi: 10.2196/12171] [Medline: 30912751]
20. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163. [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
21. de Vet HCW, Mokkink LB, Mosmuller DG, Terwee CB. Spearman-Brown prophecy formula and Cronbach’s alpha: different faces of reliability and opportunities for new applications. J Clin Epidemiol 2017 May;85:45-49. [doi: 10.1016/j.jclinepi.2017.01.013] [Medline: 28342902]
22. Hart TL, Swartz AM, Cashin SE, Strath SJ. How many days of monitoring predict physical activity and sedentary behaviour in older adults? Int J Behav Nutr Phys Act 2011 Jun 16;8:62. [doi: 10.1186/1479-5868-8-62] [Medline: 21679426]
23. Aadland E, Ylvisåker E. Reliability of objectively measured sedentary time and physical activity in adults. PLoS One 2015;10(7):e0133296. [doi: 10.1371/journal.pone.0133296] [Medline: 26192184]
24. Aguilar-Farias N, Martino-Fuentealba P, Salom-Diaz N, Brown WJ. How many days are enough for measuring weekly activity behaviours with the activPAL in adults? J Sci Med Sport 2019 Jun;22(6):684-688. [doi: 10.1016/j.jsams.2018.12.004] [Medline: 30579814]
25. Sasaki JE, Júnior JH, Meneguci J, et al. Number of days required for reliably estimating physical activity and sedentary behaviour from accelerometer data in older adults. J Sports Sci 2018 Jul;36(14):1572-1577. [doi: 10.1080/02640414.2017.1403527] [Medline: 29125022]

XSL•FO
RenderX

26. da Silva SG, Evenson KR, Ekelund U, et al. How many days are needed to estimate wrist-worn accelerometry-assessed physical activity during the second trimester in pregnancy? PLoS One 2019;14(6):e0211442. [doi: 10.1371/journal.pone.0211442] [Medline: 31246953]
27. Parker H, Burkart S, Reesor-Oyer L, et al. Feasibility of measuring screen time, activity, and context among families with preschoolers: intensive longitudinal pilot study. JMIR Form Res 2022 Sep 29;6(9):e40572. [doi: 10.2196/40572] [Medline: 36173677]
28. Reesor-Oyer L, Parker H, Burkart S, et al. Measuring microtemporal processes underlying preschoolers' screen use and behavioral health: protocol for the Tots and Tech study. JMIR Res Protoc 2022 Sep 28;11(9):e36240. [doi: 10.2196/36240] [Medline: 36169993]
29. Walch O, Alam U. Arcascope/screen-scrape. GitHub. URL: https://github.com/Arcascope/screen-scrape [accessed 2026-08-04]
30. Chronicle-android-preprocessing. GitHub. URL: https://github.com/joshculverhouse/chronicle-android-preprocessing [accessed 2026-08-04]
31. Chronicle-preprocessed-cleaning. GitHub. URL: https://github.com/joshculverhouse/chronicle-preprocessed-cleaning [accessed 2026-08-04]
32. Gamer M, Lemon J, Fellows I, Singh P. irr: various coefficients of interrater reliability and agreement. CRAN: contributed packages. 2005. URL: https://doi.org/10.32614/CRAN.package.irr [accessed 2026-08-04]
33. Spearman C. The proof and measurement of association between two things. Int J Epidemiol 2010 Oct;39(5):1137-1150. [doi: 10.1093/ije/dyq191] [Medline: 21051364]
34. Baranowski T, de Moor C. How many days was that? Intra-individual variability and physical activity assessment. Res Q Exerc Sport 2000 Jun;71 Suppl 2:74-78. [doi: 10.1080/02701367.2000.11082789] [Medline: 25680016]
35. Davison KK, Gicevic S, Aftosmes-Tobio A, et al. Fathers' representation in observational studies on parenting and childhood obesity: a systematic review and content analysis. Am J Public Health 2016 Nov;106(11):e14-e21. [doi: 10.2105/AJPH.2016.303391] [Medline: 27631735]
36. Domoff SE, Banga CA, Borgen AL, et al. Use of passive sensing to quantify adolescent mobile device usage: feasibility, acceptability, and preliminary validation of the eMoodie application. Human Behav Emerg Tech 2021 Jan;3(1):63-74. [doi: 10.1002/hbe2.247]
37. Qustodio. URL: https://www.qustodio.com/en/ [accessed 2026-08-04]
38. Bianchi SM, Milkie MA. Work and family research in the first decade of the 21st century. J Marriage Fam 2010 Jun;72(3):705-725. [doi: 10.1111/j.1741-3737.2010.00726.x]
39. Hilden P, Schwartz JE, Pascual C, Diaz KM, Goldsmith J. How many days are needed? Measurement reliability of wearable device data to assess physical activity. PLoS One 2023;18(2):e0282162. [doi: 10.1371/journal.pone.0282162] [Medline: 36827427]
40. Siebers T, Beyens I, Valkenburg PM. The effects of fragmented and sticky smartphone use on distraction and task delay. Mob Media Commun 2023;12(1):45-70. [doi: 10.1177/20501579231193941]
41. Toth R, Parry D, Emmer M. From screen time to daily rhythms. J Quant Descr Digit Media 2025;5. [doi: 10.51685/jqd.2025.019]
42. Yi H, Wu X, Liu X. When smartphones fragment the mind: exploring the links between fragmented smartphone use and anxiety among college students through distraction and procrastination. Research Square. Preprint posted online on Oct 30, 2025. [doi: 10.21203/rs.3.rs-7600425/v1]
43. Smartphone-monitoring-reliability. GitHub. URL: https://github.com/joshculverhouse/smartphone-monitoring-reliability [accessed 2026-08-04]

Abbreviations

EARS: Effortless Assessment Research System
ICC: intraclass correlation coefficient
IRB: institutional review board

XSL•FO
RenderX

Edited by L Buis; submitted 22.Aug.2025; peer-reviewed by B Singh, R Toth, U Alam; revised version received 17.Jul.2026; accepted 29.Jul.2026; published 10.Sep.2026.

Please cite as:
Culverhouse J, Adair T, Restino M, Weeks HM, Radesky JS, Weaver RG, Burkart S, Parker H, Finnegan OL, Holmes AJ, Kiely K, Ghosal R, Nelakuditi S, Beets MW, Armstrong B
Objective Smartphone Screen Time Monitoring Among Caregivers of Preschool-Aged Children: Reliability Study of Monitoring Duration
JMIR Mhealth Uhealth 2026;14:e82846
URL: https://mhealth.jmir.org/2026/1/e82846
doi:10.2196/82846

© Joshua Culverhouse, Taylor Adair, Meghan Restino, Heidi M Weeks, Jenny S Radesky, R Glenn Weaver, Sarah Burkart, Hannah Parker, Olivia L Finnegan, Anthony J Holmes, Keagan Kiely, Rahul Ghosal, Srihari Nelakuditi, Michael W Beets, Bridget Armstrong. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Distinguishing Common Digital Phenotyping and Self-Report Parameters for Monitoring and Predicting Depression: Scoping Review

Lisa Busshart[1*]; Milica Petrovic[1*], PhD; Rebeka Amin[1], PhD; Ulrich Hegerl[1,2*], Prof Dr Med

[1]Department of Psychiatry, Psychosomatics and Psychotherapy, University Hospital Frankfurt, Research Center of German Foundation for Depression and Suicide Prevention, Heinrich-Hoffmann-Str. 10, Frankfurt am Main, Germany

[2]Department for Psychiatry, Psychosomatics and Psychotherapy, Goethe University Frankfurt, Goethe Research Professorship, Frankfurt am Main, Germany

[*]these authors contributed equally

Corresponding Author:
Lisa Busshart
Department of Psychiatry, Psychosomatics and Psychotherapy, University Hospital Frankfurt, Research Center of German Foundation for Depression and Suicide Prevention, Heinrich-Hoffmann-Str. 10, Frankfurt am Main, Germany

Abstract

Background: Digital health interventions incorporating self-management strategies are increasingly used to support individuals in managing depression. These interventions often leverage self-monitoring and passive sensor-based data collection to provide personalized feedback, guiding behavioral change. With the proliferation of smartphones and wearable devices, there is growing potential to continuously collect behavioral and physiological data. However, a major limitation in the field is the lack of consolidated evidence on which specific parameters are most useful for monitoring and predicting depression-related outcomes.

Objective: This scoping review aims to identify and synthesize common digital phenotyping and self-report parameters for monitoring and predicting depression. Specifically, it addresses the methodological and knowledge gap concerning which types of sensor-based and self-reported data are most frequently used and which demonstrate predictive value in tracking changes in depressive symptoms across digital platforms.

Methods: A literature search was conducted across 4 databases, including PubMed, Embase, Cochrane Library, and the Web of Science Core Collection. Articles published between January 1, 2021, and November 26, 2025, were included. Eligible studies included adults (≥18 years) with depression confirmed through validated clinical measures and using digital approaches that collected passive sensor data, self-reports, or both. Studies focusing on comorbid disorders, nondigital interventions, or not reporting depression-related outcomes were excluded. The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines were followed, and a 5-stage methodological framework for scoping reviews was implemented. Quality assessment was performed using the Downs and Black Instrument and the Mixed Methods Appraisal Tool (MMAT).

Results: Nineteen studies were included, comprising a total of 85,193 participants. Most studies used smartphone- or wearable-based tools, with passive sensing as the predominant data source and Patient Health Questionnaire-9 (PHQ-9) as the most commonly used depression measure. Five overarching parameter categories were identified: (1) physical activity and location, (2) behavioral patterns, (3) physiological signals, (4) sleep indicators, and (5) sociability and self-reported assessments. Within these categories, 11 metrics, including step count, heart rate variability, sleep duration and mood self-ratings, were most frequently reported. Most studies used a multimodal digital phenotyping approach, integrating passive sensor-derived data with active user-reported input, enabling more individualized symptom monitoring over time.

Conclusions: This scoping review provides a novel synthesis of common digital parameters used across diverse tools for monitoring and predicting depression, moving beyond tool- or modality-specific perspectives adopted in prior reviews. Unlike existing reviews focusing on individual sensing modalities, prediction methods, or intervention effectiveness, this review maps shared parameters across observational, predictive, and interventional studies. By identifying convergent digital markers, the review supports comparability across studies and informs future model development. These findings have practical implications for the design of scalable digital mental health tools and for translating digital phenotyping into real-world clinical and self-management contexts.

(JMIR Mhealth Uhealth 2026;14:e70840) doi:10.2196/70840

XSL•FO
RenderX

KEYWORDS

depression; digital phenotyping; sensor data; objective data; self-monitoring

Introduction

Advancements in personal digital devices and sensor technologies have enabled continuous and unobtrusive data collection from individuals in real-world settings, expanding the possibilities in health care data management [1]. This has given rise to digital phenotyping, defined as the moment-by-moment quantification of individual-level behavior and aspects of physiology using data from personal smartphones and wearable devices of users [2].

The sensors commonly embedded in these devices can be broadly categorized into (1) motion sensors, (2) location sensors, and (3) environment sensors [3]. More specifically, smartphone-based sensors include linear acceleration, rotation vector, humidity, orientation, gravity, magnetic field, accelerometer, ambient temperature, gyroscope, light, pressure, microphone, proximity, and others [4]. Each device uses multiple sensors working together to provide optimal output for the user. For instance, environmental sensors, including humidity, illumination intensity, surrounding pressure, and temperature sensors enable automatic smartphone screen brightness adjustment or calculation of the dew point for the day [3]. Similarly, the geomagnetic sensor and accelerometer show user-relative location to the North Pole, providing basic data for location sensors. Motion sensors assess acceleration, speed of revolution, gravity, curl vector values, and drift, allowing tracking of multiple motions of the user, such as car direction (eg, navigation application), steps walked (eg, fitness or exercise mobile applications), and speed of movement (eg, user car speed) [3,5].

Unobtrusive tracking of health data relying on physiological parameters such as heart rate, step count, sleep, and others commonly occurs by using sensor fusion (ie, putting multiple signals together) [6]. For instance, biosensors can track physical parameters such as heart rate [7], while location sensors can disclose patient activity and potentially help infer how often outdoor activities occur, which is particularly important for individuals living with mood disorders such as depression [8]. Social patterns can be monitored through the aggregated amount of mobile phone calls and textual messages exchanged with others, pointing to social activity or potential social withdrawal, often serving as an indicator of a depressive episode [9]. Finally, sensor fusion can be used by combining multiple sensors such as accelerometer, ambient light, battery level, microphone, phone lock status, phone unlock status, and screen status for the physiological parameter of sleep that has a significant role in indicating depressive episodes [10].

Some of the earlier works that explored the field of digital phenotyping in mental health include a systematic review [11] of smartphone-based passive sensing, summarizing common data sources such as GPS, accelerometry, and phone usage metrics. Several research papers have outlined the promise of personal sensing approaches in psychiatry while also emphasizing challenges related to data interpretation and real-world implementation [12,13]. Early empirical support for the clinical relevance of smartphone-derived data has been demonstrated through significant correlations between sensor-based measures and depressive symptom severity [14].

Recent reviews have demonstrated growing interest in the use of passively collected data for depression monitoring, reflecting the rapid advancement of digital phenotyping approaches. For example, several studies have focused on specific technologies, such as smartphone-based sensing alone, or emphasized algorithm development for symptom detection rather than broader clinical integration [15-18]. Others have targeted a narrower set of parameters [19], which limits their scope for cross-study synthesis. In contrast, the present review aims to provide a comprehensive overview of the parameters most commonly used in digital interventions for depression, regardless of delivery platform, by synthesizing data across smartphone- and wearable-based sensing studies. This broader focus is intended to support standardization efforts and inform the development of more effective and clinically relevant digital mental health tools.

There is a considerable variability in the types of parameters used, the degree of personalization, and the delivery modes, making it difficult to compare studies or determine which features are most effective [19-22]. This lack of consistency makes it difficult to develop reliable digital health tools. Moreover, health practitioners and end users often lack clear guidance on which digital tools offer meaningful and reliable support.

To address such gaps, this review aims to map the landscape of passively collected parameters used in digital interventions for depression, with the goal of informing future research, standardization, and clinical application.

This review is guided by the following research questions:

1. What passively collected parameters are commonly used in digital interventions for depression to support longitudinal monitoring of depressive symptoms?
    - Which types of passively collected data are used?
    - Which parameters are most frequently used?
    - Which parameters have demonstrated predictive value for changes in symptom severity?

We define a parameter as any sensor-derived measure used to monitor or predict depression-related outcomes.

Methods

Study Design and Framework

This scoping review was conducted in accordance with the 5-stage methodological framework for scoping reviews, including (1) identifying the research question, (2) identifying relevant studies, (3) study selection, (4) data charting and synthesis, and (5) collating, summarizing, and reporting results [23]. The data were systematically collected and reported by

XSL•FO
RenderX

following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines (Checklist 1) [24].

This scoping review was not registered in PROSPERO (International Prospective Register of Systematic Reviews), as PROSPERO does not accept registrations for scoping review protocols at the time of this writing.

Data Sources and Search Strategy

We conducted a comprehensive search of 4 electronic databases, including PubMed, Embase, Cochrane Library, and Web of Science Core Collection. Study registries were searched, including ClinicalTrials.gov and the World Health Organization International Clinical Trials Registry Platform (WHO ICTRP). The findings from these registries are categorized under "Others" in the PRISMA flow chart. No online or print sources were purposefully searched or browsed outside of the 4 electronic databases. No table of contents, conference proceedings, organizational websites, or print materials were hand-searched for this review. However, reference lists of included articles were also manually screened to identify additional eligible studies, which are categorized under "Other" in the PRISMA flow chart. No databases were searched simultaneously on a single platform, and no additional studies or data were sought by contacting study authors, experts, manufacturers, or others.

Studies published between January 1, 2021, and November 26, 2025, were included to reflect the recent advancements in digital phenotyping and mobile sensor technology in mental health. Our searches applied limits for publication date and language. All searches were last conducted on November 26, 2025 across all included databases, restricted to publications in English or German. These limits were used to ensure that the review captured the most current developments in digital interventions and digital phenotyping, as the technological landscape in this field evolves rapidly. No additional limits (eg, study design filters) were applied. The searches were updated manually by rerunning all database queries prior to final analysis. No automated alerts, citation tracking tools, or continuous updating services were used. The search strategy combined controlled vocabulary where applicable (eg, medical subject headings [MeSH] in PubMed) and free-text terms related to depression, digital health interventions, sensor data, and digital phenotyping. No published search filters were used for this review, and no search strategies from previous literature reviews were adapted or reused. All search strategies were developed manually using combinations of keywords and medical subject headings relevant to depression, digital phenotyping, and digital health interventions. The search strategy did not use any existing validated filters (eg, clinical study filters or methodological filters). No formal search peer review process was conducted for this review.

A summary of keyword combinations is provided in Table 1. For detailed search strategy see PRISMA-ScR Checklist 1 containing a detailed search strategy across databases.

XSL•FO
RenderX

Table . Search strategy with combined keyword terms and date range.

Database	Keywords combined	Date range
PubMed	"Depression" AND "digital health intervention"	January 1, 2021-November 26, 2025
	"Depression" AND "digital phenotyping"	
	"Self-management" OR "Depression" AND "digital phenotyping"	
	"Depression" AND "digital phenotyping" AND "passive intervention"	
	"Depression" AND "sensor data"	
EMBASE	"Depression," "digital health intervention"	January 1, 2021-November 26, 2025
	"Depression," "digital phenotyping"	
	"Self-management" / "Depression," "digital phenotyping"	
	"Depression," "digital phenotyping," "passive intervention"	
	"Depression," "sensor data"	
Cochrane Library	"Depression" AND "digital health intervention"	January 1, 2021-November 26, 2025
	"Depression" AND "digital phenotyping"	
	"Self-management" OR "Depression" AND "digital phenotyping"	
	"Depression" AND "digital phenotyping" AND "passive intervention"	
	"Depression" AND "sensor data"	
Web of Science Core Collection	"Depression," "digital health intervention"	January 1, 2021-November 26, 2025
	"Depression," "digital phenotyping"	
	"Self-management" / "Depression," "digital phenotyping"	
	"Depression," "digital phenotyping," "passive intervention"	
	"Depression," "sensor data"	

A total of 3243 records were identified during the study selection process, screened (n=1406), and reviewed for eligibility (n=134). After applying inclusion and exclusion criteria, a total of 19 studies were included in the final analysis. Reasons for exclusion included studies without digital interventions, nonclinical populations, and studies focused on comorbid conditions or unrelated mental health disorders.

The total number of records identified within each database and other information sources was 3243 (PubMed n=518; Cochrane Library n=1256; EMBASE n=791; Web of Science n=276; and Others n=402). Relevant records were identified from 4 databases (n=518) and clinical trial registers (n=2725). After duplicate removal, 1406 records were screened. Following full-text screening and exclusions, 19 studies were included in the final review. Common reasons for exclusion included lack of intervention focus (n=27), nonclinical populations (n=13), or nondigital approaches (n=9). Records retrieved from all databases and clinical trial registries were exported into Microsoft Excel and manually reviewed to identify and remove duplicates. Deduplication was performed by comparing study titles, authors, publication years, and digital object identifiers. No automated or software-assisted deduplication tools (eg, Covidence [Veritas Health Innovation] or EndNote [Clarivate]) were used.

Eligibility Criteria

Studies including randomized controlled trials, observational studies, and mixed method primary studies were considered. Eligible studies include digital health interventions for adults (≥18 years) concretely addressing depression that is assessed primarily with clinical measures such as the Patient Health Questionnaire-9 (PHQ-9) or similar validated instruments. The population focus is mainly on participants that, at the time of study participation, were experiencing depression, noted with a clinical measurement tool and digital phenotyping methodology. Study reports on outcomes related to the assessment of digital phenotyping for depression or dedicate a substantial amount of work to such reporting. The timeframe for published studies was limited from January 1, 2021 to November 26, 2025 to capture the most recent advances in digital phenotyping and mobile sensor monitoring in mental health. This period marks a notable shift from early feasibility and pilot work to more sophisticated methods, including multisite validation, machine learning–based prediction models, and integration of wearables and smartphone sensors [25]. By focusing on this timeframe, we ensure that the review remains aligned with current technological capabilities and clinically relevant research directions.

XSL•FO
RenderX

Study Selection

Title and abstract preselection was performed by 2 authors LB and MP. Screening questions included:

1. Does the study use a digital intervention for depression delivered via a digital tool (eg, mobile app, wearable, web-based platform)?
2. Does the study report on depression outcomes assessed through clinical instruments and/or digital phenotyping?

Relevant studies were retrieved and screened in full text by one author LB. After removing duplicates, LB, MP, and RA cross-checked the full texts against the inclusion and exclusion criteria. UH served as a supervisory reviewer. In an online meeting, final study selections were presented by LB and MP, who provided justification for inclusion decisions to UH.

Inclusion Criteria

Studies were included if they met the following criteria:

1. Examined a digital intervention aimed at monitoring or improving depressive symptoms
2. Delivered via smartphone apps, wrist-worn devices, web-based platforms, offline software, or automated interactive voice response systems
3. Included adult participants (≥18 years) with depression, confirmed through validated clinical instruments (eg, PHQ-9)
4. Reported on outcomes related to monitoring and/or symptom improvement using digital phenotyping (passive sensor data and/or active user input)
5. Provided data on device use (eg, number of interventions delivered, duration of engagement)
6. Published in English or German, and accessible via open access or institutional access.

Exclusion Criteria

Studies were excluded if they:

1. Focused on depression comorbid with other conditions (eg, cancer, substance use, pregnancy, or other mental health disorders)
2. Focused on depression in the context of a viral pandemic (eg, COVID-19)
3. Studies use or report on depression based on digital interventions aimed at supporting and improving physical activity [26].

Analysis Approach

Data Extraction, Charting, and Synthesis

A structured Microsoft Excel form was developed by the author LB to extract key information from each included study. Extracted variables included:

- Study characteristics (author, year, and country)
- Participant demographics and sample size
- Type of digital intervention and delivery platform
- Sensor types and data collection methods
- Depression assessment tools (baseline and follow-up)
- Frequency and duration of interventions
- Outcomes related to depression monitoring or prediction

Once data were initially charted, it was presented in an online meeting to co-author MP and cross-checked among author LB and co-author MP. For more detailed procedures for selection of the studies, extraction, and exclusion, please see Multimedia Appendix 1.

Due to heterogeneity in study design and reporting, a narrative synthesis approach was used. Data were charted and grouped into thematic categories of parameters (see Table 2), including physical activity and location, behavioral patterns, physiological signals, sleep indicators, and sociability and self-reported input.

XSL•FO
RenderX

Table . Overview of the retrieved studies in terms of data collection method, type of digital tool, depression measures, main findings, and predictive value for depression reported.

Study ID	Reference identifier	Sample size (depression)	Data type	Digital tool type	Depression measure	Main findings	Predictive value reported
1	[27]	313	Passive + Active	Smartphone app, wearable	PHQ-9[a]	Multiple behavioral and phone use features significantly correlated with PHQ-9	Yes
2	[28]	400	Active	Computer	PHQ-9	Behavioral test metrics significantly predicted depression	Yes
3	[29]	290	Passive	Fitbit	PHQ-9	Variations in heart rate and circadian rhythm linked to depression	Yes
4	[30]	92	Passive	Smartphone app	PHQ-9	User engagement is essential for data reliability	Unclear
5	[31]	81277	Active	Smartphone app	PHQ-9	Minimal average usage; thought diary and tests most used	No
6	[32]	934	Passive	Smartphone app	PHQ-2[b]	Weak correlation between keystroke timing and symptoms	No
7	[33]	65	Passive	Smartphone app	CES-D[c]	Subject-specific standardization improved forecasting	Yes
8	[13]	60	Passive + Active	Smartphone app, Oura Ring	DASS-21[d]	Best prediction from combined smartphone + wearable + mood	Yes
9	[34]	28	Active	Computer	PHQ-8[e]	Digital social rhythm therapy improved depression	Yes
10	[35]	614	Passive	Smartphone app, Fitbit	PHQ-8	Lower engagement correlated with higher depression	Unclear
11	[36]	1	Passive	Smartphone app, wearable	Not stated	Demonstrated potential of ambient light and usage data	Unclear
12	[37]	18	Passive	Video assessment	MADRS[f], MINI[g], CSSRS[h]	Facial and speech features correlated with MADRS	Yes
13	[38]	20	Passive + Active	Smartphone app, wearable	MADRS, HDRS[i]	Self-report had higher predictive correlation than passive sensors	Yes

XSL•FO
RenderX

Study ID	Reference identifier	Sample size (depression)	Data type	Digital tool type	Depression measure	Main findings	Predictive value reported
14	[39]	623	Passive	Smartphone, Fitbit	PHQ-8	Sleep timing features correlated with depression severity	Yes
15	[20]	107	Active	Smartphone app	PHQ-9	Web-based support linked to symptom improvement	Yes
16	[40]	43	Active	Smartphone app	BDI[j]	Smartphone mood ratings correlated with standard assessments	Yes
17	[41]	48	Passive + Active	Smartphone app, wearable	PHQ-9	Higher engagement led to greater symptom improvement	Yes
18	[42]	136	Active	Smartphone app	BDI-II[k]	Significant treatment-related symptom changes	Yes
19	[43]	124	Active	Computer	DASS-21	Personalized feedback led to greater symptom reduction	Yes

[a]PHQ-9: Patient Health Questionnaire-9.

[b]PHQ-2: Patient Health Questionnaire-2.

[c]CES-D: Center for Epidemiological Studies Depression Scale.

[d]DASS-21: Depression Anxiety Stress Scale Short Form 21.

[e]PHQ-8: Patient Health Questionnaire-8.

[f]MADRS: Structured Interview Guide for the Montgomery-Åsberg Depression Rating Scale.

[g]MINI: Mini-International Neuropsychiatric Interview,

[h]CSSRS: Columbia Suicide Severity Rating Scale.

[i]HDRS: Hamilton Depression Rating Scale.

[j]BDI: Beck Depression Inventory.

[k]BDI-II: Beck Depression Inventory, Second Edition.

The frequency of each parameter category and its reported predictive value were also noted.

Quality Appraisal

To assess the quality of the included studies, we used two tools, including (1) the Downs and Black checklist [44] used for randomized controlled trials and quantitative studies and (2) the Mixed Methods Appraisal Tool (MMAT) [45] used for mixed method and nonrandomized studies.

These complementary tools allowed a consistent evaluation of methodological rigor across diverse study designs. Detailed quality assessments are included in Multimedia Appendices 2-4.

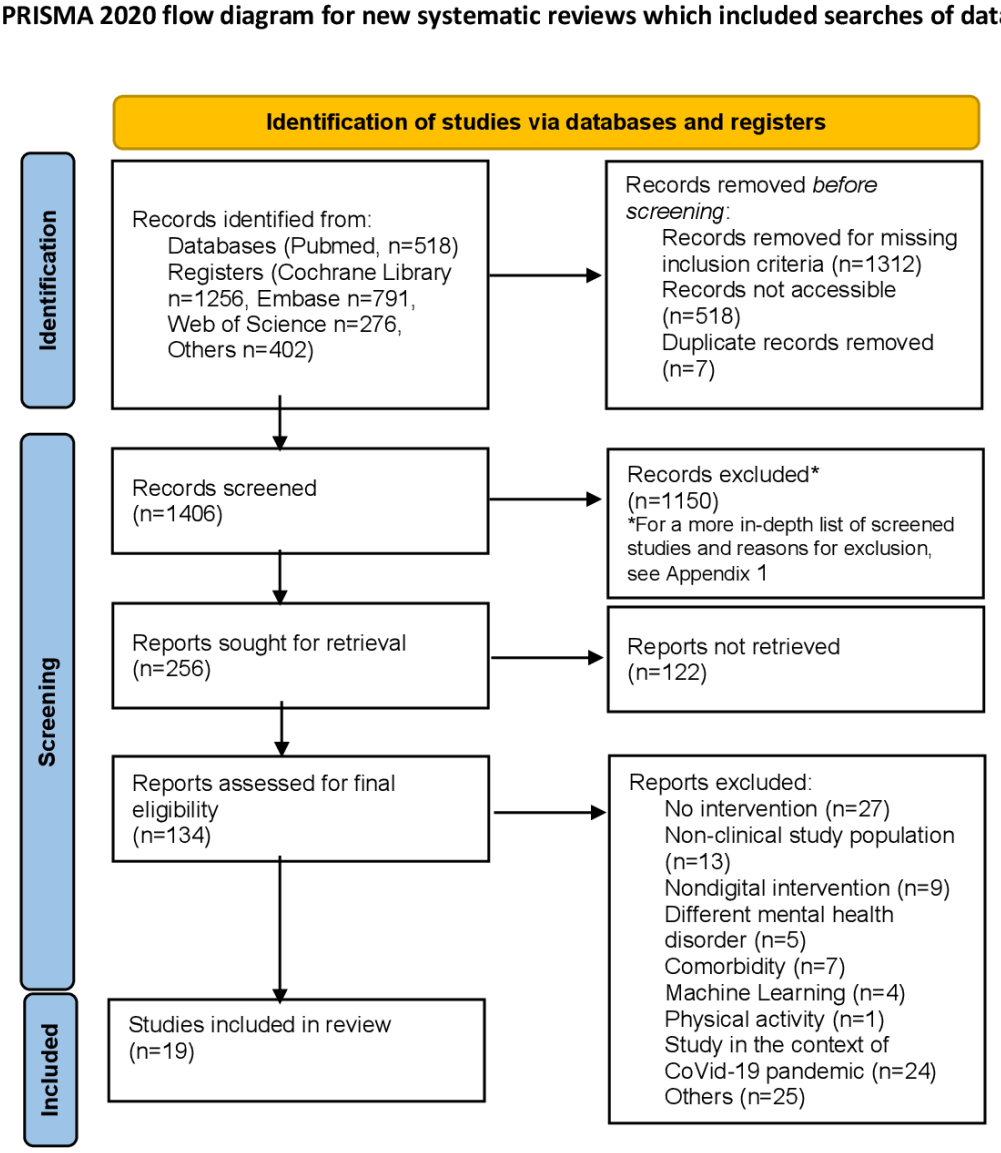
Results

Summary of Included Studies

A total of 19 studies [13,20,27-43] with 85,193 participants across studies were included, covering a diverse range of digital intervention types and data collection methods. The study selection process is shown in Figure 1.

XSL•FO
RenderX

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) flow diagram of the study selection process. Diagram adapted from PRISMA-ScR guidelines [24].



Most digital tools or platforms (n=15) used smartphone apps, either alone or in combination with wearable devices (eg, Fitbit [Fitbit, Inc], Oura Ring [Oura Health Oy]). Three studies (16%) used computer-based platforms, and one used a computer-based platform for video-based digital phenotyping. The included studies varied greatly in design and measurement approaches (see Table 3). The majority were observational studies (n=14), including cross-sectional, longitudinal cohort, retrospective, interformat validation, and feasibility designs, while 5 studies delivered a digital mental health intervention [20,34,41-43]. Three studies were randomized controlled trials [13,42,43], and 2 used a retrospective observational approach [30,39]. Several observational studies used repeated-measures designs to monitor depressive symptoms over time, although study duration and assessment frequency were heterogeneous. One study collected single-participant feasibility data [36], while another followed participants longitudinally for 43 weeks [35].

XSL•FO
RenderX

Table . Description of study design, subtype, number of studies included, and reference number.

Study design and subtype (if applicable)	N	Reference number
Randomized controlled trials (RCTs)	3	[34,42,43]
Nonrandomized interventional pilot studies		
Pilot intervention as well as feasibility with outcomes	2	[37,41]
Observational studies		
Cross-sectional	5	[28,29,31,32,38]
Prospective cohort and longitudinal cohort	5	[13,20,27,33,35]
Retrospective observational	2	[30,39]
Single-participant feasibility and case-style	1	[36]
Interformat validation	1	[40]

Data Collection Approaches

Three data collection types were noted, including passive data collection, active input, self-reports, and multimodal approaches, which combined both passive and active inputs.

Passive data collection (eg, from GPS, accelerometer, and heart rate sensors) was used in 84% of studies (n=16).

Active input and self-reports (eg, mood ratings, diary entries) were used in 58% of studies (n=11).

Multimodal approaches combining both passive and active inputs showed the highest predictive potential and were used in 47 % of studies (n=9).

Clinical Measures Used Across Studies

Validated tools used across studies included Patient Health Questionnaire versions 9, 8 and 2 (PHQ-9, PHQ-8, and PHQ-2), the Center for Epidemiologic Studies Depression Scale (CES-D), the Depression Anxiety and Stress Scale (DASS-21), the Beck Depression Inventory original and revised version (BDI-I and -II), the Montgomery-Åsbergg Depression Rating Scale (MADRS), the Hamilton Depression Rating Scale (HDRS), and the Mini-International Neuropsychiatric Interview (MINI), reflecting a broad clinical foundation across studies (see Table 2).

The PHQ-9 was the most used depression measure, used in 9 studies. The second most used depression measure included PHQ-8 (studies n=3) [34,35,39] while PHQ-2 was used in only one study [32]. Other often used measures were DASS (studies n=2) [13,43], MADRS (studies n=2) [37,38], MINI. (studies n=2) [37,38] and HDRS [38] used in only one study.

Predictive Value

Of the 19 studies, 15 (79%) [13,20,27-29,33-35,37-43] reported predictive or diagnostic associations between digital parameters and depression outcomes. Predictive value was strongest in studies that:

1. Incorporated multisource data fusion [9,13,38]
2. Applied machine learning or modeling techniques for forecasting [28,33].

Sensor Types and Platforms Breakdown

Across the reviewed studies, most digital interventions assessed or observed relied on smartphones as the primary platform, often enhanced with wearable devices (eg, Fitbit, Oura Ring) for additional collection of physiological data.

The most frequently used sensors included accelerometers, gyroscopes, GPS, light sensors, and microphone sensors.

Wearables primarily contributed heart rate and sleep cycle data, while smartphones captured mobility, screen interaction, ambient context, and social communication metrics.

Only a few studies [28,37] used computer-based platforms or video-based assessments, which focused on cognitive tasks, facial expressions, and speech expression as digital biomarkers.

Across retrieved studies the most common device used for digital data collection was a smartphone (n=13), while wearable devices such as sensors and other wrist-worn devices (eg, Fitbit) were used in 8 studies. A detailed list of devices with the main findings reported can be seen in Table 2.

Key Findings

The key findings were grouped into five major categories of passively collected parameters: (1) physical activity and location, (2) behavioral patterns, (3) physiological data, (4) sleep, and (5) sociability and self-reported assessments (see Table 2 and Figure 2). Table 2 summarizes the most commonly used passively collected parameters in digital interventions for depression. Sleep-related parameters such as total sleep time, sleep onset, and heart rate during sleep were particularly prevalent, often inferred through sensor fusion.

Similarly, physical activity metrics like step count and accelerometer data, along with location tracking (eg, distance from home), were commonly used to assess daily functioning and routine stability.

Sociability parameters, including phone call frequency, text message frequency, and app usage, were often used to approximate social withdrawal.

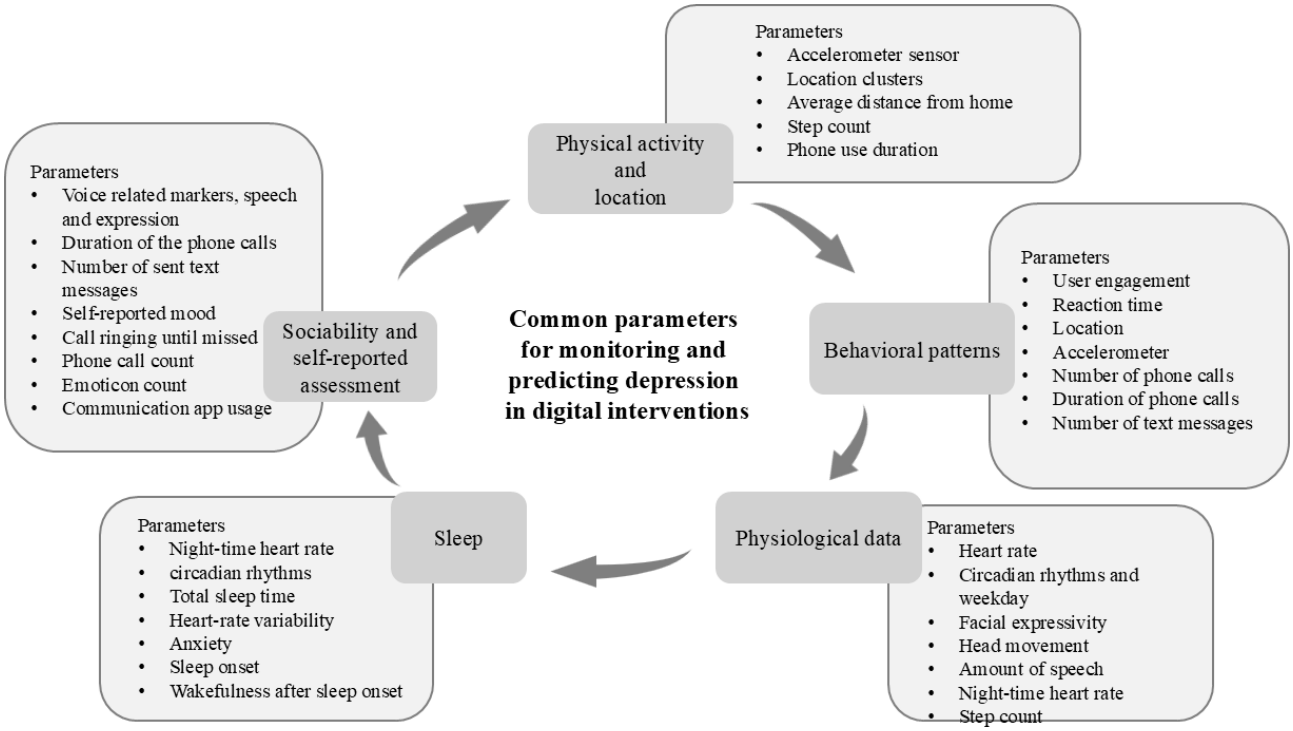
Behavioral data like reaction time or user engagement further contributed to characterizing depressive symptomatology. Figure 2 illustrates how these domains interconnect and highlights their contribution to monitoring and predicting depression.

XSL•FO
RenderX

This multidomain structure, used in 47% of studies, supports a more concrete understanding of symptom trajectories and lays a foundation for developing more reliable personalized digital interventions that incorporate digital parameters.

1. Physical Activity and location parameters—such as step count, phone use duration, and location-based data—were reported in 8 studies and captured through various digital interventions (see Table 4).
2. Behavioral patterns included parameters like user engagement, often assessed via response time to survey notifications. These were also reported in 8 studies.
3. Physiological data encompassed diverse indicators such as heart rate, circadian rhythms, and facial expressivity. These were observed in 8 studies and often overlapped conceptually with parameters categorized under “Sleep”, such as total sleep time, sleep onset, and nighttime heart rate. Sleep-related parameters were collected in 7 studies.
4. Sociability and self-reported assessments involved features inferred from communication behaviors, including the number and duration of phone calls, text message counts, and emoticon use. These were reported in 3 studies. Self-reported mood or symptom assessments, often gathered through in-app or EMA-style prompts, were present in nearly all included studies and frequently used alongside sensor data.

Figure 2. Common domains and parameters within domains used for digital phenotyping in depression monitoring.



XSL•FO
RenderX

Table . Categories included in the monitoring and assessment of depression per study.

Study ID	Sleep	Physical activity	Sociability	Location	Physiological	Other	Total parameters measured
1	Yes	Yes	Yes	No	Yes	No	34
2	No	No	No	No	No	No	—[a]
3	Yes	Yes	Yes	No	Yes	No	126
4	No	Yes	No	Yes	No	No	—
5	No	No	Yes	Yes	No	No	—
6	Yes	No	No	No	No	Yes	54
7	Yes	No	No	Yes	No	Yes	19
8	Yes	Yes	No	Yes	Yes	Yes	—
9	No	No	No	No	No	Yes	—
10	No	No	No	No	No	No	—
11	No	Yes	No	Yes	No	Yes	5
12	No	Yes	No	No	Yes	Yes	—
13	No	No	No	No	No	No	—
14	Yes	No	No	Yes	No	Yes	21
15	No	No	No	No	No	Yes	—
16	Yes	Yes	No	Yes	No	Yes	—
17	No	Yes	No	Yes	Yes	Yes	—
18	No	No	No	No	No	Yes	—
19	No	No	No	No	No	No	—

[a]Not applicable.

Across the studies, the most robust prediction of depressive symptoms emerged from models that combined smartphone and wearable sensor data with self-reported assessments. Although there was considerable heterogeneity in the parameters used, ranging from location variability and facial expressivity to sleep duration and keystroke timing, some overlap was noted. For instance, nighttime heart rate could be classified under either physiological data or sleep, depending on the study context. In such cases, we categorized the parameter according to how it was used in the original study.

Elaboration of Key Findings

Physical Activity and Location

The majority of studies that gathered passive data continually also supplied data about patient's locations. Additionally, these studies tracked patients' physical activity via accelerometer sensor. Results regarding physical activity and location parameters were homogenous. One study described a negative association between variability of locations visited and symptoms of depression (r=−0.21) [13], showing that the more active the participant is (ie, the more locations visited), the lower the symptoms of depression appear to be, which can also infer social context, as more activity could also signify more social interactions. Similarly, another study described a negative correlation between higher MADRS scores and a lower average distance from home (r=−0.25) [38], supporting the argument that greater depression severity is associated with reduced physical activity.

Behavioral Patterns

Studies categorized in behavioral patterns predominantly focused on user engagement and behavior of the patient. One of the included studies revealed that lower user engagement results in higher depression scores [41]. Although, the engagement in this case was defined as the time it took the user to respond to survey notifications. The longer the time users took to respond, the less engaged they were and the more severe their depressive symptoms were shown to be. In line with this, another study noted a correlation between higher engagement and an increase in symptom improvement, described as a decrease in at least one of the depressive symptoms and/or a clinically significant improvement [41]. User engagement can be used to assess the severity of depressive symptoms, and it is also important for gathering data continually, especially for gathering relevant data such as location and accelerometer; hence, active user engagement was found to be essential [30] also in supporting the physical activity and location category.

Apart from engagement rates, other behavioral markers differed in relation to depressive symptoms as well. A prospective cohort study, included in this review, exploring factors associated with the onset and course of mental health conditions during adolescence reported a weak negative association between most of the 54 analyzed keystroke timing features and mental health

XSL•FO
RenderX

symptoms [32]. Machine learning models using keystroke features alone were shown to be unable to predict mental health symptoms, while speech patterns and other associated voice-related markers, such as words per minute, were found to be negatively correlated (r=−0.09) with PHQ-9 [27] indicating that higher depression scores are associated with decreased speech and expression, which potentially has value for the (5) sociability and self-reported assessments category as well.

Physiological Data

Common parameters in this category were difficult to compare, as they vary in nature. Multiple physiological markers were shown to correlate with the severity of depressive symptoms across studies. For instance, one study reported significant increases in multiple digital markers, especially in facial expressivity such as fear and anger [37]. These markers were shown to decrease often in frequency, correlating to changes in the MADRS scores. Concretely, head movement and amount of speech correlated with changes in the MADRS scores [37], while other physiological parameters were associated (r=0.70) with greater severity of depressive symptom, such as greater variation of nighttime heart rate [29]. Based on the daily number of steps, depressive symptoms were associated (r=0.62) with lower regularity of weekday circadian rhythms [29]. In fact, this particular study also reported overall a limited ability of digital biomarkers to detect depression in the whole sample [29]. However, stronger correlations (r=0.8 - 1.0) were described among pairs of variables in alpha power and gamma power spectra in electroencephalography with MADRS [38], indicating that some pairs of variables in electroencephalographic streams can be measured when depression scores are increasing as well, being potentially useful as a digital biomarker in the prediction of depression.

Sleep

Multiple studies reported sleep-related data, consisting of passively gathered data or sleep-related questionnaires such as the Pittsburgh Sleep Quality Index, the Sleep Hygiene Index, or the Epworth Sleepiness Scale. Results were comparatively homogenous. One study reported a significant positive association between the total sleep time and depression (r=0.15), also waking after the sleep onset and depression (r=0.20), as well as heart rate variability and anxiety (r=0.12) [13]. Similarly, another study described a negative correlation (r=−0.12) between reported sleep duration and PHQ-9 [27]. Eventually, a stronger correlation of sleep-onset time with PHQ-8 was found cross-sectionally, and a higher correlation (r=0.24) of wakefulness after sleep-onset with PHQ-8 was also observed longitudinally [39].

Sociability and Self-Reported Assessments

In this category, studies included parameters such as self-reported mood or parameters pertaining to social behavior and sociability (eg, phone calls, text messages, and other communication apps). Since this is a rather broad range, comparison is difficult. In accordance with earlier described results, one study reported a correlation of PHQ-9 with voice-related markers such as Voice Diary words per minute, Voice Diary sentiment, and Voice Diary duration (Mood Study App, Verily Life Sciences LLC). Markers such as call ringing until missed minutes (r=0.10) showed a correlation, and phone call count (r=−0.06) showed a negative correlation as well, as did emoticon count in text messages (r=−0.06), battery percentage (r=−0.7), and ambient audio level (r=0.10) [27].

Another study described a negative correlation of high MADRS scores with lower entropy of the usage time of communication apps and lower total count of communication app usage (r=−0.42), indicating a decreasing sociability as depression scores were increasing. A strong positive correlation (r=0.9) was noted between self-reported outcomes and total MADRS score [38]. User behavior was also analyzed in a different study that described a typical person using the mental health app 3 times a day for a total of 12 minutes over the course of 90 days. The most frequently used tools in this study app were the Thought Diary (MoodTools, Inquiry Health LLC) and the self-assessment tests [31]. Significant treatment and time interactions for the BDI, GAD-7, Recovery Assessment Scale (RAS), Rosenberg Self-Esteem Scale (RSES), and Sheehan Disability Scale (SDS) were described [42]. Another study reported that participation in Cognitive Behavior Immersion (CBI) sessions via app reduced symptoms of anxiety and depression in patients (r=−1.5) [20].

A large portion of findings within this category relies on behavior data and as such is closely related to the behavioral patterns category. However, apart from user behavior, social patterns such as social support are found to be influencing the severity of depressive symptoms and the prediction thereof. Web-based social support was found to predict the improvement of depressive symptoms [20] while social-rhythm-focused digital interventions were shown to benefit patients significantly. In fact, a more rapid rate of improvement according to the PHQ-8 was shown compared to nonpersonalized interventions [34]. Additionally, one study reported a greater decline in DASS-21 scores using personalized feedback derived from the experience sampling method (ESM) compared to using nonpersonalized feedback [43].

The majority of studies apply several categories with a range of parameters used. Commonly, at least 4 categories were used for monitoring, assessing, or tracking depression across digital tools of the studies included (see Table 4). The type of device used to deliver the support was not shown as a relevant factor in effectiveness, and high comparability between smartphone-based and nonsmartphone-based questionnaire scores was reported [42]. Nevertheless, it must be noted that categories are interdependent, and all studies used parameters existing across categories, focusing on multiple aspects.

Discussion

Principal Findings

By focusing on parameters rather than platforms or algorithms, this review responds directly to calls for greater comparability in digital phenotyping research. This scoping review identified five major categories of passively collected parameters commonly used in digital interventions for depression: (1) physical activity and location, (2) behavioral patterns, (3)

XSL•FO
RenderX

physiological data, (4) sleep, and (5) sociability and self-reported assessments. Most studies used a combination of smartphone-based and wearable sensors alongside self-reports. The strongest predictive value for depression outcomes was observed in studies that integrated multiple data sources, especially when combined with machine learning or subject-specific modeling approaches. Despite promising findings, the field remains methodologically fragmented and lacks standardization in both sensor use and clinical integration.

Five overall categories distinguished in this review contain relevant parameters for monitoring, assessing, and potentially predicting depression by combining self-assessment and digital phenotyping in digital interventions. The existing findings point that the categories of (1) physical activity and location, (2) behavioral patterns, (3) physiological data, (4) sleep, and (5) sociability and self-reported assessments appear across literature as interdependent categories, used across studies to monitor and predict depression by using various passively gathered data from parameters such as accelerometer sensor, location clusters, average distance from home, step count, phone usage, number of phone calls, heart rate, sleep time, and voice-related markers, among others. Moreover, additional parameters retrieved also include 11 frequently used clinical and self-assessment measurements such as PHQ-9, PHQ-2, CES-D, DASS-21, PHQ-8, MINI, MADRS, CSSRS, HDRS, BDI, and BDI-II (see Table 2).

This synthesis aligns with the World Health Organization guideline recommendations [21] for strengthening universal health coverage, where the role of digital health interventions is addressed particularly in line with the insufficient quality parameters that can assure optimal health care support [21]. While they emphasize the need for improved quality parameters in digital health, our findings indicate that such parameters are arguably not missing but are rather scattered across literature without comparative analysis. It might be argued that current work is one of the early efforts in the field to collect or summarize such parameters for digital health interventions overall in a more structured manner.

Previous research has often examined digital phenotyping and depression in isolation [8,19,46] or focused on either digital interventions [20,34,47-49] or productive models for symptom progression [13,33] without comprehensively linking all aspects. This review aims to contribute to bridging that gap by grouping relevant parameters into more standard categories.

Retrieved parameters in this review can be used as a comprehensive guideline in determining relevant aspects to be implemented into digital health interventions for depression to combine optimal sensors and self-assessment-based tools for monitoring and predicting depression outcomes. It can be argued that the interdependence of distinguished categories provides a more solid understanding for developing optimal digital health intervention in terms of monitoring and self-management support for depression. Combining these categories by implementing described parameters can support more beneficial outcomes for the end user of digital self-management tools for depression.

Synthesis of result sections of the included studies indicated that predicting depressive symptoms by combining clinical assessment and digital phenotyping is a promising approach for further improvement of digital interventions' reliability and efficiency for individuals living with depressive symptoms and major depressive disorder. In a study using a transfer learning approach, an improvement of 12% for one-day-ahead forecasting was noted [33]. Moreover, using the MADRS score as a baseline for depression severity, the highest correlation of prediction was reported with user self-reported assessment [33]. Among other such assessments were the PHQ-2 and a model consisting of the Bond and Lader lethargic-energetic visual analog scale and the Bond and Lader interested-bored visual analog scale plus the N-Back working memory task [38].

The scores obtained in a smartphone-based intervention correlated with the ratings of nonsmartphone-based HDRS. Concretely, the smartphone-based single-item self-rating of mood correlated negatively (r=−0.54) with the BDI sum of scores as well [42], indicating that the lower the perceived mood of the user was, the higher the depression score. Therefore, participant self-reported assessments can potentially be valuable in predicting depression severity when combined with digital phenotyping for higher accuracy and reliability. In line with that, the findings support prior work showing that self-reported assessments when combined with sensor data, can meaningfully predict depressive symptom severity [34,39,48].

Reflecting on the device delivering the intervention with a combined model of smartphone features, wearable features, and self-reported mood items is also noted in one of the retrieved studies' approaches [13].

However, this review also reveals substantial variability and methodological inconsistencies across studies. Among the retrieved studies there was a notable lack of standardization in contrast with the World Health Organization recommendations guidelines [21] addressing the quality parameters. Sample sizes, study and intervention duration, sampling frequencies, and types of intervention differed considerably. This was particularly evident when efforts for comparison during the charting procedures were made, mirroring concerns in literature about standardization and reproducibility [22,30,46]. Missing data was frequent, and most studies did not report a follow-up or provide justification for demographic admissions, therefore leading to the inclusion of studies assessing depression severity with validated measures with a notably limited amount of measurement scales used across studies. Taking into consideration the chronic nature of depression, it must be noted that short-term observations, less than 6 months, might also not be sufficient to successfully record meaningful changes in symptom trajectories or assess the sustained effectiveness of digital interventions for depression.

Across the included studies, it is important to distinguish between studies that primarily examined associations and those that validated predictive models based on digital data. Only a subset of studies (6/19) developed or evaluated models to predict depression status or symptom severity from digital signals, including work on behavioral and clinical modeling, cognitive task-based digital biomarkers, wearable-derived screening

XSL•FO
RenderX

models, and sensor-plus-EMA-based forecasting [13,27-29,33,39]. The remaining studies (13/19) were observational, feasibility, validation, or intervention trials. These studies primarily characterized associations between digital parameters and depressive symptoms, evaluated feasibility and engagement, validated self-report formats, or tested the clinical effects of digital interventions without undertaking formal predictive model validation [20,30-32,34-38,40-43]. This distinction underlines that much of the evidence base still addresses which digital features relate to depression, whereas a smaller subset of studies evaluates how well these features can support robust and generalizable prediction.

Considering the rapidly developing nature of digital interventions and digital phenotyping in the mental health field, multiple challenges in conducting this review were noted. While digital phenotyping has been increasingly used to monitor symptoms and to gather vast amounts of passive data, discovering studies that combine digital phenotyping with digitally delivered interventions proved challenging. Although many studies address digital tools or passive monitoring separately, only a small percentage integrate both, further showing the fragmentation within the field. Additionally, the narrow focus of this review on depression and depressive symptoms exclusively led to discarding similar studies focused on other mental illnesses, narrowing our evidence pool. Nevertheless, the assessment of studies in our review was rather challenging due to the enormous number of articles addressing either digital interventions and depression or digital phenotyping and depression but rarely combining all these areas of interest together. This observation further aligns with literature where similar limitations have been noted previously in an umbrella review of smartphone sensors for monitoring depression [10]. In this review, studies often focused on generalized categories rather than specific parameters, restricting reproducibility and cross-study comparison.

Strengths and Limitations

This review provides a structured synthesis of passively gathered parameters used in digital interventions for depression, supplementing existing reviews that focus exclusively on either digital phenotyping [8,46] or digital interventions [1,20,34,47,49]. By incorporating both, this review offers a clearer framework for understanding how sensor-derived data and self-reported assessments can be integrated in monitoring depression.

Although many high-quality systematic reviews and meta-analyses on digital phenotyping or digital interventions have been published, there is rather limited systematic literature synthesis known to the authors of this review that includes and combines all relevant terms such as “digital intervention,” “depression,” and “digital phenotyping.”

The review allows a structured overview of all relevant categories and parameters used in digital interventions that address depression with clinical measures and digital phenotyping methodology. An important strength of this review is the categorization of parameters. It allows a comparison across 5 interdependent categories, supporting efforts in the mental health field to determine and assure reliability in parameter use for future research and intervention designs, as well as clinical use, aligning also with World Health Organization recommendations [21].

Nonetheless, several limitations of this review need to be acknowledged. First, the review’s narrow scope, focusing exclusively on studies that used both validated depression assessments and digital phenotyping, may have excluded relevant work on related conditions or interventions that did not incorporate passive data collection. While this decision was made to enhance clarity and focus, it inevitably limited the breadth of the review, as noted in other recent scoping and mixed methods reviews as well [13,45]. Second, although passive data gathering enables the collection of large volumes of data over relatively short periods, a few studies reported that only a small proportion of the collected data was directly relevant to their hypotheses, raising concerns about potential publication bias or overinterpretation of limited findings.

Included studies mostly described individual interventions in detail, although many did not report the individual parameters measured but rather focused on larger terms, which we define in our study as categories. Even though general categories across studies were comparable, the lack of information on the exact data gathered through digital phenotyping leads to limited reproducibility of these studies as well as limited comparability with similar studies, thereby raising questions about the implications of such parameters for future studies. Missing data handling was an issue encountered in numerous studies, varying from a lack of explanation regarding the drop-out rates over user-related problems or confounders to basic demographic information. Similar methodological weaknesses have been described as well in prior meta-analyses of mental health data for depression [8,30,39].

Third, technical variability also limited comparison. Concretely, retrieved studies used different sensor types and different software, which could possibly influence data reliability. Technological difficulties were reported in some studies, though their influence on the reliability of data was rarely discussed. In line with that, few studies discussed the influence of signals being generated due to reasons outside of illness context as potential confounders (eg, the influence of weather conditions on the amount of time spent at home limiting both physical and social activity). Previous work has highlighted the need to better account for such nonclinical sources of signal variability [10,38].

Finally, this review confronts challenges comparable to those of many trials researching technology or digital devices, including limited transferability of results due to the nature of study populations. The users that participated in the included studies were for the most part younger people who are commonly more familiar and comfortable with the technologies used, apart from a few studies, and had only a few issues using the interventions as intended. However, this might not be the case with an older population, thus indicating that results may not necessarily be representative, an issue that has also been raised in other studies [31,35].

Although this review identifies promising directions for standardizing and optimizing digital interventions for depression, it also underlines the need for improved methodological

XSL•FO
RenderX

consistency, demographic inclusivity, and clearer reporting of sensor-derived parameters in the future.

Conclusion

This scoping review offers an innovative contribution by systematically identifying and synthesizing common digital parameters used across heterogeneous digital tools for monitoring and predicting depression. In contrast to existing reviews that predominantly focus on specific sensing modalities, predictive algorithms, or the effectiveness of digital interventions, this review adopts a parameter-centric perspective, enabling comparison across observational, predictive, and interventional study designs. By mapping convergent parameters irrespective of device type or analytic approach, the review provides a clearer understanding of which digital features are consistently associated with depressive symptoms and outcomes. This synthesis advances the field by supporting standardization efforts, guiding future predictive model development, and reducing fragmentation in digital mental health research. In real-world settings, these findings can inform the design of interoperable digital tools, facilitate integration into clinical monitoring workflows, and support personalized, data-driven depression management beyond controlled research environments.

Acknowledgments

We would like to sincerely thank the editor and anonymous reviewers for their constructive feedback and thoughtful suggestions which helped improve the clarity and quality of this manuscript. AI tools (ChatGPT, OpenAI) were used for editing/formatting assistance in the first revision round for refining the existing paragraphs and removing redundancy; authors critically reviewed and approved all content.

Funding

This research did not receive any specific grant from funding agencies in the public or commercial sectors.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

LB and UH conceptualized the study. LB, MP, and RA performed data extraction. LB and MP conducted the data analysis, while UH supervised the analysis process and contributed to clinical interpretation which then also supported the discussion. All authors contributed to the drafting of the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Reasons for exclusion of screened studies.
[DOCX File, 29 KB - mhealth_v14i1e70840_app1.docx]

Multimedia Appendix 2
Downs and Black Instrument assessment performed for each retrieved paper.
[DOCX File, 469 KB - mhealth_v14i1e70840_app2.docx]

Multimedia Appendix 3
Mixed Methods Appraisal Tool, version 2018. Assessment performed for each retrieved paper.
[DOCX File, 413 KB - mhealth_v14i1e70840_app3.docx]

Multimedia Appendix 4
Detailed quality assessment parameters of included studies.
[DOCX File, 22 KB - mhealth_v14i1e70840_app4.docx]

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 50 KB - mhealth_v14i1e70840_app5.docx]

References

XSL•FO
RenderX

1. Chib A, Lin SH. Theoretical advancements in mHealth: a systematic review of mobile apps. J Health Commun 2018;23(10-11):909-955. [doi: 10.1080/10810730.2018.1544676] [Medline: 30449261]
2. Torous J, Kiang MV, Lorme J, Onnela JP. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR Ment Health 2016 May 5;3(2):e16. [doi: 10.2196/mental.5165] [Medline: 27150677]
3. What kinds of sensors are embedded in smartphones? [Internet]. Samsung SDS. 2024. URL: https://www.samsungsds.com/en/insights/what-kinds-of-sensors-are-embedded-in-smartphones.html [accessed 2026-02-04]
4. Perera C, Zaslavsky A, Christen P, Salehi A, Georgakopoulos D. Capturing sensor data from mobile phones using global sensor network middleware. Presented at: 2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC 2012); Sep 9-12, 2012; Sydney, NSW p. 24-29 URL: https://ieeexplore.org/document/6362778 [doi: 10.1109/PIMRC.2012.6362778]
5. Pires IM, Garcia NM, Pombo N, Flórez-Revuelta F, Spinsante S. Approach for the development of a framework for the identification of activities of daily living using sensors in mobile devices. Sensors (Basel) 2018 Feb 21;18(2):640. [doi: 10.3390/s18020640] [Medline: 29466316]
6. Celik Y, Godfrey A. Bringing it all together: wearable data fusion. NPJ Digit Med 2023 Aug 17;6(1):149. [doi: 10.1038/s41746-023-00897-6] [Medline: 37591989]
7. Draghici AE, Taylor JA. The physiological basis and measurement of heart rate variability in humans. J Physiol Anthropol 2016 Sep 28;35(1):22. [doi: 10.1186/s40101-016-0113-7] [Medline: 27680542]
8. De Angel V, Lewis S, White K, et al. Digital health tools for the passive monitoring of depression: a systematic review of methods. NPJ Digit Med 2022 Jan 11;5(1):3. [doi: 10.1038/s41746-021-00548-8] [Medline: 35017634]
9. Teo AR, Nelson S, Strange W, et al. Social withdrawal in major depressive disorder: a case-control study of hikikomori in japan. J Affect Disord 2020 Sep 1;274:1142-1146. [doi: 10.1016/j.jad.2020.06.011] [Medline: 32663943]
10. Leimhofer J, Petrovic M, Dominik A, Heider D, Hegerl U. Cross-platform availability of smartphone sensors for depression indication systems: mixed-methods umbrella review. Interact J Med Res 2025 Aug 7;14:e69686. [doi: 10.2196/69686] [Medline: 40773748]
11. Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 2018 Jan;77:120-132. [doi: 10.1016/j.jbi.2017.12.008] [Medline: 29248628]
12. Mohr DC, Lyon AR, Lattie EG, Reddy M, Schueller SM. Accelerating digital mental health research from early design and creation to successful implementation and sustainment. J Med Internet Res 2017 May 10;19(5):e153. [doi: 10.2196/jmir.7725] [Medline: 28490417]
13. Moshe I, Terhorst Y, Opoku Asare K, et al. Predicting symptoms of depression and anxiety using smartphone and wearable data. Front Psychiatry 2021;12:625247. [doi: 10.3389/fpsyt.2021.625247] [Medline: 33584388]
14. Saeb S, Zhang M, Karr CJ, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015 Jul 15;17(7):e175. [doi: 10.2196/jmir.4273] [Medline: 26180009]
15. Rohani DA, Faurholt-Jepsen M, Kessing LV, Bardram JE. Correlations between objective behavioral features collected from mobile and wearable devices and depressive mood symptoms in patients with affective disorders: systematic review. JMIR Mhealth Uhealth 2018 Aug 13;6(8):e165. [doi: 10.2196/mhealth.9691] [Medline: 30104184]
16. Choi A, Ooi A, Lottridge D. Digital phenotyping for stress, anxiety, and mild depression: systematic literature review. JMIR Mhealth Uhealth 2024 May 23;12:e40689. [doi: 10.2196/40689] [Medline: 38780995]
17. Shen S, Qi W, Zeng J, et al. Passive sensing for mental health monitoring using machine learning with wearables and smartphones: scoping review. J Med Internet Res 2025 Aug 14;27:e77066. [doi: 10.2196/77066] [Medline: 40811794]
18. Leaning IE, Ikani N, Savage HS, et al. From smartphone data to clinically relevant predictions: a systematic review of digital phenotyping methods in depression. medRxiv. Health Informatics. Preprint posted online on 2023. [doi: 10.1101/2023.10.04.23296546]
19. Ettore E, Müller P, Hinze J, et al. Digital phenotyping for differential diagnosis of major depressive episode: narrative review. JMIR Ment Health 2023 Jan 23;10:e37225. [doi: 10.2196/37225] [Medline: 36689265]
20. Ezawa ID, Hollon SD, Robinson N. Examining predictors of depression and anxiety symptom change in cognitive behavioral immersion: observational study. JMIR Ment Health 2023 Jul 14;10:e42377. [doi: 10.2196/42377] [Medline: 37450322]
21. WHO guideline recommendations on digital interventions for health system strengthening. World Health Organization. 2019. URL: http://www.ncbi.nlm.nih.gov/books/NBK541902 [accessed 2026-02-04]
22. Roberts LW, Chan S, Torous J. New tests, new tools: mobile and connected technologies in advancing psychiatric diagnosis. NPJ Digit Med 2018;1:20176. [doi: 10.1038/s41746-017-0006-0] [Medline: 31304350]
23. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
24. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
25. Torous J, Linardon J, Goldberg SB, et al. The evolving field of digital mental health: current evidence and implementation issues for smartphone apps, generative artificial intelligence, and virtual reality. World Psychiatry 2025 Jun;24(2):156-174. [doi: 10.1002/wps.21299] [Medline: 40371757]

XSL•FO
RenderX

26. Pearce M, Garcia L, Abbas A, et al. Association between physical activity and risk of depression: a systematic review and meta-analysis. JAMA Psychiatry 2022 Jun 1;79(6):550-559. [doi: 10.1001/jamapsychiatry.2022.0609] [Medline: 35416941]
27. Nickels S, Edwards MD, Poole SF, et al. Toward a mobile platform for real-world digital measurement of depression: user-centered design, data quality, and behavioral and clinical modeling. JMIR Ment Health 2021 Aug 10;8(8):e27589. [doi: 10.2196/27589] [Medline: 34383685]
28. Mandryk RL, Birk MV, Vedress S, et al. Remote assessment of depression using digital biomarkers from cognitive tasks. Front Psychol 2021;12:767507. [doi: 10.3389/fpsyg.2021.767507] [Medline: 34975656]
29. Rykov Y, Thach TQ, Bojic I, Christopoulos G, Car J. Digital biomarkers for depression screening with wearable devices: cross-sectional study with machine learning modeling. JMIR Mhealth Uhealth 2021 Oct 25;9(10):e24872. [doi: 10.2196/24872] [Medline: 34694233]
30. Currey D, Torous J. Increasing the value of digital phenotyping through reducing missingness: a retrospective review and analysis of prior studies. BMJ Ment Health 2023 Feb;26(1):e300718. [doi: 10.1136/bmjment-2023-300718] [Medline: 37197799]
31. Su L, Anderson PL. User behavior of a publicly available, free-to-use, self-guided mHealth app for depression: observational study in a global sample. JMIR Form Res 2022 Oct 25;6(10):e35538. [doi: 10.2196/35538] [Medline: 36282559]
32. Braund TA, O'Dea B, Bal D, et al. Associations between smartphone keystroke metadata and mental health symptoms in adolescents: findings from the future proofing study. JMIR Ment Health 2023 May 15;10:e44986. [doi: 10.2196/44986] [Medline: 37184904]
33. Kathan A, Harrer M, Küster L, et al. Personalised depression forecasting using mobile sensor data and ecological momentary assessment. Front Digit Health 2022;4:964582. [doi: 10.3389/fdgth.2022.964582] [Medline: 36465087]
34. Frank E, Wallace ML, Matthews MJ, et al. Personalized digital intervention for depression based on social rhythm principles adds significantly to outpatient treatment. Front Digit Health 2022;4:870522. [doi: 10.3389/fdgth.2022.870522] [Medline: 36120713]
35. Zhang Y, Pratap A, Folarin AA, et al. Long-term participant retention and engagement patterns in an app and wearable-based multinational remote digital depression study. NPJ Digit Med 2023 Feb 17;6(1):25. [doi: 10.1038/s41746-023-00749-3] [Medline: 36806317]
36. Langholm C, Kowatsch T, Bucci S, Cipriani A, Torous J. Exploring the potential of Apple SensorKit and digital phenotyping data as new digital biomarkers for mental health research. Digit Biomark 2023;7(1):104-114. [doi: 10.1159/000530698] [Medline: 37901364]
37. Abbas A, Sauder C, Yadav V, et al. Remote digital measurement of facial and vocal markers of major depressive disorder severity and treatment response: a pilot study. Front Digit Health 2021;3:610006. [doi: 10.3389/fdgth.2021.610006] [Medline: 34713091]
38. Sverdlov O, Curcic J, Hannesdottir K, et al. A study of novel exploratory tools, digital technologies, and central nervous system biomarkers to characterize unipolar depression. Front Psychiatry 2021;12:640741. [doi: 10.3389/fpsyt.2021.640741] [Medline: 34025472]
39. Sun S, Folarin AA, Zhang Y, et al. Challenges in using mHealth data from smartphones and wearable devices to predict depression symptom severity: retrospective analysis. J Med Internet Res 2023 Aug 14;25:e45233. [doi: 10.2196/45233] [Medline: 37578823]
40. Goltermann J, Emden D, Leehr EJ, et al. Smartphone-based self-reports of depressive symptoms using the Remote Monitoring Application in Psychiatry (ReMAP): interformat validation study. JMIR Ment Health 2021 Jan 12;8(1):e24333. [doi: 10.2196/24333] [Medline: 33433392]
41. Tsirmpas C, Andrikopoulos D, Fatouros P, et al. Feasibility, engagement, and preliminary clinical outcomes of a digital biodata-driven intervention for anxiety and depression. Front Digit Health 2022;4:868970. [doi: 10.3389/fdgth.2022.868970] [Medline: 35958737]
42. Ben-Zeev D, Chander A, Tauscher J, et al. A smartphone intervention for people with serious mental illness: fully remote randomized controlled trial of CORE. J Med Internet Res 2021 Nov 12;23(11):e29201. [doi: 10.2196/29201] [Medline: 34766913]
43. Suen YN, Yau JY, Wong PS, et al. Effect of brief, personalized feedback derived from momentary data on the mental health of women with risk of common mental disorders in Hong Kong: a randomized clinical trial. Psychiatry Res 2022 Nov;317:114880. [doi: 10.1016/j.psychres.2022.114880] [Medline: 37732870]
44. Downs and black checklist (DBC) [internet]. Reach Out and Read. 2024. URL: https://reachoutandread.org/metric/dbc/ [accessed 2026-02-04]
45. Hong QN, Fàbregues S, Bartlett G, et al. The mixed methods appraisal tool (MMAT) version 2018 for information professionals and researchers. EFI 2018;34(4):285-291. [doi: 10.3233/EFI-180221]
46. Martin-Key NA, Spadaro B, Funnell E, et al. The current state and validity of digital assessment tools for psychiatry: systematic review. JMIR Ment Health 2022 Mar 30;9(3):e32824. [doi: 10.2196/32824] [Medline: 35353053]
47. Kang E, Kim S, Rhee YE, Lee J, Yun YH. Self-management strategies and comorbidities in chronic disease patients: associations with quality of life and depression. Psychol Health Med 2021 Sep;26(8):1031-1043. [doi: 10.1080/13548506.2020.1838585] [Medline: 33095059]

XSL•FO
RenderX

48. Currey D, Torous J. Digital phenotyping data to predict symptom improvement and mental health app personalization in college students: prospective validation of a predictive model. J Med Internet Res 2023 Feb 9;25:e39258. [doi: 10.2196/39258] [Medline: 36757759]
49. Laban G, Morrison V, Kappas A, S Cross E. Coping with emotional distress via self-disclosure to robots: an intervention with caregivers. Int J Soc Robot 2025;17(9):1837-1870. [doi: 10.1007/s12369-024-01207-0] [Medline: 41020173]

Abbreviations

BDI: Beck Depression Inventory
BDI-II: Beck Depression Inventory, Second Edition
CES-D: Center for Epidemiological Studies Depression Scale
CSSRS: Columbia Suicide Severity Rating Scale
DASS-21: Depression Anxiety Stress Scale Short Form 21
HDRS: Hamilton Depression Rating Scale
ICTRP: International Clinical Trials Registry Platform
MADRS: Montgomery-Åsberg Depression Rating Scale
MINI: Mini-International Neuropsychiatric Interview
MMAT: Mixed Methods Appraisal Tool
PHQ-2: Patient Health Questionnaire-2
PHQ-8: Patient Health Questionnaire-8
PHQ-9: Patient Health Questionnaire-9
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
PROSPERO: International Prospective Register of Systematic Reviews
RAS: Recovery Assessment Scale
RSES: Rosenberg Self-Esteem Scale
SDS: Sheehan Disability Scale
WHO: World Health Organization

Edited by S Brini; submitted 04.Jan.2025; peer-reviewed by A Parks, A Sameh; accepted 21.Dec.2025; published 02.Mar.2026.

Please cite as:
Busshart L, Petrovic M, Amin R, Hegerl U
Distinguishing Common Digital Phenotyping and Self-Report Parameters for Monitoring and Predicting Depression: Scoping Review
JMIR Mhealth Uhealth 2026;14:e70840
URL: https://mhealth.jmir.org/2026/1/e70840
doi:10.2196/70840

© Lisa Busshart, Milica Petrovic, Rebeka Amin, Ulrich Hegerl. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 2.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluation of Smartphone Camera Positioning on Artificial Intelligence Pose Estimation Accuracy for Exercise Detection: Observational Study

Eduarda Oliosi[1,2,3], BS, MS; Soraia Ferreira[1,4,5], MS, PhD; Ana Paula Giordano[1,6], MS, PhD; Guilherme Viveiros[7], MS; José Parraca[4,5], PhD; Paulo Pereira[4,5], MS; Federico Guede-Fernández[1,2], PhD; Salomé Azevedo[1,7,8,9], PhD

[1]Value for Health CoLAB, 15 Fontes Pereira de Melo Ave, 2nd Fl, Right, Lisbon, Portugal

[2]Laboratory for Instrumentation, Biomedical Engineering and Radiation Physics (LIBPhys-UNL), Physics Department, Nova School of Science and Technology, Nova University of Lisbon, Caparica, Portugal

[3]Research Centre for Physical Activity, Health and Leisure (CIAFEL), Faculty of Sport, University of Porto, Porto, Portugal

[4]Departamento de Desporto e Saúde, Escola de Saúde e Desenvolvimento Humano, Universidade de Évora, Évora, Portugal

[5]Comprehensive Health Research Centre (CHRC), Escola de Saúde e Desenvolvimento Humano, Departamento de Ciências Médicas, Universidade de Évora, Évora, Portugal

[6]Research Unit in Business and Economics, Católica Lisbon School of Business and Economics (CUBE), Catholic University of Portugal, Lisbon, Portugal

[7]Department of Machine Learning, Dotmoovs, Braga, Portugal

[8]Department of Engineering and Management, Instituto Superior Técnico (CEG-IST), University of Lisbon, Lisbon, Portugal

[9]Comprehensive Health Research Center (CHRC), Nova Medical School, Nova University of Lisbon, Lisbon, Portugal

Corresponding Author:
Eduarda Oliosi, BS, MS
Value for Health CoLAB, 15 Fontes Pereira de Melo Ave, 2nd Fl, Right, Lisbon, Portugal

Abstract

Background: Artificial intelligence (AI)–driven pose estimation (PE) offers a scalable and cost-effective solution to track exercises in mobile health apps. However, occlusion, influenced by camera angle and distance, can reduce detection accuracy and repetition counting precision. The influence of smartphone positioning on these performance metrics remains underexplored in controlled studies.

Objective: The study aimed to examine how smartphone camera angle (front, side, and diagonal) and distance (90 cm, 180 cm, 200 cm, and 360 cm) affect detection performance and repetition counting accuracy during push-ups and squats using AI-based PE.

Methods: In this cross-sectional, within-subject study, 44 healthy university students (9 [20.5%] female participants; mean age 20.3 y, SD 0.4 y; mean BMI 23.2, SD 0.6 kg/m²) were assigned to perform either squats or push-ups. Each participant completed their assigned exercise across 12 predefined smartphone camera configurations, yielding approximately 264 squat trials (n=22) and 264 push-up trials (n=22). Each trial consisted of an average of 5 repetitions, totaling approximately 1320 repetitions per exercise. PE performance was assessed using binary classification accuracy, detection rate, and mean absolute error (MAE) for repetition counting. Generalized linear mixed-effects models evaluated classification odds, linear mixed-effects models analyzed MAE, and Tukey-adjusted post hoc tests followed significant effects.

Results: The mean detection rate was 61.1% (SD 48.8%) for push-ups and 61.5% (SD 48.7%) for squats, with MAEs of 1.08 (SD 1.78) and 1.11 (SD 1.82) repetitions, respectively. Push-ups were most accurately detected from diagonal views at 90 to 180 cm (up to 85.7% detection; MAE=0.28) and least accurately from the front at 360 cm (20%; MAE=2.70). Squats performed best from a diagonal view at 200 cm (95.5%; MAE=0.05) and worst from the side at 90 cm (0%; MAE=5). Generalized linear mixed models showed that for push-ups, the front 90 cm and diagonal 360 cm views significantly reduced classification odds compared to the side 90 cm view (P=.03 and P=.04, respectively), whereas for squats, diagonal and front views significantly outperformed side views across all distances (P<.001). Post hoc tests confirmed that for push-ups, diagonal close or mid-range views had significantly lower MAEs than far front views, and for squats, diagonal and front views at 180 to 200 cm achieved the highest accuracy and lowest MAEs (P<.05).

Conclusions: AI-based PE effectiveness for exercise tracking is significantly affected by smartphone positioning. Diagonal and frontal views at mid-range distances (180 - 200 cm) provided the highest detection accuracy and counting precision. These findings offer actionable guidance for developers, clinicians, coaches, and users optimizing mobile health exercise monitoring.

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2026;14:e82412) doi:10.2196/82412

KEYWORDS

computer vision; digital health; human pose estimation; human activity recognition; mobile health; mhealth; mobile apps; physical activity

Introduction

Physical inactivity is a major global public health concern, contributing to higher rates of chronic diseases, reduced quality of life, and increased health care costs [1-3]. While the health benefits of regular physical activity are well documented, ranging from improved cardiovascular fitness and muscle strength to enhanced mental well-being, participation levels remain insufficient worldwide [4]. Current estimates indicate that over 31% of adults and 81% of adolescents fail to meet the World Health Organization's recommended physical activity guidelines [5]. The consequences of physical inactivity are particularly severe for populations managing chronic or neurological conditions, for whom movement-based interventions can serve as both preventive and therapeutic measures [3,6,7].

In response to the demand for scalable and accessible physical activity tools, mobile health (mHealth) apps have emerged as a cost-effective means of promoting exercise participation [3]. A notable innovation in this field is pose estimation (PE), or markerless motion capture, which uses computer vision to detect joint positions from standard video footage without the need for specialized equipment [8]. Open-source PE models, such as OpenPose (Carnegie Mellon University), BlazePose (Google), and MoveNet (Google), facilitate real-time PE via smartphones or webcams. This enables the provision of remote, personalized feedback [9-13]. In biomedical contexts, PE can identify movement impairment patterns and support rehabilitation or neurological diagnoses [8]. It is also widely applied in fitness, telerehabilitation, occupational health, and sports to track joint angles, velocity, form, and repetitions [14-18]. Validation studies have demonstrated strong agreement with the gold standard motion capture [19-21], and recent studies integrating deep learning have reported repetition-counting accuracies above 90% [22-25].

Despite recent advances, most PE systems have only been evaluated in controlled laboratory environments. However, real-world mHealth apps occur in uncontrolled settings where factors such as handheld devices held at different distances and angles, and partial obstructions (eg, from pets, furniture, or body position) can significantly degrade performance [10,12,26]. These challenges are particularly pronounced for dynamic, whole-body movements such as squats and push-ups, where tracking accuracy can vary depending on body position and elevation. Although a few commercial solutions have begun to experiment with camera-based self-assessment (eg, Halo Movement [Amazon] [27] and Kaia Health GmbH [28]), there is still a lack of empirical studies on how smartphone positioning affects the reliability of PE in practical settings [10,14,27]. Improving accuracy in these contexts will likely require advances not only in 2D keypoint detection but also in our understanding of optimal camera placement and how to transform 2D keypoint data into accurate spatial representations.

This study systematically evaluates the impact of smartphone camera angle and distance on the performance of artificial intelligence (AI)–based PE when counting 2 foundational bodyweight exercises: squats and push-ups. These exercises are commonly used for fitness and rehabilitation tracking. They involve multijoint movements and present common occlusion challenges. Repetition counts are obtained from 2D PE across 12 camera configurations (3 angles × 4 distances) and benchmarked against expert-labeled ground truth. The aim is to examine how smartphone recording position, specifically camera angle and distance, affects detection performance and repetition counting accuracy, and to inform best practices for using mHealth systems in real-world environments. This work is part of the Blockchain.PT initiative (Project No. 51), which supports sustainable blockchain-based digital innovation in the health and sports sectors.

Methods

Study Design

This cross-sectional, within-subject study was designed to assess the influence of smartphone camera configurations, specifically angle and distance, on the detection of PE during squats and push-ups. These exercises were selected due to their complex multijoint biomechanics and typical self-occlusions (eg, torso obscuring limbs in push-ups).

Setting

All data were collected in a controlled laboratory environment at the School of Health and Human Development, University of Évora, Portugal, in May 2025. Environmental conditions, including lighting and temperature, were standardized across trials. Smartphone placement was controlled at predefined angles and distances to ensure reproducible video capture. The laboratory provided enough space for participants to safely perform push-ups and squats.

Participants

Participants were recruited from undergraduate courses at the University of Évora in Portugal through a targeted study announcement that specified the eligibility criteria, study procedures, and recommended attire for physical exercise during classes. Eligible volunteers were identified through convenience sampling with eligibility screening.

A total of 44 healthy university students participated and self-reported their status. Inclusion criteria were: (1) aged 18 years or over; (2) able to safely perform squats and push-ups; (3) not having used sedative or balance-impairing medications within the last 24 hours; and (4) absence of neurological or musculoskeletal conditions that could impair movement.

XSL•FO
RenderX

Ethical Considerations

The study protocol was reviewed and approved by the Ethics Committee of the University of Évora (GD/27378/2024), in accordance with institutional and international standards for human participant research. All procedures complied with the Declaration of Helsinki and the General Data Protection Regulation. Written informed consent was obtained from all participants prior to enrollment. Participants were informed about the study objectives, procedures, and potential risks and were explicitly advised of their right to withdraw at any time without penalty. To ensure privacy and confidentiality, all collected data were anonymized and coded to prevent participant identification. Data were securely stored on password-protected institutional servers accessible only to the research team. No monetary or material compensation was provided for participation in this study.

Exercise Protocol and Dataset

Participants were first briefed on the study protocol and completed a demographic questionnaire to record baseline characteristics. For the squat task, participants stood with feet shoulder-width apart and arms extended forward, performing repetitions by lowering their hips until their thighs were parallel to the floor. Male participants performed push-ups in a standard plank position, while female participants were allowed a modified knee-supported variation. All exercise procedures adhered to the 11th edition of the *American College of Sports Medicine* guidelines (2021) [29], which recommend these variations to accommodate differences in strength and fitness levels. To reduce fatigue-related performance decline and ensure high-quality movement data, participants were assigned to perform either squats or push-ups based on personal preference. Of the 44 participants, 22 (16 male participants and 6 female participants) completed the squat protocol, and 22 (19 male participants and 3 female participants) completed the push-up protocol. Each participant performed their assigned exercise across all 12 predefined camera configurations, yielding 264 squat trials (22×12) and 264 push-up trials (22×12). Each trial consisted of a continuous movement sequence, averaging approximately 5 repetitions per video [16], totaling approximately 1320 repetitions per exercise type (264 trials × ~5 reps).

Experimental Setup

To capture a representative sample of commonly used consumer devices, exercise sessions were recorded using three smartphone models, iPhone 11 (Apple Inc.), iPhone 13 (Apple Inc.), and Samsung Galaxy A52 (Samsung Electronics), each capturing video at 1080 p resolution and 30 frames per second. Smartphones were positioned horizontally on the floor at 3 fixed angles relative to the participant's body: frontal (0°), diagonal (45°), and lateral (90°). Recordings were conducted at 4 distances, including 180 cm based on prior literature [30], and 3 additional distances (90 cm, 200 cm, and 360 cm) defined by the research team to test a wider range of realistic recording conditions. This configuration resulted in 12 distinct camera setups (3 angles × 4 distances). All smartphone positions were marked and standardized to ensure consistency across trials.

This experimental setup is contextualized within existing research by Table 1, which summarizes previous studies on push-up and squat detection, detailing camera perspectives, distances, PE methods, and reported performance metrics. The comparison shows that most prior studies rely on limited viewpoints and fixed distances, restricting generalizability. In contrast, our multiangle, multidistance dataset addresses these constraints, providing a more diverse and representative resource for evaluating PE performance.

XSL•FO
RenderX

Table . Summary of state-of-the-art push-up and squat pose estimation datasets and methods.

Reference	Exercise	Camera perspectives	Camera distances	PE[n] method	Instrument	Key evaluation metrics	Dataset
Park et al [31]	Push-up	Frontal, side	Full body visible, N/R[b] value	OpenPose	2 cameras	ACC[c]=90%	Custom dataset of n=12
Youssef et al [32]	Squat	Frontal, side	N/R	BlazePose	Mobile devices + inertial sensors	ACC=94%	EJUST-SQUAT-21[o], single individual, MM-Fit[e] datasets
Hande et al [33]	Squat and others	Frontal, side	Full body visible, N/R value	OpenPose, MobileNet (Google), InceptionV3 (Google)	Single camera	ACC =~98% (MobileNet),~96% (InceptionV3)	Penn Action Dataset
Chae et al [34]	Squat	Frontal, diagonal	250 cm Kinect (Microsoft), 380 cm webcam	OpenPose, Temporal Conv1D[f], BiLSTM[g]	Kinect + webcam	ACC=85%	Custom dataset of n=52
Chariar et al [35]	Squat	Frontal	Distance N/R; 120 cm height	MediaPipe (Google), Bi-GRU[h]	2 depth cameras	ACC=94%	Custom dataset of n~50
Zhang et al [36]	Push-up	Frontal, side	N/R	MoveNet; angle-heuristic; Optical flow	N/R	Average F_1-score: angle-heuristic=0.85 (side > front); pose classification=0.94 (side > front); optical flow=0.79 (front > side)	Kaggle "Push-up Exercise" dataset
Japhne et al [24]	Push-up, squat, and others	N/R, full body visible	200 cm	OpenPose, LSTM[i]	Mobile devices	Push-up: ACC=~99%; Squat: ACC=~99%	Custom dataset of n=3
Mercadal-Baudart et al [21]	Squat and others	Frontal; multiangle validation	~300 cm radius; 150 cm height	Detectron2 (Meta Platforms, Inc), Strided Transformer	Mobile devices	RMSE[j] of joint angles versus VICON[k] (Vicon Motion Systems): <10° for most joints (shin, knee, hip, trunk, and spine), <15° for shoulder and ASIS[l] (notably front squats)	Custom dataset of n=8 - 12

[a]PE: pose estimation.

[b]N/R: not reported.

[c]ACC: accuracy.

[d]EJUST-SQUAT-21: Egypt-Japan University squat dataset 2021.

[e]MM-Fit: multimodal fitness dataset.

[f]Conv1D: one-dimensional convolutional neural network.

[g]BiLSTM: bidirectional long short-term memory.

[h]Bi-GRU: bidirectional gated recurrent unit.

[i]LSTM: long short-term memory.

[j]RMSE: root mean square error.

[k]VICON: Vicon motion capture system.

[l]ASIS: anterior superior iliac spine.

XSL•FO
RenderX

Pose Estimation and Processing

Video data were processed using a multistage PE and repetition detection pipeline designed for real-world deployment on the Dotmoovs mobile platform [37]. The pipeline was optimized for low-latency, on-device inference while maintaining sufficient accuracy for exercise recognition.

Model Selection

A lightweight 2D PE model was prioritized to ensure real-time performance on consumer smartphones and without requiring external sensors. Among candidate architectures (eg, PoseNet [Google], BlazePose, and MoveNet), MoveNet was selected due to its computational efficiency, suitability for edge inference, and enhanced generalization to fitness-related movements. MoveNet was trained on the COCO (Common Objects in Context) dataset (Microsoft) and Google's internal Active dataset, which includes annotated yoga, fitness, and dance poses exhibiting substantial motion variability [38-40]. For cloud-based processing, EvoPose2D (Huawei Technologies Co, Ltd) and ViTPose (Microsoft Research Asia) were implemented to achieve higher precision at the cost of increased computational load. This hybrid configuration balances latency, accuracy, and hardware constraints, supporting both on-device responsiveness and scalable cloud inference [39]. The models were trained on approximately 1 million samples, reserving 10,000 samples each for validation and testing, ensuring robustness and generalizability across diverse movement types.

Training Datasets and Domain Adaptation

To improve robustness for fitness-specific postures, the training pipeline incorporated multiple publicly available datasets: SMART (Sports Motion and Recognition Tasks) [41], LSP (Leeds Sports Pose) Extended [42], Penn Action [43], and MPII Human Pose [44]. Despite these datasets, complex movements, particularly floor-based exercises (eg, push-ups), remained challenging for accurate PE. This motivated the creation of DotPose, a custom internal dataset designed to complement existing datasets and improve the detection of occluded limbs and challenging postures. The combination of DotPose with public datasets further mitigates potential biases related to body types, cultural differences, exercise contexts, and environmental scenarios, thereby enhancing generalizability across diverse users and real-world conditions.

Pipeline Implementation

The PE outputs (17 keypoints per frame) were processed through a proprietary deep learning module performing two primary tasks:

1. State classification: A lightweight neural network classifies each frame into discrete movement states (eg, squat-down, squat-up, or other).
2. Repetition counting: A Markov-chain-based algorithm tracks temporal transitions between states, triggering a repetition event when a predefined sequence (eg, up-down-up) is detected.

This architecture ensures temporal stability and robustness against intermittent keypoint noise. The pipeline was benchmarked for real-time inference, with MoveNet demonstrating subsecond latency, suitable for interactive mobile feedback. Finally, all pose outputs were exported in comma-separated values format to enable statistical analysis of accuracy and classification metrics across varying camera angles and distances.

Manual Annotation

All video samples were independently annotated by 2 trained raters (an exercise physiologist and a physical therapist), who followed a predefined schema based on standardized exercise movement criteria. Each annotator labeled the exercise type and manually counted the number of valid repetitions per trial. No discrepancies were observed between raters. These annotations were then used to evaluate the accuracy of detection and repetition counting.

Statistical Analysis

All analyses were conducted in RStudio (version 4.5.0; R Foundation for Statistical Computing) using established packages including lme4, lmerTest, emmeans, and performance. Performance metrics were reported in alignment with prior benchmark studies in human activity recognition and AI-based motion analysis [26,45]. The primary outcome was a binary accuracy indicator, coded as 1 when the system's predicted repetition count exactly matched the ground truth, and as 0 otherwise. This was modeled using generalized linear mixed-effects models (GLMMs) with a logit link. The secondary outcome was the mean absolute error (MAE), defined as the absolute difference between predicted and ground-truth repetition counts per video, modeled using linear mixed effects models (LMMs) with a Gaussian distribution. For descriptive reporting purposes, we also calculated a detection rate for each video, defined as (prediction or ground truth) x100, representing the percentage of repetitions counted correctly. Both models included camera angle and distance as fixed effects, with participant ID modeled as a random intercept to account for intraindividual variability and repeated measures. When the fixed effects were statistically significant ($P<.05$), Tukey-adjusted post hoc contrasts were conducted to compare condition pairs. Odds ratios (ORs) were computed by exponentiating the log odds from the GLMM to facilitate interpretation.

Results

Participant Characteristics

Our dataset comprised 44 healthy university students from the University of Évora, Portugal, including 9 (20.5%) females and 35 (79.5%) males. The participants had a mean age of 20.31 (SD 0.40) years, a mean height of 1.74 (SD 0.09) meters, and a mean body mass of 70 kg (SD 14.72), resulting in a mean BMI of 23.16 kg/m² (SD 0.61).

Based on previous studies assessing AI-based PE for push-up and squat detection [24], we assumed a medium-to-large expected effect size (Cohen d=0.65) for power estimation. A post hoc analysis conducted in G*Power 3.1 (Heinrich Heine University Düsseldorf) indicated that the current sample size (n=44) provides approximately 0.74 statistical power to detect this effect at a 2-tailed significance level of α=.05, suggesting

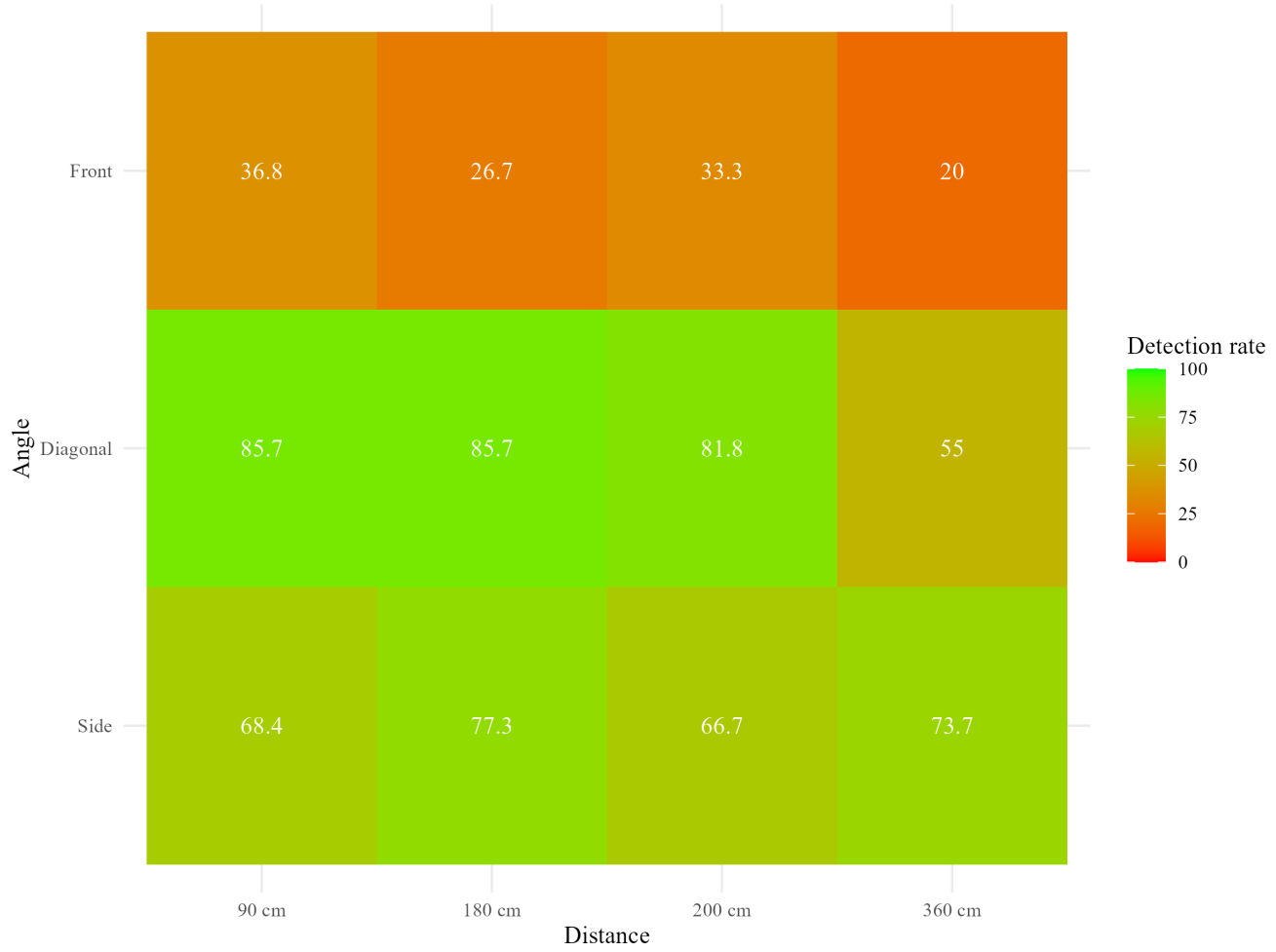
XSL•FO
RenderX

adequate sensitivity for detecting meaningful differences in PE performance across camera conditions.

Push-Up Performance

Table 2 shows detection rates and MAEs for push-ups across camera angles and distances. Overall, the mean detection rate was 61.1% (SD 48.8%), with an average MAE of 1.08 (SD 1.78) repetitions. Detection rates peaked for diagonal views at 90 cm and 180 cm (both 85.7%, SD 35.9%), as illustrated in Figure 1. The corresponding MAEs at these distances were low, at 0.29 (SD 0.90) repetitions. In contrast, the front view at 360 cm showed the lowest detection rate of 20% (SD 41.0%) and the highest MAE of 2.70 repetitions (SD 2.05).

Figure 1. Detection rate of push-ups across distances and angles.



XSL•FO
RenderX

Table . Detection rate and MAE[a] for push-ups across camera angles and distances.

Camera angle	Distance (cm)	Detection rate (%), mean (SD)	MAE, mean (SD)
Side	90	68.4 (47.8)	0.842 (1.537)
Side	180	77.3 (42.9)	0.409 (0.796)
Side	200	66.7 (48.3)	1.095 (1.972)
Side	360	73.7 (45.2)	0.789 (1.653)
Diagonal	90	85.7 (35.9)	0.286 (0.902)
Diagonal	180	85.7 (35.9)	0.286 (0.902)
Diagonal	200	81.8 (39.5)	0.227 (0.528)
Diagonal	360	55.0 (51)	1.500 (2.482)
Front	90	36.8 (49.6)	1.684 (1.974)
Front	180	26.7 (45.8)	2.067 (2.017)
Front	200	33.3 (48.8)	1.733 (1.981)
Front	360	20 (41)	2.700 (2.055)

[a]MAE: mean absolute error.

The GLMM (Table S1 in Multimedia Appendix 1) revealed a significant main effect of camera angle on accuracy. Specifically, front views had significantly lower odds of detection (OR=0.17, *P*=.03). However, no significant effects of distance alone were found. Conversely, a statistically significant interaction was found between the diagonal view and the 360 cm distance, which reduced the ability to correctly identify push-ups (*P*=.04).

Post hoc pairwise contrasts were performed to examine differences between the specific camera angles and distances (Table S2 in Multimedia Appendix 1). These contrasts confirmed that diagonal camera views at 90 cm significantly outperformed frontal views at all distances, including 90 cm (estimate=3.41, *P*=.019), 180 cm (estimate=3.67, *P*=.02), 200 cm (estimate=3.52, *P*=.02), and 360 cm (estimate=4.40, *P*<.001). There were also notable differences in the multiple mid-range diagonal and frontal configurations: the diagonal 180 cm view outperformed the frontal 180 cm view (estimate=3.64, *P*=.02), 200 cm (estimate=3.50, *P*=.03), as well as 360 cm (estimate=4.35, *P*=.002).

Likewise, the diagonal 200 cm view surpassed the frontal 360 cm view (estimate=3.96, *P*=.003). Additionally, the side views at 180 cm and 360 cm were significantly better than the frontal 360 cm view (estimate=3.586, *P*=.008, and estimate=3.174, *P*=.03, respectively). These results demonstrate that diagonal and mid-range frontal camera placements substantially increase push-up detection accuracy compared to frontal or side views at extreme distances.

Analyses of push-up repetition MAE using LMM were conducted to evaluate the error comparisons in the respective configurations. Although the main effects of angle and distance were not individually significant beyond the intercept (Table S3 in Multimedia Appendix 1), Tukey's post hoc tests (Table S4 in Multimedia Appendix 1) revealed meaningful differences between specific camera setups. Diagonal angles at mid-range distances consistently reduced counting errors compared with frontal and side angles at extreme distances. Diagonal views at 90 cm and 180 cm, for example, significantly outperformed frontal views at 180 cm and 360 cm (estimates=−1.686 to −2.395, *P*=.04). Diagonal views at 200 cm were also superior to frontal views at 360 cm (estimate=−2.428, *P*<.001). Side views at extreme distances exhibited higher MAE than frontal views at 360 cm (estimate=−1.800, *P*<.01). Overall, these results suggest that diagonal positioning at mid-range distances (approximately 180R - 200 cm) minimizes counting errors in push-ups.

Squat Performance

Table 3 shows detection rates and MAEs for squats across camera angles and distances. Overall, the mean detection rate was 61.5% (SD 48.7%), with an average MAE of 1.11 (SD 1.82) repetitions. Detection rates peaked for the diagonal view at 200 cm (95.5%, SD 21.3%), as illustrated in Figure 2. The corresponding MAE at this distance was minimal, at 0.05 (SD 0.21) repetitions. In contrast, the side view at 90 cm showed the lowest detection rate of 0% (SD 0%) as well as the highest MAE of 5 (SD 0) repetitions.

XSL•FO
RenderX

Figure 2. Detection rate of squats across distances and angles.

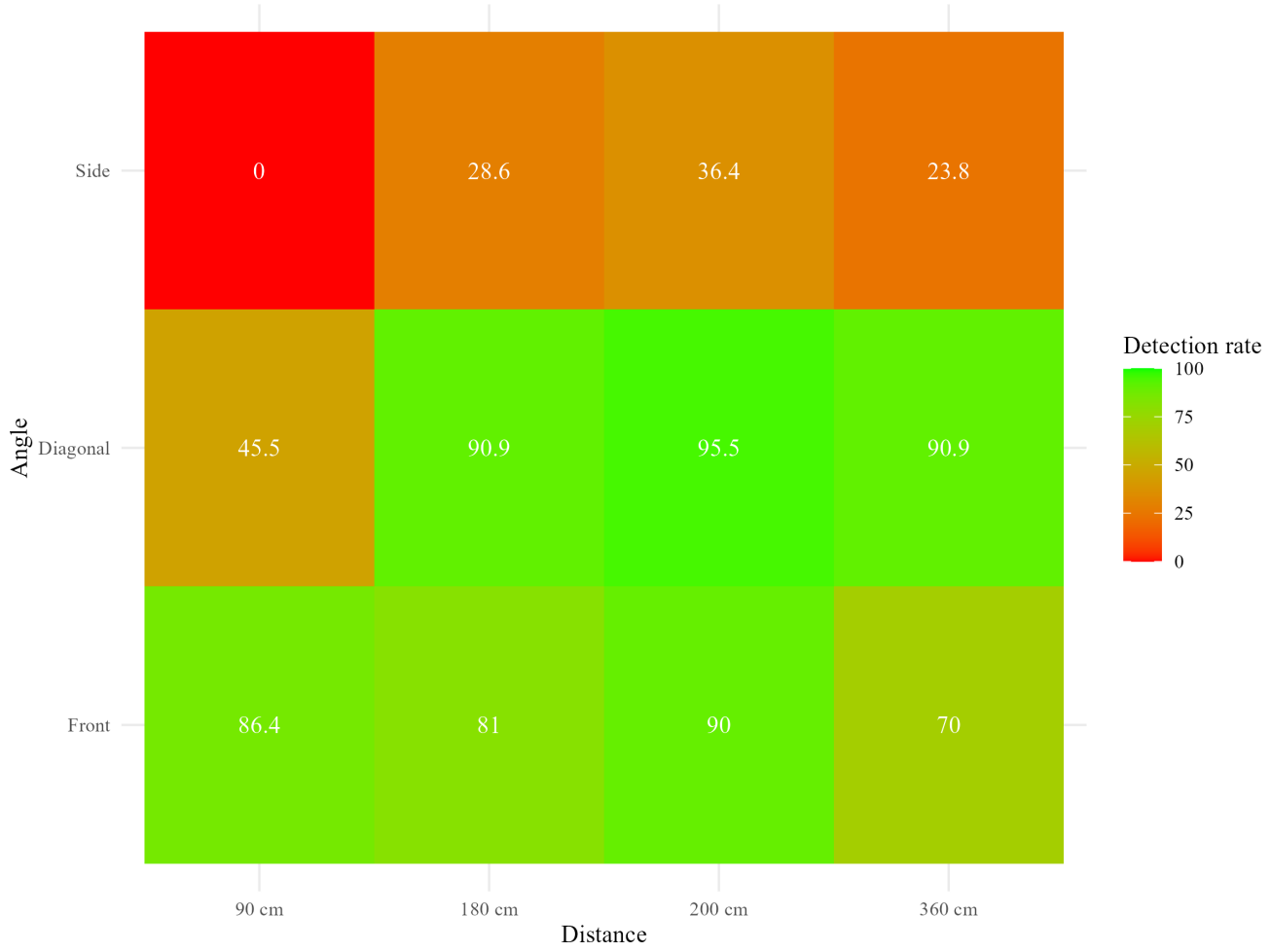


Table . Detection rate and MAE[a] for squats across camera angles and distances.

Camera angle	Distance (cm)	Detection rate (%), mean (SD)	MAE, mean (SD)
Side	90	0 (0)	5 (0)
Side	180	28.6 (46.3)	0.714 (0.463)
Side	200	36.4 (49.2)	0.636 (0.492)
Side	360	23.8 (43.6)	3.095 (2.3)
Diagonal	90	45.5 (51)	1.318 (1.673)
Diagonal	180	90.9 (29.4)	0.091 (0.294)
Diagonal	200	95.5 (21.3)	0.045 (0.213)
Diagonal	360	90.9 (29.4)	0.091 (0.294)
Front	90	86.4 (35.1)	0.318 (0.945)
Front	180	81 (40.2)	0.571 (1.248)
Front	200	90 (30.8)	0.4 (1.273)
Front	360	70 (47)	1.1 (1.889)

[a]MAE: mean absolute error.

Analyzing the binary accuracy through the GLMM (Table S5 in Multimedia Appendix 1) demonstrated that the diagonal and frontal angles significantly outperformed the side angle at 90 cm (OR=20.43, *P*<.001, and OR=23.01, *P*<.001, respectively). Detection also significantly improved at mid-range distances of 180 cm (OR=4.36, *P*=.002), 200 cm (OR=7.62, *P*<.001), and 360 cm (OR=3.054, *P*=.01) compared to 90 cm. Post hoc pairwise comparisons (Table S6 in Multimedia Appendix 1) revealed that the diagonal view at 180 cm significantly outperformed the front view at 90 cm (estimate=2.810, *P*=.02) and the front view at 180 cm (estimate=3.058, *P*=.009). However, it underperformed compared to the front view at 200

XSL•FO
RenderX

cm (estimate=−3.271, *P*=.01) and the front view at 360 cm (estimate=−4.016, *P*=.02). Side views at 180 cm and 360 cm outperformed the front view at 360 cm in correctly detecting push-ups (estimates=2.907 to 3.239, both *P*=.04). The diagonal view at 200 cm had significantly better detection accuracy than the front view at 360 cm (estimate=3.174, *P*=.03). These findings suggest that diagonal mid-range placements generally perform well, but that frontal views at longer distances (200 and 360 cm) can sometimes outperform the diagonal view at 180 cm for squat detection.

For squats, LMM analysis of MAE (Table S7 in Multimedia Appendix 1) revealed significantly lower errors for diagonal and frontal angles compared with the side (estimates were −3.682 and −4.682, respectively; both *P*<.001). All tested distances (180 cm, 200 cm, and 360 cm) were superior to the 90 cm reference (estimates ranged from −4.364 to −1.910; *P* values<.001). Several significant interactions were observed. For instance, combining a diagonal angle with 180 cm or 200 cm decreased MAE by 3.06 and 3.09 repetitions, respectively (both *P*<.001), which moderated the improvements to the main effects expected from angle and distance independently. The front × 180 cm interaction (estimate=4.547, *P*<.001) and the front × 200 cm interaction (estimate=4.460, *P*<.001) exhibited similar patterns, indicating smaller improvements than the sum of the main effects predicted. In the model, only the diagonal × 360 cm interaction was nonsignificant (*P*=.16). Several significant post hoc pairwise comparisons were identified (Table S8 in Multimedia Appendix 1). These confirmed that diagonal and frontal mid-range placements consistently minimized error. Diagonal at 200 cm produced the largest reduction (estimate=−4.96 repetitions, *P*<.001), similar to diagonal at 180 cm (estimate=−4.91 repetitions, *P*<.001). Most configurations substantially reduced MAE compared to the 90 cm side, but the magnitude of improvement depended on the specific combination of angle and distance. This highlights notable interaction effects.

Discussion

Main Findings

This study demonstrates that smartphone camera angle and distance critically affect the accuracy of PE models for detecting and counting push-ups and squats. Consistent with our results, camera placements at intermediate distances (180 - 200 cm) combined with oblique (diagonal) or frontal views generally yielded the highest detection rates and lowest counting errors. In contrast, very close-range setups (90 cm) and long-distance frontal views (360 cm) often showed reduced performance. For push-ups, diagonal views between 90 and 200 cm outperformed frontal angles, with the diagonal view at 200 cm position achieving the lowest MAE. For squats, diagonal and frontal views at 180 to 200 cm significantly outperformed side views, with diagonal 180 cm, diagonal 200 cm, and front 200 cm producing the smallest MAEs and the largest error reductions in post hoc tests. Although front 360 cm occasionally approached the accuracy of diagonal 180 cm, most mid-range configurations substantially outperformed both close-range and long-distance side views. These findings provide direct evidence that mid-range diagonal and frontal camera configurations optimize PE performance, informing best practices for smartphone-based exercise monitoring.

Comparison to Prior Work

Unlike prior lab-based, multicamera studies [11,31], our study provides practical, configuration-specific guidance for monocular smartphone setups. This approach enhances ecological validity by closely mimicking typical home or gym environments where single-camera devices are commonly used. While [17] demonstrated that full-body visibility at 300 cm supports accurate gait tracking, we show that diagonal views at 180 to 200 cm consistently optimize visibility and accuracy for dynamic strength exercises. Views at 360 cm, however, yield variable and often low detection rates (push-ups: 20% - 73.7%; squats: 23% - 90.9%), depending on the angle [46], evaluated smartphone-based distance estimation between 100 and 300 cm, the trade-offs between spatial accuracy, and usability. Our findings confirm that mid-range diagonal placements are optimal for PE performance, likely due to differences in exercise posture (horizontal for push-ups vs upright for squats).

Prior studies have evidenced smartphone apps using TensorFlow Lite and COCO-trained models for exercise counting [10,47], but have not comprehensively examined spatial configuration effects. Our results address this gap, complementing [48] who emphasized viewpoint in PE data acquisition [9], advocating for accessible PE tools beyond labs, and [12] who discussed challenges of motion tracking in naturalistic settings. By offering concrete camera setup recommendations, this work contributes actionable insights for mHealth, home fitness, telerehabilitation, sports performance, and clinical decision-making contexts.

Practical Implications for mHealth

Positioning the smartphone camera diagonally (~45°) at 180 to 200 cm significantly enhances PE accuracy without additional hardware. Mobile apps can integrate augmented reality or setup guides to assist users in achieving optimal device placement. Accurate repetition counting supports load monitoring, fatigue tracking, and muscular endurance assessment in unsupervised environments [29,49,50]. Combining spatial optimization with adaptive feedback and personalized experiences may further improve tracking reliability and user engagement, consistent with evidence from behavior change and human-computer interaction research [51,52]. This integrated approach presents a promising path for scalable, user-friendly mHealth exercise platforms.

Limitations

This study has several limitations. First, PE performance was evaluated using manual annotations rather than a gold-standard motion capture system, which may introduce variability. Second, the sample comprised healthy young adults in a controlled laboratory environment, limiting generalizability to clinical populations, older adults, or real-world contexts with variable lighting and backgrounds. Third, while only 1 PE pipeline was tested, integrating multiple public datasets and the custom DotPose dataset helped mitigate biases related to scene composition, body type, exercise variety, and environmental

XSL•FO
RenderX

scenarios. Furthermore, comparisons with other open-source PE models (eg, OpenPose, HRNet [Microsoft Research Asia], and BlazePose) are limited by variations in datasets, computational demands, and architectural design [12,35,53]. Finally, lightweight 2D models, such as MoveNet, enable near-real-time, on-device inference. These models balance speed, portability, and user engagement with the fine-grained accuracy achievable by heavier, cloud-reliant models [13]. This hybrid mobile-cloud configuration supports the robust evaluation of push-ups and squats across multiple camera setups. Nevertheless, findings may not generalize to other PE architectures, populations, or unstructured environments.

Future Directions

Future work should explore multimodal pipelines, diverse participant groups, and variable environmental conditions (eg, low-light and high-contrast conditions) to enhance robustness, applicability, and real-world relevance of mobile PE systems. While repetition counting is a fundamental first step, future mHealth systems should also evaluate movement quality. This should encompass compensatory strategies and fatigue-related adaptations [54-56]. Emerging real-time PE technologies that optimize sensor or camera placement, combined with adaptive feedback responsive to user performance, offer opportunities to enhance movement accuracy and user experience in unsupervised settings [28,48,57,58]. Longitudinal studies are needed to assess the integration of spatial setup guidance, personalization, and real-time feedback for improved usability, engagement, and long-term clinical or fitness outcomes.

Conclusions

Camera angle and distance significantly affect the accuracy of PE systems for exercise detection. For optimal performance, smartphone cameras should be positioned at mid-range distances (180 - 200 cm) with diagonal or frontal views. For push-ups, diagonal views are preferred, while for squats, both diagonal and frontal views perform well. Conversely, close-range setups (90 cm) and long-distance frontal views (360 cm) substantially reduce detection and counting accuracy. These findings provide actionable guidance for developing scalable, accurate, and user-friendly mHealth exercise tracking platforms.

Acknowledgments

The authors would like to thank the students who participated in this study for their time and commitment. They would also like to acknowledge the University of Évora for its support in facilitating the research. During the preparation of the manuscript, ChatGPT was used to suggest grammatical edits. The authors reviewed and edited all content and took full responsibility for the final version.

Funding

This work was financially supported by Project Blockchain.PT—Decentralize Portugal with Blockchain Agenda (Project 51), WP4 Sports, Leisure, and Culture, Call number 02/C05-i01.01/2022, funded by the Portuguese Recovery and Resilience Program, the Portuguese Republic, and the European Union under the framework of the Next Generation European Union Program. The funders had no involvement in the study design, data collection, analysis, interpretation of results, or the writing of the manuscript.

Data Availability

The study leveraged established benchmark datasets for human pose estimation, including COCO (Common Objects in Context) [38], SMART (Sports Motion and Recognition Tasks) [41], LSP (Leeds Sports Pose) Extended [42], Penn Action [43], and MPII Human Pose [44]. Additionally, the data collected from the 44 participants are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: EO, SF, APG
Data curation: EO, GV
Formal analysis: EO, FG-F
Investigation: EO, SF, JP, PP
Methodology: EO, GV
Project administration: SA
Supervision: APG
Writing – original draft: EO
Writing – review & editing: SF, APG, GV, JP, PP, FG-F, SA
All authors have reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

Detailed results of generalized linear mixed-effects models, linear mixed-effects models, and post-hoc pairwise comparisons for push-ups and squats.
[DOC File, 6016 KB - mhealth_v14i1e82412_app1.doc]

References

1. Dallinga J, Janssen M, van der Werf J, Walravens R, Vos S, Deutekom M. Analysis of the features important for the effectiveness of physical activity–related apps for recreational sports: expert panel approach. JMIR mHealth uHealth 2018 Jun 18;6(6):e143. [doi: 10.2196/mhealth.9459] [Medline: 29914863]
2. Kohl HW 3rd, Craig CL, Lambert EV, et al. The pandemic of physical inactivity: global action for public health. Lancet 2012 Jul;380(9838):294-305. [doi: 10.1016/S0140-6736(12)60898-8] [Medline: 23041199]
3. Kardan M, Jung A, Iqbal M, Keshtkar S, Geidl W, Pfeifer K. Efficacy of digital interventions on physical activity promotion in individuals with noncommunicable diseases: an overview of systematic reviews. BMC Digit Health 2024;2(1):40. [doi: 10.1186/s44247-024-00097-6]
4. Physical activity. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/physical-activity [accessed 2026-02-10]
5. Hallal PC, Andersen LB, Bull FC, et al. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet 2012 Jul 21;380(9838):247-257. [doi: 10.1016/S0140-6736(12)60646-1] [Medline: 22818937]
6. Buecker S, Simacek T, Ingwersen B, Terwiel S, Simonsmeier BA. Physical activity and subjective well-being in healthy individuals: a meta-analytic review. Health Psychol Rev 2021 Dec;15(4):574-592. [doi: 10.1080/17437199.2020.1760728] [Medline: 32452716]
7. Núñez-Cortés R, Salazar-Méndez J, Nijs J. Physical activity as a central pillar of lifestyle modification in the management of chronic musculoskeletal pain: a narrative review. J Funct Morphol Kinesiol 2025 May 20;10(2):183. [doi: 10.3390/jfmk10020183] [Medline: 40407467]
8. Avogaro A, Cunico F, Rosenhahn B, Setti F. Markerless human pose estimation for biomedical applications: a survey. Front Comput Sci 2023;5:1153160. [doi: 10.3389/fcomp.2023.1153160]
9. Stenum J, Cherry-Allen KM, Pyles CO, Reetzke RD, Vignos MF, Roemmich RT. Applications of pose estimation in human health and performance across the lifespan. Sensors (Basel) 2021 Nov 3;21(21):7315. [doi: 10.3390/s21217315] [Medline: 34770620]
10. El-Rajab I, Klotzbier TJ, Korbus H, Schott N. Camera-based mobile applications for movement screening in healthy adults: a systematic review. Front Sports Act Living 2025;7:1531050. [doi: 10.3389/fspor.2025.1531050] [Medline: 40416048]
11. Luangaphirom T, Lueprasert S, Kaewvichit P, Boonphotsiri S, Burapasikarin T, Siriborvornratanakul T. Real-time weight training counting and correction using MediaPipe. Adv in Comp Int 2024 Jun;4(2):3. [doi: 10.1007/s43674-024-00070-w]
12. Roggio F, Trovato B, Sortino M, Musumeci G. A comprehensive analysis of the machine learning pose estimation models used in human movement and posture analyses: a narrative review. Heliyon 2024;10(21):e39977. [doi: 10.1016/j.heliyon.2024.e39977] [Medline: 39553598]
13. Phosanarack M, Wallard L, Lepreux S, Kolski C, Avril E. Smartphone exergames with real-time markerless motion capture: challenges and trade-offs. arXiv. Preprint posted online on Jul 9, 2025. [doi: 10.48550/ARXIV.2507.06669]
14. Badiola-Bengoa A, Mendez-Zorrilla A. A systematic review of the application of camera-based human pose estimation in the field of sport and physical exercise. Sensors (Basel) 2021 Sep 7;21(18):5996. [doi: 10.3390/s21185996] [Medline: 34577204]
15. Cornman HL, Stenum J, Roemmich RT. Video-based quantification of human movement frequency using pose estimation: a pilot study. PLoS One 2021;16(12):e0261450. [doi: 10.1371/journal.pone.0261450] [Medline: 34929012]
16. Jeon H, Kim D, Kim J. Human motion assessment on mobile devices. Presented at: 2021 International Conference on Information and Communication Technology Convergence (ICTC); Oct 20-22, 2021. [doi: 10.1109/ICTC52510.2021.9621114]
17. Yang J, Park K. Improving gait analysis techniques with markerless pose estimation based on smartphone location. Bioengineering (Basel) 2024 Jan 30;11(2):141. [doi: 10.3390/bioengineering11020141] [Medline: 38391625]
18. Chen Z, Liao K, Bishop C, Bian C, Li Y. Validity of three commercial devices for recording movement velocity in the Bulgarian split squat. J Hum Kinet 2024;95:161-171. [doi: 10.5114/jhk/189365] [Medline: 39944971]
19. Stamm O, Heimann-Steinert A. Accuracy of monocular two-dimensional pose estimation compared with a reference standard for kinematic multiview analysis: validation study. JMIR mHealth uHealth 2020 Dec 21;8(12):e19608. [doi: 10.2196/19608] [Medline: 33346739]
20. van den Hoorn W, Lavaill M, Cutbush K, Gupta A, Kerr G. Comparison of shoulder range of motion quantified with mobile phone video-based skeletal tracking and 3D motion capture-preliminary study. Sensors (Basel) 2024 Jan 15;24(2):534. [doi: 10.3390/s24020534] [Medline: 38257626]
21. Mercadal-Baudart C, Liu CJ, Farrell G, et al. Exercise quantification from single camera view markerless 3D pose estimation. Heliyon 2024 Mar 30;10(6):e27596. [doi: 10.1016/j.heliyon.2024.e27596] [Medline: 38510055]
22. Ferreira B, Ferreira PM, Pinheiro G, et al. Deep learning approaches for workout repetition counting and validation. Pattern Recognit Lett 2021 Nov;151:259-266. [doi: 10.1016/j.patrec.2021.09.006]

XSL•FO
RenderX

23. Rao P, Asha CS, Rao PR. Real-time posture correction of squat exercise: a deep learning approach for performance analysis and error correction. IEEE Access 2025;13:39557-39571. [doi: 10.1109/ACCESS.2025.3545207]
24. Japhne F, Janada K, Theodorus A, Chowanda A. Fitcam: detecting and counting repetitive exercises with deep learning. J Big Data 2024;11(1):101. [doi: 10.1186/s40537-024-00915-8]
25. Sinclair A, Kautai K, Shahamiri SR. Pūioio: on-device real-time smartphone-based automated exercise repetition counting system. arXiv. Preprint posted online on Jul 22, 2023. [doi: 10.48550/ARXIV.2308.02420]
26. Fisal NR, Fathalla A, Elmanakhly DA, Salah A. Reported challenges in deep learning-based human pose estimation: a systematic review. IEEE Access 2025;13:80520-80539. [doi: 10.1109/ACCESS.2025.3567337]
27. Fanton M, Harari Y, Giffhorn M, et al. Validation of Amazon Halo Movement: a smartphone camera-based assessment of movement health. NPJ Digit Med 2022 Sep 6;5(1):134. [doi: 10.1038/s41746-022-00684-9] [Medline: 36065060]
28. Biebl JT, Rykala M, Strobel M, et al. App-based feedback for rehabilitation exercise correction in patients with knee or hip osteoarthritis: prospective cohort study. J Med Internet Res 2021 Jul 13;23(7):e26658. [doi: 10.2196/26658] [Medline: 34255677]
29. American College of Sports Medicine. In: Liguori G, Feito Y, Fountaine C, Roy BA, editors. ACSM's Guidelines for Exercise Testing and Prescription, 11th edition: Wolters Kluwer; 2021. URL: https://books.google.co.in/books?id=yjibzQEACAAJ [accessed 2026-02-10]
30. Miranda GHL. Biofotogrametria Para Fisioterapeutas [Book in Portuguese], 1st edition: Andreoli; 2014. URL: https://www.indicalivros.com/livros/biofotogrametria-para-fisioterapeutas-geraldo-henrique-lopes-miranda [accessed 2026-02-10]
31. Park HJ, Baek JW, Kim JH. Imagery based parametric classification of correct and incorrect motion for push-up counter using OpenPose. Presented at: 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE); Aug 20-21, 2020. [doi: 10.1109/CASE48305.2020.9216833]
32. Youssef F, Zaky AB, Gomaa W. Analysis of the squat exercise from visual data. Presented at: 19th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2022); Jul 14-16, 2022. [doi: 10.5220/0011347900003271]
33. Hande T, Kakirwar B, Bharadwaja AV, Kshirsagar P, Gupta A, Vijayakumar P. Correction and estimation of workout postures with pose estimation using AI. Presented at: 2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE); Jan 27-28, 2023. [doi: 10.1109/IITCEE57236.2023.10090463]
34. Chae HJ, Kim JB, Park G, O'Sullivan DM, Seo J, Park JJ. An artificial intelligence exercise coaching mobile app: development and randomized controlled trial to verify its effectiveness in posture correction. Interact J Med Res 2023 Sep 12;12:e37604. [doi: 10.2196/37604] [Medline: 37698913]
35. Chariar M, Rao S, Irani A, Suresh S, Asha CS. AI Trainer: autoencoder based approach for squat analysis and correction. IEEE Access 2023;11:107135-107149. [doi: 10.1109/ACCESS.2023.3316009]
36. Zhang X, Han SZH, Lim KYT. Designing and prototyping of AI-based real-time mobile detectors for calisthenic push-up exercise. Presented at: International Conference on Health and Social Care Information Systems and Technologies (HCist 2023); Nov 8-10, 2023. [doi: 10.1016/j.procs.2024.06.192]
37. Dotmoovs. URL: https://www.dotmoovs.com/ [accessed 2026-02-10]
38. Lin TY, Maire M, Belongie S, et al. Microsoft COCO: common objects in context. Presented at: 13th European Conference on Computer Vision (ECCV 2014); Sep 6-12, 2014. [doi: 10.1007/978-3-319-10602-1_48]
39. Votel R, Li N. Next-generation pose detection with MoveNet and TensorFlow.js. TensorFlow Blog. 2021 May 17. URL: https://blog.tensorflow.org/2021/05/next-generation-pose-detection-with-movenet-and-tensorflowjs.html [accessed 2025-10-28]
40. Kumar P, Chauhan S, Awasthi LK. Human pose estimation using deep learning: review, methodologies, progress and future research directions. Int J Multimed Info Retr 2022;11(4):489-521. [doi: 10.1007/s13735-022-00261-6]
41. Chen X, Pang A, Yang W, Ma Y, Xu L, Yu J. SportsCap: monocular 3D human motion capture and fine-grained understanding in challenging sports videos. Int J Comput Vis 2021;129(10):2846-2864. [doi: 10.1007/s11263-021-01486-4]
42. Johnson S, Everingham M. Learning effective human pose estimation from inaccurate annotation. Presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Jun 20-25, 2011. [doi: 10.1109/CVPR.2011.5995318]
43. Zhang W, Zhu M, Derpanis KG. From actemes to action: a strongly-supervised representation for detailed action understanding. Presented at: Proceedings of the IEEE International Conference on Computer Vision (ICCV) 2013; Dec 1-8, 2013. [doi: 10.1109/ICCV.2013.280]
44. Andriluka M, Pishchulin L, Gehler P, Schiele B. 2D human pose estimation: new benchmark and state of the art analysis. Presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Jun 23-28, 2014. [doi: 10.1109/CVPR.2014.471]
45. Dibenedetto G, Sotiropoulos S, Polignano M, Cavallo G, Lops P. Comparing human pose estimation through deep learning approaches: an overview. Comput Vis Image Underst 2025 Feb;252:104297. [doi: 10.1016/j.cviu.2025.104297]
46. Hamilton-Fletcher G, Liu M, Sheng D, et al. Accuracy and usability of smartphone-based distance estimation approaches for visual assistive technology development. IEEE Open J Eng Med Biol 2024;5:54-58. [doi: 10.1109/OJEMB.2024.3358562] [Medline: 38487094]

XSL•FO
RenderX

47. Jeon H, Yoon Y, Kim D. Lightweight 2D human pose estimation for fitness coaching system. Presented at: 2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC); Jun 27-30, 2021. [doi: 10.1109/ITC-CSCC52171.2021.9501458]
48. Konak O, Wischmann A, van De Water R, Arnrich B. A real-time human pose estimation approach for optimal sensor placement in sensor-based human activity recognition. Presented at: 8th International Workshop on Sensor-Based Activity Recognition and Artificial Intelligence (iWOAR 2023); Sep 21-22, 2023. [doi: 10.1145/3615834.3615848]
49. Ebben WP, Feldmann CR, Dayne A, et al. Using squat testing to predict training loads for the deadlift, lunge, step-up, and leg extension exercises. J Strength Cond Res 2008 Nov;22(6):1947-1949. [doi: 10.1519/JSC.0b013e31818747c9] [Medline: 18978614]
50. Lin II, Chen YL, Chuang LL. Test-retest reliability of home-based fitness assessments using a mobile app (R Plus Health) in healthy adults: prospective quantitative study. JMIR Form Res 2021 Dec 8;5(12):e28040. [doi: 10.2196/28040] [Medline: 34657835]
51. Baretta D, Perski O, Steca P. Exploring users' experiences of the uptake and adoption of physical activity apps: longitudinal qualitative study. JMIR mhealth uHealth 2019 Feb 8;7(2):e11636. [doi: 10.2196/11636] [Medline: 30735143]
52. Li X, Li Y. Sports training strategies and interactive control methods based on neural network models. Comput Intell Neurosci 2022;2022:7624578. [doi: 10.1155/2022/7624578] [Medline: 35295278]
53. Edriss S, Romagnoli C, Caprioli L, Bonaiuto V, Padua E, Annino G. Commercial vision sensors and AI-based pose estimation frameworks for markerless motion analysis in sports and exercises: a mini review. Front Physiol 2025;16:1649330. [doi: 10.3389/fphys.2025.1649330] [Medline: 40873758]
54. Liao Y, Vakanski A, Xian M. A deep learning framework for assessing physical rehabilitation exercises. IEEE Trans Neural Syst Rehabil Eng 2020 Feb;28(2):468-477. [doi: 10.1109/TNSRE.2020.2966249] [Medline: 31940544]
55. Bauer CM, Kankaanpää MJ, Meichtry A, Rissanen SM, Suni JH. Efficacy of six months neuromuscular exercise on lumbar movement variability - a randomized controlled trial. J Electromyogr Kinesiol 2019 Oct;48:84-93. [doi: 10.1016/j.jelekin.2019.06.008] [Medline: 31252284]
56. Woo Y, Jeong H. Exercise assessment based on human pose estimation and relative phase for real-time remote exercise system. IEEE Access 2025;13:53203-53213. [doi: 10.1109/ACCESS.2025.3551834]
57. Tharatipyakul A, Srikaewsiew T, Pongnumkul S. Deep learning-based human body pose estimation in providing feedback for physical movement: a review. Heliyon 2024 Aug 26;10(17):e36589. [doi: 10.1016/j.heliyon.2024.e36589] [Medline: 39281455]
58. Elsayed H, Hoffmann P, Günther S, et al. CameraReady: assessing the influence of display types and visualizations on posture guidance. Presented at: DIS '21: Proceedings of the 2021 ACM Designing Interactive Systems Conference; Jun 28 to Jul 2, 2021. [doi: 10.1145/3461778.3462026]

Abbreviations

AI: artificial intelligence
COCO: Common Objects in Context
GLMM: generalized linear mixed model
LMM: linear mixed effects model
LSP: Leeds Sports Pose
MAE: mean absolute error
mHealth: mobile health
OR: odds ratio
PE: pose estimation
SMART: Sports Motion and Recognition Tasks

Edited by L Buis; submitted 14.Aug.2025; peer-reviewed by LA Morales-Rosales, T Sasatani; revised version received 04.Nov.2025; accepted 27.Jan.2026; published 05.Mar.2026.

Please cite as:
Oliosi E, Ferreira S, Giordano AP, Viveiros G, Parraca J, Pereira P, Guede-Fernández F, Azevedo S
Evaluation of Smartphone Camera Positioning on Artificial Intelligence Pose Estimation Accuracy for Exercise Detection: Observational Study
JMIR Mhealth Uhealth 2026;14:e82412
URL: https://mhealth.jmir.org/2026/1/e82412
doi:10.2196/82412

XSL•FO
RenderX

© Eduarda Oliosi, Soraia Ferreira, Ana Paula Giordano, Guilherme Viveiros, José Parraca, Paulo Pereira, Federico Guede-Fernández, Salomé Azevedo. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 5.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Investigating Children's Exposure to Outdoor Food Marketing in 2 Swedish Cities Using a Smartphone App: Cross-Sectional Study

Sofia Spolander[1], BSc, MSc; Petter Fagerberg[2], MSc, PhD; Billy Langlet[1], MSc, PhD; Ioannis Sarafis[3], MSc, PhD; Ioannis Ioakimidis[1], MSc, PhD

[1]Department of Medicine, Huddinge, Karolinska Institutet, Blickagången 16, Stockholm, Sweden

[2]Karolinska Institutet, Stockholm, Sweden

[3]Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

Corresponding Author:
Sofia Spolander, BSc, MSc
Department of Medicine, Huddinge, Karolinska Institutet, Blickagången 16, Stockholm, Sweden

Abstract

Background: Childhood obesity and unhealthy dietary habits remain major public health concerns and are influenced by the surrounding food environment. Food marketing, particularly for ultraprocessed foods (UPFs), shapes children's food preferences and consumption. However, food environments are complex and constantly changing, making them difficult to map and monitor. Developing approaches that capture these dynamics is essential to understand and address children's exposure to unhealthy food marketing.

Objective: This study aimed to pilot-test a novel tool consisting of a smartphone app and dashboard designed to identify areas where children are exposed to outdoor food advertising. Additionally, the study assessed the prevalence of advertisements for UPFs, health-promoting foods, and offers in the identified areas and explored differences in exposure by city size and socioeconomic status (SES).

Methods: A cross-sectional study was performed in 2 Swedish counties. Initially, 46 children from 4 schools in areas with varying SES used a smartphone app to take pictures of food advertisements that they encountered in their everyday lives. The app also recorded the GPS locations of where the pictures were taken. Pictures with associated GPS data were automatically uploaded and visualized in a web-based dashboard, allowing for identification of areas where children see many food advertisements, so-called "hotspot areas." The identified hotspot areas were subsequently visited by 2 researchers (SS and PF), who systematically photographed all food advertisements in the areas. All pictures taken by the researchers were later analyzed based on their content of UPFs, health-promoting foods such as fruits, berries, vegetables, and seafood (FBVS), and price promotions.

Results: Based on 1308 pictures of outdoor food advertisements taken by children using the app, 34 hotspot areas were identified through the dashboard. In these areas, researchers collected 2955 pictures of outdoor food advertisements during the mapping activity. Overall, 77.5% (2291/2955) of advertisements promoted UPFs, with no significant difference between the large and small cities. In Stockholm, a higher proportion of UPFs appeared in the high-SES area compared to the low-SES area, though the proportion was high in both areas. FBVS featured in 20.8% (616/2955) of advertisements, slightly more often in Stockholm and in the high-SES area. Price promotions appeared in 23.6% (697/2955) of advertisements, mainly featured UPFs (518/697, 74.3%) and less often FBVS (142/697, 20.4%). Price promotions for UPFs were somewhat more frequent in Stockholm, and FBVS promotions were more common in the high-SES area.

Conclusions: Using a novel, child-centered, mobile health tool combining a smartphone app and dashboard, this study identified local food advertising hotspots. Most advertisements in the hotspots promoted UPFs, while a few featured FBVS, a pattern also reflected in price promotions. These trends were consistent across areas, highlighting a food marketing landscape misaligned with dietary guidelines and the need for policy action.

(JMIR Mhealth Uhealth 2026;14:e70192) doi:10.2196/70192

KEYWORDS

food environment; marketing; advertisement; ultraprocessed food; unhealthy foods; children; adolescents; obesity; social class; children-centric approach

XSL•FO
RenderX

Introduction

The prevalence of childhood obesity has increased dramatically in Sweden over the past few decades, following global trends [1,2]. This is concerning since childhood obesity is associated with an increased risk of several chronic diseases, including type 2 diabetes and cardiovascular disease, as well as various psychological problems, including low self-esteem and depression [3,4]. Moreover, childhood obesity increases the risk of obesity and developing chronic diseases in adulthood [5,6].

Poor dietary habits are a major modifiable risk factor for obesity. In recent years, increasing attention has been given to the role of the food environment in shaping poor dietary habits [7]. In parallel with the increase in obesity, the food environment has undergone rapid and widespread changes, resulting in increased access to cheap foods that are low in nutrients and high in calories, such as ultraprocessed foods (UPFs) [8]. High availability and consumption of UPFs have been associated with increased risk of obesity in both children [9] and adults [10,11], as well as an increased risk of all-cause mortality [12,13]. Conversely, the consumption of foods such as fruits, vegetables, whole grains, fish, and seafood is associated with protective effects against obesity and noncommunicable diseases [14]. This contrast points to the value of mapping food environments to inform strategies that promote healthier diets.

Food marketing represents a key element of the food environment, influencing food choices and consumption patterns. A growing number of studies show that the physical food marketing landscape is dominated by advertisements for foods associated with negative health outcomes such as UPFs [15]. A review assessing studies from 21 countries found that on average, 63% of food advertisements promoted unhealthy foods [15]. However, different studies use different terminology, including UPFs, “discretionary foods,” “noncore foods,” and “junk food,” to describe more or less the same type of unhealthy food items [15]. Studies investigating food advertisements around schools have found that the vast majority (~60%-90%) of food advertisements market unhealthy foods [16-23].

Food environments dominated by UPF marketing pose a health risk, as food marketing has been linked to increased intake, altered preferences, and purchase requests in children [24]. Experimental studies show that children exposed to food advertisements, on television or in advergames, consume more calories than those exposed to nonfood advertisements [25-27]. These effects are greater in children with overweight or obesity [25], and exposure increases the likelihood of choosing advertised products [27]. The impact of promoting healthier foods remains uncertain, as most studied products are of low nutritional value [27].

Beyond general exposure, the distribution of food advertising may also vary by socioeconomic status (SES). Our previous study conducted in 2 Stockholm neighborhoods, 1 with low SES with higher rates of childhood obesity and 1 with high SES with lower rates of childhood obesity, found that 65% of food advertisements promoted UPFs, with a higher proportion in the low-SES area (73%) compared to the high-SES area (59%) [28]. Studies from other countries have shown mixed results in regard to SES differences [15].

Mapping and measuring food environments, including exposure to food advertising, present several challenges due to their constant change and complexity [29]. This underscores the need for methods that enable continuous or repeated assessments without requiring extensive resources. To fully understand the scope and public health implications of food environments, it is also essential to use diverse measurement approaches that address multiple dimensions of the food environment [29]. Mobile health (mHealth) technologies, such as smartphone apps, offer a promising solution by facilitating real-time data collection directly from participants. These tools can capture locational, behavioral, and environmental information, allowing researchers to identify patterns of exposure and behavior in different groups or individuals and over time. Integrating mHealth approaches can therefore enhance both the efficiency and depth of environmental assessment, providing valuable insights into how food marketing and other environmental factors influence health outcomes and inform policy design.

While previous studies have investigated the outdoor food advertising landscape around schools and other locations frequently visited by children, the mapped areas have typically been defined by researchers [16-23,28,30]. In contrast, this study adopts a child-centered approach to investigate children’s exposure to outdoor food advertising.

Given this context, the objectives of this study were (1) to pilot-test a novel tool, including a smartphone app and a dashboard, designed to identify areas where children are exposed to outdoor food advertising; (2) to assess the proportion of UPFs, health-promoting foods (such as fruits and vegetables), and price promotions for food and beverages within the identified areas; and (3) to examine potential differences in exposure by city size and SES.

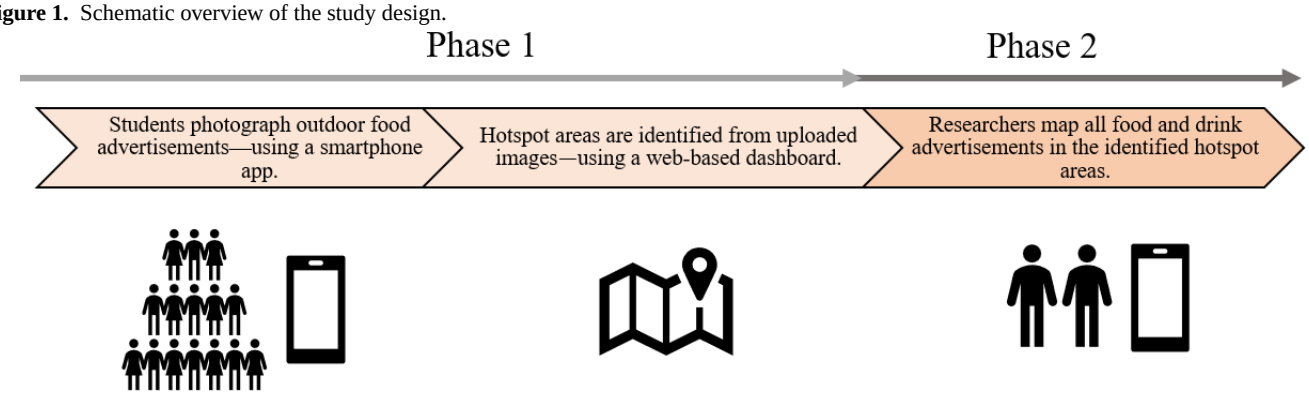
Methods

Study Design

A cross-sectional study was performed from April to June 2022. The study took place in 2 Swedish counties: Stockholm County (Stockholm) with approximately 2.5 million inhabitants (in 2024) [31] and Gävleborg County (Gävleborg), located approximately 200 km north of Stockholm, with approximately 0.3 million inhabitants (in 2024) [31]. In phase 1 of the study, a tool consisting of a mobile app and a dashboard was used to identify areas where children are exposed to many food advertisements (hotspot areas). In phase 2, researchers systematically mapped all food advertisements within the identified hotspot areas to quantify the types of food being advertised (see Figure 1 for a schematic presentation of the study design).

XSL•FO
RenderX

Figure 1. Schematic overview of the study design.



Participants and Recruitment

Children were recruited from secondary schools in Stockholm and Gävleborg. The aim was to include schools from areas with different SES in both a larger and a smaller city. Recruitment began by contacting schools in areas with varying SES levels via email and inviting them to participate. In total, 8 schools initially expressed interest, and online meetings were held to provide teachers with detailed information about the study. Ultimately, 4 schools were included based on expressed interest and scheduling availability during the study period.

In Stockholm, the capital of Sweden, areas with high and low SES were clearly distinguishable. In contrast, in Gävleborg, which is a smaller county, areas were more mixed. Consequently, clear SES differences were only present between the schools in Stockholm. The SES of the included areas was determined using the Segregation Barometer, a tool developed by the Swedish National Board of Housing, Building and Planning that incorporates statistics on education, income, health, and related factors [32].

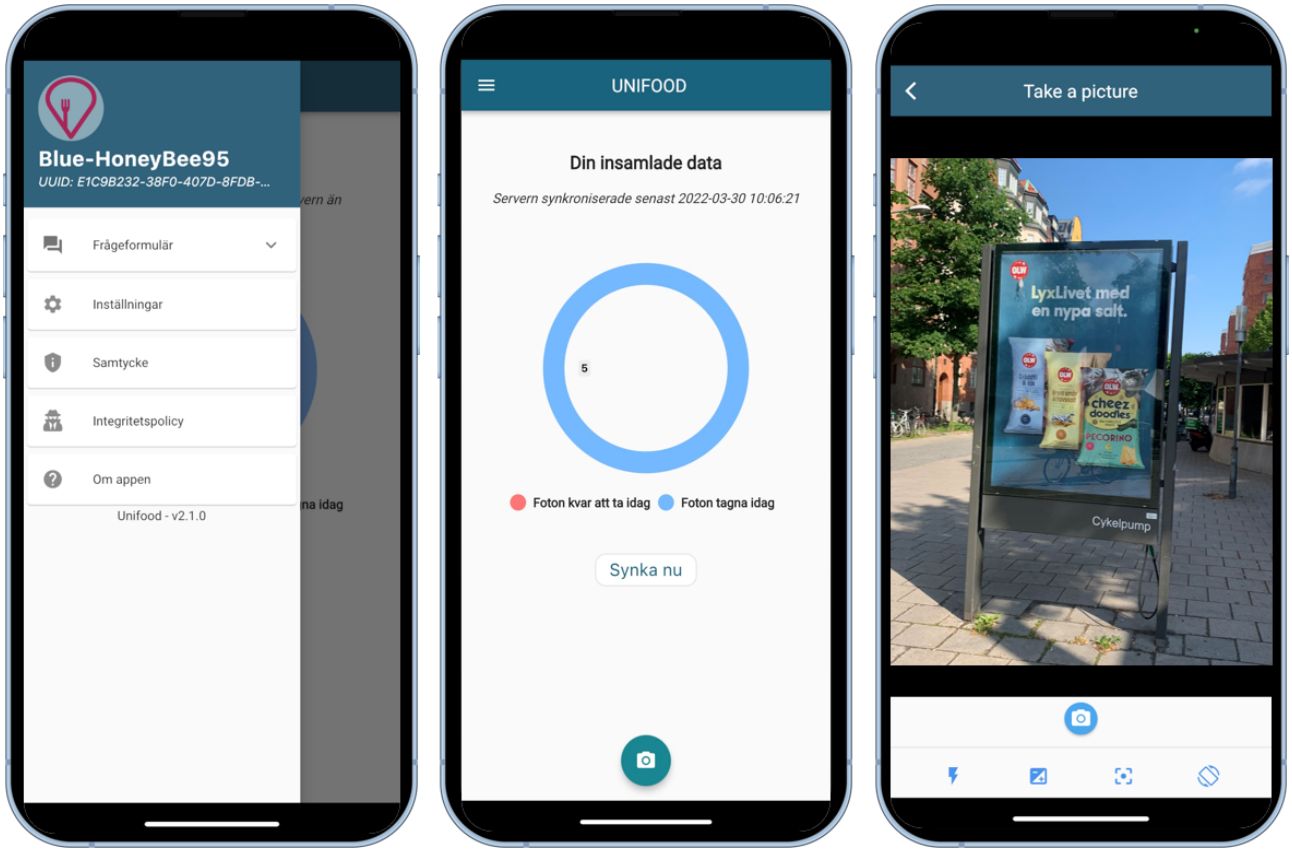
The four participating schools were located (1) in Östermalm, an urban area in central Stockholm with high SES (~83,000 inhabitants in 2021); (2) in Södertälje, an urban area in the outskirts of Stockholm with low SES (~76,000 inhabitants in 2020); (3) in central Gävle, an urban area with mixed SES; and (4) in Bomhus, an area with mixed SES in the outskirts of Gävle (~79,000 inhabitants in Gävle in 2020). Each of the 4 participating schools selected 1 or 2 school classes with students aged 13 to 14 years for participation in the study. Given the exploratory nature of the study, the sample size was determined based on feasibility and participant accessibility, rather than formal statistical calculations.

Data Collection Tools: Mobile App and Dashboard

The data collection tools for this study consisted of a custom-developed smartphone app and a purpose-built web-based dashboard. These tools were designed primarily to facilitate the identification of hotspot areas based on participant reports. The app's key feature was a camera function that allowed users to take pictures of outdoor food advertisements (Figure 2). Along with the picture itself, the app recorded the GPS location of where it was taken. When the smartphone was connected to the internet (via Wi-Fi or mobile data), the pictures with associated GPS data were automatically uploaded to a secure server and visualized in the dashboard. The app also included a function that assigned users random usernames upon registration in the app, a questionnaire in which users could answer questions about their gender, age, height, and weight (Figure 2), and visual feedback on each user's daily photo contributions (Figure 2).

XSL•FO
RenderX

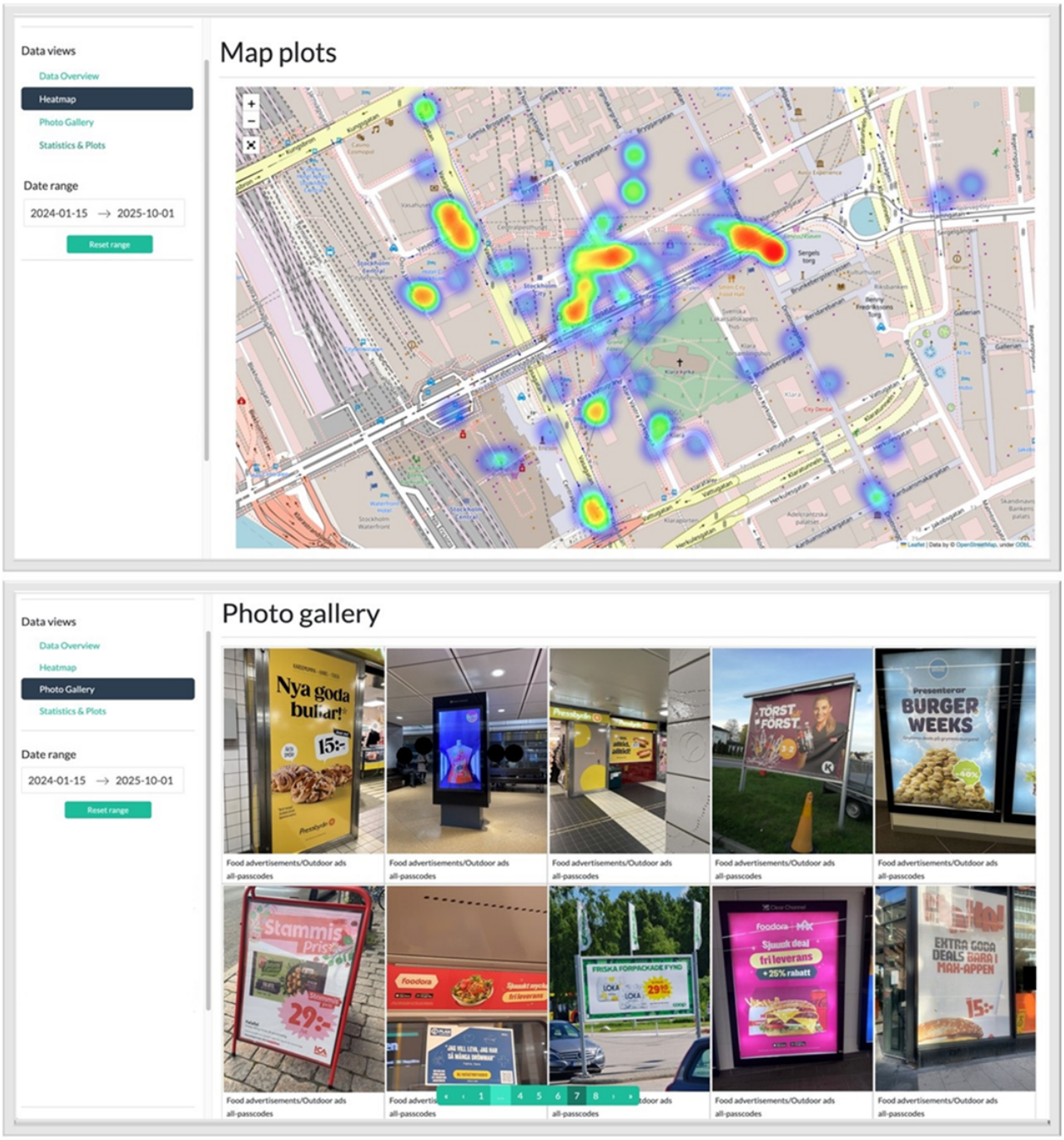
Figure 2. The mobile app used by the participants to take pictures of food advertisements. Left: the menu with information about the username and access to questionnaires and settings. Middle: the main screen of the smartphone app illustrating the daily picture contribution in blue. Right: the camera function.



The dashboard (Figure 3) was a cloud-based web service accessible through a web browser. It was developed in Python (Python Software Foundation) using Dash and the Flask web framework. Heat map visualizations were implemented using the Leaflet library. Its main purpose was to provide heat maps of aggregated GPS-tagged data collected from the app along with basic metadata filtering functionalities (eg, date range and photo category). Additionally, it included a photo gallery for quick inspection of the collected data, as well as simple statistical plots and visualizations. The app and dashboard were developed by the Multimedia Understanding Group, Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Greece.

XSL•FO
RenderX

Figure 3. The dashboard. Top: heat map visualization used to identify hotspot areas. Bottom: picture gallery.



Phase 1: Identification of Hotspot Areas Using the App and Dashboard

The study started with a session, during which researchers helped the participating children to install the app on their own smartphones. After downloading the app, children registered and answered the in-app questionnaire. Participants then received information from the researchers on how to use the app, including how to take pictures of outdoor food advertisements. They were instructed to avoid taking pictures where people's faces were visible and to avoid taking pictures at their home address. Moreover, participants were informed that the study duration was 2 weeks and that they should take a minimum of 5 pictures of outdoor food advertisements per day, but only the ones that they encounter while living their normal lives (ie, not go out of their way to find food advertisements). Participants were instructed that they could photograph the same advertisement multiple times if they encountered it on different occasions (eg, seeing it daily on their way to school), but only once per occasion. Finally, participants were instructed on what constitutes an outdoor food advertisement with examples (Figure 4). Advertisements in public transportation systems, such as in the subway or on buses, as well as general spaces in shopping malls, were considered as outdoor advertisements, whereas in-store advertisements were not.

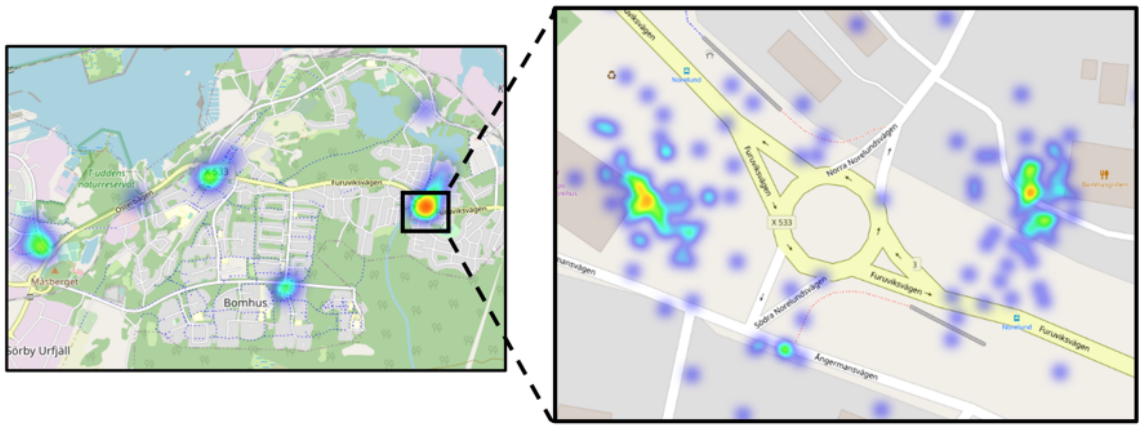
XSL•FO
RenderX

Figure 4. Different types of physical food advertisements common in the Swedish context that were considered in the study.



After 2 weeks of data collection, areas where the children reported many food advertisements, so-called "hotspot areas," could be identified via the web-based dashboard (Figure 5). An area was classified as a hotspot if more than 5 pictures of food advertisements were located within a ~50 m radius.

Figure 5. Illustration of the heat maps used to identify hotspot areas at 2 different zoom levels (displayed in the web-based dashboard).



Phase 2: Complementary Researcher Data Collection

Prior to mapping all food advertisements within the identified hotspot areas, the researchers conducted a training session to ensure consistency and reduce the risk of observer bias. During the training session, the 2 researchers independently mapped a nonhotspot area to ensure a similar approach to mapping of advertisements. Any differences in mapping strategies and definitions were discussed, resulting in a more clearly defined food advertisement mapping protocol.

After the training session, each hotspot area identified through the dashboard was converted to a square-shaped hotspot buffer, measuring either 100×100 m or 250×250 m, depending on the number and spatial distribution of pictures taken by children in that area. These buffers served as guides for researchers, indicating where to take pictures of food advertisements. Between June 7 and June 14, 2022, a total of 2 researchers systematically visited each buffer and photographed every food advertisement in all hotspot buffers using the app.

Picture Annotations

An annotation scheme previously developed by the research group [28] was adapted for this study and used as a guideline for annotating the collected pictures. All pictures of food advertisements taken by the researchers in the hotspot areas were annotated based on the content of the following foods: UPFs (including pizza, burgers, kebab, sweet snacks, salty snacks, readymade meals, sugary cereal, sweetened dairy products, sugary drinks, and artificially sweetened beverages; a full list of food and beverages included in the category is available in Multimedia Appendix 1), fruits, berries, and/or vegetables, and fish and/or other seafood. If an advertisement illustrated a meal, more than one-third of the meal had to contain fruits, berries, or vegetables to be annotated as containing fruits, berries, or vegetables (eg, salad on a hamburger or a berry on top of a dessert did not meet the criteria). Logos of well-known fast-food restaurants were annotated as fast food, whereas logos of grocery stores were not assigned to any specific food category. A single image could include multiple categories, for example, both UPFs and fruits, berries, or vegetables.

XSL•FO
RenderX

Two trained nutritionists (PF and SS) annotated all pictures independently using the VIA annotation software for images, audio, and video (Visual Geometry Group, University of Oxford). The pictures were presented in a random order to reduce the risk of location-related bias. Annotation disagreements were discussed until consensus was reached between the 2 researchers. Following this process, the categories fruits, vegetables, and/or berries and fish and other seafood were combined into the composite category "fruits, vegetables, berries, and seafood" (FBVS). Subsequently, 1 trained nutritionist (SS) annotated all pictures to identify the presence of price promotions (eg, discounts, multibuy offers, or super-size deals).

Statistical Analysis

All statistical tests were performed using R software (R Foundation for Statistical Computing). The hypothesis and analysis methods were specified before data collection.

The Pearson chi-square test was selected as the analysis method to compare the proportion of different food advertisements, due to all cells having more than 5 observations. The chi-square tests were performed to compare the proportions of advertisements containing UPFs, FBVS, and price promotions between the included schools based on (1) in what city the school was located (larger vs smaller city) and (2) the SES of the area in which the school was located (low vs high SES). For the comparison based on city size, the 2 schools in Gävleborg were compared against the 2 schools in Stockholm. For the comparison based on SES, only the 2 schools in Stockholm were compared, since there was no clear difference in SES between the 2 areas in Gävleborg.

To evaluate interrater agreement between the 2 researchers' annotations, Cohen κ was used. κ value of <0.21 was interpreted as "no agreement," 0.21‐0.39 "minimal agreement," 0.40‐0.59 "weak agreement," 0.60‐0.79 "moderate agreement," 0.80‐0.90 "strong agreement," and >0.90 as "almost perfect agreement" [33]. The interrater analysis was performed prior to the consensus discussion between the 2 researchers.

Ethical Considerations

This study was approved by the Swedish Ethical Review Authority (DNR 2021-06800-01). Prior to participation, the researchers visited children in their classrooms to provide them with detailed information about the study, including what it means to participate, that participation is voluntary, and that they are free to withdraw consent at any time. If a participant chose to withdraw their consent, they could also request the deletion of all their collected data. Children were also provided with information letters with consent forms for both themselves and their legal guardians to sign if they agreed to participate. Other than providing signed consent, there were no additional inclusion or exclusion criteria. As compensation, individuals received 1 cinema ticket for their participation, regardless of the number of pictures submitted, once the data collection period was completed. To ensure participant privacy, all collected data were pseudonymized. Upon registration in the app, each participant was automatically assigned a random username, which was linked to their uploaded data within the data collection system. As a result, the data remained fully anonymous within the system. However, a separate key file containing the link between usernames and participants' names and schools was securely stored outside the data collection system. This file was accessible only to the research team and was used exclusively to identify users who might have uploaded inappropriate content, such as images containing identifiable personal information, if necessary. All uploaded data were stored on secure servers, accessible only by selected members of the research team. Submitted pictures were reviewed, and pictures not meeting the study's criteria (eg, those unrelated to outdoor food or beverage advertisements) were permanently deleted.

Results

Participant Characteristics (Phase 1)

A total of 46 children participated in the study, and they collected a total of 1308 pictures of food and beverage advertisements using the app. The participant characteristics and the number of pictures taken by the participants in different areas can be seen in Table 1.

XSL•FO
RenderX

Table . Participant characteristics and number of pictures taken by the participants in Stockholm and Gävleborg.

School	Participants, n (%)	Sex		Age[a] (years), mean (SD)	BMI[b] (kg/m^2), mean (SD)	Total pictures taken[c], n (%)	Relevant pictures, n (% of all pictures)	Total pictures per participant, mean (SD)
		Female, n (%)	Male, n (%)					
Stockholm								
Östermalm	15 (32.6)	11 (73)	4 (27)	14.2 (0.4)	18.1 (2.0)	487 (36.7)	485 (99.5)	32.5 (33.8)
Södertälje	7 (15.2)	3 (43)	4 (57)	13.9 (0.3)	19.7 (1.9)	206 (15.5)	204 (99)	29.4 (14.4)
Total Stockholm	22 (47.8)	14 (54)	8 (46)	14.1 (0.4)	18.7 (2.1)	693 (52.2)	689 (99.4)	31.5 (28.7)
Gävleborg								
Central Gävle	5 (10.9)	0 (0)	5 (100)	14.8 (0.2)	19.8 (1.6)	147 (11.1)	144 (98)	29.4 (35.4)
Bomhus	19 (41.3)	5 (26)	14 (74)	13.8 (0.4)	20.6 (3.3)	488 (36.7)	475 (97.3)	25.7 (19.6)
Total Gävleborg	24 (52.2)	5 (21.0)	19 (79)	14.0 (0.5)	20.5 (3.1)	635 (47.8)	619 (97.5)	26.5 (22.8)
Total	46 (100)	19 (41)	27 (59)	14.0 (0.5)	19.7 (2.8)	1328 (100)	1308 (98.6)	28.9 (25.6)

[a]Data missing for 3 participants in Östermalm and 1 participant in central Gävle.

[b]Data missing for 4 participants in Östermalm, 1 participant in Södertälje, and 1 participant in central Gävle.

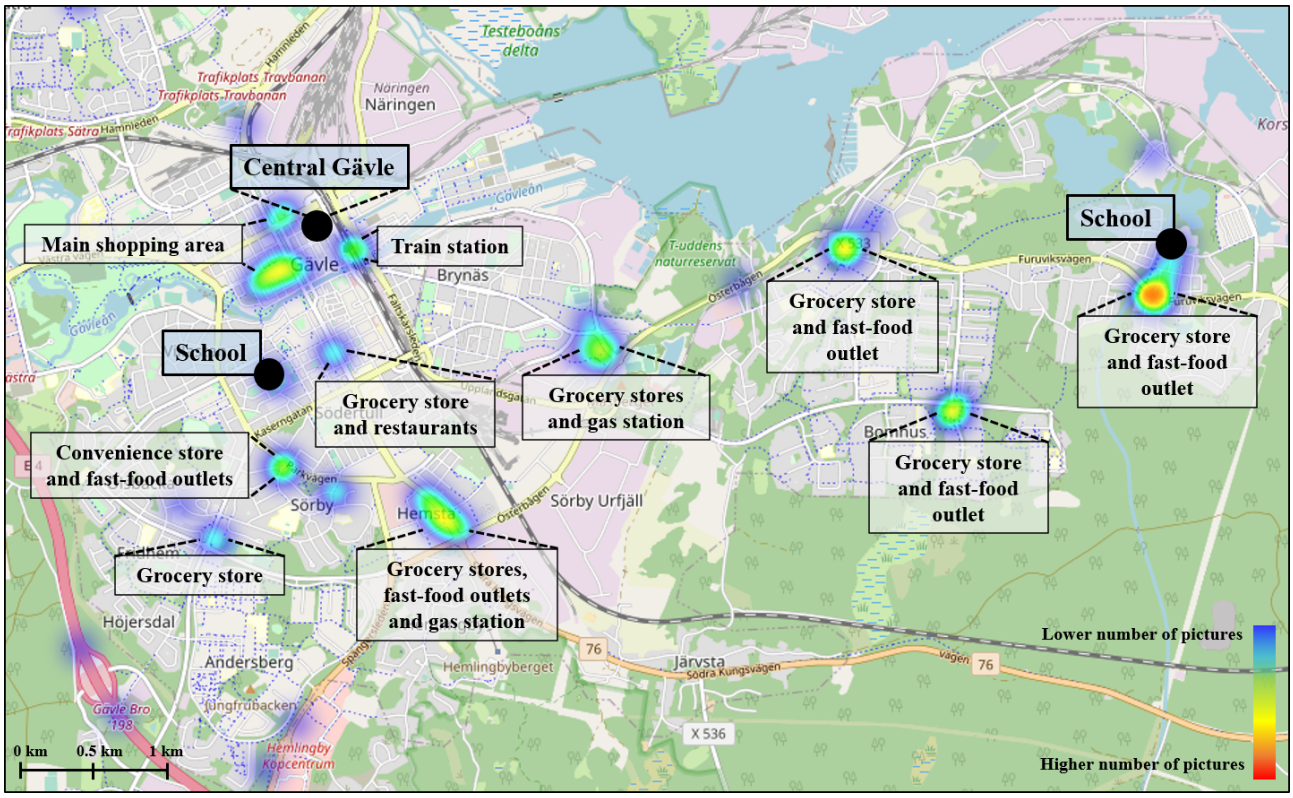
[c]Including pictures that were not in the scope of the study.

Characteristics of the Identified Hotspot Areas (Phase 1)

Across the 2 counties, 34 hotspot areas were identified using the tool, covering a total area of 740,000 m^2 (74 hectares). At least 1 hotspot area was identified near each school. These areas typically included local grocery stores, convenience stores, and/or fast-food outlets. In Gävleborg and Södertälje, additional hotspot areas were identified further from the schools, also containing similar types of food establishments. Both locations also had hotspot areas located in larger shopping areas. In central Stockholm (Östermalm school), hotspot areas were primarily concentrated in and around subway stations used by children, as well as in shopping malls and outdoor shopping areas. Figure 6 presents a map of the hotspot areas in Gävleborg extracted from the dashboard, along with examples of the types of establishments identified within these areas. Corresponding maps for Stockholm (Östermalm and Södertälje) are provided in Multimedia Appendix 2.

XSL•FO
RenderX

Figure 6. Map depicting the identified hotspot areas in Gävleborg, along with the types of establishments located within each area. The map was extracted from the research dashboard, with labels manually added to indicate the specific establishments present in the areas. The color gradient reflects the number of images associated with each hotspot, ranging from dark blue (indicating a low number of images) to red (indicating a high number of images).



Mapping of Hotspot Areas by the Researchers (Phase 2)

A total of 2955 pictures were taken during the food advertisement mapping activity conducted by the researchers. The number and total size of the identified hotspot areas as well as the number of pictures taken in the hotspot areas were greater in Stockholm compared to Gävleborg (Table 2). However, the mean number of pictures taken per hectare was similar in the 2 counties (42 pictures per hectare in Stockholm vs 37 pictures per hectare in Gävleborg).

Table . Presentation of the number and the size of the hotspot areas per city and school and the number of pictures taken by the researchers in each of the hotspot areas.

School	Hotspot areas, n/N (%)	Size (hectare)	Pictures, n/N (%)	Pictures per hectare
Stockholm	22/34 (64.7)	46.5	1948/2955 (65.9)	41.9
Östermalm	14/34 (41.2)	28.3	1350/2955 (45.7)	47.8
Södertälje	8/34 (23.5)	18.3	598/2955 (20.2)	32.8
Common for both schools	0/34 (0)	—[a]	—	—
Gävleborg	12/34 (35.3)	27.5	1007/2955 (34.1)	36.6
Central Gävle	2/34 (5.9)	2.0[b]	77/2955 (2.6)	38.5
Bomhus	5/34 (14.7)	8.0[b]	372/2955 (12.6)	46.5
Common for both schools	5/34 (14.7)	17.5	558/2955 (18.9)	31.9
Total	34/34 (100)	74.0	2955/2955 (100)	39.9

[a]Not applicable.

[b]Excluding areas common for both schools in Gävleborg.

XSL•FO
RenderX

Proportions of Different Foods in Advertisements

UPFs

The analysis of the pictures of food advertisements taken by the researchers in all hotspot areas combined showed that 77.5% (2291/2955) of all food advertisements advertised UPFs. When comparing hotspot areas in Stockholm against hotspot areas in Gävleborg, no significant difference was found in proportion of UPFs being advertised (Stockholm: 1505/1948, 77.3% vs Gävleborg: 786/1007, 78.1%; χ^2_1=0.2; *P*=.62). However, a significant difference was observed in the proportion of UPFs advertised between the low- and high-SES areas in Stockholm, with a higher proportion in the high-SES area (Södertälje: 443/598, 74.1% vs Östermalm: 1063/1350, 78.7%; χ^2_1=5.0; *P*=.03).

Fruits, Vegetables, Berries, and Seafood

Overall, 20.8% (616/2955) of all food advertisements promoted FBVS. The proportion of FBVS advertisements was higher in Stockholm compared to in Gävleborg (Stockholm: 445/1948, 22.8% vs Gävleborg: 171/1007, 17%; χ^2_1=13.8; *P*<.001) and in the high-SES area in Stockholm compared to the low-SES area (Södertälje: 116/598, 19.4% vs Östermalm: 329/1350, 24.4%; χ^2_1=5.8; *P*=.02).

Price Promotions

Of all food advertisements in all areas combined, 23.6% (697/2955) included a price promotion. The proportion of advertisements including price promotions was slightly higher in Stockholm compared to Gävleborg (Stockholm: 490/1948, 25.2% vs Gävleborg: 207/1007, 20.6%; χ^2_1=7.8; *P*=.005). No difference in the proportion of price promotions was found when comparing the low- and high-SES areas in Stockholm (Södertälje: 164/598, 27.4% vs Östermalm: 326/1350, 24.1%; χ^2_1=2.4; *P*=.12).

Of all price promotions in all areas combined, 74.3% (518/697) advertised UPFs (compared to 1773/2258, 78.5% of advertisements without a price promotion; χ^2_1=5.4; *P*=.02). The proportion of price promotions that advertised UPFs was significantly higher in Stockholm compared to Gävleborg (Stockholm: 376/490, 76.7% vs Gävleborg: 142/207, 68.6%; χ^2_1=5.1; *P*=.03). There was no significant difference in the proportion of price promotions promoting UPFs when comparing the low- and high-SES areas of Stockholm (Södertälje: 128/164, 78% vs Östermalm: 248/326, 76.1%; χ^2_1=0.2; *P*=.63).

Moreover, of all price promotions in all areas combined, 20.4% (142/697) advertised FBVS (compared to 472/2258, 21% of advertisements without price promotions; χ^2_1=0.1; *P*=.73). There was a higher proportion of advertisements containing price promotions that advertised FBVS in Stockholm compared to in Gävleborg (Stockholm: 110/490, 22.4% vs Gävleborg: 32/207, 15.5%; χ^2_1=4.4; *P*=.04). A significant difference was also observed when comparing the proportion of price promotions advertising FBVS between the low- and high-SES areas of Stockholm, with a higher proportion found in the high-SES area (Södertälje: 23/164, 14% vs Östermalm: 87/326, 26.7%; χ^2_1=10.1; *P*=.002).

Interrater Agreement

"Moderate agreement" (κ=0.762) was observed between the 2 raters for the categorization of advertisements advertising UPF out of total food advertisements. "Moderate agreement" was also observed for the categorization of fruits, berries, and/or vegetables (κ=0.720), as well as for the categorization of fish and other seafood out of total foods (κ=0.775).

Discussion

Principal Findings

This study used a novel mHealth tool, comprising an app and a dashboard, to identify hotspot areas where children are exposed to food advertisements. It also examined the content of food advertisements within the identified areas.

Using the tool, a total of 34 hotspot areas were identified. These included transportation hubs, shopping streets and malls, grocery stores, and fast-food restaurants, both in proximity to schools and further away. More than three-quarters of the food advertisements in the hotspot areas promoted UPFs, such as junk food, ready-to-eat meals, sweet and salty snacks, and sweetened drinks. Fewer than one-quarter of the advertisements advertised FBVS. Price promotions were present in nearly one-quarter of advertisements, predominantly targeting UPFs.

Comparison With Prior Work

Data Collection and Analysis Methodology

Previous research has used child-centric approaches using mHealth technology to investigate children's exposure to food advertising. For instance, the Kids'Cam study [34] combined GPS tracking and wearable cameras to objectively capture children's exposure to food environments, providing detailed and comprehensive data on children's exposure to food marketing. While such approaches generate rich information, they are resource-intensive and might raise privacy and data management challenges. This study offers an alternative approach to investigating children's exposure to food advertising, using an app for child-led documentation that is both scalable and privacy-sensitive.

Analysis Outcomes

The dominance of food advertisements for UPFs aligns with findings from Sweden and other countries [15-20]. Given that children's diets remain high in sugar, fat, and salt and low in whole grains, fruits, and vegetables, the current marketing landscape raises concerns [35]. Several countries have introduced policies to restrict unhealthy food marketing in public spaces to limit children's exposure to marketing for unhealthy foods [36,37]. The results of this study suggest that similar measures may be relevant for Sweden.

A higher proportion of UPF advertisements was observed in the high-SES area of Stockholm compared to the low-SES area, contradicting findings from a previous study by our research

XSL•FO
RenderX

group [28]. The earlier study mapped food advertisements in only 1 location per neighborhood (Skärholmen vs Östermalm) and found a higher proportion of UPF advertisements in the low-SES area. In contrast, this study included multiple locations within each area, offering a more comprehensive assessment of advertising exposure. Although the difference was statistically significant (1063/1350, 78.7% vs 443/598, 74.1%; *P*=.03), the high levels in both areas in this study suggest that the primary public health concern is the overall prevalence of UPF marketing, rather than SES-based differences.

Similar to previous studies [15,28], advertisements for health-promoting foods such as FBVS were scarce in the outdoor food advertising landscape. Other categories of health-promoting foods, such as products with whole grains or healthy fats, were not included in this study due to their extremely low prevalence (<1% of the advertisement dataset).

This study also explored price promotions, which are known to increase consumer purchases and are more frequently used for unhealthy foods compared to healthy foods [38]. While this is a recognized marketing strategy, few studies have examined its use in outdoor food advertising. One study conducted in the United States found that 19% of food advertisements around schools included price promotions but did not specify the types of foods promoted [39].

In our study, a significantly higher share of price promotions targeted UPFs in Stockholm compared to Gävleborg. This difference likely reflects the distinct characteristics of the hotspot areas. Hotspot areas in Stockholm often included subway stations with convenience stores, which frequently have price promotions for UPFs, while hotspot areas in Gävleborg more commonly featured grocery stores that generally have price promotions for a broader range of products. Moreover, within Stockholm, price promotions for fruits and vegetables were more prevalent in the high-SES area than in the low-SES area. Stockholm also had a higher overall proportion of price promotions for FBVS compared to Gävleborg. These patterns may reflect differences in consumer demand or retail marketing strategies across neighborhoods and could have potential public health implications. However, the limited number of participants submitting photos, especially in the low-SES area of Stockholm, may affect the representativeness of the identified hotspot areas, and the results should therefore be interpreted with caution.

Strengths and Limitations

A key strength of this study was the use of a smartphone app and accompanying dashboard, which enabled the identification of areas where children were exposed to food advertisements based on their movement patterns. This approach facilitated the identification of areas that would not have been selected by researchers. For instance, children in Södertälje commuted to a shopping center beyond their immediate residential area, illustrating how children often move beyond their local neighborhoods (Multimedia Appendix 2). In Sweden, where public transportation is widely accessible, particularly in urban areas, such mobility among children is common [40]. This behavior-based method captures local context and population characteristics accurately, supporting improved mapping of children's exposure to food advertising.

In addition, using a smartphone app to collect data is a convenient and cost-effective approach for this target group, as nearly all adolescents in Sweden, and in other high-income countries, own a smartphone that they carry with them throughout the day. The app used in the study allows participants to document their everyday environments in real time, minimizing the need for additional equipment or the presence of researchers. However, the approach also depends on participants' motivation, engagement, and memory to take pictures, which can vary considerably between individuals. In-app notifications might encourage continued participation and improve compliance throughout the data collection period.

Given the pilot nature of the study, the participant sample was relatively small with 46 children from 4 schools contributing pictures. In 2 schools, hotspot areas were identified from only 5 to 7 participants. Additional (or different) hotspots might have emerged with a larger group of participants. However, in Östermalm, hotspots identified from the first 6 participants closely matched those observed after all 14 contributed, suggesting that small participant groups can still provide meaningful insights.

One limitation of researcher-led mapping of food advertisements was that this study did not account for the type or size of the advertisements. For example, a large billboard promoting a single product likely has a different impact than a small poster outside a grocery store displaying multiple items. Additionally, the cross-sectional design of the study did not capture potential weekly or seasonal variation in advertising exposure.

Future Research and Implications

In future research, technology-based tools for data collection could be further developed and expanded to provide a more detailed and personalized understanding of the food environment and individual behaviors. For instance, combining outdoor advertising data with information on digital marketing exposure, such as advertisements children encounter on social media, would offer a more comprehensive picture of overall food marketing exposure. This is particularly relevant, given that 86% of Swedish children aged 8 - 19 years use social media daily, and studies indicate that they are frequently exposed to food advertisements in these digital environments [41]. Furthermore, the app could be extended beyond food advertising to capture users' perception of other environmental factors, such as food availability (fast-food restaurants, grocery stores, convenience stores, etc) and accessibility (financial and geographic). The app could also be used to record dietary habits (at a group or individual level), such as meals, snacks, and beverage consumption. Incorporating these additional data types would enable a more holistic assessment of food environments and their influence on individual behavior. Additionally, privacy-preserving features implemented in the app, such as automatic face blurring before transmission, could further improve usability and user trust and boost engagement and long-term use.

In parallel, the dashboard could be adapted accordingly to provide additional heat maps for the new environmental factors and behaviors, as well as additional analytics for examining the associations between food environment and population

XSL•FO
RenderX

behaviors (eg, by visualizing clusters of different factors on the same map). Similarly, the dashboard could be further developed to present local evolvement over time, for example, comparing data for weekdays and weekends or tracking changes in food advertisements in any given area throughout the year and across different advertising campaigns (eg, it would be expected that different types of advertisements and offers would be present in periods like Halloween or Easter). Automating the identification and classification of advertisement types using machine learning techniques could potentially improve objectivity, accuracy, and efficiency. Applying analytics at the individual level, instead of aggregated, could enable the potential of the dashboard as a tool for personalized guidance. Finally, future research could investigate whether pictures taken by children could substitute those taken by researchers.

Conclusions

Using a novel, child-centered mHealth tool combining a smartphone app and a dashboard, this study identified local food advertising hotspots. Within the identified hotspot areas, 77.5% (2291/2955) of food advertisements promoted UPFs, while only 20.8% (616/2955) featured FBVS. Price promotions were found in 23.6% (697/2955) of advertisements, with the majority promoting UPFs. The proportion of UPF advertisements was high, and FBVS low, across all areas, regardless of city size or SES. These findings reveal a food marketing landscape misaligned with dietary guidelines, underscoring the need for policy action and further research.

Acknowledgments

The authors would like to thank all the teachers and principals who have been involved with the study. This study would not have been possible without the teachers who made time for us in their already busy schedules. The authors also sincerely want to thank the Multimedia Understanding Group, led by Professor Anastasios Delopoulos at Aristotle University of Thessaloniki, for providing the digital tools that were essential for data collection and analysis in this research. ChatGPT, GPT-4 version, from OpenAI (2023), was used to improve the clarity and grammar of parts of the manuscript. No content was generated that relates to the results, interpretation, or conclusions of the study.

Funding

This study was funded by FORTE—the Swedish Research Council for Health, Working Life and Welfare (2021-00518), and by the European Union's research programs (grants 2022-1-SE01-KA220-SCH-0000899 [FoodShift Pathways] and 101104618 [PREVENT—Empowering Childhood Health, Preventing Obesity: Go Healthy for a Bright Future]). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European funding agencies.

Data Availability

The nonpicture datasets generated or analyzed during this study are not publicly available because they may require guidance for proper use but are available from the corresponding author upon reasonable request. The picture datasets are not publicly available due to privacy concerns and in accordance with the ethical approval for this study (DNR 2023-01085-01).

Authors' Contributions

Conceptualization: II, SS, and PF
Data curation: IS, PF, and SS
Formal analysis: SS and PF
Funding acquisition: II
Investigation: SS and PF
Methodology: II and SS
Project administration: II, PF, and SS
Resources: II
Software: IS and II
Supervision: II and PF
Visualization: IS, PF, and SS
Validation: II and BL
Writing—original draft: SS and BL
Writing—review and editing: SS, II, BL, PF, and IS

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

List of ultraprocessed foods included in the study.
[DOCX File, 18 KB - mhealth_v14i1e70192_app1.docx]

Multimedia Appendix 2
Hotspot maps over Södertälje and Stockholm (Östermalm).
[PDF File, 409 KB - mhealth_v14i1e70192_app2.pdf]

References

1. Obesity and overweight. World Health Organization. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2024-09-20]
2. Övervikt och fetma bland skolbarn 11–15 år fortsätter öka. Folkhälsomyndigheten. 2020. URL: https://www.folkhalsomyndigheten.se/publikationer-och-material/publikationsarkiv/oe/overvikt-och-fetma-bland-skolbarn-1115-ar-fortsatter-oka/ [accessed 2024-09-20]
3. Smith JD, Fu E, Kobayashi MA. Prevention and management of childhood obesity and its psychological and health comorbidities. Annu Rev Clin Psychol 2020 May 7;16:351-378. [doi: 10.1146/annurev-clinpsy-100219-060201] [Medline: 32097572]
4. Sahoo K, Sahoo B, Choudhury AK, Sofi NY, Kumar R, Bhadoria AS. Childhood obesity: causes and consequences. J Family Med Prim Care 2015;4(2):187-192. [doi: 10.4103/2249-4863.154628] [Medline: 25949965]
5. Abaraca-Gómez L, Abdeen ZA, Hamid ZA, Abu-Rmeileh NM, Acosta-Cazares B, Acuin C. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. Lancet 2017 Dec 16;390(10113):2627-2642. [doi: 10.1016/S0140-6736(17)32129-3] [Medline: 29029897]
6. Umer A, Kelley GA, Cottrell LE, Giacobbi P, Innes KE, Lilly CL. Childhood obesity and adult cardiovascular disease risk factors: a systematic review with meta-analysis. BMC Public Health 2017 Aug 29;17(1):683. [doi: 10.1186/s12889-017-4691-z] [Medline: 28851330]
7. Myers CA. Impact of the neighborhood food environment on dietary intake and obesity: a review of the recent literature. Curr Diab Rep 2023 Dec;23(12):371-386. [doi: 10.1007/s11892-023-01529-9] [Medline: 38008848]
8. Swinburn BA, Sacks G, Hall KD, et al. The global obesity pandemic: shaped by global drivers and local environments. Lancet 2011 Aug;378(9793):804-814. [doi: 10.1016/S0140-6736(11)60813-1]
9. De Amicis R, Mambrini SP, Pellizzari M, et al. Ultra-processed foods and obesity and adiposity parameters among children and adolescents: a systematic review. Eur J Nutr 2022 Aug;61(5):2297-2311. [doi: 10.1007/s00394-022-02873-4] [Medline: 35322333]
10. de Araújo TP, de Moraes MM, Magalhães V, Afonso C, Santos C, Rodrigues SSP. Ultra-processed food availability and noncommunicable diseases: a systematic review. Int J Environ Res Public Health 2021 Jul 10;18(14):34299832. [doi: 10.3390/ijerph18147382] [Medline: 34299832]
11. Hall KD, Ayuketah A, Brychta R, et al. Ultra-processed diets cause excess calorie intake and weight gain: an inpatient randomized controlled trial of ad libitum food intake. Cell Metab 2019 Jul 2;30(1):67-77. [doi: 10.1016/j.cmet.2019.05.008] [Medline: 31105044]
12. Taneri PE, Wehrli F, Roa-Díaz ZM, et al. Association between ultra-processed food intake and all-cause mortality: a systematic review and meta-analysis. Am J Epidemiol 2022 Jun 27;191(7):1323-1335. [doi: 10.1093/aje/kwac039] [Medline: 35231930]
13. Pagliai G, Dinu M, Madarena MP, Bonaccio M, Iacoviello L, Sofi F. Consumption of ultra-processed foods and health status: a systematic review and meta-analysis. Br J Nutr 2021 Feb 14;125(3):308-318. [doi: 10.1017/S0007114520002688] [Medline: 32792031]
14. Healthy diet. World Health Organization. 2020. URL: https://www.who.int/news-room/fact-sheets/detail/healthy-diet [accessed 2025-11-02]
15. Finlay A, Robinson E, Jones A, et al. A scoping review of outdoor food marketing: exposure, power and impacts on eating behaviour and health. BMC Public Health 2022 Jul 27;22(1):1431. [doi: 10.1186/s12889-022-13784-8] [Medline: 35896996]
16. Trapp G, Hooper P, Thornton L, et al. Children's exposure to outdoor food advertising near primary and secondary schools in Australia. Health Promot J Austr 2022 Jul;33(3):642-648. [doi: 10.1002/hpja.532] [Medline: 34418222]
17. Amevinya GS, Vandevijvere S, Kelly B, et al. Advertising of unhealthy foods and beverages around primary and junior high schools in Ghana's most urbanized and populous region. Front Public Health 2022;10:917456. [doi: 10.3389/fpubh.2022.917456] [Medline: 36211683]
18. Martin-Payo R, González-Moradas MDR, Iturrate-Bobes J, Fernández-Sutil A, Cofiño R, Fernandez-Alvarez MDM. Mapping of outdoor food and beverage advertising around Spanish schools. Nutrients 2022 Jul 31;14(15):3167. [doi: 10.3390/nu14153167] [Medline: 35956343]

XSL•FO
RenderX

19. Dia OEW, Løvhaug AL, Rukundo PM, Torheim LE. Mapping of outdoor food and beverage advertising around primary and secondary schools in Kampala city, Uganda. BMC Public Health 2021 Apr 12;21(1):707. [doi: 10.1186/s12889-021-10661-8] [Medline: 33845809]
20. Velazquez CE, Daepp MIG, Black JL. Assessing exposure to food and beverage advertisements surrounding schools in Vancouver, BC. Health Place 2019 Jul;58:102066. [doi: 10.1016/j.healthplace.2018.12.007] [Medline: 30639201]
21. Wells G, Trapp G, Wickens N, Heritage B. Powerful promotions: an investigation of the teen-directed marketing power of outdoor food advertisements located near schools in Australia. Health Promot J Austr 2024 Jan;35(1):144-153. [doi: 10.1002/hpja.724] [Medline: 37012612]
22. Ruggles PR, Thomas JE, Poulos NS, Pasch KE. School-level socioeconomic status and nutrient content of outdoor food/beverage advertisements. Int J Environ Res Public Health 2023 Sep 7;20(18):6730. [doi: 10.3390/ijerph20186730] [Medline: 37754591]
23. Wentzel A, Eichinger M, Govender L, et al. The extent of outdoor food and beverage advertising surrounding primary and secondary schools in poorly resourced townships in Cape Town, South Africa. Health Place 2024 Nov;90:103371. [doi: 10.1016/j.healthplace.2024.103371] [Medline: 39546841]
24. Boyland E, McGale L, Maden M, et al. Association of food and nonalcoholic beverage marketing with children and adolescents' eating behaviors and health: a systematic review and meta-analysis. JAMA Pediatr 2022 Jul 1;176(7):e221037. [doi: 10.1001/jamapediatrics.2022.1037] [Medline: 35499839]
25. Russell SJ, Croker H, Viner RM. The effect of screen advertising on children's dietary intake: a systematic review and meta-analysis. Obes Rev 2019 Apr;20(4):554-568. [doi: 10.1111/obr.12812] [Medline: 30576057]
26. Boyland EJ, Nolan S, Kelly B, et al. Advertising as a cue to consume: a systematic review and meta-analysis of the effects of acute exposure to unhealthy food and nonalcoholic beverage advertising on intake in children and adults. Am J Clin Nutr 2016 Feb;103(2):519-533. [doi: 10.3945/ajcn.115.120022] [Medline: 26791177]
27. Villegas-Navas V, Montero-Simo MJ, Araque-Padilla RA. The effects of foods embedded in entertainment media on children's food choices and food intake: a systematic review and meta-analyses. Nutrients 2020 Mar 31;12(4):964. [doi: 10.3390/nu12040964] [Medline: 32244299]
28. Fagerberg P, Langlet B, Oravsky A, Sandborg J, Löf M, Ioakimidis I. Ultra-processed food advertisements dominate the food advertising landscape in two Stockholm areas with low vs high socioeconomic status. Is it time for regulatory action? BMC Public Health 2019 Dec 21;19(1):1717. [doi: 10.1186/s12889-019-8090-5] [Medline: 31864323]
29. Matmiljö i sverige—en kartläggande litteraturöversikt. Folkhälsomyndigheten. 2020. URL: https://www.folkhalsomyndigheten.se/publikationer-och-material/publikationsarkiv/m/matmiljo-i-sverige/ [accessed 2025-10-20]
30. Coro DG, Schirmer K, van Rhoda M, McQueen B, Morris C. An environmental scan of food and drink advertising on public transport around Adelaide schools. Health Promot J Austr 2025 Jan;36(1):e886. [doi: 10.1002/hpja.886] [Medline: 38840470]
31. Befolkningsstatistik per kvartal/halvår/helår efter region och kön. Statistikmyndigheten. 2025. URL: https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101G/BefforandrKvRLK/ [accessed 2026-03-16]
32. Områdesstatistik. Boverket. Segregationsbarometern. 2023. URL: https://segregationsbarometern.boverket.se/om-segregationsbarometern/ [accessed 2026-03-16]
33. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282. [Medline: 23092060]
34. Signal LN, Smith MB, Barr M, et al. Kids'Cam: an objective methodology to study the world in which children live. Am J Prev Med 2017 Sep;53(3):e89-e95. [doi: 10.1016/j.amepre.2017.02.016] [Medline: 28455122]
35. Lemming EW, Moraeus L, Sipinen JP, Lindroos AK. Riksmaten ungdom 2016-17—så äter ungdomar i sverige—del 1—livsmedelskonsumtion bland ungdomar i åk 5, åk 8 och åk 2 på gymnasiet. Livsmedelsverket. 2018. URL: https://www.livsmedelsverket.se/globalassets/publikationsdatabas/rapporter/2018/2018-nr-14-riksmatenungdom-huvudrapport_del-1-livsmedelskonsumtion.pdf [accessed 2024-09-18]
36. Chung A, Zorbas C, Riesenberg D, et al. Policies to restrict unhealthy food and beverage advertising in outdoor spaces and on publicly owned assets: a scoping review of the literature. Obes Rev 2022 Feb;23(2):e13386. [doi: 10.1111/obr.13386] [Medline: 34783421]
37. Forskrift om forbud mot markedsføring av visse næringsmidler særlig rettet mot barn. Lovdata. 2025. URL: https://lovdata.no/dokument/LTI/forskrift/2025-04-25-684 [accessed 2025-11-02]
38. Bennett R, Zorbas C, Huse O, et al. Prevalence of healthy and unhealthy food and beverage price promotions and their potential influence on shopper purchasing behaviour: a systematic review of the literature. Obes Rev 2020 Jan;21(1):e12948. [doi: 10.1111/obr.12948] [Medline: 31633289]
39. Herrera AL, Pasch KE. Targeting Hispanic adolescents with outdoor food & beverage advertising around schools. Ethn Health 2018 Aug;23(6):691-702. [doi: 10.1080/13557858.2017.1290217] [Medline: 28277028]
40. Andresson M, Granberg E, Aune K, Bork O. Ungdomar reser mer med kollektivtrafik än de äldre. Svensk Kollektivtrafik. 2025. URL: https://svenskkollektivtrafik.se/2025/05/ungdomar-reser-mer-med-kollektivtrafik-an-de-aldre/ [accessed 2025-11-13]

XSL•FO
RenderX

41. Food marketing exposure and power and their associations with food-related attitudes, beliefs and behaviours: a narrative review. World Health Organization. 2022. URL: https://www.who.int/publications/i/item/9789240041783 [accessed 2025-11-06]

Abbreviations

FBVS: fruits, berries, vegetables, and seafood
mHealth: mobile health
SES: socioeconomic status
UPF: ultraprocessed food

Edited by L Buis; submitted 30.Dec.2024; peer-reviewed by JDP Matos, LA O'Hagan, T Baranowski; accepted 18.Feb.2026; published 24.Mar.2026.

Please cite as:
Spolander S, Fagerberg P, Langlet B, Sarafis I, Ioakimidis I
Investigating Children's Exposure to Outdoor Food Marketing in 2 Swedish Cities Using a Smartphone App: Cross-Sectional Study
JMIR Mhealth Uhealth 2026;14:e70192
URL: https://mhealth.jmir.org/2026/1/e70192
doi:10.2196/70192

© Sofia Spolander, Petter Fagerberg, Billy Langlet, Ioannis Sarafis, Ioannis Ioakimidis. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 24.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Impact of Sunlight and Artificial Light at Night on Sleep Stages: Evidence From a 7-Day Sensor-Based Observational Study

Andrea Montanari[1], PhD; Li Min Wang[2], PhD; Amit Birenboim[2,3], PhD; Basile Chaix[1], PhD

[1]Sorbonne Université, INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique IPLESP, Paris, France

[2]Department of Geography, The Hebrew University of Jerusalem, Mt Scopus, Jerusalem, Israel

[3]The Center for Urban Innovation, The Hebrew University of Jerusalem, Mt Scopus, Jerusalem, Israel

Corresponding Author:
Andrea Montanari, PhD
Sorbonne Université, INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique IPLESP
27 rue Chaligny
Paris, 75012
France
Phone: 33 144 738 445
Email: andrea.montanari@iplesp.upmc.fr

Abstract

Background: Exposure to circadian entrainers, such as sunlight, positively impacts sleep architecture, while exposure before bedtime to circadian disruptors, such as artificial light and smartphone use, can negatively affect sleep. However, real-world evidence from longitudinal observational studies that simultaneously capture these factors alongside electroencephalography-derived sleep stages remains limited.

Objective: This study aimed to investigate the effects of specific environmental and behavioral factors on sleep metrics and architecture by using sensor-based measurements over 7 consecutive days. Specifically, it examined day-to-day associations between (1) daytime sunlight exposure and (2) prebedtime artificial light exposure and smartphone use with selected sleep outcomes on the following night.

Methods: A total of 21 participants from the Jerusalem metropolitan area were monitored continuously using the Dreem wearable electroencephalography for sleep staging, HOBO data loggers for light exposure, the wGT3X+ triaxial accelerometer for physical activity, and a dedicated mobile app to record smartphone usage. Sleep outcomes included total sleep time (TST), sleep onset latency (SOL), and the proportions of light sleep (N1) and deep sleep (N3). Sunlight exposure was defined as the number of hours above 1000 lux during daytime, and artificial light and smartphone use before bedtime were quantified as the duration of exposure accumulated in the 2 hours preceding sleep onset. Linear mixed-effects models with a random intercept at the individual level estimated the associations between these exposures and next-night sleep outcomes, adjusting for step count and other individual covariates.

Results: The average TST was 420 (SD 85) minutes, and SOL averaged 17.6 (SD 18) minutes. Light sleep (N1) represented 6.6% (SD 2.1%) of sleep, and deep sleep (N3) accounted for 20.1% (SD 7.6%). Each additional hour of daytime sunlight exposure was associated with an increase of 10.67 (95% CI 0.6-20.7) minutes in TST the following night and with a 0.3 (95% CI –0.6 to –0.0) percentage-point decrease in light sleep (N1) percentage. No associations were found between evening artificial light exposure and sleep outcomes, while each minute of smartphone use before bedtime was linked to an increase in SOL of 0.2 (95% CI 0.0-0.4) minutes.

Conclusions: These findings emphasize the importance of daylight exposure for circadian alignment and the potential sleep-disruptive effects of evening digital engagement. This study demonstrates the feasibility and value of integrating wearable electroencephalography and environmental and behavioral sensors in naturalistic settings to uncover behavioral and environmental correlates of sleep architecture.

(JMIR Mhealth Uhealth 2026;14:e75898) doi:10.2196/75898

XSL•FO
RenderX

KEYWORDS

circadian rhythm; electroencephalogram; light exposure; sleep; smartphone usage; wearable sensors

Introduction

Sleep Dynamics

Sleep is a fundamental biological process that is crucial for maintaining overall health and well-being, playing a vital role in cognitive function, emotional regulation, metabolism, and immune system performance [1]. Insufficient or poor-quality sleep has been associated with a wide range of health issues, including cardiovascular and metabolic dysfunction, mental health disorders, cognitive decline, and increased all-cause mortality [2-5].

Sleep is a dynamic process made up of distinct stages, which collectively determine its architecture. Sleep is generally divided into non–rapid eye movement sleep, which includes light (stage 1 sleep [N1] and stage 2 sleep [N2]) and deep sleep (stage 3 sleep [N3]), and rapid eye movement (REM) sleep [6]. Each sleep stage serves critical functions: light sleep allows transitions from wakefulness into sleep and subsequent sleep stages, deep sleep supports physiological restoration and recovery, and REM sleep contributes to cognitive processing and emotional regulation [7,8]. Sleep occurs in ultradian cycles that alternate between non–rapid eye movement and REM sleep approximately every 90 minutes [9]. The first cycles are characterized by a higher prevalence of deep sleep, while REM sleep becomes increasingly prominent in the later cycles [6]. These cyclical processes are regulated by circadian rhythms of the body's internal clock, which synchronizes sleep timing with environmental cues such as light and darkness [10].

Circadian Entrainment and Disruption

Circadian entrainment refers to the alignment of the internal circadian clock with external environmental cues, such as the light-dark cycle of the day. Sunlight is the most powerful entrainer, as it modulates the release of melatonin, a hormone secreted by the pineal gland that regulates the timing of sleep and wakefulness [10]. During daylight, sunlight suppresses melatonin secretion, promoting wakefulness and alertness; as light diminishes in the evening, melatonin levels rise, facilitating sleep onset [11]. Adequate daytime sunlight exposure therefore strengthens circadian entrainment, leading to earlier sleep onset and longer total sleep time (TST) [12]. Consistent evidence shows that more time spent outdoors, exposed to natural light, is associated with a stronger circadian rhythm and better mood and sleep outcomes [13,14].

Conversely, circadian disruption occurs when external factors or individual behaviors cause misalignment between endogenous circadian rhythms and the external environment [15]. This misalignment may also disrupt the synchronization between the body's central clock, located in the suprachiasmatic nucleus, and peripheral clocks in tissues and organs, potentially contributing to dysregulation of physiological and behavioral rhythms such as hormone release, metabolism, and the sleep-wake cycle [10,15]. A key contributor to circadian disruption is artificial light at night, particularly the blue light emitted from digital screens and smartphones. This stimulation suppresses melatonin production, which normally supports circadian alignment and promotes sleep propensity, thereby delaying sleep onset and altering sleep architecture [11,16].

Behavioral circadian modulators further influence the alignment or misalignment of circadian rhythms. In addition to blue light exposure, smartphone use before bedtime also provides stimulating content (eg, social media) that may increase cognitive and physiological arousal, activating the hypothalamic-pituitary-adrenal axis and elevating cortisol levels, ultimately interfering with sleep initiation and maintenance [17]. However, there is limited and mixed evidence regarding how and to what extent smartphone use might affect sleep stages, and thus, the topic warrants further investigation [18].

Engaging in regular physical activity, particularly earlier in the day, can promote circadian entrainment by reinforcing the distinction between active daytime and restful nighttime periods [19-21]. Exercise has been associated with improved sleep onset and increased proportions of deep sleep (N3), whereas intense physical activity close to bedtime may elevate core body temperature and arousal levels, potentially delaying sleep onset and disrupting sleep architecture by reducing REM sleep and increasing light sleep [21,22]. Other health-related behaviors, such as napping, evening food intake, and the consumption of alcohol, caffeine, or medications, can also affect sleep and circadian rhythms by altering metabolism, arousal and sleep pressure, delaying sleep onset, and decreasing deep and REM sleep [23-25].

Gaps in the Literature and Rationale of the Study

Despite the recognized role of environmental and behavioral factors in sleep and circadian regulation, there are gaps in the literature concerning their combined effects on sleep architecture. Regarding the variables examined in the literature, previous studies have mainly focused on either sunlight exposure or artificial light at night in isolation, without considering the additional influence of other behavioral modulators such as smartphone use and physical activity [13,14,26,27]. For instance, time spent outdoors often correlates with higher physical activity, which also impacts sleep [28,29]. Controlling for physical activity is therefore essential when assessing the effects of sunlight exposure on sleep outcomes to avoid confounding effects.

Regarding sleep and circadian health, an increasing number of studies are adopting sensor-based methods to measure sleep and physical activity in both observational studies and clinical trials [30,31]. However, the majority of these studies rely on accelerometry to assess sleep outcomes, which, while useful for estimating general sleep metrics (eg, TST and sleep efficiency [SE]), lacks the ability to provide detailed insights into its architecture and sleep stages [32-36]. There is a scarcity of research using wearable electroencephalography devices in real-life settings outside controlled experimental environments, which offer more precise measurements of sleep stages compared to other sensors [32,37,38]. Finally, regarding study

XSL•FO
RenderX

designs, only a minority of these studies have used longitudinal frameworks with repeated measures to capture the day-to-day variability in environmental exposures, behaviors, and sleep outcomes over multiple days of tracking [39,40]. In contrast, the majority of studies have adopted cross-sectional designs, limiting their ability to move beyond mere associations and understand causal relationships through assessment of within-person processes [39,40].

Objectives

Our study addresses gaps in the literature by conducting a 7-day sensor-based investigation assessing sunlight exposure, artificial light at night, and behavioral circadian modulators. Through continuous light monitoring, smartphone use tracking, and physical activity measurement, we aim to comprehensively evaluate the effects of circadian entrainment and disruption on electroencephalography-based sleep metrics. This research not only fills a critical gap in the existing literature but also has practical implications for developing interventions and guidelines to enhance sleep health through the management of light exposure and sleep-hygiene behavioral practices.

In this study, we selected a priori TST, sleep onset latency (SOL), light sleep (N1), and deep sleep (N3) as main sleep outcomes, based on their theoretical relevance to sleep and circadian health. TST is the most used sleep measure in the literature and reflects overall sleep duration [38]. SOL measures the time taken to fall asleep, reflecting the ease of initiating sleep and showing sensitivity to presleep behaviors that may delay sleep onset [41]. N1 marks the transition from wakefulness to deeper sleep stages in each cycle and serves as an indirect marker of sleep fragmentation and efficiency [42]. N3, the most restorative sleep stage for physical health, responds strongly to daily activity and circadian alignment, making it valuable for evaluating how behavioral factors affect sleep [19,42].

First, we aim to quantify the effect of sunlight exposure during the day, as a proxy for circadian entrainment, on sleep. Second, we seek to measure the impact of potential circadian disruptors, such as exposure before bedtime to bright artificial light and smartphone use, on sleep. We hypothesize that (1) longer sunlight exposure will be associated with longer TST, a lower percentage of light sleep (N1), and a higher percentage of deep sleep (N3) and (2) exposure to bright artificial light and smartphone use close to bedtime will be associated with shorter TST, increased SOL, a higher proportion of light sleep (N1), and a lower percentage of deep sleep (N3).

Methods

Participants and Study Design

Our study included 21 participants from the Jerusalem metropolitan area (Israel) and was conducted from May 2023 to July 2024. Participants were monitored over a 7-day period using wearable sensors and smartphone apps.

Given the small sample size, a quota sampling approach was adopted to recruit participants for this study, allowing for diversity and representation across key demographic groups (ie, household income, age, and gender) without requiring strict proportionality to the broader population.

Recruitment was conducted in 2 stages: first, an Israeli panel company contacted individuals from their preestablished pool of volunteers who had previously signed up for scientific studies. The panel company confirmed participants' willingness to participate and screened them for eligibility through a questionnaire based on the study's recruitment criteria. The main inclusion criteria required participants to be aged between 18 and 65 years and to reside in the metropolitan area of Jerusalem. Exclusion criteria included engaging in any form of shift work, having mobility limitations, significant health conditions, or a clinical diagnosis of sleep or mental health disorders. In the second stage, a research assistant followed up with eligible candidates by phone to explain the study in greater detail, further inviting them to participate and initiating the scheduling of data collection.

Following initial sampling, eligible participants were onboarded through a home visit conducted by a research assistant, marking the start of the data collection period. During this visit, the research assistant administered an entry online questionnaire collecting key demographic information and explained the study protocol, emphasizing the significance of their involvement in studying urban health. Participants received an information sheet outlining the sensor charging and wearing schedule, along with a contact number for questions and support.

Ethical Considerations

The study received ethical approval from The Hebrew University of Jerusalem's Committee for the Use of Human Subjects in Research (approval number 22022023; February 2023). All participants provided informed consent prior to participation and could withdraw at any time without consequence. The study adhered to the General Data Protection Regulation and institutional ethical standards for human research, ensuring compliance with data storage, sharing, and confidentiality requirements. All collected data were pseudonymized before analysis, and identifiable information was stored separately in secure, password-protected databases. Upon successful completion of the 7-day protocol, participants were compensated with ILS 300 (US $104) as an acknowledgement of their cooperation and motivation to adhere to the study.

Sensor-Based and Mobile Measures

Sleep Assessment

Each night, participants' TST, SE, SOL, wake after sleep onset, and the duration of sleep stages (N1, N2, N3, and REM) were assessed using the Dreem Headband, a wearable electroencephalography device (Figure 1). The headband has been validated against polysomnography, the gold standard for sleep measurement, demonstrating high accuracy in sleep staging when compared to manual scoring by sleep experts [32]. This nonintrusive device provides reliable sleep assessment in real-world, home settings, providing an advantage over actigraphy devices, which are unable to differentiate sleep stages accurately [32,35,36].

XSL•FO
RenderX

Figure 1. Study Sensors: (1) Dreem 3 Headband (Beacon Biosignals); (2) wGT3X+ triaxial accelerometer (ActiGraph); (3) HOBO Data Logger (Onset Computer Corporation).



Light Exposure

Participants wore a HOBO Data Logger daily, attached to the collar of their clothing with a safety pin. This positioning was intended to approximate light exposure at the level of the eyes, where light signals are transmitted from the retina to the suprachiasmatic nucleus, the central circadian pacemaker [33]. The timing, duration, and intensity of light exposure were used to assess participants' daily circadian entrainment to sunlight and potential disruption from artificial light at night. Based on previous research, daytime sunlight exposure was operationalized as the total duration (in hours) of exposure to light intensity above 1000 lux across the day [43]. Before bedtime, artificial light exposure was defined as the total duration (in minutes) of exposure to light intensity above 30 lux across varying time windows in the 2 hours preceding sleep onset.

Physical Activity

Participants carried a wGT3X+ triaxial accelerometer on a waist belt on the right hip every day during waking hours. The accelerometer was used to assess various physical activity metrics, including step count (in units of 1000 steps), metabolic expenditure, time spent in moderate to vigorous physical activity, light physical activity, and sedentary time. Nonwear periods were identified and excluded using the Choi algorithm [44]. Physical activity metrics were calculated using established criteria, such as the Freedson Adult VM3 cut-points (2011) and the Crouter 2R regression equation [45,46]. After preliminary analyses, we selected step count over the other physical activity metrics for its greater explanatory power in capturing sleep outcome variability.

Smartphone Usage Tracking App

A designated app installed on participants' smartphones tracked the total time spent using the smartphone and the time spent using apps for the whole data collection period. The app was developed at the Department of Geography and the Center for Urban Innovation at the Hebrew University of Jerusalem. For analysis, apps were classified into categories (eg, social media, productivity, and entertainment), and the original app names were removed to ensure participant privacy. From these reports, we computed for each day the total number of minutes spent using each app category. However, preliminary analyses revealed no meaningful differences between total smartphone use and the use of social media apps in explaining sleep outcomes variability. Therefore, for our main analyses, we used total smartphone use, as it encompasses all types of smartphone activity, offering a more inclusive measure.

Statistical Methods

We used linear mixed-effects models with a random intercept at the individual level to examine the effects of circadian entrainment and disruption on sleep outcomes. This approach accounted for the nested hierarchical structure of the data (successive nights nested within participants) and controlled for between-subject variability in baseline sleep patterns.

The outcomes modeled were TST, percentages of N1 and N3, and SOL. To test the first hypothesis, that higher daytime sunlight exposure is associated with better sleep outcomes, we used daily sunlight exposure (measured in hours) as the primary predictor, with daily step count as a covariate to control for potential confounding from physical activity. To test the second hypothesis, we examined whether exposure to artificial light and smartphone use before bedtime negatively affect sleep outcomes. The primary predictors were artificial light exposure (measured as minutes of exposure above 30 lux [ie, dim light], a physiologically relevant threshold used in previous protocols [47]) and smartphone use (in minutes) in the 2 hours before bedtime. Step count in the same time window was included as a covariate to control for potential confounding from physical activity. Before selecting the 2-hour window for artificial light, we conducted a sensitivity analysis across different time windows before bedtime (5, 10, 15, 20, 25, 30, 60, and 120 minutes) to assess potential differences in their associations with sleep outcomes. As no meaningful differences were detected across these intervals, we proceeded with the most inclusive duration of 2 hours before bedtime to capture potential cumulative effects and maximize variability.

XSL•FO
RenderX

For covariate selection, we initially considered a broad set of potential confounders, refining the final model based on preliminary analyses (Table 1). The final selected covariates for all models included age group (25-29 years and 30-65 years), gender (female and male), household income per unit of consumption (low: ILS ≤5550 [US $1900] and high: ILS >5550 [US $1900]), weekday vs weekend, and season (spring, summer, autumn, and winter) to account for possible seasonal influences on sleep and light exposure.

Table 1. Participants' characteristics[a].

Characteristics	Values, n (%)
Gender	
Female	14 (67)
Male	7 (33)
Age (years)	
25-29	10 (48)
30-65	11 (52)
Education	
University	9 (43)
Professional	4 (19)
High school	8 (38)
Employment	
Unemployed	3 (14)
Employed	18 (86)
Household income (ILS 1=US $0.34)	
High (>7800)	4 (19)
Medium (5551-7800)	5 (24)
Low (≤5550)	12 (57)
Children	
No	13 (67)
Yes	7 (33)
Bed partner	
No	12 (62)
Yes	8 (38)
Citizenship	
Non-Israeli citizen	5 (24)
Israeli citizen	16 (76)
BMI (kg/m^2)	
Normal	11 (52)
Overweight	10 (48)
Neighborhood	
Suburban area	12 (67)
City center	7 (33)

[a]N=21.

Due to constraints in data availability, we tested separate models for our 2 hypotheses rather than combining all the predictors into one comprehensive model. Including all predictors in a single model would have reduced the sample size, as observations with missing data on any predictor would have been excluded, affecting model power and stability. By analyzing these factors in independent models, we maximized the use of available data to evaluate each predictor's relationship with sleep outcomes. No adjustment for multiple comparisons was performed, as models were specified a priori and focused

XSL•FO
RenderX

on a limited set of exposures and outcomes. Findings are therefore interpreted cautiously in light of the modest sample size and correlated predictors and outcomes. Statistical analyses were conducted using the R software (version 4.0.2; R Core Team; 2017).

Results

Descriptive Statistics

Of the 21 participants, 14 (67%) were female, and 11 (52%) were aged between 30 and 65 years; 16 (76%) were Israeli citizens, 9 (43%) had university education, and 12 (57%) reported a lower household income (ILS ≤5550 [US $1900]). See Table 1 for additional details.

The average TST was 420 (SD 85.4) minutes per night, ranging from 120 to 667.5 minutes across specific nights among participants. SOL averaged 17.6 (SD 18) minutes, with a range of 0.5-99 minutes across specific nights. N1 sleep accounted, on average, for 6.6% (SD 2.1%) of TST, ranging from 2.6% to 13.1%. N3 sleep reached, on average, 20.1% (SD 7.6%) of TST, with a range of 5.2%-51.1% across specific nights. See Table 2 for more sleep outcomes.

Table 2. Descriptive statistics on sleep outcomes.

Outcome	Mean (SD)	Range
TST[a] (min)	420 (85.4)	120-667.5
WASO[b] (min)	29.1 (25)	1-143
Sleep efficiency (%)	89.4 (6.7)	66.2-97.9
Sleep onset latency (min)	17.6 (18)	0.5-99
N1[c] (%)	6.6 (2.1)	2.6-13.1
N2[d] (%)	48 (7.6)	22.1-65.6
N3[e] (%)	20.1 (7.6)	5.2-51.1
REM[f] (%)	25.4 (7.9)	1.4-57.6

[a]TST: total sleep time.

[b]WASO: wake after sleep onset.

[c]N1: stage 1 sleep.

[d]N2: stage 2 sleep.

[e]N3: stage 3 sleep.

[f]REM: rapid eye movement sleep.

Participants spent an average of 1.2 (SD 1.6) hours exposed to sunlight, with a range from 0 to 8.3 hours across specific days across participants. Time spent using smartphone apps was 169 (SD 178) minutes per day on average, with social apps accounting for 44.0% (SD 21.9%) of that time. Daily step count averaged 6033 (SD 3779), ranging from 57 to 15,516 steps across specific days. See Table 3 for further details.

Table 3. Descriptive statistics on the main predictors.

Predictor	Mean (SD)	Range
Daily sunlight[a] (hours)	1.2 (1.6)	0-8.3
Before bedtime artificial light[b] (minutes)	20.9 (33.7)	0-120
Daily smartphone use[c] (minutes)	169.7 (176.9)	0-661.1
Before bedtime smartphone use[d] (minutes)	20.3 (25.2)	0-105.7
Daily step count[e] (1000 steps)	6.3 (3.9)	0.1-15.5
Before bedtime step count[f] (1000 steps)	0.6 (0.8)	0-3.8

[a]Daily sunlight: time spent exposed to light above 1000 lux during daytime hours (hours).

[b]Before bedtime artificial light: time spent exposed to light above 30 lux in the 2 hours preceding sleep onset (minutes).

[c]Daily smartphone use: total daily duration of smartphone activity (minutes).

[d]Before bedtime smartphone use: total duration of smartphone activity in the 2 hours preceding sleep onset (minutes).

[e]Daily step count: number of steps across the day (expressed in thousands).

[f]Before bedtime step count: number of steps in the 2 hours preceding sleep onset (expressed in thousands).

XSL•FO
RenderX

Sunlight Exposure and Sleep

In this study, the association between daily sunlight exposure (in hours) and key sleep outcomes the following night was analyzed, as shown in Table 4. Each additional hour of sunlight exposure was associated with an estimated increase of 10.7 (95% CI 0.6-20.7) minutes in TST. In terms of sleep architecture, sunlight exposure was associated with a decrease in the percentage of N1 sleep by 0.3 (95% CI –0.6 to –0.0) percentage points, while no association with N3 sleep was detected.

Table 4. Associations between sunlight exposure and sleep outcomes (116 observations; 21 participants).

	TST[a] (min)	N1[b] (%)	N3[c] (%)
Predictors, estimate[d] (95% CI)			
Intercept	410.58 (335.15 to 486.00)	6.41 (3.77 to 9.05)	18.14 (9.39 to 26.88)
Daily sunlight[e]	10.67 (0.64 to 20.70)	–0.29 (–0.55 to –0.03)	–0.45 (–1.49 to 0.59)
Age (30-65 years)	–53.43 (–121.57 to 14.70)	–0.79 (–3.20 to 1.61)	0.10 (–7.84 to 8.03)
Gender (male)	–28.24 (–97.45 to 40.96)	0.81 (–1.64 to 3.26)	0.68 (–7.39 to 8.75)
Season (vs autumn)			
Spring	23.35 (–59.25 to 105.94)	–1.05 (–3.86 to 1.76)	1.89 (–7.59 to 11.37)
Summer	62.64 (–9.12 to 134.40)	–0.23 (–2.76 to 2.31)	0.93 (–7.43 to 9.30)
Winter	70.37 (–31.98 to 172.72)	–0.42 (–4.02 to 3.18)	–1.34 (–13.24 to 10.56)
Weekend (vs weekday)	–2.98 (–34.47 to 28.52)	0.05 (–0.73 to 0.82)	1.71 (–1.47 to 4.89)
Household income (high)[f]	7.83 (–50.34 to 66.01)	1.00 (–0.98 to 2.98)	3.17 (–3.51 to 9.85)
Bed partners (yes)	–33.41 (–101.64 to 34.81)	1.47 (–0.89 to 3.83)	–0.49 (–8.37 to 7.38)
Other information			
σ^{2g}	4457.95	2.66	45.18
τ00 ID[h]	1743.20	2.72	26.48
ICC[i]	0.28	0.51	0.37
Marginal R^2/conditional R^2	0.11/0.36	0.11/0.56	0.06/0.41

[a]TST: total sleep time.

[b]N1: stage 1 sleep.

[c]N3: stage 3 sleep.

[d]Estimate of effect size.

[e]Daily sunlight: hours of exposure to light above 1000 lux.

[f]Household income (ILS): high (>5551); low (≤5550); ILS 1=US $0.34.

[g]σ^2: residual variance.

[h]τ00 ID: variance between individuals.

[i]ICC: intraclass correlation coefficient (variance attributed to individual differences).

A separate model was estimated to adjust for physical activity, measured as daily step count (per 1000 steps). This adjustment reduced the analytic sample by 40 observations, including all data from 2 participants, warranting separate estimation. In this model, the association between sunlight exposure and TST was unchanged (each additional hour of sunlight being associated with an increase of 11.6 (95% CI –0.5 to 23.8) minutes in TST, although the CI included 0. In contrast, higher daily step count was associated with longer TST, with an estimated increase of 6.1 (95% CI 1.0-11.1) minutes per additional 1000 steps. Further details are provided in Multimedia Appendix 1.

Circadian Disruptors and Sleep

In examining the impact of circadian disruption on sleep outcomes, we evaluated the associations of artificial light exposure, step count, and smartphone use recorded in the 2 hours prior to bedtime with key sleep outcomes the following night, as shown in Table 5. While no associations were observed between artificial light exposure or step count and TST, N1, or N3 sleep stages, smartphone use showed an association with SOL, where each additional minute of smartphone use before bedtime was associated with an increase in SOL by 0.2 (95% CI 0.0-0.4) minutes.

XSL•FO
RenderX

Table 5. Associations of artificial light, step count, and smartphone use before bedtime with sleep outcomes (75 observations; 18 participants).

	TST[a] (min)	SOL[b]	N1[c] (%)	N3[d] (%)
Predictors, estimate[e] (95% CI)				
Intercept	348.06 (235.39 to 460.72)	32.04 (8.19 to 55.88)	4.54 (1.46 to 7.63)	20.60 (8.80 to 32.41)
Before bedtime smartphone use[f]	0.15 (–0.74 to 1.04)	0.20 (0.00 to 0.39)	0.01 (–0.01 to 0.03)	–0.04 (–0.12 to 0.03)
Before bedtime artificial light[g]	0.26 (–0.69 to 1.21)	0.02 (–0.19 to 0.23)	0.01 (–0.01 to 0.04)	–0.00 (–0.09 to 0.08)
Before bedtime step count[h]	–0.01 (–0.03 to 0.02)	0.00 (–0.00 to 0.01)	0.00 (–0.00 to 0.00)	0.00 (–0.00 to 0.00)
Age (30-65 years)	–49.53 (–139.82 to 40.76)	2.64 (–16.41 to 21.70)	–0.21 (–2.70 to 2.28)	0.82 (–8.76 to 10.40)
Gender (male)	4.57 (–84.54 to 93.67)	–14.97 (–33.74 to 3.80)	1.06 (–1.41 to 3.52)	–0.84 (–10.37 to 8.69)
Season (vs autumn)				
Spring	137.13 (26.90 to 247.35)	–12.68 (–36.09 to 10.73)	–0.68 (–3.67 to 2.32)	–0.03 (–11.36 to 11.31)
Summer	108.31 (14.69 to 201.93)	–15.03 (–34.84 to 4.78)	0.31 (–2.25 to 2.88)	–1.28 (–11.10 to 8.54)
Winter	136.27 (9.96 to 262.57)	6.68 (–19.99 to 33.36)	–0.23 (–3.71 to 3.24)	–0.72 (–14.08 to 12.64)
Weekend (vs weekday)	–19.78 (–67.62 to 28.05)	7.03 (–3.58 to 17.65)	0.45 (–0.73 to 1.63)	–0.44 (–4.38 to 3.50)
Household income (high)[i]	–12.35 (–84.25 to 59.56)	–0.66 (–15.91 to 14.58)	1.40 (–0.56 to 3.36)	0.75 (–6.71 to 8.21)
Bed partners (yes)	–17.25 (–104.18 to 69.67)	–21.44 (–39.75 to –3.13)	2.35 (–0.05 to 4.75)	–3.96 (–13.25 to 5.33)
Other Information				
σ^{2}[j]	5045.09	249.22	3.05	33.68
τ00 ID[k]	1933.94	79.14	1.69	28.56
ICC[l]	0.28	0.24	0.36	0.46
Marginal R^2/conditional R^2	0.18/0.41	0.36/0.50	0.19/0.48	0.09/0.51

[a]TST: total sleep time.

[b]SOL: sleep onset latency.

[c]N1: stage 1 sleep.

[d]N3: stage 3 sleep.

[e]Estimate of effect size.

[f]Before bedtime smartphone use: minutes of smartphone use in the 120 minutes before sleep time.

[g]Before bedtime artificial light: minutes of exposure to light above 30 lux in the 120 minutes before sleep time.

[h]Before bedtime step count: number of steps in the 120 minutes before sleep time (expressed in thousands).

[i]Household income (ILS): high (>5551); low (≤5550); ILS 1=US $0.34.

[j]σ^2: Residual variance.

[k]τ00 ID: variance between individuals.

[l]ICC: intraclass correlation coefficient (variance attributed to individual differences).

Discussion

Principal Findings

Our findings support the first hypothesis that higher daytime exposure to sunlight is associated with increased TST and a reduction in light sleep (N1). Specifically, each additional hour of sunlight exposure was linked to an increase of approximately 10.7 (95% CI 0.6-20.7) minutes in TST and a decrease of 0.3 (95% CI –0.6 to –0.0) percentage points in N1 sleep. These results align with previous research suggesting that natural light exposure strengthens circadian entrainment, leading to improved sleep duration and quality [12-14]. Similarly, the potential circadian entrainment from increased sunlight exposure might have contributed to reduced N1, suggesting a quicker transition from wakefulness into deeper sleep stages [42,48]. This reduction in N1 could reflect enhanced circadian alignment, promoting a smoother progression into the other sleep stages. However, we did not observe an association between sunlight exposure and deep sleep (N3), which contrasts with our initial hypothesis. While some studies have reported that natural light exposure can enhance deep sleep [49], our results suggest that the relationship may be more nuanced. It is possible that factors such as individual differences in circadian sensitivity, genetic predispositions, or, most importantly, the timing of sunlight exposure might play a role [49-51]. The effect of sunlight

XSL•FO
RenderX

exposure on N3 might be more pronounced when it occurs shortly after wake-up time, when the sun's blue-enriched morning light exerts stronger phase-advancing effects on circadian entrainment [51,52]. Using a daily cumulative indicator of sunlight exposure could average out these timing-specific effects, potentially obscuring its impact on deep sleep, which should be further examined in future sensor-based research.

After controlling for daily step count, the point estimate of the association between sunlight exposure and TST was even slightly increased, suggesting the absence of substantial confounding by physical activity. The CI of the association included 0 in this last model, likely due to reduced statistical power from the smaller sample of 76 observations with both sunlight and physical activity data, compared to 116 in the previous model. Increased physical activity during the day might have extended TST by promoting a greater need for physiological recovery, enhancing overall sleep drive [53]. On the other hand, the independent association between sunlight exposure and a decrease in N1 sleep remained fully apparent in the model controlling for step count with fewer observations, highlighting the unique role of natural light in regulating sleep architecture beyond physical activity levels. This finding is noteworthy because it isolates the effect of sunlight from major confounders like physical activity, which is often correlated with time spent outdoors [28,29]. Additionally, this pattern suggests that physical activity may not mediate the relationship between sunlight exposure and improved sleep architecture, although a formal mediation analysis would be warranted to confirm this interpretation and to clarify the interplay between sunlight exposure, physical activity, and sleep. Remarkably, daily step count, used as a proxy for physical activity, was not associated with N1 or N3 percentages, despite evidence that higher levels of physical activity generally enhance SE and promote more restorative sleep with reduced light sleep and increased deep sleep [19,21,53].

The positive associations between daytime sunlight exposure, physical activity, and better sleep outcomes underscore the importance of natural light and movement for supporting circadian alignment and sleep health. Encouraging daily exposure to natural light and regular physical activity, individually or in combination, could be an effective strategy for improving sleep quality across the general population. These findings support public health recommendations encouraging time outdoors and physical activity to enhance sleep and circadian health, ultimately fostering overall well-being.

Contrary to our second hypothesis, we did not find any association between exposure to artificial light before bedtime and alterations in TST, SOL, or sleep architecture. This finding diverges from prior studies that have reported negative impacts of evening exposure to artificial light on sleep onset, duration, and stages with increases in light sleep and reductions in deep sleep [16,26]. This discrepancy may be due to the relatively low intensity or limited duration of artificial light exposure in our sample, which may have been insufficient to affect sleep architecture. Additionally, individual differences in light sensitivity and circadian preference may influence susceptibility to artificial light's effects on sleep [54]. Furthermore, participants may have exhibited overall good sleep hygiene, with the cumulative benefits of various positive behaviors potentially offsetting any negative impact of artificial light on sleep [55]. Finally, it is also possible that the light sensor was incorrectly worn by some of the participants during the hours just before sleep time.

Regarding smartphone use, while no associations were observed with TST, N1, or N3 sleep stages, an association with increased SOL was detected, indicating that smartphone use before bedtime may delay sleep onset. Since our model controlled for artificial light exposure, this delay likely stems not from light emitted by screens but rather from the arousal-increasing and mentally engaging nature of smartphone activities. This aligns with evidence that smartphone use before bedtime can increase cognitive arousal and hinder the presleep relaxation process critical for facilitating sleep onset [56].

The mixed findings regarding artificial light and smartphone use highlight the complexity of assessing individual sleep behaviors and their impacts. The lack of observed negative effects from artificial light before bedtime could indicate that its impact on sleep may be less pronounced or more individualized than previously thought, underscoring the need for an idiographic approach in future research to capture individual variations in light sensitivity [57]. The association between smartphone use and increased SOL suggests that smartphone engagement before bedtime can interfere with the winding-down process necessary for sleep. This points to the importance of limiting mentally stimulating activities in the evening to promote a smoother transition to sleep. Our findings support the inclusion of recommendations in sleep hygiene guidelines to reduce evening smartphone use as a strategy to maximize sleep and circadian health.

Strengths and Limitations of the Study

A major strength of this study is the use of innovative sensor-based tools to collect objective, high-resolution data on sleep architecture, light exposure, physical activity, and smartphone use in real-world settings. The use of wearable electroencephalography devices provided precise measurements of sleep stages, offering a significant advantage over traditional actigraphy, which cannot accurately differentiate between sleep stages, and an even greater advantage over self-reported sleep assessments via questionnaires commonly used in studies on the environmental determinants of sleep [32]. The continuous monitoring of light exposure using data loggers positioned near eye level allowed for accurate estimation of circadian entrainment and disruption factors. Additionally, the integration of a smartphone app to track usage patterns provided insights into a major behavioral circadian disruptor. The repeated measures design over 7 consecutive days enabled us to capture day-to-day variability and to better identify individual differences, offering a nuanced view of how environmental exposures and behaviors impact sleep architecture over time.

However, the study has several limitations. The short duration and small sample size of the study may limit the generalizability of the findings. While we accounted for seasonal variation, the 7-day monitoring period may have restricted our capacity to fully capture its influence on sleep patterns [39]. Additionally,

XSL•FO
RenderX

the limited sample size prevented us from controlling for sleep autocorrelation, a key component of sleep variability [39,57], and may have limited our ability to detect meaningful associations. Future studies with extended monitoring periods and larger samples are needed to account for temporal dynamics and enhance precision in understanding sleep variability and its determinants.

Due to missing data and limited observations, we were also unable to include all relevant predictors in a single comprehensive model, which restricted our ability to evaluate potential reciprocal confounding effects between the exposures. Furthermore, although smartphone ecological momentary assessments captured several health-related behaviors in our study, excessive missing data prevented us from using these measures to account for behaviors with known disruptive effects on sleep, such as the timing of caffeine, alcohol, medication, and food intake, napping habits, or the use of blue light–blocking lenses or screen filters [23-25]. These unmeasured confounders could have influenced the associations that we observed or our ability to identify associations that we were unable to detect. Finally, the reliance on light intensity measurements without considering the spectral composition of light limited our ability to differentiate between the effects of different light wavelengths on circadian rhythms [16,51]. Including such variables in future studies could improve the accuracy and comprehensiveness of the findings.

Despite these limitations, our study highlights the potential of sensor-based methods in advancing sleep research. By demonstrating the feasibility of integrating multiple wearable devices and a smartphone app to capture detailed environmental and behavioral data, we contribute to the development of ecologically valid approaches to studying sleep and circadian health. Serving as a proof of concept, this approach lays the groundwork for future studies with larger samples and extended monitoring periods, paving the way for more ecologically valid investigations into the complex interactions between environmental exposures, behaviors, and sleep outcomes.

Conclusions

Our findings underscore the beneficial impact of daily sunlight exposure and physical activity on enhancing sleep duration and reducing light sleep, highlighting the importance of natural light and movement for circadian alignment and overall sleep health. While artificial light before bedtime showed no association with sleep, smartphone use was linked to delayed sleep onset, potentially due to its stimulating effects. These results support public health recommendations to encourage outdoor exposure to sunlight and more physical activity while advising limited smartphone use in the evening to improve sleep and circadian health. These insights contribute to a better understanding of the factors influencing sleep architecture and underscore the potential of sensor-based methodologies in advancing sleep research.

Funding

This work is supported by the Research Executive Agency under the European Union’s Horizon 2020 Marie Skłodowska-Curie actions (H2020-MSCA-ITN; grant 956780).

Data Availability

The dataset analyzed in this study is available from the corresponding author upon request.

Authors' Contributions

Conceptualization: AM
Methodology: AM
Investigation: AM
Data curation: LMW, AB
Supervision: BC, AB
Writing – original draft: AM
Writing – review & editing: AM, LMW, AB, BC

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details of the model.
[DOCX File , 56 KB - mhealth_v14i1e75898_app1.docx]

References

1. Grandner MA. Sleep and Health. Cambridge: Academic Press; 2019.
2. Baglioni C, Nanovska S, Regen W, Spiegelhalder K, Feige B, Nissen C, et al. Sleep and mental disorders: a meta-analysis of polysomnographic research. Psychol Bull 2016;142(9):969-990 [FREE Full text] [doi: 10.1037/bul0000053] [Medline: 27416139]

XSL•FO
RenderX

3. Cappuccio F, D'Elia L, Strazzullo P, Miller M. Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies. Sleep 2010;33(5):585-592 [FREE Full text] [doi: 10.1093/sleep/33.5.585] [Medline: 20469800]
4. Grandner MA, Jackson NJ, Pak VM, Gehrman PR. Sleep disturbance is associated with cardiovascular and metabolic disorders. J Sleep Res 2012;21(4):427-433 [FREE Full text] [doi: 10.1111/j.1365-2869.2011.00990.x] [Medline: 22151079]
5. Krause AJ, Simon EB, Mander BA, Greer SM, Saletin JM, Goldstein-Piekarski AN, et al. The sleep-deprived human brain. Nat Rev Neurosci 2017;18(7):404-418 [FREE Full text] [doi: 10.1038/nrn.2017.55] [Medline: 28515433]
6. Patel AK, Reddy V, Shumway KR, Araujo JF. Physiology, Sleep Stages. Treasure Island (FL): StatPearls Publishing; 2024.
7. Moorcroft WH. Functions of sleep and NREMS. In: Understanding Sleep and Dreaming. Boston, MA: Springer US; 2013:231-259.
8. Walker MP, van der Helm E. Overnight therapy? The role of sleep in emotional brain processing. Psychol Bull 2009;135(5):731-748 [FREE Full text] [doi: 10.1037/a0016570] [Medline: 19702380]
9. Le Bon O. Relationships between REM and NREM in the NREM-REM sleep cycle: a review on competing concepts. Sleep Med 2020;70:6-16. [doi: 10.1016/j.sleep.2020.02.004] [Medline: 32179430]
10. Buijs FN, León-Mercado L, Guzmán-Ruiz M, Guerrero-Vargas NN, Romo-Nava F, Buijs RM. The circadian system: a regulatory feedback network of periphery and brain. Physiology (Bethesda) 2016;31(3):170-181 [FREE Full text] [doi: 10.1152/physiol.00037.2015] [Medline: 27053731]
11. Haim A, Zubidat AE. Artificial light at night: melatonin as a mediator between the environment and epigenome. Philos Trans R Soc Lond B Biol Sci 2015;370(1667):20140121 [FREE Full text] [doi: 10.1098/rstb.2014.0121] [Medline: 25780234]
12. Daan S, Aschoff J. The entrainment of circadian systems. In: Takahashi JS, Turek FW, Moore RY, editors. Circadian Clocks. Boston, MA: Springer US; 2001:7-43.
13. Burns AC, Saxena R, Vetter C, Phillips AJ, Lane JM, Cain SW. Time spent in outdoor light is associated with mood, sleep, and circadian rhythm-related outcomes: a cross-sectional and longitudinal study in over 400,000 UK Biobank participants. J Affect Disord 2021;295:347-352 [FREE Full text] [doi: 10.1016/j.jad.2021.08.056] [Medline: 34488088]
14. Figueiro M, Jarboe C, Sahin L. The sleep maths: a strong correlation between more daytime light and better night-time sleep. Lighting Res Technol 2021;53(5):423-435. [doi: 10.1177/14771535211005832]
15. Vetter C. Circadian disruption: what do we actually mean? Eur J Neurosci 2020;51(1):531-550 [FREE Full text] [doi: 10.1111/ejn.14255] [Medline: 30402904]
16. Hatori M, Gronfier C, Van Gelder RN, Bernstein PS, Carreras J, Panda S, et al. Global rise of potential health hazards caused by blue light-induced circadian disruption in modern aging societies. NPJ Aging Mech Dis 2017;3:9 [FREE Full text] [doi: 10.1038/s41514-017-0010-2] [Medline: 28649427]
17. Scott H, Woods HC. Understanding links between social media use, sleep and mental health: recent progress and current challenges. Curr Sleep Med Rep 2019;5(3):141-149. [doi: 10.1007/s40675-019-00148-9]
18. Combertaldi SL, Ort A, Cordi M, Fahr A, Rasch B. Pre-sleep social media use does not strongly disturb sleep: a sleep laboratory study in healthy young participants. Sleep Med 2021;87:191-202 [FREE Full text] [doi: 10.1016/j.sleep.2021.09.009] [Medline: 34627122]
19. Atoui S, Chevance G, Romain A, Kingsbury C, Lachance J, Bernard P. Daily associations between sleep and physical activity: a systematic review and meta-analysis. Sleep Med Rev 2021;57:101426. [doi: 10.1016/j.smrv.2021.101426] [Medline: 33571893]
20. Kim N, Ka S, Park J. Effects of exercise timing and intensity on physiological circadian rhythm and sleep quality: a systematic review. Phys Act Nutr 2023;27(3):52-63 [FREE Full text] [doi: 10.20463/pan.2023.0029] [Medline: 37946447]
21. Seol J, Lee J, Park I, Tokuyama K, Fukusumi S, Kokubo T, et al. Bidirectional associations between physical activity and sleep in older adults: a multilevel analysis using polysomnography. Sci Rep 2022;12(1):15399. [doi: 10.1038/s41598-022-19841-x] [Medline: 36100642]
22. Ferreira MF, Bos SC, Macedo AF. The impact of physical activity on objective sleep of people with insomnia. Psychiatry Res 2023;320:115019. [doi: 10.1016/j.psychres.2022.115019] [Medline: 36577233]
23. Garcia AN, Salloum IM. Polysomnographic sleep disturbances in nicotine, caffeine, alcohol, cocaine, opioid, and cannabis use: a focused review. Am J Addict 2015;24(7):590-598. [doi: 10.1111/ajad.12291] [Medline: 26346395]
24. Mantua J, Spencer RMC. Exploring the nap paradox: are mid-day sleep bouts a friend or foe? Sleep Med 2017;37:88-97 [FREE Full text] [doi: 10.1016/j.sleep.2017.01.019] [Medline: 28899546]
25. Oike H, Oishi K, Kobori M. Nutrients, clock genes, and chrononutrition. Curr Nutr Rep 2014;3(3):204-212 [FREE Full text] [doi: 10.1007/s13668-014-0082-6] [Medline: 25101217]
26. Chang A, Aeschbach D, Duffy JF, Czeisler CA. Evening use of light-emitting eReaders negatively affects sleep, circadian timing, and next-morning alertness. Proc Natl Acad Sci U S A 2015;112(4):1232-1237 [FREE Full text] [doi: 10.1073/pnas.1418490112] [Medline: 25535358]
27. Kaur P, Dhir A, Alkhalifa AK, Tandon A. Social media platforms and sleep problems: a systematic literature review, synthesis and framework for future research. Internet Res 2021;31(4):1121-1152. [doi: 10.1108/intr-04-2020-0187]
28. Murray K, Godbole S, Natarajan L, Full K, Hipp JA, Glanz K, et al. The relations between sleep, time of physical activity, and time outdoors among adult women. PLoS One 2017;12(9):e0182013 [FREE Full text] [doi: 10.1371/journal.pone.0182013] [Medline: 28877192]

XSL•FO
RenderX

29. Parsons AA, Ollberding NJ, Smith L, Copeland KA. Sleep matters: the association of race, bedtime, outdoor time, and physical activity with preschoolers' sleep. Prev Med Rep 2018;12:54-59 [FREE Full text] [doi: 10.1016/j.pmedr.2018.08.008] [Medline: 30181946]
30. Masanneck L, Gieseler P, Gordon WJ, Meuth SG, Stern AD. Evidence from ClinicalTrials.gov on the growth of digital health technologies in neurology trials. NPJ Digit Med 2023;6(1):23 [FREE Full text] [doi: 10.1038/s41746-023-00767-1] [Medline: 36765123]
31. de Zambotti M, Goldstein C, Cook J, Menghini L, Altini M, Cheng P, et al. State of the science and recommendations for using wearable technology in sleep and circadian research. Sleep 2024;47(4):zsad325. [doi: 10.1093/sleep/zsad325] [Medline: 38149978]
32. Arnal P, Thorey V, Debellemaniere E, Ballard M, Bou Hernandez A, Guillot A, et al. The Dreem headband compared to polysomnography for electroencephalographic signal acquisition and sleep staging. Sleep 2020;43(11):zsaa097 [FREE Full text] [doi: 10.1093/sleep/zsaa097] [Medline: 32433768]
33. Danzig R, Wang M, Shah A, Trotti LM. The wrist is not the brain: estimation of sleep by clinical and consumer wearable actigraphy devices is impacted by multiple patient- and device-specific factors. J Sleep Res 2020;29(1):e12926 [FREE Full text] [doi: 10.1111/jsr.12926] [Medline: 31621129]
34. Imtiaz SA. A systematic review of sensing technologies for wearable sleep staging. Sensors (Basel) 2021;21(5):1562 [FREE Full text] [doi: 10.3390/s21051562] [Medline: 33668118]
35. Lee XK, Chee NI, Ong JL, Teo TB, van Rijn E, Lo JC, et al. Validation of a consumer sleep wearable device with actigraphy and polysomnography in adolescents across sleep opportunity manipulations. J Clin Sleep Med 2019;15(9):1337-1346 [FREE Full text] [doi: 10.5664/jcsm.7932] [Medline: 31538605]
36. Marino M, Li Y, Rueschman MN, Winkelman JM, Ellenbogen JM, Solet JM, et al. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. Sleep 2013;36(11):1747-1755 [FREE Full text] [doi: 10.5665/sleep.3142] [Medline: 24179309]
37. de Gans C, Burger P, van den Ende E, Hermanides J, Nanayakkara PWB, Gemke RJBJ, et al. Sleep assessment using EEG-based wearables - a systematic review. Sleep Med Rev 2024;76:101951 [FREE Full text] [doi: 10.1016/j.smrv.2024.101951] [Medline: 38754209]
38. Lujan MR, Perez-Pozuelo I, Grandner MA. Past, present, and future of multisensory wearable technology to monitor sleep and circadian rhythms. Front Digit Health 2021;3 [FREE Full text] [doi: 10.3389/fdgth.2021.721919]
39. Chevance G, Minor K, Vielma C, Campi E, O'Callaghan-Gordo C, Basagaña X, et al. A systematic review of ambient heat and sleep in a warming climate. Sleep Med Rev 2024;75:101915 [FREE Full text] [doi: 10.1016/j.smrv.2024.101915] [Medline: 38598988]
40. Kim B, Branas CC, Rudolph KE, Morrison CN, Chaix B, Troxel WM, et al. Neighborhoods and sleep health among adults: a systematic review. Sleep Health 2022;8(3):322-333. [doi: 10.1016/j.sleh.2022.03.005] [Medline: 35504838]
41. Smith S, Trinder J. The effect of arousals during sleep onset on estimates of sleep onset latency. J Sleep Res 2000;9(2):129-135. [doi: 10.1046/j.1365-2869.2000.00194.x] [Medline: 10849239]
42. Shrivastava D, Jung S, Saadat M, Sirohi R, Crewson K. How to interpret the results of a sleep study. J Community Hosp Intern Med Perspect 2014;4(5):24983 [FREE Full text] [doi: 10.3402/jchimp.v4.24983] [Medline: 25432643]
43. Read SA, Vincent SJ, Tan CS, Ngo C, Collins MJ, Saw SM. Patterns of daily outdoor light exposure in Australian and Singaporean children. Transl Vis Sci Technol 2018;7(3):8 [FREE Full text] [doi: 10.1167/tvst.7.3.8] [Medline: 29862140]
44. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. Med Sci Sports Exerc 2011;43(2):357-364 [FREE Full text] [doi: 10.1249/MSS.0b013e3181ed61a3] [Medline: 20581716]
45. Crouter SE, Kuffel E, Haas JD, Frongillo EA, Bassett DR. Refined two-regression model for the ActiGraph Accelerometer. Med Sci Sports Exerc 2010;42(5):1029-1037 [FREE Full text] [doi: 10.1249/MSS.0b013e3181c37458] [Medline: 20400882]
46. Keadle SK, Shiroma EJ, Freedson PS, Lee IM. Impact of accelerometer data processing decisions on the sample size, wear time and physical activity level of a large cohort study. BMC Public Health 2014;14:1210 [FREE Full text] [doi: 10.1186/1471-2458-14-1210] [Medline: 25421941]
47. Phillips AJK, Vidafar P, Burns AC, McGlashan EM, Anderson C, Rajaratnam SMW, et al. High sensitivity and interindividual variability in the response of the human circadian system to evening light. Proc Natl Acad Sci U S A 2019;116(24):12019-12024 [FREE Full text] [doi: 10.1073/pnas.1901824116] [Medline: 31138694]
48. Ogilvie RD. The process of falling asleep. Sleep Med Rev 2001;5(3):247-270. [doi: 10.1053/smrv.2001.0145] [Medline: 12530990]
49. Wams EJ, Woelders T, Marring I, van Rosmalen L, Beersma DGM, Gordijn MCM, et al. Linking light exposure and subsequent sleep: a field polysomnography study in humans. Sleep 2017;40(12):zsx165 [FREE Full text] [doi: 10.1093/sleep/zsx165] [Medline: 29040758]
50. Jones SE, van Hees VT, Mazzotti DR, Marques-Vidal P, Sabia S, van der Spek A, et al. Genetic studies of accelerometer-based sleep measures yield new insights into human sleep behaviour. Nat Commun 2019;10(1):1585 [FREE Full text] [doi: 10.1038/s41467-019-09576-1] [Medline: 30952852]

XSL•FO
RenderX

51. Münch M, Nowozin C, Regente J, Bes F, De Zeeuw J, Hädel S, et al. Blue-enriched morning light as a countermeasure to light at the wrong time: effects on cognition, sleepiness, sleep, and circadian phase. Neuropsychobiology 2016;74(4):207-218. [doi: 10.1159/000477093] [Medline: 28637029]
52. Smith S, Trinder J. Morning sunlight can phase advance the circadian rhythm of young adults. Sleep Biol Rhythms 2005;3(1):39-41. [doi: 10.1111/j.1479-8425.2005.00153.x]
53. Chennaoui M, Arnal PJ, Sauvet F, Léger D. Sleep and exercise: a reciprocal issue? Sleep Med Rev 2015;20:59-72. [doi: 10.1016/j.smrv.2014.06.008] [Medline: 25127157]
54. Chellappa SL. Individual differences in light sensitivity affect sleep and circadian rhythms. Sleep 2021;44(2):zsaa214 [FREE Full text] [doi: 10.1093/sleep/zsaa214] [Medline: 33049062]
55. McAlpine T, Mullan B, Clarke PJF. Assessing the daily association of sleep hygiene behaviours with sleep: a between and within persons approach. J Behav Med 2024;47(2):255-270 [FREE Full text] [doi: 10.1007/s10865-023-00448-0] [Medline: 37702911]
56. Brautsch LA, Lund L, Andersen MM, Jennum PJ, Folker AP, Andersen S. Digital media use and sleep in late adolescence and young adulthood: a systematic review. Sleep Med Rev 2023;68:101742 [FREE Full text] [doi: 10.1016/j.smrv.2022.101742] [Medline: 36638702]
57. Hekler EB, Klasnja P, Chevance G, Golaszewski NM, Lewis D, Sim I. Why we need a small data paradigm. BMC Med 2019;17(1):133 [FREE Full text] [doi: 10.1186/s12916-019-1366-x] [Medline: 31311528]

Abbreviations

N1: stage 1 sleep
N2: stage 2 sleep
N3: stage 3 sleep
REM: rapid eye movement
SE: sleep efficiency
SOL: sleep onset latency
TST: total sleep time

Edited by L Buis; submitted 13.Apr.2025; peer-reviewed by R Maidstone, L Masanneck; comments to author 03.Jul.2025; revised version received 13.Apr.2026; accepted 13.Apr.2026; published 05.Jun.2026.

Please cite as:
Montanari A, Wang LM, Birenboim A, Chaix B
The Impact of Sunlight and Artificial Light at Night on Sleep Stages: Evidence From a 7-Day Sensor-Based Observational Study
JMIR Mhealth Uhealth 2026;14:e75898
URL: https://mhealth.jmir.org/2026/1/e75898
doi:10.2196/75898
PMID:

©Andrea Montanari, Li Min Wang, Amit Birenboim, Basile Chaix. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 05.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Smartwatch-Measured Clinician Perceptions of Emergency Department Crowding and Their Association With Objective Metrics: Prospective Observational Study

Ji Sim Yoon[1*]; Sookyung Park[2*], MPH; K Jane Muir[3,4], PhD; Pamela B DeGuzman[2], PhD; So Yeon Shin[5]; Hansol Chang[1,6,7], MD, PhD; Taerim Kim[6], MD, PhD; Se Uk Lee[6], MD, PhD; Sungjun Hong[1], PhD; Won Chul Cha[1,6,7], MD, PhD; Junsang Yoo[1], PhD

[1]Department of Digital Health, SAIHST, Sungkyunkwan University, Seoul, Republic of Korea

[2]School of Nursing, University of Virginia, Charlottesville, VA, United States

[3]Center for Health Outcomes and Policy Research, School of Nursing, University of Pennsylvania, Philadelphia, PA, United States

[4]Departmet of Emergency Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[5]Department of Nursing, Samsung Medical Center, Seoul, Republic of Korea

[6]Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

[7]Digital Innovation Center, Samsung Medical Center, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Junsang Yoo, PhD
Department of Digital Health
SAIHST
Sungkyunkwan University
115, Irwon-ro, Gangnam-gu
Seoul, 06355
Republic of Korea
Phone: 82 2 2148 7703
Email: junnsang@skku.edu

Abstract

Background: Emergency department (ED) crowding compromises patient safety and increases clinical workload. Existing crowding metrics capture structural indicators but not how clinicians subjectively experience crowding in real time. Clinician perception may provide a complementary measure.

Objective: This study aimed to examine the concurrent and temporal associations between clinicians' real-time ED perceived crowding (EDPC), captured via a smartwatch-based ecological momentary assessment (EMA) system, and 3 objective crowding metrics: the ED occupancy ratio (EDOR), the National Emergency Department Overcrowding Score (NEDOCS), and the ED Work Index (EDWIN).

Methods: We conducted a prospective observational study over a continuous 449-hour period (November 25, 2024, 7 AM to December 13, 2024, 11:59 PM) in a single tertiary ED operating under reduced bed capacity and attending-only staffing due to a national physician shortage. We enrolled 28 clinicians (5 attending physicians and 23 charge nurses; mean age 36.3, SD 5.9 years; 21 female and 7 male) by voluntary nonprobability sampling. Clinicians recorded EDPC hourly during active shifts on a continuous 0-200 scale anchored to the NEDOCS crowding categories, from "Not Busy" to "Dangerously Overcrowded." All valid ratings within each hourly period were aggregated into a single median EDPC value. Interrater agreement was assessed with the Gwet weighted agreement coefficient (AC2) at hourly time points with at least 2 ratings. Main analyses used Pearson correlations between hourly median EDPC and median EDOR, NEDOCS, and EDWIN. Linear mixed-effects models additionally accounted for repeated ratings within clinicians and clustering at the same hourly time point. Time-lagged correlations were exploratory. Statistical significance was set at α=.05.

Results: Across 1124 scheduled on-duty assessment opportunities, clinicians completed 1100 individual EDPC ratings (response rate 1100/1124, 97.86%). These ratings yielded 449 hourly aggregated EDPC observations, with a median of 2 (IQR 2-3; range 1-4) clinicians contributing ratings per hour. Interrater reliability was substantial (AC2=0.742, 95% CI 0.712-0.772). Objective

XSL•FO
RenderX

metrics were derived from 1846 ED visits. Hourly median EDPC correlated with median EDOR (r=0.542, 95% CI 0.473-0.604; P<.001) and NEDOCS (r=0.483, 95% CI 0.409-0.551; P<.001), but not EDWIN (r=0.060, 95% CI –0.033 to 0.152; P=.20).

Conclusions: We deployed smartwatch-based EMA to capture high-frequency, in situ clinician perceptions during active shifts, potentially reducing susceptibility to retrospective recall bias while minimizing the spatial constraints of traditional survey methods. Hourly median EDPC was moderately associated with EDOR and NEDOCS in this single tertiary ED during an exceptional staffing crisis. The association with EDWIN was less consistent and became significant only after adjustment for time-of-day patterns. The high response rate (1100/1124, 97.9%) supports the short-term feasibility of repeated wearable-based data collection during routine clinical shifts. These findings support further investigation of EDPC as a complementary operational measure but do not establish that it directly measures cognitive workload or predicts subsequent crowding.

(JMIR Mhealth Uhealth 2026;14:e103317) doi:10.2196/103317

KEYWORDS

clinician perception; crowding; ecological momentary assessment; EDOR; EDWIN; emergency department; NEDOCS; perceived crowding; real-time assessment; wearable device

Introduction

Problem

Emergency department (ED) crowding compromises the quality of care and increases clinician workload worldwide [1-3]. Crowding occurs when the demand for emergency services exceeds the available capacity of the ED or hospital resources [4], reflecting the combined impact of patient volume, acuity, bed availability, and staffing levels [2,3,5]. An aging population is bringing growing numbers of older patients whose multimorbidities and nonspecific symptoms prolong their length of stay [3,6,7]. Boarding of critically ill patients is a major driver of crowding [8-10], forcing departments to deliver intensive inpatient-level care. Despite this escalating burden, how the clinicians managing it experience crowding in real time remains poorly understood.

Review of Relevant Scholarship

Several metrics quantify different aspects of crowding [11,12]. Recent reviews, however, have emphasized that the performance and interpretability of ED crowding indices vary across hospital types and operational contexts [13], and that no universally accepted gold standard has been established [11-15]. These indices capture structural aspects of crowding, such as bed occupancy, boarding time, and acuity-weighted patient-to-staff ratios [16-18], rather than how clinicians subjectively experience crowding in real time.

Existing indices differ in what they operationalize. The National Emergency Department Overcrowding Score (NEDOCS), the most widely used metric [11,12,15], incorporates multiple operational components, including the number of patients currently in the ED, the number of admitted patients waiting for hospital beds, the total number of hospital beds, the longest waiting time from triage to bed assignment, the longest boarding time, and the number of patients receiving mechanical ventilation [17]. Based on these parameters, NEDOCS categorizes ED status into 6 levels ranging from “Not Busy” (0-20) to “Dangerously Overcrowded” (180-200) [11,12,17]. The ED occupancy ratio (EDOR) also focuses on physical occupancy, comparing active patients with operational beds [16]. Whereas EDOR and NEDOCS are based on patient counts and bed capacity, the Emergency Department Work Index (EDWIN) incorporates patient acuity and staffing levels to reflect the intensity of care required [18].

Clinicians’ subjective perception of crowding has received relatively limited empirical attention [11,14], despite growing interest in how it relates to these objective crowding measures [12,15]. Notably, perceived and objective crowding can diverge: in the original NEDOCS validation study, physicians reported feeling less rushed once the ED became severely overcrowded and all beds were full [17], suggesting that objective crowding may not align with clinicians’ subjective perceptions. This divergence has been repeatedly documented across diverse ED settings [19,20].

In this study, ED perceived crowding (EDPC) refers to clinicians’ real-time subjective perception of ED crowding, distinct from operational crowding, clinical workload, and cognitive workload. These constructs are related but not interchangeable [19], and this study directly measured only clinicians’ perceptions of ED crowding. To operationalize EDPC, we adapted the 6 anchors of the original NEDOCS outcome scale ranging from “Not Busy (0-20)” to “Dangerously Overcrowded (180-200)” as a real-time response scale [17].

Where prior work has captured clinician perception, it has typically relied on convenience samples of discrete time blocks [12,14,17] rather than systematic, repeated measurements across a full shift. In this fast-paced, highly uncertain environment [21], real-time assessment of EDPC may provide complementary operational insights. The role of frontline clinicians is critical to timely care delivery [22,23]; operational metrics alone, however, cannot capture their direct experience of crowding. Recent advances in mobile data collection, particularly ecological momentary assessment (EMA), offer a practical means to capture real-time perception while limiting post hoc rationalization or recall bias [24-26]. However, it remains unclear how clinicians’ real-time perceptions of crowding correspond to established objective crowding measures when both are assessed concurrently at high temporal resolution.

Hypothesis, Aims, and Objectives

To address this knowledge gap, we used a smartwatch-based EMA system to collect real-time EDPC data during clinical shifts. The primary aim of this study was to evaluate the concurrent associations between hourly median EDPC and the

XSL•FO
RenderX

corresponding hourly median values of the established objective crowding metrics: EDOR, NEDOCS, and EDWIN. Secondary analyses examined temporal associations between EDPC and these objective metrics. We also assessed interrater agreement among clinicians as a measurement property of EDPC and described response completion rates to characterize the feasibility of smartwatch-based EMA data collection.

Methods

Study Design

This prospective observational study combined retrospectively extracted electronic health record (EHR) data, used to derive the objective metrics, with prospectively collected clinician-reported crowding measures.

Setting

We recruited clinicians and obtained informed consent from November 12 to November 28, 2024. Data were collected in the ED of Samsung Medical Center, a tertiary referral hospital in Seoul, South Korea, from November 25, 2024, at 7 AM to December 13, 2024, at 11:59 PM, comprising 449 hourly observation time points. The study site has 1766 inpatient beds, including 67 ED beds, and manages approximately 65,000-70,000 ED visits annually. During the study period, ED operations were affected by a national medical workforce shortage following the mass resignation of residents and fellows [27,28]. Consequently, attending physicians delivered all ED care without junior medical staff, and the department operated 40 beds rather than its usual 67.

Participants

Inclusion and Exclusion

Eligible participants were attending physicians and charge nurses providing clinical care in the ED (central or acute care areas) during the study period; residents and fellows were not eligible because they were not involved in ED care at that time.

Sampling Procedures

We identified eligible clinicians from the duty roster, informed them about the study, and recruited them by voluntary nonprobability sampling; those who declined were not approached further, and written informed consent was obtained from all enrolled participants.

Of the 37 eligible clinicians, 28 (75.7%) participated, including 5 (35.7%) of 14 attending physicians and all 23 (100%) charge nurses; the remaining 9 attending physicians did not participate. Accordingly, the enrolled clinicians were not intended to constitute a representative sample of ED staff.

The observation period extended continuously from November 25, 2024, at 7 AM through December 13, 2024, at 11:59 PM, yielding 449 consecutive hourly assessment time points: 17 time points from 7 AM to 11 PM on November 25 and 24 time points per day for the subsequent 18 days. Rather than imposing artificial rater quotas or subsampling responses, we collected all responses from participating clinicians on duty. Based on the department's staffing model, each shift was staffed by at most 2 charge nurses and 1 to 3 attending physicians. Because duty periods could overlap across successive shifts, up to 3 charge nurses could be concurrently on duty during some hourly assessment periods. Because enrollment was voluntary, the number of participating clinicians on duty simultaneously varied across hourly assessment periods. Individual ratings within each hourly interval were aggregated into an hourly median EDPC value.

Participant Characteristics

Clinician demographic characteristics (age, sex, and years of clinical experience) and characteristics of ED patient visits during the study period (age, sex, triage acuity, and disposition) are reported in the Results section (Participant and Patient Characteristics).

Sample Size, Power, and Precision

No a priori sample size calculation was performed. The study size was determined by the prespecified data collection period and the number of eligible clinicians who volunteered. A total of 28 clinicians (5 physicians and 23 nurses) were enrolled, and 1100 individual EDPC ratings were obtained across 449 hourly observation time points. The precision of the estimated associations was assessed using 95% CIs. Because consecutive hourly observations were serially correlated, the 449 hourly time points were not considered statistically independent.

Assessments

Measures and Covariates

The primary outcome was EDPC, representing clinicians' perceived level of ED crowding. A raw EDPC entry was defined as a single clinician-reported rating recorded during an hourly assessment period. We then computed an hourly EDPC value by taking the median of all available entries within each hour. We chose the median value as the main hourly summary to provide a consistent measure of central tendency across hours with varying numbers of raters and to limit the influence of extreme ratings in hours with 3 or more raters. We defined the EDPC categories by anchoring scores to the 6 established NEDOCS severity categories (Not Busy, Busy, Extremely Busy, Overcrowded, Severely Overcrowded, and Dangerously Overcrowded) [17]. As the EDPC was recorded on a 0-200 numeric scale, this categorization followed the same point-based thresholds as the NEDOCS framework. This anchoring allowed direct comparison of clinicians' subjective perceptions with the objective operational thresholds defined by that framework.

The objective metrics EDOR, NEDOCS, and EDWIN [16-18] were calculated at a 1-minute temporal resolution using their respective published formulas implemented directly in R and evaluated as predictors of EDPC. These metrics represent different dimensions of ED crowding. Detailed formulas and variable definitions for all metrics are provided in Multimedia Appendix 1. Physician staffing data for the EDWIN calculation were taken from the duty roster and treated as time-varying. Physician shifts were scheduled as follows: *day* (D), 7:30 AM to 8:30 PM with 2 physicians; *evening* (E), 12 noon to 10 PM with 1 physician; and *night* (N), 8:30 PM to 7:30 AM with 2 physicians. When summarized by hourly assessment periods, under the usual staffing pattern, 2 physicians were on duty from

XSL•FO
RenderX

8 AM to 11:59 AM, 3 from 12 noon to 9:59 PM, and 2 from 10 PM to 7:59 AM the following morning. The only exceptions occurred during the N shifts beginning on November 26 and 27, 2024, when 1 physician rather than 2 was assigned to the N shift. We calculated EDWIN using the actual number of physicians scheduled on duty at each time point according to the duty roster, including these exceptions. We aggregated the objective crowding metrics into hourly summary statistics (mean, median, minimum, and maximum) for direct comparison with the hourly EDPC data.

To characterize the study population, we extracted demographic and clinical variables for each visit, including age, sex, acuity level according to the Korean Triage and Acuity Scale (KTAS), vital signs, and clinical disposition (categorized as discharge, hospital admission, transfer, death, or leaving without being seen). The latter category comprised patients who completed triage but left before entering the ED treatment area or being seen, either returning home or seeking care at another hospital. Because they remained in the ED census while physically present after triage, they contributed to the objective crowding metrics during that period.

Instrumentation

Raw EDPC entries were collected using a smartwatch-based system designed to support EMA [24-26] and to enable real-time assessment of clinicians' perceptions during active clinical shifts. Assessing perceptions in the moment was intended to reduce the recall bias associated with retrospective assessment. The research team developed a custom smartwatch application to support repeated real-time assessment within the clinical workflow. The 0-200 NEDOCS scale and its 6 crowding categories were adopted as reference anchors, giving clinicians an intuitive and standardized framework for rating their perceived crowding.

Four study smartwatches (Galaxy Watch 4 Classic; Samsung Electronics) running the custom application described above were provided for EDPC assessments. Participating clinicians wore the study smartwatches only during their on-duty shifts, and the devices were shared across successive shifts rather than assigned to individual participants. Participants did not use their personal smartwatches.

No formal pilot-testing of the EDPC instrument was conducted before the study.

Quality of Measurements

Before data collection, participants received a brief orientation on using the smartwatch application and the EDPC assessment procedure. We demonstrated the interface, including how to launch the application, enter a rating, and modify one already entered. Participants were informed that the interface displayed the 6 crowding categories based on the NEDOCS framework, with corresponding color-coded score ranges, and were instructed to use these references to rate perceived crowding once each hour.

Psychometrics

The interrater reliability of the raw EDPC entries was assessed using the Gwet weighted agreement coefficient (AC2) with ordinal weighting [29] at 447 hourly time points with at least 2 observed ratings. All available clinician ratings at each time point were included. We selected this approach because the number and composition of raters varied across hourly assessments. AC2 estimates with 95% CIs were calculated using the *irrCAC* package in R. As a complementary analysis of the continuous EDPC ratings, we fitted a crossed random-effects variance-component model with hourly time point and clinician as random intercepts to account for the changing and unbalanced rater composition.

Masking

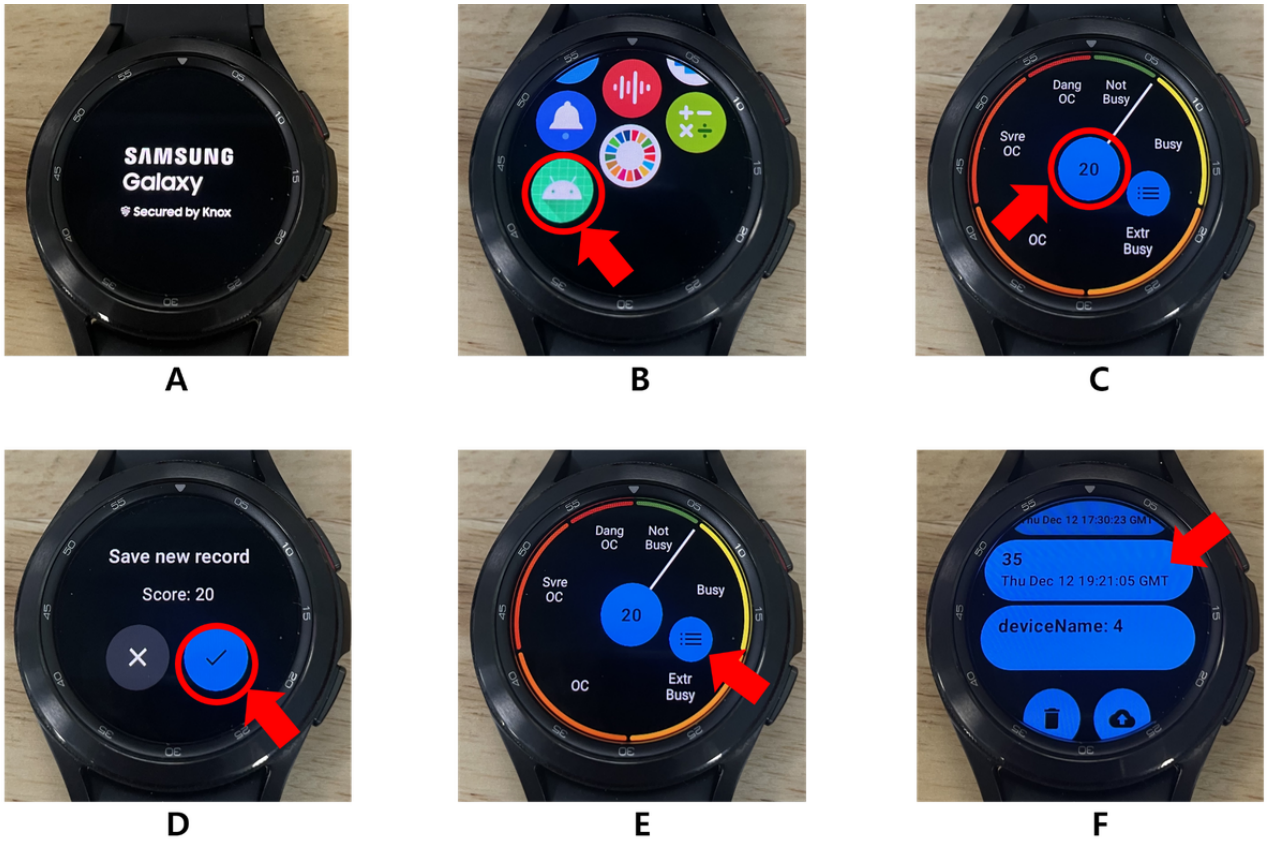
Although the NEDOCS-based categories and score ranges were displayed as reference anchors, participants could not view the actual real-time NEDOCS score while recording EDPC. Participants also had no access to an ED occupancy dashboard or to calculated operational crowding scores, including EDOR and EDWIN.

Data Collection

We prospectively collected EDPC data from enrolled clinicians using the smartwatch-based EMA system. EDPC assessments were conducted hourly. To reduce missed entries, the application delivered a gentle vibration reminder at 30 minutes past each hour. Participants entered their EDPC ratings using the bezel-controlled smartwatch interface shown in Figure 1, with score selection performed using either the rotating bezel or direct touchscreen interaction, followed by touch-based confirmation. Participants could enter one EDPC rating during each hourly assessment period and could correct an erroneous entry within the same period. Each EDPC response was assigned to the clock hour in which it was recorded, regardless of the timing of the reminder (eg, a response recorded at 11:50 AM was assigned to the 11-11:59 AM hourly period). Entries were not mandatory; therefore, clinicians could skip them during urgent clinical situations. If no response was recorded during an hourly assessment period while a clinician was on duty, that assessment opportunity was considered missing. Off-duty periods were excluded from EDPC assessment opportunities. Prompt-level completion rates were calculated as the number of completed EDPC ratings divided by the number of assessment opportunities during clinicians' on-duty periods. These rates were summarized overall, by professional role, and by individual clinician.

XSL•FO
RenderX

Figure 1. Workflow and user interface of the smartwatch-based application for recording raw emergency department perceived crowding (EDPC) entries (Galaxy Watch 4 Classic). (A) Initial smartwatch home screen. (B) Launch of the EDPC application. (C) Main EDPC input interface where clinicians selected their perceived crowding level on a 0-200 scale using either the rotating bezel or direct touchscreen interaction, with the 6 National Emergency Department Overcrowding Score (NEDOCS) crowding category labels, corresponding score ranges, and color coding displayed as visual references. The abbreviated labels displayed on the smartwatch were Extr Busy (extremely busy but not overcrowded), OC (overcrowded), Svre OC (severely overcrowded), and Dang OC (dangerously overcrowded). (D) Touch-based confirmation screen for saving the selected EDPC score. (E) Option to modify or reenter a previously recorded score. (F) Stored EDPC entry with the corresponding time stamp. The interface reproduced the NEDOCS-based category labels, score thresholds, and corresponding color coding as visual reference anchors; the actual calculated NEDOCS score at the time of assessment was not displayed. The use of these NEDOCS-based visual anchors may have introduced anchoring bias in the clinicians' EDPC ratings.



To compare the EDPC with the validated objective metrics, we extracted retrospective patient-level data for all ED visits during the study period from Samsung Medical Center's Clinical Data Warehouse (DARWIN-C). To ensure the accuracy of crowding assessment, we excluded visits that did not contribute to operational crowding: (1) cases immediately transferred to independent care units (eg, pediatric zones or delivery rooms), (2) visits that did not use ED clinical resources (eg, dead-on-arrival cases or nonclinical visits), and (3) erroneous records identified during data validation. ED and hospital bed capacities used in the calculations were obtained from institutional operational records.

Data Analysis

Data Diagnostics

Missing EDPC ratings were assessed at the individual clinician-rating level. The Little missing completely at random (MCAR) test evaluated whether missingness was completely at random.

As a sensitivity analysis, multiple imputation (MI) was performed for missing EDPC ratings using predictive mean matching to generate 50 imputed datasets. The imputation model included clinician ID, hourly median EDOR, NEDOCS, and EDWIN, sine and cosine terms for time of day, and study day. Although clinician ID and temporal covariates were included as predictors, within-clinician and serial correlation structures were not explicitly modeled. Hourly median EDPC values were recalculated for each imputed dataset, and Pearson correlation coefficients were pooled using Rubin's rules after Fisher *z* transformation.

Analytic Strategy

Clinician characteristics and key study variables are summarized using descriptive statistics. Categorical variables are presented as frequencies and percentages, and continuous variables as means (SD) or medians (IQR), as appropriate.

We performed exploratory descriptive analyses of hourly EDPC and the objective metrics to examine their distribution and diurnal patterns over the study period. We then assessed the association between hourly EDPC and each objective metric at contemporaneous time points.

The main analyses examined contemporaneous Pearson correlations between hourly median EDPC and the hourly median values of EDOR, NEDOCS, and EDWIN; no single

XSL•FO
RenderX

correlation was prespecified as the primary comparison. Exploratory analyses additionally examined Pearson correlations between hourly median EDPC and the minimum, maximum, and mean hourly values of each objective metric. Correlation strength was interpreted according to Cohen conventional benchmarks as weak (<0.3), moderate (0.3-0.5), or strong (≥0.5) [30].

In addition, we analyzed individual clinician-level EDPC ratings with linear mixed-effects models to account for the multilevel structure of the data. In the overall analysis, clinician and hourly time point were included as random intercepts to account for repeated ratings within clinicians and clustering of ratings obtained at the same time point. Profession-stratified analyses were performed separately for nurses and physicians, with clinician included as a random intercept to allow for repeated ratings within individuals.

As a post hoc sensitivity analysis, we calculated Spearman rank correlations using the same hourly summary statistics as in the Pearson analyses, to assess the robustness of the main results to departures from linearity and distributional assumptions. For the median-to-median associations, Spearman rank correlations and scatterplots were also examined overall and separately for physicians and nurses, with least absolute deviations (LAD) regression lines used to visually characterize these associations.

To account for temporal dependence, additional sensitivity analyses used generalized least squares (GLS) regression with first-order autoregressive—AR(1)—errors; sine and cosine terms for time of day simultaneously adjusted for common 24-hour diurnal patterns. Residual autocorrelation was assessed using the Ljung-Box test. In addition, we obtained percentile-based 95% CIs for the original Pearson correlation coefficients from a calendar-day block bootstrap with 10,000 resamples, resampling entire study days with replacement so that within-day temporal dependence was preserved.

As an exploratory analysis, we performed time-lagged correlation analyses by shifting each objective metric relative to hourly EDPC within a lag window of ±2 hours. A positive lag indicated that the objective metric preceded EDPC (eg, lag +1 h compared EDPC at time t with the objective metric measured at time t–1 h), whereas a negative lag indicated that the objective metric followed EDPC (eg, lag –1 h compared EDPC at time t with the objective metric measured at time t+1 h). Lag 0 represented contemporaneous measurements. This window was selected to align with the hourly sampling frequency and to capture short-term associations while avoiding more distant relationships that are less likely to be meaningful. No adjustment for multiple comparisons was applied.

The significance level for all statistical tests was set at α=.05. Analyses were performed using R (version 4.3.2; R Foundation for Statistical Computing) [31].

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki and was approved by the institutional review board (IRB) of Samsung Medical Center. Prospective collection of EDPC data from clinicians was approved under IRB number 2024-02-102, and written informed consent was obtained from all participating physicians and nurses before enrollment. Retrospective extraction of operational and clinical data from the hospital's Clinical Data Warehouse was approved under IRB number 2025-02-012, with a waiver of informed consent for the use of deidentified records. Clinician smartwatch responses and clinical data were deidentified using unique study identifiers and stored on encrypted, password-protected institutional servers to protect privacy and confidentiality. Participating clinicians were compensated KRW 3000 (US $2.10) per hour of participation. Compensation was calculated based on each clinician's total working hours during the study period. The mean participation time was 40.5 hours per clinician, corresponding to a mean compensation of approximately KRW 121,393 (US $84.96). The total compensation for all 28 participants was KRW 3,399,000 (US $2379). No identifying information about clinicians or patients is presented in the manuscript figures, user interface images, or supplementary materials.

Results

Participant Flow

Among 37 eligible clinicians, 28 (75.7%) participated in the study, including 5 (35.7%) of 14 eligible attending physicians and all 23 (100%) eligible charge nurses. Across the 449 continuous hourly observation intervals (November 25, 2024, 7 AM to December 13, 2024, 11:59 PM), clinicians completed 1100 of 1124 scheduled on-duty EDPC assessments, yielding an overall response rate of 97.86% and a missed-assessment rate of 2.14%. The completion rate was 99.45% (897/902) among nurses and 91.44% (203/222) among physicians (Table 1). The 24 missed assessments were concentrated in 7 of the 28 participants (range 1-15 missed prompts; Table S1 in Multimedia Appendix 1). Little test rejected the assumption that these data were MCAR (χ^2_3=12.3; *P*=.006). In the MI sensitivity analysis, the pooled correlations were similar to the observed-data estimates for EDOR (*r*=0.538, 95% CI 0.468-0.602; *P*<.001), NEDOCS (*r*=0.480, 95% CI 0.405-0.549; *P*<.001), and EDWIN (*r*=0.063, 95% CI –0.030 to 0.155; *P*=.19; Table S2 in Multimedia Appendix 1).

XSL•FO
RenderX

Table 1. Medical staff characteristics and emergency department perceived crowding (EDPC) assessment compliance (n=28)[a].

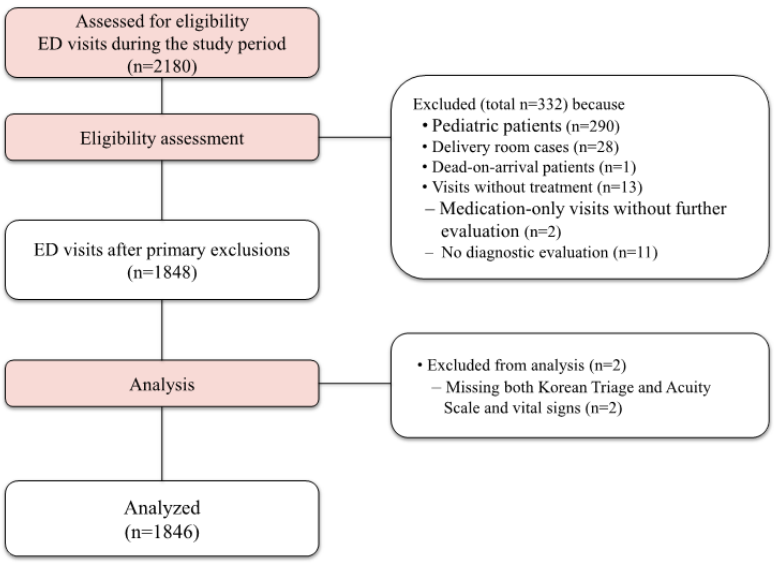
	Physician (n=5)	Nurse (n=23)	Total (n=28)
Demographics			
Age (years), mean (SD)	42.0 (6.7)	35.1 (5.0)	36.3 (5.9)
Sex, n (%)			
Female	3 (60.0)	18 (78.3)	21 (75.0)
Male	2 (40.0)	5 (21.7)	7 (25.0)
Total clinical experience (years), mean (SD)	16.2 (7.0)	10.8 (4.4)	11.8 (5.3)
EDPC assessment compliance			
Scheduled on-duty opportunities, n	222	902	1124
Completed EDPC ratings, n (%)	203 (91.44)	897 (99.45)	1100 (97.86)
Missed assessments, n (%)	19 (8.56)	5 (0.55)	24 (2.14)

[a]Clinical experience represents total years of practice rather than emergency department–specific experience. Participating physicians were attending physicians providing clinical care in the emergency department, and participating nurses were charge nurses assigned to the central or acute care areas.

Across all 449 scheduled hours, the median number of clinicians contributing ratings per hour was 2 (IQR 2-3; range 1-4). Specifically, 2 (0.4%) hours had 1 rater, 255 (56.8%) hours had 2 raters, 180 (40.1%) hours had 3 raters, and 12 hours (2.7%) had 4 raters. Hourly EDPC values were calculated as the median of all available clinician ratings within each time point, resulting in 449 aggregated EDPC observations. All 449 hourly time points had complete corresponding objective metric data and were retained for the association analyses. The 447 hourly time points with ratings from at least 2 clinicians were included in the interrater reliability analysis. The remaining 2 time points, each with only 1 clinician rating, were not eligible for that analysis but were retained for the association analyses.

During the study period, 1848 adult ED visits met the eligibility criteria. Two visits were excluded because both KTAS and vital sign data were missing, resulting in a final analytic sample of 1846 ED visits (Figure 2). “Visits without treatment” referred to visits in which patients did not undergo standard ED evaluation and consequently were not assigned a KTAS level. These included medication-only visits without further evaluation (n=2) and visits with no diagnostic evaluation (n=11). This group was distinct from patients categorized as “Left without being seen,” who completed triage but left before entering the ED treatment area or receiving treatment. Because visits without treatment were neither triaged nor included in the ED census, they were excluded from the study population and did not contribute to the objective crowding metrics.

Figure 2. Flow diagram of emergency department (ED) visits included in the study. This prospective observational study was conducted at a tertiary ED in Seoul, South Korea, from November 25 to December 13, 2024. The diagram shows ED visits assessed for eligibility, exclusions and their reasons, and the final analytic sample.



XSL•FO
RenderX

Recruitment

In total, 28 clinicians participated in this study, including 5 (17.9%) physicians and 23 (82.1%) nurses (Table 1).

Statistics and Data Analysis

Participant and Patient Characteristics

The mean age of the participating clinicians was 36.3 (SD 5.9) years, with a mean clinical experience of 11.8 (SD 5.3) years. Clinical experience varied by professional role, with an average of 16.2 (SD 7.0) years for physicians and 10.8 (SD 4.4) years for nurses. Overall, 75% (21/28) of the participants were women.

The mean patient age was 57.7 (SD 20.4) years, and 52.1% (961/1846) of visits involved male patients. KTAS level 3 was the most common acuity category (1070/1846, 58%), followed by levels 4 (397/1846, 21.5%) and 2 (311/1846, 16.8%). Most patients were alert upon arrival (1785/1846, 96.7%). Regarding disposition, 31.9% (589/1846) of visits resulted in discharge, 25.2% (466/1846) in hospital admission, and 38.8% (717/1846) in leaving without being seen. Patient characteristics and outcomes varied substantially across KTAS levels. KTAS 1-2 visits involved older and more clinically unstable patients, with high admission rates (43.5%-47.6%) and a mortality rate of 39.1% (9/23) among KTAS 1 visits. In contrast, KTAS 4-5 visits involved younger patients and substantially higher rates of leaving without being seen (71.3%-91.1%; Table S3 in Multimedia Appendix 1).

Reliability of EDPC Measurements

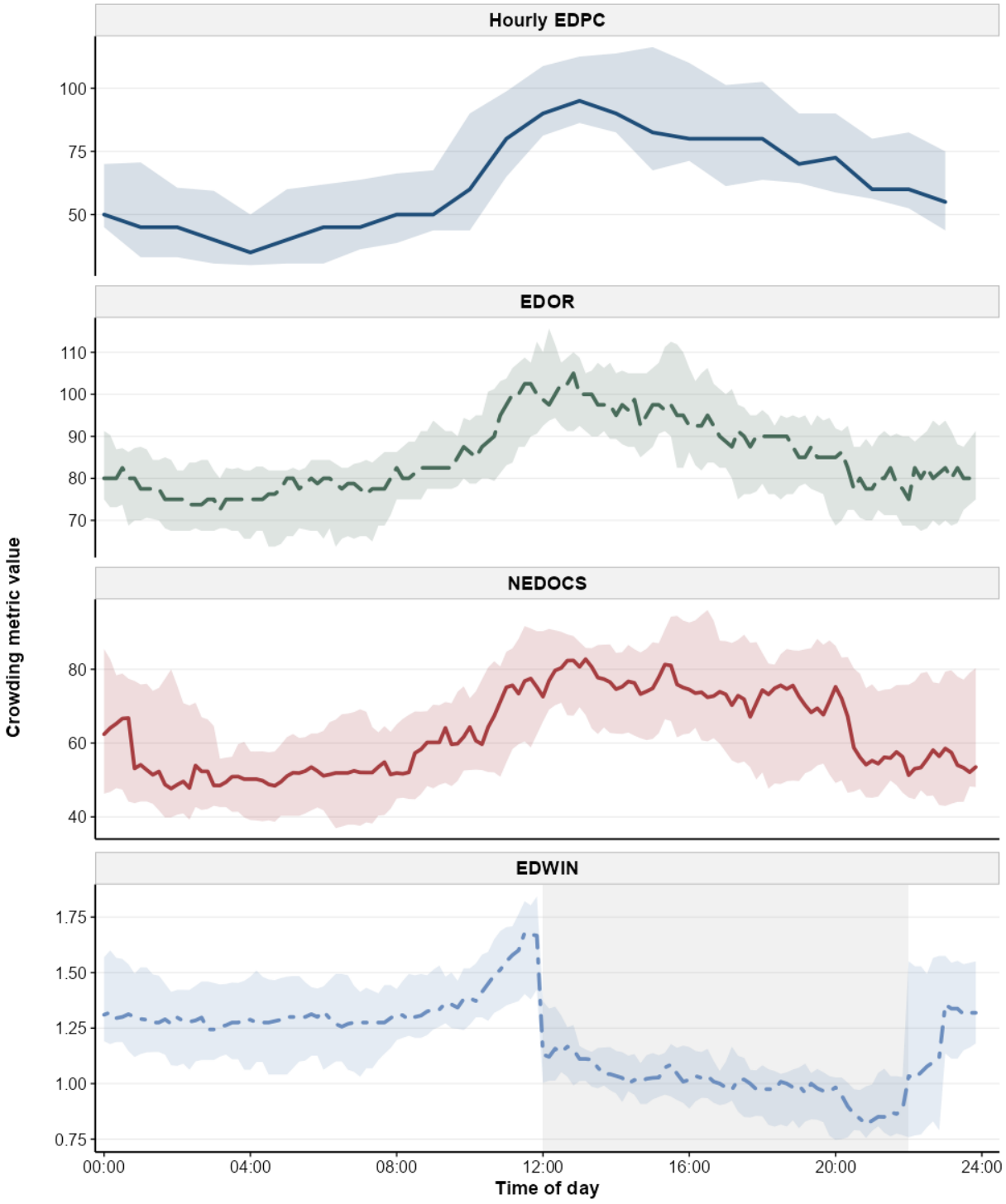
Overall and profession-specific EDPC assessment completion rates are presented in Table 1, and individual clinician-level assessment completion data are provided in Table S1 in Multimedia Appendix 1. EDPC ratings from at least 2 clinicians were available at 447 (99.6%) of the 449 hourly time points. Interrater reliability analysis demonstrated substantial agreement for raw EDPC scores (AC2=0.742, 95% CI 0.712-0.772) and almost perfect agreement after conversion to NEDOCS categories (AC2=0.811, 95% CI 0.790-0.833). In the complementary variance-component analysis of continuous EDPC ratings, 25.2% of the total variance was attributable to hourly time points, 23.5% to between-clinician differences, and 51.3% to residual variation.

Diurnal Patterns of Hourly EDPC and Objective ED Crowding Metrics

The distributions of hourly median EDPC and the hourly median values of EDOR, NEDOCS, and EDWIN are summarized in Table S4 in Multimedia Appendix 1. Both the perceived and objective crowding measures showed clear diurnal variation (Figure 3). The hourly EDPC across all raters increased after 7 AM, peaked at 1 PM, and gradually declined thereafter. Temporal patterns varied by professional role: nurse-reported hourly EDPC peaked at 1 PM, whereas physician-reported hourly EDPC remained elevated between 2 and 4 PM (Figure S1 in Multimedia Appendix 1). These role-specific patterns are presented descriptively and should be interpreted cautiously given the small number of participating physicians.

XSL•FO
RenderX

Figure 3. Diurnal patterns of hourly emergency department perceived crowding (EDPC) and objective emergency department crowding metrics. Data were collected at a tertiary ED in Seoul, South Korea, from November 25 to December 13, 2024. Lines represent median values calculated from observations pooled within each 10-minute time-of-day bin across the study period. Shaded bands represent the IQR (25th-75th percentile) of day-specific median values within each corresponding bin, indicating day-to-day variability. In the Emergency Department Work Index (EDWIN) panel, the vertical gray band marks the period when 3 physicians were usually on duty (12:00-21:59), and the unbanded areas mark the periods when 2 were (08:00-11:59 and 22:00-07:59 the following day). EDOR: emergency department occupancy rate; NEDOCS: National Emergency Department Overcrowding Score; EDWIN: Emergency Department Work Index.



Objective indices showed overall diurnal patterns similar to those of hourly EDPC, but with distinct features across metrics (Figure 3). NEDOCS showed 2 local peaks at approximately 1 and 3:30 PM. EDOR showed a late-morning peak at approximately 11:30 PM, followed by a generally declining pattern with some variability. EDWIN, in contrast, followed a different profile: a primary peak in the late morning (approximately 11:30 PM), then a rapid drop and a secondary peak in the late evening (approximately 11:20 PM).

XSL•FO
RenderX

Associations Between EDPC and Objective ED Crowding Metrics

Table 2 shows correlations between hourly EDPC and summary statistics of the 3 objective metrics. All Pearson correlation analyses were based on 449 paired hourly observations. EDPC correlated strongly with EDOR (r=0.53-0.56) and moderately to strongly with NEDOCS (r=0.47-0.51), whereas EDWIN showed no significant correlation.

Table 2. Pearson correlation between hourly emergency department perceived crowding and hourly summary statistics of objective metrics[a].

	Pearson *r* (95% CI)	*P* value
EDOR[b]		
Minimum	0.527 (0.457 to 0.591)	<.001
Maximum	0.563 (0.496 to 0.623)	<.001
Mean	0.545 (0.477 to 0.607)	<.001
Median	0.542 (0.473 to 0.604)	<.001
NEDOCS[c]		
Minimum	0.473 (0.398 to 0.542)	<.001
Maximum	0.510 (0.439 to 0.576)	<.001
Mean	0.492 (0.419 to 0.559)	<.001
Median	0.483 (0.409 to 0.551)	<.001
EDWIN[d]		
Minimum	0.014 (–0.079 to 0.106)	.77
Maximum	0.090 (–0.003 to 0.181)	.06
Mean	0.052 (–0.041 to 0.144)	.27
Median	0.060 (–0.033 to 0.152)	.20

[a]Each Pearson correlation analysis was based on 449 paired hourly observations.

[b]EDOR: emergency department occupancy ratio.

[c]NEDOCS: National Emergency Department Overcrowding Score.

[d]EDWIN: Emergency Department Work Index.

Post hoc sensitivity analyses using Spearman rank correlations yielded results consistent with the Pearson analyses. EDPC showed significant positive rank correlations with all summary statistics of EDOR (ρ=0.53-0.57; all P<.001) and NEDOCS (ρ=0.47-0.52; all P<.001), whereas EDWIN showed no significant correlations (ρ=–0.02 to 0.08; P=.09-.63; Table S5 in Multimedia Appendix 1). Scatterplots of hourly median EDPC against the hourly median EDOR, NEDOCS, and EDWIN values, with LAD regression lines and Spearman correlation coefficients, are presented overall and separately for physicians and nurses (Figure S2 in Multimedia Appendix 1).

In clinician-level multilevel analyses using all 1100 individual EDPC ratings and accounting for clustering by both clinician and hourly time point, EDPC was significantly associated with median EDOR (β=.966, 95% CI 0.840-1.091; P<.001) and median NEDOCS (β=.651, 95% CI 0.558-0.745; P<.001), but not with median EDWIN (β=3.865, 95% CI –1.761 to 9.490; P=.18; Table 3). Profession-stratified analyses showed the same pattern of statistical significance in nurses and physicians, with EDOR and NEDOCS significantly associated with EDPC and EDWIN showing no significant association in either professional group. The estimated EDWIN coefficient was positive for nurses (β=4.114) and negative for physicians (β=–.825), although neither association was statistically significant (Table 3).

Additional temporal sensitivity analyses showed that EDOR and NEDOCS remained significantly associated with hourly EDPC after simultaneously accounting for serial correlation and common time-of-day patterns using GLS AR(1) models (both P<.001). EDWIN was also significantly associated with EDPC after adjustment for both temporal components (P=.004; Table S6-A in Multimedia Appendix 1). Calendar-day block bootstrap analyses of the original Pearson correlations yielded 95% CIs excluding zero for EDOR (r=0.542, 95% CI 0.417-0.661) and NEDOCS (r=0.483, 95% CI 0.389-0.581), but not for EDWIN (r=0.060, 95% CI –0.165 to 0.234; Table S6-B in Multimedia Appendix 1).

In exploratory time-lagged correlation analyses, EDOR consistently showed positive correlations with EDPC across the ±2-hour lag window. NEDOCS followed a similar pattern, whereas EDWIN generally showed weaker associations across the lag conditions. The temporal patterns are presented in Figure S3 in Multimedia Appendix 1, and complete results, including lag direction, number of paired observations, correlation coefficients, 95% CIs, and unadjusted P values, are provided in Table S7 in Multimedia Appendix 1. P values were not adjusted for multiple comparisons because these analyses were exploratory.

XSL•FO
RenderX

Table 3. Multilevel analyses of associations between individual clinician-level emergency department perceived crowding ratings and objective emergency department crowding metrics.

Objective metric	Ratings, n	Clinicians, n	Hourly time points, n	β (95% CI)	SE	*P* value
Overall multilevel analysis[a]						
EDOR[b]	1100	28	449	.966 (0.840 to 1.091)	0.064	<.001
NEDOCS[c]	1100	28	449	.651 (0.558 to 0.745)	0.048	<.001
EDWIN[d]	1100	28	449	3.865 (–1.761 to 9.490)	2.870	.18
Profession-stratified multilevel analysis[e]						
Nurse						
EDOR	897	23	449	.985 (0.853 to 1.116)	0.067	<.001
NEDOCS	897	23	449	.641 (0.545 to 0.738)	0.049	<.001
EDWIN	897	23	449	4.114 (–1.389 to 9.616)	2.807	.14
Physician						
EDOR	203	5	191	.810 (0.620 to 1.000)	0.097	<.001
NEDOCS	203	5	191	.693 (0.549 to 0.838)	0.074	<.001
EDWIN	203	5	191	–.825 (–8.251 to 6.602)	3.789	.83

[a]Linear mixed-effects models were fitted using individual clinician-level emergency department perceived crowding ratings as the outcome and the median value of each objective crowding metric at the corresponding hourly time point as the predictor. Overall models included random intercepts for both clinician and hourly time point to account for repeated ratings within clinicians and clustering within the same time point. β represents the unstandardized regression coefficient.

[b]EDOR: emergency department occupancy ratio.

[c]NEDOCS: National Emergency Department Overcrowding Score.

[d]EDWIN: Emergency Department Work Index.

[e]Profession-stratified linear mixed-effects models included clinician as a random intercept to account for repeated ratings within clinicians. A separate hourly time point random intercept was not retained in the stratified models because within-hour replication was sparse in the physician subgroup. β represents the unstandardized regression coefficient.

Discussion

Support of Original Hypotheses (Primary Findings)

Clinician-perceived crowding, measured as the hourly EDPC, was associated with established objective crowding measures, although the magnitude and consistency of these associations differed across metrics. EDPC showed consistent associations with EDOR and NEDOCS, whereas its association with EDWIN depended on whether common time-of-day patterns were explicitly adjusted for. These findings suggest that clinician-perceived crowding may provide complementary information to objective crowding metrics. The high response rate supports the feasibility of repeated real-time EDPC assessment during routine clinical shifts.

Similarity of Results

Alignment of Hourly EDPC With Objective Crowding Measures

Hourly EDPC correlated moderately to strongly with occupancy-based measures, such as EDOR [15,16] and NEDOCS [17]. Clouston et al [12] demonstrated that ED occupancy, despite its simplicity, had a predictive accuracy similar to that of complex crowding scores such as NEDOCS and was useful for real-time tracking. Furthermore, recent before-and-after evidence indicates that transition units can reduce boarding and NEDOCS-defined crowding, illustrating that occupancy- and boarding-based measures may be particularly responsive to changes in hospital access block [32]. Our findings are consistent with this observation. Occupancy-based metrics may align well with EDPC because clinicians managing a growing number of patients directly experience the physical saturation of treatment spaces; however, this study cannot establish the specific mechanism underlying this association.

Hourly EDPC showed significant positive correlations with EDOR and NEDOCS across all 4 hourly summary statistics (minimum, maximum, mean, and median). Descriptively, correlations of EDPC with the maximum values of EDOR and NEDOCS were slightly higher than those with the corresponding mean or median values. However, these differences were small, and the dependent correlation coefficients were not formally compared; therefore, this observation should not be interpreted as evidence that clinician-perceived crowding is more strongly associated with peak values. Prior work has suggested that traditional crowding metrics may not fully capture the dynamic, rapidly fluctuating nature of ED operations [33], and that the relationship between objective crowding and clinician perception may be nonlinear [19]. These observations motivate further study of whether within-hour fluctuations or peak crowding conditions disproportionately shape clinician perceptions. Although we examined both median and peak crowding metrics,

XSL•FO
RenderX

this study cannot determine which measure best represents clinicians' subjective experience.

Comparison With Previous Real-Time Perception Studies

This study builds on the work of Boverio et al [19], who demonstrated the potential of real-time subjective assessments using EMA. Boverio et al [19] used smartphone text messages twice daily, a method limited by low measurement frequency and by the need to use a smartphone during clinical workflows, whereas our study used a smartwatch-based application with hourly prompts. Subtle wrist-vibration reminders prompted clinicians to complete hourly EDPC assessments through the application during routine clinical shifts. This approach was intended to reduce interruption and recall bias relative to retrospective, single–time point assessments, although these potential benefits were not directly measured and require further evaluation.

Patient-centered research further highlights discrepancies between objective measures and subjective perceptions of ED crowding. Wang et al [34] examined patient-perceived crowding at the end of ED encounters and observed variable agreement with objective indices. However, their study relied on a single–time point measurement, whereas our study captures clinicians' perceptions repeatedly throughout their clinical shifts, allowing for a more dynamic assessment.

Schippers et al [35] developed an AI-driven predictive model using routinely collected operational parameters to predict ED crowding, with ambulance diversion status as a binary outcome. In contrast, our study examines perceived crowding across varying levels of crowding through repeated, real-time assessments, rather than focusing on the threshold of ambulance diversion. Our findings demonstrate associations between objective crowding metrics and clinician-reported EDPC, suggesting that these metrics reflect certain aspects of clinicians' experiences of crowding but may not fully capture their subjective perceptions. Future qualitative research could explore discrepancies between operational indicators and clinicians' experiences in greater depth, providing insights into the contextual factors that shape perceived crowding beyond what quantitative measures can capture.

Interpretation

Interrater Agreement and Variability

Weighted AC2 indicated substantial ordinal agreement among clinicians' EDPC ratings (AC2=0.742). However, the complementary variance-component analysis showed that 25.2% of the total variance was attributable to hourly time points, 23.5% to between-clinician differences, and 51.3% to residual variation. These findings reflect different aspects of agreement and variability. Ordinally weighted AC2 allows partial agreement between nearby scores, whereas the variance-component analysis quantifies different sources of variability in the continuous ratings. Taken together, these findings suggest partial concordance in perceived crowding rather than interchangeability of individual clinician ratings.

Temporal Robustness and Interpretation of Associations

The associations of EDPC with EDOR and NEDOCS remained consistent across analyses accounting for serial correlation, common time-of-day patterns, and within-day temporal dependence, supporting their robustness to the temporal structure of the data. In contrast, the results for EDWIN were less consistent across the temporal sensitivity analyses. Although the original Pearson correlation and calendar-day block bootstrap analysis did not show a significant association between EDPC and EDWIN, a significant association emerged in models that explicitly adjusted for time-of-day patterns. These findings indicate that the EDPC-EDWIN association was sensitive to the adjustment for diurnal variation and should therefore be interpreted cautiously.

Occupancy-based metrics showed more consistent associations with hourly EDPC than EDWIN did. Notably, the diurnal profile of EDWIN diverged in shape from that of EDPC, EDOR, and NEDOCS: whereas the latter 3 measures peaked in the late morning to early afternoon, EDWIN exhibited an additional secondary peak in the late evening (approximately 11:20 PM) that was not mirrored by a corresponding rise in EDPC. This divergence in diurnal shape, rather than the shared presence of diurnal variation per se, may partly explain the inconsistency of the EDWIN-EDPC association across analytic approaches.

This finding must also be interpreted within the operational context of this study. Although EDWIN has the theoretical advantage of incorporating patient acuity [18], it also incorporates physician staffing, which varied over time during the study period. During a recent disruption in the physician workforce in South Korea [27], only 1 to 3 attending physicians delivered ED care, and the number of operational beds was reduced from 67 to 40. Therefore, the EDWIN findings should be interpreted in the context of this atypical staffing environment. Because EDWIN incorporates physician staffing and bed capacity directly into its formula [18], fluctuations in these structural components may influence the metric independently of clinicians' perceived crowding. This sensitivity analysis suggests that the inconsistent unadjusted association between EDWIN and EDPC in this study may be related to the exceptional staffing context and the staffing-sensitive construction of the metric; however, this study cannot establish that staffing variability was the sole or primary explanation. Considering that EDWIN aligned with staff assessments in its original development study and has been used in other settings [36], our findings should not be interpreted as evidence that EDWIN is not generally useful. Instead, staffing-sensitive metrics such as EDWIN should be applied with caution in smaller EDs and in those that adjust staffing in response to crowding.

In exploratory time-lagged analyses, EDPC showed similar correlations with objective crowding measures at lag 0 and at both +1-hour and –1-hour lags, indicating similar associations when the objective metrics preceded or followed EDPC by 1 hour. Because lagged correlations can arise from autocorrelation and shared diurnal trends rather than a genuine predictive relationship, and because predictive performance was not assessed, this finding should be interpreted as an exploratory

XSL•FO
RenderX

temporal association rather than evidence that EDPC functions as a leading indicator of crowding. Clinicians may detect early cues of worsening crowding, but confirmation of any predictive relationship would require dedicated time-series and prospective prediction analyses.

Role-Specific Differences in Perceived Crowding

Descriptively, nurse- and physician-reported EDPC showed different peak timing across the day (Figure S1 in Multimedia Appendix 1). However, in the profession-stratified multilevel analyses, EDOR and NEDOCS were significantly associated with individual EDPC ratings in both nurses and physicians, whereas EDWIN was not significantly associated with EDPC in either group. Thus, despite the descriptive differences in diurnal patterns, the pattern of associations with the objective crowding metrics was similar across professions. These findings should be interpreted cautiously given the small number of participating physicians (n=5) and require confirmation in larger, more balanced samples. Previous studies have shown that perceived crowding differs significantly across professional roles in the ED, including among nursing roles [19]. Because our study included only charge nurses and attending physicians, future research could incorporate other clinical roles, administrative staff, and patients to examine differences in perceived crowding across stakeholder groups and explore the factors underlying these differences.

Generalizability

The applicability of the assessment approach and the observed associations may vary across settings. Hourly smartwatch-based reporting could be adapted to other EDs, although the short study period limits conclusions about sustained participation and acceptability in routine practice. The magnitude and pattern of associations between EDPC and objective metrics may also vary according to staffing models, operational capacity, and patient populations. In particular, the attending-only staffing model and reduced bed capacity during the exceptional workforce disruption observed in this study may limit generalizability to EDs operating under routine conditions. Multicenter studies with longer observation periods under routine operating conditions are needed to evaluate sustained participation and the consistency of these associations across settings.

Implications

Clinical Importance of Clinician-Perceived Crowding

These findings have important implications for ED care delivery. Previous studies have consistently shown that ED crowding is associated with an increased length of stay, delays in care, reduced care quality, and poor patient safety outcomes [1,3]. Crowding arises when demand for services exceeds a department's available resources [4], and ED clinicians play a critical role in timely patient assessment [22,23], initiation of time-sensitive treatment [37], and recognition of clinical deterioration [38]. Prior literature has also characterized ED clinicians' workload as involving substantial cognitive demands and time pressure that are not fully captured by structural factors such as bed and patient numbers [39].

Objective metrics often assume static patient conditions, whereas the ED environment and patient status can change rapidly [40]. Real-time monitoring tools have been proposed to better capture these fluctuations [19,33]. In this context, hourly EDPC may provide complementary information about clinicians' real-time experience of crowding that is not captured by structural metrics alone; whether this information reflects cognitive load, safety risk, or other unmeasured dimensions of operational burden was not evaluated in this study and remains an open question for future research.

Methodological Contributions and Future Directions

The high overall response rate of 97.9% (1100 of 1124 expected on-duty assessments) supports the feasibility of this approach for frequent EDPC assessments during routine clinical shifts. This finding indicates that clinicians could complete hourly ratings consistently during the study period. However, acceptability, sustainability, and documentation burden were not directly measured relative to other assessment methods and require further evaluation. Documentation burden, the increased time and effort required to record care, is a well-known contributor to professional burnout syndrome [41,42]. Moreover, many health care technologies are abandoned when they introduce excessive complexity or demand substantial changes to established clinical routines [43]. Although the smartwatch interface was designed to minimize this burden through score selection using either the rotating bezel or direct touchscreen interaction, followed by touch-based confirmation, this study did not directly assess whether it achieved this relative to other data collection methods.

Importantly, the implementation did not require each participant to own or be assigned an individual smartwatch. Study devices were shared across successive shifts, so the number required depended on how many participating clinicians were concurrently on duty rather than on total enrollment. This device-sharing approach may reduce hardware requirements and remove the need for clinicians to own compatible devices. Nevertheless, the feasibility, cost, and scalability of this approach in larger or differently staffed EDs were not evaluated in this study and require further investigation.

Despite these methodological advantages, our approach did not entirely overcome the interruptive nature of active EMA. Raw EDPC entries may have been missed more frequently during periods when clinicians were too busy to record them [44]. In addition, as participation was voluntary, it remains uncertain whether this active recording approach is sustainable in routine practice outside the research setting.

Our findings do not imply that subjective perceptions should be used as stand-alone metrics for ED crowding. Relying solely on subjective assessments carries inherent clinical risk. As Garcia-Romero et al [20] have demonstrated, clinicians who are chronically exposed to excessive crowding tend to underestimate objectively measured crowding levels, a pattern consistent with the "normalization of deviance." To address both the interruptive limitations of EMA and individual clinicians' tendency to underestimate crowding, future research should focus on hybrid monitoring models. As a hypothesis-generating direction, future work could combine

XSL•FO
RenderX

active EDPC data with passive smartwatch-based physiological sensing (such as heart rate variability) [45], and objective EHR-based metrics in a multidimensional monitoring approach. Such an approach remains speculative at this stage; whether it could support applications such as early identification of rising operational burden or resource-reallocation alerts would require dedicated development and prospective validation before any clinical implementation could be considered.

Beyond technical feasibility, future research should use effectiveness-implementation hybrid designs to evaluate the acceptability, adoption, and long-term sustainability of smartwatch-based EDPC assessments. These studies could also investigate strategies to minimize reporting burden and protect frontline staff well-being. Furthermore, research should explore how real-time EDPC assessments could inform actionable thresholds for resource allocation and operational decision-making.

Limitations

This study had several limitations. First, it was conducted in a single tertiary hospital over a relatively short period, limiting its generalizability. Although EDPC and objective crowding metrics were measured at high temporal resolution, the limited study duration did not allow us to assess seasonal variation in ED crowding or in clinician perceptions.

Second, the EDPC instrument used NEDOCS-derived anchors, categories, and score ranges, and has not been independently validated against other subjective or objective crowding measures. Although participants could not view the actual real-time NEDOCS score while recording EDPC, the shared framework may have influenced clinicians' subjective ratings through anchoring or criterion contamination and contributed to the observed association between EDPC and NEDOCS. This design-related limitation cannot be resolved through statistical adjustment alone; accordingly, that association should not be interpreted as independent validation of either measure. Future validation studies should use neutral scale anchors that do not reproduce the categories or thresholds of the comparator metric.

Third, the study was conducted during an unprecedented national health care workforce crisis. During this study, attending physicians delivered all ED care, without residents or fellows, and as a crisis response, only 40 of the 67 ED beds were operational. This exceptional staffing and capacity context is not solely a threat to generalizability: because EDWIN incorporates physician staffing and bed capacity directly into its formula, these conditions may have altered the metric's internal behavior, complicating interpretation of its association with EDPC.

Fourth, to minimize disruption of routine clinical care, our study design permitted missed responses. Little MCAR test rejected the assumption that these responses were MCAR, indicating that missingness may have been systematically related to observed or unobserved factors. For example, clinicians may have been unable to record entries during periods of extreme workload or severe crowding. Although MI produced similar correlation estimates, the imputation model did not explicitly account for within-clinician or serial correlation, and missing-data bias remains possible. Additionally, participation in this study was voluntary, and it remains uncertain whether this active recording approach would be sustainable in routine, nonresearch clinical practice.

Fifth, multiple statistical comparisons were performed without adjustment for multiplicity, and no single primary comparison was prespecified. Therefore, the possibility of inflated type I error cannot be excluded, and the findings should be interpreted with caution.

Conclusion

This prospective observational study provides methodological and empirical insights into measuring ED crowding. Using dedicated smartwatch-based EMA, we captured clinicians' perceptions in situ at high temporal frequency during active shifts, potentially reducing susceptibility to retrospective recall bias while minimizing the spatial constraints of traditional survey methods. During an exceptional health care workforce disruption, hourly median EDPC was moderately associated with EDOR and NEDOCS, whereas its association with EDWIN was less consistent across temporal sensitivity analyses. These findings warrant further investigation of EDPC as a complementary measure of ED crowding alongside established objective metrics. However, our findings do not establish that EDPC directly measures cognitive workload or predicts subsequent crowding. The high response rate (1100/1124, 97.9%) supports the feasibility of wearable-based momentary assessments in acute care. Future studies should evaluate the generalizability of these findings across diverse ED settings and further assess the validity of EDPC as a complementary crowding measure using independently anchored scales, appropriate time-series analyses, and multicenter designs.

Acknowledgments

The authors would like to express their sincere gratitude to the emergency department clinicians who generously participated and contributed to this research.

Funding

This work was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number RS-2024-00335925). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

XSL•FO
RenderX

Data Availability

Data supporting the findings of this study are available from the corresponding author upon request. These data are not publicly available due to institutional and ethical restrictions related to patient and clinician confidentiality.

Generative AI Use

During the preparation of this manuscript, the authors used ChatGPT (GPT-5.5/GPT-6; OpenAI) and Gemini (Gemini 1.5 Pro; Google). According to the Generative AI Delegation Taxonomy [46], the following tasks were delegated to generative AI tools under full human supervision: literature search and systematization, data cleaning, proofreading and editing, translation, and reformatting. All outputs produced with these tools were thoroughly reviewed, verified, and critically evaluated by the authors. Responsibility for the final manuscript lies entirely with the authors. These tools are not listed as authors and do not bear responsibility for the final outcomes.

Authors' Contributions

Conceptualization: WCC, JY
Methodology: all authors
Investigation: JSY
Data curation: JSY
Formal analysis: JSY, SP, SH, JY
Validation: KJM, PBD, SYS, HC, TK, SUL, SH
Writing–original draft: JSY, SP, JY
Writing–review & editing: KJM, PBD, SYS, HC, TK, SUL, SH
Supervision: WCC, JY
Project administration: WCC, JY
Funding acquisition: WCC

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary methods, tables, and figures.
[DOCX File , 7019 KB - mhealth_v14i1e103317_app1.docx]

References

1. Ahmed AE, Aqeel A, Ghazwani M, Alenzi AN, Alruwaili FM, Alsharari AS, et al. Clinical and operational effects of emergency department crowding: a systematic review. Cureus 2026;18(1):e100560. [doi: 10.7759/cureus.100560] [Medline: 41625046]
2. Asplin BR, Magid DJ, Rhodes KV, Solberg LI, Lurie N, Camargo CA. A conceptual model of emergency department crowding. Ann Emerg Med 2003;42(2):173-180. [doi: 10.1067/mem.2003.302] [Medline: 12883504]
3. Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L. Emergency department crowding: a systematic review of causes, consequences and solutions. PLoS One 2018;13(8):e0203316 [FREE Full text] [doi: 10.1371/journal.pone.0203316] [Medline: 30161242]
4. American College of Emergency Physicians (ACEP). Crowding. Policy statement. Ann Emerg Med 2013;61(6):726-727. [doi: 10.1016/j.annemergmed.2013.03.037] [Medline: 23684339]
5. Kim M, Lee S, Choi M, Kim D, Yoo J, Shin TG, et al. Factors that predict emergency department length of stay in analysis of national data. Clin Exp Emerg Med 2025;12(1):35-46 [FREE Full text] [doi: 10.15441/ceem.24.309] [Medline: 39414332]
6. Haimovich AD, Berry SD, Landon BE. Prolonged emergency department stays for older US adults. JAMA Intern Med 2025;185(8):1037-1038. [doi: 10.1001/jamainternmed.2025.2006] [Medline: 40587143]
7. van Dam CS, Peters MJL, Hoogendijk EO, Nanayakkara PWB, Muller M, Trappenburg MC. Older patients with nonspecific complaints at the emergency department are at risk of adverse health outcomes. Eur J Intern Med 2023;112:86-92 [FREE Full text] [doi: 10.1016/j.ejim.2023.03.018] [Medline: 37002150]
8. Mohr NM, Wessman BT, Bassin B, Elie-Turenne MC, Ellender T, Emlet LL, et al. Boarding of critically ill patients in the emergency department. J Am Coll Emerg Physicians Open 2020;1(4):423-431 [FREE Full text] [doi: 10.1002/emp2.12107] [Medline: 33000066]
9. Oskvarek JJ, Zocchi MS, Black BS, Celedon P, Leubitz A, Moghtaderi A, US Acute Care Solutions Research Group. Emergency department volume, severity, and crowding since the onset of the coronavirus disease 2019 pandemic. Ann Emerg Med 2023;82(6):650-660. [doi: 10.1016/j.annemergmed.2023.07.024] [Medline: 37656108]

XSL•FO
RenderX

10. Jayaprakash N, Pflaum-Carlson J, Gardner-Gray J, Hurst G, Coba V, Kinni H, et al. Critical care delivery solutions in the emergency department: evolving models in caring for ICU boarders. Ann Emerg Med 2020;76(6):709-716. [doi: 10.1016/j.annemergmed.2020.05.007] [Medline: 32653331]
11. Badr S, Nyce A, Awan T, Cortes D, Mowdawalla C, Rachoin J. Measures of emergency department crowding, a systematic review. How to make sense of a long list. Open Access Emerg Med 2022;14:5-14 [FREE Full text] [doi: 10.2147/OAEM.S338079] [Medline: 35018125]
12. Clouston R, Atkinson P, Canales DD, Fraser J, Sohi D, Lee S, et al. Emergency department occupancy is useful as a simple real-time measure of crowding. CJEM 2022;24(1):23-26. [doi: 10.1007/s43678-021-00098-8] [Medline: 33748940]
13. Altun M, Kudu E. Overcrowding in emergency departments: a scoping literature review. Anatolian J Emerg Med 2025;8:94-100. [doi: 10.54996/anatolianjem.1684632]
14. Jones P, Wells S, Ameratunga S. Towards a best measure of emergency department crowding: lessons from current Australasian practice. Emerg Med Australas 2018;30(2):214-221. [doi: 10.1111/1742-6723.12868] [Medline: 28941074]
15. Mol S, Gaakeer M, van der Linden MC, Baan-Kooman ECM, Backus BE, de Ridder VA. Crowding, perceived crowding and workload in Dutch emergency departments: should we continue on the same road? Eur J Emerg Med 2023;30(4):229-230. [doi: 10.1097/MEJ.0000000000001034] [Medline: 37115965]
16. McCarthy ML, Aronsky D, Jones ID, Miner JR, Band RA, Baren JM, et al. The emergency department occupancy rate: a simple measure of emergency department crowding? Ann Emerg Med 2008;51(1):15-24, 24.e1. [doi: 10.1016/j.annemergmed.2007.09.003] [Medline: 17980458]
17. Weiss SJ, Derlet R, Arndahl J, Ernst AA, Richards J, Fernández-Frackelton M, et al. Estimating the degree of emergency department overcrowding in academic medical centers: results of the National ED Overcrowding Study (NEDOCS). Acad Emerg Med 2004;11(1):38-50 [FREE Full text] [doi: 10.1197/j.aem.2003.07.017] [Medline: 14709427]
18. Bernstein SL, Verghese V, Leung W, Lunney AT, Perez I. Development and validation of a new index to measure emergency department crowding. Acad Emerg Med 2003;10(9):938-942 [FREE Full text] [doi: 10.1111/j.1553-2712.2003.tb00647.x] [Medline: 12957975]
19. Boverio R, Signorini F, Rossi C, Bellini RIM, Zanelli C, Ghilardi GI, FENICE network. Measuring emergency department crowding: comparing the NEDOCS and Fenice crowding scores against staff perception. Intern Emerg Med 2025. [doi: 10.1007/s11739-025-04216-7] [Medline: 41385153]
20. García-Romero M, Rita-Gáfaro CG, Quintero-Manzano J, Bermon A. Escala NEDOCS vs valoración subjetiva, ¿El personal de salud en urgencias es consciente de su sobrecupo? Colomb Med 2017;48(2):53-57. [doi: 10.25100/cm.v48i2.2131]
21. Muir KJ, McHugh MD, Merchant RM, Lasater KB. Left without being seen: nurse work environment and timely outcomes in New York and Illinois emergency departments. J Emerg Nurs 2024;50(5):660-669. [doi: 10.1016/j.jen.2023.11.010] [Medline: 38127046]
22. Muir KJ, Golinelli D, Lasater KB, Aiken LH, Merchant R, McHugh MD, et al. Association of emergency department and medical-surgical nurse staffing on patient departures against medical advice. J Emerg Nurs 2026;S0099-1767(25):00510-0 [FREE Full text] [doi: 10.1016/j.jen.2025.12.009] [Medline: 41575406]
23. Heslin SM, Schwaner R, Viccellio P, King C, Rowe A, Thode H, et al. A physician-nurse team adjacent to triage shortens length of stay and left-without-being-seen rates. Clin Exp Emerg Med 2025;12(1):26-34 [FREE Full text] [doi: 10.15441/ceem.24.248] [Medline: 39237140]
24. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
25. Myin-Germeys I, Kasanova Z, Vaessen T, Vachon H, Kirtley O, Viechtbauer W, et al. Experience sampling methodology in mental health research: new insights and technical developments. World Psychiatry 2018;17(2):123-132 [FREE Full text] [doi: 10.1002/wps.20513] [Medline: 29856567]
26. de Vries LP, Baselmans BML, Bartels M. Smartphone-based ecological momentary assessment of well-being: a systematic review and recommendations for future studies. J Happiness Stud 2021;22(5):2361-2408 [FREE Full text] [doi: 10.1007/s10902-020-00324-7] [Medline: 34720691]
27. Yoon JH, Kwon IH, Park HW. The South Korean health-care system in crisis. Lancet 2024;403(10444):2589. [doi: 10.1016/S0140-6736(24)00766-9] [Medline: 38879246]
28. Park HA, Kim T, Kim HJ, Han S. Epidemiological trends in emergency department visits by age group: a report from the National Emergency Department Information System (NEDIS) of Korea, 2020-2024. Clin Exp Emerg Med 2025;12(4):405-413 [FREE Full text] [doi: 10.15441/ceem.25.291] [Medline: 41531412]
29. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. Br J Math Stat Psychol 2008;61(Pt 1):29-48. [doi: 10.1348/000711006X126600] [Medline: 18482474]
30. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
31. R: a language and environment for statistical computing. R Core Team. 2025. URL: https://www.R-project.org [accessed 2026-10-23]
32. Altun M, Kudu E, Yakin F, Korgan MB, Unal E, Demir O, et al. Impact of transition units on emergency department crowding: a before-and-after study. J Emerg Med 2026;81:68-77. [doi: 10.1016/j.jemermed.2025.12.007] [Medline: 41581296]

XSL•FO
RenderX

33. Strada A, Bravi F, Valpiani G, Bentivegna R, Carradori T. Do health care professionals' perceptions help to measure the degree of overcrowding in the emergency department? A pilot study in an Italian university hospital. BMC Emerg Med 2019;19(1):47 [FREE Full text] [doi: 10.1186/s12873-019-0259-9] [Medline: 31455226]
34. Wang H, Kline J, Jackson B, Robinson RD, Sullivan M, Holmes M, et al. The role of patient perception of crowding in the determination of real-time patient satisfaction at emergency department. Int J Qual Health Care 2017;29(5):722-727. [doi: 10.1093/intqhc/mzx097] [Medline: 28992161]
35. Schippers EE, Willemsen SP, Collée LAM, Koenraadt KLM, Schol-Gelok S. Crowding in real-time: an artificial intelligence-based model to manage crowding in the emergency department (CRAMMED Study). BMC Emerg Med 2026;26(1):232 [FREE Full text] [doi: 10.1186/s12873-026-01647-z] [Medline: 42310553]
36. Chen SM, Stewart G, Sentman E, Helwig S, Rust L, Hoffman J, et al. Validation of the emergency department work index in a pediatric freestanding community emergency department. Pediatr Emerg Care 2024;40(10):705-710. [doi: 10.1097/PEC.0000000000003247] [Medline: 39051989]
37. Peltan ID, Bledsoe J, Jacobs J, Groat D, Klippel C, Adamson M, et al. Effectiveness and safety of an emergency department code sepsis protocol: a pragmatic clinical trial. Ann Am Thorac Soc 2024;21(11):1560-1571 [FREE Full text] [doi: 10.1513/AnnalsATS.202403-286OC] [Medline: 38996086]
38. Liu Q, Xie C, Tan J, Xu L, Zhou F, Peng L. Exploring the nurses' experiences in recognising and managing clinical deterioration in emergency patients: a qualitative study. Aust Crit Care 2024;37(2):309-317. [doi: 10.1016/j.aucc.2023.06.004] [Medline: 37455210]
39. Park S, Yoo J, Lee Y, DeGuzman PB, Kang M, Dykes PC, et al. Quantifying emergency department nursing workload at the task level using NASA-TLX: an exploratory descriptive study. Int Emerg Nurs 2024;74:101424. [doi: 10.1016/j.ienj.2024.101424] [Medline: 38531213]
40. Martin N, Bergs J, Eerdekens D, Depaire B, Verelst S. Developing an emergency department crowding dashboard: a design science approach. Int Emerg Nurs 2018;39:68-76. [doi: 10.1016/j.ienj.2017.08.001] [Medline: 28865753]
41. Gesner E, Dykes PC, Zhang L, Gazarian P. Documentation burden in nursing and its role in clinician burnout syndrome. Appl Clin Inform 2022;13(5):983-990 [FREE Full text] [doi: 10.1055/s-0042-1757157] [Medline: 36261113]
42. Shanafelt TD, Dyrbye LN, Sinsky C, Hasan O, Satele D, Sloan J, et al. Relationship between clerical burden and characteristics of the electronic environment with physician burnout and professional satisfaction. Mayo Clin Proc 2016;91(7):836-848. [doi: 10.1016/j.mayocp.2016.05.007] [Medline: 27313121]
43. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
44. Vachon H, Viechtbauer W, Rintala A, Myin-Germeys I. Compliance and retention with the experience sampling method over the continuum of severe mental disorders: meta-analysis and recommendations. J Med Internet Res 2019;21(12):e14475 [FREE Full text] [doi: 10.2196/14475] [Medline: 31808748]
45. Peabody JE, Ryznar R, Ziesmann M, Gillman L. A systematic review of heart rate variability as a measure of stress in medical professionals. Cureus 2023;15(1):e34345 [FREE Full text] [doi: 10.7759/cureus.34345] [Medline: 36865953]
46. Suchikova Y, Tsybuliak N, Teixeira da Silva JA, Nazarovets S. GAIDeT (generative AI delegation taxonomy): a taxonomy for humans to delegate tasks to generative artificial intelligence in scientific research and publishing. Account Res 2026;33(3):2544331. [doi: 10.1080/08989621.2025.2544331] [Medline: 40781729]

Abbreviations

AC2: Gwet weighted agreement coefficient
AR(1): first-order autoregressive
DARWIN-C: Samsung Medical Center’s Clinical Data Warehouse
ED: emergency department
EDOR: emergency department occupancy ratio
EDPC: emergency department perceived crowding
EDWIN: Emergency Department Work Index
EHR: electronic health record
EMA: ecological momentary assessment
GLS: generalized least squares
IRB: institutional review board
KTAS: Korean Triage and Acuity Scale
LAD: least absolute deviations
MCAR: missing completely at random
MI: multiple imputation
NEDOCS: National Emergency Department Overcrowding Score

XSL•FO
RenderX

Edited by S Brini; submitted 04.Jun.2026; peer-reviewed by F Signorini, E Kudu; comments to author 08.Aug.2026; revised version received 20.Oct.2026; accepted 21.Oct.2026; published 10.Nov.2026.

Please cite as:
Yoon JS, Park S, Muir KJ, DeGuzman PB, Shin SY, Chang H, Kim T, Lee SU, Hong S, Cha WC, Yoo J
Smartwatch-Measured Clinician Perceptions of Emergency Department Crowding and Their Association With Objective Metrics: Prospective Observational Study
JMIR Mhealth Uhealth 2026;14:e103317
URL: https://mhealth.jmir.org/2026/1/e103317
doi:10.2196/103317
PMID:

©Ji Sim Yoon, Sookyung Park, K Jane Muir, Pamela B DeGuzman, So Yeon Shin, Hansol Chang, Taerim Kim, Se Uk Lee, Sungjun Hong, Won Chul Cha, Junsang Yoo. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Platform Differences in Smartphone Sensing Features for Mental Health Across Four Annual College Cohorts: Secondary Analysis

Ariz Naqvi[1*], PhD; Farah Naqvi[2*], PhD

[1]College of Business Administration, American University of the Middle East, Egaila, Kuwait

[2]School of Psychology, University of Wolverhampton, Wolverhampton, England, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Ariz Naqvi, PhD
College of Business Administration, American University of the Middle East
Block 6, Building 1, Street 250
Egaila, 54200
Kuwait
Phone: 965 97952687
Email: ariz.naqvi@aum.edu.kw

Abstract

Background: Passive sensing for mobile mental health relies on observations that can differ across smartphone platforms. Distinguishing behavioral differences from differences in observation is necessary before using these features for symptom screening or monitoring.

Objective: We quantified platform-associated differences, feature availability, sensitivity to preprocessing, and the predictive contribution of connectivity features within the Generalization of Longitudinal Behavior Modeling (GLOBEM) dataset and its collection and processing pipeline.

Methods: We analyzed 4 annual US college cohorts from 2018 to 2021, comprising 705 annual records from 497 people reported by the dataset investigators. The feature sample contained 554 iOS and 151 Android records. Without verified cross-year identity links, inference and cross-validation were performed separately within each cohort. We compared 15 raw constructs and matched transformations using Cliff δ, participant resampling, and fixed logistic regression. Phone-derived location and screen features were separated from Fitbit activity and sleep measures. Analyses included coverage, complete cases, psychological adjustment, and retrospective depressive-symptom screening using the preceding 14 days.

Results: Total Bluetooth scan counts were lower on iOS in every cohort (Cliff δ −1.00 to −0.85). Connectivity counts differed more between platforms than core phone behavioral features did (difference in mean absolute Cliff δ 0.22-0.43; all 95% CIs above zero). Smaller observed-value differences did not imply equal coverage: 41 of 103 iOS records lacked all home-time values in 2020, compared with 0 of 34 Android records. Platform-classification area under the receiver operating characteristic curve (AUROC) was 0.94 to 0.997 for 13 raw constructs, 0.34 to 0.91 after normalization, and 0.65 to 0.80 after discretization. Bluetooth differences persisted after psychological adjustment. In 661 eligible screening windows, behavioral-model AUROC ranged from 0.41 to 0.56. All behavioral, connectivity-added, and platform-added models had worse Brier scores than the intercept-only benchmark. Adding connectivity changed AUROC by 0.004 to 0.034; every paired 95% CI included zero.

Conclusions: Within GLOBEM, selected connectivity features showed larger platform-associated differences than core phone behavioral features, while availability problems also affected location. Weak screening performance limits what these comparisons can establish about downstream harm. Platform-specific coverage and outcome calibration require assessment before interpreting sensing features as clinical markers.

(JMIR Mhealth Uhealth 2026;14:e103711) doi:10.2196/103711

KEYWORDS

mobile health; passive sensing; smartphone; mental health; depressive symptoms; measurement; data quality; operating system; digital phenotyping; wearable electronic devices

XSL•FO
RenderX

Introduction

Measurement in Mobile Mental Health

A low smartphone-derived activity or communication value can reflect behavior, limited observation, or both. This ambiguity matters when mobile health (mHealth) apps use passive sensing to supplement depression screening or monitor changes between clinical assessments. Early research linked location and phone-use features to depressive symptoms [1]. Subsequent work used communication traces to model clinically assessed mood and anxiety symptoms [2], and screen and connectivity features to classify questionnaire-defined depression [3]. These studies motivate careful interpretation of the measurements supplied to a model, rather than treating a familiar feature name as evidence of comparable observation across devices.

Communication and proximity features are used as proxies for social engagement and withdrawal. Smartphone and Fitbit studies have related mobility, activity, and communication to loneliness [4]. Yet phone channels need not move together: Wang and colleagues [5] found fewer incoming contact calls but more frequent social-app use among Android users with depression-screen-positive scores. A call count is a recorded event count, and Bluetooth detections identify nearby devices rather than verified interpersonal contact. The distinction between a trace, a behavioral marker, and a clinical state is central to personal sensing models [6].

Monitoring an individual and ranking different individuals require different measurement evidence. Student studies have used repeated phone and wearable observations to model weekly symptoms and end-of-semester symptom changes [7,8]. A stable device-specific offset could affect between-person comparisons while preserving some within-person information. Intermittent observation or device replacement could affect both. Our annual platform comparisons address the former question; they do not evaluate personalized symptom trajectories.

Platform differences in observation are already documented. Halabi and colleagues [9] analyzed multimodal sensor data from approximately 3000 participants in real-world settings and found that data-quality features could identify Android versus iOS with accuracy as high as 0.98. The Reproducible Analysis Pipeline for Data Streams (RAPIDS), a toolkit for processing mobile sensing data, provides a common analysis framework, but its feature documentation describes platform-specific event handling [10]. Common output names therefore need not imply identical operational definitions. A systematic review of smartphone-only health sensing documented uneven operating-system coverage and incomplete platform reporting [11].

Prediction and measurement are distinct questions. Mobility-based depression prediction has performed differently across homogeneous student samples and heterogeneous populations [12]. The Generalization of Longitudinal Behavior Modeling (GLOBEM) benchmark also examined difficulties in generalizing longitudinal behavioral models across datasets [13]. Such findings do not identify platform as the cause. Conversely, high platform-classification performance establishes information about platform in a feature set, but does not establish that this information worsens a mental health model. Clinical relevance requires evaluation of the intended outcome and use case [14].

The GLOBEM benchmark restricted cross-dataset models to compatible feature classes and excluded Bluetooth, WiFi, and call features that were not common across its 2-institution datasets [13]. In its reimplementation of 2 earlier algorithms, most of the performance loss relative to the original studies was attributed to removing Bluetooth, WiFi, call, message, and battery features [13]. This practical loss of potentially informative communication and proximity measures motivates a feature-level comparability audit. That benchmark used 2 institutions across 2018 and 2019; the present analysis uses 4 annual cohorts from one institution and a different screening protocol.

Measurement Comparability and Predictive Use

Measurement comparability concerns observable events, feature definitions, and health inference. Goldsack and colleagues [15] distinguish technology verification, analytical validation, and clinical validation. Released features support an availability audit and outcome comparison, but simultaneous reference observations would be required to validate sensor accuracy or interchangeability.

Missingness may carry both technical and clinical information. Currey and Torous [16] found that coverage varied with app engagement and other factors in 1178 participants. In a small schizophrenia cohort, data-quality metrics were associated with symptom-related survey responses [17]. Comparing only observed means can therefore omit a relevant difference in the observation process.

Luijken and colleagues [18] showed analytically and through simulation that predictor-measurement differences between development and validation can affect performance. This establishes a possible pathway, not its presence in GLOBEM. We therefore evaluate platform recovery and symptom screening as separate tasks, with distinct outcomes and interpretations.

Between-person and within-person associations can differ. Yang and colleagues [19] separated app-use predictors into individual deviations and person averages and found different associations with loss of interest. Their Android study did not test cross-platform harmonization. It demonstrates why a classifier recovering platform from participant averages cannot establish whether deviations from an individual's baseline predict symptoms.

Objectives and Scope

We used 4 annual GLOBEM college cohorts to characterize platform-associated differences in released sensing features [20,21]. Our objectives were to locate these differences, distinguish phone-derived features from Fitbit measures, assess coverage and preprocessing, and evaluate whether adding selected connectivity features altered retrospective depressive-symptom screening. We use "platform-associated difference" for an observed contrast. We reserve "measurement artifact" for a documented incompatibility in collection or

XSL•FO
RenderX

feature meaning. We do not infer causal confounding, psychological independence, or measurement equivalence from platform classification alone.

We evaluated localization, contribution to platform prediction, sensitivity to transformation, and persistence after psychological adjustment. Coverage, matched-feature, and downstream analyses were exploratory extensions described in the "Methods" section. This secondary analysis was not preregistered and does not validate a clinical device.

Methods

Data Source and Analytical Units

The GLOBEM dataset, version 1.1, released as Multi-Year Datasets for Longitudinal Human Behavior Modeling Generalization, contains annual cohorts INS-W_1 through INS-W_4, collected during spring terms from 2018 through 2021 at a US research university. The dataset investigators report 155, 218, 137, and 195 annual participants, representing 705 person-years from 497 unique people. They collected phone data with an app built on the open-source AWARE mobile sensing framework and activity and sleep data with Fitbit, then released derived features processed with RAPIDS [20,21]. We obtained the supplied feature and survey files from the credentialed PhysioNet release [22]. No new participants were recruited for this secondary analysis.

The investigators recruited undergraduates by email and invited previous participants to return. Each study covered a 10-week spring quarter. Participants installed the app on their own phones, received a Fitbit, and completed prestudy, within-study, and poststudy surveys [20,21]. Smartphone ownership was not assigned. Individual demographic variables were unavailable in the supplied analytical files, preventing platform-specific demographic summaries and adjustment.

The release shifts dates by whole weeks to preserve privacy, within-person temporal order, and weekdays [20]. We used these dates for participant-specific survey alignment, not to compare external calendar events. Scheduled duration, released date span, and observed-day count are different quantities. An annual record represents one cohort contribution, not continuous observation for a calendar year.

Our feature sample comprised all annual identifiers in the released all-day feature tables that matched platform metadata. We validated one-to-one platform and survey joins and unique participant-date feature rows. The broader platform files contained 818 records, of which 113 had no corresponding feature record. Their absence is a coverage difference in the supplied files; we could not determine that it represented dropout. The "Results" section distinguishes metadata, feature, and screening samples.

The 705 annual records had 705 distinct literal identifier strings. Public overlap lists and their consuming code identified membership subsets but did not establish individual correspondence across years [23]. We therefore did not infer links from list position, identifier suffixes, or participant measurements. All inferential analyses and model splits were conducted within individual annual cohorts. We neither fitted pooled models assuming 705 independent people nor combined annual effects in an independence-based meta-analysis. Cross-year agreement is descriptive because some underlying people returned. A switcher count and a within-person platform-switching analysis could not be established from verified links.

Ethical Considerations

The original GLOBEM study received University of Washington Institutional Review Board approval (STUDY00003244), and participants provided informed consent before enrollment. Compensation was up to US $245 per quarter, depending on compliance [20,21]. The published consent forms described anonymized data sharing with university collaborators in 2018 and anonymous publication or disclosure to external researchers for further analysis and research in 2019-2021 [21]. We used the controlled-access, deidentified feature-level release and did not seek additional participant consent.

No separate ethics evaluation was undertaken for this secondary analysis. The rationale for not seeking additional review was that the analysis involved existing deidentified records, without participant interaction or access to direct identifiers. This rationale was informed by the definition of a human subject in the US Federal Policy for the Protection of Human Subjects (45 CFR 46.102(e)) [24]. The analysis was limited to existing, deidentified feature-level records released for research through controlled access, without participant recruitment, contact, or intervention. Access was obtained through PhysioNet registration and credentialing after completion of the required CITI Data or Specimens Only Research training, under the dataset's data-use agreement [20]. We did not have access to direct identifiers or attempt reidentification. Only aggregate findings are reported, and the supplementary materials exclude participant-level records and predictions. The source investigators restricted the release to derived features because sensor traces can reveal identities; deidentification does not eliminate all privacy risk [20].

Feature Definitions and Preprocessing

We retained the starting panel-selection rule: up to 6 all-day columns per sensor domain, selected alphabetically from keyword-matched candidates. This deterministic rule yielded 42 columns and did not use mental health outcomes or model performance. It contained 15 raw constructs, 13 normalized versions, and 14 discretized versions. We retained those raw constructs and added their 3 omitted counterparts for matched comparisons, yielding 45 columns. The raw constructs comprised 2 location features (home time and location entropy), 3 screen features (mean unlock duration overall, during exercise-classified locations, and during green-space locations), 2 Bluetooth counts, 2 incoming-call features, 2 connected-WiFi features, 2 Fitbit step summaries, and 2 Fitbit main-sleep summaries. Multimedia Appendix 1 defines each construct and records collection limitations, whereas Multimedia Appendix 2 supplies the complete column map and numerical outputs.

Alphabetical selection limits the panel's representativeness; other summaries could yield different domain contrasts. Holding

XSL•FO
RenderX

the 15 constructs fixed permits transformation comparisons without changing construct composition. We evaluated the 42-column panel as a supplementary diagnostic. The 45 matched columns represent 15 constructs in 3 forms, not independent measurements.

We distinguished 5 phone location/screen constructs from 4 Fitbit constructs. A core phone set contained home time, location entropy, and mean unlock duration without location context. The 2-event connectivity set contained total Bluetooth scans and incoming-call count. The 4-feature connectivity set additionally included scans of the least-frequently detected device and incoming calls from the most frequent contact. The latter contact feature was retained only as an artifact diagnostic: RAPIDS documentation states that iOS identifiers do not preserve repeated-contact identity [25]. It was excluded from downstream screening models.

Historical RAPIDS documentation also uses different screen-episode end points on Android and iOS [26]. Its Bluetooth provider documentation lists Android support, although the GLOBEM exports contain iOS values [27]. We could not verify the historical study-specific adaptation from derived features. Connected-WiFi documentation supports both platforms, but the supplied iOS WiFi values were entirely missing in the first 2 cohorts [28]. These observations preclude assuming that all iOS calls or Bluetooth values were absent, or attributing every observed contrast to a particular operating system restriction. Location estimates also depend on sampled records and feature construction [29].

Home time reflects sampled locations and inferred home, rather than self-reported withdrawal. Location entropy measures dispersion across observed locations in natural logarithm units. Mean unlock duration describes recorded episodes, not all attention to a phone. Bluetooth scans can repeatedly detect the same object; incoming-call counts depend on the events retained by the collection framework and do not measure all communication. Fitbit steps summarize recorded intraday intervals rather than total daily steps, and sleep measures describe wearable-estimated main sleep. The behavioral interpretation remains conditional on these definitions.

We averaged each feature over a participant's observed daily values, leaving an entirely unobserved feature missing. Each participant contributed equally to the between-person comparisons despite unequal observed-day counts. Zero remained a recorded value. Discretized labels (l/m/h or low/medium/high) were explicitly mapped to 0/1/2 before averaging; unrecognized labels triggered an error. The released discretized values are text labels, so generic numeric conversion would treat them as missing values rather than ordered categories. We used the supplied normalized and discretized features. The release defines normalization relative to each participant's median and 5th-to-95th percentile range, and discretization using participant-specific 33rd and 66th percentiles [20]. These are descriptions of released transformations, not transformations estimated from a prediction training fold.

These transformations are retrospective: their reference distributions may include observations outside a screening window, so they were excluded from downstream screening. Averaged ordinal codes describe the observed mix of low, middle, and high categories, not a recovered continuous behavioral scale. Ties and missing days can affect this summary.

Coverage and Feature Comparisons

For every feature and platform, we calculated the number of participants with any value, the proportion of missing released daily entries, observed days per participant, the proportion of observed values equal to zero, and within-person daily variance. The released tables contained 92, 97, 103, and 103 daily rows per participant across the 4 cohorts. These denominators describe the release, not verified sensor uptime or enrollment duration. A sensitivity analysis restricted dates to the interval between each participant's prestudy and poststudy questionnaires, inclusive. These questionnaire dates were not assumed to be exact enrollment boundaries.

Daily missingness divided missing entries by released daily rows; participant-level absence divided records lacking every daily value by the annual platform sample. Observed-day summaries included zero-day participants. Variance used at least 2 observations and the sample-variance denominator. Zero-valued observations remained observed, whereas missing values remained missing. Absent context-specific screen values could also reflect the absence of that context.

Within each cohort, we compared observed participant means with 2-sided Mann-Whitney tests and Cliff δ [30]. Cliff δ was defined as the probability that an iOS value exceeded an Android value minus the reverse probability; ties contributed zero. Negative values indicate lower iOS values. We used 2000 platform-stratified participant resamples for 95% percentile intervals. A bounded-difference interval replaced a degenerate bootstrap interval when all resampled estimates coincided. We applied Benjamini-Hochberg adjustment [31] separately within each cohort's raw, normalized, and discretized feature families; the latter 2 were exploratory. We did not classify small deltas as equivalence.

Cliff δ describes ordering on a −1 to 1 scale; medians and IQRs give original units. Entirely absent features were excluded from their observed-value comparisons.

Stratified resampling preserved platform counts and each person's joint feature vector. Complete empirical rank separation can yield a degenerate bootstrap distribution. The bounded-difference fallback avoids presenting that point mass as population certainty; its formula and derivation are in Multimedia Appendix 1. It does not correct selection or omitted-variable bias. Multiplicity adjustment was within cohort and feature family, without a claim of confirmatory testing or universal error control under arbitrary dependence.

To assess localization without treating correlated features as independent observations, we calculated the difference between the mean absolute Cliff δ in a fixed connectivity set and that in a fixed comparator set. Each participant resample preserved the full feature vector. The main phone-only contrast compared the 2-event connectivity set with the 3 core phone measures, excluding both Fitbit and the invalid contact-identity construct. Comparisons with all 5 phone behavioral measures, Fitbit, and

XSL•FO
RenderX

the 4 connectivity constructs were sensitivity analyses. Across-year averages of absolute deltas were descriptive and were kept distinct from absolute averages of signed deltas.

Platform Prediction and Missing-Data Sensitivity

We used fixed L2-penalized logistic regression (C=1; LIBLINEAR solver, as implemented in scikit-learn [32]) to predict iOS versus Android separately in each cohort. Models used 5 stratified folds, training-fold median imputation, and training-fold standardization. No hyperparameter selection was performed. An entirely missing training column was retained using the imputer's empty-feature handling. We calculated the area under the receiver operating characteristic curve (AUROC) from combined out-of-fold predictions, with iOS as the positive class. Values below 0.50 were reported without post hoc reversal.

Split seeds were identical across feature sets. Entirely missing training columns used the imputer's constant fallback, without held-out values.

Missingness-only models indicate entirely absent participant features, not intermittent daily absence. Missing-at-random assumptions cannot be established from observed rates alone [33].

Validation used one record per annual participant, matching the intended between-person comparison [34]. Bootstrap duplicates stayed together in a fold. We excluded leave-one-year-out validation because unidentified returning participants could occur in both training and evaluation years. No reported estimate represents validation in entirely new people at a new site or time.

The main matched comparison used 13 raw constructs excluding WiFi, their 13 normalized counterparts, and their 13 discretized counterparts. A distinct removal sequence contained 13 constructs, 12 after removing the contact feature, 10 after also removing both Bluetooth features, and 9 after removing the remaining incoming-call count. The final set comprised 5 phone behavioral and 4 Fitbit features. We also evaluated the 3 core phone measures, Fitbit alone, and indicators that participant-level features were entirely missing. WiFi-inclusive and 42-column models were diagnostic analyses rather than the primary comparison.

We repeated raw-panel classification in complete cases and compared all 3 transformations in participants complete on all 39 matched columns. Complete-case results refer to a selected population, so their difference from full-sample results cannot isolate the effect of imputation. For the main raw, normalized, discretized, and 9-feature behavioral models, uncertainty used 500 participant bootstrap samples with preprocessing and cross-validation refitted. Missingness-only and raw complete-case classifiers also used refitted resamples. Duplicate copies of a resampled participant stayed together in a fold. Other diagnostic models, including common-complete-case transformation comparisons, used conditional resampling of fixed out-of-fold predictions, explicitly labeled in the supplementary outputs. No interval treated 5-fold scores as 5 independent observations. Intervals are approximate, conditional on the observed stratum sizes, and can be unstable with small platform groups or near-perfect rankings.

Requiring all 39 columns holds participants constant across transformations. Complete-record logistic coefficients can be unbiased under specified missingness structures [35], but this does not guarantee representative AUROC. Platform-specific retention therefore remains part of the interpretation.

Overlapping training sets make fold errors dependent [36]. The outputs identify the resampling method and usable replicates; conditional intervals omit model-refitting variation.

Psychological Comparisons and Adjustment

We compared prestudy loneliness, perceived stress, resilience, emotion-regulation reappraisal and suppression, and depressive symptoms across platforms using the released common-version scores. Cohort-specific anxiety scores were compared separately: the State-Trait Anxiety Inventory–State (STAI-S) score in the first cohort and the released State-Trait Anxiety Inventory (STAI) score in the remaining cohorts. Benjamini-Hochberg adjustment covered the 7 score comparisons within each cohort. We did not interpret nonsignificant tests as psychological equivalence.

The common adjustment set comprised the 10-item University of California, Los Angeles Loneliness Scale (UCLA-10), the 10-item Perceived Stress Scale (PSS-10), the Brief Resilience Scale (BRS), Emotion Regulation Questionnaire (ERQ) reappraisal and suppression scores, and the released 9-item Center for Epidemiologic Studies Depression (CESD-9) score. Higher scores indicate more of the named construct: greater loneliness, perceived stress, resilience, reappraisal, suppression, or depressive symptoms, respectively. Anxiety used the cohort-specific released instrument rather than assuming that the state and other anxiety versions were identical. We analyzed the supplied scores, without reconstructing questionnaire items or claiming new psychometric validation in this sample.

For total Bluetooth scans and incoming-call counts, we regressed the standardized log (1 + participant mean) on an iOS indicator, first without and then with the 6 common standardized psychological scores. Both models used the same complete-case sample within each cohort. We used ordinary least squares with heteroskedasticity-consistent SEs using the HC3 covariance estimator and t-based intervals [37]. An additional specification included the cohort-specific anxiety score. The contact feature was analyzed separately as an artifact diagnostic. We checked model rank, conditioning, and leverage. This analysis tests sensitivity to measured psychological adjustment; it does not control for income, age, device price, or other unavailable ownership-related characteristics.

Android was the reference category. Coefficients describe SD differences in log-transformed participant means, not raw-count differences. Psychological coefficients were not treated as mediators.

We quantified sensitivity of the adjusted platform coefficients to omitted variables using the point-null robustness value of Cinelli and Hazlett [38]. We refitted the same regressions with classical ordinary least squares covariance for this calculation;

XSL•FO
RenderX

the main coefficient intervals retain HC3 covariance. The robustness value summarizes the equal partial R-squared an omitted factor would need with platform and the outcome, conditional on the specified controls, to explain away the coefficient. It does not estimate actual demographic confounding or test the measurement mechanism. Multimedia Appendix 1 gives the formula and assumptions.

Exploratory Depressive-Symptom Screening

We used the released end-of-term Beck Depression Inventory-II (BDI-II) label, defined by a score greater than 13, as a symptom-screening outcome rather than a clinical diagnosis. We verified agreement between the released label and the score threshold. Predictors were means of raw daily features from the 14 calendar days strictly before the questionnaire date. A feature required at least 7 observed days in that window. Participants were eligible if at least 1 of the 9 candidate predictors met this criterion. Models therefore shared an eligibility rule, with remaining missing features imputed within training folds.

A common end-of-term measure avoided mixing heterogeneous weekly targets, including the first cohort's derived labels [20]. Neither the 14-day window nor the 7-day rule was tuned against performance. Requiring one qualifying predictor retained partial multimodal records, making missingness part of the evaluated use case.

The behavioral model included the 3 core phone measures and 4 Fitbit measures. We compared this with a model adding total Bluetooth scans and incoming-call count, then a model additionally including platform. Intercept-only, platform-only, and core-phone models were supplementary comparators. Contact identity, across-dataset device-selection features, context-specific screen features, and WiFi were excluded. We used the same fixed classifier and participant-level folds, stratifying jointly by outcome and platform in 2019 to 2021 and by outcome alone in 2018, when the smallest joint subgroup had only 2 records.

We reported overall and platform-specific AUROC, average precision, Brier score, log loss, symptom prevalence, and mean predicted probability. Calibration-in-the-large was estimated only in groups with at least 20 participants and at least 5 positive and 5 negative outcomes. Paired participant resampling with model refitting estimated changes after adding connectivity and platform. Resampling preserved outcome-by-platform stratum counts, limiting the uncertainty represented for prevalence and calibration. We used Fisher exact tests for platform versus screening-label associations and, separately, for metadata records lacking all feature records. The latter 4 exploratory tests are reported without multiplicity adjustment. These exploratory models evaluate retrospective screening; presurvey feature windows do not establish prospective validity because upstream released feature construction can use broader participant context.

AUROC measures ranking, whereas average precision depends on outcome prevalence and summarizes precision across recall [39]. Brier score measures squared probability error; log loss penalizes confident errors. Lower values of both are preferable. Metrics used held-out predictions, without selecting a clinical referral threshold.

We estimated calibration-in-the-large by fitting a logistic intercept with the logit of held-out probability as an offset, fixing its coefficient at 1 [40]. We also reported observed/expected ratios, calculated as observed positive outcomes divided by the sum of predicted probabilities. Ideal values are 0 and 1, respectively. These are descriptive estimates without CIs. The minimum-size rule applied to the intercept fit; ratios remained reportable in smaller groups but were treated as unstable. Jointly estimated calibration slopes are retained in the electronic outputs as a separate diagnostic.

Reproducibility

The analysis used Python (version 3.12; Python Software Foundation), *NumPy* 2.5.3, *pandas* 3.0.6, *SciPy* 1.18.1, *scikit-learn* 1.9.1, and *statsmodels* 0.15.0. Multimedia Appendix 2 contains the executable pipeline, pinned dependencies, concise execution instructions, table-export scripts, and aggregate results. We checked sample counts, ordinal decoding, joins, Cliff δ, HC3 estimates, end point windows, and out-of-fold predictions through separate reconstructions. The "Methods" section describes the exploratory design and statistical assumptions; the study was not preregistered. Participant-level source data and intermediate predictions are not redistributed. Multimedia Appendix 3 maps the reporting items to their manuscript locations.

Results

Sample and Feature Availability

The feature sample contained 554 iOS and 151 Android annual records; Android represented 21.4% of 705 records. Its annual share ranged from 6.5% to 27.1% (Table 1). There were 669 records with end-of-term labels and 661 eligible screening windows.

XSL•FO
RenderX

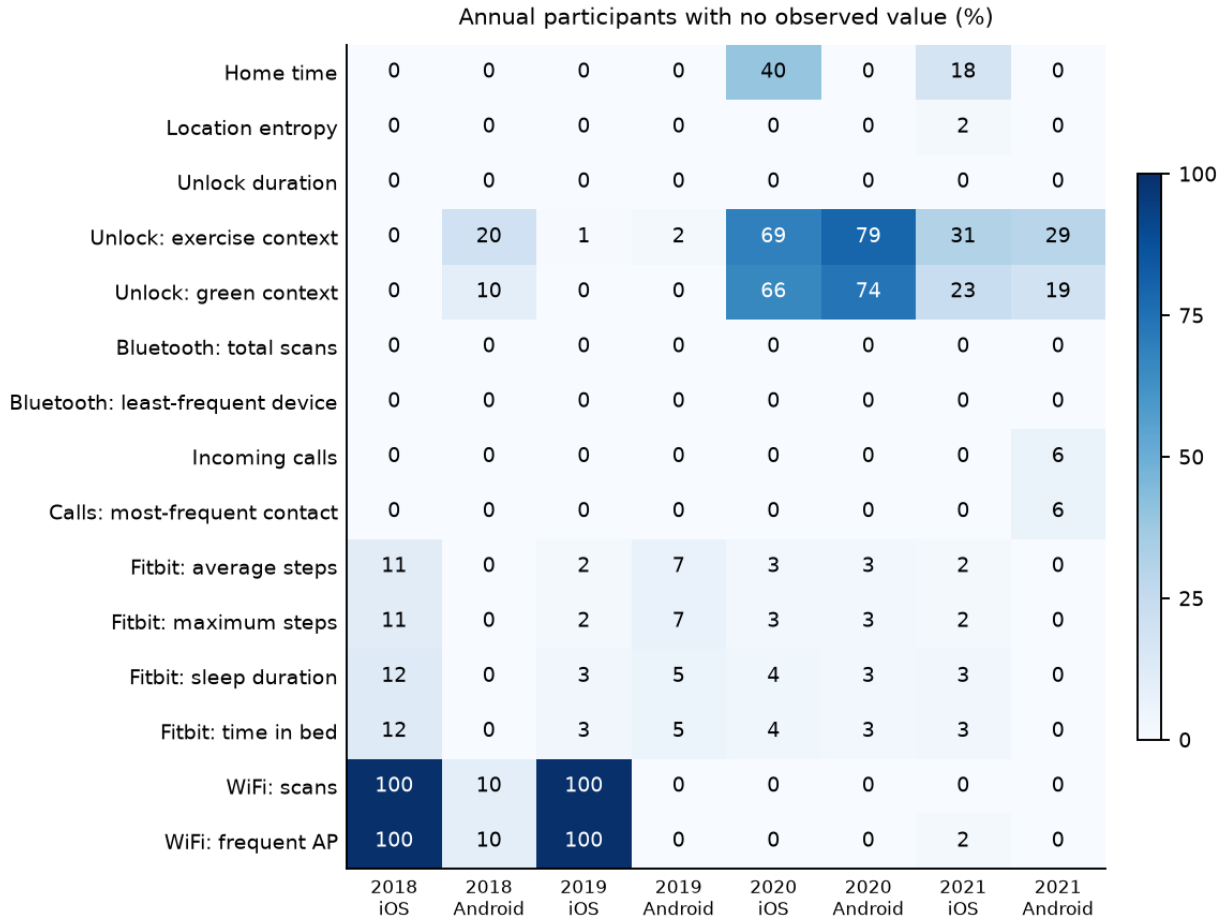
Table 1. Sample flow by platform in the secondary analysis of 4 Generalization of Longitudinal Behavior Modeling US college cohorts, 2018 to 2021[a].

Year	Platform	Metadata, n	Features, n	End-term label, n	Eligible screening, n
2018	iOS	160	145	132	131
2018	Android	41	10	10	10
2019	iOS	172	159	153	152
2019	Android	76	59	50	49
2020	iOS	114	103	102	98
2020	Android	44	34	33	33
2021	iOS	153	147	141	140
2021	Android	58	48	48	48

[a]Feature sample: 705 annual records, including 554 iOS and 151 Android (21.4%). The source reports 497 underlying people. Of 818 metadata records, 113 had no feature record; the reason is unknown. End-term labels were present for 669 feature records; 661 met the 14-day screening eligibility rule. Annual totals are not counts of distinct people across years. End-term label: Beck Depression Inventory-II score greater than 13.

Missingness differed by feature, platform, and cohort, rather than following a uniform phone-versus-wearable pattern (Figure 1; Multimedia Appendices 1 and 2). In Figure 1, absence means no value across all released daily rows, with denominators set by the annual iOS/Android samples of 145/10, 159/59, 103/34, and 147/48 for 2018 to 2021. Percentages are rounded to whole numbers, so 0% does not imply complete daily coverage. Fitbit features are wearable-derived; all others are phone-derived. This measure of missingness does not capture sensor uptime or verified enrollment coverage.

Figure 1. Participants with no observed daily value for each raw feature in a secondary analysis of Generalization of Longitudinal Behavior Modeling US college cohorts, 2018 to 2021. AP: access point.



Home time was entirely missing for 41 of 103 iOS records in 2020 and 26 of 147 in 2021, compared with 0 of 34 and 0 of 48 Android records, respectively. Both WiFi constructs were entirely absent from iOS observations in 2018 and 2019,

XSL•FO
RenderX

whereas at least one value was available for 9 of 10 and 59 of 59 Android records. Context-specific screen features also had limited coverage. Observed-day counts and within-person variance are reported separately from participant means; none of these summaries establishes equal measurement reliability.

Metadata-to-feature loss was greater for Android in every cohort. Missing feature records numbered 15/160 iOS versus 31/41 Android in 2018 (9.4% vs 75.6%; $P<.001$), 13/172 versus 17/76 in 2019 (7.6% vs 22.4%; $P=.002$), 11/114 versus 10/44 in 2020 (9.6% vs 22.7%; $P=.04$), and 6/153 versus 10/58 in 2021 (3.9% vs 17.2%; $P=.002$). These unadjusted Fisher tests describe availability in the supplied files; the cause of absence is unknown (Table S12 in Multimedia Appendix 1).

Total Bluetooth scans were observed for every annual participant on both platforms, with no zeros among observed daily values. Participant-level median imputation was therefore never triggered for this feature. Daily missingness was 31.0% versus 32.0% (iOS vs Android) in 2018, 22.1% versus 28.9% in 2019, 54.2% versus 31.5% in 2020, and 33.5% versus 28.1% in 2021 (Table S11 in Multimedia Appendix 1). Intermittent missingness could still affect participant means. Missingness-only classifier AUROCs ranged from 0.50 to 0.69. The 2020 observed-value home-time contrast uses 62 of 103 iOS records; smaller differences among retained observations do not resolve the missing-data problem.

Magnitude and Localization of Platform Differences

Total Bluetooth scan counts were lower on iOS in all 4 cohorts, with signed Cliff δ from −1.00 to −0.85. Home-time Cliff δ were also negative in every cohort, ranging from −0.32 to −0.21. The broader pattern contained smaller, variable, and sometimes reversed differences in other features. We did not combine these annual estimates into a pooled inferential effect.

The phone-only contrast comparing total Bluetooth and incoming-call counts with home time, entropy, and unlock duration was positive in each cohort. Differences in mean absolute Cliff δ ranged from 0.22 to 0.43, with all 4 participant-resampling intervals excluding zero (Table 2). This result does not depend on including Fitbit in the comparator or on the contact-identity artifact. It describes the selected feature panels, not all possible connectivity and behavioral constructs.

Table 2. Bluetooth counts and a phone-only feature contrast in 4 Generalization of Longitudinal Behavior Modeling US college cohorts, 2018 to 2021[a].

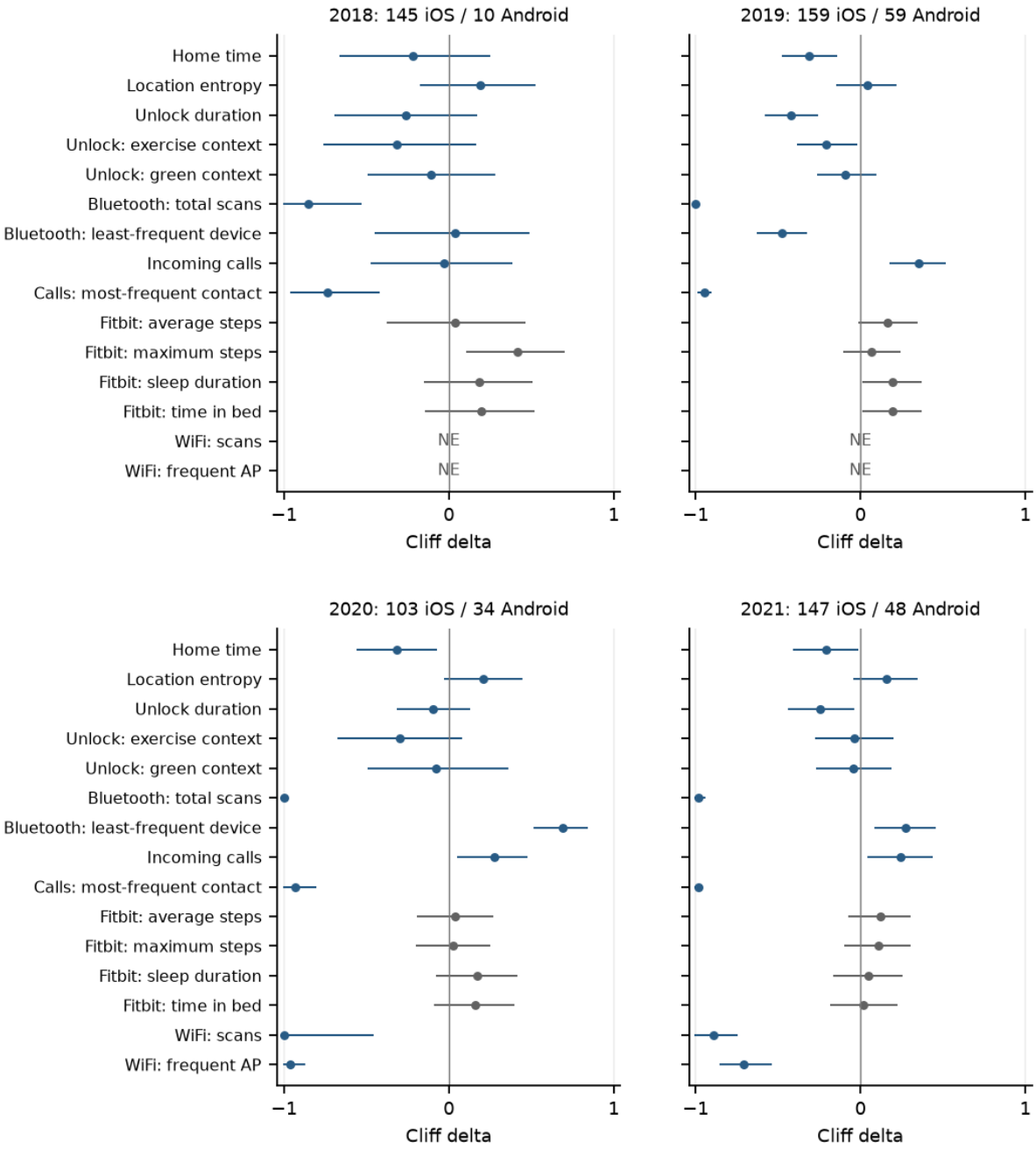
Year	iOS, n; median (IQR)	Android, n; median (IQR)	Bluetooth, Cliff δ (95% CI)	Phone-only contrast, mean absolute Cliff δ (95% CI)
2018	145; 39.3 (28.4 to 57.5)	10; 720.9 (380.8 to 1145.8)	−0.850 (−1.000 to −0.532)	0.216 (0.021 to 0.475)
2019	159; 126.8 (91.5 to 177.3)	59; 4413.3 (2735.6 to 5439.4)	−0.999 (−1.000 to −0.997)	0.416 (0.291 to 0.515)
2020	103; 5.2 (3.0 to 21.4)	34; 588.4 (445.7 to 1434.5)	−0.997 (−1.000 to −0.990)	0.427 (0.242 to 0.582)
2021	147; 12.5 (6.1 to 26.6)	48; 877.5 (324.6 to 1567.5)	−0.980 (−1.000 to −0.943)	0.409 (0.253 to 0.562)

[a]Medians and IQRs are across participant means of daily total Bluetooth scans. Cliff δ is Pr(iOS>Android) minus Pr(iOS<Android); negative values indicate lower iOS values. The phone-only contrast is mean absolute Cliff δ for total Bluetooth scans and incoming-call count minus that for home time, location entropy, and mean unlock duration. Fitbit and contact identity are excluded. Intervals use 2000 within-cohort participant resamples; overlapping people preclude independent cross-year pooling.

Figure 2 summarizes observed-value contrasts across the selected phone (blue) and Fitbit (gray) constructs. Each point is a Cliff δ, positive when iOS values are higher, and each bar is a within-cohort 95% participant-resampling interval, with a conservative bounded fallback where the bootstrap interval was degenerate. Annual platform counts appear above each panel; observed counts vary by feature and are supplied in Multimedia Appendix 2. WiFi contrasts are not estimable where iOS values are wholly missing. Because the annual cohorts share people, the panels are not independent replications.

XSL•FO
RenderX

Figure 2. Platform differences in observed participant means of raw sensing features in Generalization of Longitudinal Behavior Modeling US college cohorts, 2018 to 2021. AP: access point; NE: not estimable.



Platform Prediction and Transformations

The 13 raw constructs classified platform with out-of-fold AUROC of 0.94, 0.99, 0.997, and 0.98 across the 4 cohorts (Table 3). The 2020 estimate was below 1. Tables distinguish near-perfect rankings from exact observed rank separation. Removing all 4 connectivity constructs yielded 9-feature behavioral AUROCs of 0.78, 0.68, 0.41, and 0.48. These results do not support describing all behavioral-only models as chance classifiers or the retained features as platform-neutral.

XSL•FO
RenderX

Table 3. Out-of-fold platform area under the receiver operating characteristic curve and 95% intervals for matched transformations and behavioral features in 4 Generalization of Longitudinal Behavior Modeling US college cohorts, 2018 to 2021[a].

Year	Raw, 13	Normalized, 13	Discretized, 13	Behavioral, 9
2018	0.94 (0.78-1.00)	0.34 (0.08-0.76)	0.65 (0.24-0.84)	0.78 (0.36-0.90)
2019	0.99 (0.98-1.00)	0.91 (0.80-0.95)	0.72 (0.57-0.80)	0.68 (0.53-0.76)
2020	0.997 (0.907-1.000)	0.87 (0.70-0.94)	0.73 (0.62-0.85)	0.41 (0.26-0.64)
2021	0.98 (0.92-1.00)	0.91 (0.79-0.97)	0.80 (0.70-0.88)	0.48 (0.34-0.66)

[a]Cohort sizes are 155, 218, 137, and 195. Positive class: iOS. The 13-construct panels exclude WiFi; raw, normalized, and discretized columns represent the same constructs. The 9-feature panel contains 5 phone location/screen and 4 Fitbit features. All use fixed L2 logistic regression, 5 folds, and training-fold imputation/scaling. Intervals use 500 participant resamples with preprocessing and cross-validation refitted. They are approximate, conditional on observed platform counts. Estimates at or above 0.995 use 3 decimals; an asterisk marks an exact AUROC of 1 in the evaluated predictions. A boundary interval is conditional on the observed sample and does not certify future performance.

Matched normalized AUROCs were 0.34, 0.91, 0.87, and 0.91; discretized AUROCs were 0.65, 0.72, 0.73, and 0.80. Discretization yielded lower platform discrimination in every cohort. The first normalized estimate had a wide interval and only 10 Android records.

Missingness-only AUROC ranged from 0.50 to 0.69. Complete cases for the raw 13-construct panel numbered 135, 206, 23, and 115; the 2020 subset contained only 18 iOS and 5 Android records. Multimedia Appendix 1 reports complete-case models, the distinct removal steps, and sensitivity analyses using the 42-column panel. Multimedia Appendix 2 includes all conditional and refitted intervals and the questionnaire-bounded sensitivity results. The complete-case restriction markedly changed the available sample, especially in 2020.

Raw complete-case AUROCs were 0.89 (95% CI 0.70-1.00), 1.000* (0.994-1.000), 0.97 (0.82-1.00), and 0.99 (0.89-1.00), with refitted intervals. The asterisk marks an exact AUROC of 1 in the evaluated 2019 predictions. Requiring all 39 columns retained the same people, but used conditional intervals (Table S5 in Multimedia Appendix 1). In 2020, normalized AUROC was 0.48 among 23 complete cases versus 0.87 among all 137 records.

Restricting aggregation to dates between each participant's prestudy and poststudy questionnaires retained 143, 205, 131, and 189 annual records. Raw-panel AUROCs remained 0.93, 1.00, 0.99, and 0.98, with the 2019 unrounded estimate below 1. The corresponding 9-feature behavioral AUROCs were 0.74, 0.69, 0.42, and 0.53. Table S9 in Multimedia Appendix 1 supplies conditional intervals for this sensitivity analysis. Core-phone and Fitbit-only classifier results are reported separately in Table S4 in Multimedia Appendix 1, preserving the distinction between phone-derived behavior and wearable measurements.

Paired comparisons qualify the transformation results. In 2021, normalized minus raw AUROC was −0.066 (95% CI −0.191 to 0.014), so a reduction was not established in that cohort. Table S13 in Multimedia Appendix 1 reports every paired transformation contrast with its sign defined consistently. Below-chance estimates, including normalized AUROC 0.34 in 2018, are evaluated without reversing predictions or relabeling classes; small samples and unstable cross-validation can yield such values without establishing a reproducible inverse relationship.

Psychological Measures and Adjusted Feature Differences

None of the 7 psychological-score comparisons met the within-cohort 5% Benjamini-Hochberg threshold. The estimated differences and intervals remain relevant because the absence of a detected difference is not equivalence. Multimedia Appendix 1 reports distributions and effect sizes.

After adjustment for the 6 common psychological scores, the iOS coefficient for standardized log Bluetooth counts was −2.95, −2.12, −1.99, and −2.00 across the 4 cohorts, with all 4 intervals below zero (Table 4). The corresponding samples were 153, 216, 133, and 194. Incoming-call results differed from the contact-specific diagnostic and are reported separately. These estimates establish persistence under the specified adjustment, not that platform choice or sensing behavior is independent of personal characteristics.

XSL•FO
RenderX

Table 4. Cohort-specific platform coefficients before and after psychological adjustment in Generalization of Longitudinal Behavior Modeling US college students, 2018 to 2021[a].

Year	Outcome	n (iOS/Android)	Unadjusted β (95% CI)	Adjusted β (95% CI)	Adjusted *P*
2018	Bluetooth scans	153 (143/10)	−2.85 (−3.79 to −1.92)	−2.95 (−3.87 to −2.03)	<.001
2018	Incoming calls	153 (143/10)	−0.28 (−1.28 to 0.72)	−0.32 (−1.30 to 0.67)	.52
2019	Bluetooth scans	216 (159/57)	−2.12 (−2.24 to −2.01)	−2.12 (−2.24 to −2.01)	<.001
2019	Incoming calls	216 (159/57)	0.64 (0.38 to 0.89)	0.56 (0.30 to 0.82)	<.001
2020	Bluetooth scans	133 (99/34)	−2.00 (−2.17 to −1.83)	−1.99 (−2.15 to −1.83)	<.001
2020	Incoming calls	133 (99/34)	0.47 (0.13 to 0.81)	0.38 (0.05 to 0.70)	.02
2021	Bluetooth scans	194 (146/48)	−2.01 (−2.20 to −1.82)	−2.00 (−2.19 to −1.81)	<.001
2021	Incoming calls	191 (146/45)	0.38 (0.06 to 0.71)	0.34 (0.00 to 0.67)	.048

[a]Outcome units are SDs of log (1 + participant-mean count), standardized in the analysis sample. Android is the reference category. Models include an intercept; adjusted models add standardized prestudy 10-item University of California, Los Angeles Loneliness Scale (UCLA-10), 10-item Perceived Stress Scale (PSS-10), Brief Resilience Scale, Emotion Regulation Questionnaire reappraisal, Emotion Regulation Questionnaire suppression, and 9-item Center for Epidemiologic Studies Depression (CESD-9). Unadjusted and adjusted models use identical complete cases. Intervals and *P* values use HC3 covariance and t inference within cohort. No pooled model or cross-year person clustering is claimed. Anxiety sensitivity and contact-identity diagnostic models are in Multimedia Appendix 2.

The incoming-call coefficient was −0.32 (95% CI −1.30 to 0.67) in 2018 and positive in the 3 later cohorts after adjustment (Table 4). The 2021 incoming-call model included 191 records because 3 Android records lacked the outcome feature; its Bluetooth counterpart included 194. All primary adjusted design matrices had full rank, with 8 parameters, and residual degrees of freedom ranged from 125 to 208. Electronic outputs report anxiety-adjusted models and the most-frequent-contact diagnostic separately.

Point-null robustness values for the adjusted Bluetooth coefficient were 0.636, 0.892, 0.809, and 0.793, compared with 0.075, 0.234, 0.164, and 0.135 for incoming calls (Table S14 in Multimedia Appendix 1). These values concern equal-strength omitted associations under linear regression, not the probability of an unmeasured explanation.

Across the primary adjusted Bluetooth and incoming-call models, design-matrix condition numbers ranged from 4.57 to 9.56 and maximum observation leverage from 0.129 to 0.181. The complete diagnostic records accompany the coefficient estimates in Multimedia Appendix 2.

Exploratory Screening Performance

The 661 eligible annual windows comprised 141, 201, 131, and 188 records. Within these samples, the platform-versus-symptom-label Fisher tests had *P* values of .33, .87, .42, and .31. The first cohort's Android subgroup contained only 2 positive labels among 10 records. Subgroup estimates were therefore particularly uncertain, and its calibration-in-the-large was not estimated.

Positive symptom screens numbered 51 of 141 (36.2%), 86 of 201 (42.8%), 52 of 131 (39.7%), and 75 of 188 (39.9%). For the model including connectivity, the corresponding mean held-out predicted probabilities were 36.8%, 43.4%, 40.6%, and 40.0%. These summaries describe annual screening records; they do not count distinct diagnosed people across cohorts.

The 7-feature behavioral screening models had overall AUROCs of approximately 0.41 to 0.56. Adding total Bluetooth and incoming-call counts changed AUROC by 0.029, 0.023, 0.004, and 0.034; all paired 95% intervals included zero (Table 5). Adding platform to those models also produced intervals spanning zero. Supplementary results report Brier scores, log loss, average precision, platform-specific performance, and calibration. The models did not show consistently useful discrimination or establish a detrimental effect of connectivity features on screening.

XSL•FO
RenderX

Table 5. Exploratory end-of-term depressive-symptom screening in eligible Generalization of Longitudinal Behavior Modeling US college students, 2018 to 2021[a].

Year	n	Behavioral model AUROC[b]	Behavioral + connectivity model AUROC	Behavioral + connectivity + platform model AUROC	Connectivity AUROC change (95% CI)
2018	141	0.56	0.59	0.58	0.029 (−0.045 to 0.088)
2019	201	0.41	0.43	0.41	0.023 (−0.059 to 0.087)
2020	131	0.54	0.54	0.54	0.004 (−0.077 to 0.072)
2021	188	0.52	0.55	0.54	0.034 (−0.036 to 0.101)

[a]Outcome: BDI-II >13. Raw feature means cover the preceding 14 days, excluding the survey date, with at least 7 observations per feature. The behavioral model uses 3 core phone and 4 Fitbit features; connectivity adds total Bluetooth scans and incoming-call count; the final model also adds platform. Cohort models use the same eligible records and folds. Change is connectivity model minus behavioral model. Intervals refit preprocessing and models in 500 paired participant resamples. These are retrospective screening comparisons, not diagnoses, equivalence tests, or causal platform effects. Subgroup and calibration results are supplied in Multimedia Appendices 1 and 2.

[b]AUROC: area under the receiver operating characteristic curve.

All 3 substantive screening models had worse Brier scores than the intercept-only benchmark in each cohort. For the model including connectivity, scores were 0.233, 0.257, 0.260, and 0.253, compared with 0.231, 0.245, 0.240, and 0.240 for the benchmark. These descriptive comparisons do not establish population-level inferiority. Paired Brier-score intervals for connectivity and platform additions included zero throughout (Table S8 in Multimedia Appendix 1). Joint calibration slopes for the connectivity model were 0.39, −1.03, −0.07, and 0.04, indicating unstable or weak probability ordering.

Adding connectivity shifted Android mean predicted risk without an established AUROC gain. Android observed/expected ratios moved from 0.504 to 0.953 in 2018 (10 records), from 0.893 to 0.965 in 2019, and from 0.779 to 0.996 in 2021, but from 0.875 to 0.800 in 2020. The direction was inconsistent across cohorts, and these fixed-prediction estimates have no uncertainty intervals. They do not establish better subgroup calibration in a new sample. Table S7 in Multimedia Appendix 1 reports calibration-in-the-large and observed/expected ratios for both models and all platform groups.

Discussion

Principal Findings

Within GLOBEM, selected connectivity features carried more platform-associated variation than core phone behavioral measures, including after Fitbit features were considered separately. For mHealth apps that interpret communication or proximity as indicators of social engagement, recorded values must therefore be considered alongside their collection methods. A low Bluetooth count cannot, by itself, distinguish limited nearby-device exposure from limited observation. This finding supports feature-specific assessment before clinical interpretation, rather than treating all passive sensing as unreliable.

Smaller location and screen contrasts did not establish comparable monitoring across platforms. Home-time differences were directionally consistent, behavioral features retained platform information in some cohorts, and location-feature coverage differed between iOS and Android. In mobile monitoring between clinical assessments, an unavailable home-time feature must be distinguished from an observed low value. Similar averages among available records do not establish that an app observes behavior equally consistently on both systems.

The depressive-symptom screening experiment did not demonstrate deterioration after adding connectivity features, but weak baseline discrimination and wide paired intervals prevented conclusions about clinical usefulness or absence of harm. For mobile and ubiquitous mental health monitoring, these results support evaluating platform-specific coverage alongside performance for the intended clinical outcome. Before sensing-derived risk estimates inform follow-up or referral, their calibration and usefulness require assessment in the intended population and use case [14,15].

What the Platform Contrast Identifies

Bluetooth counts combine nearby-device exposure, scan opportunity, and feature construction. Ownership-related behavior could affect exposure; hardware and software could affect observation. These components were not observed separately. Near separation of released values therefore does not identify how much of the contrast belongs to operating-system policy, the app, the provider, or behavior.

Contact identity presents a documented construct mismatch. Without stable identifiers, repeated contacts cannot support the same social-network statistic across platforms. Exclusion from shared substantive models follows from the feature definition, not its coefficient size. Total incoming events do not require persistent contact identity. Their direction also differed: iOS values were higher in 2019 to 2021, inconsistent with a blanket claim that iOS call observations were absent. Vendor and AWARE-iOS documentation distinguish restrictions on scanning or identity from the ability to observe call events (Multimedia Appendix 1). The exact historical implementation remains unverified; third-party calling services cannot be assigned responsibility from these exports.

Implications for mHealth Measurement

Individualization can remove a stable additive offset through baseline subtraction, or a stable multiplicative difference through rescaling under suitable assumptions. It cannot restore

XSL•FO
RenderX

unrecorded events or lost contact identity. If observation changes with charging, phone state, or behavior, a fixed baseline may remain inadequate. The daily exports do not distinguish these mechanisms. With far fewer recorded Bluetooth events on iOS, centering also leaves different observation counts. It cannot be assumed to equalize measurement noise or preserve equally precise deviations from each person's baseline.

Discretization can reduce scale differences while discarding distances relevant to symptoms. Lower platform AUROC is therefore insufficient grounds for choosing a mental health representation. Validation should compare retained outcome information and measurement stability using a baseline available before prediction. Our retrospective transformation comparison leaves that prospective question open, especially after a device replacement or software update.

Smaller behavioral contrasts can prioritize features for validation without establishing interchangeability. Paired-device or reference observations would separate agreement from merely similar averages, consistent with measurement-comparison methodology [41]. Observed daily variance mixes behavioral fluctuation with measurement variation. It cannot substitute for repeated reference measurements of the same underlying behavior.

Textbox 1 summarizes practical checks that follow from these findings. Bluetooth scan counts should be excluded from a shared clinical interpretation when comparable observation cannot be justified, or retained as an explicitly evaluated sensitivity input pending measurement validation. This conditional recommendation differs from automatic removal of every connectivity feature. Software and collection choices restrict which scientific questions released traces can answer [42].

These are measurement and reporting checks, not evidence that any resulting screening model is clinically validated.

Textbox 1. Platform audit steps for passive-sensing mobile health studies.

Audit step
• Report platform counts at enrollment, feature availability, and outcome-model eligibility; account for losses between these samples.
• Separate observed zeros, intermittently missing days, entirely absent participant features, and structural noncollection. Report their denominators by platform.
• Verify event definitions and collection capabilities. Exclude or redefine constructs that cannot represent the same quantity on both platforms.
• Use platform classification as a measurement diagnostic, with matched constructs, participant-disjoint validation, and explicit missing-data checks.
• Evaluate the intended health outcome against an intercept benchmark. Compare calibration and probability error by platform alongside discrimination and uncertainty.
• Record operating system, app, hardware, and feature-provider versions, and repeat validation after consequential changes.

Relation to Prior Evidence

Platform variation is a candidate explanation for generalization difficulties, not a demonstrated cause of them [12,13]. Platform-specific modeling also has precedent: Lu and colleagues [43] fitted separate iOS and Android depression models to handle device-platform differences. Earlier connectivity-based symptom studies establish relevance, rather than validating every implementation [2,3]. Location-feature associations have also varied with weekday versus weekend aggregation [44]. Replication needs comparable sources, operational definitions, and observation windows, not just matching column labels.

Currey and Torous [45] found weak passive-only prediction and variation with analytical subset and coverage restriction. In an mHealth study involving depression and diabetes, complete-case deletion retained 950 of 2694 participant-days [46]. Neither study tests GLOBEM's mechanism. They illustrate why apparent feature quality, evaluated sample, and prediction performance must be considered together.

Cross-study coverage findings also caution against ranking platforms universally. Kiang and colleagues [47] found lower GPS noncollection on iOS across 6 digital-phenotyping studies, unlike the iOS home-time absence observed here. Raw GPS noncollection and absence of a derived home-time feature are different endpoints, collected with different systems. This contrast supports conditional interpretation, not a contradiction that either dataset alone can resolve. Shared Fitbit hardware removes one phone-sensor difference by design but does not validate every wearable output in GLOBEM.

Interpreting the Downstream Experiment

The fixed screening model made feature additions comparable but left nonlinear and personalized models untested. Weak baseline discrimination means the behavioral panel did not clearly recover the target symptom signal. Small incremental changes could reflect weak relevance, redundancy, sampling variation, or model misspecification. The null cannot distinguish these explanations. The GLOBEM benchmark also reported weak generalization, but its 2-institution datasets, feature sets, and validation tasks differ from this end-of-term experiment [13]. It cannot establish that weak screening performance here is an inevitable property of these cohorts.

A ranking improvement can coexist with poor calibration or higher probability error [40]. Clinical usefulness also depends on the consequences of decisions at specified thresholds [48]. Our models justify neither a referral cutoff nor a claim of improved patient care. Such evaluation requires a defined clinical action and an appropriate reference outcome.

XSL•FO
RenderX

Small platform groups and missing demographics constrain subgroup interpretation. Similar point estimates would not establish equal errors or benefit without adequate evaluation of specified clinical decisions.

Limitations

The platform was not randomized, and income, age, device price, and related demographic attributes were unavailable. For example, if a characteristic associated with iOS ownership also increased time away from home, it could contribute to the negative home-time contrast; the opposite behavioral association could attenuate that contrast. Direction cannot be assigned without those measurements. The Bluetooth robustness values quantify the strength of omitted associations required under the fitted linear specification, but cannot label socioeconomic explanations implausible or identify a device mechanism. They neither cover arbitrary nonlinear misspecification nor remedy selection into the available-feature sample. Demographics are available from the custodians by special request; none was supplied.

Verified cross-year person links were unavailable. Conducting analyses separately by cohort avoids fitting models that assume repeated annual records are independent, but prevents estimation of person-level random intercepts across years and within-person platform changes. The same individuals may contribute to several annual findings, so consistency across years is not independent replication. We cannot infer that there were no platform switchers. A custodian-provided cross-walk is needed for those analyses.

This is a selected college population from one US institution using one collection and processing system during 2018 to 2021. Platform ratios, hardware, operating system versions, app configurations, and user behavior may differ elsewhere. Android metadata records were more often absent from the feature files, particularly in 2018. If this reflects collection failure, the retained Android sample would selectively represent devices with recorded data. The cause is unknown, so contrasts characterize retained records rather than all enrolled phone owners. The first cohort included only 10 Android records. Complete-case selection was especially severe in 2020 and could change both effect magnitude and classifier performance. These factors limit precision and transportability, including to Android-majority settings.

The release provides derived features rather than the raw observation process. Observed days and within-person variance are coverage diagnostics, not estimates of sensor accuracy or test-retest reliability. The released date range does not identify exact enrollment, and upstream feature construction can depend on participant-level context beyond a prediction window. The historical Bluetooth implementation and context-specific screen construction could not be fully verified. Our results apply to these exports and cannot attribute every difference to the operating system alone.

The alphabetical panel rule did not prioritize commonly modeled unlock counts, unique Bluetooth devices, or outgoing calls. Holding constructs fixed makes transformation comparisons interpretable, but cannot establish that these constructs represent their domains. The screening experiment used one fixed model family, a limited feature panel, one thresholded symptom measure, and one retrospectively defined window. It did not assess diagnosis, onset forecasting, intervention effects, or deployment under a new platform distribution. Small outcome-by-platform groups constrained calibration and subgroup comparisons. Bootstrap intervals are approximate and conditional on observed stratum counts; they do not capture all uncertainty in prevalence, model specification, or a new setting. Nonsignificant paired differences therefore cannot establish clinical equivalence or fairness.

Conclusions

Within GLOBEM's 4 annual cohorts, selected connectivity features showed larger platform-associated differences than core phone behavioral features, while coverage differences also affected location. Matched transformations changed platform predictability. Because the screening models lacked demonstrated predictive usefulness, their incremental comparisons cannot establish or exclude downstream harm from platform-associated measurement differences. A shared feature name is insufficient evidence of comparable measurement. The unresolved question is which differences alter clinically useful monitoring when observation quality, owner characteristics, and device changes can be distinguished.

Acknowledgments

The authors used the Generalization of Longitudinal Behavior Modeling (GLOBEM) datasets, made available by the University of Washington and collaborating institutions through PhysioNet, and thank the original investigators for releasing this resource. The authors thank colleagues at the American University of the Middle East and the University of Wolverhampton for helpful discussions.

Data Availability

The datasets analyzed during this study are available in the PhysioNet repository (GLOBEM Dataset, version 1.1) [20]. They are not publicly available because passive-sensing data carry a risk of reidentification; access requires PhysioNet credentialing, completion of the required training, and a signed data-use agreement. Demographic variables are available only from the data custodians on special request. The analysis code and aggregate results generated during this study are included in Multimedia Appendix 2; participant-level records and predictions are not redistributed.

XSL•FO
RenderX

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Generative AI Use

ChatGPT 5.5 assisted with editing, proofreading, and citation check. The authors are responsible for the analysis and final text and for the accuracy and integrity of the work.

Authors' Contributions

Conceptualization: AN
Data curation: AN
Formal analysis: AN
Methodology: AN, FN
Software: AN
Validation: AN
Visualization: AN
Writing – original draft: AN
Writing – review & editing: AN, FN

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary methods, feature provenance, psychological comparisons, and sensitivity analyses for platform-associated sensing differences in 4 US college cohorts, 2018 to 2021.
[DOCX File , 87 KB - mhealth_v14i1e103711_app1.docx]

Multimedia Appendix 2
Executable analysis code, aggregate result tables, pinned software dependencies, and instructions for reproducing the analyses, tables, and figures.
[ZIP File (Zip Archive), 132 KB - mhealth_v14i1e103711_app2.zip]

Multimedia Appendix 3
Reporting checklist mapping for observational analyses and exploratory prediction models.
[DOCX File , 24 KB - mhealth_v14i1e103711_app3.docx]

References

1. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
2. Place S, Blanch-Hartigan D, Rubin C, Gorrostieta C, Mead C, Kane J, et al. Behavioral indicators on a mobile sensing platform predict clinically validated psychiatric symptoms of mood and anxiety disorders. J Med Internet Res 2017;19(3):e75 [FREE Full text] [doi: 10.2196/jmir.6678] [Medline: 28302595]
3. Opoku Asare K, Terhorst Y, Vega J, Peltonen E, Lagerspetz E, Ferreira D. Predicting depression from smartphone behavioral markers using machine learning methods, hyperparameter optimization, and feature importance analysis: exploratory study. JMIR Mhealth Uhealth 2021;9(7):e26540 [FREE Full text] [doi: 10.2196/26540] [Medline: 34255713]
4. Doryab A, Villalba DK, Chikersal P, Dutcher JM, Tumminia M, Liu X, et al. Identifying behavioral phenotypes of loneliness and social isolation with passive sensing: statistical analysis, data mining and machine learning of smartphone and fitbit data. JMIR Mhealth Uhealth 2019;7(7):e13209 [FREE Full text] [doi: 10.2196/13209] [Medline: 31342903]
5. Wang Y, Ren X, Liu X, Zhu T. Examining the correlation between depression and social behavior on smartphones through usage metadata: empirical study. JMIR Mhealth Uhealth 2021;9(1):e19046 [FREE Full text] [doi: 10.2196/19046] [Medline: 33404512]
6. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017;13:23-47 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
7. Wang R, Wang W, Dasilva A, Huckins JF, Kelley WM, Heatherton TF, et al. Tracking depression dynamics in college students using mobile phone and wearable sensing. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018;2(1):43. [doi: 10.1145/3191775] [Medline: 39449996]

XSL•FO
RenderX

8. Chikersal P, Doryab A, Tumminia M, Villalba DK, Dutcher JM, Liu X, et al. Detecting depression and predicting its onset using longitudinal symptoms captured by passive sensing: a machine learning approach with robust feature selection. ACM Trans Comput-Hum Interact 2021;28(1):1-41. [doi: 10.1145/3422821]
9. Halabi R, Selvarajan R, Lin Z, Herd C, Li X, Kabrit J, et al. Comparative assessment of multimodal sensor data quality collected using android and iOS smartphones in real-world settings. Sensors (Basel) 2024;24(19):6246 [FREE Full text] [doi: 10.3390/s24196246] [Medline: 39409286]
10. Vega J, Li M, Aguillera K, Goel N, Joshi E, Khandekar K, et al. Reproducible analysis pipeline for data streams: open-source software to process data collected with mobile devices. Front Digit Health 2021;3:769823 [FREE Full text] [doi: 10.3389/fdgth.2021.769823] [Medline: 34870271]
11. Trifan A, Oliveira M, Oliveira JL. Passive sensing of health outcomes through smartphones: systematic review of current solutions and possible limitations. JMIR Mhealth Uhealth 2019;7(8):e12649 [FREE Full text] [doi: 10.2196/12649] [Medline: 31444874]
12. Müller SR, Chen XL, Peters H, Chaintreau A, Matz SC. Depression predictions from GPS-based mobility do not generalize well to large demographically heterogeneous samples. Sci Rep 2021;11(1):14007 [FREE Full text] [doi: 10.1038/s41598-021-93087-x] [Medline: 34234186]
13. Xu X, Liu X, Zhang H, Wang W, Nepal S, Sefidgar Y, et al. GLOBEM: cross-dataset generalization of longitudinal human behavior modeling. Proc ACM Interact Mob Wearable Ubiquitous Technol 2022;6(4). [doi: 10.1145/3569485]
14. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. NPJ Digit Med 2019;2:88 [FREE Full text] [doi: 10.1038/s41746-019-0166-1] [Medline: 31508498]
15. Goldsack JC, Coravos A, Bakker JP, Bent B, Dowling AV, Fitzer-Attas C, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). NPJ Digit Med 2020;3:55. [doi: 10.1038/s41746-020-0260-4] [Medline: 32337371]
16. Currey D, Torous J. Increasing the value of digital phenotyping through reducing missingness: a retrospective review and analysis of prior studies. BMJ Ment Health 2023;26(1):e300718 [FREE Full text] [doi: 10.1136/bmjment-2023-300718] [Medline: 37197799]
17. Torous J, Staples P, Barnett I, Sandoval LR, Keshavan M, Onnela JP. Characterizing the clinical relevance of digital phenotyping data quality with applications to a cohort with schizophrenia. NPJ Digit Med 2018;1:15 [FREE Full text] [doi: 10.1038/s41746-018-0022-8] [Medline: 31304300]
18. Luijken K, Groenwold RHH, Van Calster B, Steyerberg EW, van Smeden M. Impact of predictor measurement heterogeneity across settings on the performance of prediction models: a measurement error perspective. Stat Med 2019;38(18):3444-3459 [FREE Full text] [doi: 10.1002/sim.8183] [Medline: 31148207]
19. Yang X, Knights J, Bangieva V, Kambhampati V. Association between the severity of depressive symptoms and human-smartphone interactions: longitudinal study. JMIR Form Res 2023;7:e42935 [FREE Full text] [doi: 10.2196/42935] [Medline: 36811951]
20. Xu X, Zhang H, Sefidgar Y, Ren Y, Liu X, Seo W, et al. GLOBEM dataset: multi-year datasets for longitudinal human behavior modeling generalization (version 1.1). PhysioNet. 2023. URL: https://physionet.org/content/globem/1.1/
21. Xu X, Zhang H, Sefidgar Y, Ren Y, Liu X, Seo W, et al. GLOBEM dataset: multi-year datasets for longitudinal human behavior modeling generalization. 2022 Presented at: Advances in Neural Information Processing Systems 35 (NeurIPS 2022), Datasets and Benchmarks Track; November 28 to December 9, 2022; New Orleans, LA p. 24655-24692 URL: https://proceedings.neurips.cc/paper_files/paper/2022/hash/9c7e8a0821dfcb58a9a83cbd37cc8131-Abstract-Datasets_and_Benchmarks.html [doi: 10.52202/068431-1790]
22. Pollard T, Moody BE, Lehman LH, Gow BJ, Fernandes C, Xie C, et al. PhysioNet as a global platform for biomedical research. Nat Health 2026;1(8):792-795. [doi: 10.1038/s44360-026-00096-z] [Medline: 42841059]
23. UW-EXP. GLOBEM. GitHub. URL: https://github.com/UW-EXP/GLOBEM/tree/4f140fc5290dc97204298cca28b956165aa0a29f [accessed 2026-09-22]
24. Definitions for purposes of this policy. Electronic Code of Federal Regulations (eCFR). URL: https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-A/part-46/subpart-A/section-46.102 [accessed 2026-10-25]
25. Phone calls. RAPIDS. URL: https://www.rapids.science/1.6/features/phone-calls/ [accessed 2026-09-22]
26. Phone screen. RAPIDS. URL: https://www.rapids.science/1.6/features/phone-screen/ [accessed 2026-09-22]
27. Phone bluetooth. RAPIDS. URL: https://www.rapids.science/1.6/features/phone-bluetooth/ [accessed 2026-09-22]
28. Phone WiFi connected. RAPIDS. URL: https://www.rapids.science/1.6/features/phone-wifi-connected/ [accessed 2026-09-22]
29. Phone locations. RAPIDS. URL: https://www.rapids.science/1.6/features/phone-locations/ [accessed 2026-09-22]
30. Cliff N. Dominance statistics: ordinal analyses to answer ordinal questions. Psychol Bull 1993;114(3):494-509. [doi: 10.1037/0033-2909.114.3.494]
31. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Methodol 1995;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
32. Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ. LIBLINEAR: a library for large linear classification. J Mach Learn Res 2008;9:1871-1874 [FREE Full text]
33. Rubin DB. Inference and missing data. Biometrika 1976;63(3):581-592. [doi: 10.2307/2335739]

XSL•FO
RenderX

34. Saeb S, Lonini L, Jayaraman A, Mohr DC, Kording KP. The need to approximate the use-case in clinical machine learning. Gigascience 2017;6(5):1-9 [FREE Full text] [doi: 10.1093/gigascience/gix019] [Medline: 28327985]
35. Bartlett JW, Harel O, Carpenter JR. Asymptotically unbiased estimation of exposure odds ratios in complete records logistic regression. Am J Epidemiol 2015;182(8):730-736 [FREE Full text] [doi: 10.1093/aje/kwv114] [Medline: 26429998]
36. Bengio Y, Grandvalet Y. No unbiased estimator of the variance of K-fold cross-validation. J Mach Learn Res 2004;5:1089-1105 [FREE Full text]
37. MacKinnon JG, White H. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. J Econom 1985;29(3):305-325. [doi: 10.1016/0304-4076(85)90158-7]
38. Cinelli C, Hazlett C. Making sense of sensitivity: extending omitted variable bias. J R Stat Soc Series B Stat Methodol 2020;82(1):39-67. [doi: 10.1111/rssb.12348]
39. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 2015;10(3):e0118432 [FREE Full text] [doi: 10.1371/journal.pone.0118432] [Medline: 25738806]
40. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. BMC Med 2019;17(1):230 [FREE Full text] [doi: 10.1186/s12916-019-1466-7] [Medline: 31842878]
41. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;327(8476):307-310. [doi: 10.1016/S0140-6736(86)90837-8]
42. Onnela JP. Opportunities and challenges in the collection and analysis of digital phenotyping data. Neuropsychopharmacology 2021;46(1):45-54 [FREE Full text] [doi: 10.1038/s41386-020-0771-3] [Medline: 32679583]
43. Lu J, Shang C, Yue C, Morillo R, Ware S, Kamath J, et al. Joint modeling of heterogeneous sensing data for depression assessment via multi-task learning. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018;2(1):1-21. [doi: 10.1145/3191753]
44. Saeb S, Lattie EG, Schueller SM, Kording KP, Mohr DC. The relationship between mobile phone location sensor data and depressive symptom severity. PeerJ 2016;4:e2537 [FREE Full text] [doi: 10.7717/peerj.2537] [Medline: 28344895]
45. Currey D, Torous J. Digital phenotyping correlations in larger mental health samples: analysis and replication. BJPsych Open 2022;8(4):e106 [FREE Full text] [doi: 10.1192/bjo.2022.507] [Medline: 35657687]
46. Sarda A, Munuswamy S, Sarda S, Subramanian V. Using passive smartphone sensing for improved risk stratification of patients with depression and diabetes: cross-sectional observational study. JMIR Mhealth Uhealth 2019;7(1):e11041 [FREE Full text] [doi: 10.2196/11041] [Medline: 30694197]
47. Kiang MV, Chen JT, Krieger N, Buckee CO, Alexander MJ, Baker JT, et al. Sociodemographic characteristics of missing data in digital phenotyping. Sci Rep 2021;11(1):15408 [FREE Full text] [doi: 10.1038/s41598-021-94516-7] [Medline: 34326370]
48. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26(6):565-574 [FREE Full text] [doi: 10.1177/0272989X06295361] [Medline: 17099194]

Abbreviations

AUROC: area under the receiver operating characteristic curve
BDI-II: Beck Depression Inventory-II
BRS: Brief Resilience Scale
CESD-9: 9-item Center for Epidemiologic Studies Depression score
ERQ: Emotion Regulation Questionnaire
GLOBEM: Generalization of Longitudinal Behavior Modeling
mHealth: mobile health
PSS-10: 10-item Perceived Stress Scale
RAPIDS: Reproducible Analysis Pipeline for Data Streams
STAI: State-Trait Anxiety Inventory
STAI-S: State-Trait Anxiety Inventory–State
UCLA-10: 10-item UCLA Loneliness Scale

XSL•FO
RenderX

Edited by A Stone; submitted 05.Jun.2026; peer-reviewed by A Tai, D Hidalgo-Mazzei; comments to author 11.Sep.2026; revised version received 25.Oct.2026; accepted 26.Oct.2026; published 16.Nov.2026.

Please cite as:
Naqvi A, Naqvi F
Platform Differences in Smartphone Sensing Features for Mental Health Across Four Annual College Cohorts: Secondary Analysis
JMIR Mhealth Uhealth 2026;14:e103711
URL: https://mhealth.jmir.org/2026/1/e103711
doi:10.2196/103711
PMID:

©Ariz Naqvi, Farah Naqvi. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 16.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

New Insights and Effectiveness of Digital Interventions for Health Anxiety: Systematic Review and Meta-Analysis

Jing Xu[1*], PhD; Jifeng Shen[1*], PhD; Guanghui Huang[1,2*], PhD; Jue Wang[3,4*], PhD

[1]Faculty of Humanities and Arts, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macau, Macau, China

[2]Zhuhai M.U.S.T. Science and Technology Research Institute, Zhuhai, Guangdong, China

[3]Faculty of Chinese Medicine, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macau, Macau, China

[4]State Key Laboratory of Mechanism and Quality of Chinese Medicine, Macau, Macau, China

[*]all authors contributed equally

Corresponding Author:
Jue Wang, PhD
Faculty of Chinese Medicine, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macau, Macau, China

Abstract

Background: Health anxiety (HA), classified in the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5*) under illness anxiety disorder (IAD) or somatic symptom disorder, is an increasingly prevalent condition associated with substantial psychological and health care burdens. Digital mental health interventions (DMHIs) offer scalable and accessible approaches to care, with digital behavioral interventions (DBIs), particularly internet-delivered cognitive and behavioral therapies, among the most frequently studied. However, their effectiveness in reducing HA has not been comprehensively synthesized.

Objective: This study aimed to systematically evaluate the effectiveness of DBIs for reducing HA in adults and to examine their effects on related psychological outcomes.

Methods: This systematic review and meta-analysis followed the PRISMA 2020 statement and Cochrane guidance. PubMed, Embase, Web of Science, ScienceDirect, PsycINFO, and the Cochrane Library were searched from database inception to November 1, 2025. Eligible studies included randomized controlled trials (RCTs), quasi-experimental studies, and cohort studies involving adults with HA. HA severity was the primary outcome. Random-effects meta-analyses were restricted to RCTs. Methodological quality was appraised using Joanna Briggs Institute tools, risk of bias in RCTs using Risk of Bias 2 (RoB 2), and certainty of evidence using Grading of Recommendations Assessment, Development, and Evaluation (GRADE).

Results: Sixteen studies involving 2581 participants were included, comprising 9 RCTs, 6 quasi-experimental studies, and 1 cohort study. Six RCTs involving 585 participants contributed to the primary meta-analysis. Digital interventions significantly reduced HA compared with control conditions (Hedges g=−0.77, 95% CI −1.25 to –0.30; P=.002), although between-study heterogeneity was substantial (I^2=86.7%). Leave-one-out analyses showed that this effect remained significant after exclusion of any individual trial. By contrast, pooled effects on illness-related attitudes, general anxiety, anxiety sensitivity, and depressive symptoms were not statistically significant and were characterized by substantial heterogeneity. Among the 9 RCTs, 4 were judged as having low overall risk of bias, 2 had some concerns, and 3 were at high risk. GRADE certainty was low for HA and very low for the secondary outcomes.

Conclusions: Digital interventions show potential for reducing core health-anxiety symptoms, but confidence in the magnitude and generalizability of the effect is limited by substantial heterogeneity, risk-of-bias concerns, and low certainty of evidence. Evidence for broader psychological outcomes remains inconclusive. Adequately powered, well-reported RCTs using rigorous outcome assessment, appropriate active comparators, and longer follow-up are needed to confirm effectiveness and clarify which patients benefit most and under what delivery conditions.

Trial Registration: PROSPERO CRD420251238746; https://www.crd.york.ac.uk/prospero/

(JMIR Mhealth Uhealth 2026;14:e91108) doi:10.2196/91108

KEYWORDS

health anxiety; anxiety disorder; hypochondria; digital intervention; digital behavioral intervention.

Introduction

Health anxiety (HA), classified in the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)* under illness anxiety disorder (IAD) or somatic symptom disorder (SSD), is a prevalent and distressing condition characterized by persistent preoccupation with having or acquiring a serious illness despite appropriate medical reassurance, a presentation

XSL•FO
RenderX

historically referred to as hypochondriasis [1]. Although the *DSM-5* classifies these presentations under somatic symptom and related disorders, HA is widely conceptualized as part of the anxiety spectrum. It shares the core features of anxiety disorders, including excessive worry, anticipatory anxiety, and avoidance behaviors [2]. In HA, these processes are directed toward illness-related threats and manifest as fear of serious disease, hypervigilance to bodily sensations, and repeated reassurance seeking. Consistent with this overlap, HA frequently co-occurs with generalized anxiety disorder, which has an estimated lifetime prevalence of 6.2%, and may represent an early or prodromal manifestation of anxiety pathology [3]. Epidemiological evidence indicates that HA affects approximately 3.4% of the general population [4], with early UK studies reporting rates of 4%-8% [5,6], and its prevalence may reach 20% among patients attending primary care and other medical settings [7]. In the wake of recent global infectious disease outbreaks, HA has become an increasingly prominent public mental health concern [8]. The pandemic and its aftermath have also reshaped the online dissemination of health information. Compared with the general population, individuals with elevated HA are more likely to respond to such information with heightened worry, fear, and negative expectations, thereby reinforcing and worsening their symptoms [9,10]. Persistent fear of serious illness and repeated monitoring of bodily symptoms can cause substantial psychological distress and functional impairment, reducing quality of life [11]. At the system level, excessive health care use, unnecessary diagnostic testing, and productivity losses place a considerable burden on health care systems [12]. Importantly, elevated levels of HA have also been observed among health care providers [13], with prevalence rates reported to be as high as 30% in medical professionals [14]. Consequently, increasing clinical and public health attention has been directed toward these populations, highlighting the urgent need for scalable and accessible interventions to mitigate the growing impact of disease-related anxiety on public health.

A critical characteristic that must be addressed in the prevention and intervention of HA is patients' anxiety sensitivity (AS) and the involuntary emergence of negative mental imagery when processing health-related information [15]. Many individuals, driven by fear of illness, frequently search online for information about diseases and available treatments; however, excessive reliance on internet-based self-diagnosis does not alleviate anxiety and may instead exacerbate it [16]. Continuous exposure to heterogeneous and often unreliable online health information environments can further contribute to psychological dysregulation. Currently, cognitive behavioral interventions have been widely used to address disease-related anxiety at the population level [17], partially reducing anxiety sensitivity and alleviating psychological distress by optimizing daily behaviors and redirecting attentional focus. In recent years, the integration of digital technologies into behavioral therapies has emerged as a promising approach to expanding the reach of mental health services. Digital mental health interventions (DMHIs) [18], including internet-based cognitive behavioral therapy (ICBT) [19], mobile health apps [20], and digitally delivered interactive exposure therapies [21], have demonstrated efficacy comparable to face-to-face interventions in treating anxiety disorders, while offering notable advantages such as cost-effectiveness [22], anonymity, and geographic flexibility [23]. Additionally, questionnaire-based evidence suggests that improving health literacy and promoting health-related behaviors play a critical role in emotional regulation among individuals with HA [24]. Within the domain of behavioral interventions, behavioral activation (BA) has been shown to effectively reduce anxiety symptoms by disrupting avoidance cycles [25] and enhancing engagement in meaningful activities, either actively or passively [26]. When implemented through digital platforms, these behavioral strategies exhibit substantial potential for HA management by enabling continuous monitoring, remote participation, and timely support during acute health-related distress [27].

Although substantial evidence supports the effectiveness of digital interventions for general anxiety and depressive disorders, a comprehensive evaluation of digital behavioral interventions (DBIs) specifically targeting HA remains lacking. Recent reviews have provided an extensive synthesis of intervention outcomes for HA among adolescents and their parents [28], and randomized controlled trials (RCTs) have examined a range of digital technological approaches. However, the evidence remains fragmented. No systematic review has yet synthesized the overall effectiveness of DBIs for reducing HA or examined whether their effects vary across populations and study designs.

This systematic review and meta-analysis addresses these gaps by evaluating the clinical effectiveness of DBIs for individuals with HA and their effects on related psychological outcomes, including depression and stress. It also examines the emerging role of DBIs as an adjunctive approach and synthesizes evidence to inform integrated health management strategies beyond conventional cognitive behavioral therapy (CBT).

Methods

Overview

This systematic review was designed in strict accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 statement [29] (Checklist 1) and the most recent guidance of the Cochrane Handbook for Systematic Reviews of Interventions [30]. To ensure transparency and reproducibility, the review protocol was prospectively registered in the PROSPERO (International Prospective Register of Systematic Reviews; registration number CRD420251238746). The research questions were formulated using the PICOS (Population, Intervention, Comparator, Outcomes, and Study design) framework [31], which provided a structured approach to clearly define the scope and focus of the review and to ensure comprehensive and systematic coverage of the topic.

Eligibility Criteria

Inclusion Criteria

Inclusion Criteria: (1) The study designs included RCTs, quasi-experimental studies (QESs), and cohort studies (CSs). Both controlled (ie, with a separate comparator group) and uncontrolled single-arm pre-post designs were considered

XSL•FO
RenderX

eligible for inclusion. All studies were required to provide original outcome data and complete procedural descriptions; (2) the participants meeting diagnostic criteria for HA, including DSM-IV hypochondriasis, *DSM-5* SSD with predominant HA features, or IAD, or participants with elevated HA confirmed using validated scales (eg, HAI ≥30, SHAI ≥18, WI-7 >14). Studies were eligible when HA was assessed as a primary target outcome; (3) digital health interventions targeting HA, defined as structured psychological interventions delivered wholly or mainly through digital technology. Eligible approaches included internet-based CBT (guided or unguided), internet-based ACT, cognitive bias modification, and other digital self-help programs, delivered via web-based platforms, mobile apps, e-books or bibliotherapy, video programs, or interactive formats such as chatbots and serious games. Comparators included waitlist, psychoeducation, usual care, or face-to-face therapy; (4) primary outcomes: severity of HA assessed using validated instruments (eg, HAI, SHAI, and Illness Attitude Scales [IASs]). Secondary outcomes: comorbid symptoms (depression and generalized anxiety), functional impairment (Sheehan Disability Scale [SDS] and World Health Organization Disability Assessment Schedule II [WHODAS-II]), quality of life, and treatment adherence/satisfaction; (5) peer-reviewed original articles published in English or Chinese, from database inception to November 1, 2025.

Exclusion Criteria

(1) Noninterventional studies (cross-sectional studies, retrospective analyses, reviews, meta-analyses), no clear definition of groups without sufficient outcome data, duplicate publications, or studies with overlapping data; (2) Studies involving minors or participants without confirmed HA (no DSM diagnosis or subclinical scale scores), or participants with severe psychiatric comorbidities, including psychotic disorders, bipolar disorder, acute suicidal ideation or behavior, or substance use disorders (active within the past 6 months). Studies were also excluded if participants had severe physical illnesses requiring urgent medical intervention or unstable medication use (antidepressants or anxiolytics adjusted within 2 months), or if concurrent psychotherapy targeting HA (eg, face-to-face CBT) was provided during the study period; (3) Intervention types: nondigital interventions (eg, behavioral therapies without digital components or pharmacotherapy alone), or interventions with unclear core components or implementation procedures; (4) Outcomes: absence of primary outcomes related to HA or IAD, incomplete or nonextractable data, or outcome measures lacking psychometric validation; (5) Non-peer-reviewed literature, including conference abstracts, dissertations, and gray literature.

Electronic Databases and Search Strategy

This review aimed to identify studies examining the effects of digital therapeutics (DTs) and digitally assisted behavioral therapies on the mental health outcomes of individuals with HA. Systematic searches were conducted in PubMed, Embase, Web of Science, ScienceDirect, PsycINFO, and the Cochrane Library from database inception to November 1, 2025. The primary search terms for the research question included health anxiety and illness anxiety (ie, “health anxiety” [All Fields] OR “hypochondriasis” [MeSH] OR “hypochondria” [All Fields]). Intervention-related terms included DT, behavioral activation (BA), and DBIs. Gray literature was explicitly excluded (Table 1). The search strategy was independently implemented by 2 reviewers, who initially screened titles and abstracts. To avoid duplication, all retrieved records were imported into the Zotero reference management software [32]. In addition to database searches, supplementary strategies were applied, including snowballing of reference lists, to enhance the comprehensiveness and depth of retrieval. In accordance with the PRISMA guidelines [33], this review primarily included RCTs, supplemented by QES and CS, to systematically evaluate the effectiveness and mechanisms of digital interventions for HA in adults. Highly similar or overlapping studies were excluded (details in Multimedia Appendix 1), and reference lists of relevant prior reviews were also screened to ensure comprehensive inclusion.

XSL•FO
RenderX

Table . Search strategies for English and Chinese databases.

Number	Search terms
1	Health anxiety (MeSH)
2	Health disorder
3	Health anxiety disorder (MeSH)
4	Health-related behavior
5	Fear of illness
6	Illness anxiety disorder (MeSH)
7	Hypochondriasis
8	Hypochondria (MeSH)
9	Somatic symptom disorder (MeSH)
10	1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9
11	Digital assistant (MeSH)
12	Digital interventions
13	Digital therapy (MeSH)
14	Online intervention
15	Internet-based
16	Remote intervention
17	11 OR 12 OR 13
18	Behavioral activation (MeSH)
19	Behavior activation (MeSH)
20	Behavioral activation therapy
21	Behavioral activation therapies
22	Digital behavioral interventions
23	Digital behavior therapy (MeSH)
24	Digital behavior therapies
25	18 OR 19 OR 20 OR 21 OR 22 OR 23 OR 24
26	10 AND 17 AND 25

Selection of Studies

Duplicate records were systematically identified and removed using Zotero. Following deduplication, all unique records underwent a structured screening process conducted by 2 reviewers. Records that were clearly unrelated to the scope of the review were initially excluded on the basis of their titles. Subsequently, the remaining publications were assessed using their titles, abstracts, and keywords. At this point, any record considered potentially relevant by at least one reviewer, or for which eligibility remained uncertain, was retained for full-text assessment.

The full texts of these potentially eligible reports were then examined in detail against the predefined inclusion and exclusion criteria. In particular, eligibility was determined according to the types of anxiety addressed, the forms and characteristics of the interventions, and the measurement scales used to evaluate the core outcomes. During this process, JX sequentially assessed each article to determine whether it met the eligibility requirements for inclusion. Where disagreements concerning full-text inclusion arose, they were resolved through discussion. However, when consensus could not be reached, GH and JW made the final decision. Overall, this sequential screening procedure enabled the systematic identification and evaluation of relevant studies, while the complete study selection process was documented and reported in accordance with the PRISMA guidelines.

Data Extraction

This review used a structured PICOS-based approach for preliminary data extraction from the included studies. Data were extracted independently by 3 reviewers. Approximately 30% of the extracted data were randomly selected and cross-checked by a fourth reviewer to assess consistency and accuracy. This multireviewer extraction and random verification process was used to assess the reliability of the extracted data. A standardized Excel-based extraction form was developed by the research team based on the PICOS framework to systematically collect core study information.

XSL•FO
RenderX

Population (P): individuals diagnosed with hypochondriasis or IAD, as well as individuals exhibiting moderate levels of HA.

Intervention (I): specific categories of DBIs, alongside detailed descriptions of their implementation, as well as forms of intervention deliverable via the internet.

Comparator (C): characteristics and descriptions of control or comparison interventions.

Outcomes (O): the primary outcome must be directly related to HA, while secondary outcome measures include, but are not limited to, other functional impairments resulting from health anxiety, as well as treatment adherence and satisfaction.

Study design (S): study type, including RCTs, QES, or CS.

Additionally, during the data extraction process, if any information was unclear or missing, the research team attempted to contact the original authors for additional data to ensure the completeness of extracted information. All extracted data were systematically organized in a standardized Excel sheet for subsequent quality assessment, data synthesis, and analysis to ensure data reliability and consistency.

Statistical Analysis

When studies were sufficiently comparable in terms of interventions, outcomes, and study designs, quantitative synthesis was undertaken. Following Cochrane methodological guidance, multiple reports of the same trial, including secondary analyses, were linked and treated as a single study [34]. When reports provided overlapping data for the same comparison, outcome, and time point, data from the primary trial report were preferentially used. Secondary reports were consulted to supplement eligible data unavailable in the primary report. Each participant was counted only once within each outcome-specific meta-analysis. All analyses were performed using custom-written code in Python (version 3.12; Python Software Foundation) [35]. For each eligible intervention–control comparison, we extracted postintervention means, SDs, and analyzed sample sizes for both groups. End point values were used consistently, as both end point and change scores are acceptable for continuous-outcome meta-analysis when appropriately analyzed [30].

The effect measure was selected according to the outcome construct and measurement scale. Health-anxiety outcomes assessed using different instruments or different versions of the same instrument, including the Health Anxiety Inventory (HAI) and its short form, the Short Health Anxiety Inventory (SHAI), were pooled using standardized mean differences (SMDs), because these measures assess the same underlying construct but have different scoring ranges [30,36]. SMDs were estimated as Hedges *g*, which applies a small-sample correction to Cohen *d* at the comparison level [37]. Mean differences (MDs) were used when the same outcome was assessed using the same instrument and scoring metric. Score directions were harmonized before analysis. Other prespecified outcomes, including illness attitudes, general anxiety, anxiety sensitivity, and depressive symptoms, were analyzed separately and were not combined with the primary health-anxiety outcome.

For comparison , the MD and its sampling variance were calculated as follows:

MDi=X¯I,i−X¯C,i

VMD,i=SDI,i2nI,i+SDC,i2nC,i

For SMD analyses, the pooled SD across the 2 groups, Hedges *g*, and its sampling variance were calculated using the RevMan algorithm as follows [38]:

SDp,i=(nI,i−1)SI,i2+(nC,i−1)SC,i2nI,i+nC,i−2

gi=|1−34(nI,i+nC,i)−9X˙I,i−X˙C,iSDp,i|

Vg,i=nI,i+nC,inI,inC,i+gi22(nI,i+nC,i−3.94)

Here, and *C* denote the digital intervention and control groups, respectively; ¯, , and *n* denote the postintervention mean (SD), and number of participants analyzed; and , and , denote the corresponding sampling variances. Before pooling, all outcome scales were oriented so that lower scores represented better outcomes; consequently, negative effect estimates indicated effects favoring the digital intervention.

Because the underlying intervention effects were expected to vary across populations, intervention formats, treatment durations, and comparator conditions, an inverse-variance random-effects model was prespecified [38,39]. The between-study variance was estimated using the DerSimonian–Laird method, and the random-effects weight assigned to comparison *i* was calculated as

wi=1Vi+τ2

where $_i$ denotes the sampling variance of the corresponding MD or SMD. Pooled effects were reported with 2-sided Wald 95% CIs, and <.05 was considered statistically significant. Model selection was not based on an ^{2} threshold, because current guidance recommends selecting the statistical model according to assumptions about the underlying effects rather than solely according to a statistical test or threshold for heterogeneity [38].

Robustness was examined using a leave-one-out sensitivity analysis. Within each outcome meta-analysis, each eligible trial-specific comparison was omitted in turn, and the random-effects meta-analysis was refitted using the remaining comparisons. The pooled effect, 95% CI, ^ 2, Cochran , 2 and comparison weights were recalculated after each omission. A trial-specific comparison was considered potentially influential if its omission materially changed the magnitude or direction of the pooled effect, caused the CI to cross the null value, or substantially altered the estimated heterogeneity [40].

Quality Appraisal of the Included Studies

The methodological quality of the included studies was appraised with the Joanna Briggs Institute (JBI) critical appraisal checklists matched to each study design [41]. The 13-item checklist for RCTs covers randomization and allocation concealment, baseline comparability, blinding of participants, treatment providers, and outcome assessors, equal treatment of the groups apart from the intervention, completeness of follow-up, analysis in the randomized groups, consistent and reliable outcome measurement, and appropriate statistical analysis and trial design. The 9-item checklist for QESs

XSL•FO
RenderX

addresses clarity of the cause-and-effect relationship, the presence of a control group, comparability of participants and of the care they received, repeated pre- and postintervention measurements, consistent and reliable outcome measurement, completeness of follow-up, and appropriate statistical analysis. The 11-item checklist for CSs addresses similarity and recruitment of the comparison groups, valid and reliable exposure measurement, identification and handling of confounders, absence of the outcome at baseline, valid and reliable outcome measurement, duration and completeness of follow-up, and appropriate statistical analysis. Following thresholds commonly applied in previous systematic reviews [42] using the JBI checklists, studies meeting ≥70% of applicable items were rated high quality, 50% - 69% moderate, and <50% as low quality. For fully applicable checklists, these correspond to total scores of 10 - 13, 7 - 9, and ≤6 (RCTs); 7 - 9, 5 - 6, and ≤4 (quasi-experimental); and 8 - 11, 6 - 7, and ≤5 (Cohort). The full wording of each item is given in the footnotes to Table 2. Each item was rated as yes, no, unclear, or not applicable. One point was assigned for each yes response and zero points; otherwise, items rated not applicable were excluded from the denominator, and overall quality was expressed as the percentage of applicable items meeting the criteria. Item-level ratings, total scores, and overall classifications for all 16 studies are presented in Table 2. Two reviewers (JX and JS) performed the appraisal independently, interrater agreement was quantified with Cohen kappa, and disagreements were resolved through discussion with a third reviewer (JW). No study was excluded on the basis of its quality rating [43]. The risk of bias of the RCTs and the certainty of the pooled evidence were further assessed.

XSL•FO
RenderX

Table . Quality of evidence in the included 16 reports based on JBI[a]'s critical appraisal tools.

Study design and year	Evaluation for evidence reported													
	1	2	3	4	5	6	7	8	9	10	11	12	13	Total score
RCT[b]														
Hedman et al (2016) [44]	Y[c]	Y	Y	N[d]	N	N	Y	Y	Y	Y	Y	Y	Y	10
Hedman et al (2011) [45]	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	11
Newby et al (2018) [46]	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	11
Axelsson et al (2020) [47]	Y	Y	Y	N	N	N	Y	Y	Y	Y	Y	Y	Y	10
Newby and McElroy (2020) [48]	Y	Y	Y	N	N	N	Y	Y	Y	Y	Y	Y	Y	10
Hoffmann et al (2021) [49]	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	11
Du et al (2024) [50]	Y	Y	Y	N	N	N	Y	Y	Y	Y	Y	Y	Y	10
O'Bryan et al (2021) [51]	Y	N	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	10
Hedman et al (2014) [52]	Y	Y	Y	N	N	N	Y	Y	Y	Y	Y	Y	Y	10
n/N (%)	(9/9) 100	(8/9) 89	(9/9) 100	(2/9) 22	(0/9) 0	(2/9) 22	(9/9) 100	(9/9) 100	(9/9) 100	(9/9) 100	(9/9) 100	(9/9) 100	(9/9) 100	—
QES[e]														
Ingeman et al (2025) [53]	Y	N	Y	Y	Y	Y	Y	N	Y	—[f]	—	—	—	7

XSL•FO
RenderX

Study design and year	Evaluation for evidence reported													
	1	2	3	4	5	6	7	8	9	10	11	12	13	Total score
Shor et al (2023) [54]	Y	N	Y	Y	Y	Y	Y	__[f]	Y	—	—	—	—	7
Sharrock et al (2021) [55]	Y	N	Y	Y	Y	Y	Y	—	Y	—	—	—	—	7
Goonesekera and Donkin (2022) [56]	Y	N	Y	Y	Y	Y	Y	N	Y	—	—	—	—	7
Hoffmann et al (2018) [57]	Y	N	Y	Y	Y	Y	Y	N	Y	—	—	—	—	7
Frydendal et al (2025) [58]	Y	N	Y	Y	Y	Y	Y	—	Y	—	—	—	—	7
n/N (%)	(6/6) 100	(0/6) 0	(6/6) 100	(6/6) 100	(6/6) 100	(6/6) 100	(6/6) 100	(6/6) 0	(6/6) 100	—	—	—	—	—
CS[g]														
Österman et al (2024) [59]	—	—	Y	Y	Y	Y	Y	Y	N	Y	Y	—	—	8
n/N (%)	0	0	100	100	100	100	100	100	0	100	100	—	—	—

[a]JBI: Joanna Briggs Institute.

[b]RCT: randomized controlled trial.

[c]Y: yes.

[d]N: no.

[e]QES: quasi-experimental studies.

[f]Not applicable.

[g]CS: cohort study.

Quality Assessment of Evidence and Risk of Bias

The risk of bias of the included RCTs was assessed with the revised Cochrane Risk of Bias 2 tool for randomized trials (RoB 2) [60]. The tool covers 5 domains. These are bias arising from the randomization process, bias due to deviations from intended interventions, bias due to missing outcome data, bias in measurement of the outcome, and bias in selection of the reported result. Each domain was judged as low risk of bias, some concerns, or high risk of bias. Because RoB 2 is applied to a specific result rather than to a trial as a whole, the assessment addressed the primary outcome of this review, namely HA symptoms at postintervention as measured with the HAI, the SHAI, or the IADs in each trial. An overall judgment was then derived for each trial following the RoB 2 algorithm. A trial was rated as having a low risk when all 5 domains were judged low risk, as some concerns when at least one domain raised some concerns without any high risk judgment, and as high risk when any domain was judged high risk or when several domains raised some concerns to a degree that substantially reduced confidence in the result. Two reviewers independently conducted all assessments in accordance with the Cochrane Handbook for Systematic Reviews of Interventions, and disagreements were resolved through discussion or consultation

XSL•FO
RenderX

with a third reviewer. The methodological quality of quasi-experimental and CSs was appraised with the JBI checklists, as described in section "Quality Appraisal of the Included Studies."

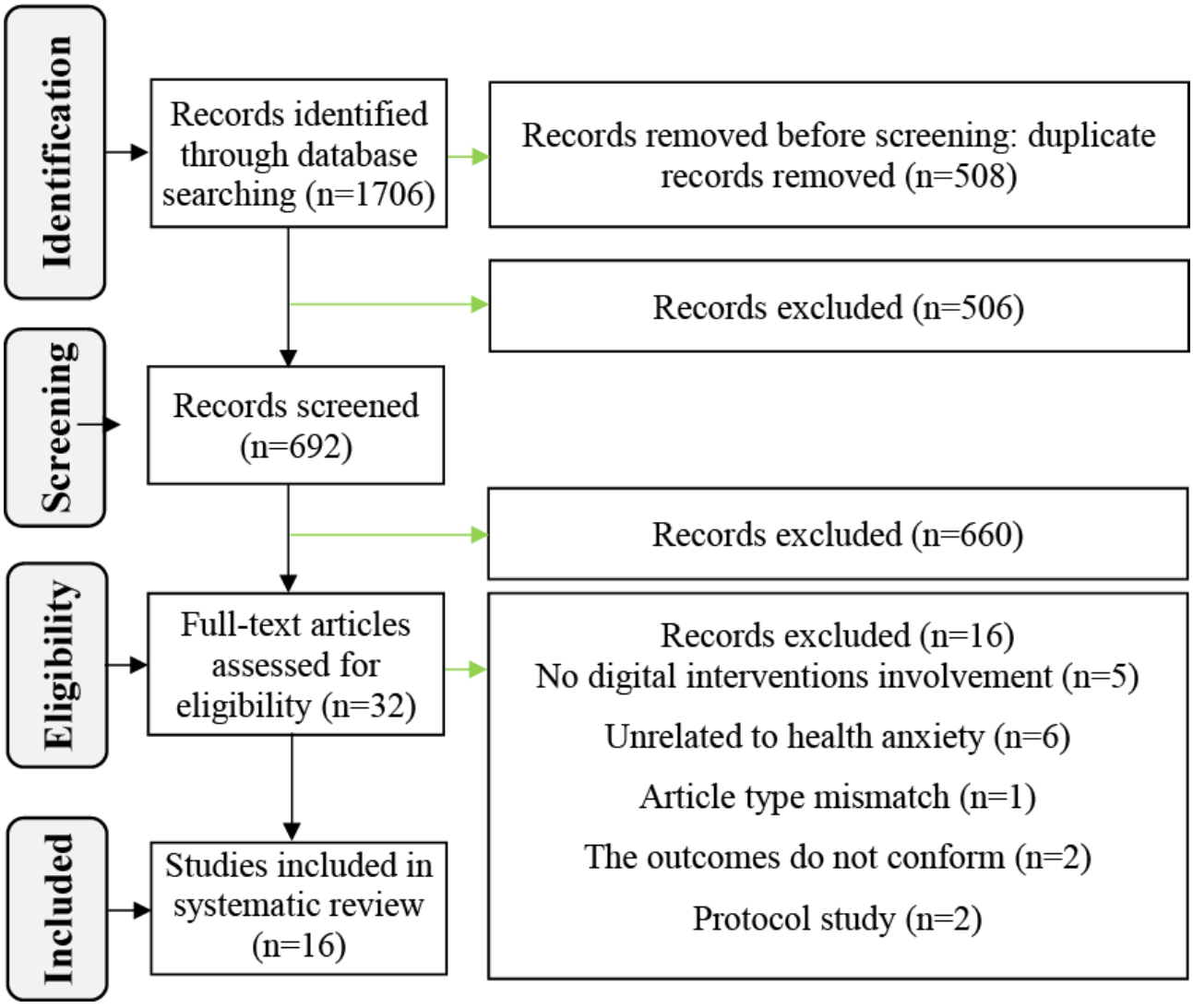
Result

Search Results and Study Characteristics

Overview

A comprehensive search was conducted across 6 electronic databases (PubMed, Embase, Web of Science, ScienceDirect, PsycINFO, and the Cochrane Central Register of Controlled Trials [CENTRAL]). A total of 1706 records were identified through the database searches. After 508 duplicate records were removed, 1198 unique records remained. Title screening excluded 506 records, leaving 692 records for abstract screening. Of these, 660 records were excluded, and the full texts of 32 articles were assessed for eligibility. Sixteen full-text articles were subsequently excluded because they involved nondigital interventions (n=6), did not target HA (n=5), had an ineligible article type (n=1), did not report eligible outcome measures (n=2), or were study protocols (n=2). Ultimately, 16 studies met the predefined eligibility criteria and were included in the systematic review. The study-selection process is presented in Figure 1.

Figure 1. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram for the included and excluded articles.



The 16 included studies were published between 2011 and 2025 and comprised 9 RCTs, 6 QESs, and one CS. Five studies were conducted in Sweden [44,45,47,52,59], 4 in Denmark [49,53,57,58], 3 in Australia [46,48,55], and 1 each in New Zealand [56], China [50], Japan [54], and the United States [51]. A total of 2581 participants were included, with individual study sample sizes ranging from 4 to 904. Intervention duration was most commonly 12 weeks (n=10) [44-49,52,57-59]. The remaining interventions comprised an 8-week program (n=1) [58], a single-session lasting 12 minutes (n=1) [54], a 14-day chatbot program (n=1) [56], a 12-day intervention (n=1) [50], a 1-week computerized intervention (n=1) [51], and a flexible program that allowed participants up to 90 days for completion (n=1) [55]. By study design, the evidence base included RCTs (n=9) [44-52], QESs (n=6) [53-58] and one CS (n=1) [59]. Study populations included adults with SSD/IAD or hypochondriasis (n=4) [44,51,52,59], severe HA (n=5) [29,45,46,52,57], HA without a specific diagnostic label (n=4) [47,50,55,56],

XSL•FO
RenderX

COVID-19-related HA (n=1) [58], HA with cyberchondria (n=1) [48], and HA associated with inflammatory bowel disease (n=1) [54], as well as parents with HA by proxy (n=1) [53]. These population categories were not mutually exclusive.

Delivery mode and intervention approach were considered separately. Regarding delivery support, guided digital interventions were evaluated in 11 studies, whereas unguided interventions were evaluated in 6 studies. One study included both guided and unguided intervention arms. Therefore, these categories were not mutually exclusive. Regarding intervention approach, 8 studies evaluated internet-delivered CBT, including exposure-based CBT. Three studies evaluated internet-delivered acceptance and commitment therapy. One study each evaluated an integrated CBT and acceptance and commitment therapy (ACT) program, a CBT-based chatbot, internet-based cognitive bias modification of interpretations, a video-testimonial intervention, and a brief computerized intervention combining psychoeducation with interpretation-bias modification.

The primary or co-primary outcomes across the included studies included core HA, illness-related attitudes, generalized anxiety, anxiety sensitivity, and depressive symptoms. Core HA was most commonly assessed using the SHAI (n=8) [45,46,48,50,51,55,56,59] and the HAI (n=4) [44,45,47,52]. Illness-related attitudes were assessed using the IAS (n=4) [44,45,47,52], generalized anxiety using the Beck Anxiety Inventory (BAI; n=4) [44,45,47,52], anxiety sensitivity using the Anxiety Sensitivity Index (ASI; n=4) [44,45,47,52], and depressive symptoms using the Montgomery–Åsberg Depression Rating Scale–Self-report (MADRS-S; n=5) [44,45,47,52,59]. Because several studies assessed more than one primary or co-primary outcome, these categories were not mutually exclusive. Detailed outcome measures and study characteristics are presented in Multimedia Appendix 2.

JBI Critical Appraisal Checklist for RCT

The methodological quality of the included randomized controlled trials was assessed using the JBI Critical Appraisal Checklist for Randomized Controlled Trials, which comprises 13 items addressing randomization, allocation concealment, blinding, group comparability, follow-up, outcome measurement, statistical analysis, and trial design. (1) Was true randomization used for assignment of participants to treatment groups? (2) Was allocation to treatment groups concealed? (3) Were treatment groups similar at the baseline? (4) Were participants blind to treatment assignment? (5) Were those delivering treatment blind to treatment assignment? (6) Were outcomes assessors blind to treatment assignment? (7) Were treatment groups treated identically other than the intervention of interest? (8) Was follow-up complete and if not, were differences between groups in terms of their follow-up adequately described and analyzed? (9) Were participants analyzed in the groups to which they were randomized? (10) Were outcomes measured in the same way for treatment groups? (11) Were outcomes measured in a reliable way? (12) Was appropriate statistical analysis used? (13) Was the trial design appropriate, and any deviations from the standard RCT design (individual randomization, parallel groups) accounted for in the conduct and analysis of the trial?

JBI Critical Appraisal Checklist for QESs

The JBI critical appraisal checklist for QESs consists of the following questions: (1) Is it clear in the study what the “cause” is and what the “effect” is (ie, there is no confusion about which variable comes first)? (2) Was there a control group? (3) Were participants included in any comparisons similar? (4) Were the participants included in any comparisons receiving similar treatment/care, other than the exposure or intervention of interest? (5) Were there multiple measurements of the outcome, both pre and post the intervention/exposure? (6) Were the outcomes of participants included in any comparisons measured in the same way? (7) Were outcomes measured in a reliable way? (8) Was follow-up complete and if not, were differences between groups in terms of their follow-up adequately described and analyzed? (9) Was appropriate statistical analysis used?

JBI Critical Appraisal Checklist for CSs

The JBI critical appraisal checklist for CSs consists of the following questions: (1) Were the 2 groups similar and recruited from the same population? (2) Were the exposures measured similarly to assign people to both exposed and unexposed groups? (3) Was the exposure measured in a valid and reliable way? (4) Were confounding factors identified? (5) Were strategies to deal with confounding factors stated? (6) Were the groups/participants free of the outcome at the start of the study (or at the moment of exposure)? (7) Were the outcomes measured in a valid and reliable way? (8) Was the follow-up time reported and sufficient to be long enough for outcomes to occur? (9) Was follow-up complete, and if not, were the reasons for loss to follow-up described and explored? (10) Were strategies to address incomplete follow-up used? (11) Was appropriate statistical analysis used?

Methodological Quality and Risk of Bias

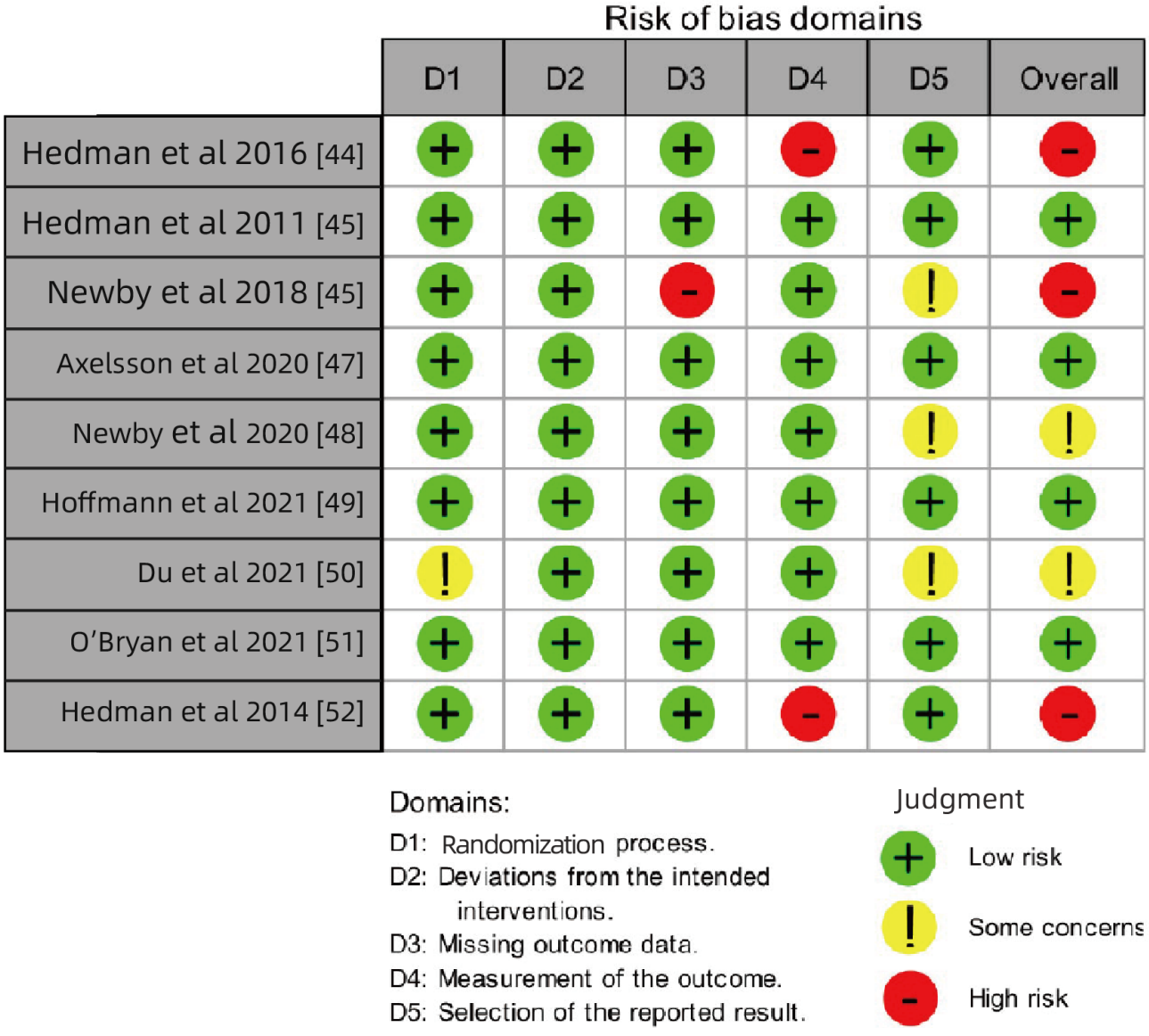
Two reviewers independently assessed the quality of the included studies (see Table 2), and the results showed an interrater agreement of 87.58% (κ=0.81). The 9 RCTs achieved total scores of 10 to 11, and all were classified as high quality. The trials by Hedman et al [45], Newby et al [46], and Hoffmann et al [49] received the highest score of 11, while the remaining studies each scored 10 [44,47,48,50-52]. Across individual assessment domains, RCTs generally demonstrated strong methodological performance in terms of randomization, baseline comparability, completeness of follow-up, consistency and reliability of outcome measurement, and appropriateness of study design. The principal weakness concerned blinding: participants were blinded in 22% of trials, reflecting the inherent difficulty of blinding in digital intervention research. Most studies implemented blinding only at specific stages rather than across the entire trial process. For QES, all 6 studies [53-58] achieved favorable quality scores, meeting high-quality standards regarding causal clarity, participant comparability, consistency and reliability of outcome measurement, and appropriateness of statistical analyses. Nonetheless, incomplete data reporting and lack of a control group remained common limitations. Quality assessment of the single CS [59] rated it as high quality (8 of 9 applicable items); items 1 and 2 were not applicable because a single treated cohort was enrolled without

XSL•FO
RenderX

an unexposed comparison group and follow-up was incomplete, which should be considered when interpreting its findings.

This review focused on HA outcomes and assessed all included RCTs using 5 criteria of the Cochrane Collaboration risk-of-bias assessment tool [61]. Of the 16 included studies, the 9 RCTs were assessed for the HA outcome using the 5-domain Cochrane RoB 2 tool (Figure 2). Overall, 4 trials were rated as having a low risk of bias, 2 as having some concerns, and 3 as having a high risk of bias. For the randomization process (D1), 8 trials were rated as low risk and one as having some concerns. All 9 trials were rated as low risk for deviations from intended interventions (D2). For missing outcome data (D3), 8 trials were rated as low risk and one as high risk. For measurement of the outcome (D4), 7 trials were rated as low risk and 2 as high risk. For selection of the reported result (D5), 6 trials were rated as low risk and 3 as having some concerns. High-risk judgments were recorded for Newby et al [46] in D3 and for Hedman et al [44] and Hedman et al [52] in D4. Some concerns were identified for Du et al [50] in D1 and for Newby et al [46], Newby and McElroy [48], and Du et al [50] in D5. Newby et al [46] were judged to be at high risk in D3 because 16 of 102 randomized participants were excluded before baseline assessment and only 69 provided posttreatment data. The authors also acknowledged that the outcomes of participants with missing data were unknown and that treatment effects might therefore have been overestimated. Hedman et al [44] was rated as having a high risk of bias in D4 because the primary outcome was assessed using an unblinded, participant-reported measure, and comparison with a waiting-list control could have influenced symptom reporting. The D4 judgment for Hedman et al [52] was similarly based on the use of an unblinded self-reported HA measure, although both groups received active interventions with comparable treatment credibility.

Figure 2. Cochrane risk of bias ratings for included randomized controlled trials (n=9) [44-52].



Results of Meta-Analysis

All 6 meta-analyses reported in this section were restricted to RCTs. The 3 quasi-experimental trials, the single CS, and the replicated single-case experimental design are summarized narratively in Table 3 and were not pooled with the RCT effect estimates. All outcomes were synthesized using prespecified random-effects models with the DerSimonian-Laird estimator

XSL•FO
RenderX

(Figure 3). Because HA was measured using both the 64-item HAI (HAI-64 [44,45,52]) and the 18-item SHAI (SHAI-18 [46,50,51]), posttreatment effects were standardized and pooled using Hedges *g*. Across 6 trials involving 585 participants, digital interventions produced a statistically significant, moderate-to-large reduction in HA relative to control conditions (Hedges g=–0.77; 95% CI –1.25 to –0.30; z=–3.17; P=.002) [44-46,50-52]. All trial-level point estimates favored digital interventions and ranged from –1.60 to –0.21, although the CIs of 3 individual trials included zero. Considerable between-trial heterogeneity was observed (τ^2=0.30; Q_5=37.55; P<.001; I^2=86.7%) indicating substantial variation in the magnitude of the intervention effects.

XSL•FO
RenderX

Table . Different digital interventions for health anxiety participants.

Guide type and study	Country	Interventions participants	Period of interventions	Primary outcome	Research type	Follow-up	Result
Guided digital therapy							
Hedman et al (2016)[a] [44]	Sweden	32 (female: 24; male: 8)	12-week	HAI[b]	RCT[c]	Yes, 6-month	Significant decrease
Hedman et al (2011) [45]	Sweden	40 (female: 28; male:12)	12-week	HAI/SHAI[d]	RCT	Yes, 6-month	Significant decrease
Hedman et al (2014) [52]	Sweden	64 (female: 64; male: 15)	12-week	HAI	RCT	Yes, 6-month	Significant decrease
Newby et al (2018) [46]	Australia	45 (female: 40; male: 5)	12-week	SHAI	RCT	Yes, 3-month	Significant decrease
Axelsson et al (2020) [47]	Sweden	102 (female: 72; male: 30)	12-week	HAI-18[e]	RCT	Yes, 12-month	Significant decrease
Newby Jand McElroy (2020) [48]	Australia	45 (female: 40; male: 5)	12-week	SHAI/CSS[f]	RCT	Yes, 3-mouth	Significant decrease
Österman et al (2024) [59]	Sweden	447 (female: 305; male: 142)	12-week	CSS	CS[g]	Yes, 6-mouth	Significant decrease
Ingeman et al 2022 [53]	Denmark	4 (female: 3; male: 1)	8-week	SHAI/HAI	Single-case experimental design	Yes, 14 - 33 days	2 patients showed significant improvement in main outcomes, while one did not; the anxiety symptoms of 2 patients were reduced
Hoffmann et al (2018) [57]	Denmark	15 (female: 12; male: 3)	12-week	WI-7[h]	CS	Yes, 3-month	Significant decrease
Frydendal et al (2025) [58]	Denmark	105 (female: 92; male: 13)	12-week	WI-6R[i]	Pragmatic study	No	Significant decrease
Hoffmann et al (2021) [49]	Denmark	101 (iAC: 53; female: 34; male: 19); iFORUM 48 (female: 32; male: 16)	12-week	WI-7	RCT	Yes, 6-month	Significant decrease
Unguided digital therapy							
Hedman et al (2016) [44]	Sweden	67 (U-ICBT: 33; female: 24; male: 9); bibliotherapy: 34 (female: 24; male: 10)	12-week	HAI	RCT	Yes, 6-month	Significant decrease
Sharrock et al (2021) [55]	Australia	904 (female: 607; male: 220; unknown: 32)	90-day	SHAI	Multiple sets of pre- and posttests	No	Significant decrease
Goonesekera and Donkin (2022) [56]	New Zealand	29 (female: 25; male: 4)	14-day	SHAI	Pilot study	Yes, 12-week	Not significant
Shor et al (2023) [54]	Japan	100 (female: 37; male: 63)	Single session (12 minutes)	VAS-A[j]	CS	No	Significant decrease
Du et al (2023) [50]	China	228 (female: 142; male: 86)	12 days	SHAI	RCT	Yes,1 mouth	Significant decrease

XSL•FO
RenderX

Guide type and study	Country	Interventions participants	Period of interventions	Primary outcome	Research type	Follow-up	Result
O'Bryan et al (2021) [51]	United States	68 (male 14; female: 54)	7-day (single session)	SHAI/ASI-3[k]	RCT	Yes, 1 week and a month	Not significant

[a]Article "Exposure-based cognitive–behavioural therapy via the internet and as bibliotherapy for somatic symptom disorder and illness anxiety disorder: the randomized controlled trial" discussed 3 intervention groups: ICBT, U-ICBT, and Bibliotherapy. Since both guided and unguided intervention formats are involved for each group, they appear twice in the article.

[b]HAI: Health Anxiety Inventory.

[c]RCT: randomized controlled trial.

[d]SHAI: Short Health Anxiety Inventory.

[e]HAI-18: Health Anxiety Inventory-18 item.

[f]CSS: Cyberchondria Severity Scale.

[g]CS:

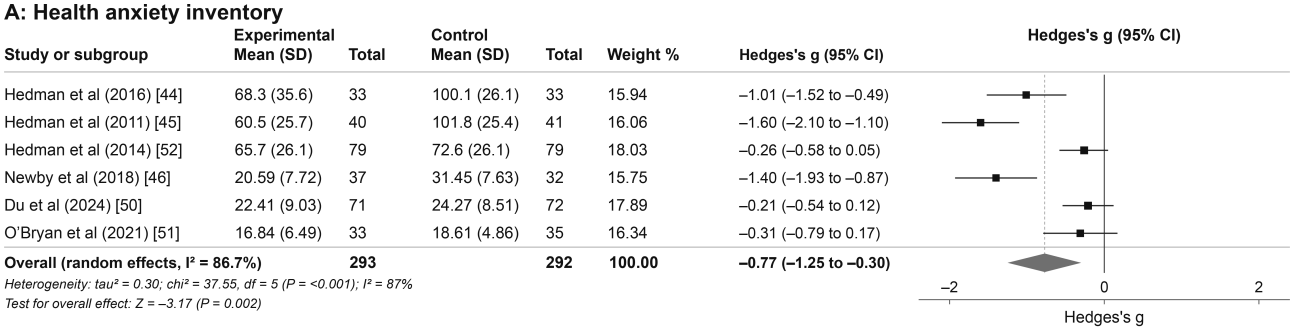
[h]WI-7: Whiteley-7 Index.

[i]WI-6R: Whiteley 6-item revised.

[j]VAS-A: visual analog scale for anxiety.

[k]ASI-3: Anxiety Sensitivity Index-3.

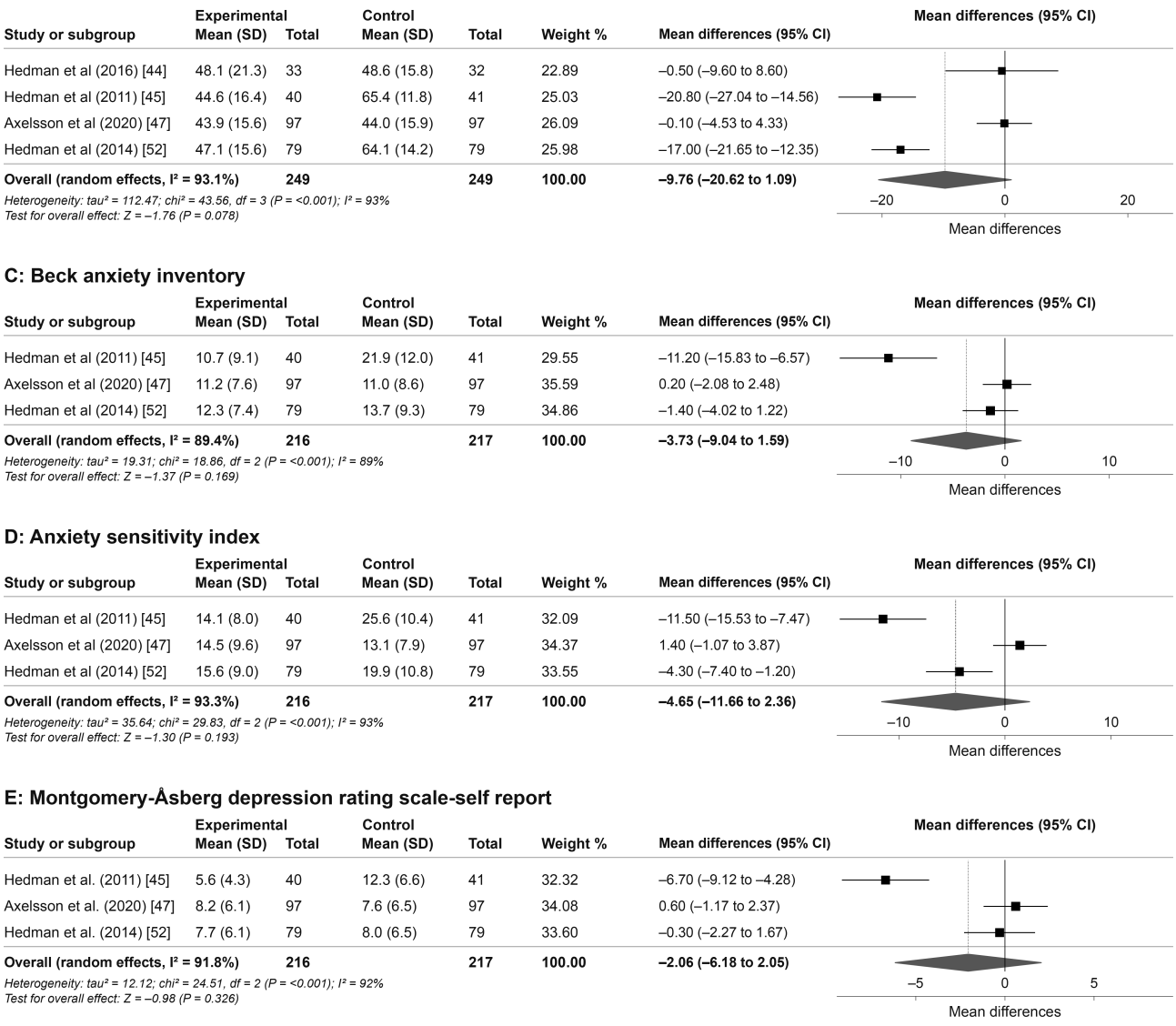
Figure 3. Effects of digital behavior interventions for health anxiety outcomes (standardized mean differences [SMDs]) [44-46,50-52].



For outcomes assessed using the same measurement instrument, mean differences (MDs) were pooled. Four trials involving 498 participants assessed illness-related attitudes using the IAS (Figure 4). Although the pooled estimate favored digital interventions, the effect was not statistically significant (MD –9.76, 95% CI –20.62 to 1.09; z=–1.76; *P*=.08) [44,45,47,52]. Similarly, no statistically significant posttreatment effect was observed for general anxiety measured using the Beck Anxiety Inventory (BAI; 3 trials, n=433; MD –3.73, 95% CI –9.04 to 1.59; z=–1.37; *P*=.17) [44,45,47,52]. Anxiety sensitivity measured using the ASI (3 trials, n=433) also did not differ significantly between digital interventions and control conditions (MD –4.65, 95% CI –11.66 to 2.36; z=–1.30; *P*=.20) [45,47,52]. Likewise, the pooled effect on depressive symptoms measured using the MADRS-S (3 trials, n=433) was not statistically significant (MD –2.06, 95% CI –6.18 to 2.05; z=-0.98; *P*=.33) [45,47,52]. Thus, although all pooled point estimates were in the direction of lower symptom scores following digital interventions, the corresponding CIs included zero.

XSL•FO
RenderX

Figure 4. Effects of digital behavioral interventions on secondary psychological outcomes (mean differences [MDs]) [44,45,47,52].



Note: pre-specified random effects (DerSimonian-Laird) model used for all outcomes.

Substantial heterogeneity was present for the IAS: τ^2=112.47; Q_3=43.56; P<.001; I^2=93.1% [44,45,47,52], BAI: τ^2=19.31; Q_2=18.86; P<.001; I^2=89.4% [45,47,52], ASI: τ^2=35.64; Q_2=29.83; P<.001; I^2=93.3% [45,47,52], and MADRS-S: τ^2=12.12; Q_2=24.51; P<.001; I^2=91.8% [45,47,52].

Leave-one-out sensitivity analyses were performed by sequentially removing each trial and recalculating the pooled effects using the same random-effects model. For the primary health-anxiety outcome, the pooled effect remained statistically significant after the exclusion of any individual trial. The recalculated effect sizes ranged from Hedges g=–0.90 to Hedges g=–0.60, and all 95% CIs excluded zero, supporting the robustness of the overall treatment effect [44-46,50-52]. The smallest pooled effect was obtained after excluding the trial by Hedman et al [45], Hedges g=–0.60, 95% CI –1.02 to –0.18, whereas the largest effect was obtained after excluding the trial by Du et al [50], Hedges g=–0.90, 95% CI –1.46 to –0.34. However, heterogeneity remained substantial across all leave-one-out iterations, with I^2 values ranging from 80.4% to 88.9%, indicating that no single trial fully explained the observed between-trial variability.

The sensitivity analyses produced less stable findings for some secondary outcomes. For the IAS, leave-one-out estimates ranged from MD –13.44 to MD –6.08, with I^2 values ranging from 85.3% to 94.9%. The pooled effect generally remained nonsignificant; however, after the exclusion of the trial by Axelsson et al [47], the effect became statistically significant, MD –13.44, 95% CI –23.14 to –3.74; I^2=85.3%. This result indicates that the pooled IAS finding was sensitive to the inclusion of the Axelsson et al [47] trial, although substantial residual heterogeneity remained.

For the BAI, all leave-one-out estimates remained statistically nonsignificant and ranged from MD –6.11 to MD –0.49, supporting the stability of the primary null finding [45,47,52]. Notably, excluding the trial by Hedman et al [45] reduced the pooled effect to MD –0.49, 95% CI –2.21 to 1.23, and eliminated the observed heterogeneity, I^2=0%. In contrast, heterogeneity remained high when either the Axelsson et al [47]

XSL•FO
RenderX

or Hedman et al [52] trial was excluded, I^2=92.3% and 94.7%, respectively. These findings suggest that the Hedman et al [45] trial accounted for a substantial proportion of the heterogeneity in the BAI analysis.

For the ASI, leave-one-out estimates ranged from MD −7.78 to MD −1.37, with residual heterogeneity remaining substantial in all analyses, I^2=87.0% - 96.5% [45,47,52]. Exclusion of the Axelsson et al [47] trial resulted in a statistically significant pooled effect, MD –7.78, 95% CI –14.83 to –0.73; I^2=87.0%, whereas the effect remained nonsignificant following the exclusion of either the Hedman et al [45] or Hedman et al [52] trial. Therefore, similar to the IAS finding, the ASI result was sensitive to the inclusion of the Axelsson et al [47] trial.

For the MADRS-S, the pooled estimates remained nonsignificant regardless of which trial was removed and ranged from MD –3.46 to MD 0.20 [45,47,52]. Excluding the trial by Hedman et al [45] shifted the pooled estimate toward the null, MD 0.20, 95% CI –1.12 to 1.51, and reduced heterogeneity from I^2=91.8% to 0%. By comparison, heterogeneity remained high after excluding the Axelsson et al [47] or Hedman et al [52] trial, I^2=93.8% and 95.6%, respectively. This pattern suggests that the Hedman et al [45] trial exerted a substantial influence on both the direction of the pooled MADRS-S estimate and the observed heterogeneity, although its exclusion did not alter the conclusion regarding statistical nonsignificance.

Overall, the leave-one-out analyses supported the robustness of the significant posttreatment effect on HA and the nonsignificant findings for the BAI and MADRS-S. In contrast, the IAS and ASI findings were less stable because the exclusion of the Axelsson et al [47] trial changed their pooled effects from nonsignificant to statistically significant. These secondary-outcome results should therefore be interpreted cautiously, particularly given the small number of contributing trials and the persistently high heterogeneity.

Although the consistently high values indicated marked between-trial heterogeneity, they should not be interpreted as evidence of moderation. No formal subgroup-difference tests or meta-regression analyses were conducted. Differences in scoring metrics between the HAI-64 and SHAI-18 were addressed by pooling standardized effect estimates, whereas the leave-one-out analyses evaluated the influence of individual trials; neither procedure constituted a moderator analysis.

Grading of Recommendations Assessment, Development, and Evaluation Assessment

This study used the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach to assess the certainty of evidence from the meta-analysis results (Table 4). Overall, the certainty of evidence for the 5 outcomes ranged from low to very low, and all 5 outcomes were informed exclusively by RCTs. Specifically, the meta-analysis of 6 RCTs regarding HA was rated as low-certainty evidence. Although the pooled estimate favored the intervention and the 95% CI did not include the null value, so that the estimate was not downgraded for imprecision, the rating was lowered because of risk-of-bias concerns across the included trials and substantial unexplained heterogeneity (I^2=86.7%). The meta-analysis for illness attitudes (IAS) was based on 4 RCTs and was rated as very low-certainty evidence, owing to risk-of-bias concerns, materially different point estimates across trials with considerable heterogeneity, and imprecision of the pooled estimate. General anxiety symptoms (BAI), anxiety sensitivity (ASI), and depressive symptoms (MADRS-S) were each assessed on the basis of 3 RCTs, and each was rated as very low-certainty evidence for the same 3 reasons: risk of bias, inconsistency, and imprecision, with the 95% CI in each case including both no effect and a potentially important benefit. No outcome was downgraded for indirectness, and publication bias was not formally assessed because fewer than 10 trials contributed to each outcome. Overall, the GRADE assessment indicates that the current evidence for digital interventions in HA remains limited. Although the pooled estimate for health anxiety favored the intervention, confidence in the effects across outcomes is constrained by methodological limitations of the contributing trials, between-study heterogeneity, and limited precision. Overall, the available evidence suggests that digital interventions may reduce HA symptoms; however, substantial uncertainty remains regarding the magnitude of the true effect and their effects on other psychological outcomes.

XSL•FO
RenderX

Table . GRADE[a] assessment of the certainty of evidence for the effects of digital interventions for health anxiety compared with control conditions.

Outcome	Number of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Participants, intervention/control (n)	Effect estimate (95% CI)	Certainty
Health Anxiety Inventory	6	Randomized trials	Serious[b]	Serious[c]	Not serious[d]	Not serious[e]	Undetected[f]	293/292 (n=585)	Hedges *g*=–0.77 (–1.25 to –0.30)	Low
Illness Attitude Scales	4	Randomized trials	Serious[b]	Serious[g]	Not serious[d]	Serious[h]	Undetected[f]	249/249 (n=498)	MD[i]=–9.76 (–20.62 to 1.09)	Very low
Beck Anxiety Inventory	3	Randomized trials	Serious[j]	Serious[g]	Not serious[d]	Serious[h]	Undetected[f]	216/217 (n=433)	MD=–3.73 (–9.04 to 1.59)	Very low
Anxiety Sensitivity Index	3	Randomized trials	Serious[j]	Serious[g]	Not serious[d]	Serious[h]	Undetected[f]	216/217 (n=433)	MD=–4.65 (–11.66 to 2.36)	Very low
Montgomery-Åsberg Depression Rating Scale-Self report	3	Randomized trials	Serious[j]	Serious[g]	Not serious[d]	Serious[h]	Undetected[f]	216/217 (n=433)	MD=–2.06 (–6.18 to 2.05)	Very low

[a]GRADE: Grading of Recommendations Assessment, Development, and Evaluation.

[b]Downgraded one level for risk of bias because a substantial proportion of the evidence was contributed by studies rated as high risk or having some concerns in the overall Risk of Bias 2 assessment.

[c]Downgraded one level for inconsistency because of substantial unexplained heterogeneity (I^2=86.7%), although the direction of effect was generally consistent.

[d]Not downgraded for indirectness because the populations, interventions, comparators, and outcomes were considered consistent with the review question.

[e]Not downgraded for imprecision because the 95% CI did not include the null value and consistently favored the intervention.

[f]Formal statistical assessment of publication bias was not performed because fewer than 10 studies contributed to each outcome; no downgrade was applied for publication bias.

[g]Downgraded one level for serious inconsistency because the point estimates differed materially across studies and considerable heterogeneity was present (I^2=89.4% - 93.3%).

[h]Downgraded one level for imprecision because the 95% CI included both no effect and a potentially important benefit.

[i]MD: mean difference.

[j]Downgraded one level for risk of bias because one of the 3 contributing studies was rated as having a high overall risk of bias.

Additionally, the digital interventions for HA can be categorized into guided and unguided DTs, each with distinct characteristics. Table 3 summarizes the analytical results of digital interventions for HA, categorized primarily by guidance type into guided and unguided digital therapies. In the guided digital therapy subgroup, the majority of studies (n=10) [44-52,59] used RCT or CS designs, with intervention durations to 12 weeks. These studies were predominantly conducted in Sweden (n=5) [44,45,47,52,59], Denmark (n=4) [49,53,57,58], and Australia (n=2) [46,48], using HAI or SHAI as primary outcomes, often in combination with the Whiteley-7 Index. A key finding was that, except for one small single-case study, all guided interventions reported significant reductions or improvements in primary outcomes, with most including follow-up periods ranging from 3 to 12 months, indicating the durability of digital intervention effects on HA.

In contrast, studies of unguided digital therapies exhibited greater heterogeneity, with intervention durations ranging from 12-week RCTs to single-session interventions (eg, 12 minutes or 7 days). These studies were conducted across a broader geographic distribution, including Sweden (n=1) [44], New Zealand (n=1) [56], the United States (n=1) [51], Australia (n=1) [55], China (n=1) [50], and Japan (n=1) [54]. Although most nonguided interventions have reported significant efficacy, some short-term pilot studies (eg, 14 days) or single-session interventions failed to show significant results. These approaches encompass formats such as unguided ICBT, self-help reading materials, video testimonials, and mobile-based iCBM-I. Additionally, 2 studies [51,56] (involving chatbots and computerized anxiety sensitivity interventions) found no significant effects. These descriptive patterns suggest that the guidance model and intervention dosage might be important factors.

XSL•FO
RenderX

The RCT included 3 intervention groups, including ICBT, unguided ICBT, and bibliotherapy. Since both guided and unguided intervention formats are involved for each group, they appear twice in the article.

Discussion

Principal Findings

This systematic review and meta-analysis examined the effectiveness of digital interventions for health anxiety across 16 studies, including 9 RCTs. Six RCTs involving 585 participants contributed to the primary meta-analysis and showed a significant reduction in health-anxiety symptoms (Hedges *g*=–0.77, 95% CI –1.25 to –0.30; *P*=.002). By contrast, pooled effects for IAS, BAI, ASI, and MADRS-S, although generally favoring digital interventions, did not reach statistical significance. Because no formal comparisons between outcomes were conducted and the analyses used different effect measures, this pattern should not be interpreted as proof that the effect on HA was statistically larger than the effects on broader psychological outcomes. Considerable heterogeneity was observed across all outcomes (I^2=86.7% - 93.3%), while GRADE certainty was low for HA and very low for the secondary outcomes. Importantly, leave-one-out analyses showed that the health-anxiety effect remained significant after exclusion of any individual trial, with estimates ranging from Hedges *g*=–0.90 to –0.60, despite persistently high heterogeneity. In comparison, IAS and ASI became significant after exclusion of Axelsson et al [47], whereas BAI and MADRS-S remained nonsignificant throughout. Taken together, these findings indicate a statistically stable effect on health-anxiety-specific symptoms across leave-one-out analyses, whereas evidence for broader psychological outcomes remains less stable and more uncertain.

This pattern is broadly consistent with the mechanisms targeted by the included interventions. HA is characterized by catastrophic interpretations of benign bodily sensations and is maintained by reassurance seeking, repeated checking, and avoidance [62,63]. CBT directly targets these processes through exposure and cognitive restructuring [64], whereas ACT seeks to reduce behavioral restriction associated with illness-related thoughts and promote engagement in meaningful activities [65]. Digital delivery may further support structured practice, symptom monitoring, and repeated therapeutic exercises [66], while its accessibility and privacy may reduce barriers to conventional psychological care [67,68]. The observation that the pooled effect was significant for HA but not for the broader psychological outcomes may also reflect measurement specificity. HAI and SHAI directly assess illness-related worry, bodily vigilance, and fear of illness and therefore closely correspond to illness-focused treatment targets, whereas BAI, ASI, and MADRS-S capture broader anxiety-related or depressive symptoms. Although these constructs overlap with HA, they are not necessarily primary treatment targets [69,70]. Accordingly, reductions in illness-related catastrophic interpretations may occur without parallel improvements in general anxiety or depressive symptoms [71]. Nevertheless, the present meta-analysis cannot determine which therapeutic components or measurement characteristics account for this differential pattern.

Beyond outcome specificity, variation in intervention intensity and clinical presentation may also have contributed to the heterogeneous findings. Intervention duration ranged from a single 12-minute session to programs allowing completion within 90 days, although structured 12-week programs were most common. From a descriptive standpoint, positive outcomes were more consistent in interventions that were longer in duration and guided; this may be linked to the greater willingness of anxious individuals to engage in “in-person social interactions” or “real-world activities” that serve to shift attention. In contrast, several brief or unguided interventions failed to yield significant results. However, duration was not formally evaluated as a moderator, and treatment length was intertwined with differences in intervention content, population, comparator, and outcome assessment. Therefore, this pattern should be regarded as hypothesis-generating rather than evidence of a dose-response relationship. Clinical heterogeneity adds further uncertainty. The included populations encompassed SSD, IAD, hypochondriasis, severe or subclinical HA, COVID-19-related HA, cyberchondria, HA associated with physical illness, and HA by proxy. Differences in baseline severity, comorbidity, and the nature of the perceived health threat may therefore have contributed to between-study variability and limit the extent to which a single pooled estimate can represent all presentations of HA.

At a broader conceptual level, this clinical diversity is relevant to the continuing debate concerning the nosological position of HA. Although *DSM-5* classifies IAD and SSD within “Somatic Symptom and Related Disorders,” HA shares several cognitive and behavioral features with generalized anxiety disorder, panic disorder, and obsessive-compulsive disorder, including threat hypervigilance, catastrophic interpretation, and excessive safety behaviors [72]. HA has therefore been proposed to fall, at least partly, within an anxiety-spectrum framework [73]. The present findings are compatible with, but do not establish, this interpretation. In particular, the favorable response of health-anxiety symptoms to interventions largely derived from established anxiety treatments, especially CBT, is consistent with previous evidence [74]. However, the absence of significant pooled effects on BAI and ASI indicates that treatment response alone cannot resolve its nosological classification. A related distinction may also exist between persistent HA and more context-dependent disease-specific fear. The included COVID-19-related study suggests that digital interventions may be applicable to both presentations [58], while longer-term evidence indicates that health-anxiety and general-anxiety trajectories may remain related over time [75]. Thus, future studies may benefit from examining whether more state-like versus trait-like characteristics influence treatment response and maintenance rather than assuming that HA represents a clinically uniform construct.

The distinction between guided and unguided interventions should likewise be interpreted cautiously. Narratively, favorable outcomes appeared relatively consistent across many guided ICBT- and internet-based acceptance and commitment therapy (iACT)–based interventions, whereas unguided interventions

XSL•FO
RenderX

showed greater variability. Therapist support may facilitate adherence, identify avoidance of therapeutic exercises, and strengthen engagement [66,76]. Nevertheless, the current evidence does not establish the superiority of guided interventions: some unguided or brief approaches also produced favorable outcomes, and direct comparisons have not consistently demonstrated significant between-group differences [44]. Moreover, guidance was confounded with intervention duration, therapeutic approach, and population characteristics. Its apparent advantage should therefore be viewed as a descriptive pattern requiring confirmation in adequately powered RCTs that directly compare otherwise equivalent guided and unguided programs. From a clinical perspective, digital interventions may consequently represent a potentially useful low-intensity or adjunctive option for HA, particularly where access to face-to-face psychological care is limited, such as in primary care, general hospital, geographically remote, or resource-constrained settings [12,77]. Their effects on health care use, long-term outcomes, and cost-effectiveness remain to be established.

Finally, implementation should consider digital health equity alongside effectiveness. Although remote delivery may reduce some barriers to psychological care, disparities in device access, internet connectivity, digital literacy, language, and socioeconomic resources may restrict participation among some populations. Accessible interfaces, culturally and linguistically adapted content, appropriate user support, and alternative routes to care should accompany future implementation. Overall, the present review provides evidence that digital interventions can reduce core health-anxiety symptoms, with the primary effect remaining robust across sensitivity analyses. The pooled effects on broader psychological outcomes were nonsignificant, while guided delivery, intervention intensity, and clinical phenotype remain plausible but unconfirmed explanations for between-study variation. Future RCTs should clarify which patients benefit most and under what delivery conditions.

Limitations and Potential of the Study

Several methodological limitations constrain confidence in these findings. First, the quantitative evidence base was limited. Although 9 RCTs were included, 6 contributed to the primary meta-analysis, and most secondary-outcome meta-analyses contained only 3 or 4 trials, restricting precision and the ability to evaluate the robustness of pooled findings. This protocol deviation should be considered when interpreting the robustness of the findings. Second, among the 9 RCTs, 4 had low overall RoB, 2 had some concerns, and 3 had high RoB. Measurement of the outcome was the most frequent source of high-risk judgments, particularly where participant-reported outcomes were used without blinding. Because participant blinding is inherently difficult in digital psychological interventions [78,79], expectancy effects, response bias, and outcome-measurement bias cannot be excluded. Future trials may strengthen assessment by combining self-report measures with blinded diagnostic evaluation and, where appropriate, behavioral, physiological, or sensor-derived indicators [44,80], while carefully designed active or sham digital comparators may help reduce differences in treatment expectancy [81].

Third, substantial statistical heterogeneity was observed, with I^2 values exceeding 80% for every pooled outcome and remaining high in most sensitivity analyses. No subgroup-difference tests or meta-regression analyses were undertaken; consequently, intervention type, guidance, duration, diagnostic status, and baseline severity cannot be regarded as established moderators. Clinical and methodological diversity may have contributed to this variability. The included populations represented several diagnostic and contextual presentations, and the interventions differed in therapeutic components, including exposure, cognitive restructuring, acceptance-based strategies, and bias modification. Differences in duration, professional guidance, adherence, comparators, and assessment timing may have further contributed to variation in the effect estimates.

Fourth, where follow-up was reported, most periods were 6 months or less, leaving the longer-term durability of digital interventions uncertain. In addition, pooled RCT estimates did not demonstrate statistically significant effects on IAS, BAI, ASI, or MADRS-S. This finding should not be generalized to every individual study and remains uncertain because of the small number of contributing trials, substantial heterogeneity, and limited precision. Publication bias could not be formally assessed because fewer than 10 studies contributed to each meta-analysis, and unpublished null findings cannot be excluded. Overall, the low GRADE certainty for the primary outcome and very low certainty for the secondary outcomes preclude strong clinical recommendations.

Despite these limitations, digital interventions remain promising accessible and scalable options for HA, particularly where face-to-face care is limited. To translate this potential into reliable clinical benefit, future research should prioritize several directions.

First, RCTs should directly compare the long-term effectiveness and cost-effectiveness of internet-delivered and face-to-face CBT. Second, predictors such as baseline severity, comorbidity, digital literacy, and treatment preference should be examined to improve matching between guided and unguided interventions. Third, rapid diagnostic tools should distinguish SSD, IAD, and nondiagnostic presentations in primary care and evaluate the added value of health-literacy education [82]. Fourth, studies should test whether somatosensory technologies improve engagement and adherence. AI should follow principles of transparency, data minimization, and clinical oversight [83]. Because severe HA may involve distress during exposure or unintended interpretations of digital content, evidence is needed on the suitability of self-guided interventions for patients with suicidal ideation, psychotic disorders, or severe physical illness. Future systems should include risk monitoring and clear escalation pathways [84] to support safe and trustworthy digital mental health care. Finally, alongside these clinical and methodological priorities, future implementation and research must consider digital health equity, including access, literacy, and cultural adaptation, to ensure that these interventions benefit diverse populations

XSL•FO
RenderX

Conclusions

The available evidence suggests that digital interventions can reduce core health-anxiety symptoms, with the pooled effect remaining significant across sensitivity analyses. However, substantial between-study heterogeneity, risk-of-bias concerns, and low certainty of evidence limit confidence in the magnitude and generalizability of this effect. Evidence for illness-related attitudes, general anxiety, anxiety sensitivity, and depressive symptoms remains inconclusive, and the current data do not establish which intervention formats, levels of guidance, or patient characteristics account for differences in treatment response. Digital interventions may therefore represent a potentially useful low-intensity or adjunctive option for HA, particularly where access to conventional psychological care is limited, but firm conclusions regarding broader psychological benefits and long-term effectiveness cannot yet be drawn. Adequately powered, well-reported RCTs using rigorous outcome assessment, appropriate active comparators, and longer follow-up are needed to confirm these findings and clarify which patients benefit most and under what delivery conditions.

Generative AI Use

AI was only used to refine language expression and check grammatical errors in this manuscript.

Funding

The authors declare that financial support was received for the research, authorship, and/or publication of this article. This study was funded by the Science and Technology Development Fund, Macau SAR (0023/2024/ITP1).

Data Availability

Data are presented in the main manuscript

Authors' Contributions

JX and JS developed the concept of the study and led the development of the protocol in collaboration with other team members. JX performed the online searches. JS and JW screened the articles for inclusion. GH and JX extracted data from the included articles. JW and JS critically appraised the studies. JX led the JBI assessment and presentation of the figures, in collaboration with JS, and other team members, who helped inform the meta-analysis, interpretation of the results, and formulation of recommendations. All coauthors reviewed and approved the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Studies excluded with reason (N=16).
[DOCX File, 23 KB - mhealth_v14i1e91108_app1.docx]

Multimedia Appendix 2
Characteristics of the included studies on digital interventions in health anxiety (n=16).
[DOCX File, 88 KB - mhealth_v14i1e91108_app2.docx]

Checklist 1
PRISMA 2020 checklist.
[PDF File, 101 KB - mhealth_v14i1e91108_app3.pdf]

References

1. Diagnostic and Statistical Manual of Mental Disorders: DSM-5, 5th edition: American Psychiatric Association; 2013. [doi: 10.1176/appi.books.9780890425596]
2. Depression and other common mental disorders: global health estimates. World Health Organization. 2017. URL: https://iris.who.int/server/api/core/bitstreams/6bab42bc-df0f-4f68-a86d-28ebedb85e42/content [accessed 2026-08-10]
3. Szuhany KL, Simon NM. Anxiety disorders: a review. JAMA 2022 Dec 27;328(24):2431-2445. [doi: 10.1001/jama.2022.22744] [Medline: 36573969]
4. Sunderland M, Newby JM, Andrews G. Health anxiety in Australia: prevalence, comorbidity, disability and service use. Br J Psychiatry 2013 Jan;202(1):56-61. [doi: 10.1192/bjp.bp.111.103960] [Medline: 22500013]
5. Tyrer P, Eilenberg T, Fink P, Hedman E, Tyrer H. Health anxiety: the silent, disabling epidemic. BMJ 2016 Apr 25;353:i2250. [doi: 10.1136/bmj.i2250] [Medline: 27112356]
6. Scarella TM, Boland RJ, Barsky AJ. Illness anxiety disorder: psychopathology, epidemiology, clinical characteristics, and treatment. Psychosom Med 2019 Jun;81(5):398-407. [doi: 10.1097/PSY.0000000000000691] [Medline: 30920464]

XSL•FO
RenderX

7. Tyrer P, Cooper S, Tyrer H, Wang D, Bassett P. Increase in the prevalence of health anxiety in medical clinics: possible cyberchondria. Int J Soc Psychiatry 2019 Nov;65(7-8):566-569. [doi: 10.1177/0020764019866231] [Medline: 31379243]
8. Cathébras P, Kohout É, Savall A, Goutte J. L'anxiété centrée sur la santé: une autre épidémie à laquelle il faut se préparer. La Revue de Médecine Interne 2024 Oct;45(10):641-648. [doi: 10.1016/j.revmed.2024.03.002]
9. Nicholas D, Huntington P, Gunter B, Withey R, Russell C. The British and their use of the web for health information and advice: a survey. Aslib Proc 2003 Dec 1;55(5/6):261-276. [doi: 10.1108/00012530310498842]
10. Loos A. Cyberchondria: too much information for the health anxious patient? J Consum Health Internet 2013 Oct;17(4):439-445. [doi: 10.1080/15398285.2013.833452]
11. Barsky AJ, Ettner SL, Horsky J, Bates DW. Resource utilization of patients with hypochondriacal health anxiety and somatization. Med Care 2001 Jul;39(7):705-715. [doi: 10.1097/00005650-200107000-00007] [Medline: 11458135]
12. Tyrer P. Recent advances in the understanding and treatment of health anxiety. Curr Psychiatry Rep 2018 Jun 22;20(7):49. [doi: 10.1007/s11920-018-0912-0] [Medline: 29931576]
13. Kamali M, Mousavi SK, Akbari F, Khodabandeh M, Moharramkhani M. Relationship between personality traits and health anxiety among nurses. J Holist Nurs Midwifery 2024 Feb 1;34(2):159-166. [doi: 10.32598/jhnm.34.2.2591]
14. Mahmood QK, Jafree SR, Jalil A, Nadir SMH, Fischer F. Anxiety amongst physicians during COVID-19: cross-sectional study in Pakistan. BMC Public Health 2021 Jan 11;21(1):118. [doi: 10.1186/s12889-020-10134-4] [Medline: 33430852]
15. Benke C, Wallenfels LM, Bleichhardt GM, Melzig CA. Health anxiety amplifies fearful responses to illness-related imagery. Sci Rep 2024 Feb 22;14(1):4345. [doi: 10.1038/s41598-024-54985-y] [Medline: 38388793]
16. Eichenberg C, Schott M. Use of web-based health services in individuals with and without symptoms of hypochondria: survey study. J Med Internet Res 2019 Jun 10;21(6):e10980. [doi: 10.2196/10980] [Medline: 31199311]
17. Dutta R. The reports are normal, but I still have a doubt: managing hypochondriasis using cognitive-behavioral interventions. Int J Cogn Behav Ther 2025;18(4):655-671. [doi: 10.1007/s41811-025-00261-4]
18. Mak WWS, Tong ACY, Fu ACM, et al. Efficacy of internet-based rumination-focused cognitive behavioral therapy and mindfulness-based intervention with guided support in reducing risks of depression and anxiety: a randomized controlled trial. Appl Psychol Health Well Being 2024 May;16(2):696-722. [doi: 10.1111/aphw.12512] [Medline: 38073271]
19. Österman S, Hentati A, Forsell E, et al. Brief digital self-care intervention for health anxiety in a Swedish Medical University Clinic: a prospective single-group feasibility study. BMJ Open 2023 Dec 21;13(12):e077376. [doi: 10.1136/bmjopen-2023-077376] [Medline: 38135329]
20. Roy A, Hoge EA, Abrante P, Druker S, Liu T, Brewer JA. Clinical efficacy and psychological mechanisms of an app-based digital therapeutic for generalized anxiety disorder: randomized controlled trial. J Med Internet Res 2021 Dec 2;23(12):e26987. [doi: 10.2196/26987] [Medline: 34860673]
21. Zeng S, Chen L, Lan S. Research on the extension of respiratory interaction modalities in virtual reality technology and innovative methods for healing anxiety disorders. Sci Rep 2025 Mar 7;15(1):7936. [doi: 10.1038/s41598-025-92419-5] [Medline: 40050356]
22. Andersson G. Internet-delivered psychological treatments. Annu Rev Clin Psychol 2016;12:157-179. [doi: 10.1146/annurev-clinpsy-021815-093006] [Medline: 26652054]
23. Andrews G, Basu A, Cuijpers P, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78. [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
24. Fang S, Mushtaque I. The moderating role of health literacy and health promoting behavior in the relationship among health anxiety, emotional regulation, and cyberchondria. Psychol Res Behav Manag 2024;17:51-62. [doi: 10.2147/PRBM.S446448] [Medline: 38196775]
25. Linehan MM. DBT Skills Training Manual Revised Edition: Guilford Press; 2025.
26. Ekers D, Webster L, Van Straten A, Cuijpers P, Richards D, Gilbody S. Behavioural activation for depression; an update of meta-analysis of effectiveness and sub group analysis. PLoS One 2014;9(6):e100100. [doi: 10.1371/journal.pone.0100100] [Medline: 24936656]
27. Kikas K, Werner-Seidler A, Upton E, Newby J. Illness anxiety disorder: a review of the current research and future directions. Curr Psychiatry Rep 2024 Jul;26(7):331-339. [doi: 10.1007/s11920-024-01507-2] [Medline: 38748190]
28. Rask CU, Duholm CS, Poulsen CM, Rimvall MK, Wright KD. Annual research review: health anxiety in children and adolescents-developmental aspects and cross-generational influences. J Child Psychol Psychiatry 2024 Apr;65(4):413-430. [doi: 10.1111/jcpp.13912] [Medline: 37909255]
29. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
30. Higgins JPT, Thomas J. In: Chandler J, editor. Cochrane Handbook for Systematic Reviews of Interventions, 2nd edition: John Wiley & Sons; 2019.
31. Amir-Behghadami M, Janati A. Population, intervention, comparison, outcomes and study (PICOS) design as a framework to formulate eligibility criteria in systematic reviews. Emerg Med J 2020 Jun;37(6):387. [doi: 10.1136/emermed-2020-209567] [Medline: 32253195]

XSL•FO
RenderX

32. Kratochvíl J. Comparison of the accuracy of bibliographical references generated for medical citation styles by EndNote, Mendeley, RefWorks and Zotero. J Acad Libr 2017 Jan;43(1):57-66. [doi: 10.1016/j.acalib.2016.09.001]
33. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8(5):336-341. [doi: 10.1016/j.ijsu.2010.02.007] [Medline: 20171303]
34. Lefebvre C, Glanville J, Briscoe S, et al. Chapter 4: searching for and selecting studies. In: Cochrane Handbook for Systematic Reviews of Interventions (Version 651): Cochrane; 2025:67-107. [doi: 10.1002/9781119536604]
35. Python Software Foundation. Python 3.12.0. Python.org. 2023 Oct 2. URL: https://www.python.org/downloads/release/python-3120/ [accessed 2026-08-10]
36. Salkovskis PM, Rimes KA, Warwick HMC, Clark DM. The health anxiety inventory: development and validation of scales for the measurement of health anxiety and hypochondriasis. Psychol Med 2002 Jul;32(5):843-853. [doi: 10.1017/s0033291702005822] [Medline: 12171378]
37. Hedges LV. Distribution theory for glass's estimator of effect size and related estimators. J Educ Stat 1981 Jun;6(2):107-128. [doi: 10.3102/10769986006002107]
38. Deeks JJ, Higgins JPT, Altman DG, Cochrane Statistical Methods Group. Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, editors. Cochrane Handbook for Systematic Reviews of Interventions, 2nd edition: John Wiley & Sons; 2019:241-284. [doi: 10.1002/9781119536604]
39. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods 2010 Apr;1(2):97-111. [doi: 10.1002/jrsm.12] [Medline: 26061376]
40. Viechtbauer W, Cheung MWL. Outlier and influence diagnostics for meta-analysis. Res Synth Methods 2010 Apr;1(2):112-125. [doi: 10.1002/jrsm.11] [Medline: 26061377]
41. Barker TH, Stone JC, Sears K, et al. The revised JBI critical appraisal tool for the assessment of risk of bias for randomized controlled trials. JBI Evid Synth 2023 Mar 1;21(3):494-506. [doi: 10.11124/JBIES-22-00430] [Medline: 36727247]
42. Xu HL, Gong TT, Song XJ, et al. Artificial intelligence performance in image-based cancer identification: umbrella review of systematic reviews. J Med Internet Res 2025 Apr 1;27:e53567. [doi: 10.2196/53567] [Medline: 40167239]
43. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Based Healthc 2015 Sep;13(3):132-140. [doi: 10.1097/XEB.0000000000000055] [Medline: 26360830]
44. Hedman E, Axelsson E, Andersson E, Lekander M, Ljótsson B. Exposure-based cognitive-behavioural therapy via the internet and as bibliotherapy for somatic symptom disorder and illness anxiety disorder: randomised controlled trial. Br J Psychiatry 2016 Nov;209(5):407-413. [doi: 10.1192/bjp.bp.116.181396] [Medline: 27491531]
45. Hedman E, Andersson G, Andersson E, et al. Internet-based cognitive-behavioural therapy for severe health anxiety: randomised controlled trial. Br J Psychiatry 2011 Mar;198(3):230-236. [doi: 10.1192/bjp.bp.110.086843] [Medline: 21357882]
46. Newby JM, Smith J, Uppal S, Mason E, Mahoney AEJ, Andrews G. Internet-based cognitive behavioral therapy versus psychoeducation control for illness anxiety disorder and somatic symptom disorder: a randomized controlled trial. J Consult Clin Psychol 2018 Jan;86(1):89-98. [doi: 10.1037/ccp0000248] [Medline: 29172593]
47. Axelsson E, Andersson E, Ljótsson B, Björkander D, Hedman-Lagerlöf M, Hedman-Lagerlöf E. Effect of internet vs face-to-face cognitive behavior therapy for health anxiety: a randomized noninferiority clinical trial. JAMA Psychiatry 2020 Sep 1;77(9):915-924. [doi: 10.1001/jamapsychiatry.2020.0940] [Medline: 32401286]
48. Newby JM, McElroy E. The impact of internet-delivered cognitive behavioural therapy for health anxiety on cyberchondria. J Anxiety Disord 2020 Jan;69:102150. [doi: 10.1016/j.janxdis.2019.102150] [Medline: 31739276]
49. Hoffmann D, Rask CU, Hedman-Lagerlöf E, Jensen JS, Frostholm L. Efficacy of internet-delivered acceptance and commitment therapy for severe health anxiety: results from a randomized, controlled trial. Psychol Med 2021 Nov;51(15):2685-2695. [doi: 10.1017/S0033291720001312] [Medline: 32404226]
50. Du X, Lai L, Shi C, et al. Internet-based cognitive bias modification of interpretation in health anxiety: a randomized controlled trial. Acta Psychologica Sinica 2024;56(10):1351. [doi: 10.3724/SP.J.1041.2024.01351]
51. O'Bryan EM, McLeish AC, Norr AM, et al. A randomized controlled trial evaluating the efficacy of a brief computerized anxiety sensitivity reduction intervention for health anxiety. J Anxiety Disord 2021 Aug;82:102425. [doi: 10.1016/j.janxdis.2021.102425] [Medline: 34082279]
52. Hedman E, Axelsson E, Görling A, et al. Internet-delivered exposure-based cognitive-behavioural therapy and behavioural stress management for severe health anxiety: randomised controlled trial. Br J Psychiatry 2014 Oct;205(4):307-314. [doi: 10.1192/bjp.bp.113.140913] [Medline: 25104835]
53. Ingeman K, Frydendal DH, Frostholm L, et al. Internet-delivered psychological treatment for parents with health anxiety by proxy: replicated randomized single-case experimental design. JMIR Form Res 2025 Oct 2;9:e65396. [doi: 10.2196/65396] [Medline: 41037769]
54. Shor J, Miyatani Y, Arita E, et al. Reducing health anxiety in patients with inflammatory bowel disease using video testimonials: pilot assessment of a video intervention. JMIR Form Res 2023 Feb 9;7:e39945. [doi: 10.2196/39945] [Medline: 36757761]

XSL•FO
RenderX

55. Sharrock MJ, Mahoney AEJ, Haskelberg H, Millard M, Newby JM. The uptake and outcomes of Internet-based cognitive behavioural therapy for health anxiety symptoms during the COVID-19 pandemic. J Anxiety Disord 2021 Dec;84:102494. [doi: 10.1016/j.janxdis.2021.102494] [Medline: 34763219]
56. Goonesekera Y, Donkin L. A cognitive behavioral therapy chatbot (Otis) for health anxiety management: mixed methods pilot study. JMIR Form Res 2022 Oct 20;6(10):e37877. [doi: 10.2196/37877] [Medline: 36150049]
57. Hoffmann D, Rask CU, Hedman-Lagerlöf E, Ljótsson B, Frostholm L. Development and feasibility testing of internet-delivered acceptance and commitment therapy for severe health anxiety: pilot study. JMIR Ment Health 2018 Apr 6;5(2):e28. [doi: 10.2196/mental.9198] [Medline: 29625957]
58. Frydendal DH, Ingeman K, Fink PK, Wellnitz KB, Frostholm L. Internet-delivered acceptance and commitment therapy for COVID-19 related health anxiety: a pragmatic study. Nord J Psychiatry 2025 Aug;79(6):401-407. [doi: 10.1080/08039488.2025.2506560] [Medline: 40491137]
59. Österman S, Axelsson E, Forsell E, et al. Effectiveness and prediction of treatment adherence to guided internet-based cognitive behavioral therapy for health anxiety: a cohort study in routine psychiatric care. Internet Interv 2024 Dec;38:100780. [doi: 10.1016/j.invent.2024.100780] [Medline: 39498477]
60. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
61. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928. [doi: 10.1136/bmj.d5928] [Medline: 22008217]
62. Barsky AJ, Silbersweig DA. The amplification of symptoms in the medically ill. J Gen Intern Med 2023 Jan;38(1):195-202. [doi: 10.1007/s11606-022-07699-8] [Medline: 35829874]
63. Kikas K, Werner-Seidler A, Corkish B, Upton E, Holden M, Newby JM. Illness anxiety disorder: a qualitative study of people with health anxiety and their experiences seeking and avoiding medical care. Br J Clin Psychol 2026 Mar;65(1):68-85. [doi: 10.1111/bjc.70005] [Medline: 40629850]
64. Weck F, Neng JMB, Schwind J, Höfling V. Exposure therapy changes dysfunctional evaluations of somatic symptoms in patients with hypochondriasis (health anxiety). A randomized controlled trial. J Anxiety Disord 2015 Aug;34:1-7. [doi: 10.1016/j.janxdis.2015.05.008] [Medline: 26093823]
65. Bailey R. Why health anxiety needs NICE clinical guidelines. Lancet Psychiatry 2024 Mar;11(3):164-165. [doi: 10.1016/S2215-0366(23)00404-2] [Medline: 38218199]
66. Hentati A, Forsell E, Ljótsson B, Kaldo V, Lindefors N, Kraepelien M. The effect of user interface on treatment engagement in a self-guided digital problem-solving intervention: a randomized controlled trial. Internet Interv 2021 Dec;26:100448. [doi: 10.1016/j.invent.2021.100448] [Medline: 34471610]
67. Greeven A, van Balkom A, Visser S, et al. Cognitive behavior therapy and paroxetine in the treatment of hypochondriasis: a randomized controlled trial. Am J Psychiatry 2007 Jan;164(1):91-99. [doi: 10.1176/ajp.2007.164.1.91] [Medline: 17202549]
68. Tyrer P, Salkovskis P, Tyrer H, et al. Cognitive-behaviour therapy for health anxiety in medical patients (CHAMP): a randomised controlled trial with outcomes to 5 years. Health Technol Assess 2017 Sep;21(50):1-58. [doi: 10.3310/hta21500] [Medline: 28877841]
69. Marcus DK, Gurley JR, Marchi MM, Bauer C. Cognitive and perceptual variables in hypochondriasis and health anxiety: a systematic review. Clin Psychol Rev 2007 Mar;27(2):127-139. [doi: 10.1016/j.cpr.2006.09.003] [Medline: 17084495]
70. Newman MG, Llera SJ, Erickson TM, Przeworski A, Castonguay LG. Worry and generalized anxiety disorder: a review and theoretical synthesis of evidence on nature, etiology, mechanisms, and treatment. Annu Rev Clin Psychol 2013;9:275-297. [doi: 10.1146/annurev-clinpsy-050212-185544] [Medline: 23537486]
71. Cooper K, Gregory JD, Walker I, Lambe S, Salkovskis PM. Cognitive behaviour therapy for health anxiety: a systematic review and meta-analysis. Behav Cogn Psychother 2017 Mar;45(2):110-123. [doi: 10.1017/S1352465816000527] [Medline: 28229805]
72. Bräscher AK, Brähler E, Häuser W, Witthöft M. Further evidence for a dimensional latent structure of health anxiety: taxometric analyses of the whiteley index based on two German representative samples. Assessment 2024 Sep;31(6):1324-1331. [doi: 10.1177/10731911231219802] [Medline: 38159038]
73. Weck F, Richtberg S, Neng JMB. Epidemiology of hypochondriasis and health anxiety: comparison of different diagnostic criteria. Curr Psychiatry Rev 2014;10(1):14-23. [doi: 10.2174/1573400509666131119004444]
74. Olatunji BO, Kauffman BY, Meltzer S, Davis ML, Smits JAJ, Powers MB. Cognitive-behavioral therapy for hypochondriasis/health anxiety: a meta-analysis of treatment outcome and moderators. Behav Res Ther 2014 Jul;58:65-74. [doi: 10.1016/j.brat.2014.05.002] [Medline: 24954212]
75. Tyrer P, Wang D, Crawford M, et al. Sustained benefit of cognitive behaviour therapy for health anxiety in medical patients (CHAMP) over 8 years: a randomised-controlled trial. Psychol Med 2021 Jul;51(10):1714-1722. [doi: 10.1017/S003329172000046X] [Medline: 32174296]
76. Buchholz JL, Abramowitz JS. The therapeutic alliance in exposure therapy for anxiety-related disorders: a critical review. J Anxiety Disord 2020 Mar;70:102194. [doi: 10.1016/j.janxdis.2020.102194] [Medline: 32007734]
77. Tyrer P, Cooper S, Crawford M, et al. Prevalence of health anxiety problems in medical clinics. J Psychosom Res 2011 Dec;71(6):392-394. [doi: 10.1016/j.jpsychores.2011.07.004] [Medline: 22118381]

XSL•FO
RenderX

78. Hall CL, Sanderson C, Brown BJ, et al. Opportunities and challenges of delivering digital clinical trials: lessons learned from a randomised controlled trial of an online behavioural intervention for children and young people. Trials 2020 Dec 9;21(1):1011. [doi: 10.1186/s13063-020-04902-1] [Medline: 33298127]
79. Juul S, Gluud C, Simonsen S, Frandsen FW, Kirsch I, Jakobsen JC. Blinding in randomised clinical trials of psychological interventions: a retrospective study of published trial reports. BMJ Evid Based Med 2021 Jun;26(3):109. [doi: 10.1136/bmjebm-2020-111407] [Medline: 32998993]
80. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120. [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
81. Boutron I, Guittet L, Estellat C, Moher D, Hróbjartsson A, Ravaud P. Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. PLoS Med 2007 Feb;4(2):e61. [doi: 10.1371/journal.pmed.0040061] [Medline: 17311468]
82. Koutsouleris N, Hauser TU, Skvortsova V, De Choudhury M. From promise to practice: towards the realisation of AI-informed mental health care. Lancet Digit Health 2022 Nov;4(11):e829-e840. [doi: 10.1016/S2589-7500(22)00153-4] [Medline: 36229346]
83. Haber Y, Hadar Shoval D, Levkovich I, et al. The externalization of internal experiences in psychotherapy through generative artificial intelligence: a theoretical, clinical, and ethical analysis. Front Digit Health 2025;7:1512273. [doi: 10.3389/fdgth.2025.1512273] [Medline: 39968063]
84. Li H, Zhang R, Lee YC, Kraut RE, Mohr DC. Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. NPJ Digit Med 2023 Dec 19;6(1):236. [doi: 10.1038/s41746-023-00979-5] [Medline: 38114588]

Abbreviations

ACT: acceptance and commitment therapy
AS: anxiety sensitivity
ASI: Anxiety Sensitivity Index
BA: behavioral activation
BAI: Beck Anxiety Inventory
CBT: cognitive behavioral therapy
CENTRAL: Cochrane Central Register of Controlled Trials
CS: cohort study
DBI: digital behavioral intervention
DMHI: digital mental health intervention
***DSM-5*:** *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*
DT: digital therapeutic
GRADE: Grading of Recommendations Assessment, Development, and Evaluation
HAI-64: 64-item Health Anxiety Inventory
iACT: internet-based acceptance and commitment therapy
IAD: illness anxiety disorder
IAS: Illness Attitude Scale
ICBT: internet-based cognitive behavioral therapy
JBI: Joanna Briggs Institute
MADRS-S: Montgomery-Åsberg Depression Rating Scale–Self
MD: mean difference
PICOS: Population, Intervention, Comparator, Outcomes, and Study design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
QES: quasi-experimental study
RCT: randomized controlled trial.
RoB 2: Risk of Bias 2
SDS: Sheehan Disability Scale
SHAI-18: 18-item Short Health Anxiety Inventory
SMD: standardized mean difference
SSD: somatic symptom disorder
WHODAS-II: World Health Organization Disability Assessment Schedule II

XSL•FO
RenderX

Edited by J Zhou; submitted 09.Jan.2026; peer-reviewed by C Cadorin, F Hong; revised version received 16.Sep.2026; accepted 16.Sep.2026; published 18.Nov.2026.

Please cite as:
Xu J, Shen J, Huang G, Wang J
New Insights and Effectiveness of Digital Interventions for Health Anxiety: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2026;14:e91108
URL: https://mhealth.jmir.org/2026/1/e91108
doi:10.2196/91108

© Jing Xu, Jifeng Shen, Guanghui Huang, Jue Wang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 18.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Gender-Affirming Voice Training Delivered by a Novel Mobile App: Mixed Methods Study

Isaac L Alter[1], AB; Christopher P Kruglik[2], MPH; Sara W Albert[1,3], BA; Rachel Coleman[1], MS, CCC-SLP; Mary Pitti[4], MS, SLP; Alexandra Li[1], BS; Katerina Andreadis[5], MS; David Bayne[4], PhD, CCC-SLP; Eseosa Odigie[6], MD; Iyra Chandra[7], BA; Valentina Mocchetti[1], MS, CCC-SLP; David Liao[1,8], MD; Keith A Chadwick[9], MD, MS; Anaïs Rameau[1], MD, MS, MPhil

[1]Weill Cornell Medical College, Department of Otolaryngology - Head and Neck Surgery, Sean Parker Institute for the Voice, 240 E 59th St, New York, NY, United States

[2]Robert Larner, M.D. College of Medicine at the University of Vermont, Burlington, VT, United States

[3]Washington University School of Medicine, St. Louis, MO, United States

[4]Department of Speech-Language Pathology and Audiology, Ithaca College, Ithaca, NY, United States

[5]Department of Population Health Sciences, New York University Grossman School of Medicine, New York, NY, United States

[6]Department of Otolaryngology-Head and Neck Surgery, Duke University School of Medicine, Durham, NC, United States

[7]Wake Forest University School of Medicine, Winston-Salem, NC, United States

[8]Department of Otolaryngology-Head and Neck Surgery, University of Pennsylvania Perelman School of Medicine, Philadephia, PA, United States

[9]Department of Surgery, Division of Otolaryngology-Head & Neck Surgery, Stony Brook University, Stony Brook, NY, United States

Corresponding Author:
Anaïs Rameau, MD, MS, MPhil
Weill Cornell Medical College, Department of Otolaryngology - Head and Neck Surgery, Sean Parker Institute for the Voice, 240 E 59th St, New York, NY, United States

Abstract

Background: Gender-affirming voice training (GAVT) has reduced gender incongruence for transgender and gender-nonconforming (TGNC) individuals, but remains limited by cost and scarcity of specialized providers. Mobile health (mHealth) is a bridge to these access gaps; yet, few high-quality, clinically informed mHealth apps exist within voice clinical sciences.

Objective: We previously created a free mobile app (Attuned; Weill Cornell Medicine and Ithaca College) to deliver clinically informed GAVT through structured instructional videos narrated by a speech-language pathologist and a laryngologist, practice exercises, and pitch-tracking tools, developed with a community advisory board of TGNC individuals. This mixed methods study aimed to assess the app's usability and users' impressions.

Methods: Adult TGNC individuals were recruited nationwide via word of mouth and advertising through lesbian, gay, bisexual, transgender, queer, or questioning (LGBTQ+) community organizations. Participants completed a baseline semistructured interview and demographics survey, used Attuned for one week, and completed a second interview regarding their impressions. Interviews were transcribed verbatim and coded inductively in MaxQDA (VERBI Software GmbH). Three authors independently generated codebooks that were reconciled via consensus, then 2 authors applied the finalized codebooks to the remaining transcripts (preconsensus agreement 95.7%; Cohen κ=0.86). Attuned was also evaluated by 2 reviewers experienced in mHealth evaluation and not involved in the app development, using the Mobile App Rating Scale (MARS), the Intercontinental Medical Statistics (IMS) Institute for Health Care Informatics Functionality Scores, the Centers for Disease Control (CDC) Modified Clear Communication Index, the Institute of Medicine (IOM) strategies for creating health-literate mHealth apps, the Patient Education Materials Assessment Tool (PEMAT), and the System Usability Scale (SUS). Readability of the didactic scripts was also assessed with 6 validated formulas.

Results: Eighteen TGNC individuals completed the study (median age 27.5 years, age range 19 - 50 years). Ten identified as women and/or transgender women, 6 as nonbinary or gender-nonconforming, and 2 as men or transgender men. Eight themes were identified. Participants identified voice as a crucial determinant of others' perceptions of their gender, with voice-related dysphoria common. Cost and difficulty finding providers were primary barriers to GAVT, and existing free resources were viewed as unstructured and untrustworthy. Postuse impressions were largely positive, particularly regarding the structured curriculum, pitch tracker, convenience, and free access, though lack of individualized feedback was a key limitation. Mean scores were 4.21/5 (SD 0.53) for MARS; IMS functionality, 6.5/11 (SD 0.7); CDC index, 82.5% (SD 24.7%); IOM strategies, 28.5/33 (SD 2.1); and

XSL•FO
RenderX

PEMAT understandability and actionability 100%, and SUS 92.5/100 (SD 10.6), with most readability formulas at or below an eighth-grade level.

Conclusions: Voice is central to gender identity for TGNC individuals; yet, structured and affordable GAVT remains largely inaccessible. Attuned was positively received by this initial cohort and scored highly on usability, health literacy, and readability, though findings are limited by potential acquiescence bias and the raters' authorship roles.

(JMIR Mhealth Uhealth 2026;14:e96102) doi:10.2196/96102

KEYWORDS

transgender persons; gender-affirming; voice training; mobile apps; qualitative research; mixed methods

Introduction

Transgender and gender-nonconforming (TGNC) individuals are people whose gender identity is distinct from the sex they were assigned at birth [1]. Prior studies have reported that as many as 96% of TGNC individuals report experiencing voice-gender incongruence, defined as a lack of alignment between an individual's gender and the perceived gender of their voice [2]. Voice has been shown to be intricately linked to TGNC individuals' self-perception of gender identity, with subsequent profound effects on quality of life [3,4]. Voice modification is thus a part of many individuals' transition; the standard of care for these patients, potentially alongside hormonal and/or surgical treatments, is gender-affirming voice training (GAVT). GAVT has a growing evidence base supporting its efficacy, both for objective voice parameters and for patient-rated satisfaction [5-10]. While many individuals focus on pitch modification for gender affirmation, many other factors, including resonance, word choice, syntax, and intonation, influence perception of vocal femininity or masculinity, solidifying the crucial importance of GAVT for optimal gender-affirming voice modification [11-13]. However, there is a relative scarcity of qualitative research on GAVT. This is a significant shortcoming given the known difficulties of quantitative measurements in this area, including poor correlation between various patient-reported outcome measures and objective acoustic measurements [5,14].

Although 70% of TGNC individuals desire gender-affirming voice care, they face barriers to health care at much higher rates than the general population [15-18]. There is thus a significant unmet need for GAVT among TGNC patients, fueled by many factors such as relative scarcity of qualified and affirming providers, difficulty obtaining insurance coverage or acquiring necessary funds to pay out-of-pocket, and the fears of many TGNC individuals in engaging with the health care system [16,19,20]. This gap also has implications for research into the demand and desires for GAVT, since both quantitative and qualitative studies are primarily limited to TGNC individuals who are already presenting for GAVT, thus missing the broad swath of individuals who are not able to access it.

Mobile health (mHealth) represents a promising avenue to reduce such gaps in access for this important care, but uptake of mHealth in voice-focused health care by laryngologists and speech-language pathologists (SLPs) remains limited [21]. A recent publication by our group identified fewer than 30 clinically relevant apps within laryngology, many of which scored poorly on evaluations of usability, health literacy, and readability [21]. The relative paucity of mHealth in our field is particularly striking given that conditions under the purview of laryngology—such as chronic cough, gastroesophageal reflux disease, dysphagia, and dysphonia—are extremely common. As the capabilities of digital health and mHealth tools grow, there remains a critical unmet need for high-quality, evidence-based, and accessible mHealth apps within laryngology.

Our group has previously described the creation of a free mobile app providing clinically informed GAVT, "Attuned" (Weill Cornell Medicine and Ithaca College) alongside a community advisory board (CAB) of TGNC individuals, with the goal of closing the gap in delivery of crucial gender-affirming health care [22]. Attuned consists of multiple modules of instructional videos covering voice physiology and laryngeal anatomy, pitch, intonation, resonance, volume, articulation, and syntax. Rather than offering separate feminization and masculinization tracks, each module presents the relevant vocal parameter as a spectrum and teaches techniques for shifting the voice in either a more feminine or more masculine direction, allowing users, including nonbinary users whose goals may not align with a binary framework, to apply the same curriculum toward their individual voice goals. The curriculum is narrated by an SLP and a laryngologist who deliver gender-affirming care in their clinical practice; this content is based on the standard of care in GAVT and was iterated with feedback from the CAB to provide maximum understandability and utility to the target user base. The app also includes other features, such as assigning "homework" assignments and practice exercises, the ability to record voice samples and track vocal pitch over time, and a piano keyboard to provide reference pitches for pitch modification.

In this study, we undertook a mixed method approach to evaluate Attuned for the first time: (1) a qualitative study investigating TGNC individuals' preexisting perceptions of their voice and voice modification as well as their reactions and feedback following the use of Attuned, and (2) a battery of validated scales to assess the app's usability, functionality, and readability.

Methods

Overview

We used a convergent mixed methods design [23], in which qualitative data (semistructured interviews) and quantitative data (validated mHealth assessment scales and readability measures) were collected in parallel and analyzed independently and integrated at the interpretation level through narrative

XSL•FO
RenderX

weaving by comparing qualitative themes with quantitative scale domains.

Qualitative Study

To maximize diverse subjective feedback, a qualitative study design was used to capture TGNC individuals' baseline attitudes about voice and gender and their reactions to the Attuned app. The qualitative component of this study is reported in accordance with the COREQ (Consolidated Criteria for Reporting Qualitative Research) [24]; the completed checklist is provided in Checklist 1. Adult (≥18 years) individuals identifying as transgender, nonbinary, or gender-nonconforming were recruited via word of mouth and advertising through lesbian, gay, bisexual, transgender, queer, or questioning (LGBTQ+) community organizations nationwide. Participants were required to have access to an iPhone, since the first version of Attuned was available only for iOS at the time of our study. Recruitment was particularly focused on representing a wide array of gender identities, and sample size was determined by thematic saturation [25,26]; in total, 18 individuals participated.

Following informed consent, participants engaged in a one-on-one semistructured interview with one of 2 members of the research team (ILA or CPK) and completed a demographic survey. They were then given access to the Attuned app for one week and instructed to experience as much didactic material as possible. Following one week, participants were re-engaged for an additional semistructured interview, focused on their perspectives of the Attuned app and its delivery of GAVT. Screenshots of the app are shown in Figure 1. Interview questions were devised by the study team to collect information regarding (1) baseline attitudes about the voice and its relationship with gender, (2) prior experience with and impressions of voice training or self-driven voice modification, (3) impressions of the Attuned app in particular, and (4) reflections on the value and shortcomings of GAVT delivered by mHealth. Interviews were conducted via Zoom (Zoom Communications, Inc), and recordings were transcribed verbatim using a combination of Zoom's built-in transcription (using a Health Insurance Portability and Accountability Act [HIPAA]–compliant Zoom platform) and the Whisper transcription app by OpenAI (which was run locally and through terminal, thus also achieving HIPAA compliance) [27]. Transcriptions were reviewed for accuracy by the individual who conducted the interview. All interviews took place between February and April 2025.

Interview transcriptions were coded using MaxQDA software (VERBI GmbH). Coding followed an inductive thematic analysis as described by Braun and Clarke [28]. Three authors (ILA, CPK, and RC) independently reviewed and coded the transcripts of the first 3 participants, both pre- and post-use interviews (ie, 6 transcripts in total). The 3 codebooks were then compared and reconciled through consensus discussion among all 3 coders to generate a finalized codebook. Prior to consensus discussion, inter-rater reliability of coding was calculated within MaxQDA, and the percent agreement was 95.7%, with a Cohen κ of 0.86, corresponding to near-perfect agreement. Two authors (ILA and CPK) then independently applied the finalized codebook to all remaining transcripts, with coding differences again reconciled via consensus discussion. No code required adjudication by a third coder. New codes arising during coding of the remaining transcripts were added to the codebook by consensus, and previously coded transcripts were reviewed again with the updated codebook. Thematic saturation was assessed following the approach of Guest et al [25], defined a priori as the point at which successive interviews yielded no new codes or themes. Codes were grouped into candidate themes and consolidated into 28 subthemes, which were reviewed and refined by the coding team into the 8 final themes reported below.

XSL•FO
RenderX

Figure 1. Logo and screenshots of Attuned app interface.



Objective Assessment of Attuned

Two authors (IC and EO), with previous experience evaluating mHealth apps via validated scales and who were not involved in the development of Attuned, completed ratings via Google Forms. In line with prior systematic assessments of mHealth apps by members of our authorship team [21], the following scales were used: (1) the Mobile App Rating Scale (MARS) [29], (2) the Intercontinental Medical Statistics (IMS) Institute for Health Care Informatics Functionality Scores [30], (3) the Centers for Disease Control (CDC) Modified Clear Communication Index [31], (4) the Institute of Medicine (IOM) strategies for creating health-literate mHealth apps [32], (5) the Patient Education Materials Assessment Tool (PEMAT) [33], and (6) the System Usability Scale (SUS) [34]. No single validated scale captures quality, functionality, health literacy, and usability, so we applied the same multiscale battery used in our group's prior systematic assessment of laryngology apps [21] to permit direct comparison with other apps. Scores were averaged between the 2 raters. For each of these scales with explicitly numerical output evaluated by 2 raters, interrater reliability was evaluated via calculation of an intraclass correlation coefficient (CA-ICC). Readability of the script used for didactic videos was also assessed using 5 validated readability formulas via Readable.io (Added Bytes Ltd) [35]: (1) Flesch-Kincaid Reading Level and Flesch-Kincaid Reading Ease [36], (2) the Gunning Fog Index [37], (3) Coleman-Liau Index [38], (4) Simple Measure of Gobbledygook (SMOG) Index [39], and (5) New Dale-Chall [40]. As previously described, it is standard to use multiple readability scales, as there is no established gold standard [41]. Each readability formula is calculated differently, focusing on various attributes of a given text selection such as word and sentence length, number of unfamiliar words, number of syllables, and sentence structure complexity. Briefly, the Flesch-Kincaid Reading Ease score ranges from 0 to 100 and is inversely related to readability, while the remaining scales describe a grade level of education necessary to understand the text [41].

Positionality Statement

The research team would like to formally acknowledge that our individual and collective identities may influence our research. We have opted to share information about the identities of the primary members of the research team who designed and conducted the majority of the present study.

ILA and CPK are both medical students and cisgender members of the LGBTQ+ community and are both advocates for gender-affirming care and the TGNC community. AR is a fellowship-trained laryngologist at the Sean Parker Institute for the Voice at the Weill Cornell Medical College, who completed her training at Stanford University and then at the University of California, Davis. She is a cisgender member of the LGBTQ+

XSL•FO
RenderX

community and an ally of the TGNC community. KAC is a fellowship-trained laryngologist at Stony Brook Medicine specializing in LGBTQ+ care, who trained at Oregon Health & Science University and then at the Sean Parker Institute for the Voice. He is a cisgender member of the LGBTQ+ community and an ally of TGNC people. MP is an SLP who trained at Ithaca College and later returned to direct the speech-language pathology clinic at the same institution. She is a cisgender woman and an LGBTQ+ ally. RC is a voice-specialized SLP, focusing on care of the professional voice and gender-affirming voice care (among other clinical specializations) at the Sean Parker Institute for the Voice, which trained at the Massachusetts General Hospital's Institute of Health Professions. She is a cisgender woman and an ally to the LGBTQ+ community. All 6 team members underwent cultural humility training prior to beginning the development process.

Ethical Considerations

This study received approval from the Weill Cornell Medical College Institutional Review Board (#22 - 09025226). Informed consent was waived for this study, as the research presented no more than minimal risk to the participants and involved no procedures for which written consent is normally required outside of the context of research. Participants were provided with a study information sheet prior to joining the study. Participants were assigned a participant ID in REDCap (Vanderbilt University), and all survey data were captured in REDCap. Interviews were conducted over Zoom, and once the interviews were transcribed and all personal identifiers were redacted from the transcript using Whisper by OpenAI, which was run offline, the audio recordings of the interviews were destroyed upon completion of the study. Audio recordings and deidentified transcripts were stored locally on a password-protected tagged device and labeled with the participant ID. Participants received US $30 for their participation in the study and were compensated by ClinCard (Greenphire LLC).

Results

Qualitative Study

Participant Characteristics

A total of 18 interviews were conducted from February to April 2025. Of the 18 participants, 10 identified as women and/or transgender women, 6 identified as nonbinary, gender-nonconforming, or gender fluid, and 2 identified as men or transgender men. Ten used the pronouns she/her, 2 used the pronouns he/him, 4 used the pronouns they/them, one used the pronouns she/they, and one used the pronouns he/she/they. Ten individuals were assigned female at birth (AFAB), 7 were assigned male at birth (AMAB), and one was intersex.

Twelve participants had a history of hormone replacement therapy. Of the 12 participants, 3 had received testosterone treatment and 9 had received one or more of the following treatments: estrogen, progesterone, and spironolactone. Twelve participants also self-reported as being under the neurodivergent umbrella.

The median participant age was 27.5 years, with ages at the time of interview ranging from 19 to 50 years. A summary of these and additional participant characteristics, including geographic regions and settings (urban, suburban, or rural) as defined by the US Census Bureau, is shown in Multimedia Appendix 1.

Overall Qualitative Results and Themes

Prior to consensus discussion, interrater reliability was calculated within MaxQDA (VERBI Software GmbH) software; percent agreement was found to be high at 95.7%, with a Cohen κ of 0.86, corresponding to near-perfect agreement. Eight overarching themes were identified (5 from pre-Attuned interviews and 3 from post-Attuned interviews) and will be described individually below: (1) the voice is essential to gender identity, (2) dissatisfaction with the voice is the norm, (3) high-quality voice training is largely inaccessible, (4) navigating and modifying the voice across social contexts, (5) current apps fall short of user needs, (6) Attuned provides a clear curriculum with useful features on users' own schedules, (7) Attuned can positively impact the voice, and (8) Attuned is limited by lack of immediate feedback and personalization.

The Voice Is Essential to Gender Identity

Exemplar quotes pertaining to the importance of voice to gender identity are shown in Table 1. Many participants emphasized that their voice was a key part of their own self-conception of their gender and gender presentation as well as their self-confidence (Q1-Q3), while also stressing its "paramount" effects on others' perceptions of their gender (Q4-Q7). The voice was identified as a "tell" or a "giveaway" when they felt they could otherwise "pass" or be identified as a cisgender person of the gender to which they had transitioned [42]. Many participants felt that their voice was a reason for being misgendered or assigned a gender different from the one with which they identified (Q8-Q11) [43]. Importantly, several participants noted that being either misgendered or identified as a transgender person may have implications for their safety.

XSL•FO
RenderX

Table . Exemplar quotations: the voice is essential to gender identity.

Subtheme	Exemplar quotations
Voice's impact on self-conception of gender	• Q1: "It's how I express myself. All the little sounds I make, everything that I say, it's just a part of who I am. It's important." (P1) • Q2: "I think having the right pitch can really help me in my transition and formation of my gender identity. It is really important in the transition process and becoming blended into society as the person and gender I identify with." (P2) • Q3: "What's most important to me is how I see myself, and just having confidence in my voice, which would make me speak up more and use my voice." (P16)
Voice's impact on others' perceptions of gender	• Q4: "I think my voice is very important because the first time I open my mouth when I'm talking to someone, they can easily assume my gender based on my voice and how I talk." (P2) • Q5: "If you know the old saying: 'If it walks like a duck, looks like a duck, and quacks like a duck, then it's a duck.' Well, I look like a duck, but I don't quack like a duck. So, yeah, it's one of those things where I've seen people turn around expecting me to be a guy because they hear my voice, and they turn around and are confused because I'm not a guy." (P16) • Q6: "I'd say it is paramount to passing in my everyday life." (P21) • Q7: "The fact that it's possible that I could be passing without it and then completely tarnish it the second I speak is something that weighs on me every single day." (P28)
Misgendering and safety	• Q8: "If people can't see me physically like on the phone or at the drive through, they say ma'am. They assume based on the way they hear my voice that I'm a woman." (P5) • Q9: "You know, there are situations where either for professionalism or for safety or just comfort, it is helpful or convenient to have voice modification available." (P11) • Q10: "But when I had shorter hair and more loosely fitting clothing, my gender became a question and you could tell because people start giving you the look and are wondering to themselves 'What's going on? What do I say?' But then, I would open my mouth and they would immediately gender me as a woman." (P12) • Q11: "Yes, I have been misgendered on the telephone. Certainly, in face-to-face situations, where I feel like I am either presenting in a nonbinary, feminine-leaning way. Or, I feel like I'm totally presenting as female, maybe not passing. So it's hard to tell if it's transphobic responses or ignorant, internal bias responses, or just responding to my voice, but I feel like my voice 'clocks' me a lot and gets me misgendered." (P35)

Dissatisfaction With the Voice Is the Norm

Exemplar quotes regarding participants' current dissatisfaction with their voice are given in Table 2. Many individuals cited particular aspects of their voice that they were dissatisfied with, including pitch being too high or too low, or a lack of control over specific sounds, while others felt that their voice was generally "not the way I would like my voice to be" (Q12-Q15). In addition, voice was identified as a source of dysphoria, defined as stress derived from incongruence between one's gender identity and the gender they were assigned at birth [44]. Some participants noted their voice was a "main cause" for their dysphoria (Q16-Q18). Dysphoria related to the voice was often severe enough to make participants "cry about it" or require them to "pause and...reset at times in conversation." Accordingly, there was a substantial desire to modify the voice, ranging from technical changes such as pitch modification to more abstract feelings of wanting to "pass" or to align their voice with "how I feel inside" (Q19-Q23). Several participants also shared their experiences with voice modification via gender-affirming hormones, particularly testosterone for voice masculinization. Some individuals feared the potentially irreversible effects of hormones, while others who had used testosterone noted that despite improvements in vocal pitch, the hormone did not allow them to achieve a voice truly in line with cisgender men, but rather often outed them as transgender (Q24-Q30).

XSL•FO
RenderX

Table . Exemplar quotations: dissatisfaction with the voice is the norm.

Subtheme	Exemplar quotations
Dissatisfaction with voice	• Q12: "I am not satisfied with much because it's not really the way I would like my voice to be." (P2) • Q13: "When I try to just let myself be heard, it tends to just go much deeper than I intended to and I don't realize until I'm mid-sentence." (P13) • Q14: "I'd say that the satisfaction with my voice has gone down over time, as the rest of me has evolved to fit my ideal image that then it just makes the voice even more important, because it's kind of like the last piece." (P16) • Q15: "I'm dissatisfied with the fact that it gives me away and the fact that it is masculine." (P28)
Voice causes dysphoria	• Q16: "It's the main cause of my dysphoria." (P21) • Q17: "It depends on the day. It depends on how I'm feeling. I'd say my voice definitely causes moderate dysphoria. But there are some times where it's mild and I forget about it. There are sometimes where it's severe and I cry about it." (P28) • Q18: "Recently I've been experiencing quite a bit of dysphoria at times surrounding my voice, to the point where I sort of pause, and have to close my eyes and reset at times in conversation." (P35)
Desire for voice modification	• Q19: "Just being able to talk to somebody in the bathroom would be really great without having anyone feel uncomfortable." (P1) • Q20: "I want my voice to be deeper." (P6) • Q21: "I would like to present in the world based on how I feel inside." (P10) • Q22: "I guess my goal would be probably to 100% pass as a female." (P16) • Q23: "I would just like to be able to speak in a softer, more breezy, feminine way, and do so naturally. It's a lot of effort at the moment to even try to think about it, and I feel silly at times, even practicing." (P35)
Impact or perceptions of gender-affirming hormones	• Q24: "It feels almost like a catch 22, like, yes, I want my voice to be deeper and all of these things. I want to do this masculinizing therapy. But people's perception of my voice being different and my vocal cords changing with the testosterone too, causing my voice to be nasally, would then actually make me feel probably more dysphoric than whatever voice my voice had been before." (P6) • Q25: "My voice was changing, and people could tell that it was changing. So their perception of my voice was like, oh, that person is trans or queer. So that was suddenly a reason that I had to start coming out, because my voice was obviously changing and things about my body were changing." (P6) • Q26: "HRT is really interesting and but not something that feels right for me at this time and it feels too permanent." (P17) • Q27: "I definitely still have the voice crack, the pubescent boy voice crack, which is not what I want to sound like." (P19) • Q28: "One of the things I've noticed in the past year on HRT is not only does my voice sound different, but I get like vocal cord exhaustion in a way that I never did before. So I talk a lot and I get more tired from it, which is a new experience." (P19) • Q29: "There's a particular sound to a trans guy kind of voice, and I feel like people hear that particular kind of quality and they notice that it's different enough and funny-sounding enough that it triggers a pause and curiosity response. And I think that is a significant part." (P19) • Q30: "But overall, in terms of just pitch in general, just talking, it does feel better." (P19)

High-Quality Voice Training Is Largely Inaccessible

Exemplar quotes pertaining to experiences and perceptions of formal or informal voice training are given in Table 3. A few participants recounted mixed impressions of formal voice training with a one-on-one coach or SLP; while such endeavors were often productive, they often failed to meet initial expectations, potentially yielding frustration after spending "so

XSL•FO
RenderX

much money on therapy" and still being "misgendered over and over and over again" (Q45-Q46). Some participants also highlighted the utility of nongender-focused voice training for better voice control, which could be leveraged in service of gender-oriented voice modification (Q47). However, the majority of participants had not had prior experience with formal voice training of any kind, citing several key barriers to accessing it, including cost ("prohibitively expensive"), lack of specialized providers, lack of convenience or difficulty with time commitment, and difficulty navigating the landscape of resources to find a voice coach (Q48-Q52). Participants also shared various impressions of free, publicly available resources such as YouTube (Google LLC) videos or posts on social media forums, including Reddit (Reddit, Inc) and TikTok (ByteDance). While such resources provided some benefit, particularly when created by fellow members of the transgender community, navigating the unstructured information from countless sources was often frustrating, and the information shared was sometimes either not what participants were looking for or perceived as unreliable (Q53-Q57). Thus, many individuals still felt a strong desire for formal voice training in order to have a professional guide them through the steps and different aspects of voice modification (Q58-Q61).

XSL•FO
RenderX

Table . Exemplar quotations: high-quality voice training is largely inaccessible.

Subtheme	Exemplar quotations
Prior experience with voice training	• Q45: "I had a couple sessions for vocal training but that was really hard to sit in front of somebody on Zoom and try to change your voice." (P1) • Q46: "I had a moment a few months ago of feeling like I spent so much money on therapy and I spent so much time and effort on this and I was just getting misgendered over and over and over again. I had a period of time where it just kept happening and it was getting really frustrating. Sometimes it bothers me. Sometimes it doesn't. And I've had panic attacks over it. But it's not a consistent thing." (P11) • Q47: "I am lucky in that I have vocal training as an actor and so I feel like I have more control of my voice than I may have otherwise." (P17)
Difficulty accessing formal voice training	• Q48: "I am working 2 jobs so I cannot make a weekly hour session with either a speech therapist or vocal trainer." (P12) • Q49: "I have considered doing vocal training. But then, you know, it's obviously not covered by insurance and that's something that I'm paying for out of pocket. And because of that, it does not become a priority." (P12) • Q50: "Then, you have to find one for a really specific thing. You might even get someone who is not the right fit. Being nonbinary, it's a little more fuzzy because I am not looking for someone to tell me, 'Here's how you speak like a man' because I am not into that." (P12) • Q51: "I'm aware that there are like people who do voice coaching specifically for people in my situation. I've looked into it, but it's prohibitively expensive." (P19) • Q52: "I'm not even sure where to find one. Like vocal coach dot com? I think I just don't know where or who or how to acquire one." (P31)
Experience with publicly available resources	• Q53: "I never had consistency with those people. Like I see a TikTok video and they say some things and I can try them on my own, but I don't have anything consistent." (P6) • Q54: "I would like to have resources to have a lower pitch feminine voice, but all the resources I have access to are attempting to teach you to do use a hyper-femme voice before saying that you can maybe drop it down later. But there's no explicit path on how to do that." (P21) • Q55: "The voice resources I've tried so far, I would give them probably like a 5 or 6 out of 10. They really haven't yielded much success, but they haven't hurt either." (P21) • Q56: "I have accessed resources in an attempt to start doing something many times. I just it's very hard for me without like any sort of structure or guidance from interacting with another human or anything like that. I need something to set the routine for me." (P28) • Q57: "I have used a lot of YouTube videos. And I have found web page apps that that will help track certain aspects of my voice, and let me know what I have to work on. Although it's always hard to know what to trust on those apps, because some of them are not very accurate." (P36)

XSL•FO
RenderX

Subtheme	Exemplar quotations
Desire for voice training	• Q58: "I think if I get good training and I can change my voice, it can help, as long as it is also coupled with my physical appearance and clothing and such things. It can help people identify my gender much faster than before." (P2) • Q59: "It would probably be really effective to have like a one-on-one voice coach who was like talking me through doing things." (P3) • Q60: "I think a voice coach would probably be one of those things that would make a humongous difference. There's the technical aspects of your voice, but there's also ways of explaining things or saying things that are considered more masculine or feminine-like. So, someone that could have an outside perspective on how I'm being read or perceived would be something I would want." (P16) • Q61: "With any sort of training, control is really nice, and when you have control and knowledge, I think it gives you a lot of autonomy." (P17)

Navigating and Modifying the Voice Across Social Contexts

Exemplar quotes regarding self-guided attempts at voice modification are shown in Multimedia Appendix 2. Many participants, largely without any formal instruction, focused primarily on attempts to alter their pitch to better align with their gender (Q31-Q33). Some also identified other aspects of voice that they sought to modify for gender-affirming purposes, such as cadence, resonance, weight, and volume (Q34-Q36). For most participants, such modifications were situational, changing their voice in professional situations or based on the individuals to whom they were speaking, such as using a "girl voice…with my friends…then at work…I use my guy voice" (Q37-Q42). Some also highlighted a fear of vocal injury or a feeling that altering their voice felt "forced" or "uncomfortable" (Q43-Q44), especially in the absence of reliable instruction for voice modification.

Current Apps Fall Short of User Needs

Exemplar quotes for participants' prior experience with voice-focused apps and their priorities for an ideal voice-focused app are shown in Multimedia Appendix 3. Several participants reported using apps largely focused on pitch identification and classification of the voice as male, androgynous, or female, though many had doubts about the accuracy and reliability of such apps (Q62-Q65). Participants were also asked to identify key priorities for their ideal voice-focused app. Common priorities included personalization for an individual's particular goals, a clean user interface, the ability to play back recordings or receive automated feedback, clear information and instructions, and structured teaching with the ability to practice (Q66-Q71).

Attuned Provides Clear Curriculum With Useful Features on Users' Own Schedules

The final 3 sections describe themes from the interviews conducted following the week during which participants used the Attuned app. Exemplar quotes regarding participants' overall satisfaction with Attuned and their perceptions of its particular strengths and benefits can be found in Table 4. All participants expressed overall positive impressions of the app, and several stated they "would definitely recommend" it to others (Q72-Q76). Two areas of particular strength were the clear structure and order with which participants were guided through the curriculum (Q77-Q80)—a significant improvement over attempts to cobble together other "scattered and unorganized" free resources—and the pitch-tracking function, which is coupled with a piano keyboard for pitch identification and pitch targeting (Q81-Q83). Participants also highlighted the convenience of app-based delivery of GAVT, allowing people to learn at their own pace and use the app in a variety of settings (Q84-Q87). Access to this free app also eliminated price as a major "sticking point" for many individuals. Finally, participants appreciated the clear delivery of instruction and emphasized the feeling of trustworthiness that comes from receiving information from a trained speech-language pathologist and a team of creators who "know what they're doing" (Q88-Q91).

XSL•FO
RenderX

Table . Exemplar quotations: Attuned provides clear curriculum with useful features, on users' own schedules.

Subtheme	Exemplar quotations
Overall satisfaction with Attuned	• Q72: "It's much better than any other voice app that I've used." (P5) • Q73: "I'm not aware of any robust voice training services available like this. And I think it's very beneficial. I would definitely recommend for anybody who was trying to have some at home practice, either in between speech therapy sessions or by itself." (P11) • Q74: "I do think getting information about your body is very empowering, and I think that is something it does very well. I think it's information a lot of people don't have and would benefit from." (P19) • Q75: "I think it's a really wonderful starting point terms of educating you on terms and the different qualities of the voice that make it up. It's amazing." (P28) • Q76: "I think that if you don't have access to an in-person voice therapist, I think the app is a lot better than having to pore through forums and YouTube videos and stuff like that." (P32)
Appreciation of clear structure	• Q77: "I thought the general structure of progressively getting more difficult and having a pretty clear line of progression was useful for me." (P3) • Q78: "It's difficult, if you're just trying to do voice training in a DIY context or on your own, because everything is so scattered and unorganized in terms of community resources. So that aspect of it was very helpful of just saying, OK, do this and this and this and this and then do these things together." (P11) • Q79: "And throughout it, I really enjoyed how it built on each aspect. Especially if somebody's coming in without much knowledge of what goes into what makes a voice work and language and all of that, it can be really informative." (P17) • Q80: "I really liked the structure. It's so important in terms of like educating you on the terms and finding the motivation. It's the only reason I had never started before just because of the overwhelming nature with zero guidance. This puts you on a certain path and it's nice." (P28)
Pitch tracking function was particularly helpful	• Q81: "I thought that the little keyboard tool to show what frequency your voice is currently at and where you want to go was really great and helpful for getting a better sense of where I want to be." (P10) • Q82: "I don't tend to think of voice in terms of like 'Oh, these are actual like pitches, and these notes, have names.' So that was very interesting to actually be able to match and see that visually." (P12) • Q83: "I liked the piano. That was really helpful. And I liked the option to record right there. Some of the exercises I wouldn't have been able to do if I couldn't both be playing the note and also trying to make a noise." (P19)
Convenience and accessibility were key benefits	• Q84: "Today I was doing it during my dog walk, or last night I was working on it while I was lying in bed. The other day I did a couple of minutes, when I had nothing to do at work. So I like that you can do little spurts." (P6) • Q85: "I could be listening to things right before I go into a meeting or spaces where I know that people are going to be meeting me for the first time. And I want to have that moment of listening to the articulation lesson or returning to a practice, whenever it is most needed." (P17) • Q86: "I think access is the biggest benefit, because that is such a sticking point for so many people. Like, even if their insurance covers it, the logistics of that are a bear." (P32) • Q87: "For many of us who identify as transgender, it is nice that it is private. I felt safe and secure doing the training in the comfort of my home." (P36)

XSL•FO
RenderX

Subtheme	Exemplar quotations
Information was clearly delivered and felt reliable	• Q88: "It was obviously designed in close partnership with people who very much know what they're doing. So that definitely comes through in a way that makes it both fundamentally useful, but also credible." (P12) • Q89: "I think it does a really good job of entry-level, accessible education around what one can control around their voice." (P17) • Q90: "It is very legitimizing to have it come from an actual speech therapist." (P32) • Q91: "The material felt digestible and easy to understand. I also found that the length of each lesson was appropriate." (P36)

Attuned Can Positively Impact the Voice

Exemplar quotations focused on Attuned's effects on the voice are given in Multimedia Appendix 4. Some participants felt that the app had already begun to affect changes in their voices, both in terms of their own satisfaction with their voice and the way their voice was perceived by others (Q92-Q94). However, the majority of participants had not yet noticed changes in their voice after only a week but remained optimistic that the tools and skills they had learned from the app would impact their voice with time and dedicated practice (Q95-Q98).

Attuned Is Limited by Lack of Immediate Feedback and Personalization

Exemplar quotations pertaining to challenges and drawbacks of using Attuned, or app-based delivery of GAVT in general, are shown in Table 5. A primary concern shared by many participants was the lack of feedback from the app, with several expressing uncertainty whether they were performing the exercises correctly; the loss of face-to-face communication, including potential physical touch from a coach, was also highlighted (Q99-Q101). Relatedly, some participants also felt that the app had a limited ability to personalize the curriculum based on their particular goals and felt that a "personal relationship" with a voice coach was also a meaningful aspect of voice training that was lost via an app (Q102-Q104). Since the app was in its initial beta-testing phase during this study, many individuals also cited specific technical glitches, which sometimes made the app difficult to navigate or operate (Q105-Q107). Finally, some participants requested that more voices, particularly trans voices or voices representing different dialects of English, be included in the instructional content, since only cisgender voices lead to participants "not hearing [themselves]" represented (Q108-Q110).

XSL•FO
RenderX

Table . Exemplar quotations: Attuned is limited by lack of immediate feedback and personalization.

Sub-Theme	Exemplar quotations
Difficult to not receive feedback from another person	• Q99: "People are really key for how one speaks. You can be really loud and resonant on your own, but if you feel nervous in front of other people, that's something that can only come with practice. Working with a coach, or even small groups of other people doing it would help break down that closing in on your throat when you feel nervous. So I think most apps run into the need for human connection at some point." (P17) • Q100: "There's no feedback of like, 'Yes, that is actually the note you're making.' I have no idea. So there's a guessing element to it. That's an element of 'I don't really know if this is right.'" (P19) • Q101: "I think the lack of feedback from someone who's actually hearing you is something that makes a difference. So, if you have the means, I would still recommend going to actual voice training over just using an app." (P32)
Lack of personalization and personal relationship	• Q102: "It's not necessarily tailored to what you would want, instead of like getting that one on one work with like a speech pathologist or someone." (P10) • Q103: "I think that obviously you have a less personal relationship. It's an app versus actually building a client provider relationship. So there is a bit of that disconnect." (P12) • Q104: "I'd say that it's a very useful tool, specifically for people who are beginning their voice training journey, because what it lacks in personalization and makes up for with general ease of use." (P21)
Beta-testing version of app had glitches	• Q105: "The app was glitchy for me. Stuff didn't load consistently." (P3) • Q106: "There were a couple of technical things about the app that I figure were because it's in its beta stage. Sometimes I would have to close the app and reopen it to get things to work." (P6) • Q107: "I will say that they did have some technical difficulties with using the app." (P10)
Diversity of voices in video content would be appreciated	• Q108: "Obviously, I think there is some level of comfort that could be achieved, even an aspirational goal that you could assign by receiving training from somebody who is themselves a trans person who has pursued and completed voice training." (P11) • Q109: "I feel like there could have been more of a diversity of the trainers, which I figure you guys also know with the app very much in testing stages. I know that there were a few things read by the more masculine person, or maybe just one. But having a more diverse set of voices would be great. And it was all in standard American. So having like AAVE[a] or other dialects would be cool and great to see." (P17) • Q110: "I would have really appreciated if there were some trans voices. Cis women's voices, I know what those are supposed to sound like. And there was an element of not hearing myself that I think we're all very used to." (P19)

[a]AAVE: African-American Vernacular English.

Objective Assessment

MARS, SUS, and IMS scales were used to evaluate quality and functionality. The overall mean MARS score averaged between the 2 raters was 4.21/5 (SD 0.53), with the following scores for each subcategory: engagement 4.40/5 (SD 0.57), functionality 4.13/5 (SD 0.18), aesthetics 4.17/5 (SD 0.71), information 4.33/5 (SD 0.47), and subjective quality 4.0/5 (SD 0.71). However, interrater reliability, as measured by CA-ICC, was poor (0.348). The mean SUS score was 92.5/100 (SD 10.6), with an excellent CA-ICC of 0.912. The IMS score was 6.5 (SD 0.7) of the 11 recommended functions. Functions that both reviewers found to be missing were the ability to share and evaluate data, the provision of guidance based on user-entered information, and the provision of communication between health professionals and patients.

Next, Attuned was assessed with regard to health literacy via the CDC, IOM, and PEMAT scales. The CDC Clear Communication Index was 82.5% (SD 24.7 percentage points), slightly below the recommended "passing" score of 90%; missing elements largely pertained to the use of language, which is likely to be used by the audience, and to the use of chunked or bulleted text. The CA-ICC for this measure was also poor (0.347). The mean IOM score was 28.5 (SD 2.1) of 33

XSL•FO
RenderX

recommended strategies implemented within Attuned, with the "engage users" and "evaluate and revise your site" domains scoring lower. Both PEMAT actionability and understandability scores were 100%, with complete agreement between reviewers with a CA-ICC of 1.00.

Readability as determined by the Flesch-Kincaid Reading Level and the Gunning Fog Index scores was 5.6 and 7.9, respectively, both below the recommended eighth-grade reading level. The Flesch Reading Ease was 72.3, corresponding to roughly a seventh-grade reading level. However, the SMOG index was 9.3, signifying that 9.3 years of school education would be necessary to read the text, and the Coleman-Liau index also resulted in an eighth-grade reading level at 8.7. On the other hand, the New Dale-Chall score was 4.5, roughly corresponding to a fourth-grade reading level.

Integration of Qualitative and Quantitative Findings

Qualitative and quantitative findings were integrated through narrative comparison and demonstrated substantial convergence. The high SUS score (92.5/100) converged with participants' qualitative descriptions of the app as clear, well-structured, and easy to navigate (Theme 6). The IMS functionality assessment identified missing functions, such as provision of guidance based on user-entered information and communication between users and professionals, that mirrored the qualitative limitation theme of absent feedback and personalization (Theme 8). Similarly, the CDC Clear Communication Index fell below the passing threshold primarily on use of the audience's own language, converging with participants' requests for trans voices and greater dialect diversity in instructional content. Divergence was minimal: the strong PEMAT and readability results were consistent with participants' descriptions of content as "digestible and easy to understand."

Discussion

Principal Findings

In this mixed methods study, we have used qualitative methodology to (1) describe baseline attitudes toward voice and vocal modification among transgender individuals and (2) evaluate impressions of a free mobile app delivering clinically informed GAVT. We have also used a wide variety of validated mHealth-focused scales to assess the mobile app for functionality, usability, health literacy, and readability. Given the increasing focus on gender-affirming health care in the United States, the importance of both research on gender affirmation and easily accessible health interventions for TGNC individuals cannot be overstated. The findings of this study help to further establish the widespread demand and need for GAVT and to understand diverse user experiences with an initial iteration of an open-access novel GAVT mobile app for planned further improvement.

GAVT alone or in combination with surgery has been repeatedly shown to improve objective measures such as fundamental frequency, listener perceptions, and patient-reported outcome measures [45-47]. However, the vast majority of existing evidence is for voice feminization, since exogenous testosterone lowers vocal pitch among transmasculine individuals taking gender-affirming hormones, though many such individuals still report voice problems or dissatisfaction with their voices [48,49]. In addition, compared to the growing quantitative literature, relatively few qualitative studies have been undertaken to deepen understanding of factors affecting patients that are not adequately reflected in quantitative measures.

Our study uncovered several key themes among our study population at baseline, prior to use of the Attuned app. Participants consistently emphasized the importance of their voices as a key facet of their complete gender identity and as a determinant of their ability to be perceived by others in line with their gender. This is in line with prior qualitative studies of both transfeminine and transmasculine individuals, which have identified both internal self-confidence and experiences of misgendering as intimately linked to the voice [50-52]. An additional qualitative study by Holmberg et al [53] also noted the key contributions of the voice toward dysphoria, similar to our findings; although they did not explicitly explore dissatisfaction with the voice in parallel, they also identified a strong motivation among their participants to seek voice modification. Importantly, our inclusion of transgender men and nonbinary individuals alongside transgender women in our study population allowed for key perspectives regarding both trepidation and relative dissatisfaction with certain aspects of testosterone's effects on the voice, highlighting that testosterone alone is sometimes insufficient for achieving vocal congruence. This solidifies the importance of further study among individuals seeking voice masculinization and further use of GAVT among this population. The phenomenon of an outwardly evident "trans guy kind of voice" that can present barriers to "passing" for transgender men is in line with existing knowledge that gender perception goes beyond simple pitch changes. A small number of previous qualitative studies have also raised similar concerns among transgender men who remain dissatisfied or continue to experience misgendering even after long-term testosterone use [52,54]. Furthermore, several nonbinary participants explicitly refrained from a binary feminization/masculinization framing of voice goals, describing both voice coaching and existing apps as engaging in "gender-boxing" that rendered them unusable (Q55, Q74). This underscores that voice modification goals cannot be inferred from gender identity alone and that both GAVT services and mHealth tools should accommodate individualized, nonbinary voice targets.

Contrary to most existing voice-related qualitative research among TGNC individuals, our study population consisted largely of individuals who desired but did not feel that they had access to GAVT. Our cohort also included individuals across a wide range of gender identities, geographic locations, ages, and prior experiences with various gender-affirming treatments. This allowed for substantial exploration of several areas that are not robustly represented in existing literature. For example, our findings regarding self-guided attempts at voice modification, including some practices that participants themselves worried could be causing vocal injury, offer an important glimpse into the vocal practices of TGNC individuals in a landscape of formal GAVT being inaccessible to many. Our findings regarding the importance of context in determining voice use and attempts at voice modification were in line with

XSL•FO
RenderX

prior research [52]. Furthermore, our study uncovered not only themes regarding the inaccessibility of formal voice training but also several patterns of frustration with both publicly available online materials and prior voice-centered mobile apps, which substantially builds on prior work by Bush et al [55]. This represents a shortcoming of prior qualitative research in this field, which has largely recruited individuals already presenting to SLPs for GAVT, thus failing to capture the broad swath of the TGNC population who do not reach this point due to previously mentioned substantial barriers.

Our results also represent a largely positive reception among the first cohort of target users of Attuned from diverse geography, age groups, and gender identities. These results should be interpreted in light of the risk of acquiescence bias, given the interviewers' membership in the development team, and are best understood as encouraging formative feedback rather than evidence of effectiveness. Despite technical glitches in the initial version of the app, participants looked favorably on the app's content, structure, and tools to support their voice modification efforts. Participants emphasized the empowering nature of receiving detailed information about the mechanics of the voice and gender perception, in line with prior qualitative studies of individuals who have undergone GAVT [56,57]. Of note, qualitative literature investigating outcomes of GAVT is remarkably sparse, despite known shortcomings of quantitative measurements in this realm, such as lack of correlation between different patient-reported outcome measures or between these measures and objective acoustic outcomes [5,14]. This suggests that existing literature may not be adequately capturing the full depth of TGNC individuals' experiences with GAVT and identifies a crucial area for further investigation.

Our study also uncovered substantial shortcomings of an app-based delivery of voice training, most prominently the lack of feedback and personalization. While certain aspects of these limitations will be intrinsic to any mobile app, advances in AI and biofeedback may allow future versions of Attuned to respond more robustly in real-time to users [58]. An additional consideration for app-based GAVT is the absence of clinical screening prior to training. In many institutional settings, laryngoscopic examination is recommended before initiating voice training to exclude underlying laryngeal pathology, whereas mobile apps are by nature accessible without professional evaluation. Attuned addresses this in part through an in-app disclaimer encouraging users with voice symptoms to seek evaluation by an otolaryngologist. Moreover, our qualitative findings suggest that in the current landscape, many TGNC individuals without access to formal care already attempt self-guided voice modification using unstructured online resources (including practices participants themselves worried could cause vocal injury). Structured, professionally designed guidance with appropriate safety recommendations may therefore represent a safer alternative to the unstructured resources many users currently rely on. It is also important to note that while some mobile apps focused on GAVT exist, most existing offerings range from free pitch-tracking with limited instructional material to structured instructional content available by subscription only [58-63]. TruVox (University of Cincinnati), a free web-based application, provides real-time pitch and resonance feedback alongside instructional content [58]. To the authors' knowledge, Attuned is distinctive in combining a free sequential video curriculum narrated by voice clinicians with practice assignments and pitch-tracking in a native mobile app developed with a CAB.

With regard to the objective assessment of Attuned, scores on the battery of validated scales were high across quality, functionality, health literacy, and readability, though these tests were performed by members of the authorship team and should be interpreted accordingly. Attuned exceeded the previously published means of laryngology-focused mobile apps across all evaluated scales, notably achieving above 4/5 on the MARS scale and all associated subcategories, 100% on the PEMAT actionability and understandability scales, and resulting in scores below the benchmark of an eighth-grade reading level for the majority of readability scales [21]. Of note, our prior systematic assessment of laryngology mobile apps [21] did not use the SUS; however, Attuned's score of 92.5 was well above the previously reported benchmarks of 68 and 80 across all digital health tools [64]. This reflects the goal of our app design and content creation process to develop a tool that would be readily available to users from around the country and with varying educational and socioeconomic backgrounds [22]. In addition, it should be noted that the version of the app subject to evaluation on these scales was the first version of the app ever made available and contained known glitches and deficits, which have already been targeted for improvement in future iterations. Future iterations of Attuned addressing these known deficits may perform differently, which will require further evaluation.

Despite its novel contributions, this study is not without limitation. First, while TGNC community members were intimately involved in the creation of Attuned and early plans for the qualitative portion of this study, we did not follow a strict community-based participatory research methodology without a governance structure, authorship opportunities, and fair compensation as previously described [22]. Given our group's aspirations toward true community-based participatory research [22], this represents a crucial area of improvement for future studies of Attuned. A new CAB is currently being assembled to guide the revision process of the app, as well as to participate as co-investigators in future prospective research using Attuned.

While our use of semistructured interviews allowed deeper exploration of important themes, the necessary trade-off is that interviews were slightly different from each other and could not be directly compared. In addition, participants may have been subject to acquiescence bias, particularly in the post-Attuned interviews, whereby individuals may have felt subtle pressure to overemphasize positive attitudes toward Attuned, knowing that the interviewers were part of the app's development team. We sought to partially mitigate this bias by asking open-ended questions to signal openness to any opinions that participants would share. Additionally, the objective assessment scales were completed by 2 authors. Although they were not involved in the app's development, they were not independent of the study team, and these ratings therefore cannot be regarded as fully unbiased. Usability measures were not collected directly from study participants, which we have

XSL•FO
RenderX

incorporated into the design of our planned prospective follow-up study. Furthermore, we did not apply exclusion criteria to potential participants (aside from requiring that they were identified within the TGNC umbrella). Our study may therefore be limited by selection bias, as participants choosing to participate in a study of GAVT are likely already interested in GAVT and are also highly motivated to pursue it; our study population—despite our best efforts to recruit broadly from across the country and to include a variety of gender identities and socioeconomic statuses—may thus have limited generalizability to the population of TGNC individuals at large. Participants' specific voice modification goals (feminization, masculinization, androgynous, or other targets) were not captured as a survey item. Although gender identity distribution offers an approximation, these do not map neatly onto one another, and explicit characterization of voice goals will be incorporated into future studies. Our study did not collect backend information regarding each participant's usage pattern or time spent on the app, which may have allowed for deeper analysis of participants' impressions of the app. Finally, both interviewers are cisgender men and members of the LGBTQ+ community but not the TGNC community; this positionality, as previously mentioned, may have impacted the responses of some participants, who may have responded differently to non-LGBTQ+ interviewers or fellow TGNC individuals.

Future directions for our team include 2 directions. First, a revision of the Attuned app based on our participants' feedback is currently underway, guided by a newly assembled CAB and addressing the technical issues identified in this study and expanding the diversity of voices represented in the instructional content. Planned future capabilities include personalized AI-driven guidance and real-time voice biofeedback to address the feedback and personalization limitations identified by participants, with development contingent on available funding and technical resources. Our team also plans a prospective, quantitative assessment of the app's efficacy, including participant-administered usability measures, as compared with standard-of-care in-person GAVT. The present study also highlighted several important future directions for the field, including prioritization of studies of transgender men or other individuals seeking voice masculinization, outreach to populations without access to traditional GAVT, and expansion of overall qualitative study of TGNC individuals either seeking or after voice modification training in order to identify shortcomings and further improve outcomes for these individuals.

Conclusion

The voice is a crucial part of identity and interpersonal interactions for TGNC individuals, and many desire tools for voice modification but have difficulty accessing high-quality GAVT. In a landscape full of unstructured free resources of varying levels of trustworthiness, there is a critical unmet need for a structured, clinically informed curriculum that can be accessible at scale. Attuned, a free mobile app developed through a user-centered approach, was positively received by an initial cohort of TGNC users. While app-based training cannot replace the individualized feedback of one-on-one voice training, these findings suggest promise in expanding access to structured, professionally developed GAVT, to be confirmed in a prospective comparative study.

Acknowledgments

No generative AI tool was used at any stage in the preparation of this manuscript.

Funding

This was works supported by the Dalio Center for Health Justice grant, Columbia University Vagelos College of Physicians and Surgeons and Weill Cornell Medicine.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

AL helped with the project administration and put project resources together, AR acquired funding, conceptualized, and supervised the study, and reviewed and edited the manuscript, CPK conducted the interviews and performed data curation and formal analysis, DB reviewed and edited the manuscript, DL conceptualized the study and performed formal analysis, EO performed formal analysis, IC performed formal analysis, ILA conducted the interviews, performed data curation, and formal analysis, and wrote the original draft of the manuscript, KAC reviewed and edited the manuscript, MP reviewed and edited the manuscript, RC reviewed and edited the manuscript, SWA helped with the project administration and put project resources together, and VM performed formal analysis. All authors reviewed the final manuscript.

Conflicts of Interest

Anaïs Rameau is an advisor for SoundHealth Inc and she is the founder of VAST Health Inc.

Multimedia Appendix 1
Demographic data for participants in qualitative study.

XSL•FO
RenderX

[DOCX File, 19 KB - mhealth_v14i1e96102_app1.docx]

Multimedia Appendix 2
Exemplar quotations: navigating and modifying the voice across social contexts
[DOCX File, 23 KB - mhealth_v14i1e96102_app2.docx]

Multimedia Appendix 3
Exemplar quotations: current apps fall short of user needs.
[DOCX File, 23 KB - mhealth_v14i1e96102_app3.docx]

Multimedia Appendix 4
Exemplar quotations: Attuned can positively impact the voice
[DOCX File, 22 KB - mhealth_v14i1e96102_app4.docx]

Checklist 1
COREQ checklist.
[PDF File, 379 KB - mhealth_v14i1e96102_app5.pdf]

References

1. Understanding transgender people: the basics. Advocates for Trans Equality (A4TE). 2023 Jan 27. URL: https://transequality.org/issues/resources/understanding-transgender-people-the-basics [accessed 2025-11-28]
2. Kennedy E, Thibeault SL. Voice-gender incongruence and voice health information-seeking behaviors in the transgender community. Am J Speech Lang Pathol 2020 Aug 4;29(3):1563-1573. [doi: 10.1044/2020_AJSLP-19-00188] [Medline: 32539455]
3. Stewart L, Oates J, O'Halloran P. "My Voice Is My Identity": the role of voice for trans women's participation in sport. J Voice 2020 Jan;34(1):78-87. [doi: 10.1016/j.jvoice.2018.05.015] [Medline: 30055980]
4. Hancock AB, Krissinger J, Owen K. Voice perceptions and quality of life of transgender people. J Voice 2011 Sep;25(5):553-558. [doi: 10.1016/j.jvoice.2010.07.013] [Medline: 21051199]
5. Chadwick KA, Coleman R, Andreadis K, Pitti M, Rameau A. Outcomes of gender-affirming voice and communication modification for transgender individuals. Laryngoscope 2022 Aug;132(8):1615-1621. [doi: 10.1002/lary.29946] [Medline: 34787313]
6. Merrick G, Figol A, Anderson J, Lin RJ. Outcomes of gender affirming voice training: a comparison of hybrid and individual training modules. J Speech Lang Hear Res 2022 Feb 9;65(2):501-507. [doi: 10.1044/2021_JSLHR-21-00056] [Medline: 35041797]
7. Carew L, Dacakis G, Oates J. The effectiveness of oral resonance therapy on the perception of femininity of voice in male-to-female transsexuals. J Voice 2007 Sep;21(5):591-603. [doi: 10.1016/j.jvoice.2006.05.005] [Medline: 16822646]
8. Chadwick KA, Liao D, Alter IL, et al. Outcomes of gender-affirming voice and communication modification training for non-binary individuals: a case series. J Voice 2026 Jan;40(1):199-205. [doi: 10.1016/j.jvoice.2023.08.005] [Medline: 37673753]
9. Quinn S, Oates J, Dacakis G. The effectiveness of gender affirming voice training for transfeminine clients: a comparison of traditional versus intensive delivery schedules. J Voice 2024 Sep;38(5):1250. [doi: 10.1016/j.jvoice.2022.03.001] [Medline: 35400554]
10. Quinn S. Intensive schedules and contemporary directions in gender affirming voice training for transfeminine clients. : La Trobe; 2022 URL: https://opal.latrobe.edu.au/articles/thesis/Intensive_Schedules_and_Contemporary_Directions_in_Gender_Affirming_Voice_Training_for_Transfeminine_Clients/23551518?file=41329467 [accessed 2026-10-04]
11. Azul D, Hancock AB. Who or what has the capacity to influence voice production? Development of a transdisciplinary theoretical approach to clinical practice addressing voice and the communication of speaker socio-cultural positioning. Int J Speech Lang Pathol 2020 Oct;22(5):559-570. [doi: 10.1080/17549507.2019.1709544] [Medline: 32063069]
12. Alekseeva M, Myachykov A, Bermudez-Margaretto B, Shtyrov Y. Neurophysiological correlates of automatic integration of voice and gender information during grammatical processing. Sci Rep 2022 Jul 30;12(1):13114. [doi: 10.1038/s41598-022-14478-2] [Medline: 35908074]
13. Chen X, Li Z, Setlur S, Xu W. Exploring racial and gender disparities in voice biometrics. Sci Rep 2022;12(1):3723. [doi: 10.1038/s41598-022-06673-y]
14. Young VN, Yousef A, Zhao NW, Schneider SL. Voice and stroboscopic characteristics in transgender patients seeking gender-affirming voice care. Laryngoscope 2021 May;131(5):1071-1077. [doi: 10.1002/lary.28932] [Medline: 32750170]
15. Gonzales G, Henning-Smith C. Barriers to care among transgender and gender nonconforming adults. Milbank Q 2017 Dec;95(4):726-748. [doi: 10.1111/1468-0009.12297] [Medline: 29226450]

XSL•FO
RenderX

16. Moog D, Timmons Sund L. Clinician and consumer perspectives on gender-affirming voice services. J Voice 2023 Sep;37(5):805. [doi: 10.1016/j.jvoice.2021.05.009] [Medline: 34167855]
17. James SE, Herman JL, Rankin S, Keisling M, Mottet L, Anafi M. The report of the 2015 US transgender survey. : National Center for Transgender Equality; 2016 URL: https://transequality.org/sites/default/files/docs/usts/USTS-Full-Report-Dec17.pdf [accessed 2026-10-04]
18. Hancock AB, Downs SC. Listening to gender-diverse people of color: barriers to accessing voice and communication care. Am J Speech Lang Pathol 2021 Sep 23;30(5):2251-2262. [doi: 10.1044/2021_AJSLP-20-00262] [Medline: 34284637]
19. DeVore EK, Gadkaree SK, Richburg K, et al. Coverage for gender-affirming voice surgery and therapy for transgender individuals. Laryngoscope 2021 Mar;131(3):E896-E902. [doi: 10.1002/lary.28986] [Medline: 32776558]
20. Kearns S, Hardie P, O'Shea D, Neff K. Instruments used to assess gender-affirming healthcare access: a scoping review. PLoS One 2024;19(6):e0298821. [doi: 10.1371/journal.pone.0298821] [Medline: 38829881]
21. Odigie E, Andreadis K, Chandra I, et al. Are mobile applications in laryngology designed for all patients? Laryngoscope 2023 Jul;133(7):1540-1549. [doi: 10.1002/lary.30465] [Medline: 36317789]
22. Alter IL, Chadwick KA, Andreadis K, et al. Developing a mobile application for gender-affirming voice training: a community-engaged approach. Laryngoscope Investig Otolaryngol 2024 Dec;9(6):e70043. [doi: 10.1002/lio2.70043] [Medline: 39679168]
23. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. Health Serv Res 2013 Dec;48(6 Pt 2):2134-2156. [doi: 10.1111/1475-6773.12117] [Medline: 24279835]
24. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
25. Guest G, Namey E, Chen M. A simple method to assess and report thematic saturation in qualitative research. PLoS One 2020;15(5):e0232076. [doi: 10.1371/journal.pone.0232076] [Medline: 32369511]
26. Morse JM. The Significance of Saturation. Qual Health Res 1995 May;5(2):147-149. [doi: 10.1177/104973239500500201]
27. Radford A, Kim JW, Xu T, Brockman G, McLeavey C, Sutskever I. Robust speech recognition via large-scale weak supervision. Presented at: Proceedings of the 40th International Conference on Machine Learning; Jul 23-29, 2023; Honolulu, Hawaii p. 28492-28518 URL: https://mlanthology.org/icml/2023/radford2023icml-robust/ [accessed 2026-10-04]
28. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
29. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
30. Patient apps for improved healthcare from novelty to mainstream. ScienceOpen,com. URL: https://www.scienceopen.com/document?vid=696d6314-87cd-43c3-9a5a-421874cbb752 [accessed 2025-11-28]
31. CDC. The CDC clear communication index. Centers for Disease Control and Prevention. 2025. URL: https://www.cdc.gov/ccindex/index.html [accessed 2025-11-28]
32. Broderick J, Devine T, Lemerise AJ, et al. Designing health literate mobile apps. NAM Perspectives 2014;4(1). [doi: 10.31478/201401a]
33. PEMAT tool for audiovisual materials (PEMAT-a/v). Agency for Healthcare Research and Quality. URL: https://www.ahrq.gov/health-literacy/patient-education/pemat-av.html [accessed 2025-11-28]
34. Brooke J. SUS: a quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry: Taylor & Francis; 1996:189-194 URL: https://digital.ahrq.gov/sites/default/files/docs/survey/systemusabilityscale%2528sus%2529_comp%255B1%255D.pdf [accessed 2026-10-04]
35. Take control of your content with ReadablePro. Readable. URL: https://readable.com/ [accessed 2025-11-28]
36. Kincaid JP, Fishburne J, Robert PR, et al. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel: defense technical information center. : Defense Technical Information Center (DTIC); 1975. [doi: 10.21236/ADA006655]
37. Gunning R. The Technique of Clear Writing: McGraw-Hill; 1968. URL: http://archive.org/details/techniqueofclear00gunn [accessed 2026-09-23]
38. Coleman M, Liau TL. A computer readability formula designed for machine scoring. J Appl Psychol 1975;60(2):283-284. [doi: 10.1037/h0076540]
39. Laughlin GHM. SMOG grading—a new readability formula. J Read 1969;12(8):639-646 [FREE Full text]
40. Chall JS. Readability Revisited: The New Dale-Chall Readability Formula. Brookline Books. 1995. URL: http://archive.org/details/readabilityrevis0000chal [accessed 2025-11-28]
41. Morse E, Odigie E, Gillespie H, Rameau A. The readability of patient-facing social media posts on common otolaryngologic diagnoses. Otolaryngol Neck Surg 2024 Apr;170(4):1051-1058. [doi: 10.1002/ohn.584] [Medline: 38018504]
42. Association AP. Guidelines for psychological practice with transgender and gender nonconforming people. American Psychologist 2015;70(9):832-864. [doi: 10.1037/a0039906]

XSL•FO
RenderX

43. Okamuro K, Card A, Barton HJ, et al. Patient and clinician perspectives on misgendering in healthcare. BMJ Qual Saf 2026 Mar 19;35(4):257-265. [doi: 10.1136/bmjqs-2024-018364] [Medline: 40484628]
44. American College of Obstetricians and Gynecologists. Health care for transgender and gender diverse individuals: ACOG Committee Opinion, Number 823. Obstet Gynecol 2021;137(3):e75-e88. [doi: 10.1097/AOG.0000000000004294]
45. Chowdhury R, Silver JA, Philippopoulos E, Kost KM. Pitch alteration techniques for transgender women: a systematic review of surgical and nonsurgical approaches. J Voice 2025 Jul 17:00257-00257. [doi: 10.1016/j.jvoice.2025.06.032] [Medline: 40681377]
46. Lanham K, Melnick BA, O'Connor MJ, et al. Efficacy and patient satisfaction in voice feminization procedures: a systematic review and meta-analysis. Otolaryngol Head Neck Surg 2025 May;172(5):1521-1538. [doi: 10.1002/ohn.1173] [Medline: 39963873]
47. Lorimer G, Rutter B. The efficacy of gender-affirming voice and communication therapy-a systematic review. J Voice 2025 Feb 13:00015-00013. [doi: 10.1016/j.jvoice.2025.01.014] [Medline: 39947968]
48. Nygren U, Nordenskjöld A, Arver S, Södersten M. Effects on voice fundamental frequency and satisfaction with voice in trans men during testosterone treatment-a longitudinal study. J Voice 2016 Nov;30(6):766. [doi: 10.1016/j.jvoice.2015.10.016] [Medline: 26678122]
49. Irwig MS. Testosterone therapy for transgender men. Lancet Diabetes Endocrinol 2017 Apr;5(4):301-311. [doi: 10.1016/S2213-8587(16)00036-X] [Medline: 27084565]
50. François F, Wolfberg J, Croegaert-Koch C, Fujiki RB, Thibeault SL. Defining goals of transfeminine individuals seeking gender-affirming voice therapy: a qualitative study. Am J Speech Lang Pathol 2025 Jul 10;34(4):2324-2333. [doi: 10.1044/2025_AJSLP-24-00369] [Medline: 40340436]
51. Jin JL, Baylor C, Teixeira J, Yorkston K, Nuara M. Reframing transgender communication in gender-affirming communication care: comfort and confidence are the main goals. Int J Speech Lang Pathol 2024 Oct;26(5):750-764. [doi: 10.1080/17549507.2023.2259124] [Medline: 37907084]
52. Azul D, Arnold A, Neuschaefer-Rube C. Do transmasculine speakers present with gender-related voice problems? Insights from a participant-centered mixed-methods study. J Speech Lang Hear Res 2018 Jan 22;61(1):25-39. [doi: 10.1044/2017_JSLHR-S-16-0410] [Medline: 29273818]
53. Holmberg J, Linander I, Södersten M, Karlsson F. Exploring motives and perceived barriers for voice modification: the views of transgender and gender-diverse voice clients. J Speech Lang Hear Res 2023 Jul 12;66(7):2246-2259. [doi: 10.1044/2023_JSLHR-23-00042] [Medline: 37263019]
54. Papeleu T, Leyns C, Alighieri C, et al. Voice and communication in transmasculine individuals one year under testosterone therapy: a qualitative study. J Voice 2026 Sep;40(5):1524. [doi: 10.1016/j.jvoice.2024.04.003] [Medline: 38679523]
55. Bush EJ, Krueger BI, Cody M, Clapp JD, Novak VD. Considerations for voice and communication training software for transgender and nonbinary people. J Voice 2024 Sep;38(5):1251. [doi: 10.1016/j.jvoice.2022.03.002] [Medline: 35382956]
56. Papeleu T, Leyns C, Motmans J, Vervalcke J, Kerckhof M, D'haeseleer E. The experiences of transgender and gender diverse individuals after intonation training for a more masculine-sounding voice: a qualitative study. J Voice 2025 Sep 4:00315-00317. [doi: 10.1016/j.jvoice.2025.08.001] [Medline: 40913008]
57. Harris KJ, Beck LA, Worth BF, Grossman RB. First-person perspectives of gender-affirming voice feminization training: a mixed methods approach. J Voice 2026 Jul;40(4):1243. [doi: 10.1016/j.jvoice.2024.03.013] [Medline: 38688779]
58. McAllister T, Eagen C, McKenna VS, et al. Real-time resonance biofeedback for gender-affirming voice training: usability testing of the TruVox web-based application. J Voice 2025 Oct 31:00420-00425. [doi: 10.1016/j.jvoice.2025.10.004] [Medline: 41176464]
59. Hawley JL, Hancock AB. Incorporating mobile app technology in voice modification protocol for transgender women. J Voice 2024 Mar;38(2):337-345. [doi: 10.1016/j.jvoice.2021.09.001] [Medline: 34706847]
60. Genderfluent - transgender voice training app. Genderfluent. URL: https://www.genderfluentapp.com/ [accessed 2026-02-27]
61. Voice Training for Gender Expression.: Voice Whiz URL: https://voicewhiz.app/trans-voice-training/ [accessed 2026-02-27]
62. VoiceShift homepage. VoiceShift. URL: https://www.voiceshift.app/ [accessed 2026-02-27]
63. Christella voiceup app: voice feminisation app. Christella Antoni. URL: https://www.christellaantoni.co.uk/transgender-voice/voiceupapp/ [accessed 2026-02-27]
64. Lewis J, Sauro J. Item benchmarks for the system usability scale. J Usability Stud 2018 May 25;13(3):158-167. [doi: 10.5555/3294033.3294037]

Abbreviations

AFAB: assigned female at birth
AMAB: assigned male at birth
CA-ICC: calculation of an intraclass correlation
CAB: community advisory board
CDC: Centers for Disease Control
COREQ: Consolidated Criteria for Reporting Qualitative Research

XSL•FO
RenderX

GAVT: gender-affirming voice training
HIPAA: Health Insurance Portability and Accountability Act
IMS: Intercontinental Medical Statistics
IOM: Institute of Medicine
LGBTQ+: lesbian, gay, bisexual, transgender, queer, or questioning
MARS: Mobile App Rating Scale
mHealth: mobile health
PEMAT: Patient Education Materials Assessment Tool
SLP: speech-language pathologist
SMOG: Simple Measure of Gobbledygook
SUS: System Usability Scale
TGNC: transgender and gender-nonconforming

Edited by A Stone; submitted 25.Mar.2026; peer-reviewed by D Shumer, T McAllister; revised version received 05.Sep.2026; accepted 08.Sep.2026; published 08.Oct.2026.

Please cite as:
Alter IL, Kruglik CP, Albert SW, Coleman R, Pitti M, Li A, Andreadis K, Bayne D, Odigie E, Chandra I, Mocchetti V, Liao D, Chadwick KA, Rameau A
Gender-Affirming Voice Training Delivered by a Novel Mobile App: Mixed Methods Study
JMIR Mhealth Uhealth 2026;14:e96102
URL: https://mhealth.jmir.org/2026/1/e96102
doi:10.2196/96102

© Isaac L Alter, Christopher P Kruglik, Sara W Albert, Rachel Coleman, Mary Pitti, Alexandra Li, Katerina Andreadis, David Bayne, Eseosa Odigie, Iyra Chandra, Valentina Mocchetti, David Liao, Keith A Chadwick, Anaïs Rameau. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 8.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Dynamics of Digital Support Platform Use Among Family Caregivers of Children With Disabilities: Longitudinal Study

Xiaojiao Yuan[1], PhD; Xi Zhao[1], BA; Meng Zhang[2], MAP; Yurou Duan[1], BA; Yizhong Shui[3], PhD

[1]School of Education and Psychology, Southwest Minzu University, Chengdu, China

[2]Health Care and Nursing College, Southwest Jiaotong University Hope College, Chengdu, China

[3]School of Economics and Management, Chengdu Normal University, No. 99 East Haike Road, Wenjiang District, Chengdu, Sichuan, China

Corresponding Author:
Yizhong Shui, PhD
School of Economics and Management, Chengdu Normal University, No. 99 East Haike Road, Wenjiang District, Chengdu, Sichuan, China

Abstract

Background: Family caregivers of children with disabilities frequently experience high levels of stress and psychological burden. Digital support platforms offer accessible channels for information and emotional support, but their long-term use patterns remain underexplored.

Objective: This study explored the longitudinal patterns of digital support platform use among family caregivers in China, with a focus on factors potentially associated with initial adoption and continued engagement.

Methods: A 3-wave survey was conducted over 6 months with 240 primary caregivers of children with disabilities in China. A 2-part latent growth curve model was used to jointly model adoption (use vs nonuse) and engagement (use frequency among users). Key demographic, psychosocial, and child-related factors were examined as predictors.

Results: Adoption increased significantly over time (probit slope=0.54, SE 0.16; *P*=.001). Higher baseline adoption was associated with male sex (β=.19; *P*=.02), higher education (β=.42; *P*<.001), better socioeconomic status (β=.26; *P*=.001), stronger social support (β=.17; *P*=.03), and more depressive symptoms (β=.18; *P*=.02). Among users, engagement frequency showed no significant average change over time (mean slope=–0.07, SE 0.07; *P*=.31). Higher initial engagement was associated with higher education (β=.36; *P*<.001) and social support (β=.20; *P*=.04). Better child rehabilitation status was associated with steeper declines in engagement over time (β=–.28; *P*=.03).

Conclusions: Adoption and engagement showed distinct longitudinal patterns and were associated with different factors. The findings suggest that reducing digital barriers and aligning support with caregivers' evolving needs may be relevant to facilitating platform use and sustained engagement. This study offers a theory-informed lens for describing digital support platform use among vulnerable caregiving populations.

(JMIR Mhealth Uhealth 2026;14:e92481) doi:10.2196/92481

KEYWORDS

digital support; family caregivers; children with disabilities; adoption; engagement

Introduction

Background

Amid rapid digitalization, family caregivers of children with disabilities are increasingly seeking support online. Substantial evidence shows high caregiving stress and poor mental health outcomes in this population [1-3]. Caregivers need reliable information about rehabilitation services, special education, and policy, as well as emotional understanding and support. However, limited professional resources and disability-related stigma often constrain offline help seeking. In this context, digital support platforms (eg, peer forums, social media groups, and specialized online communities) offer scalable, low-cost channels for information exchange and emotional support, often with anonymity [4].

While digital platforms have been shown to enhance perceived social support and improve caregiver well-being [4,5], most evidence is cross-sectional or intervention-specific. Less is known about how routine platform use changes over time and whether determinants differ for adoption vs engagement. Addressing these gaps is important for theory—clarifying how vulnerable populations engage with digital health resources—and for practice—informing design and policy that foster equitable and sustained engagement.

XSL•FO
RenderX

Use of Digital Support Platforms Among Caregivers of Children With Disabilities

Caregivers commonly use digital platforms for 2 core functions. First, information seeking: they look for practical guidance on condition management, rehabilitation options, and special education to address day-to-day challenges [6]. Second, social and emotional support: platforms connect families facing similar circumstances, facilitating experience sharing and mutual understanding that can alleviate isolation [7,8]. Qualitative evidence indicates that platforms aligning with both informational and emotional needs foster more positive engagement [9].

Evidence for benefits spans unstructured communities and structured programs. Open online communities have been associated with higher perceived social support and lower stress levels among parents of children with autism [8]. Structured interventions (eg, CaregiverTLC) can strengthen self-efficacy and coping capacity [10,11]. Reviews suggest that peer support networks may serve as a pathway to strengthening family caregivers and fostering their resilience [12]. Studies also point to potential compensatory effects, with more diverse, supportive networks linked to lower major depression risk [13].

Factors Associated With the Use of Digital Support Platforms

Caregivers' digital platform use can be understood through Andersen's behavioral model of health services use (BMHSU), which categorizes determinants into 3 domains: predisposing characteristics, enabling resources, and needs [14]. Predisposing characteristics (eg, age, gender, and education) reflect general propensities to use services. Enabling resources (eg, socioeconomic conditions, social support, and access to devices or connectivity) shape the practical feasibility of use. Needs, both perceived by individuals and evaluated by professionals, are the most proximal drivers of use. The model has demonstrated utility across traditional health services and increasingly in digital health contexts [15-17].

For caregivers, 3 key factors emerge. First, predisposing characteristics matter. Higher educational attainment is generally associated with greater use for accessing and sharing health information, while older age, lower education, and limited language proficiency are linked to barriers such as lower digital literacy and reduced access [18-21]. Gender differences in engagement remain unclear, as studies often overrepresent mothers and reflect caregiving roles [4,20].

Second, enabling resources are pivotal. Socioeconomic status—income, employment, and access to devices and connectivity—conditions capacity to engage and shapes use patterns [22]. Contextual enabling conditions, including credible information and robust privacy protections, can lower perceived risk; caregivers often prefer closed groups to safeguard privacy [23,24]. Social support further enables platform use by providing informational, emotional, and practical assistance, which helps caregivers navigate digital tools and sustains engagement [25,26].

Third, needs are immediate drivers of use. Psychological distress often drives online support seeking, with higher depressive symptoms linked to greater pursuit of emotional support [27-29]. Children's health and rehabilitation status represent key evaluated needs that sustain information seeking and support seeking over time. Even after receiving a diagnosis, many parents continue to consult multiple sources for condition details, prognosis, and management strategies, highlighting the persistent role of clinical needs in motivating digital engagement [30].

The Present Study

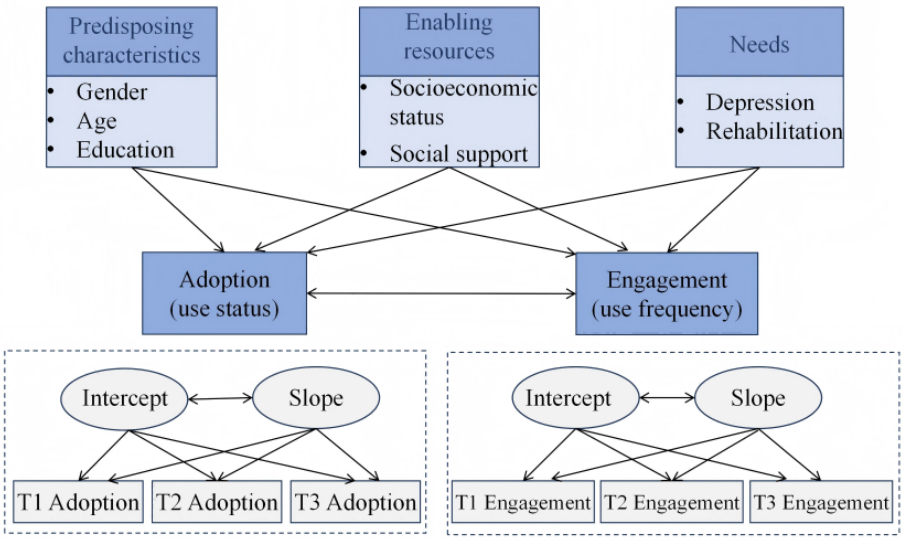
Despite growing interest in caregivers' use of digital support platforms, the literature is dominated by cross-sectional designs and short-term intervention evaluations that seldom track routine, real-world use over time. Consequently, evidence on sustained use beyond program contexts remains limited. Prior work rarely distinguishes whether caregivers use a platform (adoption) from how often they participate among users (engagement). Conflating these processes introduces zero-inflation concerns and may obscure distinct trajectories and determinants.

To address these gaps, we conducted a 3-wave, 6-month longitudinal study of 240 family caregivers of children with disabilities in China. We used a 2-part latent growth curve model to jointly estimate adoption (use vs nonuse) and engagement (frequency among users) trajectories [31,32]. This approach accommodates zero-inflated, semicontinuous data and enables direct comparison of predictors across processes.

Guided by the BMHSU [14] and tailored to the caregiving context, we examined predictors in 3 domains: predisposing characteristics (caregiver gender, age, and education), enabling resources (socioeconomic status and social support), and needs (caregivers' depressive symptoms as perceived need; children's rehabilitation status as evaluated need). The conceptual framework is presented in Figure 1.

XSL•FO
RenderX

Figure 1. Conceptual framework of the study.



We hypothesized that both adoption probability and engagement would increase over time (hypothesis 1), there would be significant interindividual variability in the initial levels and growth trajectories of both adoption and engagement (hypothesis 2), and predictors would differentially relate to adoption vs engagement (hypothesis 3).

Methods

Study Design

This study used a prospective longitudinal design with 3 survey waves. Digital support platform use was repeatedly measured at each wave using a 3-month recall window aligned with the interwave interval. All other predictors, including demographics and psychosocial variables, were assessed at baseline and used in analyses as predictors.

Participants

The inclusion criteria were as follows: (1) being the primary family caregiver (ie, the adult most responsible for the child's daily care and health decisions) of a child or adolescent aged ≤18 years with a confirmed disability diagnosis, (2) ability to complete an online smartphone survey, and (3) provision of electronic informed consent. The exclusion criteria were as follows: (1) severe cognitive or communication barriers precluding survey completion and (2) duplicate entries within the same household (1 caregiver per child).

We recruited caregivers from rehabilitation centers, special education schools, and hospitals in Sichuan, Chongqing, and Shandong, China. A total of 485 valid wave 1 responses were obtained; 259 (53.4%) caregivers consented to be recontacted. Of these, 199 and 191 valid responses were collected at waves 2 and 3, respectively. The longitudinal analytic sample comprised 240 caregivers who completed at least 1 follow-up wave. Because recruitment relied on convenience sampling via advertisements and voluntary self-registration, the number of individuals who were exposed to the recruitment materials or invited to participate was not available; therefore, a formal response rate could not be calculated.

Participant characteristics were as follows: 29% (70/240) were male, with a mean age of 37.29 (SD 6.85) years. Most (215/240, 89.6%) had completed junior high school or higher, and 33% (78/240) held a bachelor's degree or above. Subjective socioeconomic status was rated as average or below by 86.3% (207/240). The most common child disabilities were intellectual (65/240, 27%) and psychiatric (51/240, 21%). Rehabilitation status was mostly rated as fair or good (174/240, 72.7%). Additional details are presented in Table S1 in Multimedia Appendix 1.

Measures

Digital Support Platform Use

A study-developed measure assessed digital support platform use, informed by prior interviews with family caregivers to ensure cultural relevance and alignment with common platform formats. The questionnaire was reviewed by 2 subject matter experts before data collection. To reduce recall bias, respondents reported their platform use over the past 3 months, with descriptions and examples provided for each platform type. Participants indicated how often they used disability-focused digital support platforms (eg, service portals, live streams, chat rooms, social media groups, and online forums) in the past 3 months. Responses were coded as 0=never, 1=rarely (eg, less than once per month), 2=occasionally (eg, 1‐3 times per month), 3=often (eg, 1‐3 times per week), and 4=almost daily (eg, 4 or more times per week). Platform adoption was coded as 0=nonuse and 1=any use (responses 1‐4), while engagement was modeled as the frequency of use among users (responses

XSL•FO
RenderX

1 - 4). Nonusers were asked to report reasons for nonuse, including lack of time, unawareness, unfamiliarity, privacy concerns, or other (open ended).

Social Support

Perceived social support was measured at baseline using the 12-item Social Support Scale [33], which assesses family, friend, and other support. Items are rated on a 7-point scale (1=strongly disagree to 7=strongly agree). According to the scoring criteria in the original guidance, total scores of 12 to 36, 37 to 60, and 61 to 84 indicate low, moderate, and high support, respectively. In the present study, the Cronbach α for the scale was 0.92.

Depression

Depressive symptoms were measured at baseline with the Chinese version of the 9-item Patient Health Questionnaire [34,35]. The scale consists of 9 items rated on a 4-point scale (0=not at all to 3=nearly every day). Scores of 5, 10, 15, and 20 correspond to the cutoff points for mild, moderate, moderately severe, and severe depressive symptoms, respectively. The Chinese version of this scale has been previously validated in a sample of parents of children with disabilities in China [36]. In the present study, the Cronbach α for the scale was 0.94.

Demographic Characteristics

Demographic characteristics were collected at baseline for both the child and the caregiver. Child factors included the primary disability type (visual, hearing, physical, speech, intellectual, psychiatric, or multiple) and current rehabilitation status rated on a 5-point scale (1=extremely poor to 5=excellent). Caregiver factors included gender, age, education, and subjective socioeconomic status, obtained via a structured questionnaire.

Ethical Considerations

This study adhered to the ethical principles of the Declaration of Helsinki and was approved by the Academic Committee of Southwest Minzu University (SMU-202401037). All participants provided electronic informed consent prior to participation.

Procedures

The study was conducted from September 2024 to July 2025, with 3 waves of data collection at 3-month intervals. Participants were recruited through collaborating schools, hospitals, and rehabilitation centers in Sichuan Province, Chongqing Municipality, and Shandong Province. These institutions distributed posters advertising the study, and interested caregivers of children with disabilities voluntarily registered by scanning a QR code. Eligible participants who provided informed consent were invited to complete a 10-minute online survey for each wave. To prevent duplicate submissions, IP-based restrictions were implemented, and surveys completed in unrealistically short times were excluded as invalid. Follow-up surveys were distributed 3 months after the participants' initial survey date. As an incentive, participants received CNY 10 (CNY 1=US $0.15 as of September 5, 2026) for each completed wave.

Data Analysis

Analyses included participants who completed the baseline and at least 1 follow-up assessment (n=240). To examine selective attrition, baseline characteristics were compared between participants retained for longitudinal analyses and those lost to follow-up. No significant differences were found in platform use (χ^2_4=7.5; *P*=.11), age (t_{483}=0.57; *P*=.57), education (χ^2_5=5.6; *P*=.35), subjective socioeconomic status (χ^2_4=5.3; *P*=.26), disability type (χ^2_6=9.9; *P*=.13), child rehabilitation status (χ^2_4=8.2; *P*=.08), or depression (t_{483}=–1.20; *P*=.23). However, the attrition group had a higher proportion of men (χ^2_1=7.2; *P*<.01) and higher social support (t_{483}=3.97; *P*<.001). Overall, selective attrition appeared modest: most baseline characteristics did not differ, although men and those with higher social support were more likely to drop out. Missing data were handled using full information maximum likelihood (FIML).

Descriptive statistics and bivariate correlations were computed for key variables. Core analyses used a 3-wave, 2-part latent growth curve model (time coded as 0, 1, and 2) to decompose platform behavior into two components: (1) adoption (any use vs nonuse; categorical probit specification; latent intercept I_B and slope S_B) and (2) engagement (use frequency among users; measured on a 4-level ordinal scale treated as continuous in the primary model; latent intercept I_C and slope S_C). Random intercepts and slopes were allowed to covary within and across components to account for shared determinants and correlated dynamics. Engagement was reestimated as ordinal for robustness, yielding consistent results.

Unconditional models were first estimated to assess mean trajectories and variance components. Covariates (gender, age, education, socioeconomic status, social support, depressive symptoms, and child rehabilitation status) were then added to predict the 4 growth factors. All models used maximum likelihood estimation with robust standard errors in Mplus (version 7.4; Muthén & Muthén). Model fit was evaluated using chi-square statistics, Akaike information criterion (AIC), Bayesian information criterion (BIC), and sample size–adjusted BIC. Robustness checks were performed by reestimating the model with the full baseline sample (n=485) using FIML to handle longitudinal missingness across waves.

Results

Digital Platform Use Among Caregivers Across 3 Waves

Across 3 waves, caregivers reported how often they had used any disability-related digital platform in the past 3 months; nonusers additionally selected one or more reasons for nonuse. Table 1 presents the distributions of use frequency and reasons for nonuse at each wave.

XSL•FO
RenderX

Table . The use of disability-related digital platforms among caregivers.[a]

	T1[b], n (%)	T2[b], n (%)	T3[b], n (%)
Use frequency			
Never	65 (27)	47 (24)	27 (14)
Rarely	23 (10)	16 (8)	19 (10)
Occasionally	56 (23)	44 (22)	54 (28)
Often	58 (24)	66 (33)	65 (35)
Almost daily	38 (16)	26 (13)	26 (13)
Reasons for nonuse			
Limited time or energy	22 (34)	13 (28)	6 (24)
Unaware of platform availability	40 (62)	20 (44)	7 (28)
Unfamiliar with how to use platforms	4 (6)	15 (33)	10 (40)
Concerns about cybersecurity or privacy	6 (9)	10 (22)	1 (4)

[a]Percentages are based on respondents with nonmissing data at each wave. Nonuse reasons were multiple response items.

[b]T1 represents the first measurement, T2 represents the second measurement, and T3 represents the third measurement.

Descriptively, the share of caregivers reporting any use increased over time, reflected in a decline in "Never" from 27% (65/240) to 14% (27/191). The distribution also shifted toward midrange categories ("Occasionally" and "Often"), whereas the proportion reporting "Almost daily" was relatively stable, with a slight decrease from T1 to T3.

Regarding nonuse, lack of awareness of platform availability and limited time or energy were the most frequently cited barriers at T1. Over time, reports of awareness and time or energy barriers declined, whereas unfamiliarity with how to use platforms increased, suggesting a shift from access or awareness constraints toward competence-related barriers. Cybersecurity or privacy concerns peaked at T2 and declined at T3.

Descriptive Statistics and Correlations

Descriptive statistics and bivariate associations among study variables across the 3 waves are summarized in Table 2. Spearman rank correlations were used to accommodate ordinal scaling and potential nonnormality. Platform-use categories exhibited moderate temporal stability, and associations with covariates varied in magnitude and direction.

XSL•FO
RenderX

Table . Descriptive statistics and correlations of study variables.

Variables	FRE1[a,b]	FRE2[c]	FRE3[d]	GEN[e,f]	AGE[g]	EDU[h,i]	SES[j,k]	REC[l,m]	SS[n,o]	DEP[p,q]
FRE 1										
r value	1									
P value	—[r]									
FRE 2										
r value	0.60	1								
P value	<.001	—								
FRE 3										
r value	0.51	0.59	1							
P value	<.001	<.001	—							
GEN										
r value	0.19	0.22	0.21	1						
P value	.003	.001	.004	—						
AGE										
r value	−0.17	−0.22	−0.15	−0.16	1					
P value	.01	.002	.04	.02	—					
EDU										
r value	0.48	0.39	0.39	0.19	−0.29	1				
P value	<.001	<.001	<.001	.003	<.001	—				
SES										
r value	0.30	0.32	0.22	0.18	−0.19	0.34	1			
P value	<.001	<.001	.002	.005	.003	<.001	—			
REC										
r value	0.12	0.01	0.01	0.04	0.02	0.12	0.14	1		
P value	.06	.84	.87	.58	.71	.08	.03	—		
SS										
r value	0.08	0.28	0.14	0.20	−0.06	−0.03	0.16	0.07	1	
P value	.21	<.001	.05	.002	.39	.68	.01	.26	—	
DEP										
r value	0.01	0.04	−0.04	−0.02	0.06	−0.07	−0.19	0.01	−0.22	1
P value	.82	.59	.56	.78	.36	.32	.004	.93	.001	—

[a]FRE: frequency use (T1-T3).
[b]Mean 1.92 (SD 1.43).
[c]Mean 2.04 (SD 1.37).
[d]Mean 2.23 (SD 1.21).
[e]GEN: gender (0=female, 1=male).
[f]Mean 0.29 (SD 0.46).
[g]Mean 37.29 (SD 6.85).
[h]EDU: education level.
[i]Mean 3.48 (SD 1.45).
[j]SES: socioeconomic status.
[k]Mean 2.55 (SD 0.95).
[l]REC: child rehabilitation status.
[m]Mean 2.97 (SD 0.80).
[n]SS: social support.

XSL•FO
RenderX

[o]Mean 59.65 (SD 13.60).

[p]DEP: depression.

[q]Mean 10.56 (SD 6.71).

[r]Not applicable.

Two-Part Latent Growth Analysis of Digital Platform Use and Predictors

To distinguish between adoption and engagement intensity, platform use was modeled using a 2-part latent growth curve model. The binary component represented use status at each wave (b1-b3), while the continuous component captured use frequency among users (c1-c3). Unconditional growth models were first estimated separately for each part and then combined in a parallel 2-part model. In preliminary analyses, allowing the adoption slope variance (S_B) to be freely estimated resulted in numerical instability and nonconvergence, indicating insufficient information to reliably estimate between-person variability in change in the 3-wave binary adoption process. Therefore, in the primary models, we fixed Var(S_B)=0, retained a random intercept for adoption, and estimated the mean adoption slope to characterize the average change in adoption over time. The unconditional 2-part model showed acceptable fit (Pearson χ^2_4=2.2, *P*=.69; likelihood ratio χ^2_4=2.4, *P*=.67; AIC=1805.95; BIC=1851.19; adjusted BIC=1809.99).

We then estimated a conditional 2-part model incorporating gender, age, education, socioeconomic status, child rehabilitation status, social support, and caregiver depressive symptoms. Compared with the unconditional model, the conditional model showed lower AIC (1736.54 vs 1805.95) and sample size–adjusted BIC (1747.11 vs 1809.99), although BIC increased slightly (1854.88 vs 1851.19). Given the a priori inclusion of covariates and overall model performance, we retained the conditional model for inference.

In the binary part, the mean slope was positive (S_B=0.54, SE 0.16; *P*=.001), indicating increasing probability of adoption over time; the intercept variance was significant (Var(I_B)=2.76, SE 0.82; *P*=.001), evidencing substantial heterogeneity at baseline. In the continuous part, the mean intercept indicated relatively high initial frequency among users with significant between-person variability (I_C=2.52, SE 0.10; *P*<.001; and Var(I_C)=0.58, SE 0.14; *P*<.001). The mean slope was not significant (S_C=−0.07, SE 0.07; *P*=.31), but its variance approached significance (Var(S_C)=0.11, SE 0.06; *P*=.07), indicating variability in individual change rates. A strong negative correlation between the intercept and slope (*r*=−0.72; *P*<.001) suggested that users with higher initial use frequency tended to show flatter or declining trajectories over time.

Covariate effects indicated that education was a robust positive correlate of baseline adoption, with additional positive associations for socioeconomic status, social support, depressive symptoms, and male gender. For initial use frequency among users, education and social support were positive predictors, whereas other effects were small and nonsignificant. Regarding change in frequency, better child rehabilitation status predicted steeper declines over time. Parameter estimates are presented in Tables 3 and 4.

Table . Estimated means and variances of growth factors in the 2-part latent growth curve model.[a]

Latent factor	Growth factors, mean (SE)	*P* value	Growth factors, variance (SE)	*P* value
Use status intercept (I_B)	N/A[b]	N/A	2.76 (0.82)	.001
Use status slope (S_B)	0.54 (0.16)	.001	0 (fixed)	N/A
Use frequency intercept (I_C)	2.52 (0.10)	<.001	0.58 (0.14)	<.001
Use frequency slope (S_C)	−0.07 (0.07)	.31	0.11 (0.06)	.07

[a]The variance of S_B was fixed to zero due to numerical instability and nonconvergence when freely estimated. The mean of I_B is not separately estimated under the categorical (threshold or latent response, probit) parameterization.

[b]N/A: not applicable.

XSL•FO
RenderX

Table . Predictors of adoption intercept and engagement growth factors.[a]

Predictor	Adoption intercept I_B, β (SE)	*P* value	Engagement intercept I_C, β (SE)	*P* value	Engagement slope S_C, β (SE)	*P* value
Gender	0.19 (0.08)	.02	−0.04 (0.08)	.68	0.12 (0.12)	.34
Age (z-score)	−0.11 (0.07)	.11	0.00 (0.08)	.96	−0.12 (0.14)	.37
Education (z)	0.42 (0.08)	<.001	0.36 (0.10)	<.001	0.04 (0.15)	.78
Socioeconomic status (z)	0.26 (0.08)	.001	0.11 (0.09)	.21	−0.05 (0.14)	.69
Rehabilitation (z)	−0.08 (0.09)	.37	0.13 (0.08)	.10	−0.28 (0.13)	.03
Social support (z)	0.17 (0.08)	.03	0.20 (0.10)	.04	−0.09 (0.15)	.55
Depression (z)	0.18 (0.08)	.02	0.08 (0.08)	.31	−0.02 (0.12)	.89

[a]Standardized coefficients (β) with SEs and *P* values. IB uses a categorical probit specification. SB was not regressed on covariates because Var(S_B) was fixed to 0.

As a sensitivity analysis, we reestimated the 2-part model with the full baseline sample (n=485) using FIML to handle longitudinal missingness across waves. Results were largely consistent with the primary analysis, supporting the robustness of the main conclusions. Caregiver age became a significant predictor of baseline adoption in the full baseline sample, whereas the rehabilitation–engagement growth association was attenuated to a marginal level (directionally consistent). Detailed estimates are provided in Tables S2 and S3 in Multimedia Appendices 2 and 3, respectively.

Discussion

This study used a 2-part latent growth curve model to explore the dynamics and potential predictors of digital support platform use among family caregivers of children with disabilities. Findings indicated that platform adoption increased over time, while average engagement remained stable. Distinct predictors for adoption and engagement may reflect differing processes underlying initial uptake vs sustained participation.

Status and Trajectory of Digital Platform Use

At the adoption level, the positive mean slope in the binary part suggests an increasing probability of platform use over time, with adoption becoming more common. This increase may reflect the growing availability of online resources and broader shifts toward digital help seeking [37]. Notably, the slope variance (S_B) was not stably estimable and was therefore fixed to zero. Consequently, between-person variability in adoption change over time was not modeled, which may underestimate heterogeneity in adoption trajectories.

A different pattern emerged for engagement frequency. Baseline engagement frequency (I_C) showed significant between-person variability, while the mean slope (S_C) was near zero, indicating stable average engagement at the group level. Evidence for variability in S_C was marginal, suggesting some heterogeneity in individual trajectories. The negative correlation between I_C and S_C suggests differing trends by initial engagement level, possibly reflecting needs-based adjustment, regression to the mean, or ceiling effects.

Factors Associated With Digital Support Platform Use Among Caregivers

Guided by the Andersen BMHSU [14], we examined predisposing (gender, age, and education), enabling (subjective socioeconomic status and social support), and need-related (child rehabilitation status and depressive symptoms) factors.

For adoption, education emerged as a robust positive correlate, consistent with evidence linking higher educational attainment to digital literacy and earlier uptake [20]. Subjective socioeconomic status was also positively associated with adoption, aligning with research highlighting the role of digital capital and reduced barriers among higher-status groups [26,38]. Social support was another significant predictor, indicating that richer offline networks may facilitate online navigation [25]. Depressive symptoms were positively associated with adoption, supporting the notion of need-driven help-seeking behavior [29].

Interestingly, in this sample, male caregivers had a higher baseline probability of adoption, contrasting with prior studies that often report greater digital use among females [39]. However, this finding may reflect study-specific factors such as caregiver demographics, selection bias, or attrition and should not be overinterpreted as evidence of a stable gender difference. Instead, it highlights the context-dependent nature of gender effects on platform use, which may vary across settings.

For engagement frequency, only education and social support predicted baseline levels, suggesting that engagement is more closely tied to competencies and social resources than to other sociodemographic characteristics. Changes in engagement frequency were most consistently related to child rehabilitation status: better rehabilitation outcomes tended to coincide with steeper declines in use, consistent with Andersen's premise that lower perceived need is linked to reduced use [14,16]. However, this association was attenuated to a marginal level in the full baseline sample, and alternative explanations (eg, unmet evolving needs among families with improving outcomes) should be considered.

Overall, the findings partially support Andersen's model for understanding caregivers' digital platform use. At the same time, digital health use often unfolds in 2 stages—adoption and

XSL•FO
RenderX

engagement—and may also depend on technology-specific mechanisms. The technology acceptance model (TAM) and unified theory of acceptance and use of technology (UTAUT) emphasize perceived usefulness, ease of use, and facilitating conditions (eg, skills, support, and resources) [40,41], while privacy-trust research highlights privacy risk appraisal and trust [42]. Although we did not directly measure these constructs, descriptively reported barriers changed over time: "lack of awareness" declined from T1 to T3, "not knowing how to use" increased, and privacy concerns peaked at T2 but declined by T3. These patterns suggest that integrating TAM or UTAUT and privacy-trust constructs with Andersen's framework may better capture the dynamics of adoption and sustained engagement in digital health contexts.

Practical Implications

On the basis of these findings, several practical implications emerge. To support adoption and reduce inequities, interventions may prioritize caregivers with lower educational attainment, socioeconomic status, and social support. Simplified onboarding delivered through community networks, schools, and rehabilitation centers may lower entry barriers, and brief digital literacy training could address capability gaps. Targeted awareness campaigns may further promote adoption. Because adoption was higher among caregivers reporting more depressive symptoms, onboarding should also include clear pathways to emotional and psychological support, including referral options.

To support sustained engagement, digital health platforms should be adaptive, user-friendly, and trustworthy. Given stable average engagement but heterogeneous individual trajectories, personalization can adjust content, peer matching, and support intensity as caregiver needs change [43,44]. AI-based guidance may help reduce information overload and improve perceived usefulness, but implementation requires appropriate governance and safeguards [45,46]. Usability can be strengthened through caregiver-centered interface design [47], and privacy protections with transparent communication are important for maintaining trust [48].

Finally, integrating digital platforms into existing support systems may enhance reach and sustainability. Coordinated efforts across government agencies, social welfare organizations, rehabilitation centers, and special education schools can support promotion, onboarding, linkage to services, and ongoing feedback so that platforms evolve with caregivers' needs.

Limitations

This study has several limitations. First, platform use was measured with a study-specific self-report instrument informed by caregiver interviews and expert review; however, the instrument has not undergone formal psychometric validation, including evaluation of construct validity, criterion validity, and reliability. Engagement was also operationalized only as use frequency, without capturing other dimensions such as session intensity, interaction quality, or emotional engagement. Consequently, qualitatively different use patterns (eg, frequent passive browsing vs infrequent but high-quality interaction) may have been conflated, limiting interpretation of engagement trajectories.

Second, although multicenter recruitment may reduce regional bias, convenience sampling may have favored caregivers with higher needs, limiting representativeness. Differential attrition (with higher dropout among male participants and those with greater baseline social support) may further bias estimates. Attrition also reduced sample size and power for detecting longitudinal effects, so null findings should be interpreted cautiously.

Third, because participants were recruited primarily in China, generalizability to other contexts is uncertain. Cross-country differences in caregiving systems, digital literacy, internet access, and health care infrastructure may shape adoption and engagement.

Finally, although the longitudinal design supports temporal ordering, the study remains observational and measurement reactivity cannot be ruled out; completing baseline assessments may have influenced subsequent platform use. Future studies could lessen potential priming from surveys by shortening assessments and, where possible, supplementing self-report with platform use logs.

Conclusions

This study used a 2-part latent growth curve model to examine digital platform use among caregivers of children with disabilities in China. Adoption increased over time and was positively associated with education, gender, subjective socioeconomic status, social support, and depressive symptoms. Average engagement remained stable, with baseline engagement related to education and social support. Changes in engagement were most consistently associated with child rehabilitation status. Overall, these findings underscore the value of distinguishing adoption from engagement and suggest that caregiver resources and needs are related to patterns of digital platform use.

Acknowledgments

The authors used ChatGPT-4 (OpenAI) exclusively for English language polishing and formatting refinement. All revisions were carefully reviewed by the authors to ensure that the original meaning and scientific content were fully preserved.

Funding

This research was supported by the 2024 Humanities and Social Sciences Research Projects of the Ministry of Education, China (grant 24YJC840045)

XSL•FO
RenderX

Data Availability

The datasets used in this study are available from the corresponding author on reasonable request. Due to ethical and privacy considerations, and because the dataset includes sensitive information about children with disabilities and their families, the data are not publicly available.

Authors' Contributions

Conceptualization: XY, YS
Formal analysis: XY, XZ
Investigation: XY, XZ, YS
Writing – original draft: XY, XZ
Writing – review & editing: MZ, YD, YS

Conflicts of Interest

None declared.

Multimedia Appendix 1
Demographic characteristics of participants.
[DOCX File, 20 KB - mhealth_v14i1e92481_app1.docx]

Multimedia Appendix 2
Estimated means and variances of growth factors in the 2-part latent growth curve model in the full baseline sample.
[DOCX File, 18 KB - mhealth_v14i1e92481_app2.docx]

Multimedia Appendix 3
Predictors of adoption intercept and engagement growth factors in the full baseline sample.
[DOCX File, 16 KB - mhealth_v14i1e92481_app3.docx]

References

1. Rutter TL, Hastings RP, Murray CA, Enoch N, Johnson S, Stinton C. Psychological wellbeing in parents of children with Down syndrome: a systematic review and meta-analysis. Clin Psychol Rev 2024 Jun;110:102426. [doi: 10.1016/j.cpr.2024.102426] [Medline: 38652972]
2. Chen SQ, Chen SD, Li XK, Ren J. Mental health of parents of special needs children in China during the COVID-19 pandemic. Int J Environ Res Public Health 2020 Dec 18;17(24):9519. [doi: 10.3390/ijerph17249519] [Medline: 33353165]
3. van Niekerk K, Stancheva V, Smith C. Caregiver burden among caregivers of children with autism spectrum disorder. S Afr J Psychiatr 2023;29:2079. [doi: 10.4102/sajpsychiatry.v29i0.2079] [Medline: 37928940]
4. Gruebner O, van Haasteren A, Hug A, et al. Digital platform uses for help and support seeking of parents with children affected by disabilities: scoping review. J Med Internet Res 2022 Dec 6;24(12):e37972. [doi: 10.2196/37972] [Medline: 36472896]
5. Napitupulu L, Kurniawan Y. Online social support and psychological well-being of caregivers of children with autism spectrum disorder. Eur J Educ Res 2023;12(2):927-946. [doi: 10.12973/eu-jer.12.2.927]
6. Alsem MW, Ausems F, Verhoef M, Jongmans MJ, Meily-Visser JM, Ketelaar M. Information seeking by parents of children with physical disabilities: an exploratory qualitative study. Res Dev Disabil 2017 Jan;60:125-134. [doi: 10.1016/j.ridd.2016.11.015] [Medline: 27914304]
7. Khasawneh M. The influence of Twitter hashtags in building supportive online communities for parents of children with special needs. J Technol Sci Educ 2024;14(3):770. [doi: 10.3926/jotse.2537]
8. Chen L, Wang Z, Yang G. Online social support for parents of children with autism [Article in Chinese]. Chin Ment Health J 2023;37(4):299. [doi: 10.3969/j.issn.1000-6729.2023.04.005]
9. Daynes-Kearney R, Gallagher S. Online support groups for family caregivers: a qualitative exploration of social support and engagement. Br J Health Psychol 2025 Feb;30(1):e12764. [doi: 10.1111/bjhp.12764] [Medline: 39511913]
10. Fante C, Zagaria A, Dioni B, et al. Self-efficacy as a mediator between involvement in intervention and quality of life in parents of children and adolescents with autism spectrum disorder. Res Autism Spectr Disord 2024 May;113:102351. [doi: 10.1016/j.rasd.2024.102351]
11. Montoro-Rodriguez J, Ramsey J, Bilbrey AC, Kajiyama B, Thompson LW, Gallagher-Thompson D. Caregiver thrive, learn, & connect: testing the efficacy of an online psychoeducational program for family caregivers. Clin Gerontol 2024;47(1):39-49. [doi: 10.1080/07317115.2023.2232352] [Medline: 37416945]

XSL•FO
RenderX

12. Chakraborti M, Gitimoghaddam M, McKellin WH, Miller AR, Collet JP. Understanding the implications of peer support for families of children with neurodevelopmental and intellectual disabilities: a scoping review. Front Public Health 2021;9:719640. [doi: 10.3389/fpubh.2021.719640] [Medline: 34888278]
13. Dembo RS, Huntington N, Mitra M, Rudolph AE, Lachman ME, Mailick MR. Social network typology and health among parents of children with developmental disabilities: results from a national study of midlife adults. Soc Sci Med 2022 Jan;292:114623. [doi: 10.1016/j.socscimed.2021.114623] [Medline: 34891030]
14. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? J Health Soc Behav 1995 Mar;36(1):1-10. [Medline: 7738325]
15. Alturif G, Saleh W, Alsanad H, Ababio-Donkor A. Enhancing inclusive social, financial, and health services for persons with disabilities in Saudi Arabia: insights from caregivers. Healthcare (Basel) 2025 Aug 5;13(15):1901. [doi: 10.3390/healthcare13151901] [Medline: 40805934]
16. Andersen JA, Felix HC, Su D, Selig JP, Ratcliff S, McElfish PA. Factors associated with Arkansans' first use of telehealth during the COVID-19 pandemic. Int J Telemed Appl 2022;2022:5953027. [doi: 10.1155/2022/5953027] [Medline: 35775023]
17. Babitsch B, Gohl D, von Lengerke T. Re-revisiting Andersen's Behavioral Model of Health Services Use: a systematic review of studies from 1998-2011. Psychosoc Med 2012;9:Doc11. [doi: 10.3205/psm000089] [Medline: 23133505]
18. DeHoff BA, Staten LK, Rodgers RC, Denne SC. The role of online social support in supporting and educating parents of young children with special health care needs in the United States: a scoping review. J Med Internet Res 2016 Dec 22;18(12):e333. [doi: 10.2196/jmir.6722] [Medline: 28007689]
19. Kellom KS, Flaherty CM, Cacia J, et al. Provider and caregiver perspectives on telehealth assessments for autism spectrum disorder in young children: a multimethod DBPNet study exploring equity. J Dev Behav Pediatr 2023 Aug 1;44(6):e397-e411. [doi: 10.1097/DBP.0000000000001198] [Medline: 37315107]
20. Knapp C, Madden V, Wang H, Sloyer P, Shenkman E. Internet use and eHealth literacy of low-income parents whose children have special health care needs. J Med Internet Res 2011 Sep 29;13(3):e75. [doi: 10.2196/jmir.1697] [Medline: 21960017]
21. Sarkar M, Sanders LM, Kelleher KJ, Chisolm DJ. Psychosocial health, e-health literacy, and perceptions of e-health as predictors and moderators of e-health use among caregivers of children with special healthcare needs. Telemed J E Health 2016 Feb;22(2):123-131. [doi: 10.1089/tmj.2015.0028] [Medline: 26295487]
22. Zhang S, Hao Y, Feng Y, Lee NY. COVID-19 pandemic impacts on children with developmental disabilities: service disruption, transition to telehealth, and child wellbeing. Int J Environ Res Public Health 2022;19(6):3259. [doi: 10.3390/ijerph19063259] [Medline: 35328947]
23. Hermaszewska S, Sin J. End-user perspectives on the development of an online intervention for parents of children on the autism spectrum. Autism 2021 Jul;25(5):1234-1245. [doi: 10.1177/1362361320984895] [Medline: 33423522]
24. Rabba AS, Dissanayake C, Barbaro J. Development of a web-based resource for parents of young children newly diagnosed with autism: participatory research design. JMIR Pediatr Parent 2020 Sep 30;3(2):e15786. [doi: 10.2196/15786] [Medline: 32996890]
25. Titgemeyer SC, Schaaf CP. Facebook support groups for pediatric rare diseases: cross-sectional study to investigate opportunities, limitations, and privacy concerns. JMIR Pediatr Parent 2022 Jan 6;5(1):e31411. [doi: 10.2196/31411] [Medline: 34989690]
26. Rahman MI, Alam J, Khanom K, Emdad FB. Social determinants influencing internet-based service adoption among female family caregivers in Bangladesh: a sociodemographic and technological analysis. Health Sci Rep 2025 Apr;8(4):e70665. [doi: 10.1002/hsr2.70665] [Medline: 40242265]
27. Ao SH, Zhang L, Liu PL, Zhao X. Social media and partnership jointly alleviate caregivers' psychological distress: exploring the effects of online and offline connectedness. BMC Psychol 2023 Nov 14;11(1):394. [doi: 10.1186/s40359-023-01415-9] [Medline: 37964344]
28. Friedman EM, Trail TE, Vaughan CA, Tanielian T. Online peer support groups for family caregivers: are they reaching the caregivers with the greatest needs? J Am Med Inform Assoc 2018 Sep 1;25(9):1130-1136. [doi: 10.1093/jamia/ocy086] [Medline: 30016449]
29. Bizzotto N, Marciano L, de Bruijn GJ, Schulz PJ. The empowering role of web-based help seeking on depressive symptoms: systematic review and meta-analysis. J Med Internet Res 2023 Feb 2;25:e36964. [doi: 10.2196/36964] [Medline: 36729571]
30. Roche MI, Skinner D. How parents search, interpret, and evaluate genetic information obtained from the internet. J Genet Couns 2009 Apr;18(2):119-129. [doi: 10.1007/s10897-008-9198-4] [Medline: 18937062]
31. Olsen MK, Schafer JL. A two-part random-effects model for semicontinuous longitudinal data. J Am Stat Assoc 2001 Jun;96(454):730-745. [doi: 10.1198/016214501753168389]
32. Tooze JA, Grunwald GK, Jones RH. Analysis of repeated measures data with clumping at zero. Stat Methods Med Res 2002 Aug;11(4):341-355. [doi: 10.1191/0962280202sm291ra] [Medline: 12197301]
33. Wang XD, Wang XL, Ma H, editors. Handbook of Mental Health Assessment Scales: Chinese Mental Health Journal Press; 1999.
34. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

XSL•FO
RenderX

35. Wang W, Bian Q, Zhao Y, et al. Reliability and validity of the Chinese version of the Patient Health Questionnaire (PHQ-9) in the general population. Gen Hosp Psychiatry 2014;36(5):539-544. [doi: 10.1016/j.genhosppsych.2014.05.021] [Medline: 25023953]
36. Xia C, Wei T, Tang Q, Zheng H, Chen G, Lu J. Depression and associated factors among family caregivers of children with disabilities: analysis of intergenerational differences. Healthcare (Basel) 2023;11(19):2693. [doi: 10.3390/healthcare11192693] [Medline: 37830730]
37. Jia X, Pang Y, Liu LS. Online health information seeking behavior: a systematic review. Healthcare (Basel) 2021 Dec 16;9(12):1740. [doi: 10.3390/healthcare9121740] [Medline: 34946466]
38. Ramirez AV, Ojeaga M, Espinoza V, Hensler B, Honrubia V. Telemedicine in minority and socioeconomically disadvantaged communities amidst COVID-19 pandemic. Otolaryngol Head Neck Surg 2021 Jan;164(1):91-92. [doi: 10.1177/0194599820947667] [Medline: 32720844]
39. Hassan AY, Bronzini M, Lamura G. Digital technologies as sources of information for patients and caregivers during COVID-19 pandemic: a cross-sectional survey. Digit Health 2023;9:20552076231156214. [doi: 10.1177/20552076231156214] [Medline: 36908378]
40. Kim JH, Kim J, Youn BY. Using a technology acceptance model to explore the intention to use digital health technologies among people with disabilities: cross-sectional survey study. J Med Internet Res 2025 Nov 20;27:e79595. [doi: 10.2196/79595] [Medline: 41264861]
41. Yang HJ, Lee JH, Lee W. Factors influencing health care technology acceptance in older adults based on the technology acceptance model and the unified theory of acceptance and use of technology: meta-analysis. J Med Internet Res 2025 Mar 28;27:e65269. [doi: 10.2196/65269] [Medline: 40153796]
42. Shojaei P, Vlahu-Gjorgievska E, Chow YW. Enhancing privacy in mHealth applications: a user-centric model identifying key factors influencing privacy-related behaviours. Int J Med Inform 2025 Jul;199:105907. [doi: 10.1016/j.ijmedinf.2025.105907] [Medline: 40209320]
43. Shi JM, Yoo DW, Wang K, Rodriguez VJ, Karkar R, Saha K. Mapping caregiver needs to AI chatbot design: strengths and gaps in mental health support for Alzheimer's and dementia caregivers. ACM Trans Comput Healthcare 2026. [doi: 10.1145/3803549]
44. Premanandan S, Ahmad A, Cajander Å, Ågerfalk P, van Gemert-Pijnen L. Designing and evaluating IT applications for informal caregivers: scoping review. J Med Internet Res 2024 Oct 23;26:e57393. [doi: 10.2196/57393] [Medline: 39442167]
45. Fan X, Chao D, Zhang Z, Wang D, Li X, Tian F. Utilization of self-diagnosis health chatbots in real-world settings: case study. J Med Internet Res 2021 Jan 6;23(1):e19928. [doi: 10.2196/19928] [Medline: 33404508]
46. Siira E, Johansson H, Nygren J. Mapping and summarizing the research on AI systems for automating medical history taking and triage: scoping review. J Med Internet Res 2025 Feb 6;27:e53741. [doi: 10.2196/53741] [Medline: 39913918]
47. Costa Stutzel M, Filippo MP, Sztajnberg A, et al. Multi-part quality evaluation of a customized mobile application for monitoring elderly patients with functional loss and helping caregivers. BMC Med Inform Decis Mak 2019 Jul 22;19(1):140. [doi: 10.1186/s12911-019-0839-3] [Medline: 31331309]
48. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. JMIR Mhealth Uhealth 2019 Apr 16;7(4):e11223. [doi: 10.2196/11223] [Medline: 30990458]

Abbreviations

AIC: Akaike information criterion
BIC: Bayesian information criterion
BMHSU: behavioral model of health services use
FIML: full information maximum likelihood
TAM: technology acceptance model
UTAUT: unified theory of acceptance and use of technology

Edited by SD Bhattacharjee; submitted 30.Jan.2026; peer-reviewed by I Syarif, TD Mestre; revised version received 11.Jul.2026; accepted 31.Aug.2026; published 23.Sep.2026.

Please cite as:
Yuan X, Zhao X, Zhang M, Duan Y, Shui Y
Dynamics of Digital Support Platform Use Among Family Caregivers of Children With Disabilities: Longitudinal Study
JMIR Mhealth Uhealth 2026;14:e92481
URL: https://mhealth.jmir.org/2026/1/e92481
doi:10.2196/92481

XSL•FO
RenderX

© Xiaojiao Yuan, Xi Zhao, Meng Zhang, Yurou Duan, Yizhong Shui. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 23.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

User-Reported Issues With Mental Health Apps: Machine-Assisted Topic Analysis of Social Media Posts

Jack Bolter[1], MSc; Trisevgeni Papakonstantinou[2], MSc; Paulina Bondaronek[3], PhD

[1]Department of Psychology, School of Social Sciences, University of Westminster, London, United Kingdom

[2]Department of Computer Science, University College London, London, United Kingdom

[3]Institute of Health Informatics, University College London, 222 Euston Road, London, United Kingdom

Corresponding Author:
Paulina Bondaronek, PhD
Institute of Health Informatics, University College London, 222 Euston Road, London, United Kingdom

Abstract

Background: Mobile apps marketed to support mental health have become increasingly popular in recent years. Given their widespread use, it is important to identify issues that users experience while using such apps. Understanding these issues may provide insight into the safety and suitability of these apps for individuals seeking mental health support.

Objective: Unlike existing research, where user experience issues have been identified through researchers' direct analysis of apps, this study aimed to generate themes relating to user experience issues using comments from app users themselves. An additional aim was to evaluate a human-in-the-loop machine learning approach using structural topic modeling (STM) to analyze vast volumes of data gathered from X (formerly Twitter, developed by Twitter, Inc).

Methods: Data relating to five of the most popular mental health apps were collected from the X API using R. A machine-assisted thematic analysis approach combined STM with human qualitative analysis to interpret user-generated posts. An unsupervised topic-modeling approach was tested using models with 5-40 topics and differing covariates (ultimately, a model without covariates was selected). Two researchers independently conducted thematic analysis to interpret and contextualize model outputs. A structural topic model with 10 topics, each comprising 20 X posts, was selected as most appropriate for generating insights.

Results: Using R (developed by the R Core Team), 79,703 X posts were collected via the X API relating to five popular mental health apps. After negative sentiment filtering, 19,603 posts remained. Posts spanned March 2006 (the launch of X/formerly Twitter) to December 2022. Researchers collaboratively labeled the 10 topics to identify the primary user experience issue represented in each. Topic 3 was discarded due to low coherence and inconsistency in relation to app user experience, and Topic 5 was discarded because posts reflected app X account activity rather than user experience of the apps. The remaining eight topics were organized into four themes. The first theme, guidance shortfall, included difficulties following guided meditations, challenges selecting appropriate content from large libraries, and incompatibility between app use and home environments. The second theme, technical difficulties, involved subscription access issues and technical faults within apps. The third theme, heightened emotions related to app-affiliated celebrities, captured both over-excitement linked to celebrity involvement and anger directed toward specific celebrities. The final standalone theme, negative impacts of sleep self-monitoring, demonstrated users reporting that tracking sleep adversely affected sleep experience.

Conclusions: The combination of STM and human qualitative analysis of X posts identified several user-experienced issues associated with popular mental health apps, often linked to negative outcomes. This study provides evidence that STM can be combined with qualitative methods to rapidly analyze large-scale social media data and generate insights into user experience of mass-reach digital health interventions.

(JMIR Mhealth Uhealth 2026;14:e85575) doi:10.2196/85575

KEYWORDS

health behavior; mobile apps; health promotion; mHealth; human-computer interaction; mental health apps; user experiences; digital mental health; machine learning; qualitative data; Twitter; application programming interface; API; mobile health

Introduction

In England, 1 in 4 people experience mental health issues annually, with 1 in 6 reporting common problems such as anxiety or depression on a weekly basis [1,2]. Depression is the leading global cause of disability [3], and severe mental illness can reduce life expectancy by up to 20 years due to preventable physical conditions [4]. People with poor mental health may feel down, irritable, hopeless, and exhibit behavior changes such as social withdrawal, self-harm, sleep loss, increased use of drugs and alcohol, and altered eating habits [5]. Only

XSL•FO
RenderX

one-third of adults in England and Wales with common mental health conditions receive treatment [2,6], despite the World Health Organization noting globally low treatment rates [4]. Barriers to seeking help include limited knowledge about mental illness, unawareness of treatment options, and fear of stigma [7].

Given the low uptake of formal support, individuals often turn to digital alternatives. The mental health app market, valued at $5.2 billion in 2022, is projected to grow at 18% annually until 2030 [8]. Popular apps such as Calm (developed by Calm.com, Inc), Headspace (developed by Headspace, Inc), and Sleep Cycle (developed by Sleep Cycle AB) offer mindfulness, relaxation, and sleep-related features, with Calm alone exceeding 107 million downloads between 2018 and 2022 [9]. However, concerns remain regarding app effectiveness, as many apps lack empirical support, involve few health care professionals in development, and provide limited or inaccurate tracking of mental health metrics [10-15]. The unregulated nature of this market, coupled with the prioritization of venture capital-driven growth, may contribute to these shortcomings [16-18].

This study is situated within the broader landscape of mobile health (mHealth) apps, a heterogeneous category encompassing digital tools designed to support health and well-being through mobile devices [19]. Within this ecosystem, mental health apps represent a diverse subset that spans multiple functional domains, including mindfulness and meditation, sleep improvement, mood regulation, psychoeducation, and passive or active self-tracking [20]. The five apps analyzed in this study—Calm, Headspace, BetterSleep (developed by Ipnos Software Inc), Sleep Cycle, and Waking Up (developed by Waking Up, LLC)—were selected based on their status as the top-grossing mental health apps in December 2022 [9], reflecting their commercial reach and user uptake rather than functional equivalence. While Calm, Headspace, and Waking Up all foreground guided meditation and mindfulness training as core components of user engagement, they differ notably in content and delivery approaches. Headspace adopts a structured, practical approach designed to make mindfulness accessible and habit-oriented, whereas Waking Up places greater emphasis on deeper philosophical discussions and the scientific underpinnings of consciousness [21,22]. Calm, by contrast, frames mindfulness within a broader relaxation and wellness context that incorporates sleep stories and ambient soundscapes alongside meditation guidance [23]. BetterSleep occupies a hybrid position, combining relaxation, soundscapes, and sleep-focused mindfulness with elements of habit formation, while Sleep Cycle is distinct in its emphasis on sleep tracking, biometric data, and algorithmic analysis, positioning mental well-being primarily through sleep optimization rather than direct psychological intervention [24,25]. These apps therefore differ substantially in their target users, behavioral change strategies, and technological affordances, ranging from content-driven experiential engagement to data-driven self-monitoring. A comparative analysis of these apps enables examination of how varied mHealth design logics, commercial priorities, and conceptualizations of mental health are operationalized within widely adopted platforms, providing insight into the broader mental health app market rather than a single therapeutic modality.

Social media offers a rich and ecologically valid data source for evaluating user experiences of apps. X (formerly Twitter), which allows third-party data access via its API, supports large-scale qualitative research by providing real-world, unsolicited feedback [26,27]. X's arrival in 2006 predates that of any mental health apps by at least 2 years [28-30]. With around 238 million global monthly active users, the real-time and conversational nature of X results in over 80% of the platform's advertisers' inbound customer service requests arriving via the platform [31,32]. Unlike Facebook (developed by Meta Platforms, Inc) and Instagram (developed by Meta Platforms, Inc), which focus on connecting friends and sharing photos, X emphasizes text-based posts [33]. Given its functionality and comprehensive data archive, X was chosen as the data source for this study.

Natural language processing (NLP) has become an increasingly important set of methods in health research for extracting meaningful insights from unstructured text, supporting both clinical and patient-facing applications. NLP techniques have been applied across health care to analyze clinical notes, extract structured information from narrative records, and support decision-making processes, highlighting the role of text analytics in structuring and interpreting large volumes of health-related data [34]. In addition to clinical contexts, NLP has been used to analyze real-world patient and consumer language, including extracting health insights and patterns from free-text patient feedback and social media, demonstrating its utility in capturing perceptions, attitudes, and experiences related to health services and outcomes [35,36]. Recent work has shown how topic modeling, sentiment analysis, and related language processing techniques can complement traditional qualitative analyses by summarizing themes and sentiment patterns across large text datasets [35]. These applications provide context for this study's use of NLP methods to analyze user-generated commentary on mental health apps.

Machine-assisted topic analysis (MATA) combines machine learning techniques with human-led qualitative analysis to enable systematic exploration of large-scale textual datasets. In this approach, free-text data are first structured using structural topic modeling (STM), making them analytically tractable while preserving the richness of user-generated content. MATA keeps human judgment central to topic labeling, thematic synthesis, and theoretical interpretation, rather than relying solely on automated classification. Building on established MATA applications in health research [37,38], this study extends the framework to the analysis of user experiences of commercial mental health apps expressed on social media. Methodologically, this implementation demonstrates how MATA can surface and organize diverse user concerns at scale—spanning technical usability, content and operational challenges, commercial practices, and self-monitoring experiences—that would be difficult to capture through traditional thematic analysis alone due to dataset size, or through sentiment analysis alone due to its limited contextual sensitivity.

XSL•FO
RenderX

In this study, STM was used to organize user feedback on mental health apps by identifying patterns of word co-occurrence while incorporating relevant metadata. Human-led thematic analysis was subsequently applied to interpret and cluster the resulting topics [37,39-44].

Focusing on user experiences rather than researcher-led evaluations, this study targeted the 5 high-grossing mental health apps in December 2022: Calm, Headspace, BetterSleep, Sleep Cycle, and Waking Up [9]. These apps were selected due to their widespread use, financial success, and potential to generate relevant, representative data. When searching for mental health apps, health or meditation app criteria were applied, while fitness, weight control, and sports apps were excluded. Table 1 briefly summarizes the 5 apps and their individual global revenue totals for December 2022.

Table . The top 5 grossing mental health apps in December 2022, their monthly revenue, and brief summaries.

App name	Approximate gross global revenue in the month of December 2022	Focus of app
Calm	6,400,000	Mindfulness, relaxation, and sleep
Headspace	4,900,000	Mindfulness, meditation, and sleep
Sleep Cycle	1,400,000	Sleep tracking and gentle wake-up alarm clock
BetterSleep	1,300,000	Sleep tracking, education, and relaxation content
Waking Up	952,000	Meditation, mindfulness, and educational content
Total	14,952,000	—

The above table shows the top 5 grossing mental health apps in December 2022, their monthly revenue, and brief summaries of each app [9,45-49].

The aim of this study was to identify issues users experience when using commercial mental health apps by analyzing a large volume of user-generated content from X. Specific objectives were: (1) to use MATA to analyze user feedback and (2) to provide insights on the potential negative impacts of highly popular commercial mental health apps.

Methods

Study Design

MATA combines topic modeling with human qualitative analysis [37,42-44]. In this study, MATA was used to analyze X posts, identifying topics and generating representative quotes. Qualitative researchers then interpreted the outputs to synthesize the findings.

Data

X posts referencing the 5 highest-grossing health apps were collected using X's official API, accessed via an academic license and implemented using version 0.3.1 of the R package "*academictwitteR*" [50] (refer to Table 1). The search strategy involved using keywords combining each app's name with terms such as "app," "application," and "tracker," where relevant. Additionally, X posts mentioning the official app accounts' X handles were included in the search. A document showing the coding and keywords used is available in Multimedia Appendix 1. The date range for relevant posts gathered was set for X's inception (March 21, 2006) to the end of 2022.

The number of posts related to each mental health app gathered using R from the initial dataset is shown in Table 2.

Table . Number of posts related to each individual app gathered using R from the initial dataset after negative sentiment filtering.

App	Number of X posts contributed to dataset
Calm	574
Headspace	9984
Sleep Cycle	5527
BetterSleep	912
Waking Up	2606
Total	19,603

Preprocessing of the Data

Initially, 79,703 posts relating to the apps were gathered. The dataset then underwent negative sentiment lexicon filtering to identify posts expressing predominantly negative sentiment, consistent with the study's focus on user-reported issues and concerns relating to mental health apps. Sentiment filtering was performed using the Bing lexicon implemented through version 0.3.1 of the *tidytext* package [50]. The Bing lexicon is a curated list of words annotated with positive or negative polarity. Each post was tokenized into individual words, which were then matched against the lexicon. Matched words were assigned a sentiment score of +1 (positive) or −1 (negative), whereas unrecognized words were excluded from scoring. These values were summed within each post to produce a net sentiment score, where negative values indicated an overall negative tone. Posts with a negative net sentiment score were retained for subsequent STM and qualitative thematic analysis. After negative sentiment

XSL•FO
RenderX

filtering, 19,603 posts remained, and STM was applied to this dataset. Although this approach focused the analysis on expressions of negative user experience, some neutral or positive content remained within the final dataset, reflecting the inherent limitations of lexicon-based sentiment classification.

During additional data cleaning, app names were removed before STM was applied to organize the X posts into topics. Removing app names was crucial to prevent the STM from organizing X posts by name instead of the issues they presented. App names were reintroduced to the comments after STM had been completed. The data were preprocessed by first removing punctuation, symbols, and numbers, and then converting the text into individual token units. Common stop words were excluded, and stemming was applied to reduce words to their root forms, facilitating more efficient processing and analysis. In this study, no structured variables were included as covariates in the STM.

Structural Topic Model

Before running the models, diagnostic plots were generated to determine the optimal number of topics, considering both the relevant metrics and the goals of this analysis, while focusing on balancing semantic coherence and exclusivity [37]. An unsupervised topic modeling approach was evaluated by human coders, testing models with differing numbers of topics (5-40) and differing covariates in terms of coherence, residuals, and interpretability. A structural topic model with 10 topics and 20 posts per topic was identified as the most suitable for gaining insights based on the diagnostic statistics. Ultimately, a model without covariates was chosen.

We tested multiple candidate models across a range of topic numbers and evaluated their quality using established STM diagnostics, consistent with the approach in Sheen et al [38]. Specifically, we examined semantic coherence (how strongly words in a topic co-occur) and topic exclusivity (how distinct words are across topics) to assess each model's interpretability and distinctiveness. In addition to these quantitative diagnostics, candidate models were qualitatively reviewed by inspecting representative posts to determine whether the topics were meaningful and clearly interpretable. Based on this combined evaluation, a 10-topic model was selected because it provided the best balance: the topics were sufficiently distinct from one another (high exclusivity), internally coherent (high semantic coherence), and interpretable in terms of the substantive themes present in user comments on mental health apps. This approach mirrors the methodology used by Sheen et al [38], where the final number of topics was chosen based on both statistical diagnostics and qualitative interpretability rather than on a purely numeric criterion.

Output

Each of the 10 topics included as part of the output analysis was represented by the 20 X posts identified as the most semantically related to a given topic. The intention of the STM approach used in this study was that the posts within each topic would share commonalities because of the word co-occurrence of X posts within the generated topics, with the human researchers subsequently being able to label each topic by describing the main user issue evidenced by each topic. All 10 topics and quotes are provided in Multimedia Appendix 2.

Qualitative Analysis Process

After establishing the structural topic model, manual thematic analysis was performed by two researchers—Jack Bolter (JB) and an additional researcher—to provide interpretation of the topics. The researchers independently described each quote and labeled the STM-created topics based on the main issue raised by each topic. After independent analysis, the researchers discussed and agreed upon topic labels. A third researcher (PB) reviewed the analysis to increase the quality of the analysis. Where appropriate, the researchers organized the topics into major themes. Representative quotes for each topic are discussed in the Results section.

The following sequential steps describe how the researchers conducted the analysis following receipt of the STM output:

- Step 1: the researchers familiarized themselves with the data through repeated readings of the machine-generated topics and noted descriptions of each X-post.
- Step 2: the researchers independently added labels to describe each of the machine-generated topics.
- Step 3: the researchers discussed and reached consensus on major themes in their analyses, considering similarities and differences.
- Step 4: the researchers' findings and interpretations were documented in the Results and Discussion sections.

This study used MATA, a mixed methods approach that combines computational text analysis with qualitative interpretation to identify patterns in large text datasets. Following Bondaronek et al [37], MATA involves the application of automated topic modeling techniques to structure a large corpus of textual data into algorithmically derived topic groupings. These outputs are then reviewed and interpreted by researchers, who iteratively generate topic labels and develop higher-order themes. Rather than replacing human interpretation, the computational component is used to facilitate systematic exploration of large datasets, while interpretive analysis remains central to theme development and meaning-making.

While no automated bot or spam detection filters were applied during data collection, several aspects of the analytic process acted as a form of natural mitigation against the influence of such noise. Posts relating to the selected mental health apps were first organized using STM, which groups content based on patterns of semantic similarity. Given that bot or spam posts are often repetitive, promotional, or thematically inconsistent, such content is less likely to cohere meaningfully into stable topics.

No software was used to conduct the subsequent manual thematic analysis. Two researchers (JB and an additional researcher) independently reviewed the STM-generated topics and developed initial topic labels through close reading and interpretation of the data. An inductive approach was used for thematic labeling, with topic labels and themes generated directly from researcher interpretation of the STM outputs. The labels generated for each topic were based on the overall consistency and substantive meaning of the posts within a topic.

XSL•FO
RenderX

These independently generated labels were then compared and discussed. Any minor differences in interpretation were resolved through discussion and consensus. The researchers' interpretations were closely aligned, and discussion primarily served to refine and harmonize topic labels and overarching themes based on shared review of the STM outputs. Topic interpretation was therefore driven by recurring, coherent user experiences rather than individual posts. One topic was excluded entirely due to low coherence, illustrating how the human interpretive stage helped identify and remove clusters lacking a clear thematic thread, which may include random or spam-like content. In addition, not all posts within a topic contributed equally to theme definition; more detailed and thematically consistent posts were prioritized during interpretation, reducing the influence of isolated or anomalous content. While it remains possible that some bot or spam posts were present, industry estimates suggest such content accounted for approximately 10% of posts on X during the study period [51], indicating that any residual noise likely constituted a small minority and did not drive the overarching thematic findings.

Ethical Considerations

The permission to conduct the study was granted by The University of Westminster's Ethics Committee under application ID ETH2223-2307 (class 1). Users consent to X redistributing their public comments on its platform upon signing up to the platform [28]. To protect privacy, identifying details such as names and X handles were removed from reported results.

Beyond ethics approval and platform terms of service compliance, measures were taken to minimize potential harm associated with sensitive mental health-related content, including anonymization of usernames and handles. Results were primarily reported at the level of topics and themes, with a limited number of anonymized posts included as illustrative examples to support interpretation.

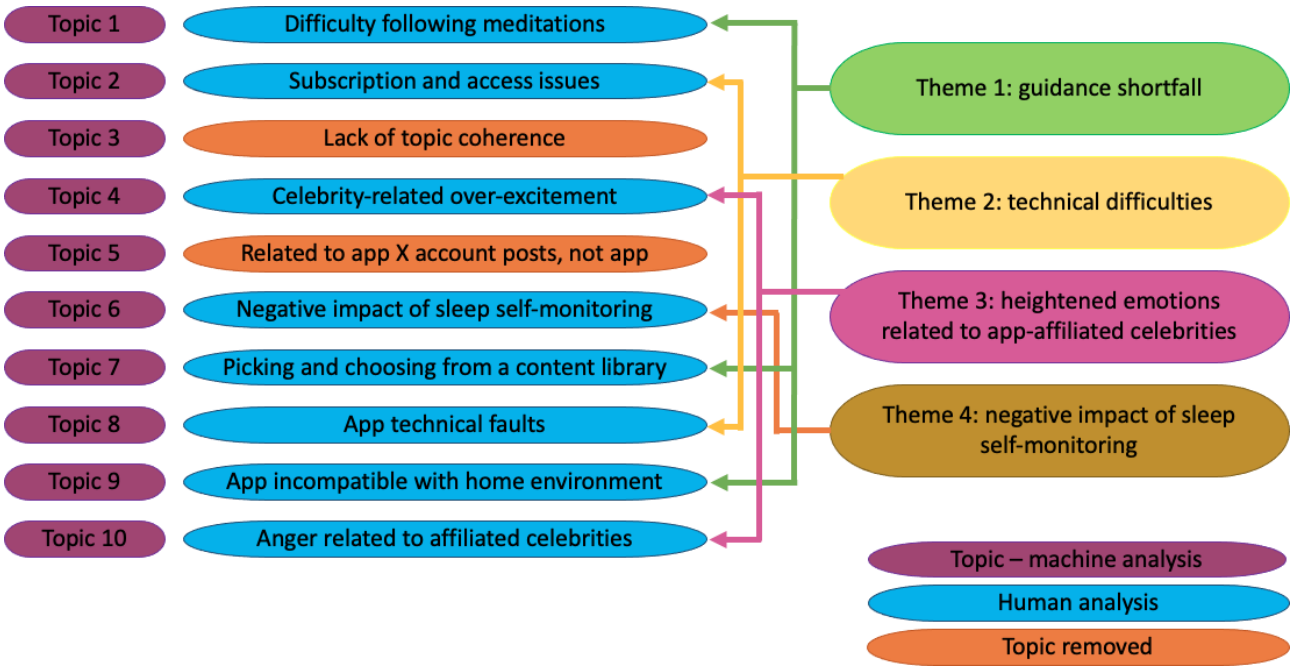
Results

Overview

Following data cleaning and negative sentiment filtering, 19,603 app-related X posts remained for STM topic organization (79,703 posts were gathered relating to the apps prior to negative sentiment filtering). Eight of the 10 subsequently STM-generated topics were included in the analysis, as these topics were coherent and provided insights relating to the issues experienced by users of the apps. The remaining 2 topics organized by the model were excluded from the analysis, 1 due to incoherence and the other due to comments being related to social media content rather than the apps themselves. As per Bondaronek et al [37], the researchers coded the X posts and assigned a label to each topic which summarized the main user issue identified. Eight of these topics were then organized into 4 themes (1 theme comprised a single stand-alone topic).

Figure 1 shows each of the topic labels and themes that the researchers produced to organize and derive meaning from the STM-generated topics.

Figure 1. Overview of the themes and topic labels derived from the structural topic model and interpreted through human thematic analysis by JB and an additional researcher. The guidance shortfall theme captures users' difficulties engaging with audio-guided content, including challenges understanding meditation instructions, selective and repetitive content use, and incompatibility with busy home environments. Technical difficulties reflect frustrations related to subscription access, in-app functionality, and limited responsiveness from app support services. Heightened emotions related to app-affiliated celebrities represent both positive excitement and negative reactions to celebrity involvement in app content and promotion. The negative impact of sleep self-monitoring theme highlights users' reliance on app-based sleep tracking, including benefits for some users alongside indications of reduced sleep self-efficacy and increased anxiety linked to monitoring practices.



XSL•FO
RenderX

Guidance Shortfall

The "guidance shortfall" main theme clustered 3 topics in which user comments demonstrated that additional guidance beyond the content within the apps would have helped users with their mindfulness and sleep hygiene practices.

Topic 1 focuses on users' difficulty following the apps' audio-guided meditations. Some quotes indicated trouble interpreting metaphors and concepts in the guided exercises, which appeared to hinder practice comprehension, while others demonstrated that users were physically uncomfortable while following the exercises.

> *I'm struggling with the idea of awareness in space and then being aware from it. I try to imagine it but all my awareness comes from my brain. I can imagine my awareness intermingling with space but in reality my awareness is only acknowledging space as aware from my brain.* [Topic 1, 18]

Topic 7 illustrates how users take the opportunity to pick and choose content from these apps and may use the same pieces of content repeatedly as opposed to following a structured course. Comments often appeared to be in response to posts from the apps' social media accounts asking X users what their favorite pieces of content on their respective apps were. For example, the following comment highlights one user's preference for two sleep podcasts on one of the apps:

> *Moon Buggy or Rainy Day Antiques... for the ZZZZs....* [Topic 7, 10]

Topic 9 was characterized by users commenting that they found their busy home environments to be incompatible with the apps. For example, one user reported difficulty following the relaxation and meditation guidance of the apps while they were aware of other noises within their home.

> *Tonight's episode was accompanied by the sounds of the hamster trying to break out of his cage, 6yr old's toy tiger roaring & 3yr old having coughing fit.* [Topic 9, 20]

Topic 9 also illustrated how busy home environments can interfere with recordings used to provide sleep quality analysis.

> *My [app name] app is supposed to record me snoring but most of the recordings are of (R)ingo snoring or him hissing at (B)owie.* [Topic 9, 5]

Technical Difficulties

The "technical difficulties" theme comprised 2 topics describing how users experienced technical problems when trying to access the apps and while using the apps.

Topic 2 was focused on subscription and access issues. Many users showed frustration due to not having gained the access levels they are entitled to despite having paid their subscription or being due access for another reason (eg, promotions allowing app access to individuals in certain professions).

> *… I'm confused. I paid £49.99 annual subscription but I still have some areas locked. Why is this please(?).* [Topic 2, 2]

Such frustration was exacerbated when users struggled to contact app staff to seek support or compensation.

> *(I have) been trying to contact you for several days by email. No response from your team. Very disappointed. My subscription fee charged was a lot higher than what was shown in−app. Please reverse this immediately...* [Topic 2, 16]

Topic 8 highlighted that some users experienced negative emotions due to technical faults encountered while using the apps.

> *I'm confounded by how LOUD the default volume is on music, SFX, stories, etc ... with no apparent way to adjust the volume, in app. If the answer=turn down iPhone volume, then alarm volume is compromised. Unusable. Am I missing something?* [Topic 8, 18]

As with topic 2, such frustration was exacerbated when users struggled to contact app support staff:

> *How do I escalate a fault on my app which I raised on 5 Aug(?) (It is) with the engineers but "Sadly, we don't currently have an estimate for when the issue might be fully resolved(.)"* [Topic 8, 20]

Heightened Emotions Related to App-Affiliated Celebrities

The main theme related to "app-affiliated celebrities" clustered 2 topics showing how users experienced heightened emotions—both positive and negative—about public figures with links to the apps. Such affiliations were generally due to celebrities either creating content for the apps or having sponsorship deals for the apps on their own platforms.

Topic 4 demonstrated that celebrity partnerships with the apps can have an effect of elevating users to states of excitement that are counter to the states of calm and relaxation on which the apps are marketed. Several X posts were all in capital letters and related to an appearance of a famous popstar on one of the apps to narrate content.

> *OMFG, I'm dead. The shoes not even on his feet. His hat the beard his outfit. OMG.* [Topic 4, 19]

While comments in topic 4 were largely positive, if perhaps excitable in tone, those in topic 10 often expressed anger in relation to celebrities affiliated with the apps.

> *(Celebrity name) (is) a fraud owned by China… the most overrated player in NBA history.* [Topic 10, 3]

Negative Impact of Sleep Self-Monitoring

Topic 6 is a stand-alone theme relating to the reliance some users seemed to have on using these apps to aid their sleep and the sleep self-monitoring they performed via the app. Many of the user comments were positive about how the apps impacted their sleep.

> *The (app name) App for iOS is a game changer. Wakes you up at the end of a cycle so as to not startle you awake from REM sleep.* [Topic 6, 3]

XSL•FO
RenderX

However, other posts hinted that users' focus on app statistics might negatively impact their sleep self-efficacy and be problematic due to excessive sleep self-monitoring.

> *Ok party people, I'm totally addicted and fascinated by the (app name) app! Always knew I sucked at sleeping but WOW my graphs r shocking!* [Topic 6, 14]

Discussion

Principal Findings

This study investigated issues users experience with mental health apps by analyzing 19,603 negative sentiment X posts related to the 5 highest-grossing mental health apps (79,703 posts relating to the apps were gathered prior to negative sentiment filtering). Using machine learning with qualitative analysis, the data were organized into 10 topics and interpreted by researchers into four main themes: guidance shortfalls, technical difficulties, heightened emotions linked to app-affiliated celebrities, and the negative impact of sleep self-monitoring. These themes encapsulated a range of richly detailed user concerns, including frustration with insufficient in-app guidance, technical obstacles, overexcitement and distress related to celebrity presence, and a potentially detrimental overemphasis on sleep data. Although users did not report direct feelings of unsafety, the absence of in-person support and the intensity of negative emotions expressed suggest that some users perceived limitations in the ability of unregulated mental health apps to meet their support needs.

Several findings from the study align with existing research. Users' difficulties with guided meditation and feelings of being unsupported resonate with Crane et al [52], who emphasize the importance of instructor feedback in mindfulness practice via the interactive inquiry section of classes, and Ruse et al [53], who report dissatisfaction stemming from the lack of human interaction in mental health apps. Users' perceptions of feeling unsupported by audio guidance during guided meditations can be interpreted through self-determination theory, which posits that sustained motivation depends on supporting basic psychological needs such as autonomy and competence; when digital guidance fails to scaffold these needs, users are less likely to internalize the practice and maintain engagement. Self-determination theory has been applied in human-computer interaction research to explain how technology design can either foster or undermine intrinsic motivation by shaping whether users feel capable, self-directed, and meaningfully engaged [54]. While audio guidance is intended to support practice, limited personalization and adaptive feedback—which are more feasible in in-person meditation classes and groups—may diminish users' sense of autonomy and competence, reducing the subjective feeling of being supported. Similarly, the observation that technical issues led to frustration and stress supports findings by Haggag et al [55] and Woods et al [56], who highlight usability problems as key barriers to engagement and therapeutic benefit. The study's insights into the risks of sleep self-monitoring—such as anxiety over sleep statistics and potential overfocus on tracking—also reflect earlier research by Baron et al [57] on orthosomnia (the idea that preoccupation with sleep quality and tracking results may paradoxically impair sleep quality).

Recent research has critically examined the user experience and interpretive practices associated with sleep tracking technologies, highlighting issues of data literacy, accuracy of app data, and optimization logic assumptions. O'Neill and Nansen [58] analyzed how mobile sleep apps do not simply monitor sleep but actively intervene in users' sleep-wake rhythms through features such as "smart wake-up" alarms and brainwave entrainment sounds. They show how these functions are underpinned by optimization logics that position sleep as something that can be measured, managed, and improved through algorithmic systems, thereby shaping how users understand and act upon their own sleep. The authors illustrate that this is problematic given the potential for these apps to produce inaccurate user sleep data. Similarly, Liang et al [59] demonstrate that users often grapple with interpreting sleep data, questioning its reliability and struggling to translate metrics into meaningful or actionable insights. Together, this literature underscores that engagement with sleep tracking apps involves complex processes of interpretation, negotiation, and trust. However, it also demonstrates that such trust can also generate user anxiety that they are not achieving optimal sleep.

In terms of novel contributions, this study found that users expressed strong emotional responses to celebrity content within mental health apps, which some posts suggested could detract from experiences of calm, focus, and emotional regulation that users associated with mindfulness practices. Existing research related to celebrity affiliations with mHealth apps has largely focused on the commercial benefits rather than their potential to interfere with therapeutic outcomes [60]. This study also highlights how apps' associations—or disassociations—with public figures can provoke strong emotional responses in users, including anger, particularly when these public figures are involved in polarizing discourse. This suggests that app branding and celebrity affiliations may be associated with socially and politically charged reactions among users. While prior work has explored the marketing benefits of celebrity endorsement, the psychological and affective implications for users remain underexplored in the mHealth app literature. The findings relating to celebrity affiliations for commercial purposes contribute to a growing body of literature that critiques the ways in which digital wellness and mindfulness apps draw from traditional Eastern philosophies—such as Buddhist practices rooted in ethical precepts such as sila—while stripping away their moral and communal dimensions [61-63]. These practices follow other recent examples of traditional mindfulness practices being reframed within Western commercial frameworks to promote individual calm, emotional regulation, or productivity, raising questions about cultural appropriation and the commodification of spirituality.

It was also found that some users repeatedly engage with familiar or favorite content. Such repeated use of the same content as opposed to following a structured program may mean users are less likely to experience meaningful behavior change. While Bol et al [64] found that autonomy and customization when using mHealth apps can increase behavioral intentions for users with a high need for autonomy, their study also

XSL•FO
RenderX

reported that, on average, customization did not lead to greater perceived control, motivation, or intention to act. While user autonomy may facilitate short-term engagement, long-term behavioral change may require more structured, theory-driven approaches, such as those informed by the Health Action Process Approach [65]. Further research is warranted to clarify the conditions under which user autonomy supports or hinders behavioral outcomes in mHealth app contexts.

Strengths and Weaknesses

A significant strength of the study lies in its methodological approach, using STM in combination with qualitative analysis to effectively synthesize insights from a vast volume of real-world data. This approach, as demonstrated by Bondaronek et al [37], proved reliable and valid. The use of X posts enhanced ecological validity by capturing unfiltered user experiences in situ [26].

A key limitation of this study lies in the lack of nuance inherent in the methodological approach used. While STM is effective for identifying broad thematic patterns across large datasets, it is not well suited to capturing the complexity and depth of individual user experiences. This is compounded by the low character limits of most X posts within the available data, which constrain user expression and may exclude more detailed or reflective insights. Moreover, the diversity of the apps analyzed—ranging from mindfulness-focused to sleep-oriented platforms—introduces variability that may limit the generalizability of specific findings across the broader category of mHealth apps. Future research could address these limitations through follow-up interviews or focus groups to help verify the findings and gain a richer understanding of the identified user experiences.

The absence of automated spam or bot detection represents a limitation of this study, as no specific bot-detection procedures were applied to the dataset prior to analysis. Consequently, the STM input data included a degree of automated or low-quality content, introducing noise into topic summarization and contributing to some less coherent or contextually relevant posts within topics. Although spam and bot detection methods may have reduced this noise and potentially improved topic coherence and interpretability, such approaches are imperfect, particularly when applied to large-scale social media datasets, meaning that some automated content may still have remained even if filtering had been undertaken. To minimize the influence of noisy data on interpretation, the analysis focused on topics that demonstrated coherence and relevance to user experiences, with an incoherent topic and another unrelated to user experiences being excluded from the subsequent analysis.

Another key limitation of this study is that it draws on publicly shared social media posts, and we cannot infer that users' experiences reported online correspond to clinically assessed mental health outcomes. The themes identified reflect users' self-reported issues when using mental health apps, rather than verified experiences of anxiety or other clinical conditions. Consequently, the findings should be interpreted as indicative of the types of issues users report, rather than evidence of therapeutic effectiveness or harm.

It is also important to recognize that X users are not fully representative of the broader population of mental health app users, which may limit the generalizability of our findings. Demographically, X's user base skews toward younger adults, with an estimated 71.6% of users being <34 years in 2025. The X user base is also male-dominated, with males comprising 63.7% of the global users in 2025 [66]. Furthermore, for the time period from which the data were gathered (2006 up until the end of 2022), X users tended to be more highly educated and comparatively affluent than the general population, with higher proportions of adults in higher income brackets and tertiary education levels [67]. These demographic profiles suggest that perspectives captured in our analysis may overrepresent the experiences of younger, more educated, and higher-income individuals, while underrepresenting older adults, lower-income groups, and those who do not engage with X as a forum for discussing app experiences.

Character limits were only increased to 4000 characters for paying subscribers after this study's data were collected, meaning that X posts in this study had character limits of 140 (for X posts posted prior to November 6, 2017) and 280 (for X posts posted from November 7, 2017, onward) [68]. Such brevity was advantageous for the manageability of the dataset, but it also limited the extent to which users were able to articulate their experiences.

The dataset was unevenly distributed across the 5 included mental health apps, with some apps generating substantially more posts than others. This variation reflects differences in app popularity and user engagement on the platform, rather than a sampling bias introduced by the study design. However, this imbalance may have influenced the relative prominence of topics identified through STM, as apps with higher volumes of user-generated content are potentially more likely to contribute greater proportions of text to the corpus. The study was designed to identify common issues associated with mental health app use more broadly, rather than to conduct app-specific comparisons, and therefore findings should be interpreted at this aggregate level. No sensitivity analyses were conducted to rebalance the dataset or isolate app-specific effects, as this was beyond the scope of the study's exploratory aims.

Implications for Research, Practice, and Policy

App developers should address the lack of guidance support identified in user reports by strengthening onboarding processes, in-app guidance, and wayfinding features. Difficulties navigating app content and uncertainty about how to engage with interventions suggest a need for clearer instructional scaffolding and more explicit signposting of core features. To ensure such guidance supports meaningful behavior change rather than surface-level engagement, interventions should be aligned with established behavior change theories such as the Health Action Process Approach [65].

User complaints relating to subscription practices and billing indicate a need for greater transparency around pricing structures, trial periods, and cancellation processes, as well as more responsive customer support mechanisms. Similarly, reports of technical instability and disrupted functionality point to the importance of improving app reliability to prevent

XSL•FO
RenderX

frustration and disengagement. Together, these issues suggest that usability and commercial design decisions may contribute to users reporting frustration, stress, and feelings of being unsupported, potentially diminishing the perceived value of mental health apps.

The findings of this study suggest that greater consideration should be given to the regulation and oversight of mental health apps. Policymakers should consider frameworks to evaluate and certify digital mental health tools, especially those that might be used in place of traditional therapeutic support. Certification schemes could distinguish evidence-based digital therapeutics from general wellness apps. Dialogue between public health bodies and app providers, as well as public information campaigns, may help users make more informed choices about mental health support and whether they should use mental health apps. Clinicians could also use the present findings to support patients in critically evaluating app-based mental health tools and, where appropriate, integrate curated lists of validated apps into care strategies to complement—not replace—traditional therapeutic support.

Future research should test the robustness of this study's STM-based findings across diverse populations and mental health contexts. Additional studies could explore whether key user issues were missed due to the analytical approach of STM. Research is needed to evaluate the psychological outcomes associated with mental health app use and to explore users' perceptions regarding the role of these apps alongside, or potentially in place of, professional support.

Conclusion

This study provides further evidence that STM can be combined with human qualitative methods to quickly analyze large volumes of X data. With data relating to the 5 most popular mental health apps analyzed via this approach, the researchers were able to expeditiously identify several issues encountered by users. This method, therefore, has the potential to be used to swiftly provide insights relating to how individuals experience various digital health interventions.

Acknowledgments

Thanks are extended to Maria Ceka, one of my fellow students in the MSc Health Psychology program at the University of Westminster, for independently performing thematic analysis of the data before collaborating with me to reach consensus on the data's topic labels and themes. Jack Bolter is a graduate of the MSc Health Psychology program at the University of Westminster. No generative AI tools were used in the preparation of this manuscript.

Funding

PB is supported by the Wellcome Trust (Project No. 300252/Z/23/Z).

Data Availability

Outputs generated by the structural topic model are available in the Multimedia Appendices. The authors are available to provide further methodological details regarding data collection and data preprocessing upon request.

Authors' Contributions

JB conducted the formal analysis and wrote the original draft of the paper, as well as reviewed and edited subsequent versions. TP was responsible for data curation and software. PB contributed to the conceptualization, methodology, supervision, and validation of the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Document illustrating the coding and keywords used to gather X posts relating to the 5 top-grossing mental health apps.
[DOCX File, 14 KB - mhealth_v14i1e85575_app1.docx]

Multimedia Appendix 2
All topics with redactions.
[DOCX File, 47 KB - mhealth_v14i1e85575_app2.docx]

References

1. McManus S, Meltzer H, Brugha T, Bebbington P, Jenkins R. Adult psychiatric morbidity in England – 2007, results of a household survey – NHS digital. : NHS Digital; 2007. URL: https://digital.nhs.uk/data-and-information/publications/statistical/adult-psychiatric-morbidity-survey/adult-psychiatric-morbidity-in-england-2007-results-of-a-household-survey [accessed 2026-07-16]

XSL•FO
RenderX

2. McManus S, Bebbington P, Jenkins R, Brugha T. Adult Psychiatric Morbidity Survey: Survey of Mental Health and Wellbeing. 2014. URL: https://webarchive.nationalarchives.gov.uk/ukgwa/20180328140249/http://digital.nhs.uk/catalogue/PUB21748 [accessed 2026-07-16]
3. Friedrich MJ. Depression Is the Leading Cause of Disability Around the World. JAMA 2017 Apr 18;317(15):1517. [doi: 10.1001/jama.2017.3826] [Medline: 28314561]
4. Mental health. World Health Organization. 2017. URL: https://www.who.int/health-topics/mental-health#tab=tab_1 [accessed 2026-07-16]
5. Depression. Mind. 2019. URL: https://www.mind.org.uk/information-support/types-of-mental-health-problems/depression/symptoms/ [accessed 2026-07-16]
6. Health status, illnesses, and other conditions (Welsh Health Survey): 2015. Welsh Government. 2015. URL: https://www.gov.wales/health-status-illnesses-and-other-conditions-welsh-health-survey-2015?tab=previous&lang=cy [accessed 2026-07-16]
7. Henderson C, Evans-Lacko S, Thornicroft G. Mental illness stigma, help seeking, and public health programs. Am J Public Health 2013 May;103(5):777-780. [doi: 10.2105/AJPH.2012.301056] [Medline: 23488489]
8. Mental Health Apps Market Size, Share and Trends 2026 to 2035. Precedence Research. 2022. URL: https://www.precedenceresearch.com/mental-health-apps-market [accessed 2023-05-16]
9. Leading health and meditation apps worldwide in December 2022, by revenue (in US dollars). AppMagic.: Statista; 2023. URL: https://www-statista-com.uow.idm.oclc.org/statistics/1239670/top-health-and-meditation-apps-by-revenue/ [accessed 2023-05-16]
10. BinDhim NF, Hawkey A, Trevena L. A systematic review of quality assessment methods for smartphone health apps. Telemed J E Health 2015 Feb;21(2):97-104. [doi: 10.1089/tmj.2014.0088] [Medline: 25469795]
11. Bondaronek P, Alkhaldi G, Slee A, Hamilton FL, Murray E. Quality of Publicly Available Physical Activity Apps: Review and Content Analysis. JMIR Mhealth Uhealth 2018 Mar 21;6(3):e53. [doi: 10.2196/mhealth.9069] [Medline: 29563080]
12. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020 Feb 1;27(2):330-340. [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
13. Ananth S. Sleep apps: current limitations and challenges. Sleep Sci 2021;14(1):83-86. [doi: 10.5935/1984-0063.20200036] [Medline: 34104344]
14. Lee-Tobin PA, Ogeil RP, Savic M, Lubman DI. Rate My Sleep: Examining the Information, Function, and Basis in Empirical Evidence Within Sleep Applications for Mobile Devices. J Clin Sleep Med 2017 Nov 15;13(11):1349-1354. [doi: 10.5664/jcsm.6814] [Medline: 28992831]
15. Daudén Roquet C, Sas C. Evaluating mindfulness meditation apps. Presented at: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems; Apr 21-26, 2018. [doi: 10.1145/3170427.3188616]
16. Longyear RL, Kushlev K. Can mental health apps be effective for depression, anxiety, and stress during a pandemic? Practice Innovations. 2021. [doi: 10.31234/osf.io/zy2ct]
17. Shah RN, Berry OO. The Rise of Venture Capital Investing in Mental Health. JAMA Psychiatry 2020 Sep 16;16. [doi: 10.1001/jamapsychiatry.2020.2847] [Medline: 32936238]
18. Kao CK, Liebovitz DM. Consumer Mobile Health Apps: Current State, Barriers, and Future Directions. PM&R 2017 May;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28284894]
19. Jacob C, Sanchez-Vazquez A, Ivory C. Understanding Clinicians' Adoption of Mobile Health Tools: A Qualitative Review of the Most Used Frameworks. JMIR Mhealth Uhealth 2020;8(7):e18072. [doi: 10.2196/18072] [Medline: 32667882]
20. Agarwal S, Jalan M, Wilcox HC, et al. Methods 2022. URL: https://www.ncbi.nlm.nih.gov/books/NBK580943 [accessed 2026-07-16]
21. Callahan C, Kimber J, Hu E, Tanner L, Kunkle S. The Real-World Impact of App-Based Mindfulness on Headspace Members With Moderate and Severe Perceived Stress: Observational Study. JMIR Mhealth Uhealth 2024 Mar 4;12:e52968. [doi: 10.2196/52968] [Medline: 38488513]
22. Headspace vs Waking Up: Which Is the Better Meditation App? Neura. 2026. URL: https://neura.health/insight/headspace-vs-waking-up-which-is-better [accessed 2026-02-03]
23. Huberty J, Puzia ME, Larkey L, Vranceanu AM, Irwin MR. Can a meditation app help my sleep? A cross-sectional survey of Calm users. PLOS ONE 2021;16(10):e0257518. [doi: 10.1371/journal.pone.0257518] [Medline: 34679078]
24. deBara D. I Tried the BetterSleep App as Part of My Nighttime Routine—Here's What Happened. Sleep Enlightened by Saatva. 2024. URL: https://www.saatva.com/blog/bettersleep-app-review [accessed 2026-07-16]
25. Peters B. Review of Sleep Cycle App: Tracking Movement to Time Awakenings. Verywell Health. 2022. URL: https://www.verywellhealth.com/sleep-cycle-iphone-app-3973921 [accessed 2026-07-16]
26. Lomborg S, Bechmann A. Using APIs for Data Collection on Social Media. The Information Society 2014 Aug 8;30(4):256-265. [doi: 10.1080/01972243.2014.915276] [Medline: 24834147]
27. Andreotta M, Nugroho R, Hurlstone MJ, et al. Analyzing social media data: A mixed-methods framework combining computational and qualitative text analysis. Behav Res Methods 2019 Aug;51(4):1766-1781. [doi: 10.3758/s13428-019-01202-8] [Medline: 30941697]

XSL•FO
RenderX

28. Arrington M. Odeo Releases Twttr. TechCrunch. 2006. URL: https://techcrunch.com/2006/07/15/is-twttr-interesting/ [accessed 2023-05-21]
29. Ricker T. Jobs: App Store launching with 500 iPhone applications, 25% free. Endgadget. 2008. URL: https://tinyurl.com/dus9z6ht [accessed 2013-05-21]
30. Takahashi D. Google releases details on android market launch. VentureBeat. 2008. URL: https://venturebeat.com/business/google-releases-details-on-android-market-launch/ [accessed 2023-05-21]
31. Dang S. Exclusive: Twitter is losing its most active users, internal documents show. Reuters. 2022. URL: https://www.reuters.com/technology/exclusive-where-did-tweeters-go-twitter-is-losing-its-most-active-users-internal-2022-10-25 [accessed 2026-07-16]
32. Cairns I. Making customer service even better on twitter. X Blog. 2016. URL: https://blog.twitter.com/en_us/a/2016/making-customer-service-even-better-on-twitter?ref=buffer.com [accessed 2023-05-21]
33. Alhabash S, Ma M. A Tale of Four Platforms: Motivations and Uses of Facebook, Twitter, Instagram, and Snapchat Among College Students? Social Media + Society 2017 Jan;3(1):1-13. [doi: 10.1177/2056305117691544]
34. Elbattah M, Arnaud É, Gignon M, Dequen G. The role of text analytics in healthcare: a review of recent developments and applications. 2021 Presented at: Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies; Feb 11-13, 2021. [doi: 10.5220/0010414508250832]
35. Feizollah A, Lin CY, O'Malley L, Thompson W, Listl S, Byrne M. The use of natural language processing to interpret unstructured patient feedback on health services: scoping review. J Med Internet Res 2025 Aug 14;27:e72853. [doi: 10.2196/72853] [Medline: 40811803]
36. Stewart J, Lu J, Goudie A, et al. Applications of natural language processing at emergency department triage: a narrative review. PLoS ONE 2023;18(12):e0279953. [doi: 10.1371/journal.pone.0279953] [Medline: 36701444]
37. Bondaronek P, Papakonstantinou T, Stefanidou C, Chadborn T. User feedback on the NHS test & trace Service during COVID-19: the use of machine learning to analyse free-text data from 37,914 England adults. Public Health Pract (Oxf) 2023 Dec;6:100401. [doi: 10.1016/j.puhip.2023.100401] [Medline: 38099087]
38. Sheen F, Porter L, Papakonstantinou T, Ceka M, Bondaronek P. Living well? The unintended consequences of highly popular commercial fitness apps through social listening using machine-assisted topic analysis: evidence from X. Br J Health Psychol 2025 Nov;30(4):e70026. [doi: 10.1111/bjhp.70026] [Medline: 41121969]
39. Blei DM. Probabilistic topic models. Commun ACM 2012 Apr;55(4):77-84. [doi: 10.1145/2133806.2133826]
40. Mishler A, Crabb ES, Paletz S, Hefright B, Golonka E. Using structural topic modeling to detect events and cluster Twitter users in the Ukrainian crisis. Commun Comput Inf Sci ;2015:639-644. [doi: 10.1007/978-3-319-21380-4_108]
41. Lösch L, Brown P, van Hunsel F. Using structural topic modelling to reveal patterns in reports on opioid drugs in a pharmacovigilance database. Pharmacoepidemiol Drug Saf 2022 Sep;31(9):1003-1006. [doi: 10.1002/pds.5502] [Medline: 35751621]
42. Wright L, Fluharty M, Steptoe A, Fancourt D. How did people cope during the COVID-19 pandemic? A structural topic modelling analysis of free-text data from 11,000 United Kingdom adults. Front Psychol 2022;13:810655. [doi: 10.3389/fpsyg.2022.810655] [Medline: 35734451]
43. Jo W, Kim Y, Seo M, Lee N, Park J. Online information analysis on pancreatic cancer in Korea using structural topic model. Sci Rep 2022 Jun 23;12(1):10622. [doi: 10.1038/s41598-022-14506-1] [Medline: 35739151]
44. Yang L. Mining and visualizing large-scale course reviews of LMOOCs learners through structural topic model. PLoS ONE 2023;18(5):e0284463. [doi: 10.1371/journal.pone.0284463]
45. Calm (version 6.23.1) [mobile application software]. App Store. 2023. URL: https://apps.apple.com/gb/app/calm-sleep-meditation/id571800810 [accessed 2023-05-25]
46. Headspace inc. headspace (3.268.0) [mobile application software]. App Store. 2023. URL: https://apps.apple.com/gb/app/headspace-mindful-meditation/id493145008 [accessed 2023-05-22]
47. Sleep cycle – sleep tracker (6.23.19) [mobile application software]. App Store. 2023. URL: https://apps.apple.com/gb/app/sleep-cycle-sleep-tracker/id320606217 [accessed 2023-05-25]
48. Ipnos software inc. bettersleep: relax and sleep (23.8.1) [mobile application software]. App Store. 2023. URL: https://apps.apple.com/gb/app/bettersleep-relax-and-sleep/id314498713 [accessed 2023-05-25]
49. Waking up: meditation & wisdom (2.15.1) [mobile application software]. App Store. 2023. URL: https://apps.apple.com/gb/app/waking-up-meditation-wisdom/id1307736395 [accessed 2023-05-25]
50. Barrie C, Ho J. academictwitteR: an R package to access the Twitter Academic Research Product Track v2 API endpoint. JOSS 2021;6(62):3272. [doi: 10.21105/joss.03272]
51. 10% of Twitter's active accounts are posting spam content, estimates GlobalData. GlobalData. 2022. URL: https://www.globaldata.com/media/business-fundamentals/10-twitters-active-accounts-posting-spam-content-says-globaldata/ [accessed 2026-02-03]
52. Crane RS, Brewer J, Feldman C, et al. What defines mindfulness-based programs? The warp and the weft. Psychol Med 2017 Apr;47(6):990-999. [doi: 10.1017/S0033291716003317] [Medline: 28031068]
53. Ruse JN, Schraube E, Rhodes P. Left to their own devices: the significance of mental health apps on the construction of therapy and care. Subjectivity 2024 Dec;31(4):410-428. [doi: 10.1057/s41286-024-00195-w]

XSL•FO
RenderX

54. Alberts L, Lyngs U, Lukoff K. Designing for sustained motivation: a review of self-determination theory in behaviour change technologies. Interact Comput 2026 Apr 13;38(3):447-468. [doi: 10.1093/iwc/iwae040]
55. Haggag O, Grundy J, Abdelrazek M, Haggag S. A large scale analysis of mHealth app user reviews. Empir Softw Eng 2022;27(7):196. [doi: 10.1007/s10664-022-10222-6] [Medline: 36246486]
56. Woods LS, Duff J, Roehrer E, Walker K, Cummings E. Patients' experiences of using a consumer mHealth app for self-management of heart failure: mixed-methods study. JMIR Hum Factors 2019 May 2;6(2):e13009. [doi: 10.2196/13009] [Medline: 31045504]
57. Baron KG, Abbott S, Jao N, Manalo N, Mullen R. Orthosomnia: are some patients taking the quantified self too far? J Clin Sleep Med 2017 Feb 15;13(2):351-354. [doi: 10.5664/jcsm.6472] [Medline: 27855740]
58. O'Neill C, Nansen B. Sleep mode: mobile apps and the optimisation of sleep-wake rhythms. FM 2019;24(6). [doi: 10.5210/fm.v24i6.9574]
59. Liang Z, Ploderer B, Liu W, et al. SleepExplorer: a visualization tool to make sense of correlations between personal sleep data and contextual factors. Pers Ubiquit Comput 2016 Nov;20(6):985-1000. [doi: 10.1007/s00779-016-0960-6]
60. Garnett C, Perski O, Beard E, Michie S, West R, Brown J. The impact of celebrity influence and national media coverage on users of an alcohol reduction app: a natural experiment. BMC Public Health 2021 Jan 6;21(1):30. [doi: 10.1186/s12889-020-10011-0] [Medline: 33407283]
61. Miyakawa M. Critique of modern instrumental mindfulness as neoliberal narcissism [thesis]. : Simon Fraser University; 2024 Jul 22 URL: https://summit.sfu.ca/item/38413 [accessed 2026-07-12]
62. Purser RE. McMindfulness: How Mindfulness Became the New Capitalist Spirituality: Repeater; 2019. URL: https://mitpressbookstore.mit.edu/book/9781912248315 [accessed 2026-07-12]
63. Solomatina I. 'Mindfulness for modern life': The blending of Orientalism and neoliberal self-care strategies in the branding of Rituals. Eur J Cul Stud 2025 Jun;28(3):849-865. [doi: 10.1177/13675494241284637]
64. Bol N, Høie NM, Nguyen MH, Smit ES. Customization in mobile health apps: explaining effects on physical activity intentions by the need for autonomy. Digit HEALTH 2019;5:2055207619888074. [doi: 10.1177/2055207619888074] [Medline: 31807312]
65. Schwarzer R. Self-Efficacy: Taylor & Francis; 2014. [doi: 10.4324/9781315800820]
66. X (Twitter) Global Users Statistics 2025 | X Demographics. The Global Statistics. 2025. URL: https://www.theglobalstatistics.com/x-twitter-global-users-statistics/ [accessed 2026-02-03]
67. Wojcik S, Hughes A. Sizing Up Twitter Users. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users [accessed 2026-07-12]
68. Reimann N. Twitter boosts character limit to 4,000 for twitter blue subscribers. Forbes. 2023. URL: https://www.forbes.com/sites/nicholasreimann/2023/02/08/twitter-boosts-character-limit-to-4000-for-twitter-blue-subscribers [accessed 2025-07-12]

Abbreviations

MATA: machine-assisted thematic analysis
mHealth: mobile health
NLP: natural language processing
STM: structural topic modeling

Edited by A Stone, L Buis; submitted 10.Oct.2025; peer-reviewed by B Nansen, M Elbattah, R Ruparel, T Yusuff; revised version received 10.Jun.2026; accepted 18.Jun.2026; published 21.Jul.2026.

Please cite as:
Bolter J, Papakonstantinou T, Bondaronek P
User-Reported Issues With Mental Health Apps: Machine-Assisted Topic Analysis of Social Media Posts
JMIR Mhealth Uhealth 2026;14:e85575
URL: https://mhealth.jmir.org/2026/1/e85575
doi:10.2196/85575

© Jack Bolter, Trisevgeni Papakonstantinou, Paulina Bondaronek. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 21.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobile Phone Access, Usage Patterns, and Perceptions of Adolescents Living With HIV on the Use of Gamified Interventions to Improve Antiretroviral Therapy Adherence in Eswatini: Qualitative Study

Londiwe D Hlophe[1], PhD; Peter S Nyasulu[1,2], PhD; Constance S Shumba[3], PhD

[1]Division of Epidemiology and Biostatistics, Faculty of Medicine and Health Sciences, Stellenbosch University, PO Box 241, Cape Town, South Africa
[2]Division of Epidemiology and Biostatistics, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
[3]Division of Epidemiology and Social Sciences, Institute for Health and Equity, Medical College of Wisconsin, Wisconsin, United States

Corresponding Author:
Londiwe D Hlophe, PhD
Division of Epidemiology and Biostatistics, Faculty of Medicine and Health Sciences, Stellenbosch University, PO Box 241, Cape Town, South Africa

Abstract

Background: Adolescents living with HIV often experience poor antiretroviral therapy (ART) outcomes due to multiple barriers affecting medication adherence. Effective self-care interventions are needed to address these challenges. Mobile phones are widely used by the adolescent population and therefore present an opportunity to enhance ART adherence using mobile phone–based interventions. However, research on mobile phone access among adolescents living with HIV, usage patterns, and perceptions of mobile phone–based interventions is limited in Eswatini.

Objective: This study aimed to explore these aspects to inform effective mobile health strategies for enhancing ART adherence among adolescents living with HIV.

Methods: We conducted a qualitative study using in-depth interviews in December 2023. A total of 29 adolescents living with HIV aged 10 to 19 years and enrolled on ART were purposively sampled and interviewed from 5 Teen Clubs in the Hhohho region of Eswatini. Interviews were audio-recorded and transcribed verbatim. Topic areas covered were mobile phone accessibility, usage patterns, and perceptions on the use of mobile phones to facilitate ART adherence. The data were analyzed using the deductive-inductive coding approach.

Results: Of the 29 participants, 15 (52%) were female, and 19 (65.5%) were aged between 15 and 19 years. The study findings indicated high mobile phone access among participants, with primary usage focused on making and receiving calls, as well as engaging with social media. Three themes emerged regarding the use of gamified interventions to support ART adherence. First, the use of gamified interventions aimed at ART adherence among adolescents living with HIV was deemed feasible based on mobile phone access and past experiences with mobile games. Second, 3 main qualities of successful gamified interventions were identified as being supportive, being educational, and ensuring secure and confidential connections with other players. Finally, confidentiality and mobile phone access factors were highlighted as potential concerns when designing gamified ART adherence interventions.

Conclusions: The findings suggest potentially high access and usage of mobile phones among adolescents living with HIV on ART in Eswatini. This provides an opportunity to leverage mobile technology to enhance ART adherence through gamified interventions. However, it is essential to carefully consider the specific needs and concerns of adolescents living with HIV in the design of these interventions to ensure their successful uptake and sustainability.

(JMIR Mhealth Uhealth 2026;14:e74207) doi:10.2196/74207

KEYWORDS

adolescents living with HIV; antiretroviral therapy; ART adherence; mobile phones; mobile health; gamified interventions; Eswatini

Introduction

Eswatini's adolescents have consistently shown lower antiretroviral therapy (ART) outcomes, such as viral load suppression (VLS), compared to the adult population. In 2023, Eswatini recorded a 96% VLS rate among adults, while adolescents aged 15 to 19 years reported a 77.8% viral suppression rate among those on ART [1,2]. Although there

XSL•FO
RenderX

has been a very slight improvement compared to 2017 data where VLS was 76% among adolescents and young people, it remains below the Joint United Nations Programme on HIV/AIDS (UNAIDS) 90% VLS target for 2020 [3,4]. Low VLS is associated with poor ART adherence, leading to opportunistic infections and subsequently AIDS-related deaths [5-7].

Nonetheless, Eswatini has made tremendous progress in combating the HIV epidemic, despite having the highest prevalence of HIV globally at 24.8% [1]. This progress has positioned the country among the first to meet the 2030 UNAIDS targets ahead of the set date [1,2]. The UNAIDS targets aim for 95% of people living with HIV to know their status, 95% of those diagnosed to be on ART, and 95% of those on ART to achieve viral suppression by 2030 [8]. Eswatini has achieved these targets through successfully implemented interventions promoting ART adherence [2]. For instance, in 2017, the country adopted the universal test and treat guidelines and coupled them with interventions promoting ART adherence counseling, resulting in the implementation of a differential model of ART care [9].

However, globally, interventions aimed at improving ART adherence have proven to be more effective in the adult population, while adolescents remain behind [10]. A systematic review conducted in 2015 highlighted that few interventions have been implemented among younger population groups globally [10]. This discrepancy is particularly pronounced in sub-Saharan Africa (SSA), which is home to 70% of the global population living with HIV and where adolescents account for 90% of the global adolescents living with HIV [8,11]. ART outcomes are also poor among the adolescent age group, resulting in 65% ART adherence and 55% VLS among adolescents in the region [12]. Despite these challenges, interventions targeting ART adherence among adolescents living with HIV are limited compared to those available for adults [13,14]. Only a few interventions, such as HIV counseling, community and facility-based support, and family-based economic empowerment programs, have been reported in previously conducted studies [10,15-17].

The potential of digital technology, particularly mobile health (mHealth), in supporting ART adherence among adolescents living with HIV remains largely underexplored, as most studies continue to combine adolescents with young adults [18,19]. For example, a systematic review of mHealth interventions targeting adolescents and young adults identified 8 studies focused on ART adherence in this population [19]. Of these, only 1 study exclusively targeted adolescents; the rest grouped adolescents with young adults, and 1 study even included children [19]. Furthermore, studies have been conducted in only 5 countries within SSA (Nigeria, Kenya, Uganda, South Africa, and Zimbabwe), highlighting the still limited implementation and evaluation of mHealth-based interventions in other high-burden settings such as Eswatini, which has the highest HIV prevalence globally [14,19]. Despite this limited focus, evidence from other African countries indicates that mHealth interventions can significantly improve adherence rates and VLS among young populations [19,20]. The positive influence of mobile phone–based interventions has been bolstered by the widespread accessibility and usage of phones among young people. As of 2024, an estimated 5.4 billion individuals owned mobile phones globally, with a total of 9.1 billion mobile phone subscriptions registered worldwide [21]. On the African continent, approximately 577 million people owned mobile phones, accounting for 1.1 billion mobile subscriptions [21]. Regarding internet access, 5.5 billion people (68% of the global population) were internet users in 2024, including 456 million individuals in Africa [21]. Notably, young people represent the most connected demographic, with 78.5% of global internet users aged between 15 and 24 years. In Africa, young people constituted 53% of internet users [21]. In 2022, Eswatini, a country with 1.3 million people, reported 1.5 million phone subscriptions [22]. Although specific data on phone ownership among adolescents are scant, a systematic review showed that young people and adolescents spend a significant amount of time on their phones, leading to phone addictions [23]. Adolescents have been reported to be particularly addicted to social media and gaming [24,25]. Studies have also reported the positive effect of mobile games in promoting medication adherence [26]. However, there is a need for tailored interventions where the end user actively participates in the design of these interventions [27].

Therefore, access to phones can be leveraged in settings such as Eswatini, where adolescents have poor ART outcomes compared to the adult population. However, there are limited data on phone ownership, accessibility, and usage among adolescents in the country. Additionally, studies have shown the importance of end-user participation in the design of health behavior interventions. We explored mobile phone access, usage patterns, and perceptions of adolescents living with HIV regarding the use of gamified interventions aimed to improve ART adherence in Eswatini.

Methods

Study Design

This was a qualitative study conducted in December 2023 using in-depth interviews (IDIs) with 29 adolescents living with HIV in Eswatini. This approach provided valuable insights from adolescents living with HIV, who have a unique understanding of the types of interventions that may be most effective for them [28].

Study Setting

The study was conducted among adolescents living with HIV accessing ART care through Teen Clubs in the Hhohho region in Eswatini. The Hhohho region is located in the northwestern part of the country and houses the capital city, accounting for 30% of the total population [29]. This region is considered the most economically advanced in Eswatini, as it has the most urbanized population and is home to the headquarters of many corporations, as well as being dominated by the forestry industry, which is among the most important sectors of the country's economy [29]. The region comprises both urban and rural settings and has 71 health facilities, including private, public, military, faith-based, and not-for-profit hospitals, health centers, public health units, and clinics [30]. These health facilities play a critical role in HIV prevention and management

XSL•FO
RenderX

through services offered from the community level to the tertiary level (hospitals). These services include the differentiated models of ART care, such as Teen Clubs [9].

Teen Clubs, a model of ART care for adolescents with disclosed HIV status, use lay health care workers also known as expert clients to provide support and education about ART and HIV [9,31]. Expert clients are HIV-positive adults whose primary role includes providing adherence counseling, tracking clients lost to follow-up, conducting pill counts to monitor medication adherence, and supporting adolescents in navigating psychosocial challenges such as stigma, suicidal ideation, and orphanhood [31]. They work collaboratively with nurses to help alleviate the clinical workload involved in managing adolescents living with HIV [31]. In 2023, there were 99 Teen Clubs with 2308 members in the country (Swaziland National AIDS Programme, National Facility Teen Clubs, Eswatini Ministry of Health, unpublished report, 2023). Of these, 31 Clubs with 1464 members were from the Hhohho region. With the assistance of the Ministry of Health, Swaziland National AIDS Program (SNAP), and the National Paediatrics HIV Care and Treatment department, 5 clubs from the Hhohho region that had been active for at least 2 years were conveniently chosen for data collection [28]. Within each club, at least 5 adolescents living with HIV meeting the inclusion criteria were purposively selected [32]. To ensure the representativeness of perceptions and experiences of all adolescents living with HIV in Teen Clubs, the selected participants included both male and female participants, adolescents from both age groups (10-14 years and 15-19 years), and at least 1 peer educator from each club [32]. Peer educators are also known as community adherence to treatment support (CATS) and are involved in educating and supporting their peers using door-to-door campaigns, especially for adolescents living with HIV lost to follow-up or with poor ART outcomes [33]. The inclusion criteria required adolescents aged 10 to 19 years who were aware of their HIV status and had been members of the Teen Club for at least 6 months at the time of data collection to ensure a comprehensive ART history.

Participant Recruitment Process

Expert clients assisted in the identification of potential participants and were referred to the researcher for further explanation of the study. Adolescents who were willing to be part of the study and met the inclusion criteria were given a detailed study information sheet in both siSwati and English. Adolescents aged below 18 years were given the information sheet, along with the researcher's contact information, consent, and assent forms to share with their guardians. Adolescents aged 18 years provided their voluntary consent to participate in the study. Interviews took place in December 2023 following the receipt of signed consent and assent forms. A total of 29 adolescents meeting our inclusion criteria were recruited. The sample size was guided by the principle of data saturation in conducting qualitative studies [32]. The sample size for achieving data saturation was informed by previous studies conducted among similar populations, using comparable data collection tools, specifically semistructured interview guides [34,35]. Data saturation was achieved when no new relevant information emerged from the participants [28]. Data saturation was assessed by evaluating the repetitiveness of responses. Specifically, data collection continued until no new or significant information emerged from the interviews.

Data Collection and Management

An IDI guide was developed based on an extensive literature review of studies examining barriers and facilitators of ART adherence, as well as interventions targeting ART adherence among adolescents living with HIV. The interview guide included topics on ART and ART adherence experience, mobile phone accessibility and usage patterns, barriers and facilitators to ART adherence, and perceptions of the use of mobile phones to facilitate ART adherence (Multimedia Appendix 1). To ensure the validity of the questions, the interview guide was pilot-tested through 2 interviews with adolescents living with HIV who were not included in the final sample [28]. The interview guide, along with all consent and assent supporting materials, was translated into siSwati to ensure comprehension by all participants. Consequently, the IDIs were conducted in either siSwati or English, depending on the participants' preferences, and were audio-recorded. Each interview lasted about 45 minutes, and interviews were conducted in private rooms provided by each health facility or other locations of participants' preference. To protect participants' identities and ensure confidentiality, personal identifiers, such as names, were not used during interviews.

The audio recordings were transcribed verbatim, and those in siSwati were translated into English. Personal identifiers of participants were not included in both the transcripts and audio files; instead, codes were used to identify participants to maintain confidentiality. Both the transcripts and audio recordings were securely stored on a password-protected computer.

Data Analysis

A deductive-inductive coding approach was used for data analysis [36]. The initial code frames were developed based on the research question being "Can mobile games be used to facilitate ART adherence?", while additional themes emerged inductively from the participants' responses. A code framework was developed, and transcripts were coded manually [28,32]. The code book was reviewed by 2 independent investigators (LDH and CSS) for comprehensiveness and appropriateness [37]. Differences were discussed and resolved collaboratively by the 2 investigators (LDH and CSS), and the final codebook was reviewed and validated by PSN.

Ethical Considerations

Approval to conduct the study was sought from the Stellenbosch University Health Research Ethics Committee (reference number: S24/02/037) and Eswatini Health and Human Research Review Board of the Ministry of Health (protocol reference number: EHHRRB143/2023). Furthermore, a letter of support from hospital management teams was obtained from the Ministry of Health, Department of Health Services. Finally, all participants and guardians in case of participants below the age of 18 were given detailed information on the study and provided consent and assent. Participants were informed that participation in the study was voluntary, and they were free to withdraw from the study at any point.

XSL•FO
RenderX

Results

Demographic Characteristics of Participants

A total of 29 participants (n=14 males and n=15 females) were included in the study. The majority (n=19, 65.5%) of the participants were aged 15 and above. The mean age of the participants was 16.5 (SD 1.08) years, with an average of 10.8 (SD 3.99) years on ART and 5 (SD 3.99) years of Club membership. The majority (n=23, 79%) of the participants were diagnosed with HIV from birth, while only 2 (7%) were diagnosed later in life (Table 1).

Table . Characteristics of participants.

Characteristics	Values, n (%)
Age group (y)	
10 - 14	10 (34.5)
15 - 19	19 (65.5)
Sex	
Female	15 (51.7)
Male	14 (48.3)
Education level	
Primary	9 (31.0)
High school	18 (62.1)
Tertiary	2 (6.9)
Age at diagnosis	
From birth	23 (79.3)
While young	4 (13.8)
Later in life	2 (6.9)
Duration on ART[a] (y)	
<5	2 (6.9)
5-10	15 (51.7)
>10	12 (41.4)
Duration in club (y)	
<2	6 (20.7)
2-5	12 (41.4)
>5	11 (37.9)

[a]ART: antiretroviral therapy.

Phone Ownership and Usage

A total of 28 (96.6%) participants either owned (n=19, 65.5%) or shared (n=9, 31%) a phone with their guardian or siblings. Among those who owned or had access to a phone, 16 (57.1%) used a smartphone, while 12 (42.9%) used a feature phone (Table 2). The majority (17/19; 89.5%) of those who owned phones were older adolescents (15-19 years old), and younger adolescents (7/9; 77.8%) were most likely to share the phone with either guardian or siblings. Only older adolescents reported uncontrolled time on the phone, while younger adolescents had their phone usage controlled by guardians.

XSL•FO
RenderX

Table . Phone ownership and usage characteristics of participants.

Characteristics	Age 10-14 years, n (%)	Age 15-19 years, n (%)	Total, n (%)
Phone ownership			
Yes	2 (20.0)	17 (89.5)	19 (65.5)
No	1 (10.0)	0 (0)	1 (3.4)
Shared	7 (70.0)	2 (10.5)	9 (31.0)
Type of phone			
Smartphone	4 (40)	12 (63.2)	16 (55.2)
Feature or basic phone	5 (50)	7 (36.8)	12 (41.4)
N/A[a]	1 (10)	0 (0)	1 (3.4)
Phone usage[b]			
Daily	6 (66.7)	19 (100)	25 (89.3)
Controlled time	3 (33.3)	0 (0)	3 (10.7)
Time spent on phone[b]			
Uncontrolled time	0 (0)	13 (68.4)	13 (46.4)
<1 hour per day	5 (55.6)	2 (10.5)	7 (25.0)
1 hour to 5 hours daily	1 (11.1)	4 (21.1)	5 (17.9)
Only on weekends and school holidays	3 (33.3)	0 (0)	3 (10.7)
Activities[b]			
Calling and receiving	9 (100)	19 (100)	28 (100)
Social media (WhatsApp, Tik-Tok, Instagram, Facebook)	3 (33.3)	12 (63.2)	15 (53.6)
Games	2 (22.2)	5 (26.3)	7 (25.0)
Researching	1 (11.1)	2 (10.5)	3 (10.7)

[a]N/A: not applicable.

[b]Number of participants with phone access.

Themes on the Perceptions of Gamified Interventions Among Adolescents Living With HIV

Overview

Three themes reflecting the perceptions of adolescents living with HIV on the use of gamified interventions to improve ART adherence emerged in this study. These were (1) feasibility of gamified interventions, (2) qualities of gamified interventions, and (3) concerns about the use of gamified interventions to enhance ART adherence (Table 3). Detailed themes with the supporting codes are provided in Multimedia Appendix 2.

XSL•FO
RenderX

Table . Themes, subthemes, and codes on the use of gamified interventions to enhance antiretroviral therapy (ART) adherence among adolescents living with HIV.

Themes and subthemes	Code
Gamified interventions are feasible	
Phone access and use	• Previous game exposure • Positive game experience
Qualities of gamified interventions	
Supportive	• Reminders • Support group
Educational	• Benefits of ART • Living with HIV
Secure and confidential connection	• Dependable or trustworthy • Prevention of unintended disclosure
Potential concerns about gamified interventions	
Confidentiality	• Unintended disclosure
Phone access factors	• Phone costs • Internet bundles cost • Network availability

Gamified Intervention Feasibility

Most participants reported that the use of gamified interventions to support ART adherence among adolescents living with HIV was feasible, given their access to mobile phones and previous exposure to games. Thus, participants reported that the success of gamified interventions is not only based on mobile phone access but also on having prior experience of playing other mobile games:

> *You see auntie, we play lots of games so this can work too but then like myself I can only play it once in a while you see because I don't own the phone … but it can help.* [Female adolescent living with HIV, aged 15]

Additionally, the practicality of a gamified intervention was perceived to be influenced by the participants' experiences with games, particularly when those experiences were positive. A participant also further noted that a gamified intervention might be more appealing to younger adolescents than older adolescents:

> *I think it can work auntie especially among the young ones because those ones enjoy playing games… myself, I am not a game person but into blogs and other social media platforms because I don't find games interesting you see.* [Male adolescent living with HIV and peer educator, aged 19]

Qualities of Gamified Intervention

Overview

Participants reported that the gamified interventions should possess certain qualities for them to yield positive results in improving ART adherence among adolescents living with HIV. The qualities of a gamified intervention included being able to provide support to the players, being educational, and finally ensuring user HIV status privacy.

Supportive

First, most participants reported that a game aimed at improving ART adherence should offer support to the players, as they identified support as a key facilitator for ART adherence:

> *You see auntie, some of us will benefit from a mobile game if it provides support as it's hard to live with HIV especially where everyone else is negative … It will be like you have something you can do and like someone who understands you and when you play you are happy.* [Male adolescent living with HIV, aged 18]

A mobile game that acts as a reminder for taking their medication would be very vital, as participants indicated that they often forget to take their medication on time:

> *You see auntie, a game where like when you play you learn in the process. Like when you play you can even remember to take your pill and know why it is important.* [Female adolescent living with HIV, aged 18]

Finally, some participants revealed that a multiplayer game could provide support through networking and sharing of experiences, especially if the game can be played in real-time with others:

> *You know a game where I can be in my home and play with someone who is in Manzini, that one can work and even better if I know the person is also living with HIV and we can even chat outside the game even though we don't know each other … sometimes talking to a stranger is better auntie*

XSL•FO
RenderX

> *because they won't gossip about you.* [Male adolescent living with HIV, aged 19]
>
> *You know those racing games auntie where you have more than one player and even better if you know the players are real … those games are the best.* [Male adolescent living with HIV, aged 17]

Educational

Participants reported that a mobile game with an educational component could benefit adolescents living with HIV on ART and improve adherence. They explained that the educational benefits would include information on the advantages of ART and guidance on living with HIV. This information would be critical in teaching players the importance of ART adherence and encouraging them to persevere despite life distractions:

> *For me a game where you are fighting using a gun and when you shoot it's an indication that what happens or the results of taking your pills.* [Female adolescent living with HIV, aged 17]
>
> *[T]he game can work for us like encourage us while teaching us about the pills That's what I can say … it can be a game like avatar you see where if you fall at this point, even your life falls … yes auntie like reminds us that even if you forget your pill, don't give up like the game when they shoot you lose a life but when you keep fighting you gain more lives, something like that.* [Male adolescent living with HIV, aged 19]

Additionally, an educational game could be crucial in guiding adolescents living with HIV as they make life choices that influence their future. These choices include staying positive about life and teaching others that living with HIV does not mean a shortened lifespan; one can still live and enjoy their life:

> *I think a game that will teach people when they play about life with HIV so that they know more and the things to do to make sure one stays well.* [Female adolescent living with HIV, aged 14]
>
> *[S]ometimes you have questions about HIV in general and you have to wait for the day you go to the health facility, you see the game can be helpful if it can answer some of the questions and you can even play with other people like children so they know more about HIV because now it's like HIV means one will die … like I use poetry to teach my friends and imagine if it's a game and we play and they learn.* [Female adolescent living with HIV, aged 17]

Secure and Confidential Connection

Participants emphasized the importance of privacy and confidentiality in mobile games, especially those incorporating social connection features. While they expressed interest in multiplayer games, they strongly preferred designs that would safeguard their privacy and prevent unintended disclosure of their HIV status during gameplay. Participants indicated a preference for playing with strangers, believing this would better protect their identity. However, if the game allowed communication between players, they suggested it would be more acceptable if the other players were also living with HIV. This would create a safe space for sharing experiences without fear of stigma, gossip, or unwanted disclosure:

> *You know a game where I can be in my home and play with someone who is in Manzini, that one can work and even better if I know the person is also living with HIV and we can even chat outside the game even though we don't know each other … sometimes talking to a stranger is better auntie because they won't gossip about you.* [Male adolescent living with HIV, aged 19]

Potential Concerns About the Gamified Interventions

Overview

Participants reported concerns regarding a mobile game aimed at supporting ART adherence among adolescents living with HIV. These concerns were critical to consider in the design of the intervention to ensure its success. Participants highlighted 2 main concerns: the possibility of the intervention leading to unintended disclosure of their HIV status and factors associated with phone access and use, which include the financial implications and network access.

Unintended Disclosure

Most participants revealed that inasmuch as they believed a mobile game can assist in supporting ART adherence, there were concerns regarding unintended disclosure if the name or some of the information could be traced to HIV or people living with HIV:

> *We play games like me I normally play games from my mom's phone, and I enjoy it a lot (laughs) but then auntie what if someone sees the game and can tell that it's about HIV and now, they will be talking about my status.* [Female adolescent living with HIV, aged 17]
>
> *A game one I can play and even if someone can see but will not be able to tell it's for people living with HIV … like my cousins can also play it without them knowing it's for us.* [Female adolescent living with HIV, aged 14]

Mobile Phone Access Factors

Mobile phone access was identified by participants as a key concern in the potential success of gamified interventions. Specifically, participants raised concerns about smartphone ownership, noting that while they had access to basic or feature phones, they did not own smartphones, which are typically required to run gamified apps:

> *Mobile games can work auntie, but problem is like myself, I don't have a phone and the one I use is not a smart phone so this will not benefit some of us.* [Female adolescent living with HIV, aged 15]

Second, participants expressed concerns that even if the games could work, the cost of internet bundles would deter their success. Most adolescents do not have income and would be unable to afford the cost of internet or data bundles that would be required for online games:

XSL•FO
RenderX

> *You see auntie, some of the adolescents here cannot afford to buy internet bundles and you know some games need you to be online.* [Male adolescent living with HIV and peer educator, aged 19]

Finally, participants cited limited mobile network coverage in some of the geographical locations in the country as making it difficult to have access to a mobile phone. They believed that the lack of mobile network coverage would hinder some of them from taking part in the games:

> *I don't know auntie, but you know there are places where there is no network … (laughs) you see at my granny's place, there is no network it's just some odd area right (laughs).* [Female adolescent living with HIV, aged 18.]

Discussion

Principal Findings

The study findings suggest that mobile phone access among adolescents living with HIV in Eswatini may be high, although this was a small qualitative sample, providing a potentially solid foundation for the implementation of mHealth interventions targeting this age group. Currently, mHealth use in the country is largely limited to calls and SMS text message reminders for clinical appointments sent by health care workers [9]. There may be opportunities to expand the scope of mobile phone–based interventions among adolescents.

Older adolescents reported owning smartphones, while younger adolescents often shared phones with siblings or caregivers. This suggests the need for tailored age-appropriate mHealth solutions, with younger adolescents requiring simpler, lower-bandwidth tools compatible with feature phones and shorter usage durations due to parental control, while older adolescents could benefit from more interactive, internet-based interventions as observed in a study conducted in Zimbabwe [38]. However, while smartphone ownership is critical for implementing internet-based mHealth solutions, such interventions require stable internet connectivity, which may be hindered by limited coverage, slow internet speeds, and high costs [38]. This is particularly important in Eswatini, where network and internet access remain limited in some places in the country [39]. As of 2024, only 50.3% of the population had internet access, and the cost of 1 GB of mobile data was US $5.82 [39]. Therefore, given that 58.9% of the population lives below US $3.65 per person per day, mHealth interventions requiring regular internet use should be implemented cautiously to avoid excluding lower-income users [40].

Additionally, phone usage patterns varied by age. All adolescents used phones primarily for calls, while older adolescents were more likely to engage with social media platforms too. This underscores the importance of designing mHealth interventions with adaptable features that align with adolescents' communication habits and preferences [41]. For instance, a study in Uganda found that adolescents preferred not only voice calls but also SMS text messaging and WhatsApp messaging, highlighting the potential value of integrating social connectivity features into mHealth interventions [41]. However, such features should be used with caution in Eswatini due to the cost of mobile services. As of 2024, calls cost approximately US $0.07 per minute, while an SMS text message cost US $0.02 in Eswatini [39].

The study further identified 3 core themes: feasibility, desired qualities, and potential concerns, which are vital to informing the design of gamified interventions. These themes offer valuable insights into the practicality and acceptability of such interventions.

In terms of feasibility, participants expressed confidence in the potential of mHealth, particularly mobile games, to improve ART adherence, despite only a quarter having prior gaming experience. This highlights the importance of considering players' prior experience in the design of gamified interventions. According to Fitzgerald et al [42], games that promote user engagement are typically those that are easy to play, intuitive, and do not require previous gaming skills or knowledge [42]. This is supported by Sardi et al [43], who emphasized that effective games should be adaptive, offering various modes of play to accommodate different player types and skill levels [43]. Consequently, involving users in the design process is critical to ensure that the resulting intervention is user-centered and tailored to their needs, preferences, and abilities [43]. Additionally, the selection of appropriate gaming elements or mechanics is crucial, as these features can significantly influence user engagement and motivation. For instance, storytelling or narrative elements help set clear expectations and create an immersive experience that puts players at ease, enhancing their willingness to engage with the game [44]. Other mechanics, such as onboarding tutorials, rewards, and leaderboards, have been shown to promote sustained engagement by guiding new players and offering in-game incentives that reinforce continued participation [44,45]. Despite participants' limited prior exposure to gaming, existing evidence supports the effectiveness of gamified mHealth interventions in improving medication adherence among adolescents [19]. These findings highlight the feasibility and potential impact of well-designed, user-centered gamified tools in supporting ART adherence among adolescents living with HIV.

Regarding the desired features or qualities of the gamified intervention, the study highlighted the importance of aligning game content and features with the preferences and needs of end users. First, participants emphasized that effective gamified interventions must include relevant, accurate, and engaging educational content, particularly on HIV and the benefits of ART. This finding aligns with a study conducted among adolescents living with HIV in the United States, where the integration of health education into games improved knowledge retention and supported positive behavior change while maintaining user engagement [46].

Second, the study revealed that incorporating a strong support system within the intervention could enhance its effectiveness in promoting ART adherence. Participants identified support in the form of medication reminders and opportunities for networking with other adolescents as essential. Literature consistently highlights the value of such support in improving ART adherence among adolescents living with HIV [47,48]. In

XSL•FO
RenderX

a gamified context, support can be delivered through features such as in-game pop-up messages reminding players of important health behaviors or through social connectivity that allows users to share progress and communicate with peers. Platforms such as social media can offer avenues for users to exchange experiences, provide encouragement, and reinforce positive behaviors. However, the inclusion of social features must be approached with caution. Participants expressed concern that social connectivity within the game could compromise their privacy and lead to unintended disclosure of their HIV status. This concern is echoed in broader literature, where privacy and confidentiality have been flagged as critical risks in mHealth interventions that incorporate social connection and collaboration features [45,49]. Although social connectivity ranks as the second most frequently used game element, it raises substantial concerns about exposing players' identities, particularly among vulnerable populations such as adolescents living with HIV [45]. These findings emphasize the need to design stigma-sensitive and developmentally appropriate interventions [49]. Adolescence is a period marked by emotional and cognitive maturation, yet formal reasoning and risk assessment abilities may still be developing [50].

Therefore, developers must carefully evaluate whether and how to include socially driven elements. Recommended approaches include the use of anonymous avatars and nicknames, while disabling features such as voice and video calls, and restricting media sharing strategies successfully implemented in the design of the PEERNaija game [51]. In the context of Eswatini, the findings suggest that integrating gamified mHealth interventions into existing service delivery structures, such as Teen Clubs, could enhance both effectiveness and user safety. Teen Clubs are trusted, adolescent-friendly platforms that provide safe spaces for sharing experiences and peer support without compromising privacy [52]. Leveraging peer educators within these settings could further enhance the intervention's reach and acceptability [52].

Finally, regarding potential concerns, the study highlighted the critical importance of privacy and confidentiality in the design of gamified interventions. Participants expressed concerns about the risk of unintended disclosure, particularly among younger adolescents who often share mobile phones with siblings or caregivers. Such unintended disclosure could lead to stigma and discrimination, which are well-documented barriers to ART adherence among adolescents living with HIV [53]. Therefore, the game's design, including its name, features, and interface, must be carefully considered to prevent any breach of confidentiality that could negatively affect adolescents' willingness to engage with the intervention [54].

In addition to privacy concerns, participants raised issues related to the financial burden of mobile gaming and limited access to network services. These concerns underscore the need to critically evaluate the economic feasibility of mHealth interventions in low-resource settings. Adolescents must not only own a mobile phone, ideally a smartphone, but also maintain consistent access to mobile data and internet services. This requirement presents a significant challenge in contexts such as Eswatini, where economic constraints and infrastructural barriers, such as poor network coverage, limited electricity, and low digital literacy, can impede the successful implementation of mHealth solutions [38]. Similar challenges have been observed in other low-income settings, where the high cost of smartphones and internet bundles has contributed to poor uptake and limited effectiveness of mHealth-based interventions [38,41,55]. To mitigate these barriers, recent innovations in offline mobile gaming have shown promise. Offline games eliminate the need for constant internet access, thereby reducing the financial burden on users and improving accessibility. Studies have shown that offline interventions are associated with higher uptake and more favorable outcomes, particularly in resource-limited environments [56]. As such, the development of offline gamified interventions could represent a practical and sustainable solution to improve ART adherence among adolescents living with HIV in Eswatini and similar contexts [57].

Nevertheless, for mHealth interventions to be applicable, engaging, feasible, and effective, they should be firmly grounded in behavior change theories [58]. These theories provide a comprehensive framework for understanding adolescent behavior, particularly during the adolescence stage, a developmental period characterized by significant psychological, social, and physiological changes that can influence ART adherence [50,59]. By applying these theories, designers can identify key determinants of adherence and gain valuable insights into how behavior change can be effectively promoted [50]. Behavior change theories also inform the selection of appropriate game elements and design features that resonate with adolescents' needs and preferences, thereby ensuring that user concerns are addressed during the design process [58]. Furthermore, they guide the integration of behavior change techniques that can enhance the intervention's overall impact on ART adherence [50,58,59].

Critically, involving adolescents in the design process is essential. Their participation enables the cocreation of game features that directly reflect their lived experiences, needs, and expectations, thereby increasing the likelihood of sustained engagement and intervention effectiveness [59]. Health behavior models, such as social cognitive theory, the information-motivation-behavioral (IMB) skills model, and motivational interviewing, have been widely applied in mHealth interventions targeting adolescents [50]. These models conceptualize ART adherence as a behavior influenced by a combination of accurate information, individual and social motivation, and the acquisition of behavioral skills. When applied effectively, they enhance adolescents' self-efficacy and lead to improved ART adherence outcomes [50]. Thus, grounding intervention design in these theories, developers can ensure that components are aligned with adolescents' realities and are both contextually relevant and capable of promoting sustainable behavior change [50,58].

Limitations and Strengths

A key limitation of this study is that it only included adolescents from one region of the country, specifically those who are members of Teen Clubs. Additionally, the small sample size is not representative of all adolescents living with HIV in the country or even within the Hhohho region. As a result, the

XSL•FO
RenderX

findings cannot be generalized to all adolescents in Eswatini. Despite these limitations, this study is the first to explore and describe mobile phone access, usage patterns, and perspectives on gamified interventions among Swati adolescents living with HIV. The findings provide valuable baseline data and serve as a foundation for future research aimed at developing and evaluating gamified interventions to enhance ART adherence among adolescents living with HIV.

Conclusion

The study indicates that mobile phone access may be high among adolescents living with HIV in Eswatini, with the majority using their phones daily for various purposes. A considerable proportion of participants expressed a belief in the feasibility of gamified interventions to support ART adherence, highlighting the importance of supportive, educational, and privacy-assured features. Nonetheless, concerns related to confidentiality and economic barriers must be addressed to ensure the successful implementation and long-term sustainability of such interventions.

Given the promise of mHealth interventions, future research should focus on the co-design of gamified interventions rooted in behavior change theories to ensure alignment with their behavioral characteristics, preferences, and lived experiences. These interventions should also be rigorously evaluated for their effectiveness within the Swati context and considered complementary strategies within existing Ministry of Health programs aimed at improving HIV-related outcomes among adolescents.

Acknowledgments

The authors thank the adolescents who participated in the study and shared their experiences.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors aknowledge the use of ChatGPT 4.0 to improve English laguage and grammar in this manuscript. After using the tool, the authors reveiwed and edited the content as needed and take full responsibility for the final product.

Authors' Contributions

Conceptualization: LDH, PSN, CSS
Data curation: LDH
Formal analysis: LDH
Methodology: LDH, PSN, CSS
Supervision: PSN, CSS
Writing – original draft: LDH
Writing – review and editing: PSN, CSS

Conflicts of Interest

None declared.

Multimedia Appendix 1
The interview guide.
[PDF File, 549 KB - mhealth_v14i1e74207_app1.pdf]

Multimedia Appendix 2
Themes and subthemes including exemplary quotes on the use of gamified interventions to enhance antiretroviral therapy adherence among adolescents living with HIV.
[PDF File, 220 KB - mhealth_v14i1e74207_app2.pdf]

References

1. Eswatini population-based HIV impact assessment 3 2021 (SHIMS3 2021): final report. : Ministry of Health, Government of the Kingdom of Eswatini; 2023 URL: https://phia.icap.columbia.edu/wp-content/uploads/2023/12/241123_SHIMS_ENG_RR3_Final-1.pdf [accessed 2026-07-04]
2. Eswatini surpasses UNAIDS fast-track targets for treatment and viral suppression. United States Department of State. 2022. URL: https://phia.icap.columbia.edu/eswatini-surpasses-unaids-fast-track-targets-for-treatment-and-viral-suppression/ [accessed 2026-07-04]
3. Swaziland HIV incidence measurement survey 2: a population-based HIV impact assessment SHIMS2 2016–2017. : Government of the Kingdom of Eswatini (Ministry of Health & CSO); 2017 URL: http://phia.icap.columbia.edu/wp-content/uploads/2017/11/Swaziland_new.v8.pdf [accessed 2026-07-04]

XSL•FO
RenderX

4. Nkambule R, Nuwagaba-Biribonwoha H, Mnisi Z, Ao T, et al. Substantial progress in confronting the HIV epidemic in Swaziland: first evidence of national impact [Abstract: MOAX0204LB]. Presented at: 9th International AIDS Society Conference on HIV Science (IAS 2017); Jul 23-26, 2017 URL: https://www.iasociety.org/sites/default/files/Conference%20reports%202011/IAS2017_conference_report.pdf [accessed 2026-07-04]
5. Byrd KK, Hou JG, Hazen R, et al. Antiretroviral adherence level necessary for HIV viral suppression using real-world data. J Acquir Immune Defic Syndr 2019 Nov 1;82(3):245-251. [doi: 10.1097/QAI.0000000000002142] [Medline: 31343455]
6. Cohen MS, Chen YQ, McCauley M, et al. Antiretroviral therapy for the prevention of HIV-1 transmission. N Engl J Med 2016 Sep 1;375(9):830-839. [doi: 10.1056/NEJMoa1600693] [Medline: 27424812]
7. Rodger AJ, Cambiano V, Bruun T, et al. Sexual activity without condoms and risk of HIV transmission in serodifferent couples when the HIV-positive partner is using suppressive antiretroviral therapy. JAMA 2016 Jul 12;316(2):171-181. [doi: 10.1001/jama.2016.5148] [Medline: 27404185]
8. Global HIV & AIDS statistics—fact sheet. UNAIDS (Joint United Nations Programme on HIV/AIDS). 2023. URL: https://www.unaids.org/en/resources/fact-sheet [accessed 2026-07-04]
9. Guidelines for differentiated service delivery. : Eswatini Ministry of Health (MoH); 2022 URL: https://cquin.icap.columbia.edu/wp-content/uploads/2022/07/Final_Policy-Guidelines-for-Eswatini-Differentiated-Service-Delivery.pdf [accessed 2026-07-04]
10. Mbuagbaw L, Sivaramalingam B, Navarro T, et al. Interventions for enhancing adherence to antiretroviral therapy (ART): a systematic review of high quality studies. AIDS Patient Care STDS 2015 May;29(5):248-266. [doi: 10.1089/apc.2014.0308] [Medline: 25825938]
11. Kharsany ABM, Karim QA. HIV infection and AIDS in sub-Saharan Africa: current status, challenges and opportunities. Open AIDS J 2016;10:34-48. [doi: 10.2174/1874613601610010034] [Medline: 27347270]
12. Hlophe LD, Tamuzi JL, Shumba CS, Nyasulu PS. Barriers and facilitators to anti-retroviral therapy adherence among adolescents aged 10 to 19 years living with HIV in sub-Saharan Africa: a mixed-methods systematic review and meta-analysis. PLoS One 2023;18(5):e0276411. [doi: 10.1371/journal.pone.0276411] [Medline: 37200399]
13. Munyayi FK, van Wyk B, Mayman Y. Interventions to improve treatment outcomes among adolescents on antiretroviral therapy with unsuppressed viral loads: a systematic review. Int J Environ Res Public Health 2022 Mar 25;19(7):3940. [doi: 10.3390/ijerph19073940] [Medline: 35409621]
14. Griffee K, Martin R, Chory A, Vreeman R. A systematic review of digital interventions to improve ART adherence among youth living with HIV in sub-Saharan Africa. AIDS Res Treat 2022;2022:9886306. [doi: 10.1155/2022/9886306] [Medline: 36199816]
15. Okonji EF, Mukumbang FC, Orth Z, Vickerman-Delport SA, Van Wyk B. Psychosocial support interventions for improved adherence and retention in ART care for young people living with HIV (10-24 years): a scoping review. BMC Public Health 2020 Dec 1;20(1):1841. [doi: 10.1186/s12889-020-09717-y] [Medline: 33261566]
16. Casale M, Carlqvist A, Cluver L. Recent interventions to improve retention in HIV care and adherence to antiretroviral treatment among adolescents and youth: a systematic review. AIDS Patient Care STDS 2019 Jun;33(6):237-252. [doi: 10.1089/apc.2018.0320] [Medline: 31166783]
17. Reif LK, Abrams EJ, Arpadi S, et al. Interventions to improve antiretroviral therapy adherence among adolescents and youth in low- and middle-income countries: a systematic review 2015-2019. AIDS Behav 2020 Oct;24(10):2797-2810. [doi: 10.1007/s10461-020-02822-4] [Medline: 32152815]
18. Cooper V, Clatworthy J, Whetham J, Consortium E. mHealth interventions to support self-management in HIV: a systematic review. Open AIDS J 2017;11(1):119-132. [doi: 10.2174/1874613601711010119] [Medline: 29290888]
19. Goldstein M, Archary M, Adong J, et al. Systematic review of mHealth interventions for adolescent and young adult HIV prevention and the adolescent HIV continuum of care in low to middle income countries. AIDS Behav 2023 May;27(Suppl 1):94-115. [doi: 10.1007/s10461-022-03840-0] [Medline: 36322217]
20. Chory A, Callen G, Nyandiko W, et al. A pilot study of a mobile intervention to support mental health and adherence among adolescents living with HIV in western Kenya. AIDS Behav 2022 Jan;26(1):232-242. [doi: 10.1007/s10461-021-03376-9] [Medline: 34292429]
21. Measuring digital development: facts and figures 2024. : International Telecommunication Union (ITU); 2024 URL: https://www.itu.int/itu-d/reports/statistics/wp-content/uploads/sites/5/2024/11/2402588_1e_Measuring-digital-development-Facts-and-Figures-2024_v4.pdf [accessed 2026-07-04]
22. Swaziland Number of Subscriber Mobile. CEIC. 2024. URL: https://www.ceicdata.com/en/indicator/eswatini/number-of-subscriber-mobile [accessed 2026-07-04]
23. Sahu M, Gandhi S, Sharma MK. Mobile phone addiction among children and adolescents: a systematic review. J Addict Nurs 2019;30(4):261-268. [doi: 10.1097/JAN.0000000000000309] [Medline: 31800517]
24. Hwang Y, Park N. Is smartphone addiction comparable between adolescents and adults? Examination of the degree of smartphone use, type of smartphone activities, and addiction levels among adolescents and adults. Int Telecommun Policy Rev 2017;24(2):59-75 [FREE Full text]
25. Jeong SH, Kim H, Yum JY, Hwang Y. What type of content are smartphone users addicted to?: SNS vs. games. Comput Human Behav 2016 Jan;54:10-17. [doi: 10.1016/j.chb.2015.07.035]

XSL•FO
RenderX

26. Majeed-Ariss R, Baildam E, Campbell M, et al. Apps and adolescents: a systematic review of adolescents' use of mobile phone and tablet apps that support personal management of their chronic or long-term physical conditions. J Med Internet Res 2015 Dec 23;17(12):e287. [doi: 10.2196/jmir.5043] [Medline: 26701961]
27. Abraham O, LeMay S, Bittner S, Thakur T, Stafford H, Brown R. Investigating serious games that incorporate medication use for patients: systematic literature review. JMIR Serious Games 2020 Apr 29;8(2):e16096. [doi: 10.2196/16096] [Medline: 32347811]
28. Grove SK, Gray JR. Understanding Nursing Research: Building an Evidence-Based Practice, 7th edition: Elsevier; 2019.
29. Eswatini national development plan 2023/24–2027/28. : Ministry of Economic Planning and Development, Government of the Kingdom of Eswatini; 2024 URL: https://www.gov.sz/images/planningministry/National-Development--Plan-2023-2028.pdf [accessed 2026-07-10]
30. Service availability and readiness assessment 2017. : Ministry of Health, Kingdom of Eswatini (with WHO support); 2017 URL: https://data-archive.hhfa.online/index.php/catalog/85/related-materials [accessed 2026-07-04]
31. Ahmed CV, Weissinger G, Teitelman A, et al. Expert client service delivery practices among adolescents living with HIV in Eswatini: a thematic analysis. Child Youth Serv Rev 2022 Jan;132:106309. [doi: 10.1016/j.childyouth.2021.106309]
32. Creswell JW, Creswell JD. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, 5th edition: Sage; 2017.
33. National Policy Guidelines FOR Community-Centred Models of ART Service Delivery (CommART) in Swaziland. : Swaziland National AIDS Programme (SNAP), Ministry of Health, Kingdom of Swaziland; 2016 URL: https://www.differentiatedservicedelivery.org/wp-content/uploads/Swaziland-Policy-Guideline-2016-1-1.pdf [accessed 2026-07-04]
34. Audi C, Jahanpour O, Antelman G, et al. Facilitators and barriers to antiretroviral therapy adherence among HIV-positive adolescents living in Tanzania. BMC Public Health 2021 Dec 13;21(1):2274. [doi: 10.1186/s12889-021-12323-1] [Medline: 34903209]
35. Kamangu JWN, Mboweni SH. Factors impacting ART adherence among HIV-positive older adolescents and younger adults in Namibia: a qualitative analysis. Open Public Health J 2024 May 23;17(1). [doi: 10.2174/0118749445299654240402033559]
36. Bingham AJ. From data management to actionable findings: a five-phase process of qualitative data analysis. Int J Qual Methods 2023 Oct;22. [doi: 10.1177/16094069231183620]
37. Cypress BS. Rigor or reliability and validity in qualitative research: perspectives, strategies, reconceptualization, and recommendations. Dimens Crit Care Nurs 2017;36(4):253-263. [doi: 10.1097/DCC.0000000000000253] [Medline: 28570380]
38. Doyle AM, Bandason T, Dauya E, et al. Mobile phone access and implications for digital health interventions among adolescents and young adults in Zimbabwe: cross-sectional survey. JMIR mHealth uHealth 2021 Jan 13;9(1):e21244. [doi: 10.2196/21244] [Medline: 33439136]
39. Information and communications technology-sector report. : Eswatini Communications Commission (ESCCOM); 2024 URL: https://www.esccom.org.sz/publications/reports/docs/Esccom sector report.pdf [accessed 2026-07-04]
40. Eswatini. World Bank Group. 2025. URL: https://www.worldbank.org/en/country/eswatini/overview [accessed 2026-07-04]
41. Tarantino N, Lartey M, Arnold T, Brown L, Kwara A, Guthrie K. Preferences for a game-based SMS adherence intervention among young people living with HIV in Ghana: a qualitative study. AIDS Behav 2022 Jul;26(7):2169-2181. [doi: 10.1007/s10461-021-03562-9] [Medline: 35066732]
42. Fitzgerald M, Ratcliffe G. Serious games, gamification, and serious mental illness: a scoping review. Psychiatr Serv 2020 Feb 1;71(2):170-183. [doi: 10.1176/appi.ps.201800567] [Medline: 31640521]
43. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Jul;71:31-48. [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]
44. Howard C, Bevins K. Game mechanics and why they are employed: what we know about gamification so far. Issues Trends Educ Technol 2018 May 1;6(1). [doi: 10.2458/azu_itet_v6i1_bevins]
45. Garett R, Young SD. Health care gamification: a study of game mechanics and elements. Tech Know Learn 2019 Sep;24(3):341-353. [doi: 10.1007/s10758-018-9353-4]
46. Whiteley L, Brown LK, Mena L, Craker L, Arnold T. Enhancing health among youth living with HIV using an iPhone game. AIDS Care 2018;30(sup4):21-33. [doi: 10.1080/09540121.2018.1503224] [Medline: 30626196]
47. Ankrah DN, Koster ES, Mantel-Teeuwisse AK, Arhinful DK, Agyepong IA, Lartey M. Facilitators and barriers to antiretroviral therapy adherence among adolescents in Ghana. Patient Prefer Adherence 2016;10:329-337. [doi: 10.2147/PPA.S96691] [Medline: 27042024]
48. Galea JT, Wong M, Muñoz M, et al. Barriers and facilitators to antiretroviral therapy adherence among Peruvian adolescents living with HIV: a qualitative study. PLoS One 2018;13(2):e0192791. [doi: 10.1371/journal.pone.0192791] [Medline: 29447226]
49. Al-Rayes S, Al Yaqoub FA, Alfayez A, et al. Gaming elements, applications, and challenges of gamification in healthcare. Inform Med Unlocked 2022;31:100974. [doi: 10.1016/j.imu.2022.100974]
50. Bezabih AM, Gerling K, Abebe W, Abeele VV. Behavioral theories and motivational features underlying eHealth interventions for adolescent antiretroviral adherence: systematic review. JMIR mHealth uHealth 2021 Dec 10;9(12):e25129. [doi: 10.2196/25129] [Medline: 34890353]

XSL•FO
RenderX

51. Ahonkhai AA, Pierce LJ, Mbugua S, et al. PEERNaija: a gamified mHealth behavioral intervention to improve adherence to antiretroviral treatment among adolescents and young adults in Nigeria. Front Reprod Health 2021;3:656507. [doi: 10.3389/frph.2021.656507] [Medline: 35237765]
52. Munyayi FK, van Wyk B. The effects of teen clubs on retention in HIV care among adolescents in Windhoek, Namibia. South Afr J HIV Med 2020;21(1):1031. [doi: 10.4102/sajhivmed.v21i1.1031] [Medline: 32158557]
53. Hightow-Weidman L, Muessig K, Knudtson K, et al. A gamified smartphone app to support engagement in care and medication adherence for HIV-positive young men who have sex with men (AllyQuest): development and pilot study. JMIR Public Health Surveill 2018 Apr 30;4(2):e34. [doi: 10.2196/publichealth.8923] [Medline: 29712626]
54. Denison JA, Banda H, Dennis AC, et al. "The sky is the limit": adhering to antiretroviral therapy and HIV self-management from the perspectives of adolescents living with HIV and their adult caregivers. J Int AIDS Soc 2015 Jan;18(1):1-6. [doi: 10.7448/IAS.18.1.19358] [Medline: 25591915]
55. Campbell BR, Choi K, Neils MG, et al. Mobile device usage by gender among high-risk HIV individuals in a rural, resource-limited setting. Telemed J E Health 2021 Jun;27(6):615-624. [doi: 10.1089/tmj.2020.0218] [Medline: 32835638]
56. Whiteley LB, Olsen EM, Haubrick KK, Odoom E, Tarantino N, Brown LK. A review of interventions to enhance HIV medication adherence. Curr HIV/AIDS Rep 2021 Oct;18(5):443-457. [doi: 10.1007/s11904-021-00568-9] [Medline: 34152554]
57. Côté J, Rouleau G, Ramirez-Garcia MP, Auger P, Thomas R, Leblanc J. Effectiveness of a web-based intervention to support medication adherence among people living with HIV: web-based randomized controlled trial. JMIR Public Health Surveill 2020 Apr 20;6(2):e17733. [doi: 10.2196/17733] [Medline: 32310145]
58. Voorheis P, Zhao A, Kuluski K, et al. Integrating behavioral science and design thinking to develop mobile health interventions: systematic scoping review. JMIR mHealth uHealth 2022 Mar 16;10(3):e35799. [doi: 10.2196/35799] [Medline: 35293871]
59. Amico KR, Mugavero M, Krousel-Wood MA, Bosworth HB, Merlin JS. Advantages to using social-behavioral models of medication adherence in research and practice. J Gen Intern Med 2018 Feb;33(2):207-215. [doi: 10.1007/s11606-017-4197-5] [Medline: 29204968]

Abbreviations

ART: antiretroviral therapy
CATS: community adherence to treatment support
IDI: in-depth interview
IMB: information-motivation-behavioral
mHealth: mobile health
SNAP: Swaziland National AIDS Program
SSA: sub-Saharan Africa
UNAIDS: Joint United Nations Programme on HIV/AIDS
VLS: viral load suppression

Edited by Z Ni; submitted 19.Mar.2025; peer-reviewed by A Adeoye, S Harrison; revised version received 18.Jul.2025; accepted 17.Aug.2025; published 24.Jul.2026.

Please cite as:
Hlophe LD, Nyasulu PS, Shumba CS
Mobile Phone Access, Usage Patterns, and Perceptions of Adolescents Living With HIV on the Use of Gamified Interventions to Improve Antiretroviral Therapy Adherence in Eswatini: Qualitative Study
JMIR Mhealth Uhealth 2026;14:e74207
URL: https://mhealth.jmir.org/2026/1/e74207
doi:10.2196/74207

© Londiwe D Hlophe, Peter S Nyasulu, Constance S Shumba. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 24.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Vaccination-Related Applications and Health Care Professionals' Observed Changes in Human Papillomavirus Vaccine Hesitancy: Cross-Sectional Survey

Manali Desai[1,2,3], MD, MPH; Shreela Sharma[2], PhD; Joel Fokom-Domgue[1,4], MD, MPH; Robert Yu[3], MCS; Wenyaw Chan[5], PhD; Onyema G Chido-Amajuoyi[6], MD; Charles Darkoh[2], PhD; Sanjay Shete[1,3,4], PhD

[1]Division of Cancer Prevention and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

[2]Center for Healthy Communities, Department of Epidemiology, The University of Texas Health Science Center at Houston School of Public Health, Houston, TX, United States

[3]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, 1515 Holcome Blvd, Houston, TX, United States

[4]Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

[5]Department of Biostatistics and Data Science, The University of Texas Health Science Center at Houston School of Public Health, Houston, TX, United States

[6]Division of Hematology and Oncology, Department of Medicine, University of Washington, Seattle, WA, United States

Corresponding Author:
Sanjay Shete, PhD
Division of Cancer Prevention and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

Abstract

Background: Digital tools are known to promote public health interventions such as vaccine delivery. The recommendation that health care professionals (HCPs) use vaccination-related mobile apps or web-based applications has contributed to improving vaccine awareness and acceptance in the United States. The state of Texas, which has one of the lowest human papillomavirus (HPV) vaccination rates, has seen a significant increase in HPV vaccine hesitancy, particularly during the COVID-19 pandemic.

Objective: This study aimed to examine the association between changes in HPV vaccine hesitancy observed by HCPs among patients in Texas and promotion of vaccination-related applications at the health care facilities where they practiced during the COVID-19 pandemic.

Methods: A population-based cross-sectional survey was administered in 2021 by the MD Anderson Cancer Center to HCPs working in Texas using email addresses obtained from the LexisNexis Master Provider Referential Database. HCPs were asked if they assessed HPV vaccination status during every patient encounter. Those who responded "Often/Always" or "Sometimes" were subsequently asked whether they observed any change ("Decreased," "No change," "Increased," or "Not sure") in HPV vaccine hesitancy during the COVID-19 pandemic. Additionally, HCPs were asked whether their practice offers HPV vaccination. Those who responded "Yes" to this question were further asked whether vaccination-related applications are promoted at the facility where they practice, with response options being "Yes," "No," or "I don't know." Logistic regression analysis was performed to examine the association between changes in HPV vaccine hesitancy observed by HCPs and promotion of vaccination-related applications at the facility where they practice.

Results: A total of 1283 HCPs completed the survey. Of the 730 HCPs who observed changes in HPV vaccine hesitancy, 51 (7%) reported a decrease in their patients' HPV vaccine hesitancy. Of these 730 HCPs, 578 (79.2%) responded to the questions regarding vaccination-related applications, of whom 104 (18%) reported that vaccination-related applications were promoted at their facilities. Compared to HCPs who reported not promoting vaccination-related applications, those who reported doing so at their facilities had significantly higher odds of observing a decrease in HPV vaccine hesitancy among patients (adjusted odds ratio [aOR] 2.48, 95% CI 1.10-5.55; *P*=.03). HCPs working at federally qualified health centers or city, county, or public health care facilities (aOR 4.02, 95% CI 1.33-12.14; *P*=.01) and HCPs who administered the HPV vaccine under standing orders at their facilities (aOR 2.91, 95% CI 1.11-7.63; *P*=.03) had significantly higher odds of observing a decrease in HPV vaccine hesitancy at their practices.

Conclusions: Our findings suggest that promoting vaccination-related applications at health care facilities in areas with high HPV vaccine hesitancy such as Texas could further decrease HPV vaccine hesitancy in the population. This may be potentially applicable across diverse health care settings, particularly in the context of pandemic preparedness.

(JMIR Mhealth Uhealth 2026;14:e77778) doi:10.2196/77778

XSL•FO
RenderX

KEYWORDS

human papillomavirus vaccination; HPV vaccination; vaccine hesitancy; mobile apps or web-based applications; digital health; COVID-19 pandemic; health policy; health care professional; health promotion

Introduction

In the United States, approximately 47,984 new cases of human papillomavirus (HPV)–associated cancers are diagnosed annually, including 26,280 cases among women and 21,704 cases among men [1]. Despite the availability of safe and effective HPV prophylactic vaccines that can help prevent these cancers and the adoption of a standardized recommendation schedule in the United States [2-5], HPV vaccination uptake remains suboptimal due to barriers including lack of knowledge, concerns about safety and efficacy, and cost concerns [6,7]. In response, government agencies have made efforts to improve HPV vaccination coverage through school- and community-based vaccination programs, health care professional (HCP)–based interventions, and cost reduction initiatives [8]. Since it was approved in the United States, the HPV vaccine was included in the Vaccines for Children program, which provides free vaccines to uninsured or underinsured children until the age of 18 years [9-11]. Additionally, the Affordable Care Act mandates that private health insurers cover vaccines recommended by the Advisory Committee on Immunization Practices, including HPV, and are prohibited from charging copays or deductibles for these vaccines when administered by in-network HCPs [12]. Despite these efforts, HPV vaccination coverage among adolescents aged 13 to 17 years was only 62.9% in 2024, which is far below the Healthy People 2030 goal of achieving 80% vaccination coverage [7,13,14].

According to the World Health Organization, vaccine hesitancy, one of the major threats to public health, is defined as a delay in accepting or refusal of safe and effective vaccines despite their availability [15,16]. In the context of HPV vaccination, vaccine hesitancy constitutes a significant barrier to both initiating and completing the HPV vaccination series [17]. The most common reasons for vaccine hesitancy are limited knowledge; beliefs that vaccination is unnecessary; concerns about vaccine safety and potential side effects; and a lack of strong, presumptive HCP recommendations [18-23]. A few studies have shown the benefits of using digital interventions to reduce vaccine hesitancy by increasing public awareness and knowledge and intentions to receive the HPV vaccine [24,25]. A quantitative study conducted among 20 adolescents and 34 parents in Philadelphia, Pennsylvania, assessed the effectiveness of the Vaccipack app, which included a vaccine tracker, inspirational stories, and an active discussion forum, designed to promote intention to receive the HPV vaccine [26]. This app, based on an integrated behavioral model, was found to have high acceptability among parents (88%) and adolescents (75%) and to be beneficial for them (82% and 85%, respectively) [26]. Another qualitative study assessing the Vax4HPV mobile app designed based on the social ecological theory was conducted among parents of children and adolescents aged 9 to 14 years in Connecticut to evaluate its benefits in reducing HPV vaccine hesitancy through informed decision-making [27]. This study reported that the implementation of this theory-based, user-centered app helped mitigate hesitancy by bridging the gap between intention to vaccinate and vaccine uptake [27]. Participants reported that the Vax4HPV app provided gender-neutral information, stronger HCP recommendations, and information on clinics providing free vaccination or scheduling appointments [27]. Additionally, a study conducted in New Mexico among 82 parents of vaccine-eligible teenagers found that mobile apps such as Vacteens reduced vaccine hesitancy among parents and their adolescent daughters aged 11 to 14 years [28]. Similarly, a clinic-based study conducted in Houston, Texas, reported that the HPVCancerFree app served as an adjunct during health care visits, providing education and support to help parents make evidence-based decisions about HPV vaccination [29]. A randomized trial among 209 parents and their sons aged 11 to 14 years evaluated the effectiveness of a tailored and engaging mobile app (TeenVac) compared to the Centers for Disease Control and Prevention (CDC) HPV vaccination pamphlet available online [30]. This study showed that using the TeenVac app led to higher completion of the HPV vaccine series than using the pamphlet [30].

Texas is the second-largest state in the United States, with one of the lowest up-to-date HPV vaccination rates (52.2%) across the country [31]. Access to preventive health care during the COVID-19 pandemic was impacted due to closed medical offices, limited availability of appointments, and concerns regarding contracting COVID-19 [32]. A study conducted among 252 parents at a children's hospital in California reported increased hesitancy regarding childhood vaccines during the COVID-19 pandemic [33]. Furthermore, a survey conducted among HCPs in Texas found that 17.1% of HCPs reported an increase in HPV vaccine hesitancy, whereas another study of pediatricians in Massachusetts found that 11% of pediatricians reported an increase in parental HPV vaccine hesitancy during the pandemic [34,35]. Compared to 2018 and 2019, substantially fewer childhood (9-12 years) and adolescent (13-17 years) vaccines were administered from March 2020 to September 2020 (ie, during the acute phase of the COVID-19 pandemic) [36]. While prior studies have reported the benefits of HPV vaccination–related applications outside major outbreaks or public health emergencies [26-29], little is known about the role that these applications may play during outbreaks such as COVID-19 or other pandemics. In this study, we examined the association between the promotion of vaccination-related applications at health care facilities and changes observed by HCPs regarding HPV vaccine hesitancy during the COVID-19 pandemic. We hypothesized that HCPs who reported promoting vaccination-related applications at their health care facilities had higher odds of observing a decrease in HPV vaccine hesitancy. Grounded in the diffusion of innovation theory, this study's findings will support the integration of digital health interventions into the package of pandemic preparedness activities in the United States.

XSL•FO
RenderX

Methods

Study Design, Data Source, and Population

We conducted a population-based cross-sectional study using data from a statewide survey collected between January 2021 and April 2021 by the University of Texas MD Anderson Cancer Center. This survey was developed to evaluate Texas-based HCPs' knowledge, perceptions, and practices regarding the HPV vaccine recommendation and was administered electronically in English and Spanish using all email addresses obtained from the LexisNexis Master Provider Referential Database [37]. The survey was sent via REDCap (Research Electronic Data Capture; Vanderbilt University) to Texas-based HCPs practicing family medicine, internal medicine, pediatrics, and gynecology; physician assistants; and nursing staff who held a current Texas license and for whom email addresses were available in the LexisNexis database. Up to 3 email reminders were sent to each participant. In case the survey was not completed after the initial invitation, the first follow-up email was sent on the eighth day (–2 to +2 days) from the initial invitation, the second follow-up email was sent on the eighth day (–2 to +2 days) from the first follow-up email, and the third follow-up email was sent 5 days (–2 to +2 days) after the second follow-up email. The estimated time to complete the survey was 10 minutes. A total of 1283 HCPs completed the survey, with a response rate of 7% (1283/18,328). To report this study, we adhered to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines to ensure ethical and scientific integrity (Checklist 1) [38].

Ethical Considerations

This study was approved by the University of Texas MD Anderson Cancer Center Ethical Review Board (protocol 2019-1257). Participation in this study was voluntary and confidential. All participants provided informed consent. The dataset underwent deidentification through the removal of all identifiable information. Each participant received a US $10 gift card for completing the survey. Additionally, after completing the survey, participants were entered into a draw for a US $1000 Amazon gift card, which was held 3 months after the survey closed.

Measures

Dependent Variable: Changes in HPV Vaccine Hesitancy Among Patients Observed by HCPs During the COVID-19 Pandemic

We examined the changes in HPV vaccine hesitancy observed by HCPs among patients during the COVID-19 pandemic. To capture the subjective impression of hesitancy, HCPs were asked the following: "At every patient encounter, do you assess HPV vaccination status?" Those who answered "Sometimes" or "Often/Always" were further asked the following: "During the COVID-19 pandemic, I have observed that HPV vaccination hesitancy..." The possible responses were "Decreased," "Increased," "No change," and "Not sure." We coded "Decreased" as 1 and "Increased or no change" as 0. Those who responded "Not sure" were excluded from this study. The reference category was defined as "Increased or no change."

Independent Variable: Promotion of Vaccination-Related Applications at the HCPs' Facilities of Practice

For this variable, HCPs were first asked the following: "Does your practice offer HPV vaccination?" Those who answered "Yes, always offered" or "Yes, but not consistently" were assessed for promoting vaccination-related applications at their facilities using the survey question "Which of the following is practiced at your facility? ...Promote vaccination-related apps," with possible responses of "Yes," "No," and "I don't know." The reference category was "No."

HCP-Related Variables (Covariates): HCPs' Sociodemographic Variables and Practice-Related Factors

We included HCPs' age (<35 years, 35 to 54 years, and ≥55 years), sex (male and female), race and ethnicity (non-Hispanic White, non-Hispanic Black, Asian, Hispanic, and other), location of practice (rural and urban) and type of facility (university hospital, teaching hospital, affiliated clinic, or employed physician practices; Federally Qualified Health Center [FQHC] or city, county, or public health care facility; or other), administration of HPV vaccination under standing orders (yes or no), use of reminders (eg, phone or mail) when HPV vaccination was due ("Yes," "No," and "I don't know"), and use of recalls (eg, computerized tracking) when HPV vaccination was past due ("Yes," "No," and "I don't know"). On the basis of the 2013 Rural-Urban Continuum Codes, we classified the location of practice for the HCPs using the zip codes of their workplaces, with Rural-Urban Continuum Codes 1 to 3 classified as urban and 4 to 9 classified as rural [39].

Patient-Related Variables: Patients' Sociodemographic Factors

As we did not collect data from patients directly, patient-related variables were obtained from the 2022 American Community Survey 5-year database [40], which allowed us to determine the race and ethnicity of the population residing in Texas at the level of the zip code where the HCPs practiced. This variable was also included in our model as a potential predictor of changes in HPV vaccine hesitancy observed by HCPs among their patients [41].

Data Analysis

Descriptive statistics were used to estimate the prevalence of promoting the use of vaccination-related applications at the HCPs' facilities and HCP- and practice-level factors stratified by HCPs' observed changes in HPV vaccine hesitancy during the COVID-19 pandemic. To that end, we used frequencies, weighted percentages, 95% CIs, and the Pearson chi-square test or Fisher exact test as appropriate. Bivariate and multivariable logistic regression analyses were used to determine the odds of observed decrease in HPV vaccine hesitancy reported by HCPs and the promotion of vaccination-related applications during the COVID-19 pandemic. In the multivariable logistic regression model, we adjusted for HCPs' age, sex, race and ethnicity, practice location, type of facility, administration of HPV vaccination under standing orders, use of reminders (eg, phone or mail) when HPV vaccination was due, use of recalls (eg, computerized tracking) when HPV vaccination was past due,

XSL•FO
RenderX

and predominant zip code–level race and ethnicity of patients. Descriptive analyses were conducted using the full sample (n=730), whereas regression analyses were based on a reduced sample (n=576) due to missing data for one or more covariates. However, for each covariate with missing data, a "missing" category was created and included in the logistic regression. We considered *P* values <.05 to be statistically significant, and all tests were 2 sided. We conducted our analysis using SAS (version 9.4; SAS Institute).

Results

Demographic Characteristics

Overall, 1283 HCPs completed the survey, with a mean age of 47.1 (SD 11.3) years. Of the 730 HCPs who responded to the survey question about changes in HPV vaccine hesitancy, 51 (7%) reported observing a decrease in HPV vaccine hesitancy, whereas 679 (93%) observed an increase or no change among patients during the COVID-19 pandemic (Table 1). Of the 730 HCPs, 578 (79.2%) responded to the survey question about promoting vaccination-related applications, of whom 104 (18%) reported promoting them at their practices. A higher proportion of HCPs included in this study were aged 35 to 54 years (441/705, 62.6%), female (556/719, 77.3%), non-Hispanic White individuals (369/730, 50.5%), and urban dwellers (699/729, 95.9%). In addition, 50.1% (282/563) used standing orders, 46.3% (274/592) used reminders, and 41% (244/595) used recalls at their facilities to remind eligible individuals to receive the recommended vaccines (Table 1). A higher proportion of HCPs included in this study practiced in zip codes with predominantly non-Hispanic White populations (424/636, 66.7%), followed by those with predominantly Hispanic populations (171/636, 26.9%; Table 1).

The proportion of HCPs who observed a decrease in HPV vaccine hesitancy among patients was higher among those who promoted vaccination-related applications (13/104, 12.5%); those who were aged 55 years or older (15/185, 8.1%), female (41/556, 7.4%), and Hispanic (13/108, 12%); those practicing in rural locations (4/30, 13.3%); those working at FQHCs or city, county, or public health care facilities (15/99, 15.2%); and those using standing orders (23/282, 8.2%) or reminders (18/274, 6.6%; Table 1). A higher proportion of HCPs who observed a decrease in HPV vaccine hesitancy practiced in zip codes with predominantly Hispanic populations (16/171, 9.4%; Table 1).

XSL•FO
RenderX

Table . Descriptive statistics of the observed changes in human papillomavirus vaccine hesitancy by the health care professionals among patients during the COVID-19 pandemic (n=730)[a]. The responses to the following survey question were excluded from this analysis: "During the COVID-19 pandemic, I have observed that...HPV vaccination hesitancy has..." - "Not sure."

Characteristic	Total, n/N (%; 95% CI)	Decreased, n/N (%; 95% CI)	Increased or no change, n/N (%; 95% CI)	*P* value[b]
Promoted vaccination-related applications				*.02*
Yes	104/578 (18.0; 14.9 - 21.1)	13/104 (12.5; 6.1 - 18.9)	91/104 (87.5; 81.1 - 93.9)	
No	402/578 (69.6; 65.8 - 73.3)	21/402 (5.2; 3.0 - 7.4)	381/402 (94.8; 92.6 - 97.0)	
I don't know	72/578 (12.5; 9.8 - 15.2)	3/72 (4.2; 0.0 - 8.8)	69/72 (95.8; 91.2 - 100.0)	
Age (y)				.64
<35	79/705 (11.2; 8.9 - 13.5)	6/79 (7.6; 1.7 - 13.5)	73/79 (92.4; 86.5 - 98.3)	
35-54	441/705 (62.6; 59.0 - 66.1)	27/441 (6.1; 3.9 - 8.4)	414/441 (93.9; 91.6 - 96.1)	
≥55	185/705 (26.2; 23.0 - 29.5)	15/185 (8.1; 4.2 - 12.1)	170/185 (91.9; 87.9 - 95.8)	
Sex				.59
Female	556/719 (77.3; 74.3 - 80.4)	41/556 (7.4; 5.2 - 9.6)	515/556 (92.6; 90.4 - 94.8)	
Male	163/719 (22.7; 19.6 - 25.7)	10/163 (6.1; 2.4 - 9.8)	153/163 (93.9; 90.2 - 97.6)	
Race or ethnicity				.10
Asian	149/730 (20.4; 17.5 - 23.3)	12/149 (8.1; 3.7 - 12.4)	137/149 (91.9; 87.6 - 96.3)	
Hispanic	108/730 (14.8; 12.2 - 17.4)	13/108 (12; 5.9 - 18.2)	95/108 (88; 81.8 - 94.1)	
Non-Hispanic Black	65/730 (8.9; 6.8 - 11.0)	4/65 (6.2; 0.3 - 12.0)	61/65 (93.8; 88.0 - 99.7)	
Non-Hispanic White	369/730 (50.5; 46.9 - 54.2)	18/369 (4.9; 2.7 - 7.1)	351/369 (95.1; 92.9 - 97.3)	
Other[c]	39/730 (5.3; 3.7 - 7.0)	4/39 (10.3; 0.7 - 19.8)	35/39 (89.7; 80.2 - 99.3)	
Location of practice				.17
Rural	30/729 (4.1; 2.7 - 5.6)	4/30 (13.3; 1.1 - 25.5)	26/30 (86.7; 74.5 - 98.9)	
Urban	699/729 (95.9; 94.4 - 97.3)	47/699 (6.7; 4.9 - 8.6)	652/699 (93.3; 91.4 - 95.1)	
Type of facility				*.003*
University hospital, teaching hospital, affiliated clinic, or employed physician practice	197/730 (27; 23.8 - 30.2)	10/197 (5.1; 2.0 - 8.1)	187/197 (94.9; 91.9-98.0)	
FQHC[d] or city, county, or public health care facility	99/730 (13.6; 11.1-16.1)	15/99 (15.2; 8.1-22.2)	84/99 (84.8; 77.8-91.9)	
Other[e]	434/730 (59.5; 55.9 - 63.0)	26/434 (6.0; 3.8 - 8.2)	408/434 (94; 91.8 - 96.2)	
Standing orders				*.01*
Yes	282/563 (50.1; 45.9-54.2)	23/282 (8.2; 5.0-11.4)	259/282 (91.8; 88.6 - 95.0)	
No	281/563 (49.9; 45.8 - 54.1)	9/281 (3.2; 1.1 - 5.3)	272/281 (96.8; 94.7 - 98.9)	
Reminders (eg, phone or mail)				.94
Yes	274/592 (46.3; 42.3-50.3)	18/274 (6.6; 3.6 - 9.5	256/274 (93.4; 90.5 - 96.4)	
No	257/592 (43.4; 39.4 - 47.4)	15/257 (5.8; 3.0 - 8.7)	242/257 (94.2; 91.3-97.0)	
I don't know	61/592 (10.3; 7.8 - 12.8)	4/61 (6.6; 0.3 - 12.8)	57/61 (93.4; 87.2 - 99.7)	
Recalls (eg, computerized tracking)				.97
Yes	244/595 (41; 37.0 - 45.0)	15/244 (6.1; 3.1 - 9.2)	229/244 (93.9; 90.8 - 96.9)	
No	263/595 (44.2; 40.2 - 48.2)	16/263 (6.1; 3.2 - 9.0)	247/263 (93.9; 91.0 - 96.8)	
I don't know	88/595 (14.8; 11.9 - 17.7)	6/88 (6.8; 1.5 - 12.1)	82/88 (93.2; 87.9 - 98.5)	
Predominant non-Hispanic White population zip code				.27
Yes	424/636 (66.7; 63.0 - 70.3)	26/424 (6.1; 3.8 - 8.4)	398/424 (93.9; 91.6 - 96.2)	

XSL•FO
RenderX

Characteristic	Total, n/N (%; 95% CI)	Decreased, n/N (%; 95% CI)	Increased or no change, n/N (%; 95% CI)	*P* value[b]
No	212/636 (33.3; 29.7 - 37.0)	18/212 (8.5; 4.7 - 12.3)	194/212 (91.5; 87.7 - 95.3)	
Predominant non-Hispanic Black population zip code				.74
Yes	36/636 (5.7; 3.9 - 7.5)	2/36 (5.6; 0.0 - 13.1)	34/36 (94.4; 86.9-100.0)	
No	600/636 (94.3; 92.5 - 96.1)	42/600 (7; 5.0 - 9.0)	558/600 (93.0; 91.0 - 95.0)	
Predominant Hispanic population zip code				.14
Yes	171/636 (26.9; 23.4 - 30.3)	16/171 (9.4; 5.0 - 13.7)	155/171 (90.6; 86.3 - 95.0)	
No	465/636 (73.1; 69.7 - 76.6)	28/465 (6; 3.9 - 8.2)	437/465 (94; 91.8 - 96.1)	

[a]Missing observations: 152 for promoted vaccination-related applications, 25 for HCP's age, 11 for provider's sex, 1 for location of practice, 167 for standing orders, 138 for reminders, 135 for recalls, 94 each for predominant non-Hispanic White, non-Hispanic Black and Hispanic populations ZIP Code.

[b]*P* values obtained from the chi-square test; *P* values significant at .05 are in italic formatting.

[c]"Other" for race and ethnicity includes American Indian and Alaska Native and others.

[d]FQHC: Federally Qualified Health Center.

[e]"Other" for type of facility includes group practice, solo practice, faith-based hospital or clinic, and others.

Relationship Between Promotion of the Use of Vaccination-Related Applications at HCPs' Facilities and Observing a Decrease in HPV Vaccine Hesitancy Among Patients

Compared to HCPs who did not report promotion of vaccination-related applications at their facilities, those who did had significantly higher odds of observing a decrease in HPV vaccine hesitancy among patients during the COVID-19 pandemic (adjusted odds ratio [aOR] 2.48, 95% CI 1.10-5.55; *P*=.03; Table 2). Additionally, compared to HCPs working at university hospitals, teaching hospitals, affiliated clinics, or employed physician practices, HCPs working at FQHCs or city, county, or public health care facilities had significantly higher odds of observing a decrease in HPV vaccine hesitancy among patients during the COVID-19 pandemic (aOR 4.02, 95% CI 1.33-12.14; *P*=.01; Table 2). Compared to HCPs who did not administer the HPV vaccine under standing orders at their facilities, those who did had significantly higher odds of observing a decrease in HPV vaccine hesitancy among patients (aOR 2.91, 95% CI 1.11-7.63; *P*=.03; Table 2).

We did not find a statistically significant association between observed decrease in HPV vaccine hesitancy during the COVID-19 pandemic reported by HCPs and HCPs' demographic characteristics, such as age, sex, race and ethnicity, location of practice, and use of reminders or recalls at their facilities, as well as patient-related characteristics, such as zip codes with predominantly non-Hispanic White, non-Hispanic Black, and Hispanic populations in which the HCPs practiced (Table 2).

XSL•FO
RenderX

Table . Logistic regression analyses of the association between the observed changes in human papillomavirus vaccine hesitancy among patients and promotion of vaccination-related applications during the COVID-19 pandemic.

Characteristic	Crude odds ratio (95% CI)	*P* value	Adjusted odds ratio (95% CI)	*P* value
Promoted vaccination-related applications				
No	Reference	Reference	Reference	Reference
Yes	2.59 (1.25 - 5.40)	*.01*	2.48 (1.10 - 5.55)	*.03*
I don’t know	0.79 (0.23 - 2.75)	.71	0.64 (0.17 - 2.48)	.52
Age (y)				
<35	Reference	Reference	Reference	Reference
35-54	0.65 (0.23 - 1.81)	.41	0.49 (0.16 - 1.51)	.21
≥55	1.07 (0.36 - 3.15)	.90	0.95 (0.30 - 3.08)	.94
Sex				
Female	Reference	Reference	Reference	Reference
Male	0.62 (0.25 - 1.51)	.29	0.45 (0.15 - 1.38)	.16
Race or ethnicity				
Asian	1.25 (0.50 - 3.18)	.63	1.40 (0.51 - 3.80)	.51
Hispanic	2.63 (1.16 - 5.97)	*.02*	2.12 (0.77 - 5.84)	.15
Non-Hispanic Black	1.18 (0.33 - 4.28)	.80	0.82 (0.13 - 5.31)	.83
Non-Hispanic White	Reference	Reference	Reference	Reference
Other[a]	1.38 (0.30 - 6.41)	.68	3.11 (0.47 - 20.42)	.24
Location of practice				
Rural	Reference	Reference	Reference	Reference
Urban	0.53 (0.15 - 1.85)	.32	0.57 (0.13 - 2.43)	.45
Type of facility				
University hospital, teaching hospital, affiliated clinic, or employed physician practice	Reference	Reference	Reference	Reference
FQHC[b] or city, county, or public health care facility	4.45 (1.74 - 11.42)	*.002*	4.02 (1.33 - 12.14)	*.01*
Other[c]	1.15 (0.46 - 2.86)	.77	0.90 (0.29 - 2.81)	.85
Standing orders				
No	Reference	Reference	Reference	Reference
Yes	2.71 (1.23 - 5.98)	*.01*	2.91 (1.11 - 7.63)	*.03*
Reminders (eg, phone or mail)				
No	Reference	Reference	Reference	Reference
Yes	1.15 (0.57 - 2.35)	.70	0.82 (0.29 - 2.33)	.71
I don’t know	1.14 (0.36 - 3.6)	.82	0.68 (0.13 - 3.60)	.65
Recalls (eg, computerized tracking)				
No	Reference	Reference	Reference	Reference
Yes	1.04 (0.5 - 2.16)	.92	0.73 (0.27 - 2.00)	.54
I don’t know	1.13 (0.43 - 3.0)	.81	1.03 (0.26 - 4.07)	.97
Predominant non-Hispanic White population zip code				
No	Reference	Reference	Reference	Reference

XSL•FO
RenderX

Characteristic	Crude odds ratio (95% CI)	*P* value	Adjusted odds ratio (95% CI)	*P* value
Yes	0.55 (0.27 - 1.13)	.10	0.81 (0.27 - 2.42)	.71
Predominant non-Hispanic Black population zip code				
No	Reference	Reference	Reference	Reference
Yes	1.04 (0.24 - 4.62)	.96	0.88 (0.26 - 3.03)	.84
Predominant Hispanic population zip code				
No	Reference	Reference	Reference	Reference
Yes	1.94 (0.93 - 4.05)	.08	0.81 (0.29 - 2.24)	.68

[a]"Other" for race and ethnicity included American Indian and Alaska Native and others.

[b]FQHC: Federally Qualified Health Center.

[c]"Other" for type of facility included group practice, solo practice, faith-based hospital or clinic, and others.

Discussion

Principal Findings

In this population-based study, we found that HCPs working in Texas who promoted vaccination-related applications at their facilities had higher odds of observing a decrease in HPV vaccine hesitancy among patients during the COVID-19 pandemic. Our study findings align with those of previous studies reporting the benefits of vaccination-related mobile apps in decreasing HPV vaccine hesitancy and thereby increasing vaccine uptake [24-29]. Recent studies have reported a rise in HPV vaccine hesitancy and decreased vaccine uptake following the COVID-19 pandemic in the United States [33-35,42]. As the HPV vaccine is safe and effective in preventing HPV-associated cancers [2,5], and considering that digital interventions have been proven effective at increasing HPV vaccine awareness and knowledge [24,25], there is a need to promote vaccination-related applications to increase awareness and knowledge and address misinformation about the HPV vaccine.

Previous studies have reported that the most common reasons for HPV vaccine hesitancy are safety concerns and lack of knowledge and absence of HCP recommendation among parents of vaccine-eligible children and among age-eligible adults, respectively [43,44]. A positive association was found between HCPs' knowledge of the HPV vaccine and initiation and completion of the vaccine series [45]. When HCPs are well informed, they can confidently communicate vaccine benefits and respond to parental concerns, which is critical for building trust and influencing vaccination decisions [46]. Mobile apps have helped users access immunization-related information, create and track vaccination schedules, and locate clinics where the public can receive vaccines [47,48]. Prior studies have focused on using mobile apps to promote HPV vaccination–related information, some of which are publicly available whereas others are still being evaluated as part of research projects. Among the mobile apps under evaluation in research studies, Vaccipack, Vacteens, Vax4HPV, HPVCancerFree, and TeenVac have demonstrated their potential to decrease HPV vaccine hesitancy [26-30]. Thus, integrating vaccination-related mobile apps into clinical practice could be a valuable tool for promoting HPV vaccination and reducing hesitancy. Furthermore, timely reminders sent by mobile apps to vaccine-eligible individuals about their vaccination dates align with the transtheoretical model of behavior change, which posits that behavior change is a gradual process that occurs in stages [49]. By providing reminders to vaccine-eligible individuals, these vaccination-related apps motivate them to move from contemplation to action, thereby improving HPV vaccination uptake [49]. The US CDC developed a publicly available app called CDC Vaccine Schedules to provide HCPs and patients with immediate access to comprehensive information on vaccine recommendations, including immunization schedules, precautions, and contraindications [50]. Other publicly available immunization-related apps include ACOG, developed by the American College of Obstetricians and Gynecologists; HPV Vaccine: Same Way, Same Day, developed by the Academic Pediatric Association and American Academy of Pediatrics; and Vaccines on the Go: What You Should Know [51]. Despite the variety of official vaccination-related applications available to HCPs, we found that a small proportion (104/578, 18%) of Texas-based HCPs promoted HPV vaccination–related applications at their health care facilities.

We also found that HCPs working at FQHCs or city, county, or public health care facilities had higher odds of observing a decrease in HPV vaccine hesitancy among patients than those working at university hospitals, teaching hospitals, affiliated clinics, or employed physician practices. This could be attributed to the greater availability of preventive health care services, such as HPV vaccination, at public facilities, which provides an opportunity for HCPs working at such practices to further counsel patients and recommend vaccination applications [52,53]. A study conducted among HCPs serving minority and low-income populations identified a lack of preventive care visits as a barrier to HPV vaccine recommendation [54]. Given that FQHCs primarily offer health care services to uninsured and low-income patients, they play a critical role in increasing the HPV vaccination uptake among socioeconomically disadvantaged populations [55]. Adolescents from more educated and higher-income households show higher rates of HPV vaccine hesitancy [56]. On the other hand, adolescents with lower income levels have higher odds of completing the HPV vaccine series [57]. These findings highlight the need for organizational policies or protocols to strengthen access to HPV

XSL•FO
RenderX

vaccination, improve patient counseling, and reduce vaccine hesitancy.

We found that HCPs who reported using standing orders for HPV vaccination at their facilities had higher odds of observing a decrease in HPV vaccine hesitancy among their patients. Standing orders authorize nurses, pharmacists, and other trained HCPs, as permitted by state law, to assess a patient's immunization status and administer vaccines in accordance with a preapproved protocol established by the medical director, physician, or other authorized practitioner [58]. A 2022 study conducted among 16 nurses, medical assistants, and HCPs working in primary care settings in the United States found that implementing standing orders for HPV vaccination increased nurses' and medical assistants' autonomy [59]. This approach led to improved access to the HPV vaccine, fewer missed opportunities, earlier vaccination, and normalization of HPV vaccination as a routine part of care [59]. As standing orders have been shown to be effective at increasing vaccination rates [60], they may be particularly effective in addressing HPV vaccine hesitancy by ensuring timely, consistent, and standardized HPV vaccination practices across health care facilities.

Strengths and Limitations

The major strength of our study is the inclusion of HCPs from diverse backgrounds and specialties. While it offers critical insights into the benefits of digital interventions during the COVID-19 pandemic, our study is not without limitations. First, we could not infer causality because the study was cross-sectional. Second, a small proportion of HCPs practicing at the health care facilities who responded to the survey observed decreases in HPV vaccine hesitancy, which may have reduced the power of our statistical analysis. This small sample size also limits the generalizability of our findings and increases the risk of falsely rejecting our study hypothesis. Third, Texas has a unique sociopolitical landscape regarding vaccine attitudes, as well as a distinct population racial, ethnic, and age distribution, limiting the generalizability of our findings to other US regions. Fourth, there is potential for information bias as changes in HPV vaccine hesitancy were self-reported and based on HCP recall. Fifth, the study question does not allow us to specify the types of applications being promoted or the mode of promotion: active (eg, discussed with patients), passive (eg, posters), or both. Sixth, we did not include HCPs comfort with digital tools, patients' socioeconomic status, and institutional culture as covariates in our study as such information was not available. Finally, the sampling frame derived from the LexisNexis database may not fully cover all Texas-based HCPs as some HCPs may have been missing from the database or lacked an email address, potentially limiting the generalizability of the findings.

Conclusions

This study demonstrated the added value of mobile apps in reducing HPV vaccine hesitancy in Texas. Our findings have implications for designing and implementing tailored digital interventions to successfully counter misinformation about HPV vaccination, especially during major outbreaks such as the COVID-19 pandemic. Future prospective studies could explore the long-term impact of these vaccination-related mobile apps in reducing missed opportunities and increasing HPV vaccination uptake among age-eligible individuals.

Acknowledgments

The authors want to thank the Texas health care professionals for participating in the study. Generative artificial intelligence was not used in any portion of manuscript writing.

Funding

The study was funded by the National Cancer Institute (P30CA016672 to S Shete), the Betty B. Marcus Chair in Cancer Prevention (awarded to S Shete), and the Duncan Family Institute for Cancer Prevention and Risk Assessment (awarded to S Shete). The funders were not involved in the study design, data analysis and interpretation, or manuscript writing.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: MD, JF-D, S Shete
Formal analysis: MD, RY
Funding acquisition: S Shete
Interpreted study data: MD, S Sharma, JF-D, OC, RY, WC, CD, S Shete
Methodology: MD, JF-D, S Shete
Supervision: S Shete
Writing—original draft: MD
Writing—review and editing: S Sharma, JF-D, OC, WC, CD, S Shete

Conflicts of Interest

None declared.

XSL•FO
RenderX

Checklist 1
STROBE checklist.
[DOCX File, 33 KB - mhealth_v14i1e77778_app1.docx]

References

1. Cancers linked with HPV each year. Centers for Disease Control and Prevention. 2025. URL: https://www.cdc.gov/cancer/hpv/cases.html [accessed 2026-4-15]
2. Joura EA, Giuliano AR, Iversen OE, et al. A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women. N Engl J Med 2015 Feb 19;372(8):711-723. [doi: 10.1056/NEJMoa1405044] [Medline: 25693011]
3. Human papillomavirus (HPV) vaccination: information for health care providers. Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/vaccines/vpd/hpv/hcp/recommendations.html [accessed 2025-09-10]
4. Restrepo J, Herrera T, Samakoses R, et al. Ten-year follow-up of 9-valent human papillomavirus vaccine: immunogenicity, effectiveness, and safety. Pediatrics 2023;152(4):e2022060993. [doi: 10.1542/peds.2022-060993]
5. Human papillomavirus (HPV) vaccine safety. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/vaccine-safety/vaccines/hpv.html#:~:text=HPV%20vaccine%20is%20safe%20and,sore%20arm%20from%20the%20shot [accessed 2026-04-15]
6. Zheng L, Wu J, Zheng M. Barriers to and facilitators of human papillomavirus vaccination among people aged 9 to 26 years: a systematic review. Sex Transm Dis 2021 Dec 1;48(12):e255-e262. [doi: 10.1097/OLQ.0000000000001407] [Medline: 33783412]
7. Pingali C, Yankey D, Elam-Evans LD, et al. Vaccination coverage among adolescents aged 13-17 years - National Immunization Survey-Teen, United States, 2024. MMWR Morb Mortal Wkly Rep 2025 Aug 14;74(30):466-472. [doi: 10.15585/mmwr.mm7430a1] [Medline: 40811113]
8. Smulian EA, Mitchell KR, Stokley S. Interventions to increase HPV vaccination coverage: a systematic review. Hum Vaccin Immunother 2016 Jun 2;12(6):1566-1588. [doi: 10.1080/21645515.2015.1125055] [Medline: 26838959]
9. About the Vaccines for Children (VFC) program. Centers for Disease Control and Prevention. 2025. URL: https://www.cdc.gov/vaccines-for-children/about/index.html [accessed 2026-04-15]
10. HPV vaccination. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hpv/vaccines/index.html [accessed 2026-4-27]
11. Markowitz LE, Dunne EF, Saraiya M, et al. Quadrivalent human papillomavirus vaccine: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep 2007 Mar 23;56(RR-2):1-24. [Medline: 17380109]
12. How to pay for vaccines. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/vaccines-adults/recommended-vaccines/how-to-pay-adult-vaccines.html [accessed 2025-09-10]
13. Pingali C, Yankey D, Chen M, et al. National vaccination coverage among adolescents aged 13-17 years - National Immunization Survey-Teen, United States, 2023. MMWR Morb Mortal Wkly Rep 2024 Aug 22;73(33):708-714. [doi: 10.15585/mmwr.mm7333a1] [Medline: 39173168]
14. Increase the proportion of adolescents who get recommended doses of the HPV vaccine — IID-08. Office of Disease Prevention and Health Promotion. URL: https://odphp.health.gov/healthypeople/objectives-and-data/browse-objectives/vaccination/increase-proportion-adolescents-who-get-recommended-doses-hpv-vaccine-iid-08 [accessed 2025-09-10]
15. Ten threats to global health in 2019. World Health Organization. URL: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 [accessed 2025-09-10]
16. Vaccine hesitancy: a growing challenge for immunization programmes. World Health Organization. 2015. URL: https://www.who.int/news/item/18-08-2015-vaccine-hesitancy-a-growing-challenge-for-immunization-programmes [accessed 2025-09-10]
17. Kyei GK, Kyei EF, Ansong R. HPV vaccine hesitancy and uptake: a conceptual analysis using Rodgers' evolutionary approach. J Adv Nurs 2025 May;81(5):2368-2381. [doi: 10.1111/jan.16653] [Medline: 39651685]
18. Beavis AL, Meek K, Moran MB, Fleszar L, Adler S, Rositch AF. Exploring HPV vaccine hesitant parents' perspectives on decision-making and motivators for vaccination. Vaccine X 2022;12:100231. [doi: 10.1016/j.jvacx.2022.100231] [Medline: 36337836]
19. Krakow M, Beavis A, Cosides O, Rositch AF. Characteristics of adolescents lacking provider-recommended human papillomavirus vaccination. J Adolesc Health 2017 May;60(5):619-622. [doi: 10.1016/j.jadohealth.2016.11.028] [Medline: 28073618]
20. Bradley-Ewing A, Lee BR, Doctor JN, Meredith G, Goggin K, Myers A. A pilot intervention combining assessment and feedback with communication training and behavioral nudges to increase HPV vaccine uptake. Hum Vaccin Immunother 2022 Dec 31;18(1):1885968. [doi: 10.1080/21645515.2021.1885968] [Medline: 34085873]
21. Sonawane K, Zhu Y, Montealegre JR, et al. Parental intent to initiate and complete the human papillomavirus vaccine series in the USA: a nationwide, cross-sectional survey. Lancet Public Health 2020 Sep;5(9):e484-e492. [doi: 10.1016/S2468-2667(20)30139-0] [Medline: 32707126]
22. Szilagyi PG, Albertin CS, Gurfinkel D, et al. Prevalence and characteristics of HPV vaccine hesitancy among parents of adolescents across the US. Vaccine 2020 Aug;38(38):6027-6037. [doi: 10.1016/j.vaccine.2020.06.074] [Medline: 32758380]

XSL•FO
RenderX

23. Jacobson RM, St Sauver JL, Griffin JM, MacLaughlin KL, Finney Rutten LJ. How health care providers should address vaccine hesitancy in the clinical setting: evidence for presumptive language in making a strong recommendation. Hum Vaccin Immunother 2020 Sep 1;16(9):2131-2135. [doi: 10.1080/21645515.2020.1735226] [Medline: 32242766]
24. Choi J, Tamí-Maury I, Cuccaro P, Kim S, Markham C. Digital health interventions to improve adolescent HPV vaccination: a systematic review. Vaccines (Basel) 2023 Jan 22;11(2):249. [doi: 10.3390/vaccines11020249] [Medline: 36851127]
25. Ou L, Chen AC, Amresh A. The effectiveness of mHealth interventions targeting parents and youth in human papillomavirus vaccination: systematic review. JMIR Pediatr Parent 2023 Nov 21;6:e47334. [doi: 10.2196/47334] [Medline: 37988155]
26. Teitelman AM, Gregory EF, Jayasinghe J, et al. Vaccipack, a mobile app to promote human papillomavirus vaccine uptake among adolescents aged 11 to 14 years: development and usability study. JMIR Nurs 2020;3(1):e19503. [doi: 10.2196/19503] [Medline: 34345789]
27. Kim K, Parekh De Campos A, Choi S. A Vax4HPV mobile application for parents of human papillomavirus vaccine-eligible children: iterative formative assessments. Comput Inform Nurs 2022 Jul 1;40(7):455-465. [doi: 10.1097/CIN.0000000000000832] [Medline: 35797614]
28. Woodall WG, Zimet G, Kong A, et al. Vacteens.org: a mobile web app to improve HPV vaccine uptake. Front Digit Health 2021;3:693688. [doi: 10.3389/fdgth.2021.693688] [Medline: 34713171]
29. Shegog R, Savas LS, Healy CM, et al. AVPCancerFree: impact of a digital behavior change intervention on parental HPV vaccine-related perceptions and behaviors. Hum Vaccin Immunother 2022 Nov 30;18(5):2087430. [doi: 10.1080/21645515.2022.2087430] [Medline: 35699953]
30. Woodall WG, Buller D, Zimet G, et al. TeenVac.org: a mobile web app to improve human papillomavirus vaccine uptake for adolescent boys. J Adolesc Health 2025 Oct;77(4):749-755. [doi: 10.1016/j.jadohealth.2025.06.015] [Medline: 40824226]
31. National Immunization Survey-Teen (NIS-Teen) 2024, Texas. Texas Department of State Health Services. URL: https://www.dshs.texas.gov/immunizations/data/surveys/nis/teens [accessed 2026-04-16]
32. Irimata KE, Pleis JR, Heslin KC, He Y. Reduced access to preventive care due to the COVID-19 pandemic, by chronic disease status and race and Hispanic origin, United States, 2020-2021. Public Health Rep 2023;138(2):341-348. [doi: 10.1177/00333549221138855] [Medline: 36524404]
33. He K, Mack WJ, Neely M, Lewis L, Anand V. Parental perspectives on immunizations: impact of the COVID-19 pandemic on childhood vaccine hesitancy. J Community Health 2022 Feb;47(1):39-52. [doi: 10.1007/s10900-021-01017-9] [Medline: 34297272]
34. Ryan GW, Miotto MB, McReynolds C, Lemon SC, Pbert L, Trivedi M. Pediatricians' perspectives on COVID-19 and HPV vaccine hesitancy. Hum Vaccin Immunother 2023 Aug 1;19(2):2225388. [doi: 10.1080/21645515.2023.2225388] [Medline: 37347712]
35. Chido-Amajuoyi OG, Pande M, Agbajogu C, Yu RK, Cunningham S, Shete S. HPV vaccination uptake, hesitancy, and refusal: observations of health-care professionals during the COVID-19 pandemic. JNCI Cancer Spectr 2022 Jul 1;6(4):pkac053. [doi: 10.1093/jncics/pkac053] [Medline: 35900184]
36. Patel Murthy B, Zell E, Kirtland K, et al. Impact of the COVID-19 pandemic on administration of selected routine childhood and adolescent vaccinations - 10 U.S. jurisdictions, March-September 2020. MMWR Morb Mortal Wkly Rep 2021 Jun 11;70(23):840-845. [doi: 10.15585/mmwr.mm7023a2] [Medline: 34111058]
37. Transform your risk decision making. LexisNexis Risk Solutions. URL: https://risk.lexisnexis.com/ [accessed 2025-09-10]
38. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007 Oct;370(9596):1453-1457. [doi: 10.1016/S0140-6736(07)61602-X] [Medline: 18064739]
39. Rural-Urban Continuum Codes. U.S. Department of Agriculture. URL: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes [accessed 2025-09-10]
40. Income in the past 12 months (in 2022 inflation-adjusted dollars). United States Census Bureau. URL: https://data.census.gov/table/ACSST1Y2022.S1901?t=Income+and+Poverty&g=010XX00US_040XX00US48,48$8600000&y=2022 [accessed 2025-12-01]
41. Walker RJ, Dawson AZ, Thorgerson A, et al. Relationship between satisfaction scores and racial/ethnic and sex concordance in primary care. Healthcare (Basel) 2023 Aug 12;11(16):2276. [doi: 10.3390/healthcare11162276] [Medline: 37628474]
42. Nguyen P, Calderon-Mora J, Singh V, Hernandez A, Roy S, Molokwu J. Impact of the COVID-19 pandemic on HPV vaccine uptake in a predominantly Hispanic border community: a retrospective cross-sectional analysis of the "Tiempo de Vacunarte Program". Arch Public Health 2024 Jun 24;82(1):96. [doi: 10.1186/s13690-024-01318-0] [Medline: 38915042]
43. Chido-Amajuoyi OG, Talluri R, Shete SS, Shete S. Safety concerns or adverse effects as the main reason for human papillomavirus vaccine refusal. JAMA Pediatr 2021 Oct 1;175(10):1074-1076. [doi: 10.1001/jamapediatrics.2021.1585] [Medline: 34180965]
44. Fokom Domgue J, Cunningham SA, Yu RK, Shete S. Reasons for not receiving the HPV vaccine among eligible adults: lack of knowledge and of provider recommendations contribute more than safety and insurance concerns. Cancer Med 2020 Jul;9(14):5281-5290. [doi: 10.1002/cam4.3192] [Medline: 32483891]

XSL•FO
RenderX

45. Rutten LJ, St. Sauver JL, Beebe TJ, et al. Clinician knowledge, clinician barriers, and perceived parental barriers regarding human papillomavirus vaccination: association with initiation and completion rates. Vaccine 2017 Jan;35(1):164-169. [doi: 10.1016/j.vaccine.2016.11.012] [Medline: 27887795]
46. Limaye RJ, Opel DJ, Dempsey A, et al. Communicating with vaccine-hesitant parents: a narrative review. Acad Pediatr 2021;21(4S):S24-S29. [doi: 10.1016/j.acap.2021.01.018] [Medline: 33958087]
47. Wilson K, Atkinson KM, Westeinde J. Apps for immunization: leveraging mobile devices to place the individual at the center of care. Hum Vaccin Immunother 2015;11(10):2395-2399. [doi: 10.1080/21645515.2015.1057362] [Medline: 26110351]
48. de Cock C, van Velthoven M, Milne-Ives M, Mooney M, Meinert E. Use of apps to promote childhood vaccination: systematic review. JMIR Mhealth Uhealth 2020 May 18;8(5):e17371. [doi: 10.2196/17371] [Medline: 32421684]
49. Hacker K, Brown E, Cabral H, Dodds D. Applying a transtheoretical behavioral change model to HIV/STD and pregnancy prevention in adolescent clinics. J Adolesc Health 2005 Sep;37(3 Suppl):S80-S93. [doi: 10.1016/j.jadohealth.2005.05.007] [Medline: 16115574]
50. CDC vaccine schedules app for healthcare providers. Centers for Disease Control and Prevention Vaccines & Immunizations. 2025. URL: https://www.cdc.gov/vaccines/hcp/imz-schedules/app.html [accessed 2025-09-10]
51. Vaccine apps. Immunize.org. URL: https://www.immunize.org/clinical/external/apps/ [accessed 2025-09-10]
52. Vaccines. Texas Department of State Health Services. URL: https://www.dshs.texas.gov/immunizations/public/vaccines [accessed 2025-09-10]
53. Community Health & Wellness Division. Harris County Public Health. URL: https://publichealth.harriscountytx.gov/Divisions-Offices/Divisions/Community-Health-Wellness-Division [accessed 2025-09-10]
54. Bruno DM, Wilson TE, Gany F, Aragones A. Identifying human papillomavirus vaccination practices among primary care providers of minority, low-income and immigrant patient populations. Vaccine 2014 Jul;32(33):4149-4154. [doi: 10.1016/j.vaccine.2014.05.058] [Medline: 24886959]
55. TACHC annual report 2024. : Texas Association of Community Health Centers URL: https://tachc.org/wp-content/uploads/2025/05/TACHC_annual-report_2024-FINAL.pdf.pdf [accessed 2026-04-16]
56. Sonawane K, Zhu Y, Damgacioglu H, et al. Factors associated with parental human papillomavirus vaccination intentions among adolescents from socioeconomically advantaged versus deprived households: a nationwide, cross-sectional survey. Lancet Reg Health Am 2024;31:100694. [doi: 10.1016/j.lana.2024.100694] [Medline: 38500960]
57. Ejezie CL, Osaghae I, Ayieko S, Cuccaro P. Adherence to the recommended HPV vaccine dosing schedule among adolescents aged 13 to 17 years: findings from the National Immunization Survey-Teen, 2019-2020. Vaccines (Basel) 2022 Apr 8;10(4):577. [doi: 10.3390/vaccines10040577] [Medline: 35455325]
58. Using standing orders for administering vaccines: what you should know. Immunize.org. URL: https://www.immunize.org/wp-content/uploads/catg.d/p3066.pdf [accessed 2025-09-10]
59. Mansfield LN, Kahn BZ, Kokitkar S, Kritikos KI, Brantz SN, Brewer NT. HPV vaccine standing orders and communication in primary care: a qualitative study. Vaccine 2024 Jul;42(19):3981-3988. [doi: 10.1016/j.vaccine.2024.05.008] [Medline: 38816304]
60. Tan LJ, VanOss R, Ofstead CL, Wetzler HP. Maximizing the impact of, and sustaining standing orders protocols for adult immunization in outpatient clinics. Am J Infect Control 2020 Mar;48(3):290-296. [doi: 10.1016/j.ajic.2019.07.023] [Medline: 31630922]

Abbreviations

aOR: adjusted odds ratio
CDC: Centers for Disease Control and Prevention
FQHC: Federally Qualified Health Center
HCP: health care professional
HPV: human papillomavirus
REDCap: Research Electronic Data Capture
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

XSL•FO
RenderX

Edited by L Buis; submitted 19.May.2025; peer-reviewed by A Ijeoma, O Taiwo, Y Deng; revised version received 26.Mar.2026; accepted 26.Mar.2026; published 14.May.2026.

Please cite as:
Desai M, Sharma S, Fokom-Domgue J, Yu R, Chan W, Chido-Amajuoyi OG, Darkoh C, Shete S
Vaccination-Related Applications and Health Care Professionals' Observed Changes in Human Papillomavirus Vaccine Hesitancy: Cross-Sectional Survey
JMIR Mhealth Uhealth 2026;14:e77778
URL: https://mhealth.jmir.org/2026/1/e77778
doi:10.2196/77778

© Manali Desai, Shreela Sharma, Joel Fokom-Domgue, Robert Yu, Wenyaw Chan, Onyema G Chido-Amajuoyi, Charles Darkoh, Sanjay Shete. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Patient and Clinician Attitudes Toward Mobile Health Apps: Qualitative Study

Cornelius A James[1,2,3], MD; Sarah L Krein[1,4], PhD, RN; Sarah Yon[4,5], MPH; Michael P Dorsch[6], MS, PharmD; John D Piette[1,4,5], MSc, PhD; Brahmajee K Nallamothu[1], MPH, MD

[1]Department of Internal Medicine, University of Michigan, 2800 Plymouth Road, Bldg. 15, Rm 430W, Ann Arbor, MI, United States

[2]Department of Pediatrics, University of Michigan, Ann Arbor, MI, United States

[3]Department of Learning Health Sciences, University of Michigan, Ann Arbor, MI, United States

[4]Department of Veterans Affairs Center for Clinical Management Research, Ann Arbor, MI, United States

[5]Department of Behavioral and Health Education, University of Michigan, Ann Arbor, MI, United States

[6]Department of Clinical Pharmacy, University of Michigan, Ann Arbor, United States

Corresponding Author:
Cornelius A James, MD
Department of Internal Medicine, University of Michigan, 2800 Plymouth Road, Bldg. 15, Rm 430W, Ann Arbor, MI, United States

Abstract

Background: Mobile health (mHealth) apps are widely available, and some have proven safe and effective for management of specific chronic conditions. Despite a high degree of interest, the potential of these technologies has yet to be realized. Patient and clinician attitudes are key factors that influence the adoption of mHealth apps but remain poorly understood, particularly in the United States.

Objective: This study aimed to identify both patient and clinician attitudes that can influence recommending and adopting mHealth apps.

Methods: Using well-established technology adoption and implementation science frameworks, this study included a deductive content analysis using a rapid qualitative analytic method. Semistructured interviews were conducted with patients and clinicians to identify technical and material, social and personal, and policy and organizational factors that can influence the recommendation or adoption of mHealth apps. The interviews and data analysis were performed between September 2023 and August 2024.

Results: Participants included 20 clinicians (n=12, 60% general internists) with a mean time in practice of 17 (SD 11.6) years, and 28 patients with a mean age of 59 (SD 12.1) years. A total of 7 categories related to patients' and clinicians' attitudes toward mHealth apps emerged: (1) apps as tools to improve health by extending care, (2) the role of apps in enhancing the patient–clinician relationship, (3) the need for simplicity and efficiency in app design, (4) the influence of prior experience with mHealth apps, (5) comfort with technology, (6) recommendations from trusted sources, and (7) education and hands-on experience. Although similar factors were considered by patients and clinicians, their views about older adults' interest and ability to use mHealth apps differed.

Conclusions: Understanding patient and clinician views about mHealth apps provides critical insights for developing approaches to facilitate their use. These findings suggest patients and clinicians share similar views about the benefits of mHealth apps. Nonetheless, clinicians' perceptions about older patients' interest and ability to use mHealth apps may negatively impact recommendation of mHealth apps and subsequent adoption by older adults.

(JMIR Mhealth Uhealth 2026;14:e77519) doi:10.2196/77519

KEYWORDS

mobile health; mHealth; digital health technology; mhealth apps; health apps; qualitative

Introduction

Mobile health (mHealth) apps can collect health information to facilitate monitoring of patients' health and behaviors throughout the day and provide real-time, personalized recommendations at key decision points (eg, when making dietary decisions). mHealth apps can increase physical activity [1] and address conditions such as depression, anxiety [2], tobacco use [3], and alcohol use disorder [4].

There are over 350,000 mHealth apps available in app stores [5], and nearly two-thirds of US adults report use of an app for health-related purposes [6]. The global mHealth market was valued at US $83.17 billion in 2023 and is expected to grow to US $249.05 billion in 2032 [7]. Despite enthusiasm for app use,

XSL•FO
RenderX

adoption has lagged in underserved populations and among persons with chronic conditions [8,9]. With an increasing life expectancy [10], increased burden of chronic diseases in the United States [11], and persisting barriers to health care access [12], mHealth apps have the potential to improve health outcomes by offering the convenience of readily available health care solutions.

Getting safe and effective apps into the hands of people who need them most, and sustaining use of these apps remains a challenge. For instance, older aged persons, women, and those with lower educational attainment are less likely to use apps, and the reasons for these disparities are unknown [13]. Additionally, it has been estimated that approximately 71% of app users discontinue use within about 90 days [14]. Further, many clinicians either recommend mHealth apps inconsistently or they do not recommend them at all [15-17].

There are technology level, individual level (eg, patient and clinician), organization level, and policy level factors that may impact the adoption of mHealth apps, with each of these factors influencing the other(s) to varying degrees [18-21]. Studies have shown that patient and clinician attitudes about mHealth apps are amongst the most influential factors that determine mHealth app use [18,21,22]. In a study by Vo et al [23], patients identified patient engagement and patient empowerment as potential strengths of mHealth apps, and trustworthiness, limited personalization (eg, relevance to their condition, culture, and accessibility) as potential weaknesses. Sarradon-Eck and colleagues [24] showed that general practitioners in France had mixed views about mHealth apps, with some adopting a favorable view and others expressing skepticism or resistance. As an example, the main perceived benefit expressed by clinicians was related to the apps' potential to facilitate completion of their work and follow-up care because of the reliable and objective measurements provided by apps. Clinicians have also expressed concerns about mHealth apps negatively impacting the patient–clinician relationship because they could lead to decreased face-to-face interactions [25]. Additionally, in a study by Dahlhausen and colleagues [26], clinicians identified their ability to effectively advise patients on their use of mHealth apps, insufficient evidence supporting use and liability as concerns.

A small number of studies have explored the views of both patients and clinicians simultaneously. For example, Byambasuren and colleagues [27] conducted semistructured interviews of general practitioners and patients to identify determinants of prescribing mHealth apps in Australian general practice settings. General practitioners identified generational differences, limited knowledge of apps that can be prescribed and trustworthy sources to access them, time required to learn to use apps, concerns about privacy, and safety and trustworthiness as potential barriers. The main barriers identified by patients were older age and usability of mHealth apps. Importantly, patients did not share clinicians' concern about data privacy and data safety, highlighting an opportunity to develop a shared understanding of these potential barriers. Jezrawi et al [28] conducted surveys of both patients and physicians, as well as focus groups with patients to understand their attitudes about mHealth apps. Physicians who prescribed apps to patients relied most on personal opinions and recommendations from patients, whereas recommendations from trusted medical organizations or doctors were the most frequently cited sources that patients considered when making decisions about an mHealth app.

Even fewer studies have directly compared the views of patients and clinicians. For instance, to understand patients' and clinicians' perceptions about mHealth apps, Liu and colleagues [29] conducted semistructured interviews with 9 patients and 14 clinicians and triangulated their data to generate 5 themes, including 1 theme indicating different concerns between the 2 groups related to mHealth apps. For example, patients were more concerned about the effectiveness of mHealth apps, and clinicians were most concerned about the usability of mHealth apps. In another study, Barber and colleagues [30] conducted semistructured interviews to elicit patients' and physicians' perspectives on the use of an mHealth app for management of osteoarthritis. Investigators found that patients were supportive of app use, but physicians were more skeptical.

While the United States has contributed more mHealth-related research than any other country in the world [31], most research specifically investigating patients' and clinicians' broader attitudes about mHealth apps is based outside of the United States [19,23,24,26,27,32-42]. This is important because non-US countries may have regulatory requirements that influence attitudes in ways that are not relevant to US consumers and clinicians [43-45]. Moreover, realizing the potential of mHealth apps to improve health outcomes requires an understanding of both the individual and shared attitudes of patients and clinicians about mHealth. This understanding may facilitate the development of interventions to foster a shared mental model about the strengths and limitations of mHealth apps to enhance shared decision-making and increase patient and clinician adoption of safe and effective mHealth apps.

The objective of this study was to investigate patient and clinician attitudes about mHealth apps and to identify factors that may influence clinicians' willingness to recommend and patients' willingness to use mHealth apps. For this study, conversations about mHealth apps were grounded by discussion of a specific mHealth app being evaluated in an ongoing randomized controlled trial (RCT). The ability to examine these questions during the deployment of an existing mHealth app—rather than a theoretical tool—provided us with a unique opportunity to generate novel insights.

Methods

Qualitative Research Approach and Paradigm

The study team included clinicians (CAJ, MPD, and BKN), individuals with expertise in qualitative methods (SLK), implementation science (SLK), health services research (BKN, SLK, and JDP), and digital health technologies (DHT) (CAJ, BKN, MPD, and JDP). All members of the study team were employed by the University of Michigan (U-M), a large midwestern public university (CAJ, SLK, MPD, JDP, BKN) or a Veteran's Health Administration hospital (SLK, JDP). Given the qualitative nature of this study and the aim of understanding

XSL•FO
RenderX

participants' subjective experiences, a constructivist research approach was used.

Study Design and Setting

We used a descriptive qualitative study design to examine patient and clinician experiences with mHealth apps [46]. The study occurred at U-M from September 2023 to August 2024. The study adhered to the standards for reporting qualitative research reporting guidelines [47].

Participants and Sample

Patients

LowSalt4Life is a mHealth app designed to decrease dietary sodium intake in patients with hypertension. The app uses geofencing and reinforcement learning technology to deliver personalized, contextual push notifications about salt intake when patients are in grocery stores and restaurants. A pilot RCT (LowSalt4Life-1—NCT03099343) showed that contextual just-in-time notifications from the LowSalt4Life app resulted in greater reduction in dietary sodium intake in adults with hypertension compared to a control group (no app) over an 8-week period [48].

Purposive sampling methods were used to recruit patients from a cohort that participated in the LowSalt4Life-2 RCT (NCT05396001), which included 200 patients in the experimental arm (received personalized LowSalt4Life app) and 200 patients in the control arm [49]. Participants randomized to use the personalized LowSalt4Life app were invited via email to participate in a semistructured interview after exiting the LowSalt4Life-2 study.

Clinicians

Purposive sampling was used to recruit clinicians who provide direct care to patients with hypertension, including dietitians, pharmacists, primary care physicians, nephrologists, and cardiologists. Clinicians were contacted via the closed, moderated U-M Division of General Medicine email listserv or through targeted, direct emails to inquire about interest in participating in the study. Additionally, snowball sampling, in which interviewees were asked to identify colleagues who could provide valuable insights into the study, was used to recruit additional clinicians. These clinicians were also emailed invitations to participate in the study.

The clinician interviews preceded the patient interviews because patients were sampled from the LowSalt4Life-2 study and completion of the study was required before patients were approached for an interview.

Data Collection

The interview guides were developed using constructs from the technology acceptance model (TAM) [50,51], the unified theory of acceptance and use of technology (UTAUT) [52,53], the theoretical domains framework [54], and the technology integration model (TIM) [55].

The patient interview guide included two domains: (1) technical and material factors and (2) social and personal factors (Multimedia Appendix 1). The clinician interview guide included three domains: (1) technical and material factors, (2) social and personal factors, and (3) policy and organizational factors (Multimedia Appendix 1). The technical and material factors domain included constructs such as perceived usefulness and perceived ease of use. Constructs in the social and personal factors domain included attitudes, self-efficacy, identity, and social influence. Finally, workflow, training, and education were the constructs included in the policy and organizational factors domain.

The patient guide included 19 questions (plus probing questions), and the clinician guide included 14 questions (plus probing questions) (supplement). Both interview guides were iteratively updated as new information was encountered during interviews, and the final guides were approved by members of the study team (CAJ, SLK, BKN, JDP, and MPD).

The interviews were conducted either virtually via the Zoom platform or by phone by a clinician (CAJ) and a nonclinician research associate (SY). Prior to interviews, clinicians viewed a brief presentation about the LowSalt4Life app to (1) facilitate a shared understanding of the general capabilities of mHealth apps, (2) evaluate their attitudes about the LowSalt4Life app, and (3) have a common reference app to allow comparison of clinician and patient attitudes about mHealth apps in general. Each interview was audio recorded and transcribed.

Data Analysis

We conducted a deductive content analysis [56-58] using a rapid qualitative analysis approach to identify subcategories and categories within the domains of the theoretical frameworks. Rapid qualitative analysis is a particularly useful method when the data collection is targeted (eg, selection of interviewees) and processes (interview protocol) are highly structured [59-61]. The rapid qualitative analysis approach allowed us to maintain rigor while also working within the constraints of another study. Rapid qualitative analysis involves three analytic steps: (1) summarization of individual transcripts, (2) consolidation of transcript summaries by participant type (eg, patient or clinician), and (3) in-depth analysis [62].

Two study team members (CAJ, SLK) developed a summary template based on the domains in the interview guide, then independently reviewed, summarized, and discussed 3 interview transcripts to test and refine the template and summarization process. Thereafter, concurrent with ongoing data collection, each interview transcript was reviewed in detail, and transcript summaries that organized the data into the relevant domains were generated by the first author (CAJ). Data from the transcript summaries were transferred to patient or clinician data matrices to allow visualization of the data across domains, constructs, and participants. A detailed review of the data matrices by several study team members (CAJ, SLK, SY) led to the identification of patient and clinician subcategories within each domain. Through triangulation, the subcategories were then aligned to create broader categories by placing the patient and clinician data in a table that allowed comparison across domains and constructs (). CAJ and SLK met biweekly to discuss the data and resolve any discrepancies. The categories were finalized during meetings with members of the research team (CAJ, SLK, SY, JDP, and BKN).

XSL•FO
RenderX

Ethical Considerations

The study protocol (HUM00238747) was approved by the U-M Institutional Review Board. All participants provided written informed consent. All the study data was deidentified and stored in a Health Insurance Portability and Accountability Act–compliant, U-M–approved electronic database to ensure privacy and confidentiality of participants. Patients received a US $25 gift card after completing the interview. Clinicians did not receive an incentive for participating in the study.

Results

Patient and Clinician Information

Prior to initiating the study, we planned to interview 20 clinicians and 20 patients. However, data saturation (no new analytical information identified by the research team) was reached after 20 clinician interviews and 28 patient interviews. Among the 20 clinicians interviewed, the mean number of years in practice was 17 (SD 11.6) years (Table 1) and most were general internists (n=12, 60%). The clinician interviews lasted between 22 to 49 minutes. The mean age of the 28 interviewed patients was 59 (SD 12.1) years, and 16 (57.1%) patients were female (Table 1). Two (7.1%) patients self-identified as Asian, 4 (14.2%) self-identified as Black/African American, 21 (75%) self-identified as White, 1 (3.6%) self-identified as unknown, and none self-identified as Hispanic. The patient interviews lasted between 21 and 63 minutes.

Table . Patient and clinician demographic data.

Characteristics	Participants
Patients (n=28)	
Age (y), mean (SD)	59 (12.1)
Sex, n (%)	
Female	16 (57.1)
Male	12 (42.8)
Race or ethnicity,[a] n (%)	
Asian	2 (7.1)
Black/African American	4 (14.2)
Hispanic	0 (0)
White	21 (75)
Unknown	1 (3.6)
Education, n (%)	
High school graduate/GED[b]	1 (3.6)
2 y degree/some college	1 (3.6)
Bachelor’s degree	12 (42.8)
Graduate degree and/or professional degree	14 (50)
Clinicians (n=20)	
Sex, n (%)	
Female	13 (65)
Male	7 (35)
Specialty, n (%)	
Cardiology	2 (10)
General medicine	12 (60)
Nephrology	3 (15)
Nutrition	1 (5)
Pharmacy	2 (10)

[a]Race or ethnicity was self-reported.

[b]GED: general educational development.

Patient and clinician attitudes about mHealth apps fit within 1 of 7 categories related to the 3 theory-based domains (Table 2). Within the technical and material factors domain, the categories identified were apps as tools to improve health by extending

XSL•FO
RenderX

care, the role of apps in enhancing the patient–clinician relationship, and the need for simplicity and efficiency in app design. Categories in the social and personal factors domain included prior experience with mHealth apps, comfort with technology, and the influence of recommendations from trusted sources. Education and hands-on experience were the category identified in the policy and organizational factors domain. These categories are described in more detail below.

Table . Categories and subcategories related to patient and clinician attitudes about mHealth[a] apps.

Domain and category	Patient subcategory	Clinician subcategory
Technical and material factors		
Tools to improve health by extending care	mHealth apps are tools to improve health	mHealth apps can extend care beyond episodic visits
Enhancing the patient–clinician relationship	mHealth apps can improve patient–clinician engagement	mHealth apps can improve the patient–clinician relationship, if beneficial
Simplicity and efficiency	mHealth apps must be simple	Patients stop using mHealth apps that are complex or require a significant time commitment.
Social and personal factors		
Prior experience with mHealth apps	Positive experience with mHealth apps but concerns about data privacy	Personal use and prior recommendation of mHealth apps, but concerns about data privacy
Comfort with technology	Confident when using technology	Confident when using and discussing technology; older adults are not confident when using technology
Recommendations from trusted sources	Influence of trusted individuals	Influence of trusted organizations and colleagues
Policy and organizational factors		
Education and hands-on experience	Access to education about an mHealth app	Use of mHealth app prior to recommending

[a]mHealth: mobile health.

Technical and Material Factors

Tools to Improve Health by Extending Care

Both patients and clinicians described 1 of the benefits of mHealth apps as the capacity for use outside traditional health care settings. For patients, mHealth apps provide them with an always accessible tool to encourage healthier behaviors.

> *So, just that tool to have with you, you know an app on your phone. It's very helpful to be able to have that information with you.* [Patient 27, female, 55-year-old]

For clinicians, mHealth apps can reinforce their health-related recommendations outside of regularly scheduled clinic visits.

> *This is almost a way to extend my advice in a more practical way that allows them to be more aware and hopefully more compliant, or make better choices...* [Clinician 8; 19 years of experience]

Enhancing the Patient–Clinician Relationship

In general, patients and clinicians believed that mHealth apps could improve the patient–clinician relationship. Patients perceived that accurate, organized data from mHealth apps could improve their engagement with their health care, including providing additional information to facilitate meaningful conversations with clinicians and changes in management.

> *Yeah, I think it'd be terrific for my provider to be able to see the data...I think that that could be helpful to fuel a good conversation with a doctor.* [Patient 23, male, 57-year-old]

Clinicians suggested that equipping patients with a tool to support their health care would enhance their relationships with patients.

> *So, either way it would augment the physician-patient relationship. So even if it [patient data] does not feed to me, I think it would still allow me to give a patient a tool that can help them with their, with supporting their medical needs.* [Clinician 8; 19 years of experience]

While clinicians recognized the potential benefits of mHealth apps on the patient–clinician relationship, they expressed concerns about remembering available apps, a need for support from staff, and managing new data streams.

> *So, incorporation of this [mHealth app] into practice would involve others in the practice to assist. Either nurse practitioner or preferably a dietitian.* [Clinician 18; 44 years of experience]

> *On the clinician perspective, okay, now that you have all this extra data...Do you have the time to take a look at it. And then how do you actually turn that into changes in your management?* [Clinician 17; 8 years of experience]

Simplicity and Efficiency

Patients identified being easy to use and requiring a low time commitment as key considerations for the use of mHealth apps. In fact, they highlighted that complexity could be an important barrier.

XSL•FO
RenderX

> *Well, if it got too techy, I wouldn't have the patience... If it gets too technical, then no, it's not going to be useful or helpful to me.* [Patient 9, female, 86-year-old]

Additionally, clinicians expressed concerns about patients starting to use apps but eventually discontinuing use for reasons including complexity, too much of a time commitment, or too much manual data entry.

> *Have to keep an eye on the utilization. I think if they find it hard to use, they won't use it, or they might use it initially, but then fall off.* [Clinician 18; 44 years of experience]

> *If you're having the data, enter everything and it's not helping you along the way gets tiresome to patients. So, they usually use it for a while and give up.* [Clinician 16; 28 years of experience]

Social and Personal Factors

Prior Experience With mHealth Apps

Patients and clinicians both talked about positive experiences with mHealth apps. Many used apps for fitness tracking, weight loss, and sleep monitoring. For example:

> *I have apps that are really important to me, and one of them is my EKG app.* [Patient 3, female, 67-year-old]

> *I mean, I think I love mobile health applications. I mean, I have a ton of them on my phone for all kinds of different things.* [Clinician 9; 1 year of experience]

However, both patients and clinicians expressed concerns about data privacy, with some patients stating that they would be unwilling to use an mHealth app that tracked specific types of data (eg, location).

> *The part that I think would be almost a deal breaker for me is the like background location tracking, or like you have it on all the time...I just would have some like privacy concerns.* [Patient 16, male, 34-year-old]

> *And so, you know, some folks will probably have a hard time with the kind of knowledge that they are quote unquote being tracked.* [Clinician 17; 8 years of experience]

Comfort With Technology

Confidence and comfort with using technology were common considerations in the use of mHealth apps for both patients and clinicians. Patients, for the most part, expressed confidence when using various types of technology and generally described themselves as tech-savvy. Older patients noted that technologies should be designed to facilitate use by older adults and expressed a willingness to learn to use health technologies if necessary.

> *I am the type where if there is a need to use an app, especially for healthcare I would learn to use the app.* [Patient 8, male, 72-year-old]

> *There are some things that I can do on a computer, some things that I can do on my tablet. And if I learn to do something, I pretty much know it.* [Patient 9, female, 86-year-old]

Clinicians also described being open to new technologies and viewed themselves as tech-savvy. In addition, they expressed confidence when discussing technologies with patients.

> *Yeah, as long as I've had some time to work with it, I can. You know I'm reasonably confident. But I'm not. I don't have a phobia of technology. I am open to it.* [Clinician 18; 44 years of experience]

On the other hand, clinicians generally did not view patients as tech-savvy, and they specifically expressed concerns related to older adults' interest and ability to effectively use mHealth apps and other technologies.

> *I have some patients where I could probably tell you right off the bat. They're not gonna be interested, just cause they're not into tech type things, you know. I have some patients that still, you know, wish they could have their flip phone. And you know, some of the old school stuff or you know, patients that just prefer to stay off their phones altogether if they can. Usually, it's my older generation of patients, although sometimes they surprise you.* [Clinician 3; 10 years of experience]

Recommendations From Trusted Sources

Trusted sources were an important factor identified by both patients and clinicians related to their willingness to use or recommend mHealth apps. A recommendation from health care providers, friends, or family members would influence patients' decisions to use an app.

> *And my doctor, if she said, I want you to try this application because I think it's gonna make an improvement of your health, then I would do it.* [Patient 7, female, 62-year-old]

Guideline recommendations from professional societies or organizations would influence clinicians' decisions to recommend an app to their patients. Additionally, some clinicians stated that if their colleagues were recommending an app to their patients, or if a respected colleague recommended an app, they would be more inclined to recommend the app to their own patients.

> *I mean, I do think something like, you know, a major guideline recommendation would be at least something, particularly if you know I'm reading the recommendation, and I see there is something behind it. Potentially like a provider that I like, or trust would be useful too.* [Clinician 1; 15 years of experience]

Policy and Organizational Factors

Education and Hands-on Experience

Another facilitating factor for use of mHealth apps was gaining familiarity with an app. Specifically, patients were interested in readily available educational resources or assistance to increase their comfort with an app and to troubleshoot any problems that arise.

XSL•FO
RenderX

> *Maybe like something where you could have a chat session with somebody to help you get the answer you need.* [Patient 4, male, 57-year-old]

Most clinicians preferred to personally use an app prior to recommending it to patients. Additionally, many clinicians described recommending mHealth apps that they had personally used.

> *I think just seeing the like interfaces on the app, seeing what the patient sees is huge for me. I don't like to like throw them blind into something.* [Clinician 4; 2 years of experience]

Discussion

Principle Findings

mHealth apps are an increasingly important part of the health care landscape and can enhance patients' autonomy and engagement with their health care. Our findings show that overall, patients and clinicians have positive attitudes about mHealth apps, including their potential usefulness in extending care outside of traditional health care settings and enhancing the patient–clinician relationship. Patients and clinicians identified several common factors to facilitate the recommendation or adoption of mHealth apps. These factors included the importance of the ease of use of an app, receiving recommendations about mHealth apps from trusted sources, and access to education about or hands-on experience with mHealth apps. Prior experience with mHealth apps and comfort using technology were also described as facilitating factors by both patients and clinicians.

Other studies have indicated that both patients and clinicians have concerns about mHealth apps eroding the patient–clinician relationship and therapeutic alliances because of less face-to-face interaction [63]. This concern is well-founded as the introduction of technology can lead to a form of automation bias in which clinicians over-rely on mHealth apps and their outputs, or inappropriate delegation of tasks to mHealth apps [64]. Clinicians will need to exercise vigilance and make nuanced decisions about the appropriateness of mHealth apps for a given task to both ensure safety and that the patient–clinician relationship remains intact. Further, measures to avoid patient overreliance on mHealth apps will need to be in place because this could also negatively impact the patient–clinician relationship [65]. In a study by Preiser and colleagues [66], general practitioners in Germany noted that a symptom checker app could result in decreased unnecessary office visits and decreased clinician workload. However, the clinicians preferred that patients err on the side of caution and seek their help early to circumvent avoidable suffering. Ultimately, technology and safe, humane care should not be treated as mutually exclusive. Consistent with the TIM [55], this will require further exploration of how mHealth apps can extend patient and clinician capabilities, and both patients and clinicians acknowledging the strengths and limitations of technology and one another [67].

While mass media campaigns [68] and direct-to-consumer marketing [69] have proven effective at influencing patient health decisions, patients in our study noted that their attitudes about an app would be most influenced by trusted individuals (eg, friends, family, and health care providers). Therefore, clinicians must actively engage with patients and their family, friends, and caregivers to encourage the adoption and sustained use of safe and effective mHealth apps [70]. Further, studies have shown that patients' adoption of health-promoting behaviors may be influenced by individuals in nontraditional health care settings. For example, in an RCT by Victor and colleagues [71], Black male patients with uncontrolled hypertension were assigned to either an intervention group in which barbers at Black-owned barbershops encouraged meetings in barbershops with specialty-trained pharmacists who prescribed antihypertensive medication, or to a control group in which the barbers encouraged lifestyle modifications and doctor appointments. While both groups experienced improvement in their blood pressure at 6 months, the improvement was much greater in the group that met with the pharmacists and were prescribed antihypertensives. This study highlights both the benefits of treating patients in their natural environments and the benefits of leveraging the influence of trusted members of the community to promote healthy behaviors. Partnerships such as this should be explored to promote the adoption and sustained use of mHealth apps that have proven safe and effective.

Clinicians in our study noted that trusted colleagues, specialty organizations, and evidence-based guidelines, each of which may be related to the subjective norms described by UTAUT, would strongly influence their decision to recommend an app to a patient. However, clinicians did not describe app certification or approval by a regulatory agency (eg, Food and Drug Administration [FDA]) as necessary for recommending an app to a patient. In other studies, clinicians have identified a lack of certification as a barrier to recommending mHealth apps. For instance, in a study by Schroeder et al [36], clinicians expressed that they view mHealth apps similarly to other medical products and that apps should be given an official certification to ensure more trust. This study was conducted in Germany, which in 2020 became the first country in the world to approve certain mHealth apps for prescription and coverage by health insurers [26]. Additionally, Germany maintains a directory of approved mHealth apps managed by the German Federal Institute for Drugs and Medical Devices known as Digitale Gesundheitsanwendungen. As of July 2024, there were 56 mHealth apps listed by the German Federal Institute for Drugs and Medical Devices, and over 274,377 prescriptions were recorded from 2020 - 2023, with Zanadio (Sidekick Health Germany), a weight loss app, prescribed most frequently during that period [72]. Another study based in the United Kingdom showed that clinicians believed the most important factor encouraging mHealth app prescription was if it had a National Health Service stamp of approval [32]. The US FDA plays a role in the regulation of mHealth apps [43] but the vast majority of mHealth apps are not regulated by the FDA. While it may not be feasible for the FDA to rigorously evaluate every app, attention should be given to developing interventions that facilitate clinician confidence in the safety and efficacy of evidence-based mHealth apps as studies have shown that many

XSL•FO
RenderX

apps in app stores with high ratings often have low usability or clinical usefulness [73].

Our study showed that patients and clinicians have differing beliefs about patients' level of tech savviness. While most patients reported using mHealth apps, clinicians generally believed that older adults would not be interested, or they would be unable to effectively use mHealth apps [63]. A recent study showed that over 81% of older adults use DHTs, and over 40% reported any use of mHealth apps [74]. Our findings suggest that clinician bias could be a barrier to recommending mHealth apps to those who may benefit most (eg, older adults). During the COVID-19 pandemic, the use of DHTs significantly increased across all age groups. However, disparities widened with older adults significantly less likely to use DHTs compared to younger adults [75]. Still, older adults' increased use of DHTs during the COVID-19 pandemic suggests that they are interested and capable of using mHealth apps, but prior to the pandemic, they were not using them for various reasons that should be explored. Moreover, views expressed by older adults in our study are consistent with this as a 72-year-old patient stated: "I am the type where if there is a need to use an app, especially for healthcare I would learn to use the app." Older adults are likely to have a greater burden of chronic disease [11] and impairments (eg, physical and cognitive) that can limit their ability to use DHTs, and this should be considered when recommending mHealth apps to patients. The results from our study suggest that clinicians should take an individualized approach to recommending mHealth apps to older adults. Notably, age moderates the constructs in the UTAUT framework (eg, performance expectancy, effort expectancy, social influence, and facilitating conditions) that predict behavior intention to use technologies [52,53]. When studying older adults' use of mHealth apps, using a framework that does not include age, but does include variables associated with the aging process (eg, memory or physical limitations) may add more nuance and avoid introducing age-related bias. Further, interventions are needed to both address clinician bias [76] and to encourage patients and their families or caregivers to advocate for themselves.

Clinicians also noted barriers to recommending mHealth apps that included getting patients oriented, time constraints, a need for support from staff, and a lack of optimized workflows to facilitate review and use of data generated by mHealth apps, which highlights the importance of the environment and situational context described in the TIM [55]. These concerns must be weighed against the benefits to patients. For instance, patient-generated data has been shown to increase patient health literacy and to help patients better engage with their health providers [77,78]. User-centered mHealth apps also have the potential to increase adherence by providing individualized, just-in-time reminders and advice during vulnerable times outside of clinics and hospitals as demonstrated in conditions such as human immunodeficiency virus and congestive heart failure [79-81]. Like other studies, clinicians in our study expressed interest in having data from mHealth apps integrated into electronic health records (EHR) for efficient visualization, analysis, and application. This type of integration would require policy and organization-level interventions that support the flow of new data streams, interoperability between systems, and application of data from mHealth apps to improve patient–clinician engagement and health outcomes. Furthermore, health care organizations and app developers would need to partner with EHR vendors to ensure that their products are interoperable, allowing seamless data exchange between the EHR and mHealth apps. Collaborations between patients, clinicians, health system leaders, and developers will be necessary to ensure that apps address relevant clinical problems, are easy to use, and simple to set up. Thus, user-centered design principles addressing the patient, caregiver, and clinician perspectives must be used to facilitate recommendation by clinicians and adoption by patients.

James et al [82] described the "4E's (ease, evidence, experience and expectations)" that inform the mindlines influencing clinicians' decisions to recommend mHealth apps. Mindlines are collectively reinforced, internalized tacit guidelines, which are informed by brief reading, but mainly (by) clinicians' interactions with each other (as well as) opinion leaders, patients, and pharmaceutical representatives and by other sources of largely tacit knowledge that build on their early training and their own and their colleagues' experiences [83]. Our study reinforces the potential influence of the 4E's described by James and raises the question of whether the mindlines concept and the 4E's informing them could be expanded to patients. However, it should be noted that a category related to expectations did not emerge in our study, which could be due to the theoretical frameworks used and the constructs used. Therefore, the role of patients' and clinicians' expectations around mHealth apps should be explored.

Limitations

Strengths of this study include using an mHealth app that was being actively studied and deployed as a common frame of reference for discussions about mHealth apps with patients and clinicians. Additionally, we combined constructs from well-established technology adoption (eg, TAM and UTAUT), and implementation science (eg, theoretical domains framework and TIM) frameworks to account for limitations of each framework, which allowed us to identify facilitators and barriers that could inform development of interventions to increase recommendation and adoption of mHealth apps by patients and clinicians.

This study also had key limitations. First, it occurred at a single academic health center in a metropolitan area. There are known geographic differences in smartphone, internet, and DHT use, with lower rates of access among patients in rural areas. Second, our study only included patients enrolled in the LowSalt4Life-2 study, a study of a single mHealth app. Hence, the results may be affected by selection bias as participants may have had a favorable view of mHealth apps and interest as indicated by their participation in the study. Third, the qualitative design of our study does not allow us to infer that the attitudes expressed by participants in our study are shared by others. Fourth, because we used a deductive analytic approach that included domains from established frameworks, we may not have identified additional relevant or novel attitudes that may have been discovered with an inductive approach. Finally, we cannot

XSL•FO
RenderX

conclude that addressing the identified patient and clinician attitudes will result in increased recommendation or adoption of mHealth apps. Our study did not measure intention to use, behavior intention, use behavior, or technology use, which are downstream constructs in the TAM, UTAUT, and TIM frameworks, respectively [50-53,55]. Thus, studies are needed to determine the impact of interventions that address the identified attitudes on patient adoption and clinician recommendation of mHealth apps.

Conclusions

In this qualitative study, patients and clinicians generally had a positive view of the potential benefits of mHealth apps. Both groups also identified several similar factors that facilitate the use of mHealth apps. However, overall, patients' and clinicians' views about older patients' interest in and ability to use technology differed. As mHealth apps are developed and implemented, careful consideration must be given to both patients' and clinicians' attitudes and agreement and disagreement between these key stakeholders. This approach will allow the development of implementation interventions that facilitate recommendation of mHealth apps by clinicians and adoption of mHealth apps by patients.

Funding

Research reported in this publication was supported by the National Heart Lung and Blood Institute of the National Institutes of Health under Award Number: R33HL155498. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the US Federal Government. Funding support for SLK was also provided through a Department of Veterans Affairs Research Career Scientist Award (RCS 11- ‐ 222).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization - CAJ, SLK, BKN, MPD, JDP
Data curation - CAJ, SY
Formal analysis - CAJ, SLK, SY, BKN, MPD, JDP
Funding acquisition – CAJ, MPD, BKN
Writing - original draft - CAJ, SLK
Writing - review & editing - CAJ, SLK, BKN, MPD, JDP, SY

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient and clinician interview guides.
[DOCX File, 37 KB - mhealth_v14i1e77519_app1.docx]

References

1. Zhang M, Wang W, Li M, Sheng H, Zhai Y. Efficacy of mobile health applications to improve physical activity and sedentary behavior: a systematic review and meta-analysis for physically inactive individuals. Int J Environ Res Public Health 2022;19(8):4905. [doi: 10.3390/ijerph19084905]
2. Graham AK, Greene CJ, Kwasny MJ, et al. Coached mobile app platform for the treatment of depression and anxiety among primary care patients: a randomized clinical trial. JAMA Psychiatry 2020 Sep 1;77(9):906-914. [doi: 10.1001/jamapsychiatry.2020.1011] [Medline: 32432695]
3. Bricker JB, Watson NL, Mull KE, Sullivan BM, Heffner JL. Efficacy of smartphone applications for smoking cessation: a randomized clinical trial. JAMA Intern Med 2020 Nov 1;180(11):1472-1480. [doi: 10.1001/jamainternmed.2020.4055] [Medline: 32955554]
4. Gustafson DH, McTavish FM, Chih MY, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572. [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
5. Wells C, Spry C. An overview of smartphone apps. Can J Health Technol 2022;2(2). [doi: 10.51731/cjht.2022.269]
6. Nearly two-thirds of US consumers are mobile health app users. EMARKETER. URL: https://www.emarketer.com/content/nearly-two-thirds-of-us-consumers-mobile-health-app-users [accessed 2025-11-28]
7. MHealth market size, share & industry analysis, by category (mobile applications {disease & treatment management [healthcare providers/insurance, medication reminders, women health and pregnancy, disease specific], wellness management [fitness, lifestyle and stress] and others} and wearables {body & temperature monitors, sleep trackers, fitness trackers, glucose monitors, and others}), by service type (monitoring services, fitness and wellness solutions, and other services),

XSL•FO
RenderX

by service provider (mhealth app companies, pharmaceutical companies, and others), and regional forecast, 2026-2034. Fortune Business Insights. URL: https://www.fortunebusinessinsights.com/industry-reports/mhealth-market-100266 [accessed 2025-12-01]
8. Hengst TM, Lechner L, Dohmen D, Bolman CA. The facilitators and barriers of mHealth adoption and use among people with a low socio-economic position: a scoping review. Digit Health 2023;9:20552076231198702. [doi: 10.1177/20552076231198702] [Medline: 37691766]
9. Paradis S, Roussel J, Bosson JL, Kern JB. Use of smartphone health apps among patients aged 18 to 69 years in primary care: population-based cross-sectional survey. JMIR Form Res 2022 Jun 16;6(6):e34882. [doi: 10.2196/34882] [Medline: 35708744]
10. Murphy SL, Kochanek KD, Xu JQ, Arias E. Mortality in the United States, 2023. Centers for Disease Control and Prevention (CDC). 2024. URL: https://www.cdc.gov/nchs/products/databriefs/db521.htm [accessed 2025-11-28]
11. Watson KB, Wiltz JL, Nhim K, Kaufmann RB, Thomas CW, Greenlund KJ. Trends in multiple chronic conditions among US adults, by life stage, Behavioral Risk Factor Surveillance System, 2013–2023. Prev Chronic Dis 2025 Apr 17;22:240539. [doi: 10.5888/pcd22.240539] [Medline: 40245168]
12. Allen EM, Call KT, Beebe TJ, McAlpine DD, Johnson PJ. Barriers to care and health care utilization among the publicly insured. Med Care 2017 Mar;55(3):207-214. [doi: 10.1097/MLR.0000000000000644] [Medline: 27579910]
13. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125. [doi: 10.2196/jmir.5604] [Medline: 28428170]
14. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022 Apr 26;24(4):e35120. [doi: 10.2196/35120] [Medline: 35471414]
15. Saxon LA. Mobile health application solutions. Circ Arrhythm Electrophysiol 2016 Feb;9(2):e002477. [doi: 10.1161/CIRCEP.115.002477] [Medline: 26810595]
16. Madujibeya I, Alreshidi S, Aroh A. Abstract 17378: “if my doctors had recommended a mobile app to me back then, I wouldn’t have been in this situation today”: identifying the barriers and facilitators to the utilization of mHealth apps to improve self-care in patients with heart failure. Circulation 2020 Nov 17;142(Suppl_3):A17378. [doi: 10.1161/circ.142.suppl_3.17378]
17. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3(1):4. [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
18. Jacob C, Sezgin E, Sanchez-Vazquez A, Ivory C. Sociotechnical factors affecting patients' adoption of mobile health tools: systematic literature review and narrative synthesis. JMIR Mhealth Uhealth 2022 May 5;10(5):e36284. [doi: 10.2196/36284] [Medline: 35318189]
19. Alsahli S, Hor SY, Lam M. Factors influencing the acceptance and adoption of mobile health apps by physicians during the COVID-19 pandemic: systematic review. JMIR Mhealth Uhealth 2023 Nov 8;11:e50419. [doi: 10.2196/50419] [Medline: 37938873]
20. Portz J, Moore S, Bull S. Evolutionary trends in the adoption, adaptation, and abandonment of mobile health technologies: viewpoint based on 25 years of research. J Med Internet Res 2024 Sep 27;26:e62790. [doi: 10.2196/62790] [Medline: 39331463]
21. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR Mhealth Uhealth 2020 Feb 20;8(2):e15935. [doi: 10.2196/15935] [Medline: 32130167]
22. Wang T, Wang W, Liang J, et al. Identifying major impact factors affecting the continuance intention of mHealth: a systematic review and multi-subgroup meta-analysis. NPJ Digit Med 2022 Sep 15;5(1):145. [doi: 10.1038/s41746-022-00692-9] [Medline: 36109594]
23. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mHealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817. [doi: 10.2196/13817] [Medline: 31293246]
24. Sarradon-Eck A, Bouchez T, Auroy L, Schuers M, Darmon D. Attitudes of general practitioners toward prescription of mobile health apps: qualitative study. JMIR Mhealth Uhealth 2021 Mar 4;9(3):e21795. [doi: 10.2196/21795] [Medline: 33661123]
25. Schroeder T, Seaman K, Nguyen A, Gewald H, Georgiou A. Enablers and inhibitors to the adoption of mHealth apps by patients - a qualitative analysis of German doctors’ perspectives. Patient Educ Couns 2023 Sep;114:107865. [doi: 10.1016/j.pec.2023.107865] [Medline: 37356116]
26. Dahlhausen F, Zinner M, Bieske L, Ehlers JP, Boehme P, Fehring L. Physicians’ attitudes toward prescribable mHealth apps and implications for adoption in Germany: mixed methods study. JMIR Mhealth Uhealth 2021 Nov 23;9(11):e33012. [doi: 10.2196/33012] [Medline: 34817385]
27. Byambasuren O, Beller E, Hoffmann T, Glasziou P. Barriers to and facilitators of the prescription of mHealth apps in Australian general practice: qualitative study. JMIR Mhealth Uhealth 2020 Jul 30;8(7):e17447. [doi: 10.2196/17447] [Medline: 32729839]

XSL•FO
RenderX

28. Jezrawi R, Balakumar S, Masud R, et al. Patient and physician perspectives on the use and outcome measures of mHealth apps: exploratory survey and focus group study. Digit Health 2022;8:20552076221102773. [doi: 10.1177/20552076221102773] [Medline: 35646382]
29. Liu P, Li X, Zhang XM. Healthcare professionals' and patients' assessments of listed mobile health apps in China: a qualitative study. Front Public Health 2023;11:1220160. [doi: 10.3389/fpubh.2023.1220160] [Medline: 37780445]
30. Barber T, Sharif B, Teare S, et al. Qualitative study to elicit patients' and primary care physicians' perspectives on the use of a self-management mobile health application for knee osteoarthritis. BMJ Open 2019 Feb 1;9(1):e024016. [doi: 10.1136/bmjopen-2018-024016] [Medline: 30782723]
31. Alkhwaiter B, Aloud M, Almezeini N. User experience in mHealth research: bibliometric analysis of trends and developments (2007–2023). JMIR Mhealth Uhealth 2025 Nov 10;13:e75909. [doi: 10.2196/75909] [Medline: 41212171]
32. Leigh S, Ashall-Payne L, Andrews T. Barriers and facilitators to the adoption of mobile health among health care professionals from the United Kingdom: discrete choice experiment. JMIR Mhealth Uhealth 2020 Jul 6;8(7):e17704. [doi: 10.2196/17704] [Medline: 32628118]
33. Byambasuren O, Beller E, Glasziou P. Current knowledge and adoption of mobile health apps among Australian general practitioners: survey study. JMIR Mhealth Uhealth 2019 Jun 3;7(6):e13199. [doi: 10.2196/13199] [Medline: 31199343]
34. Salgado T, Tavares J, Oliveira T. Drivers of mobile health acceptance and use from the patient perspective: survey study and quantitative model development. JMIR Mhealth Uhealth 2020 Jul 9;8(7):e17588. [doi: 10.2196/17588] [Medline: 32673249]
35. Sivakumar B, Lemonde M, Stein M, Goldstein S, Mak S, Arcand J. Evaluating health care provider perspectives on the use of mobile apps to support patients with heart failure management: qualitative descriptive study. JMIR Cardio 2022 Oct 26;6(2):e40546. [doi: 10.2196/40546] [Medline: 36287588]
36. Schroeder T, Haug M, Georgiou A, Seaman K, Gewald H. Evidence of how physicians and their patients adopt mHealth apps in Germany: exploratory qualitative study. JMIR Mhealth Uhealth 2024 Jan 17;12:e48345. [doi: 10.2196/48345] [Medline: 38231550]
37. Zhang X, Raghavan V, Yang X. Health perceptions and attitudes on mobile health apps in China. Health Informatics J 2023;29(4):14604582231207745. [doi: 10.1177/14604582231207745] [Medline: 37837412]
38. Alzghaibi H. Healthcare practitioners' perceptions of mHealth application barriers: challenges to adoption and strategies for enhancing digital health integration. Healthcare (Basel) 2025 Feb 25;13(5):494. [doi: 10.3390/healthcare13050494] [Medline: 40077056]
39. Sezgin E, Özkan-Yildirim S, Yildirim S. Investigation of physicians' awareness and use of mHealth apps: a mixed method study. Health Policy Technol 2017 Sep;6(3):251-267. [doi: 10.1016/j.hlpt.2017.07.007]
40. Uncovska M, Freitag B, Meister S, Fehring L. Patient acceptance of prescribed and fully reimbursed mHealth apps in Germany: an UTAUT2-based online survey study. J Med Syst 2023 Jan 27;47(1):14. [doi: 10.1007/s10916-023-01910-x] [Medline: 36705853]
41. Hossain I, Ang YN, Chng HT, Wong PS. Patients' attitudes towards mobile health in Singapore: a cross-sectional study. Mhealth 2019;5:34. [doi: 10.21037/mhealth.2019.08.07] [Medline: 31620461]
42. Nguyen AD, Frensham LJ, Baysari MT, Carland JE, Day RO. Patients' use of mobile health applications: what general practitioners think. Fam Pract 2019 Mar 20;36(2):214-218. [doi: 10.1093/fampra/cmy052] [Medline: 29873708]
43. Policy for device software functions and mobile medical applications. US Food and Drug Administration (FDA). URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/policy-device-software-functions-and-mobile-medical-applications [accessed 2025-11-28]
44. Cortez NG, Cohen IG, Kesselheim AS. FDA regulation of mobile health technologies. N Engl J Med 2014 Jul 24;371(4):372-379. [doi: 10.1056/NEJMhle1403384] [Medline: 25054722]
45. Shuren J, Patel B, Gottlieb S. FDA regulation of mobile medical apps. JAMA 2018 Jul 24;320(4):337-338. [doi: 10.1001/jama.2018.8832] [Medline: 29971339]
46. Bradshaw C, Atkinson S, Doody O. Employing a qualitative description approach in health care research. Glob Qual Nurs Res 2017;4:2333393617742282. [doi: 10.1177/2333393617742282] [Medline: 29204457]
47. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251. [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
48. Dorsch MP, Cornellier ML, Poggi AD, et al. Effects of a novel contextual just-in-time mobile app intervention (LowSalt4Life) on sodium intake in adults with hypertension: pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Aug 10;8(8):e16696. [doi: 10.2196/16696] [Medline: 32663139]
49. Dorsch MP, Dempsey W, Krambrink A, et al. A just-in-time adaptive mobile application intervention to reduce sodium intake and blood pressure in patients with hypertension: rationale and design of the LowSalt4Life 2 trial. Am Heart J 2026 Feb;292:107294. [doi: 10.1016/j.ahj.2025.107294] [Medline: 41161590]
50. Yousafzai SY, Foxall GR, Pallister JG. Technology acceptance: a meta‐analysis of the TAM: Part 1. J Model Manag 2007 Nov 6;2(3):251-280. [doi: 10.1108/17465660710834453]
51. Yousafzai SY, Foxall GR, Pallister JG. Technology acceptance: a meta‐analysis of the TAM: Part 2. J Model Manag 2007 Nov 6;2(3):281-304. [doi: 10.1108/17465660710834462]

XSL•FO
RenderX

52. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003 Sep 1;27(3):425-478. [doi: 10.2307/30036540]
53. Momani AM. The unified theory of acceptance and use of technology: a new approach in technology acceptance. Int J Sociotechnol Knowl Dev 2020;12(3):79-98. [doi: 10.4018/IJSKD.2020070105]
54. Atkins L, Francis J, Islam R, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implementation Sci 2017 Dec;12(1):77. [doi: 10.1186/s13012-017-0605-9]
55. Shaw H, Ellis DA, Ziegler FV. The technology integration model (TIM): predicting the continued use of technology. Comput Human Behav 2018 Jun;83:204-214. [doi: 10.1016/j.chb.2018.02.001]
56. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
57. Moser A, Korstjens I. Series: practical guidance to qualitative research. Part 3: sampling, data collection and analysis. Eur J Gen Pract 2018 Dec;24(1):9-18. [doi: 10.1080/13814788.2017.1375091] [Medline: 29199486]
58. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
59. Lewinski AA, Crowley MJ, Miller C, et al. Applied rapid qualitative analysis to develop a contextually appropriate intervention and increase the likelihood of uptake. Med Care 2021 Jun 1;59(Suppl 3):S242-S251. [doi: 10.1097/MLR.0000000000001553] [Medline: 33976073]
60. Nevedal AL, Reardon CM, Opra Widerquist MA, et al. Rapid versus traditional qualitative analysis using the Consolidated Framework for Implementation Research (CFIR). Implementation Sci 2021 Dec;16(1):67. [doi: 10.1186/s13012-021-01111-5]
61. Kowalski CP, Nevedal AL, Finley EP, et al. Planning for and Assessing Rigor in Rapid Qualitative Analysis (PARRQA): a consensus-based framework for designing, conducting, and reporting. Implementation Sci 2024;19(1):71. [doi: 10.1186/s13012-024-01397-1]
62. Gale RC, Wu J, Erhardt T, et al. Comparison of rapid vs in-depth qualitative analytic methods from a process evaluation of academic detailing in the Veterans Health Administration. Implementation Sci 2019 Dec;14(1):11. [doi: 10.1186/s13012-019-0853-y]
63. Giebel GD, Speckemeier C, Abels C, et al. Problems and barriers related to the use of digital health applications: scoping review. J Med Internet Res 2023 May 12;25:e43808. [doi: 10.2196/43808] [Medline: 37171838]
64. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc 2012;19(1):121-127. [doi: 10.1136/amiajnl-2011-000089] [Medline: 21685142]
65. Abdelaziz S, Garfield S, Neves AL, et al. What are the unintended patient safety consequences of healthcare technologies? A qualitative study among patients, carers and healthcare providers. BMJ Open 2024 Nov 27;14(11):e089026. [doi: 10.1136/bmjopen-2024-089026] [Medline: 39608995]
66. Preiser C, Radionova N, Ög E, et al. The doctors, their patients, and the symptom checker app: qualitative interview study with general practitioners in Germany. JMIR Hum Factors 2024 Nov 18;11:e57360. [doi: 10.2196/57360] [Medline: 39556813]
67. Warraich HJ, Califf RM, Krumholz HM. The digital transformation of medicine can revitalize the patient-clinician relationship. NPJ Digital Med 2018;1(1):49. [doi: 10.1038/s41746-018-0060-2]
68. Wakefield MA, Loken B, Hornik RC. Use of mass media campaigns to change health behaviour. The Lancet 2010 Oct;376(9748):1261-1271. [doi: 10.1016/S0140-6736(10)60809-4]
69. Donohue JM, Cevasco M, Rosenthal MB. A decade of direct-to-consumer advertising of prescription drugs. N Engl J Med 2007 Aug 16;357(7):673-681. [doi: 10.1056/NEJMsa070502] [Medline: 17699817]
70. Carman KL, Dardess P, Maurer M, et al. Patient and family engagement: a framework for understanding the elements and developing interventions and policies. Health Aff (Millwood) 2013 Feb;32(2):223-231. [doi: 10.1377/hlthaff.2012.1133] [Medline: 23381514]
71. Victor RG, Lynch K, Li N, et al. A cluster-randomized trial of blood-pressure reduction in Black barbershops. N Engl J Med 2018 Apr 5;378(14):1291-1301. [doi: 10.1056/NEJMoa1717250] [Medline: 29527973]
72. Goeldner M, Gehder S. Digital health applications (DiGAs) on a fast track: Insights from a data-driven analysis of prescribable digital therapeutics in Germany from 2020 to mid-2024. J Med Internet Res 2024 Aug 29;26:e59013. [doi: 10.2196/59013] [Medline: 39208415]
73. Singh K, Drouin K, Newmark LP, et al. Many mobile health apps target high-need, high-cost populations, but gaps remain. Health Aff (Millwood) 2016 Dec 1;35(12):2310-2318. [doi: 10.1377/hlthaff.2016.0578] [Medline: 27920321]
74. James CA, Basu T, Nallamothu BK, Kullgren JT. Use of digital health technologies by older US adults. JAMA Netw Open 2025 Jan 2;8(1):e2454727. [doi: 10.1001/jamanetworkopen.2024.54727] [Medline: 39813034]
75. Qiu Y, Huang H, Gai J, De Leo G. The effects of the COVID-19 pandemic on age-based disparities in digital health technology use: secondary analysis of the 2017–2022 Health Information National Trends Survey. J Med Internet Res 2024 Dec 4;26:e65541. [doi: 10.2196/65541] [Medline: 39631070]
76. Mace RA, Mattos MK, Vranceanu AM. Older adults can use technology: why healthcare professionals must overcome ageism in digital health. Transl Behav Med 2022 Dec 30;12(12):1102-1105. [doi: 10.1093/tbm/ibac070] [Medline: 36073770]

XSL•FO
RenderX

77. Demiris G, Iribarren SJ, Sward K, Lee S, Yang R. Patient generated health data use in clinical practice: a systematic review. Nurs Outlook 2019;67(4):311-330. [doi: 10.1016/j.outlook.2019.04.005] [Medline: 31277895]
78. Luo Y, Oh CY, Jean BS, Choe EK. Interrelationships between patients' data tracking practices, data sharing practices, and health literacy: onsite survey study. J Med Internet Res 2020 Dec 22;22(12):e18937. [doi: 10.2196/18937] [Medline: 33350960]
79. Mouselimis D, Tsarouchas A, Vassilikos VP, et al. The role of patient-oriented mHealth interventions in improving heart failure outcomes: a systematic review of the literature. Hellenic J Cardiol 2024;77:81-92. [doi: 10.1016/j.hjc.2023.11.001] [Medline: 37926237]
80. Alaboud K, Shahreen M, Islam H, et al. Clinicians' perspectives in using patient-generated health data to improve ischemic heart disease management. AMIA Jt Summits Transl Sci Proc 2022;2022:112-119. [Medline: 35854732]
81. Tran BX, Bui TM, Do AL, et al. Efficacy of a mobile phone–based intervention on health behaviors and HIV/AIDS treatment management: randomized controlled trial. J Med Internet Res 2023 Apr 27;25:e43432. [doi: 10.2196/43432] [Medline: 37104001]
82. James CA, Dorsch MP, Piette JD, Nallamothu BK. Prescribing mobile health applications. Circ Cardiovasc Qual Outcomes 2024 Jun;17(6):e010654. [doi: 10.1161/CIRCOUTCOMES.123.010654] [Medline: 38525595]
83. Wieringa S, Greenhalgh T. 10 years of mindlines: a systematic review and commentary. Implementation Sci 2015 Dec;10(1):45. [doi: 10.1186/s13012-015-0229-x]

Abbreviations

DHT: digital health technology
EHR: electronic health record
FDA: Food and Drug Administration
mHealth: mobile health
RCT: randomized controlled trial
TAM: technology acceptance model
TIM: technology integration model
U-M: University of Michigan
UTAUT: unified theory of acceptance and use of technology

Edited by A Stone; submitted 14.May.2025; peer-reviewed by RN Al-Jabali, R Koch; accepted 20.Jan.2026; published 02.Mar.2026.

Please cite as:
James CA, Krein SL, Yon S, Dorsch MP, Piette JD, Nallamothu BK
Patient and Clinician Attitudes Toward Mobile Health Apps: Qualitative Study
JMIR Mhealth Uhealth 2026;14:e77519
URL: https://mhealth.jmir.org/2026/1/e77519
doi:10.2196/77519

© Cornelius A James, Sarah L Krein, Sarah Yon, Michael P Dorsch, John D Piette, Brahmajee K Nallamothu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 2.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Spanish Version of the mHealth App Usability Questionnaire (S-MAUQ): Translation, Adaptation, and Validation Study

Antonio L Carrillo-León[1*], PhD; Juan Falgueras-Cano[1*], PhD; Esther Díaz-Mohedo[2*], PhD

[1]Department of Languages and Computer Science, University of Málaga, C/ Doctor Ortiz Ramos s/n (Campus de Teatinos), Málaga, Spain

[2]Department of Physiotherapy, University of Málaga, Málaga, Spain

[*]all authors contributed equally

Corresponding Author:
Antonio L Carrillo-León, PhD
Department of Languages and Computer Science, University of Málaga, C/ Doctor Ortiz Ramos s/n (Campus de Teatinos), Málaga, Spain

Abstract

Background: In the current digital landscape, ensuring optimal usability is one of the most crucial factors determining the success of any mobile app. Questionnaire-based usability evaluations represent a highly prevalent methodology for this purpose. To date, questionnaires have been developed to assess the general system usability; however, there are hardly any questionnaires specifically designed to assess the usability of mobile health (mHealth) apps. The most widespread, the mHealth App Usability Questionnaire (MAUQ), has been developed in 4 versions according to the type of app (interactive or standalone) and the target user (patient or provider).

Objective: The objective of this study was to translate and validate the English version of the MAUQ (standalone, for patients) into a Spanish version (S-MAUQ).

Methods: The methodology used here follows that proposed by Sousa and Rojjanasrirat, which comprises 4 stages. The initial stage of the process entails a translation, harmonization, and adaptation procedure. The second and third entailed content validation (by 10 experts) and face validation (by 12 target users), respectively, which were conducted to evaluate the relevance and clarity of the questionnaire items. The item-level content validity index, scale content validity index (S-CVI), item-level face validity index, and scale face validity index (S-FVI), as well as the modified kappa statistic (κ) were used to evaluate interrater agreement among the raters, considering the probability of agreement by chance (Pc). The fourth and final stage of the process involved the assessment of the questionnaire's reliability. A sample of 61 young adult participants installed an mHealth app (the Yazio app), used it, and responded to the S-MAUQ. The Cronbach α value for the entire questionnaire and its subscales were then calculated.

Results: For the second stage, the S-CVI was initially 0.778. We removed items #14 and #15 from the Spanish version as they were unclear and not relevant. The S-CVI changed to 0.881. The third stage had an S-FVI of 0.927, indicating that the items are clear and straightforward for the nonexpert target user to understand. Furthermore, with each κ value >0.74, the validity of the instrument is supported. The fourth stage demonstrated the reliability of the S-MAUQ with a Cronbach α value of 0.87.

Conclusions: The final version of the S-MAUQ met the validation criteria, demonstrating reliability and validity that are comparable with those of the original version. Consequently, the S-MAUQ is suitable for evaluating the usability of mHealth apps for young Spanish adults. Further research involving larger and more diverse samples is recommended.

(JMIR Mhealth Uhealth 2026;14:e64787) doi:10.2196/64787

KEYWORDS

mHealth; usability questionnaires; usability evaluation; medical informatics; mobile health; app; mobile application; translation; adaptation; validation study; questionnaire-based evaluation; methodology; Spanish version; S-MAUQ; reliability; internal consistency; kappa statistic; digital health

Introduction

mHealth is an acronym for mobile health, which is used to describe the practice of medicine and public health that is supported by mobile devices [1]. Such mobile devices encompass mobile phones, tablets, and wearable devices, including smart watches. mHealth apps can collect health information for practitioners, researchers, and patients alike and can monitor patients' vital signs in real time. They facilitate direct care using mobile telemedicine and are used in the training of and collaboration with health workers [2].

The use of mHealth apps is increasing at a considerable rate [3], as evidenced by the growing number of users or patients [4]. Moreover, the emergence of innovative eHealth and

XSL•FO
RenderX

mHealth technologies has the potential to significantly improve the cost-effectiveness of the health care system [5].

Presently, these health apps are typically developed either by researchers or, in the majority of cases, by commercial companies, without collaboration between these two groups. The lack of interaction between researchers and commercial developers in the field of health-related apps has resulted in a situation where commercially available apps have not been subjected to scientific validation and apps that have been developed from research projects are not commercially available [6].

It is imperative that mHealth apps are designed in a manner that ensures optimal *usability*. Such apps must be straightforward to use and capable of achieving their intended objectives in an efficient manner. Well-designed mHealth apps have demonstrated cost-effectiveness through the education and empowerment of patients as well as through improved treatment adherence. However, some reviews have highlighted the absence of criteria for assessing the validity of mHealth apps.

mHealth apps can be classified into two principal categories: those designed for patients and those designed for health care providers. *Patient-focused* apps are designed to assist with the management of health behaviors, whereas those oriented toward *health care providers* facilitate services such as prescriptions and education. Furthermore, mHealth apps can be classified into two additional categories: *interactive* apps, which facilitate communication with health care providers, and *standalone* apps, which include features such as reminders but lack provider interaction [7].

The optimal *usability* of apps is characterized by several key attributes, including efficiency, satisfaction, learnability, memorability, and low error rates [7,8]. However, the intrinsic constraints of mobile devices, notably their limited display area, can present obstacles to the attainment of these objectives [9].

The People at the Center of Mobile App usability model considers the user, task, and context of use to evaluate usability. This includes the cognitive load associated with tasks such as exercising while using the app, as discussed by Harrison et al [10]. To minimize errors, usability should be assessed using the following criteria: effectiveness, efficiency, satisfaction, learnability, memorability, and cognitive load [10].

A few usability scales have been evaluation to ascertain their relative importance in the context of the development of general software systems and mobile apps. For example, the Mobile App Rating Scale (MARS) [11], its user version (uMARS) [12], System Usability Scale [13,14], and Post-Study System Usability Questionnaire [15]. However, these scales are not appropriate for evaluating mHealth apps, as they were not developed with mHealth apps as a specific focus.

One of the most pivotal usability scales for assessing the usability of mHealth apps is the mHealth App Usability Questionnaire (MAUQ) [7]. This assesses the ease of use, interface, satisfaction, and usefulness of mHealth apps for end users.

The standalone for patient version of MAUQ has been translated, among other languages, into Malay, German, and Italian [16-18]. It is noteworthy that, while this research was being developed, another related study was published (a Spanish version of MAUQ standalone, for patient [19]) but adapted to breastfeeding support apps. However, it has too many differences from the original and some incorrect formulas, as will be addressed in the discussion section.

The objective of our study was to translate and validate the English version of the MAUQ (standalone, for patients) into a Spanish version (S-MAUQ) for use in future mHealth app research and development.

The importance of the translation into Spanish and its corresponding validation lies in two fundamental points: on the one hand, the large size of the Spanish-speaking community and, on the other, the cultural diversity of the type of user of these apps. Without this linguistic adaptation, these questionnaires would be out of reach for many users.

Methods

Overview of the Questionnaire

As previously stated, the MAUQ, developed and validated by Zhou and colleagues [7], is designed to assess the usability of mHealth apps among patients and health care providers. There are two versions of MAUQ, one for interactive apps and one for stand-alone apps. Each also addresses two of its potential users: patient or health care professional. We will work on MAUQ for standalone apps intended for patients. Always consider that the term “patient” refers to a person who uses an mHealth app to manage their health either for maintenance or improvement.

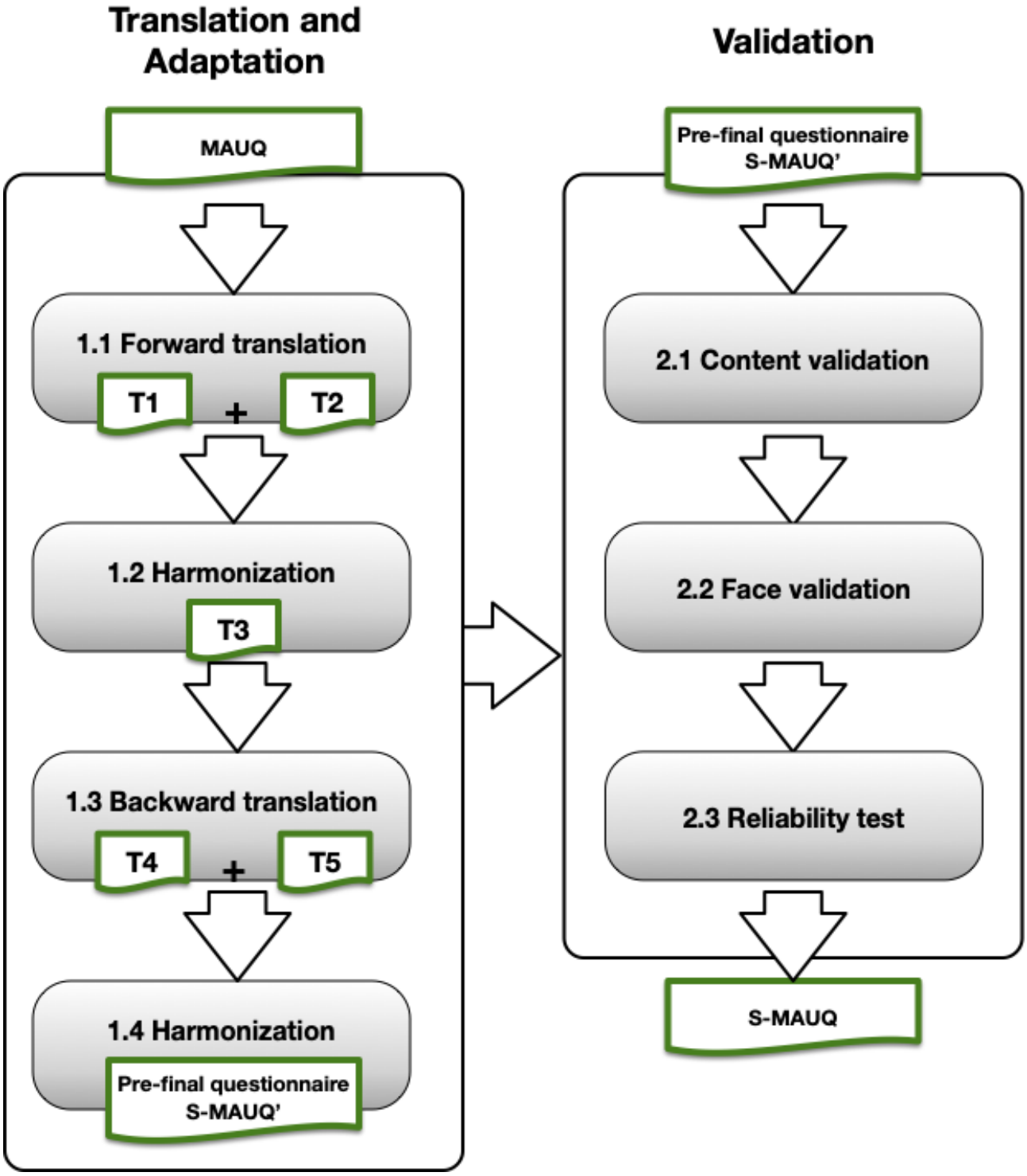
The questionnaire is comprised of 3 subscales or dimensions: ease of use (items #1 to #5), interface and satisfaction (items #6 to #12), and usefulness (items #13 to #18). Participants are required to indicate their level of agreement with each item on a 7-point Likert scale, with 1 representing strong disagreement and 7 representing strong agreement. The overall usability of the app is determined by calculating the total score and the mean score of all statements. A higher overall mean score is indicative of superior app usability. Conversely, a mean score <4 indicates suboptimal usability.

The original MAUQ (standalone, for patient version) exhibited robust internal consistency, with an overall Cronbach α value of 0.914. This indicates a high degree of correlation between the questionnaire items, indicating good reliability. The study was conducted in 2020 [7].

The methodology used to translate and validate the original MAUQ (standalone, for patient version) into Spanish followed the approach proposed by Sousa and Rojjanasrirat [20], with the objective of ensuring equivalence between the original and translated versions. The process is comprised of several steps (illustrated in Figure 1), which can be grouped into the following 4 stages: translation and adaptation, content validation, face validation, and internal consistency reliability.

XSL•FO
RenderX

Figure 1. The flow of translation, cross-cultural adaptation, and validation process for the Spanish version of the mHealth App Usability Questionnaire (S-MAUQ).



Translation and Adaptation Process

The preliminary phase of the translation process entailed the participation of two certified translators, each with a specialized understanding of both Spanish culture and linguistics. These professionals engaged in a forward translation from English to Spanish. One translator had a background in computer science, while the other had a background in health sciences. This resulted in the generation of two distinct documents designated T1 and T2.

Subsequently, the documents (T1 and T2) underwent a harmonization process. A third translator, a Spanish language teacher with more than 5 years of teaching experience, was tasked with identifying any ambiguities; discrepancies; and differences in words, sentences, grammar, and meaning. This was accomplished through a comparison of T1 and T2 with the original MAUQ. Any discrepancies and ambiguities were discussed and resolved through a process of consensus decision-making involving the third translator, the first two translators, and two members of the research team. The process

XSL•FO
RenderX

culminated in the production of an initial translated version of the MAUQ document (T3).

In the third phase of the process, the translated version of the MAUQ (T3) was provided to 2 independent native Spanish speakers who had acquired English as a second language. One individual (a computer science lecturer) was a content expert in the instrument's construct, while the other had no prior knowledge of the instrument (an English language teacher). The 2 participants then produced independent back-translated versions of the MAUQ (T4 and T5).

Last, a committee was established comprising 3 researchers, the translators involved in the initial and concluding stages, and the original developers of the MAUQ (when necessary). The committee conducted a comparative analysis of the documents with the original MAUQ, engaging in discussions via email and web-based meetings. Any ambiguities or inconsistencies in cultural meaning and expression of words and sentences and the format of the answers were addressed and resolved through consensus among the committee members. This process resulted in the development of the prefinal Spanish questionnaire, designated as S-MAUQ'.

Content Validation and Face Validation

Content validation is achieved by a panel of experts who assess the relevance and clarity of questionnaire items for measuring the usability of health-related mobile apps. Face validation is accomplished by a set of nonexperts or target users who specifically assess the clarity of the items.

Participants

For the content validation, a panel of 10 experts was formed. Our panel included 3 mobile app developers, 2 senior lecturers with a PhD in computer science, 2 usability experts, and 3 PhD senior lecturers and health-related practitioners.

For the face validation, 12 nonexpert participants were involved.

Recruitment

For the content validation, the 10 expert participants were carefully selected in order to adequately cover all areas of expertise involved. After being informed of the study's objective, all participants signed the corresponding informed consent form.

For the face validation, users were randomly selected from an openly chosen sample. The inclusion criteria were as follows: (1) native Spanish speaker, (2) older than 18 years, and (3) familiar with smartphones. The exclusion criteria were being (1) health-related practitioners, (2) mobile app developers, or (3) computer science students or professionals.

Participants were asked to complete an anonymous web form that included the informed consent, gender, date of birth, and level of education.

Setting

The group of expert participants completed their evaluations remotely and sent them to us via email. The nonexpert participants completed their evaluation online via an anonymous web form.

Study Procedures

In the content validation stage, each expert is required to assign a score on a scale of 1 to 5 to each item. A score of 1 indicates that the item was not relevant or clear, while a score of 5 indicates that the item was highly relevant and clear. Subsequently, scores of 1 to 3 were recategorized as 0, indicating irrelevance or unclarity, while scores of 4 and 5 were recategorized as 1, indicating relevance or clarity. The final categorization was conducted with the objective of obtaining dichotomous values. In numerous other studies, the range extends from 1 to 4. In contrast, in our study, the range was from 1 to 5. However, the current quality standards are more demanding, as a rating of 3 was considered unsatisfactory. In addition, there were 3 instances of unsatisfactory performance, as opposed to only 2 instances of acceptable performance.

On the other hand, in the face validation, which was less demanding with the nonexpert evaluators, the scores ranged from 1 (indicating that the item was not understandable) to 4 (indicating that the item was very understandable). Scores of 1 and 2 were recategorized as 0 (not understandable), and scores of 3 and 4 were recategorized as 1 (understandable). This was less demanding because the chosen scores were taken directly without considering the middle 3 negative values as we did with the experts.

The calculations for the 2 types of evaluation were as described in the following statements.

If we have N as the number of raters or evaluators; A_C as the number of expert raters who concur that the items are relevant, unambiguous, and intelligible; and A_F as the number of nonexpert raters who concur that the items are clear and comprehensible, the item-level content validity index (I-CVI) and the item-level face validity index (I-FVI) were calculated by dividing the propoortion of raters who agreed that an item was clear and relevant to the domain by the total number of raters.

In conclusion, it can be proposed that:

I−CVI=ACN and I−FVI=AFN

Items with an I-CVI or I-FVI of at least 0.79 were relevant, unambiguous, and understandable to the target users, indicating their potential inclusion in the questionnaire [21,22].

The scale content validity index (S-CVI) and scale face validity index (S-FVI) can be calculated as the means of the corresponding I-CVI and I-FVI. This is accomplished by dividing the sum of each item by the total number of items [23].

Moreover, it is essential to determine the level of agreement among the 10 experts. This can be accomplished by calculating the modified kappa statistic [21,24,25], which can be expressed as follows:

κCVI=I-CVI−Pc1−Pc and κFVI=I-FVI−Pc1−Pc

In this formula, P_C represents the probability of chance occurrence. The initial determination of the probability of agreement by chance for each item was determined using the

XSL•FO
RenderX

following formula (the formula for a binomial random variable) [21]:

Pc=(NA)·0.5N=N!A! (N−A)!·0.5N

This equation has its origins in the field of probability theory, specifically in the binomial distribution. The probability of obtaining precisely *A* successes in *N* independent Bernoulli trials (with the same rate *p*, in this case *1/2*) is given by the probability mass function (or discrete probability density function):

Pc=(NA) pk(1−p)N−A and for p=1/2 → Pc=(NA) 0.5N

Graphically, the interpretation criteria for the modified kappa statistic, as outlined by Wynd et al [26] and Cicchetti and Sparrow [27], is shown in Figure 2.

Figure 2. Interpretation criteria for the modified kappa statistic.



Reliability (Internal Consistency) Validation

Participants

For the internal consistency validation, the participants were selected from among students enrolled in the School of Engineering at the University of Malaga and the Campus of International Excellence (Andalucía Tech). The required sample size was calculated as 38 participants for the purposes of reliability testing, with an expected Cronbach α of 0.80 and an expected precision of 0.1 at 95% CI [28]. Given the projected rate of nonparticipation of 20%, the final sample size was determined to be 48.

Recruitment

Participants were recruited through direct invitations from the involved professors during class time. Those who agreed to take part in the study were required to complete a web-based informed consent form. Inclusion criteria were an absence of any known medical illness and ownership of a smartphone.

Setting

Participants performed the tasks on their smartphones and in their own classrooms and used a Google Form to submit their answers to the S-MAUQ.

Study Procedures

Participants were instructed to install the Yazio app on their mobile phones. Yazio, a standalone mHealth app for patients, was selected for the usability study due to its popularity. It is one of the most widely used nutrition apps in the world, helping people lead healthier lives globally. Its availability on both Android and iPhone devices [29,30] combined with its similarity to MyFitnessPal—an app frequently used in previous studies but that requires additional resources to be used—made it a suitable choice. A concise overview of the Yazio app, along with the essential steps for preliminary configuration, was displayed on the classroom projector. This procedure should be carried out exclusively at the point of the initial execution of the app following its installation.

Subsequently, participants were required to undertake the tasks detailed in Textbox 1.

XSL•FO
RenderX

Textbox 1. Sheet of tasks to be performed using the Yazio app.

1. Create initial profile when running the app for the first time (take a screenshot of the screen that appears at the end of the process, after discarding the tutorial)
2. Use the functions available in the Daily and/or Profile menu to add records or data for today:
   a. Add the following foods consumed during breakfast
      - 1 black coffee
      - 2 slices of bread (bimbo) without crust
      - Natural orange juice, 200 mL
      - Olive oil, 1 tablespoon
      - Product with barcode 8410095040362 (or scan it from the product code photo)

   b. Modify (in the profile) daily water amount goal: 2.5 liters
   c. Add water consumed: 1.75 liters
   d. Add activities (exercise completed) by pressing "More"
      - "walk, general" 30 minutes
      - "fast walk (6 km/h)" 10 minutes

   e. Add 2 notes or indications about how your day is going
      - "I slept well"
      - "bored"

   f. (Finally, take the necessary screenshots, 2 or 3, to collect the data entered)
3. Identify and note (via screenshot):
   a. The number of calories consumed during the day (due to ingested foods and drinks)
   b. Calories burned (from exercising, etc.)
   c. Calories we have left to consume to reach the day's goal
4. Explore and find a recipe that meets each of the following conditions:
   a. 1 vegan breakfast (between 150 and 200 kcal) and mark it as a favorite
   b. 1 vegan lunch (between 350 and 400 kcal) and mark it as a favorite
   c. See the recipes you have marked as favorites (and take a screenshot of them)

First, it is recommended that an initial profile be created and the desired weight goals be determined. Second, to add records or data pertaining to the current day, the functions available from the Daily and/or Profile menus need to be used. Third, it is requested that the number of calories consumed throughout the day be identified and recorded. This should include the calories derived from food and drink, the calories expended through exercise, and the remaining calories to be consumed to reach the daily goal. Fourth, it is required to conduct a comprehensive investigation and identify a recipe that aligns with the specific criteria outlined in step 4 in Textbox 1.

Following the completion of the tasks, the participants were invited to respond to the S-MAUQ in person using a Google Form.

The internal consistency of the S-MAUQ was evaluated by calculating the Cronbach α value for the entire questionnaire and its 3 subscales. A higher α value indicates greater internal reliability, and a value of ≥0.70 is generally considered acceptable as an indication of good internal reliability [31,32]. All statistical analyses were conducted using SPSS (IBM Corp), and the results were verified using Excel (Microsoft Corp).

Figure 1 provides an overview of the process of translation and cross-cultural adaptation as well as the subsequent validation stage for S-MAUQ'.

Ethical Considerations

No financial compensation was provided to participants at any stage of the study. Participation was voluntary and anonymous, except for members of the expert panel responsible for content validation, who were intentionally selected for their expertise. All participants (including the expert panel) were provided and signed an informed consent online, which included the possibility of opting out. The study adhered to the Declaration of Helsinki and was approved by the Bioethics Committee of the Medical University of Málaga (17 - 2022-H).

XSL•FO
RenderX

Results

Content Validation

Please refer to Table 1, which presents the calculated values (using Excel) of the I-CVI for relevancy and clarity, the (Pc), and the modified kappa agreement (κ) by the 10 experts for each item of the S-MAUQ. The S-CVI was calculated to be 0.778. However, as can be seen, there were 2 items (#3 and #6) with κ<0.4 (interpreted as poor agreement) and 2 others (#15 and #18) with 0.4<κ<0.6 (interpreted as fair). Considering these findings, we sought to further refine the instrument in accordance with the recommendations of the expert panel.

Table . Content validity index of item relevancy and modified kappa agreement by 10 experts in round 1.

Items	Rating of 4 or 5, n	Rating of 1 to 3, n	Item-level content validity index[a]	Probability of chance occurrence[b]	Modified kappa[c,d]	Interpretation
1	10	0	1.000	0.001	1.000	Excellent
2	9	1	0.900	0.010	0.899	Excellent
3	5	5	0.500	0.246	0.337	Poor
4	8	2	0.800	0.044	0.791	Excellent
5	9	1	0.900	0.010	0.899	Excellent
6	5	5	0.500	0.246	0.337	Poor
7	10	0	1.000	0.001	1.000	Excellent
8	8	2	0.800	0.044	0.791	Excellent
9	7	3	0.700	0.117	0.660	Good
10	8	2	0.800	0.044	0.791	Excellent
11	9	1	0.900	0.010	0.899	Excellent
12	9	1	0.900	0.010	0.899	Excellent
13	8	2	0.800	0.044	0.791	Excellent
14	7	3	0.700	0.117	0.660	Good
15	6	4	0.600	0.205	0.497	Fair
16	9	1	0.900	0.010	0.899	Excellent
17	7	3	0.700	0.117	0.660	Good
18	6	4	0.600	0.205	0.497	Fair

[a]Scale content validity index=0.778.

[b]Computed using the formula Pc = [N! / (A! * (N–A)!)] * 0.5N, where Pc is the probability of chance occurrence, n is the number of experts, and A is the number of experts who agree the items are relevant/clear.

[c]Computed using the formula κ=(item-level content validity index – Pc) / (1 – Pc), where Pc is the probability of chance occurrence.

[d]Kappa was interpreted according to the following criteria: excellent, κ>0.74; good, 0.6≤κ≤0.74; fair, 0.40≤κ≤0.59; poor, κ<0.40.

The results of the second round are presented in Table 2, which indicates an S-CVI of 0.850. However, 2 items (#14 and #15) demonstrated a markedly low level of relevance and clarity (I-CVI), accompanied by a similarly low κ (0.4≤κ≤0.6, interpreted as “fair”). Following a thorough review with our panel of experts, it was determined that both items should be removed from the preliminary Spanish questionnaire (S-MAUQ). In conclusion, the S-CVI was calculated to be 0.881 (without considering the 2 items that were subsequently eliminated).

XSL•FO
RenderX

Table . Content validity index of item relevancy and modified kappa agreement by 10 experts in round 2.

Items	Rating of 4 or 5, n	Rating of 1 to 3, n	Item-level content validity index[a]	Probability of chance occurrence[b]	Modified kappa[c,d]	Interpretation
1	10	0	1.000	0.001	1.000	Excellent
2	9	1	0.900	0.010	0.899	Excellent
3	8	2	0.800	0.044	0.791	Excellent
4	8	2	0.800	0.044	0.791	Excellent
5	10	0	1.000	0.001	1.000	Excellent
6	10	0	1.000	0.001	1.000	Excellent
7	10	0	1.000	0.001	1.000	Excellent
8	8	2	0.800	0.044	0.791	Excellent
9	7	3	0.700	0.117	0.660	Good
10	9	1	0.900	0.010	0.899	Excellent
11	10	0	1.000	0.001	1.000	Excellent
12	10	0	1.000	0.001	1.000	Excellent
13	9	1	0.900	0.010	0.899	Excellent
14	6	4	0.600	0.205	0.497	Fair
15	6	4	0.600	0.205	0.497	Fair
16	8	2	0.800	0.044	0.791	Excellent
17	8	2	0.800	0.044	0.791	Excellent
18	7	3	0.700	0.117	0.660	Good

[a]Scale content validity index=0.850.

[b]Computed using the formula Pc = [N! / (A! * (N–A)!)] * 0.5N, where Pc is the probability of chance occurrence, n is the number of experts, and A is the number of experts who agree the items are relevant/clear.

[c]Computed using the formula κ=(item-level content validity index – Pc) / (1 – Pc), where Pc is the probability of chance occurrence.

[d]Kappa was interpreted according to the following criteria: excellent, κ>0.74; good, 0.6≤κ≤0.74; fair, 0.40≤κ≤0.59; poor, κ<0.40.

Face Validation

Please refer to Table 3 for the calculated values of the I-FVI, the Pc, and the modified kappa agreement (κ) for each item. These were determined by the 12 target users. The S-FVI was calculated to be 0.927. A score >0.79 indicates that the items in the questionnaire are sufficiently clear and understandable for the nonexpert target user [27]. Furthermore, the modified kappa agreement for each item in S-MAUQ demonstrated excellent concordance (κ>0.74).

XSL•FO
RenderX

Table . Face validity index of item understandability and modified kappa agreement by 12 target users.

Items	Rating of 3 or 4, n	Rating of 1 or 2, n	Item-level face validity index[a]	Probability of chance occurrence[b]	Modified kappa[c]	Interpretation
1	12	0	1.000	0.000	1.000	Excellent
2	12	0	1.000	0.000	1.000	Excellent
3	9	3	0.750	0.054	0.740	Excellent
4	11	1	0.917	0.003	0.916	Excellent
5	11	1	0.917	0.003	0.916	Excellent
6	11	1	0.917	0.003	0.916	Excellent
7	11	1	0.917	0.003	0.916	Excellent
8	11	1	0.917	0.003	0.916	Excellent
9	10	2	0.833	0.016	0.831	Excellent
10	10	2	0.833	0.016	0.831	Excellent
11	12	0	1.000	0.000	1.000	Excellent
12	12	0	1.000	0.000	1.000	Excellent
13	12	0	1.000	0.000	1.000	Excellent
14[e]	—[f]	—	—	—	—	—
15[e]	—	—	—	—	—	—
16	12	0	1.000	0.000	1.000	Excellent
17	11	1	0.917	0.003	0.916	Excellent
18	11	1	0.917	0.003	0.916	Excellent

[a]Scale face validity index=0.927.

[b]Computed using the formula Pc = [N! / (A! * (N–A)!)] * 0.5N, where Pc is the probability of chance occurrence, n is the number of users, and A is the number of users who agree the items are understandable.

[c]Computed using the formula: =(item-level face validity index – Pc) / (1 – Pc), where Pc is the probability of chance occurrence.

[d]Kappa was interpreted according to the following criteria: excellent, $\kappa>0.74$; good, $0.6\leq\kappa\leq0.74$; fair, $0.40\leq\kappa\leq0.59$; poor, $\kappa<0.40$.

[e]Removed from the questionnaire.

[f]Not applicable.

Reliability (Internal Consistency) Validation

A total of 61 participants completed the web-based consent form and the S-MAUQ questionnaire for the purposes of the reliability study (the calculated sample size was 48). The participants were between the ages of 18 years and 22 years, and all were students (41 men, 20 women). The mean usability score (normalized to 1) for the Yazio app, as determined by the S-MAUQ, was 0.855 (SD 0.088). This indicates a high level of usability.

The Cronbach α value for the entire S-MAUQ was 0.87, and the values for the 3 subscales were 0.74 for the ease of use dimension (Q1-Q5), 0.74 for the interface and satisfaction dimension (Q6-Q12), and 0.64 for the usefulness dimension (Q13-Q16). The new numbering scheme was used for Q13-Q16, omitting the 2 deleted items.

Furthermore, the Cronbach α value for the complete questionnaire remained highly consistent when an item was deleted (see Table 4), indicating that the S-MAUQ exhibited good internal reliability.

XSL•FO
RenderX

Table . The internal consistency of the item's total statistics.

Items	Scale mean if item deleted[a]	Scale variance if item deleted	Corrected item total correlation	Cronbach α if item deleted[b]
1	89.46	87.086	0.532	0.854
2	89.07	85.129	0.635	0.850
3	89.10	85.190	0.570	0.852
4	89.07	87.562	0.470	0.857
5	89.15	84.061	0.462	0.859
6	88.90	89.323	0.345	0.863
7	89.49	87.754	0.439	0.858
8	88.97	89.432	0.403	0.860
9	89.13	87.949	0.436	0.858
10	89.28	85.404	0.593	0.851
11	89.74	79.797	0.599	0.851
12	89.13	85.949	0.702	0.849
13	89.20	88.294	0.442	0.858
14[c]	—[d]	—	—	—
15[c]	—	—	—	—
16	89.20	87.961	0.451	0.858
17	90.00	85.033	0.429	0.860
18	89.33	85.324	0.525	0.854

[a]Overall scale mean=95.21.

[b]Overall scale Cronbach α=0.87 (high internal consistency); ease of use subscale (Q1..Q5) Cronbach α=0.74; interface and satisfaction subscale (Q6..Q12) Cronbach α=0.74; usefulness subscale (Q13, Q16..Q18) Cronbach α=0.64.

[c]Removed from the questionnaire.

[d]Not appliclable.

Discussion

Principal Findings

Ensuring optimal usability of Spanish mobile apps is one of the most crucial factors for their use. Questionnaires are a common method for evaluating usability. This study outlined the process of translating, adapting, and validating the English version of the standalone MAUQ questionnaire for the patient into Spanish.

To this end, we followed the 4-step process proposed by Sousa and Rojjanasrirat [20], as it was outlined in the Methods section.

During the translation and adaptation phase, several meetings were held to ensure conceptual and linguistic equivalence for key terms such as interface, consistency, coherence, social setting, and health and well-being services. Final wording was established by consensus, often in consultation with the original questionnaire's authors.

In the content validation phase, the expert panel initially did not reach full agreement on several items. A second round of review narrowed the discrepancies but left 2 items (#14 and #15) unresolved. Both items were judged to be conceptually ambiguous and difficult to interpret, even for experts. Following a series of email consultations with the original authors on the meaning of and convenience of these items and their overlap with others, we confirmed the experts' results and their suggestions that these items could be removed from the final Spanish version.

During the face validation phase, the calculated FVI and κ coefficient yielded an S-FVI that substantially exceeded the recommended threshold, thereby ensuring the quality of the translation. The results of the study indicated that the items on the questionnaire were comprehensible and suitable for the target population.

Finally, the reliability analysis demonstrated strong internal consistency, with an overall Cronbach α of 0.87. The Spanish version maintained the original 3-dimensional structure of the English questionnaire, permitting direct cross-language comparison. Only the "usefulness" dimension was slightly affected, with an acceptable Cronbach α of 0.64 for exploratory studies. Overall, the S-MAUQ exhibited sound psychometric properties, indicating that it is a reliable and valid tool for assessing the usability of mobile health apps among Spanish-speaking patients. The final iteration of the S-MAUQ is presented in Table 5 for reference.

XSL•FO
RenderX

Table . The final Spanish version of the mHealth App Usability Questionnaire (S-MAUQ).

Item number	Question
1	La aplicación era fácil de usar.
2	Me resultó fácil aprender a utilizar la aplicación.
3	Al pasar de una pantalla a otra, la navegación era consistente, similar, previsible.
4	La aplicación me permitió usar todas las funciones que necesité (tales como visualizar e introducir información, responder a recordatorios, etc).
5	Cuando cometí un error usando la aplicación, fue fácil y rápido deshacerlo y volver a intentarlo.
6	Me gusta el diseño y aspecto de la aplicación.
7	La presentación de la información en la aplicación estaba bien organizada, y me permitió encontrar lo que buscaba.
8	La aplicación mostraba adecuadamente el progreso de mi actividad mientras interactuaba con ella.
9	Me sentiría cómodo/a utilizando esta aplicación en público, delante de otras personas.
10	El tiempo invertido en realizar las tareas deseadas me ha parecido adecuado.
11	Volvería a utilizar esta aplicación.
12	En general, estoy satisfecho/a con esta aplicación.
13	Creo que la aplicación puede ser útil para mi salud y bienestar.
14 (#16 in MAUQ)	Esta aplicación tiene todas las funciones que esperaba.
15 (#17 in MAUQ)	He podido realizar algunas tareas incluso cuando la conexión a internet era mala o no estaba disponible.
16 (#18 in MAUQ)	Esta aplicación me proporcionó una forma aceptable de recibir servicios de salud, tales como el acceso a material educativo, el seguimiento de mis actividades, o la realización de autoevaluaciones.

Comparison With Prior Research

Our findings align with prior validations of the MAUQ conducted to assess a variety of mobile apps developed for different purposes and user groups, including the MyFitnessPal in Malaysia for evaluating caloric intake, weight goals, and nutritional guidance [16]; the Left-Handed Doctor mHealth app in China for medical consultations and public health guidance [33]; the Enable app in Germany for cancer patients to register electronic patient-reported outcomes and monitor treatment side effects [17]; and the DIGICOG-MS app in Italy for cognitive assessment of individuals with multiple sclerosis [18] and patients with osteoporosis in the Persian study [34]. Collectively, our findings suggest that both the MAUQ and S-MAUQ demonstrate stable content validity across diverse contexts and could be applied to a wide range of mHealth and eHealth applications designed for different populations.

It is noteworthy that a Spanish adaptation [19] developed in parallel with our own work has not adhered to the MAUQ standard; instead, it was adapted for breastfeeding. Several items have undergone alterations and been reordered, and one item was eliminated. The original dimensions were modified from 3 to 4, but the results indicate a significant lack of internal consistency in 2 final subscales. Furthermore, there was an erroneous interpretation and implementation of the formula applied to calculate the level of relevance and clarity of the questionnaire items. Consequently, it was not possible to achieve comparable outcomes in different languages.

We also compared our work with the aforementioned studies that focus on the “standalone, for patients” version of the MAUQ. The Malaysian version of the MAUQ [16] underwent an analysis similar to ours, maintaining the 3-dimensional structure but without eliminating any element. The German version of the MAUQ [17] also performed a study similar to ours and to that of the Malaysian version. However, the creators of the German version considered eliminating 1 item but finally decided not to do so in order to maintain consistency with the original. The Italian version [18] did not undergo content or face validation; they had a panel of 6 experts who directly compared a translated version with the original questionnaire and made modifications to have a more understandable Italian version.

Limitations

In our study, as in other similar work [16], we surveyed a large sample of native Spanish-speaking students to provide a representative sample of the young adult demographic. In future research, criterion validation studies should be conducted to predict different outcomes for different domains and types of users of the S-MAUQ.

XSL•FO
RenderX

Conclusions

We addressed a significant gap in the field of health app evaluation in Spanish by translating and adapting the MAUQ, one of the most recognized usability questionnaires in this area. Despite the need to eliminate 2 items due to their lack of relevance and demonstrated confusion, S-MAUQ is a reliable tool in Spanish for evaluating the usability of mHealth apps.

Acknowledgments

We would like to thank Dr. L. Zhou for granting us permission to translate the mHealth App Usability Questionnaire (MAUQ) into the Spanish version, his help during the translation process, and clarifying some of the items in the questionnaire. Thank you also for conveying to us the interest and the various requests that he received from individuals and research groups interested in having the questionnaire available in Spanish. We also thank the panel of experts and the other participants involved in this study.

Funding

Funding for the open access charge (article processing fee [APF]): Plan Propio de la UMA - Grupo de investigación SICUMA (TIC160) - Máster de Formación Permanente de Fisioterapia en Disfunciones de Suelo Pélvico (Master Ref. 81424587), Universidad de Málaga.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization - ALCL, JFC, EDM
Formal analysis - ALCL, JFC, EDM
Investigation - ALCL, JFC, EDM
Methodology - ALCL, JFC, EDM
Supervision - ALCL, JFC, EDM
Validation - ALCL, JFC, EDM
Writing – original draft - ALCL, JFC
Writing – review & editing - ALCL, JFC, EDM

Conflicts of Interest

None declared.

References

1. Adibi S. A technology road map. In: Mobile health: Springer; 2015, Vol. 5. [doi: 10.1007/978-3-319-12817-7]
2. Pimmer C, Tulenko K. The convergence of mobile and social media: affordances and constraints of mobile networked communication for health workers in low- and middle-income countries. Mobile Media & Communication 2016 May;4(2):252-269. [doi: 10.1177/2050157915622657]
3. Ali EE, Chew L, Yap KYL. Evolution and current status of mhealth research: a systematic review. BMJ Innov 2016 Jan;2(1):33-40. [doi: 10.1136/bmjinnov-2015-000096]
4. Lunden I. 6.1B Smartphone users globally by 2020, overtaking basic phone subscriptions. TechCrunch. 2015 Jun 2. URL: https://techcrunch.com/2015/06/02/6-1b-smartphone-users-globally-by-2020-overtaking-basic-fixed-phone-subscriptions/?guccounter=12015 [accessed 2023-08-10]
5. de la Torre-Díez I, López-Coronado M, Vaca C, Aguado JS, de Castro C. Cost-utility and cost-effectiveness studies of telemedicine, electronic, and mobile health systems in the literature: a systematic review. Telemed J E Health 2015 Feb;21(2):81-85. [doi: 10.1089/tmj.2014.0053] [Medline: 25474190]
6. Ball E, Newton S, Rohricht F, et al. mHealth: providing a mindfulness app for women with chronic pelvic pain in gynaecology outpatient clinics: qualitative data analysis of user experience and lessons learnt. BMJ Open 2020 Mar 12;10(3):e030711. [doi: 10.1136/bmjopen-2019-030711] [Medline: 32165550]
7. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500. [doi: 10.2196/11500] [Medline: 30973342]
8. Nielsen J. Usability Engineering: Morgan Kaufmann; 1993.
9. Zhang D, Adipat B. Challenges, methodologies, and issues in the usability testing of mobile applications. Int J Hum Comput Interact 2005 Jul;18(3):293-308. [doi: 10.1207/s15327590ijhc1803_3]
10. Harrison R, Flood D, Duce D. Usability of mobile applications: literature review and rationale for a new usability model. J Interact Sci 2013;1(1):1. [doi: 10.1186/2194-0827-1-1]

XSL•FO
RenderX

11. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
12. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849] [Medline: 27287964]
13. Brooke J. SUS—a quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland AL, editors. Usability Evaluation in Industry: Taylor and Francis; 1996. [doi: 10.1201/9781498710411]
14. Lewis JR. The System Usability Scale: past, present, and future. International Journal of Human–Computer Interaction 2018 Jul 3;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
15. Lewis JR. Psychometric evaluation of the PSSUQ using data from five years of usability studies. Int J Hum Comput Interact 2002 Sep;14(3-4):463-488. [doi: 10.1080/10447318.2002.9669130]
16. Mustafa N, Safii NS, Jaffar A, et al. Malay version of the mHealth App Usability Questionnaire (M-MAUQ): translation, adaptation, and validation study. JMIR Mhealth Uhealth 2021 Feb 4;9(2):e24457. [doi: 10.2196/24457] [Medline: 33538704]
17. Moorthy P, Weinert L, Harms BC, Anders C, Siegel F. German version of the mHealth App Usability Questionnaire in a cohort of patients with cancer: translation and validation study. JMIR Hum Factors 2023 Nov 1;10:e51090. [doi: 10.2196/51090] [Medline: 37910144]
18. Podda J, Grange E, Susini A, et al. Italian version of the mHealth App Usability Questionnaire (Ita-MAUQ): translation and validation study in people with multiple sclerosis. JMIR Hum Factors 2024 Sep 30;11:e58079. [doi: 10.2196/58079] [Medline: 39347625]
19. Quifer-Rada P, Aguilar-Camprubí L, Gómez-Sebastià I, Padró-Arocas A, Mena-Tudela D. Spanish version of the mHealth App Usability Questionnaire (MAUQ) and adaptation to breastfeeding support apps. Int J Med Inform 2023 Jun;174:105062. [doi: 10.1016/j.ijmedinf.2023.105062] [Medline: 37037124]
20. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. J Eval Clin Pract 2011 Apr;17(2):268-274. [doi: 10.1111/j.1365-2753.2010.01434.x] [Medline: 20874835]
21. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health 2007 Aug;30(4):459-467. [doi: 10.1002/nur.20199] [Medline: 17654487]
22. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. Res Nurs Health 2006 Oct;29(5):489-497. [doi: 10.1002/nur.20147] [Medline: 16977646]
23. Zamanzadeh V, Ghahramanian A, Rassouli M, Abbaszadeh A, Alavi-Majd H, Nikanfar AR. Design and implementation content validity study: development of an instrument for measuring patient-centered communication. J Caring Sci 2015 Jun;4(2):165-178. [doi: 10.15171/jcs.2015.017] [Medline: 26161370]
24. Fleiss JL. Statistical Methods for Rates and Proportions, 2nd edition: John Wiley; 1981.
25. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960 Apr;20(1):37-46. [doi: 10.1177/001316446002000104]
26. Wynd CA, Schmidt B, Schaefer MA. Two quantitative approaches for estimating content validity. West J Nurs Res 2003 Aug;25(5):508-518. [doi: 10.1177/0193945903252998] [Medline: 12955968]
27. Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. Am J Ment Defic 1981 Sep;86(2):127-137. [Medline: 7315877]
28. Arifin WN, Unit of Biostatistics and Research Methodology, School of Medical Sciences, Universiti Sains Malaysia, Kelantan, MALAYSIA. A web-based sample size calculator for reliability studies. EIMJ 2018;10(3):67-76 [FREE Full text] [doi: 10.21315/eimj2018.10.3.8]
29. Ferrara G, Kim J, Lin S, Hua J, Seto E. A focused review of smartphone diet-tracking apps: usability, functionality, coherence with behavior change theory, and comparative validity of nutrient intake and energy estimates. JMIR Mhealth Uhealth 2019 May 17;7(5):e9232. [doi: 10.2196/mhealth.9232] [Medline: 31102369]
30. Fakih El Khoury C, Karavetian M, Halfens RJG, Crutzen R, Khoja L, Schols J. The effects of dietary mobile apps on nutritional outcomes in adults with chronic diseases: a systematic review and meta-analysis. J Acad Nutr Diet 2019 Apr;119(4):626-651. [doi: 10.1016/j.jand.2018.11.010] [Medline: 30686742]
31. Allen P, Bennett K, Heritage B. SPSS Statistics Version 22: A Practical Guide: Cengage Learning Australia; 2014.
32. Onwuegbuzie AJ, Daniel LG. A framework for reporting and interpreting internal consistency reliability estimates. Meas Eval Couns Dev 2002 Jul 1;35(2):89-103. [doi: 10.1080/07481756.2002.12069052]
33. Shan Y, Ji M, Xie W, et al. Chinese version of the Mobile Health App Usability Questionnaire: translation, adaptation, and validation study. JMIR Form Res 2022 Jul 6;6(7):e37933. [doi: 10.2196/37933] [Medline: 35793132]
34. Moulaei K, Dinari F, Sheikhtaheri A, Bahaadinbeigy K, Hajesmaeel-Gohari S. The Persian version of mhealth App Usability Questionnaire (MAUQ) for patients: a psychometric assessment study. BMC Med Inform Decis Mak 2025 Jan 23;25(1):38. [doi: 10.1186/s12911-025-02882-3] [Medline: 39849456]

XSL•FO
RenderX

Abbreviations

I-CVI: item-level content validity index
I-FVI: item-level face validity index
MARS: Mobile App Rating Scale
MAUQ: mHealth App Usability Questionnaire
mHealth: mobile health
Pc: probability of chance occurrence
S-CVI: scale content validity index
S-FVI: scale face validity index
S-MAUQ: Spanish version of the MAUQ
uMARS: MARS user version

Edited by L Buis; submitted 16.Aug.2024; peer-reviewed by A Daud, F Carrouel; revised version received 20.Dec.2025; accepted 11.Feb.2026; published 20.Apr.2026.

Please cite as:
Carrillo-León AL, Falgueras-Cano J, Díaz-Mohedo E
Spanish Version of the mHealth App Usability Questionnaire (S-MAUQ): Translation, Adaptation, and Validation Study
JMIR Mhealth Uhealth 2026;14:e64787
URL: https://mhealth.jmir.org/2026/1/e64787
doi:10.2196/64787

© Antonio L Carrillo-León, Juan Falgueras-Cano, Esther Díaz-Mohedo. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 20.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Trust and Use of Recommendations for Health Apps Among European Residents: Cross-Sectional Survey

Mariam Shokralla[1,2], MPH; Romy Fleur Willemsen[1,2], PhD; Marise Jeannine Kasteleyn[1,2], PhD; Niels Chavannes[1,2], Prof Dr; Esther Talboom-Kamp[1,3,4], MD, MBA, PhD

[1]National eHealth Living Lab, Leiden University Medical Center, Hippocratespad 21, Leiden, The Netherlands

[2]Department Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands

[3]Zuyderland Medical Center, Sittard-Geleen, The Netherlands

[4]Open University, Faculty Management Sciences, Heerlen, The Netherlands

Corresponding Author:
Mariam Shokralla, MPH
National eHealth Living Lab, Leiden University Medical Center, Hippocratespad 21, Leiden, The Netherlands

Abstract

Background: There is growing recognition of the role of health apps in addressing health care system challenges, yet app quality and evidence vary widely, and consumers have little decision support at the point of download.

Objective: This study aimed to explore which sources of recommendations European residents use and trust when choosing health apps and whether residents support government review and rating of health apps.

Methods: We conducted a cross-sectional online survey (December 7, 2022, to February 16, 2023) in 26 languages targeting residents of the European Economic Area, the United Kingdom, and Ukraine. The survey contained 11 questions covering demographics, types of apps used, sources of advice used and trusted, and views on government review and rating. We included only fully completed responses (N=1228). Descriptive statistics are presented as counts and percentages. For subgroup analyses, we dichotomized trust responses (“I trust” vs other responses) and tested associations with gender (Fisher exact test), age group, and education level (chi-square test); an adjusted significance threshold was applied (P<.004) to account for multiple testing. We followed the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) guidelines for reporting online surveys and STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) recommendations for observational studies.

Results: A total of 1228 respondents from 33 countries completed the survey; 1110 (90.4%) reported using one or more health apps. COVID-19 apps (763/1228, 62.1%) and activity apps (737/1228, 60.0%) were the most frequently used, whereas disease management (86/1228, 7.0%), diagnostic (77/1228, 6.3%), and treatment apps (70/1228, 5.7%) were the least used. Sources used to choose apps included family and friends (434/1228, 35.3%), health professionals (412/1228, 33.6%), and government or health authorities (358/1228, 29.2%). The most trusted sources were health professionals (987/1228, 80.4%), pharmacists (750/1228, 61.1%), and government or health authorities (736/1228, 59.9%). No statistically significant differences in trust by gender were observed (all P>.0038). Some differences by age and education were observed for select sources (eg, government or health authorities: χ^2_1=8.546; P=.003; family and friends: χ^2=19.133; P<.001). Overall, 1060 (86.3%) of 1228 respondents supported government review and rating of health apps (either directly or by commissioning another organization).

Conclusions: In this large multilingual European survey, most respondents reported experience with health apps and placed greatest trust in health professionals, pharmacists, and government or health authorities, yet professional recommendations were used less often than informal sources. There is clear public support for government-led review and rating schemes to guide consumer choice. Efforts to make trustworthy, easy-to-find information available at the point of download and to support health care professionals in recommending high-quality apps could help bridge the gap between trusted and used sources.

(JMIR Mhealth Uhealth 2026;14:e64468) doi:10.2196/64468

KEYWORDS

health apps; recommendations; trust; app assessment; ISO 82304-2

XSL•FO
RenderX

Introduction

Overview

Health systems in Europe face multiple challenges including aging populations, growing chronic disease burdens, rising costs, and workforce shortages. Digital health tools such as health and wellness apps ("health apps") are increasingly proposed as part of the response to these pressures [1]. The number of available health apps is large and growing (estimated >350,000 in app stores in 2020), but app quality and the underlying evidence base vary widely [2-4]. Consumers typically have limited reliable information at the point of download beyond app descriptions and star ratings, which do not reliably indicate clinical validity or safety [5-8].

One path to increasing public confidence in health apps is through recommendations from trusted sources (eg, clinicians or health authorities). Experimental and observational research indicates that who recommends an app affects the intention to download it and that clinicians are often trusted but may lack time, training, or trustworthy sources to recommend apps [9-14]. To date, evidence on which sources European residents actually use and trust is limited. This gap motivated the present cross-European survey, carried out as part of the Horizon Europe Label2Enable project, which aims to promote the newly developed [15] CEN-ISO/TS 82304-2:2021 health app quality framework [16].

Study Objectives

The primary objectives of this study were as follows:

- Identify which sources European residents use to choose health apps
- Identify which sources European residents trust when choosing health apps
- Assess whether residents think the government should review and rate health apps to support choice

The secondary objective was to explore differences in trust by gender, age group, and educational level.

Methods

Design and Participants

We conducted a cross-sectional online survey between December 7, 2022, and February 16, 2023, using the Castor electronic data capture system. Eligible participants were residents (aged ≥18 years) of the European Economic Area, the United Kingdom, or Ukraine who could read and respond to a digital survey. No further inclusion or exclusion criteria were applied. The survey was open for 10 weeks and was promoted via the Label2Enable project website, project partners, and social media channels. Because the study was exploratory and descriptive, we did not perform a sample size calculation; however, recruitment and dissemination were monitored to seek diversity across countries, age, gender, and education.

Survey Instrument

The 11-item survey (Multimedia Appendix 1) covered sociodemographic characteristics (country of residence, year of birth, gender, education level, self-reported health, and informal caregiver status); types of health apps used (17 options); sources of advice used to choose apps (15 options; multiple selections allowed); trust in the same sources (4-point Likert scale); and views on government review and rating (3 options). One open-ended question solicited final thoughts. The English version was drafted in plain language; professional translators translated the survey into 25 other languages (covering 22 of 24 official European Union languages plus Norwegian, Ukrainian, Arabic, and Turkish). Translations were validated by native speakers with expertise in digital health.

Ethical Considerations

The Medical Ethics Committee of Leiden–Delft–The Hague reviewed the study and waived the need for full ethics approval (reference 223069; November 14, 2022). Participants provided informed consent by completing an online consent statement before starting the survey; they were informed that participation was voluntary, responses were anonymous, and they could withdraw by closing the browser before submission. No personal identifiers were collected. Aggregate data were stored and analyzed in deidentified form. No compensation was offered to participants.

Translation and Dissemination

Separate language-specific links were hosted on the Label2Enable website, and social media assets were provided to project partners (eg, European Patients' Forum and EuroHealthNet) to assist dissemination. Response rates were monitored biweekly, and dissemination tactics were adjusted accordingly.

Statistical Analysis

Only fully completed surveys were included in the analysis (N=1228). Descriptive statistics were presented as counts and percentages; means and SDs were reported for continuous variables (age). To examine associations between demographic variables (gender: male, female, nonbinary, or other; age: 18 - 65 vs >65 years; education: low [high school or less or practical training] vs high [bachelor's, master's, or doctoral degree]) and trust in different sources, we dichotomized trust responses into "I trust" versus all other responses. For gender, we used the Fisher exact test when expected cell counts were small; for age and education, we used chi-square tests. Because multiple related tests were performed, we applied a Bonferroni-style adjusted significance threshold (α adj=0.05/13=.0038) [17]. All statistical analyses were performed in SPSS (version 27; IBM Corp). For qualitative data (open-ended comments) responses were translated to English using DeepL (DeepL SE) and subjected to thematic analysis following the Braun and Clarke six-step approach [18]. We followed the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) guidelines for online surveys and STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) for reporting cross-sectional observational studies.

XSL•FO
RenderX

Results

Respondents' Characteristics

A total of 1228 fully completed responses from 33 countries were included (Multimedia Appendix 1 lists countries and counts). The median age was 47.0 (mean 47.0, SD 14.8) years; 178 (14.5%) were aged 65 years or older. Most respondents were female (787/1228, 64.1%). Education was skewed toward higher education: 1001 (81.5%) of 1228 respondents reported a bachelor's, master's, or doctoral degree. Self-reported health was good or very good for 856 (69.7%) of 1228 respondents. A quarter (318/1228, 25.9%) of the respondents provided informal care at least occasionally (Table 1).

Table . Respondents' characteristics (N=1228).

Characteristics	Values
Age (y), mean (SD)	47.0 (14.8)
Age (y), n (%)	
65 or older	178 (14.5)
18 - 65	1050 (85.5)
Gender, n (%)	
Male	438 (35.7)
Female	787 (64.1)
Nonbinary	3 (0.2)
Education level, n (%)	
High school or less	121 (9.9)
Practical or vocational education	106 (8.6)
Bachelor's, master's, or doctoral or similar	1001 (81.5)
Self-reported health status, n (%)	
Poor	76 (6.2)
Fair	206 (24.1)
Good	592 (48.2)
Very good	264 (21.5)
Informal care, n (%)	
I provide informal care every day	106 (8.6)
I provide informal care at least once a week	78 (6.4)
I provide informal care, but not every week	134 (10.9)
I do not provide informal care	910 (74.1)

Use of Health Apps

Overall, 1110 (90.4%) of 1228 respondents reported using one or more health apps. The most used app types were COVID-19 apps (763/1228, 62.1%) and activity apps (737/1228, 60.0%). Disease management apps (86/1228, 7.0%), diagnostic apps (77/1228, 6.3%), and treatment apps (70/1228, 5.7%) were the least used (Table 2).

XSL•FO
RenderX

Table . Types of health apps used by respondents (multiple responses allowed; N=1228).

Type of health apps	Values, n (%)
COVID-19 app	763 (62.1)
Activity app	737 (60.0)
Health insurance app	446 (36.3)
Hospital or clinic app	417 (34.0)
Nutrition app	297 (24.2)
Personal health record app	271 (22.1)
Sleep app	270 (22.0)
Menstruation app	265 (21.6)
Vital signs app	217 (17.7)
Mindfulness app	186 (15.1)
Disease management app	86 (7.0)
Diagnostic app	77 (6.3)
Treatment app	70 (5.7)
Research app	68 (5.5)
Informal caregiver app	19 (1.5)
Other	68 (5.5)
I do not use a health app	118 (9.6)

Sources of Advice Used to Choose Health Apps

Respondents reported using recommendations from family and friends (434/1228, 35.3%), health professionals (412/1228, 33.6%), and government or health authorities (358/1228, 29.2%) more often than curated sources such as health app libraries (72/1228, 5.9%) or pharmacists (97/1228, 7.9%; Table 3).

Table . Sources of advice or tips used to choose a health app (multiple responses allowed; N=1228).

Source of advice	Values, n (%)
My family and friends	434 (35.3)
A health professional	412 (33.6)
A government or health authority	358 (29.2)
The App Store or Google Play store	291 (23.7)
A Google search	282 (23.0)
A health professional organization	247 (20.1)
Personal social media posts	183 (14.9)
A patient organization	163 (13.3)
Traditional media	155 (12.6)
A peer support group	129 (10.5)
A health app manufacturer	109 (8.9)
A pharmacist	97 (7.9)
A health app library	72 (5.9)
None of the above	88 (7.2)
Other	54 (4.4)

Trust in Sources of Advice

When asked which sources they trusted, respondents indicated the highest trust for health professionals (987/1228, 80.4%), pharmacists (750/1228, 61.1%), and government or health authorities (736/1228, 59.9%). Low trust was reported for personal social media posts (65/1228, 5.3%), app stores (144/1228, 11.7%), Google search (118/1228, 9.6%), and health app manufacturers (113/1228, 9.2%; Table 4).

XSL•FO
RenderX

Table . Trust in advice or tips for a health app from different sources (N=1228).

Source of advice	I do not trust, n (%)	I am not sure if I can or should trust, n (%)	I trust, n (%)	I do not know or have not thought about it, n (%)
A health professional	16 (1.3)	130 (10.6)	987 (80.4)	95 (7.7)
A pharmacist	50 (4.1)	283 (23.0)	750 (61.1)	145 (11.8)
A government or health authority	76 (6.2)	307 (25.0)	736 (59.9)	109 (8.9)
A health professional organization	70 (5.7)	342 (27.9)	688 (56.0)	128 (10.4)
A patient organization (website, brochure, commercial, etc)	71 (5.8)	328 (26.7)	673 (54.8)	156 (12.7)
Family and friends	103 (8.4)	506 (41.2)	512 (41.7)	107 (8.7)
A peer support group	201 (16.4)	540 (44.0)	337 (27.4)	150 (12.2)
A health app library	122 (9.9)	499 (40.6)	335 (27.3)	272 (22.1)
Traditional media	303 (24.7)	635 (51.7)	161 (13.1)	129 (10.5)
The App Store or Google Play store	370 (30.1)	591 (48.1)	144 (11.7)	123 (10.0)
A Google search	358 (29.2)	667 (54.3)	118 (9.6)	85 (6.9)
A health app manufacturer	351 (28.6)	651 (53.0)	113 (9.2)	113 (9.2)

Differences in Trust by Gender, Age, and Education

The Fisher exact test showed no statistically significant differences in trust by gender across the 13 tested sources at the α adj=.0038 level (all $P>.0038$). Chi-square tests for age and education showed some statistically significant differences for selected sources (eg, government or health authorities: $\chi^2_1=8.546$; $P=.003$; family and friends: $\chi^2_1=19.133$; $P<.001$; health app libraries: $\chi^2_1=8.249$; $P=.004$). However, when applying the adjusted threshold, only $P\leq.0038$ was considered statistically significant. Refer to Table 5 for full stratified results and *P* values.

Chi-square analyses indicated that levels of trust in most sources of advice did not differ substantially across age groups (Table 4). Notable exceptions included government and health authorities ($\chi^2_1=8.55$; $P=.003$), family and friends ($\chi^2_1=19.13$; $P<.001$), health app libraries ($\chi^2_1=8.25$; $P=.004$), app stores ($\chi^2_1=4.12$; $P=.04$), Google searches ($\chi^2_1=5.38$; $P=.02$), and health app manufacturers ($\chi^2_1=4.82$; $P=.03$), where significant differences were observed. Similarly, no consistent differences were found between education levels, although significant variation was detected for government and health authorities ($\chi^2_1=13.02$; $P<.001$) and health professional organizations ($\chi^2_1=8.07$; $P=.005$). Full test statistics for all sources of advice are presented in Table 5.

XSL•FO
RenderX

Table . Trust in sources of advice or tips for health apps, stratified by gender, age, and education.

Trusted advice source	I trust (n=1228), n (%)	Gender			Age group (y)			Education		
		Male, n (%)	Female, n (%)	*P* value	18 - 65, n (%)	>65, n (%)	*P* value	Low, n (%)	High, n (%)	*P* value
A health professional	987 (80.4)	350 (79.9)	635 (80.7)	.604	853 (81.2)	134 (75.3)	.11	174 (76.7)	813 (81.2)	.118
A pharmacist	750 (61.1)	262 (59.8)	487 (61.9)	.467	648 (61.7)	102 (57.3)	.421	129 (56.8)	621 (62.0)	.146
A government or health authority	736 (59.9)	272 (62.1)	463 (58.8)	.298	646 (61.5)	90 (50.6)	.003	112 (49.3)	624 (62.3)	<.001
A health professional organization	688 (56.0)	232 (53.0)	454 (57.7)	.247	602 (57.3)	86 (48.3)	.022	108 (47.6)	580 (57.9)	.005
A patient organization	673 (54.8)	225 (51.4)	447 (56.8)	.109	571 (54.4)	102 (57.3)	.392	120 (52.9)	553 (55.2)	.515
Family and friends	512 (41.7)	160 (36.5)	351 (44.6)	.011	463 (44.1)	49 (27.5)	<.001	81 (35.7)	431 (43.1)	.42
A peer support group	337 (27.4)	105 (24.0)	231 (29.4)	.081	289 (27.5)	48 (27.0)	.742	71 (31.3)	266 (26.6)	.152
A health app library	335 (27.3)	120 (27.4)	213 (27.1)	.313	301(28.7)	34 (19.1)	.004	51 (22.5)	284 (28.4)	.071
Traditional media	161 (13.1)	59 (13.5)	101 (12.8)	.424	136 (13.0)	25 (14.0)	.975	28 (12.3)	133 (13.3)	.701
The app store or Google Play stores	144 (11.7)	47 (10.7)	96 (12.2)	.259	130 (12.4)	14 (7.9)	.042	21 (9.3)	123 (12.3)	.199
A Google search	118 (9.6)	33 (7.5)	84 (10.7)	.06	108 (10.3)	10 (5.6)	.02	16 (7.0)	102 (10.2)	.147
A health app manufacturer	113 (9.2)	40 (9.1)	73 (9.3)	.99	104 (9.9)	9 (5.1)	.028	15 (6.6)	98 (9.8)	.134
Personal social media posts	65 (5.3)	21 (4.8)	43 (5.5)	.145	54 (5.1)	11 (6.2)	.798	11 (4.8)	54 (5.4)	.739

Views on Government Review and Rating of Health Apps

Among the 1228 respondents, a total of 1060 (86.3%) respondents indicated that either the government should review and rate health apps (n=664, 54.1%) or that the government should pay another organization to do so (n=396, 32.2%); 168 (13.7%) opposed government involvement (Table 6).

Table . Views on whether the government should review and rate health app quality (N=1228).

Response options	Values, n (%)
Yes, I think the government should review and rate health app quality	664 (54.1)
No, but I think the government should pay another organization to review and rate health app quality	396 (32.2)
No, I think the government should not review and rate health app quality and should not pay another organization to review and rate health app quality	168 (13.7)

XSL•FO
RenderX

Thematic Analysis of Open Comments

Of 1228 respondents, 213 (17.3%) provided free-text comments. Thematic analysis identified 3 main themes [19].

The first theme was exploring apps as a useful part of health care (prescription, reimbursement, and accessibility). Respondents expressed that apps should be prescribed and reimbursed to ensure their continuous improvement and that they should be made accessible to all patients. Health professionals should consider using data and apps to improve their work.

The second theme was identifying the need for quality assessment and transparency (certification and evidence), represented by this quote: “If you want something to work, you have to evaluate it.” Some referred to their role: “I would accept every recommendation first and then research the app myself and test it.” Few saw no need for an assessment: “Bad apps eventually phase themselves out as they increasingly turn out to be bogus.” or “The quality of the health apps can easily be assessed by health care professionals. I think it’s a waste of money to put much effort into this.”

The third theme revolved around who should review and rate apps (preference for independent and knowledgeable assessors). Respondents expressed mixed views; some mentioned assessors should be independent with a high ethical standard, have no interest in the app, and have extensive knowledge of both the clinical subject of the app, information and communication technology, and data privacy. Suggestions included a research institute without commercial interests with a high degree of transparency. Others indicated difficulty in identifying any trusted entity to review and rate apps. One comment summarized the 3 themes in one sentence: “Trust in health apps comes from certifications, scientific evidence, and the authorities’ approval and implementation.”

Discussion

Principal Findings

In this large multilingual European sample (N=1228), the majority reported prior or current use of health apps; COVID-19 apps and activity or wellness apps were the most common. Health professionals, pharmacists, and government or health authorities were the most trusted sources of advice, yet professional recommendations were used less often than informal sources such as family and friends. A strong majority favored government involvement in reviewing and rating apps to support consumer choice.

Used and Trusted Sources of Advice

Our findings echo prior studies from Australia and Europe that show clinicians are highly trusted but seldom provide app recommendations due to the lack of time, training, and accessible appraisal tools [9-14,20]. One could expect that people would use advice from sources they trust most. However, this discrepancy suggests a limited availability of professional recommendations despite high levels of trust in clinicians. In an Australian study, general practitioners highlighted the lack of knowledge of effective health apps and the lack of trustworthy sources to access high-quality apps as the 2 main barriers to prescribing or providing relevant health app advice [21-23]. A possible solution is to require app stores to display a health app label showing the app’s performance using standardized scales or to create curated health app libraries.

The observed high trust in pharmacists is consistent with literature pointing to pharmacists as potential contributors to digital health counseling [24]. Low public awareness and use of app libraries reflect earlier reports that curated portals are poorly known despite conceptual appeal [25]. Awareness of health app libraries remains limited. For example, one study reported that participants were unfamiliar with the concept of app portals or curated app libraries.[26] Although the idea of curated app repositories indexing government-evaluated and approved apps is appealing, existing initiatives have not yet fully met user needs. The National Health Service for instance, discontinued its central app library and instead integrated recommendations across its website to highlight apps that are already widely used and trusted. Other national initiatives include France’s *Mon Espace Santé* and Germany’s DiGA (*Digitale Gesundheitsanwendungen*; Digital Health Application) directory [27]. Despite the limited uptake among prescribers—only 4% of physicians in Germany had prescribed a DiGA as of early 2022 patient-reported experiences are encouraging: 63% reported a favorable clinical outcome, and 86% indicated willingness to use another DiGA in the future [28,29]. These findings suggest that curated health app libraries, when built on transparent and endorsed evaluation frameworks and made easy to access and navigate, could play an important role in guiding user choice and promoting the safe integration of digital health tools.

Implications for Policy and Practice

Three implications follow. First, governments and health authorities should consider transparent, evidence-based review and rating schemes (or commissioning trusted independent bodies) and ensure that appraisal results are visible at the point of download. Second, health professionals and pharmacists should be supported with accessible, up-to-date appraisal information and implementation support so they can confidently recommend high-quality apps. Third, further research should test which review and rating mechanisms most effectively change consumer behavior and health outcomes; public preference alone does not establish effectiveness.

Strengths and Limitations

Strengths include the large multilingual survey, the inclusive plain-language survey instrument, and mixed quantitative and qualitative analysis. A limitation of this study is that the sample was self-selected and skewed toward highly educated respondents, with overrepresentation from some countries (eg, the Netherlands: 403/1228, 32.8%), which limits generalizability. The mean age (47.0, SD 14.8 years) is older than many app user samples and may affect observed patterns of app types and trusted sources. These points are discussed and flagged as limitations. We did not apply poststratification weights to achieve population representativeness; given the exploratory aim, we report unweighted descriptive results but caution against overgeneralization. The thematic analysis was

XSL•FO
RenderX

performed on English translations of free-text responses, which may have influenced coding.

Conclusions

European residents in this large multilingual survey frequently use health apps and place greatest trust in clinicians, pharmacists, and government or health authorities—yet trusted sources are used less often than informal sources. There is clear public support for government-led review and rating or the commissioning of independent organizations to evaluate apps. Policymakers should prioritize transparent, accessible appraisal systems and strategies to support clinicians and pharmacists in recommending high-quality apps.

Acknowledgments

The study team thanks all consortium members of the Label2Enable project and partner organizations who assisted with recruitment and dissemination.

Funding

This study was conducted as part of the Horizon Europe Label2Enable project (Label2Enable). The project is supported by funding from the European Commission under the Horizon Europe programme. The funder had no role in study design, data collection, analysis, manuscript preparation, or decision to publish.

Data Availability

The aggregated deidentified dataset generated and analyzed during this study is available from the corresponding author on reasonable request. Multimedia appendices (including survey instrument, country list, and thematic quotes) are provided.

Authors' Contributions

Conceptualization: MS
Data curation: MJK
Formal analysis: RFW
Project administration: ET-K
Supervision: NC
Survey design: MS, RFW
Translation coordination: RFW
Thematic analysis: MS, RFW
Writing—original draft: MS
Writing—review and editing: all authors
All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The study survey, countries of respondents, and whose advice/ tips do particpants trust to choose a health app.
[DOCX File, 147 KB - mhealth_v14i1e64468_app1.docx]

References

1. Kao CK, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. PM R 2017 May;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
2. Bol N, Helberger N, Weert JCM. Differences in mobile health app use: a source of new digital inequalities? The Information Society 2018 May 27;34(3):183-193. [doi: 10.1080/01972243.2018.1438550]
3. van Velsen L, Beaujean DJMA, van Gemert-Pijnen JEWC. Why mobile health app overload drives us crazy, and how to restore the sanity. BMC Med Inform Decis Mak 2013 Feb 11;13(1):23. [doi: 10.1186/1472-6947-13-23] [Medline: 23399513]
4. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020 Feb 1;27(2):330-340. [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
5. de Chantal PL, Chagnon A, Cardinal M, Faieta J, Guertin A. Evidence of user-expert gaps in health app ratings and implications for practice. Front Digit Health 2022;4:765993. [doi: 10.3389/fdgth.2022.765993] [Medline: 35252957]
6. How reviews and ratings affect app store optimization. Business of Apps. 2019. URL: https://www.businessofapps.com/insights/reviews-ratings-affect-app-store-optimization/ [accessed 2026-03-14]

XSL•FO
RenderX

7. Kuehnhausen M, Frost VS. Trusting smartphone apps? To install or not to install, that is the question. Presented at: 2013 IEEE International Multi-Cognitive Methods in Situation Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (COGSIMA 2013); Feb 25-28, 2013. [doi: 10.1109/CogSIMA.2013.6523820]
8. Grundy Q. A review of the quality and impact of mobile health apps. Annu Rev Public Health 2022 Apr 5;43:117-134. [doi: 10.1146/annurev-publhealth-052020-103738] [Medline: 34910582]
9. Adjekum A, Blasimme A, Vayena E. Elements of trust in digital health systems: scoping review. J Med Internet Res 2018 Dec 13;20(12):e11254. [doi: 10.2196/11254] [Medline: 30545807]
10. Lupiáñez-Villanueva F, Folkvord F, Vanden Abeele M. Influence of the business revenue, recommendation, and provider models on mobile health app adoption: three-country experimental vignette study. JMIR Mhealth Uhealth 2020 Jun 4;8(6):e17272. [doi: 10.2196/17272] [Medline: 32496204]
11. Byambasuren O, Beller E, Glasziou P. Current knowledge and adoption of mobile health apps among Australian general practitioners: survey study. JMIR Mhealth Uhealth 2019 Jun 3;7(6):e13199. [doi: 10.2196/13199] [Medline: 31199343]
12. Byambasuren O, Beller E, Hoffmann T, Glasziou P. Barriers to and facilitators of the prescription of mHealth apps in Australian general practice: qualitative study. JMIR Mhealth Uhealth 2020 Jul 30;8(7):e17447. [doi: 10.2196/17447] [Medline: 32729839]
13. Alkhaldi O, McMillan B, Maddah N, Ainsworth J. Interventions aimed at enhancing health care providers' behavior toward the prescription of mobile health apps: systematic review. JMIR Mhealth Uhealth 2023 Feb 27;11(1):e43561. [doi: 10.2196/43561] [Medline: 36848202]
14. Results of Australia's health panel survey on recommendations and regulation of smartphone apps for health and wellness. : Consumers Health Forum of Australia; 2018.
15. Hoogendoorn P, Versluis A, van Kampen S, et al. What makes a quality health app-developing a global research-based health app quality assessment framework for CEN-ISO/TS 82304-2: Delphi study. JMIR Form Res 2023 Jan 23;7:e43905. [doi: 10.2196/43905] [Medline: 36538379]
16. Jafari M, Ansari-Pour N. Why, when and how to adjust your P values? Cell J 2019 Jan;20(4):604-607. [doi: 10.22074/cellj.2019.5992] [Medline: 30124010]
17. Dang A, Arora D, Rane P. Role of digital therapeutics and the changing future of healthcare. J Family Med Prim Care 2020 May;9(5):2207-2213. [doi: 10.4103/jfmpc.jfmpc_105_20] [Medline: 32754475]
18. Paradis S, Roussel J, Bosson JL, Kern JB. Use of smartphone health apps among patients aged 18 to 69 years in primary care: population-based cross-sectional survey. JMIR Form Res 2022 Jun 16;6(6):e34882. [doi: 10.2196/34882] [Medline: 35708744]
19. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
20. Bates DW, Landman A, Levine DM. Health apps and health policy: what is needed? JAMA 2018 Nov 20;320(19):1975-1976. [doi: 10.1001/jama.2018.14378] [Medline: 30326025]
21. Wykes T, Schueller SM. Why reviewing apps is not enough: Transparency for Trust (T4T) principles of responsible health app marketplaces. J Med Internet Res 2019 May 2;21(5):e12390. [doi: 10.2196/12390] [Medline: 31045497]
22. Albrecht UV. Transparency of health-apps for trust and decision making. J Med Internet Res 2013 Dec 30;15(12):e277. [doi: 10.2196/jmir.2981] [Medline: 24449711]
23. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3(1):14. [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]
24. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3:14. [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]
25. NHS apps library retired. NHS. 2024. URL: https://www.nhs.uk/ [accessed 2026-03-14]
26. Szinay D, Perski O, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of health and well-being apps and curated app portals: think-aloud and interview study. JMIR Mhealth Uhealth 2021 Apr 27;9(4):e27173. [doi: 10.2196/27173] [Medline: 33904827]
27. Essén A, Stern AD, Haase CB, et al. Health app policy: international comparison of nine countries' approaches. NPJ Digit Med 2022 Mar 18;5(1):31. [doi: 10.1038/s41746-022-00573-1] [Medline: 35304561]
28. Wangler J, Jansky M. The use of health apps in primary care-results from a survey amongst general practitioners in Germany. Wien Med Wochenschr 2021 May;171(7-8):148-156. [doi: 10.1007/s10354-021-00814-0] [Medline: 33570692]
29. Richter JG, Chehab G, Stachwitz P, et al. One year of digital health applications (DiGA) in Germany - rheumatologists' perspectives. Front Med (Lausanne) 2022;9:1000668. [doi: 10.3389/fmed.2022.1000668] [Medline: 36388899]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
DiGA: *Digitale Gesundheitsanwendungen* (Digital Health Application)
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

XSL•FO
RenderX

Edited by L Buis; submitted 21.Jul.2024; peer-reviewed by A Sriram, C Jacob, P Rasche, S Narayan; revised version received 08.Sep.2025; accepted 15.Dec.2025; published 09.Apr.2026.

Please cite as:
Shokralla M, Willemsen RF, Kasteleyn MJ, Chavannes N, Talboom-Kamp E
Trust and Use of Recommendations for Health Apps Among European Residents: Cross-Sectional Survey
JMIR Mhealth Uhealth 2026;14:e64468
URL: https://mhealth.jmir.org/2026/1/e64468
doi:10.2196/64468

© Mariam Shokralla, Romy Fleur Willemsen, Marise Jeannine Kasteleyn, Niels Chavannes, Esther Talboom-Kamp. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 9.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Reluctance to Use a Psycho-Oncology Mobile App Among Patients With Primary Breast Cancer: Retrospective Cross-Sectional Survey

Marta Pawełczak-Szastok[1], PhD; Anna Syska-Bielak[2], PhD; Aleksandra Krzywon[3], PhD; Michał Jarząb[2], MD, PhD

[1]Department of Bone Marrow Transplantation and Oncohematology, Maria Sklodowska-Curie National Research Institute of Oncology, Wybrzeże Armii Krajowej Street 15, Gliwice, Poland

[2]Center for Diagnosis and Treatment of Breast Diseases, Maria Sklodowska-Curie National Research Institute of Oncology, Gliwice, Poland

[3]Department of Biostatistics and Bioinformatics, Maria Skłodowska-Curie National Research Institute of Oncology, Gliwice, Poland

Corresponding Author:
Marta Pawełczak-Szastok, PhD
Department of Bone Marrow Transplantation and Oncohematology, Maria Sklodowska-Curie National Research Institute of Oncology, Wybrzeże Armii Krajowej Street 15, Gliwice, Poland

Abstract

Background: eHealth is an increasingly used method of health care in the field of psycho-oncology. While many reports highlight the positive impact of psychological eHealth tools, some patients refuse to use them.

Objective: This study aimed to expand knowledge of the motivation and psychoemotional functioning of patients who consciously refuse to use eHealth technology in the form of a mobile psycho-oncology app offered as part of a clinical trial. To our knowledge, this is the first study to address this topic.

Methods: A retrospective cross-sectional study was conducted between December 2022 and February 2023 to investigate the reasons why 56 patients with breast cancer refused to use the psycho-oncology mobile app offered as part of a clinical trial by the Breast Cancer Unit. The primary aim of the study was to analyze patients' self-reported reasons for not engaging with the app, while also exploring their psychoemotional functioning, including stress levels (measured using the distress thermometer), personality traits (measured using the Ten-Item Personality Inventory), coping strategies (measured using the Coping Orientation to Problems Experienced Questionnaire), and Self-efficacy (measured using the General Self-Efficacy Scale). Participants in this study declined the app intervention but agreed to participate in this separate observational study, indicating that their refusal was related to the app itself rather than to participation in clinical research in general.

Results: The patients experienced a clinically meaningful elevation in stress levels (mean 5, SD 2.1 points) and Self-efficacy (mean 32.1, SD 5.1 points). Among 5 dimensions of personality traits, patients scored highest in Agreeableness (mean 6.5, SD 0.8 stens) and Conscientiousness (mean 6.4, SD 0.9) and lowest in Neuroticism (mean 3.4, SD 1.8) (other dimensions: Extraversion [mean 5.8, SD 1.6 stens] and Openness to Experiences [mean 4.4, SD 1.5 stens]). In terms of coping with stress, patients most frequently used the strategies of Active Coping (mean 2.6, SD 0.5 points), Acceptance (mean 2.6, SD 0.6 points), and Seeking Emotional Support (mean 2.6, SD 0.6 points), and least frequently used the strategies of Psychoactive Substance Use (mean 0.2, SD 0.6 points) and Restraint (mean 0.5, SD 0.7 points). Patient responses regarding refusal to participate in app testing were divided into four categories: (1) Focus on Life Outside the Disease, (2) Focus on Disease and Treatment, (3) Denial Mechanism, and (4) Technical Issues. Statistically significant differences were found between the groups. The Focus on Life Outside the Disease group of patients had higher levels of Self-efficacy, lower Neuroticism, and more frequent use of the Positive reevaluation strategy compared to the other groups.

Conclusions: Our patients' decision not to use the eHealth psycho-oncology app was mainly influenced by characteristics suggesting better emotional coping with the disease and treatment. These factors were significantly more influential than other factors studied, particularly those related to technology. Assessing reasons for opting out of eHealth and associated psychomotional functioning may be important for improving patients' adoption of eHealth solutions.

(JMIR Mhealth Uhealth 2026;14:e71412) doi:10.2196/71412

KEYWORDS

eHealth; eHealth barriers; breast cancer; psychoemotional functioning; psycho-oncology; refusal to participate; mobile app; clinical trial

XSL•FO
RenderX

Introduction

Electronic communication–based tools (eHealth) are a method of providing health care services using web-based technologies [1]. eHealth has been developing for more than a decade [2-5]. Originally designed to expand access to health care through remote contact and reduce costs, during the COVID-19 pandemic, these solutions became increasingly common and entered the mainstream of care due to the requirements of social isolation [6]. The solutions developed during this period have proved beneficial to the health care system in terms of time, premises, and economy. Researchers emphasize widespread availability, improved health monitoring among the public, and increased awareness of healthy lifestyles and ways to prevent diseases and other undesirable psychological conditions [2,3]. Many reports indicate the positive impact of psychological eHealth tools, highlighting a number of their advantages [7-11]. eHealth solutions are also increasingly being used in specific areas of health care, such as psycho-oncology [2,12]. Their goal is to improve psychological well-being both during and after oncological illness. Most of these tools are based on the tenets of cognitive-behavioral therapy and are constantly being developed [13].

The number of publications regarding barriers to the use of eHealth technologies is lower than that outlining its advantages [14]. Most emphasize significant barriers, including inability to access the technology, lack of time for patients and clinicians, and lack of patient motivation or need to use it [14-16]. However, fears associated with the risk to privacy for health-related data, lack of trust in algorithm-based approaches, and the absence of optimal, safe, and trustworthy channels for the dissemination of distant eHealth tools are also vital potential limitations.

Few publications address the barriers associated with using eHealth solutions in the field of psycho-oncology, and even fewer provide evidence specific to cancer types [2]. In a comprehensive systematic review of eHealth technologies for breast cancer supportive care, Gyawali et al [2] identified only 2 studies reporting information on barriers to the use of eHealth interventions. These studies mention 2 groups of barriers. The first group is related to the failure to meet the necessary conditions for effective use of eHealth, including awareness of, openness to, and trust in eHealth among both patients and staff, digital competence, and a clear legal framework. The second group of barriers is mainly related to the disadvantages reported by patients during use, such as technical problems, mismatch between content and stage of treatment, lack of certain information, and discomfort with the lack of face-to-face contact.

In another recent review by Horn et al [17] on breast cancer survivors, 8 trials were identified that reported reasons for dropouts. The most frequent reason was time constraints, followed by unmet expectations of participants, technical problems, and physical and psychological health. In addition, researchers note the lack of uniform assessment scales for the usefulness of eHealth tools, and existing assessments mainly focus on technical aspects [18]. They also highlight that professionals have their own barriers to using eHealth [19].

Most studies on barriers to the use of eHealth in psycho-oncology have been conducted among patients who have used such solutions, reporting their perceived advantages and disadvantages [2,17]. We found only 1 survey that explicitly addressed patients who refuse to use such solutions [20]. This group of patients seems particularly interesting because of the growing demand for eHealth and the prospect that, in the near future, some interventions will only be possible using eHealth. Gupta et al [20] described only challenges in the recruitment process. Our goal was to expand knowledge of the motivation and psychoemotional functioning of patients who consciously refuse to use eHealth technology in the form of a mobile psycho-oncology app offered as part of a clinical trial. Specifically, we examined reasons for refusal and assessed patients' stress levels, Self-efficacy, personality traits, and coping strategies. Importantly, all participants included in this study declined the mobile app intervention but agreed to participate in a separate observational study related to psychoemotional functioning. This distinction indicates that their refusal was specific to the app itself rather than to a general reluctance to participate in clinical research. Therefore, this study aimed to explore and categorize the reasons for refusal of a mobile psycho-oncology app among patients with primary breast cancer and to examine the psychoemotional functioning associated with different refusal categories. To the best of our knowledge, this is the first study of its kind.

Methods

Materials and Procedure

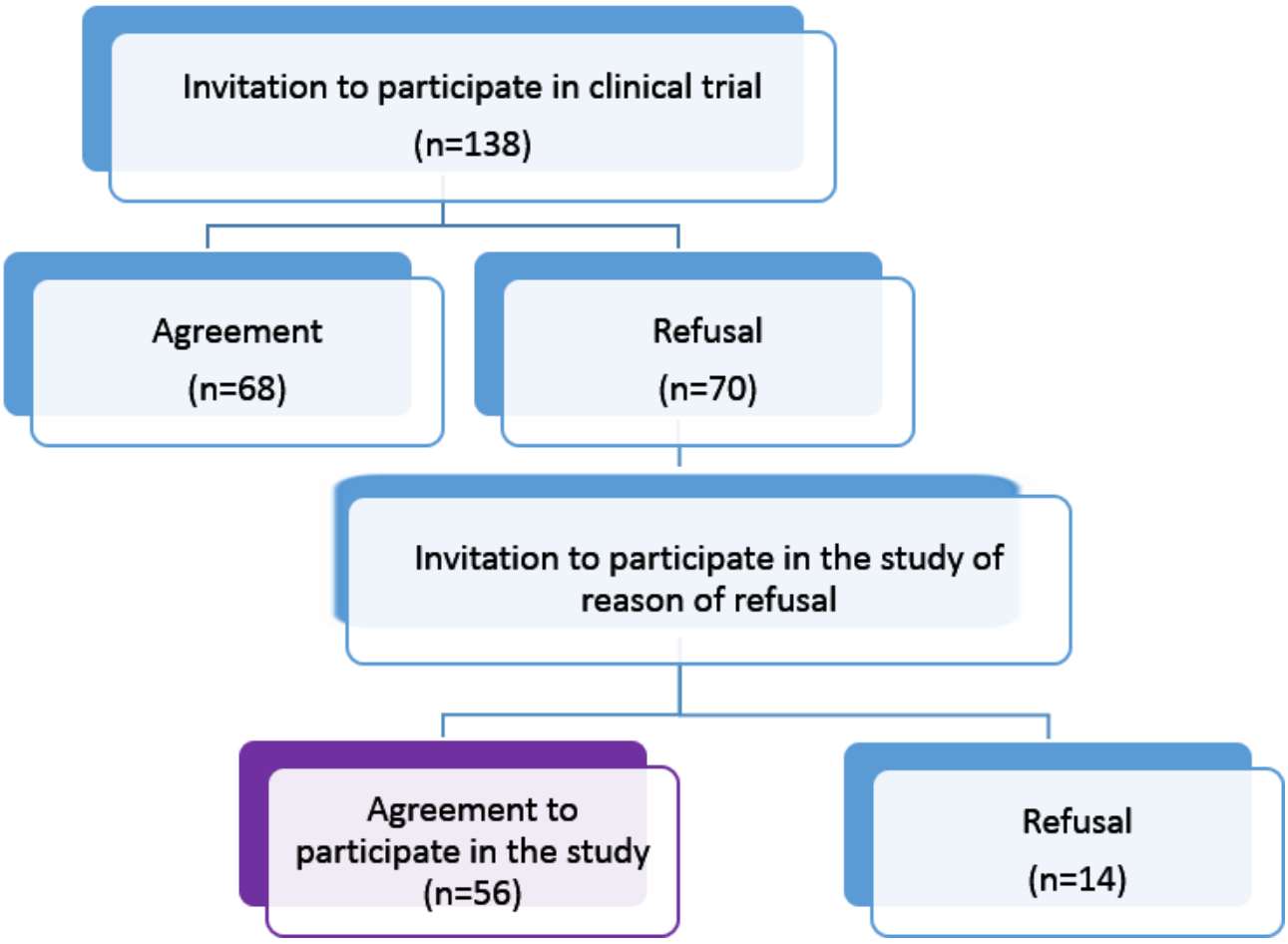
We conducted a retrospective study at the Breast Cancer Unit of the National Research Institute of Oncology in Gliwice, Poland, between February 2022 and November 2022. During this period, 138 patients with primary early-stage breast cancer at the beginning of the treatment process were offered participation in a clinical trial involving the use of a psycho-oncology mobile phone app. The clinical trial ran from February 2022 to May 2023. This clinical investigation concerned a medical device; therefore, in accordance with regulatory requirements for medical device trials, it was not assigned a clinical trial registration number (eg, European Union Drug Regulating Authorities Clinical Trials database). The app aimed to support the adaptation to the diagnosis and treatment of cancer based on cognitive-behavioral therapy. It comprised 22 visually appealing lessons, practical exercises (breathing exercises, imagery-based exercises, values-based activities, mindfulness exercises, and gratitude practice), emotion monitoring, and practical information and tips. The app was introduced to patients through a short video and was provided at no cost to those who agreed to join the clinical trial to evaluate its effectiveness. The details of the intervention are described elsewhere [21], and the final trial analysis is pending.

This report specifically focuses on patients who declined to use the mobile app within the clinical trial. Of the 138 female patients who were offered participation in the trial, 70 (51%) refused to use the app. A subsequent study involving individuals

XSL•FO
RenderX

who declined to use the app was conducted from December 2022 to February 2023. Of the 70 patients who declined, 56 agreed to participate in this study. Notably, among the remaining 14 patients who declined to participate, 6 did not attend the study appointment, and 5 provided no reason for their refusal other than unwillingness to be tested. These individuals may be considered potentially reluctant to participate in research in general. In contrast, 56 patients agreed to participate in our study, which suggests that their previous refusal was not driven by a general aversion to research but rather by specific concerns or disinterest regarding the mobile app intervention. In addition, 3 patients declined participation because of significant deterioration in health and severe pain symptoms, making it impossible to engage with a psychologist. The study recruitment and selection process is provided in Figure 1. The reporting of this observational study follows the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines.

Figure 1. Flowchart of study selection for evaluating the mobile psycho-oncology app among patients with primary breast cancer, Gliwice, Poland (2022-2023).



Measures

Demographic and Clinical Factors

Each participant answered survey questions regarding demographic issues, including age, marital status, and parental status. We reviewed patients' electronic medical records to obtain data regarding cancer diagnosis, treatment stage, date of diagnosis, and initial level of distress measured during the initial psychological interview conducted routinely for all patients at the local Breast Unit.

Reasons for Refusal

Reasons were collected using an open-ended question specifically constructed and designed for this study. Patients were asked: "Please indicate the main reason why you decided not to use the psycho-oncology mobile app offered as part of the study." Participants provided their responses in writing during an individual session conducted in the presence of a clinical psychologist, who was available to clarify instructions and provide support if needed. The assessment session took place during a subsequent hospital visit, separate from the encounter in which the patient initially declined participation in the mobile app trial. Two clinical psychologists independently coded the responses, identifying a total of 10 categories of refusal reasons. The categorization followed a data-driven process: each psychologist reviewed the answers separately and generated preliminary codes reflecting the core meaning of participants' statements. Discrepancies were resolved through consensus discussion. These initial codes were then grouped into broader thematic categories based on conceptual similarity and clinical relevance. Where applicable, the process was informed by previous literature, particularly Gupta et al [20], who described barriers to the adoption of eHealth interventions among patients with breast cancer. While most of our categories corresponded to domains identified in earlier research, one

XSL•FO
RenderX

clinically important category—reluctance to recall disease-related content—emerged uniquely from our data and is discussed further in the "Discussion" section. Importantly, the responses captured in this process referred specifically to the patients' decision to decline the mobile app, not to research participation in general. This combined approach ensured that the final categorization was both empirically grounded and clinically meaningful. The thematic categories identified through this process are summarized in Textbox 1, which presents the clinically relevant reasons for declining the psycho-oncological app.

Textbox 1. Reasons for refusal to use the psycho-oncological app within the clinical trial, Gliwice, Poland (2022 - 2023).

Lack of time—childcare responsibilities:

- Participants reported limited time due to childcare responsibilities, such as having a small child and being unable to find time for additional activities.

Lack of time—difficult family situation:

- Some respondents were engaged in caring for a disabled family member, including a disabled mother, child, or spouse.

Lack of time—demanding job:

- Several participants indicated being highly absorbed in professional work, describing their jobs as very responsible and demanding, leaving no available time.

Lack of time—frequency of medical appointments:

- Frequent medical examinations, consultations, and hospitalizations resulted in a lack of time for further commitments.

Poor physical and mental well-being:

- Respondents described mental exhaustion and poor physical condition, limiting their ability to focus on additional tasks.

Reluctance to recall disease-related content:

- Some participants did not wish to think about their illness at home and preferred to avoid disease-related content.

No mental need, feeling good:

- Several patients reported not feeling a need for psychological support or that the diagnosis did not significantly affect their well-being, reducing interest in a psycho-oncology app.

Reluctance toward new technologies:

- Some respondents expressed reluctance to use new technologies and a lack of interest in mobile apps.

Refusal to phone contact:

- Reasons included not having a private mobile phone or being unable to install non–work-related apps.

Other reasons:

- This category included various individual reasons, such as already receiving sufficient support through long-term psychotherapeutic treatment.

The process involved categorizing the responses into overarching categories based on the focus of the patients' concerns and reasons for refusal. The essential criterion for assignment to the various categories was the primary reason behind the refusal, as identified by 2 independently working psychologists. None of the responses suggested a refusal motivated by distrust or avoidance of research participation per se. This further supports the interpretation that these patients' refusals were app-specific rather than reflecting a general reluctance to engage in studies. This analysis allowed for the combination of basic categories into 5 overarching categories, providing a comprehensive understanding of the patients' perspectives and motivations for refusal: (1) Focus on Life Outside the Disease, (2) Focus on Disease and Treatment, (3) use of a Denial Mechanism in relation to content related to illness and treatment, (4) Technical Issues, and (5) Other Reasons. The process of categorizing responses with examples of patients' statements is provided in Table S1 in Multimedia Appendix 1.

The distress thermometer, a self-report measure of psychological distress, with an 11-point range from "no distress" (0) to "extreme distress" (10), resembling a thermometer [22,23], was used. Respondents were instructed to circle the number (0 - 10) that best describes their level of distress on the day of measurement. The validated Polish version was used; according to the validation results, a threshold of 5 points provides the greatest balance of sensitivity and specificity in selecting patients with severe distress.

The General Self-Efficacy Scale (GSES) by Schwarzer [24] was used to evaluate Self-efficacy beliefs. Self-efficacy is defined as the ability or level of confidence in performing

XSL•FO
RenderX

specific behaviors. The GSES consists of 10 questions, rated on a 4-point Likert scale ranging from 1 (“strongly disagree”) to 4 (“strongly agree”). The total score ranges from 10 to 40 points, with higher scores indicating a stronger sense of self-efficacy. Raw scores can be converted into standardized sten scores. The scale has demonstrated good reliability and measurement validity. The GSES has been translated into Polish by Juczyński [24].

The Ten-Item Personality Inventory (TIPI), a concise tool designed to evaluate the 5-factor personality model established by Costa and McCrae, was applied in the study [25]. TIPI was specifically developed to offer a brief (1 minute) assessment option in situations where using more comprehensive instruments would be unfeasible. The model encompasses Extraversion, Agreeableness, Conscientiousness, Emotional stability (or Neuroticism), and Openness to Experiences (10 items with five 2-item subscales). Respondents assigned a Likert score where 1 indicates “disagree strongly,” and 7 indicates “agree strongly,” with raw scores transformed into standardized sten scores. Correlations between the Polish version of TIPI and NEO-FFI confirm the convergent and discriminant validity of TIPI.

The Coping Orientation to Problems Experienced Questionnaire (Mini-COPE) was used to assess stress coping. The Mini-COPE, developed by Carver et al [26] and adapted into Polish by Juczyński and Ogińska-Bulik [27], comprises 28 items with 4 possible responses: “I haven’t been doing this at all” (0 points), “A little bit” (1 point), “A medium amount” (2 points), and “I’ve been doing this a lot” (3 points). Two items constitute one strategy, resulting in a total of 14 stress-coping strategies divided into 4 factors: Active Coping (comprising 3 strategies: Active Coping, Planning, and Positive Reinterpretation); Seeking Support (comprising 2 strategies: Seeking Emotional Support and Seeking Instrumental Support); Helplessness (comprising 3 strategies: Psychoactive Substance Use, Restraint, and Self-Blame); Avoidant Coping (comprising 3 strategies: Dealing with something else, Denial, and Venting). In addition, there are 3 strategies not belonging to any of the above-mentioned factors: Religion, Acceptance, and Humor. All strategies can be categorized into Active (effective) strategies and passive (ineffective) strategies. Active strategies include Active Coping, Seeking Support, Religion, Acceptance, and Humor, while passive strategies encompass Helplessness and Avoidant Coping. The questionnaire’s split-half reliability was 0.86, and the internal consistency for most scales was assessed as satisfactory.

Statistical Analysis

Categorical variables were summarized as frequencies and percentages. Continuous data were summarized as median values with IQR 25%-75%. For continuous variables, comparisons between more than 2 groups were performed by the Kruskal-Wallis H test and post hoc Dunn test with Benjamini-Hochberg *P* value correction. The Spearman correlation coefficient was assessed to examine correlations between variables. The classification of correlations used to evaluate the analyzed results was as follows: $0.0 \leq |r| < 0.1$ negligible correlation; $0.1 \leq |r| \leq 0.39$, weak correlation; $0.4 \leq |r| \leq 0.69$, moderate correlation; $0.7 \leq |r| \leq 0.89$, strong correlation; and $0.9 \leq r \leq 1$, very strong correlation [28]. Hierarchical clustering was based on Euclidean distance. Colors were scaled per column. A 2-sided $P<.05$ was considered statistically significant. Computations were performed in the R environment for statistical computing (version 4.0.1; R Foundation for Statistical Computing), released on June 6, 2020 [29].

Ethical Considerations

The study was approved by the Ethics Committee of the Maria Sklodowska-Curie National Research Institute of Oncology, Gliwice Branch (approval number KB/430-66/23). All procedures involving human participants were conducted in accordance with applicable ethical standards. Informed consent was obtained from all participants. Participant data were anonymized before analysis, and all personal information was handled confidentially in accordance with institutional data protection policies. No compensation was provided to participants for their involvement in the study.

Results

Participants

In the study group (N=56), all patients were diagnosed with primary early breast cancer (primary epithelial malignant tumors of the breast, at clinical stages I–III, eligible for treatment with curative intent, and with an Eastern Cooperative Oncology Group performance status of ≤2). The time since diagnosis ranged from 4 to 11 months, with most of the women undergoing preoperative systemic treatment (34/56, 60%). The age range was 32 - 67 years, with a mean age of 50 (SD 9) years. The majority of patients were married (40/56, 71%) and had at least one child (50/56, 89%). Demographic and clinical data are provided in Table 1.

XSL•FO
RenderX

Table . Sociodemographic and clinical characteristics of patients with primary breast cancer who refused to test the mobile psycho-oncology app within the clinical trial, Gliwice, Poland (2022 - 2023; N=56).

Characteristic	Value, n (%)	Value, mean (SD)	Value, median (IQR)
Age (years)	—[a]	50 (9)	48 (44-57)
Relationship status			
Married	40 (71)	—[a]	—[a]
Single	6 (11)	—[a]	—[a]
Divorced	7 (12)	—[a]	—[a]
Widowed	3 (5)	—[a]	—[a]
Children			
Yes	50 (89)	—[a]	—[a]
No	6 (11)	—[a]	—[a]
Number of children			
1	15 (27)	—[a]	—[a]
2	28 (50)	—[a]	—[a]
3	6 (11)	—[a]	—[a]
4	1 (2)	—[a]	—[a]
Average age of children	—[a]	21 (12)	23 (10-31)
Treatment stage			
During preoperative systemic treatment	35 (64)	—[a]	—[a]
After systemic therapy	20 (37)	—[a]	—[a]

[a]Not applicable.

Reasons for Refusal to Participate in Testing the App

A total of 138 women were offered participation in a clinical trial testing a psycho-oncology mobile app. Of these, 70 (51%) patients declined to use the app and were subsequently invited to participate in this observational study, and 56 of these 70 patients (40.1%, 56/138) agreed to participate in this study and provided open-ended responses regarding their reasons for refusing the intervention. As detailed in the “Methods” section, the responses were coded and grouped into categories. The most frequently given reason was lack of time due to numerous medical appointments and examinations related to cancer diagnosis and treatment (13/56, 23%). The second most frequent answer (8/56, 14%) was lack of mental need or mental well-being. Other reasons included poor physical and mental well-being (7/56, 13%), reluctance to recall content related to illness (7/56, 13%), lack of time due to difficult family situation (6/56, 11%), reluctance toward new technologies (5/56, 9%), lack of time due to caring for a young child (4/56, 7%), lack of time due to demanding professional work (3/56, 5%), and lack of private telephone (2/56, 4%). One patient (1/56, 2%) had been in psychotherapeutic contact for many years and found it sufficient support for psycho-oncological issues as well. An analysis of the answers given by 2 independent psychologists working independently of each other made it possible to combine the above 10 answers into 5 main categories, of which the most numerous was “Focus on Disease and Treatment” (20/56, 36%). Less common were the reasons associated with denial mechanisms in relation to content associated with illness and treatment (15/56, 27%), life outside the disease (13/56, 23%), Technical Issues (7/56, 13%), and Other Reasons (1/56, 2%) (Table 2).

XSL•FO
RenderX

Table . Reasons for refusal to use the mobile psycho-oncology app in the clinical trial among patients with primary breast cancer, Gliwice, Poland (2022 - 2023; N=56).

Overarching categories of reasons and basic reasons	Value, n (%)
Focus on life outside the disease	13 (23)
Lack of time—childcare responsibilities	4 (7)
Lack of time—difficult family situation	6 (11)
Lack of time—demanding job	3 (5)
Focus on the disease and treatment	20 (36)
Lack of time—frequency of medical appointments	13 (23)
Poor physical and mental well-being	7 (13)
Denial reaction	15 (27)
Reluctance to recall disease-related content	7 (13)
No mental need, feeling good	8 (14)
Technical issues	7 (13)
Reluctance toward new technologies	5 (9)
Refusal to phone contact	2 (4)
Other reasons	1 (2)

Psychological Outcomes

There were no missing values in the psychological assessments. The mean score in the patients' stress level measured with the distress thermometer was 5.0 (SD 2.1). The average level of distress in the whole group was moderate, close to the cut-off point according to Polish standards [23]. When analyzing the distribution of scores, we found that 63% of patients exceeded the cut-off point and 26% had a score of 7 or higher, indicating a significant escalation of distress. This means that at least a quarter of the patients in our study indicated a clinically significant escalation of distress. The mean score of generalized Self-efficacy measured with GSES was 32.1 (SD 5; stens: means 7.4, SD 1.9). Analysis of the 5 dimensions of personality measured using the TIPI questionnaire in patients refusing psycho-oncological eHealth support indicated the highest scores in the Conscientiousness dimension (mean 6.4, SD 0.9). The lowest score was obtained in the Neuroticism dimension (mean 3.4, SD 1.8). Mean scores for other personality traits were average: Extraversion (mean 5.8, SD 1.6), Agreeableness (mean 6.5, SD 0.8), and Openness to Experiences (mean 4.4, SD 1.5) (Table 3).

Table . Results of participants' stress level (distress thermometer [DT]), Generalized Self-Efficacy Scale (GSES), and personality trait characteristic (Ten-Item Personality Inventory [TIPI]) among patients with primary breast cancer who declined participation in the mobile psycho-oncology app clinical trial, Gliwice, Poland (2022 - 2023; N=56).

Questionnaire and variable	Value, mean (SD)	Value, median (IQR)
DT[a]		
Level of stress	5.0 (2.1)	5.0 (3.0-6.8)
TIPI[b]		
Extraversion	5.8 (1.6)	6.5 (5.4-7.0)
Agreeableness	6.5 (0.8)	6.8 (6.4-7.0)
Conscientiousness	6.4 (0.9)	7.0 (6.0-7.0)
Neuroticism	3.4 (1.8)	3.0 (2.0-4.5)
Openness to experience	4.4 (1.5)	4.5 (4.0-5.5)
GSES[c]		
Self-efficacy	32.1 (5.1)	32.5 (29.0-36.0)
Stens	7.4 (1.9)	7.5 (6.0-9.0)

[a]DT: distress thermometer.

[b]TIPI: Ten-Item Personality Inventory.

[c]GSES: General Self-Efficacy Scale.

XSL•FO
RenderX

The results in the Mini-COPE showed that patients most often used the strategies of Active Coping (mean 2.6, SD 0.5), Acceptance (mean 2.6, SD 0.6), and Seeking Emotional Support (mean 2.6, SD 0.6). Also relatively common were the strategies of Planning (mean 2.4, SD 0.6), Positive Reinterpretation (mean 2.3, SD 0.7), Seeking Instrumental Support (mean 2.2, SD 0.8), and Dealing with something else (mean 2.1, SD 0.8). The least used strategies were Psychoactive Substance Use (mean 0.2, SD 0.6), Restraint (mean 0.5, SD 0.7), and Sense of humor (mean 0.8, SD 0.7) (Table 4).

Table . Coping strategies (Coping Orientation to Problems Experienced Questionnaire [Mini-COPE]) among patients with primary breast cancer who declined to test the mobile psycho-oncology app within the clinical trial, Gliwice, Poland (2022 - 2023; N=56).

Type, dimension, and name of stress coping strategies	Value, mean (SD)	Value, median (IQR)
Active strategies mean 2.1 (SD 0.94); median 2.5 (IQR 1.5-3.0)		
Active coping	2.4 (0.65)	2.5 (2.0-3.0)
Active coping	2.6 (0.5)	3.0 (2.5-3.0)
Planning	2.4 (0.6)	2.5 (2.0-3.0)
Positive reinterpretation	2.3 (0.7)	2.5 (1.5-3.0)
Acceptance	2.6 (0.6)	3.0 (2.5-3.0)
Sense of humor	0.8 (0.7)	1.0 (0.0-1.5)
Turning to religion	1.6 (1.2)	2.0 (0.5-3.0)
Seeking support	2.4 (0.72)	2.5 (2.0-3.0)
Seeking emotional support	2.6 (0.6)	3.0 (2.0-3.0)
Seeking instrumental support	2.2 (0.8)	2.5 (1.5-3.0)
Passive strategies mean 1.1 (SD 1.0); median 1 (IQR 0.0-2.0)		
Avoidance	1.5 (0.93)	1.5 (1.0-2.0)
Dealing with something else	2.1 (0.8)	2.0 (1.5-3.0)
Denial	1.0 (0.9)	1.0 (0.0-1.5)
Venting	1.5 (0.8)	1.5 (1.0-2.0)
Helplessness	0.6 (0.86)	0.0 (0.0-1.0)
Psychoactive substance use	0.2 (0.6)	0.0 (0.0-0.0)
Restraint	0.5 (0.7)	0.0 (0.0-1.0)
Self-blame	1.1 (1.0)	1.0 (0.0-2.0)

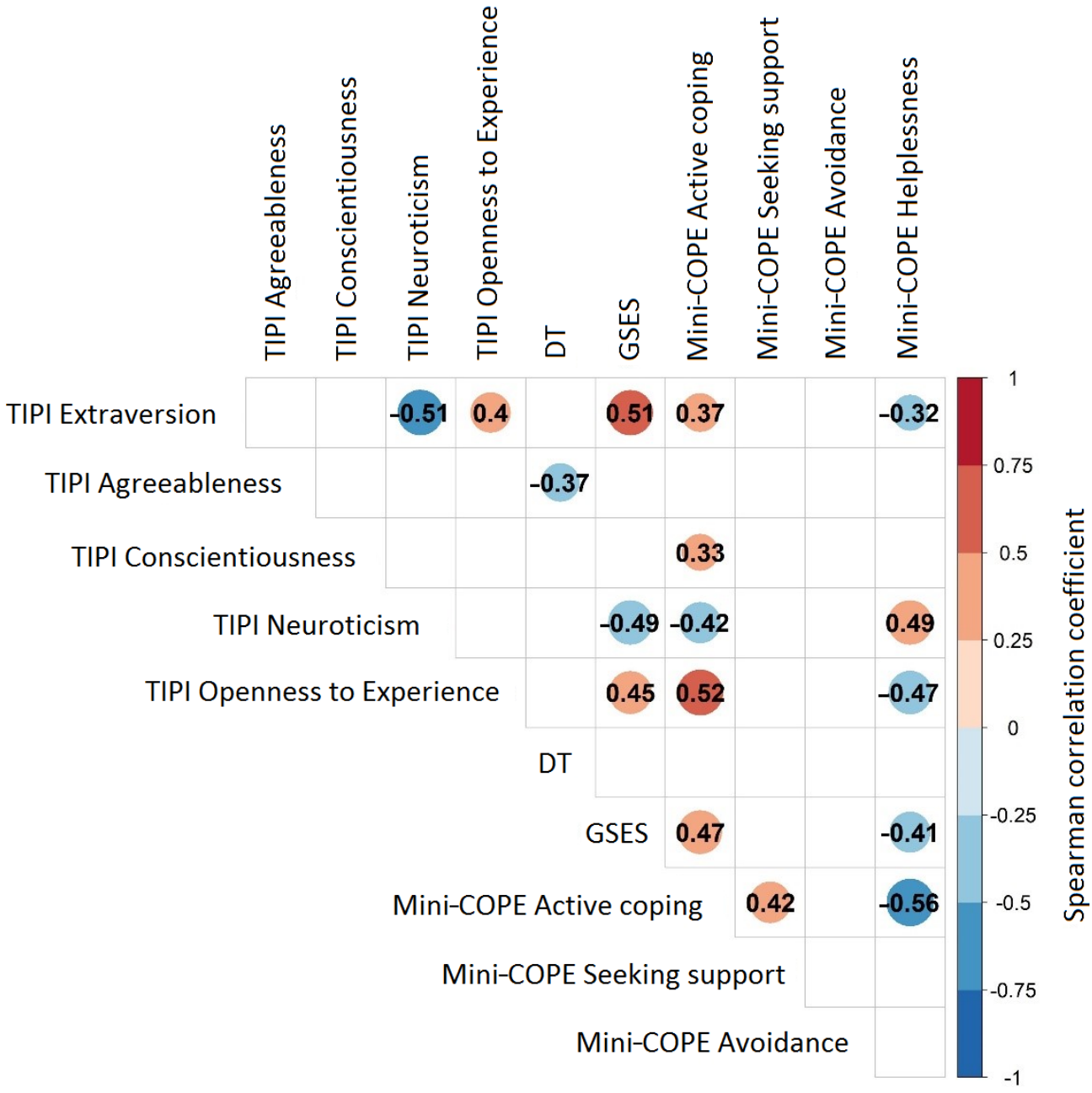
In the next step, correlations between individual variables of psychoemotional functioning were analyzed to identify possible personality profiles of patients in our study who refused to test the eHealth app. We observed moderate positive correlations between Extraversion and Self-Efficacy (r=0.51; P<.001), Helplessness and Neuroticism (r=0.49; P<.001), Self-Efficacy and Active Coping (r=0.47; P<.001), Openness to Experience and Self-Efficacy (r=0.45; P=.001), and Active Coping and Seeking Support (r=0.42; P=.003), Openness to Experience and Extraversion (r=0.4; P=.005), and Extraversion and Neuroticism (r=0.51; P<.001, Extraversion and Self-Efficacy (r=0.51; P<.001). We observed moderate negative correlations between Helplessness and Active Coping (r=−0.56; P<.001), Openness to experience and Helplessness (r=−0.47; P<.001), Neuroticism and Active Coping (r=−0.42; P=.003), and Helplessness and Self-efficacy (r=−0.41; P=.004) (Figure 2).

Heatmaps with hierarchical clustering did not identify subgroups with a specific personality-functioning profile, which indicates the homogeneity of the refusing group (Figure S1 in Multimedia Appendix 1). However, if we divided the group into 4 basic refusal categories (Focus on Life Outside the Disease, Focus on Disease and Treatment, Denial Mechanism, and Technical Issues) and analyzed each psychological test separately, we observed that patients giving different reasons for refusal differed in the severity of Neuroticism, use of positive reevaluation, and Self-efficacy. Analyses showed differences in Neuroticism between the Focus on Life Outside the Disease and Focus on the disease and treatment groups (P=.02), with the Focus on Life Outside the Disease group showing lower levels of Neuroticism. A higher level of Positive Reinterpretation was observed in the Focus on Life Outside the Disease group than in the Technical Issues group (P=.05) (Tables S2 and S3 in Multimedia Appendix 1). The results showed that, compared to the Focus on Life Outside the Disease group, patients in 2 groups had lower levels of Self-efficacy: the Focus on the Disease and Treatment (P=.03) and Denial Reaction group (P=.03) (Tables S2 and S4 in Multimedia Appendix 1). When we divided the group into 3 age groups: <45, 46-60, and >60 years, the only difference was in the use of denial strategies (P=.004). The 46 - 60 age group used this strategy more often than the ≥60 group (P=.04) (Tables S5 and S6 in Multimedia Appendix 1). There were no other differences in

XSL•FO
RenderX

psychoemotional functioning variables between the age categories.

Figure 2. Spearman correlation coefficient between stress level (distress thermometer [DT]), Generalized Self-Efficacy Scale (GSES), personality trait characteristics (Ten-Item Personality Inventory [TIPI]), and stress coping strategies (Coping Orientation to Problems Experienced Questionnaire [Mini-COPE]) among patients with primary breast cancer who declined to use the mobile psycho-oncology app within the clinical trial, Gliwice, Poland (2022-2023). A blank space indicates an insignificant correlation ($P \geq .05$).



Discussion

Principal Findings

In our study, we primarily aimed to delineate the motivations and their associations with the psychoemotional functioning of patients with primary breast cancer who consciously declined the use of eHealth technology in the form of a mobile psycho-oncology app as part of a clinical trial. In line with this objective, we examined the reasons behind the refusal to participate in the clinical trial involving the app, as expressed by the patients in an open-ended question, and evaluated their psychoemotional functioning, encompassing stress levels, self-efficacy, personality traits, and coping strategies. Importantly, all psychological assessments and reasons for refusal were collected from patients who declined the app intervention as part of a clinical trial but agreed to participate in our separate observational study, indicating that their decision was app-specific rather than a general reluctance toward research involvement. In addition, we categorized patients based on their reasons for declining participation in the study and compared the psychoemotional functioning of patients within these groups.

XSL•FO
RenderX

Refusals to Participate in a Clinical Trial Using an eHealth App

The percentage of refusals in our group is significantly higher compared to other studies [30,31]. However, there are also studies in which the refusal percentage was even higher than in our study [11]. It should be stressed that the team carrying out the clinical research on the use of the tested eHealth app did not offer participation to all consecutive patients; this could be confirmed by the upper limit of age in the tested group of patients (67 years), significantly lower than the upper age limit for patients treated by preoperative therapy in Poland. In the eligibility criteria of the trial, the skill to use the electronic tool was one of the major requirements; thus, the refusal rate is calculated in the group of patients apparently able to use it.

As mentioned previously in the Introduction section, there are limited studies that examine the reasons for rejecting the use of eHealth solutions. Gupta et al [20] listed 4 obstacles to recruiting patients for a study using a cell phone app: (1) coordination with clinic staff, (2) perceived burden among patients with breast cancer, (3) limitations on technology adoption and use, and (4) availability of resources among study staff. The reasons given by our patients corresponded to the second challenge domain identified by Gupta et al [20]. The remaining domains were less visible in our study, which may be due to a different way of constructing the refusal categories in our study compared to that of Gupta et al [20]. Among other things, when comparing these categories, we included "Unfamiliarity with mobile apps" as a "technical issue," whereas Gupta et al [20] classified it as a "perceived burden" (2nd category). Most of the reasons for refusal cited by our patients are also commonly mentioned in other studies as reasons for withdrawal from technology-based interventions [30,31]. The only exception was "reluctance to recall disease-related content," which we did not find in other studies. Meanwhile, patients reporting a reluctance to recall disease-related content are an interesting group from a clinical perspective, as they may be more likely to use a denial response, which is a risk factor for underestimating disease symptoms despite the available information and their intellectual capacity to comprehend it [32,33]. da Silva et al [34] demonstrated that patients with breast cancer made excuses to avoid check-ups, refrained from touching their breasts, and even avoided using the word "cancer." It is noteworthy that in our study, patients declaring reluctance to recall content related to illness as a reason for not using the app exhibited higher levels of Neuroticism.

It would be important to consider whether reluctance to use the app was derived from reluctance to participate in the clinical trial. However, none of the patients mentioned such a reason in the open-ended question. At the same time, these patients agreed to participate in the observational study presented in this paper, which was conducted after their refusal to test the mobile app, suggesting that it was not the mere fact of "participating in the study" that was a deterrent, but the topic itself, that is, testing the phone app.

Psychoemotional Functioning of Patients With Breast Cancer Who Declined the Use of eHealth Apps

Analysis of the results indicates that the level of stress among the patients does not appear to be the main reason for refusing to use the psycho-oncology app. Patients who declined participation may still have experienced moderate stress; however, this was not a decisive factor in their decision. This finding suggests that the refusal was more likely related to individual preferences, perceived usefulness of the app, or other psychomotor factors, rather than to low stress levels. It appears that patients who declined participation had a relatively high sense of Self-efficacy, consistent with previous Polish research [27]. It is plausible that the elevated Self-efficacy of our patients led them to forgo the use of the app, as they felt emotionally equipped to handle the challenges of illness. This conjecture aligns with Juczyński and Ogińska-Bulik's [27] assertion that heightened Self-efficacy influences an individual's positive assessment of personal resources in stressful circumstances. Moreover, the author contends that increased Self-efficacy is linked to engaging in health-related activities, a conclusion supported by other researchers [35,36]. However, the latter was contradicted by the results of our study, as our patients declined app usage despite possessing high Self-efficacy. This may be connected with the fact that their decision was more related to perceptions of the app itself than to a general disengagement from health-related behaviors or research. It is possible that they did not view app usage as a health-enhancing activity; this stresses the necessity of addressing this group both in the design of interventions and in the channels of their delivery at decisive moments.

In terms of coping with stress, patients most frequently used the strategies of Active Coping, Acceptance, and Seeking Emotional Support, and least frequently used the strategies of Psychoactive Substance Use and Restraint. This is a recurrent pattern of the most and least frequently used strategies in other Polish studies [37] by the authors of the test, which are considered to be highly effective in coping with stress.

Among personality dimensions, Neuroticism and Openness to Experience were relatively lower, which may help explain patients' reluctance to use the eHealth app. Openness to Experience is conceptualized as a willingness to accept novelty in life, including new ideas and concepts [25]. Thus, the lower Openness to Experience score of our patients may indicate a lack of readiness to accept the app as something new and unfamiliar. In addition, the low Neuroticism levels of our female patients may have been a protective factor against heightened negative emotions, thereby reducing motivation to seek psychological help. It should be noted, however, that the scores on these 2 personality dimensions, although the lowest, were within Polish normative values.

Psychoemotional Functioning in Relation to the Reasons for Refusal to Participate in the Study

Differences in the psychoemotional functioning of patients were observed based on the stated reason for refusal to participate in the study. Patients who cited focusing on life outside the disease as the reason for refusal exhibited lower levels of Neuroticism, more frequent use of positive reevaluation, and higher levels of

XSL•FO
RenderX

Self-efficacy than patients giving other reasons for declining the app intervention. These psychoemotional characteristics suggest better coping with the illness and treatment situation. Lower Neuroticism is associated with a reduced tendency to experience anxiety and sadness, greater emotional balance [25], and the ability to positively reinterpret events, leading to experiencing positive emotions despite the challenges [27]. High Self-efficacy is linked to greater self-confidence and more effective action [24]. This indicates that patients who felt too focused on their life outside the disease to participate in testing the psycho-oncology app were unlikely to require additional psychological support related to a cancer diagnosis. However, 3 other groups were identified: the "Denial Reaction" group had lower levels of Self-efficacy, the group citing technical difficulties as a reason for refusal was less likely to use the positive reappraisal technique, and the group focused on the disease and treatment had lower levels of Self-efficacy and higher levels of Neuroticism. Overall, the study group tended not to exhibit significantly elevated levels of stress and tended to have high levels of Self-efficacy. The aforementioned 3 groups of patients could potentially have needed additional psycho-oncological support that the app could have provided but did not make use of it.

Clinical Implications

Our research shows that when exploring the reasons behind patients' reluctance to use eHealth solutions, the refusing group should not be considered uniform. Our patients provided various reasons for their refusal to use a psycho-oncology mobile app, and their psychoemotional functioning differed depending on the reason. The aforementioned 3 groups of patients could have potentially benefited from using the psycho-oncological app, but the standard procedure of inviting them to use the app (direct contact by a nurse during routine pretreatment visit) did not effectively persuade them. When implementing eHealth solutions for routine psycho-oncological support of patients with breast cancer, it is essential to understand the reasons for patients' refusal to engage with such interventions, as this enables the use of more personalized incentives. Similar recommendations have been made for patients refusing to participate in clinical trials [38]. In addition, we recommend screening for psychoemotional functioning in patients with breast cancer before they are offered support in the form of an eHealth intervention. Such screening allows, first, the identification of patients who do not need additional support and, second, the presentation of screening results to patients as a personalized prompt (eg, indicating that the use of denial mechanisms or low self-efficacy potentially makes them more emotionally vulnerable during treatment, which they can effectively counteract with the support of the mobile app).

The most crucial issue is the problem of refusal rate in the context of eHealth interventions. Increasing the approval rate for novel tools is essential for understanding their population-level impact; possibly, the idea of carrying on the trial on both patients agreeing to use the tool and those refusing but who agreed to participate in this observational study would lead to the most reliable results of investigations. Nevertheless, we deliberately chose to focus this paper solely on the latter group. Our primary aim was to explore and categorize the reasons for refusing to use the eHealth app—a topic that, to our knowledge, has not yet been systematically addressed in the literature. Including comparative data at this stage might obscure the clarity of our categorization and shift the focus away from our core research objective. We acknowledge the importance of such a comparison and have collected data from users of the app; we intend to publish a separate article presenting a detailed comparison between users and nonusers. In addition, including these data here would significantly increase the length and complexity of this paper. We would like to draw attention again to the lack of standardized tools for investigating patients' reasons for refusing to use psychological eHealth tools [18]. The procedure we used to categorize patients' reasons for refusal was due to the lack of these tools. Importantly, the refusal we examined in this study pertained specifically to the use of the eHealth app, not to participation in clinical research more broadly. The results of our study, showing the different psychoemotional functioning of patients depending on the reason for refusal, indicate that this procedure was accurate and valid. We hope that this way of analyzing the reasons for patients' refusal to use psychological eHealth tools will be developed in further studies.

Moreover, our review of the literature revealed that the research tools we used are not commonly used. However, in our opinion, they are valuable and reliable as a standard data source of diagnostic information, especially considering the relatively short time required for patients to complete them. We therefore recommend these tools for psychological screening of patients with breast cancer.

Study Limitations

In our study, we used a small sample of a relatively homogeneous group of patients in a single breast cancer center. The generalizability to more diverse populations, patients in other geographic areas, or centers with different practices should be tested in the future. Poland is a country with a relatively homogeneous population of Caucasians, and in further studies, we will consider broader collaboration to increase diversity among participants.

In addition, although the intervention is briefly described as a psycho-oncology mobile app, we acknowledge the need for a more detailed presentation of its content, functionality, and interactivity. This information will be provided in a separate publication specifically dedicated to the intervention's development and implementation.

We also recognize that health care infrastructure and digital health literacy vary across countries. As such, the generalizability of our findings may be limited in settings with different eHealth accessibility or patient attitudes toward digital care. While our findings may not directly generalize to countries with highly advanced digital health systems, we believe they are especially relevant for regions where psychological eHealth solutions are still emerging and face similar adoption barriers.

We acknowledge the potential impact of self-selection bias. Our participants represent a specific subset of patients who refused to use the app but agreed to participate in the observational study. Their willingness to engage in research, despite refusing

XSL•FO
RenderX

the app intervention, may indicate characteristics that differentiate them from patients who declined both. This bias should be considered when interpreting the findings. A comparative analysis involving patients who used the app is currently in preparation and will be presented in a separate paper.

Despite these limitations, we believe that this study enriches our knowledge of the psychoemotional functioning of patients with breast cancer who refuse to use psychological eHealth tools in the form of a mobile app and their reasons for declining the app-based intervention, rather than research participation in general.

As a broader contextual note, the attitude of citizens in various countries to adopt novel solutions, especially in the technological domain, plays a key role. According to the latest Eurostat report [39], Poland is significantly below the European average in terms of technological expenditures and dissemination. However, numerous eServices are operational, including public health eServices, and the general enthusiasm for the use of various simple digital services (with ePayments at the forefront) is significant. For example, in 1 study designed by Finnish and Polish researchers, children were offered a mobile app intervention; Polish parents agreed in 92% of cases, while Finnish parents agreed in 56% [40], even though Finland rates significantly higher than Poland on the majority of Eurostat measures of technological enthusiasm. Seventy percent of Polish adults accept using web-based services for health-related issues [41], and this percentage is comparable to the level of Acceptance observed in our study. Thus, we consider our results potentially generalizable to a wider European population in countries with significant barriers to the adoption of psychological eHealth tools.

Conclusions

When incorporating eHealth solutions into the regular psycho-oncological support for patients with breast cancer, it is crucial to understand the reasons behind their refusal to use such solutions. In addition, it is important to consider these reasons in the context of the patients' psychoemotional well-being. This understanding allows for the implementation of more personalized incentives for those who could benefit from eHealth but are hesitant to use it. To achieve this, we recommend screening the patients' psychoemotional functioning.

Acknowledgments

The authors would like to thank the patients who generously agreed to participate in the observational study. We also thank the clinical staff at the Maria Sklodowska-Curie National Research Institute of Oncology, Gliwice Branch, for their assistance with data collection. ChatGPT (OpenAI) was used only for language refinement and formatting assistance. All content, study design, data interpretation, and conclusions were developed by the authors.

Funding

No external financial support or grants were received for this work.

Data Availability

Data from this study are available from the corresponding author upon reasonable request.

Authors' Contributions

MPS, ASB, and MJ contributed to the conceptualization of the study. Methodology was developed by MPS, ASB, and AK, with validation carried out by the same authors. Formal analysis was performed by AK. Investigation, resources, and data curation were managed by MPS and ASB. The original draft was written by MPS and ASB, and review and editing were conducted by MPS, ASB, AK, and MJ. Visualization was prepared by AK and MPS. Supervision was provided by MJ.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional tables and figure.
[DOCX File, 149 KB - mhealth_v14i1e71412_app1.docx]

References

1. EHealth. World Health Organization – Eastern Mediterranean Regional Office (WHO EMRO). URL: https://www.emro.who.int/health-topics/ehealth/ [accessed 2025-01-18]
2. Gyawali B, Bowman M, Sharpe I, Jalink M, Srivastava S, Wijeratne DT. A systematic review of eHealth technologies for breast cancer supportive care. Cancer Treat Rev 2023 Mar;114:102519. [doi: 10.1016/j.ctrv.2023.102519] [Medline: 36736125]

XSL•FO
RenderX

3. Lui JHL, Marcus DK, Barry CT. Evidence-based apps? A review of mental health mobile applications in a psychotherapy context. Prof Psychol Res Pr 2017;48(3):199-210. [doi: 10.1037/pro0000122]
4. Jongerius C, Russo S, Mazzocco K, Pravettoni G. Research-tested mobile apps for breast cancer care: systematic review. JMIR Mhealth Uhealth 2019 Feb 11;7(2):e10930. [doi: 10.2196/10930] [Medline: 30741644]
5. Koç Z, Kaplan E, Tanrıverdi D. The effectiveness of telehealth programs on the mental health of women with breast cancer: a systematic review. J Telemed Telecare 2024 Apr;30(3):405-419. [doi: 10.1177/1357633X211069663] [Medline: 35108131]
6. Garfan S, Alamoodi AH, Zaidan BB, et al. Telehealth utilization during the COVID-19 pandemic: a systematic review. Comput Biol Med 2021 Nov;138:104878. [doi: 10.1016/j.compbiomed.2021.104878] [Medline: 34592585]
7. Corbett T, Singh K, Payne L, et al. Understanding acceptability of and engagement with web-based interventions aiming to improve quality of life in cancer survivors: a synthesis of current research. Psychooncology 2018 Jan;27(1):22-33. [doi: 10.1002/pon.4566] [Medline: 29044832]
8. Seiler A, Klaas V, Tröster G, Fagundes CP. eHealth and mHealth interventions in the treatment of fatigued cancer survivors: a systematic review and meta-analysis. Psychooncology 2017 Sep;26(9):1239-1253. [doi: 10.1002/pon.4489] [Medline: 28665554]
9. Hernandez Silva E, Lawler S, Langbecker D. The effectiveness of mHealth for self-management in improving pain, psychological distress, fatigue, and sleep in cancer survivors: a systematic review. J Cancer Surviv 2019 Feb;13(1):97-107. [doi: 10.1007/s11764-018-0730-8] [Medline: 30635865]
10. Mitchell HR, Smith SK, Gebert R, Applebaum AJ. Mobile cognitive behavioral therapy for posttraumatic stress: diving back in after hematopoietic stem cell transplant. Psychooncology 2022 Oct;31(10):1802-1805. [doi: 10.1002/pon.6022] [Medline: 36050923]
11. Romero-Ayuso D, García-López R, Lozano-Villena C, et al. Usability of a mobile phone application to enhance activities of daily living in occupational therapy services for breast cancer survivors. Hong Kong J Occup Ther 2023 Dec;36(2):128-140. [doi: 10.1177/15691861231206489] [Medline: 38027046]
12. Badaghi N, van Kruijsbergen M, Prins J, et al. Effect of blended and unguided online delivery of mindfulness-based cognitive therapy versus care as usual on distress among cancer patients and survivors: protocol for the three-arm parallel randomized controlled buddy trial. BMC Psychol 2023 Jan 25;11(1):21. [doi: 10.1186/s40359-023-01052-2] [Medline: 36698197]
13. Lengacher CA, Reich RR, Ramesar S, et al. Feasibility of the mobile mindfulness-based stress reduction for breast cancer (mMBSR(BC)) program for symptom improvement among breast cancer survivors. Psychooncology 2018 Feb;27(2):524-531. [doi: 10.1002/pon.4491] [Medline: 28665541]
14. Mumtaz A. Prioritizing and overcoming barriers to e-Health use among elderly people: implementation of the analytical hierarchical process (AHP). J Healthc Eng 2022;2022:7852806. [doi: 10.1155/2022/7852806] [Medline: 35449830]
15. Fleddermann K, Molfenter T, Jacobson N, et al. Clinician perspectives on barriers and facilitators to implementing e-Health technology in substance use disorder (SUD) treatment facilities. Subst Abuse 2021;15:11782218211053360. [doi: 10.1177/11782218211053360] [Medline: 34720585]
16. Tyrrell JS, Redshaw CH. Physical activity in ankylosing spondylitis: evaluation and analysis of an eHealth tool. J Innov Health Inform 2016 Jul 4;23(2):169. [doi: 10.14236/jhi.v23i2.169] [Medline: 27869581]
17. Horn A, Jírů-Hillmann S, Widmann J, et al. Systematic review on the effectiveness of mobile health applications on mental health of breast cancer survivors. J Cancer Surviv 2025 Feb;19(1):1-17. [doi: 10.1007/s11764-023-01470-6] [Medline: 37906420]
18. Jacob C, Lindeque J, Klein A, Ivory C, Heuss S, Peter MK. Assessing the quality and impact of eHealth Tools: systematic literature review and narrative synthesis. JMIR Hum Factors 2023 Mar 23;10:e45143. [doi: 10.2196/45143] [Medline: 36843321]
19. Cho Y, Lavoie Smith EM, Zahrieh D, et al. Electronic patient-reported outcome data collection systems in oncology clinical trials: a survey of clinical research professionals (an Alliance study). JCO Clin Cancer Inform 2023 Sep;7:e2300007. [doi: 10.1200/CCI.23.00007] [Medline: 37677111]
20. Gupta A, Ocker G, Chow PI. Recruiting breast cancer patients for mHealth research: obstacles to clinic-based recruitment for a mobile phone app intervention study. Clin Trials 2020 Dec;17(6):675-683. [doi: 10.1177/1740774520939247] [Medline: 32660354]
21. Heinrich R, Schilling G, Wojtyna E, et al. Effects of mobile application-based cognitive behavioral therapy on psychological outcomes in women treated for breast cancer: a randomized controlled pilot trial in Germany. Psychooncology 2024 Oct;33(10):e70003. [doi: 10.1002/pon.70003] [Medline: 39439014]
22. National Comprehensive Cancer Network. Distress management. Clinical practice guidelines. J Natl Compr Canc Netw 2003 Jul;1(3):344-374. [doi: 10.6004/jnccn.2003.0031] [Medline: 19761069]
23. Życińska J, Wojtyna E, Heyda A, Syska-Bielak A. Polska adaptacja skali distress thermometer. Badania walidacyjne i normalizacyjne wśród pacjentów onkologicznych [Article in Polish]. Presented at: XXXIII Scientific Meeting of the Polish Psychological Association; Sep 11-12, 2008.
24. Juczyński Z. Narzędzia Pomiaru w Promocji i Psychologii Zdrowia [Book in Polish]: Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego; 2012.

XSL•FO
RenderX

25. Thørrisen MM, Sadeghi T. The Ten-Item Personality Inventory (TIPI): a scoping review of versions, translations and psychometric properties. Front Psychol 2023;14:1202953. [doi: 10.3389/fpsyg.2023.1202953] [Medline: 37434881]
26. Carver CS, Scheier MF, Weintraub JK. Assessing coping strategies: a theoretically based approach. J Pers Soc Psychol 1989 Feb;56(2):267-283. [doi: 10.1037//0022-3514.56.2.267] [Medline: 2926629]
27. Juczyński Z, Ogińska-Bulik N. Narzędzia Pomiaru Stresu i Radzenia Sobie Ze Stresem [Book in Polish]: Warszawa: Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego; 2009.
28. Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. Anesth Analg 2018 May;126(5):1763-1768. [doi: 10.1213/ANE.0000000000002864] [Medline: 29481436]
29. The R Project for Statistical Computing. URL: https://www.r-project.org/ [accessed 2025-01-18]
30. Rossen S, Kayser L, Vibe-Petersen J, Ried-Larsen M, Christensen JF. Technology in exercise-based cancer rehabilitation: a cross-sectional study of receptiveness and readiness for e-Health utilization in Danish cancer rehabilitation. Acta Oncol 2019 May;58(5):610-618. [doi: 10.1080/0284186X.2018.1562213] [Medline: 30698060]
31. Atema V, van Leeuwen M, Kieffer JM, et al. Efficacy of internet-based cognitive behavioral therapy for treatment-induced menopausal symptoms in breast cancer survivors: results of a randomized controlled trial. J Clin Oncol 2019 Apr 1;37(10):809-822. [doi: 10.1200/JCO.18.00655] [Medline: 30763176]
32. McWilliams N. Psychoanalytic Diagnosis: Understanding Personality Structure in the Clinical Process, 2nd edition: Guilford Press; 2011.
33. Kaminska-Winciorek G, Cybulska-Stopa B, Gajda M, Pawełczak-Szastok M, Gęga-Czarnota A. Neglected malignant neoplasms with cutaneous involvement. Postepy Dermatol Alergol 2021 Oct;38(5):916-920. [doi: 10.5114/ada.2021.108457] [Medline: 34849146]
34. Silva RD, Pinheiro CPO, Bezerra IC, Caldas JMP, Brasil CCP. Mastectomy and the meaning attributed by patients to delay in early diagnosis of breast cancer. Asian Pac J Cancer Prev 2018 Aug 24;19(8):2083-2088. [doi: 10.22034/APJCP.2018.19.8.2083] [Medline: 30139205]
35. Gu Z, Li M, Liu L, Ban Y, Wu H. The moderating effect of self-efficacy between social constraints, social isolation, family environment, and depressive symptoms among breast cancer patients in China: a cross-sectional study. Support Care Cancer 2023 Sep 28;31(10):594. [doi: 10.1007/s00520-023-08063-0] [Medline: 37768382]
36. Oswald LB, Baik SH, Buscemi J, et al. Effects of smartphone interventions on cancer knowledge and coping among Latina breast cancer survivors: secondary analysis of a pilot randomized controlled trial. J Psychosoc Oncol 2022;40(6):695-707. [doi: 10.1080/07347332.2021.1983688] [Medline: 34620035]
37. Żołnierz J, Sak J. Coping and religiosity of Polish breast cancer patients. Religions (Basel) 2023;14(5):682. [doi: 10.3390/rel14050682]
38. Miller SM, Hudson SV, Egleston BL, et al. The relationships among knowledge, self-efficacy, preparedness, decisional conflict, and decisions to participate in a cancer clinical trial. Psychooncology 2013 Mar;22(3):481-489. [doi: 10.1002/pon.3043] [Medline: 22331643]
39. Digitalisation in Europe – 2024 edition. Eurostat (European Commission). URL: https://ec.europa.eu/eurostat/web/interactive-publications/digitalisation-2024 [accessed 2025-01-18]
40. Vepsäläinen H, Skaffari E, Wojtkowska K, et al. A mobile app to increase fruit and vegetable acceptance among Finnish and Polish preschoolers: randomized trial. JMIR Mhealth Uhealth 2022 Jan 4;10(1):e30352. [doi: 10.2196/30352] [Medline: 34982718]
41. Płaciszewski K, Wierzba W, Ostrowski J, Pinkas J, Jankowski M. Use of the internet for health purposes-a national web-based cross-sectional survey among adults in Poland. Int J Environ Res Public Health 2022 Dec 6;19(23):16315. [doi: 10.3390/ijerph192316315] [Medline: 36498389]

Abbreviations

GSES: General Self-Efficacy Scale
Mini-COPE: Coping Orientation to Problems Experienced Questionnaire
TIPI: Ten-Item Personality Inventory

Edited by A Stone; submitted 18.Jan.2025; peer-reviewed by O Afolayan, Y Lee; revised version received 14.Dec.2025; accepted 15.Dec.2025; published 13.Feb.2026.

Please cite as:
Pawełczak-Szastok M, Syska-Bielak A, Krzywon A, Jarząb M
Reluctance to Use a Psycho-Oncology Mobile App Among Patients With Primary Breast Cancer: Retrospective Cross-Sectional Survey
JMIR Mhealth Uhealth 2026;14:e71412
URL: https://mhealth.jmir.org/2026/1/e71412
doi:10.2196/71412

XSL•FO
RenderX

© Marta Pawełczak-Szastok, Anna Syska-Bielak, Aleksandra Krzywon, Michał Jarząb. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 13.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Features of Mobile Health Apps for Tobacco Cessation That Appeal to Black Adults Who Use Tobacco Products: Focus Group Study

Sonia A Clark[1,2], MHA; Remi Philips[2], BS; Christine E Kistler[3], MD, MASc; Adam O Goldstein[1,2], MD, MPH; Chineme Enyioha[1,2], MD, MPH

[1]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[2]Department of Family Medicine, University of North Carolina at Chapel Hill, 590 Manning Dr, Chapel Hill, NC, United States

[3]Division of Geriatrics, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

Corresponding Author:
Chineme Enyioha, MD, MPH
Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Abstract

Background: Mobile health (mHealth) interventions show promise in supporting tobacco cessation. However, Black adults who use tobacco products are not well represented in mHealth studies for tobacco cessation, and their preferred features of mHealth apps are not well known. Identifying types of mHealth app features for tobacco cessation preferred by Black adults is critical to developing a culturally adapted app, with increased uptake by the target population.

Objective: The goal of this study was to explore culturally relevant preferences for features of smoking cessation mHealth apps among Black adults who use tobacco products.

Methods: A comprehensive list of features of mHealth apps for tobacco cessation was developed based on previous research and a review of existing mHealth literature. Through a content analysis, this list was divided into subgroups and used to develop a focus group guide. We recruited participants from Instagram, a social media platform. Eligible focus group participants included people who reported current use of a tobacco product, identified as being African American or Black, were 21 years old or older, and had access to Wi-Fi or the internet. Participants had to indicate interest in the use of an mHealth app for tobacco cessation. Participants discussed their opinions about different app features, including what features they felt would increase the use of an app by Black adults. Recordings from the focus groups were transcribed and coded deductively and inductively. We conducted a thematic content analysis of the resulting transcripts.

Results: Forty adults aged 21 - 69 (mean 43, SD 13.6) years participated in 8 focus groups. Fifty-seven percent were female, and 88% endorsed current cigarette use. Four central themes that represented app features emerged. (1) Participants wanted representation and inclusivity through personalization and featuring people with similar lived experiences, including representative images and relevant health information. (2) Participants desired the app to feature a diversity of experiences such as testimonials from individuals from different backgrounds rather than solely focusing on racial identity or excessive targeting of the Black community. (3) Participants desired accountability through trusted connections with health care professionals and other support groups within the app, as well as app tracking capability. (4) Encouragement and motivation were more salient incentives than monetary rewards.

Conclusions: Black adults who use tobacco products prefer a tobacco cessation app with features that are inclusive, relatable, supportive, and motivating. These findings can serve as the groundwork for the development of an mHealth app that will appeal to Black adults, potentially leading to increased app use, successful cessation, and health equity.

(JMIR Mhealth Uhealth 2026;14:e63340) doi:10.2196/63340

KEYWORDS

mHealth; tobacco cessation; preferred features; mobile health app; representation; inclusivity; African American; Black adult tobacco users; mobile health

Introduction

In the United States, certain racial minorities, including Black adults, face a disproportionate burden of negative health effects from tobacco use [1]. Compared to other racial and ethnic groups, Black adults who use tobacco products are more likely to experience adverse health issues because of their tobacco use, such as coronary heart disease, and have a higher morbidity

XSL•FO
RenderX

and mortality rate from other tobacco-related health conditions [2,3]. For instance, compared with Whites, Black adults have a higher mortality rate from cancer (221 vs 186 per 100,000 population) [4]. Despite starting to smoke at an older age [5] and being more likely to make a quit attempt than any other racial group, Black adults who use tobacco products are less likely to successfully quit than other racial groups [6]. Black adults also have lower odds of both quitting and using Food and Drug Administration cessation medications, despite having greater odds of using behavioral counseling/self-help materials [7].

Novel interventions, such as mobile health (mHealth) interventions, show promise in supporting tobacco cessation for Black adults. mHealth apps have previously been successfully used with Black adults to help manage conditions such as diabetes and obesity [8,9]. Compared to White adults, Black adults are more likely to use apps to inform their medical decisions [10]. mHealth apps can also increase access for Black adults who use tobacco products by reducing barriers to receiving care, like cost and transportation logistics [11,12]. One study examining willingness to participate in mHealth app research for weight management among Black women found high receptivity, frequent smartphone usage, and existing use of smartphones to access information on health and wellness, suggesting mHealth apps could be used to empower and promote participation [13].

Although a promising route to increase tobacco cessation, mHealth apps should be culturally adapted to address unique factors that impact Black adults who use tobacco products. Black adults use menthol products and smoke cigars at higher rates [14], smoke fewer cigarettes per day [15], and may be more likely to use tobacco products as a result of stress related to racial harassment [16]. One study examining differences in addiction and quitting experiences concluded that these differences may suggest that Black adults who use tobacco products experience a different “quitting process” [15] as compared to White smokers. However, Black adults who use tobacco have not been well represented in prior mHealth studies for tobacco cessation, and their preferred features of mHealth apps are not well known. In addition, research has shown that tailoring interventions in ways that incorporate values and cultural experiences for patients, especially patients from minority backgrounds, can increase engagement with the intervention [17]. For example, visual representation of an individual’s culture in an intervention, delivery or promotion of an intervention by an individual from the target population, or the use of language that the target audience can relate to can enhance the appeal of that intervention [17,18]. A critical step to developing a tobacco cessation mHealth app that Black adults are receptive to and are likely to use is to identify desirable app features.

Furthermore, the extent to which preferred features satisfy key components of the self-determination theory is also important. According to the self-determination theory, autonomy, competence, and relatedness are key factors necessary for intrinsic motivation as well as maintenance of behavior change [19]. Autonomy, described as the intrinsic need to regulate one’s own behavior without the influence of external factors, has been found to be effective in tobacco cessation [20]. Competence, also described as the need for an individual to feel confident and capable as it relates to behavior change, can be supported by app features such as an app’s ability to provide information about tobacco cessation or provision of information about a user’s progress or behavior [21]. Relatedness is described as connecting with others and can be achieved by an app that allows for social support or interaction with others [21]. Each component of the self-determination theory plays an important role in cessation-related behavior [20,22]. The goal of this study was to identify features of mHealth apps for tobacco cessation that Black adults who use tobacco products would like to see in an app and what features will make the app more inclusive [19-22].

Methods

Study Design

Between June and July of 2023, we conducted 8 focus groups with 40 Black adults who use tobacco products. Participants were recruited using the social media platform Instagram. After completing a brief screener and obtaining consent during a brief screening call, eligible participants were purposively invited to participate in focus groups. Focus groups were stratified by age groups (21 - 34, 35 - 54, and ≥55 years old) to optimize representation.

Participants

Eligible participants included Black adults who were 21 years old or older, who reported current use of one or more tobacco products in the past 30 days, spoke English, and were current US residents at the time of recruitment. Exclusion criteria included lack of access to the internet and lack of interest in the use of technology for smoking cessation.

Focus Group Guide Development

To identify which features to discuss during focus groups, we developed a comprehensive list of features of mHealth apps for tobacco cessation using previous research and a review of existing mHealth literature [23-31]. The mHealth apps involved were designed based on behavior change techniques and theories, such as the principles of acceptance and commitment therapy [24,26], persuasive technology for behavior change [27], or followed the US clinical practice guidelines [28,31]. One of the apps was developed by the National Cancer Institute [28], and another was developed to target young adults who are motivated to quit [27]. We then divided this list into different domains to explore with participants. Initial mHealth app feature domains included content, user experience, personalization, tracking, privacy and security, connections or support, and inclusivity. Within each domain, several features were included to discuss with participants, with the primary focus being whether these features were inclusive and preferred by Black adults who use tobacco products. Using guidance from experts, the resulting domains and features were organized and distilled. The list and resulting domains were further refined through focus groups using the focus group guide.

XSL•FO
RenderX

Focus Group Procedures

Using the list of domains and features of mHealth apps, the research team developed the focus group guide to address the research question: What features in mHealth apps for tobacco cessation are preferred by Black adults? The interview guide included questions and probes about participants' experience with tobacco use and cessation, and experience with mHealth apps. Next, the guide listed each of the domains and features with a corresponding question about what features they preferred and a final question to get feedback about other design features or advice participants had. The semistructured focus groups were conducted using Zoom videoconferencing software to allow for participation from different locations within the United States, and each focus group lasted an hour [32].

Analysis

The resulting focus group recordings were professionally transcribed, and transcripts were uploaded into Atlas.ti, a qualitative data analysis software (version 23). Using the focus group guide, an initial codebook was developed. The qualitative analysis team comprised 3 researchers (CE, SAC, and RP), who each had prior experience with qualitative coding and received formal qualitative and mixed methods training at the University of North Carolina at Chapel Hill. After familiarizing themselves with the codebook, 2 researchers (SAC and RP) independently coded a randomly selected transcript with the guidance of the senior researcher (CE). To establish the rigor of the analysis, we applied multiple methods to strengthen the credibility, dependability, and confirmability. First, we documented codebook revisions and applied analytic triangulation with multiple coders who reviewed transcripts independently and participated in consensus meetings to review discrepancies, add new codes, and refine definitions as needed. To assess interrater reliability, Krippendorff α was used, and the initial interrater reliability was calculated to be 0.769, indicating substantial agreement. After resolving coding conflicts, a different transcript was then coded, and "near perfect" agreement, 0.9, was reached. Upon reaching this threshold of high agreement, the team recoded the initial transcript and coded the remaining transcripts with the finalized codebook.

Furthermore, the focus groups were conducted with cultural sensitivity. For instance, the main facilitator of the focus groups was a member of the research team and is also a part of the Black community, as well as familiar with participants' cultural values [32]. We did not conduct member check-in to reduce the burden on participants and concerns for confidentiality. However, the data were presented to the larger research team, and findings were discussed in detail [33].

Using thematic content analysis, the research team discussed and identified recurring themes and concepts across the transcripts [34]. The team determined that thematic saturation was achieved after the completion of 8 focus groups [35].

Ethical Considerations

This study was approved by the institutional review board at the University of North Carolina at Chapel Hill (24 - 0798). All study participants were informed about the study aims and significance, as well as the methods. Participants gave their informed consent, retained the rights to withdraw from the study at any time, and were informed of the confidentiality and anonymity of the data and the goal to publish findings from the study. All participants received a US $50 gift card for their participation in the study.

Results

Sociodemographic and Tobacco Use Data

Forty adults aged 21-69 years, with an average age of 43 (SD 13.6) years, participated in 8 focus groups (Table 1). Most participants were female, and the majority were non-Hispanic. Many participants endorsed current cigarette use. Fifty-five percent (n=22) reported use of electronic cigarettes or other vaping devices, while 40% (n=16) endorsed use of cigars.

XSL•FO
RenderX

Table . Demographics and tobacco use characteristics of participants (N=40).

Characteristics	Value
Age (years), mean (SD)	43 (13.6)
Sex (female), n (%)	23 (57)
Race, n (%)	
Black/African American	36 (90)
Black/African American and some other race, ethnicity, or origin	4 (10)
Ethnicity (non-Hispanic), n (%)	37 (92)
Sexual orientation, n (%)	
Straight	33 (82)
Gay or lesbian	2 (5)
Bisexual	5 (13)
Current cigarette use, n (%)	
Yes	35 (88)
Current other tobacco product use, n (%)	
Yes	31 (78)
Electronic cigarettes, e-cigarettes, or other vaping devices	22 (55)
Cigars, for example, little cigars, cigarillos, or large cigars	16 (40)
Smokeless tobacco, for example, chewing tobacco, snuff, or snus	6 (15)
Water-pipe tobacco or hookah	5 (13)
Other types of oral nicotine products, such as Velo lozenges, Rogue tablets, Lucy gum, or Pixotine toothpicks	1 (3)
Other	3 (8)
None of the above	9 (23)

Four central themes emerged about what mHealth app features were desirable by Black adults who use tobacco products. Each theme had 2 - 3 subthemes that represented different domains of mHealth app features. See Table 2.

Table . Themes and subthemes with domains of mobile health app features represented.

Themes	Subthemes	Domain of app feature
Participants wanted representation and inclusivity through personalization and featuring people with similar lived experiences	• Inclusion of images to represent the target population • Personalization of the app • Inclusion of content about relevant health risks	• Representation and inclusivity • User experience • Content and inclusivity
Excessive targeting of the Black community was unappealing, and participants desired the app to feature a diversity of experiences rather than solely focusing on racial identity	• Inclusion of testimonials from people with different backgrounds • Outreach in the Black community	• Content • Social support
Participants desired accountability through trusted connections, and app tracking capability	• Connection with health professionals • Supportive anonymous community within the app • App tracking capability	• Professional support • Social or peer support • Tracking/user experience
Encouragement and motivation were more salient incentives than monetary rewards	• Nonmonetary rewards • Provision for diverse reward options	• Gamification and rewards

XSL•FO
RenderX

Theme 1: Participants Wanted Representation and Inclusivity Through Personalization and Featuring People With Similar Lived Experiences

Content and Inclusivity: Inclusion of Images to Represent the Target Population

When asked how to make the app more inclusive of Black adults, several participants emphasized the need to create an app that features Black adults.

> *If we get on the app and we see a lot of us, then that will make us wanna use it more.* [Focus group 1, Interviewee 1]

> *You gonna get more people who's like, "Hey, that applies to me. First of all, they look like me. They talk like me...They kinda think like me."* [Focus group 3, Interviewee 6]

One person lamented that it was not common to see themselves represented in existing apps.

> *You know, like I need to see people of color, you know..., connecting my eyes because it's not [in] a lot of apps out here...* [Focus group 6, Interviewee 2]

Testimonials from fellow Black individuals were seen as a way to make people want to use the app:

> *...having testimonials of other African-Americans, 'cause that would like make me at least feel more comfortable to even wanna try the app. Just seeing that other people that look like me have used it and have something positive to say.* [Focus group 4, Interviewee 3]

User Experience: Personalization of the App

Some believed that inclusivity could be achieved through personalization of tobacco use and cessation goals.

> *...[to] feel like you [are] talking to somebody that you know...maybe you could pick up from the personalized questions in beginning.* [Focus group 6, Interviewee 2]

Participants also believed personalization could be achieved through in-app customization, which can allow people to have their goals at the forefront.

> *...you should be able to customize [the app], you know, or personalize your experience with certain personal goals and..., based on your smoking habits...* [Focus group 4, Interviewee 2]

> *..Then this goal is set to be like the basis of the notification and the recommendations you receive.* [Focus group 4, Interviewee 2]

Content and Inclusivity: Inclusion of Content About Relevant Health Risks

The inclusion of content about their health risks as Black adults who use tobacco products was also suggested:

> *Include like the health aspect of it, like what smoking contributes to, like, just to add the statistics like, you know, African Americans,...make sure that you mention more of how it affects us as-as a culture.* [Focus group 7, Interviewee 5]

With personalized content, participants hope to find stories to relate to.

> *Personalize it because everyone has their own story. So, if you can find the testimonials that can represent you, that would work.* [Focus group 1, Interviewee 2]

Some indicated that having personalized content would be preferable because they have different experiences and different stressors.

> *...involve Black people [36], we have shared experiences that only [we] can understand...Like a Black mother or Black woman—if I need help, she could like advise me and pull me through.* [Focus group 4, Interviewee 4]

> *I think people that use cigarettes as a coping mechanism...I have—for example, stressors in the neighborhood...* [Focus group 3, Interviewee 2]

Theme 2: Excessive Targeting of the Black Community Was Unappealing, and Participants Desired the App to Feature a Diversity of Experiences Rather Than Solely Focusing on Racial Identity

Content: Inclusion of Testimonials From People With Different Backgrounds

While many participants voiced that the app should offer representation and inclusivity where Black adults are represented, many were adamant that the app should not focus on Black adults solely.

> *I don't wanna feel like it's directed right at me 'cause I'm Black. 'Cause I'm more than just Black.* [Focus group 2, Interviewee 2]

Rather, participants voiced that they wanted an inclusive app open to everyone struggling with smoking.

> *It should be point to everyone with smokes, despite religion, despite your-your race, despite whoever you are.* [Focus group 1, Interviewee 3]

> *...provide practical solutions accessible to users from various income levels. You know, regardless of-regardless of whether you Black or not.* [Focus group 4, Interviewee 2]

Participants suggested that one way to bring diversity into the app is to include testimonials from people from different backgrounds.

> *I would like to hear,...different people's testimonials...because we all have different stories, you know, and different reasons for doing the things that we do.* [Focus group 1, Interviewee 1]

Having the app feature real stories was viewed as a way to motivate app use:

> *...if you share like a real story of somebody that, you know, suffered bad consequences for smoking for not*

XSL•FO
RenderX

> *so long, that will probably motivate me to use the app.* [Focus group 7, Interviewee 7]

Social Support: Outreach in the Black Community

Finally, some participants suggested conducting outreach at events within the Black community to promote the app and collaborate with respected figures:

> *...have it displayed in a way that's attractive to us, and then also present it in the areas that we are...events that we go to.* [Focus group 6, Interviewee 4]

> *You need to get...somebody famous. I'mma just use Diana Ross just for name sake, right? Say she used to smoke. You get her to talk about it. We all know, in the African American community, word of mouth is the most powerful thing we have.* [Focus group 3, Interviewee 6]

Some participants suggested social media marketing:

> *...if I have like a good TikTok video of someone that's like a Black that's...reviewing or giving me something to pull me in, that's definitely gonna make me go and download it.* [Focus group 6, Interviewee 4]

Another participant reflected on their experience of trying an app because it was recommended by an influencer:

> *I actually got to know about it by...[a] Black female YouTuber and that's how I got interested...*
>
> *[Focus group 4, Interviewee 4]*

Theme 3: Participants Desired Accountability Through Trusted Connections and App Tracking Capability

Professional Support: Connection With Health Professionals

When asked whether the app should offer connections to other app users, doctors, or other health care providers, participants expressed their desire to connect in order to promote accountability. Some participants indicated talking to professionals would help keep them accountable to their goals, either by providing advice, "that would seem pretty nice to be able to...have a coach that, you know, that you can go to for advice or assistance..." [Focus group 1, Interviewee 1] or by monitoring their progress, "I think you should have some kind of feature that to share your usage with your physician and stuff like that or your coach" [Focus group 4, Interviewee 2].

Social Support: Supportive Anonymous Community Within the App

When asked about connections with other users of the app, most participants indicated interest in seeking supportive accountability within the app:

> *I think it will be better to have someone that, you know—'cause my friend doesn't smoke, so someone that understands. But to be able to hold you accountable.* [Focus group 1, Interviewee 1]

Participants expressed a desire to decrease loneliness by connecting with others who understand what they are going through:

> *I find that quitting smoking is like the loneliest thing I've ever dealt with in my life...[If] there's like a little message board type of thing inside the app where people can just write in,...what they're going through.* [Focus group 1, Interviewee 2]

Most participants preferred the community within the app to be anonymous and were interested in anonymous interactions through groups, games, and a leaderboard, "to kind of promote data security...the information of the user should be de-identified or...be on the app as anonymous user" [Focus group 4, Interviewee 2] and "...in terms of being able to see what might be called like a leaderboard or a scoreboard...it can kind of make it...a bit more interactive" [Focus group 7, Interviewee 3].

Although most participants wanted a quit buddy within the app, they did not want this quit buddy to be a family member or friend. A few people noted that friends and family can be triggers for smoking and they are not understanding, "because sometimes family and friends can be triggers to smoking" [Focus group 8, Interviewee 3] and "sometimes like she said, you can't really talk to family or friends about it, 'cause they don't—they don't do it 'cause they don't understand it..." [Focus group 1, Interviewee 1]. While most participants did not want their friends and family connected to the app, some expressed that it would be helpful to have them understand and provide accountability: "It'll give them an insight of what we're going through" [Focus group 8, Interviewee 1] and "I think it's a good idea to have, help because it could help you track your health and with how many smoking occurrences have you smoked per smoke a day, and can get your family involved and friends" [Focus group 8, Interviewee 7]. When asked if they wanted the app to connect with social media, some participants expressed fear of judgment and mean comments from other social network apps.

> *You know, there's some nasty comments, tellin' you to just get over it.* [Focus group 5, Interviewee 6]

> *I don't want that stuff flashed across Facebook and-and everybody to know what I'm tryin' to do or attemptin' to do because there's a lotta unnecessary opinions.* [Focus group 5, Interviewee 6]

Tracking and User Experience: App Tracking Capability

When asked about app tracking capabilities, most participants noted that they want to be able to track their usage to keep them accountable.

> *Say, I might smoke this many a day. How 'bout the next day—and-and-and a tracker that helps me keep accountability.* [Focus group 5, Interviewee 6]

A few participants noted that tracking their triggers through the app would allow them to learn what to avoid.

> *It help you identify your triggers, and you'll see,...what things cause you to smoke more so you could, like, lean away from it or stay from it,*

XSL•FO
RenderX

> *dependin' on what it is. So, I think it's [an] excellent idea.* [Focus group 3, Interviewee 6]

Some participants expressed a desire to have frequent notification reminders to support accountability.

> *So I need an app that's going to be a constant reminder, almost an annoyance, you know, and - and to help me.* [Focus group 8, Interviewee 2]

Most participants indicated that accountability could be supported through reminders of their goals.

> *The daily goals are there, the notifications are there to remind you of your goals and all, so, uh, that should be helpful.*
>
> *[Focus group 4, Interviewee 5]*

Theme 4: Encouragement and Motivation Were More Salient Incentives Than Monetary Rewards

Gamification and Rewards: Nonmonetary Rewards

When asked about rewards, most participants expressed that encouragement and nonmonetary rewards were motivating.

> *If you reach a goal, maybe,...you will get in this, a motivating statement, you know, as an incentive or a bunch of balloons goin', confetti exploding, you could maybe get a star.* [Focus group 2, Interviewee 1]
>
> *Maybe even just a motivational somethin', say, some kind of quote...even if it was just, "Good job. You - you did not smoke X number of hours today," or "You smoked less by X percent today." That's the reward right there.* [Focus group 8, Interviewee 3]

A few participants noted that receiving advice or feeling accomplished was rewarding.

> *I just like when [it] lets me know that I've accomplished my goals for today.* [Focus group 7, Interviewee 2]
>
> *I just need somethin' to keep me motivated, like a tip a day.* [Focus group 5, Interviewee 2]

As one participant surmised, the app should be a place, "where you can get encouragement...or you can encourage somebody" [Focus group 2, Interviewee 1].

A few participants expressed that tangible incentives are not as encouraging as the benefits of quitting for one's own health.

> *Yeah, it's not about the money, the incentives. It's what I got to do to live longer.* [Focus group 5, Interviewee 2]
>
> *I have a little girl that I gotta live for, so I'm tryin' to do better with- the smokin'...So that's...motivating me to stop.* [Focus group 6, Interviewee 7]

One participant noted that interacting with others within the app and reaching improvement milestones would be rewarding.

> *I don't really care to be rewarded, because I want to quit smoking. So, I just want the motivation there in the app, tracking...how many cigarettes that I, you know, didn't smoke or whatever, you know, form of tobacco you use. And then just-just having the motivation...* [Focus group 7, Interviewee 8]

Gamification and Rewards: Provision for Diverse Reward Options

Other participants expressed that although money is not the most important motivation, it may be important to some people; hence, allowing participants to choose their incentives was ideal. Some participants mentioned financial rewards, such as gift cards or payments.

> *Some apps give you like daily points if you log in daily. So, if you can get a reward for logging in daily and then it could add up to getting the gift card, getting, you know, PayPal transfers, I think that would be pretty enticing.* [Focus group 7, Interviewee 7]
>
> *It always works better when there's options versus you only have maybe one thing that you can do. So, it might be, oh, a cash app payout or Amazon, but then for the next person they might not really want that, they want something else.* [Focus group 7, Interviewee 4]

Discussion

Principal Findings

As more mHealth apps for tobacco cessation are introduced each year, it is critical that these apps are built to serve everyone, including priority populations. Increasingly, apps tailored for Black adults have been introduced to confront health inequities, with an emphasis on iterative design using formative work from members of the target population [37]. Themes and subthemes from the focus groups revealed a strong preference for certain features, including the type and presentation of content in the app, inclusivity, user experience, tracking capability, professional and social support, gamification, and rewards. Preferred features also aligned with components of the self-determination theory, thereby enhancing the significant role that these features may play in the usage of an mHealth app. These findings provide insight into what app features deter or promote app receptivity, and how to attract Black adults who use tobacco products to use an mHealth app for smoking cessation.

Comparison With Previous Work

In our study, we found that participants desired an app with features that highlight representation and promote inclusivity. To do this, participants recommended that an mHealth app should include testimonials, images, and videos featuring Black adults. Prior studies examining preferred features of mHealth apps for Black women have found that participants desire an app to include a directory of nearby Black female health care providers within their community [38,39]. Having racial concordance with a health care provider can help participants feel they have someone with whom they can relate [39]. One way that representation can be achieved in an mHealth app for tobacco cessation is to have a video testimonial by a Black former tobacco user included in the app, with the person sharing their experience with quitting as well as the health benefits experienced after quitting.

XSL•FO
RenderX

We also found participants desired in-app personalization of content and user experience based on their attributes, such as race, gender, and tobacco use habits and experience. Our findings suggest that to improve the user experience, mHealth apps should allow for personalization based on several factors, not just race. This may lead to a more robust experience for the app user. Further, several participants emphasized how the hypothetical mHealth app should feature a diverse range of people with different perspectives, experiences, and identities. This is consistent with findings in one study on app development where participants desired a diversity of people within an app, specifically wanting "everyone" to be included [40].

Similar to our study, previous research has shown that people are interested in an app with relevant health information and statistics [28]. This finding suggests that when health risks and statistics are personalized, the data are more relatable. Providing personalized content and feedback is associated with increased engagement with mobile app interventions [41]. For example, in an mHealth app for tobacco cessation, including statistics about the rate of cardiovascular disease or cancer for Black tobacco users compared with nontobacco users may increase engagement.

Another important finding in this study is that while participants were in favor of representation within the app, they were understandably hesitant or resistant to an app that focused solely on their racial identity. A previous study on mHealth apps for smoking cessation noted similar concern [28]. This perspective may stem from the justified mistrust of the health community and historical mistreatment of Black people in health care and research [42-44]. In a study that examined mHealth research receptivity among Black men, over a quarter (27%) of men and nearly a third of Black women (30%) identified "mistrust of researchers" as a barrier to participation [45,46].

While a more inclusive but nontargeted app is preferred, outreach strategies to reach the Black community were suggested, including marketing at events for the Black community, marketing by Black influential figures or celebrities on social media, and promotion by Black influencers. This finding suggests that having a known source may engender feelings of trust [47]. Furthermore, in recent years, health campaigns have begun working with influential figures such as pastors or influencers to promote health interventions, including vaccine campaigns [47] and physical activity and healthy eating [48,49]. Influencers can leverage their social influence and reach their audiences directly to promote health interventions [50,51], which could be used to increase app receptivity within the Black community.

In addition, with the high rate of social media use by Black adults [25] and the inclusion of social media as a tool for health promotion and intervention dissemination due to wider reach, scalability, and cost-effectiveness [52], targeted social media campaigns can be beneficial. In tobacco cessation studies, the use of social media platforms for recruitment has shown success, especially for minority populations that are typically hard to reach [49]. In one study by Bricker et al [52], tailoring of a social media campaign to disseminate an mHealth app in a way that focused on self-control and health values was found to be cost-effective and associated with high reach based on the number of clicks and conversion to app installation. The use of social media platforms to disseminate an mHealth app for tobacco cessation targeted for Black adults can increase the likelihood of awareness of the mHealth app, which may convert into more interest and greater use.

Furthermore, prior research to investigate smokers' reactions on social media platforms revealed that comments and the type of reaction such as "love" or "haha" were a reflection of the level of motivation with regard to readiness to quit [53]. This suggests that although social media can be used to enhance the visibility of mHealth apps for tobacco cessation, it can also stimulate motivational processes by serving as a space for cessation-related posts that contain culturally representative narratives, prompting users to share their own motivations and goals for quitting or positively respond to comments that are reflective of their own experiences. This way, social media is not only a tool to promote the mHealth app but can also help to assess qualitative measures such as levels of engagement and motivation to quit.

Our findings also revealed the importance of an in-app community that offers accountability through supportive connections (eg, through health care providers and quit buddy programs). Previous research on mHealth apps has found that many people desire the ability to connect with health care providers within apps [39]. One study examining supportive accountability through counselor monitoring and supportive advice found that having a trusted expert within an mHealth app intervention for smoking cessation led to increased app engagement [54]. In another intervention that used peer-to-peer exchanges between participants seeking smoking cessation resources and communications from experts with treatment-related questions for participants to discuss, doubled sustained abstinence was observed compared to those who only used nicotine patches and had access to a cessation website [55]. Further, the effectiveness of mHealth interventions can be enhanced by human and social support [56,57].

Although open to connections, participants were hesitant and less receptive to having family or social media integrated within the app due to privacy concerns. Other studies examining receptivity to social media on mHealth apps have similarly found that most participants did not want social media integration, citing privacy concerns [39]. Although privacy is a major concern, social media play a significant role by serving as a platform through which health-related behavior change interventions can be delivered and health-related connections or support groups developed in a meaningful way [58,59]. For Black adults who use tobacco products, addressing the privacy concern in a transparent way may improve receptivity to the integration of elements of social media into the intervention.

In contrast, some people prefer having emotional support, such as encouragement, from friends and family [60,61]. Studies have found that stigma against individuals who smoke can prevent them from seeking care [62]. One study found that negative comments from one's social network about their smoking habits can even lead to a decrease in quitting progress

XSL•FO
RenderX

[61]. To combat this concern, any external connections to the app should be optional.

App tracking capability was also an important feature for accountability and has been identified as a desirable feature and method to promote accountability in other studies [39,54]. Several mHealth interventions use the tracking feature to help monitor chronic health conditions and medication intake [63,64]. One qualitative study that explored experiences of Black adults who use tobacco products with a specific mHealth app for smoking cessation identified that the tracking feature was highly preferred [28]. With mHealth apps for tobacco cessation, tracking of the rate of use of tobacco products, cravings, or money saved over a period of time from less use can lead to sustained motivation and cessation-related behavior.

Lastly, we found that while some participants said that financial incentives could attract some people to use the app, many participants did not think financial incentives were necessary to attract users to the app. Prior research suggests that financial incentives can modify or increase adherence to desired behavior changes, like exercise [65], dietary changes [66], or engagement with a state Quitline [67]. However, the sustained provision of financial incentives may not be feasible for mHealth apps long term, nor are they always necessary. In one study examining motivations and barriers to participating in mHealth research among Black males, approximately 40% indicated that offering financial incentives would motivate participation, while only 16% indicated that a lack of financial incentives was a barrier to participation [45]. Similar rates were also seen among Black women, 39% and 15%, respectively [46]. Continued engagement with an app is further augmented by other considerations, like the content shared [68]. Therefore, while financial incentives are desirable, their absence may not significantly deter participation with mHealth interventions. This finding aligns with existing research that the gamification (eg, inclusion of game elements and rewards) of mobile apps can reward and motivate app users [69,70] and promote behavior change [71].

Design Implications for App Development

To enhance the translation and application of our findings to an mHealth app development, we mapped out each subtheme and domain feature into examples of concrete mHealth app features with pertinent processes from the self-determination theory. We also included examples of ethical and privacy considerations, as well as potential metrics to consider during implementation. The map, as outlined in Table 3, shows how representation and inclusivity, personalization, content, accountability, tracking and user experience, social and professional support, and nonmonetary incentives can be operationalized into actual, concrete features within an mHealth app for tobacco cessation.

XSL•FO
RenderX

Table . Design implications for mobile health app with preferred features.

Subthemes and domains of app features	Examples of concrete app features	Self-determination theory components	Privacy/ethical considerations	Potential metrics
Representation and inclusivity	• Video testimonials from other Black tobacco users • Culturally relevant language	• Relatedness • Autonomy	• Avoidance of identifying information of individuals in videos or photos • Opt-in for culturally relevant material, instead of it being the default • Informed consent • Avoidance of racial profiling in content	• Amount of time spent and frequency of viewing culturally relevant information • Rate and frequency of opt-in for culturally relevant material
Personalization	• In-app customization of goals, quit date, reminders, and prompts • Personalized diary	• Autonomy	• User-controlled input of information • Password-protected storage and opt-out option	• Achievement of personal goals • Number and frequency of diary entries
Content	• Health risk statistics relevant to the target user	• Relatedness • Competence	• Ensure credibility of the source of information • Avoidance of the use of sensitive language or medical jargon	• Completion of content sections
Accountability	• Quit buddy, daily check-in	• Autonomy • Relatedness	• User opt-in • Anonymization of profile	• Daily active use
Tracking and user experience	• Tracking tools for mood, cravings, number of cigarettes used, and days without use	• Autonomy • Competence	• User control on the type of information to track or delete • Secure or password-protected storage of tracked data	• Daily active use
Social support	• Anonymous in-app community	• Relatedness	• Anonymization of user profile • User opt-in for participation or for integration with social media	• Number of users who opt-in • Number and frequency of comments and posts on the community board
Professional support	• Quitline or access to a tobacco treatment specialist, trusted expert chat	• Competence	• Informed consent to interact with a professional • Encryption of messages	• Number and frequency of use or messages sent to an expert • Number of times the Quitline button or icon is used
Nonmonetary incentives	• Badges, points	• Competence	• Option for private mode where reward is only seen by the user • No shaming or negative language during relapse	• Rate of badges or points earned

Limitations

A limitation in this study is that findings may not be generalizable. First, we used a convenience sample online and recruited from one social media platform. This may have limited the sample size as well as the diversity of participants in terms of gender, age, educational level, and income. Conducting the focus groups online, as well as the requirement of a history of app usage or willingness to use an mHealth app in the future for participation in this study, further limits the diversity of

XSL•FO
RenderX

perspectives about features of mHealth apps for tobacco cessation. Furthermore, findings may also not be generalizable to Black adults outside of the United States due to cultural differences and other factors that may exist across countries. Obtaining the perception of Black adult tobacco users who are not receptive to the idea of mHealth apps for tobacco cessation may provide additional insight into how best to develop and market these apps in a more inclusive way.

Conclusions

Our findings confirm the need to design mHealth apps for tobacco cessation with features that are reflective of cultural relevance, inclusivity, and diversity of experiences without excessive targeting and support motivation as stated in the self-determination theory [19-21]. Consistent with Perski and colleagues [72], engagement with mHealth apps increases when design features support personal relevance, motivation to quit, and credibility. Our findings provide insight into the improvement of existing apps and future development of an mHealth app that can appeal to Black adults who use tobacco products, potentially increasing app use, engagement, and successful cessation. Findings from this research can also serve as a groundwork for future research to explore the effectiveness of an mHealth app culturally tailored with preferred features optimized to enhance engagement by Black adults who use tobacco products and to identify other factors that may improve usage and lead to better outcomes for the target population.

Funding

This project was supported by the National Institute of Drug Abuse of the National Institutes of Health (NIH) through grant K23DA057416. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Data Availability

The focus group guide is available in Multimedia Appendix 1. Due to confidentiality and ethical considerations, we are unable to publicly share full transcripts and coded data. Additional nonidentifiable information can be provided upon reasonable request to the corresponding author.

Authors' Contributions

CE, CEK, and AOG contributed to the conceptualization and design of the study. CE and SAC led the focus groups. SAC and RP analyzed the data. CE, SAC, RP, CEK, and AOG contributed to the interpretation of the findings. CE and SAC were involved in the drafting of the manuscript, while all authors were involved in the revision of the manuscript prior to submission. All authors gave approval for the final version for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Focus group guide.
[PDF File, 280 KB - mhealth_v14i1e63340_app1.pdf]

References

1. Henley SJ, Thomas CC, Sharapova SR, et al. Vital signs: disparities in tobacco-related cancer incidence and mortality - United States, 2004-2013. MMWR Morb Mortal Wkly Rep 2016 Nov 11;65(44):1212-1218. [doi: 10.15585/mmwr.mm6544a3] [Medline: 27832048]
2. Tobacco use in the Black American community. Truth Initiative. 2024 Aug 27. URL: https://truthinitiative.org/research-resources/targeted-communities/tobacco-use-african-american-community [accessed 2026-01-23]
3. Alexander LA, Trinidad DR, Sakuma KLK, et al. Why we must continue to investigate menthol's role in the African American smoking paradox.. Nicotine Tob Res 2016 Apr;18 Suppl 1(Suppl 1):S91-101. [doi: 10.1093/ntr/ntv209] [Medline: 26980870]
4. Cancer facts & figures for African American/Black people 2022-2024. : American Cancer Society; 2022 URL: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/cancer-facts-and-figures-for-african-americans/2022-2024-cff-aa.pdf [accessed 2026-01-23]
5. Health behaviors of adults: United States, 2008-2010. : National Center for Health Statistics; 2013 May Vital and Health Statistics, Series 10, Number 257 URL: https://www.cdc.gov/nchs/data/series/sr_10/sr10_257.pdf [accessed 2026-01-23]
6. Nollen NL, Cox LS, Yu Q, et al. A clinical trial to examine disparities in quitting between African-American and White adult smokers: design, accrual, and baseline characteristics. Contemp Clin Trials 2016 Mar;47:12-21. [doi: 10.1016/j.cct.2015.12.001] [Medline: 26667382]

XSL•FO
RenderX

7. Carroll DM, Cole A. Racial/ethnic group comparisons of quit ratios and prevalences of cessation-related factors among adults who smoke with a quit attempt. Am J Drug Alcohol Abuse 2022 Jan 2;48(1):58-68. [doi: 10.1080/00952990.2021.1977310] [Medline: 34752715]
8. Enyioha C, Hall M, Voisin C, Jonas D. Effectiveness of mobile phone and web-based interventions for diabetes and obesity among African American and Hispanic adults in the United States: systematic review. JMIR Public Health Surveill 2022 Feb 4;8(2):e25890. [doi: 10.2196/25890] [Medline: 35119368]
9. Willis HA, Neblett EW. Developing culturally-adapted mobile mental health interventions: a mixed methods approach. Mhealth 2023;9:1. [doi: 10.21037/mhealth-22-19] [Medline: 36760787]
10. Bhuyan SS, Lu N, Chandak A, et al. Use of mobile health applications for health-seeking behavior among US adults. J Med Syst 2016 Jun;40(6):153. [doi: 10.1007/s10916-016-0492-7] [Medline: 27147516]
11. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415. [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
12. Haskins BL, Lesperance D, Gibbons P, Boudreaux ED. A systematic review of smartphone applications for smoking cessation. Transl Behav Med 2017 Jun;7(2):292-299. [doi: 10.1007/s13142-017-0492-2] [Medline: 28527027]
13. James DCS, Harville C. Smartphone usage, social media engagement, and willingness to participate in mHealth weight management research among African American women. Health Educ Behav 2018;45(3):315-322. [doi: 10.1177/1090198117714020] [Medline: 28606004]
14. Goodwin RD, Ganz O, Weinberger AH, Smith PH, Wyka K, Delnevo CD. Menthol cigarette use among adults who smoke cigarettes, 2008-2020: rapid growth and widening inequities in the United States. Nicotine Tob Res 2023 Mar 22;25(4):692-698. [doi: 10.1093/ntr/ntac214] [Medline: 36223889]
15. Trinidad DR, Xie B, Fagan P, et al. Disparities in the population distribution of African American and non-Hispanic White smokers along the quitting continuum. Health Educ Behav 2015 Dec;42(6):742-751. [doi: 10.1177/1090198115577376] [Medline: 25794519]
16. Bennett GG, Wolin KY, Robinson EL, Fowler S, Edwards CL. Perceived racial/ethnic harassment and tobacco use among African American young adults. Am J Public Health 2005 Feb;95(2):238-240. [doi: 10.2105/AJPH.2004.037812] [Medline: 15671457]
17. Zhou ES, Revette A, Ritterband LM, et al. Developing a culturally tailored digital health intervention for insomnia in Black women. Transl Behav Med 2024 Feb 7;14(2):117-126. [doi: 10.1093/tbm/ibad056]
18. Wadi NM, Asantewa-Ampaduh S, Rivas C, Goff LM. Culturally tailored lifestyle interventions for the prevention and management of type 2 diabetes in adults of Black African ancestry: a systematic review of tailoring methods and their effectiveness. Public Health Nutr 2022 Feb;25(2):422-436. [doi: 10.1017/S1368980021003682] [Medline: 34435943]
19. Deci EL, Ryan RM. The “what” and “why” of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 2000 Oct;11(4):227-268. [doi: 10.1207/S15327965PLI1104_01]
20. Williams GC, McGregor HA, Sharp D, et al. Testing a self-determination theory intervention for motivating tobacco cessation: supporting autonomy and competence in a clinical trial. Health Psychol 2006 Jan;25(1):91-101. [doi: 10.1037/0278-6133.25.1.91] [Medline: 16448302]
21. Choi J, Noh GY, Park DJ. Smoking cessation apps for smartphones: content analysis with the self-determination theory. J Med Internet Res 2014 Feb 12;16(2):e44. [doi: 10.2196/jmir.3061] [Medline: 24521881]
22. Williams GC, Gagné M, Ryan RM, Deci EL. Facilitating autonomous motivation for smoking cessation. Health Psychol 2002 Jan;21(1):40-50. [Medline: 11846344]
23. Ubhi HK, Michie S, Kotz D, Wong WC, West R. A mobile app to aid smoking cessation: preliminary evaluation of SmokeFree28. J Med Internet Res 2015 Jan 16;17(1):e17. [doi: 10.2196/jmir.3479] [Medline: 25596170]
24. Heffner JL, Vilardaga R, Mercer LD, Kientz JA, Bricker JB. Feature-level analysis of a novel smartphone application for smoking cessation. Am J Drug Alcohol Abuse 2015 Jan;41(1):68-73. [doi: 10.3109/00952990.2014.977486] [Medline: 25397860]
25. Social media use in 2021. Pew Research Center. 2021 Apr 7. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2026-01-23]
26. Bricker JB, Copeland W, Mull KE, et al. Single-arm trial of the second version of an acceptance & commitment therapy smartphone application for smoking cessation. Drug Alcohol Depend 2017 Jan 1;170:37-42. [doi: 10.1016/j.drugalcdep.2016.10.029] [Medline: 27870987]
27. Struik LL, Bottorff JL, Baskerville NB, Oliffe J, Crichton S. Comparison of developers’ and end-users’ perspectives about smoking cessation support through the Crush the Crave app. JMIR mHealth uHealth 2019 Mar 7;7(3):e10750. [doi: 10.2196/10750] [Medline: 30843864]
28. Enyioha C, Loufman LM, Grewe ME, et al. Black smokers’ preferences for features of a smoking cessation app: qualitative study. JMIR Form Res 2023 May 30;7:e43603. [doi: 10.2196/43603] [Medline: 37252777]
29. Barroso-Hurtado M, Suárez-Castro D, Martínez-Vispo C, Becoña E, López-Durán A. Smoking cessation apps: a systematic review of format, outcomes, and features. Int J Environ Res Public Health 2021 Nov 6;18(21):11664. [doi: 10.3390/ijerph182111664] [Medline: 34770178]

XSL•FO
RenderX

30. Buller DB, Borland R, Bettinghaus EP, Shane JH, Zimmerman DE. Randomized trial of a smartphone mobile application compared to text messaging to support smoking cessation. Telemed J E Health 2014 Mar;20(3):206-214. [doi: 10.1089/tmj.2013.0169] [Medline: 24350804]
31. Iacoviello BM, Steinerman JR, Klein DB, et al. Clickotine, a personalized smartphone app for smoking cessation: initial evaluation. JMIR mHealth uHealth 2017;5(4):e56. [doi: 10.2196/mhealth.7226]
32. Brown K, Dyas J, Chahal P, Khalil Y, Riaz P, Cummings-Jones J. Discovering the research priorities of people with diabetes in a multicultural community: a focus group study. Br J Gen Pract 2006 Mar;56(524):206-213. [Medline: 16536961]
33. Hayashi P, Abib G, Hoppen N, Wolff LDG. Processual validity in qualitative research in healthcare. Inquiry 2021;58. [doi: 10.1177/00469580211060750] [Medline: 34845941]
34. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
35. Grbich C. Qualitative Data Analysis: An Introduction: SAGE Publications; 2013. [doi: 10.4135/9781529799606]
36. Agaku IT, Alpert HR. Trends in annual sales and current use of cigarettes, cigars, roll-your-own tobacco, pipes, and smokeless tobacco among US adults, 2002-2012. Tob Control 2016 Jul;25(4):451-457. [doi: 10.1136/tobaccocontrol-2014-052125] [Medline: 25899447]
37. Samarasekera U. The rise of racial minority health apps. Lancet Digit Health 2022 Apr;4(4):e218-e219. [doi: 10.1016/S2589-7500(22)00043-7] [Medline: 35337641]
38. McCall T, Ali MO, Yu F, Fontelo P, Khairat S. Development of a mobile app to support self-management of anxiety and depression in African American women: usability study. JMIR Form Res 2021 Aug 17;5(8):e24393. [doi: 10.2196/24393] [Medline: 34133313]
39. Chandler R, Hernandez N, Guillaume D, Grandoit S, Branch-Ellis D, Lightfoot M. A community-engaged approach to creating a mobile HIV prevention app for Black Women: focus group study to determine preferences via prototype demos. JMIR mHealth uHealth 2020 Jul 24;8(7):e18437. [doi: 10.2196/18437] [Medline: 32706723]
40. White MJ, Xie R, Lane H, et al. Organizational trust, usability, and inclusivity are key implementation facilitators for a proposed assets-based mobile health intervention. Transl Behav Med 2023 Jul 1;13(7):465-474. [doi: 10.1093/tbm/ibac108] [Medline: 36999807]
41. Oakley-Girvan I, Yunis R, Longmire M, Ouillon JS. What works best to engage participants in mobile app interventions and e-Health: a scoping review. Telemed J E Health 2022 Jun;28(6):768-780. [doi: 10.1089/tmj.2021.0176] [Medline: 34637651]
42. Dong L, Bogart LM, Gandhi P, et al. A qualitative study of COVID-19 vaccine intentions and mistrust in Black Americans: recommendations for vaccine dissemination and uptake. PLoS ONE 2022;17(5):e0268020. [doi: 10.1371/journal.pone.0268020]
43. Nguyen BT, Brown AL, Jones F, et al. “I’m not going to be a guinea pig”: medical mistrust as a barrier to male contraception for Black American men in Los Angeles, CA. Contraception 2021 Oct;104(4):361-366. [doi: 10.1016/j.contraception.2021.06.001] [Medline: 34118271]
44. Scharff DP, Mathews KJ, Jackson P, Hoffsuemmer J, Martin E, Edwards D. More than Tuskegee: understanding mistrust about research participation. J Health Care Poor Underserved 2010 Aug;21(3):879-897. [doi: 10.1353/hpu.0.0323] [Medline: 20693733]
45. James DCS, Harville C. Barriers and motivators to participating in mHealth research among African American men. Am J Mens Health 2017 Nov;11(6):1605-1613. [doi: 10.1177/1557988315620276] [Medline: 26634861]
46. James DCS, Harville C II, Whitehead N, Stellefson M, Dodani S, Sears C. Willingness of African American women to participate in e-Health/m-Health research. Telemed J E Health 2016 Mar;22(3):191-197. [doi: 10.1089/tmj.2015.0071] [Medline: 26313323]
47. Moore D, Mansfield LN, Onsomu EO, Caviness-Ashe N. The role of Black pastors in disseminating COVID-19 vaccination information to Black communities in South Carolina. Int J Environ Res Public Health 2022 Jul 22;19(15):8926. [doi: 10.3390/ijerph19158926] [Medline: 35897301]
48. Goodyear VA, Boardley I, Chiou SY, et al. Social media use informing behaviours related to physical activity, diet and quality of life during COVID-19: a mixed methods study. BMC Public Health 2021 Jul 6;21(1):1333. [doi: 10.1186/s12889-021-11398-0] [Medline: 34229651]
49. Pechmann C, Phillips C, Calder D, Prochaska JJ. Facebook recruitment using zip codes to improve diversity in health research: longitudinal observational study. J Med Internet Res 2020 Jun 5;22(6):e17554. [doi: 10.2196/17554] [Medline: 32501274]
50. de Vere Hunt I, Linos E. Social media for public health: framework for social media-based public health campaigns. J Med Internet Res 2022 Dec 14;24(12):e42179. [doi: 10.2196/42179] [Medline: 36515995]
51. Powell J, Pring T. The impact of social media influencers on health outcomes: systematic review. Soc Sci Med 2024 Jan;340:116472. [doi: 10.1016/j.socscimed.2023.116472] [Medline: 38070305]
52. Bricker JB, Santiago-Torres M, Mull KE, Sullivan BM, Mehrotra R. Population-level dissemination of a smoking cessation smartphone app: quasi-experimental comparison of values-based messages in social media advertisements. JMIR mHealth uHealth 2025 Jul 28;13:e71619. [doi: 10.2196/71619] [Medline: 40720905]

XSL•FO
RenderX

53. Watti J, Millner M, Siklósi K, Kiss H, Kelemen O, Pócs D. Smokers' engagement behavior on Facebook: verbalizing and visual expressing the smoking cessation process. Int J Environ Res Public Health 2022 Aug 12;19(16):9983. [doi: 10.3390/ijerph19169983] [Medline: 36011617]
54. Lepore SJ, Collins BN, Killam HW, Barry B. Supportive accountability and mobile app use in a tobacco control intervention targeting low-income minority mothers who smoke: observational study. JMIR mHealth uHealth 2021 Jul 2;9(7):e28175. [doi: 10.2196/28175] [Medline: 34255698]
55. Pechmann C, Delucchi K, Lakon CM, Prochaska JJ. Randomised controlled trial evaluation of Tweet2Quit: a social network quit-smoking intervention. Tob Control 2017 Mar;26(2):188-194. [doi: 10.1136/tobaccocontrol-2015-052768] [Medline: 26928205]
56. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30. [doi: 10.2196/jmir.1602]
57. de Dios MA, Stanton CA, Cano M, Lloyd-Richardson E, Niaura R. The influence of social support on smoking cessation treatment adherence among HIV+ smokers. Nicotine Tob Res 2016 May;18(5):1126-1133. [doi: 10.1093/ntr/ntv144] [Medline: 26116086]
58. Jacobs MA, Cobb CO, Abroms L, Graham AL. Facebook apps for smoking cessation: a review of content and adherence to evidence-based guidelines. J Med Internet Res 2014 Sep 9;16(9):e205. [doi: 10.2196/jmir.3491] [Medline: 25205129]
59. Villanti AC, Johnson AL, Ilakkuvan V, Jacobs MA, Graham AL, Rath JM. Social media use and access to digital technology in US young adults in 2016. J Med Internet Res 2017;19(6):e196. [doi: 10.2196/jmir.7303]
60. Aschbrenner KA, Naslund JA, Gill L, Bartels SJ, O'Malley AJ, Brunette MF. Preferences for smoking cessation support from family and friends among adults with serious mental illness. Psychiatr Q 2017 Dec;88(4):701-710. [doi: 10.1007/s11126-016-9485-4]
61. Aschbrenner KA, Naslund JA, Gill L, et al. Qualitative analysis of social network influences on quitting smoking among individuals with serious mental illness. J Ment Health 2019 Oct;28(5):475-481. [doi: 10.1080/09638237.2017.1340600] [Medline: 28675331]
62. Madawala S, Enticott J, Sturgiss E, Selamoglu M, Barton C. The impact of smoking status on anticipated stigma and experience of care among smokers and ex-smokers with chronic illness in general practice. Chronic Illn 2023 Sep;19(3):557-570. [doi: 10.1177/17423953221101337] [Medline: 35575240]
63. Donevant SB, Estrada RD, Culley JM, Habing B, Adams SA. Exploring app features with outcomes in mHealth studies involving chronic respiratory diseases, diabetes, and hypertension: a targeted exploration of the literature. J Am Med Inform Assoc 2018 Oct 1;25(10):1407-1418. [doi: 10.1093/jamia/ocy104] [Medline: 30137383]
64. Madujibeya I, Lennie TA, Pelzel J, Moser DK. Patients' experiences using a mobile health app for self-care of heart failure in a real-world setting: qualitative analysis. JMIR Form Res 2023 Aug 15;7:e39525. [doi: 10.2196/39525] [Medline: 37581912]
65. Mitchell MS, Goodman JM, Alter DA, et al. Financial incentives for exercise adherence in adults: systematic review and meta-analysis. Am J Prev Med 2013 Nov;45(5):658-667. [doi: 10.1016/j.amepre.2013.06.017] [Medline: 24139781]
66. Wall J, Mhurchu CN, Blakely T, Rodgers A, Wilton J. Effectiveness of monetary incentives in modifying dietary behavior: a review of randomized, controlled trials. Nutr Rev 2006 Dec;64(12):518-531. [doi: 10.1111/j.1753-4887.2006.tb00185.x] [Medline: 17274494]
67. Bourne DE, Williams R, Osbahr L, Roemhildt M, Villanti AC. Implementation of quitline financial incentives to increase counseling sessions among adults who use menthol tobacco products. Health Promot Pract 2024 Mar;25(2):167-169. [doi: 10.1177/15248399231171143] [Medline: 37118924]
68. Brower J, LaBarge MC, White L, Mitchell MS. Examining responsiveness to an incentive-based mobile health app: longitudinal observational study. J Med Internet Res 2020 Aug 10;22(8):e16797. [doi: 10.2196/16797] [Medline: 32773371]
69. Villalobos-Zúñiga G, Cherubini M. Apps that motivate: a taxonomy of app features based on self-determination theory. Int J Hum Comput Stud 2020 Aug;140:102449. [doi: 10.1016/j.ijhcs.2020.102449]
70. Tran S, Smith L, El-Den S, Carter S. The use of gamification and incentives in mobile health apps to improve medication adherence: scoping review. JMIR mHealth uHealth 2022 Feb 21;10(2):e30671. [doi: 10.2196/30671] [Medline: 35188475]
71. de Oliveira R, Cherubini M, Oliver N. MoviPill: improving medication compliance for elders using a mobile persuasive social game. In: UbiComp '10: Proceedings of the 12th ACM International Conference on Ubiquitous Computing: Association for Computing Machinery; 2010:251-260. [doi: 10.1145/1864349.1864371]
72. Perski O, Blandford A, Ubhi HK, West R, Michie S. Smokers' and drinkers' choice of smartphone applications and expectations of engagement: a think aloud and interview study. BMC Med Inform Decis Mak 2017 Feb 28;17(1):25. [doi: 10.1186/s12911-017-0422-8] [Medline: 28241759]

Abbreviations

mHealth: mobile health

XSL•FO
RenderX

Edited by L Buis; submitted 17.Jun.2024; peer-reviewed by D Pócs, E Ross, M Pai; revised version received 06.Jan.2026; accepted 08.Jan.2026; published 06.Feb.2026.

Please cite as:
Clark SA, Philips R, Kistler CE, Goldstein AO, Enyioha C
Features of Mobile Health Apps for Tobacco Cessation That Appeal to Black Adults Who Use Tobacco Products: Focus Group Study
JMIR Mhealth Uhealth 2026;14:e63340
URL: https://mhealth.jmir.org/2026/1/e63340
doi:10.2196/63340

© Sonia A Clark, Remi Philips, Christine E Kistler, Adam O Goldstein, Chineme Enyioha. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 6.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Designing mHealth Apps for Substance Use Recovery Through Real-World Co-Design and Deployment: Mixed Methods Study

Teale Masrani[1], BSc, MSc; David C Hodgins[2,3], BA, MA, PhD; Hyoun S Kim[2,4], BA, MA, PhD; Katherine Rittenbach[3,5], BA, PhD; Erika Johnson[6], BA, MSc; Geoffrey Messier[1], BEng, BSc, MSc, PhD

[1]Department of Electrical and Software Engineering, Schulich School of Engineering, University of Calgary, Calgary, AB, Canada

[2]Department of Psychology, Faculty of Arts, University of Calgary, Calgary, AB, Canada

[3]Department of Psychiatry, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[4]Department of Psychology, Faculty of Arts, Toronto Metropolitan University, Toronto, ON, Canada

[5]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[6]Department of Clinical Neurosciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Corresponding Author:
Geoffrey Messier, BEng, BSc, MSc, PhD
Department of Electrical and Software Engineering
Schulich School of Engineering
University of Calgary
2500 University Drive NW
Calgary, AB, T2N 1N4
Canada
Phone: 1 403 220 5110
Email: gmessier@ucalgary.ca

Abstract

Background: Mobile health (mHealth) apps have shown promise to support recovery from substance use disorders. However, evidence on engagement and efficacy is still inconclusive.

Objective: This study aims to identify design considerations for optimizing engagement in mHealth apps for those recovering from problematic substance use, by analyzing real-world experiences with co-designed app features.

Methods: We co-designed, deployed, and evaluated an mHealth app. Initial co-design interviews with 14 individuals in recovery led to 3 new features integrated into an existing mHealth app. The app was deployed for a 6-week trial with 53 participants using it during their daily routines without researcher supervision. Usage patterns were analyzed throughout the trial period, and follow-up interviews with 12 app users foregrounded subjective usage experiences and considerations for future design.

Results: We developed 3 new features following co-design interviews: a goal-setting feature, a craving tracker, and a meetings log. Usage metrics revealed mixed engagement, with 45.3% (24/53) of users actively engaging with the app throughout the trial. These active users opened the app 27.1 unique times on average, with a retention rate after 30 days among active users of 45.8% (11/24), exceeding the typical mobile app retention benchmark of 7% after 30 days. Interviews revealed that participants preferred app functionality to extend beyond substance use domains to support other dimensions of their lives not directly pertaining to substance use, such as general goals and daily routines. Participants further suggested that recovery apps should act as private digital journals while also providing a sense of community and connection to broader recovery ecosystems. Additionally, mHealth designs that allow users to configure their own personalized recovery pathways in the app can benefit some users who appreciate increased autonomy, while others may become overwhelmed by a lack of prescriptive guidance.

Conclusions: It is valuable to incorporate iterative co-design methodologies into digital health and recovery app research to help optimize engagement. Furthermore, recovery apps can benefit from flexible designs with customizable degrees of user autonomy. Future designers can better cater to individual user preferences by personalizing mHealth designs so that they strike a balance between system control and user control over digital recovery pathways.

(JMIR Mhealth Uhealth 2026;14:e83984) doi:10.2196/83984

KEYWORDS

co-design; digital health; mHealth; smartphone application; substance use disorder

XSL•FO
RenderX

Introduction

Mobile health (mHealth) apps have become popular tools to manage personal health, offering inexpensive and accessible health guidance, self-monitoring of progress, and interactive communication with clinicians, among other capabilities [1]. A major aspect of some individuals' personal health is use of substances, and recovery from disordered substance use. "Recovery" is defined by the individual and can have a variety of meanings, from reducing problematic use to practicing full abstinence [2,3]. Specialized mobile health apps have become popular tools to facilitate all forms of recovery, with many commercially available in iOS and Android app stores [4-6].

Prior research finds some potential for mHealth apps to facilitate recovery [7-10]. However, the effectiveness and engagement of these interventions remain unclear [8,9,11-13]. Effects are often minimal, or only notable when the app is used in conjunction with other non-app-based treatments [14-16]. Furthermore, many studies in this field use a randomized controlled trial methodology. These studies often compare any usage of the app to treatment-as-usual, and tend to overlook the details of how users engaged with particular app features and their motivations behind specific engagement patterns [8,12,17-19]. Therefore, some researchers argue that studies "should not assess the success of addiction recovery apps based only on objective measures of changes in substance use" [12], and should instead aim to understand users' subjective experiences when using such tools as a way to evaluate quality.

This body of research may additionally benefit from a greater focus on the designer's point-of-view, unpacking the details of how digital recovery tools can be designed to best support their users [17,18], rather than only studying user outcomes after the app has already been built and deployed. Careful user-experience design and research play key roles in the success of mHealth apps [12,19], as they can involve human-centered techniques such as value-sensitive design [20,21] and participatory or collaborative design (co-design) [22-24]. In the field of mobile recovery apps, however, such human-centered design techniques are scarce [13]. This study incorporates the co-design and deployment of an mHealth recovery app to investigate the real-world experiences of people dealing with problematic substance use. In doing so, we shed light on practical design considerations to optimize app engagement among this population.

There is a small but growing body of research that shares the aim of centering end users' perspectives to design engaging mHealth recovery apps [25-28]. One prior study is particularly relevant: Jones et al [29] investigated how individuals engaged with an early version of a self-tracking app, identifying considerations for future designs. Similar to our work, Jones et al deployed a working prototype of the tool to stimulate discussion with participants and gather mixed methods usage data. Notably, Jones et al [29] focused only on features that support day-to-day activity tracking, discussing themes pertaining to the types of activities participants tend to track and how they preferred to track them. Herein, we used similar methodologies in pursuit of similar aims, but importantly expanded beyond activity tracking to also investigate the design of goal-setting features and several other mHealth feature types.

Specifically, our research objective was to gain a feature-level understanding of how to create an mHealth app that engages people on their recovery journey. We achieved this objective by co-designing new app features and examining how individuals in recovery perceived and interacted with them during their day-to-day lives over a 6-week period. We took a mixed methods approach to evaluate both quantitative usage metrics and qualitative accounts of users' experiences with the new features.

Methods

Overview

We adopted a research through design (RtD) approach [30-32]. RtD involves conducting research that "employs the methods, practices, and processes of design practice, with the intention of generating new knowledge" [31]. RtD encompasses the hands-on practice of designing a new tool, evaluating its functionality, and reflecting on its design to identify implications for future technologies. In accordance with this hands-on approach, we collaborated with the developers of an existing mHealth app, Zamplo [33], to complete three phases of this work: (1) initial co-design interviews; (2) design, deployment, and a 6-week app trial; and (3) user-experience interviews.

The Zamplo mHealth App

Zamplo's existing functionality does not cater specifically to substance use recovery, as it is used as a general mHealth tool. Users of mHealth technologies tend to face difficulties using a suite of different apps with several overlapping and fragmented features, often preferring a single, flexible unified platform instead [34]. Zamplo aims to overcome these issues with an array of flexible digital health capabilities within one platform to support multiple aspects of personal health. Existing Zamplo capabilities include customizable reminders, daily personal tracking, and visualizations of medications and activities, as well as tracking of mental and physical health symptoms, and management of health documents, to-do lists, notes, and important contacts.

For this study, we configured these features to not assume the preferences of users, such as what specific health metrics they may want to track, what types of goals they will set, or how frequently they prefer to receive reminders. Instead, we configured Zamplo as an open-ended platform, allowing users to self-track, set goals, create reminders, and access several other mHealth features, in whichever ways they choose, without enforcing any planned usage patterns. This open-ended approach allowed for naturalistic observations of how individuals organically use mHealth features when not pushed to engage in a particular way.

We co-designed new features to add to the Zamplo platform and used Zamplo as a technology probe. Hutchinson et al [35] define technology probes as technological artifacts that guide researchers in understanding their users in a real world setting and inspire participants to think creatively about new potential designs [35]. Hence, Zamplo was used as a concrete artifact to

XSL•FO
RenderX

ground participants' interviews, stimulate discussions of users' experiences, and crystalize users' needs. It was a particularly well-suited technology probe as it was created and approved for the secure storage of health data, allowing extensive possible usages in the trial than available in most mHealth apps. This research consisted of three phases, as outlined below.

Phase 1: Initial Co-Design Interviews

Participants

This research began with an initial round of interviews to capture design ideas from individuals with experience recovering from substance use disorders. Participants were recruited from 5 local addiction treatment centers. Of the 23 individuals who signed up to participate, 14 attended an interview. Informed consent was obtained via an electronic consent form and CAD $20 (USD $14.76) electronic gift cards were offered to all interviewees.

Procedure

The research team collaboratively developed a semistructured interview protocol, and 45-60 minute interviews were conducted virtually by one interviewer. Participants were first asked about their recovery history and experiences using technology. This contextualized the primary component of each interview, which was a demonstration of the existing Zamplo platform, followed by inquiry into participants' impressions and ideas for future designs.

Interviews revealed participants' stated needs regarding mHealth technologies and their perceived role in recovery. While the interview protocol provided a framework for discussion topics, the direction of each interview was guided by participants [36]. As such, participants were encouraged to speak openly and creatively about their ideas for Zamplo and were prompted to provide both positive and negative feedback. To mitigate social desirability bias, the interviewer clearly stated that the app was not their own and that all critiques are welcome.

Data Analysis

Interviews were recorded using Zoom (Zoom Video Communications, Inc) and transcribed using Rev AI (Rev.com, Inc) transcription software. The interviewer then manually reviewed and corrected each transcript, before manually analyzing transcripts using the NVivo (Lumivero) platform. A generative open-coding approach was taken for each transcript, iteratively generating, refining, and categorizing original codes throughout analysis [37]. The goal of coding was to identify actionable design implications for new Zamplo features. Several design themes based on the interviews were identified and collaboratively refined via discussions with the research team. These results were relayed back to Zamplo developers to implement, launching the next phase of the research.

Phase 2: Design, Deployment, and 6-Week App Trial

Overview

Interview themes from Phase 1 laid the foundation for creative co-design discussions with the research team and the lead Zamplo developer. Discussions resulted in several low-fidelity wireframe-style prototypes [38] of potential new features, based directly on input received from participants. A new trial version of Zamplo was then deployed on iOS and Android app stores, and recruitment opened. There were two goals for this trial: first, to gather quantitative usage metrics to evaluate usage patterns and engagement; second, to provide participants with a concrete mHealth app experience to stimulate productive reflections during Phase 3 interviews. The following sections detail the methods of this trial.

Participants

Participants were recruited through posters in several treatment centers and online advertisements. Screening excluded those who were not in Canada or were not currently attempting to stop or reduce their substance use. Informed consent was obtained via an electronic consent form, after which they were provided a direct link to download Zamplo and begin their 6-week trial.

Procedure

This was a single-arm trial in which all participants were given the same mHealth intervention. Recruitment occurred from November 29, 2024, to January 17, 2025. After creating a Zamplo account, participants navigated through a text-based tutorial of the features available to them. The tutorial encouraged participants to explore all features and use the app in ways that felt most natural to them. They were further informed that they were not required to use the app to receive their monetary incentive, and all data they entered into the app will be kept entirely private. Participants were provided with an email address for further technical help, if needed. Overall, participants were left to independently engage with the app during their regular lives.

When participants reached the end of their 6-week trial, the app directed them to a final survey to provide information about their demographics and substance use history, rate the helpfulness of the app, and provide consent to be contacted for a follow up interview. They were then offered a CAD $20 (USD $14.76) electronic gift card for their participation. Those who did not submit a final survey after their trial period were considered to have dropped out of the study and were not included in the final analysis.

Data Analysis

Data were collected on the number of times participants opened the app, the duration of each app usage session, and the frequency with which participants used each individual feature. These data informed descriptive statistical analyses of engagement. Furthermore, participants were divided into categories based on demographics (gender, age, and ethnicity), recovery goals (abstinence or harm reduction), most problematic substance (alcohol, cannabis, or other illicit substances), and perceptions of Zamplo's helpfulness. Engagement was compared across groups.

Phase 3: User Experience Interviews

Overview

To gather further insight, we conducted follow-up interviews with a subset of Phase 2 app users, inquiring about their experiences with the app and suggestions for future designs. These qualitative interviews added nuance to quantitative app

XSL•FO
RenderX

usage metrics. While Phase 2 usage metrics provided valuable information on how frequently each feature was used, interviews shed light on *why* participants chose to use or not use particular app features, and other subjective experiences with the platform.

Participants

After participants completed their 6-week app trial, 42 consented to be recontacted for a follow-up interview. Researchers contacted 22 of these participants to schedule interviews. Others were not contacted due to suspected fraudulent participation, zero app usage during Phase 2, or not engaging meaningfully with the final survey, such as by skipping all questions. Of these 22 participants, 12 attended an interview. Informed consent was obtained via an electronic consent form, and CAD $20 (USD $14.76) electronic gift card incentives were offered to all interviewees.

Procedure

Interviews were 45-60 minutes in duration and conducted virtually by the same interviewer as in Phase 1. Interviews consisted of 3 main types of questions: the participant's history using apps in general during their recovery, their experiences with Zamplo during the app trial, and questions prompting the participant to brainstorm about "ideal" mHealth app functionalities to support their recovery. Both positive and negative feedback were encouraged.

Data Analysis

Interviews were recorded using Zoom, and transcribed using Rev AI transcription software. Each transcript was manually reviewed and corrected by the interviewer.

Transcripts were manually analyzed using the NVivo (Lumivero) platform, and coding was completed in 3 stages. The purpose of the first stage was to generate a codebook [39]. To do so, the interviewer conducted a preliminary round of coding. Coding was inductive [36], and after becoming familiar with transcript data and noting initial impressions, the coder generated original codes using an open-coding process [37]. Several low-level preliminary codes were created and clustered into a hierarchical structure. High-level categories were then refined to create a codebook based on collaborative discussions among the research team. A copy of the final codebook is available in Multimedia Appendix 1.

The initial coder and a second coder then used the codebook to code the full set of transcripts, making up the second stage of analysis. Both coders coded the same first transcript, then met to discuss coding disagreements and refined the codebook based on consensus. They repeated this process with a second transcript. After all remaining discrepancies in their interpretation of the codebook were resolved, each coder was randomly assigned 5 of the remaining transcripts to independently code using the codebook. Coding focused on capturing meaning as communicated by the participants (semantic coding), rather than attempting to identify implicit meaning behind participants' words (latent coding) [36,37].

Once all transcripts were coded using the codebook, the interviewer conducted a final thematic analysis of the coded transcripts, comprising the third stage of analysis. This final, high-level analysis generated meaningful themes with accompanying illustrative quotes to characterize how participants experienced Zamplo, and their values regarding mHealth apps for recovery.

Ethical Considerations

All phases of this work were approved by the University of Calgary Conjoint Faculties Research Ethics Board (REB24-0015). Informed consent was obtained via electronic consent forms for each phase. CAD $20 (USD $14.76) electronic gift cards were offered to all interviewees; participants were not required to meet any app engagement thresholds to receive the incentive. Participants were informed that Zamplo usage data were completely anonymous and confidential, and that survey data would be kept confidential with identifying information stored separately from deidentified survey data using anonymous participant identifiers.

Results

Phase 1: Co-Design Results

Overview

A total of 14 individuals participated in the initial co-design interviews. The mean interview duration was 58.9 minutes, with the longest interview lasting 107 minutes and the shortest lasting 47 minutes (13.7 hours in total). Themes from the initial interviews described several design implications for new Zamplo features. After iterative co-design discussions, 3 features were deemed feasible and worthwhile: a goal-setting feature, a craving tracker, and a meetings log. Features that were not included due to technical or resource constraints were a sobriety tracker and a mood tracker. Sobriety can instead be tracked within the goal-setting feature, and aspects of mental health can be tracked with existing Zamplo features upon user customization. The following sections detail the design and rationale of each new feature.

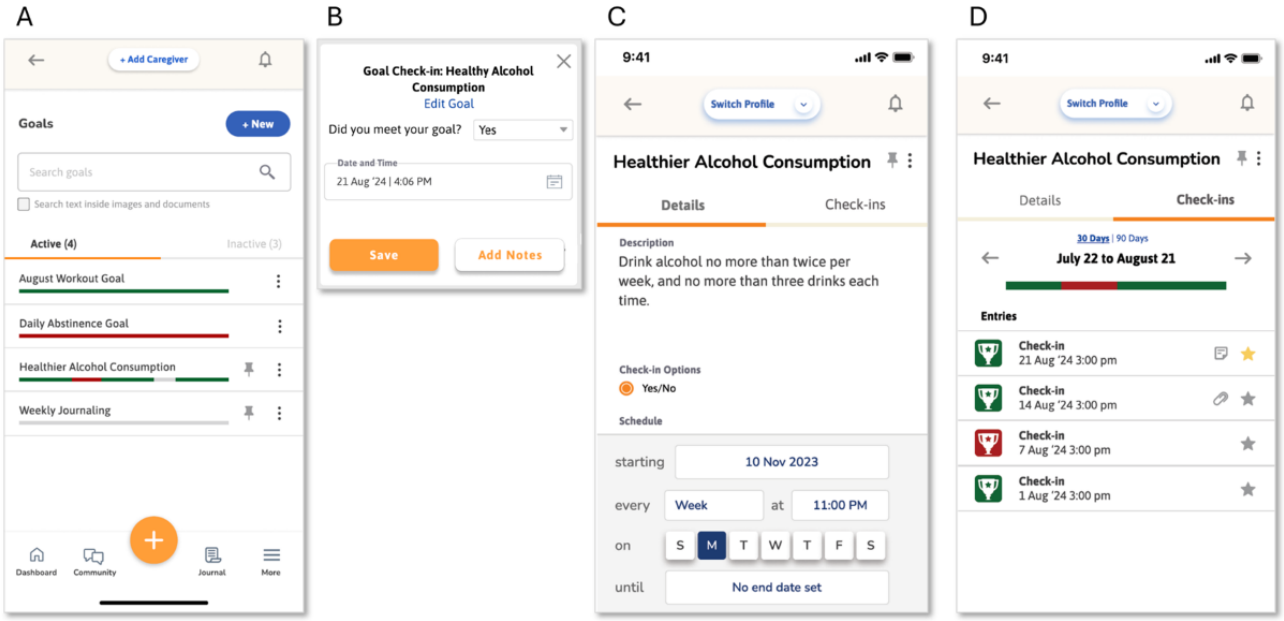
New Feature 1: Goal-Setting

During co-design interviews, participants emphasized that seeing ongoing progress of their recovery is crucial for their motivation. A visual anchor can keep them focused on and accountable to their recovery goals. However, they also described feeling frustrated and demoralized by existing sobriety tracker apps that reset and erase all previous data upon a relapse. This gave rise to the design idea of a nonconsecutive goal tracker that does not add more feelings of shame to relapses and allows imperfect adherence to goals. In contrast to sobriety trackers that only track consecutive sober days, this feature acknowledges missed days without resetting the tracker. Users can create any new goal to track, modify a goal at any time, create a flexible goal end date, and customize time intervals for regular "check-in" notifications (daily, weekly, or otherwise). At the user-specified check-in intervals for each goal, the feature generates a goal check-in notification, where users select "Yes," "No," or "Somewhat" in response to the question "Did you meet your goal?" A visual progress bar is gradually created over a 30- or 90-day timeframe for each goal, reflecting the history of the user's goal progress according to their check-ins. The progress bar displays both successful and unsuccessful days. Figure 1

XSL•FO
RenderX

shows final mock-ups of this feature. Importantly, this feature can be used to set and track goals outside of substance-use as well.

Figure 1. Goals feature mock-ups. (A) All active goals displayed with current check-in progress. Red bars indicate negative check-ins; green bars indicate positive check-ins. (B) An example of a scheduled goal check-in notification for one goal, asking if the user met their goal, with a selected option ("Yes") and a date and time. (C) Viewing or editing one particular goal, including the title, description, and customized check-in schedule. (D) Detailed history of goal check-ins for one particular goal.



New Feature 2: Craving Tracker

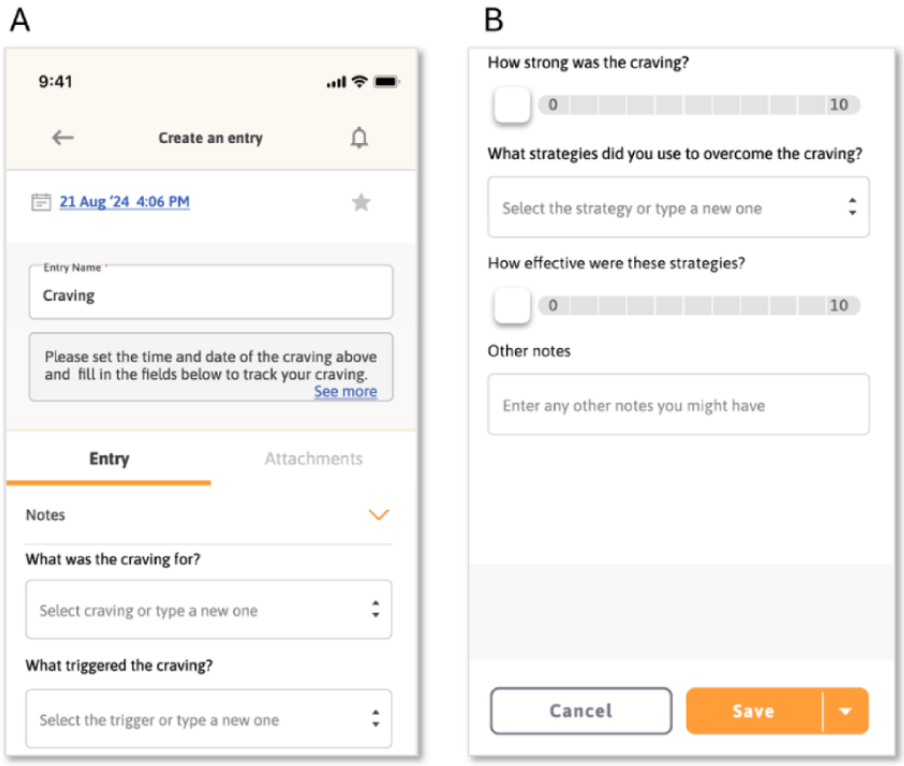
Interview analysis also revealed a desire to track cravings, along with triggers and solutions for each craving and additional health data. Participants explained that they would benefit from tracking cravings alongside several other metrics to understand potential reasons for the craving. For example, P008 described what they would like to see in a future iteration of Zamplo:

> *It'd be kind of cool to track when you're getting cravings, and then you can determine why you got that craving. Maybe because your physical activity was down, maybe because you stopped going to meetings, or maybe you're doing too many meetings, or too much physical activity or too much work. So, it'd be cool to track your balance that way so you can figure out where you go wrong.* [P008]

These types of participant sentiments led to the design of the craving tracker feature (Figure 2), which allows users to log details of each craving. The craving tracker is shown prominently on the main app dashboard for users to tap any time they experience a craving.

XSL•FO
RenderX

Figure 2. Craving tracker mock-ups. (A) The top of a new craving entry page. (B) The bottom of the entry page (scrolled down).



New Feature 3: Meetings Log

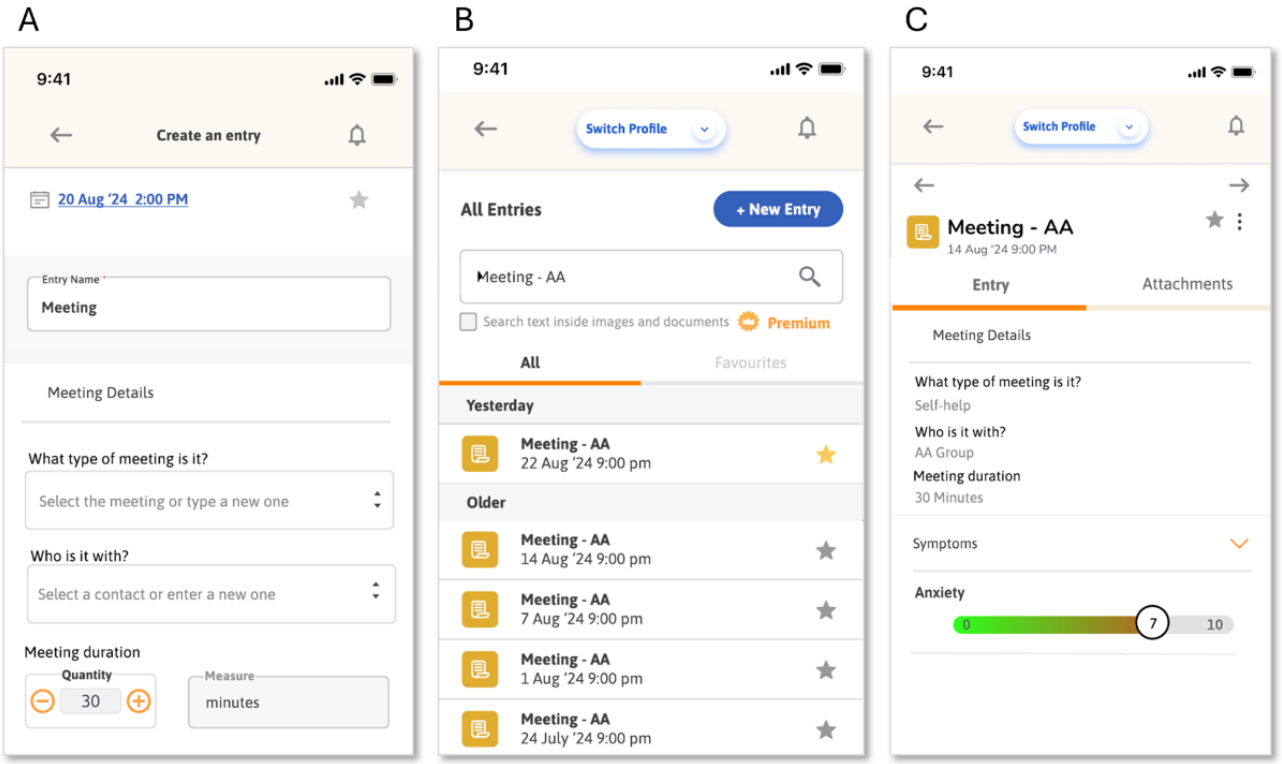
The final feature that arose from initial interviews was a meeting log. Participants described the importance of keeping track of meetings, appointments, and other treatment sessions. Participants not only needed reminders for meetings, which they can create using Zamplo's existing activity tracker, but they also wanted to log the details of each meeting to revisit in the future. P004 describes how tracking Alcoholics Anonymous meetings can boost their motivation:

> *If you documented your meetings list, you would see, 'wow, I went to a hundred meetings in the last 90 days,' or something like that. It shows how much time and effort you put in.* [P004]

Figure 3 depicts final mock-ups of the meetings log feature.

XSL•FO
RenderX

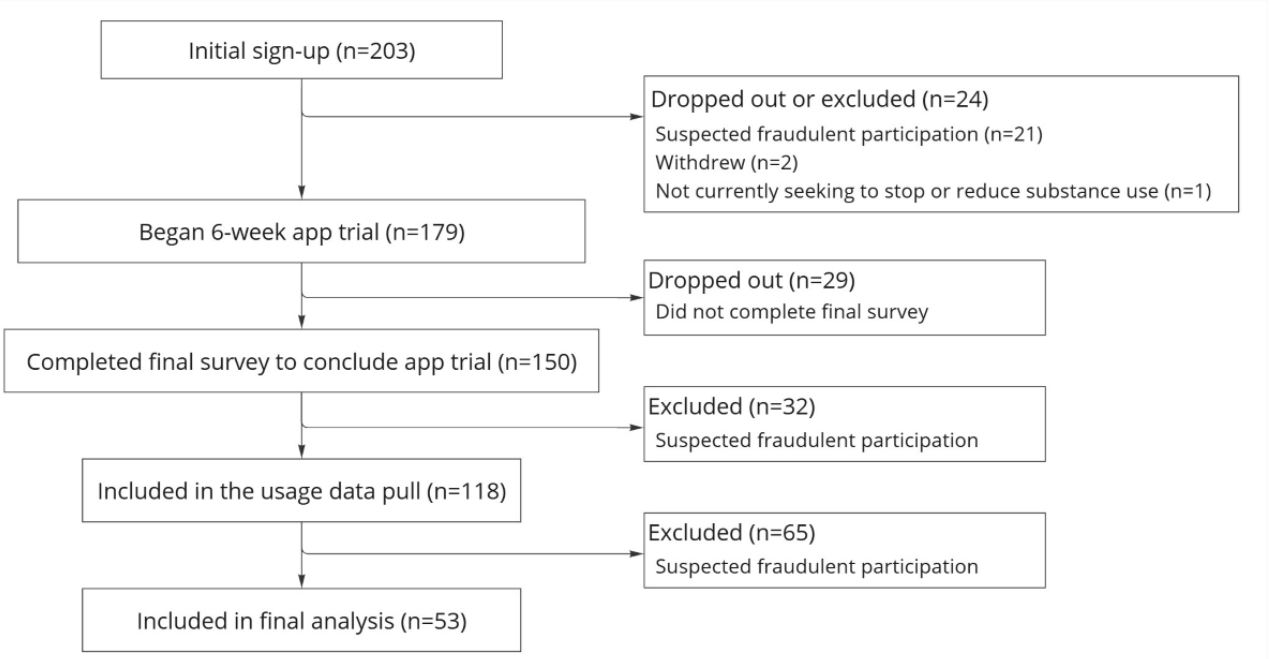
Figure 3. Meetings log mock-ups. (A) A portion of the new meeting entry page. Unshown fields include, for example, notes, mental and physical symptoms, and to-dos. (B) History of all past meeting entries. (C) Viewing the details of a previous meeting log.



Phase 2: App Usage Patterns During the 6-Week Trial

Figure 4 contains a CONSORT (Consolidated Standards of Reporting Trials) flow diagram of participation throughout this single-arm trial. Due to widespread online recruitment channels, fraudulent participation posed issues for this second research phase. Therefore, we established several authenticity screenings that effectively screened out fraudulent participation.

Figure 4. CONSORT (Consolidated Standards of Reporting Trials) diagram for Phase 2.



Of the 203 accounts initially created, 21 were suspected fraudulent due to circumventing the mandatory screening questionnaire meant to be completed before account creation. A further 32 participants were suspected fraudulent after completing their final survey after the trial, due to mismatched demographic responses when compared to initial screening, such as their age decreasing. The Zamplo usage data pull included 118 accounts. However, several accounts shared the same device address, indicating a possibility that each account did not represent a unique person. Accounts were designated

XSL•FO
RenderX

authentic if they did not share an address with other accounts. As shown in Figure 4, a total of 65 potentially fraudulent accounts were detected using this approach. The final sample size of authentic participants was 53.

App users were divided into several subcategories based on demographics, recovery approach, most problematic substance, and ratings of the app's helpfulness. These groups provide important context for design, as they demonstrate which sectors of the community found the app most useful. The final numbers of authentic users in each group are shown in Table 1.

Table 1. App user groups, with numbers of authentic and active users.

Group	Authentic users, n/N (%)	Active users[a], n/N (%)
All users	53/53 (100)	24/53 (45.3)
Gender		
Women	29/53 (54.7)	15/29 (51.7)
Men	23/53 (43.4)	8/23 (34.8)
Recovery goal		
Abstinence recovery goals	36/53 (67.9)	14/36 (38.9)
Harm reduction recovery goals	17/53 (32.1)	10/17 (58.8)
Perceived app helpfulness		
Somewhat to very helpful	35/53 (66)	13/35 (37.1)
Somewhat to very unhelpful	12/53 (22.6)	9/12 (75)
Most problematic substance		
Alcohol is most problematic substance	18/53 (34)	9/18 (50)
Cannabis is most problematic substance	14/53 (26.4)	5/14 (35.7)
Other most problematic substance	21/53 (39.6)	10/21 (47.6)
Age group (years)		
18-29	14/53 (26.4)	5/14 (35.7)
30-39	21/53 (39.6)	14/21 (66.7)
40-59	11/53 (20.8)	4/11 (36.4)
Race and ethnicity		
White	32/53 (60.4)	14/32 (43.8)
BIPOC[b]	18/53 (34.0)	7/18 (38.9)

[a]A user was considered "active" if they opened the app on 2 or more separate days, excluding the first and last days of the study period.

[b]Black, Indigenous, and People of Color.

The designation of an *active* user was created to ensure that usage data were analyzed from those who meaningfully interacted with the app. An active user was defined as someone who used the app during at least 2 separate 24-hour intervals. Usage in the first and last 24-hour periods of a person's trial did not count toward the active user designation, since those intervals included activity required for minimal study compliance. Specifically, the first 24 hours involved downloading the app to begin the trial, and the last involved following a link to conclude the trial. The number of active users in each demographic group are shown in Table 1.

While many of the group sizes in Table 1 are small, they allow for preliminary comparisons between groups. A greater proportion of women actively used the app (15/29, 51.7%) compared to men (8/23, 34.8%). Overall, the greatest number of active users were women (n=15), followed by those aged 30-39 years (n=14), those who identified as White (n=14), and those who aimed to fully abstain from substance use (n=14). However, a higher proportion of users taking a harm reduction approach to their recovery actively engaged with the app (10/17, 58.8%) compared to those taking an abstinence-based approach (14/36, 38.9%).

Table 2 depicts how often the active users in each participant group opened the app over the 42-day trial period. A "session" was logged by Zamplo each time a user opens the app for any reason, either to enter data or simply to view information or explore the interface. Table 2 includes the mean (SD), median (IQR), and top 25th percentile number of sessions for each group.

XSL•FO
RenderX

Table 2. Number of sessions among active users.

Group	Active users, n (%)	Number of sessions over 6-week app trial		
		Mean (SD)	Median (IQR)	Top 25th percentile
All users	24 (100)	27.1 (47.5)	10 (7-18)	18
Gender				
Women	15 (62.5)	37 (57.7)	10 (6-47)	47
Men	8 (33.3)	10.8 (4.5)	11 (7-16)	16
Recovery goal				
Abstinence recovery goals	14 (58.3)	13.5 (10.7)	10 (6-16)	16
Harm reduction recovery goals	10 (41.7)	46.2 (68.0)	14 (7-54)	54
Perceived app helpfulness				
Somewhat to very helpful	13 (54.2)	32.2 (57.4)	10 (6-18)	18
Somewhat to very unhelpful	9 (37.5)	24.2 (33.3)	15 (7-18)	18
Most problematic substance				
Alcohol is most problematic substance	9 (37.5)	34 (67.4)	9 (6-14)	14
Cannabis is most problematic substance	5 (20.8)	11 (3.7)	10 (9-11)	11
Other most problematic substance	10 (41.7)	29 (33.7)	16 (6-47)	47
Age group (years)				
18-29	5 (20.8)	9.8 (2.9)	10 (9-11)	11
30-39	14 (58.3)	38.1 (59.6)	10 (7-47)	47
40-59	4 (16.7)	12.5 (7.7)	15 (6-24)	24
Race and ethnicity				
White	14 (58.3)	36.4 (59.3)	15 (6-24)	24
BIPOC[a]	7 (29.2)	16.1 (16.0)	9 (6-18)	18

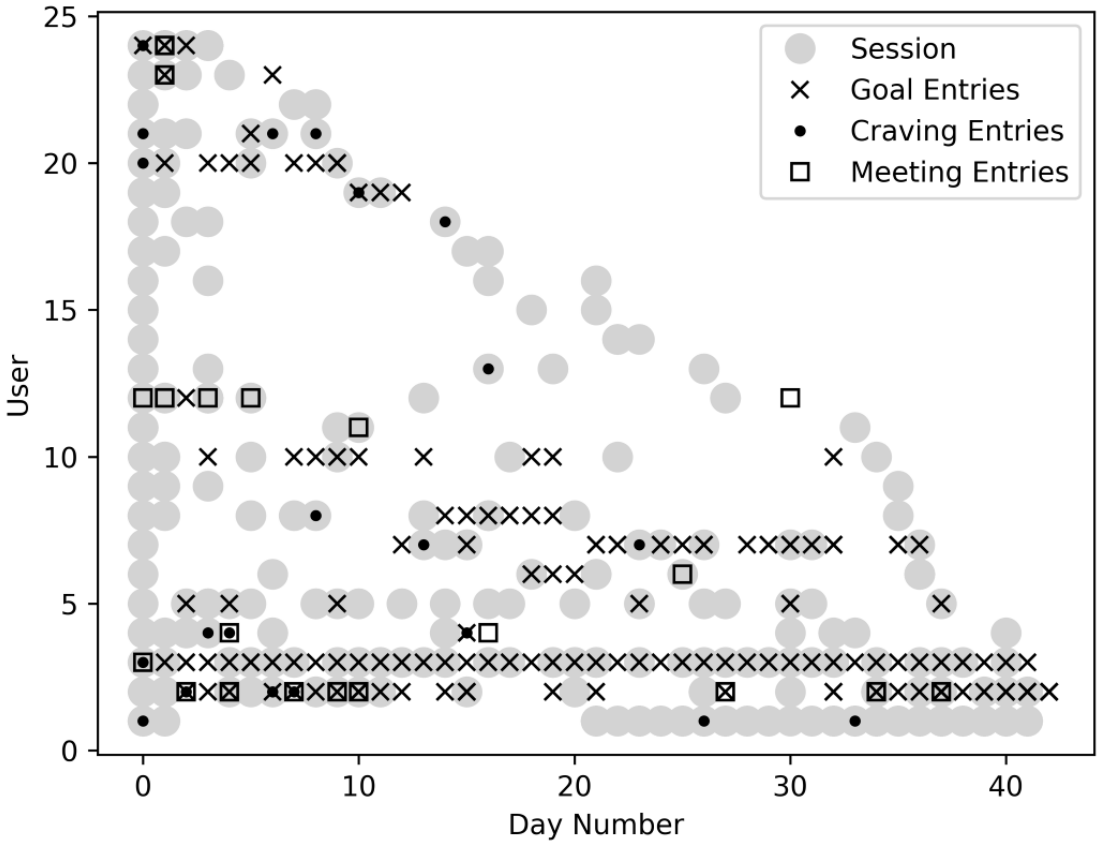
[a]Black, Indigenous, and People of Color.

While much variation can be seen in the mean and 25th percentile values, the median number of sessions is relatively stable for most groups. This suggests that a small portion of users had a particularly large number of sessions that inflated the mean and percentile values.

Figure 5 shows the temporal pattern of engagement with the app in general (sessions), as well as engagement with each of the 3 new features over the 42-day app trial. The x-axis represents the day number relative to each person's first day in the trial, and the y-axis represents user indices. Note that the user indices in Figure 5 are randomly assigned to the data in Figure 5 only and do not correspond to the participant IDs used when reporting interview data.

XSL•FO
RenderX

Figure 5. Each day in the trial when active users opened the app (sessions) and made new entries into each of the new co-designed features.



To better observe trends, users are ordered in Figure 5 according to the last day they used the app. When creating a goal, craving, or meeting entry, the user had the option to manually adjust the date that the goal was met, cravings were experienced, or the meeting occurred. Therefore, the goal, craving, and meeting data points shown in Figure 5 are based on the dates manually entered by the user, rather than the dates that the user opened the app to enter the data. Furthermore, sessions (gray circles) reflect the exact dates that each user opened the app. During each session, users may have engaged with the myriad of preexisting Zamplo features, such as activity tracking, medication tracking, creating and viewing charts of personal data, and so on.

Figure 5 depicts the relative popularity of the Zamplo features co-designed during this study compared to users' engagement with the rest of the Zamplo app. For example, users 15, 16, and 17 logged sessions of app usage but did not use the goals, cravings, or meeting features. In contrast, user 3 made nearly daily use of the app and the goal feature specifically. When viewing the full figure, it is clear that most users chose to interact with the newly designed goals, cravings, and meetings features. However, their use of these features varied widely, with some patterns of almost daily use and other patterns of intermittent and infrequent use.

The new goals feature was the most popular. More granular analysis of usage data revealed that, of the 24 active users, 17 created a single goal to track, 3 created multiple goals, and 4 did not create a goal. When users entered data into the app via the new co-designed features, 58.3% (14/24) made at least 1 goal entry, 50% (12/24) made at least 1 craving entry, and 33.3% (8/24) made at least 1 meeting entry.

A common metric for understanding app engagement is the 30-day retention rate, defined as the percentage of users still using the app 30 days after installing it. If all authentic users are included in this metric, 20.8% (11/53) of authentic users were retained after 30 days. This metric can also be calculated using only the active participants who made some effort to use the app, since the others may never have intended to use the app beyond minimal research compliance. In this case, 45.8% (11/24) of active participants were retained after 30 days.

Phase 3: User Experience Interview Findings

Overview

Qualitative interviews contextualized the above quantitative findings by revealing details of how participants experienced Zamplo. A total of 12 individuals from the app trial attended a one-on-one interview with a researcher. The mean duration of interviews was 47.6 minutes, ranging from 38 to 59 minutes (9.5 total hours). Interview participants had a mean age of 32.9 (SD 7.38) years. Participant demographics are shown in Table 3.

XSL•FO
RenderX

Table 3. Characteristics of interview participants.

Measure	Value, n (%)
Demographics	
Gender	
Women	7 (58.3)
Men	5 (41.7)
Age group (years)	
18-29	4 (33.3)
30-39	6 (50)
40-59	2 (16.7)
Ethnicity	
White	7 (58.3)
BIPOC[a]	5 (41.7)
Substance use disorder history	
Most problematic substance	
Alcohol	4 (33.3)
Prescription medications	4 (33.3)
Cannabis	2 (16.7)
Methamphetamine	1 (8.3)
Heroin, fentanyl, or other opioid	1 (8.3)
Recovery approach	
Aim to reduce substance use (harm reduction)	7 (58.3)
Aim to stop substance use (abstinence)	5 (41.7)
Time since taking first steps to address substance use disorder	
3-7 months	3 (25)
1-2 years	3 (25)
6-10 years	3 (25)
Unknown	3 (25)
Perceived app helpfulness	
Somewhat to very helpful	9 (75)
Somewhat to very unhelpful	3 (25)

[a]BIPOC: Black, Indigenous, and People of Color.

During interviews, participants described their experiences with Zamplo and other digital recovery tools, and identified their needs for future mHealth apps based on their reflections from this study. Below are findings from across these domains, focusing first on results from preliminary codebook analysis and then on 3 emergent themes from the final thematic analysis.

Codebook analysis highlighted several app features that participants repeatedly spoke highly of, prompting ideas for future mHealth apps. Favorable features included goal-setting features, life-organization features (eg, appointments, medication reminders, and to-do lists), and self-tracking and visualization features. Interviewees described beneficial mental and behavioral outcomes of interacting with such features, which included the following, in descending order of salience in the interview data: increased motivation to self-improve, feelings of accountability, greater focus on healthy habits and distraction from unhealthy habits, increased self-awareness, a sense of self-empowerment, and increased self-compassion. Functionality that was not configured in Zamplo during this study but that participants enjoyed using in other apps included gamification features (eg, rewards for progress or accumulating "points"), information and resources (eg, personally-targeted mental health tips and other educational information), community and social features (eg, group forums, chats, webinars, and media-sharing), and inspirational content (eg, regular personal affirmations and motivational quotes).

Despite the positive outcomes described by interviewees, quantitative usage metrics painted a diverse picture of engagement. A closer thematic analysis of interview data adds nuance to participants' experiences, helping contextualize our

XSL•FO
RenderX

mixed usage metrics by highlighting what participants struggled with, what they would appreciate in future tools, and potential design challenges when creating an engaging mHealth app for recovery. The following sections describe three key themes: (1) a digital journal as a touchpoint to greater recovery resources, (2) a twofold experience of Zamplo's open-ended configuration, and (3) expanding beyond substance use in recovery apps.

A Digital Journal as a Touchpoint to Greater Recovery Communities

Upon reflecting on their experiences during this app trial, participants tended to describe their needs along three categories of features, (1) personal features—journaling tools for individual self-monitoring, (2) relational features—active social spaces that enable interaction with surrounding recovery communities, and (3) informational features—frequently updating content for users to feel connected and informed about current conversations in recovery. The ideal app would bridge functionality between these 3 realms, not only providing a space for private reflection but also serving as a conduit to wider recovery ecosystems.

The first category (personal features) included goal-setting, tracking, and charting functionalities that allow participants to gain awareness and take ownership of their recovery. Such features were already configured in Zamplo during this study. The second category (relational features) included several tools that participants appreciated in other apps and would like to see integrated into a single recovery app, alongside the aforementioned personal features. These features included social networking and community spaces that allow users to share with and learn from others in recovery who are experiencing common struggles. For example, P071 requested organized peer support groups to be built into an app like Zamplo:

> *I think this is actually very important. I would actually want to have this connection with peer support groups and online community. I think that would actually be very helpful to have this online community and support groups actually around in the app.* [P071]

P032 similarly explained that their ideal recovery app would include the following:

> *…the possibility of chatting with other people experiencing the same symptoms or the same issues. Because, sometimes, it enhances the motivation, and you learn from other people's experiences.* [P032]

For relational features like these to function as participants have described, the platform would require a significant number of users who face similar challenges to join the platform so that they can begin to connect with one another. The final category (informational features) encompassed resources such as frequently-updated news postings about substance use and recovery, "tips and tricks" articles to support recovery and mental health, educational resources about substance use, inspirational stories or quotes, and personalized reading material. This type of content would provide users a sense of connectivity to the wider world of recovery. For example, P103 explained that it would be important for the app to frequently refresh with new recovery-focused content to serve as a positive pastime and increase app engagement:

> *I think I would have less anxiety, and I'd be happy to go into the app all the time, and just seeing what's new. Seeing if they added anything, or if there's any new articles or stuff to help you. Just maybe something new every day. And it's just, 'oh, look at this,' or, yeah. It just makes you happy.* [P103]

Zamplo, as it was configured during this study, did not provide this sense of connectivity with others in recovery nor did it offer personally-targeted, timely content to consume. It instead focused on providing a private digital space for participants to manage their personal recovery journeys. The minimal connection with the wider recovery world may have contributed to the mixed levels of engagement seen in quantitative usage metrics.

A Twofold Experience of Zamplo's Open-Ended Configuration

Zamplo was configured for this study with an open-ended architecture that allowed for a self-guided user experience. As such, it did not enforce that users track any specific metrics, nor did it enforce any specific types of goals to be set, or nudge user interaction according to any preset schedule. Instead, it provided participants with the means to track whichever items, set whichever goals, and schedule whichever reminders they desired. Participants were informed upon signup that they were not required to interact with Zamplo in any particular way for this study and could use the app as much or as little as they wished. This not only allowed for more naturalistic observations of usage behavior, but also surfaced a tension in participants' values regarding flexibility and control.

Zamplo's flexibility gave many participants a sense of control over their own experience. For example, P054 explained, "I liked that it was very much 'I took control of it,' and I set my own goals, and there were lots of options and reminders, if I wanted to." Similarly, P096 contrasted the flexibility of Zamplo with the rigidity of other recovery apps. Other apps often limited them to only tracking one recovery-specific goal, such as abstinence, triggering unhealthy responses when the goal is met.

> *The one [app] that was tracking my sober time, I would get too excited when it got to the month, and I'm like, "okay, let's go party." ...In this one [Zamplo], I like how I could add things and take away things. Because the other one, for me, it was just one thing, and once I reached that [goal], I felt like I had to start fresh. And I didn't see my progress for the other things, so it's like, oh, well, let's go celebrate this one. And that celebration was not always a great celebration.* [P096]

Only allowing the user to perceive their progress according to one dimension of their recovery was detrimental to P096's recovery. In contrast, Zamplo gave P096 the freedom to adapt their interface so it included a wide range of personalized goals to track.

P032 similarly described Zamplo's open-ended design favorably, explaining that they could engage with the app "at [their own] pace." P032 contrasted Zamplo with other apps that used a more strict, prescriptive approach, making them feel "pressured" if

XSL•FO
RenderX

they did not adhere to the engagement behavior enforced by the app, such as preset daily tasks to complete. Participants who appreciated the freedom to build their own experiences in Zamplo tended to describe benefits such as increased accountability, enhanced motivation, a sense of empowerment, and heightened self-awareness.

This open-ended quality of Zamplo, however, also spurred negative user experiences. Some participants described feeling overwhelmed or confused by the interface, as it did not provide the direction they would have liked. P019 described the user experience as a "free-for-all":

> *I feel like usually apps will kind of lead you through things in a certain way, and I feel like [Zamplo] was just a free-for-all. And so then it kind of just made me be like, "okay, too much going on." And I really didn't use, I don't feel like, a lot of it.* [P019]

While Zamplo's user-guided configuration gave some participants a sense of agency over their own recovery, this excerpt emphasizes that others felt burdened by the same functionality. Participants who did not appreciate Zamplo's open-ended quality often preferred a preprogrammed, prescriptive user experience.

> *It could have been easier to be like, okay, maybe "What's your goal with the app?" or "What's your goal in the next month?" Or whatever. You [the user] could say "reduce using a substance," or something. And then you could choose from a list it could prepare. ...And then it could show you, "okay, go reduce your drinks by this many," or whatever. And then it would establish some tracking starter points based on that.* [P034]

Here, P034 has ideated a user workflow that offers suggestions for recovery strategies based on personalized questions, rather than fulfilling only a passive journal-like role, as Zamplo did.

P103 additionally recalled negative experiences attempting to understand the purpose of the app: "I didn't know, like, what the meaning of the app was. Without a tutorial or, like, what am I supposed to do?" Participants did receive a step-by-step tutorial when creating an account, which included a walkthrough of each feature available to them. Still, some participants were unsure how to begin their app experience without prescribed items to track or a preset workflow to adhere to. This suggests that either the tutorial did not meet participants' needs or that they may not have engaged with it fully, possibly due to differing expectations of what a "recovery app" should provide.

A closer look at P019's interview offers an example of this mismatch in expectations. They expected a more prescriptive app design and described their experiences with a local treatment center as context for how they would expect a recovery app to function.

> *I just think that the app should guide the person, more than the person going in and really guiding it themselves. ...When I was in [anon. treatment program], they do try to keep things very basic. It's like, they structure everything for you so that you can just focus on your recovery.* [P019]

Here, P019 explained that they would prefer an app that provides a close equivalent to a formal, in-person treatment program, structuring the user experience with a top-down plan. This contrasts others' preferences for an app that allows them to set their own usage path with a multitude of options to complement in-person recovery rather than emulate it.

In summary, a tension arose from the open-ended configuration of Zamplo. The flexibility and nonprescriptive nature of Zamplo allowed some participants to feel in control of their own recovery. This led to fruitful interactions with the app's tracking and goal features, which gave rise to feelings of accountability, motivation, empowerment, and self-awareness. Meanwhile, others felt burdened by the onus of creating their own recovery regimen in the app, causing overwhelm and confusion, which may have played a role in their disengagement.

Expanding Beyond Substance Use in Recovery Apps

Finally, while the previous theme pertained to the back-end configuration of Zamplo, this theme describes the front-end subject matter available within the app. Most participants appreciated the potential to use Zamplo for purposes outside of recovery. As P096 described, Zamplo encompassed all aspects of a person's recovery journey.

> *I'm not just addiction. I'm not just mental health. I'm not just physical health. ...I am all of them, and they all impact me. Being hungry can impact my mental health, which could impact my triggers to wanting to use. It's all into one. ...This app is everything. It's not just recovery, it's not just mental health. This is a bit of everything.* [P096]

Furthermore, P054 used the goals feature to track goals not directly related to recovery, such as reading goals. They also used the goals feature to manage daily to-do items:

> *I just found it was just really easy to use and keep track of my goals, my expectations. I ended up using it for a lot of things, basically as my to-do list, and I really liked it.* [P054]

Notably, they still considered their reading goals and to-do items under the purview of "substance use" goals because they view their recovery as a holistic trajectory toward healthier daily habits. Similarly, P071 described using the fitness tracking features within Zamplo as their primary use case throughout the study, which they still considered part of their substance-use recovery regimen. The below excerpt exemplifies how this holistic app design led to a variety of tracking possibilities that all ultimately supported participants' recovery:

> *I think playing the guitar, and gratitude journaling, fitness, using medication, I think these are the things I set aside for myself in the application [Zamplo] to carry out on a daily basis. And these are the things that actually help me to avoid the usage of substances.* [P127]

However, there was a trade-off to incorporating such multifunctional, holistic tracking functionality in Zamplo. This led to an overabundance of features on the interface, each with

XSL•FO
RenderX

many potential use cases, again causing overwhelm for some participants.

> *Because I can get overwhelmed easily, and I don't like having too many options because I'm extremely indecisive, this was just way too busy for me. ...Honestly, I tried to not come into the app as frequently.* [P019]

Notably, though, those who disengaged due to feelings of overwhelm still described a preference for holistic app designs with a wide range of functionalities, rather than apps that narrowly target substance use behaviors. This preference came through in their descriptions of other apps that they have previously enjoyed and their vision of an "ideal" recovery app. Many explained that tools to assist with basic life maintenance, such as establishing daily routines or celebrating "small wins," like doing household chores, are all crucial for their recovery. However, it is essential that the digital platform provides enough clear guidance to help them navigate through the potentially overwhelming quantity of available features.

Discussion

Overview

This study resulted in the design and deployment of 3 new mHealth app features: a goal-setting feature, a craving tracker, and a meetings log. Usage analyses and interviews foregrounded quantitative and qualitative findings regarding how participants engaged with and perceived these new features and the app in general. Overall, 45.3% (24/53) of participants opened Zamplo on at least 2 separate occasions other than the required first and last days of the trial period, deeming them "active" users. These 24 active users opened Zamplo 27.1 times on average over the 42-day trial, and 45.8% (11/24) of them were still using Zamplo after 30 days.

Interviews highlighted that participants desired tools that not only act as private digital recovery journals, but also as gateways to connect with wider recovery communities and as sources of engaging and educational recovery content. A design tension also arose between prescriptive user workflows, where the app guides the user through a structured recovery plan, and open-ended platforms, where users set their own recovery regimens. Lastly, interviewees largely appreciated when apps not only focused on recovery but also included features to support them in other dimensions of their lives.

Scheibein et al [40] proposed a roadmap to achieve optimal use of digital tools in the substance use field by 2030. They identified 5 key values based on a large backcasting exercise: digital rights, evidence-based tools, user-friendliness, access or availability, and person-centeredness. They additionally warned that current digital recovery tools are heavily informed by the interests of multinational, for-profit companies, with minimal involvement from key stakeholders such as people actively struggling with substance use disorders [40]. This study has described 3 phases of a co-design and deployment process, answering this call to include end users' voices in technology design. Analysis led to several methodological insights and design implications that help to integrate the user-friendliness and person-centeredness values into recovery app development.

Reflections on Methodology

This research exemplified an end-to-end app development and testing cycle, offering a real-world case study of iterative co-design and deployment with those in recovery. As researchers strive to optimize recovery app engagement, we urge that they incorporate co-design into their methods. Valuable insights were gained during each of the 3 co-design research phases in this study. Phase 1 provided actionable takeaways for app development based on what those with recovery experience described wanting in their recovery platforms, leading to a concrete starting point for new features. In Phase 2, we gathered real-world engagement based on the features from Phase 1, revealing complex usage patterns as some users engaged and others disengaged from the app. Usage analysis then gave way to Phase 3 interviews, where nuanced design tensions came forward during one-on-one conversations with participants who could now draw upon their hands-on experience with the app.

Each phase was critical for the next phase, and they all contributed to understanding participants' perspectives on how to increase app engagement and support this population. These phases represent the beginning of a deeply iterative co-design process in which practitioners investigate end users' needs, translate their ideas into concrete tools deployed for usage, and reflect with users on the real embodiments of their original ideas. This last stage can then prompt further dialogue about future iterative designs to continue aligning more closely with users' needs.

Given the relatively low extant evidence of uptake and engagement for recovery apps, amplifying the perspectives of people with lived experience may be key to understanding why they may or may not engage with these apps. Although not all participants engaged deeply with the app during this study, still 20.8% (11/53) of all users were retained after 30 days, and 45.8% (11/24) of active users—those who went beyond minimal compliance for this study—remained engaged after 30 days. By industry standards, the typical 30-day retention rate in 2024 was 7% for all mobile apps [41] and approximately 4% for digital health apps [42]. The substantially higher retention rate in this study could be explained by participants being incentivized to take part. However, we mitigated this bias by repeatedly stating that participants are not required to interact with the app during the study to receive their incentive and are instead encouraged to use it however frequently feels helpful for them. These retention figures suggest that Zamplo features can be taken as a promising starting point for future recovery app designs. Additionally, the co-design approach may have contributed to engagement, though further controlled studies are needed to confirm this effect. This finding parallels prior literature concluding that incorporating end users' perspectives into app design can increase uptake and engagement [12]. Importantly, though, a portion of participants still disengaged from the app, necessitating further discussion of design tensions, challenges, and opportunities for future platforms.

XSL•FO
RenderX

Implications for Design

This research surfaced several design trade-offs that are crucial to consider when designing engaging mHealth apps for recovery. Interviews revealed a clear split in positive and negative experiences with Zamplo: positive experiences arose from feelings of agency and autonomy over one's own recovery, while negative experiences manifested in feelings of overwhelm and confusion. Both types of experiences resulted from the same open-ended, holistic, and flexible nature of Zamplo. As per prior research, mHealth interventions are not likely to be "one size fits all" [7]; users' needs often vary, and rigid designs are not likely to appeal to all members of the heterogeneous group of individuals in recovery. This study sheds more light on this diversity of experience and urges future designers to strike a balance between system-controlled versus user-controlled design approaches.

This tension echoes a longstanding debate in user experience research regarding the extent to which the system should impose predetermined, stepwise flows onto users, versus letting users freely set their own app workflows. While this tension is well explored in general user experience research and persuasive design domains [43], it has not been explicitly foregrounded in substance use recovery technology research, as this study has done. Indeed, this debate is particularly interesting within the recovery domain, where self-tracking technologies are often packaged as prescribed programs with fixed tracking metrics, similar to how counselors provide trusted top-down structures and routines for patients to follow. Preliminary studies in the field of human–computer interaction, however, show that app users in recovery may in fact appreciate more holistic and flexible self-tracking [25,29].

This study extends this line of inquiry by empirically demonstrating this trade-off as it relates to the recovery domain. One possible way to reconcile this tension is to design recovery apps that include both a ground-level "starter kit" with recommended features and evidence-based usage routines, while also offering additional, optional, and customizable features built on top of starter features. Users like P019, who feel overwhelmed by too much freedom, could benefit from a set of default features that reduce cognitive load and help them jumpstart their digital recovery journey. These default features could operate as a digital analogue to structured treatment programs with planned recovery pathways. For example, they could include preset goals with rigid tracking schedules, preset daily self-monitoring routines including prompts for regular check-ins and written reflections, and specific personal metrics to track on a set schedule, such as substance use quantity per day or recurring triggers. This design would need to explicitly guide the user through the interface, instructing them on how to use each feature to support their recovery, similar to a counselor's prescriptive guidance. This may be most useful for those who are new to using digital recovery apps.

Then, users like P032, who are confident to carve their own recovery pathways in a digital platform, could personalize the app to step outside the prescribed usage flows and ensure they feel in control of their recovery. This optional extended functionality could offer, for example, highly flexible goal setting with customizable goal timelines, monitoring of any chosen personal metrics, and user-determined reminders and check-in schedules to enhance self-reflection at a pace the user is comfortable with. It would be crucial for these additional features to allow tracking outside of substance-use itself to align with the present findings and the broader literature suggesting a holistic approach to recovery [29,44-46].

The distinct split between positive and negative experiences with Zamplo additionally raises questions about the characteristics of the individuals in each category. Milward et al [18] developed typologies of user engagement with harm reduction apps. They identified "trackers" who liked to track several metrics for the sole purpose of self-awareness; "cut-downers," who focused on finding a solution to their substance use and made more use of goal-setting features than self-tracking features; and "noncommitters," whose initial enthusiasm about the app quickly waned and led to disengagement.

This study found similar patterns, as illustrated in Figure 5. Several participants used Zamplo's existing tracking features but did not engage heavily with the goal-setting feature (trackers); a portion of participants displayed particularly high usage of the app in general, with special attention paid to the goal-tracking feature (cut-downers); and a portion of participants lost interest in the app after signing up for the study and never became an active user or they only engaged during the beginning of their trial period (noncommitters).

This study extends this characterization by placing more onus on designers rather than users, and suggesting that the noncommitters category does not reflect a homogenous group of people who are generally disinterested in trying to use recovery apps. We propose that this typology can be split into 2 subcategories based on the user's compatibility with the app design. One subcategory is reserved for people who are, indeed, not committed to using apps for recovery. We call this subcategory "true noncommitters." We refer to the second subcategory as "unintentional noncommitters." This group includes those who are indeed committed to finding digital recovery solutions but may lose interest in particular platforms due to ineffective designs that do not align with their nuanced recovery needs, as unpacked during this research.

Limitations and Future Work

Using existing Zamplo architecture provided us with the resources to conduct an extensive real-world app deployment study. It also provided insight into how recovery-focused features can integrate into an established mHealth platform and coexist with features not directly related to recovery. However, the preexisting Zamplo features were not the result of our co-design with end users, yet were still included in participants' experiences. This limits the internal validity of our evaluation, specifically of the co-designed features, because user experiences with those features may have been confounded by interactions with the preexisting functionalities of the platform. Additionally, parts of what participants in co-design sessions requested were not able to be fully embodied by the new co-designed features due to resource constraints and because the features had to integrate into existing Zamplo architecture.

XSL•FO
RenderX

Fraudulent participation also posed issues for this study, decreasing sample sizes in the later phases of the work. These sample sizes reduced the transferability of these findings. However, this study still highlighted several noteworthy emergent usage patterns and design tensions, despite small sample sizes. Future larger-scale research endeavors with strategies to eliminate fraudulent participation are necessary to confirm the salience of our findings.

A key finding of this study was the diversity of experience and attitudes toward using the app, as highlighted in the results of Phase 3. The small sample size of Phase 3 prevented analysis that would determine if demographics (eg, age and gender), or clinical factors (eg, recovery stage) can be linked to values and attitudes toward using a recovery app (eg, valuing autonomy versus a more prescribed user experience). Future larger-scale studies should explore such correlations in more detail. This study did reveal a subset of users inclined to feel overwhelmed by all of the available features in the Zamplo platform. Therefore, future studies using general-purpose health platforms for addictions recovery could also explore mitigating overwhelm via providing extra tutorial material.

Finally, it is important to note that this study's prime focus was capturing realistic measures of engagement with an mHealth app for recovery. However, prior work has brought into question the engagement-efficacy perspective, which assumes that higher engagement with health technology is a precursor to long term efficacy [17]. Instead, Smith et al [17] identified "productive disengagement," a phenomenon where disengaging with a health app can actually help the user in achieving their health goals. This study was unable to pinpoint every participant's exact reasons for disengagement, and instead characterized any disengagement as a negative result. As future research works to optimize engagement with recovery apps, it is important to acknowledge the engagement–efficacy assumption so that higher engagement is not always conflated with higher efficacy. Future work may benefit from investigating further into specific users who disengage with recovery apps to understand the details of their disengagement before necessarily classifying it as a negative result.

Conclusions

This study presents a detailed account of incorporating end users' design ideas into new app features to investigate engagement. We collaborated with the mHealth platform, Zamplo, to deploy 3 new app features: a goal tracker, a craving tracker, and a meeting tracker. A 6-week app trial revealed mixed engagement with the app, with 30-day retention rates substantially higher than industry benchmarks. Qualitative analysis foregrounded several mHealth features that participants perceived as valuable for their recovery. Difficult design tensions also came forward, such as balancing system control with user autonomy while mitigating cognitive load for users, as well as integrating private journal-like features into apps that also connect users to wider recovery ecosystems. This study offers early insight into the value of participatory design methods with real-world app deployment for understanding digital recovery engagement, highlighting directions for future mHealth research and design.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Qualitative codebook used by two independent coders to analyze user experience interviews.
[DOCX File , 28 KB - mhealth_v14i1e83984_app1.docx]

References

1. Liu S, Ma J, Sun M, Zhang C, Gao Y, Xu J. Mapping the landscape of digital health intervention strategies: 25-year synthesis. J Med Internet Res 2025;27:e59027 [FREE Full text] [doi: 10.2196/59027] [Medline: 39804697]
2. Zemore SE, Ziemer KL, Gilbert PA, Karno MP, Kaskutas LA. Understanding the shared meaning of recovery from substance use disorders: new findings from the what is recovery? Study. Subst Abuse 2023;17:11782218231199372 [FREE Full text] [doi: 10.1177/11782218231199372] [Medline: 37731748]
3. Brophy H, Dyson M, Katherine R. Concept analysis of recovery from substance use. Int J Ment Health Nurs 2023;32(1):117-127. [doi: 10.1111/inm.13066] [Medline: 36117317]
4. Crane D, Garnett C, Brown J, West R, Michie S. Behavior change techniques in popular alcohol reduction apps: content analysis. J Med Internet Res 2015;17(5):e118 [FREE Full text] [doi: 10.2196/jmir.4060] [Medline: 25977135]
5. Tofighi B, Chemi C, Ruiz-Valcarcel J, Hein P, Hu L. Smartphone apps targeting alcohol and illicit substance use: systematic search in commercial app stores and critical content analysis. JMIR Mhealth Uhealth 2019;7(4):e11831 [FREE Full text] [doi: 10.2196/11831] [Medline: 31008713]
6. Williamson A, Heydarshahi B, Finley-Abboud D, Massac L, Jacobson L, Christophe N. What smartphone apps exist to support recovery from opioid use disorder? A content analysis of publicly available opioid-related smartphone apps. Addict Sci Clin Pract 2025;20(1):26 [FREE Full text] [doi: 10.1186/s13722-025-00549-y] [Medline: 40075426]
7. Carreiro S, Newcomb M, Leach R, Ostrowski S, Boudreaux ED, Amante D. Current reporting of usability and impact of mHealth interventions for substance use disorder: a systematic review. Drug Alcohol Depend 2020;215:108201 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108201] [Medline: 32777691]

XSL•FO
RenderX

8. Businelle MS, Perski O, Hébert ET, Kendzor DE. Mobile health interventions for substance use disorders. Annu Rev Clin Psychol 2024;20(1):49-76 [FREE Full text] [doi: 10.1146/annurev-clinpsy-080822-042337] [Medline: 38346293]
9. O'Logbon J, Wickersham A, Williamson C, Leightley D. J Ment Health 2024;33(5):645-673 [FREE Full text] [doi: 10.1080/09638237.2023.2245902] [Medline: 37664884]
10. Johansson M, Romero D, Jakobson M, Heinemans N, Lindner P. Digital interventions targeting excessive substance use and substance use disorders: a comprehensive and systematic scoping review and bibliometric analysis. Front Psychiatry 2024;15:1233888 [FREE Full text] [doi: 10.3389/fpsyt.2024.1233888] [Medline: 38374977]
11. Staiger PK, O'Donnell R, Liknaitzky P, Bush R, Milward J. Mobile apps to reduce tobacco, alcohol, and illicit drug Use: systematic review of the first decade. J Med Internet Res 2020;22(11):e17156 [FREE Full text] [doi: 10.2196/17156] [Medline: 33231555]
12. Neale J, Bowen AM. Lessons for uptake and engagement of a smartphone app (SURE Recovery) for people in recovery from alcohol and other drug problems: interview study of app users. JMIR Hum Factors 2022;9(1):e33038 [FREE Full text] [doi: 10.2196/33038] [Medline: 35258474]
13. Östh J, Danielsson AK, Lundin A, Wennberg P, Andréasson S, Jirwe M. Keeping track of my drinking - patient perceptions of using smartphone applications as a treatment complement for alcohol dependence. Subst Use Misuse 2024;59(2):291-299 [FREE Full text] [doi: 10.1080/10826084.2023.2269578] [Medline: 37876238]
14. Kazemi DM, Borsari B, Levine MJ, Li S, Shehab M, Fang F, et al. Effectiveness of a theory-based mHealth intervention for high-risk drinking in college students. Subst Use Misuse 2020;55(10):1667-1676. [doi: 10.1080/10826084.2020.1756851] [Medline: 32394772]
15. McKay JR, Gustafson DH, Ivey M, Pe-Romashko K, Curtis B. Efficacy and comparative effectiveness of telephone and smartphone remote continuing care interventions for alcohol use disorder: a randomized controlled trial. Addiction 2022;117(5):1326-1337 [FREE Full text] [doi: 10.1111/add.15771] [Medline: 34859519]
16. Gustafson DH, Landucci G, Vjorn OJ, Gicquelais RE, Goldberg SB, Johnston DC. Effects of bundling medication for opioid use disorder with an mHealth intervention targeting addiction: a randomized clinical trial. Am J Psychiatry 2024;181(2):115-124. [doi: 10.1176/appi.ajp.20230055] [Medline: 37789744]
17. Smith W, Ploderer B, Wadley G, Webber S, Borland R. Trajectories of engagement and disengagement with a story-based smoking cessation app. 2017 Presented at: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17); May 6-11, 2017; Denver p. 3045-3056. [doi: 10.1145/3025453.3026054]
18. Milward J, Deluca P, Drummond C, Kimergård A. Developing typologies of user engagement with the BRANCH alcohol-harm reduction smartphone app: qualitative study. JMIR Mhealth Uhealth 2018;6(12):e11692 [FREE Full text] [doi: 10.2196/11692] [Medline: 30545806]
19. Dick S, O'Connor Y, Thompson MJ, O'Donoghue J, Hardy V, Wu TJ. Considerations for improved mobile health evaluation: retrospective qualitative investigation. JMIR Mhealth Uhealth 2020;8(1):e12424 [FREE Full text] [doi: 10.2196/12424] [Medline: 32012085]
20. Friedman B, Kahn PH, Borning A. Value sensitive design and information systems. In: Zhang P, Galletta D, editors. Human-Computer Interaction in Management Information Systems: Foundations. Armonk: M.E. Sharpe, Inc; 2006:55-95.
21. Borning A, Muller M. Next steps for value sensitive design. 2012 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12); May 5-10, 2012; Austin p. 1125-1134. [doi: 10.1145/2207676.2208560]
22. Muller MJ, Kuhn S. Participatory design. Commun ACM 1993;36(6):24-28. [doi: 10.1145/153571.255960]
23. Trischler J, Pervan SJ, Kelly SJ, Scott DR. The value of codesign: the effect of customer involvement in service design teams. J Serv Res 2017;21(1):75-100. [doi: 10.1177/1094670517714060]
24. Thinyane M, Bhat K, Goldkind L, Cannanure VK. Critical participatory design: reflections on engagement and empowerment in a case of a community based organization. 2018 Presented at: Proceedings of the 15th Participatory Design Conference: Full Papers - Volume 1 (PDC '18); August 20-24, 2018; Hasselt & Genk p. 1-10. [doi: 10.1145/3210586.3210601]
25. Schmitt Z, Yarosh S. Participatory design of technologies to support recovery from substance use disorders. Proc ACM Hum-Comput Interact 2018;2(CSCW):1-27. [doi: 10.1145/3274425]
26. Struik LL, Bottorff JL, Baskerville NB, Oliffe JL. The crush the crave quit smoking app and young adult smokers: qualitative case study of affordances. JMIR Mhealth Uhealth 2018;6(6):e134 [FREE Full text] [doi: 10.2196/mhealth.9489] [Medline: 29884602]
27. Eaves ER, Doerry E, Lanzetta SA, Kruithoff KM, Negron K, Dykman K. Applying user-centered design in the development of a supportive mHealth app for women in substance use recovery. Am J Health Promot 2023;37(1):56-64 [FREE Full text] [doi: 10.1177/08901171221113834] [Medline: 35815770]
28. Fleddermann K, Molfenter T, Vjorn O, Horst J, Hulsey J, Kelly B. Patient preferences for mobile health applications to support recovery. J Addict Med 2023;17(4):394-400. [doi: 10.1097/adm.0000000000001137]
29. Jones J, Yuan Y, Yarosh S. Be consistent, work the program, be present every day. Proc ACM Interact Mob Wearable Ubiquitous Technol 2021;5(4):1-26. [doi: 10.1145/3494955]
30. Zimmerman J, Stolterman E, Forlizzi J. An analysis and critique of research through design: Towards a formalization of a research approach. 2010 Presented at: Proceedings of the 8th ACM Conference on Designing Interactive Systems (DIS '10); August 16-20, 2010; Aarhus p. 310-319. [doi: 10.1145/1858171.1858228]

XSL•FO
RenderX

31. Zimmerman J, Forlizzi J. Research through design in HCI. In: Olson JS, Kellogg WA, editors. Ways of Knowing in HCI. New York: Springer New York; 2014:167-189.
32. Godin D, Zahedi M. Aspects of research through design: a literature review. 2014 Presented at: Design's Big Debates – DRS International Conference 2014; June 16-19, 2014; Umeå p. 15.
33. Advance Patient Support, Care, and Research.: Zamplo Inc URL: https://www.zamplo.org/ [accessed 2025-09-16]
34. Philip B, Abdelrazek M, Barnett S, Bonti A, Grundy J. Towards better mHealth apps: Understanding current challenges and user expectations. 2022 Presented at: roceedings of the 9th IEEE/ACM International Conference on Mobile Software Engineering and Systems (MOBILESoft '22); May 17-18, 2022; Pittsburgh p. 33-37. [doi: 10.1145/3524613.3527804]
35. Hutchinson H, Mackay W, Westerlund B, Bederson BB, Druin A, Plaisant C. Technology probes: inspiring design for and with families. 2003 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03); April 5-10, 2003; Fort Lauderdale p. 17-24. [doi: 10.1145/642611.642616]
36. Byrne D. A worked example of Braun and Clarke's approach to reflexive thematic analysis. Qual Quant 2021;56(3):1391-1412. [doi: 10.1007/s11135-021-01182-y]
37. Braun V, Clarke V. Conceptual and design thinking for thematic analysis. Qual Psychol 2022;9(1):3-26. [doi: 10.1037/qup0000196]
38. Almani A, Alrwais O. The Role of Wireframes in Enhancing User Interface Design. IRJIET 2024;8(12):134-140. [doi: 10.47001/IRJIET/2024.812020]
39. Braun V, Clarke V. Toward good practice in thematic analysis: avoiding common problems and be(com)ing a knowing researcher. Int J Transgend Health 2023;24(1):1-6 [FREE Full text] [doi: 10.1080/26895269.2022.2129597] [Medline: 36713144]
40. Scheibein F, Caballeria E, Taher MA, Arya S, Bancroft A, Dannatt L. Optimizing digital tools for the field of substance use and substance use disorders: backcasting exercise. JMIR Hum Factors 2023;10:e46678 [FREE Full text] [doi: 10.2196/46678] [Medline: 38085569]
41. Shrestha P. What makes a good app retention rate? Adjust.com. 2024. URL: https://www.adjust.com/blog/what-makes-a-good-retention-rate/ [accessed 2025-08-02]
42. Joshi S. App retention benchmarks broken down by industry. Sendbird.com. 2024. URL: https://sendbird.com/blog/app-retention-benchmarks-broken-down-by-industry [accessed 2025-08-02]
43. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009:24. [doi: 10.17705/1CAIS.02428]
44. Theodorakis Y, Hassandra M, Panagiotounis F. Enhancing substance use disorder recovery through integrated physical activity and behavioral interventions: a comprehensive approach to treatment and prevention. Brain Sci 2024;14(6):534 [FREE Full text] [doi: 10.3390/brainsci14060534] [Medline: 38928533]
45. Laudet AB. The case for considering quality of life in addiction research and clinical practice. Addict Sci Clin Pract 2011;6(1):44-55 [FREE Full text] [Medline: 22003421]
46. Eekhoudt CR, Sandhu M, Mniszak C, Goodyear T, Turuba R, Marchand K. Looking beyond drugs: a scoping review of recovery in the context of illicit substance use among adolescents and young adults. Int J Drug Policy 2024;133:104598 [FREE Full text] [doi: 10.1016/j.drugpo.2024.104598] [Medline: 39378780]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
mHealth: mobile health
RtD: research through design

Edited by N Mohamad Ali; submitted 17.Sep.2025; peer-reviewed by J Guo, R Ruparel; comments to author 07.Dec.2025; accepted 11.Jan.2026; published 11.Feb.2026.

Please cite as:
Masrani T, Hodgins DC, Kim HS, Rittenbach K, Johnson E, Messier G
Designing mHealth Apps for Substance Use Recovery Through Real-World Co-Design and Deployment: Mixed Methods Study
JMIR Mhealth Uhealth 2026;14:e83984
URL: https://mhealth.jmir.org/2026/1/e83984
doi:10.2196/83984
PMID:

©Teale Masrani, David C Hodgins, Hyoun S Kim, Katherine Rittenbach, Erika Johnson, Geoffrey Messier. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.Feb.2026. This is an open-access article distributed under the terms

XSL•FO
RenderX

of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Adapting and Validating Tools to Assess the Usability and Acceptability of mHealth Tools Among Community Health Workers in Rural Settings: Development and Usability Study

Jonathan Nkurunziza[1], MPH; Sarah Nuss[2,3], BS, MD; Eve Hiyori Estrada[4], BA; Marthe Kubwimana[1], MCP; Adeline Adwoa Boatin[2,5], MD, MPH; Laban Bikorimana[1], MPH; Richard Ribon Fletcher[6], BSc, MSc, PhD; Nissi Byiringiro[7], BE; Bethany Hedt-Gauthier[2,4*], BS, MS, PhD; Vincent Kalumire Cubaka[1,8*], MD, MMed, PhD

[1]Partners In Health, Rwanda, Kigali, Rwanda

[2]Program in Global Surgery and Social Change, Harvard Medical School, Boston, MA, United States

[3]Warren Alpert Medical School, Brown University, Providence, RI, United States

[4]Department of Global Health and Social Medicine, Harvard Medical School, Boston, MA, United States

[5]Department of Obstetrics and Gynecology, Massachusetts General Hospital, Boston, MA, United States

[6]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, United States

[7]Insightiv LTD, Kigali, Rwanda

[8]University of Global Health Equity, Burera, Rwanda, Department of Community Health and Social Medicine, Kigali, Rwanda

[*]these authors contributed equally

Corresponding Author:
Jonathan Nkurunziza, MPH
Partners In Health, Rwanda
3 KG 546 St, Kacyiru-Rwanda
Kigali, Po. Box 3432
Rwanda
Phone: 250 788 308 289
Email: jnkurunziza3@gmail.com

Abstract

Background: Mobile health (mHealth) apps are increasingly leveraged to support community health workers (CHWs) in delivering high-quality care, particularly in low- and middle-income countries. However, despite the proliferation of mHealth tools, few have been implemented at scale, partly due to limited attention to usability and acceptability among end users. In sub-Saharan Africa, mHealth tools designed for CHWs often lack systematic evaluation using validated instruments tailored to local contexts. Without such assessments, it is difficult to ensure that these tools can be integrated effectively into CHW workflows and scaled sustainably.

Objective: This study aimed to adapt and validate existing mHealth usability and acceptability assessment tools to be contextually appropriate for CHWs in Rwanda. Specifically, we sought to ensure contextual appropriateness for CHWs supporting postoperative home follow-up for women after cesarean delivery. The resulting tool was designed for use in an implementation study of a novel CHW-led mHealth app.

Methods: This study was conducted in the Kirehe district, Rwanda, from October 2022 to March 2023. We adapted 2 established tools—the mHealth App Usability Questionnaire and selected items from the Practitioner Opinion (Acceptability) Scale—and added new items that reflect core functions of the CHW-focused mHealth app. All items were translated into Kinyarwanda and simplified to align with CHWs' educational levels. We conducted a three-stage validation that consisted of (1) content validity testing with 8 local and international experts using a recommended content validity index threshold of >0.78; (2) face validity testing with 10 CHWs using a recommended face validity index threshold of ≥0.60; and (3) reliability testing using responses from 30 CHWs, with a Cronbach α coefficient of ≥0.70 indicating acceptable internal consistency.

Results: Of the 25 items assessed, 22 (88%) achieved a content validity index score of >0.78 for both clarity and relevance. The face validity index across all 22 items was 0.991, indicating strong comprehensibility and relevance to CHWs. Internal consistency was high: the Cronbach α was 0.86 for the mHealth App Usability Questionnaire items, 0.73 for the Practitioner Opinion (Acceptability) Scale items, and 0.87 for the newly developed questions. The final tool—named the Community Health

XSL•FO
RenderX

Worker mHealth Usability and Acceptability Assessment Tool—included 22 items with strong content validity, face validity, and internal reliability.

Conclusions: This study presents a rigorously adapted and validated tool for assessing mHealth usability and acceptability among CHWs in Rwanda. The Community Health Worker mHealth Usability and Acceptability Assessment Tool can guide future evaluations of mHealth interventions in similar contexts and serve as a model for localizing mHealth assessment tools in low- and middle-income country settings to ensure fit-for-purpose implementation.

(JMIR Mhealth Uhealth 2026;14:e64916) doi:10.2196/64916

KEYWORDS

mobile health; digital health; Africa; global surgery; community health workers; adaptation and validation; usability; acceptability

Introduction

Mobile health (mHealth) apps have emerged as important tools to support health care workers in delivering high-quality care, particularly in low- and middle-income countries. Over the past 2 decades, an increasing number of mHealth solutions have been developed to assist community health workers (CHWs) in performing critical tasks across a range of health interventions [1,2]. CHWs, who often have limited formal training, nonetheless play a vital role in screening, diagnosing, and managing health conditions at the community level [3]. mHealth apps targeting CHWs are typically designed to support specific workflows or interventions. However, to ensure that these tools are adopted and used effectively, it is essential to evaluate their usability and acceptability among intended users. Robust assessment can guide the refinement of interface design and core functionalities, ultimately improving user experience and the quality of care delivered. Despite the increasing deployment of CHW-focused mHealth tools, few published studies have quantitatively assessed their usability or acceptability in sub-Saharan Africa (SSA) [4-6], and to our knowledge, there are no validated quantitative tools that have been developed for this purpose.

There are several tools that assess the usability of digital interventions or mHealth apps among target users outside of CHWs. Examples of usability assessment questionnaires geared toward patients include the Mobile App Rating Scale, a tool to measure the quality of apps commonly used in the health domain [7]; the System Usability Scale, developed to assess the usability of health care innovations [8]; and the Health Information Technology Usability Evaluation Scale, a tool used to assess the usability of mHealth technologies in groups of adults living with chronic diseases [9]. Tools that assess usability among clinical providers are the Telehealth Usability Tool [10] and the mHealth App Usability Questionnaire (MAUQ) [11]. The Telehealth Usability Tool was designed to evaluate the interface quality, reliability, and satisfaction of computer-based telehealth videoconferencing software, whereas the MAUQ was developed to evaluate mHealth apps.

Evaluation of user acceptability is also a critical aspect in the development of mHealth apps. The Practitioner Opinion (Acceptability) Scale (POAS) is an example of a tool designed to evaluate the acceptability of decision aids during the development process and early evaluation stages [12]. Other examples of acceptability tools to evaluate mHealth technology include the Acceptability E-scale, [13], the technology acceptance model [14], and the User Engagement Scale [15]. However, none of these were developed to assess acceptability among CHWs.

In this paper, we describe our process of adapting existing tools to develop and validate a quantitative tool to assess CHWs' acceptability of an mHealth app and its usability. This validated tool was specifically developed to assess the usability and acceptability of a new mHealth app to support CHW-led comprehensive home-based follow-up care among women who had delivered via cesarean section in rural Rwanda.

Methods

Study Design

Between October 2022 and March 2023, we adapted and validated a tool to assess the usability and acceptability of a novel mHealth screening app in a sample of target end users (CHWs). The mHealth app supports comprehensive home-based follow-up care among women who have delivered via cesarean section. The usability and acceptability assessment tool we developed is referred to as the Community Health Worker mHealth Usability and Acceptability Assessment Tool (CHW-MUAAT).

Study Setting

This study was implemented in the Kirehe district located in Eastern Province, rural Rwanda. The Kirehe district is represented by Kirehe District Hospital, with 19 health centers and 2448 CHWs. This hospital operates under the Rwanda Ministry of Health and is supported by Inshuti Mu Buzima, a Rwandan sister organization of the US-based nongovernmental organization Partners In Health since 2005. This study was conducted as part of the study titled "mHealth - community health worker tool for comprehensive postcesarean follow-up in rural Rwanda," which was being conducted in the same area.

Selection of the Existing Tools and Initial Questions

After reviewing multiple existing tools to assess usability and acceptability of mHealth or digital health interventions, our team decided to move forward with 2 existing tools due to the relevant information they provided to our study. To assess usability among CHWs, we adapted the MAUQ for stand-alone mHealth apps, which was developed and validated in English by Zhou et al [11]. The original MAUQ comes in 4 versions depending on the type of app (interactive or stand-alone) and the target users (patients or health care providers). Given that our mHealth app is stand-alone and geared toward CHWs, we

XSL•FO
RenderX

adapted the MAUQ for stand-alone mHealth apps used by health care providers. This version of the MAUQ includes 14 questions. For acceptability, we selected the questions most relevant to our study from the POAS, validated by O'Connor and Cranney [12]; we started with 4 questions that were deemed relevant out of the 15 from this tool. In addition, we developed and added 7 usability questions; these questions were drafted by the core study team members to reflect core functions of the mHealth screening app under development. These functions were not captured in the MAUQ or POAS questions and underwent the validation steps described in subsequent sections. To assess the reading level of our assessment tool, we intentionally designed all items to meet a below 8th-grade reading level, following widely accepted best practices in health communication. This decision was guided by recommendations from the National Institutes of Health (NIH), which emphasize using plain language, short and clear sentences, and familiar vocabulary in survey instruments to promote comprehension and reduce respondent burden. These NIH guidelines support creating materials that are accessible across diverse literacy levels and help minimize misunderstanding or response error (NIH Plain Language Guidance [16]).

Translation and Adaptation

After ensuring its comprehensibility with a reading level equivalent to sixth grade, the CHW mHealth usability and acceptability assessment tool underwent a cross-cultural adaptation process, which consisted of translation into Kinyarwanda, back translation, and harmonization, as described by Beaton et al [17]. A study team member fluent in Kinyarwanda and English translated the adapted questionnaire from English into Kinyarwanda. After translation, an independent translator with skills in the health domain and who did not have access to the original version of the assessment questions translated the questions back into English. The study team and translator met to review and complete a harmonization process to help identify any ambiguity, errors, or confusion in the translated version and ensure that the tool was appropriate for the cultural context.

Validity and Reliability Testing

Validity testing comprised 2 assessments: content validity testing and face validity testing, whereas reliability was evaluated through internal consistency analysis.

Content Validity

Content validity assessed the relevancy and clarity of the tool among content experts. Eight content experts, including app developers, study team members, and clinicians, were conveniently asked to rate the relevancy and clarity of items from the CHW-MUAAT on a scale from 1 to 4 (1="not relevant" or "not clear"; 4="very relevant" or "very clear"). Content validity was established using the content validity index (CVI). The CVI was calculated using the number of experts who agreed that the item was relevant or clear (scores of 3 or 4) divided by the total number of experts. The scores for each individual item are referred to as the "item CVI." On the basis of recommended CVI thresholds from the literature [18], items with an item CVI of >0.78 were deemed content valid and remained on the survey. Those with an item CVI score of ≤0.78 were removed.

Face Validity

Face validity was established by assessing the clarity and comprehensibility of the translated tool among target end users. In total, 10 target users (CHWs) from the Kirehe district were selected at random and asked to rate the clarity and comprehensibility of the translated items on the assessment tool on a scale from 1 to 4 (1="not at all clear" or "not at all understandable"; 4="very clear" or "very understandable"). The face validity index (FVI) was calculated by dividing the number of CHWs who agreed that the item was understandable (score of 3 or 4) by the total number of CHWs (n=10). On the basis of recommendations from the literature [19], an FVI above 0.74 was considered excellent, an FVI of 0.60 to 0.74 was considered good, and an FVI of 0.54 to 0.59 was considered fair. Items with an FVI of ≥0.60 were considered acceptable and kept.

Reliability Testing

After the content and face validity testing, the adapted tool was used to assess the usability and acceptability of the mHealth app among 30 CHWs. The 30 CHWs received explanations on the use of the mHealth CHW app and were asked to respond to all questions on the app during 3 sample patient vignette role-plays led by a study team facilitator. After the sample role-play, the CHWs were asked to answer questions from the CHW-MUAAT. The full results of this assessment are presented elsewhere. For this study, we used the data from reliability testing to assess the internal consistency of the tool. The internal consistency of the tool was evaluated by calculating the Cronbach α coefficient [20]. A higher α value suggests greater internal consistency reliability, with an α value greater than 0.70 deemed to indicate good reliability [21,22].

Ethical Considerations

This tool adaptation and validation study falls under a larger study titled "mHealth - community health worker tool for comprehensive postcesarean follow-up in rural Rwanda," which received human participant ethical review approval from the Rwanda National Ethics Committee (109/RNEC/2022) and the Institutional Review Board of the Harvard Faculty of Medicine (22-1025). For the reliability testing, all CHWs provided informed consent before participating in the mHealth usability and acceptability assessments and were able to opt out of participation in the study activities. Data used in measuring validity and reliability were deidentified. Participants were compensated with 6000 Rwandan francs (approximately US $4.15) for their time and travel related to the study.

Results

The first adapted tool included 14 questions from the MAUQ, 4 POAS questions, and 7 questions added by the study team, yielding a total of 25 questions. All questions were modified for a sixth-grade reading level and underwent cross-cultural translation. The final questions for this first adaptation are presented in Table 1.

XSL•FO
RenderX

Table 1. Adaptation of original questions to sixth-grade reading level and cross-language translation.

Source and original question	Adapted question to sixth-grade English reading level (back translation)	Adapted question to Kinyarwanda
MAUQ[a]		
"The app was easy to use."	"The app is easy to use."	"Iyi apurikasiyo iroroshye kuyikoresha."
"It was easy for me to learn to use the app."	"It is easy for me to learn to use the app."	"Biranyorohera kwiga gukoresha iyi apurikasiyo y'ikoranabuhanga."
"The navigation was consistent when moving between screens."	"The navigation is consistent when moving between screens."	"Kugenda mpinduranya aho nkorera ni ibintu bihamye cyangwa bikorwa mu byryo bumwe mu gukoresha iyi apurikasiyo."
"The interface of the app allowed me to use all the functions (such as entering information, responding to reminders, viewing information) offered by the app."	"I can use all the functions (such as entering information, responding to reminders, viewing information) in the app."	"Nshobora gukoresha fongisiyo zose / ibice byose (nko kwinjiza amakuru, gusubiza ibijyanye no kwibutsa, kureba amakuru)."
"Whenever I made a mistake using the app, I could recover easily and quickly."	"Whenever I make a mistake on the app, it is easy and quick to correct it."	"Igihe cyose nkoze ikosa cyangwa nibeshye mu gukoresha apurikasiyo, biroroshye kandi biruhuta kurikosora."
"I like the interface of the app."	"I like the interface of the app."	"Nakunze uburyo apurikasiyo igaragara mu kuyikoresha."
"The information in the app was well organized, so I could easily find the information I needed."	"The information in the app is well organized, so I can easily find the information I need."	"Amakuru yashyizwe neza muri apurikasiyo ku buryo nshobora kubonamo ayo nkeneye ku buryo bworoshye."
"The app adequately acknowledged and provided information to let me know the progress of my action."	"The app does a good job of letting me know my progress."	"Apurikasiyo imfasha kumenya aho ngeze nkora."
"The amount of time involved in using this app has been fitting for me."	"The time it takes to use this app works well for me."	"Umwanya bimfata mu gukoresha iyi apurikasiyo mu kazi urakwiye kuri njye."
"Overall, I am satisfied with this app."	"Overall, I am satisfied with this app."	"Muri rusange nyuzwe niyi apurikasiyo."
"The app improved my access to delivering healthcare services."	"The app improves my ability to deliver healthcare services."	"Apurikasiyo itezimbere ubushobozi bwanjye bwo gutanga serivisi z'ubuzima."
"The app helped me manage my patients' health effectively."	"The app helps me manage my patients' health effectively."	"Apurikasiyo imfasha mu kwita neza ku buzima bw'abarwayi banjye mu buryo buboneye."
"This app has all the functions and capabilities I expected it to have."	"This app has all the functions and capabilities I expected it to have."	"Iyi apurikasiyo ifite imikorere n'ubushobozi nateganyaga ko ifite."
"I could use the app even when the Internet connection was poor or not available."	"I can use the app even when the Internet connection was poor or not available."	"Nshobora gukoresha apurikasiyo nubwo umurongo wa interineti waba ari muke cyangwa ari ntawo."
POAS[b]		
"It will be easy for me to understand how to use the app."	"It is easy for me to understand how to use the app."	"Biranyoroheye kumva uburyo bwo gukoresha apurikasiyo."
"This strategy is compatible with the way I think things should be done."	"The app is a good fit for the way I conduct visits to women after cesarean section."	"Apurikasiyo ikwiranye neza n'uburyo bwo gusura abagore nyuma yo kubyara babazwe."
"Using this strategy will save me time."	"The app will save me time during home visits to women who delivered by cesarean section."	"Apurikasiyo izamfasha kuzigama igihe nakoreshaga mugihe cyo gusura mu rugo ababyeyi babyaye babazwe."
"There is a high probability that using this strategy may cause or result in more benefit than harm."	"I think using this mHealth app will be more helpful than hurtful."	"Ndatekereza ko gukoresha iyi apurikasiyo ya mHealth bizafasha cyane kuruta guteza ibyago."
Questions added by the study team members		
—[c]	"This app would be useful for my home visits to women after cesarean section."	"Iyi apurikasiyo yaba ingirakamaro mu gusura mu rugo ababyeyi nyuma yo kubyara babazwe."
—	"It is easy for me to log into the app with my username and password."	"Biranyorohera kwinjira muri apurikasiyo ukoresheje izina ryanjye n'ijambo ryibanga."
—	"It is easy for me to find the correct patient using the patient name ID."	"Biranyorohera kubona umurwayi nyawe nkoresheje ijambo riranga umurwayi."

XSL•FO
RenderX

Source and original question	Adapted question to sixth-grade English reading level (back translation)	Adapted question to Kinyarwanda
—	"It is easy for me to take or retake a picture of the wound using the app."	"Biranyorohera gufata ndetse no kongera gufata ifoto y'igisebe nkoresheje apurikasiyo."
—	"It is easy for me to read the recommendations given by the app."	"Biranyorohera gusoma amabwiriza yatanzwe na apurikasiyo."
—	"It is easy for me to communicate with women when using the app."	"Biranyorohera kuvugana n'mubyeyi mugihe nkoresheje apurikasiyo."
—	"I feel comfortable using the app during visit for follow-up of women who delivered by C-section."	"Ndumva mbohokewe cyangwa nifitiye ikizere mu gukoresha iyi apurikasiyo mugihe cyo gusura mu rugo kugirango nkurikirane abagore babyaye babazwe."

[a]MAUQ: mHealth App Usability Questionnaire.

[b]POAS: Practitioner Opinion (Acceptability) Scale.

[c]Not applicable.

A total of 10 content experts participated in the content validity survey. In total, 88% (22/25) of the items received an item CVI score of >0.78 on both the clarity and relevance scales (Table 2). Three items had an item CVI lower than this threshold and were removed. The 3 items that were removed from the first adapted set of questions were "The app does a good job of letting me know my progress," "The navigation is consistent between screens," and "I can use all the functions (such as entering information, responding to reminders, and viewing information) in the app."

All the remaining 22 items received individual FVI scores of >0.60 for clarity and relevance and, thus, were retained in the full final tool (Table 2). The FVI score for clarity across all questions was 0.991 and was the same for relevance. The full final tool is available in Table 3. The Cronbach α value was 0.86 for the 12 MAUQ usability questions, 0.73 for the 4 POAS acceptability questions, and 0.87 for the questions added by the study team, indicating the high internal consistency reliability of these questions.

XSL•FO
RenderX

Table 2. Content and face validity testing of items considered for the Community Health Worker mHealth Usability and Acceptability Assessment Tool.

Item	Content validity testing with content expert panel[a]		Face validity testing with CHWs[b,c]	
	Clarity I-CVI[d]	Relevance I-CVI	Clarity FVI[e]	Comprehension FVI
"The app is easy to use."	1	1	1	1
"It is easy for me to learn to use the app."	0.9	1	1	1
"The navigation is consistent when moving between screens."	0.6	0.8	—[f]	—
"I can use all the functions (such as entering information, responding to reminders, viewing information) in the app."	0.6	0.9	—	—
"Whenever I make a mistake on the app, it is easy and quick to correct it."	0.9	1	1	1
"I like the interface of the app."	0.9	0.9	1	1
"The information in the app is well organized, so I can easily find the information I need."	0.9	1	1	1
"The app does a good job of letting me know my progress."	0.9	0.7	—	—
"The time it takes to use this app works well for me."	0.9	1	1	1
"Overall, I am satisfied with this app."	1	0.9	1	1
"The app improves my ability to deliver healthcare services."	1	1	1	1
"The app helps me manage my patients' health effectively."	0.9	0.9	1	1
"This app has all the functions and capabilities I expected it to have."	1	1	1	1
"I can use the app even when the Internet connection was poor or not available."	0.8	1	0.8	0.8
"This app would be useful for my home visits to women after cesarean section."	1	0.9	1	1
"It is easy for me to log into the app with my username and password."	1	1	1	1
"It is easy for me to find the correct patient using the patient name ID."	1	1	1	1
"It is easy for me to take or retake a picture of the wound using the app."	1	1	1	1
"It is easy for me to read the recommendations given by the app."	1	1	1	1
"It is easy for me to communicate with women when using the app."	0.9	1	1	1
"I feel comfortable using the app during home visit for follow up of women who delivered by caesarean section."	1	1	1	1
"It is easy for me to understand how to use the app."	1	0.8	1	1
"The app is a good fit for the way I conduct visits to women after cesarean section."	0.8	0.9	1	1
"The app will save me time during home visits to women who delivered by cesarean section."	1	1	1	1
"I think using this mHealth app will be more helpful than hurtful."	0.9	0.9	1	1

[a]In total, 10 content experts from the study team participated in content validity testing. Items with a clarity item content validity index or relevance item content validity index of <0.78 were removed from the final tool.

[b]In total, 10 community health workers participated in face validity testing. Items with a face validity index of <0.60 were removed from the final tool.

[c]CHW: community health worker.

[d]I-CVI: item content validity index.

XSL•FO
RenderX

[e]FVI: face validity index.

[f]Not applicable.

Table 3. Contents of the Community Health Worker mHealth Usability and Acceptability Assessment Tool[a].

Item number	Statement	"Strongly disagree" score	"Disagree" score	"Neutral" score	"Agree" score	"Strongly agree" score
1	"The app is easy to use."	1	2	3	4	5
2	"It is easy for me to learn to use the app."	1	2	3	4	5
3	"Whenever I make a mistake on the app, it is easy and quick to correct it."	1	2	3	4	5
4	"I like the interface of the app."	1	2	3	4	5
5	"The information in the app is well organized, so I can easily find the information I need."	1	2	3	4	5
6	"The time it takes to use this app works well for me."	1	2	3	4	5
7	"Overall, I am satisfied with this app."	1	2	3	4	5
8	"The app improves my ability to deliver healthcare services."	1	2	3	4	5
9	"The app helps me manage my patients' health effectively."	1	2	3	4	5
10	"This app has all the functions and capabilities I expected it to have."	1	2	3	4	5
11	"Once I have logged in, I can use the app even when the Internet connection was poor or not available."	1	2	3	4	5
12	"This app would be useful for my home visits to women after cesarean section."	1	2	3	4	5
13	"It is easy for me to log into the app with my username and password."	1	2	3	4	5
14	"It is easy for me to find the correct patient using the Patient name ID."	1	2	3	4	5
15	"It is easy for me to take or retake a picture of the wound using the app."	1	2	3	4	5
16	"It is easy for me to read the recommendations given by the app."	1	2	3	4	5
17	"It is easy for me to communicate with women when using the app."	1	2	3	4	5
18	"I feel comfortable using the app during home visit for follow up of women who delivered by caesarean section."	1	2	3	4	5
19	"It is easy for me to understand how to use the app."	1	2	3	4	5
20	"The app is a good fit for the way I conduct visits to women after cesarean section."	1	2	3	4	5
21	"The app will save me time during home visits to women who delivered by cesarean section."	1	2	3	4	5
22	"I think using this mHealth app will be more helpful than hurtful."	1	2	3	4	5

[a]Questions 1 to 12 are from the mHealth App Usability Questionnaire (MAUQ). Questions 13 to 18 were developed to assess the usability of core functions of the mobile health community health worker app not captured by the MAUQ or Practitioner Opinion (Acceptability) Scale (POAS). Questions 19 to 22 were adapted from the POAS.

XSL•FO
RenderX

Discussion

Principal Results

In this study, we adapted 2 existing tools, the MAUQ and POAS, to evaluate CHWs' perceptions of the usability and acceptability of a novel mHealth app for comprehensive home-based follow-up care among women who have delivered via cesarean section in Rwanda. Our final tool included 22 items in total: 12 (55%) out of the 15 original items from the MAUQ tool, all 4 (18%) of the items that we pulled from the POAS, and 6 (27%) new items added by our study team. The final items had high content validity among content experts and high face validity and internal reliability among CHWs.

While there is an increasing number of mHealth apps designed to support CHWs across SSA, very few are operating at a large scale and beyond isolated research settings [2,23,24]. One of the key barriers to scale-up is the limited understanding of whether these tools are perceived as usable and acceptable by the CHWs expected to adopt them. Without sufficient usability and acceptability, even well-designed tools are unlikely to integrate effectively into CHW workflows. Although a few studies in SSA have examined these dimensions among CHWs [2,4-6], most have not used adapted or validated tools, limiting the reliability and comparability of their findings. A recent review found that 70% of researcher-developed usability questionnaires for mHealth apps were not validated—and those validated were not tailored to the community health context [25]. Validated assessment frameworks are critical to draw robust conclusions about user experience and determining the potential for mHealth tools to be well integrated into routine practice.

This quantitative CHW-MUAAT scale provides a standardized, scalable way to summarize usability and acceptability constructs, which facilitates comparison of potential digital tools in one environment or of the functionality of a single tool in different contexts [26]. However, complementary qualitative data can offer deeper insights into the contextual factors that shape CHW perceptions of usability or acceptability. For example, a mixed methods study identified that CHWs in Uganda had high rates of digital health acceptance but low digital tool use due to low smartphone ownership [27]; similarly, a study of *YendaNafe*, a mobile app to support CHW activities in Malawi, found an overall positive acceptance of the app but that there were several facilitators of and barriers to actual app use [28]. For this study, our qualitative results indicated general acceptability and emphasized that strong CHW-patient relationships and trust were needed for successful implementation (full results available elsewhere [29]).

To meaningfully inform the integration of an mHealth app into CHWs' workflows, usability and acceptability assessments must be tailored to the local context and to the needs and capacities of end users. In our study, this required careful linguistic and cognitive adaptation of existing measurement tools. Given that most CHWs in Rwanda have completed only primary school education, our team revised the wording of several items to ensure that they were easily understandable. For example, the original item "There is a high probability that using this strategy may cause or result in more benefit than harm" was simplified to "I think using this mHealth app will be more helpful than hurtful" [30]. Additionally, all items were translated into Kinyarwanda, the primary language spoken by CHWs, with attention to clarity and cultural appropriateness.

Beyond linguistic adaptation, we also modified the content of existing tools to better capture usability and acceptability as related to the specific task—home-based follow-up care after cesarean delivery. This included rewording items from the MAUQ and POAS tools and adding step-specific questions relevant to CHWs' real-world experience. For example, questions about ease of logging into the app were added given that many CHWs have limited experience with smartphones. This consideration is important in Rwanda, where, as of 2024, although 80% of households own a mobile phone, only 34% own a smartphone [31]. By demonstrating a structured approach to adapting and contextualizing usability and acceptability assessment tools, this study offers both a practical instrument and a replicable methodology for future CHW-focused mHealth evaluations in similar settings across the region.

This instrument can be adapted for other teams interested in assessing the usability and acceptability of a particular CHW mHealth tool with appropriate adaptation and validation. First, they may need to identify usability questions tailored to community health interventions, similarly to how our team added 7 usability questions specific to our postcesarean monitoring app. These questions will need to undergo reliability and validity assessments. The full tool should then be piloted in their settings, whereas smaller sample sizes may be warranted given the baseline work presented in this paper.

Limitations

We acknowledge that the adaptation and validation of the tool has some limitations that should be considered. First, this study had a small sample size of content experts and CHWs. Because of the overall high scores and consistent conclusions, we are confident in the validity of the final tool developed, but future validation studies will likely need larger sample sizes if there is more variability in participant responses. Second, this tool focuses on a quantitative assessment that cannot pick up nuanced responses. Our final evaluation includes the quantitative responses collected using this tool during the usability and acceptability testing of the app. However, we do believe that these standardized assessments can help those evaluating mHealth apps identify areas for improvement and test against prespecified targets to determine whether the mHealth app is ready for broader use. Finally, this tool was designed for a specific intervention—a mobile phone app developed for CHWs to use during the home visit follow-up of women who delivered via cesarean section—and was only validated in the Kinyarwanda language. Other research groups should proceed cautiously and likely conduct their own validation studies when applying this tool to a different intervention or in a different language.

Conclusions

Our team is using the final tool as a core component of evaluating an mHealth app used by CHWs to support women

XSL•FO
RenderX

after cesarean delivery. The processes and tool described in this paper serve as a model for others conducting CHW usability and acceptability assessments of mHealth apps in SSA. Such assessments will be critical for moving toward better integration of innovative solutions into CHW activities.

Acknowledgments

The study team acknowledges the support of Partners In Health/Inshuti Mu Buzima; Kirehe District Hospital; and Mulindi, Nasho, and Mushikiri Health Centers' leadership and the community health workers for their voluntary participation. The authors declare the use of generative artificial intelligence (GenAI) in the research and writing process. According to the Generative Artificial Intelligence Delegation Taxonomy (2025), limited proofreading and editing tasks were delegated to GenAI tools under full human supervision. The GenAI tool used was ChatGPT (OpenAI). Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to restrictions on data sharing in accordance with the Rwanda Data Protection and Privacy Law but are available from the corresponding author on reasonable request.

Funding

This study is part of a study funded by National Institutes of Health grant NIH/FIC-5R21HD103052-02. Author SN is supported by the Fogarty International Center and National Institute of Mental Health of the National Institutes of Health under award D43 TW010543.

Authors' Contributions

Conceptualization: SN, JN, EHE, BH-G, VKC
Data curation: JN, MK, SN, EHE
Funding acquisition: BH-G
Methodology: SN, JN, EHE, BH-G, VKC
Supervision: BH-G, VKC
Writing—original draft: JN
Writing—review and editing: JN, SN, EHE, MK, AB, RRF, NB, LB, BH-G, VKC
BH-G and VKC were principal investigators. All authors reviewed and approved of the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Braun R, Catalani C, Wimbush J, Israelski D. Community health workers and mobile technology: a systematic review of the literature. PLoS One 2013;8(6):e65772 [FREE Full text] [doi: 10.1371/journal.pone.0065772] [Medline: 23776544]
2. Early J, Gonzalez C, Gordon-Dseagu V, Robles-Calderon L. Use of mobile health (mHealth) technologies and interventions among community health workers globally: a scoping review. Health Promot Pract 2019 Nov;20(6):805-817. [doi: 10.1177/1524839919855391] [Medline: 31179777]
3. Glenton C, Javadi D, Perry HB. Community health workers at the dawn of a new era: 5. Roles and tasks. Health Res Policy Syst 2021 Oct 12;19(Suppl 3):128 [FREE Full text] [doi: 10.1186/s12961-021-00748-4] [Medline: 34641903]
4. Chang LW, Njie-Carr V, Kalenge S, Kelly JF, Bollinger RC, Alamo-Talisuna S. Perceptions and acceptability of mHealth interventions for improving patient care at a community-based HIV/AIDS clinic in Uganda: a mixed methods study. AIDS Care 2013;25(7):874-880 [FREE Full text] [doi: 10.1080/09540121.2013.774315] [Medline: 23452084]
5. Lacroze E, Frühauf A, Nordmann K, Rampanjato Z, Muller N, De Neve JW, et al. Usability and acceptance of a mobile health wallet for pregnancy-related healthcare: a mixed methods study on stakeholders' perceptions in central Madagascar. PLoS One 2023;18(1):e0279880 [FREE Full text] [doi: 10.1371/journal.pone.0279880] [Medline: 36595530]
6. Medhanyie AA, Little A, Yebyo H, Spigt M, Tadesse K, Blanco R, et al. Health workers' experiences, barriers, preferences and motivating factors in using mHealth forms in Ethiopia. Hum Resour Health 2015 Jan 15;13(1):2 [FREE Full text] [doi: 10.1186/1478-4491-13-2] [Medline: 25588973]
7. Martin-Payo R, Carrasco-Santos S, Cuesta M, Stoyan S, Gonzalez-Mendez X, Fernandez-Alvarez MD. Spanish adaptation and validation of the User Version of the Mobile Application Rating Scale (uMARS). J Am Med Inform Assoc 2021 Nov 25;28(12):2681-2686 [FREE Full text] [doi: 10.1093/jamia/ocab216] [Medline: 34613400]
8. Ensink CJ, Keijsers NL, Groen BE. Translation and validation of the System Usability Scale to a Dutch version: D-SUS. Disabil Rehabil 2024 Jan;46(2):395-400. [doi: 10.1080/09638288.2022.2160837] [Medline: 36573399]

XSL•FO
RenderX

9. Schnall R, Cho H, Liu J. Health information technology usability evaluation scale (Health-ITUES) for usability assessment of mobile health technology: validation study. JMIR Mhealth Uhealth 2018 Jan 05;6(1):e4 [FREE Full text] [doi: 10.2196/mhealth.8851] [Medline: 29305343]
10. Parmanto B, Lewis ANJ, Graham KM, Bertolet MH. Development of the Telehealth Usability Questionnaire (TUQ). Int J Telerehabil 2016;8(1):3-10 [FREE Full text] [doi: 10.5195/ijt.2016.6196] [Medline: 27563386]
11. Zhou L, Bao J, Setiawan IM, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
12. O'Connor AM, Cranney A. User manual - acceptability. Ottawa Hospital Research Institute. 2002. URL: https://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_Acceptability.pdf [accessed 2026-01-20]
13. Micoulaud-Franchi JA, Sauteraud A, Olive J, Sagaspe P, Bioulac S, Philip P. Validation of the French version of the Acceptability E-scale (AES) for mental E-health systems. Psychiatry Res 2016 Mar 30;237:196-200. [doi: 10.1016/j.psychres.2016.01.043] [Medline: 26809367]
14. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
15. Holdener M, Gut A, Angerer A. Applicability of the user engagement scale to mobile health: a survey-based quantitative study. JMIR Mhealth Uhealth 2020 Jan 03;8(1):e13244 [FREE Full text] [doi: 10.2196/13244] [Medline: 31899454]
16. Eltorai AEM, Ghanian S, Adams CA, Born CT, Daniels AH. Readability of patient education materials on the American Association for Surgery of Trauma website. Arch Trauma Res 2014 Jun;3(2):e18161 [FREE Full text] [doi: 10.5812/atr.18161] [Medline: 25147778]
17. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000 Dec 15;25(24):3186-3191. [doi: 10.1097/00007632-200012150-00014] [Medline: 11124735]
18. Shi J, Mo X, Sun Z. [Content validity index in scale development]. Zhong Nan Da Xue Xue Bao Yi Xue Ban 2012 Feb;37(2):152-155. [doi: 10.3969/j.issn.1672-7347.2012.02.007] [Medline: 22561427]
19. Zamanzadeh V, Ghahramanian A, Rassouli M, Abbaszadeh A, Alavi-Majd H, Nikanfar AR. Design and implementation content validity study: development of an instrument for measuring patient-centered communication. J Caring Sci 2015 Jun;4(2):165-178 [FREE Full text] [doi: 10.15171/jcs.2015.017] [Medline: 26161370]
20. Mat Nawi FA, Tambi AM, Samat MF, Mustapha WM. A review on the internal consistency of a scale: the empirical example of the influence of human capital investment on malcom baldridge quality principles in TVET institutions. Asian People J 2020;3(1):19-29 [FREE Full text] [doi: 10.37231/apj.2020.3.1.121]
21. Mustafa N, Safii NS, Jaffar A, Sani NS, Mohamad MI, Abd Rahman AH, et al. Malay version of the mHealth App Usability Questionnaire (M-MAUQ): translation, adaptation, and validation study. JMIR Mhealth Uhealth 2021 Feb 04;9(2):e24457 [FREE Full text] [doi: 10.2196/24457] [Medline: 33538704]
22. Allen P, Bennett K, Heritage B. SPSS Statistics Version 22: A Practical Guide. Victoria, Australia: Cengage Australia; 2014.
23. Huang F, Blaschke S, Lucas H. Beyond pilotitis: taking digital health interventions to the national level in China and Uganda. Global Health 2017 Jul 31;13(1):49 [FREE Full text] [doi: 10.1186/s12992-017-0275-z] [Medline: 28756767]
24. O'Donnell A. Commentary on Harder et al. (2020): ensuring the sustainability of mHealth in low- and middle-income countries-how do we cure 'pilotitis'? Addiction 2020 Jun 19;115(6):1061-1062. [doi: 10.1111/add.14986] [Medline: 32072704]
25. Lim PC, Lim YL, Rajah R, Zainal H. Usability questionnaire for standalone or interactive mobile health applications: a systematic review. BMC Digit Health 2025 Apr 01;3:11. [doi: 10.1186/s44247-025-00150-y]
26. An Q, Kelley MM, Hanners A, Yen PY. Sustainable development for mobile health apps using the human-centered design process. JMIR Form Res 2023 Aug 25;7:e45694 [FREE Full text] [doi: 10.2196/45694] [Medline: 37624639]
27. Chraish M, Oyama C, Aoki Y, Andrew D, Nishio M, Shi S, et al. Bridging the gap between community health workers' digital health acceptance and actual usage in Uganda: exploring key external factors based on technology acceptance model. PLOS Digit Health 2025 Nov 19;4(11):e0001099. [doi: 10.1371/journal.pdig.0001099] [Medline: 41259341]
28. Kachimanga C, Mulwafu M, Ndambo MK, Harare J, Murkherjee J, Kulinkina AV, et al. Experiences of community health workers on adopting mHealth in rural Malawi: a qualitative study. Digit Health 2024 May 15;10:20552076241253994 [FREE Full text] [doi: 10.1177/20552076241253994] [Medline: 38757088]
29. Estrada EH, Hedt-Gauthier B, Nkurunziza J, Kubwimana M, Nuss S, Forbes C, et al. Community health workers' usability and acceptability of an mHealth tool for post-cesarean assessments: a mixed-methods study in rural Rwanda. BMC Pregnancy Childbirth 2025 Dec 23 [FREE Full text] [doi: 10.1186/s12884-025-08155-1] [Medline: 41436987]
30. Condo J, Mugeni C, Naughton B, Hall K, Tuazon MA, Omwega A, et al. Rwanda's evolving community health worker system: a qualitative assessment of client and provider perspectives. Hum Resour Health 2014 Dec 13;12:71 [FREE Full text] [doi: 10.1186/1478-4491-12-71] [Medline: 25495237]
31. Seventh integrated household living conditions survey. National Institute of Statistics of Rwanda. 2025. URL: https://www.statistics.gov.rw/sites/default/files/documents/2025-04/EICV_7_booklet%20for%20dissemination.pdf [accessed 2025-07-14]

XSL•FO
RenderX

Abbreviations

CHW: community health worker
CHW-MUAAT: Community Health Worker mHealth Usability and Acceptability Assessment Tool
CVI: content validity index
FVI: face validity index
MAUQ: mHealth App Usability Questionnaire
mHealth: mobile health
NIH: National Institutes of Health
POAS: Practitioner Opinion (Acceptability) Scale
SSA: sub-Saharan Africa

Edited by A Stone; submitted 30.Jul.2024; peer-reviewed by Å Grönlund; comments to author 30.May.2025; revised version received 05.Dec.2025; accepted 09.Dec.2025; published 20.Feb.2026.

Please cite as:
Nkurunziza J, Nuss S, Estrada EH, Kubwimana M, Boatin AA, Bikorimana L, Fletcher RR, Byiringiro N, Hedt-Gauthier B, Cubaka VK
Adapting and Validating Tools to Assess the Usability and Acceptability of mHealth Tools Among Community Health Workers in Rural Settings: Development and Usability Study
JMIR Mhealth Uhealth 2026;14:e64916
URL: https://mhealth.jmir.org/2026/1/e64916
doi:10.2196/64916
PMID:

©Jonathan Nkurunziza, Sarah Nuss, Eve Hiyori Estrada, Marthe Kubwimana, Adeline Adwoa Boatin, Laban Bikorimana, Richard Ribon Fletcher, Nissi Byiringiro, Bethany Hedt-Gauthier, Vincent Kalumire Cubaka. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 20.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Lived Experiences of Older Adults Using Wearables With Real-Time Feedback: Phenomenological Study

Tien-Ying Lu[1], MS; Aurora Rosato[1], MS; Seraina Anne Dual[1], PhD; Sanna Kuoppamäki[1], PhD

Department of Biomedical Engineering and Health Systems, KTH Royal Institute of Technology, Stockholm, Sweden

Corresponding Author:
Tien-Ying Lu, MS
Department of Biomedical Engineering and Health Systems
KTH Royal Institute of Technology
Hälsovägen 11 C
Stockholm, SE-100 44
Sweden
Phone: 46 724446680
Email: tyl@kth.se

Abstract

Background: Wearable devices with real-time feedback (WRFs) provide increasing opportunities to enhance physical activity and improve rehabilitation through collecting and processing health-related data. Real-time feedback (RTF) from the device is expected to result in a more dynamic, coordinated, and synchronous rhythmic activity, defined as step-by-step movements mediated by the real-time heart rate feedback. However, age-specific characteristics in the user engagement with WRFs integrating real-time audio feedback have largely remained unexplored.

Objective: This study explores the lived experiences of older adults using wearables with RTF to uncover motivations, aspirations, and hindering factors in their engagement with WRFs in rhythmic activity. The study explores narratives that older adults articulate in their previous use of wearables for physical activity, their experiences with WRFs during rhythmic activity, and their meaning-making of the interactive features enhancing the synchronization of the movement during rhythmic activity.

Methods: The study was conducted as a qualitative interview study with 18 older adults who used a WRF for rhythmic activity during a 3-week period in their home environment. The wearable used in the study is a chest-band sensor device that helps users to synchronize their steps with their heartbeat through the provision of real-time audio feedback. The material consists of semistructured interviews before and after using the device. Material from the semistructured interviews was analyzed with an interpretative phenomenological analysis method.

Results: The study identified four main themes characterizing older adults' lived experiences with wearables, which are (1) use of wearable technologies without RTF in daily life, (2) embodied rhythmic negotiation with RTF, (3) interpretation of health data with RTF, and (4) temporal trajectories of device engagement with RTF. Older adults demonstrated intentional distancing from wearable technologies rather than simple disuse, prioritizing authentic bodily experiences over external validation. Their engagement was fundamentally relational, mediated through trusted social networks, and required dialogical support for data interpretation. Device-guided movement synchronization created contextually situated challenges that varied significantly based on environmental demands, individual bodily capacity, and exercise routines. Extended temporal engagement transformed participants' relationships with the technology from initial disruption to potential integration, with RTF serving as a bridge toward enhanced embodied awareness when carefully designed.

Conclusions: The study concludes that RTF from the device can enhance synchronization and bodily awareness, but meaningful engagement requires adaptive designs that respect older adults' authentic movement practices, accommodate their relational approach to technology validation, and allow sufficient time for embodied competency development.

(JMIR Mhealth Uhealth 2026;14:e71509) doi:10.2196/71509

KEYWORDS

wearables; physical activity; multimodal feedback; synchronization; older adults; interview study

XSL•FO
RenderX

Introduction

Background

Physical activity (PA) is an important predictor for active and healthy aging; yet, few older adults meet the recommended levels of PA [1]. Wearable devices with real-time feedback (WRFs) having sensors provide increasing opportunities to enhance PA through collecting and processing health-related data [2,3]. Despite their potential, the long-term user engagement with wearables remains a challenge [4-6]. In particular, age-specific characteristics in the user engagement with interactive wearables with real-time feedback (RTF) have remained largely unexplored.

This study explores the use of wearables with and without RTF to uncover the lived experiences of older adults in their engagement with interactive wearables. WRFs can be defined as body-worn technologies that implement technologies, such as sensors, to measure biomedical information, combined with modalities, such as haptic or auditory feedback, that help users to adjust the movement to improve the performance in PA [7]. Rhythmic activity can be defined as a form of PA that incorporates elements such as auditory feedback, music, and social interaction, which involve coordinated movements synchronized with music [8-10]. In this study, we define rhythmic activity as a step-by-step movement (eg, walking or running) that is mediated by a device-guided synchronization of step and heart rate.

Older adults, defined as an age group of 65 years and older, are considered an important user group of WRFs due to the potential of wearables to improve PA, rehabilitation, and cardiovascular health [1,11-14]. Due to demographic changes accompanied by population aging, wearables as digital health interventions are expected to contribute to health promotion and prevention of disease among this demographic [15]. However, the adherence with these interventions still remains a challenge [4-6]. Older adults' adherence with wearables is mediated by their health, motivations for PA, and digital skill literacy, which can impact the use of wearables that require dexterity, mobility, or vision [16,17]. Older adults emphasize enjoyment of social interaction as a motivation to be physically active [1], and wearables can enhance this through the provision of visual or auditory feedback [12]. The perceived appeal of visual or auditory feedback, however, is individual and mediated by the lived experiences in later life. This suggests that affordances embedded in wearables are not directed only toward improving performance, but also delivering a sense of self-awareness [13]. Considering the individual differences in older adults' motivations, it becomes important to recognize their lived experiences when exploring their adherence to WRFs.

The aim of this study is to explore the lived experiences of older adults using wearables with RTF in rhythmic activity. The study explores narratives that older adults articulate in their previous use of wearables for PA, their experiences with WRFs during rhythmic activity, and their meaning-making of the interactive features enhancing the synchronization of the movement during rhythmic activity.

Wearables With RTF

Wearable technologies are designed and used for enhancing PA in health promotion, preventative health, and treatment of health conditions [18]. As *self-care technologies*, wearables allow the user to take responsibility for managing their own health independently, without external intervention from the health care provider [15,19,20]. As a *medical technology*, wearables can be used before medical diagnosis to monitor PA levels, and detect and predict the onset of clinical conditions [21].

Wearable technologies include a variety of applications, such as smart watches and chest bands, and these technologies integrate features, such as instructions, self-monitoring, and reminders, along with customizable options [22,23]. These interactive features, such as goal setting, feedback provision, and motivation tracking, are increasingly implemented in day-to-day applications [23]. RTF refers to an interaction mechanism that provides haptic, auditory, or visual feedback to the user simultaneously with the user's movement, instead of after the movement or activity has ended [2,3]. Machine learning techniques have enabled accurate and rapid human gesture recognition, providing RTF to users [24]. WRFs have been investigated among athletes in the prevention of injuries, improvements of performance [2,25], as well as in rehabilitation [26]. For older adults, RTF has been shown as a preferred feature in wearables [11]. In comparison to other interactive features, RTF, as a rather novel feature, has nevertheless been less implemented in commercially available wearables [2].

RTF enables reactions from the user while the action or movement is ongoing. Auditory feedback allows users to be observant of the environment and fits into more dynamic settings compared with both visual and tactile guidance [27]. Through sound interaction, users can coordinate and stabilize movements with internal or external imposed rhythms [28]. Auditory feedback can be used in a variety of physical activities, ranging from guiding speed to correct movements [25], supporting coupling steps and breathing in day-to-day running exercises [27], to guiding static breathing exercises [29]. Beat-by-beat stepping to the heart rate has shown improved heart function in athletes; however, without assessing motivational aspects of this interaction [30]. As an alternative to the auditory cues, the use of music can increase motivation and engagement in cardiovascular rehabilitation [31].

Factors Influencing Adherence With Wearables Among Older Adults

Wearables are accessible and easy to use, but the adherence, defined as long-term user acceptance and engagement with the device, remains a challenge [4-6]. The ownership of wearables is defined by sociodemographic factors, such as age and gender. Younger groups with higher levels of education and full-time employment are more likely to own a wearable compared with other demographic groups [32]. Older adults' motivations for using a wearable may also significantly differ from other age cohorts regarding the perceived value and ease-of-use of the device [11,14,33-35].

Long-term adherence to wearables is determined by a combination of individual motivation, ease-of-use, and perceived

XSL•FO
RenderX

value of the wearable, which subsequently predicts whether or not the device will be integrated into the user's life [11,33]. Older adults must be motivated by the usefulness of the device and its added value to PA [11,36]. Formal and informal support and guidance in having access to the device and continuing its use are connected to long-term adoption [37]. These individual factors are expected to differ between older adults, resulting in heterogeneous use of wearables depending on various individual and social determinants [10,16,17].

Emotional, sensory, and social influences also impact the user engagement with wearables [13,38,39]. Long-term attachment with wearables requires that these devices elicit positive emotional responses among users [40]. When wearables are framed as an assistive technology that aims to compensate for age-related functional or cognitive losses, older adults might abandon these devices due to (negative) user images [36]. Wearable devices can be perceived as motivating or condescending depending on how freely older adults can make autonomous choices about PA [34].

Accordingly, older adults face barriers to wearable adoption, including issues about usability, affordability, and accessibility, which can lead to social isolation and marginalization from digital health innovations [41]. At the same time, digital technologies offer significant opportunities for older adults, who often have more leisure time and are financially independent, to seek health-related information and maintain self-care [15,42]. Fewer phenomenological studies are focused on exploring older adults' experiences with rather novel wearable technologies in their home environments. This methodological approach allows us to capture the essence of subjective experiences and the perceived value they derive from using wearables through first-person descriptions [41-44] and the meanings they attribute to the interactions with wearables [15,45]. The research questions (RQs) are the following:

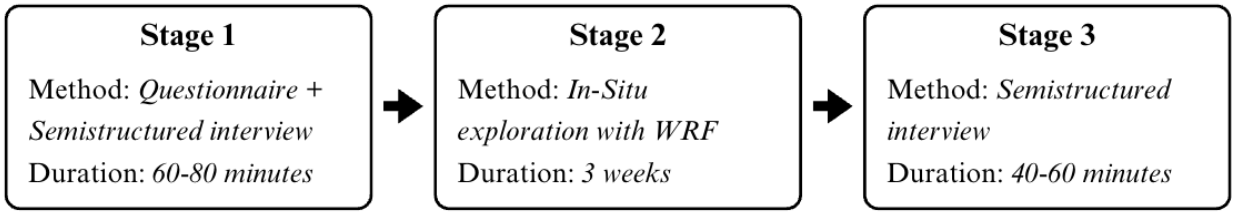
- RQ1: What kind of narratives do older adults articulate in the use of wearables for PA?
- RQ2: How do older adults experience the use of wearables with RTF in rhythmic activity?
- RQ3: What kind of meanings do older adults associate with the interactive features of a WRF during rhythmic activity?

Methods

Study Procedure

This study consisted of three stages: (1) semistructured interview and a questionnaire before using WRF, (2) a 3-week use of WRF for PA (referred to as in-situ study), and (3) a semistructured interview after using WRF. Figure 1 presents the method and duration of each stage.

Figure 1. Method and duration for each study stage.



In the first stage, we conducted a semistructured interview for participants to investigate their motivation for PA, self-perceived health and well-being, previous experience with technologies, and previous interactions with wearables for PA. Participants were also given a questionnaire measuring their PA level, self-perceived health, and sociodemographic characteristics.

In the second stage, participants were asked to use a WRF for a 3-week period in their home environment. They were asked to use the device at least 3 times per week during their PA routines, including any rhythmic activity, such as walking, running, or jogging. Each participant was shown how to use the device, and the device was set up and connected to their mobile phones. They also received detailed written instructions for using the device. A 3-week period for in-situ study was selected to ensure the feasibility of the procedure as a common period in digital health intervention studies [46].

In the third stage, we conducted semi-interviews to explore participants' lived experiences with the device, focusing on their self-perceived perceptions of the device and its impact on motivation and adherence with rhythmic activity.

Recruitment of Participants

Participants in this study were recruited through convenience sampling at the Swedish School of Sport and Health Sciences. An open invitation was distributed to those who attended the physical exercise sessions 3 times a week at the Swedish School of Sport and Health Sciences. Since the WRF device was originally designed for athletes to improve their performance, we recruited physically active older adults to ensure that the device was considered meaningful and safe for the user group to allow a phenomenological approach to their experiences [47,48]. Other inclusion criteria were (1) participants aged older than 65 years, (2) able to exercise independently without assistive devices, (3) own a mobile phone that can download the app and could carry the phone during PA, (4) able to adequately see and hear (with or without a hearing aid) the wearable devices' RTF, and (5) able to read and speak English to follow the WRF's instructions.

In total, 21 older adults were recruited, and 18 of them completed all 3 stages in the study. The sample size was determined based on the principles of qualitative research and purposive sampling, where having 20 participants is often

XSL•FO
RenderX

considered a saturation point after which little or no new information is generated from 1 participant group [49].

Data collection

Questionnaires

At the beginning of this study, a questionnaire designed and based on guidelines from the Swedish National Board of Health and Welfare [50,51] was distributed for older adults to subjectively assess their health status and PA (Multimedia Appendix 1). Weekly PA time with multiple-choice questions for both high- and low-intensity activities was reported. Self-perceived health is measured using a 5-point scale covering self-perceived physical health. Other information includes age, gender, marital and family status, living arrangements, and finally, a section for additional comments. In this study, the questionnaire results were used as background information on the participants' health and PA.

Semistructured Interview

We conducted 2 rounds of semistructured interviews with the participants before and after they used a WRF for 3 weeks. All interviews, lasting approximately 45-60 minutes, were audio-recorded and later transcribed for analysis. Interviews were conducted by 2 researchers, and they were collected as individual in-person interviews. Interview guidelines were prepared for both interviews to ensure that the procedure was similar in all interviews (Multimedia Appendices 2 and 3).

In the first round of interviews, participants were asked about their life situations and daily routines in PA, including questions about their PA patterns, motivations, challenges, and perceived health. Questions were asked about their previous use of digital health technologies and wearables. For those who had used wearables previously, we asked about the contexts, benefits, and challenges of wearables use, while nonusers of wearables were asked about their expectations and willingness to adopt such technologies. This interview also touched upon the social aspects of wearables for PA, including preferences for group exercise and peer advice on technology adoption. In this round of interviews, we gathered comprehensive qualitative data about participants' relationships with both PA and technology (Multimedia Appendix 2).

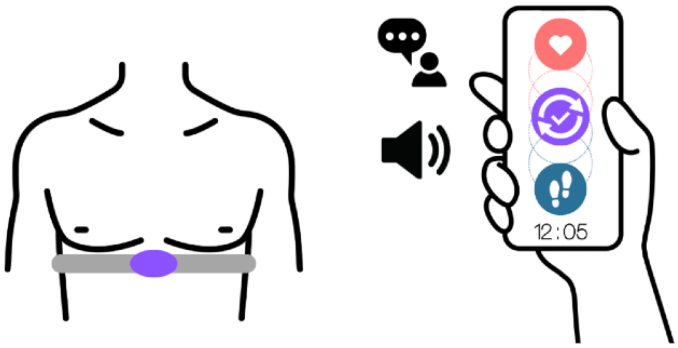
After participants had used the WRF for 3 weeks, they attended the second round of interviews. Participants were asked about their experiences, perceptions, and attitudes toward the device based on their usage of it in rhythmic activity (walking, jogging, or running). Participants were encouraged to reflect in detail on their interactions with the device, including possible challenges. Questions relating to user experience sought to understand participants' motivations and emotional responses toward WRF and capture the subjective, contextualized characteristics of their engagement. We probed participants to evaluate the RTF and whether or not it was seen as supportive. Usability-focused questions assessed the ease of use and technical barriers. Finally, we addressed acceptance and future use of WRFs to inspect their openness to long-term adoption for more advanced interactions, such as conversational features or personalized feedback mechanisms (Multimedia Appendix 3).

In-Situ Study of the Device

In-situ study is valuable for exploring health-related experiences because it honors participants' situated knowledge and lived contexts [52,53]. Engaging with the embodied and sensory dimensions helps to reveal nuanced insights into how contextual factors shape health and well-being [54,55]. Moreover, this method allows participants to actively take control of their experiences and foster their agency [56,57]. This aligns with our intention to offer participants more flexibility and autonomy throughout the study.

The WRF used in this study, Counterpace (Pulson Inc), is a commercially available novel wearable that guides wearers to step precisely between heartbeats during rhythmic activity. Its chest-worn electrocardiography sensor detects the heart state (contract or relax), and an accelerometer detects the walking state (stepping). The connected mobile app provides visual information (step rate, heart rate, and synchronization time and percentage). It also provides voice instructions and metronome beats for wearers (Figure 2). By synchronizing step pacing with heartbeats, the device promotes a step-by-step movement that we define as *rhythmic activity* in this study. There are 2 main reasons why this device was chosen. First, this WRF is developed for rhythmic activity, which is expected to positively influence cardiovascular health [30], with possibly greater benefits for older adults due to increased vascular stiffness [58]. However, its usage and benefits for older adults remain underexplored. Second, although this WRF was originally developed for elite runners and high-performance training, its design has not considered accessibility or inclusivity.

Figure 2. Schematic diagram of the WRF—it consists of a chest-worn heart rate and step rate sensors and a mobile app, which provides both real-time visual and auditory feedback. WRF: wearable device with real-time feedback.



XSL•FO
RenderX

By introducing this technology to older adults, we aim to understand whether its potential benefits could be motivating and meaningful for this population, and to uncover the specific challenges they encounter in using it. Their activity and heart rate were recorded by the mobile app and stored in an anonymized format on a cloud-based service. Participants were encouraged to contact the researchers for technical support throughout the study.

Participants

Of the 21 recruited participants, 18 completed all stages of the study. Table 1 presents their sociodemographic characteristics. The majority (11/18) were between the ages of 70-75 years, with 3 participants aged 66-67 years, and 1 participant older than 80 years, with a mean age of 72.1 years. Most participants were females (12/18). For the PA level, the majority (15/18) reported 90 minutes or more of high-intensity activity per week, with 2 participants reporting an intermediate level (30-89 minutes), and only 1 reported a low level (0-30 minutes). Most participants (10/18) also engaged in 90 minutes or more of low-intensity PA per week. Moreover, 6 participants reported high perceived physical health (scores 1-2), while 4 reported low perceived physical health (scores 5-6).

Table 1. Participants' sociodemographic characteristics, including age, gender, physical activity level, and perceived health (n=18).

ID	Age (y)	Sex	High-intensity PA[a] level (minutes per week)	Low-intensity PA level (minutes per week)	Perceived physical health[b] (score 1-6)
P1	70	Female	≥120	30-89	High
P2	74	Male	90-119	30-89	Medium
P3	70	Female	90-119	30-89	High
P4	70	Male	≥120	90-119	Medium
P5	78	Female	≥120	≥120	Low
P6	72	Male	90-119	30-89	Medium
P7	76	Male	30-89	30-89	Medium
P8	71	Female	≥120	0-30	Medium
P9	82	Male	0-30	30-89	Medium
P10	70	Female	90-119	90-119	Medium
P11	70	Female	90-119	30-89	Medium
P12	72	Male	≥120	≥120	High
P13	78	Female	≥120	≥120	High
P14	72	Female	30-89	90-119	Medium
P15	67	Female	≥120	90-119	High
P16	66	Female	≥120	≥120	Low
P17	72	Female	≥120	90-119	Low
P18	67	Female	90-119	≥120	Low

[a]PA: physical activity.

[b]Perceived physical health was assessed by asking participants, "How satisfied are you with your health?" using a 6-point Likert scale, where 1 indicated very satisfied, and 6 indicated not satisfied at all. The responses were grouped as follows: high=scores 1-2, medium=scores 3-4, and low=scores 5-6.

Analytical Approach

This study used a phenomenologically informed qualitative approach, drawing on the principles of interpretative phenomenological analysis (IPA) [47,59] to explore how older adults make sense of their experiences with wearable technology in PA. IPA is particularly suited to examining how individuals interpret, reflect on, and assign meaning to personally significant experiences, with an emphasis on embodiment, temporality, and the situated nature of meaning-making.

While the early stages of the analysis were guided by a thematic analysis conducted by 3 researchers to navigate a rather large and diverse dataset [60,61], the final interpretative process was anchored in phenomenological inquiry. The goal was not simply to identify recurring patterns but to understand participants' lived experiences in depth, including how they perceived, emotionally responded to, and negotiated the role of wearable devices in their embodied PA routines.

The analysis followed a recursive and idiographic process, beginning with repeated readings of each interview transcript to immerse in the participants' narratives. After preliminary notes were made, segments that revealed how participants made sense of their relationship with WRF were identified and grouped. These segments were later interpreted within the broader biographical, social, and emotional context, maintaining an interpretative stance that sought to balance empathic understanding with critical reflection, a process central to IPA's double hermeneutic process. Although initial thematic groupings

XSL•FO
RenderX

were constructed, these were reevaluated through a phenomenological lens to ensure that the final themes captured the depth, nuance, and individuality of participants' meaning-making. Particular attention was paid to heterogeneity among the narratives articulated by older adults, not as demographic variance, but as diverse experiential positions shaped by bodily awareness, social scaffolding, emotional responses, and contextual life circumstances. The final themes represent not only recurring patterns but rich, interpretative accounts of how wearable technologies were encountered, negotiated, incorporated, or even resisted in older adults' daily lives.

Ethical Considerations

The research has been approved by the Swedish Ethical Review Authority (references 2023-00426-01 and 2024-06934-01). All participants gave informed verbal and written consent for participation in the study and for publication of anonymized extractions from the data. Before participants signed the informed consent, they were given both written and verbal information about the effects of adapting this device in their PA routines and potential risks in participating in the study. Participants were given the possibility to ask questions and withdraw from the study at any time. The privacy and confidentiality of participants were protected by giving all participants a pseudonym (numeric ID) that was used in the analysis and reporting of the results. Participants were given a small financial compensation of a gift card worth 500 SEK (US $54) after they completed their participation in the study. The amount of financial compensation was not specified in the recruitment material.

Results

Overview

The analysis revealed four superordinate themes: (1) use of wearable technologies without RTF in daily life, (2) embodied rhythmic negotiation with RTF, (3) interpretation of health data with RTF, and (4) temporal trajectories of device engagement with RTF (Table 2). The results capture the embodied, temporal, and situated nature of participants' meaning-making processes, revealing how relationships with technology evolved. A detailed analysis table can be found in Multimedia Appendix 4. The findings demonstrate the depth and heterogeneity of lived experiences, showing how wearable adoption is deeply personal and contextual, shaped by participants' ongoing negotiation between technological mediation and their intuitive, embodied understanding of PA.

Table 2. Themes from the interpretative phenomenological analysis.

Superordinate theme	Subordinate theme
Use of wearables without RTF[a] in daily life	• Embodied autonomy and resistance to digital mediation • Contextual and temporal meaning-making • Tensions and ambivalent emotional responses
Wearables with RTF: embodied rhythmic negotiation	• Disrupted natural flow • Learning and guided transformation • Raising bodily awareness through real-time multisensory feedback
Wearables with RTF: interpretation of data	• Relational meaning-making of health data • Technological barriers hindering PA enhancement
Wearables with RTF: temporal trajectories of device engagement	• Disruption and emerging connection in limited engagement • Awareness and growth in moderate engagement • From discipline to integration in extended engagement

[a]RTF: real-time feedback.

Use of Wearables Without RTF in Daily Life

Overview

Participants expressed diverse narratives with wearable technologies, shaped by personal values, bodily autonomy, and lived experience. Many resisted using digital tools, such as tracking apps, prioritizing internal bodily awareness over external data. Engagement was also influenced by social networks and emotional responses, such as stress or pressure. Technology use reflected intentional choices rooted in identity, context, and trust.

Embodied Autonomy and Resistance to Digital Mediation

The lived experience of wearable engagement was characterized by participants' assertion toward embodied autonomy over digital health mediation. This theme represents a profound stance where participants actively resisted and defended their authentic relationship with their bodies against what is perceived as technological intrusion. This theme emphasizes that participants' adoption of wearables goes beyond simple preference, revealing a deeply held commitment to preserving their agency and embodied experience in health and wellness practices.

Most participants owned one or more wearables but deliberately remained disconnected from them, manifesting as "ownership without utilization," where possession of the technology coexisted with intentional non-engagement. For instance, in P2's statement, we see more than simple disuse; we see intentional distancing from technology.

> *I have a smartwatch, but I do not want to use it during physical activity. I keep it somewhere at home.* [P2, male, 74 years]

XSL•FO
RenderX

The way participants physically removed themselves from their devices reflects a conscious choice to make the technology irrelevant in their daily lives, expressing a more profound desire to protect their genuine, lived experience of the body and to maintain psychological boundaries against letting technology shape their bodily awareness.

The experience of external pressure from family, friends, or PA instructors to adopt wearables revealed another dimension of resistance. One participant's steadfast position of remaining uninterested was evident as he was,

> *…uninterested in adopting wearables, even though my son uses a smartwatch and tries to persuade me to wear it too. I do not think it is that important to know how many steps (I took during my walks) because I already feel it roughly.* [P4, male, 70 years]

This illuminated how this resistance was not rooted in inaccessibility, but rather in a principled stance about the characteristics of health monitoring. This resistance is grounded in an understanding that external validation of bodily states has less value compared to their own physical experience.

For some participants, PA was experienced as unencumbered bodily engagement, where additional equipment was perceived as interference with pure movement. One participant articulated this essence with particular clarity:

> *For me, that is why I like jogging and running. I do not want... (to use wearables). It should be easy to train my body... It should be just as easy as possible to move around.* [P10, female, 70 years]

Her lived experience revealed how wearable technology was perceived as disrupting the unmediated nature of exercise. Participants' preference for exercise as "easy" was not about avoiding difficulty, but about maintaining a natural and straightforward connection between mind and body during exercise, suggesting that participants viewed their rejection of wearables as a protective measure for something fundamental to their identity and well-being.

Moreover, the lived experience of health monitoring was characterized by participants' profound trust in their own bodily awareness over technological feedback, revealing a sense of self-awareness grounded in direct physical experience rather than external measurement. One participant explained:

> *I have an Apple Watch, which I do not use much... I look at my heart rate because of the type of exercise I do. But I am kind of structured anyway.* [P14, female, 72 years]

Her experience illuminated how participants privileged their embodied and intuitive understanding of their bodies over external technological validation. Their narrative suggested that they possessed an internal system for understanding their physical needs and responses, rendering external monitoring redundant. This represented not only the rejection of health monitoring but also a commitment to internally generated, experientially based health awareness that participants trusted more than digital data.

Contextual and Temporal Meaning-Making

Participants' engagement with wearables unfolded as profoundly contextual and temporal, where meaning and utility shifted across different life circumstances and personal preferences. Participants' interactions with technology were not static, but a dynamic process of meaning-making that they continually reflected upon in relation to contextual needs and values. The temporal dimension of this engagement also suggests that relationships with wearable technology were biographically embedded within participants' life stories and subject to reinterpretation over time.

Contextually, the technology preference emerges through participants' definition of what "useful health monitoring" entails, revealing highly individualized criteria. One participant articulated:

> *(I found) a basic step-counter in the form of a phone-based application sufficient to track most basic physical activities.* [P1, female, 70 years]

Her use of the word "sufficient" was particularly intriguing, which shows that she had developed precise measurements for what constitutes adequate, allowing her to be free from external recommendations or marketing messages about optimal health tracking. This preference revealed how participants proactively set their own boundaries around technological complexity and determined what digital tools aligned with their technological sophistication, necessity, and manageability.

Additionally, selective engagement with personal data interpretation emerged as participants demonstrated discernment about which aspects of wearable technology resonated with their personal understanding of health and movement. One participant described his experience:

> *(During walking and running) It is fun to see how far you have gone, how fast you have gone, what kind of mileage you have got…* [P4, male, 70 years]

This illustrates how enjoyment emerged from specific data points that connected with his personal measurement of a meaningful activity. His emphasis on "fun" shows that the engagement was not just driven by arbitrary fitness imperatives but by his intrinsic interests in tracking and inquisitiveness in his performance. This selective engagement demonstrates how participants acted as curators of their technological experience by choosing to interact with elements that enhance their PA.

On the negative side, abandonment or pause of usage occurred due to temporal change in their relationships with wearables, disclosing how the purpose and utility of these devices were subject to changing life circumstances, which vary from the loss of family members, injury, chronic diseases, relocation, and so on. Another participant described how her usage of wearables shifted over time:

> *I used to have one (Garmin watch) many years ago when I was running. It was useful then, but now I do not use it anymore.* [P13, female, 78 years]

The phrase "*it was useful then*" implies that usefulness was not an inherent property of the technology, but rather a mixture of the device's capabilities and the participant's needs at that

XSL•FO
RenderX

particular period in her life. Ultimately, the disengagement with wearable devices was the natural evolution over time, and could not be indicative of their usability or utility.

Finally, we discovered a technological curiosity among participants who have not tried any wearable devices, as evidenced by their openness to digital health tools, particularly those that offer well-structured guidance for PA. A participant expressed that,

> *I appreciate and I believe in technology. And I had tried several fitness apps previously…* [P3, female, 70 years]

This highlights a broader exploratory relationship with digital health resources. The experimental approach to "try and find" a new tool indicated that participants do not regard technology adoption as a firm commitment to 1 tool, thereby retaining control over their engagement. This sense of curiosity also reflects active meaning-making, as participants need various technologies and can identify those that resonate with them, rather than simply consuming them passively.

Tensions or Ambivalent Emotional Responses

The use of wearables among older adults was significantly influenced by their social and emotional aspects. This subtheme illuminated the social nature of technology adoption among this demographic, which was often embedded within and mediated through personal networks and communities. Similarly, their engagement was marked by complex, often conflicting feelings, demonstrating that wearables are not a neutral tool but an emotionally charged artefact that can have a profound impact on participants' sense of self and their relationship with their bodies.

The social dimension of wearable adoption was notably influenced by participants' trust toward others, who served as mediators of acceptance and attachment. One participant described how her wearable usage began through professional guidance:

> *I got to know about my smart ring from a health engineer. She is also very skilled in various aspects of health. And she was the one who suggested this ring.* [P17, female, 72 years]

P17's emphasis on the health engineer's expertise suggested that this referral carried weight in her decision-making because it came from individuals whose judgment she believed. The technology gained legitimacy through respected social connections, and at the same time, this mediation represented not simply an information transfer but also the validation of technological benefits.

Some participants also mentioned attending webinars and lectures on new technologies, suggesting that older adults' technology adoption was facilitated through community-based learning environments where they could explore within supportive contexts. This furthermore shows that older adults' engagement with wearables was fundamentally relational rather than individualistic. This social scaffolding appeared crucial for older adults who might otherwise feel intimidated by or skeptical of wearable technology, providing both practical guidance and emotional support for technology exploration.

Participants demonstrated remarkable emotional awareness in expressing these complex relationships with wearables. We found that accessing and tracking health data is a double-edged sword, raising paradoxical feelings. Participants, such as P6 (male, 72 years) and P15 (female, 67 years), reported feeling negative emotions, including stress, confusion, and anxiety, after noticing an unexpected decline in their PA levels. Their experience exemplifies how data can disrupt their perceived well-being. More explicitly, P15's experience illuminated the tension between the potential benefits and its emotional costs:

> *I have a step calculator (on the phone). But I do not want to check it all the time. It is competitive and stressful... I think that is stressful.* [P15, female, 67 years]

P15's experience revealed how the pressure of goal-oriented tracking conflicted with her authentic relationship to PA, leading her to limit her engagement. The repetition of "stressful" in her statement emphasized that constant monitoring could hinder her view of the technologies. Her description of tracking as "competitive" suggested that it is hard to disconnect the quantitative data from external standards, such as suggested steps per day, even though well-being should be very individual and personalized.

Wearables With RTF: Embodied Rhythmic Negotiation

Overview

Participants experienced RTF as both disruptive and transformative. While initial use challenged authentic movement patterns, some developed new bodily awareness. Engagement was shaped by sensory feedback, usability, and the need for support in interpreting health data. The technology offered both opportunities for connection and barriers to meaningful use.

Disrupted Natural Flow

Participants' experiences with WRF revealed a tension between guided synchronization and their authentic rhythm. To synchronize their heart rate and step rate, there was a disruption to their PA flow. Participants became conscious about how they move, and their exercise session became a deliberate practice. For participants, the device's requirements created a confrontation with their bodily rhythms. As P8 (female, 71 years) reflected, the expectation to "*go with your heart*" triggered a realization about her relationship with her body:

> *I have realised that I have never followed my heart. I walk much faster.* [P8, female, 71 years]

This shows how the synchronization challenge exposed participants' disconnection from their internal rhythms.

Moreover, a sense of inauthenticity and bodily disharmony prevailed when they attempted synchronization. Participants described feeling caught between complex demands, including their original moving patterns, environmental constraints, and the device's rhythmic imperatives. All the demands combined led to a cognitive burden for participants, which requires

XSL•FO
RenderX

constant attention and internal negotiation. Participants experienced this burden as fragmenting their usual unrestricted movement. As P18 (female, 67 years) described, "*even momentary lapses in attention resulted in failing synchronisation*," revealing how the device demanded a sustained cognitive effort with bodily rhythm that felt foreign to them.

Participants also experienced synchronization widely across different movement contexts. Flat terrain emerged as providing the most conducive environment, while stairs and uneven surfaces created barriers to synchronization. This suggests that the experience of rhythmic alignment is deeply contextual, shaped by the interplay between bodily capacity, environmental demands, and technological requirements. Environmental factors thus intensified their cognitive burden. Icy conditions, uneven terrain, and slopes created more layers of complexity that further estranged participants from feeling natural. P13's (female, 78 years) experience of "*when walking downhill, I need to slow down my steps to remain in sync*" illustrates how environmental contexts compounded the sense of imposition.

Learning and Guided Transformation

Despite cognitive struggles, some participants experienced a meaningful transformation in their relationship with synchronization. Their learning experience emerged as a gradual integration of guided rhythm in PA. P3 (female, 70 years) described the experience "*becoming like a dance*" once synchronized, suggesting a qualitative shift from effortful practice to embodied harmony. This transformation necessitated a reorientation in how participants perceived their bodies and movement. The warming-up process described by P3 also suggests that synchronization required not just physical preparation, but also an attentional preparation that allowed for the integration of heart rhythm awareness into movement consciousness.

The desire expressed by some participants (P12 and P14) to extend synchronization to other activities not included in the connected app, such as gym workouts or group settings, indicates an emerging appreciation for the device's potential in various scenarios. However, they noted a lack of guidance for these contexts as problematic, and customized guidance according to each activity will be needed, highlighting how rhythmic experiences could be further explored.

Raising Bodily Awareness Through Real-Time Multisensory Feedback

For participants, audio feedback served more than a technical feature; it is a medium that cultivates and maintains embodied awareness. The sound guidance raises present-moment consciousness, drawing participants into a close connection with their rhythmic movements. P3 explained the bridge between technological instruction and bodily understanding:

> *I focused on the beat and the sound. And there it (the synchronization) was much better.* [P3, female, 70 years]

This audio-mediated awareness helps to orient the walking experience into a contemplative practice. The synchronization between sound, step, and heart rate interweaves into a dynamic movement, where participants learned to listen and resonate not just with their ears but with their entire moving body.

Notably, this dynamic movement was fragile, easily disrupted by environmental distractions or cognitive overload, as mentioned in the Disrupted Natural Flow section, revealing how audio-guided synchronization demanded sustained and focused presence. To help with audio-guided synchronization, the integration of visual feedback created a richer source of engagement and accomplishment. Visual cues, including the changing numbers and colors, functioned as an immediate representation of participants' efforts, making abstract concepts of "being in sync" intuitively observable. With the interactive app interfaces and sounds, the compound feedback system gains a more meaningful and motivating guiding quality.

The device's capacity to provide a real-time compound feedback functioned as an extension of participants' bodily awareness during exercise. P13's experience demonstrated how RTF could recalibrate habitual movement patterns: "*Otherwise, I would walk too fast all the time*." This suggests that the device helps develop enhanced self-awareness, helping participants recognize and modify unconscious movement tendencies. For P1, the discovery of elevated heart rate during brief exercise periods created a sudden awareness of her body's responses that had previously remained hidden. This demonstrates how technological feedback could generate moments of surprise, revealing aspects that might otherwise remain unconscious. The real-time nature of the feedback opened up opportunities for immediate embodied learning, where wearers integrated technological information into the lived experience.

Wearables With RTF: Interpretation of Data

Overview

Participants' relationships with health data revealed intricacy between curiosity, confusion, and desire for meaning-making. The pulse data held significance as a direct representation of their most intimate bodily rhythm by making their heartbeat visible and quantifiable. This transformation of internal and felt experience into external and observable data created both fascination and interpretive challenges.

Relational Meaning-Making of Health Data

The participants struggle to understand how the data represent their health. Participants thus expressed a need for consultation and interpretation of data. P18's reflection captured this essential need:

> *It's fun (to get the data). But with the data, I would like to have a discussion (with others or professionals), so I can also understand the data.* [P18, female, 67 years]

This desire revealed that health data was not merely informational but deeply relational for these participants. The meaning of their physiological information emerged not in isolation but through conversation, interpretation, and contextualization within their broader life experiences.

XSL•FO
RenderX

The absence of data interpretation left participants with pieces of embodied information that lacked coherent meaning or actionable significance. P6's confusion, "*I do not understand why I get this information*," also highlighted how health data without interpretive support could create disorientation rather than empowerment. This suggests that meaningful engagement with data requires not just data provision but spaces for sense-making, where technological information becomes integrated into participants' knowledge of their own well-being.

Technological Barriers Hindering PA Enhancement

Participants' encounters with technical challenges revealed how technological complexity could fracture the intended engagement. P7's struggle with volume settings exemplified how technical barriers could create frustration. Also, P11's frustration revealed how technological barriers could create alienation from one's own embodied data, transforming what should be empowering information into a source of confusion and inadequacy.

> *I could not find information, like my heart rate, since I didn't know where to look for it.* [P11, female, 70 years]

These moments of technological breakdown highlighted the fragility of the wearer-device relationship and the importance of seamless integration for meaningful embodied engagement, so that the technological exclusion will not alienate the participants from the device's capabilities due to the technical limitations. This tension highlights the mismatch between technological promise for enhanced bodily awareness and the reality of technical complexity for wearable devices.

Wearables With RTF: Temporal Trajectories of Device Engagement

Overview

During the 3-week study period, the 17 participants (P1-P18, excluding P8) showed highly variable engagement with the device. While the recommended usage was 30 minutes per session, 3 times weekly (270 minutes total), actual usage time ranged from 51% to over 700% among the participants. P2 used the device least (137 minutes, 51% of recommended), while P9 showed the highest engagement (1905 minutes, 706% of recommended). On average, participants used the device for 720 minutes, nearly 3 times the recommended amount (267%).

We grouped participants into 3 categories based on total usage time—limited engagement (≤150% of recommendation) included 6 participants (P2, P3, P5, P11, P13, and P14), moderate engagement 150%-300% of recommendation) comprised the majority with 7 participants (P1, P4, P10, P12, P15, P16, and P18), and high engagement (≥300% of recommendation) included 4 participants (P6, P7, P9, and P17). The threshold of 150% and 300% was chosen based on the empirical distribution of the data and interpretability considerations. When we examined the actual usage patterns, they emerged as a natural break point that meaningfully separated participants with different engagement levels with the device.

Session patterns also varied considerably, with some participants preferring fewer but longer sessions (eg, P4 and P9 averaged over 60 minutes per session) while others engaged more frequently with shorter durations (eg, P15 and P16). Detailed engagement data and participant groupings are provided in Multimedia Appendix 5.

The 3 groups of device usage time became a lens through which different dimensions of the lived experience unfolded. Wearer-device relationships evolved with usage time. Limited engagement often brought frustration, moderate use led to reflection, and extended engagement fostered mastery and integration. Over time, some users shifted from resistance to routine, showing how long-term interaction could transform technology from a burden into a meaningful part of PA.

Disruption and Emerging Connection in Limited Engagement

For participants who engaged themselves for shorter durations (≤148% of the recommended 270 minutes, ie, ≤400 minutes), the experience was characterized by a critical awareness of bodily disruption and cognitive demand. P11 (97%) described her encounter with the device as requiring her to alter her way of exercising:

> *I am a fast walker, so walking too fast got uncomfortable (to synchronize with my heart rate). It required too much effort and focus to follow the exact goals.* [P11, female, 70 years]

This disruption of habitual movement patterns created a tension between her natural bodily rhythm and the technologically mediated pace. P13 (112%) experienced a similar disruption of her natural walking cadence, describing the necessity to walk more slowly as counterintuitive:

> *My heartbeat often was between 70 and 80, and that is really slow to walk. But it was hard to walk that slowly... You have got to go really slowly to be in time, in sync.* [P13, female, 78 years]

Despite this challenge, she found meaning in the device's auditory feedback, which became a source of guidance and positive effect:

> *It is helpful when (the sounds) tell you when you are not in sync... There is a happy telephone sound when you are in sync.* [P13, female, 78 years]

P3 (132%), despite her limited usage time, experienced a profound transformation in her relationship with the device over time. Her lived experience evolved from initial struggle to proficiency:

> *The first week, I was 69% synced. The following week, I achieved 95% sync, and in the last week, I reached 97% sync.* [P3, female, 70 years]

This progression revealed how even limited temporal engagement could yield significant shifts in competence and connection with the technology.

Awareness and Growth in Moderate Engagement

Among participants who used the device between 148% and 333% of the recommended dosage (approximately 400-900

XSL•FO
RenderX

minutes), no particular positive or negative feelings were articulated toward the device. In this group, participants expressed a neutral observational attitude toward their synchronization performance. Interestingly, their awareness of the synchronization fluctuations during each session sometimes raises curiosity for them to reflect on their embodied experience. For example, P18 (156%) was fascinated by the dynamic, ever-changing nature of synchronization:

> *That is interesting to see how it fluctuates all the time... it really went, from like 61 (to) 68 and then it was 64 and then it went up to 78 and it was really up and down.* [P18, female, 67 years]

This temporal awareness revealed how moderate engagement fostered a contemplative viewpoint toward her own bodily processes.

Throughout the entire study period, P12 (272%) also experienced a changing synchronization rate, marking a journey of progression. His extended engagement with the device brought joy in improvement and a sense of achievement for him:

> *In the beginning, this was a hard day. Only 21%... but now this is 90%! Perfect.* [P12, male, 72 years]

Extended usage thus for P12 became a narrative of personal growth and technological competence.

From Discipline to Integration in Extended Engagement

All participants who engaged with the device for extended periods (≥333% of the recommended dosage, ie, over 900 minutes approximately) developed strategies and demonstrated discipline when interacting with it, creating a nearly self-training process.

P16 (348%) expresses that she put in intense focus, which demanded a withdrawal from environmental engagement:

> *You must be so focused, extremely focused. If you see a nice sky or some flowers, you are not in focus again... You need to always concentrate on this device instead of this environment.* [P16, female, 66 years]

Her experience revealed the demanding nature of long-term synchronization, requiring a narrowing of perceptual attention.

P7 (567%) experienced the longest engagement with the device and developed a profoundly different relationship with the technology. For him, extended usage transformed the device from a demanding technological interface into a meditative companion:

> *In the end, I found that sound pleasant. It calmed me down... It reminds me of meditation... I did not have to think about anything.* [P7, male, 76 years]

His extended temporal engagement allowed for the emergence of a peaceful relationship with a familiar, internalized synchronization process.

Discussion

Principal Findings

In this study, we have explored the lived experiences of physically active older adults in using wearables for rhythmic activity, with a specific attention to RTF as an interaction modality. The study explored older adults' previous use of wearables to track their PA, adherence with WRF during rhythmic activity, and the interactive features of WRF in enhancing the synchronization between step rate and heart rate. The study identified four main themes characterizing older adults' lived experiences with wearables that are (1) use of wearable technologies without RTF in daily life, (2) embodied rhythmic negotiation with RTF, (3) interpretation of health data with RTF, and (4) temporal trajectories of device engagement with RTF.

While previous studies on wearables with older adults have recognized diverse motivations for adoption [11,12,33-35], this study reveals the specific mechanisms underlying this diversity in the context of real-time rhythmic guidance. Rather than treating heterogeneity as a static characteristic, we demonstrate how varied engagement patterns emerge through the dynamic intersection of personal values, embodied experiences, and technological demands. This study's contribution extends beyond documenting diversity to explaining how older adults' relationships with wearable technology are fundamentally shaped by their commitment to preserving authentic bodily experiences, their reliance on social networks for technology validation, and their need for contextually adaptive feedback systems that respect individual rhythmic patterns.

The study shows that the discontinuous use of wearables, accompanied by self-reported challenges in the device-guided synchronization, can be partly explained by the mismatch between technological capabilities and older adults' inherent values and aspirations in later life. Factors such as comparison, competition, and goal setting that are often inherently built into commercial wearables [62] are less important for this demographic, for whom it is often more important to maintain the existing PA level through inner self-motivation. Simultaneously, device-guided synchronization requires older adults to change their natural walking rhythm and pace, which may result in a sense of discomfort and frustration. These age-specific requirements for achieving device-guided synchronization have largely remained unexplored in previous studies [30,58]. To overcome these challenges, the study proposes that interaction modalities that guide users to increase their bodily awareness may help older adults to find the right walking pace authentically.

The study addresses that older adults' desire to use wearable devices stems mostly from their health-related goals and preferences, rather than their need for constant digital monitoring. These personal goals and preferences may show high variance in later life, which can partly explain the heterogeneity in driving factors to use wearables. Importantly, the findings reveal more than simple disuse but an intentional distancing from technology. This distancing aligns with research showing that older adults may abandon wearables when they

XSL•FO
RenderX

are framed as assistive technologies that compensate for age-related losses, leading to negative user images [36]. Moreover, wearables can be perceived as motivating or condescending depending on how freely older adults are able to make autonomous choices about PA [34]. As a result, technology adoption among older adults involves complex value-based decision-making processes that prioritize the preservation of unmediated bodily experience over external validation.

Individuals' readiness to accept wearable technology varies depending on their interests, experiences, and sociocultural background [32,63]. This study indicates that personal goals, often focused on the enjoyment of PA, were stated as the primary motivator to use wearables. These personal goals may sometimes contradict the device-related features that emphasize self-monitoring or goal setting in PA [34], which may partly explain the low or limited use of wearables in PA among older adults.

Previous research has explored the use of wearables by focusing on older adults' motivations and driving factors at the individual level or from the device-centric perspective [11,33,35]. The study adds to this line of research by demonstrating the social aspects and motivations in the use of wearables for PA. The results demonstrate that technology adoption emerged not as an individual consumer choice, but as a socially embedded practice mediated through relationships, community connections, and shared experiences. This relational orientation, evidenced through participants' learning via community seminars, family introductions, and the emotional significance attached to devices, challenges dominant narratives that frame wearable technology use as primarily self-directed and autonomous. This finding supports previous research indicating that formal and informal support and guidance in accessing and continuing device use are connected to long-term adoption [37]. The relational dimension extends beyond initial adoption to ongoing meaning-making, where device data interpretation requires dialogical engagement with professionals or trusted others. The combination of individual motivation, ease-of-use, and perception of added value ultimately determines whether devices will be integrated into users' lives [11,33]. This shows that the role of family members or friends, also known as "warm experts" [17], remains significant in the uptake of wearables. Consequently, social support in the uptake of wearables, which is often unequally distributed in later life [64], could further diversify the heterogeneous use of wearables in PA.

The negative narratives we observed serve as a vital reminder that technology and health data can alter participants' relationships with their bodies, shifting their view from an intuitive and instinctive feeling to one of external judgment and evaluation. The fact that decreased activity levels led to stress, rather than simply providing information or motivation, suggests that when technology introduces an evaluative aspect to PA, it might be emotionally challenging if not carefully designed. This emotional dimension is critical, as research demonstrates that long-term attachment with wearables requires devices to elicit positive emotional responses among users [40], and that emotional, sensory, and social influences significantly impact user engagement with wearables [13,38,39]. A long-term, motivating engagement with wearables also requires participants to develop strategies for managing the emotional impact, involving selective engagement and conscious boundary-setting around metrics. Fostering personalized approaches that preserve their sense of agency will be the key to selectively engaging with technologies that align with their values and needs.

Lived Experiences of Wearables With RTF

The study reveals the significant age-related variations in user experience and adaptation to the wearable with RTF. Central to these findings is the recognition that rhythmic synchronization is deeply contextual, shaped by the interplay between bodily capacity, environmental demands, and technological requirements. The results demonstrate how synchronization challenges were not uniform technical difficulties but varied significantly based on environmental factors (flat terrain vs stairs and uneven surfaces), individual bodily characteristics (natural walking pace and heart rate variability), and temporal factors (time needed for adaptation). This aligns with research showing that through sound interaction, users can coordinate and stabilize movements with internal or external imposed rhythms [28], and that auditory feedback allows users to be observant of the environment and fits into more dynamic settings compared with visual and tactile guidance [27]. This contextual nature means that synchronization cannot be understood as a purely technical challenge, but rather as a complex negotiation between multiple intersecting factors that are dynamically configured in each exercising session.

Regarding synchronization challenges, older adults encountered difficulties synchronizing their steps to the audio feedback designed to align with their heartbeats. These challenges stemmed from a mismatch between the device's requirements and the cognitive load, as well as from environmental constraints. Participants reported that synchronizing often required them to walk at an unnaturally slow pace, which conflicted with their habitual walking styles. This resulted in a perception that the device's requirements were intrusive and disruptive to the natural flow of walking. Despite these challenges, older adults demonstrated an ability to improve their synchronization over time. This improvement, however, was not uniform across participants and appeared to depend on the individual's personal exercise routines.

The study indicates that device-guided synchronization in rhythmic activity can be associated with older adults' previous routines and rhythms, and thus connected to temporalities of PA. These insights suggest that while initial engagement with the device may be frustrating, persistence and strategic adjustments could enhance the experience of synchronization for some users. These adjustments, however, require "bodily knowledge and know-how" [65] to be able to adapt the device to already existing exercise routines. As Phoenix and Bell [66] highlight, disruptions in the natural movement rhythms may be associated with bodily discomfort and emotions, including symptoms associated with chronic health conditions connected to aging. These findings highlight that the device-guided synchronization, despite being a physiological mechanism between heart and gait cycles [30,58], is always connected to wider routines, rhythms, and temporalities in PA, as well as

XSL•FO
RenderX

older adults' previous exercise background and physical capabilities. In this study, the interconnectedness of synchronization challenges in relation to the environmental or activity-related factors was repeatedly pronounced in participants' statements. This interconnectedness between synchronization and contextual and environmental factors may explain the reason for participants' frequent statements of synchronization challenges, despite the fact that the device itself was considered relatively easy to use.

Overall, the device created both disruption and possibility, disrupting established movement patterns while opening space for new forms of bodily awareness and rhythmic experience. The challenge of synchronization revealed the depth of participants' disconnection from their internal rhythms and their capacity for developing integration when given appropriate support and practice. This dual nature is inherent to RTF mechanisms, which provide haptic, auditory, or visual feedback simultaneously with user movement rather than after the activity has ended [2,3]. Notably, RTF has been identified as a preferred feature among older adults in wearable technologies [11]. In this sense, wearable devices with RTF function not merely as monitoring tools, but as mediators that can either fragment or enhance the relationship between consciousness and embodied movement, depending on how they are designed and implemented.

Meaning-Making of Multisensory RTF

The meanings associated with the interactive features of WRFs extended beyond functional utility. The study addresses the potential of real-time audio feedback to help older adults overcome synchronization challenges and adapt to device-related functionalities over time.

First, RTF in auditory format can help older adults to reduce cognitive load in synchronization, which was considered to require an intense focus on finding and maintaining the right walking pace. Previous studies, respectively, have explored the auditory feedback in relation to increasing the enjoyment in walking activity, and showed that auditory cues could enhance the health benefit from walking among older adults through acting as a positive reinforcement to a rhythmic stepping pattern [67]. However, the degree of adherence to these feedback modalities depends strongly on individual characteristics, such as technological familiarity, cognitive capacity, and even sensory limitations. According to this study, older adults with more previous experience with technology tend to engage more easily with both visual and auditory feedback.

Second, the study shows that compound feedback, which combines visual and aural cues, improves synchronization and has a positive impact on the overall experience. Although its effectiveness may vary depending on individual needs and preferences, this strategy is especially helpful in fostering positive emotions like motivation and enjoyment. In order to accommodate a greater range of sensory abilities and preferences, wearable device design must be adaptive to an individual's skills and capabilities. For example, older adults with hearing impairments may find auditory feedback less effective, while those with visual impairments may prefer larger visual cues.

Finally, temporal engagement with the device was not merely quantitative but fundamentally shaped the qualitative nature of participants' lived experiences. Shorter engagements were characterized by disruption and effortful adaptation, while extended engagements allowed for the development of more integrated, contemplative, or mastery-oriented relationships with the technology. The duration of usage emerged as a critical factor in determining whether participants experienced the device as a disruptive intervention or as an integrated tool for embodied self-awareness, and potentially a natural component in PA. This temporal dimension is particularly significant given that adherence, defined as long-term user acceptance and engagement with devices, remains a persistent challenge in wearable technology research [4-6]. The individual factors of motivation, ease-of-use, and perceived added value that determine device integration into users' lives [11,33] appear to require extended time periods for meaningful development. To sum up, technology adoption among older adults requires sufficient time for the development of new embodied competencies and the gradual integration of technological feedback into existing movement practices. The transformation from initial resistance or struggle to potential acceptance suggests that wearable technology design should accommodate extended adaptation periods rather than expecting immediate integration.

Limitations and Future Research

The study gathered older adults with relatively high levels of PA and high satisfaction with their physical health. Therefore, their user experience with the device could potentially be influenced by their previous PA level. The participants used the device only for a 3-week period, and the long-term user engagement remained unexplored. Despite these limitations, this study has extended previous research by considering the lived experiences of older adults in device-guided rhythmic activity, with a specific focus on audio feedback during synchronization of step and heart rate. Future research should explore device-guided synchronization and adherence to RTF with a more diverse group of older adults. This would ensure that wearable technologies for self-monitoring can reach their expected health benefits for those older adults who are at risk of cardiovascular disease.

Conclusions

This study reveals that older adults' engagement with wearable devices for rhythmic activity is characterized by 4 fundamental dimensions that challenge conventional assumptions about technology adoption in later life. First, rather than simple disuse, older adults demonstrate intentional distancing from wearable technologies, actively resisting digital mediation to preserve authentic relationships with PA. Second, their engagement is fundamentally relational rather than individualistic, emerging through trusted social networks and requiring ongoing dialogical support for meaningful data interpretation. Third, the experience of rhythmic alignment is deeply contextual, shaped by dynamic negotiations between bodily capacity, environmental demands, and technological requirements that vary significantly across individuals and situations. Fourth, temporal trajectories of engagement reveal that meaningful relationships with wearable

XSL•FO
RenderX

technology develop through extended interaction periods, with different engagement durations producing qualitatively distinct experiences ranging from disruption to integration.

These findings extend previous research on wearable technology use among older adults by moving beyond motivational factors to reveal the underlying experiential mechanisms that shape engagement patterns. The study demonstrates that device-guided synchronization creates both disruption and possibility, requiring extended temporal engagement for participants to develop integrated relationships with technological feedback. While RTF shows potential for enhancing bodily awareness, its effectiveness depends on designs that accommodate older adults' commitment to embodied autonomy, their relational approach to technology validation, and the contextual nature of their rhythmic experiences. Future wearable technologies must therefore prioritize adaptive, socially embedded, and temporally sensitive design approaches that respect rather than override older adults' established movement practices and values.

Acknowledgments

We would like to thank all research participants for their invaluable contribution to the study. We would like to thank Jie Che for her participation in the data collection.

Funding

The research has been funded by the Kamprad Family Foundation (reference 20233117) and the Åke Wiberg Foundation.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

TYL: writing – original draft, writing – review and editing, visualization, data curation, and formal analysis
JC: data curation and formal analysis
AR: data curation, formal analysis, and writing – original draft
SAD: conceptualization, writing – original draft, supervision, and funding acquisition
SK: conceptualization, writing original draft, writing – review and editing, supervision, funding acquisition, and project administration.
All authors have accepted the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire Design.
[DOCX File , 13 KB - mhealth_v14i1e71509_app1.docx]

Multimedia Appendix 2
Pre-interview template.
[DOCX File , 17 KB - mhealth_v14i1e71509_app2.docx]

Multimedia Appendix 3
Post-interview template.
[DOCX File , 15 KB - mhealth_v14i1e71509_app3.docx]

Multimedia Appendix 4
Thematic analysis (themes, quotes, and commentary).
[DOCX File , 11 KB - mhealth_v14i1e71509_app4.docx]

Multimedia Appendix 5
Device use among participants (n=18).
[DOCX File , 9 KB - mhealth_v14i1e71509_app5.docx]

References

XSL•FO
RenderX

1. Devereux-Fitzgerald A, Powell R, Dewhurst A, French DP. The acceptability of physical activity interventions to older adults: A systematic review and meta-synthesis. Soc Sci Med 2016;158:14-23 [FREE Full text] [doi: 10.1016/j.socscimed.2016.04.006] [Medline: 27104307]
2. Van Hooren B, Plasqui G, Meijer K. The effect of wearable-based real-time feedback on running injuries and running performance: A randomized controlled trial. Am J Sports Med 2024;52(3):750-765 [FREE Full text] [doi: 10.1177/03635465231222464] [Medline: 38287728]
3. Van Hooren B, Willems P, Plasqui G, Meijer K. Changes in running economy and running technique following 6 months of running with and without wearable-based real-time feedback. Scand J Med Sci Sports 2024;34(1):e14565 [FREE Full text] [doi: 10.1111/sms.14565] [Medline: 38268070]
4. Burford K, Golaszewski NM, Bartholomew J. "I shy away from them because they are very identifiable": A qualitative study exploring user and non-user's perceptions of wearable activity trackers. Digit Health 2021;7:20552076211054922 [FREE Full text] [doi: 10.1177/20552076211054922] [Medline: 34820134]
5. Nelson EC, Verhagen T, Vollenbroek-Hutten MMR, Noordzij ML. When it is part of me, I can do it: Using embodied empowerment to predict adherence to wearable self-care technology. Computers in Human Behavior 2024;158:108314 [FREE Full text] [doi: 10.1016/j.chb.2024.108314]
6. Sousa P, Ferreira P, Ventura F, Dixe M, Martinho R, Dias S, et al. Using m-health and wearable technologies to promote self-management and treatment adherence among cardiovascular disease patients: A clinical trial protocol. European Journal of Cardiovascular Nursing 2023;22(1):zvad064.015, [FREE Full text] [doi: 10.1093/eurjcn/zvad064.015]
7. Qiu H, Wang X, Xie F. A survey on smart wearables in the application of fitness. 2017 Presented at: IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech); 2017 Nov 15; New York City p. 303-307 URL: https://doi.org/10.1109/DASC-PICom-DataCom-CyberSciTec.2017.64 [doi: 10.1109/dasc-picom-datacom-cyberscitec.2017.64]
8. Colverson A, Barsoum S, Cohen R, Williamson J. Rhythmic musical activities may strengthen connectivity between brain networks associated with aging-related deficits in timing and executive functions. Exp Gerontol 2024;186:112354 [FREE Full text] [doi: 10.1016/j.exger.2023.112354] [Medline: 38176601]
9. Ma C, Li M, Li R, Wu C. The effect of rhythmic movement on physical and cognitive functions among cognitively healthy older adults: A systematic review and meta-analysis. Arch Gerontol Geriatr 2023;104:104837 [FREE Full text] [doi: 10.1016/j.archger.2022.104837] [Medline: 36257163]
10. Kian KTC, Chang S. Groove into ageing: Exploring the effects of rhythmic exercise on the well-being of older adults. Archives of Gerontology and Geriatrics Plus 2025;2(1):100111 [FREE Full text] [doi: 10.1016/j.aggp.2024.100111]
11. Moore K, O'Shea E, Kenny L, Barton J, Tedesco S, Sica M, et al. Older adults' experiences with using wearable devices: Qualitative systematic review and meta-synthesis. JMIR Mhealth Uhealth 2021;9(6):e23832 [FREE Full text] [doi: 10.2196/23832] [Medline: 34081020]
12. Zhang Z, Giordani B, Margulis A, Chen W. Efficacy and acceptability of using wearable activity trackers in older adults living in retirement communities: A mixed method study. BMC Geriatr 2022;22(1):231 [FREE Full text] [doi: 10.1186/s12877-022-02931-w] [Medline: 35313819]
13. Vargemidis D, Gerling K, Abeele VV, Geurts L. Performance and Pleasure: Exploring the perceived usefulness and appeal of physical activity data visualizations with older adults. ACM Trans Access Comput 2023;16(3):1-35 [FREE Full text] [doi: 10.1145/3615664]
14. Vargemidis D, Gerling K, Spiel K, Abeele VV, Geurts L. Wearable physical activity tracking systems for older adults—A systematic review. ACM Trans Comput Healthcare 2020;1(4):1-37 [FREE Full text] [doi: 10.1145/3402523]
15. McCarren L, Kuoppamäki S. Design preferences, routines, and well-being of older adults using voice-guided digital mindfulness: Qualitative interview study. JMIR Hum Factors 2025;12:e67533 [FREE Full text] [doi: 10.2196/67533] [Medline: 40460434]
16. Taipale S, Oinas T, Karhinen J. Heterogeneity of traditional and digital media use among older adults: A six-country comparison. Technology in Society 2021;66:101642 [FREE Full text] [doi: 10.1016/j.techsoc.2021.101642]
17. Hänninen R, Taipale S, Luostari R. Exploring heterogeneous ICT use among older adults: The warm experts' perspective. New Media & Society 2021;23(6):1584-1601 [FREE Full text] [doi: 10.1177/1461444820917353]
18. Akinosun AS, Polson R, Diaz-Skeete Y, De Kock JH, Carragher L, Leslie S, et al. Digital technology interventions for risk factor modification in patients with cardiovascular disease: Systematic review and meta-analysis. JMIR Mhealth Uhealth 2021;9(3):e21061 [FREE Full text] [doi: 10.2196/21061] [Medline: 33656444]
19. Bove LA. Increasing patient engagement through the use of wearable technology. The Journal for Nurse Practitioners 2019;15(8):535-539 [FREE Full text] [doi: 10.1016/j.nurpra.2019.03.018]
20. Kang HS, Exworthy M. Wearing the future-wearables to empower users to take greater responsibility for their health and care: Scoping review. JMIR Mhealth Uhealth 2022;10(7):e35684 [FREE Full text] [doi: 10.2196/35684] [Medline: 35830222]
21. Canali S, Schiaffonati V, Aliverti A. Challenges and recommendations for wearable devices in digital health: Data quality, interoperability, health equity, fairness. PLOS Digit Health 2022;1(10):e0000104 [FREE Full text] [doi: 10.1371/journal.pdig.0000104] [Medline: 36812619]

XSL•FO
RenderX

22. McKay FH, Slykerman S, Dunn M. The app behavior change scale: Creation of a scale to assess the potential of apps to promote behavior change. JMIR Mhealth Uhealth 2019;7(1):e11130 [FREE Full text] [doi: 10.2196/11130] [Medline: 30681967]
23. McKay FH, Wright A, Shill J, Stephens H, Uccellini M. Using health and well-being apps for behavior change: A systematic search and rating of apps. JMIR Mhealth Uhealth 2019;7(7):e11926 [FREE Full text] [doi: 10.2196/11926] [Medline: 31274112]
24. Pyun KR, Kwon K, Yoo MJ, Kim KK, Gong D, Yeo WH, et al. Machine-learned wearable sensors for real-time hand-motion recognition: Toward practical applications. Natl Sci Rev 2024;11(2):nwad298 [FREE Full text] [doi: 10.1093/nsr/nwad298] [Medline: 38213520]
25. Godbout A. & Boyd, J. E.. Rhythmic sonic feedback for speed skating by real-time movement synchronization. International Journal of Computer Science in Sport. 11 2012:37-51.
26. Schwerz de Lucena D, Rowe JB, Okita S, Chan V, Cramer SC, Reinkensmeyer DJ. Providing real-time wearable feedback to increase hand use after stroke: A randomized, controlled trial. Sensors (Basel) 2022;22(18):6938 [FREE Full text] [doi: 10.3390/s22186938] [Medline: 36146287]
27. Vincent VR, Eric H, Thomas F, Alexander M. Breath tools: Exploring the effects on adherence and user experience of 4 sounds assisting runners with coupling breath to steps. : Association for Computing Machinery; 2023 Presented at: In Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23); 2023 Apr 19; New York, NY, USA p. 1-7 URL: https://doi.org/10.1145/3544549.3585736 [doi: 10.1145/3544549.3585736]
28. Fink PW, Foo P, Jirsa VK, Kelso JA. Local and global stabilization of coordination by sensory information. Exp Brain Res 2000 Sep;134(1):9-20. [doi: 10.1007/s002210000439] [Medline: 11026721]
29. Harris JV, Sarah F, Ricardo GO. Sonic respiration: Controlling respiration rate through auditory biofeedback. 2014 Presented at: Conference on Human Factors in Computing Systems - Proceedings; 2024 Apr 26; Barcelona, Spain. [doi: 10.1145/2559206.2581233]
30. Constantini K, Stickford ASL, Bleich JL, Mannheimer PD, Levine BD, Chapman RF. Synchronizing gait with cardiac cycle phase alters heart rate response during running. Medicine and science in sports and exercise 2018;50(5):1046-1053 [FREE Full text] [doi: 10.1249/mss.0000000000001515]
31. Chair SY, Zou H, Cao X. A systematic review of effects of recorded music listening during exercise on physical activity adherence and health outcomes in patients with coronary heart disease. Ann Phys Rehabil Med 2021;64(2):101447 [FREE Full text] [doi: 10.1016/j.rehab.2020.09.011] [Medline: 33130038]
32. Shandhi MMH, Singh K, Janson N, Ashar P, Singh G, Lu B, et al. Assessment of ownership of smart devices and the acceptability of digital health data sharing. NPJ Digit Med 2024;7(1):44 [FREE Full text] [doi: 10.1038/s41746-024-01030-x] [Medline: 38388660]
33. Peng W, Li L, Kononova A, Cotten S, Kamp K, Bowen M. Habit formation in wearable activity tracker use among older adults: Qualitative study. JMIR Mhealth Uhealth 2021;9(1):e22488 [FREE Full text] [doi: 10.2196/22488] [Medline: 33464216]
34. Leese J, MacDonald G, Backman CL, Townsend A, Nimmon L, Li LC. Experiences of wearable technology by persons with knee osteoarthritis participating in a physical activity counseling intervention: Qualitative study using a relational ethics lens. JMIR Mhealth Uhealth 2021;9(11):e30332 [FREE Full text] [doi: 10.2196/30332] [Medline: 34766912]
35. Farivar S, Abouzahra M, Ghasemaghaei M. Wearable device adoption among older adults: A mixed-methods study. Int J Inf Manage 2020;55:102209 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102209] [Medline: 32834339]
36. Wang Z, Fang D, Liu X, Zhang L, Duan H, Wang C, et al. Consumer acceptance of sports wearables: The role of products attributes. Sage Open 2023;13(3):21582440231182653 [FREE Full text] [doi: 10.1177/21582440231182653]
37. Pang C, Collin Wang Z, McGrenere J, Leung R, Dai J, Moffatt K. Technology adoption and learning preferences for older adults: Evolving perceptions, ongoing challenges, and emerging design opportunities. 2021 Presented at: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems; 2021 May 07; New York N Y United States p. 1-13 URL: https://doi.org/10.1145/3411764.3445702 [doi: 10.1145/3411764.3445702]
38. Fayazi N, Frankel L. Creating emotional attachment with assistive wearables. In: Stephanidis C, Kurosu M, Kurosu H, Reinerman-Jones DL, editors. HCI International 2020-Late Breaking Papers: Multimodality and Intelligence. Cham Switzerland: Springer International Publishing; 2020:73-88.
39. Van DGMJ, Bol N, Weert JCMV. Age differences in preferences for emotionally-meaningful versus knowledge-related appeals. Communications 2021;46(2):205-228 [FREE Full text] [doi: 10.1515/commun-2019-0108]
40. Desmet PMA, Roeser S. Emotions in design for values. In: Van Den Hoven J, Vermaas PE, Van De Poel I, editors. Handbook of Ethics, Values, and Technological Design. Netherlands: Springer; 2015:203-219.
41. Desai S, McGrath C, McNeil H, Sveistrup H, McMurray J, Astell A. Experiential value of technologies: A qualitative study with older adults. IJERPH 2022;19(4):2235 [FREE Full text] [doi: 10.3390/ijerph19042235]
42. Trocchia PJ, Janda S. A phenomenological investigation of internet usage among older individuals. The Journal of Consumer Marketing 2000;17(7):605-616 [FREE Full text] [doi: 10.1108/07363760010357804]

XSL•FO
RenderX

43. Burns M, Bally J, Burles M, Holtslander L, Peacock S. Constructivist grounded theory or interpretive phenomenology? Methodological choices within specific study contexts. International Journal of Qualitative Methods 2022;21:16094069221077758 [FREE Full text] [doi: 10.1177/16094069221077758]
44. Davidsen AS. Phenomenological approaches in psychology and health sciences. Qual Res Psychol 2013;10(3):318-339 [FREE Full text] [doi: 10.1080/14780887.2011.608466] [Medline: 23606810]
45. Hill R, Betts LR, Gardner SE. Older adults' experiences and perceptions of digital technology: (Dis)empowerment, wellbeing, and inclusion. Computers in Human Behavior 2015;48:415-423 [FREE Full text] [doi: 10.1016/j.chb.2015.01.062]
46. Hewitt S, Sephton R, Yeowell G. The effectiveness of digital health interventions in the management of musculoskeletal conditions: Systematic literature review. J Med Internet Res 2020;22(6):e15617 [FREE Full text] [doi: 10.2196/15617] [Medline: 32501277]
47. Smith JA, Osborn M. Interpretative phenomenological analysis. In: Smith JA, editor. Qualitative Psychology: A Practical Guide to Research Methods. USA: Sage Publications; 2003:51-80.
48. Dawe J, Cavicchiolo E, Palombi T, Baiocco R, Antoniucci C, Pistella J, et al. Measuring self-efficacy for exercise among older adults: Psychometric properties and measurement invariance of a brief version of the self-efficacy for exercise (SEE) scale. Healthcare (Basel) 2024;12(16):1642 [FREE Full text] [doi: 10.3390/healthcare12161642] [Medline: 39201200]
49. Green J, Thorogood N. Qualitative Methods for Health Research. USA: SAGE Publications; 2018.
50. Olsson SJS, Ekblom Ö, Andersson E, Börjesson M, Kallings LV. Categorical answer modes provide superior validity to open answers when asking for level of physical activity: A cross-sectional study. Scand J Public Health 2016;44(1):70-76 [FREE Full text] [doi: 10.1177/1403494815602830] [Medline: 26392418]
51. König HH, Bernert S, Angermeyer MC, Matschinger H, Martinez M, Vilagut G, The ESEMeD/MHEDEA 2000 Investigators. Comparison of population health status in six european countries: Results of a representative survey using the EQ-5D questionnaire. Medical Care 2009;47(2):255-261 [FREE Full text] [doi: 10.1097/mlr.0b013e318184759e]
52. Ross NJ, Renold E, Holland S, Hillman A. Moving stories: Using mobile methods to explore the everyday lives of young people in public care. Qualitative Research 2009;9(5):605-623 [FREE Full text] [doi: 10.1177/1468794109343629]
53. Evans J, Jones P. The walking interview: Methodology, mobility and place. Applied Geography 2011;31(2):849-858 [FREE Full text] [doi: 10.1016/j.apgeog.2010.09.005]
54. Low KEY. The sensuous city: Sensory methodologies in urban ethnographic research. Ethnography 2015;16(3):295-312 [FREE Full text] [doi: 10.1177/1466138114552938]
55. Carpiano RM. Come take a walk with me: The "go-along" interview as a novel method for studying the implications of place for health and well-being. Health Place 2009;15(1):263-272 [FREE Full text] [doi: 10.1016/j.healthplace.2008.05.003] [Medline: 18606557]
56. Glenn NM, Lapalme J, McCready G, Frohlich KL. Young adults' experiences of neighbourhood smoking-related norms and practices: A qualitative study exploring place-based social inequalities in smoking. Soc Sci Med 2017;189:17-24 [FREE Full text] [doi: 10.1016/j.socscimed.2017.07.021] [Medline: 28778059]
57. Eisenberg ME, Garcia CM, Frerich EA, Lechner KE, Lust KA. Through the eyes of the student: What college students look for, find, and think about sexual health resources on campus. Sex Res Soc Policy 2012;9(4):306-316 [FREE Full text] [doi: 10.1007/s13178-012-0087-0]
58. Novak V, Hu K, Vyas M, Lipsitz LA. Cardiolocomotor coupling in young and elderly people. J Gerontol A Biol Sci Med Sci 2007;62(1):86-92 [FREE Full text] [doi: 10.1093/gerona/62.1.86] [Medline: 17301044]
59. Eatough V, Smith JA, Willig C, Stainton-Rogers W. Interpretative phenomenological analysis. In: In The SAGE Handbook of Qualitative Research in Psychology. USA: SAGE Publications; 2012:179-194.
60. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005;15(9):1277-1288 [FREE Full text] [doi: 10.1177/1049732305276687] [Medline: 16204405]
61. Bengtsson M. How to plan and perform a qualitative study using content analysis. NursingPlus Open 2016;2:8-14 [FREE Full text] [doi: 10.1016/j.npls.2016.01.001]
62. Mercer K, Li M, Giangregorio L, Burns C, Grindrod K. Behavior change techniques present in wearable activity trackers: A critical analysis. JMIR Mhealth Uhealth 2016;4(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4461] [Medline: 27122452]
63. Li J, Ma Q, Chan AH, Man SS. Health monitoring through wearable technologies for older adults: Smart wearables acceptance model. Appl Ergon 2019;75:162-169 [FREE Full text] [doi: 10.1016/j.apergo.2018.10.006] [Medline: 30509522]
64. Kuoppamäki S, Hänninen R, Taipale S. Enhancing older adults' digital inclusion through social support: A qualitative interview study. In: Tsatsou P, editor. Vulnerable People and Digital Inclusion. Cham Switzerland: Springer International Publishing; 2022:211-230.
65. Parviainen J, Aromaa J. Bodily knowledge beyond motor skills and physical fitness: A phenomenological description of knowledge formation in physical training. Sport, Education and Society 2017;22(4):477-492 [FREE Full text] [doi: 10.1080/13573322.2015.1054273]
66. Phoenix C, Bell SL. Beyond "Move More": Feeling the rhythms of physical activity in mid and later-life. Soc Sci Med 2019;231:47-54 [FREE Full text] [doi: 10.1016/j.socscimed.2018.05.006] [Medline: 29759825]

XSL•FO
RenderX

67. Mate KK, Abou-Sharkh A, Morais JA, Mayo NE. Real-Time auditory feedback-induced adaptation to walking among seniors using the heel2toe sensor: Proof-of-concept study. JMIR Rehabil Assist Technol 2019;6(2):e13889 [FREE Full text] [doi: 10.2196/13889] [Medline: 31825320]

Abbreviations

IPA: interpretative phenomenological analysis
PA: physical activity
RQ: research question
RTF: real-time feedback
WRF: wearable device with real-time feedback

Edited by L Buis; submitted 20.Jan.2025; peer-reviewed by K Burford, S Lin, M Fraser; comments to author 28.May.2025; revised version received 20.Feb.2026; accepted 20.Feb.2026; published 29.Apr.2026.

Please cite as:
Lu TY, Rosato A, Dual SA, Kuoppamäki S
Lived Experiences of Older Adults Using Wearables With Real-Time Feedback: Phenomenological Study
JMIR Mhealth Uhealth 2026;14:e71509
URL: https://mhealth.jmir.org/2026/1/e71509
doi:10.2196/71509
PMID:

©Tien-Ying Lu, Aurora Rosato, Seraina Anne Dual, Sanna Kuoppamäki. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 29.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Smartphone Apps for Cannabis Cessation: Quality Assessment and Content Analysis

Siddharth Seth[1,2], BHSc; Sumedha Kushwaha[2,3], MDS; Reshma Prashad[4], PhD; Michael Chaiton[2,3,5], PhD

[1]Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[2]Centre for Addiction and Mental Health, Toronto, ON, Canada

[3]Institute of Medical Science, University of Toronto, Toronto, ON, Canada

[4]Faculty of Health Sciences, Ontario Tech University, Oshawa, ON, Canada

[5]Dalla Lana School of Public health, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Siddharth Seth, BHSc
Temerty Faculty of Medicine
University of Toronto
1 King's College Cir.
Toronto, ON, M5S 3K3
Canada
Phone: 1 437 980 7379
Email: sethsiddharth786@gmail.com

Abstract

Background: Over the past 2 decades, global rates of cannabis use have risen significantly, especially among young adults. This has corresponded to an increase in cannabis-related problems and hospitalizations. Thus, there has been significant interest in developing new interventions that can help facilitate cannabis cessation and reduce hospitalization rates. Specifically, mobile apps have emerged as scalable and accessible stand-alone or adjunct interventions that can help individuals with cannabis use disorders.

Objective: This study aimed to evaluate the quality of free cannabis cessation apps available on both the Apple App Store and Google Play Store, focusing on the analysis of their features, content, and adherence to evidence-based practices.

Methods: A systematic search was conducted in April 2023 using a variety of keywords. The apps were deemed eligible if they were free, available in English, accessible on both the Apple App Store and the Google Play Store, and related to cannabis cessation. Eligible apps were used for at least 1 month and were rated on the Mobile App Rating Scale by 2 reviewers. Interrater reliability was excellent, with a weighted Cohen κ of 0.893 (95% CI 0.835-0.943).

Results: Four apps were included in the analysis, namely, “Grounded–Quit Weed,” “Quit Weed,” “Marijuana Addiction Calendar,” and “Marijuana Anonymous.” The mean overall quality score of the apps was 3.4 out of 5, indicating poor to acceptable quality. The apps scored the highest on the “functionality” section and the lowest on the “information” section. Of the 4 apps, 3 focused on tracking cannabis use and duration of abstinence, whereas 1 focused on peer support. A limited number of cannabis cessation apps were identified, and those that were available were of low quality due to a lack of evidence-based information.

Conclusions: This study is the first to evaluate the current availability and quality of mobile apps designed for cannabis cessation. Unlike previous research that broadly assessed cannabis-related mobile apps, this study focuses on the limited number of free cannabis cessation tools, reflecting what is most available to the general population. The findings highlight a significant gap between the growing demand for virtual cessation tools and the quality of existing options. With the rising global prevalence of cannabis use disorders, there is an increasing need for robust, accessible, and evidence-based therapeutic options. While mobile health apps may be a viable option to support cannabis cessation, the current landscape is limited by poor quality apps and a lack of evidence-based information. From a real-world perspective, this study highlights the need for users to exercise caution when relying on current cannabis cessation apps and underscores the urgent need for the development and evaluation of new evidence-based digital interventions.

(JMIR Mhealth Uhealth 2026;14:e58908) doi:10.2196/58908

XSL•FO
RenderX

KEYWORDS

cannabis; cessation; mobile health interventions; mHealth interventions; mobile apps

Introduction

On October 17, 2018, Canada legalized recreational cannabis through the Cannabis Act, making it the second country in the world, after Uruguay, to do so at the national level [1]. The act was initially passed to protect public health and safety by restricting youth access to cannabis, preventing inducement to use, reducing illicit production, providing access to a quality-controlled supply of cannabis, and enhancing public awareness of the health risks of cannabis use [2]. However, the legislation appears to have led to a marked increase in cannabis use across Canada [3,4]. In fact, in 2017, the Canadian Tobacco, Alcohol and Drugs Survey reported that the prevalence of past-year cannabis use was 15% [5]. This contrasts sharply with findings from the updated Canadian Substance Use Survey, which reported that 32.4% of the population used cannabis in the past 12 months in 2023 [6]. Furthermore, the 2021 Canadian Cannabis Survey indicated that 17% of Canadians aged 16 years and older reported using cannabis in the past 30 days, with an average consumption of 14.3 days [4]. Another study on cannabis legalization from 2001 to 2019 revealed significant increases in daily cannabis use, past 12-month use, and cannabis-related problems, with the prevalence of cannabis-related problems in Canada rising from 6% to 14% between 2004 and 2019 [7]. In 2019, there were 23.80 million cases of cannabis use disorder globally, with the age-adjusted highest number of incidences observed in high-income North America (121/100,000) [8].

A study combining 19 iterations of the Centre for Addiction and Mental Health Monitor Surveys using a pre-post design found that young adults (aged 18-34 years) had higher rates of cannabis-related problems, such as dependence, anxiety, depressive disorders, and physical health effects, than their older counterparts [7,9]. Although younger individuals continue to exhibit higher levels of cannabis use, the largest rise in consumption since legalization was observed among middle-aged and older adults (those aged 45-64 years and ≥65 years, respectively) [7]. It was concluded that the Cannabis Act was strongly associated with an increased likelihood of cannabis use, daily cannabis use, and cannabis-related problems, including dependence [7]. This finding is supported by Ontario Cannabis Store reports stating that total cannabis sales in Ontario increased by 182% (from 35 million grams to 99 million grams) between the first and second fiscal years [10,11].

According to the Canadian Institute for Health Information, cannabis was the leading contributor to hospitalization rates for substance use among youth aged 10 to 24 years (40%) between 2017 and 2018 [12]. For those aged 25 and older, cannabis was associated with 11% of hospitalizations [12]. Given the age-standardized rate of 104 per 100,000 for youth hospital stays caused by cannabis harm, it is clear that significant resources are required to address youth-specific substance use needs, including timely access to prevention, intervention, and treatment services [12]. Nonetheless, addiction treatment wait times have increased over the last decade, with the average wait time for residential programs reaching 50 days by 2019. Many clients were reported to have been hospitalized, incarcerated, to have attempted suicide, or to have died while awaiting treatment [13].

Various methods can be used to facilitate cannabis addiction treatment, such as cognitive behavioral therapy, which aids individuals in identifying and correcting problematic behaviors, including addiction [14]. Contingency management is another approach that monitors target behaviors and provides or removes tangible rewards to reinforce positive behaviors [14]. Although no medications are currently approved for cannabis use disorder therapy, some are being trialed to help control withdrawal symptoms, and newer clinical trials are focusing on allosteric modulators to inhibit cannabis's rewarding effect [15].

Mobile phones have revolutionized communication technology, enabling users to access various apps for different purposes. As of 2021, this technology reached a global audience of 6.4 billion individuals and continues to grow, making it pervasive and widely used [16]. Current literature has documented the use of mobile apps for alcohol and tobacco dependence, and research suggests that these apps may promote successful quit attempts among e-cigarette users or vapers [17-19]. The available apps focus on habit tracking, therapy, or education as their primary mode of effectiveness. However, there has been limited research into mobile apps that can be used to facilitate cannabis cessation. This is a pressing issue, as cannabis use is rising globally [20]. In 2014, a content analysis of apps about cannabis on the Apple App Store and the Google Play Store was conducted; however, despite their extensive searching, they only found 1 app that was aimed at promoting cannabis cessation [21]. Since then, the mobile health (mHealth) field has grown significantly, and thousands of new apps have been added to the Apple App Store and the Google Play Store. With the growing field of mHealth, the increasing prevalence of mobile phone use, and the role of technology in disseminating cannabis-related news and information, this study aimed to examine the quality of cannabis cessation apps available in the marketplace, their content and features, popularity among users, and adherence to evidence-based practices. This study primarily focused on free English-language mobile apps available in both the Apple App Store and the Google Play Store.

Methods

Search Strategy

Using the search terms "cannabis cessation," "cannabis quit," "quit weed," "marijuana stop," "stop pot," "weed cessation," and "ganja," the primary and secondary authors searched the Canadian Apple App Store and Google Play Store in April 2023 for smartphone apps targeting cannabis cessation. Searches were conducted directly within the embedded mobile app store platforms rather than their corresponding websites to ensure that the results reflected how users typically discover and download apps.

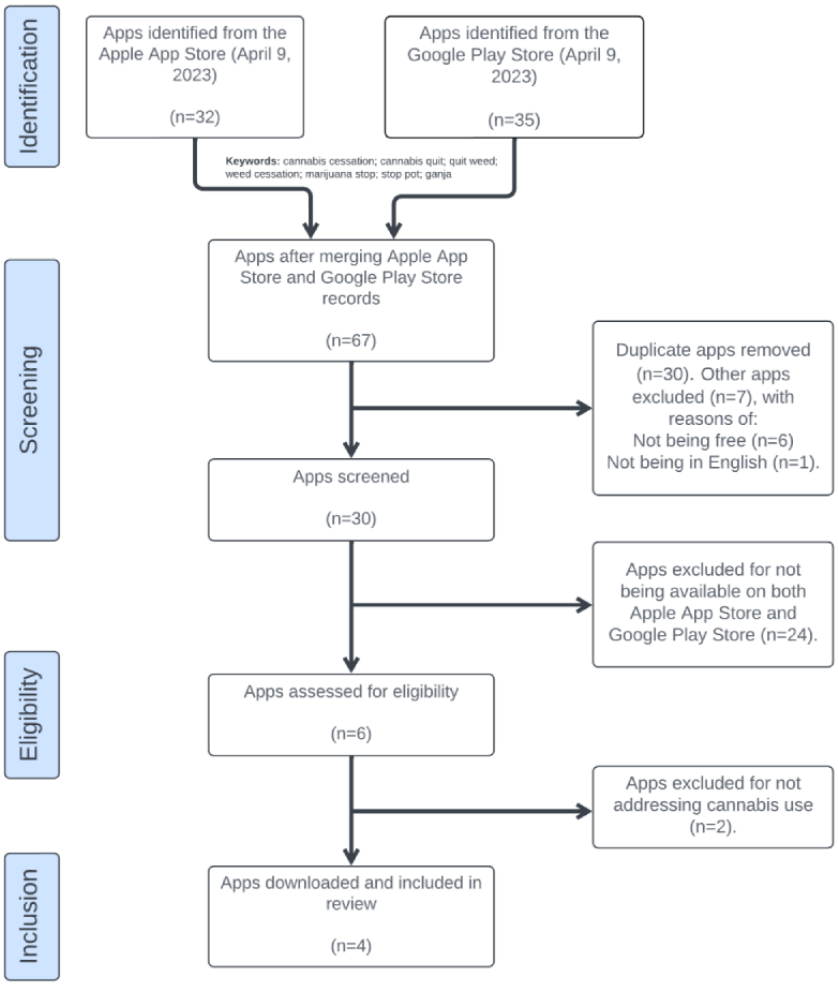
XSL•FO
RenderX

Eligibility Criteria

Mobile apps related to cannabis cessation or cannabis behavior modification were included. Apps that were not available for free download or were not in English were excluded. Only free apps were included as they tend to have higher use and accessibility for the general population than paid apps. Apps that were not available on both the Google Play Store and the Apple App Store were also excluded. The Google Play Store and the Apple App Store were selected as the 2 markets of interest, as these platforms are responsible for the largest market share of users [22].

Once ineligible apps were removed through a preliminary screening, the remaining apps were examined by checking their descriptions, profiles, relevance to cannabis cessation, and functionality. Those that did not meet the required criteria were excluded. The remaining eligible apps were then independently reviewed by the primary and secondary authors through a 2-pronged screening approach. First, the profiles of the apps were examined on the online app stores, and second, an ancestry search of developer websites, profiles, and overall online presence was conducted when information was available. Apps that were confirmed to have no association with cannabis cessation were excluded, resulting in a final sample of 4 apps. A detailed summary of the inclusion and exclusion process can be seen in Figure 1.

Figure 1. Flow diagram showing the identification, screening, eligibility assessment, and inclusion of cannabis cessation mobile apps for further Mobile App Rating Scale analysis.



Assessment of Quality, Contents and Features, and Popularity Among Users

To ensure the accuracy and reliability of our results, each app was rigorously tested over a short period (up to 1 month). This allowed us to classify the apps based on quality, features, and content and to perform the necessary evaluation of these tools. A variety of metrics, including the number of downloads, the number of reviews, and the overall user ratings, were also collected when available. We exclusively assessed the apps on Apple devices. This was because the Apple ecosystem provides a more consistent experience across devices, minimizing the variability associated with different hardware and software configurations found in Android devices.

The quality of each app was evaluated using the Mobile App Rating Scale (MARS), which is a multidimensional measure that is designed to rate the quality of mHealth apps [23]. The MARS has been specifically designed to evaluate mHealth apps and has been validated as a reliable measure of health app quality [23]. The scale comprises 5 categories—engagement,

XSL•FO
RenderX

functionality, esthetics, information quality, and subjective quality—with a total of 23 collective items rated using a 5-point scale. For items that could not be adequately assessed, an option of "not applicable" was available. The MARS also includes a subjective quality category; however, it was not factored into the final score.

To identify the strengths and weaknesses of each app, mean scores were calculated for each of the 4 objective categories, while the sum of all 4 category scores provided an overall quality score for each app. However, given the variability among health-related apps, the MARS does not provide a defined threshold for the type of scores a high-quality app should obtain on the scale. The use of the MARS and the overall methodology were based on a previous study that assessed the quality and content of mobile vaping cessation apps [19]. A MARS total mean score of 3.0 was established as the cutoff for apps with acceptable quality [19].

Data Extraction and Quality Assessment

Two raters tested the MARS with a randomly selected app. Results from the test app showed substantial agreement, with minor differences (≤2 points) that were discussed and resolved. The remaining apps were then independently evaluated and once again showed excellent agreement between the 2 reviewers (≤2 points). The weighted Cohen κ score was 0.893 (95% CI 0.835-0.943).

Ethical Considerations

This paper does not include any studies involving human participants or animals conducted by the authors; therefore, informed consent was not required. Institutional review board approval, privacy, or compensation statements do not apply to this study. All information is publicly available.

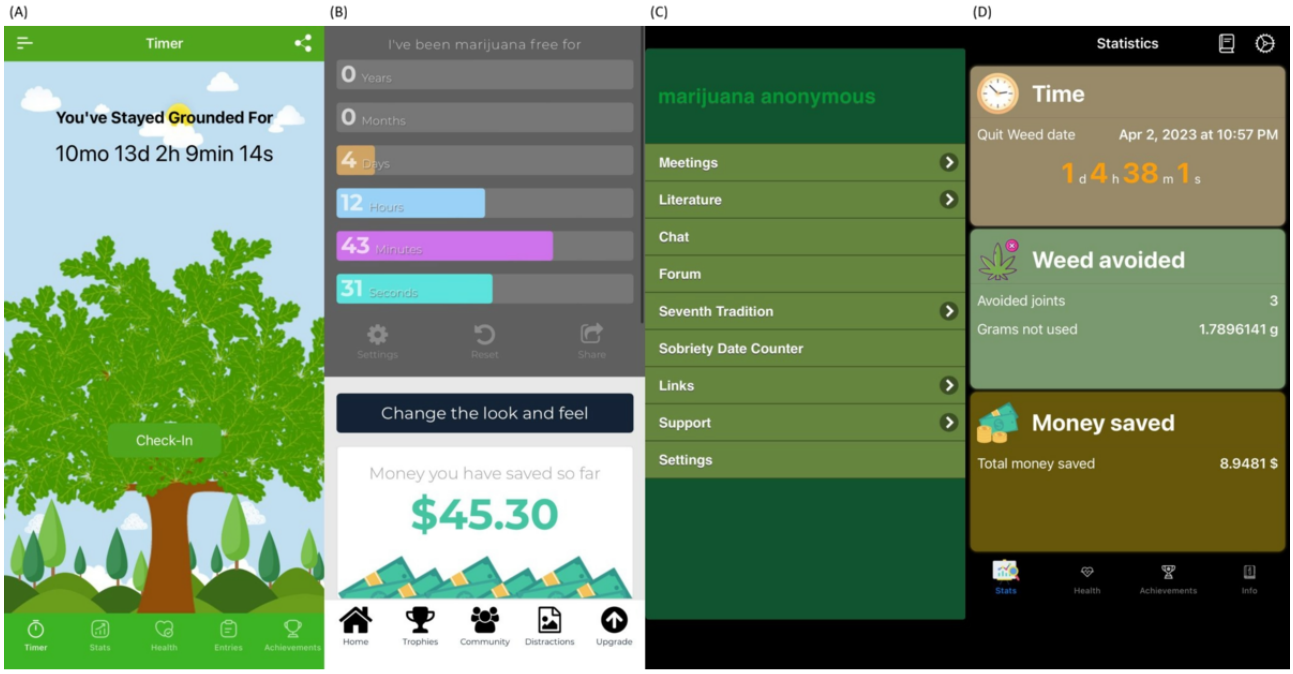
Results

Overview

Following a comprehensive screening process, 4 cannabis cessation mobile apps were identified. The full app selection and screening process can be seen in Figure 1.

On the basis of the inclusion criteria, we selected the following four apps: (1) "Grounded–Quit Weed," (2) "Quit Weed," (3) "Marijuana Addiction Calendar," and (4) "Marijuana Anonymous." Figure 2 provides a visual overview of the 4 apps included in the analysis.

Figure 2. Visual overview of the 4 mobile apps included in the analysis. (A) Grounded–Quit Weed, (B) Marijuana Addiction Calendar, (C) Marijuana Anonymous, and (D) Quit Weed.



General Assessment of Mobile Apps

The overall mean score for the MARS app quality was 3.4, indicating a poor to acceptable quality of the apps. Among the 4 apps, "Grounded–Quit Weed" and "Quit Weed" received the highest score (3.7 each), while "Marijuana Anonymous" received the lowest score (2.8).

On average, the functionality category received the highest mean score (4.3), followed by engagement (3.2), esthetics (3.2), and information (2.8). The full app score breakdown can be seen in Table 1.

XSL•FO
RenderX

Table 1. Characteristics of mobile apps based on publicly available information from the Apple App Store and Google Play Store, along with average Mobile App Rating Scale (MARS) scores. The maximum possible average score is 5.

Characteristic	Grounded–Quit Weed	Marijuana Addiction Calendar	Marijuana Anonymous	Quit Weed
App classification	Tracker	Tracker	Support meetings	Tracker
User rating on Apple App Store	4.7	4.4	2.8	4.9
User rating on Google Play Store	4.4	3.1	4.1	4.6
Number of downloads on Google Play Store	>50,000	>10,000	>10,000	>50,000
Engagement	3.6	4.0	1.7	3.4
Functionality	4.3	4.0	4.0	5.0
Esthetics	4.3	2.7	2.3	3.3
Information quality	2.8	2.3	3.0	3.0
MARS total mean score	3.7	3.3	2.8	3.7

Grounded–Quit Weed

The “Grounded–Quit Weed” was one of the highest scoring apps based on the MARS analysis. It was also the most downloaded app on the Google Play Store. The app began by asking the user a series of questions regarding their cannabis use, specifically about the number of grams used per day, the price per gram, and the quit date. The app also asked users to specify their goal, whether it was quitting cannabis completely, taking a tolerance break, or simply tracking use. The app then used a timer to track the amount of time an individual spent abstaining from cannabis use. Additionally, the app reported the approximate money saved and the amount of cannabis avoided.

The app also featured a journaling component, which allowed users to “check in” by recording whether they used cannabis, the quantity they used, the total amount of money spent, the number and intensity of urges, the type of equipment used, and the date and time of each entry. This enabled users to monitor and analyze the frequency and intensity of their urges to help identify possible triggers. The feature for monitoring withdrawal symptoms and recording detailed trigger information was only available in the paid version of the app and thus was not evaluated. The app included a “Motivation” subsection, which provided users with daily rotating quotes. Additionally, the app included an achievements tracker and a visual representation of user progress, depicted by a tree that grew larger with continued abstinence.

The “Grounded–Quit Weed” app included a “Quit Guide” section that offered limited educational information on cannabis cessation. This section divided the quitting process into 3 distinct phases, providing brief descriptions of each phase, including common side effects, associated thoughts, and guidance on identifying the user’s current stage. However, the app did not provide citations or references to support this phase-based framework. Additionally, the app featured a “Library” section linking to the Grounded420 website, which covered topics such as “What is CBD?” and “What is CHS?” Of these, only the “What is CHS?” paper included citations. Further exploration of the website revealed a range of blog posts on cannabis cessation; however, many of these papers lacked references, which brings into question the credibility and evidence base of the content provided.

The app advertised a community chat feature; however, this was available only in its premium paid version and thus was not evaluated.

Quit Weed

The “Quit Weed” app used a similar approach to the “Grounded–Quit Weed” app by tracking the time users spent abstaining from cannabis use. Users could enter their average weed use, cost, and quit date in the settings. This app then tracked the approximate amount of cannabis avoided and the approximate money saved. This app also tracked the side effects users may experience during the quitting process and the duration required for those side effects to resolve. While there were references and citations in this section, many were not from peer-reviewed journal papers.

Similar to “Grounded–Quit Weed,” this app had an “Achievements” section, which gave badges based on the number of joints avoided, grams of cannabis avoided, and money saved. The app also seemed to have a motivation section; however, this feature was locked behind a paywall and thus was not evaluated. The app also seemed to have information on recovery phases and associated side effects; however, this content was also blocked behind a paywall and thus was not evaluated.

Marijuana Addiction Calendar

The “Marijuana Addiction Calendar” also used a timer to track the amount of time users spent abstaining from cannabis use. Similar to “Grounded–Quit Weed”, the app asked for information about the amount of money spent on marijuana per day, the amount of time spent smoking marijuana, the user’s quit date, and the reason for quitting. The app also calculated the money and time saved by not smoking.

The app also had a “Distractions” section, which contained a limited selection of images, sounds, games, and workouts. However, there were only 2 categories of images and 1 category of sounds that were available in the free version. The rest of the images, sounds, and games were restricted to the paid version and thus were not evaluated. The workout section was actually

XSL•FO
RenderX

an advertisement for another mobile app and was also not evaluated. Additionally, this app had a "Trophies" section associated with duration of abstinence, ranging from 1 week up to 15 years.

The app also had a community chat feature, which allowed users to post in the forum after signing in with their Apple account. The forum was not very active, with only a few users using this feature weekly. Discussions and posts typically centered on individuals' initial attempts at quitting; documenting their setbacks; and seeking advice, strategies, and encouragement. Users are also able to like, reply to, or flag other user's comments as inappropriate. However, users with free memberships were limited in how much they could post. The app also had a secondary feature that allowed users to submit testimonials about the effectiveness of this app on their quitting journey. These testimonials were then displayed to all users. The app also showed frequent full-page advertisements when switching between pages or tabs within the app.

Marijuana Anonymous

The "Marijuana Anonymous" app functioned primarily as a networking tool to locate licensed Marijuana Anonymous meetings nearby. Although it included a feature to track abstinence duration, this feature was minimal, allowing users to enter only a start date. The app did not provide any educational content on cannabis cessation but instead had a range of documents that described the purpose of "Marijuana Anonymous," the tenets of the organization, stories from individuals in the organization, and information about their 12-step abstinence process. It was noted that these documents and texts had heavy religious and biblical undertones. Although the primary intended function of the app was to help users locate active meetings, the location-based search feature was nonfunctional at the time of evaluation, limiting its utility. Consequently, the app displayed all scheduled meetings without filtering by user location.

A summary of app features and premium membership pricing can be seen in Table 2.

Table 2. App features available in each analyzed mobile app.

Feature	Grounded–Quit Weed	Quit Weed	Marijuana Addiction Calendar	Marijuana Anonymous
Time since cessation	✓	✓	✓	✓
Money saved	✓	✓	✓	—[a]
Time saved by avoiding smoking	—	—	✓	—
Cannabis avoided	✓	✓	—	—
Goal setting	✓	—	✓	—
Withdrawal symptom tracking	✓[b]	✓	—	—
Journaling	✓	—	—	—
Motivation	✓	✓[b]	✓	—
Achievements or trophies	✓	✓	✓	—
Educational content	✓	—	—	✓[c]
Community content (chats or meetings)	✓[b]	✓[b]	✓	✓
Pricing for premium (CAD $1=US $0.75 as of July 2023)	CAD $7.49/month; CAD $35.99/year	CAD $13.99 (one-time payment)	CAD $6.49/month; CAD $52.99/year	—

[a]Not applicable.

[b]Premium content that has not been evaluated.

[c]Educational content is about the Marijuana Anonymous method of quitting rather than about cannabis cessation in general.

Discussion

Principal Findings

This is the first study that aimed to evaluate the effectiveness of mobile apps for cannabis cessation and assess the current mobile app market in this regard. The study identified 4 apps available on both the Apple App Store and the Google Play Store for cannabis cessation, namely, "Grounded–Quit Weed," "Quit Weed," "Marijuana Anonymous," and "Marijuana Addiction Calendar." The apps were evaluated using the MARS scale and received an average score of 3.4. The "Grounded–Quit Weed" and the "Marijuana Addiction Calendar" apps received the highest score, whereas "Marijuana Anonymous" received the lowest score. "Grounded–Quit Weed," "Quit Weed," and "Marijuana Addiction Calendar" were use tracking apps, whereas "Marijuana Anonymous" was primarily focused on connecting individuals with local Marijuana Anonymous meetings. The results of this study offer valuable insights into the state of the mobile app market for cannabis cessation. Specifically, these results showcase the relative undersaturation of the mobile app field for cannabis cessation. When compared to smoking or vaping cessation mobile apps, it was found that there were significantly fewer apps for cannabis cessation [18,19].

XSL•FO
RenderX

The apps evaluated in this study had significant variability in their comprehensiveness, accessibility, evidence base, and overall quality. While "Grounded–Quit Weed" offered the most extensive range of features, it, like the other apps, lacked appropriate citations and references to support its educational content, limiting the credibility of the information provided. This pattern was reflected in the MARS analysis, in which the apps scored the lowest in the "information" category, suggesting a lack of evidence-based content within the apps. This is likely because clinicians were not involved in the development of these apps, and instead, the apps were focused on creating an engaging and functional experience rather than one that was empirically grounded. This is a critical gap in the field, as recent research has shown that evidence-based mobile apps for behavioral modification, both as stand-alone tools and as part of an intervention, have advantages over standard health interventions [24]. Despite the overall positive correlation between mobile apps and behavioral modifications in the literature, the heterogeneity of results suggests that more research is needed to determine the specific types of mobile apps and features that may yield the most significant benefits for users seeking cannabis cessation support [24].

Use tracking was the most commonly used strategy among both the analyzed and excluded apps. Most apps primarily focused on tracking the duration of cannabis cessation and reporting auxiliary information such as money saved. Although some other features were identified, such as the in-person community meetings associated with the "Marijuana Anonymous" app, a greater variety of strategies and tools is needed within the existing set of apps. Recent research has demonstrated the high efficacy of personalized medical apps, such as artificial intelligence chatbots that adapt to users, in changing health behaviors [25]. The apps should be designed to adapt to users' moods, preferences, and patterns of use to provide personalized health care support and maximize effectiveness [25]. However, more extensive randomized controlled trials are needed to draw definitive conclusions [25].

Strengths

This study has several strengths, particularly as the first systematic analysis of cannabis cessation apps. It adds to the growing body of research on alcohol, tobacco, and vaping cessation mobile apps and informs future research and practice in developing and analyzing cessation-focused apps. On the basis of our findings, we propose several recommendations for improving existing apps and developing future apps to better support the cessation process. Specifically, we suggest that researchers focus on providing adequate, evidence-based information for users and offering personalized experiences. Additionally, we recommend the use of multiple tools and methods to help facilitate cannabis cessation, including the use of artificial intelligence and apps that adapt to the users' moods and preferences [25]. This study lays the groundwork for further research on technology-based interventions, particularly mobile apps and chatbots, to help users quit cannabis.

Limitations

It is essential to acknowledge the limitations of this study. First, the small sample size limits the generalizability of the findings. However, this also demonstrates the need for further development of mHealth tools for cannabis use disorders. Furthermore, the evaluation of health care mobile apps is a recent endeavor, and thus, additional research is required to develop tools for accurate quality assessments. Additionally, given the rapid pace of mobile app development, it is possible that new apps released after our search were not captured in this review. Only free features were evaluated in this paper, as paid features were beyond the scope of this analysis. However, this approach was justified as free features and apps generally have larger audiences, higher use, and broader accessibility. The search was limited to the Canadian app stores, which could have possibly excluded apps available exclusively in other countries. Finally, the apps were analyzed only on Apple devices and not on Android devices. This meant some apps may not score comparatively on platforms running alternative operating systems. This was especially concerning as the apps were scored lower on the Google Play Store than on the Apple App Store.

Future Direction

Future research should be done to determine the clinical impact of these apps in regard to different users, ethnicities, and socioeconomic backgrounds. While usability and interface design were evaluated through the "Esthetics" and "Functionality" domains of the MARS analysis, the actual impact of these apps on the end users, particularly in promoting behavior change or supporting cannabis cessation, remains undetermined.

This study is the first to systematically evaluate freely available cannabis cessation mobile apps across major app platforms. Overall, this study found that the effectiveness and quality of currently available mobile apps for cannabis cessation are low, suggesting that more research is needed to develop evidence-based mobile apps for cannabis cessation. Unlike previous studies that focused on any apps associated with cannabis, this study particularly helps identify critical gaps in evidence-based content for mHealth interventions for cannabis cessation [21]. This study directly informs the need for the development of high-quality, evidence-based digital interventions and highlights the need for caution when patients use or clinicians recommend existing cannabis cessation apps. As cannabis use continues to rise globally, the creation and rigorous evaluation of mHealth apps may play a pivotal role in expanding access to effective cessation support and reducing cannabis-related harm.

Conclusions

This study identified and evaluated 4 mobile apps related to cannabis cessation. On the basis of the MARS analyses conducted, there is a critical gap in the current landscape of cannabis cessation mobile apps. The limited number of available apps, coupled with their low scores in the "information" category, underscores the need for more robust, scientifically tested digital interventions. As cannabis use and cannabis use disorders continue to rise globally, the development of high-quality mHealth tools may prove to be valuable in supporting the cessation process. Therefore, we recommend the development of more robust mobile apps that incorporate strong,

XSL•FO
RenderX

evidence-based approaches to cannabis cessation, to better support individuals seeking to reduce or quit cannabis use.

Acknowledgments

The authors attest that no generative artificial intelligence technology was used in the generation of text, figures, or other content in this manuscript.

Funding

The authors declare that no financial support was received for this study.

Data Availability

Data sharing is not applicable to this study as no datasets were generated or analyzed during this study.

Authors' Contributions

Conceptualization: SK, SS
Methodology: SK, SS
Investigation: SK, SS
Formal Analysis: SK, SS
Supervision: MC
Writing—original draft: SS
Writing—review and editing: MC, RP, SK

Conflicts of Interest

None declared.

References

1. Hawley P, Gobbo M, Afghari N. The impact of legalization of access to recreational cannabis on Canadian medical users with cancer. BMC Health Serv Res 2020 Oct 27;20(1):977 [FREE Full text] [doi: 10.1186/s12913-020-05756-8] [Medline: 33109169]
2. Cannabis Act S.C. 2018, c. 16. Government of Canada. 2018. URL: https://laws-lois.justice.gc.ca/eng/acts/c-24.5/page-1.html [accessed 2025-07-10]
3. Taking stock of progress: cannabis legalization and regulation in Canada. Government of Canada. 2022. URL: https://www.canada.ca/en/health-canada/programs/engaging-cannabis-legalization-regulation-canada-taking-stock-progress/document.html [accessed 2023-04-30]
4. Canadian Cannabis Survey 2021: summary. Government of Canada. 2021. URL: https://www.canada.ca/en/health-canada/services/drugs-medication/cannabis/research-data/canadian-cannabis-survey-2021-summary.html [accessed 2023-04-30]
5. Canadian Tobacco, Alcohol and Drugs Survey (CTADS): summary of results for 2017. Government of Canada. 2017. URL: https://www.canada.ca/en/health-canada/services/canadian-alcohol-drugs-survey/2017-summary.html [accessed 2025-07-10]
6. Drug and alcohol use in Canada: general population (ages 15+). Government of Canada. 2025. URL: https://health-infobase.canada.ca/substance-use/csus/ [accessed 2025-07-10]
7. Imtiaz S, Nigatu YT, Ali F, Douglas L, Hamilton HA, Rehm J, et al. Cannabis legalization and cannabis use, daily cannabis use and cannabis-related problems among adults in Ontario, Canada (2001-2019). Drug Alcohol Depend 2023 Mar 01;244:109765 [FREE Full text] [doi: 10.1016/j.drugalcdep.2023.109765] [Medline: 36652851]
8. Shah K, Farwa UE, Vanaparti A, Patel S, Kanumuri M, Vashishth O, et al. Global epidemiology of cannabis use disorders and its trend from 1990 to 2019: benchmarking analysis of the global burden of disease study. J Family Med Prim Care 2024 Mar;13(3):881-889 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_824_23] [Medline: 38736814]
9. Nigatu YT, Hamilton HA. CAMH monitor eReport 2022: substance use, mental health and well-being among Ontario adults. Centre for Addiction and Mental Health. 2022. URL: https://www.camh.ca/-/media/files/pdfs---camh-monitor/camh-monitor-2022_ereport_dec-19_final-pdf.pdf [accessed 2025-07-10]
10. A year in review 2019-2020. Ontario Cannabis Store. URL: https://cdn.shopify.com/s/files/1/2636/1928/files/OCS-InsightsReport_2019-2020_8f607a7b-21b0-4718-adb3-7cd60a2f08ea.pdf [accessed 2026-03-24]
11. A year in review April 1, 2020 – March 31, 2021. Ontario Cannabis Store. URL: https://cdn.shopify.com/s/files/1/2636/1928/files/OCS-InsightsReport_____2020-21.pdf [accessed 2026-03-24]

XSL•FO
RenderX

12. Original quantitative research – setting the baseline: a description of cannabis poisonings at a Canadian pediatric hospital prior to the legalization of recreational cannabis. Canadian Institute for Health Information. 2019. URL: https://tinyurl.com/59x6wynk [accessed 2025-05-25]
13. Lysyk B. Addictions-related treatment wait times, emergency department visits and deaths rising despite increased spending: Auditor General. Office of the Auditor General of Ontario. 2019 Dec 4. URL: https://www.auditor.on.ca/en/content/news/19_newsreleases/2019news_v1_3.02.pdf [accessed 2026-03-24]
14. Sherman BJ, McRae-Clark AL. Treatment of cannabis use disorder: current science and future outlook. Pharmacotherapy 2016 May 24;36(5):511-535 [FREE Full text] [doi: 10.1002/phar.1747] [Medline: 27027272]
15. Marijuana. National Institute on Drug Abuse (NIDA). 2018. URL: https://nida.nih.gov/sites/default/files/1380-marijuana.pdf [accessed 2026-04-09]
16. Liu JC, Ellis DA. Editorial: eating in the age of smartphones: the good, the bad, and the neutral. Front Psychol 2021 Dec 6;12:796899 [FREE Full text] [doi: 10.3389/fpsyg.2021.796899] [Medline: 34938246]
17. Tofighi B, Chemi C, Ruiz-Valcarcel J, Hein P, Hu L. Smartphone apps targeting alcohol and illicit substance use: systematic search in in commercial app stores and critical content analysis. JMIR Mhealth Uhealth 2019 Apr 22;7(4):e11831 [FREE Full text] [doi: 10.2196/11831] [Medline: 31008713]
18. Thornton L, Quinn C, Birrell L, Guillaumier A, Shaw B, Forbes E, et al. Free smoking cessation mobile apps available in Australia: a quality review and content analysis. Aust N Z J Public Health 2017 Dec;41(6):625-630. [doi: 10.1111/1753-6405.12688] [Medline: 28749591]
19. Sanchez S, Kundu A, Limanto E, Selby P, Baskerville NB, Chaiton M. Smartphone apps for vaping cessation: quality assessment and content analysis. JMIR Mhealth Uhealth 2022 Mar 28;10(3):e31309 [FREE Full text] [doi: 10.2196/31309] [Medline: 35343904]
20. Wang Q, Qin Z, Xing X, Zhu H, Jia Z. Prevalence of cannabis use around the world: a systematic review and meta-analysis, 2000-2024. China CDC Wkly 2024 Jun 21;6(25):597-604 [FREE Full text] [doi: 10.46234/ccdcw2024.116] [Medline: 38933041]
21. Ramo DE, Popova L, Grana R, Zhao S, Chavez K. Cannabis mobile apps: a content analysis. JMIR Mhealth Uhealth 2015 Aug 12;3(3):e81 [FREE Full text] [doi: 10.2196/mhealth.4405] [Medline: 26268634]
22. Pires IM, Marques G, Garcia NM, Flórez-Revuelta F, Ponciano V, Oniani S. A research on the classification and applicability of the mobile health applications. J Pers Med 2020 Feb 27;10(1):11 [FREE Full text] [doi: 10.3390/jpm10010011] [Medline: 32120849]
23. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
24. Iribarren SJ, Akande TO, Kamp KJ, Barry D, Kader YG, Suelzer E. Effectiveness of mobile apps to promote health and manage disease: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2021 Jan 11;9(1):e21563 [FREE Full text] [doi: 10.2196/21563] [Medline: 33427672]
25. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence-based chatbots for promoting health behavioral changes: systematic review. J Med Internet Res 2023 Feb 24;25:e40789 [FREE Full text] [doi: 10.2196/40789] [Medline: 36826990]

Abbreviations

MARS: Mobile App Rating Scale
mHealth: mobile health

Edited by S Brini; submitted 21.Apr.2024; peer-reviewed by J Westmaas, MDM Fernández Álvarez; comments to author 23.Feb.2025; accepted 12.Feb.2026; published 14.May.2026.

Please cite as:
Seth S, Kushwaha S, Prashad R, Chaiton M
Smartphone Apps for Cannabis Cessation: Quality Assessment and Content Analysis
JMIR Mhealth Uhealth 2026;14:e58908
URL: https://mhealth.jmir.org/2026/1/e58908
doi:10.2196/58908
PMID:

©Siddharth Seth, Sumedha Kushwaha, Reshma Prashad, Michael Chaiton. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction

XSL•FO
RenderX

in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Mobile Health App Acceptance in Japan's Aging Society: Multigroup Structural Equation Modeling Based on the Extended Unified Theory of Acceptance and Use of Technology and eHealth Literacy Frameworks

Yasue Fukuda[1], PhD; Koji Fukuda[2], PhD

[1]Faculty of Pharmaceutical Sciences, Suzuka University of Medical Science, Suzuka, Mie, Japan

[2]Faculty of Political Science and Economics, Waseda University, Tokyo, Shinjuku-ku, Japan

Corresponding Author:
Yasue Fukuda, PhD
Faculty of Pharmaceutical Sciences
Suzuka University of Medical Science
3500-3 Minamitamagaki-cho
Shiroko Campus
Suzuka, Mie, 513-8670
Japan
Phone: 81 593737030
Fax: 81 593681271
Email: yfukuda@suzuka-u.ac.jp

Abstract

Background: Mobile health (mHealth) technologies are increasingly promoted as tools for chronic disease management and healthy aging, yet adoption remains persistently uneven across demographic groups. Japan, where 29.1% of the population is 65 years or older—the highest proportion globally—exemplifies the challenges of mHealth promotion in super-aging societies. Despite high smartphone penetration (90.1%) and active national digital transformation initiatives, only 21.6% of Japanese adults report regular mHealth app use, with marked disparities by age and sex.

Objective: This study examined determinants of mHealth acceptance by extending the unified theory of acceptance and use of technology to incorporate eHealth literacy, self-efficacy, perceived risk, distrust, and health-related factors (health status and health interest). Age- and sex-specific differences in acceptance mechanisms were also investigated using multigroup structural equation modeling (SEM).

Methods: We conducted a cross-sectional online survey in November 2023 with 960 Japanese adults sampled across 7 age strata (aged 18-27 years to aged ≥78 years). SEM tested hypothesized relationships among 9 constructs. Health status and health interest were included as observed covariates. Multigroup SEM with configural, metric, and structural invariance testing examined age- and sex-specific differences, and binary logistic regression identified predictors of current mHealth app use.

Results: The structural model demonstrated good fit (χ^2/df=2.06; comparative fit index 0.953; Tucker-Lewis index 0.945; root mean square error of approximation 0.047) and explained 71.6% of the variance in behavioral intention. Effort expectancy (β=0.404), facilitating conditions (β=0.349), and performance expectancy (β=0.188) were the primary proximal predictors of behavioral intention. Social influence exerted strong upstream effects on effort expectancy (β=0.811), eHealth literacy (β=0.507), and self-efficacy (β=0.422). Health interest positively influenced performance expectancy (β=0.133), whereas neither health interest nor health status showed a significant direct effect on distrust. Distrust did not directly predict behavioral intention in the overall sample. Multigroup analyses identified 5 significant age differences and 5 sex differences. eHealth literacy increased distrust among young adults but reduced perceived risk among middle-aged and older adults. Self-efficacy negatively predicted performance expectancy among young adults yet positively predicted it among middle-aged and older adults. Distrust significantly reduced behavioral intention only among middle-aged adults.

Conclusions: mHealth acceptance in Japan's aging society is characterized by stable proximal determinants of behavioral intention alongside heterogeneous upstream belief formation processes that vary systematically by age and sex. Health interest, rather than health status, emerged as the key contextual driver of perceived usefulness. At the theoretical level, this study clarifies

XSL•FO
RenderX

how eHealth literacy, self-efficacy, and distrust function as age- and sex-contingent antecedents within an extended unified theory of acceptance and use of technology framework. At the practical level, these findings highlight the need for trust-centered, demographically tailored, and literacy-sensitive strategies to promote equitable mHealth adoption in rapidly aging societies.

(JMIR Mhealth Uhealth 2026;14:e87832) doi:10.2196/87832

KEYWORDS

aging society; distrust; eHealth literacy; health interest; Japan; mHealth; mobile health; self-efficacy; structural equation modeling; Unified Theory of Acceptance and Use of Technology; UTAUT

Introduction

Background

Mobile health (mHealth) technologies have emerged as critical tools for health care delivery, particularly in aging societies facing resource constraints and increasing chronic disease burdens [1,2]. Japan, with 29.1% of its population aged ≥65 years and projected to reach 38.4% by 2065 [3], exemplifies these demographic challenges [4]. Despite widespread smartphone penetration of 90.1% in 2022 and advanced digital infrastructure [5], mHealth adoption remains paradoxically low, with only 21.6% of Japanese adults reporting regular use of health-related apps [6,7]. This paradox reveals fundamental gaps in understanding technology acceptance mechanisms, particularly in aging populations where mHealth could deliver the greatest health benefits [8,9].

Existing technology acceptance research has predominantly examined younger, technologically proficient users in Western contexts [10,11], creating a critical knowledge gap regarding older adults' adoption mechanisms in rapidly aging Asian societies. Moreover, traditional technology acceptance models often overlook psychosocial factors specific to health technologies, privacy concerns [12,13], system distrust [14,15], and eHealth literacy [16,17], which may operate differently across age cohorts and cultural contexts. Understanding these age-differentiated, context-specific acceptance mechanisms is essential for designing evidence-based interventions to promote mHealth adoption among the world's fastest-aging populations [18,19].

Research Gaps and Theoretical Contributions

Despite extensive unified theory of acceptance and use of technology (UTAUT) applications to diverse technologies [20-22], several critical limitations persist in mHealth adoption research. Most existing studies have examined general populations or younger cohorts [23,24], with insufficient attention to older adults' distinct needs, capabilities, and concerns in aging societies where they constitute the primary target population for disease management apps yet demonstrate the lowest adoption rates [25,26]. In Japan's context, prior mHealth acceptance research has primarily focused on specific demographic segments to enable in-depth investigation. For example, Cao et al [27] examined mHealth acceptance among Japanese young adults using an extended UTAUT model, finding that trust, performance expectancy, and effort expectancy directly influenced behavioral intention, while health consciousness and social influence operated indirectly. Building upon these valuable insights, a comprehensive examination across the full adult lifespan within a single study remains needed to identify life stage–specific intervention strategies and enable population-level policy formulation [28,29].

Furthermore, research rarely examines how acceptance mechanisms vary across demographic subgroups within the same population [30,31]. While some studies compare age or sex groups separately, few use rigorous multigroup structural equation modeling (SEM) to test whether theorized relationships operate identically across subgroups [32,33], preventing identification of demographic-specific barriers and facilitators [34,35]. Additionally, few studies distinguish between disease-management apps and wellness-oriented apps, despite potentially distinct adoption drivers [36,37].

This study complements existing research through 3 theoretical contributions. First, we integrate privacy-trust dimensions by incorporating perceived risk [25,38] and distrust [30,39] as distinct constructs—dimensions inadequately addressed in core UTAUT yet critical in health contexts involving sensitive personal data [40,41]. Importantly, we theorize distrust not merely as a direct barrier to behavioral intention but as a mediating mechanism undermining other acceptance perceptions, proposing that skepticism operates through both direct and indirect pathways [42,43]. This dual-pathway conceptualization advances prior research treating distrust solely as a direct predictor [14,30].

Second, we theorize eHealth literacy [16,17] as a multifaceted construct that simultaneously enhances acceptance perceptions while potentially increasing critical risk evaluation. This dual-function conceptualization challenges prevailing assumptions that digital literacy uniformly reduces barriers [44,45]. We hypothesize that eHealth literacy may paradoxically increase distrust among digitally sophisticated users capable of critical evaluation—an alternative mechanism unexplored in prior research [46,47].

Third, we extend UTAUT by theorizing social influence as a comprehensive distal predictor affecting not only core UTAUT constructs but also self-efficacy, eHealth literacy, and distrust [20,48,49]. This comprehensive pathway specification tests whether social influence operates through multiple mechanisms simultaneously—a theoretical proposition inadequately examined in prior research [20,48].

Theoretical Model and Research Hypotheses

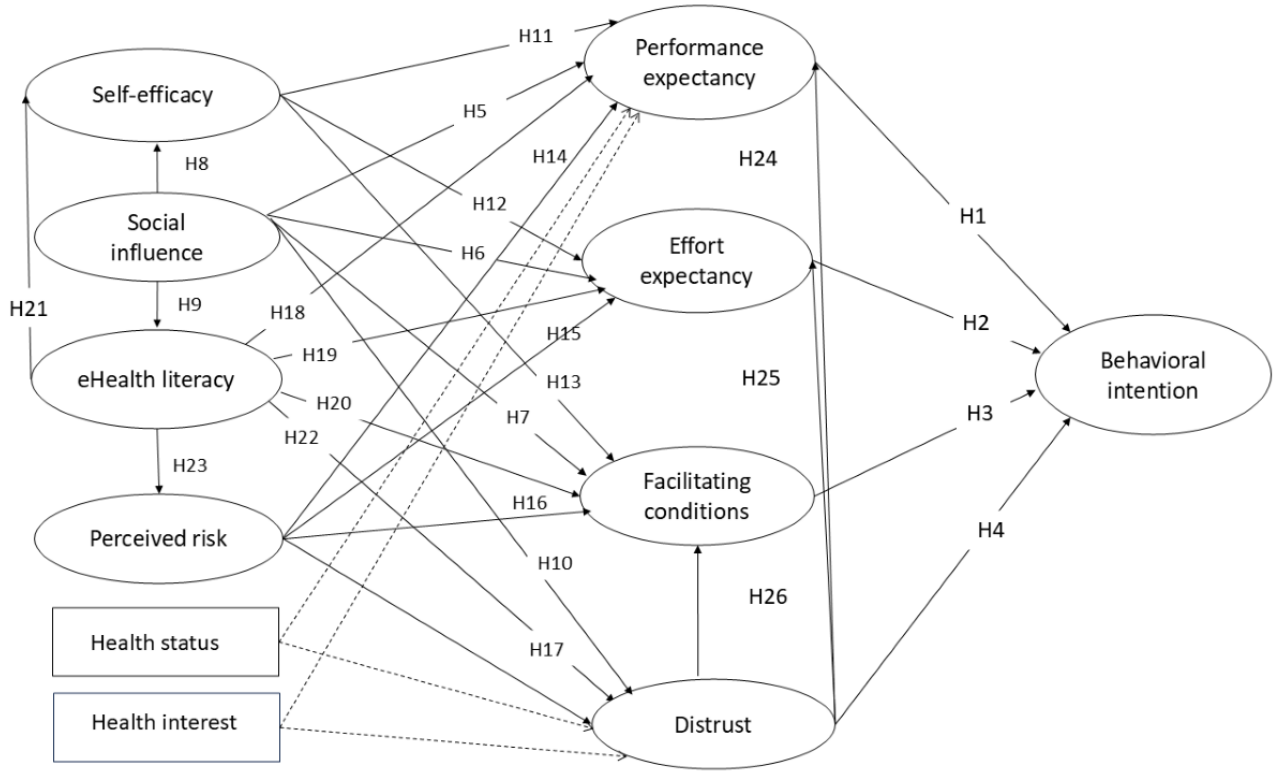
Conceptual Framework

The theoretical model integrates core UTAUT constructs with privacy-trust dimensions and digital literacy constructs to comprehensively explain mHealth acceptance mechanisms

XSL•FO
RenderX

[20,50]. The model comprises 8 constructs organized into 3 layers: social influence, self-efficacy, perceived risk, and eHealth literacy function as distal predictors; performance expectancy, effort expectancy, facilitating conditions, and distrust serve as proximal predictors; and behavioral intention represents the outcome construct [51,52]. This multilayered specification enables examination of direct effects, indirect effects through mediation pathways, and total effects across demographic subgroups and app types [53,54]. Figure 1 presents the theoretical model with all hypothesized relationships.

Figure 1. Extended unified theory of acceptance and use of technology research model incorporating eHealth literacy, self-efficacy, perceived risk, distrust, and health-related covariates (H1-H26). Dotted lines indicate covariate effects.



Research Hypotheses

Based on UTAUT [20,50] and theoretical extensions [17,25,30], we propose 26 hypotheses specifying relationships among constructs [55,56].

- H1: performance expectancy positively influences behavioral intention to use mHealth apps [20,57]. Users perceiving greater health benefits demonstrate higher adoption intentions.
- H2: effort expectancy positively influences behavioral intention to use mHealth apps [20,57]. Users perceiving ease of use demonstrate higher adoption intentions.
- H3: facilitating conditions positively influence behavioral intention to use mHealth apps [20,57]. Users perceiving adequate infrastructure support demonstrate higher adoption intentions.
- H4: distrust negatively influences behavioral intention to use mHealth apps [30,58,59]. System skepticism directly discourages adoption.
- H5: social influence positively influences performance expectancy [20,48]. Social endorsements enhance perceived benefits through observational learning.
- H6: social influence positively influences effort expectancy [20,48]. Peer success experiences enhance ease-of-use perceptions.
- H7: social influence positively influences facilitating conditions [20,48]. Social networks provide information about support resources.
- H8: social influence positively influences self-efficacy [48,60]. Observing others' success enhances confidence through vicarious experience.
- H9: social influence positively influences eHealth literacy [48,49]. Social environments facilitate learning through knowledge exchange.
- H10: social influence negatively influences distrust [49]. Trusted endorsements reduce system skepticism through trust transfer.
- H11: self-efficacy positively influences performance expectancy [60,61]. Confident users anticipate successfully extracting maximum app value.
- H12: self-efficacy positively influences effort expectancy [60,61]. Confident users perceive technology tasks as less effortful.
- H13: self-efficacy positively influences facilitating conditions [60]. Confident users perceive support infrastructure as more adequate.
- H14: perceived risk negatively influences performance expectancy [25,38]. Privacy concerns introduce costs that offset anticipated benefits.

XSL•FO
RenderX

- H15: perceived risk negatively influences effort expectancy [25,38]. Privacy concerns increase perceived complexity through risk management burden.
- H16: perceived risk negatively influences facilitating conditions [38,62]. Privacy concerns signal potential infrastructure inadequacies.
- H17: perceived risk positively influences distrust [62,63]. Privacy concerns amplify general system skepticism.
- H18: eHealth literacy positively influences performance expectancy [16,17]. Digital health competencies enable informed benefit recognition.
- H19: eHealth literacy positively influences effort expectancy [16,17]. Digital literacy reduces learning curves through transferable skills.
- H20: eHealth literacy positively influences facilitating conditions [17,64]. Digital competencies expand perceived scope of adequate support.
- H21: eHealth literacy positively influences self-efficacy [16,64]. Digital health competencies enhance confidence through mastery experiences.
- H22: eHealth literacy negatively influences distrust [46,47]. Digital understanding may reduce skepticism, though sophisticated users may paradoxically identify system vulnerabilities.
- H23: eHealth literacy negatively influences perceived risk [44,45]. Digital knowledge enables accurate risk assessment that typically reduces threat overestimation.
- H24: distrust negatively influences performance expectancy [65,66]. Skepticism introduces doubts about whether promised benefits will materialize.
- H25: distrust negatively influences effort expectancy [65,66]. Skepticism elevates perceived complexity through suspicion of deliberately confusing design.
- H26: distrust negatively influences facilitating conditions [66,67]. Skepticism reduces confidence that adequate support truly exists.

Together, H24-H26 specify distrust's indirect pathways complementing the direct effect in H4, capturing how system skepticism operates through both immediate discouragement and systematic erosion of acceptance perceptions.

In addition, self-rated health status and health interest were incorporated as observed exogenous covariates predicting performance expectancy and distrust. This specification allows examination of whether perceived health condition and baseline health motivation shape core acceptance beliefs, thereby addressing the role of health-related need in mHealth adoption.

Research Objectives

This study pursued 3 objectives. First, we tested an extended UTAUT model incorporating perceived risk, distrust, eHealth literacy, and health-related covariates (health status and health interest), evaluating all 26 hypothesized relationships.

Second, we examined demographic heterogeneity by testing whether structural relationships differ across age groups (young adults aged 18-39 years, middle-aged adults aged 40-59 years, and older adults aged ≥60 years) and sex groups using multigroup SEM.

Third, we explored app-type-specific adoption mechanisms for different mHealth apps.

Hypotheses were specified for the overall model, whereas subgroup and app-type analyses were conducted as comparative extensions.

Methods

Ethical Considerations

This cross-sectional online survey study examined determinants of mHealth adoption in Japan's aging society between November 11 and 18, 2023. Ethical approval was obtained from the Waseda University Ethics Committee (number 2023-250), and the study complied with Japan's Act on the Protection of Personal Information [41]. Electronic informed consent was obtained from all participants through the online survey platform. Participants received a small number of reward points corresponding to approximately ¥10 or less than US $0.10 for completing the survey. No personally identifiable information was collected, and study reporting followed the CHERRIES (Checklist for Reporting Results of Internet E-Surveys; Multimedia Appendix 1) guidelines [68]. Survey data were stored on secure, password-protected servers accessible only to the research team and were exported to encrypted files for analysis. All analyses used deidentified data, and results are reported only in aggregate so that no individual participant can be identified.

Participants and Sampling

Sample size was determined using 2 approaches. For population estimation, normal approximation with finite-population correction for Japan (126 million) required a minimum n=385 (95% confidence; 5% margin of error; P=.50). For SEM (9 constructs; 31 indicators; 26 paths), a minimum n=310 was required (10:1 observations-to-parameters ratio) [69,70]. For multigroup analyses across 3 age groups and 2 sexes, 100 to 150 participants per subgroup ensured adequate power [32,33]. The target sample (N=960) exceeded all requirements and provided power >0.80 for detecting moderate effects (β≥0.20). Age was collected as a continuous variable (in years) and was also categorized into seven 10-year age strata for quota sampling and descriptive purposes.

Participants were recruited from the Cross Marketing, Inc (Tokyo, Japan) internet research panel using quota sampling across 7 age strata (18-27 years, 28-37 years, 38-47 years, 48-57 years, 58-67 years, 68-77 years, and ≥78 years) to ensure representation across the adult lifespan.

Of 23,434 invited panelists, a total of 2203 consented to participate. Data quality was ensured through multiple procedures, including the exclusion of responses with inconsistent demographic information and extremely short completion times. In addition, an attention-check item was embedded within the 5-point Likert scale questionnaire (eg, "Please select 'strongly agree' for this item"). Respondents who selected any option other than the instructed response were considered inattentive and excluded from the final sample.

XSL•FO
RenderX

After applying these criteria, a total of 960 valid responses were retained for analysis.

For inferential analyses, these 7 strata were collapsed into 3 theoretically meaningful groups: young adults (aged 18-39 years), middle-aged adults (aged 40-59 years), and older adults (aged ≥60 years), ensuring adequate statistical power for multigroup SEM. We chose 3 life-stage groups rather than all seven 10-year strata to avoid very small subgroup sizes, which would have produced unstable parameter estimates and less interpretable age-specific patterns in the SEM.

Participants were not required to have prior experience with mHealth apps, as the study assessed both behavioral intention and current use. However, all participants necessarily had internet access and basic digital literacy due to the online survey format.

Measurement Instruments

Participants reported demographic variables (age, sex, marital status, educational level, and current use of health monitoring apps). Technology acceptance constructs were measured using 9 established scales, each via 5-point Likert items (1=strongly disagree and 5=strongly agree). The following latent constructs were assessed: performance expectancy, effort expectancy, social influence, facilitating conditions, privacy risk, self-efficacy, distrust, behavioral intention, and eHealth literacy (eHealth Literacy Scale).

All survey items and their sources are provided in Table S1 in Multimedia Appendix 2. Items served as observed variables for latent constructs in SEM. Measurement validity and reliability were evaluated using confirmatory factor analysis (CFA).

Statistical Analysis

All analyses were performed using IBM SPSS Statistics (version 30) and IBM SPSS AMOS (version 30) [71]. Statistical significance was set at $P<.05$. Descriptive statistics summarized participant characteristics. Internal consistency was assessed using Cronbach α [72]. Multicollinearity was assessed using the variance inflation factor <5 [71,72].

CFA tested measurement validity, establishing adequacy via standardized factor loadings [72,73], average variance extracted (AVE≥0.50), and composite reliability (CR≥0.70). Discriminant validity was determined using the Fornell-Larcker criterion [72].

SEM using maximum likelihood estimation was performed to test hypothesized paths [71,74]. In the structural model, self-rated health status and health interest were specified as observed covariates predicting performance expectancy and distrust to account for individual differences in perceived health condition and baseline health motivation. Model fit was evaluated using multiple indices: $\chi^2/df<3$, CFI>.90, TLI>.90, goodness-of-fit index (GFI)>.90, adjusted GFI (AGFI)>.90, and root mean square error of approximation (RMSEA)≤.08 [72,73,75,76]. Descriptive statistics and inter-construct correlations appear in Table S2 in Multimedia Appendix 3.

Multi-group SEM examined sex and age differences with configural, metric, and structural invariance testing [32,33]. Chi-square difference tests ($\Delta\chi^2$) evaluated group differences [77,78]. Binary logistic regression analysis identified predictors of mHealth app use and behavioral intention by app type, reporting odds ratios (ORs) and 95% CIs [71,72]. Binary logistic regression analysis was used for app-specific use patterns because the outcome variable was dichotomous (current use vs nonuse) and the number of current users per app type was insufficient for stable SEM estimation (blood pressure: n=173; heart rate: n=118; physical activity: n=110). For the app-specific logistic regression analyses, construct scores were calculated as the mean scores of the corresponding questionnaire items and used as independent variables.

Results

Participant Characteristics

Among 960 valid respondents (476/960, 49.6% male; 484/960, 50.4% female; mean age 51.8, SD 17.8 years), participants were stratified into young adults (aged 18-39 years; 298/960, 31%), middle-aged adults (aged 40-59 years; 275/960, 28.6%), and older adults (aged ≥60 years; 387/960, 40.3%). For descriptive purposes, Table 1 reports age distributions across the original seven 10-year strata used for quota sampling, whereas subsequent inferential analyses (multigroup SEM and logistic regression) are based on the 3 life-stage groups (young adults, middle-aged adults, and older adults). Educational attainment was high, with university or graduate degrees representing the largest category (482/960, 50.2%), followed by high school graduates (272/960, 28.3%) and junior college graduates (156/960, 16.3%). Most participants were married (633/960, 65.9%), reported health interest (786/960, 81.9%), and self-assessed their health status as good or somewhat better (526/960, 54.8%).

XSL•FO
RenderX

Table 1. Demographic characteristics of participants (N=960).

Demographic characteristics	Frequency, n (%)
Age (years)	
18-27	130 (13.5)
28-37	138 (14.4)
38-47	138 (14.4)
48-57	143 (14.9)
58-67	136 (14.2)
68-77	138 (14.4)
≥78	137 (14.3)
Sex	
Male	476 (49.6)
Female	484 (50.4)
Education	
Junior high school graduate	31 (3.2)
High school graduate	272 (28.3)
Junior college or vocational school graduate	156 (16.3)
University graduate or master's-level graduate	482 (50.2)
I do not want to answer	19 (2)
Marital status	
Unmarried	318 (33.1)
Married	633 (65.9)
I do not want to answer	9 (0.9)
Interest in health	
Not interested at all	22 (2.3)
Not very interested	40 (4.2)
Neither one nor the other	104 (10.8)
Somewhat curious	387 (40.3)
Very interested	399 (41.6)
Do not know	8 (0.8)
Health status	
Bad	26 (2.7)
Somewhat worse	153 (15.9)
Neither one nor the other	252 (26.3)
Somewhat better	364 (37.9)
Good	162 (16.9)
I do not want to answer	3 (0.3)

eHealth Literacy

Table 2 provides the results on eHealth literacy by sex, age, and educational background.

The overall mean eHealth literacy score was 22.84 (SD 6.31). No significant sex differences were observed (male: 22.90, SD 6.72 and female: 22.79, SD 5.98), nor were there significant differences by educational level. However, age-related differences were identified, with participants aged 28-37 years (mean 24.42, SD 6.22) demonstrating the highest eHealth literacy scores, while those aged ≥78 years showed the lowest scores (mean 20.98, SD 6.14).

XSL•FO
RenderX

Table 2. eHealth literacy and demographic characteristics of participants.

Demographic characteristics and eHealth literacy	Mean (SD)
Age (years)	
18-27	23.1077 (7.20922)
28-37	24.4203 (6.22429)
38-47	23.6739 (6.00871)
48-57	22.7063 (6.58298)
58-67	22.5515 (5.53082)
68-77	22.4493 (6.28644)
≥78	20.9781 (6.14347)
Sex	
Male	22.8971 (6.72309)
Female	22.7851 (5.98059)
Education	
Junior high school graduate	20.7097 (7.52415)
High school graduate	22.4044 (5.85354)
Junior college or vocational school graduate	22.859 (6.63271)
University graduate or master's-level graduate	23.3631 (6.39248)
I do not want to answer	19.1579 (6.27396)
Total	22.8406 (6.3053)

Descriptive Statistics for Study Constructs

Descriptive statistics (mean and SD) for all constructs and questionnaire items are provided in Table S2 in Multimedia Appendix 3. Across constructs, mean values for behavioral intention, distrust, and performance expectancy tended to be higher than for perceived risk and social influence, indicating broadly positive expectations for mHealth alongside some skepticism toward information reliability.

Current Use of mHealth Apps

Regarding mHealth use, a substantial majority (688/960, 71.7%) reported no current use, while 272/960 (28.3%) used at least 1 app. Among those, blood pressure monitoring showed the highest usage (173/960, 18.0%), followed by heart rate monitoring (118/960, 12.3%) and physical activity tracking (110/960, 11.5%). Detailed usage rates for each type of mHealth app are presented in Multimedia Appendix 4.

Usage patterns differed significantly by age group, with older adults demonstrating higher adoption of disease-management apps, while younger adults showed greater engagement with wellness-oriented apps.

Measurement Model Validation

Table 3 summarizes the convergent and discriminant validity indices (standardized loading ranges, AVE, CR, Cronbach α, maximum shared variance, and discriminant validity judgments) for all constructs included in the measurement model. The measurement model demonstrated satisfactory psychometric properties, with all standardized factor loadings exceeding 0.60 and most exceeding 0.70 (Table S3 in Multimedia Appendix 5 and Table S2 in Multimedia Appendix 3).

Convergent validity was supported, as most constructs achieved AVE values above the recommended threshold of 0.50. Self-efficacy and distrust were at the threshold level (approximately 0.50) but remained acceptable given their CR>0.70 and substantial factor loadings. To further address the marginal AVE values for self-efficacy and distrust, we estimated alternative CFA and structural models in which the lowest-loading indicator of each construct was removed. These reduced-item models yielded slightly higher AVE estimates and marginally improved fit indices (eg, CFI≈0.95; RMSEA≈0.05), but the overall pattern and significance of structural paths remained essentially unchanged. We therefore retained the original 3-item scales for self-efficacy and distrust in the main model to preserve content validity and comparability with prior research, while interpreting their structural effects with appropriate caution.

Discriminant validity was confirmed using the Fornell-Larcker criterion, as all constructs demonstrated AVE values greater than their maximum shared variance.

XSL•FO
RenderX

Table 3. Measurement model indices and discriminant validity (N=960). Discriminant validity was confirmed via the Fornell-Larcker criterion (average variance extracted [AVE] > maximum shared variance [MSV]). Detailed factor loadings are available in Table S2 in Multimedia Appendix 3. AVE values for self-efficacy and distrust were near the 0.50 threshold but were considered acceptable given composite reliability (CR) >0.70 and substantial factor loadings.

Construct	Standardized loading range	AVE	CR	Cronbach α	MSV	Discriminant validity
Self-efficacy	0.656-0.823	0.501	0.752	0.731	0.214	Supported
Social influence	0.767-0.824	0.644	0.844	0.871	0.54	Supported
Privacy risk	0.659-0.921	0.658	0.852	0.83	0.035	Supported
Performance expectancy	0.788-0.860	0.687	0.868	0.868	0.578	Supported
Effort expectancy	0.799-0.825	0.659	0.795	0.779	0.578	Supported
Distrust	0.612-0.773	0.5	0.749	0.739	0.035	Supported
Facilitating conditions	0.693-0.848	0.65	0.847	0.827	0.578	Supported
Behavioral intention	0.841-0.901	0.761	0.905	0.907	0.543	Supported
eHealth literacy	0.745-0.872	0.665	0.941	0.935	0.367	Supported

Overall Structural Model

The hypothesized extended UTAUT model is shown in Figure 1. The extended structural equation model including health interest and health status demonstrated good overall fit to the data: χ^2_{452}=932.4; *P*<.001; χ^2/df=2.1; GFI=0.889; AGFI=0.863; CFI=0.953; TLI=0.945; RMSEA=0.047 (90% CI 0.043-0.052). Although the chi-square statistic was statistically significant, which is common in large samples, the ratio of chi-square to degrees of freedom remained within an acceptable range. In addition, both incremental and absolute fit indices met or exceeded commonly recommended thresholds, indicating adequate model fit.

The structural model explained a substantial proportion of variance in behavioral intention (R^2=0.716). The explanatory power for endogenous constructs varied across variables, ranging from 0.4% for perceived risk to 82.1% for effort expectancy. Facilitating conditions (R^2=0.666) and performance expectancy (R^2=0.628) also showed substantial explained variance. These results indicate that the extended UTAUT framework provides strong explanatory power for mHealth acceptance in the study sample.

The structural path coefficients for the overall sample are presented in Table 4.

Effort expectancy exerted the strongest positive effect on behavioral intention (β=0.404; *P*<.001), followed by facilitating conditions (β=0.349; *P*<.001) and performance expectancy (β=0.188; *P*<.001), supporting H1-H3. In contrast, distrust did not have a statistically significant effect on behavioral intention (β=−0.027; *P*=.45); therefore, H4 was not supported. Together, these findings indicate that perceived ease of use, infrastructural support, and expected health benefits are the primary proximal determinants of mHealth adoption intention in the overall sample, while distrust does not exert a direct influence.

Social influence emerged as the most powerful distal predictor in the model. It showed strong positive effects on effort expectancy (β=0.811; *P*<.001), facilitating conditions (β=0.548; *P*<.001), and performance expectancy (β=0.662; *P*<.001), supporting H5-H7. Social influence also significantly enhanced self-efficacy (β=0.422; *P*<.001) and eHealth literacy (β=0.507; *P*<.001), supporting H8 and H9. In addition, social influence significantly reduced distrust (β=−0.248; *P*<.001), supporting H10. These results indicate that social environments strongly shape perceptions of usefulness, ease of use, infrastructural support, digital competencies, and trust in mHealth technologies.

eHealth literacy demonstrated a complex pattern of effects. Higher eHealth literacy significantly improved facilitating conditions (β=0.323; *P*<.001), self-efficacy (β=0.204; *P*<.001), performance expectancy (β=0.097; *P*=.04), and effort expectancy (β=0.164; *P*<.001), supporting H18-H21. However, it did not significantly reduce perceived risk (β=−0.063; *P*=.21), leading to rejection of H23. Notably, eHealth literacy increased distrust (β=0.150; *P*=.02), yielding a statistically significant but directionally reversed effect relative to H22. These findings suggest that individuals with higher digital health competencies may develop a more critical awareness of potential risks and limitations in digital health systems.

Perceived risk consistently undermined acceptance-related perceptions. It significantly increased distrust (β=0.265; *P*<.001), supporting H17, and negatively affected facilitating conditions (β=−0.112; *P*=.003), effort expectancy (β=−0.117; *P*=.002), and performance expectancy (β=−0.079; *P*=.04), supporting H14-H16. These findings indicate that privacy and security concerns simultaneously increase distrust while reducing perceived usefulness, ease of use, and infrastructural support for mHealth technologies.

Self-efficacy did not exhibit significant direct effects on the core UTAUT constructs in the overall model. Its paths to performance expectancy (β=0.100; *P*=.06), effort expectancy (β=−0.022; *P*=.67), and facilitating conditions (β=0.046; *P*=.37) were not significant, and therefore H11-H13 were not supported. These results suggest that self-efficacy primarily functions as an outcome of social influence and eHealth literacy rather than as a direct determinant.

XSL•FO
RenderX

Table 4. Structural path coefficients for overall sample (N=960). P values <.05 were considered statistically significant.

Category	Path	β	*P* value	Hypothesis	Result
BI[a]	EE[b] → BI	0.404	<.001	H2	Supported
BI	PE[c] → BI	0.188	<.001	H1	Supported
BI	FC[d] → BI	0.349	<.001	H3	Supported
BI	DT[e] → BI	–0.027	.45	H4	Not supported
SI[f]	SI → EE	0.811	<.001	H6	Supported
SI	SI → FC	0.548	<.001	H7	Supported
SI	SI → PE	0.662	<.001	H5	Supported
SI	SI → SE[g]	0.422	<.001	H8	Supported
SI	SI → eHL[h]	0.507	<.001	H9	Supported
SI	SI → DT	–0.248	<.001	H10	Supported
eHL	eHL → FC	0.323	<.001	H20	Supported
eHL	eHL → SE	0.204	<.001	H21	Supported
eHL	eHL → DT	0.15	.02	H22	Supported (opposite direction)
eHL	eHL → PE	0.097	.04	H18	Supported
eHL	eHL → PR[i]	–0.063	.21	H23	Not supported
eHL	eHL → EE	0.164	<.001	H19	Supported
PR	PR → DT	0.265	<.001	H17	Supported
PR	PR → FC	–0.112	.003	H16	Supported
PR	PR → EE	–0.117	.002	H15	Supported
PR	PR → PE	–0.079	.04	H14	Supported
SE	SE → PE	0.1	.06	H11	Not supported
SE	SE → EE	–0.022	.67	H12	Not supported
SE	SE → FC	0.046	.37	H13	Not supported
DT	DT → EE	0.003	.94	H25	Not supported
DT	DT → PE	–0.017	.70	H24	Not supported
DT	DT → FC	0.125	.006	H26	Not supported
Health-related covariates	Health status → PE	–0.016	.64	—[j]	—
Health-related covariates	Health interest → PE	0.133	<.001	—	—
Health-related covariates	Health status → DT	0.049	.33	—	—
Health-related covariates	Health interest → DT	–0.054	.29	—	—

[a]BI: behavioral intention.
[b]EE: effort expectancy.
[c]PE: performance expectancy.
[d]FC: facilitating conditions.
[e]DT: distrust.
[f]SI: social influence.
[g]SE: self-efficacy.
[h]eHL: eHealth literacy.
[i]PR: perceived risk.
[j]Not applicable.

XSL•FO
RenderX

Distrust showed limited direct effects. It did not significantly influence effort expectancy (β=0.003; *P*=.94) or performance expectancy (β=−0.017; *P*=.70); therefore, H25 and H24 were not supported. Contrary to H26, distrust showed a small but significant positive effect on facilitating conditions (β=0.125; *P*=.006), indicating that the predicted negative relationship was not confirmed; therefore, H26 was not supported. This unexpected direction suggests that distrust may prompt compensatory information-seeking behavior rather than withdrawal from support structures.

Health-related covariates showed selective effects. Health interest significantly increased performance expectancy (β=0.133; *P*<.001), indicating that individuals with greater health interest perceive higher utility in mHealth technologies. In contrast, self-rated health status did not significantly influence performance expectancy (β=−0.016; *P*=.64), and neither health interest nor health status significantly predicted distrust (β=−0.054; *P*=.29 and β=0.049; *P*=.33, respectively). These covariate effects were included for control purposes and were not part of the formal hypothesis testing framework.

Sex-Stratified Analysis

The multigroup structural equation model for sex demonstrated acceptable to good model fit. The chi-square statistic was significant (χ^2_{904}=1994.9; *P*<.001), which is expected given the large sample size; however, the ratio of chi-square to degrees of freedom was within an acceptable range (χ^2/df=2.21).

Incremental fit indices indicated good model fit (CFI=0.944; TLI=0.935; incremental fit index (IFI)=0.944). The RMSEA value was 0.035 (90% CI 0.033-0.038; P_{close}>.99), indicating excellent fit. Although the GFI (0.882) and AGFI (0.853) were slightly below the conventional threshold, these indices are known to be sensitive to sample size and model complexity.

Overall, these results suggest that the model provides a satisfactory representation of the data and is appropriate for subsequent multigroup comparisons by sex.

The explanatory power of the model remained high across sex groups. The model explained a substantial proportion of variance in behavioral intention, accounting for 71.7% of the variance among males and 70.1% among females. Sex-stratified structural path coefficients are presented in Table 5.

First, social influence significantly reduced distrust among males (β=−0.325; *P*<.001) but not among females (β=−0.100; *P*=.16; $\Delta\chi^2$=5.7; *P*=.02). This finding indicates that normative social cues function as more effective trust-building mechanisms for male users.

Second, eHealth literacy significantly reduced perceived risk among females (β=−0.210; *P*<.001) but not among males (β=−0.080; *P*=.11; $\Delta\chi^2$=5.2; *P*=.02). This suggests that digital health competencies play a stronger role in mitigating privacy and security concerns among women.

Third, distrust exerted significantly stronger effects on effort expectancy among females (β=0.287; *P*<.001) than among males (β=0.031; *P*=.48; $\Delta\chi^2$=17.3; *P*<.001). This indicates that distrust more strongly shapes perceived ease of use among female users.

Fourth, distrust also had a significantly stronger effect on performance expectancy among females (β=0.226; *P*<.001) than among males (β=−0.020; *P*=.65; $\Delta\chi^2$=12.6; *P*<.001), suggesting that distrust influences perceived usefulness primarily among women.

Fifth, performance expectancy exerted significantly stronger effects on behavioral intention among females (β=0.388; *P*<.001) than among males (β=0.200; *P*<.001; $\Delta\chi^2$=4.0; *P*=.04). These results indicate that perceived health benefits constitute a more central driver of mHealth adoption intentions for female users.

In contrast, no significant sex difference was observed for the path from distrust to behavioral intention ($\Delta\chi^2$=1.7; *P*=.19), although the effect was statistically significant only among females.

XSL•FO
RenderX

Table 5. Sex-stratified path coefficients. Chi-square difference tests identified 5 statistically significant sex differences (Δχ2>3.8; P<.05).

Pathway and path direction		Male				Female				Chi-square difference (*df*); *P* value
		β	SE	T value	*P* value	β	SE	T value	*P* value	
SI[a] pathways										
eHL[b]	SI → eHL	0.493	0.061	9.436	<.001	0.433	0.057	7.765	<.001	0.9 (1); .34
SE[c]	SI → SE	0.406	0.065	6.497	<.001	0.324	0.068	5.145	<.001	—[d]
DT[e]	SI → DT	–0.325	0.071	–4.287	<.001	–0.1	0.055	–1.395	.16	5.7 (1); .02
PE[f]	SI → PE	0.616	0.061	10.31	<.001	0.59	0.058	9.719	<.001	—
EE[g]	SI → EE	0.77	0.06	12.08	<.001	0.699	0.055	11.702	<.001	0.9 (1); .34
FC[h]	SI → FC	0.559	0.066	9.601	<.001	0.672	0.063	12.536	<.001	3.4 (1); .07
eHL pathways										
SE	eHL → SE	0.22	0.052	3.802	<.001	0.264	0.062	4.452	<.001	—
PR	eHL → PR[i]	–0.08	0.061	–1.585	.11	–0.21	0.071	–4.029	<.001	5.2 (1); .02
DT	eHL → DT	0.135	0.052	2.081	.04	0.26	0.052	3.718	<.001	1.6 (1); .21
PE	eHL → PE	0.131	0.04	2.825	.005	0.097	0.049	1.823	.07	1.0 (1); .32
EE	eHL → EE	0.201	0.037	4.425	<.001	0.181	0.045	3.648	<.001	—
FC	eHL → FC	0.329	0.046	6.958	<.001	0.273	0.053	5.885	<.001	—
PR pathways										
DT	PR → DT	0.259	0.037	4.676	<.001	0.178	0.032	3.004	.003	—
FC	PR → FC	–0.140	0.032	–3.540	<.001	–0.143	0.032	–3.821	<.001	—
EE	PR → EE	–0.118	0.026	–3.093	.002	–0.212	0.028	–5.08	<.001	1.7 (1); .20
PE	PR → PE	–0.054	0.028	–1.377	.17	–0.165	0.031	–3.647	<.001	0.7 (1); .40
SE pathways										
DT	SE → DT	0.096	0.065	1.32	.19	0.013	0.051	0.181	.86	—
PE	SE → PE	0.142	0.05	2.737	.006	–0.027	0.048	–0.496	.62	—
EE	SE → EE	–0.007	0.045	–0.139	.89	0.017	0.043	0.342	.73	—
FC	SE → FC	0.044	0.055	0.954	.34	–0.029	0.051	–0.628	.53	—
DT pathways										
EE	DT → EE	0.031	0.045	0.705	.48	0.287	0.06	5.795	<.001	17.3 (1); <.001
FC	DT → FC	0.129	0.055	2.841	.004	0.102	0.064	2.481	.01	—
PE	DT → PE	–0.020	0.049	–0.453	.65	0.226	0.064	4.414	<.001	12.6 (1); <.001
BI[j] pathways										
BI	PE → BI	0.2	0.064	3.881	<.001	0.388	0.067	7.362	<.001	4.0 (1); .04
BI	FC → BI	0.336	0.062	5.998	<.001	0.177	0.068	2.652	.008	1.4 (1); .23
BI	EE → BI	0.358	0.087	5.43	<.001	0.325	0.103	4.109	<.001	0.6 (1); .45
BI	DT → BI	–0.051	0.048	–1.404	.16	–0.102	0.065	–2.486	.01	1.7 (1); .19
Health-related covariates										
PE	health status → PE	–0.017	0.022	–0.495	.62	–0.063	0.022	–1.702	.09	—
DT	health status → DT	0.049	0.03	0.957	.34	–0.136	0.026	–2.608	.009	—
PE	health interest → PE	0.132	0.023	3.785	<.001	0.172	0.027	4.608	<.001	—
DT	health interest → DT	–0.053	0.031	–1.044	.30	–0.014	0.03	–0.27	.79	—

XSL•FO
RenderX

[a]SI: social influence.
[b]eHL: eHealth literacy.
[c]SE: self-efficacy.
[d]Not applicable.
[e]DT: distrust.
[f]PE: performance expectancy.
[g]EE: effort expectancy.
[h]FC: facilitating conditions.
[i]PR: perceived risk.
[j]BI: behavioral intention.

Age-Stratified Analysis

For age-stratified multigroup SEM, participants were categorized into 3 groups: young adults (aged 18-39 years; n=298), middle-aged adults (aged 40-59 years; n=275), and older adults (aged ≥60 years; n=387).

The multigroup structural equation model across the 3 age groups demonstrated acceptable to good fit: χ^2_{1356}=2762.4; *P*<.001; χ^2/df=2.04; GFI=0.849; AGFI=0.812; normed fit index 0.871; CFI=0.929; TLI=0.917; IFI=0.930; RMSEA=0.033 (90% CI 0.031-0.035; P_{close}>.99).

Although GFI and AGFI were slightly below the conventional threshold of 0.90, these indices are known to be sensitive to model complexity and sample size. In contrast, incremental fit indices (CFI, TLI, and IFI) exceeded recommended thresholds, and RMSEA indicated excellent model fit. Taken together, these results suggest that the age-stratified model provides an adequate and robust representation of the observed data.

The explanatory power of the model remained high across age groups. The model explained 77.1% of the variance in behavioral intention among young adults, 67.5% among middle-aged adults, and 69.1% among older adults (Table 6).

First, social influence exerted significantly different effects on performance expectancy across age groups ($\Delta\chi^2_2$=11.76; *P*<.01). The effect was strongest among young adults (β=0.919; *P*<.001), followed by middle-aged adults (β=0.617; *P*<.001) and older adults (β=0.574; *P*<.001), indicating that social cues play a more prominent role in shaping perceived usefulness among younger individuals.

Second, eHealth literacy showed significant age differences in its effect on perceived risk ($\Delta\chi^2_2$=13.7; *P*<.01). Specifically, eHealth literacy significantly reduced perceived risk among middle-aged adults (β=−0.187; *P*=.005) and older adults (β=−0.231; *P*<.001), whereas no significant association was observed among young adults (β=0.061; *P*=.36). This suggests that digital health competencies function as a risk-mitigating factor primarily in older populations.

Third, self-efficacy exhibited the largest age-related heterogeneity in its effect on performance expectancy ($\Delta\chi^2_2$=17.1; *P*<.001). Notably, self-efficacy negatively influenced performance expectancy among young adults (β=−0.324; *P*<.001), while showing positive effects among middle-aged adults (β=0.161; *P*=.02) and older adults (β=0.110; *P*=.05). This reversal suggests that confidence in one's digital abilities may lead to more critical evaluations of mHealth usefulness among younger users.

Fourth, distrust demonstrated significant age differences in its effect on effort expectancy ($\Delta\chi^2_2$=13.5; *P*<.01). Distrust strongly increased perceived difficulty of use among young adults (β=0.349; *P*<.001), whereas this relationship was not significant among middle-aged adults (β=0.075; *P*=.22) or older adults (β=0.039; *P*=.37).

Fifth, distrust also showed significant age differences in its effect on performance expectancy ($\Delta\chi^2_2$=10.7; *P*<.01). Distrust positively influenced performance expectancy among young adults (β=0.249; *P*<.001), while no significant effects were observed among middle-aged adults (β=−0.002; *P*=.97) or older adults (β=0.022; *P*=.66), indicating divergent interpretations of distrust across age groups.

In contrast, no significant age differences were observed in the effects of performance expectancy ($\Delta\chi^2_2$=1.7; *P*>.05), facilitating conditions ($\Delta\chi^2_2$=3.6; *P*>.05), or distrust ($\Delta\chi^2_2$=9.1; *P*>.01) on behavioral intention, and age differences in the direct effect of distrust on behavioral intention were confined to middle-aged adults. Thus, while distrust showed some age-contingent influence on intention, the core positive determinants of mHealth adoption (performance expectancy, effort expectancy, and facilitating conditions) remained largely invariant across age groups.

These findings indicate that the proximal determinants of mHealth adoption intention are largely invariant across age groups, despite substantial heterogeneity in upstream cognitive and perceptual mechanisms. Health-related covariates showed a consistent pattern across age groups. Health interest significantly increased performance expectancy in all 3 age groups (young adults: β=0.133; *P*=.002; middle-aged adults: β=0.090; *P*=.05; older adults: β=0.202; *P*<.001), with the strongest effect among older adults, whereas self-rated health status did not significantly predict performance expectancy or distrust in any group.

XSL•FO
RenderX

Table 6. Age-stratified path coefficients. Chi-square difference tests identified 5 statistically significant age differences (Δχ2>6.0; P<.05).

Pathway and path direction	Young adults			Middle-aged adults			Older adults			Age, chi-square difference (*df*); *P* value
	β	T value	*P* value	β	T value	*P* value	β	T value	*P* value	
SI[a] pathways										
eHL[b] ← SI	0.593	7.698	<.001	0.31	4.459	<.001	0.456	7.437	<.001	__[c]
DT[d] ← SI	–0.234	–2.285	.02	–0.215	–2.676	.007	–0.159	–2.228	.03	—
SE[e] ← SI	0.602	6.286	<.001	0.353	4.798	<.001	0.229	3.273	.001	—
PE[f] ← SI	0.919	8.165	<.001	0.617	8.86	<.001	0.574	9.026	<.001	11.8 (2); <.01
EE[g] ← SI	1.03	8.219	<.001	0.745	10.11	<.001	0.707	11.738	<.001	—
FC[h] ← SI	0.599	6.62	<.001	0.626	9.581	<.001	0.651	11.115	<.001	—
eHL pathways										
PR[i] ← eHL	0.061	0.912	.36	–0.187	–2.79	.005	–0.231	–4.14	<.001	13.7 (2); <.01
DT ← eHL	0.366	3.664	<.001	0.119	1.513	.13	0.089	1.274	.20	—
SE ← eHL	0.166	2.041	.04	0.288	4.055	<.001	0.243	3.601	<.001	—
PE ← eHL	0.103	1.31	.19	0.09	1.551	.12	0.101	1.815	.07	—
EE ← eHL	–0.009	–0.11	.91	0.217	3.62	<.001	0.25	5.015	<.001	—
FC ← eHL	0.322	4.51	<.001	0.339	6.011	<.001	0.209	4.045	<.001	—
PR pathways										
DT ← PR	0.153	2.116	.03	0.25	3.303	<.001	0.186	3.026	.002	—
FC ← PR	–0.108	–2.202	.03	–0.234	–4.77	<.001	–0.126	–2.951	.003	—
EE ← PR	–0.16	–2.948	.003	–0.098	–1.832	.07	–0.139	–3.444	<.001	—
PE ← PR	–0.184	–3.53	<.001	–0.114	–2.134	.03	0.009	0.186	.85	—
Self–efficacy pathways										
PE ← SE	–0.324	–3.384	<.001	0.161	2.399	.02	0.11	1.998	.05	17.1 (2); <.001
EE ← SE	–0.222	–2.271	.02	0.089	1.331	.18	0.061	1.307	.19	—
FC ← SE	–0.024	–0.305	.76	0.027	0.45	.65	0.014	0.275	.78	—
Distrust pathways										
EE ← DT	0.349	4.906	<.001	0.075	1.225	.22	0.039	0.899	.37	13.5 (2); <.01
FC ← DT	0.204	3.541	<.001	0.036	0.67	.50	0.09	1.985	.05	—
PE ← DT	0.249	3.94	<.001	–0.002	–0.036	.97	0.022	0.438	.66	10.7 (2); <.01
BI[j] pathways										
BI ← EE	0.377	3.259	.001	0.231	2.301	.02	0.419	5.63	<.001	1.5 (2); .48
BI ← PE	0.272	3.16	.002	0.216	3.025	.002	0.284	5.299	<.001	1.7 (2); .43
BI ← FC	0.352	4.158	<.001	0.42	4.595	<.001	0.254	3.964	<.001	3.6 (2); .17
BI ← DT	–0.113	–2.209	.03	–0.171	–3.159	.002	0.028	0.702	.48	9.1 (2); .01
Health-related covariates										
PE ← health status	–0.039	–0.911	.36	–0.043	–0.925	.36	–0.022	–0.53	.60	—
DT ← health status	–0.09	–1.39	.16	–0.009	–0.126	.90	–0.04	–0.713	.49	—
PE ← health interest	0.133	3.126	.002	0.09	1.957	.05	0.202	4.707	<.001	—
DT ← health interest	0.017	0.268	.79	–0.029	–0.423	.67	0.091	1.634	.10	—

[a]SI: social influence.

[b]eHL: eHealth literacy.

XSL•FO
RenderX

[c]Not applicable.
[d]DT: distrust.
[e]SE: self-efficacy.
[f]PE: performance expectancy.
[g]EE: effort expectancy.
[h]FC: facilitating conditions.
[i]PR: perceived risk.
[j]BI: behavioral intention.

App-Specific Use Patterns: Binary Logistic Regression Analysis

Binary logistic regression analysis examined predictors of current use across 3 app types: blood pressure monitoring (n=173), heart rate monitoring (n=118), and physical activity tracking (n=110), using the original 7 age strata as categorical predictors. All models demonstrated adequate fit (Hosmer-Lemeshow $P>.05$) with a Nagelkerke R^2 of 0.14 for each model. Table 7 presents ORs and 95% CIs.

Table 7. App-specific predictors of current mHealth use.

Predictor	Blood pressure (n=173)	Heart rate (n=118)	Physical activity (n=110)
Behavioral constructs, OR[a] (95% CI)			
Behavioral intention	1.40 (1.08-1.80)[b]	1.77 (1.30-2.41)[c]	2.47 (1.76-3.47)[c]
eHealth literacy	1.85 (1.42-2.42)[c]	1.84 (1.35-2.53)[c]	1.57 (1.14-2.17)[d]
Demographics, OR (95% CI)			
Sex (Male; ref)	0.74 (0.52-1.05)	0.89 (0.59-1.33)	0.82 (0.54-1.25)
Age group (years; ref=18-27), OR (95% CI)			
28-37	0.64 (0.29-1.40)	0.62 (0.27-1.44)	0.97 (0.44-2.13)
38-47	0.88 (0.42-1.86)	1.21 (0.57-2.59)	1.23 (0.57-2.65)
48-57	1.25 (0.61-2.56)	0.94 (0.41-2.11)	1.58 (0.74-3.40)
58-67	2.34 (1.18-4.61)[b]	1.26 (0.57-2.77)	1.03 (0.45-2.39)
68-77	3.15 (1.63-6.12)[c]	2.37 (1.15-4.89)[b]	1.59 (0.74-3.45)
≥78 years	3.76 (1.93-7.32)[c]	1.91 (0.90-4.08)	1.06 (0.46-2.45)
Model fit			
Hosmer-Lemeshow $\chi^2(df)$; P value	6.92 (8); .55	3.55 (8); .90	7.76 (8); .46

[a]OR: odds ratio.
[b]$P<.05$.
[c]$P<.001$.
[d]$P<.01$.

Behavioral intention and eHealth literacy emerged as consistent significant predictors across all 3 app types. For blood pressure monitoring, behavioral intention (OR 1.40, 95% CI 1.08-1.80; $P=.01$) and eHealth literacy (OR 1.85, 95% CI 1.42-2.42; $P<.001$) significantly predicted current use. Sex was not a significant predictor (OR 0.74, 95% CI 0.52-1.05). Age demonstrated a pronounced gradient compared with users aged 18-27 years, those aged 58-67 years (OR 2.34, 95% CI 1.18-4.61; $P=.02$), aged 68-77 years (OR 3.15, 95% CI 1.63-6.12; $P<.001$), and aged ≥78 years (OR 3.76, 95% CI 1.93-7.32; $P<.001$) exhibited substantially elevated odds of use, reflecting the clinical relevance of hypertension management among older populations.

For heart rate monitoring, behavioral intention (OR 1.77, 95% CI 1.30-2.41; $P<.001$) and eHealth literacy (OR 1.84, 95% CI 1.35-2.53; $P<.001$) significantly predicted current use. Sex was not significant (OR 0.89, 95% CI 0.59-1.33). Age effects were less pronounced than for blood pressure monitoring, with only users aged 68-77 years demonstrating significantly elevated odds (OR 2.37, 95% CI 1.15-4.89; $P=.02$), suggesting broader age-range appeal for heart rate monitoring compared with blood pressure monitoring.

Physical activity tracking demonstrated the strongest behavioral intention effect among the 3 apps (OR 2.47, 95% CI 1.76-3.47; $P<.001$), with eHealth literacy also significantly predicting use (OR 1.57, 95% CI 1.14-2.17; $P=.006$). Sex (OR 0.82, 95% CI 0.54-1.25) and all age groups showed no significant differences compared with the reference group, indicating cross-demographic adoption patterns driven by intrinsic

XSL•FO
RenderX

motivation and digital health competency rather than age-related clinical necessity.

Discussion

Overview

This study examined the determinants of mHealth acceptance in Japan using an extended UTAUT framework that incorporated distrust, perceived risk, and eHealth literacy, while additionally modeling health status and health interest as covariates of performance expectancy and distrust. Overall, the findings confirmed the central role of the core UTAUT constructs—performance expectancy, effort expectancy, and facilitating conditions—in shaping behavioral intention, while also demonstrating that demographic heterogeneity in Japan is concentrated primarily in upstream cognitive and evaluative processes rather than in the final intention-formation stage [20-22,50,52,54,55]. This distinction is important because it suggests that the basic decision logic of mHealth acceptance is relatively stable, even in a superaged society, whereas the pathways through which individuals form beliefs about usefulness, ease of use, trust, and risk differ substantially by age and sex.

Principal Findings and Theoretical Contributions

Consistent with prior UTAUT-based research, performance expectancy, effort expectancy, and facilitating conditions emerged as robust predictors of behavioral intention in the overall model [20-22,50,52,54,55,79]. This pattern is also broadly consistent with earlier studies on mHealth adoption among older adults and chronic disease populations [23,24,28,29]. However, this study extends this literature by showing that, although these proximal determinants remain central, age differences were not statistically significant for the direct effects of performance expectancy and facilitating conditions on behavioral intention. This suggests that the core decision structure linking beliefs to intention may be more stable across the life course than is often assumed. Put differently, age appears to matter less for how people translate beliefs into intention than for how those beliefs are initially formed.

A major theoretical contribution of this study is the expanded role of social influence. In conventional UTAUT, social influence is usually modeled as one determinant of intention or belief formation [20,21]. In this study, however, social influence exerted broad and upstream effects on eHealth literacy, self-efficacy, performance expectancy, effort expectancy, and facilitating conditions, while also reducing distrust. This broader role exceeds many Western UTAUT applications [20,22] and is consistent with the Japanese context, where family members, peers, and health care professionals often function as key agents of reassurance, persuasion, and health-related decision support [18,19,27]. The age-stratified analyses further showed that the effect of social influence on performance expectancy was strongest among young adults. This finding suggests that, in contemporary Japan, younger users may evaluate the usefulness of mHealth partly through socially mediated expectations and peer-based information environments rather than solely through individual instrumental reasoning. In this sense, social influence in Japan may operate not only as a normative force but also as a mechanism of informal digital health education.

A second major contribution concerns eHealth literacy. Previous studies have generally positioned eHealth literacy as a uniformly beneficial facilitator of digital health use [16,17,45,46,64,80-82]. In Japan, the Japanese eHealth Literacy Scale has demonstrated strong internal consistency and construct validity in a large internet-based survey, with higher scores among middle-aged adults and frequent internet users [83]. Our results support this view only partially. Among middle-aged and older adults, higher eHealth literacy significantly reduced perceived risk, which is consistent with both international research and Japanese evidence linking eHealth literacy to improved health knowledge and screening-related practices [16,17,64]. However, this pattern did not extend to younger adults. In the younger group, eHealth literacy did not reduce perceived risk and was accompanied by more critical appraisals in related pathways. This suggests that eHealth literacy is better conceptualized as a context-dependent cognitive resource: it may reduce uncertainty in populations with lower baseline digital familiarity, but among digitally experienced users, it may instead sharpen awareness of privacy vulnerability, data misuse, or functional limitations. This interpretation aligns with research on the privacy paradox and digitally engaged patients [12,25,47,84,85].

A third theoretical contribution lies in the age-differentiated role of self-efficacy. In many prior studies, self-efficacy has shown positive associations with perceived usefulness and ease of use [20,57,60,61]. In this study, however, self-efficacy did not significantly predict performance expectancy, effort expectancy, or facilitating conditions in the overall model, and age-stratified analyses revealed a more complex pattern. Most notably, self-efficacy negatively predicted performance expectancy among younger adults but positively predicted it among middle-aged and older adults. This reversal suggests that digitally confident younger users may hold higher expectations for digital services and therefore evaluate mHealth apps more critically when the apps fail to meet their standards, whereas among middle-aged and older adults, self-efficacy plays the more conventional enabling role by lowering psychological barriers and making benefits more salient [23,26,28]. At the same time, the weak direct effect of self-efficacy in the overall model and its conceptual proximity to eHealth literacy indicate that self-efficacy in mHealth acceptance may operate more indirectly and contextually than previous models assume, partly overlapping with eHealth literacy as a mediating cognitive resource [16,17,64,82].

A fourth contribution is the clearer differentiation between distrust and perceived risk. Prior information systems and digital health research have often linked trust, distrust, and risk closely, but not always modeled them as distinct pathways [12-15,30,39,42,58,59,62,63]. In this study, perceived risk and distrust behaved differently. Perceived risk consistently undermined acceptance-related beliefs, whereas distrust showed selective and demographically contingent effects. In the overall model, distrust did not significantly reduce behavioral intention, but it influenced effort expectancy and performance expectancy in subgroup analyses, particularly among women and younger adults. These findings suggest that distrust should not be reduced

XSL•FO
RenderX

to a mere proxy for perceived risk. Rather, distrust appears to function as a broader evaluative orientation toward the technology, platform, or data environment, while perceived risk reflects a more specific anticipation of privacy or security harm [14,30,58,59]. This conceptual separation is valuable for digital health research, where users may simultaneously recognize objective risks and still maintain—or fail to maintain—confidence in the system's legitimacy and safety [12,13,40-43,86,87]. Consistent with the additional reduced-item models, the selective and mainly indirect role of distrust appears to be driven more by its structural position in the extended UTAUT framework than by weaknesses in the measurement model, reinforcing the need to interpret distrust as a nuanced, context-dependent barrier rather than a uniformly dominant deterrent to mHealth adoption.

Contrary to H26, distrust showed a small but significant positive effect on facilitating conditions. This unexpected finding may reflect compensatory information-seeking behavior, whereby skeptical users actively engage with available support structures to verify system safety rather than withdrawing from them [88]. In addition, the strong role of social influence and the differentiated effects of distrust are congruent with culturally specific notions such as sekentei (concern for social evaluation) and anshin (sense of safety and peace of mind), which shape how Japanese users balance institutional assurances, peer endorsement, and perceived vulnerability when deciding whether to adopt mHealth services [89,90].

Sex and Age Heterogeneity in the Japanese Context

The sex-stratified results revealed systematic, rather than incidental, heterogeneity. Compared with women, men showed a stronger negative pathway from social influence to distrust, suggesting that social endorsement and interpersonal reassurance play a larger role in trust formation among male users. By contrast, women showed a stronger negative pathway from eHealth literacy to perceived risk and a stronger positive pathway from performance expectancy to behavioral intention. These findings are broadly consistent with prior research suggesting that women are more attentive to privacy, risk, and benefit evaluation in digital contexts, whereas men may rely more heavily on social and structural cues [31,79,90]. In the Japanese context, this implies that male users may respond particularly well to social endorsement and visible legitimacy, whereas female users may be more influenced by the extent to which a system appears beneficial, transparent, and safe.

The age-stratified results were especially informative. Significant age differences were identified in the pathways from social influence to performance expectancy, eHealth literacy to perceived risk, self-efficacy to performance expectancy, and distrust to both effort expectancy and performance expectancy. In contrast, no statistically significant age differences were observed in the direct effects of performance expectancy or facilitating conditions on behavioral intention. This pattern strongly suggests that age-related heterogeneity in Japan is concentrated in the formation of beliefs rather than in the final translation of those beliefs into adoption intention.

Among young adults, social influence had the strongest effect on performance expectancy, and distrust significantly affected both effort expectancy and performance expectancy. This suggests that younger users do not simply accept or reject mHealth technologies based on distrust alone; rather, distrust may stimulate more active evaluation of usability and utility. Such a pattern is compatible with the idea of "informed skepticism," in which digitally experienced users are not necessarily more trusting, but instead more critical and selective [12,47,79,84]. Among middle-aged adults, eHealth literacy reduced perceived risk, and self-efficacy positively influenced performance expectancy, suggesting a more instrumental and balanced evaluative style. This may reflect the life-stage demands of work, family responsibility, and health management, which make reliability and usefulness especially salient. This interpretation is also consistent with recent Japanese evidence showing that mHealth adoption among workers is associated with health-related motivations and daily-life relevance [7]. Among older adults, eHealth literacy also reduced perceived risk, but distrust no longer affected behavioral intention. This suggests that older adults may rely more on literacy-based reassurance and practical value than on distrust-sensitive rejection. Rather than indicating a simple deficit, this pattern implies a functional reorganization of evaluation in later life, consistent with earlier studies emphasizing usability support, family assistance, and the need for age-sensitive design [9,23,26,28,82,91].

Taken together, these findings help explain a persistent Japanese paradox. Japan is characterized by high life expectancy, universal health coverage, extensive digital infrastructure, and rapid population aging, yet mHealth adoption remains limited [3-7]. Our results suggest that this underuse is not explained simply by a lack of infrastructure or a universal failure of perceived usefulness. Rather, it stems from demographic differences in how usefulness, trust, literacy, and risk are interpreted and weighted. In that sense, this study extends prior Japanese work by moving from a descriptive discussion of barriers to a structural account of how those barriers are organized within the acceptance process.

Health-Related Covariates and Contextual Interpretation

An important addition prompted by the revision process was the inclusion of health status and health interest as covariates of performance expectancy and distrust. These analyses strengthened the interpretation of the model in meaningful ways. Health interest positively predicted performance expectancy in the overall model and across sex and age subgroups, with particularly strong effects among older adults. This indicates that users who are more concerned about their health are more likely to perceive mHealth technologies as beneficial. This result is intuitively plausible and aligns with earlier research suggesting that health motivation is a key antecedent of engagement with preventive and self-management applications [1,36,37].

By contrast, self-rated health status showed weaker and less consistent effects. It did not significantly predict performance expectancy in the overall model or across most subgroups, although in some subgroup analyses it was associated with lower distrust. These findings suggest that subjective health concern

XSL•FO
RenderX

may be a more important driver of acceptance-related beliefs than objective or self-perceived health condition per se. This is a useful refinement of the "clinical necessity" argument: it is not simply being less healthy that matters, but being more attentive to one's health and more motivated to manage it. Accordingly, claims about "pragmatic acceptance" among older adults should be interpreted in relation to health interest and literacy rather than assumed solely from age or illness burden.

App-Type Specificity

The app-type-specific analyses further emphasized that mHealth adoption is not monolithic. Disease-management apps were more strongly associated with older adults and with contexts of clinical necessity, consistent with prior telemonitoring and chronic disease management studies [2,37,80,81,92,93]. By contrast, wellness and physical activity apps showed broader cross-demographic appeal and appeared to be driven more by behavioral engagement than by medical need [1,36,37]. This distinction is important because it implies that "performance expectancy" may not mean the same thing across app categories. In disease management, perceived usefulness is likely tied to symptom control, physician recommendation, and safety; in wellness apps, it may reflect motivation, convenience, enjoyment, or self-optimization [1,36,44,56]. Future models of mHealth acceptance should therefore be more explicit about app context rather than assuming that adoption mechanisms are uniform across all digital health tools.

Practical Implications

Several practical implications follow. First, trust-centered strategies are essential. Because distrust and perceived risk operate differently across demographic groups, improving app functionality alone will not be sufficient. Transparent data governance, visible compliance with privacy regulations, third-party security certification, and clear communication about data use are especially important in the Japanese setting, where public sensitivity to personal information protection remains high [12,13,25,40,41,86,87].

Second, interventions should combine universal and targeted strategies. Universal approaches can focus on improving performance expectancy, effort expectancy, and facilitating conditions, because these remain stable predictors of intention across age groups. However, upstream mechanisms require tailoring. For younger users, transparent governance and user control may be more effective than generic promotional messages. For middle-aged users, interventions should emphasize practical value, reliability, and low-risk integration into everyday life. Even among digitally reachable older adults with relatively high eHealth literacy, digital literacy support, simplified interfaces, and family- or community-assisted onboarding remain important for sustaining engagement, while interventions targeting offline or low-connectivity seniors will require additional outreach and support mechanisms [9,23,26,28,82].

Third, implementation in Japan should make explicit use of interpersonal and institutional actors. The structural breadth of social influence observed in this study indicates that peers, family members, workplaces, and health care professionals are not merely dissemination channels but active components of the acceptance process [18,19,27]. This has implications for public health campaigns, workplace wellness programs, and physician-endorsed disease management systems.

Limitations and Future Research

Several limitations should be noted. First, the study relied on an online panel, which likely underrepresents individuals with limited internet access or lower digital engagement, especially among older adults [68,94]. Accordingly, the older group in this study should be interpreted primarily as digitally reachable older adults rather than as fully representative of the Japanese older population as a whole. In our sample, the mean eHealth literacy score was 22.84 (SD 6.31), which is somewhat higher than the 20.46 (SD 5.64) reported in a recent Japanese internet-based study using the eHealth Literacy Scale in a public health context [83]. Because both studies relied on online samples, these values likely overestimate eHealth literacy in the broader Japanese population, particularly among offline or low-connectivity seniors. Relative to the broader Japanese population, nondigital seniors may have lower eHealth literacy, greater usability barriers, and stronger dependence on facilitating conditions, meaning that the present findings may underestimate acceptance barriers in the most vulnerable groups. Future research should therefore use mixed-mode survey designs and community-based recruitment strategies—such as outreach through local clinics, senior centers, and home-visit services—to reach digitally excluded older adults who are absent from online panels and to better capture their acceptance mechanisms and support needs. Although our overall structural model showed acceptable fit according to incremental and parsimony-adjusted indices (CFI=0.953; TLI=0.945; RMSEA=0.047), the GFI (0.889) and AGFI (0.863) fell slightly below the conventional 0.90 threshold specified in the Methods section. Given the well-performing CFI, TLI, and RMSEA values and the known sensitivity of GFI and AGFI to sample size and model complexity, we interpret the structural model as adequately fitting while acknowledging this limitation.

Second, the cross-sectional design precludes causal inference and limits our ability to distinguish age-related differences from cohort or period effects [95]. Third, although the model explained a substantial proportion of variance in behavioral intention, future research should incorporate actual usage behavior and longitudinal follow-up to distinguish initial intention from sustained engagement [1,36,44,56]. Fourth, while this study identified meaningful age- and sex-related heterogeneity, future work should continue to test measurement robustness and explore whether similar structural patterns hold in other East Asian contexts or in more clinically specific populations [32-35,77,78]. Finally, because habit and hedonic motivation were not included, future research focusing on established users should examine whether UTAUT2 constructs become more relevant in later stages of mHealth engagement [50,56].

Conclusion

This study provides robust empirical evidence that mHealth acceptance in Japan's aging society is shaped by a dual structure consisting of stable proximal determinants and heterogeneous

XSL•FO
RenderX

upstream belief-formation processes. While performance expectancy, effort expectancy, and facilitating conditions consistently predict behavioral intention across demographic groups, age and sex differences emerge primarily in how these beliefs are constructed through social influence, eHealth literacy, distrust, and perceived risk.

By extending the UTAUT framework to incorporate distrust, perceived risk, and eHealth literacy, and by integrating health status and health interest as contextual covariates, this study advances the theoretical understanding of digital health acceptance beyond conventional models. In particular, the findings demonstrate that eHealth literacy functions as a context-dependent cognitive resource, self-efficacy exhibits age-contingent effects—including a reversal pattern among younger users—and distrust operates through distinct mechanisms that are not reducible to perceived risk. These results challenge the assumption of uniform acceptance mechanisms and highlight the importance of distinguishing between belief formation and intention formation.

From a policy and practice perspective, the findings indicate that universal promotion strategies for mHealth are insufficient in aging societies. Instead, effective implementation requires trust-centered, demographically tailored, and context-sensitive approaches. For younger users, strategies should address informed skepticism through transparency and user control; for middle-aged users, emphasis should be placed on practical value and risk mitigation; and for older adults, interventions should prioritize literacy support, usability, and social or familial facilitation.

In the Japanese context—characterized by rapid population aging, advanced digital infrastructure, and strong institutional health care systems—these differentiated mechanisms help explain the persistent gap between technological availability and actual mHealth use. Addressing this gap requires not only improving technological functionality but also strengthening trust, enhancing digital health literacy, and leveraging social and institutional support structures.

Overall, this study contributes to the growing literature on digital health by demonstrating that mHealth acceptance is fundamentally shaped by the interaction of psychological, social, and demographic factors. Future research should further examine longitudinal adoption patterns, incorporate behavioral usage data, and explore culturally specific constructs to refine theory and inform more effective digital health interventions in aging societies.

Acknowledgments

The authors thank the participants for their involvement in the study.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Funding

This work was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI (grant number 20K10359) and (grant number 21K01308). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

YF contributed to conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, and writing of the original draft, while KF contributed to formal analysis, investigation, validation, and writing, review, and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist.
[PDF File (Adobe PDF File), 165 KB - mhealth_v14i1e87832_app1.pdf]

Multimedia Appendix 2
Construct items and descriptive statistics (mean and SD).
[DOCX File , 21 KB - mhealth_v14i1e87832_app2.docx]

Multimedia Appendix 3
Descriptive statistics and correlations.
[DOCX File , 19 KB - mhealth_v14i1e87832_app3.docx]

XSL•FO
RenderX

Multimedia Appendix 4
Supplementary Figure 1. Usage rates of various mobile health applications.
[PNG File , 28 KB - mhealth_v14i1e87832_app4.png]

Multimedia Appendix 5
Confirmatory factor analysis results: factor loadings, reliability, and validity indices for constructs and items.
[DOCX File , 22 KB - mhealth_v14i1e87832_app5.docx]

References

1. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
2. Jiang Y, Liu F, Guo J, Sun P, Chen Z, Li J. Evaluating an intervention program using WeChat for patients with chronic obstructive pulmonary disease: randomized controlled trial. J Med Internet Res 2020;22(4):e17089 [FREE Full text] [doi: 10.2196/17089] [Medline: 32314971]
3. Annual report on the ageing society [summary]: FY2024. Cabinet Office, Government of Japan. 2024 Jun. URL: https://www8.cao.go.jp/kourei/english/annualreport/2024/pdf/2024.pdf [accessed 2026-05-16]
4. World report on ageing and health. World Health Organization. 2015 Sep 29. URL: https://www.who.int/publications/i/item/9789241565042 [accessed 2026-06-03]
5. Information and communications in Japan: white paper. Ministry of Internal Affairs and Communications. 2022. URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/eng/WP2022/2022-index.html [accessed 2026-04-22]
6. Information and communications in Japan: white paper. Ministry of Internal Affairs and Communications. 2021. URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/eng/WP2021/2021-index.html [accessed 2026-05-16]
7. Ozaki I, Nishijima M, Shibata E, Zako Y, Chiang C. Factors related to mhealth app use among Japanese workers: cross-sectional survey. JMIR Hum Factors 2024;11:e54673 [FREE Full text] [doi: 10.2196/54673] [Medline: 39454195]
8. Laurino M, Alfì G, Billeci L, Bortone I, Buzzigoli E, Cecchi A. Healthy aging: the INTECMAN project. Aging Clin Exp Res 2021;33(7):2011-2015. [doi: 10.1007/s40520-019-01444-8] [Medline: 31894564]
9. Wang Q, Liu J, Zhou L, Tian J, Chen X, Zhang W. Usability evaluation of mHealth apps for elderly individuals: a scoping review. BMC Med Inform Decis Mak 2022;22(1):317 [FREE Full text] [doi: 10.1186/s12911-022-02064-5] [Medline: 36461017]
10. Apolinário-Hagen J, Paganin G, Simbula S. Editorial: current status of and future directions for assessing technology acceptance for digital (mental) health interventions. Front Digit Health 2024;6:1467297 [FREE Full text] [doi: 10.3389/fdgth.2024.1467297] [Medline: 39228866]
11. Yang M, Al Mamun A, Gao J, Rahman MK, Salameh AA, Alam SS. Predicting m-health acceptance from the perspective of unified theory of acceptance and use of technology. Sci Rep 2024;14(1):339 [FREE Full text] [doi: 10.1038/s41598-023-50436-2] [Medline: 38172184]
12. Alhammad N, Alajlani M, Abd-Alrazaq A, Arvanitis T, Epiphaniou G. Patients and stakeholders' perspectives regarding the privacy, security, and confidentiality of data collected via mobile health apps in Saudi Arabia: protocol for a mixed method study. JMIR Res Protoc 2024 May 22;13:e54933 [FREE Full text] [doi: 10.2196/54933] [Medline: 38776540]
13. Nebeker C, Harlow J, Espinoza J. mHealth privacy and ethical concerns: a scoping review. NPJ Digit Med 2023;6:78. [doi: 10.1038/s41746-023-00850-2]
14. Mcknight DH, Carter M, Thatcher JB, Clay PF. Trust in a specific technology. ACM Trans Manage Inf Syst 2011;2(2):1-25. [doi: 10.1145/1985347.1985353]
15. Söllner M, Hoffmann A, Leimeister JM. Why different trust relationships matter for information systems users. Eur J Inf Sys 2017;25(3):274-287. [doi: 10.1057/ejis.2015.17]
16. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
17. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. J Med Internet Res 2006;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
18. Hofstede G, Hofstede GJ, Minkov M. Cultures and Organizations: Software of the Mind. 3rd Edition. New York, NY: McGraw-Hill; 2010.
19. Xing Y, Ma W, Shi N, Wang Y. Latent variable models for analyzing children's processing ability. Jpn Psychol Res 2014;56(3):243-253. [doi: 10.1111/jpr.12048]
20. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
21. Dwivedi YK, Rana NP, Jeyaraj A, Clement M, Williams MD. Re-examining the unified theory of acceptance and use of technology (UTAUT): towards a revised theoretical model. Inf Syst Front 2017;21(3):719-734. [doi: 10.1007/s10796-017-9774-y]

XSL•FO
RenderX

22. Marangunić N, Granić A. Technology acceptance model: a literature review from 1986 to 2013. Univ Access Inf Soc 2014;14(1):81-95. [doi: 10.1007/s10209-014-0348-1]
23. Hoque R, Sorwar G. Understanding factors influencing the adoption of mHealth by the elderly: an extension of the UTAUT model. Int J Med Inform 2017;101:75-84. [doi: 10.1016/j.ijmedinf.2017.02.002] [Medline: 28347450]
24. Dou K, Yu P, Deng N, Liu F, Guan Y, Li Z. Patients' acceptance of smartphone health technology for chronic disease management: a theoretical model and empirical test. JMIR Mhealth Uhealth 2017;5(12):e177 [FREE Full text] [doi: 10.2196/mhealth.7886] [Medline: 29212629]
25. Anderson CL, Agarwal R. The digitization of healthcare: Boundary risks, emotion, and consumer willingness to disclose personal health information. Inf Sys Res 2011;22(3):469-490. [doi: 10.1287/isre.1100.0335]
26. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: a literature based framework (MOLD-US). Int J Med Inform 2018;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
27. Cao J, Kurata K, Lim Y, Sengoku S, Kodama K. Social acceptance of mobile health among young adults in Japan: an extension of the UTAUT model. Int J Environ Res Public Health 2022;19(22):15156 [FREE Full text] [doi: 10.3390/ijerph192215156] [Medline: 36429875]
28. Cimperman M, Makovec Brenčič M, Trkman P. Analyzing older users' home telehealth services acceptance behavior-applying an extended UTAUT model. Int J Med Inform 2016;90:22-31. [doi: 10.1016/j.ijmedinf.2016.03.002] [Medline: 27103194]
29. Alam MZ, Hoque MR, Hu W, Barua Z. Factors influencing the adoption of mHealth services in a developing country: a patient-centric study. Int J Inf Manag 2020;50:128-143. [doi: 10.1016/j.ijinfomgt.2019.04.016]
30. McKnight DH, Choudhury V, Kacmar C. Developing and validating trust measures for e-commerce: an integrative typology. Inf Sys Res 2002;13(3):334-359. [doi: 10.1287/isre.13.3.334.81]
31. Venkatesh V, Morris MG. Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. MIS Q 2000;24(1):115-139. [doi: 10.2307/3250981]
32. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equ Modeling 2002;9(2):233-255. [doi: 10.1207/S15328007SEM0902_5]
33. Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. Organ Res Methods 2000;3(1):4-70. [doi: 10.1177/109442810031002]
34. Meade AW, Johnson EC, Braddy PW. Power and sensitivity of alternative fit indices in tests of measurement invariance. J Appl Psychol 2008;93(3):568-592. [doi: 10.1037/0021-9010.93.3.568] [Medline: 18457487]
35. Little TD, Card NA, Slegers DW, Ledford EC. Representing contextual effects in multiple-group MACS models. In: Modeling Contextual Effects in Longitudinal Studies. Mahwah: Lawrence Erlbaum Associates; 2007:121-147.
36. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
37. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: an overview of systematic reviews. PLoS One 2017;12(3):e0173160 [FREE Full text] [doi: 10.1371/journal.pone.0173160] [Medline: 28249025]
38. Dinev T, Hart P. An extended privacy calculus model for E-commerce transactions. Inf Syst Res 2006;17(1):61-80. [doi: 10.1287/isre.1060.0080]
39. Thatcher JB, McKnight DH, Baker EW, Arsal RE, Roberts NH. The role of trust in postadoption IT exploration: an empirical examination of knowledge management systems. IEEE Trans. Eng. Manage 2011;58(1):56-70. [doi: 10.1109/tem.2009.2028320]
40. Lucivero F, Hallowell N, Johnson S, Prainsack B, Samuel G, Sharon T. The COVID-19 pandemicthe data protection regime: legitimacy, trustunderstanding in the UK. Big Data Soc 2022;9(1):1-14. [doi: 10.1177/20539517221092714]
41. Act on the Protection of Personal Information (APPI). Ministry of Justice: Japanese Law Translation.: Personal Information Protection Commission URL: https://www.japaneselawtranslation.go.jp/en/laws/view/4241/en [accessed 2026-04-22]
42. Lankton NK, McKnight DH, Tripp J. Technology, humanness, and trust: rethinking trust in technology. JAIS 2015;16(10):880-918. [doi: 10.17705/1jais.00411]
43. Vance A, Elie-Dit-Cosaque C, Straub DW. Examining trust in information technology artifacts: the effects of system quality and culture. J Manag Inf Syst 2014;24(4):73-100. [doi: 10.2753/mis0742-1222240403]
44. Cho J. The impact of post-adoption beliefs on the continued use of health apps. Int J Med Inform 2016;87:75-83. [doi: 10.1016/j.ijmedinf.2015.12.016] [Medline: 26806714]
45. Hua Z, Yuqing S, Qianwen L, Hong C. Factors influencing ehealth literacy worldwide: systematic review and meta-analysis. J Med Internet Res 2025;27:e50313 [FREE Full text] [doi: 10.2196/50313] [Medline: 40063939]
46. Paige SR, Krieger JL, Stellefson M, Alber JM. eHealth literacy in chronic disease patients: an item response theory analysis of the eHealth literacy scale (eHEALS). Patient Educ Couns 2017;100(2):320-326 [FREE Full text] [doi: 10.1016/j.pec.2016.09.008] [Medline: 27658660]
47. Lupton D. The digitally engaged patient: self-monitoring and self-care in the digital health era. Soc Theory Health 2013;11(3):256-270. [doi: 10.1057/sth.2013.10]
48. Hsu SC. Revisiting eHealth literacy in the AI health era: benefits and risks. J Med Internet Res 2023;25:e46629. [doi: 10.2196/46629]

XSL•FO
RenderX

49. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
50. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157-178. [doi: 10.2307/41410412]
51. Huang W, Ong WC, Wong MKF, Ng EYK, Koh T, Chandramouli C. Applying the UTAUT2 framework to patients' attitudes toward healthcare task shifting with artificial intelligence. BMC Health Serv Res 2024;24(1):455 [FREE Full text] [doi: 10.1186/s12913-024-10861-z] [Medline: 38605373]
52. Philippi P, Baumeister H, Apolinário-Hagen J, Ebert DD, Hennemann S, Kott L. Acceptance towards digital health interventions - model validation and further development of the unified theory of acceptance and use of technolog. Internet Interventions 2021;26:100459. [doi: 10.1016/j.invent.2021.100459]
53. Batool S, Darzi MA, Ahmed T. Adoption of mobile health applications based on UTAUT2: a study Of North-Indian states. IJFMR 2023;5(5):99-110. [doi: 10.36948/ijfmr.2023.v05i05.7370]
54. Schomakers E, Lidynia C, Vervier LS, Calero Valdez A, Ziefle M. Applying an extended UTAUT2 model to explain user acceptance of lifestyle and therapy mobile health apps: survey study. JMIR Mhealth Uhealth 2022;10(1):e27095 [FREE Full text] [doi: 10.2196/27095] [Medline: 35040801]
55. Calegari LP, Tortorella GL, Fettermann DC. Getting connected to M-Health technologies through a meta-analysis. Int J Environ Res Public Health 2023;20(5):4369 [FREE Full text] [doi: 10.3390/ijerph20054369] [Medline: 36901379]
56. Limayem M, Hirt SG, Cheung CMK. How habit limits the predictive power of intention: the case of information systems continuance. MIS Q 2007;31(4):705-737. [doi: 10.2307/25148817]
57. Kim HW, Chan HC, Gupta S. Value-based adoption of mobile internet: an empirical investigation. Decis Support Syst 2007;43(1):111-126. [doi: 10.1016/j.dss.2005.05.009]
58. Komiak SYX, Benbasat I. The effects of personalization and familiarity on trust and adoption of recommendation agents. MIS Q 2006;30(4):941-960. [doi: 10.2307/25148760]
59. Kim DJ, Ferrin DL, Rao HR. A trust-based consumer decision-making model in electronic commerce: the role of trust, perceived risk, and their antecedents. Decis Support Syst 2008;44(2):544-564. [doi: 10.1016/j.dss.2007.07.001]
60. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
61. Compeau DR, Higgins CA. Computer self-efficacy: development of a measure and initial test. MIS Q 1995;19(2):189-211 [FREE Full text] [doi: 10.2307/249688]
62. Pavlou PA. Consumer acceptance of electronic commerce: integrating trust and risk with the technology acceptance model. Int J Electron Commer 2014;7(3):101-134. [doi: 10.1080/10864415.2003.11044275]
63. Featherman MS, Pavlou PA. Predicting e-services adoption: a perceived risk facets perspective. Int J Hum Comput Stud 2003;59(4):451-474. [doi: 10.1016/S1071-5819(03)00111-3]
64. Mitsutake S, Shibata A, Ishii K, Oka K. Association of eHealth literacy with colorectal cancer knowledge and screening practice among internet users in Japan. J Med Internet Res 2012;14(6):e153 [FREE Full text] [doi: 10.2196/jmir.1927] [Medline: 23149453]
65. Lowry PB, Vance A, Moody G, Beckman B, Read A. Explaining and predicting the impact of branding alliances and web site quality on initial consumer trust of e-commerce web sites. J Manag Inf Syst 2008;24(4):199-224. [doi: 10.2753/mis0742-1222240408]
66. Gefen D, Straub DW. Managing user trust in B2C e-services. e-Service J 2003;2(2):7-24. [doi: 10.1353/esj.2003.0011]
67. Chiu C, Wang ET. Understanding web-based learning continuance intention: the role of subjective task value. Inf Manag 2008;45(3):194-201. [doi: 10.1016/j.im.2008.02.003]
68. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet E-surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
69. Wolf EJ, Harrington KM, Clark SL, Miller MW. Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. Educ Psychol Meas 2013;76(6):913-934 [FREE Full text] [doi: 10.1177/0013164413495237] [Medline: 25705052]
70. Kline RB. Principles and Practice of Structural Equation Modeling. New York: Guilford Press; 2016.
71. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis. 8th Edition. Andover: Cengage Learning; 2018.
72. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 1981;18(1):39-50. [doi: 10.1177/002224378101800104]
73. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 1999;6(1):1-55. [doi: 10.1080/10705519909540118]
74. Byrne BM. Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming. New York: Routledge; 2016.
75. Bagozzi RP, Yi Y. On the evaluation of structural equation models. J Acad Mark Sci 1988;16(1):74-94. [doi: 10.1007/bf02723327]
76. Bentler PM. Comparative fit indexes in structural models. Psychol Bull 1990;107(2):238-246 [FREE Full text] [doi: 10.1037/0033-2909.107.2.238] [Medline: 2320703]

XSL•FO
RenderX

77. Dimitrov DM. Testing for factorial invariance in the context of construct validation. Meas Eval Couns Dev 2017;43(2):121-149. [doi: 10.1177/0748175610373459]
78. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. Struct Equ Modeling 2007;14(3):464-504. [doi: 10.1080/10705510701301834]
79. Zhang X, Guo X, Lai KH, Guo F, Li C. Understanding gender differences in m-health adoption: a modified theory of reasoned action model. Telemed J E Health 2014;20(1):39-46. [doi: 10.1089/tmj.2013.0092] [Medline: 24161004]
80. Tao D, LeRouge C, Smith KJ, De Leo G. Defining information quality into health websites: a conceptual framework of health website information quality for educated young adults. JMIR Hum Factors 2017;4(4):e25 [FREE Full text] [doi: 10.2196/humanfactors.6455] [Medline: 28986336]
81. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S. eHealth literacy and web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
82. Guo X, Sun Y, Wang N, Peng Z, Yan Z. The dark side of elderly acceptance of preventive mobile health services in China. Electron Mark 2012;23(1):49-61. [doi: 10.1007/s12525-012-0112-4]
83. Mitsutake S, Shibata A, Ishii K, Okazaki K, Oka K. Developing Japanese version of the eHealth literacy scale (eHEALS). Nihon Koshu Eisei Zasshi 2011;58(5):361-371. [Medline: 21905612]
84. Awad NF, Krishnan MS. The personalization privacy paradox: an empirical evaluation of information transparency and the willingness to be profiled online for personalization. MIS Q 2006;30(1):13-28. [doi: 10.2307/25148715]
85. Bansal G, Zahedi FM, Gefen D. Do context and personality matter? Trust and privacy concerns in disclosing private information online. Inf Manag 2016;53(1):1-21. [doi: 10.1016/j.im.2015.08.001]
86. Becker MW. Tracking emerging privacy and security risks in mHealth: Privacy, security, and safety of digital information in health. Yearb Med Inform 2016(1):62-66. [doi: 10.15265/IY-2016-037]
87. Hong W, Thong JYL. Internet privacy concerns: an integrated conceptualization and four empirical studies. MIS Q 2013;37(1):275-298. [doi: 10.25300/misq/2013/37.1.12]
88. Gefen D, Karahanna E, Straub DW. Trust and TAM in online shopping: an integrated model. MIS Q 2003;27(1):51-90. [doi: 10.2307/30036519]
89. Yamamoto T, Chiba N, Magata F, Takahashi K, Ueda H, Sekiya N. Joint research on anshin in internet usage. NTT Technical Review 2011;9(10):50-57. [doi: 10.53829/ntr201110fa8]
90. Garbarino E, Strahilevitz M. Gender differences in the perceived risk of buying online and the effects of receiving a site recommendation. J Bus Res 2004;57(7):768-775. [doi: 10.1016/s0148-2963(02)00363-6]
91. Steinhubl SR, Waalen J, Edwards AM, Ariniello LM, Mehta RR, Ebner GS. Effect of a home-based wearable continuous ECG monitoring patch on detection of undiagnosed atrial fibrillation: the mSToPS randomized clinical trial. JAMA 2018;320(2):146-155 [FREE Full text] [doi: 10.1001/jama.2018.8102] [Medline: 29998336]
92. Omboni S, Gazzola T, Carabelli G, Parati G. Clinical usefulness and cost effectiveness of home blood pressure telemonitoring: meta-analysis of randomized controlled studies. J Hypertens 2013;31(3):455-67; discussion 467. [doi: 10.1097/HJH.0b013e32835ca8dd] [Medline: 23299557]
93. Sun Y, Wang N, Guo X, Peng Z. Understanding the acceptance of mobile health services: a comparison and integration of alternative models. J Electron Commer Res 2013;14(2):183-200 [FREE Full text]
94. Bethlehem J. Selection bias in web surveys. Int Stat Rev 2010;78(2):161-188. [doi: 10.1111/j.1751-5823.2010.00112.x]
95. Schaie KW. A general model for the study of developmental problems. Psychol Bull 1965;64:92-107. [doi: 10.1037/h0022371] [Medline: 14320080]

Abbreviations

AGFI: adjusted GFI
AVE: average variance extracted
CFA: confirmatory factor analysis
CFI: comparative fit index
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
CR: composite reliability
GFI: goodness-of-fit index
IFI: incremental fit index
mHealth: mobile health
OR: odds ratio
RMSEA: root mean square error of approximation
SEM: structural equation modeling
TLI: Tucker-Lewis index
UTAUT: unified theory of acceptance and use of technology

XSL•FO
RenderX

Edited by R Chandrasekaran; submitted 15.Nov.2025; peer-reviewed by Y Chibaa, M Chatrangsan; comments to author 22.Feb.2026; revised version received 23.Apr.2026; accepted 02.May.2026; published 09.Jun.2026.

Please cite as:
Fukuda Y, Fukuda K
Mobile Health App Acceptance in Japan's Aging Society: Multigroup Structural Equation Modeling Based on the Extended Unified Theory of Acceptance and Use of Technology and eHealth Literacy Frameworks
JMIR Mhealth Uhealth 2026;14:e87832
URL: https://mhealth.jmir.org/2026/1/e87832
doi:10.2196/87832
PMID:

©Yasue Fukuda, Koji Fukuda. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 09.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Chinese Version of the mHealth App Usability Questionnaire for Health Care Providers: Cross-Sectional Translation and Validation Study

Mengxia Chen[1*], PhD; Wenjun Gao[2*], PhD; Yijie Pan[3*], MSCN; Mengdi Wang[4], MSCN; Bing Yu[3], MSCN; Dongmei Li[5], MSCN; Huanhuan Hu[5], BSN; Jing Zheng[5], BSN; Lingjuan Zhang[3*], PhD; Yuan Gao[1*], PhD

[1]Department of Nursing, the First Medical Center of Chinese People's Liberation Army General Hospital, Beijing, China

[2]Nursing College, Naval Medical University, Shanghai, China

[3]Education and Scientific Research Department of Clinical Nursing, Changhai Hospital, Shanghai, China

[4]Department of Emergency and Critical Care Medicine, The 961th Hospital of Joint Logistics Support Force, Qiqihaer, China

[5]Neurovascular Center, Changhai Hospital, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Yuan Gao, PhD
Department of Nursing, the First Medical Center of Chinese People's Liberation Army General Hospital
No.28, Fuxing Road
Beijing, 100853
China
Phone: 86 66875500
Email: gaoyuanzd@163.com

Abstract

Background: With the rapid proliferation of mobile health (mHealth) apps in clinical practice, ensuring their usability for health care providers is critical to optimize workflow integration and service delivery. The mHealth App Usability Questionnaire (MAUQ) is a validated tool for assessing usability, but its applicability in Chinese health care settings remains untested. Cross-cultural adaptation and validation of the MAUQ for interactive apps are essential to address this gap.

Objective: This study aimed to translate the MAUQ (for interactive apps for health care providers) into Chinese (MAUQ-C), adapt it to the Chinese health care context, and validate its psychometric properties for future mHealth research and use in China.

Methods: This cross-sectional study followed international guidelines for cross-cultural adaptation, including forward translation, backward translation, expert review, and pretesting. A convenience sample of 382 health care providers (eg, physicians and nurses) using mHealth apps was recruited from 2 hospitals in China. Reliability was evaluated using the Cronbach α, test-retest reliability, and split-half reliability. Validity was assessed through content validity index, exploratory factor analysis, and confirmatory factor analysis.

Results: The MAUQ-C retained the original 21-item structure and 3 dimensions. Content validity index values ranged from 0.833 to 1.000 (Scale-Content Validity Index/Average=0.992). Exploratory factor analysis identified a 3-factor model explaining 63.48% of the total variance. Confirmatory factor analysis confirmed a good model fit (minimum discrepancy/df=1.810; cumulative fit index=0.950; root-mean-square error of approximation=0.058). The Cronbach α for the total scale was 0.919, with dimension-specific values of 0.928, 0.877, and 0.886. Test-retest reliability was 0.881, and split-half reliability was 0.726. These results confirm that the MAUQ-C effectively measures usability across the dimensions of usability and satisfaction, system information arrangement, and efficiency.

Conclusions: The MAUQ-C is a reliable and valid instrument for evaluating mHealth app usability among Chinese health care providers. It can serve as a critical tool to guide the development and optimization of mHealth solutions tailored to the Chinese health care system.

(JMIR Mhealth Uhealth 2026;14:e86783) doi:10.2196/86783

KEYWORDS

mobile health app; mHealth app; questionnaire translation; usability; health care provider; Chinese

XSL•FO
RenderX

Introduction

The substantial user base for internet medical services of 418 million by 2024 in China provides a critical foundation for mobile health (mHealth) [1]. The World Health Organization defines mHealth as the "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices" [2]. Driven by advancing public health literacy and internet technology, this sector has experienced explosive growth and become extensively integrated into personal health management ecosystems [3,4], where mHealth apps typically offer multifaceted functionalities, including monitoring and recording health data, disease management [5], and medical guidance [6], among others [7].

However, nearly half a million mHealth apps developed remain underused [8], largely due to the absence of effective usability evaluations. This gap contributes to skepticism among clinicians, scholars, and patients regarding the reliability of mHealth programs [9]. High-usability mHealth apps typically exhibit 5 core attributes: efficiency, favorable user perception, ease of learning, memorability, and low error rates [10]. While usability questionnaires remain the predominant method for evaluating usability, the most commonly used instruments—such as the System Usability Scale (SUS) and Post-Study System Usability Questionnaire (PSSUQ) [11]—are primarily designed for expert-based objective evaluation, not specifically to assess mHealth app usability. In addition, mHealth apps offer a unique and readily accessible platform not only for the patient but also for the health care provider. To address the crucial need for evaluating the utility and usability from both patient and health care provider perspectives, the mHealth App Usability Questionnaire (MAUQ) was crafted by Zhou et al [12] in 4 different versions categorized based on the nature of the app (interactive or stand-alone) and the intended user (patient or health care provider). The MAUQ's unique design—differentiating between patient and health care provider versions and focusing on interactive app functionalities—makes it particularly suitable for evaluating how mHealth tools integrate into the complex clinical workflows of health care professionals. By focusing on dimensions such as system information arrangement and efficiency, the MAUQ provides actionable insights into clinical usability that broader quality assessment tools may not fully capture.

The necessity of a health care provider–specific version of the MAUQ is grounded in the fundamental differences between the use contexts of patients and health care professionals. While patients primarily focus on ease of use and personal health management, health care providers operate within complex clinical environments characterized by high cognitive load, time pressure, and the critical responsibility for clinical decision-making [3]. The patient version of the MAUQ, which emphasizes personal health outcomes and self-care, fails to capture these professional-specific dimensions such as the integration of mHealth tools into existing clinical workflows and the impact of app usability on patient safety [8]. Therefore, a validated health care provider–specific version is essential to evaluate whether an app facilitates clinical practice. Relying on the patient version of the MAUQ for professional users would lead to a misalignment between the tool's assessment criteria and the actual professional requirements of health care providers.

The different versions of the MAUQ have been translated into different languages and validated, such as into Chinese (interactive apps, patient version) [8], Spanish (stand-alone apps, patient version) [13], Persian (interactive apps, patient version) [14], Malay (stand-alone apps, patient version) [15], and German (stand-alone apps, patient version) [16]. However, China has not yet introduced a Chinese version of the MAUQ (MAUQ-C; interactive apps for health care providers). Therefore, the present study aimed to translate and verify the health care provider version of the MAUQ for interactive apps into Chinese.

Methods

Study Design

This was a cross-sectional study using the MAUQ. The original English version was developed to assess the usability of mHealth apps among patients and health care providers with a strong internal consistency [12]. The MAUQ for interactive apps for health care providers was used in this study. The MAUQ contains a total of 21 items and 3 dimensions: usability and satisfaction (8 items), system information arrangement (6 items), and efficiency (7 items) scored on a 7-point Likert scale ranging from 1 ("extremely strongly agree") to 7 ("extremely strongly disagree"). The usability of the app is the average score of all items (total score divided by the number of items). The closer the average score is to 1, the higher the app's usability is. A sample size of 382 participants was determined to ensure a robust participant-to-item ratio, which exceeds the minimum requirement of 10:1 for exploratory factor analysis (EFA), thereby ensuring the stability of the factor structure.

Inclusion and Exclusion Criteria

The inclusion criteria were (1) health care providers (physicians or nurses) currently working in the participating hospitals, (2) individuals who had used an mHealth app for at least 1 month to provide medical services, and (3) those who voluntarily consented to participate. Exclusion criteria were (1) individuals with less than 1 year of clinical experience and (2) those who were unable to complete the online questionnaire for technical or personal reasons.

Translation and Adaptation Process

Translation

Permission from the author of the original questionnaire was obtained through email. The translation model by Brislin [17] was adopted. The first step is forward translation, where 2 translators with native English-language proficiency independently translated the original MAUQ into Chinese versions (C1 and C2). The research team then reconciled these versions through discussion with both translators to produce a consensus Chinese draft (C3). The second step is back translation, where 2 new translators independently translated C3 back into English (E1 and E2). The research team harmonized E1 and E2 into a single English version (E3). The

XSL•FO
RenderX

third step is cross-validation, where the research team and all translators compared E3 with the original MAUQ to evaluate semantic equivalence. Discrepancies were resolved by refining C3. The revised MAUQ-C was finalized after incorporating all necessary amendments.

Experts' Comments

The expert panel consisted of 6 professionals (4 clinical health care providers, 1 software engineer, and 1 English teacher), all of whom held a master's degree or above and possessed intermediate or senior professional titles. Their expertise spanned clinical nursing, software engineering, and medical education, ensuring a comprehensive evaluation of the questionnaire's content validity. On the basis of their theoretical knowledge and practical experience, each expert evaluated the accuracy of translation, content comprehension, language expression habits, and consistency with the cultural background of each item one by one. For suggestions that were not easy to understand, the researcher had an in-depth discussion with the experts and made detailed records. Finally, the group summarized all the opinions from experts. The expert panel was selected to ensure a balance between clinical practice and technical implementation. Although the panel lacked dedicated psychometricians, the experts' deep familiarity with mHealth research and clinical nursing ensured that the items were culturally adapted and clinically relevant.

Prediction Test

A preliminary survey was conducted among 10 health care providers at a tertiary hospital using convenience sampling. The survey aimed to assess respondents' understanding of the questionnaire and their perceptions of its items. Feedback and suggestions from participants were documented, and the final MAUQ-C for interactive apps was developed by integrating experts' recommendations and respondents' suggestions.

Research Objects and Tools

This cross-sectional descriptive study was conducted in 1 third-class hospital. The convenience sampling method was adopted to select health care providers from a hospital who used an mHealth app (UHealth) to provide medical services. All participants gave informed consent and voluntarily took part in the study. UHealth is an app designed for the population with chronic diseases, providing comprehensive management such as medication reminders, online consultation, and disease knowledge [18]. The inclusion criteria were physicians or nurses who had ever used the UHealth app. The research tools included the MAUQ-C and health care workers' general information, including their age, sex, marital status, educational level, and professional title.

Verification of the Questionnaire

Item Analysis

The central tendency of the responses to each item was first examined. If a single response option accounted for over 80% of the total responses for an item, the item was considered to have weak discriminative power and was thus eliminated. Next, independent-sample 2-tailed *t* tests were conducted to compare mean scores between high- and low-score groups (top 27% vs bottom 27% of the total scores), with items showing no significant between-group differences ($P>.05$) being removed. Finally, Pearson correlation analysis was performed between individual item scores and the total questionnaire score, and items with poor correlation ($r<0.30$; $P>.05$) were excluded to ensure construct validity.

Reliability and Validity Testing

Six experts in mHealth were invited to evaluate the relevance of the questionnaire contents. A 4-point Likert scale was used to evaluate the correlation between each item (1="not related"; 4="strongly related"). The content validity index (CVI) was calculated. Principal component analysis and maximum variance orthogonal rotation were used for EFA. Items with a load factor of less than 0.40 were deleted [19]. The correlation matrix between each dimension and the total score of the questionnaire was tested, and the internal correlation analysis was carried out. Confirmatory factor analysis (CFA) was undertaken to verify whether the structure of the questionnaire was consistent with the theoretical structure of the original questionnaire. The reliability of the questionnaire was evaluated using the Cronbach α coefficient, test-retest reliability, and split-half reliability.

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki and was approved by the ethics committee of Changhai Hospital (approval CHE2023-061). All participants provided informed consent before taking part in the study. On the first page of the online survey, we provided a detailed introduction to the study and an informed consent form. Participants were required to click a box stating, "I agree to participate in this research of my own volition" to proceed. The Wenjuanxing platform automatically logged their consent, and participants were free to opt out at any time by closing the survey. To ensure privacy and confidentiality, all collected data were deidentified and stored in a secure, password-protected database accessible only to the research team. No personally identifiable information was collected or disclosed, and no compensation was provided.

Results

General Information

A total of 382 questionnaires were forwarded to health care workers in the hospital who used an mHealth app to provide medical care services for patients. Of these 382 health care workers, 248 (64.9%) were male, and 134 (35.1%) were female. Most users (n=363, 95%) were under 50 years of age. Table 1 provides further details.

XSL•FO
RenderX

Table 1. General characteristics of the study participants (N=382).

Category and classification	Participants, n (%)
Sex	
Male	248 (64.9)
Female	134 (35.1)
Age (y)	
<30	92 (24.1)
30-40	184 (48.2)
40-50	87 (22.8)
≥50	19 (5)
Length of service (y)	
≤5	114 (29.8)
5-10	145 (38)
10-15	68 (17.8)
≥15	55 (14.4)
Professional rank and title	
Junior	53 (13.9)
Intermediate	208 (54.5)
Senior	121 (31.7)

Item Analysis

The results (Table 2) indicated a significant difference between the high- and low-score groups (t_{211}=−14.497; *P*<.001). Items with no statistically significant difference (critical ratio<3 or *P*<.05) between the high- and low-score groups would be deleted [20]. However, the results showed that there were significant differences (critical ratio=8.594-16.613; *P*<.001) in each item between the high- and low-score groups; thus, no item was deleted.

The Pearson correlation coefficient was used to assess the correlation among 21 items. Items with a correlation coefficient below 0.4 or a *P* value below .05 would be deleted. As shown in Table 3, all the items were significant, and a positive correlation was found among the values of the assessment phase. The correlation coefficient between the score of each item and the total score of the questionnaire was calculated (0.505-0.743; *P*<.001); thus, all items were retained.

XSL•FO
RenderX

Table 2. Results of the item analysis.

Item	Expert evaluation (score of 1-4), mean (SD)		*t* test (*df*)	*P* value
	Low-score group (n=107)	High-score group (n=106)		
1	1.41 (0.49)	3.44 (1.42)	−14.497 (211)	<.001
2	1.53 (0.64)	3.54 (1.38)	−14.120 (211)	<.001
3	1.48 (0.54)	3.59 (1.36)	−15.382 (211)	<.001
4	1.39 (0.66)	3.53 (1.43)	−14.486 (211)	<.001
5	1.63 (0.62)	3.26 (1.49)	−10.803 (211)	<.001
6	1.25 (0.46)	3.59 (1.44)	−16.613 (211)	<.001
7	1.55 (0.62)	3.45 (1.35)	−13.593 (211)	<.001
8	1.45 (0.55)	3.30 (1.28)	−14.198 (211)	<.001
9	1.50 (0.64)	3.45 (1.46)	−12.997 (211)	<.001
10	1.65 (0.65)	3.13 (1.28)	−11.029 (211)	<.001
11	1.62 (0.58)	3.38 (1.22)	−13.905 (211)	<.001
12	1.65 (0.71)	3.18 (1.37)	−10.609 (211)	<.001
13	1.55 (0.59)	3.32 (1.32)	−13.086 (211)	<.001
14	1.51 (0.62)	3.13 (1.41)	−11.224 (211)	<.001
15	1.69 (0.85)	3.06 (1.27)	−9.561 (211)	<.001
16	1.80 (0.75)	3.04 (1.27)	−8.963 (211)	<.001
17	1.72 (0.69)	2.89 (1.26)	−8.618 (211)	<.001
18	1.60 (0.64)	2.96 (1.14)	−11.039 (211)	<.001
19	1.54 (0.64)	2.79 (1.31)	−9.162 (211)	<.001
20	1.66 (0.69)	2.94 (1.21)	−9.784 (211)	<.001
21	1.70 (0.81)	2.88 (1.20)	−8.594 (211)	<.001

XSL•FO
RenderX

Table 3. Pearson correlation analysis results.

Item	Correlation coefficient	*P* value
1	0.727	<.001
2	0.656	<.001
3	0.743	<.001
4	0.685	<.001
5	0.589	<.001
6	0.696	<.001
7	0.641	<.001
8	0.697	<.001
9	0.647	<.001
10	0.543	<.001
11	0.622	<.001
12	0.505	<.001
13	0.656	<.001
14	0.510	<.001
15	0.582	<.001
16	0.594	<.001
17	0.527	<.001
18	0.606	<.001
19	0.602	<.001
20	0.594	<.001
21	0.538	<.001

Validity Testing of the Questionnaire

As shown in Table 4, the score range of the CVI in the questionnaire was 0.833 to 1 (minimum value≥0.78) [20], indicating that the items were comprehensible, and the CVI of the total score of the questionnaire was 0.992.

XSL•FO
RenderX

Table 4. Content validity evaluation results of the Chinese health care provider version of the mHealth App Usability Questionnaire.

Item	Score by experts						Experts who provided a score of 3 or more, n	I-CVI[a]
	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6		
1	4	4	4	4	4	4	6	1.000
2	4	3	4	4	4	4	6	1.000
3	2	3	3	4	4	4	5	0.833
4	4	4	4	4	3	4	6	1.000
5	4	4	3	3	3	4	6	1.000
6	4	4	3	4	4	4	6	1.000
7	4	4	4	4	4	4	6	1.000
8	4	4	4	4	4	4	6	1.000
9	4	4	4	3	4	3	6	1.000
10	4	4	4	4	3	3	6	1.000
11	4	4	4	3	4	3	6	1.000
12	4	3	4	4	4	3	6	1.000
13	4	4	4	4	4	4	6	1.000
14	4	4	3	4	4	4	6	1.000
15	4	4	4	3	3	4	6	1.000
16	4	4	4	4	4	4	6	1.000
17	4	4	4	4	4	4	6	1.000
18	4	4	4	4	4	4	6	1.000
19	4	4	4	4	4	4	6	1.000
20	3	4	4	4	4	4	6	1.000
21	4	4	4	3	4	4	6	1.000

[a]I-CVI: item-level content validity index.

Structural Validity

Exploratory Factor Analysis

The validity of the questionnaire was assessed through EFA of the values of the Kaiser-Meyer-Olkin (KMO) test, the Bartlett test of sphericity, common degree (common factor variance), load factor, and other indicators. The KMO test was conducted before factor analysis to evaluate whether the data were suitable for EFA. The KMO value of this study was 0.932 (>0.900), which indicated that the questionnaire had an acceptable validity. At the same time, the Bartlett test of sphericity was carried out to determine whether the correlation coefficient between items was significant. The value of the Bartlett test was $P<.001$, indicating that the questionnaire data were very suitable for EFA. According to the maximum variance method, the factors with eigenvalues greater than 1 were extracted from the 21 items, resulting in the extraction of 3 common factors with eigenvalues greater than 1. The cumulative contribution rate of these 3 common factors was 63.48% (>60%), indicating that the scale structure was more consistent with the theoretical structure. Detailed results are shown in Table 5.

The orthogonal rotation method was used to analyze the crushed stone diagram (Figure 1). The dimensions of each item were the same as those in the original version and, ultimately, resulted in an MAUQ-C for health care providers with 21 items including 3 dimensions: usability and satisfaction (8 items), system information arrangement (6 items), and usefulness (7 items).

The results of EFA are shown in Table 6.

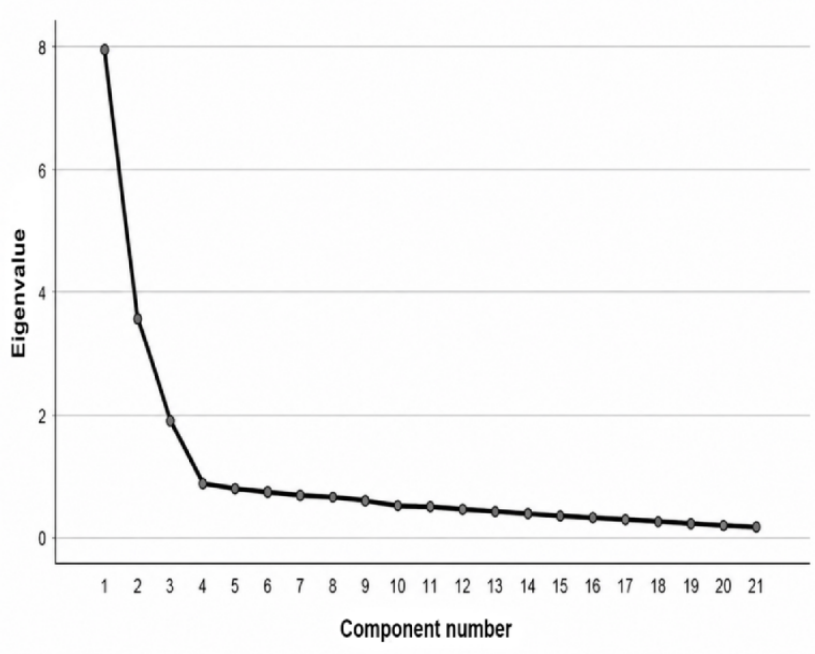
XSL•FO
RenderX

Table 5. Results for the total variance interpretation.

Item	Initial eigenvalues			Squares of the extracted loadings			Squares of the transverse loadings		
	Total	Variance (%)	Cumulation (%)	Total	Variance (%)	Cumulation (%)	Total	Variance (%)	Cumulation (%)
1	7.907	37.653	37.653	7.907	37.653	37.653	5.496	26.170	26.170
2	3.554	16.924	54.577	3.554	16.924	54.577	4.129	19.660	45.830
3	1.870	8.903	63.48	1.870	8.903	63.48	3.707	17.650	63.48
4	0.797	3.796	67.276	__[a]	—	—	—	—	—
5	0.721	3.432	70.708	—	—	—	—	—	—
6	0.674	3.208	73.916	—	—	—	—	—	—
7	0.618	2.945	76.861	—	—	—	—	—	—
8	0.599	2.854	79.715	—	—	—	—	—	—
9	0.556	2.648	82.363	—	—	—	—	—	—
10	0.464	2.211	84.574	—	—	—	—	—	—
11	0.453	2.157	86.731	—	—	—	—	—	—
12	0.423	2.016	88.747	—	—	—	—	—	—
13	0.399	1.898	90.644	—	—	—	—	—	—
14	0.346	1.647	92.291	—	—	—	—	—	—
15	0.306	1.456	93.747	—	—	—	—	—	—
16	0.281	1.338	95.086	—	—	—	—	—	—
17	0.262	1.247	96.332	—	—	—	—	—	—
18	0.236	1.125	97.457	—	—	—	—	—	—
19	0.211	1.003	98.461	—	—	—	—	—	—
20	0.179	0.851	99.312	—	—	—	—	—	—
21	0.145	0.688	100.00	—	—	—	—	—	—

[a]Not applicable.

Figure 1. Fragmentation graph of the scale factors of the Chinese health care provider version of the mHealth App Usability Questionnaire.



XSL•FO
RenderX

Table 6. Loading matrix of the scale factors of the Chinese health care provider version of the mHealth App Usability Questionnaire.

Item	Factor loadings		
	Factor 1	Factor 2	Factor 3
1	*0.862* [a]	0.080	0.194
2	*0.842*	0.128	0.111
3	*0.828*	0.128	0.239
4	*0.776*	0.082	0.171
5	*0.787*	0.102	0.002
6	*0.853*	0.108	0.119
7	*0.827*	0.083	0.096
8	*0.673*	0.039	0.429
9	0.195	0.249	*0.749*
10	0.162	0.125	*0.649*
11	0.228	0.193	*0.689*
12	0.047	0.129	*0.771*
13	0.219	0.234	*0.767*
14	0.086	0.313	*0.676*
15	0.095	*0.706*	0.253
16	0.173	*0.661*	0.211
17	0.000	*0.743*	0.159
18	0.063	*0.687*	0.244
19	0.097	*0.848*	0.238
20	0.210	*0.711*	0.128
21	0.034	*0.764*	0.030

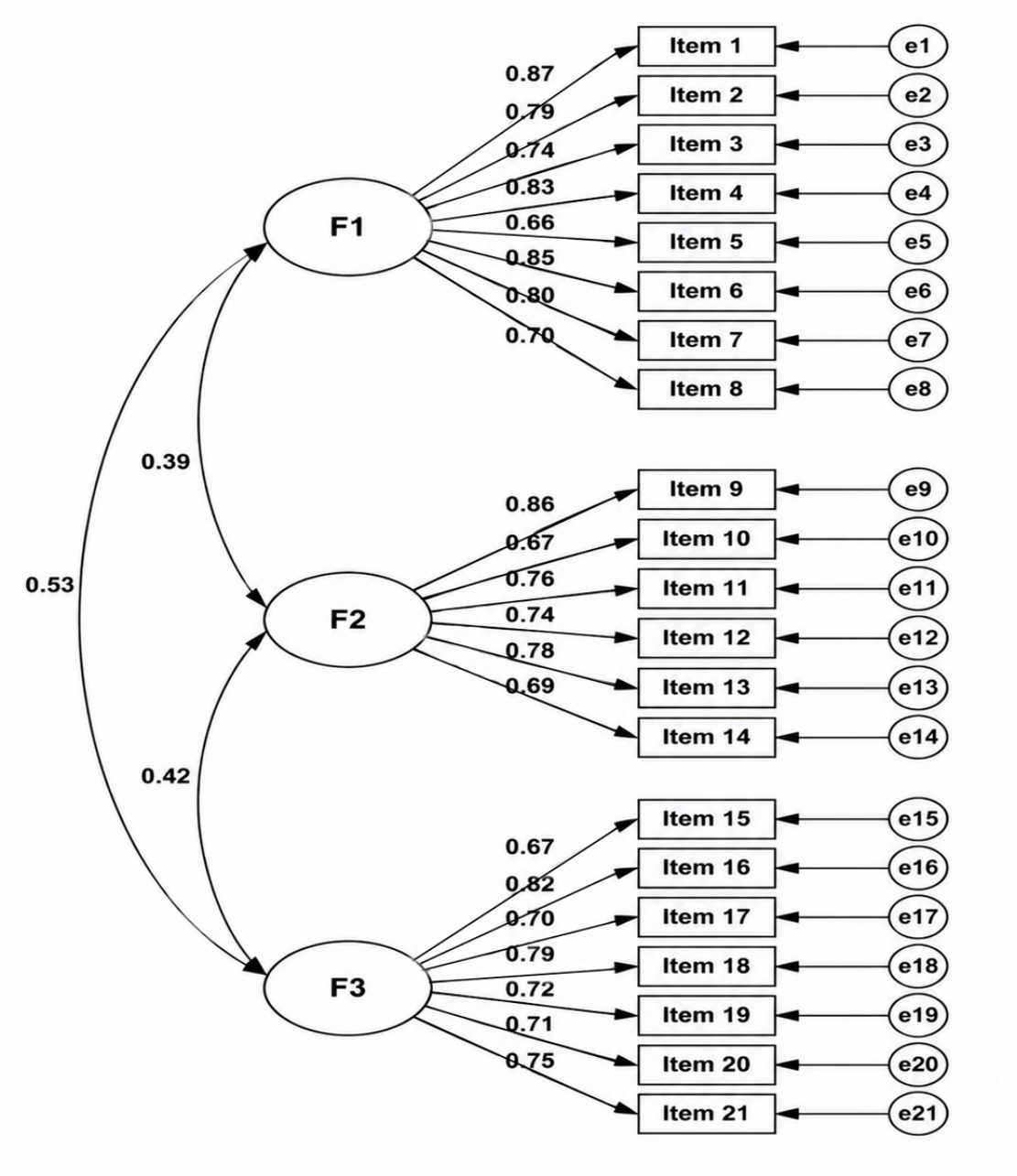
[a]The italicized items have a sufficiently strong and meaningful association with their assigned latent factors in the measurement model.

Confirmatory Factor Analysis

As shown in Figure 2, in this model, the minimum discrepancy/df value was 1.810; the root-mean-square error of approximation was 0.058; the incremental fit index and cumulative fit index values were both 0.950; the goodness-of-fit index, normed fit index, and relative fit index values were 0.887, 0.895, and 0.881, respectively; and the parsimony normed fit index value was 0.793. All the goodness-of-fit indexes met the general standard, which proves that the fitness of the model was good.

XSL•FO
RenderX

Figure 2. The confirmatory factor analysis model for the Chinese version of the mHealth App Usability Questionnaire. e: error, residual; F: factor.



Reliability Testing of the Questionnaire

The Cronbach α coefficient of the MAUQ-C was 0.919. A higher α value suggests greater internal reliability, and more than 0.700 is acceptable as good internal reliability. The values for test-retest reliability and split-half reliability were 0.881 and 0.726, respectively. The reliability of each dimension of the questionnaire is shown in Table 7.

Table 7. Questionnaire reliability analysis results.

	Cronbach α	Test-retest reliability	Split-half reliability
Dimension 1	0.928	0.759	0.933
Dimension 2	0.877	0.874	0.875
Dimension 3	0.886	0.745	0.903
Overall questionnaire	0.919	0.881	0.726

XSL•FO
RenderX

Discussion

Principal Findings

mHealth apps that provide health information and comprehensive management may be helpful both for patients and health care workers, and their use should be encouraged. The evaluations of health care providers as the main users of mHealth apps are also critical. Thus, we translated the original MAUQ (interactive apps for health care providers) into Chinese and validated the modified questionnaire in our study. This study represents the first cross-cultural adaptation and validation of the MAUQ for interactive apps for health care providers into Chinese, addressing a critical gap in usability assessment tools for medical professionals in China's mHealth landscape.

The Chinese adaptation process of the health care provider version of the MAUQ was standardized. Before the Chinese adaptation of this scale, authorization was obtained, and the preparatory work was standardized. The translation model by Brislin [17] was strictly followed. The translators and back translators all had a master's degree or above. For the disputed items in each version, decisions on modification were made through collective discussions by the research group, so the translation process was standardized and accurate. After the initial Chinese-adapted scale was developed, experts in the fields of medicine and English were invited to conduct cross-cultural adjustments from different professional perspectives. Revisions were made while ensuring the equivalence of the Chinese-adapted scale and the original scale to the greatest extent so as to ensure accuracy, comprehensibility of the item content, and its consistency with the local culture and further improve the adaptability of the scale. Then, a preliminary validation was carried out among medical staff. Methods such as item analysis and reliability and validity tests were used to test the internal consistency and effectiveness. The results of the presurvey study showed that the items of this Chinese-adapted scale were clear, which was convenient for the research participants to understand and answer. Therefore, the Chinese adaptation process of this scale was scientific and standardized, and the results were reliable.

Reliability evaluation can reflect the stability and consistency of the scale. In this study, the Cronbach α was used for consistency testing, and test-retest reliability was used for stability testing. The results showed that the Cronbach α of the MAUQ-C for health care providers was 0.919 (>0.8). The Cronbach α values of the 3 dimensions of the scale were 0.928, 0.877, and 0.886, all above 0.8, indicating that the Chinese-adapted scale had good internal consistency. The overall test-retest reliability of the scale was 0.881, and the test-retest reliability of each dimension was 0.759, 0.874, and 0.745, all above 0.7, indicating that this Chinese-adapted scale had good stability. Validity evaluation can show whether the scale truly and accurately reflects the content it surveys. In this study, methods such as content validity and factor analysis were used to test the validity of the Chinese-adapted scale. The content validity results showed that the I-CVI (0.833-1) of each item and the Scale-Content Validity Index/Average (0.992) of the scale met the standards, indicating that the experts had a high degree of recognition of the Chinese-adapted items and their corresponding measurement content and each item could effectively reflect the usability of the procedure, which indicated that the content validity of the MAUQ-C for health care providers was good. EFA extracted 3 common factors, which were consistent with the dimension division of the original scale, indicating that the item results were reasonable and the structural validity was good. The CFA results showed that the model fit met the standards (minimum discrepancy/df=1.810; root-mean-square error of approximation=0.058; incremental fit index=0.950; comparative fit index=0.950; goodness-of-fit index=0.887; normed fit index=0.895; relative fit index=0.881), indicating that the Chinese-adapted scale was consistent with the original scale in structure. These psychometric properties are consistent with previous validation studies of the MAUQ in other languages [8,11,15]. For instance, the 3-factor structure identified in our study aligns with the original development study by Zhou et al [12], confirming that the usability dimensions—usability and satisfaction, system information arrangement, and efficiency—are robust across different cultural and linguistic contexts. This cross-cultural stability suggests that the MAUQ-C is a reliable instrument for measuring usability among Chinese health care providers, offering a solid foundation for further research in this population.

Owing to the rapid development of the internet, mHealth apps have also developed rapidly. As the usability evaluation of apps can effectively identify and correct problems in and thereby optimize their operation, this method has become a common step in the app development process. In addition, the usability of apps also affects patients' use experience and the final intervention effect. Therefore, scholars at home and abroad have paid more and more attention to the usability evaluation of mHealth apps in recent years [10]. Unlike widely used instruments such as the SUS, Software Usability Measurement Inventory, or Questionnaire for User Interaction Satisfaction—which focus on technical usability for general software or websites—the MAUQ is uniquely tailored to mHealth contexts. Its differentiation between patient and health care provider versions and emphasis on interactive app functionalities addresses the specialized needs of health care professionals, whose perspectives are critical for optimizing the clinical workflow integration of mHealth tools.

The robust psychometric properties of the MAUQ-C suggest that usability is not a monolithic construct but is highly context dependent. Unlike general usability scales (eg, the SUS), the MAUQ-C's focus on system information arrangement and efficiency reflects the specific cognitive and operational demands placed on health care providers. The mechanism underlying this validity lies in the instrument's alignment with the clinical workflow; when an app reduces the cognitive load of data entry or information retrieval, app users perceive higher usability. This study's original contribution is that, first, it provides a psychometrically sound instrument tailored to the Chinese clinical environment and, second, it shifts the focus of mHealth evaluation from generic technical performance to professional-specific utility, which is a prerequisite for the successful integration of digital health into routine clinical practice.

XSL•FO
RenderX

Debates in mHealth usability research have highlighted the tension between subjective self-reported measures and objective behavioral metrics. While objective data (eg, task completion time and error rates) provide precise insights into technical performance, they often fail to capture the user's cognitive load, satisfaction, and perceived workflow integration—dimensions that are critical for long-term app adoption in clinical settings. Our study contributes to this debate by demonstrating that the MAUQ-C provides a reliable subjective assessment that complements rather than replaces objective performance data. However, we acknowledge that relying solely on self-reported instruments may be subject to recall and social desirability bias. Future research should aim to triangulate MAUQ-C scores with objective use logs to provide a more holistic view of mHealth usability.

Unlike previous system usability evaluation tools, the MAUQ was not only developed for mHealth apps. Its 4 original versions are also carefully divided according to whether the app interacts with patients, and the evaluation objects are divided into 2 categories: patients and medical staff, which is more targeted. While the patient version of the MAUQ has been adapted into Chinese, this study fills the unmet need for a validated health care worker–specific tool. This is significant because health care providers serve as both end users and intermediaries in mHealth interventions, making their usability feedback essential for enhancing app adoption and effectiveness in clinical settings.

Limitations

This study has several limitations. First, the evaluation was restricted to an app for populations with chronic diseases; future research should validate the MAUQ-C across diverse mHealth contexts to ensure broader generalizability. Second, the use of convenience sampling in 2 hospitals may introduce selection bias; thus, multicenter randomized studies are recommended to enhance representativeness. Third, the cross-sectional design precludes the assessment of long-term usability or the impact of app updates; longitudinal studies are needed to capture evolving user perceptions. In addition, conducting EFA and CFA on the same dataset may have inflated model fit indexes. Future validation should use a split-sample approach or independent datasets to provide a more rigorous confirmation of the factor structure. Finally, this study lacks an assessment of convergent and discriminant validity. Future research should compare the MAUQ-C with established usability instruments such as the SUS or the PSSUQ to further demonstrate its unique utility and construct validity in measuring mHealth usability for health care professionals.

Conclusions

The MAUQ-C for interactive apps for health care providers is a psychometrically sound and culturally adapted tool ideal for assessing mHealth app usability in Chinese clinical settings. Its validation advances mHealth research and practice by providing health care professionals, developers, and researchers with a reliable instrument to drive user-centered design and improve the quality of digital health interventions.

Acknowledgments

The authors sincerely thank Dr Leming Zhou for kindly granting permission to translate the mHealth App Usability Questionnaire into Chinese. They also extend their gratitude to all experts who provided valuable insights during the translation and cultural adaptation process. Finally, the authors would like to express their appreciation to the participants for their time and contribution to this study. Generative AI was used for language polishing and grammar correction.

Funding

This study received no funding.

Data Availability

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Authors' Contributions

MC, YG, and LZ contributed to conceptualization. YP, JZ, HH, and BY contributed to data collection. MW, DL, MC, YP, and WG contributed to data analysis. MW, DL, MC, and YP contributed to data interpretation and manuscript drafting. BY and WG contributed to critical revision of the manuscript.

Conflicts of Interest

None declared.

References

1. The 55th statistical report on China's internet development. China Internet Network Information Center. URL: https://www.100ec.cn/Public/Upload/file/20250120/1737335916881378.pdf [accessed 2025-01-20]

XSL•FO
RenderX

2. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013 Aug 06;1(2):160-171 [FREE Full text] [doi: 10.9745/GHSP-D-13-00031] [Medline: 25276529]
3. Nezamdoust S, Abdekhoda M, Ranjbaran F, Azami-Aghdash S. Adopting mobile health applications by nurses: a scoping review. J Res Nurs 2022 Aug;27(5):480-491 [FREE Full text] [doi: 10.1177/17449871221077080] [Medline: 36131693]
4. Yang X, Kovarik CL. A systematic review of mobile health interventions in China: identifying gaps in care. J Telemed Telecare 2021 Jan;27(1):3-22 [FREE Full text] [doi: 10.1177/1357633X19856746] [Medline: 31319759]
5. Ifejika NL, Bhadane M, Cai CC, Noser EA, Grotta JC, Savitz SI. Use of a smartphone-based mobile app for weight management in obese minority stroke survivors: pilot randomized controlled trial with open blinded end point. JMIR Mhealth Uhealth 2020 Apr 22;8(4):e17816 [FREE Full text] [doi: 10.2196/17816] [Medline: 32319963]
6. Kaddumukasa M, Nakibuuka J, Mugenyi L, Namusoke O, Birungi D, Kabaala B, et al. Feasibility study of a targeted self-management intervention for reducing stroke risk factors in a high-risk population in Uganda. J Neurol Sci 2018 Mar 15;386:23-28 [FREE Full text] [doi: 10.1016/j.jns.2017.12.032] [Medline: 29406961]
7. Al-Arkee S, Mason J, Lane DA, Fabritz L, Chua W, Haque MS, et al. Mobile apps to improve medication adherence in cardiovascular disease: systematic review and meta-analysis. J Med Internet Res 2021 May 25;23(5):e24190 [FREE Full text] [doi: 10.2196/24190] [Medline: 34032583]
8. Zhao S, Cao Y, Cao H, Liu K, Lv X, Zhang J, et al. Chinese version of the mHealth App Usability Questionnaire: cross-cultural adaptation and validation. Front Psychol 2022 Feb 2;13:813309 [FREE Full text] [doi: 10.3389/fpsyg.2022.813309] [Medline: 35185732]
9. Hajesmaeel-Gohari S, Khordastan F, Fatehi F, Samzadeh H, Bahaadinbeigy K. The most used questionnaires for evaluating satisfaction, usability, acceptance, and quality outcomes of mobile health. BMC Med Inform Decis Mak 2022 Jan 27;22(1):22 [FREE Full text] [doi: 10.1186/s12911-022-01764-2] [Medline: 35081953]
10. Zhou L, Bao J, Parmanto B. Systematic review protocol to assess the effectiveness of usability questionnaires in mHealth app studies. JMIR Res Protoc 2017 Aug 01;6(8):e151 [FREE Full text] [doi: 10.2196/resprot.7826] [Medline: 28765101]
11. Shan Y, Ji M, Xie W, Li R, Qian X, Zhang X, et al. Chinese version of the Mobile Health App Usability Questionnaire: translation, adaptation, and validation study. JMIR Form Res 2022 Jul 06;6(7):e37933 [FREE Full text] [doi: 10.2196/37933] [Medline: 35793132]
12. Zhou L, Bao J, Setiawan IM, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
13. Quifer-Rada P, Aguilar-Camprubí L, Gómez-Sebastià I, Padró-Arocas A, Mena-Tudela D. Spanish version of the mHealth App Usability Questionnaire (MAUQ) and adaptation to breastfeeding support apps. Int J Med Inform 2023 Jun;174:105062 [FREE Full text] [doi: 10.1016/j.ijmedinf.2023.105062] [Medline: 37037124]
14. Moulaei K, Dinari F, Sheikhtaheri A, Bahaadinbeigy K, Hajesmaeel-Gohari S. The Persian version of mHealth App Usability Questionnaire (MAUQ) for patients: a psychometric assessment study. BMC Med Inform Decis Mak 2025 Jan 23;25(1):38 [FREE Full text] [doi: 10.1186/s12911-025-02882-3] [Medline: 39849456]
15. Mustafa N, Safii NS, Jaffar A, Sani NS, Mohamad MI, Abd Rahman AH, et al. Malay version of the mHealth App Usability Questionnaire (M-MAUQ): translation, adaptation, and validation study. JMIR Mhealth Uhealth 2021 Feb 04;9(2):e24457 [FREE Full text] [doi: 10.2196/24457] [Medline: 33538704]
16. Moorthy P, Weinert L, Harms BC, Anders C, Siegel F. German version of the mHealth App Usability Questionnaire in a cohort of patients with cancer: translation and validation study. JMIR Hum Factors 2023 Nov 01;10:e51090 [FREE Full text] [doi: 10.2196/51090] [Medline: 37910144]
17. Brislin RW. Comparative research methodology: cross‐cultural studies. Int J Psychol 1976 Oct;11(3):215-229 [FREE Full text] [doi: 10.1080/00207597608247359]
18. Hu H, Zheng J, Lu X, Zhang L, Li D. Design and development of a brain and heart health management mini-program. Chin J Mod Nurs 2024;30(14):1851-1856 [FREE Full text] [doi: 10.3760/cma.j.cn115682-20231015-01507]
19. Ferrando PJ, Lorenzo-Seva U. Exploratory item factor analysis: some additional considerations. Ann Psychol 2014;30(3):1170-1175 [FREE Full text] [doi: 10.6018/analesps.30.3.199991]
20. Zamanzadeh V, Ghahramanian A, Rassouli M, Abbaszadeh A, Alavi-Majd H, Nikanfar AR. Design and implementation content validity study: development of an instrument for measuring patient-centered communication. J Caring Sci 2015 Jun 01;4(2):165-178 [FREE Full text] [doi: 10.15171/jcs.2015.017] [Medline: 26161370]

Abbreviations

CFA: confirmatory factor analysis
CVI: content validity index
EFA: exploratory factor analysis
KMO: Kaiser-Meyer-Olkin
MAUQ: mHealth App Usability Questionnaire

XSL•FO
RenderX

MAUQ-C: Chinese version of the mHealth App Usability Questionnaire
mHealth: mobile health
PSSUQ: Post-Study System Usability Questionnaire
SUS: System Usability Scale

Edited by A Stone, L Buis; submitted 30.Oct.2025; peer-reviewed by F Carrouel, K Becerro de Bengoa Losa, S Sahin; comments to author 27.Apr.2026; revised version received 29.Jun.2026; accepted 30.Jun.2026; published 29.Jul.2026.

Please cite as:
Chen M, Gao W, Pan Y, Wang M, Yu B, Li D, Hu H, Zheng J, Zhang L, Gao Y
Chinese Version of the mHealth App Usability Questionnaire for Health Care Providers: Cross-Sectional Translation and Validation Study
JMIR Mhealth Uhealth 2026;14:e86783
URL: https://mhealth.jmir.org/2026/1/e86783
doi:10.2196/86783
PMID:

©Mengxia Chen, Wenjun Gao, Yijie Pan, Mengdi Wang, Bing Yu, Dongmei Li, Huanhuan Hu, Jing Zheng, Lingjuan Zhang, Yuan Gao. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 29.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Design Preferences for Mental Health Apps Among Nondigitally Native Adults With Chronic Pain: Qualitative Analysis

Abby L Cheng[1], MD, MSc; Christine Y Gou[1], MD; Adriana Martin[1], MSc, JD; Sarah M Hartz[2], MD, PhD; Joanna Abraham[3], PhD

[1]Division of Musculoskeletal Physical Medicine and Rehabilitation, Department of Orthopaedic Surgery, Washington University in St. Louis, St. Louis, MO, United States

[2]Department of Psychiatry, Washington University in St. Louis, St. Louis, MO, United States

[3]Department of Anesthesiology & Institute for Informatics, Washington University in St. Louis, St. Louis, MO, United States

Corresponding Author:
Abby L Cheng, MD, MSc
Division of Musculoskeletal Physical Medicine and Rehabilitation
Department of Orthopaedic Surgery
Washington University in St. Louis
Campus Box MSC 8233-0004-05
660 S Euclid Ave
St. Louis, MO, 63110
United States
Phone: 1 314 747 8489
Email: chengal@wustl.edu

Abstract

Background: On a population level, mental health apps are accessible and effective. However, nondigitally native adults with chronic pain are a large and growing population who have been neglected during the development process of these interventions. Although technology use is rapidly growing among this population, their engagement with mobile health–related apps is lagging because usability is often not optimized for their needs and preferences.

Objective: This study aimed to identify design preferences and determinants of engagement with mental health apps by nondigitally native adults who have chronic pain and coexisting symptoms of depression or anxiety.

Methods: In this qualitative study, participants completed a semistructured interview regarding their experience with, and perceptions of, mobile devices, apps, and digital health interventions. Participants were 45 years or older; scored ≥10 on the 9-item Patient Health Questionnaire, 7-item Generalized Anxiety Disorder, or both; endorsed pain on most days or every day in the past 3 months; and were living in the United States. The interview guide was informed by the Consolidated Framework for Implementation Research and the Behavioral Intervention Technology model. Codes were organized into themes. Recruitment continued until thematic saturation was achieved.

Results: A total of 42 participants were interviewed (mean age 57, SD 8 years; n=32, 76% women). Participants strongly preferred apps that are free, describe strong privacy policies, and add functional value to their lives. They were more motivated by “real-life” goal achievement and tangible health improvements than by gamification within an app, and many were wary of allowing apps to passively collect certain types of data, especially to make inferences about their mental health. Most participants were unaware of apps designed to address chronic pain, but they were interested in the concept, particularly to help track their mood and pain and then identify associations between their symptoms, app activity, and other life events. Despite their daily use of apps, many participants described frequent challenges related to app navigation. While on-demand access to app tools was preferred, most participants appreciated the potential value of occasional push notifications if their timing was thoughtful and customizable. Participants cautioned against an overly cheerful, “infantile,” or informal tone for an app that addresses serious issues such as mental health and chronic pain.

Conclusions: Mental health apps for nondigitally native adults should highlight tangible health improvements that can be achieved from app engagement (more so than gamification), potentially using a multidomain tracking feature, if appropriate. This population is available to receive just-in-time adaptive interventions, but the frequency and timing should be thoughtful, customizable, and not wholly reliant on passively collected personal data. Health-related apps designed to address conditions that are more common with increasing age should account for these preferences.

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2026;14:e87358) doi:10.2196/87358

KEYWORDS

anxiety; chronic pain; depression; digital health; digital mental health intervention; engagement; mHealth; middle-age; mobile app; navigation; older adult; usability; user-centered design

Introduction

Chronic pain commonly coexists with depression and anxiety, and the prevalence of chronic pain increases with age. In the United States, nearly one-third of middle-aged and older adults are living with chronic pain, and up to half of them have depression, anxiety, or both [1-4]. Furthermore, chronic pain reduces the effectiveness of stand-alone mental health treatment unless a person's pain is simultaneously addressed [5,6]. Regardless of the etiology, chronic pain is associated with increased rates of recurrent depressive and anxiety episodes, reduced rates of remission, and reduced efficacy of antidepressant medications [7-10]. For these patients, addressing their mental health and pain experience simultaneously is more effective than addressing either problem in isolation [11].

Additionally, technology use is rapidly growing in these age groups, but engagement with digital health interventions is lagging because usability is often not optimized for their needs and preferences [2,7-10,12-23]. At this time, the majority of middle-aged and older adults are nondigital natives, meaning they were born before 1980 and grew up before immediate access to technology was widespread [24,25]. Compared to younger (and digitally native) users, this population more often has a greater interest in interventions that address multimorbidity; they prefer different screen layouts due to age-related changes in visual acuity, eye tracking patterns, and dexterity (eg, simple and flat navigation structure, large labeled buttons, and minimal need for typing); and they may benefit from more "push" notifications because they check their mobile devices less frequently. Some also have age-related cognitive deficits that reduce their ability to navigate through apps [20,26-31]. Although these age-related preference differences are now well-established, the user design of commercially available digital health apps has often continued to lag behind accepted design recommendations [32].

On a population level, digital mental health interventions, including mobile apps, effectively reduce depression and anxiety symptoms, and they have produced clinical outcomes that are equivalent to face-to-face treatment [17,19,33,34]. They also improve access to mental health support by reducing common barriers due to stigma, transportation, cost, and provider shortages that too often limit access to trained mental health professionals [35,36]. In recent years, the increasing availability of artificial intelligence has only accelerated public interest in using technology to address mental health symptoms [37-41]. Nevertheless, some populations have been neglected during the development process of mental health apps [42], and nondigitally native adults with chronic pain are one such understudied, yet growing and significant population [20].

The purpose of this study was to narrow this evidence gap by exploring design preferences of nondigitally native adults who have chronic pain and coexisting symptoms of depression or anxiety, with particular emphasis on examining determinants of engagement with mental health apps. This study is the first phase of a 3-phase project using the Discover, Design/Build, and Test (DDBT) framework to design and launch a mental health app that is optimized for the usability needs and design preferences of this population [43,44].

Methods

Study Design

In this formative qualitative study, each participant completed a single semistructured interview regarding their first-hand experience with, and perceptions of, mobile devices, apps, and digital health interventions [45,46]. Participant interviews occurred between April 3, 2024, and October 2, 2024. Data analysis occurred in 2024 and 2025. The study details are described in accordance with the Consolidated Criteria for Reporting Qualitative Research (COREQ) guidelines [47]. Of note, study participants also completed usability testing of a mental health app and provided feedback on specific usability-related features as part of the 3-phase project. These findings are reported separately [48].

Ethical Considerations

Prior to participant recruitment, approval of the study and all related procedures was obtained from the Washington University Institutional Review Board (IRB; 202311024). Participant consent was obtained via a waiver of written informed consent, and participants' data were maintained in secure databases and were deidentified before sharing with anyone outside the IRB-approved study team. Participants received a US $30 gift card as remuneration for participation.

Participants

To be eligible, potential participants had to be at least 45 years of age, report at least moderate depression or anxiety symptoms (9-item Patient Health Questionnaire [PHQ-9] score of ≥10, 7-item Generalized Anxiety Disorder [GAD-7] score of ≥10, or both), and endorse chronic pain (defined as pain on most days or every day in the past 3 months) [2,49,50]. Potential participants were excluded if they endorsed an active mental health crisis that warranted escalation of care (eg, active suicidal ideation and psychosis), had cognitive impairment that would interfere with communication with the interviewer, or lived outside the United States.

Recruitment

Potential participants were recruited from a variety of sources including from a chronic musculoskeletal pain clinic at an academic medical center (ie, printed flyers in clinic, phone call or email from the research team), email and social media advertisement via a university-based volunteer research

XSL•FO
RenderX

participant registry (ie, Facebook and Twitter/X), and targeted advertisement in the electronic newsletter of a free online health community (The Mighty). Initial eligibility screening was performed using a self-reported questionnaire hosted by REDCap (Vanderbilt University), followed by final verification by a study team member [51,52]. Recruitment and interviews continued until thematic saturation had been achieved.

Guiding Frameworks

Development of the interview guide was informed by the CFIR (Consolidated Framework for Implementation Research) and the Behavioral Intervention Technology (BIT) model [53,54]. The CFIR was used to ensure that all domains that could impact users' engagement with a mental health app were systematically addressed (ie, characteristics of the app user [individual], the user's home environment [inner setting], the user's social and structural support environment [outer setting], the user's preferred design features of the app [adaptable periphery], the app's therapeutic content [intervention], and the process of identifying, downloading, and onboarding to a new app [implementation process]) [53]. The BIT model was used to ensure that participants were prompted to comment on how an app's tools [elements], user design [characteristics], and intended timing of interactions [workflow] may impact their engagement with the app's behavioral interventions [54].

Interview Process

Semistructured interviews were conducted and recorded using audio/visual conferencing via a secure Zoom platform, each lasting approximately 45 to 60 minutes. The virtual interview format facilitated more accessible participation for people with mobility limitations related to pain, other medical conditions, transportation barriers, or geographic distance from the study site. Participants were also asked to answer sociodemographic descriptive questions via an electronic REDCap questionnaire, either before or after completing the interview. During the interviews, only the interviewer and participant were present. The interviewer described the rationale and purpose of the study, and then participants were asked questions from the interview guide (Multimedia Appendix 1).

Analysis

Audio recordings of the interviews were professionally transcribed [55], cleaned, and chunked. A preliminary codebook was developed by the senior investigator using a deductive approach guided by the CFIR and BIT models (ALC). The codebook and coding tree were refined using an inductive approach as 3 research team members coded 5 transcripts together (ALC, AM, and CYG). At that time, >90% consensus was reached, and there were fewer than 5 discrepancies between coders. After these discrepancies were discussed and resolved through consensus, the codebook was refined accordingly, and the rest of the transcripts were each coded by a single team member (AM or CYG) using NVivo 15 (Lumivero). Periodic coding cross-checks were performed across team members to ensure consistency. Next, we used thematic analysis to capture patterns in the data by examining relationships between codes [56]. Codes were organized into themes by a primary and secondary research team member. Participant recruitment, data collection via interviews, and data analysis were iterative and continued until thematic saturation was reached [56]. This was defined as the point at which no new themes or insights emerged from analysis of the last 3 interviews or related transcripts. Finalization of the thematic analysis was performed as a group discussion among 3 research team members (AM, CYG, and ALC), using the BIT model as a framework for organization. To optimally facilitate rapid intervention development and refinement as part of our 3-phase project, we aimed to identify and highlight both overarching themes and specific feature preferences among this population that could be implemented and iteratively tested using the DDBT framework.

Reflexivity, Rigor, and Reproducibility

The senior investigator was a female physician scientist with a clinical background in physiatry and musculoskeletal medicine (ALC). The interviews were conducted by a senior female research coordinator who has clinical research experience working with older adults and in the field of mental health (AM). The interviewer was not involved in clinical care, and participants' only contact with the interviewer was related to study recruitment, consent, and participation. The third member of the coding and analysis team was a female physiatry resident trainee (CYG). Additionally, a clinician scientist with mental health clinical expertise (SMH) and a health IT expert (JA) were involved in finalizing the study design, interview guide, and interpretation of themes. Although formal member checking was not possible due to the single-session nature of participant involvement, informal member checking was performed by the interviewer by paraphrasing participants' statements during interviews to ensure the participants' thoughts and opinions were correctly understood. Additionally, preliminary themes from early interviews were iteratively shared with later participants, who generally affirmed their face validity.

Results

Overview

Of 62 people who completed the screening questionnaire and were determined to be eligible, 42 participants who represented a range of sociodemographic backgrounds were invited and participated in the study, at which point thematic saturation had been achieved (mean age 57, SD 8 years; n=32, 76% women; at least n=28, 67% non-Hispanic White; n=40, 96% used the internet at least multiple times a day; Table 1). We identified themes that cut across the CFIR domains and related to factors that impact participants' general use of mobile devices and apps, in addition to determinants of engagement with health-related apps and specifically mental health apps.

XSL•FO
RenderX

Table 1. Participant characteristics (N=42).

Characteristic	Value
Age (years), median (IQR; range)	55 (51-61; 45-76)
Sex[a], n (%)	
Female	32 (76)
Male	10 (24)
Race, n (%)	
White	30 (71)
Black or African American	9 (21)
Other	1 (2)
More than one race	1 (2)
Prefer not to answer	1 (2)
Ethnicity, n (%)	
Hispanic or Latino	3 (7)
Not Hispanic or Latino	36 (86)
Unknown or prefer not to answer	3 (7)
Employment status, n (%)	
Disabled	13 (31)
Working full-time	11 (26)
Retired	10 (24)
Working part-time	4 (10)
Unemployed	2 (5)
Prefer not to answer	2 (5)
Frequency of internet use[b], n (%)	
Most of the day	12 (29)
Multiple times a day	28 (67)
At least once a day	1 (2)
Less than once a day	1 (2)
Duration of chronic pain (years), median (IQR; range)	12 (8-20; 2-65)
High-impact chronic pain present[c], n (%)	33 (80)
BPI[d] Pain Interference score, median (IQR; range)	7.1 (6-8.1; 2.4-9)
PHQ-9[e] score, median (IQR; range)	14 (12-17; 7-22)
GAD-7[f] score, median (IQR; range)	12 (9-15; 1-21)
Prior mental health treatments, n (%)	
Medication	30 (71)
Psychiatrist, psychologist, or therapist	28 (67)
Self-management (meditation, mindfulness, deep breathing, etc)	21 (50)
Support group	10 (24)
Mental health app	4 (10)
None	3 (7)

[a]Sex and gender identity were concordant for all participants.

[b]Self-reported internet use was captured by the Pew Digital Savviness Classifier [57].

[c]"High impact chronic pain" is defined as pain that limits life or work activities on most or every day over the past 3 months [2].

XSL•FO
RenderX

[d]The Brief Pain Inventory (BPI) Pain Interference score ranges from 0 to 10 [58]. Higher scores indicate greater pain interference.

[e]The 9-item Patient Health Questionnaire (PHQ-9) is a 9-item screening measure for depressive symptoms [49]. Scores range 0-27. Higher scores suggest more depressive symptoms.

[f]The 7-item Generalized Anxiety Disorder (GAD-7) is a 7-item screening measure for anxiety symptoms [50]. Scores range 0-21. Higher scores suggest more anxiety symptoms.

General Use of Mobile Devices and Apps

Users Are Constantly Connected

Our nondigitally native adult participants reported having their mobile devices within arm's reach essentially during all waking hours, largely out of necessity for daily life activities such as finances and urgent communication, as well as for leisure especially related to social connection (Table 2). Most participants reported typically responding to mobile device messages immediately or at least within the same day.

XSL•FO
RenderX

Table 2. General patterns of mobile device and app use among nondigitally native adults who have chronic pain and coexisting symptoms of depression or anxiety.

Theme	Representative quote
Tools	
Life navigation: participants typically used apps "because they have to" *for daily life* (eg, for driving directions, banking, taxes, shopping, coupons, alarm clocks, current events, and having a contact method in case of emergencies).	• "I pretty much use my phone for absolutely every aspect of my life...work, grocery list, like everything." (48-year-old male)
Leisure: participants also frequently used apps that facilitate *social connection and entertainment* (eg, social media, video chatting, taking pictures, shopping, and, less commonly, to play games).	• "Keep in touch with other people, movies and TV, sometimes books, sometimes for my grandkids." (60-year-old female)
User design	
Usability challenges with mobile devices: participants commonly experienced challenges using their mobile devices due to *dexterity difficulty* with their screen's size and *forgetting passwords*.	• "Remembering my password is definitely [a challenge]. Sometimes finding the right one, because if you put in an AT&T one, sometimes there's three or four of them and it's like, 'Which one is the one that I should actually be downloading?'" (60-year-old female)
Usability challenges with apps: participants also reported challenges navigating within some apps, understanding *how to take advantage of all an app's features*, and *reducing unwanted notifications* for some apps.	• "Some of the apps are confusing with how they want you to move around, and they don't really explain it." (60-year-old female) • "Trying to figure out everything that the app does—to make sure I'm using the maximum of everything that it does." (53-year-old female) • "Some of my apps give me notifications that I don't really want. Figuring out how to turn off unwanted notifications can be a challenge." (64-year-old female)
Troubleshooting: when participants have trouble navigating an app, they typically *first "push around" on their own* to troubleshoot. Next, they search the internet for assistance. Finally, they reach out to their social circle or a professional for navigation assistance if needed.	• "Pretty much in this order—I will try it on my own. Then I will look up a video because it's like, 'I'm determined to do this.' But then if it's something too complicated, then usually I call one of my kids." (53-year-old female)
Timing of interactions	
Availability for alerts: most participants kept their smartphones *almost always within arm's reach*, except maybe silenced while sleeping or driving.	• "I rarely turn it off." (60-year-old female)
Response time: participants typically responded to mobile device messages *immediately or within the same day*.	• "[I respond] when I see them, if I need to." (50-year-old female)
General app preferences	
Pros: participants preferred apps that are *free*, *functionally meaningful*, personally relevant, and ideally have good brand reputation.	• "I don't typically pay for a lot of the apps. There's usually free versions of something out there that you can find...There is a medication that I use that they have you download the app for, and there's no point to it. There's no purpose to it, so it's just taking up space on my phone." (49-year-old female)
Cons: participants were less likely to use apps that raise *privacy concerns* about how their data will be used or that frequently display advertisements. Some participants were also less likely to use apps that require high data usage or a stable internet connection.	• "Making sure tracking and invasion of privacy is not an issue." (72-year-old male)

Preference for Free, Useful, Trusted Apps

Participants strongly preferred apps that were free. They also gravitated toward apps that were functionally meaningful, personally relevant, and well-established with a good reputation. They reported being less likely to use apps that do not provide sufficient reassurance that users' privacy and data will be protected. Many also disliked frequent advertisements within apps.

Challenges With Navigation and Notifications

Participants reported commonly experiencing usability challenges with mobile devices due to difficulty with dexterity and forgetting passwords. They also described challenges navigating some apps, particularly with regard to understanding how to use all of an app's features and reducing unwanted notifications from apps. When usability challenges are encountered, participants most frequently reported troubleshooting by "pushing around" on their own, followed

XSL•FO
RenderX

by searching the internet for assistance (eg, Google and YouTube), and then reaching out to their social circle or professional assistance as a last resort.

Determinants of Engagement With Health-Related Apps: Interest in Content and Tracking Capabilities That Result in "Real-Life" Goal Achievement

Participants most frequently reported using health-related apps for education (eg, to learn new information), accountability (eg, for assigned tasks and reminders), and community (eg, to connect with people who are living with similar challenges; Table 3). Tracking capability was a common theme among desired functions of health-related apps, whether to track physical activity, biometrics, mood, food intake, or another health-related measure. Nevertheless, participants insisted on clear communication and ability to choose which types of passive data are collected and for what purposes, due to concerns related to privacy and "creepiness" (eg, acceptable to passively collect step count but not location). Most participants also did not consider gamification with rewards within a health-related app to be a meaningful motivator of behavior change or engagement. However, participants did report being more likely to value and use a health-related app if it is recommended or hosted by a medical professional. Regarding the timing of app interaction, participants highly appreciated the capability to customize the time and frequency of push notifications and other prompts from the app.

XSL•FO
RenderX

Table 3. Determinants of engagement with health-related apps among nondigitally native adults who have chronic pain and coexisting symptoms of depression or anxiety.

Theme	Representative quote
Tools	
Goals: participants used health-related apps for *education, accountability, and community*.	• "I primarily enjoy getting information from them." (61-year-old female) • "I've done [a fitness app]...and [a mental health app]. I know what I need to do each day, and I don't have to worry about remembering it. So I can log in, and each day I know I need to tackle these kind of tasks." (49-year-old female) • "I use [an online health community app]. A lot of people have the same experiences I do. So that's why I like it. The same thing that I go through—people not believing me—it gives you a lot of insights." (51-year-old female)
Function: in descending order, participants most commonly used apps for activity/fitness tracking; blood sugar, blood pressure, and weight tracking; mental health, mindfulness, and mood tracking; food/calorie tracking; medication refills and communication with a health care provider; and sleep assistance.	• "I have a [fitness tracking] app for monitoring my steps and water. My oncologist, once I was finished with treatment, wanted me to make sure I walked 10 to 15,000 steps a day." (60-year-old female)
Tracking: participants predominantly used health-related apps to assist with tracking one or more elements of their health.	• "The step counter is really probably the only one I use regularly." (50-year-old female)
User design	
Selective enthusiasm for passive data: the majority of participants were *open to passive collection of some types of data but not others* due to privacy concerns. If an app does passively collect data, they strongly voiced a need for the app to clearly describe the purpose and extent upfront, with an option to opt out.	• "I tend to steer away from apps that ask me that kind of information because I'm not comfortable being tracked in where I am, what I'm doing. I just find it intrusive. I mean, I put my weight on there, my height, it has my blood pressure and my heartrate—so that's pretty personal, but location—I do not share... [Passive data tracking] is a little creepy." (66-year-old female) • "I want to be able to pick when I use [passive data]. I don't want it to be able to use it and pick me. I mean, I need reminders to use [the app], but not by my location." (62-year-old female)
Low enthusiasm for gamification: most participants *did not consider embedded games or rewards as a meaningful motivator* to engage with a health-related app.	• "I don't know if I'm a typical guy or maybe not, but I don't like games. When it comes to something like this, I'm not motivated by a reward. I would rather see [all my tool] options than, 'Good boy, good boy.' ...When it comes to counting my steps, it is nice at the end of the night to have hit my target or beyond. So in that sense, yes, an emotional reward is a good thing. But the reward is that I've accomplished what I want. I don't need something extra on top of that." (72-year-old male)
Timing of interactions	
Participants strongly preferred the ability to *customize their timing of interaction* with the app (eg, time and frequency of push notifications and ability to turn off push notifications).	• "I like if it allows you to set what time you want [for check-ins]." (64-year-old female) • "I would like the option [of changing the check in frequency]." (55-year-old female)
General app preferences	
Source: participants were more likely to use a health-related app if it is *recommended by a medical professional* or directly offered from their medical provider/hospital system.	• "It's normally based on advice from a medical professional that I would download something of this type." (51-year-old male)

Determinants of Engagement With Mental Health Apps That Address Chronic Pain

Explanation of How "An App Can Help With Pain"

Many participants were unaware of apps that were designed to address chronic pain (in addition to mental health), nor did they understand how an app could address chronic pain (Table 4). When it was explained that such an app could guide users through pain coping strategies, participants were generally enthusiastic and expected that such an app would be helpful. Given the relative unawareness, participants expected that early in the onboarding and orientation process, such an app would provide education on the neuroscience of chronic pain and how it could help them manage and cope with chronic pain. They also expressed interest in tools related to mindfulness, cognitive behavioral strategies to improve pain acceptance, and gentle physical exercise. They generally preferred the app to guide them through a structured curriculum initially, but also to allow for self-guided exploration of app tools, if desired.

XSL•FO
RenderX

Table 4. Preferred features of a mental health app that addresses chronic pain among nondigitally native adults.

Theme	Representative quote
Tools	
Features	
• Participants were interested in having access to a variety of tools including education about chronic pain neuroscience, mindfulness exercises to help manage pain, strategies to improve pain coping and acceptance, pain and mood tracking capabilities, and gentle physical exercise. They appreciated immediate, 24/7 access to management tools.	• "I think it's probably going to be some kind of a help [to have a] mindfulness activity to help control and manage pain." (56-year-old female) • "Give me a summary of how my week was...pain, mood—both depression and anxiety—which I'm sure part of my anxiety is from the pain. Yeah." (66-year-old female) • "Last night I had trouble sleeping because of my lower back, and it's progressively getting worse. I have injections next week and so on, and I'm really looking forward to them. But in the meantime—getting up in the middle of the night and just saying, 'My back hurts, maybe there's this app I can check out.' Could be interesting." (51-year-old male)
Mood tracking	
• Participants were interested in periodic, brief, self-report assessments to track their mental health over time (eg, 4 questions as opposed to 16 questions). Participants preferred self-reporting their mood, as opposed to the app using passively collected data to infer their mood (eg, with call logs, step count, and GPS data). • Participants would be more motivated to complete self-assessments if the results are used to tailor suggestions of additional tools for them. • Participants voiced mixed sentiments regarding whether they would want their health care team to see responses to their self-reported mental health assessments.	• "If it asked, 'Hey, how was your week? Did you have any kind of issues that you need to talk about?' And then if you're honest with the app, then maybe it can touch base a little bit sooner." (48-year-old male) • "[Passive data collection to infer mood] is a little creepy." (66-year-old female) • "It can be a double-edged sword. For example, if you answer incorrectly, 'Do you feel like harming yourself?' Is somebody [from your healthcare team] going to say, 'Wait a minute, there are specific laws that say we have to take action and you have to be under care now.' If it's chronic pain, that's different. But if it's mental health, there's a big stigma with it." (51-year-old male)
Other tracking feature	
• Participants were interested in a feature to track associations between their pain symptoms, mental health symptoms, physical activity and life events, and engagement/tasks completed within the app. Their goal from this complex tracking feature would be to identify what app tools and general behaviors were helpful to reduce their symptoms. Some participants would also want to show this data to their health care team.	• "Yeah, maybe my activity level is higher. Maybe my emotional level is better. I would think tracking all of those things that correlate directly to my pain level would be good...Something where you can really see the results of the app working with you or for you would be good, but I think it would feel a little disconnected without also tracking any kind of pain scale." (56-year-old female)
Methods to encourage sustained engagement	
• Participants expected they would be motivated to sustain engagement with the app if they were instructed to set goals and had pending tasks on a to-do list, if the app helped them track their tasks completed and corresponding symptom improvement, and if the app suggested new features for them to try. They were generally not in favor of using access to new features as a reward/motivator for continued engagement. • Participants were also in favor of a free trial period of paid content before deciding whether to purchase the app (or an upgrade). There was a slight preference for a content-limited trial rather than a time-limited trial (eg, 2-4 weeks).	• "Is there stuff that you could add that you're working on in the next week, or maybe things that you didn't fully complete?" (60-year-old male) • "Suggest a tool, something that's actually useful as opposed to a gold star. Yes. Yes." (53-year-old female) • "Yeah. Having a trial and then going to paid content after the trial—that would determine for me, 'How well is this working?'" (60-year-old female)
User design	
Onboarding	
• Participants requested a clear explanation regarding how the app can help with chronic pain and mental health (eg, coping and management strategies, rather than a focus on pain severity reduction).	• "I am confused, and I don't know how this is going to work. Yeah, I guess my knees and my back are preventing me from exercising. Is that what you're talking about? Mental and physical health going together?" (66-year-old female) • "I guess...intervene in letting pain take over. I think by checking in or doing mindfulness things, you could possibly alleviate chronic pain." (56-year-old female)

XSL•FO
RenderX

Theme	Representative quote
Tone	
• Some participants thought an overly cheerful, personal tone presented by a cartoon-like avatar was less suitable for older adults and felt artificial. Similarly, participants had mixed opinions on the use of emojis throughout the app experience.	• "These kinds of issues, we don't want to make light of them, and I think a [cartoon-like chatbot avatar] kind of makes it a little childish. If this was geared toward children and teens, ok. But for adults, I think it needs to have something a little more serious... I think I would be more comfortable with more of a question-and-answer screen—more like journal prompts than a text conversation." (60-year-old female) • "For me, [emojis are] very unnecessary. [They're] infantile. I mean, but I'm old. So a lot of this stuff is goofy now, that wouldn't be goofy to other people. If a kid is using it, they might like that." (69-year-old female) • "I think an emoji is like putting an expression to a sentence. So I think it's amazing." (52-year-old female)
User flow	
• Participants generally preferred for the app to initially guide them through a structured curriculum (such as for a duration of 8 weeks) but to also allow for self-guided exploration through the app, if desired.	• "It seems like once you become familiar with what it is, self-guided would definitely be preferable." (56-year-old female) • "I think eight weeks would be good because it takes a significant amount of time to change habits or incorporate new habits." (54-year-old female)
Timing of interactions	
Notifications to prompt engagement	
• Most participants preferred to initiate engagement with the app on their own, but many participants were also welcoming or at least accepting of occasional notifications to encourage engagement if relevant. • Most participants expressed that once or twice daily notifications would be acceptable, and some participants preferred an even greater frequency of notifications. For once daily notifications, evening check-ins were more often preferred to morning check-ins. • Participants voiced varied preferences regarding the option for notification frequency to be adaptive based on their engagement. Some participants were in favor of pausing notifications (eg, for 3-7 days), while others did not want a pause in notifications unless they were specifically asked first.	• "If it recognizes that you struggled, and if you're honest with the app, then maybe it can touch base a little bit sooner." (48-year-old male) • "[Check-ins twice a day] is okay with me because I don't do anything. I'm in pain, I don't work, and all I do is go back and forth to the doctors." (55-year-old female) • "Like, you could say, 'Dismiss for three days,' or however long it is. Or let's say you're going on vacation, you could set a time limit to dismiss for seven days or however long it needs to go. Give me the option to click that button, then I can go from there." (49-year-old female) • With regards to pausing notifications: "Don't forget about me now. Everybody else in the world does. And don't give up on me. I look at that as giving up on me. Don't do that. Don't leave me." (62-year-old female)
General app preferences	
Improved awareness	
• While most participants were aware of apps designed to address mental health, many participants were unaware of apps that address chronic pain. They were also unclear as to "how an app can help with pain."	• "I had no idea there were any [apps to help manage pain]. I'm wondering what [a pain app] can do because I have no clue." (61-year-old male)
Other stressors to address	
• Participants also suggested additional stressors that should be addressed by an app that aims to improve mental health and chronic pain. These stressors commonly included other health concerns, finances, and low motivation. Mobility limitations, shame, attention issues, frustration, loneliness, time pressures, and work were also suggested.	• "It could be other health challenges. You could have something related other than just pain." (53-year-old female) • "Finances play a big part of everything you do because if you don't have it, you can't do it." (55-year-old female)

Interest in Self-Reported Symptom Tracking

Participants were particularly interested in a tracking feature that would help identify associations between their pain, mental health symptoms, daily activities and events, and engagement with the app. To track mood, participants generally preferred completing periodic, brief self-assessments, rather than having the app use passively collected data, such as call logs or step count, to infer their mood, due to concerns about privacy and the accuracy of inferences. Participants reported being motivated to complete these self-assessments if the results lead to a more personalized and relevant app experience. They were somewhat

XSL•FO
RenderX

conflicted as to whether they wanted their self-reported symptoms to be shared with their health care team. While many participants anticipated added clinical value from sharing data such as a pain diary, some of these same participants (in addition to others) expressed concern regarding how mental health symptom reports may be handled by licensed health care providers. They feared that a clinician may misinterpret their symptom reports and feel obligated as a mandatory reporter to act on concern for potential imminent danger, even if none exists.

Encouraging But Mature Tone

Several participants cautioned against an overly personal or cheerful tone for an app that addresses chronic pain and mental health because it felt artificial and less suitable for their age group. Similarly, participants had mixed opinions on the use of a cartoon-like avatar or emojis throughout the app experience. Some felt they were "infantile," whereas others felt they enhanced communication. To maximize the relevance of an app that addresses chronic pain and mental health, participants suggested additional stressors that should also be acknowledged and ideally addressed, such as other health issues, financial pressures, and low motivation.

Customized Frequency of Notifications

Most participants preferred to initiate engagement with the app on their own, but many participants were also welcoming, or at least accepting, of notifications to encourage engagement if relevant. Participants typically described once- or twice-daily notifications as being most appropriate (especially evening notifications). Most participants were also open to an adaptive frequency of notifications based on their degree of engagement. However, some participants voiced concern about feeling like the app "ghosted [them] just like people do" if notifications suddenly stopped without their explicit request.

Motivated by Real-World Results

Regarding sustained engagement with an app that addresses chronic pain and mental health, participants expected to be motivated to continue engaging if they had specific goals and a to-do list within the app, especially if they could see a correlation between task completion within the app and symptom improvement. Participants were not in favor of using access to new app tools as a reward for engagement because they wanted to have access to whichever tool they thought was most relevant at the time they needed it. Instead, participants were generally open to a free trial period of app content, which may then motivate them to purchase access to additional content (or continued access to the original content).

Discussion

Principal Results

This qualitative study investigated design preferences and determinants of engagement with mental health apps by nondigitally native (middle-aged and older) adults who have chronic pain. Similar to findings across many other user populations, our participants in this population were generally avid mobile device users, and they strongly preferred to use apps that are free of charge, describe reassuring privacy policies, and add clear functional value to their lives [59]. Participants were also generally enthusiastic about health-related apps, particularly for tracking/logging purposes, and while on-demand access to app tools was clearly preferred, most participants also appreciated the potential value of occasional push notification "engagement reminders" if the notification timing was thoughtful and customizable [59]. Building on existing literature, this study also contributes some novel insights that are more unique to this nondigitally native population with chronic pain, when compared to younger users and users without chronic pain. Specifically, despite their daily use of mobile apps, many participants described frequent usability challenges, particularly related to app orientation and navigation (not just due to technical issues such as an app crashing). They were also more consistently motivated by "real-life" goal achievement and tangible health improvements than by gamification within an app, and many were wary of allowing apps to passively collect data, especially for the purpose of making inferences about their mental health. Furthermore, the majority of participants were unaware of apps designed to address chronic pain, but they showed a high level of interest in the concept. They were particularly interested in an infrastructure to track both their mood and pain symptoms and then identify associations between their symptoms, their activity on the app, and other life events. Finally, the participants cautioned against an overly cheerful, "infantile," or informal tone for an app that is intended to support adult users with serious issues related to mental health and chronic pain.

Comparison With Prior Work

The constellation of preferences that participants expressed presents a challenge when designing and delivering a health-related app that meets both their stated preferences and needs. Specifically, participants stated that they prefer not to be excessively bothered by push notifications. Yet, they also prefer to self-report their symptoms rather than have their symptoms inferred from passive data, and they describe low motivation as a common challenge in daily life—which is consistent with previous studies that have observed reduced app engagement among users with a high severity of depressive symptoms and fatigue (such as due to chronic pain) [59]. Of note, depression and fatigue are also barriers to engagement with nondigital health interventions [60].

Ultimately, we conclude that despite growing technological potential for passively collected data to infer changes in users' mental and physical health accurately [61,62], feedback from our participants suggests that at least at this time, digital health interventions for middle-aged and older adults should not rely heavily upon passively collected data. Rather, careful iteration and customization of push notification frequency, timing, and content have consistently proven to be a key strategy for optimizing engagement with digital health interventions among people across the age spectrum [63]. For this population, effective notifications may include reminders about available tools that the user previously found helpful, previews about pain neuroscience principles that could meaningfully impact the user's mindset, and/or insights into how app engagement could help address functional challenges previously reported

XSL•FO
RenderX

by the user. Adaptive timing of notification delivery based on users' demonstrated engagement patterns may also improve acceptance and engagement [64], and any use of passively collected data (beyond app engagement data) should likely be offered as an opt-in feature, rather than a default feature. Quantitative analysis of engagement patterns and symptom severity in response to varied push notification strategies is warranted.

Another challenge is that participants preferred apps that are both free of charge but are also not filled with obtrusive advertisements. This is consistent with preferences among younger age groups in which 67% of people report only using free apps and an additional 22% only use apps that cost $5 or less [65]. Third-party reimbursement of digital health interventions, either as stand-alone therapeutics or as part of a treatment bundle offered by a health system, could address these financial preferences and also add legitimacy to the intervention in the eyes of potential users.

Participants' strong interest in health-related tracking capabilities poses a unique conundrum for chronic pain-related apps, as well. While tracking metrics such as step count and sleep quality are essentially universally encouraged by health care professionals, an app that frequently prompts a user to report their pain intensity could unintentionally increase pain rumination and subsequent discomfort and impairment [66-68]. This phenomenon is supported by prior studies among different target user populations, which have simultaneously reported high popularity of pain tracking features and mixed feedback regarding the helpfulness of these features [69,70]. Accordingly, we propose that a chronic pain-related app may consider enabling the user to systematically log pain *interference* (and potentially add free-text notes about pain severity), rather than prompting the user to frequently reflect on pain intensity.

To have the greatest impact specifically for people who are struggling with chronic pain, we also need to improve awareness about the potential and availability of digital health interventions to deliver evidence-based pain management techniques such as cognitive behavioral therapy and mindfulness exercises. Based on participants' responses in this study, middle-aged and older adults may be most interested and accepting of such an intervention if it is introduced and recommended by a trusted member of their health care team. To most effectively increase awareness, these discussions could occur in primary care and orthopedic clinics, as well as mental health care settings [71]. Feasible, sustainable uptake of this workflow would likely be facilitated by use of established billing codes for behavioral health integration, digital mental health treatment, and/or remote therapeutic monitoring.

Strengths

The broad potential applicability of our findings across multiple health domains is a strength of this study. Although the therapeutic purpose and content of various health-related apps may be somewhat distinct, common themes related to target users' preferred engagement techniques and usability features exist. Some findings from this study can be used to inform optimal design of other digital mental health interventions and/or other health-related apps for nondigitally native middle-aged and older adults, not just related to chronic pain but potentially for other purposes such as medication monitoring, organ failure symptom monitoring (eg, for heart failure, kidney disease, etc), or encouragement to engage in social interaction or physical activity.

Limitations

A limitation of this study is that the study population preferentially represents middle-aged, White females from the United States who use the internet at least multiple times daily. Therefore, we may not have fully captured themes that are uniquely relevant for older adults and for adults from other geographical or sociotechnical backgrounds. Nevertheless, even though the majority of participants reported daily internet use, we still identified usability challenges and specific opportunities to tailor digital health experiences for this group. Of note, our findings represent participants' *stated* usability and engagement preferences; quantitative assessment of participants' *actual* engagement with apps that are concordant versus discordant with their stated preferences was outside the scope of this study.

Conclusions

Nondigitally native adults who are avid mobile device users still often encounter usability challenges and report design preferences that likely differ somewhat from those of younger users. To optimally meet the preferences and needs of this population, digital health interventions should emphasize tangible health improvements (more so than gamification) that can be achieved by engaging with the intervention, potentially using a multidomain tracking feature if appropriate. This population is generally available to receive real-time messages such as for just-in-time adaptive interventions, but the frequency and timing of any push notifications should still be thoughtful and customizable. This population remains wary of allowing passive data collection of their personal information. In conclusion, digital health interventions designed to address conditions that are more common with increasing age should account for these preferences.

Acknowledgments

The authors thank Dr Patricia Areán for her contributions to the study conception. Generative AI was not used in any portion of conducting this study or generating this manuscript.

XSL•FO
RenderX

Data Availability

The dataset generated and analyzed during this study is available in the National Institute of Mental Health (NIMH) Data Archive (NDA) [72].

Funding

This study was funded by the National Institute of Mental Health (grant R01MH131989; principal investigator ALC). The funder did not play a role in study design or data collection, management, analysis, or interpretation; writing of the report; or the decision to submit the report for publication.
Use of Washington University's instance of Research Electronic Data Capture for this study was supported by the Siteman Cancer Center's National Cancer Institute (NCI) Cancer Center Support Grant P30CA091842 and by the Washington University Institute of Clinical and Translational Sciences Grant UL1TR002345 from the National Center for Advancing Translational Sciences (NCATS).

Authors' Contributions

Conceptualization: ALC, JA
Methodology: ALC, CYG, SMH, JA
Formal analysis: ALC, CYG, AM, SMH, JA
Investigation: AM
Data curation: AM
Supervision: ALC, JA
Funding acquisition: ALC
Writing—original draft: ALC, CYG, AM
Writing—review and editing: ALC, CYG, AM, SMH, JA
Validation: CYG, AM

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File , 29 KB - mhealth_v14i1e87358_app1.docx]

References

1. Gureje O, Von Korff M, Simon GE, Gater R. Persistent pain and well-being: a World Health Organization study in primary care. JAMA 1998;280(2):147-151. [doi: 10.1001/jama.280.2.147] [Medline: 9669787]
2. Dahlhamer J, Lucas J, Zelaya C, Nahin R, Mackey S, DeBar L, et al. Prevalence of chronic pain and high-impact chronic pain among adults - United States, 2016. MMWR Morb Mortal Wkly Rep 2018;67(36):1001-1006 [FREE Full text] [doi: 10.15585/mmwr.mm6736a2] [Medline: 30212442]
3. Ruoff GE. Depression in the patient with chronic pain. J Fam Pract 1996;43(6 Suppl):S25-S33. [Medline: 8969710]
4. Zelaya CE, Dahlhamer JM, Lucas JW, Connor EM. Chronic pain and high-impact chronic pain among U.S. adults, 2019. NCHS Data Brief 2020(390):1-8. [Medline: 33151145]
5. Sharpe L, McDonald S, Correia H, Raue PJ, Meade T, Nicholas M, et al. Pain severity predicts depressive symptoms over and above individual illnesses and multimorbidity in older adults. BMC Psychiatry 2017;17(1):166 [FREE Full text] [doi: 10.1186/s12888-017-1334-y] [Medline: 28472936]
6. Sheng J, Liu S, Wang Y, Cui R, Zhang X. The link between depression and chronic pain: neural mechanisms in the brain. Neural Plast 2017;2017:9724371 [FREE Full text] [doi: 10.1155/2017/9724371] [Medline: 28706741]
7. DeVeaugh-Geiss AM, West SL, Miller WC, Sleath B, Gaynes BN, Kroenke K. The adverse effects of comorbid pain on depression outcomes in primary care patients: results from the ARTIST trial. Pain Med 2010;11(5):732-741. [doi: 10.1111/j.1526-4637.2010.00830.x] [Medline: 20353408]
8. Roughan WH, Campos AI, García-Marín LM, Cuéllar-Partida G, Lupton MK, Hickie IB, et al. Comorbid chronic pain and depression: shared risk factors and differential antidepressant effectiveness. Front Psychiatry 2021;12:643609 [FREE Full text] [doi: 10.3389/fpsyt.2021.643609] [Medline: 33912086]
9. Gerrits MMJG, van Oppen P, Leone SS, van Marwijk HWJ, van der Horst HE, Penninx BW. Pain, not chronic disease, is associated with the recurrence of depressive and anxiety disorders. BMC Psychiatry 2014;14:187 [FREE Full text] [doi: 10.1186/1471-244X-14-187] [Medline: 24965597]
10. Gerrits MMJG, Vogelzangs N, van Oppen P, van Marwijk HWJ, van der Horst H, Penninx BWJH. Impact of pain on the course of depressive and anxiety disorders. Pain 2012;153(2):429-436. [doi: 10.1016/j.pain.2011.11.001] [Medline: 22154919]

XSL•FO
RenderX

11. Kamper SJ, Apeldoorn AT, Chiarotto A, Smeets RJEM, Ostelo RWJG, Guzman J, et al. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain: Cochrane systematic review and meta-analysis. BMJ 2015;350:h444 [FREE Full text] [doi: 10.1136/bmj.h444] [Medline: 25694111]
12. Behrens-Wittenberg E, Wedegaertner F. A return-to-work intervention for prematurely retired depression or anxiety disorder patients. Front Psychiatry 2021;12:662158 [FREE Full text] [doi: 10.3389/fpsyt.2021.662158] [Medline: 34149480]
13. Karvonen-Gutierrez CA. The importance of disability as a health issue for mid-life women. Womens Midlife Health 2015;1:10 [FREE Full text] [doi: 10.1186/s40695-015-0011-x] [Medline: 30766697]
14. Building the caregiver workforce our aging world needs. Global Coalition on Aging, Home Instead. 2021. URL: https://globalcoalitiononaging.com/wp-content/uploads/2021/06/GCOA_HI_Building-the-Caregiving-Workforce-Our-Aging-World-Needs_REPORT-FINAL_July-2021.pdf [accessed 2026-07-20]
15. Wolff JL, Nicholas LH, Willink A, Mulcahy J, Davis K, Kasper JD. Medicare spending and the adequacy of support with daily activities in community-living older adults with disability: an observational study. Ann Intern Med 2019;170(12):837-844 [FREE Full text] [doi: 10.7326/M18-2467] [Medline: 31132789]
16. Griffith LE, Raina P, Levasseur M, Sohel N, Payette H, Tuokko H, et al. Functional disability and social participation restriction associated with chronic conditions in middle-aged and older adults. J Epidemiol Community Health 2017;71(4):381-389. [doi: 10.1136/jech-2016-207982] [Medline: 27754857]
17. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. J Med Internet Res 2016;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
18. Dryman MT, McTeague LM, Olino TM, Heimberg RG. Evaluation of an open-access CBT-based internet program for social anxiety: patterns of use, retention, and outcomes. J Consult Clin Psychol 2017;85(10):988-999. [doi: 10.1037/ccp0000232] [Medline: 28650193]
19. Mohr DC, Kwasny MJ, Meyerhoff J, Graham AK, Lattie EG. The effect of depression and anxiety symptom severity on clinical outcomes and app use in digital mental health treatments: meta-regression of three trials. Behav Res Ther 2021;147:103972 [FREE Full text] [doi: 10.1016/j.brat.2021.103972] [Medline: 34600398]
20. Fortuna KL, Torous J, Depp CA, Jimenez DE, Areán PA, Walker R, et al. A future research agenda for digital geriatric mental healthcare. Am J Geriatr Psychiatry 2019;27(11):1277-1285 [FREE Full text] [doi: 10.1016/j.jagp.2019.05.013] [Medline: 31196619]
21. Wang S, Bolling K, Mao W, Reichstadt J, Jeste D, Kim H, et al. Technology to support aging in place: older adults' perspectives. Healthcare (Basel) 2019;7(2):60 [FREE Full text] [doi: 10.3390/healthcare7020060] [Medline: 30974780]
22. Perrin A. Mobile technology and home broadband 2021. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/2021/06/03/mobile-technology-and-home-broadband-2021/ [accessed 2026-07-20]
23. Bhattarai P, Phillips JL. The role of digital health technologies in management of pain in older people: an integrative review. Arch Gerontol Geriatr 2017;68:14-24. [doi: 10.1016/j.archger.2016.08.008] [Medline: 27584871]
24. Schumacher R. Gran got tech: inclusivity and older adults. J Usability Stud 2023;18(2):62-67 [FREE Full text]
25. Prensky M. Digital natives, digital immigrants part 1. On Horiz 2001;9(5):1-6. [doi: 10.1108/10748120110424816]
26. Bartels SJ, DiMilia PR, Fortuna KL, Naslund JA. Integrated care for older adults with serious mental illness and medical comorbidity: evidence-based models and future research directions. Psychiatr Clin North Am 2018;41(1):153-164 [FREE Full text] [doi: 10.1016/j.psc.2017.10.012] [Medline: 29412843]
27. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: a literature based framework (MOLD-US). Int J Med Inform 2018;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
28. Wagner N, Hassanein K, Head M. The impact of age on website usability. Comput Human Behav 2014;37:270-282. [doi: 10.1016/j.chb.2014.05.003]
29. Romano Bergstrom JC, Olmsted-Hawala EL, Jans ME. Age-related differences in eye tracking and usability performance: website usability for older adults. Int J Hum Comput Interact 2013;29(8):541-548. [doi: 10.1080/10447318.2012.728493]
30. Americans check their phones 96 times a day. Asurion. 2019. URL: https://www.asurion.com/press-releases/americans-check-their-phones-96-times-a-day/ [accessed 2026-07-20]
31. Liu N, Yin J, Tan SS, Ngiam KY, Teo HH. Mobile health applications for older adults: a systematic review of interface and persuasive feature design. J Am Med Inform Assoc 2021;28(11):2483-2501 [FREE Full text] [doi: 10.1093/jamia/ocab151] [Medline: 34472601]
32. Gomez-Hernandez M, Ferre X, Moral C, Villalba-Mora E. Design guidelines of mobile apps for older adults: systematic review and thematic analysis. JMIR Mhealth Uhealth 2023;11:e43186 [FREE Full text] [doi: 10.2196/43186] [Medline: 37733401]
33. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018;47(1):1-18 [FREE Full text] [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
34. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]

XSL•FO
RenderX

35. Mojtabai R, Olfson M, Sampson NA, Jin R, Druss B, Wang PS, et al. Barriers to mental health treatment: results from the National Comorbidity Survey Replication. Psychol Med 2011;41(8):1751-1761 [FREE Full text] [doi: 10.1017/S0033291710002291] [Medline: 21134315]
36. Schure M, McCrory B, Tuchscherer Franklin K, Greist J, Weissman RS. Twelve-month follow-up to a fully automated internet-based cognitive behavior therapy intervention for rural adults with depression symptoms: single-arm longitudinal study. J Med Internet Res 2020;22(10):e21336 [FREE Full text] [doi: 10.2196/21336] [Medline: 33006561]
37. Carlo AD, Hosseini Ghomi R, Renn BN, Strong MA, Areán PA. Assessment of real-world use of behavioral health mobile applications by a novel stickiness metric. JAMA Netw Open 2020;3(8):e2011978 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.11978] [Medline: 32744628]
38. Carlo AD, Hosseini Ghomi R, Renn BN, Areán PA. By the numbers: ratings and utilization of behavioral health mobile applications. NPJ Digit Med 2019;2:54 [FREE Full text] [doi: 10.1038/s41746-019-0129-6] [Medline: 31304400]
39. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883] [Medline: 34505369]
40. Torous J, Linardon J, Goldberg SB, Sun S, Bell I, Nicholas J, et al. The evolving field of digital mental health: current evidence and implementation issues for smartphone apps, generative artificial intelligence, and virtual reality. World Psychiatry 2025;24(2):156-174 [FREE Full text] [doi: 10.1002/wps.21299] [Medline: 40371757]
41. Abd-Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M. Perceptions and opinions of patients about mental health chatbots: scoping review. J Med Internet Res 2021;23(1):e17828 [FREE Full text] [doi: 10.2196/17828] [Medline: 33439133]
42. Badr J, Motulsky A, Denis J. Digital health technologies and inequalities: a scoping review of potential impacts and policy recommendations. Health Policy 2024;146:105122 [FREE Full text] [doi: 10.1016/j.healthpol.2024.105122] [Medline: 38986333]
43. Cheng AL, Abraham J, Hartz SM, Laber EB, Miller JP. Evaluating and optimizing just-in-time adaptive interventions in a digital mental health intervention (Wysa for Chronic Pain) for middle-aged and older adults with chronic pain: protocol for a series of randomized trials. JMIR Res Protoc 2025;14:e77532 [FREE Full text] [doi: 10.2196/77532] [Medline: 40961490]
44. Lyon AR, Munson SA, Renn BN, Atkins DC, Pullmann MD, Friedman E, et al. Use of human-centered design to improve implementation of evidence-based psychotherapies in low-resource communities: protocol for studies applying a framework to assess usability . JMIR Res Protoc 2019;8(10):e14990 [FREE Full text] [doi: 10.2196/14990] [Medline: 31599736]
45. Beaton DE, Clark JP. Qualitative research: a review of methods with use of examples from the total knee replacement literature. J Bone Joint Surg Am 2009;91 Suppl 3:107-112. [doi: 10.2106/JBJS.H.01631] [Medline: 19411508]
46. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2008;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
47. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
48. Cheng AL, Gou CY, Martin A, Hartz SM, Abraham J. Optimizing usability of digital health interventions for non-digitally native adults: a framework-based approach to develop just-in-time adaptive interventions (JITAIs) and other adaptations. JMIR Hum Factors (Forthcoming) 2026:e87365. [doi: 10.2196/87358]
49. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
50. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
51. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap Consortium: building an international community of software platform partners. J Biomed Inform 2019;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
52. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
53. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
54. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
55. Rev. URL: https://www.rev.com/ [accessed 2026-07-26]
56. Ahmed SK, Mohammed RA, Nashwan AJ, Ibrahim RH, Abdalla AQ, M. Ameen BM, et al. Using thematic analysis in qualitative research. J Med Surg Public Health 2025;6:100198. [doi: 10.1016/j.glmedi.2025.100198]

XSL•FO
RenderX

57. Colbenson GA, Cook DA, Stephenson CR. Learner engagement, teaching effectiveness, and digital proficiency in in-person versus livestream continuing medical education. J Contin Educ Health Prof 2026;46(2):88-94. [doi: 10.1097/CEH.0000000000000622] [Medline: 41143745]
58. Cleeland CS, Ryan KM. Pain assessment: global use of the brief pain inventory. Ann Acad Med Singap 1994;23(2):129-138. [Medline: 8080219]
59. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
60. Bijkerk LE, Spigt M, Oenema A, Geschwind N. Engagement with mental health and health behavior change interventions: an integrative review of key concepts. J Contextual Behav Sci 2024;32:100748. [doi: 10.1016/j.jcbs.2024.100748]
61. Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 2018;77:120-132 [FREE Full text] [doi: 10.1016/j.jbi.2017.12.008] [Medline: 29248628]
62. Zarate D, Stavropoulos V, Ball M, de Sena Collier G, Jacobson NC. Exploring the digital footprint of depression: a PRISMA systematic literature review of the empirical evidence. BMC Psychiatry 2022;22(1):421 [FREE Full text] [doi: 10.1186/s12888-022-04013-y] [Medline: 35733121]
63. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022;24(5):e35371 [FREE Full text] [doi: 10.2196/35371] [Medline: 35612886]
64. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
65. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. JMIR Mhealth Uhealth 2019;7(4):e11223 [FREE Full text] [doi: 10.2196/11223] [Medline: 30990458]
66. Fonseca das Neves J, Kornacka M, Serra E, Rollin N, Kosinski T, Maréchal V, et al. The impact of rumination on fibromyalgia pain after physical activity: an experimental study. Sci Rep 2023;13(1):20523 [FREE Full text] [doi: 10.1038/s41598-023-47414-z] [Medline: 37993555]
67. Vertsberger D, Talmon A, Ziadni M, Kong J, Darnall BD, Manber R, et al. Intensity of chronic low back pain and activity interference: a daily diary study of the moderating role of cognitive pain coping strategies. Pain Med 2023;24(4):442-450 [FREE Full text] [doi: 10.1093/pm/pnac151] [Medline: 36214626]
68. Grau LE, Larkin K, Lalloo C, Stinson JN, Zempsky WT, Ball SA, et al. Perspectives on adapting a mobile application for pain self-management in neurofibromatosis type 1: results of online focus group discussions with individuals living with neurofibromatosis type 1 and pain management experts. BMJ Open 2022;12(7):e056692 [FREE Full text] [doi: 10.1136/bmjopen-2021-056692] [Medline: 35840301]
69. Shah N, Borrelli B, Kumar D. Perceptions about smartphone-based interventions to promote physical activity in inactive adults with knee pain - a qualitative study. Disabil Rehabil Assist Technol 2024;19(6):2221-2228. [doi: 10.1080/17483107.2023.2272854] [Medline: 37873670]
70. Bostrøm K, Varsi C, Eide H, Børøsund E, Kristjansdottir Ó, Schreurs KMG, et al. Engaging with EPIO, a digital pain self-management program: a qualitative study. BMC Health Serv Res 2022;22(1):577 [FREE Full text] [doi: 10.1186/s12913-022-07963-x] [Medline: 35488295]
71. Cheng AL, Leo AJ, Calfee RP, Dy CJ, Armbrecht MA, Abraham J. What are orthopaedic patients' and clinical team members' perspectives regarding whether and how to address mental health in the orthopaedic care setting? A qualitative investigation of patients with neck or back pain. Clin Orthop Relat Res 2023;481(7):1415-1429 [FREE Full text] [doi: 10.1097/CORR.0000000000002513] [Medline: 36480637]
72. Digital mental health intervention for middle-aged and older adults with depression and/or anxiety and coexisting chronic pain. National Institute of Mental Health Data Archive. URL: https://doi.org/10.15154/6x0x-5b27 [accessed 2026-07-26]

Abbreviations

BIT: Behavioral Intervention Technology
CFIR: Consolidated Framework for Implementation Research
COREQ: Consolidated Criteria for Reporting Qualitative Research
DDBT: Discover, Design/Build, and Test
GAD-7: 7-item Generalized Anxiety Disorder
IRB: institutional review board
PHQ-9: 9-item Patient Health Questionnaire

XSL•FO
RenderX

Edited by A Stone; submitted 07.Nov.2025; peer-reviewed by Z Zhang, M Hirschi; comments to author 24.Jun.2026; accepted 20.Jul.2026; published 11.Aug.2026.

Please cite as:
Cheng AL, Gou CY, Martin A, Hartz SM, Abraham J
Design Preferences for Mental Health Apps Among Nondigitally Native Adults With Chronic Pain: Qualitative Analysis
JMIR Mhealth Uhealth 2026;14:e87358
URL: https://mhealth.jmir.org/2026/1/e87358
doi:10.2196/87358
PMID:

©Abby L Cheng, Christine Y Gou, Adriana Martin, Sarah M Hartz, Joanna Abraham. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Accessing Germany's Electronic Patient Record Through Health Insurer–Provided Apps: Qualitative Interview Study

Susann May[1,2], Dr rer medic, MPH; Manuela Marquardt[3], MA; Gina Barzen[1,4], Dr med; Frances Seifert[1], MSc; Mary Nyarko[1], MPH; Antonia Niemeyer[1], MPH; Sonja Gscheidle[1], MSc; Liane Schenk[3], Prof Dr; Sebastian Spethmann[1,4*], Prof Dr; Felix Muehlensiepen[1,2*], PD, Dr rer medic

[1]German Heart Center at Charité – University Medical Center Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Berlin, Germany

[2]User Experience in Digital Health Lab, Center for Health Services Research, Brandenburg Medical School Theodor Fontane, Rüdersdorf, Germany

[3]Institute of Medical Sociology and Rehabilitation Science, Charité - Universitätsmedizin Berlin, Berlin, Berlin, Germany

[4]DZHK (German Centre for Cardiovascular Research), Berlin, Germany

[*]these authors contributed equally

Corresponding Author:
Susann May, Dr rer medic, MPH
German Heart Center at Charité – University Medical Center Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin
Chariteplatz 1
Berlin, Berlin, 10117
Germany
Phone: 49 30 450 513
Email: susann.may@dhzc-charite.de

Abstract

Background: Nationwide electronic health records (EHRs) are intended to strengthen patient empowerment and improve information continuity across health care settings. In Germany, the electronic patient record (elektronische Patientenakte [ePA]) transitioned from an opt-in to an opt-out model in 2025. Initial survey data suggest that technical complexity and access requirements may hinder engagement; yet, little is known about how users experience this access process.

Objective: This study explored users' experiences of setting up and authenticating access to insurer-provided ePA apps and examined the barriers affecting successful access.

Methods: We conducted a qualitative interview study with individuals covered by statutory health insurance. From 61 interviews conducted within the broader ePA4all project, 23 were selected because participants had attempted to access the ePA through an insurer-provided application and described at least 1 concrete action or event during setup or authentication. Semistructured telephone interviews were conducted between August and December 2025. Data were analyzed using Kuckartz's structured qualitative content analysis. The process phases were developed inductively from the material. All 23 interviews were independently coded by 2 researchers, with differences resolved through consensus. A second analytic step examined how barriers interacted and accumulated across the access pathway.

Results: The sample included 21 participants who had successfully accessed the ePA and 2 who had initiated but abandoned setup or authentication. Setup and authentication were experienced not as a single technical step but as an interdependent process comprising 5 phases: orientation, technical entry, verification, interruption, and transition to initial use. During orientation, incomplete or outdated information left requirements and procedural steps unclear. Technical entry required users to coordinate different apps, interfaces, and device requirements. Verification involved multiple credentials, identification services, and partly repeated security steps. Technical failures, administrative inconsistencies, postal procedures, and waiting periods interrupted progression and sometimes required users to repeat previously completed steps. After successful access, an unclear immediate benefit could limit the transition to initial use. Across phases, participants relied on digital competence, prior experience, persistence, and self-efficacy to navigate the process. Difficulties were also reported by participants with high affinity for technology. Barriers could accumulate across phases: insufficient orientation increased subsequent coordination work, failures produced interruptions and repeated attempts, and prolonged effort contributed to frustration or disengagement.

XSL•FO
RenderX

Conclusions: Access to the opt-out ePA through insurer-provided apps is shaped by cumulative demands rather than by isolated technical problems. Formal provision of an EHR does not ensure that users can successfully complete the access process. Implementation should prioritize clear preregistration guidance, continuity across applications and identification services, resumable procedures, actionable error messages, and accessible support. More broadly, onboarding and authentication should be treated as integral stages in the implementation of patient-facing digital health infrastructures rather than as neutral technical prerequisites.

Trial Registration: German Clinical Trials Register DRKS00037053; https://www.drks.de/search/de/trial/DRKS00037053

International Registered Report Identifier (IRRID): RR2-10.1177/20552076251407130

(JMIR Mhealth Uhealth 2026;14:e95294) doi:10.2196/95294

KEYWORDS

accessibility; digital health; EHR; electronic health record; health; patient empowerment; qualitative interviews

Introduction

As part of the digital transformation within health care systems, implementation of patient-centered health technologies into treatment pathways is increasing. These developments include the global introduction of nationwide electronic health records (EHRs) aimed at empowering individuals to manage their own health and health data more competently. EHRs are intended to improve the availability of health information to patients and encourage their active participation in care processes [1], thereby contributing to improving the quality of medical care and helping health care systems meet future challenges [1]. Patient empowerment constitutes a central guiding principle of these digital health initiatives, aiming to strengthen patients' self-efficacy and confidence in managing their health and navigating interactions with health care professionals (HCPs) [1-4].

Despite these intended benefits, the acceptance and active use of national EHR systems cannot be taken for granted. Data protection, data security, and trust in the transparent handling of health data are key determinants of the acceptance or rejection of national EHR systems [5,6]. Concerns about unauthorized access and data breaches represent some of the most frequently reported barriers [7,8]. Furthermore, usability challenges, such as complex interfaces, unclear navigation, and limited perceived value or benefit, reduce individuals' motivation to use digital health records [9].

International evidence from patient portals indicates that formal availability does not automatically translate into successful enrollment and access. In a survey of patients who had not enrolled in a portal, almost two-thirds had not attempted enrollment, most commonly because of insufficient information or motivation; negative attitudes and computer-related obstacles were additional reasons [10]. A qualitative study of the US Veterans Health Administration's national patient portal further showed that its multistep enrollment pathway, including online registration, in-person identity verification, and additional activation steps, created administrative and logistical barriers for patients [11]. These findings suggest that onboarding and authentication are not merely preliminary technical tasks but implementation stages that can determine whether patient-facing digital health infrastructures are reached in practice. However, the interaction and accumulation of barriers across these stages remain insufficiently understood.

In Germany, the electronic patient record (elektronische Patientenakte [ePA]) was introduced in 2021 [12]. At that time, individuals who had statutory health insurance could decide whether they wanted to apply for and use an ePA provided by their respective health insurance fund. In January 2025, the system transitioned from an opt-in to an opt-out model [13]. Since then, an ePA has been automatically created for all individuals covered by statutory health insurance, with the aim of enabling the storage and exchange of medical documents, facilitating cross-sectoral information flow, and strengthening patients' roles within the care process [14]. Although the ePA is established automatically, its use remains voluntary. Individuals who do not wish to use the ePA may actively object to its use. Given the central role of statutory health insurance within the German health system, this regulatory change has broad practical relevance. Approximately 75 million of Germany's 83.5 million inhabitants are covered by statutory health insurance, and 93 independent statutory health insurance funds (as of January 2026) [15] operate autonomously within a common legal framework [16]. Each statutory health insurer provides its members with its own ePA application. Access to the ePA is provided through apps offered by the respective statutory health insurers. Consequently, the ePA constitutes a national infrastructure, whereas the interfaces and parts of the setup process are implemented through multiple insurer-provided apps. While functional and technical requirements are defined by law and embedded within the national telematics infrastructure [17], differences remain in interface design, user guidance, and the specific implementation of the registration process. Consequently, access to the ePA may vary depending on the respective health insurer.

Initial findings on the use of the ePA indicate that it has thus far been used only to a limited extent. In a survey conducted in 2025, a total of 1500 individuals covered by statutory health insurance were asked about their use of the ePA. Twelve percent of respondents reported having actively set up their ePA and used it at least once. In contrast, 79% were classified as passive users who had not independently accessed their ePA to date, while 7% stated that they had objected to its use [18]. Passive users were further queried about their reasons for nonuse. Eleven percent selected the following response: too complicated, inconvenient, application does not work, no/old mobile phone [18]. These findings suggest that the uptake of the ePA may, at least in part, be hindered by challenges related to access requirements and technical prerequisites. In particular, these

XSL•FO
RenderX

findings point to the importance of the initial setup and authentication process as a potential barrier to use. However, this process has so far received little systematic attention.

The setup of the ePA application is a complex process involving multiple technical and organizational requirements. In terms of actor-network theory, ePA authentication can be conceptualized as an obligatory passage point (OPP) [19]: as a necessary bottleneck, it bundles the interactions of heterogeneous actors (patient, smartphone, ID card, electronic ID [eID], application ecosystem, network connection, and backend infrastructure) and makes access to the ePA dependent on these elements being successfully aligned in a multistage sequence. A disruption in one component may therefore interrupt the entire access pathway, even when all other requirements have been met. From this perspective, authentication requires users not only to operate an application but also to coordinate technical systems, identification procedures, and organizational requirements. From this perspective, access to the ePA is not a simple technical step but a sociotechnical process that may generate barriers at different stages. Despite this conceptualization, empirical research has so far paid little attention to how users actually experience this process.

To date, there have been no studies on the subjective experiences of individuals during the ePA application setup process, mainly due to the recent introduction of the opt-out model in Germany. As a result, there is currently little insight into how users experience the registration process or at which specific stages problems may arise. Previous studies have primarily focused on the usability of digital health technologies themselves rather than on the processes of accessing or setting up these systems [8,20]. However, the effective use of the ePA presupposes that access to the ePA application is both available and manageable for users. Against this background, a deeper understanding of user experiences across the entire registration process is essential. By systematically examining challenges along this process, critical points that may hinder successful access can be identified. This study therefore explored how individuals with statutory health insurance experienced setup and authentication for their insurer-provided ePA application and which barriers affected successful access to the ePA. Initial use after authentication was considered only insofar as participants described the immediate transition from formal access to engagement with the application.

Methods

Ethical Considerations

All ethical issues were addressed. All study procedures were approved by the responsible ethics committee. All methods were performed in accordance with relevant guidelines and regulations. This study involves human participants and was approved by the Ethics Committee of the Charité Berlin (reference ID: EA1/333/24). Participants were informed verbally and in writing about the purpose, procedure, and significance of the study, as well as the associated benefits and risks, and were given the opportunity to ask questions. They were also informed that they had the right to withdraw their consent to participate in the study at any time, either verbally or in writing, without giving reasons. They were also informed that personal data would be collected and stored, whereby the data would be anonymized, but no data would be published that could be used to identify the individual. For security reasons, the data received from participants were always stored in a password-protected folder on a secure desktop computer. Written consent was obtained after participants had the opportunity to ask questions. Patients were not involved in the design of this study. Participants were offered €50 (approximately US $59.48) as an incentive for their participation in the study.

Overview

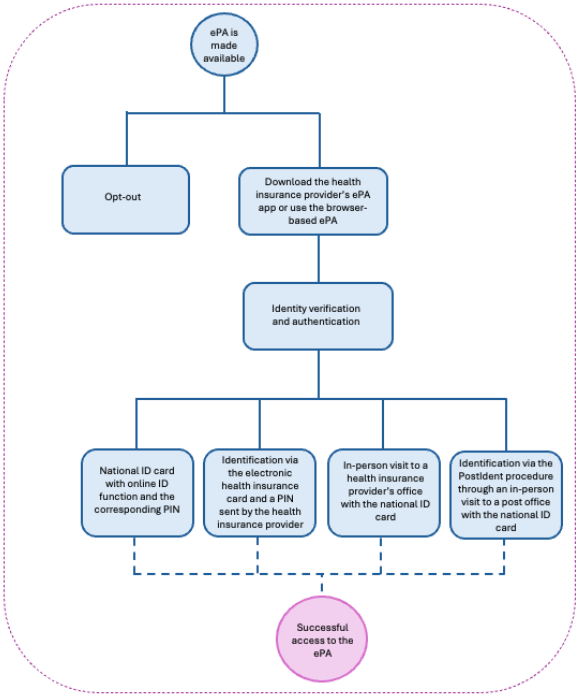
The aim of the study was to investigate how individuals with statutory health insurance experienced setup and authentication for their insurer-provided ePA application in Germany. The study focused on the process from the users' perspective, including perceived barriers and challenges from initial orientation to successful access to the ePA. Experiences after authentication were included only when they concerned the immediate transition to initial use. This study is part of the ePA4all project [21], which examines facilitators and barriers to the adoption of the ePA in Germany.

Accessing the ePA Through Insurer-Provided Applications

Active access to the ePA is primarily provided through an application offered by the individual's statutory health insurer and requires a multistep setup and authentication process. Alternatively, access via a stationary device (desktop computer) is possible if no smartphone or tablet is used. Insured individuals may grant access rights to a representative (eg, a family member). The authorized representative can view and manage the ePA on behalf of the insured person via the application. Initial access requires identity verification. Verification can be completed either via the electronic identity card in combination with a PIN or via the electronic health card in combination with a personal identification number (PIN). The PIN for the electronic health card must be requested from the health insurer. Issuance of the PIN requires prior identity verification, which can be conducted either in person at a branch office of the health insurer or through a postal identification procedure (PostIdent). In addition to the electronic health card, insured individuals may apply for a digital identity (GesundheitsID) through their health insurer. Following the application, authentication via electronic identity card or electronic health card and PIN is likewise required. After successful initial registration, simplified authentication methods may be used for subsequent logins, such as biometric procedures (eg, facial recognition), provided the device supports these functions. Available authentication options may vary depending on the device and operating system (Figure 1). The ePA is also established for children and adolescents. Until the age of 15 years, the ePA is managed by the parents or another legal guardian, who may also object to its use on behalf of the child. Upon reaching the age of 15 years, adolescents may independently manage their electronic patient record and may also exercise their own right to object [22].

XSL•FO
RenderX

Figure 1. Depiction of the elektronische Patientenakte (ePA) access and authentication process (own illustration). PIN: personal identification number; PostIdent: postal identification procedure.



Participants

Participants were recruited with the support of regional Associations of Statutory Health Insurance Physicians in Germany, including the Association of Statutory Health Insurance Physicians in Westphalia-Lippe, the Association of Statutory Health Insurance Physicians in Schleswig-Holstein, the Association of Statutory Health Insurance Physicians in Lower Saxony, and the Association of Statutory Health Insurance Physicians in Berlin. These associations contacted general practitioner (GP) practices within their respective regions and invited them to support the recruitment process. Participating GP practices received study flyers and written information materials describing the aim, procedures, and inclusion criteria of the study. The materials were made available to patients in the practices. Persons who were interested in participating were invited to contact the research team directly. Registration for study participation was possible via email or telephone. Participants were selected using purposive sampling [23]. Individuals were eligible to participate in the study if they met the following inclusion criteria:

- Legally insured under the German health insurance system
- Aged 18 years or older
- Experience as an active ePA user (defined as individuals who had successfully completed authentication and accessed the ePA through their insurer-provided application at least once) or passive ePA user (defined as individuals for whom an ePA had been created but who did not actively use the application)
- Intention to access the ePA through an insurer-provided application
- Willingness to participate in a qualitative interview

In total, 61 interviews were conducted within the broader ePA4all project [21]. All 61 interviews from the broader ePA4all dataset were screened for first-hand accounts of setup and authentication for insurer-provided ePA apps. Interviews were included if participants had personally attempted to access the ePA through an insurer-provided application and described at least 1 concrete action or event during this process. Based on this criterion, 23 interviews were selected for in-depth analysis. The remaining 38 interviews were excluded because participants had not personally attempted to access the ePA through an insurer-provided application or did not describe a concrete action or event during setup or authentication; no other exclusion reasons were applied. Of the 23 included participants, 21 were active users. The 2 passive users had initiated setup or authentication but abandoned the process before successfully accessing the ePA.

Research Team and Reflexivity

The interview team had academic backgrounds in health sciences and sociology. SM and FM had extensive experience in conducting and analyzing qualitative interviews. All interviewers had prior research experience in digital health and the German ePA, which may have sensitized them to implementation-related barriers and challenges. To reflect on the potential influence of these prior perspectives, interpretations and coding decisions were discussed within the research team and documented in analytic memos. None of the interviewers had a prior personal or clinical relationship with the participants. Transcripts and findings were not returned to participants because participant validation was not included in the study design. Instead, the transparency and dependability of the interpretation were supported through independent coding by 2 researchers and subsequent consensus discussions.

XSL•FO
RenderX

Data Collection

Data were collected through semistructured qualitative interviews. An interview guide was developed based on the study objectives and existing research on EHRs and was refined iteratively during data collection. The interview guide addressed experiences with the setup of and authentication for insurer-provided ePA apps, perceived barriers, and the immediate transition to initial use or nonuse after access (Multimedia Appendix 1). The interview guide was piloted with the first 3 participants, and no revisions were required because the questions were understood and elicited relevant accounts of setup and authentication. The interview guide therefore remained unchanged during subsequent data collection. Interviews were conducted by researchers (SM, FM, AN, MN, and FS) via telephone between August and December 2025. All interviews were conducted in German, audio-recorded with informed consent, and transcribed verbatim. Transcripts were anonymized prior to analysis. In addition, brief field notes were documented after each interview to capture contextual impressions and support interpretation of the interview situation; these notes were not included in the formal analysis. Saturation was assessed during the in-depth analysis of the 23 interviews included in the analytic subsample rather than used as a stopping criterion for data collection in the broader ePA4all study.

Data Analysis

Qualitative analysis of the interviews was performed iteratively by SM and FM based on structured qualitative content analysis by Kuckartz [24] using MAXQDA Analytics Pro 2022 (version 22.1.0; Verbi GmbH). During initial familiarization with and coding of the broader dataset, all 61 interviews were screened for relevance to the present research question. The analytic subsample was selected before the subsequent in-depth analysis of the setup and authentication process. During initial coding of the 23 included interviews, recurring topics were grouped into inductively developed main categories and subcategories. The 5 process phases were developed from the coded material and were not predefined by the structure of the interview guide. The coding system was iteratively reviewed and revised by comparing coded passages within and across interviews. Category definitions, coding rules, and illustrative examples were documented in a codebook. Analytic decisions, revisions to the category system, and points discussed during consensus meetings were documented in analytic memos. Both researchers (SM and FM) independently coded all 23 interviews using the agreed coding system. Differences in coding or category interpretation were discussed until consensus was reached. Where necessary, category definitions were clarified and the relevant material was recoded. This procedure was intended to strengthen the transparency and dependability of the analysis.

In a second analytic step, recurring patterns of access barriers across the coded material were systematically compared. Beyond phase-specific challenges, the analysis focused on the relationships between successive phases of the access process. Through iterative comparison of cases and categories, we examined how difficulties arising in one phase affected participants' progression through subsequent phases. Particular attention was paid to recurring sequences involving unclear requirements, increased coordination work, technical or administrative failure, interruption, repeated attempts, and disengagement. This second step resulted in a process-oriented interpretation of how barriers could interact and accumulate across the authentication pathway. The concept of authentication as an obligatory passage point served as a sensitizing perspective for examining how users, devices, applications, identification services, and administrative procedures had to be coordinated and how disruptions in one component affected the wider access process.

Within the analytic subsample, saturation was considered reached when the analysis of additional interviews produced no new process phases, categories, or relevant dimensions of the setup and authentication experience [25].

The data were analyzed in German. To present the findings, significant excerpts from the transcriptions were chosen as representative quotes. These quotes were translated into English and incorporated into the manuscript. The translated quotations were reviewed against the original German transcripts by bilingual members of the research team to ensure that their meaning and tone were retained. The manuscript was prepared in accordance with the COREQ (Consolidated Criteria for Reporting Qualitative Research; checklist provided in Multimedia Appendix 2) [26].

Results

Sociodemographic Characteristics

Of the 23 participants, 21 were active ePA users and 2 were passive ePA users. The 2 passive users included in the analysis had initiated setup or authentication but abandoned the process before successfully accessing the ePA. The mean age of the participants was 47.9 (SD 17.3; range 21-80) years. Most participants were male (male: n=14; female: n=9). The mean duration of the interviews was 35 (SD 8.3; range 18-50) minutes. The participants were insured with 11 different health insurance companies. Detailed characteristics of the study participants are shown in Table 1.

XSL•FO
RenderX

Table 1. Sample characteristics and ATI[a] score (computed as the mean of ATI items; scale range 1-6; categorized as low: 1.00-2.66, moderate: 2.67-4.32, and high: 4.33-6.00).

ID	User type; age (years); sex	Population size (number of inhabitants), n	Chronic diseases	Operating system	ATI-S[b] score and interpretation, mean (SD)
ePA_1a	Active; 30-39; male	3.8 million	No	iOS	5.5 (0.58); high
ePA_2a	Active; 20-29; male	60,000	No	iOS	5.0 (1.15); high
ePA_3a	Active; 30-39; female	12,000	Yes	iOS	5.5 (0.58); high
ePA_4a	Active; 50-59; male	12,000	No	Android	5.5 (0.58); high
ePA_5a	Active; 50-59; female	5000	Yes	Android	3.75 (0.96); moderate
ePA_6a	Active; 60-69; male	60,000	No	iOS	5.0 (1.15); high
ePA_7a	Active; 60-69; male	14,000	Yes	Android	6.0 (0.00); high
ePA_8a	Active; 60-69; female	6000	Yes	iOS	4.75 (0.96); high
ePA_9a	Active; 30-39; male	3.8 million	No	iOS	4.75 (1.50); high
ePA_10a	Active; 20-29; male	1 million	Yes	iOS	5.0 (1.15); high
ePA_11a	Active; 40-49; female	3.8 million	No	iOS	5.5 (0.58); high
ePA_12a	Active; 40-49; female	3.8 million	No	iOS	4.75 (0.96); high
ePA_13a	Active; 30-39; female	500,000	No	iOS	3.5 (0.58); moderate
ePA_14a	Active; 50-59; female	30,000	Yes	iOS	4.5 (S0.58); high
ePA_15a	Active; 80-89; male	1000	Yes	iOS	5.25 (0.50); high
ePA_16a	Active; 40-49; female	3000	Yes	iOS	4.5 (0.58); high
ePA_17a	Active; 30-39; male	550,000	No	iOS	5.0 (2.00); high
ePA_18a	Active; 60-69; male	26,000	No	Android	4.5 (1.73); high
ePA_19a	Active; 20-29; male	3.8 million	No	iOS	4.5 (1.73); high
ePA_20a	Active; 20-29; male	600,000	Yes	iOS	4.25 (0.96); high
ePA_21a	Active; 40-49; male	170,000	Yes	Android	5.5 (0.58); high
ePA_1p	Passive; 60-69; female	3.8 million	Yes	Android	3.5 (1.29); moderate
ePA_2p	Passive; 80-89; male	3.8 million	Yes	Android	2.5 (0.58); low

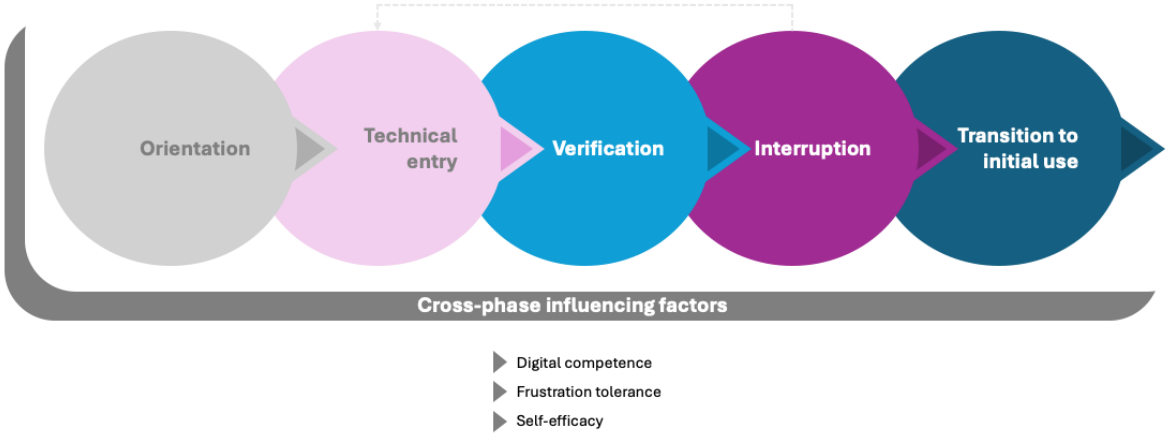
[a]ATI: Affinity for Technology Interaction.

[b]ATI-S: Affinity for Technology Interaction Short Scale.

The setup and authentication required to access the ePA through insurer-provided apps do not take place in a single step but unfold across several consecutive phases. Challenges rarely occur in isolation, but may interact and accumulate throughout the process. In the following, these barriers are traced along the individual process phases. Figure 2 visualizes the multistep access process of the ePA as a sequential process consisting of 5 inductively derived phases: orientation, technical entry, verification, interruption, and transition to initial use.

XSL•FO
RenderX

Figure 2. Summary of the experience of the elektronische Patientenakte (ePA) setup and usage process from the perspective of individuals with statutory health insurance.



Phase 1: Orientation

In the first phase, barriers arise even before the actual registration begins, as potential users do not perceive the process as a clearly plannable and reliable sequence of steps but rather as an uncertain endeavor with unclear requirements and insufficient guidance. The available information is experienced as inconsistent and outdated.

> *Well, there is a video, and the video seems very outdated. (...) Apparently, the health insurance company did not update it.* [ePA_13a_37_female]

The orientation gap does not only concern the timeliness of information but also the coverage of concrete use cases. This becomes particularly evident in special situations, such as setting up the ePA application for children.

> *I had no idea how I was supposed to do that [install the ePA app for my child]. I simply didn't know.* [ePA_13a_37_female]

As a result, entry into the process is not structured by a clear and comprehensible procedure but rather by an unsystematic search for relevant information, in which key courses of action remain unclear (eg, setting up the application for children or other persons). Consequently, the registration process often begins without a clear understanding of which steps are required in which order and what prerequisites need to be met.

Phase 2: Technical Entry

The phase of technical entry is characterized by structural challenges. After downloading the ePA application, entry is experienced as a fragmented sequence of different applications, procedural paths, and decision points, the underlying logic of which is only partially comprehensible to users. The need to switch between various apps, identification options, and technical interfaces generates early feelings of overload and uncertainty about the next step. This experience is described as highly burdensome.

> *At that time, I was really somewhat overwhelmed by the whole process; I experienced it as really very labor-intensive.* [ePA_12a_49_female]

The metaphor of an "authentication jungle" illustrates the perception of a confusing and difficult-to-navigate system.

> *I just got into this authentication jungle.* [ePA_1a_34_male]

Entry into the system is experienced as a process in which users have to orient themselves independently, without being actively supported by the system architecture. This fragmentation does not only concern the number of steps but also their organizational and technical distribution. Entry does not take place via an integrated platform but through a combination of the main application, identification services, and, in some cases, separate service applications.

> *Yes, this whole registration process, right, with [name of health insurance company], you have to download two apps (...) and then you also have to open a second app. (...) Especially since you can't even get into the patient record without opening the other app, so it's basically duplicated.* [ePA_7a_60_male]

As a result, a high level of coordination is required from users right from the beginning, as they must relate various technical and administrative components to one another. A lack of transparency regarding which component fulfills which function intensifies the feeling of being caught in an opaque procedure whose logic is not immediately apparent.

In addition, technical access requirements act as independent barriers that can partially prevent entry altogether. Several participants report that the process failed due to unmet device requirements or security settings.

> *And then it said, you don't have NFC on your device, you can't use this device.* [ePA_2p_80_male]

Such requirements often appear unexpectedly and are usually neither transparently communicated in advance nor easily resolved in the short term, especially when they concern hardware-related or security-relevant settings. Access to the ePA thus depends on the availability of suitable end devices. Participants whose devices did not meet these requirements were unable to proceed with the access process using those devices.

XSL•FO
RenderX

At the same time, some participants report a smooth entry when technical requirements are met and prior experience is available. In these cases, entry is perceived as uncomplicated.

> *Well, everything went smoothly.* [ePA_3a_36_female]

However, these positive experiences are limited to specific constellations and do not represent a systematic characteristic of the procedure.

Phase 3: Verification

Identity verification is not experienced as a single, clearly delineated step but as a complex sequence of multiple, partly redundant security measures, the internal logic of which is only partially comprehensible to users. Over the course of the process, various identification tools, passwords, PINs, one-time codes, and external verification services have to be coordinated. The parallel use of different security mechanisms increases cognitive load and leads to frustration.

> *And overall, there was this constant verification and logging in. And all these codes you get by email. It's just a lot.* [ePA_9a_35_male]

At the same time, uncertainty arises as to whether the current step has been completed correctly and whether it will result in sustainable progress within the process. Identification is thus not experienced as a stable transition point but as a fragile intermediate state that can collapse at any time.

> *I felt like in the story of the hare and the hedgehog. You're constantly being sent back and forth.* [ePA_2p_80_male]

This fragility becomes particularly evident in the frequently reported technical breakdowns during verification. Several participants describe that redirects to external identification services or internal verification modules do not function reliably.

> *Then, in that verification process, the page just didn't load.* [ePA_9a_35_male]

In other cases, the process breaks down completely.

> *The process technically crashed for me, and then I basically had to do it several times.* [ePA_17a_32_male]

Such breakdowns not only terminate the current step but often lead to previously completed verification stages losing their validity. Users are therefore forced to repeat the entire identification process, regardless of how far they had already progressed. The associated uncertainty increases emotional pressure, as each additional attempt is accompanied by the expectation of renewed failure.

In addition to technical disruptions, administrative challenges also act as independent barriers in this phase, for example, due to inconsistencies in personal data.

> *That didn't work either, because somewhere the addresses weren't exactly the same.* [ePA_19a_23_male]

These administrative challenges are experienced as particularly burdensome, as they cannot be overcome through personal technical skills or increased effort.

Phase 4: Interruption

Even after successfully completing individual steps, access to the ePA is often not enabled immediately but delayed due to postponed feedback, postal mailings, or time-limited activations. As a result, the registration process is not experienced as a continuous flow but as a sequence of separate steps that extend over several days or weeks.

> *It wasn't just one day; it was several days, because I (...) still had to wait for an activation letter.* [ePA_17a_32_male]

These temporal interruptions force users to repeatedly resume the process. After each pause, they must reorient themselves, recall previously completed steps, and reconstruct the current status. Repeated interruptions have a direct impact on motivation. The initial willingness to invest effort into the setup diminishes over time when the process stretches across longer periods. The need to re-enter a complex technical procedure after days or weeks is perceived as an additional burden and often results in abandonment.

> *I also failed at some point during authentication and then I just gave up.* [ePA_1a_34_male]

In combination with the previously described technical and administrative barriers, these interruptions contribute substantially to the perception of the setup process as lengthy and disproportionately cumbersome.

> *I experienced that as extremely complicated (...). Very, very rigid.* [ePA_17a_32_male]

Phase 5: Transition to Initial Use

Participants' accounts also provided limited insight into what occurred immediately after successful setup and authentication. Formal access did not necessarily lead to initial or continued engagement with the respective ePA application. Despite completing the access process, some participants reported only sporadic or no subsequent use. In the transition to actual use, it becomes apparent that successfully completing registration and authentication does not automatically lead to sustainable integration of the ePA application into users' everyday lives. Instead, the step from formal access authorization to actual, routinized use often fails to materialize. Despite overcoming multiple hurdles, participants report only sporadic or completely absent use of the application. For these participants, completion of setup marked the end of an effortful access process but not necessarily the beginning of regular application use.

A central reason for this lack of subsequent engagement was the limited perception of an immediate benefit. Many participants are unable to identify a clear added value after successful setup that would justify continued engagement with the ePA.

> *I installed it, but I never used it, because I didn't see any benefit in it.* [ePA_13a_37_female]

The app was not always experienced as an integral part of medical care or personal health management but rather as an optional add-on whose immediate relevance was unclear. Consequently, successful access did not necessarily result in initial or continued engagement with the respective application.

XSL•FO
RenderX

Cross-Phase Influencing Factors

Across all phases, participants drew on individual coping resources to navigate unclear instructions, coordinate multiple technical components, resolve unexpected problems, and resume interrupted procedures. These resources included digital competence, prior experience, persistence, and self-efficacy. This was explicitly reflected upon by participants.

> *If I didn't have a certain persistence or digital affinity, I probably would have given up.* [ePA_6a_67_male]

The registration process was therefore often experienced as an individual problem-solving project without systematic support.

> *I basically muddled through it on my own.* [ePA_12a_49_female]

Accumulation of Barriers Across the Access Process

Comparison across the 5 phases indicated that barriers could become consequential through their accumulation rather than through any single requirement. Unclear or outdated information during orientation meant that some participants entered the process without knowing which applications, documents, devices, or authentication procedures would be required. During technical entry and verification, they then had to coordinate multiple applications, identification services, passwords, PINs, and technical prerequisites. When technical failures, inconsistent administrative data, or postal procedures interrupted progress, previously completed steps sometimes had to be repeated. Each interruption therefore created additional coordination work and required participants to reorient themselves within the process.

These recurring sequences provide a process-oriented explanation of how access could become increasingly burdensome. Technical or administrative problems did not only affect the step in which they occurred; they could also prolong the pathway, require repeated attempts, and contribute to frustration or disengagement. Individual resources such as digital competence, prior experience, persistence, and self-efficacy helped participants navigate these demands but did not prevent technical or organizational failures. The findings therefore locate access difficulties in the alignment of user actions, technical components, and organizational procedures. When these components were not coherently connected, participants had to compensate through additional coordination and problem-solving work.

Variation Across Insurers, Devices, and Authentication Pathways

Participants were insured with 11 different statutory health insurers and used both iOS and Android devices. Cross-case comparison indicated that the central process-related difficulties were not confined to a single insurer or operating system. Unclear guidance, coordination between multiple technical components, repeated verification steps, and interruptions were described across different application and device constellations. No consistent pattern indicating systematically greater difficulties with either iOS or Android was identified. Other barriers were conditional on the authentication route used. Near-field communication (NFC) capability was relevant when electronic identity cards or electronic health cards were used for authentication. The need to request a health card PIN or wait for an activation letter applied only to pathways requiring these credentials. Problems with redirects or switching between apps arose when external identification services or separate authentication apps were involved. Administrative inconsistencies, such as differences in address data, could prevent verification independently of device type. Because individual insurers were represented by only a small number of participants and authentication routes were not distributed systematically across the sample, the study was not designed to compare or rank specific insurers, apps, operating systems, or authentication methods. The findings therefore distinguish between broadly recurring and pathway-specific barriers but do not support conclusions about the relative performance of individual implementations.

Discussion

Principal Findings

This study examined how individuals with statutory health insurance experienced setup and authentication for insurer-provided ePA apps and which challenges affected successful access to the ePA. The results show that setting up the ePA is not perceived as a single, completed step but rather as a sequential process that spans several stages separated in terms of time and organization. The fifth phase, transition to initial use, should be interpreted as an immediate downstream consequence of the access process rather than as an analysis of sustained adoption or integration into everyday practice. Informational, technical, administrative, and temporal difficulties may interact across these stages, requiring users to repeatedly coordinate components, resolve problems, and resume interrupted procedures. The results are consistent with international studies on EHRs, which also report heterogeneous use and complex barriers to access [27].

At the beginning of the access process, participants encountered limitations related to incomplete, unclear, or outdated information. This finding is consistent with international research emphasizing the importance of information for access to patient-facing EHR systems [6,7,28]. Within the present process model, orientation was not merely a preliminary question of awareness. It influenced whether participants understood which applications, documents, technical requirements, and authentication steps would be required. Insufficient orientation could therefore increase the coordination work needed during technical entry and verification and contribute to the accumulation of burdens across subsequent phases.

The inductively identified phases indicate that challenges do not arise from isolated technical problems alone but rather from the interaction between technical systems, organizational structures, and user practices. The ePA registration process comprises several interrelated components: digital infrastructures (eg, applications, authentication, and identity services), organizational actors and procedures (eg, health insurers, identity verification processes, and postal delivery of access credentials), as well as the work required from users to understand, coordinate, and repeatedly navigate these components. When

XSL•FO
RenderX

these components were not coherently aligned, difficulties could extend beyond the phase in which they initially occurred. Unclear requirements increased the coordination required during technical entry, technical or administrative failures interrupted verification, and waiting periods required users to resume the process after temporal gaps. Repeated attempts and prolonged effort could subsequently contribute to frustration and disengagement. The phases should therefore not be understood as independent categories or as a uniformly linear sequence, but as an interdependent process in which difficulties may accumulate and return users to earlier stages. These findings are consistent with insights from implementation research, which demonstrate that the successful introduction of digital technologies depends not only on technical functionality but, to a significant extent, on their organizational embedding and the provision of clear, comprehensible user guidance [29].

Importantly, substantial difficulties were also reported by participants with moderate or high affinity for technology. The findings therefore cannot be explained solely by limited digital competence. Participants' knowledge, prior experience, self-efficacy, and persistence helped them navigate the process, but these resources could not compensate for incompatible devices, malfunctioning applications, inconsistent administrative data, or delayed activation procedures. Access should therefore be understood as emerging from the interaction between users' available resources and the informational, procedural, technical, and organizational demands created by the access process. These findings can be situated within the eHealth Literacy Framework, which conceptualizes engagement with digital health as an interaction between individual capabilities and the characteristics of digital systems and services [30]. The present study does not propose an alternative or extended literacy framework. Rather, it provides a process-oriented account of how this interaction unfolds during setup and authentication. User resources influenced participants' ability to interpret instructions, coordinate applications, solve problems, and resume interrupted procedures. At the same time, process design and technical and organizational conditions determined the demands they encountered and whether progression through the pathway was possible.

Conceptualizing authentication as a multistage obligatory passage point highlights that access depends on the successful alignment of users, devices, applications, identification services, health insurers, and administrative procedures [19]. The findings indicate that these components did not always operate as a coherent access arrangement. Participants had to bridge gaps between them by identifying the correct application, interpreting system feedback, coordinating authentication tools, resolving administrative inconsistencies, or resuming procedures after delays. Technical and administrative disruptions therefore became user-facing access barriers that required additional coordination and problem-solving work. Authentication should consequently be understood not as a neutral technical prerequisite but as a sociotechnical implementation stage that structures whether individuals can actively access a formally available national EHR infrastructure.

The findings raise equity-relevant concerns but do not provide direct evidence of differences in access between population groups. Participants described substantial demands related to digital competence, persistence, problem-solving, and access to compatible devices. Because the sample predominantly comprised motivated participants with moderate or high affinity for technology, most of whom ultimately accessed the ePA successfully, the study cannot determine how these demands affect groups with fewer resources. It is plausible that the identified requirements may be particularly consequential for individuals with fewer digital, material, linguistic, cognitive, or social resources. However, this represents an inference from the observed process demands rather than a direct empirical finding. This interpretation is consistent with international research showing that digital health technologies can reinforce disparities when effective access depends on unequally distributed resources [31-33]. Further research specifically sampling and comparing diverse population groups is needed to examine these equity implications.

Practical Implications

The findings indicate that barriers to ePA access are embedded in the overall architecture and sequencing of the registration process rather than in isolated technical malfunctions. Implementation efforts should therefore focus on reducing cumulative burdens that emerge across stages. The practical implications were structured according to the 5 phases of the access process and the cross-phase factors identified in the analysis. For each phase, 1 priority measure was linked directly to the corresponding empirical finding. An additional cross-phase measure addresses participants' reliance on individual coping resources throughout the pathway (Table 2).

XSL•FO
RenderX

Table 2. Practical implications to support access to the ePA[a] application.

Process phase	Empirical finding	Practical implications
Phase 1: orientation	Participants began registration without a clear understanding of the required applications, documents, PINs[b], device functions, or available authentication routes. Information was sometimes incomplete, outdated, or insufficient for specific situations.	Provide centralized, up-to-date, and pathway-specific guidance, including a preregistration checklist of all technical and administrative requirements.
Phase 2: technical entry	Switching between applications, identification services, and technical interfaces increased coordination work. Device requirements such as NFC[c] capability were not always communicated before the process began.	Minimize transitions between applications and clearly communicate device and system requirements before users enter the authentication pathway.
Phase 3: verification	Multistep verification, technical failures, unclear feedback, and inconsistencies in administrative data prevented users from completing authentication or required repeated attempts.	Simplify verification procedures and provide actionable error messages that identify the problem and explain the next step or available support.
Phase 4: interruption	Postal procedures, waiting periods, and technical disruptions interrupted progression and required users to reconstruct or repeat previously completed steps.	Preserve completed steps and enable users to resume the process after interruptions, supported by clear progress and status information.
Phase 5: transition to initial use	Successful registration did not necessarily lead to initial or continued use when participants could not identify an immediate practical benefit.	Provide a brief postregistration orientation showing relevant initial use cases and the information or functions available in the ePA application.
Cross-phase factors	Across the pathway, participants relied on digital competence, prior experience, persistence, and self-efficacy to understand requirements, resolve problems, and resume interrupted procedures. The process was often experienced as an individual problem-solving task without systematic support.	Provide consistent and easily accessible support throughout the access pathway, including telephone or in-person assistance, to reduce reliance on individual coping resources.

[a]ePA: elektronische Patientenakte.

[b]PIN: personal identification number.

[c]NFC: near-field communication.

Strength and Limitations

To the best of our knowledge, this is the first qualitative study to examine the setup and authentication process of the opt-out ePA in Germany from the perspective of people with statutory health insurance. The study provides detailed insights into process-related burdens that are often missed in survey-based research. The qualitative approach enabled the identification of cumulative barriers and their interaction across successive process phases and provided a nuanced understanding of how structural features of digital implementation influence access and the transition to initial use. At the same time, several limitations warrant consideration. The analysis is based on a purposively selected analytic subsample from a larger dataset, which limits the transferability of the findings beyond the study context. Participation was voluntary, and the sample predominantly comprised motivated participants with moderate or high affinity for technology. Moreover, 21 of the 23 participants had ultimately gained active access to the ePA application. The study may therefore underrepresent individuals who did not attempt registration, abandoned the process at an early stage, had limited digital experience, or lacked access to suitable devices or support. Consequently, the findings are well suited to demonstrating that substantial access difficulties can occur even among technologically confident and predominantly successful users, but they do not support conclusions about the prevalence of dropout or the experiences of socially, linguistically, cognitively, or digitally disadvantaged groups. The results are based on retrospective self-reports and may therefore be influenced by recall bias, particularly regarding the exact sequence of authentication steps, error messages, and waiting periods. The interviews captured participants' subjective reconstruction of the access process rather than direct observations of their interactions with the applications. Furthermore, the study reflects experiences from an early phase of opt-out implementation; subsequent procedural or technical changes may alter user experiences over time. Because the interviews did not systematically examine use over an extended period, the findings do not support conclusions about sustained adoption or the long-term integration of the ePA into everyday health management. Since the interviews were conducted in 2025 and additional or different registration options have been introduced since then, the results reflect the registration procedures in effect at that time.

Conclusion

This study explores access to the opt-out ePA as a multistage process rather than a single technical act. Obstacles may interact and accumulate in the areas of orientation, technical access, verification, interruption, and the transition to use, reflecting the interplay of infrastructure design, organizational processes, and individual coping resources. Formal availability therefore does not ensure straightforward access. More coherent process design, reduced fragmentation, and accessible support are needed to limit the extent to which successful access depends on individual coping resources. The equity implications of these process requirements should be examined in studies specifically designed to include and compare populations with different

XSL•FO
RenderX

levels of digital, material, linguistic, cognitive, and social resources.

Acknowledgments

The authors thank all interview partners and participants for their valuable contributions and time. The authors would like to thank the Associations of Statutory Health Insurance Physicians in Westphalia-Lippe, Schleswig-Holstein, Lower Saxony, and Berlin for their valuable support in facilitating participant recruitment. During manuscript preparation, generative AI was used exclusively for language editing purposes, including translation into English and grammatical correction.

Data Availability

The qualitative interview data generated and analyzed during this study are not publicly available because they contain sensitive information and may permit participant identification. Public sharing would not be compatible with the conditions of the ethical approval and participants' informed consent. Requests concerning access to selected anonymized data may be directed to the corresponding author and will be considered subject to ethical, legal, and data protection requirements. For further questions regarding the reuse of data, please contact the corresponding author.

Funding

This study was funded by the Innovation Committee at the Federal Joint Committee (Innovation Fund; grant number 01VSF24032). The funder had no involvement in the study design, data collection, analysis, interpretation of the findings, or preparation of the manuscript.

Authors' Contributions

Conceptualization: SM, MM, GB, FS, MN, AN, SG, LS, SS, FM
Data curation: SM, MM, FM
Formal analysis: SM, MM, FM
Funding acquisition: SM, SS, FM
Investigation: SM, FS, MN, AN, FM
Methodology: SM, MM, GB, FS, MN, AN, SG, LS, SS, FM
Project administration: SS
Supervision: SM, FM
Validation: SM, FM
Visualization: SM, MM, FM
Writing—original draft: SM, MM, FM
Writing—review and editing: GB, FS, MN, AN, SG, LS, SS

Conflicts of Interest

SS is the editor in chief of *JMIR Cardio* at the time of this publication. SS had no involvement in the editorial review and processing of this manuscript. FM is an associate editor of *JMIR Cardio* at the time of this publication. FM had no involvement in the editorial review and processing of this manuscript.

Multimedia Appendix 1
Interview guide.
[PDF File (Adobe PDF File), 21 KB - mhealth_v14i1e95294_app1.pdf]

Multimedia Appendix 2
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[PDF File (Adobe PDF File), 149 KB - mhealth_v14i1e95294_app2.pdf]

References

1. Alomar D, Almashmoum M, Eleftheriou I, Whelan P, Ainsworth J. The impact of patient access to electronic health records on health care engagement: systematic review. J Med Internet Res 2024;26:e56473 [FREE Full text] [doi: 10.2196/56473] [Medline: 39566058]
2. Zarcadoolas C, Vaughon WL, Czaja SJ, Levy J, Rockoff ML. Consumers' perceptions of patient-accessible electronic medical records. J Med Internet Res 2013;15(8):e168 [FREE Full text] [doi: 10.2196/jmir.2507] [Medline: 23978618]
3. Gee PM, Paterniti DA, Ward D, Soederberg Miller LM. e-Patients perceptions of using personal health records for self-management support of chronic illness. Comput Inform Nurs 2015;33(6):229-237. [doi: 10.1097/CIN.0000000000000151] [Medline: 25899440]

XSL•FO
RenderX

4. Rappaport J. Terms of empowerment/exemplars of prevention: toward a theory for community psychology. Am J Community Psychol 1987;15(2):121-148. [doi: 10.1007/BF00919275] [Medline: 3604997]
5. Hollo Z, Martin DE. An equitable approach to enhancing the privacy of consumer information on in Australia. Health Inf Manag 2023;52(1):37-40. [doi: 10.1177/18333583211019764] [Medline: 34132130]
6. Holt M, MacGibbon J, Smith AKJ, Broady TR, Davis MDM, Newman CE. Knowledge of Australia's my health record and factors associated with opting out: results from a national survey of the Australian general population and communities affected by HIV and sexually transmissible infections. PLOS Digit Health 2023;2(3):e0000200 [FREE Full text] [doi: 10.1371/journal.pdig.0000200] [Medline: 36857326]
7. Tapuria A, Porat T, Kalra D, Dsouza G, Xiaohui S, Curcin V. Impact of patient access to their electronic health record: systematic review. Inform Health Soc Care 2021;46(2):192-204 [FREE Full text] [doi: 10.1080/17538157.2021.1879810] [Medline: 33840342]
8. Son EH, Nahm ES. Adult patients' experiences of using a patient portal with a focus on perceived benefits and difficulties, and perceptions on privacy and security: qualitative descriptive study. JMIR Hum Factors 2023;10:e46044 [FREE Full text] [doi: 10.2196/46044] [Medline: 37490316]
9. Carini E, Villani L, Pezzullo AM, Gentili A, Barbara A, Ricciardi W, et al. The impact of digital patient portals on health outcomes, system efficiency, and patient attitudes: updated systematic literature review. J Med Internet Res 2021;23(9):e26189 [FREE Full text] [doi: 10.2196/26189] [Medline: 34494966]
10. Goel MS, Brown TL, Williams A, Cooper AJ, Hasnain-Wynia R, Baker DW. Patient reported barriers to enrolling in a patient portal. J Am Med Inform Assoc 2011;18 Suppl 1(Suppl 1):i8-12 [FREE Full text] [doi: 10.1136/amiajnl-2011-000473] [Medline: 22071530]
11. Fix GM, Hogan TP, Amante DJ, McInnes DK, Nazi KM, Simon SR. Encouraging patient portal use in the patient-centered medical home: three stakeholder perspectives. J Med Internet Res 2016;18(11):e308 [FREE Full text] [doi: 10.2196/jmir.6488] [Medline: 27876686]
12. Patientendaten-Schutz-Gesetz (PDSG). Bundesministerium für Gesundheit. URL: https://www.bundesgesundheitsministerium.de/patientendaten-schutz-gesetz.html [accessed 2026-08-26]
13. Gesetz zur Beschleunigung der Digitalisierung des Gesundheitswesens (Digital-Gesetz—DigiG). Bundesministerium der Justiz. URL: https://www.recht.bund.de [accessed 2026-01-27]
14. Die elektronische Patientenakte für alle. Bundesministerium für Gesundheit. URL: https://gesund.bund.de/en/the-electronic-patient-record [accessed 2026-01-27]
15. Die gesetzlichen Krankenkassen. GKV-Spitzenverband. URL: https://www.gkv-spitzenverband.de/krankenversicherung/kv_grundprinzipien/alle_gesetzlichen_krankenkassen/alle_gesetzlichen_krankenkassen.jsp [accessed 2026-01-06]
16. Das Prinzip der Selbstverwaltung. Bundesministerium für Gesundheit. URL: https://www.bundesgesundheitsministerium.de/gesundheitswesen-selbstverwaltung.html [accessed 2026-01-27]
17. ePA für alle. gematik GmbH. URL: https://www.gematik.de/anwendungen/epa-fuer-alle [accessed 2026-07-27]
18. Bevölkerungsbefragung zur elektronischen Patientenakte (ePA). Bundesbeauftragte für den Datenschutz und die Informationsfreiheit. 2025. URL: https://www.bfdi.bund.de/SharedDocs/Downloads/DE/Themen/Datenbarometer/ePA-Bericht.pdf [accessed 2026-01-27]
19. Callon M. Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. Sociol Rev 1984;32(1_suppl):196-233. [doi: 10.1111/j.1467-954x.1984.tb00113.x]
20. Lyles CR, Nelson EC, Frampton S, Dykes PC, Cemballi AG, Sarkar U. Using electronic health record portals to improve patient engagement: research priorities and best practices. Ann Intern Med 2020;172(11 Suppl):S123-S129 [FREE Full text] [doi: 10.7326/M19-0876] [Medline: 32479176]
21. May S, Muehlensiepen F, Barzen G, Seifert F, Marquardt M, Bunar C, et al. ePA4all-exploring facilitators and barriers to the adoption of the electronic health record in Germany: a study protocol for a mixed-methods analysis. Digit Health 2026;12:20552076251407130 [FREE Full text] [doi: 10.1177/20552076251407130] [Medline: 41732184]
22. Die elektronische Patientenakte. Bundesministerium für Gesundheit. URL: https://gesund.bund.de/die-elektronische-patientenakte#nutzung-der-epa [accessed 2026-01-27]
23. Etikan I, Musa SA, Alkassim RS. Comparison of convenience sampling and purposive sampling. Am J Theor Appl Stat 2016;5(1):1-4. [doi: 10.11648/j.ajtas.20160501.11]
24. Kuckartz U. Qualitative Inhaltsanalyse. 4th Edition. Weinheim, Germany: Beltz Juventa; 2018.
25. Hennink MM, Kaiser BN, Marconi VC. Code saturation versus meaning saturation: how many interviews are enough? Qual Health Res 2017;27(4):591-608 [FREE Full text] [doi: 10.1177/1049732316665344] [Medline: 27670770]
26. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
27. Tsai CH, Eghdam A, Davoody N, Wright G, Flowerday S, Koch S. Effects of electronic health record implementation and barriers to adoption and use: a scoping review and qualitative analysis of the content. Life (Basel) 2020;10(12):327 [FREE Full text] [doi: 10.3390/life10120327] [Medline: 33291615]
28. Powell KR. Patient-perceived facilitators of and barriers to electronic portal use: a systematic review. Comput Inform Nurs 2017;35(11):565-573. [doi: 10.1097/CIN.0000000000000377] [Medline: 28723832]

XSL•FO
RenderX

29. Fennelly O, Cunningham C, Grogan L, Cronin H, O'Shea C, Roche M, et al. Successfully implementing a national electronic health record: a rapid umbrella review. Int J Med Inform 2020;144:104281 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104281] [Medline: 33017724]
30. Norgaard O, Furstrand D, Klokker L, Karnoe A, Batterham R, Kayser L, et al. The e-health literacy framework: a conceptual framework for characterizing e-health users and their interaction with e-health systems. Knowl Manag E-Learn 2015;7(4):522-540. [doi: 10.34105/j.kmel.2015.07.035]
31. Yao R, Zhang W, Evans R, Cao G, Rui T, Shen L. Inequities in health care services caused by the adoption of digital health technologies: scoping review. J Med Internet Res 2022;24(3):e34144 [FREE Full text] [doi: 10.2196/34144] [Medline: 35311682]
32. Vesperi W, Ventura M, Cristofaro CL, Melina AM. Digital evolution and emerging inequalities in healthcare: a scoping review through the lens of knowledge management. BMC Health Serv Res 2025;25(1):1371 [FREE Full text] [doi: 10.1186/s12913-025-13302-7] [Medline: 41102765]
33. Western MJ, Smit ES, Gültzow T, Neter E, Sniehotta FF, Malkowski OS, et al. Bridging the digital health divide: a narrative review of the causes, implications, and solutions for digital health inequalities. Health Psychol Behav Med 2025;13(1):2493139 [FREE Full text] [doi: 10.1080/21642850.2025.2493139] [Medline: 40276490]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
EHR: electronic health record
eID: electronic ID
ePA: elektronische Patientenakte
GesundheitsID: digital identity
GP: general practitioner
HCP: health care professional
NFC: near-field communication
OPP: obligatory passage point
PIN: personal identification number
PostIdent: postal identification procedure

Edited by L Buis; submitted 13.Mar.2026; peer-reviewed by K Becerro de Bengoa Losa, J Schwarz; comments to author 08.Jul.2026; revised version received 22.Jul.2026; accepted 24.Jul.2026; published 14.Sep.2026.

Please cite as:
May S, Marquardt M, Barzen G, Seifert F, Nyarko M, Niemeyer A, Gscheidle S, Schenk L, Spethmann S, Muehlensiepen F
Accessing Germany's Electronic Patient Record Through Health Insurer–Provided Apps: Qualitative Interview Study
JMIR Mhealth Uhealth 2026;14:e95294
URL: https://mhealth.jmir.org/2026/1/e95294
doi:10.2196/95294
PMID:

©Susann May, Manuela Marquardt, Gina Barzen, Frances Seifert, Mary Nyarko, Antonia Niemeyer, Sonja Gscheidle, Liane Schenk, Sebastian Spethmann, Felix Muehlensiepen. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Mobile Heat Early Warning System for Pregnant and Postpartum Women in South Africa: Prospective Observational Evaluation Study

Clara Sofia Heil[1], MSc; Shobna Sawry[2], MSc; Susanne Frennert[1], PhD; Jean Le Roux[2], MSc; Pascalia Ozida Munyewende[2], PhD; Sibusiso Mkwananzi[2], PhD; Ijeoma Solarin[2], MPH; Perkins Watambwa[3], MSc; Koen van der Sanden[4], MSc; Isabelle Lange[5,6], PhD; Günter Alce[1], PhD; Veronique Filippi[5], PhD; Stanley Luchters[3,7,8], MD, MSc, PhD; Chuansi Gao[1], PhD

[1]Aerosol and Climate Laboratory, Division of Ergonomics and Aerosol Technology, Department of Design Sciences, Faculty of Engineering, Lund University, Lund, Skåne, Sweden

[2]Climate and Health, Implementation Science, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, Gauteng, South Africa

[3]Centre for Sexual Health and HIV AIDS Research, Harare, Harare, Zimbabwe

[4]Netherlands Organisation for Applied Scientific Research, Soesterberg, South Holland, The Netherlands

[5]London School of Hygiene & Tropical Medicine, London, England, United Kingdom

[6]Department of Neurology, Technical University of Munich, Munich, Bavaria, Germany

[7]Department of Public Health and Primary Care, Ghent University, Ghent, Flanders, Belgium

[8]Liverpool School of Tropical Medicine, Liverpool, England, United Kingdom

Corresponding Author:
Clara Sofia Heil, MSc
Aerosol and Climate Laboratory, Division of Ergonomics and Aerosol Technology, Department of Design Sciences
Faculty of Engineering
Lund University
Box 188, Se-221 00
Formstråket 13
Lund, Skåne, 22362
Sweden
Phone: 46 725853154
Email: clara.heil@design.lth.se

Abstract

Background: Early warning systems (EWS) have the potential to reduce heat-related health risks, yet few are designed specifically for vulnerable populations. Pregnant women and infants, often cared for by postpartum women, are particularly susceptible to heat-related illnesses. The MotherHeat Alert mobile app is a heat EWS specifically developed for pregnant and postpartum women.

Objective: This study aimed to assess the usability and effectiveness of the MotherHeat Alert in the Tshwane District, South Africa.

Methods: This prospective observational evaluation study included a cohort of 201 participants (100 pregnant and 101 postpartum) who were followed over a 6-month period. In addition, a questionnaire was administered at 3 time points (baseline, midpoint, and end point) to collect self-reported information on health, living situation, and heat exposure, among other parameters. Usability was evaluated in the end point questionnaire through a validated Post-Study System Usability Questionnaire (PSSUQ), using descriptive statistics. Effectiveness was assessed through changes in self-reported heat-protective behaviors between baseline, midpoint, and end point, as well as through end point assessments of heat-health knowledge increase and self-reported behavior change attributable to the intervention. Generalized estimating equations were used to analyze repeated measures of heat-protective behaviors, while descriptive statistics were used to summarize end point–reported knowledge gains and behavior change.

Results: Usability scores indicated high overall usability, with mean PSSUQ scores of 1.76 (SD 0.60) among pregnant women and 1.37 (SD 0.45) among postpartum women. Similar patterns were observed across most PSSUQ subscales, including system usefulness, interface quality, and information quality. Most self-reported behavior change due to alerts received through the app

XSL•FO
RenderX

was rated as large to very large by the pregnant women (47/70, 67%) and postpartum women (68/82, 83%). Nevertheless, comparing behavior change at baseline and end point mainly highlighted a decrease in protective behavior in both pregnant and postpartum cohorts. Self-reported knowledge increase among pregnant and postpartum women was reported as large to very large following the use of the EWS (55/70, 79% and 72/82, 88%, respectively). Importantly, however, attrition was substantial, particularly among pregnant participants, reaching 30% (30/100) in the pregnant cohort and 19% (19/101) in the postpartum cohort by end point.

Conclusions: The MotherHeat Alert app demonstrated positive outcomes in usability. Effectiveness showed conflicting results, with positive self-reports on behavior change and knowledge increase, yet a decrease in protective heat behaviors when comparing baseline and midpoint with end point results, likely due to seasonality effects. Nevertheless, these findings highlight the potential of tailored EWS to support heat adaptation among pregnant and postpartum populations. Successful implementation and scalability require careful consideration of contextual factors, further assessment of long-term usability and integration into existing health infrastructure.

(JMIR Mhealth Uhealth 2026;14:e102863) doi:10.2196/102863

KEYWORDS

pregnancy; postpartum period; hot temperature; mobile apps; digital health; maternal health; global warming; South Africa

Introduction

Heat exposure causes an estimated 546,000 deaths annually worldwide [1], and contributes to increases in morbidity, particularly among high-risk populations [2] such as pregnant women. With heat extremes becoming longer, more frequent, and more intense, heat extremes currently classified as rare will become increasingly commonplace [3,4]. Pregnant women face heightened physiological strain due to elevated core body temperature and metabolic demands, while postpartum women bear responsibility for protecting heat-vulnerable newborns who cannot thermoregulate effectively. Many women in Africa have limited or no protection against heat during pregnancy [5], outlining the urgent need to identify effective prevention and response measures in these regions [3]. Heat action plans and forecasts, including heat-related early warning systems (EWS), issue early alerts and advisories [6,7] and have been demonstrated to promote heat adaptation and save lives [7-10]. However, current EWS designed to reduce heat-related health risks remain largely generic [9,10]. Despite heat-health risks in pregnancy, no heat-health EWS has been systematically designed and evaluated for pregnancy and the perinatal period.

EWS are widely recognized as cost-effective tools for climate adaptation, integrating different aspects of risk knowledge, monitoring and warning, and communication and response capacity [8]. A central challenge for EWS effectiveness is risk communication; systems must ensure that people at risk receive, comprehend, and respond to the alerts [11] communicated at the community and individual level [12]. For EWS to function effectively across different geographies and cultures, they should integrate social and cultural perceptions shaping heat awareness, risk perception, protective behaviors, and coping strategies, while providing science-based and locally relevant warnings [8]. Additionally, warnings issued without advice on responses are less effective [13]. Evidence from sub-Saharan Africa suggests that digital health interventions are most successful when they are simple, low-cost, and aligned with local contexts and languages [8,14,15]. Successful implementation further depends on user attitudes and behaviors and cultural acceptance [16]. Interventions must therefore be people-centered, context-sensitive, and coproduced with communities and local partners to ensure warnings are understood and acted upon [8,14,15]. Integrating such systems into existing digital health strategies may further support adoption and sustained engagement [14].

While there seems to be some awareness regarding heat warnings and high temperature forecasts among the general population, taking protective actions is less common [17], particularly among individuals who do not see themselves as at risk [13]. There is a demand for assessing the effectiveness of interventions to determine their ability to change knowledge, awareness, and behavior. Limited awareness of heat-induced health risks poses a major challenge for heat EWS [10,18]. Increasing the understanding of these risks and adopting small behavior changes could significantly reduce heat-induced morbidity and mortality [18]. Pregnant women exposed to cues to action are more likely to exhibit positive adaptation behaviors in response to heat, influenced by the awareness of vulnerabilities and their risk perception [18]. Therefore, effective communication of information enhances awareness while also significantly leading to adoption of adaptive behaviors [18]. Furthermore, research stresses the importance of assessing the effectiveness of heat EWSs, not just overall, but of specific interventions, particularly targeting vulnerable populations [10]. Improving the design and implementation of EWS directly enhances their effectiveness [10] and usability, which refers to user satisfaction within a defined intended use and context of use [19].

This study evaluates the usability and effectiveness of a mobile heat EWS app implemented in a South African context, with messages co-designed with pregnant and postpartum women. It is part of the larger Heat Indicators of Global Health (HIGH) Horizon project [20], which developed this EWS to help protect pregnant and postpartum women from heat exposure [21]. We assess whether the warning system is perceived as useful by its intended users, whether the mHealth app can improve awareness of the adverse health consequences of heat during and shortly after pregnancy, and whether it supports behavior change toward more protective heat-health practices.

XSL•FO
RenderX

Methods

Study Design and Setting

We conducted an intervention cohort study to evaluate the usability and effectiveness of a personalized heat-health EWS, MotherHeat Alert, among 100 pregnant and 101 postpartum women in South Africa. As part of the larger HIGH Horizons project [20], the study was implemented between October 2024 and August 2025 at a community health center (CHC) in Tshwane, Gauteng Province, South Africa. The district is predominantly urban, with large informal settlements. Housing consists of formal houses, backyard rooms, and informal dwellings, most of which lack consistent access to cooling infrastructure. The telecom industry reports that almost 98% of mobile connections in South Africa are connected via 3G, 4G, or 5G mobile networks [22]. Average daily outdoor temperature was estimated using Copernicus Climate Data for coordinates near the health facility during the peak summer period (December 2024 to January 2025), when temperatures are typically highest in the study setting. During this period, average monthly maximum temperatures ranged from 26.9°C to 27.4°C, with 19 days measuring above 30°C and 3 days above 35°C.

Study Participants

A total of 2 cohorts were enrolled in this study:

1. Pregnant women: Confirmed pregnant, in the second trimester of pregnancy as assessed by a registered health worker, aged 18 years and older, and attending antenatal care services at the CHC.
2. Postpartum women: Women aged 18 years and older, with a live-birth infant in the past 3 months or women with a child aged 6-9 months of age, and recruited from the postnatal care services or well-baby and immunization clinics at the CHC.

Additionally, all participants needed to own a smartphone that could run the MotherHeat Alert app, intend to remain in the study area for the next 6 months, and be able and willing to provide written informed consent for study participation.

Sampling and Recruitment

Convenience sampling was used to recruit participants, with efforts made to cover diversity in age and demographic background. Recruitment flyers and posters were displayed in the maternity unit and health facilities, containing brief details about the study and contact details. Languages spoken throughout recruitment were English, Ndebele, Sepedi, Setswana, Zulu, and Sesotho.

Study Visits

Questionnaire data were collected at 3 study visits (baseline, 2 months, and 6 months). The enrollment visit (~60 to 90 minutes) included informed consent, administration of the baseline questionnaire, and a 30-minute training session to install and configure the MotherHeat Alert app and show participants how to use the app. User profiles were set up within the app to include age, body weight, and height. Follow-up visits occurred at 2 and 6 months post enrollment, January-March and June-August, respectively. Slight changes were made between trial registration and the implementation of the study by introducing visit windows to reduce the risk of missed visits and participant dropouts. Each of these sessions lasted approximately 60 minutes and included the questionnaire administration. At the 6-month visit, participants also completed the Post-Study System Usability Questionnaire (PSSUQ) to assess the usability of the EWS (listed in Multimedia Appendix 1). At any time during the study period, participants could complete a short in-app survey, consisting of 7 questions. This survey was optional and assessed heat symptoms, perceived heat, and app usability, collecting real-time submission of user responses.

Data Collection and Analyses

Quantitative data were collected through structured, interviewer-administered questionnaires using Research Electronic Data Capture (REDCap; Vanderbilt University) [23] and through the continuous in-app data collection. Questionnaires captured sociodemographic information, medical and obstetric history, knowledge and protective behaviors linked to heat risks, data on the economic burden and app feasibility, and acceptability and usability, particularly at the end point visits. The in-app data included timestamps of interactions, app usage metrics, and responses to usability and heat-related questions. For this study, usability, knowledge increase, and behavior change related to the intervention were of main interest. Furthermore, the EWS captured heat parameters, which were used to analyze 2 in-app survey items, as described in more detail in the following section. The current Universal Thermal Climate Index (UTCI), a comprehensive heat stress index integrating air temperature, humidity, solar radiation, and wind speed and including models estimating clothing insulation and metabolic heat production [24], was used to capture the environmental heat exposure in this study. The app retrieved local weather forecast data daily at approximately 6 AM, but an alert was only generated and issued when the predefined heat threshold was met. Alert generation was determined locally on participants' devices.

Outcome Measures: Usability and Effectiveness

Usability was assessed through a validated PSSUQ administered at the end point visit. The PSSUQ is a 16-item scale that assesses user satisfaction through an overall evaluation of the system used [25]. Besides its overall score, it has 3 subscales that have been assessed individually: system usefulness, information quality, and interface quality. All scales have a maximum possible score of 7, with lower numbers indicating better usability [25]. Effectiveness in the context of this study refers to the ability of MotherHeat Alert to change knowledge and awareness of heat-health risks [26]. This was assessed through comparisons between end point and baseline scores. Effectiveness was therefore evaluated based on (1) heat-protective behavior change, for instance, through the question "Thinking about the last day when it was really hot, what did you do to keep yourself cool?" and (2) knowledge increase on heat and health following the intervention period, for instance, through the question "From your use of the app, do you think that your knowledge about the effects of heat on your health has increased?" Table S1 in Multimedia Appendix

XSL•FO
RenderX

1 provides a visual illustration of the usability and effectiveness concepts and how these were assessed. All analyses were conducted using SPSS version 30.0 (IBM Corp).

Statistical Analysis

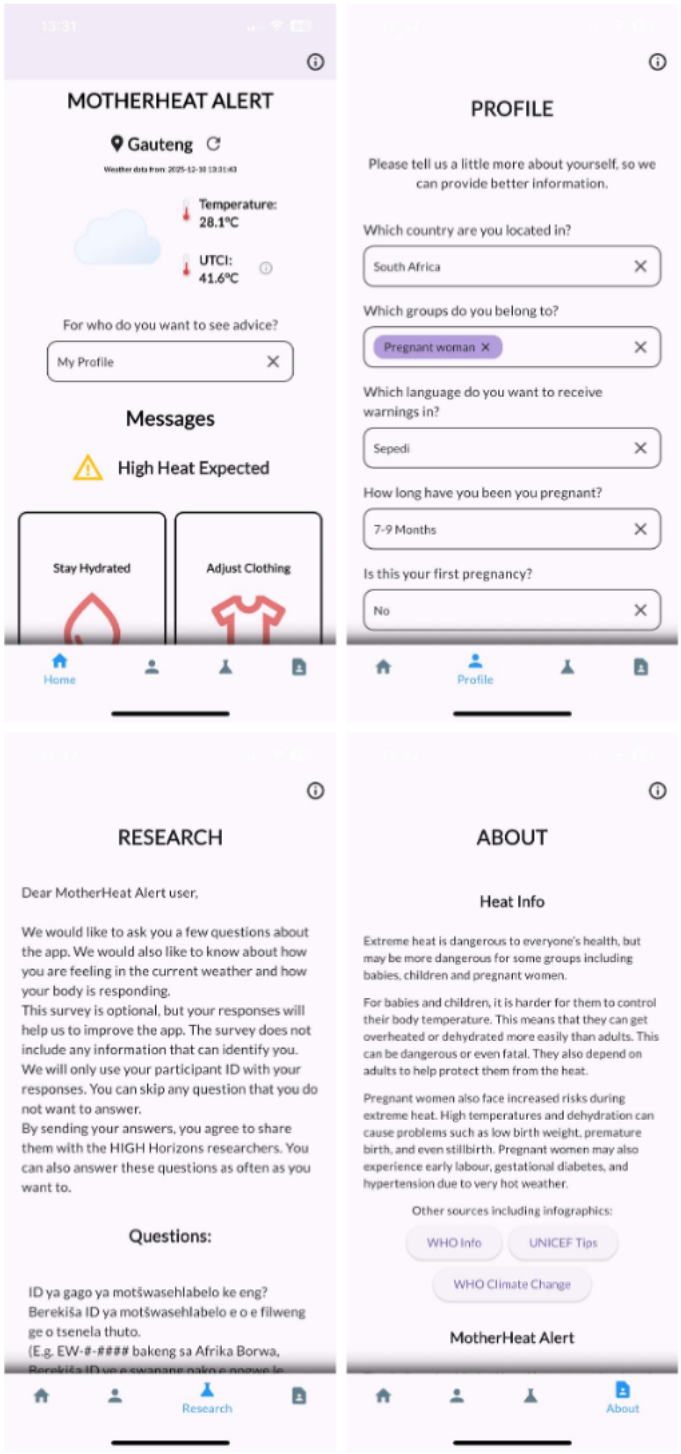
Most self-reported usability and knowledge increase perceptions were assessed at end point, using descriptive statistics. Binary logistic generalized estimating equations (GEE) models were used on 2 questionnaire items, namely "Thinking about the last day when it was really hot, what did you do to keep yourself cool?" and "Thinking about the last day when it was really hot, what did you do to protect your baby from the heat?" to assess whether a significant behavior change was detected by comparing baseline and end point results. We further used 2 methods to assess the in-app survey items: a generalized linear mixed model (GLMM) to evaluate survey items with large discrepancies in repeated responses and a linear mixed effects model with random intercepts, which accounts for substantial between-individual differences at baseline (Table S1 in Multimedia Appendix 1 provides further details regarding choice of statistical model and time of data collection).

The Heat Health EWS: MotherHeat Alert

The MotherHeat Alert app is a mobile EWS for Android (Google LLC) and iOS (Apple Inc.) users that provides real-time, localized heat warnings with tailored health advice to improve awareness of heat-related health risks and promote protective behavior [26]. The app served as an intervention that integrates environmental data (UTCI temperature, air temperature, humidity, solar radiation, and wind) with the user's gestational status to deliver personalized alerts and recommend specific, locally adapted protective actions (refer to Figure 1 for app interface). It used a 7-day UTCI daily forecast and issued push notifications at 6 AM whenever at least one of the following 7 days was predicted to exceed the predefined heat alert threshold of 27.3°C UTCI. Individual parameters including age and BMI could also be added to the EWS to further personalize messaging. Participants selected profiles, choosing between pregnancy or postpartum stages, which in turn determined the content and timing of heat alerts and messages. At a minimum, the EWS aimed to increase heat risk awareness and support behavior change, such as using specific cooling techniques, more suitable clothes, and how to stay hydrated, among others, during hot, high-risk periods. In this study, the EWS app was provided free of charge and delivered response messages with locally applicable guidance in local languages.

XSL•FO
RenderX

Figure 1. The MotherHeat Alert interface. From left to right, top to bottom: the home screen, the profile screen, the research screen, and the about screen.



Patient and Public Involvement

A co-design process was used to develop the messages used in the MotherHeat Alert app. This included photovoice workshops with pregnant and postpartum women recruited from the study site to develop messages based on their lived experiences [27]. The draft messages were then presented to community health workers based at the CHC in 2 focus group discussion sessions to enable triangulation and further contextual validation. Further refinements were made in consultation with expert reviewers. This process helped ensure that the EWS was grounded in lived experiences while being informed by professional expertise.

Ethical Considerations

Both the University of Witwatersrand Human Research Ethics Committee (HREC) in South Africa (approval number 240606)

XSL•FO
RenderX

and district approval were granted to conduct the study in a Tshwane health facility. The study was registered with the South African National Clinical Trial Registry (DOH-27-032025-5623), as part of the European Union (EU) and United Kingdom–funded HIGH Horizons project. All participants provided written informed consent. Data were fully anonymized or pseudonymized. Each participant was assigned a unique participant identification number. Participants were reimbursed ZAR400 (±US $25) for attending each study visit.

Results

Overview

In total, 280 women were screened for eligibility. Of these, 209 (74.6%) met the inclusion criteria and provided informed consent. Reasons for nonenrollment include not having compatible smartphones (26/280, 9.3%), not planning to remain in the area for 6 months (14/280, 5%), or not willing and/or able to provide written consent (51/280, 18.2%; note that reasons are not mutually exclusive). Of the 209 women who were enrolled, 8 discontinued early: 5 were unreachable after enrollment, 1 declined to install the app, and 2 withdrew after enrollment without providing a reason. Ultimately, 201 participants (100 pregnant and 101 postpartum) were included in the analyses; however, 25 (12.4%) did not complete the midpoint visit, and 50 (24.9%) did not complete the end point visit. Enrollments occurred between October 2024 and January 2025, while 2-month follow-up (midpoint) and 6-month follow-up (end point) visits were conducted between January 2025 and March 2025 and June 2025 and August 2025, respectively. For the pregnant cohort, the number of participants retained was 100 at baseline, 93 at midpoint, and 70 at end point. For the postpartum cohort, 101 were included at baseline, 83 at midpoint, and 82 at end point. Due to the privacy-by-design architecture of the system, alert generation was not recorded on the central server; therefore, the number and timing of alerts displayed to individual participants could not be determined.

At baseline, the median age for both pregnant and postpartum participants was 29 (IQR pregnant women 25-34; IQR postpartum women 24.5-33) years. Most participants completed high or secondary school (117/201, 58.2%), and half were unemployed but looking for work (100/201, 49.8%). Most participants (193/201, 96.0%) reported owning a phone that was not shared with other family members. Almost all (n=200) participants reported using apps on their phone, mostly for social media (198/201, 98.5%) and messaging (92/201, 45.8%; Table 1). Furthermore, 55 of 201 (27.4%) participants reported having a pregnancy or baby monitoring app on their phone, and 79 of 201 (39.3%) participants received pregnancy or baby growth and development messages on their phones (eg, MomConnect). At baseline, prior to the MotherHeat Alert intervention, the largest proportion of participants reported never discussing heat-related health concerns with the community (96/201, 47.7%; Table 2), and only a few participants mentioned knowing about any community programs or public health messages about heat safety (9/201, 4.5%; Table 2). Among the postpartum cohort, very few participants reported knowing about the effect of heat on themselves during pregnancy or the postnatal period (10/101, 9.9%) and their baby (9/101, 8.9%; Table 2). Around half of the participants (102/201, 50.7%) reported that it is hotter inside compared to outdoors (Table 2). A large proportion (65/201, 32.3%) reported wanting to take additional particular heat-protective measures but were unable to do so.

XSL•FO
RenderX

Table 1. Descriptive statistics on education, employment, phone ownership, and general app usage at baseline.

Characteristics	Pregnant, n (%)	Post partum, n (%)
Maternal age (years)		
≤20	0 (0)	5 (5)
21-25	23 (23)	18 (17.8)
26-30	30 (30)	31(30.7)
31-40	23 (23)	33 (32.7)
>40	24 (24)	14 (13.9)
Education		
Completed only primary school	1 (1)	0 (0)
Some high or secondary school	17 (17)	19 (18.8)
Completed high or secondary school	54 (54)	63 (62.4)
Some tertiary, university, or college	20 (20)	8 (7.9)
Completed tertiary, university, or college	8 (8)	11 (10.9)
Total	100 (100)	101 (100)
Employment[a]		
Employed full-time	35 (34.4)	28 (27.7)
Employed part-time	10 (9.8)	3 (3)
Self-employed	7 (6.9)	0 (0)
Not employed but looking for work	42 (41.2)	58 (57.4)
Not employed and not able to work	0 (0)	7 (6.9)
Not employed and not looking for work	2 (2)	0 (0)
Homemaker	5 (4.9)	5 (5)
Student	1 (1)	0 (0)
Total	102 (100)	101 (100)
Phone ownership and app usage[a]		
Phone used solely by participant	96 (96)	97 (96)
Phone shared with other family members	4 (4)	4 (4)
Using apps on phone	100 (100)	100 (99)
Messaging	39 (39)	53 (52.5)
Health	16 (16)	20 (19.8)
Weather	30 (30)	35 (33.7)
News	9 (9)	15 (14.9)
Social media	98 (98)	100 (99)
Other	3 (3)	2 (2)

[a]Notice that this does not add up to the total number of participants per cohort, since response options were not mutually exclusive.

XSL•FO
RenderX

Table 2. Self-reported indoor heat-health parameters by pregnant and postpartum cohort at baseline prior to installation of the app.

Participants' heat-health information and awareness prior to app installation	Pregnant (n=100), n (%)	Postpartum (n=101), n (%)
Discuss health concerns with community		
Never	46 (46)	50 (49.5)
Rarely	3 (3)	6 (5.9)
Sometimes	41 (41)	28 (27.7)
Often	5 (5)	11 (10.9)
Always	5 (5)	6 (5.9)
Availability of heat-health information and protective measures		
Community programs or public health messages about heat safety in the area	4 (4)	5 (5.0)
Protective action that you would have liked to do, but were unable to	26 (26)	39 (38.6)
Heat-health knowledge postpartum		
Have you heard about how hot weather might affect your health?	—[a]	10 (9.9)
Have you heard about how hot weather might affect your baby's health?	—	9 (8.9)
Indoor heat perception		
Indoor is cooler	38 (38)	46 (45.5)
Indoor is hotter	49 (49)	53 (52.9)
Just as hot inside as outside	10 (10)	2 (2)
Do not know	3 (3)	0 (0)

[a]Not available.

Usability

With lower PSSUQ scores indicating better usability, the summed overall score of PSSUQ (questions Q1-Q16) in the pregnant cohort ranged between 1 and 3.44 (mean 1.76, SD 0.60; n=70) and 1 and 2.85 (mean 1.37, SD 0.45; n=81) in the postpartum cohort (Figure 2). The PSSUQ score on system usefulness (Q1-Q6) ranged between 1 and 2.83 (mean 1.34, SD 0.51; n=70) for the pregnant cohort and 1 and 2.83 (mean 1.28, SD 0.39; n=81) for the postpartum cohort. The PSSUQ score on information quality (Q7-Q12) was higher for the pregnant cohort and ranged between 1 and 5.33 (mean 2.52, SD 1.19; n=70) but lower for the postpartum cohort, ranging from 1 to 3.67 (mean 1.48, SD 0.64; n=81). Finally, the PSSUQ score on interface quality (Q12-Q16) ranged between 1 and 4.0 (mean 1.31, SD 0.63; n=70) for pregnant and 1 and 3.25 (mean 1.35, SD 0.53; n=81) for the postpartum cohort (Figure 3). To account for the large dropout rate across the study (Figure 4), we conducted a sensitivity analysis for the total usability-related scores, comparing baseline to the end point. Overall, 2 different scenarios were assessed, one assuming that dropouts occurred due to low usability ratings and the other assuming that dropouts occurred due to other external reasons.

Baseline comparisons between participants who dropped out and those who remained in the study showed no significant difference in age (independent-samples 2-tailed *t* test), self-rated general health, presence of chronic health conditions, perceived indoor heat, housing type (formal vs informal dwelling), or perceived seriousness of heat for themselves or their infants (chi-square tests; all $P<.05$). There was some indication of differences by educational attainment, with a nonsignificant overall association (chi-square test, $P=.09$), but evidence of a significant linear trend ($P=.02$), suggesting lower dropout rates among participants with higher levels of education. Furthermore, sensitivity analyses were conducted to evaluate the effect of missing end point usability data under alternative assumptions. In the worst-case scenario, dropouts were assigned the maximum PSSUQ score of 7, representing the poorest usability and assuming that withdrawal was related to poor user experience. In the as-usual scenario, dropouts were assigned scores comparable to those who completed end point assessment, assuming that dropouts occurred for reasons unrelated to usability. Results were robust under the as-usual assumption (mean 1.53, SD 0.52, 95% CI 1.46-1.61; n=201). Under the worst-case assumption, the mean score was higher (mean 2.91, SD 2.41, 95% CI 2.57-3.24; n=201), indicating poorer usability and greater variability; however, it did not change the overall interpretation of the findings.

For the in-app question "Was the message you received relevant for your situation" (yes or no), current UTCI significantly predicted the likelihood of participants finding messages relevant: for every 1°C increase, the odds of reporting that a message was relevant increased by ±5% (odds ratio [OR] 1.05, 95% CI 1.01-1.09). The model highlighted significant between-person variability in baseline relevance rating (Var=3.59; $P<.001$). For the in-app question "On a scale of 1 to 5, please rate how easy you have found it to use the app" (1=easy, 5=difficult), a comparison of novice users with experienced users assessing whether the perceived ease of use changed with experience showed that the effect of time was

XSL•FO
RenderX

statistically significant ($F_{1,6335}$=76.18; P<.001), indicating that perceived difficulty decreased as participants used the MotherHeat Alert over time. Importantly, however, this statistically significant change was very small in magnitude (marginal R^2=0.007), where most variation can be explained by the differences across users (conditional R^2=0.63). This suggests that 0.7% of the variance in difficulty can be explained through the repeated use over time.

Figure 2. Violin plot showing Post-Study System Usability Questionnaire total score (Q1-16) stratified by pregnant and postpartum cohort. PSSUQ: Post-Study System Usability Questionnaire.

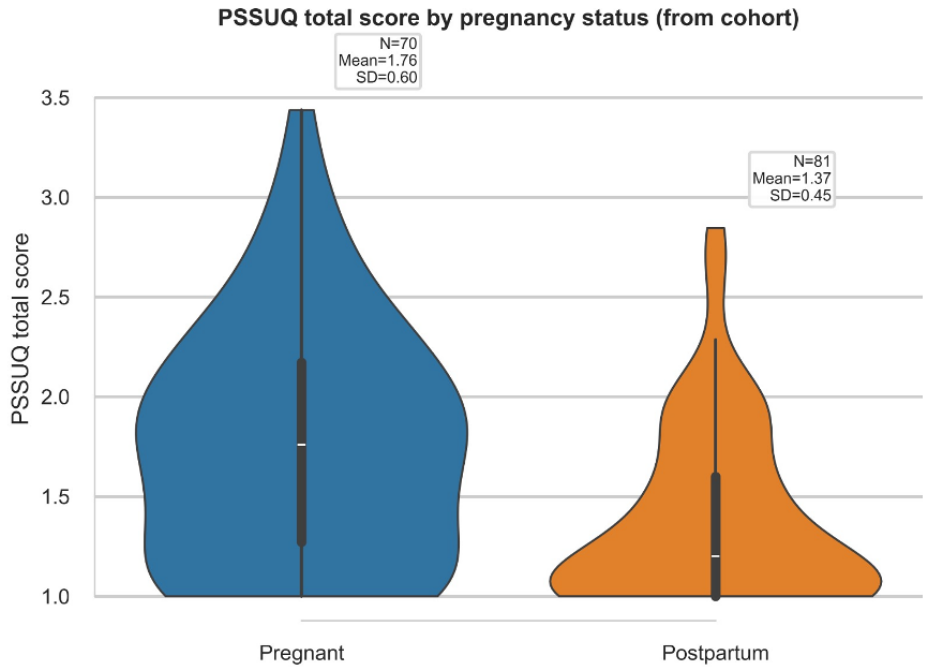
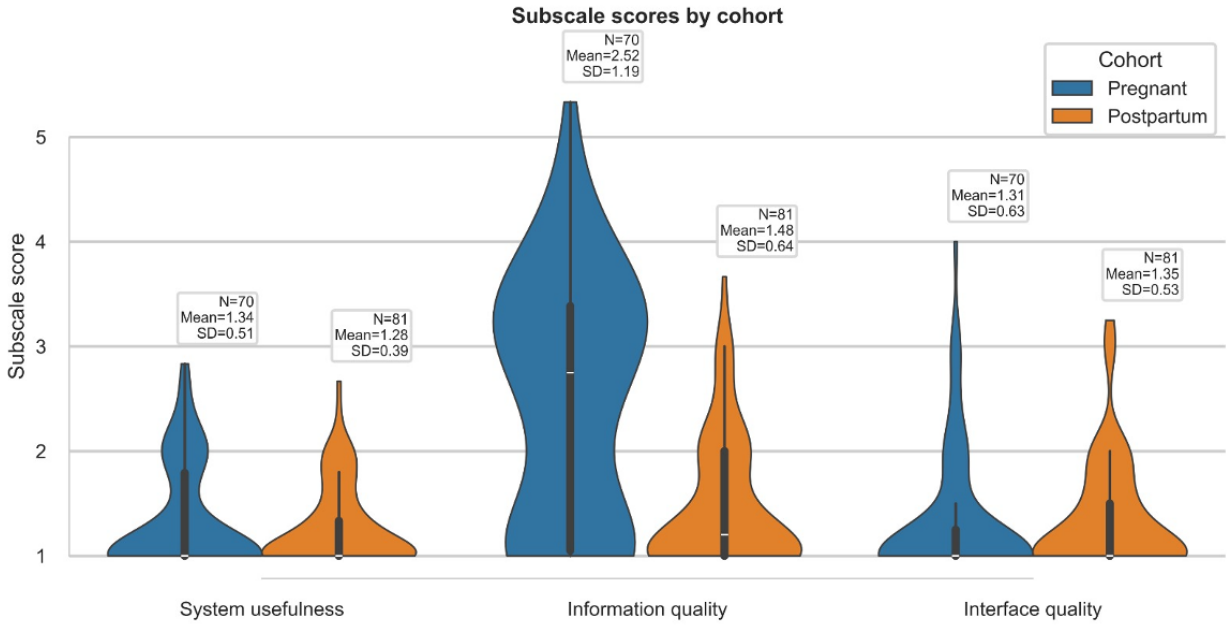


Figure 3. Violin plot comparing the PSSUQ subscales, system usefulness, information quality and interface quality. Results are stratified amongst pregnant and postpartum cohorts. PSSUQ: Post-Study System Usability Questionnaire.



XSL•FO
RenderX

Figure 4. Dropout rate across the study; at baseline, midpoint (2-month follow up) and end point (6-month follow up) for pregnant cohort and postpartum cohort.



Effectiveness

Knowledge Increase

The largest proportion in both pregnant and postpartum cohorts self-reported a large to very large knowledge increase (55/70, 78.6% and 72/82, 87.8%, respectively) on the effects of heat on pregnant women's health (Table 3). The postpartum cohort also indicated a large to very large knowledge increase (72/82, 87.8%) on the self-reported effects of heat on the baby's health.

Table 3. Self-reported knowledge increase due to early warning systems app reported by pregnant and postpartum cohort.

Self-reported knowledge increases due to EWS[a] use	Values
Pregnant cohort: effects of heat on pregnancy health, n (%)	
No-slight increase	2 (2.8)
Moderate increase	13 (18.6)
Large-very large increase	55 (78.6)
Do not know	0 (0)
Total	70 (100)
Postpartum cohort: effects of heat on infants' health, n (%)	
No-slight increase	0 (0)
Moderate increase	9 (11)
Large-very large increase	72 (87.8)
Do not know	1 (1.2)
Total	82 (100)

[a]EWS: early warning systems.

Behavior Change

The item assessing "Thinking about the last day when it was really hot, what did you do to keep yourself cool?" contained 17 potential response options. In a GEE model, after adjusting for maternal age and education, a significant behavior change was noted for 8 options comparing end point (6-month follow-up after app introduction) to baseline (preapp introduction) in paired analyses (Table 4). For both pregnant (n=70) and postpartum cohorts (n=82), the following behaviors showed statistically significant shifts toward less frequent use after introduction of the app: drinking more water, drinking more fluids, bathing more, using more light and loose clothing, opening windows, and wearing a hat. For the pregnant women cohort, eating more fruits and increased use of a fan also decreased significantly between baseline and end point. For the postpartum cohort, working less and staying indoors decreased significantly between baseline and end point. A total of 4 response options—using Mcako or clay or ibovu (a type of clay that is high in minerals and traditionally used as skin protection from sun exposure), air conditioning, using an umbrella, and did nothing—could not be analyzed due to an insufficient number of cases across groups for statistical testing. Importantly, all the comparisons between baseline and end point indicate a negative change, highlighting that less protective heat action had been taken at end point compared to baseline (Table 4). Comparing midpoint (2-month follow-up after app introduction) to baseline, a similar, yet less pronounced trend was discerned, with 2 response options showing a significant decrease in protective behavior in the pregnant cohort, including using a fan and wearing light and loose clothing. In the postpartum cohort, 7 behaviors indicated a significant decrease, including drinking other fluids, working less, bathing more frequently, wearing light and loose clothing, opening windows, wearing a hat, and staying indoors (Table S2 in Multimedia Appendix 1). In this GEE model, 2 covariates were controlled for: maternal age and education [28-31] since both have consistently been associated with more protective health behavior. Effect estimates of a stratified analysis of these covariates can be found in Table S3 in Multimedia Appendix 1 for end point results and Table S4 in Multimedia Appendix 1 for midpoint results. Statistical significance of the correlations did not differ between crude and adjusted models. At end point, almost all participants from the

XSL•FO
RenderX

pregnant (66/69, 96%) and postpartum (72/82, 88%) cohorts indicated that these measures were taken because they were alerted by the app.

The item "Thinking about the last day when it was really hot, what did you do to protect your baby from the heat?" contained 13 answer options, where giving the baby more water, giving the baby more fruit, and seeking shaded or cool areas yielded a positive behavior change, whereas dressing the baby in light and loose clothing and opening windows showed a decrease at end point in the adjusted model (Table 5). Again, the midpoint results yielded a similar pattern, where seeking shaded or cool areas and other measures showed an increase in this adaptive behavior, and dressing the baby in light and loose clothing, opening windows, and keeping the baby indoors showed a decrease in this behavior in the adjusted model (Tables S5 and S6 in Multimedia Appendix 1). Most postpartum women (72/82, 87.8%) indicated that these protective measures were taken because they were alerted by the app. The largest proportion of self-reported behavior change due to the alerts received through the app rated this change as large to very large: 67.2% (47/70) in the pregnant cohort and 82.9% (68/82) in the postpartum cohort (Table S7 in Multimedia Appendix 1).

The in-app question "Did you follow the advice provided? (yes or no)," which was linked to current UTCI, significantly predicted the likelihood of following advice; for every 1°C increase in current UTCI, the odds of following advice increased by ±14% (OR 1.14, 95% CI 1.09-1.20). The model highlighted significant between-person variability in the baseline following advice rating (random intercept var=5.072; *P*<.001).

Table 4. Pregnant and postpartum effect estimates from a generalized estimating equation assessing how pregnant and postpartum women kept themselves cool during hot days, comparing end point to baseline results by asking "Thinking about the last day when it was really hot, what did you do to keep yourself cool?"

Behavior[a]	Pregnant			Postpartum		
	Baseline (n=100), n	End point (n=70), n	aOR[b] (95% CI)[c]	Baseline (n=101), n	End point (n=82), n	aOR (95% CI)
Drank more water	95	66	0.09 (0.04-0.23)	+91	78	0.37 (0.16-0.86)
Drank other fluids	17	8	0.42 (0.18-0.97)	72	51	0.40 (0.23-0.72)
Ate more fruits	17	8	0.42 (0.19-0.92)	41	42	1.04 (0.64-1.70)
Sought shaded or cool areas	21	15	0.65 (0.32-1.33)	71	65	0.75 (0.46-1.23)
Went to cooler buildings	12	12	1.00 (0.41-2.42)	40	39	0.96 (0.57-1.62)
Used Mcako[d]/clay/ibovu	2	0	—[e]	1	6	6.34 (0.73-55.10)
Worked less	15	17	1.16 (0.58-2.30)	70	49	0.40 (0.25-0.65)
Bathed more frequently	36	23	0.53 (0.28-0.99)	82	68	0.47 (0.25-0.90)
Used fan	44	15	0.22 (0.11-0.42)	48	42	0.78 (0.47-1.31)
Used air conditioning	6	0	—	8	10	1.28 (0.49-3.37)
Wore light and loose clothing	36	24	0.56 (0.32-0.99)	93	72	0.20 (0.09-0.46)
Opened windows	50	37	0.58 (0.35-0.98)	98	68	0.06 (0.02-0.19)
Put feet or hands in cool water	8	4	0.48 (0.15-1.52)	34	24	0.61 (0.36-1.05)
Wore a hat	26	8	0.25 (0.10-0.58)	76	62	0.51 (0.29-0.88)
Used umbrella	41	0	—	79	0	—
Stayed indoors	40	52	1.64 (0.95-2.84)	74	58	0.47 (0.26-0.84)
Did nothing	0	0	—	0[e]	0	—

[a]Adjusted for maternal age and education.

[b]aOR: adjusted odds ratio.

[c]Model having insufficient data to execute the analysis.

[d]Mcako is a type of clay high in minerals that is traditionally used as skin protection from sun exposure.

[e]Not available.

XSL•FO
RenderX

Table 5. Postpartum cohort generalized estimating equation effect estimates assessing how mothers kept their babies cool during hot days, comparing end point to baseline results by asking: "Thinking about the last day when it was really hot, what did you do to protect your baby from heat?"

Behavior[a]	Baseline (n=101), n	End point (n=82), n	OR[b] (95% CI)
Breastfed more frequently	57	58	1.04 (0.64-1.70)
Gave baby more water	55	77	2.73 (1.57-4.75)
Gave baby more fruits	17	27	1.86 (1.05-3.28
Used Mcako[c]/clay/ibovu	0	0	—[d]
Sought shaded/cool areas	43	59	1.96 (1.17-3.28)
Bathed baby more frequently	68	67	0.96 (0.53-1.72)
Used fan	17	23	1.49 (0.73-3.03)
Used air conditioning	3	2	—
Dressed baby in light and loose clothing	91	74	0.29 (0.14-0.63)
Opened windows	86	69	0.37 (0.18-0.73)
Kept baby indoors	73	64	0.65 (0.35-1.20)
Did nothing	0	0	—
Other	2	12	6.69 (1.42-31.63)

[a]Adjusted for covariates: maternal age and education.

[b]OR: odds ratio.

[c]Mcako is a type of clay high in minerals that is used as skin protection or often also ingested.

[d]Not applicable; represents a model having insufficient data to execute the analysis.

Discussion

Principal Findings

This study assessed the usability and effectiveness of an EWS aimed at reducing heat-related health risks among pregnant and postpartum women.

Usability results indicated that the MotherHeat Alert was considered useful and was well accepted by users. The effectiveness findings of the MotherHeat Alert suggest that engagement with the EWS led to self-reported increases in adoption of heat-protective knowledge and behavior. Comparisons of actual self-reported heat-protective behaviors at baseline and end point showed that these behaviors largely decreased by the end of the 6-month intervention period. This decrease could also be identified to a lesser extent at the 2-month midpoint results. Importantly, these results should be interpreted with caution, as midpoint and end point assessments were conducted during cooler months, introducing seasonal variation as an important confounder. In addition, approximately one quarter of participants were lost to follow-up between baseline and end point.

Usability of the MotherHeat Alert

Successful digital health interventions rely on effective integration, adoption, and sustained use by their intended users [32,33]. The Technology Acceptance Model emphasizes perceived usefulness and ease of use as key determinants of technology adoption [34,35]. The positive usability and knowledge outcomes observed in this study support the relevance of these principles to heat-related EWS. Usability, assessed using the PSSUQ questionnaire, was high overall, with system usefulness, information quality, and interface quality all scoring above average compared to norms reported by Lewis and James [25]; the system usefulness subscale scored highest and the information quality subscale lowest. Although similar patterns were observed between pregnant and postpartum cohorts, postpartum women consistently rated the EWS's usability higher, compared to pregnant women, and a greater variability in responses was typically observed in the pregnant cohort. A total of 2 additional in-app usability indicators, "perceived relevance of warnings" and "ease of use," showed significant changes over time. The "perceived relevancy of warnings" increased with rising temperatures, with a 5% increase in likelihood per 1°C rise, suggesting that alerts aligned well with users' lived experiences. "Ease of use" improved over time, indicating that with repeated use and increased familiarity, the EWS was perceived as easier to use, a trend that has been seen in other studies [36,37].

Effectiveness of the MotherHeat Alert

Most intervention studies in African settings focus on health promotion and disease prevention, aiming to improve health knowledge and encourage behavior change [14]. These generally report positive results [14], which is reflected in the outcomes of this study where women expressed interest in weather apps and general involvement with the EWS. Education and awareness building are among the most common and effective apps of digital health interventions [32]. Behavior change is therefore a suitable measure of effectiveness, as education and knowledge shaping are widely used and effective behavior change techniques [38,39]. High engagement with the EWS supports its effectiveness, as 95.7% (66/69) of pregnant participants and 87.8% (72/82) of postpartum women reported taking protective actions in response to alerts. Effectiveness,

XSL•FO
RenderX

measured through self-reported knowledge increase, indicated that at end point, 78.6% (55/70) of pregnant women showed a large to very large knowledge increase on the effects of heat on their own health. For the postpartum cohort, 87.8% (72/82) reported a large to very large knowledge increase of the effects of heat on their own and their baby's health. The warnings issued through the MotherHeat Alert resulted in preventive actions, suggesting that the recommended measures were practical and feasible for participants to carry out.

Nevertheless, when comparing baseline to end point outcomes, 9 of 17 behaviors reduced significantly among pregnant participants; a similar, but less pronounced trend was observed when comparing midpoint to baseline. Seasonal variation is an important confounder when interpreting the observed decrease in protective behaviors. Baseline assessments were conducted from October 2024 to January 2025, covering the transition period into warmer and early summer months; midpoint assessments from January to March 2025, spanning the later summer period and the early transition into cooler months; and end point assessments from June to August 2025, during the cooler months. Baseline and end point assessments therefore occurred during substantially different climatic conditions, and changes in heat-protective behaviors cannot be attributed to the intervention alone but must be interpreted in the context of seasonal variation. This would also explain why the reduction in heat-protective behavior was less pronounced comparing midpoint to baseline than end point to baseline. As days became cooler, there was no need for pregnant and postpartum women to apply cooling measures as the ambient temperatures were not high enough to warrant adaptive behavior. Fewer cooling behaviors during cooler months should therefore not necessarily be interpreted as poorer adherence or reduced intervention effectiveness but may instead represent an appropriate behavioral response to lower heat exposure. This interpretation is supported by our in-app data, which showed greater adherence to warning advice at higher temperatures, indicating that perceived risk intensifies responsiveness. As the MotherHeat Alert and heat-health apps and EWS more broadly are most relevant during warmer periods, strategies to reengage and retain users ahead of following heat seasons should therefore be considered. Future evaluations should compare behavioral outcomes during similar climatic conditions, for example by conducting baseline and follow-up assessments within the same warm season or across multiple summers for clearer findings and interpretation. For protective behaviors exhibited by postpartum women directed toward infants, a similar negative trend, highlighting less protective behavior at end point, could be identified in some behavior responses, yet a positive trend, indicating more protective behavior, could be seen in others. Among postpartum women, the ability to work less and stay indoors may have been affected by the urgency of returning to employment after childbirth or a need to seek employment as financial responsibilities have increased.

The observed reductions in self-reported protective behaviors should not necessarily be interpreted as indicating reduced engagement with heat-protective practices. Although end point assessments took place during the cooler season, participants were asked to report behaviors undertaken on "the last day when it was really hot", rather than behaviors undertaken at the time of assessment. During the cooler season, this may have required participants to recall behaviors from further back in time than during the warmer periods, potentially affecting the accuracy of their responses. In addition, social desirability may have influenced self-reported behaviors, particularly at baseline when participants may have been more inclined to report behaviors perceived as desirable or appropriate. Overall, it is further likely that at the start of the intervention, participants were more engaged and motivated to use a new technology, which may have temporarily increased both attention to the app and self-reported adoption of heat-protective behaviors. However, over time, the novelty of participation likely diminished as attention fatigue set in, and actual behaviors may have returned to preintervention patterns, contributing to lower observed heat-protective practices at midpoint and end point. Therefore, the observed changes may reflect a combination of seasonality, social desirability, and reduced motivation over time, rather than an intervention effect alone.

Limitations

This study has several limitations. Participation was limited to women [14] attending health facilities and owning smartphones, introducing potential selection bias and thereby excluding some of the most vulnerable populations. Convenience sampling through antenatal clinics may have introduced additional selection bias, so findings may not be representative of the broader population of pregnant and postpartum women in South Africa and should therefore be generalized with caution. As the MotherHeat Alert is a text-based EWS, which delivers warnings through written messages, illiterate individuals were likely excluded; literacy rates were not assessed. To increase inclusivity, EWS apps should ideally rely more on graphic messages or audio prompts. Additionally, local infrastructure constraints, such as unstable electricity supply to charge mobile phones, unstable network, shutting down phones to preserve battery power [16], or deleting the app to preserve storage space on end-devices, may have affected notifications, alerts, and system performance. Recruitment in this study was constrained as some women lacked smartphones, while others had devices with limited storage. Installing the app often required deleting photos or other apps, which some participants were unwilling to do. Furthermore, the EWS app required mobile data for daily temperature synchronization and warning delivery. Future EWS should ideally function without the requirement for mobile data or operate through agreements with telecommunication providers to waive data charges (ie, zero-rated app). Finally, all outcomes were self-reported, introducing the possibility of response and social desirability bias. Given the enrollment procedures, participants were aware of project aims, which may have influenced their responses. Furthermore, participant engagement in this study may have been influenced by monetary reimbursement, which may suggest that women in real-world settings may be less likely to install or use the app given these barriers.

Given that the study intended to assess usability and effectiveness over the short- and long-term, 6 months of follow-up was required. This resulted in seasonal timing of data collection becoming a limitation, where baseline data was

XSL•FO
RenderX

collected during warmer months, while end point measures occurred during colder months. This may partly explain the decline in reported protective behaviors. Furthermore, given that the proportion of adaptive behaviors reported at baseline was high, particularly among postpartum women, participants may already have adapted to heat exposure, such that explicit warnings added limited new information. Dropout rates in this study were substantial, and usability outcomes were therefore further examined using sensitivity analyses under 2 assumptions: a worst-case scenario, in which participants discontinued due to poor usability, and an as-usual scenario, in which dropouts were unrelated to EWS usability. Baseline characteristics did not differ significantly between participants who dropped out and those who remained, suggesting no clear evidence of systematic differences, thereby supporting the as-usual scenario. In addition, local contextual factors indicate that some participants may have discontinued the follow-up because women may return to their place or country of origin after receiving maternity care in South Africa [40]. It is also important to note that women reported wanting to take additional heat-protective behavior at baseline, but reported they were unable to do so. This finding, however, reflected their general perceived ability to protect themselves from heat, rather than their ability to follow the specific behavioral advice provided through the MotherHeat Alert, which was codeveloped with intended users to ensure that the recommended actions were practical and feasible within their context. Furthermore, the project-based design and limited timeframe restricted conclusions about long-term usability beyond the novelty effect of a new intervention, and about sustainability and scalability, which should be prioritized prior to consideration for mass roll-out. This structural limitation presents a common limitation of intervention studies in African contexts [14]. Evidence on cross-country uptake of the EWS, scalability, and continuity of digital health interventions in Africa remains weak, with most existing studies limited to pilot or small-scale implementations [14,34]. Finally, it should be noted that, to the best of our knowledge, the PSSUQ has not yet been validated in South Africa, although previous research has used the PSSUQ in South Africa [41] and other African countries [42].

Future Directions

Given the borderless nature of heat risk and its increasing severity, the development of a global heat EWS should be a long-term goal [8]. Despite their potential, few countries have heat-specific EWS, and no global system exists for the specific vulnerable groups. This represents a critical gap, particularly since the United Nations' Early Warning for All initiative aims to ensure effective EWS coverage by 2027 [43]. Currently, South Africa still issues heat warnings using absolute air temperature thresholds unrelated to health responses [44]. Future systems should be built around specific heat-health indicators, locally relevant, and integrated heat stress thresholds and codeveloped communication strategies. Greater attention is needed to gendered and intersecting vulnerabilities to ensure access to warnings [15]. Future research should prioritize rigorous impact evaluations, cost-effectiveness analyses, and qualitative studies on implementation context [14,32]. In addition, simplicity of design and modest objectives have repeatedly been associated with successful implementation in African contexts, since these products were least affected by contextual challenges [14]. Greater emphasis is needed on integrating the digital interventions into the existing health care infrastructure and addressing scalability [45]. Given the constraints of limited memory on mobile phones, together with the high use of social media among participants, stand-alone apps for heat-health warnings may not be the most efficient approach. Integrating these warnings into social media platforms may therefore be a more feasible and effective strategy.

Digital health interventions are generally received with optimism [14]. As mobile phone use is rapidly increasing in sub-Saharan Africa, the digital health technologies designed for this context follow suit [34]. Although evidence on long-term effectiveness and impact remains limited, this study demonstrates high effectiveness and usability of a heat-related EWS. Education and training have shown to reduce heat stress and illness [39], as shown in a study where digital technologies supporting maternal health in sub-Saharan Africa have contributed to improved maternal and infant health decision-making [16].

Conclusion

Overall, our study strengthens the growing evidence that well-designed, user-centered digital health technologies, including heat EWS, have high usability outcomes and contribute to knowledge gains in the context of climate-related health risks. However, the contribution of EWSs to improved health outcomes in broader contexts will depend on whether they are perceived as useful, easy to use, and whether social, cultural, and policy contexts have been considered during their implementation. This study demonstrates the potential of a co-designed heat-health EWS tailored toward pregnant and postpartum women to support protective responses to heat exposure. The intervention demonstrated high self-rated usability and perceived effectiveness among users. However, the observed decrease in heat-protective behaviors when comparing reported behavior at baseline compared to end point assessments should be interpreted with caution, as seasonal variation was linked to the follow-up visits through the study design, making it impossible to attribute behavioral changes to the intervention alone. Digital health technologies may play an important role in strengthening climate adaptation responses for maternal health. Focusing on the usability and effectiveness of EWS interventions, considering contextual relevance, assessing long-term usability beyond the novelty effect, and ensuring access for everyone by integrating these interventions into existing health systems is required to achieve this potential.

Acknowledgments

We thank Rose Refilwe Lamola, Celeste Madondo, Nondumiso Precious Mbatha, Mbalenhle Felicity Mvundla, and Kgomotso Reginah Mabidikama for their contributions to the implementation of the MotherHeat Alert and data collection. We also

XSL•FO
RenderX

acknowledge Birgit Kerstens for her continued organizational support within the HIGH Horizons project. ChatGPT (OpenAI) was used for refinement, correction, and editing of the manuscript to improve clarity of language.

Funding

This study was funded by the European Union's Horizon Framework Programme under grant agreement number 101057843. Project partner London School of Hygiene & Tropical Medicine is funded by UK Research and Innovation's Innovate UK agency (reference number 10038478). The funder played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to sensitive health data and participant confidentiality but are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: CSH, SF, GA, CG
Data curation: CSH, SS, JLR, POM, SM, KVDS, GA, CG
Formal analysis: CSH, PW
Funding acquisition: VF, SL, CG
Investigation: CSH, SS, JLR, POM, SM
Methodology: CSH, SF, GA, CG
Project administration: SL, CG
Resources: SS, JLR, POM, SM
Software: KVDS, GA, BK
Supervision: SF, GA, CG
Validation: CSH, SF, GA, CG
Visualization: CSH, PW
Writing – original draft: CSH
Writing – review & editing: CSH, SS, SF, POM, SM, IS, KVDS, IL, GA, SL, CG

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables, figures, and questionnaire items related to the usability and effectiveness of the MotherHeat Alert Early Warning System.
[DOCX File , 627 KB - mhealth_v14i1e102863_app1.docx]

References

1. Romanello M, Walawender M, Hsu S, Moskeland A, Palmeiro-Silva Y, Scamman D, et al. The 2025 report of the Lancet Countdown on health and climate change: climate change action offers a lifeline. Lancet 2025 Dec 13;406(10521):2804-2857. [doi: 10.1016/S0140-6736(25)01919-1] [Medline: 41175887]
2. Ballester J, Quijal-Zamorano M, Méndez Turrubiates RF, Pegenaute F, Herrmann FR, Robine JM, et al. Heat-related mortality in Europe during the summer of 2022. Nat Med 2023;29(7):1857-1866 [FREE Full text] [doi: 10.1038/s41591-023-02419-z] [Medline: 37429922]
3. Jay O, Capon A, Berry P, Broderick C, de Dear R, Havenith G, et al. Reducing the health effects of hot weather and heat extremes: from personal cooling strategies to green cities. Lancet 2021;398(10301):709-724. [doi: 10.1016/S0140-6736(21)01209-5] [Medline: 34419206]
4. 2024 State of Climate Services. Geneva, Switzerland: World Meteorological Organization; 2024.
5. Chersich MF, Scorgie F, Filippi V, Luchters S, Climate ChangeHeat-Health Study Group. Increasing global temperatures threaten gains in maternal and newborn health in Africa: a review of impacts and an adaptation framework. Int J Gynaecol Obstet 2023;160(2):421-429 [FREE Full text] [doi: 10.1002/ijgo.14381] [Medline: 35906840]
6. Toloo GS, Fitzgerald G, Aitken P, Verrall K, Tong S. Are heat warning systems effective? Environ Health 2013;12:27 [FREE Full text] [doi: 10.1186/1476-069X-12-27] [Medline: 23561265]
7. Ebi KL, Vanos J, Baldwin JW, Bell JE, Hondula DM, Errett NA, et al. Extreme weather and climate change: population health and health system implications. Annu Rev Public Health 2021;42:293-315 [FREE Full text] [doi: 10.1146/annurev-publhealth-012420-105026] [Medline: 33406378]

XSL•FO
RenderX

8. Brimicombe C, Runkle JD, Tuholske C, Domeisen DIV, Gao C, Toftum J, et al. Preventing heat-related deaths: the urgent need for a global early warning system for heat. PLOS Clim 2024;3(7):e0000437. [doi: 10.1371/journal.pclm.0000437]
9. Li T, Chen C, Cai W. The global need for smart heat-health warning systems. Lancet 2022;400(10362):1511-1512. [doi: 10.1016/S0140-6736(22)01974-2] [Medline: 36522205]
10. Hess JJ, Ebi KL. Iterative management of heat early warning systems in a changing climate. Ann N Y Acad Sci 2016;1382(1):21-30. [doi: 10.1111/nyas.13258] [Medline: 27788557]
11. Oberai M, Xu Z, Bach A, Forbes C, Jackman E, O'Connor F, et al. A digital heat early warning system for older adults. NPJ Digit Med 2025;8(1):114 [FREE Full text] [doi: 10.1038/s41746-025-01505-5] [Medline: 39972034]
12. Lumbroso D, Brown E, Ranger N. Stakeholders' perceptions of the overall effectiveness of early warning systems and risk assessments for weather-related hazards in Africa, the Caribbean and South Asia. Nat Hazards 2016;84(3):2121-2144. [doi: 10.1007/s11069-016-2537-0]
13. Toloo G, FitzGerald G, Aitken P, Verrall K, Tong S. Evaluating the effectiveness of heat warning systems: systematic review of epidemiological evidence. Int J Public Health 2013;58(5):667-681. [doi: 10.1007/s00038-013-0465-2] [Medline: 23564031]
14. Krah EF, de Kruijf JG. Exploring the ambivalent evidence base of mobile health (mHealth): a systematic literature review on the use of mobile phones for the improvement of community health in Africa. Digit Health 2016;2:2055207616679264 [FREE Full text] [doi: 10.1177/2055207616679264] [Medline: 29942576]
15. Trahan A, Walshe R, Mehta V. Extreme heat, gender, and access to preparedness measures: an analysis of the heatwave early warning system in Ahmedabad, India. Int J Disaster Risk Reduct 2023;99:104080 [FREE Full text]
16. Maliwichi P, Chigona W, Sowon K. Appropriation of mHealth interventions for maternal health care in sub-saharan Africa: hermeneutic review. JMIR Mhealth Uhealth 2021;9(10):e22653 [FREE Full text] [doi: 10.2196/22653] [Medline: 34612835]
17. Sheridan SC. A survey of public perception and response to heat warnings across four North American cities: an evaluation of municipal effectiveness. Int J Biometeorol 2007;52(1):3-15. [doi: 10.1007/s00484-006-0052-9] [Medline: 17024398]
18. Kc A, Maharjan S, Basnet O, Malla H, Gurung R, Pokharel SM, et al. Development, validation and reliability of scales and items for heat wave risk assessment of pregnant women. Int J Biometeorol 2024;68(11):2205-2214. [doi: 10.1007/s00484-024-02738-x] [Medline: 39207507]
19. Tandon A, Cobb B, Centra J, Izmailova E, Manyakov NV, McClenahan S, et al. Human factors, human-centered design, and usability of sensor-based digital health technologies: scoping review. J Med Internet Res 2024;26:e57628 [FREE Full text] [doi: 10.2196/57628] [Medline: 39546781]
20. Gao C, Heil C, Eggeling J, Alce G, Luchters S, Machingura F, et al. HIGH horizons study group. HIGH horizons - protocol for testing of the EWS among pregnant and postpartum women, young children and health workers. Zenodo 2024 [FREE Full text]
21. Gao C, Alce G, Heil C, van der Sanden K, Kingma B. HIGH Horizons Study Group. HIGH Horizons - MotherHeat Alert: an early warning system app for pregnant and postpartum women, infants and health workers. Zenodo. 2025. URL: https://doi.org/10.5281/zenodo.16992435 [accessed 2026-08-29]
22. Digital 2025: South Africa. Datareportal. URL: https://datareportal.com/reports/digital-2025-south-africa [accessed 2026-08-26]
23. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap) - a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
24. Jendritzky G, de Dear R, Havenith G. UTCI--why another thermal index? Int J Biometeorol 2012;56(3):421-428. [doi: 10.1007/s00484-011-0513-7] [Medline: 22187087]
25. Lewis JRJ. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int J Hum Comput Interact 1995;7(1):57-78. [doi: 10.1080/10447319509526110]
26. Gao Chuansi HC, Jakob E, Günter A, Stanley L, Fortunat M, Lameck K, et al. HIGH Horizons - protocol for testing of the EWS among pregnant and postpartum women, young children and health workers. Zenodo. 2024. URL: https://zenodo.org/records/13149904 [accessed 2026-09-13]
27. Solarin I, Scorgie F, Lamola RR, Madondo C, Lange IL, Munyewende P, et al. Through their lens: a photovoice study exploring pregnant and postpartum women's experiences, coping strategies and barriers to heat adaptation in an urban township in South Africa. Glob Public Health 2026;21(1):2612452 [FREE Full text] [doi: 10.1080/17441692.2025.2612452] [Medline: 41510673]
28. Peters E, Baker DP, Dieckmann NF, Leon J, Collins J. Explaining the effect of education on health: a field study in Ghana. Psychol Sci 2010;21(10):1369-1376. [doi: 10.1177/0956797610381506] [Medline: 20739672]
29. Xiong LC, Caparos S, Gueraud S, Augusto C, Habib M. Are health protective behaviors promoted by anticipated regret? J Health Psychol 2025;30(10):2798-2804. [doi: 10.1177/13591053241298516] [Medline: 39584493]
30. Khare S, Hajat S, Kovats S, Lefevre CE, de Bruin WB, Dessai S, et al. Heat protection behaviour in the UK: results of an online survey after the 2013 heatwave. BMC Public Health 2015;15:878 [FREE Full text] [doi: 10.1186/s12889-015-2181-8] [Medline: 26357923]

XSL•FO
RenderX

31. Akompab D, Bi P, Williams S, Grant J, Walker I, Augoustinos M. Heat waves and climate change: applying the health belief model to identify predictors of risk perception and adaptive behaviours in Adelaide, Australia. Int J Environ Res Public Health 2013;10(6):2164-2184 [FREE Full text] [doi: 10.3390/ijerph10062164] [Medline: 23759952]
32. Hoque MR, Rahman MS, Nipa NJ, Hasan MR. Mobile health interventions in developing countries: a systematic review. Health Informatics J 2020;26(4):2792-2810 [FREE Full text] [doi: 10.1177/1460458220937102] [Medline: 32691659]
33. Seager J. Disasters are gendered: what's new. Dordrecht: Springer; Jan 01, 2014:265-281.
34. Opoku D, Stephani V, Quentin W. A realist review of mobile phone-based health interventions for non-communicable disease management in sub-Saharan Africa. BMC Med 2017;15(1):24 [FREE Full text] [doi: 10.1186/s12916-017-0782-z] [Medline: 28162090]
35. Meyer AJ, Armstrong-Hough M, Babirye D, Mark D, Turimumahoro P, Ayakaka I, et al. Implementing mHealth interventions in a resource-constrained setting: case study from Uganda. JMIR Mhealth Uhealth 2020;8(7):e19552 [FREE Full text] [doi: 10.2196/19552] [Medline: 32673262]
36. Shen Z, Zhang L, Xiao X, Li R, Liang R. Icon familiarity affects the performance of complex cognitive tasks. Iperception 2020;11(2):2041669520910167 [FREE Full text] [doi: 10.1177/2041669520910167] [Medline: 32180935]
37. Casaló L, Flavián C, Guinalíu M. The role of perceived usability, reputation, satisfaction and consumer familiarity on the website loyalty formation process. Comput Human Behav 2008;24(2):325-345 [FREE Full text] [doi: 10.1016/j.chb.2007.01.017]
38. Thomas Craig KJ, Morgan LC, Chen C, Michie S, Fusco N, Snowdon JL, et al. Systematic review of context-aware digital behavior change interventions to improve health. Transl Behav Med 2021;11(5):1037-1048 [FREE Full text] [doi: 10.1093/tbm/ibaa099] [Medline: 33085767]
39. Boonruksa P, Maturachon T, Kongtip P, Woskie S. Heat stress, physiological response, and heat-related symptoms among Thai sugarcane workers. Int J Environ Res Public Health 2020;17(17):6363 [FREE Full text] [doi: 10.3390/ijerph17176363] [Medline: 32882881]
40. Clouse K, Vermund SH, Maskew M, Lurie MN, MacLeod W, Malete G, et al. Mobility and clinic switching among postpartum women considered lost to HIV care in South Africa. J Acquir Immune Defic Syndr 2017;74(4):383-389 [FREE Full text] [doi: 10.1097/QAI.0000000000001284] [Medline: 28225717]
41. Farao J, Malila B, Conrad N, Mutsvangwa T, Rangaka MX, Douglas TS. A user-centred design framework for mHealth. PLoS One 2020;15(8):e0237910 [FREE Full text] [doi: 10.1371/journal.pone.0237910] [Medline: 32813711]
42. Gannon H, Larsson L, Chimhuya S, Mangiza M, Wilson E, Kesler E, et al. Development and implementation of digital diagnostic algorithms for neonatal units in Zimbabwe and Malawi: development and usability study. JMIR Form Res 2024;8:e54274 [FREE Full text] [doi: 10.2196/54274] [Medline: 38277198]
43. Early Warnings for All. United Nations Office for Disaster Risk Reduction. 2022. URL: https://www.undrr.org/implementing-sendai-framework/sendai-framework-action/early-warnings-for-all [accessed 2026-09-18]
44. Kapwata T, Abdelatif N, Scovronick N, Gebreslasie MT, Acquaotta F, Wright CY. Identifying heat thresholds for South Africa towards the development of a heat-health warning system. Int J Biometeorol 2024;68(2):381-392 [FREE Full text] [doi: 10.1007/s00484-023-02596-z] [Medline: 38157021]
45. Ibrahim AM, Jahanfar S. Effectiveness and equity of mHealth apps for preeclampsia management in LMICs: a rapid review protocol. PLoS One 2024;19(11):e0313655 [FREE Full text] [doi: 10.1371/journal.pone.0313655] [Medline: 39536068]

Abbreviations

CHC: community health center
EU: European Union
EWS: early warning system
GEE: generalized estimating equations
GLMM: generalized linear mixed model
HIGH: Heat Indicators of Global Health
HREC: Human Research Ethics Committee
OR: odds ratio
PSSUQ: Post-Study System Usability Questionnaire
UTCI: Universal Thermal Climate Index

XSL•FO
RenderX

Edited by A Stone; submitted 29.May.2026; peer-reviewed by A Dozier, K Clouse; comments to author 10.Jul.2026; revised version received 21.Aug.2026; accepted 21.Aug.2026; published 25.Sep.2026.

Please cite as:
Heil CS, Sawry S, Frennert S, Le Roux J, Munyewende PO, Mkwananzi S, Solarin I, Watambwa P, van der Sanden K, Lange I, Alce G, Filippi V, Luchters S, Gao C
Mobile Heat Early Warning System for Pregnant and Postpartum Women in South Africa: Prospective Observational Evaluation Study
JMIR Mhealth Uhealth 2026;14:e102863
URL: https://mhealth.jmir.org/2026/1/e102863
doi:10.2196/102863
PMID:

©Clara Sofia Heil, Shobna Sawry, Susanne Frennert, Jean Le Roux, Pascalia Ozida Munyewende, Sibusiso Mkwananzi, Ijeoma Solarin, Perkins Watambwa, Koen van der Sanden, Isabelle Lange, Günter Alce, Veronique Filippi, Stanley Luchters, Chuansi Gao. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 25.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Tablet-Based Neurocognitive Tests for Ugandan Adolescents Living With HIV: Cross-Sectional Validation Study

Reuben N Robbins[1,2], PhD; Angel Nanteza[3,4], MPhil; Joy Louise Gumikiriza-Onoria[5], PhD; Anthony Santoro[2], PhD; Christopher M Ferraris[2,6], DSW; Christine Karungi[7], BS; Daphne Tsapalas[8], BA; Nana Asiedu[2], MPH; Courtney Kirsch[1,2], MA; Mei Tan[1,2], PhD; Jun Liu[2], PhD; Itziar Familiar-Lopez[9], MPH, MD, PhD; Victor Musiime[7,10], MBChB, MMED, PhD; Sahera Dirajlal-Fargo[11,12], DO

[1]Department of Psychiatry, Vagelos College of Physicians and Surgeons, Columbia University, 1051 Riverside Dr, New York, NY, United States

[10]Department of Paediatrics, College of Health Sciences, Makerere University, Kampala, Central Region, Uganda

[11]Feinberg School of Medicine, Northwestern University, Chicago, Illinois, Philippines

[12]Lurie Children's Hospital, Chicago, IL, United States

[2]New York State Psychiatric Institute, New York, NY, United States

[3]Uganda-CWRU Research Collaboration, Kampala, Uganda

[4]Butabika Hospital, Kampala, Central Region, Uganda

[5]Department of Psychiatry, College of Health Sciences, Makerere University, Kampala, Central Region, Uganda

[6]School of Social Policy & Practice, University of Pennsylvania, Philadelphia, PA, United States

[7]Joint Clinical Research Centre, Kampala, Central Region, Uganda

[8]University of California, Irvine, Irvine, CA, United States

[9]Department of Psychiatry, Michigan State University, Lansing, MI, United States

Corresponding Author:
Reuben N Robbins, PhD
Department of Psychiatry, Vagelos College of Physicians and Surgeons, Columbia University, 1051 Riverside Dr, New York, NY, United States

Abstract

Background: Adolescents living with perinatally acquired HIV (PHIV) remain at risk of neurocognitive difficulties, yet access to valid neurocognitive tests is limited in sub-Saharan Africa (SSA), where most of this population resides. Tests that can be administered by nonspecialist personnel are culturally and linguistically appropriate for SSA's diverse populations and are brief enough for routine use are needed to expand access, but few such tools exist.

Objective: We adapted *NeuroScreen*, a brief tablet-based neurocognitive test battery, into Luganda and evaluated its validity and acceptability for use among Ugandan adolescents with PHIV.

Methods: In this cross-sectional validation study, 100 Ugandan adolescents aged 11 to 20 years (50 with PHIV; 50 HIV-negative controls) completed the Luganda- or English-language *NeuroScreen* battery and a commonly used traditional test battery. Participants were recruited from an existing longitudinal cardiovascular cohort at the Joint Clinical Research Centre in Kampala, Uganda. We examined convergent validity with correlations between the 2 batteries' individual and global *z*-scores, known-groups validity with unadjusted and age-adjusted group comparisons, criterion validity by examining classification agreement between the 2 batteries quantified with Cohen κ, and acceptability through self-reported technology use and ease of use.

Results: *NeuroScreen* and the traditional battery global *z*-scores were strongly correlated (r=0.75; P<.001). Correlations between tests assessing similar domains were generally small to moderate, and were strongest for processing speed, attention, and working memory. Adolescents with PHIV scored significantly lower than controls on global scores for both *NeuroScreen* (P=.008) and the traditional battery (P=.003), even after adjusting for age. Agreement between batteries in classifying participants ≤1 SD (κ=0.56) and ≤2 SD (κ=0.66) below the global *z*-score mean was moderate, with both batteries classifying more adolescents with PHIV than controls below these thresholds. Although fewer than half of participants (42/97, 43%) had previously used a tablet, most rated *NeuroScreen* as easy to use, and prior tablet experience was not independently associated with *NeuroScreen* performance after age adjustment.

Conclusions: The Luganda-adapted *NeuroScreen* demonstrated convergent, known-groups, and criterion validity in Ugandan adolescents with PHIV. By delivering automated administration, automated scoring, and standardized multimedia instructions in the local language on a portable tablet, *NeuroScreen* addresses several persistent barriers to neurocognitive assessment in low-resource settings, including shortages of specialist personnel, the time burden of traditional batteries, and the scarcity of

XSL•FO
RenderX

culturally and linguistically appropriate tools. The *NeuroScreen* battery was found to be highly acceptable among adolescents with little or no prior tablet experience. These findings add to a growing body of evidence supporting tablet-based assessment as a practical and scalable mobile health approach to expanding access to neurocognitive testing. Future research should evaluate test-retest reliability, sensitivity to change over time, normative benchmarks, and implementation within real-world clinical workflows.

(JMIR Mhealth Uhealth 2026;14:e103288) doi:10.2196/103288

KEYWORDS

HIV infection; neuropsychological tests; adolescent; Uganda; mobile applications; neurocognitive disorders; mobile health; mHealth; validity; NeuroScreen; psychometrics

Introduction

Globally, there are approximately 1.57 million (1,200,000-1,900,000) adolescents (10-19 y) living with HIV, most with perinatally acquired HIV (PHIV) and living in low- and middle-income countries (LMICs), with approximately 85% living in sub-Saharan Africa (SSA) [1,2]. Greater access to more effective antiretroviral therapy (ART) has greatly improved their health outcomes [3,4], allowing adolescents with HIV to live well into adulthood [5,6]. Despite this public health success, adolescents with PHIV remain at high risk for a variety of complications, including neurocognitive problems [7,8].

As many as 46% of adolescents with PHIV experience neurocognitive problems [9], despite viral suppression [10,11]. Although both the persistence of HIV and its downstream effects in the central nervous system are thought to play a key role in these complications [12], low socioeconomic status, limited educational opportunities, early life adversity, and other comorbidities are also likely contributing factors [11,13,14]. PHIV typically affects the neurocognitive domains of processing speed and executive functioning [9]. Neurocognitive problems have been associated with poor academic performance, suboptimal ART adherence, increased mental health problems, poor decision-making, and engagement in HIV transmission and other risk behaviors (eg, unprotected sex) [11,15-29].

Early detection of neurocognitive problems can allow for ART regimen adjustments, tailored support approaches, better coping, and earlier intervention [30]. Without timely intervention, these problems may accumulate, limiting adolescents' future independence, employment potential, and quality of life [31]. However, detection of neurocognitive problems is challenging in LMICs for several reasons [32,33]. First, many LMICs, including those in SSA, have a limited number of psychologists and/or neuropsychologists trained to administer, score, and interpret traditional neuropsychological test batteries to measure and diagnose neurocognitive problems [34]. For example, in Uganda, where the current study took place, there are only 0.04 psychologists per 100,000 people, compared with 31.3 per 100,000 in the United States [35,36]. The precise number of neuropsychologists and/or psychologists with specialized neuropsychology training in Uganda is not known.

Second, access to culturally and linguistically appropriate neurocognitive tests is limited. Most tests to date were designed for use by English speakers in high-income countries and, as such, may only be available in English, which may be inappropriate for other cultures and contexts [37-39]. While English is often considered an official language of government and taught in schools in many LMICs, including Uganda, many communities rely on their indigenous languages and local dialects for everyday activities and may not be fully fluent in the officially recognized languages [37,40]. Administering English-language tests in communities that prefer or rely on other languages not only produces invalid test results but also violates ethical practices for psychologists and standardized testing practices [41,42]. Finally, traditional test batteries often require multiple hours to complete—time and resources that overburdened and resource-constrained health care systems in LMICs do not have.

Computerized neurocognitive tests, particularly those that are tablet-based, show promise for use in LMICs and may help make testing more accessible to LMIC populations [43-45]. Compared with traditional tests, computerized tests are highly automated, and many have been designed for administration by a wide variety of personnel who do not have specialized training. Tablet-based tests are also highly portable, and tablets are considerably more affordable than desktop or laptop computers [46,47].

NeuroScreen is a tablet-based battery of neurocognitive tests designed for resource-limited settings [48-50]. It has been validated for use in adult populations with HIV in South Africa [48] and among patients with first-episode psychosis in Uganda [51], as well as with adolescents and young adults with PHIV in Thailand [49] and the United States [39]. This cross-sectional validation study evaluated the construct validity (ie, convergent validity, known-groups validity, and criterion validity), as well as acceptability, of a Luganda-language–adapted *NeuroScreen* for use among adolescents in Uganda living with PHIV. We hypothesized that performance on the *NeuroScreen* battery would be positively correlated with performance on a traditional battery of neurocognitive tests assessing similar domains, that adolescents with PHIV would perform worse than controls on both batteries, and that the 2 batteries would show agreement in the proportion of participants scoring ≤1 SD and ≤2 SD below the global mean.

Methods

NeuroScreen Translation Process From English to Luganda

Prior to data collection, adaptation of the English *NeuroScreen* test instructions and content into Luganda was completed. We

XSL•FO
RenderX

followed a "group consensus" [52] process developed in our laboratory and informed by the International Test Commission [53,54], the American Educational Research Association [41], Nguyen et al [55], and Brislin [56], along with our research team's experience adapting neurocognitive and other measures cross-culturally. Our process began with generating a set of forward translations (English to Luganda) by a professional translator. Next, 4 bilingual (English and Luganda) individuals (1 Master's-level psychologist, 2 research psychometrists, and 1 research assistant) with experience administering traditional neurocognitive tests in Uganda completed blinded, individual back-translations. Back-translators were asked to make notes and identify any Luganda translations that were not easily understood. Then the US-based team reviewed the back-translations to identify differences that could affect the meaning and understandability of the instructions (ie, consistency with the original English and across the back-translations), including any translation issues identified by the translators. A series of 4 meetings with the United States and Ugandan teams was convened to review all translations, with special attention paid to those with identified issues. Final Luganda translations were generated after consensus of the translators was achieved. Some examples of issues addressed were overly technical words and English words without an equivalent in Luganda. For example, the Luganda word for circle is "obukulungwa" but was not readily known by any of the bilingual translators. During the group consensus, the bilingual translators reported that they could infer the word from the context in which it was used (instruction sets for the Trail Making tests) but had not encountered it before. Two of the bilingual translators reported looking up the word in a Luganda dictionary and confirming that it meant circle. Hence, the team decided that an Anglicized version, "esaako," would be much more widely understood by Luganda speakers. Once translations were finalized, audio recordings were made, synced with video instructions, and added to the *NeuroScreen* app.

Participants

In this cross-sectional validation study with a known-groups comparison, participants were 100 Ugandan adolescents (11-20 y): 50 living with PHIV and 50 HIV-negative controls recruited from the same communities as those with PHIV. All participants were recruited from a longitudinal study of cardiovascular health in adolescents with PHIV (K23HD088295; PI: SDF). Participants were included if they were (1) 11 to 20 years old, (2) enrolled in the cardiovascular study, and (3) had documented PHIV or HIV-negative status; participants with PHIV were additionally required to (4) have been on an ART regimen for the last 6 months, and (5) be virally suppressed (<200 copies/mL) at the time of enrollment in the parent study. For the primary correlation of interest (global *z*-scores between the 2 batteries), a sample size of 100 provides ≥80% power to detect a correlation of ≥0.28 in absolute value (a small effect size).

Procedures

This study was carried out at the Joint Clinical Research Centre (JCRC) in Kampala, Uganda, between February 2021 and November 2022. Parents or caregivers of potential participants were approached by study counselors at the JCRC and given information about the study. Trained research assistants collected participants' sociodemographic information before administering the *NeuroScreen* battery on a Samsung Galaxy Tab A 10.5-inch tablet. Participants were given the option to complete *NeuroScreen* in their language of preference, English or Luganda. After completing *NeuroScreen* and taking a short break, participants completed a traditional battery of pencil-and-paper neurocognitive tests and questionnaires.

Ethical Considerations

This study was reviewed and approved by the Uganda National Council for Science and Technology (#HS2266), the JCRC Research Ethics Committee (Uganda), the New York State Psychiatric Institute Institutional Review Board (IRB #8029), and the University Hospitals Cleveland Medical Center IRB (#11-16-18). All procedures were conducted in accordance with the ethical standards of these committees and with the 1964 Declaration of Helsinki and its later amendments. This study was reported according to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines. The completed STROBE checklist is provided as Checklist 1.

Written informed consent was obtained from all participants aged 18 years and older. For participants younger than 18 years, written informed consent was obtained from a parent or caregiver, and written assent was obtained from the adolescent. Participants and caregivers were informed that participation was voluntary, that they could decline to answer any question, and that they could withdraw from the study at any time without penalty or loss of benefits to which they were otherwise entitled. All study data were deidentified and stored using unique study identification numbers rather than personal identifiers. Data were kept on password-protected, access-restricted secure servers available only to authorized members of the study team.

Participants were compensated 186,000 UGX (approximately US $50 at the time of the study) for their time and travel. The amount was determined in accordance with local norms and approved by the reviewing ethics committees.

Measures

Sociodemographics

Sociodemographic information from caregivers and participants included participants' age, sex, years of education, and household income. Current HIV medical data were obtained from participants' research records and included the most recent viral load and CD4 cell count, as well as the nadir CD4 cell count and years on ART.

Technology Use and NeuroScreen Acceptability

Experience and comfort using digital technologies (ie, desktops and laptops, smartphones with touchscreens, and tablets) were collected with a 12-item questionnaire that asked participants about their experience with computers, smartphones, and tablets (eg, "Have you ever used a smartphone with a touchscreen before?," "How often do you use a smartphone with a touchscreen?," and "How comfortable are you using a smartphone with a touchscreen?"). Responses were either yes/no, frequency-based (eg, every day, 1 - 3 times per week,

XSL•FO
RenderX

or less than once a month), or on a 6-point Likert-type scale (eg, very comfortable to very uncomfortable).

NeuroScreen

NeuroScreen was administered on Samsung Galaxy Tab A 10.5-inch tablets running Android version 10 or 11 by either our Ugandan psychologist (JLG-O) or our research assistant (AN). Both received approximately 2 hours of direct training in the administration of *NeuroScreen*, in addition to a review of the user manual and multiple (~10) practice assessments. *NeuroScreen* was administered in a private and quiet office space and took, on average, 25 minutes to complete. The test battery is embedded in a graphical user interface that allows the examiner to enter participant data and administer the tests. To reduce potential bias, the examiner read standardized test instructions presented on the screen for the verbal learning and memory and verbal working memory tests; embedded instructional videos in the local language were played for the remaining tests. The app automatically cycled through the battery of tests in the same sequence and forced delayed recall of the verbal memory test after a 5-minute delay. After completing all subtests, results were automatically uploaded to a secure server. Each test is automatically scored, thus reducing potential bias. Each test yielded a primary score that was used for analysis. The battery (Table 1) included 5 tests of processing speed, 1 test of executive functioning, 1 test of verbal learning and delayed recall, 1 test of motor speed (finger tapping with dominant and nondominant hands), and 1 test of working memory (number span forward and backward) for a total of 12 scores. For details, see Robbins et al [49].

Table . *NeuroScreen* and traditional test batteries by domain and test.

Domain	*NeuroScreen* battery	Traditional battery
Learning	• Verbal list learning (VLL)	• KABC-II[a]: Atlantis (ATL) [57] • KABC-II: Hand movements (HM) [57]
Memory	• Verbal list learning delayed recall (VLLD)	• KABC-II: Atlantis delayed (ATLD) [57]
Working memory	• Number span forward (NSP-NF) • Number span backward (NSPNB)	• WISC-IV: Digit span forward (DSF) [58] • WISC-IV: Digit span backward (DSB) [58]
Executive functioning	• Trail Making Test 2 (TM2)	• Children's Color Trails Test 2 (CCTT2) [59] • KABC-II: Pattern reasoning (PR) [57] • KABC-II: Rover (RV) [57]
Processing speed	• Visual discrimination 1 (VD1) • Visual discrimination 2 (VD2)	• WISC-IV: Symbol search (SS) [58] • WISC-IV: Coding (CD) [58]
Motor speed	• Trail Making Test 1 (TM1) • Number speed test (NSPD) • Trail Making Test 3 (TM3) • Finger tapping dominant hand (FTD) • Finger tapping nondominant hand (FTND)	• CCTT1 [59]

[a]KABC-II: Kaufman Assessment Battery for Children, Second Edition.

Traditional Test Battery

Participants were administered a battery of traditional tests from commonly used neuropsychological test batteries that have been used in research studies with Ugandan adolescents (Table 1) [60-62]. The tests were administered on paper and scored by either a Ugandan Master's-level clinical psychologist (JLG-O) with expertise in neuropsychological assessment or a trained research assistant (AN) supervised by the psychologist. The research assistant received individualized training in the administration and scoring of the traditional battery by our Ugandan psychologist and a postdoctoral clinical psychology fellow with advanced training in clinical neuropsychology (AS). Individualized training included review of test manuals, observation of assessments, practice assessments, and directly supervised assessments (by JLG-O). Only after approval from the Ugandan psychologist was the research assistant allowed to complete assessments without directly observed supervision. To reduce potential bias, only standardized instruction sets were read, and quality assurance methods were conducted throughout the entirety of the study, such that all completed test protocols were reviewed by the US-based postdoctoral fellow (AS). Tests were administered in a private, quiet office space, with test

XSL•FO
RenderX

instructions delivered to children in the local language, and sessions took about 120 to 180 minutes to complete.

The measures assessed 6 neurocognitive domains of functioning: processing speed, working memory, learning, memory, motor speed, and executive functioning. Because we anticipated that most participants would be younger than 18 years and closer to 16 years, we decided to use tests for the younger age range so that all participants would take the same tests. The published upper age limits of the tests are as follows: Kaufman Assessment Battery for Children, Second Edition (KABC-II) [57]—18 years, Children's Color Trails Test (CCTT) [59]—17 years, and Wechsler Intelligence Scales for Children, Fourth Edition (WISC-IV) [58]—16 years. All participants received the same test battery regardless of age; 34 were older than 16 years, including 17 older than 17 years, and 11 older than 18 years. To assess potential ceiling effects among participants older than the upper age limit of these tests, scatterplots were generated and inspected (no evidence of ceiling effects for participants older than the stated age cap for each test was observed).

Data Analysis

Univariate statistics were used to examine participant demographics, technology use, and acceptability. Prior to analyses, sample-based *z*-scores were computed for each test in both batteries; completion-time raw scores for Trail Making 1, 2, and 3, Number Speed, CCTT1, and CCTT2 were reverse-coded so that higher scores indicated better performance. Global *z*-scores were computed by averaging all available *z*-scores for each test in each battery for each participant. To evaluate the convergent validity of *NeuroScreen*, Pearson correlation coefficients (and 95% CI) were computed between the global *z*-scores and each test within each battery. Kendall rank correlation coefficients were also computed for nonnormally distributed tests; however, they did not affect the overall interpretation of the data, and thus, we report only the Pearson correlation coefficients. HIV status served as the grouping variable, and age (in years), sex (male or female), and education (in years) as covariates in adjusted analyses. Known-groups validity was examined with between-HIV group comparisons across each test in each battery, as well as the global *z*-scores, using 2-tailed *t* tests and age-adjusted analyses of covariance. To examine whether prior tablet experience influenced *NeuroScreen* performance, we compared participants who had previously used a tablet with those who had never used a tablet before using *NeuroScreen* across *NeuroScreen* tests and the global *z*-score, with and without adjustment for age and false discovery rate (FDR) correction. Criterion validity was assessed using Cohen κ by comparing the proportion of participants scoring ≤1 SD and ≤2 SD below the global *z*-score of each battery. Between-battery analyses were restricted to participants with complete data on all tests in both batteries (n=95), and missing data were not imputed. To control for multiple comparisons, the FDR [63] was calculated for each battery's between-group comparisons. Data were analyzed using R 4.4.1 (R Core Team, 2024) [64].

Results

Participant Characteristics

One hundred participants were included in the final analysis. Participants' mean age (M_{age}) was 16.6 (SD 1.97; minimum=12.24, maximum=20.80) years; 52% (n=52) were male participants; 92% (n=92) were right-handed. Participants' mean years of education were 8 (SD 2.36) years. PHIV and control participants did not differ significantly on any demographic variables, except for age (*P*=.03), with the PHIV group being almost 1 year older, on average, than the control group (Table 2). Just about two-thirds of participants in each group chose to complete the Luganda version of *NeuroScreen* over the English version. Five participants were excluded from between-battery analyses owing to incomplete data on one or more tests (4 participants were missing one test on the *NeuroScreen* battery, and 1 participant was missing 2 tests on the traditional battery; these were due to errors in administration). Excluded and included participants did not differ in age, sex, education, or HIV status. For between-battery comparisons, a final total sample of 95 was used, which represented complete data on all tests for both batteries. The mean completion time for the *NeuroScreen* battery was 24.69 (SD 2.80) minutes, with the shortest completion time logged as 20 minutes and the longest as 39 minutes.

XSL•FO
RenderX

Table . Participant characteristics.

Variable	Full sample (n=100)	PHIV[a] (n=50)	Controls (n=50)	*t* test or chi-square test (*df*)	*P* value	*d*
Sex (male), n/N (%)	52/100 (52)	27/50 (54)	25/50 (50)	0.04 (1)[f]	.84	—[b]
Age (y), mean (SD)	16.16 (1.97)	16.60 (2.04)	15.72 (1.81)	2.27 (98)[g]	.03	−0.46
Education (y), mean (SD)	8.05 (2.35)	7.86 (2.70)	8.24 (1.95)	−0.81 (98)[g]	.42	0.16
Monthly income (UGX[c]), mean (SD)	332,597 (311,843)	273,415 (254,181)	400,000 (358,536)	−1.80 (75)[g]	.08	0.41
Living in extreme poverty[d], n/N (%)	34/75 (45)	20/41 (49)	14/36 (39)	0.76 (1)[f]	.38	—
Ever used tablet before[e], n/N (%)	42/97 (43)	17/50 (34)	25/47 (53)	3.63 (1)[f]	.06	—
Own a tablet, n/N (%)	4/97 (4)	1/50 (2)	3/47 (6)	0.33 (1)[f]	.57	—
Frequency of tablet use, mean (SD)	2.12 (1.45)	2.23 (1.57)	2.00 (1.38)	−0.64 (40)[g]	.53	−0.20
Ease of using tablet[e], mean (SD)	1.77 (0.79)	1.82 (0.85)	1.72 (0.72)	0.64 (94)[g]	.53	−0.13
NeuroScreen language (Luganda), n (%)	66/100 (66)	32/50 (64)	34/50 (68)	0.04 (1)[f]	.83	—

[a]Among participants with perinatally acquired HIV (PHIV), the mean age at ART initiation was 4.79 (SD 3.09) years, the mean nadir CD4 count was 713 (SD 520) cells/mm³, the mean most recent CD4 count was 1048 (SD 452) cells/mm³, and 44/47 (94%) had a most recent viral load of <200 copies/mL.

[b]Not applicable.

[c]1 US$=3781 UGX at the time of writing.

[d]Extreme poverty defined as <2 US$ per day.

[e]Three participants were missing data on technology use.

[f]Chi-square test.

[g]*t* test.

Technology Use

Table 2 describes the sample's experience using tablet devices. Less than half of the sample (42/97, 43%) had ever used a tablet before using the *NeuroScreen* tablet on the day of their study visit. More controls (25/47, 53%) had used tablets than adolescents with PHIV (17/50, 34%), but this was not statistically significant. Of the 42 participants who reported having used a tablet before, 4 (4%) reported currently owning a tablet. Despite limited tablet use and ownership, the mean ease of use rating for the *NeuroScreen* tablet was in the "very easy" to "easy" range (mean 1.79, SD 0.79). There were no differences between groups on this rating. Just under half the sample (48/97, 49%) had ever used a desktop or laptop computer before, and 91% (88/97) reported having used a smartphone with a touchscreen before. There were no between-group differences in either computer or smartphone use. In an unadjusted analysis examining test performance between those who had ever used a tablet and never used a tablet, participants who had used a tablet before performed significantly better on the visual discrimination 1 and visual discrimination 2 tests (t_{91}=−2.36; P=.02 and t_{95}=−2.09; P=.04, respectively). However, after adjusting for age and FDR, no differences remained significant.

Convergent Validity

Pearson correlation coefficients were computed between the global *z*-scores and individual tests in each battery (Multimedia Appendix 1). The correlation between the 2 global *z*-scores was large and statistically significant (r=0.75; P<.001; 95% CI 0.64-0.83). Correlations between individual tests were generally in the small-to-moderate range, supporting convergent validity between *NeuroScreen* and the traditional battery.

For tests assessing processing speed, correlation coefficients ranged from 0.23 to 0.49. For example, *NeuroScreen* Trail Making 1 was moderately associated with the CCTT1 from the traditional battery (r=0.47; P<.001; 95% CI 0.29-0.61). Similarly, coding and symbol search from the traditional battery

XSL•FO
RenderX

showed moderate correlations with *NeuroScreen* visual discrimination 1 (r=0.33; P=.001; 95% CI 0.14-0.5; and r=0.44; P=.001; 95% CI 0.26-0.59, respectively), visual discrimination 2 (r=0.40; P<.001; 95% CI 0.22-0.55 and r=0.49; P<.001; 95% CI 0.32-0.62, respectively), and number speed (r=0.33; P=.001; 95% CI 0.14-0.49 and r=0.44; P<.001; 95% CI 0.26-0.58, respectively).

For working memory, *NeuroScreen* number span forward was moderately correlated with digit span forward (r=0.34; P=.001; 95% CI 0.15-0.5), as was *NeuroScreen* number span backward with digit span backward (r=0.41; P<.001; 95% CI 0.23-0.56). For learning and memory, *NeuroScreen* verbal learning was modestly correlated with the Atlantis paired-associate learning (r=0.32; P=.001; 95% CI 0.13-0.49), whereas the correlation between delayed memory measures was weaker, with *NeuroScreen* verbal delayed recall showing only a small, nonsignificant association with Atlantis delayed recall (r=0.13; P=.21; 95% CI −0.07 to 0.31). Verbal learning was moderately correlated with hand movements (r=0.36; P<.001; 95% CI 0.18-0.52). For executive function tests, *NeuroScreen* Trail Making 2 and CCTT2 were moderately correlated (r=0.37; P<.001; 95% CI 0.19-0.53). None of the other traditional battery executive function tests correlated with the *NeuroScreen* Trail Making 2 test. The *NeuroScreen* battery did not contain any problem-solving or reasoning tests equivalent to the KABC-II Rover and pattern reasoning subtests; those 2 tests had mostly small correlations with the *NeuroScreen* processing speed and working memory tests. The traditional battery also did not have a pure motor speed test like *NeuroScreen*'s finger tapping, which had small correlations with traditional tests of processing speed, working memory, and learning.

Known-Groups Validity

We examined whether the *NeuroScreen* and traditional tests, and their global *z*-scores, detected performance differences between groups by HIV status. Across both batteries, participants with PHIV showed poorer performance than controls, although not all test-level differences were statistically significant (Multimedia Appendix 2). This pattern was also evident in the global scores: on the traditional battery, participants with PHIV had a significantly lower mean global *z*-score than controls (mean −0.17, SD 0.62 vs mean 0.17, SD 0.51; P=.003; M_{diff} =−0.34; 95% CI −0.57 to −0.12). *NeuroScreen* also showed the same significantly lower mean global *z*-score (mean −0.15, SD 0.60 vs mean 0.12, SD 0.52; P=.02; M_{diff}=−0.27; 95% CI −0.50 to −0.04).

Domain-level findings were broadly consistent across the traditional and *NeuroScreen* batteries (Multimedia Appendix 2). Processing speed showed the strongest cross-battery differentiation, with participants with PHIV performing worse than controls on 4 of 8 measures overall (P=.001 to P=.02). In working memory, both batteries showed the same general pattern of lower performance among participants with PHIV, although no between-group differences were statistically significant. For learning and memory, *NeuroScreen* and traditional measures yielded parallel results, with poorer PHIV performance in unadjusted analyses (learning P=.04 to P=.047; memory P=.04), although these differences were attenuated after adjustment. In motor speed, neither battery differentiated the groups (all P≥.34). In executive functioning, both batteries detected poorer PHIV performance on the trail-making switching-type tests in unadjusted analyses, although only the traditional battery difference was statistically significant (P=.07 and P=.01, respectively). After adjustments, the between-group difference on the traditional test was attenuated.

Criterion Validity

We examined classification agreement between batteries and HIV groups for participants with global *z*-scores ≤−1 SD and ≤−2 SD below the mean. Using the traditional battery, 15 of 95 participants were classified as ≤−1 SD, compared to 11 of 95 on *NeuroScreen*. Across both batteries, more participants living with PHIV than HIV-negative controls were classified as having global *z*-scores ≤−1 SD. The traditional battery classified significantly more participants with PHIV as ≤−1 SD (13 PHIV and 2 controls; χ^2_1=9.86; P=.001) than the *NeuroScreen* battery (7 PHIV and 4 controls; χ^2_1=1.0, P=.32). Agreement between batteries was as follows: 8 participants were classified as ≤−1 SD by both batteries, 7 were identified only by the traditional battery, and 3 only by *NeuroScreen*. This yielded a κ coefficient of 0.56 (95% CI 0.31-0.80).

At the global *z*-score ≤−2 SD below the mean level, the traditional battery classified 2 of 95 participants as ≤−2 SD, compared with 4 of 95 on *NeuroScreen*; neither difference was statistically significant. Agreement between batteries was as follows: 2 participants were classified as ≤−2 SD by both batteries; 2 additional participants were classified as ≤−2 SD by *NeuroScreen* only; no participants were classified as ≤−2 SD only by the traditional battery. This yielded a κ coefficient of 0.66 (95% CI 0.22-1.0).

Discussion

In this study, we adapted *NeuroScreen* for use with English- and Luganda-speaking Ugandan adolescents and examined its construct and criterion validity compared with a widely used traditional test battery. Overall, the findings support the use of *NeuroScreen* as a brief, tablet-based approach for assessing neurocognitive functioning in this population. Global performance on *NeuroScreen* was strongly correlated with global performance on the traditional battery, and adolescents with PHIV performed more poorly than HIV-negative controls on both batteries. Tests assessing similar domains had small to moderate correlations.

Convergent validity was most clearly supported by the strong, statistically significant correlation between the global *z*-scores derived from *NeuroScreen* and the traditional battery (r=0.75). Correlation coefficients between conceptually similar processing speed tests were among the highest (r=0.37-0.48), with conceptually similar tests of attention and working memory having similar correlations (r=0.30-0.47), and a conceptually similar executive functioning test showing a moderate correlation (r=0.34). These findings are encouraging because an exact one-to-one correspondence between computerized and traditional tests is not necessarily expected, even when they are intended to assess similar domains [65,66]. Differences in

XSL•FO
RenderX

modality, response demands, timing parameters, and stimulus format may account for the modest, rather than large, test-level correlations [65-67].

Furthermore, the modest test-level correlations are likely a function of the *NeuroScreen* tests being considerably shorter than the traditional battery tests (eg, coding has a time limit of 120 s vs 45 s for the *NeuroScreen* visual discrimination 2 test, and thus the latter yields fewer scorable responses). Hence, each *NeuroScreen* test comprises fewer items than its traditional counterpart and therefore, by classical test theory, carries less measurement information and lower reliability, meaning briefer tests will correlate more weakly with longer traditional tests even when both assess closely related constructs [68-70]. This is the pattern we observed, and for this reason, the global scores, which aggregate scores across many subtests and thus represent greater length and reliability, are a more appropriate index of convergent validity. This also carries an important implication for *NeuroScreen*. The approximately 25-minute *NeuroScreen* battery shares approximately 56% of the variance with the traditional battery (r=0.75 or r^2=0.56), which requires 120 to 180 minutes to administer (and more time to score and interpret). *NeuroScreen* captures much of the same signal as the traditional battery at a small fraction of the assessment burden—an efficiency especially valuable in low-resource settings.

NeuroScreen was able to distinguish performance between adolescents with PHIV and controls. Adolescents with PHIV had significantly worse global performance on *NeuroScreen* compared with controls. This finding was the same for the traditional battery. Adolescents with PHIV performed worse than controls on all but 2 individual tests and significantly worse on 4 tests in the *NeuroScreen* battery, compared with 6 tests in the traditional battery, and this pattern mostly remained after age adjustment. When adjusting for the FDR on unadjusted comparisons, only 3 of the traditional battery tests remained significant. When adjusting for the FDR on age-adjusted comparisons, only 2 of the traditional battery tests remained significant. Notably, the magnitude of the group differences was similar across batteries, suggesting that *NeuroScreen* was sensitive to the neurocognitive difficulties associated with PHIV in this sample. These findings are consistent with prior work from Uganda and other settings showing that youth with PHIV remain at risk for neurocognitive difficulties despite access to ART [61,62,71]. They also extend prior *NeuroScreen* validation work in PHIV-affected youth and Ugandan clinical populations by demonstrating similar performance with a Luganda-adapted version among Ugandan adolescents [39,49,51].

At the domain level, the most consistent differences between participants with PHIV and controls were observed in processing speed and executive functioning. This pattern is broadly in line with the literature on HIV-associated neurocognitive difficulties in youth, for whom slowed processing speed and executive inefficiencies are among the most commonly reported deficits [9,71,72]. In contrast, differences in learning and memory were significant in unadjusted analyses but were attenuated after age adjustment. Working memory differences were in the expected direction but did not reach statistical significance. These findings suggest that *NeuroScreen* may be especially useful for detecting the domains most consistently affected in PHIV, while also highlighting that not all cognitive domains are equally discriminating in a relatively small sample of virally suppressed youth. Importantly, the absolute differences in mean global *z*-scores between groups were modest, which may reflect the relatively good clinical status of this cohort, all of whom were virally suppressed. Even so, both batteries detected meaningful between-group differences, indicating that subtle but measurable neurocognitive vulnerability persists in this population.

The classification results provide additional support for criterion validity. Both batteries identified more participants with PHIV than controls at ≤1 SD and ≤2 SD below the sample mean global *z*-scores. Between-group differences in classification rates were only statistically significant in the traditional battery at ≤1 SD below the mean threshold. There was a moderate level of agreement between the batteries in their classifications. Given the modest sample size and the small number of participants falling below the classification thresholds, these analyses should be interpreted cautiously. Still, the convergence between the 2 batteries suggests that *NeuroScreen* may have utility not only for continuous assessment of neurocognitive performance but also for identifying youth who may warrant more comprehensive follow-up evaluation.

A major strength of *NeuroScreen* in this context is its feasibility for scale-up. Neurocognitive assessment in many low-resource settings is constrained by shortages of trained specialists, limited availability of culturally and linguistically appropriate tools, and the time burden of traditional batteries. *NeuroScreen* addresses several of these barriers by standardizing administration, automating scoring, and reducing administration time. In its current form, *NeuroScreen* only requires approximately 25 minutes to complete and provides information in line with traditional neuropsychological test batteries, which require considerably more time to complete and score (some batteries can require 2 - 6 h just for administration). This is particularly relevant in Uganda and similar LMIC settings, where specialist neuropsychology resources are scarce. The successful Luganda adaptation also underscores the importance of culturally and linguistically grounded test development. Rather than simply translating instructions literally, the adaptation process aimed to preserve meaning and comprehensibility for local users, which is essential for valid assessment [55].

Acceptability findings further support the promise of *NeuroScreen* for use in Uganda. Although many participants had never previously used a tablet, most reported that the app was easy to use. This is notable because familiarity with digital devices is often raised as a concern when introducing tablet-based assessments in low-resource settings [67,73,74]. In this sample, limited prior tablet experience did not appear to prevent successful completion of, or favorable impressions of, the assessment. These findings suggest that, with appropriate orientation and localized instructions, *NeuroScreen* can be acceptable even among youth with limited exposure to this specific technology platform.

Several limitations should be acknowledged. First, this was a relatively small cross-sectional study, which limited statistical

XSL•FO
RenderX

power, particularly for impairment classification analyses and finer-grained domain comparisons. Second, the traditional battery used as the comparator is not a true gold standard for neurocognitive assessment in this context. As in many LMIC settings, available traditional neuropsychological tests themselves have limitations related to cultural fit, language, and normative data. Accordingly, this study should be understood as evaluating *NeuroScreen* against a commonly used battery of tests rather than against a definitive reference standard. Relatedly, because no published normative data exist for adolescent populations in Uganda for either battery's tests, sample-based *z*-scores were used, which can inflate agreement. Third, because all participants with PHIV were virally suppressed and were recruited from an existing urban research cohort, findings may not generalize to youth with less stable HIV disease, those not engaged in care, or more socioeconomically diverse populations. Fourth, several of the traditional measures were administered to participants who exceeded the tests' published upper age limits (KABC-II, 18 y; CCTT, 17 y; WISC-IV, 16 y). Because we anticipated that most participants would be near 16 years of age and wanted every participant to complete an identical battery, we selected the younger-age forms for the full sample. Administering tests beyond their normed age ranges can affect the validity of age-referenced interpretation and may compress scores near the top of the range, potentially attenuating between-group differences. To assess this, we inspected scatterplots of raw scores against age for each affected test and found no evidence of ceiling effects among older participants. Because our analyses relied on sample-based *z*-scores rather than the tests' published norms, the influence of exceeding the normative age ranges on our findings is likely limited, though it cannot be fully excluded. Fifth, to optimize study workflow, we did not counterbalance the order of battery administration. As such, order, practice, or fatigue effects on the second (traditional) battery cannot be ruled out. Sixth, we note that the data were collected 4 to 5 years ago. Nonetheless, we do not believe this would have much bearing on interpretation of the findings because the study evaluates psychometric properties of the adapted tests, which are not time-sensitive. In addition, the versions of the *NeuroScreen* and the traditional battery tests have not changed. Furthermore, the clinical profile of the virally suppressed Ugandan adolescents with PHIV has remained largely unchanged. Moreover, the shortage of neuropsychology specialists and valid neuropsychological tests for Luganda-speaking populations also remains. Finally, because this study focused on validity and acceptability, additional work is still needed to establish clinical cut-points, longitudinal sensitivity, and predictive utility for functional outcomes such as school performance, medication adherence, and daily functioning.

Despite these limitations, the present findings have important implications. *NeuroScreen* offers a brief, scalable, and locally adapted approach to neurocognitive assessment that showed a similar pattern of group differences to a traditional neuropsychological test battery. In settings where the use of traditional neuropsychological test batteries is not feasible or practical, a tool such as *NeuroScreen* could expand access to screening and help identify youth who may benefit from monitoring, referral, or intervention. Future studies should examine its performance in larger and more diverse samples, assess test-retest reliability and sensitivity to change over time, and determine how best to integrate *NeuroScreen* into routine HIV care for adolescents in SSA.

Acknowledgments

The authors would like to thank the participants and their caregivers who participated in this research. The authors would also like to thank Betty Nyangomo and Mary Nyakato for their assistance in the translation process. The authors used Claude Opus 4.8 to improve the readability of portions of the manuscript. All AI-assisted text was reviewed and edited by the authors, who take full responsibility for the content.

Funding

This work was supported by the National Institutes of Health (K23HD088295 to SDF; R21TW011683 to RNR; R01HD113450 to RNR, SDF, and VM; and P30MH043520). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

The data sets generated and analyzed during this study are not publicly available due to privacy and ethical restrictions on participant data, but are available in deidentified form from the corresponding author upon reasonable request, subject to the study's ethics approval and data-sharing agreements.

Authors' Contributions

Conceptualization: RNR, SD-F
Data curation: CMF, C Kirsch, JL
Formal analysis: JL
Funding acquisition: SD-F, RNR, VM
Investigation: AN, JLG-O, C Karungi
Methodology: RNR, SD-F, VM

XSL•FO
RenderX

Project administration: CMF, C Kirsch, C Karungi
Supervision: RNR, SD-F, VM
Writing – original draft: AN, CMF, C Kirsch, MT, RNR
Writing – review & editing: AN, JLG-O, AS, C Karungi, CMF, DT, C Kirsch, NA, MT, JL, VM, SD-F
All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Pearson correlation matrix between *NeuroScreen* and traditional battery tests.
[DOCX File, 19 KB - mhealth_v14i1e103288_app1.docx]

Multimedia Appendix 2
Mean test performance across all tests by HIV status.
[DOCX File, 13 KB - mhealth_v14i1e103288_app2.docx]

Checklist 1
STROBE checklist.
[DOCX File, 18 KB - mhealth_v14i1e103288_app3.docx]

References

1. AIDSinfo. 2023. URL: https://aidsinfo.unaids.org/ [accessed 2024-01-01]
2. Adolescent HIV prevention. UNICEF data (monitoring the situation of children and women). 2023. URL: https://data.unicef.org/topic/hivaids/adolescents-young-people/ [accessed 2024-01-01]
3. Romley JA, Juday T, Solomon MD, Seekins D, Brookmeyer R, Goldman DP. Early HIV treatment led to life expectancy gains valued at $80 billion for people infected in 1996-2009. Health Aff (Millwood) 2014 Mar;33(3):370-377. [doi: 10.1377/hlthaff.2013.0623] [Medline: 24590933]
4. Global report: UNAIDS report on the global AIDS epidemic 2013. : Joint United Nations Programme on HIV/AIDS (UNAIDS); 2013 URL: https://www.unaids.org/sites/default/files/media_asset/UNAIDS_Global_Report_2013_en_1.pdf [accessed 2026-09-10]
5. Khalifa A, Stover J, Mahy M, Idele P, Porth T, Lwamba C. Demographic change and HIV epidemic projections to 2050 for adolescents and young people aged 15-24. Glob Health Action 2019;12(1):1662685. [doi: 10.1080/16549716.2019.1662685] [Medline: 31510887]
6. Patel K, Hernán MA, Williams PL, et al. Long-term effectiveness of highly active antiretroviral therapy on the survival of children and adolescents with HIV infection: a 10-year follow-up study. Clin Infect Dis 2008 Feb 15;46(4):507-515. [doi: 10.1086/526524] [Medline: 18199042]
7. Alford K, Vera JH. Cognitive impairment in people living with HIV in the ART era: a review. Br Med Bull 2018 Sep 1;127(1):55-68. [doi: 10.1093/bmb/ldy019] [Medline: 29868901]
8. Wei J, Hou J, Su B, et al. The prevalence of Frascati-criteria-based HIV-associated neurocognitive disorder (HAND) in HIV-infected adults: a systematic review and meta-analysis. Front Neurol 2020;11:581346. [doi: 10.3389/fneur.2020.581346] [Medline: 33335509]
9. Phillips N, Amos T, Kuo C, et al. HIV-associated cognitive impairment in perinatally infected children: a meta-analysis. Pediatrics 2016 Nov;138(5):e20160893. [doi: 10.1542/peds.2016-0893] [Medline: 27940772]
10. Kerr SJ, Puthanakit T, Malee KM, et al. Increased risk of executive function and emotional behavioral problems among virologically well-controlled perinatally HIV-infected adolescents in Thailand and Cambodia. J Acquir Immune Defic Syndr 2019 Nov 1;82(3):297-304. [doi: 10.1097/QAI.0000000000002132] [Medline: 31335589]
11. Laughton B, Cornell M, Boivin M, Van Rie A. Neurodevelopment in perinatally HIV‐infected children: a concern for adolescence. J Int AIDS Soc 2013 Jan;16(1):18603. [doi: 10.7448/IAS.16.1.18603]
12. Zayyad Z, Spudich S. Neuropathogenesis of HIV: from initial neuroinvasion to HIV-associated neurocognitive disorder (HAND). Curr HIV/AIDS Rep 2015 Mar;12(1):16-24. [doi: 10.1007/s11904-014-0255-3] [Medline: 25604237]
13. Sherr L, Croome N, Parra Castaneda K, Bradshaw K, Herrero Romero R. Developmental challenges in HIV infected children—an updated systematic review. Child Youth Serv Rev 2014 Oct;45:74-89. [doi: 10.1016/j.childyouth.2014.03.040]
14. Smith R, Wilkins M. Perinatally acquired HIV infection: long-term neuropsychological consequences and challenges ahead. Child Neuropsychol 2015;21(2):234-268. [doi: 10.1080/09297049.2014.898744] [Medline: 24697320]
15. Donald KA, Hoare J, Eley B, Wilmshurst JM. Neurologic complications of pediatric human immunodeficiency virus: implications for clinical practice and management challenges in the African setting. Semin Pediatr Neurol 2014 Mar;21(1):3-11. [doi: 10.1016/j.spen.2014.01.004] [Medline: 24655398]

XSL•FO
RenderX

16. Elkington KS, Robbins RN, Bauermeister JA, Abrams EJ, McKay M, Mellins CA. Mental health in youth infected with and affected by HIV: the role of caregiver HIV. J Pediatr Psychol 2011 Apr;36(3):360-373. [doi: 10.1093/jpepsy/jsq094] [Medline: 20947561]
17. Ettenhofer ML, Hinkin CH, Castellon SA, et al. Aging, neurocognition, and medication adherence in HIV infection. Am J Geriatr Psychiatry 2009 Apr;17(4):281-290. [doi: 10.1097/JGP.0b013e31819431bd] [Medline: 19307857]
18. Ettenhofer ML, Foley J, Castellon SA, Hinkin CH. Reciprocal prediction of medication adherence and neurocognition in HIV/AIDS. Neurology 2010 Apr 13;74(15):1217-1222. [doi: 10.1212/WNL.0b013e3181d8c1ca] [Medline: 20220123]
19. Gadow KD, Angelidou K, Chernoff M, et al. Longitudinal study of emerging mental health concerns in youth perinatally infected with HIV and peer comparisons. J Dev Behav Pediatr 2012 Jul;33(6):456-468. [doi: 10.1097/DBP.0b013e31825b8482] [Medline: 22772819]
20. Garvie PA, Zeldow B, Malee K, et al. Discordance of cognitive and academic achievement outcomes in youth with perinatal HIV exposure. Pediatr Infect Dis J 2014 Sep;33(9):e232-e238. [doi: 10.1097/INF.0000000000000314] [Medline: 25361033]
21. Heaton RK, Franklin DR, Ellis RJ, et al. HIV-associated neurocognitive disorders before and during the era of combination antiretroviral therapy: differences in rates, nature, and predictors. J Neurovirol 2010 Feb;17(1):3-16. [doi: 10.1007/s13365-010-0006-1] [Medline: 21174240]
22. Herting MM, Uban KA, Williams PL, et al. Default mode connectivity in youth with perinatally acquired HIV. Medicine (Baltimore) 2015 Sep;94(37):e1417. [doi: 10.1097/MD.0000000000001417] [Medline: 26376381]
23. Hinkin CH, Castellon SA, Durvasula RS, et al. Medication adherence among HIV+ adults: effects of cognitive dysfunction and regimen complexity. Neurology 2002 Dec 24;59(12):1944-1950. [doi: 10.1212/01.wnl.0000038347.48137.67] [Medline: 12499488]
24. Hinkin CH, Hardy DJ, Mason KI, et al. Medication adherence in HIV-infected adults: effect of patient age, cognitive status, and substance abuse. AIDS 2004 Jan 1;18(Suppl 1):S19-S25. [doi: 10.1097/00002030-200418001-00004] [Medline: 15075494]
25. Mellins CA, Brackis-Cott E, Leu CS, et al. Rates and types of psychiatric disorders in perinatally human immunodeficiency virus-infected youth and seroreverters. J Child Psychol Psychiatry 2009 Sep;50(9):1131-1138. [doi: 10.1111/j.1469-7610.2009.02069.x] [Medline: 19298479]
26. Mellins CA, Tassiopoulos K, Malee K, et al. Behavioral health risks in perinatally HIV-exposed youth: co-occurrence of sexual and drug use behavior, mental health problems, and nonadherence to antiretroviral treatment. AIDS Patient Care STDS 2011 Jul;25(7):413-422. [doi: 10.1089/apc.2011.0025] [Medline: 21992620]
27. Nachman S, Chernoff M, Williams P, Hodge J, Heston J, Gadow KD. Human immunodeficiency virus disease severity, psychiatric symptoms, and functional outcomes in perinatally infected youth. Arch Pediatr Adolesc Med 2012 Jun 1;166(6):528-535. [doi: 10.1001/archpediatrics.2011.1785] [Medline: 22312169]
28. Sirois PA, Chernoff MC, Malee KM, et al. Associations of memory and executive functioning with academic and adaptive functioning among youth with perinatal HIV exposure and/or infection. J Pediatric Infect Dis Soc 2016 Dec;5(suppl 1):S24-S32. [doi: 10.1093/jpids/piw046] [Medline: 27856673]
29. Smith R, Chernoff M, Williams PL, et al. Impact of HIV severity on cognitive and adaptive functioning during childhood and adolescence. Pediatr Infect Dis J 2012 Jun;31(6):592-598. [doi: 10.1097/INF.0b013e318253844b] [Medline: 22592486]
30. Wagstaff RA, Mullens AB, Daken K, et al. HIV associated neurocognitive disorder screening and diagnosis pathways in Australia: a scoping review and international implications. AIDS Care 2024 Aug;36(8):1029-1040. [doi: 10.1080/09540121.2024.2343768] [Medline: 38685763]
31. Cohen S, Ter Stege JA, Geurtsen GJ, et al. Poorer cognitive performance in perinatally HIV-infected children versus healthy socioeconomically matched controls. Clin Infect Dis 2015 Apr 1;60(7):1111-1119. [doi: 10.1093/cid/ciu1144] [Medline: 25516183]
32. Barber TJ, Bradshaw D, Hughes D, et al. Screening for HIV-related neurocognitive impairment in clinical practice: challenges and opportunities. AIDS Care 2014 Feb;26(2):160-168. [doi: 10.1080/09540121.2013.819401] [Medline: 23869661]
33. Vastag Z, Fira-Mladinescu O, Rosca EC. HIV-associated neurocognitive disorder (HAND): obstacles to early neuropsychological diagnosis. Int J Gen Med 2022;15:4079-4090. [doi: 10.2147/IJGM.S295859] [Medline: 35450033]
34. Rathod S, Pinninti N, Irfan M, et al. Mental health service provision in low- and middle-income countries. Health Serv Insights 2017;10:1178632917694350. [doi: 10.1177/1178632917694350] [Medline: 28469456]
35. Lin L, Conroy J, Christidis P. Datapoint: which states have the most licensed psychologists? American Psychological Association. 2020. URL: https://www.apa.org/monitor/2020/01/datapoint-states [accessed 2026-09-10]
36. Mental health ATLAS 2017 member state profile. : World Health Organization; 2018 URL: https://cdn.who.int/media/docs/default-source/mental-health/mental-health-atlas-2017-country-profiles/ug.pdf [accessed 2024-01-01]
37. Aghvinian M, Santoro AF, Gouse H, et al. Taking the test: a qualitative analysis of cultural and contextual factors impacting neuropsychological assessment of Xhosa-speaking South Africans. Arch Clin Neuropsychol 2021 Aug 31;36(6):976-980. [doi: 10.1093/arclin/acaa115] [Medline: 33236088]
38. Robbins RN, Brown H, Ehlers A, et al. A smartphone app to screen for HIV-related neurocognitive impairment. J Mob Technol Med 2014 Feb 1;3(1):23-26. [doi: 10.7309/jmtm.3.1.5] [Medline: 24860624]

XSL•FO
RenderX

39. Robbins RN, Kluisza L, Liu J, et al. Construct validity supports use of a novel, tablet-based neurocognitive assessment for adolescents and young adults affected by perinatal HIV from vulnerable communities in the United States. AIDS Behav 2021 Apr;25(4):1185-1191. [doi: 10.1007/s10461-020-03099-3] [Medline: 33180252]
40. Shakib MK. The position of language in development of colonization. J Lang Cult 2011;2(7):117-123. [doi: 10.5897/JLC.9000028]
41. Standards for educational and psychological testing. : American Educational Research Association; 2014 URL: https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf [accessed 2026-09-10]
42. Ethical principles of psychologists and code of conduct. American Psychological Association. 2017. URL: https://www.apa.org/ethics/code [accessed 2026-09-10]
43. McHenry MS, Balogun KA, McDonald BC, Vreeman RC, Whipple EC, Serghides L. In utero exposure to HIV and/or antiretroviral therapy: a systematic review of preclinical and clinical evidence of cognitive outcomes. J Int AIDS Soc 2019 Apr;22(4):e25275. [doi: 10.1002/jia2.25275] [Medline: 30983111]
44. Parsey CM, Schmitter-Edgecombe M. Applications of technology in neuropsychological assessment. Clin Neuropsychol 2013;27(8):1328-1361. [doi: 10.1080/13854046.2013.834971] [Medline: 24041037]
45. Wilson S, Milanini B, Javandel S, Nyamayaro P, Valcour V. Validity of digital assessments in screening for HIV-related cognitive impairment: a review. Curr HIV/AIDS Rep 2021 Dec;18(6):581-592. [doi: 10.1007/s11904-021-00585-8] [Medline: 34820750]
46. Lee H, Baniqued PL, Cosman J, et al. Examining cognitive function across the lifespan using a mobile application. Comput Human Behav 2012 Sep;28(5):1934-1946. [doi: 10.1016/j.chb.2012.05.013]
47. Possin KL, Moskowitz T, Erlhoff SJ, et al. The brain health assessment for detecting and diagnosing neurocognitive disorders. J Am Geriatr Soc 2018 Jan;66(1):150-156. [doi: 10.1111/jgs.15208] [Medline: 29355911]
48. Robbins RN, Gouse H, Brown HG, et al. A mobile app to screen for neurocognitive impairment: preliminary validation of NeuroScreen among HIV-infected South African adults. JMIR Mhealth Uhealth 2018 Jan 5;6(1):e5. [doi: 10.2196/mhealth.9148] [Medline: 29305338]
49. Robbins RN, Santoro AF, Ferraris C, et al. Adaptation and construct validity evaluation of a tablet-based, short neuropsychological test battery for use with adolescents and young adults living with HIV in Thailand. Neuropsychology 2022 Nov;36(8):695-708. [doi: 10.1037/neu0000851] [Medline: 35980694]
50. Integration of mental health and HIV interventions: key considerations. : Joint United Nations Programme on HIV/AIDS, World Health Organization; 2022 URL: https://iris.who.int/server/api/core/bitstreams/831587a7-3fab-4a70-86c1-6e179a112638/content [accessed 2026-09-10]
51. Asiedu N, Mwesiga EK, Akena D, et al. Evaluating construct and criterion validity of NeuroScreen in assessing neurocognition among hospitalized Ugandan first-episode psychosis patients. Schizophr Res Cogn 2022 Jun;32:100276. [doi: 10.1016/j.scog.2022.100276] [Medline: 36593996]
52. Asiedu N, Robbins R, Santoro A, et al. Adapting tablet-based neuropsychological tests for adolescents and young adults with perinatally acquired HIV in Uganda: the translation process for two languages. Presented at: 16th International Workshop on HIV Treatment, Pathogenesis and Prevention Research in Resource-Limited Settings (INTEREST 2022); May 10-13, 2022 URL: https://academicmedicaleducation.com/programs/interest-2022 [accessed 2026-09-14]
53. Gregoire J. ITC guidelines for translating and adapting tests (second edition). Int J Test 2018 Apr 3;18(2):101-134. [doi: 10.1080/15305058.2017.1398166]
54. International Test Commission (ITC). ITC guidelines for the large-scale assessment of linguistically and culturally diverse populations. Int J Test 2019 Oct 2;19(4):301-336. [doi: 10.1080/15305058.2019.1631024]
55. Nguyen CM, Rampa S, Staios M, et al. Neuropsychological application of the International Test Commission guidelines for translation and adapting of tests. J Int Neuropsychol Soc 2024 Aug;30(7):621-634. [doi: 10.1017/S1355617724000286] [Medline: 39291438]
56. Brislin RW. The wording and translation of research instruments. In: Field Methods in Cross-Cultural Research: Sage Publications, Inc; 1986:137-164 URL: https://psycnet.apa.org/record/1987-97046-005 [accessed 2026-09-10]
57. Kaufman AS, Kaufman NL. Kaufman Assessment Battery for Children (KABC-II), 2nd edition: American Guidance Service; 2004. URL: https://www.pearsonclinical.in/products/programs/kaufman-assessment-battery-for-children-second-edition-normative-update.html [accessed 2026-09-17]
58. Wechsler D. Wechsler Intelligence Scale for Children (WISC-IV), 4th edition: The Psychological Corporation; 2003. URL: https://www.pearsonclinical.in/products/programs/wechsler-intelligence-scale-for-children-fourth-edition-india-complete-kit.html [accessed 2026-09-17]
59. Llorente AM, Williams J, Satz P, D'Elia LF. Children's Color Trails Test (CCTT): Psychological Assessment Resources; 2003. URL: https://www.parinc.com/products/CCTT [accessed 2026-09-17]
60. Bangirana P, Musisi S, Boivin MJ, et al. Malaria with neurological involvement in Ugandan children: effect on cognitive ability, academic achievement and behaviour. Malar J 2011 Nov 3;10:334. [doi: 10.1186/1475-2875-10-334] [Medline: 22047193]

XSL•FO
RenderX

61. Nalwanga D, Musiime V, Bangirana P, et al. Neurocognitive function among HIV-infected children on protease inhibitor -based versus non-protease inhibitor based antiretroviral therapy in Uganda: a pilot study. BMC Pediatr 2021 Apr 26;21(1):198. [doi: 10.1186/s12887-021-02676-2] [Medline: 33902506]
62. Ruel TD, Boivin MJ, Boal HE, et al. Neurocognitive and motor deficits in HIV-infected Ugandan children with high CD4 cell counts. Clin Infect Dis 2012 Apr;54(7):1001-1009. [doi: 10.1093/cid/cir1037] [Medline: 22308272]
63. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 1995 Jan 1;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
64. Ritz C, Streibig JC. Bioassay analysis using R. J Stat Softw 2005;12:1-22. [doi: 10.18637/jss.v012.i05]
65. Björngrim S, van den Hurk W, Betancort M, Machado A, Lindau M. Comparing traditional and digitized cognitive tests used in standard clinical evaluation—a study of the digital application minnemera. Front Psychol 2019;10:2327. [doi: 10.3389/fpsyg.2019.02327] [Medline: 31681117]
66. Spreij LA, Gosselt IK, Visser-Meily JMA, Nijboer TCW. Digital neuropsychological assessment: feasibility and applicability in patients with acquired brain injury. J Clin Exp Neuropsychol 2020 Oct;42(8):781-793. [doi: 10.1080/13803395.2020.1808595] [Medline: 32878561]
67. Lunardini F, Luperto M, Romeo M, et al. Supervised digital neuropsychological tests for cognitive decline in older adults: usability and clinical validity study. JMIR Mhealth Uhealth 2020 Sep 21;8(9):e17963. [doi: 10.2196/17963] [Medline: 32955442]
68. Nunnally JC, Bernstein IH. Psychometric Theory, 3rd edition: McGraw-Hill; 1994.
69. Embretson SE, Reise SP. Item Response Theory for Psychologists: Lawrence Erlbaum Associates; 2000. [doi: 10.4324/9781410605269]
70. Smith GT, McCarthy DM, Anderson KG. On the sins of short-form development. Psychol Assess 2000;12(1):102-111. [doi: 10.1037/1040-3590.12.1.102] [Medline: 10752369]
71. Nichols SL, Robbins RN, Rampa S, Malee KM. Neurocognitive outcomes following perinatal human immunodeficiency virus infection. Clin Perinatol 2024 Dec;51(4):865-879. [doi: 10.1016/j.clp.2024.08.004] [Medline: 39487025]
72. Van den Hof M, Ter Haar AM, Scherpbier HJ, et al. Neurocognitive development in perinatally human immunodeficiency virus–infected adolescents on long-term treatment, compared to healthy matched controls: a longitudinal study. Clin Infect Dis 2020 Mar 17;70(7):1364-1371. [doi: 10.1093/cid/ciz386] [Medline: 31106812]
73. Jenkins A, Lindsay S, Eslambolchilar P, Thornton IM, Tales A. Administering cognitive tests through touch screen tablet devices: potential issues. J Alzheimers Dis 2016 Oct 4;54(3):1169-1182. [doi: 10.3233/JAD-160545] [Medline: 27567875]
74. McHenry MS, Mukherjee D, Bhavnani S, et al. The current landscape and future of tablet-based cognitive assessments for children in low-resourced settings. PLOS Digit Health 2023 Feb;2(2):e0000196. [doi: 10.1371/journal.pdig.0000196] [Medline: 36821551]

Abbreviations

ART: antiretroviral therapy
CCTT: Children's Color Trails Test
FDR: false discovery rate
IRB: institutional review board
JCRC: Joint Clinical Research Centre
KABC-II: Kaufman Assessment Battery for Children, Second Edition
LMIC: low- and middle-income country
NCI: neurocognitive impairment
PHIV: perinatally acquired HIV
SSA: sub-Saharan Africa
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
WISC-IV: Wechsler Intelligence Scales for Children, Fourth Edition

Edited by L Buis; submitted 02.Jun.2026; peer-reviewed by D Vance, P Akhigbe; revised version received 26.Jul.2026; accepted 27.Jul.2026; published 13.Oct.2026.

Please cite as:
Robbins RN, Nanteza A, Gumikiriza-Onoria JL, Santoro A, Ferraris CM, Karungi C, Tsapalas D, Asiedu N, Kirsch C, Tan M, Liu J, Familiar-Lopez I, Musiime V, Dirajlal-Fargo S
Tablet-Based Neurocognitive Tests for Ugandan Adolescents Living With HIV: Cross-Sectional Validation Study
JMIR Mhealth Uhealth 2026;14:e103288
URL: https://mhealth.jmir.org/2026/1/e103288
doi:10.2196/103288

XSL•FO
RenderX

© Reuben N Robbins, Angel Nanteza, Joy Louise Gumikiriza-Onoria, Anthony Santoro, Christopher M Ferraris, Christine Karungi, Daphne Tsapalas, Nana Asiedu, Courtney Kirsch, Mei Tan, Jun Liu, Itziar Familiar-Lopez, Victor Musiime, Sahera Dirajlal-Fargo. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 13.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobile Health Apps for Individuals Who Are Blind or Have Low Vision: Systematic Review of Evidence, Accessibility, and App Quality

Carlos Mora[1,2], MSci; Ester Solé[1,2,3], PhD; Pere Llorens-Vernet[1,2,3], PhD; Jordi Miró[1,2,3], PhD

[1]Research Center for Behavior Assessment (CRAMC), Department of Psychology, Universitat Rovira i Virgili, Carretera de Valls s/n, Tarragona, Catalonia, Spain

[2]Unit for the Study and Treatment of Pain–ALGOS, Chair in Pediatric Pain, Tarragona, Catalonia, Spain

[3]Institut d'Investigació Sanitària Pere Virgili, Tarragona, Catalonia, Spain

Corresponding Author:
Jordi Miró, PhD
Research Center for Behavior Assessment (CRAMC), Department of Psychology, Universitat Rovira i Virgili, Carretera de Valls s/n, Tarragona, Catalonia, Spain

Abstract

Background: Individuals who are blind or have low vision experience disproportionate burdens in both physical and psychological functioning, and mobile health apps may help address some of these needs. However, their availability, accessibility, and empirical support for this population remain unclear.

Objective: This systematic review aimed to identify mobile health apps relevant to supporting physical and psychological functioning in individuals who are blind or have low vision, summarize the empirical evidence available for these apps, and examine their accessibility and quality.

Methods: We used a multicomponent approach that combined (1) identification of apps through organizational recommendations; (2) a systematic search of the scientific literature, conducted and reported in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines; and (3) an exploratory accessibility and app quality appraisal using the Web Content Accessibility Guidelines 2.1 and the Mobile App Development and Assessment Guide.

Results: A total of 19 unique mobile health apps were identified after combining and deduplicating apps identified through organizational recommendations and scientific literature. Fifteen apps were identified through organizational recommendations, and 14 empirical studies evaluating 11 apps were included in the qualitative synthesis. Only a subset of these studies specifically involved blind or low-vision participants; several evaluated apps in broader clinical or general populations where visual impairment status was not reported. The identified apps focused primarily on physical functioning, particularly medication management, glucose monitoring, pharmaceutical information access, and health care access. No apps primarily designed to support psychological functioning in individuals who are blind or have low vision were identified. Of 14 studies, 6 (43%) were rated as being of good methodological quality. The exploratory user-informed accessibility appraisal showed variability across apps: MyPills received the most favorable accessibility profile, with "High" ratings across all 4 Web Content Accessibility Guidelines principles, whereas several organizationally recommended apps received low ratings across most principles. Mobile App Development and Assessment Guide scores ranged from 8/39 to 34/39, with Medisafe scoring highest.

Conclusions: The current landscape of mobile health apps for individuals who are blind or have low vision remains limited, uneven in accessibility and quality, and supported by a limited and heterogeneous evidence base. The absence of identified apps targeting psychological functioning represents a particularly important gap. Improving the development and implementation of mobile health apps for individuals who are blind or have low vision will require more than increasing the number of available apps; it will require apps that are accessible from the outset, rigorously evaluated, and developed with the people they are intended to serve.

Trial Registration: PROSPERO CRD42023448243; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023448243

(JMIR Mhealth Uhealth 2026;14:e85809) doi:10.2196/85809

KEYWORDS

blindness; visually impaired persons; mobile apps; mHealth; telemedicine; mental health; self-management; accessibility

XSL•FO
RenderX

Introduction

In 2019, the World Health Organization estimated that at least 2.2 billion people worldwide are blind or have low vision [1], and the number of individuals living with blindness is projected to reach 115 million by 2050. Visual impairment is therefore a major global public health issue with substantial implications for health, functioning, and quality of life [2].

Research consistently shows that individuals who are blind or have low vision report poorer physical and mental health outcomes than their sighted peers [3-6]. For example, a recent review [7] concluded that chronic pain, particularly neuropathic ocular pain and back pain, appears to be common in this population. Visual impairment is also a well-established risk factor for falls and fractures [8,9], which may in turn contribute to loss of independence, permanent disability, and increased mortality, especially in older adults [10]. Mental health outcomes are similarly affected. Adults with visual impairments experience higher rates of depression [11] and anxiety [4] than those with normal vision, and greater severity of visual impairment has also been associated with increased risk of suicidal thoughts and behaviors [12,13].

Despite the elevated risks, access to health care remains limited for individuals with disabilities, including those with visual impairments [14]. Previous research has documented persistent disparities in preventive care, health care usage, and access to appropriate treatment compared with sighted individuals [2,15,16]. For example, although older adults with visual impairments present higher rates of clinically significant depression and anxiety [17], many never receive appropriate treatment [18,19]. Even in settings where accessibility is legally mandated, implementation is inconsistent [20]. Access to effective treatment is further restricted by the limited accessibility of many medical technologies. For example, insulin pumps and other diabetes-related devices often lack basic accessibility features such as voice guidance and accessible instructions [21], despite the relevance of diabetes in this population.

Mobile health apps (ie, software designed for smartphones, tablets, and other portable devices [22]) have emerged as promising tools to support the management of chronic noncommunicable diseases, such as chronic pain [23,24], diabetes [25], hypertension [26], depression [27], and anxiety [28]. For individuals who are blind or have low vision, mobile apps can promote autonomy, enhance communication with health care professionals and peers, and improve personalized care and overall quality of life [29-33]. However, many commercially available apps are developed primarily for commercial rather than clinical purposes and lack rigorous empirical support [34]. Limited involvement of key stakeholders, particularly end users [35], may further contribute to mistrust, privacy concerns, and the dissemination of inaccurate or even potentially harmful information [36,37]. Some reviews and empirical studies have begun to describe the broader landscape of mobile app usability for visually impaired users. For example, Al-Razgan and colleagues [29] conducted a systematic literature review on the usability of mobile apps for people with visual impairments, identifying key themes such as accessibility, assistive technologies, navigation, and interaction design. Their review provides an important overview of the general usability challenges faced by visually impaired users in mobile environments. However, mobile health apps, particularly those intended to support the physical and psychological functioning of individuals who are blind or have low vision, received far less systematic attention.

To our knowledge, no systematic review has specifically examined mobile health apps relevant to individuals who are blind or have low vision with a focus on their potential role in supporting physical and psychological functioning, the empirical evidence supporting their use, and the extent to which such apps meet current accessibility and quality standards. Conducting such a review has important implications for both research and clinical practice: it can help identify promising tools, clarify important gaps in the evidence base, and inform the development and recommendation of apps that are both effective and accessible for this population.

Accordingly, this systematic review addressed three research questions: (1) Which mobile health apps relevant to individuals who are blind or have low vision can be identified through organizational recommendations and the scientific literature? (2) What empirical evidence is available regarding the usability, accessibility, and health-related outcomes of these apps? (3) What do the exploratory user-informed accessibility appraisal and app-quality assessment suggest about the accessibility and quality of the identified apps?

Methods

Procedure

Overview

This study followed a multicomponent review design intended to address complementary dimensions of the same research problem. First, we identified mobile health apps recommended by organizations serving individuals who are blind or have low vision in order to characterize apps currently promoted within this community. Second, we conducted a systematic review of the scientific literature to identify and examine empirical studies evaluating mobile health apps relevant to this population. Apps identified through organizational recommendations and the scientific literature were then combined and deduplicated by app name, developer, intended function, and platform availability to determine the final number of unique apps included in the review. This review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [38] for systematic reviews and meta-analyses and was conducted in accordance with the preregistered protocol in the PROSPERO [39] database (CRD42023448243). No substantive deviations from the protocol occurred; any minor refinements were limited to improvements in reporting clarity and did not affect the scope, eligibility criteria, or analytical approach. All studies were analyzed using a qualitative synthesis approach. Two independent reviewers (CM and JM) assessed study eligibility using a 2-step process that involved initial screening of titles

XSL•FO
RenderX

and abstracts followed by full-text evaluation. Discrepancies were resolved through discussion, and when necessary, a third reviewer was consulted to reach consensus. Third, we carried out an exploratory, user-informed accessibility appraisal of apps that were available for direct review. This final component was not designed as a stand-alone primary study or as a formal usability trial intended for statistical generalization but rather as a complementary appraisal to provide practical insight into how the identified apps could be accessed and experienced by potential users. Accordingly, the accessibility appraisal was intended to contextualize the evidence synthesis by bridging the gap between published descriptions of apps and their real-world accessibility in use. As this review was based exclusively on previously published literature and did not involve the collection of primary data or direct participation of human subjects, approval from an Institutional Review Board (IRB) or Research Ethics Committee was not required.

Identification of Apps Recommended by Organizations Serving Individuals With Visual Impairments

To identify mobile health apps currently recommended to individuals who are blind or have low vision, we conducted a targeted review of websites and digital resources from organizations serving this population. These organizations often act as key sources of information and dissemination channels for assistive technologies and digital tools within the visually impaired community. Organizations from different geographical regions were considered, including international, national, and regional organizations that provide resources and support for individuals with visual impairments. Examples include the American Foundation for the Blind, the Royal National Institute of Blind People, the World Blind Union, the Spanish National Organization for the Blind (ONCE), and Vision Australia. The websites of these organizations were systematically examined to identify mobile health apps intended to support physical and psychological functioning in individuals who are blind or have low vision and those that were recommended or listed as potentially useful for this population. When relevant information was available, we recorded basic descriptive characteristics of each app, including the app name, platform availability (ie, iOS or Android), intended purpose, and any description provided by the recommending organization.

Systematic Search of the Scientific Literature on Mobile Health Apps Intended to Support Physical and Psychological Functioning

To identify empirical studies on mobile health apps intended to support the physical and psychological functioning of individuals who are blind or have low vision, we conducted a systematic search of the scientific literature. The search was performed in 3 major databases: ScienceDirect, PubMed or MEDLINE, and other National Center for Biotechnology Information (NCBI) resources. The search strategy combined terms related to mobile apps and visual impairment. Specifically, the following keywords were used: (“app*” OR “mobile” OR “smartphone”) AND (“visually impaired” OR “visual disability” OR “blind*” OR “low vision” OR “ocular*” OR “eye*”), using asterisks as wild cards to capture variations of each term. A supplementary app-name search was also conducted for apps identified through organizations serving individuals with visual impairments, combined with the same keyword strategy. The initial search was conducted in ScienceDirect, PubMed or MEDLINE, and other NCBI resources and covered records published through December 2025. In April 2026, the search was updated and extended across all 6 databases included in the final review: ScienceDirect, PubMed or MEDLINE, PubMed Central, Scopus, Web of Science, and the ACM Digital Library. All 6 databases were searched from inception through April 2026. Thus, the April 2026 update both extended the final coverage date of the databases included in the initial search and broadened database coverage by adding Scopus, Web of Science, and the ACM Digital Library. The updated search strategy was developed through an iterative refinement process. Initial exploratory searches using broad terms yielded a high proportion of irrelevant records due to the ambiguity of certain keywords. Therefore, search terms were progressively refined by prioritizing more specific descriptors related to mobile apps and visual impairment and by combining these with methodological terms (eg, “evaluation” and “empirical study”) where appropriate. This approach aimed to balance sensitivity and specificity, ensuring adequate coverage while reducing noise. The search was restricted to peer-reviewed journal papers. Gray literature and conference proceedings were not included as primary evidence sources, as the review focused on empirical studies with formal evaluation components to ensure methodological consistency and comparability across studies. The full database-specific search strategies used in the updated search, including complete search strings, Boolean operators, filters, and records retrieved per source, are provided in Multimedia Appendix 1. The supplementary app-name search was intended to identify empirical studies on apps already recommended in practice, while the broader database search aimed to capture additional relevant apps not identified through organizational sources.

To maximize comprehensiveness, we also manually screened the reference lists of eligible papers for additional studies. Studies were included if they met the following criteria: (1) evaluated a mobile app intended as relevant to individuals who are blind or have low vision, because it was designed for this population, recommended by organizations serving this population, or identified in the scientific literature as potentially relevant; (2) reported empirical findings related to the app; (3) were published in a peer-reviewed journal; (4) were available in full text; and (5) were written in English or Spanish. Studies were eligible even when the empirical evaluation was conducted in a broader clinical or general population, provided that the app itself had been identified as relevant to individuals who are blind or have low vision. Data extraction was performed using a structured template developed for this review. Extracted variables included study characteristics (authors, year, country, and design), participant characteristics, app name and purpose, health domain addressed, key accessibility and usability findings, and the main outcomes reported. Study designs were categorized into broader groups (eg, observational and experimental) to facilitate consistent reporting and comparison across heterogeneous studies. The data extraction form used in this review, including all extracted variables, is provided in Multimedia Appendix 2.

XSL•FO
RenderX

Assessment of Methodological Quality and Risk of Bias

Methodological quality and risk of bias were assessed by 2 reviewers (CM and JM) using the version of the National Institutes of Health (NIH) Quality Assessment Tool appropriate to each study design. Discrepancies were discussed and resolved by consensus following guidance from the Cochrane Handbook for Systematic Reviews of Interventions. The final ratings for all included studies were based on this consensus process.

For each study, the NIH criteria were rated as "yes," "no," "cannot determine," or "not applicable," as appropriate. "Cannot determine" was used when insufficient information was available to judge an item, and "not applicable" when a criterion was not relevant to the study design. To obtain an overall indicator of methodological rigor, "yes" responses were scored as 1 point, whereas all other responses ("no," "not reported," "cannot determine," or "not applicable") were scored as 0, and item scores were summed across criteria. Consistent with approaches used in previous systematic reviews [40-42], total scores were categorized as poor (0 - 6), fair (7-10), or good (11-14). For ease of interpretation, the resulting ratings are presented both at the item level and as an overall summary classification.

Exploratory Accessibility and Functionality Appraisal of the Identified Mobile Apps

This appraisal was conceived as an exploratory complement to the review. Consequently, the resulting ratings should be interpreted as illustrative observations derived from participant experiences and consensus discussions rather than as a formal usability study or as independently generalizable evidence. The accessibility and functionality of the identified apps were appraised using the Web Content Accessibility Guidelines (WCAG 2.1, 2018) [43] framework focusing on the 4 POUR principles: (1) Perceivable (ie, information and user interface elements must be easily perceivable by users), (2) Operable (ie, all elements should be fully interactive, free of distracting animations, or gestures that require the use of multiple fingers), (3) Understandable (ie, information and interface functions should be clear and easy to understand), and (4) Robust (ie, apps should be compatible with different devices and operating systems, including features such as adjustable font sizes). In practical terms, the accessibility appraisal examined whether relevant app information and interface elements were perceivable (eg, accessible labels, readable content, and discernible feedback), whether navigation and controls were operable (eg, keyboard or screen-reader accessibility, predictable navigation, and actionable controls), whether content and interaction flows were understandable (eg, clarity of instructions, consistency, and comprehensibility of interface elements), and whether the app appeared robust in its compatibility with assistive technologies and platform accessibility services. Although WCAG was originally developed for web content, its principles have also been applied and adapted to mobile and nonweb software environments in World Wide Web Consortium (W3C) guidance and prior accessibility research [43,44]. Because some WCAG wording is web-oriented, criteria were interpreted in relation to mobile app environments following established W3C guidance for nonweb Information and Communication Technologies and mobile apps. Thus, references to web pages were considered at the level of app screens or views, and accessibility judgments were made in relation to native app interactions and compatibility with mobile assistive technologies. Following this approach, we applied WCAG 2.1 to native mobile apps to assess barriers and facilitators relevant to individuals who are blind or have low vision, particularly those related to screen reader compatibility, interface clarity, navigation, and interaction with core app functions.

To assess the accessibility and functionality of the identified apps, we conducted a user-based appraisal guided by the WCAG 2.1 framework. Four adult volunteers with visual impairments participated in the evaluation (2 blind and 2 with reduced visual acuity). Participants were recruited through convenience sampling based on accessibility and prior experience using mobile assistive technologies. No formal inclusion or exclusion criteria were applied beyond having lived experience of visual impairment and familiarity with smartphone use. Participants were instructed to download, install, and test the core functionalities of each app (eg, setting medication reminders and navigating menus) using the native accessibility features available on their devices. Following individual testing, participants discussed their experiences in a group session and appraised each app in terms of ease of use, clarity of information, compatibility with screen readers or other assistive technologies, and overall performance across the 4 WCAG principles. Based on this process, each app received an overall rating of low, moderate, or high for each WCAG principle. When disagreement occurred, the most stringent rating was retained in order to prioritize barriers encountered by blind users. These ratings therefore reflect a consensus-based synthesis of the participants' practical experience with the apps rather than a formal psychometric assessment, standardized accessibility measurement, or clinical evaluation. No interrater reliability, test-retest reliability, or formal validation procedure was conducted for this exploratory component. A detailed operationalization of WCAG 2.1 principles, including examples, is provided in Multimedia Appendix 3. In addition, we reviewed scientific literature related to the accessibility and functionality of the identified apps to complement the user-based appraisal. Given the small and diagnostically homogeneous sample, this component was intended to enrich the review with real-world end user perspectives rather than to generate generalizable estimates of usability.

App Quality Appraisal Using the Mobile App Development and Assessment Guide

Each identified app was also appraised using the Mobile App Development and Assessment Guide (MAG) [35]. The MAG is a structured framework developed to support the assessment of mobile health apps across multiple domains relevant to quality, including usability, privacy, security, appropriateness, safety, and technology performance. The guide includes 39 items organized into seven categories: (1) Usability, (2) Privacy Policy Information, (3) Privacy Features, (4) Security, (5) Appropriateness and Suitability, (6) Safety, and (7) Technology Performance [34]. Two independent researchers (CM and PLl-V) reviewed each app against the 39 MAG criteria. Each item was rated as "yes" when the criterion was met, "no" when

XSL•FO
RenderX

it was not met, and "not applicable" when the criterion could not reasonably be assessed for a given app. Agreement between reviewers was approximately 90% before consensus discussions. Disagreements were generally minor and primarily reflected different interpretations of individual MAG criteria. These discrepancies were resolved through discussion and consensus review before final scores were assigned. The full list of MAG criteria used in the evaluation is presented in Table 1. In addition to the item-level appraisal, an overall summary score was calculated for each app based on the number of MAG criteria met, in order to facilitate comparison across apps. Because the MAG does not provide predefined cutoff points for interpreting total scores, descriptive score bands were created for this review. Apps meeting approximately two-thirds or more of the MAG criteria were classified as higher app quality (≥27/39), apps meeting approximately one-half to less than two-thirds of the criteria were classified as moderate app quality (20-26/39), and apps meeting fewer than one-half of the criteria were classified as lower app quality (≤19/39). These categories were used only to aid interpretation of the descriptive MAG scores and should not be interpreted as validated MAG thresholds.

XSL•FO
RenderX

Table . Mobile App Development and Assessment Guide (MAG) criteria used to appraise the quality of mobile health apps for individuals who are blind or have low vision.

Item	Statement	MAG category
1	The app was tested by potential users before being made available to the public.	Usability
2	It is easy to use.	Usability
3	The app features are adapted to the purposes of the app.	Usability
4	Functionality is adjusted to the targeted user.	Usability
5	The app is designed to be accessible for people with physical impairments.	Usability
6	The app uses lay language that can be easily understood by the targeted user.	Usability
7	The user can easily access all resources and features.	Usability
8	The app provides information about the terms and conditions for the use of the app.	Privacy Policy Information
9	The app provides information about privacy policies.	Privacy Policy Information
10	The app provides information about commercial agreements with third parties or the absence of them.	Privacy Policy Information
11	The app informs users if their data must be transferred to third parties for the app to operate.	Privacy Policy Information
12	The app indicates a recommended user age based on its content.	Privacy Policy Information
13	The app includes a simple and understandable version of the terms and conditions of use, and privacy policies.	Privacy Policy Information
14	The app gives a warning when an update influences sensitive data.	Privacy Features
15	The app provides the option to only collect user data that are essential for the app to operate.	Privacy Features
16	The app guarantees users' rights on stored personal data.	Privacy Features
17	The app provides the user with the option to refuse the transfer of data to third parties.	Privacy Features
18	The app requires the user to give her or his consent to the terms and conditions of use and privacy policies.	Privacy Features
19	The app notifies the user when other resources on the mobile device, such as her or his accounts or social network profiles, are accessed.	Privacy Features
20	The app guarantees the user’s privacy.	Privacy Features
21	The app offers to delete the user’s data at any time.	Privacy Features
22	The app informs the user about the security measures in place.	Security
23	The app includes password management mechanisms.	Security
24	The authorization and authentication mechanisms work correctly.	Security
25	The app validates the user’s identity to access confidential data.	Security

XSL•FO
RenderX

Item	Statement	MAG category
26	The app informs the user when it detects and identifies cybersecurity vulnerabilities, possible threats, and the risk of being exploited that may affect the data.	Security
27	The app informs the user that when a third-party component is updated, the updated version is inspected, and the risks are evaluated.	Security
28	The benefits and advantages of using the app are explained.	Appropriateness and Suitability
29	Experts have participated in the development of the app.	Appropriateness and Suitability
30	The app informs the user that it is based on scientific findings.	Appropriateness and Suitability
31	The app informs the user that it is developed in accordance with ethical principles and values.	Appropriateness and Suitability
32	The possible health risks for the user are identified.	Safety
33	The app complies with regulatory standards as a medical device.	Safety
34	The user is warned when adverse events are identified.	Safety
35	The app recommends that users always consult a specialist if they have questions, or it warns users that the app is not meant to replace the services provided by a health professional.	Safety
36	The app works correctly.	Technology Performance
37	Functions are correctly retrieved after events external to the app occur.	Technology Performance
38	The app uses mobile device resources efficiently (eg, battery, CPU, memory, data, network, etc), giving it a level of performance high enough not to interfere with the user's experience.	Technology Performance
39	The app has a data recovery system in case of loss.	Technology Performance

Results

Overview

The results are organized according to the 3 research questions and the corresponding methodological components of the study. We first describe the mobile health apps identified through organizational recommendations and the scientific literature. We then summarize the empirical evidence available for these apps, including an assessment of the risk of bias and quality of these studies. Finally, we present the exploratory accessibility and functionality appraisal and the app quality assessment of the identified apps.

Mobile Health Apps Identified Through Organizational Recommendations

After combining and deduplicating apps identified through organizational recommendations and the scientific literature, 19 unique mobile health apps were included in the review. Fifteen apps were identified through organizational recommendations and 11 were evaluated in the empirical studies included in the qualitative synthesis. These counts should not be summed because 7 apps were identified through both sources, 8 were identified only through organizational recommendations, and 4 were identified only through empirical studies. Apps were counted as unique when they represented distinct apps after comparison by app name, developer, intended function, and platform availability. Among the 15 apps identified through organizational recommendations, all were freely available, and most (14/15, 93%) were compatible with both iOS and Android platforms. Across the 19 unique apps included in the review, none specifically targeted psychological functioning or mental health concerns. Six apps (ie, Medisafe, MyFitnessPal, SACYL Cita Previa, Salud Informa, SCSalud, and SESCAM) were recommended by both Fundación ONCE and the American Foundation for the Blind, whereas the remaining apps were endorsed by a single organization.

Apps identified only through the scientific literature were retained as additional unique apps when they represented distinct mobile apps not already captured through organizational recommendations. Across the 15 apps identified through

XSL•FO
RenderX

organizational recommendations, 2 broad functional categories were identified. First, several apps supported health self-management or health-related information access, including medication scheduling, medication information retrieval, glucose monitoring, nutrition tracking, or physical activity support. These included apps with medication scheduling and reminders (eg, Medisafe and Alarma de Medicamentos), medication information access or voice-based output (eg, ScripTalk and Medicamento Accesible Plus), glucose monitoring or diabetes self-management functions (eg, Glucose Buddy Diabetes Tracker, OneTouch Reveal, DiabetesPal, and FreeStyle LibreLink), and nutrition or weight management support (eg, MyFitnessPal). Second, several apps primarily supported health care access or administrative functions, such as appointment scheduling, prescription management, or integration with regional public health systems (eg, SACYL Cita Previa, Salud Informa, SCSalud, SESCAM, GVA + Salut and Astursalud Cita Previa). These apps are linked to specific Spanish regional public health systems and are generally restricted to individuals receiving care within those systems. They are therefore not broadly available or directly transferable to users in other national or regional health care contexts. These administrative apps were explicitly identified through the organizational recommendation component and may provide meaningful access to appointments, prescriptions, health records, and other services within their intended local settings. However, they should be interpreted as a separate category of locally contextualized health care access tools rather than as equivalent to internationally available clinical self-management apps, such as medication management, glucose monitoring, or health information tools. The main features and functional categories of these apps are summarized in Table 2.

XSL•FO
RenderX

Table . Main features of mobile health apps identified through organizational recommendations for individuals who are blind or have low vision.

Mobile app	Functional category[a]	Organization	Operating system	Developer	Languages	Key features	Price	Downloads, n
Alarma de Medicamentos	Medication management	Fundación ONCE	Android/iOS	Aplicativos legais LTDA-ME	English, Spanish, Portuguese, French, German, Italian, Russian, Korean, Turkish, and others	Customizable medication reminders and dosage tracking	Free with in-app purchases	>1 million
Astursalud Cita Previa	Health care access or administrative	Fundación ONCE	Android	Gobierno del Principado de Asturias	Spanish	Manage primary care appointments within the Asturias (Spain) regional public health system	Free	76,000
DiabetesPal	Diabetes self-management	American Foundation for the Blind	Android/iOS	Pascal Freiburghaus	English and German	Diabetes management: track glucose, insulin, and meals; sync across devices via iCloud, Dropbox, or WebDAV	Free to download with in-app purchases	>50,000 (Android)
FreeStyle LibreLink	Glucose monitoring	Fundación ONCE	Android/iOS	Abbott Diabetes Care Inc	Multilingual (includes English, Spanish, French, Italian, and more)	Glucose monitoring with sensor technology; provides historical trends; compatible with LibreView platform	Free to download with in-app purchases	>1 million
Glucose Buddy Diabetes Tracker	Diabetes self-management	American Foundation for the Blind	Android/iOS	Azumio Inc	English	Logs blood glucose, meals, and medications; integrates with Apple Health and Dexcom devices; provides detailed reports and optional coaching	Free with in-app purchases	>1 million
GVA + Salut	Health care access or administrative	Fundación ONCE	Android/iOS	Generalitat Valenciana	Spanish and Catalan from València	Appointment scheduling, medical records, prescriptions, telemedicine, and access to services within the Valencian (Spain) regional public health system	Free	2.4 million

XSL•FO
RenderX

Mobile app	Functional category[a]	Organization	Operating system	Developer	Languages	Key features	Price	Downloads, n
Medicamento Accesible Plus	Medication information access	Fundación ONCE	Android/iOS	ILUNION Accesibilidad, Fundación ONCE, Fundación Vodafone España, and Consejo General de Colegios Farmacéuticos.	Spanish	Accessible medication information (indications, dosage, and warnings); includes food intolerance, pregnancy safety, and doping risk	Free	>50,000 (Android)
Medisafe	Medication management	American Foundation for the Blind/ONCE	Android/iOS	Medisafe Inc	Available in multiple languages, including English, Spanish, French, German, and more	Pill reminders and medication scheduling, drug interaction alerts, appointment tracking, and MedFriend caregiver support	Free with in-app purchases	>10 million (Android)
MyFitnessPal	Nutrition or weight management	American Foundation for the Blind/ONCE	Android/iOS	MyFitnessPal Inc	English and additional languages for localized versions	Food and nutrition tracking, barcode scanner, and macro- and micronutrient analysis	Free with optional in-app purchases for premium features	>200 million (worldwide)
OneTouch Reveal	Glucose monitoring	American Foundation for the Blind	Android/iOS	LifeScan Inc	English, French, German, Italian, Spanish, Portuguese, Dutch, Czech, Japanese, Slovenian, Polish, and Vietnamese	Tracks blood sugar readings via integration with OneTouch Verio Reflect and Verio Flex meters, allows manual data entry, and provides visual summaries, cloud synchronization, and web integration	Free	>1 million
SACYL Cita Previa	Health care access or administrative	American Foundation for the Blind/ONCE	Android/iOS	Junta de Castilla y León	Spanish	Manage, book, consult, and cancel appointments at health centers within the Castilla y León region (Spain); locate health centers; and manage multiple health cards	Free	>1 million

XSL•FO
RenderX

Mobile app	Functional category[a]	Organization	Operating system	Developer	Languages	Key features	Price	Downloads, n
Salud Informa	Health care access or administrative	American Foundation for the Blind/ONCE	Android/iOS	Gobierno de Aragón	Spanish	Manage medical appointments, view treatment sheets, receive notifications, and manage health cards within the Aragón (Spain) regional public health system	Free	No data
SCSalud	Health care access or administrative	American Foundation for the Blind/ONCE	Android/iOS	Gobierno de Canarias	Spanish	Appointment management, medical record consultation, and health card information within the Canary Islands (Spain) regional public health system	Free	>500,000
SESCAM	Health care access or administrative	American Foundation for the Blind/ONCE	Android/iOS	Castilla-La Mancha Health Service	Spanish	Manage medical appointments, access health reports, and use personalized services within the Castilla-La Mancha (Spain) regional public health system	Free	>100,000

XSL•FO
RenderX

Mobile app	Functional category[a]	Organization	Operating system	Developer	Languages	Key features	Price	Downloads, n
ScripTalk	Medication information access	World Blind Union	Android/iOS	En-Vision America Inc	Afrikaans, German, Amharic, Armenian, Bengali, Burmese, Bulgarian, Cambodian, Catalan, Czech, Chinese, Korean, Croatian, Danish, Slovak, Slovenian, Estonian, Filipino, Finnish, French, Greek, Haitian, Hebrew, Hindi, Hungarian, Indonesian, English, Italian, Japanese, Laotian, Latvian, Lithuanian, Malay, Marathi, Dutch, Nepali, Norwegian Bokmal, Pashto, Persian, Polish, Portuguese, Romanian, Russian, Serbian, Somali, Spanish, Swahili, Swedish, Thai, Turkish, Ukrainian, Vietnamese, and Arabic	Audio descriptions of prescription labels via RFID[b] technology; reads drug name, dosage, and warnings	Free	>1000 (Android; no data for iOS)

[a]Functional categories distinguish apps primarily intended for health self-management or health-related information access from apps primarily intended to support health care access or administrative functions.

[b]RFID: radio frequency identification.

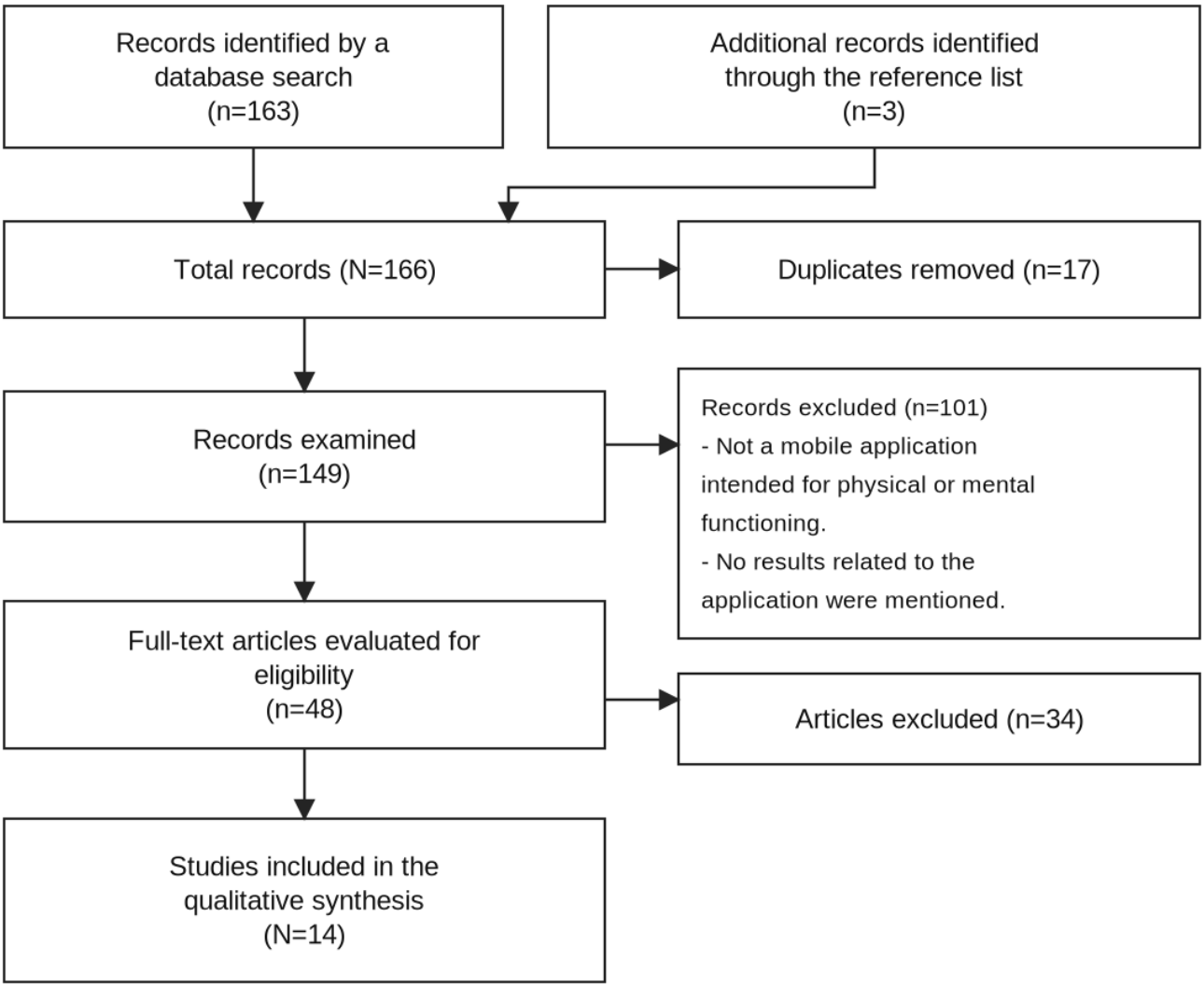
Studies on Health-Related Apps in Scientific Databases

The database search was conducted in 2 phases. The initial search covered records published through December 2025 in ScienceDirect, PubMed or MEDLINE, and other NCBI resources. In April 2026, the search was updated across these 3 databases and expanded to include Scopus, Web of Science, and the ACM Digital Library; all 6 databases were searched from inception through April 2026. Across these database searches, 163 records were identified. In addition, 3 papers were identified through manual reference screening, yielding a total of 166 records. After removing 17 duplicates, 149 titles and abstracts of papers were screened. Of these, 101 were excluded, leaving 48 papers for full-text review. Following eligibility assessment, 14 studies were included in the qualitative synthesis.

An additional expanded search was also conducted across 6 databases to further examine the coverage of the search strategy: ScienceDirect, PubMed or MEDLINE, PubMed Central, Scopus, Web of Science, and the ACM Digital Library. This expanded search retrieved 1371 records: ScienceDirect (n=222), PubMed or MEDLINE (n=22), PubMed Central (n=54), Scopus (n=579), Web of Science (n=158), and ACM Digital Library (n=336). After duplicate removal and initial screening, 40 papers were selected for further review. Four papers underwent full-text eligibility assessment; however, none met the inclusion criteria. Therefore, the expanded search did not change the final number of included studies. Figure 1 presents the PRISMA flow diagram illustrating the paper selection process.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart illustrating the main study selection process for health-related mobile apps targeting individuals who are blind or have low vision.



The 14 empirical studies examined 11 different mobile health apps identified as relevant to individuals who are blind or have low vision, most often in the context of medication management, diabetes, hypertension, nutrition, or physical activity. However, the population focus of the evidence varied. Some studies specifically involved participants who were blind or had low vision, whereas others evaluated apps recommended for, or potentially relevant to, blind and low-vision users but were conducted in broader clinical or general populations where visual impairment status was not reported. Not all apps recommended by organizations had corresponding scientific evidence, and conversely, not all apps identified in the scientific literature were endorsed by organizations.

These studies used a range of methodologies: cross-sectional designs (6/14, 42%), observational studies (6/14, 42%), and randomized clinical trials (2/14, 14%). Most studies were conducted in the United States (6/14, 42%). Table 3 provides an overview of the 14 studies and the 11 apps evaluated. Because the included studies differed substantially in population focus and outcome type, Table 3 distinguishes between studies involving blind or low-vision participants and studies conducted in broader clinical or general populations where visual impairment status was not reported, and between studies evaluating objective or clinical outcomes and those focused primarily on usability, feasibility, satisfaction, or accessibility-related outcomes. Most of the identified apps were developed specifically to support medication management and access to pharmaceutical information. Despite geographic diversity, several core features were common: (1) voice output for medication identification and instructions; (2) barcode or Global Trade Item Number scanning to recognize medication packaging; (3) accessible search interfaces, usable by text, voice, or image; and (4) reminder systems for dosing and medication refills. These apps were primarily developed for Android platforms and followed user-centered design principles. Many included adjustable user interfaces (eg, text size, spoken descriptions, and voice navigation). Overall, the evidence was heterogeneous in both population focus and outcome type. Studies involving blind or low-vision participants primarily reported findings related to usability, feasibility, medication information access, perceived autonomy, or satisfaction with apps such as ScripTalk, MyPills, FarmaceuticApp, and Ru Tan Ya [5,44,45]. Studies conducted in broader clinical or general populations reported more objective or large-scale outcomes, including medication adherence and blood pressure control for Medisafe [5,46], glycemic metrics for FreeStyle LibreLink [47] and OneTouch Reveal [48,49], weight management behavior

XSL•FO
RenderX

for MyFitnessPal [50], and dietary assessment accuracy for MyFitnessPal [51]. However, findings from broader populations should be interpreted as indirectly relevant to blind and low-vision users unless visual impairment status was specifically assessed. In addition, some barriers were noted. For example, the study of Medicamento Accesible Plus highlighted low digital literacy among older adults as a barrier to effective app use [52].

XSL•FO
RenderX

Table . Overview of empirical studies identified on mobile health apps identified as relevant to individuals who are blind or have low vision by population relevance and outcome type in chronological order of publication.

Authors (year), country	Purposes of the study	Apps	Principal functions	Population relevance to blind or low-vision users[a]	Languages	Sample of study	Type of study	Technical characteristics	Outcome type	Results
Spektor et al (2015) [53], United States	Evaluate effectiveness of ScripTalk for reading prescription labels	ScripTalk	Reads prescription labels aloud	Direct evidence: BLV[b]-specific sample	English	12 veterans with visual impairments	Field study	Allows iOS devices with Near Field Communication capability to read ScripTalk Talking Labels.	Medication access and adherence-related outcome	Improved treatment adherence and independence in medication use
Bee et al (2016) [54], Singapore	To evaluate the app for insulin use in insulin-naive patients with type 2 diabetes	DiabetesPal	Insulin adjustment based on glycemic goals	Indirect evidence: broader clinical population; BLV status not reported	English	62 participants (32 intervention, 30 control); type 2 adult patients with diabetes	Pilot randomized clinical trial	Diabetes data recording (eg, glucose, insulin, medication, blood pressure, etc); graph generation	Insulin titration feasibility and safety	Safe and feasible for insulin titration; no severe hypoglycemia recorded
Morawski et al (2018) [46], United States	To evaluate Medisafe's impact on antihypertensive adherence and blood pressure control	Medisafe	Medication reminders, adherence reports, and caregiver support	Indirect evidence: broader clinical population; BLV status not reported	17 languages, including Spanish, English, Arabic, Danish, Dutch, Finnish, French, German, Hebrew, Italian, Japanese, Korean, Portuguese, Russian, Simplified Chinese, and Turkish	412 adults with uncontrolled hypertension	Randomized clinical trial	Adherence tracking; MedFriend support; multicondition tracking	Medication adherence and blood pressure control	Improved adherence significantly compared with the control group
González Oñate et al (2018) [52], Spain	Evaluate accessibility of medication information for older adults	Medicamento Accesible Plus	Search and read medicine information	Indirect evidence: general older adult population; BLV status not reported	Spanish	Older adults in Spain	Descriptive study	Barcode scanning; health database access	Accessibility and digital literacy	Identified digital divide, highlighting the need for promoting digital literacy

XSL•FO
RenderX

Authors (year), country	Purposes of the study	Apps	Principal functions	Population relevance to blind or low-vision users[a]	Languages	Sample of study	Type of study	Technical characteristics	Outcome type	Results
Gordon et al (2019) [55], United States	To analyze weight loss goal setting on MyFitnessPal	MyFitnessPal	Diet and exercise tracking tool	Indirect evidence: general app-user population; BLV status not reported	23 languages, including Spanish, English, French, German, Italian, Portuguese, Swedish, Danish, Norwegian, Bokmal, Dutch, Russian, Korean, Japanese Simplified, Chinese Traditional, Chinese Filipino, Indonesian, Malay, Polish, and Turkish	1.4 million users	Observational study	Meal planning; activity tracking; goal customization	Weight management behavior	Effective tool for behavioral change; consistent users had better weight loss outcomes
Nedovic et al (2019) [44], Switzerland	To examine how the app can support medication management for blind and visually impaired people	MyPills	Identification of medication packages, voice output of drug name and intake instructions, and voice output of package inserts	Mixed evidence: sample included blind participants and sighted participants	Different languages (not specified)	Two blind and 4 sighted individuals (aged 30 - 70 years)	Cross-sectional study	Reminders for drug intake and reordering; audio access to package inserts; RFID[c] scanning with GTIN[d] support	Usability and medication management support	Users quickly understood the app; blind users found package scanning useful
Shade et al (2019) [50], United States	To explore usability and usefulness of Medisafe among older adults living in rural communities	Medisafe	Medication reminders	Indirect evidence: general older adult population; BLV status not reported	17 languages, including Spanish, English, Arabic, Danish, Dutch, Finnish, French, German, Hebrew, Italian, Japanese, Korean, Portuguese, Russian, Simplified Chinese, and Turkish	13 older adults residing in rural communities	Feasibility study	Adherence tracking; MedFriend support; multicondition tracking	Usability, usefulness, and feasibility	No usability issues, useful for medication tracking and adherence

XSL•FO
RenderX

Authors (year), country	Purposes of the study	Apps	Principal functions	Population relevance to blind or low-vision users[a]	Languages	Sample of study	Type of study	Technical characteristics	Outcome type	Results
Madrigal-Cadavid et al (2020) [5], Colombia	To design and develop a mobile drug information app for individuals with visual disabilities	FarmaceuticApp	Medication search (barcode, voice, or text command), access medication information, and nearby pharmacies/hospitals locator	Direct evidence: BLV-specific sample	Spanish	48 participants with visual disabilities (age range: 18 - 60 years)	Cross-sectional descriptive study	Voice descriptions; scalable fonts; configurable accessibility settings; counterfeit detection; prescription history	User satisfaction and perceived usefulness	Rated 4 to 5 on a 1 to 5 numerical rating scale; all participants stated they would use it for medication adherence
Nimmolrat et al (2021) [45], Thailand	To analyze, design, and develop a pharmaceutical app for users with visual impairment and test usability	Ru Tan Ya	Search medicinal information, timer, map to pharmacies, medication history	Direct evidence: BLV-specific sample	Thai	60 volunteers	Experimental study	Touch screen; voice user interface design; slide rule and navigation techniques functions; optimized text or image access	Usability and satisfaction	Usability was low but most users were satisfied
Kao et al (2022) [56], United States	To evaluate glucose monitoring effectiveness using the app	FreeStyle LibreLink	Blood glucose monitoring and trend analysis	Indirect evidence: broader diabetes population; BLV status not reported	English	14,000 FreeStyle Libre users; patients with type 1 and 2 diabetes	Observational study	History view; blood glucose alarms; diet, insulin use, and exercise tracking; share data with doctor	Glucose metrics	Statistically significant improvements in blood glucose metrics compared with traditional reader users
Grady et al (2023) [47], United States–United Kingdom	To evaluate Bluetooth meter (OneTouch Verio Reflect) together with a mobile app (OneTouch Reveal) for diabetes control	OneTouch Reveal	Glucose monitoring, trend alerts, health data logging	Indirect evidence: broader diabetes population; BLV status not reported	12 languages, including Spanish, German, Czech, Slovenian, French, English, Italian, Japanese, Dutch, Polish, Portuguese, and Vietnamese	55,000 users; patients with type 1 and 2 diabetes	Retrospective observational study	Pattern detection; step tracking; glucose reports	Glycemic outcomes	Statistically significant improvements in glycemia: fewer highs or lows; better average blood glucose levels

XSL•FO
RenderX

Authors (year), country	Purposes of the study	Apps	Principal functions	Population relevance to blind or low-vision users[a]	Languages	Sample of study	Type of study	Technical characteristics	Outcome type	Results
Merino et al (2023) [48], different countries	To evaluate OneTouch Reveal (OTR Pro) for remote diabetes management during the COVID-19 pandemic	OneTouch Reveal	Monitoring glucose levels, reminders, track health data	Indirect evidence: broader diabetes population; BLV status not reported	12 languages, including Spanish, German, Czech, Slovenian, French, English, Italian, Japanese, Dutch, Polish, Portuguese, and Vietnamese	1500 users during virtual consultations	Descriptive observational study	Detects blood glucose patterns; tracks daily steps; creates reports for therapeutic decisions	Clinical practice or diabetes management support	Facilitated clinical practice; ease of secure data sharing; enhanced therapeutic decisions
Banal et al (2024) [49], Philippines	To evaluate validity and intercoder reliability of MyFitnessPal for dietary assessment	MyFitnessPal	Diet and exercise tracking tool	Indirect evidence: broader diabetes population; BLV status not reported	21 languages, including Spanish, English, French, German, Italian, Portuguese, Swedish, Danish, Norwegian, Bokmal, Dutch, Russian, Korean, Japanese Simplified, Chinese Traditional, Chinese Filipino, Indonesian, Malay, Polish, and Turkish	37 Filipino adults with obesity	Observational study	Meal planning; activity tracking; dietary intake estimations	Dietary assessment validity and reliability	Useful for diet tracking; some limitations in nutritional estimate accuracy

XSL•FO
RenderX

Authors (year), country	Purposes of the study	Apps	Principal functions	Population relevance to blind or low-vision users[a]	Languages	Sample of study	Type of study	Technical characteristics	Outcome type	Results
Choi et al (2025) [51], United States	Develop an accessible conversational mHealth app (Personalized and Conversational Health Agent) to support physical activity in blind and low-vision individuals	Personalized and Conversational Health Agent	Personalized and Conversational Health Agent is a conversational, voice-based health assistant that provides personalized recommendations for physical activity, enables self-monitoring of health metrics (steps and heart rate), and facilitates interactive communication via a voice user interface	Direct evidence: BLV-specific expert evaluation	English with voice-based natural language interaction (ie, via a conversational agent using a large language model)	Two blind accessibility experts evaluated the prototype	Design and development study with qualitative expert evaluation	Uses iOS with large language model integration; voice user interface; multimodal feedback (audio and haptic via Apple Watch); uses physical activity assessment and activity guidelines for Americans to provide physical activity recommendations via an algorithm	Prototype accessibility and expert evaluation	Successful development of a functional prototype with accessible user interface, effective voice interaction, and multimodal feedback; includes 2 experts to test the prototype; limitations identified included response latency, a lack of personalized physical activity feedback, and the need for clearer goal-setting explanations

[a]"Direct evidence" refers to studies that specifically included blind or low-vision participants or expert evaluators. "Mixed evidence" refers to studies including both blind or low-vision and sighted participants. "Indirect evidence" refers to studies evaluating apps identified as relevant to blind or low-vision users but conducted in broader clinical or general populations where visual impairment status was not reported.

[b]BLV: blind or low vision.

[c]RFID: radio frequency identification.

[d]GTIN: Global Trade Item Number.

Assessment of Bias and Quality of Evidence

The interrater reliability in the assessment of methodological quality was high, with 98% agreement between reviewers across all included studies. This high level of agreement reflects the use of a structured appraisal framework with clearly defined criteria, as well as prior calibration between reviewers before the formal assessment. Minor discrepancies, mainly due to different interpretations of the criteria, were resolved through discussion, without the need to involve a third reviewer. A considerable proportion of the studies included in the review presented a moderate or high risk of bias, and out of 14 studies, only 6 (42%) met the criteria for good methodological quality: DiabetesPal [49,54,55], Medisafe [52], MyFitnessPal [50,51], FreeStyle LibreLink [47], and OneTouch Reveal [48]. Most papers reported a clear objective, defined population, and inclusion and exclusion criteria [47,48,56]. However, many did not report participation rates or whether assessors were blinded [45,52]. Table 4 provides a summary of this information. Table 4 should be interpreted as a structured overview of methodological quality across NIH criteria. The overall quality rating reflects the number of criteria met, whereas "Cannot determine" indicates insufficient information to judge an item and "Not applicable" indicates that the criterion was not relevant to the study design.

XSL•FO
RenderX

Table . Methodological quality of studies—the National Institutes of Health Quality Assessment Tool[a].

Authors (year)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Quality rating[b]
Banal et al (2024) [49]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good
Bee et al (2016) [54]	Yes	Yes	CD[c]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	CD	Yes	CD	Good
Choi et al (2025) [51]	Yes	Yes	NA[d]	Yes	Yes	NA	NA	NA	CD	NA	CD	NA	NA	NA	Poor
González Oñate and Fanjul Peyró (2018) [52]	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	NA	No	Fair
Gordon et al (2019) [55]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good
Grady et al (2023) [47]	Yes	Yes	CD	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	CD	Yes	CD	Good
Kao et al (2022) [56]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	NA	Yes	Yes	Good
Madrigal-Cadavid et al (2020) [5]	Yes	Yes	Yes	Yes	Yes	NA	NA	NA	Yes	NA	Yes	CD	NA	NA	Fair
Merino et al (2023) [48]	Yes	Yes	CD	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	CD	CD	CD	Fair
Morawski et al (2018) [46]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Good
Nedovic et al (2019) [44]	Yes	Yes	CD	Yes	No	NA	NA	NA	Yes	NA	Yes	No	Yes	No	Poor
Nimmolrat et al (2021) [45]	Yes	Yes	CD	Yes	Yes	NA	NA	NA	Yes	NA	Yes	CD	NA	NA	Poor

XSL•FO
RenderX

Authors (year)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Quality rating[b]
Shade et al (2019) [50]	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	Fair
Spektor et al (2015) [53]	Yes	Yes	CD	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	NA	CD	Fair

[a]These items are as follows: 1=study objectives stated, 2=study population defined, 3=eligible participation rate (at least 50%), 4=participant selection and inclusion and exclusion criteria uniformity, 5=sample size sufficient and/or described, 6=exposure measured prior to outcome, 7=sufficient time frame for association between exposure and outcome, 8=inclusion exposure level, 9=exposure measures valid and reliable, 10=multiple exposure measurements, 11=outcome measures valid and reliable, 12=outcome assessors blinded, 13=Loss to follow-up postbaseline (20% or less), and 14=confounders measured and adjusted between exposure and outcome.

[b]Quality ratings were derived using the National Institutes of Health Quality Assessment Tool. Overall ratings were based on the total number of criteria met: poor=0 - 6 “Yes” responses, fair=7 - 10 “Yes” responses, and good=11 - 14 “Yes” responses.

[c]CD: cannot determine (insufficient information available to judge the criterion).

[d]NA: not applicable (criterion not relevant to the study design).

Accessibility, Functionality, and Quality of the Identified Mobile Apps

We conducted an exploratory user-informed accessibility appraisal of the identified apps using the WCAG 2.1 principles: Perceivable, Operable, Understandable, and Robust. These ratings should be interpreted as illustrative user-informed observations rather than as formal accessibility certification, psychometric assessment, or standardized measurement outcomes. The appraisal suggested variability in accessibility across the apps. Within this exploratory user-informed appraisal, MyPills received the most favorable accessibility profile, obtaining “High” ratings across all 4 WCAG 2.1 [43] principles. Participants highlighted its compatibility with screen readers, clear instructions, and reliable performance across devices. In contrast, 6 apps (ie, Astursalud Cita Previa, Glucose Buddy Diabetes Tracker, SESCAM, SCSalud, SACYL Cita Previa, and MyFitnessPal) received “Low” ratings on most or all WCAG principles within this exploratory appraisal. Common issues included poor support for assistive technologies, visually cluttered interfaces without alternatives, and navigation challenges for screen reader users. Examples of accessibility barriers identified during the appraisal included unlabeled controls, insufficient clarity of instructions, inconsistent navigation flows, and limited compatibility with screen-reader interaction. Conversely, apps with clearer menu structures, accessible labeling, and more predictable navigation were judged more favorably. Of the mobile apps identified, 2 (ie, FarmaceuticApp and Ru Tan Ya) could not be downloaded or assessed because they were not available for download in the app stores at the time of assessment. Accessibility ratings of the remaining apps are summarized in Table 5.

XSL•FO
RenderX

Table . Exploratory user-informed accessibility appraisal of mobile apps for users who are blind or have low vision based on the WCAG[a] 2.1 principles[b].

Name of app	Principle 1: Perceivable	Principle 2: Operable	Principle 3: Understandable	Principle 4: Robust
Alarma de Medicamentos	High	High	High	Moderate
Astursalud Cita Previa	Low	Low	Low	Low
DiabetesPal	Moderate	Moderate	Moderate	Low
FreeStyle LibreLink	Moderate	Moderate	Moderate	Low
Glucose Buddy Diabetes Tracker	Low	Moderate	Low	Low
GVA + Salut	Moderate	Moderate	Moderate	Low
Medicamento Accesible Plus	Moderate	Moderate	Moderate	Low
Medisafe	High	High	High	Moderate
MyFitnessPal	Low	Moderate	Low	Low
MyPills	High	High	High	High
OneTouch Reveal	Moderate	High	Moderate	Moderate
Personalized conversational health agent	High	Moderate	Moderate	Low
SACYL Cita Previa	Low	Low	Low	Low
Salud Informa	Moderate	Low	Moderate	Low
SCSalud	Low	Low	Low	Low
SESCAM	Low	Low	Low	Low
ScripTalk	High	High	High	Moderate

[a]WCAG: Web Content Accessibility Guidelines.

[b]Ratings were derived from an exploratory user-informed appraisal involving 4 participants (2 blind and 2 with low vision) and should not be interpreted as formal accessibility certification, psychometric assessment, or standardized measurement outcomes.

App quality was assessed using the MAG [34], which includes 39 criteria covering usability, privacy, security, safety, appropriateness, and technology performance. The appraisal showed substantial variability across apps. Medisafe scored highest, meeting 34 of 39 criteria, whereas SACYL Cita Previa scored lowest, meeting 8 of the 39 criteria. Table 6 presents a summary of the MAG total score and the descriptive score band assigned to each app for this review. Higher total scores indicate that a greater number of quality criteria were met, whereas lower scores reflect more limited compliance across MAG domains. The complete item-level MAG appraisal is provided in Multimedia Appendix 4.

XSL•FO
RenderX

Table . Summary of app-quality appraisal using the Mobile App Development and Assessment Guide (MAG).

App	MAG total score of 39, n[a]	Descriptive MAG-based interpretation[b]
Medisafe	34	Higher app quality
Medicamento Accesible Plus	33	Higher app quality
ScripTalk	32	Higher app quality
OneTouch Reveal	31	Higher app quality
FreeStyle LibreLink	29	Higher app quality
Glucose Buddy Diabetes Tracker	29	Higher app quality
MyFitnessPal	28	Higher app quality
MyPills	27	Higher app quality
Salud Informa	25	Moderate app quality
DiabetesPal	24	Moderate app quality
GVA + Salut	22	Moderate app quality
SCSalud	18	Lower app quality
SESCAM	18	Lower app quality
Astursalud Cita Previa	16	Lower app quality
Alarma de Medicamentos	13	Lower app quality
Personalized Conversational Health Agent	12	Lower app quality
SACYL Cita Previa	8/39	Lower app quality

[a]The MAG includes 39 criteria. Higher scores indicate that a greater number of app-quality criteria were met.

[b]Descriptive score bands were created for this review because the MAG does not provide predefined cutoff points for interpreting total scores: higher app quality ≥27/39 criteria met, moderate app quality 20-26/39 criteria met, and lower app quality ≤19/39 criteria met. These categories were used only to aid interpretation of the descriptive MAG scores and should not be interpreted as validated MAG thresholds. The full item-level MAG appraisal is provided in Multimedia Appendix 4.

Discussion

Principal Findings

This systematic review examined the current landscape of mobile health apps intended to support the physical and psychological functioning of individuals who are blind or have low vision. Overall, the findings suggest that this field remains limited in scope, uneven in accessibility and quality, and only partially supported by empirical evidence.

Three main findings emerged. First, the number of identified apps was small, and only a subset had been evaluated in peer-reviewed studies. Second, the available apps focused primarily on physical functioning, particularly medication management and health data tracking, whereas no apps specifically targeting psychological functioning were identified. Third, accessibility and app quality varied considerably, with only a small number of apps performing well across both WCAG and MAG criteria. These findings are broadly consistent with a recent systematic review on the usability of mobile apps for individuals with visual impairments [29]. Although that review did not focus specifically on health-related apps or on physical and psychological functioning, it similarly concluded that the usability of mobile apps for visually impaired users remains limited due to device-related barriers (eg, camera-based interfaces) and user interaction challenges (eg, difficulties interpreting graphical user interfaces) [29].

With respect to the first research question, the apps identified through organizational recommendations, and the systematic search of scientific literature reflected a relatively narrow functional landscape. Most supported medication use, medication information access, glucose monitoring, nutrition tracking, or physical activity support. A separate subset of apps primarily supported health care access or administrative functions, such as appointment scheduling, prescription management, or access to regional health system information. These latter tools may be useful for navigating health care services, particularly in local contexts, but they are functionally distinct from clinical self-management apps such as medication management, glucose monitoring, nutrition tracking, or physical activity tools. Overall, the identified apps represent a restricted subset of the possible ways in which mobile health apps might support functioning in individuals who are blind or have low vision. As a result, they do not reflect a broader conceptualization of functioning that includes psychological well-being, self-efficacy, autonomy, and participation in daily life.

Importantly, the characterization of this landscape as limited should not be understood as a simple consequence of the number of apps identified or of the inclusion of geographically restricted administrative tools. Rather, it reflects the narrow range of functions addressed across the identified apps, the limited number supported by population-specific empirical evidence,

XSL•FO
RenderX

the variability observed in accessibility and app quality, and the absence of apps primarily addressing psychological functioning.

A particularly important finding was the absence of apps specifically targeting psychological functioning. This should be interpreted cautiously, as it reflects an absence of identified apps rather than definitive evidence that none exists. Some apps may include psychological or mental health–related components; however, within the scope of this review, we did not identify apps primarily designed for, or empirically evaluated in, blind and low-vision populations with a specific focus on psychological or mental health functioning. Nonetheless, this gap is clinically meaningful given the elevated burden of depression, anxiety, and psychological distress in individuals who are blind or have low vision [19]. This observation is consistent with Inal and colleagues [57], who reviewed 42 usability studies of mobile mental health technologies and found that none included users with visual impairments. Taken together, these findings suggest that psychological functioning remains insufficiently addressed in both app development and research for this population. By contrast, app-based interventions addressing pain-related symptoms and quality of life have been developed for other chronic pain populations. For example, Fibroline incorporates a self-administered cognitive behavioral intervention designed for young people with fibromyalgia or chronic widespread pain, illustrating how psychological and symptom management components can be integrated into a mobile health app outside the context of blindness and low vision [24,58].

With respect to the second research question, the empirical evidence base was limited in both size and scope. In this review, "evidence" refers to the existence of empirical studies examining an app and should not be interpreted as formal validation in a psychometric, regulatory, or certification sense. Only a minority of identified apps had been examined in empirical studies, and the methodological quality of those studies was variable. Importantly, the available evidence differed in its relevance to blind and low-vision users and in the type of outcomes assessed. Some studies directly involved participants who were blind or had low vision, whereas others evaluated apps recommended for this population but did so in broader clinical or general populations where visual impairment status was not reported. Therefore, evidence of effectiveness or usability in general populations should be interpreted as indirectly relevant rather than as population-specific evidence for blind and low-vision users. Importantly, the available evidence also differed according to the type of outcomes that were assessed. Some studies reported objective or clinical outcomes, including medication adherence and blood pressure control, glucose metrics, weight management behavior, or dietary assessment accuracy. Other studies focused primarily on usability, feasibility, acceptability, satisfaction, medication information access, or perceived usefulness. Consequently, the current evidence base is better understood as comprising 2 complementary forms of evidence: studies evaluating objective or clinical outcomes, predominantly conducted in broader populations, and studies examining usability, accessibility, or user experience, more commonly conducted with participants who were blind or had low vision. This distinction is important because the existence of an empirical study does not necessarily imply demonstrated clinical effectiveness, and findings from broader populations should not be interpreted as direct evidence of effectiveness, accessibility, or usability for blind and low-vision users. Studies conducted in broader clinical or general populations examined objective or clinical outcomes for Medisafe [46], including medication adherence and blood pressure control; FreeStyle LibreLink [56] and OneTouch Reveal [47,48], including glycemic metrics; and MyFitnessPal [49,55], including weight management behavior and dietary assessment accuracy. By contrast, studies involving blind or low-vision participants more commonly examined usability, feasibility, medication information access, satisfaction, perceived autonomy, or prototype accessibility for apps such as ScripTalk [53], MyPills [44], FarmaceuticApp [5], Ru Tan Ya [45], and PCHA [51]. Accordingly, the literature provides some evidence of clinical or behavioral benefit for selected apps in broader populations and some direct evidence of usability or perceived value among blind or low-vision users but very limited evidence combining population-specific evaluation with objective health-related outcomes. The review also identified only partial overlap between apps recommended by organizations and apps supported by empirical evidence. That is to say, several apps endorsed by organizations lacked corresponding scientific evaluation, whereas some apps described in the literature were not recommended by organizations. This discrepancy is important because organizational recommendations may shape what users, families, and clinicians perceive as credible or useful. The findings therefore highlight the need for closer alignment between research, practice, and dissemination processes so that recommended tools are not only accessible but also supported by evidence.

With respect to the third research question, accessibility and app quality were inconsistent across the identified apps. Only a small number of apps appeared to perform well across the exploratory WCAG-informed appraisal and the MAG assessment, whereas others showed substantial limitations. In the exploratory user-informed appraisal, some apps endorsed by organizations appeared to present potential accessibility barriers, including unlabeled graphical elements, limited screen reader compatibility, and complex navigation structures. These observations should be interpreted cautiously because they were derived from a small exploratory appraisal rather than from a formal accessibility validation study. Nevertheless, they are consistent with previous work in accessibility, human-computer interaction, and assistive technology, which has highlighted persistent problems related to interface complexity, visually dependent interaction models, and inadequate compatibility with assistive tools [29,59-61]. Our findings extend that literature by showing that these barriers remain highly relevant even in apps with apparent health value or organizational endorsement.

The user-informed accessibility appraisal also suggests that technical accessibility alone may not be sufficient to ensure meaningful use. Even when apps were theoretically compatible with assistive features, their practical value could still be limited by poor onboarding, inadequate support, unfamiliar interaction patterns, or broader digital literacy barriers. Accessibility,

XSL•FO
RenderX

therefore, should not be understood solely as technical compliance but also as the extent to which an app can be used effectively in real-world contexts by its intended users.

At the same time, some apps demonstrated comparatively strong accessibility and overall quality. For example, within the exploratory WCAG-informed appraisal, MyPills received the most favorable accessibility profile, obtaining "High" ratings across all 4 WCAG principles, whereas Medisafe achieved the highest MAG score (34/39), suggesting particularly strong performance in broader app quality domains. More generally, the studies with the strongest designs reported improvements in medication outcomes, such as medication adherence, self-management, and user satisfaction. However, these findings should be interpreted cautiously, given the limited number of studies, their relatively small samples, and the generally high digital literacy of participants. By contrast, many other apps received low accessibility or quality ratings, and the methodological quality of the supporting studies was mixed overall. Several apps endorsed by organizations still showed substantial accessibility limitations across 1 or more WCAG domains and performed less well on the MAG appraisal. For example, despite being recommended by ONCE and the American Foundation for the Blind, MyFitnessPal relied heavily on visually structured content, complex menu hierarchies, and limited screen reader compatibility, which may reduce usability for individuals who are blind or have low vision. This illustrates a potentially important mismatch between organizational endorsement and real-world accessibility. More broadly, even apps that appeared technically accessible could still be difficult to use in practice because of limited onboarding, insufficient guidance, low awareness, and digital literacy barriers. Taken together, these findings suggest that technical accessibility alone is insufficient; effective dissemination, onboarding, and real-world usability are also essential to achieve meaningful impact.

These findings have several important implications. From a clinical perspective, they suggest that only a small number of apps currently offer reasonably credible and accessible tools for supporting physical functioning, particularly medication self-management. Clinicians may wish to be aware of these options but should also recognize that endorsement or availability does not necessarily imply accessibility, quality, or evidence of benefit. At the same time, the absence of identified apps targeting psychological functioning leaves an important gap in the digital support currently available for this population. From a development, policy, and research perspective, the findings point to the need for more rigorous and transparent standards for app development, evaluation, and recommendation. Apps intended for individuals who are blind or have low vision should be developed through participatory and inclusive co-design processes involving end users, clinicians, and accessibility specialists, and should be evaluated not only for intended usefulness but also for real-world accessibility and usability. This perspective is consistent with recent developments reflected in the WCAG 3.0 Working Draft [62]. Although this document is not a ratified standard, it illustrates an ongoing shift from a predominantly technical, criteria-based approach toward a more flexible framework centered on users' diverse accessibility needs. In addition, future research should prioritize apps that address psychological functioning in ways that are both accessible and clinically meaningful and should include larger and more diverse samples, stronger study designs, and outcomes that extend beyond feasibility or satisfaction to capture meaningful indicators of functioning. Greater attention should also be given to accessible onboarding, guidance, and support features to facilitate sustained real-world use. Finally, future work should examine how emerging assistive and AI-enabled functionalities can be integrated into mobile health apps without increasing complexity or introducing new accessibility barriers.

Limitations

This review has several limitations that should be acknowledged. First, the identified studies were heterogeneous in design, geography, sample size, participant characteristics, and outcome measures, limiting comparability and precluding meta-analysis. In addition, although the present review included studies with qualitative and mixed methods components, it was not designed as a formal qualitative evidence synthesis. Applying frameworks such as those proposed in the Cochrane-Campbell Handbook for Qualitative Evidence Synthesis would entail a substantially different review design but may represent a valuable direction for future research focused on in-depth analysis of user experience and interaction design in this field. Second, restricting the search to English and Spanish publications may have excluded relevant studies in other languages, and despite a comprehensive search strategy, some studies may have been missed. In addition, some apps identified through the literature could not be directly assessed because they were not available for download at the time of review, which may introduce a degree of availability bias. Moreover, identifying apps through organizational recommendations may have introduced selection bias, as not all organizations publish or regularly update such recommendations. This process also introduced geographic and contextual limitations. Several organizationally identified apps were region-specific health care access or administrative tools linked to regional public health systems, which may limit their relevance to international readers and make them functionally distinct from clinical self-management apps. Therefore, the number and nature of identified apps should be interpreted in light of the sources consulted and the organizational recommendation process. Finally, the accessibility appraisal should be interpreted as a complementary and exploratory component intended to contextualize the review findings. It was not designed as a stand-alone usability study, formal accessibility validation, or psychometric assessment. Therefore, its findings should be understood as illustrative user-informed observations of potential accessibility barriers rather than as generalizable estimates of usability or definitive measurements of accessibility across the broader population of users who are blind or have low vision. However, its use is consistent with prior accessibility literature and with W3C guidance supporting the application of WCAG principles to mobile and nonweb software.

Despite the study's limitations, this review shows that the mobile health apps identified through organizational recommendations and the scientific literature remain limited in scope,

XSL•FO
RenderX

heterogeneous in function, uneven in accessibility and quality, and supported by a limited and heterogeneous empirical base. While a small number of apps appear promising, major gaps remain in accessibility, quality, and especially support for psychological functioning. Moving the field forward will require more than simply increasing the number of available apps. Future efforts should focus on developing apps that are accessible from the outset, rigorously evaluated, and developed in collaboration with the people they are designed to support.

Acknowledgments

No generative AI tools were used at any stage in the preparation of this manuscript.

Funding

This work was partly funded by grants from the Spanish Ministry of Economy, Industry and Competitiveness (RED2022-134869-T), the European Regional Development Fund (ERDF), and the Government of Catalonia (AGAUR; 2021SGR-730). ES is a Serra Húnter Fellow at the Department of Psychology, Universitat Rovira i Virgili, Catalonia, Spain. JM's work is supported by ICREA-Acadèmia. The Chair in Pediatric Pain is supported by Fundación Grünenthal and ANUBIS.

Authors' Contributions

Conceptualization: JM (lead), CM (supporting)
Methodology: JM (lead), CM (supporting)
Investigation: CM (lead), JM (supporting), PLl-V (supporting)
Data curation: CM (lead), JM (supporting), PLl-V (supporting)
Writing – original draft: JM (lead), CM (supporting)
Writing – review & editing: JM (lead), CM (supporting), ES (supporting), PLl-V (supporting)
Supervision: JM (lead)
Project administration: JM (lead)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of updated search.
[DOCX File, 33 KB - mhealth_v14i1e85809_app1.docx]

Multimedia Appendix 2
Data extraction template.
[DOCX File, 26 KB - mhealth_v14i1e85809_app2.docx]

Multimedia Appendix 3
Operationalization and scoring of Web Content Accessibility Guidelines (WCAG) 2.1 principles in the appraisal of mobile health apps.
[DOCX File, 27 KB - mhealth_v14i1e85809_app3.docx]

Multimedia Appendix 4
Quality assessment of health-related mobile apps for individuals who are blind or have low vision, based on the Mobile App Development and Assessment Guide criteria.
[DOCX File, 28 KB - mhealth_v14i1e85809_app4.docx]

Checklist 1
PRISMA checklist.
[DOCX File, 32 KB - mhealth_v14i1e85809_app5.docx]

References

1. World Health Organization. World Report on Vision: World Health Organization; 2019.
2. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013 Jun 14;15(6):e120. [doi: 10.2196/jmir.2600] [Medline: 23770578]

XSL•FO
RenderX

3. Court H, McLean G, Guthrie B, Mercer SW, Smith DJ. Visual impairment is associated with physical and mental comorbidities in older adults: a cross-sectional study. BMC Med 2014 Oct 17;12:181. [doi: 10.1186/s12916-014-0181-7] [Medline: 25603915]
4. Demmin DL, Silverstein SM. Visual impairment and mental health: unmet needs and treatment options. Clin Ophthalmol 2020;14:4229-4251. [doi: 10.2147/OPTH.S258783] [Medline: 33299297]
5. Madrigal-Cadavid J, Amariles P, Pino-Marín D, Granados J, Giraldo N. Design and development of a mobile app of drug information for people with visual impairment. Res Social Adm Pharm 2020 Jan;16(1):62-67. [doi: 10.1016/j.sapharm.2019.02.013] [Medline: 30853509]
6. Quinn TJ, Livingstone I, Weir A, et al. Accuracy and feasibility of an Android-based digital assessment tool for post stroke visual disorders-the StrokeVision app. Front Neurol 2018;9:146. [doi: 10.3389/fneur.2018.00146] [Medline: 29643830]
7. Mora C, Jensen MP, Miró J. Chronic pain in individuals who have visual impairments: a systematic review. J Vis Impair Blind 2026 Mar;120(2):117-133. [doi: 10.1177/0145482X251392565]
8. Coleman AL, Cummings SR, Ensrud KE, et al. Visual field loss and risk of fractures in older women. J Am Geriatr Soc 2009 Oct;57(10):1825-1832. [doi: 10.1111/j.1532-5415.2009.02432.x] [Medline: 19702619]
9. Hong T, Mitchell P, Burlutsky G, Samarawickrama C, Wang JJ. Visual impairment and the incidence of falls and fractures among older people: longitudinal findings from the Blue Mountains Eye Study. Invest Ophthalmol Vis Sci 2014 Nov 1;55(11):7589. [doi: 10.1167/iovs.14-14262]
10. Kuang TM, Tsai SY, Hsu WM, Cheng CY, Liu JH, Chou P. Visual impairment and falls in the elderly: the Shihpai Eye Study. J Chin Med Assoc 2008 Sep;71(9):467-472. [doi: 10.1016/S1726-4901(08)70150-3] [Medline: 18818140]
11. Zhang X, Bullard KM, Cotch MF, et al. Association between depression and functional vision loss in persons 20 years of age or older in the United States, NHANES 2005-2008. JAMA Ophthalmol 2013 May;131(5):573-581. [doi: 10.1001/jamaophthalmol.2013.2597] [Medline: 23471505]
12. Lam BL, Christ SL, Lee DJ, Zheng DD, Arheart KL. Reported visual impairment and risk of suicide: the 1986-1996 national health interview surveys. Arch Ophthalmol 2008 Jul;126(7):975-980. [doi: 10.1001/archopht.126.7.975] [Medline: 18625946]
13. Palbo A, Roed Rasmussen ML, Hansen MS, Subhi Y. Association between visual impairment and suicide: a systematic review and meta-analysis. Acta Ophthalmol 2024 May;102(3):263-273. [doi: 10.1111/aos.15730] [Medline: 37395145]
14. Cupples ME, Hart PM, Johnston A, Jackson AJ. Improving healthcare access for people with visual impairment and blindness. BMJ 2012 Jan 30;344(jan30 1):e542. [doi: 10.1136/bmj.e542] [Medline: 22290101]
15. Kim HN. Mobile health technology accessible to people with visual impairments. J Technol Pers Disabil 2019;7 [FREE Full text]
16. Rodin A, Shachak A, Miller A, Akopyan V, Semenova N. Mobile apps for eye care in Canada: an analysis of the iTunes store. JMIR Mhealth Uhealth 2017 Jun 14;5(6):e84. [doi: 10.2196/mhealth.7055] [Medline: 28615154]
17. Evans JR, Fletcher AE, Wormald RPL. Depression and anxiety in visually impaired older people. Ophthalmology 2007 Feb;114(2):283-288. [doi: 10.1016/j.ophtha.2006.10.006]
18. Sturrock BA, Xie J, Holloway EE, et al. The prevalence and determinants of desire for and use of psychological support in patients with low vision. Asia Pac J Ophthalmol (Phila) 2014;3(5):286-293. [doi: 10.1097/APO.0000000000000012] [Medline: 26107915]
19. van der Aa HPA, Hoeben M, Rainey L, van Rens GHMB, Vreeken HL, van Nispen RMA. Why visually impaired older adults often do not receive mental health services: the patient's perspective. Qual Life Res 2015 Apr;24(4):969-978. [doi: 10.1007/s11136-014-0835-0] [Medline: 25398494]
20. Withers J, Speight C. Health care for individuals with hearing loss or vision loss. N C Med J 2017 Mar;78(2):107-112. [doi: 10.18043/ncm.78.2.107]
21. Burton DM, Uslan MM, Blubaugh MV, Clements CW 3rd. Are current insulin pumps accessible to blind and visually impaired people? J Diabetes Sci Technol 2009 May 1;3(3):613-618. [doi: 10.1177/193229680900300327] [Medline: 20144301]
22. Martín ISM, Fernández MG, Yurrita LC. Mobile applications for nutrition, dietetics and healthy habits; analysis and consequences of an increasing trend [Article in Spanish]. Nutr Hosp 2014;30(1):15-24. [doi: 10.3305/nh.2014.30.1.7398]
23. de la Vega R, Roset R, Castarlenas E, Sánchez-Rodríguez E, Solé E, Miró J. Development and testing of painometer: a smartphone app to assess pain intensity. J Pain 2014 Oct;15(10):1001-1007. [doi: 10.1016/j.jpain.2014.04.009] [Medline: 24854065]
24. Miró J, Lleixà-Daga M, de la Vega R, Llorens-Vernet P, Jensen MP. A mobile application to help self-manage pain severity, anxiety, and depressive symptoms in patients with fibromyalgia syndrome: a pilot study. Int J Environ Res Public Health 2022 Sep 23;19(19):12026. [doi: 10.3390/ijerph191912026] [Medline: 36231327]
25. El-Sappagh S, Ali F, El-Masri S, Kim K, Ali A, Kwak KS. Mobile health technologies for diabetes mellitus: current state and future challenges. IEEE Access 2018;7:21917-21947. [doi: 10.1109/ACCESS.2018.2881001]
26. Thangada ND, Garg N, Pandey A, Kumar N. The emerging role of mobile-health applications in the management of hypertension. Curr Cardiol Rep 2018 Jul 26;20(9):78. [doi: 10.1007/s11886-018-1022-7] [Medline: 30046971]

XSL•FO
RenderX

27. Martín-Martín J, Muro-Culebras A, Roldán-Jiménez C, et al. Evaluation of Android and Apple Store depression applications based on Mobile Application Rating Scale. Int J Environ Res Public Health 2021 Nov 27;18(23):12505. [doi: 10.3390/ijerph182312505] [Medline: 34886232]
28. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22. [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
29. Al-Razgan M, Almoaiqel S, Alrajhi N, et al. A systematic literature review on the usability of mobile applications for visually impaired users. PeerJ Comput Sci 2021;7:e771. [doi: 10.7717/peerj-cs.771] [Medline: 34901428]
30. de la Vega R, Miró J. mHealth: a strategic field without a solid scientific soul. a systematic review of pain-related apps. PLoS One 2014;9(7):e101312. [doi: 10.1371/journal.pone.0101312] [Medline: 24999983]
31. Groen G, Jörns-Presentati A, Dessauvagie A, et al. Development of a mobile application for detection of adolescent mental health problems and feasibility assessment with primary health care workers. Issues Ment Health Nurs 2022 Nov;43(11):1046-1055. [doi: 10.1080/01612840.2022.2124003] [Medline: 36205922]
32. Horton EL, Renganathan R, Toth BN, et al. A review of principles in design and usability testing of tactile technology for individuals with visual impairments. Assist Technol 2017;29(1):28-36. [doi: 10.1080/10400435.2016.1176083] [Medline: 27187665]
33. Morris JT, Sweatman WM, Jones ML. Smartphone use and activities by people with disabilities: user survey 2016. J Technol Pers Disabil 2017 [FREE Full text]
34. Llorens-Vernet P, Miró J. The Mobile App Development and Assessment Guide (MAG): Delphi-based validity study. JMIR Mhealth Uhealth 2020 Jul 31;8(7):e17760. [doi: 10.2196/17760] [Medline: 32735226]
35. Llorens-Vernet P, Miró J. Standards for mobile health-related apps: systematic review and development of a guide. JMIR Mhealth Uhealth 2020 Mar 3;8(3):e13057. [doi: 10.2196/13057] [Medline: 32130169]
36. Huckvale K, Prieto JT, Tilney M, Benghozi PJ, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med 2015 Dec;13(1):1-13. [doi: 10.1186/s12916-015-0444-y]
37. Miró J, Llorens-Vernet P. Assessing the quality of mobile health-related apps: interrater reliability study of two guides. JMIR Mhealth Uhealth 2021 Apr 19;9(4):e26471. [doi: 10.2196/26471] [Medline: 33871376]
38. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
39. PROSPERO: international prospective register of systematic reviews. : Centre for Reviews and Dissemination, University of York; 2025. URL: https://www.crd.york.ac.uk/prospero/ [accessed 2026-10-06]
40. Farley JB, Stein J, Keogh JWL, Woods CT, Milne N. The relationship between physical fitness qualities and sport-specific technical skills in female, team-based ball players: a systematic review. Sports Med Open 2020 Apr 15;6(1):18. [doi: 10.1186/s40798-020-00245-y] [Medline: 32297147]
41. Garriga-Cazorla H, Roman-Juan J, Martí L, Solé E, Martínez-Leal R, Miró J. Chronic pain in autism. Clin J Pain 2025;41(8). [doi: 10.1097/AJP.0000000000001296]
42. Martí L, Garriga-Cazorla H, Roman-Juan J, Miró J. Family-related factors in children and adolescents with chronic pain: a systematic review. Eur J Pain 2025 Jul;29(6):e70038. [doi: 10.1002/ejp.70038] [Medline: 40407120]
43. Web Content Accessibility Guidelines (WCAG) 21 W3C recommendation. World Wide Web Consortium (W3C). 2018 Jun 5. URL: https://www.w3.org/TR/WCAG21/ [accessed 2025-10-02]
44. Nedovic T, Umeri-Sali N, Denecke K. Supporting blind and visually impaired persons in managing their medication. Stud Health Technol Inform 2019 Sep 3;267:189-196. [doi: 10.3233/SHTI190826] [Medline: 31483272]
45. Nimmolrat A, Khuwuthyakorn P, Wientong P, Thinnukool O. Pharmaceutical mobile application for visually-impaired people in Thailand: development and implementation. BMC Med Inform Decis Mak 2021 Jul 16;21(1):217. [doi: 10.1186/s12911-021-01573-z] [Medline: 34271932]
46. Morawski K, Ghazinouri R, Krumme A, et al. Association of a smartphone application with medication adherence and blood pressure control: the MedISAFE-BP randomized clinical trial. JAMA Intern Med 2018 Jun 1;178(6):802-809. [doi: 10.1001/jamainternmed.2018.0447] [Medline: 29710289]
47. Grady M, Cameron H, Holt E. Sustained improvements in readings in-range using an advanced Bluetooth® connected blood glucose meter and a mobile diabetes app: real-world evidence from more than 55,000 people with diabetes. Diabetes Ther 2023 Jun;14(6):1023-1035. [doi: 10.1007/s13300-023-01415-3] [Medline: 37138183]
48. Merino Torres JF, Grady M, Lopez Alba A, et al. Clinical practice experiences using a professional diabetes management ecosystem during COVID. J Diabetes Sci Technol 2023 Mar;17(2):390-399. [doi: 10.1177/19322968211065778] [Medline: 34957884]
49. Banal MG, Bongga D, Angbengco JM, Amarra S, Panlasigui L. MyFitnessPal smartphone application: relative validity and intercoder reliability among dietitians in assessing energy and macronutrient intakes of selected Filipino adults with obesity. BMJ Nutr Prev Health 2024;7(1):54-60. [doi: 10.1136/bmjnph-2023-000770] [Medline: 38966115]
50. Shade M, Boron J, Manley N, Kupzyk K, Pullen C. Ease of use and usefulness of medication reminder apps among rural aging adults. J Community Health Nurs 2019;36(3):105-114. [doi: 10.1080/07370016.2019.1630960] [Medline: 31291770]

XSL•FO
RenderX

51. Choi S, Seo JY, Krishnan A, Kamath S, Kitsiou S, Haegele J. Development of a personalized conversational health agent to enhance physical activity for blind and low-vision individuals. Mhealth 2025;11:29. [doi: 10.21037/mhealth-24-60] [Medline: 40755937]
52. González Oñate C, Fanjul Peyró C. Mobile applications for the elderly: a study of their current strategy [Article in Spanish]. RIFIE 2018;47(1):107. [doi: 10.17811/rifie.47.1.2018.107-112]
53. Spektor T, Nikolic N, Lekakh O, Gaynes BI. Efficacy of ScripTalk automated prescription label reader and veterans with visual impairments. J Vis Impair Blind 2015 Sep;109(5):412-418. [doi: 10.1177/0145482X1510900511]
54. Bee YM, Batcagan-Abueg APM, Chei CL, et al. A smartphone application to deliver a treat-to-target insulin titration algorithm in insulin-naive patients with type 2 diabetes: a pilot randomized controlled trial. Diabetes Care 2016 Oct;39(10):e174-e176. [doi: 10.2337/dc16-0419] [Medline: 27506223]
55. Gordon M, Althoff T, Leskovec J. Goal-setting and achievement in activity tracking apps: a case study of MyFitnessPal. 2019 May 13 Presented at: WWW '19: The World Wide Web Conference; May 13-17, 2019; San Francisco CA USA p. 571-582. [doi: 10.1145/3308558.3313432]
56. Kao K, Brandner L, Virdi N. Comparison of glucose metrics between users of CGM readers and CGM-connected apps. J Diabetes Sci Technol 2022 Jan;16(1):259-260. [doi: 10.1177/19322968211044141] [Medline: 34493103]
57. Inal Y, Wake JD, Guribye F, Nordgreen T. Usability evaluations of mobile mental health technologies: systematic review. J Med Internet Res 2020 Jan 6;22(1):e15337. [doi: 10.2196/15337] [Medline: 31904579]
58. de la Vega R, Roset R, Galán S, Miró J. Fibroline: a mobile app for improving the quality of life of young people with fibromyalgia. J Health Psychol 2018 Jan;23(1):67-78. [doi: 10.1177/1359105316650509] [Medline: 27225285]
59. Theodorou P, Tsiligkos K, Meliones A, Filios C. An extended usability and UX evaluation of a mobile application for the navigation of individuals with blindness and visual impairments outdoors-an evaluation framework based on training. Sensors (Basel) 2022 Jun 16;22(12):4538. [doi: 10.3390/s22124538] [Medline: 35746320]
60. Kane SK, Morris MR, Perkins AZ, Wigdor D, Ladner RE, Wobbrock JO. Access overlays: improving non-visual access to large touch screens for blind users. UIST'11: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology 2011 Oct:273-282. [doi: 10.1145/2047196.2047232]
61. Grussenmeyer W, Folmer E. Accessible touchscreen technology for people with visual impairments. ACM Trans Access Comput 2017 Jun 30;9(2):1-31. [doi: 10.1145/3022701]
62. W3C Accessibility Guidelines (WCAG) 30 (working draft). World Wide Web Consortium (W3C). 2026 Mar 3. URL: https://www.w3.org/TR/2026/WD-wcag-3.0-20260303/ [accessed 2026-04-10]

Abbreviations

MAG: Mobile App Development and Assessment Guide
NCBI: National Center for Biotechnology Information
NIH: National Institutes of Health
ONCE: Spanish National Organization for the Blind
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
W3C: World Wide Web Consortium
WCAG: Web Content Accessibility Guidelines

Edited by NM Ali; submitted 13.Oct.2025; peer-reviewed by N Ross, SS Kamath; revised version received 24.Jul.2026; accepted 04.Aug.2026; published 12.Nov.2026.

Please cite as:
Mora C, Solé E, Llorens-Vernet P, Miró J
Mobile Health Apps for Individuals Who Are Blind or Have Low Vision: Systematic Review of Evidence, Accessibility, and App Quality
JMIR Mhealth Uhealth 2026;14:e85809
URL: https://mhealth.jmir.org/2026/1/e85809
doi:10.2196/85809

© Carlos Mora, Ester Solé, Pere Llorens-Vernet, Jordi Miró. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 12.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Frontline Health Workers' Perspectives of the World Health Organization Skin Neglected Tropical Diseases App in Kenya: Qualitative Study on AI-Embedded mHealth Implementation

Emily E V Quilter[1], MSc; Ruth Nyangacha[2], PhD; Esther Kinyeru[3], MBChB, MSc; Aïna Fuster Casanovas[4,5], PhD; Amberly Brigden[1], PhD; Kenton O'Hara[6], PhD; Diana Atieno[2], BSc; José A Ruiz-Postigo[7], PhD; Carme Carrion[4,8], PhD

[1]Digital Health and Care, University of Bristol, One Cathedral Square, Trinity Street, Bristol, United Kingdom

[2]Center for Traditional Medicine and Drug Research, Kenyan Medical Research Institute, Nairobi, Kenya

[3]Department for Neglected Tropical Disease, Disease Control and Prevention Cluster, World Health Organization, Nairobi, Kenya

[4]eHealth Lab Research Group, eHealth Center, School of Health Sciences, Universitat de Catalunya, Barcelona, Spain

[5]Vall d'Hebron University Hospital, Institut Català de la Salut, Innovation Unit, Quality, Processes and Innovation Directorate, Barcelona, Spain

[6]School of Computer Science, University of Bristol, Bristol, United Kingdom

[7]Department of Control of Neglected Tropical Diseases, Prevention, Treatment and Care Unit, World Health Organization, Geneva, Switzerland

[8]Department of Medical Sciences, School of Medicine, Universitat de Girona, Girona, Spain

Corresponding Author:
Emily E V Quilter, MSc
Digital Health and Care, University of Bristol, One Cathedral Square, Trinity Street, Bristol, United Kingdom

Abstract

Background: Skin neglected tropical diseases (NTDs) pose significant diagnostic and management challenges in resource-limited settings due to constrained dermatological expertise, frontline health worker (FHW) training, and limited access to diagnostic resources. Mobile health apps with artificial intelligence (AI)–enabled diagnostic imaging capabilities have the potential to enhance clinical decision-making and professional development at the primary care level. The World Health Organization (WHO) skin NTD mobile app uses convolutional neural networks to analyze images of skin lesions and generate differential diagnoses, intended to be used alongside clinical history and examination, to support FHWs in identifying 12 skin NTDs and 24 common skin conditions. Beyond clinical decision support, the app also aims to upskill FHWs in the recognition and management of these diseases. However, the success of such tools depends on understanding users' needs and the realities of implementation in diverse clinical contexts.

Objective: This study aimed to explore FHWs' perspectives on the real-world use and impact of the AI-embedded WHO Skin NTDs app on diagnostic workflows, dermatological understanding, clinical decision-making, and FHW-patient interactions across diverse health care delivery settings in Kenya.

Methods: This qualitative study involved 36 FHWs from 5 skin NTD-endemic counties in Kenya. Following a training workshop, FHWs integrated the app into routine clinical workflows from June to October 2024. Data were collected through 15 semistructured interviews (each 30 - 45 minutes) and 4 focus group discussions (1 - 1.5 hours) exploring FHW experiences across diverse health care delivery contexts. All sessions were audio-recorded, transcribed verbatim, and thematically analyzed using NVivo (QSR International), using a bottom-up inductive coding approach.

Results: FHWs reported that the app facilitated a shift from habitual referral to more proactive case management at the local-level facility, reinforcing clinical ownership and positioning them as local dermatology reference points. It was perceived to enhance diagnostic confidence, strengthen patient trust, and encourage community engagement. Some FHWs described how the app helped mitigate situations for patient stigma due to decreased reliance on public colleague consultations. However, technical limitations (eg, internet dependency and algorithmic errors) constrained consistent use. While most FHWs used the app in line with its intended role as an assistive tool, a minority reported situations of diagnostic deferral to the AI output, highlighting potential considerations of clinical autonomy.

Conclusions: The WHO Skin NTDs app shows strong potential to strengthen frontline dermatological capacity that aligns with WHO strategies to decentralize NTD care and promote "skin health for all." Our findings underscore the importance of embedding such tools within ethical and pedagogical frameworks that protect clinical autonomy and foster sustainable capacity building. Further research will examine real-world use in situ to guide context-specific governance, ensuring that this AI-embedded tool enhances—rather than displaces—clinical reasoning and epistemic authority.

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2026;14:e81829) doi:10.2196/81829

KEYWORDS

mHealth; artificial intelligence; clinical decision support; skin neglected tropical diseases; frontline health workers; digital health; Kenya; qualitative research; health systems; World Health Organization; human-AI collaboration

Introduction

Neglected tropical diseases (NTDs) are predominantly diseases of those in poverty and impact over 1 billion people worldwide [1]. Skin NTDs (which predominantly affect the skin) constitute a significant subset, accounting for nearly half of all NTDs and disproportionately affect marginalized populations in remote rural areas [2]. The World Health Organization (WHO) has categorized skin NTDs together due to their coendemicity, similar initial symptom presentation, common differential diagnosis, and the potential for integrated multidisease control interventions [1]. Despite their widespread prevalence, skin NTDs are often underreported, resulting in an underestimation of their true prevalence and burden on global health [3].

The consequences of skin NTDs extend beyond physical symptoms. Chronic illness, disability, and visible disfigurement often intersect with stigma, social exclusion, and economic hardship [4,5]. These factors delay care-seeking and contribute to ongoing transmission, disease progression, and long-term disability [6]. In many settings, especially in sub-Saharan Africa, access to specialized dermatological care is minimal—with ratios of 1 dermatologist per 500,000 to 1 million people not uncommon. As a result, frontline health workers (FHWs) often serve as the first and only point of contact for patients with skin conditions [7,8].

While FHWs play a critical role in the early identification and management of skin conditions, most have only minimal formal dermatology training [9,10]. For some of the rarer skin NTDs, this knowledge gap is further compounded by limited opportunities for experiential learning and skill retention [11]. While visual examination remains the most accessible diagnostic method, it requires pattern recognition and contextual clinical judgment—capabilities often undersupported in primary care [12]. Furthermore, advanced diagnostic methods such as polymerase chain reaction and microscopy, the gold-standard for diagnosing many skin NTDs, are often unavailable to FHWs at the grassroots level, where most disease detection and management occur [2,11,13].

To address this gap, the WHO has begun to promote a more integrated approach to skin NTD care through its 2021 - 2030 road map, which frames “skin health for all” as foundational to universal health coverage [1]. The strategy emphasizes decentralized care, crosscutting interventions, and capacity-building for FHWs. This agenda was reinforced at the 2025 World Health Assembly where a resolution was adopted recognizing skin diseases as a significant public health concern [14]. Unlike some NTDs managed through mass drug administration, most skin NTDs require individualized diagnosis and longer-term follow-up, increasing the need for clinical decision support at the frontline [2,15]. Enhancing FHW diagnostic dermatological capacity is therefore a strategic priority for reaching skin NTD control and elimination targets [1].

Digital health technologies, particularly mobile health (mHealth) tools, offer promising opportunities to strengthen frontline dermatological care through capacity building [16-18]. More specifically, by embedding artificial intelligence (AI), such as image-based diagnostic algorithms, within these mHealth tools, there is the potential to support FHWs by providing real-time clinical decision support, fostering diagnostic confidence, and facilitating experiential learning in situ [18,19] (in this study, we use this term AI-embedded to refer to mHealth tools that incorporate machine learning models as a defined component of their functionality).

In response to long-standing challenges in the recognition and management of skin conditions at the primary care level, the WHO Skin NTDs app was developed in collaboration with the WHO and No Leprosy Remains [20,21]. The app is designed to assist FHWs in recognizing, documenting, and managing skin NTDs and other common skin conditions, functioning as an assistive clinical decision support tool to enhance clinical diagnostic confidence and capacity [20-22]. The app includes both educational and AI-embedded components, the latter providing image-based differential diagnoses to support clinical reasoning in situ. A more detailed description of the app and its functionality is provided in the Methods section.

However, understanding how such tools are integrated into real-world clinical workflows, and how they may influence clinical reasoning, referral patterns, professional autonomy, and patient experience, is critical. AI-embedded tools in health care can introduce both opportunities and risks, particularly if they shift clinical responsibility or impact human judgment [23]. Evaluating whether such tools are used as intended, and within the boundaries of clinical competence, is therefore central to responsible mHealth innovation and implementation [23]. This study also seeks to examine how the app does or does not support its intended educational use and values, providing evidence to inform its usefulness, and identifying opportunities for iterative design and implementation improvements.

This study explores the use and acceptability of the WHO Skin NTDs app within frontline dermatological care in Kenya, a country with a pluralistic, complex health system and significant gaps in dermatological service provision. Health care delivery in Kenya operates across multiple levels (national referral, county-level, and community-based) [24]. Kenya is endemic for at least 9 skin NTDs including chromoblastomycosis, cutaneous leishmaniasis, post–kala-azar dermal leishmaniasis, leprosy, lymphatic filariasis, mycetoma, scabies, sporotrichosis, and Tungiasis [25,26]. Common skin conditions such as superficial fungal infections are highly prevalent, affecting an estimated 11.57% of the population [27]. Facility-level studies indicate even higher burdens at the point of care; for example,

XSL•FO
RenderX

a level 5 hospital in Kiambu County reported eczema in 25.5% of dermatological consultations [28].

This study therefore asks: How do FHWs perceive, interpret, and engage with the AI-embedded component of the WHO Skin NTDs app in real-world clinical settings? Additionally, what do these interactions reveal about the pedagogical, ethical, and systems-level dynamics of deploying AI-embedded clinical decision support tools in primary dermatological care within resource-limited settings?

Methods

The COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist was used to guide reporting, available in Checklist 1 [29].

Study Design

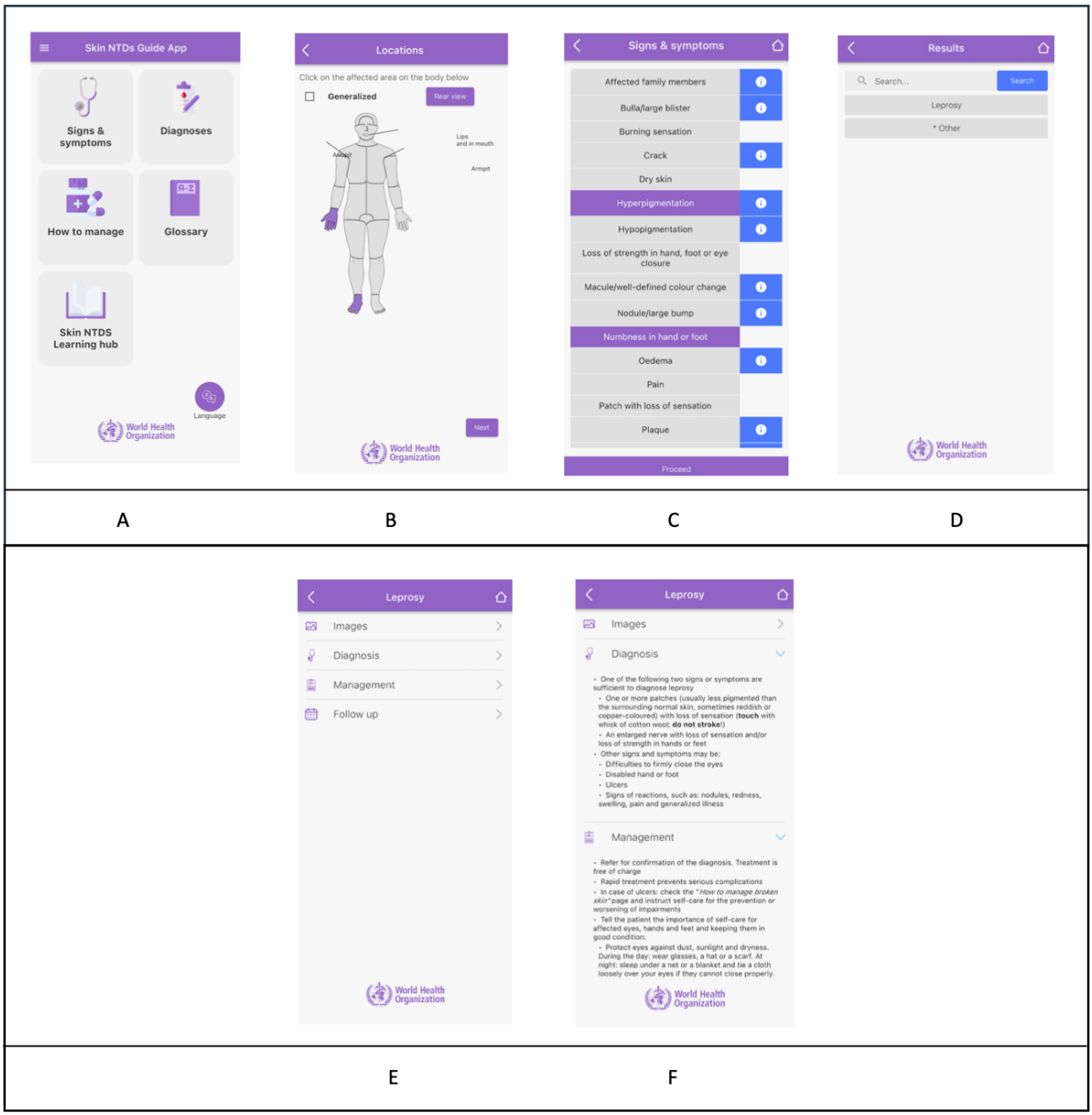
This qualitative study explored the experiences of FHWs integrating the WHO Skin NTDs app into routine clinical workflows across 5 counties in Kenya—Baringo, Nakuru, West Pokot, Kwale, and Kajiado, purposefully selected by the Kenyan Ministry of Health due to their endemicity for skin NTDs. Participants represented FHWs across Kenya's primary health care system, including national referral (level 5), county-level (levels 3 and 4), and community-level (level 2) facilities. Facilities were intentionally sampled across multiple levels of the health system to capture a diversity of clinical and community contexts, resource availability, and workflow structures shaping dermatological care delivery. These levels differ in staffing (eg, nurse-led [level 2] vs clinician-led [levels 3 - 5] services), access to diagnostic and treatment resources, opportunities for peer consultation, and referral pathways. The aim of sampling across multiple facility levels and regions in this way was not to support any systematic formal comparison between levels. Rather, sampling across different facilities and regions provided an opportunity for particular distinguishing factors of these varied settings to be invoked by participants, as appropriate, in their explanatory accounts of usage, perceived usefulness, and impact. Data were collected through focus group discussions (FGDs) and semistructured interviews (SSIs) with participating FHWs. This qualitative component forms part of a broader assessment study on the performance of the app's AI algorithm [30].

WHO Skin NTDs App

The WHO Skin NTDs app comprises two integrated components: (1) an educational component providing structured guidance on diagnosis and management, and (2) an AI-embedded component that generates differential diagnoses from uploaded images of skin lesions. The educational component includes interactive skin lesion localization and condition-specific information, clinical images, and management guidance adapted from the WHO training guide on skin NTDs for FHWs (Figure 1) [5]. This can be accessed during or outside clinical consultations to support learning and decision-making.

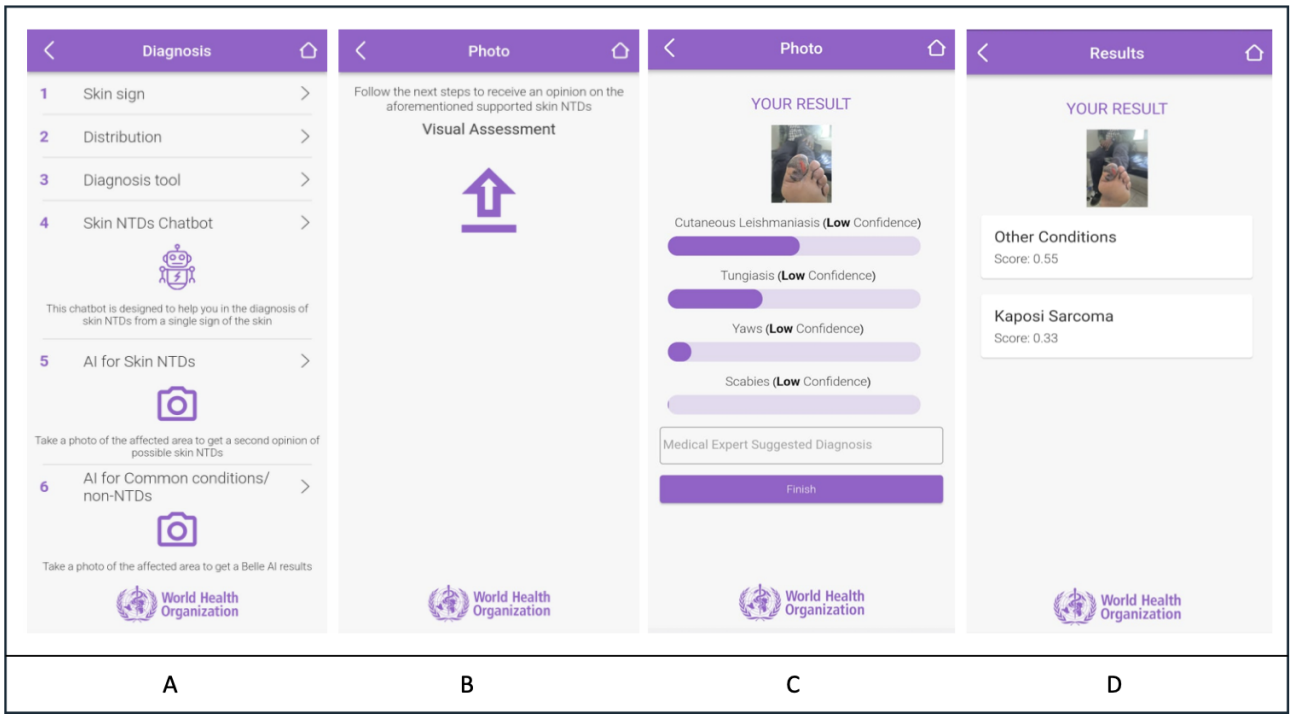
XSL•FO
RenderX

Figure 1. Screenshots from the educational component of the World Health Organization (WHO) Skin NTD app. This component coexists with the artificial intelligence–embedded diagnostic functionality (shown in Figure 2) in the current version of the app. (A) Main home screen of the WHO Skin NTDs app; (B) Body map interface used to indicate the anatomical location of skin lesions; (C) Signs and symptoms interface, where users can select multiple features based on the patient's clinical presentation (or for exploratory learning outside the clinical encounter); (D) Results screen displaying potential conditions (eg, leprosy) based on the selected lesion location and symptoms; (E) Condition-specific menu (eg, leprosy) displaying sections for images, diagnosis, management, and follow-up; (F) Example educational content page providing clinical information on diagnosis and management. NTDs: Neglected Tropical Diseases.



XSL•FO
RenderX

Figure 2. Screenshots from the AI-embedded component of the World Health Organization skin NTD app. (A) Interface of the top-level screen for the AI-embedded component, showing 2 entry points for image upload corresponding to the skin NTD and common skin condition algorithms. (B) Image upload interface for the skin NTD algorithm, where users can capture a photograph within the app or upload an existing image from the device gallery. (C) Example output displaying AI-generated differential diagnoses with associated confidence levels for the skin NTD algorithm. (D) Example output displaying AI-generated differential diagnoses for the common skin condition algorithm; "Other conditions" denotes outputs falling outside the predefined set of 24 common skin conditions included in the model. AI: artificial intelligence; NTDs: Neglected Tropical Diseases.



Within the "Signs & symptoms" component, users first indicate the anatomical location of the lesion using a body map interface (Figure 1A). They then select relevant signs and symptoms from a structured list (Figure 1B and C) and based on these inputs, the app generates a list of potential conditions, which users can explore further to access detailed information on diagnosis and management (Figure 1D-F).

The AI-embedded component is based on convolutional neural network models that analyze and classify uploaded skin lesion images. For each uploaded image, the model generates a small set of differential diagnoses with associated confidence levels (low, medium, and high) across 12 skin NTDs and 24 common skin conditions [21,22] (Figure 2). Given that over 80% of skin conditions encountered by FHWs are common skin conditions, their inclusion is essential for clinical usefulness [31].

The WHO Skin NTDs app is designed to be used during routine patient consultations, although no standardized protocol governs its integration within clinical workflows. In practice, use is flexible and embedded within the consultation itself, typically following initial history taking and physical examination. The anticipated workflow involves FHWs capturing an image of the skin lesion with the phone camera, uploading it to both the skin NTD and the common skin condition algorithms, and reviewing AI-generated differential diagnoses. These outputs are then interpreted alongside patient history and clinical examination findings. The app's outputs are not prescriptive; rather, they are intended to support clinical reasoning, enabling FHWs to refine or rule out potential conditions. However, given the app's finite disease library, the true diagnosis may not always be represented.

The algorithm was trained on a dataset comprising 5760 images of skin NTDs alongside 16,577 images of other skin conditions, 2469 images of healthy skin, and 50,000 non–skin images to support classification boundaries [30]. The training set included a diverse range of skin tones, with approximately 70% of images representing dark skin types addressing key concerns around bias in dermatology AI [30,32]. A more detailed account of the app's design, development process, and training approach is provided elsewhere [30].

Participants and Recruitment

Participants were nondermatologist FHWs who routinely manage patients with skin NTDs or common skin conditions at the primary care level. Five counties were purposively selected in collaboration with the Ministry of Health, Kenya. Within each county, health facilities and affiliated FHWs were identified with support from the County Departments of Health. From this sampling frame, 50 FHWs were randomly selected and invited by telephone to participate in an initial training and deployment workshop in April 2024.

Training and Deployment

The training and deployment workshop, held at a centrally located training center in Kenya, introduced participants to the WHO Skin NTDs app. The training combined short didactic sessions with practical exercises. Content included (1) step-by-step instruction on app navigation, image capture, and interpretation of AI-generated differential diagnoses, and (2) guidance on consent practices. Participants engaged in supervised practical sessions where they used the app on peers

XSL•FO
RenderX

to familiarize themselves with consent practices, image capture, and result interpretation.

Following the first training workshop, participants were added to a WhatsApp group with members of the study team to facilitate ongoing technical support and troubleshooting during the deployment period. Participants were invited to integrate the app into their routine clinical practice over a deployment period of approximately 6 months. After the 6-month deployment period, a second evaluation session was convened in November 2024, again at a central location in Kenya. This session focused on understanding participant reflections on their experience using the app and comprised the main qualitative data collection activities. During this session, FGDs and SSIs were conducted to explore participants' experiences using the app, including perceived benefits, challenges, and contextual influences on use. Of the original 50 FHWs invited, 47 attended the first training workshop and participated in the deployment phase. Thirty-six attended the second evaluation workshop and contributed to the qualitative data collection.

Ethical Considerations

This study was approved by the research ethics committee at the Open University of Catalonia, with the Kenyan Medical Research Institute (KEMRI) protocol number KEMRI/SERU/CTMDR/122/4959, Open University of Catalonia reference number CE23-RC21, and the Kenya National Commission for Science, Technology and Innovation license number NACOSTI/P/24/35733. Written informed consent was obtained from all participating FHWs. The FHWs were informed that participation was voluntary and that they could withdraw from the study at any time without any consequences. All FHWs participating in the second workshop provided written consent for audio recording and the use of data for analysis and publication. In addition, FHWs obtained written informed consent from patients before using the WHO Skin NTDs app during clinical consultations during the 6-month study period. Patients were informed that participation was voluntary and that their standard of care would not be affected by refusal to participate. All data were deidentified prior to analysis to ensure participant confidentiality. No personally identifiable information is reported in this study. FHWs were not financially compensated for participating in this study, although travel and subsistence costs (11,000 Kenyan Shillings [1 Kenyan Shilling=US $0.0077 as of June 26, 2026] per FHW, based on government rates) were provided to support workshop attendance. Mobile data bundles (5000 Kenyan Shillings per FHW) were also provided to support app use within clinical workflow.

Data Collection

Data were collected through 15 SSIs (each lasting 30 - 45 minutes) and 4 FGDs (1 - 1.5 hours, 8 - 10 FHW in each). The SSIs enabled a focused exploration of individual FHW perspectives, while the FGDs offered opportunity to encourage debate and consensus building, highlighting areas of agreement and contrast in perspectives on the WHO Skin NTDs app use within diverse clinical workflows. The interviewers (EQ, RN, CC, AF, and DA) used semistructured topic guides to guide FGDs and SSIs, which were reviewed by all authors prior to the workshop (Multimedia Appendices 1 and 2). All FHWs participated in FGDs, and 15 FHWs were purposively selected to participate in SSIs to ensure a balanced mix across counties.

Both SSIs and FGDs were used to enable complementary insights. SSIs allowed for in-depth exploration of personal experiences, with participants asked to recount specific instances of app use in situ. This yielded rich, contextualized accounts of how the app was integrated into individual diagnostic decision-making, including nuanced challenges and variations in use. FGDs, by contrast, enabled group-level reflection, providing a space to examine shared practices, compare perspectives, and identify points of consensus or disagreement regarding the app's usefulness role in workflow and broader systemic challenges. This combination of methods was intended to enhance both the depth and breadth of understanding regarding app integration in routine care.

All interviews and FGDs were conducted in English (an official language in Kenya). Each FGD was cofacilitated by a local research scientist from KEMRI (RN and DA) and a researcher from the Global North (EEVQ and CC). All FGDs and SSIs were recorded on portable Dictaphones and stored on a secured university-licensed cloud service (OneDrive). Field notes were taken by all researchers during and after both FGDs and SSIs and were used to support the thematic analysis.

Data Analysis

Audio recordings were transcribed verbatim by the corresponding author (EEVQ), supporting deep familiarization with the data. Transcripts were analyzed thematically using NVivo (version 14; QSR International) software. Given that this is the first qualitative assessment of the AI component of the WHO Skin NTDs app usability and acceptability, a bottom-up inductive coding approach was selected to support the iterative construction of themes grounded in participant narratives. This exploratory approach aimed to capture nuanced insights that may inform the application of appropriate theoretical frameworks in future studies, as the evidence base on the app's integration expands. EEVQ conducted initial coding of all transcripts. A subset of transcripts was independently double coded by AB and KO. Coding discrepancies were resolved through team discussion, and the codebook was iteratively refined. The full dataset was analyzed using reflexive thematic analysis, with codes inductively organized into 4 higher-order themes within the research team [33].

The derived themes from the analysis made reference to behavioral patterns and interdependencies that lent themselves to communication through systems-thinking visualizations. With this in mind, we constructed 2 causal loop diagrams (CLDs) grounded in FHW narratives to provide a structured way to visualize how system-level dermatological workflow factors interact in practice, both before and after app integration [34]. The intention here is not to generate formal theory but rather to offer descriptive representations of diagnostic and referral pathways and contextual factors shaping their progression as revealed through the FHW accounts. In this respect, the CLDs function as an analytic and communication tool to scaffold the interpretation and presentation of the overall thematic narrative.

XSL•FO
RenderX

Results

Overview

The 36 participants who attended the second workshop in November 2024 comprised FHWs from 5 counties in Kenya: Nakuru (n=11), a fairly urbanized county situated in the Rift Valley; Baringo (n=8), also situated in the Rift Valley, characterized by semiurban and rural areas; Kajiado (n=6), bordering Tanzania in southern Kenya with a seminomadic population; West Pokot (n=8), a remote rural county bordering Uganda, predominantly inhabited by nomadic pastoralist communities; and Kwale (n=3), a largely rural, coastal county in southeastern Kenya [35-39]. Participant characteristics are described in Table 1.

XSL•FO
RenderX

Table . Participant details: ID, county, job role, and level of health care facility (specific county names have been removed to protect participant anonymity).

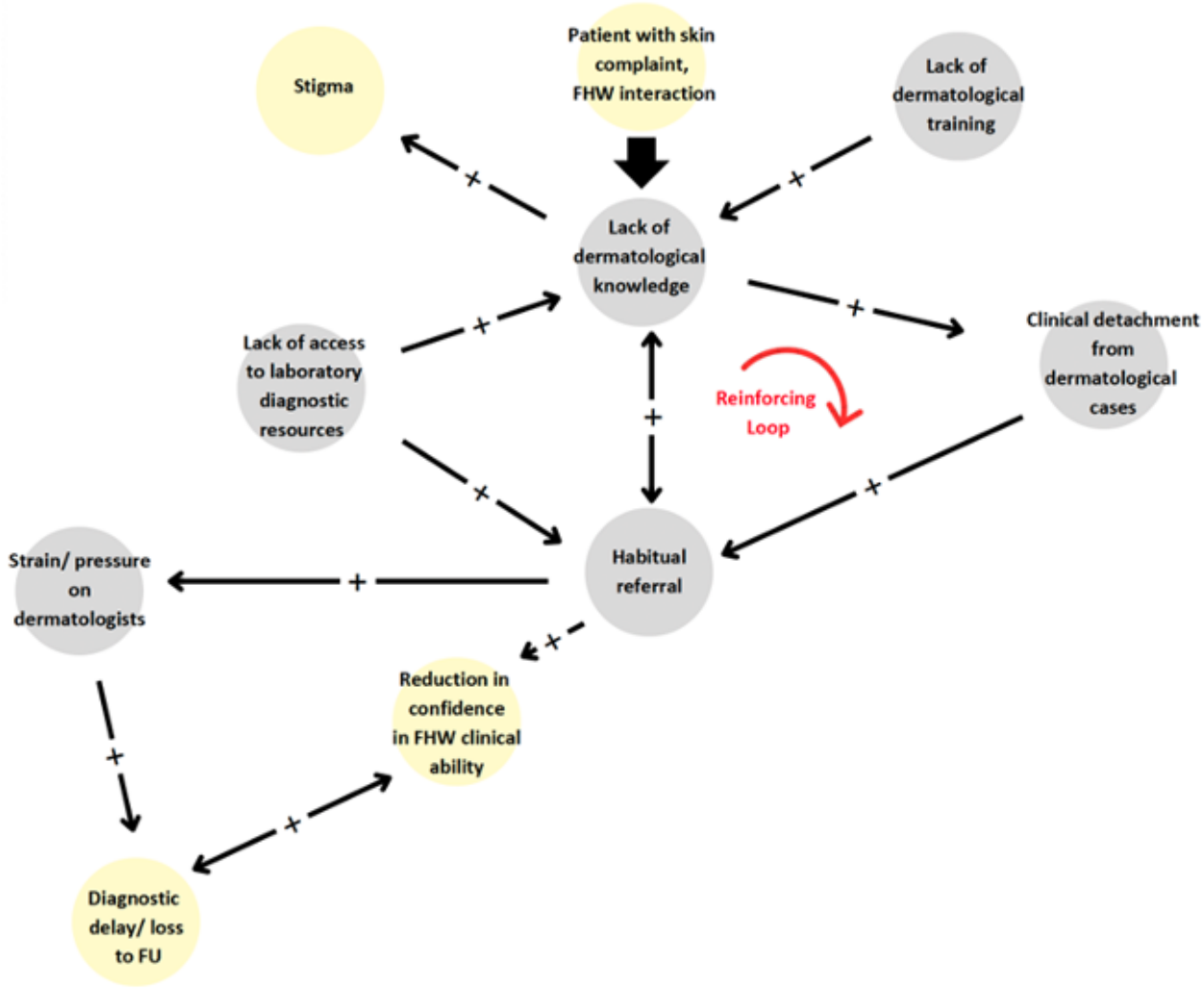
Participant ID	Job role	Level of health care facility
1	Public health officer	Level 5
2	Nurse manager	Coordinator
3	Clinical officer	Level 4
4	Clinician	Level 4
5	Clinician	Level 5
6	Clinician	Level 5
7	Clinician	Level 5
8	Clinician	Level 5
9	Nurse	Level 5
10	Clinician	Level 4
11	Clinical officer	Level 4
12	Clinician	Level 4
13	Clinical officer	Level 5
14	Nurse	Level 5
15	Community health nurse	Level 2
16	Nurse	Level 2
17	Public health officer	Level 3
18	Clinician	Level 5
19	Clinician	Level 2
20	Clinician	Level 2
21	Clinician	Level 4
22	Clinical officer	Level 3
23	Clinical officer	Level 3
24	Nurse	Level 2
25	Clinical officer	Level 5
26	Clinical officer	Level 5
27	Doctor	Level 4
28	Clinical officer	Level 4
29	Nurse	Level 2
30	Clinical officer	Level 5
31	Clinical officer	Level 5
32	Clinician	Level 5
33	Clinician	Level 2
34	Pharmacy technician	Level 3
35	Clinical officer	Level 4
36	Clinician	Coordinator

The thematic analysis highlighted shifts in dermatological care delivery, representing the complex dynamics before and after app integration, visualized through 2 CLDs (Figures 3 and 4). The CLDs are grounded in FHW narratives, with illustrative quotes supporting each node (Multimedia Appendix 3). Yellow nodes represent patient-related factors, gray nodes denote FHW-related elements, green arrows indicate changes postapp implementation, and reinforcing loops perpetuate challenges or amplify improvements. Connections depict the relationships between these factors, as articulated by FHWs, indicating whether they have a positive or negative influence. The diagrams synthesize FHW perspectives across multiple facility levels

XSL•FO
RenderX

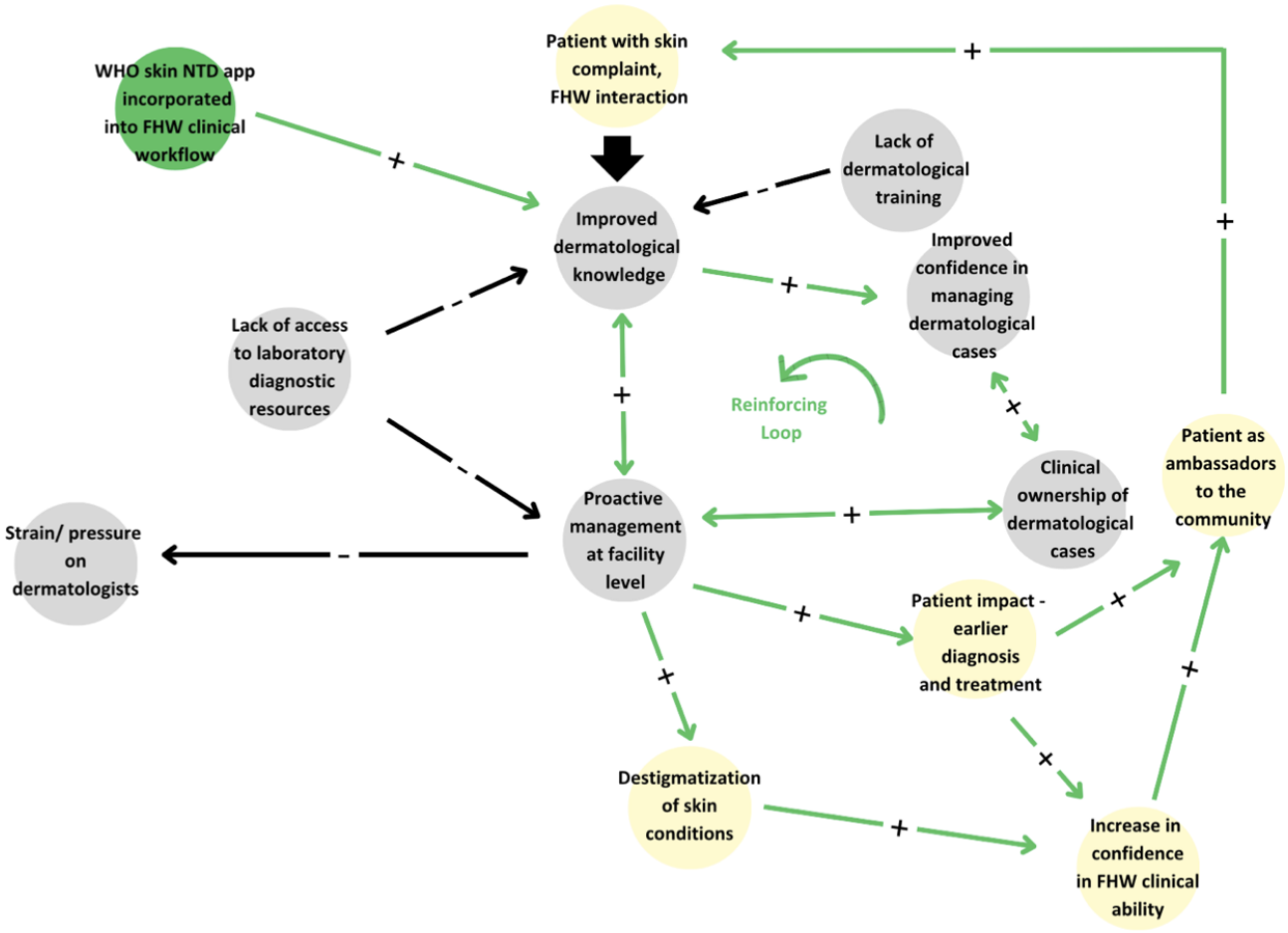
rather than disaggregated by facility levels. However, we acknowledge that the key elements of aggregated narratives will have a differential impact across the different facility.

Figure 3. Causal loop diagram: preapp context, reported systemic challenges in dermatological care. FHW: frontline health worker; FU: follow-up.



XSL•FO
RenderX

Figure 4. Causal loop diagram: postapp integration and shifts in clinical workflow. FHW: frontline health worker; NTD: Neglected Tropical Diseases; WHO: World Health Organization.



Before app integration, FHWs reported patterns of habitual referral of dermatological cases, where nonspecialist dermatological knowledge often limited immediate diagnostic assessment capability, combined with limited access to diagnostic resources and specialist training that would enable necessary upskilling (Figure 3). This resulted in FHWs expressing a sense of clinical detachment, reinforcing a cycle where FHWs avoided managing dermatological cases at their local facility; instead, depending upon habitual referral to higher-level facilities and specialist services. FHWs expressed that this referral dependency strained dermatology services, which they associated with diagnostic delay and subsequent patient attrition (loss to follow-up). The diagram further highlights a consequent reinforcing feedback loop—the more FHWs referred cases, the less experience they gained in managing them, further diminishing confidence and knowledge in dermatological care.

After app integration, FHWs described a shift toward more immediate dermatological case management at their local-level facility, interrupting the cycle of habitual referral to higher-level facilities and specialist services (Figure 4). The app was perceived as enhancing dermatological knowledge acquisition and development, reinforcing a sense of clinical confidence and greater feeling of ownership of dermatological cases. Local facility case management was perceived to facilitate earlier diagnosis and treatment, which FHWs linked to improved patient trust in their clinical abilities. Additionally, FHWs highlighted the app's role in destigmatization of skin conditions, as greater diagnostic confidence reduced the need for voyeurism of collaborative deliberation over patients. FHWs spoke of patients becoming "ambassadors" within communities, actively promoting care-seeking behavior based on their positive experiences with the app. The diagram further highlights a consequent reinforcing feedback loop—as FHWs gained dermatological knowledge, their confidence in managing cases increased, fostering greater clinical ownership. This, in turn, reduced reliance on specialist referrals as FHWs managed more cases at the facility level, with potential implications for strengthened primary dermatological care.

These system-level causal-loop representations (Figures 3 and 4) offer high-level synthesis of FHW-identified factors influencing the transformation of care and referral pathways. Building on this, we present a deeper dive into the phenomenological accounts of FHW-reported app use and the perceived transformation of practice experienced arising from its adoption.

From Habitual Referral to Local Facility–Based Ownership of Dermatological Cases

The integration of the WHO Skin NTDs app into diagnostic workflows was seen to facilitate a shift from habitual referral to local facility management. FHWs described this transition as improving patient retention and fostering wider community engagement (Figures 3 and 4).

XSL•FO
RenderX

Pre - app: Clinical Detachment and a Clinical Model of Habitual Onward Referral

Before the introduction of the app, FHWs described a sense of clinical detachment from dermatological patients, often perceiving these cases as outside their professional remit (Figure 3). This detachment was largely attributed to "very little knowledge about skin or maybe dermatology" [Participant 23, clinical officer, level 3].

This lack of dermatological expertise seemed to contribute to a "clearing and referring" culture in which dermatological cases were routinely categorized as "clearly not for me" [Participant 3]. Consequently, patients were often referred to higher-level facilities and specialist services without comprehensive clinical assessment (Figure 3). High patient volumes further exacerbated these practices whereby referral, in lieu of clinical knowledge, contributed to a more efficient throughput to help manage overall patient load: "sometimes 120, 130, 150 [patients a day]...you have a lot of patients to attend, it's not only the skin ones" [Participant 13, clinical officer, level 5]. Collectively, FHWs spoke of viewing dermatological conditions as inherently complex and firmly within the domain of specialist dermatologists, reinforcing a reliance on specialist referral as a default rather than a considered clinical choice:

> *The patient with skin NTDs...we didn't want much to know much about it since we didn't know how to make the diagnosis...we didn't want to learn...It was difficult to make diagnosis and in terms of management, it was still difficult for us to manage those conditions...we were not in a position to make that diagnosis...These were conditions that were supposed to be managed by specialist. Specialised people, not by us. Yeah, ours was clearing [patients] and referring...and we wouldn't have followed up.* [Participant 3, clinical officer, level 4]

Notably, some FHWs framed habitual referrals not as an inability to manage but as a need for diagnostic "*confirmation*" [Participant 23], highlighting the potential role of confidence in enabling local facility–based management.

Post - app: The App's Perceived Role in Shifting Toward Clinical Ownership and Local Facility Management

FHWs perceived that the WHO Skin NTDs app disrupted habitual referral patterns by providing clinical decision support, reframing dermatological conditions as more manageable within primary care. Where appropriate, this shift seemed to encourage more immediate and confident management at the local-level facility (Figure 4):

> *We used to refer all patients with skin conditions. But now with the app, we try to manage the patients without sending all of them to the dermatologist.* [Participant 32, clinician, level 5]

Many of the health workers described a developing sense of clinical ownership, actively positioning themselves as local dermatological experts with colleagues now seeking their consultation for dermatological cases:

> *From the time we started using this app, it has, really had a positive impact. Because in our health facility, if anyone comes across a patient with skin lesion, he or she will first send the patient to me. Even if I am off, they will call me..."come and see there is a patient with a skin lesion, come and use the app, we see what is the problem.*" [Participant 21, clinician, level 4]

Impact of Local Facility Management on Patient Retention and Engagement in the Full Illness Trajectory

Before the introduction of the app, FHWs described how their habitual referral patterns would "delay the management or the treatment of this patient." The reasons for this are that the specialist dermatology clinics to which the patients were being referred were infrequent, happening maybe "once in a month or once in a week" [Participant 13, clinical officer, level 5]. As such, patients would be required to wait until the clinics were available and take further time out to travel. In response, many FHWs (notably from level 2 and 3 facilities) described how they informally adopted WhatsApp to enable a limited form of remote dermatological consultations as a workaround. These kinds of practices point to the broader informal integration of mobile phones into everyday health care delivery—an improvised form of telehealth that enables more immediate access to specialist input without requiring patients to travel—in principle, easing the burden of referral. However, FHWs also highlighted the fragility of this workaround, particularly when "sometimes they [the dermatologist] don't even reply" [Participant 19]. Reliance on asynchronous responses from often-unavailable specialists was therefore seen, at times, as a potential source of delay in patient management:

> *OK, I think it [WHO skin NTD app] also saves time since like people who are in a rural facility, when they come across a skin disease and they don't have knowledge about it, they call a dermatologist. And the dermatologist says "take a photo and send it through Whatsapp." So it takes time to get the results. But with the app, it helps diagnosis and you get the patient directly into the system.* [Participant 34, pharmacy technician, level 3]

FHWs widely reported that the app's immediacy of decision support was seen as important in enabling earlier treatment at the local-level facility, reducing the reliance on informal WhatsApp workarounds (Figure 4). Significantly, the subsequent reduction in assessment and treatment delays was felt to be further beneficial in reducing the potential health complications associated with treatment delay:

> *It [WHO skin NTD app] is a key app for any primary healthcare worker, because one, it's going to help in giving the treatment to the patients...it's going to stop delay in giving the patient treatment. And that will reduce the complication associated with each particular condition depending on each condition you are dealing with. "Cause as we delay, there are problems of that condition.*" [Participant 18, clinician, level 5]

Managing dermatological patients at the local-level facility was widely reported as "cost-effective for our clients" [Participant

XSL•FO
RenderX

2], reducing both direct costs (eg, transport and consultation fees) and indirect costs (eg, loss of earnings due to travel and appointment wait times). FHWs linked these cost savings to improved patient retention, with fewer patients "defaulting" from care [Participant 2]:

> *The impact to my patient...it has helped so much because it will reduce the cost because generally rather than the patient having to go and look for a dermatologist in Nairobi or in another place, you are able to make that diagnosis [at the facility level] and help that patient.* [Participant 13, clinical officer, level 5]

Some FHWs also reported a sense that patient confidence in local facility–based care improved with app use. They described that for some patients, seeing the app used in their consultation through digital imaging of their skin lesion appeared to foster "a lot of hopes—that maybe here the treatment will come..." [Participant 9, nurse, level 5]. However, there were also reports that some patients remained skeptical of the app, particularly those with prior experiences of misdiagnosis. This skepticism was often expressed as hesitancy around image capture. As one FHW described: "One [patient with a skin condition] was saying 'I have been facing several dermatologists without any improvement, so why would I assist you with consent to take a photo from part of my body?'" [Participant 25, clinical officer, level 5]. Similarly, others described that prolonged illness trajectory appeared to shape patient perceptions of the app's value "someone has maybe a lesion for a very long time, then...'here you are, you want to take a photo of that lesion. So, what difference will it make?'" [Participant 22, clinical officer, level 3].

Impact of More Immediate Local Facility Management on Community Engagement

Multiple FHWs described an unexpected outcome of earlier facility-level management. Having experienced the app's role in clinical decision support, FHWs reported that patients then encouraged others to seek similar care, emerging as informal community "ambassadors" [Participant 25], seemingly facilitating peer-driven health-seeking behavior (Figure 4):

> *And also to the community, once you use the app in diagnosing one patient, he or she will go on, referring another one, "go to the facility, the nurse will take a photo.*" [Participant 29, nurse, level 2]

FHW associated this community advocacy with "hope," driven by the belief "that there is something that will come out of that [the app]" [Participant 24, nurse, level 2].

From Limited Dermatological Knowledge to Perceptions of Increased Knowledge and Confidence in Case Management

Findings suggest that the shift from habitual referral to local facility case management was driven by FHWs' perception of improved dermatological knowledge, in turn enhancing clinical confidence (Figures 3 and 4). The majority of FHWs suggested that the app did not replace clinical judgment but seemed to reinforce it—functioning as an assistive tool that encouraged active engagement with differential dermatological diagnoses.

Knowledge Acquisition and Development Through Diagnostic Comparison and Self-Directed Education

The app's provision of differential diagnoses appeared to encourage curiosity, with many FHWs describing how this broadened their clinical awareness of skin conditions (Figure 4). Encountering unfamiliar or unexpected diagnostic suggestions seemed to prompt more active clinical reasoning, including questioning, comparison, and reflection on local disease patterns. As one FHW described:

> *...it has opened my mind more, giving me more learning on a lot of exposure to many conditions...you take a photo, it says Yaws, and Yaws is not common in that region. So you get suspicious...why? So you have to ask several questions....* [Participant 13, clinical officer, level 5]

FHWs also described engaging in self-directed learning in response to these prompts, seeking additional information through books and online resources when presented with unfamiliar conditions. One participant noted that the app "challenged [their] knowledge" and motivated them "to go and read...to understand more about the disease" [Participant 33, clinician, level 2].

Beyond knowledge acquisition, FHWs noted a deliberate process of comparing the app's differential diagnosis— referred to by some as "impressions" [Participant 3]—with their own clinical assessments. This comparative process involved integrating patient history, physical examination, and app output to support diagnostic reasoning: "...once I ask the app what is their opinion, it's it will help me see if my clinical judgement is next to what the app is actually saying" [Participant 18, clinician, level 5]. FHWs emphasized how this process enabled them to refine, rather than replace, their clinical judgment. The app was described as providing "room" to rule out or reconsider working diagnoses based on likelihood and clinical context [Participant 21].

While the app was widely perceived as supportive, most FHWs recognized that its outputs were not infallible. As one participant noted, it should be used as "support...and then you add to the knowledge you have" [Participant 19]. However, engagement with the app varied, with a minority of FHWs expressing high levels of reliance on its outputs, at times treating them as definitive and "100% this is the skin condition" [Participant 1, public health officer, level 5]. In cases of diagnostic discrepancy, FHWs reported prioritizing peer consultation over defaulting to the app's output, emphasizing the continued importance of clinical judgment and collaborative decision-making. As one FHW reflected: "You have to still have your own clinical judgment...If you don't agree [with the app output]...you must consult" [Participant 6].

Confidence Gains as an Outcome of Incremental Knowledge Acquisition and Development

FHWs reported their "confidence building up" in dermatological case management as they perceived development in their dermatological knowledge through continued app use [Participant 19, clinician, level 2].

XSL•FO
RenderX

> *So it [the app] has impacted me positively because now I'm very confident to tackle any skin condition rather than before. Before seeing that patient, you can see the patient from far and then you are thinking me how am I going to manage this patient now? Where can I start, from what you see? But when this app came, you have that confidence to see that patient and come up with a diagnosis.* [Participant 11, clinical officer, level 4]

Importantly, this confidence was not derived from blind reliance on the app's output but from the alignment between the app's suggestions and FHWs' clinical judgment. When the app's diagnostic suggestions corroborated their own clinical assessments, the majority of FHWs reported an increased "morale" and motivation to engage more deeply with patient management:

> *The positive part is that when you come into contact with a client, and then you use the app, and the app says the exact diagnosis you have been thinking of clinically and then you feel comfortable in managing the patient. And then that morale of doing more for the patient comes in.* [Participant 12, clinician, level 4]

This confidence was further reinforced by the app's availability as a "backup" [Participant 23] tool, particularly in settings where dermatological expertise and resources were scarce:

> *If I have something that I'm not very sure, I'm not worried or I'm not anxious because I know I have the app, which is a tool that can assist me... I'm confident because it's like knowing there's somebody here who can, you can always consult and refer to...I'm sure that there's somebody who can help me. And that's somebody is the app.* [Participant 23, clinical officer, level 3 (Figure 3).]

Reduced Dependency on Public Deliberation Increases Patient Trust and Alleviates Stigmatization of Patient as a Clinical Spectacle

Before app integration, FHWs reported that one consequence of limited dermatological knowledge was the felt need to call multiple colleagues, "maybe even 2-3 people to come to the same room and discuss the patient" [Participant 13]. While perhaps intended as a means of clinical support, many FHWs reflected on how this public deliberation created awkward and uncomfortable situations for patients, reinforcing stigma around dermatological conditions (Figure 3):

> *Before, every time they [the patient] comes, we call each other "come confirm if this is leishmaniasis, come confirm if this is leish"...By the time you are calling more people to come and see, you see, the patient feels stigmatised.* [Participant 2, nurse manager, coordinator]

Beyond stigma, FHWs observed that patients often perceived these public deliberations—conducted in front of the patient (and distinct from standard peer consultation outside the consultation space)—as a sign of clinical uncertainty, reducing trust in their ability to provide appropriate care. Participant 2 elaborated:

> *...if a patient comes in and you are not sure of the diagnosis, the patient may feel that this clinician is not confident. Why has the clinician have to call others to come and verify the diagnosis, it means that the clinician is not confident enough...even the clinician doesn't know what I'm suffering from. They will not even feel confident with the treatment that you write for them.*

However, with the app as "a partner that can help you make a diagnosis," FHWs described an increased confidence to make clinical decisions more independently, "instead of calling a colleague, several clinicians to come" [Participant 15]. FHWs reported that this shift contributed to a reduction in stigma, as patients were no longer subjected to group deliberations during diagnosis:

> *Before we get hold of this app, when you see a condition where you suspect this is an NTD, you have to call your colleague, maybe even 2 - 3 people to come on to come to the same room and discuss the patient. So, for now at least, there is that reduction in the stigma because you are only with a patient. So, there is a lot of confidentiality.* [Participant 13, clinical officer, level 5]

Opportunities to Extend Knowledge Development Beyond Clinical Decision Support to Full Illness Trajectory

While the app was reported to support dermatological learning and diagnostic reasoning, many FHWs reported that it "is limiting in terms of management… it does not give us information on treatment." This limitation added complexity to clinical workflows and often required FHWs to "use the app alongside Google...to find how to treat" [Participant 8, clinician, level 5]. Although the app does contain management guidance in a separate section, FHWs envisioned embedding treatment and management guidance directly after diagnostic suggestion to help reinforce incremental learning across the full illness trajectory. Participant 8 went on to explain: "...it would be better if, for the purpose of frontline workers who really don't have a lot of knowledge in dermatology, every information comes to the app so that once you have said that you have uploaded the photo, you get a diagnosis, get treatment."

The WHO Skin NTDs App Facilitating Patient Involvement in the Diagnostic Process

Findings show that the integration of the WHO Skin NTDs app into clinical workflows appeared to shift diagnosis from a clinician-driven process to a more participatory model, actively involving patients in their care. Initial narratives from FHWs indicate that the app seemed to function as both a clinical decision support tool and a patient engagement mechanism.

Strengthening Patient Involvement Through Diagnostic Transparency and Participation

FHWs described how, for some patients, seeing the app's interface and diagnostic suggestion appeared to reinforce engagement in their own care:

XSL•FO
RenderX

> *This app, when I used it on my patients, they were a bit happy because when I showed them the diagnosis [on the app], we could share together. So, most of them were actually happy because of the technology.* [Participant 5, clinician, level 5]

The app also appeared to promote more interactive consultations, with some patients becoming increasingly engaged in discussions about their skin conditions and expressing curiosity about the diagnostic process:

> *This time around when you are using the app you explain to the patient more and also the patient was curious to know more about, about the app and sometimes also they were curious about the skin diseases. Maybe they could ask you, "have you ever seen maybe a skin disease like this with anyone else, or is it? Or is it me alone? So, there was a lot of discussion between the patient and me now. They'll tell you..."this time around the way you have explained to me, I think I'll get well."* [Participant 22, clinical officer, level 3]

Beyond sharing diagnostic suggestions, FHWs felt facilitated by the app in engaging patients in real-time diagnostics, making consultations more collaborative. In turn, FHWs attributed this increased participation with strengthened patient confidence:

> *Participant 3 [clinical officer, level 4]: ...the best thing is to involve your patient in the [diagnostic] process.*
>
> *Interviewer: And the app allows you to do that?*
>
> *Participant 3: Of course, whenever you're using it, it is our phone. You know, the way we sit together, the way we sit with our patients in our facility. My patient will come. OK we sit at an angle we normally use a system [computer system], this patient wherever he's sitting he will have to see the system...It's almost next to each other...So, the patient can see everything you're doing...so at least the patient will follow the process, and by the end of the day, he'll see the diagnosis that you get from the app. So, it is still build confidence in this patient.*

Technical Limitations and Their Impact on Patient Involvement in the Diagnostic Process

FHWs described varied experiences with internet connectivity, which shaped their ability to use the app in real time and, in turn, influenced the participatory nature of consultations (the AI-embedded component of the app currently functions online only). Some reported no issues "in terms of connectivity" [Participant 3], while others anticipated connectivity issues but developed workarounds such as identifying areas with stronger signal:

> *My first impression was since the app is using data bundles, and connectivity in my area is a bit poor. So, I thought it would give me a really hard time uploading the photos, which wasn't the case so much because I designated an area that was a bit faster.* [Participant 24, nurse, level 2]

However, in more rural settings, FHWs described network dependence as a major limitation, preventing the app from being used consistently: "I think the challenging part of this app is the problems when there is no network...so you cannot use the app... in rural facility...you cannot assist the patient using the same app" [Participant 12, clinician, level 4].

Beyond connectivity, specific technical failures disrupted patient involvement more directly. One recurrent issue was the unsettling error message "not human skin" [Participant 36], which appeared during attempts to capture images in live consultations. Despite multiple adjustments, this issue often persisted: "The first time I got the response of not a human skin. I kept on trying several times, like increasing the light, maybe the light was low, changing postures...I tried several times, but still it failed" [Participant 32, clinician, level 5]. Such breakdowns seemed to introduce moments of uncertainty during consultations, undermining both interactional flow and patient confidence. As one FHW reflected:

> *The problem of course is when the patient looks at you it also gives stigma to those patients. Including myself...you are wondering "what I am doing, am I doing the right thing." It's the patient and me, and the patient trusts me, and I look like this thing is not working then the patient also thinks 'he does not know what he is doing, using this app.* [Participant 1, public health officer, level 5]

In some accounts, some FHWs attributed this error message to the technology's failure to recognize darker skin tones:

> *Another disadvantage as part of it is when you take a photo, especially in our place, you see the African skin, the app will tell you that's not a human skin.* [Participant 12, clinician, level 4]

Interconnectedness: Expanding the Role of the WHO Skin NTDs App

FHWs envisioned the WHO Skin NTDs app evolving beyond its current role as an assistive tool for capacity building within the individual clinical encounter. Their recommendations highlighted its potential to function as part of an interconnected health system, linking clinical dermatological care with broader health system infrastructure.

Connecting Clinical Services and Integration Into Local Digital Infrastructure

The majority of FHWs emphasized the importance of embedding the app within existing digital health systems, particularly Kenya's electronic medical record (EMR), to streamline clinical workflows and facilitate referrals. Integration was viewed as important for both sustainability and routine care.

> *Using the computer system [EMR], it could improve the referral system for so many clients from skin NTDs to the facilities. So, there maybe should be a connection between the app in their phone and the gadget in the facility.* [Participant 17, public health officer, level 3]
>
> *The Kenya Government is encouraging the digitalisation of the health system...So, any system*

XSL•FO
RenderX

> *which is coming, any digital app which is coming, it has to be interoperable with those systems of government [Kenyan EMR]. Those ones, it's just helping in terms of data management, not in terms of diagnosis.* [Participant 10, clinician, level 4]

These insights highlight a potential opportunity for additional interoperability, with the app positioned not as a stand-alone tool but as part of an integrated digital ecosystem supporting sustainable service delivery.

Connecting Clinical Services to Broader Public Health

FHWs commented how greater interoperability could potentially facilitate closer alignment with national health systems in ways that could further benefit broader public health. Beyond point-of-care decision support, the app was imagined as a tool for real-time disease surveillance, resource mobilization, and data-informed planning.

> *It [the app] will also help in resource mobilisation. Because at the community level, at the lower level, level 1, level 2, level 3, you will be able to know those diagnosis using the app and the clinician judgment. So through the county, our leadership will be able to know 'this area has this number of cutaneous [leishmaniasis], and then they will pull the resources to that area so that they can deal with that condition.* [Participant 15, community health nurse, level 2]

While not framed as formal design principles, these insights point to the importance of designing AI-embedded mHealth tools that are responsive to existing digital infrastructures and aligned with routine clinical workflows, rather than functioning as stand-alone applications. Further ethnographic work will enable more fine-grained articulation of how such integration is enacted in practice.

Discussion

Principal Findings

This study presents a comprehensive understanding of FHWs' experience and perception of the WHO Skin NTDs app in Kenya. The majority of FHWs described the app as facilitating the following: (1) validating their clinical judgment, which reduced dependency on referral pathways and improved patient retention and community engagement; (2) incremental knowledge acquisition and growing confidence in managing dermatological conditions independently; and (3) enhanced patient engagement and perceived trust in the diagnostic process.

These findings contribute to a growing body of evidence on how mHealth tools are being integrated into resource-limited frontline health systems and the ethical and pedagogical challenges this raises, particularly given the app's intended function as an educational and capacity-building tool [21,30,40]. By contrasting FHWs perceptions of pre- and postapp clinical workflows (Figures 3 and 4) and integrating FHW narratives (Multimedia Appendix 3) with thematic analysis, we identify potential system-level challenges associated with dermatological care and specific mechanisms of change facilitated by the app. These insights inform both app refinement and broader implementation strategies to maximize pedagogical value and ensure ethical integration of mHealth tools in frontline health care systems. Taken together, they suggest that the app is having a positive impact on dermatological care within the studied context and is, in many cases, being used in alignment with its educational remit. However, there remains a need to examine edge use cases and potential risks—particularly where usage may extend beyond the app's original design parameters [23].

Clinical Improvisation and AI-Supported Reasoning

This study highlights the central role of clinical improvisation—real-time adaptation of clinical workflows and care strategies—as a core component of dermatological care in resource-limited settings [41,42]. In sub-Saharan Africa, FHWs routinely navigate complex social, political, infrastructural, and economic constraints, necessitating clinical improvisation [41]. FHWs identified multiple systemic barriers impacting dermatological care, including limited diagnostic resources, inadequate training, scarce dermatology expertise, and high patient volumes (Figure 3). In response, they described improvisational practices such as habitual referral, informal WhatsApp consultations, and public deliberation of clinical cases (Figure 3). Crucially, these practices are not merely indicative of knowledge gaps or health system limitations; rather, they reflect locally adapted strategies developed under constraints, described elsewhere as the "craft" of care under constraint [41].

However, while such strategies may enable care delivery where formal systems fall short, our findings suggest that some may also inadvertently limit opportunities for building and maintaining dermatological knowledge and case management skills. In this study, FHWs linked habitual referral and reliance on external consultation to diagnostic delays, patient attrition, and stigma–patterns observed across similar African contexts [43,44]. These findings suggest that some adaptive strategies may unintentionally reinforce ongoing dependency on specialist services by reducing opportunities for local dermatological knowledge development and case management experience.

Our findings suggest that integrating the WHO Skin NTDs app into clinical workflows may help reshape improvisational practices by facilitating more reflective diagnostic reasoning and greater confidence in local facility case management. While some of these benefits may also arise through other forms of clinical decision support, the AI-enabled nature of the app allows FHWs to compare image-based differential diagnoses with their own clinical assessments in real time.

In the broader literature, AI-embedded tools are ideally situated within a model of human-AI symbiosis: a collaborative relationship where the clinician's reasoning complements the algorithm's suggestions [45,46]. Most FHWs in our study described using the app in these terms—referring to it as a "partner" [Participant 15], akin to a second reader that supports and moderates, rather than supplants, their clinical judgment. FHWs reflected that this perceived partnership enabled more confident and independent patient management at the local-level facility, reducing reliance on improvisational workarounds such as habitual referral. This could be interpreted through the lens

XSL•FO
RenderX

of self-efficacy theory, which posits that an individual's engagement in a behavior is related to their belief in their ability to perform such tasks [47]. In this context, the app appears to function as a mechanism for enhancing FHWs' perceived self-efficacy in dermatological care, supporting their ability to act on clinical decisions locally. This reflects wider literature on how AI-embedded clinical decision support systems can be, and should be, designed to augment rather than replace human reasoning [46,48].

However, our findings also suggest that when the app's output did not align with the FHW's clinical judgment, FHWs often reverted to referral to specialist dermatologists. While this can be understood as an appropriate response to clinical uncertainty, it also highlights an important opportunity for future design refinement. AI-embedded decision support tools may benefit from incorporating structured support mechanisms to guide arbitration in such cases, ensuring that the burden of resolving diagnostic discrepancies does not fall solely on the FHW. Such mechanisms could further enhance the app's pedagogical value by facilitating reflection and learning through structured resolution pathways, rather than relying on referral as a default form of risk mitigation.

Ethical Tensions: Maintaining Clinical Autonomy in AI-Supported Care

Our study raises a potential ethical concern with respect to how some FHWs may inadvertently use the app beyond its intended remit as an educational clinical decision support tool, potentially engaging in care beyond their clinical competence [23]. One FHW, for example, explicitly described deferring entirely to the app's diagnostic suggestions. This pattern may reflect elements of automation bias, whereby users place undue trust in algorithmic outputs over their own clinical judgment [49]. Given the limited dermatological knowledge reported by FHWs, such potential overreliance may raise medicolegal and ethical concerns [23].

Some scholars caution that as AI systems are perceived to outperform human judgment, they may begin to displace human epistemic authority [50]. This risk is echoed in broader literature, which highlights that AI tools may contribute to clinician deskilling, diminished critical reasoning, and reduced capacity to navigate complex or atypical cases through overreliance [48-53]. While often discussed as a future-facing concern, our findings suggest that elements of this dynamic may already be emerging in contexts where dermatological expertise is limited [50]. In such cases, for some FHWs, the app's output may be seen not as suggestive but as directive—an inversion of its intended role.

Meaningful human-AI collaboration depends on users having sufficient knowledge to question and validate algorithmic suggestions—a process that relies on training, confidence, and the time and institutional space to exercise clinical judgment [46]. Many FHWs in our study described using the app in this way, comparing AI-generated differentials with their own clinical assessments and reporting perceived knowledge gains through continued app use. Such experiences may, over time, strengthen their capacity to engage critically with AI outputs. Conversely, where foundational dermatological knowledge is limited, the same system may increase the risk of inappropriate reliance on algorithmic suggestions.

This presents a paradox in the positioning of the WHO Skin NTD app. The tool is explicitly designed to support capacity building among FHWs in response to limited dermatological expertise [21,22,30]. Yet, for the app to be used responsibly, FHWs must possess sufficient baseline knowledge to critically interrogate and contextualize its outputs. As our results suggest, this prerequisite cannot be assumed and is likely to vary significantly across different health system contexts, facility levels, and individual FHWs. To help mitigate this risk, explainability of AI outputs is critical: presenting outputs in transparent, interpretable ways can better support FHWs in understanding and validating diagnostic suggestions [54]. Other human-centered design principles, such as offering contextual guidance and supporting stepwise reasoning, could further assist FHWs in maintaining epistemic authority while engaging with the app. Features already present in the app, such as graded confidence levels associated with differentials, may help surface diagnostic uncertainty in ways that encourage critical engagement [54-56].

In addition to this epistemic tension, it is critical to recognize that the dynamics of AI use in clinical practice are not shaped by individual knowledge alone. Broader contextual factors, including patient load, time constraints, and infrastructural limitations, also influence how AI outputs are interpreted and acted upon [45]. This reinforces the importance of designing and adapting AI tools and implementation strategies in ways that explicitly protect and promote clinical autonomy, both in individual decision-making and across broader health care delivery [45,57,58].

Three areas of further work are needed. First, direct observation of app use in situ will be critical to understanding how it is conceptualized in practice, how broader contextual factors shape its use and whether it is functioning as intended (as a clinical decision support tool used for educational purposes within the clinical encounter), or whether some usage patterns risk reclassification. This will also be critical for elucidating how patients perceive the app's role in the collaborative context of FHW-patient interaction. Second, targeted design adaptations, such as embedded disclaimers or periodic training modules, should be explored to reinforce the app's role as a nondirective support tool and bolster FHWs' dermatological knowledge to safeguard clinical autonomy. Third, development of a context-specific professional ethical guidance framework for health care AI use (PEG-AI) may be necessary to ensure the safe, responsible, and ethically robust integration of the app into frontline care [23]. PEG-AI offers a context-sensitive governance approach for AI-embedded tools, essential for ensuring ethical implementation in resource-limited settings and safeguarding clinical autonomy, regardless of medical device classification [23].

Maintaining the use of the app within the boundaries of clinical competencies is therefore not solely a function of the tool itself but depends on the interaction between design features (eg, explainability, confidence indicators, and embedded guidance)

XSL•FO
RenderX

and broader system-level supports, including training, workflow integration, and context-specific ethical governance.

Situated Learning

In our study, only 1 FHW had ever heard of the WHO manual "Recognising Neglected Tropical Diseases Through Changes on the Skin: A Training Guide for Frontline Health Workers" [5]—a training guide specifically developed to build dermatological capacity among FHWs in resource-limited settings. Interestingly, the entire contents of the app are derived from this manual. However, our findings suggest that the format in which the information is both accessed and delivered shapes engagement and learning. While the manual offers comprehensive guidance, its didactic, decontextualized format perhaps separates learning from real-world clinical application. By contrast, the app delivers this same content in ways that are tailored to individual clinical cases, offering real-time feedback and follow-up information that are directly tied to patient encounters. Embedded in the relational space of the consultation, the app appears to support experiential, situated learning [59]. This aligns with well-established principles from the situated learning literature which emphasizes that learning is most effective when embedded within authentic contexts of practice [60-62].

In optimal conditions, the affordance of a continuously connected mobile device likely enhances this mode of learning. However, in our study, limited internet connectivity emerged as a barrier to consistent app use. Ensuring that the app's AI algorithm functions offline will be essential for equitable and sustained integration—an improvement the WHO Department for NTDs is currently exploring.

System-Level Impact and Referral Pathway

Existing literature suggests that over 80% of dermatological cases encountered by FHWs at the primary care level are common skin conditions, which are generally simpler to diagnose and manage [31]. In our study, FHWs reported that the app supported their ability to identify and manage these common skin conditions, often confirming their clinical judgment and enabling local facility management. However, when confronted with complex cases—particularly suspected skin NTDs—our findings show that FHWs remained more likely to defer to specialist referral, especially when the app's diagnostic output could not be corroborated by their own limited dermatological knowledge. This pattern aligns with a key dimension of the PEG-AI framework which emphasizes that AI-embedded tools should be used within the boundaries of clinical competencies [23].

In this context, the WHO Skin NTDs app appears to strengthen local dermatological care by enabling FHWs to manage common skin conditions more confidently, potentially reducing unnecessary referrals. This may help alleviate pressure on overburdened dermatology services, allowing specialists to focus on more complex or rare presentations, particularly skin NTDs.

However, this shift in clinical workflow introduces algorithmic implications. As FHWs increasingly retain patients at the facility level, the accuracy of the algorithm becomes more consequential. In such cases, it may be preferable for the algorithm to prioritize sensitivity over specificity (ie, to favor false positives over false negatives) to minimize the risk of missed cases, particularly as reliance on referral as a default pathway has declined (Figures 3 and 4).

This shift also raises important system-level considerations. While strengthening that facility-level management may improve access to care and reduce pressure on referral systems, it may also introduce new risks. If only suspected skin NTDs cases are escalated—while structural barriers to referral (such as distance, cost, and dermatologist shortages) remain—patients with complex dermatological conditions may still face delayed diagnosis and discontinuity of care, particularly in rural and underserved areas.

Assessing the full impact of AI-embedded tools such as the WHO Skin NTDs app—both intended and unintended—requires moving beyond the individual clinical encounter to examine how such tools reshape referral pathways, clinical workflows, and health system capacity. A systems-level process evaluation will be essential to understand these dynamics and their implications for patient outcomes, including health economic considerations [63,64]. This aligns with broader calls to evaluate mHealth interventions not only at the point of care but across the full health system to surface both enabling factors and unintended consequences [65].

Illness Trajectories

A key challenge frequently reported by FHWs was the lack of immediate management and follow-up support following the app's diagnostic suggestions. While the app provides AI-generated diagnostic suggestions instantaneously, the accompanying disease management guidance—although present—is in a separate section on the app and not seamlessly integrated into the clinical decision support workflow. This design choice may inadvertently reinforce a fragmented view of care, privileging diagnosis as an endpoint rather than as part of a broader continuum that includes, for example, management, follow-up, and psychosocial support [41]. This is particularly relevant, given the chronic course and complex illness trajectories associated with many common skin conditions and skin NTDs [66].

Enhancing the app's capacity to comprehensively support FHWs across the full illness trajectory—by embedding management guidance directly after diagnostic suggestion—could significantly broaden its value. The section of the app describing how to diagnose and treat each skin disease incorporates links to targeted training modules or embedded decision pathways (such as those aligned with WHO Academy resources) to better equip FHWs to deliver holistic, sustained care that responds to the complex and chronic illness trajectories of dermatological conditions [67].

Limitations

This study has several limitations. First, findings reflect the perspectives of FHWs only; patient perspectives on the app's influence on care, agency, and experience were not explored. This is a key aim that will be addressed in forthcoming ethnographic fieldwork. Second, FHW responses may have

XSL•FO
RenderX

been influenced by social desirability bias, given some of the research team's involvement with the app's development. However, independent facilitation was used to mitigate this risk [68,69]. Third, reported shifts in care may reflect short-term novelty effects rather than sustained practice change [70].

While this study aimed to generate a systems-level understanding of the app's implementation, future research should complement this with detailed patient pathway mapping and comparative studies across diverse contexts to assess transferability [62]. Of the 50 FHWs initially enrolled, 47 participated in the training and deployment phase, and 36 FHWs attended the follow-up workshop where qualitative data collection occurred (72% follow-up). This is notable, given the wide geographical spread of participants and the logistical barriers involved in attending a centralized workshop. All 5 counties were represented among those who attended. However, it is possible that nonattendees had lower engagement with the app during the deployment phase, although no definitive reasons were provided for their absence. This potential attrition bias should be considered when interpreting findings.

Conclusions

This study offers preliminary insights into the ways FHWs perceive the AI component of the WHO Skin NTDs app is reshaping frontline dermatological care in a complex, resource-limited health system. FHWs reported that the app supported more confident local-level facility management, disrupted habitual referral patterns, and fostered incremental knowledge acquisition. Importantly, our findings highlight the ethical and pedagogical complexities of integrating AI-embedded tools into frontline care. While FHWs reported that the app often strengthened clinical reasoning and learning, risks of overreliance and unintended shifts in clinical autonomy remain, particularly where baseline dermatological knowledge is limited. Future research should explore these dynamics through in situ observation and pathway-level evaluations. With continued attention to both pedagogical value and ethical governance, the WHO Skin NTD app has strong potential to advance the WHO's vision of equitable, person-centered skin health for all.

Acknowledgments

The authors would like to thank all partners involved in this study–the World Health Organization (WHO), Open University of Catalonia, Kenyan Medical Research Institute, the Kenyan Ministry of Health, and their funders at the Special Programme for Research and Training in Tropical Diseases (TDR). The authors are especially grateful to all FHWs who participated in the study, the health care facilities that hosted the research, and the patients who consented to the use of the WHO Skin Neglected Tropical Diseases app as part of their clinical consultations. Generative artificial intelligence was not used in any portion of the manuscript writing.

Funding

This study was funded by TDR, the Special Programme for Research and Training in Tropical Diseases, and the World Health Organization.

Data Availability

The datasets generated or analyzed during this study are not publicly available, as their primary value lies in the interpretive account developed through thematic analysis, rather than in reuse of raw data. All key findings and interpretations are presented in full within this published paper.

Authors' Contributions

JAR-P conceptualized and oversaw the study and secured project funding. CC was the principal investigator of the assessment study and developed the study protocol. RN served as site principal investigator and, together with EK, coordinated participant identification and recruitment, acting as the primary point of contact for frontline health workers. EEVQ led the qualitative component of the study, including development of the interview and focus group guides. Focus group discussions were facilitated by EEVQ, CC, RN, and DA. Semistructured interviews were conducted by EEVQ, CC, RN, DA, and AFC. EEVQ conducted the thematic analysis, with a subset of transcripts double coded by AB and KO. EEVQ drafted the manuscript. All authors reviewed and approved the final version. The participants did not have a professional or personal relationship with the researchers prior to the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Semistructured interview topic guide.
[DOCX File, 18 KB - mhealth_v14i1e81829_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Focus group discussion topic guide.
[DOCX File, 17 KB - mhealth_v14i1e81829_app2.docx]

Multimedia Appendix 3
Supporting quotes for causal loop diagrams.
[DOCX File, 25 KB - mhealth_v14i1e81829_app3.docx]

Checklist 1
COREQ (Consolidated Criteria for Reporting Qualitative Research).
[DOCX File, 39 KB - mhealth_v14i1e81829_app4.docx]

References

1. Ending the neglect to attain the sustainable development goals: a road map for neglected tropical diseases 2021-2030. World Health Organization. URL: https://www.who.int/publications/i/item/9789240010352 [accessed 2024-09-03]
2. Yotsu RR, Fuller LC, Murdoch ME, et al. A global call for action to tackle skin-related neglected tropical diseases (skin NTDs) through integration: an ambitious step change. PLoS Negl Trop Dis 2023 Jun 15;17(6):e0011357. [doi: 10.1371/journal.pntd.0011357] [Medline: 37319139]
3. Timothy JWS, Rogers E, Halliday KE, et al. Quantifying population burden and effectiveness of decentralized surveillance strategies for skin-presenting neglected tropical diseases, Liberia. Emerg Infect Dis 2022 Sep;28(9):1755-1764. [doi: 10.3201/eid2809.212126] [Medline: 35997318]
4. Alderton DL, Ackley C, Trueba ML. The psychosocial impacts of skin-neglected tropical diseases (SNTDs) as perceived by the affected persons: a systematic review. PLoS Negl Trop Dis 2024 Aug 2;18(8):e0012391. [doi: 10.1371/journal.pntd.0012391] [Medline: 39093848]
5. Recognizing neglected tropical diseases through changes on the skin: a training guide for frontline health workers. World Health Organization. URL: https://www.who.int/publications/i/item/9789241513531 [accessed 2024-09-18]
6. Stangl AL, Earnshaw VA, Logie CH, et al. The health stigma and discrimination framework: a global, crosscutting framework to inform research, intervention development, and policy on health-related stigmas. BMC Med 2019 Feb 15;17(1):31. [doi: 10.1186/s12916-019-1271-3] [Medline: 30764826]
7. Ahmat A, Okoroafor SC, Kazanga I, et al. The health workforce status in the WHO African Region: findings of a cross-sectional study. BMJ Glob Health 2022 May;7(Suppl 1):e008317. [doi: 10.1136/bmjgh-2021-008317] [Medline: 35675966]
8. Faye O, Bagayoko CO, Dicko A, et al. A teledermatology pilot programme for the management of skin diseases in primary health care centres: experiences from a resource-limited country (Mali, West Africa). Trop Med Infect Dis 2018 Aug 17;3(3):88. [doi: 10.3390/tropicalmed3030088] [Medline: 30274484]
9. Agarwal S, Perry HB, Long LA, Labrique AB. Evidence on feasibility and effective use of mHealth strategies by frontline health workers in developing countries: systematic review. Trop Med Int Health 2015 Aug;20(8):1003-1014. [doi: 10.1111/tmi.12525] [Medline: 25881735]
10. Community health worker programmes in the WHO African region: evidence and options-policy brief. World Health Organization. URL: https://www.afro.who.int/publications/community-health-worker-programmes-who-african-region-policy-brief [accessed 2024-09-04]
11. WHO expert committee on leprosy: eighth report. World Health Organization. URL: https://iris.who.int/items/da598737-ff32-4d11-b679-e5033641cacd [accessed 2024-09-18]
12. Hay RJ, Asiedu K. Skin-related neglected tropical diseases (skin NTDs)—a new challenge. Trop Med Infect Dis 2018 Dec 25;4(1):4. [doi: 10.3390/tropicalmed4010004] [Medline: 30585179]
13. Bharadwaj M, Bengtson M, Golverdingen M, Waling L, Dekker C. Diagnosing point-of-care diagnostics for neglected tropical diseases. PLoS Negl Trop Dis 2021 Jun 17;15(6):e0009405. [doi: 10.1371/journal.pntd.0009405] [Medline: 34138846]
14. Seventy-eighth World Health Assembly— daily update: 24 May 2025. World Health Organization. URL: https://www.who.int/news/item/24-05-2025-seventy-eighth-world-health-assembly---daily-update--24-may-2025 [accessed 2025-06-16]
15. Webster JP, Molyneux DH, Hotez PJ, Fenwick A. The contribution of mass drug administration to global health: past, present and future. Philos Trans R Soc Lond B Biol Sci 2014 May 12;369(1645):20130434. [doi: 10.1098/rstb.2013.0434] [Medline: 24821920]
16. Hengst TM, Lechner L, Dohmen D, Bolman CAW. The facilitators and barriers of mHealth adoption and use among people with a low socio-economic position: a scoping review. Digit Health 2023 Sep 5;9:20552076231198702. [doi: 10.1177/20552076231198702] [Medline: 37691766]
17. WHO Global Observatory for eHealth. MHealth: new horizons for health through mobile technologies: second global survey on eHealth. : World Health Organization; 2011 URL: https://iris.who.int/handle/10665/44607 [accessed 2026-07-04]

XSL•FO
RenderX

18. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23. [doi: 10.2196/mhealth.8873] [Medline: 29343463]
19. Ciecierski-Holmes T, Singh R, Axt M, Brenner S, Barteit S. Artificial intelligence for strengthening healthcare systems in low- and middle-income countries: a systematic scoping review. NPJ Digit Med 2022 Oct 28;5(1). [doi: 10.1038/s41746-022-00700-y] [Medline: 36307479]
20. Frej A, Cano M, Ruiz-Postigo JA, et al. Assessing the quality of the World Health Organization's Skin NTDs App as a training tool in Ghana and Kenya: protocol for a cross-sectional study. JMIR Res Protoc 2022 Dec 8;11(12):e39393. [doi: 10.2196/39393] [Medline: 36480252]
21. Cano M, Ruiz-Postigo JA, Macharia P, et al. Evaluating the World Health Organization's SkinNTDs app as a training tool for skin neglected tropical diseases in Ghana and Kenya: cross-sectional study. J Med Internet Res 2024 Apr 30;26(1):e51628. [doi: 10.2196/51628] [Medline: 38687587]
22. Quilter EEV, Butlin CR, Carrion C, Ruiz-Postigo JA. The WHO Skin NTD mobile application – a paradigm shift in leprosy diagnosis through Artificial Intelligence? LEPROSY 2024;95(2):1-3. [doi: 10.47276/lr.95.2.2024030]
23. Smith H, Ives J. Developing professional ethical guidance for healthcare AI use (PEG-AI): an attitudinal survey pilot. AI Soc 2025 Oct;40(7):5439-5455. [doi: 10.1007/s00146-025-02276-z] [Medline: 41788979]
24. Muga R, Kizito P, Mbayah MM, Gakuruh T. Overview of the health system in Kenya. In: Kenya Service Provision Assessment Survey 2004: National Coordinating Agency for Population and Development, Ministry of Health (Nairobi, Kenya), ORC Macro; 2005. URL: https://dhsprogram.com/pubs/pdf/spa8/02chapter2.pdf [accessed 2026-07-04]
25. Elson L, Kamau C, Koech S, et al. National prevalence and risk factors for tungiasis in Kenya. Infect Dis Poverty 2023 Sep 18;12(1):85. [doi: 10.1186/s40249-023-01131-x] [Medline: 37723532]
26. Ngere I, Gufu Boru W, Isack A, et al. Burden and risk factors of cutaneous leishmaniasis in a peri-urban settlement in Kenya, 2016. PLoS One 2020 Jan 23;15(1):e0227697. [doi: 10.1371/journal.pone.0227697] [Medline: 31971945]
27. Ratemo SN, Denning DW. Burden of fungal infections in Kenya. Mycology 2023 Apr 27;14(2):142-154. [doi: 10.1080/21501203.2023.2204112] [Medline: 37152847]
28. Njenga W, Mwancha-Kwasa M, Fossa K, et al. Skin care habits of patients with eczema at Kiambu Level 5 Hospital, Kiambu county, Kenya: a retrospective cross-sectional study. JAAD Int 2024 May 7;16:214-220. [doi: 10.1016/j.jdin.2024.04.011] [Medline: 39045170]
29. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
30. Serrano Pons J, Romero-Lopez A, Muñoz I, et al. The World Health Organization skin neglected tropical diseases app: a dynamic capacity building training tool enhanced with artificial intelligence. J Invest Dermatol 2026 May 6:S0022-202X(26)01185-1. [doi: 10.1016/j.jid.2026.03.043] [Medline: 42102907]
31. Hay RJ, Johns NE, Williams HC, et al. The global burden of skin disease in 2010: an analysis of the prevalence and impact of skin conditions. J Invest Dermatol 2014 Jun;134(6):1527-1534. [doi: 10.1038/jid.2013.446] [Medline: 24166134]
32. Benčević M, Habijan M, Galić I, Babin D, Pižurica A. Understanding skin color bias in deep learning-based skin lesion segmentation computer methods and programs in biomedicine. Comput Methods Programs Biomed 2024 Mar;245:108044. [doi: 10.1016/j.cmpb.2024.108044] [Medline: 38290289]
33. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2008 Jul 21;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
34. Barbrook-Johnson P, Penn AS. Causal loop diagrams. In: Systems Mapping: How to Build and Use Causal Models of Systems: Springer; 2022. URL: https://link.springer.com/book/10.1007/978-3-031-01919-7 [accessed 2026-06-26] [doi: 10.1007/978-3-031-01919-7]
35. County statistical abstract Nakuru County. : Kenya National Bureau of Statistics; 2015 URL: https://www.knbs.or.ke/wp-content/uploads/2023/09/2015-County-Statistical-Abstracts-Nakuru.pdf [accessed 2026-06-23]
36. County statistical abstract Baringo County. : Kenya National Bureau of Statistics; 2015 URL: https://www.knbs.or.ke/wp-content/uploads/2023/09/2015-County-Statistical-Abstracts-Baringo.pdf [accessed 2026-06-23]
37. County statistical abstract Kajiado County. : Kenya National Bureau of Statistics; 2015 URL: https://www.knbs.or.ke/wp-content/uploads/2023/09/2015-County-Statistical-Abstracts-Kajiado.pdf [accessed 2026-06-23]
38. Statistical abstract 2021. : Kenya National Bureau of Statistics; 2021 URL: https://www.knbs.or.ke/wp-content/uploads/2023/09/2021-Statistical-Abstract.pdf [accessed 2026-06-23]
39. County statistical abstract Kwale County. : Kenya National Bureau of Statistics; 2015 URL: https://www.knbs.or.ke/wp-content/uploads/2023/09/2015-County-Statistical-Abstracts-Kwale.pdf [accessed 2026-06-23]
40. Meyer AJ, Armstrong-Hough M, Babirye D, et al. Implementing mHealth interventions in a resource-constrained setting: case study from Uganda. JMIR Mhealth Uhealth 2020 Jul 13;8(7):e19552. [doi: 10.2196/19552] [Medline: 32673262]
41. Livingston J. Improvising Medicine. : Duke University Press; 2013. [doi: 10.1215/9780822395768]
42. McKenna K, Leykum LK, McDaniel RR. The role of improvising in patient care. Health Care Manage Rev 2013;38(1):1-8. [doi: 10.1097/HMR.0b013e31823ea9c7] [Medline: 22261668]

XSL•FO
RenderX

43. Kapepo M, Van Belle JP, Weimann E. WhatsApp as an improvisation of health information systems in Southern African public hospitals: a socio-technical perspective. Procedia Comput Sci 2025;256:805-815. [doi: 10.1016/j.procs.2025.02.182]
44. Ogundaini O, de la Harpe R, McLean N. Unintended consequences of technology-enabled work activities experienced by healthcare professionals in tertiary hospitals of sub-Saharan Africa. Afr J Sci Technol Innov Dev 2022 Jun 7;14(4):876-885. [doi: 10.1080/20421338.2021.1899556]
45. Matthiesen S, Diederichsen SZ, Hansen MKH, et al. Clinician preimplementation perspectives of a decision-support tool for the prediction of cardiac arrhythmia based on machine learning: near-live feasibility and qualitative study. JMIR Hum Factors 2021 Nov 26;8(4):e26964. [doi: 10.2196/26964] [Medline: 34842528]
46. Dlugatch R, Georgieva A, Kerasidou A. AI-driven decision support systems and epistemic reliance: a qualitative study on obstetricians' and midwives' perspectives on integrating AI-driven CTG into clinical decision making. BMC Med Ethics 2024 Jan 6;25(1):6. [doi: 10.1186/s12910-023-00990-1] [Medline: 38184595]
47. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Pshycol Rev 1977 Mar;84(2):191-215. [doi: 10.1016/0146-6402(78)90002-4] [Medline: 847061]
48. Adegbesan A, Akingbola A, Aremu O, Adewole O, Amamdikwa JC, Shagaya U. From scalpels to algorithms: the risk of dependence on artificial intelligence in surgery. J Med Surg Public Health 2024 Aug;3:100140. [doi: 10.1016/j.glmedi.2024.100140]
49. Romeo G, Conti D. Exploring automation bias in human–AI collaboration: a review and implications for explainable AI. AI & Soc 2026 Jan;41(1):259-278. [doi: 10.1007/s00146-025-02422-7]
50. Hatherley JJ. Limits of trust in medical AI. J Med Ethics 2020 Jul;46(7):478-481. [doi: 10.1136/medethics-2019-105935] [Medline: 32220870]
51. Jeon Y, Hwang C, Chen XA. Empowering medical data labeling for non-experts with DANNY: enhancing accuracy and mitigating over-reliance on AI. 2025 Presented at: Proceedings of the 30th International Conference on Intelligent User Interfaces; Mar 24-27, 2025; Cagliari, Italy p. 624-640. [doi: 10.1145/3708359.3712161]
52. Klingbeil A, Grützner C, Schreck P. Trust and reliance on AI — An experimental study on the extent and costs of overreliance on AI. Comput Human Behav 2024 Nov;160(C):108352. [doi: 10.1016/j.chb.2024.108352]
53. Lee MH, Chew CJ. Understanding the effect of counterfactual explanations on trust and reliance on AI for human-AI collaborative clinical decision making. Proc ACM Hum-Comput Interact 2023 Oct 4;7(CSCW2):1-22. [doi: 10.1145/3610218]
54. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak 2020 Nov 30;20(1):1-9. [doi: 10.1186/s12911-020-01332-6] [Medline: 33256715]
55. Tonekaboni S, Joshi S, McCradden MD, Goldenberg A. What clinicians want: contextualizing explainable machine learning for clinical end use. arXiv. Preprint posted online on May 13, 2019. [doi: 10.48550/arXiv.1905.05134]
56. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. JAMA 2018 Dec 4;320(21):2199-2200. [doi: 10.1001/jama.2018.17163] [Medline: 30398550]
57. Ghosh P, Posner KL, Hyland SL, et al. Framing machine learning opportunities for hypotension prediction in perioperative care: a socio-technical perspective. ACM Trans Comput Hum Interact 2023 Sep 23;30(5):1-33. [doi: 10.1145/3589953]
58. Zhang M, Ehrmann D, Mazwi M, Eytan D, Ghassemi M, Chevalier F. Get to the point! Problem-based curated data views to augment care for critically ill patients. 2022 Apr 29 Presented at: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems; Apr 29 to May 5, 2022; New Orleans, LA USA p. 1-13. [doi: 10.1145/3491102.3501887]
59. Giles AK, Pitonyak JS, George-Paschal L, Piernik-Yoder B, Taff SD. Situated learning. In: Ikiugu M, Taff SD, Kantartzis S, Pollard N, editors. Routledge Companion to Occupational Therapy: Theories, Concepts and Models: Routledge; 2025. [doi: 10.4324/9781003526766]
60. Jeng YL, Hong ML, Lu HT. The development of the mobile situated learning application based on microlocation technology. Presented at: 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI); Jul 9-13, 2017. [doi: 10.1109/IIAI-AAI.2017.108]
61. Pfeiffer VDI, Gemballa S, Jarodzka H, Scheiter K, Gerjets P. Situated learning in the mobile age: mobile devices on a field trip to the sea. Res Learn Technol 2011 Dec;17(3):187-199. [doi: 10.1080/09687760903247666]
62. O'Brien BC, Battista A. Situated learning theory in health professions education research: a scoping review. Adv Health Sci Educ Theory Pract 2020 May;25(2):483-509. [doi: 10.1007/s10459-019-09900-w] [Medline: 31230163]
63. Moore GF, Audrey S, Barker M, et al. Process evaluation of complex interventions: medical research council guidance. BMJ 2015 Mar 19;350:h1258. [doi: 10.1136/bmj.h1258] [Medline: 25791983]
64. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013 Aug 6;1(2):160-171. [doi: 10.9745/GHSP-D-13-00031] [Medline: 25276529]
65. Maar MA, Yeates K, Perkins N, et al. A framework for the study of complex mHealth interventions in diverse cultural settings. JMIR Mhealth Uhealth 2017 Apr 20;5(4):e47. [doi: 10.2196/mhealth.7044] [Medline: 28428165]
66. Chandler DJ, Fuller LC. The skin-a common pathway for integrating diagnosis and management of NTDs. Trop Med Infect Dis 2018 Sep 10;3(3):101. [doi: 10.3390/tropicalmed3030101] [Medline: 30274497]
67. WHO academy. World Health Organization. URL: https://www.who.int/about/who-academy [accessed 2025-06-05]

XSL•FO
RenderX

68. Grimm P. Social desirability bias. Wiley International Encyclopedia of Marketing. 2010 Dec 15. URL: https://onlinelibrary.wiley.com/doi/full/10.1002/9781444316568.wiem02057 [accessed 2025-06-05]
69. Mathieu C. Performance appraisal. In: Dark Personalities in the Workplace: Academic Press; 2021:71-96.
70. Diaper G. The Hawthorne Effect: a fresh examination. Educ Stud 1990 Jan;16(3):261-267. [doi: 10.1080/0305569900160305]

Abbreviations

AI: artificial intelligence
CLD: causal loop diagram
EMR: electronic medical record
FGD: focus group discussion
FHW: frontline health worker
KEMRI: Kenyan Medical Research Institute
mHealth: mobile health
NTD: Neglected Tropical Disease
PEG-AI: professional ethical guidance framework for health care AI use
SSI: semistructured interview
WHO: World Health Organization

Edited by L Buis; submitted 04.Aug.2025; peer-reviewed by B Eapen, L Orii; revised version received 12.Jun.2026; accepted 15.Jun.2026; published 14.Jul.2026.

Please cite as:
Quilter EEV, Nyangacha R, Kinyeru E, Fuster Casanovas A, Brigden A, O'Hara K, Atieno D, Ruiz-Postigo JA, Carrion C
Frontline Health Workers' Perspectives of the World Health Organization Skin Neglected Tropical Diseases App in Kenya: Qualitative Study on AI-Embedded mHealth Implementation
JMIR Mhealth Uhealth 2026;14:e81829
URL: https://mhealth.jmir.org/2026/1/e81829
doi:10.2196/81829

© Emily E V Quilter, Ruth Nyangacha, Esther Kinyeru, Aïna Fuster Casanovas, Amberly Brigden, Kenton O'Hara, Diana Atieno, José A Ruiz-Postigo, Carme Carrion. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Feasibility and Acceptability of a Mobile App and Wearable Device for Collecting Mental Health Survey and Passively Sensed Data Among Health Care Workers in Kenya: Mixed Methods Pilot Study

Linda Khakali[1], MA; Andrew Aballa[1], MSc; Dorcas G Mwigereri[2], MSc; Eileen M Weinheimer-Haus[3,4], PhD; Willie Njoroge[1], MSc; Rachel Maina[1], PhD; Amos Bunde[5], BSc; Peris Musitia[6], MA; Moses K Nyongesa[1], PhD; James Orwa[2], PhD; Jasmit Shah[1], PhD; Zhenke Wu[7], PhD; Anthony K Ngugi[2], PhD; Lukoye Atwoli[1,8], MD, PhD; Srijan Sen[9], MD, PhD; Candace Kolars[3,4], MPH, PhD; Akbar K Waljee[3,4], MD, MSc; Amina Abubakar[6,10], PhD; Elena Frank[9*], PhD; Zul Merali[1*], PhD

[1]Brain and Mind Institute, Aga Khan University Nairobi, 3rd Parklands Avenue, Nairobi, Kenya

[10]Neurosciences Unit, Kenya Medical Research Institute, Wellcome Trust Research Programme, Kilifi, Kenya

[2]Department of Population Health, Aga Khan University Nairobi, Nairobi, Kenya

[3]Center for Global Health Equity, University of Michigan, Ann Arbor, MI, United States

[4]Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, MI, United States

[5]Computing and Data Innovation Office, Aga Khan University Nairobi, Nairobi, Kenya

[6]Institute for Human Development, Aga Khan University Nairobi, Nairobi, Kenya

[7]Department of Biostatistics, University of Michigan, Ann Arbor, MI, United States

[8]Department of Medicine, Medical College East Africa, Aga Khan University Nairobi, Nairobi, Kenya

[9]Michigan Neuroscience Institute, University of Michigan, Ann Arbor, MI, United States

[*]these authors contributed equally

Corresponding Author:
Linda Khakali, MA
Brain and Mind Institute, Aga Khan University Nairobi, 3rd Parklands Avenue, Nairobi, Kenya

Abstract

Background: Mobile apps and wearable devices may help to facilitate early detection of mental health conditions by providing objective, real-time data to supplement other forms of feedback and diagnoses. Few studies have investigated the acceptability and feasibility of using a mobile app to track survey- and wearable-based data in mental health research in Sub-Saharan Africa.

Objective: This pilot study evaluated the feasibility and acceptability of using a mobile app and wearables to capture mental health–based survey data and passively sensed data among Kenyan health care workers.

Methods: A mixed methods study was conducted among health care workers employed at 4 hospitals in Nairobi, Kenya, over 30 days. A mobile app was used to collect and integrate active (baseline questionnaire and daily mood) and passive (wearable) data. The baseline questionnaire gathered information on sociodemographics, work environment, and mental health assessments on depression, anxiety, personality, early family environment, posttraumatic stress disorder, and substance use. A wearable device was used to gather data on steps, heart rate, and sleep. Qualitative interviews were conducted post trial to gain in-depth insights into participants' experiences during the study.

Results: Fifty-one participants enrolled in the pilot study. They were primarily nurses (47%) and female (70%), with a median (IQR) age of 32 (29-36) years. Attrition over 30 days was low, with only one participant dropping out due to device malfunction, which was a broken screen. Completeness of the baseline survey was high, with participants completing 96.1% of the questions. Further, 58% of the daily mood ratings were completed over the 30 days. For the wearable measures, participants submitted steps, heart rate, and sleep data on 93%, 73%, and 51% of study days, respectively. The proportion of days the wearable was worn for over 10 hours was 63%. Interviews revealed 2 primary themes. The first was intrinsic and extrinsic motivation; participants indicated that they liked having their health metrics tracked and receiving congratulatory messages from the app, encouraging increased step counts. The second theme was technical and usability challenges; 48% (10/21) of the participants reported discomfort wearing the watch while sleeping and challenges with synchronization of data due to the nonautomated nature of the process. Participants suggested additional prompts to remind them to complete the daily mood question.

Conclusions: This pilot study demonstrates the feasibility of deploying mental health surveys, collecting data through wearable devices, and integrating such data within a single mobile platform under real-world infrastructure constraints. Health care workers

XSL•FO
RenderX

in Kenya were willing to provide sensitive information through mental health assessments using a mobile app. To improve adherence, future studies should consider addressing some contextual factors such as daily prompts, enhanced data synchronization methods, and comfort concerns to improve adherence, especially during sleep.

(JMIR Mhealth Uhealth 2026;14:e77761) doi:10.2196/77761

KEYWORDS

mental health; health care workers; mHealth; wearable device; mobile app; Kenya; low- and middle-income countries; Africa; mobile health

Introduction

Globally, health care workers (HCWs) play a crucial role in the health care system, in support of a wide array of responsibilities from patient care to supporting critical operational functions [1]. Human resources are limited in health care systems located in low- and middle-income countries (LMICs), resulting in strain on the professionals in the system [2]. Additionally, the nature of the work involves exposure to potentially traumatic events which may place HCWs at greater risk for psychological distress [3]. A meta-analysis including HCWs from 50 countries found that 42% reported anxiety symptoms, 33% reported depressive symptoms, and 42% reported insomnia, highlighting the high burden of these mental disorders [4]. In Kenya, similar findings were reported showing HCWs to be at risk of mental disorders such as depression at 32%, anxiety at 36%, and harmful alcohol use at 43.9%, and this was exacerbated during the COVID pandemic [5,6].

Prevention and early treatment of mental health conditions and psychological distress have the potential to reduce the impact of these disorders on the workforce and improve productivity and quality of output [7]. The traditional way of diagnosing mental health conditions is through questionnaires and scorecards during a patient visit [8]. However, point-in-time (eg, during an appointment with a health care provider) and retrospective questionnaires can result in the loss of valuable data informing diagnostic conclusions [9]. There is a growing interest in the use of digital technologies, such as smartphone apps and wearables, to augment traditional mental health care, particularly in LMICs where resources are limited for surveillance, diagnosis, and treatment. These approaches leverage streams of passively collected data captured from personal digital sensors to infer, monitor, and predict changes in health states over time [10-12].

Wearable devices capture objective indicators of sleep, physical activity, and heart rate, and changes in these variables have been closely linked to mental health states. For example, short or fragmented sleep has been associated with depression and anxiety, reduced physical activity has been linked to depressive symptoms, and elevated resting heart rate or reduced heart rate variability can reflect autonomic dysregulation observed in stress-related mood disorders [13-18]. Together, these measures provide real-time, ecological data that complement self-reported questionnaires and supplement longitudinal monitoring of mental health trajectories.

Wearable devices and mobile health (mHealth) technologies have been used in high-income countries to assess mental health indicators in various populations, including medical residents [18]. However, the integration of wearable device information with traditional questionnaires into a data collection platform in Africa has been scant, partly due to the infrastructure cost and limited availability of devices [19]. Despite these limitations, the continued increase in mobile phone access in LMICs such as Kenya provides an opportunity to digitize data collection through mobile apps [20]. In Kenya, studies have been conducted on the use of wearable devices for different health conditions and climate factors, with limited studies focused on mental health factors [21]. Thus, understanding the feasibility and acceptability of these platforms is important in assessing readiness for deployment of digitized platforms for mental health research in countries such as Kenya.

This pilot study evaluated the feasibility and acceptability of using a mobile technology app and wearables to collect both survey-based and passively sensed data among Kenyan HCWs to inform the deployment of a larger mHealth study. The primary focus was on the practical ability to deploy surveys, sustain device wear, and synchronize wearable data through an integrated app under real-world infrastructure constraints.

Methods

Research Design

This study used a sequential explanatory mixed methods approach where we conducted quantitative survey and wearable data collection for 30 days, after which qualitative interviews were completed to provide context to the quantitative outcomes. Our results are reported in line with the CONSORT (Consolidated Standards of Reporting Trials) extension for randomized pilot and feasibility trials (Multimedia Appendix 1) [22].

Study Participants and Sampling Methods

HCWs, including nurses, doctors, psychologists, pharmacists, radiographers, physiotherapists, and nutritionists, were recruited from 4 urban health care facilities in Nairobi, Kenya from July to August 2022 (Table 1). Study inclusion criteria were aged 18 years or older, employment as an HCW, and smartphone ownership.

XSL•FO
RenderX

Table . Description of health care facilities in Nairobi, Kenya[a,b].

Facility	Description
1	A level 5 tertiary private health care facility with 1088 health care workers.
2	A level 6 national referral hospital with 665 health care workers.
3	A level 5 county referral hospital, with 700 health care workers.
4	A level 5 obstetric and delivery referral hospital, with 396 health care workers.

[a]Health care facilities associated with a 30-day mixed methods pilot study focused on the feasibility and acceptability of using a mobile app to track survey- and wearable-based data in mental health research among Kenyan health care workers.

[b]Level 5 facilities are secondary referral hospitals that offer diagnostic and specialized care and more comprehensive medical care. Level 6 facilities are national referral and teaching hospitals.

Recruitment

Recruitment strategies have been previously described [23]. Briefly, in-person sensitization meetings were conducted in study sites. During these sessions, HCWs were informed about the study objectives, duration, and study procedures. Additional recruitment strategies included email blasts, posters, and pamphlets placed around the health care facilities, and snowballing through participants who had already registered. Those interested in participating were invited to complete an online registration form that collected names, work cadre, email addresses, and phone numbers for enrollment. To increase reach, the registration form, along with brochures containing frequently asked questions sheets and QR codes for registering, was circulated in departmental WhatsApp (Meta Platforms) groups, Continuous Medical Education meetings, and virtual sensitization meetings. In total, 356 HCWs expressed interest in the study.

Of these, 60 individuals were randomly selected and stratified by facility using a random number generator. The target sample size for the pilot was n=50; thus, 60 individuals were contacted with an anticipated participation rate of 80%. The randomly selected individuals were contacted by email and invited to enroll in the pilot study within 7 days. Participants enrolled using the MyDataHelps (CareEvolution) mobile app. Within a week, 51 individuals provided informed consent and were enrolled in the study. Qualitative exit interviews were performed at the end of the study, with a subset of participants (n=22) selected through convenience sampling. Participants were recruited from individuals who were readily available and had participated in the broader study, rather than being selected through random sampling. Potential participants were identified based on their availability and willingness to participate in follow-up qualitative interviews. Representation across professional cadres and institutions was desired; however, the availability of participants was a key factor for inclusion due to demanding schedules. Participants were contacted by phone to schedule an appointment to meet face-to-face with the researchers to review the consent form. Of the 22 who were contacted, 1 participant declined to participate in the interviews.

Survey and Wearable Data Collection Procedures

Following study enrollment, participants completed a baseline questionnaire that included information on sociodemographics, work environment, and the following assessments: (1) the Neuroticism, Extraversion, Openness to Experience Five-Factor Inventory to assess domains of personality in 5 areas [24]; (2) Risky Families Questionnaire to assess the degree of physical, mental, and emotional distress experienced in their homes during childhood and adolescence [25]; (3) Generalized Anxiety Disorder-7 to measure anxiety scores [26]; (4) Patient Health Questionnaire-9 to assess symptoms of depression [27]; (5) posttraumatic stress disorder (PTSD) checklist to assess presence and severity of PTSD symptoms [28]; and (6) World Health Organization Alcohol, Smoking and Substance Involvement Screening Test to assess substance use–related health risks and substance use disorders [29].

Following submission of the baseline questionnaire, participants received a Fitbit Inspire 2 (Fitbit Inc) and were instructed to wear the device continuously for the duration of the 30-day study. Participants were instructed on how to use wrist mode, verify that the sleep feature was on, and ensure syncing of the data with the mobile apps. Additionally, participants were advised to charge the device as soon as a low battery notification appeared. The battery life on Fitbit is up to 10 days on a single charge and charges to full in about 1 to 2 hours. Sleep, step count, heart rate, and mood rating were collected daily for 30 days.

The mobile app, MyDataHelps, was used to collect and integrate active (baseline questionnaire and daily mood) and passive (wearable) data. The app was available on Android and iOS operating systems, and on the web. Step-by-step instructions were provided for the web-based baseline questionnaire. The app provided a daily prompt to complete the mood rating (scale of 1-10). Wearable data were uploaded daily to the mobile app and integrated into the data collection platform through automatic synchronization from the Fitbit device to the Fitbit app on the participant's mobile phone. To ensure the data synchronization from the Fitbit app to the MyDataHelps app, participants were required to have their mobile phone connected to the internet and Bluetooth enabled. The data synchronization process retrieved data stored on the device for up to 7 days. When gaps in data upload exceeded 3 days, the app sent a prompt to remind participants to synchronize their data.

Exit Interviews

At the end of the 30-day pilot study period, a subset of participants was interviewed by members of the research team (LK, WN, and A Aballa) at the participants' respective health

XSL•FO
RenderX

facilities. The interview guide consisted of questions on user experience, challenges, and recommendations regarding the content and technology.

The interviews were conducted by trained social scientists. Prior to the interview, the interviewers completed a 3-day training on research ethics and the interview guide. A pretest was conducted with HCWs who were not part of the study, and the interview guide was later revised to ensure the flow and clarity of the questions. The interviews took approximately 40 to 70 minutes.

Data Analysis

Feasibility of the pilot study was evaluated using baseline questionnaire response time, data completeness, recruitment results, and study attrition. For the baseline questionnaire, data completeness was calculated based on the proportion of visible questions answered, which excludes branching questions. Data completeness was also analyzed based on the number of days participants completed the daily mood rating and uploaded any steps, heart rate (heart rate intraday count), and sleep (duration) data. The average daily step count was also reported, excluding days with zero step count. The number of days participants wore the wearable for at least 10 hours was calculated using heart rate intraday minute count, as studies have suggested this is the minimal wear time to indicate a full day of use [30]. Descriptive statistics were calculated using R (version 4.1.1; R Core Team), and summary statistics were presented as mean and SD or median and IQR for continuous data, and frequencies and percentages for categorical data.

Acceptability of the baseline questionnaire and wearables was based on user experiences provided in qualitative exit interviews.

The interviews were recorded and transcribed verbatim, and a deductive thematic analysis approach was used to analyze the data with the support of NVivo Software (version 12; QSR International). Interviewers created field notes to capture participants' impressions and observations during the interviews. The analysis followed the key stages of thematic analysis: familiarization with the data, generating initial codes, searching for themes, reviewing and refining themes, and defining and naming themes [31]. In the initial phase, transcripts were read thoroughly, and data were coded line by line, with concepts and ideas identified. As the analysis progressed, codes were refined and grouped into potential themes, which were then reviewed and further developed to reflect deeper meanings. In the final phase, the themes were clearly defined and organized to capture the core ideas of the study based on the objectives of the study. The coding matrices and final themes were reviewed and refined by LK, A Aballa, WN, and RM to ensure the accuracy and depth of the analysis.

Ethical Considerations

The study obtained approvals from Aga Khan University Institutional Scientific Ethics and Review Committee (reference number 2021/IERC-166(V2)), the University of Nairobi and the Kenyatta National Hospital Ethics Review Committee (reference number KNH-ERC/A/473), regulatory bodies including the National Commission for Science Technology and Innovation (778557) and the Nairobi County and Nairobi Metropolitan Services (SMS; reference number EOP/NMS/HS/167), and hospital administrators in each study site. The study was conducted in accordance with the Declaration of Helsinki and written and informed consent was obtained from all participants. The consent form provided information on the study objective, willingness to participate, data safety, and confidentiality. For qualitative interviews, the consent included exit interview procedures and consent to be audio recorded. All participant data were deidentified using study identifier and not individual names, and data were securely transferred to the institution server. Due to budget constraints in maintaining prepaid data packages in LMICs, participants were compensated (KSh 500 or US $3.58 equivalent) for the purchase of prepaid data packages that were needed to support timely synchronization. Qualitative participants were compensated (KSh 1000 or US $7.06 equivalent) for their time. All participants were able to keep the wearable device at the end of the study.

Results

Participant Characteristics

The participants were primarily female (70%), with 69% reporting attainment of a bachelor's or master's degree (Table 2). The median age of the participants was 32 (IQR 29 - 36) years, and nurses (47%) were the largest health care cadre. The median duration of employment in health care was 8 (IQR 5 - 12) years. The retention rate for the 30-day study duration was 98% (n=50). Characteristics of those who participated in the qualitative interviews are in Table 2 and Figure 1.

XSL•FO
RenderX

Table . Characteristics of enrolled participants and of those who subsequently took part in an exit interviews[a].

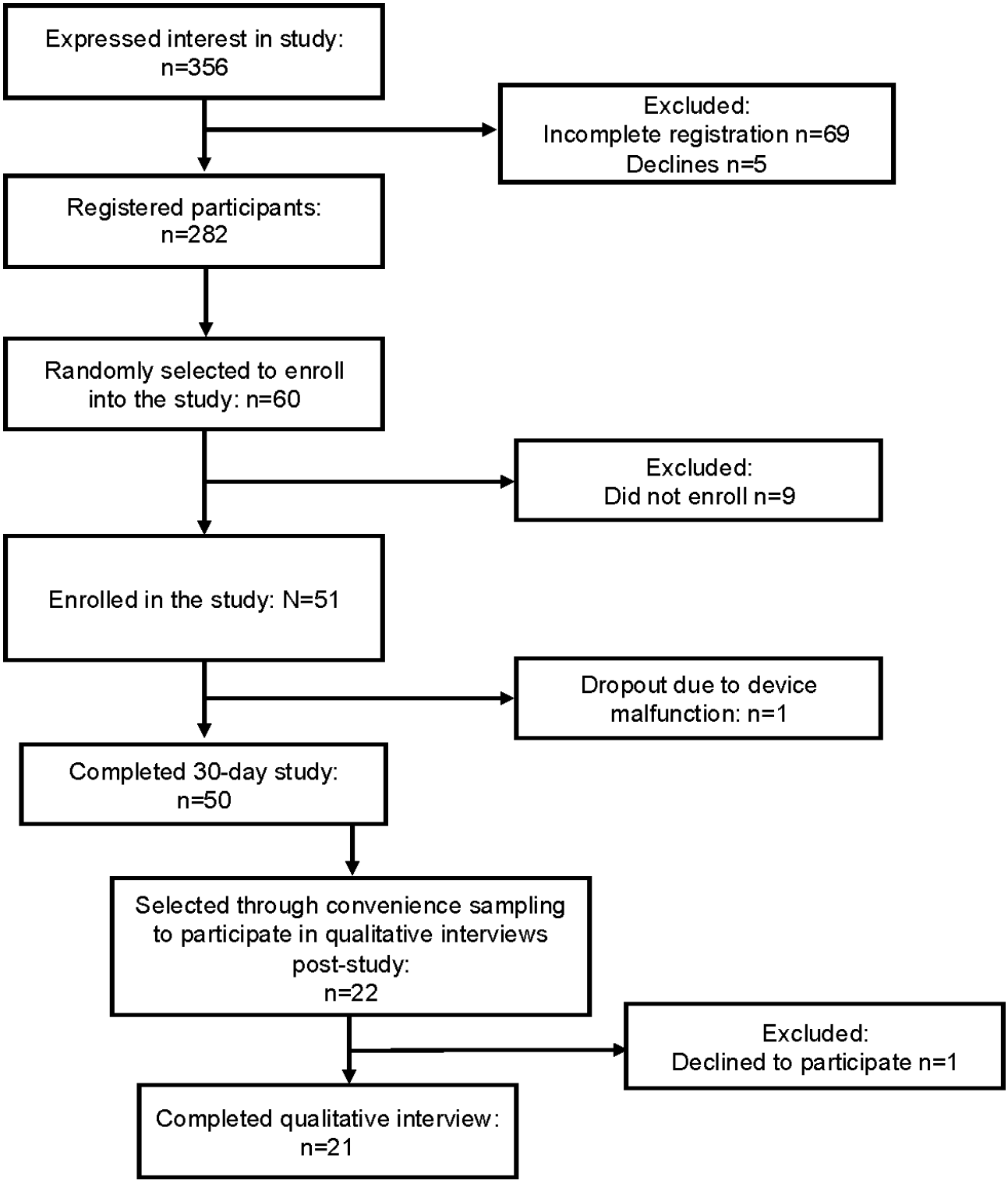
Characteristic	All participants (N=51)	Exit interview participants (n=21)
Age (years), median (IQR)	32 (29 - 36)	33 (28 - 37)
Sex, n (%)		
Male	15 (30.0)	8 (38.1)
Female	36 (70.6)	13 (61.9)
Cadre, n (%)		
Doctors	10 (19.6)	1 (4.8)
Nurses	24 (47.1)	12 (57.1)
Others[b]	17 (33.3)	8 (38.1)
Years working in health care, median (IQR)	8 (5-12)	9 (7-13)
Sexual orientation, n (%)		
Heterosexual	48 (94.1)	20 (95.2)
Other or prefer not to say	3 (5.9)	1 (4.8)
Marital status, n (%)		
Not in a committed relationship	12 (23.5)	7 (35.0)
In a committed relationship	11 (21.6)	1 (5.0)
Married	26 (51.0)	12 (60.0)
Separated, divorced, or widowed	1 (2.0)	0 (0)
Highest education		
Diploma	16 (25.5)	8 (38.1)
Bachelor’s degree	25 (49.0)	10 (47.6)
Master’s degree	7 (13.7)	3 (14.3)
Other[b]	6 (11.8)	0 (0)
Residence		
Rural	1 (2.0)	1 (4.8)
Suburban	4 (7.8)	0
Urban	45 (88.2)	20 (95.2)
Prefer not to say	1 (2.0)	0
Facility		
1	18 (35)	6 (29)
2	15 (29)	9 (43)
3	10 (20)	3 (14)
4	8 (16)	3 (14)

[a]Participants enrolled in a 30-day mixed methods pilot study focused on the feasibility and acceptability of using a mobile application to track survey- and wearable-based data in mental health research among Kenyan health care workers.

[b]Nutritionists, pharmacists, physiotherapists, and psychologists.

XSL•FO
RenderX

Figure 1. Enrollment flow of participants. Health care workers participated in a 30-day mixed methods pilot study to assess the feasibility and acceptability of using a mobile app and wearables to capture survey-based and passively sensed data in Kenya.



Feasibility of Mobile App and Wearable for Data Collection

The median time to complete the baseline questionnaire was 1 (IQR 1 - 2, range 1-11) day, from the date the link was received to submission. All participants (N=51) completed 100% of questions in the Risky Families Questionnaire, Generalized Anxiety Disorder-7, Patient Health Questionnaire-9, and PTSD assessments, while 92% (n=47) and 84.4% (n=42) completed all questions in the Neuroticism, Extraversion, Openness to experience Five-Factor Inventory and World Health Organization Alcohol, Smoking and Substance Involvement Screening Test assessment, respectively.

The wearable was worn for >10 hours on 63% of study days (n=50). Nearly all participants uploaded step (n=50), heart rate (n=46), and sleep (n=45) data within 5 days of receiving the wearable. On average, participants completed 58.4% of the daily mood questions over the 30 days. Data on steps, heart rate, and

XSL•FO
RenderX

sleep were uploaded on 92.6%, 76.6%, and 49.8% of the days, respectively. The average number of daily steps was 8180 (SD 4872).

To further characterize variability in adherence, we examined the median (IQR) and distribution of the number of days each participant submitted data (Table 3 and Multimedia Appendix 1). Approximately half of the participants submitted heart rate (n=26) and step count (n=31) data on all 30 study days, indicating higher adherence for these 2 passively collected metrics (Multimedia Appendix 1). In contrast, submissions for mood and sleep and the number of days with >10 hours of data (ie, wear time) were more variable across participants and across days. Histograms of the number of submission days show a more dispersed pattern for these measures; however, the distributions were not strongly skewed toward a small subset of highly adherent individuals.

Table . Completeness of longitudinal data (n=50)[a,b].

Variable	Days, median (IQR)
Wear time >10 hours	21 (13 - 28)
Steps	30 (28 - 30)
Heart rate	30 (18 - 30)
Sleep	15 (7 - 23)
Mood	18 (13 - 23)

[a]Participants enrolled in a 30-day mixed methods pilot study focused on the feasibility and acceptability of using a mobile app to track survey- and wearable-based data in mental health research among Kenyan health care workers.

[b]Completeness was based on the average number of days over the 30-day pilot that participants wore the wearable for >10 hours; uploaded any step, heart rate, and sleep data; and completed the daily mood question.

Acceptability of Mobile App and Wearable for Data Collection

Theme 1: Experiences With Study Protocols

Qualitative interviews with participants on baseline questionnaires and protocols elicited both negative and positive experiences. The participants had mixed perceptions of the administration of the baseline questionnaire. The positive experience was attributed to the questionnaire's simple language, questions interlinked, making it easy to follow through, and flexible to complete at one's convenience due to the online format, while the challenge was attributed to the questionnaire being lengthy and challenging to complete in one sitting.

> *The questions were simplified so they were easy to understand and of course relate... the questions tied to the study.* [Pharmacist]

> *I think the baseline survey was so long. I didn't finish at once. I'm guessing maybe I started when I was a bit busy, but it was long. I almost gave up.* [Facility 1, male medical officer]

Theme 2: Intrinsic and Extrinsic Motivation

Participants expressed that the wearable and MyDataHelps app significantly increased their awareness of personal health behaviors, such as steps taken, water intake (which was an optional feature), and sleep patterns. The wearable had additional metrics that could be activated by participants, but the study focused on heart rate, sleep, and activity rate. The wearable device encouraged goal setting and helped track progress, providing users with regular reminders and alerts to stay on track.

> *It has come in handy in terms of reminding me how many steps I have done in a day... it updates me when to take my water.* [Physiotherapist]

> *It actually made me cautious about my sleep patterns and sleep hygiene.* [Clinical officer]

Congratulatory messages and goal reminders from the app positively affected participants' motivation. Many felt encouraged by progress notifications, which fostered a sense of accomplishment and supported mental wellness. This motivation extended to family and social contexts, as some participants began to actively engage in physical activities with family members.

> *I can walk, and it tells me congratulations you've done this many during the day apart from checking my mood, it's also checking my health.* [Nurse]

Additionally, the app was reported to facilitate self-monitoring, primarily through mood-tracking and daily health prompts. Participants accepted the mobile app because of its ability to track their mood at the end of each day, which helped them reflect on their mental health.

> *This feature encouraged daily self-reflection, allowing users to evaluate their day and develop an awareness of patterns in their mood.* [Nurse]

The wearable was acceptable based on features such as water resistance, as well as the ability to keep power for a long time. The straps were found to be light and easy to wear throughout the day and night. These features were encouraging for participants to have the wearable on throughout the day and night with ease and comfort.

> *The good thing is it keeps power very well. I charge maybe once in every two weeks.* [Nurse]

> *When I realized it had a function of water lock then I decided that I will not be removing it at all, so when I'm going to shower, I activate water lock, so I have it all the time.*

> *[Nurse]*

XSL•FO
RenderX

Theme 3: Technical and Usability Challenges

Despite the value of wearable devices, 48% (10/21) of the participants reported challenges with wearable design and technical requirements. Continuous Bluetooth connectivity was necessary for data syncing, which participants found inconvenient, especially when forgetting to activate Bluetooth for extended periods, sometimes leading to data collection gaps.

> *The only thing is that if you're not online or not connected with Bluetooth, you might not be able to enter data.* [Nurse]

Participants recommended providing additional reminders and motivational prompts to ensure adherence to the study. Two participants expressed a desire for more color varieties for the straps and to receive a replacement if the wearable broke, as replacements were not provided in this pilot.

Most of the participants noted some discomfort sleeping with the wearable on their wrist to record their sleep patterns in the initial days, but this improved with time.

> *It felt uncomfortable, you had had the watch the whole day, and sometimes at night, I would try but would wake up and remove it(watch) because it just felt uncomfortable.* [Nurse]

> *Then being plastic sometimes just gets sort of uncomfortable. So, you'd want to change from one hand to another.* [Nurse]

Participants encountered technical challenges with the mobile app, primarily regarding the requirement for consistent internet and Bluetooth connectivity. The app required participants to log in daily at a set time, but limited internet data constrained consistent access, which could delay synchronization of data to the mobile app or lead to a missed daily prompt. Participants struggled with the Bluetooth connectivity for synchronization of data and wished that the app could only work with internet access to synchronize.

> *It's only that it won't synchronize data when either your battery is low.* [Nurse]

> *The only thing is that if you're not online or not connected with Bluetooth, you might not be able to enter data* [Nurse]

Recommendations were provided on simplifying the app's connectivity requirements to reduce the dependency on Bluetooth and facilitate prepaid mobile data packages. Suggestions included the possibility of offline functionality for data logging, which could sync automatically when internet access is restored.

> *Bluetooth is necessary for pairing, but I wish it was not required. Just the bundle would have been enough because most of the time you realize you are online on your phone but using Bluetooth is a habit that most of us do not have.* [Clinical officer]

Discussion

Principal Findings

This pilot study supports the feasibility of deploying mental health surveys, collecting data through wearables, and integrating wearable and survey data within a single mobile platform under real-world infrastructure constraints among Kenyan HCWs. Participant retention and completion of the initial questionnaire were high, indicating strong initial engagement and comfort reporting on mental health symptoms. Daily adherence was moderate; approximately half of the participants submitted heart rate and step count data on all 30 study days, while mood, steps, and wear time (≥10 h/d) were more variable across participants and over time. Sleep data were only captured on about half of all study days. Together, these metrics suggest that while continuous daily engagement was not fully achieved, the high retention and collection of multidimensional data support that these technologies are logistically viable in similar LMIC settings. Moreover, participants reported largely positive experiences with the platform, especially appreciating the self-monitoring aspect and feedback features, further supporting the feasibility and acceptability of this approach.

Mental health research in HCWs is critical to identifying and reducing stress, burnout, and psychological harm in those who care for others, which in turn improves worker well-being and patient care. This pilot study revealed that HCWs were willing to provide sensitive mental health information, such as depression, anxiety, posttraumatic stress disorder, and substance use, within a day of enrollment. Studies in Africa have highlighted that the absence of native terminology for mental health and the existence of unwritten nondisclosure norms make it difficult to openly initiate conversations about mental health [32]. Willingness to provide sensitive information on mental health is a critical step in identifying people at risk for mental health disorders. In this pilot, the web-based mode of administration and the preenrollment sensitization process, which helped build rapport between the study team and participants while addressing their concerns and questions about data confidentiality, may have mitigated some stigma-related barriers to disclosure as other studies have suggested [33,34].

The use of wearables in mental health research to infer, monitor, and predict changes in health states over time is an emerging area. On average, studies that have used wearables to collect data among HCWs with small sample sizes or shorter duration of data collection have reported a retention rate of 83% - 98% [35,36], which aligns with the retention rate in this pilot. In our study, participants reported that motivation to know their health information was a contributor to high retention, and other studies have shown motivational support to be a contributor in reducing attrition [37]. In addition, all 51 participants completed the baseline questionnaire, a rate that was 2 times higher than a similar study conducted in the United States where the completion rate was 48% [38]. The differences could be due to the varied recruitment approaches. Unlike the US-based study that used a mainly monetary incentive-driven approach to recruit participants, our study used a multipronged recruitment

XSL•FO
RenderX

approach, including increased physical contact through sensitization, where we had one-on-one discussions with participants as opposed to sharing information via email or website, which could be more impersonal reimbursement for internet connectivity, email, and phone notifications. Similar studies have reported that adopting such a comprehensive approach during recruitment of participants for web-based surveys leads to better engagement [39].

In terms of adherence to daily data collection, our findings are consistent with results from similar digital health studies in both low- and high-income settings. For example, a 5-day pilot in Uganda reported high compliance with wearable use and daily diary completion, despite challenges with comfort during sleep [40]. Similarly, a 3-week study in rural Kenya demonstrated high step and sleep data completeness, but limited heart rate data availability [41]. The wear time observed in our study (63% of days with over 10 h of usage) is within the typical range of 50% - 70% reported in free-living environments [30]. Additionally, mood surveys were completed on two-thirds of the days, which is encouraging given the known challenges of sustaining engagement in digital health monitoring. For example, the average 30-day retention rates for mobile apps in general are less than 6% [42,43]. Collectively, these findings highlight the feasibility of digital monitoring among HCWs in LMICs.

Several barriers likely contributed to moderate adherence to the daily data collection, particularly for sleep and daily mood entries. Participants cited discomfort wearing the device overnight, contributing to reduced sleep data availability. Similar issues have been documented in other wearable studies [40,41] and highlight the importance of comfort for extended wear. Technical challenges with the wearable included instances where the device was accidentally switched from wrist mode to clip mode, which turns off sleep monitoring. Participants were guided on how to check if the wearable was in wrist mode during the initial onboarding to minimize potential data loss. However, this is an important consideration for future studies that leverage wearable devices with the functionality to switch modes. Regarding daily mood, reporting may have been impacted by notification delivery failures. Push notifications relied on continuous internet access, which was not always available. Interestingly, studies reporting higher adherence to daily mood diaries have reported using SMS as opposed to push notifications [44], which may circumvent the need for Wi-Fi. Participants also reported forgetfulness, competing work demands, and survey fatigue, which are consistent with prior research indicating that survey delivery mode, timing, and user workload influence daily self-report adherence [40,45]

Despite these challenges, our findings suggest strong participant engagement. Qualitative feedback revealed largely positive perceptions about using the mobile app and wearable device. Many participants appreciated tracking their health metrics and receiving motivational messages (eg, congratulatory prompts for meeting step count goals), indicating that these features likely enhanced engagement. Future studies that use wearables in mental health research may need to consider that beyond serving as an assessment tool, the wearable itself may also influence participant behavior. Increased awareness of sleep patterns and physical activity may motivate behavior change, which may function as an unintended intervention and potentially confer mental health benefits.

Limitations

The study was conducted on a small sample size, which is appropriate for a pilot study, but limits our ability to detect subgroup or temporal comparisons. In addition, the population was limited to 4 urban hospitals with an overrepresentation of nurses, which may limit the generalizability of the findings to rural areas and other cadres. In addition, the trial duration was only 1 month, which captures feasibility in the short term, but adherence patterns might change over longer periods. It remains unclear whether participants would continue using the app and wearable for an extended period of time (eg, months). Qualitative feedback was obtained from a subset of participants who completed the study; those who had more difficulties might have been less likely to participate in the interviews, potentially leading to a positive bias in qualitative results. Collectively, this pilot focused on the practical ability to deploy surveys, sustain device wear, and synchronize wearable data under real-world infrastructure constraints, rather than on deriving clinically interpretable mental health metrics such as heart rate variability or complete daily physical activity profiles, which were beyond the scope of this pilot. Future studies are warranted to better link wearable-derived measures with mental health outcomes and to evaluate clinically interpretable metrics in larger, more comprehensive studies. Despite these limitations, the study provides valuable preliminary evidence to inform the scaling of similar digital mental health studies in HCWs in LMICs.

Conclusion

This pilot study provides justification for the use of a mobile app and wearable in mental health research in LMICs. The high survey completion rate and study retention, as well as qualitative acceptability, reflect strong engagement and suggest that such technologies are logistically viable in similar contexts. Future studies looking to scale similar approaches should consider enhancing prompts, improving device comfort, and ensuring reliable internet connectivity or finding an alternative solution.

Acknowledgments

We would like to acknowledge the health care facilities and administration staff who provided permission to set up and conduct the study, the Center for Global Health Equity, University of Michigan for their expertise and support in conceptualizing the study design.

XSL•FO
RenderX

The authors used ChatGPT (GPT4.0; OpenAI) to help organize author-written content and improve manuscript structure and flow. No artificial intelligence (AI) tool was used to generate data, perform analyses, interpret findings, or create references.

Funding

Research reported in this publication was supported by the Office of The Director, National Institutes of Health (OD), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the National Institute of Mental Health (NIMH), and the Fogarty International Center (FIC) of the National Institutes of Health under award number U54TW012089 (A Abubakar and AKW). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: AKW, A Abubakar, EF, SS ZM
Data curation: LK, A Aballa, DM, EH, WN, RM, AB, PM, JO, JS, ZW, AKN, LA
Formal analysis: LK, A Aballa, DM, EH, PM, EF, AKN
Funding acquisition: AKW, A Abubakar
Investigation: LK, A Aballa, WN, RM, ZM, LA
Methodology: LK, EH, AKN, LA, AKW, A Abubakar, EF, ZM
Project administration: LK, RM, EH, MKN, AKW, A Abubakar, EF, ZM
Resource: AKN, LA, SS, CK, AKW, A Abubakar, EF, ZM
Supervision: CK, ZW, AKN, LA, AKW, A Abubakar, EF, ZM
Validation: LK, A Aballa, DM, EH, WN, RM, AB, PM, MKN, JO, JS, ZW, AKN, LA, SS, CK, AKW, A Abubakar, EF, ZM
Visualization: LK, A Aballa, EH
Writing original draft: LK, A Aballa, EH, WN, PM, CK, EF
Writing- reviewing, and editing: LK, A Aballa, DM, EH, WN, RM, AB, PM, MKN, JO, JS, ZW, AKN, LA, SS, CK, AKW, A Abubakar, EF, ZM

Conflicts of Interest

None declared.

Multimedia Appendix 1
Distributions of data completeness.
[DOCX File, 327 KB - mhealth_v14i1e77761_app1.docx]

References

1. Alzamanan AM, Almareh NF, Alyami AAM, et al. The role of healthcare workers in modern medicine: a comprehensive review of challenges, contributions, and future directions. J Ecohumanism 2024;3(7). [doi: 10.62754/joe.v3i7.4666]
2. Ahmat A, Okoroafor SC, Kazanga I, et al. The health workforce status in the WHO African Region: findings of a cross-sectional study. BMJ Glob Health 2022 May;7(Suppl 1):e008317. [doi: 10.1136/bmjgh-2021-008317] [Medline: 35675966]
3. Walton M, Murray E, Christian MD. Mental health care for medical staff and affiliated healthcare workers during the COVID-19 pandemic. Eur Heart J Acute Cardiovasc Care 2020 Apr;9(3):241-247. [doi: 10.1177/2048872620922795] [Medline: 32342698]
4. Aymerich C, Pedruzo B, Pérez JL, et al. COVID-19 pandemic effects on health worker's mental health: systematic review and meta-analysis. Eur Psychiatry 2022 Jan 21;65(1):e10. [doi: 10.1192/j.eurpsy.2022.1] [Medline: 35060458]
5. Kwobah EK, Mwangi A, Patel K, Mwogi T, Kiptoo R, Atwoli L. Mental disorders among health care workers at the early phase of COVID-19 pandemic in Kenya; findings of an online descriptive survey. Front Psychiatry 2021;12:665611. [doi: 10.3389/fpsyt.2021.665611] [Medline: 34366913]
6. Jaguga F, Kwobah EK, Mwangi A, et al. Harmful alcohol use among healthcare workers at the beginning of the COVID-19 pandemic in Kenya. Front Psychiatry 2022;13:821610. [doi: 10.3389/fpsyt.2022.821610] [Medline: 35295779]
7. Søvold LE, Naslund JA, Kousoulis AA, et al. Prioritizing the mental health and well-being of healthcare workers: an urgent global public health priority. Front Public Health 2021;9:679397. [doi: 10.3389/fpubh.2021.679397] [Medline: 34026720]
8. Roe D, Mazor Y, Gelkopf M. Patient-reported outcome measurements (PROMs) and provider assessment in mental health: a systematic review of the context of implementation. Int J Qual Health Care 2022 Mar 31;34(Supplement_1):ii28-ii39. [doi: 10.1093/intqhc/mzz084]

XSL•FO
RenderX

9. Fung CH, Hays RD. Prospects and challenges in using patient-reported outcomes in clinical practice. Qual Life Res 2008 Dec;17(10):1297-1302. [doi: 10.1007/s11136-008-9379-5] [Medline: 18709564]
10. Marconcin P, Werneck AO, Peralta M, et al. The association between physical activity and mental health during the first year of the COVID-19 pandemic: a systematic review. BMC Public Health 2022 Feb 1;22(1):209. [doi: 10.1186/s12889-022-12590-6] [Medline: 35101022]
11. Stautland A, Jakobsen P, Fasmer OB, et al. Reduced heart rate variability during mania in a repeated naturalistic observational study. Front Psychiatry 2023;14:1250925. [doi: 10.3389/fpsyt.2023.1250925] [Medline: 37743991]
12. Yang FN, Liu TT, Wang Z. Functional connectome mediates the association between sleep disturbance and mental health in preadolescence: a longitudinal mediation study. Hum Brain Mapp 2022 Apr 15;43(6):2041-2050. [doi: 10.1002/hbm.25772] [Medline: 35040524]
13. Daghlas I, Lane JM, Saxena R, Vetter C. Genetically proxied diurnal preference, sleep timing, and risk of major depressive disorder. JAMA Psychiatry 2021 Aug 1;78(8):903-910. [doi: 10.1001/jamapsychiatry.2021.0959] [Medline: 34037671]
14. Matcham F, Carr E, Meyer N, et al. The relationship between wearable-derived sleep features and relapse in major depressive disorder. J Affect Disord 2024 Oct 15;363:90-98. [doi: 10.1016/j.jad.2024.07.136] [Medline: 39038618]
15. Jacobson NC, Feng B. Digital phenotyping of generalized anxiety disorder: using artificial intelligence to accurately predict symptom severity using wearable sensors in daily life. Transl Psychiatry 2022 Aug 17;12(1):336. [doi: 10.1038/s41398-022-02038-1] [Medline: 35977932]
16. Zhang Y, Stewart C, Ranjan Y, et al. Large-scale digital phenotyping: Identifying depression and anxiety indicators in a general UK population with over 10,000 participants. J Affect Disord 2025 Apr;375:412-422. [doi: 10.1016/j.jad.2025.01.124]
17. Jo YT, Lee SW, Park S, Lee J. Association between heart rate variability metrics from a smartwatch and self-reported depression and anxiety symptoms: a four-week longitudinal study. Front Psychiatry 2024;15:1371946. [doi: 10.3389/fpsyt.2024.1371946] [Medline: 38881544]
18. Horwitz A, Czyz E, Al-Dajani N, et al. Utilizing daily mood diaries and wearable sensor data to predict depression and suicidal ideation among medical interns. J Affect Disord 2022 Sep 15;313:1-7. [doi: 10.1016/j.jad.2022.06.064] [Medline: 35764227]
19. Dai B, Larnyo E, Tetteh EA, Aboagye AK, Musah AAI. Factors affecting caregivers' acceptance of the use of wearable devices by patients with dementia: an extension of the unified theory of acceptance and use of technology model. Am J Alzheimers Dis Other Demen 2020;35:1533317519883493. [doi: 10.1177/1533317519883493] [Medline: 31679390]
20. Maree J, Piontak R, Omwansa T, Shinyekwa I, Njenga K. Developmental uses of mobile phones in Kenya and Uganda. SSRN J 2013. [doi: 10.2139/ssrn.2323429]
21. Barteit S, Boudo V, Ouedraogo A, et al. Feasibility, acceptability and validation of wearable devices for climate change and health research in the low-resource contexts of Burkina Faso and Kenya: study protocol. PLoS One 2021;16(9):e0257170. [doi: 10.1371/journal.pone.0257170] [Medline: 34591893]
22. Thabane L, Hopewell S, Lancaster GA, et al. Erratum to: Methods and processes for development of a CONSORT extension for reporting pilot randomized controlled trials. Pilot Feasibility Stud 2016;2(1):35. [doi: 10.1186/s40814-016-0078-7] [Medline: 27976755]
23. Aballa A, Mwigereri DG, Zhao Z, et al. Depressive symptoms and associated factors among Kenyan health care workers. JAMA Psychiatry 2026 Feb 1;83(2):216-218. [doi: 10.1001/jamapsychiatry.2025.3727] [Medline: 41369956]
24. Murray G, Rawlings D, Allen NB, Trinder J. NEO Five-Factor Inventory scores: psychometric properties in a community sample. Meas Eval Couns Dev 2003 Oct;36(3):140-149. [doi: 10.1080/07481756.2003.11909738]
25. Taylor SE, Lerner JS, Sage RM, et al. Risky Families Questionnaire. PsycTESTS Dataset 2011 Sep 12. [doi: 10.1037/t03616-000]
26. Dhira TA, Rahman MA, Sarker AR, Mehareen J. Validity and reliability of the Generalized Anxiety Disorder-7 (GAD-7) among university students of Bangladesh. PLoS One 2021;16(12):e0261590. [doi: 10.1371/journal.pone.0261590] [Medline: 34914811]
27. Mwangi P, Nyongesa MK, Koot HM, Cuijpers P, Newton CRJC, Abubakar A. Validation of a Swahili version of the 9-item Patient Health Questionnaire (PHQ-9) among adults living with HIV compared to a community sample from Kilifi, Kenya. J Affect Disord Rep 2020 Dec;1:100013. [doi: 10.1016/j.jadr.2020.100013] [Medline: 33313580]
28. Bovin MJ, Marx BP, Weathers FW, et al. Psychometric properties of the PTSD Checklist for Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (PCL-5) in veterans. Psychol Assess 2016 Nov;28(11):1379-1391. [doi: 10.1037/pas0000254] [Medline: 26653052]
29. Humeniuk R, Henry-Edwards S, Ali R, Poznyak V, Monteiro MG, et al. The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): manual for use in primary care. World Health Organization. 2010. URL: https://iris.who.int/handle/10665/44320 [accessed 2025-03-20]
30. Chan A, Chan D, Lee H, Ng CC, Yeo AHL. Reporting adherence, validity and physical activity measures of wearable activity trackers in medical research: a systematic review. Int J Med Inform 2022 Apr;160:104696. [doi: 10.1016/j.ijmedinf.2022.104696] [Medline: 35121356]
31. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

XSL•FO
RenderX

32. Jumbe S, Nyali J, Simbeye M, Zakeyu N, Motshewa G, Pulapa SR. “We do not talk about it”: engaging youth in Malawi to inform adaptation of a mental health literacy intervention. PLoS One 2022;17(3):e0265530. [doi: 10.1371/journal.pone.0265530] [Medline: 35349575]
33. Bidonde J, Meneses-Echavez JF, Hafstad E, Brunborg GS, Bang L. Methods, strategies, and incentives to increase response to mental health surveys among adolescents: a systematic review. BMC Med Res Methodol 2023 Nov 16;23(1):270. [doi: 10.1186/s12874-023-02096-z] [Medline: 37974067]
34. Ericson A, Bonuck K, Green LA, Conry C, Martin JC, Carney PA. Optimizing survey response rates in graduate medical education research studies. Fam Med 2023 May;55(5):304-310. [doi: 10.22454/FamMed.2023.750371] [Medline: 37310674]
35. Mendelsohn D, Despot I, Gooderham PA, Singhal A, Redekop GJ, Toyota BD. Impact of work hours and sleep on well-being and burnout for physicians-in-training: the Resident Activity Tracker Evaluation Study. Med Educ 2019 Mar;53(3):306-315. [doi: 10.1111/medu.13757] [Medline: 30485496]
36. Jevsevar DS, Molloy IB, Gitajn IL, Werth PM. Orthopaedic surgeon physiological indicators of strain as measured by a wearable fitness device. J Am Acad Orthop Surg 2021 Dec 15;29(24):e1378-e1386. [doi: 10.5435/JAAOS-D-21-00078] [Medline: 33999882]
37. Egilsson E, Bjarnason R, Njardvik U. Usage and daily attrition of a smartphone-based health behavior intervention: randomized controlled trial. JMIR Mhealth Uhealth 2023 Jun 26;11:e45414. [doi: 10.2196/45414] [Medline: 37358888]
38. Li SX, Halabi R, Selvarajan R, et al. Recruitment and retention in remote research: learnings from a large, decentralized real-world study. JMIR Form Res 2022 Nov 14;6(11):e40765. [doi: 10.2196/40765] [Medline: 36374539]
39. Meyer VM, Benjamens S, Moumni ME, Lange JFM, Pol RA. Global overview of response rates in patient and health care professional surveys in surgery: a systematic review. Ann Surg 2022 Jan 1;275(1):e75-e81. [doi: 10.1097/SLA.0000000000004078] [Medline: 32649458]
40. Nielsen K, Mobley K, Culbreth R, Palmier J, Nabulya A, Swahn MH. Feasibility and acceptability of wearable devices and daily diaries to assess sleep and other health indicators among young women in the slums of Kampala, Uganda. Digit Health 2024;10:20552076241288754. [doi: 10.1177/20552076241288754] [Medline: 39430703]
41. Matzke I, Huhn S, Koch M, et al. Assessment of heat exposure and health outcomes in rural populations of western Kenya by using wearable devices: observational case study. JMIR Mhealth Uhealth 2024 Jul 4;12:e54669. [doi: 10.2196/54669] [Medline: 38963698]
42. Tafradzhiysk N. Mobile app retention rates (2025). Business of Apps. URL: https://www.businessofapps.com/guide/mobile-app-retention/ [accessed 2026-05-21]
43. Lipschitz JM, Pike CK, Hogan TP, Murphy SA, Burdick KE. The engagement problem: a review of engagement with digital mental health interventions and recommendations for a path forward. Curr Treat Options Psychiatry 2023 Sep;10(3):119-135. [doi: 10.1007/s40501-023-00297-3] [Medline: 38390026]
44. Kalmbach DA, Fang Y, Arnedt JT, et al. Effects of sleep, physical activity, and shift work on daily mood: a prospective mobile monitoring study of medical interns. J Gen Intern Med 2018 Jun;33(6):914-920. [doi: 10.1007/s11606-018-4373-2] [Medline: 29542006]
45. Huhn S, Matzke I, Koch M, et al. Using wearable devices to generate real-world, individual-level data in rural, low-resource contexts in Burkina Faso, Africa: a case study. Front Public Health 2022;10:972177. [doi: 10.3389/fpubh.2022.972177] [Medline: 36249225]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
HCW: health care worker
LMIC: low- and middle-income country
mHealth: mobile health
PTSD: posttraumatic stress disorder

Edited by A Stone; submitted 04.Jun.2025; peer-reviewed by KE Nielsen, M Pulier; revised version received 30.Mar.2026; accepted 31.Mar.2026; published 12.Jun.2026.

Please cite as:
Khakali L, Aballa A, Mwigereri DG, Weinheimer-Haus EM, Njoroge W, Maina R, Bunde A, Musitia P, Nyongesa MK, Orwa J, Shah J, Wu Z, Ngugi AK, Atwoli L, Sen S, Kolars C, Waljee AK, Abubakar A, Frank E, Merali Z
Feasibility and Acceptability of a Mobile App and Wearable Device for Collecting Mental Health Survey and Passively Sensed Data Among Health Care Workers in Kenya: Mixed Methods Pilot Study
JMIR Mhealth Uhealth 2026;14:e77761
URL: https://mhealth.jmir.org/2026/1/e77761
doi:10.2196/77761

XSL•FO
RenderX

© Linda Khakali, Andrew Aballa, Dorcas G Mwigereri, Eileen M Weinheimer-Haus, Willie Njoroge, Rachel Maina, Amos Bunde, Peris Musitia, Moses K Nyongesa, James Orwa, Jasmit Shah, Zhenke Wu, Anthony K Ngugi, Lukoye Atwoli, Srijan Sen, Candace Kolars, Akbar K Waljee, Amina Abubakar, Elena Frank, Zul Merali. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 12.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Feasibility and Acceptability of mPallCare, a Digital Health Intervention for People Living With Advanced Cancer in a Refugee Settlement in Uganda: Mixed Methods Study

Eve Namisango[1], MSc, PhD; Agatha Aduro[2], MSc; William Goodman[2], BSc, MSc, PhD; Shaunna Burke[3], MA, PhD; Raphael Ryabonye[4]; Nickson Mutaasa[4]; Timothy Muyami[4]; Elizabeth Nabirye[5]; Dennis Olodi[6]; Viola Ederu[7]; Bassey Ebenso[2], MBBCh, MPH, PhD; Omolola Salako[8], MBBS; Kehinde Okunade[8], MBBS, MSc; Desiree Azizoddin[9], PsyD; Mhoira Leng[10], MBChB; Karl Lorenz[11], MD, MSHS; Felix Muehlensiepen[12], MPH, PhD, Dr rer medic habil; Richard A Powell[13], MA, MSc; Matthew Allsop[2], BSc, PhD, CPsychol

[1]Digital Health Hub, College of Medicine, University of Lagos, Lagos, Nigeria

[10]Palliative Care in Humanitarian Situations and Emergencies, Kampala, Uganda

[11]Stanford Medicine, Stanford, CA, United States

[12]Center for Health Services Research, Brandenburg Medical School Theodor Fontane, Neuruppin, Germany

[13]Imperial College London, London, United Kingdom

[2]African Palliative Care Association, Kampala, Uganda

[3]Academic Unit of Palliative Care, Leeds Institute of Health Sciences, University of Leeds, Worsley Building, Leeds, United Kingdom

[4]School of Biological Sciences, University of Leeds, Leeds, United Kingdom

[5]International Rescue Committee, Kampala, Uganda

[6]Mulago Palliative Care Unit, Mulago Hospital, Kampala, Uganda

[7]Uganda Cancer Society, Kampala, Uganda

[8]Yumbe Regional Referral Hospital, Yumbe, Uganda

[9]Dana-Farber Cancer Institute, Boston, MA, United States

Corresponding Author:
Matthew Allsop, BSc, PhD, CPsychol
African Palliative Care Association, Kampala, Uganda

Abstract

Background: Palliative care is a key component of comprehensive humanitarian health; yet, access and service capacity remain limited in displacement settings, where fragile health systems struggle to meet the complex needs of people living with advanced illness. Digital health technologies have the potential to enhance the reach and delivery of palliative care; yet, their feasibility and acceptability in humanitarian settings remain underexplored.

Objective: We evaluated the feasibility and acceptability of mPallCare, a mobile health intervention integrating patient-reported symptom and outcome monitoring with a clinician dashboard, to support palliative care delivery in the Bidibidi Refugee Settlement, Uganda.

Methods: A 6-week, uncontrolled, exploratory concurrent mixed methods feasibility study was conducted, involving 32 participants with advanced cancer. Community health workers (ie, village health teams) used the mobile app to document patient-reported symptoms and multidimensional outcomes, which were accessible to clinical teams via a dashboard. Following the use of mPallCare, patient and clinical team participants participated in face-to-face interviews. Data collected via mPallCare were analyzed using descriptive statistics to assess feasibility (ie, compliance with reporting, with a feasibility threshold of ≥65% of scheduled reports), and interview data from a subsample of patient and clinical team participants were analyzed using framework analysis to assess acceptability.

Results: Participants completed 84.9% (163/192) of symptom reports and 59.4% (266/448) of outcome reports, with a combined 67% (429/640) of all scheduled reports completed. A modest decline in engagement with report submissions occurred across the 6-week study period. Commonly reported symptoms included headache (27/32, 84.4%), muscle pain (27/32, 84.4%), and dizziness (26/32, 81.3%). Interview findings indicated strong acceptability among patients and clinicians, who described improved communication, enhanced symptom management, and greater continuity of care. Reported challenges included initial navigation difficulties, limited translation accuracy, and technical synchronization issues. Participants and clinical leaders identified the

XSL•FO
RenderX

potential for integrating mPallCare within Uganda's district health information system to strengthen data use and visibility of palliative care within health reporting structures.

Conclusions: mPallCare is a feasible and acceptable digital health intervention for palliative care in a humanitarian setting. While initial uptake was high, sustaining engagement over time may require simplified reporting processes, enhanced language accessibility, and optimizing the mobile app's connectivity and usability. This feasibility phase highlights key priorities for scale-up, including integration with existing health information systems and adaptation for sustained, equitable use across low-resource and displaced populations.

(JMIR Mhealth Uhealth 2026;14:e73483) doi:10.2196/73483

KEYWORDS

digital health; palliative care; mobile health; humanitarian settings; refugee health

Introduction

Globally, record numbers of people are living in displacement from natural and man-made disasters, leading to an increase in humanitarian and fragile settings [1]. Refugees' and migrants' experiences in their origin country, alongside their migration journey and subsequent entry to a host country, can influence physical and mental health needs [2]. Humanitarian settings are further associated with higher mortality and morbidity than the general population [1,3,4], alongside high levels of mental health conditions that can further negatively affect health outcomes [1]. These conditions, by the nature of their disease course, require long-term engagement through management, promotion, and prevention, alongside treatment, rehabilitation, and an integrated palliative care approach [5]. In practice, however, humanitarian agencies in emergencies often focus on "lifesaving" interventions, with little consideration for preventative, curative, or palliative measures for conditions not regarded as immediately life-threatening [6,7]. This apparent dichotomy in service provision means people living with advancing conditions endure suffering, with some referred to as the "expectant" (ie, expected to die, without any form of palliation or end-of-life care) [7,8]. This directly contravenes humanitarian principles, which explicitly require the prevention and alleviation of human suffering [9]. Moreover, displacement settings of all types have increasingly taken on some form of semipermanence, with marginal prospects for pathways to resolution, leaving millions of people spending many years to decades living in these conditions [10].

There is a crucial need to integrate palliative care into both emergency responses and as part of care delivery for people living in humanitarian and fragile settings [11]. Although the need for health services research in humanitarian settings remains significant [12], a paucity of data remains due to the unique and sometimes dangerous nature of humanitarian aid settings, a lack of familiarity by many academics, and the fact that research is often not a key priority for humanitarian organizations [1]. An emerging research area is the role of digital technologies, which are becoming a pivotal part of humanitarian responses, providing access to often limited and critical support in the context of crises. Digital tools are being increasingly used to address persistent challenges in the humanitarian sector, such as inclusion [13]. For palliative care in low-resource settings, digital technologies can provide a means of augmenting care and providing timely data to inform clinical and policymaker decision-making. To enhance care continuity, innovative digital health technologies can be leveraged in humanitarian crises and displaced populations [14] to support key tenets of palliative care delivery, including information sharing, decision-making, and communication [15]. Digital health approaches are common across palliative care providers [16], can improve service delivery around components such as pain management [17], and are acceptable across key stakeholders that include patients, caregivers, health and care professionals, and policymakers [18]. Digital health approaches, such as remote symptom monitoring, are also increasing across low-resource settings to support palliative care delivery [19].

This study arises from a research program focused on developing digital health for palliative care in sub-Saharan Africa. This study was initiated during the COVID-19 pandemic, when physical distancing was required to reduce transmission and digital approaches were encouraged in humanitarian settings [20]. A consortium of key care delivery actors in Uganda and the wider Africa region was involved, including the African Palliative Care Association (APCA), the United Nations Refugee Agency, the International Rescue Committee, and national and local care delivery organizations supporting palliative care. This study aimed to determine the feasibility and acceptability of a mobile phone app and clinician dashboard to support remote monitoring of the symptoms and outcomes of patients receiving palliative care. We present the findings of this novel digital health approach and lessons from its pilot testing to support palliative care delivery in the Bidibidi Refugee Settlement in Uganda.

Methods

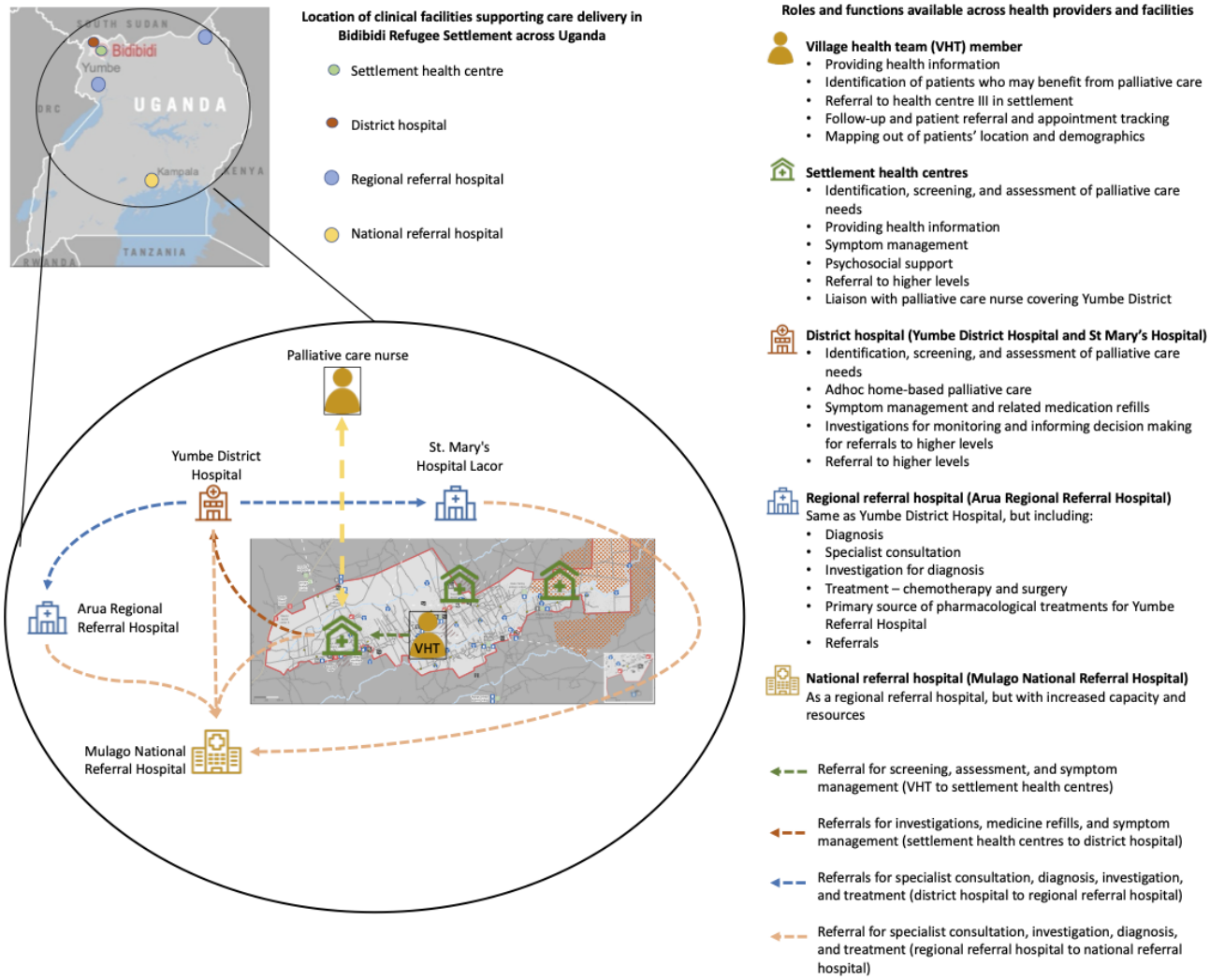
Overview

An uncontrolled, 6-week, exploratory concurrent mixed methods pilot feasibility and acceptability study was conducted, supplementing existing palliative care delivery in the Bidibidi Refugee Settlement. This design enabled complementary insights: quantitative findings identified patterns of engagement, while qualitative data explained underlying reasons and contextual influences. Uganda hosts one of the largest refugee populations globally, hosting more than 1.74 million refugees and asylum-seekers [21]. Most refugees come from South Sudan and the Democratic Republic of the Congo [21]. Bidibidi Refugee Settlement covers more than 250 km^2 and is the largest settlement in the country, hosting more than 246,000 refugees

XSL•FO
RenderX

across 5 areas, or zones, with a large majority of the population comprising women and children. Palliative care delivery within the settlement is currently overseen by one palliative care nurse based at a former district hospital (now regional referral hospital) in the nearest town, Yumbe (Figure 1). The palliative care nurse liaises with village health team (VHT) members and health care providers across the settlement. VHTs comprise community volunteers who help refugees and other community members access health services through, for example, providing health information and serving as a link between the village and health services. VHTs support palliative care delivery by identifying people with palliative care needs in the community and facilitating their access to care, in liaison with facility-based health workers in the settlement and the palliative care nurse. Opioid analgesics, including oral morphine, are available through the clinical teams within the settlement with palliative care nurse oversight. The structure of service delivery and referral relationships is summarized in Figure 1.

Figure 1. A schematic outlining the role of the village health team in identifying and supporting patients receiving palliative care in liaison with a palliative care nurse who supports referrals to a range of services available across various facilities, reflecting provision at the time of the study, with Yumbe District Hospital now a referral hospital with a hospital-based palliative care unit.



Intervention Development

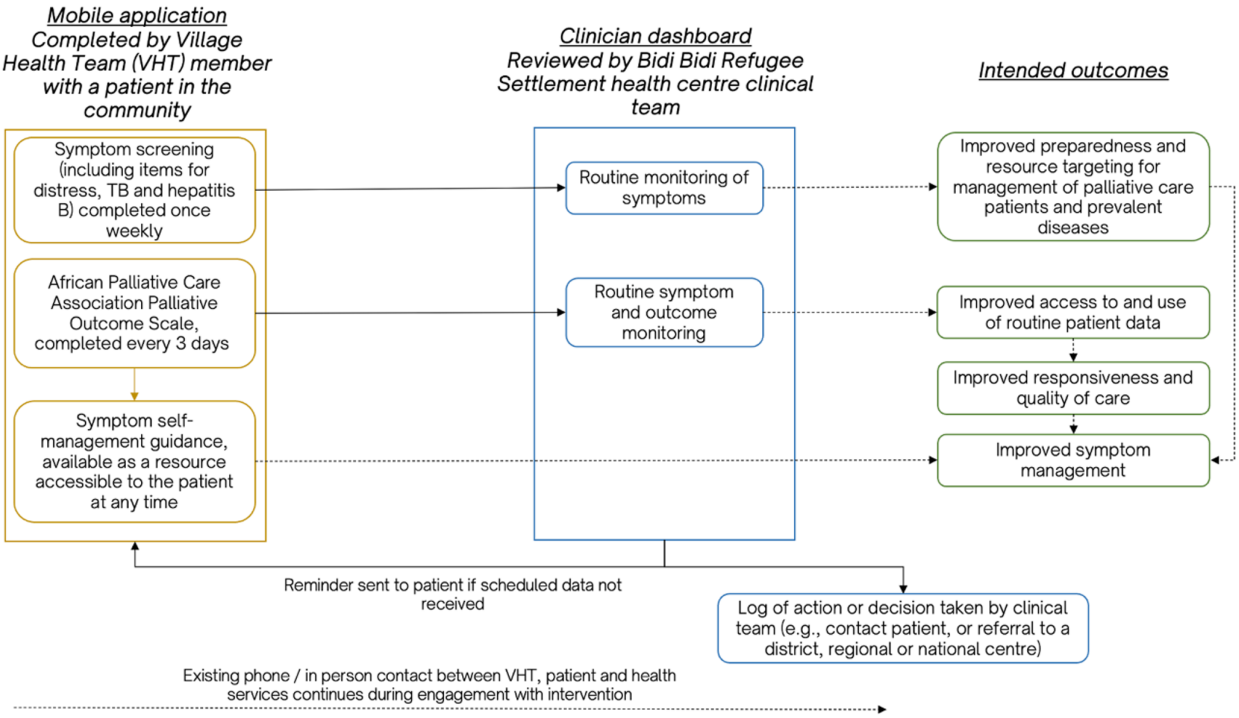
The study is positioned within the first 2 stages of the Medical Research Council Framework for the Development and Testing of Complex Interventions [22], that is, to develop a novel intervention and to determine its feasibility and potential effectiveness in humanitarian settings. The project team included applied health researchers (University of Leeds, Makerere University, and Palliative Care Education and Research Consortium), APCA, palliative care providers (New Life Hospice, Yumbe Regional Referral Hospital Palliative Care Unit, and National South Sudan Palliative Care Association), a patient advocacy organization (the Uganda Cancer Society), and a digital technology company (Universal Doctors).

An initial platform structure using a mobile phone app and clinician dashboard was developed following guidelines for digital health use in palliative care in Africa previously developed by the team (Figure 2) [18]. This included the need to support reporting on physical, psychological, social, spiritual and practical symptoms, the provision of information about living with illness and enabling communication with palliative care services [18]. The resulting platform gathers patient-level symptom and outcome data via the mobile phone app, in addition to providing access to self-management guidance and support (ie, guidance on the management of common symptoms using both video- and text-based content). This approach is underpinned by evidence highlighting that routine symptom

XSL•FO
RenderX

and outcome monitoring and the provision of self-management resources can improve symptom management and quality of care delivery for people with palliative care needs [23,24].

Figure 2. Schematic to show the structure of the mobile phone app and dashboard discussed with key stakeholders, alongside preliminary hypothesized and intended outcomes.



To refine and finalize the digital health approach, an advisory group was formed and consulted. The group was facilitated by the APCA and included clinic staff from New Life Hospice and Yumbe Regional Referral Hospital Palliative Care Unit, Ministry of Health representatives, global health security and infectious disease experts from the Infectious Diseases Institute in Uganda, patient representatives supported by the Uganda Cancer Society, and gender, diversity, and inclusion experts. During a stakeholders workshop in Yumbe District, Uganda, and hosted by the APCA, the following topics were discussed: alignment of the intervention and its content with the refugee population, review and development of symptom and outcome measures that could be collected via the mobile phone app, and developing an implementation plan that built on and strengthened existing activity, while ensuring alignment with the Uganda Health Sector Integrated Refugee Response Plan [25]. The Refugee Response Plan recognizes the need for parallel provision of services across refugee and host communities, highlighting health technologies and health information as key pillars in developing health care for refugee populations. Following the meeting, the intervention was refined, with the symptom checklist expanded to ensure the inclusion of symptoms indicative of tuberculosis and hepatitis B, both prevalent in the settlement. Furthermore, options for the inclusion of symptom and outcome scales were discussed, including which items would be included in the mobile phone app, balancing established protocols for using the tools and existing workflows of VHTs in the community setting.

The mPallCare Intervention

mPallCare comprises 2 components: a mobile phone app to gather patient-reported data that is completed by a VHT member (ie, a volunteer community health worker) from the settlement clinical team, and a web-based clinical dashboard that displays the data gathered from the mobile app for clinical teams (as shown in Figure 2). Together, using the mobile app, a VHT and patient complete 2 forms: a symptom screening questionnaire and a multidimensional outcome measure designed for palliative care in Africa [26]. The symptom screening questionnaire was developed during the advisory group workshop and included symptoms commonly reported by patients receiving palliative care (eg, pain, dyspnea, nausea, and vomiting). Items also included symptoms indicative of COVID-19, 4 symptom screening items for tuberculosis as recommended by the World Health Organization [27], and 2 items related to symptoms that could be suggestive of hepatitis B (ie, clay-colored stools and yellowing of the skin; refer to Multimedia Appendix 1 for symptom screening items). Respondents were asked to respond (yes or no) to a total of 17 items, reporting whether they had been present over the previous week. The multidimensional outcome scale used is the APCA Palliative Outcome Scale (POS) [26], which assesses key palliative care domains using numeric rating scales. It includes 7 patient-focused questions that include the following:

1. Pain: captured using a scale (0=not at all; 5=overwhelmingly)
2. Other symptoms: “Have you been affected by other symptoms (such as nausea, breathlessness, etc)?" (0=not at all; 5=overwhelmingly)

XSL•FO
RenderX

3. Concern: Have you been feeling worried about your illness in the past 3 days? (0=not at all; 5=overwhelming worry)
4. Feelings: Over the past 3 days, have you been able to share how you are feeling with your family or friends? (0=not at all; 5=yes, I've talked freely)
5. Life: Over the past 3 days, have you felt that life was worthwhile? (0=no, not at all; 5=yes, all the time)
6. Peace: Over the past 3 days, have you felt at peace? (0=no, not at all; 5=yes, all the time)
7. Advice: Have you had enough help and advice for your family to plan for the future? (0=not at all; 5=as much as wanted)

Data reported via the mobile app are then displayed on a clinical dashboard, enabling a settlement clinical team to monitor patient symptoms and outcomes routinely, determine any changes or problematic symptoms, and devise an appropriate response plan. This may involve coordinating with external palliative care specialists located beyond the settlements to ensure comprehensive cancer care. The symptom and outcome reports were available in the mobile phone app in 4 languages (English, French, Swahili, and Juba Arabic), developed through forward and backward translation by professional translation services in Uganda. Aligned with Figure 2, this approach seeks to support the timely identification and management of symptoms and issues, as well as inform the coordination of, and referrals to, additional services and clinical sites when needed. This may potentially influence reductions in any symptom burden experienced by patients, as well as increase coordination and care continuity.

mPallCare also includes self-management resources developed by the research team, including an expert in physical activity in advanced disease (SB). Text- and graphic-based self-management content was developed for pain (eg, managing medications, relaxation, and distraction; adapted for the context from earlier work by the team) [28], breathlessness, fatigue and energy, and mobility. Following the submission of symptom and outcome reports using the mobile phone app, self-management related to any reported symptoms is displayed (eg, reporting pain displays pain self-management information). However, patients and health workers could access self-management information for all symptoms at any time as a standalone resource within the mobile phone app. Once reports are submitted, the data are available immediately via the clinic dashboard. The dashboard presents individual lines for each patient report received, with a flag system (ie, a red flag placed next to a patient report on the main dashboard) where any symptom or outcome item has been reported. Any member of the clinical team can review a report submitted by a VHT. On reviewing a report, the clinical team member is asked to detail what action they will take (ie, contact the patient, speak to a colleague, or do nothing), and there is space to add additional notes. This feature was included to enable other clinic team members to understand actions in response to reviewing a report.

A strong focus was placed on the user-friendliness of the graphic interface on the mobile phone app by the software developers. The technical requirements of the mobile phone app were devised around the setting in which its use was intended and made available for an Android smartphone with Android OS 4.0.3 or above. A secure server was set up to receive data reported via the mobile phone app. A web-based backend was developed to allow the clinical team members to log in and view the data submitted. All patient data submitted was deidentified before analysis to protect privacy and confidentiality.

Study Design

The mixed methods study sought to determine the feasibility of collecting palliative care patient-reported outcome data and symptomatology mapping, alongside qualitative investigation of the acceptability of the intervention to target users.

Sample

Participants were selected using a purposive convenience sampling approach, with all eligible patients on existing palliative care lists invited consecutively until the target sample was reached. The most proximate entry point of care for patients receiving palliative care in Bidibidi Refugee Settlement was from the camp-based health facilities via the Arua Regional Referral Hospital. From the hospital, patients are typically referred to New Life Hospice (based in Arua, Uganda) or Yumbe Regional Referral Hospital palliative care unit. We noted that it was common for New Life Hospice to screen camp-based patients from Bidibidi Refugee Settlement who were then referred to Yumbe Regional Referral Hospital palliative care unit, which is closer to their residence. Eligible participants were identified collaboratively with the incumbent health service provider (International Rescue Committee) from existing patient lists of those receiving care, generating an initial list of 32 eligible camp-based patients on the palliative care program from the 2 services, including both refugees and displaced Ugandan nationals. Both refugees and Ugandan nationals were included to reflect the integrated service delivery model operating across settlement and host communities, consistent with Uganda's national policy promoting equitable access to health services. This approach aligns with broader strategic frameworks developed by the United Na Country Team in Uganda and the Ministry of Health to strengthen integration between host and refugee populations and promote peaceful coexistence. This list was continually updated during the recruitment period to account for attrition through patient deaths or relocation from the settlement. The eligibility criteria used to generate the list are outlined in Textbox 1. The final sample of 32 participants reflects all patient participants approached who met the criteria and consented during the recruitment period.

XSL•FO
RenderX

Textbox 1. Eligibility criteria for patient study participants.

Inclusion criteria

- Aged 18 years or older
- South Sudanese or Ugandan nationals
- Receiving hospice care or palliative care from Arua Regional Referral Hospital, Yumbe District Hospital, or New Life Hospice
- Living with advanced cancer, defined as those with metastatic cancer (where, if possible, determined through histological, cytological, or radiological evidence) and/or receiving anticancer therapy with palliative intent
- Estimation of expected prognosis between 3 and 12 months
- Able to read and understand one of the languages available on the mobile phone app (ie, English, French, Swahili, or Juba Arabic) or have a carer who can read and understand one of the languages available
- Able to provide informed consent to participate and have sufficient cognitive ability to provide informed consent

Exclusion criteria

- Patients who are too ill to endure study procedures
- Dyads with visual impairments who are unable to read screen-based information
- Patients who may wish to return to their home countries over the next 3 months

Procedure

Before the commencement of patient enrollment, a training session was delivered for VHTs in May 2022. VHTs were provided with training manuals developed to support navigation and use of the mobile phone app, and an in-person training session took place. The training was attended by representatives from the regional referral hospital, the district health office, the Ministry of Health of Uganda and the Office of the United Nations High Commissioner for Refugees to ensure awareness of the project and commitment to its delivery. The care team for the settlement approached potential participants and generated a list of people who agreed to participate in the study. VHTs supporting participants across the different zones in the settlement attended the training session ahead of the study commencement. In total, 20 VHTs were trained on the study objectives, ethical principles (confidentiality and informed consent), procedures for administering symptom and outcome tools, and step-by-step use of the mPallCare app (research team of the University of Leeds, African Palliative Care Association, and UniversalDoctor) and dashboard. An IT specialist from the APCA led practical sessions that included simulated data entry (ie, using the app for submitting the patient outcome and symptom reports), troubleshooting exercises, and accessing self-management symptom guidance. From the training, it was highlighted that there was a risk that devices provided for the study could be at risk of being stolen, so they were retained in health centers rather than VHT homes when they were not being used. VHTs also requested a paper copy of the self-management resources, which were provided for the study. During the training, the initial list of identified patients was discussed and confirmed in partnership with representatives from Yumbe District Hospital, the International Rescue Committee, Arua Regional Referral Hospital, and New Life Hospice. Initial testing focused on advanced cancer, a common disease group in populations in refugee settlements in Uganda, alongside chronic hepatitis B, diabetes, severe injuries, and sickle cell anemia [29].

Following training, VHTs supported the identification and recruitment of eligible patients. Active recruitment and enrollment of patients took place between August and October 2022. VHTs worked with clinical teams to review health facility patient lists to identify those who met the eligibility criteria. Where eligible patients were identified, a VHT discussed the study, what participation entailed, potential risks and benefits of participation, and the process of data flow from the mobile phone app to the clinic, including where data are stored. If the patient agreed to participate, consent was obtained with a signature or thumbprint. We sought to recruit between 30 and 35 patients receiving palliative care, aligned with sample size requirements for feasibility studies [30]. For those who enrolled in the study, a baseline demographic questionnaire was completed capturing a participant's age, gender, highest level of education, home location (zone and block or TIN number), primary diagnosis, family size, Demographic and Health Surveys measure of socioeconomic status [31], and whether the patient had access to a mobile phone (as a proxy for familiarity with digital devices and to determine the future feasibility of patient-reported data). The Eastern Cooperative Oncology Group Scale of Performance Status [32] is also captured at this point, which describes a person's functional status based on their capacity for self-care, engagement in daily activities, and physical capabilities such as walking and working. A grade is provided from 0 (fully active, able to carry on all predisease performance without restriction) to 5 (dead), with graduated levels of function between the two values. Responses to the baseline demographic questionnaire were recorded by VHTs using a purpose-designed Microsoft Excel spreadsheet with logic and consistency checks to reduce errors.

Following the capture of sociodemographic information, the VHT worked through both the outcome and symptom reports with a patient. Following initial completion, it was expected that a VHT would complete 1 symptom report per week for each patient over the 6-week study period (6 reports×32 participants=192 expected symptom reports) and 1 outcome

XSL•FO
RenderX

report every 3 days (approximately 14 reports×32 participants=448 expected outcome reports). In total, 640 reports were planned across the study period. VHTs visited patients every 3 days to complete and update the outcome report, and once weekly to complete the symptom screening report for up to 6 weeks. During the study period, an IT specialist from the APCA monitored registrations of new patient participants to ensure active recruitment and responded to any queries related to device functionality and data synchronization.

Following the completion of 6 weeks of reporting, clinical officers from Bidibidi Refugee Settlement who were members of the research team (RR, NM, or TM; all male) conducted semistructured face-to-face interviews with patient participants at their homes. The semistructured format allowed for consistent coverage of key areas—such as patient participants' experience of completing reports with VHTs, exploring the perception of how data would be used, and any perceived influence it had on the care they received. Interviewers remained flexible around the presence of family members and caregivers during interviews. An interpretivist paradigm was deemed the most suitable approach to the qualitative component of the study, due to the nature of the exploration of feasibility and acceptability and the necessity of understanding individual experiences of mPallCare from the viewpoint of those who interacted and used the platform. This perspective emphasizes the significance of subjective interpretations, personal perceptions, and the meanings individuals attribute to their experiences [33].

A purposive sampling strategy was used to ensure representation of patient participants with diverse demographic and clinical characteristics relevant to the study aims, drawn purposively from the same feasibility cohort. We sought to capture variation across age and diagnoses (ie, different cancer types) in a subsample of 10 patients. We sought to understand their experience of the mobile phone app being used to obtain symptom and outcome data. Alongside patient interviews, clinical staff and VHTs involved at each site (n=10) were purposively sampled by site (geographic zones within the Bidibidi Refugee Settlement) and role (ie, VHT, clinical officer, and nurse) and participated in face-to-face interviews to explore their views of how the app had supported patient care and the experience of its implementation in routine care. Our sampling approach did not seek to achieve data saturation but rather capture a breadth of perspectives that could provide practical insights into user experience and inform future implementation and scale-up. This approach focused on identifying variation and key issues influencing acceptability and feasibility. Interviews were conducted by a female member of the research team with specialist palliative care clinical expertise and undertaking an MSc in Palliative Medicine (LN). Patient and clinical staff participants were not known to the research team. The research team members were fluent in the languages used during interviews. All interviews were audio-recorded and transcribed verbatim in the language used, then translated into English where required by the researcher who conducted the interview and included in parentheses alongside the original transcription. Transcripts were then checked against audio files by a second member of the research team involved in conducting interviews to ensure accuracy in the translation and transcription.

Analysis

All statistical analyses were conducted in SPSS (version 28.0; IBM Corp). Descriptive statistics were run on the demographic and clinical characteristics of the individuals who completed outcome and symptom reports. Compliance with the outcome and symptom reporting protocol was reported in terms of completed reports. Graphical representations were developed to chart app usage over the 6 weeks. Individual mean scores were calculated, as well as a grand mean for each outcome variable. The change in the outcome measures of Concern, Feelings, Life, Peace, and Advice was charted over time. Feasibility is defined as participants completing ≥65% of scheduled reports, reflecting a pragmatic, context-appropriate target informed by prior studies reporting completion rates between 60% and 80% in studies adopting digital health approaches in the collection of patient-reported outcomes for palliative care populations: for example, completion rates of 58% for weekly patient-reported online questionnaires [34], 73% for mobile phone–based data capture [35], and 78.4% of outcomes data via a patient portal [36]. While there are no comparable studies in the context of humanitarian settings, the conservative threshold reflects broader digital health implementation challenges common in the context of humanitarian settings [37], including availability, access, affordability, and rights. Acceptability of the intervention was explored through both patient and clinical team interviews on the experience of using the mPallCare over the study period.

Interview transcripts from patient and clinical staff were imported into NVivo 12 (QSR International) for framework analysis [38]. This approach guided a structured, transparent, and systematic comparison across participant groups while remaining grounded in participants' accounts. The framework approach also enabled the organization of data around predefined domains of interest (ie, usability, communication, and integration with care) while permitting the inductive identification of emergent themes reflecting user experience and contextual factors. MA and AA analyzed the interview transcripts. Separate analyses were undertaken for transcripts derived from patient interviews and those completed with clinical team staff and VHTs. For both sets of data, the same approach was adopted. A coding frame was developed using both deductive and inductive approaches, with an initial framework developed from a selection of 4 transcripts. For both, these were selected to ensure variation in the clinical sites and participants (for patients, varying by sex and age, and for clinical team members by role). The developed frameworks were discussed and, where necessary, adapted by the wider team, after which the remainder of the transcripts were analyzed. Where differences or uncertainty arose, these were discussed with the wider team. Comparative analysis in the patient and clinical team frameworks enabled common themes to be identified, as well as sample divergences. Illustrative quotes are reported for each theme, alongside the study participant ID and site, to demonstrate reporting from across the sample breadth.

Ethical Considerations

The protocol, interview guide, and consent materials were reviewed and approved by the Hospice Africa Uganda Research

XSL•FO
RenderX

Ethics Committee (reference HAU-2021‐4) and complied with the principles outlined in the Declaration of Helsinki. All study participants received verbal and written information about the study objectives, procedures, potential risks, and their right to withdraw at any time without consequence to their care. For participants with limited literacy, consent was read aloud in their preferred language and consent was recorded through signature or thumbprint. All data were deidentified prior to analysis, stored on encrypted servers, and accessible only to the research team. No financial compensation was provided to participants. The paper has been compiled in accordance with the COREQ (Consolidated Criteria for Reporting Qualitative Research; Checklist 1), ensuring transparency and completeness in the reporting of study design, conduct, and analysis.

Results

Overview

In total, 32 participants with advanced cancer took part, with the majority of these being female (n=19, 59.4%), Christian (n=19, 59.4%), and married (n=17, 53.1%). The ages of the participants ranged from 18 to 77 years, with an average of 48.5 (SD 16.9) years. The primary caregiver was most often their spouse (n=16, 50%), but children and parents were also commonly reported. Most participants also did not have access to a mobile phone (n=18, 56.3%). An initial list of participants comprising camp-based patients on the palliative care program included additional participants to accommodate challenges in recruitment. Participants initially identified on the list included those who died (n=6) and those who left the settlement (n=4) before study commencement. A breakdown of the demographic and clinical characteristics of the participants can be found in Table 1.

Table . Demographic and clinical characteristics (N=32).

Variables	Values
Age (y), mean (SD)	48.5 (16.9)
ECOG[a] functional performance, mean (SD)	1.6 (0.8)
Family size, mean (SD)	9.2 (2.9)
Nationality, n (%)	
South Sudanese	13 (40.6)
Ugandan	19 (59.4)
Sex, n (%)	
Male	13 (40.6)
Female	19 (59.4)
Religion, n (%)	
Christian	19 (59.4)
Muslim	13 (40.6)
Marital status, n (%)	
Married	17 (53.1)
Single, divorced or separated, or widowed	15 (46.9)
Highest level of education, n (%)	
None	22 (68.8)
Primary or secondary	10 (31.3)
Caregiver relationship, n (%)	
Spouse	16 (50.0)
Child	10 (31.3)
Parent	6 (18.8)
Access to mobile phone, n (%)	
No	18 (56.3)
Yes	14 (43.8)

[a]ECOG: Eastern Cooperative Oncology Group.

XSL•FO
RenderX

Outcome and Symptom Report Data

All patient participants (N=32) completed symptom and outcome reports. From a total of 192 possible symptom reports (32 participants × 6 time points), participants completed 163 reports (84.9%), an average of 5.1 reports per participant, with a range of 2 to 11 reports. From a total of 448 possible outcome reports (32 participants × 14 time points), participants completed 266 reports (59.4%), an average of 8.3 reports per person, with a range of 3 to 13 reports. In total, a combined 67% (429/640) of all scheduled reports were completed.

Figure 3 presents the frequency of responses from participants over the course of the study, which shows that as the study progressed, responses declined. Detailed findings from patient and clinical team interviews are provided below. Decline in engagement was reported relating to practical and technical barriers. Early difficulties with navigation and translation occasionally delayed data entry, and intermittent connectivity and dashboard errors resulted in missing or repeated reports. VHTs also described periods when device availability or competing health outreach duties limited the frequency of visits to patients. Collectively, operational factors may have contributed to the reduced number of submissions observed toward the end of the study.

Figure 3. Frequency of responses submitted via the patient mobile phone app across the 6-week study period.

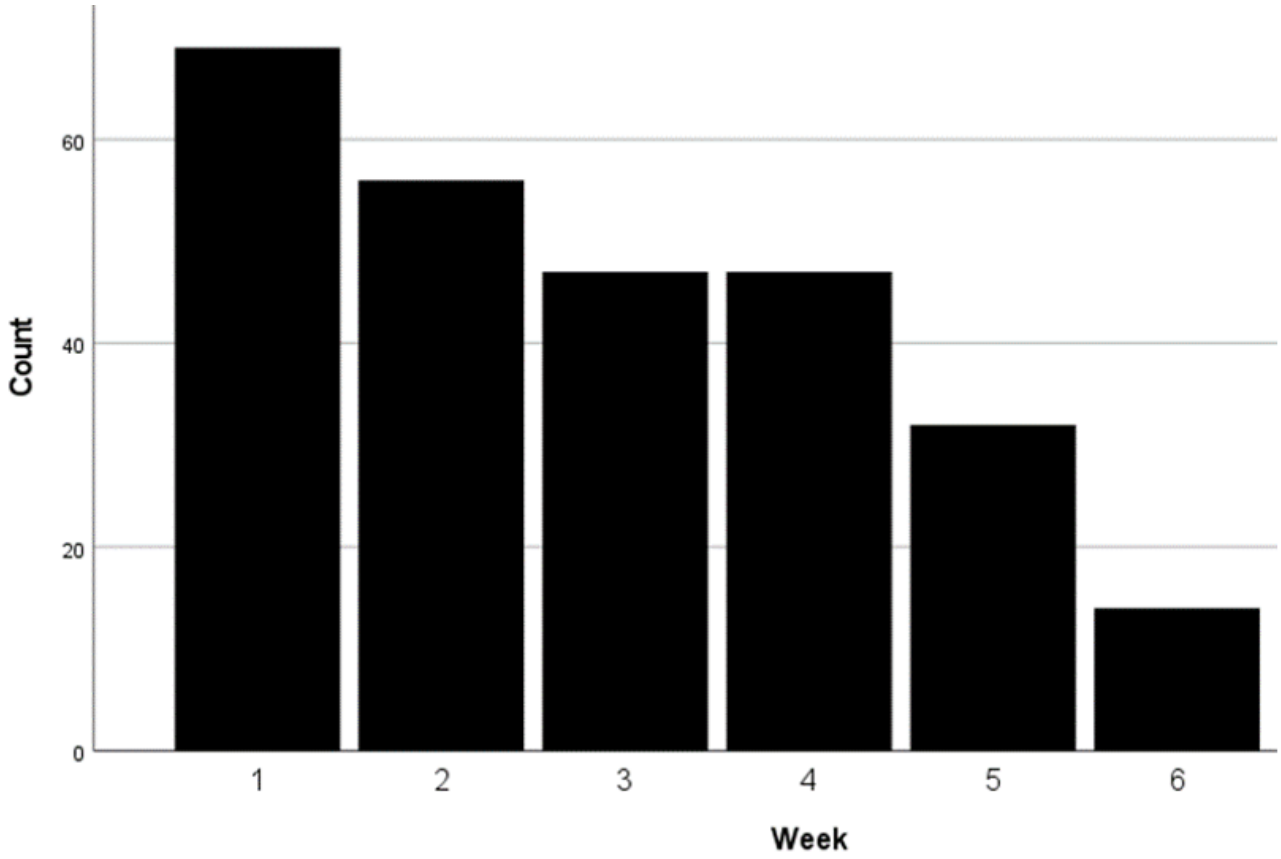
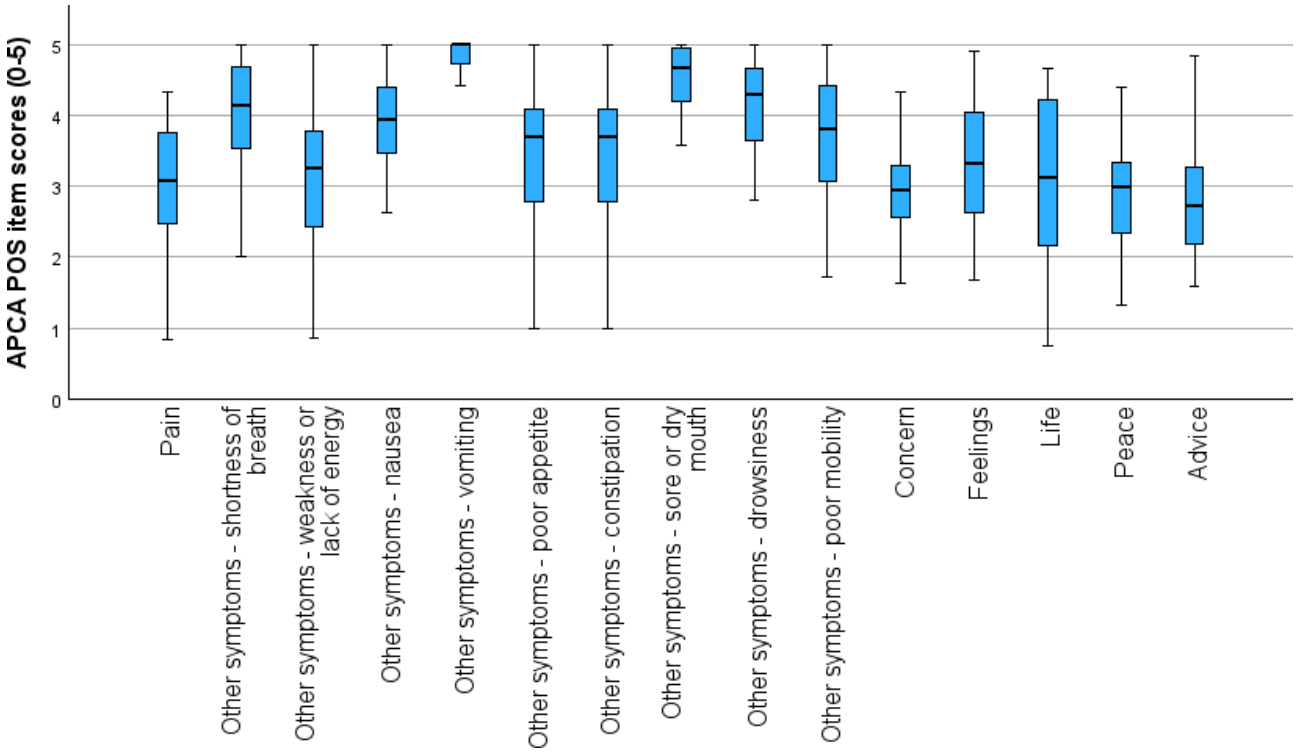


Figure 4 presents the individual mean scores of each outcome measure across the study period. The symptoms of vomiting, sore or dry mouth, and drowsiness had the highest scores, whereas the symptoms of weakness or lack of energy, constipation, and poor appetite had the lowest symptom scores, although these also had a wider range of scores. The scores for the measures for Feelings, Concern, Life, Peace, and Advice were clustered around 3.0. However, Figure 5 suggests that each of the outcomes of Pain, Feelings, Life, Peace, and Advice improves throughout the study but that Concerns around their illness increased over the study. However, these graphs also indicate there is substantial variability in individual scores, with some individual scores decreasing or increasing over time.

XSL•FO
RenderX

Figure 4. Boxplots showing the mean and range of scores of participants across the outcome measures, which include pain, other symptoms (shortness of breath, weakness or lack of energy, nausea, vomiting, poor appetite, constipation, sore or dry mouth, drowsiness, and poor mobility), Concern, Feelings, Life, Peace, and Advice. Higher scores indicate a more positive outcome and lower scores a more negative outcome. APCA: African Palliative Care Association; POS: Palliative Outcome Scale.



XSL•FO
RenderX

Figure 5. Changes in participants' self-reported scores on the African Palliative Care Association Palliative Outcome Scale domains over the study period, with each gray line representing an individual participant's fitted trajectory of change in scores over time (in days). The blue line represents the overall fitted trajectory for the sample. Higher scores indicate a more positive outcome in each domain (A=Pain, B=Concern, C=Feelings, D=Life, E=Peace, F=Advice). APCA: African Palliative Care Association; POS: Palliative Outcome Scale.

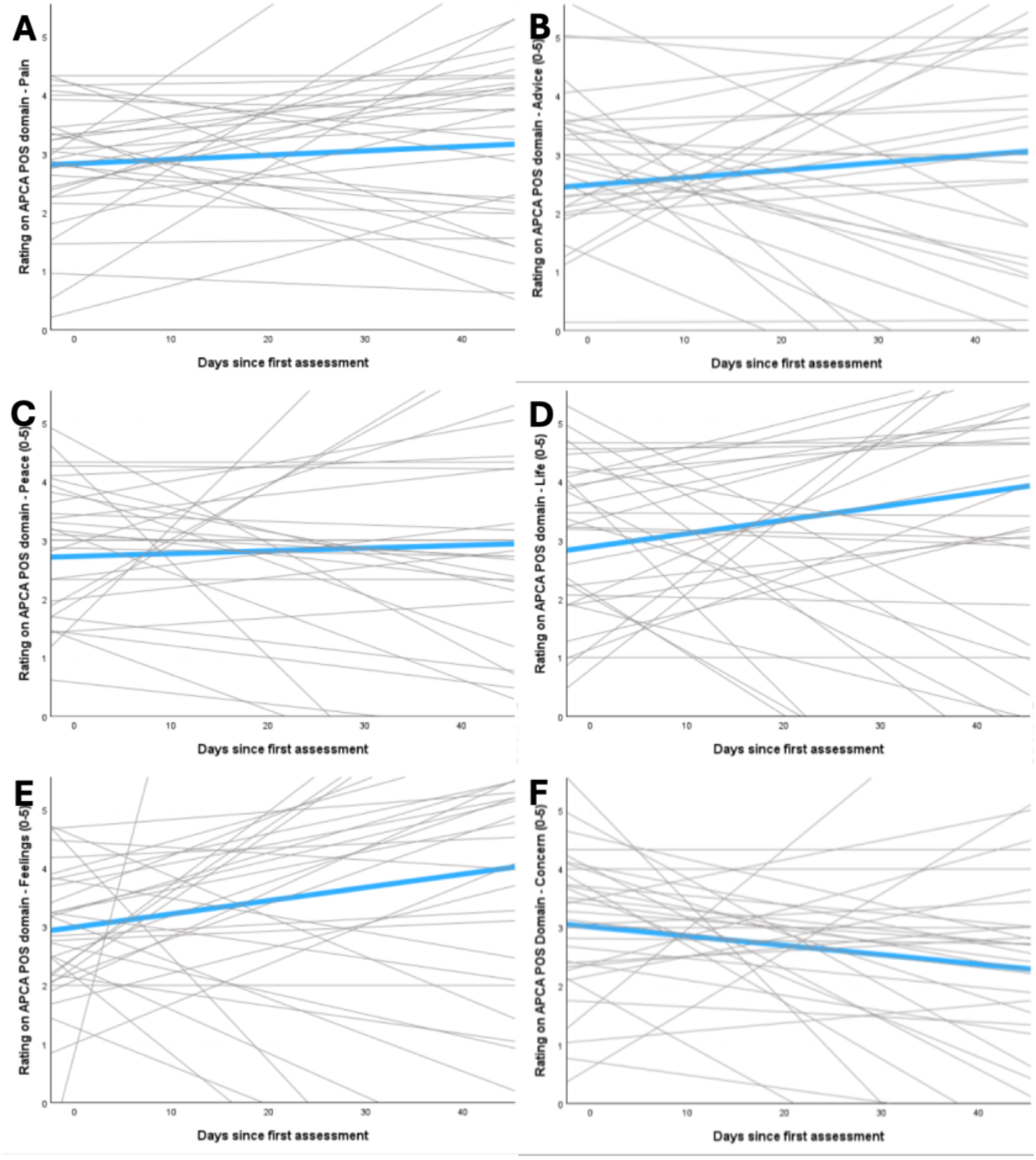


Table 2 outlines the symptoms reported by participants over the course of the study. Across all participants, the most commonly reported symptoms were headache, muscle pain, and dizziness; the least reported were colored stools, yellowing skin, and diarrhea, with fatigue not reported at all. The most reported symptoms over the course of the study were headaches, chest pain, weight loss, and abdominal pain, with diarrhea and foot blisters being the least commonly reported.

XSL•FO
RenderX

Table . Symptoms reported by participants.

Symptom	Number of participants reporting symptom (%)	Number of times symptom reported	Of those reporting symptom, average number reported
Headache	27 (84.4)	90	3.3
Muscle pain	27 (84.4)	73	2.7
Dizziness	26 (81.3)	76	2.9
Chills	25 (78.1)	59	2.4
Weight loss	24 (75.0)	75	3.1
Chest pain	24 (75.0)	75	3.1
Loss of appetite	24 (75.0)	69	2.9
Night sweats	22 (68.8)	35	1.6
Abdominal pain	22 (68.8)	68	3.1
Fever	20 (62.5)	40	2.0
Shortness of breath	19 (59.4)	46	2.4
Confusion	18 (56.3)	46	2.6
Cough	17 (53.1)	38	2.2
Eye soreness	14 (43.8)	32	2.3
Loss of smell	13 (40.6)	25	1.9
Hoarse voice	13 (40.6)	24	1.9
Painful throat	10 (31.3)	17	1.7
Nausea and vomiting	9 (28.1)	16	1.8
Welts or swelling	9 (28.1)	16	1.8
Prolonged cough	6 (18.8)	13	2.2
Feet blisters	6 (18.8)	8	1.3
Diarrhea	3 (9.4)	4	1.3
Colored stools	2 (6.3)	3	1.5
Yellowing skin	2 (6.3)	5	2.5
Fatigue	0 (0.0%)	--	--

Interviews With Patients and Clinical Teams

In total, 20 qualitative interviews were conducted; 10 with patient participants and 10 with members of the clinical team, including VHT members, nurses, and clinical officers. From the patient interviews, three key themes were generated from the analysis: (1) ease of use and perception of app usefulness, (2) understanding of the questions and perception of data use, and (3) impact of the intervention on health and care. From health professional interviews, three themes were generated from the analysis, namely (1) perception of impact of the intervention on patient care, (2) experiences of implementation, and (3) data robustness and reporting. Representative quotations are presented within the Results to illustrate key findings within each theme. The duration of interviews ranged from 20 minutes to 1 hour and 33 minutes for patient participants and 18 to 44 minutes for clinical team interviews. Exemplar quotes from patient and clinical team interviews are presented in Table 3.

XSL•FO
RenderX

Table . Exemplar quotes to accompany the narrative of findings from patient and clinical team interviews.

Theme	Supporting quotes
Patient interview findings	
Ease of use and perception of app usefulness	• Quote 1: “The fact that it was my first encounter with a smartphone, I found it challenging in the beginning, however with time with the help of the VHT I was able to somehow learn how to use the phone.” (Patient participant 4, female 77, cervical cancer) • Quote 2: “...to be able to communicate better to a wider population especially those who don’t understand English, picture illustrations would be added to make people understand better.” (Patient participant 2, female 40, cancer of the eye) • Quote 3: “I found some difficulties, especially with some of the words which I couldn’t as well understand. However, with the help of the VHT, they were interpreted. My suggestion is that the app should be changed to a language that we usually use here, for example, Juba Arabic and Bari to make it simpler.” (Patient participant 2, female 40, cancer of the eye)
Understanding of the questions and perception of data use	• Quote 4: “...some other things were okay to use that method and others were hard, for example, say, if say I’m worried, : “I found some difficulties, especially with some of the words which I couldn’t as well understand. However, with the help of the VHT, they were interpreted. My suggestion is that the app should be changed to a language that we usually use here, for example, Juba Arabic and Bari to make it simpler.” (Patient participant 2, female 40, cancer of the eye) • Quote 5: “...On about two occasions I received a call from the doctors saying that I needed to be reviewed, this made me think that they’d based on what I’d sent them to make that decision, hence, according to me, the scores when seen by the doctors were used make decisions on how I felt. I was told that they looked at the scores and indeed they had meaning, green, yellow and red; all had different meanings. The scores that I had sent at that time were yellow, I was told yellow was a sign that my sickness was increasing and so, there was need to review me so that I don’t enter red or get worse.” (Patient participant 2, female 40, cancer of the eye)

XSL•FO
RenderX

Theme	Supporting quotes
Impact of the intervention on health and care	• Quote 6: "...I began becoming open to the doctors and felt it easy to tell me my plight as I felt it, the fear of the doctors went away while using the app, I could say that the communication gap really went away during this time." (Patient participant 6, male 40, colorectal cancer) • Quote 7: "For sure, during that period, the VHT used to visit me frequently, sometimes 3 or 4 times to check on me, that made me feel closer to him and know him more. During those times, drugs used to come on time, and it became hard to miss as the doctors were always in close contact with me especially when I had not been able to reach the health center. Even the times when the doctor calls you to the health center, I used not to delay he knows how I feel so I just go come back in no time." (Patient participant 4, female 77, cervical cancer) • Quote 8: "It has created a strong bond between me and my doctors. For example, I am better understood in the health facility because I'm constantly monitored by the doctors. This as well led to me not delaying in the facility because the doctors already know how I feel even before I get there, this saves me the trouble of explaining myself every time when I come to the facility." (Patient Participant 10, female 40, breast cancer) • Quote 9: "There's that part that talked about how to manage a difficulty in breathing especially when working. When I followed the advice, the app gave me on how to control my breathing, that really improved my symptoms, and I was able to even become more productive. The symptom management further helped me realize that I can live with the disease which initially I thought it was a death sentence to me." (Patient participant 10, female 40, breast cancer) • Quote 10: "There was a question that accessed my mental health status, especially how I related with family members, if I talk to them and how close I was, this reminded me to always talk to them in case I'm not feeling okay or worried." (Patient Participant 2, female 40, cancer of the eye)
Clinical team interview findings	

XSL•FO
RenderX

Theme	Supporting quotes
Perception of impact of the intervention on patient care	• Quote 11: "I think I mentioned before we were able to stay in touch with patients even before their appointments; before we would only interact with these patients on their appointments right now they are being seen twice so the frequency with which you're interacting with the patient and progress is more by about 4 times instead of a one month review, you can review your patient like 4 times a month; a patient is interacting with you more not only on their treatments and what but also on other issues like psycho-socio, you're visiting them so definitely the connection was increased." (Health team participant 2, clinical officer) • Quote 12: "...it changed somehow because it was, I was actually now dealing more with, relying on what the VHT tells me and you know sometimes this information as you were asking whether they benefitted the patients to some extent you would wish this patient talks directly and you listen from this patient so that you get the quality of the information." (Health team participant 3, palliative care nurse)
Experiences of implementation	• Quote 13: "The design of the app was good but the challenge I found there was the severity of the sickness, like slightly, moderate severe, more severe those things in the beginning, translating it in the local language was a challenge. But now I myself I started making research from my colleagues; what do you mean by moderate in our language, severe and slightly and after coming up with the meaning of those word in our local language, I would then translate to the patients, then it became familiar but, in the beginning, these were a bit challenging." (Health team participant 8, village health team member) • Quote 14: "That challenge came with one of the patients; the patient was started on the program from the dashboard he just disappeared from the app, from the VHT side the information is there on the app again they told me that was a technical problem otherwise that thing of VHT not submitting a report; No. they have submitted the reports." (Health team participant 3, palliative care nurse)
Data robustness and reporting	• Quote 15: "Then the option, on the option of the actions to do for the patient depending on the assessment, if we could broaden it a bit. We had where either you go speak to the client as a professional, then you are going to call and then others; so, in the others is where you would put referrals. I think if we could have like referrals captured separately or linkage captured separately so that patients are linked to other sources of service like psycho-socio support; others were linked to protection; others were linked to livelihood so if we could have where we can capture this then it would be great for the design." (Health team participant 8, village health team member) • Quote 16: "...if we can along the way if you can find a way of having it reflect some elements of the DHIS too in the information system so that we can actually have this data fed into our reporting into the ministry as well. I know that will take improvements on both sides; an improvement onto updating and making our data collection for palliative care in the Ministry of Health system also up to date because right now it also looks shallow; it is also shallow so maybe that will take two sides so that when you have this patient's data you can easily extract it to inform the district and the Ministry of Health in terms of programs." (Health team participant 2, clinical officer)

Patient Interview Findings

Theme 1: Ease of Use and Perception of App Usefulness

The mobile phone app was accessed with the help of the VHT, in keeping with the intended use. Initial difficulties in navigating the app and understanding the input of data were easily overcome with support from a VHT and with repeated use (Quote 1 in Table 3). Some participants reported that self-management resources could be enhanced through the inclusion of visual aids such as pictures, as well as improving the format and presentation of information (ie, increasing font size and wider availability of accessibility options, such as pictorial and audio formats; Quote 2 in Table 3). All participants found the mobile phone app useful and particularly appreciated that symptom and outcome items were easy to understand, requiring only minimal explanations from the VHT, and relevant to their condition (Quote 3 in Table 3). Some participants indicated that translation into additional local languages may also improve the app.

XSL•FO
RenderX

Theme 2: Understanding of the Questions and Perception of Data Use

Participants generally agreed that the questions were easy to understand and context-appropriate, addressing their most troubling symptoms, such as pain, as well as what participants perceived as often-overlooked mental health concerns. Although the symptom and outcome items were easy to understand, the process of providing numeric ratings was unfamiliar and reported as difficult by some participants (Quote 4 in Table 3). However, participants reported confidence that data gathered via the mobile app would be accessed and used by clinical teams. This was reinforced when submitted reports were referred to by clinicians in subsequent meetings when discussing symptom management (Quote 5 in Table 3).

Theme 3: Impact of the Intervention on Health and Care

Some participants reported the intervention made respondents feel closer and more connected to their health care teams, making them feel more comfortable sharing symptoms and concerns with them (Quote 6 in Table 3). Participants also felt the app made the clinicians more knowledgeable about them as individuals, as well as their unique conditions. This was perceived as translating to quicker decision-making and improved access to medication and treatment (Quotes 7 and 8 in Table 3). The use of the app also contributed to patients' agency, as they felt better informed and empowered to self-manage symptoms and to communicate the same with caregivers, family members, and the hospital team, where necessary. Some participants reported beneficial outcomes from reviewing and applying the self-management guidance (Quote 9 in Table 3). The intervention was found to have supported a focus on the mental health and well-being of participants. Some participants reported that having easily accessible information about their health condition helped to allay worries and anxieties and made them feel that help was available if required. The mobile phone app was reported to have fostered better interactions between patients, their caregivers and health workers, enabling them to recognize the need to seek help and receive it in a timely manner (Quote 10 in Table 3).

Clinical Team Interview Findings

Theme 1: Perception of Impact of the Intervention on Patient Care

All participants from clinical teams reported that the intervention was beneficial to patient care. Participants reported feeling more connected with patients and valued the ability to assess changes promptly in a patient's condition that could enable a timely response (Quote 11 in Table 3). However, the routine reporting by VHTs for some participants meant feeling more detached from patients, having instead to rely on reports provided by VHTs (Quote 12 in Table 3).

Theme 2: Experiences of Implementation

A period of adaptation was required for VHTs when using the mobile phone app, particularly for determining appropriate words for translating symptom and outcome items into local languages (Quote 13 in Table 3). This was overcome in collaboration with colleagues, with increased understanding of items by patients. The platform experienced some technical challenges during the study, with some participants reporting that reports submitted to the platform could not be viewed (Quote 14 in Table 3). In some instances, this led to delays and extra costs relating to patient care, as the VHTs often had to be called in to verify information or repeat the visit to the patient.

Theme 3: Data Robustness and Reporting

The clinical team members reviewing reports felt that additional options could be added to indicate which action was taken in response to reviewing a report. This included options such as being able to capture interaction or activity with other service providers. This was seen as a way of improving the usefulness of the dashboard (Quote 15 in Table 3). Some participants queried the wider use and linkage of data with existing government health information systems. Future integration with district health information systems was also viewed as useful, enabling the integration of palliative care data in existing processes for the Ministry of Health of Uganda (Quote 16 in Table 3). Alongside considering integration, a few participants raised concerns about data privacy and the safeguarding of personal data.

Discussion

Principal Findings

The findings from this study highlight the feasibility and acceptability of mPallCare, a digital health intervention for symptom monitoring and outcome reporting among patients accessing palliative care in a refugee settlement in Uganda. Initial high engagement with symptom and outcome reporting occurred, though engagement declined modestly over the study period. Initial challenges relating to navigation and translation of content, intermittent technical issues, connectivity limitations and competing community health duties may have contributed to this decline. Commonly reported symptoms included headaches, muscle pain, and dizziness. While there was varied completion of outcome assessments, overall trends for the study cohort suggested improvements in outcomes relating to pain, ability to share feelings, life feeling worthwhile, feeling at peace (ie, an item relating to aspects of perception of self and world, relationship to others and spiritual beliefs) and accessing help and advice. Patients reported that the intervention fostered a stronger connection with VHT volunteers and clinical teams, improved symptom management, and enhanced patient empowerment. However, challenges such as initial difficulties with navigating the content of the mobile phone app, translation issues, and occasional technical problems—most commonly data synchronization errors between the VHT app and the clinician dashboard—were identified. Health care professionals reported that the intervention facilitated more timely clinical decision-making, though reliance on intermediary reporting by VHTs sometimes limited direct patient interaction. The initial exploration of the feasibility and acceptability of the intervention suggests that it may have the potential to improve patient-centered care, but considerations for accessibility, training, and integration with existing health services remain crucial. While there were no clear differences in feasibility or overall acceptability across patient subgroups, variations in

XSL•FO
RenderX

experience were noted between clinical roles. For example, nurses emphasized how reliance on intermediary reporting by VHTs occasionally limited direct engagement with patients, while VHTs described increased confidence and improved communication skills over time. These differences suggest the importance of exploring the delineation of complementary roles and feedback mechanisms between VHTs and clinical staff during future implementation.

In supporting the systematic collection of structured, patient-reported data, the intensive collection of longitudinal data by mPallCare can help to determine the temporal progression of patient-reported outcomes and better address individuals' changing needs. For example, data can directly inform decision-making by care teams, alongside supporting preparation for community visits (eg, ensuring adequate medical supplies to support identified issues with symptom management). The approach also provides a means of undertaking both comparative effectiveness research and quality assessment relating to service delivery. This aligns with the research imperative to develop the evidence base underpinning care delivery in the context of humanitarian emergencies and crises [39] and builds on emerging applications of digital technologies, such as virtual training for palliative care [6] and more general development of approaches, including electronic health records [40]. Routinely captured data, such as symptom and outcomes data gathered via mPallCare, also provides opportunities to support policymaker decision-making, providing quality and timely data that can inform planning, decision-making and resource allocation in low-resource settings [41,42]. Investment in digital approaches such as mPallCare can support the collection of useful, relevant and reliable data that can help to characterize the current delivery of palliative care and guide the development of effective health services [43] across humanitarian emergencies and crises. This can ensure that the delivery of care is data-driven and reflects the needs of the population [44].

Having established the feasibility and acceptability of mPallCare, the platform is currently being evaluated across 20 additional clinical teams in Uganda and additional countries in Africa, including both refugee settlements and clinical care facilities outside refugee settlements. The initial development and testing of mPallCare reported in this manuscript highlighted multiple technical and user considerations for its future development. The design and accessibility of the intervention continue to be refined, including increasing the number of languages available (including local languages such as Bari, Ma'di, and Lingala, which have been incorporated) and the ease with which additional languages can be added. Such consideration of the accessibility and contextual relevance of a digital health intervention is essential to facilitate user engagement [45]. The decline in outcome reporting, alongside a slight increase in reported outcomes for most items, requires further research. The decline in outcome reporting over the 6-week study period warrants further reflection. Several factors may have contributed to this pattern. First, disease progression and declining physical function among participants with advanced illness likely limited their ability to engage consistently toward later data collection points. Similar declines in engagement have been observed in other mobile health interventions for people with serious illness, where symptom burden and reduced energy constrain sustained participation [35,46]. Second, competing priorities and workload among VHT members, who were responsible for data collection across multiple community programs, may have reduced the frequency of visits and the submission of reports. Third, technology fatigue and waning motivation once the initial novelty subsides—commonly noted in digital health studies—may have contributed to attrition, especially where perceived direct benefits to users were limited [47,48]. Intermittent network connectivity and temporary dashboard access problems also presented practical barriers. These findings mirror broader evidence from low-resource and humanitarian settings, where infrastructure limitations, connectivity challenges, and high clinical workloads hinder sustained digital engagement [37,49]. Future iterations of mPallCare will explore ways to mitigate these barriers, such as streamlined reporting schedules, automated reminders, stronger feedback mechanisms between clinicians and community teams, and increased user feedback on clinical action taken in response to submitted data.

In terms of the observed upward trends in Feelings, Peace, and Life, formal minimal clinically important differences (MCIDs) for the APCA POS have not been established. The palliative patient-reported outcome measures literature recommends interpreting change using anchor- and distribution-based approaches and explicitly estimating MCIDs for context-specific use [26,50,51]. Recent work for related instruments (eg, the Integrated Palliative Care Outcome Scale) is beginning to define MCIDs at the domain level (eg, dyspnea [52]) but comparable thresholds are not yet available for APCA POS. Accordingly, our findings should be interpreted as directional evidence consistent with expected benefits of structured symptom and outcome monitoring and regular follow-up—namely, improvements in psychological and existential well-being—rather than definitive proof of clinically important change. Future studies could prespecify MCID estimation for APCA POS (eg, using patient global impression of change anchors or distribution-based rules such as the SD of baseline or change scores), to enable a formal determination of clinical significance in humanitarian and low-resource settings.

Sustainability of digital health interventions in humanitarian settings remains a key challenge. The long-term viability of mPallCare may depend on factors critical to scale-up and sustainability, including ongoing technical support, training of community health workers, reliable internet access, and integration with national information systems to ensure interoperability and avoid duplication [53]. Sustaining engagement also requires ongoing supervision and feedback loops between VHTs and clinical teams, supported by consistent funding for device upkeep, data storage, and connectivity. These resource requirements mirror lessons from other digital health programs, where interventions that became embedded within Ministry of Health systems, supported by local champions and recurrent budget lines, were most likely to persist beyond pilot phases [54]. Successful integration will further depend on addressing key operational enablers and constraints. These include ensuring interoperability with existing digital systems

XSL•FO
RenderX

such as District Health Information System 2 (DHIS2), maintaining compliance with national data protection policies, and sustaining Ministry of Health engagement for oversight and governance. Establishing clear data sharing agreements and local technical stewardship will be essential for scaling mPallCare responsibly and sustainably. In Uganda, mPallCare continues to be developed in partnership with the Ministry of Health of Uganda to ensure alignment with priorities for palliative care and refugee health, while exploring routes to integrate mPallCare with existing health information systems. This reflects both stakeholder enthusiasm and a feasible next step identified during qualitative interviews, where several clinical participants suggested integration with the national DHIS2 to strengthen data use and visibility of palliative care within national reporting structures. Furthermore, future work will evaluate the impact of mPallCare on patient-level outcomes, including symptom burden, quality of life, and timeliness of clinical review. This may require an approach such as a pragmatic, longitudinal mixed methods study design that incorporates repeated measures of patient-reported outcomes and clinician response times, with potential progression to a cluster randomized controlled trial following optimization.

Although evidence on digital health to support palliative care in humanitarian settings remains limited, our findings resonate with emerging work demonstrating how technology can strengthen communication and continuity of care in crisis contexts. In Lebanon, a mobile health tool for noncommunicable disease management among Syrian refugees improved documentation and patient satisfaction, though uptake was constrained by fragmented systems and parallel reporting [55]. Similarly, virtual learning networks have been found to build palliative care capacity among humanitarian health workers in Bangladesh, yet implementation was hindered by unstable connectivity and high staff turnover [6]. These studies, like the present one, highlight both the promise and fragility of digital innovation in settings characterized by workforce pressures and infrastructure limitations. This tension extends to telemedicine initiatives [56], which can be constrained by ad hoc funding and weak evaluation frameworks; yet evaluation is essential to avoid entrenching inequities [57]. mPallCare contributes to this body of evidence, illustrating that digital platforms can feasibly support palliative care when coupled with community engagement and integration within existing systems. Sustainability, however, remains central as most digital health projects undertaken in low- and middle-income countries falter beyond pilot phases [53]. For mPallCare, longevity may depend on embedding within national infrastructures such as DHIS2, maintaining technical and workforce capacity, and ensuring stable financing for maintenance and connectivity.

As mPallCare continues to develop, ethical considerations will be central to sustainability given its role in the context of often severe resource constraints. Routine symptom and outcome monitoring may highlight numerous issues for which a clear service response is not available. For example, outcome items relating to advice, concerns and being at peace were all scored relatively low by participants, yet there are many challenges in responding to these issues, such as accessing psychological support in the context of refugee settlements [58]. This creates a moral tension between the humanitarian principles of beneficence and nonmaleficence in resource-constrained contexts, where raising awareness of unmet need risks compounding distress if appropriate support cannot be mobilized [8,11]. During the study, interim mitigation was achieved through escalation pathways: VHTs could refer urgent clinical issues directly to the regional palliative care nurse, who provided intermittent outreach and psychosocial support. However, these mechanisms were limited by staffing and distance, highlighting the structural constraints of delivering equitable palliative care in humanitarian settings [11]. Future phases of mPallCare may address this ethical challenge through aligning system-level digital monitoring and response capacity. This could include, for example, developing simplified triage algorithms that prioritize symptoms or concerns requiring immediate action. Embedding such approaches within broader capacity-building efforts—particularly training frontline workers in basic psychological first aid and ensuring referral linkages for complex cases—will be essential to ensure that data collection translates into meaningful and safe care responses. Recognizing such ethical tensions, strategies to mitigate unmet needs are being incorporated into ongoing work. These include establishing partnerships with existing psychosocial and mental health programs operating within refugee settlements, embedding referral pathways into the digital platform, and curating evidence-informed digital self-help materials accessible directly through the mobile app. In addition, collaboration with the Ministry of Health aims to ensure that symptom and outcome data inform broader service planning, enabling scarce resources to be targeted toward the most pressing needs. Such measures seek to balance the ethical imperative of monitoring suffering with a commitment to providing or facilitating meaningful forms of support.

Implementation strategies and their evaluation form an essential element of digital health intervention development [59]. Continued evaluation to optimize how mPallCare can influence service responses within highly constrained resources will be essential to optimize its use in care delivery, alongside further understanding its role in enabling access to contextually appropriate self-management support. Additionally, the influence of different models of care delivery, staffing, local and broader policies, and case mix of patients beyond people living with advanced cancer can be explored through future scale-up and evaluation. Findings from Bidibidi Refugee Settlement provide valuable insights, but they will not be directly generalizable to all humanitarian settings. Contextual factors such as population stability, digital infrastructure, language diversity, and existing palliative care provision will influence implementation elsewhere. Nevertheless, the core principles of participatory design, health worker–led data collection, and integration with existing health systems may be broadly transferable to other displacement contexts. Furthermore, this is a small, uncontrolled feasibility study conducted over a 6-week period; findings are descriptive and not intended to infer efficacy. The limited sample size and reliance on VHT-facilitated reporting may have influenced patient responses. Furthermore, we used an uncontrolled, 6-week feasibility study, where findings are descriptive and not intended to infer efficacy or long-term sustainability. The short follow-up

XSL•FO
RenderX

period and absence of a control group limit the generalizability of results beyond the study setting and timeframe. Future research using longitudinal or controlled designs will be required to evaluate sustainability and clinical impact. Despite these limitations, the study provides essential early evidence to guide refinement and larger-scale evaluation of mPallCare.

Conclusion

This study underscores the feasibility of mPallCare, a palliative care–focused digital health intervention evaluated in the context of a refugee settlement in a low-resource setting. Routine symptom and outcome data collection was feasible, and key areas for further development were identified. The accessibility of mPallCare continues to be developed, including adding additional languages and planning for integration with existing health information systems. While this initial evaluation suggests mPallCare may foster improved patient-provider communication and support symptom management, challenges such as declining engagement over time, technical issues, and balancing the roles of VHTs and clinical teams must be addressed to optimize its implementation in clinical care. Ongoing adaptation and refinement will be critical to ensure mPallCare can support care delivery across diverse humanitarian and low-resource settings. Future research should continue to ensure continued evaluation throughout the ongoing scaling of mPallCare while maintaining a focus on usability and alignment with clinical workflows, including accommodating differing levels of staffing and resources to respond to identified symptoms and outcomes. The future success of mPallCare may be determined by its ability to support varied services and resource constraints, align with national health information systems, and effectively support patient-centered decision-making for palliative care delivery in the context of humanitarian and low-resource settings.

Acknowledgments

The authors thank the African Palliative Care Association (APCA), the Uganda Ministry of Health, and Village Health Teams (VHTs) for their collaboration and support throughout this study. We also extend our thanks to the humanitarian and clinical partners, including the International Rescue Committee, Yumbe Regional Referral Hospital and New Life Hospice. Generative artificial intelligence tools were not used in the preparation of this manuscript.

The authors thank the African Palliative Care Association (APCA), the Uganda Ministry of Health, and Village Health Teams (VHTs) for their collaboration and support throughout this study. We also extend our thanks to the humanitarian and clinical partners, including the International Rescue Committee, Yumbe Regional Referral Hospital and New Life Hospice. Special thanks also to Nickson Mutaasa, who provided help and support in the organisation and delivery of the research project.

Funding

MA is supported by Yorkshire Cancer Research University Academic Fellowship funding. RAP was supported by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration Northwest London. The views expressed are those of the author and not necessarily those of the NIHR or the Department of Health and Social Care. Funders had had no role in study design, data collection, analysis, or manuscript preparation.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to ethical restrictions, as permission to use data from the study for secondary analyses and other purposes was not sought from participants but are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: MA
Methodology: EN (lead), AA, WG, BE, MA
Data curation: MA, WG
Formal analysis: AA, WG, FM, RAP, MA
Investigation: EN (lead), AA, WG, SB, RR, NM, TM, DO, VE
Resources: EN, SB
Validation: RR, NM, TM, EN, DO, VE, ML
Supervision: ML, KL, MA
Project administration: MA
Funding acquisition: EN, MA
Writing – original draft: EN (lead), AA, MA
Writing – review & editing: EN, AA, WG, SB, RR, NM, TM, DO, VE, BE, OS, KO, DA, ML, KL, FM, RAP, MA

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Symptom screening items used in mPallCare.
[DOCX File, 17 KB - mhealth_v14i1e73483_app1.docx]

Checklist 1
COREQ checklist.
[DOCX File, 26 KB - mhealth_v14i1e73483_app2.docx]

References

1. Kohrt BA, Mistry AS, Anand N, Beecroft B, Nuwayhid I. Health research in humanitarian crises: an urgent global imperative. BMJ Glob Health 2019;4(6):e001870. [doi: 10.1136/bmjgh-2019-001870] [Medline: 31798999]
2. Refugee and migrant health. World Health Organization. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/refugee-and-migrant-health [accessed 2026-04-29]
3. Aebischer Perone S, Martinez E, du Mortier S, et al. Non-communicable diseases in humanitarian settings: ten essential questions. Confl Health 2017;11(1):17. [doi: 10.1186/s13031-017-0119-8] [Medline: 28932259]
4. Amara AH, Aljunid SM. Noncommunicable diseases among urban refugees and asylum-seekers in developing countries: a neglected health care need. Global Health 2014 Apr 3;10(1):24. [doi: 10.1186/1744-8603-10-24] [Medline: 24708876]
5. Munday D, Kanth V, Khristi S, Grant L. Integrated management of non-communicable diseases in low-income settings: palliative care, primary care and community health synergies. BMJ Support Palliat Care 2019 Dec;9(4):e32. [doi: 10.1136/bmjspcare-2018-001579] [Medline: 30389694]
6. Doherty M, Lynch-Godrei A, Azad T, et al. Using virtual learning to develop palliative care skills among humanitarian health workers in the Rohingya refugee response in Bangladesh. J Med Educ Curric Dev 2022;9:23821205221096099. [doi: 10.1177/23821205221096099] [Medline: 36032810]
7. Powell RA, Schwartz L, Nouvet E, et al. Palliative care in humanitarian crises: always something to offer. Lancet 2017 Apr;389(10078):1498-1499. [doi: 10.1016/S0140-6736(17)30978-9]
8. Krakauer EL, Daubman BR, Aloudat T. Integrating palliative care and symptom relief into responses to humanitarian crises. Med J Aust 2019 Sep;211(5):201-203. [doi: 10.5694/mja2.50295] [Medline: 31373022]
9. Daubman BR, Khan F, Slater SE, Krakauer EL. Save lives and relieve suffering: the twin imperatives of humanitarian response and the role of palliative care. Front Oncol 2023;13:1120380. [doi: 10.3389/fonc.2023.1120380] [Medline: 36937419]
10. Forced migration or displacement. Migration Data Portal. 2024. URL: https://www.migrationdataportal.org/themes/forced-migration-or-displacement [accessed 2026-04-29]
11. Schwartz L, Nouvet E, de Laat S, et al. Aid when “there is nothing left to offer”: experiences of palliative care and palliative care needs in humanitarian crises. PLOS Glob Public Health 2023;3(2):e0001306. [doi: 10.1371/journal.pgph.0001306] [Medline: 36962993]
12. Bozorgmehr K, Gold AW. Access to, continuity and coordination of healthcare for refugees: emerging challenges and topics for health services research. In: Wensing M, Ullrich C, editors. Foundations of Health Services Research: Principles, Methods, and Topics: Springer International Publishing; 2023:271-285. [doi: 10.1007/978-3-031-29998-8_22]
13. Bryant J. Digital technologies and inclusion in humanitarian response. : ODI; 2022 URL: https://media.odi.org/documents/Digital_inclusion_synthesis.pdf [accessed 2026-04-30]
14. Shrestha A, Alawa J, Ashworth H, Essar MY. Innovation is needed in creating electronic health records for humanitarian crises and displaced populations. Front Digit Health 2022;4:939168. [doi: 10.3389/fdgth.2022.939168] [Medline: 36300131]
15. Finucane AM, O’Donnell H, Lugton J, Gibson-Watt T, Swenson C, Pagliari C. Digital health interventions in palliative care: a systematic meta-review. NPJ Digit Med 2021 Apr 6;4(1):64. [doi: 10.1038/s41746-021-00430-7] [Medline: 33824407]
16. Allsop MJ, Namisango E, Powell RA. A survey of mobile phone use in the provision of palliative care services in the African region and priorities for future development. J Telemed Telecare 2019 May;25(4):230-240. [doi: 10.1177/1357633X18767187] [Medline: 29635966]
17. Namisango E, Ntege C, Luyirika EBK, Kiyange F, Allsop MJ. Strengthening pharmaceutical systems for palliative care services in resource limited settings: piloting a mHealth application across a rural and urban setting in Uganda. BMC Palliat Care 2016 Feb 19;15(1):20. [doi: 10.1186/s12904-016-0092-9] [Medline: 26895882]
18. Nkhoma KB, Ebenso B, Akeju D, et al. Stakeholder perspectives and requirements to guide the development of digital technology for palliative cancer services: a multi-country, cross-sectional, qualitative study in Nigeria, Uganda and Zimbabwe. BMC Palliat Care 2021 Jan 4;20(1):4. [doi: 10.1186/s12904-020-00694-y] [Medline: 33397321]
19. Salako O, Enyi A, Miesfeldt S, et al. Remote symptom monitoring to enhance the delivery of palliative cancer care in low-resource settings: emerging approaches from Africa. Int J Environ Res Public Health 2023 Dec 17;20(24):7190. [doi: 10.3390/ijerph20247190] [Medline: 38131741]
20. Allsop MJ, El Zakhem A, Garrigue N, Namisango E, Osman H, Piriz G, et al. Palliative care in the COVID-19 pandemic briefing note: the role of telemedicine: remote access to patients, caregivers and health workers. : International Association

XSL•FO
RenderX

of Hospice and Palliative Care; 2020 URL: https://www.researchgate.net/publication/349698 Palliative Care in the COVID-19 Pandemic Briefing Note The role of telemedicine Remote access to patient care and health workers?channel=doi&linkId=5d8f429f1f67d72b84ow_file=true [accessed 2026-04-30]
21. Refugees and nationals by district. UNHCR. URL: https://data.unhcr.org/en/country/uga [accessed 2026-05-22]
22. Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061. [doi: 10.1136/bmj.n2061] [Medline: 34593508]
23. ElMokhallalati Y, Mulvey MR, Bennett MI. Interventions to support self-management in cancer pain. PR9 2018;3(6):e690. [doi: 10.1097/PR9.0000000000000690]
24. Wittich L, Tsatsaronis C, Kuklinski D, et al. Patient-reported outcome measures as an intervention: a comprehensive overview of systematic reviews on the effects of feedback. Value Health 2024 Oct;27(10):1436-1453. [doi: 10.1016/j.jval.2024.05.013] [Medline: 38843978]
25. Health sector integrated refugee response plan. : Ministry of Health of Uganda; 2019 URL: https://globalcompactrefugees.org/sites/default/files/2019-12/Uganda-%20Health%20sector%20integrated%20refugee%20response%20plan%2C%202019-2024_0.pdf [accessed 2026-05-02]
26. Harding R, Selman L, Agupio G, et al. Validation of a core outcome measure for palliative care in Africa: the APCA African Palliative Outcome Scale. Health Qual Life Outcomes 2010 Jan 25;8(1):10. [doi: 10.1186/1477-7525-8-10] [Medline: 20100332]
27. WHO consolidated guidelines on tuberculosis: module 5: management of tuberculosis in children and adolescents. World Health Organization. 2022. URL: https://www.who.int/publications/i/item/9789240046764 [accessed 2026-05-02]
28. Bennett MI, Allsop MJ, Allen P, et al. Pain self-management interventions for community-based patients with advanced cancer: a research programme including the IMPACCT RCT. Programme Grants Appl Res 2021;9(15):1-92. [doi: 10.3310/pgfar09150]
29. Buyinza N, Ryan C, Olet D, Lwanga O, Acuda W. Understanding the need, challenges and opportunities for integrating palliative care services into humanitarian health interventions among health care workers serving refugee communities in Uganda. Nur Primary Care 2025;9(3):1-7. [doi: 10.33425/2639-9474.1286]
30. Browne RH. On the use of a pilot sample for sample size determination. Stat Med 1995 Sep 15;14(17):1933-1940. [doi: 10.1002/sim.4780141709] [Medline: 8532986]
31. Rutstein SO, Staveteig S. Making the Demographic and Health Surveys Wealth Index Comparable: DHS Methodological Reports 9: United States Agency for International Development; 2014.
32. Oken MM, Creech RH, Tormey DC, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol 1982 Dec;5(6):649-655. [doi: 10.1097/00000421-198212000-00014] [Medline: 7165009]
33. Green J, Thorogood N. Qualitative Methods for Health Research 2018. URL: https://sk.sagepub.com/book/mono/qualitative-methods-for-health-research-4e/toc [accessed 2026-04-29] [doi: 10.4135/9781036234386]
34. Yang GM, Lee Y, Ke Y, Neo PSH, Cheung YB. Feasibility of weekly Integrated Palliative Care Outcome Scale (IPOS) within a stepped care model. J Pain Symptom Manage 2024 Sep;68(3):e174-e182. [doi: 10.1016/j.jpainsymman.2024.06.004] [Medline: 38878909]
35. Pavic M, Klaas V, Theile G, Kraft J, Tröster G, Guckenberger M. Feasibility and usability aspects of continuous remote monitoring of health status in palliative cancer patients using wearables. Oncology (Williston Park, NY) 2020 Jun 23;98(6):386-395. [doi: 10.1159/000501433]
36. Fridriksdottir N, Ingadottir B, Skuladottir K, Zoëga S, Gunnarsdottir S. Supportive digital health service during cancer chemotherapy: single-arm before-and-after feasibility study. JMIR Form Res 2023 Dec 22;7:e50550. [doi: 10.2196/50550] [Medline: 38015268]
37. Matlin SA, Hanefeld J, Corte-Real A, et al. Digital solutions for migrant and refugee health: a framework for analysis and action. Lancet Reg Health Eur 2025 Mar;50:101190. [doi: 10.1016/j.lanepe.2024.101190] [Medline: 39816782]
38. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13(1):24047204. [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
39. Integrating palliative care and symptom relief into the response to humanitarian emergencies and crises: a WHO guide. : World Health Organization; 2018 URL: https://iris.who.int/server/api/core/bitstreams/75a499d3-bf99-456b-a48e-6d5e5043e81f/content [accessed 2026-05-2]
40. Draugelis SD, Brown EC, Donahue DA, et al. Development and implementation of an electronic health record system for use in humanitarian emergencies, disaster response, and conflict zones. PLOS Glob Public Health 2025;5(1):e0003124. [doi: 10.1371/journal.pgph.0003124] [Medline: 39879195]
41. Mutatina B, Basaza R, Obuku E, Lavis JN, Sewankambo N. Identifying and characterising health policy and system-relevant documents in Uganda: a scoping review to develop a framework for the development of a one-stop shop. Health Res Policy Syst 2017 Feb 6;15(1):7. [doi: 10.1186/s12961-017-0170-3] [Medline: 28166798]

XSL•FO
RenderX

42. Schwartz JI, Guwatudde D, Nugent R, Kiiza CM. Looking at non-communicable diseases in Uganda through a local lens: an analysis using locally derived data. Global Health 2014 Nov 19;10(1):77. [doi: 10.1186/s12992-014-0077-5] [Medline: 25406738]
43. Allsop MJ, Kabukye J, Powell RA, Namisango E. Routine data and minimum datasets for palliative cancer care in sub-saharan africa: their role, barriers and facilitators. In: Silbermann M, editor. Palliative Care for Chronic Cancer Patients in the Community: Global Approaches and Future Applications: Springer International Publishing; 2021:179-192. [doi: 10.1007/978-3-030-54526-0_15]
44. Harding R, Selman L, Powell RA, et al. Research into palliative care in sub-Saharan Africa. Lancet Oncol 2013 Apr;14(4):e183-e188. [doi: 10.1016/S1470-2045(12)70396-0] [Medline: 23561750]
45. Naderbagi A, Loblay V, Zahed IUM, et al. Cultural and contextual adaptation of digital health interventions: narrative review. J Med Internet Res 2024 Jul 9;26:e55130. [doi: 10.2196/55130] [Medline: 38980719]
46. DeGroot L, Gillette R, Villalobos JP, et al. Feasibility of a digital palliative care intervention (Convoy-Pal) for older adults with heart failure and multiple chronic conditions and their caregivers: a waitlist randomized control trial. BMC Palliat Care 2024 Oct 1;23(1):234. [doi: 10.1186/s12904-024-01561-w] [Medline: 39354453]
47. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022 Apr 26;24(4):e35120. [doi: 10.2196/35120] [Medline: 35471414]
48. Eaton C, Vallejo N, McDonald X, et al. User engagement with mHealth interventions to promote treatment adherence and self-management in people with chronic health conditions: systematic review. J Med Internet Res 2024 Sep 24;26:e50508. [doi: 10.2196/50508] [Medline: 39316431]
49. Benson J, Brand T, Christianson L, Lakeberg M. Localisation of digital health tools used by displaced populations in low and middle-income settings: a scoping review and critical analysis of the Participation Revolution. Confl Health 2023 Apr 15;17(1):20. [doi: 10.1186/s13031-023-00518-9] [Medline: 37061703]
50. Bausewein C, Daveson BA, Currow DC, et al. EAPC White Paper on outcome measurement in palliative care: improving practice, attaining outcomes and delivering quality services - Recommendations from the European Association for Palliative Care (EAPC) Task Force on Outcome Measurement. Palliat Med 2016 Jan;30(1):6-22. [doi: 10.1177/0269216315589898] [Medline: 26068193]
51. Chyr LC, Sauers EG, Dy SM, Waldfogel JM. The application of minimal clinically important differences in palliative research: interpretation of results of a systematic review. J Pain Symptom Manage 2022 Dec;64(6):e363-e371. [doi: 10.1016/j.jpainsymman.2022.08.008] [Medline: 36002121]
52. Matsuda Y, Mori M, Tokoro A, et al. Minimal clinically important difference in the Integrated Palliative care Outcome Scale for cancer dyspnea. J Pain Symptom Manage 2025 Oct;70(4):335-340. [doi: 10.1016/j.jpainsymman.2025.06.010] [Medline: 40550290]
53. Ebenso B, Namisango E, Abejirinde IO, Allsop MJ. Editorial: the scale-up and sustainability of digital health interventions in low- and middle-income settings. Front Digit Health 2025;7:1634223. [doi: 10.3389/fdgth.2025.1634223] [Medline: 41058973]
54. Dilhani WNS, Mitchell S, Dale J, Toor K, Javaid M, MacArtney JI. A mixed-methods systematic review investigating the use of digital health interventions to provide palliative and end-of-life care for patients in low- and middle-income countries. Palliat Care Soc Pract 2024 Apr;18:26323524241236965. [doi: 10.1177/26323524241236965]
55. Doocy S, Paik K, Lyles E, et al. Pilot testing and implementation of a mHealth tool for non-communicable diseases in a humanitarian setting. PLoS Curr 2017 Jun 5;9:ecurrents.dis.e98c648aac93797b1996a37de099be74. [doi: 10.1371/currents.dis.e98c648aac93797b1996a37de099be74] [Medline: 28744410]
56. Parkes P, Pillay TD, Bdaiwi Y, et al. Telemedicine interventions in six conflict-affected countries in the WHO Eastern Mediterranean region: a systematic review. Confl Health 2022 Dec 14;16(1):64. [doi: 10.1186/s13031-022-00493-7] [Medline: 36517869]
57. Beduschi A. Harnessing the potential of artificial intelligence for humanitarian action: opportunities and risks. Int rev Red Cross 2022 Apr;104(919):1149-1169. [doi: 10.1017/S1816383122000261]
58. Grasser LR. Addressing mental health concerns in refugees and displaced populations: is enough being done? RMHP 2022;Volume 15(null):909-922. [doi: 10.2147/RMHP.S270233]
59. Ross J, Stevenson F, Dack C, et al. Developing an implementation strategy for a digital health intervention: an example in routine healthcare. BMC Health Serv Res 2018 Oct 19;18(1):794. [doi: 10.1186/s12913-018-3615-7] [Medline: 30340639]

Abbreviations

APCA: African Palliative Care Association
COREQ: Consolidated Criteria for Reporting Qualitative Research
DHIS2: District Health Information System 2
MCID: minimal clinically important difference
POS: Palliative Outcome Scale
VHT: village health team

XSL•FO
RenderX

Edited by L Buis; submitted 05.Mar.2025; peer-reviewed by C Wang, S Agondeze; revised version received 11.Nov.2025; accepted 02.Apr.2026; published 26.Jun.2026.

Please cite as:
Namisango E, Aduro A, Goodman W, Burke S, Ryabonye R, Mutaasa N, Muyami T, Nabirye E, Olodi D, Ederu V, Ebenso B, Salako O, Okunade K, Azizoddin D, Leng M, Lorenz K, Muehlensiepen F, Powell RA, Allsop M
Feasibility and Acceptability of mPallCare, a Digital Health Intervention for People Living With Advanced Cancer in a Refugee Settlement in Uganda: Mixed Methods Study
JMIR Mhealth Uhealth 2026;14:e73483
URL: https://mhealth.jmir.org/2026/1/e73483
doi:10.2196/73483

© Eve Namisango, Agatha Aduro, William Goodman, Shaunna Burke, Raphael Ryabonye, Nickson Mutaasa, Timothy Muyami, Elizabeth Nabirye, Dennis Olodi, Viola Ederu, Bassey Ebenso, Omolola Salako, Kehinde Okunade, Desiree Azizoddin, Mhoira Leng, Karl Lorenz, Felix Muehlensiepen, Richard A Powell, Matthew Allsop. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 26.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Terminal Digit Preference and Threshold Avoidance in Digital Blood Pressure Measurements During Pregnancy: Secondary Analysis of Data From the CLIP and PRECISE Cohorts

Peter von Dadelszen[1,2], MBChB, DPhil, FRCOG; Akshdeep Sandhu[2,3], MSc; Jeffrey N Bone[2], PhD; Rahat N Qureshi[4], MBBS, FRCOG; Olukayode A Dada[5], PhD; Ashalata A Mallapur[6], MBBS; Marleen Temmerman[7], MD, PhD; Marie-Laure Volvert[1], PhD; Joseph Waiswa[8], PhD; Moses Mukhanya[7], BCS; Hiten D Mistry[1], PhD; Shivaprasad S Goudar[9], MBBS, MD; Hawanatu Jah[10], MD; Angela Koech[7], MBChB, MMed; Charfudin Sacoor[11], MSc; Anifa Vala[11], BVMS; Khatia Munguambe[11], PhD; Anna Roca[10], PhD; Olelekan O Adetoro[5], MBBS; Esperanca Sevene[11], MD, PhD; Marianne Vidler[2,3], PhD; Mrutyunjaya B Bellad[9], MBBS, MD; John Sotunsa[12], MBBS, MPH; Hannah J Blencowe[8], MBBS, PhD; Zulfiqar A Bhutta[13], MBBS, PhD; Umberto d'Alessandro[10], MD, PhD; John Mark Ansermino[3], MBBCh, PhD; Guy A Dumont[14], PhD; Laura A Magee[1], MD, MSc, FRCPC; CLIP Trials Study Group[15,17]; PRECISE Network[15,18]

[1]Department of Women and Children's Health, School of Life Course and Population Sciences, Faculty of Life Sciences and Medicine, King's College London, 6th Floor, Addison House, Guy's Campus, Great Maze Pond, London, United Kingdom

[10]MRC Unit The Gambia at the London School of Hygiene and Tropical Medicine, Banjul, Fajara, The Gambia

[11]Centro de Investigação em Saúde da Manhiça, Manhiça and Department of Clinical Pharmacology, Faculdade de Medicina, Universidade Eduardo Mondlane, Manhiça, Mozambique

[12]Department of Obstetrics and Gynaecology, Babcock University Teaching Hospital, Ilishan-Remo, Ogun State, Nigeria

[13]Centre for Global Child Health, Hospital for Sick Children, Toronto, ON, Canada

[14]Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada

[15]See Acknowledgments

[17]

[18]

[2]Department of Obstetrics and Gynaecology, University of British Columbia, Vancouver, BC, Canada

[3]BC Children's Hospital Research Institute, Vancouver, BC, Canada

[4]Centre of Excellence for Women & Child Health, Aga Khan University, Karachi, Sindh, Pakistan

[5]Centre for Research in Reproductive Health, Sagamu, Ogun State, Nigeria

[6]S Nijalingappa Medical College and HSK Hospital & Research Centre, Bagalkot, Karnataka, India

[7]Centre of Excellence in Women and Child Health (East Africa), Aga Khan University Nairobi, Nairobi, Kenya

[8]MARCH Centre, London School of Hygiene & Tropical Medicine, London, United Kingdom

[9]Women's and Children's Health Research Unit, KLE Academy of Higher Education and Research, Jawaharlal Nehru Medical College, Belagavi, India

Corresponding Author:
Peter von Dadelszen, MBChB, DPhil, FRCOG
Department of Women and Children's Health, School of Life Course and Population Sciences, Faculty of Life Sciences and Medicine, King's College London, 6th Floor, Addison House, Guy's Campus, Great Maze Pond, London, United Kingdom

Abstract

Background: Screening for, detecting, and managing pregnancy hypertension is a core function of antenatal care. To reduce both training requirements and the risks of measurement error in blood pressure (BP) values, automated and semiautomated BP devices have been validated in pregnant women with normal BP and pregnant women with hypertension and introduced for serial antenatal measurement of BP.

Objectives: The study aimed to (1) determine whether or not repeated BP measurements reduced the presence of terminal digit preference and (2) discern whether or not there was evidence of threshold avoidance in the Community-Level Interventions for Preeclampsia (CLIP) trials compared with the purely observational Pregnancy Care Integrating Translational Science, Everywhere (PRECISE) cohorts.

XSL•FO
RenderX

Methods: The BP 3AS1-2 and CRADLE Vital Signs Alert low-cost Microlife BP devices were used by trained research staff in the CLIP trials conducted in India, Mozambique, Nigeria (pilot trial only), and Pakistan and the PRECISE cohorts of unselected pregnant women and nonpregnant women of reproductive age recruited in the Gambia, Kenya, and Mozambique. Both devices algorithmically calculate systolic blood pressure and diastolic blood pressure values displayed on digital read-outs. All BP readings were entered manually into a digital platform, which averaged them as the BP for that visit; the first and second readings were averaged unless they were more than 10 mm Hg different, which triggered a third reading, and the second and third readings were averaged.

Results: A total of 51,875 participants had their BP measured 438,404 times. Using raw BP values, there was terminal digit preference (129,539/911,500, 14.21% vs 10%; *P*<.001 values ended in zero). A total of 28,929 out of 437,446 (6.61%) dBP values were 62 mm Hg, compared with 9310 of 195,349 (4.77%) from the averaged values (*P*<.001); errors were obviated by averaging BP values. There was evidence of both threshold preference and avoidance in the CLIP trials and the PRECISE cohort.

Conclusions: Given the excess of 62 mm Hg values, there is a shared inherent algorithmic error in the calculation of dBP in the BP 3AS1-2 and CRADLE Vital Signs Alert devices. Averaged BP measurements are important to reduce the impact of user errors in manually recording BP values. We recommend that automated and semiautomated BP devices should be connected wirelessly to automatically transfer readings to digital health records to further optimize care.

Trial Registration: ClinicalTrials.gov NCT01911494; https://clinicaltrials.gov/study/NCT01911494

(JMIR Mhealth Uhealth 2026;14:e73307) doi:10.2196/73307

KEYWORDS

Africa; blood pressure devices; pregnancy; South Asia; terminal digit preference; threshold avoidance; women of reproductive age

Introduction

Pregnancy hypertension (defined as systolic blood pressure [sBP] of at least 140 mm Hg or diastolic blood pressure [dBP] of at least 90 mm Hg) is associated with approximately 46,000 maternal and 500,000 perinatal deaths annually; more than 99% of those deaths occur in low-income and middle-income countries [1,2]. Pregnancy hypertension is classified as chronic (pre-existing) hypertension, gestational hypertension, and preeclampsia [2]. Screening for, detecting, and managing pregnancy hypertension is a core function of antenatal care [2,3].

To improve access to BP and other vital signs measurement in low-income and middle-income countries, the semiautomated, low-cost Microlife BP 3AS1-2 and CRADLE Vital Signs Alert (VSA) devices were developed and have been validated in pregnant women with normal BP and pregnant women with hypertension [4,5]. In the context of the CLIP (Community-Level Interventions for Preeclampsia) trials, per protocol use of the Microlife BP 3AS1-2 device within a digital health package deployed by community health workers conducting antenatal contacts in women's homes cost-effectively improved maternal and perinatal survival [6-11].

There are a number of caveats when measuring BP. First, individuals should rest for at least 5 minutes, and their left arm should be supported, with the cuff at the level of their heart. Second, at least 2 measurements of BP should be taken, and if there is discordance between the first and second reading, at least 1 additional measurement should be taken, with the final 2 values of both sBP and dBP being averaged [12]. BP readings that are not wirelessly connected to a digital application are susceptible to a number of errors, such as terminal digit preference (ie, disproportionately enumerating numbers that end in "0" or "5") and threshold avoidance (ie, enumerating numbers just below the threshold that mandates a clinical response) [13].

In this study, we studied all the sBP and dBP values of the pregnant women who participated in the CLIP trials (including the Nigerian pilot trial) and the nonpregnant women of reproductive age and pregnant women who participated in the PRECISE (Pregnancy Care Integrating Translational Science, Everywhere) network. We aimed to (1) determine whether or not repeated BP measurements reduced the presence of terminal digit preference; and (2) discern whether or not there was evidence of threshold avoidance in the CLIP trials compared with the purely observational PRECISE cohort.

Methods

Study Design

This is a secondary analysis of data collected in the 4 countries and 27 intervention clusters of the CLIP cluster randomized controlled trials [14], in India (Karnataka: n=6) [6], Pakistan (Sindh Province: n=10) [7], Mozambique (Maputo and Gaza Provinces: n=6) [8], and Nigeria (Ogun State: n=5) [9], and from all participants in the 3-country PRECISE Network [15], in Gambia (n=3 primary health centers, Farafenni District), Kenya (n=2 primary health centers, Kilifi County), and Mozambique (n=2 primary health centers, Maputo Province; Table S2 in Multimedia Appendix 1). A STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist is provided (Checklist 1).

The CLIP Trials

In brief, pregnant women (aged 15 - 49 y in India, Pakistan, and Nigeria; and 12 - 49 y in Mozambique) were enrolled in the CLIP trials when they first declared their pregnancy and following informed consent [6-9]. The CLIP intervention consisted of community engagement and community health

XSL•FO
RenderX

worker-provided mobile health-guided clinical assessment, initial treatment, and referral to facility [14].

The CLIP intervention was implemented in primarily rural areas of India (February 2014-October 2016), Pakistan (February 2014-December 2016), Nigeria (March 2014-January 2016), and Mozambique (February 2015-February 2017). Within each country, clusters consisted of an established unit of the health system (ie, primary health center in India, union council in Pakistan, administrative post in Mozambique, and local government area in Nigeria).

In intervention clusters, community health workers (accredited social health activists, India; agentes polivalentes elementares, Mozambique; nurses, midwives, community health extension workers, and health assistants, Nigeria; and lady health workers, Pakistan) were trained to provide mobile health–guided contacts that were pregnancy hypertension-oriented, antenatal and postpartum care at home (India, Pakistan, and Mozambique) or at a primary health center (Nigeria). Women in control clusters received usual care, advocated by the World Health Organization. In none of the study countries did community health workers usually either manage pregnancy hypertension or carry with them blood pressure measurement devices or antihypertensive medication.

Digital health-guided clinical assessments were recommended every 4 weeks until 28 weeks of gestation, every 2 weeks from 28 to 35 weeks, weekly from 36 weeks until delivery, once within 24 hours of birth, and postnatally around postpartum days 3, 7, and 14; in Nigeria, visits were opportunistic when women attended the primary health center [16,17].

For generalizability, as sBP can be identified by the return of a radial pulse without the requirement for a stethoscope, the CLIP trial protocol selected a sBP of at least 140 mm Hg as the threshold at which community health workers would respond by initiating referral. In addition, community health workers were directed by the app to either administer oral methyldopa 750 mg for sBP of 160 mm Hg or higher; administer intramuscular magnesium sulfate 10 g for suspected severe preeclampsia (miniPIERS [mini Pre-eclampsia Integrated Estimate of Risk] for an adverse maternal outcome of at least 25%, severe systolic hypertension [at least 160 mm Hg], eclampsia, stroke, or vaginal bleeding); or refer the woman to a comprehensive emergency obstetric care facility for suspected preeclampsia or increased risk of stillbirth (4+ dipstick proteinuria value, or absent fetal movements for at least 12 h) [10,14].

The PRECISE Observational Cohorts

Briefly, in each participating country, 2 cohorts of women were recruited to optimize the identification of social, clinical, and biomarker determinants of the placental disorders of pregnancy, with purposeful sampling of both urban-dwelling and rural-dwelling women [18].

First, unselected pregnant women planning to give birth in the facility (primary health center) were approached and recruited at the time of their booking visit for antenatal care [18]. Consenting women answered an in-depth questionnaire regarding social determinants and clinical history, underwent a targeted examination (eg, biometry, BP, and pulse oximetry), and provided biological samples (blood, urine, and vaginal swabs). Participants were followed up (1) for a second antenatal research visit in the third trimester, at least 4 weeks after the first visit (questionnaire, targeted examination, and phlebotomy); (2) the birth episode (both antepartum and postpartum; questionnaire, targeted examination, phlebotomy, vaginal swabs, placental, membrane, and cord blood samples); and (3) 6 weeks to 6 months postpartum (with their infant, if alive; questionnaire, targeted maternal and infant examination, phlebotomy, vaginal swabs, and heel-prick samples).

Second, for comparison and to understand the lives and biology of sub-Saharan African nonpregnant women of reproductive age (16 - 49 y); such women were identified as either accompanying the pregnant women described in Table 1 or attending for family planning services (Kenya and Mozambique) or recruited in their homes by research staff accompanying the district health surveillance system team (the Gambia). Consenting women answered an in-depth questionnaire regarding social determinants and clinical history, underwent a targeted examination (eg, biometry, BP, and pulse oximetry), and provided biological samples (blood, urine, and vaginal swabs) [18].

XSL•FO
RenderX

Table . Characteristics and pregnancy outcomes (where relevant) of the study cohorts.

Characteristics	CLIP[a] trials (n=32,053)	PRECISE[b] pregnancy cohorts (n=6932)	PRECISE WRA[c] cohorts (n=1825)
Age (y)[d], median (IQR)	25 (22-30)	25 (21-30)	28 (23-34)
Nulliparous, n (%)	8782 (27.4)	2112 (30.47)	274 (15.01)
Marital status, n (%)			
Never married or single	1029 (3.21)	1148 (16.56)	445 (24.38)
Married or cohabiting	26 032 (981.22)	5684 (82)	1295 (70.96)
Separated or divorced	208 (0.65)	68 (0.98)	58 (3.18)
Widowed	128 (0.4)	32 (0.46)	27 (1.48)
No formal education, n (%)	14,926 (46.57)	1243 (17.93)	500 (27.4)
Probability of poverty, median (IQR)	—[e]	20.3 (6.1-36.9)	31.7 (13.9-51.3)
GA[f] at enrollment (wk), median (IQR)	22 (15-29)	21 (15-27)	—
Women with birth outcome data			
Outcome (n=23,265 for CLIP; n=5832 for PRECISE)			
GA at delivery (wk), median (IQR)	39.0 (37.0-40.4)	39.0 (37.3-40.4)	—
Maternal death, n (%)	43 (0.18)	8 (0.14)	—
Miscarriage or TOP[g] (<20 wk), n (%)	601 (2.58)	21 (0.36)	—
Stillbirth (≥20 wk), n (%)	819 (3.52)	182 (3.12)	—
Neonatal death, n (%)	862 (3.71)	27 (0.46)	—

[a]CLIP: Community-Level Interventions for Preeclampsia.

[b]PRECISE: Pregnancy Care Integrating Translational Science, Everywhere.

[c]WRA: nonpregnant women of reproductive age.

[d]Maternal age standardized to age at expected date of delivery for pregnant women.

[e]Not applicable.

[f]GA: gestational age.

[g]TOP: termination of pregnancy.

Neither patients nor the public were involved in the design, or conduct, or reporting, or dissemination plans of this audit.

BP Measurement Protocol

The Microlife BP 3AS1-2 device was used in the Indian, Mozambican, and Pakistani CLIP trials, and its successor, the CRADLE VSA, in the Nigerian CLIP pilot trial, and the PRECISE pregnancy and nonpregnancy cohorts. Both devices are inflated manually and use the steady deflationary phase to detect BP, algorithmically calculating systolic and diastolic values from the measured mean arterial pressure and displaying the sBP and dBP and heart rate values on digital read-outs on the front of the device.

The research staff (community health workers in CLIP and nurses in PRECISE) were trained to have women rest seated for 5 minutes, and then to measure BP in a standardized fashion, with the left arm supported and the cuff at the level of the heart, at least twice. All BP readings were entered manually into a study-specific digital platform, which averaged them as the BP for that visit; the first and second readings were averaged unless they were more than 10 mm Hg different, in which instance a third reading was requested by the study digital platform, and the second and third readings were averaged [12].

Statistical Analyses

All BP readings were used for the analyses.

We determined the incremental change in sBP and dBP between the first and second measurements, the second and third measurements, and the first and third measurements, if a third measurement was taken.

Individual BP value frequencies were displayed for both sBP and dBP for each study site, cumulatively for both the CLIP trials and the PRECISE cohorts, and cumulatively for the combined dataset. This was to identify if there was any evidence of terminal digit preference and threshold avoidance. Through that step, evidence of terminal digit preference was confirmed, overreporting of the dBP value of 62 mm Hg became apparent (see the “Results” section), and, in the CLIP dataset, suspicion of threshold avoidance at 140 mm Hg systolic was observed.

XSL•FO
RenderX

We compared the directly observed counts for dBP of 62 mm Hg versus the averaged count for dBP of 62 mm Hg from the duplicate measurement (ie, 1 and 2 of two measurements or 2 and 3 of three measurements) using a Fisher exact test and the Koopman asymptotic score to calculate 95% CI for the relative risk. The percentage of values ending in "0" and "5" was calculated and compared with the expected value of 10% (assuming a uniform distribution of digits) using both raw and averaged measurements using a 1-sample test for proportion.

To assess threshold avoidance in the CLIP dataset (where an intervention was associated with systolic hypertension), we fit a segmented regression model to the number of BP measurements per unit. Specifically, we used a Poisson model with the following covariates: BP level, threshold (above and below 140 mm Hg), and an interaction between these terms. We display fitted values from these models graphically and use a likelihood ratio test comparing models with and without the segmentation to assess evidence of the possible discontinuity corresponding to threshold avoidance. We compared these results with the CLIP dBP values and PRECISE sBP and dBP values, where no relevant clinical decision thresholds were set.

All analyses were performed in R statistical software version 4.0.3 (R Foundation for Statistical Computing). $P<.01$ was used for statistical significance to adjust for multiple comparisons.

Ethical Considerations

The CLIP trials were approved by the University of British Columbia Research Ethics Board (H12-03497) and within each country (MDC/IECHSR/2013‐14/A, KLE University, Belgavi, India; 2590-Obs-ERC-13, Aga Khan University, Karachi, Pakistan; 219/CNBS/13, the Mozambique Ministry of Health, National Bioethics Committee for Health, Maputo, Mozambique; and OOUTH/DA.326/T/1/, Olabisi Onabanjo University Teaching Hospital, Sagamu, Nigeria). The PRECISE study was approved by King's College London (HR-17/18‐7855) and within each country (2018/REC-74, Aga Khan University Hospital, Nairobi, Kenya; SCC 1619, the Gambia Government/the Medical Research Council, the Gambia Joint Committee, Banjul, the Gambia; and 545/CNBS/18, the Mozambique Ministry of Health, National Bioethics Committee for Health, Maputo, Mozambique). All women provided signed informed consent before taking part in either the CLIP trials or the PRECISE study. All data were deidentified and anonymized. CLIP trial participants were not compensated, other than support for urgent transfers to health facilities when complications were noted or per protocol when risks were identified. PRECISE participants were supported to attend study visits (transport and meals) and were supported for urgent transfers to health facilities as indicated by clinical concerns.

Results

Data were available for 51,875 eligible women, who were recruited from communities and primary health centers in 6 countries (the Gambia, India, Kenya, Mozambique, Nigeria, and Pakistan), with a median of 2841 (IQR 610‐7862) women per included cohort. In total, 911,500 discrete sBP or dBP values were available for analysis.

Some individual-level characteristics of the participating women varied between cohorts (Table 1; Table S2 in Multimedia Appendix 1). Women tended to be in their mid-to-late 20s, were variably nulliparous (range 6.4%-39.5%), and, if pregnant, were recruited from 10.8 to 42.6 weeks' gestation. Pregnant participants generally delivered at term. Miscarriage and termination of pregnancy rates reflect generally late booking for antenatal care, as most miscarriages occur before 12 weeks of gestation. Maternal mortality varied from 0% to 0.27%, stillbirths from 2.2% to 5.1%, and neonatal deaths from 0.32% to 5% (Table 1; Table S2 in Multimedia Appendix 1).

First BP measurements tended to be higher compared with subsequent measurements, but not consistently (Table 2; Table S3 in Multimedia Appendix 1). 25,693 of 183,504 (14%) measurements were taken in triplicate. The raw and averaged BP profiles from the CLIP trials and PRECISE Network (Figure 1) examined BP distributions, terminal digit preference, and the unanticipated rate of dBP values at 62 mm Hg.

XSL•FO
RenderX

Table . Characteristics of blood pressure values.

Characteristics	CLIP[a] trials (n=43,118 pregnancies)	PRECISE[b] pregnancy cohorts (n=6932 pregnancies)	PRECISE WRA[c] cohorts (n=1825 women)
BP[d] values, n/N (%)			
2 values	157,811/183,304 (86)	17,249/19,625 (87.89)	1389/1822 (76.23)
3 values	25,693/183,304 (14)	2376/19,625 (12.11)	433/1822 (23.77)
Differences in BP values (mm Hg), median (IQR)			
sBP1[e,f]-sBP2[g]	1 (–2 to 5)	3 (0 to 8)	2 (–2 to 6)
sBP1-sBP3[h]	1 (–2 to 5)	3 (0 to 8)	2 (–2 to 6)
sBP2-sBP3	0 (–4 to 4)	0 (–4 to 5)	–1 (–9 to 6)
dBP1[f,i]-dBP2	1 (–2 to 4)	2 (0 to 6)	1 (–1 to 4)
dBP1-dBP3	1 (–2 to 4)	2 (0 to 6)	1 (–1 to 4)
dBP2-dBP3	0 (–3 to 3)	0 (–2 to 4)	0 (–3 to 4)
Terminal digit preference—raw values (sBP), n/N (%)			
Values ending in “0” (N)	57,343/392,759 (14.6)	6393/41,055 (15.57)	468/4513 (10.37)
Values ending in “5” (N)	44,614/392,759 (11.36)	3782/41,055 (9.21)	388/4513 (8.6)
Terminal digit preference—raw values (dBP), n/N (%)			
Values ending in “0”	58,677/392,319 (14.96)	6245/41,048 (15.21)	409/4079 (10.03)
Values ending in “5”	45,470/392,319 (11.59)	3699/41,048 (9.01)	406/4079 (9.95)
Terminal digit preference—averaged values (sBP), n/N (%)			
Values ending in “0”	18,832/182,636 (10.31)	869/10,911 (7.96)	130/1801 (7.22)
Values ending in “5”	17,159/182,636 (9.4)	868/10,911 (7.96)	139/1801 (7.72)
Terminal digit preference—averaged values (dBP), n/N (%)			
Values ending in “0”	18,808/182,637 (10.3)	819/10,911 (7.51)	149/1801 (8.27)
Values ending in “5”	17,302/182,637 (9.47)	847/10,911 (7.76)	142/1801 (7.88)
Algorithm issues—raw values, n/N (%)			
dBP=62 mm Hg	26,474/392,319 (6.75)	2281/41,048 (5.56)	174/4079 (4.27)
Algorithm issues—averaged values, n/N (%)			
dBP=62 mm Hg	8774/182,637 (4.8)	495/10,911 (4.54)	41/1801 (2.28)

[a]CLIP: Community-Level Interventions for Preeclampsia.

[b]PRECISE: Pregnancy Care Integrating Translational Science, Everywhere.

[c]WRA: nonpregnant women of reproductive age.

[d]BP: blood pressure.

[e]sBP: systolic blood pressure.

[f]BP1: first value.

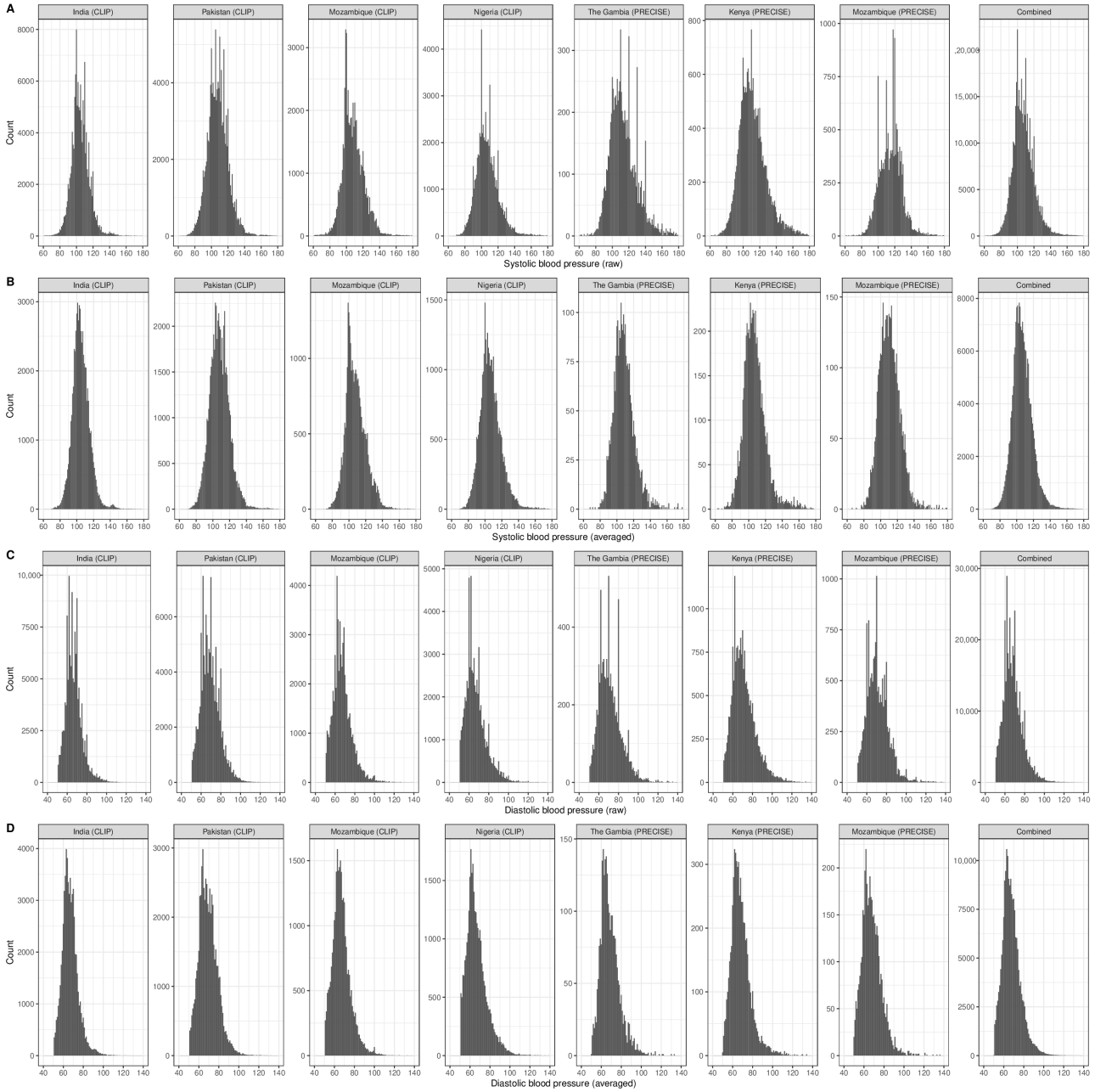
[g]BP2: second value.

[h]BP3: third value (if taken).

[i]dBP: diastolic blood pressure.

XSL•FO
RenderX

Figure 1. Blood pressure value profiles from the CLIP trials and PRECISE Network. (A) Raw systolic blood pressure values; (B) raw diastolic blood pressure values; (C) averaged systolic blood pressure values; (D) averaged diastolic blood pressure values. CLIP: Community-Level Interventions for Preeclampsia; PRECISE: Pregnancy Care Integrating Translational Science, Everywhere.

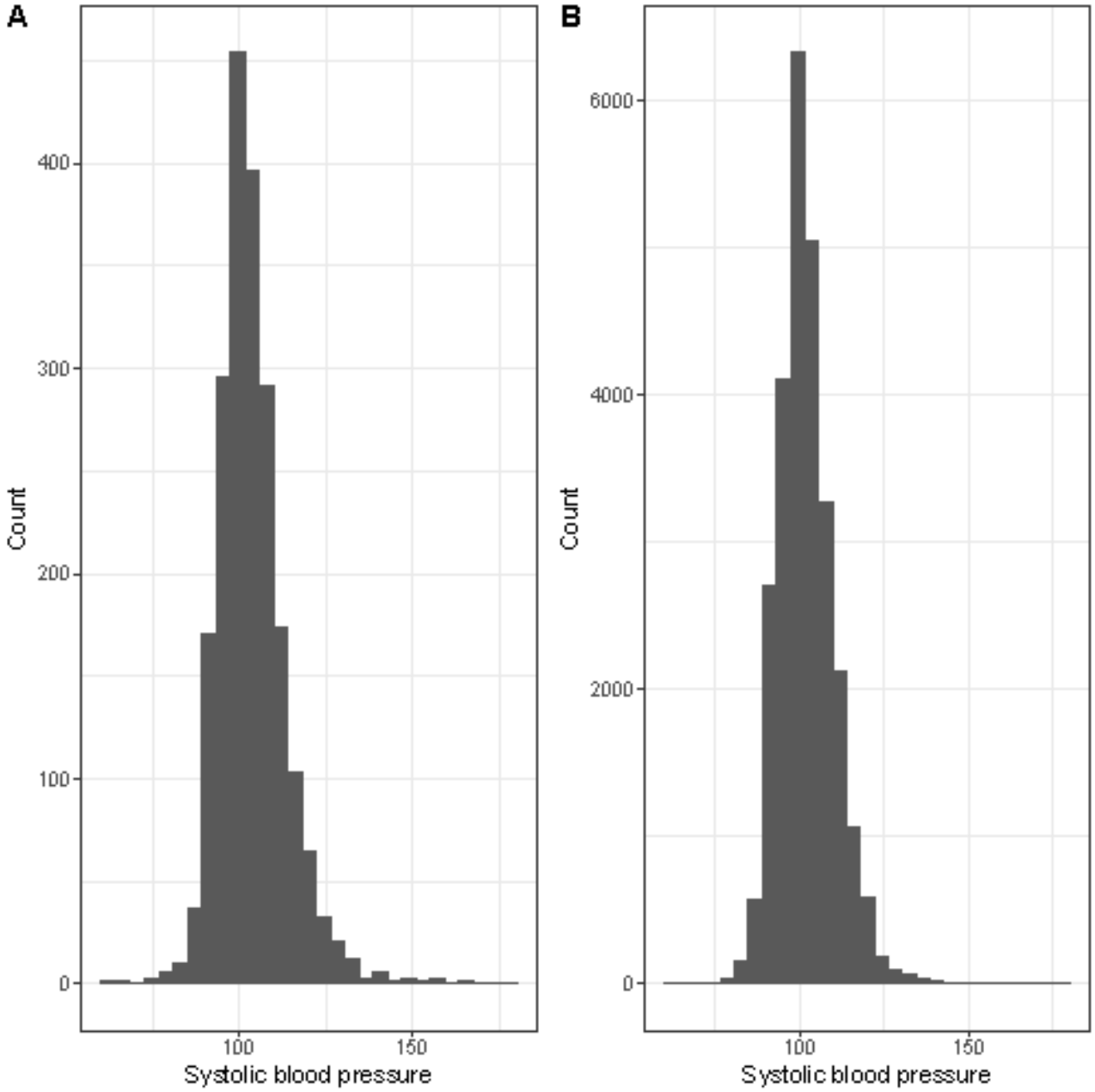


There was clear evidence of terminal digit preference for raw values ending in "0" (129,539/911,500, 14.2%) and "5" (98,359/911,500, 10.8%) versus 10% expected (Table 2; Table S3 in Multimedia Appendix 1; *P*<.001 for both "0" and "5"); these were obviated in the averaged values for "0" (39,607/390,697, 10.1%) and "5" (36,457/390,697, 9.3%).

There was a greater than expected frequency of dBP values of 62 mm Hg (raw observations: 28,929/437,446, 6.6% vs averaged observations: 9310/195,349, 4.8%, *P*<.01; relative risk 1.094, 95% CI 1.09 - 1.10; Table 2; Table S3 in Multimedia Appendix 1; Figure 1). These errors in data recording into digital platforms were largely obviated by averaging the sBP and dBP values (first and second or, if required, second and third; Table 2; Table S3 in Multimedia Appendix 1; Figure 1). The sBP values corresponding to these 62 mm Hg diastolic values varied considerably and were approximately normally distributed (Figure 2); in both cohorts, very few of these values were at least 140 mm Hg (53, 0.2% in CLIP; 17, 0.7% in PRECISE) or at least 160 mm Hg (0.1% in both cohorts).

XSL•FO
RenderX

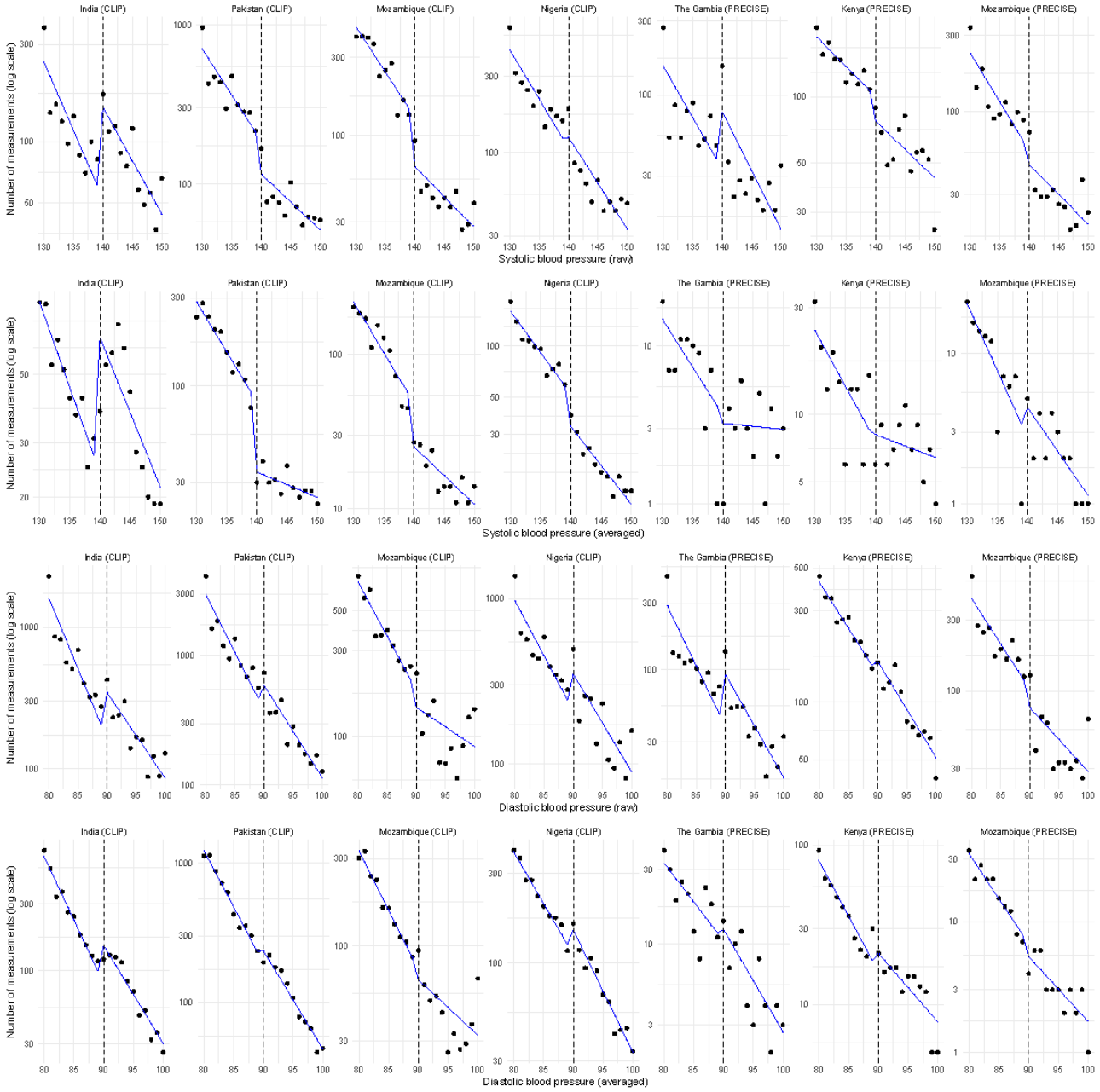
Figure 2. Systolic values corresponding to diastolic blood pressure of 62 mm Hg in (A) CLIP (Community-Level Interventions for Preeclampsia; 55 and 31 of 26,474 ≥140 mm Hg and ≥160 mm Hg, respectively) and (B) PRECISE (Pregnancy Care Integrating Translational Science, Everywhere; 15 and 2 of 2094 ≥140 mm Hg and ≥160 mm Hg, respectively).



For sBP readings, there was evidence of both threshold (140 mm Hg) preference (India; *P*=.02) and threshold avoidance in Mozambique (CLIP, *P*=.007; PRECISE, *P*=.006) and Pakistan (*P*<.001), but for neither the Gambia (*P*=.31), Kenya (*P*=.85), nor Nigeria (*P*=.50; Figure 3). These observations were modified by averaging values (the Gambia *P*=.05; India *P*=.47; Kenya *P*=.05; Mozambique CLIP *P*=.01, PRECISE *P*=.35; Nigeria *P*=.76; and Pakistan *P*<.001; Figure 3). For dBP readings, there was evidence of threshold (90 mm Hg) preference (India *P*<.001; Pakistan *P*<.001) and avoidance (Mozambique CLIP *P*<.001, PRECISE *P*=.008), but neither for the Gambia (*P*=.11), Kenya (*P*=.70), nor Nigeria (*P*=.05; Figure 3). Generally, these observations were modified by averaging values (the Gambia *P*=.44; India *P*<.001; Kenya *P*=.06; Mozambique CLIP *P*<.001, PRECISE *P*=.49; Nigeria *P*=.08; Pakistan *P*=.03; Figure 3).

XSL•FO
RenderX

Figure 3. Threshold preference and avoidance in the CLIP (Community-Level Interventions for Preeclampsia) trials and PRECISE (Pregnancy Care Integrating Translational Science, Everywhere) network. The blue lines are fitted values from segmented Poisson regression models.



Discussion

Principal Results

Using BP data from 51,875 prospectively recruited eligible women living in 5 low- and middle-income countries, 911,500 discrete sBP or dBP values were analyzed in this study. Our protocol of repeating (up to 3 times) and averaging BP measurements (values 1 and 2 of 2, or 2 and 3 of 3) clearly improves the quality of data and clinical guidance that are derived from use of these 2 devices. This was true for the terminal digit preference and the algorithmic issue that resulted in the excess of dBP values of 62 mm Hg, but not for threshold avoidance.

Comparison With Prior Work

Reassuringly for clinical use of these two devices, 62 mm Hg diastolic readings were rarely (≥0.7%) associated with a sBP value of at least 140 mm Hg, and only 0.1% were associated with systolic readings of at least 160 mm Hg that increase the risk for fatal intracranial bleeding [19]. In addition, it might be useful to assess whether or not the 62 mm Hg issue is shared between other Microlife automated and semiautomated BP devices.

Within the CLIP trials, the addition of the traffic light display may underlie why there was no evidence of threshold avoidance in the CLIP Nigeria pilot trial. The CLIP Nigeria pilot trial was the sole part of the CLIP endeavor that used the then-new CRADLE VSA device [9], compared with the BP 3AS1-2 device used in the other CLIP trials. In India (threshold

XSL•FO
RenderX

preference), and Mozambique and Pakistan (threshold avoidance) were clearly seen for systolic, but less so diastolic, hypertension, as the protocol was sBP-driven [6-8,10]. The display of an amber or red light that was visible to others in the room when the BP was taken may have provided important shared information to guide shared decision-making about initiating transfers of care and, in some cases, methyldopa and magnesium sulfate therapy. However, both systolic threshold preference and threshold avoidance were observed in Gambian and Mozambican, but not Kenyan, PRECISE cohort. Diastolic preference and avoidance were both seen, but were of lower magnitude than for sBP values. It would be interesting to understand why the Gambian research staff and the Indian Accredited Social Health Activist workers preferred a value diagnostic of hypertension, while Mozambican research staff and *Agentes Polivalentes Elementares* and Pakistani lady health workers displayed threshold avoidance. Were there differences in the training about the importance of detecting hypertension, or were there concerns about the additional workload associated with making such a diagnosis [15,20-22]?

Previously in the CLIP dataset, we have identified that lowering both sBP and dBP thresholds by 5 mm Hg (ie, 135 mm Hg and 85 mm Hg, respectively, for nonsevere hypertension; and 155 mm Hg and 105 mm Hg, respectively, for severe hypertension), for measurements taken in community settings would improve the detection of pregnant women at risk of the complications associated with pregnancy hypertension, especially preeclampsia [23].

Strengths and Limitations

The strengths of this study include prospective data collection by trained individuals, a shared standardized BP measurement protocol, geographical collection of data from women living in 2 South Asian and 4 sub-Saharan countries (2 from west, and one each from east and southern Africa), and its large sample size. The two study designs, cluster randomized controlled trials (CLIP) and prospective observational cohort (PRECISE), provided additional insights into the role of threshold avoidance with and without the traffic light display introduced in the CRADLE VSA device.

Our limitations include the absence of a wirelessly connected device for comparison, by which means we could have tested the hypothesis that the terminal digit preference and threshold avoidance could be overcome by wireless data collection onto digital health platforms. In addition, we were unable to examine the reasons behind the differences in response to BP thresholds (ie, preference: the Gambia, India vs avoidance: Mozambique, Pakistan vs none: Kenya, Nigeria) between study sites. Such insights would help to overcome threshold avoidance, in particular, in the interim before wireless connectivity between BP devices and digital health platforms becomes routine. Threshold avoidance is a safety signal as women at risk may not have been referred when appropriate.

Conclusions

In summary, the BP 3AS1-2 and CRADLE VSA have been used reliably in a number of other clinical trials and studies with largely positive results [24-27]. They have provided generally reliable, low-cost BP monitors that can be safely deployed from households to facilities in low- and middle-income countries. The addition of the traffic lights in the CRADLE VSA appears to have reduced the risk of threshold avoidance. However, in this study, we have identified opportunities for further improvements that relate to terminal digit preference and the algorithm issues around 62 mm Hg. In the interim, clinicians and fellow researchers should use our protocol for using these devices 2 to 3 times to reduce errors.

Acknowledgments

The authors thank the governments of the Gambia; India; Kilifi County, Kenya; Mozambique; Ogun State, Nigeria; and Sindh Province, Pakistan, for their permission to integrate the CLIP trials and PRECISE study into their health systems with in-kind support. The authors thank the families of the 152 women, 2777 fetuses, and 2688 neonates who died during the study period and who were willing to share their stories despite their grief.

The members of the CLIP Trials Study Group are as follows:

CLIP Mozambique:

Esperança Sevene; Eusébio Macete; Khátia Munguambe; Charfudin Sacoor; Anifa Vala; Helena Boene; Felizarda Amose; Rosa Pires; Zefanias Nhamirre; Marta Macamo; Rogério Chiaú; Analisa Matavele; Faustino Vilanculo; Ariel Nhancolo; Silvestre Cutana; Ernesto Mandlate; Salésio Macuacua; Cassimo Bique; Sibone Mocumbi; Emília Gonçalves; Sónia Maculuve; Ana Ilda Biz; Dulce Mulungo; Orvalho Augusto; Paulo Filimone; Vivalde Nobela; Corsino Tchavana; Cláudio Nkumbula.

CLIP Pakistan:

Rahat Qureshi; Zulfiqar A Bhutta; Zahra Hoodbhoy; Farrukh Raza; Sana Sheikh; Javed Memon; Imran Ahmed; Amjad Hussain.

CLIP India:

Mrutunjaya B Bellad; Umesh S Charantimath; Shivaprasad S Goudar; Geetanjali M Katageri; Avinash J Kavi; Amit P Revankar; Ashalata A Mallapur; Umesh Y Ramdurg; Shashidhar G Bannale; Vaibhav B Dhamanekar; Geetanjali I Mungarwadi; Narayan V Honnungar; Bhalachandra S Kodkany; Anjali M Joshi; Uday S Kudachi; Sphoorthi S Mastiholi; Chandrappa C Karadiguddi; Gudadayya S Kengapur; Namdev A Kamble; Keval S Chougala.

XSL•FO
RenderX

CLIP Trial Steering Committee:

J Mark Ansermino; Ana Pilar Betrán; Richard Derman; Shafik Dharamsi; France Donnay; Sharla Drebit; Guy Dumont; Susheela M Engelbrecht; Veronique Fillipi; Tabassum Firoz; William Grobman; Marian Knight; Ana Langer; Simon Lewin; Gwyneth Lewis; Craig Mitton; Nadine Schuurman; Andrew H Shennan; Joel Singer; James G Thornton; Hubert Wong.

CLIP Trial Executive Committee:

Olalekan O Adetoro; Mrutunjaya M Bellad; Zulfiqar A Bhutta; Peter von Dadelszen; Shivaprasad S Goudar; Jerker Liljestrand; Laura A Magee; Ashalata A Mallapur; Khátia Munguambe; Beth A Payne; Rahat Qureshi; Charfudin Sacoor; Esperança Sevene; Sumedha Sharma; John O Sotunsa; Marianne Vidler.

CLIP Data Safety and Monitoring Board (DSMB):

Romano Nkumbwa Byaruhanga; Brian Darlow; Eileen Hutton; Mario Merialdi; Lehana Thabane.

The members of the PRECISE Network are as follows:

PRECISE The Gambia:

Umberto D'Alessandro; Anna Roca; Hawanatu Jah; Abdoulie Bah; Yorro Bah; Brahima Diallo; Gibril Gabbidon; Lawrence Gibba; Yahaya Idris; Fatoumata Kongira; Melisa Martinez-Alvarez; Modou FS Ndure; Baboucarr Njie; Andrew Prentice; Abdul Sesay; Sambou Suso; Fatima Touray.

PRECISE Kenya:

Marleen Temmerman; Angela Koech; Consolata Juma; Thomas Mendy; Moses Mukhanya; Isaac Mwaniki; Grace Mwashigadi; Joseph Mutunga; Emily Mwadime; Marvin Ochieng; Patricia Okiro; Geoffrey Omuse; Onesmus Wanje.

PRECISE Mozambique:

Esperança Sevene; Anifa Vala; Helena Boene; Carla Carrilho; Salesio Macuacua; Sonia Maculuve; Inacio Mandomando; Lazaro Quimice; Corssino Tchavana.

PRECISE Supporting Institutions and Investigators:

Prestige Tatenda Makanga; Liberty Makacha; Reason Mlambo; Peter von Dadelszen; Laura A. Magee; Rachel Craik; Marie-Laure Volvert; Hiten Mistry; Meriel Flint-O'Kane; Amber Strang; Marina Daniele; Lucilla Poston; Rachel Tribe; Sophie Moore; Tatiana Salisbury; Aris Papageorghiou; Alison Noble; Hannah Blencowe; Veronique Filippi; Joy Lawn; Matt Silver; Joseph Akuze; Ursula Gazeley; Ben Barratt; Judith Cartwright; Guy Whitley; Sanjeev Krishna; Marianne Vidler; Jing (Larry) Li; Jeffrey N Bone; Mai-Lei (Maggie) Woo Kinshella; Domena Tu; Ash Sandhu; Kelly Pickerill; Donna Russell.

Funding

Funding was received from the University of British Columbia, a grantee of the Bill & Melinda Gates Foundation (OPP1017337), and the UK Research and Innovation Global Challenges Research Fund (through the Medical Research Council, MR/P027938/1). The funders of these studies had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Authors' Contributions

The corresponding author and LAM designed the study and wrote the first draft of the manuscript. The corresponding author had access to all data in the study and final responsibility for the decision to submit for publication. JNB was the primary statistician and data visualization for these analyses. AS was primarily responsible for data curation for both studies. All African and South Asian authors were involved in project management and primary data collection, with central coordination by MV (CLIP trials) and MLV and HDM (PRECISE). All authors contributed to the review and editing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Research group members and detailed cohort and blood pressure descriptions.
[PDF File, 243 KB - mhealth_v14i1e73307_app1.pdf]

Checklist 1
STROBE checklist.
[PDF File, 216 KB - mhealth_v14i1e73307_app2.pdf]

XSL•FO
RenderX

References

1. Magee LA, Nicolaides KH, von Dadelszen P. Preeclampsia. N Engl J Med 2022 May 12;386(19):1817-1832. [doi: 10.1056/NEJMra2109523] [Medline: 35544388]
2. Magee LA, Brown MA, Hall DR, et al. The 2021 international society for the study of hypertension in pregnancy classification, diagnosis & management recommendations for international practice. Pregnancy Hypertens 2022 Mar;27:148-169. [doi: 10.1016/j.preghy.2021.09.008] [Medline: 35066406]
3. WHO recommendations on antenatal care for a positive pregnancy experience. : World Health Organization; 2016 URL: https://www.who.int/publications/i/item/9789241549912 [accessed 2025-12-18]
4. Nathan HL, de Greeff A, Hezelgrave NL, Chappell LC, Shennan AH. An accurate semiautomated oscillometric blood pressure device for use in pregnancy (including pre-eclampsia) in a low-income and middle-income country population: the Microlife 3AS1-2. Blood Press Monit 2015 Feb;20(1):52-55. [doi: 10.1097/MBP.0000000000000086] [Medline: 25243711]
5. Nathan HL, Vousden N, Lawley E, et al. Development and evaluation of a novel vital signs alert device for use in pregnancy in low-resource settings. BMJ Innov 2018 Oct;4(4):192-198. [doi: 10.1136/bmjinnov-2017-000235] [Medline: 30319784]
6. Bellad MB, Goudar SS, Mallapur AA, et al. Community level interventions for pre-eclampsia (CLIP) in India: a cluster randomised controlled trial. Pregnancy Hypertens 2020 Jul;21:166-175. [doi: 10.1016/j.preghy.2020.05.008] [Medline: 32554291]
7. Qureshi RN, Sheikh S, Hoodbhoy Z, et al. Community-level interventions for pre-eclampsia (CLIP) in Pakistan: a cluster randomised controlled trial. Pregnancy Hypertens 2020 Oct;22:109-118. [doi: 10.1016/j.preghy.2020.07.011] [Medline: 32777710]
8. Sevene E, Sharma S, Munguambe K, et al. Community-level interventions for pre-eclampsia (CLIP) in Mozambique: a cluster randomised controlled trial. Pregnancy Hypertens 2020 Jul;21:96-105. [doi: 10.1016/j.preghy.2020.05.006] [Medline: 32464527]
9. Adepoju AA, Vidler M, Akadri AA, et al. The ability and safety of community-based health workers to safely initiate lifesaving therapies for pre-eclampsia in Ogun State, Nigeria: an analysis of 260 community treatments with MgSO4 and/or methyldopa. Pregnancy Hypertens 2021 Aug;25:179-184. [doi: 10.1016/j.preghy.2021.05.005] [Medline: 34175582]
10. von Dadelszen P, Bhutta ZA, Sharma S, et al. The community-level interventions for pre-eclampsia (CLIP) cluster randomised trials in Mozambique, Pakistan, and India: an individual participant-level meta-analysis. Lancet 2020 Aug 22;396(10250):553-563. [doi: 10.1016/S0140-6736(20)31128-4] [Medline: 32828187]
11. Bone JN, Khowaja AR, Vidler M, et al. Economic and cost-effectiveness analysis of the community-level interventions for pre-eclampsia (CLIP) trials in India, Pakistan and Mozambique. BMJ Glob Health 2021 May;6(5):e004123. [doi: 10.1136/bmjgh-2020-004123] [Medline: 34031134]
12. Daskalopoulou SS, Khan NA, Quinn RR, et al. The 2012 Canadian hypertension education program recommendations for the management of hypertension: blood pressure measurement, diagnosis, assessment of risk, and therapy. Can J Cardiol 2012 May;28(3):270-287. [doi: 10.1016/j.cjca.2012.02.018] [Medline: 22595447]
13. Kottke TE, Anderson JP, Zillhardt JD, et al. Association of an automated blood pressure measurement quality improvement program with terminal digit preference and recorded mean blood pressure in 11 clinics. JAMA Netw Open 2022 Aug 1;5(8):e2229098. [doi: 10.1001/jamanetworkopen.2022.29098] [Medline: 36044216]
14. Payne BA, Bhutta ZA, Magee LA, et al. Protocol 13PRT/9313: the Community Level Interventions for Pre-eclampsia (CLIP) trials: four prospective cluster randomised controlled trials comparing a package of interventions directed towards improving maternal and perinatal outcomes related to pre-eclampsia with current standards of care (NCT01911494). The Lancet. URL: https://www.thelancet.com/protocol-reviews/13PRT-9313
15. von Dadelszen P, Flint-O'Kane M, Poston L, et al. The PRECISE (PREgnancy Care Integrating translational Science, Everywhere) Network's first protocol: deep phenotyping in three sub-Saharan African countries. Reprod Health 2020 Apr 30;17(Suppl 1):51. [doi: 10.1186/s12978-020-0872-9] [Medline: 32354357]
16. Dunsmuir DT, Payne BA, Cloete G, et al. Development of mHealth applications for pre-eclampsia triage. IEEE J Biomed Health Inform 2014 Nov;18(6):1857-1864. [doi: 10.1109/JBHI.2014.2301156] [Medline: 25375683]
17. Lim J, Cloete G, Dunsmuir DT, et al. Usability and feasibility of PIERS on the move: an mHealth app for pre-eclampsia triage. JMIR Mhealth Uhealth 2015 Apr 17;3(2):e37. [doi: 10.2196/mhealth.3942] [Medline: 25887292]
18. Craik R, Akuze J, Volvert ML, et al. PREgnancy Care Integrating translational Science, Everywhere (PRECISE): a prospective cohort study of African pregnant and non-pregnant women to investigate placental disorders - cohort profile. BMJ Open 2025 May 11;15(5):e091831. [doi: 10.1136/bmjopen-2024-091831] [Medline: 40350193]
19. Martin JN Jr, Thigpen BD, Moore RC, Rose CH, Cushman J, May W. Stroke and severe preeclampsia and eclampsia: a paradigm shift focusing on systolic blood pressure. Obstet Gynecol 2005 Feb;105(2):246-254. [doi: 10.1097/01.AOG.0000151116.84113.56] [Medline: 15684147]
20. Charanthimath U, Katageri G, Kinshella MLW, et al. Community health worker evaluation of implementing an mHealth application to support maternal health care in rural India. Front Glob Womens Health 2021;2:645690. [doi: 10.3389/fgwh.2021.645690] [Medline: 34816198]

XSL•FO
RenderX

21. Boene H, Valá A, Kinshella MLW, et al. Implementation of the PIERS on the move mHealth application from the perspective of community health workers and nurses in rural Mozambique. Front Glob Womens Health 2021;2:659582. [doi: 10.3389/fgwh.2021.659582] [Medline: 34816216]
22. Kinshella MLW, Sheikh S, Bawani S, et al. "Now You Have Become Doctors": lady health workers' experiences implementing an mHealth application in rural Pakistan. Front Glob Womens Health 2021;2:645705. [doi: 10.3389/fgwh.2021.645705] [Medline: 34816199]
23. von Dadelszen P, Bone JN, Sandhu A, et al. Choosing blood pressure thresholds to inform pregnancy care in the community: an analysis of cluster trials. BJOG 2023 Sep;130(10):1275-1285. [doi: 10.1111/1471-0528.17465] [Medline: 37092252]
24. Vousden N, Lawley E, Nathan HL, et al. Effect of a novel vital sign device on maternal mortality and morbidity in low-resource settings: a pragmatic, stepped-wedge, cluster-randomised controlled trial. Lancet Glob Health 2019 Mar;7(3):e347-e356. [doi: 10.1016/S2214-109X(18)30526-6] [Medline: 30784635]
25. Beardmore-Gray A, Vousden N, Seed PT, et al. Planned delivery or expectant management for late preterm pre-eclampsia in low-income and middle-income countries (CRADLE-4): a multicentre, open-label, randomised controlled trial. Lancet 2023 Jul 29;402(10399):386-396. [doi: 10.1016/S0140-6736(23)00688-8] [Medline: 37393919]
26. Bright S, Moses F, Ridout A, et al. Scale-up of a novel vital signs alert device to improve maternity care in Sierra Leone: a mixed methods evaluation of adoption. Reprod Health 2023 Jan 6;20(1):6. [doi: 10.1186/s12978-022-01551-2] [Medline: 36609353]
27. Kuhrt K, Seed PT, Shennan AH. A prospective case control study to evaluate shock index for identifying patients at risk of clinically important malaria in refugee settings. Trop Doct 2023 Apr;53(2):233-236. [doi: 10.1177/00494755221134975] [Medline: 36654493]

Abbreviations

BP: blood pressure
CLIP: Community-Level Interventions for Preeclampsia
dBP: diastolic blood pressure
PRECISE: Pregnancy Care Integrating Translational Science, Everywhere
sBP: systolic blood pressure
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
VSA: Vital Signs Alert

Edited by L Buis; submitted 01.Mar.2025; peer-reviewed by A Adeoye, X Zhou; revised version received 01.Dec.2025; accepted 01.Dec.2025; published 03.Jun.2026.

Please cite as:
von Dadelszen P, Sandhu A, Bone JN, Qureshi RN, Dada OA, Mallapur AA, Temmerman M, Volvert ML, Waiswa J, Mukhanya M, Mistry HD, Goudar SS, Jah H, Koech A, Sacoor C, Vala A, Munguambe K, Roca A, Adetoro OO, Sevene E, Vidler M, Bellad MB, Sotunsa J, Blencowe HJ, Bhutta ZA, d'Alessandro U, Ansermino JM, Dumont GA, Magee LA, CLIP Trials Study Group, PRECISE Network
Terminal Digit Preference and Threshold Avoidance in Digital Blood Pressure Measurements During Pregnancy: Secondary Analysis of Data From the CLIP and PRECISE Cohorts
JMIR Mhealth Uhealth 2026;14:e73307
URL: https://mhealth.jmir.org/2026/1/e73307
doi:10.2196/73307

© Peter von Dadelszen, Akshdeep Sandhu, Jeffrey N Bone, Rahat N Qureshi, Olukayode A Dada, Ashalata A Mallapur, Marleen Temmerman, Marie-Laure Volvert, Joseph Waiswa, Moses Mukhanya, Hiten D Mistry, Shivaprasad S Goudar, Hawanatu Jah, Angela Koech, Charfudin Sacoor, Anifa Vala, Khatia Munguambe, Anna Roca, Olelekan O Adetoro, Esperanca Sevene, Marianne Vidler, Mrutyunjaya B Bellad, John Sotunsa, Hannah J Blencowe, Zulfiqar A Bhutta, Umberto d'Alessandro, John Mark Ansermino, Guy A Dumont, Laura A Magee, CLIP Trials Study Group, PRECISE Network. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 3.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Community-Based Usability Study of an AI-Enabled Oral Cancer Screening App Operated by Village Health Volunteers: Mixed Methods Study

Mansuang Wongsapai[1], DDS, PhD; Kornwipa Wudtijureepun[1], BSc; Thawatchai Suthachai[1], BTec; Decha Tamdee[2], PhD; Jitjiroj Ittichaicharoen[3], DDS, PhD; Patiwet Wuttisarnwattana[4], PhD; Siriwan Suebnukarn[5], DDS, PhD

[1]Ministry of Public Health of Thailand, Intercountry Centre for Oral Health, Chiang Mai, Thailand

[2]Faculty of Nursing, Chiang Mai University, Chiang Mai, Thailand

[3]Faculty of Dentistry, Chiang Mai University, Chiang Mai, Thailand

[4]Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand

[5]Faculty of Dentistry, Thammasat University, Piyachart Building, 9th Floor, Phaholyothin Road, Pathum Thani, Thailand

Corresponding Author:
Siriwan Suebnukarn, DDS, PhD
Faculty of Dentistry, Thammasat University, Piyachart Building, 9th Floor, Phaholyothin Road, Pathum Thani, Thailand

Abstract

Background: Oral cancer is a major public health concern in low- and middle-income countries, where access to specialist care and early detection remains limited. Mobile health technologies supported by artificial intelligence (AI) offer a scalable approach to extend screening services into underserved communities. In Thailand, village health volunteers (VHVs) are key frontline workers who provide preventive services and bridge gaps between rural populations and specialist care.

Objective: This study aimed to describe the technical development of RiskOCA (Risk Assessment for Oral Cancer using Artificial Intelligence), a smartphone-based, AI-assisted oral cancer screening platform, and evaluate its usability when deployed by VHVs in a rural Thai province.

Methods: RiskOCA was developed using a 3-tier architecture comprising a patient-facing interface for risk factor profiling and guided imaging, an embedded deep learning engine (DeepLab v3+with ResNet-50 backbone) for lesion detection, classification, and segmentation, and a secure specialist portal for expert review of all cases. The AI model was trained on 2226 annotated intraoral images and validated for real-world use. Field testing was conducted in the Phu Kamyao district, Phayao province, where 1242 adults (≥40 y) were screened with assistance from VHVs. Usability was evaluated through a structured 25-item questionnaire completed by 250 VHVs, with responses rated on a 5-point Likert scale.

Results: The AI model achieved a mean classification accuracy of 93.22% (SD 0.88%) across 3 diagnostic categories. Usability evaluation indicated high satisfaction across all domains, with an overall mean score of 4.17 out of 5. The highest ratings were for the app's impact on older adult surveillance (mean 4.30), while all domains were rated "satisfied" or "very satisfied."

Conclusions: RiskOCA demonstrated strong technical performance and high user acceptance among VHVs, supporting its feasibility for community-based oral cancer screening. By integrating AI-assisted triage with expert review, the platform has the potential to reduce diagnostic delays, expand screening coverage, and serve as a scalable model for oral cancer prevention in resource-limited settings.

(JMIR Mhealth Uhealth 2026;14:e83738) doi:10.2196/83738

KEYWORDS

mobile apps; cancer screenings; prevention; early detection of cancer; health inequity

Introduction

Background

Oral cancer continues to be a major global health burden, particularly in low- and middle-income countries where access to early detection and specialist care remains limited. In Southeast Asia, oral health is often underprioritized in public health policy, despite its growing impact on population well-being. In 2020, the region reported more than 900 million cases of oral disease and nearly 100,000 deaths due to lip and oral cavity cancers—conditions that disproportionately affect rural populations [1]. In response, the World Health Assembly and the World Health Organization Southeast Asia Regional Committee have called for oral health to be integrated into universal health coverage and noncommunicable disease frameworks [2]. Nevertheless, geographic, financial, and

XSL•FO
RenderX

infrastructural barriers continue to prevent timely access to screening and care in remote communities.

The early detection of oral squamous cell carcinoma (OSCC) and oral potentially malignant disorders (OPMD) is essential to reducing disease burden, improving survival outcomes, and preserving function. However, timely diagnosis is often hindered by low health literacy, insufficient clinical workforce capacity, and delayed referrals [3,4]. In rural Thailand and similar settings, the challenge lies not only in detection but also in delivering equitable access to specialist assessment and follow-up. One promising strategy to address this gap is task-shifting, wherein screening responsibilities are extended to trained community health workers—such as Thailand's village health volunteers (VHVs)—who serve as trusted frontline actors in noncommunicable disease surveillance and public health promotion [5].

Thailand's VHVs, recognized as a cornerstone of its universal health coverage success, play a pivotal role in linking underserved communities with formal health care systems [6]. Their responsibilities include conducting home visits, collecting health data, and facilitating screening initiatives across a range of conditions, including hypertension, diabetes, and cervical cancer. Leveraging this existing infrastructure for oral cancer screening could offer a scalable, sustainable approach to early detection—particularly if supported by digital health tools that enable accurate assessment and specialist consultation.

Mobile health (mHealth) technologies have emerged as effective solutions to expand health care access in resource-limited settings. Mobile apps can support self-screening, risk stratification, and community-based triage and have shown promise across various domains, including infectious disease surveillance and chronic care [7,8]. In the context of oral cancer, smartphone-based platforms can enable nonspecialists to identify suspicious lesions and trigger referral pathways earlier in the disease course. Yet, despite advances in design and accuracy, few tools have been evaluated for usability or real-world effectiveness in frontline settings, particularly among nonexpert users, such as VHVs [9].

Recent developments in artificial intelligence (AI)—especially deep learning—have demonstrated strong diagnostic potential in analyzing clinical images of oral lesions [10,11]. However, most AI models remain isolated in research settings, with limited translation to community health workflows. To have a public health impact, these technologies must be embedded within systems that are accessible, usable, and clinically meaningful for those operating outside hospital environments [12].

To address this challenge, we developed RiskOCA (Risk Assessment for Oral Cancer using Artificial Intelligence), a smartphone-based app designed to support AI-assisted screening and referral for oral cancer [13]. The platform integrates risk factor profiling, guided intraoral imaging, and lesion analysis via deep learning, with all submitted cases reviewed by oral cancer specialists. Crucially, RiskOCA was designed using a user-centered approach to enable VHVs—as well as other frontline workers—to participate in the early detection process, thereby bridging the gap between remote communities and oral cancer specialists.

Objectives

This paper presents the technical development and usability evaluation of RiskOCA, with a focus on its system architecture, user interface design, and field-based testing among VHVs in a rural Thai province. By examining how VHVs interact with the app and assess its usability, the study provides insights into the potential of AI-integrated mHealth tools to strengthen community-based cancer screening and reduce diagnostic delays in resource-limited settings.

Methods

Ethical Considerations

This study comprised three main components: (1) system development, which included the architecture, AI model, platform technology stack, and data security protocols; (2) user interface and experience design, guided by a user-centered approach to ensure accessibility and cultural relevance; and (3) user evaluation, which involved field-based testing among community participants and VHVs in a rural Thai province. All procedures were conducted in accordance with ethical standards for research involving human participants.

The study was approved by the Regional Health Promotion Center 1, Department of Health, Ministry of Public Health, Thailand (approval No. 20/2565, dated April 29, 2022). Informed consent was obtained from all participants prior to participation using both written and verbal processes conducted in the local language (Thai). Written informed consent was documented using an Institutional Review Board–approved consent form. In addition, trained VHVs provided a standardized verbal explanation of the study purpose, procedures, potential risks and benefits, data handling practices, and the voluntary nature of participation.

The consent process explicitly informed participants that their clinical images and associated data would be used for AI-based risk assessment and may be included in anonymized datasets for research and model development. Participants were informed that no personally identifiable information (PII) would be used for AI training, that participation could be declined or withdrawn at any time without consequence, and that no financial compensation was provided.

Verbal consent was recorded by study staff using a consent checklist confirming that key information had been explained and understood. Participant understanding was assessed through interactive dialogue, during which participants were encouraged to ask questions and were asked to confirm their understanding of the core elements of the study, including data use and privacy protections, before enrollment proceeded.

To ensure confidentiality, all PII (eg, names, national ID numbers, and contact details) was removed or stored separately prior to analysis. The research team only had access to the fully anonymized data. Participants did not receive any compensation for their data being included in this study.

XSL•FO
RenderX

System Development

Architecture Overview

RiskOCA is a smartphone-based digital platform designed to support oral cancer risk assessment and guided self-examination, particularly among adults in underserved and remote communities. The system architecture comprises four core components: (1) an individual risk profiling module, which collects personal health data and behavioral risk factors (eg, tobacco and alcohol use, betel nut chewing, denture wear); (2) a user-facing mobile web app, which guides individuals step by step in capturing high-quality intraoral images using smartphone cameras; (3) an embedded AI engine, a deep learning model that detects and segments potentially malignant lesions in submitted images; and (4) a specialist review portal, which enables trained oral medicine professionals to verify AI-generated findings and provide clinical recommendations.

Images flagged as high-risk by either the AI system or the reviewing specialist are referred for follow-up, which may include a clinical examination and histopathological confirmation. All user data are securely transmitted and stored in compliance with national ethical and privacy guidelines. By integrating automated triage, guided self-screening, expert validation, and follow-up pathways, RiskOCA offers a scalable framework to enhance the early detection and surveillance of oral cancer in low-resource settings.

Platform Technology Stack

RiskOCA was built on a 3-tier architecture, which separates the system into the presentation layer (frontend), app layer (business logic), and data layer (storage and management). This modular architecture facilitates scalability, maintainability, and efficient system updates.

At the presentation tier, the front-end interface was developed using the Ionic framework, enabling cross-platform deployment on both Android and iOS devices from a single codebase. The interface uses standard web technologies (HTML5, CSS3, and JavaScript) to ensure responsiveness and consistent user experience across device types.

The application tier incorporates an application programming interface layer implemented in PHP (version 7.4.5), which manages all business logic, handles interactions between the frontend and the backend, and processes communications with the AI engine and clinician portal.

The data tier is hosted on an Apache web server and uses a MySQL relational database for storing user records, clinical metadata, AI results, and clinician annotations. The database is configured with utf8_general_ci character encoding to support multilingual content, including Thai and English. Database management is performed via phpMyAdmin (version 5.0.2), which facilitates tasks, such as schema definition, data linking, and automated backups. Together, this stack—Ionic (user interface/user experience), PHP (logic or application programming interface), and MySQL (data storage)—provides a robust, scalable, and cost-effective foundation for RiskOCA's deployment in real-world health systems.

AI Model Description and Performance

A central feature of RiskOCA is its integration of a deep learning model to enable automated, image-based screening for oral lesions. The model is based on the DeepLab v3+ architecture with a ResNet-50 backbone, a widely used configuration for semantic segmentation tasks. It was optimized for intraoral images captured in nonclinical environments, accommodating variability in lighting, positioning, and mucosal anatomy [14]. The AI system performs three key functions: (1) lesion detection—identifying the presence of any abnormal lesion in the image; (2) classification—categorizing lesions into one of 3 diagnostic classes: normal oral mucosa, OPMD, or OSCC; and (3) segmentation—generating a pixel-level mask that outlines the lesion boundaries to assist with visual interpretation and confirmatory review.

The dataset included 2591 images, each accompanied by expert-annotated segmentation masks. Images were acquired from multiple oral cavity regions using both clinical-grade single-lens reflex cameras and consumer mobile phone cameras, reflecting real-world variability in image acquisition. All images were resized to 342×512 pixels before being input to the model. To mitigate potential data leakage, substantially similar or near-duplicate intraoral photographs from the same participant were identified through manual review and consensus by 2 board-certified oral and maxillofacial pathologists and removed prior to model training and cross-validation. The model was finally trained and evaluated on 2226 labeled intraoral images (normal: n=667, OPMD: n=1310, OSCC: n=249).

A 5-fold stratified cross-validation was used to evaluate the robustness and generalizability of the DeepLab v3+model with a ResNet-50 backbone. In each fold, 20% of the data served as the held-out test set, while the remaining 80% were split into training (70%) and validation (10%). Model selection was based on validation loss, and the final performance for each fold was evaluated on the independent test set. To derive an image-level screening label from the segmentation output, a lightweight postprocessing rule was applied to summarize pixel-level predictions into one of the 3 diagnostic categories. Specifically, the predicted label was assigned according to the dominant predicted lesion class in the segmentation map, with the suppression of very small and isolated predicted regions to reduce spurious micro-segmentations. This strategy provides both an interpretable lesion outline and a single screening label suitable for community-based triage.

Across the 5 cross-validation folds, the model achieved an overall classification accuracy of 93.22 (SD 0.88%). Class-specific performance was stable and strong for normal and OPMD and clinically meaningful for OSCC. Normal classification achieved a precision of 90.82% (SD 2.41%), recall of 92.65% (SD 1.65%), and F_1-score of 91.70% (SD 1.17%). OPMD classification achieved a precision of 94.96% (SD 0.67%), recall of 94.73% (SD 1.93%), and F_1-score of 94.83% (SD 0.88%). For OSCC, the model achieved a precision of 90.78% (SD 2.16%), recall of 86.78% (SD 5.17%), and F_1-score of 88.67% (SD 3.23%).

XSL•FO
RenderX

The pooled confusion matrix across all held-out test images from the 5 folds (n=2226) is shown in Table 1. Most images were correctly classified, with diagonal counts of 618 normal, 1241 OPMD, and 216 OSCC cases. The predominant OSCC error mode was misclassification as OPMD (24/249, 9.6%), whereas the misclassification of OSCC as normal occurred less frequently (9/249, 3.6%). In the context of community-based screening with specialist review, the misclassification of OSCC as OPMD would still flag abnormality and prompt follow-up, whereas OSCC misclassified as normal represents the more clinically concerning miss category.

Table . Pooled confusion matrix (all test images across 5 folds; counts).

Ground truth	Prediction		
	Normal, n (%)	OPMD[a], n (%)	OSCC[b], n (%)
Normal	618 (92.7)	42 (6.3)	7 (1.0)
OPMD	54 (4.1)	1241 (94.7)	15 (1.2)
OSCC	9 (3.6)	24 (9.6)	216 (86.8)

[a]OPMD: oral potentially malignant disorders.

[b]OSCC: oral squamous cell carcinoma.

Segmentation performance was consistent across folds, with a pixel-wise micro mean intersection over union of 92.32% (SD 0.55%) and a micro mean F_1-score of 96.01% (SD 0.30%). Per-class pixel intersection over union was the highest for normal (95.75%, SD 0.28%) and remained moderate for lesion classes (OPMD: 69.30%, SD 1.28%; OSCC: 65.15%, SD 4.34%). The lower and more variable intersection over union for OSCC indicates greater heterogeneity in lesion boundaries, which is expected given the diversity of tumor morphology and extent.

Security and Privacy

RiskOCA was designed with a security-first approach, emphasizing data confidentiality, integrity, and user consent. Prior to using the platform, participants must review and electronically consent to a clearly worded form detailing the study purpose, data handling policies, potential risks, and the voluntary nature of participation.

PII, including full name, national identification number, contact information, and demographic details, is collected during user registration to support clinical follow-up, referral coordination, and linkage with health care services. However, no PII is attached to image data at the point of image submission or during AI-based risk assessment. Instead, each participant is assigned a unique alphanumeric study ID, which is used to link clinical images and questionnaire data while decoupling them from direct personal identifiers.

To further protect privacy, all submitted images are stripped of embedded metadata, including GPS location, device identifiers, and timestamps, prior to storage and analysis. All data transmissions occur over HTTPS using TLS 1.3 encryption. PII and anonymized clinical data are stored in logically separated, encrypted MySQL databases, accessible only to authorized personnel under role-based access controls.

Clinical reviewers and AI developers are provided access only to anonymized cases and are not able to view names, national ID numbers, or contact information. All system access, AI outputs, and expert reviews are logged and auditable. RiskOCA complies with national data protection regulations and aligns with international best practices for health data governance. Participants may request the deletion of their data at any time via an embedded secure request form on the platform.

User Interface and Experience Design

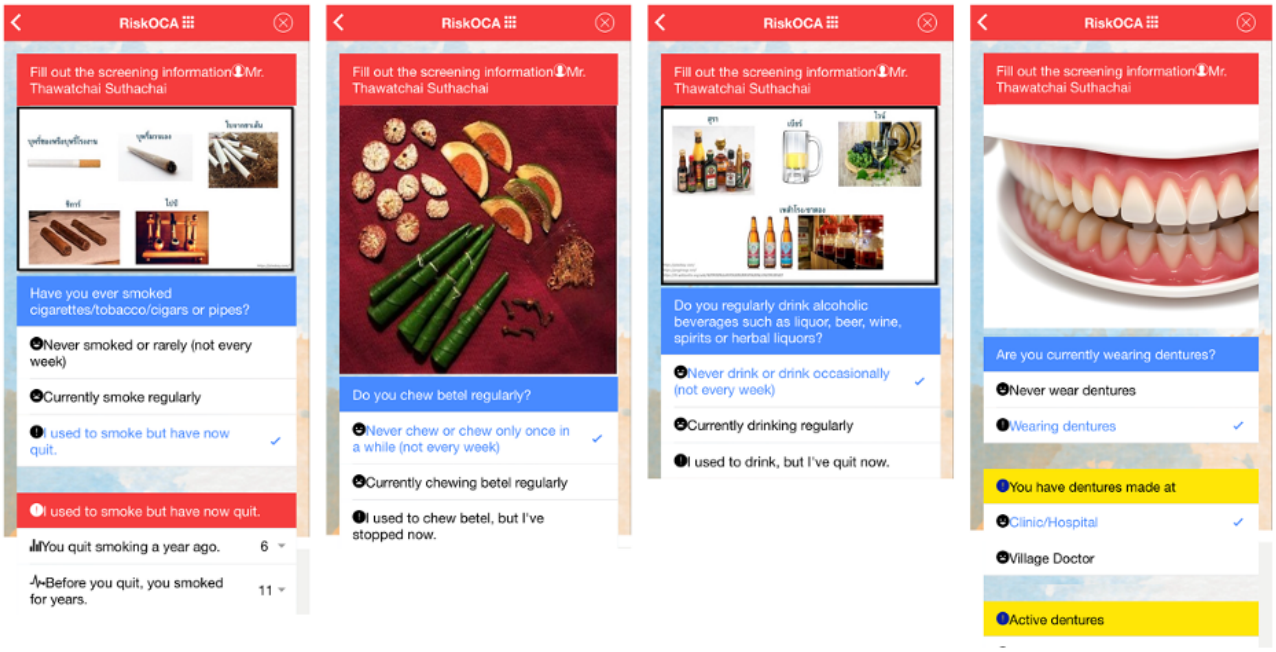
Patient Interface for Risk Assessment

The RiskOCA platform was designed using a user-centered design approach, with a strong emphasis on accessibility, cultural relevance, and ease of navigation for diverse user populations, particularly in rural and underserved areas. The patient-facing interface supports guided data entry, risk assessment, and health education through intuitive, mobile-optimized screens. The app is designed to be used independently by patients or with assistance from VHVs, ensuring inclusivity for users with varying levels of digital literacy.

Upon launching the app, users are prompted to enter their full name and national ID number for clinical identification and follow-up purposes, which is essential for individuals undergoing screening for oral potentially malignant disorders or oral cancer. Real-time input validation is implemented to prevent incorrect or incomplete ID entry. Once validated, users proceed by selecting their province, district, and sub-district, and then enter additional demographic information, including phone number, gender, date of birth, and ethnicity. These entries are facilitated through standardized input menus, including dropdowns and scrollable selection wheels optimized for mobile devices (Figure 1).

XSL•FO
RenderX

Figure 1. Oral cancer risk assessment user interface.



Following demographic data entry, users are guided through a structured oral cancer risk assessment, consisting of four culturally and clinically relevant domains [15]:

1. Tobacco smoking: Users are asked if they have ever smoked cigarettes, cigars, pipes, or used tobacco leaves. Based on their response—"Never or only occasionally," "Currently smoke regularly," or "Used to smoke but quit"—the interface dynamically adapts to show follow-up questions. For current users, sliders allow the input of number of units smoked per day and duration of smoking (years). Former users are asked how many years ago they quit and for how long they smoked.
2. Betel nut chewing: Users indicate whether they have regularly chewed betel nut. If applicable, the app collects data on frequency per day and duration in years, again using slider-based inputs. Those who quit are asked about cessation timelines.
3. Alcohol consumption: The app asks whether the user currently drinks alcoholic beverages, such as spirits, beer, wine, or herbal liquors. For current drinkers, follow-up fields collect details about the type of alcohol, frequency (daily, weekly, and monthly), and duration. Former drinkers are asked to specify the duration of prior use and time since quitting.
4. Wearing dentures: Users report whether they currently wear dentures. Those who do are asked about the source of fabrication (clinic or hospital vs local practitioner), fit quality (well-fitting or loose), and duration of denture use in years.

Each question concludes with a visual confirmation indicator (✓) and a "Confirm" button, ensuring users understand and acknowledge their responses before proceeding.

After completing the risk assessment, the app displays personalized feedback based on the presence of any risk factors. If risk factors are identified, a warning message is shown: "You should receive an oral screening by a dentist, or dental health professional." Users are prompted to enter their current address for follow-up and are provided with a link to self-examination instructions for oral cancer. If no risk factors are present, the message displayed is: "You are not in a risk group, but it is recommended to have an oral health check-up every 6 months." The app then concludes with a thank-you message.

Throughout the interface, the design prioritizes simplicity and responsiveness, using mobile-friendly controls such as sliders, scroll menus, and contextual prompts. This ensures that the app can be comfortably used by individuals with varying levels of digital literacy.

Clinical Screening Interface

The RiskOCA platform includes a secure, role-specific interface designed for the members of the oral health care team—including licensed dentists and trained VHVs—to facilitate oral cancer screening. To ensure authenticated and traceable access, users must complete a 1-time registration process by providing their full name, national identification number, affiliated health care facility, and, for dentists, their professional license number.

Upon successful login, users are directed to the patient entry screen, where they input the full name and national identification number of the individual undergoing screening. Once patient identification is confirmed, VHVs—under dentist supervision—can record lesion findings by selecting from 9 predefined anatomical sites commonly associated with OPMDs and oral cancer:

1. Upper and lower lips
2. Right and left buccal mucosa
3. Upper and lower gingiva (gums)
4. Retromolar trigone (area behind the last molars)
5. Hard palate
6. Soft palate
7. Dorsal and lateral surfaces of the tongue

XSL•FO
RenderX

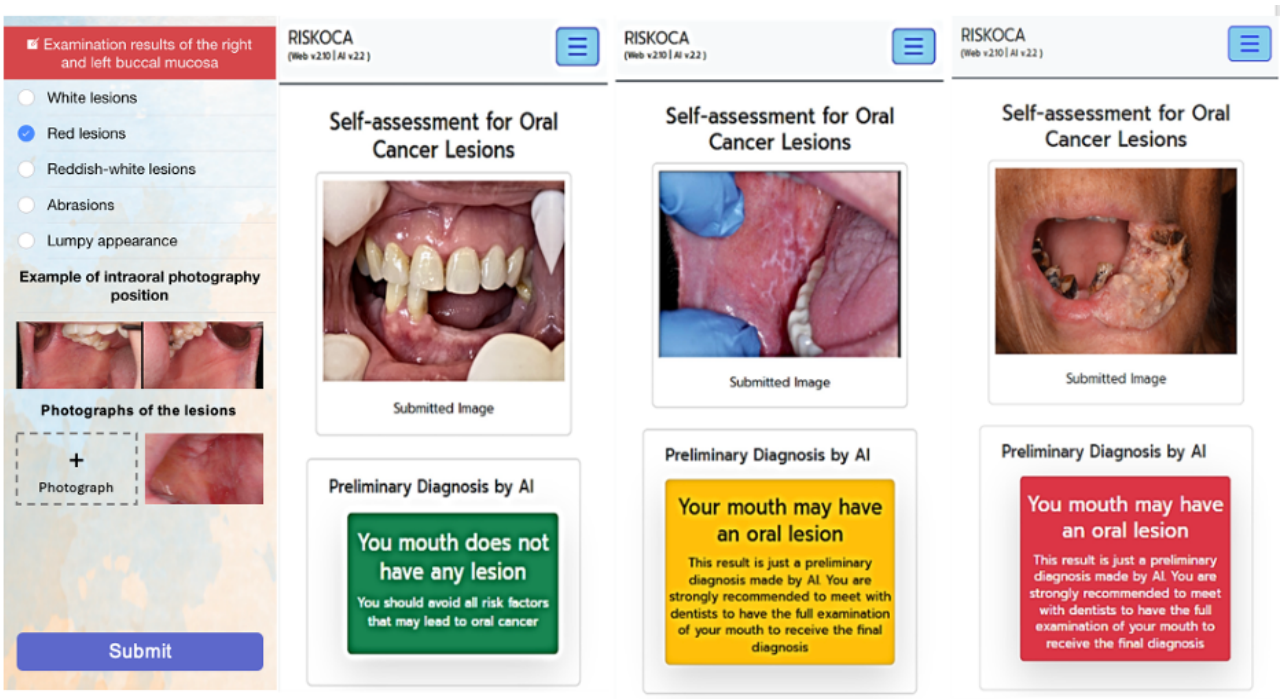
8. Ventral (underside) of the tongue
9. Floor of the mouth

Each lesion is then visually categorized into 1 of 5 clinical types:

1. White lesion
2. Red lesion
3. Mixed red and white lesion
4. Ulcerative lesion
5. Mass or lump

Photographs of the lesions are captured using a smartphone under natural lighting, with auxiliary light sources used as needed. Images captured were obtained exclusively using mobile devices operated by VHVs. To reduce variability and support consistent image quality, VHVs received structured training on capturing intraoral photographs from 9 predefined anatomical sites, with specific guidance on lighting conditions, image focus, framing, and resolution requirements prior to uploading images to the RiskOCA platform. Despite these measures, differences in clinical context and acquisition conditions between training and field images may influence model performance and are acknowledged as a consideration for generalizability. These images are securely transmitted to the Intercenter Center for Oral Health server, where they undergo AI-based evaluation and expert review to support the referral and care pathway (Figure 2).

Figure 2. Oral examiner and artificial intelligence (AI) analysis interface.



AI Output and Patient Feedback Interface

Upon image submission, the embedded AI model performs real-time analysis to classify the case into 1 of the 3 risk categories: no visible lesion, suspected OPMD, and suspected OSCC. These preliminary classifications are immediately displayed to the user in the form of a simple, color-coded message and visual overlay indicating lesion location (if applicable). To ensure clinical safety and reduce false positives or negatives, all images—regardless of AI output—are forwarded to participating oral health specialists for manual validation through the dentist interface.

If the AI detects a suspected OPMD or OSCC, the app flags the case and initiates a follow-up protocol. Users with flagged findings receive a notification recommending further evaluation. In parallel, system administrators contact the participant directly and facilitate referral to the nearest designated oral cancer center for in-person examination and, if necessary, biopsy confirmation.

To support public awareness and encourage proactive behavior, the platform also includes embedded educational content, including short videos and visual guides. These materials explain the etiology of oral cancer, common early warning signs (eg, white or red patches, ulcers, and unexplained lumps), and evidence-based self-care recommendations. This dual function—automated risk assessment and patient education—aims to increase health literacy while expediting the path to diagnosis and treatment in high-risk individuals.

Expert Review and Diagnostic Confirmation

All images analyzed by the AI model—regardless of classification (normal, OPMD, or OSCC)—were subjected to independent expert review. Two board-certified oral medicine specialists, each with more than 5 years of clinical experience in the diagnosis and management of OPMD and oral cancer, participated in the expert review process.

The specialists independently reviewed all submitted images through a secure clinician portal without access to AI-generated classifications or outputs. Reviews were based on standard clinical visual assessment of intraoral photographs. Cases identified by either the AI model or specialists as suspicious for OPMD or OSCC were referred for further clinical evaluation and biopsy confirmation when indicated.

XSL•FO
RenderX

Among the 1242 adult community members screened, the AI model flagged 8 cases as suspected OPMD and 8 cases as suspected OSCC. Independent specialist review identified 10 cases of OPMD and 6 cases of OSCC. Biopsy-confirmed diagnoses demonstrated that 0.81% (n=10) of the participants had OPMD and 0.48% (n=6) had OSCC with clinically significant lesions.

There were no instances in which normal cases were misclassified as normal by either the AI model or the specialists, and no OSCC cases were missed by either approach. Two cases initially classified by the AI model as OSCC were subsequently identified by specialists and biopsy-confirmed as OPMD, representing the primary source of discordance. No cases classified as normal by the AI model or specialists were later found to harbor a clinically significant disease.

For cases deemed suspicious based on the AI model or specialists, the RiskOCA administrative team contacted the participants and coordinated referral to the nearest designated oral cancer center. There, patients underwent comprehensive clinical evaluation, including biopsy and histopathological analysis, which served as the diagnostic gold standard. Histopathological confirmation allowed for definitive diagnosis, and in several cases, the final outcomes extended beyond the initial screening categories of OPMD and OSCC—highlighting the complexity and spectrum of oral mucosal pathology encountered in community-based screenings. This multitiered workflow—integrating AI triage, expert review, and histological confirmation—ensures clinical accuracy while maintaining accessibility for early detection in at-risk populations.

User Evaluation

Community Participants for Screening

A total of 1242 adult community members residing in the Phu Kamyao district, Phayao province, were recruited to undergo oral cancer screening using the RiskOCA app. This rural district was selected due to its limited access to oral health care services and the high prevalence of behavioral risk factors associated with oral cancer. Participants were eligible if they were aged 40 years or older, had permanent residency in the study area, were capable of providing informed consent, and were able to use a smartphone independently or with assistance from a trained volunteer.

Prior to the screening activities, all VHVs participated in a structured training session designed to ensure safe and consistent use of the RiskOCA app. The training consisted of a 1.5-hour session that included (1) a lecture on basic oral cancer knowledge, common clinical presentations, and associated risk factors; (2) an introduction to the RiskOCA app interface and workflow for risk assessment; and (3) hands-on practice using a mouth simulation model. During the hands-on component, VHVs were trained—under dentist supervision—to capture standardized intraoral photographs and to record lesion-related findings within the RiskOCA app. Following the training, VHVs facilitated participant recruitment, introduced the RiskOCA app, and supported community members during image capture and questionnaire completion as needed.

Village Health Volunteers for Usability Evaluation

A total of 250 VHVs participated in the usability evaluation of the RiskOCA app. These VHVs were the frontline users responsible for guiding community members through the self-screening process, and they played a key role in oral cancer risk surveillance in the region.

All VHVs were recruited from the same rural district and had prior experience with basic health education and mobile technology. As part of the study, they received hands-on training in using the app, including functionalities related to user registration, image capture, data entry, and interpretation of AI-generated outputs.

Following their use of the app in real-world screening activities, VHVs completed a structured usability questionnaire designed to assess the app's interface, functionality, alert system, ease of use, and perceived value for oral cancer surveillance. Their feedback was used to evaluate the app's practicality, acceptability, and potential for broader community-based deployment.

Usability Evaluation

To evaluate the usability and user satisfaction of the RiskOCA app, a structured questionnaire was administered to all participating village health volunteers following their use of the platform. The questionnaire assessed key domains, including interface design, functional performance, automated alert features, ease of use, perceived impact on community surveillance, and overall satisfaction.

User satisfaction and usability were assessed using a 25-item questionnaire adapted from the Mobile App Usability Questionnaire (MAUQ) [16]. The MAUQ aligns with established usability principles, including effectiveness, efficiency, and satisfaction, as outlined in ISO 9241‐210 human-centered design standards. The original MAUQ framework was supplemented with context-specific items addressing oral cancer risk assessment and older adult surveillance to reflect the intended use of the RiskOCA platform in community-based screening. The content validity of the adapted instrument was evaluated by a panel of 6 experts, including public health specialists, oral cancer clinicians, and mHealth or informatics experts. Item-level content validity index values ranged from 0.87 to 0.92, and the scale-level content validity index was 0.91. Minor wording revisions were made following expert review to enhance clarity and contextual appropriateness. A pilot test involving 25 community participants was conducted to assess clarity, feasibility, and reliability. The average time required to complete the questionnaire was approximately 5 minutes, and no major difficulties were reported. Internal consistency reliability was evaluated using Cronbach α, with an overall coefficient of 0.89 and domain-specific values ranging from 0.87 to 0.92, indicating good-to-excellent reliability across all domains.

The instrument was organized into six domains based on thematic relevance:

XSL•FO
RenderX

1. General usage: This domain evaluated users' perceptions of the screen layout and the logical sequencing of on-screen functions during navigation.
2. Functionality: This domain covered the ability to enter and retrieve data during and after examinations, collect risk factor information, raise user awareness regarding oral cancer, and summarize screening outcomes across all oral anatomical sites.
3. Automated alert system: This domain assessed the system's ability to generate timely prompts, such as identifying missing or incomplete records, assisting with national identification number verification, and alerting users when data were omitted.
4. Ease of use: Items in this domain focused on minimizing data entry burden, supporting note-taking features for enhanced documentation, and evaluating the system's technical responsiveness (eg, loading speed and screen transitions).
5. Impact on older adult surveillance: This domain explored the app's utility in identifying risk factors, reducing the cost of screening and care, and improving data reliability compared to traditional paper–based methods.
6. Overall satisfaction: This final domain gauged users' general impressions of the app, its usefulness in promoting older adult participation, its effectiveness compared to conventional methods, and their willingness to recommend or reuse the app in future screening activities.

Each item was rated on a 5-point Likert scale, with response options ranging from 1=very dissatisfied, 2=dissatisfied, 3=moderately satisfied, 4=satisfied, to 5=very satisfied. This scale enabled the quantifiable measurement of subjective perceptions regarding the app's effectiveness, usability, and practical relevance in a community screening context. This comprehensive Likert-scale–based evaluation enabled the robust analysis of the system's usability across multiple dimensions. The approach also accounted for users with varying levels of digital literacy, including those less experienced with smartphone technology. Mean scores were calculated for each domain as well as for the full scale. Descriptive statistics were used to analyze participant responses across all 25 items. In addition, subgroup analyses were conducted to explore differences in satisfaction scores by age group and self-rated smartphone usage ability. Mean scores across subgroups were compared using one-way ANOVA. Statistical significance was set at $P<.05$. All analyses were conducted using SPSS (version 22.0; SPSS Inc).

Results

General Information of App Users

This satisfaction assessment was conducted as part of the project titled "Development of a Mobile Application for Self-Surveillance of Potentially Malignant and Oral Cancer Lesions Among the Elderly and At-Risk Groups." The evaluation took place at the Phu Kamyao district, Phayao province, Thailand. According to Table 2, all 250 training participants were village health volunteers. The majority (50.8%) of them were aged between 30 and 39 years and had a self-rated ability and experience in using smartphones at a "fair" level (49.6%).

Table . General information of app users (n=250).

General information	Users, n (%)
Age (y)	
20 - 29	25 (10)
30 - 39	127 (50.8)
40 - 49	54 (21.6)
50 - 59	32 (12.8)
≥60	12 (4.8)
Total	250 (100)
Smartphone usage ability and experience	
Beginner	95 (38.0)
Fair	124 (49.6)
Moderate	24 (9.6)
Good	7 (2.8)
Expert	0 (0)
Total	250 (100)

User Satisfaction Results

The analysis of the mean satisfaction scores across all 6 domains revealed consistently positive user perceptions of the app. The impact on older adult surveillance domain received the highest average score (mean 4.30, SD 0.75), suggesting that users strongly agreed that the app enhanced their ability to identify risk factors, improved data reliability compared to paper-based methods, and contributed to reducing the cost of screening and care. The overall satisfaction domain also achieved a high rating (mean 4.21, SD 0.68), indicating strong endorsement of the

XSL•FO
RenderX

app's usefulness, ease of use, and potential for broader implementation and recommendation (Table 3).

XSL•FO
RenderX

Table . Percentage of user satisfaction levels toward the oral cancer screening app among older adult participants (n=250).

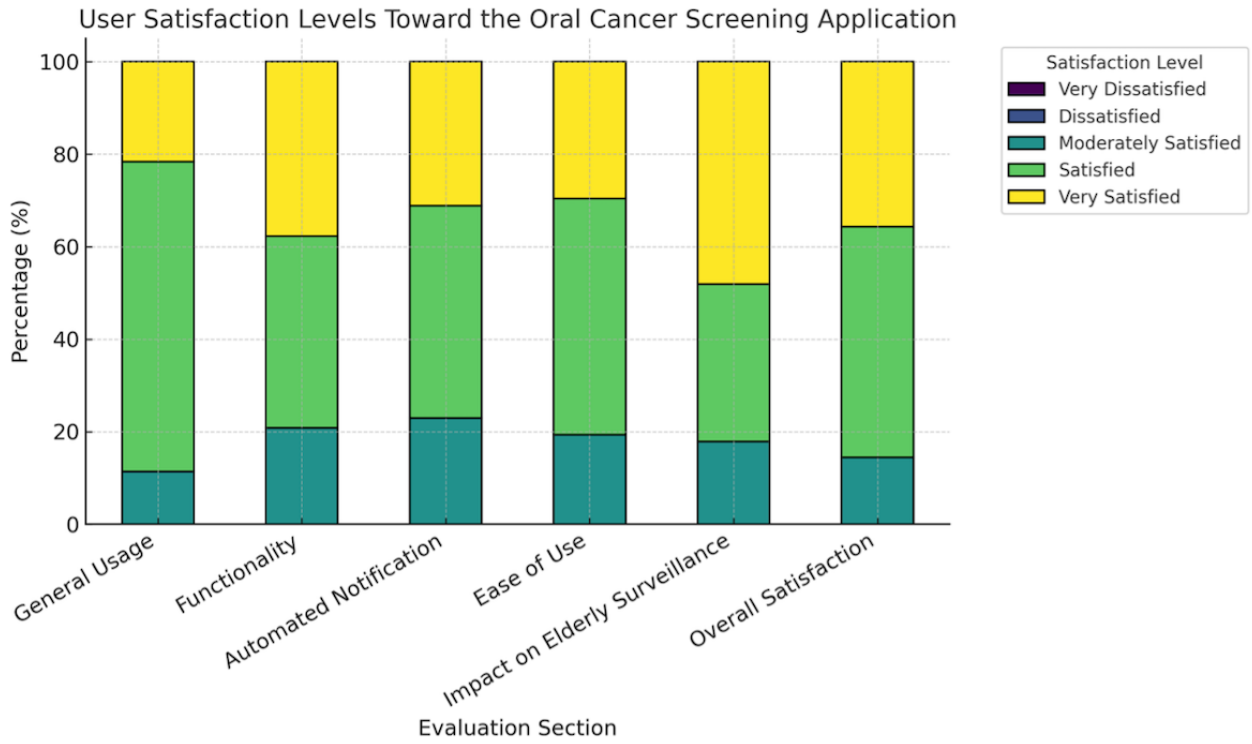
Section and item	Very dissatisfied, n (%)	Dissatisfied, n (%)	Moderately satisfied, n (%)	Satisfied, n (%)	Very satisfied, n (%)	Average score (out of 5)
General usage						
1. Screen layout	0 (0)	0 (0)	34 (13.6)	153 (61.2)	63 (25.2)	4.12
2. Order of functions displayed	0 (0)	0 (0)	23 (9.2)	182 (72.8)	45 (18.0)	4.09
Subtotal (1-2)	0 (0)	0 (0)	57 (11.4)	335 (67.0)	108 (21.6)	4.10
Functionality						
3. Data entry during screening	0 (0)	0 (0)	56 (22.4)	98 (39.2)	96 (38.4)	4.16
4. Retrieval of past records	0 (0)	0 (0)	55 (22.0)	121 (48.4)	74 (29.6)	4.08
5. Collecting risk factor data	0 (0)	0 (0)	34 (13.6)	108 (43.2)	108 (43.2)	4.30
6. Raising awareness for oral cancer prevention	0 (0)	0 (0)	63 (25.2)	88 (35.2)	99 (39.6)	4.14
Subtotal (3-6)	0 (0)	0 (0)	208 (20.8)	415 (41.5)	377 (37.7)	4.17
Automated notification system						
7. Alerts for data entry issues	0 (0.0)	0 (0)	63 (25.2)	112 (44.8)	75 (30)	4.05
8. Identifying ID and detecting data changes	0 (0)	0 (0)	45 (18.0)	131 (52.4)	74 (29.6)	4.12
9. Alert for incomplete entries	0 (0)	0 (0)	64 (25.6)	101 (40.4)	85 (34.0)	4.08
Subtotal (7-9)	0 (0)	0 (0)	172 (22.93)	344 (45.87)	234 (31.20)	4.08
Ease of use						
10. Minimal data entry required	0 (0)	0 (0)	51 (20.4)	125 (50)	74 (29.6)	4.09
11. Additional notes entry	0 (0)	0 (0)	49 (19.6)	132 (52.8)	69 (27.6)	4.08
12. App responsiveness	0 (0)	0 (0)	45 (18.0)	126 (50.4)	79 (31.6)	4.14
Subtotal (10-12)	0 (0)	0 (0)	145 (19.33)	383 (51.07)	222 (29.6)	4.10
Impact on older adult surveillance						
13. Risk factor screening support	0 (0)	0 (0)	12 (4.8)	110 (44.0)	128 (51.2)	4.46
14. Reducing screening and care costs	0 (0)	0 (0)	70 (28.0)	67 (26.8)	113 (45.2)	4.17
15. Reliability compared to paper-based recording	0 (0)	0 (0)	52 (20.8)	78 (31.2)	120 (48.0)	4.27
Subtotal (13-15)	0 (0)	0 (0)	134 (17.87)	255 (34.0)	361 (48.13)	4.30
Overall satisfaction						
16. Helps assess oral cancer risk	0 (0)	0 (0)	20 (8.0)	145 (58.0)	85 (34.0)	4.26
17. Promotes elderly participation	0 (0)	0 (0)	36 (14.4)	58 (23.2)	156 (62.4)	4.48
18. Improves quality over paper methods	0 (0)	0 (0)	45 (18.0)	126 (50.4)	79 (31.6)	4.14
19. Easy to use	0 (0)	0 (0)	31 (12.4)	119 (47.6)	100 (40)	4.28
20. Willingness to reuse the app	0 (0)	0 (0)	29 (11.6)	78 (31.2)	143 (57.2)	4.46
21. Recommend to 1 - 5 colleagues	0 (0)	0 (0)	99 (39.6)	118 (47.2)	33 (13.2)	3.74
22. Recommend to 6 - 30 colleagues	0 (0)	0 (0)	38 (15.2)	123 (49.2)	89 (35.6)	4.20
23. Recommend to more than 30 colleagues	0 (0)	0 (0)	9 (3.6)	167 (66.8)	74 (29.6)	4.26
24. Recommend to entire team	0 (0)	0 (0)	30 (12.0)	156 (62.4)	64 (25.6)	4.14

XSL•FO
RenderX

Section and item	Very dissatisfied, n (%)	Dissatisfied, n (%)	Moderately satisfied, n (%)	Satisfied, n (%)	Very satisfied, n (%)	Average score (out of 5)
25. Overall satisfaction with the app	0 (0)	0 (0)	25 (10)	155 (62.0)	70 (28.0)	4.18
Subtotal (16-25)	0 (0)	0 (0)	362 (14.48)	1245 (49.8)	893 (35.72)	4.21
Overall mean score						4.17

The stacked bar chart (Figure 3) illustrates the percentage of older participants reporting different levels of satisfaction with the oral cancer screening app. Responses were collected using a 5-point Likert scale (very dissatisfied to very satisfied) across 6 key domains: general usage, functionality, automated notification, ease of use, impact on older adult surveillance, and overall satisfaction. The majority of users reported being "satisfied" or "very satisfied" in all domains, with the highest satisfaction observed in the "impact on elderly surveillance" domain. No participants reported being dissatisfied or very dissatisfied.

Figure 3. User satisfaction levels toward the oral cancer screening app across 6 evaluation domains (n=250).



The domains of functionality (mean 4.17, SD 0.75), ease of use (mean 4.10, SD 0.69), and general usage (mean 4.10, SD 0.57) were all rated in the "very good" range, reflecting high user satisfaction with the system's operational features, visual structure, and user interface design. The automated alert system domain scored slightly lower (mean 4.08, SD 0.73) but still within the "very good" threshold, suggesting that users appreciated the built-in prompts and notifications but may have encountered occasional challenges with error messages or input validation mechanisms.

Overall, the combined mean score across all domains was 4.17 (SD 0.68), indicating a high level of satisfaction among users. These findings affirm that the app was not only functionally effective but also well received by users, including those with limited digital proficiency.

Subgroup Analysis of User Satisfaction

Subgroup analyses were conducted to examine differences in overall satisfaction scores by age group and self-rated smartphone experience (Table 4). No statistically significant differences in overall satisfaction were observed across age groups or across levels of smartphone experience. These findings indicate that user satisfaction with the RiskOCA app was consistent regardless of age or prior smartphone proficiency. Importantly, mean satisfaction scores across all subgroups remained within the "very good" range, suggesting that the app was well accepted even among VHVs with limited smartphone experience. This consistency supports the usability of the app across a heterogeneous community health workforce.

XSL•FO
RenderX

Table . Mean overall satisfaction scores by age group and smartphone experience among village health volunteers (n=250)[a].

Subgroup	Number, n	Mean (SD)	*P* value
Age group (y)			.27
20 - 29	25	4.19 (0.71)	
30 - 39	127	4.21 (0.85)	
40 - 49	54	4.20 (0.75)	
50 - 59	32	4.11 (0.62)	
≥60	12	4.15 (0.53)	
Smartphone experience			.31
Beginner	95	4.14 (0.70)	
Fair	124	4.16 (0.63)	
Moderate or good	31	4.21 (0.59)	

[a]*P* values were calculated using 1-way ANOVA.

Implementation Challenges and Practical Considerations

Although overall usability was high, several implementation challenges were observed during field deployment. Among the 250 VHVs participating in the usability evaluation, all were able to complete the core screening workflow, including participant data entry, intraoral image capture, and image submission. Based on field observations, the complete screening process was typically completed within approximately 15 minutes per participant, including cases where assistance was required.

No errors in data entry or image submission were observed during the evaluation period. However, a subset of VHVs required assistance when using the RiskOCA app. The most commonly observed challenges included difficulty in capturing adequately focused intraoral images (n=25), inconsistent lighting conditions during image acquisition (n=12), and unfamiliarity with smartphone interface navigation (n=9). In such instances, real-time assistance was provided by supervising dental staff or trained team members, particularly during the initial use of the app.

Minor technical issues, including temporary internet connectivity interruptions and delayed image uploads, were occasionally encountered in areas with unstable network coverage. These issues were typically resolved through repeated upload attempts or by completing submissions once connectivity was restored. No critical system failures, data losses, or breaches of data integrity were observed during the study period.

Discussion

This study presents a comprehensive evaluation of the technical development and usability of RiskOCA, a smartphone-based oral cancer screening app deployed among VHVs in a rural Thai setting. The findings demonstrate that the app is both feasible and well-accepted, with consistently high satisfaction scores across domains, such as functionality, ease of use, and overall system performance. Importantly, the platform was designed not only as a digital tool but as part of a broader community–based strategy to strengthen the early detection of oral potentially malignant disorders and oral cancer in resource-limited settings.

User acceptance is a critical determinant of success in the development and deployment of mHealth apps [17]. Even technically robust systems may fail to achieve impact if end users find them difficult to use, culturally inappropriate, or poorly integrated into their daily workflows. Previous studies on digital health tools have highlighted that usability, perceived usefulness, and trust strongly influence adoption, particularly in community-based and resource-limited settings [18,19]. In the context of oral cancer screening, where nonspecialist users, such as VHVs, play a central role, ensuring high levels of satisfaction and ease of use is essential for long-term sustainability. The positive feedback from VHVs in this study, therefore, provides an encouraging indication that RiskOCA is acceptable to frontline health workers and has the potential to be successfully scaled in real-world public health systems.

The involvement of VHVs as primary users underscores the value of leveraging task-shifting models in public health. Thailand has long integrated VHVs into its universal health coverage framework, and their role in RiskOCA parallels the use of community health workers in other global initiatives, such as mobile-based cervical cancer screening in sub-Saharan Africa [20] or breast cancer awareness apps in South Asia [21]. Similar to these initiatives, RiskOCA illustrates how mHealth interventions can expand screening coverage, raise the awareness of modifiable risk factors, and create a direct digital pathway between at-risk populations and oral health specialists.

From an implementation perspective, RiskOCA was designed as a modular mHealth platform to support scalability and future functional extensions beyond oral cancer triage. Key design considerations include a flexible system architecture, standardized data structures, and role-based user interfaces that can accommodate additional AI models, clinical workflows, and screening use cases. This design approach enables the potential integration of AI-assisted screening for other oral and maxillofacial conditions, such as oral potentially malignant disorders surveillance, periodontal risk assessment, or referral

XSL•FO
RenderX

prioritization for other noncommunicable oral diseases. Anticipating such extensions also underscores the importance of governance frameworks, clinician oversight, and continuous model validation to ensure safety, equity, and clinical relevance as new AI functionalities are introduced within community-based mHealth systems.

A key strength of RiskOCA lies in its potential for scalability and sustainability. The modular architecture allows adaptation across regions, while its integration into existing VHV workflows ensures cost-effectiveness and cultural alignment. Moreover, the dual-layer verification—AI-assisted triage supplemented by universal expert review—enhances diagnostic reliability and reduces the likelihood of false negatives, thereby supporting more timely referral and treatment. By embedding digital innovation into community-driven workflows, RiskOCA has the capacity to reduce diagnostic delays and narrow inequities in specialist access for rural populations.

Despite these strengths, several limitations should be acknowledged. First, only participants flagged as abnormal by the AI system and/or specialists underwent biopsy confirmation; therefore, not all 1242 participants received a gold-standard diagnostic assessment. As a result, true sensitivity and specificity of RiskOCA in the community setting cannot be fully determined, and undetected false-negative cases may exist. Second, variability in image quality—due to differences in lighting, positioning, and smartphone hardware—may affect AI performance. While the deep learning model demonstrated strong accuracy, the greater interpretability of AI outputs is needed to improve user trust and clinical decision support. The current reliance on stable internet connectivity may limit deployment in areas with poor network coverage. Addressing these issues will be critical for large-scale implementation. Another limitation of this study is that usability evaluation relied primarily on self-reported satisfaction measures, and objective usability metrics—such as task completion time per participant, standardized error rates, and formal usability testing protocols—were not systematically recorded. Future studies should incorporate structured usability testing frameworks to quantitatively assess efficiency, learnability, and error recovery, particularly as the platform is scaled to broader community and clinical settings. In addition, this study did not include a direct comparison between AI-assisted screening and traditional oral cancer screening methods conducted by VHVs. Future research should therefore consider randomized controlled trials in which VHVs are assigned to use either the RiskOCA app or conventional visual oral examination protocols. Such comparative designs would allow the rigorous evaluation of screening effectiveness, including detection rates of suspicious lesions, referral accuracy, diagnostic concordance with dentists, and biopsy-confirmed outcomes. Evidence from these trials would be essential to establish the added clinical value of AI-enabled screening over existing community–based practices.

Although a formal cost-effectiveness analysis was beyond the scope of this usability study, the RiskOCA app has the potential to offer cost advantages compared with traditional oral cancer screening pathways. The development of RiskOCA was funded by the Ministry of Public Health of Thailand, with the intention that the app be provided free of charge to users and community health workers. As a result, no direct costs are incurred by villagers or VHVs for app access or routine screening activities. In contrast, traditional oral cancer screening often requires patients to travel to health care facilities or necessitates outreach visits by dentists and health care teams to community settings, both of which involve transportation costs, personnel time, and logistical coordination. By enabling VHVs to conduct preliminary screening using widely available smartphones, RiskOCA may reduce unnecessary referrals, optimize the use of specialist resources, and lower indirect costs associated with delayed diagnosis. Nonetheless, future studies incorporating randomized controlled designs should include formal economic evaluations to quantify cost-effectiveness, taking into account implementation, maintenance, training, and health care system–level outcomes. In addition, structured questionnaire surveys and mixed method evaluations capturing the perspectives, perceived challenges, and recommendations of both VHVs and dentists would provide valuable implementation insights, supporting contextual adaptation and scale-up of similar mHealth initiatives across diverse health care settings.

This study was conducted in a single rural district in northern Thailand, which may limit the generalizability of the findings to other regions with different demographic characteristics, health care infrastructures, digital literacy levels, or patterns of oral cancer risk. Although the setting reflects a high-need population with limited access to oral health care, further studies are warranted to evaluate the performance, usability, and implementation feasibility of the RiskOCA platform in urban settings, other rural regions, and across different health system contexts. Multisite and cross-regional evaluations will be essential to confirm the scalability and adaptability of this community-based AI-assisted screening approach.

Future development should also focus on larger-scale deployment and longitudinal evaluation to validate RiskOCA's impact on referral rates, confirmed diagnoses, and patient outcomes. Integration with national eHealth records could support coordinated care and surveillance, while the development of multilingual interfaces and culturally adapted educational modules would enhance accessibility for ethnolinguistic minorities. Together, these efforts will be essential to ensuring RiskOCA's adaptability, effectiveness, and sustainability in diverse low- and middle-income country contexts.

In summary, this study demonstrates that RiskOCA is technically feasible and acceptable to frontline health volunteers. By bridging community-level screening with specialist expertise, the platform represents a scalable and policy-relevant model for strengthening early oral cancer detection in underserved populations. Its successful deployment highlights the broader potential of AI-integrated mHealth systems to transform community-based cancer prevention and control globally.

XSL•FO
RenderX

Acknowledgments

The authors thank all the authorized individuals as well as community health volunteers of the Phu Kamyao district, Phayao province, Thailand, providing important information and helping to collect data.

Funding

This research was supported in part by the National Research Council of Thailand, Chiang Mai University, and the Intercountry Centre for Oral Health (Department of Health, Ministry of Public Health, Thailand).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

MW, KW, TS, DT, JI, PW, and SS conceptualized and designed the study. MW and SS supervised the study. PW was responsible for the artificial intelligence model development and evaluation. DT performed the statistical analysis. MW, TS, and SS drafted the manuscript, and all authors critically reviewed and approved the final version for publication.

Conflicts of Interest

None declared.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021 May;71(3):209-249. [doi: 10.3322/caac.21660] [Medline: 33538338]
2. Global Cancer Observatory (GCO): Cancer Today. International Agency for Research on Cancer, World Health Organization. URL: https://gco.iarc.who.int/today [accessed 2026-02-18]
3. Warnakulasuriya S. Global epidemiology of oral and oropharyngeal cancer. Oral Oncol 2009;45(4-5):309-316. [doi: 10.1016/j.oraloncology.2008.06.002] [Medline: 18804401]
4. Warnakulasuriya S. Oral potentially malignant disorders: a comprehensive review on clinical aspects and management. Oral Oncol 2020 Mar;102:104550. [doi: 10.1016/j.oraloncology.2019.104550] [Medline: 31981993]
5. Papachristou Nadal I, Aramrat C, Jiraporncharoen W, et al. Process evaluation protocol of a cluster randomised trial for a scalable solution for delivery of Diabetes Self-Management Education in Thailand (DSME-T). BMJ Open 2021 Dec 9;11(12):e056141. [doi: 10.1136/bmjopen-2021-056141] [Medline: 34887283]
6. Nawsuwan K, Singweratham N, Suriya N, Oupra R. Solving area-based problems through intellectual empowerment: a model for developing village health volunteers' and public health officials' competencies in producing academic work. Sci Rep 2025 May 20;15(1):17537. [doi: 10.1038/s41598-025-01997-x] [Medline: 40394183]
7. Nagao T, Warnakulasuriya S. Screening for oral cancer: future prospects, research and policy development for Asia. Oral Oncol 2020 Jun;105:104632. [doi: 10.1016/j.oraloncology.2020.104632] [Medline: 32315954]
8. Kanmodi KK, Salami AA, Shah K, Zohoori FV, Nnyanzi LA. The types and effectiveness of mobile health applications used in improving oral cancer knowledge: a mixed methods systematic review. Health Sci Rep 2024;7(11):e70171. [doi: 10.1002/hsr2.70171] [Medline: 39474342]
9. Donkor A, Ayitey JA, Adotey PN, et al. Mobile-based application interventions to enhance cancer control and care in low- and middle-income countries: a systematic review. Int J Public Health 2023;68:1606413. [doi: 10.3389/ijph.2023.1606413] [Medline: 38125709]
10. Warin K, Limprasert W, Suebnukarn S, Jinaporntham S, Jantana P, Vicharueang S. AI-based analysis of oral lesions using novel deep convolutional neural networks for early detection of oral cancer. PLoS ONE 2022;17(8):e0273508. [doi: 10.1371/journal.pone.0273508] [Medline: 36001628]
11. Acharya A, Bhat S, Sultana N, et al. AI-based analysis of oral lesions using novel deep convolutional neural networks for early detection of oral cancer. Oral Oncol 2022;134:106192. [doi: 10.1016/j.oraloncology.2022.106192] [Medline: 35995005]
12. Guo J, Li B. The application of medical artificial intelligence technology in rural areas of developing countries. Health Equity 2018;2(1):174-181. [doi: 10.1089/heq.2018.0037] [Medline: 30283865]
13. Wongsapai M, Wudtijureepun K, Tamdee D, Suthachai T, Wuttisarnwattana P, Suebnukarn S. Oral cancer surveillance in remote areas using a smartphone digital platform. Stud Health Technol Inform 2025 Aug 7;329:1898-1899. [doi: 10.3233/SHTI251270] [Medline: 40776287]
14. Wuttisarnwattana P, Wongsapai M, Theppitak S, et al. Precise identification of oral cancer lesions using artificial intelligence. Stud Health Technol Inform 2024 Aug 22;316:1096-1097. [doi: 10.3233/SHTI240601] [Medline: 39176572]
15. Irani S. New insights into oral cancer-risk factors and prevention: a review of literature. Int J Prev Med 2020;11:202. [doi: 10.4103/ijpvm.IJPVM_403_18] [Medline: 33815726]

XSL•FO
RenderX

16. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth app usability questionnaire (MAUQ): development and validation study. JMIR mHealth uHealth 2019 Apr 11;7(4):e11500. [doi: 10.2196/11500] [Medline: 30973342]
17. Schliemann D, Tan MM, Hoe WMK, et al. mHealth interventions to improve cancer screening and early detection: scoping review of reviews. J Med Internet Res 2022 Aug 15;24(8):e36316. [doi: 10.2196/36316] [Medline: 35969450]
18. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172. [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
19. Adnan A, Irvine RE, Williams A, Harris M, Antonacci G. Improving acceptability of mHealth apps-the use of the technology acceptance model to assess the acceptability of mHealth apps: systematic review. J Med Internet Res 2025 May 7;27:e66432. [doi: 10.2196/66432] [Medline: 40334265]
20. Brinkel J, Krämer A, Krumkamp R, May J, Fobil J. Mobile phone-based mHealth approaches for public health surveillance in sub-Saharan Africa: a systematic review. Int J Environ Res Public Health 2014 Nov 12;11(11):11559-11582. [doi: 10.3390/ijerph111111559] [Medline: 25396767]
21. Hamid F, Roy T. Unveiling sociocultural barriers to breast cancer awareness among the South Asian population: case study of Bangladesh and West Bengal, India. JMIR Hum Factors 2025 Jan 10;12:e53969. [doi: 10.2196/53969] [Medline: 39793010]

Abbreviations

AI: artificial intelligence
MAUQ: Mobile App Usability Questionnaire
mHealth: mobile health
OPMD: oral potentially malignant disorder
OSCC: oral squamous cell carcinoma
PII: personally identifiable information
RiskOCA: Risk Assessment for Oral Cancer using Artificial Intelligence
VHV: village health volunteer

Edited by R Chandrasekaran; submitted 08.Sep.2025; peer-reviewed by K Nozaki, O Igunma; revised version received 06.Feb.2026; accepted 09.Feb.2026; published 09.Mar.2026.

Please cite as:
Wongsapai M, Wudtijureepun K, Suthachai T, Tamdee D, Ittichaicharoen J, Wuttisarnwattana P, Suebnukarn S
A Community-Based Usability Study of an AI-Enabled Oral Cancer Screening App Operated by Village Health Volunteers: Mixed Methods Study
JMIR Mhealth Uhealth 2026;14:e83738
URL: https://mhealth.jmir.org/2026/1/e83738
doi:10.2196/83738

© Mansuang Wongsapai, Kornwipa Wudtijureepun, Thawatchai Suthachai, Decha Tamdee, Jitjiroj Ittichaicharoen, Patiwet Wuttisarnwattana, Siriwan Suebnukarn. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 9.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Timed Up-and-Go Test Using Instrumented Insoles to Monitor Tibial or Malleolar Fracture Healing in Patients With Healing Fractures or Nonunion: Longitudinal and Cross-Sectional Exploratory Study

Elke Warmerdam, PhD; Lea-Marie Burger, MD; Bergita Ganse, MD, PhD

Innovative Implant Development (Fracture Healing), Departments and Institutes of Surgery, Saarland University, Kirrberger Str. 100, Building 57, Homburg, Saarland, Germany

Corresponding Author:
Bergita Ganse, MD, PhD
Innovative Implant Development (Fracture Healing), Departments and Institutes of Surgery, Saarland University, Kirrberger Str. 100, Building 57, Homburg, Saarland, Germany

Abstract

Background: The timed up-and-go (TUG) test is a mobility-related functional task commonly performed in clinical practice for a variety of clinical applications. By instrumenting the TUG test with sensors, creating an instrumented TUG (iTUG), additional parameters can be extracted, which may add clinical value. Instrumented insoles are among the most promising wearable devices that are currently being explored for clinical purposes.

Objective: The study aims to assess additional parameters obtained by instrumenting the TUG test with pressure-sensing insoles for monitoring fracture healing and to analyze whether these parameters reveal differences between patients with and without union of tibial or malleolar fractures.

Methods: Patients with tibial or malleolar fractures and matched controls were enrolled. Patients' TUG test data were collected as follows: (1) longitudinally—at 6 weeks, 3 months, and 6 months after surgery—and (2) cross-sectionally—at 6 months after surgery when patients were referred for nonunion treatment. All participants performed the TUG test using instrumented insoles. TUG tests were video-recorded and analyzed in terms of the time taken to complete each part of the TUG test (conventional TUG). The insole data-derived outcome measures were maximal force during sit-to-stand and stand-to-sit movements, stance time and stride time during walking, and asymmetry of the previously mentioned parameters (iTUG).

Results: Twenty-seven patients with fracture, a similar number of healthy adults, and 9 patients with fracture nonunion were assessed. All TUG parameters improved throughout the healing process (all $P<.001$). The greatest improvements were detected between 6 weeks and 3 months (all $P<.001$). Among the classical TUG parameters, between 3 and 6 months, only the total TUG time, walking time, stand-to-sit time, and total number of steps significantly improved ($P=.004$, $P=.01$, $P=.03$, and $P=.002$). Based on the additional iTUG measures, 6 months after surgery, the loading of the injured side was reduced in patients compared with healthy controls during sit-to-stand and stand-to-sit movements ($P<.001$). Six months after surgery, patients with fracture nonunion had significantly poorer performance in terms of all of the TUG test parameters, except for stance time, than patients with fracture union.

Conclusions: This study indicates the potential for using functional outcomes to identify patients at risk of nonunion. Additional tasks and functional outcomes require further investigation.

Trial Registration: German Clinical Trials Registry DRKS00025108; https://drks.de/search/en/trial/DRKS00025108

(JMIR Mhealth Uhealth 2026;14:e92493) doi:10.2196/92493

KEYWORDS

mobility; nonunion; tibia; wearables; rehabilitation; mHealth; digital medicine

Introduction

Fracture Nonunion

Bone fracture nonunion is a serious complication associated with high socioeconomic costs [1]. It is estimated that diaphyseal nonunion of the tibia and femur occurs in 4.6% to 8% of patients following intramedullary nailing of closed fractures performed in accordance with current standards, whereas the risk is even higher in open fractures [1]. Risk factors for nonunion include severe fractures (eg, open fractures and multiple fractures), a high BMI, smoking, and alcoholism [2]. While women experience fractures more frequently, men are more prone to

XSL•FO
RenderX

nonunion [2]. The nonunion rate is greatest in the navicular bone, tibia and fibula, and femur [2]. In routine clinical practice, radiographs are used to monitor fracture healing. Considering that delayed healing can lead to nonunion, patients are often monitored for an extended period, and interventions such as ultrasound stimulation or shock wave therapy are sometimes prescribed with the aim of inducing union [3]. Revision surgery with a bone graft is usually arranged once healing within a reasonable time seems unlikely [4]. Considering that mineralization of the callus in and around the fracture gap does not take place immediately, radiographs can only show bone healing with a time delay [5]. Moreover, radiographs require X-rays, which are carcinogenic [6]. To find better alternatives, functional outcome measures, such as patient-reported outcome measures, have been suggested to monitor fracture healing [7-9]. However, no correlation was shown between differences in adverse event rates and differences in patient-reported outcome measures in orthopedic trauma randomized controlled trials [10]. Although analyses of movement quality are rarely performed in daily clinical practice for the quantification of actual physical function, they may be suitable for predicting fracture nonunion, as discussed in the literature [11-13].

The TUG Test

The timed up-and-go (TUG) test is a simple, short functional assessment that covers multiple mobility-related aspects. It has high reliability and validity [14]. This test includes rising from a chair, walking (straight and turning), and sitting down. The outcome measure of this test is the time needed to complete the test. When the TUG test is video-recorded, additional outcome measures can be assessed via video analysis, including the time it takes to perform each subtask and the number of steps [14]. The TUG test is known to be a reliable measure in different populations, such as older adults and patients with a history of a stroke, spinal cord injury, or cerebral palsy [15]. Endurance training, resistance training, and physical therapy are known to reduce the time it takes for older adults, patients with chronic stroke, and patients with Parkinson disease to perform the TUG test [16-18]. The time it takes to complete the TUG test can provide information about future fracture risk above that provided by bone mineral density, as it has been shown in 2 studies [19,20].

In recent decades, sensors, such as inertial measurement units and instrumented insoles, have been increasingly used during TUG tests [21,22]. These additions make it possible to extract additional outcome parameters, such as gait parameters [23,24], joint kinematics [25], angular velocities [26], and accelerations [27]. Compared to the traditional TUG test, the predictive value of the instrumented TUG (iTUG) was higher for fall prediction owing to the additional extracted, mainly temporal and velocity, parameters [22]. Patients with idiopathic normal-pressure hydrocephalus, Parkinson disease, and orthopedic conditions may complete the iTUG test as a means to screen for treatment and disease-related changes in mobility [21]. The results of the iTUG test after total knee or hip joint replacement improved over the course of rehabilitation and provided additional information [28,29]. However, it is not yet clear whether changes during fracture healing can similarly be quantified with the iTUG test or whether there are advantages of the iTUG test over the conventional TUG test.

Improvements After Injury

Instrumented gait analysis with motion capture systems or wearables can be used to quantify improvements in the gait pattern throughout the healing process of tibial and malleolar fractures [11,30]. There is also potential to analyze other mobility-related movements, such as standing and rising from a chair. In particular, the analysis of the vertical ground reaction force of these movements is of interest, as patients can choose how to distribute the load between the injured and uninjured sides, which is not the case with walking once they can walk without crutches. When patients begin to walk without crutches after the injury, the improvements in maximal loading during walking are rather small because they will always reach loading values of 100% of their body weight during the single-support phase. The normative values for loading lie between 100% of body weight at slow speeds and 125% at fast speeds [31]. Therefore, other movements might be even more sensitive for monitoring fracture healing at later stages of healing.

Hypotheses

Hence, this study analyzed whether the TUG test performed with instrumented insoles can be used to monitor the healing status of tibial and malleolar fractures and whether it has advantages over the conventional TUG test. This information could aid in the process of deciding when to revise or intervene during the fracture healing process. It was hypothesized that (1) longitudinal changes in mobility throughout the tibial fracture healing can be assessed in terms of iTUG test results; (2) 6 months after a tibial fracture, the performance of several iTUG parameters returns to that of sex- and age-matched healthy controls; and (3) several parameters measured in the iTUG test can distinguish between patients with and without union 6 months after a tibial fracture.

Methods

This longitudinal, observational, exploratory study was registered with the German Clinical Trials Register (DRKS-ID: DRKS00025108). The STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist is provided as a download (Checklist 1).

Participants

Patients with tibial or malleolar fractures who underwent surgical treatment at Saarland University Hospital between February 2022 and January 2024 were included and assigned to the longitudinal patient group. Patients who were younger than 18 years could not provide informed consent, had limited mobility before the fracture, had other injuries that could affect mobility, or were pregnant were excluded. Patients were recruited during their hospital stay after undergoing surgical treatment for their fracture. Radiographs in 2 planes obtained 6 months after the fracture were examined by experienced trauma surgeons to determine whether the fracture had healed.

In addition, a group of patients with fracture nonunion was recruited from those treated in the outpatient nonunion clinic.

XSL•FO
RenderX

As only less than 10% of patients develop nonunion, a longitudinal study would take several years to reach a sufficient number of patients with fracture nonunion. At Saarland University Hospital, where this study was conducted, patients who were originally treated in other hospitals can present to seek advice and further treatment for their nonunion in the specialized nonunion outpatient clinic, which serves a larger region. To increase the number of patients with fracture nonunion in this study, the authors added a cross-sectional group of patients from this outpatient group, for whom longitudinal measurements were impossible, because they had initially been treated elsewhere. These patients received only one measurement, as they underwent their revision surgery shortly after, and their further treatment was again conducted at the hospital where they had originally been treated. The inclusion criteria for this cross-sectional nonunion group were surgery 4 to 10 months earlier and nonunion as diagnosed by an experienced trauma surgeon. The exclusion criteria were below 18 years of age, inability to provide informed consent, limited mobility before the fracture, other injuries that could affect mobility, or pregnancy.

Healthy controls were sex- and age-matched to the longitudinal patient group. These participants were recruited among the spouses of the patients, hospital employees, and acquaintances of the study team and were measured between February 2022 and April 2024. Healthy participants were excluded if they were below 18 years of age, were unable to provide written informed consent, or had limited mobility. They underwent only a single measurement, as changes were not expected over time without a fracture.

Study Protocol

Patients in the longitudinal patient group were assessed during each outpatient visit, which was usually scheduled approximately 6 weeks, 3 months, and 6 months after surgery. The first visit had to take place between 30 and 60 days after surgery, the second visit between 75 and 105 days, and the third visit between 135 and 235 days. Only patients who took part in all 3 assessments within these time frames were included in the analysis. In the patients with fracture nonunion group and the healthy control group, only one measurement was performed. During the experiment, all participants wore instrumented insoles (OpenGO insoles, Moticon GmbH) containing 16 pressure sensors in both shoes. Before the start of each measurement, the insoles were fitted in accordance with the shoe size of each participant and calibrated to each participant's body weight using the insole software calibration function. All patients and participants wore sports shoes with a soft sole. The participants were asked to sit down on a wooden chair measuring 45.7 centimeter in height with their back against the backrest of the chair. At the start of the TUG test, they had to rise from the chair, walk for 3 meters, turn around a marked spot on the floor, walk back to the chair, and sit down again with their back against the backrest. They were instructed to perform this assessment at their preferred speed and to turn in their preferred direction. If patients still used crutches in their daily lives, they were allowed to use crutches to perform the task. The tests were video-recorded (HERO7 Silver, GoPro Inc).

Data Analysis

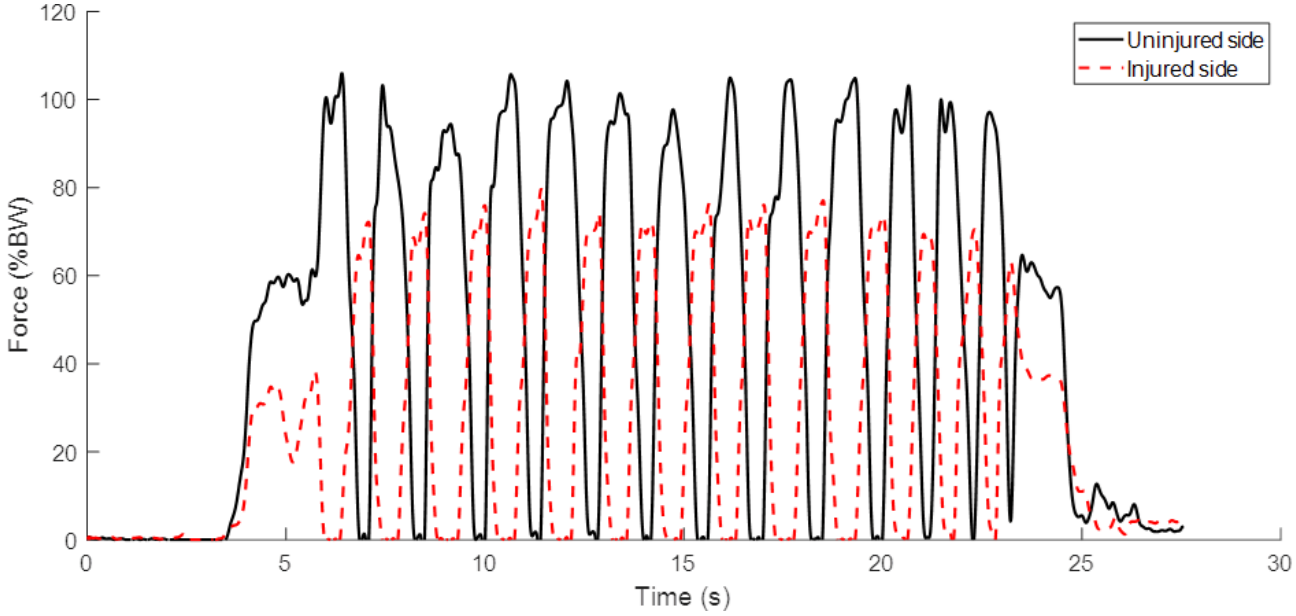
The total TUG time, time per subtask, and the number of steps were extracted from the video recordings using the open-source software Media Player Classic–Home Cinema (MPC-HC, version 1.7.13; open source, MPC Community Forum) based on Savoie et al [32], but with slight adjustments. The first video frame in which the patient moved the upper part of the trunk forward marked the start of the TUG test. The frame in which they started to walk (heel rise, which was followed by toe-off) marked the start of the next subtask. The moment the patient placed the supporting leg for the first single step, in which the hips or legs began a turning movement in their preferred turn direction, marked the start of the turn. The end of the turn was analogously defined as the touchdown of the swing leg after the last individual step in which the hips or legs were still in a turning motion. Since the turn and sit-down movements at the end of the TUG test are generally performed in combination, the start of this stand-to-sit movement was defined as the landing of the supporting leg on the first single step, when the hips or legs were again in a turning movement. The test was completed when the patient's back touched the backrest. The numbers of left and right steps were counted from the start of the TUG test (forward movement of the trunk) until the end of the TUG test (patient's back touched the backrest) and then summed.

The raw data from the insoles were downloaded with the manufacturer's software. A custom MATLAB (version 2021b; MathWorks) script was written to extract the sit-to-stand and stand-to-sit movements and the steps. Data were filtered with a fourth-order Butterworth filter with a 6-Hertz cutoff frequency. The sit-to-stand and stand-to-sit movements were manually detected in the ground reaction force curves of both feet, whereafter the peak in the signal was detected and used as the maximal force during sit-to-stand and stand-to-sit movements (see Figure 1 for an example of the forces during the TUG test). The manual detection of these movements was performed to ensure that these movements were also detected when no force was applied to the insole on the injured side or when the data from both feet were not correctly synchronized by the software. A threshold of 30 N in the vertical force data was used to detect the initial and final contact of each step that occurred between the sit-to-stand and stand-to-sit movements. The vertical force data were normalized to the percentage of body weight to allow for comparisons between participants. From the sit-to-stand and stand-to-sit movements, the maximal force was extracted for both the injured and uninjured legs. From the walking phase, the maximal force and stance and stride times were extracted for each stride. An average of these parameters was calculated per side. Additionally, the asymmetry of all of these parameters was calculated as a percentage using the following equation:

$$\text{Asymmetry} = (\text{uninjured side} - \text{injured side}) 0.5 \times (\text{uninjured side} + \text{injured side}) \times 100$$

XSL•FO
RenderX

Figure 1. Vertical ground reaction forces of a 76-year-old female patient with a malleolar fracture, measured 76 days after surgery during the timed up-and-go test. BW: body weight.



Statistical Analyses

Statistical analyses were performed with JASP (version 0.17.3.0). The data were tested for a normal distribution with the Shapiro-Wilk test. Differences in demographics between patients with union and healthy controls, as well as between patients with union and those without union, were assessed with an independent *t* test in the case of a normal distribution; otherwise, a Mann-Whitney *U* test was used. A repeated-measures ANOVA was performed to analyze whether the extracted parameters of the longitudinal patient group improved throughout the healing phase, with the extracted TUG parameters as the within-subject factors. Effect sizes were calculated according to eta-squared. Post hoc testing was performed with Bonferroni corrections. The parameters assessed at 6 months after surgery in the group of longitudinal patients with union were compared with those of healthy controls via an independent *t* test or a Mann-Whitney *U* test in cases of violation of normality. Additionally, the longitudinal patient group at 6 months was compared with patients with nonunion at 6 months after surgery via an independent *t* test or a Mann-Whitney *U* test in the absence of a normal distribution. Effect sizes were calculated using the Cohen *d* method. The magnitude of Cohen *d* was categorized as follows: small (d=0.2), medium (d=0.5), and large (d≥0.8). Significance was set at a *P* value <.05 for all tests.

Ethical Considerations

The study protocol was approved by the Institutional Review Board of the Saarland Medical Board (Ärztekammer des Saarlandes, Germany, application number 30/21). Written informed consent was obtained from all participants before the start of the measurements. Participants had the opportunity to opt out at any time without needing to provide reasons. This observational, prospective longitudinal cohort study was conducted in accordance with the Declaration of Helsinki. The data were pseudonymized. The participants did not receive any financial or nonfinancial compensation for their participation in the study.

Results

Demographics

Twenty-seven patients (Table 1) with union of a tibial or malleolar fracture were assessed 3 times during the healing process. Twenty-seven healthy adults, matched for sex and age with the longitudinal patient group, served as the control group. Additionally, 9 patients with nonunion were included in the nonunion group. There were no significant differences in demographics between the longitudinal patient group and the control group or between the longitudinal patient group and the patients with fracture nonunion group (Table 1).

XSL•FO
RenderX

Table . Demographics of all groups.

Parameter	Longitudinal patients with union	Healthy controls	Patients with fracture nonunion	*P* value	
				Longitudinal patients vs controls	Longitudinal patients vs patients with fracture nonunion
N (% female)	27 (44)	27 (44)	9 (33)	—[a]	—
Age (y), mean (SD)	54 (13)	54 (13)	59 (18)	.92	.11
Height (cm), mean (SD)	175 (9)	173 (9)	178 (9)	.67	.52
Weight (kg), mean (SD)	83 (13)	79 (15)	80 (14)	.13	.62
BMI	27 (4)	26 (4)	25 (4)	.62	.29
Fracture site (proximal tibia/tibial shaft/distal tibia/malleolar)	9/5/1/12	—	1/8/0/0	—	—

[a]Not applicable.

Longitudinal Changes in the Performance of the (Instrumented) TUG Test

The longitudinal patient group was assessed on days 42, 86, and 182, on average, after surgery. At 6 weeks after surgery, 22% (6/27) of the patients performed the test without crutches. At 3 months, this number increased to 85% (23/27), and at 6 months, none of the patients used crutches anymore. Repeated-measures ANOVA revealed significant changes in all parameters throughout healing. Post hoc tests revealed that all parameters significantly improved between 6 weeks and 3 months after surgery. Between 3 and 6 months, only the total TUG time, walking time, stand-to-sit time, and total number of steps significantly improved (Table 2 and Figures 2 and 3).

XSL•FO
RenderX

Table . Differences between time points in the longitudinal measurements (N=27).[a]

Parameter	Mean (SD)			*P* value	Effect size (η^2)	Post hoc *P* value		
	T1[b]	T2[c]	T3[d]			T1 vs T2	T1 vs T3	T2 vs T3
TUG[e] parameters extracted from video recordings								
Total time (s)	19.6 (4.5)	14.3 (3.8)	11.8 (2.6)	<.001[f]	0.696[f]	<.001[f]	<.001[f]	.004[f]
Sit-to-stand time (s)	2.8 (1.1)	1.6 (0.8)	1.3 (0.4)	<.001[f]	0.521[f]	<.001[f]	<.001[f]	.42
Walking time (s)	7.8 (2.7)	5.7 (2.0)	4.4 (1.2)	<.001[f]	0.549[f]	<.001[f]	<.001[f]	.01[f]
Turning time (s)	3.8 (0.8)	2.8 (0.8)	2.5 (0.5)	<.001[f]	0.553[f]	<.001[f]	<.001[f]	.22
Stand-to-sit time (s)	5.3 (1.1)	4.2 (1.1)	3.6 (1.0)	<.001[f]	0.555[f]	<.001[f]	<.001[f]	.03[f]
Total number of steps	18.5 (3.0)	16.2(2.8)	14.7 (2.2)	<.001[f]	0.602[f]	<.001[f]	<.001[f]	.002[f]
TUG parameters extracted from instrumented insoles								
Maximal force sit-to-stand (%BW[g])	24.7 (15.2)	44.4 (10.6)	43.0 (11.0)	<.001[f]	0.493[f]	<.001[f]	<.001[f]	>.99
Maximal force walking (%BW)	43.4 (29.6)	94.3 (16.5)	101.7 (13.1)	<.001[f]	0.737[f]	<.001[f]	<.001[f]	.50
Stance time walking (%)	51.3 (14.5)	67.6 (4.4)	67.7 (6.1)	<.001[f]	0.547[f]	<.001[f]	<.001[f]	>.99
Stride time walking (s)	1.7 (0.4)	1.4 (0.2)	1.3 (0.2)	<.001[f]	0.522[f]	<.001[f]	<.001[f]	.35
Maximal force stand-to-sit (%BW)	23.3 (18.1)	44.9 (13.0)	45.4 (11.0)	<.001[f]	0.524[f]	<.001[f]	<.001[f]	>.99
Asymmetry maximal force sit-to-stand (%)	102.0 (55.3)	36.8 (35.7)	31.6 (38.4)	<.001[f]	0.471[f]	<.001[f]	<.001[f]	>.99
Asymmetry maximal force walking (%)	89.1 (60.1)	9.8 (20.3)	0.9 (10.4)	<.001[f]	0.669[f]	<.001[f]	<.001[f]	>.99
Asymmetry stance time (%)	44.3 (33.7)	6.5 (9.1)	4.9 (14.2)	<.001[f]	0.529[f]	<.001[f]	<.001[f]	>.99
Asymmetry maximal force stand-to-sit (%)	107.3 (59.3)	45.4 (37.0)	31.5 (41.3)	<.001[f]	0.540[f]	<.001[f]	<.001[f]	.55

[a]The results of the repeated-measures ANOVA for the longitudinal patient group with union are shown.

[b]T1: 6 weeks.

[c]T2: 3 months.

[d]T3: 6 months.

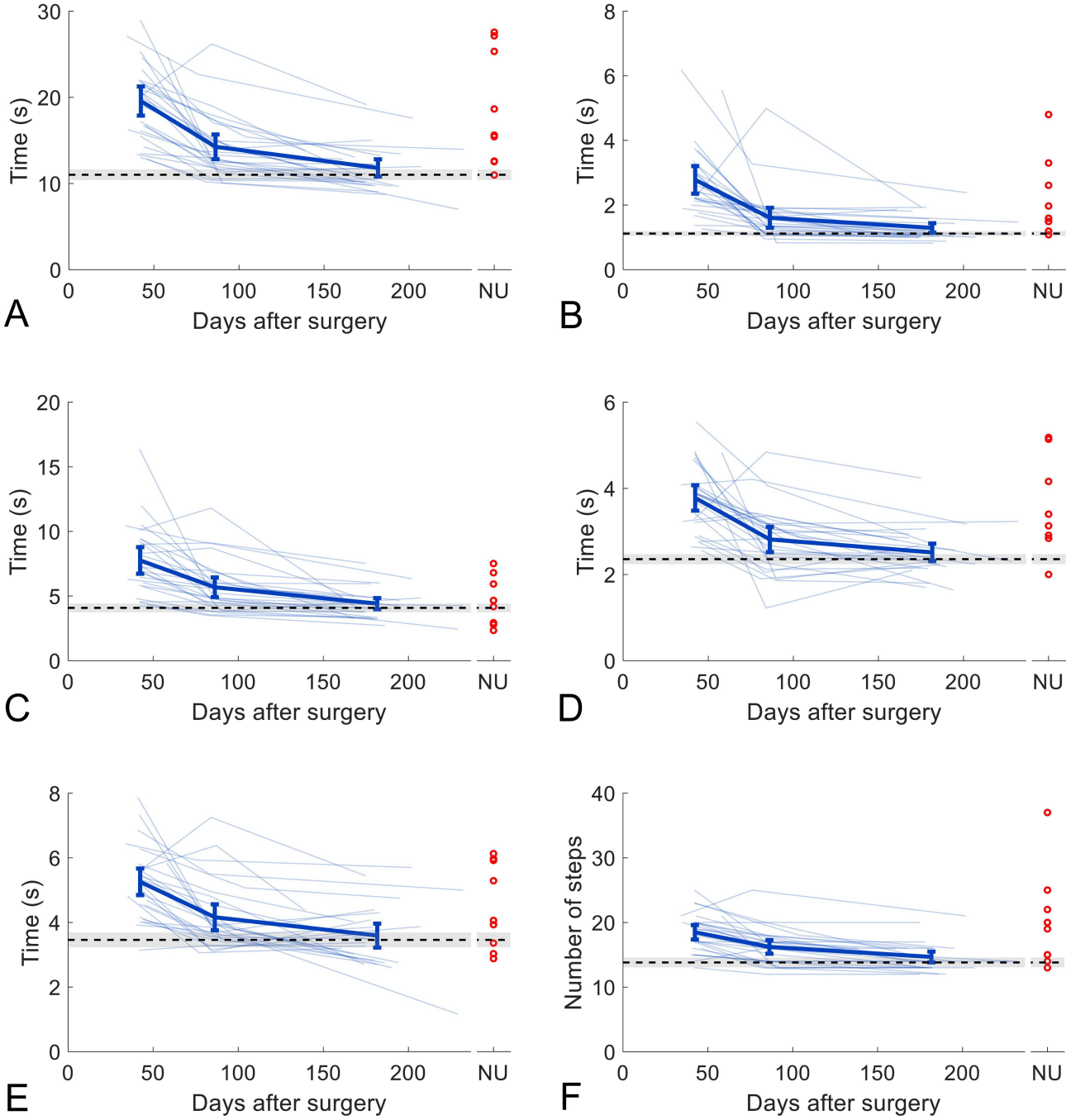
[e]TUG: timed up-and-go.

[f]These values represent significant results.

[g]BW: body weight.

XSL•FO
RenderX

Figure 2. Timed up-and-go test parameters extracted from the video recordings. (A) Total time; (B) sit-to-stand time; (C) walking time; (D) turning time; (E) stand-to-sit time; (F) total number of steps. The thin blue lines represent the results for each patient, and the thicker blue line represents the average of the longitudinal patients. The whiskers show the 95% CIs of the longitudinal data. The horizontal black dotted line represents the average of the matched controls, and the 95% CIs are in light gray. The red circles are the data points of the patients with fracture nonunion. NU: nonunion group.



The maximal force during walking had the greatest effect size, followed by the total TUG time (Table 2). The remaining effect sizes of the parameters extracted from the video recordings are comparable to those of the parameters extracted from the instrumented insoles. However, between 3 and 6 months after surgery, significant changes were observed only in the following video-based parameters: the total time, walking time, stand-to-sit time, and the total number of steps.

XSL•FO
RenderX

Figure 3. Timed up-and-go test parameters extracted from instrumented insoles. (A) Maximal force sit-to-stand; (B) maximal force walking; (C) stance time walking; (D) stride time walking; (E) maximal force stand-to-sit; (F) asymmetry maximal force sit-to-stand; (G) asymmetry maximal force walking; (H) asymmetry stance time; (I) asymmetry maximal force stand-to-sit. The thin blue lines represent the results of each patient, and the thicker blue line represents the average of the longitudinal patients. The whiskers show the 95% CIs of the longitudinal data. The horizontal black dotted line represents the average of the matched controls, and the 95% CIs are in light gray. The red circles are the data points of the patients with fracture nonunion. BW: body weight; NU: nonunion group.

XSL•FO
RenderX

Group Differences in the Performance of the TUG Test

Six months after surgery, the maximal force during the sit-to-stand and stand-to-sit movements of patients with union was significantly lower than that of the matched controls, and the asymmetry of these parameters and the stance time were significantly greater in the patients with union (Table 3). The other TUG test parameters were similar to those of healthy controls. The comparison of patients with and without union at an average of 182 and 200 days after surgery, respectively, revealed significant differences in almost all the parameters, except for stance time and stance time asymmetry (Table 3). It should be noted that 2 of the patients with fracture nonunion performed the test with crutches, whereas the patients with fracture nonunion all performed the test without crutches at 6 months after surgery. The parameters extracted from the videos showed the 3 greatest effect sizes of all parameters. These were the total time, the time walking, and the time turning. A large effect was also observed for the stand-to-sit time (Cohen d>0.8). Nonetheless, the effect sizes of the force-related parameters were also large, except for the asymmetry of the maximal force during walking. Among the forces, the largest effects were found for the maximal force sit-to-stand, maximal force walking, and the asymmetry in the maximal force sit-to-stand, but their effect sizes were smaller than those of the video-related parameters.

XSL•FO
RenderX

Table . Comparisons between longitudinal patients with union 6 months after surgery and sex- and age-matched healthy controls, as well as between longitudinal patients with union and patients with fracture nonunion 6 and 10 months after surgery, respectively.

Parameter	Mean (SD)			Patients with union vs controls		Patients with union vs patients with fracture nonunion	
	Union group at 6 months (N=27)	Healthy controls (N=27)	Nonunion group at 6 months (n=9)	*P* value	Effect size (Cohen *d*)	*P* value	Effect size (Cohen *d*)
TUG[a] parameters extracted from video recordings							
Total time (s)	11.8 (2.6)	11.0 (1.6)	18.4 (6.6)	.19	−0.359	<.001[b]	−1.675[b]
Sit-to-stand time (s)	1.3 (0.4)	1.1 (0.2)	2.2 (1.2)	.06[c]	−0.299	.005[b,c]	−0.634[b]
Walking time (s)	4.4 (1.2)	4.1 (0.9)	8.0 (3.6)	.26	−0.313	<.001[b]	−1.764[b]
Turning time (s)	2.5 (0.5)	2.4 (0.3)	3.8 (1.2)	.19	−0.359	<.001[b]	−1.689[b]
Stand-to-sit time (s)	3.6 (1.0)	3.5 (0.6)	4.5 (1.3)	.55	−0.164	.03[b]	−0.851[b]
Total number of steps	14.7 (2.2)	13.8 (2.0)	19.9 (7.6)	.11[c]	−0.248	.03[b,c]	−0.477[b]
TUG parameters extracted from instrumented insoles							
Maximal force sit-to-stand (%BW[d])	43.0 (11.0)	55.8 (8.3)	26.3 (9.7)	<.001[b]	1.313[b]	<.001[b]	1.567[b]
Maximal force walking (%BW)	101.7 (13.1)	98.7 (7.0)	81.5 (16.4)	.30	−0.286	<.001[b]	1.451[b]
Stance time walking (%)	67.7 (6.1)	69.8 (2.9)	65.9 (7.5)	.12[c]	0.246	.18[c]	0.309
Stride time walking (s)	1.3 (0.2)	1.2 (0.1)	1.5 (0.3)	.20[c]	−0.203	.02[b,c]	−0.514[b]
Maximal force stand-to-sit (%BW)	45.4 (11.0)	60.5 (7.0)	33.1 (12.2)	<.001[b]	1.623[b]	.008[b]	1.092[b]
Asymmetry maximal force sit-to-stand (%)	31.6 (38.4)	−9.1 (28.6)	94.0 (37.9)	<.001[b]	−1.202[b]	<.001[b]	−1.632[b]
Asymmetry maximal force walking (%)	0.9 (10.4)	0.2 (7.1)	21.8 (23.7)	.85[c]	−0.032	.001[b,c]	−0.704[b]
Asymmetry stance time (%)	4.9 (14.2)	−2.7 (3.7)	6.9 (15.9)	<.001[b,c]	−0.638[b]	.13[c]	−0.350
Asymmetry maximal force stand-to-sit (%)	31.5 (41.3)	−9.0 (24.2)	84.1 (36.4)	<.001[b]	−1.198[b]	.002[b]	−1.306[b]

[a]TUG: timed up-and-go.

[b]These values represent significant results of the independent *t* test.

[c]Mann-Whitney *U* test in case of the absence of a normal distribution.

[d]BW: body weight.

Discussion

Principal Results

All TUG test parameters, extracted from both video recordings and instrumented insoles, improved throughout the healing process of tibial and malleolar fractures. Six months after surgery, there were still significant differences between patients and healthy controls. Compared with that of matched controls, the maximal loading on the injured side was significantly lower during sit-to-stand and stand-to-sit movements. No differences in video-based parameters were detected between patients and

XSL•FO
RenderX

healthy controls. Patients with fracture nonunion performed significantly worse in 13 of the 15 assessed parameters of the TUG test than patients with union at 6 months after surgery.

Longitudinal Findings

The longitudinal patient group showed significant improvements in all the assessed parameters between 6 weeks and 3 months. However, between 3 months and 6 months, only 4 TUG parameters improved, all of which were video-based parameters. Most of the changes in the TUG test parameters occurred during the first 3 months. Similar findings were reported in patients with hip fractures, whose time to perform the TUG test mainly improved during the first 3 months [33]. For monitoring improvement over time, the conventional TUG test might be sufficient. However, based solely on the TUG test parameters extracted from the video recordings in this study, the performance of patients returned to normal within 6 months. Nonetheless, by instrumenting the TUG test, we were able to show that the maximal loading during sit-to-stand and stand-to-sit movements did not return to control-like values within 6 months. The time required to perform the TUG test returned to normal values within 6 months; however, the patients still tended to unload the injured side during the sit-to-stand and stand-to-sit movements, activities where they can choose how to distribute their load. The iTUG provided limited additional information compared to the video-based TUG, but the differences in loading during sit-to-stand and stand-to-sit movements may provide valuable information. Physical therapy should focus on reducing the asymmetry during these movements to prevent future complaints because of an asymmetric movement pattern. The asymmetric loading pattern could be used for further therapy planning.

The maximal loading parameters during sit-to-stand and stand-to-sit movements showed only very limited nonsignificant improvements between 3 and 6 months, and at 6 months, they were still significantly different from controls. It has previously been reported that the gait pattern 6 months after a tibial shaft fracture was still significantly different from that of healthy controls [34]. Nonetheless, no differences in temporal gait parameters were found in another cohort assessed 5 years after a tibial shaft fracture [35]. Therefore, the recovery time may be prolonged and last longer than 6 months.

Sit-to-stand and stand-to-sit movements might be valuable for monitoring the progress of fracture healing throughout the rehabilitation process because patients can choose how to distribute their weight between the injured and uninjured sides. Shortly after surgery, most patients receive partial-weight-bearing (20 kg) instructions for the first 6 weeks. At 6 weeks, patients are allowed to increase the loading of the leg. As soon as patients start to walk without walking aids (usually somewhere between 6 weeks and 3 months), they load at least 100% of their body weight on the injured side because of single-limb support phases during walking. Therefore, from that moment on, the improvements in maximal loading will be only minor. Since sit-to-stand and stand-to-sit movements are normally performed with loading on both legs, patients can choose to load the injured leg less during these movements. This is indeed the case in this study, as there is still considerable asymmetry between the loading of the injured and uninjured legs at 6 months after surgery during the sit-to-stand and stand-to-sit movements. It is not clear whether this is because of existing complaints such as pain or because patients are still very careful or even afraid to load the injured leg during these movements. Another possible cause is unawareness of unloading due to habituation. Both young and older healthy adults have good perception of weight distribution during the sit-to-stand task [36]. However, how patients with lower extremity injuries perceive their weight distribution is unknown. Furthermore, a cause for the lower maximal force could be related to decreases in strength and power after fracture. After a hip fracture, at least up to 13 weeks after surgery, women have reduced leg extension power on the injured side [37]. It is also known that the ground reaction force during the sit-to-stand movement is related to leg strength and power in older adults [38]. Therefore, potentially lower strength in the injured leg could also have led to the lower maximal force during sit-to-stand and stand-to-sit movements. The actual reason for less maximal loading of the injured side during these movements is crucial to planning physical therapy throughout the rehabilitation process.

Nonunion

The differences between patients with and without union suggest that the iTUG test might have potential to identify patients at risk of nonunion earlier than is currently possible [39,40]. However, further investigation is warranted in longitudinal studies examining the relationship between retrospective TUG results and the diagnosis of union vs nonunion. There were no changes in the instrumented insole data between 3 and 6 months, and there were only small changes in the video-extracted parameters. It might be possible to identify patients at high risk of developing nonunion at 3 months, which may provide additional information when deciding whether there is an indication for revision surgery. Based on the effect sizes, the most suitable parameters would be the total time to perform the TUG test, the total walking time, the turning time, and the force-related parameters from the instrumented insoles during the sit-to-stand and stand-to-sit movements. If patients at risk of developing nonunion can be identified earlier, additional treatment could be started earlier during the rehabilitation process, potentially preventing nonunion and the accompanying decrease in quality of life [41] and increase in societal cost [42]. Other possible noninvasive treatments that are known to improve fracture healing in more than 80% of patients are extracorporeal shockwave therapy [43], pulsed electromagnetic fields [44], and low-intensity pulsed ultrasound [45].

Alternative Tests Using Smart Insoles

Since this study has shown that sit-to-stand and stand-to-sit movements are valuable for monitoring rehabilitation progress after tibial and malleolar fractures, different assessments, such as the 5 times sit-to-stand test and the 30-second sit-to-stand test, may also be applicable for monitoring rehabilitation progress and determining whether the patients' performance returns to that of healthy adults. These tests involve multiple consecutive sit-to-stand and stand-to-sit movements, and the changes observed between consecutive sit-to-stand movements may differ throughout the rehabilitation process. This group of

XSL•FO
RenderX

parameters could have additional clinical value and should also be explored. Differences between similar repeated movements could result from fatigue or pain and are therefore expected to be greater at the beginning of the rehabilitation phase or in patients with nonunion.

The generalizability of the results cannot be judged from the present findings, but the authors believe that similar findings may be obtained from other patient collectives.

Limitations

A limitation of this study is that different types of fractures were assessed together in the analysis. Owing to the small sample of the tibial fractures, it was not possible to analyze each fracture type separately. The different fracture types might have altered the results. However, these differences are expected to be minor. In addition, only patients without mobility limitations before the fracture were included in this study, but previous minor changes in gait and movement may have affected the findings in individual patients. It is therefore important to consider the patient's mobility level before the fracture and surgery when monitoring rehabilitation. Overall, despite this being among the largest sample sizes for a nonunion group, the sample size is still small, and future studies should aim to collect larger sample sizes. Besides that, the patients with fracture nonunion were measured in a larger time frame (4-10 mo after surgery) and compared to patients with union whose measurements were taken about 6 months after surgery. This difference might have led to a chronological bias. However, since all patients with fracture nonunion had a persisting fracture gap, both groups had a similar fracture-healing status, and the effect is therefore expected to be small. Due to the small sample size, there is a risk of bias due to multiple analyses. It should be noted that muscle strength measurements are impossible in patients with fracture, as they pose a risk to the healing fracture and may result either in refracture or in implant failure. Instrumented insoles have been criticized for delivering estimates rather than measurements due to measurement error. Data have been shown to be influenced by age, body weight, body height, hand grip strength [46], the walking slope [47], and the walking surface [48]. The data delivered by instrumented insoles may be less accurate than the data recorded by force plates, and this fact should be acknowledged when interpreting the results of this study. In future studies, the pain level could be assessed each time, which was not done in this study, as the authors frequently observe patients with a nonunion who do not report any pain. However, to quantify these experiences, future studies could include such questions.

Conclusions

Both the conventional TUG test and the iTUG test can be used to monitor functional changes throughout the healing process of tibial and malleolar fractures. Instrumenting the TUG test with pressure-sensing insoles provided additional information, which made it possible to detect differences in loading during sit-to-stand and stand-to-sit movements between patients assessed 6 months after their surgery and controls. Most of the extracted TUG test parameters differed between the patients with and without union 6 months after surgery. These differences could be detected with both the conventional TUG test and the iTUG test. The TUG test might have the potential to identify patients at risk of developing nonunion earlier. This study indicates the potential of using functional outcomes to identify patients at risk of nonunion. Further investigation of additional tasks and functional outcomes is needed.

Acknowledgments

The authors acknowledge Dr Sonja Baumgartner, Jan Kattanek, Andrey Lange, Dr Jan Laqua, Diana Mergen, and Dr Caroline Stoll for assistance with patient recruitment and measurements. No generative AI was used in the manuscript.

Funding

This work was funded by the Werner Siemens Foundation for the project "Smart Implants 2.0." The funder had no involvement in the study design, data collection, analysis, interpretation, or the writing of the manuscript.

Data Availability

The dataset analyzed during the present study is available from the corresponding author upon reasonable request. Access may be granted based on a collaboration agreement. The requesting institution needs to fall within the eligibility criteria under German data protection law.

Authors' Contributions

Conceptualization: EW, BG
Data curation: LMB
Formal analysis: EW
Investigation: EW, LMB
Project administration: BG
Resources: BG
Supervision: BG
Visualization: EW
Writing – original draft: EW, BG

XSL•FO
RenderX

Writing – review & editing: LMB

Conflicts of Interest

None declared.

Checklist 1
STROBE checklist.
[PDF File, 158 KB - mhealth_v14i1e92493_app1.pdf]

References

1. Rupp M, Biehl C, Budak M, Thormann U, Heiss C, Alt V. Diaphyseal long bone nonunions—types, aetiology, economics, and treatment recommendations. Int Orthop 2018 Feb;42(2):247-258. [doi: 10.1007/s00264-017-3734-5] [Medline: 29273837]
2. Zura R, Xiong Z, Einhorn T, et al. Epidemiology of fracture nonunion in 18 human bones. JAMA Surg 2016 Nov 16;151(11):e162775. [doi: 10.1001/jamasurg.2016.2775] [Medline: 27603155]
3. Searle HKC, Lewis SR, Coyle C, Welch M, Griffin XL. Ultrasound and shockwave therapy for acute fractures in adults. Cochrane Database Syst Rev 2023 Mar 3;3(3):CD008579. [doi: 10.1002/14651858.CD008579.pub4] [Medline: 36866917]
4. Gómez-Barrena E, Ehrnthaller C. Long bone uninfected non-union: grafting techniques. EFORT Open Rev 2024 May 10;9(5):329-338. [doi: 10.1530/EOR-24-0032] [Medline: 38726992]
5. Blokhuis TJ, de Bruine JH, Bramer JA, et al. The reliability of plain radiography in experimental fracture healing. Skeletal Radiol 2001 Mar;30(3):151-156. [doi: 10.1007/s002560000317] [Medline: 11357453]
6. Mettler FA, Huda W, Yoshizumi TT, Mahesh M. Effective doses in radiology and diagnostic nuclear medicine: a catalog. Radiology 2008 Jul;248(1):254-263. [doi: 10.1148/radiol.2481071451] [Medline: 18566177]
7. Carney DD, Vyas PS, Hicks JJ, et al. Effect of postoperative immobilization time on PROMIS scores and clinical outcomes in ankle fracture patients. Foot Ankle Orthop 2023 Jan;8(1):24730114221151080. [doi: 10.1177/24730114221151080] [Medline: 36741682]
8. Houwen T, de Munter L, Lansink KWW, de Jongh MAC. There are more things in physical function and pain: a systematic review on physical, mental and social health within the orthopedic fracture population using PROMIS. J Patient Rep Outcomes 2022 Apr 6;6(1):34. [doi: 10.1186/s41687-022-00440-3] [Medline: 35384568]
9. Williams BR, McCreary DL, Parikh HR, Albersheim MS, Cunningham BP. Improvement in functional outcomes after elective symptomatic orthopaedic implant removal. J Am Acad Orthop Surg Glob Res Rev 2020 Sep;4(9):e20.00137. [doi: 10.5435/JAAOSGlobal-D-20-00137] [Medline: 32890009]
10. Parry JA, Salimi M, Stacey SC, Mauffrey C. There is no correlation between differences in adverse events and differences in patient-reported outcome measures in orthopaedic trauma randomized controlled trials: a systematic review. Injury 2024 Nov;55(11):111909. [doi: 10.1016/j.injury.2024.111909] [Medline: 39332228]
11. Warmerdam E, Orth M, Pohlemann T, Ganse B. Gait analysis to monitor fracture healing of the lower leg. Bioengineering (Basel) 2023 Feb 15;10(2):255. [doi: 10.3390/bioengineering10020255] [Medline: 36829749]
12. Warmerdam E, Laqua J, Kattanek J, Ganse B. Differences between tibial or malleolar fracture types and union or nonunion in spatiotemporal and kinematic gait parameters throughout healing: an observational study. Ann Biomed Eng 2026 Apr;54(4):1164-1174. [doi: 10.1007/s10439-025-03937-2] [Medline: 41423533]
13. Pajchert Kozłowska A, Pawik Ł, Szelerski Ł, et al. Dynamic gait parameters in patients with nonunion of the tibia following treatment with the Ilizarov method. J Biomech Eng 2021 Nov 1;143(11):111004. [doi: 10.1115/1.4051432] [Medline: 34114601]
14. Kamnardsiri T, Thawinchai N, Parameyong A, et al. Conventional video-based system for measuring the subtask speed of the timed up and go test in older adults: validity and reliability study. PLoS One 2023;18(6):e0286574. [doi: 10.1371/journal.pone.0286574] [Medline: 37267315]
15. Christopher A, Kraft E, Olenick H, Kiesling R, Doty A. The reliability and validity of the timed up and go as a clinical tool in individuals with and without disabilities across a lifespan: a systematic review. Disabil Rehabil 2021 Jun;43(13):1799-1813. [doi: 10.1080/09638288.2019.1682066] [Medline: 31656104]
16. Gamborg M, Hvid LG, Dalgas U, Langeskov-Christensen M. Parkinson's disease and intensive exercise therapy—an updated systematic review and meta-analysis. Acta Neuro Scandinavica 2022 May;145(5):504-528. [doi: 10.1111/ane.13579] [Medline: 34997759]
17. Nguyen PT, Chou LW, Hsieh YL. Proprioceptive neuromuscular facilitation-based physical therapy on the improvement of balance and gait in patients with chronic stroke: a systematic review and meta-analysis. Life (Basel) 2022 Jun 13;12(6):882. [doi: 10.3390/life12060882] [Medline: 35743913]
18. Ren Y, Lin C, Zhou Q, Yingyuan Z, Wang G, Lu A. Effectiveness of virtual reality games in improving physical function, balance and reducing falls in balance-impaired older adults: a systematic review and meta-ana. Arch Gerontol Geriatr 2023 May;108:104924. [doi: 10.1016/j.archger.2023.104924] [Medline: 36680968]

XSL•FO
RenderX

19. Jeong SM, Shin DW, Han K, et al. Timed up-and-go test is a useful predictor of fracture incidence. Bone 2019 Oct;127:474-481. [doi: 10.1016/j.bone.2019.07.018] [Medline: 31323429]
20. Zhu K, Devine A, Lewis JR, Dhaliwal SS, Prince RL. " 'Timed up and go' test and bone mineral density measurement for fracture prediction. Arch Intern Med 2011 Oct 10;171(18):1655-1661. [doi: 10.1001/archinternmed.2011.434] [Medline: 21987195]
21. Böttinger MJ, Labudek S, Schoene D, et al. "TiC-TUG": technology in clinical practice using the instrumented timed up and go test—a scoping review. Aging Clin Exp Res 2024 Apr 27;36(1):100. [doi: 10.1007/s40520-024-02733-7] [Medline: 38676844]
22. Ortega-Bastidas P, Gómez B, Aqueveque P, Luarte-Martínez S, Cano-de-la-Cuerda R. Instrumented timed up and go test (iTUG)—more than assessing time to predict falls: a systematic review. Sensors (Basel) 2023 Mar 24;23(7):3426. [doi: 10.3390/s23073426] [Medline: 37050485]
23. Seo M, Shin MJ, Park TS, Park JH. Clinometric gait analysis using smart insoles in patients with hemiplegia after stroke: pilot study. JMIR mHealth uHealth 2020 Sep 10;8(9):e22208. [doi: 10.2196/22208] [Medline: 32909949]
24. Smith E, Cunningham C, Greene BR, McCarthy Persson U, Blake C. Detecting subtle mobility changes among older adults: the quantitative timed up and go test. Aging Clin Exp Res 2021 Aug;33(8):2157-2164. [doi: 10.1007/s40520-020-01733-7] [Medline: 33098079]
25. Kataoka Y, Saito Y, Takeda R, et al. Evaluation of lower-limb kinematics during timed up and go (TUG) test in subjects with locomotive syndrome (LS) using wearable gait sensors (H-Gait system). Sensors (Basel) 2023 Jan 6;23(2):687. [doi: 10.3390/s23020687] [Medline: 36679484]
26. Picardi M, Redaelli V, Antoniotti P, et al. Turning and sit-to-walk measures from the instrumented timed up and go test return valid and responsive measures of dynamic balance in Parkinson's disease. Clin Biomech (Bristol) 2020 Dec;80:105177. [doi: 10.1016/j.clinbiomech.2020.105177] [Medline: 32979787]
27. Buraschi R, Pollet J, Villafañe JH, Piovanelli B, Negrini S. Temporal and kinematic analyses of timed up and go test in chronic low back pain patients. Gait Posture 2022 Jul;96:137-142. [doi: 10.1016/j.gaitpost.2022.05.027] [Medline: 35635989]
28. Bloomfield RA, Williams HA, Broberg JS, Lanting BA, McIsaac KA, Teeter MG. Machine learning groups patients by early functional improvement likelihood based on wearable sensor instrumented preoperative timed-up-and-go tests. J Arthroplasty 2019 Oct;34(10):2267-2271. [doi: 10.1016/j.arth.2019.05.061] [Medline: 31255408]
29. Perelgut ME, Polus JS, Lanting BA, Teeter MG. The effect of femoral stem collar on implant migration and clinical outcomes following direct anterior approach total hip arthroplasty. Bone Joint J 2020 Dec;102-B(12):1654-1661. [doi: 10.1302/0301-620X.102B12.BJJ-2019-1428.R1] [Medline: 33249909]
30. Warmerdam E, Wolff C, Orth M, Pohlemann T, Ganse B. Long-term continuous instrumented insole-based gait analyses in daily life have advantages over longitudinal gait analyses in the lab to monitor healing of tibial fractures. Front Bioeng Biotechnol 2024;12:1355254. [doi: 10.3389/fbioe.2024.1355254] [Medline: 38497053]
31. Senden R, Marcellis R, Willems P, Witlox M, Meijer K. Normative 3D gait data of healthy adults walking at three different speeds on an instrumented treadmill in virtual reality. Data Brief 2024 Apr;53:110230. [doi: 10.1016/j.dib.2024.110230] [Medline: 38445200]
32. Savoie P, Cameron JAD, Kaye ME, Scheme EJ. Automation of the timed-up-and-go test using a conventional video camera. IEEE J Biomed Health Inform 2020 Apr;24(4):1196-1205. [doi: 10.1109/JBHI.2019.2934342] [Medline: 31403450]
33. Beckmann M, Bruun-Olsen V, Pripp AH, Bergland A, Smith T, Heiberg KE. Recovery and prediction of physical function 1 year following hip fracture. Physiother Res Int 2022 Jul;27(3):e1947. [doi: 10.1002/pri.1947] [Medline: 35332627]
34. Kröger I, Müßig J, Brand A, et al. Recovery of gait and function during the first six months after tibial shaft fractures. Gait Posture 2022 Jan;91:66-72. [doi: 10.1016/j.gaitpost.2021.09.199] [Medline: 34653876]
35. Larsen P, Eriksen CB, Stokholm R, Elsoe R. Results following prolonged recovery show satisfactory functional and patient-reported outcome after intramedullary nailing of a tibial shaft fracture: a prospective 5-year follow-up cohort study. Arch Orthop Trauma Surg 2021 Aug;141(8):1303-1310. [doi: 10.1007/s00402-020-03608-y] [Medline: 32951059]
36. Lauzière S, Brière A, Nadeau S. Perception of weight-bearing distribution during sit-to-stand task in healthy young and elderly individuals. Percept Mot Skills 2010 Aug;111(1):187-198. [doi: 10.2466/03.15.25.PMS.111.4.187-198] [Medline: 21058599]
37. Portegijs E, Sipilä S, Rantanen T, Lamb SE. Leg extension power deficit and mobility limitation in women recovering from hip fracture. Am J Phys Med Rehabil 2008 May;87(5):363-370. [doi: 10.1097/PHM.0b013e318164a9e2] [Medline: 18303470]
38. Tsuji T, Tsunoda K, Mitsuishi Y, Okura T. Ground reaction force in sit-to-stand movement reflects lower limb muscle strength and power in community-dwelling older adults. Int J Gerontol 2015 Jun;9(2):111-118. [doi: 10.1016/j.ijge.2015.05.009]
39. Özkan S, Nolte PA, van den Bekerom MPJ, Bloemers FW. Diagnosis and management of long-bone nonunions: a nationwide survey. Eur J Trauma Emerg Surg 2019 Feb;45(1):3-11. [doi: 10.1007/s00068-018-0905-z] [Medline: 29335752]
40. Qvist AH, Væsel MT, Jensen CM, Jakobsen T, Jensen SL. Minimal pain decrease between 2 and 4 weeks after nonoperative management of a displaced midshaft clavicle fracture is associated with a high risk of symptomatic nonunion. Clin Orthop Relat Res 2021 Jan 1;479(1):129-138. [doi: 10.1097/CORR.0000000000001411] [Medline: 32675585]

XSL•FO
RenderX

41. Brinker MR, Hanus BD, Sen M, O'Connor DP. The devastating effects of tibial nonunion on health-related quality of life. J Bone Joint Surg Am 2013 Dec 18;95(24):2170-2176. [doi: 10.2106/JBJS.L.00803] [Medline: 24352770]
42. Hak DJ, Fitzpatrick D, Bishop JA, et al. Delayed union and nonunions: epidemiology, clinical issues, and financial aspects. Injury 2014 Jun;45 Suppl 2:S3-S7. [doi: 10.1016/j.injury.2014.04.002] [Medline: 24857025]
43. Birnbaum K, Wirtz DC, Siebert CH, Heller KD. Use of extracorporeal shock-wave therapy (ESWT) in the treatment of non-unions. A review of the literature. Arch Orthop Trauma Surg 2002 Jul;122(6):324-330. [doi: 10.1007/s00402-001-0365-4] [Medline: 12136295]
44. Assiotis A, Sachinis NP, Chalidis BE. Pulsed electromagnetic fields for the treatment of tibial delayed unions and nonunions. A prospective clinical study and review of the literature. J Orthop Surg Res 2012 Jun 8;7:24. [doi: 10.1186/1749-799X-7-24] [Medline: 22681718]
45. Leighton R, Phillips M, Bhandari M, Zura R. Low intensity pulsed ultrasound (LIPUS) use for the management of instrumented, infected, and fragility non-unions: a systematic review and meta-analysis of healing proportions. BMC Musculoskelet Disord 2021 Jun 11;22(1):532. [doi: 10.1186/s12891-021-04322-5] [Medline: 34116673]
46. Wolff C, Steinheimer P, Warmerdam E, et al. Effects of age, body height, body weight, body mass index and handgrip strength on the trajectory of the plantar pressure stance-phase curve of the gait cycle. Front Bioeng Biotechnol 2023;11:1110099. [doi: 10.3389/fbioe.2023.1110099] [Medline: 36873371]
47. Wolff C, Steinheimer P, Warmerdam E, et al. Characteristic changes of the stance-phase plantar pressure curve when walking uphill and downhill: cross-sectional study. J Med Internet Res 2024 May 8;26:e44948. [doi: 10.2196/44948] [Medline: 38718385]
48. Warmerdam E, Burger LM, Mergen DF, Orth M, Pohlemann T, Ganse B. The walking surface influences vertical ground reaction force and centre of pressure data obtained with pressure-sensing insoles. Front Digit Health 2024;6:1476335. [doi: 10.3389/fdgth.2024.1476335] [Medline: 39583997]

Abbreviations

iTUG: instrumented timed up-and-go
MPC-HC: Media Player Classic–Home Cinema
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
TUG: timed up-and-go

Edited by A Stone, L Buis; submitted 30.Jan.2026; peer-reviewed by N Lofgren, T Ibara, V Santos; revised version received 20.Jul.2026; accepted 20.Jul.2026; published 21.Sep.2026.

Please cite as:
Warmerdam E, Burger LM, Ganse B
Timed Up-and-Go Test Using Instrumented Insoles to Monitor Tibial or Malleolar Fracture Healing in Patients With Healing Fractures or Nonunion: Longitudinal and Cross-Sectional Exploratory Study
JMIR Mhealth Uhealth 2026;14:e92493
URL: https://mhealth.jmir.org/2026/1/e92493
doi:10.2196/92493

© Elke Warmerdam, Lea-Marie Burger, Bergita Ganse. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 21.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Feasibility of a Stool State Check App Using AI During Bowel Preparation Before Colonoscopy: Multicenter Prospective Study (SCAN Study)

Atsushi Inaba[1], MD, PhD; Kensuke Shinmura[1], MD, PhD, MPH; Shozo Osera[2], MD, PhD; Takahiro Yamada[2], MD; Hiroaki Kon[3], MD; Maki Kanazawa[3], CLS; Naoki Sugimura[4], MD, PhD; Yasushi Sano[4], MD, PhD; Hiroko Hosaka[5], MD, PhD; Toshio Uraoka[5], MD, PhD; Daiki Sato[6], MD, PhD; Yusuke Yoda[6], MD; Hiroyuki Takamaru[7], MD, PhD; Yutaka Saito[7], MD, PhD; Toshihiko Gocho[8], MD, PhD; Atsushi Katagiri[8], MD, PhD; Kazuhisa Yamaguchi[9], MD, PhD; Takahisa Matsuda[9], MD, PhD; Atsuki Imai[10], MD; Hitomi Fujimoto[10], MD; Hiroki Matsuzaki[11], ME; Nobuyoshi Takeshita[11], MD, PhD; Masashi Wakabayashi[12], MSc; Hiroaki Ikematsu[1,13], MD, PhD; Tomonori Yano[1,14], MD, PhD

[1]Department of Gastroenterology and Endoscopy, National Cancer Center Hospital East, 6-5-1 Kashiwanoha, Kashiwa, Chiba, Japan

[10]Department of Gastrointestinal Medicine, Aizawa Hospital, Matsumoto, Nagano, Japan

[11]Jmees, Inc., Kashiwa, Chiba, Japan

[12]Biostatistics Division, Center for Research Administration and Support, National Cancer Center, Tokyo, Japan

[13]Department of Gastroenterology, IMSUT Hospital, Institute of Medical Science, The University of Tokyo, Tokyo, Japan

[14]Medical Device Innovation Center, National Cancer Center Hospital East, Kashiwa, Japan

[2]Department of Gastroenterology, Saku Central Hospital Advanced Care Center, Saku, Nagano, Japan

[3]Kon Internal Medicine Clinic, Chiba, Japan

[4]Gastrointestinal Center, Sano Hospital, Kobe, Hyogo, Japan

[5]Department of Gastroenterology and Hepatology, Gunma University Graduate School of Medicine, Maebashi, Gunma, Japan

[6]Department of Endoscopy, Saitama Cancer Center, Ina, Saitama, Japan

[7]Endoscopy Division, National Cancer Center Hospital, Tokyo, Japan

[8]Division of Gastroenterology, Department of Medicine, Showa Medical University School of Medicine, Tokyo, Japan

[9]Department of Internal Medicine Division of Gastroenterology and Hepatology, Omori Medical Center Toho University, Tokyo, Japan

Corresponding Author:
Kensuke Shinmura, MD, PhD, MPH
Department of Gastroenterology and Endoscopy, National Cancer Center Hospital East, 6-5-1 Kashiwanoha, Kashiwa, Chiba, Japan

Abstract

Background: Optimal bowel preparation (BP) is crucial for a successful colonoscopy. Although multiple factors influence BP quality, including patient adherence to laxatives and dietary instructions, the stool state during BP should be properly evaluated to perform a colonoscopy of sufficient quality. Therefore, we developed a smartphone app to evaluate a patient's stool state during BP and a viewer to enable real-time monitoring by medical staff.

Objective: This study aimed to assess the feasibility of performing colonoscopies of appropriate quality using the app-and-viewer system.

Methods: This prospective observational study was conducted between November 2022 and December 2023, involving patients scheduled for colonoscopy at 10 Japanese institutions, comprising 6 tertiary hospitals, 3 regional general hospitals, and 1 community-based clinic. Patients who (1) underwent a colonoscopy at participating institutions, (2) were aged between 20 and 70 years, and (3) owned smartphones compatible with Android or iOS were included in the study. The patients downloaded the app on their smartphones and captured images of their stools during BP, while the medical staff reviewed the evaluation of the stools by the app via the viewer system. The primary end point was defined as the proportion of patients with a Boston Bowel Preparation Scale (BBPS) score of ≥6 among those who successfully used the app. Secondary end points included mean BBPS score, rate of an excellent BBPS score (≥8), adenoma detection rate, cecal intubation rate, and withdrawal time in negative colonoscopy. Additionally, we evaluated the usability of the app, medical staff workload burden with the app, and viewer usage via questionnaire surveys.

Results: A total of 343 patients were enrolled, and 326 were ultimately included in the analysis. Overall, 99.1% (323/326, 95% CI 97.3%-99.8%) of the patients achieved the primary end point. The mean BBPS score was 8.5 (SD 1.0), and the proportion of

XSL•FO
RenderX

excellent BBPS scores was 87.4% (285/326). The adenoma detection rate, cecal intubation rate, and mean withdrawal time in negative colonoscopy were 46.9% (153/326, 95% CI 41.4%-52.5%), 99.7% (325/326, 95% CI 98.3%-99.9%), and 10.7 (SD 5.9) minutes, respectively. In the questionnaire survey, 98.5% (321/326) of the patients reported that the tutorial was easy to understand, 96.0% (313/326) found stool image capture easy, and 87.8% (286/326) reported reduced anxiety regarding BP. Furthermore, 90.5% (295/326) of the patients indicated that they would like to use the app again for future colonoscopies. Among medical staff, 92.5% (62/67) considered the viewer system necessary, 89.6% (60/67) found it easy to use, and 89.6% (60/67) reported a reduction in workload burden.

Conclusions: AI-based stool state assessment using the app and the viewer during BP was feasible across diverse BP methods and clinical environments. Favorable BP outcomes and high usability among patients and medical staff support the potential use of this approach in real-world colonoscopy practice.

(JMIR Mhealth Uhealth 2026;14:e91274) doi:10.2196/91274

KEYWORDS

artificial intelligence; bowel preparation; colonoscopy; mobile apps; smartphone; feasibility studies; workload; medical staff; surveys and questionnaires

Introduction

The latest cancer statistics report high mortality and morbidity rates for colorectal cancer (CRC) worldwide [1]. Colonoscopy aids in the detection and resection of precancerous lesions, thereby reducing CRC-related morbidity and mortality [2-8]. Colonoscopy quality indicators include the adenoma detection rate (ADR), cecal intubation rate, withdrawal time, and bowel preparation (BP) [9-11]. Notably, the quality of BP is crucial for the other indicators to be maintained appropriately [12-15]. For instance, BP is associated with an improvement in ADR and with a decreased risk of CRC mortality [16].

Multiple factors affect BP quality, including patient adherence to laxatives and dietary instructions [17-19]. Accordingly, enhancing patient education using booklets, visual aids, short messages, or smartphone apps has proven effective [20-25]. Additionally, evaluating the stool state during BP is important to achieve appropriate BP quality. In Japan, BP is commonly performed at home or in the hospital. At home, patients use visual scales to assess their stool; in hospitals, the assessment may be conducted by the medical staff, contingent on the facility's protocols. In both methods, colonoscopy is initiated when the stool is clear and watery. However, evaluating the stool state during BP can be burdensome for patients and medical staff, and an accurate evaluation may sometimes be difficult. Some patients may also feel embarrassed about having their stool evaluated by the medical staff; likewise, the medical staff may experience psychological discomfort when assessing the patient's stool. Therefore, we previously developed a smartphone app incorporating AI to evaluate stool state during BP to address these issues, and the effectiveness of this app was demonstrated in a single-center study [26]. While the app was effective in assessing BP at the hospital, its feasibility at home or under diverse laxatives and toilet environments remains unknown. Furthermore, we developed a viewer system that connects to the app and enables the medical staff to easily monitor the patient's stool state during routine tasks without directly checking the app screen. This system is intended to improve workflow efficiency in clinical settings.

Accordingly, in this study, we aimed to assess the feasibility of performing colonoscopy of appropriate quality using our app and viewer system.

Methods

Study Design

This multicenter prospective study was conducted at 10 Japanese institutions, comprising 6 tertiary hospitals, 3 regional general hospitals, and 1 community-based clinic, between November 2022 and December 2023. Consecutive patients scheduled to undergo colonoscopy at participating institutions during the study period were screened for eligibility. Eligible patients were approached by the medical staff at each institution and invited to participate in the study.

Ethical Considerations

This study was conducted in accordance with the principles outlined in the Declaration of Helsinki and the Japanese Ethical Guidelines for Clinical Studies Involving Human Subjects. The study protocol was approved by the Institutional Review Board of the National Cancer Center (2022‐023). Written informed consent was obtained from all participants prior to study enrollment. All data were deidentified before analysis to protect participant privacy and confidentiality. Participants did not receive any financial compensation for participation in this study.

Patient Selection

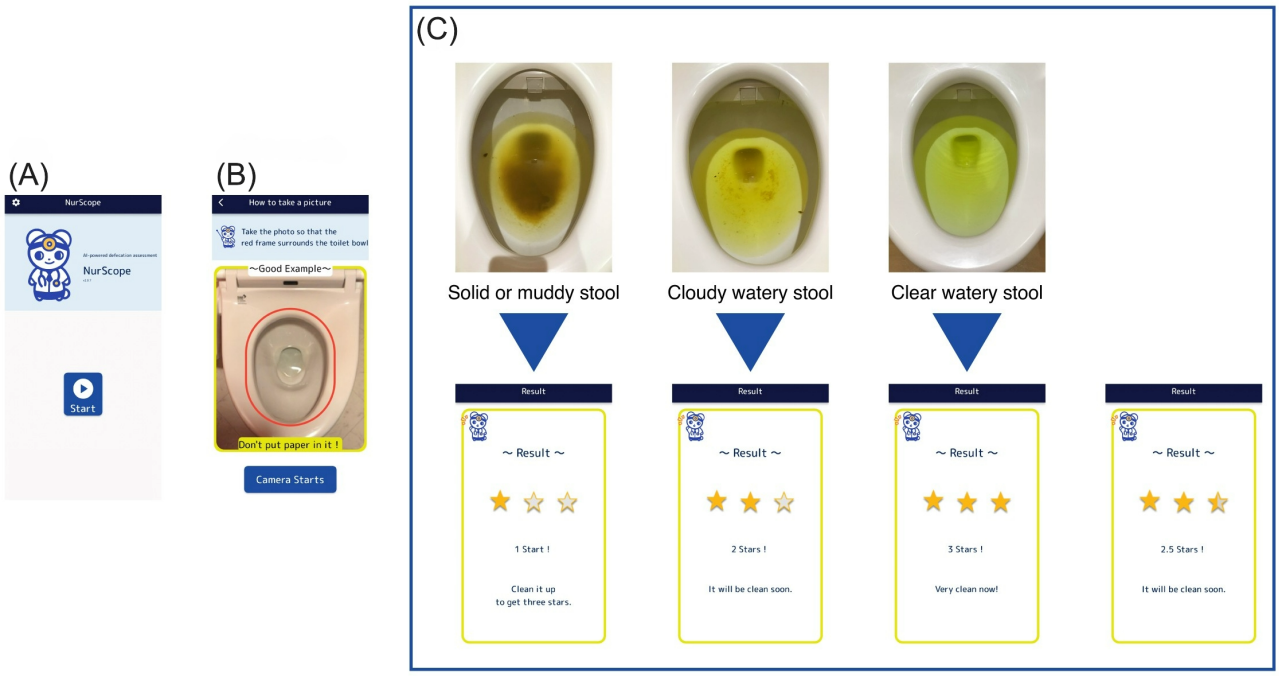
Patients who met the following criteria were included in the study: patients (1) who underwent colonoscopy at participating institutions, (2) aged between 20 and 70 years, and (3) who owned smartphones with Android or iOS compatibility. Patients were excluded from the study if they met any of the following criteria: (1) had a known entire or subentire circumferential CRC, (2) previously underwent colorectal resection (excluding appendectomy), (3) had inflammatory bowel disease, (4) had colorectal disease with active bleeding, (5) had an Eastern Cooperative Oncology Group performance status of ≥2, or (6) had severe dementia.

XSL•FO
RenderX

Instructions to Use the App and the Viewer

The screen displayed in the app is shown in Figure 1. The user was asked to start the app and tap the "Start" button. The user was presented with a screen explaining how to capture a picture. The user was cautioned to take a picture such that the red frame displayed on the screen aligned with the toilet bowl. Then, the user was asked to tap the button "Camera starts" to switch to the screen for taking pictures and touch the camera symbol to capture a picture. The captured stool images were graded on a 4-star scale: 1 star (solid or muddy stool), 2 stars (cloudy watery stool), 3 stars (clear watery stool), and 2.5 stars. Images evaluated as 1 or 2 stars indicated that the stools were insufficient to perform colonoscopy, whereas those evaluated as 3 stars indicated that the stools were sufficiently clear to perform colonoscopy. Images evaluated as 2.5 stars indicated clear watery stools excreted immediately after solid or muddy stools, or clear watery stools excreted despite a total defecation frequency of less than 5 times: this 2.5-star rating was included to ensure the reliability of the 3-star evaluation.

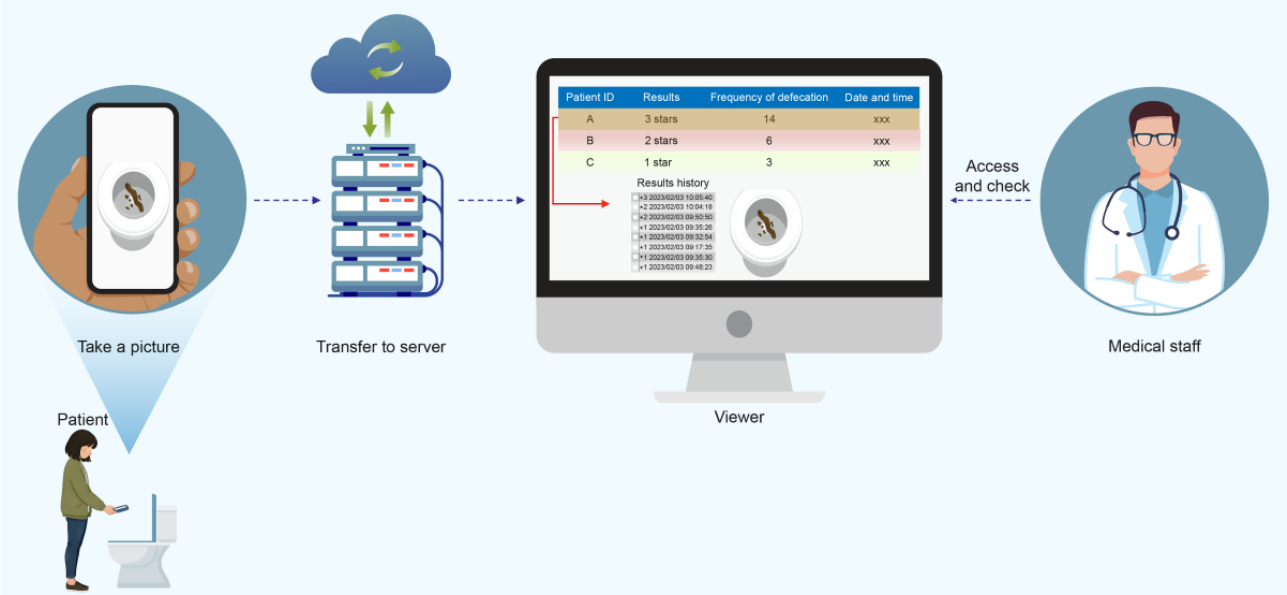
Figure 1. App screen display. (A) Start screen, (B) tutorial screen before taking pictures, and (C) screen displaying the result of stool image evaluation.



The viewer system is shown in Figure 2. Medical staff accessed the viewer using a dedicated personal computer independent of the electronic medical record system at each institution. A unique patient ID was issued when a patient downloaded the app. The viewer screen showed the lists of patient ID, number of defecations, and evaluation of the latest stool state by the app. The medical staff selected the patient ID and checked the number of defecations, stool images, and the app's evaluation. Further, the medical staff who participated in the study noted the patient ID in advance and checked the patient's stool state and the app's evaluation through the viewer during the BP, as necessary.

XSL•FO
RenderX

Figure 2. Overview diagram of the viewer system. Medical staff can check the number and timing of the patient's defecations, as well as the app's evaluation results, by entering the patient ID into the viewer.



Bowel Preparation

BP procedures followed the usual practices established at each institution. The method of laxative intake (same-day or split-dose method), the location where BP was performed, and the consumption of a low-fiber diet were determined by each institution. Patients received picosulfate, sennoside, magnesium oxide, or elobixibat the day before colonoscopy, as needed. On the day of colonoscopy, the patient was administered one of the following laxatives (Table S1 in Multimedia Appendix 1): polyethylene glycol with ascorbic acid (MOVIPREP; EA Pharma Co), magnesium citrate (MAGCOROL; Horii Pharmaceutical Ind), polyethylene glycol (NIFLEC; EA Pharma Co), oral sulfate solution (SULPREP; Fuji Pharma), or magnesium citrate plus sodium picosulfate (PICOPREP; Nippon Chemiphar Co). BP solutions were generally administered on the day of colonoscopy. However, when BP was performed over 2 days, including the day before and the day of colonoscopy, the regimen was determined according to the protocol of each participating institution. Prior to BP, patients downloaded the app to their smartphones and reviewed the instructions for use displayed within the app. Once BP was initiated, patients used the app to capture images of their stools every time they defecated. Colonoscopy was performed when the app presented 3 stars within 3 hours of administration. Contrastingly, when the app did not display 3 stars within 3 hours after BP initiation, additional procedures (such as additional laxative administration or enema) were conducted at the discretion of the doctor. If the app displayed 3 stars after additional procedures, a colonoscopy was subsequently performed. Colonoscopy was performed at the discretion of the doctor for patients who could not tolerate these procedures or who were not presented with 3 stars even after additional procedures.

End Points

Considering that the quality of colonoscopy strongly depends on the degree of bowel cleansing, the Boston Bowel Preparation Scale (BBPS) was used to assess the quality of colonoscopy in this study [27]. The BBPS score was recorded by the endoscopists who performed the colonoscopy. Endoscopists were not blinded to study participation status or app evaluations. The primary end point was defined as the proportion of patients with a BBPS score of ≥6 among patients who successfully used the app. The secondary end points were the mean BBPS score; rate of excellent BBPS score (BBPS score ≥8); mean BBPS score of each segment (right colon, transverse colon, and left colon); ADR; cecal intubation rate; mean withdrawal time in negative colonoscopy; rate of fair, good, or excellent on the Aronchick scale [28]; mean time from laxative administration to colonoscopy initiation; proportion of patients with 3 stars, 2.5 stars, 2 stars, and 1 star for the app's evaluation; proportion of cases with a BBPS score ≥6 when colonoscopy was performed for each evaluation in the app; diagnostic performance of the app for predicting adequate BP (BBPS score ≥6); mean BBPS score for each evaluation; and adverse events in colonoscopy. Additionally, we surveyed patients and medical staff who participated in this study to assess the usability and effectiveness of the app and viewer system using a questionnaire. These end points were also evaluated in patients who could successfully use the app, referring to patients who could launch the app by themselves, capture their stool images, and confirm the app's evaluation. Successful app use was defined as obtaining at least 5 appropriate stool images during BP. If patients missed a stool image, they could self-report it using a dedicated function within the app. Cases were still considered successful even if some images were of poor quality, captured the stool only partially, or contained nonstool objects, provided that at least 5 appropriate stool images were ultimately obtained. We also calculated the proportion of patients who failed to use the app among all enrolled patients.

Statistical Analysis

For the primary end point, the target value was defined as 95% or above according to the European guideline [10]. The threshold value for the proportion of patients with a BBPS score

XSL•FO
RenderX

of ≥6 among patients who successfully used the app was 90%. Considering an expected value of 95%, a one-sided alpha of 2.5%, and a power of 90%, the number of patients required was calculated to be 316 by a method based on binomial distribution. The target number of patients was set to 340 to account for patients who could not use the app successfully. All statistical analyses were performed using SAS (version 9.4).

Results

Patient Characteristics

A total of 343 patients were enrolled in this study. After the exclusion of 17 patients, 326 eligible patients were included in the final analysis. The median patient age was 57 (IQR 48-63; range 26 - 69) years, and 57.7% (n=188) were men (Table 1). Among the 326 patients, 99.4% (n=324) underwent same-day BP, and 68.4% (n=223) underwent BP at home. In addition, 8.6% (n=28) of the patients had a history of inadequate BP during a previous colonoscopy.

XSL•FO
RenderX

Table . Baseline characteristics of eligible patients (n=326).

Characteristics	Values
Age (y), median (IQR; range)	57 (48-63; 26 - 69)
Sex, n (%)	
Male	188 (57.7)
Female	138 (42.3)
BMI (kg/m^2), mean (SD)	23.1 (3.9)
Regular use of laxatives, n (%)	
Yes	17 (5.2)
No	309 (94.8)
Regular use of tricyclic antidepressant, n (%)	
Yes	1 (0.3)
No	325 (99.7)
History of abdominal (noncolorectal) surgery, n (%)	
Yes	58 (17.8)
No	268 (82.2)
Comorbidities, n (%)	
Diabetes mellitus	25 (7.7)
Liver cirrhosis	1 (0.3)
Cerebral stroke	7 (2.1)
Cerebral hemorrhage	1 (0.3)
Parkinson disease	0 (0)
History of inadequate bowel preparation, n (%)	
Yes	28 (8.6)
No	194 (59.5)
N/A due to no prior colonoscopy	96 (29.4)
Unknown	8 (2.5)
Indication of colonoscopy, n (%)	
Screening	148 (45.4)
Surveillance	114 (35.0)
Diagnostic	34 (10.4)
Symptomatic	26 (8.0)
Other	4 (1.2)
Methods of bowel preparation, n (%)	
Same-day bowel preparation	324 (99.4)
Split-dose bowel preparation	2 (0.6)
Location where the bowel preparation was performed, n (%)	
Patients' home	223 (68.4)
Hospital	103 (31.6)
Taking a low-fiber diet, n (%)	
Yes	221 (67.8)
No	105 (32.2)
Laxative regimen, n (%)	
Polyethylene glycol with ascorbic acid	227 (69.6)

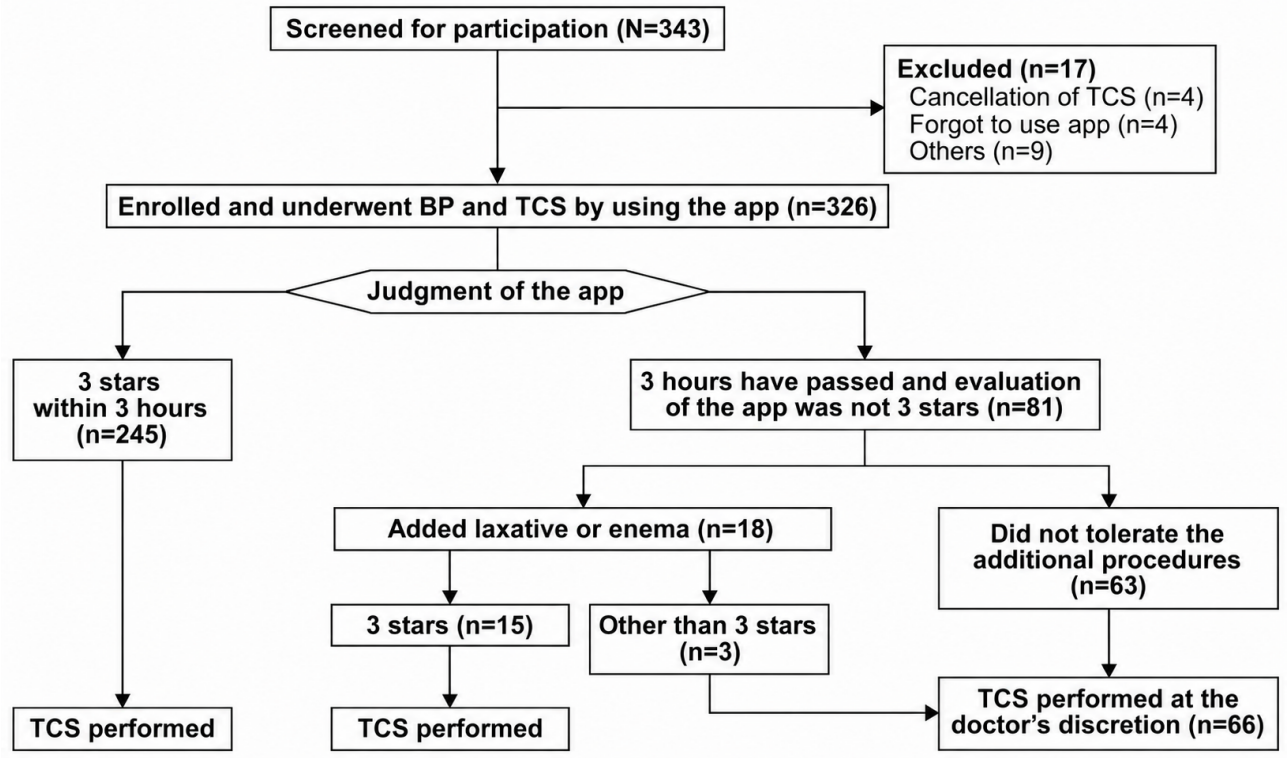
XSL•FO
RenderX

Characteristics	Values
Magnesium citrate	48 (14.7)
Polyethylene glycol	28 (8.6)
Oral sulfate solution	21 (6.4)
Magnesium citrate plus sodium picosulfate	2 (0.7)

Study Flow

As shown in Figure 3, 326 patients successfully used the app and underwent colonoscopy following the app's evaluation. Furthermore, 245 patients were conferred 3 stars within 3 hours of the start of laxative administration and subsequently underwent colonoscopy, whereas 81 patients did not receive 3 stars even 3 hours after laxative initiation. Of these, 63 patients who could not tolerate the additional procedure underwent colonoscopy at the discretion of the doctor, whereas 18 patients underwent additional procedures. Of these 18 patients, 15 were further evaluated with 3 stars, and a colonoscopy was performed. The remaining 3 patients who were not graded with 3 stars underwent colonoscopy at the discretion of the doctor.

Figure 3. Status of the patients in the study flow from laxative administration to colonoscopy. BP: bowel preparation; TCS: total colonoscopy.



Outcome

Among the entire cohort, 99.1% (323/326, 95% CI 97.3%-99.8%) of the patients had a BBPS score of ≥6 (Table 2), indicating that the primary end point was met because the lower limit of the 95% CI exceeded the prespecified threshold of 90%. The mean BBPS score was 8.5 (SD 1.0), and the proportion of excellent BBPS score was 87.4% (285/326). The mean BBPS scores of the right, transverse, and left colons were 2.8 (SD 0.5), 2.9 (SD 0.3), and 2.9 (SD 0.4), respectively. The proportion of fair, good, or excellent Aronchick scale scores was 98.8% (322/326). The ADR, cecal intubation rate, and mean withdrawal time in negative colonoscopy were 46.9% (153/326, 95% CI 41.4%-52.5%), 99.7% (325/326, 95% CI 98.3%-99.9%), and 10.7 (SD 5.9) minutes, respectively. The proportions of patients with 3 stars, 2.5 stars, 2 stars, and 1 star for the app's evaluation were 79.8% (260/326), 5.8% (19/326), 12.6% (41/326), and 1.8% (6/326), respectively. The proportion of cases with a BBPS score of ≥6 when colonoscopy was performed for each evaluation in the app was 100% (260/260), 100% (19/19), 95.1% (39/41), and 83.3% (5/6) for the 3 stars, 2.5 stars, 2 stars, and 1 star, respectively. The mean BBPS score for each evaluation outcome was 8.6 (SD 0.8), 7.9 (SD 1.2), 8.4 (SD 1.2), and 7.5 (SD 2.9) for 3 stars, 2.5 stars, 2 stars, and 1 star, respectively. The mean time from laxative administration to colonoscopy initiation was 305 (SD 89) minutes. The incidence rates of adverse events in colonoscopy, particularly bleeding, perforation, and ischemic colitis, were 0.6% (2/326), 0% (0/326), and 0% (0/326), respectively. The diagnostic performance of the app for predicting a BBPS score of ≥6 is summarized in Table S2 in Multimedia Appendix 1. The sensitivity, specificity, positive predictive value, and negative predictive value were 80.5%, 100%, 100%, and 4.5%, respectively.

XSL•FO
RenderX

Table . Outcomes of patients who used the app successfully (n=326).

Outcomes	Values
Primary end point	
Patients scoring BBPS[a] score ≥6, n/N (%)	323/326 (99.1)
Secondary end points	
BBPS score, mean (SD)	8.5 (1.0)
Excellent bowel preparation (BBPS score ≥8) rate, n/N (%)	285/326 (87.4)
BBPS score of each segment, mean (SD)	
Right colon	2.8 (0.5)
Transverse colon	2.9 (0.3)
Left colon	2.9 (0.4)
Aronchick scale, n/N (%)	
1 (Excellent)	255/326 (78.2)
2 (Good)	53/326 (16.3)
3 (Fair)	14/326 (4.3)
4 (Poor)	3/326 (0.9)
5 (Inadequate)	1/326 (0.3)
Adenoma detection rate, n/N (%)	153/326 (46.9)
Cecal intubation rate, n/N (%)	325/326 (99.7)
Withdrawal time in disease-free patients (min), mean (SD)	10.7 (5.9)
Patients for app evaluation, n/N (%)	
3 stars	260/326 (79.8)
2.5 stars	19/326 (5.8)
2 stars	41/326 (12.6)
1 star	6/326 (1.8)
Rate of BBPS score ≥6 for each app evaluation, n/N (%)	
3 stars	260/260 (100)
2.5 stars	19/19 (100)
2 stars	39/41 (95.1)
1 star	5/6 (83.3)
BBPS score for each app evaluation, mean (SD)	
3 stars	8.6 (0.8)
2.5 stars	7.9 (1.2)
2 stars	8.4 (1.2)
1 star	7.5 (2.9)
Time from laxative administration to starting colonoscopy (min), mean (SD)	305 (89)
Adverse events, n/N (%)	
Bleeding	2/326 (0.6)
Perforation	0/326 (0)
Ischemic colitis	0/326 (0)

[a]BBPS: Boston Bowel Preparation Scale.

XSL•FO
RenderX

Questionnaire Survey

Table 3 presents the results of the survey to assess the usability and effectiveness of the app and viewer system using a questionnaire administered to patients. Overall, 96.0% (313/326) of the patients answered that downloading the app was easy, 98.5% (321/326) reported that the description screen was easy to understand, and 96.0% (313/326) expressed that they could take pictures of their stool without any problems. Of the 15.6% (51/326) of patients who felt uncomfortable with the medical staff checking their defecation, 98.0% (50/51) of those answered that using the app would solve this problem. Furthermore, 37.7% (123/326) of the patients answered that they had no confidence in evaluating their own stool state, and 87.8% (108/123) of those patients answered that using the app would solve this problem. Ultimately, 90.5% (295/326) of the patients indicated a preference for using the app during their next colonoscopy.

Table . Questionnaire survey for patients (n=326).

Questionnaire results	Participants, n/N (%)
Question 1. Did you download the app without any problems?	
Yes	313/326 (96.0)
No	13/326 (4.0)
Question 2. Was the app tutorial screen easy to understand?	
Yes	321/326 (98.5)
No	5/326 (1.5)
Question 3. Was it easy to take stool images?	
Yes	313/326 (96.0)
No	13/326 (4.0)
Question 4. Was the evaluation of the app easy to understand?	
Yes	310/326 (95.1)
No	16/326 (4.9)
Question 5. Would you feel uncomfortable if your stool was checked by medical staff during the BP[a]?	
Yes	51/326 (15.6)
No	275/326 (84.4)
Did you feel that the app would improve your discomfort?[b]	
Yes	50/51 (98.0)
No	1/50 (2.0)
Question 6. Do you have any concerns about being able to evaluate your own defecation during the BP?	
Yes	123/326 (37.7)
No	203/326 (62.3)
Did you feel that your anxiety would be relieved by the app?[c]	
Yes	108/123 (87.8)
No	15/123 (12.2)
Question 7. Would you like to use the app for your next colonoscopy?	
Yes	295/326 (90.5)
No	31/326 (9.5)

[a]BP: bowel preparation.

[b]This question was administered to participants who answered "Yes" to question 5.

[c]This question was administered to participants who answered "Yes" to question 6.

Table 4 presents the results of the questionnaire survey administered to the medical staff. Overall, 83.6% (56/67) of the medical staff checked the patients' stool state during daily work. Among them, 17.9% (10/56) answered that they felt a psychological burden, and 55.4% (31/56) answered that they felt a physical burden in checking patients' stool state in their daily work. Notably, 89.6% (60/67) of the medical staff indicated that the implementation of the app would alleviate these burdens. For the viewer system, 89.6% (60/67) of the medical staff reported that it was easy to use, while 92.5% (62/67) indicated that it was necessary for BP.

XSL•FO
RenderX

Table . Questionnaire survey for medical staff (n=67).

Questionnaire results	Participants, n/N (%)
Question 1. Do you check the patients' stool state during BP[a] in your daily work?	
Yes	56/67 (83.6)
No	11/67 (16.4)
Do you have a psychological burden to check patients' stool state?[b]	
Yes	10/56 (17.9)
No	46/56 (82.1)
Do you have a workload to check patients' stool state?[b]	
Yes	31/56 (55.4)
No	25/56 (44.6)
Question 2. Did you feel that using the app during BP would reduce the workload of checking patients' stools?	
Yes	60/67 (89.6)
No	7/67 (10.4)
Question 3. How was the operability of the viewer screen displaying?	
Easy	60/67 (89.6)
Difficult	7/67 (10.4)
Question 4. Did you feel that the viewer screen was necessary?	
Yes	62/67 (92.5)
No	5/67 (7.5)

[a]BP: bowel preparation.

[b]This question was administered to participants who answered "Yes" to question 1.

Discussion

This multicenter prospective study successfully evaluated the performance of the app across various types of laxatives, toilet bowls, and diverse patient populations. Among 326 patients who successfully used the app, 99.1% (n=323) patients achieved a BBPS score ≥6, which exceeded the set expected value. Additionally, several key quality indicators, including an ADR of 46.9% (n=153) and a cecal intubation rate of 99.7% (n=325), exceeded the target values recommended by American and European guidelines [9,10]. These results demonstrated the feasibility of app-guided BP management while maintaining high-quality colonoscopy indicators across multiple institutions and environments.

Notably, this study demonstrated the effective use of our app across a wide range of clinical settings. Specifically, we validated our app under diverse BP protocols and toilet environments and demonstrated that it could be applied to BP management in various clinical contexts. The following 2 main findings of this study support the versatility of the app, which is critical for its implementation in real-world clinical practice.

First, we evaluated the app's assessment of BP using a variety of laxatives. In Japan, 6 major types of laxatives are commonly used for BP [29], of which 5 were used in this study. Additionally, the timing of administration, use of a low-fiber diet, and administration of laxatives on the day before colonoscopy were determined according to the protocols of each institution. Regardless of the type of laxative used and a wide variety of BP methods, a stool state evaluation rated as 3 stars using the app indicated that colonoscopy could be performed with a BBPS score ≥6 in 100% (260/260) of cases. Moreover, the mean BBPS score of patients rated as 3 stars was ≥8 (SD 0.8; excellent BBPS score). Even in real-world clinical settings, which exhibit variability in the types of laxatives used and BP methods, the 3-star rating by the app was shown to be highly correlated with an excellent BBPS score. These findings suggest that the app can assist BP across a broad range of institutional practices (eg, clinics and general hospitals) and diverse patient needs (eg, preference and tolerability for specific laxatives and low-fiber diets) while maintaining an adequate quality of colonoscopy. Although the app evaluations were associated with BP quality, the app was not designed to directly predict BBPS score. Instead, the app was intended to evaluate stool state to verify whether BP was sufficient to proceed with colonoscopy. In the present study, 95.1% (39/41) of the patients with a 2-star evaluation and 83.3% (5/6) of those with a 1-star evaluation still achieved a BBPS score ≥6. Consistent with these findings, the app showed a high positive predictive value (100%) but a low negative predictive value (4.5%) for predicting a BBPS score of ≥6 above. These results suggest that while a 3-star evaluation reliably identified patients with adequate BP (BBPS score ≥6), lower star evaluations did not necessarily indicate inadequate BP (BBPS score <6). Although the number of patients with a 1-star evaluation was small, these findings suggest that a substantial proportion of patients with lower star

XSL•FO
RenderX

evaluations ultimately achieved adequate BP. This may reflect limitations of the star-based evaluation, as well as the imperfect relationship between stool state immediately before colonoscopy and overall bowel cleanliness assessed by the BBPS. Nevertheless, further refinement of the stool-state assessment algorithm may improve its ability to distinguish between different stool states during BP. In the present study, some patients with 1-star or 2-star evaluations still achieved a BBPS score of ≥6, suggesting that additional preparation may have been unnecessary in a subset of cases. Therefore, there remains room for improvement in the diagnostic discrimination of the app, particularly in reducing potential additional preparation (eg, additional laxative intake or enema). To further clarify the diagnostic characteristics of the app, sensitivity, specificity, positive predictive value, and negative predictive value for predicting a BBPS score of ≥6 are summarized in Table S2 in Multimedia Appendix 1.

Second, the study included both hospital-based and home-based BP settings, reflecting diverse real-world environments with potential variability in toilet conditions. Unlike our previous study [26], this study also included patients who underwent BP at home. Evaluations for BP at hospitals were conducted using toilet bowls available at each of the 10 participating institutions. Meanwhile, evaluations for BP at home were performed using toilets installed in each patient's home. Thus, the app assessed a wide variety of toilet bowls. In both settings, when the evaluation of the app was 3 stars, colonoscopy was performed with a BBPS score of ≥6 in 100% (260/260) of the cases. Furthermore, the mean BBPS score for patients rated as 3 stars was ≥8 (SD 0.8) in both settings, indicating that an excellent BBPS score was achieved. Therefore, even across a wide variety of toilet environments with differences in shape, color tone, and lighting conditions, the app was able to evaluate stool state consistency; when the stool was rated as 3 stars, colonoscopy could be performed with an adequate BBPS score. These findings suggest that the app may be applicable across diverse clinical and home-based environments and support its potential feasibility in real-world clinical practice.

Using a similar approach to this study, 2 recent reports have demonstrated the effectiveness of AI-based smartphone apps in improving BP quality [30,31]. In previous studies, the developed app had a binary classification function directly determining whether the bowel cleansing quality was adequate based on the stool state. Consequently, the mean BBPS score achieved with those apps was 7.2 (SD 1.4) and 6.7 (SD 1.3), respectively. In contrast, our app evaluated each defecation individually during the BP process, considering the number of defecations. The app was used to assess stool state throughout the BP process, followed by colonoscopy. In our study, a mean BBPS score of 8.5 (SD 1.0) was observed, which was higher than the values reported in previous studies. These findings suggest that it is difficult to accurately predict the quality of BP solely based on a single stool state and that evaluating all defecations may support more appropriate BP management throughout the BP process.

The observed rate of the BBPS score of ≥6 in this study was notably high (99.1%, 323/326), exceeding both the minimum (≥90%) and target (≥95%) standards recommended by the European guideline [9,10]. Several factors may have contributed to this finding. First, patients aged 70 years or older were excluded from this study. Because advanced age is a known risk factor for inadequate BP, this exclusion may have contributed to the favorable BBPS outcomes. However, this exclusion criterion was not intended to select patients with better BP quality. Instead, this decision was based on the low smartphone ownership rates among older individuals in Japan at the time of study protocol development (38.3% among individuals in their 70s and 11.0% among those in their 80s), as well as our previous single-center experience showing that all patients who were unable to appropriately operate the app were aged 70 years or older [26,32]. Therefore, older patients were excluded because they were considered more likely to have difficulty appropriately using the app in this study. In addition, the use of the app itself and the awareness that stool images and app evaluations could be reviewed by medical staff through the viewer system may have increased patient awareness and adherence to BP instructions (ie, observation effect).

Although the mechanisms underlying the potential benefits of the app were not directly evaluated in this study, several possible explanations and future perspectives may be considered. Several mechanisms may explain how the app could support BP management. First, the viewer system enabled real-time monitoring of stool state during BP, allowing medical staff to review temporal changes and provide timely instructions when needed. Second, the use of the app itself may have promoted behavioral reinforcement and improved adherence to BP instructions. In addition, future expansion of app functions, such as automated reminders for examination schedules and dietary instructions, may further support behavioral modification during BP. Third, the app may help reduce uncertainty in BP assessment. In Japan, stool state is commonly evaluated during BP; however, assessment methods may vary across institutions and between patients and medical staff. By providing standardized stool state assessment and visualization of temporal changes in stool state, the app may help reduce variability and uncertainty in BP management.

Although our app evaluates the stool state over time during BP, the responsibility for deciding whether to proceed with colonoscopy based on the app's evaluations or to implement additional procedures if the BP is insufficient lies with the medical staff. Therefore, a viewer system was introduced in this study to enable an efficient review of temporal changes in stool state and the corresponding evaluations during BP. In the questionnaire survey, most medical staff reported no issues with the usability of the viewer and responded that the viewer was necessary during BP. Particularly, the reasons cited for the necessity of the viewer were that its use enabled real-time monitoring of BP, allowing timely instructions for patients during the BP process, and that it had the potential to reduce patient waiting times before colonoscopy. Thus, by enabling a standardized assessment of the stool state and real-time monitoring, the combined use of the app and viewer system introduces a new framework for quality assurance in BP.

Finally, the usability of our app was evaluated using questionnaires administered to patients and medical staff. Most patients were able to use the app without any issues, and

XSL•FO
RenderX

discomfort with having their stool state assessed by medical staff, as well as anxiety about evaluating their own stool state, was alleviated using the app. The observed reduction in patient anxiety during BP suggests that the app may also help lower the psychological barriers to undergoing colonoscopy. Furthermore, answers from medical staff suggested that using the app might help reduce their physical and psychological burdens. This reduction implies that the app can support a more efficient clinical workflow. In health care systems experiencing workforce shortages, especially in aging populations, tools that facilitate task shifting and reduce manual workloads are essential. In this regard, our app may provide a practical solution for supporting efficient BP management without compromising the quality of care.

Nevertheless, this study has some limitations. First, the study was not designed as a randomized controlled trial and therefore lacks direct comparison with BP without the use of the app. Future randomized controlled trials comparing outcomes with and without the app are essential to validate the clinical usage of the app more robustly. Second, we did not specifically assess the usability of the app in patients aged 70 years or older. This subset of patients was excluded from the study because smartphone ownership rates among older individuals in Japan were relatively low at the time of study initiation, and many older patients were unfamiliar with smartphone app operation. In addition, in our previous single-center study, all patients who were unable to appropriately operate the app were aged 70 years or older. However, smartphone ownership among older individuals in Japan has increased substantially in recent years. According to national data from 2024, smartphone ownership rates have risen to 67.5% among individuals in their 70s and 30.7% among those in their 80s [33]. Therefore, with further improvements in user interface design and usability support, the applicability of the app to older patients and populations with lower digital literacy may expand in the future. Third, this study was conducted exclusively in Japan. Therefore, the generalizability of the results to other countries remains unclear. Furthermore, because BP protocols and toilet environments differ across countries and regions, further validation studies are needed to confirm whether the app can achieve similar performance outside Japan. Fourth, endoscopists who assigned BBPS were not blinded to study participation status or app evaluations. Therefore, the possibility of observer bias in BBPS assessment cannot be excluded because the primary end point relied on endoscopist-assigned BBPS. Although all participating endoscopists received standardized instruction on BBPS assessment, including washing and suctioning of residues before scoring, such training cannot completely eliminate the possibility of assessment bias in a nonblinded study. Fifth, the primary end point analysis included only patients who successfully used the app. Therefore, the observed results may overestimate the feasibility of app-guided BP management in real-world settings.

In conclusion, the proportion of patients with a BBPS score of ≥6 for colonoscopy using the app exceeded the expected value in addition to the threshold value. In this multicenter prospective study, AI-based stool state assessment using the app and viewer during BP was feasible across diverse BP methods and clinical environments. Favorable BP outcomes and high usability among patients and medical staff support the potential usage of this approach in real-world colonoscopy practice.

Acknowledgments

The authors thank all participating patients and nurses who answered the questionnaire. The authors also thank Junko Iwase for supporting data management; Natsuko Iwasaki, Mayumi Yamada, and Yuya Ikeda of the Clinical Research Support Office, National Cancer Center Hospital East, for generously providing the EDC system and for their invaluable support with data management; and Astellas Pharma Inc. for supporting the implementation of this study. The authors also wish to thank Editage [34] for English language editing.

Generative AI tools (ChatGPT, OpenAI) were used to assist with language editing and improvement of manuscript readability. The authors reviewed and edited all AI-generated content and take full responsibility for the final content of the manuscript.

Funding

Funding was provided by Astellas Pharma, Inc.

Data Availability

Due to patient privacy and ethical restrictions, the data underlying this study cannot be made publicly available. Deidentified data may be available from the corresponding author upon reasonable request and with approval from the institutional review board.

Authors' Contributions

Conceptualization: A Inaba, KS
Data curation: A Inaba, KS
Formal analysis: MW
Funding acquisition: KS
Investigation: A Inaba, KS, SO, TY, HK, MK, NS, Y Sano, HH, TU, DS, YY, HT, Y Saito, TG, AK, KY, TM, A Imai, HF, HI, TY

XSL•FO
RenderX

Methodology: A Inaba, KS, MW
Project administration: KS, A Inaba
Resources: SO, TY, HK, MK, NS, Y Sano, HH, TU, DS, YY, HT, Y Saito, TG, AK, KY, TM, A Imai, HF, HI, TY
Software: HM
Supervision: TY
Validation: HM
Visualization: A Inaba, KS
Writing – original draft: A Inaba
Writing – review and editing: A Inaba, KS, SO, TY, HK, MK, NS, Y Sano, HH, TU, DS, YY, HT, Y Saito, TG, AK, KY, TM, A Imai, HF, HM, NT, MW, HI, TY

Conflicts of Interest

Dr Tatsuro Murano (who served as a reviewer) worked in the same department with the authors at the National Cancer Center Hospital East from November 2022 until his retirement in March 2023. The review did not unduly affect the decision to publish the article. A Inaba received honorarium from Fuji Pharma Co., Ltd. All other authors have no conflicts of interest to declare.

Multimedia Appendix 1
Laxative administration protocols and diagnostic performance of the app.
[DOCX File, 18 KB - mhealth_v14i1e91274_app1.docx]

References

1. Arnold M, Abnet CC, Neale RE, et al. Global burden of 5 major types of gastrointestinal cancer. Gastroenterology 2020 Jul;159(1):335-349.e15. [doi: 10.1053/j.gastro.2020.02.068] [Medline: 32247694]
2. Singh H, Turner D, Xue L, Targownik LE, Bernstein CN. Risk of developing colorectal cancer following a negative colonoscopy examination: evidence for a 10-year interval between colonoscopies. JAMA 2006 May 24;295(20):2366-2373. [doi: 10.1001/jama.295.20.2366] [Medline: 16720822]
3. Baxter NN, Warren JL, Barrett MJ, Stukel TA, Doria-Rose VP. Association between colonoscopy and colorectal cancer mortality in a US cohort according to site of cancer and colonoscopist specialty. J Clin Oncol 2012 Jul 20;30(21):2664-2669. [doi: 10.1200/JCO.2011.40.4772] [Medline: 22689809]
4. Nishihara R, Wu K, Lochhead P, et al. Long-term colorectal-cancer incidence and mortality after lower endoscopy. N Engl J Med 2013 Sep 19;369(12):1095-1105. [doi: 10.1056/NEJMoa1301969] [Medline: 24047059]
5. Lieberman DA, Weiss DG, Harford WV, et al. Five-year colon surveillance after screening colonoscopy. Gastroenterology 2007 Oct;133(4):1077-1085. [doi: 10.1053/j.gastro.2007.07.006] [Medline: 17698067]
6. Brenner H, Chang-Claude J, Seiler CM, Hoffmeister M. Long-term risk of colorectal cancer after negative colonoscopy. J Clin Oncol 2011 Oct 1;29(28):3761-3767. [doi: 10.1200/JCO.2011.35.9307] [Medline: 21876077]
7. Bretthauer M, Løberg M, Wieszczy P, et al. Effect of colonoscopy screening on risks of colorectal cancer and related death. N Engl J Med 2022 Oct 27;387(17):1547-1556. [doi: 10.1056/NEJMoa2208375] [Medline: 36214590]
8. Sano Y, Hotta K, Matsuda T, et al. Endoscopic removal of premalignant lesions reduces long-term colorectal cancer risk: results from the Japan Polyp Study. Clin Gastroenterol Hepatol 2024 Mar;22(3):542-551.e3. [doi: 10.1016/j.cgh.2023.07.021] [Medline: 37544420]
9. Rex DK, Anderson JC, Butterly LF, et al. Quality indicators for colonoscopy. Am J Gastroenterol 2024 Sep 1;119(9):1754-1780. [doi: 10.14309/ajg.0000000000002972] [Medline: 39167112]
10. Kaminski MF, Thomas-Gibson S, Bugajski M, et al. Performance measures for lower gastrointestinal endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) Quality Improvement Initiative. Endoscopy 2017 Apr;49(4):378-397. [doi: 10.1055/s-0043-103411] [Medline: 28268235]
11. Hassan C, Bretthauer M, Kaminski MF, et al. Bowel preparation for colonoscopy: European Society of Gastrointestinal Endoscopy (ESGE) guideline. Endoscopy 2013;45(2):142-150. [doi: 10.1055/s-0032-1326186] [Medline: 23335011]
12. Clark BT, Rustagi T, Laine L. What level of bowel prep quality requires early repeat colonoscopy: systematic review and meta-analysis of the impact of preparation quality on adenoma detection rate. Am J Gastroenterol 2014 Nov;109(11):1714-1723. [doi: 10.1038/ajg.2014.232] [Medline: 25135006]
13. Sulz MC, Kröger A, Prakash M, Manser CN, Heinrich H, Misselwitz B. Meta-analysis of the effect of bowel preparation on adenoma detection: early adenomas affected stronger than advanced adenomas. PLoS One 2016;11(6):e0154149. [doi: 10.1371/journal.pone.0154149] [Medline: 27257916]
14. Guo R, Wang YJ, Liu M, et al. The effect of quality of segmental bowel preparation on adenoma detection rate. BMC Gastroenterol 2019 Jul 8;19(1):119. [doi: 10.1186/s12876-019-1019-8] [Medline: 31286888]
15. Alvi H, Rasheed T, Shaikh MA, Ali FS, Zuberi BF, Samejo AA. Impact of bowel preparation on caecal intubation time during colonoscopy. Pak J Med Sci 2019;35(6):1516-1519. [doi: 10.12669/pjms.35.6.1031] [Medline: 31777485]

XSL•FO
RenderX

16. Zessner-Spitzenberg J, Waldmann E, Rockenbauer LM, et al. Impact of bowel preparation quality on colonoscopy findings and colorectal cancer deaths in a nation-wide colorectal cancer screening program. Am J Gastroenterol 2024 Oct 1;119(10):2036-2044. [doi: 10.14309/ajg.0000000000002880] [Medline: 39007693]
17. Martel M, Barkun AN, Menard C, Restellini S, Kherad O, Vanasse A. Split-dose preparations are superior to day-before bowel cleansing regimens: a meta-analysis. Gastroenterology 2015 Jul;149(1):79-88. [doi: 10.1053/j.gastro.2015.04.004] [Medline: 25863216]
18. Chen E, Chen L, Wang F, Zhang W, Cai X, Cao G. Low-residue versus clear liquid diet before colonoscopy: an updated meta-analysis of randomized, controlled trials. Medicine (Baltimore) 2020 Dec 4;99(49):e23541. [doi: 10.1097/MD.0000000000023541] [Medline: 33285772]
19. Pan P, Gu L, Zhao S, et al. Prepackaged formula low-residue diet vs. self-prepared low-residue diet before colonoscopy: a multicenter randomized controlled trial. Front Med (Lausanne) 2023;10:1029493. [doi: 10.3389/fmed.2023.1029493] [Medline: 37035340]
20. Spiegel BMR, Talley J, Shekelle P, et al. Development and validation of a novel patient educational booklet to enhance colonoscopy preparation. Am J Gastroenterol 2011 May;106(5):875-883. [doi: 10.1038/ajg.2011.75] [Medline: 21483463]
21. Tae JW, Lee JC, Hong SJ, et al. Impact of patient education with cartoon visual aids on the quality of bowel preparation for colonoscopy. Gastrointest Endosc 2012 Oct;76(4):804-811. [doi: 10.1016/j.gie.2012.05.026] [Medline: 22840295]
22. Walter B, Klare P, Strehle K, et al. Improving the quality and acceptance of colonoscopy preparation by reinforced patient education with short message service: results from a randomized, multicenter study (PERICLES-II). Gastrointest Endosc 2019 Mar;89(3):506-513.e4. [doi: 10.1016/j.gie.2018.08.014] [Medline: 30138612]
23. Kang X, Zhao L, Leung F, et al. Delivery of instructions via mobile social media app increases quality of bowel preparation. Clin Gastroenterol Hepatol 2016 Mar;14(3):429-435.e3. [doi: 10.1016/j.cgh.2015.09.038] [Medline: 26492848]
24. Guo B, Zuo X, Li Z, et al. Improving the quality of bowel preparation through an app for inpatients undergoing colonoscopy: a randomized controlled trial. J Adv Nurs 2020 Apr;76(4):1037-1045. [doi: 10.1111/jan.14295] [Medline: 31840286]
25. Walter B, Frank R, Ludwig L, et al. Smartphone application to reinforce education increases high-quality preparation for colorectal cancer screening colonoscopies in a randomized trial. Clin Gastroenterol Hepatol 2021 Feb;19(2):331-338.e5. [doi: 10.1016/j.cgh.2020.03.051] [Medline: 32240835]
26. Inaba A, Shinmura K, Matsuzaki H, et al. Smartphone application for artificial intelligence-based evaluation of stool state during bowel preparation before colonoscopy. Dig Endosc 2024 Dec;36(12):1338-1346. [doi: 10.1111/den.14827] [Medline: 39031797]
27. Lai EJ, Calderwood AH, Doros G, Fix OK, Jacobson BC. The Boston bowel preparation scale: a valid and reliable instrument for colonoscopy-oriented research. Gastrointest Endosc 2009 Mar;69(3 Pt 2):620-625. [doi: 10.1016/j.gie.2008.05.057] [Medline: 19136102]
28. Aronchick CA, Lipshutz WH, Wright SH. Validation of an instrument to assess colon cleansing. Am J Gastroenterol 1999;94:2667 [FREE Full text]
29. Tamai N, Sumiyama K. Optimal bowel preparation for colonoscopy. Dig Endosc 2025 Feb;37(2):139-146. [doi: 10.1111/den.14914] [Medline: 39229776]
30. Lu YB, Lu SC, Huang YN, et al. A novel convolutional neural network model as an alternative approach to bowel preparation evaluation before colonoscopy in the COVID-19 era: a multicenter, single-blinded, randomized study. Am J Gastroenterol 2022 Sep 1;117(9):1437-1443. [doi: 10.14309/ajg.0000000000001900] [Medline: 35973166]
31. Zhu Y, Zhang DF, Wu HL, et al. Improving bowel preparation for colonoscopy with a smartphone application driven by artificial intelligence. NPJ Digit Med 2023 Mar 14;6(1):41. [doi: 10.1038/s41746-023-00786-y] [Medline: 36918730]
32. Information and communications in Japan: white paper [Report in Japanese]. : Ministry of Internal Affairs and Communications (MIC); 2020 URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r02/pdf/index.html [accessed 2026-05-26]
33. Survey on trends in communications usage 2024 [Report in Japanese]. : Ministry of Internal Affairs and Communications (MIC); 2025 URL: https://www.soumu.go.jp/johotsusintokei/statistics/data/250530_1.pdf [accessed 2026-07-24]
34. Editage. URL: https://www.editage.com/ [accessed 2026-07-24]

Abbreviations

ADR: adenoma detection rate
BBPS: Boston Bowel Preparation Scale
BP: bowel preparation
CRC: colorectal cancer

XSL•FO
RenderX

Edited by L Buis; submitted 15.Jan.2026; peer-reviewed by C Ramprasad, T Murano; revised version received 24.Jun.2026; accepted 25.Jun.2026; published 01.Sep.2026.

Please cite as:
Inaba A, Shinmura K, Osera S, Yamada T, Kon H, Kanazawa M, Sugimura N, Sano Y, Hosaka H, Uraoka T, Sato D, Yoda Y, Takamaru H, Saito Y, Gocho T, Katagiri A, Yamaguchi K, Matsuda T, Imai A, Fujimoto H, Matsuzaki H, Takeshita N, Wakabayashi M, Ikematsu H, Yano T
Feasibility of a Stool State Check App Using AI During Bowel Preparation Before Colonoscopy: Multicenter Prospective Study (SCAN Study)
JMIR Mhealth Uhealth 2026;14:e91274
URL: https://mhealth.jmir.org/2026/1/e91274
doi:10.2196/91274

© Atsushi Inaba, Kensuke Shinmura, Shozo Osera, Takahiro Yamada, Hiroaki Kon, Maki Kanazawa, Naoki Sugimura, Yasushi Sano, Hiroko Hosaka, Toshio Uraoka, Daiki Sato, Yusuke Yoda, Hiroyuki Takamaru, Yutaka Saito, Toshihiko Gocho, Atsushi Katagiri, Kazuhisa Yamaguchi, Takahisa Matsuda, Atsuki Imai, Hitomi Fujimoto, Hiroki Matsuzaki, Nobuyoshi Takeshita, Masashi Wakabayashi, Hiroaki Ikematsu, Tomonori Yano. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 1.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Outcomes of Mobile Health Use in Sinonasal Surgery: Retrospective Cohort Study

Heli Majeethia[1], BS, MD; Akshay R Prabhakar[1], BSA; Justina R Varghese[1], BA; Najm S Khan[2], MBS; Roshan Dongre[3], BS; Vincent Provasek[4], PhD; Faizaan I Khan[3], BS; Zain Mehdi[1], BA; Omar G Ahmed[1], MD; Masayoshi Takashima[1], MD

[1]Department of Otolaryngology - Head & Neck Surgery, Houston Methodist Academic Institute, 6550 Fannin Street, Houston, TX, United States

[2]Rutgers New Jersey Medical School, Rutgers, The State University of New Jersey, New Brunswick, NJ, United States

[3]School of Engineering Medicine, Texas A&M University, Houston, TX, United States

[4]Texas A&M College of Medicine, Texas A&M University, Houston, TX, United States

Corresponding Author:
Masayoshi Takashima, MD
Department of Otolaryngology - Head & Neck Surgery, Houston Methodist Academic Institute, 6550 Fannin Street, Houston, TX, United States

Abstract

Background: Mobile health (mHealth) technologies are increasingly integrated into perioperative care to enhance patient engagement and communication. Prior studies in surgical and medical specialties suggest that mHealth platforms may be associated with reductions in hospital stay and readmissions; however, evidence supporting their impact in otolaryngology, particularly in sinonasal surgery, remains limited.

Objective: The aim of this study was to evaluate the association between perioperative enrollment in CareSense, a patient-facing mHealth platform, and postoperative health care utilization outcomes, including hospital readmissions, emergency department (ED) visits, and length of stay (LOS), among adults undergoing sinonasal surgery.

Methods: This is a retrospective cohort study performed at a single tertiary care academic medical center between May 2021 and January 2024. All adult patients (≥18 years) who underwent sinonasal surgery with two fellowship-trained rhinologists during the study period were included. CareSense was offered to all patients at the time of surgical scheduling, and enrollment was voluntary. Patients were categorized into CareSense participants and nonparticipants. Primary outcomes were all-cause hospital readmissions and ED visits within 30, 60, and 90 days following surgery. Secondary outcomes included the length of hospital stay among readmitted patients. Clinical, demographic, and outcome data were obtained through retrospective electronic health record review. Univariate analyses compared outcomes between groups, and multivariable logistic regression using generalized estimating equations was performed to estimate the association between CareSense participation and outcomes while adjusting for age, sex, hypertension, and diabetes.

Results: A total of 1135 patients were included, of whom 340 (30%) enrolled in CareSense and 795 (70%) did not. Compared with nonparticipants, CareSense participants had lower adjusted odds ratio (OR) for readmission for any cause at 30 days (OR 0.24, 95% CI 0.08 - 0.75; *P*=.007), 60 days (OR 0.40, 95% CI 0.19 - 0.83; *P*=.01), and 90 days (OR 0.54, 95% CI 0.29 - 0.99; *P*=.04). Among patients who were readmitted, mean LOS was shorter in the CareSense group than in the nonparticipating group (0.17 vs 1.68 d; *P*<.001). The majority of readmissions in both cohorts were unrelated to complications of the index sinonasal procedure.

Conclusions: This study demonstrates the benefit of CareSense in lowering postoperative readmission rates and LOS for sinonasal surgery patients, illustrating the role of medical health technology in improving patient care and quality outcomes. Perioperative enrollment in a patient-facing mHealth platform was associated with lower postoperative health care utilization and shorter hospital length of stay following sinonasal surgery. Given the voluntary nature of enrollment and the observational design, these findings should be interpreted as observation findings and hypothesis-generating for prospective studies to more definitively assess the causal impact of mHealth interventions and to identify which components of digital perioperative care most effectively improve outcomes in otolaryngologic surgery.

(JMIR Mhealth Uhealth 2026;14:e75403) doi:10.2196/75403

KEYWORDS

mobile health; telehealth; patient-centered care; patient satisfaction; patient empowerment; patient engagement; patient involvement; hospital stay; communication programs

XSL•FO
RenderX

Introduction

Continuous improvement in patient care and clinical outcomes remains a central goal in health care, and advances in digital technology have created new opportunities to support this aim by improving communication, education, and patient engagement. The World Health Organization defines mobile health (mHealth) as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistance, and other wireless devices" [1,2]. As mobile phone usage steadily increases worldwide, leveraging this development offers the potential for rapid dissemination of health information and remote patient engagement, which may enhance medical care as evidenced during the COVID-19 pandemic [3].

Over the past decade, mHealth interventions have been increasingly incorporated into routine care across multiple specialties [4]. Emerging evidence suggests that these platforms may be associated with improvements in health care utilization outcomes, including reduced length of stay and lower readmission rates in orthopedic and cardiovascular populations [5-8]. In otolaryngology, mHealth tools have been adopted for perioperative communication and symptom monitoring; however, evidence evaluating their association with postoperative health care utilization remains limited, particularly for patients undergoing sinonasal surgery [9-11].

CareSense is a patient-facing mHealth platform designed to support perioperative care through automated educational content and reminders delivered via text message or email, with parallel clinician-facing dashboards for monitoring patient-reported information. The platform aims to improve patient preparedness and adherence with visits and surgeries by facilitating perioperative education and communication, while enabling tracking of patient progress and satisfaction. The aim of this study was to evaluate the association between perioperative enrollment in CareSense and postoperative health care utilization outcomes, including hospital readmissions, emergency department visits, and length of hospital stay, among adults undergoing sinonasal surgery.

Methods

Study Population

A retrospective observational cohort study was conducted at a single tertiary academic medical center between May 2021 and January 2024. Beginning May 2021, the Department of Otolaryngology–Head and Neck Surgery at Houston Methodist Hospital began offering CareSense (MedTrak, Inc) to all patients undergoing sinonasal surgery. Clinical, demographic, and outcome data were obtained through retrospective review of the electronic health record (EHR). All patients ≥18 years of age who underwent sinonasal surgery with two fellowship-trained rhinologists were included. Patients were divided into 2 groups: those who elected to participate in CareSense and those who did not. Enrollment in CareSense was voluntary and occurred at the time of surgical scheduling.

Ethical Considerations

This study was reviewed and approved by the Houston Methodist Research Institute Institutional Review Board (PRO00036908) with a waiver of informed consent due to its retrospective design and use of existing clinical data. All data were analyzed in a deidentified format to protect patient privacy and confidentiality. No compensation was provided to the participants.

Mobile Health Technology

CareSense is a digital platform used to deliver standardized perioperative educational content and automated reminders to patients undergoing surgery. Enrolled patients received procedure-specific "pathways," defined as preset configurations of communication and information delivery tailored to the planned surgery (eg, sinus surgery or septoplasty). Each pathway provided surgery-specific perioperative instructions and educational materials, including preoperative preparation, postoperative expectations, warning signs of complications, and medication courses. Message timing and content were predefined and delivered consistently across enrolled patients via text message or email. Communication formats included brief text notifications, longer-form educational emails, and physician-defined survey questions, with selected responses generating alerts within a clinician-facing Care Navigator dashboard. Details of the sinus surgery and septoplasty pathways are provided in Multimedia Appendices 1 and 2, respectively. Patients without access to a mobile phone or email address were not eligible for enrollment.

Outcome Measures

The primary outcomes evaluated were readmission rates and emergency visits to any of our hospital system sites within 30, 60, and 90 days of discharge following sinonasal surgery. Readmissions were defined as any unplanned inpatient or observation admission occurring after surgery. Emergency department (ED) visits were defined as encounters that originated in the ED and resulted in discharge without inpatient admission. Outcome ascertainment was performed through system-wide EHR review, capturing events occurring within the same integrated health system. The secondary outcome was postoperative length of stay (LOS) among patients who were readmitted, used as a surrogate for the complexity and acuity of the readmission. Outcomes were assessed at fixed postoperative intervals; variation in individual follow-up duration beyond these windows was not evaluated. All patients were followed through the EHR for a minimum of 90 days postoperatively within the integrated health system, with outcome ascertainment limited to encounters captured during these predefined intervals.

Statistical Analysis

Demographic and comorbidity data were available for all patients. Differences in baseline patient characteristics between CareSense participants and nonparticipants were assessed using an independent-samples *t*-test for continuous variables and χ^2 tests for categorical variables. Univariate analyses using χ^2 testing were performed to compare outcome measures between the two cohorts. Given the low frequency of procedure-related

XSL•FO
RenderX

events, analyses of surgery-related readmissions were conducted descriptively rather than through inferential subgroup modeling. To partially address confounding related to voluntary enrollment, multivariable logistic regression using generalized estimated equations (GEE) was used to estimate the association between CareSense participation and postoperative outcomes while adjusting for prespecified covariates, including age, sex, hypertension, and diabetes. Covariates were prespecified a priori based on clinical relevance and prior literature evaluating postoperative health care utilization [12-14]. Odds ratios (ORs) with 95% CIs were reported. All statistical tests were two-sided, with an α level of .05 considered statistically significant. Univariate analyses were performed using Jamovi [15], and multivariable analyses were conducted in RStudio (version 4.3; R Core Team, 2023).

Results

Study Population

A total of 1135 patients underwent sinonasal surgery from May 2021 to January 2024. Among these patients, 340 (30%) elected to participate in CareSense, whereas 795 (70%) did not participate. Enrollment in CareSense increased over time, with an 80.7% increase from 2021 to 2024.

Of the 340 patients enrolled in CareSense, 202 (59%) were assigned to the sinus surgery pathway and 138 (41%) to the septoplasty with turbinate reduction pathway (Multimedia Appendix 1 and Multimedia Appendix 2, respectively). Patient demographics and baseline characteristics are summarized in Table 1.

XSL•FO
RenderX

Table . Sample patient characteristics of nonparticipating and participating patient groups.

Variables	Patients, n (%)		*P* value
	Nonparticipating (n=795)	Participating (n=340)	
Mean age in years (SD)	51.1 (17.8)	47.1 (17.2)	<.001
Sex			.04
Female	375 (47.2)	138 (40.6)	
Male	420 (52.8)	202 (59.4)	
Ethnicity			
Declined	14 (1.8)	14 (4.1)	.02
Hispanic or Latino	133 (16.7)	49 (14.4)	.33
Not Hispanic or Latino	648 (81.5)	277 (81.5)	.99
Race			
Asian	73 (9.2)	41 (12)	.14
Black	109 (13.7)	37 (10.9)	.19
White	555 (69.8)	237 (69.7)	.97
Declined	48 (6)	20 (5.9)	.92
Other	10 (1.3)	5 (1.5)	—[a]
Comorbidity			
Asthma	176 (22.1)	78 (22.9)	.77
Atherosclerosis	127 (16.0)	42 (12.4)	.12
COPD[b]	64 (8.1)	26 (7.6)	.82
Diabetes type 2	150 (18.9)	41 (12.1)	.005
GERD[c]	286 (36.0)	115 (33.8)	.49
Heart failure	52 (6.5)	16 (4.7)	.23
Hypertension	365 (45.9)	127 (37.4)	.008
Hyperlipidemia	339 (42.6)	135 (39.7)	.36
Illicit drug use			
Yes	27 (3.4)	11 (3.2)	.89
No	710 (89.3)	314 (92.4)	.11
Alcohol use			
Yes	395 (49.7)	187 (55.0)	.10
No	364 (45.8)	148 (43.5)	.48
Tobacco use			
Yes	49 (6.2)	20 (5.9)	.86
Quit	168 (21.1)	79 (23.2)	.43
Never	547 (68.8)	238 (70.0)	.69
Smoking or tobacco use			
Current	49 (6.2)	21 (6.2)	.86
Former	165 (20.8)	77 (22.6)	.48
Never	550 (69.2)	240 (70.6)	.64

[a]Not available.

[b]COPD: chronic obstructive pulmonary disease.

[c]GERD: gastroesophageal reflux disease.

XSL•FO
RenderX

Overall, the mean age of participating patients was 47.1 (SD 17.2) years, and that of nonparticipating patients was 51.1 (SD 17.8) years (*P*<.001). Males accounted for 420 of 795 (52.8%) patients in the nonparticipating group and 202 of 340 (59.4%) patients in the participating group (*P*=.04). There were no statistical differences in race and ethnicity between groups.

The nonparticipating group had a higher prevalence of hypertension (365/795, 45.9% vs 127/340, 37.4%, *P*=.008) and type 2 diabetes (150/795, 18.9% vs 41/340, 12.0%, *P*=.005). There were no significant differences between groups with respect to asthma, atherosclerosis, chronic obstructive pulmonary disease, gastroesophageal reflux disease, heart failure, and hyperlipidemia. Self-reported alcohol use, tobacco use, and illicit drug use did not significantly differ between groups.

Outcome Measures

Readmission rates are presented in Table 2. At 30 days postoperatively, readmission rates were 5.8% (46/795) in the nonparticipating group and 1.2% (4/340) in the participating group (*P*<.001). At 60 days, readmission rates were 7.9% (63/795) in the nonparticipating group and 2.9% (10/340) in the participating group (*P*=.002). At 90 days, readmission rates were 8.9% (71/795) in the nonparticipating group and 4.1% (14/340) in the participating group (*P*=.005).

Table . Comparing the likelihood of readmissions and emergency visits between patients who were and were not enrolled in CareSense.

Outcome variables	Patients, n (%)		*P* value
	Control (n=795)	Intervention (n=340)	
Readmission (days)			
30	46 (5.8)	4 (1.2)	<.001
60	63 (7.9)	10 (2.9)	.002
90	71 (8.9)	14 (4.1)	.005
Emergency visit (days)			
30	27 (3.4)	4 (1.2)	.04
60	40 (5.0)	8 (2.4)	.04
90	52 (6.5)	11 (3.2)	.03

ED visit rates demonstrated a similar pattern. At 30 days, ED visit rates were 3.4% (27/795) in the nonparticipating group and 1.2% (4/340) in the participating group (*P*=.04). At 60 days, visit rates were 5.0% (40/795) in the nonparticipating group and 2.4% (8/340) in the participating group (*P*=.04). At 90 days, visit rates were 6.5% (52/795) in the nonparticipating group and 3.2% (11/340) in the participating group (*P*=.03).

To further characterize postoperative health care utilization, an exploratory descriptive analysis was performed examining readmissions related to complications of the index sinonasal procedure. Among the 202 patients enrolled in the CareSense sinus surgery pathway, 14 (7%) experienced a readmission, compared with 7 (5%) of the 138 patients enrolled in the septoplasty pathway (*P*=.45). Of all readmissions in the participating group, 6 of 25 (24%) were related to postoperative complications of the index procedure, including 5 cases of uncontrolled epistaxis and 1 nasoseptal abscess. In the nonparticipating group, 21 of 125 readmissions (17%) were procedure-related (*P*=.41), most commonly due to uncontrolled epistaxis (18 cases), with additional causes including sinusitis flares requiring intravenous antibiotics (2 cases) and septal hematoma (1 case). When related to the index procedure, readmissions were predominantly attributable to uncontrolled epistaxis in both cohorts. Unrelated admissions included asthma exacerbations, chest pain, abdominal pain, and COVID-19–related illness.

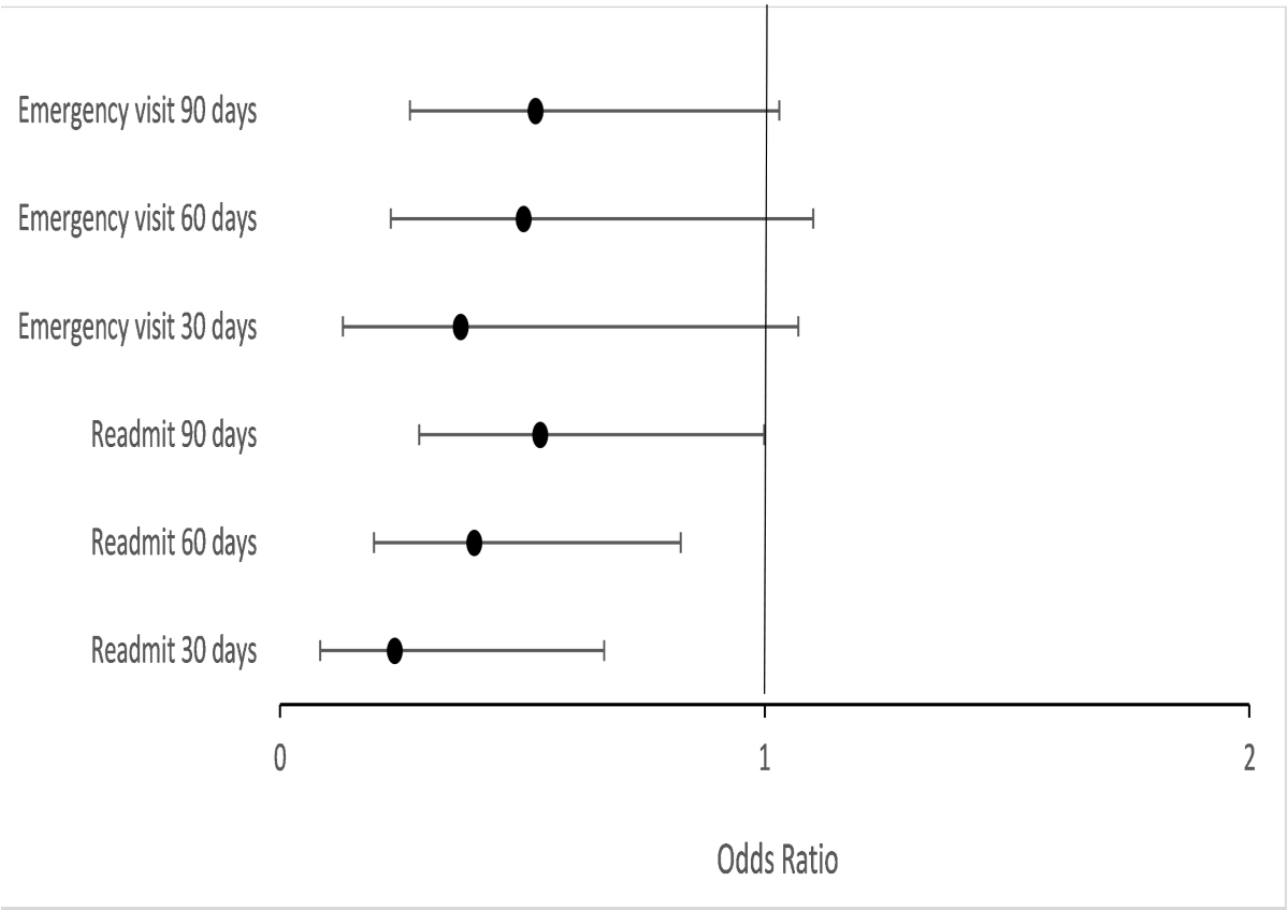
The mean postoperative LOS among patients who were readmitted was 1.68 days in the nonparticipating group compared with 0.17 days in the participating group (*P*<.001).

Multivariable Analysis

After adjusting for age, sex, hypertension, and diabetes, multivariable logistic regression demonstrated that CareSense participation was associated with lower odds of readmission at 30 days (OR 0.24, 95% CI 0.08 - 0.75; *P*=.007), 60 days (OR 0.4, 95% CI 0.19 - 0.83; *P*=.01), and 90 days (OR 0.54, 95% CI 0.29 - 0.99; *P*=.04) when compared to the nonparticipating group (Figure 1). Although not statistically significant, CareSense participation trended toward an association with lower odds of emergency department visits at 30 days (OR 0.37, 95% CI 0.13 - 1.07; *P*=.07), 60 days (OR 0.5, 95% CI 0.23 - 1.1; *P*=.09), and 90 days (OR 0.53, 95% CI 0.27 - 1.03; *P*=.06) when compared to the nonparticipating group (Figure 1).

XSL•FO
RenderX

Figure 1. Forest plot illustrating the odds of readmissions and emergency visits between patients who were and were not enrolled in CareSense at 30, 60, and 90 days.



Discussion

The rapid expansion of mHealth technologies has raised important questions regarding their role in improving clinical outcomes and health care utilization. Prior studies have reported associations between mHealth interventions and improved outcomes across multiple medical and surgical domains, including chronic pulmonary disease, heart failure, diabetes, and orthopedic surgery [5,16,17]. These effects have been attributed to improved access to health information, enhanced communication between patients and care teams, and increased patient engagement, which may collectively support self-management and appropriate health care utilization [18,19]. However, evidence evaluating mHealth interventions in otolaryngologic surgery, particularly with respect to postoperative health care utilization, remains limited.

In this retrospective observational study, perioperative enrollment in CareSense was associated with lower rates of all-cause hospital readmissions and ED visits at 30, 60, and 90 days following sinonasal surgery, as well as shorter hospital LOS among patients who were readmitted. These findings suggest that the use of a perioperative patient engagement platform may be associated with reduced postoperative health care utilization in this population. From a health systems perspective, even modest reductions in readmissions and length of stay may have meaningful implications for cost and resource utilization, given that a single inpatient hospital day in the United States is estimated to cost between US $2000 and $3000 [20].

Notably, the majority of readmissions observed in both cohorts were unrelated to the index sinonasal procedure. This finding suggests that the observed associations may not reflect a direct effect on surgical complications, but rather broader differences in postoperative care-seeking behavior and health system utilization. Importantly, procedure-related readmissions were uncommon and were predominantly driven by postoperative epistaxis in both cohorts, suggesting that differences in overall health care utilization were not driven by an increase in surgery-specific complications. Enrollment in CareSense may serve as a proxy for patient engagement, with participating patients potentially more likely to manage symptoms proactively, adhere to discharge instructions, or seek outpatient guidance before presenting for acute care. The platform's structured perioperative education and automated communication may also help establish expectations and provide reassurance during recovery, thereby reducing uncertainty-driven utilization of emergency services. One study of a perioperative mHealth intervention demonstrated reductions in postoperative care needs and improvements in patient self-management, even when effects on disease-specific outcomes were modest [21]. These findings support the hypothesis that digital perioperative engagement may influence health care utilization by addressing information gaps and postoperative uncertainty rather than directly preventing surgical

XSL•FO
RenderX

complications. Collectively, these features may support patient confidence and self-efficacy before and after surgery, contributing to reduced reliance on acute care services in the postoperative period.

Baseline differences between participating and nonparticipating patients indicate that there should be cautious interpretation. Patients who enrolled in CareSense were younger and had lower rates of hypertension and diabetes, characteristics that may independently influence postoperative outcomes and health care utilization. Although multivariable adjustment was performed, residual confounding is likely in any observational study of voluntary technology adoption. The demographic profile of CareSense participants aligns with prior literature demonstrating greater uptake of digital health tools among younger patients [22,23]. Interestingly, a higher proportion of CareSense participants were male, which contrasts with prior studies reporting higher digital health engagement among younger women [22,23]. Additionally, nonparticipating patients had higher rates of chronic comorbidities such as hypertension and diabetes mellitus, which are known to adversely impact postoperative outcomes in several surgical populations, although an association with sinonasal surgery outcomes has not been well established [24-27]. Importantly, even after adjusting for these measured differences, CareSense participation remained associated with lower odds of postoperative readmission at multiple time points. Despite these differences, CareSense participation remained associated with lower odds of readmission across multiple time points after adjustment, consistent with prior reports of CareSense-associated benefits in orthopedic surgery populations [5].

This study has limitations. Selection bias is inherent to voluntary mHealth enrollment, and unmeasured factors such as health literacy, social support, and baseline health care engagement may have influenced the observed associations. Outcomes were limited to encounters captured within a single integrated health system, and healthcare utilization outside the system was not assessed. Procedure-related readmissions accounted for a small proportion of events, limiting the power to evaluate surgery-specific complications. Additionally, sinonasal disease–specific patient-reported outcome measures and endoscopic findings were unavailable, precluding assessment of symptom severity or disease control as mediators of postoperative utilization. Patients without access to a mobile phone or email address were not eligible for CareSense enrollment, raising the possibility that unmeasured social determinants of health contributed to group differences.

Despite these limitations, this study provides evidence from a large real-world cohort that perioperative enrollment in a patient-facing mHealth platform is associated with reduced postoperative health care utilization following sinonasal surgery. These findings are observational and hypothesis-generating, and they highlight the need for prospective multicenter studies to more definitively evaluate causal relationships with pre-specified subgroups, characterize engagement intensity, and identify which components of digital perioperative care are most effective. Future studies could prospectively assess subgroup effects by surgical procedure, baseline health status, and measured engagement metrics (eg, message opens or survey completion), to better understand which patients derive the greatest benefit from perioperative mHealth support. As mHealth platforms become increasingly integrated into routine surgical care, understanding their role in optimizing postoperative outcomes and health care utilization will be essential.

Acknowledgments

We gratefully acknowledge the support of Peter and Eileen Zwart. Their commitment to improving Otolaryngology patient care has greatly supported this project and contributed to its successful completion. Generative AI was not used in any portion of the manuscript writing.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to institutional restrictions on sharing identifiable or potentially identifiable electronic health record data but are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: HM (lead), ARP (equal)
Data curation: HM (lead), FK (supporting)
Formal analysis: NSK (lead), HM (supporting)
Funding acquisition: MT
Investigation: NSK (lead), FIK (supporting), RD (supporting)
Methodology: HM (lead), MT (supporting), OGA (supporting)
Project administration: MT (lead), VP (equal), OGA (supporting)
Resources: MT

XSL•FO
RenderX

Software: VP (lead), RD (supporting)
Supervision: MT (lead), OGA (equal)
Validation: HM (lead), ARP (supporting), ZM (supporting)
Visualization: HM (lead), ARP (supporting), JRV (supporting)
Writing – original draft: HM (lead), NSK (supporting), FIK (supporting), ZM (supporting)
Writing – review & editing: ARP (lead), JRV (supporting), HM (supporting), MT (supporting).

Conflicts of Interest

MT is a consultant for Neurent Medical, Aerin Medical, Medtronic ENT, and LivaNova. OGA is a consultant for Aerin Medical, Medtronic ENT, and LivaNova. The authors have no other funding, financial relationships, or conflicts of interest to disclose.

Multimedia Appendix 1
CareSense Sinus Surgery Pathway Content.
[DOCX File, 35 KB - mhealth_v14i1e75403_app1.docx]

Multimedia Appendix 2
CareSense Septoplasty with Turbinate Reduction Pathway Content.
[DOCX File, 33 KB - mhealth_v14i1e75403_app2.docx]

References

1. The Global Health Observatory. mHealth programmes are sponsored by government.: World Health Organization URL: https://www.who.int/data/gho/indicator-metadata-registry/imr-details/4774
2. Bradway M, Carrion C, Vallespin B, et al. mHealth Assessment: Conceptualization of a Global Framework. JMIR Mhealth Uhealth 2017 May 2;5(5):e60. [doi: 10.2196/mhealth.7291] [Medline: 28465282]
3. Macbeth R. COVID-19 an opportunity for collaborative approaches to digital technology. Health Informatics New Zealand. URL: https://www.hinz.org.nz/news/492852/COVID-19-an-opportunity-for-collaborative-approaches-to-digital-technology.htm
4. Ghafur S, Schneider EC. Engaging patients using digital technology — learning from other industries. NEJM Catalyst 2019 Jun;5(3). [doi: 10.1056/CAT.19.0654]
5. Bruce CR, Harrison P, Nisar T, et al. Assessing the impact of patient-facing mobile health technology on patient outcomes: retrospective observational cohort study. JMIR Mhealth Uhealth 2020 Jun 26;8(6):e19333. [doi: 10.2196/19333] [Medline: 32589161]
6. Zhang J, Nie X, Yang X, Mei Q, Xiang X, Cheng L. A meta-analysis of effectiveness of mobile health interventions on health-related outcomes in patients with heart failure. J Cardiovasc Med (Hagerstown) 2024 Aug 1;25(8):587-600. [doi: 10.2459/JCM.0000000000001631] [Medline: 38949146]
7. Zhu Y, Zhao Y, Wu Y. Effectiveness of mobile health applications on clinical outcomes and health behaviors in patients with coronary heart disease: a systematic review and meta-analysis. Int J Nurs Sci 2024 Apr;11(2):258-275. [doi: 10.1016/j.ijnss.2024.03.012] [Medline: 38707688]
8. Boitano TKL, Gardner A, Chu DI, Leath CA III, Straughn JM Jr, Smith HJ. Use of a mobile health patient engagement technology improves perioperative outcomes in gynecologic oncology patients. Gynecol Oncol 2023 Nov;178:23-26. [doi: 10.1016/j.ygyno.2023.09.008] [Medline: 37742507]
9. Shah M, Douglas J, Carey R, et al. Reducing ER visits and readmissions after head and neck surgery through a phone-based quality improvement program. Ann Otol Rhinol Laryngol 2021 Jan;130(1):24-31. [doi: 10.1177/0003489420937044] [Medline: 32567391]
10. Garcia A, Chari DA, Stankovic KM, Lee DJ, Kozin ED, Franck KH. Implementation of mobile audiometry during the COVID-19 pandemic. Otolaryngol Head Neck Surg 2022 Sep;167(3):465-468. [doi: 10.1177/01945998211051588] [Medline: 34609938]
11. Ning Y, Dong Z, Jia Z, et al. Development of mobile health-based interventions to promote physical activity in patients with head and neck cancer: a qualitative study. Front Public Health 2023;11:1260804. [doi: 10.3389/fpubh.2023.1260804] [Medline: 38074711]
12. Dumitra T, Ganescu O, Hu R, et al. Association between patient activation and health care utilization after thoracic and abdominal surgery. JAMA Surg 2021 Jan 1;156(1):e205002. [doi: 10.1001/jamasurg.2020.5002] [Medline: 33146682]
13. Knight SR, Ng N, Tsanas A, Mclean K, Pagliari C, Harrison EM. Mobile devices and wearable technology for measuring patient outcomes after surgery: a systematic review. NPJ Digit Med 2021 Nov 12;4(1):157. [doi: 10.1038/s41746-021-00525-1] [Medline: 34773071]
14. Nguyen N, Leveille E, Guadagno E, Kalisya LM, Poenaru D. Use of mobile health technologies for postoperative care in paediatric surgery: a systematic review. J Telemed Telecare 2022 Jun;28(5):331-341. [doi: 10.1177/1357633X20934682] [Medline: 32605411]

XSL•FO
RenderX

15. The Jamovi Project. 2024. URL: https://www.jamovi.org [accessed 2026-06-17]
16. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23. [doi: 10.2196/mhealth.8873] [Medline: 29343463]
17. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth interventions to support self-management of long-term conditions: a systematic metareview of diabetes, heart failure, asthma, chronic obstructive pulmonary disease, and cancer. J Med Internet Res 2017 May 17;19(5):e172. [doi: 10.2196/jmir.6688] [Medline: 28526671]
18. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mhealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817. [doi: 10.2196/13817] [Medline: 31293246]
19. Müssener U. Digital encounters: human interactions in mHealth behavior change interventions. Digit Health 2021;7:20552076211029776. [doi: 10.1177/20552076211029776] [Medline: 34262783]
20. AHA Annual Survey Database, copyright 2022 by Health Forum, LLC, an affiliate of the American Hospital Association. AHA Data & Insights. 2023. URL: https://www.ahadata.com/aha-annual-survey-database [accessed 2026-06-17]
21. Wang TF, Huang RC, Yang SC, Chou C, Chen LC. Evaluating the effects of a mobile health app on reducing patient care needs and improving quality of life after oral cancer surgery: quasiexperimental study. JMIR Mhealth Uhealth 2020 Jul 27;8(7):e18132. [doi: 10.2196/18132] [Medline: 32716303]
22. Laing SS, Alsayid M, Ocampo C, Baugh S. Mobile health technology knowledge and practices among patients of safety-net health systems in Washington State and Washington, DC. J Patient Cent Res Rev 2018;5(3):204-217. [doi: 10.17294/2330-0698.1622] [Medline: 31414005]
23. Gallagher R, Roach K, Sadler L, et al. Mobile technology use across age groups in patients eligible for cardiac rehabilitation: survey study. JMIR Mhealth Uhealth 2017 Oct 24;5(10):e161. [doi: 10.2196/mhealth.8352] [Medline: 29066425]
24. Venkatesan S, Myles PR, Manning HJ, et al. Cohort study of preoperative blood pressure and risk of 30-day mortality after elective non-cardiac surgery. Br J Anaesth 2017 Jul 1;119(1):65-77. [doi: 10.1093/bja/aex056] [Medline: 28633374]
25. Bamba R, Gupta V, Shack RB, Grotting JC, Higdon KK. Evaluation of diabetes mellitus as a risk factor for major complications in patients undergoing aesthetic surgery. Aesthet Surg J 2016 May;36(5):598-608. [doi: 10.1093/asj/sjv241] [Medline: 27069242]
26. Rubel KE, Rodman CP, Jones A, et al. The effect of diabetes mellitus on post-operative outcomes in patients undergoing endoscopic sinus surgery. Ann Otol Rhinol Laryngol 2023 Mar;132(3):317-321. [doi: 10.1177/00034894221091348] [Medline: 35473357]
27. Filimonov A, Chung SY, Wong A, Brady JS, Baredes S, Eloy JA. Effect of diabetes mellitus on postoperative endoscopic sinus surgery outcomes. Int Forum Allergy Rhinol 2017 Jun;7(6):584-590 [FREE Full text] [doi: 10.1002/alr.21906]

Abbreviations

ED: Emergency Department
EHR: Electronic Health Record
GEE: generalized estimated equations
LOS: length of stay
mHealth: mobile health

Edited by L Buis; submitted 02.Apr.2025; peer-reviewed by M Tuna, PH Chang; revised version received 21.Jan.2026; accepted 14.May.2026; published 21.Jul.2026.

Please cite as:
Majeethia H, Prabhakar AR, Varghese JR, Khan NS, Dongre R, Provasek V, Khan FI, Mehdi Z, Ahmed OG, Takashima M
Outcomes of Mobile Health Use in Sinonasal Surgery: Retrospective Cohort Study
JMIR Mhealth Uhealth 2026;14:e75403
URL: https://mhealth.jmir.org/2026/1/e75403
doi:10.2196/75403

© Heli Majeethia, Akshay R Prabhakar, Justina R Varghese, Najm S Khan, Roshan Dongre, Vincent Provasek, Faizaan I Khan, Zain Mehdi, Omar G Ahmed, Masayoshi Takashima. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 21.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Long-Term Engagement With a Physical Activity App Among Gynecologic Cancer Survivors in the LETSGO Trial: 12-Month Prospective, Multicenter, Quasi-Experimental Study

Sindre Herskedal Fosstveit[1,2], PhD; Ingvild Vistad[3,4], MD, PhD; Mette Skorstad[3,5], MD, PhD; Sveinung Berntsen[1,2], PhD

[1]Department of Sport Science and Physical Education, Faculty of Health and Sport Sciences, University of Agder, Universitetsveien 25, Kristiansand, Norway

[2]Research Unit, Sorlandet Hospital, Kristiansand, Norway

[3]Department of Gynecology and Obstetrics, Sorlandet Hospital, Kristiansand, Norway

[4]Institute of Clinical Medicine, University of Oslo, Oslo, Norway

[5]Medical Department, Clinical Institute II, University of Bergen, Bergen, Norway

Corresponding Author:
Sindre Herskedal Fosstveit, PhD
Department of Sport Science and Physical Education, Faculty of Health and Sport Sciences, University of Agder, Universitetsveien 25, Kristiansand, Norway

Abstract

Background: Physical activity (PA) alleviates many treatment-related side effects in gynecologic cancer survivors, yet long-term PA levels remain low. Mobile health interventions can support self-management and increase PA levels; however, evidence from real-world, year-long engagement with smartphone apps in this population is still limited.

Objective: The aim of this study is to describe 12-month user engagement with the PA component of a smartphone app implemented within a partially nurse-led routine follow-up in a real-world cohort of gynecologic cancer survivors.

Methods: This descriptive study analyzed server-generated log data from the intervention arm of the prospective, multicenter, quasi-experimental LETSGO (Lifestyle and Empowerment Techniques in Survivorship of Gynecologic Oncology) trial (NCT04122235). Between December 2019 and July 2022, 378 cancer survivors (ovarian, endometrial, cervical, vulvar, or vaginal cancer) from 5 Norwegian hospitals were enrolled in the intervention arm and were offered the app plus a Garmin Vivofit 4 activity tracker alongside standard consultations. Primary outcomes for this study were (1) weekly PA registrations (objective step counts from the activity tracker and self-reported PAs) and (2) temporal patterns of step logging during each cancer survivor's first 52 weeks postenrollment. Secondary analyses compared baseline characteristics of app users (≥2 wk of PA logging) and nonusers.

Results: Of 378 eligible participants (mean age 63, SD 13 y; BMI 28.5, SD 6.4 kg/m²), 272 (72%) logged at least 2 weeks of PA, and 225 (60%) synchronized objective step data. Mean daily steps were 5657 (SD 2799; median 5533, IQR 3499 - 7498). Step tracking dominated app use (mean 21, SD 16 logged wk), followed by self-reported walking (mean 23, SD 19 wk) and resistance training (mean 15, SD 16 wk). Weekly step-logger counts fell around study weeks 16, 32, and 44, but rebounded by 30 to 50 participants within 3 weeks, indicating episodic rather than permanent disengagement. App users were younger (mean difference –6.7 y, 95% CI –9.6 to –3.9; P<.001), more often employed (χ^2_1=12.7; P<.001), and more likely to have higher education (χ^2_2=13.6; P<.001) than nonusers. Tumor type and treatment modality were not associated with engagement.

Conclusions: In a routine-care setting, nearly three-quarters of gynecologic cancer survivors engaged repeatedly with an app-supported PA module over 12 months, although mean PA levels were modest and participation was more prevalent among younger, employed, and more highly educated participants. Engagement followed an ebb-and-flow pattern, suggesting that built-in re-engagement prompts and equity-focused onboarding are needed to sustain and broaden participation. These findings support the feasibility of blended mobile health follow-up while highlighting the importance of adaptive strategies to promote long-term PA adherence and bridge the digital divide among cancer survivors.

Trial Registration: ClinicalTrials.gov NCT04122235; https://clinicaltrials.gov/study/NCT04122235

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2021-050930

(JMIR Mhealth Uhealth 2026;14:e84838) doi:10.2196/84838

XSL•FO
RenderX

KEYWORDS

survivorship care; mobile health; mHealth; eHealth; mobile app; step count; engagement; adherence; self-management; nurse-led follow-up; real-world evidence

Introduction

Survivorship rates after gynecologic cancers are rising globally, leading to a growing number of women living with long-term treatment sequelae [1]. Gynecologic cancer survivors frequently experience treatment-related late effects that can significantly impact their quality of life, with fatigue, neuropathy, lymphedema, sexual dysfunction, cognitive impairment, anxiety, and depression being among the most commonly reported [2-8].

Regular physical activity (PA) has been shown to alleviate many late effects, providing both psychological and physiological benefits [9]. In addition, PA is associated with improved health-related quality of life and a reduced risk of cancer-related morbidity, making it a crucial component of survivorship care [10,11]. Consistent with this, objectively measured PA has recently been associated with a lower risk of cardiovascular disease in cancer survivors [12]. In recent years, increasing emphasis has been placed on promoting self-management in cancer rehabilitation, with a focus on fostering a healthy lifestyle [13], of which PA is a central element. Despite these recognized benefits and the growing focus on self-managed survivorship care, PA engagement among gynecologic cancer survivors remains low, often due to treatment-related side effects, fear of injury, and limited referral pathways to exercise professionals [14-16]. This persists despite contemporary practice guidance recommending routine brief PA promotion and formal referral within multidisciplinary oncology care [17].

Mobile health (mHealth) interventions, including smartphone apps and wearable PA trackers, have emerged as promising tools for promoting PA among cancer survivors [18]. These technologies facilitate real-time feedback, goal setting, and self-monitoring, which may enhance motivation and adherence to PA recommendations [19]. A systematic review by Khoo et al [20] found that mHealth interventions effectively increased PA levels in cancer survivors. In practice, though, these benefits depend on continued use. Engagement with mHealth typically declines after an early peak, and maintaining use over time has proved more difficult than achieving initial uptake [21,22]. The decline is rarely steady: many patients stop for a period and then start again, so looking only at average use can mask meaningful patterns of disengagement and re-engagement [23]. Even so, real-world evidence on how cancer survivors engage with these tools over the long term (≥12 mo) remains limited, particularly among gynecologic cancer survivors [18,24].

Existing evidence on mHealth interventions in populations with cancer is primarily based on controlled trials with selected, potentially more motivated populations, which may limit the generalizability of findings to broader cancer survivor groups [25]. Uptake of digital health also tends to be lower among older, less-educated, and lower-income survivors, widening a digital divide that real-world programs need to address [26]. To address this evidence-practice gap, the LETSGO (Lifestyle and Empowerment Techniques in Survivorship of Gynecologic Oncology) project was developed as a pragmatic program embedded in routine, partly nurse-led follow-up across multiple Norwegian hospitals, with app support to promote self-management among gynecologic cancer survivors [27]. A key feature of the app was its PA component, which included goal setting, PA tracking, and structured exercise programs tailored to users with varying levels of prior PA experience. Implemented without exclusions based on baseline PA levels, LETSGO enabled evaluation of real-world engagement with digital PA tools in a clinically heterogeneous cohort of gynecologic cancer survivors.

Accordingly, the aim of this study was to explore user engagement with the PA component of the LETSGO app over the first 12 months of follow-up. Specifically, we examined how participants interacted with the PA module by assessing PA logging frequency, the types of PA recorded, and patterns of objectively recorded steps over time.

Methods

Study Design and Setting

This study is part of the LETSGO project, a prospective, multicenter, quasi-experimental clinical cohort study (ClinicalTrials.gov identifier: NCT04122235) evaluating a follow-up program for patients with gynecologic cancer, which includes alternating nurse-led and physician-led follow-up consultations at the intervention hospitals [28]. The nurses focused on coaching techniques and were supported by a mobile app. The complete study spans 3 years, comparing the combined intervention (nurse-led consultations and a mobile app) with standard follow-up care at control hospitals [27]. Patients followed the intervention or the standard pathway according to their treatment hospital rather than through individual randomization [28].

This study provides a descriptive analysis of 1-year usage data from the PA component of the LETSGO app. Data were collected from participants in the intervention group from 5 hospitals, who had access to the LETSGO app as part of their follow-up care. Usage patterns and PA logging behavior were examined based on app-generated log data collected over each participant's first 12 months in the study. Server log data (eg, PA registrations in the app) were collected automatically through a secure server using an encrypted connection.

Inclusion and Exclusion Criteria

Participants were eligible if they were aged 18 years or older, were within 3 months of completing primary treatment for a histologically verified gynecologic malignancy (ovarian, endometrial, cervical, vulvar, or vaginal cancer), and had entered the survivorship follow-up phase of care.

Participants were excluded if they had disease progression at the time of enrollment, were participating in other clinical trials, required ongoing intravenous therapy treatment, or had

XSL•FO
RenderX

undergone trachelectomy for cervical cancer. Any participant who experienced cancer recurrence or progression during the study was subsequently excluded and referred for appropriate treatment.

Participant Characteristics

The 378 enrolled participants were gynecologic cancer survivors with endometrial, ovarian, cervical, vulvar, or vaginal cancer.

Sampling Procedures

Between December 2019 and July 2022, all eligible patients entering survivorship follow-up at the 5 intervention hospitals were identified and invited consecutively by clinical staff, using information from medical records and multidisciplinary team meetings; no random selection or prespecified quota was applied. The 378 enrolled participants received access to the LETSGO app as part of their routine follow-up care.

Sample Size, Power, and Precision

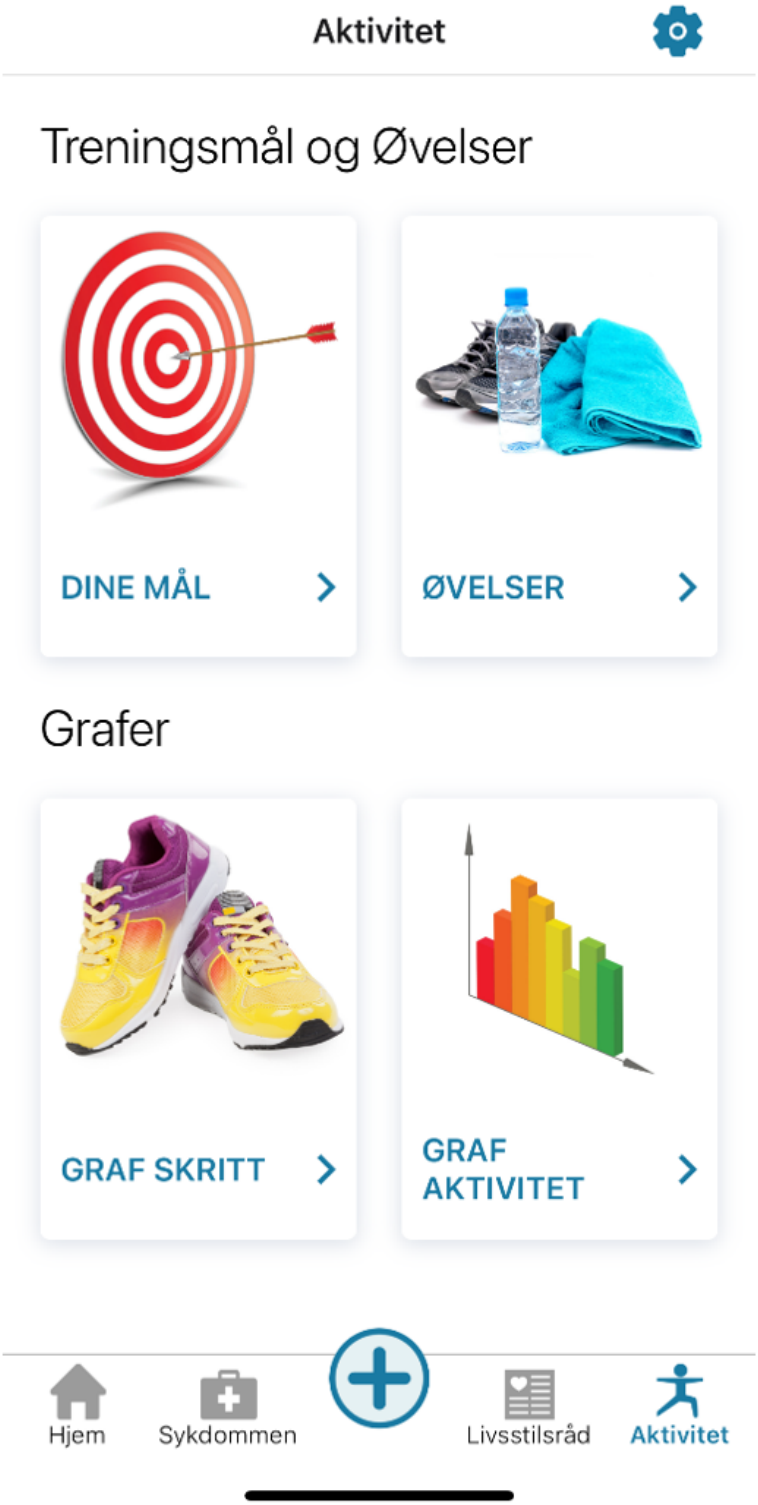
No a priori sample-size calculation was performed for this analysis; the parent LETSGO trial was powered for its primary outcome of change in patient empowerment, which required 377 participants per group [28]. As a descriptive study, this analysis included all 378 survivors enrolled in the intervention arm during the recruitment period, representing the full available population.

Ethical Considerations

The LETSGO study was approved by the Regional Committee of Medical Research Ethics in Norway (2019/11093) and received additional approval from the institutional review board and the data protection officer at each study site. The study was conducted in accordance with Norwegian data protection regulations, and all participants provided written informed consent. Participants were informed of their right to withdraw from the study at any time without consequences for their further treatment. All data were deidentified prior to analysis, with participants assigned a unique study identifier; no directly identifiable personal information was included in the analytic dataset. No financial compensation was provided to participants for their involvement in the study. No identifiable participants or users appear in any figure or the supplementary material in Multimedia Appendix 1; Figure 1 shows only the app interface.

XSL•FO
RenderX

Figure 1. Overview of the physical activity interface of the LETSGO (Lifestyle and Empowerment Techniques in Survivorship of Gynecologic Oncology) smartphone app used by gynecologic cancer survivors. The interface includes sections for setting training goals (*Dine Mål*), accessing structured exercise programs (*Øvelser*), and monitoring activity data through graphical representations of step count (*Graf Skritt*) and logged activities (*Graf Aktivitet*).



The LETSGO Intervention

A detailed description of the intervention hospitals and their roles in the LETSGO trial is provided in the protocol paper by Vistad et al [27]. In this study, only aspects relevant to the implementation of the digital follow-up program and the PA component are presented.

Throughout the follow-up period, participants alternated between nurse-led and physician-led consultations, with nurses conducting every second scheduled consultation. The nurse-led consultations incorporated motivational strategies to promote PA engagement and support participants in managing late effects of cancer treatment. The program's intervention model draws on self-determination theory [29], which posits that sustained

XSL•FO
RenderX

behavior change is promoted when 3 basic psychological needs are supported: autonomy, competence, and relatedness.

At their first nurse-led consultation, participants were introduced to the LETSGO app with several features, including structured information on PA, goal-setting features, and progress tracking. The app was supplemented by a Garmin activity tracker (Garmin Vivofit 4; Garmin Ltd), which enabled users to track their step counts and synchronize PA data with the app. During this consultation, nurses assessed participants' prior PA levels and collaboratively set individualized PA goals tailored to their motivation and perceived barriers. Nurses assisted participants in downloading the app and pairing their Garmin Vivofit 4 activity tracker with the app during this initial consultation.

At follow-up consultations, nurses reviewed each participant's app-logged PA data together with the participant, discussed progress toward individualized goals, and adjusted targets as needed. All study nurses had completed a 2-day intensive training course covering PA counseling, individualized goal setting, autonomy-supportive communication, and motivational interviewing techniques, reinforced by a mandatory e-learning program. Nurses applied an autonomy-supportive communication approach inspired by motivational interviewing [30], emphasizing participant empowerment and self-efficacy in PA adherence. Within the self-determination theory framework, autonomy was supported through this communication style and through participant-led goal setting; competence was fostered through individualized, graded exercise prescriptions and self-monitoring of progress via the app; and relatedness was maintained through the ongoing nurse-led consultations that provided a consistent relational context for PA support. Participants without smartphones (n=35) were provided with printed material containing the same PA guidance.

PA Component of the App

The PA component of the LETSGO app was designed to support gynecologic cancer survivors in maintaining or increasing their PA levels during follow-up care. The app, available on smartphones and tablets, provided participants with structured exercise programs, goal-setting functionality, and self-monitoring tools to facilitate engagement in PA.

The PA goals and exercises section included 2 features: "your goals" and "exercises" (Figure 1). In the goal-setting module, participants recorded their weekly PA goals. The exercise module included structured resistance and endurance training programs, accompanied by written instructions and video demonstrations, suitable for both beginner and advanced levels. Detailed descriptions of the app's training content, including resistance and endurance training modules, are provided in Multimedia Appendix 1.

The activity graphs section included a "step graph" that displayed step-count data and an "activity graph" that showed the frequency and distribution of different logged PAs (Figure 1).

The PA component of the LETSGO app incorporated several behavior change techniques (BCTs) from the BCT Taxonomy v1 [31]. Participants engaged in self-monitoring of behavior by reviewing weekly step counts and logging PA frequency through in-app graphs. Goal-setting (behavior) was supported through individual PA targets set collaboratively with the nurse at each consultation and revisited at follow-up. Graded task difficulty was embedded in the structured resistance and endurance exercise programs. Prompts and cues were delivered via push notifications, encouraging weekly activity logging.

Measures and Covariates

The primary outcome of this study was participant engagement with the PA component of the LETSGO app over each participant's first 12 months in the study. For descriptive analyses, an active week was defined as a study week with at least one logged PA registration, including tracker-synchronized step data and/or self-reported PA. Weekly activity was used as the unit of analysis to describe the frequency of app interaction over time.

In addition to overall engagement, the study examined the types of PAs registered in the app, including steps, jogging, cycling, resistance training, yoga or Pilates, swimming, group classes, and other self-reported activities. Step registrations were objectively measured using a Garmin Vivofit 4 activity tracker, while all other PAs were self-reported by participants. For step-based analyses, Garmin Vivofit 4 data were available in the LETSGO server log as weekly synchronized registrations. For each such week, the server provided a precomputed average daily step variable, defined as the total number of steps recorded during that calendar week divided by 7. Thus, the denominator for weekly average daily steps was always 7 calendar days within a synchronized week.

Secondary outcomes included differences in demographic characteristics between participants who engaged with the app (logged at least 2 weeks of PA app usage) and those who did not, as well as temporal patterns in step-logging frequency and the total number of daily steps recorded over the 12-month period. This threshold was chosen pragmatically to distinguish repeated use from no use or a single, isolated week of use.

Baseline covariates considered as potential correlates of engagement were age, body weight, BMI, educational level, employment status, tumor type, and treatment modality. Age (from the date of birth), tumor type, and treatment modality were obtained from the patients' medical records, and educational level and employment status were obtained from baseline questionnaires [27,28]; BMI was calculated from weight and height (kilograms per square meter).

Data Analysis

All statistical analyses were conducted using SPSS (version 29; IBM Corp), and figures were created using GraphPad Prism (version 8.0.1; GraphPad Software). Descriptive statistics were used to summarize demographic and clinical characteristics. The normality of continuous variables was assessed using histograms and measures of skewness and kurtosis. Independent *t* tests (2-tailed) were performed for normally distributed variables, while the Mann-Whitney *U* test was used for nonnormally distributed variables. Associations between categorical variables were examined using the Pearson chi-squared test. Statistical significance was set at $P<.05$ for all analyses. The methods were structured and reported in

XSL•FO
RenderX

accordance with the APA Journal Article Reporting Standards for Quantitative Research (JARS-Quant) [32]. This study was reported in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guideline for cohort studies (Checklist 1). Missing baseline data were quantified; the missing-data mechanism was assessed with the Little test for data missing completely at random, and a multiple-imputation sensitivity analysis (m=20, chained equations, pooled with Rubin rules) was performed for the variables with missing data.

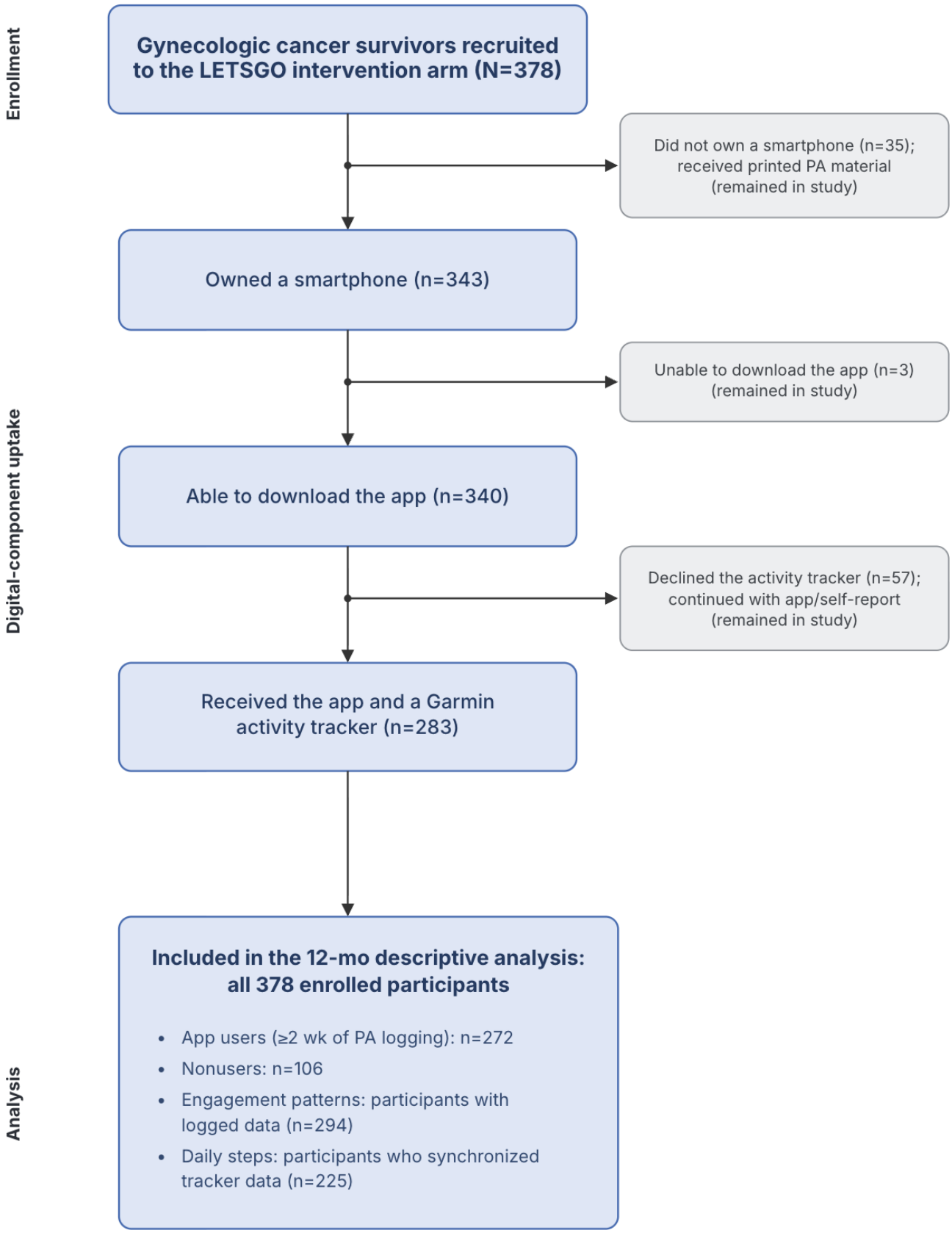
Results

Study Participants

Figure 2 displays participant flow through the digital pathway of the LETSGO intervention. Of the 378 recruited participants, 343 (90.7%) owned a smartphone, and 340 (89.9%) were able to download the app. A total of 283 (74.9%) received both the app and a Garmin activity tracker.

XSL•FO
RenderX

Figure 2. Participant flow through the LETSGO (Lifestyle and Empowerment Techniques in Survivorship of Gynecologic Oncology) intervention arm. The chart shows enrollment, uptake of digital components (smartphone, app, and Garmin activity tracker), and analysis. Counts beside the main pathway indicate the number of participants who did not take up a given component; all 378 enrolled participants were included in the 12-month descriptive analysis. PA: physical activity.



The baseline demographic and clinical characteristics, overall, and by app-engagement group, are summarized in Table 1.

Participants who engaged with the PA component of the app were significantly younger than those who did not (mean difference –6.7 y, 95% CI –9.6 to –3.9; t_{376}=–4.6; P<.001). No

XSL•FO
RenderX

significant differences were observed for body weight (mean difference 0.2 kg, 95% CI −4.5 to 4.8; *P*=.94) or BMI (mean difference −0.3 kg/m², 95% CI −1.9 to 1.3; *P*=.73).

There were significant associations between educational level (χ^2_2=13.6; *P*<.001) and engagement with the app for PA, and between employment status (χ^2_1=12.7; *P*<.001) and engagement with the app for PA. There were no significant associations of cancer diagnosis or treatment modality with engagement with the app for PA.

Missing baseline data for educational level, employment status, and weight or BMI are summarized in Table 1; all other variables (age, tumor type, and treatment modality) were complete.

Table . Baseline demographic and clinical characteristics of gynecologic cancer survivors in the LETSGO[a] intervention arm, overall, and by app-engagement group[b].

Variable	Total (N=378)	App user (n=272)	Non–app user (n=106)	*P* value
Age (y), mean (SD)	63 (13)	61 (12)	68 (13)	<.001
BMI (kg/m^2), mean (SD)	28.5 (6.4)	28.4 (6.5)	28.7 (6.3)	.73
Missing, n	41	14	27	
Body weight (kg), mean (SD)	79 (19)	79 (19)	79 (17)	.94
Tumor type, n (%)				.27
Endometrial	221 (59)	154 (57)	67 (63)	
Ovarian	92 (25)	73 (27)	19 (18)	
Cervical	58 (15)	41 (15)	17 (16)	
Vulvar or vaginal	7 (2)	4 (2)	3 (3)	
Treatment modality, n (%)				.31
Surgery only	213 (56)	147 (54)	66 (62)	
Chemotherapy and radiotherapy	19 (5)	16 (6)	3 (3)	
Surgery and chemotherapy	121 (32)	93 (34)	28 (26)	
Other combinations[c]	25 (7)	16 (6)	9 (9)	
Employment[d], n (%)				<.001
Yes	120 (35)	105 (40)	15 (19)	
No	219 (65)	154 (60)	65 (81)	
Missing, n	39	13	26	
Education level[d], n (%)				<.001
Primary	50 (15)	28 (11)	22 (28)	
Secondary	156 (46)	125 (48)	31 (39)	
Medium or higher	133 (39)	106 (41)	27 (34)	
Missing, n	39	13	26	

[a]LETSGO: Lifestyle and Empowerment Techniques in Survivorship of Gynecologic Oncology.

[b]Percentages may not sum to 100 due to rounding.

[c]Includes chemotherapy only or any other combination of treatments, including hormonal treatment.

[d]Not all participants reported their employment and education levels; thus, the total is lower than the total number of participants. App user: a participant who logged physical activity (PA) in the app in ≥2 distinct wk during their first 52-wk postenrollment (counting objective step synchronizations and/or self-reported PA registrations). Non–app user: a participant with <2 wk of PA logging in the app (ie, 0 or 1 wk).

PA Logging and Volume

Table 2 presents the number of participants who recorded different PAs in the LETSGO app on a weekly basis, along with the mean number of PA registrations per participant, where each registration represents 1 week of recorded activity.

Step tracking was the most frequently used feature, recorded by 225 (60%) participants, with an average of 21 (SD 16) weeks of registration. Walking (n=53; mean 23, SD 19 wk) and resistance training (n=39; mean 15, SD 16 wk) were the most commonly reported self-logged activities. Other types of PA,

XSL•FO
RenderX

such as bicycling, yoga, and swimming, were logged by fewer participants, with an average of 10 to 30 participants.

Table . Engagement with objective step tracking and self-reported physical activity types in the LETSGO[a] app over 52 weeks, among gynecologic cancer survivors (N=378)[b].

Modality	Participants, n (%)	Weeks logged per participant	
		Mean (SD)	Median (IQR)
Objective (wearable-synchronized)			
Steps	225 (60)	21 (16)	18 (7 - 35)
Self-reported			
Walking	53 (14)	23 (19)	21 (3 - 41)
Resistance training	39 (10)	15 (16)	7 (2 - 21)
Bicycling	30 (8)	13 (14)	8 (3 - 17)
Other	24 (6)	15 (17)	10 (3 - 25)
Group classes	14 (4)	7 (11)	3 (2 - 8)
Jogging	10 (3)	12 (11)	10 (4 - 17)
Yoga or Pilates	9 (2)	12 (13)	9 (1 - 17)
Swimming	9 (2)	7 (6)	6 (2 - 12)

[a]LETSGO: Lifestyle and Empowerment Techniques in Survivorship of Gynecologic Oncology.

[b]Percentages are based on the total cohort (N=378). Weeks logged represents the number of distinct weeks during which a participant recorded the modality, averaged across participants who ever logged it.

Table 3 presents the weekly duration of performed PAs recorded in the LETSGO app. Participants registered an average of 5657 daily steps (SD 2799, 95% CI 5291 - 6023; median 5533, IQR 3499 - 7498). Walking had the highest reported average weekly duration of 172 (SD 116) minutes, followed by other self-reported activities (118, SD 80 min) and yoga or Pilates (93, SD 78 min). Resistance training, group classes, and bicycling ranged from 54 (SD 36) to 74 (SD 30) minutes per week, while jogging and swimming had the lowest average durations, at 37 (SD 20) and 40 (SD 16) minutes, respectively.

Table . Weekly physical activity volume by modality recorded in the LETSGO[a] app over 52 weeks, among gynecologic cancer survivors (N=378)[b].

Modality	Participants, n	Weekly dose	
		Mean (SD)	Median (IQR)
Objective, steps/d			
Steps	225	5657 (2799)	5533 (3499 - 7498)
Self-reported, min/wk			
Walking	53	172 (116)	147 (92 - 230)
Resistance training	39	61 (38)	52 (34 - 73)
Bicycling	30	54 (36)	53 (24 - 80)
Other	24	118 (80)	99 (55 - 167)
Group classes	14	74 (30)	75 (52 - 89)
Jogging	10	37 (20)	32 (28 - 43)
Yoga or Pilates	9	93 (78)	61 (46 - 99)
Swimming	9	40 (16)	35 (34 - 47)

[a]LETSGO: Lifestyle and Empowerment Techniques in Survivorship of Gynecologic Oncology.

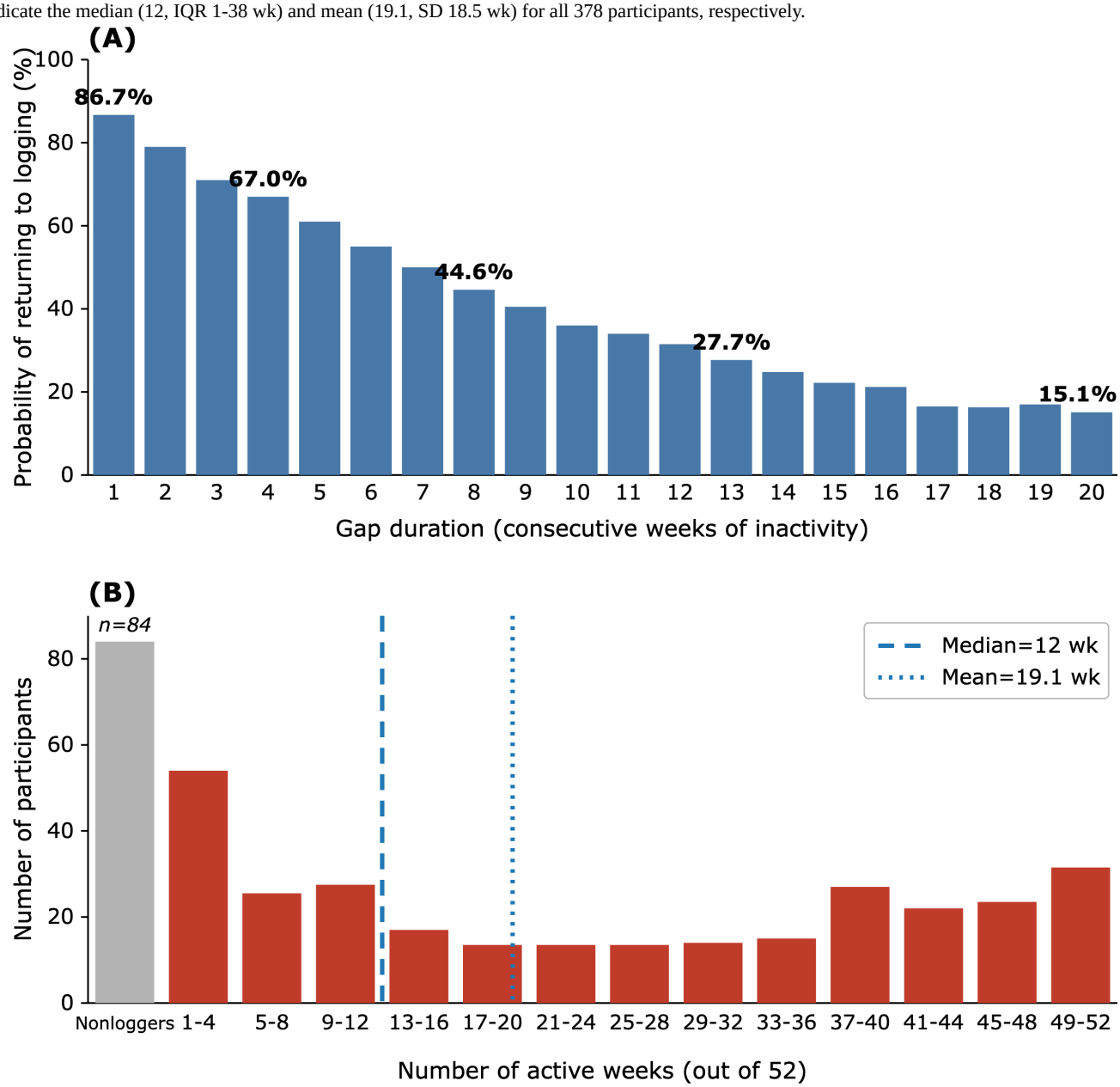
[b]n is the number of participants who logged that modality at least once over 52 weeks (for steps, the number of participants who synchronized step data ≥1 time). The weekly dose for self-reported modalities is the average number of minutes per week during weeks when that modality was logged, averaged across participants who ever logged it. Steps are the participant-level mean daily steps across weeks with synchronized data. All values reflect weeks in which physical activity was recorded and are not averaged across the entire 52-week follow-up.

XSL•FO
RenderX

Engagement Patterns Over Time

Among all 378 enrolled participants, the median number of active logging weeks was 12 (IQR 1 - 38; mean 19.1, SD 18.5) out of 52 available weeks. The distribution was bimodal, with concentrations at both the low end (0 - 4 active wk, n=138) and the high end (41 - 52 active wk, n=76; Figure 3B). Among the 294 participants who logged at least one activity registration, the median was 23.5 (IQR 8 - 41) active weeks, and only 8 (3%) logged every week without interruption.

Figure 3. Episodic engagement with the physical activity component of the LETSGO (Lifestyle and Empowerment Techniques in Survivorship of Gynecologic Oncology) app over the first 12 months (52 wk) among gynecologic cancer survivors (N=378). (A) Probability of returning to activity logging after an inactivity gap of a given duration (n=294 participants with logged data). Participants who experienced a gap were classified as having "returned" if at least 1 additional log registration followed the gap, or as "not returned" if the gap extended to the end of the 52-week follow-up. (B) Distribution of the total number of active logging weeks (out of 52) across all 378 enrolled participants. The gray bar represents the 84 participants who never logged any activity. Red bars show the distribution among the 294 participants with logged data in 4-week intervals. Dashed and dotted lines indicate the median (12, IQR 1-38 wk) and mean (19.1, SD 18.5 wk) for all 378 participants, respectively.



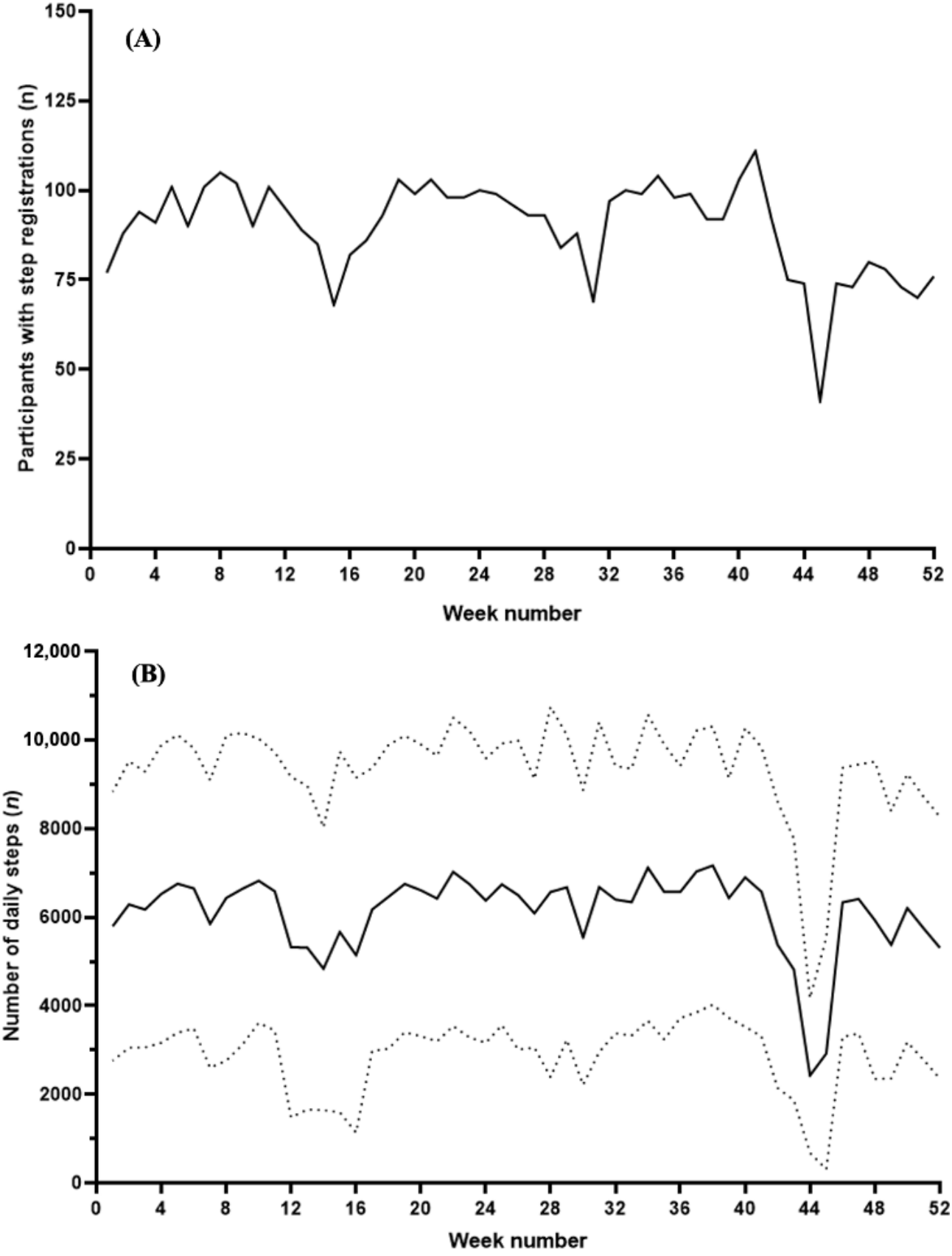
Re-engagement after inactivity gaps was common. After gaps of 1 or more weeks, 86.7% (248/286) of participants returned to logging. Return probability remained above 50% for gaps up to 7 weeks and was 27.7% (33/119) for gaps of 13 or more weeks (Figure 3A).

Figure 4A presents the number of participants who registered their step counts over the first 12 months of the study. Weekly registrations remained stable with fluctuations, showing declines around weeks 16, 32, and 44, followed by partial recovery.

Figure 4B displays the average number of daily steps recorded by the LETSGO app. Step counts fluctuated throughout the study period, with declines occurring during the same weeks as the drop in registrations. The lowest step counts were observed in the final quarter of the study.

XSL•FO
RenderX

Figure 4. Weekly objective step engagement over the first 12 months (52 wk) among gynecologic cancer survivors in the LETSGO (Lifestyle and Empowerment Techniques in Survivorship of Gynecologic Oncology) intervention arm. (A) Number of participants registering step data each week. (B) Weekly mean daily steps. (B) The solid line represents the weekly mean number of daily steps, and the dotted lines represent the mean and 1 SD. Across the 225 participants who synchronized step data, the overall mean was 5657 daily steps (SD=2799).



XSL•FO
RenderX

Sensitivity Analysis

The Little test did not reject the assumption that the baseline data were missing completely at random (χ^2_4=7.13; *P*=.13). Because missingness was nonetheless somewhat more frequent among nonusers than among app users, we performed a multiple-imputation sensitivity analysis (*m*=20 imputations; chained equations; pooled with Rubin rules); the associations between engagement and educational level and employment status remained significant in each imputed dataset and after pooling (education: combined *P*=.001; employment: *P*<.001).

Discussion

Principal Findings

This 12-month real-world follow-up study of 378 gynecologic cancer survivors shows that almost three-quarters of those offered the LETSGO app engaged with its PA component at least twice, and that 225 participants synchronized their objective step counts via an activity tracker. Mean daily steps remained modest (≈5700, SD 2800) and fluctuated over time, with notable dips around weeks 16, 32, and 44 before a partial rebound. App engagement was higher among younger, employed, and more highly educated participants, whereas tumor type and treatment modality were not associated with usage.

Variability in Digital Adherence

Despite evidence that mHealth interventions can increase PA levels [20], engagement remains highly variable between individuals, across studies, and over time. A 2024 systematic review by Montalescot et al [33] highlights the complexity of adherence in digital cancer care, noting that most demographic and clinical factors, including age, sex, and tumor type, show inconsistent associations with app usage. Moreover, sustaining long-term adherence to mHealth interventions remains challenging, as documented in a recent review by Amagai et al [21], which is consistent with the "law of attrition" proposed by Eysenbach [22] for eHealth interventions.

Findings from the present LETSGO cohort sharpen this picture of heterogeneity. Of the 378 participants offered the app, 272 (72%) engaged with the PA module at least twice, and 225 (60%) synchronized tracker–derived step counts during their first 12 months of individual follow-up. Participants predominantly engaged in low-impact, ambulatory PAs, with walking accounting for the majority of all recorded sessions. Yet engagement was far from uniform: users registered a mean of 21 (SD 16) weeks of step data, but the range spanned from a single upload to more than 50 weeks. Among those who logged walking, the average duration was approximately 172 minutes per week. In contrast, muscle-strengthening activities were reported less frequently, with a mean of 15 (SD 16) logged weeks and an average of approximately 61 minutes per week, rarely meeting the recommended minimum of 2 or more sessions per week [9]. Other modalities, including cycling, yoga or Pilates, swimming, and group exercise classes, were recorded sporadically by a small subset of participants (≤30). These usage patterns may reflect both technological and behavioral factors. First, automated step-count synchronization reduces the self-monitoring burden relative to manual registration, which likely contributed to the higher adherence to step tracking observed in this study [34]. Second, resistance training is a comparatively complex modality, often requiring equipment, technique mastery, and greater self-efficacy, especially for beginners. Accordingly, cancer survivors often prefer walking, which is perceived as simple, safe, and equipment-independent [16,35].

The objectively measured mean of 5657 steps per day places the cohort in the "low-active" range [36]. This step volume is below the approximate threshold of 7000 steps per day, estimated to align with the 150 minutes of moderate-intensity PA per week [36] recommended for cancer survivors by the American College of Sports Medicine and the American Society of Clinical Oncology [9,37]. However, it is comparable to those reported in other real-world accelerometer-based studies in mixed cancer populations, which typically range from 4500 to 6000 steps per day [38,39]. Epidemiological data indicate that most of the reduction in all-cause mortality risk among cancer survivors is realized at approximately 5000 to 6000 steps per day [39]. Notably, the cohort's mean of 5657 steps per day falls within this range, indicating that participants were likely engaging in a level of PA associated with meaningful health benefits in this population. Collectively, these findings highlight that PA logging was dominated by walking and that targeted strategies to increase resistance training frequency may be necessary to achieve the full spectrum of benefits outlined in current exercise-oncology guidelines [9,37].

Episodic Use and Re-Engagement

Real-world data from other studies confirm that engagement often wanes and is episodic rather than continuous. In the 12-month IMPACT randomized controlled trial of a mobile PA program for young adult cancer survivors, median weekly log-ins fell from 89% (23/26) of study weeks in months 0 to 6 to 42% (11/26; IQR 2 - 22 wk) during months 7 to 12, with only a subset of participants sustaining regular use [40,41]. High attrition is likewise a recurring theme in the broad review by Amagai et al [21] of 62 mHealth trials, most of which were only "partially successful" at preserving long-term engagement. The LETSGO data align with these findings: the weekly number of participants uploading step data fell at several points, most notably around study weeks 16, 32, and 44, yet recovered by 30 to 50 participants within the next 1 to 3 weeks (Figure 4). These rebounds imply that a meaningful proportion of participants who became "inactive" subsequently re-engaged, illustrating that lapses frequently reflect temporary disengagement rather than permanent dropout. This ebb-and-flow dynamic suggests that blended-care models are well positioned to support re-engagement.

This analysis quantified this re-engagement pattern directly. Among the 294 participants with any logged data, 286 experienced a gap of at least 1 week; of these, 86.7% (248/286) returned to logging after a gap of at least 1 week, and the return probability remained above 50% for gaps up to 7 consecutive weeks. Even after gaps exceeding a full follow-up quarter (≥13 wk), 27.7% of participants resumed logging (Figure 3A). The distribution of active weeks was bimodal, with clusters at both

XSL•FO
RenderX

the low end (0 - 4 wk, n=138) and the high end (41 - 52 wk, n=76), indicating 2 distinct engagement profiles rather than a uniform decline (Figure 3B). These patterns are consistent with emerging evidence that mHealth engagement is better characterized as a dynamic, nonlinear process than as a simple trajectory of decline [23]. In a systematic review and meta-analysis of app-based chronic disease interventions, Meyerowitz-Katz et al [42] reported a pooled dropout rate of 43%, yet this binary framing of "retained versus dropped out" does not capture the cyclical re-engagement observed here. Perski et al [23] have similarly argued that engagement with digital behavior change interventions involves a dynamic interplay between usage behavior and subjective experience that fluctuates within and across individuals over time. These data lend empirical support to this conceptualization and suggest that the conventional emphasis on preventing attrition may be usefully complemented by strategies that facilitate re-engagement after inevitable lapses, such as the just-in-time adaptive interventions described by Nahum-Shani et al [43].

Because each cancer survivor's "week 0" corresponded to their enrollment date over the study period, these drops cannot be ascribed to calendar seasons but instead indicate phase-specific motivational cycles. Consistent with the mixed evidence in the review by Montalescot et al [33], tumor type and treatment modality were not associated with app use, yet analyses identified modest socioeconomic gradients: app users were younger (by ≈6.7 y) and more likely to be employed or highly educated than nonusers (Table 1). These gradients underline the persistent digital divide that many mHealth survivorship programs must still address. The BCTs embedded in the LETSGO app—self-monitoring, goal setting, graded exercise prescription, and behavioral prompting—are consistent with engagement-sustaining mechanisms identified in systematic reviews of mHealth PA interventions in cancer survivors [44,45]. The progressive structure of these techniques may partly explain the declining-but-sustained engagement pattern observed over 12 months, as goal review and graded task progression provide ongoing behavioral reinforcement beyond the initial novelty period of app adoption.

Implications for Intervention Development

Together, the literature and these findings converge on 3 imperatives. First, engagement trajectories must be monitored continuously, recognizing that cancer survivors' needs and motivation ebb and flow. Second, interventions should build in flexible re-engagement mechanisms, whether automated prompts, personalized feedback, or brief human contact, to draw inactive users back at critical junctures. Third, broad eligibility alone is insufficient; programs should include targeted digital-literacy assistance, alternative low-tech access routes, and tailored outreach to ensure that socioeconomically disadvantaged cancer survivors can initiate and sustain use on equal terms with their more advantaged peers.

The observed sociodemographic gradients in app use, with older, less educated, and unemployed participants consistently less likely to engage, likely reflect intersecting barriers including lower digital literacy, reduced familiarity with wearable technology, and differential access to smartphones [46]. These findings align with broader evidence of a persistent digital divide in cancer survivorship care, where older age, lower education, and socioeconomic disadvantage are consistently associated with reduced uptake of digital health tools [46,47]. Importantly, device access alone does not guarantee meaningful engagement, as barriers to uptake and sustained use operate independently of access [47]. Future survivorship programs should incorporate equity-focused implementation strategies, such as assisted device onboarding, nurse-guided digital literacy support, simplified app interfaces, and nondigital participation pathways for those unable or unwilling to engage with mHealth tools.

Strengths and Limitations

One of the main strengths of this study is its pragmatic, real-world implementation: LETSGO was embedded in routine nurse-led survivorship care and accepted participants regardless of baseline PA level or smartphone ownership. This yielded a large, clinically heterogeneous cohort (N=378) that mirrors the breadth of gynecologic cancer follow-up, enhancing external validity. The study also combined objectively synchronized step data with self-reported PA logs, providing complementary assessments of total PA level and mode diversity while reducing recall bias. In addition, the 12-month duration, longer than most oncology mHealth trials [18,24], made it possible to examine engagement trajectories across a whole year and to document the characteristic ebb-and-flow pattern of digital adherence. Finally, evaluating the app within a blended model of autonomy-supportive nurse consultations plus digital self-monitoring offers implementation insights that are immediately transferable to other survivorship settings.

Although the LETSGO study used a longitudinal, quasi-experimental design with matched control hospitals, this analysis focuses only on the intervention arm's PA module; therefore, it cannot establish causal relationships between app use and changes in PA or health outcomes. Wear-time compliance was not verified; therefore, gaps in step data may reflect tracker nonwear, true inactivity, or technical issues with the device, app, or server synchronization. In addition, weekly average daily steps were derived from server-generated weekly step totals divided by 7 calendar days; therefore, average daily step counts may have been underestimated. All nonstep activities were self-reported, making their duration estimates subject to recall and social-desirability bias. Despite open eligibility, engagement remained higher among younger, employed, and better-educated cancer survivors, indicating a residual digital divide and potential selection bias. Some baseline data were missing (10% - 11% for BMI, educational level, and employment status), somewhat more often among nonusers than app users, and although the Little test did not reject the assumption that these data were missing completely at random, we used multiple imputation as sensitivity analysis, which reproduced the socioeconomic associations, indicating that the findings are robust to missing data. Phase-specific dips in engagement could not be separated from contextual factors, such as treatment milestones or holidays, because analysis weeks were aligned with individual enrollment dates. Data collection spanned from December 2019 to July 2022, overlapping substantially with the COVID-19 pandemic (Norwegian lockdowns: March 2020-September 2021); pandemic-related

XSL•FO
RenderX

restrictions on movement, facilities, and clinical workflows may have affected both PA levels and app engagement in ways that cannot be disentangled from individual engagement trajectories. Finally, the study took place within the Norwegian health care system; extrapolating results to regions with different infrastructures or lower smartphone penetration should be approached with caution.

Overall, the study provides a realistic demonstration of embedding a blended mHealth intervention in routine survivorship care while also highlighting methodological and equity-related challenges that future controlled and implementation-focused research will need to address.

Conclusions

The LETSGO program showed that a heterogeneous cohort of gynecologic cancer survivors engaged with an app-supported PA module for up to 12 months. Although mean PA levels remained modest (mean 5657, SD 2799 steps/d; low-active range) and engagement fluctuated, nearly three-quarters of the participants interacted with the PA component of the app at least twice, and 60% (n=225) synchronized their objective step data with the app. Engagement lapses were typically temporary, with recurrent dips followed by re-engagement, indicating a dynamic of ebb and flow rather than definitive dropout. App use centered on walking, whereas resistance training was less frequent, with the recommended 2 or more sessions per week. Socioeconomic gradients in use, favoring younger, employed, and more highly educated participants, highlight the persistent digital divide. Collectively, these findings confirm the feasibility of pragmatic, blended digital follow-up in survivorship care, while also signaling the need for adaptive re-engagement strategies and equity-focused implementation supports, alongside targeted content to increase adherence to resistance training.

Acknowledgments

The authors express their gratitude to all the gynecologic cancer survivors who participated in the LETSGO (Lifestyle and Empowerment Techniques in Survivorship of Gynecologic Oncology) study and contributed their time and data. They are also thankful to the dedicated clinical staff at participating hospitals for their collaboration in implementing the intervention as part of routine follow-up care. Special thanks are extended to the nurse coordinators for their integral role in supporting participant engagement and ensuring the smooth integration of the digital tools into standard care. Finally, the authors acknowledge the contributions of the LETSGO project team and the technical partners involved in app development and data management.

The authors declare the use of generative AI (GenAI) in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: proofreading and editing. The GenAI tool used was Claude (Opus 4.8). Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

Funding

The LETSGO (Lifestyle and Empowerment Techniques in Survivorship of Gynecologic Oncology) study was funded by the Norwegian Cancer Society (grant 198057), the UNI Foundation (grant 3287), and the South-Eastern Norway Regional Health Authority (grant 2019073). The funders had no role in the study design, data collection, data analysis, interpretation of data, or preparation of the manuscript.

Data Availability

The data that support the findings of this study are available upon reasonable request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Authors' Contributions

SB and IV conceptualized the study, registered the protocol, supervised the execution of the project, and assisted with manuscript writing. MS contributed to the development of the app, implemented the intervention, and assisted with manuscript writing. SHF analyzed and interpreted the data and wrote the manuscript. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Resistance and endurance training modules.
[DOCX File, 16 KB - mhealth_v14i1e84838_app1.docx]

Checklist 1
STROBE checklist.
[DOCX File, 34 KB - mhealth_v14i1e84838_app2.docx]

XSL•FO
RenderX

References

1. Coleman MP, Allemani C. Global surveillance of cancer survival trends up to 2014 (CONCORD-3). Eur J Public Health 2018 Nov 1;28(suppl_4). [doi: 10.1093/eurpub/cky212.623]
2. Vistad I, Cvancarova M, Fosså SD, Kristensen GB. Postradiotherapy morbidity in long-term survivors after locally advanced cervical cancer: how well do physicians' assessments agree with those of their patients? Int J Radiat Oncol Biol Phys 2008 Aug 1;71(5):1335-1342. [doi: 10.1016/j.ijrobp.2007.12.030] [Medline: 18355976]
3. Vistad I, Cvancarova M, Kristensen GB, Fosså SD. A study of chronic pelvic pain after radiotherapy in survivors of locally advanced cervical cancer. J Cancer Surviv 2011 Jun;5(2):208-216. [doi: 10.1007/s11764-011-0172-z] [Medline: 21259075]
4. Vistad I, Kristensen GB, Fosså SD, Dahl AA, Mørkrid L. Intestinal malabsorption in long-term survivors of cervical cancer treated with radiotherapy. Int J Radiat Oncol Biol Phys 2009 Mar 15;73(4):1141-1147. [doi: 10.1016/j.ijrobp.2008.05.064] [Medline: 18760883]
5. Vistad I, Fosså SD, Kristensen GB, Dahl AA. Chronic fatigue and its correlates in long-term survivors of cervical cancer treated with radiotherapy. BJOG 2007 Sep;114(9):1150-1158. [doi: 10.1111/j.1471-0528.2007.01445.x] [Medline: 17655733]
6. Steen R, Dahl AA, Hess SL, Kiserud CE. A study of chronic fatigue in Norwegian cervical cancer survivors. Gynecol Oncol 2017 Sep;146(3):630-635. [doi: 10.1016/j.ygyno.2017.05.028] [Medline: 28552254]
7. Liavaag AH, Dørum A, Fosså SD, Tropé C, Dahl AA. Controlled study of fatigue, quality of life, and somatic and mental morbidity in epithelial ovarian cancer survivors: how lucky are the lucky ones? J Clin Oncol 2007 May 20;25(15):2049-2056. [doi: 10.1200/JCO.2006.09.1769] [Medline: 17513809]
8. Cianci S, Rosati A, Capozzi VA, et al. Quality of life and sexual functioning of patient affected by endometrial cancer. Minerva Med 2021 Feb;112(1):81-95. [doi: 10.23736/S0026-4806.20.07081-0] [Medline: 33104302]
9. Campbell KL, Winters-Stone KM, Wiskemann J, et al. Exercise guidelines for cancer survivors: consensus statement from International Multidisciplinary Roundtable. Med Sci Sports Exerc 2019 Nov;51(11):2375-2390. [doi: 10.1249/MSS.0000000000002116] [Medline: 31626055]
10. Thomaier L, Jewett P, Brown K, et al. The associations between physical activity, neuropathy symptoms and health-related quality of life among gynecologic cancer survivors. Gynecol Oncol 2020 Aug;158(2):361-365. [doi: 10.1016/j.ygyno.2020.05.026] [Medline: 32499071]
11. Friedenreich CM, Cook LS, Wang Q, et al. Prospective cohort study of pre- and postdiagnosis physical activity and endometrial cancer survival. J Clin Oncol 2020 Dec 1;38(34):4107-4117. [doi: 10.1200/JCO.20.01336] [Medline: 33026939]
12. Jiang C, Li Z, Guo B, et al. Wearable device-measured physical activity and incident cardiovascular disease in cancer survivors. Br J Sports Med 2025 May 2;59(10):706-714. [doi: 10.1136/bjsports-2024-108734] [Medline: 40074236]
13. Hernandez Silva E, Lawler S, Langbecker D. The effectiveness of mHealth for self-management in improving pain, psychological distress, fatigue, and sleep in cancer survivors: a systematic review. J Cancer Surviv 2019 Feb;13(1):97-107. [doi: 10.1007/s11764-018-0730-8] [Medline: 30635865]
14. Toohey K, Paterson C, Coltman CE. Barriers and enablers to participation in physical activity among women diagnosed with ovarian cancer. J Cancer Surviv 2024 Aug;18(4):1252-1263. [doi: 10.1007/s11764-023-01366-5] [Medline: 37171718]
15. Mizrahi D, Naumann F, Broderick C, Samara J, Ryan M, Friedlander M. Quantifying physical activity and the associated barriers for women with ovarian cancer. Int J Gynecol Cancer 2015 May;25(4):577-583. [doi: 10.1097/IGC.0000000000000349] [Medline: 25628105]
16. Elshahat S, Treanor C, Donnelly M. Factors influencing physical activity participation among people living with or beyond cancer: a systematic scoping review. Int J Behav Nutr Phys Act 2021 Apr 6;18(1):50. [doi: 10.1186/s12966-021-01116-9] [Medline: 33823832]
17. Spence RR, Sandler CX, Newton RU, Galvão DA, Hayes SC. Physical activity and exercise guidelines for people with cancer: why are they needed, who should use them, and when? Semin Oncol Nurs 2020 Oct;36(5):151075. [doi: 10.1016/j.soncn.2020.151075] [Medline: 33008685]
18. Haberlin C, O'Dwyer T, Mockler D, Moran J, O'Donnell DM, Broderick J. The use of eHealth to promote physical activity in cancer survivors: a systematic review. Support Care Cancer 2018 Oct;26(10):3323-3336. [doi: 10.1007/s00520-018-4305-z] [Medline: 29909476]
19. Roberts AL, Fisher A, Smith L, Heinrich M, Potts HWW. Digital health behaviour change interventions targeting physical activity and diet in cancer survivors: a systematic review and meta-analysis. J Cancer Surviv 2017 Dec;11(6):704-719. [doi: 10.1007/s11764-017-0632-1] [Medline: 28779220]
20. Khoo S, Mohbin N, Ansari P, Al-Kitani M, Müller AM. mHealth interventions to address physical activity and sedentary behavior in cancer survivors: a systematic review. Int J Environ Res Public Health 2021 May 28;18(11):5798. [doi: 10.3390/ijerph18115798] [Medline: 34071342]
21. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022 Apr 26;24(4):e35120. [doi: 10.2196/35120] [Medline: 35471414]
22. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11. [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

XSL•FO
RenderX

23. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
24. Salisbury CE, Hyde MK, Cooper ET, Stennett RC, Gomersall SR, Skinner TL. Physical activity behaviour change in people living with and beyond cancer following an exercise intervention: a systematic review. J Cancer Surviv 2023 Jun;17(3):569-594. [doi: 10.1007/s11764-023-01377-2] [Medline: 37074621]
25. Wang L, Langlais CS, Kenfield SA, et al. mHealth interventions to promote a healthy diet and physical activity among cancer survivors: a systematic review of randomized controlled trials. Cancers (Basel) 2022 Aug 6;14(15):3816. [doi: 10.3390/cancers14153816] [Medline: 35954479]
26. Benzo RM, Tetrick MK, Krok-Schoen JL, et al. Digital information-seeking behaviors among cancer survivors: associations with sociodemographic determinants, cancer history, and perceived health. J Cancer Surviv 2025 Nov 6. [doi: 10.1007/s11764-025-01927-w] [Medline: 41196508]
27. Vistad I, Skorstad M, Demmelmaier I, et al. Lifestyle and Empowerment Techniques in Survivorship of Gynaecologic Oncology (LETSGO study): a study protocol for a multicentre longitudinal interventional study using mobile health technology and biobanking. BMJ Open 2021 Jul 12;11(7):e050930. [doi: 10.1136/bmjopen-2021-050930] [Medline: 34253678]
28. Vistad I, Skorstad M, Berntsen S, et al. Empowerment and quality of life in gynecological cancer survivors: outcomes from a multicenter quasi-experimental cohort study from Norway (the LETSGO trial). Cancer 2025 Aug 15;131(16):e70040. [doi: 10.1002/cncr.70040] [Medline: 40781852]
29. Ryan RM, Deci EL. Self-determination theory. In: Maggino F, editor. Encyclopedia of Quality of Life and Well-Being Research: Springer; 2023:1-7. [doi: 10.1007/978-3-031-17299-1_2630]
30. Miller WR, Rollnick S. Meeting in the middle: motivational interviewing and self-determination theory. Int J Behav Nutr Phys Act 2012 Mar 2;9:25. [doi: 10.1186/1479-5868-9-25] [Medline: 22385872]
31. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
32. Appelbaum M, Cooper H, Kline RB, Mayo-Wilson E, Nezu AM, Rao SM. Journal article reporting standards for quantitative research in psychology: the APA Publications and Communications Board task force report. Am Psychol 2018 Jan;73(1):3-25. [doi: 10.1037/amp0000191] [Medline: 29345484]
33. Montalescot L, Baussard L, Charbonnier E. Factors associated with digital intervention engagement and adherence in patients with cancer: systematic review. J Med Internet Res 2024 Dec 11;26:e52542. [doi: 10.2196/52542] [Medline: 39661976]
34. Keats MR, Yu X, Sweeney Magee M, et al. Use of wearable activity-monitoring technologies to promote physical activity in cancer survivors: challenges and opportunities for improved cancer care. Int J Environ Res Public Health 2023 Mar 8;20(6):4784. [doi: 10.3390/ijerph20064784] [Medline: 36981693]
35. Farrokhzadi L, Dhillon HM, Goumas C, Young JM, Cust AE. Physical activity correlates, barriers, and preferences for women with gynecological cancer. Int J Gynecol Cancer 2016 Oct;26(8):1530-1537. [doi: 10.1097/IGC.0000000000000790] [Medline: 27488217]
36. Tudor-Locke C, Craig CL, Brown WJ, et al. How many steps/day are enough? For adults. Int J Behav Nutr Phys Act 2011 Jul 28;8:79. [doi: 10.1186/1479-5868-8-79] [Medline: 21798015]
37. Schmitz KH, Campbell AM, Stuiver MM, et al. Exercise is medicine in oncology: engaging clinicians to help patients move through cancer. CA Cancer J Clin 2019 Nov;69(6):468-484. [doi: 10.3322/caac.21579] [Medline: 31617590]
38. D'Silva A, Gardiner PA, Boyle T, Bebb DG, Johnson ST, Vallance JK. Associations of objectively assessed physical activity and sedentary time with health-related quality of life among lung cancer survivors: a quantile regression approach. Lung Cancer 2018 May;119:78-84. [doi: 10.1016/j.lungcan.2018.03.010] [Medline: 29656757]
39. Hyde ET, Evenson KR, Bandoli GE, et al. Accelerometer-measured physical activity, sedentary behavior, and mortality among cancer survivors: the Women's Health Accelerometry Collaboration. JNCI Cancer Spectr 2025 Apr 30;9(3):pkaf034. [doi: 10.1093/jncics/pkaf034] [Medline: 40139695]
40. Valle CG, Diamond MA, Heiling HM, et al. Physical activity maintenance among young adult cancer survivors in an mHealth intervention: twelve-month outcomes from the IMPACT randomized controlled trial. Cancer Med 2023 Aug;12(15):16502-16516. [doi: 10.1002/cam4.6238] [Medline: 37317660]
41. Valle CG, Diamond MA, Heiling HM, et al. Effect of an mHealth intervention on physical activity outcomes among young adult cancer survivors: the IMPACT randomized controlled trial. Cancer 2023 Feb 1;129(3):461-472. [doi: 10.1002/cncr.34556] [Medline: 36444676]
42. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283. [doi: 10.2196/20283] [Medline: 32990635]

XSL•FO
RenderX

43. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
44. Finne E, Glausch M, Exner AK, Sauzet O, Stölzel F, Seidel N. Behavior change techniques for increasing physical activity in cancer survivors: a systematic review and meta-analysis of randomized controlled trials. Cancer Manag Res 2018;10:5125-5143. [doi: 10.2147/CMAR.S170064] [Medline: 30464612]
45. Ester M, Eisele M, Wurz A, McDonough MH, McNeely M, Culos-Reed SN. Current evidence and directions for future research in ehealth physical activity interventions for adults affected by cancer: systematic review. JMIR Cancer 2021 Sep 20;7(3):e28852. [doi: 10.2196/28852] [Medline: 34542415]
46. Moon Z, Zuchowski M, Moss-Morris R, Hunter MS, Norton S, Hughes LD. Disparities in access to mobile devices and e-health literacy among breast cancer survivors. Support Care Cancer 2022 Jan;30(1):117-126. [doi: 10.1007/s00520-021-06407-2] [Medline: 34236506]
47. Fareed N, Swoboda CM, Jonnalagadda P, Huerta TR. Persistent digital divide in health-related internet use among cancer survivors: findings from the Health Information National Trends Survey, 2003-2018. J Cancer Surviv 2021 Feb;15(1):87-98. [doi: 10.1007/s11764-020-00913-8] [Medline: 32671557]

Abbreviations

BCT: behavior change technique
JARS-Quant: APA Journal Article Reporting Standards for quantitative research
LETSGO: Lifestyle and Empowerment Techniques in Survivorship of Gynecologic Oncology
mHealth: mobile health
PA: physical activity
RCT: randomized controlled trial
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by S Brini; submitted 25.Sep.2025; peer-reviewed by CA Miaskowski, N Lukkahatai; revised version received 01.Jul.2026; accepted 02.Jul.2026; published 13.Aug.2026.

Please cite as:
Fosstveit SH, Vistad I, Skorstad M, Berntsen S
Long-Term Engagement With a Physical Activity App Among Gynecologic Cancer Survivors in the LETSGO Trial: 12-Month Prospective, Multicenter, Quasi-Experimental Study
JMIR Mhealth Uhealth 2026;14:e84838
URL: https://mhealth.jmir.org/2026/1/e84838
doi:10.2196/84838

© Sindre Herskedal Fosstveit, Ingvild Vistad, Mette Skorstad, Sveinung Berntsen. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 13.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Predictors of Mobile Health App Acceptance Among Nurses in Kashan, Iran: Cross-Sectional Study

Narjes Mirabootalebi[1,2*], PhD; Felix Holl[3,4*], PhD; Walter Swoboda[3], MD; Zahra Tagharrobi[5], PhD; Fatemeh Rangraz Jeddi[2,6], PhD; Hossein Akbari[7], PhD

[1]Social Determinants in Health Promotion Research Center, Hormozgan Health Institute, Hormozgan University of Medical Sciences, Bandar Abbas, Iran

[2]Health Information Management Research Center, Kashan University of Medical Sciences, Kashan, Iran

[3]DigiHealth Institute, Neu-Ulm University of Applied Sciences, Neu-Ulm, Germany

[4]Leibniz Science Campus Digital Public Health, Bremen, Germany

[5]Trauma Nursing Research Center, Kashan University of Medical Sciences, 5th of Qotb-e Ravandi Blvd, Kashan, Iran

[6]Department of Health Information Management and Technology, Allied Medical Sciences Faculty, Kashan University of Medical Sciences, Kashan, Iran

[7]Social Determinants of Health (SDH) Research Center, Kashan University of Medical Sciences, Kashan, Iran

[*]these authors contributed equally

Corresponding Author:
Zahra Tagharrobi, PhD
Trauma Nursing Research Center, Kashan University of Medical Sciences, 5th of Qotb-e Ravandi Blvd, Kashan, Iran

Abstract

Background: The integration of mobile health apps (MHAs) into nursing practice is essential for improving efficiency, conserving resources, and enhancing patient care. To promote the acceptance of these apps in clinical care, it is important to assess current acceptance levels and identify influencing factors.

Objective: This study investigated the acceptance of MHAs and related factors among nurses in hospitals affiliated with Kashan University of Medical Sciences, Iran.

Methods: In this cross-sectional study, 250 nurses were selected via stratified random sampling in Kashan in 2022. Data were collected using the Nurses' Mobile Health Device Acceptance Scale (NMHDA-S) and the Probable Predictors Questionnaire. The data were analyzed using SPSS (version 16; IBM Corp), 1-way ANOVA, independent *t* test, Pearson correlation coefficients, and multiple linear regression.

Results: The findings showed that nurses had a mean acceptance score of 4.207 (SD 0.740) on a Likert scale from 1 to 7 (95% CI 4.115 - 4.299) regarding the use of MHAs. Multiple linear regression analysis indicated that 4 variables significantly predicted acceptance of MHAs (R^2=0.230, $F_{4,245}$=18.308, P<.001): interest in participating in relevant educational programs (β=0.392, P<.001), encouragement from health care professionals to use MHAs (β=0.133, P=.02), male participant (β=−0.116, P=.04), and duration of daily internet use (β=0.117, P=.04).

Conclusions: Although MHA acceptance among clinical nurses in Kashan is moderate, adoption can be enhanced through targeted training, facilitated internet access, and gender-specific incentive policies, particularly for female staff.

(JMIR Mhealth Uhealth 2026;14:e93888) doi:10.2196/93888

KEYWORDS

mobile apps; acceptance; nurses; attitude; Iran

Introduction

The increasing prevalence of smartphones and mobile health apps (MHAs) among health care professionals highlights the growing integration of health information technology into clinical practice [1]. Nurses are encouraged to incorporate various forms of health information technology into their clinical practice for several reasons [2]. One primary reason is the nature of the nursing profession, which requires unrestricted mobility within the clinical setting [3]. On the other hand, the global shortage of nurses is a major challenge [4]. A study conducted in 2020 revealed that most nurses (52.4%) had a negative attitude toward adopting and using MHAs [5]. In addition, a study conducted in Iran [6] reported that approximately 63.4% of nurses did not use MHAs on their smartphones.

The absence of MHAs in the clinical setting can lead to several challenges associated with the use of paper. These challenges include significant time consumption, frequent visits to nursing

XSL•FO
RenderX

stations for recording and retrieving information, and an increased risk of forgetting important details. Moreover, the use of paper increases the likelihood of errors in data recording and management, which may have serious consequences for patient safety. In this context, implementing MHAs at the patient's bedside is essential and represents an inevitable solution. These technologies have led to fundamental improvements in the quality of nursing care and are straightforward and easily accessible [7]. Therefore, health care managers across various levels need to adopt MHAs to increase the quality-of-care services and address the issues arising from traditional paper-based systems. The initial step in addressing this challenge involves assessing the current acceptance of MHAs among nurses and identifying the factors that influence this acceptance [8].

The concept of acceptance concerning technologies such as MHAs is framed primarily through behavioral models, with the expectation-confirmation model (ECM) serving as the foundation for this study. To capture the unique aspects of the nursing context thoroughly, our ECM framework intentionally incorporates established components from the technology acceptance model and the unified theory of acceptance and use of technology (UTAUT). These include constructs such as perceived ease of use and social influence. This integrated theoretical approach operationalizes acceptance by measuring a range of multidimensional factors, thus ensuring that the model is directly aligned with the clinical environment, including effectiveness, usage, performance, and anxiety related to new technology [9,10].

Numerous studies have explored the factors influencing the acceptance of MHAs. However, a notable limitation of these studies is the inconsistent prioritization of identified factors among health care professionals [8,11,12]. For example, Lwoga and Lwoga [13] and Chopdar et al [14] emphasized perceived security risk, whereas Lau et al [15] reported no significant correlation between this factor and the acceptance of MHAs. Sezgin et al [16] emphasized the characteristics of the user interface as an important factor influencing the use of MHAs. However, the results of Hsiao and Chen [17] revealed that this factor does not have a significant influence. In addition, there is a discrepancy in the literature regarding the influence of organizational characteristics on the use of MHAs. While Zhou et al [18] demonstrated a significant effect, Hsiao and Chen [19] noted that this factor was not significant.

Despite the expanding body of literature on MHAS acceptance, most studies have been conducted in developed countries with well-established health information infrastructures. In contrast, developing nations such as Iran face distinct infrastructural, sociocultural, and economic challenges during the digital transformation of their health care systems. Kashan, a prominent center for medical education and clinical care in central Iran, offers an ideal environment for investigating these transitional dynamics. The 5 hospitals affiliated with Kashan University of Medical Sciences include a wide range of clinical wards and diverse nursing demographics, making them an ideal setting for a comprehensive examination of digital health adoption in a resource-limited environment. Given inconsistencies in the existing literature on the impact of attitudes on performance and the lack of comprehensive studies in such contexts, this research was conducted to assess the acceptability of using MHAs and related factors among clinical nurses across all 5 hospitals in Kashan, Iran. The results of this study can serve as a foundation for developing educational interventions and policy decisions to advance the implementation of digital health in nursing practice.

Methods

Study Design and Setting

The present study is a descriptive cross-sectional study. In this study, the sample size was calculated on the basis of a pilot study conducted with 30 nurses, which produced a mean acceptance score of 4.055 (SD 0.722). With a confidence level of 95% and an accuracy estimate (d) defined as a conservative fraction of the SD ($d = 0.124 \times 0.722$), consistent with established methodological guidelines for achieving high precision in continuous data [20,21], Cochran formula was used to calculate the required sample size, resulting in a total of 250 participants (Equation 1).

The inclusion criteria included Iranian citizenship, a university degree in nursing, no known psychological disorders based on self-reports (such as mood disorders or anxiety disorders), at least 6 months of clinical experience, current employment in clinical roles in inpatient departments, and consent to participate in the study. Refusal to complete the data collection instruments was also considered as an exclusion criterion. A stratified random sample by department was conducted among eligible nurses in different clinical wards of hospitals affiliated with Kashan University of Medical Sciences. This sampling process was conducted over 2 months, from August to October 2022.

$$n=z2\times SD2d2$$

Equation 1 shows the Cochrane formula for sample size.

Data Collection Process

After approval from the Research Council and the Ethics Committee of Kashan University of Medical Sciences, the first author of the article approached the nursing offices of 5 hospitals (Shahid Beheshti, Naghavi, Matini, Kargarnejad, and Seyed Al Shohada Hospitals) with an introductory letter. After the study procedure was explained and permission from hospital management was secured, a comprehensive list was compiled detailing all clinical departments, including the names and total numbers of active nursing staff in each unit. To ensure proportional representation in line with the inclusion criteria and the target sample size, a proportionate number of eligible nurses from each department was then selected via simple random sampling. After consent was obtained from the head nurses, the first author of the article visited the departments at the beginning of each shift in accordance with the work schedules of the selected nurses. After providing consent, the selected nurses were instructed to complete the data collection instruments. The instruments were collected at the end of each shift. If the scales were not completed within the specified time frame, an agreement was reached with the respective nurse regarding the time of submission. In cases where the selected sample was unavailable or uncooperative, another nurse from

XSL•FO
RenderX

the same department was randomly selected. All data collection instruments were completed by the nurses themselves and were based on self-reports.

Data Collection Instruments

In the present study, data were collected via the probable predictors of acceptance of MHAs by nurses using the Nurses' Mobile Health Device Acceptance Scale (NMHDA-S) [8] (Multimedia Appendix 1).

The probable predictors of acceptance of mobile health programs, according to the nurses' questionnaire, are divided into 4 sections:

1. Personal and professional information, including age, gender, marital status, education level, department, work experience, frequent shift work, having young or adolescent children, and holding a managerial position
2. Knowledge of MHAs, assessed via a single self-reported item rated on a 7-point scale, specifically asking: "Please indicate your level of knowledge regarding MHAs on a scale of 1 to 7"
3. A history of familiarity with MHAs, including observations of health care professionals or acquaintances using MHAs, participation in relevant courses, specialized university training, encouragement to use MHAs by health care professionals, involvement in relevant research, mandatory use of specific MHAs in the workplace, instances of patient inquiries about MHAs, the presence of individuals with educational or professional backgrounds in information technology or related fields among first-degree relatives or close friends, and their willingness to engage with topics about new technologies
4. Use of MHAs, including professional and personal use, smartphone ownership, duration of smartphone usage, device operating system, types of devices used, duration of daily internet use, and interest in participating in relevant educational programs

This questionnaire was developed in Persian by the researchers of the present article on the basis of a comprehensive literature review, including relevant articles and questionnaires, as well as a survey of nursing and health information management professionals. The qualitative content validity was confirmed by 10 faculty members from Kashan University of Medical Sciences. The reliability of the questionnaire was evaluated via the test-retest method with a group of 20 nurses over a 1-week interval. The degree of agreement between the 2 measurements was determined by calculating the intraclass correlation coefficient for quantitative variables (such as knowledge of MHAs) and the κ statistic for categorical variables (such as familiarity factors with MHAs). Notably, all the responses regarding personal and professional variables remained consistent across both assessments. The intraclass correlation coefficients for the quantitative variables were found to exceed 0.95, whereas the κ statistics for the categorical variables surpassed 0.97.

The NMHDA-S was used to evaluate the acceptance of MHAs among nurses. This scale was originally developed in Persian on the basis of the constructs of the ECM and was psychometrically evaluated by Mirabootalebi et al [8] in 2024. The NMHDA-S consists of 33 items categorized into 5 subscales: usage and performance (10 items), social influence (7 items), perceived ease of use (5 items), effectiveness (7 items), and new technology anxiety (4 items). The respondents rated each item on a 7-point Likert-type scale. To determine the total acceptance score, the mean of all the items is calculated, resulting in a continuous score ranging from 1 to 7, where higher values indicate greater acceptance. Subscale scores are derived via the same mean-based procedure. The scale's construct validity has been confirmed, and it demonstrates high internal consistency, with Cronbach α and McDonald ω coefficients of 0.938 and 0.953, respectively. These indices were estimated to be above 0.88 for all the subscales [8].

Data Analysis

The data were analyzed using SPSS version 16. Acceptance scores were rated on a 7-point scale (1 to 7), and the 95% CI was calculated for the target population. Skewness and kurtosis indices were used to assess the data distribution, with a range of ±2 indicating a normal distribution. To prioritize the scale factors, a stepwise multiple linear regression analysis was performed to determine the contributions of various factors to acceptance, including usage and performance, social influence, perceived ease of use, effectiveness, and new technology anxiety.

The effects of the factors were analyzed in 2 phases to identify predictors. In the first phase, univariate analyses, including independent *t* tests, one-way ANOVA, and Pearson correlation coefficients, were conducted to assess the relationship between each factor and MHA acceptance. Variables with a significance level of <0.20 were subsequently entered into a stepwise multiple linear regression model. To ensure the validity of the regression results, key assumptions were rigorously evaluated. Multicollinearity was assessed using variance inflation factors (VIFs) and tolerance indices, while autocorrelation among the residuals was examined using the Durbin-Watson statistic. The normality and homoscedasticity of the residuals were evaluated using objective statistical measures—specifically, skewness and kurtosis values within ±2 were considered indicative of normal distributions. This analysis was further supported by visual inspection of residual histograms and normal probability (P-P) plots. For all analyses, a *P* value of <.05 was considered statistically significant.

Ethical Considerations

All necessary approvals were obtained from the Research Council (number 039, dated 2022) and the Ethics Committee of Kashan University of Medical Sciences (code 40138). The first author explained the study to each nurse individually in a quiet, private, and comfortable room within the department. The key points discussed included the purpose of the study, the expected actions required for participation, the estimated time needed to complete the instruments, measures to ensure the confidentiality of the data, the absence of negative consequences if they choose not to participate in the study, the individuals and departments overseeing the study, and their right to withdraw from participation at any time. After these

XSL•FO
RenderX

explanations, each nurse provided verbal consent and then signed a written informed consent form.

Results

Participant Characteristics

Among the 306 eligible nurses, 250 (81.7%) participated in the study, resulting in the analysis of 250 samples. The mean age of the participants was 35.624 (SD 7.582) years. Table 1 presents the personal and professional information, knowledge of MHAs, history of familiarity with MHAs, and usage of MHAs. The mean knowledge score of MHAs was calculated to be 2.976 (SD 1.210) on a 7-point scale (1 to 7).

XSL•FO
RenderX

Table . Characteristics of the participants, Kashan, 2022 (N=250).

Variables	Values
Personal and professional information	
Gender, n (%)	
Man	44 (17.6)
Woman	206 (82.4)
Marital status, n (%)	
Married	201 (80.4)
Single	47 (18.8)
Divorced	2 (0.8)
Education, n (%)	
Bachelor	217 (86.8)
Master	33 (13.2)
Department, n (%)	
Emergency	23 (9.2)
Internal	31 (12.4)
Surgery	50 (20)
Intensive unit	51 (20.4)
Pediatrics	7 (2.8)
Operating room	34 (13.6)
Other	54 (21.6)
Frequent shift work, n (%)	
Morning	114 (45.6)
Evening	29 (11.6)
Night	36 (14.4)
Rotating	71 (28.4)
Having young or adolescent children, n (%)	
Yes	102 (40.8)
No	148 (59.2)
Having managerial position, n (%)	
Yes	49 (19.6)
No	201 (80.4)
Age (y), mean (SD)	35.624 (7.582)
Work experience (y), mean (SD)	11.594 (6.722)
History of familiarity with MHAs[a]	
Observations of health care professionals or acquaintances using MHAs, n (%)	
Yes	91 (36.4)
No	159 (63.6)
Participation in relevant courses, n (%)	
Yes	67 (26.8)
No	183 (73.2)
Specialized university training, n (%)	
Yes	47 (18.8)
No	203 (81.2)

XSL•FO
RenderX

Variables	Values
Encouragement to use MHAs by health care professionals, n (%)	
Yes	57 (22.8)
No	193 (77.2)
Involvement in relevant research, n (%)	
Yes	27 (10.8)
No	223 (89.2)
Mandatory use of specific MHAs in the workplace, n (%)	
Yes	51 (20.4)
No	199 (79.6)
Instances of patient inquiries about MHAs, n (%)	
Yes	60 (24)
No	190 (76)
Presence of individuals with professional backgrounds in information technology or related fields among first-degree relatives or close friends, n (%)	
Yes	69 (27.6)
No	181 (72.4)
Willingness to engage with topics about new technologies [1-5], mean (SD)	3.184 (0.985)
Duration of smartphone usage (y), (n=231), mean (SD)	9.446 (3.340)
Interest in participating in relevant educational programs [1-5], mean (SD)	3.036 (0.937)
Use of MHAs, n (%)	
Professional use	
Yes	65 (26)
No	185 (74)
Personal use	
Yes	73 (29.2)
No	177 (70.8)
Smartphone ownership	
Yes	231 (92.4)
No	19 (7.6)
Device operating system (n=231)	
Android	207 (82.8)
Apple iOS	24 (9.6)
Type of devices used	
iPhone	19 (7.6)
Mobile phone	224 (89.6)
Mobile phone and tablet	7 (2.8)
Duration of daily internet use	
Less than 1 hour	49 (19.6)
1 to 2 hours	91 (36.4)
2 to 4 hours	62 (24.8)
More than 4 hours	48 (19.2)

XSL•FO
RenderX

Variables	Values
NMHDA-S[b] score (total and subscales), mean (SD)	
Usage and performance (on a scale of 1 to 7)	4.401 (0.952)
Social influence (on a scale of 1 to 7)	4.211 (1.219)
Perceived ease of use (on a scale of 1 to 7)	4.351 (1.384)
Effectiveness (on a scale of 1 to 7)	4.471 (1.185)
New technology anxiety (on a scale of 1 to 7)	3.602 (1.246)
Total score of NMHDA-S (on a scale of 1 to 7)	4.207 (0.740)

[a]MHA: mobile health app.

[b]NMHDA-S: Nurses' Mobile Health Device Acceptance Scale.

Acceptance Levels

The mean acceptance score for using MHAs in nursing was 4.207 (SD 0.740; on a scale of 1 to 7), indicating a moderate level of acceptance among clinical nurses. This rate was estimated to be 4.207 (95% CI 4.115-4.299) within the target population. All quantitative variables in the current study had a normal distribution based on the criteria included in the methodology (skewness and kurtosis within ±2). Stepwise multiple linear regression was used to prioritize the subscales of the acceptance scale. The results revealed that the contributions of each subscale, including usage and performance, social influence, perceived ease of use, effectiveness, and new technology anxiety, to explaining the variations in acceptance ratings were 1.9%, 5.5%, 12.6%, 76.9%, and 3.1%, respectively. Consequently, effectiveness proved to be the most important subscale.

Predictors of Acceptance

A significant univariate association was found between acceptance scores and several factors, including gender (P=.03), education level (P=.001), observations of health care professionals or acquaintances using MHAs (P<.001), participation in relevant courses (P=.002), encouragement to use MHAs by health care professionals (P<.001), instances of patient inquiries about MHAs (P=.002), the presence of individuals with educational or professional backgrounds in information technology or related fields among first-degree relatives or close friends (P<.001), professional (P=.001), and personal use (P=.03), duration of daily internet use (P=.009), interest in participating in relevant educational programs (P<.001), willingness to engage with topics about new technologies (P<.001), and knowledge of MHAs (P=.002; Tables 2 and 3).

XSL•FO
RenderX

Table . Acceptance scores for mobile health app (MHA) use in care according to categorical predictor variables among Iranian nurses, Kashan, 2022 (N=250).

Variable	Acceptance score, mean (SD)	*t* test (*df*) or *F* test (*df*)	*P* value
Gender		2.235[a] (248)	.03
Woman	4.159 (0.723)		
Man	4.432 (0.783)		
Marital status		4.437[b] (2, 2.863)	.13
Married	4.181 (6.990)		
Single	4.288 (0.899)		
Divorced	4.904 (0.305)		
Education		−3.368[a] (248)	.001
Bachelor	4.147 (0.718)		
Master	4.603 (0.769)		
Department		0.795[c] (6, 243)	.58
Emergency	4.047 (0.619)		
Internal	4.243 (0.861)		
Surgery	4.330 (0.702)		
Intensive unit	4.217 (0.675)		
Pediatrics	4.232 (0.421)		
Operating room	4.023 (0.719)		
Other	4.244 (0.846)		
Frequent shift work		1.786[c] (3, 246)	.15
Morning	4.103 (0.760)		
Evening	4.172 (0.678)		
Night	4.290 (0.766)		
Rotating	4.347 (0.705)		
Having young or adolescent children		−0.735[a] (248)	.44
Yes	4.166 (0.697)		
No	4.236 (0.769)		
Having managerial position		1.044[a] (248)	.30
Yes	4.306 (0.716)		
No	4.183 (0.745)		
Observations of health care professionals or acquaintances using MHAs		3.860[a] (248)	<.001
Yes	4.440 (0.741)		
No	4.074 (0.708)		
Participation in relevant courses		3.206[a] (248)	.002
Yes	4.451 (0.768)		
No	4.118 (0.711)		
Specialized university training		0.944[a] (248)	.35
Yes	4.311 (0.866)		
No	4.183 (0.708)		

XSL•FO
RenderX

Variable	Acceptance score, mean (SD)	*t* test (*df*) or *F* test (*df*)	*P* value
Encouragement to use MHAs by health care professionals		3.700[a] (248)	<.001
Yes	4.518 (0.770)		
No	4.115 (0.707)		
Involvement in relevant research		0.822[a] (248)	.42
Yes	4.350 (0.981)		
No	4.190 (0.706)		
Mandatory use of specific MHAs in the workplace		0.610[a] (248)	.54
Yes	4.264 (0.742)		
No	4.193 (0.741)		
Instances of patient inquiries about MHAs		3.195[a] (248)	.002
Yes	4.468 (0.752)		
No	4.125 (0.718)		
Presence of individuals with professional backgrounds in IT or related fields among first-degree relatives or close friends		3.606[a] (248)	<.001
Yes	4.474 (0.700)		
No	4.105 (0.731)		
Professional use		3.414[a] (248)	.001
Yes	4.710 (0.767)		
No	4.114 (0.709)		
Personal use		2.229[a] (248)	.03
Yes	4.368 (0.808)		
No	4.141 (0.701)		
Smartphone ownership		1.783[a] (248)	.08
Yes	4.231 (0.752)		
No	3.918 (0.504)		
Device operating system		0.803[a] (248)	.42
Android	4.245 (0.757)		
iOS	4.114 (0.710)		
Type of devices used		1.661[b] (2, 13.303)	.23
iPhone	4.367 (0.423)		
Mobile	4.181 (0.728)		
Mobile and tablet	4.626 (1.479)		
Duration of daily internet use (h)[d,e]		3.975[c] (3, 246)	.009
<1 (A)	3.948 (0.676)		
1 - 2 (B)	4.275 (0.708)		
2 - 4 (C)	4.142 (0.760)		

XSL•FO
RenderX

Variable	Acceptance score, mean (SD)	*t* test (*df*) or *F* test (*df*)	*P* value
>4 (D)	4.427 (0.769)		

[a]Independent *t* test.

[b]ANOVA (Welch statistic).

[c]ANOVA (*F* statistic).

[d]Pairwise comparison: A and B: *P*=.06; A and C: *P*=.65; A and D: *P*=.008; B and C: *P*=.84; B and D: *P*=.80; D and C: *P*=.23.

[e]Post hoc (Gabriel).

Table . Correlation of acceptance scores for using mobile health apps (MHAs) in care with quantitative variables among Iranian nurses, Kashan, 2022 (N=250).

Variable	*r* values[a]	*P* value
Age (y)	–0.01	.87
Work experience (y)	0.052	.41
Duration of smartphone usage (y; n=231)	0.128	.04
Interest in participating in relevant educational programs [1-5]	0.419	<.001
Willingness to engage with topics about new technologies [1-5]	0.276	<.001
Knowledge of MHAs [1-7]	0.196	.002

[a]Pearson coefficient.

A stepwise multiple linear regression was performed to further examine the relationships between the independent variables and acceptance scores. The VIFs for all the variables were <5, and their tolerances exceeded 0.10, indicating an absence of multicollinearity. The Durbin-Watson statistic was 1.943, which falls within the acceptable range of 1.5 to 2.5. Before conducting the final regression analysis, the underlying assumptions were thoroughly verified. The Durbin-Watson statistic was within the acceptable range, confirming the absence of autocorrelation. The VIF values indicated no multicollinearity. For the regression residuals, both skewness and kurtosis were within the acceptable range of ±2, suggesting normality. Additionally, examination of residual histograms and normal P-P plots further confirmed that the assumptions of normality and homoscedasticity were fully satisfied. Consequently, the assumptions of multiple linear regression were satisfied, confirming the validity of the model. The results show that the simultaneous inclusion of 4 variables in the model is significant ($F_{4,245}$=18.308, *P*<.001) and explains 23% of the observed variance in acceptance scores. These variables include interest in participating in relevant educational programs, encouragement to use MHAs by health care professionals, gender, and duration of daily internet use. The variable with the greatest influence is interest in participating in relevant educational programs, which accounts for 17.6% of the observed changes. Interest in participating in relevant educational programs, encouragement by health care professionals to use MHAs, and duration of daily internet use positively influence acceptance. In addition, being a man is associated with greater acceptance (Table 4).

XSL•FO
RenderX

Table . Results of multiple linear regression to determine the predictive factors influencing the acceptance of using mobile health apps (MHAs) among Iranian nurses.

Model[a]	R^2	B[b] (95% CI)	SE	β[c]	*t* value	*P* value	Collinearity statistics	
							Tolerance	VIF[d]
Constant	—[e]	3.886 (3.202 to 4.570)	0.347	—	11.184	<.001	—	—
Interest in participating in relevant educational programs (scoring 1 - 5)	0.176	0.310 (0.221 to 0.398)	0.045	0.392	6.881	<.001	.966	1.035
Encouragement to use MHAs by health care professionals[f]	0.024	0.235 (0.035 to 0.435)	0.102	0.133	2.310	.02	.943	1.060
Gender[g]	0.017	−0.226 (−0.442 to −0.010)	0.110	−0.116	−2.059	.04	.981	1.019
Duration of daily internet use (h)	0.013	0.086 (0.003 to 0.168)	0.042	0.117	2.049	.04	.961	1.041

[a]R^2=0.230, $F_{4,245}$=18.308, P<.001.

[b]Unstandardized coefficients.

[c]Standardized coefficients.

[d]VIF: variance inflation factors.

[e]Not applicable.

[f]Reference group=no.

[g]Reference group=man.

When controlling for other variables, the average acceptance score for MHAs among female nurses was 0.226 points lower than that of male nurses. In contrast, nurses who received encouragement from health care professionals to use MHAs experienced an average increase of 0.235 points in their acceptance scores. For each 1-unit increase in interest in participating in related educational programs, the acceptance score increased by an average of 0.310 points. Additionally, each extra hour of daily internet usage corresponds to an average increase of 0.086 points in the acceptance score. In this model, the unstandardized coefficients (B) represent the absolute change in the dependent variable per unit change in the independent variable, measured in the original units of the independent variable. The standardized coefficients (β) are dimensionless values that normalize the variables by their SDs, allowing them to be assessed on a common scale. While B quantifies the absolute magnitude of change, β is used to interpret the relative importance and compare the predictive power of the independent variables, regardless of their measurement units. According to the β coefficients in Table 4, interest in participating in relevant educational programs has the highest relative explanatory power in the model.

Discussion

Acceptance of MHAs Among Nurses

The present study aimed to investigate the acceptance of MHAs and associated factors among clinical nurses in Iran. The principal findings indicate moderate acceptance of MHAs among the surveyed nurses, with a mean score of 4.207 (SD 0.740) out of 7.

A study conducted by Shanmugapriya et al [22] in India reported that 71.6% of nursing students strongly agreed, 27.8% agreed, and only 0.6% expressed no opinion regarding their acceptance of MHAs. In addition, Kürtüncü et al [7] reported a mean total score of 124.69 (SD 15.98; ranging from 0 - 128) for Turkish nurses on the MHAS acceptance model, indicating a high level of acceptance in this group. The researchers explained that the reason for the high level of acceptance among Turkish nurses was related to their high use of smartphones [7]. In Iran, the acceptance of MHAs among nurses was found to be moderate, possibly due to the lack of necessary infrastructure, absence of established guidelines, and insufficient knowledge [6]. In simpler terms, nurses are still not sufficiently equipped for digital transformation in health care and face several challenges, including limited digital skills, restrictions on smartphone usage, and financial constraints. By examining these specific behavioral factors in a resource-limited context such as Iran, this study

XSL•FO
RenderX

goes beyond incremental findings to provide a comprehensive understanding of the unique challenges of digital health adoption in developing health care systems. It is recommended that incentives be offered to nurses to help offset the costs of using their personal smartphones for clinical tasks [23].

The findings of the current study suggest that the dimensions of "effectiveness" and "perceived ease of use" contribute significantly to the variations in acceptance scores. In contrast, "usage and performance" was responsible for the least variation in these scores. These results are consistent with those of previous studies. Hsiao and Chen [19] highlighted the impact of maturity and perceived usefulness on technology acceptance. Recent evidence on the adoption of health care technology indicates that perceived usefulness and perceived ease of use play significant roles in technology acceptance and the intention to use these innovations [24]. Additionally, recent studies focusing on nurses and health care professionals emphasize that perceived usefulness is an essential factor influencing both technology acceptance and the intention to use it [24,25]. In a review paper, Nezamdoust et al [26] identified 8 key themes influencing nurses' use of MHAs: ease of use, usefulness, security and privacy, feasibility and performance, design and user interface, effectiveness, infrastructure, adaptability, and social norms. The findings of this and other studies indicate that perceived usefulness, perceived ease of use, and maturity are important determinants of clinical nurses' acceptance of MHAs. Owing to living in the information and technology era and recognizing the benefits of using MHAs, nurses are increasingly inclined to use similar systems. Therefore, it is clear that perceived ease of use and usefulness play important roles in the acceptance of MHAs and ultimately encourage their increased or continued use.

In the present study, "usage and performance" made the smallest contribution to explaining the difference in caregiver acceptance of MHAs. Consistent with these findings, Hsiao and Chen [19] concluded that individual habits do not significantly influence the acceptance or performance of MHAs. In addition, Sezgin et al [16] reported no significant relationships between outcome acceptance, habits, self-efficacy, expected performance, creativity, and community influence on behavioral intentions. Chen et al [27] also concluded that security risk has a negative effect on the intention to use. However, some studies support the rejection of the hypothesis of the present study. Recent findings indicate that privacy and performance risks negatively impact trust and the intention to adopt digital health technologies, suggesting that perceived risk reduces acceptance [24]. Studies by Lwoga and Lwoga [13] and also Chopdar et al [14] have emphasized the influence of perceived security risk. The factors influencing usage and performance include personal habits, security risk, and social approval. The underutilization of MHAs can be attributed to several challenges, such as the absence of robust policies, lack of practical and user-friendly apps, and nurses' unawareness of the potential benefits of MHAs. These issues likely play an important role in the resulting usage scores.

On the basis of the results of the present study, "interest in participating in relevant educational programs," "encouragement to use MHAs by health care professionals," "gender," and "duration of internet use per day" were identified as factors influencing nurses' acceptance of using MHAs in their nursing practice. The greatest contribution was "interest in participating in relevant educational programs," which was significantly positively correlated with the acceptance rate.

Kaya et al [28] reported a significant association between in-service training and changes in nurses' attitudes toward themselves, computers, and information technology devices. This technology requires training and attention to organizational and environmental factors to increase technology anxiety among nurses. In addition, it is important to create a supportive organizational culture that promotes technical skills and provides adequate time and resources for learning [29,30]. In a 2019 study [31], 69.5% of nurses could not use technology effectively and correctly. One of the main reasons for this is nurses' lack of computer skills. Addressing this issue through continuous education for nurses can help bridge the generational and educational divide, ultimately alleviating their anxiety about using new technologies. Lifelong education for nurses plays a critical role in improving their acceptance of technology, especially in the face of constant changes in patient care, advancements in knowledge, and evolving practices and policies. In addition, supervisor and peer support can encourage nurses to embrace information technology by actively promoting MHA, effectively integrating it into clinical practice, and recognizing the resulting performance improvements.

The present study revealed that encouraging the use of MHAs among health care professionals has a positive and significant effect on acceptance rates. Recent evidence indicates that social factors significantly influence users' intention to adopt technology, whereas attitudes remain a key contributor in these contexts [32]. Recent studies have shown that social influence significantly predicts the intention to use emerging technologies [33]. The findings of the study by Alasmari and Zhang [34] are consistent with those of the present study. The acceptance of MHAs by nurses is significantly influenced by various factors, including use by nursing colleagues, other health care professionals, and care managers. For example, colleagues who integrate smartphones into their clinical practice can be role models for nurses and encourage them to use similar technologies. In addition, the positive experiences and satisfaction of colleagues and managers with MHAs can help reduce nurses' technology-related anxiety and reservations and thus increase their overall acceptance of these tools. In particular, sharing positive experiences with experts can be a strong motivator for nurses. When selecting MHAs, nurses usually consult more experienced colleagues, often giving considerable weight to their recommendations. Consequently, a positive view from peers can significantly increase nurses' acceptance of technologies. When colleagues endorse specific technologies, this promotes a sense of approval among health care professionals. Conversely, peer resistance to MHAs can have a negative effect on acceptance and usage.

In the statistical model outlined in the present study, the "duration of daily internet usage" is found to have a positive effect on the acceptance score. Shanmugapriya et al [22] reported that there is a statistically significant correlation between the acceptance of MHAs and the time spent on nursing

XSL•FO
RenderX

studies (internet-based learning). A 2020 study [35] identified several significant predictors of acceptance of MHAs among patients with cancer, including age, education level, and access to the internet. In addition, the duration of daily internet use increased among nurses using smartphones, which in turn impacted their acceptance of these apps. Smartphones provide easy access to the internet and allow nurses to access important online resources, such as treatment and medication guidelines, directly from their workplace. This capability enhances the ease of access and utilization of digital information in the clinical setting. In addition, smartphones facilitate communication with colleagues, physicians, and patients via various platforms, such as email, messaging applications, and telemedicine services.

This study revealed that men had a greater level of acceptance. A 2019 study [34] revealed that men tended to use mobile learning technologies more strongly than women did. In addition, a study by Nunes et al [36] confirmed the significant influence of men on the acceptance of MHAs. In this context, Al-Azawei [37] pointed to gender difference as a moderator in the relationship between the model constructs. However, in this study, this moderating effect was found to be minimal with respect to the relationship between e-learning self-efficacy and learning management system acceptance. Self-efficacy had a stronger effect on the intention to use learning management system in men than in women, suggesting that men have relatively higher levels of acceptance [37]. However, a 2024 study [38] examining students' intentions to use online learning through educational technology applications based on the UTAUT model revealed a significant influence of women on acceptance. Recent evidence suggests that both gender and age can affect the relationship between self-efficacy and technology adoption, with self-efficacy remaining a crucial factor in users' intentions to adopt technology [39]. Furthermore, the findings indicate that specific computer effectiveness has a more pronounced effect on perceived ease of use than does general remaining efficacy. Among cofactors, general computer self-efficacy emerged as a significant predictor of perceived usefulness for both men and women. Additionally, while perceptions of general computer self-efficacy were lower, it still had a notable influence on the intention to use technology. Conversely, specific computer self-efficacy demonstrated a stronger indirect effect on behavioral intentions among women, although both genders reported similar levels of self-efficacy [40]. The results can serve as a reference for hospitals and clinical training organizations. When integrating nursing technologies into nurses' clinical practice, it is important to consider user characteristics and relevant moderators when specific groups are targeted through MHAs. Studies suggest that men are more at ease with and acclimatized to technology than women are. In addition, societal expectations around gender roles can influence how people perceive and interact with technology. In a country with cultural characteristics similar to those of Iran, it seems that women's greater anxiety about new technology and men's greater belief in their self-efficacy contribute to these results. Further research is needed, as there is a lack of robust evidence in this area for different societies. Analyzing the underlying causes in Iranian society highlights the need for qualitative approaches.

Importantly, several factors may contribute to the discrepancies between the results of the current study and those of previous related research. These factors include the timing and location of the studies, the type of statistical tests, the presence or absence of established national guidelines, differences in the relevant target groups, and the specific characteristics of those populations.

The substantial sample size and the ability to conduct multiple analyses are notable strengths of the current study. Several factors may play important roles for health care professionals, including task mobility, individual performance, technological maturity [19], self-efficacy, responsiveness [41], organizational support, environmental and organizational influences, adaptability [17], and acceptance of MHA; however, these factors were not explored in the current study. In this context, the final multiple linear regression model (Table 4) explained only 23% of the variance in the acceptance score, indicating that 77% of the variance is explained by other variables not included in the model. Consequently, the interpretation of these results should be approached with more caution, as the identified predictors provide only a partial explanation of MHAS acceptance. Furthermore, the timing of sampling during the COVID-19 crisis disrupted the usual structure of hospital departments, changed the composition of human resources, and significantly increased the occupational stress among nurses in this setting [42]. As a result, the unwillingness of a considerable number of nurses to participate in the study probably affected the generalizability of the findings, which is recognized as a limitation of this study. It is recommended that a study be conducted under normal conditions, free from the constraints imposed by pandemic crises, to address the limitations of the current study. In addition, conducting a qualitative study with a grounded theory approach would be valuable for an in-depth investigation of the factors influencing the acceptance of MHAs by nurses in different societies, including Iran.

This study has several limitations. First, the data were collected concurrently with the COVID-19 crisis, during which the nursing staff experienced a high workload and occupational stress [42], potentially affecting participation and responses. A substantial proportion of eligible nurses were unwilling to cooperate, which may have introduced response bias. Furthermore, the participants' knowledge of MHAs was assessed using a single-item question, which may not have captured the full multidimensionality of this construct. It is strongly recommended that future studies evaluate this variable via more appropriate, comprehensive, and highly precise measurement tools.

Conclusions

The present study revealed moderate acceptance of MHAs among nurses in hospitals affiliated with Kashan University of Medical Sciences. The key factors that positively and significantly contribute to the acceptance rate include "interest in participating in relevant educational programs," "encouragement from health care professionals to use MHAs," and "duration of daily internet use." In addition, the acceptance rate is significantly higher among men than women. Facilitating interest in participation in relevant training courses, alongside

XSL•FO
RenderX

the provision of adequate internet access and the implementation of incentive policies targeting all nursing professionals, especially women, may increase the acceptance of MHAs in the delivery of care. Therefore, health policy makers and nursing managers should recognize the importance of this issue when designing continuing education programs for nurses, formulating policies related to incentive structures, and establishing supportive infrastructure in hospitals. By implementing effective strategies, they can improve the acceptance of MHAs among nurses in the clinical setting.

Acknowledgments

The authors would like to thank all the nurses who participated in the study and the Vice Chancellor for Research and Technology of Kashan College of Medical Sciences. The authors acknowledge the generous scientific support of the Clinical Research Development Unit of Matini/Kargarnejad Hospitals, Kashan University of Medical Sciences, Kashan, Iran. Generative AI tool was not used in any stage of preparing the article.

Funding

The present article is based on the findings of the first author's doctoral dissertation, which was approved by the Research Council of Kashan University of Medical Sciences and Health Services (number 039, dated 2022) and conducted with financial support from the Vice-Chancellor for Research and Technology of Kashan University of Medical Sciences.

Data Availability

The datasets analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: ZT, HA, NM, FH, WS, FRJ
Data curation: NM
Formal analysis: ZT, HA, NM
Investigation: NM
Methodology: ZT, HA
Writing – original draft: NM
Writing – review & editing: NM, ZT, HA, FH, WS, FRJ

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data collection instruments (in Persian).
[DOCX File, 29 KB - mhealth_v14i1e93888_app1.docx]

References

1. de Jong A, Donelle L, Kerr M. Nurses' use of personal smartphone technology in the workplace: scoping review. JMIR Mhealth Uhealth 2020 Nov 26;8(11):e18774. [doi: 10.2196/18774] [Medline: 33242012]
2. Bautista JR, Rosenthal S, Lin TTC, Theng YL. Psychometric evaluation of the Smartphone for Clinical Work Scale to measure nurses' use of smartphones for work purposes. J Am Med Inform Assoc 2018 Aug 1;25(8):1018-1025. [doi: 10.1093/jamia/ocy044] [Medline: 29788375]
3. Ehrler F, Wipfli R, Teodoro D, Sarrey E, Walesa M, Lovis C. Challenges in the implementation of a mobile application in clinical practice: case study in the context of an application that manages the daily interventions of nurses. JMIR Mhealth Uhealth 2013 Jun 12;1(1):e7. [doi: 10.2196/mhealth.2344] [Medline: 25100680]
4. Hsiao SJ, Li YC, Chen YL, Ko HC. Critical factors for the adoption of mobile nursing information systems in Taiwan: the nursing department administrators' perspective. J Med Syst 2009 Oct;33(5):369-377. [doi: 10.1007/s10916-008-9199-8] [Medline: 19827263]
5. Agyei DD, Adzobu P. Factors influencing professional nurses' acceptance and use of mobile medical apps in Ghana. J Inf Technol Manag 2020;12(1):27-42. [doi: 10.22059/jitm.2020.286522.2389]
6. Rahimi R, Fathian A, Khoundabi B, Sadooghiasl A. The use of mobile health applications for professional purposes by nurses during the COVID-19 pandemic. Hayat 2022;28(1):14-27 [FREE Full text]
7. Kürtüncü M, Kurt A, Arslan N. An examination of nurses' acceptance of mobile health applications. Clin Exp Health Sci 2021;11(4):865-870. [doi: 10.33808/clinexphealthsci.905574]

XSL•FO
RenderX

8. Mirabootalebi N, Meidani Z, Akbari H, et al. Design and psychometric evaluation of Nurses' Mobile Health Device Acceptance Scale (NMHDA-scale): application of the Expectation-Confirmation Theory. JMIR Hum Factors 2024 Sep 17;11:e55324. [doi: 10.2196/55324] [Medline: 39288375]
9. Mirabootalebi N, Holl F, Meidani Z, et al. Determinants of nurses' continuance intention to use mobile health apps in clinical nursing practice: structural equation modeling to extend the expectation-confirmation model. JMIR Nurs 2025 Oct 31;8(1):e68048. [doi: 10.2196/68048] [Medline: 41172281]
10. Cao K, Wang P, Zhang C, Zhao J. Multidimensional factors influencing continuance usage intention of university library self-service systems: an empirical analysis based on an extended TAM-UTAUT. PLoS One 2026;21(3):e0346014. [doi: 10.1371/journal.pone.0346014] [Medline: 41911293]
11. Fauzi A, Wandira R, Sepri D, Hafid A. Exploring students' acceptance of Google Classroom during the COVID-19 pandemic by using the technology acceptance model in West Sumatera universities. Electron J e Learn 2021;19(4):233-240. [doi: 10.34190/ejel.19.4.2348]
12. Wandira R, Fauzi A, Nurahim F. Analysis of factors influencing behavioral intention to use cloud-based academic information system using extended technology acceptance model (TAM) and expectation-confirmation model (ECM). J Inf Syst Eng Bus Intell 2024;10(2):179-190. [doi: 10.20473/jisebi.10.2.179-190]
13. Lwoga ET, Lwoga NB. User acceptance of mobile payment: the effects of user‐centric security, system characteristics and gender. Electron J Inf Syst Dev Ctries 2017 Jul;81(1):1-24. [doi: 10.1002/j.1681-4835.2017.tb00595.x]
14. Chopdar P, Korfiatis N, Sivakumar VJ, Lytras MD. Mobile shopping apps adoption and perceived risks: a cross-country perspective utilizing the unified theory of acceptance and use of technology. Comput Human Behav 2018 Sep;86:109-128. [doi: 10.1016/j.chb.2018.04.017]
15. Lau MM, Lam AYC, Cheung R, Leung TF. Understanding determinants of customer behavioral intention in using mobile payment at convenience stores. In: Chen KC, Ma Y, editors. IC4E '19: Proceedings of the 10th International Conference on E-Education, E-Business, E-Management and E-Learning: Association for Computing Machinery; 2019:357-362. [doi: 10.1145/3306500.3306549]
16. Sezgin E, Özkan-Yildirim S, Yildirim S. Understanding the perception towards using mHealth applications in practice. Inf Dev 2018 Mar;34(2):182-200. [doi: 10.1177/0266666916684180]
17. Hsiao JL, Chen RF. Critical factors influencing physicians' intention to use computerized clinical practice guidelines: an integrative model of activity theory and the technology acceptance model. BMC Med Inform Decis Mak 2016 Jan 16;16(1):3. [doi: 10.1186/s12911-016-0241-3] [Medline: 26772169]
18. Zhou T, Lu Y, Wang B. Integrating TTF and UTAUT to explain mobile banking user adoption. Comput Human Behav 2010 Jul;26(4):760-767. [doi: 10.1016/j.chb.2010.01.013]
19. Hsiao JL, Chen RF. Understanding determinants of health care professionals' perspectives on mobile health continuance and performance. JMIR Med Inform 2019 Mar 18;7(1):e12350. [doi: 10.2196/12350] [Medline: 30882353]
20. Cohen J. Statistical Power Analysis for the Behavioral Sciences: Routledge; 2013. [doi: 10.4324/9780203771587]
21. Verma J, Verma P. Determining Sample Size and Power in Research Studies: Springer; 2020. [doi: 10.1007/978-981-15-5204-5]
22. Shanmugapriya K, Seethalakshmi A, Zayabalaradjane Z, Rani NRV. Mobile technology acceptance among undergraduate nursing students instructed by blended learning at selected educational institutions in South India. J Educ Health Promot 2023;12(1):45. [doi: 10.4103/jehp.jehp_488_22] [Medline: 37113422]
23. Abane AM, Mariwah S, Owusu SA, Kasim A, Robson E, Hampshire K. Mobile phone use and the welfare of community health nurses in Ghana: an analysis of unintended costs. World Dev Perspect 2021 Sep;23:100317. [doi: 10.1016/j.wdp.2021.100317] [Medline: 34568642]
24. Hussain A, Zhiqiang M, Li M, et al. The mediating effects of perceived usefulness and perceived ease of use on nurses' intentions to adopt advanced technology. BMC Nurs 2025 Jan 9;24(1):33. [doi: 10.1186/s12912-024-02648-8] [Medline: 39789568]
25. Lim AY, Park JY. Factors influencing the acceptance of information communication technology among hospital nurses in a tertiary hospital. Korean J Health Promot 2024 Jun;24(2):67-73. [doi: 10.15384/kjhp.2024.00059]
26. Nezamdoust S, Abdekhoda M, Ranjbaran F, Azami-Aghdash S. Adopting mobile health applications by nurses: a scoping review. J Res Nurs 2022 Aug;27(5):480-491. [doi: 10.1177/17449871221077080] [Medline: 36131693]
27. Chen SL, Chen JH, Chang SC. Understanding the antecedents of individuals intention of using cloud services. J Econ Manag 2017;13(2):139-166 [FREE Full text]
28. Kaya H, Işık B, Bodur G. The relationship between in-service training nurses' attitudes toward computer in health care and computer anxiety in Turkey. J Hum Sci 2014;11(2):948. [doi: 10.14687/ijhs.v11i2.2712]
29. Rajamani S, Hultman G, Bakker C, Melton GB. The role of organizational culture in health information technology implementations: a scoping review. Learn Health Syst 2021 Jul;6(3):e10299. [doi: 10.1002/lrh2.10299] [Medline: 35860317]
30. Mitchell JI, Gagné M, Beaudry A, Dyer L. The role of perceived organizational support, distributive justice and motivation in reactions to new information technology. Comput Human Behav 2012 Mar;28(2):729-738. [doi: 10.1016/j.chb.2011.11.021]
31. Orhan I, Serin EK. Use of health technologies by nurses and their thoughts on technology. Int J Caring Sci 2019;12(1):416-422 [FREE Full text]

XSL•FO
RenderX

32. Shahzad MF, Xu S, Baheer R. Assessing the factors influencing the intention to use information and communication technology implementation and acceptance in China's education sector. Humanit Soc Sci Commun 2024;11(1):1-15. [doi: 10.1057/s41599-024-02777-0]
33. Korchak A, Al Murshidi G, Getman A, et al. The role of social influence in generative artificial intelligence ChatGPT adoption intentions among undergraduate and graduate students. Innov Educ Teach Int 2025 Sep 3;62(5):1559-1573. [doi: 10.1080/14703297.2025.2496942]
34. Alasmari T, Zhang K. Mobile learning technology acceptance in Saudi Arabian higher education: an extended framework and a mixed-method study. Educ Inf Technol 2019 May;24(3):2127-2144. [doi: 10.1007/s10639-019-09865-8]
35. Potdar R, Thomas A, DiMeglio M, et al. Access to internet, smartphone usage, and acceptability of mobile health technology among cancer patients. Support Care Cancer 2020 Nov;28(11):5455-5461. [doi: 10.1007/s00520-020-05393-1] [Medline: 32166381]
36. Nunes A, Limpo T, Castro SL. Acceptance of mobile health applications: examining key determinants and moderators. Front Psychol 2019;10:2791. [doi: 10.3389/fpsyg.2019.02791] [Medline: 31920836]
37. Al-Azawei A. The moderating effect of gender differences on learning management system acceptance: a multi-group analysis. Ital J Educ Technol 2019;27(3):257-278. [doi: 10.17471/2499-4324/1088]
38. Gupta S, Bajaj R, Das B, Kumar V, Gupta D. Gender as a moderator to students' behavioral intention toward online learning. Int J Educ Reform 2024:10567879241278969. [doi: 10.1177/10567879241278969]
39. Sakdavong JC, Peyrègne M. How gender and age mediate the effect of self-efficacy on digital learning performance. Commun Comput Inf Sci 2025 [FREE Full text]
40. Bao Y, Xiong T, Hu Z, Kibelloh M. Exploring gender differences on general and specific computer self-efficacy in mobile learning adoption. J Educ Comput Res 2013 Jul;49(1):111-132. [doi: 10.2190/EC.49.1.e]
41. Zhang X, Han X, Dang Y, Meng F, Guo X, Lin J. User acceptance of mobile health services from users' perspectives: the role of self-efficacy and response-efficacy in technology acceptance. Inform Health Soc Care 2017 Mar;42(2):194-206. [doi: 10.1080/17538157.2016.1200053] [Medline: 27564428]
42. Ghaderi Z, Tagharrobi Z, Sooki Z, Sharifi K. Predictive factors of occupational stress among nurses during the COVID-19 pandemic: a cross-sectional study in Kashan, Iran. BMC Nurs 2024 May 9;23(1):313. [doi: 10.1186/s12912-024-01967-0] [Medline: 38724986]

Abbreviations

ECM: expectation-confirmation model
MHA: mobile health app
NMHDA-S: Nurses' Mobile Health Device Acceptance Scale
VIF: variance inflation factor

Edited by NM Ali; submitted 21.Feb.2026; peer-reviewed by L González, S Anwar; revised version received 12.Jun.2026; accepted 13.Jun.2026; published 31.Jul.2026.

Please cite as:
Mirabootalebi N, Holl F, Swoboda W, Tagharrobi Z, Rangraz Jeddi F, Akbari H
Predictors of Mobile Health App Acceptance Among Nurses in Kashan, Iran: Cross-Sectional Study
JMIR Mhealth Uhealth 2026;14:e93888
URL: https://mhealth.jmir.org/2026/1/e93888
doi:10.2196/93888

© Narjes Mirabootalebi, Felix Holl, Walter Swoboda, Zahra Tagharrobi, Fatemeh Rangraz Jeddi, Hossein Akbari. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 31.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluation of the Impact of a Mobile App (LoAD Calc) on the Calculation of Maximum Safe Doses of Local Anesthetics: Randomized Controlled Trial

Alexandra Stefani[1], MD; Pietro Elias Fubini[1], MD; Tal Sarah Beckmann[1], MD; Caroline Flora Samer[2,3], MD; Georges Louis Savoldelli[1,3], MEd, MD; Mélanie Suppan[1,3], MSc, MD

[1]Division of Anaesthesiology, Department of Acute Care Medicine, Geneva University Hospitals, Rue Gabrielle-Perret-Gentil 4, Geneva, Switzerland

[2]Division of Clinical Pharmacology and Toxicology, Department of Acute Care Medicine, Geneva University Hospitals, Geneva, Switzerland

[3]Department of Anaesthesiology, Pharmacology, Intensive Care and Emergency Medicine, Faculty of Medicine, University of Geneva, Geneva, Switzerland

Corresponding Author:
Mélanie Suppan, MSc, MD
Division of Anaesthesiology, Department of Acute Care Medicine, Geneva University Hospitals, Rue Gabrielle-Perret-Gentil 4, Geneva, Switzerland

Abstract

Background: Local anesthetics (LAs) are widely used in clinical practice, but their administration carries the risk of LA systemic toxicity, a potentially fatal complication. Calculating maximum safe doses is complex and error-prone when considering multiple patient-specific factors. Most existing tools and mobile apps have not been scientifically validated and may propose unsafe doses. The Local Anesthetics Dose Calculator (LoAD Calc) mobile app was developed at the Geneva University Hospitals to calculate maximum safe doses by integrating comprehensive patient parameters.

Objective: The primary objective was to evaluate the overdose rate using the LoAD Calc app compared with traditional calculation methods. Secondary objectives were calculation time, user confidence, app usability, and interpractitioner variability.

Methods: This single-center unblinded randomized controlled trial recruited 50 anesthesiologists to use either LoAD Calc or their usual calculation methods. Assessment was conducted in-person using a web-based platform. Participants completed 10 clinical vignettes featuring various patient characteristics, comorbidities, and medication interactions requiring maximum safe dose calculations for lidocaine, levobupivacaine, and ropivacaine. The primary outcome was the overdose rate, defined as exceeding the maximum safe doses calculated using a comprehensive set of rules incorporating all patient-specific factors. Secondary outcomes included overdose rates using 2 simplified weight-based methods, underdose rates, calculation time, and self-reported measures of confidence and app usability. Overdose rates were compared using chi-square tests with 95% CIs; calculation time and confidence were assessed using Student *t* test.

Results: All 50 participants completed the study. Using the full set of calculation rules, the overdose rate was significantly lower when LoAD Calc was used (30/250, 12.0% vs 198/250, 79.2%; relative risk [RR] 0.15, 95% CI 0.11 - 0.21; *P*<.001). This was consistent across all 3 LAs: ropivacaine (5/100, 5% vs 64/100, 64%; *P*<.001), levobupivacaine (19/100, 19% vs 90/100, 90%; *P*<.001), and lidocaine (6/50, 12% vs 44/50, 88%; *P*<.001). When considering simplified weight-based methods, overdose rates remained significantly lower when LoAD Calc was used (actual weight: 8/250, 3.2% vs 95/250, 38.0%; RR 0.08, 95% CI 0.04 - 0.17; *P*<.001; ideal body weight: 10/250, 4.0% vs 151/250, 60.4%; RR 0.07, 95% CI 0.04 - 0.12; *P*<.001). Underdose rates were similar (25/250, 10.0% vs 27/250, 10.8%; *P*=.77). Calculation time was also similar (mean 111, SD 56 s vs mean 115, SD 89 s; *P*=.53). LoAD Calc users reported higher confidence (mean 8.5, SD 0.9 vs mean 6.3, SD 1.7; *P*<.001), and the app achieved a median System Usability Scale score of 98 (IQR 95-100).

Conclusions: LA overdoses were significantly less likely when the LoAD Calc app was used in simulated clinical vignettes, without increasing calculation time or underdose rates. Given its high usability score and the improvement in practitioner confidence, clinical implementation could be considered and may contribute to improving patient safety and dose calculation standardization.

International Registered Report Identifier (IRRID): RR2-10.2196/53679

(JMIR Mhealth Uhealth 2026;14:e89236) doi:10.2196/89236

KEYWORDS

local anesthetics; dose calculation; toxicity; mobile apps; safety

XSL•FO
RenderX

Introduction

Local anesthetics (LAs) are widely used by clinicians across various settings, including operating rooms and office-based procedures [1,2]. In anesthesia practice, local and regional anesthesia have become integral components of perioperative management for elective surgical procedures [3]. While generally safe and effective, LA overdose can expose patients to serious complications, particularly local anesthetic systemic toxicity (LAST), which manifests as neurological and cardiac toxicity and can be potentially fatal [4,5].

To prevent LAST, the American Society of Regional Anesthesia and Pain Medicine (ASRA) recommends several safety measures, including ultrasound guidance, incremental and fractional injection with intermittent aspiration, the use of intravascular markers, and the administration of the lowest effective LA dose [6].

In clinical practice, dose calculations are typically performed according to empirical guidelines, with many clinicians relying on mental calculations or basic calculators. However, the mental calculation of maximum safe LA doses is challenging and error-prone, particularly when multiple variables must be considered simultaneously, including patient weight, age, comorbidities, LA type and concentration, injection site vascularity, and potential drug interactions [7,8]. When LA mixtures are used, complexity grows substantially, thereby significantly increasing the risk of calculation errors. To address this challenge, tools such as nomograms and mobile apps have been developed to facilitate dose calculations [9]. However, most have not been scientifically validated for clinical implementation, and preliminary assessments suggest that they allow the calculation of doses that could exceed safe limits and potentially lead to LAST [10].

A mobile health (mHealth) app, Local Anesthetics Dose Calculator (LoAD Calc), was developed at the Geneva University Hospital to assist anesthesiologists in calculating safe maximum LA doses [10]. This app integrates various patient parameters and enables calculations for LA mixtures. The hypothesis was that automating LA dose calculation would reduce overdose rates without increasing the risk of underdosing or calculation time. The primary objective of this study was to evaluate whether the LoAD Calc app reduces overdose rates compared to traditional calculation methods. Secondary objectives included assessing the impact on calculation time, user confidence, app usability, and whether the app standardizes LA dosing in clinical practice.

Methods

Study Design and Participants

This single-center, parallel-group randomized controlled trial was conducted at the Geneva University Hospitals. The study protocol was previously published [11]. This paper was designed according to the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online TeleHealth) checklist (Checklist 1) [12]. Relevant elements from the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) were included due to the use of online questionnaires [13]. The trial was conducted in an unblinded design, as it was not possible to conceal the intervention type from participants.

All resident physicians and registrars in the anesthesiology department were eligible for participation. All participants were employed at the Geneva University Hospitals, a tertiary academic center, and were therefore subject to the same institutional workload and mandatory continuing medical education requirements. Regional anesthesia is routinely performed as part of standard perioperative care at this institution. The sole exclusion criterion was current or prior use of the LoAD Calc app, determined through screening questions.

Ethical Considerations

The regional ethics committee (Commission Cantonale d'Éthique de la Recherche [CCER], Geneva, Switzerland) issued a declaration of no objection (CCER 2022 - 01577), as the study did not fall within the scope of Swiss federal law on human research [14]. This exemption was appropriate as the study involved only health care professionals in a simulated setting, with no patient data collected. All participants received a written information sheet and provided explicit consent before participating. Consent was obtained through an online platform, where participants clicked an "I agree to participate" button before proceeding. They were informed that participation was entirely voluntary and that they could withdraw at any time without consequence. Unique anonymous identifiers were used throughout; no personal identifying information was linked to study responses. The investigator (MS) who analyzed the data had no means of linking the anonymous identifiers to individual participants, and the final dataset used for analysis was fully anonymized. Participants received no financial compensation or other material incentives. The only benefit offered was the advancement of scientific knowledge.

LoAD Calc App Description

LoAD Calc is a multiplatform mobile app (iOS and Android) developed at the Geneva University Hospitals, as previously described in detail [10]. Briefly, the user selects the LAs and their concentrations, enters the patient's weight and height, and selects sex and the presence of relevant dose-modifying factors, including age over 70 years, comorbidities (renal insufficiency defined as glomerular filtration rate <50 mL/min, hepatic insufficiency defined as prothrombin time <50%, severe heart failure defined as left ventricular ejection fraction ≤30%, and pregnancy), and concomitant medications known to decrease LA metabolism (major CYP1A2 and CYP3A inhibitors). If one dose-modifying factor is present, the maximum dose is reduced by 20%; if 2 or more are present, it is reduced by 30%. The results are displayed as maximum safe volumes in milliliters, which is the clinically relevant unit for anesthesiologists. The app works offline and was developed using the Dart programming language with the Flutter open-source development kit (Google LLC). Screenshots of the app interface are provided in Multimedia Appendix 1. Version 1.2.1 of LoAD Calc was evaluated in this trial, and no updates or changes were made to the app's functionality or content during the study period. The complete calculation rules and algorithms are

XSL•FO
RenderX

detailed in this paper and the referenced development paper [10]. The source code is available upon request from the corresponding author.

Clinical Vignettes

Ten clinical vignettes (Multimedia Appendix 2) were developed by one of the authors (PEF), featuring realistic scenarios requiring LA dose calculations for 3 commonly used agents: lidocaine, levobupivacaine, and ropivacaine. The vignettes included various patient characteristics, comorbidities, and medication interactions requiring dose adjustments. These vignettes have already been described in a previous paper [15]. For each vignette, 3 authors (PEF, TSB, and MS) were required to determine the maximum dose of LA the simulated patient should receive according to the rules used to develop the LoAD Calc app, without using the app. Any disagreement prompted a review of the vignette. Final vignette approval was only possible if a consensus was reached.

Sample Size and Randomization

The sample size calculation was performed using Stata (version 17.0; StataCorp LLC). A minimum clinically meaningful difference of 10% in overdose rates was assumed based on clinical judgment, with an SD of 10% and a power of 90%, yielding a required sample size of 46 participants (23 per group). This was rounded to 50 to account for potential attrition.

Participants were randomized into one of 2 groups. The control group used their usual methods for calculating maximum safe LA doses, while the LoAD Calc group used the LoAD Calc mobile app preinstalled on a standardized dedicated smartphone. Randomization was stratified based on participants' positions (resident vs registrar) using Stata's balanced randomization mechanism by MS. Unique identifiers consisting of 12 random alphanumeric characters were generated and prerandomized to the respective groups. These identifiers served as login credentials for the web-based platform (described later) and automatically assigned participants to their respective groups (resident vs registrar and LoAD Calc vs traditional method).

Recruitment

Participants were recruited one at a time from the anesthesiology department of the Geneva University Hospitals between February 2024 and February 2025 and were temporarily relieved from their clinical duties. Participants were enrolled by MS, TSB, AS, and PEF. Potential participants first received an information sheet summarizing the study details and data protection measures. The information sheet did not identify which calculation method was the intervention of interest, maintaining a neutral presentation to minimize participant bias regarding the comparison. If they consented to participate, they were accompanied by an investigator to a dedicated room equipped with a computer. Participants were asked to put their personal phones in airplane mode and to give their professional phones to a colleague to avoid interruptions during the approximately 1-hour session. They were then asked to randomly choose a sealed, opaque envelope containing the credentials necessary to log in.

Online Platform

A dedicated web-based platform was developed by one of the authors (MS) using the Joomla! 4.3 content management system (Open Source Matters). The platform was hosted on a Swiss server (Kreativmedia GmbH) and secured using RSFirewall 3 (RSJoomla) and AdminTools 7 (Akeeba Ltd). Prior to implementation, the platform design and questionnaire interface were reviewed and tested by all authors to ensure usability and technical functionality. The study home page provided general information about the research and enabled participants to login. An initial questionnaire, managed using Shondalai Community Surveys 6 (Bulasikku Technologies Pvt Ltd), facilitated participant consent collection, exclusion criterion verification, and demographic data gathering. Then, an introductory screen displayed information regarding the clinical vignettes participants were about to encounter and specified the calculation method they were to use (LoAD Calc app for the experimental group vs the participant's preferred method for the control group). At this stage, those allocated to the LoAD Calc group were given a smartphone preinstalled with the LoAD Calc app. Participants received no prior training or instruction on the app and used it for the first time during the study session. The clinical vignettes were presented one by one in randomized order using Shondalai Community Quiz 7. A timer automatically started once each vignette page was loaded and stopped upon validation of the entered dose. After completing the vignettes, participants allocated to the LoAD Calc group completed the validated French version of the System Usability Scale (SUS) [16,17]. Participants allocated to the control group were asked which methods they used to calculate LA doses. Finally, both groups answered a question based on a 10-point Likert scale to assess their confidence in the method they used to carry out the maximum safe LA dose calculations, ranging from 1 (absolutely not confident) to 10 (perfectly confident). All participants were invited to provide open-ended comments or feedback about their experience with the calculation method or app if they wished to do so.

All data were stored in an encrypted MySQL-compatible database (MariaDB 10, MariaDB Foundation).

Overdose and Underdose Definitions

When using LAs, anesthesiologists decide to administer a specific volume (mL) rather than a particular quantity (mg). Thus, although toxicity is related to the quantity of LA used, it is clinically more relevant to determine the maximum volume (mL) of LA that should be used for a particular case.

To determine the maximum safe dose of LA for each vignette, quantities (mg) were computed using 3 different calculation methods before being converted to milliliters according to the LA's concentration. The comprehensive method utilized the full set of calculation rules incorporated in the LoAD Calc app as previously described [10], including patient-specific factors such as age, comorbidities, and drug interactions. Two simplified methods were also employed: one based on actual weight (AW) and the other on ideal body weight (IBW) using the usual formula:

XSL•FO
RenderX

Maximum dose (mg)=Weight (AW∨IBW) (kg)×Dose limit for chosen LA (mg/kg)

This multimethod approach reflects the complexity of LA dosing in clinical practice, where various calculation methods coexist. While the comprehensive, most conservative method served as the primary evaluation criterion due to its rigorous safety assessment, the simplified methods were used to reflect common clinical practices, where weight-based calculations are more commonly employed without taking other factors into account [18].

If the volume obtained by converting the maximum safe quantities of LA into milliliters was an integer, it was considered the maximum safe volume without any further change. If the result was a decimal value, it was first rounded down to the nearest integer to be consistent with the rules used to develop the LoAD Calc app. However, to avoid being too conservative, a value of 1 mL was then added, and the final value (either in milliliters or its equivalent in milligrams) was considered the maximum safe dose for the purpose of this study. Any dose higher than this value was considered an overdose. All maximum safe doses are reported in Table 1.

An underdose was defined as the maximum acceptable volume minus 20% (or its corresponding LA quantity in milligrams), rounded down to the nearest integer. This was an empirical choice since anesthetic underdose can only be determined clinically.

Table . Patient characteristics and maximum safe local anesthetic doses according to actual weight (AW), ideal body weight (IBW), and the full set of calculation rules.

Vignette (unit)	Weight (kg)	Height (cm)	Local anesthetic assessed	Mixture	Relevant characteristics	Maximum safe dose		
						Full rules	AW	IBW
1 (mg)	55.0	174	Ropivacaine	No	None	165	165	210
2 (mL)	51.5	159	Levobupivacaine	Yes	>70 years	10	14	14
3 (mg)	60.0	170	Lidocaine	Yes	Pregnancy	50	80	90
4 (mL)	70.0	170	Levobupivacaine	No	>70 years, renal failure	19	28	27
5 (mg)	120.0	180	Levobupivacaine	No	CYP1A2 inhibitor	115	240	145
6 (mL)	80.0	175	Ropivacaine	Yes	Liver cirrhosis	19	40	29
7 (mg)	60.0	175	Levobupivacaine	No	>70 years, heart failure	86.25	120	142.5
8 (mL)	75.0	170	Lidocaine	Yes	None	8	12	8
9 (mg)	90.0	160	Ropivacaine	No	>70 years	127.5	270	157.5
10 (mL)	80.0	157	Ropivacaine	No	None	30	48	30

Outcome Measures

The primary outcome was the overall overdose rate computed using the full calculation rules. The secondary outcomes were the overall overdose rate according to AW and IBW, the overdose rate according to each LA studied, the overall underdose rate, and the variability of results between participants.

The time taken to complete these calculations, the app's usability (according to the SUS), and the physicians' confidence (on a score ranging from 0 to 10) in the method they used to determine the maximum safe LA doses were also analyzed.

Statistical Analysis

Data were extracted from the encrypted MySQL-compatible database (MariaDB) into CSV files and imported into Stata 19 BE (StataCorp LLC) for data curation and statistical analysis. The primary statistical analysis was conducted by MS, who was blinded to group assignment.

Descriptive statistics were used to detail participants' characteristics. Normality was checked first graphically and then, in case of doubt, using the Shapiro-Wilk test. Values were reported accordingly.

The chi-square test was used to compare overdose and underdose rates; risk ratios with 95% CIs were also reported. The time required to complete the vignettes and participant confidence in the methods they used to determine LA doses were assessed using Student *t* test. Box plots were used to graphically display variability.

A post hoc sensitivity analysis was performed to assess the potential influence of outliers on the primary outcome. Participants with at least 1 dose value exceeding 3 SDs above their group's mean for that vignette were excluded.

Missing values were reported as such. No imputation method was used. Double-sided *P* values less than .05 were considered significant.

Results

Fifty anesthesiologists (30 residents and 20 registrars) consented to participate. The flow of participants through the trial is

XSL•FO
RenderX

presented in Figure 1. Only 14% (7/50) had already heard about the LoAD Calc app. None of them had ever installed or used the app before participating in the study. Thus, all 50 anesthesiologists were included. Their characteristics are detailed in Table 2.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of participant enrollment, randomization, allocation, follow-up, and analysis. LoAD Calc: Local Anesthetics Dose Calculator.

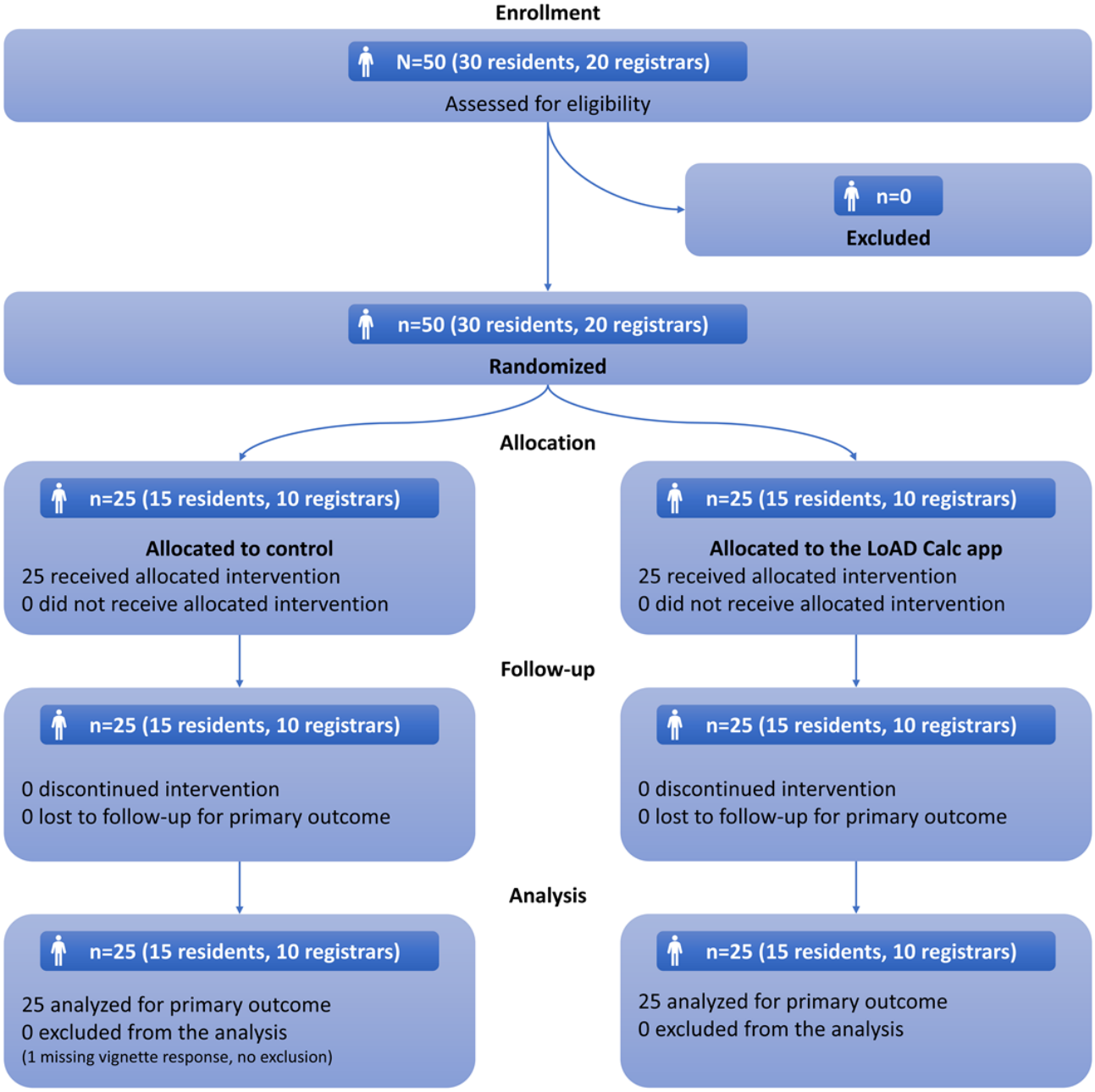


Table . Participant characteristics.

Characteristics	Overall (N=50)	Control group (n=25)	LoAD Calc[a] group (n=25)
Residents, n (%)	30 (60)	15 (60)	15 (60)
Gender (woman), n (%)	23 (46)	10 (40)	13 (52)
Age (y), median (IQR)	34 (32-37)	33 (32-35)	35 (33-38)
Years since diploma, median (IQR)	8 (6-10)	7 (6-10)	9 (6-10)
Years of anesthesiologic practice, median (IQR)	5 (2-7)	5 (3-6)	5 (2-8)

[a]LoAD Calc: Local Anesthetics Dose Calculator.

XSL•FO
RenderX

One participant, who belonged to the control group, did not answer one of the 10 clinical vignettes. All other participants answered all questions. The proportion of missing data was thus 0.2% (1/500 total responses).

When considering the full calculation rules, the overdose rate was significantly higher in the control group than in the LoAD Calc group (198/250, 79.2% vs 30/250, 12.0%; relative risk [RR] 0.15, 95% CI 0.11 - 0.21; *P*<.001). This difference was consistent for all 3 LAs (Table 3).

A post hoc sensitivity analysis, excluding participants with at least 1 dose value exceeding 3 SDs above their group's mean for that vignette (n=4 in the control group, n=6 in the LoAD Calc group), yielded consistent results for the primary outcome (control: 165/210, 78.6% vs LoAD Calc: 15/190, 7.9%; RR 0.10, 95% CI 0.06 - 0.16; *P*<.001).

Table . Overdose rates according to the full set of calculation rules.

Local anesthetic	Control, n (%)	LoAD Calc[a], n (%)	Relative risk (95% CI)	*P* value
Ropivacaine (n=100)	64 (64)	5 (5)	0.08 (0.03 - 0.19)	<.001
Levobupivacaine (n=100)	90 (90)	19 (19)	0.21 (0.14 - 0.32)	<.001
Lidocaine (n=50)	44 (88)	6 (12)	0.14 (0.06 - 0.29)	<.001
Overall (N=250)	198 (79.2)	30 (12)	0.15 (0.11 - 0.21)	<.001

[a]LoAD Calc: Local Anesthetics Dose Calculator.

The underdose rate was similar in both groups (27/250, 10.8% in the control group vs 25/250, 10.0% in the LoAD Calc group; RR 0.93, 95% CI 0.55 - 1.55; *P*=.77).

When AW was the only parameter taken into account, the overdose rate was 38.0% (95/250) in the control group vs 3.2% (8/250) in the LoAD Calc group (RR 0.08, 95% CI 0.04 - 0.17; *P*<.001). A similar difference was observed when considering IBW, with an overdose rate of 60.4% (151/250) in the control group vs 4.0% (10/250) in the LoAD Calc group (RR 0.07, 95% CI 0.04 - 0.12; *P*<.001).

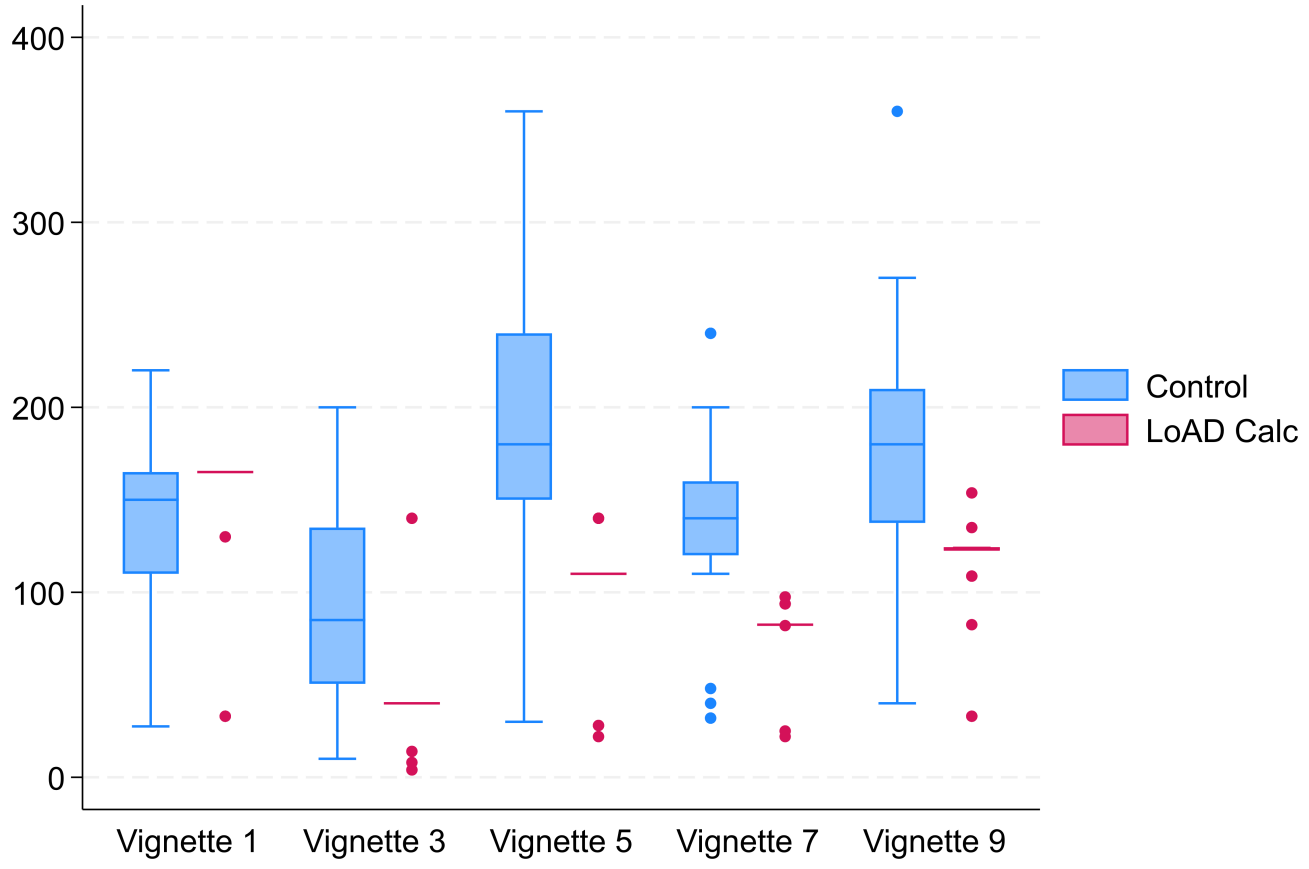
There was no difference in the mean time required to complete vignettes (mean 115, SD 89 s in the control group vs mean 111, SD 56 s in the LoAD Calc group; *P*=.53).

Participants felt significantly more confident when using the LoAD Calc app than when using other methods (mean 8.5, SD 0.9 vs mean 6.3, SD 1.7; *P*<.001).

The median SUS score (rated by the 25 participants who used the LoAD Calc app) was 98 (IQR 95-100).

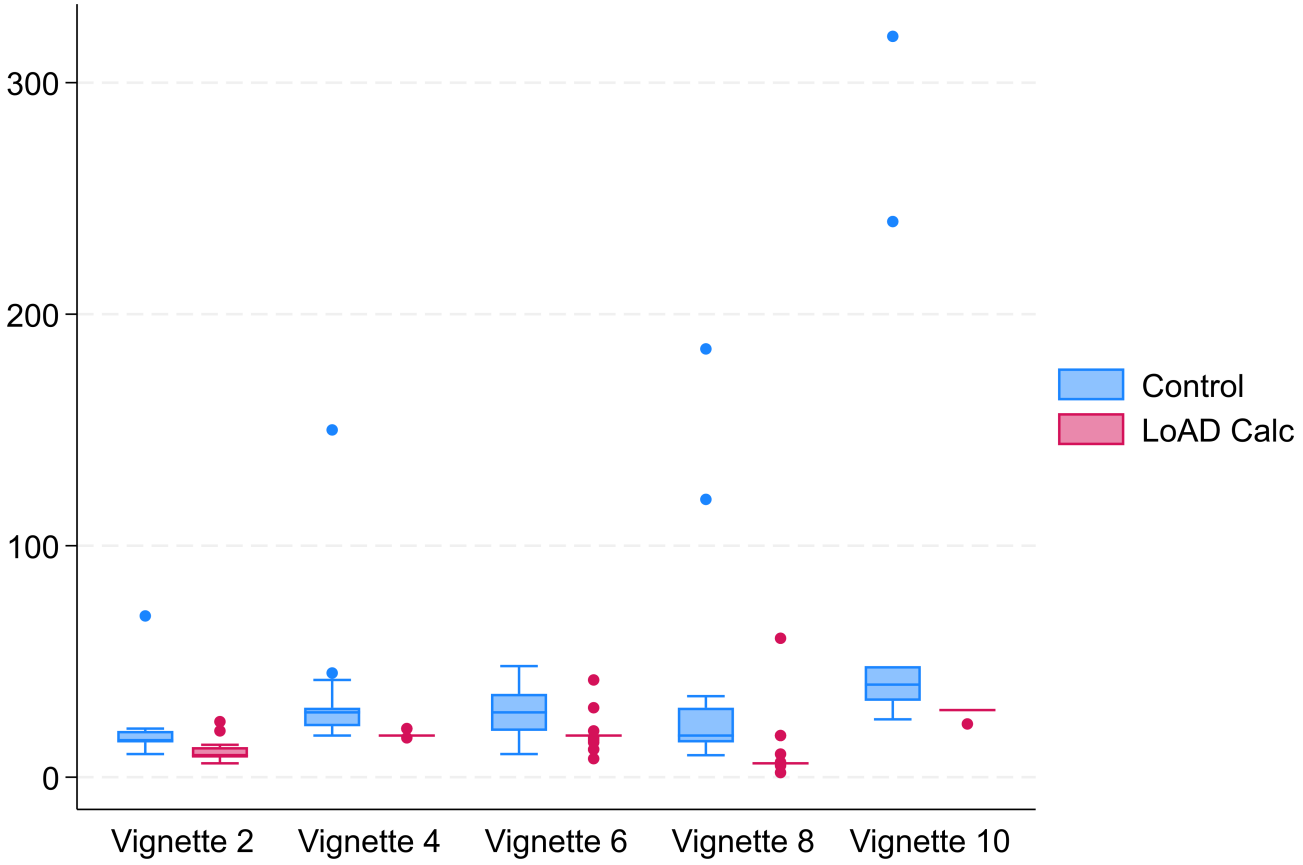
Figure 2 (quantities) and Figure 3 (volumes) show that variability was more limited when the LoAD Calc app was used.

Figure 2. Box plot of the quantities (in mg) recorded by study participants. LoAD Calc: Local Anesthetics Dose Calculator.



XSL•FO
RenderX

Figure 3. Box plot of the volumes (in mL) recorded by study participants. LoAD Calc: Local Anesthetics Dose Calculator.



Discussion

Principal Results

Using the LoAD Calc app significantly reduced overdose rates compared to traditional methods. This difference was consistent across all 3 LAs studied and remained significant even when using simplified calculation methods based solely on AW or IBW. These results demonstrate that the app's benefit is not an artifact of the chosen reference standard but extends to more commonly used calculation approaches, confirming the initial hypothesis that this mHealth app supports the safer calculation of maximum LA doses in simulated scenarios without increasing the risk of underdosing.

The reduction in overdose rates observed using LoAD Calc could contribute to reducing the risk of LAST, even though potentially unsafe doses do not systematically result in toxicity [19]. By standardizing dose calculations and systematically integrating patient-specific risk factors, the app could constitute an additional safety barrier in locoregional anesthesia practice. In addition, this mHealth app could also reduce interpractitioner variability and contribute to improving the overall quality and standardization of care [20].

Several tools have been developed to assist practitioners in calculating LA doses. The Williams nomogram and the WiMP formula (weight×maximum dose/percentage concentration) [9,21], while useful, do not allow for the integration of all patient-specific variables necessary for optimal calculation. Similarly, existing mobile apps such as MDCalc's Anesthetic Dosing Calculator [22], the Podiatry Institute's LA Toxic Dose Calculator [23], and SafeLocal by Johns Hopkins Digital calculate doses based on actual body weight without adjustment [24], do not account for comorbidities or drug interactions, and have not been validated for clinical practice. Preliminary assessments of these apps suggest that most permit the calculation of doses exceeding safe limits and could potentially lead to LAST [10].

Regarding weight calculation recommendations, there is currently significant heterogeneity in the literature. Different weight calculation methods have been proposed, including IBW calculated using formulas such as Devine [9], adjusted body weight [25], or lean body weight as recommended by the 3rd ASRA Practice Advisory [6], which can be calculated using the Boer formula [26]. These variations highlight the need for clarification and standardization of recommendations. The rules integrated into LoAD Calc, although very restrictive and prioritizing safety, may require adaptation based on the evolution of international consensus to promote wider adoption.

In contrast to these existing tools, the LoAD Calc app distinguishes itself by its ability to simultaneously integrate multiple patient-specific parameters: IBW, cardiac, renal or hepatic insufficiency, pregnancy, advanced age, and potential drug interactions. Each parameter is clearly defined in the app, enabling standardized use. Moreover, unlike nomograms and other apps that do not address this complexity, LoAD Calc allows the calculation of maximum safe doses for a mixture of 2 LAs, a particularly useful functionality in clinical practice but complex to perform mentally [27].

XSL•FO
RenderX

The results also demonstrate excellent app usability, with a median SUS score of 98 (IQR 95-100), reflecting very high user satisfaction. This high usability likely stems from the iterative user-centered design approach employed during app development [28], which prioritized clinicians' needs and workflow integration. App use did not increase the time required to complete calculations, a crucial element for clinical acceptability [29]. Furthermore, participants reported significantly higher confidence in their calculations when using the app, suggesting that the tool may enhance clinicians' situational awareness and help avoid calculation errors during locoregional anesthesia procedures.

The mHealth apps have become an integral part of modern clinical practice and are already well adopted by anesthesiologists [30]. These tools enable rapid access to information, particularly at the patient's bedside, and offer valuable support for improving daily medical practice and decision-making [31]. By standardizing complex calculations and reducing cognitive load, such apps may also decrease practitioner stress [32], which could ultimately improve both the quality of care and clinician well-being.

Several perspectives for future development, research, and implementation can be envisioned. The LoAD Calc app is currently based on 3 LAs, namely lidocaine, levobupivacaine, and ropivacaine, and could be expanded to include additional agents. Future versions could also allow mixing of more than 2 LAs and incorporate repeated doses with time intervals, common scenarios in obstetric anesthesia or when combining spinal anesthesia with peripheral nerve blocks. Integration with electronic medical records could facilitate the automatic retrieval of patient data. While currently targeting anesthesiologists, the tool could benefit other specialties using LAs daily, such as surgeons, ophthalmologists, or dentists, though adaptation to their specific practices would be necessary. Real-world observational studies are necessary to evaluate the app's impact on clinical outcomes, particularly LAST incidence, as well as user satisfaction and sense of safety. Long-term evaluation should assess adoption patterns, adherence factors, and the standardization of dose calculations in practice. Multicenter studies would establish the generalizability of results across different contexts, while cost-effectiveness analysis would guide larger-scale implementation decisions. For wider adoption beyond research settings, the app would require recognition as a medical device with appropriate regulatory approval.

Limitations

Several limitations must be acknowledged. First, the trial was conducted exclusively within the Anesthesiology Department of Geneva University Hospitals over a limited period. Since all participating physicians belonged to the same institution, their responses may have been influenced by shared departmental practices or organizational culture, which may limit the generalizability of the results. Second, participants were placed in a simulated environment using clinical vignettes. Therefore, real-world clinical conditions might influence how clinicians calculate maximum LA doses differently than observed in this controlled setting. Time pressure, frequent interruptions, and the complexity of real clinical situations could modify observed performances. Third, some observed outliers with unexpectedly high values likely resulted from participants confusing milligram and milliliter units when entering their answers, despite clear instructions specifying which unit to use for each vignette. A post hoc sensitivity analysis excluding these participants confirmed that our primary conclusions were not affected by these extreme values. Another limitation lies in the choice to include only anesthesiologists, either in training or already certified, many of whom use LAs daily. The perception and use of the LoAD Calc app could differ in other medical specialties with different practices and needs regarding local anesthesia. Extending app use will therefore require understanding the specific needs of other clinical disciplines and adapting the tool accordingly.

Conclusion

This study demonstrates that LoAD Calc, an mHealth app designed to calculate maximum safe LA doses including mixtures, significantly reduces calculation-related overdose rates in simulated scenarios compared to traditional calculation methods, without increasing calculation time and with excellent usability. Clinicians also reported greater confidence when using LoAD Calc compared to traditional methods. These findings suggest that the app may support safer and more standardized LA dose calculations in clinical practice. However, as this was a vignette-based simulation study, further research in real-world clinical settings is needed to determine whether these improvements translate into enhanced patient safety and reduced risk of LAST.

Acknowledgments

No external acknowledgments are applicable. No AI-generated content was included in the manuscript.

Funding

The authors declared no financial support was received for this work.

Data Availability

The data that support the findings of this study are openly available in YARETA [33].

XSL•FO
RenderX

Authors' Contributions

Conceptualization: AS, PEF, TSB, MS
Data curation: AS, MS
Formal analysis: AS, MS
Investigation: AS, PEF, MS
Methodology: AS, PEF, TSB, MS
Project administration: MS
Software: MS
Supervision: MS, CFS, GLS
Validation: MS, CFS, GLS
Visualization: AS, MS
Writing – original draft: AS, MS
Writing – review and editing: AS, PEF, TSB, CFS, GLS, MS

Conflicts of Interest

TSB, CFS, GLS, and MS were involved in the development of the LoAD Calc app, which served as the intervention evaluated in this study. This app was developed solely for research purposes, is not commercialized, and will not be commercialized in the future.

Multimedia Appendix 1
Screenshots of the LoAD Calc app.
[PNG File, 721 KB - mhealth_v14i1e89236_app1.png]

Multimedia Appendix 2
Clinical vignettes.
[DOCX File, 31 KB - mhealth_v14i1e89236_app2.docx]

Checklist 1
CONSORT-eHEALTH checklist (V 1.6).
[PDF File, 1224 KB - mhealth_v14i1e89236_app3.pdf]

References

1. Sagir A, Goyal R. An assessment of the awareness of local anesthetic systemic toxicity among multi-specialty postgraduate residents. J Anesth 2015 Apr;29(2):299-302. [doi: 10.1007/s00540-014-1904-9] [Medline: 25138822]
2. Denke NJ. Local anesthetic systemic toxicity (LAST): more common than you think. J Emerg Nurs 2024 May;50(3):336-341. [doi: 10.1016/j.jen.2024.02.002] [Medline: 38705705]
3. Elmaleh Y. Anesthésie locorégionale: comment les solutions et les techniques se sont adaptées aux pratiques [Article in French]. Le Praticien en Anesthésie Réanimation 2023 Dec;27(6):397-404. [doi: 10.1016/j.pratan.2023.10.010]
4. On'Gele MO, Weintraub S, Qi V, Kim J. Local anesthetics, local anesthetic systemic toxicity (LAST), and liposomal bupivacaine. Anesthesiol Clin 2024 Jun;42(2):303-315. [doi: 10.1016/j.anclin.2023.11.011] [Medline: 38705678]
5. Gitman M, Fettiplace MR, Weinberg GL, Neal JM, Barrington MJ. Local anesthetic systemic toxicity: a narrative literature review and clinical update on prevention, diagnosis, and management. Plast Reconstr Surg 2019;144(3):783-795. [doi: 10.1097/PRS.0000000000005989]
6. Neal JM, Barrington MJ, Fettiplace MR, et al. The third American Society of Regional Anesthesia and Pain Medicine practice advisory on local anesthetic systemic toxicity: executive summary 2017. Reg Anesth Pain Med 2018 Feb;43(2):113-123. [doi: 10.1097/AAP.0000000000000720] [Medline: 29356773]
7. Rosenberg PH, Veering BT, Urmey WF. Maximum recommended doses of local anesthetics: a multifactorial concept. Reg Anesth Pain Med 2004;29(6):564-575. [doi: 10.1016/j.rapm.2004.08.003] [Medline: 15635516]
8. Riff C, Le Caloch A, Dupouey J, et al. Local anesthetic plasma concentrations as a valuable tool to confirm the diagnosis of local anesthetic systemic toxicity? A report of 10 years of experience. Pharmaceutics 2022 Mar 26;14(4):708. [doi: 10.3390/pharmaceutics14040708] [Medline: 35456542]
9. Williams DJ, Walker JD. A nomogram for calculating the maximum dose of local anaesthetic. Anaesthesia 2014 Aug;69(8):847-853. [doi: 10.1111/anae.12679] [Medline: 24820093]
10. Suppan M, Beckmann TS, Gercekci C, et al. Development and preliminary validation of LoAD Calc, a mobile app for calculating the maximum safe single dose of local anesthetics. Health Care 2021 Jun 25;9(7):799. [doi: 10.3390/healthcare9070799]

XSL•FO
RenderX

11. Fubini PE, Savoldelli GL, Beckmann TS, Samer CF, Suppan M. Impact of a mobile app (LoAD Calc) on the calculation of maximum safe doses of local anesthetics: protocol for a randomized controlled trial. JMIR Res Protoc 2024 Jan 3;13:e53679. [doi: 10.2196/53679] [Medline: 38170571]
12. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
13. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet E-surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34. [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
14. 810.30 Federal Act of 30 September 2011 on research involving Human Beings (Human Research Act, HRA). Fedlex. URL: https://www.admin.ch/opc/en/classified-compilation/20061313/index.html [accessed 2024-06-01]
15. Suppan M, Fubini PE, Stefani A, Gisselbaek M, Samer CF, Savoldelli GL. Performance of 3 conversational generative artificial intelligence models for computing maximum safe doses of local anesthetics: comparative analysis. JMIR AI 2025 May 13;4:e66796. [doi: 10.2196/66796] [Medline: 40605845]
16. Brooke J. SUS: a “quick and dirty” usability scale. In: Usability Evaluation In Industry, 1st edition: CRC Press; 1996:189-194.
17. Gronier G, Baudet A. Psychometric evaluation of the F-SUS: creation and validation of the French version of the system usability scale. Int J Hum Comput Interact 2021 Oct 2;37(16):1571-1582. [doi: 10.1080/10447318.2021.1898828]
18. Beckmann TS, Samer CF, Wozniak H, Savoldelli GL, Suppan M. Local anaesthetics risks perception: a web-based survey. Heliyon 2023 Jan 15;10(1):e23545. [doi: 10.1016/j.heliyon.2023.e23545] [Medline: 38187280]
19. Douglas RN, Niesen AD, Johnson RL, et al. A single center descriptive study of local anesthetic dose in knee arthroplasty: was there evidence of local anesthetic systemic toxicity? J Clin Anesth 2024 Oct;97:111534. [doi: 10.1016/j.jclinane.2024.111534] [Medline: 38943851]
20. Taenzer AH, Sites BD, Kluger R, Barrington M. Settled science or unwarranted variation in local anesthetic dosing? An analysis from an International Registry of Regional Anesthesiology. Reg Anesth Pain Med 2019 Sep 6;44(11):998-1002. [doi: 10.1136/rapm-2019-100650] [Medline: 31494594]
21. Diver AJ. The ‘WiMP’ formula for local anaesthetic volume calculation. J Plast Reconstr Aesthet Surg 2008 Jul;61(7):743. [doi: 10.1016/j.bjps.2007.09.057]
22. MDCalc. URL: https://www.mdcalc.com/ [accessed 2025-10-17]
23. Local anesthetic toxic dose calculator. Scribd. URL: https://www.scribd.com/document/448872869/Local-Anesthetic-Toxic-Dose-Calculator [accessed 2026-07-08]
24. SafeLocal. App Store. URL: https://apps.apple.com/us/app/safelocal/id1440999841 [accessed 2026-07-07]
25. Carabalona JF, Delwarde B, Duclos A, et al. Serum concentrations of lidocaine during bariatric surgery. Anesth Analg 2020;130(1):e5-e8. [doi: 10.1213/ANE.0000000000003905]
26. Caruso D, De Santis D, Rivosecchi F, et al. Lean body weight - tailored iodinated contrast injection in obese patient: Boer versus James formula. Biomed Res Int 2018;2018:8521893. [doi: 10.1155/2018/8521893] [Medline: 30186869]
27. French J, Sharp L. Local anaesthetics. Ann R Coll Surg Engl 2012 Mar;94(2):76-80. [doi: 10.1308/003588412X13171221502185]
28. Farao J, Malila B, Conrad N, Mutsvangwa T, Rangaka MX, Douglas TS. A user-centred design framework for mHealth. PLoS ONE 2020;15(8):e0237910. [doi: 10.1371/journal.pone.0237910] [Medline: 32813711]
29. Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons for physicians not adopting clinical decision support systems: critical analysis. JMIR Med Inform 2018 Apr 18;6(2):e24. [doi: 10.2196/medinform.8912] [Medline: 29669706]
30. Smith S, Houghton A, Mockeridge B, van Zundert A. The internet, apps, and the anesthesiologist. Health Care 2023 Nov 20;11(22):3000. [doi: 10.3390/healthcare11223000]
31. Siebert JN, Bloudeau L, Combescure C, et al. Effect of a mobile app on prehospital medication errors during simulated pediatric resuscitation: a randomized clinical trial. JAMA Netw Open 2021 Aug 2;4(8):e2123007. [doi: 10.1001/jamanetworkopen.2021.23007] [Medline: 34459905]
32. Lacour M, Bloudeau L, Combescure C, et al. Impact of a mobile app on paramedics’ perceived and physiologic stress response during simulated prehospital pediatric cardiopulmonary resuscitation: study nested within a multicenter randomized controlled trial. JMIR Mhealth Uhealth 2021 Oct 7;9(10):e31748. [doi: 10.2196/31748] [Medline: 34617916]
33. Suppan M. LA RCT load calc vs usual. Yareta. URL: https://doi.org/10.26037/yareta:idtxttl4yncyrjmxp354tmqwfi [accessed 2026-07-07]

Abbreviations

ASRA: American Society of Regional Anesthesia and Pain Medicine
AW: actual weight
CCER: Commission Cantonale d'Éthique de la Recherche
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
CONSORT-eHEALTH: Consolidated Standards of Reporting Trials Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth

XSL•FO
RenderX

IBW: ideal body weight
LA: local anesthetic
LAST: local anesthetic systemic toxicity
LoAD Calc: Local Anesthetics Dose Calculator
RR: relative risk

Edited by A Stone; submitted 09.Dec.2025; peer-reviewed by K Pang, L Xie; revised version received 30.Jun.2026; accepted 30.Jun.2026; published 30.Jul.2026.

Please cite as:
Stefani A, Fubini PE, Beckmann TS, Samer CF, Savoldelli GL, Suppan M
Evaluation of the Impact of a Mobile App (LoAD Calc) on the Calculation of Maximum Safe Doses of Local Anesthetics: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e89236
URL: https://mhealth.jmir.org/2026/1/e89236
doi:10.2196/89236

© Alexandra Stefani, Pietro Elias Fubini, Tal Sarah Beckmann, Caroline Flora Samer, Georges Louis Savoldelli, Mélanie Suppan. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 30.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Deep Neural Network for Interpreting Wearable Electrocardiogram Data in Atrial Fibrillation: Prospective Observational Diagnostic Accuracy Study

Olli A Rantula[1,2,3], MSc, MD; Jukka A Lipponen[4], PhD; Jari Halonen[1,3], MD, PhD; Helena Jäntti[5], MD, PhD; Tuomas T Rissanen[1,6], MD, PhD; Mika P Tarvainen[4], PhD; Noora S Naukkarinen[1,2,3], MD; Eemu-Samuli Seljola[7,8], MD, PhD; Onni E Santala[1,9], MD, PhD; Jagdeep Sedha[1,2,3], MD; Tero J Martikainen[10], MD, PhD; Juha E K Hartikainen[1,3], MD, PhD

[1]School of Medicine, Faculty of Health Sciences, University of Eastern Finland, Yliopistonranta 1, PO BOX 1627, Kuopio, Finland

[10]Department of Prehospital Emergency Care, Acute Services, Kuopio University Hospital, Kuopio, Finland

[2]Doctoral School, Faculty of Health Sciences, University of Eastern Finland, Kuopio, Finland

[3]Heart Center, Kuopio University Hospital, Kuopio, Finland

[4]Department of Technical Physics, Faculty of Science and Forestry, University of Eastern Finland, Kuopio, Finland

[5]Department of Anaesthesiology and Intensive Care, North Karelia Central Hospital, Joensuu, Finland

[6]Heart Center, North Karelia Central Hospital, Joensuu, Finland

[7]Faculty of Medicine, University of Helsinki, Helsinki, Finland

[8]Conservative Ward Department, Mikkeli Central Hospital, Mikkeli, Finland

[9]Anesthesiology and Intensive Care, Kuopio University Hospital, Kuopio, Finland

Corresponding Author:
Olli A Rantula, MSc, MD
School of Medicine, Faculty of Health Sciences, University of Eastern Finland, Yliopistonranta 1, PO BOX 1627, Kuopio, Finland

Abstract

Background: Atrial fibrillation (AF) and atrial flutter (AFL) are common arrhythmias associated with the risk of ischemic stroke, which can be reduced with anticoagulation therapy. Thus, early diagnosis of AF and AFL is essential. However, diagnosis may be challenging due to the paroxysmal and asymptomatic nature of these arrhythmias.

Objective: Current diagnostic workflows involve time-consuming and resource-intensive manual review of noisy signals and prolonged recordings. We evaluated a mobile system that combines a wireless wearable single-lead chest strap electrocardiogram (ECG) and a novel deep neural network (DNN)–based artificial intelligence (AI) method for detecting AF/AFL episodes, AF/AFL burden, and rhythm change and estimated the delay in the detection of rhythm change from AF/AFL to sinus rhythm. We also assessed the rhythm classification performance.

Methods: A total of 116 patients with recent-onset AF or AFL undergoing cardioversion were monitored using a mobile single-lead chest strap ECG system. Simultaneously, a 3-lead Holter ECG served as the reference. The DNN-based AI analyzed the single-lead chest strap ECG data to detect AF/AFL, non-AF/AFLrhythm, and noninterpretable segments, as well as to estimate AF/AFL burden and detect rhythm change. Performance metrics included sensitivity, specificity, positive predictive value, negative predictive value, and intraclass correlation coefficient for AF and AFL burden estimation.

Results: The sensitivity and specificity for detecting AF/AFL were 91.9% (204.9/223.0 h) and 99.6% (242.4/243.5 h), respectively. The sensitivity for detecting AF was 96.2% (191.5/199.0 h), whereas it was 55.8% (13.4/24.0 h) for detecting AFL. The positive predictive value and negative predictive value for AF/AFL detection were 99.5% (204.9/206.0 h) and 93.1% (242.4/260.5 h), respectively. The intraclass correlation coefficient between the AF/AFL burden estimated by the DNN-based AI method and that derived from the physician-interpreted reference ECG was 0.96 (95% CI: 0.94 - 0.97; P<.001). Rhythm change detection occurred within 1 minute in most cases.

Conclusions: The mobile single-lead chest strap ECG system powered by a DNN-based AI algorithm demonstrated strong performance in detecting AF, estimating AF burden, and recognizing rhythm change to sinus rhythm. This AI-driven approach enables automated and accurate rhythm analysis, supporting clinical decision-making. Further validation in real-world ambulatory settings is warranted.

(JMIR Mhealth Uhealth 2026;14:e82475) doi:10.2196/82475

XSL•FO
RenderX

KEYWORDS

atrial fibrillation; single-lead electrocardiogram; ambulatory monitoring; automatic rhythm detection; artificial intelligence; deep neural network

Introduction

Atrial fibrillation (AF) is the most common cardiac tachyarrhythmia worldwide. It is characterized by uncoordinated electrical activation of the atria and irregular rhythm. AF is often accompanied by atrial flutter (AFL). Both AF and AFL reduce quality of life and are associated with significant morbidity and mortality [1-6]. The most severe complication related to AF is ischemic stroke. At least two-thirds of AF-related strokes can be prevented with proper anticoagulation treatment if AF is diagnosed early enough [1,6-10]. AFL is also associated with increased stroke risk, although the risk is somewhat lower than the risk related to AF [11]. Therefore, a similar risk evaluation and anticoagulation policy is recommended for AF and AFL [8]. With an aging population, the prevalence of AF and AFL is increasing, and screening for AF in patients at high risk for stroke is recommended [8,12-15]. The European Society of Cardiology guidelines recommend opportunistic screening for AF in individuals aged over 65 years [8]. The diagnosis of AF and AFL requires a high-quality electrocardiogram (ECG) recording which must be interpreted by a physician [8]. However, the diagnosis of AF and AFL is challenging, as they are often paroxysmal and frequently asymptomatic. Traditional 12-lead and ambulatory ECGs are limited by their time frame and are far from ideal prolonged everyday monitoring [8].

The diagnosis of AF and AFL in current clinical practice typically prompts the initiation of anticoagulation therapy if specific risk factors for stroke are present [8,9]. In addition to the presence of AF and AFL, AF and AFL burden contribute to the stroke risk. AF and AFL burden refers to the cumulative time spent in AF or AFL during a monitoring period. It reflects both the frequency and duration of AF episodes. This measure has been proposed as one of the metrics for guiding personalized AF and AFL management [8].

Several types of wearable sensors are available and capable of capturing ECG signals. These may be used for arrhythmia detection in paroxysmal AF and AFL. However, longer recordings generate more data, and interpretation represents a substantial workload for physicians, increasing resource use and health care costs.

Particularly, increasing the duration of recordings require the use of automatic and reliable data analytics to reduce manual effort and minimize resource waste. Various automated algorithms are available for analyzing extended ECG data, and several studies have evaluated their effectiveness in arrhythmia detection [16-22]. However, data from wearable devices, which are usually equipped with dry electrodes, often contain extensive noise and artifacts, which can negatively impact detection accuracy and further increase the workload required to analyze the recordings. The development of advanced artificial intelligence (AI) has provided new tools for automated data analysis.

In this study, we used a mobile single-lead chest strap ECG system that combines a wireless, mobile phone–connected consumer-grade ECG device with an automatic deep neural network (DNN)–based artificial intelligence (AI) method. The DNN-based method was originally developed for analyzing 3-lead ambulatory ECG recordings to detect AF and AFL.

The aim of this study was to evaluate the diagnostic performance of this DNN-based method when applied to wearable single-lead ECG data, including its ability to detect AF/AFL episodes; estimate AF/AFL burden; identify rhythm changes and associated detection delays; and classify recordings into the AF/AFL, non-AF/AFL, and noninterpretable categories.

Methods

Study Population and Participation Criteria

This study is part of a larger research project Wearables and Biomarkers in Detecting Atrial Fibrillation (WB-AF). It is registered in the ClinicalTrials database (NCT04917653).

Data collection was conducted at the emergency care department of Kuopio University Hospital. A total of 291 volunteer patients were screened between June 2022 and July 2023. The inclusion criteria were recent-onset (<48 h) AF or AFL confirmed by a 12-lead ECG, age ≥18 years, and being scheduled for cardioversion during the index treatment period at the emergency department. The type of cardioversion included but was not limited to electrical synchronized defibrillation, antiarrhythmic medication, and spontaneous rhythm change. Successful rhythm change was not a requirement for inclusion.

The exclusion criteria were BMI ≥35 kg/m^2, a cardiac pacemaker, or a medical condition requiring immediate treatment. Patients with BMI ≥35 kg/m² were excluded for practical reasons related to device fit and stability, as the chest strap ECG sensor may not reliably fit or remain in place in individuals with severe obesity. This exclusion was applied to ensure proper functioning of the recording setup.

Initially, 127 patients were recruited in the study. Eleven patients were excluded due to technical issues related to data acquisition. In 9 cases, data were missing because of a bug or instability in the measurement application. In 2 cases, problems related to device use led to complete corruption of the recorded data. The final study population comprised 116 patients with AF or AFL. Of these, 109 patients successfully resumed sinus rhythm (SR) and had both AF/AFL and SR data available, while 7 patients had only AF/AFL data due to unsuccessful cardioversion.

Clinical data were extracted from medical records and patient interviews. Symptoms reported during ECG recordings were documented using a structured questionnaire designed for this study. The CHA_2DS_2-VA score (congestive heart failure, hypertension, age ≥75 years, diabetes, prior stroke or transient ischemic attack, vascular disease, age 65-74 years) was used to estimate the risk of stroke. This scoring system is in line with

XSL•FO
RenderX

clinical guidelines and accounts for several risk factors. In this system, 2 points each are assigned for age ≥75 years and a history of stroke or transient ischemic attack, and 1 point each is assigned for age between 65 and 74 years, congestive heart failure, hypertension, diabetes mellitus, and vascular disease. In accordance with clinical guidelines, a score of ≥2 indicates an elevated stroke risk and most likely indicates the need for anticoagulant therapy [8].

Mobile Single-Lead Chest Strap ECG System and Automated DNN-Based AI Method for Rhythm Detection

Our mobile single-lead chest strap ECG system consisted of 3 main components: a wearable single-lead chest strap ECG sensor, a dedicated mobile phone application for data collection, and a DNN-based AI method for analytics. The ECG sensor, along with an elastic chest-worn belt designed for consumer use (Movesense Medical, Movesense Ltd.), recorded single-lead ECG signals at a sampling rate of 256 Hz. Data were transmitted wirelessly via Bluetooth to a mobile phone, which ran an in-house–developed mobile application responsible for data collection. The data were then forwarded to a DNN-based AI method operating on a dedicated computing platform.

In this study, we used a DNN model based on a residual network architecture developed in our research group. A similar residual network has been successfully applied to ECG arrhythmia classification tasks in earlier work, including the study by Hannun et al [23].

The DNN-based AI method was trained using ECG data from a total of 15,816 recordings obtained from 15,242 patients across 7 heterogeneous datasets collected using different clinical devices. Of these, 10,248 recordings originated from publicly available annotated AF/AFL datasets, and 5568 recordings were derived from locally collected clinical datasets from Kuopio University Hospital and from our previous studies. The training data consisted of multilead ECG recordings, including standard 12-lead ECGs and ambulatory ECG recordings with durations ranging from 30 minutes to 24 hours (Holter-type recordings), reflecting typical real-world clinical conditions. Recordings were acquired using different ECG systems with sampling rates ranging from 125 to 500 Hz and most commonly included leads I, II, and V3, although lead configurations varied slightly between datasets. Across the training dataset, 3180 patients had AF or AFL, while the remaining recordings represented non-AF/AFL rhythms. The training dataset did not include any single-lead ECG recordings.

The model processes raw ECG input in 10-second segments, each consisting of 3 input channels sampled at 125 Hz. In the original model architecture, these channels correspond to 3 distinct ECG leads. Because the chest strap device used in this study provides only a single-lead ECG signal, the same waveform was replicated across all 3 input channels when applying the model to chest strap recordings. This design choice was made to preserve compatibility with the original, previously validated 3-channel network architecture without modifying the model structure or retraining the network.

No additional patient-related information was required, enabling fully automated analysis based solely on the ECG waveform. Each 10-second segment was classified into 1 of 3 categories: AF/AFL; non-AF/AFL, including SR; or noninterpretable due to poor signal quality.

The model produces an independent rhythm classification for each 10-second segment. To improve clinical relevance and reduce short detections, the model output was subsequently postprocessed to account for temporal continuity of rhythms. Specifically, consecutive 10-second segments classified as AF/AFL were combined into continuous episodes. AF/AFL detections shorter than 30 seconds (fewer than 3 consecutive AF/AFL-positive 10-second segments) were excluded to align with commonly used guideline-based definitions that require at least 30 seconds of continuous single-lead ECG for AF diagnosis [8]. Closely spaced AF/AFL segments separated by short gaps were merged into unified episodes. Performance metrics were calculated on the basis of the total duration of correctly and incorrectly classified 10-second segments and are reported as hours correctly classified divided by total hours (h/h). AF burden was defined as the total duration of AF/AFL during the ECG recording period.

Reference for Rhythm Detection and Quality Analysis

Written informed consent was obtained from all patients after eligibility was confirmed in the emergency department. Immediately after consent was obtained, patients were equipped with both the mobile single-lead chest strap ECG device and a 3-lead ambulatory Holter ECG device, and continuous ECG monitoring was initiated.

ECG monitoring continued throughout the treatment period, including during cardioversion and subsequent observation, and the devices were removed prior to hospital discharge. All analyses of algorithm performance were performed retrospectively after data collection was complete.

The concurrently recorded 3-lead Holter ECG was used as the reference standard. Reference ECG data were recorded at a sampling frequency of 250 Hz using an ambulatory Holter device (Faros 360, Bittium), capturing leads I, II, and V3. Synchronization between the single-lead and 3-lead recordings was initially based on timestamps and was manually verified by aligning the recordings to the documented time of cardioversion.

All reference ECG recordings were manually interpreted by an experienced physician trained in ECG interpretation. The physician annotated rhythm categories (AF, AFL, and SR) and the timing of cardioversion. The same physician also visually assessed the single-lead chest strap ECG recordings for signal quality. Periods of at least 10 seconds considered noninterpretable due to poor signal quality were labeled as noninterpretable, whereas shorter noisy periods were assumed to represent the same rhythm as that observed immediately before and after the segment. These physician-labeled annotations served as the reference for evaluating the performance of the DNN-based method in detecting noninterpretable segments.

XSL•FO
RenderX

Statistical Methods

Rhythm detection performance was evaluated using both time-based and episode-based (patient-level) approaches.

In the time-based analysis, the mobile single-lead chest strap ECG data were analyzed as consecutive 10-second intervals. Each interval classified by the DNN-based AI method was compared with the corresponding reference rhythm derived from the 3-lead Holter ECG. A classification was considered correct if the rhythm category (AF/AFL, non-AF/AFL, or noninterpretable) matched the reference for the same time interval.

In the episode-based analysis, ECG recordings were evaluated at the patient level to assess whether AF or AFL was detected during reference-confirmed AF/AFL episodes. A patient was considered correctly identified as AF/AFL-positive if the DNN-based method detected at least 1 AF/AFL segment within a reference-confirmed AF/AFL episode. Performance was assessed against the reference using sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and Youden index. For episode-based AF/AFL detection, the Youden index was calculated to summarize the overall balance between sensitivity and specificity. Sensitivity was also calculated separately for AF and AFL.

Performance metrics included sensitivity, specificity, PPV, NPV, F_1-score, and Cohen kappa. For proportion-based metrics (sensitivity, specificity, PPV, and NPV), 95% CIs were calculated using Wilson score intervals. For F_1-scores, 95% CIs were estimated using nonparametric bootstrap resampling (2000 iterations). Cohen kappa was used to assess the agreement between rhythm categories (AF/AFL, non-AF/AFL, or noninterpretable) assigned by the DNN-based AI method and the reference.

The AF/AFL burden estimated by the DNN-based method was compared with the reference AF/AFL burden derived from the 3-lead ECG using Bland-Altman analysis, reporting the mean difference (bias) and 95% limits of agreement (mean difference ± 1.96 SD). Agreement was further assessed using the intraclass correlation coefficient (ICC) based on a 2-way mixed-effects model with absolute agreement and single measurements (ICC model 3.1). Bland-Altman and scatter plots were used for visualization [24].

Delay in rhythm change detection was defined as the time difference between the first SR segment identified in the reference 3-lead ECG after conversion from AF/AFL and the time point at which the DNN-based method first classified the rhythm as SR in the single-lead ECG after postprocessing. Delays were calculated for each AF/AFL-to-SR transition event and summarized descriptively. A 2-sided significance level of $P<.05$ was used.

Ethical Considerations

This study received approval from the Ethics Committee of the Northern Savo Hospital District (Approval number: 2061/2020). Prior to enrollment, all participants were provided with detailed written information about the study procedures, and written informed consent was obtained from each participant. Participation was voluntary, and participants had the right to withdraw from the study at any time. This study complied with the ethical principles set forth in the Declaration of Helsinki for research involving human participants.

All data were anonymized prior to analysis, and no identifiable personal information was included in the dataset. Data were handled and stored in accordance with applicable data protection regulations. Participants did not receive any financial compensation for participation in this study.

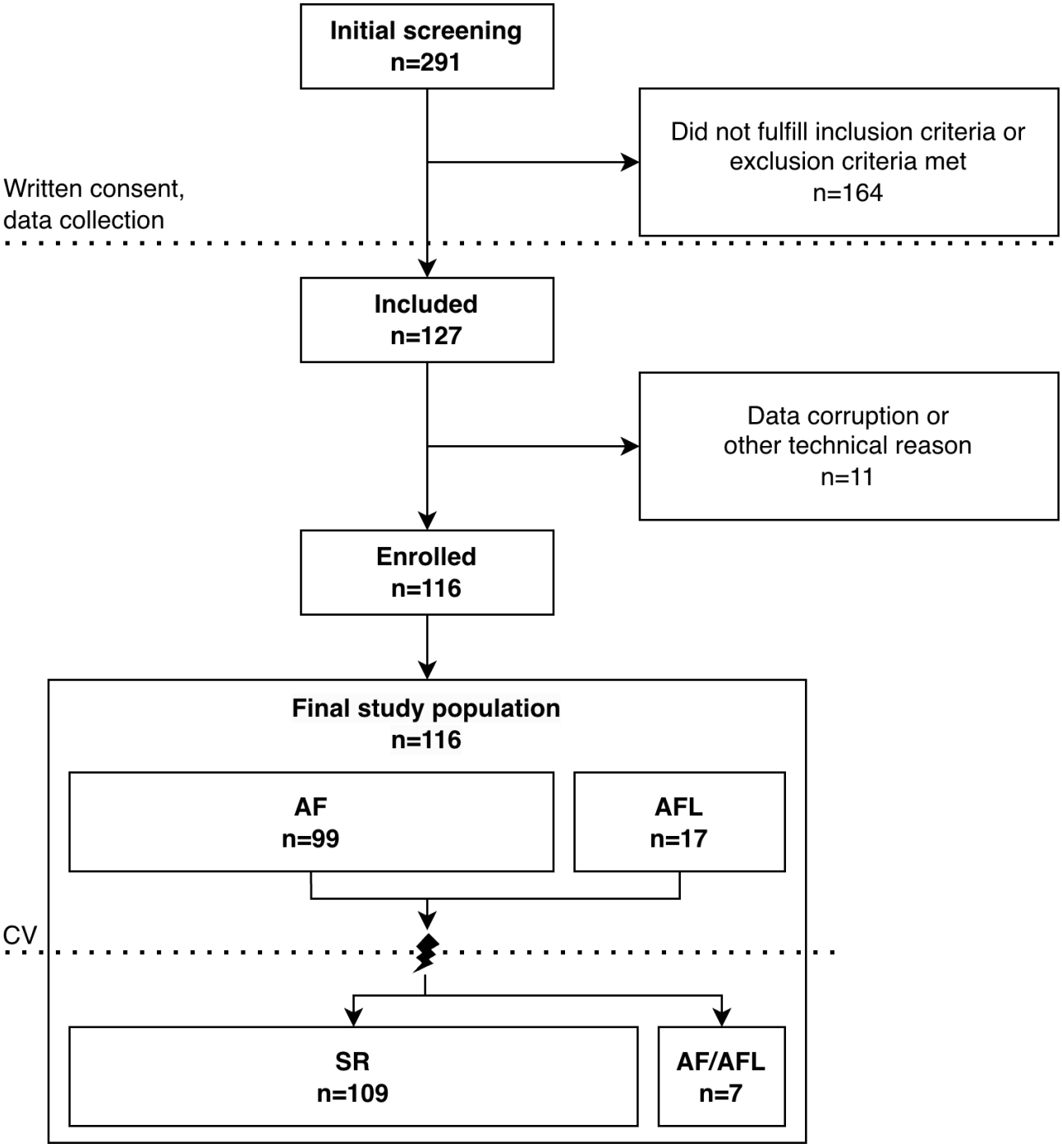
Results

Study Population

The final study population comprised data collected from 116 patients diagnosed with AF (n=99) or AFL (n=17) who were scheduled for SR restoration. The data collection flowchart is presented in Figure 1. SR was achieved in 109 patients. The mean age was 65.1 (SD 11.4) years (range 24-90 years), the mean BMI 27.5 (SD 4.2) kg/m² (range 18.8-34.9 kg/m²), and 60.3% (70/116) of the patients were male (Table 1). Hypertension was the most prevalent comorbidity present in 51.7% (60/116) patients, 90.5% (105/116) had a prior history of AF or AFL, and 9.5% (11/116) presented with first-diagnosed AF/AFL. The most frequently reported symptoms prior to hospital admission were palpitations (88.8%, 103/116), fatigue (50%, 58/116), and decline of general condition (44.8%, 52/116). The CHA_2DS_2-VA score was 0, 1, and ≥2 for 27 (23.3%), 32 (27.6%), and 57 (49.1%) of the 116 patients, respectively.

XSL•FO
RenderX

Figure 1. Data collection flowchart. AF, atrial fibrillation; AFL, atrial flutter; CV, cardioversion; SR, sinus rhythm.



XSL•FO
RenderX

Table . Patient demographics (N=116).

	Value
Characteristic	
Age (years), mean (SD)	65.1 (11.4)
Weight (kg), mean (SD)	83.0 (15.9)
Height (cm), mean (SD)	173.6 (9.4)
BMI (kg/m^2), mean (SD)	27.5 (4.2)
Male sex, n (%)	70 (60.3)
Recording	
Overall duration (min), mean (SD)	241.2 (93.6)
AF[a]/AFL[b] duration (min), mean (SD)	122.0 (87.1)
SR[c] duration (min), mean (SD)	119.2 (46.6)
Medical history	
Previous AF diagnosis, n (%)	105 (90.5)
Hypertension, n (%)	60 (51.7)
Coronary artery disease, n (%)	18 (15.5)
Structural heart disease[d], n (%)	17 (14.6)
Diabetes, n (%)	16 (13.8)
Other arrhythmia[e], n (%)	8 (6.9)
Prior cardiac surgery, n (%)	8 (6.9)
Congestive heart failure, n (%)	2 (1.7)
Medication	
Beta blocker, n (%)	92 (79.3)
Anticoagulation therapy, n (%)	74 (63.8)
Flecainide, n (%)	15 (12.9)
Amiodarone, n (%)	4 (3.4)
Digoxin, n (%)	1 (0.9)
Symptoms during to hospital admission	
Palpitation, n (%)	103 (88.8)
Fatigue, n (%)	58 (50)
General condition decline, n (%)	52 (44.8)
Respiratory distress, n (%)	23 (19.8)
Chest pain, n (%)	15 (12.9)
Thromboembolic risk	
CHA_2DS_2VA[f]=0, n (%)	27 (23.3)
CHA_2DS_2VA=1, n (%)	32 (27.6)
CHA_2DS_2VA≥2, n (%)	57 (49.1)

[a]AF: atrial fibrillation.

[b]AFL: atrial flutter.

[c]SR: sinus rhythm.

[d]Presence of any physician-documented structural heart disease in the medical records.

[e]Previously diagnosed arrhythmia excluding atrial fibrillation or atrial flutter.

[f]The CHA_2 DS_2 -VA score estimates stroke risk in patients with AF/AFL and includes the following components: congestive heart failure (1 point), hypertension (1 point), age 65 - 75 years (1 point), diabetes mellitus (1 point), vascular disease (1 point), prior stroke or transient ischemic attack (2

XSL•FO
RenderX

points), and age >75 years (2 points).

ECG Data Characteristics, Quality, and Rhythm Classification

A total of 466.3 hours of concurrent single-lead chest strap ECG and reference 3-lead ECG data were collected, which included 223.0 hours (47.8%) of AF/AFL, consisting of 199.0 hours (42.7%) of AF and 24.0 hours (5.1%) of AFL as defined by the reference. Additionally, 219.0 hours (46.9%) of data represented SR based on the reference ECG. The physician identified 24.5 hours (5.3%) of single-lead chest strap ECG data as noninterpretable (Table 2).

Table . Confusion matrix for rhythm classification by the deep neural network (DNN)–based artificial intelligence method compared with the reference 3-lead Holter electrocardiogram[a].

DNN prediction	AF[b] (Holter)[c]	AFL[d] (Holter)	Non-AF/AFL (Holter)	Noninterpretable (physician)
AF/AFL	191.5 (41.1)	13.4 (2.9)	0.06 (0.0)	1.0 (0.2)
Non-AF/AFL	2.6 (0.6)	10.2 (2.2)	214.1 (45.9)	1.9 (0.4)
Noninterpretable	4.9 (1)	0.4 (0.1)	4.8 (1)	21.6 (4.6)

[a]Values are reported as number of hours (%) of total recording time (ie, 466.3 h).

[b]AF: atrial fibrillation.

[c]Holter: Holter monitor.

[d]AFL: atrial flutter.

The time-based performance of the DNN-based AI method in detecting AF/AFL, non-AF/AFL, and noninterpretable episodes is presented in the confusion matrix (Table 2), providing an overview of the classification results. The DNN-based AI method correctly classified 204.9 of 466.3 hours (43.9%) as AF/AFL. Misclassification into this category was minimal, comprising 0.06 of 466.3 hours (0%) of non-AF/AFL rhythm and 1.0 of 466.3 hours (0.2%) of noninterpretable data (Table 2). For SR, the DNN-based AI method correctly identified 214.1 of 466.3 hours (45.9%) as non-AF/AFL but misclassified 12.8 of 466.3 hours (2.7%) of actual AF/AFL and 1.9 of 466.3 hours (0.4%) of actual noninterpretable data into this category.

Additionally, the DNN-based AI method classified 31.7 of 466.3 hours (6.8%) of the data as noninterpretable, whereas the physician classified 24.5 of 466.3 hours (5.3%) as noninterpretable on the basis of visual assessment of the single-lead ECG (Table 2). Of the algorithm-labeled noninterpretable data, 21.6 of 466.3 hours (4.6%) were also considered noninterpretable by the physician.

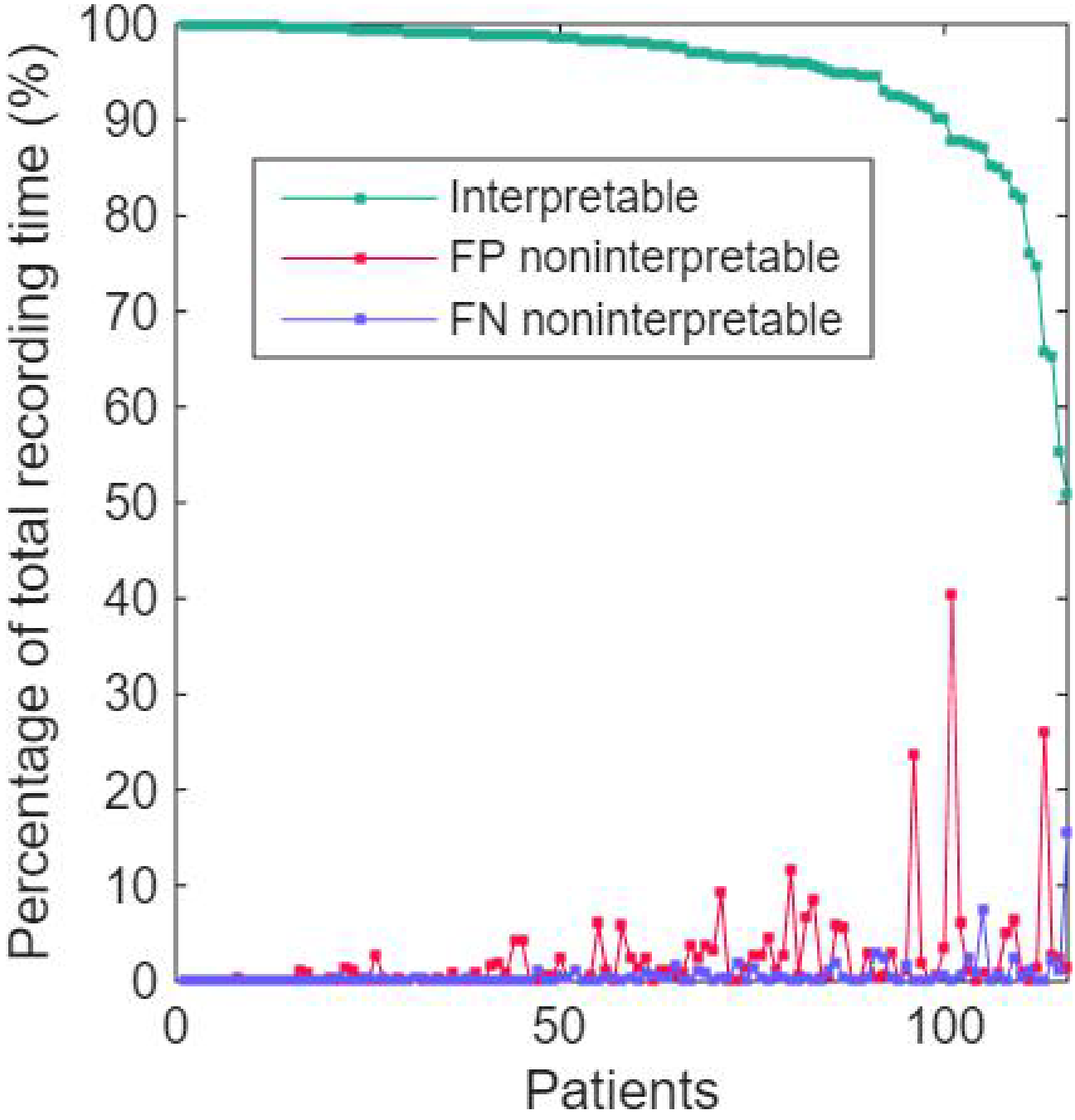
The difference was due to different signal quality thresholds and handling of short noisy segments between the algorithm and the physician. The physician labeled noisy periods of at least 10 seconds as noninterpretable, whereas shorter noisy segments were assumed to represent the surrounding rhythm. The DNN-based method evaluated each 10-second segment independently and could classify individual segments as noninterpretable even when adjacent segments were interpretable.

The DNN-based method assigned a rhythm label of AF/AFL and non-AF/AFL to 1.0 (0.2%) and 1.9 (0.4%) of 466.3 hours, respectively, which the physician considered noninterpretable due to poor signal quality. Conversely, the algorithm classified 4.9 hours (1.1%) during AF, 4.8 hours (1%) during SR, and 0.4 hours (0.1%) during AFL as noninterpretable despite the reference rhythm being known from the Holter ECG (Table 2). Importantly, this difference did not meaningfully affect the AF/AFL burden estimation, as reflected by the high agreement between methods (ICC 0.96).

For most participants, the proportion of single-lead chest strap ECG data considered interpretable on the basis of physician review exceeded 95% of the total recording time, and false-positive and false-negative noninterpretable classifications were generally low (Figure 2). In 3 patients, more than 20% of the recorded data were falsely classified as noninterpretable by the DNN-based AI method. In these patients, the total recording duration ranged from 1.0 to 5.0 hours, of which 0.4 to 2.2 hours were falsely classified as noninterpretable.

XSL•FO
RenderX

Figure 2. Proportion of interpretable single-lead electrocardiogram (ECG) data and classification errors per patient. Each data point represents 1 patient (N=116). Patients are ordered from left to right according to decreasing proportion of ECG data considered interpretable on the basis of physician review of the single-lead chest strap recordings. The y-axis shows the percentage of total recording time per patient. The green curve represents the proportion of data considered interpretable according to physician review. Red points indicate false-positive noninterpretable classifications (segments classified as noninterpretable by the DNN but considered interpretable by physician review), and blue points indicate false-negative noninterpretable classifications (segments classified as interpretable by the DNN but considered noninterpretable by physician review). FP, false positive; FN, false negative.



Performance of the DNN-Based AI Method for Time-Based Rhythm Detection

The performance of the DNN-based AI method in detecting rhythm categories showed an overall accuracy of 94.4% (440.3/466.3 h). Cohen kappa was 0.90, indicating excellent agreement between the DNN-based AI method and the reference.

For AF/AFL detection, the DNN-based AI method achieved a time-based sensitivity of 91.9% (204.9/223.0 h) and a specificity of 99.6% (242.4/243.5 h), with a PPV of 99.5% (204.9/206.0 h) and an NPV of 93.1% (242.4/260.5 h), yielding an F_1-score of 0.96 (Table 3). More specifically, the DNN-based AI method achieved a time-based detection sensitivity of 96.2% (191.5/199.0 h) for AF and 55.8% (13.4/24.0 h) for AFL, indicating stronger performance in identifying AF than AFL. For SR detection, the DNN-based AI method achieved a sensitivity of 97.8% (214.1/219.0 h) and a specificity of 94.1% (232.8/247.5 h), with a PPV of 93.6% (214.1/228.8 h) and an NPV of 97.9% (232.8/237.7 h), yielding an F_1-score of 0.96 (Table 3).

XSL•FO
RenderX

The performance of the mobile single-lead chest strap ECG system compared with the physician's interpretation for detecting noninterpretable cases was lower, with a time-based sensitivity of 88.2% (21.6/24.5 h) and a specificity of 97.7% (431.9/442.0 h; Table 3). This category had a PPV of 68.1% (21.6/31.7 h) and an NPV of 99.3% (431.9/434.8 h), yielding an F_1-score of 0.77. These results highlight the challenges in detecting and categorizing ambiguous or unclear cases. Detection results for the noninterpretable data category of the DNN-based AI method are presented in Figure 2.

Table . Performance metrics of the deep neural network (DNN)–based artificial intelligence method for rhythm detection from single-lead chest strap electrocardiogram data.

Metric[a]	AF[b]/AFL[c]	Non-AF/AFL	Noninterpretable
Sensitivity	204.9/223.0 (91.9, 95% CI[d] 91.6 - 92.2)	214.1/219.0 (97.8 95% CI 97.6 - 97.9)	21.6/24.5 (88.2, 95% CI 87.2 - 88.8)
Sensitivity for AF[e]	191.5/199.0 (96.2, 95% CI 96 - 96.5)	—[f]	—
Sensitivity for AFL[e]	13.4/24.0 (55.8, 95% CI 54 - 57.6)	—	—
Specificity	242.4/243.5 (99.6, 95% CI 99.5 - 99.6)	232.8/247.5 (94.1, 95% CI 93.9 - 94.4)	431.9/442.0 (97.7, 95% CI 97.6 - 97.8)
PPV[g]	204.9/206.0 (99.5, 95% CI 99.4 - 99.5)	214.1/228.8 (93.6, 95% CI 93.3 - 93.9)	21.6/31.7 (68.1, 95% CI 66.6 - 69.6)
NPV[h]	242.4/260.5 (93.1, 95% CI 92.8 - 93.3)	232.8/237.7 (97.9, 95% CI 97.7 - 98.1)	431.9/434.8 (99.3, 95% CI 99.3 - 99.4)
F_1-score[i]	0.96 (95% CI 0.95 - 0.96)	0.96 (95% CI 0.95 - 0.96)	0.77 (95% CI 0.76 - 0.78)
Proportion of data	222.9/466.3 (47.8)	219.0/466.3 (46.9)	24.5/466.3 (5.3)

[a]Performance metrics are based on 10-second segments and reported as correctly classified hours divided by total hours (%).

[b]AF: atrial fibrillation.

[c]AFL: atrial flutter.

[d]95% CIs for sensitivity, specificity, PPV, and NPV were calculated using Wilson score intervals.

[e]Because the DNN outputs a combined AF/AFL class, AF- and AFL-specific results can be reported only as sensitivities (proportion of AF or AFL time classified as AF/AFL).

[f]Not applicable.

[g]PPV: positive predictive value.

[h]NPV: negative predictive value.

[i]95% CIs for F_1-scores were estimated using bootstrap resampling.

Performance of the DNN-Based AI Method for Detection of AF/AFL Episodes, Rhythm Change, and AF/AFL Burden

For episode detection, the sensitivity and specificity of AF/AFL detection were 97.4% (113/116) and 98.2% (107/109), respectively, with a PPV of 98.3% (113/115) and an NPV of 97.3% (107/110). When calculated solely for AF cases (excluding AFL), the sensitivity was 100% (100/100) for AF episode detection and 81% (13/16) for AFL episode detection (Table 4). The Youden index for AF/AFL detection was 0.96. The delay in rhythm change detection after rhythm change was less than 1 minute in most cases (Figure 3).

XSL•FO
RenderX

Table . Atrial fibrillation (AF)/atrial flutter (AFL) episode detection and AF/AFL burden estimation.

Metric	Value
Episode-based confusion matrix for DNN[a] AF/AFL detection	
True Positive AF/AFL episode detection	113
False Positive AF/AFL episode detection	2
False Negative AF/AFL episode detection	3
True Negative AF/AFL episode detection	107
Episode-based performance metrics[b]	
Sensitivity for AF/AFL, episodes/episodes (%)	113/116 (97.4, 95% CI 92.7 - 99.1)
Specificity for AF/AFL, episodes/episodes (%)	107/109 (98.2, 95% CI 93.6 - 99.5)
PPV[c] for AF/AFL, episodes/episodes (%)	113/115 (98.3, 95% CI 93.9 - 99.5)
NPV[d] for AF/AFL, episodes/episodes (%)	107/110 (97.3, 95% CI 92.3 - 99.1)
Sensitivity for AF, episodes/episodes (%)	100/100 (100, 95% CI 96.3 - 100)
Sensitivity for AFL, episodes/episodes (%)	13/16 (81, 95% CI 57 - 93)
Youden index for AF/AFL	0.96
Mean AF/AFL burden estimated by DNN (h), mean (SD)	1.8 (1.5)
Mean AF/AFL burden by Holter electrocardiogram (h), mean (SD)	2.0 (1.5)
ICC[e] for AF/AFL burden	0.96 (95% CI 0.94 - 0.97; *P*<.001)

[a]DNN: deep neural network.

[b]Episode-based performance metrics are derived from the 2×2 confusion matrix (TP, FP, FN, and TN) shown in the table and are reported as episodes correctly classified divided by the total number of episodes (%).

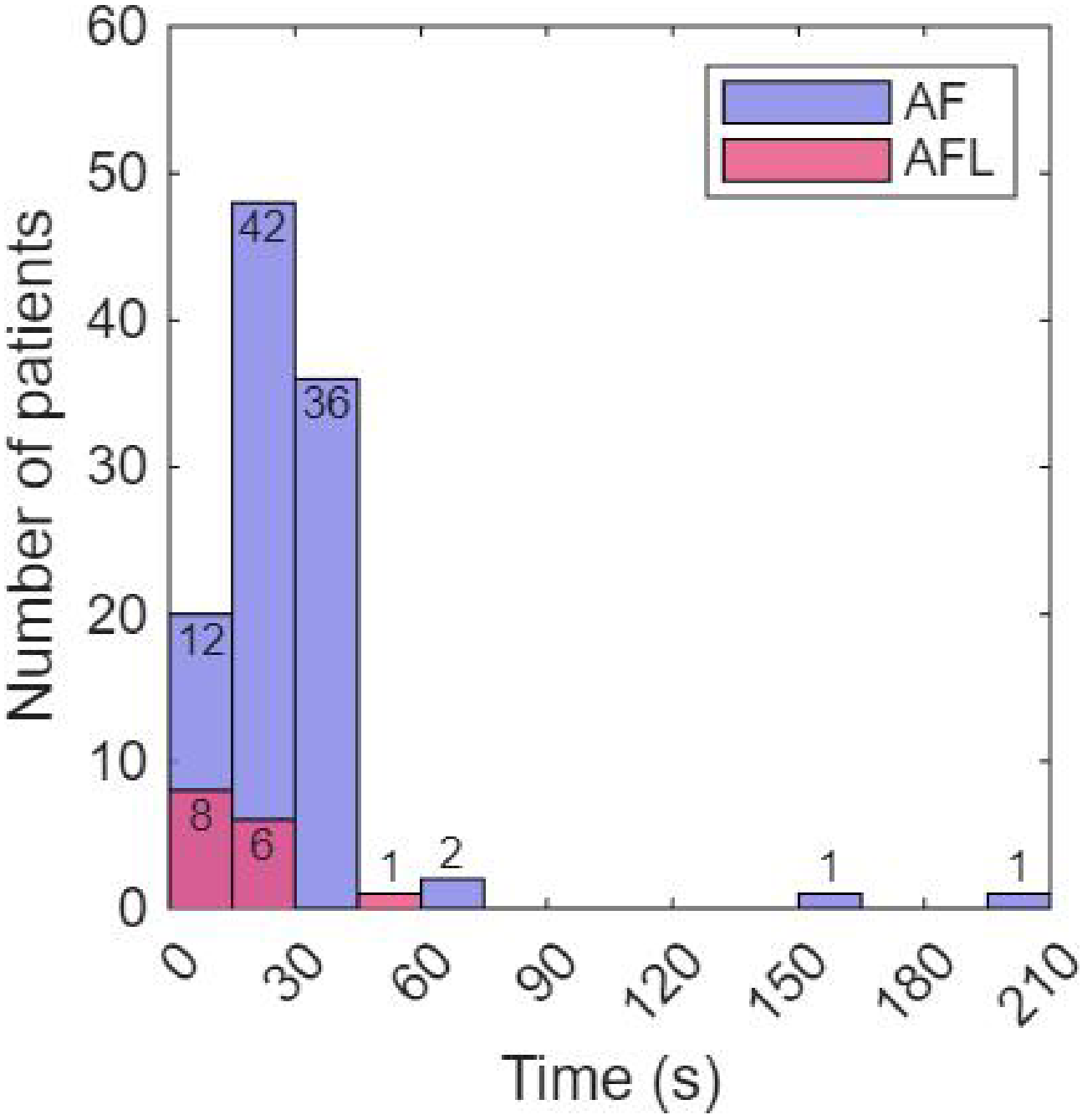
[c]PPV: positive predictive value.

[d]NPV: negative predictive value.

[e]ICC: intraclass correlation coefficient.

XSL•FO
RenderX

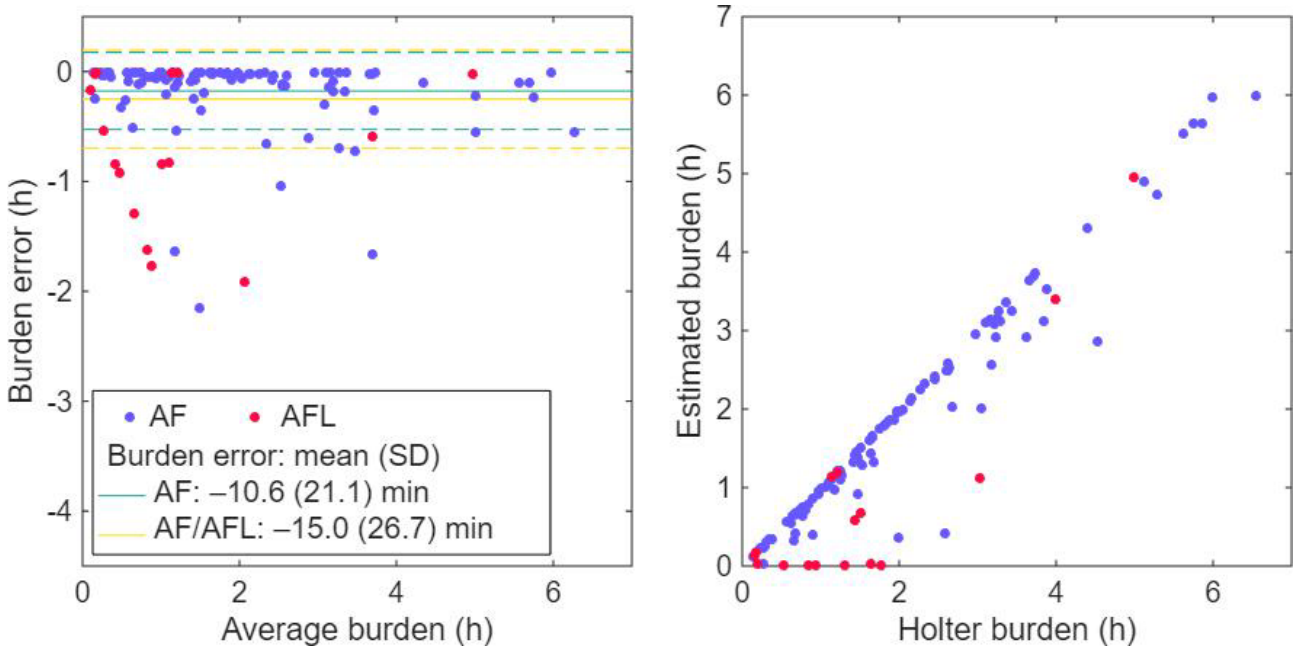
Figure 3. Distribution of delay in rhythm change recognition from AF/AFL to sinus rhythm following cardioversion or spontaneous rhythm change. AF, atrial fibrillation; AFL, atrial flutter.



The mean AF/AFL burden estimated by the mobile single-lead chest strap ECG system was 1.8 (SD 1.5) hours, whereas the corresponding burden derived from the reference ECG was 2.0 (SD 1.5) hours (Table 4). The mean difference in AF/AFL burden was −15.0 (SD 26.7) minutes, and for AF alone, it was −10.6 (SD 21.1) minutes. The Bland-Altman analysis demonstrated a mean bias of −0.25 hours for AF/AFL burden estimation, with 95% limits of agreement ranging from −1.12 to 0.62 hours. The Bland-Altman plot and scatter plot comparing the AF/AFL burden estimates between the DNN-based method and the reference are shown in Figure 4. Agreement between the methods was high, with an ICC of 0.96 (95% CI 0.94 - 0.97; P<.001) (Table 4). Visual inspection of the Bland-Altman plot did not suggest the presence of proportional bias across the range of AF/AFL burden values.

XSL•FO
RenderX

Figure 4. Bland-Altman plot and scatter plot of atrial fibrillation (AF) and atrial flutter (AFL) burden estimated from single-lead chest strap electrocardiogram (ECG) using a deep neural network–based artificial intelligence method compared with the reference 3-lead Holter ECG. The left panel shows the Bland-Altman plot, and the right panel shows the correlation between the estimated and reference AF/AFL burden.



Discussion

Principal Findings

The main finding of this study was that the DNN-based AI method applied to data obtained from a mobile chest strap ECG system demonstrated high diagnostic accuracy for detecting AF episodes, identifying rhythm changes from AF to SR, and estimating AF burden, with adequate sensitivity and specificity and strong agreement with the reference method. The mobile single-lead chest strap ECG system detected AF episodes more accurately than AFL episodes.

Our findings demonstrate high performance in the use of wearable devices and AI algorithms for AF detection, comparable to or exceeding results reported in previous studies [16-21,25-31]. For instance, the BASEL Wearable Study demonstrated that various consumer-grade smart devices achieved reasonable sensitivity (up to 85%) and specificity (up to 75%) for AF detection, although the rate of inconclusive recordings by the automated detection method was relatively high (ranging from 17% to 26%, depending on the device). However, manual review indicated that a majority of the tracings were interpretable [25]. Similarly, Pan et al [32] showed that using longer, 7-day recordings, smartwatch-based single-lead ECG can detect AF episodes with a duration of >5 minutes with a sensitivity of 89.7% and specificity of 67.4%. Moreover, the systematic review and meta-analysis by Manetas-Stavrakakis et al [31], which evaluated the diagnostic accuracy of AI-based technologies for AF detection, reported pooled sensitivity and specificity values of 92.3% and 96.2% for single-lead ECG, respectively.

Recent clinical studies further support the feasibility of artificial intelligence–based atrial fibrillation detection using single-lead ECG. De Guio et al [33] demonstrated enhanced detection of AF using a cloud-based AI platform applied to single-lead ECG recordings, highlighting the clinical applicability of AI-assisted rhythm analysis in simplified ECG configurations. Zhao et al [34] also developed a temporal convolutional network model for AF detection using single-lead ECG and reported high diagnostic accuracy, supporting the applicability of advanced neural network architectures to wearable ECG data.

In comparison, our study demonstrated even higher sensitivity (96.2%) for AF detection and comparable sensitivity (91.9%) and specificity (99.5%) for AF/AFL detection using the mobile single-lead chest strap ECG system, further encouraging the potential of consumer-grade devices in clinical settings in providing more information for diagnostics about AF estimation and the burden of AF.

The DNN-based AI method was trained using ECG data from a diverse patient population worldwide, incorporating various device types and lead configurations, including 3- and 4-lead Holter monitors and 12-lead ECG recordings. The model was evaluated using a single-lead ECG configuration that was not represented in the training data, as the chest strap lead does not correspond directly to any of the training leads. This setup allowed assessment of whether the model had learned features that generalize beyond specific lead configurations. The observed performance suggests that the model captures AF/AFL-related features that are transferable across different electrode placements, which may support its applicability across heterogeneous wearable ECG devices.

However, AFL detection was less reliable than AF detection, as reflected by the substantially lower sensitivity for AFL. This finding is clinically relevant and likely reflects the inherent limitations of single-lead ECG morphology in recognizing atrial flutter. Unlike AF, which is characterized by highly irregular RR intervals, AFL often presents with more regular ventricular response and subtle atrial activity patterns that can be difficult to distinguish from SR when only a single lead is available. In

XSL•FO
RenderX

addition, the DNN-based model used in this study was primarily trained on 3-lead ECG data, which provide richer spatial information on atrial activity. The transition from a 3-lead training environment to a single-lead wearable setting may therefore limit the ability to identify characteristic flutter wave morphology.

Furthermore, single-lead wearable ECG recordings are more susceptible to motion artifacts and noise, which may affect the detection of AFL compared to AF and contribute to a higher proportion of noninterpretable segments. Similar challenges have been reported in previous studies, where AI models trained on multilead ECG data demonstrated reduced performance when applied to single-lead recordings, particularly for arrhythmias with overlapping waveform characteristics [35,36]. From a clinical perspective, these findings suggest that while the system performs robustly for AF detection and AF burden estimation, caution is warranted when interpreting potential AFL cases. In situations where AFL is suspected or diagnostic uncertainty remains, confirmation using higher-quality or multilead ECG recordings may be necessary. These results also highlight the need for further development of AI models specifically optimized for single-lead ECG data to improve AFL detection in wearable monitoring settings.

The high sensitivity and specificity for AF detection suggest that the DNN-based AI method is robust and reliable for identifying AF episodes when data quality is sufficient. Accurate detection of rhythm changes between SR and AF is essential for identifying paroxysmal AF. In AF screening, the primary aim is to detect AF episodes and rhythm changes from SR to AF. Additionally, for assessing AF burden, it is important to detect both AF episodes and rhythm changes in both directions, from SR to AF and from AF to SR. In this study, only rhythm changes from AF to SR were detected due to the study protocol. However, given the high time-based detection performance with an accuracy of 94.4%, we assume that the DNN-based AI method is also capable of detecting rhythm changes from SR to AF.

AF burden is recognized in current clinical guidelines as an important predictor of stroke risk and a relevant factor in clinical decision-making [8]. Therefore, estimating AF burden from single-lead ECG data is particularly valuable for the continuous monitoring and management of AF and may serve as a complementary tool in clinical evaluation. Pan et al [32] showed sufficient correlation regarding AF burden between single-lead ECG and reference Holter ECG with an ICC of 0.927 over a 7-day monitoring period. Moreover, Hennings et al [37] demonstrated a high correlation (Pearson correlation coefficient 0.998) between an AI-based tool used to detect AF burden and manual assessment of AF burden in 7-day Holter ECG recording, with minimal negative bias of AF burden. Similarly, in our study, the ICC was 0.96, with a low rhythm change recognition time and low negative bias. In most cases, rhythm changes were detected immediately or at least within 1 minute after cardioversion. The high level of agreement between the AF burden measurements obtained using the mobile single-lead chest strap ECG system and the reference Holter ECG reinforces the validity of this approach for single-lead ECG usage. Notably, we used a dry-electrode ECG instead of a Holter monitor. This method is particularly suitable for long-term monitoring and is more comfortable than Holter monitors with potentially irritating wet electrodes. Moreover, the use of ECG allows for the diagnosis of AF according to current clinical guidelines [8].

Study Limitations

Despite the promising results, there are some limitations to consider. The study population was limited to patients with recent-onset AF or AFL who presented to the emergency department, which does not represent the broader population with paroxysmal or persistent AF/AFL with fluctuating rhythm. Patients were permitted to move freely during the study recordings. However, while hospitalized, they were primarily lying in the supine position, which affects the conclusions regarding the use of the devices in real-world settings. Additionally, studies involving a more diverse patient population and longer monitoring periods in real-world ambulatory settings are needed to validate the generalizability and long-term performance of this approach.

Conclusions

The high sensitivity and specificity for AF detection and strong correlation for AF burden suggest that the mobile chest strap ECG system, in which a DNN-based AI method is used with single-lead chest strap ECG, is promising for identifying AF episodes, rhythm change, and AF burden. A majority of the data was analyzable, and no complaints or discomfort regarding the wearables or study devices were reported.

Our study indicates the potential of consumer-grade single-lead ECG devices combined with an advanced detection algorithm for the detection and monitoring of AF/AFL, independent of the specific ECG device. This approach offers promising complementary tools for the continuous monitoring, diagnosis, and personalized management of AF and AFL.

Acknowledgments

We thank the nursing and medical staff at Kuopio University Hospital, with special thanks to Lasse Ollikainen, MD; Otso Heinonen, BM; Alina Ruhanen, BM; and Patrik Toivanen, BM, from the School of Medicine, University of Eastern Finland, for their efforts in data acquisition.

Funding

This study was supported by the Research Committee of the Kuopio University Hospital Catchment Area for the State Research Funding (project 5101137 and project 507T044). Olli Rantula received research support from the Finnish Medical Foundation,

XSL•FO
RenderX

the Finnish Foundation for Cardiovascular Research, the Ida Montin Foundation, the Aarne Koskelo Foundation, the Paulo Foundation, the Paavo Nurmi Foundation, and the Paavo Ilmari Ahvenainen Foundation.

Data Availability

The source data for this article cannot be shared publicly due to privacy concerns regarding the patients who participated in the study. The data will be made available upon reasonable request to the corresponding author.

Authors' Contributions

OAR contributed to conceptualization, methodology, data collection, formal analysis, data curation, visualization, and writing of the original draft. JAL contributed to conceptualization, methodology, software development, formal analysis, data curation, visualization, and review and editing of the manuscript. MPT contributed to conceptualization and software development. JH, HJ, TTR, JS, ESV, and OES contributed to conceptualization, methodology, and review and editing of the manuscript. NSN contributed to conceptualization, data collection, and review and editing of the manuscript. TJM and JEKH contributed to supervision, resource provision, conceptualization, methodology, and review and editing of the manuscript.

Conflicts of Interest

OAR, JAL, JH, HJ, TTR, MPT, ESV, OES, and TJM are shareholders in Heart2Save, a company that designs electrocardiogram-based medical equipment for arrhythmia detection. JAL, HJ, and OES report personal fees from Heart2Save. No other conflicts of interest have been declared.

References

1. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. Stroke 1991 Aug;22(8):983-988. [doi: 10.1161/01.str.22.8.983] [Medline: 1866765]
2. Steinberg BA, Kim S, Fonarow GC, et al. Drivers of hospitalization for patients with atrial fibrillation: results from the Outcomes Registry for Better Informed Treatment of Atrial Fibrillation (ORBIT-AF). Am Heart J 2014 May;167(5):735-742. [doi: 10.1016/j.ahj.2014.02.003] [Medline: 24766985]
3. Shamloo AS, Dagres N, Müssigbrodt A, et al. Atrial fibrillation and cognitive impairment: new insights and future directions. Heart Lung Circ 2020 Jan;29(1):69-85. [doi: 10.1016/j.hlc.2019.05.185] [Medline: 31262618]
4. Schnabel RB, Michal M, Wilde S, et al. Depression in atrial fibrillation in the general population. PLoS One 2013;8(12):e79109. [doi: 10.1371/journal.pone.0079109] [Medline: 24324579]
5. Piccini JP, Passman R, Turakhia M, Connolly AT, Nabutovsky Y, Varma N. Atrial fibrillation burden, progression, and the risk of death: a case-crossover analysis in patients with cardiac implantable electronic devices. Europace 2019 Mar 1;21(3):404-413. [doi: 10.1093/europace/euy222] [Medline: 30462208]
6. Morillo CA, Banerjee A, Perel P, Wood D, Jouven X. Atrial fibrillation: the current epidemic. J Geriatr Cardiol 2017 Mar;14(3):195-203. [doi: 10.11909/j.issn.1671-5411.2017.03.011] [Medline: 28592963]
7. Yang SY, Huang M, Wang AL, et al. Atrial fibrillation burden and the risk of stroke: a systematic review and dose-response meta-analysis. World J Clin Cases 2022 Jan 21;10(3):939-953. [doi: 10.12998/wjcc.v10.i3.939] [Medline: 35127908]
8. Van Gelder IC, Rienstra M, Bunting KV, et al. 2024 ESC Guidelines for the management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). Eur Heart J 2024 Sep 29;45(36):3314-3414. [doi: 10.1093/eurheartj/ehae176] [Medline: 39210723]
9. Kaplan RM, Koehler J, Ziegler PD, Sarkar S, Zweibel S, Passman RS. Stroke risk as a function of atrial fibrillation duration and CHA2DS2-VASc score. Circulation 2019;140(20):1639-1646. [doi: 10.1161/circulationaha.119.041303]
10. Ferro JM. Cardioembolic stroke: an update. Lancet Neurol 2003 Mar;2(3):177-188. [doi: 10.1016/s1474-4422(03)00324-7] [Medline: 12849239]
11. Brugada J, Katritsis DG, Arbelo E, et al. 2019 ESC guidelines for the management of patients with supraventricular tachycardia: the task force for the management of patients with supraventricular tachycardia of the European Society of Cardiology (ESC): developed in collaboration with the Association for European Paediatric and Congenital Cardiology (AEPC). Eur Heart J 2020;41(5):655-720. [doi: 10.1093/eurheartj/ehz827] [Medline: 31504425]
12. Lindberg T, Wimo A, Elmståhl S, Qiu C, Bohman DM, Sanmartin Berglund J. Prevalence and incidence of atrial fibrillation and other arrhythmias in the general older population: findings from the Swedish national study on aging and care. Gerontol Geriatr Med 2019;5:2333721419859687. [doi: 10.1177/2333721419859687] [Medline: 31276022]
13. Linz D, Gawalko M, Betz K, et al. Atrial fibrillation: epidemiology, screening and digital health. Lancet Reg Health Eur 2024 Feb;37:100786. [doi: 10.1016/j.lanepe.2023.100786] [Medline: 38362546]
14. Zoni-Berisso M, Lercari F, Carazza T, Domenicucci S. Epidemiology of atrial fibrillation: European perspective. Clin Epidemiol 2014;6(1):213-220. [doi: 10.2147/CLEP.S47385] [Medline: 24966695]
15. Hariri E, Hachem A, Sarkis G, Nasr S. Optimal duration of monitoring for atrial fibrillation in cryptogenic stroke: a nonsystematic review. Biomed Res Int 2016;2016:5704963. [doi: 10.1155/2016/5704963] [Medline: 27314027]

XSL•FO
RenderX

16. Santala OE, Lipponen JA, Jäntti H, et al. Necklace-embedded electrocardiogram for the detection and diagnosis of atrial fibrillation. Clin Cardiol 2021 May;44(5):620-626. [doi: 10.1002/clc.23580] [Medline: 33629410]
17. Santala OE, Lipponen JA, Jäntti H, et al. Continuous mHealth patch monitoring for the algorithm-based detection of atrial fibrillation: feasibility and diagnostic accuracy study. JMIR Cardio 2022 Jun 21;6(1):e31230. [doi: 10.2196/31230] [Medline: 35727618]
18. Santala OE, Halonen J, Martikainen S, et al. Automatic mobile health arrhythmia monitoring for the detection of atrial fibrillation: prospective feasibility, accuracy, and user experience study. JMIR Mhealth Uhealth 2021 Oct 22;9(10):e29933. [doi: 10.2196/29933] [Medline: 34677135]
19. Rajakariar K, Koshy AN, Sajeev JK, Nair S, Roberts L, Teh AW. Accuracy of a smartwatch based single-lead electrocardiogram device in detection of atrial fibrillation. Heart 2020 May;106(9):665-670. [doi: 10.1136/heartjnl-2019-316004] [Medline: 31911507]
20. Tieleman RG, Plantinga Y, Rinkes D, et al. Validation and clinical use of a novel diagnostic device for screening of atrial fibrillation. Europace 2014 Sep;16(9):1291-1295. [doi: 10.1093/europace/euu057] [Medline: 24825766]
21. Hartikainen S, Lipponen JA, Hiltunen P, et al. Effectiveness of the chest strap electrocardiogram to detect atrial fibrillation. Am J Cardiol 2019 May 15;123(10):1643-1648. [doi: 10.1016/j.amjcard.2019.02.028] [Medline: 30878151]
22. Eerikäinen LM, Bonomi AG, Schipper F, et al. Comparison between electrocardiogram- and photoplethysmogram-derived features for atrial fibrillation detection in free-living conditions. Physiol Meas 2018 Aug 8;39(8):084001. [doi: 10.1088/1361-6579/aad2c0] [Medline: 29995641]
23. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med 2019 Jan;25(1):65-69. [doi: 10.1038/s41591-018-0268-3] [Medline: 30617320]
24. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310. [Medline: 2868172]
25. Mannhart D, Lischer M, Knecht S, et al. Clinical validation of 5 direct-to-consumer wearable smart devices to detect atrial fibrillation: BASEL wearable study. JACC Clin Electrophysiol 2023 Feb;9(2):232-242. [doi: 10.1016/j.jacep.2022.09.011] [Medline: 36858690]
26. Harmon DM, Sehrawat O, Maanja M, Wight J, Noseworthy PA. Artificial intelligence for the detection and treatment of atrial fibrillation. Arrhythm Electrophysiol Rev 2023;12:e12. [doi: 10.15420/aer.2022.31] [Medline: 37427304]
27. Papalamprakopoulou Z, Stavropoulos D, Moustakidis S, Avgerinos D, Efremidis M, Kampaktsis PN. Artificial intelligence-enabled atrial fibrillation detection using smartwatches: current status and future perspectives. Front Cardiovasc Med 2024;11:1432876. [doi: 10.3389/fcvm.2024.1432876] [Medline: 39077110]
28. Avram R, Ramsis M, Cristal AD, et al. Validation of an algorithm for continuous monitoring of atrial fibrillation using a consumer smartwatch. Heart Rhythm 2021 Sep;18(9):1482-1490. [doi: 10.1016/j.hrthm.2021.03.044] [Medline: 33838317]
29. Boriani G, Glotzer TV, Santini M, et al. Device-detected atrial fibrillation and risk for stroke: an analysis of >10,000 patients from the SOS AF project (Stroke preventiOn Strategies based on Atrial Fibrillation information from implanted devices). Eur Heart J 2014 Feb;35(8):508-516. [doi: 10.1093/eurheartj/eht491] [Medline: 24334432]
30. Petryszyn P, Niewinski P, Staniak A, et al. Effectiveness of screening for atrial fibrillation and its determinants: a meta-analysis. PLoS One 2019;14(3):e0213198. [doi: 10.1371/journal.pone.0213198] [Medline: 30893323]
31. Manetas-Stavrakakis N, Sotiropoulou IM, Paraskevas T, et al. Accuracy of artificial intelligence-based technologies for the diagnosis of atrial fibrillation: a systematic review and meta-analysis. J Clin Med 2023 Oct 17;12(20):6576. [doi: 10.3390/jcm12206576] [Medline: 37892714]
32. Pan Y, Chen E, Jie S, et al. Continuous atrial fibrillation monitoring using a wearable smartwatch: using long-term Holter as reference. Digit Health 2025;11:20552076251314105. [doi: 10.1177/20552076251314105] [Medline: 39866888]
33. De Guio F, Rienstra M, Lillo-Castellano JM, et al. Enhanced detection of atrial fibrillation in single-lead electrocardiograms using a Cloud-based artificial intelligence platform. Heart Rhythm 2025 Jul;22(7):1667-1674. [doi: 10.1016/j.hrthm.2024.12.048] [Medline: 39800092]
34. Zhao X, Zhou R, Ning L, Guo Q, Liang Y, Yang J. Atrial fibrillation detection with single-lead electrocardiogram based on temporal convolutional network–ResNet. Sensors (Basel) 2024;24(2):398. [doi: 10.3390/s24020398]
35. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. Lancet 2019 Sep 7;394(10201):861-867. [doi: 10.1016/S0140-6736(19)31721-0] [Medline: 31378392]
36. Clifford G, Liu C, Moody B, et al. AF classification from a short single lead ECG recording: the physionet computing in cardiology challenge 2017. In: 2017 Computing in Cardiology Conference (CinC): IEEE; 2017:1-4. [doi: 10.22489/CinC.2017.065-469]
37. Hennings E, Coslovsky M, Paladini RE, et al. Assessment of the atrial fibrillation burden in Holter electrocardiogram recordings using artificial intelligence. Cardiovasc Digit Health J 2023 Jan 27;4(2):41-47. [doi: 10.1016/j.cvdhj.2023.01.003] [Medline: 37101946]

XSL•FO
RenderX

Abbreviations

AF: atrial fibrillation
AFL: atrial flutter
AI: artificial intelligence
CHA_2 DS_2 -VA score: congestive heart failure, hypertension, age ≥75 years, diabetes, prior stroke or transient ischemic attack, vascular disease, age 65-74 years
DNN: deep neural network
ECG: electrocardiogram
ICC: intraclass correlation coefficient
KUH: Kuopio University Hospital
NPV: negative predictive value
PPV: positive predictive value
SR: sinus rhythm

Edited by SD Bhattacharjee; submitted 20.Aug.2025; peer-reviewed by P Taiwo, S Suzuki; revised version received 26.Jan.2026; accepted 26.Feb.2026; published 23.Apr.2026.

Please cite as:
Rantula OA, Lipponen JA, Halonen J, Jäntti H, Rissanen TT, Tarvainen MP, Naukkarinen NS, Seljola ES, Santala OE, Sedha J, Martikainen TJ, Hartikainen JEK
A Deep Neural Network for Interpreting Wearable Electrocardiogram Data in Atrial Fibrillation: Prospective Observational Diagnostic Accuracy Study
JMIR Mhealth Uhealth 2026;14:e82475
URL: https://mhealth.jmir.org/2026/1/e82475
doi:10.2196/82475

© Olli A Rantula, Jukka A Lipponen, Jari Halonen, Helena Jäntti, Tuomas T Rissanen, Mika P Tarvainen, Noora S Naukkarinen, Eemu-Samuli Seljola, Onni E Santala, Jagdeep Sedha, Tero J Martikainen, Juha E K Hartikainen. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 23.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Definition and Test-Retest Reliability of a Monitoring Method Integrating Accelerometric Actigraphy and Bluetooth Indoor Location Tracking Applied in a Long-Term Residential Unit for Persons With Dementia: Longitudinal Observational Study

Marco Rabuffetti[1], DrEng; Pietro Davide Trimarchi[1], PhD; Alessia Gallucci[1], PhD; Ilaria Carpinella[1], DrEng; Elena Kisel[1], MD; Maria Patrizia Andriani[1], MSc; Ennio De Giovannini[2], MD; Gaia Bailo[1], DrEng; Fabrizio Giunco[1], MD; Maurizio Ferrarin[1], PhD

[1]IRCCS (Istituto di Ricovero e Cura a Carattere Scientifico) Fondazione Don Carlo Gnocchi Ente del Terzo Settore, via Capecelatro, 66, Milano, Italy
[2]Centro Medico Riabilita Cooperativa Sociale Mano Amica Onlus, Schio, Italy

Corresponding Author:
Alessia Gallucci, PhD
IRCCS (Istituto di Ricovero e Cura a Carattere Scientifico) Fondazione Don Carlo Gnocchi Ente del Terzo Settore, via Capecelatro, 66, Milano, Italy

Abstract

Background: Dementia has an impact on the physical activities performed daily in a social context. Sleeping and resting, in general, are also affected by dementia. Monitoring techniques based on miniaturized wearable sensors and on sensorized environments allow for actigraphic recordings and location tracking. The availability of contemporaneous physical activities profile led to quantify, in the social actigraphy approach, the level of correlation between individuals living in the same environment.

Objective: This study has two main objectives: (1) to define a methodology for actigraphic recordings, based on wearable accelerometers, and on location tracking, based on Bluetooth wearable technology, and to apply it in a well-defined social context, a long-term care residential unit for people with dementia; and (2) to quantify test-retest reliability of the indices obtained by the monitoring methodology.

Methods: Persons with dementia living in the long-term care unit have been equipped with miniaturized wearable sensors, an accelerometer at their dominant wrist, and a Bluetooth beacon at their ankle for 7 days. The raw recordings allowed for computing indices related to physical activity intensities, to the occurrence of walking bouts, to the efficiency of sleep and waking phases, to social interactions between individuals, and to locations preferably occupied. The 7-day session was repeated at short (3 weeks) and long (3 months) terms in order to quantify the test-retest reliability of the indices.

Results: Twenty-five persons with dementia were enrolled, 4 of them dropped out, and valid data were obtained, in the different sessions, from 19 to 21 individuals of the recruited group. Control data from 10 age-matched healthy participants were derived from published datasets. As a group, compared with age-matched healthy participants, persons with dementia showed a comparable duration of phases of no activity and of light activity (energy cost lower than 3 metabolic equivalents of tasks [METs]), a relevantly lower duration (−84.3%) of phases of moderate activity (energy cost ranging from 3 to 6 METs), and substantial absence (−100%) of phases of vigorous activity (larger than 6 METs); moreover, daytime and nighttime were characterized by comparable wake and sleep, respectively, efficiency; finally, as to the social interactions, persons with dementia showed a lower correlation of their motor activity profiles (−53.1%). The test-retest reliability was excellent for physical activity indices (intraclass correlation coefficients ranging from 0.76 to 0.98), good for social indices (0.65 - 0.67), excellent for sleep or wake efficiency (0.74 - 0.89), and fair for location tracking indices (0.37 - 0.78).

Conclusions: The considered methodology, particularly concerning accelerometry, proved to be feasible, informative, and with a good to excellent test-retest reliability. Interestingly, the methodology clearly identified behaviors, such as wandering, in a minority of individuals inside this study's group of persons with dementia, thus supporting a possible clinical use of the methodology.

(JMIR Mhealth Uhealth 2026;14:e70188) doi:10.2196/70188

KEYWORDS

dementia; long-term residency unit; physical activity; location tracking; wearable sensor; social actigraphy

XSL•FO
RenderX

Introduction

The number of people with dementia is estimated to increase from 57.4 million cases worldwide in 2019 to 152.8 million cases in 2050, with the female-to-male ratio of 1.69 in 2019 being substantially stable and the projected increases in cases largely attributed to population growth and population aging [1].

One of the most relevant aspects in the management of individuals with dementia, even in mild-to-moderate stages, is related to movement [2]. The transition from a life without illness to a life dealing with alterations in spatial-temporal orientation or critical judgment and perception of dangers requires a delicate balance between the freedom the person still demands (moving around the house, walking, or going out) and the dangers they are exposed to, such as the risk of falls, disorientation, and domestic and road accidents. For these reasons, both in domestic and residential settings, there is a significant use of physical or pharmacological restraint, which sometimes may also be improper or untimely [3-5]. Alternatively, or as an integration to such restraints, there are environmental and behavioral management strategies, involving architectural or environmental adaptations that reduce risks and dangers, and/or facilitation, validation, and orientation support strategies [6]. Accordingly, special Alzheimer Units ("Nuclei Alzheimer"), long-term residential wards specific for persons with dementia, introduced by the Italian region Lombardy legislation in the mid nineties [7], are specifically designed and organized to promote free movement and activity in a safe and supervised environment, and where specific interventions may be carried out with safety and efficiency.

In order to support supervision of motor behaviors [8], particularly in nursing settings, easily manageable methodologies can help the caregivers. Current technologies based on wearable accelerometers support activity tracking [9,10]. Interestingly, these approaches may support group studies which provide information about the interplay among persons, thus implementing the so-defined social actigraphy [11]. Additionally, systems based on wearable Bluetooth beacons and antennas [12], similar to the contact tracing methods adopted during the COVID-19 pandemic [13], can support indoor location tracking in order to identify the room or environment the participant is currently in [14-16]. Such possibilities of using continuous measures of the quantity and quality of movement and tracking of location represent a relevant contribution to assessments based on tests, surveys, and questionnaires [17]. The integration of an activity tracking system and a location tracking system has already been considered in order to compute activity-related indices specific to indoor location. This approach proved to be able to better distinguish between at-risk patients who can gain independence vs the patients who are rehospitalized [18], and, when applied to a longitudinal study, provided information concerning patients receiving rehabilitation at a nursing facility [19].

When considering instrumented monitoring of activity and location, it is important to underline the inherent variability associated with free behavior, which requires us to monitor human behavior for a relevant time duration, starting from a single day [20], the basic circadian cycle, up to 1 week [11] and more [21], to obtain reliable and valid outcome indices converging to stable values [22,23]. Still, outcome indices show interindividual and intraindividual variability that would require proper experiments to be characterized in terms of test-retest reliability. Such variability thresholds allow for the identification of only significant deviations that may be predictive of critical situations, such as abnormal movement patterns or deviations from usual behaviors, suggestive of a changed, eventually worsened, condition. Intraindividual test-retest variability has already been considered for physical activity questionnaires on persons with dementia [24-26] and for actigraphic outcome indices related to movement [27-29].

The objective of the present study is to define a composite method for actigraphic monitoring [11,30] and indoor location tracking in a community residence and to perform an experimental campaign in an Alzheimer Unit, a long-term residential care facility for persons with dementia, to quantify the test-retest reliability of the resulting indices.

Methods

Ethical Considerations

This study and related documents were approved by the local Ethics Committee (Comitato Etico IRCCS [Istituto di Ricovero e Cura a Carattere Scientifico] Regione Lombardia, Sezione IRCCS Fondazione Don Carlo Gnocchi, ID 07_07/04/2021, April 7, 2021, amended on May 26, 2023). Due to severe dementia, informed consents were obtained from legal tutors of the participants. Any data or information was recorded in an anonymized form. No compensation was provided to participants.

Participants

Participants were recruited from the Alzheimer Unit of the Palazzolo Institute, Don Carlo Gnocchi Foundation, Milan. As it is a pilot study, no minimum sample size has been identified, but all resident guests of the experimental site were contacted for possible enrollment.

The inclusion criteria were an autonomous locomotor capacity (even with the help of walking aids such as, for example, a cane or a crutch) and the frequentation of common areas of the residential unit. The exclusion criteria were the presence of behavioral disorders incompatible with the research objectives (delirium, psychomotor agitation requiring psychoactive pharmacological interventions, acute clinical events limiting autonomous movement, and nonneurological pathologies strongly limiting motor activities), the need for assistance for walking by a caregiver, any dermatological problems preventing contact with wearable sensors, any confined conditions or quarantine for COVID-19 risk management.

Reference data from age-matched healthy participants were extracted from previously published datasets [11,30] in order to build a control group with a comparable average age.

XSL•FO
RenderX

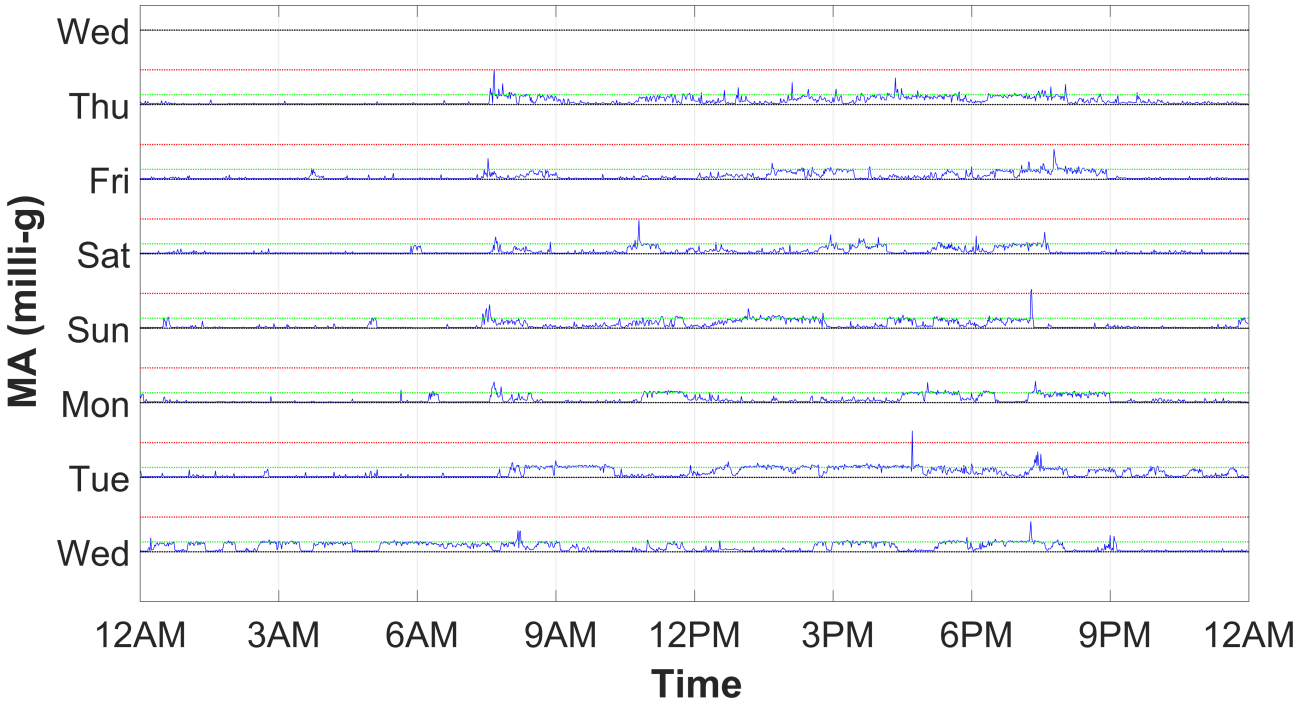
Clinical and Functional Assessment

The chosen assessment battery included Hendrich II Fall Risk Model [31], Tinetti Scale [32] when applicable, Barthel Index [33], Mini-Mental State Examination [34], Clinical Dementia Rating [35], and Scheda di Osservazione Intermedia di Assistenza [36]. Moreover, an Event Monitoring Diary (clinical events, changes in pharmacological and nonpharmacological interventions, falls, discomfort, and/or incongruous interaction with sensors, and temporarily leaving the ward) was compiled during the experiment.

Actigraphic Monitoring

Participants wore, at their dominant wrist, a wristwatch-like waterproof device including a triaxial accelerometer for 1 week. Accelerometric raw data (measurement unit is one thousandth of earth gravity constant g, termed in the following as milli-g), sampled at 50 Hz, were stored in a resident memory (Geneactiv, Activinsights). The device was not to be removed unless the participant was willing to. The 3 acceleration components allowed for the computation of the vector norm, from which the motor activity (MA) index was derived as the SD of the vector norm in 1-minute epochs [37], thus resulting in a 7-day MA profile composed of 1440 values per day (Figure 1).

Figure 1. Seven-day profile of the MA index of a guest of the long-term care residential unit. Y-grid dotted lines in green and red mark, respectively, values 100 (transition value from LPA to MPA) and 350 (transition value from MPA to VPA) milli-g above black dotted baselines. LPA: light physical activity; MA: motor activity; milli-g: measurement unit is one thousandth of Earth gravity constant g; MPA: moderate physical activity; VPA: vigorous physical activity.

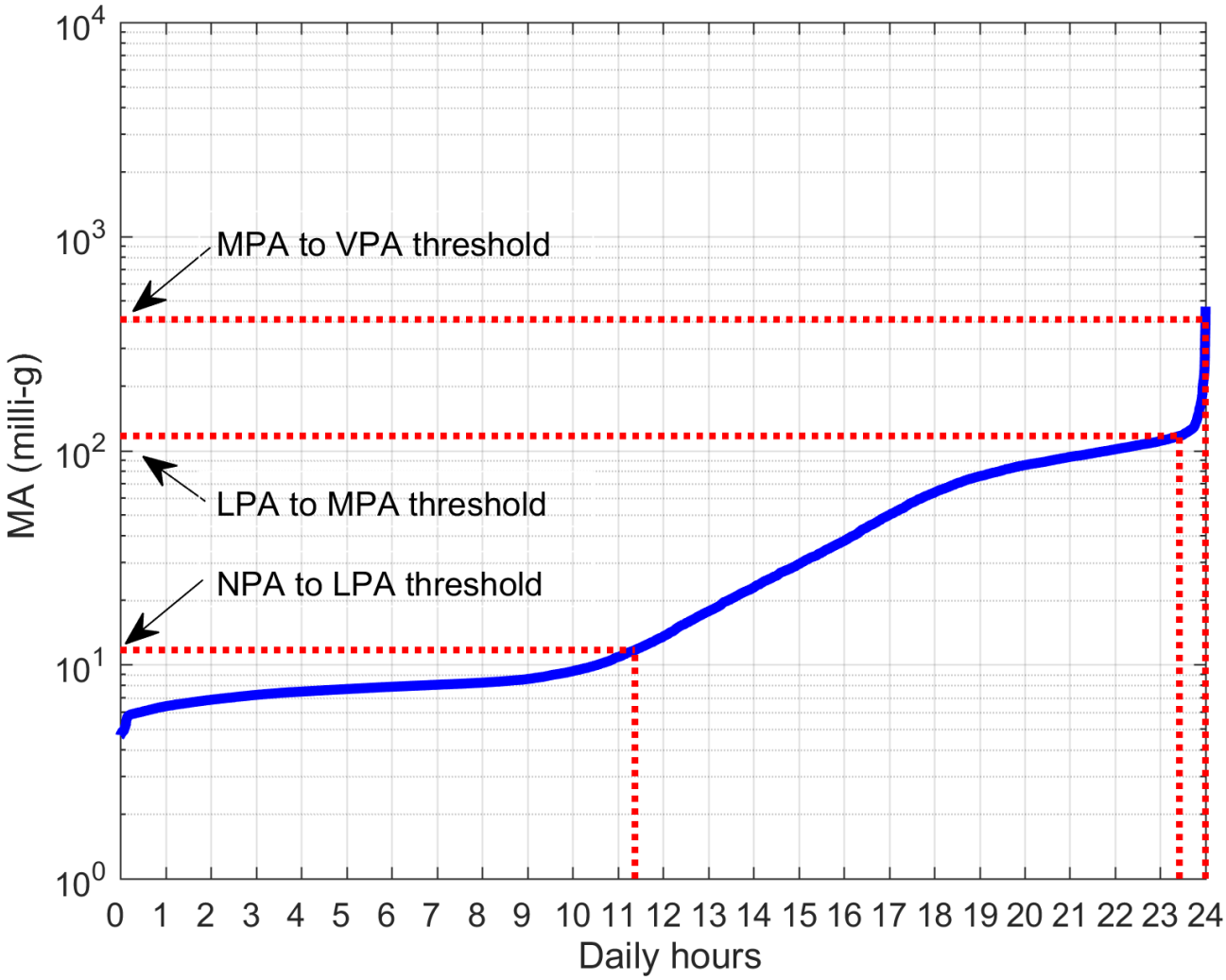


A sorting operation on the MA profile produced a distribution plot (Figure 2) which, through proper thresholds (10 milli-g for transition from no to light physical activity [LPA], 100 milli-g for light to moderate, and 350 milli-g for moderate to vigorous, for a sensor located on the nondominant wrist) [38], eventually corrected for dominant side (previous threshold values multiplied by $Coeff_{Dominant}$=1.16926), allowed computation of physical activity indices: the percentage duration of the phases with no physical activity (NPA), LPA, moderate physical activity (MPA), and vigorous physical activity (VPA).

An additional analysis was performed on the raw acceleration profile to detect the occurrence of regular patterns that are likely to be associated with a pseudoperiodic function such as locomotion. The raw signal was windowed in 4-second epochs, and an autocorrelation analysis was applied to the signal window. The possible occurrence of lagged peaks of the autocorrelation function was searched, adopting a threshold of 0.7 (a perfect periodic phenomenon has a unitary peak value, while white noise is denoted by null values of the autocorrelation function), which had been set, based on trial-and-error attempts [39]. The duration of such pseudoperiodic behavior (PPB) phases was quantified by their percentage duration (PPB).

XSL•FO
RenderX

Figure 2. The distribution of the MA values measured for 7 days is presented in Figure 1. The predefined MA thresholds (identified by values on the y-axis) allow for quantifying the duration (as an interval on the x-axis) on a daily basis of the phases of NPA, LPA, MPA, and VPA (this latter one is almost absent in the shown example as well as in the whole dataset). LPA: light physical activity; MA: motor activity; MPA: moderate physical activity; NPA: no physical activity; VPA: vigorous physical activity.



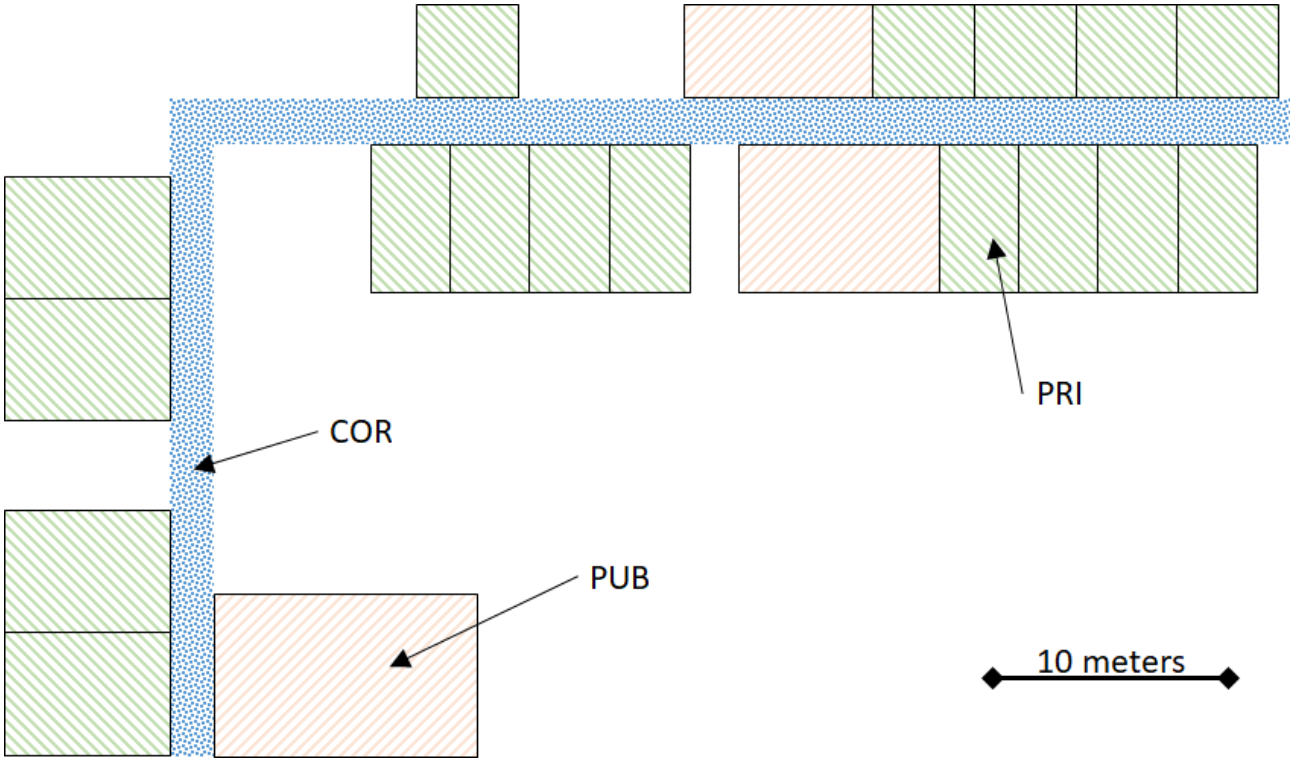
Two more indices were computed to quantify the quality of nighttime sleep and daytime wake time: nighttime sleep efficiency (SEff), the percentage duration of NPA phases between 11 PM and 7 AM, the a priori "night" time interval, which is related to SEff [40]; analogously, daytime wake efficiency (WEff), the percentage duration of LPA, MPA, or VPA (ie, the complementary duration of no activity phases) between 9 AM and 7 PM, the a priori "day" time interval, which quantifies WEff. The choice of the strict timing for "day" and "night" time intervals was based on the daily routines established within the long-term care unit: all guests are expected to be in bed during the "night" interval and awake during the "day" interval, while this assumption is not true for transition intervals (7 AM-9 AM and 7 PM-11 PM).

Location Tracking

The chosen location tracking system is composed of individual wearable devices and a network of environment fixed-position antennas (TapMyLife Inc). Monitored participants wore a miniaturized Bluetooth beacon (dimensions 32 mm × 23 mm × 3.3 mm, weight 4.0 grams) at one of their ankles by means of a biocompatible plastic bracelet. Such a device continuously emits a unique ID code, possibly received by a Bluetooth receiving antenna positioned nearby (according to the Bluetooth technology, approximately up to 10 meters in open space). All environments where participants resided (private bedrooms and any common spaces they spend time in, such as dining rooms, living rooms, corridors, etc) were equipped with Bluetooth receiving antennas: 1 antenna for each private room (PRI; 13 double rooms and 4 single rooms), 2 antennas for each large public room (3 PUBs), and 7 antennas along the main corridor, as shown in Figure 3. The frequency of the Bluetooth emitter is partly determined by the asynchronous Bluetooth protocol, generally in the magnitude order of 1 Hz. As a single Bluetooth emission can be received by several antennas, a proprietary signal algorithm identified the closest antenna and updated the beacon location (ie, the identified receiving antenna position) only when the beacon switched to a new antenna. Therefore, the source raw data provided by the location tracking system was natively asynchronous, and the current location of a beacon was determined by the position of the last antenna that detected the beacon itself. In order to summarize the location data, the percentage values of time spent in PRIs, common living rooms (PUBs), and corridors were computed.

XSL•FO
RenderX

Figure 3. Schematic map of the long-term care residential unit where the monitoring experiments took place. Different colors or patterns allow for identification of PRIs, PUBs, where social daily activities take place, and corridors. COR: corridor; PRI: private room; PUB: public room.



Experimental Sessions

The single monitoring experiment lasted for 7 days. Three experimental sessions were performed in July (Session A), September (Session B), and October (Session C) 2023. For short-term repeatability assessment (3-week time interval), Sessions B and C were compared. For long-term repeatability assessment (3-month time interval), the Sessions A and C were considered.

Outcome Variables

The variables computed starting from the actigraphic recordings were:

- NPA, LPA, MPA, and VPA: the percentage duration of NPA (1 metabolic equivalent of task [MET], MA≤10), LPA (1 - 3 MET, 10<MA≤100), MPA (3 - 6 MET, 100<MA≤350), and VPA (>6 MET, MA>350) [38], where MET stands for metabolic equivalent of task, a standardized unit of measurement for the energy cost of physical activity, being 1 MET equal to the metabolic energy expended while sitting at rest;
- PPB: the percentage duration of pseudoperiodic behavior, whose occurrence was identified in 4-second epochs of the acceleration norm in which the autocorrelation coefficient was larger than 0.7 [39];
- CC: the cross-correlation coefficient between contemporaneous physical activity MA profiles in couples of participants [11], its unit is a real value from 0 (absence of cross-correlated physical activities) to 1 (perfectly synchronous physical activity);
- SEff: nighttime sleep efficiency, the percentage of time between 11 PM and 7 AM in which MA was classified as NPA [40];
- WEff: daytime wake efficiency, the percentage of time between 9 AM and 7 PM in which MA was classified as LPA, MPA, or VPA;
- PRI, PUB, and corridor: the percentage duration of time spent in PRIs, that is, bedrooms (PRI), in PUBs, that is, social spaces where guests of the long-term care residential unit had their meals and performed leisure activities (PUB), and in public spaces, that is, main corridors.

Indices were computed on the 7-day-long recording. When the full-length recordings included a portion of scratch recording for any reason, for example, because of participants taking off the sensor, the scratch portion was removed from the recording, and the remaining time duration was further reduced to the highest possible multiple of 24 hours. The previous algorithm allowed for the computation of the indices on recordings whose length was a multiple of the basic 24-hour circadian cycle.

Statistical Analyses

The difference between outcome variables as obtained from control data and from pooled experimental data was quantified as percentage variation of median values and presented as boxplots. The difference was assessed by the Wilcoxon rank sum test (as implemented by “ranksum” function in MATLAB software, version R2016a, The MathWorks Inc), a nonparametric test for 2 independent populations.

Test-retest reliability was presented as scatterplots and assessed through the intraclass correlation coefficient (ICC). ICC was computed with a 95% CI, using a 2-way mixed-effect, absolute agreement, single measurement model. Significance level was set at 0.05. The significance level was corrected for multiple comparisons according to Bonferroni-Holm.

XSL•FO
RenderX

The adopted guidelines for interpreting ICC [41] meant “poor” for values less than 0.40, “fair” for values between 0.40 and 0.59, “good” for values between 0.60 and 0.74, and “excellent” for values between 0.75 and 1.00.

Results

The setup of technical infrastructure (BLE antennas) took place at the beginning of 2023. The data collection phase lasted from June to October 2023. The 30 guests of the long-term care residential unit were considered for this study. It was not obtained or was not possible to obtain informed consent from 5 persons, who were therefore not recruited for this study. A total of 25 guests of the long-term care residential unit provided informed consent, through their legal tutors, to participate in this study. Their demographic data and clinical scores are summarized in Table 1. Noticeably, some clinical tests were not administered (marked as “n.adm.” in the table) to some persons because of their inability to receive them. At the time of the first experiment, 4 participants denied their participation by refusing to be equipped with the sensors. Therefore, the experimental group included 21 participants (15 women and 6 males); their age was 83.4 years on average, SD 10.1 years, ranging from 69 to 97 years plus 1 younger outlier aged 55 years. Four participants rejected the wrist sensor after a variable duration from 2 days to 6 days (but the collected data lasted more than 24 hours, thus they were allowed to be included in the analysis). At the time of the second experiment, 2 more participants denied their participation and dropped out, and a third one rejected the wrist sensor after 2 days of data collection. At the time of the third experiment, 2 participants rejected the wrist sensor after 1 or 2 days, and 1 participant removed the wrist sensor, and it was lost. Therefore, the dataset for the analysis included recordings from 21 participants for the location tracking, while the actigraphic assessment included 21 participants in the first experiment, 19 participants in the second experiment, and 18 participants in the third experiment. All the observed sensors’ rejections were clearly expressed by the participants with signs of discomfort about that “alien” presence at their wrist or ankle. Though it was not possible to clarify with them the reason why, while possible explanations, such as a tight strap or bracelet, were excluded by the experimenter’s supervision. Therefore, the rejections were considered a behavioral aspect of the dementia condition.

XSL•FO
RenderX

Table . Demographic data and clinical scores of this study's participants.

Subject ID	Sex (F[a]/M[b])	Age (y)	CDR[c] [35]	MMSE[d] [34]	Barthel [33]	SOSIA[e] [36]	Tinetti et al [32]	Hendrich II Fall Risk [31]
S01[f]	F	84	3	3	44	7	22	6
S02	M	69	3	N.adm.[g]	32	5	22	7
S03[f]	F	79	3	N.adm.	55	7	26	6
S04	F	89	3	16	57	5	21	6
S05	M	88	3	N.adm.	59	7	21	7
S06	F	85	3	N.adm.	34	7	23	5
S07	F	91	3	N.adm.	37	5	21	5
S08	F	82	3	N.adm.	40	7	23	6
S09	F	80	3	N.adm.	33	5	23	5
S10	F	97	3	9.4	73	7	23	5
S11	F	87	3	11.2	60	7	21	6
S12	F	79	3	N.adm.	26	5	N.adm.	5
S13	M	87	3	N.adm.	29	5	N.adm.	7
S14	F	88	3	7	61	7	21	6
S15	F	94	2	13	59	7	24	6
S16[f]	M	63	3	N.adm.	44	5	22	7
S17[f]	F	90	3	N.adm.	56	5	22	5
S18	M	69	3	N.adm.	20	5	N.adm.	6
S19	F	84	3	7	68	7	23	6
S20	M	75	3	N.adm.	26	5	20	7
S21	F	85	3	16	61	5	22	6
S22	F	92	3	N.adm.	34	7	20	7
S23	F	55	3	N.adm.	31	5	19	6
S24	F	96	3	N.adm.	26	5	N.adm.	8
S25	M	80	3	3	34	5	23	7

[a]F: female.

[b]M: male.

[c]CDR: Clinical Dementia Rating.

[d]MMSE: Mini Mental State Examination.

[e]SOSIA: Scheda di Osservazione Intermedia di Assistenza.

[f]Indicates participants dropped out before Session A experiment.

[g]N.adm.: not administered.

No major health episode possibly influencing mobility was observed (and recorded in the Event Monitoring Diary) among the participants during the 3-month span in which experiments took place.

The control group included 10 healthy persons living in their own homes, 5 males and 5 females, aged 82.1 years on average (SD 2.1 years), ranging from 80 to 87 years, and were extracted from the dataset of a previously published study [11,30]. For the interaction outcomes (CC index), reference data from 40 healthy individuals (20 females and 20 males), whose mean age was 72 years, SD 7.6 years, ranging from 58 to 87 years) from another study [11] were considered.

The summary of all results is presented in Table 2.

XSL•FO
RenderX

Table . Summary statistics of resulting indices.

Index	Controls median (IQR)	Cases session A median (IQR)	Cases session B median (IQR)	Cases session C median (IQR)	Wilcoxon raw *P* value (after Holm correction)	Short-term ICC[a] median (95% range)	Long-term ICC median (95% range)
NPA[b]	37.0 (7.0)	50.4 (21.9)	44.1 (16.9)	48.8 (17.7)	.01 (n.s.[c])	0.92 (0.81 - 0.97)	0.88 (0.71 - 0.95)
LPA[d]	37.9 (4.1)	38.6 (8.8)	41.1 (17.7)	40.2 (16.4)	.20 (n.s.)	0.86 (0.66 - 0.95)	0.76 (0.45 - 0.90)
MPA[e]	23.3 (8.4)	2.7 (8.0)	3.7 (0.3)	3.9 (11.9)	<.001 (<.05)	0.96 (0.90 - 0.98)	0.84 (0.62 - 0.94)
VPA[f]	0.2 (0.5)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	<.001 (<.05)	N.c.[g]	N.c.
PPB[h]	0.8 (1.6)	0.3 (3.2)	0.7 (5.9)	1.0 (4.1)	.68 (n.s.)	0.83 (0.59 - 0.93)	0.95 (0.87 - 0.98)
CC[i]	0.28 (0.08)	0.14 (0.13)	0.14 (0.17)	0.13 (0.11)	<.001 (<.05)	0.67 (0.58 - 0.75)	0.65 (0.55 - 0.73)
SEff[j]	82.8 (4.7)	75.0 (24.6)	74.7 (29.2)	77.1 (25.4)	.09 (n.s.)	0.89 (0.73 - 0.96)	0.78 (0.50 - 0.91)
WEff[k]	88.9 (10.7)	70.3 (28.1)	78.3 (22.1)	75.0 (15.3)	.03 (n.s.)	0.74 (0.42 - 0.89)	0.81 (0.57 - 0.93)
PRI[l]	N/A[m]	59.5 (5.2)	52.9 (10.0)	56.0 (6.7)	N/A	0.37 (0.07 - 0.68)	0.48 (0.06 - 0.75)
PUB[n]	N/A	35.3 (7.5)	29.6 (10.9)	29.8 (14.0)	N/A	0.78 (0.54 - 0.91)	0.48 (0.06 - 076)
COR[o]	N/A	5.3 (12.1)	16.7 (15.3)	17.0 (19.9)	N/A	0.59 (0.22 - 0.81)	0.47 (0.05 - 0.75)

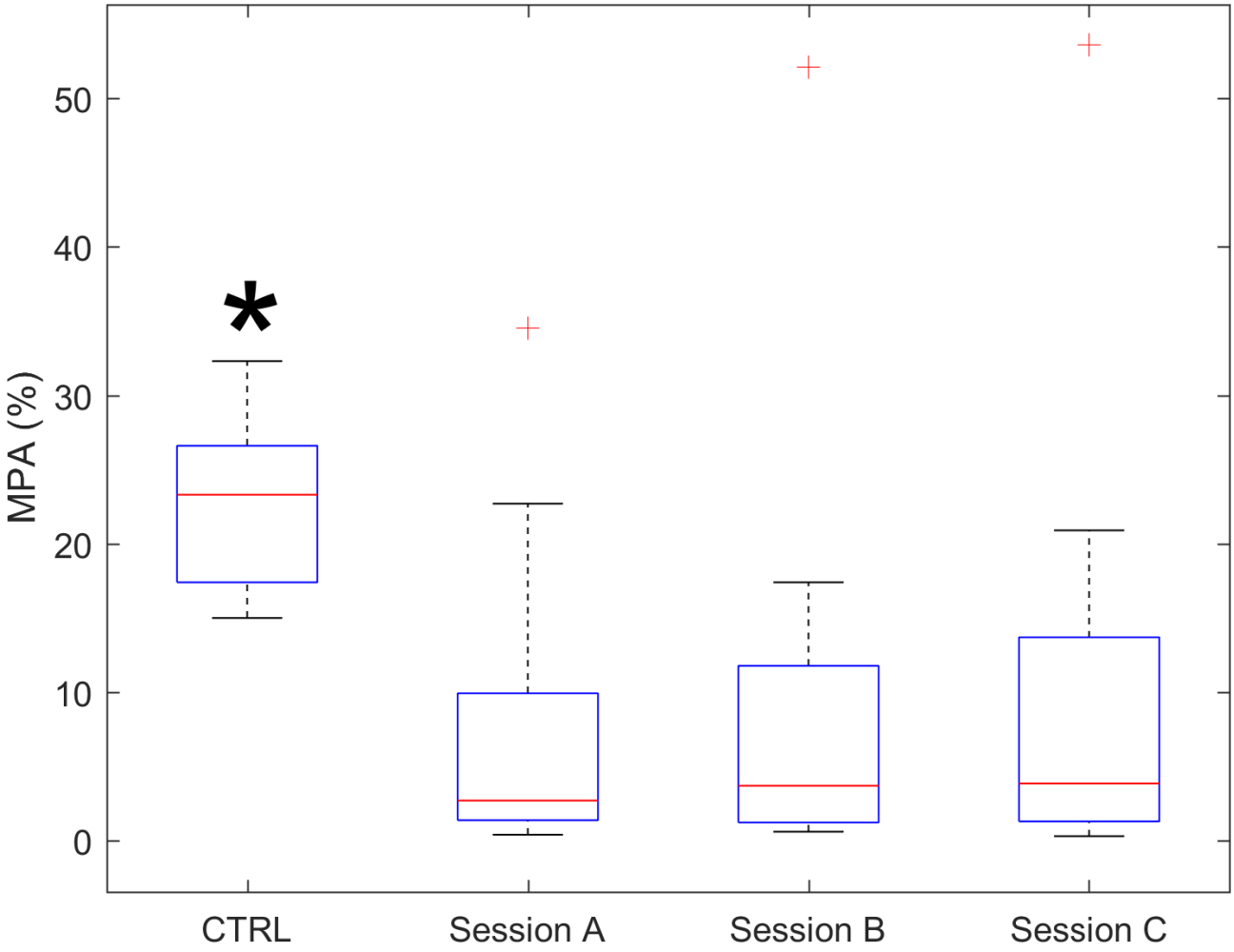
[a]ICC: intraclass correlation coefficient.
[b]NPA: no physical activity.
[c]n.s.: not significant.
[d]LPA: light physical activity.
[e]MPA: moderate physical activity.
[f]VPA: vigorous physical activity.
[g]N.c.: not computed.
[h]PPB: pseudoperiodic behavior.
[i]CC: cross-correlation coefficient.
[j]SEff: sleep efficiency.
[k]WEff: wake efficiency.
[l]PRI: private room.
[m]N/A: not available.
[n]PUB: public room.
[o]COR: corridor.

The analytical presentation of selected indices' values is provided in the form of a boxplot, which also displays any potential outliers. The ICC values are interpreted as "excellent" when larger than 0.75, "good" when between 0.60 and 0.74, "fair" when between 0.40 and 0.59, and "poor" otherwise.

As to indices related to levels of physical activity (NPA, LPA, MPA, and VPA), the experimental group, compared to control data, showed significantly lower (−84.3%) moderate activity (MPA, Figure 4) and absent vigorous activity, while lightly longer durations of NPA and light activity (LPA) were not significantly different from control data. Both short-term and long-term ICC of indices related to levels of physical activity evidenced excellent test-retest reliability.

XSL•FO
RenderX

Figure 4. MPA index for CTRL and for the 3 sessions of the experimental group. Red "+" marks indicate outlier individuals. The asterisk indicates that control data have significantly (P<.05) higher values than the experimental group. CTRL: control data; MPA: moderate physical activity.

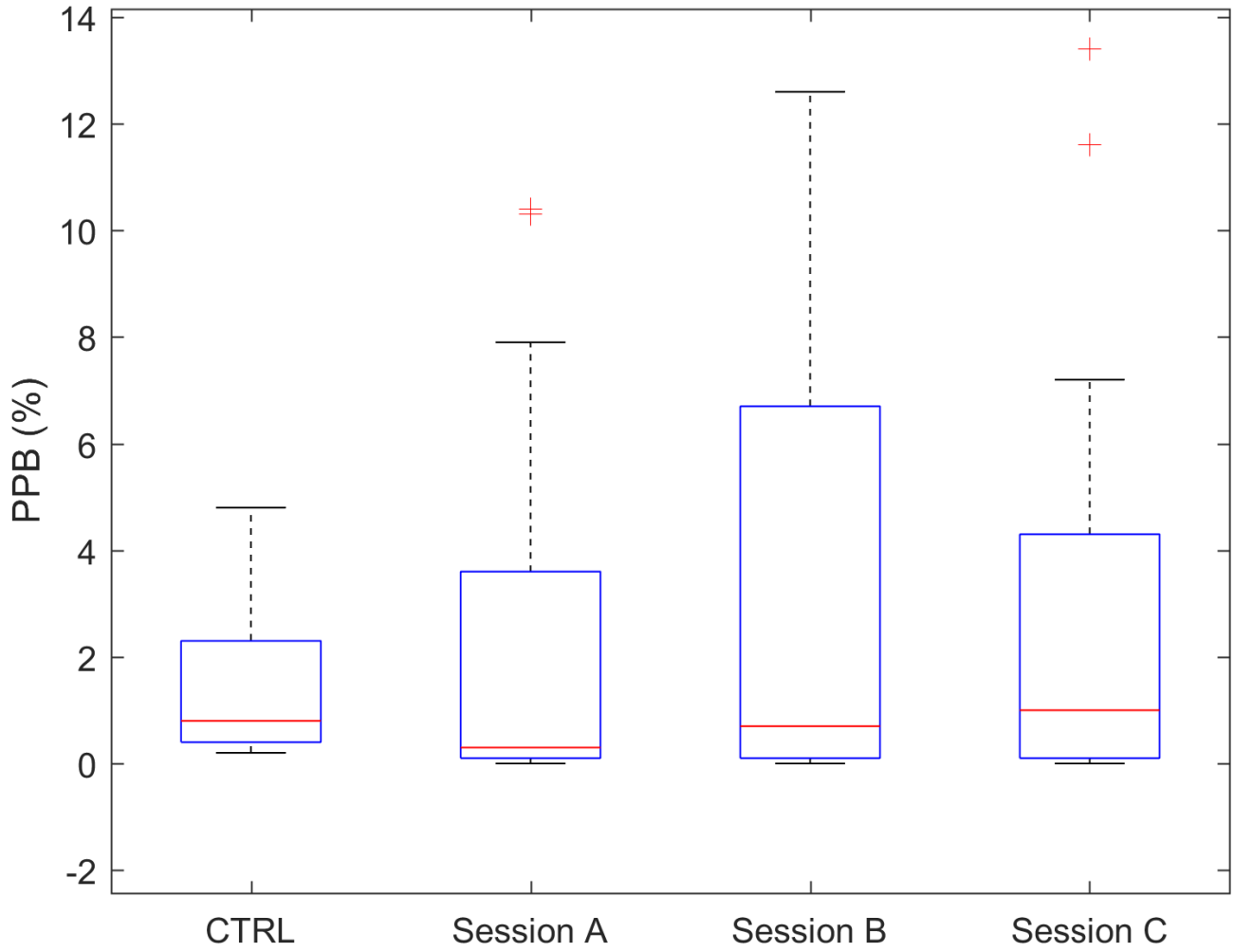


The PPB index, related to the percentage of time spent in steady walking, did not show a statistically significant difference between the control and experimental group, though in the latter group, the median values of the index were reduced by −37.5%. A perusal of the index values' distributions (Figure 5) lets us also detect the presence of a large number of cases showing larger values, and sometimes extremely larger values, than normal PPB values, in contrast with the average trend. The latter observation accounts for the so-called "wandering" behavior adopted by a subgroup of cases, who spent most of their time walking along the unit's corridor. The test-retest reliability of the PPB index was excellent.

The CC index, quantifying the accordance between the activity profiles of different participants, is compared with control data obtained from healthy age-matched individuals not living together (retrieved from the study by Rabuffetti et al [11]). The CC values (Figure 6) of the experimental group were significantly smaller, about half of the control values' (−53.1%), and demonstrated a good test-retest reliability.

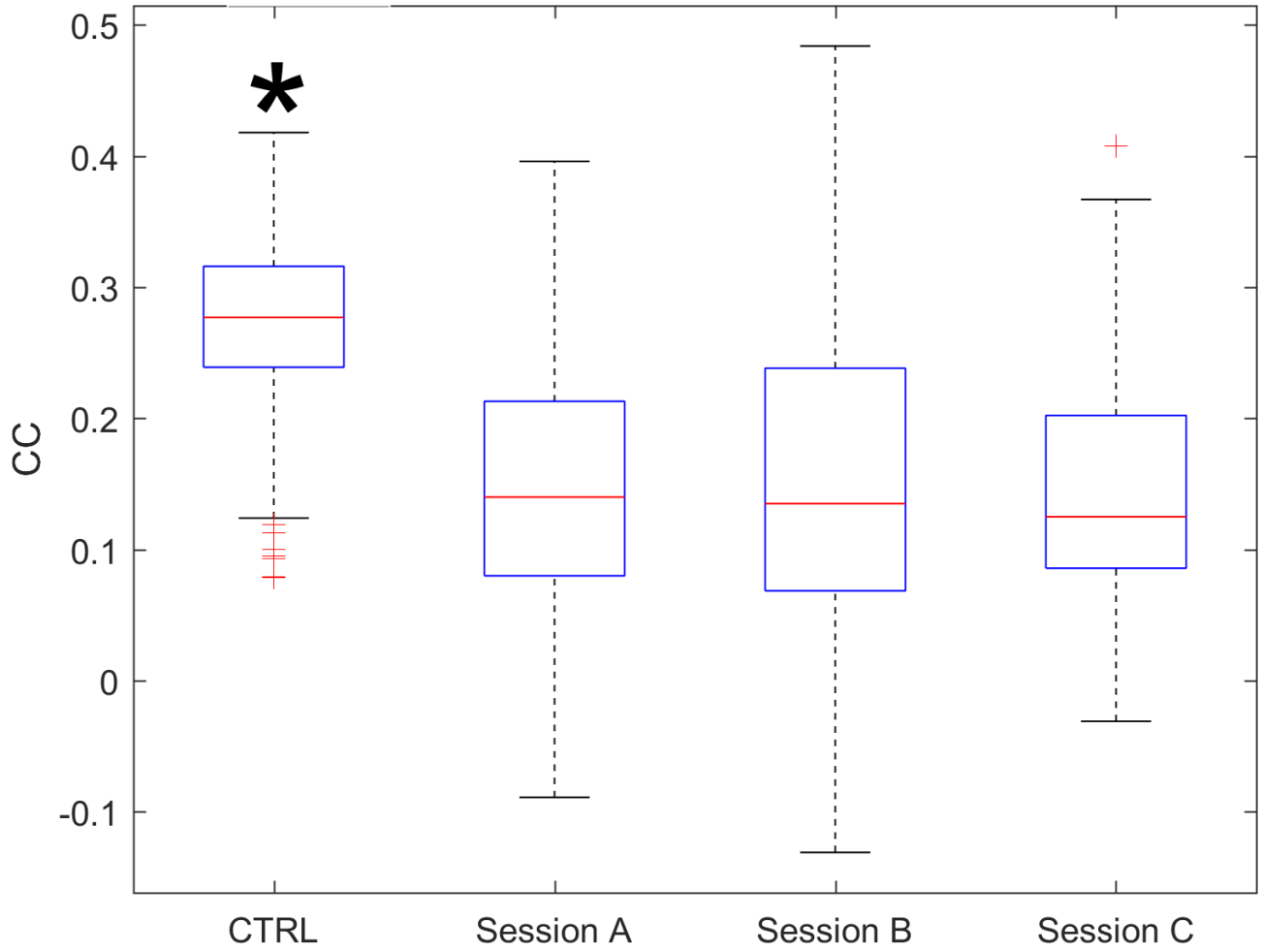
XSL•FO
RenderX

Figure 5. PPB index for CTRL and for the 3 sessions of the experimental group. Red "+" marks indicate outlier individuals. CTRL: control data; PPB: pseudoperiodic behavior.



XSL•FO
RenderX

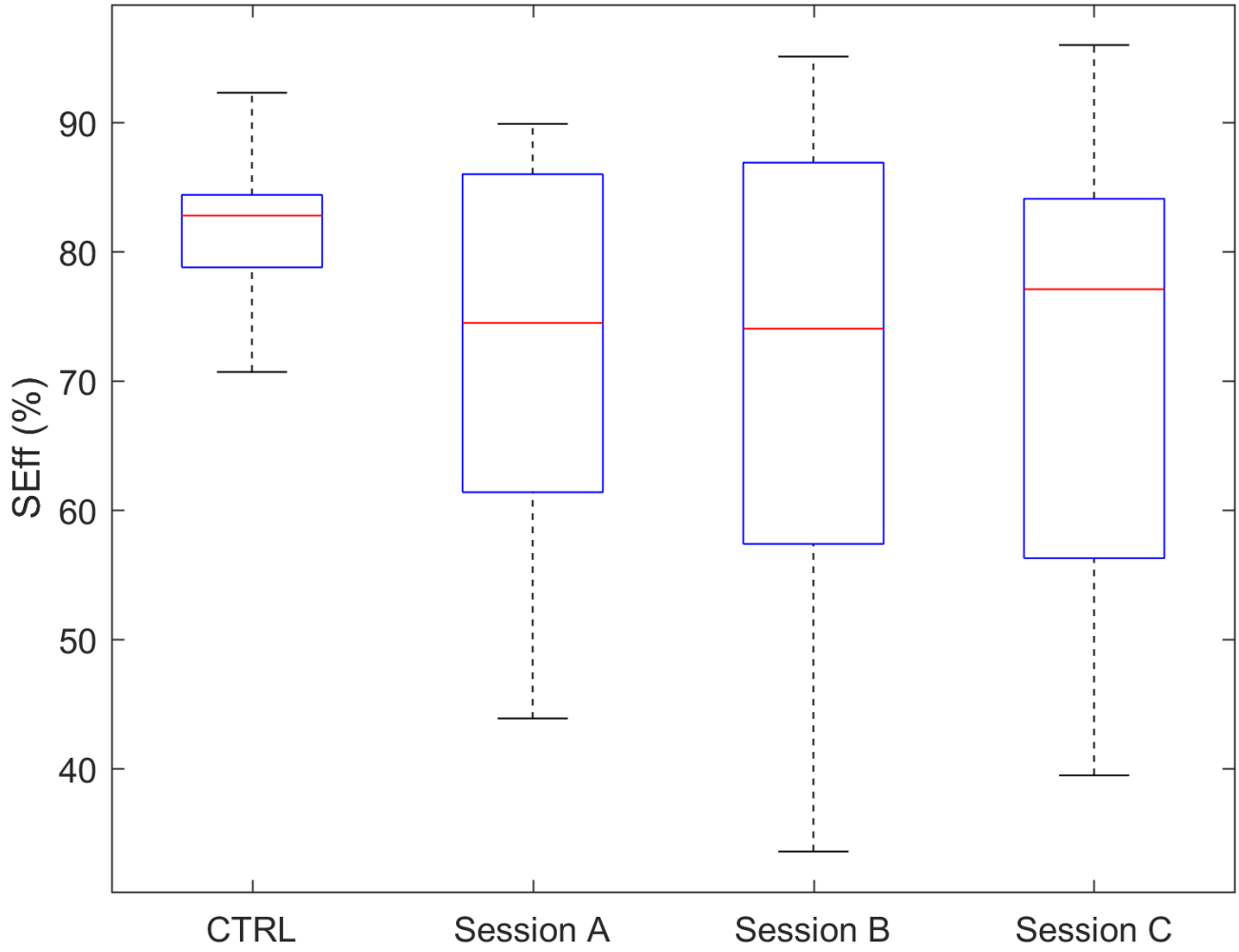
Figure 6. CC index for CTRL and for the 3 sessions of the experimental group. Red "+" marks indicate outlier individuals. The asterisk indicates that control data have significantly (P<.05) higher values than the experimental group. CTRL: control data; CC: cross-correlation coefficient.



The indices quantifying efficiency of wake time (WEff) and sleep time (SEff), that is, the percentage of the a priori fixed duration of, respectively, daytime and nighttime, showed a tendency toward a smaller efficiency in the experimental group than in controls, which failed to achieve statistical significance after correction for multiple comparisons. Though a perusal of values' distributions for SEff (Figure 7) and WEff (Figure 8) pointed out the occurrence of a not negligible number of individuals showing fairly reduced efficiency indices' values. Interestingly, although these 2 indices did not show a between-group significant difference, their test-retest reliability proved to be excellent (ranging from 0.74 to 0.89).

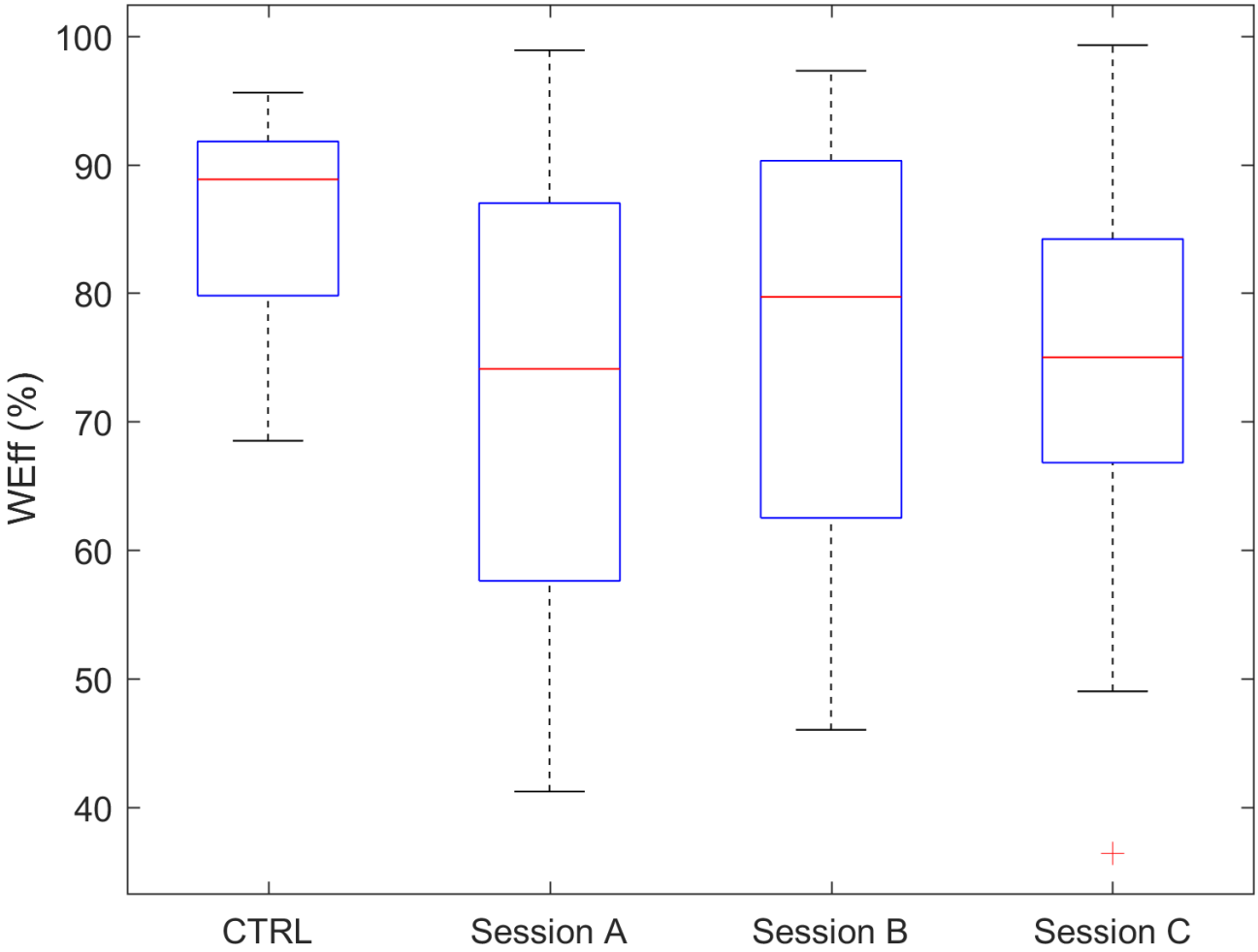
XSL•FO
RenderX

Figure 7. SEff index for CTRL and for the 3 sessions of the experimental group. CTRL: control data; SEff: sleep efficiency.



XSL•FO
RenderX

Figure 8. WEff index for CTRL and for the 3 sessions of the experimental group. CTRL: control data; WEff: wake efficiency.



The indices related to location tracking (PRI, PUB, and corridor) could not be compared to control data, which were unavailable due to the specific nature of the measurement. There was no significant intersession difference, and the test-retest reliability was generally fair (it was good just in 1 of 6 ICC values).

These location tracking indices' boxplots and the between-sessions scatterplot are available as a slide presentation in Multimedia Appendix 1.

Discussion

Features of the Proposed Monitoring Methodology

Instrumented methods for physical activity detection and tracking are valuable tools for studying individuals with dementia [42], complementing self-report measures such as questionnaires and surveys.

The presented method has provided a quantification of various aspects concerning physical activities during wake time, sleep, and social interactions among the components of the participants' group. Besides the characterization of average behaviors as quantified by the various indices, it is worth noting that distinctive individual behaviors, different from the average patterns, were consistently identified as well, as shown also by the high test-retest reliability of the computed indices. In fact, this sensitivity aspect supports the use of such methods not only in group studies but also in clinical settings where accurate individual characterization and management are essential.

The indices considered in this study were intended to depict a multidimensional figure of each measured participant. Along with indices related to levels of physical activity (NPA, LPA, MPA, and VPA) [38], indices concerning sustained walking (PPB), social interaction (CC) [11], quality of nighttime rest (SEff), and daytime vitality (WEff) were included.

The indices' definitions are generally self-explanatory. Only the rigid identification of wake-time as daytime and sleep-time as nighttime adopted in the wake and SEff indices may be regarded as a critical choice. An alternative choice would be a functional identification of wake epochs and sleep epochs [43]. In fact, the relatively simple original definition of SEff in sleep studies (percentage of time asleep on the total time spent lying in bed, with these 2 measures obtained, respectively, from EEG (electroencephalography) polysomnography and observation log) required a redefinition of the indices when wrist accelerometry was introduced [44]. Interestingly, a literature review [45] has highlighted that the use of wrist accelerometry in sleep studies offers several advantages, such as high participant compliance, but also drawbacks, including the potential under- or overestimation of sleep indices. In this study, a rigid timeline approach was adopted to meet the requirements of clinical partners, who were specifically interested in monitoring residents' compliance with the daily routines

XSL•FO
RenderX

established within the long-term care unit. According to these routines, all guests are expected to be in bed during the "night" interval (11 PM-7 AM) and awake during the "day" interval (9 AM-7 PM). The "day" and "night" time-intervals are separated by 2 transition time intervals, during which "go-to-bed" (7 PM-11 PM) and "get-out-of-bed" (7 AM-9 AM) events may occur. Interestingly, these events do not occur simultaneously across the guests. Consequently, the inclusion of the transition intervals within either the "day" or "night" intervals would have implied the risk of false positive or false negative classifications. Importantly, omitting the transition intervals does not compromise the reliability of the efficiency indices, which are computed over time windows of 8 hours ("night") and 10 hours ("d"). These durations are consistent with the minimum actigraphic recording time of 8 hours required to obtain reliable indices [46]. Therefore, we consider the rigid timeline approach reasonable and justify the inclusion of the transition intervals as a strategy to improve the reliability of the computed indices.

Features of the Implemented Experiments

The indices obtained in the experimental group were compared with proper age-matched normative data extracted from already available datasets [11,30]. The only exceptions are for the CC index whose normative data were from a younger group (mean age 72, SD 7.6, range 58-87 years) [11], and for the location tracking indices which were based on a specific local network of BLE antennas and therefore did not allow neither the recording of normative data in the long-term care residential unit, nor the option to set up a duplicate of the residential unit for healthy older adults, was considered.

The installation of the system of BLE antennas was performed in the weeks preceding the experiments without major interference with the daily flow of activities in the care unit. The accuracy of the system (ie, the correct classification of the room where the specific BLE beacon was in) was obtained as a result of a calibration procedure following the installation. The experimental data collection was carried out smoothly without any potentially confounding events, particularly, as described by the clinical logs, the participants constantly resided in the long-term care unit without leaving it. The participants occasionally received nursing assistance in their activities, particularly hygiene-related ones, in which partly passive movements could be interpreted as LPAs. However, these nursing activities were limited in time (few minutes), therefore were not considered to alter the experimental outcome significantly. All accelerometric recordings were a posteriori checked for identifying sensor removal (a constant baseline clearly distinguishes these from actual measurement, even from a still sleeping participant) and eventually tagged to exclude these time intervals from further analyses.

Discussion on Outcome Indices Compared With Reference Data

The indices related to physical activity (NPA, LPA, MPA, and VPA) evidenced significant differences between the group of persons with dementia and age-matched healthy participants: a slightly larger, though not significantly different after *P* value correction, duration of phases of NPA, a comparable LPA, a significantly reduced duration of MPA, and a substantially absent VPA. This figure is compatible with a commonly expected behavior in a nursing ward: increased sleep or rest, even in a seated position, particularly during the day [47], and drastically reduced activities requiring a relevant physical involvement [48]. A perusal of the individual data allowed us to identify individuals who showed activity profiles that differ from average behavior: particularly, 2 cases evidenced a very low value of NPA, which was compensated in 1 case (S18) by large values of LPA, while the other case (S23) evidenced the largest values of MPA. Interestingly, S18 had the lowest Barthel score among participants, while S23 was the youngest and had the lowest Tinetti score.

The index PPB, related to a pseudoperiodic motor behavior, is to be interpreted as an index of steady walking. Indeed, while the PPB index was, on average, not significantly different between controls and this study's group, a small subgroup of participants showed higher than normal PPB (S25, S23, and S07 above all controls in all 3 sessions, S13, S06, and S09 above at least 9 of 10 controls in all 3 sessions). The motor behavior of these participants resembles the so-called "wandering" behavior [49,50], a continuous, uninterrupted walking along the allowed paths, such as up and down a main corridor. Specifically, in this study's group, 6 individuals had this wandering behavior, clearly detectable in PPB scatterplots. Interestingly, this behavior is very consistent and can be observed also in repeated sessions, weeks, or months apart.

The CC index, quantifying the accordance between MA profiles of any possible couple of participants, is the key variable in the so-called social actigraphy approach [11] that aims at identifying congruent synchronized behaviors induced by the social proximity (in the original paper due to the marriage or cohabitation, in this study due to the common belonging to the community living in the long-term residential unit). In the presented data, the CC index showed significantly lower values in the experimental group, compared to reference data published in literature [11]. This result was initially unexpected, since a simple assumption implies that a tight social community, such as the one composed of guests of a nursing home, should be characterized by higher between-subjects CC values. The presented data evidenced, on the contrary, that participants with dementia show larger between-subjects differences in the MA profiles (corresponding to lower CC values) than observed in healthy participants. This outcome is surprising when the attention is all upon the social context, with common activities such as having meals together or participating in socialization events, since it would imply a certain "temporal synchronization" of motor behaviors. On the contrary, this result could be explained by the fact that people with dementia are known to often show apathy and withdraw from social interactions [51], and sometimes individuals with dementia may even show sociopathic traits [52].

The index SEff showed, on average, comparable values between healthy participants and persons with dementia. Interestingly, a large number of this study's group was characterized by lower-than-normal values, thus evidencing a troubled management of the hours eligible for sleep or at least rest. The repeatability of these lower-than-normal values lets us identify a subgroup of the guests of the residential ward that need special

XSL•FO
RenderX

attention concerning the management of the nighttime. Without entering into specific details, several actions can be performed, including a revision of pharmacological or nonpharmacological interventions [53-55]. Nonetheless, the present method lets us consider also more practical choices such as a change of the bedroom and related roommate. The index WEff evidenced a tendency, not statistically significant after *P* value correction, of lower-than-normal values for people with dementia, implying reduced percentage of activity time and increased resting, possibly daytime napping [56]. Symmetrically with previous comments on SEff, it is possible to identify a small subgroup that evidenced a higher-than-normal WEff, thus presenting themselves as the guests who never sleep during the day and are always active.

The indices related to location tracking (PRI, PUB, and corridor), that is, the percentages of total time spent in specific indoor locations, have no reference normative data. The interpretation of the observed values relies on comparison with other indices. The PRI values, representing the time spent in PRIs, were found to be slightly larger than NPA, the time characterized by no activity. This result is consistent with the fact that PRIs were mostly used for resting and sleeping. Conversely, the sum of PUB and corridor durations (ie, the time spent in any public space) was found to be slightly lower than the sum of LPA, MPA, and VPA durations (ie, the time spent at any level of activity), thus implying that MA might occur even in PRIs.

Test-Retest Reliability of Outcome Indices

The ICC values of the physical activity indices (NPA, LPA, and MPA) were excellent either in the short-term or in the long-term, with just a slight decrease of the values in the latter case. No ICC was computed for the VPA because vigorous activity was substantially absent in all cases. When considering PPB as a proxy for the amount of performed steady gait, the ICC was excellent in both short-term and long-term, and, to our surprise, it scored a higher value in the long-term. We are not willing to interpret this latter point as significant of some underlying phenomenon; rather, we consider it the result of a stochastic effect. Lower, but still good, ICC values were observed for the social actigraphy index CC. In fact, the CC values were substantially unchanged in all sessions, both after 3 weeks and after 3 months. As the CC values depicted a figure of a substantial absence of relations among the MA profiles of participants, even compared with control data, our interpretation is that people with dementia living in long-term care residences show signs of social withdrawal in their lifestyles as well as peculiar behaviors that may occur, such as the wandering one that largely differs from the inactive one which characterizes the majority of guests. Excellent ICC values were observed for SEff in the short-term. Indeed, participants tended to keep their night habits, which included either high SEff behaviors or persons who spent a large part of the night asleep. It is to be noted that the clinical management rules of the experimental residential unit excluded a general use of antipsychotics [57], thus increasing the possibilities for a person to be free to stay awake and even wander, as it was seldom observed, during the night. The test-retest reliability of WEff was between good and excellent. As for night habits, even during the daytime, participants had different behaviors (some stayed awake the whole day and others had frequent day naps). Those behaviors were individually consistent and repeatable. The location tracking variables, that is, the percentage time spent in different locations inside the residence unit, showed mainly fair values of reliability. The relatively lower values of the location tracking indices' ICC disagreed with the higher ICC values of the accelerometer-based indices. It is an open question how much of this low reliability of the location tracking indices is due to the actual, and different from the acceleration, nature of measured variables, and how much is due to a different behavioral performance. As, occasionally, we observed that the system delayed the detection of a displacement of a participant, that is, of the Bluetooth tag he wore, from one space to another one, we believe that these technical occurrences might have decreased the overall reliability of the Bluetooth measure, leading therefore to a decrease in the reliability of the variables which were derived from that system.

The obtained ICC values should be compared to ICC values found in the literature. When considering indices obtained by a questionnaire, a published study reported ICC ranging from 0.53 to 0.68 for total sedentary time, while for more focused domain-specific sedentary times, ICC ranged from 0.36 to 0.76 [24]. Another study concerning a self-administered questionnaire quantifying physical activity indices reported ICC values ranging from 0.56 to 0.91 for physical activity indices [25]. A questionnaire administered to caregivers of persons with dementia scored an ICC ranging from 0.87 to 0.95 [26]. When considering sensor-based assessments, a study on patients with knee osteoarthritis reported ICC ranging from 0.70 to 0.93 for physical activity indices [29]. In a gait monitoring approach in older adult participants, it was observed that higher repeatability was observed among the prefrail or frail group (ICC>0.78) compared to nonfrail individuals (0.39<ICC<0.55) [28]. In a study, in which the accelerometer data were processed with different algorithms, the test-retest reliability of the physical activity indices was generally very good to excellent (ICC=0.70 - 0.90) [27]. All considered, the test-retest reliability found in the present study is comparable with the already published literature.

Limitations

The presented experiment has some limitations. First, the location tracking method required the installation of a hardware infrastructure embedded in the building, which is not movable to another setting (this limitation obviously does not apply to the wrist accelerometer setup). As a consequence, control data obviously cannot be obtained from healthy age-matched individuals living in the same environment. This limitation might be overcome in the future, in the light of a rapidly evolving technological advancement, if detection antennas for an ultrawide band become ubiquitous indoors [58] or with AI-enabled surveillance cameras, provided that privacy concerns are solved [59]. Second, the activity tracking method had a compliance which was not 100% because some individuals did not agree to wear the wrist sensors or removed them during the experiment, resulting in a generic discomfort, though the clinical partners had preliminarily warned about these refusals as a consequence of the dementia affecting the unit residents. In the

XSL•FO
RenderX

latter cases, when the wrist sensors measured for more than 24 hours, the data were retained for multiples of 24-hour durations, and outcome variables were computed on the available full days. Nonetheless, such occurrences were seldom observed. Despite the authorization given by the legal tutors, our participants, persons with dementia, were still able to refuse to participate or at least to reject the wrist or ankle sensors because they felt disturbed by their presence. The dimension and location of the sensors might play a role in their refusal, so it may be expected that future technological developments, letting smaller sensors be embedded in the garments or fixed into adhesive patches, will reduce this noncompliance. Interestingly, compliance problems arose mostly for the wrist sensor, as the participant could perceive and grasp it. Third, the control group for the accelerometric data was built from an available dataset of healthy older adults living in the community [11,30] to be comparable as to average age, but the age range was narrower (80 - 87 years), 9 of 25 participants were above the maximum control age and 7 of 25 participants were below the minimum control age. This is due to the difficulty in recruiting healthy older adults in their nineties living in the community [60] while younger healthy older adults were excluded in order to have a matched average age. Fourth, the choice of monitoring activities with wrist accelerometry (adopted in 22% of the published actigraphic studies) was critically compared with the option of performing waist accelerometry (in 48% of the studies) [61], as an accelerometer on the waist is closer to the body center of mass and, therefore, considered more informative about physically engaging activities. The wrist choice was justified by the relevance of upper limb movements in older adult residents in a long-term care unit, as locomotor activities might be quite reduced, and by feasibility reasons, as an accelerometer positioned on the lower back with a belt would greatly interfere with autonomous bathroom activities of older adults, would not have had a stable position, and would be disturbing when lying in the bed. Moreover, the integration of a gyroscope sensor into the monitoring watch was not useful, and therefore dropped, as gyroscopes are mainly involved in activity classification, which was not an objective of this study.

Conclusions

In conclusion, the method involving activity tracking by a wearable accelerometer and location tracking through the detection of a wearable Bluetooth beacon by a network of Bluetooth antennas was proven to be feasible. The computed indices quantified the activities and social behaviors of the long-term care unit guests, providing a novel multidimensional characterization of the individual guest behavior, which considers the physical activities, the WEff or SEff, as well as the social interactions in terms of agreement between the activity profiles among participants. Though the results about activity levels in persons with dementia are in line with intuitive knowledge (ie, people with dementia move less than age-matched healthy participants), the assessment provided novel evidence against intuitive knowledge when quantifying the strict and continuous social interactions involved in a residential long-term unit. Such interactions, quantified by the CC index, resulted in lower values than those observed among healthy older adult couples not even living together. Apart from behavioral evidence, the fundamental novel information provided by this study consists of the quantification of the test-retest reliability of the computed indices, which resulted in good to excellent values, thus possibly supporting future power analyses in designing longitudinal studies on the activity tracking and localization of persons affected by health problems or frailty, and particularly by dementia or cognitive decline. These test-retest features complement the ones already available in the literature concerning studies on healthy adults [27], on walking bouts in frail older adults [28], and in people with orthopedic problems [29]. The results presented here will possibly support future studies able to characterize the effects of aging and changing social situations with a longitudinal design, or the effect of a possible therapy in a randomized controlled trial design. Interestingly, the methodology proved to be able to clearly identify behaviors, such as, for example, the wandering behavior, expressed by a minority of individuals inside this study's group of persons with dementia, thus supporting a possible use of the methodology in clinical settings, either as a way to characterize the individual or a means to detect changes in their behavior. Moreover, the computed indices might support a clinical team in improving the guests' experience [62] and in tailoring a personalized and coordinated across-subject rehabilitation program.

Acknowledgments

The authors wish to thank Piero Votta, Prof; Alberto Marzegan, DrEng; and Virginia Bandini, DrEng, for their help in the experimental setup installation. No generative artificial intelligence tools were used in this study.

Funding

The authors from IRCCS (Istituto di Ricovero e Cura a Carattere Scientifico) Fondazione Don Carlo Gnocchi ETS were funded by Italian Ministry of Health, under the Aging Network of Italian Research Hospitals (IRCCS; project: "Next Generation Promising [NGP]: a New Network Approach to Frailty, Multimorbidity and Agerelated Disease in the Era of Precision Medicine" RCR-202-223682286) and under "Ricerca Corrente 2025-2027" programme. The funders had no involvement in this study's design, data collection, analysis, interpretation, or the writing of this paper.

XSL•FO
RenderX

Data Availability

The dataset supporting this study can be provided by contacting the corresponding author and following an agreement with the Milano research unit.

Authors' Contributions

Conceptualization: MR, PDT, AG, FG, MF
Data curation: MR, EK, MPA, EDG, GB
Formal analysis: MR, IC, MF
Funding acquisition: AG, FG, MF
Investigation: MR, PDT, EK, MPA, EDG, GB, MF
Methodology: MR, PDT, IC, EDG, MF
Project administration: FG, MF
Resources: AG, FG, MF
Software: MR, IC, GB
Supervision: AG, FG, MF
Validation: MR, PDT, IC, MF
Visualization: MR, IC, GB, MF
Writing – original draft: MR, PDT, AG, IC, MF
Writing – review & editing: MR, PDT, AG, IC, EK, MPA, EDG, GB, FG, MF

Conflicts of Interest

None declared.

Multimedia Appendix 1
Boxplots and scatterplots are not included in the main text.
[PPTX File, 466 KB - mhealth_v14i1e70188_app1.pptx]

References

1. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. Lancet Public Health 2022 Feb;7(2):e105-e125. [doi: 10.1016/S2468-2667(21)00249-8] [Medline: 34998485]
2. Jehu DA, Pottayil F, Dong Y, Zhu H, Sams R, Young L. Exploring the association between physical activity and cognitive function among people living with dementia. J Alzheimers Dis 2025 Jun;105(4):1037-1046. [doi: 10.3233/JAD-230594] [Medline: 38363607]
3. Muñiz R, López-Alvarez J, Agüera-Ortiz L, Perea L, Olazarán J. Syndrome-based prescription to optimize psychotropics: are CHROME criteria a game changer? Front Psychiatry 2021;12:662228. [doi: 10.3389/fpsyt.2021.662228] [Medline: 33967863]
4. Korkatti-Puoskari N, Tiihonen M, Caballero-Mora MA, et al. Therapeutic dilemma's: antipsychotics use for neuropsychiatric symptoms of dementia, delirium and insomnia and risk of falling in older adults, a clinical review. Eur Geriatr Med 2023 Aug;14(4):709-720. [doi: 10.1007/s41999-023-00837-3] [Medline: 37495836]
5. Andres EB, Chaudhry I, Balasubramanian I, Poco L, Malhotra C, PISCES study group. Caregiver-reported use of physical restraints among community-dwelling older adults with severe dementia in Singapore. J Am Geriatr Soc 2024 Jun;72(6):1817-1823. [doi: 10.1111/jgs.18797] [Medline: 38424687]
6. Möhler R, Richter T, Köpke S, Meyer G. Interventions for preventing and reducing the use of physical restraints for older people in all long-term care settings. Cochrane Database Syst Rev 2023 Jul 28;7(7):CD007546. [doi: 10.1002/14651858.CD007546.pub3] [Medline: 37500094]
7. Attivazione nuclei alzheimer in residenze sanitarie assistenziali. bollettino ufficiale della regione lombardia [article in italian]. : 3° supplemento straordinario, numero 16; 1995 Apr 22 URL: https://www.ats-brianza.it/images/udoss/normaspecifica/RSA/D.g.r.%2064515_28febb1995.pdf [accessed 2026-05-14]
8. den Ouden M, Bleijlevens MHC, Meijers JMM, et al. Daily (in)activities of nursing home residents in their wards: an observation study. J Am Med Dir Assoc 2015 Nov 1;16(11):963-968. [doi: 10.1016/j.jamda.2015.05.016] [Medline: 26155723]
9. Straiton N, Alharbi M, Bauman A, et al. The validity and reliability of consumer-grade activity trackers in older, community-dwelling adults: a systematic review. Maturitas 2018 Jun;112:85-93. [doi: 10.1016/j.maturitas.2018.03.016] [Medline: 29704922]
10. Neikrug AB, Chen IY, Palmer JR, et al. Characterizing behavioral activity rhythms in older adults using actigraphy. Sensors (Basel) 2020 Jan 19;20(2):549. [doi: 10.3390/s20020549] [Medline: 31963889]

XSL•FO
RenderX

11. Rabuffetti M, De Giovannini E, Ferrarin M. To live together is to move together: social actigraphy applied to healthy elderly people. Sensors (Basel) 2022 Aug 11;22(16):6011. [doi: 10.3390/s22166011] [Medline: 36015772]
12. Morgan AA. On the accuracy of BLE indoor localization systems: an assessment survey. Comput Electr Eng 2024 Sep;118:109455. [doi: 10.1016/j.compeleceng.2024.109455]
13. Sattler F, Ma J, Wagner P, et al. Risk estimation of SARS-CoV-2 transmission from bluetooth low energy measurements. NPJ Digit Med 2020;3:129. [doi: 10.1038/s41746-020-00340-0] [Medline: 33083564]
14. Baronti P, Barsocchi P, Chessa S, Mavilia F, Palumbo F. Indoor Bluetooth low energy dataset for localization, tracking, occupancy, and social interaction. Sensors (Basel) 2018 Dec 17;18(12):4462. [doi: 10.3390/s18124462] [Medline: 30562934]
15. Surian D, Kim V, Menon R, Dunn AG, Sintchenko V, Coiera E. Tracking a moving user in indoor environments using Bluetooth low energy beacons. J Biomed Inform 2019 Oct;98:103288. [doi: 10.1016/j.jbi.2019.103288] [Medline: 31513890]
16. Moatamed B, Shahmohammadi F, Ramezani R, Naeim A, Sarrafzadeh M. Low-cost indoor health monitoring system. 2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN) . [doi: 10.1109/BSN.2016.7516252]
17. Fjeldsoe BS, Winkler EAH, Marshall AL, Eakin EG, Reeves MM. Active adults recall their physical activity differently to less active adults: test-retest reliability and validity of a physical activity survey. Health Promot J Austr 2013 Apr;24(1):26-31. [doi: 10.1071/HE12912] [Medline: 23575586]
18. Ramezani R, Zhang W, Xie Z, et al. A combination of indoor localization and wearable sensor-based physical activity recognition to assess older patients undergoing subacute rehabilitation: baseline study results. JMIR mHealth uHealth 2019 Jul 10;7(7):e14090. [doi: 10.2196/14090] [Medline: 31293244]
19. Ramezani R, Zhang W, Roberts P, et al. Physical activity behavior of patients at a skilled nursing facility: longitudinal cohort study. JMIR mHealth uHealth 2022 May 23;10(5):e23887. [doi: 10.2196/23887] [Medline: 35604762]
20. Toba MN, Pagliari C, Rabuffetti M, et al. Quantitative assessment of motor neglect. Stroke 2021 May;52(5):1618-1627. [doi: 10.1161/STROKEAHA.120.031949] [Medline: 33657852]
21. Aadland E, Ylvisåker E. Reliability of the Actigraph GT3X+ accelerometer in adults under free-living conditions. PLoS ONE 2015;10(8):e0134606. [doi: 10.1371/journal.pone.0134606] [Medline: 26274586]
22. Dillon CB, Fitzgerald AP, Kearney PM, et al. Number of days required to estimate habitual activity using wrist-worn GENEActiv accelerometer: a cross-sectional study. PLoS One 2016;11(5):e0109913. [doi: 10.1371/journal.pone.0109913] [Medline: 27149674]
23. Sasaki JE, Júnior JH, Meneguci J, et al. Number of days required for reliably estimating physical activity and sedentary behaviour from accelerometer data in older adults. J Sports Sci 2018 Jul;36(14):1572-1577. [doi: 10.1080/02640414.2017.1403527] [Medline: 29125022]
24. Wijndaele K, DE Bourdeaudhuij I, Godino JG, et al. Reliability and validity of a domain-specific last 7-d sedentary time questionnaire. Med Sci Sports Exerc 2014 Jun;46(6):1248-1260. [doi: 10.1249/MSS.0000000000000214] [Medline: 24492633]
25. Potchana K, Saengsuwan J, Kittipanya-Ngam P. Validity and test-retest reliability of a thai stroke physical activity questionnaire. J Stroke Cerebrovasc Dis 2021 Aug;30(8):105907. [doi: 10.1016/j.jstrokecerebrovasdis.2021.105907] [Medline: 34130103]
26. Durepos P, Akhtar-Danesh N, Sussman T, Ploeg J, Boerner K, Kaasalainen S. Evaluation of the caring ahead: preparing for end-of-life with dementia questionnaire. J Am Med Dir Assoc 2021 Oct;22(10):2108-2114. [doi: 10.1016/j.jamda.2021.06.032] [Medline: 34302752]
27. Sirard JR, Forsyth A, Oakes JM, Schmitz KH. Accelerometer test-retest reliability by data processing algorithms: results from the Twin Cities Walking Study. J Phys Act Health 2011 Jul;8(5):668-674. [doi: 10.1123/jpah.8.5.668] [Medline: 21734312]
28. Pradeep Kumar D, Wendel C, Mohler J, Laksari K, Toosizadeh N. Between-day repeatability of sensor-based in-home gait assessment among older adults: assessing the effect of frailty. Aging Clin Exp Res 2021 Jun;33(6):1529-1537. [doi: 10.1007/s40520-020-01686-x] [Medline: 32930988]
29. Johnston W, Scharfe M, Ryan P, et al. Defining a 'valid' measurement and test-retest reliability of actigraphy based digital endpoints of physical activity in knee osteoarthritis. Osteoarthr Cartil 2024 Apr;32:S237-S238. [doi: 10.1016/j.joca.2024.02.346]
30. Rabuffetti M, De Giovannini E, Carpinella I, Lencioni T, Fornia L, Ferrarin M. Association of 7-day profiles of motor activity in marital dyads with one component affected by Parkinson's disease. Sensors (Basel) 2023 Jan 17;23(3):1087. [doi: 10.3390/s23031087] [Medline: 36772127]
31. Hendrich AL, Bender PS, Nyhuis A. Validation of the Hendrich II Fall Risk Model: a large concurrent case/control study of hospitalized patients. Appl Nurs Res 2003 Feb;16(1):9-21. [doi: 10.1053/apnr.2003.YAPNR2] [Medline: 12624858]
32. Tinetti ME, Williams TF, Mayewski R. Fall risk index for elderly patients based on number of chronic disabilities. Am J Med 1986 Mar;80(3):429-434. [doi: 10.1016/0002-9343(86)90717-5] [Medline: 3953620]
33. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. Md State Med J 1965 Feb;14(61-5):61-65. [Medline: 14258950]

XSL•FO
RenderX

34. Folstein MF, Folstein SE, McHugh PR. “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
35. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology (ECronicon) 1993 Nov;43(11):2412-2414. [doi: 10.1212/wnl.43.11.2412-a] [Medline: 8232972]
36. Dotti C, Casale G, Zacchi V, Lovaglio P, Fazzone U. La Classificazione SOSIA degli anziani ospiti delle residenze sanitario-assistenziali lombarde [SOSIA classification of the frail elderly in nursing homes of region of Lombardy] [Article in Italian]. Ann Ig 2006;18(5):439-451. [Medline: 17089959]
37. Rabuffetti M, Meriggi P, Pagliari C, Bartolomeo P, Ferrarin M. Differential actigraphy for monitoring asymmetry in upper limb motor activities. Physiol Meas 2016 Oct;37(10):1798-1812. [doi: 10.1088/0967-3334/37/10/1798] [Medline: 27653815]
38. Gervasoni E, Anastasi D, Di Giovanni R, et al. Physical activity in non-disabled people with early multiple sclerosis: a multicenter cross-sectional study. Mult Scler Relat Disord 2022 Aug;64:103941. [doi: 10.1016/j.msard.2022.103941] [Medline: 35691235]
39. Rabuffetti M, Steinach M, Lichti J, et al. The association of fatigue with decreasing regularity of locomotion during an incremental test in trained and untrained healthy adults. Front Bioeng Biotechnol 2021;9:724791. [doi: 10.3389/fbioe.2021.724791] [Medline: 34900952]
40. Reed DL, Sacco WP. Measuring sleep efficiency: what should the denominator be? J Clin Sleep Med 2016 Feb;12(2):263-266. [doi: 10.5664/jcsm.5498] [Medline: 26194727]
41. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994;6(4):284-290. [doi: 10.1037/1040-3590.6.4.284]
42. Gallucci A, Trimarchi PD, Abbate C, et al. ICT technologies as new promising tools for the managing of frailty: a systematic review. Aging Clin Exp Res 2021 Jun;33(6):1453-1464. [doi: 10.1007/s40520-020-01626-9] [Medline: 32705589]
43. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. Sleep 1992 Oct;15(5):461-469. [doi: 10.1093/sleep/15.5.461] [Medline: 1455130]
44. Galasso L, Calogiuri G, Castelli L, et al. A novel index for actigraphic studies: the core sleep efficiency. Chronobiol Int 2025 Jun;42(6):817-824. [doi: 10.1080/07420528.2025.2510508] [Medline: 40454465]
45. Conley S, Knies A, Batten J, et al. Agreement between actigraphic and polysomnographic measures of sleep in adults with and without chronic conditions: a systematic review and meta-analysis. Sleep Med Rev 2019 Aug;46:151-160. [doi: 10.1016/j.smrv.2019.05.001] [Medline: 31154154]
46. Airlie J, Forster A, Birch KM. An investigation into the optimal wear time criteria necessary to reliably estimate physical activity and sedentary behaviour from ActiGraph wGT3X+ accelerometer data in older care home residents. BMC Geriatr 2022 Feb 17;22(1):136. [doi: 10.1186/s12877-021-02725-6] [Medline: 35177023]
47. Ramsey KA, Loveland P, Rojer AGM, et al. Geriatric rehabilitation inpatients roam at home! A matched cohort study of objectively measured physical activity and sedentary behavior in home-based and hospital-based settings. J Am Med Dir Assoc 2021 Dec;22(12):2432-2439. [doi: 10.1016/j.jamda.2021.04.018] [Medline: 34022152]
48. Moyle W, Jones C, Murfield J, et al. Levels of physical activity and sleep patterns among older people with dementia living in long-term care facilities: a 24-h snapshot. Maturitas 2017 Aug;102:62-68. [doi: 10.1016/j.maturitas.2017.05.015] [Medline: 28610686]
49. Cipriani G, Lucetti C, Nuti A, Danti S. Wandering and dementia. Psychogeriatrics 2014 Jun;14(2):135-142. [doi: 10.1111/psyg.12044] [Medline: 24661471]
50. Adekoya AA, Guse L. Wandering behavior from the perspectives of older adults with mild to moderate dementia in long-term care. Res Gerontol Nurs 2019 Sep 1;12(5):239-247. [doi: 10.3928/19404921-20190522-01] [Medline: 31158296]
51. Lozupone M, Dibello V, Sardone R, et al. Social dysfunction and apathy: transdiagnostic domains in late-life cognitive disorders. J Alzheimers Dis 2024;100(s1):S57-S61. [doi: 10.3233/JAD-240556] [Medline: 39031368]
52. Cipriani G, Borin G, Vedovello M, Di Fiorino A, Nuti A. Sociopathic behavior and dementia. Acta Neurol Belg 2013 Jun;113(2):111-115. [doi: 10.1007/s13760-012-0161-7] [Medline: 23180469]
53. Eggermont LHP, Blankevoort CG, Scherder EJA. Walking and night-time restlessness in mild-to-moderate dementia: a randomized controlled trial. Age Ageing 2010 Nov;39(6):746-749. [doi: 10.1093/ageing/afq115] [Medline: 20823127]
54. Aryankhesal A, Blake J, Wong G, et al. Sleep disturbance in people living with dementia or mild cognitive impairment: a realist review of general practice. Br J Gen Pract 2024 Apr;74(741):e233-e241. [doi: 10.3399/BJGP.2023.0171] [Medline: 38499365]
55. Wilfling D, Calo S, Dichter MN, Meyer G, Möhler R, Köpke S. Non-pharmacological interventions for sleep disturbances in people with dementia. Cochrane Database Syst Rev 2023 Jan 3;1(1):CD011881. [doi: 10.1002/14651858.CD011881.pub2] [Medline: 36594432]
56. Li P, Gao L, Yu L, et al. Daytime napping and Alzheimer’s dementia: a potential bidirectional relationship. Alzheimers Dement 2023 Jan;19(1):158-168. [doi: 10.1002/alz.12636] [Medline: 35297533]
57. Mühlbauer V, Möhler R, Dichter MN, Zuidema SU, Köpke S, Luijendijk HJ. Antipsychotics for agitation and psychosis in people with Alzheimer’s disease and vascular dementia. Cochrane Database Syst Rev 2021 Dec 17(12):CD013304. [doi: 10.1002/14651858.CD013304.pub2] [Medline: 34918337]

XSL•FO
RenderX

58. Huang Y, Cao B, Wang A. Design a novel algorithm for enhancing UWB positioning accuracy in GPS denied environments. Sci Rep 2024 Oct 12;14(1):23895. [doi: 10.1038/s41598-024-74773-y] [Medline: 39402136]
59. Park KW, Mirian MS, McKeown MJ. Artificial intelligence-based video monitoring of movement disorders in the elderly: a review on current and future landscapes. Singapore Med J 2024 Mar 1;65(3):141-149. [doi: 10.4103/singaporemedj.SMJ-2023-189] [Medline: 38527298]
60. Rabuffetti M, Zemp DD, Tettamanti M, Quadri PL, Ferrarin M. Stabilization after postural transitions in the elderly: experimental study on community-dwelling subjects and nursing home residents. Gait Posture 2022 Jan;91:105-110. [doi: 10.1016/j.gaitpost.2021.10.011] [Medline: 34673445]
61. Evenson KR, Scherer E, Peter KM, Cuthbertson CC, Eckman S. Historical development of accelerometry measures and methods for physical activity and sedentary behavior research worldwide: a scoping review of observational studies of adults. PLoS ONE 2022;17(11):e0276890. [doi: 10.1371/journal.pone.0276890] [Medline: 36409738]
62. Fahsold A, Schmüdderich K, Roes M, Verbeek H, Palm R, Holle B. The perspective of residents living with dementia toward their built environment: a walking interview study in German nursing homes. Innov Aging 2024;8(7):igae058. [doi: 10.1093/geroni/igae058] [Medline: 39022667]

Abbreviations

BLE: Bluetooth Low Energy
CC: cross-correlation coefficient
EEG: electroencephalography
ETS: Ente del Terzo Settore
ICC: intraclass correlation coefficient
IRCCS: Istituto di Ricovero e Cura a Carattere Scientifico
LPA: light physical activity
MA: motor activity
MET: Metabolic Equivalent of Task
milli-g: one-thousandth of an unit gravity acceleration (1 g about equal to 9.81 m/s^2
MPA: moderate physical activity
n.adm.: not administered (said of a clinical test)
NPA: no physical activity
PPB: pseudo periodic behavior
PRI: private room
PUB: public room
SEff: sleep efficiency
VPA: vigorous physical activity
WEff: wake efficiency

Edited by L Buis; submitted 17.Dec.2024; peer-reviewed by J Ruokolainen, M Liu; revised version received 18.Feb.2026; accepted 18.Feb.2026; published 21.May.2026.

Please cite as:
Rabuffetti M, Trimarchi PD, Gallucci A, Carpinella I, Kisel E, Andriani MP, De Giovannini E, Bailo G, Giunco F, Ferrarin M Definition and Test-Retest Reliability of a Monitoring Method Integrating Accelerometric Actigraphy and Bluetooth Indoor Location Tracking Applied in a Long-Term Residential Unit for Persons With Dementia: Longitudinal Observational Study
JMIR Mhealth Uhealth 2026;14:e70188
URL: https://mhealth.jmir.org/2026/1/e70188
doi:10.2196/70188

© Marco Rabuffetti, Pietro Davide Trimarchi, Alessia Gallucci, Ilaria Carpinella, Elena Kisel, Maria Patrizia Andriani, Ennio De Giovannini, Gaia Bailo, Fabrizio Giunco, Maurizio Ferrarin. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 21.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Failed Implementation of Mobile Access to Electronic Health Records in Home Care: Qualitative Study in Sweden

Lovisa Jäderlund Hagstedt[1,2], MD; Helena Hvitfeldt[1,2], PhD; Maria Hägglund[1,3,4], PhD

[1]Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

[2]Tiohundra AB, Stockholm, Sweden

[3]Department of Women's and Children's Health, Uppsala University, Box 256, Uppsala, Sweden

[4]Uppsala University Hospital, Uppsala, Sweden

Corresponding Author:
Maria Hägglund, PhD
Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

Abstract

Background: Digitalization and mobile health (mHealth) technologies hold promise for improving home care delivery. However, many mHealth initiatives fail to achieve their goals. Understanding the reasons behind these failures is critical for informing the successful implementation of mHealth in primary and home care settings.

Objective: This study aimed to explore the implementation process of a tablet computer with an mHealth app providing mobile access to the electronic health record (EHR) in home care, identifying barriers and facilitators to its uptake.

Methods: A tablet with EHR access was introduced at 4 primary care centers and 1 municipal home care organization in Sweden. Participants were nurses and physicians working at the study sites. Focus group discussions and interviews were conducted to obtain a rich understanding of implementation-related issues experienced by the health care professionals. Qualitative content analysis was conducted using the Consolidated Framework for Implementation Research to guide interpretation.

Results: Eighteen health care professionals (16 nurses and 2 physicians) participated in the study. The implementation of the mHealth app was largely unsuccessful. Key barriers included limited functionality of the app, technological immaturity, and unstable infrastructure. Organizational context influenced uptake, especially due to differing EHR systems and varying levels of user engagement. Users who were involved in the development process were more positive, despite the absence of certain functionalities, while those excluded struggled with adoption. Long development and implementation timelines and limited training reduced enthusiasm and negatively affected user engagement. Additional challenges included insufficient implementation planning, lack of leadership engagement, and inadequate resources for support and training.

Conclusions: For mHealth implementations to succeed, tools must meet users' needs and integrate seamlessly with existing eHealth ecosystems and infrastructures. Premature implementations can lead to change fatigue and diminish future engagement. Investments in user-centered design, thorough testing, organizational readiness, and sustained support are essential to realize the potential of mHealth in home care.

(JMIR Mhealth Uhealth 2026;14:e69590) doi:10.2196/69590

KEYWORDS

home care; primary care; work environment; CFIR; mHealth; eHealth; qualitative evaluation; electronic health record

Introduction

As populations age, the need for primary care and home care increases [1], yet resources remain scarce, with serious recruitment challenges concerning educated personnel [2,3]. According to the Swedish Association of Local Authorities and Regions, in the future, such care must be provided closer to the patients. Further, a holistic approach, greater use of digitalization, more collaboration among departments and specialties, and improved patient access are considered necessary. This sentiment is echoed in the World Health Organization's global strategy on digital health, where the purpose of the strategy is to "strengthen health systems through the application of digital health technologies for consumers, health professionals, health care providers and industry towards empowering patients and achieving the vision of health for all" [4]. Electronic health records (EHRs) are used in everyday care throughout Sweden [5]. Most regions use one EHR system throughout hospitals, primary care, and psychiatric care, making patient records available in the entire region. Elder care, however, is the responsibility of municipalities that fall under different legislation [6], and often use separate EHR systems [7-9].

XSL•FO
RenderX

The concept of mobile health (mHealth) has emerged as a subcategory of eHealth and can be defined as medical and public health practice supported by mobile devices [5,10]. The use of smartphones and tablets in health care is increasing. The use of mobile devices for home care in Swedish municipalities has increased significantly in the past few years; in 2020, 41% of the municipalities provided health care staff the possibility to read and write information using a mobile device, compared to 5% in 2015 [11,12]. By 2024, 70% of the municipalities used mobile devices in home care [12].

Further digitalization and the introduction of eHealth and mHealth have been seen as means to optimize health care [4]. Expectations are high that they will be the solution to many of the problems that exist in health care today [13]. However, many projects are unsuccessful [14], and challenges of digitalization have been raised [15]. There are several reasons why implementations of eHealth in health care fail [16,17]. By studying and analyzing why an eHealth or mHealth implementation is unsuccessful, we could learn how successful digitalization in primary and home care can be achieved.

A Cochrane review of mHealth apps showed that the complexities of health care delivery and human interactions defy simplistic conclusions on how health workers will perceive and experience their use of mHealth. Perceptions reflect the interplay between technology, contexts, and human attributes. Detailed descriptions of the program, implementation processes, and contexts, alongside effectiveness studies, will help to unravel this interplay to formulate hypotheses regarding the effectiveness of mHealth [18]. Despite the increasing use of mHealth [19], limited research on mobile EHRs exists, mostly focusing on in-hospital use for bedside EHR access [20,21]. Studies indicate that usability issues of mobile EHRs combined with low digital literacy negatively impact nurses' adoption of the tools [21], and may even have a negative impact on nurses' work environment despite the aim to better support their work [22].

This study was designed to evaluate an mHealth app that provides health care professionals (HCPs) with mobile access to the EHR and describe its impact on work processes and work environment. The project was, however, terminated prematurely since HCP did not adopt and use the system. The collected data were instead analyzed to understand the reasons for the implementation failure.

We aim to understand the reasons for an unsuccessful mHealth implementation and potential success factors by describing lessons learned from the implementation process of a tablet computer with an mHealth app that provides mobile access to EHR in home care.

Methods

Overview

A qualitative approach has been applied to understand the HCPs' expectations before implementation and their experiences of using the system. Individual interviews and focus group interviews have been used to obtain a rich understanding of the implementation-related issues. Due to the limited number of users of the mHealth app, a qualitative approach was deemed suitable to gain an in-depth understanding of the factors causing the failed implementation. The Consolidated Framework for Implementation Research (CFIR) by Damschroder et al [23] was used to guide our analysis.

Study Setting

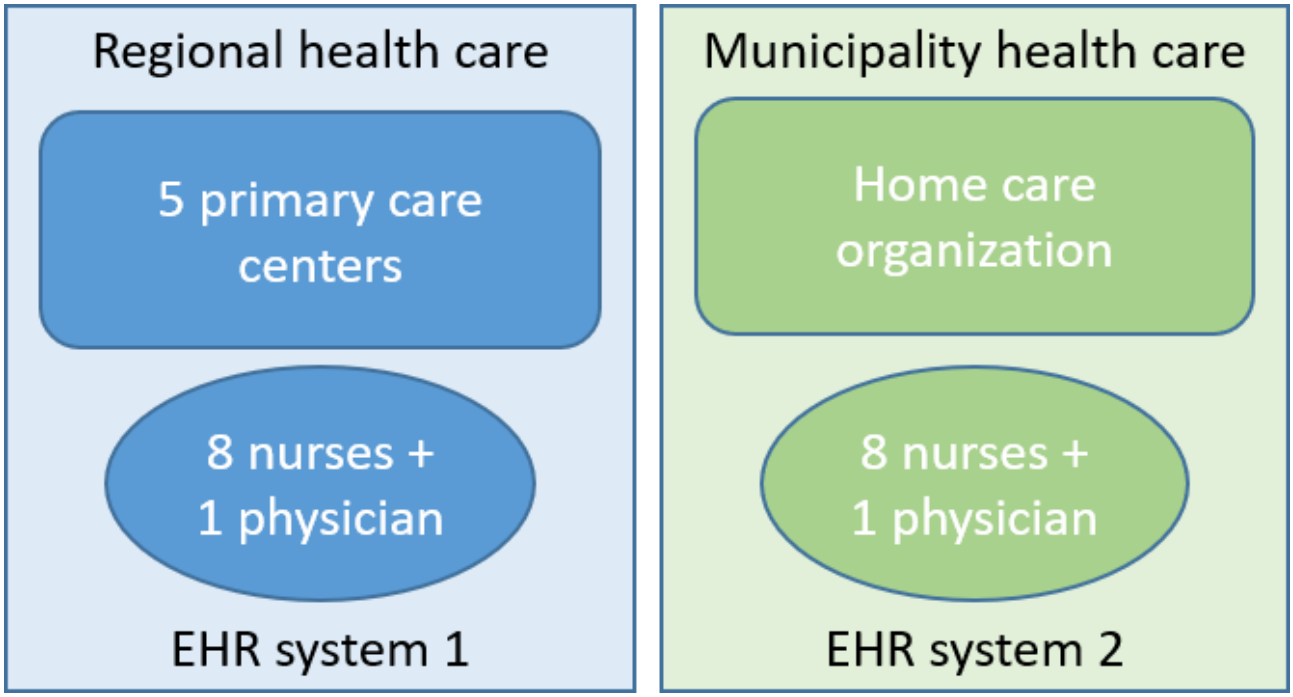
The study was conducted in a public municipality home care organization (MHCO) and the home care sections of 4 public academic primary care centers (PCCs) in a rural part of Region Stockholm.

The home care organization provides medical and supportive care for a diverse range of patients, predominantly older adults but also younger individuals with medical needs. The MHCO has the main responsibility for home care, whereas the PCCs provide some home care, for example, the first 2 weeks after discharge from the hospital and the support and service for persons covered under the Swedish Act Concerning Support and Service for Persons with Certain Functional Impairments. At the time of this study, there was no interaction between the PCCs and the MHCO.

The PCCs all use the same EHR system, accessible through the mHealth app, whereas the MHCO has a different EHR system, lacking accessibility through the mHealth app. Figure 1 presents an overview of the study setting.

XSL•FO
RenderX

Figure 1. Study setting describing the regional and municipality settings and included study participants. EHR: electronic health record.



The Intervention

The intervention consists of a tablet computer with several apps (eg, decision support, online prescription tools, and online supplies ordering), and the new mHealth app that provides mobile access to the EHR (example screenshots from the app are in Multimedia Appendix 1). In the following text, we will refer to the latter as "the mHealth app" and use the term "the mobile tool" when referring to the tablet with all its apps.

The mHealth app on the tablet computer was developed after an initiative by professionals (2 doctors and 2 nurses) at one of the participating academic PCCs. They saw a need for easy access to information, including patient data, and to carry out all parts of their home care work, including documentation, while in the patient's home. The HCP took an active part in an initial need and requirement analysis, and the mHealth app was then developed from their ideas. The first author was involved in some of the app's needs and requirement analysis. Professionals provided feedback on functionality and design but were not further involved in the mHealth app's development. No structured usability assessments were performed.

The project was carried out through a collaboration between academic PCCs and the IT unit of Region Stockholm, along with the private company Chorus, which was responsible for the technical development of the app. As an innovation project, each participating entity covered its own costs. The aim of the pilot project was to evaluate the tool's effectiveness and its potential for broader implementation within the region, with the long-term goal of generating both clinical benefits and economic value. The mobile tool was tested at selected academic PCCs, and the responsibility for approval and rollout was shared between the involved academic PCCs and the IT unit of Region Stockholm.

The mHealth app provided mobile access to the EHR used in regional home care. This included, for example, access to patient record entries, information about the patients to be visited during the day, and the possibility to write patient record entries while in the patient's home, a task otherwise done after returning to the office, as previously described in our study outlining the matching of the functionality of the app to the home care work processes [24]. While the previously published study described the work process in primary care in detail, matching it to the functionality of the study, it did not go into details on the users' experiences and reasons for the failed implementation. Due to technical difficulties, not all parts of the EHR were accessible through the mHealth app. The app did not provide information on patients' medications, referrals, or radiology results because the EHR from which the mHealth app collected the medical data did not support this output. However, it was possible to access information on the patients' previous outpatient and inpatient care visits, diagnoses, and laboratory tests. Thereby, some, but not all the HCPs' needs were met by the mHealth app. The mHealth app was intended to be further developed, and the implementation was therefore initiated.

The tablet computer also provided access to several clinical decision support systems and a mobile web-based prescription system that is used for some of the patients in home care, thus enabling the HCP to manage part of those patients' medications. Thereby, the diverse functions of the tablet computer provided HCPs with a multifaceted mobile tool for home care visits.

There were no costs involved for the users. Use of the mobile tool was strongly encouraged but voluntary.

The mobile tool was first implemented in the home care section of the PCC, which was part of the development of the mHealth app and in one of the other PCCs. In both participating PCCs where the mobile tool was first implemented, some HCPs were avid advocates for the mHealth app.

XSL•FO
RenderX

At a later stage, the mobile tool was implemented in the MHCO and the remaining PCCs. There were no advocates for the mHealth app at either of these workplaces. The mHealth app was further developed later to improve stability and implemented and used in in-hospital care for bedside record access in the Stockholm region. In 2020, Region Stockholm ceased all support of the mHealth app, and it is currently not in commercial operation.

Data Collection

Data for this study were collected through focus groups and semistructured interviews with nurses and doctors working in home care where the mobile tool was introduced. The first author (LJH, female) was responsible for the data collection, with the assistance of 2 Master's students (both female, health informatics students). All 3 had both training in and experience of qualitative data collection, and were supported by the last author (MH, female) with extensive experience in qualitative research. LJH was working as a general practitioner in the region at the time of the study and was well-known to all study participants. An overview of the timeline and data collection is presented in Table 1. Before implementation, a workshop was held where questions were asked regarding the time spent on administrative work related to home visits, experienced stress, use of decision support tools, and the experienced quality of documentation and communication between HCPs.

Table . Overview of the data collection timeline, including the method used, setting, and time for data collection.

Data collection	PCC1[a]	PCC2	PCC3	PCC4	MHCO[b]
Development of an app	2014 - 2015	—[c]	—	—	—
Focus group on expectations	November 2015	May 2016	April 2017	April 2017	April-May 2017
Individual interviews	May 2016	February 2017	February 2018	March 2018	February 2018
Focus group on completion of study	October 2016	—	—	—	—

[a]PCC: primary care center.

[b]MHCO: municipality home care organization.

[c]No data were collected using that method at that specific unit.

The mobile tool was introduced and demonstrated during a workshop where participants had the opportunity to practice using it. The demonstration was conducted by guiding participants through the different functionalities step by step, ensuring a comprehensive understanding of its features and applications. This hands-on approach facilitated engagement and allowed attendees to gain practical experience with the tool.

The original plan was to conduct data collection through (1) preimplementation focus groups, (2) individual interviews after 6 months of use, and (3) a second round of focus group interviews after 12 months of use. The different qualitative methods were chosen to gather a rich dataset; the focus group interviews enabled discussion and exchange of opinions between participants, whereas the interviews allowed more detailed exploration of individual HCPs' experiences of using the mobile tool. We also recognized that individual interviews provide an opportunity for participants to express thoughts and concerns they might be reluctant to share in a focus group setting with other participants present.

Interview guides were developed for individual and focus group interviews (Multimedia Appendices 2 and 3). All interviews and focus groups were recorded and transcribed. Notes were also taken during focus groups and included in the material. All interviews were held in Swedish. The individual interviews lasted 20 - 30 minutes and the focus group interviews lasted 1 - 1.5 hours. Data collection was conducted where most convenient for the participants, that is, at the PCCs or the MHCO's office. Data collection was conducted during working hours. Participant checking of the transcribed interviews was not done to reduce the burden on the study participants.

Preimplementation focus group interviews were held in all the PCCs and the MHCO, where the positive and negative expectations were discussed and reflected upon by HCP who would use the mobile tool.

Individual interviews with nurses and physicians were conducted according to plan to capture the staff's experiences using the mobile tool and its impact on work processes.

Interviews were conducted when the mobile tool had been used for at least 6 months. All interviews were conducted by the first author and the Master's students in the research team. The Master's students conducted interviews in the MHCO. One of the Master's students continued working in the project as a research assistant after graduation and also contributed to the data analysis.

All eligible nurses and physicians working in the setting were invited and accepted to participate in the study. Study participants were approached via email and in-person by the first author through the managers of the MHCO and PCCs. The same participants at each site participated in both the focus groups and interviews.

A second set of focus group interviews was planned at the end of the study, when the mobile tool had been used for 1 year. The PCC involved in the mHealth app's development started testing earlier than the other clinics, and a focus group interview was held a year after they started using the mobile tool to capture their long-term experiences. When the 6-month interviews were conducted at the other clinics, they had stopped using the mobile tool, and the decision was made not to conduct the planned

XSL•FO
RenderX

focus group interviews, as they would not gain further experiences of using the mobile tool.

Data Analysis

All interviews were transcribed verbatim and analyzed using a deductive content analysis approach [25]. The CFIR [23] was used to structure the analysis. The data analysis performed in this study does not include any quantification of responses regarding specific categories.

CFIR identifies 5 domains that are essential to implement an intervention [23]:

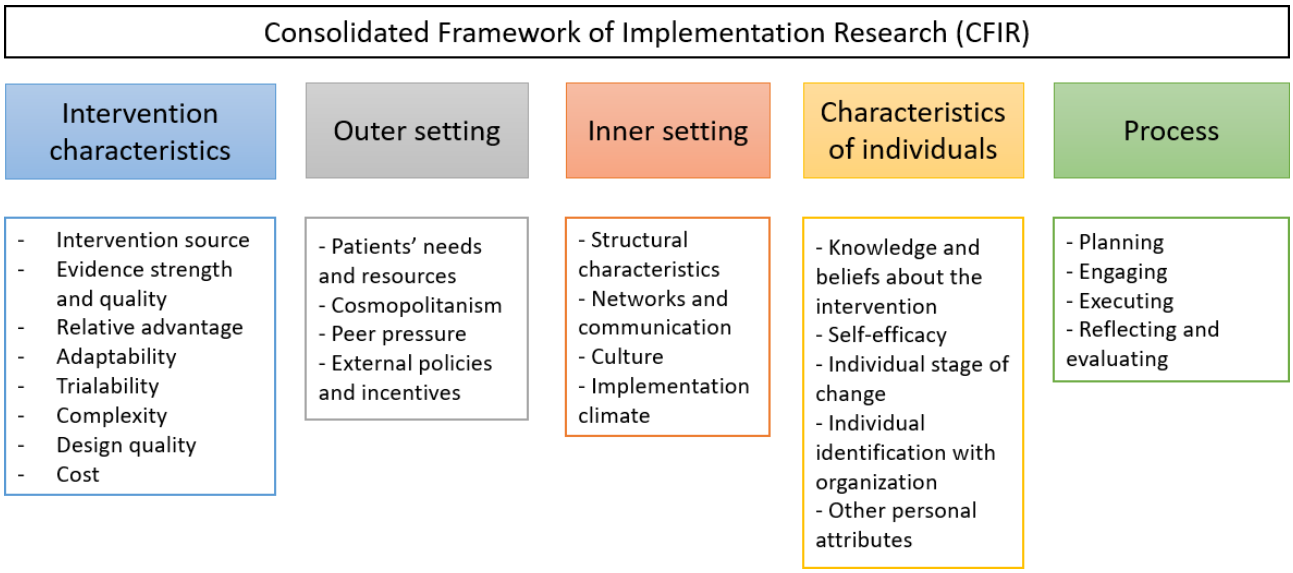
1. Intervention characteristics—features of the intervention itself, such as complexity, evidence strength and quality, and adaptability.
2. Outer setting—the external environment influencing implementation, such as external policies and incentives.
3. Inner setting—organizational factors such as culture, leadership, and available resources within the implementing organization.
4. Characteristics of individuals—traits and actions of the individuals involved in the implementation, including knowledge, beliefs, and self-efficacy.
5. Process—the planning, execution, and reflection processes involved in implementing the intervention.

In CFIR, a domain represents a broad category of factors or determinants that affect the implementation process. A construct, on the other hand, refers to a specific factor within a domain that provides more detailed insight into the aspects of implementation. An overview of the CFIR domains and their constructs is presented in Figure 2.

The 3 authors all actively participated in the analysis. Initial coding was performed by LJH, MH, and one of the Master's students who continued working on the project as a research assistant after graduating. First, all materials were read and reread, and meaning-bearing units were identified. Each coder coded and categorized the meaning-bearing units individually into 1 of the 5 domains of the CFIR framework. HH joined the analysis at this stage, and the 3 authors discussed the coding until a consensus was reached on the interpretation. Where appropriate, relevant CFIR constructs were used to signify our findings, but we did not systematically search for constructs in the material.

Due to the relatively small number of participants, achieving full data saturation was not possible. However, we ensured sufficient information power by including all eligible HCPs across multiple roles and settings. Thematic consistency across discussions suggests key insights were captured despite the limited number of study participants.

Figure 2. An overview of the Consolidated Framework for Implementation Research, including the 5 domains and their constructs.



Ethical Considerations

The Research Ethics Committee at Karolinska Institutet has granted ethical approval for this study (DNR 2015/1457-31/5). All study participants provided written or oral informed consent before participating in the study. In accordance with the Swedish Ethical Review Authorities' requirements, participants were informed that their participation was voluntary and that they could choose to revoke their participation at any time. Transcribed interviews and focus groups were deidentified to protect the participants' privacy, and confidentiality was preserved when reporting results. Study participants did not receive any compensation for their time.

Results

Overview

The 18 participants in the study were all nurses and physicians working in home care in either the home care section of the participating PCCs or the MHCO. A total of 8 nurses and 1 physician from the PCCs (PCC1=2 nurses and 1 physician, PCC2=1 nurse, PCC3=3 nurses, and PCC4=2 nurses), and 8 nurses and 1 physician from the MHCO participated in the study (Figure 1). Participants were predominantly female (17/18, 94.5%), and age ranged between 43 and 64 years. Due to the small number of participants in the study, we chose not to collect

XSL•FO
RenderX

and report further detailed characteristics of the individual study participants to ensure their anonymity.

Using the 5 domains of CFIR, we present our analysis of the HCPs' expectations before implementation and their experiences during the implementation of the mobile tool. We identify the key barriers causing the implementation to fail and the aspects that had a positive influence.

Intervention Characteristics

Before the implementation, HCPs' expectations were high, yet the interviews after 3 and 6 months of using the app revealed that many of the expectations were unmet. Textbox 1 presents an overview of the constructs related to the characteristics of the intervention, with further details from the qualitative analysis presented below.

Textbox 1. Identified categories related to the Consolidated Framework for Implementation Research domain intervention characteristics, divided by expectations before and experiences after implementation.

Expectations before implementation

- Relative advantage
  - Reduction of cognitive workload
  - Time-efficient and timely
  - Real-time access to information
  - Lack of information and functionality
- Complexity
  - Increased complexity and double documentation with many mobile apps
  - Workarounds to adapt to the system may increase cognitive workload
- Design quality
  - Slow and frustrating
  - Immature and unstable

Experiences after use

- Relative advantage
  - Reduction of the cognitive workload
    - Finish work on the site
    - No need for reminders
  - Access to important functions and information
  - Missing functions
  - Reduces unnecessary lab tests and orders
  - Note formatting is worse
- Complexity
  - Adaptations of work processes are required
    - Condensed documentation
    - Preliminary documentation on site
- Design quality
  - Poor design or unclear layout
  - Design improved during the project
  - Immature and unstable

XSL•FO
RenderX

Preimplementation Expectations

HCP had high hopes that the mobile tool would support their work and serious concerns that the technology would not meet their expectations.

Relative Advantage

During the preimplementation focus group interviews, the HCP expressed high expectations that the new mobile tool would help reduce their cognitive workload by allowing them to complete tasks (eg, ordering materials or sending referrals) and finish documentation at the patients' home:

> *To complete the documentation before leaving the patient's home. It would be a dream. You forget so much on the way back to the office.* [Nurse, MHCO]

Real-time access to information would make it possible to make more informed decisions and reduce the need to call colleagues to find information, for example, from the EHR. However, there were also concerns:

> *It's faster to write a post-it note. If it [the mobile tool] doesn't provide everything we need, it may end up in the office* [Nurse, PCC 1]

Staff already knew that some important functionality (eg, the medication list and referrals) would not be available, and concerns were that this would limit the app's usefulness:

> *Frankly, it feels a bit underdeveloped. Several things are missing, like the medication list and x-ray results, which is important when you visit the patient and need to discuss prior events or plans.* [Physician, PCC 1]

Complexity

There were also concerns that missing functionality would result in workarounds and the need for double documentation that might increase the administrative and cognitive workload, especially concerning medications and the lack of necessary templates in the medical record:

> *Above all, I would like access to the medication list. I know I won't, but it would make my job much easier since I organize the patients' pill box dispensers in their homes.* [Nurse, PCC2]

Design Quality

Before the implementation, there were some apprehensions regarding the system's maturity and that it might be slow and frustrating:

> *...that the software is only half-developed and not fully functional. (Physician, PCC 1). Several HCP were worried that the testing would be hampered due to immature technology: If it's malfunctional you won't use it much.* [Nurse, PCC 1]

Intervention Experiences

Overview

The semistructured interviews and the focus group showed that the users' needs and requirements were not sufficiently met by the mobile tool, largely because the mHealth app was not sufficiently adapted to the users' needs. There was, however, a notable difference between how the mobile tool was perceived by the HCP who worked where the intervention was first implemented and who had taken a more active role in the needs analysis and requirements specification, compared to those working where the intervention was later implemented. Those involved in the mHealth app's development and who were part of the initial implementation considered the app much more useful than those who had not been part of the development.

Relative Advantage

One of the nurses responded as follows:

> *Patient visits have become more complete, new questions that arise can be answered quickly.* [Nurse, PCC 1]

The HCP experienced that the mobile tool facilitated work processes and found it fulfilling to be able to answer questions from patients immediately. Another positive effect of the intervention was that it reduced the cognitive workload:

> *I don't need to bring notes that I will later have to remember to document separately on every patient.* [Nurse, PCC 2]

However, not all needs and requirements were met. The concerns about missing functions that were expressed before the implementation were confirmed. Several of the participants stated that the mHealth app needed further development and additional functions:

> *Honestly, it feels undeveloped. Several things are missing, like the medication list and x-ray results, which are important when you're in the patient's home.* [Physician, PPC 1]

Complexity

The limitations in functionality, with missing templates, lack of dictation functionality, and technical challenges with the smaller keyboard on the tablet computer, proved to be a confining factor:

> *The templates are a bit too, well...wrong. So, it becomes very limited, and then it feels easier to take notes by hand.* [Nurse, PCC 1]

Many of the HCPs also felt that their documentation became more concise, and that the on-site documentation in the app was preliminary and needed further development when they returned to the office:

> *The documentation may be a little shorter, sometimes the record entries become shorter.* [Nurse, PCC 1]

For some, the limitations in documentation functionality resulted in them not using that at all:

> *And still, for me, it is faster to dictate than to write. So, I do not write any record entries in the app. I use it to read record entries in the patient's home. Then it works well for the most part.*" [Physician, PCC 1]

XSL•FO
RenderX

Design Quality

The users were not impressed by the app's layout, especially the design of the laboratory results:

> *But you get for example all the hemoglobin results one after the other. It's difficult to read. Seven or eight rows of results for each blood test.* [Nurse, PCC 1]

The users did, however, feel that the design improved somewhat during the project.

Most of the HCPs experienced difficulties with the functionality and deemed the app immature and unstable.

> *Sometimes the internet connection has been bad. Sometimes the application does not work despite an adequate internet connection. It has happened that I have been able to access the medical record, but then it just stops working. I tap and tap but nothing happens. That's it.* [Physician, PCC 1]

Another user stated:

> *All too often it has not worked. Either the internet connection has been inadequate or there is no explanation why I can't log in.* [Nurse, PCC 1]

Outer Setting

In this domain, the expectations of the intervention coincided with the actual experiences of the intervention to a large extent. Textbox 2 shows the categories identified in the outer setting domain.

Textbox 2. Identified categories related to the Consolidated Framework for Implementation Research domain outer setting, divided by expectations before and experiences after implementation.

Expectations before implementation

- Technical infrastructure
  - Previous experiences among the staff with poor internet connectivity in the home care areas caused worry
- Patient needs and resources
  - Health care professionals looked forward to being able to use the mobile device as a tool for learning and patient communication
  - Being able to answer the patients' questions immediately

Experiences after use

- Technical infrastructure
  - Poor internet connectivity
- Patient needs and resources
  - Supports better work processes and improves patient safety.
  - Facilitates patient participation.

Preimplementation Expectations

HCPs were expressing concerns regarding the internet connectivity in the area but were also looking forward to using the mobile tool to meet the needs of their patients.

Technical Infrastructure

Network connectivity is limited in parts of the rural areas of the study setting, and since many of the HCPs had previous experiences with problems due to poor internet connectivity, concerns were expressed before the implementation:

> *The fear is the network connection. That's the thing, and it's hard to solve.* [Nurse, PCC 2]

Patient Needs and Resources

HCPs were hoping that the mobile tool would enable them to improve their communication with patients as well as the patients' participation in their care by using it as both a tool for learning:

> *And you may be able to show patients how to easily access sites for healthcare e-services* [Nurse, PCC 1]

and a means to answer questions asked by the patient:

> *Medical record entries may be useful when you're in the patient's home. Like if a patient asks, ' What did the doctor say at the last visit? I did not understand.'* [Nurse, MHCO]

Intervention Experiences

The HCP's expectations regarding the outer setting were fulfilled to a high degree.

Technical Infrastructure

The limited network connectivity was a real challenge when using the mobile tool. One of the users stated:

> *Our damn bad network connection has proven to be a huge, huge problem.* [Nurse, PCC 2]

Patient Needs and Resources

The staff experienced that the mobile tool supported better work processes and thus improved patient satisfaction:

XSL•FO
RenderX

> *communication with the patient runs smoother because I'm able to answer questions on-site, eg, concerning lab results.* [Nurse, PCC 1]

However, when the mHealth app did not work, it also sometimes rendered other, less positive questions from patients:

> *I've appreciated being able to answer a question that the patient has asked. But sometimes you get questions about the application when it doesn't work. A patient asked me - But if it does not work, why do you use it?* [Physician, PCC 1]

Inner Setting

Overview

All HCPs experienced that the mHealth app was not adapted to their workflow. Therefore, the mobile tool was not seen as entirely useful. This was especially apparent in the MHCO and for the participating physicians since important functions for their workflow were lacking to a larger extent than for the nurses in the PCCs. There were also implications that the implementation process would have benefited from a stronger commitment from the organization and some of the HCPs from the MHCO experienced that there was not enough time for training and preparation before the implementation. Textbox 3 gives an overview of the categories in this domain.

Textbox 3. Identified categories related to the Consolidated Framework for Implementation Research domain inner setting, divided by expectations before and experiences after implementation.

Expectations before implementation

- Structural characteristics
  - Concerns that there would be no time to complete the additional tasks required to manage the system
  - Concerns that the lack of home care visits would make the system less useful and rarely used

Experiences after use

- Structural characteristics
  - With more home care visits, the need for the system increased
- Networks and communications
  - Unclear responsibility and support functions
- Implementation climate
  - Same job, new technology

Preimplementation Expectations

Overview

There were no mentions of the inner setting during the preimplementation workshops, and predominantly concerns were raised.

Structural Characteristics

Staff from the PCCs expressed concerns that the app might not be very useful and that there would be few opportunities for usage due to the few home care visits in primary care. There were also concerns that the management of the app would create additional tasks that there would not be any time for:

> *We might not have time for the increased administration.* [Nurse, MHCO]

Intervention Experiences

The concerns were mostly addressed, but issues related to the support within the organization were also experienced.

Structural Characteristics

Staff who had used the mobile tool for some time reflected that it would be more useful with additional home care visits:

> *If I was to have homecare visits a whole day, day after day, then I think it would be good to be able to do all the documentation immediately. Then I think it really would improve patient safety.* [Nurse, PCC1]

Networks and Communications

The responsibility for technical support of the mobile tool and mHealth app was unclear to many of the participating HCP:

> *But I don't know where to turn for help when it doesn't work. Therefore, I have unfortunately not used it for more than a month.* [Nurse, PCC2]

Another person stated:

> *Well, it's like; who can you ask?!* [Nurse, MHCO]

Implementation Climate

In many ways, the implementation of the mobile tool did not change the HCP's perceptions of their work:

> *It's simply a new technology for the same job.* [Physician, PCC 1]

XSL•FO
RenderX

Characteristics of Individuals

As previously described, the user's tolerance for challenges differed. Some continued using the mobile tool despite difficulties, whereas others stopped using it. The categories related to this domain are presented in Textbox 4.

Textbox 4. Identified categories related to the Consolidated Framework for Implementation Research domain characteristics of individuals, divided by expectations before and experiences after implementation.

Expectations before implementation

- Knowledge and beliefs about the intervention
  - Concerns that the technology will steal time and focus from the patient
  - Concerns of becoming too dependent on the technology
  - Concerns that the project will amount to nothing
- Self-efficacy
  - To dare to try the new technology
- Individual stage of change
  - Looking forward to getting to test
- Individual identification with organization
  - Worry about personal responsibility for the devices

Experiences after use

- Knowledge and beliefs about the intervention
  - Positive that this type of intervention could provide benefit
  - Unreliable technology has required patience and persistence
  - Needs further development and additional functions
- Self-efficacy
  - Embarrassing when technology does not work
  - Improved work processes improve patient safety, credibility, and work satisfaction
  - Confidence in using mobile technology
- Individual stage of change
  - Not used to this way of working
  - Creates a need for new knowledge
  - Have not used the technology

Preimplementation Expectations

Before the implementation began, users expressed eagerness to test the tools they had been waiting for, despite the already described concerns relating to, for example, internet connectivity and mHealth functionality.

Knowledge and Beliefs About the Intervention

HCPs expressed concerns about whether the mobile tool would influence their relationship with the patients, worrying that the technology would shift time and focus from the patient to the technology. They expressed that "there might be too much focus on technology during the visits" (physician, PCC1), and that there could be a "risk for less contact with the patient if the tablet takes more time" (nurse, PCC4).

Another concern expressed was the risk of becoming too dependent on the technology:

> *To become dependent on the tablet, and then when it breaks down you don't know what to do.* [Nurse, MHCO]

Participants also expressed concerns that the technology might never be ready for use or that its functionality would be reduced to the point that the system would become unusable.

Self-Efficacy

Before implementation, the HCPs' expectations varied. Some were confident in their ability to use it, while others were a bit more hesitant and stressed that daring to try would be important:

> *One way is to just start using the tablet; perhaps it'll go better than we think…* [Nurse, MHCO]

XSL•FO
RenderX

Individual Stage of Change

Similarly, individuals' readiness to try the technology varied, but most were eager to start testing the mobile tools.

> *To dare to start using it and see what happens. I think there are several areas of use. You can find all sorts of online services [mentions Swedish e-health services such as the national patient portal, portals for ordering materials, and online prescription tools]. And perhaps you can help the patient to find the patient portal [1177.se] by using the tablet.* [Nurse, PCC 3]

Those involved from the start were more positive and eager to test the mobile tools, having waited a long time for the pilot to start, whereas the others were more hesitant.

Individual Identification With the Organization

Concerns regarding personal responsibility for the device were only raised during the preimplementation focus groups. Participants had concerns that they might forget the tablet at the patient's home, drop it, and break the screen. They worried this could be either a waste of resources for the organization or that they would personally be responsible for the costs.

Intervention Experiences

Although the participants' characteristics did not necessarily change during the implementation, their self-efficacy and beliefs about the intervention appear to have been affected.

Knowledge and Beliefs About the Intervention

The study participants remained positive that the intervention could provide benefits, for example, by making work processes more time efficient and improving patient safety. One of the participants who had used the tool the most said:

> *I don't have to prepare, which usually takes 5 minutes, to print eg, the list of medications, and the latest lab results* [Physician, PCC1]

Care planning in the patients' home was also experienced as more efficient:

> *I can finish it at the patient's home, with the patient. You can discuss what to do next, and then that's done.* [Physician, PCC 1]

To complete the work while in the field was experienced as reducing the risk of forgetting tasks or losing information:

> *[...] and for the patients and the staff [at the nursing home], that they don't have to remind me, 'did you remember that? now this hasn't arrived, why not?,' 'Oh, I didn't order that, I forgot, or I didn't see the note….' It feels more organized.* [Nurse, PCC 1]

However, dealing with unreliable technology required patience and persistence from HCP.

> *The biggest challenge has been not to lose your patience. It's like well, now it doesn't work. I must try the next time again, and the time after that. And that has been difficult since I usually only do this kind of work once a week, so it has been a challenge to not just give up but to instead try and try again.* [Nurse, PCC 1]

Only the most enthusiastic users had the patience to continue trying despite the unstable technology. The other users, especially those who worked in the MHCO, gave up on the first hurdle and never used the mobile tool.

Self-Efficacy

Unstable technology also affects the HCP's sense of professionalism and could be a cause for embarrassment in front of patients:

> *It feels embarrassing when it doesn't work when you use the tablet. I can live with that, but it's a little embarrassing. Then it would've been better not to have it [the mobile tool]* [Physician, PCC 1]

One of the HCPs had problems using the mobile tool in the presence of the patient's relatives:

> *That makes me stressed because the technology is new to me. When five people ask me questions at once. I haven't been able to handle that. [regarding not finding the necessary information on the mobile tool].* [Physician, PCC 1]

Those who did use the mobile tool found that it improved work processes, patient safety, credibility, and work satisfaction.

> *What's been very positive is that I can finish my work in the patient's home. I know that I won't forget anything. The patient won't be left without medications, which is good. Because it has been stressful before, I haven't always had the time to finish the work immediately when I'd returned to the workplace… so, I find it to be satisfactory as well, to be able to finish the job and not have any loose threads.* [Nurse, PCC 1]

They also felt confident using the mobile technology:

> *It's been good. It's easy to connect and I think it works well. I've not experienced any challenges* [Nurse, PCC 2]

However, there was a significant discrepancy between the two contexts, where the HCPs in the PCCs were much more positive. In the MHCO, most of the HCPs did not use the mobile tool at all.

Individual Stage of Change

As mentioned earlier, not all HCPs used the mobile tool:

> *I may have a maximum of five [patients]. And then I might as well remember that information or make some notes until I dictate [the record entries]* [Physician, PCC 1]

One reason was that the users were not used to this way of working. In some cases, the users did not like how record entries were presented:

> *I could edit the record entries later if I only write a few keywords [while in the patient's home]. But I don't know, I don't find the entries aesthetically*

XSL•FO
RenderX

> *appealing. I think they look much better when I do it my way.* [Nurse, PCC 2]

In other cases, they had no prior experience writing directly in the medical record, being used to dictating all record entries:

> *Well, I don't have that much experience. I know how to use the iPad, but it feels like I only use it to read record entries, not to write them myself. I usually don't [write record entries] so I don't do it here either.* [Physician, PCC 1]

The users believed there were other possible areas where the mobile tool could be useful:

> *I think it's more that you don't have enough imagination. You need to find other examples of what you can use it for. You just must use it more and discover even better ways of using it.* [Nurse, PCC1]

It was believed that it would be better to use the mobile tool regularly to reach its full potential:

> *If I had to use this every day, then maybe I would become more comfortable with it* [Nurse, PCC 1]

Process

The implementation process was not addressed in interviews with HCP, as this was not the focus of the original project. We will reflect on the implementation process in the discussion.

Discussion

Principal Findings

The implementation of the mHealth app was unsuccessful, despite reports of positive experiences and potential benefits. The main reason for the failed implementation was that the app did not sufficiently support the users' needs. Immature technology and unstable technical infrastructure also contributed. The findings from this implementation study highlight the critical importance of aligning digital health interventions with user needs and existing infrastructure, particularly in the context of aging populations and resource-constrained health care systems. For digitalization to be effective and sustainable, strategic approaches must prioritize usability, cross-organizational compatibility, and user involvement throughout development and implementation. Moreover, adequate resourcing—including time, training, and leadership engagement—is essential to prevent change fatigue and to ensure long-term adoption and impact of mHealth technologies.

Before implementation, HCPs had high expectations that the app would facilitate work processes and reduce cognitive workload by allowing real-time access to information. There were also concerns, mainly due to prior experiences of poor internet connectivity and immature and unstable technology, but also that the app might increase the workload, shift time, and focus from the patient, and that the project would amount to nothing. Ultimately, both expectations and concerns were confirmed. Another preimplementation concern expressed by some HCPs was that they would become too dependent on the new technology. This was not brought up again postimplementation, and we interpret this as being due to the failed implementation. The concern could have negatively influenced the HCP's willingness to adopt the technology, but likely this was not the most important factor for the failed implementation.

Some users, particularly those involved in the mHealth app's development, found value in the app despite its shortcomings and kept trying time after time, undeterred by the immature technology. However, users who worked in the MHCO and had not been involved in the design process gave up at the first hurdle and did not use the mobile tool. Data on usage of the mHealth app, for example, the number of log-ins per day or number of active users, were not available during the implementation process. Access to such data could potentially have guided the implementation process so that action could have been taken when usage was low [26].

End user involvement in the design process, often referred to as participatory design [27-29], has proven to be essential in both ensuring that the software meets the end user's needs and supports their work processes, and increases acceptance among the users [30,31]. In addition to the difference in participation during the development process, the MHCO also used another EHR system in their daily work that was not integrated with the mobile system. The benefit of being able to at least begin documentation in the patient's home was lost to them, further lowering their incentives to engage with the system. Lack of interoperability and integration between eHealth systems has previously been identified as an important barrier to successful implementation [16]. Physicians who made few home visits and used dictation for EHR documentation in everyday work found the mHealth app less useful. Again, the importance of capturing the perspectives of all potential end users in the design process was highlighted [31,32].

There are several lessons to be learned from this. First, that outcome depends on context, and second, that user engagement in both the development and the implementation processes is crucial for a successful implementation. In our study, both lack of user involvement [30,33] and poor interoperability [7,17] hurt the acceptability of the system, contributing to a system that did not support the user's work processes [24]. The importance of an interoperable digital health ecosystem is stressed in the World Health Organization digital health strategy, defining it as a "digital interoperable information technology infrastructure that is primarily used by the health care community across all care settings, in particular by health care providers, health service providers and patients" [4]. With such an infrastructure in place, the integration and implementation of the mHealth app would likely have been more successful. The lengthy development and implementation process also affected user engagement negatively. Studies of user involvement's relationship with user satisfaction have indicated that user involvement can overcome frustration from delays and prolonged projects [30]. In our study, we did see higher user satisfaction and tolerance among users who had been more actively involved in the design process, but it was not enough to overcome the implementation challenges completely. Additionally, for less engaged staff, the barrier to start using a new technology was high, and they also had a greater need for additional support, whereas users with high engagement, often

XSL•FO
RenderX

referred to as “champions” in the implementation science literature, kept trying regardless and often found new areas of use for the technology. The most common reason for discarding an eHealth service is issues with the technology [19,34]. eHealth services often are premature and thus do not adequately support the work in home care [19]. Technical issues [19] include everything from a lack of suitable infrastructure to usability issues, the latter referring to whether the product provides the right utility and the degree to which it helps users reach their goals in an effective, efficient, and safe manner, while still being easy to learn and remember how to use [35]. Usability of the mHealth app should have been evaluated more extensively before the implementation, for example, through user-driven scenario-based usability testing [36], or expert- or user-driven heuristic evaluations [37,38].

In our case, considerable problems with internet connectivity, immature technology, and a lack of functionality contributed to the low app usage.

Other studies and reviews confirm the challenges described by the users in this study. User involvement, interoperability, reliability of connection, and technology and infrastructure can be considered facilitators if handled appropriately in the implementation of eHealth, or barriers if they are not considered [12,17]. In a review of 43 studies of mHealth implementation experiences from primary health care services, the main results were that health workers appreciated mHealth when it improved feedback, speed, and workflow [17]. Challenges stated were that mHealth sometimes created more work, resulting in some preferring paper instead. Some health workers found the decision support opportunities useful, and others thought it threatened their clinical skills [18]. The aspect of workflow was found substantial for the success or failure of eHealth interventions particularly if workload increased, if workflow was interrupted, and if there was an alignment with clinical processes [17].

CFIR Suitability for mHealth Implementation

CFIR is developed to study implementation processes, not specifically new technology. We chose to include “technical infrastructure” as a new construct in the domain “Outer setting” since the effect of new technology depends on its functionality. A more specific focus on the usability of the eHealth or mHealth intervention could also be included under intervention characteristics, but we chose to use the “design quality” construct here.

The implementation process is crucial for the success of an implementation and for a successful implementation, all stages of the implementation must be properly executed. Our study was not originally designed to study the implementation process, and subsequently, material on some aspects of the implementation process is missing since no questions regarding those aspects were asked in the interviews. We see, however, that the implementation of the mobile tool most likely would have benefited from better planning, execution, support functions, and resources. When implementing complex interventions, feasibility and pilot studies are essential [35,39]. This is especially important when implementing mHealth where lack of usability is often the cause of failed implementations. This could be further specified in the CFIR domain “process” when studying new technology. A specific challenge of implementing the mHealth app in a rural home care area was the physical distance between units and the challenges of keeping day-to-day contact with and providing support to the HCP. The need for a specific facilitator role, that is, a person dedicated to supporting the implementation process and ensuring problems are identified and addressed, was apparent, as is also stressed in other implementation theories, for example, the Promoting Action on Research Implementation in Health Services (PARIHS) framework [40]. In line with the integrated-Promoting Action on Research Implementation in Health Services (i-PARIHS) framework and other implementation models, the facilitator role typically requires a combination of technical, organizational, and interpersonal competencies. Core competencies include basic technical troubleshooting and familiarity with the intervention, skills in user training and coaching, and the ability to communicate effectively across different organizational levels, including leadership. Responsibilities often span the entire implementation process: supporting planning and readiness activities before rollout, providing hands-on assistance and problem-solving during implementation, and ensuring ongoing support and feedback loops during maintenance. While the exact scope and emphasis of these tasks must be adapted to the local context and resources, these elements are commonly highlighted as critical for successful implementation.

There is a strong belief in the potential for eHealth and the introduction of new technology to revolutionize health care and that it will be necessary for health care and primary care in the future [4]. The introduction of new technology, however, often fails [17,41], and there is a risk that premature implementation of new technology will exhaust HCP for future projects and implementation. The challenges and concerns must be considered to move forward successfully [15]. According to the CFIR [23,42] and i-PARIHS [40] frameworks, the individuals’ previous experiences and beliefs about the intervention will impact the implementation’s success. When HCPs are exposed to immature digital solutions with poor usability, we may risk a negative impact on future implementation and innovation projects. To avoid this, we conclude that it is important to ensure that new technology has the necessary functionality and stability before implementation and that there are routines in place in the organization to guide decision-making and implementation processes. Structured contingency planning for technical failures, such as providing clear fallback procedures, timely support, and communication strategies, could sustain user engagement by reducing frustration, maintaining workflow continuity, and reinforcing trust in the implementation process despite technological setbacks.

Strengths and Limitations

The study was guided by the CFIR framework, which is a well-established model for understanding implementation processes [23]. This adds theoretical rigor and structure to the analysis, ensuring that a broad range of influential factors is considered. In our analysis, we used the original CFIR framework which was available when we began our analysis. More recent versions of the framework [42] have added new constructs; however, the domains remain the same, and we

XSL•FO
RenderX

believe that using the most recent version of CFIR would not have substantially influenced our results.

The use of interviews (both individual and focus group) allowed for an in-depth exploration of the HCPs' experiences and perspectives. This is particularly useful given the small number of users of the mHealth app, making qualitative methods appropriate for capturing rich, detailed insights. The study also collected data from both nurses and physicians across different settings (PCCs and MHCO), which ensured a variety of perspectives were included in the analysis. The small sample size (8 nurses and 1 physician from each of the two groups) may limit the generalizability of the findings to other settings or health care systems. Additionally, the majority of participants were female, which may impact the diversity of perspectives, yet it reflects the current workforce in Swedish home care. To maintain anonymity, the study does not report detailed demographic information about participants, such as years of experience, which could be useful for understanding how these factors might influence perceptions of the mHealth app. We also do not include any quantification of responses, which means that while rich qualitative data were captured, the relative importance of different factors or themes was not assessed. This could limit the ability to weigh certain barriers or facilitators more heavily than others; however, considering the qualitative nature of our dataset, we argue that this is the best approach as quantifying the results may be rather misleading. Using a validated instrument to measure health care professionals' self-efficacy could be a way to further understand how self-efficacy impacts adoption and abandonment, and how it may change over time during an implementation project. This should be explored further in future research.

A strength of the study is that the data collection spanned several years, and follow-up interviews after 6 months of using the mobile tool captured longer-term experiences and effects, offering insights into sustained usage and challenges. However, the study took place from 2015 to 2017, with final data collection in 2018, which could render the results of the study out of date. Yet the findings remain highly relevant in today's digital health landscape. Despite technological advancements, the fundamental implementation challenges we identified, such as poor alignment with clinical workflows, limited user engagement, inadequate training and support, and insufficient organizational readiness, persist in contemporary mHealth and eHealth initiatives. Implementation projects continue to experience implementation challenges across diverse settings, also in large-scale initiatives. For example, recent EHR rollouts in Sweden [43], Norway [44], and Finland and Denmark [45] encountered major disruptions due to user dissatisfaction, lack of preparation, and technological shortcomings. These cases underscore the enduring importance of addressing the sociotechnical and organizational dimensions of digital health implementations.

The first author was involved in the needs and requirement analysis for the mHealth app and also conducted the interviews. This dual role might introduce potential bias into the data collection and analysis process, as prior involvement may influence how the data are interpreted or what aspects are emphasized. Throughout the analysis process, we have been vigilant about this, and the involvement of Master's students with no prior connection to the project facilitated an unbiased data collection. The first author's familiarity with the context and the app has also been an enabling factor, to both data collection and by facilitating insights into both the design and implementation processes that would otherwise have been challenging to obtain in retrospect.

These strengths and limitations provide important context for interpreting the findings of the study and considering its applicability to broader implementation efforts.

Conclusions

We conclude that new technology must be stable and have the desired functionality before implementation for success. The functionality of a new technology must support the users' needs and be sufficiently integrated with other IT systems in the health care organization. How the implementation process is executed is important, but the implementation will not be successful if the new technology does not properly support the users' needs. In the case presented here, implementation failed because the technology was unstable and thus unreliable, not sufficiently supportive of the users' needs, and not adapted to the users' workflow. There were, however, positive experiences too, and mHealth still has a strong potential to support HCPs working in home care. To achieve this potential, we must ensure that sufficient resources are allocated to both design and development, evaluation and feasibility studies, and support and engagement during the implementation of mHealth in the future.

Acknowledgments

We thank Gabor Revay and Richard Scholvin at the Department of eHealth and IT, Region Stockholm, Health Care Service, Sweden, for their support and work on the project. We also thank the Master's students Sofia Rosborg (who continued working as a research assistant in the project) and Malin Lindberg for their contributions to the project.

We used generative artificial intelligence technology Grammarly (Grammarly Inc) to enhance the grammatical accuracy and vocabulary richness.

Funding

This work received funding from Innovationsfonden (20151143), FORTE – the Swedish Research Council for Health, Working Life and Welfare supports "MobEVAL" (2017- - 01962), and AFA Försäkring "ePrIm" (190210).

XSL•FO
RenderX

Data Availability

The datasets generated and analyzed during this study are not publicly available due to the qualitative nature of the materials and the terms of the ethical approval, but are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: MH, LJH
Data curation: MH, LJH
Formal analysis: LJH, MH, HH
Funding acquisition: MH, LJH, HH
Investigation: LJH
Methodology: MH, HH, LJH
Project administration: MH
Supervision: MH, HH
Visualization: MH
Writing – original draft: MH, LJH, HH
Writing – review & editing: LJH, MH, HH

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the mobile health app.
[DOCX File, 998 KB - mhealth_v14i1e69590_app1.docx]

Multimedia Appendix 2
Interview guide focus group interview—expectations of the intervention.
[DOCX File, 13 KB - mhealth_v14i1e69590_app2.docx]

Multimedia Appendix 3
Interview guide—evaluation after the intervention.
[DOCX File, 14 KB - mhealth_v14i1e69590_app3.docx]

References

1. Lanzieri G. Long-term population projections at national level. Eurostat Stat Focus 2006 Feb 28 [FREE Full text]
2. Russo G, Perelman J, Zapata T, Šantrić-Milićević M. The layered crisis of the primary care medical workforce in the European region: what evidence do we need to identify causes and solutions? Hum Resour Health 2023 Jul 14;21(1):55. [doi: 10.1186/s12960-023-00842-4] [Medline: 37443059]
3. Health and care workforce in Europe: time to act. : World Health Organization; 2022 URL: https://iris.who.int/handle/10665/362379 [accessed 2025-03-02]
4. Global strategy on digital health 2020-2025. World Health Organization (WHO). 2021. URL: https://www.who.int/health-topics/digital-health#tab=tab_1 [accessed 2025-12-05]
5. Rapport ehälsa och IT i landstingen 2018 - eHälsa och IT i landstingen maj 2018 Inventering på - Studocu [Website in Swedish]. URL: https://www.studocu.com/sv/document/hogskolan-dalarna/arytmi-kurs/rapport-ehalsa-och-it-i-landstingen-2018/15115952 [accessed 2025-03-02]
6. Lind L, Sundvall E, Karlsson D, Shahsavar N, Ahlfeldt H. Requirements and prototyping of a home health care application based on emerging JAVA technology. Int J Med Inform 2002 Dec 18;68(1-3):129-139. [doi: 10.1016/s1386-5056(02)00071-0] [Medline: 12467797]
7. Hägglund M, Scandurra I, Koch S. Studying intersection points - an analysis of information needs for shared homecare of elderly patients. J Inf Technol Healthc 2009;7(1):23-42 [FREE Full text]
8. Koch S, Hägglund M, Scandurra I, Information Resources Management Association (IRMA) (ed.). Informatics and socio-technical challenges when designing solutions for integrated ecare. In: Healthcare Ethics and Training: Concepts, Methodologies, Tools, and Applications: IGI Global Scientific Publishing; 2017. [doi: 10.4018/978-1-5225-2237-9.ch011]
9. Hägglund M, Scandurra I, Koch S. Supporting citizen-centered care for seniors - experiences from two swedish research projects. Presented at: 2012 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS); Jun 20-22, 2012. [doi: 10.1109/CBMS.2012.6266377]
10. MHealth: new horizons for health through mobile technologies: second global survey on ehealth. World Health Organization. URL: https://iris.who.int/handle/10665/44607 [accessed 2025-03-02]

XSL•FO
RenderX

11. E-hälsa och välfärdsteknik i kommunerna 2017 [Report in Swedish]. : Socialstyrelsen; 2017 URL: https://www.socialstyrelsen.se/contentassets/3bdc30a5fa3f4adbb4f3e279848cbc19/2017-4-22.pdf [accessed 2025-11-15]
12. E-hälsa och välfärdsteknik i kommunerna 2024 [Report in Swedish]. : Socialstyrelsen; 2024 URL: https://www.socialstyrelsen.se/contentassets/4c3331fe414a4749a3e57790aab363b5/2024-5-9099.pdf [accessed 2025-11-15]
13. van der Kleij R, Kasteleyn MJ, Meijer E, et al. SERIES: eHealth in primary care. Part 1: Concepts, conditions and challenges. Eur J Gen Pract 2019 Oct;25(4):179-189. [doi: 10.1080/13814788.2019.1658190] [Medline: 31597502]
14. Vision e-hälsa 2025 [Report in Swedish]. : Regeringskansliet (Socialdepartementet) and Sveriges Kommuner och Landsting; 2025 URL: https://ehalsa2025.se/wp-content/uploads/2021/02/vision-e-halsa-2025-overenskommelse.pdf [accessed 2025-12-20]
15. Öberg U, Orre CJ, Isaksson U, Schimmer R, Larsson H, Hörnsten Å. Swedish primary healthcare nurses' perceptions of using digital eHealth services in support of patient self-management. Scand J Caring Sci 2018 Jun;32(2):961-970. [doi: 10.1111/scs.12534] [Medline: 28960451]
16. Lennon MR, Bouamrane MM, Devlin AM, et al. Readiness for delivering digital health at scale: lessons from a longitudinal qualitative evaluation of a National Digital Health Innovation Program in the United Kingdom. J Med Internet Res 2017 Feb 16;19(2):e42. [doi: 10.2196/jmir.6900] [Medline: 28209558]
17. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of eHealth interventions: systematic review of the literature. J Med Internet Res 2018 May 1;20(5):e10235. [doi: 10.2196/10235] [Medline: 29716883]
18. Odendaal WA, Anstey Watkins J, Leon N, et al. Health workers' perceptions and experiences of using mHealth technologies to deliver primary healthcare services: a qualitative evidence synthesis. Cochrane Database Syst Rev 2020 Mar 26;3(3):CD011942. [doi: 10.1002/14651858.CD011942.pub2] [Medline: 32216074]
19. Rydenfält C, Persson J, Erlingsdottir G, Johansson G. eHealth services in the near and distant future in Swedish home care nursing. Comput Inform Nurs 2019 Jul;37(7):366-372. [doi: 10.1097/CIN.0000000000000536] [Medline: 31135467]
20. Thompson BW. The transforming effect of handheld computers on nursing practice. Nurs Adm Q 2005;29(4):308-314. [doi: 10.1097/00006216-200510000-00004] [Medline: 16260994]
21. Hsu SC, Liu CF, Weng RH, Chen CJ. Factors influencing nurses' intentions toward the use of mobile electronic medical records. Comput Inform Nurs 2013 Mar;31(3):124-132. [doi: 10.1097/NXN.0b013e318270100b] [Medline: 23114391]
22. Heponiemi T, Kaihlanen AM, Gluschkoff K, et al. The association between using a mobile version of an electronic health record and the well-being of nurses: cross-sectional survey study. JMIR Med Inform 2021 Jul 6;9(7):e28729. [doi: 10.2196/28729] [Medline: 34255704]
23. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implementation Sci 2009 Dec;4(1):50. [doi: 10.1186/1748-5908-4-50]
24. Rosborg S, Lindberg M, Ramukumba M, Jäderlund Hagstedt L, Hägglund M. Exploring mHealths fit to workflow in homecare - a case study in Sweden. Stud Health Technol Inform 2019 Aug 9;265:54-59. [doi: 10.3233/SHTI190137] [Medline: 31431577]
25. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
26. Haugen HA. Beyond Implementation: A Prescription for Lasting EMR Adoption: Magnusson Skor; 2010.
27. Bødker S, Iversen OS. Staging a professional participatory design practice: moving PD beyond the initial fascination of user involvement. 2002 Presented at: NordiCHI '02: Proceedings of the second Nordic conference on Human-computer interaction; Oct 19-23, 2002; Aarhus Denmark p. 11-18. [doi: 10.1145/572020.572023]
28. Kushniruk A, Nøhr C. Participatory design, user involvement and health IT evaluation. Stud Health Technol Inform 2016;222:139-151. [Medline: 27198099]
29. Bano M, Zowghi D, da Rimini F. User satisfaction and system success: an empirical exploration of user involvement in software development. Empir Software Eng 2017 Oct;22(5):2339-2372. [doi: 10.1007/s10664-016-9465-1]
30. Berg M. Implementing information systems in health care organizations: myths and challenges. Int J Med Inform 2001 Dec;64(2-3):143-156. [doi: 10.1016/s1386-5056(01)00200-3] [Medline: 11734382]
31. Scandurra I, Hägglund M, Koch S. From user needs to system specifications: multi-disciplinary thematic seminars as a collaborative design method for development of health information systems. J Biomed Inform 2008 Aug;41(4):557-569. [doi: 10.1016/j.jbi.2008.01.012] [Medline: 18394969]
32. Scandurra I, Hägglund M, Koch S. Application of the multi-disciplinary thematic seminar method in two homecare cases - a comparative study. Stud Health Technol Inform 2008;136:597-602. [Medline: 18487796]
33. Preece J, Sharp H, Rogers Y. Interaction Design: Beyond Human-Computer Interaction: John Wiley & Sons; 2015.
34. Scandurra I, Hägglund M, Persson A, Ahlfeldt RM. Disturbing or facilitating?--on the usability of Swedish eHealth systems 2013. Stud Health Technol Inform 2014;205:221-225. [Medline: 25160178]
35. Craig P, Dieppe P, Macintyre S, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655. [doi: 10.1136/bmj.a1655] [Medline: 18824488]
36. Scandurra I, Hägglund M, Koch S, Lind M. Usability laboratory test of a novel mobile homecare application with experienced home help service staff. Open Med Inform J 2008;2:117-128. [doi: 10.2174/1874431100802010117] [Medline: 19415140]

XSL•FO
RenderX

37. Scandurra I, Hägglund M, Engström M, Koch S. Heuristic evaluation performed by usability-educated clinicians: education and attitudes. Stud Health Technol Inform 2007;130:205-216. [Medline: 17917194]
38. Scandurra I, Hägglund M, Moström D, Koch S. Heuristic evaluation extended by user analysis: a fast and efficient method to identify potential usability problems in health information systems. J Inf Technol Healthc 2006;4(5):317-325 [FREE Full text]
39. Berg M. Patient care information systems and health care work: a sociotechnical approach. Int J Med Inform 1999 Aug;55(2):87-101. [doi: 10.1016/s1386-5056(99)00011-8] [Medline: 10530825]
40. Harvey G, Kitson A. PARIHS revisited: from heuristic to integrated framework for the successful implementation of knowledge into practice. Implement Sci 2016 Mar 10;11(1):33. [doi: 10.1186/s13012-016-0398-2] [Medline: 27013464]
41. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. BMJ 2010 Nov 16;341:c5814. [doi: 10.1136/bmj.c5814] [Medline: 21081595]
42. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated Consolidated Framework for Implementation Research based on user feedback. Implement Sci 2022 Oct 29;17(1):75. [doi: 10.1186/s13012-022-01245-0] [Medline: 36309746]
43. Scrapping the millennium: introduction of a health record in Sweden fails. Heise Online. 2025. URL: https://www.heise.de/en/news/Scrapping-the-millennium-introduction-of-a-health-record-in-Sweden-fails-10323142.html [accessed 2025-08-05]
44. Heggelund MO, Hussain SS, Farshchian BA. Information overload – a case study of using an integrated electronic health record system in the emergency room. SSRN. Preprint posted online on Jun 21, 2024. [doi: 10.2139/ssrn.4866896]
45. Hertzum M, Ellingsen G, Cajander Å. Implementing large-scale electronic health records: experiences from implementations of Epic in Denmark and Finland. Int J Med Inform 2022 Nov;167:104868. [doi: 10.1016/j.ijmedinf.2022.104868] [Medline: 36194994]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
EHR: electronic health record
HCP: health care professionals
i-PARIHS: integrated-Promoting Action on Research Implementation in Health Services
MHCO: municipality home care organization
mHealth: mobile health
PARIHS: Promoting Action on Research Implementation in Health Services
PCC: primary care center

Edited by L Buis; submitted 03.Dec.2024; peer-reviewed by W Odendaal, Y Chu; revised version received 23.Nov.2025; accepted 24.Nov.2025; published 23.Jan.2026.

Please cite as:
Jäderlund Hagstedt L, Hvitfeldt H, Hägglund M
Failed Implementation of Mobile Access to Electronic Health Records in Home Care: Qualitative Study in Sweden
JMIR Mhealth Uhealth 2026;14:e69590
URL: https://mhealth.jmir.org/2026/1/e69590
doi:10.2196/69590

© Lovisa Jäderlund Hagstedt, Helena Hvitfeldt, Maria Hägglund. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 23.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluating the Feasibility of an Electronic Patient-Reported Outcomes Platform Integrating Electronic Health Records and a Mobile Messaging App in Breast Cancer Radiotherapy: Retrospective Cross-Sectional Study

Jong Yun Baek[1*], MD; Haeyoung Kim[1*], MD, PhD; Won Kyung Cho[1], MD; Nalee Kim[1], MD, PhD; Tae Hoon Lee[1], MD, PhD; Won Chul Cha[2,3], MD, PhD

[1]Department of Radiation Oncology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
[2]Department of Digital Health, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul, Republic of Korea
[3]Data Innovation Center, Samsung Medical Center, Seoul, Republic of Korea, Seoul, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Haeyoung Kim, MD, PhD
Department of Radiation Oncology, Samsung Medical Center, Sungkyunkwan University School of Medicine
81 Irwon-ro, Gangnam-gu, Seoul
Seoul, 06351
Republic of Korea
Phone: 82 2 3410 2612
Fax: 82 2 3410 2619
Email: haeyoung0131.kim@samsung.com

Abstract

Background: Integrating electronic patient-reported outcomes (ePROs) into electronic health records (EHRs) can enhance the quality of patient care. However, collecting longitudinal ePRO data throughout treatment and posttreatment surveillance remains challenging in patients with breast cancer. To address this, we implemented an automated system that enables ePRO acquisition and seamless integration into the EHR. The system delivers questionnaire weblinks via a mobile messaging app, allowing patients to complete ePROs before clinic visits, with responses automatically transferred to the EHR.

Objective: This study aimed to assess patient response rates to the ePRO system and identify key factors influencing the response rate among patients with breast cancer who received radiotherapy and postradiotherapy follow-up.

Methods: We conducted a retrospective analysis of prospectively collected ePRO data by using the BREAST-Q questionnaire, a validated patient-reported outcome measure for breast surgery, from patients who received adjuvant radiotherapy at our institution between May 2023 and April 2024. At a preradiotherapy or postradiotherapy visit, each patient was asked to complete the questionnaire via a weblink sent to their mobile messaging app, KakaoTalk. The questionnaire was dispatched from minutes to several days before each visit. The response rate was calculated as the percentage of patients whose responses were successfully recorded in the EHR among those who were requested to respond. A complete response (CR) was defined as completion of all required questionnaire items. CR rates were analyzed according to clinical factors using univariate and multivariate logistic regression.

Results: Data from 1488 patients were analyzed, encompassing 2431 encounters (median 1, IQR 1-2 per patient). The median age of the patients was 51 (range 23-83) years, with 65.1% (n=968) patients aged 40 to 59 years. Comorbidities were present in 15% (223/1488) of the patients. The CR rate for the first, second, and third ePRO encounters was 89.9% (1338/1488), 98.3% (735/748), and 97.3% (180/185), respectively. Among first-time respondents, younger patients had a significantly higher CR rate (patients aged <60 years: 100/1104, 90.9%; patients aged ≥60 years: 334/384, 87%; *P*=.03). The timing of the questionnaire dispatch also affected the CR rate (*P*<.001). The CR rate was the highest when questionnaires were sent more than 1 hour before the visit (547/583, 93.3%) or in the afternoon of the previous day (505/545, 92.7%) and the lowest when sent 2 or more days before (100/130, 76.9%) or within 1 hour before the appointment (92/112, 81.7%). Both age (*P*=.006) and timing (*P*<.001) remained significant in the multivariate analysis.

XSL•FO
RenderX

Conclusions: This study demonstrates the feasibility of integrating ePRO into EHR through a mobile messaging app–based system, with high patient adherence. Response rates were significantly influenced by patient age and the timing of questionnaire dispatch. These findings provide practical insight for optimizing ePRO implementation in routine oncology care.

(JMIR Mhealth Uhealth 2026;14:e67514) doi:10.2196/67514

KEYWORDS

patient-reported outcomes; mobile app; mHealth; mobile messaging app; electronic health records; eHealth; telemedicine; radiotherapy; breast neoplasms

Introduction

Background

Breast radiotherapy is an essential element in the management of breast cancer, as it enables breast conservation by eliminating microscopic tumor foci following tumor resection and prevents locoregional recurrence, leading to improved survival [1,2]. With advances in multimodality treatment, survival outcomes for patients with breast cancer have substantially improved over the past decades [3,4]. As survival extends, radiotherapy regimens have evolved to minimize treatment-related toxicity while maintaining tumor control [5]. Therefore, capturing toxicity profiles and patient satisfaction is critical for optimizing radiotherapeutic approaches. To achieve this, the incorporation of patient-reported outcomes (PROs) has become increasingly important for individualized counseling and shared decision-making [6].

Despite growing consensus on the value of electronic PROs (ePROs), their routine adoption in oncology remains limited. Collecting longitudinal ePRO data across active treatment and long-term follow-up is particularly challenging, as survivors of breast cancer often require monitoring for more than a decade after completion of primary treatment [7,8]. Previous studies have highlighted several barriers to sustained ePRO use, including workflow burden on clinicians, lack of integration with electronic health records (EHRs), and patient fatigue over repeated reporting [9-12]. Most existing ePRO systems operate as stand-alone or web portal–based platforms, requiring separate log-ins or additional applications, which hinders seamless use during clinical visits [9-11]. As a result, adherence to ePRO completion varies widely, ranging from 27% to 95% across populations with cancer [13-17], and evidence on how to maintain high adherence in daily oncology practice is still lacking.

To address these gaps, we developed and implemented an automated EHR-integrated ePRO system that delivers questionnaires via KakaoTalk (Kakao Corp), the most widely used mobile messaging app in South Korea [18]. This study evaluated the feasibility of this platform by assessing response rates and identifying clinical and contextual factors associated with ePRO adherence among patients receiving postoperative adjuvant breast radiotherapy and follow-up. For an assessment of the feasibility of the ePRO system, the response rate was considered the primary indicator, as it reflects both patient adherence and the sustainability of the platform in routine clinical practice. Consistent with previous ePRO implementation studies [9,10,14,15,17], the response rate has been widely used as a practical measure of feasibility, representing patient engagement, system usability, and operational sustainability within clinical workflows. A consistently high response rate across visits would indicate that the system can be feasibly integrated into long-term follow-up, whereas lower or variable rates would highlight barriers that require further optimization. In addition, factors influencing feasibility were examined by analyzing both clinical and contextual variables that might affect adherence. Clinical variables, such as age, type of surgery, histology, and comorbidities, were considered because they may influence treatment-related experiences and patients' ability to engage with the ePRO systems. Contextual factors, including previous exposure to other ePRO systems and the timing of questionnaire requests, were analyzed because they are expected to directly influence adherence and response behavior.

Study Objective

This study aimed to evaluate the feasibility of integrating an ePRO system directly into the EHR through a mobile messaging app by assessing patient adherence and identifying clinical and contextual factors associated with ePRO completion among patients with breast cancer undergoing postoperative adjuvant radiotherapy and follow-up.

Methods

Participants and Study Design

This study was conducted at the Samsung Medical Center, a large tertiary referral hospital in South Korea, with a comprehensive cancer center located in Seoul. Located in an urban area, the institution serves as a nationwide referral center, providing care to a high volume of patients with cancer from both urban and rural regions. As of 2021, approximately 31,000 patients—representing about 11% of all patients with cancer in South Korea—were treated at the Samsung Medical Center [19]. Between May 2023 and April 2024, ePRO questionnaires were distributed to patients visiting the Department of Radiation Oncology for postoperative adjuvant radiation treatment for breast cancer. These visits included both preradiotherapy evaluations and postradiotherapy follow-ups for surveillance. A weblink connected to the ePRO questionnaires was sent to each individual's KakaoTalk mobile messaging app before their appointment with the attending physician, and patients were asked to respond to the questionnaires through the app. KakaoTalk is a free mobile messaging app used by more than 90% of the Korean population across all age groups and has been widely incorporated into both personal and institutional services. Its functions extend beyond instant messaging to include secure log-in verification, payment systems, and government or health care notifications [20,21]. Because of its

XSL•FO
RenderX

nationwide penetration and user familiarity, the app enables seamless integration of digital health tools, such as ePRO questionnaires, without requiring additional software installation or user registration. The dispatch of the questionnaires was managed by nurses or physicians through a system deployed in the EHR before each patient's visit. The timing of sending the questionnaire varied depending on the sender's preference, ranging from days to minutes before the visit, with no predefined time points. On the day of the visit, an outpatient receptionist or nurse checked whether the questionnaires had been completed. If the questionnaires were incomplete, the receptionist or nurse briefly asked the patient to complete them before the physician's session.

During the study period, the adaptation of the ePRO platform into daily clinical practice varied among the radiation oncologists in our department. Some began using the platform at the beginning of the study period, while others adopted it later. Once a radiation oncologist initiated use of the system, all eligible patients assigned to them were requested to complete the ePRO questionnaires, regardless of individual characteristics. As a result, not all patients with breast cancer visiting our department were uniformly invited to complete ePRO questionnaires. Instead, the number of patients with breast cancer undergoing radiotherapy who were requested to submit ePROs increased toward the latter part of the study period.

After an outpatient visit for preradiotherapy evaluation, patients received radiation treatment according to our institutional protocol. Radiotherapy was administered once daily for 5 consecutive days, with 3 to 19 fractionations over a period of 1 to 4 weeks. Fractionation schedules were determined based on tumor stage, surgery types, the inclusion of regional nodal irradiation, and other risk factors. After completing the treatment, patients were followed up for 2 to 3 weeks after treatment and subsequently every 6 months. ePRO data were collected at the preradiotherapy visit, the immediate postradiotherapy visit at 2 to 3 weeks after treatment, and every 6 months thereafter. As each patient was required to respond to a questionnaire at each hospital visit, 1 or more questionnaires were requested to be completed by each patient during the period of this study.

A System for ePRO Collection and Storage in the EHR

A system for ePRO acquisition and integration with the EHR was developed as an in-house model at our institution. The system links our hospital's EHR with the individual's mobile messaging app for collecting and storing ePRO data. Data entry is performed through the messaging app on the patient's mobile phone, while the data presentation and storage are conducted in our hospital's EHR. A section dedicated to ePRO is integrated into the EHR, allowing physicians and other medical staff to send new ePRO questionnaires on request and view each patient's responses at any time (Figure S1 in Multimedia Appendix 1). When medical staff select ePRO questionnaires and dispatch them through the EHR, a weblink for the ePRO questionnaires is sent to the patient's messaging app. The patient can open the link by entering their date of birth and submit their response to the questionnaire, which is in the form of checkboxes (Figure S2 in Multimedia Appendix 1). The patient's responses are automatically transferred to the ePRO section of the EHR and stored in the hospital's data warehouse. This process of entering, transferring, and storing ePRO data occurs simultaneously in real time, enabling physicians to view the content and time stamp of the data in the EHR.

For the PRO instrument used in this study, we used the Korean version of BREAST-Q (version 2.0) postoperative scale, including modules for breast-conserving therapy, mastectomy, and reconstruction. Among the domains of these modules, the following were used for our patients: satisfaction with breasts, satisfaction with implants, physical well-being of the chest or upper body, and adverse effects of radiation [22,23]. Patients who visited for preradiotherapy evaluation were asked to complete a questionnaire without the domain of adverse effects of radiation, while those attending for postradiotherapy surveillance received a questionnaire that included the domain.

Assessment of the Response Rate and Influencing Factors

Response rate was calculated as the percentage of patients whose responses were successfully recorded in the EHR among those who were requested to respond to the questionnaires. The response rate for each survey encounter was assessed and compared according to the number of encounters, from the first to the third. Because the current analysis was based on surveys conducted over 1 year, most of our patients encountered the questionnaires 1 to 3 times according to the scheduled follow-up interval. We classified response status into 3 categories: complete response (CR; all questions answered), partial response (PR; at least 1 question answered but not all), and no response (NR). In addition, when analyzing significant factors influencing CR, we divided our patients into 2 groups: complete responders and noncomplete responders, with partial responders and nonresponders merged into the noncomplete responder group.

To determine the significant factors influencing CR, we compared the CR rate according to various factors, focusing only on patients who encountered the survey for the first time. Specifically, to assess the impact of questionnaire request timing before a visit appointment, the timing was categorized into 5 groups: within 1 hour of the appointment time (≤1 hour on the day), more than 1 hour before the appointment on the visit day (>1 hour on the day), in the afternoon of the day before the appointment (PM the day before), in the morning of the day before the appointment (AM the day before), and 2 or more days before the appointment day (≥2 days before). In addition, patients who had previous experiences responding to questionnaires requested from other departments in our hospital were categorized as previous other ePRO (+), while those without the experiences were categorized as previous other ePRO (–). Finally, patients with any of the following diseases were categorized as having comorbidity: diabetes, cardiovascular disease, chronic pulmonary disease, hepatic disease, renal disease, or other cancers.

Statistical Analysis

For the univariate analysis, Fisher exact test was used to compare the CR according to clinical and contextual variables, including age, type of surgery, histology, comorbidities, previous

XSL•FO
RenderX

ePRO experience, and the timing of questionnaire requests. Age was categorized using a 60-year cutoff for univariate analysis and treated as a continuous variable in multivariate models. Patients with missing data for a given variable were excluded from the study. Factors with $P<.10$ in univariate analyses were included in the multivariate logistic regression model to adjust for potential confounding effects. All P values were 2-sided, with $P<.05$ considered statistically significant. All analyses were performed using SPSS software (version 27; IBM Corp).

Ethical Considerations

This study was approved by the institutional review board of the Samsung Medical Center (institutional review board approval number 2024-07-147-001). The analysis was conducted retrospectively using deidentified ePRO data collected during routine clinical care, and the need for informed consent was waived by the institutional review board. All patient information was deidentified before analysis to ensure privacy and confidentiality. No personal identifiers were included in the downloaded dataset. No compensation was provided to patients for participation, as the data were collected as part of standard clinical procedures without additional burden or intervention.

Results

Patients' Characteristics

A total of 2334 patients with breast cancer attended our department during the study period. Of the 2334 patients, 1491 (63.9%) with nonmetastatic breast cancer were invited to complete the ePRO survey. This difference reflects the gradual adoption of the ePRO system among physicians, as the platform was implemented in stages and not all attending radiation oncologists had begun using it at the start of the study period. A total of 3 (0.2%) patients were excluded from the analysis due to missing time records of their questionnaire responses, resulting in 1488 (99.8%) patients being included in this study. Among them, 740 (49.7%) encountered 1 survey, 563 (37.8%) encountered 2 surveys, 175 (11.8%) encountered 3 surveys, and 10 (0.7%) encountered 4 surveys, resulting in 2431 survey encounters.

The characteristics of the 1488 patients are summarized in Table 1. The median age was 51 years, with most of the patients (n=968, 65.1%) aged between 40 and 59 years, and 41 (2.8%) patients were aged 75 years or older. Most patients underwent breast-conserving surgery (n=1186, 79.7%) and had invasive carcinoma (n=1342, 90.2%). Comorbidities were found in 223 (14.9%) patients, including diabetes (n=98, 6.6%); cardiovascular disease (n=67, 4.5%); chronic disease of the liver, lung, or kidney (n=63, 4.2%); and other cancers (n=26, 1.7%).

XSL•FO
RenderX

Table 1. Patients' characteristics (N=1488).

Characteristics	Patients
Age (y), median (range)	51 (24-85)
Age (y), n (%)	
≥20 to <40	136 (9.1)
≥40 to <60	968 (65.1)
≥60	384 (25.8)
Sex, n (%)	
Female	1487 (99.9)
Male	1 (0.1)
Type of surgery, n (%)	
Breast-conserving surgery	1186 (79.7)
Mastectomy without reconstruction	129 (8.7)
Mastectomy with reconstruction	173 (11.6)
Histology, n (%)	
Ductal carcinoma in situ	146 (9.8)
Invasive carcinoma	1342 (90.2)
Type of visit, n (%)	
Preradiotherapy evaluation visit	946 (63.6)
Postradiotherapy follow-up visit	542 (36.4)
Comorbidity, n (%)	
Yes	223 (15)
No	1265 (85)
Experience of other electronic patient-reported outcomes, n (%)	
Yes	392 (26.3)
No	1096 (73.7)
Timing of the questionnaire requests[a], n (%)	
≤1 h on the day	115 (7.7)
>1 h on the day	586 (39.4)
PM the day before	545 (36.6)
AM the day before	112 (7.5)
≥2 d before	130 (8.7)

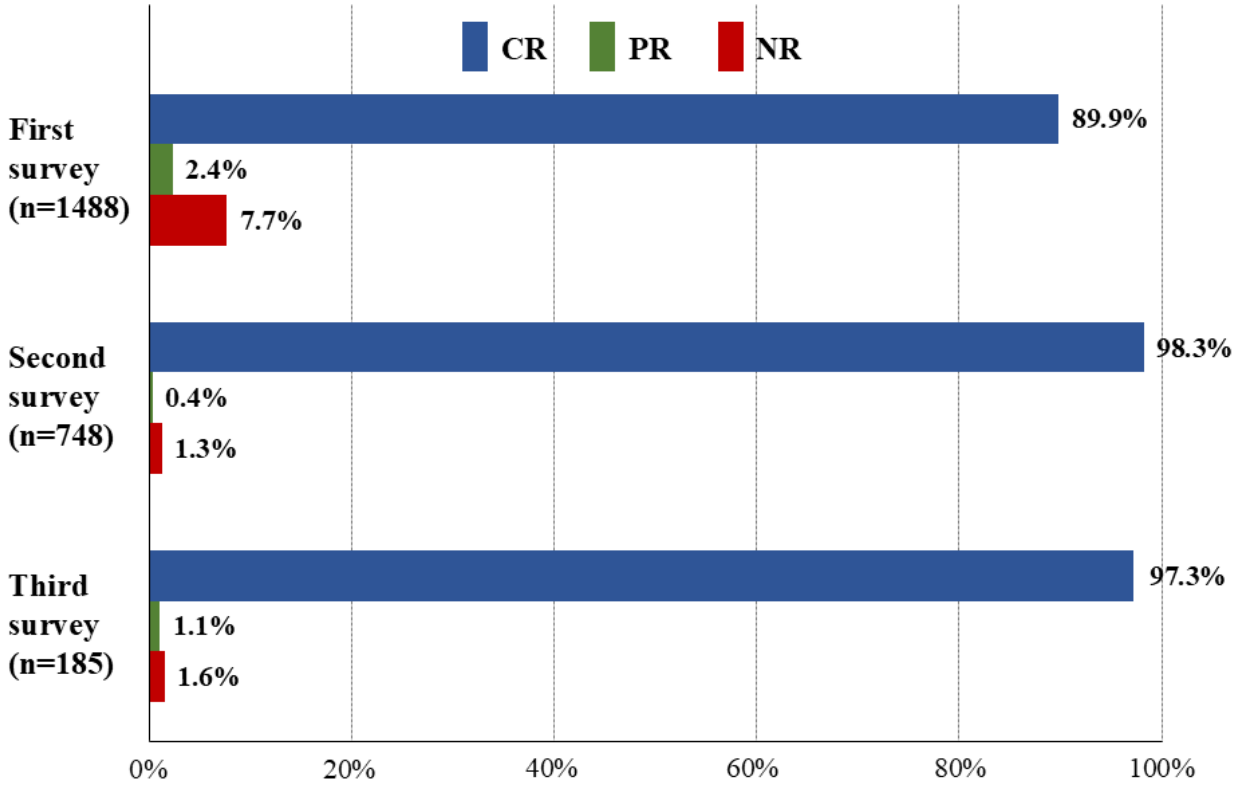
[a]The timing of the questionnaire requests before a visit appointment was categorized as follows: within 1 hour of the appointment time (≤1 hour on the day), more than 1 hour before the appointment on the visit day (>1 hour on the day), in the afternoon of the day before the appointment (PM the day before), in the morning of the day before the appointment (AM the day before), and 2 or more days before the appointment day (≥2 days before).

Response Rate and Influencing Factors

Of the 1488 patients who encountered the questionnaires for the first time, 1338 (89.9%) exhibited CR, 35 (2.4%) submitted PR, and 115 (7.7%) did not respond to the questionnaire. The response status and rates according to the number of survey encounters are shown in Figure 1.

XSL•FO
RenderX

Figure 1. Rates of complete response (CR), partial response (PR), and no response (NR) for electronic patient-reported outcomes surveys across first, second, and third requests.



In the univariate analysis of factors influencing CR, the following were statistically significant: patient's age, type of visit, and the timing of the questionnaire requests before a visit appointment (Table 2). The CR rate was 90.9% (1004/1104) in patients younger than 60 years and 87% (334/384) in those aged 60 years or older (P=.03), indicating that older age was associated with lower completion. A significantly lower CR rate was also observed in patients receiving the questionnaire at the preradiotherapy visit (837/946, 88.5%) compared to the postradiotherapy visits (501/542, 92.4%; P=.02).

XSL•FO
RenderX

Table 2. Univariate analysis of factors affecting the complete response to the electronic patient-reported outcome questionnaire (N=1488).

Characteristics	CR[a], n (%)	Non-CR, n (%)	*P* value
Age (y)			.03
<60	1004 (90.9)	100 (9.1)	
≥60	334 (87)	50 (13)	
Type of surgery			.10
Breast-conserving surgery	1057 (89.1)	129 (10.9)	
Mastectomy without reconstruction	122 (94.6)	7 (5.4)	
Mastectomy with reconstruction	159 (91.9)	14 (8.1)	
Histology			.77
Ductal carcinoma in situ	133 (91.1)	13 (8.9)	
Invasive carcinoma	1205 (89.8)	137 (10.2)	
Type of visit			.02
Preradiotherapy evaluation visit	837 (88.5)	109 (11.5)	
Postradiotherapy follow-up visit	501 (92.4)	41 (7.6)	
Comorbidity			.55
Yes	198 (88.8)	25 (11.2)	
No	1140 (90.1)	125 (9.9)	
Experience of other electronic patient-reported outcome			.85
Yes	354 (90.3)	38 (9.7)	
No	984 (89.8)	112 (10.2)	
			<.001
≤1 h on the day	94 (81.7)	21 (18.3)	
>1 h on the day	547 (93.3)	39 (6.7)	
PM the day before	505 (92.7)	40 (7.3)	
AM the day before	92 (82.1)	20 (17.9)	
≥2 d before	100 (76.9)	30 (23.1)	

[a]CR: complete response.

[b]The timing of the questionnaire requests before a visit appointment was categorized as follows: within 1 hour of the appointment time (≤1 hour on the day), more than 1 hour before the appointment on the visit day (>1 hour on the day), in the afternoon of the day before the appointment (PM the day before), in the morning of the day before the appointment (AM the day before), and 2 or more days before the appointment day (≥2 days before).

Notably, the timing of the questionnaire requests had a strong influence on CR. Patients who received the questionnaire more than 1 hour before the appointment or in the afternoon of the previous day showed the highest CR rates at 93.3% (547/583) and 92.7% (505/545), respectively. In contrast, the lowest rates were seen in those who received the questionnaire 2 or more days before (100/130, 76.9%) and within 1 hour of the appointment (94/115, 81.7%; *P*<.001).

In the multivariate analysis, age and the timing of the questionnaire requests remained significant influencing factors for CR. Old age (odds ratio 0.98, 95% CI 0.96-0.99; *P*=.006) and questionnaire request timing of ≤1 hour on the day of the visit, AM the day before, or ≥2 days before the appointment (odds ratio 0.32, 95% CI 0.22-0.45; *P*<.001) were significantly associated with a lower CR rate (Table 3).

XSL•FO
RenderX

Table 3. Multivariate analysis of factors affecting the complete response to the electronic patient-reported outcome questionnaire.

Characteristics	Odds ratio (95% CI)	*P* value
Age (y): continuous	0.98 (0.96-0.99)	.006
Type of surgery: breast-conserving surgery vs mastectomy or reconstruction	0.64 (0.39-1.05)	.08
Type of visit: preradiotherapy evaluation visit vs postradiotherapy follow-up visit	1.33 (0.90-1.97)	.15
Timing of the questionnaire requests[a]: ≤1 h on the day or AM the day before or ≥2 d before vs >1 h on the day or PM the day before	0.32 (0.22-0.45)	<.001

[a]The timing of ePRO requests before a visit appointment was categorized as follows: within 1 hour of the appointment time (≤1 hour on the day), more than 1 hour before the appointment on the visit day (>1 hour on the day), in the afternoon of the day before the appointment (PM the day before), in the morning of the day before the appointment (AM the day before), and 2 or more days before the appointment day (≥2 days before).

Discussion

Principal Findings

Patients using our ePRO system, which links a commercial mobile messaging app with our hospital's EHR, showed an 89.9% (1338/1488) CR rate for the BREAST-Q questionnaires. Responses of patients who visited our radiation oncology department for breast cancer management were successfully recorded. The response rate to the questionnaires increased as the number of survey encounters increased. Age of 60 years or older was associated with a lower rate of CR; however, 87% (334/384) of participants of this age group provided appropriate responses to the questionnaires delivered through their mobile messaging app, even if it was their first time encountering the questionnaires using the app. Notably, the timing of the questionnaire requests significantly influenced the CR rate, with a higher CR rate of more than 92% observed when the questionnaires were requested more than 1 hour before the scheduled visit or in the afternoon of the day before the appointment. Given these findings, our ePRO system shows potential as a feasible platform for ePRO collection and integration with the hospital's EHR. In addition, the factors identified as significantly affecting the CR rate of the questionnaires can be used to guide improvements in responses to ePRO questionnaires in daily practice.

Comparison With Prior Work

The significance of PRO measurement in oncology care has been increasingly emphasized in recent years [24]. In this regard, the European Society for Medical Oncology released a clinical practice guideline concerning the use of PRO in the continuum of cancer care, emphasizing the essential role of symptom monitoring via PRO measurements [25]. ePRO offers several advantages, such as greater patient preference and acceptability, lower human resource costs, and higher data quality [9,11]. Various forms of ePRO collection platforms, including web-based and app-based systems, have been developed and used [13,14,26]. In addition to the data collection system, integrating the data into the EHR is essential to facilitate the incorporation of ePRO into clinical practice [9-12].

In this analysis, we found favorable patient adherence to our ePRO system, with more than 89.9% (1338/1488) CR rate for the ePRO questionnaires, even at the first encounter. This promising result is attributed to adopting a system that uses the KakaoTalk messaging app, which was familiar to our patients [18]. Patients were able to access the questionnaires using the existing app on their smartphone, without the need to install an additional app for ePRO. Because the message with the questionnaire link was sent under the hospital's name, our patients likely accepted it confidently, without concerns about cybercrimes. Moreover, given that more than 94% of Koreans own a smartphone, ePRO questionnaires delivered via the mobile app could effectively encourage responses from our patient population [27]. According to previous studies, ePRO adherence rate among patients with cancer has been reported to range between 27% and 95% [13-17]. In a study conducted in the United States, ePRO adherence rates ranged from 27% to 70%, following administration either on-site via tablet or remotely via a patient portal. There were significant differences in response rates depending on patient age and race, with older patients aged 65 years and older and non-White individuals being negatively associated with adherence to ePRO [14]. Meanwhile, a Japanese study reported a 95% ePRO adherence rate via a mobile messaging app, LINE, from 40 participants, which was similar to our patients' adherence rate [16]. Considering that LINE is used by more than 70% of the Japanese population, the familiarity with the ePRO acquisition tool likely contributed to the high ePRO adherence in their study [28]. Taken together, these findings suggest that selecting appropriate tools for ePRO administration based on respondents' demographic or cultural characteristics is essential for achieving favorable ePRO acceptance.

Older age has been reported to be significantly associated with lower adherence to ePRO [13,14,17]. In a prospective study conducted among French patients aged 75 years and older, 26% of the participants accepted ePRO, which was conducted remotely using a web-based app [13]. More than 52% of the participants did not respond to the ePRO due to technological issues, such as a lack of internet access or discomfort with using the internet [13]. In addition, a study performed in the United States showed that patients aged 65 years or older exhibited a 6% decrease in adherence to ePRO, which was a significant difference compared to younger patients [14]. Similarly, in our study, patients aged 60 years or older showed a significantly lower CR rate for ePRO than those younger than 60 years. However, considering that 87% (334/384) of the older participants completed the questionnaires, the ePRO acceptance using our system appears favorable even among older patients. In our study, ePRO was requested from all patients attending

XSL•FO
RenderX

our department, without selection based on their smartphone possession or daily internet use. Furthermore, given that more than 90% of Koreans aged 60 years or older use smartphones [29,30] and most are reported to be familiar with KakaoTalk [31], the ePRO acceptance among our older participants likely reflects the real-world feasibility of implementing ePRO in clinical practice for older Korean patients, particularly when it is delivered through the familiar messaging app. In the meantime, we also found that 13% of patients in this older age group did not properly respond to ePRO, suggesting that there is room for improvement in enhancing ePRO adherence among older patients. It is uncertain why these patients did not respond to the ePRO measurements, as we did not assess the reasons for nonresponse to the questionnaires. However, referencing previous studies, various factors have been identified that affect ePRO acceptance in older patients, including frailty level, socioeconomic status, technological barriers, and the modes of ePRO administration [13,30,32]. Future studies are needed to determine the causes of nonadherence to ePRO and to provide the most appropriate ePRO collection modalities based on the individual characteristics of older Korean patients undergoing cancer treatments.

Optimizing ePRO Response Through Timing

Interestingly, we found that the timing of ePRO requests was significantly related to the patient's CR rate. The most appropriate time for requesting ePRO questionnaires was either more than 1 hour before the appointment or in the afternoon on the day before the scheduled visit. This finding suggests that patients may feel more comfortable and have sufficient time to review and respond to ePRO requests when they are delivered within this timeframe. Delivering ePRO questionnaires more than 2 days before a scheduled visit may have hindered appropriate responses, possibly due to difficulties in locating our ePRO request message among other personal messages. As our patients' ePROs were collected remotely using a mobile app and their completeness was rechecked on-site in our clinic, it is likely that our ePRO acceptance rate is higher than that reported in settings relying solely on remote ePROs collection. This may be indirectly supported by the CR rate of 76.9% (100/130) among our patients who received ePRO questionnaires more than 2 days before an appointment. However, from another perspective, the combination of remote ePRO collection via a messaging app, requested within a specific timeframe, and on-site feedback appears to be an effective strategy for maximizing ePRO acceptance, as indicated by a CR rate of more than 93.8% (547/583) among our patients.

Strengths and Limitations

A key strength of this study is the provision of real-world data on ePRO adherence and the identification of significant factors influencing adherence among 1488 patients with nonmetastatic breast cancer following implementation of an in-house ePRO platform. Because our ePRO platform uses a messaging app that is largely familiar to Koreans, we observed a favorable acceptance for the ePRO. However, we acknowledge the limitations of this study. Our data were retrospectively derived from a single institution over a 1-year data collection period. Therefore, the data may be insufficient to capture long-term ePRO adherence among general patients with breast cancer. In addition, reasons for PR or NR to ePRO questionnaires were not available, as this analysis was conducted retrospectively. Because the causes of incomplete ePRO are important for identifying areas of improvement in enhancing ePRO adherence, further assessments are necessary among those submitting PR or NR to ePRO questionnaires in future studies. Furthermore, because our ePRO platform relies on a specific mobile messaging app—KakaoTalk, which is predominantly used in South Korea—its general acceptance in other countries remains uncertain. Therefore, the generalizability of our findings to other populations with different digital habits may be limited. In addition, technological barriers or disparities in mobile access may have also influenced patient participation and response accuracy. Moreover, as all our ePRO data were self-reported, reporting bias may have been introduced. Future research involving institutions that use different platforms or serve more diverse populations is needed to explore the generalizability of our findings.

Conclusions

The collection of ePRO data and its integration into EHR was successful with our ePRO platform, achieving an overall CR rate of 89.9% (1338/1488). Patient age and the specific timeframe for ePRO requests were significant factors influencing complete ePRO acceptance. Patients aged 60 years or older showed significantly lower ePRO adherence. In addition, a specific timeframe, including more than 1 hour before a clinical visit or in the afternoon on the day before the appointment, was associated with a significantly higher CR rate for ePRO. These factors are expected to improve ePRO acceptance among patients with nonmetastatic breast cancer.

Funding

This study was supported by grants from the National Research Foundation of Korea funded by the Ministry of Education (RS-2023-00252283) and from the Samsung Medical Center (SMO1240501).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

HK, WKC, and NK planned the study. HK, WKC, NK, and THL conducted the electronic patient-reported outcomes survey. JYB carried out data collection and analysis. JYB and HK drafted the manuscript. WCC designed the electronic patient-reported outcomes system. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots illustrating the workflow of the electronic health record–integrated electronic patient-reported outcome system and the patient questionnaire interface.
[DOCX File , 522 KB - mhealth_v14i1e67514_app1.docx]

References

1. Early Breast Cancer Trialists' Collaborative Group (EBCTCG), Darby S, McGale P, Correa C, Taylor C, Arriagada R, et al. Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10,801 women in 17 randomised trials. Lancet 2011 Nov 12;378(9804):1707-1716 [FREE Full text] [doi: 10.1016/S0140-6736(11)61629-2] [Medline: 22019144]
2. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Radiotherapy to regional nodes in early breast cancer: an individual patient data meta-analysis of 14 324 women in 16 trials. Lancet 2023 Nov 25;402(10416):1991-2003 [FREE Full text] [doi: 10.1016/S0140-6736(23)01082-6] [Medline: 37931633]
3. Wojtyla C, Bertuccio P, Wojtyla A, La Vecchia C. European trends in breast cancer mortality, 1980-2017 and predictions to 2025. Eur J Cancer 2021 Jul;152:4-17 [FREE Full text] [doi: 10.1016/j.ejca.2021.04.026] [Medline: 34062485]
4. Giaquinto AN, Sung H, Miller KD, Kramer JL, Newman LA, Minihan A, et al. Breast cancer statistics, 2022. CA Cancer J Clin 2022 Nov 03;72(6):524-541 [FREE Full text] [doi: 10.3322/caac.21754] [Medline: 36190501]
5. Shah C, Al-Hilli Z, Vicini F. Advances in breast cancer radiotherapy: implications for current and future practice. JCO Oncol Pract 2021 Dec;17(12):697-706. [doi: 10.1200/OP.21.00635] [Medline: 34652952]
6. Alcorn SR, Corbin KS, Shumway DA. Integrating the patient's voice in toxicity reporting and treatment decisions for breast radiotherapy. Semin Radiat Oncol 2022 Jul;32(3):207-220. [doi: 10.1016/j.semradonc.2022.01.010] [Medline: 35688519]
7. Pedersen RN, Esen BO, Mellemkjær L, Christiansen P, Ejlertsen B, Lash TL, et al. The incidence of breast cancer recurrence 10-32 years after primary diagnosis. J Natl Cancer Inst 2022 Mar 08;114(3):391-399 [FREE Full text] [doi: 10.1093/jnci/djab202] [Medline: 34747484]
8. De Rose F, Meduri B, De Santis MC, Ferro A, Marino L, Colciago RR, et al. Rethinking breast cancer follow-up based on individual risk and recurrence management. Cancer Treat Rev 2022 Sep;109:102434 [FREE Full text] [doi: 10.1016/j.ctrv.2022.102434] [Medline: 35933845]
9. Cracchiolo JR, Arafat W, Atreja A, Bruckner L, Emamekhoo H, Heinrichs T, et al. Getting ready for real-world use of electronic patient-reported outcomes (ePROs) for patients with cancer: a National Comprehensive Cancer Network ePRO Workgroup paper. Cancer 2023 Aug 15;129(16):2441-2449 [FREE Full text] [doi: 10.1002/cncr.34844] [Medline: 37224181]
10. Hassett MJ, Cronin C, Tsou TC, Wedge J, Bian J, Dizon DS, et al. eSyM: an electronic health record-integrated patient-reported outcomes-based cancer symptom management program used by six diverse health systems. JCO Clin Cancer Inform 2022 Jan;6:e2100137 [FREE Full text] [doi: 10.1200/CCI.21.00137] [Medline: 34985914]
11. Meirte J, Hellemans N, Anthonissen M, Denteneer L, Maertens K, Moortgat P, et al. Benefits and disadvantages of electronic patient-reported outcome measures: systematic review. JMIR Perioper Med 2020 Apr 03;3(1):e15588 [FREE Full text] [doi: 10.2196/15588] [Medline: 33393920]
12. Salmani H, Nasiri S, Ahmadi M. The advantages, disadvantages, threats, and opportunities of electronic patient-reported outcome systems in cancer: a systematic review. Digit Health 2024;10:20552076241257146 [FREE Full text] [doi: 10.1177/20552076241257146] [Medline: 38812853]
13. Cancel M, Sauger C, Biogeau J, Dardaine-Giraud V, Lecomte T, Solub D, et al. FASTOCH: feasibility of electronic patient-reported outcomes in older patients with cancer-a multicenter prospective study. J Clin Oncol 2024 Aug 01;42(22):2713-2722. [doi: 10.1200/JCO.23.02150] [Medline: 38709983]
14. Takvorian SU, Anderson RT, Gabriel PE, Poznyak D, Lee S, Simon S, et al. Real-world adherence to patient-reported outcome monitoring as a cancer care quality metric. JCO Oncol Pract 2022 Sep;18(9):e1454-e1465. [doi: 10.1200/OP.21.00855] [Medline: 35675586]
15. Takala L, Kuusinen TE, Skyttä T, Kellokumpu-Lehtinen P, Bärlund M. Electronic patient-reported outcomes during breast cancer adjuvant radiotherapy. Clin Breast Cancer 2021 Jun;21(3):e252-e270 [FREE Full text] [doi: 10.1016/j.clbc.2020.10.004] [Medline: 33229222]

XSL•FO
RenderX

16. Hayashida T, Nagayama A, Seki T, Takahashi M, Matsumoto A, Kubota A, et al. Feasibility study on collecting patient-reported outcomes from breast cancer patients using the LINE-ePRO system. Cancer Sci 2022 May;113(5):1722-1730 [FREE Full text] [doi: 10.1111/cas.15329] [Medline: 35279907]
17. El Shafie RA, Weber D, Bougatf N, Sprave T, Oetzel D, Huber PE, et al. Supportive care in radiotherapy based on a mobile app: prospective multicenter survey. JMIR Mhealth Uhealth 2018 Aug 30;6(8):e10916 [FREE Full text] [doi: 10.2196/10916] [Medline: 30166275]
18. Digital 2024: South Korea. Kepios. URL: https://datareportal.com/reports/digital-2024-south-korea [accessed 2025-05-29]
19. Cancer center outcomes and introduction. Samsung Medical Center Cancer Hospital. URL: http://www.samsunghospital.com/home/cancer/intro.do?view=OUTCOMES [accessed 2025-05-29]
20. KakaoTalk overview. Kakao Corp. URL: https://www.kakaocorp.com/service/KakaoTalk?lang=en [accessed 2026-01-19]
21. Kim B. South Korea's Megacorp and super app: Kakao's paths to market dominance. Media Cult Soc 2024 Nov 16;47(4):641-659. [doi: 10.1177/01634437241294207]
22. Seth I, Seth N, Bulloch G, Rozen WM, Hunter-Smith DJ. Systematic review of Breast-Q: a tool to evaluate post-mastectomy breast reconstruction. Breast Cancer (Dove Med Press) 2021;13:711-724 [FREE Full text] [doi: 10.2147/BCTT.S256393] [Medline: 34938118]
23. Liu LQ, Branford OA, Mehigan S. BREAST-Q measurement of the patient perspective in oncoplastic breast surgery: a systematic review. Plast Reconstr Surg Glob Open 2018 Aug;6(8):e1904 [FREE Full text] [doi: 10.1097/GOX.0000000000001904] [Medline: 30254830]
24. Howell D, Molloy S, Wilkinson K, Green E, Orchard K, Wang K, et al. Patient-reported outcomes in routine cancer clinical practice: a scoping review of use, impact on health outcomes, and implementation factors. Ann Oncol 2015 Sep;26(9):1846-1858 [FREE Full text] [doi: 10.1093/annonc/mdv181] [Medline: 25888610]
25. Di Maio M, Basch E, Denis F, Fallowfield LJ, Ganz PA, Howell D, ESMO Guidelines Committee. The role of patient-reported outcome measures in the continuum of cancer clinical care: ESMO Clinical Practice guideline. Ann Oncol 2022 Sep;33(9):878-892 [FREE Full text] [doi: 10.1016/j.annonc.2022.04.007] [Medline: 35462007]
26. Basch E, Schrag D, Henson S, Jansen J, Ginos B, Stover AM, et al. Effect of electronic symptom monitoring on patient-reported outcomes among patients with metastatic cancer: a randomized clinical trial. JAMA 2022 Jun 28;327(24):2413-2422 [FREE Full text] [doi: 10.1001/jama.2022.9265] [Medline: 35661856]
27. Yoon JS. Smartphone usage in South Korea - statistics and facts. Statista. URL: https://www.statista.com/topics/5340/smartphone-usage-in-south-korea/#topicOverview [accessed 2025-05-29]
28. Why is LINE the most popular social media app in Japan? Digital Marketing for Asia. URL: https://www.digitalmarketingforasia.com/why-line-is-the-most-popular-social-media-app-in-japan/ [accessed 2025-05-29]
29. Kang H, Baek J, Chu SH, Choi J. Digital literacy among Korean older adults: a scoping review of quantitative studies. Digit Health 2023;9:20552076231197334 [FREE Full text] [doi: 10.1177/20552076231197334] [Medline: 37654708]
30. Lee H, Choi JY, Kim SW, Ko KP, Park YS, Kim KJ, et al. Digital health technology use among older adults: exploring the impact of frailty on utilization, purpose, and satisfaction in Korea. J Korean Med Sci 2024 Jan 08;39(1):e7 [FREE Full text] [doi: 10.3346/jkms.2024.39.e7] [Medline: 38193326]
31. Kim K, Lee B, Park Y, Jung EY, Kim SK, Suh DH, et al. Factors encouraging mobile instant messaging service use in medical education. PeerJ 2019;7:e7275 [FREE Full text] [doi: 10.7717/peerj.7275] [Medline: 31328039]
32. Appleyard SE, Larkin MJ, Stewart EM, Minton O, Gilbert DC. Digital medicine in men with advanced prostate cancer - a feasibility study of electronic patient-reported outcomes in patients on systemic treatment. Clin Oncol (R Coll Radiol) 2021 Dec;33(12):751-760. [doi: 10.1016/j.clon.2021.04.008] [Medline: 33966948]

Abbreviations

CR: complete response
EHR: electronic health record
ePRO: electronic patient-reported outcome
NR: no response
PR: partial response
PRO: patient-reported outcome

XSL•FO
RenderX

Edited by L Buis; submitted 14.Oct.2024; peer-reviewed by ME Arayici, KL Mauco; comments to author 01.Apr.2025; accepted 16.Jan.2026; published 05.Feb.2026.

Please cite as:
Baek JY, Kim H, Cho WK, Kim N, Lee TH, Cha WC
Evaluating the Feasibility of an Electronic Patient-Reported Outcomes Platform Integrating Electronic Health Records and a Mobile Messaging App in Breast Cancer Radiotherapy: Retrospective Cross-Sectional Study
JMIR Mhealth Uhealth 2026;14:e67514
URL: https://mhealth.jmir.org/2026/1/e67514
doi:10.2196/67514
PMID:

©Jong Yun Baek, Haeyoung Kim, Won Kyung Cho, Nalee Kim, Tae Hoon Lee, Won Chul Cha. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 05.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Use of an AI Image Recognition–Based Smartphone App in Personalized Bowel Preparation Before Colonoscopy: Prospective Randomized Controlled Trial

Fei Han[1*], MM; Shilin Gao[1*], MM; Hanqiang Zhan[1], MM; Nan Wang[1], MM; Zhaoxiang Song[1], MM; Jinlin Xie[1], MM; Yuhan Hou[1], MM; Yimeng Deng[1], MM; Jianning Yao[1], MD

The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

[*]these authors contributed equally

Corresponding Author:
Jianning Yao, MD
The First Affiliated Hospital of Zhengzhou University
1 East Construction Road
Zhengzhou, 450000
China
Phone: 86 13733183434
Email: rjyy@zzu.edu.cn

Abstract

Background: Sufficient bowel preparation is critical for increasing the quality of colonoscopy. However, current bowel preparation guidelines have limitations. We used a constructed and validated convolutional neural network model to assist patients in assessing the adequacy of bowel preparation and guide the use of laxatives based on individual variations.

Objective: This research intended to explore the influence of AI-guided bowel preparation on patients' use of laxatives and bowel cleansing quality.

Methods: We conducted a prospective randomized controlled trial that included patients who had a colonoscopy at the First Affiliated Hospital of Zhengzhou University and used sodium phosphate solution for bowel cleansing. Included individuals were randomly categorized into 2 groups: the AI-guided group, which used an AI program to assess the adequacy of bowel cleansing and stopped the use of laxatives once the assessment was passed, and the control group, which evaluated bowel preparation adequacy based on standard practices and used the recommended doses of laxatives. The primary outcomes were the amount of laxatives used and the rate of adequate bowel cleansing based on the Boston Bowel Preparation Scale (BBPS). Other outcomes included overall BBPS scores, BBPS scores by colon region, adenoma detection rate (ADR), polyp detection rate, and advanced ADR, as well as adverse reactions and compliance.

Results: This study included 513 patients (AI-guided group: n=255, 49.7%; control group: n=258, 50.3%). The volume of bowel-cleansing solution consumed by patients in the AI-guided group was 1245.5 (SD 377.2) mL, significantly lower than the 1586.2 (SD 50.4) mL in the control group (P<.001). Reduced use of laxatives was independently predicted by younger age, lower BMI, and higher defecation frequency. The rate of adequate bowel preparation in the AI group was not inferior to that in the control group (238/255, 93.3% vs 229/258, 88.8%, respectively; 1-sided P<.001 for noninferiority). No significant differences were observed in segmental or total BBPS scores, ADR, polyp detection rate, and advanced ADR between the 2 groups. Additionally, the medication adherence rate in the AI-guided group was significantly higher than in the control group (247/255, 96.9% vs 236/258, 91.5%, respectively; P=.009), whereas the incidence of adverse reactions was notably lower (5/255, 2% vs 15/258, 5.8%, respectively; P=.02).

Conclusions: Using the AI-driven bowel preparation program to assist patients in personalized bowel preparation can reduce unnecessary use of laxatives while achieving results comparable to those of standard practices.

Trial Registration: ClinicalTrials.gov NCT06610630; https://clinicaltrials.gov/study/NCT06610630

(JMIR Mhealth Uhealth 2026;14:e77701) doi:10.2196/77701

KEYWORDS

bowel preparation; colonoscopy; artificial intelligence; AI; smartphone

XSL•FO
RenderX

Introduction

Colorectal cancer (CRC) is a malignant neoplasm that originates from the epithelial lining of the colon and rectum, with incidence and mortality rates rising each year [1]. Approximately 90% of CRCs develop from colorectal polyps, particularly adenomatous polyps [2,3]. If polyps are removed during the polyp stage via colonoscopy, this can prevent 70% to 90% of CRCs [4,5]. However, there is currently a high rate of missed polyps during colonoscopy. Research indicates that the miss rate for polyps after colonoscopy might reach 22% to 28%, with adenoma miss rates ranging from 12% to 26% [6]. These missed polyps and adenomas increase the risk of developing CRC [7,8]. Thus, it is evident that high-quality colonoscopy is essential for preventing CRC [9].

Quality control indicators for colonoscopy encompass adenoma detection rate (ADR), withdrawal duration, cecal intubation rate, and bowel preparation quality [10,11]. Among these, sufficient bowel preparation is fundamental for enhancing colonoscopy quality [12]. Although guidance through images and patient education can generally meet the requirements for bowel preparation, 20% to 25% of patients still have inadequate bowel preparation [13,14]. The underlying reason is that the adequacy of bowel preparation currently relies heavily on the patients' subjective judgment, which can vary between observers. Therefore, a more objective and consistent assessment of bowel preparation, along with personalized guidance, is urgently needed in clinical practice.

AI has emerged as an auxiliary tool for this purpose. AI applications can use image recognition technology to assess bowel preparation adequacy in real time through a smartphone, allowing for objective assessment of bowel preparation and personalized guidance for cleansing. This approach aims to improve patients' bowel preparation and optimize their cleansing experience. In this study, we hypothesized that not all patients need to consume the recommended doses of laxatives to achieve adequate bowel preparation. We used a developed AI platform to evaluate the adequacy of bowel cleansing in real time. Once the assessment was passed, the use of laxatives was stopped. This study sought to investigate the use of laxatives under AI direction and evaluate its impact on the quality of bowel preparation in patients.

Methods

Study Design

This was a prospective randomized controlled trial conducted from October 2024 to March 2025 at the First Affiliated Hospital of Zhengzhou University. This study included patients undergoing colonoscopy who were using sodium phosphate for bowel preparation. This study was registered in ClinicalTrials.gov with the registration number NCT06610630.

Ethical Considerations

This study was approved the First Affiliated Hospital of Zhengzhou University's ethics committee (2024-KY-1219-001) and aligned with the Declaration of Helsinki. All participants provided written informed consent prior to enrollment via a form that explicitly outlined study objectives, procedures, potential risks and benefits, and participants' unrestricted right to withdraw at any time without penalty. All participant data were fully anonymized before analysis to protect confidentiality. Participants did not receive any financial compensation or other incentives for taking part in this study.

Study Population

The research population consisted of individuals who had painless colonoscopy at the First Affiliated Hospital of Zhengzhou University from October 2024 to March 2025 using sodium phosphate for bowel preparation. Patients were recruited from the appointment registration office of the Digestive Endoscopy Center. During the study period, patients scheduled for painless afternoon colonoscopy and prescribed sodium phosphate solution for bowel preparation were systematically screened by trained research staff according to the predefined inclusion and exclusion criteria. Eligible patients must (1) use sodium phosphate for bowel cleansing, (2) perform bowel preparation in the morning and undergo colonoscopy in the afternoon, and (3) be able to use a smartphone autonomously or with help. The ability to use a smartphone was assessed during participant screening by trained staff. After standardized verbal instructions and a brief demonstration of the AI program, participants were evaluated on whether they, either independently or with assistance from an accompanying family member or caregiver, could complete the key steps required for using the program, including scanning the QR code, taking stool photographs, uploading images, and checking the feedback results. Patients who were unable to complete these steps and had no available assistant were considered unable to use the smartphone-based program. Individuals were excluded from the study if they refused to participate or met one or more of the following conditions: (1) contraindications to oral sodium phosphate solution, such as renal failure, congestive heart failure, bowel obstruction, congenital megacolon, or ascites; (2) allergy to components of sodium phosphate; and (3) history of colorectal surgery.

AI Bowel Preparation Assessment Program

We used an AI platform developed based on a convolutional neural network (CNN) model accessible via QR code scanning with a mobile phone and provided by Xiamen Innovision Medical Technology in China. The CNN model was constructed for the intelligent evaluation of images of patients' bowel preparation defecation. During the model training phase, a total of 4302 images were collected, and the size of the training set was determined through accuracy analysis. An incremental training strategy was adopted where 250 images were added to the training set each time, and the 10-fold cross-validation method was used to evaluate model performance under different sample sizes. The results showed that, with the increase in the number of training samples, model accuracy improved in stages: it increased rapidly in the initial stage, and when the sample size reached approximately 2500, accuracy increased to 97%. Thereafter, the slope of the curve approached 0, indicating that model performance gradually stabilized. Further verification demonstrated that, when the training set contained 3000 images or more, model accuracy could stably exceed 95%, and the total

XSL•FO
RenderX

number of images used for final modeling exceeded 4000 [15]. To further evaluate its diagnostic performance, the model was tested on an independent dataset of 750 images that was not used for model training. The CNN model achieved a sensitivity of 93.02%, specificity of 92.56%, and overall accuracy of 92.80% for binary classification of bowel preparation images, supporting its reliability for real-time assessment of bowel cleansing adequacy.

Image annotation was completed by a nurse with 20 years of clinical experience who uniformly labeled the images to classify them into 2 categories: “qualified” (passing) or “unqualified” (not passing). During model initialization, parameters were randomly configured; in the training process, annotated images were input into the network to iteratively optimize the parameters. After training, the CNN model could realize automatic binary classification of newly uploaded bowel preparation images. To enhance clinical practicality, this study encapsulated the CNN model into a web service based on the MobileNetV3 architecture so as to meet the simultaneous use needs of a large number of patients and achieve real-time provision evaluation results. Compared with other CNN architectures, MobileNetV3 effectively addresses the efficiency bottleneck of traditional classification models through a lightweight network design (smaller scale and fewer parameters).

During use, after scanning the QR code, the system displays examples of how to take a photo of the stool and instructions for uploading the image to the system. After the image is uploaded, patients automatically receive an evaluation outcome of “pass” or “not pass.” Regarding those who have a “not pass” outcome, the system provides prompts to guide them in taking remedial measures, such as increasing physical activity, gently massaging the abdomen, and increasing fluid intake to enhance the efficacy of bowel cleansing.

Bowel Preparation

Unblinded researchers used the SPSS (version 26.0; IBM Corp) software to generate the block randomization sequence with a block size set to 4. The allocation results were stored by independent statisticians (nonresearch staff). Independent third-party personnel prepared opaque sealed envelopes numbered by ID (containing group assignment slips) and delivered them to staff at the endoscopy registration office. The researchers had no prior knowledge of the group allocations. Staff at the appointment registration office randomly assigned participants to the AI-guided group or control group based on the allocation results, with a 1:1 allocation ratio between the 2 groups. All participants were given a bowel preparation instruction sheet when scheduling their colonoscopy, and trained physicians provided verbal explanations of the content, including recommendations for a low-fiber diet prior to the procedure, instructions for preparing the laxative solution, and methods for assessing the sufficiency of bowel preparation. All participants adhered to a low-fiber diet the day before bowel preparation and used sodium phosphate oral solution for cleansing. The standard sodium phosphate bowel preparation protocol is as follows: the first bowel preparation is scheduled for 8 AM, during which 45 mL of sodium phosphate solution are diluted in water to prepare approximately 800 mL of bowel-cleansing solution, which should be consumed within 30 minutes. The second dose is administered at 10 AM using the same method. After starting bowel preparation, patients in the AI group uploaded images of their stool using the app each time they had a bowel movement. Once the stool image was uploaded, the app displayed the assessment results as follows. On the basis of the uploaded image, a “pass” result was given if the assessment indicated that the preparation met the requirements for the procedure, and patients could proceed to the hospital for their appointment. An assessment of “not pass” was given when, based on the uploaded image, the assessment indicated that the preparation did not meet the requirements. Recommendations included increasing physical activity; gently massaging the abdomen in a clockwise direction; and, if necessary, increasing the use of laxatives. Patients were instructed to take another photo in their next bowel movement for reassessment until they received a “pass” result. Once a “pass” result was determined, patients were advised to stop taking the laxatives. In the control group, individuals assessed their bowel preparation based on the provided instructions without the assistance of the AI app.

Data Collection

The primary outcomes were the actual volume of bowel-cleansing solution consumed and the proportion of patients achieving adequate bowel preparation. The volume of bowel-cleansing solution was recorded as the total amount of prepared sodium phosphate solution ingested before colonoscopy. Reduced laxative use was defined as consumption of at least 25% less than the standard 1600-mL bowel-cleansing solution. Endoscopists and nurses were completely unaware of the group assignments. Outcome assessors—specifically, 2 independent endoscopists—had no access to group information, and patients were only exposed to the interventions designated for their respective groups. Two endoscopists evaluated the patients’ bowel preparation status using the Boston Bowel Preparation Scale (BBPS) [12,16]. The ultimate decision was made by a senior endoscopist if they could not come to an agreement. The definition of sufficient bowel preparation was a total BBPS score of 6 or higher, with each segment score being 2 or higher [17]. Secondary outcomes included total BBPS score, segmental BBPS scores, polyp detection rate (PDR), ADR, advanced ADR (aADR), medication adherence, patient-reported acceptance of the AI program, and adverse events. PDR was defined as the proportion of patients with at least one detected polyp, and ADR was defined as the proportion of patients with at least one detected adenoma. Advanced adenomas were defined as adenomas larger than 10 mm, those with high-grade dysplasia, or those with villous features [18]. Medication adherence in the AI-guided group was defined as following the AI program instructions, including discontinuing medication after receiving a “pass” result or implementing remedial measures after receiving a “not pass” result. Medication adherence in the control group was defined as completing the full prescribed dose of bowel-cleansing solution, namely, 1600 mL. Patient-reported acceptance of the AI program was assessed in the AI-guided group based on whether patients could successfully use the program without additional instructions

XSL•FO
RenderX

and whether they were willing to use the AI program again for future bowel preparation. Adverse events were recorded through patient self-report and structured inquiry before colonoscopy. These included symptoms occurring during bowel preparation, such as nausea, vomiting, abdominal pain, bloating, dizziness, and palpitations. The incidence of adverse events was defined as the proportion of patients reporting at least one of these symptoms. Regarding clinical characteristics, daily defecation frequency was categorized into 3 levels: low frequency (fewer than 3 times per week), medium frequency (2 times per day to 3 times per week), and high frequency (more than 2 times per day).

Statistical Analysis

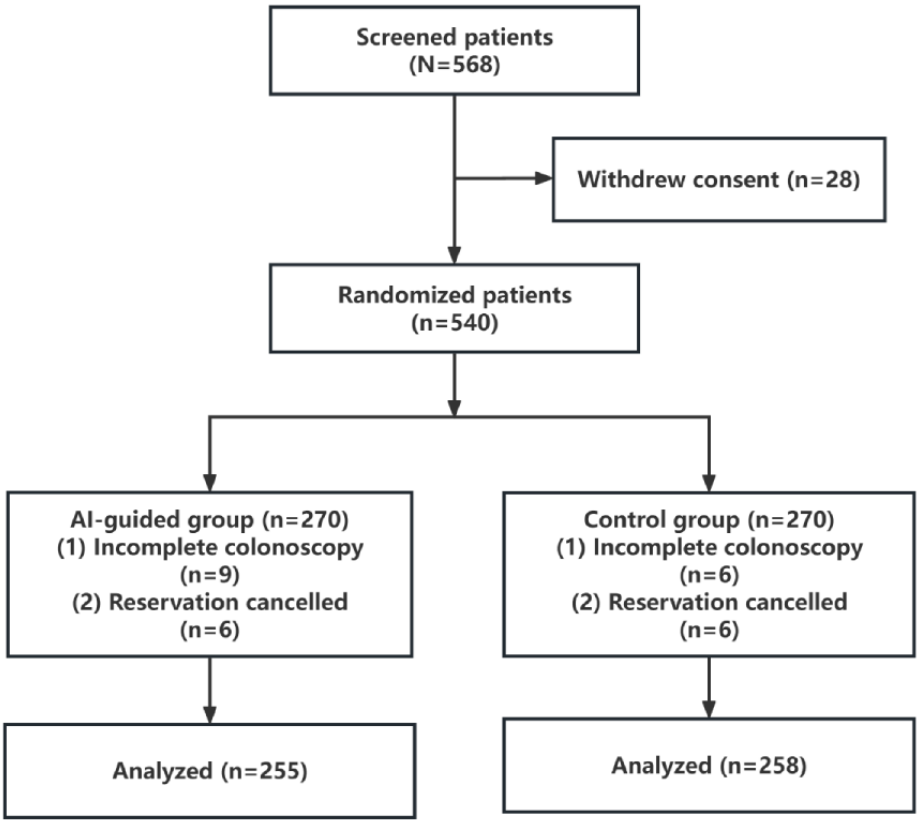
The sample size was calculated based on data from previous studies and pilot test results. We set the adequate bowel preparation rate for the control group at 85% assuming an adequate bowel preparation rate of 90% for the AI group. The noninferiority margin was set at −5 percentage points. Using PASS (version 11; NCSS, LLC) for calculations, we established that a sample size of 230 per group would achieve 90% statistical power and a significance level of .025 to detect noninferiority between the groups. Considering a 10% attrition rate, at least 512 patients were required. Data analysis was conducted using SPSS (version 26.0) and R (version 4.5.0; R Foundation for Statistical Computing). Categorical data were reported as counts and percentages, and the chi-square test, continuity-corrected chi-square test, or Fisher exact test were used to compare the 2 groups. For continuous variables, normal distribution was represented as means and SDs, whereas skewed distribution was reported as medians with IQRs. Group comparisons were made using *t* tests, Welch *t* tests, or nonparametric tests (Mann-Whitney *U* tests). A 2-tailed *P* value of less than .05 was considered statistically significant for conventional between-group comparisons. We used a one-sided test to evaluate the noninferiority of AI-guided bowel preparation vs the control group. Noninferiority was concluded if the lower bound of the 95% CI for the difference in adequate bowel preparation rate did not cross the –5–percentage point margin. To find predictors of reduced laxative use in the AI-guided group, logistic regression analysis was used.

Results

Patient Characteristics

A total of 540 patients were randomly assigned to both groups, of whom 27 (5%) were excluded due to incomplete cecal intubation for reasons such as rectal cancer. Ultimately, the study comprised a total of 513 patients (n=255, 49.7% in the AI-guided group and n=258, 50.3% in the control group; Figure 1). There were no significant variations among the 2 groups' baseline features, including age, sex, BMI, indications for colonoscopy (screening, surveillance, and diagnosis), daily defecation frequency, and relevant medical history (history of abdominal surgery, coronary heart disease, diabetes, and hypertension; Table 1).

Figure 1. Flow diagram for patient inclusion.



XSL•FO
RenderX

Table 1. Patient baseline characteristics.

	AI-guided group (n=255)	Control group (n=258)	*P* value
Age (y), mean (SD)	47.77 (13.20)	49.53 (11.81)	.11
Male, n (%)	125 (49)	142 (55)	.17
BMI (kg/m^2), mean (SD)	23.84 (2.75)	24.33 (3.08)	.06
Colonoscopy indication, n (%)			.84
Screening	81 (31.8)	77 (29.8)	
Surveillance	66 (25.9)	72 (27.9)	
Diagnostic	108 (42.4)	109 (42.2)	
Defecation frequency, n (%)			.39
Low	52 (20.4)	41 (15.9)	
Medium	181 (71)	191 (74)	
High	22 (8.6)	26 (10.1)	
Previous surgery (abdominal or pelvic), n (%)	50 (19.6)	61 (23.6)	.27
Medical history, n (%)			
Diabetes	36 (14.1)	40 (15.5)	.66
Hypertension	51 (20)	59 (22.9)	.43
Coronary disease	17 (6.7)	19 (7.4)	.76

Volume of Laxatives Consumed

In the control group, all participants (258/258, 100%) consumed the recommended dosage of sodium phosphate solution for bowel preparation. In the AI-guided group, 48.6% (124/255) of the patients reduced their use of laxatives. The average volume of laxatives consumed in the AI-guided group was 1245.5 (SD 377.2) mL, which was significantly lower than the 1586.2 (SD 50.4) mL consumed in the control group (P<.001). The multivariate analysis revealed that reduced use of laxatives was independently predicted by younger age, lower BMI, and higher defecation frequency (Table 2).

XSL•FO
RenderX

Table 2. Predictive factors for reducing the use of laxatives.

	Univariate analysis		Multivariate analysis	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value
Age	0.957 (0.937-0.976)	*<.001* [b]	0.964 (0.942-0.986)	*.001*
Sex				
Male	1	—[c]	—	—
Female	1.191 (0.729-1.948)	.49	—	—
BMI	0.719 (0.642-0.805)	*<.001*	0.739 (0.656-0.832)	*<.001*
Previous surgery (abdominal or pelvic)				
Yes	1	—	—	—
No	1.713 (0.909-3.227)	.10	—	—
Defecation frequency				
Low	1	—	1	—
Medium	3.241 (1.622-6.476)	*.001*	2.595 (1.209-5.572)	*.01*
High	10.200 (3.140-33.136)	*<.001*	10.200 (3.365-43.729)	*<.001*
Diabetes				
Yes	1	—	—	—
No	1.586 (0.772-3.260)	.21	—	—
Hypertension				
Yes	1	—	—	—
No	1.194 (0.645-2.210)	.57	—	—
Coronary disease				
Yes	1	—	—	—
No	0.493 (0.177-1.377)	.18	—	—

[a]OR: odds ratio.

[b]Italics indicate statistical significance.

[c]Not applicable.

Bowel Preparation Quality and Colonoscopy Results

No significant difference was observed in the proportion of patients achieving sufficient bowel preparation between the AI-guided group and the control group (238/255, 93.3% vs 229/258, 88.8%, respectively; *P*=.07). The noninferiority analysis demonstrated that the AI-guided group was noninferior to the control group as the CI did not exceed the noninferiority margin of –5 percentage points and the one-sided *P* value was statistically significant (*P*<.001; 95% CI −0.0038 to 0.0971). Additionally, the average BBPS scores for each colon region, ADR, PDR, and aADR did not differ significantly between the 2 groups (Table 3).

XSL•FO
RenderX

Table 3. Quality of bowel preparation and colonoscopy outcomes in both groups.

	AI-guided group (n=255)	Control group (n=258)	*P* value
Bowel preparation quality, n (%)			.07
Adequate	238 (93.3)	229 (88.8)	
Inadequate	17 (6.7)	29 (11.2)	
BBPS[a] score, mean (SD)			
Total	7.44 (1.20)	7.26 (1.22)	.09
Right colon	2.23 (0.56)	2.14 (0.57)	.07
Transverse colon	2.68 (0.47)	2.68 (0.47)	>.99
Left colon	2.53 (0.50)	2.44 (0.55)	.06
Adenoma detection rate, n (%)	89 (34.9)	88 (34.1)	.85
Polyp detection rate, n (%)	112 (43.9)	116 (45.0)	.81
Advanced adenoma detection rate, n (%)	16 (6.3)	8 (3.1)	.09

[a]BBPS: Boston Bowel Preparation Scale (each colon segment of the BBPS ranges from 0 to 3, and the total BBPS score ranges from 0 to 9, with higher scores indicating better bowel preparation quality).

Patient-Reported Acceptance, Medication Adherence, and Incidence of Adverse Events

A total of 94.9% (242/255) of patients were able to successfully use the AI bowel preparation program without additional instructions. In total, 96.5% (246/255) of patients expressed a willingness to use the AI bowel preparation program for their next bowel cleansing. In addition, the medication adherence rate in the AI-guided group was significantly higher than that in the control group (247/255, 96.9% vs 236/258, 91.5%, respectively; *P*=.009), whereas the incidence of adverse reactions was notably lower (5/255, 2% vs 15/258, 5.8%, respectively; *P*=.02). Specifically, the incidence of reported adverse reactions in the AI-guided group vs the control group was as follows: 1.6% (4/255) vs 4.7% (12/258) for nausea (*P*=.045), 1.2% (3/255) vs 3.9% (10/258) for vomiting (*P*=.052), and 1.2% (3/255) vs 4.7% (12/258) for bloating (*P*=.02).

Discussion

In this research, we used a developed and validated AI bowel preparation program to assess stool status in real time and guide patients in personalized bowel cleansing. In the AI-guided group, 48.6% (124/255) of the participants reduced the use of laxatives. The proportions of adequate bowel cleansing, BBPS scores, and ADR did not vary significantly between the AI-guided group and the control group. This suggested that AI-guided bowel cleansing can provide patients with a more personalized bowel preparation plan, reducing unnecessary use of laxatives.

In previous studies, to help patients prepare their bowels better, a number of therapies have been investigated. For example, visual aids [19] and educational videos [20] have been used to guide bowel cleansing in a more understandable manner. Additionally, patient education has been reinforced through phone calls [21,22] or smartphone apps [23-25] to enhance bowel preparation quality. However, these studies have primarily focused on enhancing patient education and do not provide real-time assistance for patients to assess the sufficiency of their bowel preparation. In recent years, research has suggested that AI-driven smartphone apps can evaluate bowel preparation adequacy in real time, thereby improving patients' bowel preparation [26,27]. However, these studies have only focused on how AI programs can provide recommendations for additional laxative use for patients with inadequate bowel preparation. Different individuals may respond differently to laxatives. For instance, patients with constipation or older adults may require more laxatives, whereas those with normal bowel habits or irritable bowel syndrome may achieve adequate preparation with less. Currently, limited research has explored the feasibility of an AI bowel preparation program providing tailored recommendations for increasing or decreasing laxative use based on individual responses.

We proposed a new approach for bowel preparation based on an AI bowel preparation program. After a successful AI assessment, patients could stop taking laxatives, whereas those who did not pass would take additional remedial measures to achieve adequate preparation. The final results showed that participants in the AI-guided group could perform bowel preparation in a way comparable to conventional practices after consuming an average of approximately 1.25 L of bowel-cleansing solution. This significantly reduced the volume of cleansing solution used compared to the 3-L or 4-L polyethylene glycol solution or the 1.6-L sodium phosphate solution recommended by guidelines and other studies [28-30]. This suggests that the AI bowel preparation program has significant potential for guiding individualized laxative use. In the logistic regression analysis, younger age, lower BMI, and higher defecation frequency were identified as independent predictive factors for reduced laxative use. Future research could establish a predictive model for bowel-cleansing agent dosage by increasing the sample size and considering other relevant factors, providing patients with reasonable laxative use plans in situations in which using AI programs is not feasible.

In terms of bowel preparation quality across the 2 groups, the AI-guided group exhibited superior rates of sufficient bowel

XSL•FO
RenderX

preparation and BBPS scores compared to the control group. However, these differences lacked statistical significance. Despite thorough explanations and education, some patients may not have used the AI bowel preparation program correctly. Data indicated that 1.2% (3/255) of the patients took a photo of an empty toilet bowl, which could have led to inaccuracies in the AI assessment. Additionally, some patients who did not pass the assessment after taking the laxatives failed to follow the recommended remedial measures for bowel preparation. Ultimately, only 4.7% (12/255) of the patients undertook additional remedial actions and postponed their colonoscopy, whereas other patients who did not pass the assessment proceeded with their scheduled colonoscopy, which may have led to an underestimation of the effectiveness of the AI-guided program. The colonoscopy results, such as ADR, PDR, and aADR, did not exhibit substantial variations between the 2 groups. Previous studies have suggested that enhanced educational methods such as telephone or WeChat education could achieve higher ADR [21,23]. A possible reason for this is that bowel preparation quality was comparable in both groups, and the ADR among the endoscopists at our center was relatively high (89/255, 34.9% in the AI-guided group and 88/258, 34.1% in the control group), which may have led to an inability to observe differences in ADR and other colonoscopy outcomes.

In the AI-guided group, patients had a generally positive attitude toward the use of the AI program. Most were able to successfully use the AI program for bowel preparation after a simple explanation and introduction. Nevertheless, some older individuals were excluded from the study because they were unable to use smartphones or unable to use them with assistance from others. To address this issue, first, the future program interface should be designed to be more concise, with visual operation instructions to ensure that it is easy to understand and use. Second, for older patients who still cannot master smartphone operations, their immediate family members (or accompanying personnel) could be asked to assist with image uploads and feedback checks. Additionally, some patients experienced issues with the display of evaluation results due to network problems, but they successfully used the program after troubleshooting their network and devices. To address network obstacles, an “offline evaluation–subsequent synchronization” dual-mode approach should be developed in the future. The app should support locally storing encrypted images and perform offline analysis using lightweight CNNs, with automatic data synchronization enabled once network connectivity is restored. Regarding adverse reactions, AI-guided laxative dose reduction lowered the incidence of nausea, vomiting, and abdominal distension, with no severe adverse events reported. This benefit arose from the AI model’s ability to precisely identify optimal bowel cleanliness thresholds through real-time image analysis, thereby avoiding overmedication in nonconstipated patients under fixed-dose regimens. Mechanistically, this reduced intestinal distension and vagal nerve stimulation. Furthermore, patients in the AI-guided group had a higher medication adherence, indicating that the AI program has high usability in clinical practice and significant potential for broader application.

This research has limitations. First, this study exclusively included patients using sodium phosphate for bowel preparation, a decision driven by our research design. However, this approach may restrict the broader applicability of our findings to other laxatives (eg, polyethylene glycol). Future studies should expand the sample size and incorporate multiple laxative regimens to enhance the clinical generalizability of AI-driven bowel cleanliness assessment tools. Second, the operation of the AI bowel preparation program relies on the support of smartphones and wireless networks, which means that some individuals who cannot use smartphones, particularly certain older patients, were excluded from the study, potentially leading to selection bias.

In conclusion, this study presents the use of an AI bowel preparation program to guide patients in personalized bowel preparation, which can reduce unnecessary use of laxatives while achieving results comparable to those obtained with standard practices. Promoting this program in clinical practice could help improve the efficiency and comfort of bowel preparation for patients.

Acknowledgments

The authors would like to thank the assistance provided by Xiamen Innovision Medical Technology in the development of the AI app. Generative AI was not used in the writing of the manuscript.

Funding

This study was supported by the Key Scientific Research Project of Colleges and Universities in Henan Province (grant 23A320020).

Data Availability

The data that support the findings of this study are available from the corresponding author on reasonable request.

Authors' Contributions

All authors contributed to study conception and design. FH contributed to writing—original draft preparation. FH, SG, and JY contributed to writing—review and editing. SG and HZ contributed to methodology. FH and NW contributed to formal analysis. ZS and JX contributed to investigation. YH and YD contributed to data curation. JY contributed to funding acquisition. HZ and ZS contributed to resources. SG and HZ contributed to supervision, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2873 KB - mhealth_v14i1e77701_app1.pdf]

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
2. He X, Hang D, Wu K, Nayor J, Drew DA, Giovannucci EL, et al. Long-term risk of colorectal cancer after removal of conventional adenomas and serrated polyps. Gastroenterology 2020 Mar;158(4):852-61.e4 [FREE Full text] [doi: 10.1053/j.gastro.2019.06.039] [Medline: 31302144]
3. Dekker E, Tanis PJ, Vleugels JL, Kasi PM, Wallace MB. Colorectal cancer. Lancet 2019 Oct 19;394(10207):1467-1480. [doi: 10.1016/S0140-6736(19)32319-0] [Medline: 31631858]
4. Zauber AG, Winawer SJ, O'Brien MJ, Lansdorp-Vogelaar I, van Ballegooijen M, Hankey BF, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N Engl J Med 2012 Feb 23;366(8):687-696 [FREE Full text] [doi: 10.1056/NEJMoa1100370] [Medline: 22356322]
5. Saini SD. Colonoscopy with polypectomy reduces colorectal cancer mortality in patients with adenomas. Gastroenterology 2012 Oct;143(4):1112-1113. [doi: 10.1053/j.gastro.2012.08.018] [Medline: 22917862]
6. Zhao S, Wang S, Pan P, Xia T, Chang X, Yang X, et al. Magnitude, risk factors, and factors associated with adenoma miss rate of tandem colonoscopy: a systematic review and meta-analysis. Gastroenterology 2019 May;156(6):1661-74.e11. [doi: 10.1053/j.gastro.2019.01.260] [Medline: 30738046]
7. Martínez ME, Baron JA, Lieberman DA, Schatzkin A, Lanza E, Winawer SJ, et al. A pooled analysis of advanced colorectal neoplasia diagnoses after colonoscopic polypectomy. Gastroenterology 2009 Mar;136(3):832-841 [FREE Full text] [doi: 10.1053/j.gastro.2008.12.007] [Medline: 19171141]
8. Samadder NJ, Curtin K, Tuohy TM, Pappas L, Boucher K, Provenzale D, et al. Characteristics of missed or interval colorectal cancer and patient survival: a population-based study. Gastroenterology 2014 Apr;146(4):950-960. [doi: 10.1053/j.gastro.2014.01.013] [Medline: 24417818]
9. Sawicki T, Ruszkowska M, Danielewicz A, Niedźwiedzka E, Arłukowicz T, Przybyłowicz KE. A review of colorectal cancer in terms of epidemiology, risk factors, development, symptoms and diagnosis. Cancers (Basel) 2021 Apr 22;13(9):2025 [FREE Full text] [doi: 10.3390/cancers13092025] [Medline: 33922197]
10. Saito Y, Oka S, Kawamura T, Shimoda R, Sekiguchi M, Tamai N, et al. Colonoscopy screening and surveillance guidelines. Dig Endosc 2021 May;33(4):486-519. [doi: 10.1111/den.13972] [Medline: 33713493]
11. Rex DK, Schoenfeld PS, Cohen J, Pike IM, Adler DG, Fennerty MB, et al. Quality indicators for colonoscopy. Am J Gastroenterol 2015 Jan;110(1):72-90. [doi: 10.1038/ajg.2014.385] [Medline: 25448873]
12. Hassan C, East J, Radaelli F, Spada C, Benamouzig R, Bisschops R, et al. Bowel preparation for colonoscopy: European Society of Gastrointestinal Endoscopy (ESGE) guideline - update 2019. Endoscopy 2019 Aug;51(8):775-794 [FREE Full text] [doi: 10.1055/a-0959-0505] [Medline: 31295746]
13. Liu Z, Zhang MM, Li YY, Li LX, Li YQ. Enhanced education for bowel preparation before colonoscopy: a state-of-the-art review. J Dig Dis 2017 Feb;18(2):84-91. [doi: 10.1111/1751-2980.12446] [Medline: 28067463]
14. Lebwohl B, Kastrinos F, Glick M, Rosenbaum AJ, Wang T, Neugut AI. The impact of suboptimal bowel preparation on adenoma miss rates and the factors associated with early repeat colonoscopy. Gastrointest Endosc 2011 Jun;73(6):1207-1214 [FREE Full text] [doi: 10.1016/j.gie.2011.01.051] [Medline: 21481857]
15. Lu YB, Lu SC, Huang YN, Cai ST, Le PH, Hsu FY, et al. A novel convolutional neural network model as an alternative approach to bowel preparation evaluation before colonoscopy in the COVID-19 era: a multicenter, single-blinded, randomized study. Am J Gastroenterol 2022 Sep 01;117(9):1437-1443. [doi: 10.14309/ajg.0000000000001900] [Medline: 35973166]
16. Lai EJ, Calderwood AH, Doros G, Fix OK, Jacobson BC. The Boston Bowel Preparation Scale: a valid and reliable instrument for colonoscopy-oriented research. Gastrointest Endosc 2009 Mar;69(3 Pt 2):620-625 [FREE Full text] [doi: 10.1016/j.gie.2008.05.057] [Medline: 19136102]
17. Calderwood AH, Jacobson BC. Comprehensive validation of the Boston Bowel Preparation Scale. Gastrointest Endosc 2010 Oct;72(4):686-692 [FREE Full text] [doi: 10.1016/j.gie.2010.06.068] [Medline: 20883845]
18. Gupta S, Lieberman D, Anderson JC, Burke CA, Dominitz JA, Kaltenbach T, et al. Recommendations for follow-up after colonoscopy and polypectomy: a consensus update by the US Multi-Society Task Force on colorectal cancer. Am J Gastroenterol 2020 Mar;115(3):415-434 [FREE Full text] [doi: 10.14309/ajg.0000000000000544] [Medline: 32039982]

XSL•FO
RenderX

19. Calderwood AH, Lai EJ, Fix OK, Jacobson BC. An endoscopist-blinded, randomized, controlled trial of a simple visual aid to improve bowel preparation for screening colonoscopy. Gastrointest Endosc 2011 Feb;73(2):307-314. [doi: 10.1016/j.gie.2010.10.013] [Medline: 21168840]
20. Pillai A, Menon R, Oustecky D, Ahmad A. Educational colonoscopy video enhances bowel preparation quality and comprehension in an inner city population. J Clin Gastroenterol 2018 Jul;52(6):515-518. [doi: 10.1097/MCG.0000000000000893] [Medline: 28742732]
21. Liu X, Luo H, Zhang L, Leung FW, Liu Z, Wang X, et al. Telephone-based re-education on the day before colonoscopy improves the quality of bowel preparation and the polyp detection rate: a prospective, colonoscopist-blinded, randomised, controlled study. Gut 2014 Jan;63(1):125-130. [doi: 10.1136/gutjnl-2012-304292] [Medline: 23503044]
22. Lee YJ, Kim ES, Choi JH, Lee KI, Park KS, Cho KB, et al. Impact of reinforced education by telephone and short message service on the quality of bowel preparation: a randomized controlled study. Endoscopy 2015 Nov;47(11):1018-1027. [doi: 10.1055/s-0034-1392406] [Medline: 26182387]
23. Kang X, Zhao L, Leung F, Luo H, Wang L, Wu J, et al. Delivery of instructions via mobile social media app increases quality of bowel preparation. Clin Gastroenterol Hepatol 2016 Mar;14(3):429-35.e3. [doi: 10.1016/j.cgh.2015.09.038] [Medline: 26492848]
24. Walter B, Frank R, Ludwig L, Dikopoulos N, Mayr M, Neu B, et al. Smartphone application to reinforce education increases high-quality preparation for colorectal cancer screening colonoscopies in a randomized trial. Clin Gastroenterol Hepatol 2021 Feb;19(2):331-8.e5. [doi: 10.1016/j.cgh.2020.03.051] [Medline: 32240835]
25. Lorenzo-Zúñiga V, Moreno de Vega V, Marín I, Barberá M, Boix J. Improving the quality of colonoscopy bowel preparation using a smart phone application: a randomized trial. Dig Endosc 2015 Jul;27(5):590-595. [doi: 10.1111/den.12467] [Medline: 25708251]
26. Inaba A, Shinmura K, Matsuzaki H, Takeshita N, Wakabayashi M, Sunakawa H, et al. Smartphone application for artificial intelligence-based evaluation of stool state during bowel preparation before colonoscopy. Dig Endosc 2024 Dec;36(12):1338-1346. [doi: 10.1111/den.14827] [Medline: 39031797]
27. Zhu Y, Zhang DF, Wu HL, Fu PY, Feng L, Zhuang K, et al. Improving bowel preparation for colonoscopy with a smartphone application driven by artificial intelligence. NPJ Digit Med 2023 Mar 14;6(1):41 [FREE Full text] [doi: 10.1038/s41746-023-00786-y] [Medline: 36918730]
28. ASGE Standards of Practice Committee, Saltzman JR, Cash BD, Pasha SF, Early DS, Muthusamy VR, et al. Bowel preparation before colonoscopy. Gastrointest Endosc 2015 Apr;81(4):781-794. [doi: 10.1016/j.gie.2014.09.048] [Medline: 25595062]
29. Johnson DA, Barkun AN, Cohen LB, Dominitz JA, Kaltenbach T, Martel M, et al. Optimizing adequacy of bowel cleansing for colonoscopy: recommendations from the US Multi-Society Task Force on Colorectal Cancer. Am J Gastroenterol 2014 Oct;109(10):1528-1545. [doi: 10.1038/ajg.2014.272] [Medline: 25223578]
30. Unal S, Doğan UB, Oztürk Z, Cindoruk M. A randomized prospective trial comparing 45 and 90-ml oral sodium phosphate with X-Prep in the preparation of patients for colonoscopy. Acta Gastroenterol Belg 1998;61(3):281-284. [Medline: 9795453]

Abbreviations

aADR: advanced adenoma detection rate
ADR: adenoma detection rate
BBPS: Boston Bowel Preparation Scale
CNN: convolutional neural network
CRC: colorectal cancer
PDR: polyp detection rate

Edited by A Stone; submitted 18.May.2025; peer-reviewed by P Pan, Y Kishida; comments to author 16.Sep.2025; accepted 26.Jun.2026; published 14.Jul.2026.

Please cite as:
Han F, Gao S, Zhan H, Wang N, Song Z, Xie J, Hou Y, Deng Y, Yao J
Use of an AI Image Recognition–Based Smartphone App in Personalized Bowel Preparation Before Colonoscopy: Prospective Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e77701
URL: https://mhealth.jmir.org/2026/1/e77701
doi:10.2196/77701
PMID:

XSL•FO
RenderX

©Fei Han, Shilin Gao, Hanqiang Zhan, Nan Wang, Zhaoxiang Song, Jinlin Xie, Yuhan Hou, Yimeng Deng, Jianning Yao. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Dietary and Exercise Relapse Management in Mobile Health Apps for Weight and Glycemic Control: Descriptive Qualitative Study

Ahlam Saleh Alhajri, PhD; Afnan I Alandanoosi, PhD; Hala Hazam Al-Otaibi, PhD

Department of Food and Nutrition Sciences, College of Agricultural and Food Sciences, King Faisal University, PO Box 400 Hofuf, Al Ahsa, Saudi Arabia

Corresponding Author:
Ahlam Saleh Alhajri, PhD
Department of Food and Nutrition Sciences, College of Agricultural and Food Sciences, King Faisal University, PO Box 400 Hofuf, Al Ahsa, Saudi Arabia

Abstract

Background: Mobile health (mHealth) apps for weight and glycemic control support self-monitoring and behavior change, but engagement often declines over time. Although relapse in diet and exercise is common, little is known about how users experience relapse within digital contexts or how apps should respond.

Objective: This study aimed to explore how users of mHealth apps for weight and glycemic control experience, manage, and wish to be supported during dietary and exercise relapses.

Methods: A descriptive qualitative design was adopted. Semistructured online interviews were conducted with 26 adults who had used weight or glycemic control apps for at least 3 months and reported at least 1 relapse. Data were analyzed using inductive thematic analysis.

Results: Seven themes emerged: stress and reduced capacity as conditions surrounding relapse; cultural, temporal, and environmental misalignment; psychological disengagement and avoidance; social and relational influences; relapse as emotional disengagement rather than simply a behavioral lapse; informal relapse management and self-directed recalibration; and desire for compassionate, context-aware support. Participants described relapse in relation to stress, grief, caregiving and occupational demands, religious observances, festivals, shift work, travel, and constrained food environments. Relapse was frequently experienced as emotional disengagement, characterized by guilt, shame, identity threat, and loss of meaning, and was often accompanied by avoidance of the app or monitoring devices. Participants managed relapse through self-directed strategies such as temporal resets, goal adjustment, partial engagement, and social support, but reported few structured relapse-recovery tools within apps. They expressed a strong preference for flexible goals, context-aware modes, mood-sensitive feedback, neutral language, and gentle reentry pathways.

Conclusions: Participants' accounts suggest that relapse in mHealth contexts involves emotional responses, contextual constraints, and changing self-management capacity. Flexible goals, neutral feedback, context-sensitive support, and graded reentry are promising design targets for relapse-responsive apps, but their effects on engagement and clinical outcomes require prospective evaluation.

(JMIR Mhealth Uhealth 2026;14:e94792) doi:10.2196/94792

KEYWORDS

mobile health; dietary; weight management; glycemic control; digital health design

Introduction

The major chronic diseases, such as type 2 diabetes, cardiovascular diseases (hypertension, heart disease, and stroke), obesity, chronic obstructive pulmonary disease, and osteoarthritis, with high global prevalence, continue to be major public health challenges, and are associated with mortality risk [1-3]. Lifestyle modification, dietary regulation, and regular physical activity are a few self-management and preventive techniques that can result in improved health parameters, such as weight reduction and glycemic control [4-7]. However, adopting these behaviors in the long term could be challenging and may result in relapses, that is, returning to previous maladaptive behaviors after a period of improvement [8-10]. In weight management and diabetes self-care contexts, relapses may include neglecting or avoiding self-monitoring, reduced or absent physical activity, and adopting previous or unhealthy eating patterns [11,12].

Mobile health (mHealth), with the integration of effective and efficient technologies such as AI, has been observed to be a scalable tool for facilitating dietary tracking, physical activity monitoring, glucose level logging and monitoring, customized feedback, and tailored behavioral coaching [13,14]. However, studies [15-18] have indicated that mHealth interventions can

XSL•FO
RenderX

produce modest but significant improvements in weight and glycemic outcomes. Some of the key features of such apps include self-monitoring, goal setting, prompts, and personalized feedback, which can be related to behavior change theories such as Social Cognitive Theory and Self-Determination Theory [19].

Despite their promising benefits, sustained engagement with mHealth apps has been found to decline over time [20,21]. Attrition rates in this context are high, and discontinuation of self-monitoring is strongly associated with poorer weight and glycemic outcomes [22]. Most digital interventions, such as mHealth apps, are designed to promote adherence and behavior initiation but do not extensively focus on user experiences during relapse periods or how the app should respond effectively when lapses occur [23-25]. According to the Relapse Prevention Model, relapse is a process often influenced by high-risk situations, coping responses, and self-efficacy [26]. Furthermore, psychological and emotional factors such as guilt, shame, and reduced motivation can further aggravate disengagement following the relapse period [27].

Therefore, understanding relapses not only as a failure but also as a predictable aspect of behavior change is crucial for designing responsive mHealth apps. However, relapse-responsive features such as dynamic goal adjustment, compassionate messaging, and structured "restart" pathways are poorly addressed in intervention studies [28,29], and empirical evidence on how users experience and wish to be supported during dietary and exercise relapses in mHealth contexts is limited [30,31].

Sustained regulatory behaviors such as dietary planning, physical activity, and self-monitoring are essential for achieving long-term weight management and glycemic regulation [32,33]. mHealth apps for weight management and diabetes self-management have reflected moderate improvements in clinical and behavioral outcomes. However, their impact is usually weakened over time. Studies [15,17,34-36] focusing on mHealth apps that combine diet and physical activity monitoring with tailored feedback, goal setting, and, in some cases, professional coaching have reflected average weight losses of 1 to 2 kg at 3 to 12 months and modest glycated hemoglobin reductions. However, engagement with mHealth apps usually decreases or becomes null, reflecting that intervention content is not well aligned with the cyclical nature of behavior change and relapse.

Studies focusing on behavior change frameworks in relation to weight loss and maintenance digital interventions [37-42] identified that initial success is contributed by factors such as self-monitoring, feedback, goal setting, action planning, and shaping knowledge. However, these frameworks specify only generic messages to resume tracking but do not specify how apps should respond when users miss logs, exceed dietary targets, or stop exercising. Structured relapse-management plans, strategies to bolster autonomous motivation, and emotional self-regulation have been identified as key aspects for achieving sustained maintenance but are not effectively integrated into mHealth apps [37-39].

Furthermore, studies [43,44] have observed that digital apps can enhance awareness, knowledge, and perceived control among users in relation to weight and diabetes management but highlight the need to address complex emotional responses to lapses. It has also been observed that approaches such as reminders and feedback not only support accountability but also trigger guilt and shame among users when goals are not met, which can lead to disengagement from the app and behavior change efforts [27,45]. In relation to diabetes self-management, it was observed that patients value supportive, nonjudgmental communication from health care providers in the event of high glycemic levels [46]. Similar supportive approaches need to be translated into digital apps, but this requires careful design to avoid exacerbating feelings of failure. Some of the challenges identified by users include tracking burden, usability issues, competing life demands, and the feeling that apps do not adapt to their changing capabilities, illness episodes, or psychological factors such as mood that precipitate relapse [47-49]. Studies focusing on user-centered designs highlight the importance of autonomy support, flexible goal recalibration, and normalization of setbacks [50,51].

Research on relapse in diet and physical activity digital interventions highlights that lapses are a common phenomenon, triggered by emotional distress, social situations, environmental cues, and disruptions to routine. A few skills, such as self-efficacy, coping planning, and reframing lapses as learning opportunities, were found to be associated with effective lapse recovery and sustained maintenance of physical activity and weight loss [52]. However, most digital apps assume linear progress, providing limited scaffolding for users to navigate cyclical patterns of engagement and disengagement. Furthermore, there is a lack of research examining how individuals using weight and glycemic control apps interpret and respond to relapses in diet and exercise, how app features shape these experiences, and what forms of digital support they would find acceptable and helpful during relapse episodes.

Considering these gaps, this study aims to explore how users of mHealth apps for weight and glycemic control experience, manage, and wish to be supported during dietary and exercise relapses. Specifically, it examines the relationships among changing personal circumstances, emotional responses to app-generated feedback, and disengagement from digital self-management. By translating these experiences into proposed relapse-responsive app features and behavior change strategies, the study seeks to inform interventions that support recovery and reengagement rather than assuming uninterrupted adherence.

Methods

Study Settings and Participants

A descriptive qualitative design was adopted to gain a deeper understanding of how users of mHealth apps for weight and glycemic control experience, manage, and wish to be supported during dietary and exercise relapses. A qualitative approach was selected because the study focused on a multidimensional construct—lived experiences, perceptions, emotional responses, coping strategies, and support needs during relapse episodes—that requires in-depth interpretation of users'

XSL•FO
RenderX

perspectives. Semistructured online interviews were used to gather participants' accounts of relapse in the context of mHealth app use. Participants included adult users of mHealth apps for weight management and/or glycemic control.

Recruitment and Sampling

A purposive sampling strategy was adopted to ensure the inclusion of diverse participants in different contexts. Inclusion criteria were as follows: (1) use of a mobile app for weight management and/or glycemic control for at least 3 months and (2) self-reported experience of at least 1 noticeable relapse in diet and/or exercise during that period (eg, cessation of logging, reduced physical activity, or reversion to previous eating patterns). Variation was sought in age, gender, app type, and health goals (weight management and/or glycemic control) to capture diverse perspectives. Participants were recruited through online advertisements posted on social media platforms, digital health forums, diabetes and weight management community groups, and university mailing lists.

An initial target of 17 interviews was informed by methodological guidance on sample sizes for qualitative interview studies [53]. Recruitment continued beyond this target to ensure variation in age, gender, occupation, health goal, app type, and relapse context. During analysis, the research team reviewed new codes and themes after each batch of interviews. By approximately interview 24, no substantively new domains of experience were identified, although later interviews continued to provide contextual variation and elaboration. All eligible individuals invited to interview agreed to participate, and no participants withdrew or discontinued their interviews. Interviews continued to 26 to strengthen diversity and depth rather than to increase statistical precision. The final sample of 26 participants is therefore described as providing adequate information for the study's descriptive aims.

Reporting Guideline

This study was reported with reference to the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist. The completed checklist is provided in Checklist 1.

Research Team and Reflexivity

Interviews were conducted by the primary researcher, the corresponding author, who had no prior relationship with participants. The interviewer introduced the study as independent research examining users' experiences of relapse in mHealth apps. The corresponding author, a female associate professor, conducted the interviews. She had prior experience in multiple research projects, data collection, and the application of research methods.

The research team discussed their assumptions concerning relapse, self-management, and digital health during study design and analysis. These discussions supported reflexive awareness of how researchers' expectations could influence the interpretation of participants' accounts and the framing of themes. Reflexivity was therefore used to make the researchers' interpretive role explicit rather than to claim that researcher influence could be eliminated. During analysis, initial codes and differing interpretations were discussed with reference to the transcripts, and analytic decisions were documented in an audit trail. These procedures supported transparent interpretation and helped limit the uncritical application of preexisting assumptions to participants' experiences.

Interview Guide and Data Collection

The semistructured interview guide was developed from the study objectives and relevant literature on relapse, self-management, and user experience of digital health tools. It was reviewed by 2 academic professors in the eHealth department and piloted with 3 eligible users who were not included in the final sample. Minor revisions were made to improve the clarity and sequencing of questions.

Before each interview, participants were informed that the study concerned experiences of dietary and exercise relapse while using apps for weight management and/or glycemic control. They received information about the interview procedure, voluntary participation, confidentiality, and their rights as research participants. Written informed consent was obtained before interviewing, including consent to audio recording and permission to publish anonymized quotations and nonidentifying participant characteristics. Participants were also informed about the anticipated interview duration and their right to decline individual questions or discontinue the interview. No nonparticipants were present during the interviews; only the participant and the interviewer attended each session.

Semistructured interviews were conducted through secure online videoconferencing apps. The average interview duration ranged from 40 to 50 minutes. Each participant was interviewed once; no repeat interviews were conducted. No field notes were taken during or after the interviews. Interviews were guided by questions focusing on key components, examining how individuals using weight and glycemic control apps interpret and respond to relapses in diet and exercise, how app features shape these experiences, and what forms of digital support they would find acceptable and helpful during relapse episodes. The interviews were audio recorded and conducted in Arabic with the consent of participants.

Data Analysis

Audio-recorded interviews were transcribed, translated into English, and checked by the authors to preserve the meaning of participants' accounts. Data were analyzed using inductive thematic analysis organized around Braun and Clarke's 6 phases: familiarization, initial coding, development of candidate themes, review of themes, definition and naming of themes, and report production. NVivo (version 14; Lumivero) supported data organization, coding, and theme refinement. Codes and themes were developed through engagement with participants' accounts rather than imposed from a predefined theoretical framework.

Two researchers independently coded all transcripts. Initial codes were compared during analytic meetings, and differences were resolved through discussion and reference to the transcripts. The team maintained an audit trail documenting code development, theme refinement, and decisions about theme boundaries. Following Braun and Clarke's clarification of the diversity of thematic analysis approaches [54], this procedure is reported as an inductive thematic analysis using independent

XSL•FO
RenderX

coding and discussion-based resolution, rather than characterized as their contemporary reflexive thematic analysis approach. The 6-phase framework organized the analysis, while the reported coding procedures describe how it was implemented in this study.

Coding was nonexclusive: participants could contribute to multiple themes and subthemes. Consequently, participant counts were used descriptively and were not interpreted as prevalence estimates. The analysis focused on patterns of meaning across accounts, including differences between digital disengagement and continued health-related activity. The final sample was considered adequate for the study's descriptive aims on the basis of the breadth and contextual depth of the accounts, as described in the *Recruitment and Sampling* subsection.

Trustworthiness

Trustworthiness was addressed through procedures supporting credibility, dependability, confirmability, and transferability. Credibility was supported by a review of the interview guide by 2 academic professors, pilot interviews with 3 eligible users outside the final sample, verification of translated transcripts by the authors, participant review of interview transcripts, and the presentation of participant quotations to substantiate interpretations. Transcripts were returned to participants for review and feedback, and a small number of changes were made in response to their comments. This process supported the accuracy of the interview records used for analysis. The reporting of contrasting experiences, including continued physical activity despite withdrawal from app monitoring, also helped preserve variation within the accounts.

Dependability was supported by a documented interview-guide development process and an audit trail recording code development, theme refinement, and decisions about theme boundaries. Confirmability was supported by linking interpretations to transcript material, discussing researchers' assumptions and differing interpretations, and documenting analytic decisions. These procedures made the interpretation process more transparent; they do not imply that the findings were independent of researcher judgment.

Transferability was supported by describing the recruitment strategy, eligibility criteria, participant characteristics, health goals, and cultural, social, and occupational circumstances surrounding relapse. These contextual details enable readers to assess the relevance of the findings to other populations and settings. The purposive sample was intended to provide diverse accounts rather than statistical representativeness, and transferability beyond this Saudi Arabian sample requires attention to contextual differences.

Ethical Considerations

This study adhered to strict ethical guidelines, including voluntary participation, confidentiality, and informed consent. Ethical approval (KFU-REC-2026-FEB--ETHICS4011) was obtained from the Research Ethics Committee at King Faisal University, Saudi Arabia, before conducting the interviews. Participants were fully informed about the interview procedure and their rights before starting the interviews. Written informed consent was obtained from all participants before the interviews. The consent process included permission to use anonymized quotations and nonidentifying participant characteristics in publications. Identifying details were removed or generalized before analysis and reporting. Data were anonymized or deidentified, and access was restricted to the research team.

Results

Participants' Sociodemographic Characteristics, Health Goals, and Patterns of Relapse Triggers

As shown in Table 1, a total of 26 individuals participated in the semistructured interviews. Participants were aged between 19 and 70 years and represented a diverse group, including students, working professionals, shift workers, homemakers, and retirees. Female representation was slightly higher (n=14) than male representation (n=12). Diverse health goals were observed among the participants, with 11 participants focused primarily on weight management, 8 on type 2 diabetes management, 5 on prediabetes, 1 on polycystic ovary syndrome (PCOS)–related weight concerns, and 1 on muscle gain/bulking. Various types of relapses were observed among the participants. Work- and stress-related relapse was prominent across 5 participants, followed by cultural and celebratory contexts (n=5), digital design–induced relapse (n=4), grief- and loss-related relapse (n=3), injury- or health flare–related relapse (n=3), psychological and emotional eating (n=3), monitoring burnout or anxiety-driven relapse (n=2), structural or occupational environmental barrier–based relapse (n=2), and goal incongruence–based relapse (n=1). Analyzing the relapse types, relapse triggers were distributed across emotional (8/26, approximately 31%), contextual/cultural (5/26, 19%), structural/occupational (2/26, 8%), injury/health-related (3/26, 12%), digital design–related (4/26, 15%), and social influence factors (including peer pressure and isolation; 8% - 12%), reflecting the multidimensional nature of relapse, which is influenced by personal, social, cultural, and technological factors.

XSL•FO
RenderX

Table . Participant demographics.

Participant ID	Age group (y)	Gender	Health condition/goal	Relapse context (brief description)
P01	30 - 39	Female	Weight loss/prediabetes	Work stress–induced dietary and exercise cessation
P02	40 - 49	Male	Type 2 diabetes	Family wedding–related dietary binge and exercise lapse
P03	20 - 29	Female	PCOS[a]-related weight issues	Hormonal flare-up leading to dietary cravings and yoga cessation
P04	50 - 59	Male	Type 2 diabetes	Grief-induced comfort eating and walking avoidance
P05	30 - 39	Female	Weight/gestational diabetes history	Child exam stress causing exercise neglect
P06	20 - 29	Male	Fitness/weight (bulking)	Injury-related gym halt and macronutrient deviations
P07	60 - 69	Female	Type 2 diabetes	Arthritis flare and festive sweets overindulgence
P08	30 - 39	Male	Prediabetes/weight	Night shift disruptions and junk food binges
P09	20 - 29	Female	Weight loss	Digital disengagement following dietary lapse
P10	50 - 59	Male	Type 2 diabetes	Contextual/cultural disruption (Ramadan)
P11	30 - 39	Female	Weight loss	Injury-related activity reduction with algorithmic misalignment
P12	40 - 49	Female	Weight/prediabetes	Stress-induced emotional eating and selective nonlogging
P13	20 - 29	Male	Weight gain (muscle)	Goal incongruence and underconsumption during stress
P14	60 - 69	Female	Type 2 diabetes	Grief-related disengagement from tracking
P15	30 - 39	Male	Weight loss	Competitive burnout and leaderboard disengagement
P16	40 - 49	Female	Glycemic control	Monitoring burnout and anxiety-driven avoidance
P17	40 - 49	Male	Type 2 diabetes	Environmental/structural dietary and logging avoidance
P18	20 - 29	Female	Weight management	Psychological (“The Streak Trap”)
P19	60 - 69	Male	Prediabetes	Cultural invisibility during family festivals
P20	19 - 29	Female	Weight management	Social/binging (peer pressure and alcohol)
P21	30 - 39	Male	Weight/hypertension	Decision fatigue (exercise and dietary)
P22	30 - 39	Female	Prediabetes	Chaos-induced glucose monitoring relapse
P23	40 - 49	Male	Weight management	The “Ostrich Effect” (exercise/data avoidance)

XSL•FO
RenderX

Participant ID	Age group (y)	Gender	Health condition/goal	Relapse context (brief description)
P24	20 - 29	Female	Weight management	Emotional eating feedback loop
P25	60 - 69	Male	Type 2 diabetes	Social isolation (dietary and testing)
P26	40 - 49	Female	Prediabetes/weight	Seasonal/holiday maintenance dietary relapse

[a]PCOS: polycystic ovary syndrome.

The analysis identified 7 overlapping themes: (1) stress and reduced capacity as conditions surrounding relapse; (2) cultural, temporal, and environmental misalignment; (3) psychological disengagement and avoidance; (4) social and relational influences; (5) relapse as emotional disengagement rather than simply a behavioral lapse; (6) informal relapse management and self-directed recalibration; and (7) desire for compassionate, context-aware support (see Table S1 in Multimedia Appendix 1). The themes are presented as interpretive patterns in participants' accounts. Participant counts were used only descriptively and were not treated as prevalence estimates. Coding was nonexclusive: a participant could contribute to more than 1 theme or subtheme, and subtheme counts therefore do not sum to the total sample.

Stress and Reduced Capacity as Conditions Surrounding Relapse

Participants commonly described relapse in the context of stress, exhaustion, grief, caregiving demands, academic or occupational pressure, and reduced physical capacity. These accounts suggested that relapse was often experienced as a reduction in available emotional or cognitive resources rather than as a lack of knowledge about healthy behavior.

A software engineer described a period of intense occupational pressure:

> *A major project launch meant 14-hour days, no time for meal prep, and biryani deliveries became my savior.*

She explained that logging and exercise stopped during this period, despite her previous success with both activities. Similarly, an intensive care nurse described relapse as a consequence of limited cognitive capacity after prolonged shifts:

> *Relapse isn't always about a lack of willpower; sometimes it's about a lack of bandwidth.*

The account linked missed logging and convenience eating to exhaustion and decision fatigue rather than to a change in health goals.

Caregiving demands produced a similar disruption. A participant caring for 3 children described family emergencies and mealtimes as circumstances in which her own monitoring routine was displaced:

> *When a kid gets sick, or the toddler has a tantrum during dinner, my health routine is the first thing to go out the window.*

A school teacher also described examination periods as a time when stress eating increased and logging became uncomfortable. Together, these accounts positioned relapse within competing demands and temporary reductions in capacity.

Physical and health-related changes also altered what participants could do. Participants with an injury, arthritis flare, or PCOS described reducing exercise or changing eating patterns while their capacity was affected. One participant with a knee injury stated,

> *I didn't stop moving. I switched to gentle yoga and stretching. But because those activities didn't 'count' toward the app's metrics, they felt invisible.*

This account shows that participants did not always experience behavioral disengagement and digital disengagement as the same process.

Cultural, Temporal, and Environmental Misalignment

Participants described relapse when the routines assumed by an app did not fit religious observances, festivals, family events, seasonal schedules, shift work, travel, or constrained food environments. These accounts did not necessarily indicate rejection of health goals. Instead, participants described the app as becoming less relevant when daily routines changed.

A participant managing type 2 diabetes described Ramadan as a period when eating times shifted and glucose readings became more variable. He stated,

> *Relapse wasn't rebellion; it was misalignment.*

The app continued to provide advice based on fixed meal timing, which he experienced as inconsistent with his religious practice. He wanted the app to recognize that

> *My lived experience wasn't recognized by the digital system guiding me.*

Cultural food practices were also described as a source of difficulty. A retired participant explained that the app's database did not adequately represent homemade biryani and festival foods:

> *I felt like the app was telling me that my culture and my health were incompatible.*

He described abandoning logging after repeated uncertainty about how to record traditional meals. Another participant described family weddings and social meals as events that interrupted exercise and dietary routines. Rather than avoiding such events, he wished for an event mode that would support flexible tracking and recovery after celebrations.

XSL•FO
RenderX

Work schedules and travel created additional environmental barriers. A shift worker described night rotations, disrupted sleep, and irregular meals. A long-haul truck driver described limited food choices during a winter storm and stated that the app seemed "designed for someone with a kitchen and a routine, not for someone whose life happens in a 70-foot vehicle." These accounts indicate that relapse was sometimes associated with structural constraints, including limited access to suitable food, irregular work hours, travel, and the physical environment in which self-management occurred.

Seasonal demands also affected engagement. A teacher described school holidays and end-of-term events as periods when weight loss was unrealistic. She wanted a maintenance option because

> *When the goal becomes 'perfection or nothing', people choose 'nothing' every time the holidays roll around.*

This account illustrates how rigid goals could become difficult to sustain during predictable periods of contextual change.

Psychological Disengagement and Avoidance

Participants described a process in which an initial lapse was followed by avoidance of the app, monitoring devices, or feedback. Avoidance took several forms, including not logging, removing or hiding devices, ignoring glucose readings, selectively recording food, or postponing reengagement until behavior appeared more acceptable.

A marketing executive described opening the app after several takeout meals and then closing it without recording anything: "It wasn't forgetfulness—it was avoidance." She associated the app with failure and described the loss of the reflective pause that logging had previously created. A participant managing weight described a similar process as "The Ostrich Effect," explaining, "If I don't wear the watch, the data doesn't exist." In both accounts, avoidance reduced immediate emotional discomfort but also interrupted monitoring.

Visual feedback and streak systems were particularly salient. One participant stated that red graphs and notifications produced guilt and led her to hide the app. Another described a 75-day streak ending after one missed entry:

> *If I can't have a perfect score, why bother playing the game?*

The participant interpreted the loss of the streak as a loss of status and subsequently stopped using the app for an extended period.

Some participants described selective nonlogging rather than complete disengagement. A school teacher explained that color-coded food categories made recording certain foods uncomfortable: "Red meant wrong." She began eating foods without recording them and later stopped logging altogether because she did not want a record of "bad days." A participant using a continuous glucose monitor described the opposite pattern: repeated checking of minor fluctuations increased anxiety until she avoided the app. Thus, disengagement could follow both undermonitoring and excessive monitoring.

The findings also showed that digital disengagement did not always correspond to complete behavioral disengagement. A participant recovering from injury continued gentle activity even though the app classified her reduced-intensity exercise as insufficient. Another participant continued occasional walking after withdrawing from a competitive leaderboard. These accounts suggest that app data may not fully represent participants' physical or self-management activities during relapse periods.

Social and Relational Influences

Participants described relapse as embedded in relationships and social settings. Family meals, weddings, peer drinking, parenting responsibilities, bereavement, and social isolation all influenced whether participants could maintain routines or felt able to reengage with an app.

A university student described tracking as socially uncomfortable during evenings with friends:

> *If I pull out my phone to track a drink, I feel like a total buzzkill.*

She experienced a conflict between being healthy and maintaining her social identity. Her account suggested that the app made these identities feel incompatible during social occasions.

Family contexts could either disrupt or support reengagement. A business owner described negotiating home workouts with his wife after a wedding-related lapse. A widowed participant stated that her daughter's encouragement helped her return to glucose monitoring:

> *Human connection restarted what technology couldn't.*

Another older participant living alone described relapse as connected to loneliness and requested a check-in system that could alert a trusted person when prolonged disengagement occurred.

Participants also described social features that were not uniformly helpful. Competitive leaderboards initially motivated one participant but later became discouraging when he fell behind. He explained,

> *Relapse here wasn't about inactivity—it was discouragement.*

These accounts indicate that social accountability may support reengagement when it is experienced as relational and supportive, but may contribute to withdrawal when it is experienced as comparison or public failure.

Relapse as Emotional Disengagement

Participants frequently described relapse using emotional terms, including guilt, shame, discouragement, hopelessness, loss of meaning, identity threat, and indifference. These accounts did not imply that every participant understood relapse in the same way. Rather, they illustrate a recurring interpretation in which emotional responses shaped how a lapse was understood and whether the participant returned to the app.

XSL•FO
RenderX

A software engineer described the experience as a threat to her identity: "I'm a failure at this too." A participant with PCOS described feeling that her body had betrayed her during a flare-up. A participant managing diabetes after bereavement stated,

> *Relapse, in that period, was existential. It wasn't rebellion or ignorance. It was emotional depletion.*

These accounts framed relapse as a disruption in meaning and self-understanding, not simply as a failure to perform a health behavior.

Participants also described a difference between a behavioral lapse and an emotional withdrawal from self-management. One participant summarized this distinction as follows:

> *For me, relapse is less about food and more about losing the reflective habit that the app initially helped build.*

Another participant continued gentle physical activity after an injury but considered herself digitally unsuccessful because the app did not recognize the activity. Such accounts suggest that participants evaluated relapse through both their lived behavior and the app's representation of that behavior.

Informal Relapse Management and Self-Directed Recalibration

Participants generally described managing relapse through improvised, self-directed strategies rather than through structured recovery tools within the apps. Strategies included restarting on a chosen date, reducing goals, switching to simpler activities, recording only selected measures, using journaling or professional advice, and seeking support from family or friends.

Several participants described temporal resets, such as restarting on Monday or after a holiday. A participant experiencing social eating described a "Weekend Void" followed by an intention to "start fresh on Monday." Another participant used journaling and discussions with clinicians to manage a health-related flare-up. These strategies helped participants regain a sense of control but were not usually supported by an explicit relapse pathway in the app.

Participants also recalibrated goals when full adherence was unrealistic. Examples included using lighter activity during injury, adjusting goals during night shifts, choosing maintenance rather than weight loss during holidays, and simplifying food recording during long work shifts. A participant working in intensive care wanted an app that would reduce demands during high-capacity periods and ask for a simpler account of eating rather than exact measurements.

Partial engagement was another form of self-management. Participants sometimes measured glucose, continued some movement, or recorded selected meals while avoiding complete tracking. One participant described restarting gradually but found partial entries discouraging because the app treated them as incomplete. These accounts suggest that participants were attempting to maintain a connection with self-management, even when they could not meet the app's preferred level of monitoring.

Desire for Compassionate and Context-Aware Support

Participants consistently described a preference for support that recognized emotional states, changing circumstances, health conditions, and temporary limitations. Their suggestions included pause or crisis modes, event and festival modes, maintenance goals, injury-sensitive plans, mood tracking, neutral feedback, simplified logging, accessible interfaces, and gentle reentry pathways.

A participant experiencing work stress wanted "a pause mode for high-stress weeks" with a smaller goal, such as logging 1 meal. A participant managing type 2 diabetes during Ramadan wanted seasonal or cultural adaptation. A participant with an injury wanted to record that she was recovering so that the app could recalibrate activity expectations. These suggestions shared the assumption that support should adapt to changing capacity rather than continue to evaluate users against a fixed routine.

Participants also requested emotional and relational support. Suggestions included mood tracking, reflection prompts, grief resources, peer stories, family support, check-in buddies, and community connections. A participant experiencing emotional eating stated,

> *When the app only treats the 'what' and not the 'why', it's only doing half the job.*

An older participant living alone wanted the app to function as "a bridge to other people, not just a digital diary."

Gentle reentry was a central feature of participants' desired support. Participants described wanting fresh-start options, archived difficult periods, blind weighing, reset streaks, maintenance modes, and neutral language. One participant wished for a system that would separate logging from judgment, while another wanted to "archive" a difficult week without deleting previous progress. Overall, participants described relapse-responsive support as flexible, nonpunitive, and attentive to the circumstances in which health behaviors occur.

The overall findings indicate that participants experienced relapse as a dynamic interaction among stress, reduced capacity, contextual misalignment, psychological responses to digital feedback, and social circumstances. Participants described self-directed attempts to recover but reported few structured app features for managing relapse. Their accounts pointed to the importance of flexible goals, culturally and contextually responsive interfaces, neutral feedback, and supportive reentry. These findings are interpretive and should not be read as estimates of the frequency of these experiences among all mHealth app users.

Discussion

Principal Findings and Contribution to mHealth

The findings provide deeper insights into how users experience, interpret, and manage dietary and exercise relapse within mHealth apps for weight and glycemic control. Consistent with the Relapse Prevention Model, participants described stress, social events, health-related limitations, and contextual instability as circumstances surrounding relapse. However, their accounts also highlighted emotional disengagement, including

XSL•FO
RenderX

guilt, shame, discouragement, and identity disruption. These findings complement research on goal-discrepant feedback and mHealth engagement by illustrating how users may interpret app-generated information during periods of difficulty.

An important distinction was that digital disengagement did not always correspond to complete behavioral disengagement. Some participants continued walking, adapted their exercise, or monitored selected health indicators despite withdrawing from an app. Consequently, an absence of app entries may indicate tracking burden, discomfort with feedback, or a mismatch between app expectations and users' circumstances rather than abandonment of health goals. This finding adds a user-centered perspective to research examining disengagement from self-monitoring and declining mHealth engagement [20-22].

Participants reported limited structured support for recovering from relapse within the apps they used. Their accounts suggest a possible sequence in which an initial lapse is followed by negative self-evaluation, avoidance of feedback, and delayed reentry. However, the qualitative design does not establish mediation or causality, and these accounts should not be interpreted as evidence that specific features caused disengagement. Instead, they identify potentially important processes for investigation and highlight the need to distinguish support for behavior initiation from support for recovery and maintenance [40].

Emotional Responses and Recovery After Relapse

Participants' accounts suggest that relapse-responsive support should address the emotional meaning attributed to setbacks as well as the behaviors themselves. Broken streaks, unmet targets, and evaluative displays were sometimes interpreted as evidence of personal failure or loss of previous progress. These experiences provide context for research examining engagement following goal-discrepant feedback. Rather than treating feedback as inherently motivating, developers should consider how its language, visual presentation, and timing may be experienced by users whose capacity has temporarily declined.

A practical implication is to combine self-monitoring with supportive reflection and gradual recovery. An app could acknowledge previous progress, invite users to identify the circumstances surrounding a lapse, and support the selection of a manageable next action. For example, a user returning after a stressful week might choose to record one meal or complete a short activity rather than immediately restore all previous targets. These proposals translate participants' existing strategies of partial engagement and goal recalibration into structured support and are conceptually consistent with the role of self-efficacy and coping in weight maintenance.

Compassionate communication should not imply that health goals are unimportant. Rather, it could separate information about behavior from judgments about personal worth. Neutral wording, recognition of partial completion, and preservation of accumulated progress may make reentry more acceptable for users who experience all-or-nothing responses to missed targets. These are user-informed design proposals, however, and their effects on emotional responses, self-management, and sustained engagement require evaluation.

Relapse-Responsive Features and Behavior Change Strategies

The findings identify opportunities to translate established behavior change strategies into relapse-responsive app features. Research on digital weight-management interventions has examined self-monitoring, goal setting, feedback, and persuasive design approaches [37-40]. This study extends this discussion by considering how these strategies could be adapted when users experience stress, emotional distress, injury, or disruption to their usual routines. Table 2 summarizes the relationships between the study findings, proposed relapse-responsive app features, corresponding behavior change strategies, and priorities for evaluation.

XSL•FO
RenderX

Table . Proposed relapse-responsive app features and behavior change strategies.

Study finding	Proposed app feature	Corresponding behavior change strategy	Evaluation priority
Stress, exhaustion, and reduced capacity	Optional low-demand mode with simplified logging, fewer nonessential prompts, and smaller goals	Graded tasks, goal review, and action planning	Perceived tracking burden and maintenance of feasible self-management
Cultural, temporal, and environmental misalignment	User-selected Ramadan, celebration, travel, or shift-work settings and culturally relevant food entries	Context-specific planning and problem solving	Cultural acceptability, usability, and clinical appropriateness
Shame and avoidance following evaluative feedback	Neutral language, recognition of partial completion, and preservation of accumulated progress	Supportive feedback and reflection on setbacks	Emotional responses, avoidance, and willingness to resume monitoring
All-or-nothing responses to missed targets	Graded reentry without mandatory completion of missing records or punitive streak resets	Graded tasks and review of achievable goals	Time to reengagement and sustained self-management after a lapse
Social isolation or discouraging comparison	Optional trusted-person check-ins, noncompetitive peer support, and user-controlled sharing	Emotional and practical social support	Acceptability, perceived autonomy, privacy, and supportive contact
Injury or health-related limitations	Capacity-sensitive activity goals and temporary maintenance options within clinically appropriate boundaries	Goal adjustment and problem solving	Safety, feasibility, and recognition of adapted health-related activity

For participants experiencing reduced capacity, an optional low-demand mode could simplify food recording, reduce nonessential prompts, and support smaller, achievable goals. Goal review and graded tasks could allow users to maintain some connection with self-management without requiring full adherence immediately. Similarly, temporary maintenance goals or capacity-sensitive activity targets could recognize that weight loss or usual exercise intensity may not be feasible during every period. These proposals align with user-centered approaches to adaptive interfaces and personalized recommendations, but their effectiveness during relapse remains to be established [50].

Gentle reentry pathways could address the difficulty participants described when attempting to resume app use. Possible features include restarting without retrospectively completing missing records, retaining previous achievements after a broken streak, and allowing users to select a limited set of activities before restoring their usual routine. Such features would recognize recovery as a process rather than a single decision to resume tracking. Importantly, these proposals should complement, but not replace, clinically appropriate self-management guidance.

Social support features also require differentiation between supportive contact and competitive comparison. Participants described family encouragement and trusted-person contact as helpful, whereas leaderboards could become discouraging when they fell behind. Optional check-ins, noncompetitive peer support, and user-controlled sharing may therefore offer acceptable recovery pathways for some users. Their design should preserve choice and privacy rather than assume that increased visibility or accountability benefits everyone.

Cultural and Contextual Adaptation

Cultural and contextual misalignment was a prominent concern in participants' accounts. Religious observances, celebrations, traditional foods, shift work, travel, and constrained food environments altered the feasibility of fixed tracking expectations. These findings complement qualitative research examining the fit between digital self-management tools and users' everyday circumstances [47-49]. They suggest that personalization should extend beyond demographic characteristics to changing routines, available resources, and self-management capacity.

For example, user-selected Ramadan, celebration, travel, or shift-work settings could adjust optional reminder schedules and support context-specific planning. Culturally relevant food entries and simplified recording options could reduce difficulties associated with logging traditional meals. However, such features should accommodate individual variation rather than assume that all users within a cultural or religious group follow the same practices. The relevance of these adaptations should be established through user involvement, consistent with user-centered development approaches [50].

Adaptation should also begin with user confirmation rather than automatic interpretation of missing data as relapse. A missed entry alone cannot establish whether a person is struggling, successfully adapting their behavior, or choosing not to use the app. A neutral invitation to select a smaller goal, change a schedule, or pause optional reminders could provide support while preserving autonomy. This distinction is particularly important given participants' accounts of continued health-related activity despite reduced digital engagement.

Broader Implications and Future Evaluation

The broader implication for mHealth is that sustained self-management should not be equated with uninterrupted app interaction. Research has documented challenges in maintaining engagement and sustaining the benefits of app-based interventions [20,21,34-36]. These findings suggest that useful digital support may involve helping users navigate disruption, preserve feasible behaviors, and return after a lapse rather than prioritizing continuous tracking or perfect streaks.

XSL•FO
RenderX

Relapse-responsive features should therefore be evaluated using outcomes that distinguish digital activity from health behavior. In addition to app use, relevant outcomes could include time to reengagement, recovery after a lapse, perceived tracking burden, emotional responses to feedback, and sustained self-management. Weight and glycemic outcomes should also be examined where appropriate, given evidence linking self-monitoring adherence with weight-management outcomes. Increased interaction alone would not establish benefit if it also increased anxiety or burden [22].

Implementation must also distinguish reduced interaction demands from reduced clinically necessary care. Low-demand modes could simplify optional logging or motivational reminders, but features used in glycemic-control apps would require clinical review to ensure that essential monitoring and safety functions remain appropriate. Mood-related information and trusted-person notifications would similarly require clear consent, privacy protections, and user control. These are considerations for future development rather than outcomes assessed in this study.

Overall, the findings provide a user-informed basis for developing flexible goals, neutral feedback, contextual adaptation, optional relational support, and graded reentry pathways. Participatory development and prospective evaluation are needed to determine their acceptability, safety, and effectiveness. The study's contribution is therefore to identify how digital support might remain useful during periods of reduced capacity and disrupted routines, rather than to demonstrate that these proposed features improve engagement or clinical outcomes.

Limitations and Future Research

Several limitations should be considered. Interviews were conducted in Arabic and translated into English; subtle changes in meaning may have occurred despite author verification. The purposive sample comprised 26 adults using apps for weight and/or glycemic control in Saudi Arabia. Although participants differed in age, gender, occupation, health goals, and relapse circumstances, the study does not support statistical generalization. Recruitment through online communities and university mailing lists may also have favored individuals willing and able to discuss digital self-management.

Transferability depends on similarities between the study context and other settings, particularly concerning religious observances, food practices, family relationships, occupational schedules, and access to suitable food or activity opportunities. Retrospective accounts are also susceptible to recall and social desirability influences. The analysis involved researcher interpretation supported by independent coding, transcript-based discussion, and documented analytic decisions; these procedures do not eliminate interpretive influence. Although interview transcripts were returned to participants for feedback and a small number of changes were made in response, this process supported the accuracy of the interview records rather than establishing the validity of the researchers' thematic interpretations.

The study did not independently assess app functionality, usage logs, behavioral trajectories, or clinical outcomes. Participants' reports of limited recovery support should therefore not be interpreted as an objective assessment of all apps or proof that such features were absent. Likewise, the study does not establish causal relationships among app feedback, emotional responses, and disengagement. The proposed relapse-responsive features are user-informed design hypotheses requiring participatory development and prospective evaluation.

Conclusions

Participants described dietary and exercise relapse in mHealth contexts as involving emotional disengagement, changing capacity, contextual misalignment, and social circumstances, rather than simply failure to meet behavioral targets. Digital disengagement did not always indicate abandonment of health-related activity. These findings identify opportunities for relapse-responsive support through flexible goals, neutral feedback, culturally appropriate adaptation, optional relational support, and graded reentry. Their acceptability, safety, and effectiveness require evaluation; this study provides a user-informed basis for developing such interventions rather than evidence of their clinical benefit.

Acknowledgments

The authors thank all participants for sharing their experiences. The authors also acknowledge the support of King Faisal University.

Generative AI Use

Generative AI, Paperpal (version 2.1.2), was used solely to assist with language editing and drafting. The authors reviewed, revised, and take full responsibility for all content, accuracy, citations, and conclusions in the manuscript.

Funding

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (KFU261183).

Data Availability

The deidentified interview data are not publicly available because they contain potentially identifiable personal information and because the consent and ethics approval did not permit unrestricted public sharing. Deidentified data may be available from the

XSL•FO
RenderX

corresponding author on reasonable request, subject to approval by the relevant ethics committee and applicable data-protection requirements.

Authors' Contributions

All 3 authors contributed substantially to this work. They were responsible for the study conception and design, data collection, analysis and interpretation of results, and manuscript preparation. All authors reviewed and approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Key themes emerged from the data analysis.
[DOCX File, 19 KB - mhealth_v14i1e94792_app1.docx]

Checklist 1
COREQ checklist.
[DOCX File, 19 KB - mhealth_v14i1e94792_app2.docx]

References

1. Hacker K. The burden of chronic disease. Mayo Clin Proc Innov Qual Outcomes 2024 Feb;8(1):112-119. [doi: 10.1016/j.mayocpiqo.2023.08.005] [Medline: 38304166]
2. Fenta ET, Eshetu HB, Kebede N, et al. Prevalence and predictors of chronic kidney disease among type 2 diabetic patients worldwide, systematic review and meta-analysis. Diabetol Metab Syndr 2023 Nov 28;15(1):245. [doi: 10.1186/s13098-023-01202-x] [Medline: 38012781]
3. Al Wachami N, Guennouni M, Iderdar Y, et al. Estimating the global prevalence of chronic obstructive pulmonary disease (COPD): a systematic review and meta-analysis. BMC Public Health 2024 Jan 25;24(1):297. [doi: 10.1186/s12889-024-17686-9] [Medline: 38273271]
4. Gropper SS. The role of nutrition in chronic disease. Nutrients 2023 Jan 28;15(3):664. [doi: 10.3390/nu15030664] [Medline: 36771368]
5. Núñez-Cortés R, Salazar-Méndez J, Nijs J. Physical activity as a central pillar of lifestyle modification in the management of chronic musculoskeletal pain: a narrative review. J Funct Morphol Kinesiol 2025 May 20;10(2):183. [doi: 10.3390/jfmk10020183] [Medline: 40407467]
6. Andy A. Empowering lifestyle changes in preventing and managing type 2 diabetes. J Diabetes Res Rev Rep 2022;4(2):1. [doi: 10.47363/JDRR/2022(4)E101]
7. Nisar K, Naz K, Ikram A, Rasheed I. Analyzing the exercise self-efficacy effect on diabetes self-management through a mediating and moderating mechanism: a quantitative study. Int J Healthc Manag 2025 Apr 3;18(2):378-385. [doi: 10.1080/20479700.2024.2310377]
8. Budny A, Janczy A, Szymanski M, Mika A. Long-term follow-up after bariatric surgery: key to successful outcomes in obesity management. Nutrients 2024 Dec 21;16(24):4399. [doi: 10.3390/nu16244399] [Medline: 39771020]
9. Grothe KB, Koball AM, Ames GE, Czerwony BA. Long-term weight maintenance and dietary and behavioral management of weight recurrence. In: Bhasker AG, Craggs-Dino L, O'Kane M, Jain V, editors. Handbook of Bariatric Nutrition: Springer; 2025:299-315. [doi: 10.1007/978-981-96-5099-6_17]
10. Aggarwal I, O'Brien MJ, Pauls AM, et al. Renewal of challenging behavior in an intensive outpatient clinic: replication and extension to task changes. J Appl Behav Anal 2026 Apr;59(2):e70057. [doi: 10.1002/jaba.70057] [Medline: 41674209]
11. Hermanns N, Ehrmann D, Shapira A, Kulzer B, Schmitt A, Laffel L. Coordination of glucose monitoring, self-care behaviour and mental health: achieving precision monitoring in diabetes. Diabetologia 2022 Nov;65(11):1883-1894. [doi: 10.1007/s00125-022-05685-7] [Medline: 35380233]
12. Pérez-Osorio KC, Jiménez-Nieto YA, González-Diaz JE, Rodríguez-Vásquez G, Rodríguez-Parada A, Juárez-Rivera O. Mobile apps to improve diabetes mellitus self-care: a scoping review. In: Mejía J, Muñoz M, Rocha A, Espinosa-Faller FJ, Trejo-Sanchez JA, editors. New Challenges in Software Engineering: Springer; 2025:405-431. [doi: 10.1007/978-3-031-90310-6_26]
13. Spinean A, Carniciu S, Mladin OA, Serafinceanu C. The transformative power of mHealth apps: empowering patients with obesity and diabetes - a narrative review. J Med Life 2024 Dec;17(12):1030-1035. [doi: 10.25122/jml-2024-0340] [Medline: 39877040]

XSL•FO
RenderX

14. Prasad G, Padhiary M, Hoque A, Kumar K. AI-driven personalized nutrition apps and platforms for enhanced diet and wellness. In: Alhussaini Hamad AM, Soni R, editors. Food in the Metaverse and Web 3.0 Era: Intersecting Food, Technology, and Culture: IGI Global Scientific Publishing; 2025:125-158. [doi: 10.4018/979-8-3693-9025-2.ch006]
15. Chew HSJ, Koh WL, Ng J, Tan KK. Sustainability of weight loss through smartphone apps: systematic review and meta-analysis on anthropometric, metabolic, and dietary outcomes. J Med Internet Res 2022 Sep 21;24(9):e40141. [doi: 10.2196/40141] [Medline: 36129739]
16. Cui M, Wu X, Mao J, Wang X, Nie M. T2DM self-management via smartphone applications: a systematic review and meta-analysis. PLoS One 2016;11(11):e0166718. [doi: 10.1371/journal.pone.0166718] [Medline: 27861583]
17. Cavero-Redondo I, Martinez-Vizcaino V, Fernandez-Rodriguez R, Saz-Lara A, Pascual-Morena C, Álvarez-Bueno C. Effect of behavioral weight management interventions using lifestyle mHealth self-monitoring on weight loss: a systematic review and meta-analysis. Nutrients 2020 Jul 3;12(7):1977. [doi: 10.3390/nu12071977] [Medline: 32635174]
18. Li S, Zhou Y, Tang Y, et al. Behavior change resources used in mobile app-based interventions addressing weight, behavioral, and metabolic outcomes in adults with overweight and obesity: systematic review and meta-analysis of randomized controlled trials. JMIR mHealth uHealth 2025 Aug 19;13:e63313. [doi: 10.2196/63313] [Medline: 40829125]
19. Mazzotti VL, Shogren KA, Stewart-Ginsburg JH, et al. The goal setting challenge app: promoting self-determination through technology. Remedial Spec Educ 2024 Feb;45(1):3-17. [doi: 10.1177/07419325221147698]
20. Mustafa AS, Ali N, Dhillon JS, Alkawsi G, Baashar Y. User engagement and abandonment of mHealth: a cross-sectional survey. Healthcare (Basel) 2022 Jan 24;10(2):221. [doi: 10.3390/healthcare10020221] [Medline: 35206837]
21. Eaton C, Vallejo N, McDonald X, et al. User engagement with mHealth interventions to promote treatment adherence and self-management in people with chronic health conditions: systematic review. J Med Internet Res 2024 Sep 24;26:e50508. [doi: 10.2196/50508] [Medline: 39316431]
22. Carpenter CA, Eastman A, Ross KM. Consistency with and disengagement from self-monitoring of weight, dietary intake, and physical activity in a technology-based weight loss program: exploratory study. JMIR Form Res 2022 Feb 18;6(2):e33603. [doi: 10.2196/33603] [Medline: 35179513]
23. Deniz-Garcia A, Fabelo H, Rodriguez-Almeida AJ, et al. Quality, usability, and effectiveness of mHealth apps and the role of artificial intelligence: current scenario and challenges. J Med Internet Res 2023 May 4;25:e44030. [doi: 10.2196/44030] [Medline: 37140973]
24. Giebel GD, Speckemeier C, Abels C, et al. Problems and barriers related to the use of digital health applications: scoping review. J Med Internet Res 2023 May 12;25:e43808. [doi: 10.2196/43808] [Medline: 37171838]
25. Eaves ER, Doerry E, Lanzetta SA, et al. Applying user-centered design in the development of a supportive mHealth app for women in substance use recovery. Am J Health Promot 2023 Jan;37(1):56-64. [doi: 10.1177/08901171221113834] [Medline: 35815770]
26. Larimer ME, Palmer RS, Marlatt GA. Relapse prevention. An overview of Marlatt's cognitive-behavioral model. Alcohol Res Health 1999;23(2):151-160. [Medline: 10890810]
27. Hurley L, O'Shea NG, Nezami BT, Valle CG, Tate DF. Fluctuations in mHealth engagement following receipt of goal-discrepant feedback messages. Digit Health 2025;11:20552076241312569. [doi: 10.1177/20552076241312569] [Medline: 39801582]
28. Coenen T, Maerevoet M, Chen S, et al. Requirements and concerns of individuals remitted from depression for an early relapse detection mHealth app: focus group study. JMIR mHealth uHealth 2025 Oct 23;13:e67141. [doi: 10.2196/67141] [Medline: 41130588]
29. Wills F, Sanders D. Process difficulties and endings in CBT. In: Cognitive Behaviour Therapy: Foundations for Practice, 3rd edition: SAGE Publications; 2013:244-270. [doi: 10.4135/9781526435651.n10]
30. Wilkinson H, McGraw C, Chung K, Kyratsis Y. "Can I exercise? Would it help? Would it not?": exploring the experiences of people with relapsing remitting multiple sclerosis engaging with physical activity during a relapse: a qualitative study. Disabil Rehabil 2023 Jun;45(13):2095-2106. [doi: 10.1080/09638288.2022.2084774] [Medline: 35727957]
31. Swencionis C, Smith-Wexler L, Lent MR, et al. Triggers of lapse and relapse of diet and exercise in behavioral weight loss. Obesity (Silver Spring) 2019 Jun;27(6):888-893. [doi: 10.1002/oby.22437] [Medline: 31033215]
32. Żurawski T, Bartosiewicz A. Dietary and behavioral strategies for weight loss and weight loss maintenance: a narrative review. Nutrients 2025 Dec 19;18(1):12. [doi: 10.3390/nu18010012] [Medline: 41515130]
33. Hu T, Bao Y. Long-term management of body weight and physiological mechanisms of weight regain. J Mol Cell Biol 2026 Jul 3;18:mjaf062. [doi: 10.1093/jmcb/mjaf062] [Medline: 41489473]
34. Ufholz K, Werner J. The efficacy of mobile applications for weight loss. Curr Cardiovasc Risk Rep 2023;17(4):83-90. [doi: 10.1007/s12170-023-00717-2] [Medline: 36974130]
35. Szinay D, Forbes CC, Busse H, DeSmet A, Smit ES, König LM. Is the uptake, engagement, and effectiveness of exclusively mobile interventions for the promotion of weight‐related behaviors equal for all? A systematic review. Obes Rev 2023 Mar;24(3):e13542. [doi: 10.1111/obr.13542] [Medline: 36625062]
36. Gemesi K, Winkler S, Schmidt-Tesch S, Schederecker F, Hauner H, Holzapfel C. Efficacy of an app-based multimodal lifestyle intervention on body weight in persons with obesity: results from a randomized controlled trial. Int J Obes 2024 Jan;48(1):118-126. [doi: 10.1038/s41366-023-01415-0] [Medline: 38017117]

XSL•FO
RenderX

37. Yen HY, Jin G, Chiu HL. Smartphone app-based interventions targeting physical activity for weight management: a meta-analysis of randomized controlled trials. Int J Nurs Stud 2023 Jan;137:104384. [doi: 10.1016/j.ijnurstu.2022.104384] [Medline: 36379104]
38. Villinger K, Wahl DR, Boeing H, Schupp HT, Renner B. The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: a systematic review and meta-analysis. Obes Rev 2019 Oct;20(10):1465-1484. [doi: 10.1111/obr.12903] [Medline: 31353783]
39. Burke LE, Bizhanova Z, Conroy MB, et al. Adherence to self-monitoring and behavioral goals is associated with improved weight loss in an mHealth randomized-controlled trial. Obesity (Silver Spring) 2025 Mar;33(3):478-489. [doi: 10.1002/oby.24234] [Medline: 39962997]
40. Asbjørnsen RA, Smedsrød ML, Solberg Nes L, et al. Persuasive system design principles and behavior change techniques to stimulate motivation and adherence in electronic health interventions to support weight loss maintenance: scoping review. J Med Internet Res 2019 Jun 21;21(6):e14265. [doi: 10.2196/14265] [Medline: 31228174]
41. Tay A, Hoeksema H, Murphy R. Uncovering barriers and facilitators of weight loss and weight loss maintenance: insights from qualitative research. Nutrients 2023 Mar 6;15(5):1297. [doi: 10.3390/nu15051297] [Medline: 36904294]
42. Li SJ, Zhou Y, Ma HM, et al. Effects of mobile applications on weight-related, behavior and metabolic outcomes for adults with overweight and obesity: a systematic review and meta-analysis. Stud Health Technol Inform 2024 Jul 24;315:386-391. [doi: 10.3233/SHTI240175] [Medline: 39049288]
43. Walle AD, Ferede TA, Shibabaw AA, et al. Willingness of diabetes mellitus patients to use mHealth applications and its associated factors for self-care management in a low-income country: an input for digital health implementation. BMJ Health Care Inform 2023 May;30(1):e100761. [doi: 10.1136/bmjhci-2023-100761] [Medline: 37236653]
44. Sands DZ. Beyond the EHR: how digital health tools foster participatory health and self-care for patients with diabetes. Am J Med Open 2023 Dec;10:100043. [doi: 10.1016/j.ajmo.2023.100043] [Medline: 39035248]
45. Milne-Ives M, Homer SR, Andrade J, Meinert E. Potential associations between behavior change techniques and engagement with mobile health apps: a systematic review. Front Psychol 2023;14:1227443. [doi: 10.3389/fpsyg.2023.1227443] [Medline: 37794916]
46. Solh Dost L, Gastaldi G, Schneider MP. The power of active listening to address medication non-adherence during care transition: a case report of a polypharmacy patient with type 2 diabetes. Pharmacy (Basel) 2025 Apr 30;13(3):64. [doi: 10.3390/pharmacy13030064] [Medline: 40407502]
47. Baptista S, Wadley G, Bird D, Oldenburg B, Speight J, My Diabetes Coach Research Group. User experiences with a type 2 diabetes coaching app: qualitative study. JMIR Diabetes 2020 Jul 17;5(3):e16692. [doi: 10.2196/16692] [Medline: 32706649]
48. Breland JY, Agha K, Mohankumar R. Adoption and appropriateness of mHealth for weight management in the real world: a qualitative investigation of patient perspectives. JMIR Form Res 2021 Dec 8;5(12):e29916. [doi: 10.2196/29916] [Medline: 34889761]
49. Kabeza CB, Harst L, Schwarz PEH, Timpel P. A qualitative study of users' experiences after 3 months: the first Rwandan diabetes self-management Smartphone application "Kir'App". Ther Adv Endocrinol Metab 2020;11:2042018820914510. [doi: 10.1177/2042018820914510] [Medline: 32523670]
50. Wang W, Grundy J, Khalajzadeh H, Madugalla A, Obie HO. Designing adaptive user interfaces for mHealth applications targeting chronic disease: a user-centered approach. ACM Trans Softw Eng Methodol 2026 Feb 28;35(2):1-57. [doi: 10.1145/3731750]
51. Chatterjee A, Prinz A, Gerdes M, Martinez S, Pahari N, Meena YK. ProHealth eCoach: user-centered design and development of an eCoach app to promote healthy lifestyle with personalized activity recommendations. BMC Health Serv Res 2022 Sep 4;22(1):1120. [doi: 10.1186/s12913-022-08441-0] [Medline: 36057715]
52. Latner JD, McLeod G, O'Brien KS, Johnston L. The role of self-efficacy, coping, and lapses in weight maintenance. Eat Weight Disord 2013 Dec;18(4):359-366. [doi: 10.1007/s40519-013-0068-1] [Medline: 24078407]
53. Squire CM, Giombi KC, Rupert DJ, Amoozegar J, Williams P. Determining an appropriate sample size for qualitative interviews to achieve true and near code saturation: secondary analysis of data. J Med Internet Res 2024 Jul 9;26:e52998. [doi: 10.2196/52998] [Medline: 38980711]
54. Braun V, Clarke V. Toward good practice in thematic analysis: avoiding common problems and be(com)ing a knowing researcher. Int J Transgend Health 2023;24(1):1-6. [doi: 10.1080/26895269.2022.2129597] [Medline: 36713144]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
mHealth: mobile health
PCOS: polycystic ovary syndrome

XSL•FO
RenderX

Edited by L Buis; submitted 06.Mar.2026; peer-reviewed by J Hill, M Ershadmanesh, V Giannouli; revised version received 09.Oct.2026; accepted 21.Oct.2026; published 17.Nov.2026.

Please cite as:
Alhajri AS, Alandanoosi AI, Al-Otaibi HH
Dietary and Exercise Relapse Management in Mobile Health Apps for Weight and Glycemic Control: Descriptive Qualitative Study
JMIR Mhealth Uhealth 2026;14:e94792
URL: https://mhealth.jmir.org/2026/1/e94792
doi:10.2196/94792

© Ahlam Saleh Alhajri, Afnan I Alandanoosi, Hala Hazam Al-Otaibi. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 17.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effects of a Hypertension Management Mobile App on Urinary Sodium Excretion in Patients With Chronic Kidney Disease: Randomized Controlled Trial

Takayuki Kawaoka[1*], MD, PhD; Yusuke Sakaguchi[1*], MD, PhD; Tatsufumi Oka[1], MD, PhD; Yohei Doi[1], MD, PhD; Ryohei Yamamoto[1,2], MD, PhD; Isao Matsui[1], MD, PhD; Masayuki Mizui[1], MD, PhD; Jun-Ya Kaimori[1], MD, PhD; Yoshitaka Isaka[1], MD, PhD

[1]Department of Nephrology, Graduate School of Medicine, The University of Osaka, 2-2-D11, Yamada-oka, Suita, Japan

[2]Health and Counseling Center, The University of Osaka, Toyonaka, Japan

[*]these authors contributed equally

Corresponding Author:
Yusuke Sakaguchi, MD, PhD
Department of Nephrology, Graduate School of Medicine, The University of Osaka, 2-2-D11, Yamada-oka, Suita, Japan

Abstract

Background: Excessive salt intake is detrimental to the kidneys. Nevertheless, salt restriction is often suboptimal in patients with chronic kidney disease (CKD). Smartphone app–based interventions might help reduce salt intake by supporting self-monitoring and behavior change at scale. However, clinical trials evaluating these interventions for salt reduction are limited, particularly in CKD populations.

Objective: This study investigated whether a hypertension management app could reduce urinary sodium excretion in patients with CKD.

Methods: This open-label, single-center, randomized clinical trial included 101 patients with CKD who had a history of hypertension and estimated 24-hour urinary sodium excretion of 100 mmol or greater. Patients in the intervention group used CureApp HT, a smartphone app designed to manage hypertension through lifestyle modifications and self-monitoring, particularly for salt restriction. The app delivered daily, individualized guidance tailored to each patient's lifestyle. Patients also received lifestyle counseling by nephrologists during outpatient visits. The control group received lifestyle counseling alone. The intervention period was 12 weeks. The primary outcome was the change in estimated 24-hour urinary sodium excretion from baseline to week 12, calculated from spot urine samples using the Tanaka method. Key secondary outcomes included office blood pressure, brachial-ankle pulse wave velocity, urinary protein-to-creatinine ratio, and plasma brain natriuretic peptide. The analysis was conducted in the intention-to-treat population, using a mixed-effects model for repeated measures.

Results: A total of 101 patients were randomly assigned to the intervention group (n=51) or the control group (n=50). The median (IQR) app engagement rate, calculated by dividing the number of days patients recorded blood pressure in the app by the total intervention period, was 96% (73%-99%). The mean (SD) baseline estimated glomerular filtration rate and 24-hour urinary sodium excretion were 38 (18) mL/min/1.73 m^2 and 145 (33) mmol, respectively. A higher proportion of patients in the intervention group reported that their salt intake behaviors had "significantly improved" or "somewhat improved" by the intervention than those in the control group (35/46, 76% vs 18/47, 38%; $P<.001$). However, the mean change in estimated 24-hour urinary sodium excretion during the intervention period did not differ significantly between groups (1.4, 95% CI −12.0 to 14.7 mmol in the intervention group vs 2.5, 95% CI −10.7 to 15.6 mmol in the control group; between-group difference −1.1, 95% CI −19.8 to 17.7 mmol; $P=.92$). Secondary outcomes were not significantly different between groups. These outcomes were not altered even in a subgroup of patients reporting improved self-reported salt intake behaviors.

Conclusions: The smartphone app did not reduce salt intake in patients with CKD, despite a substantial improvement in self-reported salt intake behaviors. Enhancing the intervention intensity may be necessary to effectively bridge the intention-behavior gap.

Trial Registration: Japan Registry of Clinical Trials jRCTs052220164; https://jrct.mhlw.go.jp/en-latest-detail/jRCTs052220164

(JMIR Mhealth Uhealth 2026;14:e68447) doi:10.2196/68447

KEYWORDS

smartphone app–based intervention; salt restriction; chronic kidney disease; hypertension; lifestyle modification

XSL•FO
RenderX

Introduction

Excessive salt intake is a major public health concern [1], contributing to a growing number of noncommunicable diseases, such as hypertension and cardiovascular diseases, as well as mortality [2-6]. Sodium load is also detrimental to the kidney, independent of blood pressure levels [7]. Cohort studies showed an association between higher urinary sodium excretion and an elevated risk of kidney outcomes even after adjustment for blood pressure levels [8,9]. In a meta-analysis of randomized controlled trials (RCTs), short-term low-salt diets reduced albuminuria by 36% with only a modest decrease in blood pressure levels among patients with early stages of chronic kidney disease (CKD) [10]. Salt restriction may also enhance proteinuria-lowering effects of renin-angiotensin system inhibitors and improve kidney outcomes [11,12]. Thus, limiting salt intake is an essential part of dietary therapy for patients with CKD, potentially offering kidney protection beyond blood pressure control. Accordingly, the Kidney Disease Improving Global Outcomes guideline recommends a daily sodium intake of less than 2 g for patients with hypertensive CKD [13].

In real-world clinical settings, however, salt restriction is often challenging. The average urinary sodium excretion in patients with CKD has been reported to be well above the guideline recommendation (124.6-160.9 mmol/24 h [2.9-3.7 g/24 h]) [8,14-17]. Lifestyle modifications through dietary consultation and motivational interviewing guided by health care professionals have not shown consistent long-term efficacy for salt restriction [18-22]. Notably, such time-consuming individualized therapy would be costly and difficult to scale for the growing CKD population.

Smartphone app–based interventions have emerged as promising therapeutic options for hypertension. These apps aim to facilitate behavior changes and have the potential to be delivered at scale. However, RCTs examining the efficacy of these apps are scarce [23]. A recent meta-analysis showed a moderate reduction in 24-hour urinary sodium excretion (−0.39 g, 95% CI −0.50 to −0.27) by technology-supported behavior change interventions, including smartphone apps, although patients with CKD were excluded [24]. CureApp HT is a hypertension management app that provides individualized health care advice for lifestyle modification. In the HERB Digital Hypertension 1 pivotal phase III trial, a 12-week use of CureApp HT reduced blood pressure and self-reported salt intake among hypertensive patients who had not used antihypertensive medications [25]. Since the majority of participants in this trial had normal kidney function, it remains uncertain whether CureApp HT would confer similar benefits to patients with CKD who are managed by nephrologists. Here, we evaluated the efficacy of CureApp HT in patients with CKD who were treated at a tertiary referral hospital and exhibited elevated urinary sodium excretion.

Methods

Study Design

This was an open-label, single-center, RCT examining efficacy of CureApp HT, a smartphone app for hypertension management, on urinary sodium excretion in patients with CKD. The intervention period lasted 12 weeks, followed by a 12-week post-intervention period to assess sustainability of treatment effects. A detailed schedule is shown in Figure S1 in Multimedia Appendix 1. Patients were enrolled from the nephrology outpatient department at Osaka University Hospital, Japan, between February 2023 and November 2023. This study was conducted and reported in accordance with the CONSORT eHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (Checklist 1).

Ethical Considerations

The study protocol was approved by the University of Osaka Clinical Research Review Board (approval number 22015) and registered at Japan Registry of Clinical Trials (jRCTs052220164) on February 6, 2023. All participants provided written informed consent. Participants' data were deidentified with unique identification codes and maintained in password-protected, access-restricted institutional databases. No financial or other compensation was provided to participants. The trial was conducted in accordance with the Declaration of Helsinki.

Eligibility Criteria

The eligibility criteria were (1) aged 18 years or older, (2) CKD stages G1 to G5, (3) a history of hypertension (systolic blood pressure 140 mmHg or higher and/or diastolic blood pressure 90 mmHg or higher and/or use of antihypertensive medications), and (4) an estimated 24-hour urinary sodium excretion based on the Tanaka method of 100 mmol or greater [26].

Patients were excluded if (1) they were unable to handle a smartphone due to cognitive impairment or impaired finger dexterity or (2) they already used other health care smartphone apps.

Randomization

Eligible patients were randomly assigned to the intervention group or control group in a 1:1 ratio. Randomization was performed using a computer-generated random number list with a permuted block size of 4, stratified by age (≥75 or <75 y) and estimated 24-hour urinary sodium excretion (≥170 or <170 mmol). The randomization procedure was conducted by one of the investigators who was unaware of participants' information and had no involvement in patient enrollment.

Interventions

Patients assigned to the intervention group used CureApp HT (CureApp Inc.), a smartphone app designed to facilitate hypertension management through lifestyle modifications and self-monitoring, for 12 weeks. This app was approved by the Pharmaceuticals and Medical Devices Agency as software as a medical device in Japan and was included in the national health insurance coverage on September 1, 2022. They also received lifestyle counseling by nephrologists as per the Clinical Practice Guideline for CKD 2018 issued by the Japanese Society of Nephrology [27]. This lifestyle counseling included general recommendations on salt restriction (targeting <6 g/d), moderation of alcohol consumption, and other standard dietary and behavioral advice. Patients in the control group received

XSL•FO
RenderX

lifestyle counseling during outpatient visits, typically once every 1 to 3 months, without access to the app.

To minimize the risk of cross-contamination, all patient recruitment and follow-up were conducted individually through routine outpatient visits. No group education sessions or group-based activities were provided. Patients in both the intervention and control groups were not informed of each other's group assignments. Furthermore, the app was accessed privately by patients in the intervention group on their personal smartphones.

Dose adjustment of antihypertensive medications, diuretics, oral antihyperglycemic medications, and potassium binders was not permitted throughout the intervention period, unless deemed clinically necessary by attending physicians.

Patients in the intervention group were instructed to download CureApp HT to their smartphones and activate it using prescription codes. They entered personal data, including age, sex, body weight, lifestyle behaviors, social background, and daily home blood pressure. These data were transferred to the cloud system, which generated individualized lifestyle modification programs.

CureApp HT is designed to enhance patients' awareness of hypertension and facilitate lifestyle modifications and self-management through the following 3 steps:

- Step 1 (Education): Patients acquire knowledge about hypertension and lifestyle modification methods through interactive conversations with a virtual nurse within the app.
- Step 2 (Implementation): The app provides patients with personalized action plans, including salt restriction, weight loss, exercise, improvement of sleep quality, mitigation of mental stress, and reduction in alcohol consumption.
- Step 3 (Habituation): Patients set their own behavioral goals and cultivate habits with the support of the app.

Through a specialized web application, attending physicians monitored patients' daily blood pressure levels, as entered by patients in the app, and tracked the progress of the 3 steps mentioned above.

Primary Outcome

The primary outcome was the change in estimated 24-hour urinary sodium excretion during the intervention period, measured at baseline, week 4, and week 12. The estimated 24-hour urinary sodium excretion was calculated using the Tanaka method based on spot urine samples [26]: Estimated 24-hour urinary sodium excretion (mg) = 23 × 21.98 × {[urinary Na (mEq/L)/(urinary creatinine (mg/dL)×10)] × predicted 24-hour urinary creatinine excretion (mg/day)}$^{0.392}$

Predicted 24-hour urinary creatinine excretion (mg/day) = [−2.04 × age (years)] + [14.89 × weight (kg)] + [16.14 × height (cm)] − 2244.45

Secondary Outcomes

Secondary outcomes were selected to explore the potential impact of the intervention on kidney and cardiovascular risk markers, including the changes in the following parameters during the intervention period, measured at baseline and week 12: (1) office blood pressure, (2) brachial-ankle pulse wave velocity (baPWV), (3) urinary protein-to-creatinine ratio (UPCR), (4) urinary potassium-to-creatinine ratio, (5) urinary sodium-to-potassium ratio, (6) estimated glomerular filtration rate (eGFR), (7) brain natriuretic peptide (BNP), (8) body weight, and (9) the number of antihypertensive medications.

Office blood pressure was measured using an automated oscillometric blood pressure monitor (Omron HEM-907). Patients self-measured their blood pressure after resting quietly in a seated position for 5 minutes without attendance of medical personnel. The average value of 2 measurements at 1-minute intervals was used for statistical analysis.

The baPWV was measured using a noninvasive automated device (Omron BP-203RPEIII) after patients rested for 5 minutes in a supine position. This device measured baPWV twice simultaneously on the right and left sides. The average value of these 4 measurements was used for statistical analysis.

eGFR was calculated using the Japanese standard formula: 194 × creatinine$^{-1.094}$ × age$^{-0.287}$ (if female, × 0.739) [28].

Questionnaire Survey on Salt Intake Behaviors

After the 12-week postintervention period, patients were asked to assess how much their lifestyle behaviors had improved by responding to the following question: "To what extent has your lifestyle related to (1) salt intake, (2) exercise, (3) sleep quality, (4) mental stress, and (5) alcohol consumption improved as a result of participating in this trial?." Patients selected one of the following options: "significantly improved," "somewhat improved," "not much improved," or "not improved at all." Regarding research objectives, only the salt-intake item was included in a between-group comparison in the present analysis.

Statistical Analysis

A previous observational study reported that every 40 mmol/24 h increase in urinary sodium excretion was associated with a 10% higher hazard of cardiovascular events in patients with CKD [9]. Based on this finding, we considered a 40 mmol/24 h decrease in urinary sodium excretion to be clinically relevant. According to the 2-arm pilot trial sample size calculation [29], 96 patients (48 in each group) were required to provide 90% power, with a 2-sided type I error level of .05, to detect a between-group difference of 40 mmol/24 h with a SD of 60 mmol/24 h. To account for a 5% dropout rate, we planned to recruit 100 patients.

Between-group comparisons were conducted using the Student *t* test or Mann-Whitney test for continuous variables and the chi-square test or Fisher exact test for categorical variables, as appropriate.

The primary analysis was conducted in the intention-to-treat population, defined as all randomized patients who used CureApp HT for at least one day or received at least one lifestyle counseling session. We employed a mixed-effects model for repeated measures (MMRM) with an unstructured covariance matrix to compare between-group differences in the primary and secondary outcomes. Fixed effects in the MMRM were treatment groups, categoric time, treatment-by-time interaction,

XSL•FO
RenderX

and stratification factors (age and baseline estimated 24-h urinary sodium excretion). No multiplicity adjustments for each outcome were made in this proof-of-concept study. There were no missing data on baseline estimated 24-hour urinary sodium excretion. The missing data after baseline were not imputed but were instead handled directly within the MMRM.

Prespecified subgroup analyses were performed based on baseline age (below or above the median), baseline estimated 24-hour urinary sodium excretion (below or above the median), and app engagement rates calculated by dividing the number of days patients recorded blood pressure in the app by the total intervention period (≥71.4% or <71.4%). Effect modifications were tested by incorporating treatment-by-subgroup interactions into the MMRM.

Sensitivity analyses were performed after excluding patients (1) who did not reach step 3 and (2) whose family members assisted with app usage due to patients' difficulty in operating a smartphone.

An exploratory analysis was conducted in which patients were categorized into 4 groups using the results of the questionnaire survey on salt intake behaviors ("not improved at all," "not much improved," "somewhat improved," and "significantly improved"). A linear regression model was used to evaluate the association between the self-reported behavior category and the change in estimated 24-hour urinary sodium excretion.

All statistical analyses were performed using the Stata/IC software (version 18.0; Stata Corp). A *P* value of <.05 was considered statistically significant.

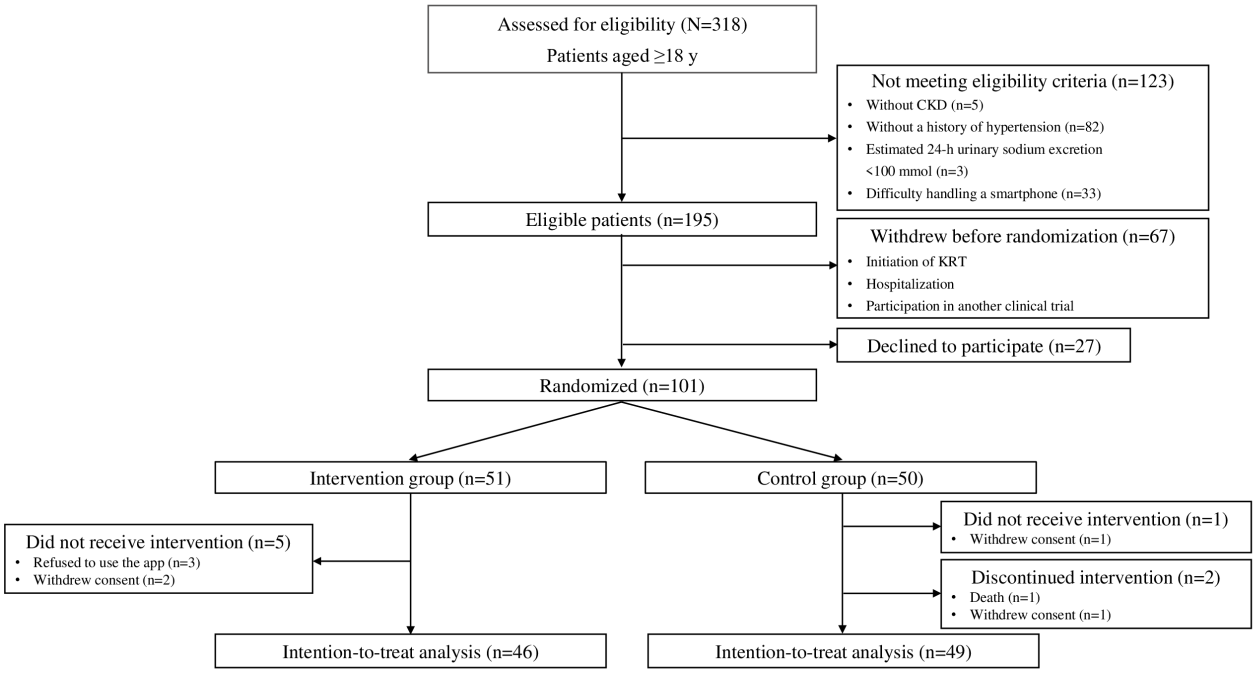
Results

Patient Characteristics

A total of 101 patients were randomly assigned to the intervention group (n=51) or the control group (n=50) (Figure 1). In the intervention group, 3 patients refused to use the app and 2 patients withdrew consent before using the app, while no patients dropped out during the intervention period. In the control group, 1 patient withdrew consent before the intervention period, and 2 patients dropped out during the intervention period.

The baseline characteristics are shown in Table 1. The mean (SD) age, blood pressure, and eGFR were 66 (14) years, 131 (19)/79 (14) mmHg, and 38 (18) mL/min/1.73 m^2, respectively. Patients in the intervention group showed slightly lower systolic blood pressure and eGFR than those in the control group. The estimated 24-hour urinary sodium excretion at baseline was 145 (33) mmol and was well-balanced between groups.

Figure 1. Flow diagram of the study. CKD: chronic kidney disease; KRT: kidney replacement therapy.



XSL•FO
RenderX

Table . Baseline characteristics of the study patients.

Characteristics	Total (N=101)	Intervention group (n=51)	Control group (n=50)	*P* value[a]
Age, years, mean (SD)	66 (14)	67 (14)	66 (14)	.56
Male, n (%)	63 (62)	29 (57)	34 (68)	.25
BMI, kg/m^2, mean (SD)	23.9 (3.6)	23.9 (3.7)	23.9 (3.5)	.98
Systolic BP[b], mmHg, mean (SD)	131 (19)	128 (20)	134 (16)	.10
Diastolic BP, mmHg, mean (SD)	79 (14)	79 (15)	79 (14)	.76
Diabetes mellitus, n (%)	36 (36)	15 (29)	21 (42)	.19
Cardiovascular comorbidities, n (%)	20 (20)	9 (18)	11 (22)	.58
CKD[c] stages, n (%)				.14
G2	12 (12)	5 (10)	7 (14)	
G3	53 (52)	26 (51)	27 (54)	
G4	28 (28)	14 (27)	14 (28)	
G5	8 (8)	6 (12)	2 (4)	
Smoking history, n (%)				.33
Never smoked	52 (51)	30 (59)	22 (44)	
Former smoker	40 (40)	17 (33)	23 (46)	
Current smoker	9 (9)	4 (8)	5 (10)	
Hemoglobin, g/dL, mean (SD)	12.9 (1.8)	12.7 (1.7)	13.1 (1.9)	.22
Albumin, g/dL, mean (SD)	4 (0.4)	4 (0.3)	4.1 (0.4)	.43
Potassium, mEq/L, mean (SD)	4.5 (0.5)	4.5 (0.6)	4.5 (0.5)	.88
Creatinine, mg/dL, mean (SD)	1.7 (0.9)	1.8 (1.1)	1.5 (0.6)	.09
eGFR[d], mL/min/1.73 m^2, mean (SD)	38 (18)	35 (18)	40 (18)	.18
Calcium, mg/dL, mean (SD)	9.4 (0.4)	9.5 (0.5)	9.4 (0.4)	.63
Phosphate, mg/dL, mean (SD)	3.5 (0.7)	3.6 (0.6)	3.4 (0.7)	.20
CRP[e], mg/dL, median (IQR)	0.1 (0.0-0.2)	0.1 (0.0-0.2)	0.1 (0.0-0.1)	.62
UPCR[f], g/gCre, median (IQR)	0.28 (0.10-0.74)	0.28 (0.09-0.97)	0.27 (0.10-0.53)	.62
BNP[g], pg/mL, median (IQR)	28 (11-63)	27 (9-43)	34 (15-69)	.29
Estimated 24-h urinary sodium excretion, mmol, mean (SD)	145 (33)	143 (29)	147 (37)	.56

[a]*P* values indicate differences between intervention and control groups.

[b]BP: blood pressure.

[c]CKD: chronic kidney disease.

[d]eGFR: estimated glomerular filtration rate.

[e]CRP: C-reactive protein.

[f]UPCR: urinary protein-to-creatinine ratio.

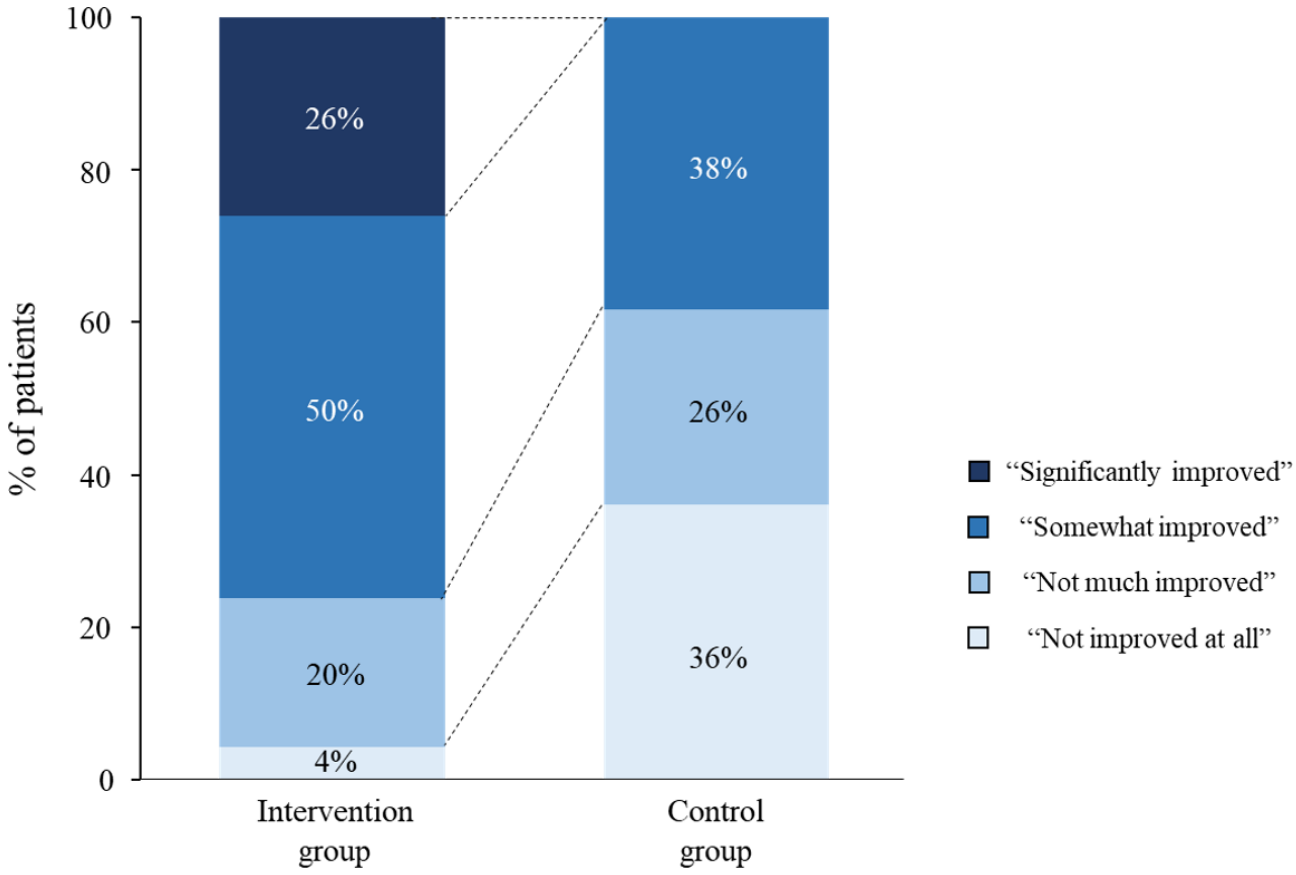
[g]BNP: B-type natriuretic peptide.

XSL•FO
RenderX

Questionnaire Survey on Salt Intake Behaviors

In the intervention group, 35 of 46 (76%) patients reported that their salt intake behavior had “significantly improved” or “somewhat improved” by the intervention, compared with 18 of 47 (38%) in the control group (χ^2_3=24.87; *P*<.001; Figure 2, Table S1 in Multimedia Appendix 1).

Figure 2. Questionnaire survey on salt intake behaviors. Self-reported salt intake behaviors were significantly improved in the intervention group compared with the control group. Patients were asked, "To what extent has your lifestyle related to salt intake improved as a result of participating in this trial?" In the intervention group, 35 of 46 (76%) patients reported that their salt intake behavior had “significantly improved” or “somewhat improved” compared with 18 of 47 (38%) in the control group (*P*<.001; χ^2_3=24.87).



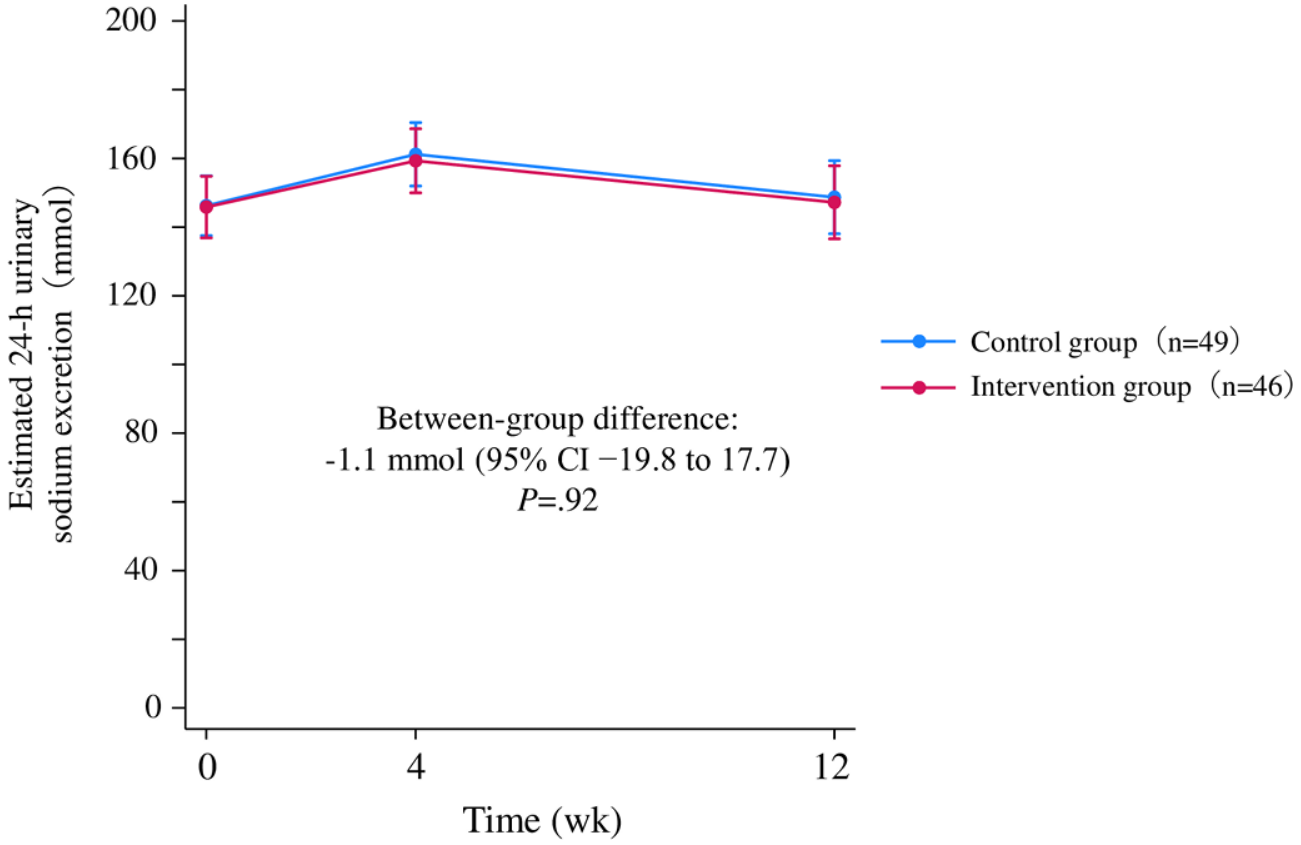
Primary Outcome

The mean change in estimated 24-hour urinary sodium excretion from baseline to week 12 was 1.4 (95% CI −12.0 to 14.7) mmol in the intervention group and 2.5 (95% CI −10.7 to 15.6) mmol in the control group, with no significant between-group difference (−1.1 mmol, 95% CI −19.8 to 17.7; *P*=.92) (Figure 3, Table 2). Age, baseline estimated 24-hour urinary sodium excretion, and app engagement rates did not significantly modify the results (Table S2 in Multimedia Appendix 1).

After the 12-week postintervention period, the mean change in the estimated 24-hour urinary sodium excretion from baseline was not significantly different between the groups; 4.7 (95% CI −8.9 to 18.4) and 13.9 (95% CI 0.4-27.3) mmol in the intervention and control groups, respectively (between-group difference, −9.1 mmol, 95% CI −28.3 to 10.0; *P*=.35).

XSL•FO
RenderX

Figure 3. Estimated 24-hour urinary sodium excretion during the intervention period. Data are shown for the intention-to-treat population (46 patients in the intervention group and 49 patients in the control group). A mixed-effects model for repeated measures with an unstructured covariance matrix was used to compare between-group changes. The model included age, baseline estimated 24-hour urinary sodium excretion, randomized groups, time, and interaction between randomized groups and time as fixed effects, and subjects as a random effect. Error bars represent 95% CIs.



XSL•FO
RenderX

Table . Summary of the primary and secondary outcomes.[a]

Outcome	Intervention group (n=46)			Control group (n=49)			Difference-in-changes[b] (95% CI)	*P* value[c]
	Baseline, mean (95% CI)	Week 12, mean (95% CI)	Within-group change[d], mean (95% CI)	Baseline, mean (95% CI)	Week 12, mean (95% CI)	Within-group change[d], mean (95% CI)		
Primary outcome								
Estimated 24-h urinary sodium excretion, mmol	145.8 (136.8 to 154.8)	147.2 (136.6 to 157.9)	1.4 (−12 to 14.7)	146.2 (137.5 to 154.9)	148.7 (138.1 to 159.3)	2.5 (−10.7 to 15.6)	−1.1 (−19.8 to 17.7)	.92
Secondary outcomes								
Office SBP[e], mmHg	129 (125 to 133)	130 (126 to 134)	1 (−4 to 6)	131 (127 to 135)	129 (126 to 133)	−2 (−7 to 3)	3 (−4 to 10)	.43
Office DBP[f], mmHg	79 (77 to 81)	78 (76 to 80)	−1 (-4 to 2)	79 (77 to 81)	77 (75 to 80)	−2 (−5 to 1)	1 (−4 to 5)	.72
Mean baPWV[g], cm/s	1687 (1648 to 1726)	1708 (1668 to 1748)	21 (−35 to 76)	1697 (1659 to 1735)	1726 (1687 to 1765)	30 (−25 to 84)	−9 (−87 to 69)	.82
UPCR[h], g/gCre	0.75 (0.63 to 0.87)	0.79 (0.66 to 0.91)	0.04 (−0.14 to 0.21)	0.74 (0.62 to 0.86)	0.91 (0.78 to 1.03)	0.16 (−0.01 to 0.33)	−0.13 (−0.37 to 0.12)	.31
uK/uCre[i], mmol/gCre	0.46 (0.41 to 0.52)	0.41 (0.35 to 0.46)	−0.06 (−0.14 to 0.02)	0.44 (0.38 to 0.49)	0.46 (0.40 to 0.51)	0.02 (−0.06 to 0.10)	−0.08 (−0.19 to 0.03)	.18
uNa/uK[j]	3.22 (2.87 to 3.57)	3.30 (2.95 to 3.65)	0.08 (−0.42 to 0.57)	3.27 (2.93 to 3.61)	3.11 (2.75 to 3.46)	−0.16 (−0.65 to 0.33)	0.24 (−0.46 to 0.94)	.50
eGFR[k], mL/min/1.73 m^2	38 (37 to 39)	38 (37 to 39)	0 (−1.4 to 1.5)	38 (37.1 to 39)	39.3 (38.3 to 40.4)	1.3 (−0.1 to 2.7)	−1.3 (-3.3 to 0.8)	.22
BNP[l], pg/mL	54.2 (42 to 66.4)	32.1 (19.9 to 44.3)	−22.1 (−39.3 to −4.8)	56.8 (45 to 68.6)	49.8 (37.6 to 62)	−7 (−24 to 10)	−15.1 (−39.3 to 9.1)	.22
Body weight, kg	64.1 (63.8 to 64.5)	64.4 (64.1 to 64.8)	0.3 (−0.2 to 0.8)	64.1 (63.8 to 64.5)	64.3 (63.9 to 64.6)	0.2 (−0.3 to 0.7)	0.2 (−0.6 to 0.9)	.66
The number of antihypertensive medications	1.22 (1.19 to 1.26)	1.25 (1.21 to 1.29)	0.03 (−0.02 to 0.08)	1.22 (1.19 to 1.26)	1.22 (1.19 to 1.26)	0 (−0.05 to 0.05)	0.03 (−0.05 to 0.10)	.46

[a]Results were obtained from the mixed-effects models for repeated measures (MMRM) with an unstructured covariance matrix. Fixed effects in the MMRM were treatment group, categoric time, the treatment-by-time interaction, age, and baseline values of each outcome variable. Data presented as predicted mean (95% CI).

[b]Between-group differences in the changes from baseline to week 12. Positive values indicate that the intervention group exhibited greater increases than the control group.

[c]*P* values for difference-in-changes.

[d]Positive values indicate an increase from baseline to week 12.

[e]SBP: systolic blood pressure.

[f]DBP: diastolic blood pressure.

[g]baPWV: brachial-ankle pulse wave velocity.

[h]UPCR: urinary protein-to-creatinine ratio.

[i]uK/uCre: urinary potassium-to-creatinine ratio.

[j]uNa/uK: urinary sodium-to-potassium ratio.

[k]eGFR: estimated glomerular filtration rate.

[l]BNP: B-type natriuretic peptide.

XSL•FO
RenderX

Secondary Outcomes

There were no significant between-group differences in the changes in blood pressure, baPWV, UPCR, urinary potassium-to-creatinine ratio, urinary sodium-to-potassium ratio, eGFR, BNP, body weight, and the number of antihypertensive medications during the intervention period (Table 2, Figure S2 in Multimedia Appendix 1). No adverse events related to app use were reported.

The median (IQR) app engagement rate was 96% (73%-99%). At the end of the intervention period, 35 of 46 (76%) patients who used the app at least once completed Step 1, and 22 of 46 (48%) completed both Step 1 and Step 2.

Sensitivity Analyses

The results were not substantially altered after excluding patients who did not reach Step 3 (24/46, 52%) or whose family members helped use the app (1/46, 2%) (Table S3 in Multimedia Appendix 1).

Exploratory Analysis

Participants were categorized into 4 groups according to the questionnaire survey on salt intake behaviors at week 12. There were no significant linear trends between self-reported salt intake behavior and the change in urinary sodium excretion in either the intervention or control group (Table S4 in Multimedia Appendix 1).

Discussion

Overview

In this single-center, open-label, RCT involving patients with CKD, we observed a substantial improvement in self-reported salt intake behaviors after 12 weeks of CureApp HT use. However, the app did not improve the primary outcome, estimated 24-hour urinary sodium excretion, a surrogate measure of actual salt intake. Secondary outcomes, including blood pressure, baPWV, UPCR, and BNP, also did not differ significantly between groups. Notably, these outcomes did not improve even in a subgroup of patients who reported significant improvements in salt intake behaviors, suggesting a gap between self-awareness and actual behaviors regarding dietary salt intake. Bridging this gap will be a major challenge that digital therapeutics need to address.

Principal Findings

Salt restriction is essential in the management of CKD to achieve optimal blood pressure control. Associations between urinary sodium excretion and kidney outcomes have been reported to be independent of blood pressure levels [8,9]. This suggests that mechanisms beyond blood pressure elevation may contribute to salt-induced kidney damage. Basic studies have shown that a high-salt diet accelerates kidney damage through several mechanisms, including Rac1-mediated mineralocorticoid receptor activation [7], inhibition of tissue remodeling by macula densa cells [30], and inflammation and fibrosis induced by Y-box binding protein-1 [31]. These preclinical findings imply that salt restriction may confer kidney protection even when blood pressure levels are normal. Although this concept has not been clinically tested, short-term low-salt diets significantly reduced albuminuria with only modest impacts on blood pressure levels [10,11]. Additionally, salt restriction may enhance the kidney-protective effects of renin–angiotensin system inhibitor [12]. Therefore, although our patients had well-controlled blood pressure levels, salt restriction may still be beneficial in reducing residual salt toxicity to the kidney. Thus, urinary sodium excretion is a clinically relevant outcome in this trial.

We identified a substantial discrepancy between patients' self-reported salt restriction behaviors and objectively measured urinary sodium excretion. A similar finding was reported in the HERB Digital Hypertension 1 trial, in which CureApp HT significantly improved self-reported salt intake but had minimal impact on the urinary sodium-to-creatinine ratio [25]. These findings suggest that intentions do not necessarily translate into behavioral changes, highlighting the presence of an intention-behavior gap. Psychological behavior theories, such as the Theory of Planned Behavior and Self-Determination Theory, assume that intention is the proximal determinant of behavior. However, a "gap" exists between intention and behavior [32]. There is only a modest correlation between intentions and actual behaviors [33,34]; intention accounts for only 27% of the variance in dietary habits [35], and individuals often underestimate their salt intake [36,37]. This gap is influenced by several moderators, including (1) social environment, (2) behavioral inertia, and (3) sodium palatability. The social environment plays a significant role in determining salt intake behaviors. Even when patients recognize the importance of salt restriction, it is not feasible in situations where access to fresh, low-sodium foods is limited and processed foods (including prepackaged and fast food) are prevalent. In Japan, many traditional foods, such as soy sauce and pickled vegetables, are high in salt. These social and environmental factors may impede patients' intention to adhere to a low-salt diet despite their increased awareness of salt restriction. Public health policies and social support must be implemented, in collaboration with multiple stakeholders such as the food industry, to create an environment where low-salt foods are more readily accessible [38].

Behavioral inertia is another barrier that exacerbates the intention-behavior gap. Established eating habits are a putative determinant of human salt intake [39]. A prior study found no association between the intention to change salt intake habits and the actual purchasing behavior of low-salt foods [40]. Because our patients were older, they were likely to maintain long-established dietary and cooking habits, which are inherently resistant to change, although age did not significantly modify the results in subgroup analysis. Regular face-to-face care from nephrologists may have led patients to assume that their current salt intake was already acceptable, even though their salt intake was beyond the target range. This false perception may reinforce patients' psychological resistance to behavioral modifications for salt restriction proposed by the app. Importantly, health benefits from salt restriction are not always immediately apparent. The delayed rewards of long-term health benefits obtained from salt restriction, in exchange for immediate

XSL•FO
RenderX

gratification of enjoying salty foods, further complicate maintaining salt restriction as a regular habit [40].

Finally, sodium palatability may also contribute to the intention-behavior gap among patients with CKD. Under sodium-sated conditions, humans exhibit an attraction toward low concentrations of sodium but an aversion toward high concentrations of sodium [41]. However, patients with CKD typically have a higher threshold for salty taste perception than healthy individuals [42,43], and their aversive response to high sodium concentrations may be diminished [44]. Such an altered salty taste perception, coupled with an increased tolerance for excess salt, likely enhances sodium palatability of these patients, making dietary salt restriction even more challenging.

To bridge the intention-behavior gap regarding salt restriction, it may be necessary to optimize the intensity of intervention. Greaves et al [45] reported that intervention intensity, such as the number of sessions and total contact time, was related to self-reported dietary changes. For example, in the Trials of Hypertension Prevention, which demonstrated substantial efficacy of comprehensive education and counseling on salt restriction, intensive and meticulous interventions were conducted (8 group counseling sessions and 2 individual sessions over the first 3 mo, followed by periodic group meetings and individual telephone or in-person contacts throughout the follow-up) [46]. In patients with CKD, theory-based behavioral interventions, using both face-to-face and web-based approach (e-coaching and group meetings), have shown only modest and temporary effects on salt restriction, potentially due to a relatively low intensity of these interventions [20,21]. The intensity of the smartphone-based digital intervention might be even weaker, particularly for patients with CKD who likely have a large intention-behavior gap, owing to the absence of human involvement. Therefore, enhancing intervention intensity in smartphone-based approaches, potentially by incorporating more interactive and individualized elements, may be essential to improve their effectiveness in promoting sustained salt restriction.

Limitations

This trial has several limitations, one of which is the open-label design. However, blinding was impractical given the nature of the intervention. Since the measurement of urinary sodium excretion was objective, potential bias from the open-label design was likely mitigated. Second, the sample size was modest and the trial duration was relatively short. However, given the neutral result of the primary outcome, it is unlikely that a larger and longer-term trial would have produced a positive result. Third, although the accuracy of the Tanaka method for estimating 24-hour urinary sodium excretion from a spot urine test has been validated in Japanese patients with CKD [47], its accuracy may still be inferior to that of a 24-hour urine collection. However, if dietary salt intake had been meaningfully reduced, secondary outcomes, including blood pressure, UPCR, and BNP, would likely have shown improvement. Fourth, the questionnaire survey on salt intake behaviors may have been subject to social-desirability bias, potentially leading to an overestimation of the proportion of patients reporting improvements. However, the substantial difference between the intervention and control groups suggests that the app did have some impact at least on patients' intention. Fifth, the study did not assess the potential time lag between knowledge acquisition and actual behavioral change, which may have contributed to the limited correspondence between self-reported behavior and objective outcomes. Finally, since our study included patients who were willing to use the app, the generalizability of our findings to the broader real-world population, particularly outside of Japan, is uncertain.

Conclusions

CureApp HT, a hypertension management app that provides individualized advice and self-monitoring for lifestyle modifications and salt restriction, was not effective in reducing urinary sodium excretion among patients with CKD, despite substantial improvement in self-reported salt intake behaviors. Secondary outcomes, including blood pressure, baPWV, UPCR, and BNP, also did not differ significantly between groups, further indicating that the app did not improve other cardiovascular and kidney risk markers. These results should be interpreted with caution due to the open-label design and the relatively short duration of the trial. Nonetheless, these findings suggest the presence of an intention-behavior gap for salt restriction, which might be addressed by enhancing the intensity of smartphone app–based interventions.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Data Availability

The Ethics Committee of Osaka University Hospital restricts public sharing of individual participant data because of privacy and ethical issues. The dataset and STATA program code used in this study are available on reasonable request to the corresponding author.

Authors' Contributions

Conceptualization: TK, YS, IM, JYK, YI
Data curation: TK, YS
Formal analysis: TK, YS
Investigation: TK, YS

XSL•FO
RenderX

Resources: YS, YD, IM, MM, JYK
Supervision: YS, IM, JYK, YI
Visualization: TK, YS
Writing – original draft: TK
Writing – review & editing: TK, YS, TO, YD, RY, IM, MM, JYK, YI

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire results on salt intake, subgroup and sensitivity analyses, and study design.
[DOCX File, 159 KB - mhealth_v14i1e68447_app1.docx]

Checklist 1
CONSORT eHEALTH checklist (version 1.6.1).
[PDF File, 1130 KB - mhealth_v14i1e68447_app2.pdf]

References

1. Mozaffarian D, Fahimi S, Singh GM, et al. Global sodium consumption and death from cardiovascular causes. N Engl J Med 2014 Aug 14;371(7):624-634. [doi: 10.1056/NEJMoa1304127] [Medline: 25119608]
2. Ellison DH, Welling P. Insights into salt handling and blood pressure. N Engl J Med 2021 Nov 18;385(21):1981-1993. [doi: 10.1056/NEJMra2030212] [Medline: 34788509]
3. Farquhar WB, Edwards DG, Jurkovitz CT, Weintraub WS. Dietary sodium and health: more than just blood pressure. J Am Coll Cardiol 2015 Mar 17;65(10):1042-1050. [doi: 10.1016/j.jacc.2014.12.039] [Medline: 25766952]
4. Mente A, O'Donnell M, Rangarajan S, et al. Urinary sodium excretion, blood pressure, cardiovascular disease, and mortality: a community-level prospective epidemiological cohort study. Lancet 2018 Aug 11;392(10146):496-506. [doi: 10.1016/S0140-6736(18)31376-X] [Medline: 30129465]
5. Strazzullo P, D'Elia L, Kandala NB, Cappuccio FP. Salt intake, stroke, and cardiovascular disease: meta-analysis of prospective studies. BMJ 2009 Nov 24;339:b4567. [doi: 10.1136/bmj.b4567] [Medline: 19934192]
6. Afshin A, Sur PJ, Fay KA, et al. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet 2019 May;393(10184):1958-1972. [doi: 10.1016/S0140-6736(19)30041-8]
7. Kawarazaki W, Nagase M, Yoshida S, et al. Angiotensin II- and salt-induced kidney injury through Rac1-mediated mineralocorticoid receptor activation. J Am Soc Nephrol 2012 Jun;23(6):997-1007. [doi: 10.1681/ASN.2011070734] [Medline: 22440899]
8. He J, Mills KT, Appel LJ, et al. Urinary sodium and potassium excretion and CKD progression. J Am Soc Nephrol 2016 Apr;27(4):1202-1212. [doi: 10.1681/ASN.2015010022] [Medline: 26382905]
9. Mills KT, Chen J, Yang W, et al. Sodium excretion and the risk of cardiovascular disease in patients with chronic kidney disease. JAMA 2016;315(20):2200-2210. [doi: 10.1001/jama.2016.4447] [Medline: 27218629]
10. McMahon EJ, Campbell KL, Bauer JD, Mudge DW, Kelly JT. Altered dietary salt intake for people with chronic kidney disease. Cochrane Database Syst Rev 2021 Jun 24;6(6):CD010070. [doi: 10.1002/14651858.CD010070.pub3] [Medline: 34164803]
11. Kwakernaak AJ, Krikken JA, Binnenmars SH, et al. Effects of sodium restriction and hydrochlorothiazide on RAAS blockade efficacy in diabetic nephropathy: a randomised clinical trial. Lancet Diabetes Endocrinol 2014 May;2(5):385-395. [doi: 10.1016/S2213-8587(14)70030-0] [Medline: 24795252]
12. Lambers Heerspink HJ, Holtkamp FA, Parving HH, et al. Moderation of dietary sodium potentiates the renal and cardiovascular protective effects of angiotensin receptor blockers. Kidney Int 2012 Aug;82(3):330-337. [doi: 10.1038/ki.2012.74] [Medline: 22437412]
13. Cheung AK, Chang TI, Cushman WC, et al. Executive summary of the KDIGO 2021 clinical practice guideline for the management of blood pressure in chronic kidney disease. Kidney Int 2021 Mar;99(3):559-569. [doi: 10.1016/j.kint.2020.10.026] [Medline: 33637203]
14. Fan L, Tighiouart H, Levey AS, Beck GJ, Sarnak MJ. Urinary sodium excretion and kidney failure in nondiabetic chronic kidney disease. Kidney Int 2014 Sep;86(3):582-588. [doi: 10.1038/ki.2014.59] [Medline: 24646858]
15. Kang M, Kang E, Ryu H, et al. Measured sodium excretion is associated with CKD progression: results from the KNOW-CKD study. Nephrol Dial Transplant 2021 Feb 20;36(3):512-519. [doi: 10.1093/ndt/gfaa107] [Medline: 32582942]
16. McQuarrie EP, Traynor JP, Taylor AH, et al. Association between urinary sodium, creatinine, albumin, and long-term survival in chronic kidney disease. Hypertension 2014 Jul;64(1):111-117. [doi: 10.1161/HYPERTENSIONAHA.113.03093] [Medline: 24732890]

XSL•FO
RenderX

17. Ogata S, Akashi Y, Sakusabe T, et al. A multiple 24-hour urine collection study indicates that kidney function decline is related to urinary sodium and potassium excretion in patients with chronic kidney disease. Kidney Int 2022 Jan;101(1):164-173. [doi: 10.1016/j.kint.2021.10.030] [Medline: 34774553]
18. van Zuilen AD, Bots ML, Dulger A, et al. Multifactorial intervention with nurse practitioners does not change cardiovascular outcomes in patients with chronic kidney disease. Kidney Int 2012 Sep;82(6):710-717. [doi: 10.1038/ki.2012.137] [Medline: 22739979]
19. Hwang JH, Chin HJ, Kim S, et al. Effects of intensive low-salt diet education on albuminuria among nondiabetic patients with hypertension treated with olmesartan: a single-blinded randomized, controlled trial. Clin J Am Soc Nephrol 2014 Dec 5;9(12):2059-2069. [doi: 10.2215/CJN.01310214] [Medline: 25332317]
20. Meuleman Y, Hoekstra T, Dekker FW, et al. Sodium restriction in patients with CKD: a randomized controlled trial of self-management support. Am J Kidney Dis 2017 May;69(5):576-586. [doi: 10.1053/j.ajkd.2016.08.042] [Medline: 27993433]
21. Humalda JK, Klaassen G, de Vries H, et al. A self-management approach for dietary sodium restriction in patients with CKD: a randomized controlled trial. Am J Kidney Dis 2020 Jun;75(6):847-856. [doi: 10.1053/j.ajkd.2019.10.012] [Medline: 31955921]
22. O'Callaghan CA, Camidge C, Thomas R, et al. Evaluation of a simple low-cost intervention to empower people with CKD to reduce their dietary salt intake: OxCKD1, a multicenter randomized controlled trial. Kidney360 2023 Jul 1;4(7):890-898. [doi: 10.34067/KID.0000000000000160] [Medline: 37254243]
23. Ali SH, Luo R, Li Y, Liu X, Tang C, Zhang P. Application of mobile health technologies aimed at salt reduction: systematic review. JMIR Mhealth Uhealth 2019 Apr 17;7(4):e13250. [doi: 10.2196/13250] [Medline: 30994467]
24. Yan YY, Chan LML, Wang MP, Kwok JYY, Anderson CS, Lee JJ. Technology-supported behavior change interventions for reducing sodium intake in adults: a systematic review and meta-analysis. NPJ Digit Med 2024 Mar 18;7(1):72. [doi: 10.1038/s41746-024-01067-y] [Medline: 38499729]
25. Kario K, Nomura A, Harada N, et al. Efficacy of a digital therapeutics system in the management of essential hypertension: the HERB-DH1 pivotal trial. Eur Heart J 2021 Oct 21;42(40):4111-4122. [doi: 10.1093/eurheartj/ehab559] [Medline: 34455443]
26. Tanaka T, Okamura T, Miura K, et al. A simple method to estimate populational 24-h urinary sodium and potassium excretion using a casual urine specimen. J Hum Hypertens 2002 Feb;16(2):97-103. [doi: 10.1038/sj.jhh.1001307] [Medline: 11850766]
27. Evidence-based clinical practice guideline for CKD 2018. : Japanese Society of Nephrology; 2021 URL: https://cdn.jsn.or.jp/data/CKD2018.pdf [accessed 2026-02-18]
28. Matsuo S, Imai E, Horio M, et al. Revised equations for estimated GFR from serum creatinine in Japan. Am J Kidney Dis 2009 Jun;53(6):982-992. [doi: 10.1053/j.ajkd.2008.12.034] [Medline: 19339088]
29. Whitehead AL, Julious SA, Cooper CL, Campbell MJ. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. Stat Methods Med Res 2016 Jun;25(3):1057-1073. [doi: 10.1177/0962280215588241] [Medline: 26092476]
30. Gyarmati G, Shroff UN, Riquier-Brison A, et al. Neuronally differentiated macula densa cells regulate tissue remodeling and regeneration in the kidney. J Clin Invest 2024 Apr 10;134(11):e174558. [doi: 10.1172/JCI174558] [Medline: 38598837]
31. Bernhardt A, Krause A, Reichardt C, et al. Excessive sodium chloride ingestion promotes inflammation and kidney fibrosis in aging mice. Am J Physiol Cell Physiol 2023 Aug 1;325(2):C456-C470. [doi: 10.1152/ajpcell.00230.2023] [Medline: 37399499]
32. Sheeran P, Webb TL. The intention-behavior gap. Soc Personal Psychol Compass 2016 Sep;10(9):503-518. [doi: 10.1111/spc3.12265]
33. Rhodes RE, Dickau L. Experimental evidence for the intention-behavior relationship in the physical activity domain: a meta-analysis. Health Psychol 2012 Nov;31(6):724-727. [doi: 10.1037/a0027290] [Medline: 22390739]
34. Feil K, Fritsch J, Rhodes RE. The intention-behaviour gap in physical activity: a systematic review and meta-analysis of the action control framework. Br J Sports Med 2023 Oct;57(19):1265-1271. [doi: 10.1136/bjsports-2022-106640] [Medline: 37460164]
35. McEachan RRC, Conner M, Taylor NJ, Lawton RJ. Prospective prediction of health-related behaviours with the theory of planned behaviour: a meta-analysis. Health Psychol Rev 2011 Sep;5(2):97-144. [doi: 10.1080/17437199.2010.521684]
36. Du X, Fang L, Xu J, et al. The association of knowledge, attitudes and behaviors related to salt with 24-h urinary sodium, potassium excretion and hypertensive status. Sci Rep 2022 Aug 16;12(1):13901. [doi: 10.1038/s41598-022-18087-x] [Medline: 35974077]
37. Takahashi N, Tanabe K, Adachi T, et al. Awareness of salt restriction is not reflected in the actual salt intake in Japanese hypertensive patients. Clin Exp Hypertens 2015;37(5):388-392. [doi: 10.3109/10641963.2014.987392] [Medline: 25496382]
38. Contreras Navarro A, Gallagher K, Griffin S, Leydon CL, Perry IJ, Harrington JM. Systematic review on the impact of salt-reduction initiatives by socioeconomic position to address health inequalities in adult populations. Nutr Rev 2025 Mar 1;83(3):e1090-e1100. [doi: 10.1093/nutrit/nuae088] [Medline: 38976594]
39. Leshem M. Biobehavior of the human love of salt. Neurosci Biobehav Rev 2009 Jan;33(1):1-17. [doi: 10.1016/j.neubiorev.2008.07.007] [Medline: 18708089]

XSL•FO
RenderX

40. Mørk T, Lähteenmäki L, Grunert KG. Determinants of intention to reduce salt intake and willingness to purchase salt-reduced food products: evidence from a web survey. Appetite 2019 Aug 1;139:110-118. [doi: 10.1016/j.appet.2019.04.018] [Medline: 31034859]
41. Zhang Y, Pool AH, Wang T, et al. Parallel neural pathways control sodium consumption and taste valence. Cell 2023 Dec 21;186(26):5751-5765. [doi: 10.1016/j.cell.2023.10.020] [Medline: 37989313]
42. McMahon EJ, Campbell KL, Bauer JD. Taste perception in kidney disease and relationship to dietary sodium intake. Appetite 2014 Dec;83:236-241. [doi: 10.1016/j.appet.2014.08.036] [Medline: 25192896]
43. Kusaba T, Mori Y, Masami O, et al. Sodium restriction improves the gustatory threshold for salty taste in patients with chronic kidney disease. Kidney Int 2009 Sep;76(6):638-643. [doi: 10.1038/ki.2009.214] [Medline: 19516246]
44. Okuno-Ozeki N, Kohama Y, Taguchi H, et al. Aversion to a high salt taste is disturbed in patients with CKD. Kidney Int Rep 2024 May;9(5):1254-1264. [doi: 10.1016/j.ekir.2024.02.1393] [Medline: 38707820]
45. Greaves CJ, Sheppard KE, Abraham C, et al. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011 Feb 18;11:119. [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
46. Kumanyika SK, Hebert PR, Cutler JA, et al. Feasibility and efficacy of sodium reduction in the trials of hypertension prevention, phase I. Trials of Hypertension Prevention Collaborative Research Group. Hypertension 1993 Oct;22(4):502-512. [doi: 10.1161/01.hyp.22.4.502] [Medline: 8406655]
47. Ogura M, Kimura A, Takane K, et al. Estimation of salt intake from spot urine samples in patients with chronic kidney disease. BMC Nephrol 2012 Jun 8;13:36. [doi: 10.1186/1471-2369-13-36] [Medline: 22682402]

Abbreviations

baPWV: brachial-ankle pulse wave velocity
BNP: brain natriuretic peptide
CKD: chronic kidney disease
CONSORT eHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
eGFR: estimated glomerular filtration rate
MMRM: mixed-effects model for repeated measures
RCT: randomized controlled trial
UPCR: urinary protein-to-creatinine ratio

Edited by L Buis; submitted 08.Nov.2024; peer-reviewed by N Koizumi, PW de Leeuw; accepted 05.Feb.2026; published 16.Mar.2026.

Please cite as:
Kawaoka T, Sakaguchi Y, Oka T, Doi Y, Yamamoto R, Matsui I, Mizui M, Kaimori JY, Isaka Y
Effects of a Hypertension Management Mobile App on Urinary Sodium Excretion in Patients With Chronic Kidney Disease: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e68447
URL: https://mhealth.jmir.org/2026/1/e68447
doi:10.2196/68447

© Takayuki Kawaoka, Yusuke Sakaguchi, Tatsufumi Oka, Yohei Doi, Ryohei Yamamoto, Isao Matsui, Masayuki Mizui, Jun-Ya Kaimori, Yoshitaka Isaka. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 16.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Mobile App–Based Individualized Nonpharmacological Intervention for Behavioral and Psychological Symptoms in Dementia: Pilot Randomized Controlled Trial

Eunhee Cho[1], PhD; Minhee Yang[1], PhD; Sinwoo Hwang[1,2], MSN; Eunkyo Kim[1,3], MSN; Jungwon Cho[1,3], MSN; Min Jung Kim[4], PhD

[1]Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University, 50-1, Yonsei-ro, Seodaemun-gu, Seoul, Republic of Korea

[2]Army Cadet Military School, Goesan, Republic of Korea

[3]College of Nursing and Brain Korea 21 FOUR Project, Yonsei University, Seoul, Republic of Korea

[4]College of Nursing, Ewha Womans University, Seoul, Republic of Korea

Corresponding Author:
Minhee Yang, PhD
Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University, 50-1, Yonsei-ro, Seodaemun-gu, Seoul, Republic of Korea

Abstract

Background: Behavioral and psychological symptoms of dementia (BPSD) are common and negatively impact both individuals with dementia and their families. Although nonpharmacological interventions are recommended as the first-line treatments, their use in community settings is limited by access and caregiver resources. Existing approaches are often not individualized and rely on institutional or clinician-led delivery.

Objective: We developed a caregiver-initiated and individualized multimodal mobile app. The app delivers tailored nonpharmacological interventions—such as music therapy, exercise, and reminiscence therapy—based on each user's preferences and functional abilities. This study aimed to evaluate the effectiveness of this mobile app–based intervention in reducing BPSD in community-dwelling persons living with dementia.

Methods: This study used a single-blinded randomized controlled trial design. Participants were recruited from an outpatient clinic of a tertiary hospital, a dementia care center, and 5 home care service centers. A total of 36 dyads participated, each comprising a community-dwelling person living with dementia aged 60 years or older and their primary family caregiver. The dyads were randomly allocated to either the intervention or control group. The intervention group received a caregiver-initiated, multimodal, mobile app–based individualized intervention for 4 weeks, whereas the control group continued with usual care. The primary outcomes were overall BPSD, agitated behavior, and depression. The secondary outcomes were nighttime sleep efficiency and caregiver competency in managing BPSD. Assessments were conducted at baseline, immediately after the intervention, and at a 2-week follow-up.

Results: Of the 36 randomized dyads, 33 were included in the final analysis. Although the intervention group showed greater reductions in overall BPSD, agitated behavior, and depression after the intervention, no significant group-by-time interaction effects were observed in the total sample. In the subgroup analysis of participants with clinically significant baseline BPSD, a statistically significant improvement in overall BPSD was found in favor of the intervention group (β=–12.885, 95% CI –24.530 to –1.240; *P*=.03). No significant effects were observed for either nighttime sleep efficiency or competence in the management of BPSD.

Conclusions: A mobile app–based individualized intervention may offer a flexible, caregiver-initiated approach to managing BPSD in home-care settings. While overall effects were limited, exploratory subgroup findings provided meaningful insights, indicating potential benefits for those with higher baseline symptom severity. The results highlight the need for further research on adaptive personalization and optimized intervention delivery to enhance the clinical effectiveness of digital dementia care.

Trial Registration: Clinical Research Information Service KCT0008713; https://tinyurl.com/83xaa5ab

(JMIR Mhealth Uhealth 2026;14:e79469) doi:10.2196/79469

KEYWORDS

behavioral and psychological symptoms of dementia; dementia; mobile app; nonpharmacological intervention; randomized controlled trial

XSL•FO
RenderX

Introduction

Cognitive decline is a common symptom of dementia, with behavioral and psychological symptoms of dementia (BPSD) recognized as the core features of the condition [1]. The nature of BPSD varies depending on the type and stage of the condition; however, these symptoms can appear at any time [1]. According to recent literature, nearly 90% of persons living with dementia experience BPSD [2]. Among these, symptoms such as apathy, depression, anxiety, irritability, agitation, and aggression are commonly observed in persons living with dementia [3], and often coexist with sleep disturbances, which are highly prevalent in this population [4-6]. In particular, a person living with dementia frequently exhibits reduced nighttime sleep efficiency and disrupted rest–activity rhythms, characterized by fragmented nocturnal sleep and frequent nighttime awakenings [4,5]. These disturbances are not simply co-occurring conditions but interact bidirectionally with BPSD, jointly contributing to increased mortality [7] and decreased quality of life [8]. Furthermore, BPSD and nocturnal sleep disruptions significantly increase caregiver burden [8-11] and care costs [8]. Therefore, implementing effective interventions targeting BPSD has become critical in dementia care. Such interventions can alleviate caregiver burden at the individual level and mitigate the broader societal impact of dementia care.

Nonpharmacological interventions are recommended as a first-line strategy for managing BPSD owing to their comparable efficacy to pharmacological treatments while minimizing adverse effects [12,13]. A wide range of nonpharmacological approaches has been developed to address the cognitive, psychological, social, and environmental dimensions of dementia care [12-14]. Interventions such as music therapy, exercise, and reminiscence therapy have demonstrated effectiveness in managing BPSD [14-17] and may be applied to support their prevention and management. Despite the growing emphasis on nonpharmacological interventions for BPSD, the COVID-19 pandemic has exacerbated the exclusion of community-dwelling persons living with dementia from essential care services. In particular, access to nonpharmacological activities was restricted by global social distancing measures [18-20]. These challenges persist beyond the pandemic, particularly among family caregivers, who often face limited access to therapy, logistical challenges in scheduling sessions, insufficient funding, and poor local accessibility [21].

In light of the constraints on home-based nonpharmacological interventions during and after the pandemic, digital approaches for remote delivery have gained growing importance [22]. Digital technologies, such as virtual reality, robotics, and mobile apps, have been widely applied to deliver diverse nonpharmacological interventions for persons living with dementia. These technologies have enabled activity-based programs, exercise, interactive activities, and coaching or counseling support [23-25]. A systematic review found that reminiscence therapy delivered through mobile apps enhanced social interaction [23], and a pet robot in long-term care settings helped alleviate BPSD [24]. However, most of these digital tools have been implemented in long-term care settings or require operation by clinicians or trained staff [25]. Among these, artificial intelligence–based cognitive stimulation programs and digital reminiscence therapies have been delivered by care staff in institutional environments [26,27]. A range of apps and digital platforms enables clinicians and formal caregivers to assess, record, and manage BPSD, with limited applicability to home settings [28-30]. Although mobile apps are also used to provide family caregivers with training on disease-related information, caregiving skills, and coping strategies [31,32], adherence to nonpharmacological interventions remains low [21,33]. Moreover, existing apps rarely enable family caregivers to directly deliver these interventions in real-world home settings, highlighting a critical gap in current digital care approaches. While these interventions have shown beneficial effects on BPSD, their impact could be further enhanced through greater individualization [25,26,34]. Therefore, there is a need for enhanced digital tools to support caregiver-led interventions and tailor strategies to the diverse contexts of persons living with dementia.

To address this gap, we developed a mobile app to empower family caregivers to deliver tailored interventions in real-world home settings. The app's usability and user acceptability were evaluated during its developmental phase among persons living with dementia and their caregivers [35]. The app guides caregivers through a structured process that includes the assessment of symptoms and potential triggers and recommends tailored nonpharmacological strategies that caregivers can implement. Because BPSD management often exceeds the competence of family caregivers, the app was also designed to gradually build their skills and confidence in delivering these interventions. Building on these findings, we designed a randomized controlled trial to assess the effectiveness of this mobile app–based intervention for persons living with dementia in a community setting. We hypothesized that it would alleviate the severity and frequency of BPSD compared with usual care. The objective of this randomized controlled trial was to evaluate the effectiveness of a mobile app that delivers individualized nonpharmacological interventions for reducing BPSD. The study also examined its effects on nighttime sleep efficiency and caregiver competence.

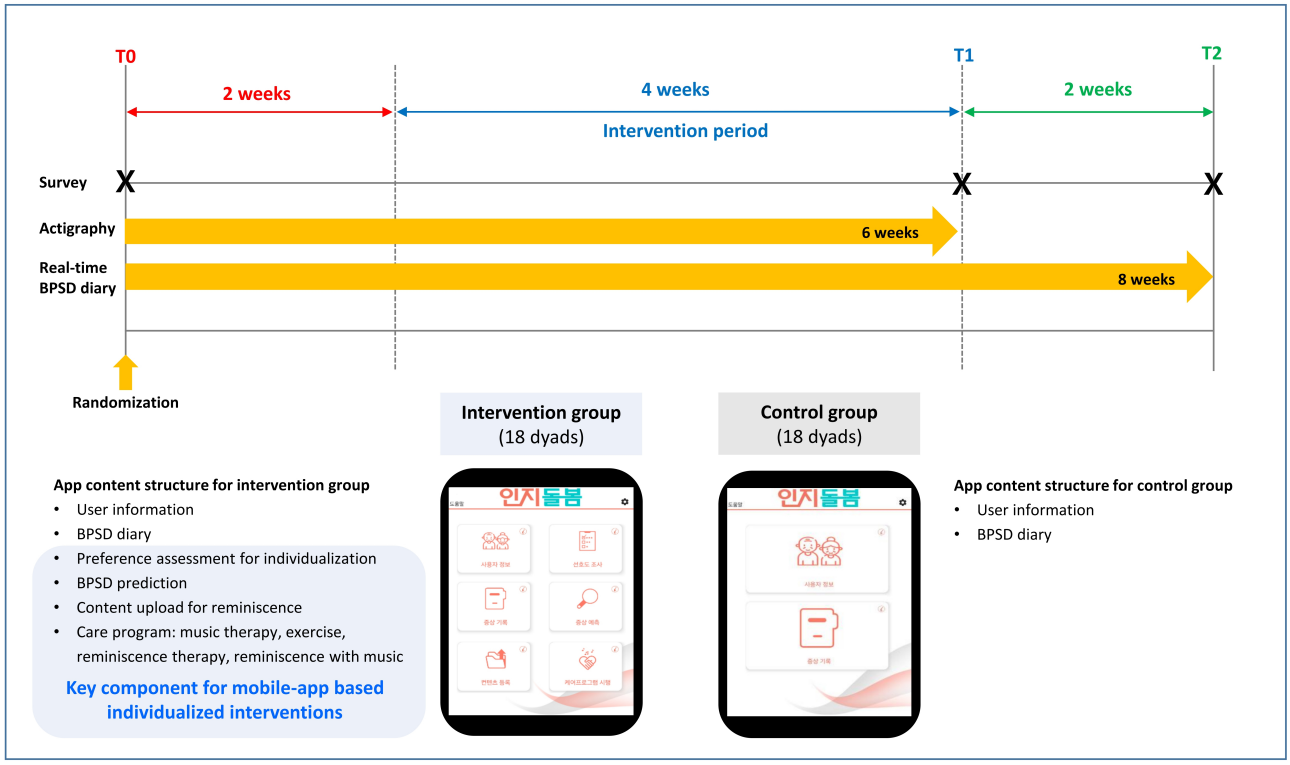
Methods

Study Design and Setting

This was a single-blind, parallel, randomized controlled trial designed to compare mobile app–based individualized nonpharmacological interventions with usual care. Participants were recruited from the neurology outpatient clinic of a tertiary hospital, a regional dementia care center, and 5 home care service centers. Figure 1 presents an overview of the study timeline and app components for each group.

XSL•FO
RenderX

Figure 1. Study timeline and app components by group. BPSD: behavioral and psychological symptoms of dementia.



Trained research assistants visited participants to collect outcome data and baseline information according to the study timeline and provided education on app use. They remained available to answer any questions and ensure completeness while family caregivers completed self-administered paper questionnaires. Outcome assessments were conducted at baseline (T0), 4 weeks postintervention (T1), and at 2 weeks of follow-up (T2). The trial was prospectively registered with the Clinical Research Information Service (registration number: KCT0008713). The study protocol has been published previously [36].

Participants

The study sample included persons living with dementia in a community setting and their family caregivers as a dyad. Eligibility was assessed by trained research assistants. The inclusion criteria for persons living with dementia were (1) aged 60 years or older, (2) any type of dementia, (3) Korean version of the Mini-Mental State Examination score <24, (4) exhibited BPSD at least once per week, and (5) lived in Seoul or Gyeonggi-do Province with their family caregivers. The inclusion criteria for family caregivers were (1) primary caregivers of persons living with dementia, and (2) no difficulties in reading and writing Korean. The participants were enrolled between October 2023 and March 2024, based on the eligibility criteria.

Ethical Considerations

This study was approved by the institutional review board of Yonsei University Health System (IRB number 4-2023-0660). Written informed consent was obtained from both persons living with dementia and their primary family caregivers after a detailed explanation of the study objectives and procedures. In cases where persons living with dementia were unable to provide consent due to impaired decision-making capacity, consent was obtained from a surrogate decision-maker. Given the dependency and vulnerability of the study population, special ethical considerations were applied throughout the study period. To ensure confidentiality, all participants were assigned identification numbers, and access to the collected data was restricted to the research team only. A small token of appreciation was provided based on the study participation period. Participants who completed the entire study received approximately US $146 (KRW ₩ 200,000) per dyad.

Sample Size

The sample size was calculated using Cohen's power analysis (repeated measure analysis of variance within factor: α=.05, power level=0.8, effect size=0.25, groups=2, and number of measurements=3). A minimum sample size of 28 participants was required to achieve a statistical power of 0.80. As the intervention was primarily implemented by caregivers, an anticipated 30% attrition rate was applied, based on previous intervention studies with dementia caregivers [37,38]. Accordingly, 36 patient–caregiver dyads were recruited from both groups.

Randomization and Blinding

Participants were randomly assigned to the intervention and control groups in a 1:1 allocation ratio. A researcher who was not in contact with the participants created a randomization list using a computer program (STATA version 13.1; StatCorp LP). Block randomization (with a block size of 4 to assign patients to 9 groups) was used to evenly allocate the dyads to the intervention and control groups. The generated randomization list was sealed in an envelope and stored. After participants provided consent and the study commenced, group allocation was disclosed only to the researcher responsible for data

XSL•FO
RenderX

collection and mobile app training. Persons living with dementia and their caregivers were not informed of their group allocation to minimize bias. Both the intervention and control groups used the same mobile app interface and wore actigraphy devices during the study period. The only distinction was that the intervention group used both symptom diaries and individualized nonpharmacological interventions through the app, whereas the control group used only symptom diaries via the app.

Intervention

All participants were provided with a tablet device (Galaxy Tab A8; Samsung) preinstalled with a mobile app. Both the intervention and control groups were instructed to record a real-time diary of BPSD throughout the study period using the app, allowing continuous monitoring of symptoms and possible triggers.

Only the intervention group received a 4-week individualized nonpharmacological intervention program via the app. Caregivers were instructed to initiate the intervention at least once daily but could use it more frequently or flexibly according to their needs. The program included 4 types of activities: music therapy, exercise, reminiscence therapy, and reminiscence therapy with music. To enhance individualization, the trained research assistants collected information on personal backgrounds, preferences, and physical functioning at baseline. The app also integrates individualized content, such as family-uploaded photographs, videos, individualized music playlists, and tailored exercise programs. Additionally, the intervention content was updated based on caregivers' requests to reflect their changing preferences and needs. Family caregivers in the intervention group autonomously selected and initiated interventions based on the app's recommendations and observed the response of persons living with dementia during the 4-week period.

In addition, the app incorporated a machine-learning–based BPSD prediction model developed and validated in our previous studies [35,39]. The model was trained using data collected from 187 community-dwelling persons living with dementia, with data from 35 others used for external validation. Information collected included demographic data, health status, and personality traits prior to dementia onset. Actigraphy devices tracked participants' sleep and activity, and caregivers documented daily BPSD symptoms and possible triggers in a symptom diary. Multiple algorithms were tested, including logistic regression, random forest, gradient boosting machine, and support vector machine. Among these, the gradient boosting machine demonstrated the best predictive performance, with average AUC values above 0.8 for most symptoms [39]. Based on these findings, the gradient boosting model was implemented within the mobile app to provide BPSD prediction. Furthermore, we conducted an umbrella review to identify evidence-based nonpharmacological approaches with demonstrated effectiveness for each BPSD to guide the selection of appropriate interventions [14]. Based on this review, we developed evidence-based recommendations, which were integrated into the app. Thus, this app recommends tailored nonpharmacological interventions based on BPSD subtypes predicted through machine learning. Meanwhile, the control group did not receive any nonpharmacological interventions through the app and continued routine care throughout the same period.

Measurements

Primary outcomes were assessed at 3 time points based on the observation of primary family caregivers. Secondary outcomes regarding nighttime sleep efficiency and caregiver competence in managing BPSD were also collected at baseline and at the 4-week postintervention period.

Primary Outcomes

The Korean version of the Neuropsychiatric Inventory (NPI) was used to assess overall dementia-related symptoms. The total score is calculated by multiplying the frequencies and severities of the domains, with a higher score indicating more BPSD [40]. Agitated behaviors were measured using the Korean version of the Cohen-Mansfield Agitation Inventory, which consists of 29 items that assess the frequency of agitated behaviors. The total score ranges from 29 to 203, with higher scores indicating a higher frequency of agitated behavior [41]. Depressive symptoms were evaluated using the Korean version of the Cornell Scale for Depression, which contains 19 items on depressive signs and symptoms [42]. The total score on this scale ranges from 0 to 38, with a higher score indicating a more depressed status.

Secondary Outcomes

The activity monitor wGT3X-BT (ActiGraph, LLC), a validated actigraphy device for assessing circadian rhythms in persons living with dementia [43], was used in this study. Persons living with dementia were instructed to wear the wrist-worn actigraphy device as consistently as possible, particularly at night. Sleep data were collected over 2 distinct periods: during the 2 weeks prior to the intervention (baseline period) and throughout the 4-week intervention period. The initial 2-week period was used to collect the data necessary for predicting BPSD patterns and establishing baseline information. The subsequent 4-week period was used to evaluate the effects of the intervention. Nighttime sleep data from these periods were extracted and analyzed to monitor changes in sleep patterns. Sleep efficiency was operationalized as nighttime sleep efficiency, given that it serves as a core indicator of overall sleep quality in persons living with dementia [44]. Previous studies have demonstrated that insufficient nighttime sleep is closely associated with the exacerbation of BPSD [4-6,45] and increased caregiver burden [11], highlighting its clinical relevance in dementia care. Nighttime sleep efficiency was calculated for the period between 8 PM and 8 AM as the ratio of total sleep time to time in bed, consistent with the operational definitions used in previous studies [5].

Our app was designed to provide a comprehensive care framework that includes symptom assessment, prediction of BPSD, recommendation of tailored nonpharmacological strategies, and delivery of interventions. This structured process was expected to enhance caregivers' competence in managing BPSD, thereby contributing to the effectiveness and sustainability of the intervention. Accordingly, caregiver competence was assessed using the competence scale in managing BPSD [46]. This is a self-report tool for measuring

XSL•FO
RenderX

knowledge, skills, and attitudes in the management of BPSD. In addition to the 28 core items, an additional question assessed family caregivers' general competency.

Engagement Measures

As both groups completed BPSD diaries for 8 weeks, the number of diary entries was obtained from in-app logs. During the 4-week intervention period, app usage data in the intervention group were extracted from the logs, including usage frequency and mean usage time for each program. The app dose for each participant was operationalized as the number of days used, the total number of sessions, and the mean session duration. Engagement level was defined as the degree to which the persons living with dementia actively participated in the app-guided nonpharmacological intervention. It was rated by family caregivers after each session on a 5-point Likert scale ranging from very bad (1) to very good (5).

Basic Information

Sociodemographic variables for persons living with dementia included sex, age, years of education, marital status, and dementia diagnosis. Cognitive function was assessed with the Korean version of Mini-Mental State Examination-2, with scores below 24 out of 30 indicating cognitive impairment [47]. Activities of daily living were measured using the Korean Activities of Daily Living scale [48], a 7-item measure of basic physical functions (eg, dressing, face washing, bladder/bowel control, eating, toilet use, and mobility). Higher scores reflect greater dependence, with scores ≥8 indicating dependency. Sociodemographic variables for family caregivers included sex, age, years of education, and relationship with persons living with dementia. Family caregiver–perceived intimacy was assessed using a single Likert item ("How close would you describe your relationship with the person living with dementia?") rated on a 5-point scale from very high (1) to very low (5). For analysis, scores of 1 - 2 were categorized as high intimacy, 3 as moderate intimacy, and 4 - 5 as low intimacy. Caregiver subjective health was assessed with a single 5-point item from very healthy (1) to very unhealthy (5), and for analysis, scores of 1 or 2 were categorized as healthy, and 3 to 5 as moderate-to-unhealthy.

Statistical Analyses

All statistical analyses were performed using SPSS Statistics for Windows (version 27.0; IBM Corp., Armonk, NY). All statistical tests were 2-tailed with a significance level of 0.05. Descriptive statistics were used to summarize participants' general characteristics. Baseline homogeneity between the intervention and control groups was assessed using the Pearson chi-square test, Fisher exact test, and the independent *t* test. We applied a mixed-effects model to examine whether changes in outcomes over time differed between the intervention and control groups across the 3 measurement time points (T0, baseline; T1, postintervention; and T2, follow-up). The model included fixed effects for group, time, and their interaction (group×time). The baseline value of each outcome variable was included as a covariate to adjust for initial differences between participants. A subject-level random intercept was specified to account for within-participant correlations arising from repeated measures. Time was modeled as a categorical factor to avoid assuming linearity. In this specification, each group×time coefficient reflects the time-specific difference-in-difference, that is, the between-group difference in change from baseline. This approach enabled us to (1) incorporate all available data under the missing-at-random assumption, (2) model within-subject covariance structures to improve estimation efficiency, and (3) perform an omnibus Wald/F test of overall trajectory differences. This reduced the risk of multiplicity compared with conducting separate per-visit analyses. Furthermore, as BPSD are expected to change gradually over time, alternative covariance structures were compared. Specifically, compound symmetry and first-order autoregressive structures were evaluated, and model fit was compared using the Akaike Information Criterion and the Bayesian Information Criterion. The first-order autoregressive structure provided the better fit and was therefore adopted. Although unstructured and independent structures converged, the final Hessian was not positive definite, and therefore, they were not considered further. To improve the accuracy of inference with a relatively small sample, degrees of freedom were adjusted using the Kenward–Roger method.

In addition, due to heterogeneity in baseline BPSD, we conducted an exploratory subgroup analysis of participants with baseline NPI scores ≥11, based on a commonly used threshold for clinically significant BPSD [49]. This analysis aimed to explore whether the intervention exerted greater effects in participants with more severe baseline symptoms. Within this subgroup, the same linear mixed-effects model including group, time, and group×time was fitted to examine whether the intervention effect differed among participants with higher baseline symptom burden. This model was chosen to ensure consistency with the main analysis and to account for repeated measures while estimating time-specific differences in intervention effects within this clinically meaningful subgroup.

Results

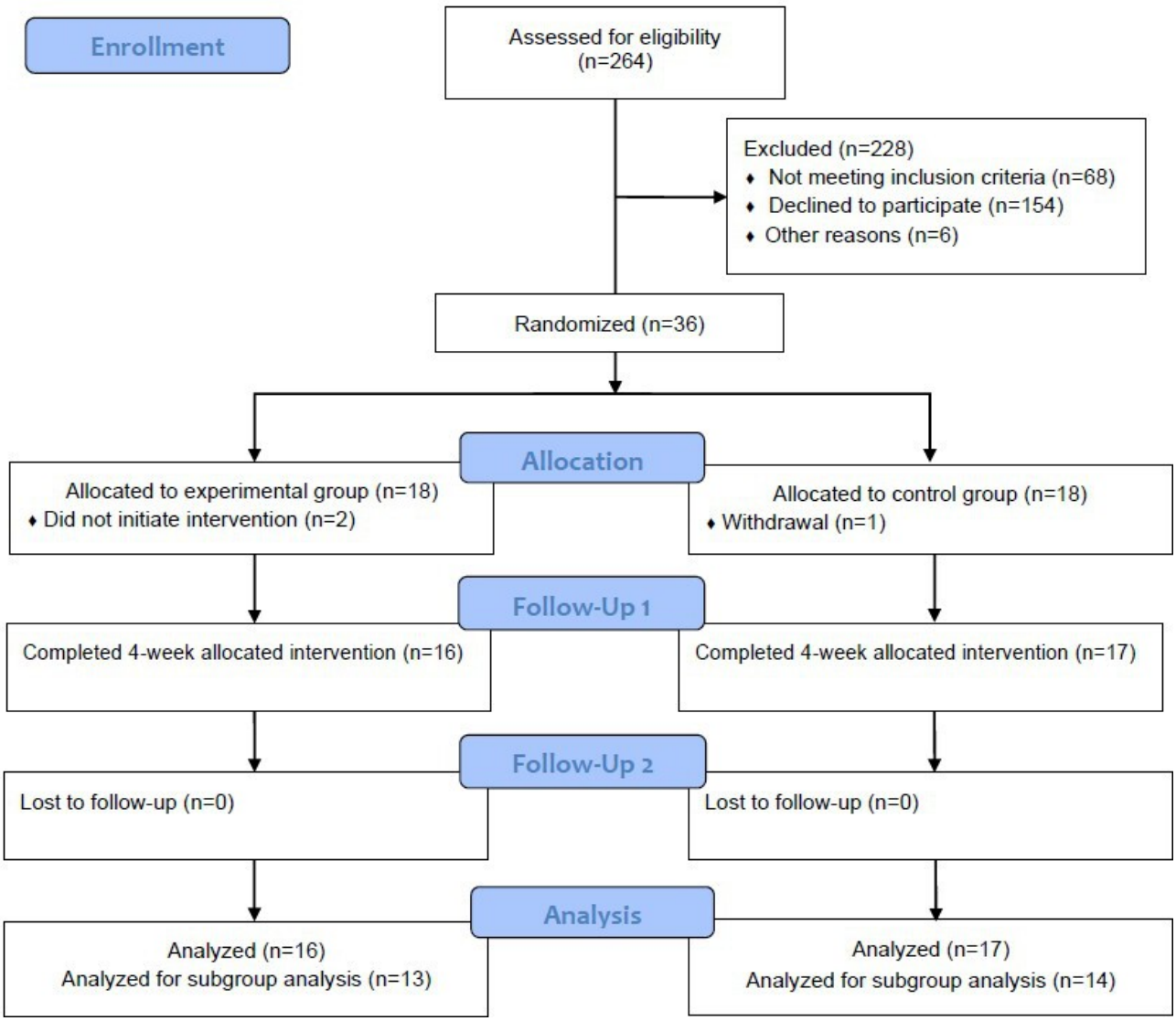
Participant Baseline Characteristics

Participants were recruited between October 2023 and March 2024, with follow-up being completed by April 2024. As shown in Figure 2, 264 participants were assessed for eligibility across 7 institutions (a neurology outpatient clinic in a tertiary hospital, a dementia care center, and 5 home care service centers). Of these, 36 dyads were enrolled and randomly allocated to either the intervention (n=18) or the control group (n=18). One dyad from the control group withdrew consent for personal reasons before the intervention. Two dyads in the intervention group withdrew because the intervention could not be initiated, as the persons living with dementia refused to participate due to worsening BPSD. Therefore, data from 33 dyads were included in the analysis. The participants' baseline characteristics are presented in Table 1. The mean age of the participants was 82.70 (SD 7.10) years, and most of them (24/33, 72.7%) were female. A homogeneity test of the persons living with dementia showed no significant differences between the groups at baseline. As shown in Table 2, the mean age of the family caregivers was 65.03 (SD 11.28) years, and most of them (27/33, 81.8%) were

XSL•FO
RenderX

female. Family caregivers had provided care for their relatives with dementia for an average of 4.39 (SD 4.37) years, offering approximately 16.32 (SD 8.58) hours of care per day.

Figure 2. Study flow chart of participants.



XSL•FO
RenderX

Table . Participants' characteristics at baseline (N=33).

Characteristics	Total (N=33)	Intervention (n=16)	Control (n=17)	*t* test (*df*)	Chi-square (*df*)	*P* value
Sex, n (%)				—[a]	0.248 (1)	.71[b]
Male	9 (27.3)	5 (31.3)	4 (23.5)			
Female	24 (72.7)	11 (68.8)	13 (76.5)			
Age (years), mean (SD)	82.70 (7.10)	84.19 (6.15)	81.29 (7.81)	−1.178 (31)	—	.25
Marital status, n (%)					0.047 (1)	.83
Married	20 (60.6)	10 (62.5)	10 (58.8)			
Widowed	13 (39.4)	6 (37.5)	7 (41.2)			
Education years, n (%)				—	0.029 (1)	.87
≤6 years	17 (51.5)	8 (50)	9 (52.9)			
>6 years	16 (48.5)	8 (50)	8 (47.1)			
Diagnosis, n (%)				—	0.732 (1)	.48[b]
AD[c]	12 (36.4)	7 (43.8)	5 (29.4)			
Others	21 (63.6)	9 (56.3)	12 (70.6)			
MMSE[d], mean (SD)	16.67 (5.78)	15.06 (7.12)	18.18 (3.76)	1.556 (22.461)	—	.13
ADL[e], mean (SD)	12.24 (4.24)	12.38 (4.67)	12.12 (3.94)	−0.171 (31)	—	.87
Overall BPSD[f], mean (SD)	30.64 (22.48)	30.69 (21.05)	30.59 (24.39)	−0.012 (31)	—	.99
Agitated behavior, mean (SD)	48.24 (18.07)	46.19 (16.53)	50.18 (19.71)	0.628 (31)	—	.54
Depression, mean (SD)	9.88 (6.10)	9.25 (4.70)	10.47 (7.28)	0.576 (27.537)	—	.57
Nighttime sleep efficiency (%), mean (SD)	94.87(2.38)	95.38 (2.73)	94.38 (1.95)	−1.209 (31)	—	.24

[a]Not applicable.

[b]Fisher exact test.

[c]AD: Alzheimer disease.

[d]MMSE: Mini-Mental State Examination.

[e]ADL: activities of daily living.

[f]BPSD: behavioral and psychological symptoms of dementia.

XSL•FO
RenderX

Table . Family caregivers' characteristics at baseline (N=33).

Characteristics	Total (n=33)	Intervention (n=16)	Control (n=17)	*t* test (df)	Chi-square (*df*)	*P* value
Sex, n (%)				—[a]	0.007 (1)	>.99[b]
Male	6 (18.2)	3 (18.8)	3 (17.6)			
Female	27 (81.8)	13 (81.3)	14 (82.4)			
Age (years), mean (SD)	65.03 (11.28)	64.88 (11.09)	65.18 (11.80)	0.076 (31)	—	.94
Education years, n (%)				—	0.004 (1)	>.99[b]
≤6 years	4 (12.1)	2 (12.5)	2 (11.8)			
>6 years	29 (87.9)	14 (87.5)	15 (88.2)			
Relationship, n (%)				—	0.047 (1)	.83
Spouse	13 (39.4)	6 (37.5)	7 (41.2)			
Child	20 (60.6)	10 (62.5)	10 (58.8)			
Caregiver–perceived intimacy with persons living with dementia , n (%)				—	0.170 (1)	>.99[b]
High	28 (84.8)	14 (87.5)	14 (82.4)			
Moderate	5 (15.2)	2 (12.5)	3 (17.6)			
Caregiving period (years), mean (SD)	4.39 (4.37)	4.84 (3.48)	3.97 (5.14)	−0.568 (31)	—	.57
Caregiving hours per day, mean (SD)	16.32 (8.58)	15.54 (9.35)	17.05 (8.01)	0.498 (31)	—	.62
Subjective health status, n (%)				—	0.750 (1)	.39
Healthy	16 (48.5)	9 (56.3)	7 (41.2)			
Moderate-unhealthy	17 (51.5)	7 (43.8)	10 (58.8)			
Competence in managing BPSD[c,d], mean (SD)	102.48 (13.95)	102.00 (15.41)	103.00 (12.74)	0.196 (29)	—	.85

[a]Not applicable.

[b]Fisher exact test.

[c]Competence in managing BPSD was assessed in 15 control group participants owing to missing data.

[d]BPSD: behavioral and psychological symptoms of dementia.

Effects of Mobile App–Based Individualized Interventions on Outcome Variables

Table 3 presents the changes in mean scores and group×time effects for each outcome variable, with baseline scores included as covariates in the linear mixed model analysis. Descriptive trends indicated reductions in overall BPSD, agitated behavior, and depression in both groups after the 4-week intervention. The linear mixed model analysis revealed no statistically significant group×time effects for overall BPSD (β=–8.647, 95% CI –18.737 to 1.442; *P*=.09), agitated behavior (β=–1.290, 95% CI –7.236 to 4.655; *P*=.67), or depression (β=.121, 95% CI –2.794 to 3.037; *P*=.93).

XSL•FO
RenderX

Table . Effects of the mobile app–based individualized nonpharmacological intervention on outcome variables.

Outcome variables	Intervention (n=16)	Control (n=17)	Estimate[a] (95% CI)	*P* value
Overall BPSD[b], mean (SD)				
T1[c]-T0[d]	−11.00 (18.46)	−2.35 (13.13)	−8.647 (−18.737 to 1.442)	.09
T2[e]-T0	−8.44 (20.37)	−3.76 (17.89)	−4.673 (−16.689 to 7.343)	.44
Agitated behavior, mean (SD)				
T1-T0	−1.94 (7.45)	−0.65 (7.56)	−1.290 (−7.236 to 4.655)	.67
T2-T0	−4.38 (10.26)	−3.24 (10.69)	−1.140 (−7.913 to 5.634)	.74
Depression, mean (SD)				
T1-T0	−0.94 (5.69)	−1.06 (4.41)	0.121 (−2.794 to 3.037)	.93
T2-T0	−2.13 (4.88)	−0.59 (3.26)	−1.537 (−4.832 to 1.758)	.36
Nighttime sleep efficiency (%), mean (SD)				
T1-T0	−0.57 (1.42)	−0.52 (1.95)	−0.054 (−1.271 to 1.163)	.93
Competence in managing BPSD, mean (SD)				
T1-T0	6.00 (12.78)	0.07 (12.19)	5.933 (−6.979 to 18.846)	.36

[a]All models were adjusted for baseline scores as covariates. The control group at baseline (T0) was used as the reference.

[b]BPSD: behavioral and psychological symptoms of dementia.

[c]T1: after 4 weeks.

[d].T0: baseline

[e]T2: 2-week follow-up.

At the 2-week postintervention period, overall BPSD increased in the intervention group, while the control group maintained a downward trend. This difference, however, did not reach statistical significance (β=–4.673, 95% CI –16.689 to 7.343; *P*=.44). Agitated behaviors continued to decrease in both groups, with no statistically significant interaction effect (β=–1.140, 95% CI –7.913 to 5.634; *P*=.74). Depressive symptoms increased in the control group but showed a downward trend in the intervention group. However, no statistically significant group×time effect was observed (β=–1.537, 95% CI –4.832 to 1.758; *P*=.36).

Regarding secondary outcomes, no statistically significant group×time interaction effect on nighttime sleep efficiency was observed after the intervention (β=−.054, 95% CI −1.271 to 1.163; *P*=.93). Family caregivers' level of competence in managing BPSD increased in the intervention group; however, the group×time effect was not statistically significant (β=5.933, 95% CI –6.979 to 18.846; *P*=.36). No harm was recorded during the study period in either group.

Subgroup Analysis of Outcomes

Table 4 presents a subgroup analysis among participants with baseline NPI scores ≥11, a cutoff previously identified as indicating clinically significant BPSD. Greater reductions in overall BPSD, agitated behavior, and depression were observed in the intervention group over the study period. However, the linear mixed model analysis controlling for baseline scores showed that only the group×time interaction for overall BPSD was statistically significant after the 4-week intervention (β=–12.885, 95% CI –24.530 to –1.240; *P*=.03).

XSL•FO
RenderX

Table . Subgroup analysis of outcomes in participants with Neuropsychiatric Inventory≥11 at baseline.

Outcome variables	Intervention (n=13)	Control (n=14)	Estimate (95% CI)	*P* value
Overall BPSD[a], mean (SD)				
T1-T0[b,c]	−15.38 (16.70)	−2.50 (14.53)	−12.885 (−24.530 to −1.240)	.03[d]
T2-T0[e]	−12.23 (20.30)	−3.86 (19.85)	−8.374 (−22.221 to 5.474)	.23
Agitated behavior, mean (SD)				
T1-T0	−2.08 (8.26)	−0.07 (8.17)	−2.005 (−9.116 to 5.105)	.57
T2-T0	−6.00 (10.60)	−3.14 (11.66)	−2.857 (−10.944 to 5.230)	.48
Depression, mean (SD)				
T1-T0	−1.92 (5.41)	−1.07 (4.68)	−0.852 (−4.073 to 2.369)	.60
T2-T0	−3.31 (3.95)	−0.07 (3.22)	−3.236 (−6.753 to 0.281)	.07
Nighttime sleep efficiency (%), mean (SD)				
T1-T0	−0.57 (1.56)	−0.42 (1.95)	−0.152 (−1.558 to 1.255)	.83
Competence in managing BPSD, mean (SD)				
T1-T0	6.85 (23.76)	0.77 (12.99)	6.077 (−9.423 to 21.576)	.43

[a]BPSD: behavioral and psychological symptoms of dementia.

[b]T1: after 4 weeks.

[c]T0: baseline.

[d]Statistically significant values ($P<.05$).

[e]T2: 2-week follow-up.

Use of Behavioral and Psychological Symptoms of Dementia Diary and App-Based Individualized Nonpharmacological Interventions

During the 4-week intervention period, the frequency and intensity of use varied across participants. The total number of days the intervention was used ranged from 6 to 25, and the total number of sessions ranged from 9 to 45. The average session duration ranged from 11 to 54 minutes. On average, participants engaged with the intervention on approximately 15 days, completing about 22 sessions, with a mean session duration of around 30 minutes.

Tables 5 and 6 present the logs of the BPSD diary records and mobile app–based individualized nonpharmacological interventions. Over the 8-week study period, the total number of diary records was 1819 in the intervention group and 1324 in the control group. The mean number of records per participant was higher in the intervention group than in the control group. Music therapy was the most frequently used intervention in the intervention group, demonstrating the highest usage time (mean 38.21, SD 33.46 min) and engagement level (mean 3.84, SD 0.91). Although exercise therapy had a shorter average usage time (23.12, SD 15.37 min), it recorded the second-highest frequency of use (94/357, 26.3%). Reminiscence therapy had a relatively low usage frequency (82/357, 23.0%) and engagement (mean 3.49, SD 1.03). However, its average usage time exceeded the overall intervention mean of 31.40 (SD 28.79) min. In contrast, reminiscence therapy with music showed the lowest usage frequency (29/357, 8.1%) and shortest average usage time (12.76, SD 9.14 min); however, it received a relatively positive engagement rating (mean 3.69, SD 0.85).

Table . Summary of behavioral and psychological symptoms of dementia diary records.

	Control	Intervention
Total number of records	1324	1819
Number of records		
Mean (SD)	77.76 (35.17)	113.81 (166.42)
Range	51 - 164	42 - 735

XSL•FO
RenderX

Table . Summary of mobile app-based nonpharmacological intervention usage.

Type of intervention[a]	Frequency, n (%)	Usage time, mean (SD)	Engagement level, mean (SD)
Total	357 (100)	31.40 (28.79)	3.71 (0.93)
Music therapy	152 (42.6)	38.21 (33.46)	3.84 (0.91)
Exercise	94 (26.3)	23.12 (15.37)	3.71 (0.84)
Reminiscence therapy	82 (23.0)	34.88 (30.90)	3.49 (1.03)
Reminiscence with music	29 (8.1)	12.76 (9.14)	3.69 (0.85)

[a]Intervention usage data were collected only for the intervention group.

Exploratory analyses were conducted to examine the relationship between intervention dose and changes in outcome variables. Participants were categorized into high-dose and low-dose groups based on the median values of usage metrics (total frequency and mean usage time). Dose-response exploratory plots are presented in Multimedia Appendix 1. Exploratory comparisons showed no significant differences in pre-post changes of outcome variables between participants with higher and lower intervention usage. This result remained consistent even among participants with baseline NPI scores ≥11.

Discussion

Principal Results

Unlike previous studies that primarily implemented clinician-led interventions or were conducted in institutional settings, this study assessed a fully caregiver-driven app-based intervention in real-world community environments. Family caregivers were able to independently monitor symptoms and deliver tailored nonpharmacological interventions at home. In this way, this study demonstrated the feasibility of a decentralized, technology-enabled dementia care model. This approach aligns with the growing demand for accessible and scalable solutions that reduce the reliance on formal care systems. No statistically significant differences were observed between the intervention and control groups in overall BPSD, agitated behavior, or depression after the intervention. However, an exploratory subgroup analysis of participants with more severe BPSD (NPI≥11) suggested a potential benefit (β=–12.885; *P*=.03). Given the pilot nature of this study, these findings should be regarded as preliminary and exploratory and not interpreted as definitive. Nevertheless, even mild reductions in BPSD can have clinical significance, as they may help reduce caregiver burden, delay institutionalization, and improve quality of life for both persons living with dementia and their families.

Comparison With Prior Work

Several factors may account for these small effects. First, regular assessment and documentation of BPSD, including symptom type, severity, and triggers, has been widely recognized as an effective strategy for managing these symptoms [29,50-54]. However, previous studies have often relied on digital databases, such as electronic medical records, which rely on retrospective documentation with limited contextual detail [55,56]. In contrast, the mobile app was used in this study to collect real-time, caregiver-reported data in the home setting. This approach allowed for more precise identification and characterization of BPSD and their triggers as they occurred. The NPI, a commonly used tool for assessing BPSD, has been adapted for daily diaries in mobile apps, minimizing recall bias and enabling users to record and monitor symptoms in real time [52,53]. This tool facilitates easier access to symptom information and enhances symptom awareness, offering therapeutic value through improved recognition and monitoring [53]. In this study, both the intervention and control groups were instructed to assess and document BPSD, including the type, frequency, and severity, using a mobile app modeled after the NPI. Moreover, family caregivers were asked to record potential triggers of BPSD, such as pain, discomfort, or environmental stressors. This approach is consistent with the Need-driven Dementia-compromised Behavior model [57], which views BPSD as an expression of unmet needs triggered by proximal factors. By regularly recording these symptoms and their potential triggers, family caregivers may be better equipped to recognize and preempt these unmet needs. This allows them to address potential issues early, before they escalate into more severe behavioral manifestations [58]. Overall, systematic documentation of BPSD and their triggers via a mobile app serves not only as a real-time monitoring tool but also as a key component of proactive, person-centered dementia care. Such an approach, therefore, aligns with dementia care frameworks, which emphasize that comprehensive assessment should guide tailored nonpharmacological strategies [54].

Second, it is important to consider the individual abilities of persons living with dementia when applying nonpharmacological interventions. Factors such as sensory function, attention span, and mobility limitations can significantly influence the appropriateness and effectiveness of the intervention [59]. To address individual needs, our study included 4 types of mobile app interventions, allowing for flexible selection according to user characteristics. We also made efforts to incorporate individual preferences into the intervention planning process, as tailoring individual preferences may assist in treatment adherence [58]. Preferences were continuously updated based on caregivers' observations of persons living with dementia's responses during the sessions. However, it is worth noting that the tailoring efforts in this study relied primarily on caregiver-reported observations of preferences rather than direct input from the persons living with dementia. To complement these efforts, the difficulty level of each activity could be adjusted according to the user's cognitive and physical function to ensure appropriate challenge and accessibility. In our app, individualized multimedia content such as photos, videos, and music reflecting users' memories

XSL•FO
RenderX

and preferences was incorporated. Future updates could further expand this feature to enhance engagement and emotional connection. This can also be achieved by analyzing and incorporating in-app records, including caregiver-entered engagement data and log data on intervention use. Furthermore, integrating external factors, including real-time data from wearable sensors, environmental monitoring, or smart home technologies, may further improve the adaptability of the intervention [60,61].

Finally, unlike traditional nonpharmacological interventions, which typically maintain a consistent intensity based on the facilitator's planned structure [15-17], our approach allowed greater flexibility. Participants initiated sessions according to their own preferences. This flexible delivery model may have led to variability in outcomes and how participants engage with the program [62,63]. Consistent with this design feature, we observed considerable variation in usage metrics, including total usage days, number of sessions, and average duration per session. These findings indicate differences in intervention fidelity, which may explain the heterogeneous use patterns observed across different components of the program, including music and reminiscence therapy. Music therapy was the most frequently used intervention in this study, likely due to its ease of use and flexible format [64]. Despite its popularity, however, it did not lead to significant reductions in BPSD. This finding echoes prior findings that family-delivered music interventions show limited effects, whereas those provided by trained professionals yield clearer benefits [33]. In contrast, reminiscence therapy in our study showed the lowest levels of usage and engagement, possibly reflecting greater usability challenges than those associated with music- or exercise-based interventions. Yet, recent evidence indicates that digital reminiscence therapy can be more effective than traditional formats [65].

Our exploratory analyses showed no significant differences between higher- and lower-usage participants, indicating that dosage alone may not drive clinical effects. A previous study reported that factors such as intervention quality, caregiver capacity, and contextual support likely play a more critical role in shaping outcomes [66]. In institutional settings, professional caregivers such as nurses can integrate cognitive stimulation activities into daily routines. By contrast, in home environments, limited expertise and resources among family caregivers may reduce intervention fidelity and impact [67]. To address this, structured support should be provided in community settings, so that caregivers can use the app independently while still receiving regular guidance from trained professionals or volunteers [58,68]. Such support, together with age-friendly interface improvements and user education, may help enhance digital competencies and promote more consistent engagement. These improvements, in turn, could increase the effectiveness of caregiver-initiated, app-based nonpharmacological interventions [66,68]. Broader contextual factors, such as differences in broadband access, digital literacy, and cultural familiarity, may also affect how interventions are used and how well they work. In particular, the fact that many caregivers in our study were older adults may have influenced their ability to consistently use the app. Therefore, tailored implementation strategies are needed for different care settings [66].

Implications for Practice and Policy

This study provides preliminary evidence on the potential of caregiver-driven, technology-enabled dementia care within the Korean context. The findings suggest that integrating structured guidance and support could enhance caregiver capacity, ensuring a more effective and sustainable implementation of nonpharmacological interventions in home settings. Given the global rise in the prevalence of dementia and reliance on informal caregivers, these exploratory results may inform future efforts to adapt and refine digital intervention frameworks.

Future Research

Building on the current findings, future studies should consider larger and more diverse samples and examine the effects of stratifying participants according to baseline BPSD severity. Moreover, the development of adaptive personalization algorithms that incorporate direct and continuous monitoring of persons living with dementia's responses may enhance intervention responsiveness and reduce caregiver burden. Such systems should also dynamically adapt to factors such as modality of use, thereby revealing intervention preferences and informing recommendations for optimal dosage. In addition, tailoring recommendations based on cognitive and physical function as well as BPSD severity may further refine the personalization of nonpharmacological interventions. Finally, evaluating the implementation of this intervention in long-term care or rural settings may extend its utility and scalability. The BPSD diary itself may also serve as a monitoring strategy that helps alleviate symptoms. Therefore, a 3-arm randomized controlled trial including (1) no diary and no intervention, (2) diary monitoring only, and (3) diary monitoring with the intervention could provide robust evidence on their individual and combined effects.

Limitations

Despite its strengths, this study had several limitations. First, the sample size was relatively small and restricted to urban community-dwelling participants in South Korea, which may limit the generalizability of the findings. While South Korea is one of the most digitally advanced countries [69], the study participants were predominantly from urban areas. As such, the observed feasibility and engagement levels may not fully reflect outcomes in populations with limited broadband access or in rural settings where digital resources are less developed. In addition, the sample size was calculated a priori based on a medium effect size. However, the wide confidence intervals raise the possibility of a type II error and suggest that the study may have been underpowered to detect smaller-than-expected effects. Accordingly, these findings should be regarded as preliminary and interpreted with caution. Second, the personalization of interventions was based on caregiver-reported observations, with limited direct input from persons living with dementia. This reliance on proxy reports may have introduced reporting bias and restricted the extent to which the intervention reflected the lived experiences and preferences of persons living with dementia themselves. Finally, nighttime sleep efficiency,

XSL•FO
RenderX

included as a secondary outcome, was already high at baseline (94%), which is comparable to or even higher than levels reported in previous studies involving persons living with dementia [44]. This suggests that participants in our study exhibited relatively preserved sleep quality at baseline. A potential ceiling effect may therefore have constrained the magnitude of observable change. As a result, the interpretation of sleep-related outcomes should take this limitation into account.

Conclusions

This study evaluated a mobile app–based individualized nonpharmacological intervention for managing BPSD in community-dwelling persons living with dementia. Although improvements were observed, they did not reach statistical significance in the overall analysis. Exploratory findings indicated potential benefits for participants with more severe baseline BPSD; however, these results should be interpreted with caution. Nevertheless, the findings highlight the potential value of empowering family caregivers to deliver flexible, person-centered care at home through digital tools. Future research should refine personalized strategies and optimize intervention delivery by using usage data. Despite its limitations, including a small sample size and variability in engagement, this study shows the potential of caregiver-driven, app-based, individualized nonpharmacological intervention for BPSD in the home and community setting. Continued efforts to integrate digital interventions within community-based care frameworks, accompanied by education and guidance for family caregivers, may contribute to the development of sustainable person-centered dementia care approaches.

Funding

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (grant number RS-2020-NR049581). MY was supported by the 2025 Kyu-Hee Lee Post-Doctoral Fellowship from Yonsei University College of Nursing. The funders had no involvement in the study design, data collection, analysis, interpretation, or the writing of the manuscript.

Data Availability

Data related to the findings reported in this manuscript may be made available by the corresponding author upon reasonable request.

Authors' Contributions

EC contributed to the conceptualization, methodology, funding acquisition, and writing of the original draft, as well as the review and editing of the manuscript. SH, EK, and JC contributed to writing the original draft and to the review and editing of the manuscript and were also involved in formal analysis and data curation. MJK contributed to writing the original draft and to the review and editing of the manuscript. MY contributed to writing the original draft and to the review and editing of the manuscript, as well as to methodology, formal analysis, and data curation. All authors read and approved the final version of the manuscript.

Multimedia Appendix 1
Baseline-post changes according to total frequency and mean usage time within the intervention group.
[DOCX File, 274 KB - mhealth_v14i1e79469_app1.docx]

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 1269 KB - mhealth_v14i1e79469_app2.pdf]

References

1. Arvanitakis Z, Shah RC, Bennett DA. Diagnosis and management of dementia: review. JAMA 2019 Oct 22;322(16):1589-1599. [doi: 10.1001/jama.2019.4782] [Medline: 31638686]
2. Laganà V, Bruno F, Altomari N, et al. Neuropsychiatric or behavioral and psychological symptoms of dementia (BPSD): focus on prevalence and natural history in Alzheimer's disease and frontotemporal dementia. Front Neurol 2022;13:832199. [doi: 10.3389/fneur.2022.832199] [Medline: 35812082]
3. Kwon CY, Lee B. Prevalence of behavioral and psychological symptoms of dementia in community-dwelling dementia patients: a systematic review. Front Psychiatry 2021;12:741059. [doi: 10.3389/fpsyt.2021.741059] [Medline: 34744832]
4. Zhang Y, Ren R, Yang L, et al. Sleep in Alzheimer's disease: a systematic review and meta-analysis of polysomnographic findings. Transl Psychiatry 2022;12(1):136. [doi: 10.1038/s41398-022-01897-y]
5. Ishimaru D, Tanaka H, Nagata Y, Ogawa Y, Fukuhara K, Nishikawa T. Aspects of rest-activity rhythms associated with behavioral and psychological symptoms in moderate and severe dementia: results of a cross-sectional analysis. Alzheimer Dis Assoc Disord 2023;37(4):322-327. [doi: 10.1097/WAD.0000000000000584] [Medline: 37824629]

XSL•FO
RenderX

6. Kuzmik A, Boltz M, BeLue R, Galvin JE, Arendacs R, Resnick B. Factors associated with sleep quality in hospitalized persons with dementia. Alzheimer Dis Assoc Disord 2022;36(3):253-258. [doi: 10.1097/WAD.0000000000000502] [Medline: 36001764]
7. Bränsvik V, Granvik E, Minthon L, Nordström P, Nägga K. Mortality in patients with behavioural and psychological symptoms of dementia: a registry-based study. Aging Ment Health 2021 Jun;25(6):1101-1109. [doi: 10.1080/13607863.2020.1727848] [Medline: 32067466]
8. Tampi RR, Bhattacharya G, Marpuri P. Managing behavioral and psychological symptoms of dementia (BPSD) in the era of boxed warnings. Curr Psychiatry Rep 2022 Sep;24(9):431-440. [doi: 10.1007/s11920-022-01347-y] [Medline: 35781675]
9. Kim B, Noh GO, Kim K. Behavioural and psychological symptoms of dementia in patients with Alzheimer's disease and family caregiver burden: a path analysis. BMC Geriatr 2021 Mar 5;21(1):160. [doi: 10.1186/s12877-021-02109-w] [Medline: 33663416]
10. Sakka M, Goto J, Kita S, Sato I, Soejima T, Kamibeppu K. Associations among behavioral and psychological symptoms of dementia, care burden, and family-to-work conflict of employed family caregivers. Geriatr Gerontol Int 2019 Jan;19(1):51-55. [doi: 10.1111/ggi.13556] [Medline: 30556370]
11. Tractenberg RE, Singer CM, Cummings JL, Thal LJ. The sleep disorders inventory: an instrument for studies of sleep disturbance in persons with Alzheimer's disease. J Sleep Res 2003 Dec;12(4):331-337. [doi: 10.1046/j.0962-1105.2003.00374.x] [Medline: 14633245]
12. Dyer SM, Harrison SL, Laver K, Whitehead C, Crotty M. An overview of systematic reviews of pharmacological and non-pharmacological interventions for the treatment of behavioral and psychological symptoms of dementia. Int Psychogeriatr 2018 Mar;30(3):295-309. [doi: 10.1017/S1041610217002344] [Medline: 29143695]
13. Meng X, Sun J, Du Z, et al. Comparisons of nonpharmacological caregiver interventions for behavioural and psychological symptoms of dementia: a systematic review and network meta-analysis. Int J Nurs Stud 2025 Oct;170:105153. [doi: 10.1016/j.ijnurstu.2025.105153] [Medline: 40639204]
14. Cho E, Lee JY, Yang M, Jang J, Cho J, Kim MJ. Symptom-specific non-pharmacological interventions for behavioral and psychological symptoms of dementia: an umbrella review. Int J Nurs Stud 2024 Nov;159:104866. [doi: 10.1016/j.ijnurstu.2024.104866] [Medline: 39163681]
15. Pedersen SKA, Andersen PN, Lugo RG, Andreassen M, Sütterlin S. Effects of music on agitation in dementia: a meta-analysis. Front Psychol 2017;8:742. [doi: 10.3389/fpsyg.2017.00742] [Medline: 28559865]
16. Barreto PDS, Demougeot L, Pillard F, Lapeyre-Mestre M, Rolland Y. Exercise training for managing behavioral and psychological symptoms in people with dementia: a systematic review and meta-analysis. Ageing Res Rev 2015 Nov;24(Pt B):274-285. [doi: 10.1016/j.arr.2015.09.001] [Medline: 26369357]
17. Saragih ID, Tonapa SI, Yao CT, Saragih IS, Lee BO. Effects of reminiscence therapy in people with dementia: a systematic review and meta-analysis. J Psychiatr Ment Health Nurs 2022 Dec;29(6):883-903. [doi: 10.1111/jpm.12830] [Medline: 35348260]
18. Cohen G, Russo MJ, Campos JA, Allegri RF. Living with dementia: increased level of caregiver stress in times of COVID-19. Int Psychogeriatr 2020 Nov;32(11):1377-1381. [doi: 10.1017/S1041610220001593] [Medline: 32729446]
19. Kwon HS, Sung W, Park KU, et al. Impact of the COVID-19 pandemic on mortality and loss to follow-up among patients with dementia receiving anti-dementia medications. Sci Rep 2024 Apr 5;14(1):7986. [doi: 10.1038/s41598-024-58316-z] [Medline: 38575660]
20. Wang H, Li T, Barbarino P, et al. Dementia care during COVID-19. Lancet 2020 Apr 11;395(10231):1190-1191. [doi: 10.1016/S0140-6736(20)30755-8] [Medline: 32240625]
21. Ross SD, Ziegert N, Rodriguez FS. Implementation of non-pharmacological interventions in dementia care: family caregiver perspective. Home Health Care Manag Pract 2024 Feb;36(1):20-30. [doi: 10.1177/10848223231174226]
22. Sohn M, Yang J, Sohn J, Lee JH. Digital healthcare for dementia and cognitive impairment: a scoping review. Int J Nurs Stud 2023 Apr;140:104413. [doi: 10.1016/j.ijnurstu.2022.104413] [Medline: 36821951]
23. Hoel V, Feunou CM, Wolf-Ostermann K. Technology-driven solutions to prompt conversation, aid communication and support interaction for people with dementia and their caregivers: a systematic literature review. BMC Geriatr 2021 Mar 4;21(1):157. [doi: 10.1186/s12877-021-02105-0] [Medline: 33663401]
24. Chan DKY, Chan LKM, Kuang YM, Le MNV, Celler B. Digital care technologies in people with dementia living in long-term care facilities to prevent falls and manage behavioural and psychological symptoms of dementia: a systematic review. Eur J Ageing 2022 Sep;19(3):309-323. [doi: 10.1007/s10433-021-00627-5] [Medline: 36052197]
25. Cho E, Shin J, Seok JW, et al. The effectiveness of non-pharmacological interventions using information and communication technologies for behavioral and psychological symptoms of dementia: a systematic review and meta-analysis. Int J Nurs Stud 2023 Feb;138:104392. [doi: 10.1016/j.ijnurstu.2022.104392] [Medline: 36434931]
26. Hird N, Osaki T, Ghosh S, Palaniappan SK, Maeda K. Enabling personalization for digital cognitive stimulation to support communication with people with dementia: pilot intervention study as a prelude to AI development. JMIR Form Res 2024;8:e51732. [doi: 10.2196/51732]
27. Moon S, Park K. The effect of digital reminiscence therapy on people with dementia: a pilot randomized controlled trial. BMC Geriatr 2020 May 6;20(1):166. [doi: 10.1186/s12877-020-01563-2] [Medline: 32375661]

XSL•FO
RenderX

28. Ishikawa A, Onchi E, Yamanaka K. Usability of behavioural recording application (BRA) for coping with behavioural and psychological symptoms of dementia. Psychogeriatrics 2025 Jul;25(4):e70061. [doi: 10.1111/psyg.70061] [Medline: 40555407]
29. Burns K, Jayasinha R, Brodaty H. Evaluation of an electronic app developed to assist clinicians in the management of behavioral and psychological symptoms of dementia (BPSD). Int J Human-Computer Interact 2017 Nov 2;33(11):902-910. [doi: 10.1080/10447318.2017.1296075]
30. Gitlin LN, Bouranis N, Kern V, et al. WeCareAdvisor, an online platform to help family caregivers manage dementia-related behavioral symptoms: an efficacy trial in the time of COVID-19. J Technol Behav Sci 2022;7(1):33-44. [doi: 10.1007/s41347-021-00204-8] [Medline: 33786370]
31. Shin Y, Kim SK, Kim Y, Go Y. Effects of app-based mobile interventions for dementia family caregivers: a systematic review and meta-analysis. Dement Geriatr Cogn Disord 2022;51(3):203-213. [doi: 10.1159/000524780] [Medline: 35609526]
32. Jagoda T, Dharmaratne S, Rathnayake S. Informal carers' information needs in managing behavioural and psychological symptoms of people with dementia and related mHealth applications: a systematic integrative review to inform the design of an mHealth application. BMJ Open 2023 May 11;13(5):e069378. [doi: 10.1136/bmjopen-2022-069378] [Medline: 37169501]
33. Baker FA, Pac Soo V, Bloska J, et al. Home-based family caregiver-delivered music and reading interventions for people living with dementia (HOMESIDE trial): an international randomised controlled trial. eClinicalMedicine 2023 Nov;65:102224. [doi: 10.1016/j.eclinm.2023.102224]
34. Gitlin LN, Kales HC, Lyketsos CG. Nonpharmacologic management of behavioral symptoms in dementia. JAMA 2012 Nov 21;308(19):2020. [doi: 10.1001/jama.2012.36918]
35. Cho E, Yang M, Jang J, et al. A mobile app to predict and manage behavioral and psychological symptoms of dementia: development, usability, and users' acceptability. Digit Health 2024 Sep;10:20552076241249277. [doi: 10.1177/20552076241249277]
36. Cho E, Yang M, Kim MJ, Hwang S, Kim E, Cho J. Effectiveness of a mobile app-based individualized non-pharmacological intervention on behavioral and psychological symptoms of dementia in community-dwelling older adults: study protocol for a randomized control trial. J Korean Gerontol Nurs 2024 Aug;26(3):248-256. [doi: 10.17079/jkgn.2024.00521] [Medline: 41584880]
37. Hill AM, Moyle W, Slatyer S, et al. Nurse telephone support for caregivers of older adults at hospital discharge: a randomized clinical trial. JAMA Netw Open 2024 Oct 1;7(10):e2441019. [doi: 10.1001/jamanetworkopen.2024.41019] [Medline: 39453654]
38. Teles S, Ferreira A, Seeher K, Fréel S, Paúl C. Online training and support program (iSupport) for informal dementia caregivers: protocol for an intervention study in Portugal. BMC Geriatr 2020 Jan 8;20(1):10. [doi: 10.1186/s12877-019-1364-z] [Medline: 31914936]
39. Cho E, Kim S, Heo SJ, et al. Machine learning-based predictive models for the occurrence of behavioral and psychological symptoms of dementia: model development and validation. Sci Rep 2023 May 18;13(1):8073. [doi: 10.1038/s41598-023-35194-5] [Medline: 37202454]
40. Choi SH, Na DL, Kwon HM, Yoon SJ, Jeong JH, Ha CK. The Korean version of the neuropsychiatric inventory: a scoring tool for neuropsychiatric disturbance in dementia patients. J Korean Med Sci 2000 Dec;15(6):609-615. [doi: 10.3346/jkms.2000.15.6.609] [Medline: 11194184]
41. Suh GH. Agitated behaviours among the institutionalized elderly with dementia: validation of the Korean version of the Cohen-Mansfield Agitation Inventory. Int J Geriatr Psychiatry 2004 Apr;19(4):378-385. [doi: 10.1002/gps.1097] [Medline: 15065232]
42. Lee Y, Lee YS, Hong GRS. Validation of Korean version of The Cornell Scale for depression in dementia among older adults in long-term care facilities. J Korean Gerontol Nurs 2015;17(3):142-151. [doi: 10.17079/jkgn.2015.17.3.142]
43. Camargos EF, Louzada FM, Nóbrega OT. Wrist actigraphy for measuring sleep in intervention studies with Alzheimer's disease patients: application, usefulness, and challenges. Sleep Med Rev 2013 Dec;17(6):475-488. [doi: 10.1016/j.smrv.2013.01.006] [Medline: 23669093]
44. Hjetland GJ, Nordhus IH, Pallesen S, et al. An actigraphy-based validation study of the sleep disorder inventory in the nursing home. Front Psychiatry 2020;11:173. [doi: 10.3389/fpsyt.2020.00173] [Medline: 32231600]
45. Benca R, Herring WJ, Khandker R, Qureshi ZP. Burden of insomnia and sleep disturbances and the impact of sleep treatments in patients with probable or possible Alzheimer's disease: a structured literature review. J Alzheimers Dis 2022;86(1):83-109. [doi: 10.3233/JAD-215324] [Medline: 35001893]
46. Cheon H, Song JA, Kim J, Jung S. The competence scale in managing behavioral and psychological symptoms of dementia (CS-MBPSD) for family caregivers: instrument development and cross-sectional validation study. Int J Nurs Stud 2022 May;129:104205. [doi: 10.1016/j.ijnurstu.2022.104205] [Medline: 35344838]
47. Kang Y, Jahng S, Kim S, Korean Dementia Association. Korean Mini-Mental State Examination, 22th edition: Insight of Psychology; 2020.
48. Won CW, Rho YG, Kim SY, Cho BR, Lee YS. The validity and reliability of the Korean Activities of Daily Living (K-ADL) scale. J Korean Geriatr Soc 2002;6(2):98-106 [FREE Full text]

XSL•FO
RenderX

49. Nunes PV, Schwarzer MC, Leite REP, et al. Neuropsychiatric inventory in community-dwelling older adults with mild cognitive impairment and dementia. J Alzheimers Dis 2019;68(2):669-678. [doi: 10.3233/JAD-180641] [Medline: 30856109]
50. Braly T, Muriathiri D, Brown JC, Taylor BM, Boustani MA, Holden RJ. Technology intervention to support caregiving for Alzheimer's disease (I-CARE): study protocol for a randomized controlled pilot trial. Pilot Feasibility Stud 2021 Jan 11;7(1):23. [doi: 10.1186/s40814-020-00755-2] [Medline: 33431027]
51. Morganti F, Soli A, Savoldelli P, Belotti G. The neuropsychiatric inventory-diary rating scale (NPI-Diary): a method for improving stability in assessing neuropsychiatric symptoms in dementia. Dement Geriatr Cogn Dis Extra 2018;8(3):306-320. [doi: 10.1159/000490380] [Medline: 30323831]
52. Pozzi FE, Calì L, D'Antonio F, et al. BPSDiary study protocol: a multi-center randomized controlled trial to compare the efficacy of a BPSD diary vs. standard care in reducing caregiver's burden. Front Dement 2023;2:1301280. [doi: 10.3389/frdem.2023.1301280] [Medline: 39081982]
53. Rangseekajee P, Aphisitphinyo S, Paholpak P, et al. Mobile application for monitoring behavioral and psychological symptoms of dementia in patients with moderate to severe dementia. Geriatr Gerontol Int 2021 Jun;21(6):472-477. [doi: 10.1111/ggi.14164] [Medline: 33851502]
54. James IA, Gray K, Moniz-Cook E, Lee K, Reichelt K, Richardson J. Behavioural and psychological symptoms of dementia: a new framework for holistic understanding and non-pharmacological management. BJPsych Advances 2022 Jan;28(1):11-20. [doi: 10.1192/bja.2021.12]
55. Halpern R, Seare J, Tong J, Hartry A, Olaoye A, Aigbogun MS. Using electronic health records to estimate the prevalence of agitation in Alzheimer disease/dementia. Int J Geriatr Psychiatry 2019 Mar;34(3):420-431. [doi: 10.1002/gps.5030] [Medline: 30430642]
56. Eikelboom WS, Singleton EH, van den Berg E, et al. The reporting of neuropsychiatric symptoms in electronic health records of individuals with Alzheimer's disease: a natural language processing study. Alzheimers Res Ther 2023 May 12;15(1):94. [doi: 10.1186/s13195-023-01240-7] [Medline: 37173801]
57. Algase DL, Beck C, Kolanowski A, et al. Need-driven dementia-compromised behavior: an alternative view of disruptive behavior. Am J Alzheimers Dis (Columbia) 1996 Nov;11(6):10-19. [doi: 10.1177/153331759601100603]
58. Scales K, Zimmerman S, Miller SJ. Evidence-based nonpharmacological practices to address behavioral and psychological symptoms of dementia. Gerontologist 2018 Jan 18;58(suppl_1):S88-S102. [doi: 10.1093/geront/gnx167] [Medline: 29361069]
59. Wang G, Albayrak A, van der Cammen TJM. A systematic review of non-pharmacological interventions for BPSD in nursing home residents with dementia: from a perspective of ergonomics. Int Psychogeriatr 2019 Aug;31(8):1137-1149. [doi: 10.1017/S1041610218001679] [Medline: 30334500]
60. Husebo BS, Heintz HL, Berge LI, Owoyemi P, Rahman AT, Vahia IV. Sensing technology to monitor behavioral and psychological symptoms and to assess treatment response in people with dementia. a systematic review. Front Pharmacol 2019;10:1699. [doi: 10.3389/fphar.2019.01699] [Medline: 32116687]
61. Anderson MS, Homdee N, Bankole A, et al. Behavioral interventions for Alzheimer's management using technology: home-based monitoring. Curr Geri Rep 2020 Jun;9(2):90-100. [doi: 10.1007/s13670-020-00312-y]
62. Kerssens C, Kumar R, Adams AE, et al. Personalized technology to support older adults with and without cognitive impairment living at home. Am J Alzheimers Dis Other Demen 2015 Feb;30(1):85-97. [doi: 10.1177/1533317514568338] [Medline: 25614507]
63. Park E, Park H, Kim EK. The effect of a comprehensive mobile application program (CMAP) for family caregivers of home‐dwelling patients with dementia: a preliminary research. Jpn J Nurs Sci 2020 Oct;17(4):e12343. [doi: 10.1111/jjns.12343]
64. Sousa L, Neves MJ, Moura B, Schneider J, Fernandes L. Music-based interventions for people living with dementia, targeting behavioral and psychological symptoms: a scoping review. Int J Geriatr Psychiatry 2021 Nov;36(11):1664-1690. [doi: 10.1002/gps.5568] [Medline: 34097789]
65. Pu Y, Zhang G, You S, et al. Reminiscence therapy delivery formats for older adults with dementia or mild cognitive impairment: a systematic review and network meta-analysis. Int J Nurs Stud 2025 Aug;168:105085. [doi: 10.1016/j.ijnurstu.2025.105085]
66. Conway A, Ryan A, Harkin D, Mc Cauley C, Goode D. A review of the factors influencing adoption of digital health applications for people living with dementia. Digit Health 2023;9:20552076231162985. [doi: 10.1177/20552076231162985] [Medline: 36937696]
67. Guicheteau J, Seven Ü, Boes J, et al. Characteristics of complex non-pharmacological interventions for cognitive stimulation in people with mild to moderate dementia in nursing homes: a systematic review. Geriatr Nurs 2025;66(Pt A):103586. [doi: 10.1016/j.gerinurse.2025.103586] [Medline: 40829529]
68. Ramirez M, Duran MC, Pabiniak CJ, et al. Family caregiver needs and preferences for virtual training to manage behavioral and psychological symptoms of dementia: interview study. JMIR Aging 2021 Feb 10;4(1):e24965. [doi: 10.2196/24965] [Medline: 33565984]
69. Larson JF. Network-centric digital development in Korea: origins, growth and prospects. Telecomm Policy 2017 Nov;41(10):916-930. [doi: 10.1016/j.telpol.2017.03.007]

XSL•FO
RenderX

Abbreviations

BPSD: behavioral and psychological symptoms of dementia
NPI: Neuropsychiatric Inventory

Edited by L Buis; submitted 22.Jun.2025; peer-reviewed by D Liu, I Adefolaju, KW Tay, ME Heidari; revised version received 14.Nov.2025; accepted 08.Mar.2026; published 07.Apr.2026.

Please cite as:
Cho E, Yang M, Hwang S, Kim E, Cho J, Kim MJ
A Mobile App–Based Individualized Nonpharmacological Intervention for Behavioral and Psychological Symptoms in Dementia: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e79469
URL: https://mhealth.jmir.org/2026/1/e79469
doi:10.2196/79469

© Eunhee Cho, Minhee Yang, Sinwoo Hwang, Eunkyo Kim, Jungwon Cho, Min Jung Kim. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 7.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Contribution of Longitudinal Mobile Health Measures in the Dynamic Track of Patients With Major Depressive Disorder: Multiple Centers, Prospective Cohort Study Using Functional Data Analysis and Machine Learning

Rou Zhong[1,2*], PhD; Nanxi Li[1,2*], MD, PhD; Le Xiao[1,2], MD, PhD; Lei Feng[1,2], PhD; Yuan Feng[1,2], MM; Gang Wang[1,2], MD; Xuequan Zhu[1,2], MD

[1]Beijing Key Laboratory of Intelligent Drug Research and Development for Mental Disorders, National Clinical Research Center for Mental Disorders, National Center for Mental Disorders, Beijing Anding Hospital, Capital Medical University, No. 5 Ankang Hutong, Xicheng District, Beijing, China

[2]Advanced Innovation Center for Human Brain Protection, Capital Medical University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Xuequan Zhu, MD
Beijing Key Laboratory of Intelligent Drug Research and Development for Mental Disorders, National Clinical Research Center for Mental Disorders, National Center for Mental Disorders, Beijing Anding Hospital, Capital Medical University, No. 5 Ankang Hutong, Xicheng District, Beijing, China

Abstract

Background: Continuous follow-up for patients with major depressive disorder (MDD) is essential for treatment decisions and a better prognosis. There remains limited evidence regarding the critical issue of depression variation trajectory prediction using mobile health (mHealth) measures. Moreover, the temporal dynamics of mHealth measures have not been fully modeled in previous studies, and the poor patient adherence to mHealth records poses great challenges to the dynamic feature modeling.

Objective: This study aimed to examine the contribution of mHealth measures in predicting depression variation trajectory for patients with MDD, with full consideration of the temporal dynamics of mHealth measures.

Methods: A total of 229 patients with MDD from a multiple-center, prospective cohort were included. A 12-week follow-up was conducted involving the collection of the Hamilton Depression Rating Scale (HAMD-17), along with patient-reported outcomes (Immediate Mood Scaler and Altman Self-Rating Mania Scale) via mobile devices and sleep duration through wearable wristbands. We used functional data analysis to extract dynamic features from the sparse mHealth records, rather than aggregating the data to a single scalar summary measure through collapsing over time. Subsequently, 3 machine learning models were applied to predict the depression variation trajectory classes based on the baseline characteristics and these extracted dynamic features.

Results: Based on the variation of HAMD-17 scores within 12 weeks, the participants were labeled into 4 classes through the *k*-means algorithm. The classes included stable decline (n=93), fluctuate decline (n=44), fast decline (n=60), and delayed and fluctuate (n=32), in light of the shape of depression trajectories. With both baseline features and dynamic features of the mHealth measures, accuracy rates for the overall data were 54.35%, 60.87%, and 56.52%, for the stable decline patients were 78.95%, 84.21%, and 73.68%, for the nonstable decline patients were 59.26%, 62.96%, and 70.37% based on the 3 machine learning models, respectively. The results were significantly superior to the prediction obtained without mHealth measures (with an overall accuracy below 50%) and only showed a marginal reduction in accuracy relative to the ideal prediction with assessment obtained from clinical visits. Moreover, in the construction of the most accurate prediction model, dynamic features of the Immediate Mood Scaler, the Altman Self-Rating Mania Scale, and sleep duration emerged as the most influential predictors, ranking first, third, and fourth, respectively, in terms of their relative importance.

Conclusions: Longitudinal mHealth measures show potential in depression variation trajectory monitoring for patients with MDD even under poor patient adherence. Our work provides practical help in alleviating the follow-up burden for patients with MDD and validates the effectiveness of mHealth measures in clinical applications.

(JMIR Mhealth Uhealth 2026;14:e81397) doi:10.2196/81397

KEYWORDS

major depressive disorder; mobile health; patient-reported outcome; digital phenotype; functional data analysis; machine learning

XSL•FO
RenderX

Introduction

Depression is a common mental disorder with a high lifetime prevalence, ranked as one of the most crucial causes of disease burden worldwide [1,2]. Despite the availability of evidence-based treatments, including antidepressants and psychotherapy, more than 60% of patients with major depressive disorder (MDD) exhibit inadequate response with their first treatment, often requiring either an alternative antidepressant or adjunctive treatment [3]. The early reduction of depressive symptoms observed in the first 2 - 4 weeks of antidepressant treatment has been proposed as a potential predictor of treatment success, possibly reflecting psychological resilience in patients with MDD [4], and the predictive efficacy of early response varies across antidepressants. Leveraging time-varying individual characteristics, including demographics, clinical phenotypes, and social functioning [5], holds promise for refining treatment algorithms. An important question raised in this context is how to capture the temporal dynamics of depression severity, which remains a persistent challenge in both clinical practice and research.

It is burdensome for both patients and doctors to conduct continuous and frequent face-to-face psychometric evaluations. Fortunately, with the increasing use of smartphones and wearable devices in modern life, it has become possible to achieve real-time monitoring of physical activity, sleep, and self-assessment for patients with MDD through mobile health (mHealth) technologies [6-8]. The mHealth measures obtained from smartphones and wearable devices can be summarized into 2 types, which are self-reported data (patient-reported outcomes [PROs] such as questionnaire scores) and passively collected data (digital phenotype data such as sleep data, activity data, and phone usage data) [9]. Smartphones are superior in capturing mood changes, while wearable devices are effective in daily activity records [10]. The incorporation of records from both smartphones and wearable devices is beneficial for clinical evaluation and provides a new perspective for depression monitoring [7].

For the potential of self-reported data recorded by smartphone in the depressive symptoms monitoring, Torous et al [11] examined the correlation between the Patient Health Questionnaire-9 scores recorded by smartphone and paper for the patients with MDD, demonstrating the capability of digital depressive symptom monitoring. Moreover, Goltermann et al [9] also revealed the validity of smartphone-based monitoring of depressive symptoms through the analysis of the agreement between smartphone-based and non–smartphone-based assessments. These studies highlight the effectiveness of smartphone-based self-reported data in tracking depressive symptoms.

Digital phenotype data have been widely applied in mental health studies and have shown the ability to strengthen the management of depressive symptoms [12,13]. As sleep problems occur frequently in patients with MDD [14], an increasing number of studies focused on the relationship between remote sleep records and depression symptoms [15,16]. Moreover, Lim et al [17] used sleep-wake pattern data exclusively in the prediction of mood episodes. Further, compared with other types of digital phenotype data, sleep data are easier to collect and less affected by other interfering factors [18]. These findings revealed the potential of remote sleep records in enhancing depression monitoring and prediction.

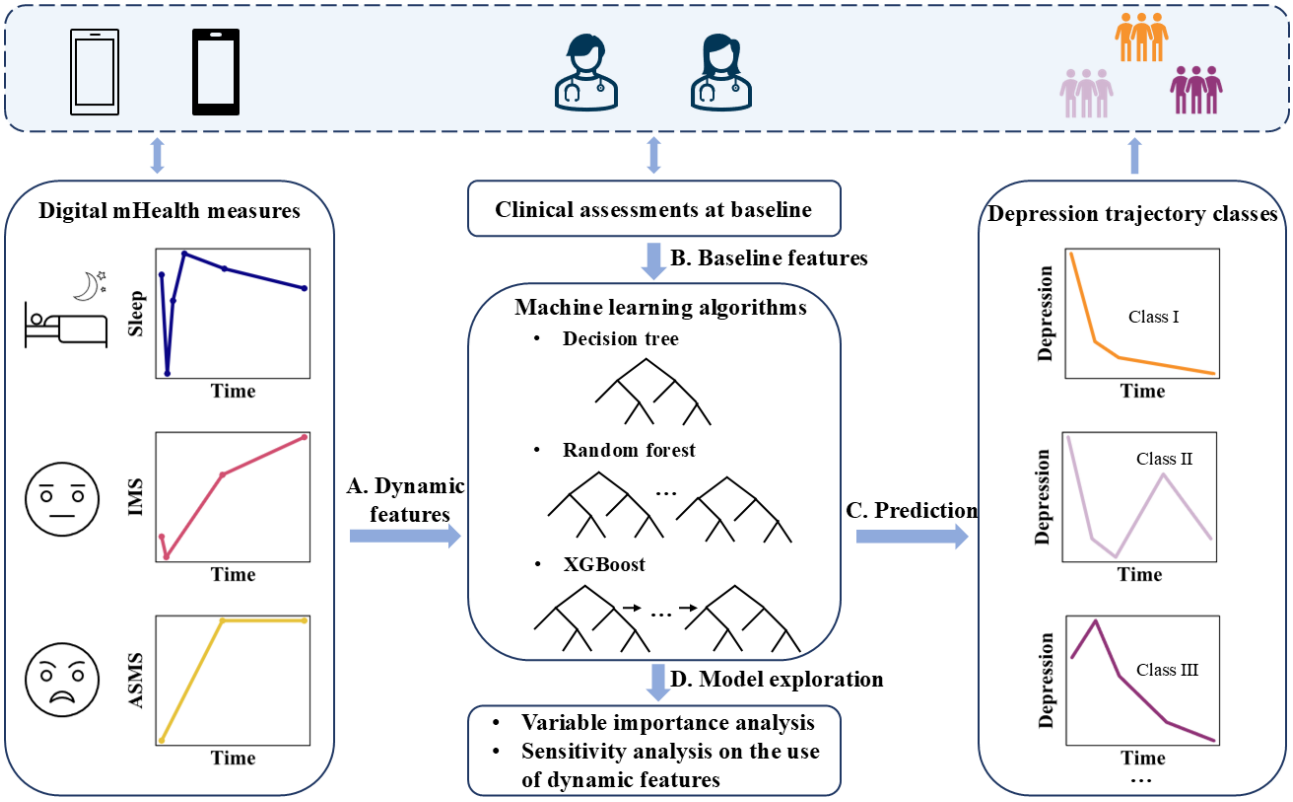
As depressive symptoms are constantly evolving and show heterogeneity over time, capturing the temporal dynamics of depression severity is essential [19]. However, given the great importance of depression dynamics, studies on the prediction of depression variation trajectory through mHealth measures were scarce. Bai et al [20] discussed the prediction of mood swings for patients with MDD using passively collected data from smartphones and wearable devices. Price et al [21] used digital phenotype data in the prediction of depression symptom variability. The definition of depression variation was still vague in existing mHealth literature and lack intuitiveness and clinical significance. And there is a growing demand for ascertaining the contribution of mHealth measures in depression variation trajectory prediction.

Additionally, fusing the temporal dynamic features of smartphone- and wearable device–based data in the prediction of depression trajectory is also crucial due to the dynamic nature of personalized behaviors [22-24]. However, the records from smartphones and wearable devices are always sparse and irregular, owing to poor study adherence and unsatisfactory patient compliance [25-27]. Patients may accidentally switch off the app or forget to report their states. The sparsity of the mHealth records poses additional challenges to the capture of dynamic features for both self-reported data and digital phenotype data, which calls for the clinical application of novel statistical methods.

The primary objective of this study is to examine the contribution of longitudinal mHealth measures in predicting depression variation trajectory for patients with MDD. By applying a functional data analysis method to the longitudinal records of sleep duration, Immediate Mood Scaler (IMS) scores, and Altman Self-Rating Mania Scale (ASMS) scores, the dynamic features of mHealth measures (including both PROs and digital phenotypes) were obtained (Figure 1A). The depression trajectory class was predicted (Figure 1C) based on the dynamic features of mHealth measures and clinical assessments at baseline (Figure 1B) through machine learning algorithms. The investigation of classification accuracy with and without the longitudinal records of PROs and digital phenotypes, as well as the variable importance analysis and sensitivity analysis (Figure 1D), enables the detection for the contribution of the dynamic features of mHealth measures in depression variation trajectory prediction for patients with MDD.

XSL•FO
RenderX

Figure 1. Study framework for detecting the contribution of longitudinal mHealth measures in depression variation trajectory prediction. ASMS: Altman Self-Rating Mania Scale; IMS: Immediate Mood Scaler; mHealth: mobile health.



Methods

Ethical Considerations

This study used data from an epidemiological, noninterventional, cohort study conducted between February 2019 and April 2020 at 4 psychiatric hospitals or units in general hospitals in China. The study was approved by the Independent Medical Ethics Committee Board of Beijing Anding Hospital and the other 3 sites (ethical approval no. 2018-119-201917FS-2). All patients provided written informed consent to participate in the study. To ensure confidentiality, each participant was assigned an identification number. Moreover, participants received ¥100 (US $15.5) for each follow-up visit.

Participants

Study inclusion and exclusion criteria have been reported [20]. Generally, inclusion criteria for the study included outpatients (1) aged 18 - 60 years and (2) diagnosed with MDD according to the *Diagnostic and Statistical Manual of Mental Disorders* (fourth edition) criteria. Exclusion criteria included patients who had axis I primary psychiatric diagnosis other than MDD or had substance abuse. Patients in this cohort were followed up at 2, 4, 8, and 12 weeks.

Measures

Demographics and Clinical Characteristics

The participants' basic social-demographic and clinical characteristics were collected through the structured questionnaire designed for this study.

Clinical-Rated Outcomes

The Hamilton Depression Rating Scale (HAMD-17) [28] was used to measure depressive symptoms. The total HAMD-17 score ranges from 0 to 52, with a higher score indicating greater severity of depression.

The Hamilton Anxiety Scale (HAMA) [29] was used to measure anxiety symptoms. The total HAMA score ranges from 0 to 56, and a higher score indicates the severity of anxiety.

The Sheehan Disability Scale [30] was used to measure the social functional impairment due to MDD that interferes with work/school, social life/leisure activities, and family life/home responsibilities. The participants are invited to rate the extent to which each domain is impaired by their symptoms using a 10-point scale (0=not at all impaired, 1 - 3=mildly impaired, 4 - 6=moderately impaired, 7 - 9=markedly impaired, and 10=extremely impaired).

Patient-Reported Outcomes

The Chinese version of the IMS was used to assess current mood symptoms [31,32]. It is composed of 22 questions about the respondent's state of mood. Each item was rated on a 7-point scale with complementary antonyms positioned at opposite ends. A higher summed score of the 22 items reflects a higher level of severity of the patient's depressive symptoms. Upon enrollment, participants were instructed to complete the IMS scale on their smartphones on a daily basis.

The ASMS was administered as a self-assessment tool to evaluate the intensity of manic symptoms [33]. The ASMS comprises five key symptom domains: (1) elevated mood, (2) heightened self-esteem, (3) diminished need for sleep, (4) rapid

XSL•FO
RenderX

or pressured speech, and (5) psychomotor agitation. Patients rated each item by choosing from 5 response options, ranging from 0 (symptom not present) to 4 (highly severe). Participants were instructed to complete weekly administrations of the ASMS using their smartphones.

Digital Phenotype

Participants were instructed and trained to wear wristbands during sleep for sleep-related data collection during the 12-week follow-up. The data collected would first be stored locally and uploaded to our server when the user connects with the wristband using the Mood Mirror app [20]. Key sleep characteristics, particularly total sleep duration and slow-wave sleep time, were systematically extracted and incorporated into the analytical model.

Statistical Analysis

Clustering Analysis

In this paper, we mainly focused on the variation of depression severity at 5 consecutive visits within 12 weeks. Therefore, the class labels for the patients were expected to be related to the variation of depression severity. To determine the class label for each patient, the *k*-means clustering algorithm was used. The *k*-means clustering algorithm is an unsupervised learning method that can divide samples into *k* clusters based on the cluster means. We computed the difference of HAMD-17 scores between each adjacent visit (week 2 to week 0, week 4 to week 2, week 8 to week 4, and week 12 to week 8) for each patient and input these features into the clustering analysis. That means patients who had a similar changing trend of HAMD-17 score would be assigned to the same cluster. We implemented the *k*-means algorithm through the kmeans function in R statistical software, version 4.4.1. The number of clusters was selected by considering both the silhouette coefficients and the clinical significance of the clustering results. The silhouette coefficients were computed by the silhouette function of the cluster R package, version 2.1.6. After comparing the results obtained by k=2,3,4,5, we chose the cluster number as 4 and further defined the class labels as stable decline, fluctuate decline, fast decline, and delayed and fluctuate.

Feature Extraction by Functional Principal Component Analysis

Within the 12 weeks of monitoring, the median numbers of days with a record of IMS score, ASMS score, and sleep duration were 5 (IQR 4-8), 4 (IQR 3-4), and 50 (IQR 28-63), respectively. Figure S1 in Multimedia Appendix 1 shows the heatmap of the observation size for each participant in the record of IMS scores, ASMS scores, and sleep duration. It is shown that the records for all the mHealth measures were severely sparse, especially for the IMS scores and ASMS scores, due to poor patient adherence.

To extract the dynamic features of the IMS scores, ASMS scores, and sleep duration from baseline to week 12 for each patient, functional data analysis methods were used [34]. Through functional data analysis methods, the temporally dependent PROs and digital phenotype records have not to be collapsed into a single value, and the temporal dynamic information can be retained [35]. In specific, to adapt to the sparseness of the data caused by poor patient adherence, a method called principal components analysis through conditional expectation (PACE) was used [36]. The PACE approach is a commonly used functional principal component (FPC) analysis method that is designed for sparsely and irregularly observed functional data. This approach uses a local linear smoother for the estimation of the mean function and the covariance function, and then it computes the FPC scores based on conditional expectations. The PACE approach borrows the strength of the entire sample in the computation, so it can make full use of the information from the dataset. By using the fdapace R package with version 0.6.0, we implemented the PACE approach for the records of IMS score, ASMS score, and sleep duration, respectively, and the estimated first FPC scores were treated as the extracted dynamic features.

Machine Learning Models

To predict the variation trajectory class of depression severity for the patients, various machine learning methods were used, such as decision tree, random forest, and XGBoost (Extreme Gradient Boosting). In specific, a decision tree completes classification by recursively selecting the best decision rules to partition the data, and the decision paths can be summarized through a tree structure. Random forest and XGBoost are both ensemble learning methods, but with different ensemble patterns. Random forest depends on many independent decision trees, while XGBoost trains a sequence of decision trees in an iterative way. To alleviate the overfit problems for the random forest and XGBoost methods, the node size (minimum size of terminal nodes) was selected as 10 for the random forest model, and the maximum depth of a tree was set as 3 for the XGBoost model. We conducted the above machine learning methods using the rpart, randomForest, and xgboost R packages with versions 4.1.24, 4.7‐1.2, and 1.7.11.1, respectively.

Predictive Performance Assessment

To find out how the records of IMS score, ASMS score, and sleep duration contribute to the classification of depression variation, we compared the classification results obtained from different input information. In specific, we considered the prediction based on 4 types of input scenarios. Scenario 1 uses only baseline features; scenario 2 integrates baseline features with the FPC scores of the PROs and digital phenotype records; scenario 3 combines baseline features and HAMD-17 and HAMA scores assessed at week 2; scenario 4 incorporates baseline features, FPC scores of the PROs and digital phenotype records, and HAMD-17 and HAMA scores at week 2. Here, scenario 2 was our central focus, while scenario 1 was used to assess the prediction without mHealth measures. Moreover, scenarios 3 and 4 were introduced to simulate the ideal cases with assessments obtained from clinical visits. By comparing scenario 2 with scenarios 3 and 4, we can shed light on how the predictive performance in scenario 2 approaches the ideal cases.

Moreover, the performance was evaluated by the classification accuracy for the overall data, stable decline patients, and nonstable decline patients. The 95% CIs of the classification accuracy were obtained through the Wilson method based on Hmisc R package. The dataset was randomly split into a training

XSL•FO
RenderX

set (80%) and a test set (20%). Given the class imbalance, we randomly oversampled the fluctuate decline, fast decline, and delayed and fluctuate classes to make their class size 80% of the size of the stable decline. The oversampling procedure was implemented within the training set, ensuring a strict separation between the training set and test set.

Variable Importance

The importance of the considered predictive factors was assessed for the decision tree, random forest, and XGBoost to illustrate the contribution of mHealth measures in the model construction. Specifically, Gini impurity reduction [37] was used to compute variable importance for the decision tree [38] and random forest models [39], while the Gain metric [40] was used for the XGBoost model. For the implementation, the rpart package, randomForest package, and xgboost package in R were used, respectively. Moreover, we also conducted a permutation test to estimate the significance of variable importance for the random forest model through permuting the response variable 1000 times [41]. A *P* value less than .05 indicates a significant variable contribution. The rfPermute R package with version 2.5.5 was used for the computation.

Sensitivity Analysis

Further, to investigate the advantage of the functional data analysis method, we also evaluate the predictive performance when the mean values of IMS score, ASMS score, and sleep duration were used for model construction, rather than their FPC scores. The dynamic features of the PROs and digital phenotype were disrupted because the records were roughly summarized as a single value. Hence, it is hypothesized that using the FPC scores would yield a more accurate prediction.

Results

Characteristics of Participants

A total of 229 patients with MDD were included in the study (Figure 2), with 78 males (34.1%), 151 females (65.9%), and a mean age of 28.7 years. According to the variation of HAMD-17 scores, the participants were labeled as stable decline, fluctuate decline, fast decline, and delayed and fluctuate, respectively. The number of classes was selected as k=4 with the consideration of both silhouette coefficients (0.182) and clinical significance. Clinical significance refers to the capacity of the chosen clusters to distinguish 4 clinically recognizable patterns of symptom reduction over 12 weeks, categorizing the longitudinal patient responses into stable decline, fluctuate decline, fast decline, and delayed and fluctuate trajectories.

XSL•FO
RenderX

Figure 2. Flowchart illustrating participant inclusion. With 358 patients enrolled in the multiple centers, prospective cohort, 229 patients were finally included in our study. mHealth: mobile health.

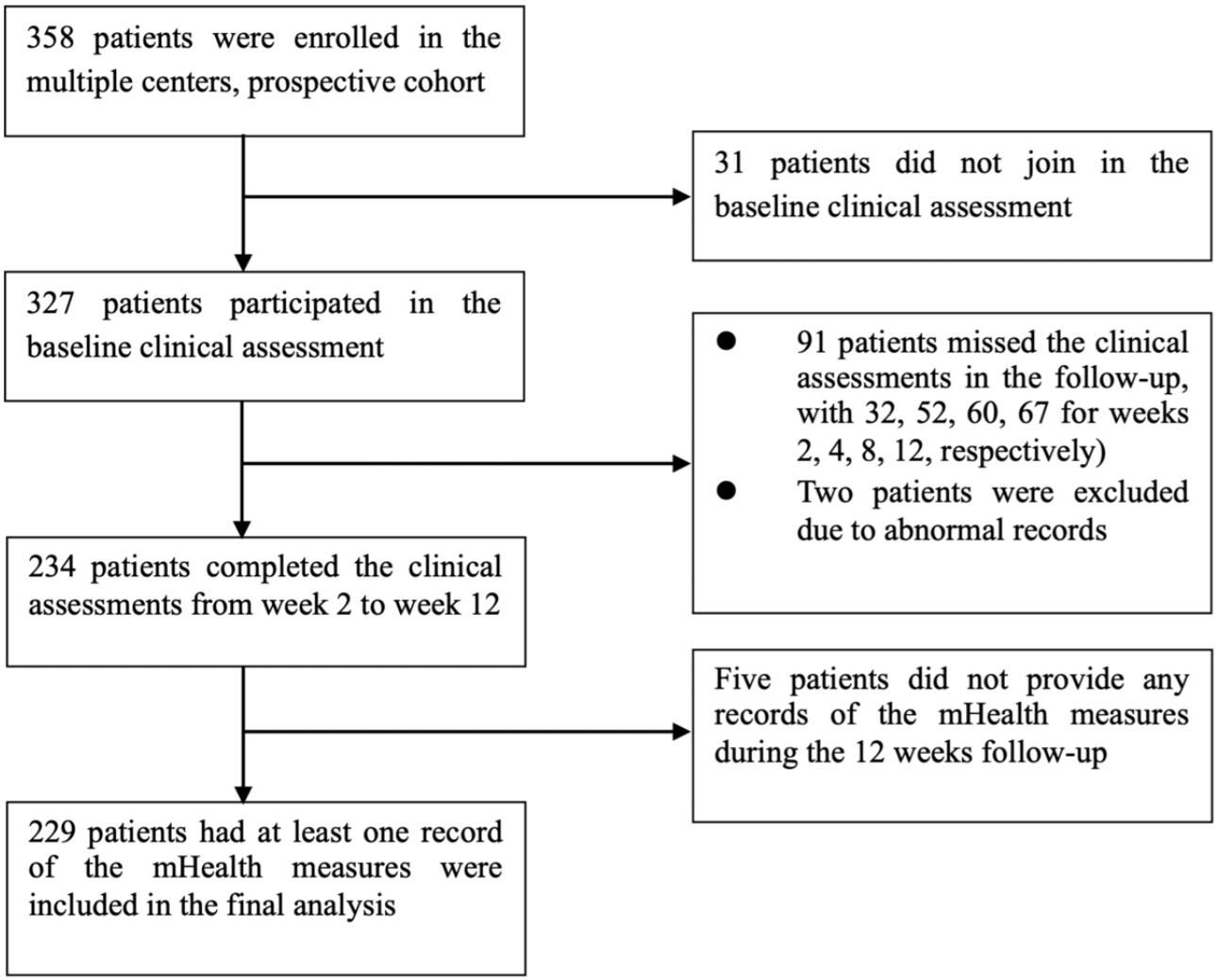


Figure 3 shows the variation of the mean HAMD-17 scores from baseline to week 12 for each class (refer to Multimedia Appendix 1 for HAMD-17 scores figures when k=2,3,5), and Figure S4 in Multimedia Appendix 1 shows that patients within the same class exhibited the same temporal trend of depression severity. In specific, the stable decline class contained 93 (40.6%) patients with MDD, who were more likely to improve steadily during the study. There were 44 (19.2%) patients with MDD in the fluctuate decline class. It is shown that the HAMD-17 scores of participants in the fluctuate decline class might go through an upward fluctuation before week 2 and begin to decrease afterwards. The fast decline class included 60 (26.2%) patients with MDD, and the HAMD-17 scores of these patients showed a rapid decline before week 2. Moreover, 32 (14%) patients with MDD were assigned to the delayed and fluctuate class. It is evident that the HAMD-17 scores of these patients showed a delayed decline and were prone to exhibit a drastic rise at week 8. The symptomatic rebound at week 8 indicates a critical clinical juncture for evaluating treatment efficacy in depression [42]. This pattern suggests a unique phenotype of therapeutic instability where early progress fails to persist. Such findings emphasize the importance of extended clinical monitoring to identify patients at risk of midstage relapse.

XSL•FO
RenderX

Figure 3. The mean HAMD-17 scores from baseline to week 12 for patients in stable decline, fluctuate decline, fast decline, and delayed and fluctuate classes. HAMD-17: Hamilton Depression Rating Scale.

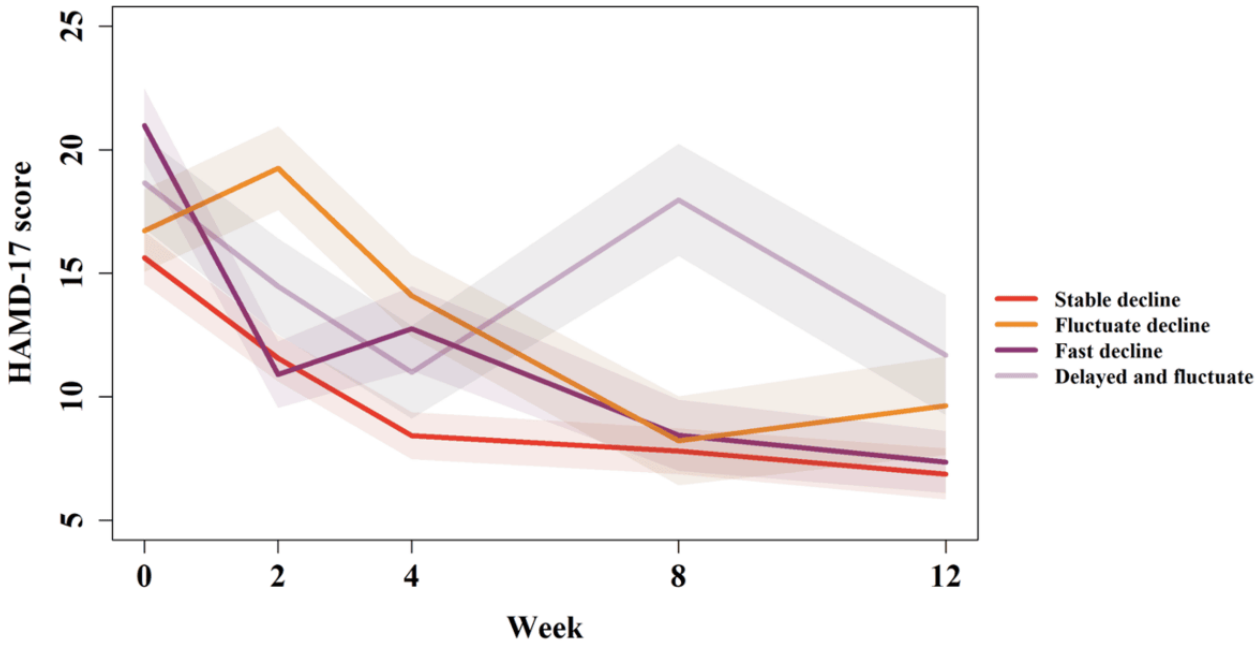


Table 1 reports the demographic and clinical characteristics of the patients with MDD in each class. It can be observed that the HAMD-17 and HAMA scores, as well as the changes in HAMD-17 scores, were significantly different among these 4 classes for each week. In addition, the average IMS scores during the follow-up were also significantly different among these 4 classes. Moreover, there was no significant difference in sex, age, BMI, and baseline Quality of Life Enjoyment and Satisfaction Questionnaire—Short Form, Sheehan Disability Scale, IMS, and ASMS scores among these 4 classes. Refer to Multimedia Appendix 1 for the figures of mean curves of IMS scores, ASMS scores, and sleep duration for each class, and for the demographic characteristics of patients in each participating center.

XSL•FO
RenderX

Table . Demographic and clinical characteristics of patients with major depressive disorder in the study.

Features	All (n=229)	Stable decline (n=93)	Fluctuate decline (n=44)	Fast decline (n=60)	Delayed and fluctuate (n=32)	Statistics[a] (*df*=3)	*P* value
Sex, n (%)						1.74	.63
Male	78 (34.1)	29 (31.2)	17 (38.6)	23 (38.3)	9 (28.1)		
Female	151 (65.9)	64 (68.8)	27 (61.4)	37 (61.7)	23 (71.9)		
Age (years), mean (SD)	28.7 (8.9)	30.4 (10.2)	27.0 (7.4)	28.4 (8.4)	27.0 (6.3)	3.73	.29
Baseline BMI, mean (SD)	22.2 (3.9)	22.3 (4.0)	21.8 (3.9)	22.2 (3.7)	22.6 (3.9)	1.25	.74
Baseline HAMD-17[b], mean (SD)	17.7 (5.9)	15.6 (5.2)	16.7 (5.7)	21.0 (5.9)	18.7 (5.3)	28.33	<.001
Baseline HAMA[c], mean (SD)	15.6 (6.7)	13.8 (5.5)	15.8 (7.0)	17.8 (7.6)	16.7 (6.5)	10.84	.01
Baseline Q-LES-Q-SF[d], mean (SD)	44.9 (9.4)	46.1 (10.1)	43.1 (7.9)	44.9 (9.7)	43.8 (8.9)	4.65	.20
Baseline SDS[e], mean (SD)	13.9 (6.5)	13.0 (6.7)	13.5 (5.4)	14.6 (6.8)	15.4 (6.6)	4.13	.25
Work/school, mean (SD)	5.0 (2.7)	4.9 (2.8)	4.8 (2.5)	5.1 (2.7)	5.3 (2.8)	1.21	.75
Social life/leisure activities, mean (SD)	4.8 (2.5)	4.4 (2.6)	4.7 (2.2)	5.2 (2.5)	5.2 (2.3)	4.75	.19
Family life/home responsibilities, mean (SD)	4.1 (2.6)	3.8 (2.7)	4.0 (2.3)	4.3 (2.8)	4.9 (2.4)	5.46	.14
Baseline IMS[f], mean (SD)	71.9 (25.8)	73.4 (24.5)	64.0 (24.1)	76.3 (29.3)	70.1 (23.0)	6.01	.11
Baseline ASMS[g], mean (SD)	2.7 (2.3)	2.8 (2.4)	2.1 (1.6)	3.1 (2.4)	2.1 (2.2)	7.02	.07
HAMD-17 at week 2, mean (SD)	13.3 (6.0)	11.6 (4.6)	19.2 (5.7)	10.9 (5.3)	14.5 (5.6)	54.45	<.001
HAMD-17 at week 4, mean (SD)	11.0 (6.0)	8.4 (4.7)	14.1 (5.6)	12.8 (6.7)	11.0 (5.4)	35.46	<.001
HAMD-17 at week 8, mean (SD)	9.5 (6.4)	7.8 (4.6)	8.2 (6.1)	8.4 (5.7)	18.0 (6.6)	47.51	<.001
HAMD-17 at week 12, mean (SD)	8.2 (5.9)	6.9 (5.0)	9.6 (6.7)	7.4 (5.0)	11.7 (7.0)	15.91	.001
HAMD-17 changes up to week 2, mean (SD)	4.4 (5.6)	4.1 (3.1)	–2.5 (3.7)	10.1 (4.1)	4.2 (4.6)	133.71	<.001
HAMD-17 changes up to week 4, mean (SD)	6.7 (6.0)	7.2 (4.8)	2.6 (5.8)	8.2 (7.0)	7.7 (5.3)	26.22	<.001

XSL•FO
RenderX

Features	All (n=229)	Stable decline (n=93)	Fluctuate decline (n=44)	Fast decline (n=60)	Delayed and fluctuate (n=32)	Statistics[a] (*df*=3)	*P* value
HAMD-17 changes up to week 8, mean (SD)	8.2 (6.8)	7.8 (5.1)	8.5 (7.4)	12.5 (5.7)	0.7 (5.5)	63.66	<.001
HAMD-17 changes up to week 12, mean (SD)	9.5 (6.7)	8.7 (5.5)	7.1 (7.7)	13.6 (5.8)	7.0 (6.5)	34.21	<.001
HAMA at week 2, mean (SD)	11.9 (6.7)	10.5 (5.5)	16.3 (7.0)	10.3 (6.4)	13.1 (7.4)	25.21	<.001
HAMA at week 4, mean (SD)	10.5 (6.3)	8.3 (5.2)	13.3 (6.2)	11.8 (6.8)	10.3 (6.3)	23.89	<.001
HAMA at week 8, mean (SD)	8.6 (6.6)	7.0 (5.3)	7.2 (5.5)	8.2 (5.5)	15.6 (8.4)	29.51	<.001
HAMA at week 12, mean (SD)	7.8 (6.4)	6.8 (5.5)	9.1 (7.9)	6.9 (5.4)	10.8 (7.2)	10.31	.02
Average IMS during follow-up, mean (SD)	76.8 (18.8)	79.6 (16.2)	70.9 (15.0)	79.4 (22.0)	72.0 (22.2)	12.29	.006
Average ASMS during follow-up, mean (SD)	3.3 (1.9)	3.5 (1.8)	2.9 (1.5)	3.5 (2.3)	2.8 (1.8)	4.77	.19

[a]Chi-square test for sex and Kruskal-Wallis rank sum test for other features.

[b]HAMD-17: Hamilton Depression Rating Scale.

[c]HAMA: Hamilton Anxiety Scale.

[d]Q-LES-Q-SF: Quality of Life Enjoyment and Satisfaction Questionnaire—Short Form.

[e]SDS: Sheehan Disability Scale.

[f]IMS: Immediate Mood Scaler.

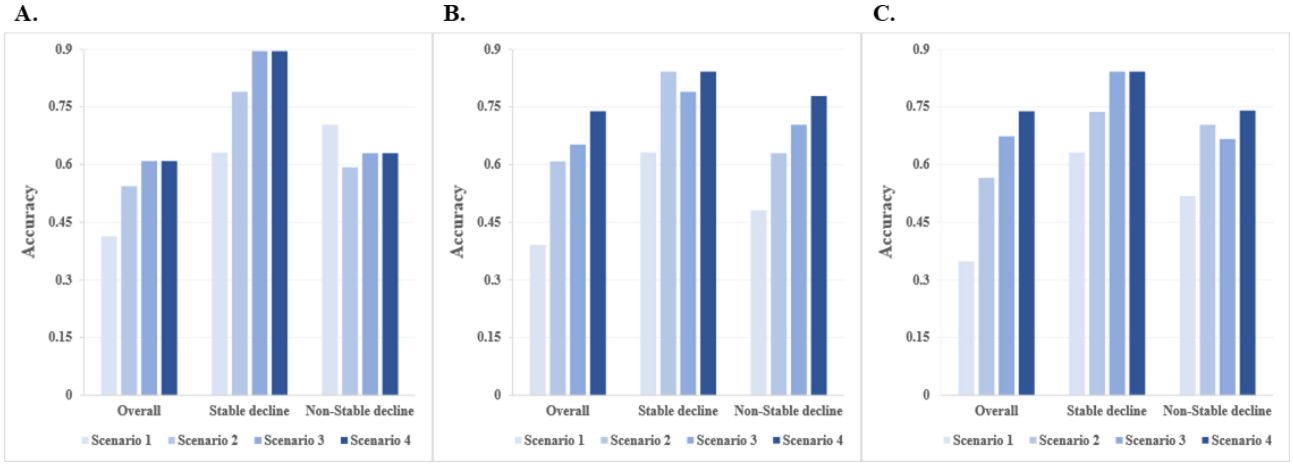
[g]ASMS: Altman Self-Rating Mania Scale.

Prediction of Depression Variation Pattern With Digital mHealth Measures

Figure 4 and Table S2 in Multimedia Appendix 1 show the classification results obtained from various input information by different machine learning approaches. It is shown that by using baseline features and FPC scores of the PROs and digital phenotype records (scenario 2), decision tree, random forest, and XGBoost all give reasonable classification accuracy for the overall data (54.35%, 60.87%, and 56.52%), stable decline patients (78.95%, 84.21%, and 73.68%), and nonstable decline patients (59.26%, 62.96%, and 70.37%). On the other hand, when the models were constructed without dynamic features of the digital mHealth measures (scenario 1), all considered machine learning approaches show poor classification efficacy, which indicates the great importance of digital mHealth measures in the monitoring of depression variation. Further, comparing the predictive performance in scenario 2 with that in scenarios 3 and 4, though the classification accuracy was elevated, the improvement was not as obvious as that achieved by changing from scenario 1 to scenario 2. Moreover, despite a marginal reduction compared to the ideal clinical scenarios, scenario 2 offers practical advantages in terms of scalability and reduces assessment burden for continuous follow-ups.

XSL•FO
RenderX

Figure 4. Classification accuracy for the overall data, stable decline patients, and nonstable decline patients with different input information based on the machine learning algorithms: (A)decision tree, (B) random forest, and (C) XGBoost. Scenario 1: baseline features; scenario 2: baseline features and FPC scores of the PROs and digital phenotype records; scenario 3: baseline features and HAMD-17 and HAMA scores at week 2; scenario 4: baseline features, FPC scores of the PROs and digital phenotype records, and HAMD-17 and HAMA scores at week 2. FPC: functional principal component; HAMA: Hamilton Anxiety Scale; HAMD-17: Hamilton Depression Rating Scale; PRO: patient-reported outcome; XGBoost: Extreme Gradient Boosting.



Contribution of Digital mHealth Measures

Figure S5 and Table S3 in Multimedia Appendix 1 show the variable importance for decision tree, random forest, and XGBoost in the prediction of depression variation patterns using baseline features and FPC scores of the PROs and digital phenotype records (scenario 2). For both random forest and XGBoost, which yielded more accurate classification, the dynamic feature of IMS score records was of paramount importance in the model construction. Moreover, FPC scores of the ASMS score and sleep duration also played a dominant role in the modeling, closely following the baseline HAMD-17 score. Table S4 in Multimedia Appendix 1 reports the *P* values of the permutation test, which confirm the significance of the contribution of both baseline HAMD-17 score and mHealth measures. The variable importance results demonstrated the great contribution of digital mHealth measures in the prediction of depression variation.

Sensitivity Analysis

Table S5 in Multimedia Appendix 1 reports the classification results of decision tree, random forest, and XGBoost with the use of baseline features and the mean values of IMS score, ASMS score, and sleep duration, rather than the FPC scores of them (scenario 2). Compared with introducing the dynamic features of the mHealth measures, using the mean values led to inferior prediction, especially for decision trees (overall data: 54.35% vs 47.83%; stable decline: 78.95% vs 47.37%; nonstable decline: 59.26% vs 59.26%) and random forests (overall data: 60.87% vs 50.00%; stable decline: 84.21% vs 78.95%; nonstable decline: 62.96% vs 55.56%). The results implied the necessity of considering the temporal dynamics of the digital mHealth measures.

Discussion

Principal Findings

This study evaluated the contribution of digital mHealth measures, including PROs and digital phenotype, to the prediction of the variation of depression within 12 weeks for patients with MDD. Based on the variation of HAMD-17 scores at 5 visits within 12 weeks, each patient was labeled as stable decline, fluctuate decline, fast decline, or delayed and fluctuate. The study used functional data analysis methods to address the severe sparsity of PROs and digital phenotype records caused by poor patient adherence, thereby effectively extracting their dynamic features. The classification results showed that by combining the baseline features with the dynamic information of PROs and digital phenotypes, the depression variation pattern can be identified with reasonable accuracy. Overall, our study demonstrated that the dynamic features of mHealth measures extracted by the functional data analysis method can be an effective alternative to clinical visits for depression variation monitoring. The findings can alleviate the burdens for both patients and doctors.

For the detection of latent trajectory classes regarding depression severity, growth mixture modeling was widely used in the previous studies [43-45]. However, growth mixture modeling was more likely to capture the similarity in the magnitude of depression severity, rather than its changing pattern. In our study, the patients with MDD were classified into 4 classes based on *k*-means clustering through the HAMD-17 score differences. Within each class, the patients exhibited the same temporal trend of depression severity. The classes were further named based on the dynamic characteristics of the depression symptom trajectories. In specific, stable decline implies a stable improvement of depression symptoms, fluctuate decline indicates a fluctuation of depression symptoms in the early stage and an alleviation at the end, fast decline represents a high baseline severity and a rapid improvement of depression symptoms, while delayed and fluctuate means a delayed improvement and sharp variation of depression symptoms.

Our findings revealed that the dynamic features of digital mHealth measures can reflect the depression variation of patients with MDD. By recording the trajectories of IMS scores, ASMS scores, and sleep duration using the smartphone and wristband, the dynamic characteristics of depression severity can be evaluated without strong dependence on continuous tracking of HAMD-17 scores. Moreover, results from the variable

XSL•FO
RenderX

importance analysis highlighted the contribution of the mHealth measures in the depression variation prediction.

Further, the sensitivity analysis indicated that taking into account the temporal dynamics of the mHealth measures is beneficial. In our study, the dynamic features of PROs and digital phenotypes were fully characterized through a functional data analysis method, rather than aggregating the records to a single scalar summary measure (eg, mean or median) by collapsing over time. Functional data analysis is popular in the study of continuously recorded data, such as physical activity data [46] and electroencephalography data [35], which supports the use of functional data analysis methods in our research. However, classical functional data analysis methods are only suitable for densely collected data. To extract dynamic features of the sparsely collected PROs and digital phenotypes, the PACE method was applied, which is a commonly used method for the principal component analysis of sparsely and irregularly observed functional data [36]. Collectively, the benefits of using functional data analysis methods can be summarized in two aspects: (1) avoiding the loss of information related to temporal dynamics of the PROs and digital phenotype records and (2) conquering the problem of sparse observations caused by poor patient adherence without excluding any participants with a low observation size (or in other words, with a high percentage of missing observations).

Comparison With Previous Work

Previous studies have demonstrated the potential of mHealth measures in predicting symptom severity of patients with MDD [6,23,47]. To further monitor the depression stability, Bai et al [20] defined two types of variation patterns, steady and swing, based on the patient-reported Patient Health Questionnaire-9 scores. However, such PRO records cannot substitute for the gold-standard HAMD-17 scores evaluated by clinicians in the longitudinal monitoring of patients with MDD. In our study, the more credible clinician-reported scores were used for outcome labeling, and PROs were incorporated as predictive factors, underscoring the need for more sophisticated use of PRO records. Moreover, Ikäheimonen et al [48] also considered variation of depression by partitioning the participants into 4 classes: declines, increases, remains depressed, and remains nondepressed. Although their model based on smartphone behavioral data achieved an overall accuracy of 75%, the predicting efficacy for classes with state changes (declines and increases) was invalid with a precision lower than 50%. Compared with their work, we refined the depression variation labels to facilitate the detection of specific temporal patterns, and our model yielded a more robust predictive performance for patients with fluctuating clinical courses. This improvement likely stems from our application of the functional data analysis to the mHealth records, which can effectively characterize the continuous underlying transition patterns that traditional summary statistics might overlook.

The core contribution of our study lies in the methodological integration of dual temporal dynamics for depression severity and mHealth measures, and the capacity to handle extremely sparse mHealth records. First, our definition of the outcome labels effectively characterized the temporal evolution of depression severity, enabling the identification of the inherent heterogeneity in depression progression over time. Moreover, the dynamic features of mHealth measures were extracted to predict the changing patterns of depression trajectory, forming an innovative dual-dynamic framework that is more practical in real-world psychiatric monitoring. Last but not least, our study made it feasible to obtain an acceptable depression monitoring using mHealth records with great irregularity and sparsity, which significantly enhances the ecological validity of our model and mitigates the adherence-related barriers that hinder clinical application of mHealth measures.

In recent years, high-resource biological markers derived from neuroimaging [49] and electrophysiology [50] have gained prominence in predicting antidepressant response due to their insights into underlying mechanisms. Compared with these markers, low-cost mHealth measures possess the ability of high-frequency, continuous recording within a naturalistic environment, facilitating scalable real-world monitoring and the dynamic capture of longitudinal symptom variations that static biomarkers may overlook. Given the complementary strengths, the integration of high-resource biological markers with real-world mHealth data represents a promising avenue for personalized psychiatry that warrants further investigation.

Strengths and Limitations

The strengths of our study can be summarized as follows. First, this study contained a long period of continuous monitoring for a large number of patients with MDD. Second, our study overcame the missing and sparse problems that universally existed in natural observations with the use of functional data analysis methods, which also facilitated the characterization of the temporal dynamics of PROs and digital phenotype records. Third, our study effectively reduced the dependence of patient adherence and eased the heavy burden of clinical follow-up in the prediction of depression variation trajectory.

Our study still has several limitations. One limitation is that only the IMS score, ASMS score, and sleep duration were continuously monitored and were used in the prediction of depression severity. Additional collection of more indicators of PROs and digital phenotype would be beneficial [27]. The other limitation is that only HAMD-17 scores were used in the construction of the labels of depression variation, which may be insufficient. Therefore, collecting more indicators of depressive symptoms and putting them together would be advantageous for the characterization of the change of depression severity. In addition, as the entire sample was used in the *k*-means clustering analysis to ensure a robust definition of clinical phenotypic labels, an implicit data leakage was introduced into the model assessment. So, the results, which may be inclined toward optimism, should be interpreted with caution. Moreover, constrained by the limited and imbalanced sample sizes of the participating centers, the predictive performance exhibited variability across heterogeneous clinical environments. This suboptimal performance underscores the inherent challenges of model generalizability. Thus, further research is required to enhance the robustness of the model when applied to diverse clinical settings. Last, while the predicted depression variation pattern can guide further

XSL•FO
RenderX

treatment for the patient with MDD, it was not able to provide an earlier warning within 12 weeks, which is worth further research.

Conclusions

Our study demonstrated the potential of mHealth measures, as an alternative to clinical visits, in the monitoring of depression variation, even under poor patient adherence. The results suggested that dynamic features of PROs and digital phenotypes, combined with baseline clinical features, can provide clinically meaningful information for depression variation patterns. Our findings help reduce dependence on strict patient adherence and alleviate the burdens of frequent clinical follow-ups, which facilitates a more practical monitoring of depression trajectory variations.

Funding

This study was funded by the Beijing Municipal Science & Technology Commission (Z221100007422010).

Authors' Contributions

Co-Corresponding Author Gang Wang
Beijing Key Laboratory of Intelligent Drug Research and Development for Mental Disorders, National Clinical Research Center for Mental Disorders, National Center for Mental Disorders, Beijing Anding Hospital, Capital Medical University, Beijing, China
No. 5 Ankang Hutong, Xicheng District, Beijing, China
Phone Number: 86 13466604224
Email: gangwangdoc@ccmu.edu.cn

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional statistical results.
[DOCX File, 3764 KB - mhealth_v14i1e81397_app1.docx]

References

1. Kessler RC, Berglund P, Demler O, et al. The epidemiology of major depressive disorder. JAMA 2003 Jun 18;289(23):3095. [doi: 10.1001/jama.289.23.3095]
2. Malhi GS, Mann JJ. Depression. Lancet 2018 Nov 24;392(10161):2299-2312. [doi: 10.1016/S0140-6736(18)31948-2] [Medline: 30396512]
3. Rush AJ, Trivedi MH, Wisniewski SR, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am J Psychiatry 2006 Nov;163(11):1905-1917. [doi: 10.1176/ajp.2006.163.11.1905] [Medline: 17074942]
4. Wagner S, Engel A, Engelmann J, et al. Early improvement as a resilience signal predicting later remission to antidepressant treatment in patients with major depressive disorder: systematic review and meta-analysis. J Psychiatr Res 2017 Nov;94:96-106. [doi: 10.1016/j.jpsychires.2017.07.003] [Medline: 28697423]
5. de Vries YA, Roest AM, Bos EH, Burgerhof JGM, van Loo HM, de Jonge P. Predicting antidepressant response by monitoring early improvement of individual symptoms of depression: individual patient data meta-analysis. Br J Psychiatry 2019 Jan;214(1):4-10. [doi: 10.1192/bjp.2018.122] [Medline: 29952277]
6. Sun S, Folarin AA, Zhang Y, et al. Challenges in using mHealth data from smartphones and wearable devices to predict depression symptom severity: retrospective analysis. J Med Internet Res 2023 Aug 14;25:e45233. [doi: 10.2196/45233] [Medline: 37578823]
7. Zhuparris A, Maleki G, van Londen L, et al. A smartphone- and wearable-based biomarker for the estimation of unipolar depression severity. Sci Rep 2023 Nov 1;13(1):18844. [doi: 10.1038/s41598-023-46075-2] [Medline: 37914808]
8. Beltrán JM, Jacob Y, Mehta MM, et al. Digital measures of activity and motivation impact depression and anxiety in the real world. NPJ Digit Med 2025 May 10;8(1):268. [doi: 10.1038/s41746-025-01669-0] [Medline: 40348910]
9. Goltermann J, Emden D, Leehr EJ, et al. Smartphone-based self-reports of depressive symptoms using the remote monitoring application in psychiatry (ReMAP): interformat validation study. JMIR Ment Health 2021 Jan 12;8(1):e24333. [doi: 10.2196/24333] [Medline: 33433392]
10. Yang M, Ngai ECH, Hu X, et al. Digital phenotyping and feature extraction on smartphone data for depression detection. Proc IEEE 2025;112(12):1773-1798. [doi: 10.1109/JPROC.2025.3542324]
11. Torous J, Staples P, Shanahan M, et al. Utilizing a personal smartphone custom app to assess the Patient Health Questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder. JMIR Ment Health 2015;2(1):e8. [doi: 10.2196/mental.3889] [Medline: 26543914]

XSL•FO
RenderX

12. Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 2016 Jun;41(7):1691-1696. [doi: 10.1038/npp.2016.7] [Medline: 26818126]
13. Yim SJ, Lui LMW, Lee Y, et al. The utility of smartphone-based, ecological momentary assessment for depressive symptoms. J Affect Disord 2020 Sep 1;274:602-609. [doi: 10.1016/j.jad.2020.05.116] [Medline: 32663993]
14. Pandi-Perumal SR, Monti JM, Burman D, et al. Clarifying the role of sleep in depression: a narrative review. Psychiatry Res 2020 Sep;291:113239. [doi: 10.1016/j.psychres.2020.113239]
15. Aledavood T, Torous J, Triana Hoyos AM, Naslund JA, Onnela JP, Keshavan M. Smartphone-based tracking of sleep in depression, anxiety, and psychotic disorders. Curr Psychiatry Rep 2019 Jun 4;21(7):49. [doi: 10.1007/s11920-019-1043-y] [Medline: 31161412]
16. Matcham F, Carr E, Meyer N, et al. The relationship between wearable-derived sleep features and relapse in major depressive disorder. J Affect Disord 2024 Oct 15;363:90-98. [doi: 10.1016/j.jad.2024.07.136] [Medline: 39038618]
17. Lim D, Jeong J, Song YM, et al. Accurately predicting mood episodes in mood disorder patients using wearable sleep and circadian rhythm features. NPJ Digit Med 2024 Nov 18;7(1):324. [doi: 10.1038/s41746-024-01333-z] [Medline: 39557997]
18. Hu Y, Chen J, Chen J, Wang W, Zhao S, Hu X. An ensemble classification model for depression based on wearable device sleep data. IEEE J Biomed Health Inform 2024 May;28(5):2602-2612. [doi: 10.1109/JBHI.2023.3258601] [Medline: 37030745]
19. Nemesure MD, Collins AC, Price GD, et al. Depressive symptoms as a heterogeneous and constantly evolving dynamical system: idiographic depressive symptom networks of rapid symptom changes among persons with major depressive disorder. J Psychopathol Clin Sci 2024 Feb;133(2):155-166. [doi: 10.1037/abn0000884] [Medline: 38271054]
20. Bai R, Xiao L, Guo Y, et al. Tracking and monitoring mood stability of patients with major depressive disorder by machine learning models using passive digital data: prospective naturalistic multicenter study. JMIR Mhealth Uhealth 2021;9(3):e24365. [doi: 10.2196/24365]
21. Price GD, Heinz MV, Song SH, Nemesure MD, Jacobson NC. Using digital phenotyping to capture depression symptom variability: detecting naturalistic variability in depression symptoms across one year using passively collected wearable movement and sleep data. Transl Psychiatry 2023 Dec 9;13(1):381. [doi: 10.1038/s41398-023-02669-y] [Medline: 38071317]
22. Wichers M. The dynamic nature of depression: a new micro-level perspective of mental disorder that meets current challenges. Psychol Med 2014 May;44(7):1349-1360. [doi: 10.1017/S0033291713001979] [Medline: 23942140]
23. Zhang Y, Folarin AA, Sun S, et al. Longitudinal relationships between depressive symptom severity and phone-measured mobility: dynamic structural equation modeling study. JMIR Ment Health 2022 Mar 11;9(3):e34898. [doi: 10.2196/34898] [Medline: 35275087]
24. Leaning IE, Ikani N, Savage HS, et al. From smartphone data to clinically relevant predictions: a systematic review of digital phenotyping methods in depression. Neurosci Biobehav Rev 2024 Mar;158:105541. [doi: 10.1016/j.neubiorev.2024.105541]
25. Onnela JP. Opportunities and challenges in the collection and analysis of digital phenotyping data. Neuropsychopharmacology 2021 Jan;46(1):45-54. [doi: 10.1038/s41386-020-0771-3] [Medline: 32679583]
26. Zou B, Zhang X, Xiao L, et al. Sequence modeling of passive sensing data for treatment response prediction in major depressive disorder. IEEE Trans Neural Syst Rehabil Eng 2023;31:1786-1795. [doi: 10.1109/TNSRE.2023.3260301] [Medline: 37030733]
27. Aledavood T, Luong N, Baryshnikov I, et al. Multimodal digital phenotyping study in patients with major depressive episodes and healthy controls (mobile monitoring of mood): observational longitudinal study. JMIR Ment Health 2025 Feb 21;12:e63622. [doi: 10.2196/63622] [Medline: 39984168]
28. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960 Feb;23(1):56-62. [doi: 10.1136/jnnp.23.1.56] [Medline: 14399272]
29. Hamilton M. The assessment of anxiety states by rating. Br J Med Psychol 1959;32(1):50-55. [doi: 10.1111/j.2044-8341.1959.tb00467.x] [Medline: 13638508]
30. Leu SH, Chou JY, Lee PC, et al. Validity and reliability of the Chinese version of the Sheehan Disability Scale (SDS-C). Asia Pac Psychiatry 2015 Jun;7(2):215-222. [doi: 10.1111/appy.12182] [Medline: 25847187]
31. Nahum M, Van Vleet TM, Sohal VS, et al. Immediate Mood Scaler: tracking symptoms of depression and anxiety using a novel mobile mood scale. JMIR Mhealth Uhealth 2017 Apr 12;5(4):e44. [doi: 10.2196/mhealth.6544] [Medline: 28404542]
32. Chen X, Feng Z, Li N, Xiao L, Chen X, Zhu X. The Chinese version of the immediate mood scaler (IMS): a study evaluating its validity, reliability, and responsiveness in patients with MDD in China. BMC Psychiatry 2025 Jan 7;25(1):20. [doi: 10.1186/s12888-024-06418-3]
33. Altman EG, Hedeker D, Peterson JL, Davis JM. The Altman Self-Rating Mania Scale. Biol Psychiatry 1997 Nov 15;42(10):948-955. [doi: 10.1016/S0006-3223(96)00548-3] [Medline: 9359982]
34. Ramsay JO, Silverman BW. Functional Data Analysis: Springer; 2005.
35. Ballard ED, Greenstein D, Reiss PT, et al. Functional changes in sleep-related arousal after ketamine administration in individuals with treatment-resistant depression. Transl Psychiatry 2024 Jun 4;14(1):238. [doi: 10.1038/s41398-024-02956-2] [Medline: 38834540]

XSL•FO
RenderX

36. Yao F, Müller HG, Wang JL. Functional data analysis for sparse longitudinal data. J Am Stat Assoc 2005 Jun;100(470):577-590. [doi: 10.1198/016214504000001745]
37. Breiman L, Friedman J, Olshen RA, Stone CJ. Classification and Regression Trees: Chapman and Hall/CRC; 1984. [doi: 10.1201/9781315139470]
38. Therneau TM, Atkinson EJ. An introduction to recursive partitioning using the RPART routines. 2026 Mar 26. URL: https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf [accessed 2026-04-20]
39. Breiman L. Manual on setting up, using, and understanding random forests V3.1. : University of California, Berkeley URL: https://www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf [accessed 2026-04-20]
40. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining: Association for Computing Machinery; 2016:785-794. [doi: 10.1145/2939672.2939785]
41. Archer E. rfPermute: estimate permutation p-values for random forest importance metrics. R package version 2.5.5. 2025 Jul 23. URL: https://cran.r-project.org/web/packages/rfPermute/rfPermute.pdf [accessed 2026-04-20]
42. Berlim MT, Turecki G. Definition, assessment, and staging of treatment—resistant refractory major depression: a review of current concepts and methods. Can J Psychiatry 2007 Jan;52(1):46-54. [doi: 10.1177/070674370705200108]
43. Uher R, Muthén B, Souery D, et al. Trajectories of change in depression severity during treatment with antidepressants. Psychol Med 2010 Aug;40(8):1367-1377. [doi: 10.1017/S0033291709991528] [Medline: 19863842]
44. Scott K, Lewis CC, Marti CN. Trajectories of symptom change in the treatment for adolescents with depression study. J Am Acad Child Adolesc Psychiatry 2019 Mar;58(3):319-328. [doi: 10.1016/j.jaac.2018.07.908]
45. Meng F, Ou W, Zhao X, et al. Identifying latent subtypes of symptom trajectories in major depressive disorder patients and their predictors. Eur Arch Psychiatry Clin Neurosci 2025 Jun;275(4):1177-1187. [doi: 10.1007/s00406-024-01883-z] [Medline: 39223324]
46. Zablocki RW, Hartman SJ, Di C, et al. Using functional principal component analysis (FPCA) to quantify sitting patterns derived from wearable sensors. Int J Behav Nutr Phys Act 2024 Apr 26;21(1):48. [doi: 10.1186/s12966-024-01585-8] [Medline: 38671485]
47. Zhang Y, Folarin AA, Sun S, et al. Relationship between major depression symptom severity and sleep collected using a wristband wearable device: multicenter longitudinal observational study. JMIR Mhealth Uhealth 2021 Apr 12;9(4):e24604. [doi: 10.2196/24604] [Medline: 33843591]
48. Ikäheimonen A, Luong N, Baryshnikov I, et al. Predicting and monitoring symptoms in patients diagnosed with depression using smartphone data: observational study. J Med Internet Res 2024 Dec 3;26:e56874. [doi: 10.2196/56874] [Medline: 39626241]
49. Zhukovsky P, Trivedi MH, Weissman M, Parsey R, Kennedy S, Pizzagalli DA. Generalizability of treatment outcome prediction across antidepressant treatment trials in depression. JAMA Netw Open 2025 Mar 3;8(3):e251310. [doi: 10.1001/jamanetworkopen.2025.1310] [Medline: 40111362]
50. Ip CT, de Bardeci M, Kronenberg G, et al. EEG-vigilance regulation is associated with and predicts ketamine response in major depressive disorder. Transl Psychiatry 2024 Jan 26;14(1):64. [doi: 10.1038/s41398-024-02761-x] [Medline: 38272875]

Abbreviations

ASMS: Altman Self-Rating Mania Scale
FPC: functional principal component
HAMA: Hamilton Anxiety Scale
HAMD-17: Hamilton Depression Rating Scale
IMS: Immediate Mood Scaler
MDD: major depressive disorder
mHealth: mobile health
PACE: principal components analysis through conditional expectation
PRO: patient-reported outcome

Edited by Z Liang; submitted 28.Jul.2025; peer-reviewed by KM Choi, S Tedesco; revised version received 26.Mar.2026; accepted 02.Apr.2026; published 11.May.2026.

Please cite as:
Zhong R, Li N, Xiao L, Feng L, Feng Y, Wang G, Zhu X
Contribution of Longitudinal Mobile Health Measures in the Dynamic Track of Patients With Major Depressive Disorder: Multiple Centers, Prospective Cohort Study Using Functional Data Analysis and Machine Learning
JMIR Mhealth Uhealth 2026;14:e81397
URL: https://mhealth.jmir.org/2026/1/e81397
doi:10.2196/81397

XSL•FO
RenderX

© Rou Zhong, Nanxi Li, Le Xiao, Lei Feng, Yuan Feng, Gang Wang, Xuequan Zhu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Digital Structured Education With Behavioral Nudge Tools for Adults With Type 2 Diabetes: Multicenter Randomized Controlled Trial

Yan Lin[1,2], MN; Yingchun Zeng[3], PhD; Kaining Chen[4], MM; Zongcun Chen[2], PhD; Caihua Ye[2], MN; Ying Zhou[5], BSN; Qiwei Zhou[1], MN; Chengying Yu[1], MN; Vivien Xi Wu[3], PhD; Samuel Seidu[6,7], PhD; Bin Li[4], BSN; Xinjun Jiang[1,8], PhD

[1]School of Nursing, Hainan Medical University, No 3 Xueyuan Road, Longhua district, Haikou, Hainan, China

[2]The Second Affiliated Hospital of Hainan Medical University, Haikou, China

[3]National University of Singapore, Singapore, Singapore

[4]Hainan General Hospital, Haikou, China

[5]928th Hospital of PLA Joint Logistics Support Force, Haikou, China

[6]Leicester Diabetes Centre, Leicester, United Kingdom

[7]Leicester General Hospital, Leicester, United Kingdom

[8]Key Laboratory of Emergency and Trauma of Ministry of Education, Engineering Research Center for Hainan Biological Sample Resources of Major Diseases, Key Laboratory of Tropical Cardiovascular Diseases Research of Hainan Province, The First Affiliated Hospital of Hainan Medical University, Haikou, China

Corresponding Author:
Xinjun Jiang, PhD
School of Nursing, Hainan Medical University, No 3 Xueyuan Road, Longhua district, Haikou, Hainan, China

Abstract

Background: Digital interventions offer scalable alternatives to traditional face-to-face diabetes education, but often face challenges related to inconsistent clinical effectiveness, and declining user engagement. However, whether a digital structured education program integrated with behavioral nudge tools can improve metabolic, behavioral, and psychological outcomes in adults with type 2 diabetes remains unclear.

Objective: This study aimed to evaluate the effectiveness of a digital structured education program integrated with behavioral nudge tools in improving metabolic, behavioral, and psychological outcomes among adults with type 2 diabetes.

Methods: This multicenter randomized controlled trial was conducted in the endocrinology departments of 4 hospitals in China. Adults with type 2 diabetes were randomly assigned to an intervention group receiving a digital structured education program integrated with behavioral nudge tools (n=146) or a control group receiving standard digital diabetes education (n=147). Assessments were conducted at baseline and 12-week follow-up. The primary outcome was hemoglobin A_{1c} (HbA_{1c}) at 12 weeks, adjusted for baseline HbA_{1c}, and study center. Secondary outcomes included fasting blood glucose (FBG), weight, BMI, waist circumference, blood pressure, lipid profiles, self-management behaviors, self-efficacy, and habit strength.

Results: Among 293 participants (mean age 49.19, SD 10.02 y), 287 (97.9%) completed follow-up. At 12 weeks, the intervention group demonstrated significantly greater improvements than the control group in HbA_{1c} (adjusted mean difference −0.38%, 95% CI −0.68% to −0.09%; *P*=.01), FBG (adjusted mean difference −0.75, 95% CI −1.27 to −0.44 mmol/L; *P*<.001), weight (adjusted mean difference −0.84, 95% CI −1.61 to −0.07 kg; *P*=.03), BMI (adjusted mean difference −0.38, 95% CI −0.65 to −0.11 kg/m²; *P*=.01), systolic blood pressure (adjusted mean difference −2.71, 95% CI −4.62 to −0.79 mm Hg; *P*=.01), diastolic blood pressure (adjusted mean difference −2.92, 95% CI −4.47 to −1.37 mm Hg; *P*<.001), and total cholesterol (adjusted mean difference −0.27, 95% CI −0.48 to −0.05 mmol/L; *P*=.02). The intervention was also associated with significantly greater improvements in self-management behaviors, self-efficacy, and habit strength (all *P*<.05).

Conclusions: Digital structured education integrated with behavioral nudge tools improved metabolic outcomes and strengthened psychological and behavioral determinants of self-management among adults with type 2 diabetes over a 12-week period. These findings suggest that a digital structured education program integrated with behavioral nudge tools may enhance diabetes self-management beyond standard digital diabetes education. Further studies with longer follow-up and real-world implementation are warranted to evaluate the sustainability, generalizability, and long-term clinical impact of this integrated intervention.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2400082373; https://tinyurl.com/3djbnt36

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2026;14:e91407) doi:10.2196/91407

KEYWORDS

type 2 diabetes; digital structured education; behavioral nudges; self-management behaviors; randomized controlled trial

Introduction

Background

Type 2 diabetes is a major global public health challenge and one of the fastest-growing chronic diseases. The International Diabetes Federation estimates that 589 million adults aged 20-79 years were living with diabetes in 2024, including approximately 148 million in China, where type 2 diabetes accounts for 90%-95% of cases [1]. Diabetes contributes to more than 3.4 million deaths annually worldwide and generates nearly US $1 trillion in health care expenditures, posing substantial clinical, economic, and societal burdens [1]. In China, the high prevalence of type 2 diabetes and its complications underscores an urgent need for effective and scalable management strategies.

Optimal diabetes management requires maintained glycemic and lipid control, blood pressure regulation, and long-term adherence to self-management behaviors [2,3]. However, control rates for blood glucose, blood pressure, and lipids remain suboptimal in mainland China, partly due to limited access to structured education and insufficient long-term engagement in self-management [4,5]. Improving maintained behavior change remains a central challenge in chronic disease management. Behavior change is increasingly understood through the lens of dual-process theory, which describes maintained 2 interacting systems: a reflective system that supports deliberate, goal-directed decision-making and an automatic system that drives habitual behavior through environmental cues [6-8].

Digital health interventions have emerged as scalable tools for diabetes management, providing remote education, monitoring, and personalized feedback through mobile and web-based platforms [9,10]. Evidence indicates that digital approaches can improve glycemic control, reduce health care usage, and enhance patient empowerment, particularly in resource-limited settings [7,11-13]. However, despite their potential, existing evidence on digital interventions reveals inconsistent findings regarding efficacy, with some studies highlighting challenges in maintaining engagement and a reliance on passive features like reminders rather than comprehensive behavioral strategies [9,14].

Behavioral Nudges in Digital Diabetes Management

Behavioral nudge tools, grounded in nudge theory, represent an evolving innovation in health behavior change, subtly altering environments or choices to promote healthier decisions without limiting autonomy [15-17]. These tools have been applied in various chronic disease contexts, including obesity and cardiovascular health, showing promise in fostering habits through unconscious cues, such as default options or timely prompts [17,18]. Nevertheless, digital structured education programs integrated with behavioral nudge tools remain underexplored, particularly in China, where adoption is limited and often disconnected from structured educational frameworks.

A digital structured education program integrated with behavioral nudge tools may better address both reflective and automatic drivers of behavior by combining complementary intervention components. Digital education enhances knowledge, self-efficacy, and intentional self-management, while behavioral nudge tools target automatic processes to build habits, such as through gamified prompts or environmental cues that encourage maintained actions like physical activity or medication adherence [19-21]. By simultaneously targeting these complementary behavioral processes, such integrated programs may reduce cognitive load, facilitate maintained behavior change, and potentially achieve greater improvements than standard digital diabetes education alone.

Knowledge Gap

Despite these synergies, significant gaps persist in the integration of nudge tools with digital education for adults with type 2 diabetes. Limited high-quality evidence from multicenter randomized controlled trials evaluating digital structured education programs integrated with behavioral nudge tools, especially in diverse populations like those in China, hinders understanding of their effectiveness and cultural adaptability [22]. Additionally, there is a scarcity of studies examining how these integrated interventions affect psychological outcomes (eg, habit strength and self-efficacy) alongside metabolic ones, and how they can be scaled equitably amid varying digital literacy levels.

Study Objective

To address these gaps, we conducted a multicenter randomized controlled trial to evaluate the effectiveness of a digital structured education with behavioral nudge tools (DSE-BN) program compared with standard digital diabetes education developed by the Chinese Diabetes Association among adults with type 2 diabetes in China. Guided by dual-process theory, the DSE-BN program was designed to target both reflective and automatic behavioral processes through the integration of digital structured education and behavioral nudge tools. We hypothesized that, compared with standard digital diabetes education, the DSE-BN program would produce greater improvements in metabolic outcomes (hemoglobin A_{1c} [HbA_{1c}], fasting blood glucose [FBG], weight, BMI, blood pressure, and lipid profiles), behavioral outcomes (diabetes self-management behaviors), and psychological outcomes (self-efficacy and habit strength) among adults with type 2 diabetes in China.

Methods

Study Design and Setting

This multicenter randomized clinical trial was conducted across 3 tertiary hospitals in Hainan Province, China—Hainan Provincial People's Hospital and the First and Second Affiliated Hospitals of Hainan Medical University. To ensure adequate recruitment, 928th Hospital was added as an additional study

XSL•FO
RenderX

site after trial initiation; no other substantive changes were made to the trial design, outcomes, or analytical plan.

Participants were recruited between March and September 2024 and were randomly assigned in a 1:1 ratio to either the intervention or control group. Follow-up assessments of both benefits and harms were completed at 12 weeks, between June 2024 and December 2025. An interim analysis was conducted midway through the trial to monitor data quality, recruitment progress, and participant safety. Given the low-risk nature of the behavioral education intervention, no formal stopping rules were predefined, and the trial was completed as planned without early termination. Reporting followed the CONSORT (Consolidated Standards of Reporting Trials) checklist.

Participants

Eligible participants were adults aged 18 to 65 years with type 2 diabetes diagnosed according to the 2020 Chinese guideline for the prevention and treatment of type 2 diabetes mellitus [23], receiving lifestyle interventions alone or combined with oral agents, cognitively and communicatively intact, and able to use a smartphone with guidance. Exclusion criteria included severe diabetes-related complications, mental or cognitive disorders, mobility impairment, recent cancer treatment, or participation in other trials.

The sample size was calculated using PASS (Power Analysis and Sample Size; version 15.0; NCSS) with the tests for 2 means in a multicenter randomized design procedure, using HbA_{1c} as the primary outcome. Assuming a between-group difference of 0.3% and an SD of 0.85% [10], with a 2-sided α of .05 and 80% power, the required sample size was 227 participants. Because the PASS multicenter randomized design procedure requires specification of an intraclass correlation coefficient (ICC), and no empirical estimate of the ICC was available from previous comparable studies, an ICC of 0.10 was adopted as a conservative design assumption, following the recommendation provided in the PASS software documentation, to account for potential within-site correlation. Allowing for a 20% attrition rate, the target sample size was 284 participants. Ultimately, 293 participants were enrolled, exceeding the prespecified target sample size.

Randomization and Blinding

Participants were randomly allocated in a 1:1 ratio to the intervention or control group using a computer-generated randomization sequence created with SPSS (version 26.0; IBM Corp). Individual-level simple randomization was adopted. Neither block randomization nor stratified randomization by study center was implemented; therefore, allocation block size was not applicable. After eligible participants had been recruited at each study site, an independent researcher centrally assigned participants according to the pregenerated randomization sequence. The randomization process was independent of participant recruitment and outcome assessment, and allocation concealment was maintained throughout the allocation process, thereby minimizing the risk of selection bias across study sites.

Due to the nature of the behavioral intervention, blinding of participants and intervention nurses was not feasible. However, physicians responsible for routine clinical care and outcome assessment, as well as data analysts, were blinded to group allocation to minimize potential performance and detection bias. In addition, study center was included as a covariate in the primary statistical analyses to account for potential between-center differences.

Intervention

Intervention Implementation

Each study site was supported by at least 1 physician and 1 nurse, operating within a centrally coordinated intervention team. Physicians were responsible for clinical diagnosis and participant recruitment, while nurses delivered the intervention and managed informed consent, baseline and follow-up data collection, registration on the self-management support system (SMASS), and ongoing participant support. Study procedures and intervention delivery were coordinated centrally to ensure consistency across sites.

To standardize implementation of the intervention across study sites, all intervention personnel completed uniform, centralized training before study initiation. Training covered intervention content, delivery procedures, use of the SMASS platform, and fidelity requirements. Competency assessments were conducted following training, and only personnel meeting predefined competency criteria were authorized to deliver the intervention. Ongoing supervision and cross-site communication were used to maintain consistency of implementation throughout the study period.

Intervention Group

Participants in the intervention group received the DSE-BN program, a nurse-developed, theory-informed digital structured education program incorporating an enhanced educational workflow, shared digital support features, and behavioral nudge tools. Grounded in dual-process theory, the DSE-BN program was designed to influence both reflective and automatic behavioral processes. Reflective behavior change was targeted through structured digital education aimed at improving diabetes-related knowledge and enhancing self-efficacy via progressive goal setting, structured knowledge learning, repetition and review, timely positive feedback, experience sharing, and reminders, while automatic behavioral processes were addressed through subtle environmental modifications using behavioral nudge tools embedded in participants' everyday environments.

The DSE-BN program comprised two integrated components: (1) a theory-informed digital structured education program incorporating an enhanced educational workflow and shared digital support features and (2) behavioral nudge tools. The digital structured education component was designed to facilitate reflective behavior change through structured educational activities and enhanced educational workflow, whereas the behavioral nudge component targeted automatic behavioral processes by providing environmental cues embedded within participants' daily living environments. These 2 components were implemented simultaneously throughout the intervention period as complementary elements of a single integrated intervention.

XSL•FO
RenderX

To facilitate comparison, the key components and distinguishing features of the DSE-BN and control interventions are summarized in Table 1.

Table . Comparison of the key components and distinguishing features of the DSE-BN[a] and standard digital diabetes education interventions.

Intervention component or feature	Intervention group (DSE-BN)	Control group (standard digital diabetes education)
Intervention overview		
Overall intervention	A nurse-developed, theory-informed digital structured education program integrated with behavioral nudge tools	Standard digital diabetes education based on the Chinese Diabetes Society educational materials
Behavioral theory	Dual-process theory	Not explicitly theory-informed
Primary objective	Promote sustainable diabetes self-management through an enhanced educational workflow, shared digital support features, and environmental behavioral nudges	Deliver standardized diabetes education to improve diabetes-related knowledge and self-management
Educational design		
Educational content	Theory-informed digital structured education	Standard diabetes education based on the Chinese Diabetes Society educational materials
Educational structure	4 structured weekly digital learning modules with an enhanced educational workflow	4 weekly digital educational modules
Delivery platform	SMASS[b] digital platform	SMASS digital platform
Shared digital support features		
Sequential learning	✓	✓
Discussion area for peer interaction and professional support	✓	✓
Automated learning reminders	✓	✓
Rule-based tailored feedback	✓	✓
Data-driven early-warning alerts (presented using a traffic-light system for metabolic status)	✓	✓
Enhanced educational workflow		
Preparatory questions before each module	✓	—[c]
Review questions after each module	✓	—
Immediate feedback on review questions	✓	—
Behavioral nudge component		
Behavioral nudge tools	✓ (Self-monitoring reminder sticker, portion-control plate, footprint exercise mat, foot-care reminder image, and medication reminder sticker)	—

[a]DSE-BN: digital structured education with behavioral nudge tools.

[b]SMASS: self-management support system.

[c]Not applicable.

Digital Structured Education

The digital structured education component incorporated structured educational content, an enhanced educational workflow, and shared digital support features delivered individually and remotely through 4 weekly modules via the SMASS platform, enabling participants to complete the program independently in their home environments using their smartphones over 4 consecutive weeks (1 session per week; 9 - 25 min per module; total education time approximately 74 min) [11,12,24]. Module content covered diabetes knowledge and self-monitoring, diet management, physical activity and foot care, and regular examinations and follow-up (Table 2). Educational materials included prerecorded videos, interactive quizzes, and learning checkpoints. Modules were completed sequentially, with progression restricted until completion of the current module.

Following completion of baseline assessments, participants commenced the DSE-BN program. Each module followed a standardized learning workflow, beginning with brief preparatory questions before video viewing. Educational videos were presented in a preset order and could not be skipped or

XSL•FO
RenderX

fast-forwarded to ensure completion of all learning content. After each module, participants completed review questions and received immediate feedback through encouraging language and supportive visual cues (eg, emojis) to reinforce knowledge acquisition and active learning.

Table . Components and delivery of the digital structured education with behavioral nudge tools (DSE-BN) program[a].

Contents of each module	Tools: design rationale and description
Module 1: knowledge about diabetes and self-monitoring	
This module introduces the importance of diabetes management and facilitates patients' self-monitoring. It covers normal blood glucose ranges, hyperglycemia, and hypoglycemia, emphasizing the significance of routine self-monitoring and accurate record-keeping. Patients are encouraged to recognize early warning signs, actively engage in managing their blood glucose levels by self-monitoring.	• Self-monitoring reminder sticker[b]: Developed as a reminder sticker to be placed in prominent household locations or near medications, serving to prompt patients while also reinforcing the blood glucose and urine monitoring components of the structured education program.
Module 2: diet management	
Building on module 1, this session highlights the role of weight management in diabetes care. Patients are introduced to different nutrients and calorie, values using visual aids such as food images. Practical exercises include evaluating and calculating daily dietary intake. Patients are encouraged to set personalized dietary goals and develop a structured meal plan tailored to their needs.	• Nudge plate[c]: A real, 3-color round plate was developed for home dietary management. The plate adopts a traffic-light scheme, divided into 1 red zone, 2 yellow zones, and 1 green zone. Rim markings correspond directly to the food classification cards used in the structured education program. The green zone is designated for low-calorie, vegetables, the yellow zones for moderate-calorie, foods such as staple grains, lean meat, eggs, poultry, and root vegetables, and the red zone for high-calorie, items such as sugary or fatty foods. Notably, the red zone is half the size of the other colored zones. By incorporating familiar traffic-light cues and linking them with the structured education cards, the plate provides an intuitive visual reminder that encourages patients to increase vegetable consumption, consume moderate portions of staple foods and lean proteins, and limit the intake of high-calorie, foods.
Module 3: physical activity and foot care	
This module reviews the dietary principles discussed previously and emphasizes the importance of regular exercise in diabetes management. Patients are guided to set achievable exercise goals and create individualized activity plans. The session also includes education on foot care, with demonstrations of foot exercises to reduce complications and encouragement to integrate these practices into daily routines. Patients are encouraged to conduct foot care and take exercise.	• Footprint exercise mat[b]: Designed as a green footprint exercise mat, with a starting point labeled “6000 steps per day.” Placed at the household entrance, it serves as a daily prompt for physical activity, creating a supportive proximal home microenvironment while reinforcing the exercise component of the structured education program. • Foot care reminder image[b]: Developed as a reminder image to be displayed at home, prompting patients to perform daily foot care and providing guidance on proper techniques, while reinforcing the foot care component of the structured education program.
Module 4: regular examinations and follow-up	
The final module revisits the key points of physical activity and foot care, then expands on the importance of regular medical examinations. Patients are educated about appropriate timing, essential laboratory tests, and screening for diabetes-related complications. Medication management is introduced, with a focus on adherence and understanding prescribed treatments. Patients are encouraged to maintain their goals and attend regular follow-up visits with health care providers.	• Medication reminder sticker[b]: Developed as a reminder sticker to be placed in the home’s drinking area, serving to prompt patients while reinforcing the medication adherence component of the structured education program.

[a]Digital structured education was delivered through the self-management support system platform as 1 session per week for 4 consecutive weeks (9-25 minutes per module; total education time approximately 74 minutes). Behavioral nudge tools were introduced at the beginning of the intervention and used continuously in participants’ home environments throughout the 12-week intervention to provide ongoing environmental cues reinforcing healthy self-management behaviors. Refer to the study by Liu et al [24] for the structured education program.

[b]China Copyright Protection Center. Copyright registration certificate for the work “Self-Management Behavior Illustrations.” Registration No. 国作登字-2024-F-00022150. Beijing: China Copyright Protection Center; 2024 [25].

[c]Nudge plate is a patent product [26].

Beyond the enhanced educational workflow described above, participants in the DSE-BN group also received the same digital support features available to the control group. These shared digital support features included a discussion area for peer interaction and professional support, automated reminders, rule-based tailored feedback, and data-driven early-warning alerts. Within the discussion area of the SMASS platform, participants could post questions or comments and interact with health care professionals and other participants, thereby facilitating peer interaction and professional support. Automated

XSL•FO
RenderX

reminders encouraged timely completion of the learning modules. Rule-based tailored feedback was generated according to participants' learning progress and self-reported metabolic indicator records uploaded through the SMASS platform and was accompanied by encouraging messages to reinforce self-management behaviors. In addition, participants' metabolic status was displayed using a traffic-light system (green, yellow, and red) to facilitate timely recognition of abnormal results and encourage appropriate self-management actions.

Behavioral Nudge Tools

The behavioral nudge component consisted of a set of environmental cueing tools (Table 2), including a self-monitoring reminder sticker, a nudge plate, a footprint exercise mat, a foot care reminder image, and a medication reminder sticker. These tools were designed to subtly encourage healthy behaviors by modifying participants' everyday environments while preserving individual autonomy.

Standardized instructional videos embedded within the SMASS platform guided participants on how to place and use each behavioral nudge tool appropriately in their homes and daily living environments according to their individual routines. These environmental cues remained in participants' home environments throughout the intervention period, providing continuous environmental cues to reinforce healthy self-management behaviors.

Intervention Fidelity

Intervention fidelity was ensured through standardized educational content, centralized training of intervention personnel, restricted module progression, automated reminders, and system-based tracking of learning activities. Engagement with the digital education component was monitored using SMASS system logs. Among the 146 participants allocated to the DSE-BN group, platform records verified that 142 participants completed all 4 educational modules, 1 participant completed 3 modules, and 3 participants completed 1 or fewer modules. Implementation of the behavioral nudge tools was confirmed through participant-submitted photographs of home-based installations reviewed by research nurses. No substantive modifications were made to the DSE-BN program during the study period.

Control Group

Participants in the control group received standard digital diabetes education delivered through the SMASS platform, with educational content developed by the study team based on educational materials from the Chinese Diabetes Association. The control program consisted of 4 weekly sessions (15 - 30 min per session; total approximately 103 min) covering diabetes basics, pharmacotherapy, glucose monitoring, complications, and lifestyle management. WeChat (Tencent Holdings Ltd) was used only for routine communication and follow-up management and was not part of the intervention delivery.

Similar to the DSE-BN group, participants completed the educational modules through the SMASS platform and had access to the same shared digital support features, including sequential learning, a discussion area for interaction with health care professionals and other participants, automated reminders, rule-based tailored feedback, and data-driven early-warning alerts throughout the intervention period. However, unlike the DSE-BN program, the control intervention did not include the enhanced educational workflow, including preparatory questions before video learning and postmodule review questions with immediate feedback, or home-based behavioral nudge tools designed to provide home-based environmental cues for behavior change.

Among the 147 participants in the control group, completion of the 4 digital education modules was tracked through the platform. Platform records indicated that 113 participants completed all 4 modules, 25 completed 3 modules, 6 completed 2 modules, and 3 completed 1 or fewer modules.

All participants in both groups continued to receive routine diabetes care, including outpatient visits, laboratory monitoring, and pharmacological treatment as prescribed by their physicians.

Outcome Measures

All metabolic, psychosocial, and behavioral outcome data were collected at baseline and at the 12-week follow-up (March 2024 to December 2025). The primary outcome was HbA_{1c} level at 12 weeks. Secondary metabolic outcomes of blood lipid profiles (total cholesterol [TC], triglycerides, high-density lipoprotein cholesterol [HDL], and low-density lipoprotein cholesterol [LDL]) were extracted from medical records at each study site. Additional secondary outcomes, including FBG, blood pressure, weight, waist circumference (WC), BMI, self-efficacy, habit strength, self-management behaviors, and other outcomes of medications, were assessed and recorded by trained research nurses using reliable and validated instruments.

Diabetes self-efficacy was assessed using the validated Chinese version of the Self-Efficacy for Diabetes (C-SED) scale [27]. The original scale was developed by Lorig and colleagues [28] in the United States and was subsequently translated and culturally adapted into Chinese by Wei et al [27]. The C-SED comprises 9 items covering 4 domains: diet, physical activity, blood glucose management, and disease control. Each item is rated on a 5-point Likert scale, ranging from 1 ("not at all confident") to 5 ("completely confident"). The mean score ranges from 1 to 5, with higher scores indicating greater diabetes self-efficacy. The Chinese version has demonstrated good reliability in previous studies, and the Cronbach α in this study was 0.87.

Habit strength was assessed using the Self-Report Habit Index (SRHI) [29,30]. The original SRHI consists of standardized descriptive statements that can be adapted to specific behavioral contexts (eg, "XXX is something I do automatically"). In this study, the SRHI was adapted to assess 5 diabetes self-management behaviors: dietary management, physical activity, blood glucose monitoring, foot care, and medication adherence. Following contextual adaptation and expert review, a 20-item questionnaire was developed, comprising 4 items for each behavior. All items were rated on a 5-point Likert scale, ranging from 1 ("strongly disagree") to 5 ("strongly agree"). Total scores range from 20 to 100, with higher scores indicating

XSL•FO
RenderX

stronger habit strength [30]. The Cronbach α in this study was 0.94.

Diabetes self-management behaviors were assessed using the validated Chinese version of the Summary of Diabetes Self-Care Activities questionnaire [31]. The questionnaire comprises 11 items across 6 domains: general diet (2 items), specific diet (2 items), physical activity (2 items), blood glucose monitoring (2 items), foot care (2 items), and medication adherence (1 item). Each item assesses the number of days per week (0 - 7 d) on which the corresponding self-management behavior was performed. Total scores range from 0 to 77, with higher scores indicating better diabetes self-management behaviors. Item 4 is reverse scored, whereas all remaining items are scored in the positive direction. The Cronbach α in this study was 0.78.

Harms were defined as severe hypoglycemia, characterized by events requiring assistance from another person or medical attention. These events were assessed systematically at the 12-week follow-up visit through structured participant interviews and review of medical records, and supplemented by nonsystematic participant self-reports during the study period. All reported events were documented and verified by site investigators.

Demographic and clinical characteristics were collected using a self-designed questionnaire. Information collected included age, sex, ethnicity, medical insurance, educational background, employment status, marital status, cohabitation group, individual monthly income, smoking status, alcohol consumption, and electronic health management. Clinical characteristics included type 2 diabetes duration, type 2 diabetes medication duration, types of medications, and diabetes-related complications.

Quality Control

To ensure consistency and methodological rigor across study sites, standardized operating procedures were used to guide participant recruitment, intervention delivery, data collection, and follow-up. All study personnel received uniform training before study initiation, and adherence to the study protocol was reinforced throughout the trial.

Intervention fidelity was supported through system-based controls embedded within the SMASS platform, including restricted module progression, automated reminders, and engagement prompts (eg, pre- and postmodule quizzes and timed pauses). Nurses monitored participants' engagement and progress on a weekly basis, provided reminders as needed, and assisted with scheduling follow-up assessments. To minimize potential contamination between study groups, intervention materials and behavioral nudge tools were accessible only to participants randomized to the intervention group, and participants in the intervention and control groups were followed up at separate time points.

Quality assurance procedures were implemented at both the site and central levels. Monthly site monitoring was conducted to assess protocol adherence and resolve implementation issues. Data accuracy and completeness were ensured through double data entry and independent expert review. Any discrepancies were resolved by consensus. When necessary, original source data were verified by contacting participants, and records were corrected accordingly.

Ethical Considerations

The trial was registered with the Chinese Clinical Trial Registry (ChiCTR2400082373) on March 27, 2024. The study protocol is available at the registry website, and deidentified study data are available from the corresponding author upon reasonable request. Ethical approval was granted by the Ethics Committee of Hainan Medical University (HYLL-2023 - 461), and the study was conducted in accordance with the Declaration of Helsinki. Before enrollment, all participants received detailed information regarding the study objectives, study procedures, potential risks and benefits, and measures to protect their privacy and confidentiality. Written informed consent was obtained from all participants before participation, and participants were informed of their right to withdraw from the study at any time without affecting their medical care.

To protect participants' privacy and confidentiality, all study data were deidentified before analysis and securely stored, with access restricted to authorized research personnel only. Participants received free access to the SMASS platform throughout the study and could consult the research team at no cost during the intervention period. Participants in the intervention group also received the behavioral nudge tools free of charge, including a self-monitoring reminder sticker, a nudge plate, a footprint exercise mat, a foot care reminder image, and a medication reminder sticker. No monetary compensation was provided for participation in the study.

Statistical Analysis

Analyses were conducted using SPSS (version 27.0) under the intention-to-treat principle. At the 12-week follow-up, outcome data were missing for 6 of the 293 (2%) participants. Missing values were observed for weight, BMI, WC, FBG, HbA_{1c}, lipid profiles, self-efficacy, diabetes self-management behaviors, and habit strength. These missing outcome data were handled using multiple imputation before the primary analyses. A total of 5 imputed datasets were generated via a Markov chain Monte Carlo approach with 10 iterations per dataset. Predictive mean matching was applied for continuous variables, and logistic regression was used for categorical variables. Final estimates were derived by pooling results across datasets according to Rubin's rules.

Baseline characteristics were summarized using means and SDs for normally distributed continuous variables or medians (IQRs) for nonnormally distributed variables, as appropriate, and numbers (percentages) for categorical variables. Between-group comparisons of baseline characteristics were conducted using independent-samples *t* tests, chi-square tests, Fisher exact tests, or Mann-Whitney *U* tests, as appropriate.

The primary analysis compared HbA_{1c} at 12 weeks between the intervention and control groups using ANCOVA, with adjustment for baseline HbA_{1c} and study center as a fixed effect. The same baseline- and center-adjusted ANCOVA approach was applied to secondary outcomes. Sensitivity analyses for both primary and secondary outcomes at 12 weeks were

XSL•FO
RenderX

conducted using ANCOVA with adjustment for baseline values only.

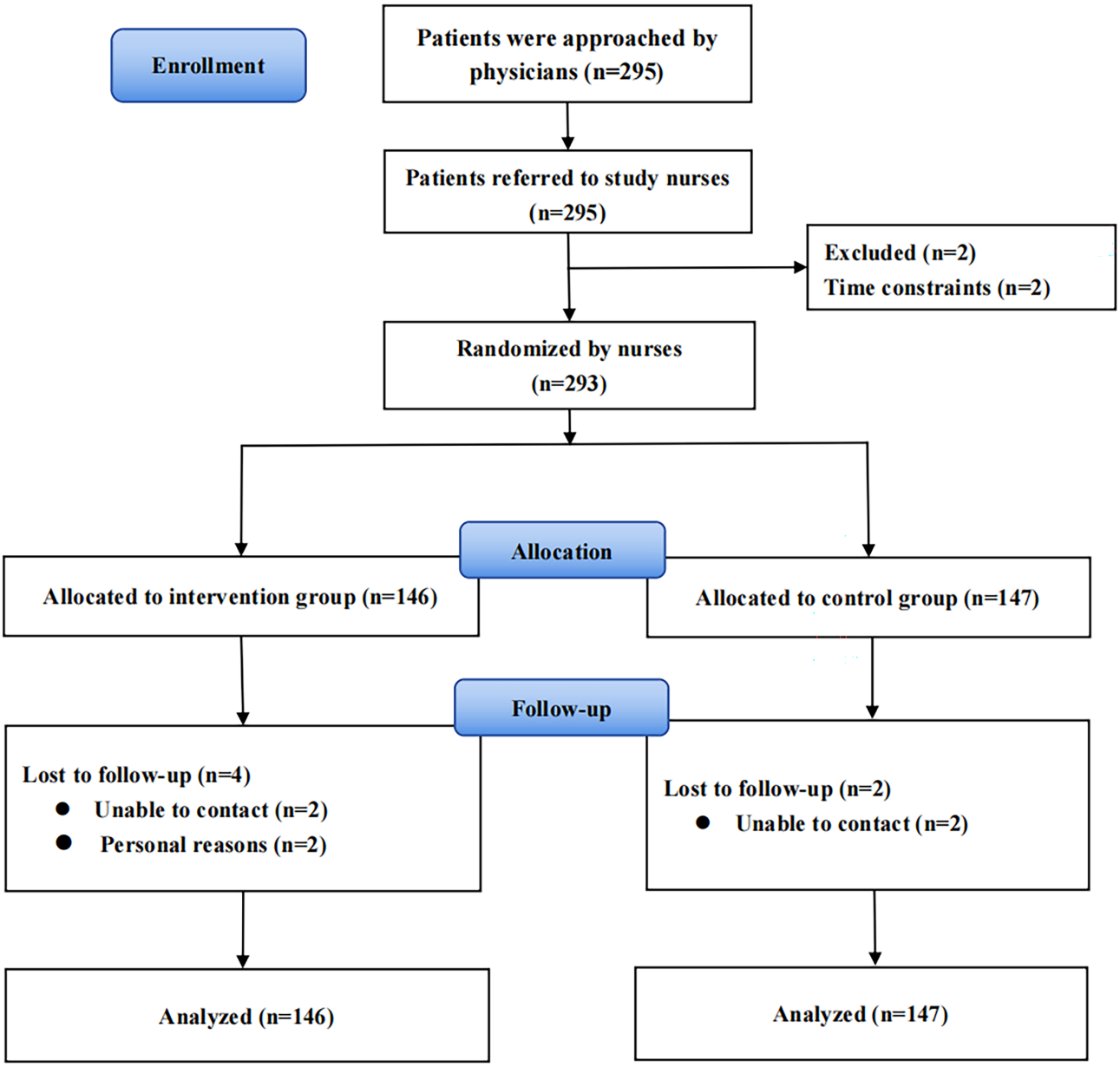
Between-group comparisons of other outcomes, including achievement of an HbA_{1c} reduction ≥0.5% and differences in medication use at 12 weeks, were performed using chi-square tests or Mann-Whitney *U* tests, as appropriate. Within-group comparisons of primary and secondary outcomes between baseline and 12 weeks were conducted using paired *t* tests or Wilcoxon signed-rank tests, as appropriate. All tests were 2-sided, with a *P* value <.05 considered statistically significant.

Results

Overview

A total of 295 participants met the inclusion criteria, of whom 293 (99.3%) consented and completed baseline data collection. Participants were randomized to the intervention group receiving digital structured education with nudging tools (n=146) or the control group receiving standard digital health education (n=147). Moreover, 4 participants in the intervention group and 2 in the control group were lost to follow-up, each completing fewer than 2 modules (Figure 1).

Figure 1. The study flowchart.



Baseline Characteristics

The 293 participants had a mean age of 49.19 (SD 10.02) years; 191 (65.2%) were male, and 281 (95.9%) were of Han ethnicity. Most participants were married (n=254, 86.7%), employed (n=211, 72%), and had completed at least junior high school education (n=157, 53.6%). About two-thirds (n=196, 66.9%) were covered by urban employee or urban-rural resident medical insurance. The mean duration of type 2 diabetes was 3.94 (SD 4.39) years, with a mean treatment duration of 3.36 (SD 4.32) years; 208 (71%) reported no diabetes-related complications. Overall, 184 (62.8%) had not previously used electronic devices for disease management. Baseline characteristics of the 2 groups are presented in Table 3.

XSL•FO
RenderX

Table . Characteristics of patients with type 2 diabetes mellitus (N=293).

Characteristics	IG[a] (n=146)	CG[b] (n=147)	*t* test (*df*) or chi-square (*df*)	*P* value
Age (y), mean (SD)	48.14 (10.25)	50.22 (9.72)	−1.784 (291)[c]	.08
Sex, n (%)			0.3 (1)[d]	.59
Male	93 (63.7)	98 (66.7)		
Female	53 (36.3)	49 (33.3)		
Ethnicity, n (%)			<0.001 (1)[d]	.99
Han Ethnicity	140 (95.9)	141 (95.9)		
Other	6 (4.1)	6 (4.1)		
Medical insurance, n (%)			0.7 (1)[d]	.41
UEI[e] and URI	101 (69.2)	95 (64.6)		
CMS[f] and others	45 (30.8)	52 (35.4)		
Educational background, n (%)			0.2 (1)[d]	.68
≤Primary school	66 (45.2)	70 (47.6)		
≥Junior high school	80 (54.8)	77 (52.4)		
Employment status, n (%)			1.6 (1)[d]	.21
Employed and others	110 (75.3)	101 (68.7)		
Retired	36 (24.7)	46 (31.3)		
Marital status, n (%)			1.4 (1)[d]	.24
Married	130 (89)	124 (84.3)		
Others	16 (11)	23 (15.7)		
Cohabitation group, n (%)			5.0 (3)[d]	.17
Spouses living together	110 (75.3)	98 (66.7)		
Children living together	18 (12.3)	29 (19.7)		
Parents living together	10 (6.9)	7 (4.8)		
Living alone or other	8 (5.5)	13 (8.8)		
Individual income per month (RMB[g]), n (%)			0.8 (1)[d]	.33
≤2000	53 (36.3)	61 (41.5)		
>2000	93 (63.7)	86 (58.5)		
Smoking, n (%)			2.2 (1)[d]	.14
Yes	67 (45.9)	55 (37.4)		
No	79 (54.1)	92 (62.6)		
Alcohol consumption, n (%)			0.3 (1)[d]	.60
Yes	63 (43.2)	59 (40.1)		
No	83 (58.9)	88 (59.9)		
Electronic health management, n (%)			1.9 (1)[d]	.17
Yes	60 (41.1)	49 (33.3)		
No	86 (58.9)	98 (66.7)		
Type 2 diabetes duration (y), mean (SD)	3.78 (4.42)	4.09 (4.37)	−0.559 (291)[c]	.55

XSL•FO
RenderX

Characteristics	IG[a] (n=146)	CG[b] (n=147)	*t* test (*df*) or chi-square (*df*)	*P* value
Medication duration (y), mean (SD)	3.21 (4.45)	3.52 (4.20)	−0.617 (291)[c]	.54
Antihyperglycemic medications, n (%)			2.4 (3)[d]	.50
0	25 (17.1)	24 (16.3)		
1	28 (19.2)	39 (26.5)		
2	69 (47.3)	64 (43.6)		
≥3	24 (16.4)	20 (13.6)		
Antihypertensive medications, n (%)			3.8 (2)[d]	.25
0	128 (87.7)	135 (91.8)		
1	16 (11)	12 (8.2)		
≥2	2 (1.3)	0 (0)		
Lipid-lowering medications, n (%)			5.7 (2)[d]	.06
0	96 (65.8)	112 (76.2)		
1	50 (34.2)	34 (23.1)		
≥2	0 (0)	1 (0.7)		
Complications[h], n (%)			1.3 (1)[d]	.26
Yes	38 (26)	47 (32)		
No	108 (74)	100 (68)		

[a]IG: intervention group.

[b]CG: control group.

[c]represents *t* test.

[d]represents chi-square values.

[e]UEI and URI: Urban employee insurance and residential insurance.

[f]CMS and others: Cooperative Medical Scheme and others.

[g]The exchange rate used was RMB 7.11=US $1.

[h]Complications: diabetes-related complications.

Between-Group Comparisons at 12 Weeks

At follow-up, the intervention group demonstrated greater reductions in HbA_{1c} (adjusted mean difference −0.38, 95% CI −0.68 to −0.09; *P*=.01), FBG (adjusted mean difference −0.75, 95% CI −1.27 to −0.44; *P*<.001), weight (adjusted mean difference −0.84, 95% CI −1.16 to −0.07; *P*=.03), BMI (adjusted mean difference −0.38, 95% CI −0.65 to −0.11; *P*=.01), systolic blood pressure (SBP; adjusted mean difference −2.71, 95% CI −4.62 to −0.79; *P*=.01), and diastolic blood pressure (DBP; adjusted mean difference −2.92, 95% CI −4.47 to −1.37, *P*<.001), TC (adjusted mean difference −0.27, 95% CI −0.48 to −0.05; *P*=.02), compared with the control group (Table 4). Moreover, the proportion of participants with an HbA_{1c} reduction ≥0.5% was significantly higher than in the control group (70.54% vs 57.82%; χ^2_1=5.2; *P*=.02).

XSL•FO
RenderX

Table . Baseline and 12-week outcomes, within-group changes, and adjusted between-group comparisons in participants with type 2 diabetes (N=293).

Outcomes	Baseline, mean (SD)	Week 12, mean (SD)	Within-group[a]		Adjusted between-group[b]	
			Mean difference (95% CI)	*P* value	Mean difference (95% CI)	*P* value
HbA_{1c}[c,d] (%)					−0.38 (−0.68 to −0.09)	.01
IG[e] (n=146)	8.19 (2.14)	6.76 (1.31)	−1.43 (−1.76 to −1.11)	<.001		
CG[f] (n=147)	8.11 (1.89)	7.12 (1.47)	−0.99 (−1.28 to −0.69)	<.001		
FBG[g] (mmol/L)					−0.75 (−1.27 to −0.44)	<.001
IG (n=146)	8.05 (3.23)	6.44 (1.32)	−1.61 (−2.08 to −1.14)	<.001		
CG (n=147)	8.43 (3.06)	7.29 (2.18)	−1.14 (−1.61 to −0.67)	<.001		
Weight (kg)					−0.84 (−1.61 to −0.07)	.03
IG (n=146)	66.57 (13.38)	64.79 (12.55)	−1.78 (−2.41 to −1.15)	<.001		
CG (n=147)	66.18 (11.07)	65.17 (10.74)	−1.01 (−1.53 to −0.49)	<.001		
BMI (kg/m^2)					−0.38 (−0.65 to −0.11)	.01
IG (n=146)	24.32 (3.35)	23.64 (3.10)	−0.68 (−0.90 to −0.47)	<.001		
CG (n=147)	24.47 (3.30)	24.12 (3.24)	−0.35 (−0.53 to −0.16)	<.001		
WC[h] (cm)					−0.44 (−1.19 to 0.31)	.25
IG (n=146)	85.90 (10.23)	84.51 (9.68)	−1.39 (−1.98 to −0.80)	<.001		
CG (n=147)	86.57 (8.83)	85.48 (8.50)	−1.09 (−1.62 to −0.57)	<.001		
SBP[i] (mm Hg)					−2.71 (−4.62 to −0.79)	.01
IG (n=146)	126.21 (16.48)	123.94 (12.41)	−2.27 (−3.98 to −0.56)	.01		
CG (n=147)	130.45 (15.87)	128.99 (13.33)	−1.46 (−3.16 to 0.25)	.09		
DBP[j] (mm Hg)					−2.92 (−4.47 to −1.37)	<.001
IG (n=146)	78.05 (11.35)	75.95 (8.83)	−2.11 (−3.32 to −0.90)	.001		
CG (n=147)	80.31 (10.31)	79.98 (9.96)	−0.33 (−1.73 to 1.06)	.64		
TC[k] (mmol/L)					−0.27 (−0.48 to −0.05)	.02
IG (n=146)	4.82 (1.41)	4.19 (1.26)	−0.63 (−0.84 to −0.43)	<.001		
CG (n=147)	4.66 (1.47)	4.33 (1.27)	−0.33 (−0.49 to −0.17)	<.001		

XSL•FO
RenderX

Outcomes	Baseline, mean (SD)	Week 12, mean (SD)	Within-group[a]		Adjusted between-group[b]	
			Mean difference (95% CI)	*P* value	Mean difference (95% CI)	*P* value
TG[l] (mmol/L)					−0.18 (−0.41 to 0.05)	.12
IG (n=146)	2.70 (2.43)	2.18 (1.53)	−0.52 (−0.80 to −0.25)	<.001		
CG (n=147)	2.55 (1.73)	2.27 (1.37)	−0.28 (−0.46 to −0.10)	.003		
HDL[m] (mmol/L)					0.08 (−0.01 to 0.16)	.08
IG (n=146)	1.28 (0.65)	1.42 (0.55)	0.14 (0.05 to 0.22)	.002		
CG (n=147)	1.24 (0.37)	1.33 (0.39)	0.08 (0.03 to 0.13)	.001		
LDL[n] (mmol/L)					−0.16 (−0.34 to 0.02)	.08
IG (n=146)	2.85 (1.05)	2.53 (1.05)	−0.32 (−0.48 to −0.16)	<.001		
CG (n=147)	2.97 (1.02)	2.74 (1.01)	−0.22 (−0.35 to −0.10)	<.001		
Behaviors[o]					5.90 (3.67 to 8.14)	<.001
IG (n=146)	34.10 (12.76)	49.13 (10.45)	15.03 (12.41 to 17.65)	<.001		
CG (n=147)	34.30 (14.19)	43.66 (9.80)	9.36 (7.03 to 11.69)	<.001		
Self - efficacy					2.90 (1.79 to 4.01)	<.001
IG (n=146)	25.42 (5.41)	31.28 (4.75)	5.86 (4.74 to 7.00)	<.001		
CG (n=147)	24.47 (5.52)	28.23 (5.09)	3.76 (2.78 to 4.75)	<.001		
Habit Strength					6.34 (4.32 to 8.37)	<.001
IG (n=146)	57.86 (13.08)	73.79 (8.81)	15.93 (13.69 to 18.16)	<.001		
CG (n=147)	55.78 (12.45)	67.03 (10.04)	11.26 (9.31 to 13.21)	<.001		

[a]Within-group mean differences represent changes from baseline to week 12.

[b]Adjusted between-group mean differences were estimated using ANCOVA, with baseline values and study center included as covariates.

[c]Primary outcome.

[d]HbA_{1c}: hemoglobin A_{1c}.

[e]IG: intervention group.

[f]CG: control group.

[g]FBG: fasting blood glucose.

[h]WC: waist circumference.

[i]SBP: systolic blood pressure.

[j]DBP: diastolic blood pressure.

[k]TC: total cholesterol.

[l]TG: triglycerides.

[m]HDL: high - density lipoprotein cholesterol.

[n]LDL: low - density lipoprotein cholesterol.

[o]Behaviors: diabetes self - management behaviors.

Compared with controls, the intervention group achieved greater improvements in self-management behaviors (adjusted mean difference 5.90, 95% CI 3.67 - 8.14; *P*<.001), self-efficacy (adjusted mean difference 2.90, 95% CI 1.79 - 4.01; *P*<.001), and habit strength (adjusted mean difference 6.34, 95% CI 4.32 - 8.37; *P*<.001). No significant between-group differences were observed for WC, triglycerides, HDL, or LDL (Table 4). No severe hypoglycemia occurred or nonsignificant difference was observed among medications among 2 groups (Table S1 in Multimedia Appendix 1).

XSL•FO
RenderX

Sensitivity analyses adjusting for baseline values only yielded results that were largely consistent with the primary analyses adjusted for both baseline values and study center. The direction and statistical significance of the primary and secondary outcomes remained unchanged, with the exception of weight, for which the between-group difference was attenuated and no longer statistically significant (Table S2 in Multimedia Appendix 1).

Within-Group Changes From Baseline to 12 Weeks

Both 2 groups showed significant improvements in HbA_{1c}, FBG, weight, BMI, WC, and blood lipids of TC, triglycerides, HDL, and LDL levels from baseline to 12 weeks. Blood pressure of SBP and DBP improved significantly only in the intervention group, with no significant changes observed in the control group. From baseline to follow-up, both groups demonstrated significant gains in self-management behaviors, self-efficacy, and habit strength (Table 4).

Discussion

Principal Findings

This multicenter randomized trial demonstrated that the DSE-BN program improved HbA_{1c}, FBG, weight, BMI, blood pressure, and TC compared with standard digital diabetes education. The intervention was also associated with improvements in diabetes self-management behaviors, self-efficacy, and habit strength, highlighting the potential value of an integrated digital structured education program incorporating behavioral nudge tools and complementary digital support features to simultaneously support reflective and automatic behavioral processes in diabetes self-management.

Interpretation of Findings

The observed improvements in glycemic outcomes in the intervention group, including a mean reduction of 0.38% in HbA_{1c} and decreases in FBG, are consistent with previous evidence suggesting that app-based diabetes education may support glycemic control by facilitating lifestyle modification and self-management behaviors [32]. Although the absolute HbA_{1c} reduction was modest, a significantly greater proportion of participants in the intervention group achieved an HbA_{1c} reduction of ≥0.5% compared with the control group, indicating that the intervention was capable of producing clinically meaningful improvements in a subset of participants. It is important to note that the mean HbA_{1c} reduction observed in this study falls below the 0.5% threshold commonly considered clinically meaningful at the individual level [33], and the relatively short follow-up period limits inference regarding long-term glycemic benefit. Nevertheless, modest HbA_{1c} reductions have been shown to be relevant at the population level. For example, findings from the UK Prospective Diabetes Study suggest that each 1% decrease in HbA_{1c} is associated with substantial reductions in diabetes-related complications [34]. While extrapolation should be interpreted with caution, the findings of this study suggest that even small improvements achieved through low-intensity, scalable digital interventions may have public health relevance if maintained over time.

In contrast, pharmacological agents like glucagon-like peptide-1 receptor agonists typically achieve HbA_{1c} reductions exceeding 1.2% [35]. Improvements observed in the control group are consistent with evidence that digital interventions can enhance cognitive-behavioral processes and health literacy [36,37]. Notably, although the intervention group received a more streamlined educational curriculum than the control group, it achieved superior improvements across several metabolic and behavioral outcomes. This finding may reflect the added value of the integrated intervention package, which combined digital structured education with multiple behavioral support components, including instructional videos, interactive quizzes, preparatory and postmodule review questions with immediate feedback, peer interaction, automated reminders, rule-based tailored feedback, data-driven early-warning alerts, and home-based behavioral nudge tools. Together, these integrated features may have reinforced both intentional self-management and habit formation, thereby supporting long-term engagement beyond education alone. Importantly, no significant between-group differences were observed in the use of antihyperglycemic medications at the 12-week follow-up, and no severe hypoglycemic events were reported, supporting the safety of the intervention and indicating that the observed effects were unlikely to be driven by pharmacological changes.

The design of the DSE-BN program was informed by dual-process theory, which proposes that health behaviors are influenced by both reflective and automatic behavioral processes [38]. Rather than relying on education alone, the intervention integrated digital structured education with complementary digital support features and behavioral nudge tools to reinforce healthy behaviors in participants' home environments. By simultaneously strengthening reflective processes through structured education and supporting automatic processes through home-based behavioral nudge tools and digital behavioral support, the intervention may have facilitated the transition from intentional self-management to more consistent daily habits. These behavioral changes may, in turn, have contributed to the observed improvements in weight, BMI, blood pressure, and TC through better adherence to physical activity, medication use, and healthy dietary behaviors [39]. Consistent with this interpretation, no significant between-group differences were observed in the use of antihypertensive or lipid-lowering medications at the 12-week follow-up, suggesting that these cardiometabolic benefits were also unlikely to be driven by medication changes. Although this trial evaluated the overall effectiveness of an integrated intervention package rather than the independent effects of individual components, the findings suggest that integrating behavioral nudge tools within a broader digital structured education program may provide additional benefits beyond education alone.

Psychological enhancements in self-efficacy, self-management, and habit strength echo previous research [40], with structured education addressing personal and environmental factors per social cognitive theory [41]. From the perspective of dual-process theory, the integrated intervention may have further strengthened automatic behavioral processes while reinforcing the effects of reflective learning [26,42]. This integration of reflective and automatic behavioral processes may have

XSL•FO
RenderX

facilitated the transition from conscious goal-setting to more habitual self-management behaviors, providing a theory-informed approach to promoting long-term diabetes self-management [43,44]. Future studies should investigate the neural mechanisms underlying behavioral changes following this dual-process intervention and evaluate the relative contribution of individual intervention components [45].

Consistent with American Diabetes Association and European Association for the Study of Diabetes guidelines recommending HbA_{1c} levels below 7% to mitigate complications [46], this nonpharmacological intervention serves as an adjunct for patients near targets or in combination with medications, enhancing accessibility in real-world practice.

Implications for Digital Health Practice

This multicenter randomized trial demonstrates that an integrated digital structured education program incorporating behavioral nudge tools and complementary behavioral support strategies can enhance diabetes self-management by simultaneously supporting reflective and automatic behavioral processes. The intervention was associated with improvements in HbA_{1c}, weight, BMI, blood pressure, TC, self-management behaviors, self-efficacy, and habit strength, addressing common limitations of digital interventions such as declining engagement and suboptimal adherence.

By extending support into patients' home environments, this low-cost, technology-enabled approach shows strong potential for scalable, sustainable deployment in routine diabetes care, particularly in resource-constrained settings. Its favorable safety profile and independence from pharmacological changes further support its feasibility as an adjunct to standard treatment. These findings underscore the potential value of integrated, theory-informed digital interventions that integrate reflective learning with support for automatic behavior formation, providing a scalable approach to sustainable diabetes self-management.

Study Limitations

Several limitations of this study should be acknowledged. Blinding of participants was infeasible due to the intervention's behavioral nature, which may have introduced expectation bias. Outcomes were assessed only at 12 weeks, limiting inference regarding the sustainability of intervention effects over the longer term. Accordingly, improvements observed across multiple secondary outcomes should be interpreted cautiously. This relatively short follow-up period may also partly explain the attenuation of the between-group difference in weight observed in sensitivity analyses, suggesting that weight-related outcomes—given their multifactorial determinants—may be more sensitive to model specification than glycemic and behavioral outcomes. Recruitment from a single Chinese city limited population diversity, and the educated, urban, insured sample may reduce generalizability, particularly regarding socioeconomic, rural-urban, and digital literacy disparities that could exacerbate exclusion. In addition, because this study evaluated the overall effectiveness of an integrated intervention package, it was not designed to isolate the effects of individual intervention components. Future dismantling studies are warranted to identify the active components and optimize the design of theory-informed digital behavioral interventions. Despite these limitations, the multicenter randomized design, standardized intervention delivery, and the theory-informed dual-process framework bolster internal validity. Future studies with longer follow-up and more diverse populations are warranted to confirm and extend these findings.

Conclusions

This study demonstrates that an integrated digital structured education program incorporating behavioral nudge tools can improve glycemic control, weight, BMI, blood pressure, TC, and self-management in adults with type 2 diabetes. Beyond clinical outcomes, the intervention was associated with improvements in psychological and behavioral factors, including self-efficacy and habit formation, which are critical for supporting short- to medium-term adherence in chronic disease management. These findings suggest that an integrated digital structured education program incorporating behavioral nudge tools provides a scalable, theory-informed approach that bridges intentional behavior change and habit formation, with potential to support sustainable diabetes self-management in routine care.

Acknowledgments

The authors sincerely thank all participants for their valuable contributions to this study.

The authors confirm that no generative AI tools were used in the writing or preparation of this manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (Youth Program, grant 82304262), the Hainan Provincial Natural Science Foundation of China (grant 825RC766), and the China Scholarship Council (grant 202408460019).

Data Availability

The deidentified data supporting the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available because they contain information that could compromise participant privacy and are subject to the conditions of the informed consent provided by the participants.

XSL•FO
RenderX

Authors' Contributions

Investigation, data curation, formal analysis, writing – original draft: YL
Formal analysis, writing – original draft, writing – review & editing: Y Zeng
Investigation, data curation: KC, ZC, C Ye, Y Zhou, QZ, C Yu
Writing – review & editing: VWW, SS, BL
Conceptualization, methodology, resources, supervision, funding acquisition, writing – original draft, writing – review & editing: XJ

The authors confirm that no generative artificial intelligence (AI) tools were used in the writing or preparation of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary results.
[DOCX File, 23 KB - mhealth_v14i1e91407_app1.docx]

Checklist 1
CONSORT checklist.
[PDF File, 144 KB - mhealth_v14i1e91407_app2.pdf]

References

1. Genitsaridi I, Salpea P, Salim A, et al. 11th edition of the IDF Diabetes Atlas: global, regional, and national diabetes prevalence estimates for 2024 and projections for 2050. Lancet Diabetes Endocrinol 2026 Feb;14(2):149-156. [doi: 10.1016/S2213-8587(25)00299-2] [Medline: 41412135]
2. Jiang XJ, Jiang H, Chen Y, et al. The effectiveness of a self-efficacy-focused structured education program (SSEP) in improving metabolic control and psychological outcomes of type 2 diabetes patients: a 12-month follow-up of a multicenter randomized controlled trial. Diabetes Metab Syndr Obes 2021;14:305-313. [doi: 10.2147/DMSO.S290029] [Medline: 33519219]
3. Strawbridge LM, Lloyd JT, Meadow A, Riley GF, Howell BL. One-year outcomes of diabetes self-management training among Medicare beneficiaries newly diagnosed with diabetes. Med Care 2017 Apr;55(4):391-397. [doi: 10.1097/MLR.0000000000000653] [Medline: 27753746]
4. Du JQ, Li YT, Li HY, Huang Y. Status quo and influencing factors of self-management behaviors in patients with type 2 diabetes mellitus. Chin Nurs Res 2022;16(36):2904-2908 [FREE Full text]
5. Ma XY. Self-managed family support in rural type 2 diabetes patients based on social support theory. AMU 2024. [doi: 10.26921/d.cnki.ganyu.2024.001250]
6. Bargh JA. The Four Horsemen of Automaticity: Awareness, Intention, Efficiency, and Control in Social Cognition: Psychology Press; 2014:1-40.
7. Hollands GJ, Marteau TM, Fletcher PC. Non-conscious processes in changing health-related behaviour: a conceptual analysis and framework. Health Psychol Rev 2016 Dec;10(4):381-394. [doi: 10.1080/17437199.2015.1138093] [Medline: 26745243]
8. Moors A, De Houwer J. Automaticity: a theoretical and conceptual analysis. Psychol Bull 2006 Mar;132(2):297-326. [doi: 10.1037/0033-2909.132.2.297] [Medline: 16536645]
9. Agarwal P, Mukerji G, Desveaux L, et al. Mobile app for improved self-management of type 2 diabetes: multicenter pragmatic randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 10;7(1):e10321. [doi: 10.2196/10321] [Medline: 30632972]
10. Jia W, Zhang P, Zhu D, et al. Evaluation of an mHealth-enabled hierarchical diabetes management intervention in primary care in China (ROADMAP): a cluster randomized trial. PLoS Med 2021 Sep;18(9):e1003754. [doi: 10.1371/journal.pmed.1003754] [Medline: 34547030]
11. Jiang XJ, Liu YL, Li MZ. Structured education in patients with diabetes: a literature review. Chin J Nurs Educ 2019;16(12):897-900.
12. Jiang XJ, Jiang H, Lu YH, et al. The effectiveness of a self-efficacy-focused structured education programme on adults with type 2 diabetes: a multicentre randomised controlled trial. J Clin Nurs 2019 Sep;28(17-18):3299-3309. [doi: 10.1111/jocn.14908] [Medline: 31066115]
13. Jiang X, Wang J, Lu Y, Jiang H, Li M. Self-efficacy-focused education in persons with diabetes: a systematic review and meta-analysis. Psychol Res Behav Manag 2019;12:67-79. [doi: 10.2147/PRBM.S192571] [Medline: 30774486]

XSL•FO
RenderX

14. Boels AM, Rutten G, Zuithoff N, de Wit A, Vos R. Effectiveness of diabetes self-management education via a smartphone application in insulin treated type 2 diabetes patients - design of a randomised controlled trial ('TRIGGER study'). BMC Endocr Disord 2018 Oct 22;18(1):74. [doi: 10.1186/s12902-018-0304-9] [Medline: 30348142]
15. Li R, Zhang Y, Cai X, et al. The nudge strategies for weight loss in adults with obesity and overweight: a systematic review and meta-analysis. Health Policy 2021 Dec;125(12):1527-1535. [doi: 10.1016/j.healthpol.2021.10.010] [Medline: 34772518]
16. Thaler RH. Misbehaving The Making of Behavioral Economics: WW Norton & Company; 2015.
17. Sunstein CR. The ethics of nudging. SSRN 2015. [doi: 10.2139/ssrn.2526341]
18. Wu Q, Zhang R, Tao L, et al. Nudge theories and strategies influencing adult health behaviors and outcomes in COPD management: a systematic review. Front Public Health 2024;12:1404590. [doi: 10.3389/fpubh.2024.1404590] [Medline: 39564365]
19. Hollands GJ, Bignardi G, Johnston M, et al. The TIPPME intervention typology for changing environments to change behaviour. Nat Hum Behav 2017;1(8):140. [doi: 10.1038/s41562-017-0140]
20. Zhang Y, Yang Y, Huang Q, Zhang Q, Li M, Wu Y. The effectiveness of lifestyle interventions for diabetes remission on patients with type 2 diabetes mellitus: a systematic review and meta-analysis. Worldviews Evid Based Nurs 2023 Feb;20(1):64-78. [doi: 10.1111/wvn.12608] [Medline: 36480153]
21. Ren JWD. Health governance, top-level design, policy tools and practical experience. J Tianjin Admin Inst 2019;21(3):86-95.
22. Zhao YL, Yao XM, Gao RM. Advancements in the Application of Structured Education Programs in Type 2 Diabetes. Contemp Med Chin 2020;20(26):192-194 [FREE Full text]
23. Chiness Diabetes Society, Zhu D. Guideline for the prevention and treatment of type 2 diabetes mellitus in China(2020 edition). Practical Internal Medicine 2021;41:08. [doi: 10.3760/cma.j.cn311282-20210304-00142]
24. Liu Y, Jiang X, Jiang H, Lin K, Li M, Ji L. A culturally sensitive nurse-led structured education programme in patients with type 2 diabetes. Int J Nurs Pract 2019 Oct;25(5):e12757. [doi: 10.1111/ijn.12757] [Medline: 31261444]
25. China Copyright Protection Center. Self-management behavior illustrations [in chinese]. 2024 assignee.
26. Jiang X, Li M, Zhang C. A type of tableware to aid in diet management. 2021 Jun 25 inventors URL: https://patents.google.com/patent/CN213524739U/zh [accessed 2026-08-19]
27. Wei J. Research on Relationship between Diabetes Knowledge, Self-Efficacy and Self-Management among Rural Elderly Patients: Hangzhou Normal University; 2013.
28. English evaluation tools. Self-Management Resource Center. URL: https://selfmanagementresource.com/resources/evaluation-tools/english-evaluation-tools/ [accessed 2026-08-19]
29. Gardner B, de Bruijn GJ, Lally P. A systematic review and meta-analysis of applications of the self-report habit index to nutrition and physical activity behaviours. Ann Behav Med 2011 Oct;42(2):174-187. [doi: 10.1007/s12160-011-9282-0] [Medline: 21626256]
30. Lin Y, Ye C, Xing S, et al. Habit strength as a partial explanation of the association between self-efficacy and self-management behaviors in type 2 diabetes: a multicenter cross-sectional study. Front Public Health 2026;14:1870512. [doi: 10.3389/fpubh.2026.1870512] [Medline: 42415764]
31. Wan QQ, Shang SM, Lai XB, Pan J. Study on the reliability and validity of summary of diabetes self-care activities for type 2 diabetes patients. Chin J Pract Nurs 2008;24(7):26-27 [FREE Full text]
32. Wu Y, Yao X, Vespasiani G, et al. Mobile app-based interventions to support diabetes self-management: a systematic review of randomized controlled trials to identify functions associated with glycemic efficacy. JMIR Mhealth Uhealth 2017 Mar 14;5(3):e35. [doi: 10.2196/mhealth.6522] [Medline: 28292740]
33. Santos S, Pentzek M, Altiner A, et al. HbA1c as a shared treatment goal in type 2 diabetes? A secondary analysis of the DEBATE trial. BMC Prim Care 2023 May 13;24(1):115. [doi: 10.1186/s12875-023-02067-9] [Medline: 37173620]
34. Ferreira SRG. Revisiting clinical trials on glycemic control and cardiovascular risk. Diabetol Metab Syndr 2009 Sep 28;1(1):12. [doi: 10.1186/1758-5996-1-12] [Medline: 19825201]
35. Gronroos NN, Swift C, Frazer MS, et al. Real-world HbA1c changes among type 2 diabetes mellitus patients initiating treatment with a 1.0 mg weekly dose of semaglutide for diabetes. J Health Econ Outcomes Res 2024;11(2):118-124. [doi: 10.36469/001c.124111] [Medline: 39507603]
36. Lu D, Jiang W, Ai D, et al. An interventional study of digital therapy on blood glucose and weight management in Chinese patients with type 2 diabetes. J Diabetes Investig 2025 Mar;16(3):475-480. [doi: 10.1111/jdi.14376] [Medline: 39825624]
37. Ostini R, Kairuz T. Investigating the association between health literacy and non-adherence. Int J Clin Pharm 2014 Feb;36(1):36-44. [doi: 10.1007/s11096-013-9895-4] [Medline: 24293337]
38. Verplanken B, Orbell S. Reflections on past behavior: a self-report index of habit strength. J Appl Soc Psychol 2006;33(6):1313-1330. [doi: 10.1111/j.1559-1816.2003.tb01951.x]
39. Long T, Zhang Y, Zhang Y, et al. Making diet management easier: the effects of nudge-based dietary education and tableware in individuals with both T2DM and overweight/obesity: a 2 × 2 cluster randomized controlled trial. Nutrients 2025 May 3;17(9):1574. [doi: 10.3390/nu17091574] [Medline: 40362883]
40. Zeidi IM, Morshedi H, Alizadeh Otaghvar H. A theory of planned behavior-enhanced intervention to promote health literacy and self-care behaviors of type 2 diabetic patients. J Prev Med Hyg 2020 Dec;61(4):E601-E613. [doi: 10.15167/2421-4248/jpmh2020.61.4.1504] [Medline: 33628967]

XSL•FO
RenderX

41. Mehta G, Hirji A. The outcome of structured education in adults with type 2 diabetes mellitus and substance use disorder: a literature review. Can J Diabetes 2020 Aug;44(6):487-493. [doi: 10.1016/j.jcjd.2020.05.010] [Medline: 32792102]
42. Payne JB, Moyer A. Social psychological theories and nudges as tools for health promotion. Am J Bioeth 2019 May;19(5):74-76. [doi: 10.1080/15265161.2019.1588417] [Medline: 31090526]
43. Corbit LH, Balleine BW. The role of prelimbic cortex in instrumental conditioning. Behav Brain Res 2003 Nov 30;146(1-2):145-157. [doi: 10.1016/j.bbr.2003.09.023] [Medline: 14643467]
44. Yin HH, Knowlton BJ, Balleine BW. Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. Eur J Neurosci 2004 Jan;19(1):181-189. [doi: 10.1111/j.1460-9568.2004.03095.x] [Medline: 14750976]
45. de Wit S, Corlett PR, Aitken MR, Dickinson A, Fletcher PC. Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. J Neurosci 2009 Sep 9;29(36):11330-11338. [doi: 10.1523/JNEUROSCI.1639-09.2009] [Medline: 19741139]
46. American Diabetes Association Professional Practice Committee. 6. Glycemic goals and hypoglycemia: standards of care in diabetes-2025. Diabetes Care 2025 Jan 1;48(1 Suppl 1):S128-S145. [doi: 10.2337/dc25-S006] [Medline: 39651981]

Abbreviations

C-SED: Chinese version of the Self-Efficacy for Diabetes
CONSORT: Consolidated Standards of Reporting Trials
DBP: diastolic blood pressure
DSE-BN: digital structured education with behavioral nudge tools
FBG: fasting blood glucose
HbA_{1c}: hemoglobin A_{1c}.
HDL: high-density lipoprotein cholesterol
ICC: intraclass correlation coefficient
LDL: low-density lipoprotein cholesterol
PASS: Power Analysis and Sample Size
SBP: systolic blood pressure
SMASS: self-management support system
SRHI: Self-Report Habit Index
TC: total cholesterol
WC: waist circumference

Edited by L Buis; submitted 23.Jan.2026; peer-reviewed by S Castro-Barquero, YA Rias; revised version received 19.Jul.2026; accepted 24.Jul.2026; published 01.Sep.2026.

Please cite as:
Lin Y, Zeng Y, Chen K, Chen Z, Ye C, Zhou Y, Zhou Q, Yu C, Wu VX, Seidu S, Li B, Jiang X
Digital Structured Education With Behavioral Nudge Tools for Adults With Type 2 Diabetes: Multicenter Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e91407
URL: https://mhealth.jmir.org/2026/1/e91407
doi:10.2196/91407

© Yan Lin, Yingchun Zeng, Kaining Chen, Zongcun Chen, Caihua Ye, Ying Zhou, Qiwei Zhou, Chengying Yu, Vivien Xi Wu, Samuel Seidu, Bin Li, Xinjun Jiang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 1.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Radiographic Knee Osteoarthritis Identification Using Wearable Plantar Pressure: Secondary Cross-Sectional Analysis of Interpretable Framework Integrating Biomechanical Descriptors and the Multiscale Synergy Index

Guanyu Xin, BEng; Hongbo Yao, PhD; Jingyuan Qiao, MEng; Huilin Yao, MEng; Siyuan Li, BEng; Lu Liu, MEng; Jingyi Zhang, MEng; Ke Qin, MEng; Lin Shu, PhD

School of Future Technology, South China University of Technology, 777 Xingye Avenue East, Panyu District, Guangzhou, Guangdong, China

Corresponding Author:
Hongbo Yao, PhD
School of Future Technology, South China University of Technology, 777 Xingye Avenue East, Panyu District, Guangzhou, Guangdong, China

Abstract

Background: Radiographic knee osteoarthritis (ROA) is associated with abnormal plantar loading and altered gait coordination. Insole-based plantar-pressure sensing offers a practical wearable approach to ROA identification, but existing methods typically use either conventional biomechanical descriptors or end-to-end temporal models. An interpretable subject-level framework that captures both local abnormalities and distributed coordination is, therefore, needed.

Objective: This study aimed to develop an interpretable subject-level framework that combines conventional biomechanical descriptors with coordination-level plantar information to identify ROA.

Methods: In secondary cross-sectional analysis, using a previously reported Pearl River Osteoarthritis Cohort plantar-pressure dataset from our group, we analyzed 92 participants (43 with ROA and 49 without ROA; mean age 63.0, SD 8.4 y). The source cohort was recruited from clinical and nearby community sources at Zhujiang Hospital of Southern Medical University, Guangzhou, China, between January 2022 and February 2023. We developed an interpretable participant-level classification framework integrating conventional biomechanical descriptors with a coordination-level Plantar Synergy Index (PSI). Participant-wise stratified split allocated 72 participants for 5-fold cross-validation, and 20 for locked independent evaluation. Feature selection used Fisher-score ranking (k=50). All point estimates are accompanied by 2000-iteration bootstrap 95% CIs, and paired model comparisons used DeLong tests. All statistical tests used α=.05. Additional analyses examined feature stability, phase-resolved PSI patterns, robustness to perturbation and design variation, and comparisons with representative end-to-end temporal baselines.

Results: The fusion model achieved the best primary classification performance on the independent test set, with an F_1-score of 0.842 (95% CI 0.615 - 1.000) and accuracy of 0.850 (95% CI 0.700 - 1.000). It also showed balanced performance across additional metrics, with area under the receiver operating characteristic curve (AUROC) of 0.818 (95% CI 0.576 - 1.000), precision of 0.800 (95% CI 0.538 - 1.000), recall of 0.889 (95% CI 0.636 - 1.000), and specificity of 0.818 (95% CI 0.571 - 1.000). DeLong testing showed that fusion outperformed PSI-only at the AUROC level (P=.02). Stable retained descriptors were dominated by biomechanical features, with a reproducible PSI component. Phase-resolved analyses suggested distributed stance-phase coordination differences. Robustness and baseline comparisons supported the stability of the framework, but the wide CIs indicate that the findings remain preliminary.

Conclusions: This study introduces an interpretable participant-level representation that integrates local biomechanical loading with multiscale interregional temporal coordination, extending plantar-pressure approaches based on summary descriptors or end-to-end models. Across evaluated settings, the framework showed competitive identification and preliminary robustness. Local biomechanical descriptors constituted the main discriminative basis, whereas PSI contributed a reproducible coordination-level component whose incremental predictive value remains to be established. This broadens plantar-pressure analysis from local loading to distributed coordination. External multicenter prospective validation is required before evaluating the framework as an imaging adjunct for functional assessment and longitudinal monitoring.

(JMIR Mhealth Uhealth 2026;14:e101026) doi:10.2196/101026

KEYWORDS

radiographic knee osteoarthritis; plantar pressure; gait biomechanics; Plantar Synergy Index; subject-level identification; feature-selection stability

XSL•FO
RenderX

Introduction

Osteoarthritis is a chronic disease characterized by whole-joint degeneration. It causes pain, reduced activity, and functional decline, and is a major global cause of disability [1]. Knee osteoarthritis is one of the most common forms, and its burden continues to rise, highlighting the need for scalable and repeatable assessment tools that can capture functional change and support rehabilitation management and longitudinal follow-up [2].

In clinical practice, imaging—especially X-ray—is widely used to assess knee osteoarthritis, and the Kellgren-Lawrence (K-L) scale is one of the most common radiographic grading systems [3]. Here, we focus on radiographic knee osteoarthritis (ROA), using radiographic status as the reference standard for classification-oriented model development and evaluation. However, radiographic grades mainly reflect static structural changes and do not directly capture functional abnormalities during walking and daily activities. Moreover, a mismatch between imaging findings and symptoms, including pain, has been reported repeatedly [4,5]. Together, these limitations suggest that imaging alone may be insufficient for rehabilitation-oriented functional assessment, motivating the exploration of wearable functional biomarkers that reflect gait-related impairment [1].

Gait biomechanics is closely linked to the functional manifestations of knee osteoarthritis [6]. Recent studies have further linked ROA-related functional impairment to altered plantar-pressure behavior, gastrocnemius muscle-tendon-unit stiffness, and 3D gait biomechanics [7,8]. Plantar pressure is particularly informative because it reflects both load distribution and gait control, providing interpretable measures such as pressure distribution, peak pressure, and center-of-pressure progression. Wearable insole systems are especially attractive due to their portability and ease of deployment, enabling data collection in both clinical and daily-life settings [9-12]. Earlier intelligent-footwear research demonstrated daily-life monitoring of spatial and temporal plantar pressure, in-shoe conditions, center-of-pressure trajectories, and acceleration in people with diabetes [12]. Recent digital insole studies further suggest promising potential for detecting ROA-related gait signatures [13].

Current plantar-pressure modeling for knee osteoarthritis largely follows 2 paradigms, each with distinct strengths and limitations. The first relies on handcrafted biomechanical features, which are typically interpretable but often emphasize localized loading characteristics or outcome-level summaries. As a result, they may not adequately capture how plantar regions coordinate and covary over time during gait. The second applies end-to-end deep learning directly to pressure sequences, which is more flexible for temporal pattern learning but can be less stable and less interpretable in medical settings with limited sample sizes [14]. However, within plantar-pressure–based ROA identification, an important unresolved challenge is how to represent pressure signals in a way that preserves biomechanical interpretability while also capturing distributed interregional coordination beyond conventional summary descriptors.

Previous plantar-pressure studies have identified ROA-related abnormalities in anatomically meaningful foot regions, including reduced hallux loading, altered heel pressure distribution, and increased mediolateral center-of-pressure excursion [15-18]. However, most studies have relied on summary-level biomechanical descriptors and have not examined how plantar regions coordinate with one another over the gait cycle. On the same 92-participant Pearl River Osteoarthritis Cohort (PROC) cohort, previous work by our group distinguished ROA from non-ROA using handcrafted features at the window level [19] and predicted functional performance at the participant level [20], but neither incorporated interregional coordination structure. Whether coordination-level information captured from the same plantar-pressure signals provides discriminative power beyond conventional biomechanical features therefore remains an open question—one that the growing adoption of wearable-based biomechanical assessment makes increasingly timely [21].

This evidence, across both the broader literature and our own previous work, points to a specific representation gap in plantar-pressure–based ROA classification, that is, existing frameworks either provide interpretable biomechanical anchors or flexible temporal modeling, but none explicitly incorporate interregional coordination structure. In this study, we treat conventional biomechanical descriptors and interregional coordination patterns as 2 complementary components of plantar-pressure representation. Accordingly, in this secondary methodological analysis of a previously reported plantar-pressure cohort, we aimed to determine whether multiscale plantar coordination information, quantified by the Plantar Synergy Index (PSI), provides participant-level discriminatory information beyond conventional biomechanical descriptors for ROA identification. We hypothesized that combining PSI with biomechanical descriptors would improve threshold-dependent classification performance while preserving interpretability under a strict participant-wise, leakage-controlled evaluation protocol. To test this hypothesis, we developed an interpretable framework and evaluated it using a locked independent test set, feature-stability analyses, phase-resolved PSI visualization, robustness analyses, and comparisons with representative end-to-end temporal baselines.

Methods

Study Design, Data Source, and Participants

Study Design and Data Source

This study was a secondary cross-sectional analysis of a plantar-pressure dataset previously reported in related studies [19,20]. The analytic dataset was derived from the PROC and included 92 participants, comprising 43 participants with ROA and 49 participants with non-ROA. ROA was defined according to the original radiographic criterion used in the source study, namely modified K-L grade ≥2.

The source cohort was recruited from the PROC study at Zhujiang Hospital of Southern Medical University, Guangzhou, China, between January 2022 and February 2023. Participants in the source PROC cohort were recruited from osteoarthritis

XSL•FO
RenderX

outpatient clinics and nearby communities; therefore, the recruitment procedure is best characterized as nonprobability clinical or community-based recruitment rather than random population sampling. No new participant recruitment or additional plantar-pressure data collection was conducted for the secondary analysis of this study. This study used the available bilateral in-shoe plantar-pressure recordings and ROA status for participant-level ROA identification.

A participant-wise stratified split was used for model development and final evaluation, allocating 72 participants to the development set and 20 participants to the locked independent test set. The *Journal Article Reporting Standards* (JARS) participant flowchart documenting cohort flow, data completeness, and the development or test split is provided in Figure S1 in Multimedia Appendix 1.

Eligibility and Participant Characteristics

The original source cohort involved adults with knee osteoarthritis who were able to walk independently, as described in the related cohort report [20]. For the analysis of this study, inclusion required available bilateral wearable in-shoe plantar-pressure recordings, available ROA or non-ROA status, and complete demographic and anthropometric variables, including age, height, weight, and BMI.

All 92 participants in the available source dataset met these analytic inclusion requirements. Baseline demographic and anthropometric characteristics stratified by ROA status are reported in Table S1 in Multimedia Appendix 1. The ROA group was older than the non-ROA group (mean 66.6, SD 7.5 y vs mean 59.8, SD 7.9 y; Cohen d=−0.885; Mann-Whitney P<.001) and had higher BMI (mean 25.4, SD 3.6 kg/m^2 vs mean 23.7, SD 3.1 kg/m^2; Cohen d=−0.504; Welch t test P=.02). Height and weight did not differ significantly between groups. Individual-level sex distribution and K-L grade were not available in the dataset of this analysis and are therefore reported descriptively based on the available source information rather than used as adjustment covariates.

Measures and Covariates

The primary outcome was binary ROA status, defined as modified K-L grade ≥2 for the classification task of this study. The primary predictors were participant-level plantar-pressure descriptors derived from conventional biomechanical features, PSI-derived plantar coordination features, or their fusion. Detailed preprocessing, windowing, PSI construction, biomechanical feature definitions, and participant-level aggregation procedures are described in the subsections under the Methods section and Multimedia Appendices 2-4. Available demographic and anthropometric covariates included age, height, weight, and BMI.

Sample Size, Power, Precision, and Missing Data

The analytic sample size was determined by the available PROC source dataset rather than by a new prospective sample-size calculation. The final analytic sample included all 92 available participants. Given the modest sample size, especially the 20-participant locked independent test set, we emphasized estimation precision rather than relying on point estimates alone. CIs and paired area under the receiver operating characteristic curve (AUROC) comparisons were used to support precision-aware interpretation, as described in the Statistical Analysis subsection.

No data were missing; all 92 participants completed the full bilateral plantar-pressure acquisition, and no pressure windows were missing. The 4 demographic and anthropometric variables used in this analysis, namely age, height, weight, and BMI, were complete for all participants.

Overlap Statement With Previous Cohort Studies

This study uses the same 92-participant PROC plantar-pressure dataset previously used in related studies by our group [19,20]. It overlaps with those studies in participant cohort, wearable in-shoe pressure-sensing system, bilateral raw plantar-pressure recordings, and available clinical or radiographic source information. There is also partial authorship overlap with the previous cohort publications.

However, this study differs in analytical objective, feature representation, outcome modeling, and evaluation protocol. Previous work focused on conventional biomechanical descriptors for window-level ROA classification [19] or participant-level functional-performance prediction [20]. In contrast, this study evaluates whether multiscale PSI-derived plantar coordination descriptors provide complementary participant-level information when combined with conventional biomechanical descriptors under a strict participant-wise, leakage-controlled classification protocol. A structured comparison with previous plantar-pressure and gait-based studies is provided in Multimedia Appendix 5.

Accordingly, the contribution of this manuscript is methodological and representational. No new cohort, device, deployment setting, or longitudinal validation is claimed.

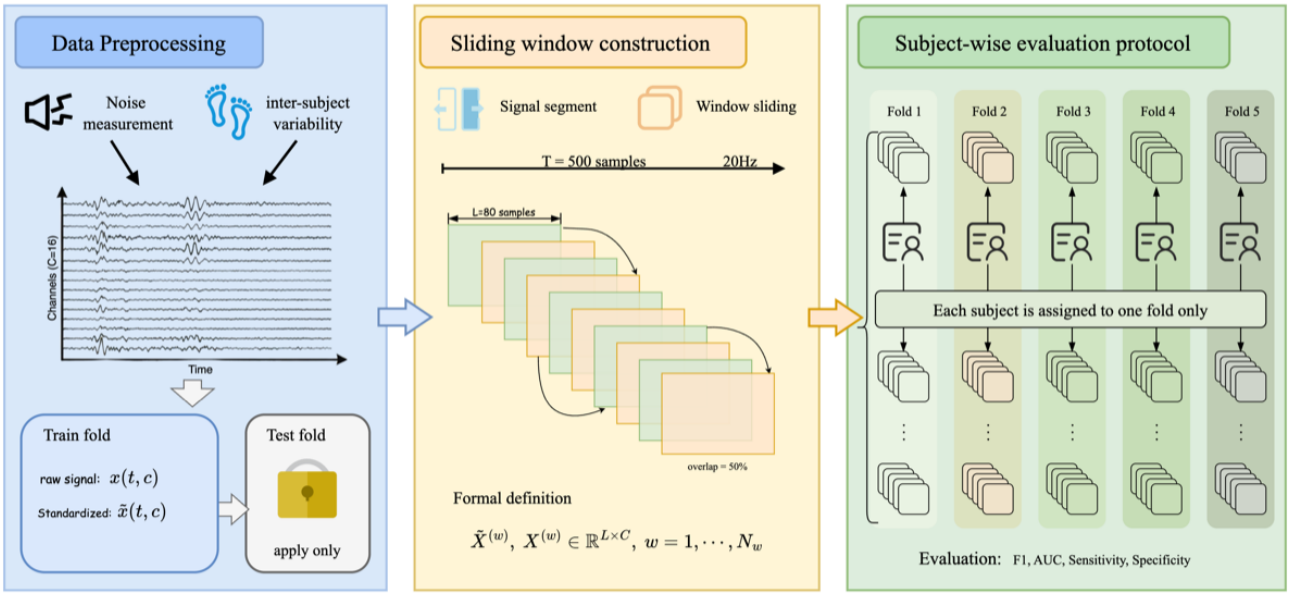
Plantar-Pressure Acquisition and Preprocessing

Plantar-pressure data were collected using a bilateral in-shoe wearable system, with 1 insole per foot. Each insole contained 8 discrete pressure sensors, yielding 16 channels sampled at 20 Hz. The detailed 16-channel sensor-to-anatomy mapping and region definitions are provided in Multimedia Appendix 2.

Figure 1 illustrates the leakage-controlled preprocessing and evaluation pipeline used in this study. The full workflow consisted of following 3 steps: participant-wise data splitting, construction of parallel signal representations, and sliding-window construction. All windows from the same participant were assigned to the same subset throughout, preventing repeated-participant leakage across model development and final testing and reducing optimistic bias in performance estimation [22,23]. We first performed a participant-level stratified split, yielding an independent test set of 20 participants (11 non-ROA and 9 ROA) and a development set of 72 participants. All subsequent model-development procedures were confined to the development set.

XSL•FO
RenderX

Figure 1. Leakage-controlled data preprocessing pipeline. All windows from the same participant were kept within a single subset throughout the workflow. Representation construction, preprocessing, and model development were performed without using information from the independent test set.



For each participant, a fixed-length steady-walking segment of 500 samples was retained from the original plantar-pressure recording. From this same segment, 2 signal representations were constructed for subsequent analyses—the raw plantar-pressure signal in its original scale, and a participant-wise channel-wise standardized representation. Let μc and σc denote the mean and SD of channel c computed from that participant's steady-walking signal. We first standardize each channel within a participant:

(1)x~(t,c)=x(t,c)−μcσc+

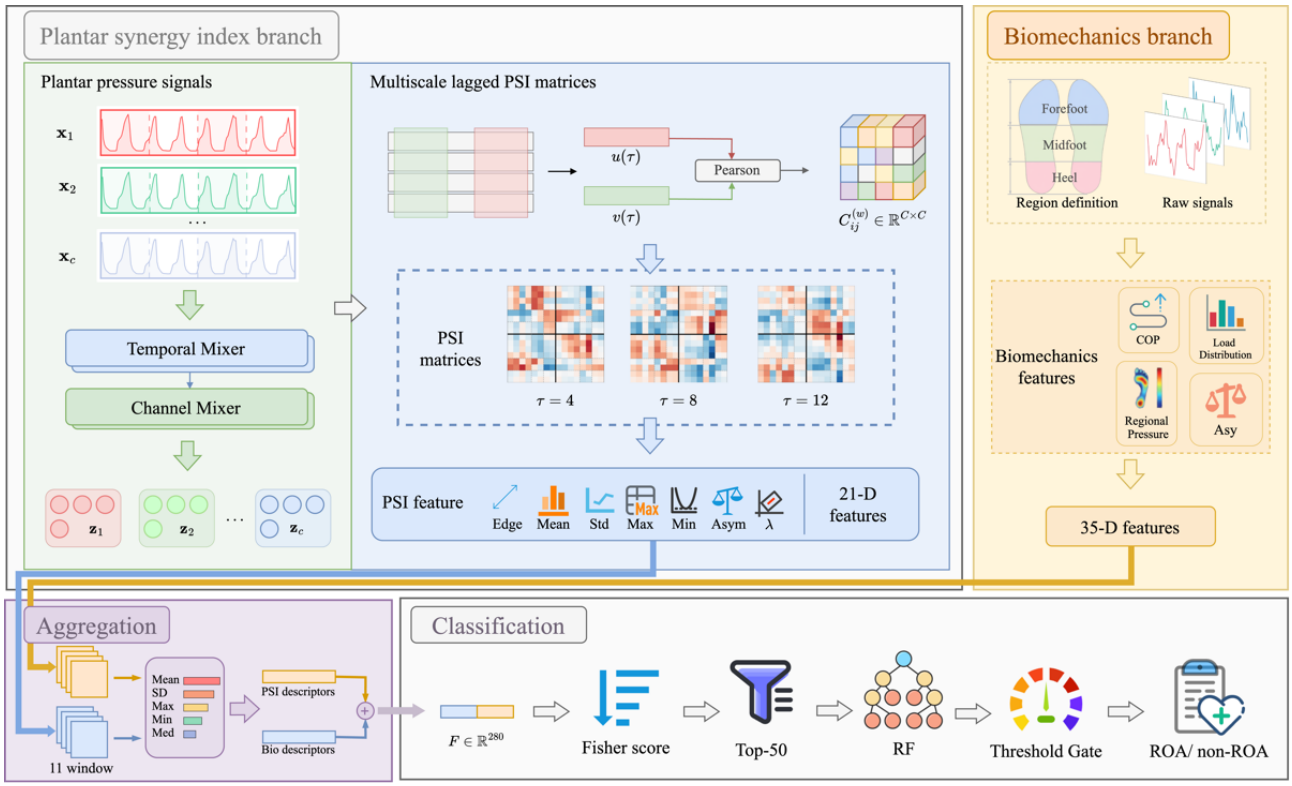
where =10-6 ensured numerical stability. The same representation-construction procedure was applied to both the development and independent test sets. After representation construction, the retained steady-walking segment was segmented into overlapping windows to capture short-term dynamics. We set the window length to L=80 samples (4 s at 20 Hz) and the stride to s=40 samples (2 s), yielding 50% overlap. This duration was chosen to span approximately 3 - 4 complete gait cycles at typical walking cadences (1.0 - 1.2 Hz), balancing within-window stability for PSI estimation against temporal resolution. This procedure produced Nw=11 windows per participant, so that all participants contributed the same number of window-level samples. Each window was denoted as X(w)∈ RL×C for w=1,…,Nw.

Window-Level Feature Extraction

We extracted 2 complementary window-level feature sets from the windowed plantar-pressure signals (ie, conventional biomechanical features and PSI-based coordination features; Figure 2). In the PSI branch, we aimed to capture distributed coordination patterns across plantar regions that may not be adequately characterized by conventional biomechanical descriptors alone. The PSI-related analyses were based on the participant-wise standardized representation. Estimating PSI directly from the window X~(w) can be sensitive to sensor noise, gait-phase misalignment, and interparticipant amplitude variation, particularly in small cohorts. To improve robustness while preserving a lightweight formulation, we introduced a mixer-style representation transform inspired by the token-channel mixing strategy of multilayer perceptron (MLP)-mixer and its time-series adaptations [24,25].

XSL•FO
RenderX

Figure 2. Two-branch plantar-pressure representation framework for subject-level ROA (radiographic knee osteoarthritis) identification. The biomechanics branch operates on raw pressure windows, whereas the PSI (Plantar Synergy Index) branch operates on the subject-wise standardized representation. Window-level features from both branches are aggregated to subject level for feature selection and final classification.



Within each development split, this transform was pretrained using only non-ROA windows from the corresponding training subset in a self-supervised one-step-ahead prediction task, consistent with predictive representation learning for sequential data [26,27]. The resulting module alternates temporal and channel mixing to produce a stable intermediate representation before PSI construction. Specifically, the transformed signal was defined as:

(2)Z(w)=M2 (M1 (X~(w))), Z(w)∈ RL×C

where M1 and M2 implement temporal mixing and channel mixing, respectively, using lightweight MLPs shared across channels and time steps to control complexity. PSI matrices were then constructed for each window over a multiscale lag set T = {4, 8, 12}, chosen to span short-to-long coordination offsets under the current sampling rate. The sensitivity of this design choice was further examined in the design-variation analyses. Each PSI matrix was defined from lag-aligned interregional correlation, motivated by the lead-lag structure commonly observed in plantar-pressure signals [28,29]. For window w, let $z_i^{(w)(t)}$ denote channel i at time t in $Z^{(w)}$. For any τ ∈ T, we formed a lead-lag aligned signal pair:

(3)ui(w)(τ)=[zi(w)(1+τ),...,zi(w)(L)],vj(w)(τ)=[zj(w)(1),...,zj(w)(L−τ)]

and defined the PSI matrix C(w)(τ)∈ RC×C as

(4)Cij(w)(τ)=corr(ui(w)(τ),vj(w)(τ)),τ∈ T

where corr(·,·) is Pearson correlation over length L-τ.

From each PSI matrix C(w)(τ), we extracted a compact set of PSI-derived summary features to characterize global coupling structure at each lag scale. These features capture complementary aspects of PSI structure, including edge density, distributional summaries (mean, SD, maximum, minimum), lagged asymmetry, and dominant spectral organization. The full definitions, mathematical formulations, and naming rules for PSI are provided in Multimedia Appendix 3. Thus, PSI features from all 3 lag scales were included in the candidate representation space before downstream feature selection.

Meanwhile, the biomechanics branch computed conventional plantar-pressure features directly from the raw pressure window X(w), providing interpretable anchors such as overall statistics, center-of-pressure (CoP), load ratios, and asymmetry indices. The complete feature list (35 features), channel mappings, region definitions, and formulas are provided in the Multimedia Appendix 2. This branch outputs window-level biomechanics feature vector fBio(w)∈ R35. For each window, the PSI and biomechanics features were concatenated to form the window-level feature vector f(w)=[fPSI(w);fBio(w)].

Participant-Level Modeling

To obtain 1 risk score per participant, we aggregated window-level features into a participant-level representation using fixed summary statistics. For a scalar window feature sequence {uk(w)}w=1Nw, we computed 5 statistics (mean, SD, maximum, minimum, and median) and stacked them as

(5)Ak,m=Statm({uk(w)}w=1Nw),m=1,…,5

Applying this aggregation to PSI features fPSI(w)∈ R21 and biomechanics features fBio(w)∈ R35 yielded F∈ R105 and G∈ R175, respectively, and the final participant-level feature vector was x=[F;G]∈ R280.

XSL•FO
RenderX

Given the limited cohort size relative to the 280-dimensional representation, the resulting feature space was prone to redundancy and overfitting. We therefore used Fisher score ranking as a simple filter-based selection step to retain the most discriminative participant-level features. For feature i, the score was defined as

(6)Fscore(i)=(μi,1−μi)2+(μi,0−μi)2(σi,1)2+(σi,0)2+

Here, μi,c and σi,c denote the class-specific mean and standard deviation for class c∈ {0,1}, μi denotes the pooled mean, and =10-6 is a small constant for numerical stability. The Fisher score quantifies between-class separability relative to within-class variance. The top-50 setting was fixed before locked independent test-set evaluation according to the development protocol. The top-k sensitivity analysis reported below was conducted to examine whether this feature count lay within a stable performance range, rather than to select the final feature count using the locked test set.

Leakage-Controlled Model Development and Evaluation

The downstream classifier was a random forest (RF), chosen for its robustness in small-sample settings and its ability to capture nonlinear feature interactions [30]. To avoid information leakage in threshold selection, the operating threshold was determined within the development set by participant-level 5-fold cross-validation. In each fold, all model-development steps, including Fisher score ranking, top-50 retention, and RF fitting, were performed using only the fold-specific training partition. Each validation participant contributed 1 aggregated participant-level feature vector and 1 corresponding RF probability as an out-of-fold estimate, ensuring that the probability used for threshold selection was generated without using that participant in representation learning, feature selection, or classifier fitting. The final threshold was selected from these pooled out-of-fold predictions to maximize the score.

After this internal development step, Fisher score ranking was reapplied once to the full development set to construct the final top-50 panel, and the RF classifier was refit on all 72 development participants using the locked feature panel and locked operating threshold before 1-time evaluation on the independent test set. Thus, cross-validation was used for internal model development rather than for reporting averaged final performance, whereas the independent test set was reserved for a single locked evaluation. Further implementation details are provided in Multimedia Appendix 4.

We additionally benchmarked the proposed fusion model against 3 representative end-to-end temporal baselines to examine how it compares with alternative modeling paradigms under the same fixed-length input, development or test split, and participant-level evaluation protocol. InceptionTime was included as a strong convolutional baseline for time-series classification [31], Medformer as a Transformer baseline whose multigranularity patching design is well-suited to medical time-series classification [32], and TodyNet as a graph-based baseline that directly models multivariate temporal dynamics through dynamic graph construction and temporal graph learning [33]. For fairness, each baseline was trained with its own tuned model-specific hyperparameter configuration, while the shared data handling and evaluation setup are summarized in Multimedia Appendix 4.

Statistical Analysis

All evaluations were performed at the participant level, with 1 ROA risk score and 1 binary decision assigned to each participant. We used the F_1-score as the primary metric:

(7)F1=2·Precision·RecallPrecision+Recall

In addition to F1, we reported AUROC, sensitivity, specificity, and precision to provide a more comprehensive characterization of discrimination performance. Given the limited test-set size (n=20), point estimates alone could overstate differences between models. We therefore computed 2000-iteration bootstrap 95% CIs for all metrics, and supplemented the bootstrap analysis with paired DeLong tests to evaluate whether observed AUROC differences were statistically distinguishable from chance. To assess whether the primary model performance could be explained primarily by available demographic and anthropometric variables, we conducted a supplementary model-level covariate sensitivity analysis. First, we trained a demographics-only RF model using age, height, weight, and BMI under the same participant-wise evaluation protocol. Second, we added these variables to the Fusion feature set before Fisher-score selection and RF classification. The complete covariate sensitivity analysis protocol is detailed in Multimedia Appendix 6.

To better understand which participant-level descriptors contributed consistently under resampling, we examined feature-selection stability across the 5 development folds. Within each fold, Fisher score ranking was applied to the training partition using the same feature-selection procedure as in the main pipeline. For each candidate feature, we summarized the number of folds in which it was retained and its mean within-fold Fisher score. For interpretation, retained features were further grouped by descriptor family, lag scale, and anatomical sensor region according to the insole topology.

To characterize robustness, we evaluated the fusion, bio-only, and PSI-only models under the following 3 perturbation categories: temporal scaling, random channel dropout, and additive Gaussian noise. Temporal scaling was tested at 0.8×, 1.0×, and 1.2× relative to the native sampling rate. Random channel dropout was evaluated at masking rates (0.0, 0.1, 0.2, and 0.3), and additive Gaussian noise was evaluated at σ=0.00, 0.01, 0.03, and 0.05 applied to the standardized signal. Design sensitivity was further examined by varying the sparsity threshold in PSI matrix construction, the lag configuration (single-scale vs multiscale), and the window length using 3 s, 4 s, and 5 s windows with 50% overlap. Phase-resolved PSI patterns were visualized by aligning trajectories to the normalized stance phase using 5% bins and pooling them at the group level within each functional module and lag setting. These perturbation and design-sensitivity analyses were used to assess the stability of the selected configuration and were not used to reselect the feature count, window length, lag configuration, sparsity threshold, classifier, or operating threshold.

XSL•FO
RenderX

Ethical Considerations

The study was approved by the Institutional Review Board (IRB) of Zhujiang Hospital, Southern Medical University, Guangzhou (IRB 2019-KY-016 - 02). All participants were enrolled from the PROC study and provided written informed consent before data collection; the original consent permitted secondary analysis of deidentified data without additional consent. Participants received financial compensation of approximately US $30 per participant. No identifiable images of individual participants appear in this manuscript or its supplementary materials.

Results

Primary Classification Performance

We first compared PSI-only, bio-only, and fusion models under the leakage-controlled development or test protocol. As summarized in Table 1, PSI-only showed limited standalone classification performance, with a low F_1-score, AUROC, accuracy, and precision despite relatively high recall. Bio-only provided a strong baseline with substantially improved discrimination and a balanced sensitivity-specificity profile. Fusion showed numerically higher F_1-scores (0.842 vs 0.778), accuracy (0.850 vs 0.800), precision (0.800 vs 0.778), and recall (0.889 vs 0.778) than bio-only, while AUROC and specificity were identical between the 2 models.

Table . Participant-level classification performance on the independent test set for PSI[a]-only, bio-only, and fusion models.

Metric	PSI-only	Bio-only	Fusion
F_1-score	0.609	0.778	0.842
AUROC[b]	0.525	0.818	0.818
Accuracy	0.550	0.800	0.850
Precision	0.500	0.778	0.800
Recall	0.778	0.778	0.889
Specificity	0.364	0.818	0.818

[a]PSI: Plantar Synergy Index.

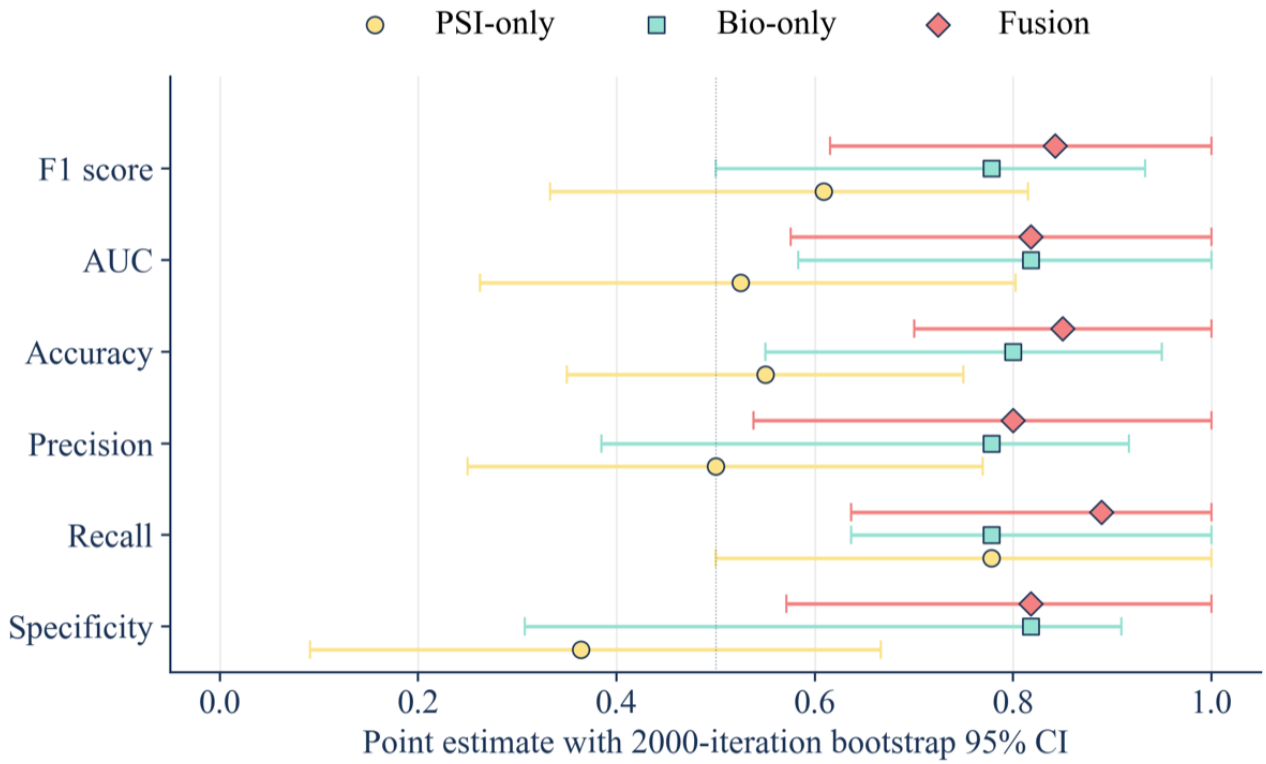
[b]AUROC: area under the receiver operating characteristic curve.

Bootstrap 95% CIs for all metrics are shown in Figure 3. The intervals were wide and overlapping across branches; for instance, the F_1-score spanned 0.500 - 0.948 for bio-only and 0.615 - 1.000 for fusion. Paired DeLong tests showed that both fusion and bio-only outperformed PSI-only at the AUROC level (P=.02 and P=.04, respectively), whereas fusion did not differ from bio-only in AUROC (P>.99). Together, these results suggest that PSI alone is insufficient as a standalone classifier, while fusion showed a directionally favorable threshold-dependent profile relative to bio-only. Given the wide and overlapping CIs, this pattern should be interpreted cautiously as preliminary evidence that PSI may complement biomechanical descriptors rather than as definitive superiority over bio-only. The complete bootstrap confidence intervals and paired DeLong comparisons are provided in Multimedia Appendix 7.

The covariate sensitivity analysis is summarized in Multimedia Appendix 6. The demographics-only RF model performed substantially worse than the fusion model, indicating that age, height, weight, and BMI alone could not reproduce the primary plantar-pressure–based classification performance. Adding these demographic and anthropometric variables to the fusion feature set did not materially improve the F_1-score, AUROC, or specificity. Within the augmented feature space, age and BMI were retained by Fisher score ranking, whereas height and weight were not retained. These findings suggest that the fusion model was not simply a demographic proxy, although residual confounding cannot be excluded.

XSL•FO
RenderX

Figure 3. Bootstrap 95% CIs for six metrics across PSI (Plantar Synergy Index)-only, Bio-only, and Fusion (2000 iterations, n=20). Point estimates and CIs are shown for each branch–metric combination; wide and overlapping intervals reflect the limited test-set size. These intervals support cautious interpretation of between-branch differences, particularly for Fusion versus Bio-only. AUC: area under the curve.



Cross-Fold Feature-Selection Stability and Representative Stable Descriptors

To examine whether the locked top-50 feature-retention setting lay within a stable performance range, we conducted a top-k scan across 15 values; results are illustrated in Figure 4. Fusion performance at k=50 (F_1-score=0.842) lay within a high-performance plateau, with similarly strong F_1-score and accuracy at neighboring k values. At smaller k values, performance was limited by an insufficient number of retained features, whereas larger panels around k>=90 - 100 showed degraded classification performance, with F_1-score dropping to 0.667 at k=100. These results suggest that k=50 was located within a relatively stable operating range rather than representing an isolated favorable point. However, this analysis was used to assess the robustness of the selected configuration and was not used for final feature-count tuning.

To evaluate whether the descriptors selected during the 5-fold development procedure were reproducible under sample variation, we analyzed feature-selection stability across the 5 development folds. In each fold, Fisher score ranking was applied to the fold-specific training partition using the same selection procedure as in the main pipeline, and the top 50 descriptors were retained. Each descriptor was then summarized by its cross-fold selection rate and mean within-fold Fisher score, and descriptors retained in at least 3 of 5 folds were regarded as stable. This cross-fold view was used for interpretation because, in the present small-sample setting, the single top-50 panel obtained from the full development set may be sensitive to sample composition, whereas repeatedly retained descriptors are more likely to reflect reproducible discriminative structure [34-36].

XSL•FO
RenderX

Figure 4. Top-k sensitivity scan for Fisher-score feature retention. Left: F_1-score across retained feature counts (k ∈ {10, ., 280}). Right: Accuracy across retained feature counts. PSI (Plantar Synergy Index)-only, Bio-only, and Fusion are shown in each panel. The analysis evaluates whether the prespecified Top-50 configuration lies within a stable performance range, rather than serving as final model tuning.

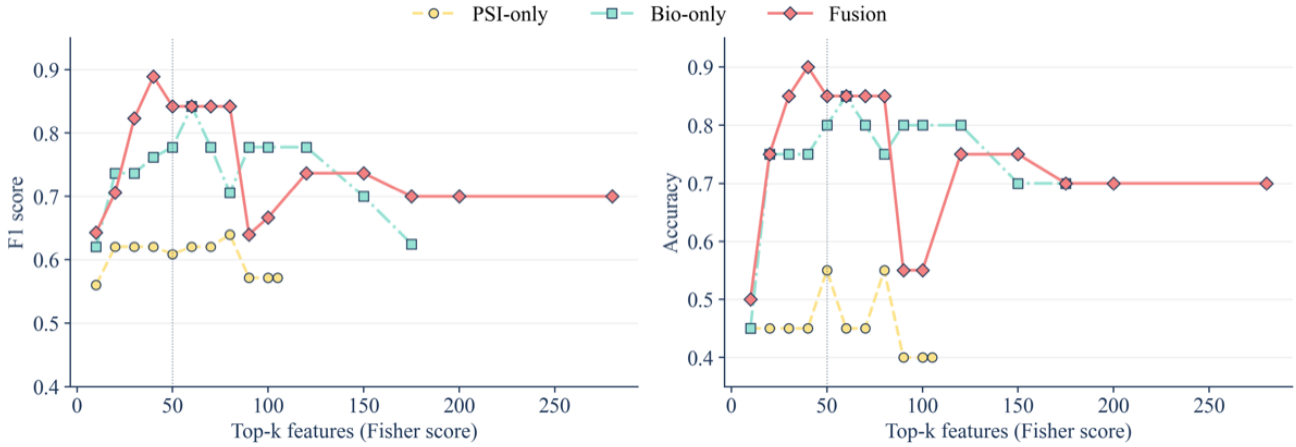
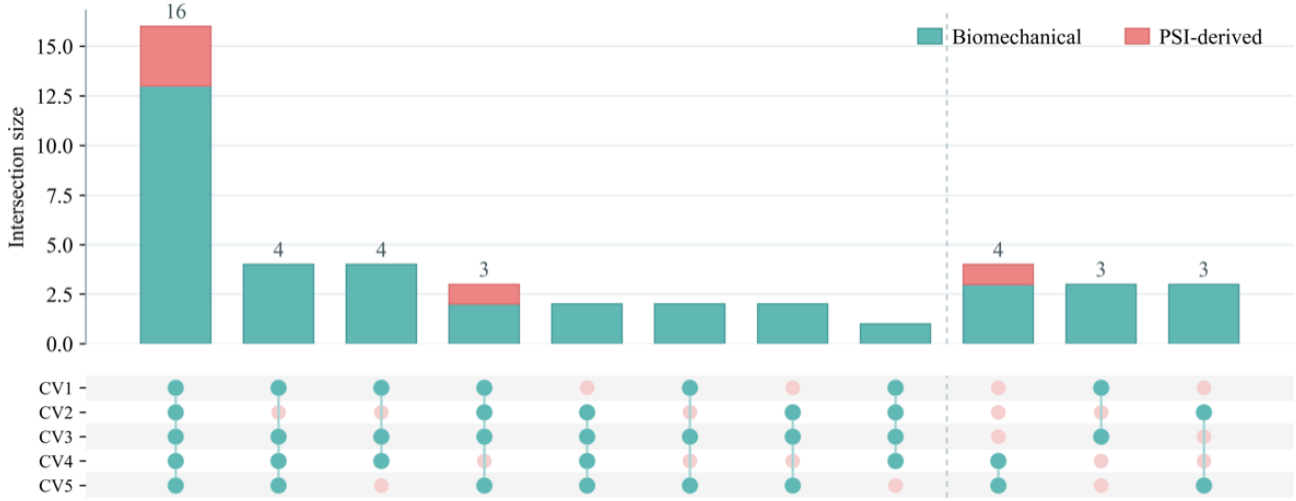


Figure 5 summarizes the cross-fold overlap structure of selected descriptors across the 5 development folds. A substantial shared overlap, together with several additional higher-overlap intersections, indicates that the retained descriptor set was not purely fold-specific but contained a reproducible core. Importantly, the higher-overlap intersections were dominated by biomechanical descriptors, supporting conventional biomechanics as the primary anchor of the stable feature set. PSI-derived descriptors appeared less frequently and in smaller numbers, but they were still recurrent across multiple higher-overlap intersections, indicating a complementary yet reproducible coordination-level contribution. Lower-overlap intersections were also present, suggesting some sensitivity of the selected panel to sample composition, although these intersections were generally smaller and more heterogeneous. Importantly, these stable descriptors were used for interpretation rather than as a fixed final feature panel. For final model construction, Fisher score ranking was reapplied once to the full development set, and the locked top-50 panel was selected from all candidate descriptors using all 72 development participants.

To further illustrate the main discriminative signals within this reproducible subset, Table 2 highlights representative descriptors that were both relatively stable across folds and highly ranked by Fisher score.

The representative examples make the reproducible structure more concrete and highlight two complementary, clinically meaningful patterns. First, the biomechanical examples are concentrated in interpretable domains, such as asymmetry, CoP, and gait timing, reinforcing conventional biomechanics as the primary anchor of the stable signal. This pattern is consistent with previous evidence that knee osteoarthritis is associated with abnormal plantar-pressure distribution, altered load redistribution, and CoP-related gait changes [6,16-18]. Second, PSI-derived descriptors were also retained among the stable high-ranking features, indicating that coordination-level information was recurrently represented alongside conventional biomechanical descriptors. Taken together, these representative descriptors indicate that ROA-related discrimination in the present cohort was anchored in clinically interpretable abnormalities of plantar loading, load redistribution, and whole-foot progression.

Figure 5. Cross-fold overlap structure of selected descriptors across the 5 development folds. Each column shows one descriptor-overlap pattern across the 5 development folds. Bar height indicates the overlap size, and stacked colors indicate the numbers of biomechanical and PSI (Plantar Synergy Index)-derived descriptors. For clarity, only a subset of lower-overlap patterns (<3 folds) is displayed. Stable retained descriptors were mainly biomechanical, with a smaller recurrent PSI-derived component.



XSL•FO
RenderX

Table . Representative high-ranking and relatively stable subject-level descriptors selected from the reproducible subset across development folds. Selection rate denotes the proportion of development folds in which a descriptor was retained, and mean fold-wise Fisher score summarizes its average discriminative ranking within folds.

Descriptor	Selection rate	Fisher score, mean (SD)
asymmetry_std_sensor_5_std	1.0	7.92 (1.59)
asymmetry_mean_sensor_2_std	1.0	6.01 (3.82)
PSI[a]-Eigen-S-Min	1.0	5.62 (2.28)
PSI-Mean-S-Min	1.0	4.20 (1.70)
cop_variability_min	0.8	3.65 (0.42)
gait_timing_min	1.0	3.63 (1.46)
PSI-Edge-L-Std	0.6	3.17 (1.54)
PSI-Edge-L-Min	0.6	3.11 (0.60)

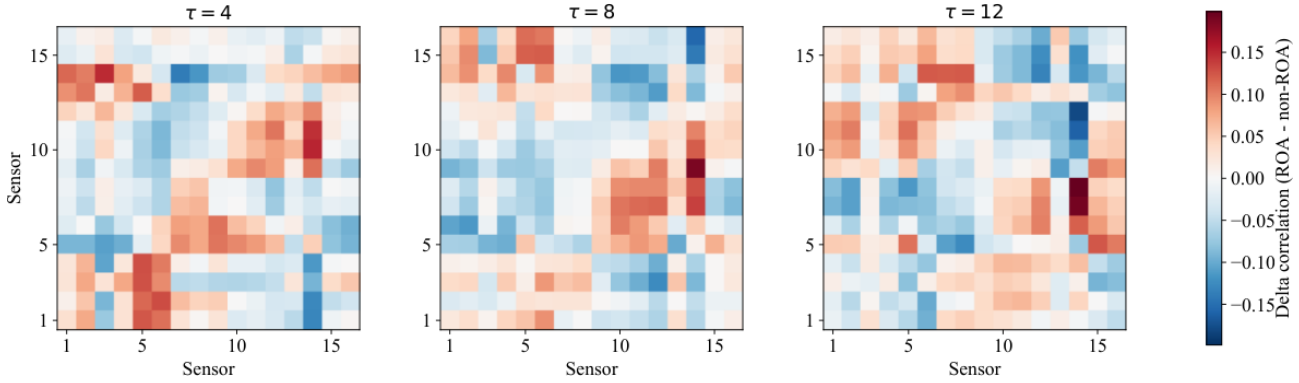
[a]PSI: Plantar Synergy Index.

Phase-Resolved PSI Patterns Across Lag Scales and Functional Modules

To further interpret the coordination-level information captured by PSI, we examined whether ROA-related PSI differences showed coherent organization across lag scales, stance phase, and functional plantar modules. Because PSI was designed to characterize interregional temporal coupling rather than local pressure magnitude, the analysis was structured as a progressive visualization from sensor-pair spatial organization to phase-resolved stance dynamics, and finally to module-level difference trajectories. This design allowed us to assess whether PSI captured a distributed coordination pattern associated with ROA rather than isolated or nonspecific correlation differences.

We first visualized sensor-pair delta-correlation matrices, defined as ROA minus non-ROA, at representative lag settings. As illustrated in Figure 6, the delta-correlation matrices revealed spatially distributed group differences across multiple sensor pairs, rather than a focal abnormality confined to a single plantar region. The pattern also varied across lag settings, indicating that ROA-related PSI alterations depended on the temporal offset used to characterize interregional coupling. At shorter lags, the observed differences may reflect altered near-synchronous load sharing among plantar regions, whereas longer lags may capture delayed coordination related to forward load transfer and bilateral stance progression. These findings suggest that PSI differences were not merely a consequence of local pressure magnitude changes, but reflected lag-dependent reorganization of plantar coordination.

Figure 6. Sensor-pair delta-correlation heatmaps across representative lag settings. Heatmaps show the difference in lag-aligned sensor-pair correlation between ROA (radiographic knee osteoarthritis) and non-ROA groups (ROA minus non-ROA) at τ=4,8and12. Sensors 1 - 8 and 9 - 16 correspond to homologous plantar regions on the left and right feet, respectively. PSI (Plantar Synergy Index) differences were distributed across sensor pairs and varied by lag, rather than being confined to a single sensor region.



The spatially distributed and lag-dependent delta-correlation patterns in Figure 6 suggested that ROA-related PSI differences were linked to the temporal organization of plantar coordination during walking. However, the heatmap representation does not indicate when during stance these coordination differences emerged. We therefore further aligned PSI trajectories to the normalized stance phase (0% - 100%) to examine the temporal expression of the observed lag-dependent coordination changes. As illustrated in Figure 7, group-level PSI trajectories differed between ROA and non-ROA across all examined lag settings. Across broad portions of stance, the ROA curves remained lower than the non-ROA curves, suggesting reduced or less coherent interregional plantar coordination in the ROA group. From a gait-phase perspective, this reduction was not limited to initial contact or early loading, but extended across mid-stance and late stance, where controlled CoP progression, forward load transfer, and preparation for push-off become increasingly important. The magnitude and temporal expression of this separation varied with lag scale; shorter lags emphasized more immediate interregional coupling, whereas longer lags showed

XSL•FO
RenderX

greater fluctuation near late stance and push-off. These lag-dependent patterns suggest that the proposed PSI may capture ROA-related compensatory changes in plantar load redistribution, particularly during forward load transfer and propulsion, when coordinated forefoot and heel-to-midfoot interactions become biomechanically important.

Figure 7. Phase-resolved PSI (Plantar Synergy Index) patterns for ROA (radiographic knee osteoarthritis) and non-ROA groups across lag scales. Stance phase was normalized to 0% - 100% and summarized in 5% bins. Curves represent group-level PSI summaries pooled across modules within each phase bin; shaded bands indicate 95% CIs. PSI differences emerged primarily in mid-to-late stance, consistent with altered load redistribution and propulsion-phase coordination in ROA.

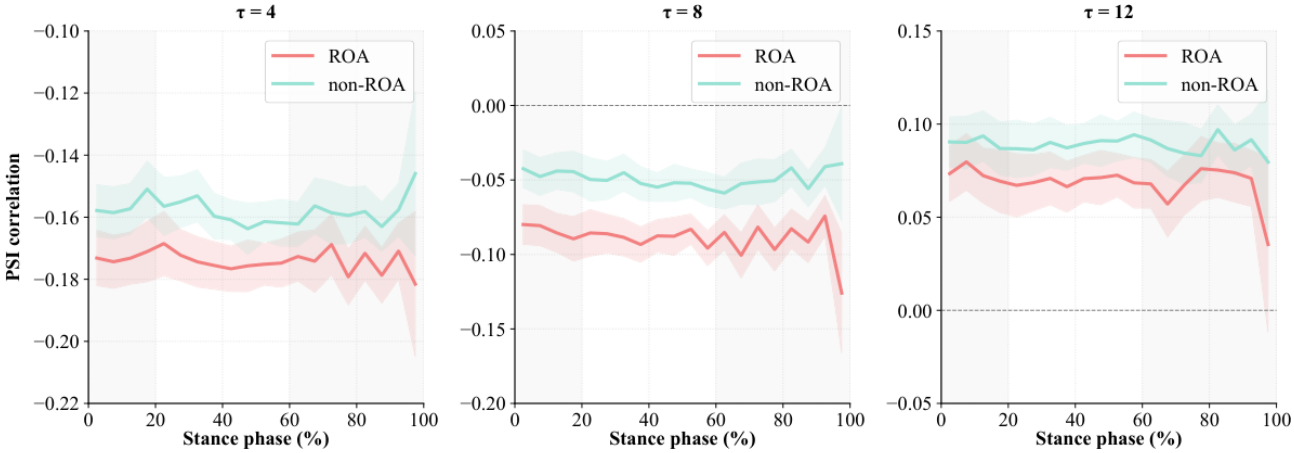
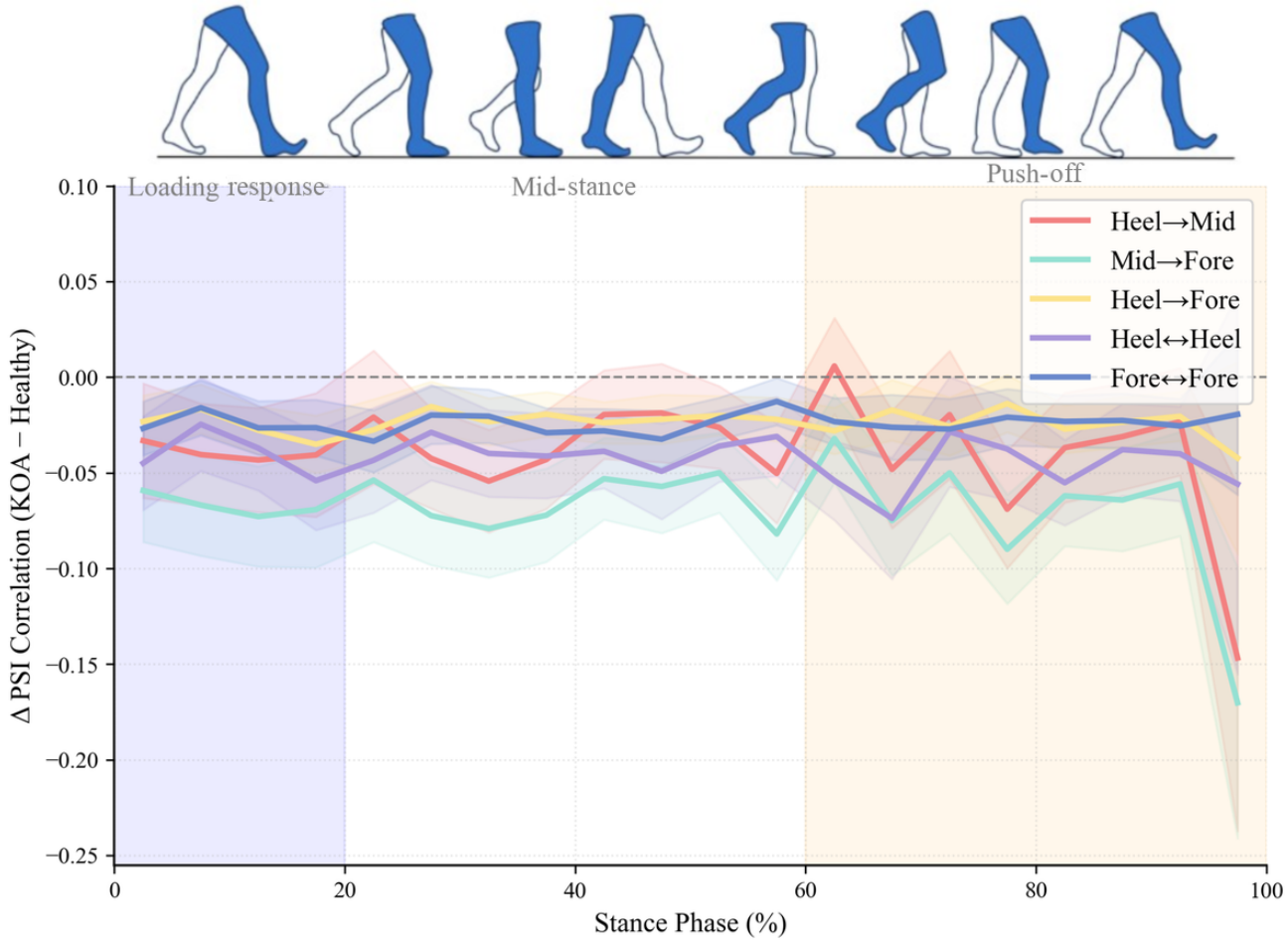


Figure 8 further shows phase-resolved PSI difference trajectories at a representative lag (τ=0.4s), where the vertical axis represents ΔPSI=PSIROA-PSInon-ROA. Across most of stance, the major module-level trajectories remained below zero, indicating that PSI was generally lower in ROA than in non-ROA for these interregion couplings. The persistence of negative differences across broad stance intervals suggests that ROA-related coordination changes were not restricted to a single gait event, but were expressed throughout stance-phase load transfer. This group difference persisted across broad portions of stance and was especially evident in forefoot-related and heel-to-mid couplings, while larger fluctuations near late stance and push-off suggested greater between-participant heterogeneity when forward load transfer and propulsion became dominant [6,16,17].

XSL•FO
RenderX

Figure 8. Phase-resolved PSI (Plantar Synergy Index) difference trajectories at the representative lag τ=0.4 s. Curves show ΔPSI=PSIROA-PSInon-ROA for different inter-region functional modules across the normalized stance phase. Shaded background regions indicate major stance sub-phases. Cross-module PSI differences persisted across multiple stance-phase bins, indicating distributed inter-regional coupling changes rather than isolated within-module effects. ROA: radiographic knee osteoarthritis.



Taken together, Figures 6-8 indicate a coherent progression of PSI evidence: ROA-related differences were spatially distributed across sensor pairs, temporally expressed across the normalized stance phase, and functionally organized across interregion plantar modules. Importantly, these phase-resolved analyses were intended to characterize the organization of PSI differences, rather than to imply that all stance bins, lag scales, or modules contributed equally to the final classifier. The final retained descriptors were still determined by the participant-level feature-selection procedure.

Robustness to Perturbation and Design Variation

We further examined whether PSI-related information remained useful under perturbation and design variation. Across temporal scaling, channel dropout, and Gaussian noise, fusion generally maintained the highest or among the highest participant-level F_1-scores among the 3 models (Table 3). Performance was strongest at the native and mildly stretched temporal scales (1.0× and 1.2×, both F_1-score=0.842), with only a modest reduction under compression (F_1-score=0.824 at 0.8×). Under channel dropout and Gaussian noise, fusion also maintained a generally favorable profile relative to both PSI-only and bio-only models. These perturbation analyses suggest preliminary robustness of the combined representation under moderate signal variation, although the small locked test set limits strong comparative claims.

XSL•FO
RenderX

Table . Robustness of PSI-only, bio-only, and fusion models under temporal and signal perturbations. Reported values are subject-level F_1-scores on the independent test set.

Setting and level	PSI[a]-only	Bio-only	Fusion
Temporal scaling			
0.8×	0.593	0.778	0.824
1.0×	0.609	0.778	0.842
1.2×	0.571	0.824	0.842
Channel dropout			
0.0	0.609	0.778	0.842
0.1	0.588	0.762	0.824
0.2	0.615	0.727	0.818
0.3	0.571	0.692	0.762
Gaussian noise			
0.00	0.609	0.778	0.842
0.01	0.600	0.778	0.824
0.03	0.571	0.727	0.800
0.05	0.556	0.667	0.818

[a]PSI: Plantar Synergy Index.

As shown in Table 4, across design variations, fusion showed F_1-scores values within a relatively narrow range over sparsity thresholds (F_1-score=0.824 - 0.842) and lag settings (F_1-score=0.818 - 0.842), while showing greater sensitivity to window length (F_1-score=0.761 - 0.842). The 4-second window (80 samples at 20 Hz, spanning approximately 3 - 4 complete gait cycles at typical walking cadences of 1.0 - 1.2 Hz) showed the numerically highest F_1-score among the examined window lengths. Shorter 3-second windows, covering only 1 - 2 gait cycles, may have yielded less reliable correlation estimates for PSI matrix construction because the reduced sample size per window limited the statistical stability of the lag-aligned Pearson correlations. Conversely, 5-second windows, although providing more samples per correlation estimate, may have smoothed phase-specific coordination changes across stance subphases, reducing the ability to resolve temporally localized PSI differences between ROA and non-ROA groups. The 4-second window therefore appeared to provide a suitable balance between within-window statistical stability for correlation-based PSI estimation and temporal resolution across stance subphases.

Within the lag analysis, the multiscale configuration (T={4, 8, 12}, corresponding to 0.20, 0.40, and 0.60 s at 20 Hz) showed the numerically highest performance among the examined lag settings (F_1-score=0.842), compared with each single-scale alternative. A single lag may not simultaneously capture rapid intrafoot pressure propagation between adjacent sensors and slower cross-region coordination reorganization spanning multiple functional modules. The multiscale design therefore provided a broader candidate coordination space—spanning short-to-intermediate coupling timescales relevant to mid-to-late stance gait dynamics—from which downstream Fisher score selection could retain informative lag-specific descriptors across all 3 scales. This exploratory pattern is consistent with the idea that the multiscale design may capture complementary coupling timescales, but it requires confirmation in larger independent cohorts. Overall, these exploratory perturbation and design-sensitivity analyses suggest that the selected configuration was not uniquely dependent on a single perturbation or design setting. These findings should be interpreted as preliminary robustness checks rather than as independent optimization or validation of the model design.

XSL•FO
RenderX

Table . Sensitivity of the fusion model to design variations. Reported values are participant-level F_1-scores on the locked independent test set under alternative window lengths, lag configurations, and sparsity thresholds. These analyses assess robustness of the selected design rather than serving as final model tuning.

Setting and level	Fusion
Window length	
3 s (50%)	0.761
4 s (50%)	0.842
5 s (50%)	0.800
Lag configuration	
Single-τ=4	0.824
Single-τ=8	0.818
Single-τ=12	0.824
Multiscale	0.842
Sparsity threshold	
θ=0.01	0.824
θ=0.02	0.842
θ=0.05	0.842

Benchmarking Against Representative Convolutional Neural Network, Transformer, and Graph-Based Baselines

Table 5 shows that fusion achieved the strongest overall participant-level performance among the evaluated models. Among the baselines, InceptionTime [31] was the strongest competitor and showed a relatively balanced classification profile, but remained below fusion in accuracy, F_1-score, AUROC, precision, and recall, while matching the best specificity. Medformer [32] achieved high recall comparable with fusion, but with lower AUROC, lower precision, and markedly reduced specificity, indicating a more false-positive−prone decision profile. TodyNet [33] showed perfect recall but near-zero specificity, yielding the weakest overall discrimination and suggesting a strong bias toward positive predictions under the present data scale.

Overall, these results suggest that, under the current data scale and evaluation setting, fusion showed a directionally favorable threshold-dependent profile relative to the evaluated end-to-end baselines. Relative to InceptionTime, the strongest baseline, fusion showed numerically higher accuracy, F_1-score, AUROC, precision, and recall while matching specificity, but these differences should be interpreted cautiously given the small locked test set. Medformer preserved high recall but at the cost of lower specificity and lower overall discrimination, whereas TodyNet showed a marked false-positive tendency with poor ranking quality. Together, these comparisons suggest that the proposed representation-based framework remained competitive with representative end-to-end baselines in this proof-of-concept evaluation.

Table . Participant-level comparison between the proposed fusion model and representative end-to-end baselines on the independent test set. All models used the same participant-wise development/test split, fixed-length input, and participant-level evaluation protocol.

Model	Accuracy	F_1-score	AUROC[a]	Precision	Recall	Specificity
InceptionTime	0.800	0.778	0.802	0.778	0.778	0.818
Medformer	0.700	0.727	0.707	0.615	0.889	0.545
TodyNet	0.450	0.621	0.485	0.450	1.000	0.000
Fusion	0.850	0.842	0.818	0.800	0.889	0.818

[a]AUROC: area under the receiver operating characteristic curve.

Discussion

Principal Results

This study addressed a key representation gap in plantar-pressure–based ROA identification by introducing a coordination-aware framework that integrates local biomechanical loading with multiscale interregional temporal information. The combined framework achieved competitive participant-level ROA identification, but it did not show consistent superiority over bio-only across performance metrics. The biomechanics branch made the strongest individual contribution, whereas PSI showed limited standalone discrimination. However, PSI captured complementary coordination patterns that were not represented by conventional biomechanical descriptors, and a reproducible PSI component

XSL•FO
RenderX

was retained within the combined representation. Robustness and design-sensitivity analyses provided preliminary evidence that the findings were not confined to a single evaluated configuration. Accordingly, the principal innovation of this study is representational and methodological rather than cohort- or device-based; it explicitly combines interpretable local loading descriptors with multiscale interregional temporal coordination under a participant-level, leakage-controlled evaluation protocol.

The branch-wise findings help explain the primary source of discrimination. The strong contribution of the biomechanics branch is consistent with previous evidence that ROA-related plantar-pressure abnormalities occur in anatomically meaningful regions, particularly the hallux and heel, and are also reflected in CoP stability and variability [17,18]. This interpretation is further supported by recent studies associating ROA-related functional impairment with altered plantar-pressure behavior, gastrocnemius muscle-tendon-unit stiffness, and 3D gait biomechanics [7,8]. These patterns are biomechanically plausible because structural degeneration, pain-related gait adaptation, and altered neuromuscular control at the knee may be associated with compensatory load redistribution and changes in stance progression [6-8,15-18]. Experimental modulation of tibialis anterior activity has also been shown to alter plantar pressure and gastrocnemius activity during gait, supporting the sensitivity of plantar loading to neuromuscular modulation [37]. The findings of this study therefore reinforce the importance of anatomically grounded loading descriptors for participant-level ROA identification.

The phase-resolved and fold-wise analyses help clarify what PSI may contribute beyond these local loading descriptors. The observed PSI differences were distributed across multiple functional modules and lag settings rather than being confined to isolated local extrema. In addition, a core group of PSI descriptors was repeatedly retained across the development folds, supporting the reproducibility of these selected patterns within the present dataset in accordance with established feature-stability principles [34,35]. Although PSI is not a direct measure of joint kinematics or electromyographic muscle activity, its distributed temporal patterns are compatible with studies reporting altered kinematic synergy, muscle synergy, and muscle-activation patterns in individuals with KOA during walking [38-40]. PSI may therefore provide an interpretable coordination-level description of bilateral load transfer and whole-foot progression. However, feature stability and phase-resolved interpretability do not by themselves establish incremental predictive value beyond the biomechanics branch. Given the limited standalone discrimination of PSI and the absence of consistent superiority of fusion over bio-only, the findings of this study should not be interpreted as definitive evidence of additional predictive or clinical value.

Comparisons with the end-to-end baselines place these findings in a broader modeling context. The evaluated models represented complementary learning paradigms but exhibited different operating profiles: InceptionTime was the strongest and most balanced competitor, Medformer favored sensitivity over specificity, and TodyNet showed lower specificity and a higher false-positive tendency in the present locked test set. The value of the proposed framework therefore lies less in definitive performance superiority over every alternative than in combining competitive participant-level identification with an explicit and interpretable organization of plantar-pressure information before classification. The robustness and design-sensitivity analyses further suggested that the main findings were not entirely dependent on a single analytical configuration, although these analyses remain internal and exploratory. Together with the structured comparison in Multimedia Appendix 5, these results support the positioning of the framework as a representational and methodological contribution rather than a cohort- or device-based advance.

To provide a screening-context interpretation of the locked-test operating characteristics, we translated sensitivity and specificity into positive predictive value and negative predictive value across assumed community ROA prevalence levels in Multimedia Appendix 8 [41]. This supplementary analysis illustrates how disease prevalence could affect the potential rule-in and rule-out value of the framework. It should not, however, be interpreted as evidence of clinical deployment readiness because the underlying operating characteristics were estimated from a small internal test set rather than from a representative prospective screening population.

Limitations

Several limitations should be acknowledged. The sample size was modest, and all participants were drawn from a single previously reported cohort, limiting the precision and external generalizability of the performance estimates. Although the locked test set was separated from model development, it remained part of the same source cohort and therefore represents an internal held-out proof-of-concept evaluation rather than external validation. Contemporary prediction-model guidance emphasizes that independent validation should be adequately powered to estimate performance with sufficient precision [14,42]. The findings of this study should consequently be interpreted as preliminary and require confirmation in larger independent cohorts.

Important clinical and demographic factors, including age, BMI, and affected-side heterogeneity, may influence plantar-pressure patterns [15]. The supplementary covariate sensitivity analysis indicated that the available demographic variables alone did not reproduce the performance of the fusion model and that their inclusion did not materially improve its performance (Multimedia Appendix 6). Nevertheless, this analysis cannot exclude residual confounding or establish the independent and interactive effects of age and BMI in the present modest sample. These questions will require larger, more carefully stratified cohorts with richer clinical covariate information. Furthermore, to our knowledge, no publicly available dataset combined confirmed ROA labels with suitable in-shoe plantar-pressure recordings for direct external validation. Available public resources lacked either ROA annotations or compatible pressure measurements, precluding cross-dataset benchmarking [43,44]. Multicenter efforts to establish open datasets with imaging-grade clinical labels and standardized plantar-pressure recordings would substantially strengthen future evaluation.

XSL•FO
RenderX

This study also focused on binary ROA identification rather than disease-severity stratification or longitudinal progression. PSI depends on sensor layout, signal quality, and preprocessing choices, which may affect the validity and comparability of plantar-pressure measurements [10]; its transferability across devices and clinical settings therefore remains uncertain. Recent Soft Science studies demonstrate advances in biocompatible piezoelectric, gel-based, and electrospun flexible pressure sensors for walking, human-motion, gait, and continuous health monitoring [45-47]. These developments strengthen the technological basis for wearable plantar sensing, but they also underscore the need to verify signal stability, durability, environmental resilience, and cross-platform calibration before clinical translation [45-47]. In addition, the proposed coordination-level interpretation was not directly validated using concurrent joint-kinematic, electromyographic, or other neuromuscular measurements. Future studies should validate the framework prospectively in larger multicenter cohorts, evaluate its robustness across sensor platforms and patient subgroups, examine its association with direct biomechanical and neuromuscular measurements, and extend it toward severity modeling and longitudinal prediction.

Conclusions

This study introduces an interpretable participant-level representation that integrates local biomechanical loading with multiscale interregional temporal coordination, extending plantar-pressure approaches centered primarily on summary descriptors or less interpretable end-to-end models. Complementary evaluations provided convergent evidence of competitive participant-level ROA identification and preliminary robustness across the evaluated settings, while showing that local biomechanical descriptors constituted the main discriminative basis and that PSI offered a reproducible coordination-level component whose incremental predictive value remains to be established. By broadening plantar-pressure analysis from local loading to distributed gait coordination, the framework provides a methodological basis for more interpretable functional characterization of ROA. If prospectively validated in external multicenter cohorts, it could warrant further evaluation as an adjunct to imaging for functional assessment and longitudinal monitoring. Prospective multicenter validation is required before clinical translation.

Acknowledgments

We thank all participants for their valuable contributions. Hongbo Yao and LS are co-corresponding authors.

The authors declare the use of generative AI (GenAI) tools during manuscript preparation and the July 2026 revision. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: literature search and systematization, proofreading and editing, reformatting, quality assessment, identification of limitations, recommendations, and publication support. The GenAI tools used were ChatGPT 5.5 Thinking (OpenAI; accessed July 2026) and DeepSeek V4 Pro (DeepSeek; accessed July 2026).

Specifically, the tools assisted with language editing; condensation and restructuring of manuscript text; refinement of the Abstract, Discussion, and Conclusions; checking reporting consistency and whether claims had adequate citation support; identifying limitations; recommending manuscript revisions; and preparing draft responses to reviewers. The tools were not used to formulate the research questions or hypotheses, design the study, collect or analyze data, generate code, create figures, or independently interpret the results or determine the scientific conclusions.

The authors critically reviewed and revised all AI-assisted outputs and independently verified factual statements, references, calculations, statistical interpretations, and journal requirements. Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes. This declaration was submitted collectively by all authors.

Funding

This work was supported by the National Natural Science Foundation of China (82572955).

Data Availability

The data analyzed in this study were originally collected and reported. Access to the original dataset is available from the corresponding author of that study upon reasonable request. The analysis code is openly available at GitHub repository [48].

Authors' Contributions

Conceptualization: GX
Methodology: GX, Huilin Yao
Formal analysis: GX, JQ
Investigation: GX, LL, KQ
Validation: GX, LL, SL
Software: GX, JZ
Visualization: GX, Hongbo Yao, Huilin Yao, JQ

XSL•FO
RenderX

Supervision: Hongbo Yao, LS
Writing – original draft: GX
Writing – review & editing: GX, HBY, LS

Conflicts of Interest

None declared.

Multimedia Appendix 1
Participant flow and baseline characteristics.
[DOCX File, 1040 KB - mhealth_v14i1e101026_app1.docx]

Multimedia Appendix 2
Sensor-to-anatomy mapping and biomechanical feature definitions.
[DOCX File, 25 KB - mhealth_v14i1e101026_app2.docx]

Multimedia Appendix 3
Explicit forms of Plantar Synergy Index.
[DOCX File, 20 KB - mhealth_v14i1e101026_app3.docx]

Multimedia Appendix 4
Implementation and training hyperparameters.
[DOCX File, 22 KB - mhealth_v14i1e101026_app4.docx]

Multimedia Appendix 5
Cross-literature comparison.
[DOCX File, 28 KB - mhealth_v14i1e101026_app5.docx]

Multimedia Appendix 6
Covariate sensitivity analysis.
[DOCX File, 110 KB - mhealth_v14i1e101026_app6.docx]

Multimedia Appendix 7
Bootstrap confidence intervals and DeLong comparisons.
[DOCX File, 19 KB - mhealth_v14i1e101026_app7.docx]

Multimedia Appendix 8
Predictive values across community prevalence levels.
[DOCX File, 238 KB - mhealth_v14i1e101026_app8.docx]

References

1. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. The Lancet 2019 Apr;393(10182):1745-1759. [doi: 10.1016/S0140-6736(19)30417-9]
2. Steinmetz JD, Culbreth GT, Haile LM, et al. Global, regional, and national burden of osteoarthritis, 1990–2020 and projections to 2050: a systematic analysis for the Global Burden of Disease Study 2021. The Lancet Rheumatology 2023 Sep;5(9):e508-e522. [doi: 10.1016/S2665-9913(23)00163-7]
3. Kohn MD, Sassoon AA, Fernando ND. Classifications in brief: Kellgren-Lawrence classification of osteoarthritis. Clin Orthop Relat Res 2016 Aug;474(8):1886-1893. [doi: 10.1007/s11999-016-4732-4] [Medline: 26872913]
4. Bedson J, Croft PR. The discordance between clinical and radiographic knee osteoarthritis: a systematic search and summary of the literature. BMC Musculoskelet Disord 2008 Sep 2;9:116. [doi: 10.1186/1471-2474-9-116] [Medline: 18764949]
5. Hannan MT, Felson DT, Pincus T. Analysis of the discordance between radiographic changes and knee pain in osteoarthritis of the knee. J Rheumatol 2000 Jun;27(6):1513-1517. [Medline: 10852280]
6. Favre J, Jolles BM. Gait analysis of patients with knee osteoarthritis highlights a pathological mechanical pathway and provides a basis for therapeutic interventions. EFORT Open Rev 2016 Oct;1(10):368-374. [doi: 10.1302/2058-5241.1.000051] [Medline: 28461915]

XSL•FO
RenderX

7. Wang W, Li F, Guo J, Zhang Z. Changes in gastrocnemius MTU stiffness and their correlation with plantar pressure in patients with knee osteoarthritis. Front Bioeng Biotechnol 2024;12:1378031. [doi: 10.3389/fbioe.2024.1378031] [Medline: 38784765]
8. Pan J, Xie Z, Shen H, Luan J, Zhang X, Liao B. Three-dimensional gait biomechanics in patients with mild knee osteoarthritis. Sci Rep 2025;15(1):32061. [doi: 10.1038/s41598-025-17398-z]
9. Patel S, Park H, Bonato P, Chan L, Rodgers M. A review of wearable sensors and systems with application in rehabilitation. J Neuroeng Rehabil 2012 Apr 20;9:21. [doi: 10.1186/1743-0003-9-21] [Medline: 22520559]
10. Abdul Razak AH, Zayegh A, Begg RK, Wahab Y. Foot plantar pressure measurement system: a review. Sensors 2012;12(7):9884-9912. [doi: 10.3390/s120709884]
11. Diamond LE, Grant T, Uhlrich SD. Osteoarthritis year in review 2023: biomechanics. Osteoarthr Cartil 2024 Feb;32(2):138-147. [doi: 10.1016/j.joca.2023.11.015]
12. Shu L, Mai KY, Tao XM, et al. Monitoring diabetic patients by novel intelligent footwear system. Presented at: 2012 International Conference on Computerized Healthcare (ICCH); Dec 17-18, 2012. [doi: 10.1109/ICCH.2012.6724478]
13. Wipperman MF, Lin AZ, Gayvert KM, et al. Digital wearable insole-based identification of knee arthropathies and gait signatures using machine learning. Elife 2024 Apr 30;13:e86132. [doi: 10.7554/eLife.86132] [Medline: 38686919]
14. Collins GS, Dhiman P, Ma J, et al. Evaluation of clinical prediction models (part 1): from development to external validation. BMJ 2024 Jan 8;384:e074819. [doi: 10.1136/bmj-2023-074819] [Medline: 38191193]
15. Panyarachun P, Angthong C, Jindasakchai P, Rajbhandari P, Rungrattanawilai N. Abnormal foot pressure in older adults with knee osteoarthritis: a systematic review. Eur Rev Med Pharmacol Sci 2022 Sep;26(17):6236-6241. [doi: 10.26355/eurrev_202209_29646] [Medline: 36111924]
16. Lidtke RH, Muehleman C, Kwasny M, Block JA. Foot center of pressure and medial knee osteoarthritis. J Am Podiatr Med Assoc 2010;100(3):178-184. [doi: 10.7547/1000178] [Medline: 20479447]
17. Saito I, Okada K, Nishi T, et al. Foot pressure pattern and its correlation with knee range of motion limitations for individuals with medial knee osteoarthritis. Arch Phys Med Rehabil 2013 Dec;94(12):2502-2508. [doi: 10.1016/j.apmr.2013.07.017] [Medline: 23916619]
18. He YJ, Zheng XL, Wang DF, et al. Static and dynamic plantar pressure distribution in 94 patients with different stages of unilateral knee osteoarthritis using the Footscan® platform system: an observational study. Med Sci Monit 2023 Jan 12;29:e938485. [doi: 10.12659/MSM.938485] [Medline: 36632023]
19. Li G, Li S, Xie J, et al. Identifying changes in dynamic plantar pressure associated with radiological knee osteoarthritis based on machine learning and wearable devices. J Neuroeng Rehabil 2024 Apr 3;21(1):45. [doi: 10.1186/s12984-024-01337-6] [Medline: 38570841]
20. Xie J, Li S, Song Z, et al. Functional monitoring of patients with knee osteoarthritis based on multidimensional wearable plantar pressure features: cross-sectional study. JMIR Aging 2024 Nov 25;7:e58261. [doi: 10.2196/58261] [Medline: 39586093]
21. Qualter JM, Chertok AN, Buhler A, Laende EK, Costello KE. Osteoarthritis year in review 2025: biomechanics. Osteoarthr Cartil 2026 Jun;34(6):767-775. [doi: 10.1016/j.joca.2026.01.003]
22. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics 2006 Feb 23;7:91. [doi: 10.1186/1471-2105-7-91] [Medline: 16504092]
23. Cawley GC, Talbot NLC. On over-fitting in model selection and subsequent selection bias in performance evaluation. J Mach Learn Res 2010;11:2079-2107. [doi: 10.5555/1756006.1859921]
24. Tolstikhin IO, Houlsby N, Kolesnikov A, et al. MLP-mixer: an all-MLP architecture for vision. arXiv. Preprint posted online on May, 2021. [doi: 10.48550/arXiv.2105.01601]
25. Chen SA, Li CL, Arik SO, Yoder N, Pfister T. TSMixer: an all-MLP architecture for time series forecasting. Transactions on Machine Learning Research. Preprint posted online on 2023. [doi: 10.48550/arXiv.2303.06053]
26. Li Y, Vinyals O. Representation learning with contrastive predictive coding. arXiv 2018. [doi: 10.48550/arXiv.1807.03748]
27. Zhang K, Wen Q, Zhang C, et al. Self-supervised learning for time series analysis: taxonomy, progress, and prospects. IEEE Trans Pattern Anal Mach Intell 2024 Oct;46(10):6775-6794. [doi: 10.1109/TPAMI.2024.3387317] [Medline: 38598381]
28. Bendat JS, Piersol AG. Random data: analysis and measurement procedures. In: Hoboken, 4th edition: Wiley; 2010.
29. Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Time Series Analysis: Forecasting and Control, 5th edition: Wiley; 2015.
30. Breiman L. Random forests. Mach Learn 2001 Oct;45(1):5-32. [doi: 10.1023/A:1010933404324]
31. Ismail Fawaz H, Lucas B, Forestier G, et al. InceptionTime: finding AlexNet for time series classification. Data Min Knowl Disc 2020 Nov;34(6):1936-1962. [doi: 10.1007/s10618-020-00710-y]
32. Wang Y, Huang N, Li T, Yan Y, Zhang X. Medformer: a multi-granularity patching transformer for medical time-series classification. Presented at: Advances in Neural Information Processing Systems 37; Dec 10-15, 2024; Vancouver, BC, Canada p. 36314-36341. [doi: 10.52202/079017-1145]
33. Liu H, Yang D, Liu X, et al. TodyNet: temporal dynamic graph neural network for multivariate time series classification. Inf Sci (Ny) 2024 Aug;677:120914. [doi: 10.1016/j.ins.2024.120914]
34. Meinshausen N, Bühlmann P. Stability selection. J R Stat Soc Series B Stat Methodol 2010 Sep 1;72(4):417-473. [doi: 10.1111/j.1467-9868.2010.00740.x]

XSL•FO
RenderX

35. Nogueira S, Sechidis K, Brown G. On the stability of feature selection algorithms. J Mach Learn Res 2018;18(174):1-54. [doi: 10.5555/3122009.3242031]
36. Parvandeh S, Yeh HW, Paulus MP, McKinney BA. Consensus features nested cross-validation. Bioinformatics 2020 May 1;36(10):3093-3098. [doi: 10.1093/bioinformatics/btaa046] [Medline: 31985777]
37. Moriguchi M, Maeshige N, Ueno M, Yoshikawa Y, Terashi H, Fujino H. Modulation of plantar pressure and gastrocnemius activity during gait using electrical stimulation of the tibialis anterior in healthy adults. PLOS ONE 2018;13(5):e0195309. [doi: 10.1371/journal.pone.0195309] [Medline: 29746498]
38. Anan M, Tokuda K, Tanimoto K, Sawada T. Coordination of joint movement during gait in knee osteoarthritis: insights from uncontrolled manifold analysis. J Biomech 2024 Nov;176:112305. [doi: 10.1016/j.jbiomech.2024.112305] [Medline: 39260234]
39. Taniguchi M, Umehara J, Yamagata M, et al. Understanding muscle coordination during gait based on muscle synergy and its association with symptoms in patients with knee osteoarthritis. Clin Rheumatol 2024 Feb;43(2):743-752. [doi: 10.1007/s10067-023-06852-w] [Medline: 38133793]
40. Wang M, Zhang C, Yang Z, Cheng T, Lan C, Mo F. Muscle activation patterns and gait changes in unilateral knee osteoarthritis patients: a comparative study with healthy controls. Clin Rheumatol 2024 Sep;43(9):2963-2972. [doi: 10.1007/s10067-024-07057-5] [Medline: 39088118]
41. Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. Front Public Health 2017;5:307. [doi: 10.3389/fpubh.2017.00307] [Medline: 29209603]
42. Riley RD, Snell KIE, Archer L, et al. Evaluation of clinical prediction models (part 3): calculating the sample size required for an external validation study. BMJ 2024 Jan 22;384:e074821. [doi: 10.1136/bmj-2023-074821] [Medline: 38253388]
43. Chatzaki C, Skaramagkas V, Tachos N, et al. The smart-insole dataset: gait analysis using wearable sensors with a focus on elderly and Parkinson's patients. Sensors (Basel) 2021 Apr 16;21(8):2821. [doi: 10.3390/s21082821] [Medline: 33923809]
44. Larracy R, Phinyomark A, Salehi A, et al. A dataset of high-resolution plantar pressures for gait analysis across varying footwear and walking speeds. Sci Data 2025 Aug 13;12(1):1415. [doi: 10.1038/s41597-025-05792-1] [Medline: 40804054]
45. Kim NI, Lee JM, Moradnia M, et al. Biocompatible composite thin-film wearable piezoelectric pressure sensor for monitoring of physiological and muscle motions. Soft Sci 2022;2(2):8. [doi: 10.20517/ss.2022.06]
46. Sun G, Wang P, Jiang Y, Sun H, Meng C, Guo S. Recent advances in flexible and soft gel-based pressure sensors. Soft Sci 2022;2(4):17. [doi: 10.20517/ss.2022.16]
47. Shahzad A, Wei JA, Su X, et al. Recent advances in electrospun fiber-based flexible pressure sensors for next-gen healthcare applications: a review. Soft Sci 2026;6(1):5. [doi: 10.20517/ss.2025.76]
48. ROA identification from plantar pressure. GitHub. URL: https://github.com/XinGuanyu/ROA_identification [accessed 2026-08-28]

Abbreviations

AUROC: area under the receiver operating characteristic curve
CoP: center of pressure
IRB: Institutional Review Board
JARS: *Journal Article Reporting Standards*
K-L: Kellgren-Lawrence
MLP: multilayer perceptron
PROC: Pearl River Osteoarthritis Cohort
PSI: Plantar Synergy Index
RF: random forest
ROA: radiographic knee osteoarthritis

Edited by S Brini; submitted 11.May.2026; peer-reviewed by CR Torres-SanMiguel, FJ Gallegos-Funes, H Zhang; revised version received 30.Jul.2026; accepted 04.Aug.2026; published 11.Sep.2026.

Please cite as:
Xin G, Yao H, Qiao J, Yao H, Li S, Liu L, Zhang J, Qin K, Shu L
Radiographic Knee Osteoarthritis Identification Using Wearable Plantar Pressure: Secondary Cross-Sectional Analysis of Interpretable Framework Integrating Biomechanical Descriptors and the Multiscale Synergy Index
JMIR Mhealth Uhealth 2026;14:e101026
URL: https://mhealth.jmir.org/2026/1/e101026
doi:10.2196/101026

XSL•FO
RenderX

© Guanyu Xin, Hongbo Yao, Jingyuan Qiao, Huilin Yao, Siyuan Li, Lu Liu, Jingyi Zhang, Ke Qin, Lin Shu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Mediating Role of Sleep Quality in Associations of Smartphone Addiction and Time-Specific Use With Prenatal Depression and Anxiety: Cross-Sectional Observational Study

Xiangxin Zhang[1*], MBChB; Mixia Ouyang[2*], MD; Xiaoli Wu[2], MD; Xinghua Yang[1], MD; Xinyi Wang[1], MBChB; Yuxinqiao Cao[1], MBChB; Mingmin Xu[2], MD; Yanling Shu[3], PhD; Mingyang Wu[1], PhD

[1]Department of Maternal and Child Health, Xiangya School of Public Health, Central South University, No. 172 Tongzipo Road, Changsha, Hunan, China

[2]Hunan Provincial Maternal and Child Health Hospital, Changsha, Hunan, China

[3]Department of Maternal and Child Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China

[*]these authors contributed equally

Corresponding Author:
Mingyang Wu, PhD
Department of Maternal and Child Health, Xiangya School of Public Health, Central South University, No. 172 Tongzipo Road, Changsha, Hunan, China

Abstract

Background: Prenatal depression and anxiety have emerged as pressing global public health challenges. In the digital era, excessive smartphone use may exert complex effects on maternal mental health. Prior research has predominantly focused on adolescents, whereas evidence among pregnant women, particularly regarding total weekly use and use during critical circadian windows, such as after waking and before bedtime, remains limited. Furthermore, the mediating role of sleep quality in these associations has yet to be systematically examined.

Objective: This study aimed to investigate the associations of smartphone addiction and time-specific smartphone use with prenatal depression and anxiety, and to assess the mediating role of sleep quality in these associations.

Methods: This cross-sectional observational study enrolled 537 pregnant women. Smartphone use characteristics, sleep quality, and prenatal depression and anxiety were assessed using the Smartphone Addiction Scale-Short Version (SAS-SV), objective total weekly smartphone use recorded by the device, and self-reported duration of smartphone use after waking and duration of bedtime smartphone use; the Pittsburgh Sleep Quality Index (PSQI), the Patient Health Questionnaire-9 (PHQ-9), and the Generalized Anxiety Disorder-7 (GAD-7), respectively. Binary logistic regression was used to estimate associations, restricted cubic spline (RCS) models were used to explore nonlinear patterns, and exploratory mediation analyses were used to examine whether sleep quality statistically accounted for part of the observed associations.

Results: The prevalence of moderate-to-severe prenatal depression and anxiety was 8.8% (47/537) and 5.8% (31/537), respectively. After adjusting for covariates, smartphone addiction was associated with higher odds of depression (odds ratio [OR] 3.461, 95% CI 1.771 - 7.100) and anxiety (OR 7.738, 95% CI 3.064 - 23.172). Duration of smartphone use after waking ≥61 minutes and duration of bedtime smartphone use ≥91 minutes were also associated with higher odds of affective disorders. RCS analyses further revealed distinct nonlinear patterns across different exposure dimensions: duration of smartphone use after waking displayed U-shaped and J-shaped associations with depression and anxiety, respectively, while total weekly use displayed a U-shaped association with anxiety. Exploratory mediation analyses showed that the indirect pathways through sleep quality were statistically significant for the associations of bedtime smartphone use, weekly smartphone use, and smartphone addiction with prenatal depression, with mediation proportions of 44.89% (0.0065/0.0145), 31.30% (0.0056/0.0183), and 31.48% (0.0415/0.1315), respectively. The corresponding proportions in the anxiety models were 35.83% (0.0044/0.0123), 27.05% (0.0039/0.0144), and 22.73% (0.0271/0.1187).

Conclusions: Smartphone addiction and excessive smartphone use after waking and before bedtime were associated with higher odds of prenatal depression and anxiety. Sleep quality may statistically account for part of these associations, but longitudinal studies are needed to establish temporal ordering and causal pathways.

(JMIR Mhealth Uhealth 2026;14:e100064) doi:10.2196/100064

XSL•FO
RenderX

KEYWORDS

smartphone addiction; smartphone use; sleep quality; antenatal anxiety; antenatal depression; prenatal depression; prenatal anxiety; time-specific smartphone use; restricted cubic spline; mediating effect

Introduction

Pregnancy represents a critically vulnerable period in a woman's life course, both physiologically and psychologically [1-3]. During this period, women not only endure physiological challenges such as dramatic hormonal fluctuations and early pregnancy symptoms, but also face multiple psychosocial stressors, including shifts in social roles, restructuring of family relationships, and concerns about fetal well-being [4,5]. Although pregnancy is generally regarded as a positive life event, it is frequently accompanied by significant affective disturbances. A systematic review reported that the global average prevalence of prenatal depression is approximately 28.5%, which is substantially higher than that of postpartum depression [6]. In mainland China, a systematic review encompassing 271 studies and 369,477 women found that the average prevalence of prenatal anxiety reached 17.4% [7]. These common prenatal affective disorders, if left unaddressed, can substantially increase the risk of adverse obstetric outcomes, such as preterm birth, low birth weight, and preeclampsia [8,9].

In the digital era, smartphones have become an integral extension of daily life. For pregnant women, smartphones serve as important tools for accessing health information and maintaining social support. Healthy smartphone use can, to some extent, alleviate the sense of isolation and uncertainty brought about by the temporary disruption of social roles [10]. Conversely, the fragmented entertainment offered by smartphones can easily trigger overuse, blurring the self-perceived boundary between functional use and pathological dependence [11]. Against this backdrop, the positive mental health effects of smartphone use may be reversed, turning it into a potential risk factor that jeopardizes maternal prenatal mental health [12].

Smartphone addiction, also referred to as problematic smartphone use, is defined as excessive dependence on smartphones, characterized by core features of behavioral addiction such as withdrawal anxiety, increased tolerance, and impaired social functioning [13,14]. In recent years, the adverse effects of smartphone addiction on mental health have been widely documented across diverse populations. Large-scale studies among adolescents, in particular, have demonstrated significant positive associations of problematic smartphone use with depressive and anxiety symptoms as well as sleep disturbance [15-17]. Moreover, several longitudinal studies have shown that baseline levels of smartphone addiction prospectively predict the worsening of anxiety symptoms and the continued deterioration of sleep quality during follow-up, suggesting that smartphone addiction may not merely be a consequence of psychological distress but a potential cause that precipitates or exacerbates affective disorders [18,19].

Although previous research has established negative associations between smartphone addiction and mental health in adolescents [20,21], studies examining the association between smartphone addiction and prenatal affective disorders among pregnant women, a population characterized by marked physiological and psychological fluctuations, remain scarce. In addition, few studies have distinguished total smartphone use time from use during specific periods. This distinction may be theoretically relevant: circadian rhythm disruption plays a key role in the onset, recurrence, and treatment response of mood disorders, and the central role of circadian rhythms in emotional regulation is widely recognized [22,23]. Screen exposure during different circadian phases may influence emotional states through distinct neurobiological pathways. Specifically, bedtime screen exposure is closely associated with reduced sleep quality [24], whereas screen exposure after waking may have different effects on the daytime stress regulation system [25]. Compared with absolute usage duration, whether screen exposure during key circadian windows, such as bedtime and after waking, is associated with prenatal depressive and anxiety symptoms has yet to be systematically investigated.

In recent years, studies on adolescents have shown that smartphone use is associated with reduced sleep quality and emotional problems [20,26,27]. Experimental studies have further found that nighttime blue light exposure can prolong sleep latency and cause a circadian phase shift by suppressing melatonin secretion [28], while the cognitive and emotional arousal triggered by immersive screen exposure may also interfere with sleep onset and maintenance [29]. It is worth noting that sleep quality in pregnant women undergoes a physiological deterioration as pregnancy progresses, with the prevalence of sleep disturbances rising significantly from the first trimester to the third trimester [30]. This vulnerability may stem from the combined effects of hormonal fluctuations, increased nocturia, and physical discomfort [31]. Furthermore, poor sleep quality is associated with an increased risk of prenatal depression and anxiety [32,33], and its neural mechanisms may involve the effects of sleep deprivation on prefrontal cortex function and amygdala reactivity [34,35]. Based on this prior evidence and biological plausibility, it is reasonable to infer that sleep quality may mediate the relationship between smartphone use and prenatal affective disorders. However, current research on this topic remains scarce, particularly studies that comprehensively assess multiple dimensions, such as smartphone use at specific time points and smartphone addiction.

To address this research gap, this study aims to investigate the associations between smartphone addiction and usage during specific time periods and prenatal depression and anxiety, and to explore the mediating effect of sleep quality on these associations, to provide evidence-based support for the development of digital health guidelines and clinical intervention strategies during pregnancy.

XSL•FO
RenderX

Methods

Study Population

This study used a cross-sectional design and was conducted in the obstetrics outpatient clinic and inpatient ward of Hunan Provincial Maternal and Child Health Hospital. Participants were recruited using convenience sampling. The survey was administered by uniformly trained interviewers, who explained the study's purpose and procedures to eligible pregnant women while they were waiting to give birth in the admission area or in the ward. After obtaining informed consent, the interviewers provided one-on-one guidance to help participants scan a QR code to complete the electronic questionnaire, which was filled out and submitted on the spot. Investigators provided necessary explanations and assistance but did not interfere with the respondents' answers. Inclusion criteria were as follows: (1) permanent residents of Changsha, (2) all prenatal checkups completed at the study hospital, (3) age ≥18 years, (4) singleton pregnancy. Exclusion criteria were (1) severe cognitive impairment or mental illness preventing effective communication or understanding of the questionnaire content, and (2) presence of legally reportable infectious diseases such as HIV, syphilis, or hepatitis B. Participants who met all inclusion criteria and none of the exclusion criteria were formally enrolled in the study after signing the informed consent form.

All data in this study were collected by having participants scan a QR code to complete a structured questionnaire. All key research variables in the questionnaire were marked as required fields; participants could not submit the questionnaire unless they had completed all required items. Consequently, none of the variables analyzed in this study had missing values, and all 537 cases ultimately included in the analysis were valid questionnaires.

Ethical Considerations

The study protocol, encompassing participant recruitment procedures, informed consent procedures, data collection, and quality control methods, was reviewed and approved by the Ethics Committee of the Xiangya School of Public Health, Central South University (approval number: XYGW-2024 - 19) and the Ethics Committee of the Hunan Provincial Maternal and Child Health Care Hospital (approval number: 2023-S189). After fully understanding the study's objectives, procedures, potential risks, and benefits, all participants voluntarily signed a paper informed consent form and were informed of their right to withdraw from the study at any time without condition. No remuneration or compensation was provided to any participant in this study.

Exposure Measurement

Smartphone addiction tendency was assessed using the Smartphone Addiction Scale-Short Version (SAS-SV). The scale consists of 10 items rated on a 6-point Likert scale (1="strongly disagree" to 6="strongly agree"), yielding a total score ranging from 10 to 60, with higher scores indicating a greater risk of smartphone addiction. According to the developers' recommendation, a cut-off score of 33 was adopted to define a high risk of addiction for females. The SAS-SV has demonstrated good psychometric properties, with an internal consistency Cronbach α coefficient of 0.91 [36]. The scale has been widely applied and validated in Chinese populations and is suitable for the rapid screening of smartphone dependence among adults [37].

Time-specific smartphone use was assessed using self-developed items, including duration of bedtime smartphone use and duration of smartphone use after waking. Bedtime smartphone use was defined as the cumulative duration of smartphone use from the moment participants lay down in bed with the intention to sleep until they actually fell asleep. Smartphone use after waking was defined as the cumulative duration of use from the moment participants awoke in the morning until they got out of bed. Both definitions were clearly stated in the questionnaire and were verbally explained in a standardized manner by trained investigators before questionnaire completion. Investigators specifically emphasized that the reported duration should only include time spent actively using the phone while lying in bed, rather than the total time spent in bed including nonuse states. Meanwhile, participants were instructed to recall their typical usage patterns over the past week rather than relying on a single day's experience. Both measures were specific to nighttime sleep periods and did not include daytime naps, and were recorded in minutes. Weekly smartphone use was obtained by asking participants to check the average daily usage duration of all applications over the past week through the screen time management feature built into their smartphone system settings (corresponding to "Screen Time" or "Healthy Phone Usage Time" or "App Management Time" in Android and Apple iOS systems, respectively), and was recorded in hours. According to the current distribution of data and previous studies on smartphone use duration, we classified participants into 4 consecutive groups as follows: (1) duration at bedtime [25]: 0 - 30 minutes, 31 - 60 minutes, 61 - 90 minutes, and ≥91 minutes; (2) duration after waking [38]: 0 - 1 minutes, 2 - 10 minutes, 11 - 60 minutes, and ≥61 minutes; (3) duration of total week [39,40]: 0 - 14 hours/week, 15 - 28 hours/week, 29 - 49 hours/week, and ≥50 hours/week.

Participants' sleep quality over the past month was assessed using the Pittsburgh Sleep Quality Index (PSQI). Developed by Buysse et al [41], the PSQI comprises 7 dimensions: subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleep medication, and daytime dysfunction. The total score, ranging from 0 to 21, is calculated as the sum of all component scores. Based on the threshold established by previous studies, a total score of ≥7 was defined as poor sleep quality in this study [42,43].

Outcome Measurement

Depressive symptoms over the past 2 weeks were assessed using the Patient Health Questionnaire-9 (PHQ-9). Developed by Kroenke et al [44] based on the DSM-IV diagnostic criteria for depression, the PHQ-9 comprises nine items. Each item asks respondents to rate the frequency of each symptom over the past 2 weeks on a 4-point Likert scale (0="not at all" to 3="nearly every day"), yielding a total score ranging from 0 to 27, with higher scores indicating more severe depressive

XSL•FO
RenderX

symptoms. The PHQ-9 has demonstrated good reliability and validity among Chinese pregnant women [45]. As prior research has shown that moderate-to-severe depressive symptoms may lead to more adverse outcomes, a cut-off score of ≥10 was adopted in this study to define moderate-to-severe depressive symptoms, in line with previous studies [46,47]. Hereafter, the term "prenatal depression" refers to moderate-to-severe prenatal depressive symptoms.

Anxiety symptoms were measured using the Generalized Anxiety Disorder-7 (GAD-7). Developed by Spitzer et al [48], the GAD-7 is designed for rapid screening of generalized anxiety symptoms over the past 2 weeks. The scale consists of 7 self-reported items, each requiring respondents to rate the frequency of anxiety symptoms over the past 2 weeks on a 4-point Likert scale (0="not at all" to 3="nearly every day"), with total scores ranging from 0 to 21, higher scores reflecting more severe anxiety symptoms. Given that moderate-to-severe anxiety symptoms may be associated with more detrimental effects, a cut-off score of ≥10 was used to define moderate-to-severe anxiety, consistent with previous research [49,50]. Hereafter, the term "prenatal anxiety" refers to moderate-to-severe prenatal anxiety symptoms.

Covariates

Based on previous studies [12,51-55], the following variables that may influence smartphone use and prenatal mental health outcomes were included as covariates in the analyses, covering 3 domains: sociodemographic characteristics, behavioral and lifestyle factors, and health and pregnancy-related characteristics. Sociodemographic characteristics included maternal age (years), maternal educational attainment (junior high school or below, high school or vocational high school, and college or above), and paternal educational attainment (junior high school or below, high school or vocational high school, and college or above). Behavioral and lifestyle factors included prepregnancy smoking status within the 6 months before pregnancy (yes or no), prepregnancy alcohol consumption within the 6 months before pregnancy (yes or no), frequency of physical exercise during pregnancy (<once per week, 1 - 3 times per week, 4 - 6 times per week, and ≥ once per day), and frequency of social media posting on platforms such as WeChat (Tencent) and Xiaohongshu (Xingyin Information Technology Ltd.) during pregnancy (less than once every 6 months, 2 - 5 times per year, 1 - 2 times per month, 3 - 4 times per month, and several times per week). Health- and pregnancy-related characteristics included gestational age (weeks), prepregnancy BMI (kg/m^2), gestational weight gain (kg), gestational diabetes mellitus (yes or no), abortus (times), and gravida (times).

Statistical Analysis

Descriptive statistics were computed for the dataset, with continuous variables presented as means and SD and categorical variables expressed as counts and percentages. Between-group comparisons were performed using independent-samples *t* tests for continuous variables and chi-square tests for categorical variables. Pearson correlation coefficients were used to assess the bivariate associations among total weekly smartphone use duration, duration of smartphone use after waking, duration of bedtime smartphone use, SAS-SV score, PHQ-9 score, and GAD-7 score.

Subsequently, binary logistic regression models were fitted separately for each smartphone-use dimension and outcome. Two models were constructed to verify the robustness of the associations: Model 1 was a crude model that assessed the initial associations without adjusting for any covariates; Model 2 was a fully adjusted model that included maternal age, maternal educational attainment, paternal educational attainment, prepregnancy BMI, gestational weight gain, gestational age, gestational diabetes mellitus, gravida, abortus, prepregnancy smoking status, prepregnancy alcohol consumption, frequency of physical exercise during pregnancy, and frequency of social media posting during pregnancy as covariates. Results were expressed as odds ratios (ORs) with their 95% CIs.

To further explore potential nonlinear dose-response relationships between smartphone use, sleep quality, and prenatal affective disorders, restricted cubic spline (RCS) functions were fitted on the basis of the above logistic regression models. Following previous studies, 4 knots were placed at the fifth, 35th, 65th, and 95th percentiles, and the median of each independent variable was set as the reference value [56-58]. Finally, mediation models were established to exploratorily examine the indirect pathways of sleep quality in the associations between multidimensional smartphone use and prenatal affective disorders. In our mediation analysis, we clearly distinguished between the total effect, the direct effect, and the indirect effect based on the effect decomposition framework. We used the continuous scores of the PHQ-9 and GAD-7 as outcome variables to preserve the informational integrity of the original scales and enhance the statistical power for detecting indirect effects. Point estimates for direct and indirect effects were obtained using linear regression models, and their 95% CIs were estimated via 1000 Monte Carlo simulations; the mediation proportion was calculated by dividing the indirect effect by the total effect. All models were adjusted for all abovementioned covariates. In addition, to examine the impact of outcome definition on the stability of mediation results, we conducted a sensitivity analysis based on a binary outcome (PHQ-9 and GAD-7 score ≥10). Given the cross-sectional design of this study, the results of the mediation analysis should be interpreted as an exploratory decomposition of association pathways rather than evidence of causal mechanisms.

Given the limited number of events for moderate-to-severe depression (n=47) and anxiety (n=31), two sensitivity analyses were conducted to assess the robustness of the main findings. First, Firth penalized maximum likelihood estimation was applied to refit all models to reduce small-sample bias in rare-events logistic regression [59]. Second, the threshold for the PHQ-9 and GAD-7 outcomes was lowered from ≥10 to ≥5 (ie, including those with mild and above depression and anxiety symptoms) to evaluate whether the observed associations were sensitive to the choice of the cutoff for symptom severity.

Data analyses were performed using SPSS 26.0 (IBM Corp) and R 4.4.1 (R Foundation for Statistical Computing). The primary R packages used included *dplyr*, *rms*, *mediation*, and

XSL•FO
RenderX

ggplot2. The significance level was set at α=.05, and *P*<.05 was considered statistically significant.

Results

Baseline Characteristics of Participants

Table 1 presents the baseline characteristics of the study participants. A total of 537 pregnant women were included, with a mean age of 31.7 (4.0) years. Moderate-to-severe prenatal depression and anxiety were present in 47 (8.8%) and 31 (5.8%) participants, respectively. The mean SAS-SV score (SD) in the moderate-to-severe depression group was 36.4 (9.4), significantly higher than that in the nonmoderate-to-severe depression group (30.8, SD 8.9). The moderate-to-severe depression group also had longer weekly screen time and significantly longer durations of smartphone use after waking and before bedtime compared with the nonmoderate-to-severe depression group (*P*=.02 and *P*=.002, respectively). In addition, the PSQI score in the moderate-to-severe depression group was significantly higher than that in the nonmoderate-to-severe depression group (*P*<.001). The pattern of differences between the moderate-to-severe anxiety group and the nonmoderate-to-severe anxiety group was generally similar. The overall detection rate of sleep disturbance was 46.2% (n=248), and the detection rates in both the moderate-to-severe depression group and the moderate-to-severe anxiety group were significantly higher than those in their respective control groups (*P*=.003 and *P*=.008, respectively).

XSL•FO
RenderX

Table . Baseline characteristics of the study participants (N=537)[a].

Variables	Total	Prenatal depression (PHQ-9[b] score≥10)			Prenatal anxiety (GAD-7[c] score≥10)		
		No	Yes	*P* value	No	Yes	*P* value
Age, mean (SD) years	31.7 (4.0)	31.7 (4.0)	32.0 (3.6)	.60	31.7 (4.0)	32.3 (4.5)	.40
BMI, mean (SD) kg/m²	21.0 (2.9)	21.1 (2.9)	20.8 (3.0)	.48	21.0 (2.9)	21.2(2.8)	.78
Gravida, mean (SD) times	2.0 (1.1)	2.0 (1.1)	2.1 (1.4)	.29	2.0 (1.1)	2.0 (1.2)	.85
Abortus, mean (SD) times	1.6 (0.9)	1.5 (0.8)	1.7 (1.0)	.30	1.6 (0.9)	1.7 (0.8)	.55
Weight gain during pregnancy, mean (SD) kg	15.2 (5.9)	15.3 (5.9)	14.0 (5.4)	.13	15.2 (5.8)	14.6 (7.0)	.56
Gestational age, mean (SD) weeks	38.87 (1.43)	38.86 (1.44)	38.98 (1.39)	.86	38.87 (1.43)	38.91 (1.46)	.94
Gestational diabetes mellitus, n (%)				.58			.46
No	441 (82.1)	401 (81.8)	40 (85.1)		414 (81.8)	27 (87.1)	
Yes	96 (17.9)	89 (18.2)	7 (14.9)		92 (18.2)	4 (12.9)	
Maternal education attainment, n (%)				.02[d]			.61
Junior high school or below	19 (3.5)	15 (3.1)	4 (8.5)		17 (3.4)	2 (6.5)	
High school or vocational high school	43 (8.0)	36 (7.3)	7 (14.9)		40 (7.9)	3 (9.7)	
College or above	475 (88.5)	439 (89.6)	36 (76.6)		449 (88.7)	26 (83.9)	
Paternal education attainment, n (%)				.57			.29
Junior high school or below	30 (5.6)	26 (5.3)	4 (8.5)		29 (5.7)	1 (3.2)	
High school or vocational high school	58 (10.8)	52 (10.6)	6 (12.8)		52 (10.3)	6 (19.4)	
College or above	449 (83.6)	412 (84.1)	37 (78.7)		425 (84.0)	24 (77.4)	
Prepregnancy smoking, n (%)				>.99			.09
No	511 (95.2)	466 (95.1)	45 (95.7)		484 (95.7)	27 (87.1)	
Yes	26 (4.8)	24 (4.9)	2 (4.3)		22 (4.3)	4 (12.9)	
Prepregnancy alcohol consumption, n (%)				.48			.41
No	373 (69.5)	343 (70.0)	30 (63.8)		354 (70.0)	19 (61.3)	
Yes	164 (30.5)	147 (30.0)	17 (36.2)		152 (30.0)	12 (38.7)	
Frequency of physical exercise, n (%)				.25			.30
Less than once a week	223 (41.5)	197 (40.2)	26 (55.3)		206 (40.7)	17 (54.8)	
1 - 3 times a week	199 (37.1)	186 (38.0)	13 (27.7)		192 (37.9)	7 (22.6)	
4 - 6 times a week	62 (11.5)	58 (11.8)	4 (8.5)		59 (11.7)	3 (9.7)	

XSL•FO
RenderX

Variables	Total	Prenatal depression (PHQ-9[b] score≥10)			Prenatal anxiety (GAD-7[c] score≥10)		
		No	Yes	*P* value	No	Yes	*P* value
Once a day or more	53 (9.9)	49 (10.0)	4 (8.5)		49 (9.7)	4 (12.9)	
Frequency of social media posts, n (%)				.07			.33
Less than once every six months	171 (31.8)	159 (32.4)	12 (25.5)		161 (31.8)	10 (32.3)	
2 - 5 times a year	144 (26.8)	125 (25.5)	19 (40.4)		132 (26.1)	12 (38.7)	
1 - 2 times a month	147 (27.4)	138 (28.2)	9 (19.1)		143 (28.3)	4 (12.9)	
3 - 4 times a month	53 (9.9)	50 (10.2)	3 (6.4)		49 (9.7)	4 (12.9)	
Several times a week	22 (4.1)	18 (3.7)	4 (8.5)		21 (4.2)	1 (3.2)	
PSQI[e] total score, mean (SD) points	7.6 (3.3)	7.3 (3.2)	9.9 (3.6)	<.001	7.4 (3.3)	9.9 (3.4)	<.001
Poor sleep quality, n (%)				.003			.008
No	289 (53.8)	274 (55.9)	15 (31.9)		280 (55.3)	9 (29.0)	
Yes	248 (46.2)	216 (44.1)	32 (68.1)		226 (44.7)	22 (71.0)	
Weekly smartphone use, mean (SD) hours	31.3 (27.6)	30.8 (26.8)	36.4 (34.2)	.18	30.8 (26.8)	39.0 (37.4)	.11
SAS-SV[f] total score, mean (SD) points	31.3 (9.1)	30.8 (8.9)	36.4 (9.4)	<.001	30.8 (9.0)	38.3 (7.6)	<.001
Smartphone addiction, n (%)				<.001			<.001
No	309 (57.5)	295 (60.2)	14 (29.8)		303 (59.9)	6 (19.4)	
Yes	228 (42.5)	195 (39.8)	33 (70.2)		203 (40.1)	25 (80.6)	
Smartphone use after waking, mean (SD) minutes	29.1 (50.6)	27.4 (50.9)	46.3 (44.2)	.02	28.1 (51.0)	45.0 (40.3)	.07
Bedtime smartphone use, mean (SD) minutes	73.1 (53.4)	70.9 (52.1)	96.5(61.6)	.002	71.3 (52.4)	102.9 (62.4)	.001

[a]Continuous variables (age, BMI, gravida, abortus, weight gain during pregnancy, gestational age, PSQI total score, weekly smartphone use, SAS-SV total score, smartphone use after waking, and bedtime smartphone use) were analyzed using an independent samples *t* test, while categorical variables (gestational diabetes mellitus, maternal education attainment, paternal education attainment, prepregnancy smoking, prepregnancy alcohol consumption, frequency of physical exercise, frequency of social media posts, poor sleep quality, and smartphone addiction) were analyzed using the chi-square test.

[b]PHQ-9: Patient Health Questionnaire-9.

[c]GAD-7: Generalized Anxiety Disorder-7.

[d]Statistical significance (*P*<.05).

[e]PSQI: Pittsburgh Sleep Quality Index.

[f]SAS-SV: Smartphone Addiction Scale-Short Version.

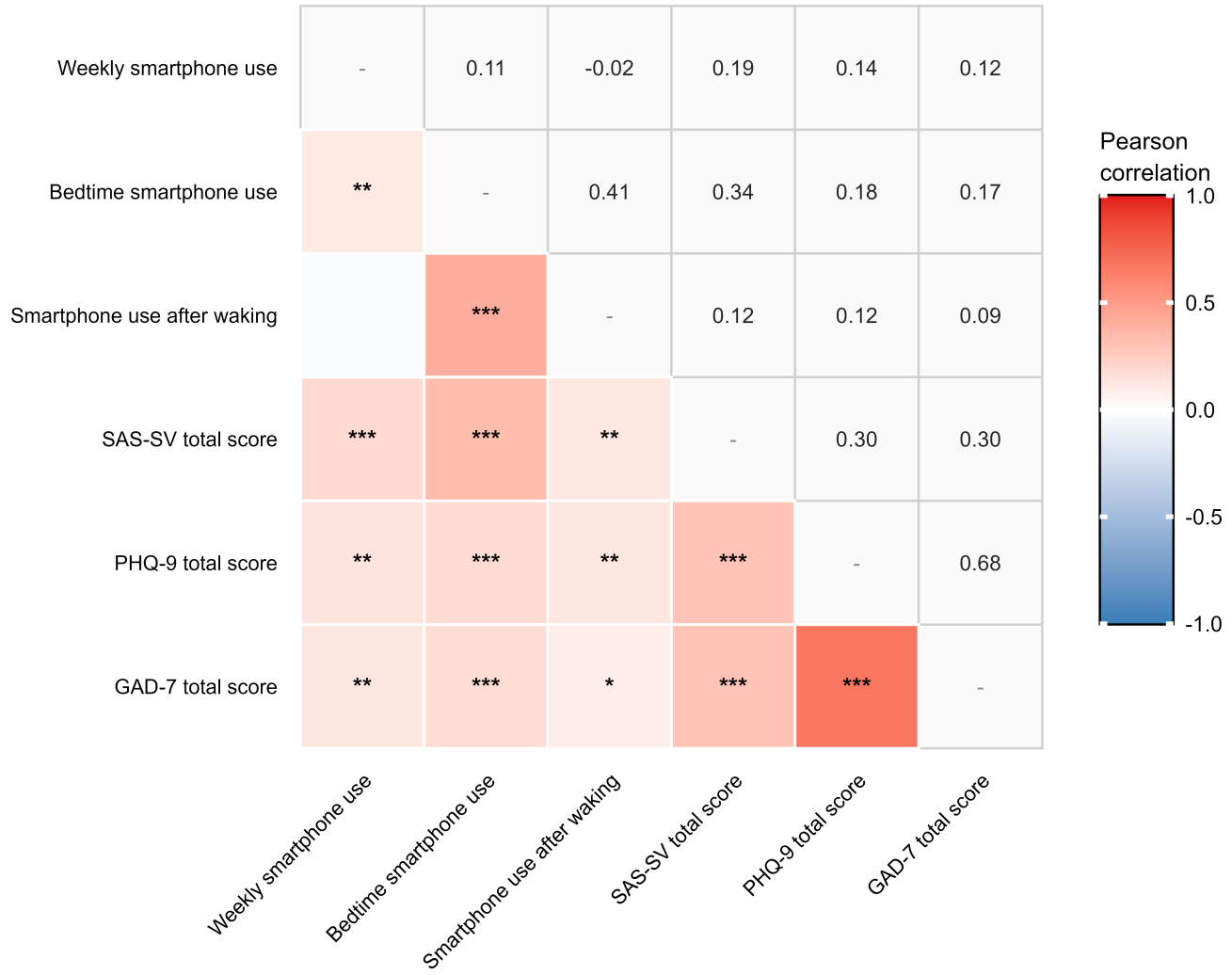
The Correlation between Different Dimensions of Smartphone Use and Prenatal Affective Disorders

To preliminarily explore the bivariate associations between multidimensional smartphone use characteristics during pregnancy (SAS-SV total score, weekly smartphone use, bedtime smartphone use, and smartphone use after waking) and prenatal affective disorders (PHQ-9 and GAD-7 total scores), Pearson correlation analysis was performed, with the results presented in Figure 1. The correlation matrix showed that both

XSL•FO
RenderX

the SAS-SV total score and all dimensions of smartphone use duration were significantly positively correlated with PHQ-9 and GAD-7 total scores (P<.05), providing preliminary evidence that excessive digital media exposure was related to prenatal affective disorders.

Figure 1. Correlation between different dimensions of smartphone use and prenatal affective disorders. These values are Pearson correlation coefficients (r). The dash "-" on the diagonal indicates that self-correlations were not computed (not applicable). *: P value <.05; **: P value <.01; ***: P value <.001. GAD-7: Generalized Anxiety Disorder-7; PHQ-9: Patient Health Questionnaire-9; SAS-SV: Smartphone Addiction Scale-Short Version.



The Associations of Different Smartphone Use Dimensions and Sleep Quality With Prenatal Affective Disorders

The results of the independent associations between various dimensions of smartphone use and prenatal affective disorders are detailed in Table 2. Smartphone addiction was associated with higher odds of both outcomes in the crude and adjusted models. In the fully adjusted model, the OR was 3.461 (95% CI 1.771 - 7.100) for prenatal depression and 7.738 (95% CI 3.064 - 23.172) for prenatal anxiety. For bedtime use, compared with 0 - 30 minutes, use of ≥91 minutes was associated with higher odds of depression (OR 3.716, 95% CI 1.467 - 10.377) and anxiety (OR 5.141, 95% CI 1.568 - 20.808). For use after waking, ≥61 minutes was associated with higher odds of depression (OR 4.758, 95% CI 1.736 - 13.773) and anxiety (OR 10.902, 95% CI 2.606 - 58.482). Weekly use showed a nonmonotonic categorical pattern: compared with 0 - 14 hours/week, the 29-49 hours/week group had lower odds of depression (OR 0.378, 95% CI 0.119 - 1.012) and anxiety (OR 0.221, 95% CI 0.033 - 0.868), whereas the ≥50 hours/week group demonstrated elevated odds estimates for both outcomes (depression OR 1.687, 95% CI 0.795 - 3.597; anxiety OR 1.979, 95% CI 0.830 - 4.795). Sensitivity analyses using Firth penalized maximum likelihood estimation and an alternative threshold (PHQ-9 and GAD-7 score ≥5) yielded results consistent with the main findings in terms of direction, supporting the robustness of our results (Tables S1 and S2 in Multimedia Appendix 1).

RCS analyses showed that bedtime smartphone use and SAS-SV score had monotonic positive associations with both outcomes, without statistically significant evidence of nonlinearity. Smartphone use after waking showed a U-shaped association with depression (P for nonlinearity=.01) and a J-shaped association with anxiety (P for nonlinearity=.009). Weekly smartphone use showed a U-shaped association with anxiety (P for nonlinearity=.01), whereas its nonlinear association with depression was not statistically significant (Figures 2 and 3).

XSL•FO
RenderX

Table . Association between different dimensions of smartphone use and prenatal affective disorders[a].

Exposure	Prenatal depression (PHQ-9[b] score≥10)			Prenatal anxiety (GAD-7[c] score≥10)		
	n/N	Model 1[d], OR[e] (95% CI)	Model 2[f], OR (95% CI)	n/N	Model 1[d], OR (95% CI)	Model 2[f], OR (95% CI)
SAS-SV[g] total score (points)	—[h]	1.069 (1.035 - 1.105)	1.070 (1.033 - 1.110)	—	1.091 (1.049 - 1.137)	1.104 (1.056 - 1.158)
No	14/309	Ref[i]	Ref	6/309	Ref	Ref
Yes	33/228	3.566 (1.897 - 7.041)	3.461 (1.771 - 7.100)	25/228	6.219 (2.671 - 16.984)	7.738 (3.064 - 23.172)
Weekly smartphone use (hours)	—	1.099 (0.951 - 1.256)	1.120 (0.959 - 1.295)	—	1.139 (0.961 - 1.327)	1.184 (0.990 - 1.398)
0 - 14	18/189	Ref	Ref	13/189	Ref	Ref
15 - 28	6/87	0.704 (0.247 - 1.749)	0.711 (0.238 - 1.880)	2/87	0.319 (0.049 - 1.187)	0.362 (0.053 - 1.471)
29 - 49	5/133	0.371 (0.120 - 0.958)	0.378 (0.119 - 1.012)	2/133	0.207 (0.032 - 0.765)	0.221 (0.033 - 0.868)
50+	18/128	1.555 (0.772 - 3.133)	1.687 (0.795 - 3.597)	14/128	1.663 (0.751 - 3.708)	1.979 (0.830 - 4.795)
Smartphone use after waking (minutes)	—	1.143 (1.004 - 1.296)	1.171 (1.000 - 1.349)	—	1.123 (0.956 - 1.279)	1.207 (0.991 - 1.433)
0 - 1	8/152	Ref	Ref	3/152	Ref	Ref
2 - 10	9/133	1.306 (0.485 - 3.578)	1.399 (0.498 - 4.011)	7/133	2.759 (0.750 - 13.006)	3.375 (0.840 - 17.270)
11 - 60	16/187	1.694 (0.723 - 4.284)	1.637 (0.667 - 4.303)	13/187	3.732 (1.176 - 16.503)	4.833 (1.375 - 23.253)
61+	14/65	4.941 (1.998 - 13.025)	4.758 (1.736 - 13.773)	8/65	6.971 (1.941 - 32.674)	10.902 (2.606 - 58.482)
Bedtime smartphone use (minutes)	—	1.248 (1.079 - 1.437)	1.238 (1.046 - 1.461)	—	1.294 (1.092 - 1.520)	1.361 (1.109 - 1.666)
0 - 30	7/170	Ref	Ref	4/170	Ref	Ref
31 - 60	13/148	2.242 (0.892 - 6.114)	1.964 (0.746 - 5.548)	7/148	2.060 (0.609 - 7.999)	2.386 (0.665 - 9.938)
61 - 90	8/84	2.451 (0.850 - 7.228)	2.073 (0.685 - 6.380)	7/84	3.773 (1.106 - 14.760)	4.116 (1.115 - 17.548)
91+	19/135	3.814 (1.618 - 10.035)	3.716 (1.467 - 10.377)	13/135	4.422 (1.523 - 15.999)	5.141 (1.568 - 20.808)

[a]Continuous exposure variables were SAS-SV total score, weekly smartphone use, smartphone use after waking, and bedtime smartphone use. For continuous exposures, ORs represent the change in odds per 1-point increase in SAS-SV total score, per 14-hour increase in weekly smartphone use, and per 30-minute increase in smartphone use after waking and bedtime smartphone use. A total SAS-SV score of 33 or higher is defined as smartphone addiction. Categorical smartphone use variables used the first category as the reference (Ref).

[b]PHQ-9: Patient Health Questionnaire-9.

[c]GAD-7: Generalized Anxiety Disorder-7.

[d]Model 1: crude model.

[e]OR: odds ratio.

[f]Model 2: adjustment for maternal age, maternal educational attainment, paternal educational attainment, prepregnancy smoking status, prepregnancy alcohol consumption, frequency of physical exercise during pregnancy, frequency of social media posts on platforms, prepregnancy BMI, weight gain during pregnancy, gravida, abortus, gestational age, and gestational diabetes mellitus.

[g]SAS-SV: Smartphone Addiction Scale-Short Version.

[h]Not available.

[i]Ref: reference.

XSL•FO
RenderX

Figure 2. The dose-response relationships of (A) Smartphone Addiction Scale-Short Version total score, (B) smartphone use after waking, (C) bedtime smartphone use, (D) weekly smartphone use, and (E) Pittsburgh Sleep Quality Index total score with prenatal depression. The analyses in Figures (A) to (D) were adjusted for maternal age, maternal educational attainment, paternal educational attainment, pre-pregnancy smoking status, prepregnancy alcohol consumption, frequency of physical exercise during pregnancy, frequency of social media posts on platforms, prepregnancy body mass index, weight gain during pregnancy, gravida, abortus, gestational age, and gestational diabetes mellitus. The analyses in Figure (E) were adjusted for maternal age, maternal educational attainment, paternal educational attainment, pre-pregnancy smoking status, pre-pregnancy alcohol consumption, frequency of physical exercise during pregnancy, frequency of social media posts on platforms, pre-pregnancy body mass index, weight gain during pregnancy, gravida, abortus, gestational age, gestational diabetes mellitus, and weekly smartphone use. OR: odds ratio; PSQI: Pittsburgh Sleep Quality Index; SAS-SV: Smartphone Addiction Scale-Short Version.

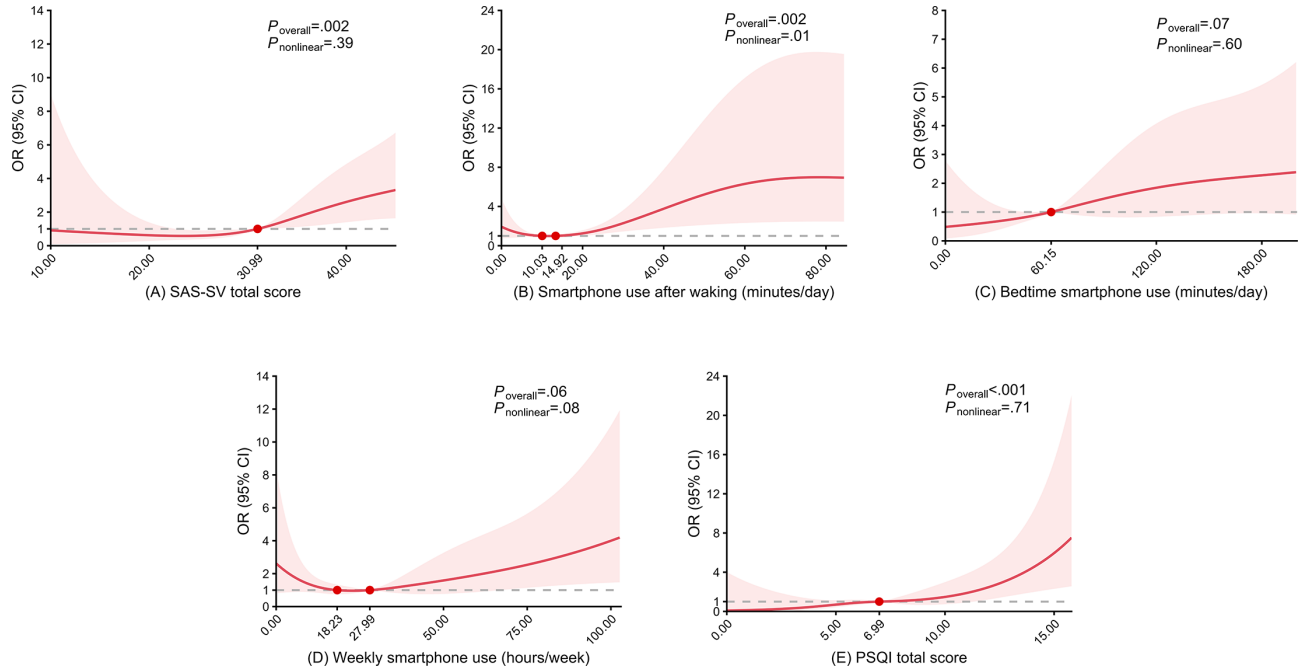
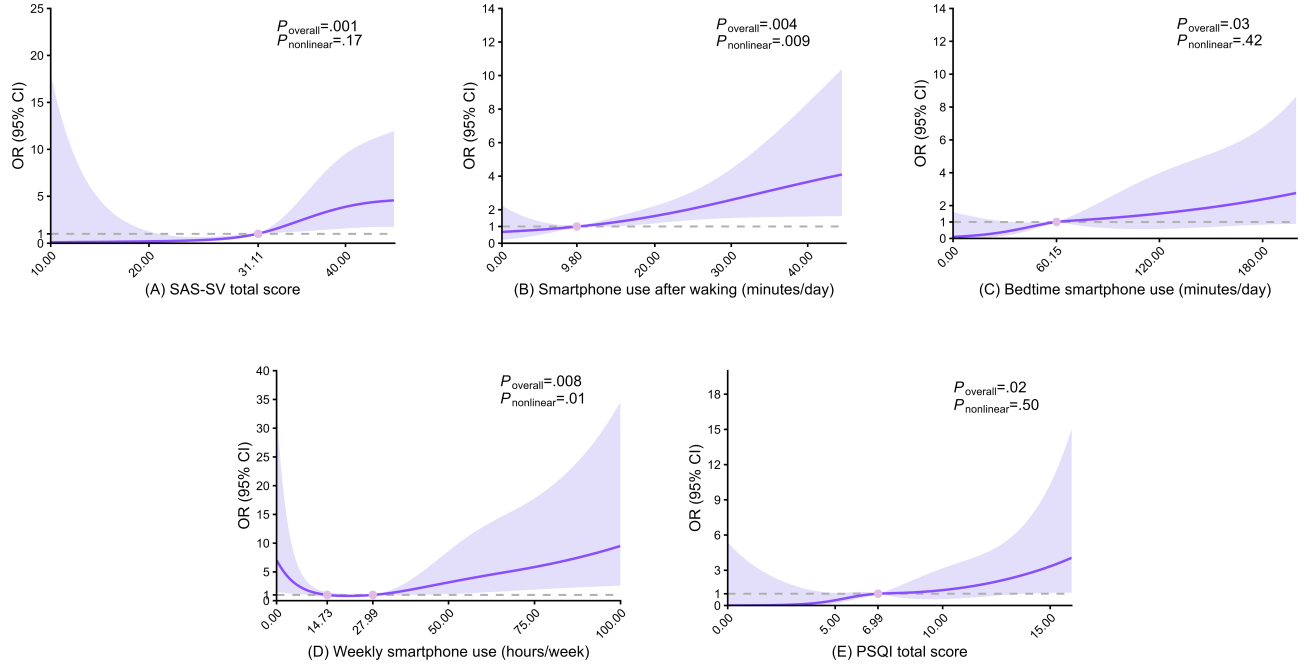


Figure 3. The dose-response relationships of (A) Smartphone Addiction Scale-Short Version total score, (B) smartphone use after waking, (C) bedtime smartphone use, (D) weekly smartphone use, and (E) Pittsburgh Sleep Quality Index total score with prenatal anxiety. The analyses in Figures (A) to (D) were adjusted for maternal age, maternal educational attainment, paternal educational attainment, pre-pregnancy smoking status, pre-pregnancy alcohol consumption, frequency of physical exercise during pregnancy, frequency of social media posts on platforms, pre-pregnancy body mass index, weight gain during pregnancy, gravida, abortus, gestational age, and gestational diabetes mellitus. The analyses in Figure (E) were adjusted for maternal age, maternal educational attainment, paternal educational attainment, pre-pregnancy smoking status, pre-pregnancy alcohol consumption, frequency of physical exercise during pregnancy, frequency of social media posts on platforms, pre-pregnancy body mass index, weight gain during pregnancy, gravida, abortus, gestational age, gestational diabetes mellitus, and weekly smartphone use. OR: odds ratio; PSQI: Pittsburgh Sleep Quality Index; SAS-SV: Smartphone Addiction Scale-Short Version.



XSL•FO
RenderX

Furthermore, the independent association between sleep quality and prenatal affective disorders was evaluated, with the results shown in Table 3. In the fully adjusted model, when the PSQI score was analyzed as a continuous variable, each 1-point increase in the score was associated with significantly higher odds of prenatal depression (OR 1.243, 95% CI 1.130 - 1.373) and anxiety (OR 1.220, 95% CI 1.087 - 1.374). The analysis using a categorical variable yielded consistent results: pregnant women with sleep disturbance (PSQI ≥7) had 2.732 times the odds of depression (OR 2.732, 95% CI 1.389 - 5.592) and 2.702 times the odds of anxiety (OR 2.702, 95% CI 1.174 - 6.657) compared with those with normal sleep. RCS analyses revealed that the PSQI score was positively associated with both prenatal depression and anxiety, although the nonlinear trends did not reach statistical significance (Figure 2E and Figure 3E).

Table . Association between sleep quality and prenatal affective disorders[a].

Exposure	Prenatal depression (PHQ-9[b] score≥10)		Prenatal anxiety (GAD-7[c] score≥10)	
	Model 1[d], OR[e] (95% CI)	Model 2[f], OR (95% CI)	Model 1[d], OR (95% CI)	Model 2[f], OR (95% CI)
PSQI[g] total score	1.247 (1.142 - 1.367)	1.243 (1.130 - 1.373)	1.237 (1.115 - 1.377)	1.220 (1.087 - 1.374)
Poor sleep quality				
No	Ref	Ref	Ref	Ref
Yes	2.706 (1.453 - 5.259)	2.732 (1.389 - 5.592)	3.029 (1.410 - 7.055)	2.702 (1.174 - 6.657)

[a]For the continuous exposure variable (PSQI total score), the effect size reflects the change in the odds associated with a 1-point increase in the PSQI total score. A total PSQI score of 7 or higher is defined as indicating poor sleep quality.

[b]PHQ-9: Patient Health Questionnaire-9.

[c]GAD-7: Generalized Anxiety Disorder-7.

[d]Model 1: crude model.

[e]OR: odds ratio.

[f]Model 2: adjustment for maternal age, maternal educational attainment, paternal educational attainment, prepregnancy smoking status, prepregnancy alcohol consumption, frequency of physical exercise during pregnancy, frequency of social media posts on platforms, prepregnancy BMI, weight gain during pregnancy, gravida, abortus, gestational age, gestational diabetes mellitus, and weekly smartphone use.

[g]PSQI: Pittsburgh Sleep Quality Index.

Mediating Effect Analysis

In this study, sleep quality was examined as a potential mediator in the associations between multidimensional smartphone use and prenatal affective disorders (Figures 4 and 5). Exploratory mediation analyses showed that the indirect pathways via sleep quality were statistically significant for the associations of bedtime smartphone use, weekly smartphone use, and smartphone addiction with prenatal depression, with mediation proportions of 44.89% (0.0065/0.0145; *P*<.001; Figure 4C), 31.30% (0.0056/0.0183; *P*=.04; Figure 4D), and 31.48% (0.0415/0.1315; *P*<.001; Figure 4A), respectively. Similar indirect patterns were observed in the anxiety models, with corresponding mediation proportions of 35.83% (0.0044/0.0123; *P*<.001; Figure 5C), 27.05% (0.0039/0.0144; *P*=.04; Figure 5D), and 22.73% (0.0271/0.1187; *P*<.001; Figure 5 B1). However, the indirect pathways from smartphone use after waking to depression and anxiety through sleep quality did not reach statistical significance (*P*=.12; Figure 4B and Figure 5B). Sensitivity mediation analyses using binary outcomes (PHQ-9 or GAD-7≥10) showed that the direction of indirect effects across pathways remained consistent with the main analysis, supporting the robustness of our mediation findings (Figures S1 and S2 in Multimedia Appendix 1).

XSL•FO
RenderX

Figure 4. The mediation effect of sleep quality on the association between multidimensional smartphone use (Smartphone Addiction Scale-Short Version total score, smartphone use after waking, bedtime smartphone use, and weekly smartphone use) and prenatal depression using continuous outcomes (Patient Health Questionnaire-9 scores). Models were adjusted for maternal age, maternal educational attainment, paternal educational attainment, prepregnancy smoking status, pre-pregnancy alcohol consumption, frequency of physical exercise during pregnancy, frequency of social media posts on platforms, pre-pregnancy body mass index, weight gain during pregnancy, gravida, abortus, gestational age, and gestational diabetes mellitus. *: *P* value <.05. ACME: average causal mediation effect; ADE: average direct effect; SAS-SV: Smartphone Addiction Scale-Short Version.

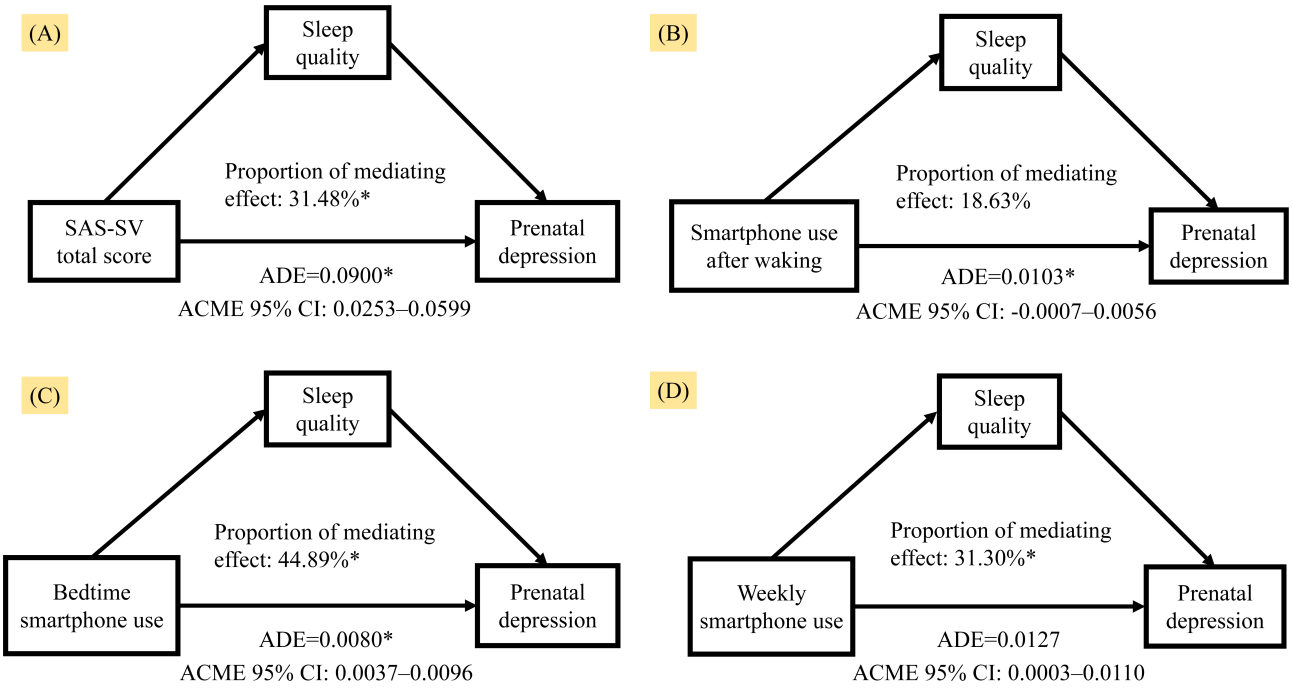
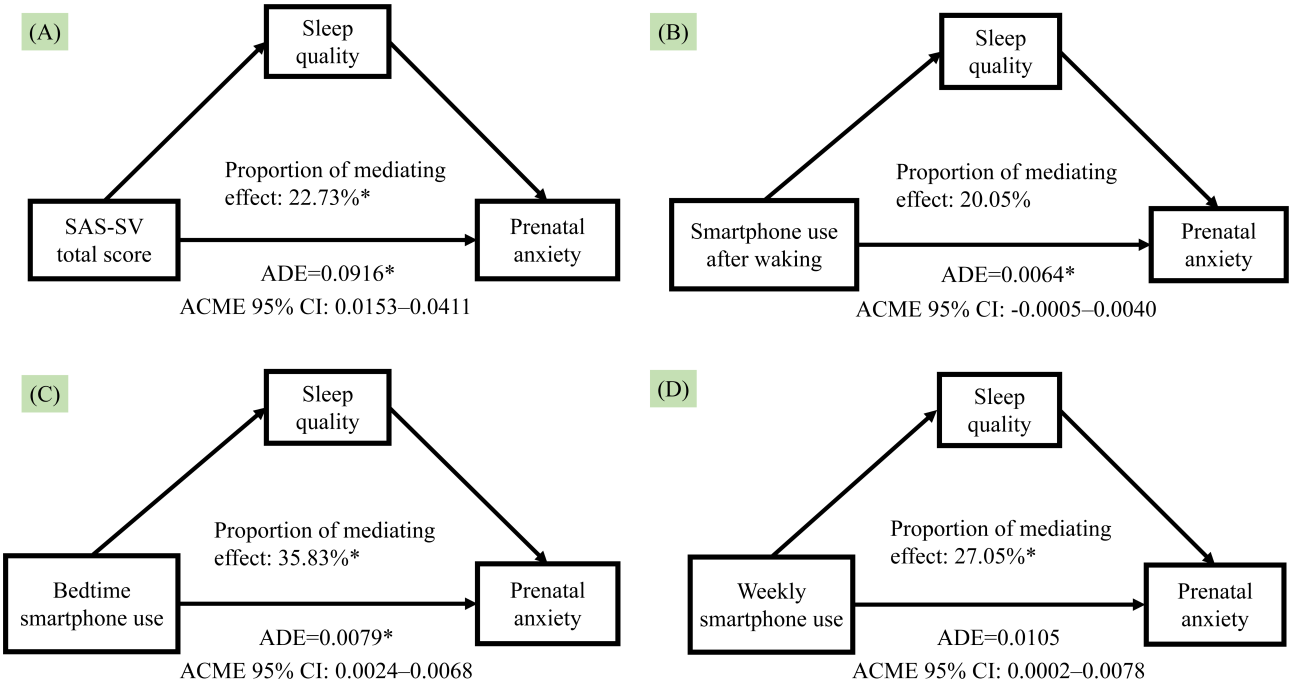


Figure 5. The mediation effect of sleep quality on the association between multidimensional smartphone use (Smartphone Addiction Scale-Short Version total score, smartphone use after waking, bedtime smartphone use, and weekly smartphone use) and prenatal anxiety using continuous outcomes (Generalized Anxiety Disorder-7 scores). Models were adjusted for maternal age, maternal educational attainment, paternal educational attainment, prepregnancy smoking status, prepregnancy alcohol consumption, frequency of physical exercise during pregnancy, frequency of social media posts on platforms, prepregnancy body mass index, weight gain during pregnancy, gravida, abortus, gestational age, and gestational diabetes mellitus. *: *P* value <.05. ACME: average causal mediation effect; ADE: average direct effect; SAS-SV: Smartphone Addiction Scale-Short Version.



XSL•FO
RenderX

Discussion

Principal Findings

This study systematically explored the associations of multidimensional smartphone use characteristics with prenatal depression and anxiety, and quantitatively validated the mediating role of sleep quality in these associations. The results showed that smartphone addiction and excessive use after waking and before bedtime were significantly associated with higher odds of prenatal affective disorders, and these associations remained robust after adjustment for multidimensional covariates. RCS analyses further revealed distinct dose–response relationships across different exposure dimensions: smartphone use after waking showed U-shaped and J-shaped associations with prenatal depression and anxiety, respectively; bedtime smartphone use and SAS-SV score showed monotonic positive associations. In addition, sleep quality partially mediated the associations of bedtime use, total weekly use, and smartphone addiction with prenatal depression and anxiety, with the mediation proportions varying across models.

The observed positive associations between smartphone addiction and prenatal depression and anxiety are consistent with findings from adolescents and college students [60-62]. In a meta-analysis encompassing 17 studies and 24,019 college students, Tang et al [63] reported that the pooled OR for the association between problematic smartphone use and depressive symptoms was 2.40 (95% CI 2.19 - 2.63), which is lower than the estimate found in this study among pregnant women. This discrepancy suggests that pregnant women may be more susceptible to the adverse mental health effects of smartphone addiction. The potential mechanisms may involve 2 levels. From a neuroendocrine perspective, the dramatic fluctuations in pregnancy-related hormones such as estradiol, progesterone, and cortisol have already rendered the hypothalamic-pituitary-adrenal axis and the emotional regulation circuits of the limbic system in a fragile state of equilibrium; the withdrawal symptoms accompanying smartphone addiction may constitute an additional stress load that further disrupts an emotional homeostasis already operating at a compensatory margin [64,65]. From a sociobehavioral perspective, individuals with addiction tendencies tend to allocate a large amount of time to virtual social environments, thereby encroaching on the time budget that could otherwise be devoted to seeking family emotional support, offline health care visits, and self-care behaviors in the real world, consequently exacerbating the sense of social isolation that is characteristic of the prenatal period [66].

Notably, the associations of time-specific exposure demonstrated marked heterogeneity in this study. In the logistic regression analyses, bedtime smartphone use exceeding 91 minutes and smartphone use after waking exceeding 61 minutes were both significantly associated with higher odds of prenatal depression and anxiety. Previous studies have confirmed that bedtime and nocturnal smartphone use are associated with increased blue light exposure, which in turn has been linked to suppression of melatonin secretion and circadian phase delay [67-69]. However, the underlying mechanisms of digital exposure during the period after waking remain unclear. One possible explanation is that exposure to information on smartphones immediately after waking, such as work notifications and social media updates, may be associated with additional morning psychosocial stressors, which could in turn be related to fluctuations in daytime emotional states. Previous literature has suggested an association between smartphone use and alterations in neuroendocrine function, such as the cortisol awakening response [70]; however, as this study did not directly measure cortisol or other biomarkers, such pathways remain to be verified in future research. Therefore, the mechanisms linking smartphone use after waking to prenatal affective disorders should be interpreted with caution.

In this study, pregnant women with either very low or very high levels of smartphone use exhibited relatively higher odds of depression and anxiety, whereas those with a weekly use duration of 29 - 49 hours had relatively lower effect estimates. This nonlinear pattern is similar to the U-shaped trend previously observed in the association between social media use and mental health [20]. From a clinical perspective, the use duration within this range may reflect pregnant women's active needs for accessing pregnancy-related health information and maintaining social connections, rather than pathological overreliance [10]. Previous studies have suggested that daily smartphone use exceeding 4 hours is significantly associated with prenatal anxiety [12], which appears to contrast with the findings of this study. This discrepancy may indicate that the relationship between smartphone duration and mental health is moderated by multiple factors such as purpose of use, content, and individual susceptibility, rather than being a simple linear accumulation. In addition, the wide CIs for this point estimate and the inability to completely rule out residual confounding warrant a cautious interpretation of the findings.

The RCS analyses further revealed differences in nonlinear patterns across exposure dimensions. The dose-response relationship for duration of smartphone use after waking could not be adequately described by a simple monotonic linear function, with curves showing J-shaped and U-shaped patterns for anxiety and depression, respectively. The association between weekly total use duration and anxiety exhibited a U-shaped curve, whereas its nonlinear trend with depression was not statistically significant, suggesting that the between-group differences observed in Table 2 may partly arise from fluctuations introduced by the categorization strategy. Both duration of bedtime smartphone use and smartphone addiction score showed monotonically increasing associations with depression and anxiety, suggesting that higher exposure levels or greater addiction severity may be associated with higher odds of these outcomes. Thus, the adverse mental health correlates of smartphone use are not solely dependent on cumulative screen time but exhibit notable time-specific differences. Compared with focusing only on total weekly use duration, precise identification of and intervention on screen exposure during critical circadian windows may hold greater clinical value.

Exploratory mediation analyses further indicated that sleep quality partially accounted for the associations between multidimensional smartphone use and prenatal affective

XSL•FO
RenderX

disorders, which is consistent with the mechanisms previously observed in nonpregnant populations [20,26,27]. Previous studies have demonstrated that excessive use of electronic devices such as smartphones is associated with increased blue light exposure, which in turn has been shown to suppress the normal secretory rhythm of pineal melatonin, prolong sleep onset latency, and delay circadian phase [67-69]. In addition, research has suggested that cognitive and emotional arousal triggered by immersive screen exposure before bedtime may interfere with sleep initiation and maintenance [29]. For pregnant women who already experience physiological sleep fragmentation, these effects may be further amplified. Previous research has also indicated that sleep deprivation is associated with impaired prefrontal cortex function and increased amygdala reactivity, which may weaken prefrontal-amygdala connectivity and contribute to emotion dysregulation [34,71]; insufficient sleep may also be associated with higher odds of depression and anxiety through activation of the sympathetic nervous system and inflammatory signaling pathways [72,73]. In this study, the mediation proportions of sleep quality in the associations of bedtime use, weekly use, and smartphone addiction with prenatal depression were 44.89% (0.0065/0.0145), 31.30% (0.0056/0.0183), and 31.48% (0.0415/0.1315), respectively; for prenatal anxiety, the corresponding proportions were 35.83% (0.0044/0.0123), 27.05% (0.0039/0.0144), and 22.73% (0.0271/0.1187). These findings suggest that sleep quality may represent a relevant pathway in these associations. Notably, the indirect pathways through sleep quality for smartphone use after waking were not statistically significant for either depression or anxiety. The specificity of these patterns suggests that the associations between digital device use at different times of day and prenatal mental health may operate through distinct biobehavioral pathways.

The association between prolonged and time-specific smartphone use and higher odds of prenatal affective disorders underscores the urgency of digital health management during pregnancy as an important public health issue. In this study, RCS analyses revealed a nonlinear association between smartphone use after waking and prenatal affective disorders, and identified a U-shaped pattern in the relationship between weekly smartphone use and affective disorders. This finding is consistent with the U-shaped association between social media use and mental health observed by Singh et al [74] in adolescent populations, suggesting that moderate smartphone use may be associated with lower emotional distress, whereas excessive use may be related to poorer mental health. These findings provide preliminary evidence for the development of prenatal digital health guidelines. These findings suggest that prenatal mental health interventions should not simply advocate a blanket reduction in smartphone use, but should instead shift toward more precise risk-stratification strategies, such as prioritizing screening for smartphone addiction tendencies and targeting screen exposure behaviors during high-risk periods, specifically after waking and before bedtime. Furthermore, the mediation proportion of sleep quality suggests that improving sleep may be a key intervention target for interrupting the pathway linking digital exposure to affective disorders. Cognitive behavioral interventions targeting bedtime smartphone use behavior may achieve the dual benefits of improving sleep and stabilizing mood simultaneously.

This study has several limitations. First, the study was conducted at the Hunan Provincial Maternal and Child Health Care Hospital, and the study population was limited to permanent residents who completed their entire prenatal care at that hospital; nearly 90 percent (475/537) of them held a college degree or higher, which may have influenced smartphone usage patterns: pregnant women with higher educational attainment may be more inclined to use smartphones to access prenatal health information, but they may also face a higher cognitive load from work-related information processing. Their smartphone usage behavior may reflect a dual characteristic of both health promotion and overreliance, which in turn limits the generalizability of this study's conclusions to pregnant women with lower educational levels or those in rural areas. Second, the cross-sectional design limits our ability to infer causality; although the pathways in the mediation analysis were established based on evidence from previous longitudinal studies and biological plausibility, we cannot rule out the possibility of reverse pathways. Pregnant women may increase their smartphone use, particularly during periods of nighttime wakefulness, to pass the time or alleviate distress due to preexisting poor sleep quality or symptoms of depression and anxiety. This reverse causality may violate the temporal sequence assumption of the mediation model and could introduce bias into the estimates of indirect effects.

Furthermore, assessments of smartphone usage duration during specific time periods, smartphone addiction, and sleep quality relied on self-report questionnaires, which may introduce recall bias. In this study, total weekly usage duration was objectively recorded using the screen time feature built into the smartphone operating system; however, for usage durations after waking and before bedtime, self-reporting was used as an alternative because the existing operating system could not export data for specific time periods separately. This inconsistency in measurement methods may have some impact on the comparability of the results. Additionally, although we adjusted for several key covariates in our models, some theoretically important confounders (such as caffeine intake, dietary habits, sleep patterns, and occupation type) were not included in the analysis; their omission may lead to residual confounding, which could affect the precision of the effect estimates. Although this study included gestational diabetes mellitus as a covariate to control for the potential impact of common pregnancy complications, other unmeasured complications and lifestyle factors may still constitute residual confounding. However, given that the primary objective of this study was to explore the direction of the association and dose-response trends between smartphone usage patterns and prenatal affective disorders, the impact of these omissions on the directionality and statistical significance of the core conclusions is likely limited; nevertheless, more comprehensive covariate control would help yield more accurate effect estimates. Given the limited sample size in the anxiety group, the CIs for some estimates are relatively wide, indicating limited precision; therefore, the results should be interpreted with caution. Future large-scale, multicenter, prospective longitudinal cohort studies

XSL•FO
RenderX

are needed to integrate objective data on screen time tracking and sleep monitoring, comprehensively collect the aforementioned covariates, and explore mediating mechanisms within a longitudinal research framework to further validate the findings of this study.

Conclusion

The results of this study indicate that smartphone addiction and excessive smartphone use after waking and before bedtime were significantly associated with elevated odds of prenatal depression and anxiety, and distinct dose-response patterns exist across different exposure-outcome combinations. Additionally, exploratory mediation analyses indicated that sleep quality may statistically account for part of these associations. Future longitudinal studies are needed to establish temporal precedence and to test the directional assumptions underlying the mediation model.

Acknowledgments

The authors express their gratitude to all the participants and staff involved in this study. Their efforts and contributions are vital to the success of the research.

Funding

The study was funded through grants from Hunan Province Health Research Project (grant number 20255507, 20254300), the Natural Science Foundation of Hunan Province (2025JJ80661), the National Clinical Key Specialty Cultivation Project (Nursing Key Specialty), and the National Natural Science Foundation of China (Grant No. 42407585).

Data Availability

The datasets generated and analyzed during this study are not publicly available due to their containing sensitive information about participants' social media use and mental health, and to protect participant confidentiality as stipulated by the ethics approval. However, deidentified data underlying the findings presented in this study are available from the corresponding author (MYW) upon reasonable request. Requestors will need to submit a methodologically sound proposal and sign a data access agreement.

Authors' Contributions

Conceptualization: XXZ, MXOY, XLW, XHY, XYW, YXQC, MMX, YLS, MYW
Data curation: MXOY, XLW, MMX, YLS, MYW
Formal analysis: XXZ, MXOY, MMX, YLS, MYW
Funding acquisition: MXOY, MMX, XLW, YLS, MYW
Investigation: MXOY, XLW, MMX
Methodology: XXZ, YLS, MYW
Project administration: MXOY, XLW, MMX, YLS, MYW
Resources: MXOY, XLW, MMX, YLS, MYW
Software: XXZ, MXOY, MMX, YLS, MYW
Supervision: MXOY, MMX, YLS, MYW
Validation: XXZ, MXOY, XLW, XHY, XYW, YXQC, MMX, YLS, MYW
Visualization: XXZ, MYW
Writing–original draft: XXZ, MYW
Writing–review & editing: XXZ, MXOY, MMX, YLS, MYW
Authors XZ and MO are co-first authors and contributed equally to this manuscript.
MX, YS, and WY are co-corresponding authors of this manuscript and can be reached at:
MX: 1183559609@qq.com; YS: shushu_1012@163.com

The authors have no conflicts of interest to declare.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Results of sensitivity mediation analyses.
[DOCX File, 435 KB - mhealth_v14i1e100064_app1.docx]

References

1. Howard LM, Khalifeh H. Perinatal mental health: a review of progress and challenges. World Psychiatry 2020 Oct;19(3):313-327. [doi: 10.1002/wps.20769] [Medline: 32931106]

XSL•FO
RenderX

2. Dagher RK, Bruckheim HE, Colpe LJ, Edwards E, White DB. Perinatal depression: challenges and opportunities. J Womens Health (Larchmt) 2021 Feb;30(2):154-159. [doi: 10.1089/jwh.2020.8862] [Medline: 33156730]
3. Lim G. Perinatal depression. Curr Opin Anaesthesiol 2021 Jun 1;34(3):233-237. [doi: 10.1097/ACO.0000000000000998] [Medline: 33935170]
4. McCarthy M, Houghton C, Matvienko-Sikar K. Women's experiences and perceptions of anxiety and stress during the perinatal period: a systematic review and qualitative evidence synthesis. BMC Pregnancy Childbirth 2021 Dec 6;21(1):811. [doi: 10.1186/s12884-021-04271-w] [Medline: 34865642]
5. Matvienko-Sikar K, Dijk WV, Dockray S, Leahy-Warren P. Modifiable and vulnerability factors for maternal stress and anxiety in the first 1000 days: an umbrella review and framework. Women Birth 2025 Jul;38(4):101941. [doi: 10.1016/j.wombi.2025.101941] [Medline: 40544557]
6. Al-Abri K, Edge D, Armitage CJ. Prevalence and correlates of perinatal depression. Soc Psychiatry Psychiatr Epidemiol 2023 Nov;58(11):1581-1590. [doi: 10.1007/s00127-022-02386-9] [Medline: 36646936]
7. Yang L, Sun J, Nan Y, et al. Prevalence of perinatal anxiety and its determinants in mainland China: a systematic review and meta-analysis. J Affect Disord 2023 Feb 15;323:193-203. [doi: 10.1016/j.jad.2022.11.075] [Medline: 36442655]
8. Sakowicz A, Allen E, Alvarado-Goldberg M, Grobman WA, Miller ES. Association between antenatal depression symptom trajectories and preterm birth. Obstet Gynecol 2023 Apr 1;141(4):810-817. [doi: 10.1097/AOG.0000000000005125] [Medline: 36897146]
9. Gelaye B, Rondon MB, Araya R, Williams MA. Epidemiology of maternal depression, risk factors, and child outcomes in low-income and middle-income countries. Lancet Psychiatry 2016 Oct;3(10):973-982. [doi: 10.1016/S2215-0366(16)30284-X] [Medline: 27650773]
10. Aftab N. Social media use during pregnancy: risks and benefits for women's health. Cureus 2025 Aug;17(8):e90410. [doi: 10.7759/cureus.90410] [Medline: 40970070]
11. Costantino C, Mazzucco W, Bonaccorso N, et al. A cross-sectional study on smartphone uses among pregnant women attending childbirth classes in the metropolitan area of Palermo, Italy: the stop-phone study. Ann Ig 2023;35(3):319-330. [doi: 10.7416/ai.2022.2543] [Medline: 36190310]
12. Guo T, Yu Y, Wang X, Sun Y, Yang C. Correlates of antenatal anxiety: smartphone use, depressive symptoms, and hypertensive disorders in a cross-sectional study in Southwest China. Front Med (Lausanne) 2025;12:1682499. [doi: 10.3389/fmed.2025.1682499] [Medline: 41158470]
13. Loleska S, Pop-Jordanova N. Is smartphone addiction in the younger population a public health problem? Pril (Makedon Akad Nauk Umet Odd Med Nauki) 2021 Dec 30;42(3):29-36. [doi: 10.2478/prilozi-2021-0032] [Medline: 35032372]
14. Lin CY, Ratan ZA, Pakpour AH. Collection of smartphone and internet addiction. BMC Psychiatry 2023 Jun 14;23(1):427. [doi: 10.1186/s12888-023-04915-5] [Medline: 37316810]
15. Ye XL, Zhang W, Zhao FF. Depression and internet addiction among adolescents: a meta-analysis. Psychiatry Res 2023 Aug;326:115311. [doi: 10.1016/j.psychres.2023.115311] [Medline: 37348449]
16. Lopes LS, Valentini JP, Monteiro TH, et al. Problematic social media use and its relationship with depression or anxiety: a systematic review. Cyberpsychol Behav Soc Netw 2022 Nov;25(11):691-702. [doi: 10.1089/cyber.2021.0300] [Medline: 36219756]
17. Nikolic A, Bukurov B, Kocic I, et al. Smartphone addiction, sleep quality, depression, anxiety, and stress among medical students. Front Public Health 2023;11:1252371. [doi: 10.3389/fpubh.2023.1252371] [Medline: 37744504]
18. Kang Y, Liu S, Yang L, et al. Testing the bidirectional associations of mobile phone addiction behaviors with mental distress, sleep disturbances, and sleep patterns: a one-year prospective study among Chinese college students. Front Psychiatry 2020;11:634. [doi: 10.3389/fpsyt.2020.00634] [Medline: 32765310]
19. Wang Y, Luo Y, Yan N, et al. The longitudinal relationships between problematic smartphone use and anxiety symptoms among Chinese college students: a cross-lagged panel network analysis. Addict Behav 2025 Jan;160:108170. [doi: 10.1016/j.addbeh.2024.108170] [Medline: 39305708]
20. Zou Q, Qin J, Zhang L, et al. Time spent on social media applications in relation to depressive symptoms during emerging adulthood and the mediating role of sleep quality: cross-sectional observational study. J Med Internet Res 2025 Dec 19;27:e75337. [doi: 10.2196/75337] [Medline: 41418287]
21. Xiao Y, Meng Y, Brown TT, Keyes KM, Mann JJ. Addictive screen use trajectories and suicidal behaviors, suicidal ideation, and mental health in US youths. JAMA 2025 Jul 15;334(3):219-228. [doi: 10.1001/jama.2025.7829] [Medline: 40531519]
22. Geoffroy PA, Maruani J. Chronobiology of mood disorders: the role of the biological clock in depression and bipolar disorder. Biol Psychiatry 2025 Dec 1;98(11):830-841. [doi: 10.1016/j.biopsych.2025.05.005] [Medline: 40381827]
23. Wang XL, Serchov T. Rhythms of mood: bidirectional regulation of molecular clocks and neurotransmission in affective disorders. Neurosci Biobehav Rev 2026 Jul;186:106700. [doi: 10.1016/j.neubiorev.2026.106700] [Medline: 42019704]
24. Zhong C, Masters M, Donzella SM, Diver WR, Patel AV. Electronic screen use and sleep duration and timing in adults. JAMA Netw Open 2025 Mar 3;8(3):e252493. [doi: 10.1001/jamanetworkopen.2025.2493] [Medline: 40146105]
25. Höhn C, Schmid SR, Plamberger CP, et al. Preliminary results: the impact of smartphone use and short-wavelength light during the evening on circadian rhythm, sleep and alertness. Clocks Sleep 2021 Jan 22;3(1):66-86. [doi: 10.3390/clockssleep3010005] [Medline: 33499010]

XSL•FO
RenderX

26. Lima Santos JP, Soehner AM, Biernesser CL, Ladouceur CD, Versace A. Role of sleep and white matter in the link between screen time and depression in childhood and early adolescence. JAMA Pediatr 2025 Sep 1;179(9):1000-1008. [doi: 10.1001/jamapediatrics.2025.1718] [Medline: 40549406]
27. Feng Z, Diao Y, Ma H, et al. Mobile phone addiction and depression among Chinese medical students: the mediating role of sleep quality and the moderating role of peer relationships. BMC Psychiatry 2022 Aug 23;22(1):567. [doi: 10.1186/s12888-022-04183-9] [Medline: 35999533]
28. Dresp-Langley B, Hutt A. Digital addiction and sleep. Int J Environ Res Public Health 2022 Jun 5;19(11):6910. [doi: 10.3390/ijerph19116910] [Medline: 35682491]
29. León Méndez M, Padrón I, Fumero A, Marrero RJ. Effects of internet and smartphone addiction on cognitive control in adolescents and young adults: a systematic review of fMRI studies. Neurosci Biobehav Rev 2024 Apr;159:105572. [doi: 10.1016/j.neubiorev.2024.105572] [Medline: 38320657]
30. Mislu E, Kumsa H, Tadesse S, et al. Sleep quality disparities in different pregnancy trimesters in low- and middle-income countries: a systematic review and meta-analysis. BMC Pregnancy Childbirth 2024 Oct 1;24(1):627. [doi: 10.1186/s12884-024-06830-3] [Medline: 39354392]
31. Gupta R, Rawat VS. Sleep and sleep disorders in pregnancy. Handb Clin Neurol 2020;172:169-186. [doi: 10.1016/B978-0-444-64240-0.00010-6] [Medline: 32768087]
32. Elahi M, Sanin KI, Rahman MS, et al. Association between mental health and sleep quality among forcibly displaced Rohingya pregnant women living in Cox's Bazar, Bangladesh: a cross-sectional observation. BMC Psychiatry 2026 Jan 16;26(1):103. [doi: 10.1186/s12888-025-07561-1] [Medline: 41546038]
33. Pauley AM, Moore GA, Mama SK, Molenaar P, Symons Downs D. Associations between prenatal sleep and psychological health: a systematic review. J Clin Sleep Med 2020 Apr 15;16(4):619-630. [doi: 10.5664/jcsm.8248] [Medline: 32003734]
34. Hyndych A, El-Abassi R, Mader ECJ. The role of sleep and the effects of sleep loss on cognitive, affective, and behavioral processes. Cureus 2025 May;17(5):e84232. [doi: 10.7759/cureus.84232] [Medline: 40525051]
35. Sullivan EC, McCall C, Henderson LM, Cairney SA. Disentangling sleep's role in emotion processing. Neuropsychologia 2025 Dec 15;219:109269. [doi: 10.1016/j.neuropsychologia.2025.109269] [Medline: 40907885]
36. Kwon M, Kim DJ, Cho H, Yang S. The smartphone addiction scale: development and validation of a short version for adolescents. PLoS ONE 2013;8(12):e83558. [doi: 10.1371/journal.pone.0083558] [Medline: 24391787]
37. Luk TT, Wang MP, Shen C, et al. Short version of the Smartphone Addiction Scale in Chinese adults: psychometric properties, sociodemographic, and health behavioral correlates. J Behav Addict 2018 Dec 1;7(4):1157-1165. [doi: 10.1556/2006.7.2018.105] [Medline: 30418073]
38. Lee JJ, Wang MP, Luk TT, Guo N, Chan SSC, Lam TH. Associations of electronic device use before and after sleep with psychological distress among Chinese adults in Hong Kong: cross-sectional study. JMIR Ment Health 2020 Jun 11;7(6):e15403. [doi: 10.2196/15403] [Medline: 32525489]
39. Woo KS, Bong SH, Choi TY, Kim JW. Mental health, smartphone use type, and screen time among adolescents in South Korea. Psychol Res Behav Manag 2021;14:1419-1428. [doi: 10.2147/PRBM.S324235] [Medline: 34548824]
40. Twenge JM, Campbell WK. Associations between screen time and lower psychological well-being among children and adolescents: evidence from a population-based study. Prev Med Rep 2018 Dec;12:271-283. [doi: 10.1016/j.pmedr.2018.10.003] [Medline: 30406005]
41. Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
42. Li C, Cai HB, Zhou Q, Zhang HQ, Wang M, Kang HC. Sleep disorders in the acute phase of coronavirus disease 2019: an overview and risk factor study. Ann Gen Psychiatry 2023 Jan 31;22(1):3. [doi: 10.1186/s12991-023-00431-8] [Medline: 36721207]
43. Huang Y, Jiang ZJ, Deng J, Qi YJ. Sleep quality of patients with postoperative glioma at home. World J Clin Cases 2020 Oct 26;8(20):4735-4742. [doi: 10.12998/wjcc.v8.i20.4735] [Medline: 33195641]
44. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
45. Cheng J, Yu Y, Peng Y, et al. Psychometric validation of the Patient Health Questionnaire-9 for pregnant women: a comprehensive longitudinal study in China. J Affect Disord 2026 Aug 1;406:121712. [doi: 10.1016/j.jad.2026.121712] [Medline: 41935747]
46. Wu D, Chen S, Zhong X, Zhang J, Zhao G, Jiang L. Prevalence and factors associated with antenatal depressive symptoms across trimesters: a study of 110,584 pregnant women covered by a mobile app-based screening programme in Shenzhen, China. BMC Pregnancy Childbirth 2024 Jul 16;24(1):480. [doi: 10.1186/s12884-024-06680-z] [Medline: 39014317]
47. Cheng J, Ren Z, Rehman S, et al. Comparative assessment of Patient Health Questionnaire-9 and Edinburgh Postnatal Depression Scale for screening antenatal depression among Chinese pregnant women. J Affect Disord 2025 May 1;376:104-112. [doi: 10.1016/j.jad.2025.02.004] [Medline: 39909161]
48. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

XSL•FO
RenderX

49. Gong Y, Zhou H, Zhang Y, et al. Validation of the 7-item Generalized Anxiety Disorder scale (GAD-7) as a screening tool for anxiety among pregnant Chinese women. J Affect Disord 2021 Mar 1;282:98-103. [doi: 10.1016/j.jad.2020.12.129] [Medline: 33401129]
50. Liu H, Huang F, Gao Y, et al. Network analysis of depression and anxiety symptoms and their associations with cognitive fusion among pregnant women. BMC Psychiatry 2025 May 26;25(1):537. [doi: 10.1186/s12888-025-06978-y] [Medline: 40420034]
51. Mu L, Yu H, Qian G. Associations between physical activity and prenatal depression and anxiety symptoms: a cross-sectional study. Front Public Health 2025;13:1666312. [doi: 10.3389/fpubh.2025.1666312] [Medline: 41426672]
52. Wang R, Cong S, Sha L, et al. Association between social networking site use intensity and depression among Chinese pregnant women: cross-sectional study. J Med Internet Res 2023 Mar 15;25:e41793. [doi: 10.2196/41793] [Medline: 36920458]
53. Yang K, Wu J, Chen X. Risk factors of perinatal depression in women: a systematic review and meta-analysis. BMC Psychiatry 2022 Jan 27;22(1):63. [doi: 10.1186/s12888-021-03684-3] [Medline: 35086502]
54. Herbert D, Young K, Pietrusińska M, MacBeth A. The mental health impact of perinatal loss: a systematic review and meta-analysis. J Affect Disord 2022 Jan 15;297:118-129. [doi: 10.1016/j.jad.2021.10.026] [Medline: 34678403]
55. Toscano M, Royzer R, Castillo D, Li D, Poleshuck E. Prevalence of depression or anxiety during antepartum hospitalizations for obstetric complications: a systematic review and meta-analysis. Obstet Gynecol 2021 May 1;137(5):881-891. [doi: 10.1097/AOG.0000000000004335] [Medline: 33831928]
56. Kang D, Zhao D, Ryu S, et al. Perceived stress and non-alcoholic fatty liver disease in apparently healthy men and women. Sci Rep 2020 Jan 8;10(1):38. [doi: 10.1038/s41598-019-57036-z] [Medline: 31913326]
57. Okosieme OE, Usman D, Taylor PN, et al. Cardiovascular morbidity and mortality in patients in Wales, UK with resistance to thyroid hormone β (RTHβ): a linked-record cohort study. Lancet Diabetes Endocrinol 2023 Sep;11(9):657-666. [doi: 10.1016/S2213-8587(23)00155-9] [Medline: 37475119]
58. Durrleman S, Simon R. Flexible regression models with cubic splines. Stat Med 1989 May;8(5):551-561. [doi: 10.1002/sim.4780080504] [Medline: 2657958]
59. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Stat Med 2002 Aug 30;21(16):2409-2419. [doi: 10.1002/sim.1047] [Medline: 12210625]
60. Li Y, Li G, Liu L, Wu H. Correlations between mobile phone addiction and anxiety, depression, impulsivity, and poor sleep quality among college students: a systematic review and meta-analysis. J Behav Addict 2020 Oct 12;9(3):551-571. [doi: 10.1556/2006.2020.00057] [Medline: 32903205]
61. Matar Boumosleh J, Jaalouk D. Depression, anxiety, and smartphone addiction in university students- a cross sectional study. PLoS One 2017;12(8):e0182239. [doi: 10.1371/journal.pone.0182239] [Medline: 28777828]
62. Shiferaw BD, Tang J, Wang Y, et al. Impact of digital addiction on youth health: a systematic review and meta-analysis. J Behav Addict 2025 Sep 30;14(3):1129-1158. [doi: 10.1556/2006.2025.00081] [Medline: 40928886]
63. Tang S, Chen R, Ma Z, Li X, Chen J, Zhao J. Associations of problematic smartphone use with depressive symptoms and suicidal ideation in university students before and after the COVID-19 outbreak: a meta-analysis. Addict Behav 2024 May;152:107969. [doi: 10.1016/j.addbeh.2024.107969] [Medline: 38290322]
64. Duthie L, Reynolds RM. Changes in the maternal hypothalamic-pituitary-adrenal axis in pregnancy and postpartum: influences on maternal and fetal outcomes. Neuroendocrinology 2013;98(2):106-115. [doi: 10.1159/000354702] [Medline: 23969897]
65. Zhang C, Xu D, Luo H, et al. Prenatal xenobiotic exposure and intrauterine hypothalamus-pituitary-adrenal axis programming alteration. Toxicology 2014 Nov 5;325:74-84. [doi: 10.1016/j.tox.2014.08.015] [Medline: 25194749]
66. Guo N, Wang MP, Luk TT, et al. The association of problematic smartphone use with family well-being mediated by family communication in Chinese adults: a population-based study. J Behav Addict 2019 Sep 1;8(3):412-419. [doi: 10.1556/2006.8.2019.39] [Medline: 31394912]
67. Dufier JL, Touitou Y. Light pollution. Rev Prat 2022 Feb;72(2):141-146. [Medline: 35289519]
68. Minich DM, Henning M, Darley C, Fahoum M, Schuler CB, Frame J. Is melatonin the “next vitamin D”?: A review of emerging science, clinical uses, safety, and dietary supplements. Nutrients 2022 Sep 22;14(19):3934. [doi: 10.3390/nu14193934] [Medline: 36235587]
69. Oh JH, Yoo H, Park HK, Do YR. Analysis of circadian properties and healthy levels of blue light from smartphones at night. Sci Rep 2015 Jun 18;5:11325. [doi: 10.1038/srep11325] [Medline: 26085126]
70. Afifi TD, Zamanzadeh N, Harrison K, Acevedo Callejas M. WIRED: The impact of media and technology use on stress (cortisol) and inflammation (interleukin IL-6) in fast paced families. Comput Human Behav 2018 Apr;81:265-273. [doi: 10.1016/j.chb.2017.12.010]
71. Ben Simon E, Vallat R, Barnes CM, Walker MP. Sleep loss and the socio-emotional brain. Trends Cogn Sci 2020 Jun;24(6):435-450. [doi: 10.1016/j.tics.2020.02.003] [Medline: 32299657]
72. Ballesio A. Inflammatory hypotheses of sleep disturbance - depression link: update and research agenda. Brain Behav Immun Health 2023 Aug;31:100647. [doi: 10.1016/j.bbih.2023.100647] [Medline: 37408788]

XSL•FO
RenderX

73. Meerlo P, Sgoifo A, Suchecki D. Restricted and disrupted sleep: effects on autonomic function, neuroendocrine stress systems and stress responsivity. Sleep Med Rev 2008 Jun;12(3):197-210. [doi: 10.1016/j.smrv.2007.07.007] [Medline: 18222099]
74. Singh B, Zhou M, Curtis R, Maher C, Dumuid D. Social media use and well-being across adolescent development. JAMA Pediatr 2026 Mar 1;180(3):288-297. [doi: 10.1001/jamapediatrics.2025.5619] [Medline: 41525086]

Abbreviations

GAD-7: Generalized Anxiety Disorder-7
OR: odds ratio
PHQ-9: Patient Health Questionnaire-9
PSQI: Pittsburgh Sleep Quality Index
RCS: restricted cubic spline
SAS-SV: Smartphone Addiction Scale-Short Version

Edited by A Stone; submitted 02.May.2026; peer-reviewed by K Pan, L Wang, P Dou; revised version received 17.Sep.2026; accepted 17.Sep.2026; published 30.Oct.2026.

Please cite as:
Zhang X, Ouyang M, Wu X, Yang X, Wang X, Cao Y, Xu M, Shu Y, Wu M
The Mediating Role of Sleep Quality in Associations of Smartphone Addiction and Time-Specific Use With Prenatal Depression and Anxiety: Cross-Sectional Observational Study
JMIR Mhealth Uhealth 2026;14:e100064
URL: https://mhealth.jmir.org/2026/1/e100064
doi:10.2196/100064

© Xiangxin Zhang, Mixia Ouyang, Xiaoli Wu, Xinghua Yang, Xinyi Wang, Yuxinqiao Cao, Mingmin Xu, Yanling Shu, Mingyang Wu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 30.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Multidomain Outcomes of a Digital Self-Management Program in Rural Adults With Chronic Conditions: Single-Arm Prospective Interventional Study

Nathan K Smith[1], PhD; Gabriela Ilie[1,2,3], PhD; Cody MacDonald[1], MSc; James Richard Ballentyne Langley[4], BHSc; Robert David Harold Rutledge[3], MD

[1]Department of Community Health and Epidemiology, Faculty of Medicine, Dalhousie University, 5790 University Avenue, Halifax, NS, Canada

[2]Department of Urology, Faculty of Medicine, Dalhousie University, Halifax, NS, Canada

[3]Department of Radiation Oncology, Faculty of Medicine, Dalhousie University, Halifax, NS, Canada

[4]Health Sciences Department, Queen's University, Kingston, ON, Canada

Corresponding Author:
Gabriela Ilie, PhD
Department of Community Health and Epidemiology, Faculty of Medicine, Dalhousie University, 5790 University Avenue, Halifax, NS, Canada

Abstract

Background: Chronic disease and comorbid mental health problems disproportionately affect rural and underserved populations. Digital self-management programs are potentially scalable, but evaluations often focus on narrow outcomes.

Objective: This study examined 12-month longitudinal changes in psychological and multidomain outcomes among Pictou County Patient Empowerment Program participants and explored co-occurring changes and psychological moderators.

Methods: This single-arm prospective interventional study enrolled 182 community-dwelling adults with ≥1 chronic condition in Pictou County, Nova Scotia, between October 2023 and August 2025 through community-based outreach. Median age was 60 (IQR 51 - 68) years, 155 of 182 (85.2%) participants were women, and the median number of chronic conditions was 3 (IQR 1 - 3). The initial 6-month digital intervention delivered brief daily modules targeting physical activity, nutrition, stress regulation, sleep, habit formation, and social connection, followed by a 6-month lower-intensity maintenance period during which monthly group videoconference sessions continued to be offered. Outcomes were assessed at baseline, 6 months, and 12 months. Primary outcomes were illness perceptions (Brief Illness Perception Questionnaire), perceived stress (10-item Perceived Stress Scale [PSS-10]), disease-related worry (10-item Questionnaire on Distress in Cancer Patients), life satisfaction (Satisfaction With Life Scale), and mental well-being (World Health Organization-Five Well-Being Index [WHO-5]). Secondary outcomes assessed the physical, behavioral, and psychosocial domains. Generalized estimating equations estimated population-averaged longitudinal changes, adjusted for age, sex, relationship status, comorbidity burden, and baseline depression or anxiety medication use. Exploratory models examined co-occurring changes and moderation. Tests were 2-sided with α=.05.

Results: Across outcomes and follow-up assessments, completion ranged from 58.2% (95% CI 51.0%-65.2%; 106/182) to 65.4% (95% CI 58.2% - 71.9%; 119/182). At 12 months, illness perceptions decreased (β=−7.3, 95% CI −9.7 to −4.9), perceived stress decreased (β=−2.4, 95% CI −3.5 to −1.3), disease-related worry decreased (β=−6.3, 95% CI −7.8 to −4.7), life satisfaction increased (β=2.5, 95% CI 1.4 - 3.5), and mental well-being increased (β=10.4, 95% CI 7.2 - 13.5; all P<.001). Secondary outcomes improved across fatigue, sleep, diet quality, sedentary behavior, spiritual well-being, mental health–related quality of life, weight, and BMI. Exploratory analyses showed that improvements in fatigue, sleep quality, and illness perceptions co-occurred with greater well-being and life satisfaction and lower perceived stress and disease-related worry at both follow-up assessment points. Clinically meaningful stress improvement (≥4-point PSS-10 reduction) was achieved by 36.4% (95% CI 27.3%-45.6%, 39/107) and 43.4% (95% CI 34.0%-52.8%, 46/106) participants at 6 and 12 months, respectively. At 12 months, 43.5% (95% CI 34.2%-52.9%, 47/108) participants reported clinically meaningful improvement in mental well-being (≥10-point WHO-5 increase).

Conclusions: Participation in this low-bandwidth, home-based digital self-management program was associated with sustained psychological, physical, and behavioral improvements among rural adults with chronic conditions. These multidomain findings support randomized controlled evaluation of scalable digital self-management approaches as adjuncts to rural chronic disease care.

Trial Registration: ClinicalTrials.gov NCT06064110; https://clinicaltrials.gov/study/NCT06064110

(JMIR Mhealth Uhealth 2026;14:e95283) doi:10.2196/95283

XSL•FO
RenderX

KEYWORDS

rural health services; mental health; patient empowerment; home-based interventions; chronic disease; psychological distress

Introduction

Chronic disease represents a growing public health challenge. In 2021, approximately 45% of Canadians were living with at least 1 major chronic condition, and 1 in 12 had 3 or more comorbid chronic diseases [1]. Globally, chronic diseases account for an estimated 75% of non–pandemic-related deaths and are projected to generate cumulative health care costs exceeding US $47 trillion by 2030 [2,3]. Chronic disease frequently co-occurs with adverse mental health outcomes, including depression, anxiety, and psychological distress, contributing to reduced quality of life, symptom exacerbation, and higher risk of downstream morbidity and mortality [4,5].

Rural populations may be particularly vulnerable to this clustering of physical and mental health burden. Compared with urban populations, rural residents experience a higher prevalence of chronic disease [6] and face substantial barriers to accessing specialized care, including mental health services. Even among individuals prescribed pharmacological treatment for mental health conditions, access to follow-up and supportive services declines across urban, rural-adjacent, and rural-nonadjacent settings [7]. These disparities underscore the need for accessible, scalable approaches to chronic disease management that extend beyond disease-specific biomedical care.

Effective self-management is a central component of chronic disease care. Clinical practice guidelines increasingly emphasize self-management education and support as core elements of long-term disease control [8]. Evidence from systematic reviews and meta-analyses demonstrates that self-management programs improve quality of life, self-efficacy, and depressive symptoms [9]; that interventions incorporating psychotherapeutic components yield superior physiological outcomes [10]; and that multicomponent approaches targeting multiple behaviors simultaneously are more effective than single-behavior interventions [11]. Despite this evidence, many chronic disease programs omit outcomes and supports that patients identify as important, including stress regulation, emotional well-being, and social connection [12]. Social support in particular influences self-management both directly and indirectly through empowerment and psychological resilience [13]. Even in the context of emerging pharmacotherapies for obesity-related and cardiometabolic conditions, experts increasingly emphasize the need for integrated psychosocial and behavioral support to sustain benefits and mitigate adverse psychological effects [14].

Although many chronic diseases are preventable or modifiable, health systems remain largely organized around reactive, episodic care, prioritizing treatment of acute disease escalation rather than sustained self-management and prevention. Models such as the Expanded Chronic Care Model emphasize proactive support, patient activation, and community-based resources, yet scalable implementations of these principles remain limited, particularly for populations living with multimorbidity and in rural settings. As a result, many individuals experience fragmented care that addresses downstream symptoms without adequately targeting upstream behavioral and psychosocial processes that shape chronic disease trajectories over time.

Digital health interventions offer a promising and scalable approach to address these gaps [15,16]. Importantly, digital self-management programs can extend behavioral and psychosocial support beyond traditional clinical encounters by embedding brief, scalable guidance into patients' daily routines, enabling continuous engagement with health behaviors between health care visits. Such interventions can extend care beyond traditional clinical encounters; support continuity and personalization; and incorporate outcomes that matter to patients, including sleep, physical activity, and psychosocial well-being [15,17]. By addressing shared behavioral and psychosocial risk factors across conditions, digital self-management programs have the potential to activate patients in their own care, support daily decision-making, and improve adherence and longer-term outcomes.

While digital interventions for chronic disease have demonstrated benefits for mental health outcomes, the mechanisms through which these improvements occur remain poorly understood. Meta-analyses of digital interventions for depression suggest that change occurs through multiple indirect pathways, including cognitive, behavioral, self-related, physiological, and mindfulness-related processes [18,19]. However, most individual studies are narrow in scope, mechanistic insights outside depression are limited, and few evaluations examine multidomain change over extended periods. Mechanistic understanding is particularly sparse in rural populations and in digital interventions addressing chronic disease more broadly. Identifying co-occurring changes across psychological, physical, behavioral, and social domains, as well as factors that modify response, is essential for optimizing intervention design and implementation.

To address these gaps, the Pictou County Patient Empowerment Program (PictouCountyPEP) was developed as a 6-month, comprehensive digital self-management program for adults living with chronic conditions in a rural setting [20]. A prior digital health evaluation established the feasibility, usability, and sustained mental health benefits of this low-bandwidth delivery model in rural adults with chronic disease [20]. The present study extends this work by evaluating longer-term change across psychological, physical, behavioral, and social domains and by examining whether improvements occur in theoretically coherent patterns relevant to chronic disease self-management.

The program builds on an empowerment-based intervention model previously evaluated in cancer populations [21,22] and was adapted to remove cancer-specific content while retaining the core behavioral and psychosocial components. Rather than targeting a single symptom or disease-specific pathway, the program was designed to influence modifiable upstream processes related to stress regulation, coping, self-efficacy, health behaviors, and social connection that are expected to covary with psychological distress and chronic disease burden

XSL•FO
RenderX

over time. Chronic diseases frequently cluster and progress through shared upstream processes, including stress dysregulation, physical inactivity, sleep disruption, social isolation, and reduced self-efficacy. Interventions that target these shared processes may be preventive in orientation, with potential to attenuate escalation and accumulation of multimorbidity, even when not designed to prevent specific diagnoses.

The primary objective of this study was to quantify longitudinal changes in psychological outcomes over 12 months following participation in PictouCountyPEP. Concurrent changes in physical, behavioral, and social outcomes were examined to contextualize psychological change. Secondary objectives were to characterize changes across physical, behavioral, and social outcomes; examine whether changes in selected candidate factors were associated with changes in psychological outcomes; and explore whether baseline psychological characteristics modified longitudinal change. These analyses were exploratory and hypothesis-generating and were intended to identify candidate pathways for evaluation in future controlled studies rather than establish causal mechanisms.

Understanding longitudinal psychosocial, physical, and behavioral changes associated with participation in a comprehensive digital self-management program can inform clinical decision-making in rural primary care, support health system planning for chronic disease management, and guide adaptation for populations at higher risk of multimorbidity. Identifying potential psychological pathways of change may further inform the optimization, personalization, and scale-up of digital empowerment interventions in rural and underserved regions.

Methods

Study Design

This study reports outcomes from the PictouCountyPEP, a single-arm, prospective, 6-month digital behavioral intervention delivered to adults living with chronic health conditions in Pictou County, Nova Scotia, Canada (October 2023 – August 2025). The study was designed as a pragmatic, real-world evaluation of feasibility, engagement, and longitudinal change under community delivery conditions rather than as a controlled efficacy trial; accordingly, observed changes over time cannot be attributed solely to the intervention.

Setting

Pictou County is a predominantly rural region with approximately 44,000 residents and limited access to specialty health services [23].

Participant Inclusion and Exclusion Criteria

Adults aged 18 years or older residing in Pictou County and living with 1 or more self-reported chronic conditions were eligible to participate. Eligibility was not restricted to a prespecified list of chronic disease diagnoses; the program was intentionally designed for adults living with chronic conditions across diagnostic categories and for those with multimorbidity. Eligibility required the ability to safely engage in low- to moderate-intensity physical activity, assessed using the Physical Activity Readiness Questionnaire for Everyone (PAR-Q+) [24], English proficiency, and access to the internet and email. No specific chronic disease categories were otherwise excluded; however, individuals with a current or recent cancer diagnosis were excluded because the intervention model has been evaluated separately in cancer populations in prior trials [21,22].

Recruitment and Sampling Procedures

Recruitment occurred through community-based outreach, including local media, social media, flyers, and partner organizations [20]. Eligible participants were enrolled consecutively as they presented for participation during the recruitment period.

Ethical Considerations

Participants provided electronic informed consent via the REDCap eConsent framework. No financial compensation for study participation was provided. Data were collected and managed in accordance with Nova Scotia Health privacy policies and the requirements of Canada's Personal Information Protection and Electronic Documents Act (PIPEDA). No images identifying individual participants are included in the manuscript or supplementary materials. The study was approved by the Nova Scotia Health Research Ethics Board (REB 1029744) prior to participant enrollment and was registered on ClinicalTrials.gov (NCT06064110).

Data Collection and Assessment Schedule

Self-reported surveys were administered using REDCap [25]. Participants completed assessments at baseline, 6 months, and 12 months.

Intervention

PictouCountyPEP is a home-based, multicomponent digital intervention designed to support self-management and mental well-being among adults living with chronic health conditions. The intervention was adapted from previously evaluated patient empowerment program models and modified to remove all cancer-specific education. The full protocol has been published previously [20].

The program delivered daily automated emails over a 26-week period, linking participants to brief instructional videos hosted by authors GI and RDHR. The videos were approximately 4 to 6 minutes in duration and accessible asynchronously via a public video-hosting platform. The content addressed 6 behavioral domains: aerobic physical activity, strength training or yoga, relaxation and stress management, dietary guidance, habit formation, and relationship and social connection skills. The content was modular, allowing flexible engagement according to individual needs and symptom burden.

Exercise guidance aligned with the national physical activity recommendations, encouraging 150 to 300 minutes per week of moderate-intensity aerobic activity and strength or flexibility training at least twice weekly. The content emphasized pacing, gradual progression, and symptom monitoring, with encouragement to tailor activities in consultation with health care providers. Dietary guidance promoted a predominantly plant-based pattern with reduced intake of ultra-processed foods,

XSL•FO
RenderX

consistent with Canada's Food Guide [26]. The stress management content incorporated mindfulness-based practices, breathing techniques, and cognitive strategies. The relationship modules focused on communication, boundary setting, and social connection, while the habit formation content emphasized small, sustainable behavioral changes.

The intervention was designed for low-bandwidth delivery and mobile accessibility to minimize barriers related to internet speed, device availability, or digital literacy. No specialized applications, wearable devices, or synchronous participation was required. Engagement was encouraged through daily reminders and weekly compliance surveys. Optional components included a peer support option, monthly live group videoconference sessions, and occasional in-person community events. These components were offered to enhance social connection and accountability but were not required or standardized. The intervention was delivered under real-world conditions to maximize external validity and inform future scale-up. Given the single-arm interventional design, this study follows the TREND (Transparent Reporting of Evaluations with Nonrandomized Designs) guidelines for nonrandomized evaluations of behavioral interventions [27]. The daily digital intervention content was delivered during the initial 6 months. During months 7 to 12, participants entered a lower-intensity maintenance period in which monthly group videoconference sessions continued to be offered, but daily intervention emails were no longer delivered. The 12-month assessment therefore reflects outcomes following the initial 6-month daily intervention and the subsequent 6-month lower-intensity maintenance period.

Because the intervention content was delivered asynchronously and participants could engage flexibly with individual components, standardized individual-level measures of total intervention exposure or module completion were not available for the present analysis; consequently, a formal dose-response analysis was not conducted.

Outcomes and Measures

Primary and Secondary Outcomes and Measures

Primary outcomes were prespecified psychological measures reflecting illness perceptions, well-being, perceived stress, life satisfaction, and disease-related worry. Participants completed validated self-report measures assessing psychological, physical, behavioral, and social outcomes. All multi-item composite scales were scored using complete item-level data. References for all measures and cut points can be found in the Multimedia Appendix 1.

Primary Psychological Measures

Illness perceptions were assessed using the 9-item Brief Illness Perception Questionnaire (B-IPQ), which measures the perceived consequences, timeline, personal control, treatment control, symptom identity, concern, emotional response, and illness coherence related to chronic illness [28]. Each item is rated on a 0 to 10 numeric scale, with 3 items being reverse-coded. Higher scores indicate more threatening or negative illness perceptions. A total score was derived by summing all items, and an improvement of ≥7 was considered clinically meaningful [29].

Psychological well-being was measured using the World Health Organization-Five Well-Being Index (WHO-5), a 5-item instrument assessing positive mood, vitality, and general interest in daily life over the preceding 2 weeks [30]. Raw scores (range 0 - 25) were multiplied by 4 to yield a total score ranging from 0 to 100, with higher scores indicating better well-being. Scores below 50 were used to indicate poor well-being in binary analyses [31].

Perceived stress was assessed using the 10-item Perceived Stress Scale (PSS-10), which measures the extent to which situations in one's life are appraised as stressful, unpredictable, and uncontrollable during the past month [32]. Items are rated from 0 ("never") to 4 ("very often"), with 4 items being reverse-coded. The items were summed to produce a total score ranging from 0 to 40, with higher scores indicating greater perceived stress. Scores ≥20 were used to classify high perceived stress, consistent with prior studies [33].

Life satisfaction was measured using the Satisfaction With Life Scale (SWLS), a 5-item instrument assessing global cognitive judgments of satisfaction with one's life [34]. Items are rated from 1 ("strongly disagree") to 7 ("strongly agree") and summed to yield total scores ranging from 5 to 35, with higher scores indicating greater life satisfaction. For responder analyses, an improvement of ≥5 points was used as an interpretive threshold informed by the scale author's score categories [35].

Disease-related worry was assessed using the 10-item Questionnaire on Distress in Cancer Patients (QSC-R10), which measures illness-related fears and concerns relevant to individuals living with chronic conditions [36]. Each item assesses whether a concern applies and the degree of distress it causes, rated from 0 to 5. Items were summed to produce total scores ranging from 0 to 50, with higher scores indicating greater disease-related worry. A cut point of>14 was used to classify clinically relevant need for psychosocial support in exploratory analyses [36].

Secondary Physical, Psychosocial, and Behavioral Measures

BMI was derived from self-reported weight and height. BMI (kg/m^2) was calculated as weight (kg) divided by height (m) squared. Outcomes included continuous change in weight (lb) and BMI, as well as categorical indicators of clinically meaningful weight loss (≥3%), consistent with previous trials and recommended guidelines [37,38].

Health-related quality of life was assessed using the 12-item Short Form Health Survey (SF-12), yielding norm-based Physical Component Summary (PCS) and Mental Component Summary (MCS) scores standardized to a mean of 50 and SD of 10 in the general population [39,40]. Higher scores indicate better health-related quality of life. Score changes of ≥3 points were considered clinically meaningful, in line with previous research [41].

Fatigue was assessed using the 13-item Functional Assessment of Chronic Illness Therapy—Fatigue scale (FACIT-F) [42],

XSL•FO
RenderX

which evaluates fatigue severity and its impact on daily functioning. Items are rated from 0 ("not at all") to 4 ("very much") and summed after reverse scoring, such that higher scores indicate less fatigue. A change in score of ≥3 points was used to define clinically meaningful improvement [43].

Sleep quality was measured using the Pittsburgh Sleep Quality Index (PSQI), which yields a global score ranging from 0 to 21, with higher scores indicating poorer sleep quality [44]. A global score >5 was used to identify poor sleep [44].

Sedentary behavior was assessed using the adult Sedentary Behavior Questionnaire (SBQ), which captures time spent in common sedentary activities during weekdays and weekends [45]. Reported hours were converted to total daily sedentary time, with values exceeding 24 hours being truncated. Daily weekday and weekend hours were summed to obtain weekly sedentary hours. High sedentary behavior was defined as ≥8 hours per day (equivalent to ≥56 hours per week), consistent with prior studies [46].

Perceived social support was measured using the 19-item Medical Outcomes Study Social Support Survey (MOS-SSS), yielding emotional and informational, tangible, affectionate, positive interaction, and overall support scores [47]. Items are rated from 1 to 5 and averaged according to scoring guidelines, with higher scores indicating greater perceived support.

Diet quality was assessed using the Rapid Eating Assessment for Participants—Shortened version (REAP-S), a 13-item measure of dietary habits over the past week [48]. Items are scored from 1 to 3, yielding total scores ranging from 13 to 39, with higher scores indicating healthier dietary patterns.

Spiritual well-being was measured using the 12-item Functional Assessment of Chronic Illness Therapy—Spiritual Well-Being scale (FACIT-SP-12), which assesses meaning, peace, and faith [49]. Items are scored from 0 to 4 (reversed where scoring instructions indicate) and summed to yield total scores ranging from 0 to 48, with higher scores indicating greater spiritual well-being.

Alcohol consumption was assessed using self-reported drinking and binge drinking frequency over the preceding 6 months. Binge drinking was defined as consuming 5 or more drinks containing any kind of alcohol within a 2-hour period [50]. Original categorical responses were converted into continuous frequency scores to retain variability and maximize statistical power, with values ranging from 0 (none) to 9 (daily) [51].

Cannabis use was assessed using self-reported frequency of use over the preceding 6 months and self-identification as a regular user. Categorical frequency responses were converted into a continuous score ranging from 0 (no use) to 4 (8 or more uses), using the same rationale applied to alcohol frequency scoring [51].

Measurement Reliability

Internal consistency of all multi-item scales was assessed at baseline using Cronbach α with 95% CI to characterize measurement reliability in the study population prior to intervention exposure (Table S1 in Multimedia Appendix 1).

Covariates

Covariates included age, sex, relationship status, comorbidity burden measured by the Charlson Comorbidity Index, and use of medication for depression or anxiety at baseline. These variables were selected a priori based on theoretical relevance and prior evidence from related patient empowerment program trials identifying their associations with mental health outcomes and intervention response [20-22].

Sample Size, Power, and Precision

This phase 2 interventional study enrolled 182 participants to evaluate feasibility, engagement, and longitudinal changes across a broad set of health outcomes. The sample size was determined pragmatically based on anticipated recruitment capacity within the study region and time frame rather than a formal a priori power calculation, consistent with the study's phase 2 implementation-focused objectives.

Statistical Analysis

All participants who consented and completed baseline assessments were included in the analytic cohort. Primary longitudinal analyses incorporated all available observed outcome data from participants who contributed data at the relevant assessment time points.

Longitudinal changes in psychosocial, physical, and behavioral outcomes were examined using generalized estimating equations (GEEs), which estimate population-averaged effects while accounting for within-participant correlation across repeated measurements. Continuous outcomes were modeled using an identity link with Gaussian working variance, and binary outcomes defined using established clinical thresholds (Multimedia Appendix 1) were modeled using a logit link with binomial working variance. An exchangeable working correlation structure was specified for all models, reflecting the limited number of equally spaced follow-up assessments, and robust (sandwich) SEs were used for inference to ensure valid estimation under potential misspecification of the working correlation. GEEs were selected to estimate population-averaged effects aligned with the study objective of evaluating overall program-associated change rather than subject-specific trajectories.

Time was modeled as a categorical variable (baseline, 6 months, and 12 months). All models were adjusted a priori for baseline age, sex, relationship status, comorbidity burden (Charlson Comorbidity Index), and use of medication for depression or anxiety. Model-based estimated marginal means and corresponding SEs were derived at each time point, and pairwise contrasts were used to estimate differences between baseline and follow-up assessments, with 95% CI reported. Primary inference for continuous outcomes was based on time indicator coefficients from these models, representing adjusted population-averaged mean differences from baseline.

Exploratory interaction models were estimated for psychological outcomes only, by including baseline characteristics interacted with time, to examine whether baseline factors modified trajectories of psychological change. These analyses were

XSL•FO
RenderX

prespecified as exploratory and interpreted as hypothesis-generating.

Exploratory analyses examined associations between within-participant changes in selected candidate factors and changes in psychological outcomes. These analyses were intended to identify patterns of co-occurring change and generate hypotheses regarding potential pathways rather than to test formal statistical mediation or establish causal mechanisms. Residualized change scores from baseline to follow-up were calculated, and linear regression models were fitted to assess associations between changes in physical, psychosocial, and substance use measures and changes in psychological well-being (WHO-5), perceived stress (PSS-10), life satisfaction (SWLS), and disease-related worry (QSC-R10). Models were adjusted for baseline outcome values and prespecified covariates. Results are interpreted as exploratory associations between longitudinal changes rather than as estimates of mediation or causal effects.

Clinically meaningful improvement was further examined using responder analyses based on established thresholds. Two responder definitions were applied: magnitude-based change, defined as achieving a prespecified minimally important difference regardless of baseline status, and categorical improvement, defined as movement from a clinically worse category at baseline to a more favorable category at follow-up, restricted to participants eligible to improve. Responder proportions at 6 and 12 months were calculated with 95% CI. Exploratory univariate logistic regression models were used to examine associations between baseline characteristics and responder status.

Missing Data and Sensitivity Analyses

Missing follow-up data were evaluated using several complementary approaches. In unadjusted comparisons, participants with incomplete follow-up had significantly lower baseline well-being (WHO-5; *P*=.03) and life satisfaction (SWLS; *P*=.01) than participants with complete follow-up, whereas no statistically significant differences were observed for age, sex, comorbidity burden, relationship status, baseline depression/anxiety medication use, illness perceptions (B-IPQ), perceived stress (PSS-10), or disease-related worry (QSC-R10) (Table S2 in Multimedia Appendix 1). Complete follow-up was defined as completion of assessments at baseline, 6 months, and 12 months; follow-up was nonmonotonic, with some participants missing the 6-month assessment but returning at 12 months. Missingness was further evaluated using logistic regression models for incomplete follow-up, including all prespecified covariates and primary psychological outcomes as predictors. None of the included variables was statistically significantly associated with incomplete follow-up in the multivariable model. Primary analyses used covariate-adjusted GEE models and included all available observed outcome data, regardless of follow-up status. To evaluate the robustness of findings to incomplete follow-up, sensitivity analyses were conducted using both complete case analyses and analyses using stabilized inverse probability weights for attrition. Findings were highly consistent across the primary, complete case, and weighted analyses, indicating that the principal results were not materially altered by the analytic approach to incomplete follow-up.

All analyses were conducted using R software (version 4.5.1; R Foundation for Statistical Computing) [52]. Statistical tests were 2-sided, with a significance level of α=.05. Given the number of outcomes assessed, no formal adjustment for multiple comparisons was applied; results for secondary and exploratory analyses are therefore interpreted with emphasis on effect sizes, CI, and consistency of findings rather than statistical significance alone, and these findings are considered hypothesis-generating rather than confirmatory. Full model–based estimated marginal means and associated SEs for continuous outcomes are provided in Multimedia Appendix 1.

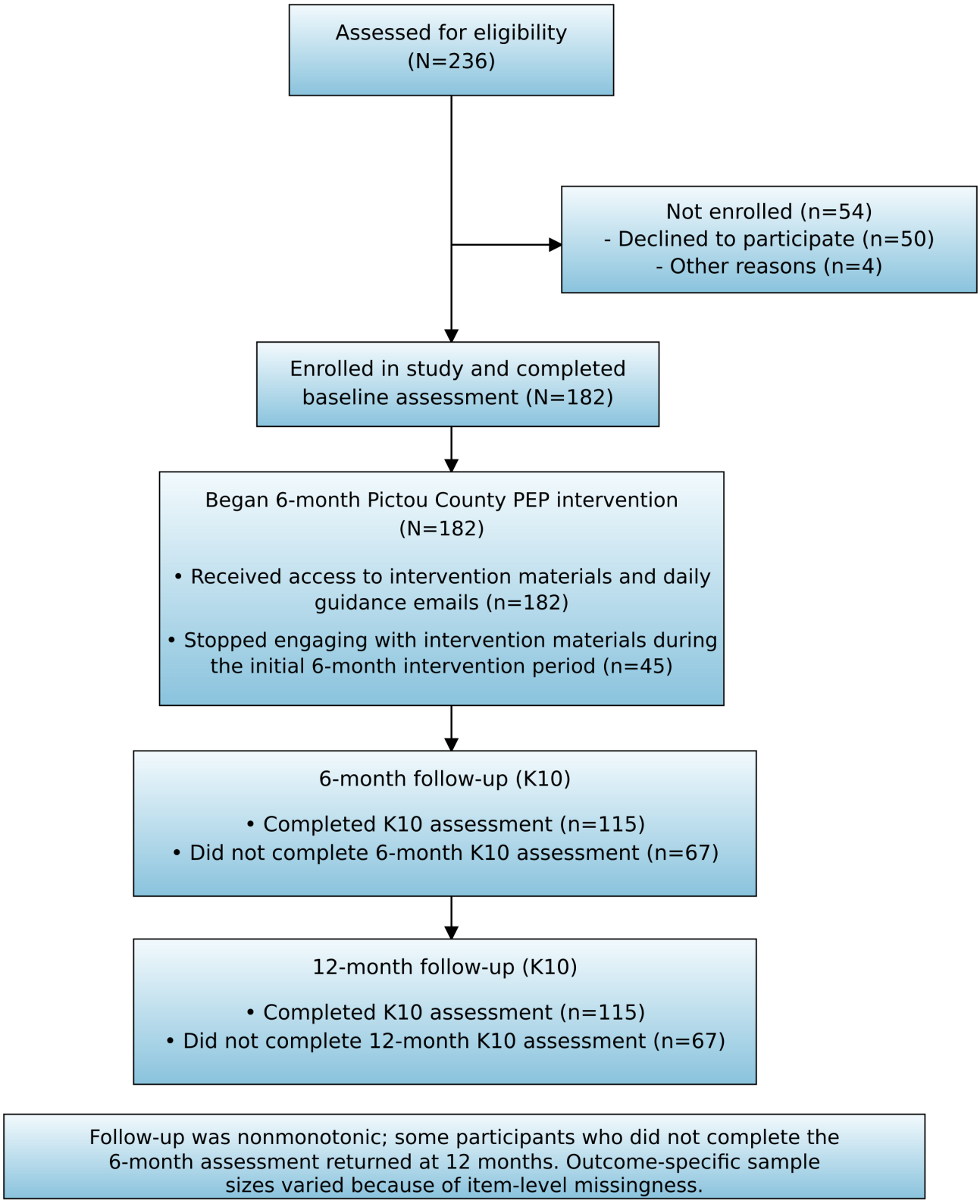
Results

Participants and Baseline Characteristics

A total of 182 individuals enrolled and completed the baseline assessment. Across outcomes and follow-up assessments, outcome-specific completion ranged from 106 of 182 (58.2%) participants to 119 of 182 (65.4%) participants (Figure 1; Table S4 in Multimedia Appendix 1). Baseline characteristics are shown in Table 1. Participants had a median age of 60 (IQR 51 - 68) years; of the 182 participants, 155 (85.2%) were women, and 89 (48.9%) resided in the Municipality of Pictou County. The median number of diagnosed chronic conditions was 3 (IQR 1 - 3), and the median age-adjusted Charlson Comorbidity Index was 2 (IQR 1 - 3). The baseline symptom burden was substantial: of 182 participants, 52.7% (n=96) screened positive for low well-being (WHO-5<50), 41.2% (n=75) screened positive for high perceived stress (PSS-10≥20), and 48.4% (n=88) screened positive for psychological distress. Additional baseline sociodemographic, clinical, and health care use characteristics are reported in Table S3 in Multimedia Appendix 1. The follow-up status for each outcome is reported in Table S4 Multimedia Appendix 1.

XSL•FO
RenderX

Figure 1. Participant flow is presented in the JARS-adapted participant flow diagram. JARS: Journal Article Reporting Standards; PEP: patient empowerment program.



XSL•FO
RenderX

Table . Baseline sociodemographic and clinical characteristics of participants enrolled in the Pictou County Patient Empowerment Program (N=182)[a].

Characteristic	Value
Age (years), median (IQR)	60 (51-68)
Sex assigned at birth, n (%)	
Male	27 (14.8)
Female	155 (85.2)
Gender (self-report), n (%)	
Male	27 (14.8)
Female	155 (85.2)
Sexual orientation, n (%)	
Heterosexual	168 (92.3)
Gay (self-identified)	1 (0.6)
Bisexual	7 (3.8)
Other/unsure/prefer not to answer	6 (3.3)
Race, n (%)	
White	175 (96.2)
Black	3 (1.6)
South Asian	1 (0.6)
Other	3 (1.6)
Education, n (%)	
High school or other certification	110 (60.4)
University education or higher	72 (39.6)
Household income (CAD $)[b], n (%)	
<50,000	38 (20.9)
50,000 - 99,999	71 (39.0)
≥100,000	55 (30.2)
Do not know/prefer not to answer	18 (9.9)
Employment status, n (%)	
Working full- or part-time	90 (49.5)
Retired/not working	92 (50.5)
Currently partnered, n (%)	131 (72.0)
Weight 2 years before enrollment (lb), median (IQR)	170 (145-206)
Current weight (lb), median (IQR)	194 (165-235)
BMI (kg/m^2), median (IQR)	32 (27-38)
Obesity (BMI≥30), n (%)	107 (58.8)
Number of chronic conditions, median (IQR)	3 (1-3)
Charlson Comorbidity Index (age-adjusted), median (IQR)	2 (1-3)
Rural residence (Municipality of Pictou County), n (%)	89 (48.9)
Current smoking, n (%)	
Daily	6 (3.3)
Less than daily	2 (1.1)
Not at all	174 (95.6)
Cannabis use (past 6 months), n (%)	

XSL•FO
RenderX

Characteristic	Value
Never	77 (42.3)
Former use	61 (33.5)
1 - 7 times	28 (15.4)
≥8 times	16 (8.8)
Alcohol use (past 6 months), n (%)	
Lifetime abstainer	4 (2.2)
Former drinker	31 (17.0)
≤monthly	52 (28.6)
Weekly	45 (24.7)
≥3 days per week	31 (17.0)
Has a regular primary care provider, n (%)	136 (74.7)
Any doctor visit in past 12 months, n (%)	172 (94.5)
Prescribed medication for anxiety or depression, n (%)	76 (41.8)
Screened positive for psychological distress, n (%)	88 (48.4)
Physical activity level, n (%)	
Not very active	95 (52.2)
Moderately active	61 (33.5)
Very active	26 (14.3)

[a]Values are median (IQR) or n (%); percentages may not sum to 100 due to rounding.

[b]CAD $1=US $0.73 as of August 29, 2025 (Bank of Canada).

Primary Outcomes: Longitudinal Changes in Psychological Outcomes

In adjusted population-averaged GEE models, all primary psychological outcomes improved from baseline to 6 months, with effects sustained or larger at 12 months (Table 2; Figure 2). Illness perceptions (B-IPQ) decreased at 6 months (β=−5.1, 95% CI −7.3 to −3.0; *P*<.001) and 12 months (β=−7.3, 95% CI −9.7 to −4.9; *P*<.001). Perceived stress (PSS-10) declined at 6 months (β=−2.2, 95% CI −3.2 to −1.2; *P*<.001) and 12 months (β=−2.4, 95% CI −3.5 to −1.3; *P*<.001), and disease-related worry (QSC-R10) decreased at 6 months (β=−3.7, 95% CI −5.2 to −2.3; *P*<.001) and 12 months (β=−6.3, 95% CI −7.8 to −4.7; *P*<.001). Life satisfaction (SWLS) increased at 6 months (β=1.5, 95% CI 0.5-2.6; *P*=.004) and 12 months (β=2.5, 95% CI 1.4-3.5; *P*<.001). Well-being (WHO-5) increased at 6 months (β=10.1, 95% CI 7.2-13.0; *P*<.001) and 12 months (β=10.4, 95% CI 7.2-13.5; *P*<.001). Mental health–related quality of life (SF-12 MCS) improved at 6 months (β=3.5, 95% CI 1.8-5.2; *P*<.001) and 12 months (β=4.3, 95% CI 2.4-6.2; *P*<.001).

XSL•FO
RenderX

Table . Adjusted population-averaged mean differences from baseline in continuous outcomes at 6 and 12 months (generalized estimating equation models).[a]

Outcome	β (vs baseline) at 6 months	95% CI	*P* value	β (vs baseline) at 12 months	95% CI	*P* value
B-IPQ[b]	−5.1	−7.3 to −3.0	<.001	−7.3	−9.7 to −4.9	<.001
PSS-10[c]	−2.2	−3.2 to −1.2	<.001	−2.4	−3.5 to −1.3	<.001
QSC-R10[d]	−3.7	−5.2 to −2.3	<.001	−6.3	−7.8 to −4.7	<.001
SWLS[e]	1.5	0.5 to 2.6	.004	2.5	1.4 to 3.5	<.001
WHO-5[f]	10.1	7.2 to 13.0	<.001	10.4	7.2 to 13.5	<.001
FACIT-SP-12[g]	2.9	1.8 to 4.0	<.001	4.6	3.0 to 6.2	<.001
MOS-SSS[h]	0.19	0.09 to 0.30	<.001	0.10	−0.01 to 0.21	.08
SBQ[i]	−5.7	−10.0 to −1.4	.01	−8.7	−13.2 to −4.2	<.001
FACIT-F[j]	3.6	2.0 to 5.2	<.001	5.0	3.3 to 6.8	<.001
REAP-S[k]	3.0	2.4 to 3.7	<.001	2.6	2.0 to 3.3	<.001
PSQI[l]	−1.8	−2.4 to −1.2	<.001	−2.2	−2.8 to −1.5	<.001
SF-12 PCS[m]	1.8	0.3 to 3.3	.02	2.9	1.0 to 4.9	.003
SF-12 MCS[n]	3.5	1.8 to 5.2	<.001	4.3	2.4 to 6.2	<.001
BMI	−1.2	−1.8 to −0.7	<.001	−1.7	−2.5 to −0.9	<.001
Weight (lb)	−7.2	−10.4 to −4.0	<.001	−10.3	−15.2 to −5.4	<.001
Drink score	−0.5	−0.8 to −0.2	.002	−0.6	−0.9 to −0.4	<.001
Binge score	−0.21	−0.42 to 0.00	.047	−0.03	−0.31 to 0.26	.90
Cannabis score	−0.13	−0.27 to 0.01	.06	−0.12	−0.27 to 0.02	.09

[a]β values represent adjusted mean differences from baseline. Models were adjusted for age, sex, comorbidity burden, relationship status, and psychotropic medication use.

[b]B-IPQ: Brief Illness Perception Questionnaire.

[c]PSS-10: Perceived Stress Scale.

[d]QSC-R10: Questionnaire on Distress in Cancer Patients.

[e]SWLS: Satisfaction With Life Scale.

[f]WHO-5: World Health Organization-Five Well-Being Index.

[g]FACIT-SP-12: Functional Assessment of Chronic Illness Therapy—Spiritual Well-Being scale.

[h]MOS-SSS: Medical Outcomes Study Social Support Survey.

[i]SBQ: Sedentary Behavior Questionnaire.

[j]FACIT-F: Functional Assessment of Chronic Illness Therapy—Fatigue scale.

[k]REAP-S: Rapid Eating Assessment for Participants—Shortened.

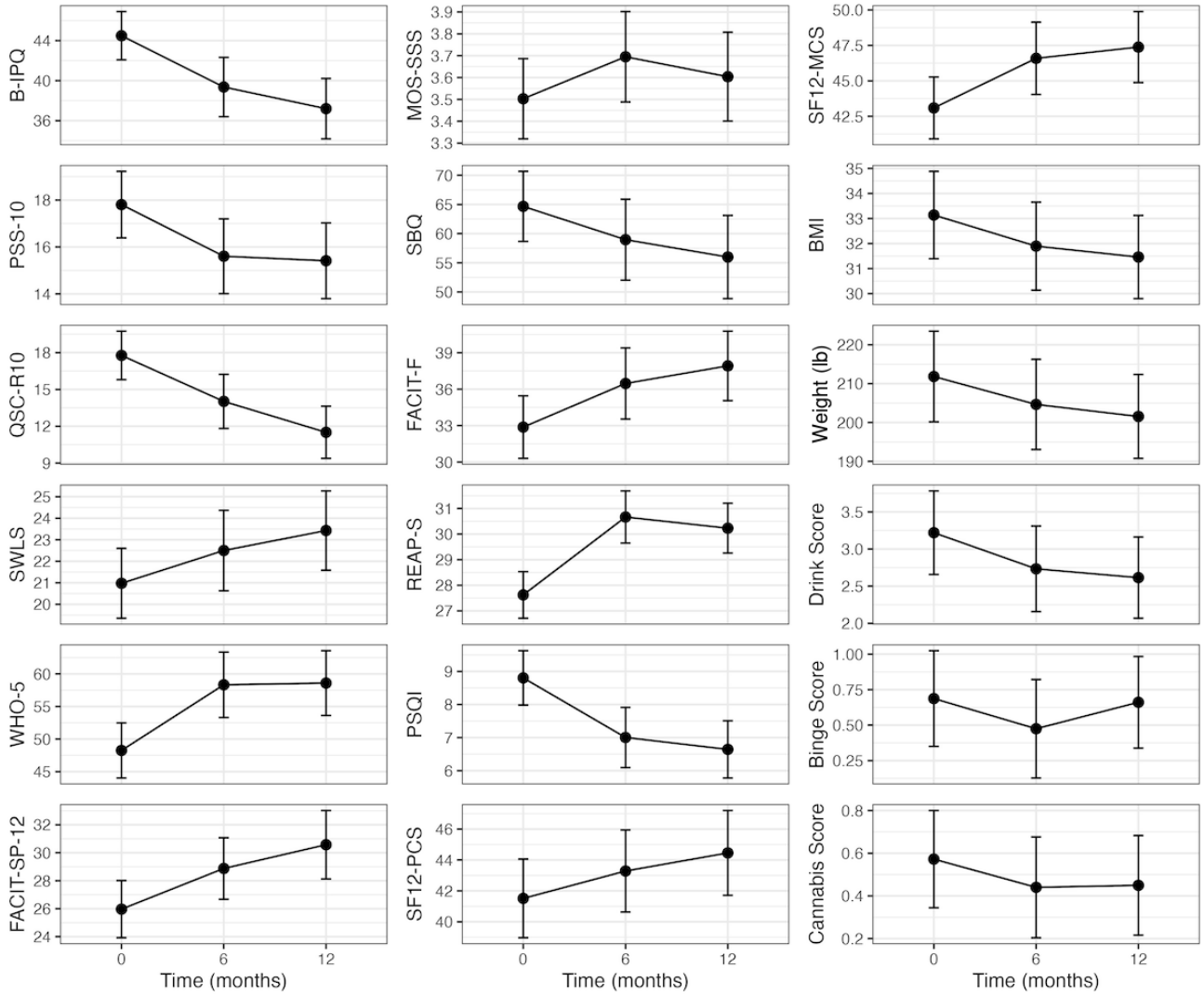
[l]PSQI: Pittsburgh Sleep Quality Index.

[m]SF-12 PCS: 12-item Short Form Health Survey Physical Component Summary.

[n]SF-12 MCS: 12-item Short Form Health Survey Mental Component Summary.

XSL•FO
RenderX

Figure 2. Adjusted marginal means (95% CI) at baseline (0 months), 6 months, and 12 months for continuous outcomes. Marginal means were estimated from adjusted population-averaged generalized estimating equation models including age, sex, comorbidity burden, relationship status, and baseline use of medication for depression or anxiety. Lower scores represent improvement for B-IPQ, PSS-10, QSC-R10, SBQ, PSQI, BMI, weight, and substance use frequency measures, whereas higher scores represent improvement for the positively scored measures. B-IPQ: Brief Illness Perception Questionnaire; FACIT-F: Functional Assessment of Chronic Illness Therapy—Fatigue scale; FACIT-SP-12: Functional Assessment of Chronic Illness Therapy—Spiritual Well-Being scale; MOS-SSS: Medical Outcomes Study Social Support Survey; PSQI: Pittsburgh Sleep Quality Index; PSS-10: Perceived Stress Scale; QSC-R10: Questionnaire on Distress in Cancer Patients; REAP-S: Rapid Eating Assessment for Participants—Shortened; SBQ: Sedentary Behavior Questionnaire; SF-12 MCS: 12-item Short Form Health Survey Mental Component Summary; SF-12 PCS: 12-item Short Form Health Survey Physical Component Summary; SWLS: Satisfaction With Life Scale; WHO-5: World Health Organization-Five Well-Being Index.



Secondary Outcomes: Multidomain Health Changes

Several secondary outcomes improved over follow-up (Table 2; Figure 2). Fatigue (FACIT-F) improved at 6 months (β=3.6, 95% CI 2.0-5.2; *P*<.001) and 12 months (β=5.0, 95% CI 3.3-6.8; *P*<.001), and sleep quality (PSQI) improved at 6 months (β=−1.8, 95% CI −2.4 to −1.2; *P*<.001) and 12 months (β=−2.2, 95% CI −2.8 to −1.5; *P*<.001). Diet quality (REAP-S) improved at 6 months (β=3.0, 95% CI 2.4-3.7; *P*<.001) and 12 months (β=2.6, 95% CI 2.0-3.3; *P*<.001), and sedentary behavior (SBQ) decreased at 6 months (β=−5.7, 95% CI −10.0 to −1.4; *P*=.01) and 12 months (β=−8.7, 95% CI −13.2 to −4.2; *P*<.001). Spiritual well-being (FACIT-SP-12) improved at 6 months (β=2.9, 95% CI 1.8-4.0; *P*<.001) and 12 months (β=4.6, 95% CI 3.0-6.2; *P*<.001). Perceived social support (MOS-SSS) increased at 6 months (β=0.2, 95% CI 0.1-0.3; *P*<.001) but was not clearly different from baseline at 12 months (β=0.1, 95% CI −0.01 to 0.21; *P*=.08). Physical health–related quality of life (SF-12 PCS) improved at 6 months (β=1.8, 95% CI 0.3-3.3; *P*=.02) with a larger improvement at 12 months (β=2.9, 95% CI 1.0-4.9; *P*=.003). Weight and BMI decreased at both follow-ups (weight, 6 months: β=−7.2 lb, 95% CI −10.4 to −4.0, *P*<.001; 12 months: β=−10.3 lb, 95% CI −15.2 to −5.4, *P*<.001; BMI, 6 months: β=−1.2, 95% CI −1.8 to −0.7, *P*<.001; 12 months: β=−1.7, 95% CI −2.5 to −0.9, *P*<.001). Alcohol frequency decreased at 6 months (Drink score β=−0.5, 95% CI −0.8 to −0.2; *P*=.002) and 12 months (β=−0.6, 95% CI −0.9 to −0.4; *P*<.001). Binge drinking frequency showed a small reduction at 6 months (β=−0.21, 95% CI −0.42 to 0.00; *P*=.047) but was not clearly different from baseline at 12 months (β=−0.03, 95% CI −0.31 to 0.26; *P*=.90), and cannabis use frequency did not show a clear change at either follow-up (6 months: β=−0.13, 95% CI −0.27 to 0.01; *P*=.06; 12 months: β=−0.12, 95% CI −0.27 to 0.02; *P*=.09). Model-estimated

XSL•FO
RenderX

marginal means and full contrasts for all continuous outcomes are provided in Tables S5 and S6 in Multimedia Appendix 1.

Binary Clinical Threshold Outcomes

Relative to baseline, the odds of low well-being (WHO-5<50) were lower at 6 months (odds ratio [OR] 0.17, 95% CI 0.11-0.26; *P*<.001) and 12 months (OR 0.15, 95% CI 0.10-0.24; *P*<.001) (Table 3). The odds of high perceived stress (PSS-10≥20) were lower at 6 months (OR 0.31, 95% CI 0.21-0.48; *P*<.001) and 12 months (OR 0.20, 95% CI 0.13-0.33; *P*<.001). The odds of clinically elevated disease-related worry (QSC-R10>14) decreased at 6 months (OR 0.39, 95% CI 0.26-0.57; *P*<.001) and 12 months (OR 0.24, 95% CI 0.15-0.37; *P*<.001). Poor sleep (PSQI>5) decreased at 6 months (OR 0.45, 95% CI 0.30-0.68; *P*<.001) and 12 months (OR 0.33, 95% CI 0.21-0.52; *P*<.001), and high sedentary behavior (≥8 hours/day; ≥56 hours/week) decreased at 6 months (OR 0.59, 95% CI 0.39-0.89; *P*=.01) and 12 months (OR 0.43, 95% CI 0.28-0.65; *P*<.001). Obesity prevalence (BMI≥30) did not change at 6 months (OR 0.88, 95% CI 0.72-1.07; *P*=.20) or 12 months (OR 0.88, 95% CI 0.67-1.15; *P*=.30).

Table . Adjusted odds ratios (ORs) for meeting clinical threshold outcomes at 6 and 12 months vs baseline (GEE[a] models[b]).

Outcome (threshold)	OR (vs baseline), 6 months	95% CI	*P* value	OR (vs baseline), 12 months	95% CI	*P* value
Low well-being (WHO-5[c]<50)	0.17	0.11-0.26	<.001	0.15	0.10-0.24	<.001
High perceived stress (PSS-10[d]≥20)	0.31	0.21-0.48	<.001	0.20	0.13-0.33	<.001
Clinically elevated worry (QSC-R10[e]>14)	0.39	0.26-0.57	<.001	0.24	0.15-0.37	<.001
Obesity (BMI≥30)	0.88	0.72-1.07	.20	0.88	0.67-1.15	.30
Poor sleep (PSQI[f]>5)	0.45	0.30-0.68	<.001	0.33	0.21-0.52	<.001
High sedentary time (SBQ[g]≥56 hours/week)	0.59	0.39-0.89	.011	0.43	0.28-0.65	<.001

[a]GEE: generalized estimating equation.

[b]ORs represent adjusted odds ratios relative to baseline. Models were adjusted for age, sex, comorbidity burden, relationship status, and psychotropic medication use.

[c]WHO-5: World Health Organization-Five Well-Being Index.

[d]PSS-10: Perceived Stress Scale.

[e]QSC-R10: Questionnaire on Distress in Cancer Patients.

[f]PSQI: Pittsburgh Sleep Quality Index.

[g]SBQ: Sedentary Behavior Questionnaire.

Responder analyses showed clinically meaningful improvement across multiple domains (Table S7 in Multimedia Appendix 1). At 6 and 12 months, respectively, 39 of 107 (36.4%, 95% CI 27.3%-45.6%) and 46 of 106 (43.4%, 95% CI 34.0%-52.8%) participants achieved a ≥4-point reduction in perceived stress (PSS-10), while 53 of 108 (49.1%, 95% CI 39.6%-58.5%) and 47 of 108 (43.5%, 95% CI 34.2%-52.9%) achieved a ≥10-point improvement in mental well-being (WHO-5). At 12 months, clinically meaningful improvements were also observed for illness perceptions (B-IPQ; 54/108, 50.0%, 95% CI 40.6%-59.4%), disease-related worry (QSC-R10; 56/111, 50.5%, 95% CI 41.1%-59.8%), fatigue (FACIT-F; 62/109, 56.9%, 95% CI 47.6%-66.2%), physical health–related quality of life (SF-12 PCS; 60/111, 54.1%, 95% CI 44.8%-63.3%), mental health–related quality of life (SF-12 MCS; 53/111, 47.7%, 95% CI 38.5%-57.0%), and ≥3% weight loss (52/115, 45.2%, 95% CI 36.1%-54.3%).

Exploratory Analyses: Co-Occurring Changes in Candidate Factors and Psychological Outcomes

Exploratory adjusted models examined associations between changes in prespecified candidate factors and changes in psychological outcomes (Table 4). Improvements in fatigue were consistently associated with improvements in well-being, life satisfaction, perceived stress, and disease-related worry. From baseline to 6 months, improvement in FACIT-F scores was associated with improvement in WHO-5 (β=0.9, 95% CI 0.6 to 1.2; *P*<.001), SWLS (β=0.3, 95% CI 0.2-0.4; *P*<.001), PSS-10 (β=−0.4, 95% CI −0.5 to −0.3; *P*<.001), and QSC-R10 (β=−0.6, 95% CI −0.8 to −0.5; *P*<.001) scores.

XSL•FO
RenderX

Table . Exploratory associations between change in candidate factors and change in psychological outcomes from baseline to 6 and 12 months (adjusted GEE[a] models).[b]

Interval and candidate factor (Δ)	ΔWHO-5[c]		ΔSWLS[d]		ΔPSS-10[e]		ΔQSC-R10[f]	
	β (95% CI)	*P* value	β (95% CI)	*P* value	β (95% CI)	*P* value	β (95% CI)	*P* value
Baseline to 6 months								
Fatigue (FACIT-F[g])	0.9 (0.6 to 1.2)	<.001	0.3 (0.2 to 0.4)	<.001	−0.4 (−0.5 to −0.3)	<.001	−0.6 (−0.8 to −0.5)	<.001
Sleep quality (PSQI[h])	−2.2 (−3.1 to −1.3)	<.001	−0.8 (−1.2 to −0.5)	<.001	0.8 (0.4 to 1.1)	<.001	1.4 (0.9 to 1.8)	<.001
Mental HRQoL[i] (SF-12 MCS[j])	0.9 (0.6 to 1.2)	<.001	0.3 (0.2 to 0.4)	<.001	−0.4 (−0.5 to −0.3)	<.001	−0.4 (−0.6 to −0.3)	<.001
Social support (MOS-SSS[k])	8.2 (3.8 to 12.6)	<.001	3.3 (1.6 to 5.0)	<.001	−2.4 (−4.0 to −0.8)	.004	−2.4 (−4.9 to −0.02)	.049
Sedentary behavior (SBQ[l])	−0.10 (−0.22 to 0.03)	.12	−0.06 (−0.11 to −0.02)	.005	0.05 (0.01 to 0.10)	.01	0.13 (0.07 to 0.20)	<.001
Diet quality (REAP-S[m])	1.1 (0.2 to 1.9)	.01	0.4 (0.1 to 0.7)	.02	−0.5 (−0.8 to −0.2)	<.001	−0.8 (−1.2 to −0.4)	<.001
Spiritual well-being (FACIT-SP-12[n])	1.1 (0.7 to 1.5)	<.001	0.5 (0.3 to 0.6)	<.001	−0.4 (−0.5 to −0.2)	<.001	−0.4 (−0.6 to −0.2)	.001
Illness perceptions (B-IPQ[o])	−0.6 (−0.8 to −0.4)	<.001	−0.2 (−0.3 to −0.1)	<.001	0.3 (0.2 to 0.3)	<.001	0.4 (0.3 to 0.5)	<.001
Baseline to 12 months								
Fatigue (FACIT-F)	1.3 (1.0 to 1.6)	<.001	0.2 (0.1 to 0.4)	<.001	−0.4 (−0.5 to −0.3)	<.001	−0.6 (−0.7 to −0.5)	<.001
Sleep quality (PSQI)	−2.8 (−3.7 to −1.9)	<.001	−0.7 (−1.0 to −0.4)	<.001	0.7 (0.4 to 1.0)	<.001	1.4 (1.0 to 1.8)	<.001
Mental HRQoL (SF-12 MCS)	0.8 (0.5 to 1.2)	<.001	0.2 (0.1 to 0.3)	<.001	−0.3 (−0.4 to −0.2)	<.001	−0.3 (−0.5 to −0.1)	<.001
Social support (MOS-SSS)	8.5 (3.6 to 13.4)	<.001	2.7 (1.2 to 4.2)	<.001	−1.9 (−3.7 to −0.1)	.04	−4.3 (−6.6 to −2.0)	<.001
Sedentary behavior (SBQ)	−0.17 (−0.31 to −0.03)	.02	−0.06 (−0.11 to −0.02)	.008	0.03 (−0.02 to 0.08)	.30	0.12 (0.05 to 0.18)	<.001
Diet quality (REAP-S)	1.6 (0.7 to 2.6)	.001	0.4 (0.06 to 0.7)	.02	−0.4 (−0.8 to −0.1)	.02	−0.8 (−1.2 to −0.3)	.001
Spiritual well-being (FACIT-SP-12)	1.3 (1.0 to 1.6)	<.001	0.4 (0.3 to 0.5)	<.001	−0.3 (−0.4 to −0.2)	<.001	−0.5 (−0.7 to −0.3)	<.001
Illness perceptions (B-IPQ)	−0.5 (−0.8 to −0.3)	<.001	−0.2 (−0.3 to −0.1)	<.001	0.2 (0.1 to 0.3)	<.001	0.4 (0.3 to 0.5)	<.001

[a]GEE: generalized estimating equation.

XSL•FO
RenderX

[b]β values represent adjusted associations between change in the candidate mediator (Δ) and change in the psychological outcome (Δ) over the specified interval. Models were adjusted for age, sex, relationship status, Charlson Comorbidity Index, psychotropic medication use, and baseline outcome values. Analyses are exploratory and hypothesis-generating.

[c]WHO-5: World Health Organization-Five Well-Being Index.

[d]SWLS: Satisfaction With Life Scale.

[e]PSS-10: Perceived Stress Scale.

[f]QSC-R10: Questionnaire on Distress in Cancer Patients.

[g]FACIT-F: Functional Assessment of Chronic Illness Therapy—Fatigue scale.

[h]PSQI: Pittsburgh Sleep Quality Index.

[i]HRQoL: health-related quality of life.

[j]SF-12 MCS: 12-item Short Form Health Survey Mental Component Summary.

[k]MOS-SSS: Medical Outcomes Study Social Support Survey.

[l]SBQ: Sedentary Behavior Questionnaire.

[m]REAP-S: Rapid Eating Assessment for Participants—Shortened.

[n]FACIT-SP-12: Functional Assessment of Chronic Illness Therapy—Spiritual Well-Being scale.

[o]B-IPQ: Brief Illness Perception Questionnaire.

Improvements in sleep quality were similarly associated with improvements across psychological outcomes. From baseline to 6 months, improvement in PSQI scores was associated with improvement in WHO-5 (β=−2.2, 95% CI −3.1 to −1.3; *P*<.001), SWLS (β=−0.8, 95% CI −1.2 to −0.5; *P*<.001), PSS-10 (β=0.8, 95% CI 0.4 to 1.1; *P*<.001), and QSC-R10 (β=1.4, 95% CI 0.9 to 1.8; *P*<.001).

More favorable illness perceptions were also associated with improvements across psychological outcomes. From baseline to 6 months, improvement in B-IPQ scores was associated with improvement in WHO-5 (β=−0.6, 95% CI −0.8 to −0.4; *P*<.001), SWLS (β=−0.2, 95% CI −0.3 to −0.1; *P*<.001), PSS-10 (β=0.3, 95% CI 0.2-0.3; *P*<.001), and QSC-R10 (β=0.4, 95% CI 0.3-0.5; *P*<.001). Similar patterns of association were observed at 12 months (Table 4). Additional exploratory associations between changes in candidate factors and psychological outcomes are reported in Tables S8 and S9 in Multimedia Appendix 1.

Exploratory Analyses: Moderators of Psychological Trajectories

Exploratory moderation analyses assessed whether baseline characteristics modified trajectories of psychological change (Table 5). Higher baseline perceived stress was associated with larger reductions in perceived stress at 12 months (time × baseline PSS-10: β=−0.4, 95% CI −0.6 to −0.3; *P*<.001) and greater improvement in well-being (time × baseline PSS-10 for WHO-5: β=0.6, 95% CI 0.2-1.1; *P*=.007). Higher baseline well-being was associated with smaller improvement in well-being (time × baseline WHO-5: β=−0.2, 95% CI −0.3 to −0.1; *P*<.001), and higher baseline life satisfaction was associated with smaller gains in life satisfaction (time × baseline SWLS: β=−0.4, 95% CI −0.5 to −0.2; *P*<.001). Baseline illness perceptions did not modify trajectories of well-being (time × baseline B-IPQ: β=0.13, 95% CI −0.10 to 0.37; *P*=.30) or perceived stress (time × baseline B-IPQ for PSS-10: β=−0.06, 95% CI −0.15 to 0.04; *P*=.20). Full interaction outputs are provided in Table S10 in Multimedia Appendix 1.

XSL•FO
RenderX

Table . Baseline psychological moderators of longitudinal change in psychological outcomes (baseline to 12 months).[a]

Baseline psychological characteristic	Outcome	Direction of moderation	Interaction β (95% CI)	*P* value
Perceived stress (PSS-10[b])	WHO-5[c]	Higher baseline stress associated with greater improvement in well-being	0.6 (0.2 to 1.1)	.007
	PSS-10	Higher baseline stress associated with larger reductions in perceived stress	−0.4 (−0.6 to −0.3)	<.001
Illness perceptions (B-IPQ[d])	WHO-5	No evidence of effect modification	0.13 (−0.10 to 0.37)	.30
	PSS-10	No evidence of effect modification	−0.06 (−0.15 to 0.04)	.20
Psychological well-being (WHO-5)	WHO-5	Higher baseline well-being associated with smaller gains	−0.2 (−0.3 to −0.1)	<.001
Life satisfaction (SWLS[e])	SWLS	Higher baseline life satisfaction associated with smaller gains	−0.4 (−0.5 to −0.2)	<.001

[a]β coefficients represent time × baseline characteristic interaction terms from adjusted generalized estimating equation models, reflecting differential change from baseline to 12 months. Models were adjusted for age, sex, relationship status, Charlson Comorbidity Index, and psychotropic medication use. Only prespecified psychological moderators with stable and interpretable interaction effects are shown.

[b]PSS-10: Perceived Stress Scale.

[c]WHO-5: World Health Organization-Five Well-Being Index.

[d]B-IPQ: Brief Illness Perception Questionnaire.

[e]SWLS: Satisfaction With Life Scale.

Discussion

Principal Findings

This secondary analysis of PictouCountyPEP found sustained improvements across multiple psychological, behavioral, and physical health domains over 12 months in rural adults living with chronic conditions. Improvements were observed across all primary psychological outcomes and were accompanied by favorable changes in fatigue, sleep, diet quality, sedentary behavior, spiritual well-being, and mental health–related quality of life. Exploratory analyses further suggested that psychological improvements co-occurred with changes in fatigue, sleep, illness perceptions, social support, and spiritual well-being, pointing to plausible multidomain pathways through which digital self-management interventions may influence well-being over time.

Exploratory analyses further showed that improvements in several candidate factors, notably fatigue, sleep quality, illness perceptions, social support, and spiritual well-being, were associated with improvements in psychological outcomes. These patterns of co-occurring improvement may help identify candidate multidomain pathways for future investigation but should not be interpreted as evidence of statistical mediation or causal mechanisms. Establishing whether these factors mediate intervention-related psychological outcomes will require an appropriate temporal and causal analytic framework, ideally within a controlled study designed prospectively to test these pathways.

Social support and binge drinking frequency were less clearly improved at 12 months, and improvements in diet quality were somewhat attenuated, but the overall pattern remained favorable. Binary threshold and responder analyses were broadly concordant with the continuous outcome models. Obesity status was the main exception: although the odds of meeting the binary obesity threshold did not change, continuous BMI and weight outcomes improved, as did responder analyses based on clinically meaningful weight loss. Program engagement, adherence, and acceptability have been reported elsewhere and are therefore not repeated here [20].

These findings contribute to the growing literature on digital health interventions for chronic disease self-management in community settings [53]. Importantly, this study evaluated the program under real-world delivery conditions rather than within a tightly controlled clinical trial environment. Participants accessed the intervention remotely using their own devices, engagement occurred asynchronously, and optional components were used flexibly. These characteristics provide context for interpreting the observed outcomes and for future evaluation of feasibility, accessibility, and acceptability in routine community settings [53]. Such implementation-oriented evidence is particularly relevant for digital health interventions intended to reach geographically dispersed populations [20,53].

The sustained psychological improvements observed at 12 months contrast with several recent systematic reviews and meta-analyses of digital interventions, which commonly report attenuation of mental health effects over time. For example, digital behavioral activation interventions have shown reductions in depressive symptoms at shorter follow-up, but no significant

XSL•FO
RenderX

pooled benefit at 12 months; pooled anxiety outcomes were not significantly improved at the assessed follow-up points [54]. Digital peer support interventions have similarly demonstrated weaker mental health effects with longer follow-up [55], and digital exercise interventions have shown improvements at the end of the intervention that are not consistently maintained at later follow-up [56]. By contrast, psychological improvements in PictouCountyPEP remained evident at 12 months, during the lower-intensity maintenance period. Although direct comparisons across interventions and follow-up designs should be made cautiously, these findings suggest that comprehensive, multicomponent approaches warrant further evaluation for their potential to support longer-term psychological benefits. These findings complement evidence that interventions targeting multiple health behaviors can support chronic disease management [11] and indicate a need to evaluate their longer-term psychological outcomes.

The durability of effects may also reflect intervention design choices intended to reduce common barriers to digital engagement in rural settings, including low-bandwidth delivery, brief daily content, and minimal onboarding requirements [57]. Although engagement mechanisms were not directly tested in the present study, this microdose format may support habit formation and continued participation among individuals managing fluctuating symptoms and competing life demands [58]. These findings are also relevant to the well-described mismatch between health care systems historically oriented toward acute, reactive care and the proactive, ongoing support required for effective chronic disease self-management [59,60]. Many determinants most closely linked to symptom burden and progression unfold between clinical encounters, including sleep disruption, fatigue, stress, dietary patterns, social connection, and illness perceptions [60]. In this context, PictouCountyPEP may function as a low-burden adjunct that supports daily self-regulation and patient capability in routine living, complementing rather than replacing biomedical care.

Exploratory moderation analyses addressed an important question for digital self-management programs: for whom do these interventions appear to work best? Participants with higher baseline perceived stress experienced larger reductions in stress and greater improvements in well-being over 12 months, whereas those with higher baseline well-being or life satisfaction showed smaller gains in those same domains. This pattern suggests that the program may confer greater benefit among individuals with greater baseline psychological need. We found no consistent evidence that demographic characteristics substantially altered psychological trajectories, although these exploratory analyses may have been underpowered to detect smaller subgroup differences.

The exploratory mechanistic analyses point to several plausible pathways through which psychological improvement may occur. One pathway may involve modification of upstream cognitive processes, including illness perceptions that shape coping and behavioral responses, consistent with Leventhal's common-sense model of self-regulation [61]. Another may reflect increased engagement in meaningful, reinforced daily activities supported by social connection and sustained routines, aligning with behavioral activation frameworks [62]. More broadly, the observed pattern is consistent with interdependent changes across energy, sleep, cognition, and social resources that collectively support psychological recovery [61-63]. These findings extend earlier work on Prostate Cancer Patient Empowerment Program (PC-PEP), in which self-efficacy and illness perceptions explained improvements in psychological distress [64]. However, these analyses are exploratory and should not be interpreted as evidence of causal mediation. Future studies should formally test these candidate pathways using longitudinal mediation models or experimental designs.

Although this study was not designed to assess disease incidence, the improvements observed across stress, sleep, diet quality, sedentary behavior, and social support are relevant to secondary prevention in chronic disease. These domains represent potentially modifiable targets relevant to chronic disease self-management and health [60,65]. Future controlled studies with longer follow-up should examine whether these changes translate into downstream clinical outcomes, multimorbidity accumulation, or health service use.

Beyond the observed health improvements, PictouCountyPEP builds on a digital delivery model already implemented in PC-PEP across Canada and internationally. The program delivers core content through automated email prompts and brief asynchronous video modules, limiting the need for clinicians to deliver individual sessions. An international phase 4 implementation trial of PC-PEP reported similar follow-up completion across rural and urban participants, high program acceptability, and sustained improvements in mental health [66]. This implementation experience supports the scalability of the delivery model and provides a foundation for extending it to rural chronic disease self-management. Broader implementation of PictouCountyPEP should build on this experience while accounting for local digital access, staffing arrangements, and integration with existing services [57,67].

From a health system perspective, scalable self-management programs that can be delivered at low cost and without reliance on specialized personnel are of increasing interest, particularly in rural and resource-constrained settings [66,68]. An economic evaluation within the PC-PEP randomized controlled trial estimated program delivery costs of CAD $200.07 (CAD $1=US $0.73 as of August 29, 2025) per participant and found that early delivery improved outcomes while lowering overall health care costs [68]. These findings provide an economic rationale for broader implementation of the delivery model. A formal cost-effectiveness analysis of PictouCountyPEP from the health care system perspective is currently underway to assess downstream resource use and longer-term value. These findings also have implications for how digital self-management programs could be integrated into chronic disease care pathways. In many health systems, patients with multimorbidity receive episodic clinical care but limited ongoing behavioral or psychosocial support between visits [59,60]. Digital empowerment programs may complement traditional care by providing continuous, low-intensity behavioral guidance that patients can access independently. For primary care teams serving rural populations, such programs could function as an adjunct resource that reinforces lifestyle recommendations, supports stress regulation, and promotes patient activation

XSL•FO
RenderX

without substantially increasing clinical workload. Work is currently underway in collaboration with Nova Scotia Health to explore potential implementation pathways for the program within existing care structures. Together with the ongoing cost-effectiveness analysis, these efforts aim to determine whether digitally delivered self-management programs can offer a sustainable and cost-efficient complement to traditional chronic disease care models.

Nursing and Primary Care Implications

Although PictouCountyPEP was not designed as a nurse-delivered intervention, its asynchronous, home-based structure has potential implications for nursing and primary care practice in rural settings. Nurse-led remote digital approaches have been used to support adults with chronic conditions through health education, remote support, monitoring, and other forms of digitally enabled care [69]. More broadly, digital interventions are increasingly being evaluated as adjuncts to chronic disease self-management, although their effectiveness varies across conditions, intervention components, and self-care domains [65]. Rather than requiring nurses to deliver the intervention content directly, a program of this type could provide a standardized adjunct to health education and self-management support, with nurses and other primary care professionals potentially identifying appropriate patients, reinforcing individualized self-management goals, recognizing emerging needs during routine encounters, and directing patients to additional clinical or community services when required. This model may be particularly relevant in settings where clinical time and access to specialized chronic disease services are constrained because the core educational and behavioral content is delivered outside the clinical encounter and does not require continuous clinician involvement [69].

These implementation questions are particularly relevant in the present rural context: at baseline, 19 of 182 (10.4%) participants reported receiving primary care from another provider such as a nurse practitioner, while 46 of 182 (25.3%) participants reported having no regular primary care provider (Table S3 in Multimedia Appendix 1). Importantly, however, these potential nursing and primary care roles were not evaluated in the present study. Future implementation research should examine how referral to digital self-management programs can be incorporated into rural primary care and nursing workflows, including which patients are most appropriate for referral, whether some patients require additional adherence or digital navigation support, and whether integration can extend self-management support without increasing clinical workload.

Limitations

This study has several limitations. First, the study was conducted in a single, predominantly rural region, and the sample was predominantly female (85.2%) and White (96.2%), limiting generalizability to men and more ethnoculturally diverse populations. Eligibility also required English proficiency and access to the internet and email; consequently, the findings cannot be assumed to generalize to non–English-speaking populations or to individuals experiencing greater barriers to digital access or with lower digital health literacy. Although the intervention was intentionally designed for low-bandwidth delivery and did not require specialized applications, wearable devices, or synchronous participation, these design features do not eliminate potential digital access or digital literacy barriers. Generalizability to other rural regions may also depend on local sociodemographic characteristics, health care access, digital infrastructure, and community resources [57,67]. Similarly, the findings should not be assumed to apply directly to urban populations, where health service availability, referral pathways, and opportunities for in-person chronic disease support may differ. Evaluation in more diverse populations and geographic settings is therefore needed. Second, the single-arm design precludes causal attribution of the observed longitudinal changes to PictouCountyPEP. Without a concurrent control group, changes associated with natural disease trajectories, secular or time effects, regression to the mean, unmeasured co-interventions, and residual confounding cannot be separated from changes occurring during program participation. Although covariate adjustment and sensitivity analyses strengthen the assessment of longitudinal associations, they do not overcome this design limitation. Future randomized controlled trials should evaluate the efficacy of PictouCountyPEP relative to an appropriate comparator and determine whether the observed multidomain changes are attributable to the intervention. Third, all outcomes were self-reported and may therefore be subject to recall, reporting, and social desirability biases. The study also did not assess longer-term clinical or health service outcomes, such as hospitalization or mortality; consequently, the observed changes in patient-reported outcomes should not be interpreted as evidence of effects on these outcomes. Fourth, participants with incomplete follow-up had lower baseline well-being and life satisfaction, raising the possibility of attrition-related bias. However, none of the included variables was significantly associated with incomplete follow-up in the adjusted model, and the primary findings were consistent in complete case and weighted sensitivity analyses. Fifth, because multiple outcomes and exploratory analyses were examined without formal adjustment for multiple comparisons, some statistically significant findings may reflect chance. Accordingly, secondary and exploratory findings should be interpreted with caution, with emphasis on consistency and effect size rather than statistical significance alone. Finally, responder analyses were only possible for measures with available thresholds for meaningful change; for some outcomes, such thresholds remain poorly defined in general chronic disease populations. Future research should prioritize the development and validation of clinically meaningful benchmarks across relevant outcomes in chronic disease settings. Notwithstanding these limitations, strengths of the study include its comprehensive multidomain scope, use of validated instruments, prospective longitudinal design, and analytic approach incorporating all available observed outcome data, supported by complete case and inverse probability–weighted sensitivity analyses.

Conclusion

PictouCountyPEP is a digitally delivered self-management program associated with sustained improvements across multiple psychological, behavioral, and physical health domains in rural adults living with chronic conditions. These findings suggest that low-burden, multicomponent digital interventions may

XSL•FO
RenderX

support sustained improvements in mental well-being alongside favorable changes in related behavioral and psychosocial outcomes. The program's low-bandwidth, asynchronous delivery model highlights its potential scalability and relevance for rural and underserved populations, where access to ongoing behavioral and psychosocial support is limited. By addressing shared, upstream determinants of chronic disease across conditions, digitally delivered self-management programs may complement traditional care by extending support into patients' daily lives without increasing clinical workload. Future research should evaluate long-term clinical and health system outcomes, including health service use and cost-effectiveness, and prospectively test the candidate pathways suggested by these exploratory analyses to inform intervention optimization and implementation at scale.

Generative AI Use

No generative AI tools were used in the preparation, writing, or editing of this manuscript.

Funding

This study was funded by the Aberdeen Health Foundation and the Faculty of Medicine at Dalhousie University. The funders had no role in study design, conduct, or reporting.

Data Availability

Deidentified data will be made available upon request to the corresponding author, per institutional policies, and a data access process in compliance with patient privacy and protection research acts (Nova Scotia Health Authority Research Ethics Board).

Authors' Contributions

Conceptualization: NKS, GI, RDHR
Data curation: NKS, GI, CM
Formal analysis: NKS, GI, CM
Funding acquisition: GI, RDHR
Investigation: NKS, GI, CM, RDHR, JRBL
Methodology: NKS, GI, CM, RDHR
Project administration: GI, CM, RDHR, JRBL
Resources: GI, RDHR
Software: NKS, GI, RDHR
Supervision: GI, RDHR
Validation: NKS, GI, RDHR, CM
Visualization: NKS, JRBL
Writing – original draft: NKS, GI
Writing – review & editing: NKS, GI, RDHR, JRBL

No generative artificial intelligence (AI) tools were used in the preparation, writing, or editing of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Tables depicting the internal consistency of instruments at baseline, baseline characteristics by follow-up status, additional baseline characteristics, follow-up status for each outcome at 6 and 12 months, model-estimated marginal means by time point for continuous outcomes, longitudinal change from baseline, results of responder analyses, associations between changes in candidate factors and changes in outcomes from baseline to 6 months and 12 months, and effect modification by baseline psychological characteristics.
[DOCX File, 84 KB - mhealth_v14i1e95283_app1.docx]

References

1. A glimpse at the health of canadians. Statistics Canada. 2023. URL: https://www.statcan.gc.ca/o1/en/plus/5102-glimpse-health-canadians [accessed 2025-10-22]
2. Noncommunicable diseases. World Health Organization. 2025. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2025-12-22]
3. Hacker K. The burden of chronic disease. Mayo Clin Proc Innov Qual Outcomes 2024 Feb;8(1):112-119. [doi: 10.1016/j.mayocpiqo.2023.08.005] [Medline: 38304166]
4. Akif A, Qusar MMAS, Islam MR. The impact of chronic diseases on mental health: an overview and recommendations for care programs. Curr Psychiatry Rep 2024 Jul;26(7):394-404. [doi: 10.1007/s11920-024-01510-7] [Medline: 38767815]

XSL•FO
RenderX

5. Zu B, Wang N, Fan L, et al. A study on the impact of chronic diseases and depressive symptoms comorbidity on the risk of cognitive impairment in middle-aged and older adults people based on the CHARLS database. Front Public Health 2025;13:1558430. [doi: 10.3389/fpubh.2025.1558430] [Medline: 40084208]
6. The Daily - Key findings from the Health of Canadians report, 2024. Statistics Canada. 2025. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/250305/dq250305a-eng.htm [accessed 2025-10-22]
7. Kirby JB, Zuvekas SH, Borsky AE, Ngo-Metzger Q. Rural residents with mental health needs have fewer care visits than urban counterparts. Health Aff (Millwood) 2019 Dec;38(12):2057-2060. [doi: 10.1377/hlthaff.2019.00369] [Medline: 31794321]
8. Diabetes Canada Clinical Practice Guidelines Expert Committee, Sherifali D, Berard LD, Gucciardi E, MacDonald B, MacNeill G. Self-management education and support. Can J Diabetes 2018 Apr;42 Suppl 1:S36-S41. [doi: 10.1016/j.jcjd.2017.10.006] [Medline: 29650109]
9. Huang Y, Li S, Lu X, Chen W, Zhang Y. The effect of self-management on patients with chronic diseases: a systematic review and meta-analysis. Healthcare 2024;12(21):2151. [doi: 10.3390/healthcare12212151]
10. Farris O, Orbell S, Lamarche VM, Smith R. Promoting self-management in chronic disease: a systematic review and meta-analysis of behaviour change interventions for patients on dialysis. Health Psychol Rev 2025 Jun;19(2):368-408. [doi: 10.1080/17437199.2025.2471775] [Medline: 40035473]
11. Silva CC, Presseau J, van Allen Z, et al. Effectiveness of interventions for changing more than one behavior at a time to manage chronic conditions: a systematic review and meta-analysis. Ann Behav Med 2024 May 23;58(6):432-444. [doi: 10.1093/abm/kaae021] [Medline: 38721982]
12. Sunol R, González-González AI, Valli C, et al. Self-management interventions for adults living with obesity to improve patient-relevant outcomes: an evidence map. Patient Educ Couns 2023 May;110:107647. [doi: 10.1016/j.pec.2023.107647] [Medline: 36739705]
13. Lin C, Zhu X, Wang X, et al. The impact of perceived social support on chronic disease self-management among older inpatients in China: the chain-mediating roles of psychological resilience and health empowerment. BMC Geriatr 2025 Apr 26;25(1):284. [doi: 10.1186/s12877-025-05902-z] [Medline: 40287628]
14. Miller CK, Young-Hyman D, Wooldridge JS, Piatt GA, Hawkins J. Goal integration of diabetes self-management education and psychosocial support with pharmacotherapy for weight loss in adults with type 2 diabetes and obesity. Clin Diabetes 2025;43(4):562-569. [doi: 10.2337/cd25-0020] [Medline: 41140718]
15. Maita KC, Maniaci MJ, Haider CR, et al. The impact of digital health solutions on bridging the health care gap in rural areas: a scoping review. Perm J 2024 Sep 16;28(3):130-143. [doi: 10.7812/TPP/23.134] [Medline: 39135461]
16. Snoswell CL, Chelberg G, De Guzman KR, et al. The clinical effectiveness of telehealth: a systematic review of meta-analyses from 2010 to 2019. J Telemed Telecare 2023 Oct;29(9):669-684. [doi: 10.1177/1357633X211022907] [Medline: 34184580]
17. Shomali M, Mora P, Aleppo G, et al. The critical elements of digital health in diabetes and cardiometabolic care. Front Endocrinol (Lausanne) 2024;15:1469471. [doi: 10.3389/fendo.2024.1469471] [Medline: 39351525]
18. Domhardt M, Mennel V, Angerer F, et al. Processes of change in digital interventions for depression: a meta-analytic review of cognitive and behavioral mediators. Behav Res Ther 2025 Jun;189:104735. [doi: 10.1016/j.brat.2025.104735] [Medline: 40222276]
19. Angerer F, Mennel V, Grund S, et al. Mechanisms of change in digital interventions for depression: a systematic review and meta-analysis of six mediator domains. J Affect Disord 2025 Jan 1;368:615-632. [doi: 10.1016/j.jad.2024.09.055] [Medline: 39284530]
20. Ilie G, Murphy S, MacDonald C, Rutledge RDH. Digital delivery of a 6-month home-based empowerment program improves mental health in rural adults with chronic conditions: a single-arm prospective interventional trial. Digit Health 2026;12:20552076251413351. [doi: 10.1177/20552076251413351] [Medline: 41585304]
21. Ilie G, Rendon R, Mason R, et al. A Comprehensive 6-mo Prostate Cancer Patient Empowerment Program Decreases Psychological Distress Among Men Undergoing Curative Prostate Cancer Treatment: A Randomized Clinical Trial. Eur Urol 2023 Jun;83(6):561-570. [doi: 10.1016/j.eururo.2023.02.009] [Medline: 36822969]
22. Ilie G, Knapp G, Davidson A, et al. The Cancer Patient Empowerment Program: A Comprehensive Approach to Reducing Psychological Distress in Cancer Survivors, with Insights from a Mixed-Model Analysis, Including Implications for Breast Cancer Patients. Cancers (Basel) 2024 Oct 2;16(19):3373. [doi: 10.3390/cancers16193373] [Medline: 39409993]
23. Census Profile, 2021 Census of Population. Statistics Canada. 2023. URL: https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/index.cfm?Lang=E [accessed 2026-03-06]
24. Warburton DER, Jamnik VK, Bredin SSD, Gledhill N. The physical activity readiness questionnaire for everyone (PAR-Q+) and electronic physical activity readiness medical examination (ePARmed-X+). Health Fit J Can 2011;4(2):3-17. [doi: 10.14288/hfjc.v4i2.103]
25. Lawrence CE, Dunkel L, McEver M, et al. A REDCap-based model for electronic consent (eConsent): moving toward a more personalized consent. J Clin Transl Sci 2020 Apr 3;4(4):345-353. [doi: 10.1017/cts.2020.30] [Medline: 33244416]
26. Canada's food guide: healthy eating recommendations. Government of Canada. 2019. URL: https://www.canada.ca/en/health-canada/services/food-guide/explore/healthy-eating-recommendations.html [accessed 2025-11-19]

XSL•FO
RenderX

27. Des Jarlais D, Lyles C, Crepaz N. TREND statement checklist. CDC Stacks. 2004. URL: https://stacks.cdc.gov/view/cdc/149677
28. Broadbent E, Petrie KJ, Main J, Weinman J. The brief illness perception questionnaire. J Psychosom Res 2006 Jun;60(6):631-637. [doi: 10.1016/j.jpsychores.2005.10.020] [Medline: 16731240]
29. Schwartz MN, Rimland CA, Quinn KA, et al. Utility of the brief illness perception questionnaire to monitor patient beliefs in systemic vasculitis. J Rheumatol 2020 Dec 1;47(12):1785-1792. [doi: 10.3899/jrheum.190828] [Medline: 32238516]
30. The World Health Organization-Five Well-Being Index (WHO-5). World Health Organization. 2024. URL: https://www.who.int/publications/m/item/WHO-UCN-MSD-MHE-2024.01 [accessed 2026-10-23]
31. Domenech A, Kasujee I, Koscielny V, Griffiths CEM. Systematic review of the use of the WHO-5 well-being index across different disease areas. Adv Ther 2025 Aug;42(8):3657-3677. [doi: 10.1007/s12325-025-03266-9] [Medline: 40506676]
32. Yılmaz Koğar E, Koğar H. A systematic review and meta-analytic confirmatory factor analysis of the perceived stress scale (PSS-10 and PSS-14). Stress Health 2024 Feb;40(1):e3285. [doi: 10.1002/smi.3285] [Medline: 37341705]
33. Plantinga LC, Bowling CB, Hoge C, et al. Physical performance in a diverse, population-based cohort of individuals with systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2024 Apr;76(4):517-525. [doi: 10.1002/acr.25266] [Medline: 37885120]
34. Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. J Pers Assess 1985 Feb;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
35. Diener E. Understanding scores on the Satisfaction with Life Scale. 2006 URL: https://labs.psychology.illinois.edu/~ediener/Documents/Understanding%20SWLS%20Scores.pdf [accessed 2026-10-23]
36. Book K, Marten-Mittag B, Henrich G, et al. Distress screening in oncology-evaluation of the Questionnaire on Distress in Cancer Patients-short form (QSC-R10) in a German sample. Psychooncology 2011 Mar;20(3):287-293. [doi: 10.1002/pon.1821] [Medline: 20669340]
37. Wharton S, Lau DCW, Vallis M, et al. Obesity in adults: a clinical practice guideline. CMAJ 2020 Aug 4;192(31):E875-E891. [doi: 10.1503/cmaj.191707] [Medline: 32753461]
38. Foley P, Steinberg D, Levine E, et al. Track: a randomized controlled trial of a digital health obesity treatment intervention for medically vulnerable primary care patients. Contemp Clin Trials 2016 May;48:12-20. [doi: 10.1016/j.cct.2016.03.006] [Medline: 26995281]
39. Ware JE, Kosinski M, Keller SD. A 12-item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34(3):220-233. [doi: 10.1097/00005650-199603000-00003]
40. Shah CH, Brown JD. Reliability and validity of the Short-Form 12 item version 2 (SF-12v2) health-related quality of life survey and disutilities associated with relevant conditions in the US older adult population. J Clin Med 2020 Feb 29;9(3):661. [doi: 10.3390/jcm9030661] [Medline: 32121371]
41. Joly F, Kosinski M, Shafer F, Wardecki M, Karaouni A, Hill Q. Clinically important change in SF-12v2 physical (PCS) and mental (MCS) component summary scores for patients with cold agglutinin disease: an analysis using the phase 3 CARDINAL and CADENZA studies. Blood 2021 Nov 5;138(Supplement 1):2002. [doi: 10.1182/blood-2021-147759]
42. Cella D, de la Loge C, Fofana F, et al. The Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-Fatigue) scale in patients with axial spondyloarthritis: psychometric properties and clinically meaningful thresholds for interpretation. J Patient Rep Outcomes 2024 Aug 12;8(1):92. [doi: 10.1186/s41687-024-00769-x] [Medline: 39133438]
43. Scippa K, Mangrum R, Martin S, et al. Establishing content validity of the FACIT-fatigue scale in patients living with warm autoimmune hemolytic anemia: results from qualitative interview studies. Blood 2023 Nov 2;142(Supplement 1):3753. [doi: 10.1182/blood-2023-174965]
44. Buysse DJ, Reynolds CF III, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
45. Sansano-Nadal O, Wilson JJ, Martín-Borràs C, et al. Validity of the sedentary behavior questionnaire in European older adults using English, Spanish, German and Danish versions. Meas Phys Educ Exerc Sci 2022 Jan 2;26(1):1-14. [doi: 10.1080/1091367X.2021.1922910]
46. Motuma A, Gobena T, Roba KT, Berhane Y, Worku A. Sedentary behavior and associated factors among working adults in eastern Ethiopia. Front Public Health 2021;9:693176. [doi: 10.3389/fpubh.2021.693176] [Medline: 34557467]
47. Dao-Tran TH, Lam LT, Balasooriya NN, Comans T. The Medical Outcome Study Social Support Survey (MOS-SSS): a psychometric systematic review. J Adv Nurs 2023 Dec;79(12):4521-4541. [doi: 10.1111/jan.15786] [Medline: 37449790]
48. Segal-Isaacson CJ, Wylie-Rosett J, Gans KM. Validation of a short dietary assessment questionnaire: the Rapid Eating and Activity Assessment for Participants short version (REAP-S). Diabetes Educ 2004;30(5):774-781. [doi: 10.1177/014572170403000512] [Medline: 15510530]
49. Peterman AH, Fitchett G, Brady MJ, Hernandez L, Cella D. Measuring spiritual well-being in people with cancer: the functional assessment of chronic illness therapy—Spiritual Well-being Scale (FACIT-Sp). Ann Behav Med 2002;24(1):49-58. [doi: 10.1207/S15324796ABM2401_06] [Medline: 12008794]
50. Understanding binge drinking. National Institute on Alcohol Abuse and Alcoholism (NIAAA). URL: https://www.niaaa.nih.gov/publications/brochures-and-fact-sheets/binge-drinking [accessed 2025-12-18]

XSL•FO
RenderX

51. Pasta DJ. Learning when to be discrete: continuous vs categorical predictors. SAS Global Forum. 2009 Jan 10 URL: https://support.sas.com/resources/papers/proceedings09/248-2009.pdf
52. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2025. URL: https://www.R-project.org/ [accessed 2026-10-23]
53. Pong C, Tseng R, Tham YC, Lum E. Current implementation of digital health in chronic disease management: scoping review. J Med Internet Res 2024 Dec 12;26:e53576. [doi: 10.2196/53576] [Medline: 39666972]
54. Jia E, Macon J, Doering M, Abraham J. Effectiveness of digital behavioral activation interventions for depression and anxiety: systematic review and meta-analysis. J Med Internet Res 2025 Jun 17;27:e68054. [doi: 10.2196/68054] [Medline: 40526910]
55. Yeo G, Fortuna KL, Lansford JE, Rudolph KD. The effects of digital peer support interventions on physical and mental health: a review and meta-analysis. Epidemiol Psychiatr Sci 2025 Feb 13;34:e9. [doi: 10.1017/S2045796024000854] [Medline: 39945388]
56. Zangger G, Bricca A, Liaghat B, et al. Benefits and harms of digital health interventions promoting physical activity in people with chronic conditions: systematic review and meta-analysis. J Med Internet Res 2023 Jul 6;25:e46439. [doi: 10.2196/46439] [Medline: 37410534]
57. Sahakyan S, Akseer S, Abrahamyan L, et al. Barriers and facilitators to implementing digital health technologies for remote management of NCDs in rural areas: mixed methods systematic review. J Med Internet Res 2026 Jul 31;28:e93875. [doi: 10.2196/93875] [Medline: 42536988]
58. Singh B, Murphy A, Maher C, Smith AE. Time to form a habit: a systematic review and meta-analysis of health behaviour habit formation and its determinants. Healthcare 2024;12(23):2488. [doi: 10.3390/healthcare12232488]
59. Grudniewicz A, Gray CS, Boeckxstaens P, De Maeseneer J, Mold J. Operationalizing the chronic care model with goal-oriented care. Patient 2023 Nov;16(6):569-578. [doi: 10.1007/s40271-023-00645-8] [Medline: 37642918]
60. Allegrante JP, Wells MT, Peterson JC. Interventions to support behavioral self-management of chronic diseases. Annu Rev Public Health 2019 Apr 1;40:127-146. [doi: 10.1146/annurev-publhealth-040218-044008] [Medline: 30601717]
61. Hagger MS, Orbell S. The common sense model of illness self-regulation: a conceptual review and proposed extended model. Health Psychol Rev 2022 Sep;16(3):347-377. [doi: 10.1080/17437199.2021.1878050] [Medline: 33461402]
62. Santopetro N, Jones D, Garron A, Meyer A, Joyner K, Hajcak G. Examining a fully automated mobile-based behavioral activation intervention in depression: randomized controlled trial. JMIR Ment Health 2024 Aug 30;11:e54252. [doi: 10.2196/54252] [Medline: 39212598]
63. Scott AJ, Webb TL, Martyn-St James M, Rowse G, Weich S. Improving sleep quality leads to better mental health: a meta-analysis of randomised controlled trials. Sleep Med Rev 2021 Dec;60:101556. [doi: 10.1016/j.smrv.2021.101556] [Medline: 34607184]
64. MacDonald C, Ilie G, Kephart G, et al. Mediating effects of self-efficacy and illness perceptions on mental health in men with localized prostate cancer: a secondary analysis of the Prostate Cancer Patient Empowerment Program (PC-PEP) randomized controlled trial. Cancers (Basel) 2024 Jun 27;16(13):2352. [doi: 10.3390/cancers16132352] [Medline: 39001414]
65. Longhini J, Pedrotti D, Foladori F, et al. Effectiveness of digital health interventions to improve self-care in patients with chronic diseases: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2026 Jun 9;28:e88708. [doi: 10.2196/88708] [Medline: 42263266]
66. MacNevin W, Smith NK, Ilie G, et al. Equitable reach, engagement, and mental health outcomes of a Digital Prostate Cancer Survivorship Program across rural and urban settings: results from an international phase 4 trial. JU Open Plus 2026;4(6):e00057. [doi: 10.1097/JU9.0000000000000458]
67. MacAskill W, Gill P, Woloszczuk C, et al. A systematic review of the scope and impact of rural primary healthcare innovations using digital health technology. BMJ Open 2026 Jan 29;16(1):e097874. [doi: 10.1136/bmjopen-2024-097874] [Medline: 41611458]
68. Nuyens A, Ilie G, Rendon RA, et al. Evaluating the cost-effectiveness of the Prostate Cancer Patient Empowerment Program: a comprehensive health economic analysis from a randomized controlled trial. Can Urol Assoc J 2025 Dec;19(12):410-419. [doi: 10.5489/cuaj.9222] [Medline: 40893098]
69. Kilfoy A, Chu C, Krisnagopal A, et al. Nurse-led remote digital support for adults with chronic conditions: a systematic synthesis without meta-analysis. J Clin Nurs 2025 Mar;34(3):715-736. [doi: 10.1111/jocn.17226] [Medline: 38894583]

Abbreviations

B-IPQ: Brief Illness Perception Questionnaire
FACIT-F: Functional Assessment of Chronic Illness Therapy—Fatigue scale
FACIT-SP-12: 12-item Functional Assessment of Chronic Illness Therapy—Spiritual Well-Being scale
GEE: generalized estimating equation
MCS: Mental Component Summary (of the 12-item Short Form Health Survey)
MOS-SSS: Medical Outcomes Study Social Support Survey
PAR-Q+: Physical Activity Readiness Questionnaire for Everyone

XSL•FO
RenderX

PC-PEP: Prostate Cancer Patient Empowerment Program
PCS: Physical Component Summary (of the 12-item Short Form Health Survey)
PictouCountyPEP: Pictou County Patient Empowerment Program
PIPEDA: Personal Information Protection and Electronic Documents Act
PSQI: Pittsburgh Sleep Quality Index
PSS-10: 10-item Perceived Stress Scale
QSC-R10: 10-item Questionnaire on Distress in Cancer Patients
REAP-S: Rapid Eating Assessment for Participants—Shortened version
REB: Research Ethics Board
SBQ: Sedentary Behavior Questionnaire
SF-12: 12-item Short Form Health Survey
SWLS: Satisfaction With Life Scale
TREND: Transparent Reporting of Evaluations with Nonrandomized Designs
WHO-5: World Health Organization-Five Well-Being Index

Edited by S Brini; submitted 19.Mar.2026; peer-reviewed by A Ramadas, X Wu; revised version received 07.Oct.2026; accepted 08.Oct.2026; published 30.Oct.2026.

Please cite as:
Smith NK, Ilie G, MacDonald C, Langley JRB, Rutledge RDH
Multidomain Outcomes of a Digital Self-Management Program in Rural Adults With Chronic Conditions: Single-Arm Prospective Interventional Study
JMIR Mhealth Uhealth 2026;14:e95283
URL: https://mhealth.jmir.org/2026/1/e95283
doi:10.2196/95283

© Nathan K Smith, Gabriela Ilie, Cody MacDonald, James Richard Ballentyne Langley, Robert David Harold Rutledge. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 30.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluating City-Cut Geographic Risk Classification for a Pandemic Travel-Record System in Shanghai, China: Retrospective Observational Analysis

Rowan Wang[1*], PhD; Baofeng Shi[2*], PhD; Shenghan Li[3*], BMgt; Cong Xie[4*], PhD; Qingpeng Zhang[5,6*], PhD

[1]College of Business, Southern University of Science and Technology, Shenzhen, Guangdong, China

[2]College of Economics and Management, Northwest A&F University, Xianyang, Shaanxi, China

[3]Carlson School of Management, University of Minnesota Twin Cities, Minneapolis, MN, United States

[4]China Economic Research Institute, Liaoning University, Shenyang, Liaoning, China

[5]Department of Pharmacology and Pharmacy, LKS Faculty of Medicine, The University of Hong Kong, Pokfulam Road, Hong Kong, China (Hong Kong)

[6]Musketeers Foundation Institute of Data Science, The University of Hong Kong, Hong Kong, China (Hong Kong)

[*]all authors contributed equally

Corresponding Author:
Qingpeng Zhang, PhD
Department of Pharmacology and Pharmacy, LKS Faculty of Medicine, The University of Hong Kong, Pokfulam Road, Hong Kong, China (Hong Kong)

Abstract

Background: Smartphone-delivered travel-record and health-code systems were deployed across China during the COVID-19 pandemic to classify risk based on travel history and regulate access to venues and travel. The geographic classification rule is a key design choice for these systems. A city-boundary rule (city-cut) is administratively convenient but may group areas with different mobility patterns under one risk label, whereas a fixed-radius rule (radius-cut) offers greater spatial precision but may exclude more distant areas connected by routine travel. Whether a classification rule aligns with observed mobility is a design question distinct from evaluating a deployed system's epidemiological effects.

Objective: This study examined whether observed mobility in Shanghai supported a city-cut or a radius-cut geographic design for a mobile health travel-record system and whether mobility-derived communities corresponded to administrative districts.

Methods: We conducted a retrospective observational analysis of 1.44 billion deidentified China Unicom mobility records collected for two 46-day periods in 2019 and 2020. The analytical unit was the daily origin-area movement proportion. We calculated within-Shanghai, confirmed external-city, and nonadjacent-district movement proportions; for each origin area, we paired them by calendar day. We constructed grid-level transition matrices, used spectral clustering followed by k-means++, and checked the groupings across repeated initializations and selected dates.

Results: Across 46 daily city-average observations in 2019, the intracity proportion was 97.2% (95% *t* CI 97.1% - 97.3%). Nonadjacent-district movement was 13.7% (95% *t* CI 13.5% - 13.9%), compared with 2.8% (95% *t* CI 2.7% - 2.9%) for confirmed external-city movement. A paired, 2-tailed Newey-West heteroskedasticity- and autocorrelation-consistent (HAC) *t* test (α=.05) gave a mean difference of 10.9 (95% HAC CI 10.7 - 11.1; *P*<.001) percentage points. At the Chongming border, the corresponding difference was –3.1 (95% HAC CI –4.1 to –2.1; *P*<.001) percentage points.

Conclusions: This study moves the question from postdeployment evaluation to predeployment design, treating geographic calibration as an upstream mobile health decision. It differs from prior work that assessed uptake, privacy, or epidemiological outcomes after fielding, and from mobility research that treated movement data as a predictor, an outcome, or a model input rather than as a criterion for choosing a boundary rule. The contribution is threefold: it formulates a boundary-fit question that can be answered before a system is built; it provides a reproducible, auditable workflow (daily origin-area proportions, a paired HAC-based comparison, and a mobility-clustering check) that agencies and developers can rerun on their own data; and it demonstrates the workflow from end to end in a large municipality. Practically, a city-level unit is a defensible starting reference where movement is predominantly internal, whereas border areas warrant separate local assessment. Agencies should document where a boundary fits local movement and where exceptions may be needed. Decisions about restrictions also require epidemiological evidence and assessment of proportionality, public acceptability, and social costs.

(JMIR Mhealth Uhealth 2026;14:e64941) doi:10.2196/64941

XSL•FO
RenderX

KEYWORDS

COVID-19; human mobility; nonpharmaceutical intervention; epidemic modeling; epidemic control

Introduction

During the COVID-19 pandemic in mainland China, health-code and travel-record systems used smartphone-delivered information to regulate access to venues and travel [1-3]. These systems made the choice of geographic unit consequential: a city-level risk label could affect people whose recorded travel history included that city. Location-based digital health tools also raised concerns about privacy, consent, and governance [1,4-9]. A central design question is the spatial scale at which a travel-history system should assign a risk label. An administrative city boundary is simple to communicate and implement, whereas a fixed-radius rule is geographically narrower [1-3]. The best fit depends partly on how people move across and within the relevant boundaries. A broad label can encompass places with different movement patterns, while a narrow label can divide locations connected by routine travel. The design task is therefore to make the geographic assumption explicit and assess its consistency with observed connectivity before using it to classify travel histories.

We examine this question using mobility data from Shanghai, China. Shanghai is a large and internally heterogeneous municipality with 16 districts [10]. The city provides a useful comparison between a city boundary and a fixed-radius rule: residents may live far from an index location within Shanghai, while some locations just beyond the municipal boundary may be geographically close. This setting makes it necessary to distinguish geographic distance from membership in the same administrative unit. Examining the city as a whole, alongside areas near its municipal boundary, can reveal whether an aggregate pattern obscures local differences. The internal district structure also provides an administrative reference against which to consider functional mobility regions. We evaluate both designs against observed mobility without assuming that either applies universally.

First, we ask whether observed movement within Shanghai is more frequent than confirmed movement to other cities, including for border-origin areas. Second, we ask whether movement between nonadjacent Shanghai districts is substantial and more frequent than confirmed external-city movement, including for border-origin areas. We refer to these as the intracity and nonadjacent-district comparisons. The first concerns how much recorded movement remains within the proposed city unit. The second concerns connectivity across more widely separated parts of that unit, which is relevant to whether a geographically narrow label would capture the scale of routine movement. Border-origin areas are especially informative because proximity to another jurisdiction may weaken the fit of a uniform municipal label. Both comparisons describe mobility; neither tests infection transmission nor intervention effectiveness. These questions connect work on digital public-health systems with research on mobility and epidemic spread [11].

Location-based digital public-health tools, including digital contact tracing, geofenced quarantine monitoring, and travel-record systems, became prominent components of COVID-19 governance [12-14]. Mobile health (mHealth) scholarship identifies client registration, tracking of vital events, and data collection as core application types [15]. A scoping review distinguishes contact tracing and location-based risk assessment from population-level outbreak detection and mobility tracking [14]. This distinction helps locate the present question within digital public health: the geographic rule used to summarize a travel history should be evaluated in relation to that function, without treating it as a measure of individual contact or infection. Prior evaluations have examined adoption and epidemiological outcomes after deployment [7,16-19], while work on health-code systems has emphasized their operational and governance implications [1-3,6,20,21]. Studies in this literature show that design and governance shape how digital systems operate [6,8,9], but seldom test the geographic unit that determines who will be flagged before the system is deployed. Selecting the boundary before deployment differs from estimating whether a deployed tool changed infection outcomes. The question is whether a proposed spatial boundary fits observed mobility at the scale the system is intended to manage.

Large-scale mobile-phone datasets have been used to characterize human mobility during the COVID-19 pandemic, study the effects of movement restrictions, and inform public-health action [11,22-31]. Related research has examined spatial connectivity and the role of movement in infectious-disease models [32-36]. Work on the public-health use of mobile-phone data also emphasizes the need to match data and analysis to the decision being supported, while addressing access, coordination, and privacy constraints [26]. For a travel-record system, the relevant design question is whether the chosen unit captures observed connections within its boundary and distinguishes them from connections across it. This application requires an explicit link between a mobility measure and the geographic classification decision it is intended to inform. These studies generally treat mobility as a predictor, an outcome, or a model input rather than as evidence for choosing the geographic boundary of a digital travel-record system.

Network studies also show that interaction data can delineate functional regions that may differ from administrative borders [37,38]. Such functional regions summarize interaction patterns and provide a way to examine whether administrative divisions align with those patterns [37,38]. Alignment would support their use as an operational reference, whereas divergence would identify places where classification or coordination may require closer attention. Neither outcome alone establishes an appropriate restriction because a mobility partition and a public-health decision serve different purposes. Community-detection and spectral-clustering methods [39-42] can inform boundary design: mobility-based clustering can be compared with official geography, while aggregate movement shares can measure movement within and across a proposed

XSL•FO
RenderX

unit. Mobility evidence cannot resolve legal, ethical, or epidemiological questions, but it can make the assumptions behind a geographic rule explicit.

To examine this design question, we use deidentified, aggregated Shanghai mobile-phone mobility records. We evaluate city-cut and radius-cut designs using daily intracity, confirmed external-city, border-origin, nonadjacent-district, and center-bound movement proportions. We also apply a prespecified spectral clustering procedure to assess how closely the resulting communities correspond to district boundaries. This approach evaluates geographic calibration before deployment and complements postdeployment evaluations of digital interventions. The study objectives are to assess whether within-city and nonadjacent-district mobility patterns in municipal-border areas support the use of a city-level geographic unit and to examine the correspondence between mobility-derived communities and administrative districts.

Methods

Study Context and Setting

This study evaluates the geographic design of China's antipandemic travel-record card system in the Shanghai setting. The record card was a smartphone-based travel-history credential that displayed cities visited during the previous 14 days and could trigger restrictions when a listed city was classified as medium or high risk. Figure 1 illustrates the card interface used to communicate these city-level risk labels. Shanghai was selected because it is a large, highly connected municipality bordered by several major neighboring cities, making it a useful case for comparing city-cut and radius-cut risk classification. Figure 2 shows Shanghai's district structure and surrounding cities, which define the geographic context for the mobility analysis.

Figure 1. Chinese antipandemic travel-record card. Source: author-generated illustrative screenshot based on the record-card interface; no personal identifiers are displayed.



XSL•FO
RenderX

Figure 2. Map of Shanghai and surrounding areas. Source: author-generated map based on Shanghai administrative boundary data.



Research Design

This was a retrospective observational secondary-data study. It involved no intervention allocation, randomization, or follow-up of individual participants. All conditions were naturally observed; the investigators did not manipulate mobility conditions or assign individuals or geographic areas to groups. The design comprised repeated daily observations during 2 prespecified 46-day windows in 2019 and 2020, as well as comparisons across naturally occurring geographic origin areas and movement categories.

Data Collection Procedures

China Unicom processed and deidentified the supplied operator records before author access was granted. The authors received aggregated records for the prespecified 2019 and 2020 study windows and applied the documented visit, movement, and daily origin-area aggregation procedures.

Inclusion and Exclusion

No participants were recruited. The provider preprocessed and deidentified the records before they were accessed by the authors. All supplied records within the study windows were used; the authors did not apply individual-level inclusion or exclusion criteria.

Unit of Analysis and Data Characteristics

The unit of analysis was a daily origin-area movement proportion, not an individual participant. The study used deidentified, aggregated movement records and did not collect

XSL•FO
RenderX

participant characteristics. The data did not contain individual names, phone numbers, device identifiers, or individual-level trajectories.

Sampling Procedures

The dataset was supplied by China Unicom for the specified study windows in Shanghai. It is an operator-based data source rather than a probability sample of Shanghai residents. All obtained records were retained after provider deidentification and the documented movement-construction procedures.

Sample Size, Power, and Precision

No a priori power analysis was performed because the analysis used all supplied aggregate records. Precision is reported using 2-tailed 95% *t* CIs for the mean of daily origin-area proportions across 46 days, based on day-to-day variation, with heteroskedasticity- and autocorrelation-consistent (HAC) intervals used as a serial-correlation-robustness analysis.

Measures and Covariates

The prespecified mobility measures were intracity movement, confirmed external-city movement, nonadjacent-district movement, center-bound movement, and border-origin status. Administrative-district, center-area, and municipal-border labels were assigned through the grid crosswalk. No participant-level covariates were available or analyzed.

Data Quality and Missing Data

No unavailable origin or destination codes were used in the analysis. State 9999 represented confirmed external-city journeys rather than missing geography, and no imputation was performed.

We obtained proprietary, deidentified, aggregated mobile-phone signaling records from China Unicom, a major Chinese mobile service provider. The data provide a large operator-based mobility dataset, but we do not assume that China Unicom subscribers represent the overall Shanghai population. This potential coverage bias is addressed as a study limitation.

China Unicom's deidentified mobile-phone signaling records capture device connections to base stations, which were used to reconstruct aggregated daily visits and movement segments. The dataset covers all records supplied for Shanghai from October 10 to November 24, 2019, and the corresponding dates in 2020. The movement-location records used in the analysis did not include records with missing origin or destination codes. Confirmed movements outside Shanghai were instead represented by the external-city state (9999).

Based on suggestions from location-tracking experts at China Unicom, we used the following settings:

The map of Shanghai was divided into 2213 grid units of 2 km × 2 km. If a device remained continuously in a grid for more than 30 minutes, we counted this as a visit to that grid. This 30-minute threshold was used to exclude brief pass-through observations, such as travel by car or train, without a substantive stay. When a device had 2 consecutive visits (eg, grid A followed by grid B), we counted this as a movement from grid A to grid B. If a device left Shanghai and remained in another city for more than 30 minutes, we used its last observed grid in Shanghai (eg, grid A) and counted this as a movement from grid A to another city. This counting procedure was repeated for each calendar day.

If a device stayed in 1 grid (eg, grid A) throughout a calendar day, we counted this as local movement within grid A.

In total, the supplied dataset contained 1.44 billion observed movement segments. The main text reports the 2019 baseline results, and Multimedia Appendix 1 reports the parallel 2020 analysis.

Data preprocessing followed a clean-append-merge workflow implemented in Stata (StataCorp LLC). Outbound Shanghai journeys were aggregated by day and origin grid and coded with the confirmed external-city destination state (9999); inbound journeys were aggregated by day and destination grid and coded with the confirmed external-city origin state (9999). These records were appended to within-Shanghai grid-to-grid movements. The 30-minute active-population file was merged by day and origin grid, and movement probability was calculated as prob = wnumpop30, where wnum is the movement count and pop30 is the corresponding 30-minute active population. Each movement record carried an origin and a destination grid identifier (s_{fnid} and e_{fnid}), where *fnid* is the numeric code of a 2-km grid cell. The reserved value 9999 denotes the external-city state rather than a Shanghai grid. For inbound records originating outside Shanghai (s_{fnid}=9999), the external-state denominator was set using the archived population-scaling rule; outbound records leaving Shanghai (e_{fnid}=9999) retained the 30-minute active-population denominator of their Shanghai origin grid. For reproducibility, the fixed external-city denominator used for inbound records (s_{fnid}=9999) was constructed by scaling the total Shanghai grid active population by the ratio of the population outside Shanghai to the Shanghai population; the formula, parameter values, and reference date are provided in Multimedia Appendix 1. Finally, origin and destination grids were merged with the grid-to-administrative-region crosswalk and the center-area, border-band, and external-city labels.

Visit histories were processed for each calendar day. A stay of more than 30 minutes in a 2 km × 2 km grid was treated as a visit. Two consecutive visits generated 1 directed movement segment. A day with no grid change was counted as a local movement. A device observed to leave Shanghai and remain in another city for more than 30 minutes was recorded from its last Shanghai grid to a confirmed external-city state; this coding did not denote missing data.

Statistical Analysis

We defined 4 movement types: intra was movement within Shanghai, including local movement; inter was movement to the confirmed external-city state; intra ld was movement between nonadjacent Shanghai districts; and intra tc was movement toward one of the 7 central districts (Changning, Hongkou, Huangpu, Jing'an, Putuo, Xuhui, and Yangpu).

The analytical unit was the daily origin-area proportion. For every district, border-origin area, and city-average series, we reconstructed 46 calendar-day proportions and reported their

XSL•FO
RenderX

mean with a conventional 2-tailed 95% *t* CI, calculated from the same 46 daily proportions. As a robustness analysis for possible serial correlation across consecutive days, we also calculated Newey-West HAC intervals from an intercept-only model with lag 3, selected by the conventional rule 4(N100)29 for *N*=46 [43]. The HAC results were substantively consistent with the conventional 95% *t* CIs. Hypothesis comparisons used a paired, 2-tailed Newey-West HAC *t* test of the 46 daily differences, d_t=Intra ld_t-$Inter_t$, with null mean (d_t)=0 and α=.05. Movement segments were not treated as independent observations. Inference was based on the 46 ordered daily proportions; the HAC analysis allowed heteroskedasticity and autocorrelation through lag 3.

For a daily proportion series, the conventional interval was calculated as x− ± t0.975,45s46, where x- is the mean of the 46 daily proportions and s is their SD. It is therefore a conventional descriptive interval based on day-to-day variation, not an individual-movement sampling interval. The Newey-West HAC covariance estimator adjusts the estimated uncertainty for heteroskedasticity and autocovariances through lag 3, reducing reliance on independence of adjacent daily proportions; it was reported as a robustness analysis rather than as proof that all temporal dependence had been removed [43,44].

City-border areas were defined as grids within Shanghai's boundary districts that were within 8 km, equivalent to four 2-km grid units, of the municipal boundary. This definition is used consistently throughout the main analysis and in Multimedia Appendix 1. We then repeated the movement-proportion calculations for movements originating within these border areas.

The intracity long-distance metric is an operational measure of nonadjacent-district connectivity, not proof of a direct, nonstop trip. Because visits are recorded only after a stay of more than 30 minutes, a segment may include short, unrecorded pass-through travel, including through central areas. Conversely, a stay of more than 30 minutes in an intermediate grid splits a journey into separate segments.

Cluster Analysis

We explored whether the tested mobility-clustering specification produced an administratively interpretable subcity partition. This was an exploratory analysis: it assessed the correspondence of a specified clustering procedure to district geography, rather than adjudicating among all possible subcity classifications.

We used a spectral Markov-chain clustering approach [41,42]. The primary input was the October 10, 2019, Shanghai grid-level transition dataset, with 2213 Shanghai grid states plus the confirmed external-city state (9999). We summed repeated origin-destination edges, constructed a sparse off-diagonal transition matrix from observed probabilities, and assigned the residual probability to each diagonal stay entry. States with no observed off-diagonal movement were removed, and the reduced row-stochastic matrix was rebuilt before eigendecomposition. We computed 8 eigenvectors using eigs(T2,8,'largestabs') and clustered the real-valued second eigenvector. The primary specification used *k*=16. In an audited reimplementation, k-means++ initialization, Lloyd iterations (maximum 1000), 100 restarts per seeded fit, and 50 fixed seeds (5000 initializations) were used. We also examined a 46-day transition-probability average and selected daily matrices across the 2019 and 2020 windows. Repeated fits were stable to initialization, but the tested specifications showed only partial and spatially heterogeneous correspondence with districts; they did not exactly reproduce the district partition and do not rule out other subcity classifications.

Ethical Considerations

This study used secondary, deidentified, and aggregated mobile-phone signaling records provided by China Unicom under data-use and privacy restrictions. The authors did not recruit participants; interact with individuals; or access names, phone numbers, device identifiers, or individual-level trajectories. An institutional ethics review was not required for this secondary analysis because the dataset available to the authors had been deidentified before analysis, and all results are reported only as aggregated district-level, border-area, or grid-cluster summaries. No additional informed consent was obtained for this secondary analysis, and no compensation was provided because there was no participant recruitment or direct participation. The article, figures, and supplementary materials do not contain identifiable individual/user information.

Results

Unit of Analysis and Participant Flow

No participants were recruited or followed. The analytical flow comprised provider-supplied, deidentified movement records, daily origin-area aggregation, construction of 46 daily proportions for each series, and summary and paired comparison of those proportions.

Recruitment

The study used secondary deidentified aggregate data and involved no recruitment, enrollment, or participant compensation.

Statistics and Data Analysis

The exact inferential procedure, analytical unit, number of daily observations, pairing, sidedness, conventional 95% *t* CIs, and HAC robustness analysis are reported in the *Statistical Analysis* subsection. Tables 1 and 2 present the reconstructed means and conventional 95% *t* CIs.

XSL•FO
RenderX

Table . Movement proportions by Shanghai origin district, 2019.[a]

District	Intra[b] (%), mean (95% CI)	Inter[c] (%), mean (95% CI)	Intra ld[d] (%), mean (95% CI)	Intra tc[e] (%), mean (95% CI)
Baoshan	98.2 (98.2 - 98.3)	1.8 (1.7 - 1.8)	14.8 (14.5 - 15.1)	15.2 (14.8 - 15.7)
Changning	98.0 (97.8 - 98.1)	2.0 (1.9 - 2.2)	24.3 (24.1 - 24.5)	18.6 (18.5 - 18.8)
Chongming	93.7 (93.6 - 93.8)	6.3 (6.2 - 6.4)	11.2 (10.8 - 11.5)	3.9 (3.7 - 4.0)
Fengxian	99.0 (99.0 - 99.1)	1.0 (0.9 - 1.0)	4.7 (4.6 - 4.8)	3.3 (3.2 - 3.4)
Hongkou	98.5 (98.3 - 98.7)	1.5 (1.3 - 1.7)	15.1 (14.7 - 15.4)	19.3 (18.9 - 19.7)
Huangpu	98.5 (98.4 - 98.6)	1.5 (1.4 - 1.6)	23.4 (23.2 - 23.6)	18.0 (17.8 - 18.2)
Jiading	95.0 (94.9 - 95.0)	5.0 (5.0 - 5.1)	10.7 (10.6 - 10.8)	11.4 (11.2 - 11.7)
Jing'an	98.5 (98.4 - 98.6)	1.5 (1.4 - 1.6)	19.5 (19.3-19.7)	19.0 (18.7 - 19.2)
Jinshan	94.7 (94.6 - 94.8)	5.3 (5.2 - 5.4)	9.9 (9.7 - 10.1)	3.5 (3.3 - 3.6)
Minhang	97.6 (97.5 - 97.8)	2.4 (2.2 - 2.5)	10.7 (10.4 - 10.9)	13.1 (12.8 - 13.4)
Pudong	98.1 (98.0 - 98.1)	1.9 (1.9 - 2.0)	10.1 (9.9 - 10.2)	11.0 (10.7 - 11.2)
Putuo	96.4 (96.1 - 96.6)	3.6 (3.4 - 3.9)	21.5 (21.1 - 21.8)	18.6 (18.3 - 19.0)
Qingpu	91.8 (91.4 - 92.2)	8.2 (7.8 - 8.6)	13.3 (13.1 - 13.4)	8.2 (8.1 - 8.3)
Songjiang	98.2 (98.1 - 98.2)	1.8 (1.8 - 1.9)	11.4 (11.2 - 11.6)	7.1 (6.9 - 7.2)
Xuhui	98.7 (98.6 - 98.7)	1.3 (1.3 - 1.4)	17.2 (17.0 - 17.5)	19.0 (18.6 - 19.4)
Yangpu	99.0 (98.9 - 99.1)	1.0 (0.9 - 1.1)	18.1 (17.8 - 18.5)	18.8 (18.4 - 19.1)
City average	97.2 (97.1 - 97.3)	2.8 (2.7 - 2.9)	13.7 (13.5 - 13.9)	12.6 (12.4 - 12.9)

[a]Values are means (conventional 95% *t* CIs) of 46 daily origin-area proportions. Corresponding Newey-West heteroskedasticity- and autocorrelation-consistent CIs, which allow for serial correlation across days, were evaluated as a robustness analysis.

[b]Intra: movements within Shanghai, including local movements.

[c]Inter: confirmed movements to the external-city state.

[d]Intra ld: movements between nonadjacent Shanghai districts.

[e]Intra tc: movements toward the 7 central districts.

Table . Movement proportions by Shanghai municipal-border origin area, 2019.[a]

Area	Intra[b] (%), mean (95% CI)	Inter[c] (%), mean (95% CI)	Intra ld[d] (%), mean (95% CI)	Intra tc[e] (%), mean (95% CI)
Baoshan city border	92.0 (91.9 - 92.1)	8.0 (7.9 - 8.1)	10.7 (10.6 - 10.9)	5.3 (5.2 - 5.5)
Chongming city border	84.6 (84.3 - 84.9)	15.4 (15.1 - 15.7)	12.3 (11.9 - 12.7)	4.2 (4.0 - 4.5)
Jiading city border	91.4 (91.3 - 91.5)	8.6 (8.5 - 8.7)	10.3 (10.2 - 10.4)	8.9 (8.8 - 9.0)
Jinshan city border	93.3 (93.2 - 93.4)	6.7 (6.6 - 6.8)	11.4 (11.2 - 11.6)	4.1 (3.9 - 4.2)
Qingpu city border	90.7 (90.5 - 90.8)	9.3 (9.2 - 9.5)	10.3 (10.1 - 10.4)	5.8 (5.7 - 5.8)
City average	91.0 (90.9 - 91.1)	9.0 (8.9 - 9.1)	10.8 (10.6 - 11.0)	6.1 (6.0 - 6.2)

[a]Values are means (conventional 95% *t* CIs) of 46 daily origin-area proportions. Corresponding Newey-West heteroskedasticity- and autocorrelation-consistent CIs, which allow for serial correlation across days, were evaluated as a robustness analysis.

[b]Intra: movements within Shanghai, including local movements.

[c]Inter: confirmed movements to the external-city state.

[d]Intra ld: movements between nonadjacent Shanghai districts.

[e]Intra tc: movements toward the 7 central districts.

Ancillary Analyses

Ancillary analyses applied the same daily proportion workflow to a parallel 2020 window and examined the 46-day average and selected daily transition matrices. These analyses were used to assess the robustness of descriptive mobility and clustering patterns across dates.

Adverse Events

This secondary analysis involved no interventions, participant recruitment, or participant-facing study procedures.

XSL•FO
RenderX

Table 1 reports mean daily mobility proportions for each origin district, and Table 2 reports the same quantities for movements originating in the 8-km municipal border band. Parentheses contain conventional 95% *t* CIs, based on 46 daily proportions rather than SDs. HAC intervals using the same daily series were examined as a serial-correlation-robustness analysis. The long-distance measure captures nonadjacent-district connectivity at the grid-day resolution and should not be interpreted as evidence that every movement was nonstop.

Across the 16 districts in 2019, intracity movement proportions exceeded 90%; the city average was 97.2% (conventional 95% *t* CI 97.1% - 97.3%). The city-average intracity long-distance proportion was 13.7% (conventional 95% *t* CI 13.5% - 13.9%), compared with 2.8% (conventional 95% *t* CI 2.7% - 2.9%) for confirmed external-city movement.

For district-origin series, the paired mean difference between intracity long-distance and external-city movement was positive in every district and at the city average. The city-average difference was 10.9 (95% HAC CI 10.7 - 11.1; 2-tailed paired HAC *t* test $P<.001$) percentage points. The 2019 border-origin results were also positive except in Chongming, where external-city movement was more frequent than nonadjacent-district movement. The corresponding 2020 city-border average was not different from 0 (mean difference –0.1 percentage points, 95% HAC CI –0.3 to 0.1; $P=.52$), with negative differences in Chongming and Jiading. These exceptions qualify the citywide pattern and motivate local assessment near municipal boundaries.

The border-origin analysis is interpreted with deliberate caution. Chongming is geographically distinct, and its border-origin series had more confirmed external-city movement than nonadjacent-district movement in 2019. Therefore, border-area results do not support a uniform inference for every municipal edge.

Taken together, the movement results provide mobility-based evidence relevant to city-cut classification in Shanghai, while identifying boundary-area exceptions. They do not, by themselves, establish legal, ethical, social, epidemiological, or intervention-effectiveness legitimacy.

The same workflow was applied to the 2020 data; Multimedia Appendix 1 provides the 2020 point estimates and conventional 95% *t* CIs. The corresponding HAC analysis produced substantively consistent conclusions.

Mobility Communities and District Geography

Figure 3 presents the 16-cluster spectral partition for the October 10, 2019, transition matrix. Colors identify mobility communities, and district boundaries are overlaid for comparison. The map shows partial geographic correspondence in several peripheral areas but no exact one-to-one recovery of all district boundaries. Repeated-initialization and selected-date checks produced stable partitions, but this tested procedure does not identify a unique optimal subcity partition. The corresponding 46-day average map is provided in Figure S1 in Multimedia Appendix 1.

The corresponding 46-day average partitions for 2019 and 2020 are presented in Figures S1 and S2 in Multimedia Appendix 1.

The clustering results are exploratory evidence of functional connectivity rather than a ready-made policy boundary. They are most informative when used to compare the tested mobility partition with existing administrative geography.

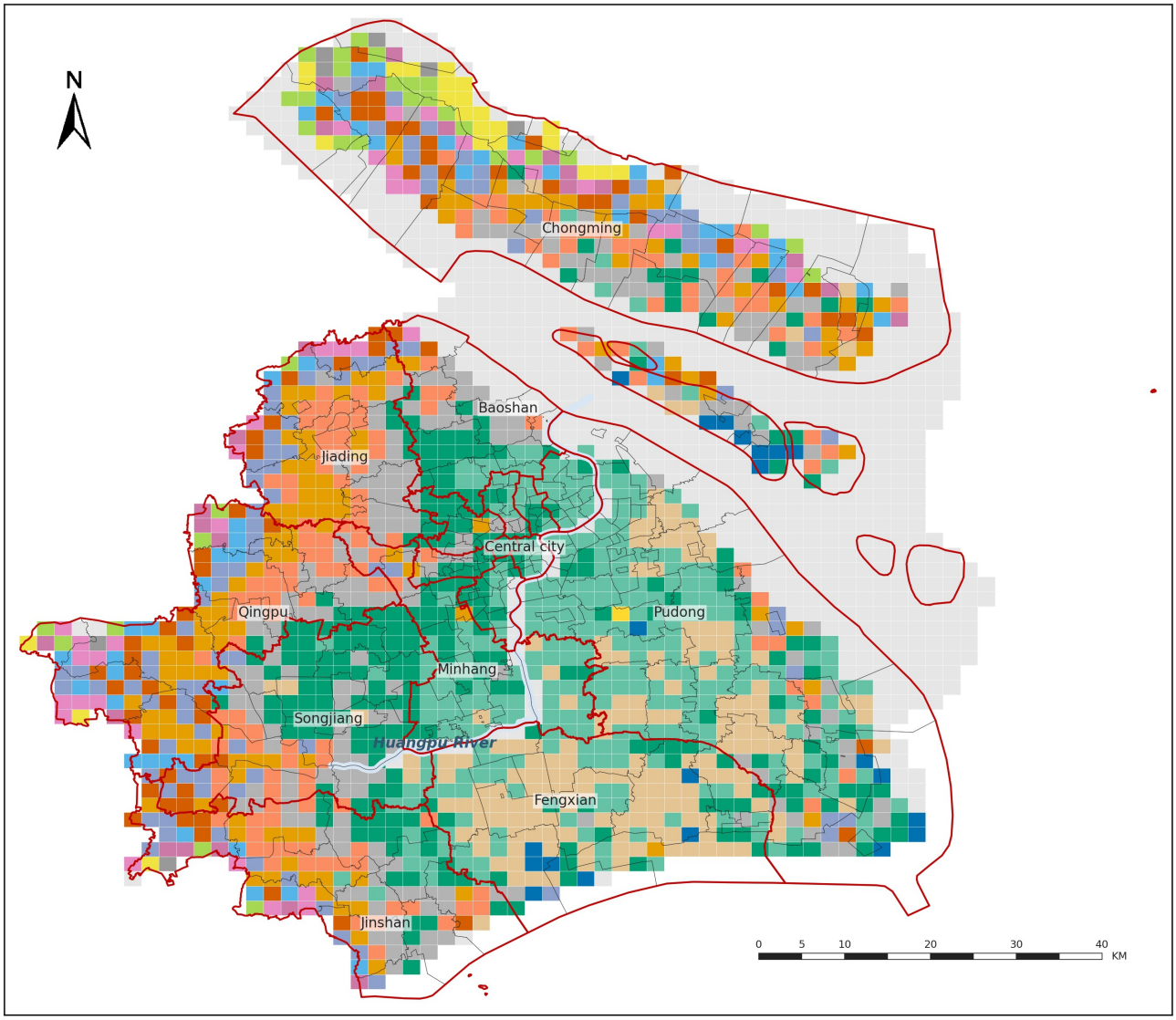
The single-day map shows broad, spatially coherent mobility regions, including partial correspondence with district geography in peripheral areas. In the 46-day average map (Figure S1 in Multimedia Appendix 1), several western areas remain visibly separated by district geography, whereas Fengxian and Pudong form a more integrated cross-district community, and the central urban area is more internally connected. These visual patterns identify places where cross-district coordination may be relevant; they do not confirm direct routes or infection transmission.

Some peripheral areas, including parts of Songjiang and Chongming, form more spatially concentrated blocks than the central city. Other blocks combine grids from multiple districts. This heterogeneity is consistent with districts being a practical first-order administrative reference in some settings, rather than a uniquely optimal or perfectly self-contained mobility partition.

The tested spectral-clustering specifications did not reproduce existing districts or simple district groupings as an exact map. This result does not rule out other subcity classifications, which would require purpose-built objectives, geographic constraints, and independent validation before policy use.

XSL•FO
RenderX

Figure 3. Spectral clustering of the October 10, 2019, Shanghai grid-level transition matrix (k=16). Colors identify mobility communities and district boundaries are overlaid. Source: authors' calculations from deidentified China Unicom mobility records and Shanghai grid and district-boundary data.



Discussion

Principal Findings

This study examined whether daily mobility patterns provide evidence relevant to the city-cut geographic design of a digital travel-record system. In 2019, movement within Shanghai was common across all origin districts, and the city-average nonadjacent-district movement proportion exceeded the confirmed external-city movement by 10.9 (95% HAC CI 10.7 - 11.1) percentage points. The 2020 analysis showed a similar district-level pattern, although border-origin comparisons were heterogeneous. These estimates concern observed mobility connectivity, not infection transmission or the effectiveness of a travel-restriction policy. The clustering analysis complements rather than overturns the city-level result. Under the reported spectral/Fiedler-vector/k-means specification, repeated fits produced stable mobility communities with partial and spatially heterogeneous correspondence to district geography. The 46-day average map retains some western district-oriented separation while showing tighter Fengxian-Pudong and central-area communities (Figure S1 in Multimedia Appendix 1). The clusters did not exactly reproduce the district map, but that finding neither rules out district-based management nor excludes other data-driven subcity classifications.

Comparison With Prior Work

Prior mobility research has used digital-trace data to characterize movement, model spatial transmission, or evaluate how restrictions changed mobility [11,22-24,27-36]. Our study differs by using observed mobility to examine an upstream system-design choice: the geographic unit used to flag travel histories. This distinction is important because a boundary rule determines the spatial scope of a digital intervention before effects on transmission or uptake can be measured. Network and community-detection studies show that interaction data can reveal functional regions that sometimes diverge from official boundaries [37,38]. Building on that work, we use grid-level transition probabilities to compare a city boundary with a radius-based alternative and to assess whether a tested subcity clustering procedure corresponds to administratively usable districts. The contribution is an empirical, reproducible boundary-calibration workflow rather than a model of disease spread or a claim that mobility alone determines a policy's

XSL•FO
RenderX

value. The results should not be read as evidence that mobility connectivity causes rapid transmission, that city-cut policies reduce infections, or that city-wide restrictions are proportionate. Those questions require linked epidemiological, behavioral, legal, and equity evidence [6,16-19]. The present study instead documents a mobility pattern that policymakers could consider alongside those other forms of evidence when defining geographic risk units.

Location-based mHealth tools were widely deployed during the COVID-19 pandemic, and evaluations of these systems have examined uptake, privacy practices, and epidemiological outcomes [7-9,16-19]. Our analysis contributes earlier in the design sequence by making the geographic classification rule empirically inspectable. It shows how aggregate mobility data can be used to test whether a proposed administrative boundary is broadly consistent with the movement scale the tool is intended to manage. For Shanghai, high within-city mobility and substantial nonadjacent-district movement provide a mobility-based rationale for considering a city-level unit, with clear caution for some border-origin areas. At a subcity scale, districts offer an administratively available first-order reference, but central and corridor-like connections may need cross-district coordination. This interpretation is deliberately narrower than a claim of policy effectiveness or broad legitimacy.

Real-World Implications

The most direct real-world application of this workflow is at the point where a location-based mHealth system is specified rather than after it has been fielded. Travel-record and health-code systems require a geographic unit before a single record can be classified, and that unit determines who will be flagged [1-3,20,21]. The analysis reported here shows that this choice can be treated as an explicit, testable parameter: daily origin-area movement proportions and a paired HAC comparison give a documented and reproducible basis for adopting or rejecting a candidate boundary. Development teams can therefore record the mobility assumptions behind a geographic rule in the same way they record data retention or consent settings, producing an auditable design trail that reviewers, regulators, and the public can inspect [8,9]. For public-health agencies, the results support treating a city-level boundary as a first-order operational reference in settings where observed movement is predominantly internal. In Shanghai in 2019, the city-average intracity proportion was 97.2% (95% *t* CI 97.1% - 97.3%) and nonadjacent-district movement exceeded confirmed external-city movement by 10.9 (95% HAC CI 10.7 - 11.1) percentage points, so the movement that a city-level rule is intended to govern occurred largely inside the boundary that the rule would draw. Agencies elsewhere can apply the same daily proportion construction to test whether that condition holds locally before adopting a comparable rule and can repeat the test as travel patterns change, rather than importing a boundary convention from another jurisdiction.

The border-origin results carry a more specific operational message. Comparisons for the 8-km municipal border band were heterogeneous, and Chongming showed a negative difference, meaning that confirmed external-city movement exceeded nonadjacent-district movement for that origin area. A uniform city label would misdescribe the movement of such areas. Practical responses include treating designated border zones as separate classification units, applying a lower-intensity label within them, or coordinating classification with the adjacent jurisdiction. The clustering results point the same way at a finer scale: mobility communities showed only partial correspondence with district geography, so central and corridor-like areas may warrant cross-district coordination even where a city-level unit is retained [37,38].

More importantly, the workflow is not specific to Shanghai, to a single operator's data, or to COVID-19. It requires only aggregated origin-destination counts, an activity denominator, and a candidate boundary, inputs that are available to operators and statistical agencies in many settings [25,26]. It can be applied to other administrative units, to other hazards for which movement-based risk labeling is proposed, and to periodic recalibration as travel patterns shift. Where comparable data cannot be obtained, the same reasoning can still be used qualitatively to state, and later test, the mobility assumption on which a proposed boundary depends. At last, these implications should be read within the study's evidentiary limits. The analysis describes observed mobility. It does not measure infections, evaluate whether a city-cut rule reduced transmission, or establish that any restriction was legally authorized, ethically proportionate, or publicly acceptable [16-19]. Mobility evidence is one input among several, and this workflow is offered as a way to make that one input explicit and checkable, not as a decision procedure for policy.

Strengths and Limitations

This study's strengths are the large deidentified operator dataset, the reconstruction of daily origin-area proportions, the parallel 2019 and 2020 analyses, and transparent operational definitions for the movement measures. The grid-level records allow movement to be examined across Shanghai while retaining the spatial detail needed to distinguish district and municipal-border patterns. Applying the same geographic definitions and daily aggregation procedures in both years provides a consistent basis for comparing the 2 study periods. The district-level and border-origin comparisons help reveal local differences that would be obscured by a citywide average, while the clustering analysis provides a complementary view of how observed connectivity relates to administrative geography. Documenting the visit threshold, movement categories, and transition-matrix construction makes the analytical choices explicit and supports the assessment of the approach in settings with comparable data.

Despite these strengths, several limitations should be acknowledged. First, reliance on 1 mobile service provider may introduce bias if its users differ from the broader population, despite China Unicom's substantial market share [35]. Second, the visit definition requires a stay of more than 30 minutes in a 2 km × 2 km grid. This threshold reduces pass-through noise but may miss shorter epidemiologically relevant encounters. Third, the nonadjacent-district metric captures connectivity at the resolution of recorded visits and may include unobserved pass-through travel. Fourth, this study provides a mobility-based empirical justification for city-cut classification, but it does not evaluate epidemiological effectiveness, legal authority, ethical

XSL•FO
RenderX

proportionality, public acceptability, or the social costs of restrictions. Future research should combine mobility data with epidemiological and policy-outcome data to assess these broader dimensions.

Conclusions

This study introduces a mobility-based approach to an upstream design choice in location-based mHealth: how to define the geographic unit used for travel-risk classification. Unlike studies that assess transmission or intervention outcomes after deployment, it uses daily grid-to-grid mobility to assess whether a proposed boundary is consistent with observed movement. In Shanghai, the results support considering a city-level unit as a mobility-based reference while identifying border-area and subcity exceptions that require local assessment. The approach can inform more transparent geographic calibration of digital public-health tools, but it does not establish infection effects, legal legitimacy, or the overall appropriateness of any restriction policy. The innovation of this work lies in relocating the analytic question from postdeployment evaluation to predeployment design. It differs from existing studies in 2 respects. Evaluations of deployed location-based mHealth tools have assessed uptake, privacy practices, and epidemiological outcomes [7-9,16-19], and prior mobility research has treated movement data as a predictor, an outcome, or a model input rather than as a criterion for choosing a boundary rule [11,22-24,27-36]. This study contributes to the field a reproducible and auditable calibration workflow, combining daily origin-area movement proportions with paired HAC inference and transition-matrix clustering that can be applied to any candidate geographic unit in any setting with comparable aggregated mobility data. In real-world terms, the findings indicate that where observed movement is predominantly internal, a city-level unit is a defensible starting reference for travel-risk classification, while border-origin areas such as Chongming require separate local assessment before a uniform city label is applied, and areas whose mobility communities span administrative lines may require cross-district coordination. For developers and public-health agencies, the practical implication is that a boundary rule can be stated, tested, and documented before deployment rather than justified retrospectively.

Generative AI Use

The authors declare the use of generative AI (GenAI) in the research and writing process. According to the Generative AI Delegation Taxonomy (GAIDeT; 2025), GenAI tools were used under full human supervision for code generation, algorithm development for data analysis, visualization, reproducibility testing, text generation, proofreading and editing, reformatting, and publication support. The GenAI tool used was OpenAI Codex (GPT-5). The authors retain full responsibility for the final manuscript. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes. The declaration was submitted under the collective responsibility of the authors.

The original data-processing and analytical code was developed manually by the authors. GenAI was subsequently used to assist with auxiliary code generation and the reproduction and verification of selected prespecified analyses, including the recalculation of summary estimates and CIs, sensitivity checks, figure and table preparation, language editing, and the preparation of draft responses. The tool had access only to deidentified, aggregated mobility files; no identifiable individual-level data were provided. The authors independently reviewed and verified all AI-assisted code and outputs and retained responsibility for the research design, data processing, statistical analysis, interpretation, and the final manuscript.

Funding

This research was supported by the Strategic Topic Fund (Grant No. STG2/P-705/24-R) of the Research Grants Council of Hong Kong, the HKU-100 Scholar Fund of The University of Hong Kong, and the Startup Research Grant of Liaoning University.

Data Availability

The data used in this study were obtained from China Unicom and are subject to data use agreements and privacy restrictions. Therefore, the data are not publicly available. Aggregated results supporting the findings of this study are included in the paper and supplementary materials. Researchers interested in accessing similar data may contact China Unicom directly, subject to its data access policies and approval procedures.

All mobility results are reported as aggregated district-level, border-area, or grid-cluster summaries. Analytical definitions, preprocessing steps, and construction of the transition matrix are described in the *Methods* section and Multimedia Appendix 1 to support reproducibility.

Conflicts of Interest

None declared.

Multimedia Appendix 1
This appendix reports the parallel 2020 analysis using the same 46-day daily proportion construction and conventional 95% *t* CIs. Corresponding heteroskedasticity- and autocorrelation-consistent analyses are reported as robustness checks.
[DOCX File, 2474 KB - mhealth_v14i1e64941_app1.docx]

XSL•FO
RenderX

References

1. Liang F. COVID-19 and health code: how digital platforms tackle the pandemic in China. Soc Media Soc 2020;6(3):2056305120947657. [doi: 10.1177/2056305120947657] [Medline: 34192023]
2. Wu J, Wang J, Nicholas S, Maitland E, Fan Q. Application of big data technology for COVID-19 prevention and control in China: lessons and recommendations. J Med Internet Res 2020 Oct 9;22(10):e21980. [doi: 10.2196/21980] [Medline: 33001836]
3. Mao Z, Yao H, Zou Q, Zhang W, Dong Y. Digital contact tracing based on a graph database algorithm for emergency management during the COVID-19 epidemic: case study. JMIR mHealth uHealth 2021 Jan 22;9(1):e26836. [doi: 10.2196/26836] [Medline: 33460389]
4. Whitelaw S, Mamas MA, Topol E, Van Spall HGC. Applications of digital technology in COVID-19 pandemic planning and response. Lancet Digit Health 2020 Aug;2(8):e435-e440. [doi: 10.1016/S2589-7500(20)30142-4] [Medline: 32835201]
5. Golinelli D, Boetto E, Carullo G, Nuzzolese AG, Landini MP, Fantini MP. Adoption of digital technologies in health care during the COVID-19 pandemic: systematic review of early scientific literature. J Med Internet Res 2020 Nov 6;22(11):e22280. [doi: 10.2196/22280] [Medline: 33079693]
6. Ienca M, Vayena E. On the responsible use of digital data to tackle the COVID-19 pandemic. Nat Med 2020 Apr;26(4):463-464. [doi: 10.1038/s41591-020-0832-5] [Medline: 32284619]
7. Grekousis G, Liu Y. Digital contact tracing, community uptake, and proximity awareness technology to fight COVID-19: a systematic review. Sustain Cities Soc 2021 Aug;71:102995. [doi: 10.1016/j.scs.2021.102995] [Medline: 34002124]
8. Jiang J, Zheng Z. Personal information protection and privacy policy compliance of health code apps in China: scale development and content analysis. JMIR mHealth uHealth 2023 Nov 14;11:e48714. [doi: 10.2196/48714] [Medline: 37990813]
9. Bardus M, Al Daccache M, Maalouf N, Al Sarih R, Elhajj IH. Data management and privacy policy of COVID-19 contact-tracing apps: systematic review and content analysis. JMIR mHealth uHealth 2022 Jul 12;10(7):e35195. [doi: 10.2196/35195] [Medline: 35709334]
10. Statistical table of administrative divisions of Shanghai (as of December 31, 2019) [Article in Chinese]. Shanghai Civil Affairs Bureau. 2020 Jan 7. URL: https://mzj.sh.gov.cn/MZ_zhuzhan116_0-2-8-15-58-77/20200519/MZ_zhuzhan116_43817.html [accessed 2026-09-19]
11. Kraemer MUG, Yang CH, Gutierrez B, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science 2020 May 1;368(6490):493-497. [doi: 10.1126/science.abb4218] [Medline: 32213647]
12. Ferretti L, Wymant C, Kendall M, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 May 8;368(6491):eabb6936. [doi: 10.1126/science.abb6936] [Medline: 32234805]
13. Budd J, Miller BS, Manning EM, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020 Aug;26(8):1183-1192. [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]
14. Wirth FN, Johns M, Meurers T, Prasser F. Citizen-centered mobile health apps collecting individual-level spatial data for infectious disease management: scoping review. JMIR mHealth uHealth 2020 Nov 10;8(11):e22594. [doi: 10.2196/22594] [Medline: 33074833]
15. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013 Aug;1(2):160-171. [doi: 10.9745/GHSP-D-13-00031] [Medline: 25276529]
16. Kendall M, Tsallis D, Wymant C, et al. Epidemiological impacts of the NHS COVID-19 app in England and Wales throughout its first year. Nat Commun 2023 Feb 22;14(1):858. [doi: 10.1038/s41467-023-36495-z] [Medline: 36813770]
17. Jeon S, Rainisch G, Harris AM, et al. Estimated cases averted by COVID-19 digital exposure notification, Pennsylvania, USA, November 8, 2020-January 2, 2021. Emerg Infect Dis 2023 Feb;29(2):426-430. [doi: 10.3201/eid2902.220959] [Medline: 36639132]
18. Wymant C, Ferretti L, Tsallis D, et al. The epidemiological impact of the NHS COVID-19 app. Nature 2021 Jun;594(7863):408-412. [doi: 10.1038/s41586-021-03606-z] [Medline: 33979832]
19. Liu M, Zhou S, Jin Q, Nishimura S, Ogihara A. Effectiveness, policy, and user acceptance of COVID-19 contact-tracing apps in the post-COVID-19 pandemic era: experience and comparative study. JMIR Public Health Surveill 2022 Oct 27;8(10):e40233. [doi: 10.2196/40233] [Medline: 36190741]
20. Ye Q, Zhou J, Wu H. Using information technology to manage the COVID-19 pandemic: development of a technical framework based on practical experience in China. JMIR Med Inform 2020 Jun 8;8(6):e19515. [doi: 10.2196/19515] [Medline: 32479411]
21. Nakamoto I, Wang S, Guo Y, Zhuang W. A QR code-based contact tracing framework for sustainable containment of COVID-19: evaluation of an approach to assist the return to normal activity. JMIR mHealth uHealth 2020 Sep 7;8(9):e22321. [doi: 10.2196/22321] [Medline: 32841151]
22. Chinazzi M, Davis JT, Ajelli M, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020 Apr 24;368(6489):395-400. [doi: 10.1126/science.aba9757] [Medline: 32144116]
23. Hu T, Wang S, She B, et al. Human mobility data in the COVID-19 pandemic: characteristics, applications, and challenges. Int J Digit Earth 2021;14(9):1126-1147. [doi: 10.1080/17538947.2021.1952324]

XSL•FO
RenderX

24. Xia J, Yin K, Yue Y, et al. Impact of human mobility on COVID-19 transmission according to mobility distance, location, and demographic factors in the Greater Bay Area of China: population-based study. JMIR Public Health Surveill 2023 Apr 26;9:e39588. [doi: 10.2196/39588] [Medline: 36848228]
25. Buckee CO, Balsari S, Chan J, et al. Aggregated mobility data could help fight COVID-19. Science 2020 Apr 10;368(6487):145-146. [doi: 10.1126/science.abb8021] [Medline: 32205458]
26. Oliver N, Lepri B, Sterly H, et al. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. Sci Adv 2020 Jun;6(23):eabc0764. [doi: 10.1126/sciadv.abc0764] [Medline: 32548274]
27. Ohsawa Y, Sun Y, Sekiguchi K, et al. Risk index of regional infection expansion of COVID-19: moving direction entropy study using mobility data and its application to Tokyo. JMIR Public Health Surveill 2024 Aug 21;10:e57742. [doi: 10.2196/57742] [Medline: 39037745]
28. Kishore K, Jaswal V, Verma M, Koushal V. Exploring the utility of Google Mobility Data during the COVID-19 pandemic in India: digital epidemiological analysis. JMIR Public Health Surveill 2021 Aug 30;7(8):e29957. [doi: 10.2196/29957] [Medline: 34174780]
29. Sahu KS, Dubin JA, Majowicz SE, Liu S, Morita PP. Revealing the mysteries of population mobility amid the COVID-19 pandemic in Canada: comparative analysis with Internet of Things-based thermostat data and Google Mobility insights. JMIR Public Health Surveill 2024 Mar 20;10:e46903. [doi: 10.2196/46903] [Medline: 38506901]
30. Huang Q, Kang YS. Mathematical modeling of COVID-19 control and prevention based on immigration population data in China: model development and validation. JMIR Public Health Surveill 2020 May 25;6(2):e18638. [doi: 10.2196/18638] [Medline: 32396132]
31. Kurita J, Iwasaki Y. Effect of long-distance domestic travel ban policies in Japan on COVID-19 outbreak dynamics during dominance of the ancestral strain: ex post facto retrospective observation study. Online J Public Health Inform 2024 Apr 22;16:e44931. [doi: 10.2196/44931] [Medline: 38648635]
32. Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. Proc Natl Acad Sci U S A 2009 Dec 22;106(51):21484-21489. [doi: 10.1073/pnas.0906910106] [Medline: 20018697]
33. Pullano G, Alvarez-Zuzek LG, Colizza V, Bansal S. Characterizing US spatial connectivity and implications for geographical disease dynamics and metapopulation modeling: longitudinal observational study. JMIR Public Health Surveill 2025 Feb 18;11:e64914. [doi: 10.2196/64914] [Medline: 39965190]
34. Wesolowski A, Eagle N, Tatem AJ, et al. Quantifying the impact of human mobility on malaria. Science 2012 Oct 12;338(6104):267-270. [doi: 10.1126/science.1223467] [Medline: 23066082]
35. Liu C, Holme P, Lehmann S, Yang W, Lu X. Nonrepresentativeness of human mobility data and its impact on modeling dynamics of the COVID-19 pandemic: systematic evaluation. JMIR Form Res 2024 Jun 28;8:e55013. [doi: 10.2196/55013] [Medline: 38941609]
36. Chu AM, Chan JN, Tsang JT, Tiwari A, So MK. Analyzing cross-country pandemic connectedness during COVID-19 using a spatial-temporal database: network analysis. JMIR Public Health Surveill 2021 Mar 29;7(3):e27317. [doi: 10.2196/27317] [Medline: 33711799]
37. Ratti C, Sobolevsky S, Calabrese F, et al. Redrawing the map of Great Britain from a network of human interactions. PLoS One 2010 Dec 8;5(12):e14248. [doi: 10.1371/journal.pone.0014248] [Medline: 21170390]
38. Sobolevsky S, Szell M, Campari R, Couronné T, Smoreda Z, Ratti C. Delineating geographical regions with networks of human interactions in an extensive set of countries. PLoS One 2013;8(12):e81707. [doi: 10.1371/journal.pone.0081707] [Medline: 24367490]
39. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech 2008 Oct 9;2008(10):P10008. [doi: 10.1088/1742-5468/2008/10/P10008]
40. von Luxburg U. A tutorial on spectral clustering. Stat Comput 2007 Dec;17(4):395-416. [doi: 10.1007/s11222-007-9033-z]
41. Fiedler M. Algebraic connectivity of graphs. Czech Math J 1973;23(2):298-305. [doi: 10.21136/CMJ.1973.101168]
42. Liu N, Stewart WJ. Markov chains and spectral clustering. In: Hummel KA, Hlavacs H, Gansterer W, editors. Performance Evaluation of Computer and Communication Systems. Milestones and Future Challenges. Lecture Notes in Computer Science, vol 6821: Springer; 2011:87-98. [doi: 10.1007/978-3-642-25575-5_8]
43. Newey WK, West KD. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 1987 May;55(3):703. [doi: 10.2307/1913610]
44. Newey WK, West KD. Automatic lag selection in covariance matrix estimation. Rev Econ Stud 1994 Oct 1;61(4):631-653. [doi: 10.2307/2297912]

Abbreviations

HAC: heteroskedasticity- and autocorrelation-consistent
mHealth: mobile health

XSL•FO
RenderX

Edited by S Brini; submitted 31.Jul.2024; peer-reviewed by A Girma, D Nalmpantis; revised version received 27.Sep.2026; accepted 30.Sep.2026; published 09.Nov.2026.

Please cite as:
Wang R, Shi B, Li S, Xie C, Zhang Q
Evaluating City-Cut Geographic Risk Classification for a Pandemic Travel-Record System in Shanghai, China: Retrospective Observational Analysis
JMIR Mhealth Uhealth 2026;14:e64941
URL: https://mhealth.jmir.org/2026/1/e64941
doi:10.2196/64941

© Rowan Wang, Baofeng Shi, Shenghan Li, Cong Xie, Qingpeng Zhang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 9.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Impact of Mobilization Facilitated by Wearable Device Enhanced Patient Monitoring/Electrophysiology Pod–Based Feedback on Postoperative Complications Following Colorectal Cancer Surgery: Randomized Controlled Trial

Yang Meng[1*], MS; Fengyan Fan[1*], BS; Yumei Ma[2*], MS; Nong Yan[3], MS; Huan Wang[1], MS; Zhen Zhang[1], MS; Yiting Wang[1], MS; Hailong Dong[1], MD, PhD; Huang Nie[1], MD, PhD

[1]Department of Anesthesiology and Perioperative Medicine, Fourth Military Medical University Xijing Hospital, No.127 Changle West Road, Xi'an, China

[2]Department of Anesthesiology, No.905 Hospital of PLA NAVY, Shanghai, China

[3]Mindray Medical International Ltd, Shenzhen, China

[*]these authors contributed equally

Corresponding Author:
Huang Nie, MD, PhD
Department of Anesthesiology and Perioperative Medicine, Fourth Military Medical University Xijing Hospital, No.127 Changle West Road, Xi'an, China

Abstract

Background: Enhanced recovery after surgery (ERAS) guidelines recommend early postoperative mobilization to reduce complications, but adherence is often suboptimal, highlighting the need for effective tools to monitor and encourage movement. The Mindray enhanced patient monitoring (ePM)/electrophysiology (ep) pod, capable of tracking activity, vital signs, sleep, and pain, offers high-precision postoperative monitoring and is well-suited for research on activity feedback.

Objective: The study aims to assess whether wearable device-based (ePM/ep pod) activity feedback could reduce postoperative complications within 30 days of colorectal cancer (CRC) surgery.

Methods: We conducted an open-label, evaluator-blind, randomized controlled trial involving patients aged ≥18 years scheduled for CRC surgery. Patients were randomized to a feedback group or a control group. Both groups were set the same target activity time postoperatively based on ERAS guidelines. The feedback group received real-time visual feedback of movement time daily through the ePM/ep pod device, while the control group did not receive feedback. The primary outcome was the comprehensive complication index (CCI) within postoperative 30 days. Secondary outcomes included daily activity time, pain Numeric Rating Scale scores for rest and movement during the first 3 postoperative days, length of stay, percentage of reaching the scheduled mobilization target, 30-day postoperative mortality rate, and the times of first exhaust and defecation.

Results: Two hundred thirty-nine patients were recruited between February 2023 and September 2023, with 216 randomized (n=108 for each group). There was no significant difference in CCI within 30 postoperative days between the control group (median CCI 0, range 0 - 20.90) and the activity feedback group (median CCI 0, range 0 - 12.20). The estimated mean difference was −0.59 (95% CI −3.56 to 2.38; *P*=.66). Sensitivity analysis excluding patients with low device compliance did not alter these findings. No significant differences between groups were found in daily activity time, length of hospital stay, or pain scores. Post hoc analysis revealed significant negative correlations between 30-day CCI and activity on the second day after operation (*r*=−0.166) and the third day after operation (POD3) (*r*=−0.264; *P*<.05 for both). Linear regression indicated that POD3 activity significantly reduced CCI (β=−.025; 95% CI −0.045 to −0.006; *P*=.01), with peak CCI reduction at 215 minutes of activity.

Conclusions: In the context of ERAS, this study found no evidence that activity stimulation based on feedback from the wearable device (ePM/ep pod) could reduce 30-day postoperative CCI in patients undergoing CRC surgery. However, the ePM/ep pod could accurately record daily activity duration, which may be negatively correlated with CCI on POD3.

Trial Registration: Chinese Clinical Trial Register ChiCTR2300068107; https://www.chictr.org.cn/showproj.html?proj=189756

(JMIR Mhealth Uhealth 2026;14:e70534) doi:10.2196/70534

KEYWORDS

wearable device; postoperative complications; ERAS; colorectal cancer; postoperative mobilization; enhanced recovery after surgery

XSL•FO
RenderX

Introduction

Colorectal cancer (CRC) is the second most common cancer in China and ranks third globally in mortality rate [1,2], which is an urgent health problem. Surgery is the first-line treatment of CRC. Unfortunately, 50% of patients undergoing surgery develop postoperative complications [3], leading to increased medical expenses, longer hospital stays, and delay of postoperative recovery [4]. Previous studies have shown that postoperative complications are an independent risk factor related to poor 5-year survival and the increased overall recurrence rate [5-7].

Enhanced recovery after surgery (ERAS) is committed to reducing surgical trauma and stress, decreasing postoperative morbidity, and accelerating postoperative recovery of patients [8-10]. ERAS guidelines emphasize the importance of early mobilization after surgery [11], which can reduce adverse outcomes such as pulmonary complications, decreased skeletal muscle strength, thromboembolism, and insulin resistance [12]. It was confirmed that ERAS was safe and effective in patients undergoing CRC surgery [13,14], and early mobilization after operation was the key to the success of ERAS [15]. However, more than half of the patients did not mobilize as recommended in the ERAS guidelines [16], which meant a reliable measurement tool was needed to not only accurately monitor the postoperative activity of patients, but also to increase their enthusiasm for mobilization. One of the possibilities is to use a wearable device [17,18].

The development of wearable devices allows objective measurements of, for example, activity steps or activity time, which can be used as a feedback tool to urge patients to achieve their intended mobilization goals [17-19]. A study found it was safe to use an activity tracker with feedback after gynecological operations and may increase patients' physical activity [20]. Several wearable devices (such as Apple Watch, Fitbit, and Garmin) that provide combined automatic and visual feedback have been identified as effective tools for cancer survivors to raise awareness of self-activity, stimulate behavior change, and improve physical activity [21,22]. However, the clinical effect on recovery or promoting movement after operation of the wearable devices has not been sufficiently evaluated in patients with CRC. The enhanced patient monitoring (ePM)/electrophysiology (ep) pod developed by Mindray supports activity time statistics, continuous vital sign monitoring, sleep monitoring, pain scores, and trend recording, has a high accuracy of postoperative activity monitoring [23], and is suitable for related research.

The present study aims to determine whether the activity incentive feedback based on a wearable device (ePM/ep pod) could reduce postoperative complications of patients with CRC within 30 days after their operations. We hypothesized that postoperative complications would be reduced in the activity feedback group compared to the usual postoperative care group.

Methods

Study Design and Patients

It was an open-label, evaluator-blind, randomized controlled study with 2 parallel groups.

Patients aged ≥18 years and scheduled for CRC surgery were included. The exclusion criteria were patients who (1) had American Society of Anesthesiologists (ASA) grade >III, (2) had preoperative immobility or inability to walk unaided affected by neural, muscular, or skeletal diseases, (3) had participated in other clinical studies in the past 3 months, (4) had planned reoperation within 30 days after the initial operation, and (5) had expected to be transferred to the intensive care unit after their operations.

Ethical Considerations

The Ethics Committee of the First Affiliated Hospital of the Air Force Military Medical University approved this randomized controlled trial (approval no. KY20222299-F-1), and the study protocol was registered with the Chinese Clinical Trial Registry (ChiCTR2300068107) on July 2, 2023. All procedures complied with the Declaration of Helsinki and institutional regulations (The Ethics Committee of the First Affiliated Hospital of the Air Force Military Medical University, Medical Ethics Committee Charter) [24]. Written informed consent was obtained from all participants prior to enrollment, which included detailed explanation of the study and their right to withdraw at any time. Upon registration, each patient was assigned a unique study number, and all subsequent data analyses were performed using a deidentified dataset to ensure participant privacy and confidentiality. No financial compensation was provided for participation.

Interventions

All patients in the trial were randomly assigned (1:1) to a feedback group or a control group. Both groups were set the same target activity time postoperatively and thus performed activities for 2 hours on the first day after operation (POD1), 4 hours on the second day after operation (POD2), and 6 hours on the third day after operation (POD3), according to the ERAS guidelines and our ERAS pathway for CRC [10]. All patients were fitted with a wireless wearable device that continuously recorded their daily ambulation durations (ePM 12M monitor placed at the bedside or ep pod monitor placed on their wrist or blood pressure pod monitor placed on their upper arm).

In the feedback group, patients received real-time visual feedback of the movement time every day through the display of the ePM. According to the wards' routine care based on ERAS guidelines, patients were reminded to engage in activities combined with ePM/ep pod feedback to meet predefined daily step goals and also informed about the gap between the active time and the target time.

In the control group, patients' movement time was set to not appear on the display of the ePM. Patients were not given feedback, and they were only encouraged to carry out activities according to the current care standards based on ERAS guidelines.

XSL•FO
RenderX

Prevention and Handling of Adverse Events

If the following incidents occur, activity supervision should be suspended immediately: participants experience pain intensity >4 (from 0 to -10), electrocardiogram abnormalities (signs of arrhythmia or acute ischemia), indoor air oxygen saturation <88%, systolic blood pressure <90 mmHg, symptoms of standing intolerance (ie, dizziness, nausea, and blurred vision), exercise weakness (ie, difficulty standing), etc. In the next scheduled mobilization, if the participant is considered to be in stable physical condition after clinical evaluation, a new mobilization attempt will be conducted. If serious adverse events such as falling out of bed or falling occur, in addition to suspending activity supervision, the competent physician should also be asked to evaluate the patient's injury and terminate the study if necessary. The researcher should immediately fill out the serious adverse event report and report it to the hospital ethics committee.

Randomization and Blinding

Patients were randomized using block random groups (sized blocks of 2 and 4) generated by specialized staff. All random assignments were made and sealed to an envelope by intervention staff using Statistical Analysis Software. The randomization process was performed by an independent anesthesiologist at the end of the surgical procedure, separated from the ward's nursing staff and outcome assessors to reduce potential bias.

It was impractical to blind the intervention staff in view of the intervening measure, but contrary to the outcome evaluators. To ensure blinding during outcome assessments in the surgical ward, a schedule was established to stagger the intervention and evaluation implementation. Ward staff were informed about the activity monitoring study but were not provided with any specific hypothetical details to prevent potential bias from influencing the study results. Given the intervention characteristics, blinding the patients was also deemed to be unfeasible. To minimize performance bias, patients were informed that we were evaluating 2 mobilization schemes on the understanding that neither scheme was expected to confer any potential advantage.

Data Collection

Preoperative baseline data were collected before an operation as follows: demographic data; preoperative medication and abdominal surgery history, and preoperative radiotherapy and chemotherapy history; vital signs; available preoperative examinations; and preoperative physical activity ability assessed by the International Physical Activity Questionnaire (IPAQ) [25]. Operation-related data collected on the day of surgery included surgical duration, intraoperative bleeding, site, type, and catheter placement status. Postoperative data were collected after each operation as follows: complications according to the Clavien-Dindo classification during hospitalization and 1 week, 2 weeks, and 1 month after the operation; the wearable device ep pod record of the activity time; digital pain Numeric Rating Scale (NRS) scores of rest and exercise; length of hospitalization; time of first exhaust and defecation; and opioids consumption; when discharged from hospital, patients in the intervention group were asked about concerns for feedback on activity time and whether they would supplement activities.

Outcome Measurement

Primary Outcome

The primary outcome was the comprehensive complication index (CCI) [26] within 30 days after surgery. The CCI measured the overall complication calculated for each patient, ranging from 0 (no complication) to 100 (death), including any minor complications instead of focusing on specific ones. It was thus an objective measure with comprehensive and accurate evaluation [27-29]. When calculating CCI, all postoperative complications were evaluated and classified based on the Clavien-Dindo classification, then converted into a weighted continuous scale including quantity and severity. Finally, the CCI value was evaluated using the CCI calculator available [30].

Secondary Outcomes

Daily activity time was set from 12 AM to 11:59 PM based on the data uploaded by ePM/ep pod until the 4th day or until discharge, whichever came first. The patient's compliance was considered to be low [31] on the condition that the patient wore it for less than 8 hours between 8 AM and 8 PM. The degree of pain for rest and exercise in the first 3 days after operation was scored by the NRS [32]. NRS used 0 - 10 points to represent different degrees of pain scored by patients according to their subjective feelings, 0 for no pain and 10 for severe pain. The length of stay was set as the actual length of stay after a patient's operation. The percentage of reaching the scheduled mobilization target was calculated. The patient's activity was considered to be not up to standard if the daily activity time after their operation did not reach the target time. The mortality rate of 30 days after operation was also counted. The first exhaust and defecating times were subject to the patient's own assessment, meaning that an obvious feeling of exhaust or defecation was reported.

Sample Size Calculation

The sample size calculation was based on the primary outcome measured by the CCI 30 days after operation. CCI represents the measurement data that follows a normal distribution, as detailed in the previous trials on postoperative complications [33,34]. The average postoperative CCI of CRC was 10 (SD 5). Assuming that the visual feedback of the movement time could reduce CCI by 25%, the average CCI of the feedback group was set to 7.5 (SD 5) [35]. Based on the above effect size, Power Analysis and Sample Size 2015 Software was used to calculate that a minimum of 172 patients (n=86 per group) would provide 90% power to detect a 25% decrease of CCI and an α of .05 (2-sided test), assuming the feedback group was more favorable compared to the control group [35]. To compensate for a potential 20% dropout rate, 216 patients were required to be included (n=108 per group).

Statistical Analysis

We followed the intention-to-treat principle analyzing all random patients in the baseline analysis. The analysis results of the study were based on the modified intention-to-treat

XSL•FO
RenderX

(MITT) set. Statistical analyses were performed using SPSS version 23.0. The 2-sided significance level was set at a *P* value <.05.

Continuous variables with independent access to mean and SD were initially subjected to the Kolmogorov-Smirnov test to assess their normality. Otherwise, the Shapiro-Wilk test was chosen. The Levene test was used for the analysis of homogeneity of variance. Data with skewed distributions are reported as medians with IQR, while data adhering to a normal distribution are given as means with SD. Categorical variables are expressed as frequencies, percentages, and ratios. For baseline comparisons, normally distributed continuous variables with comparable variance were analyzed using Student *t* test. In contrast, continuous variables with skewed distributions or uneven variance were analyzed with the Mann-Whitney *U* test. Categorical variables were compared using either the chi-square or Fisher exact tests.

The impact of the primary outcome CCI within 30 days after an operation was analyzed using the Student *t* test or the Mann-Whitney *U* test. Except for the MITT analysis, subjects with low device compliance (wearing ePM/ep pod for less than 8 h from 8 AM to 8 PM) were excluded for sensitivity analysis of the primary outcome. In addition, subgroup analyses (preselected age, ASA, preoperative chemoradiotherapy, surgical site, and preoperative mobility) were carried out. Simultaneously, we examined the interaction between the treatment effect and these 5 subgroup variables. In addition, we statistically described the distribution of postoperative complications at all levels in the 2 groups and performed a chi-square test between the groups. As a post hoc analysis, the Spearman/point-biserial correlation coefficient was used to determine the correlation between CCI and the postoperative movement time. In the preliminary analysis, the linear regression model showed a good fitting effect. Based on this result, we conducted a linear regression analysis to assess the relationship between the movement time of POD3 (for patients discharged on POD3, the activity time on POD2 was used to fill in the data) and CCI, adjusted for potential factors such as age, ASA, sex, preoperative chemoradiotherapy, and BMI, all factors which could influence postoperative recovery.

In analyses of secondary outcomes, the Mann-Whitney *U* or Student *t* tests were used to analyze the other measurement data except for repeated measurements. The chi-square test was used to analyze the percentage of patients who achieved the predetermined mobilization goal. Binary logistic regression analysis was used to determine the determinants of patient cohorts to achieve mobilization goals. The regression model considered the following variables: age, sex, ASA grade, surgical site, and the IPAQ [25]. Repeated measurement of the NRS pain score, using a generalized linear mixed model with multinomial distribution, was used for within-group comparisons.

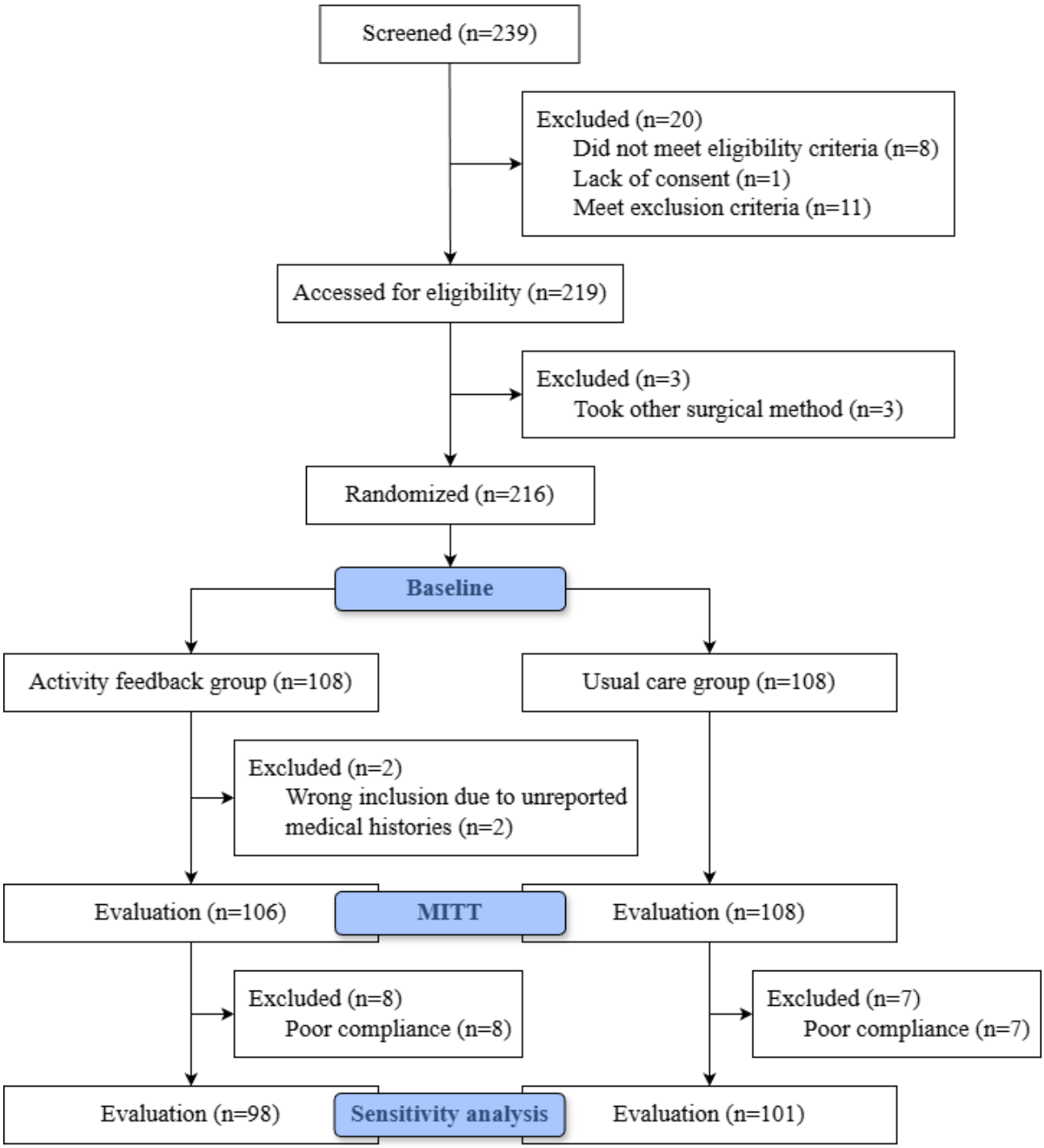
Results

A total of CRC 239 patients were recruited between February 2023 and September 2023. Of these, 216 were randomized, 108 into usual care and 108 with activity feedback. Two patients who met the exclusion criteria were inadvertently enrolled due to unreported medical histories and were subsequently excluded from the MITT analysis (Figure 1). Patients in each randomized group had similar baseline and perioperative characteristics (Table 1).

As shown in Table 2, no significant differences were found in CCI within 30 postoperative days between the control group, with a median CCI of 0 (range, 0 - 20.90), and the activity feedback group, with a median CCI of 0 (range, 0 - 12.20). The estimated mean difference between the 2 groups was −0.59 (95% CI −3.56 to 2.38; *P*=.66). Removing 15 patients with low device compliance as a sensitivity analysis, the estimated mean difference was −0.98 (95% CI −4.07 to 2.10; *P*=.54). The subgroup analysis results revealed no significant interaction between the two groups on CCI within 30 days and all the subgroup variables (Figure 2 and Checklist 1).

XSL•FO
RenderX

Figure 1. Flowchart showing patient inclusion and exclusion.



XSL•FO
RenderX

Table . Baseline and perioperative characteristics.

Patient characteristics	Usual care (n=108)	Activity feedback (n=108)	Mann-Whitney *U* test	Chi-square (*df*)	*t* test (*df*)	ρ	Continuity correction factor	*P* value
Age (y)	60.5 (54.25 - 68.75)	63 (55-68)	5998.5	—	—	—	—	.72
Sex, n (%)			—	0.020 (1)	—	—	—	.89
Male	67 (62)	68 (62)						
Female	41 (38)	40 (38)						
BMI[a] (kg/m²), mean (SD)	23.5 (3.3)	23.6 (3.2)	—	—	0.125 (214)	—	—	.90
ASA[b] grade, n (%)			—	—	—	0.016	—	.82
I	1 (0.9)	2 (1.9)						
II	97 (89.8)	94 (87.0)						
III	10 (9.3)	12 (11.1)						
Smoking history, n (%)	22 (20.4)	19 (17.6)	—	0.271 (1)	—	—	—	.60
POC[c], n (%)	46 (42.6)	44 (40.7)	—	0.076 (1)	—	—	—	.78
Hypertension, n (%)	34 (31.5)	33 (30.6)	—	0.022 (1)	—	—	—	.88
Diabetes, n (%)	20 (18.5)	12 (11.1)	—	2.348 (1)	—	—	—	.13
CHD[d], n (%)	4 (3.7)	9 (8.3)	—	2.046 (1)	—	—	—	.15
Preoperative chemoradiotherapy, n (%)	15 (13.9)	12 (11.1)	—	0.381 (1)	—	—	—	.54
IPAQ[e], n (%)			—	—	—	0.053	—	.43
High	23 (21.3)	20 (18.5)						
Medium	69 (63.9)	68 (63.0)						
Low	16 (14.8)	20 (18.5)						
Surgical site, n (%)			—	1.506 (1)	—	—	—	.22
Colon	46 (42.6)	55 (50.9)						
Rectum	62 (57.4)	53 (49.1)						
Laparoscopic, n (%)	103 (95.4)	105 (97.2%)	—	—	—	—	0.130	.72
Converted, n (%)	1 (0.9)	0	—	—	—	—	0.000	>.99
Surgery duration (min), median (IQR)	150 (125 - 183.75)	155 (130-185)	6031.5	—	—	—	—	.66
Bleeding volume (ml), median (IQR)	50 (20-50)	50 (20-50)	5989.5	—	—	—	—	.71
POEM[f] (mg), mean (SD)	56.3 (27.9)	58.1 (31.3)	—	—	0.446 (214)	—	—	.66

[a]BMI: Body Mass Index.

[b]ASA: American Society of Anesthesiologists.

XSL•FO
RenderX

[c]POC: preoperative complications.

[d]CHD: coronary heart disease.

[e]IPAQ: International Physical Activity Questionnaire.

[f]POEM: postoperative equivalent morphine consumption. Sufentanil 0.01 mg equal to morphine 10 mg; Hydromorphone 1.5 mg equal to morphine 10 mg; Nalbuphine 10 mg equal to morphine 10 mg [36].

Table . Between-group difference in CCI[a] 30 days after operation.

	Usual care (n=108)	Activity feedback (n=106)	Mean difference	*P* value
CCI, median (IQR)	0 (0 to 20.90)	0 (0 to 12.20)	−0.59 (−3.56 to 2.38)	.66
CCI[b], median (IQR)	0 (0 to 20.90)	0 (0 to 14.38)	−0.98 (−4.07 to 2.10)	.54
Maximum complication level, n (%)				
0	63 (58.3)	64 (60.4)	—[c]	—
1	14 (13)	17 (16)	—	—
2	26 (24.1)	21 (19.8)	—	—
3[b]	0	1 (0.9)	—	—
3[d]	5 (4.6)	1 (0.9)	—	—
4[b]	0	1 (0.9)	—	—
4[d]	0	1 (0.9)	—	—
5 (death)	0	0	—	—
Severe complications[e], n (%)	5 (4.6)	4 (3.8)	—	—

[a]CCI: comprehensive complications index.

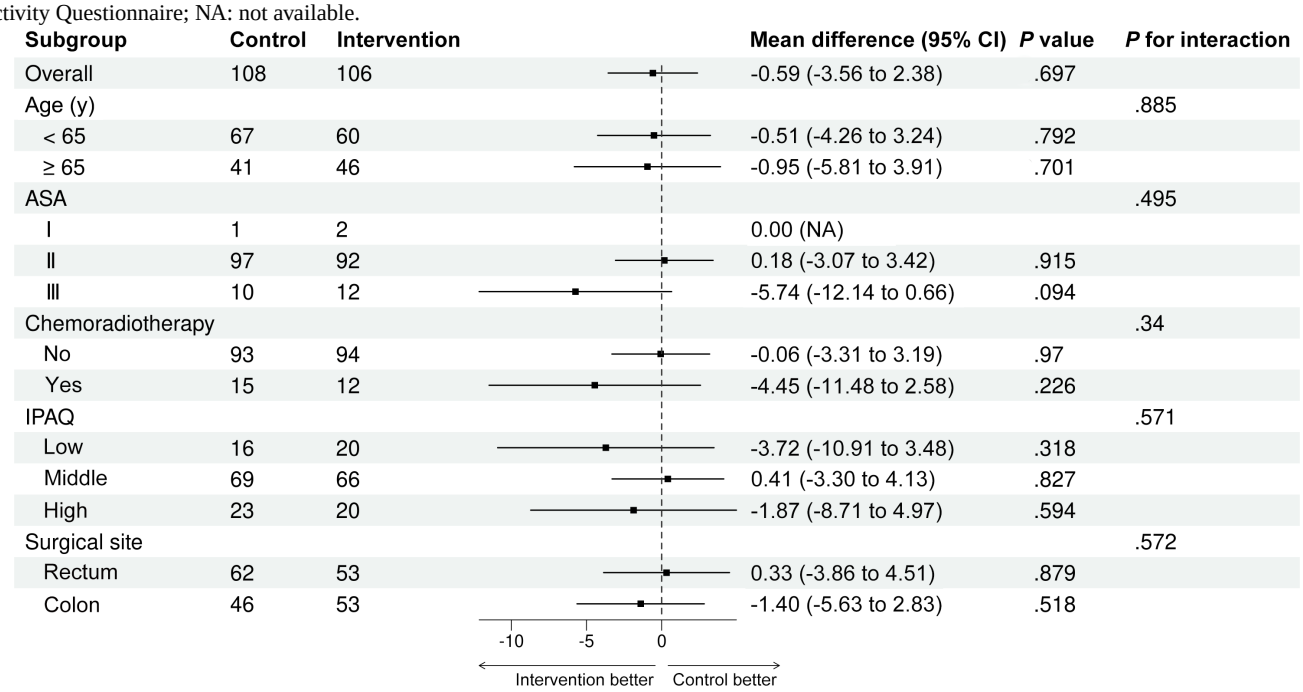
[b]Removing 15 patients with low device compliance.

[c]Not available.

[d]CCI grading was according to the classification of Clavien-Dindo classification.

[e]Severe complication defined as complication ≥3 [35].

Figure 2. Subgroup analysis forest plot for primary outcome indicators. ASA: American Society of Anesthesiologists; IPAQ: International Physical Activity Questionnaire; NA: not available.



XSL•FO
RenderX

The rate of no complications accounted for 58.3% (63/108) of patients in the usual care group and 60.4% (64/106) in the activity feedback group, while severe complications accounted for 4.6% (5/108) and 3.8% (4/106), respectively (Table 2). Both groups had the highest proportion of grade 2 complications. In addition, the incidence of urinary retention was the highest (15/108, 13.9% in the usual care group and 11/106, 10.4% in the activity feedback group) among the operative complications. The incidence rates of complications of infectious, respiratory system, and cardiovascular systems were 6.5% (7/108) vs 9.4% (11/106), 5.5% (6/108) vs 2.8% (3/106), and 1.9% (1/108) vs 1.8% (2/106), respectively, in the usual care group and in the activity feedback group (Table 3). On the analysis of specific complications, by the Fisher exact test, the incidence rate for pneumonia in the routine care group was 4.6% (5/108) and that in the activity feedback group was 1.9% (2/106), *P*=.45; the incidence rate for ileus in the routine care group was 0.9% (1/108) and that in the activity feedback group was 3.8% (4/106), *P*=.21, which indicated that the intervention had no significant effect on the incidence of these specific complications.

Table . Type of complications 30 days after operation.

	Usual care (n=108), n (%)	Activity feedback (n=106), n (%)
Cardiovascular system		
Acute myocardial infarction	0 (0)	1 (0.9)
Arrhythmia	2 (1.9)	1 (0.9)
Respiratory system		
Pneumonia	5 (4.6)	2 (1.9)
Atelectasis	1 (0.9)	0 (0)
Pleural effusion	0	1 (0.9)
Infectious complications		
Wound infection	3 (2.8)	6 (5.7)
Intraperitoneal or retroperitoneal abscess	0	2 (1.9)
Other infections[a]	4 (3.7)	3 (2.8)
Operative complication		
Anastomotic leakage	6 (5.6)	5 (4.7)
Ileus	1 (0.9)	4 (3.8)
Hemorrhage	0 (0)	3 (2.8)
Anemia	8 (7.4)	4 (3.8)
Urinary retention	15 (13.9)	11 (10.4)
Fat liquefaction	3 (2.8)	2 (1.9)
Other gastrointestinal complications[b]	6 (5.6)	7 (6.6)

[a]Other infections: including leukocytosis (white blood cells) >12 and temperature >38°C, shingles.

[b]Other gastrointestinal complications, including diarrhea, constipation, and gastroparesis.

As shown in Table 4, the 2 groups had no significant difference in daily activity time. Similarly, there was no difference in the rate of reaching the target between groups. The usual care and activity feedback groups demonstrated no significant difference in the length of postoperative hospital stay, with a median of 4 days for both groups (*P*=.99), in the first exhaust time (37.13, IQR 22.73-58.77 h vs 38.29, IQR 22.15-57.49 h, respectively, in the usual care group and in the activity feedback group; *P*=.89), and in the first defecating time (44.50, IQR 24.04-68.88 h vs 44.13, IQR 25.81-67.44 h; *P*=.91, respectively, in the usual care group and in the activity feedback group; Table 5). The results of the generalized linear mixed model (Table 6) revealed that the NRS scores of rest and activity showed a declining trend over time, with a significant time effect that was independent of the patient groups (*P*<.001). Moreover, there were no significant interactions between the intervention and the NRS scores at rest (*P*=.54) or during movement (*P*=.89).

In post hoc analysis, there was a significant negative correlation between CCI within the 30th postoperative day and the activity duration on POD2 (r=−0.166; *P*=.02) and POD3 (r=−0.264; *P*=.002). A linear regression analysis showed that the movement time of POD3 was negatively correlated with CCI within the 30th postoperative day, and the standardized β was −.025 (95% CI −.045 to −.006; *P*=.01). Additionally, the longer the activity time of POD3 before 215 minutes, the faster the CCI decreased; however, beyond 215 minutes, the rate of the decline slowed (Figure 3).

XSL•FO
RenderX

Table . Daily activity time and the rate of reaching the target.

	Usual care	Activity feedback	Mean difference/odds ratio (95% CI)	*P* value
Activity time, min				
POD1[a], median (IQR)	105 (74 to 133) (n=107)	114 (65 to 131) (n=105)	−0.7 (−13.4 to 12.0)[b]	.89
POD2[c], median (IQR)	174 (117 to 225) (n=103)	183 (134.5 to 229.5) (n=101)	5.2 (−13.1 to 23.5)[b]	.58
POD3[d], median (IQR)	201 (132 to 248) (n=71)	195 (125.8 to 260.5) (n=70)	−3.8 (−31.8 to 24.1)[b]	.73
Reaching the target				
POD1, n/N (%)	40/107 (37.4)	46/105 (43.8)	0.8 (0.4 to 1.4)[e]	.34
POD2, n/N (%)	20/103 (19.4)	21/101 (20.8)	1.1 (0.5 to 2.2)[e]	.81
POD3, n/N (%)	4/71 (5.6)	2/70 (2.9)	0.5 (0.1 to 3.0)[e]	.69

[a]POD1: the first day after operation.
[b]Mean difference (95% CI).
[c]POD2: the second day after operation.
[d]POD3: the third day after operation.
[e]Odds ratio (95% CI).

Table . Postoperative recovery index.

	Usual care (n=108), median (IQR)	Activity feedback (n=106), median (IQR)	Mean difference (95% CI)	*P* value
LOS[a] (d)	4 (3 to 5)	4 (4 to 5)	−0.4 (−1.49 to 0.69)	.47
First exhaust time (h)	37.13 (22.73 to 58.77)	38.29 (22.15 to 57.49)	0.13 (−8.07 to 8.34)	.98
First defecating time (h)	44.50 (24.04 to 68.88)	45.13 (25.81 to 67.44)	−2.99 (−14.00 to 8.03)	.60

[a]LOS: length of postoperative hospital stay.

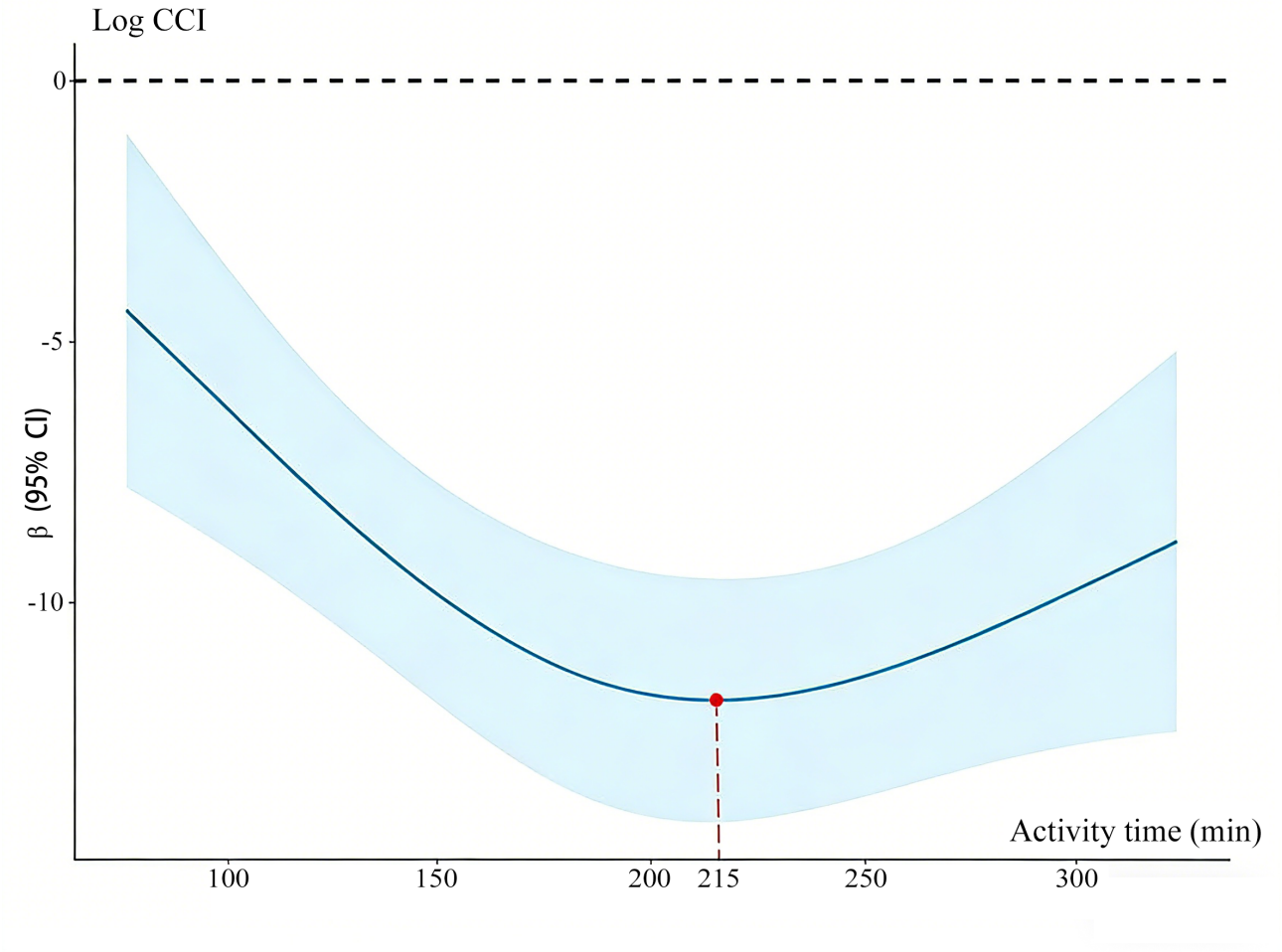
Table . Pain scores after surgery at rest or movement.

Pain scores	Usual care (n=108), median (IQR)	Activity feedback (n=106), median (IQR)	Fixed effect *P* value		
			Group	Time	Group*time
NRS[a] at rest			.95	.007	.54
POD1[b]	1 (0 - 2)	1 (0 - 2)			
POD2[c]	0 (0 - 2)	0 (0 - 2)			
POD3[d]	1 (0 - 1.75)	1 (0 - 1)			
NRS at movement			.32	.001	.89
POD1	2 (1 - 4)	2 (2 - 4)			
POD2	2 (1 - 3.75)	2 (1 - 4)			
POD3	2.5 (2 - 3)	2.5 (1.75 - 3)			

[a]NRS: Numeric Rating Scale.
[b]POD1: the first day after operation.
[c]POD2: the second day after operation.
[d]POD3: the third day after operation.

XSL•FO
RenderX

Figure 3. Multivariate restricted cubic spline analysis of activity time and comprehensive complications index (CCI).



Discussion

Principal Findings

In this study, the activity incentive feedback based on a wearable device (ePM/ep pod) did not reduce the CCI of patients with CRC within postoperative 30 days. The differences between groups in terms of daily activity time, length of postoperative hospital stay, intestinal function recovery, and NRS pain scores were also not significant.

There may be the following reasons for the lack of difference in CCI and daily activity durations between the 2 groups. First, this study was conducted in a surgical ward with relatively deepening ERAS guidelines, where preoperative education and routine activity supervision may have facilitated effective postoperative activities even in the absence of the feedback from wearable devices. Second, all the subjects in the present study received the same activity encouragement by medical staff according to the routine care and were set the same daily activity target; therefore, patients in the control group with high compliance might obtain activity feedback through other means. Third, laparoscopy accounted for more than 95% of CRC surgery in the present study, which was similar to the type of surgery used in another study that yielded negative results regarding the hypothesis that early postoperative activity could reduce complications [15]. Future studies should investigate whether such incentive feedback would be effective for patients experiencing more intense surgical stimulation or possessing less ERAS education. Previous studies using step count feedback for mobility-enhancing showed conflicting

The assessment of early postoperative activity achievement in patients with CRC directly reflects the practical outcomes of the intervention strategies employed in this study. The activity compliance rate of patients with CRC on POD1 was 37.4% (in the usual care group) and 43.8% (in the activity monitoring feedback group), which was much lower than expected, indicating that even with the feedback mechanism, there was only a marginal improvement in patients' adherence to the activity targets. Based on the current 6 hours recommended by ERAS guidelines, this study set a less active target of 2 hours for POD 0. In addition, the maximum activity time on POD1 was 310 minutes, which was much lower than the 6 hours recommended by ERAS guidelines. The relatively low compliance rates suggest that factors beyond the provision of activity feedback may be at play in influencing patients' postoperative behavior, such as the intensity and appropriateness of the activity targets set. The insights gained from analyzing the achievement of early postoperative activity not only provide a deeper understanding of the limitations of the feedback intervention but also highlight the need for more tailored and patient-centered approaches to enhance postoperative mobilization. The postoperative activity time was measured by an electronic parameter objectively recorded by a wearable device, and the measurement method may be different from the

XSL•FO
RenderX

considerations of previous guidelines. This may be the reason why the postoperative activity time was inconsistent with most previous studies.

Previous studies [37,38] have found patients undergoing major visceral (abdominal) surgery may not be able to achieve the mobilization goals in the early postoperative period. An observational pilot study [39] pointed out that only 50% of patients undergoing major visceral surgery could walk until POD2. The cumulative duration of postoperative activity per day ranged from 15 to 155 minutes, and only 40% of patients were satisfied with activity or physical exercise. There is a discrepancy between ERAS goals and clinical practice. Another study [40] reported that only a few patients (23.5%) achieved the mobilization goal of 6 hours (ERAS guidelines) of activity on the first day after CRC surgery. Ramírez et al [41] set the early activity goal for patients with CRC as sitting for at least 6 hours on the first postoperative day. However, the achievement rate was only 44.6%. In general, the mobilization goals described in most studies were either unclearly defined, imprecisely measured, or inaccurately monitored.

In addition, the early activity goals in ERAS did not differentiate between disease types, and radical abdominal tumor surgery is far more invasive than other simple abdominal organ partial removal. Patients with tumors were often elderly and had comorbidities, which can limit physical fitness and motivation. A study by Reed et al [42] on activity tracking after bariatric surgery found a significant negative correlation between age and the steps on POD1. Our study presented similar findings in that age was an independent factor affecting the achievement of activity on POD1. The older the patients were, the lower the possibility of reaching the activity goal. Furthermore, age negatively affected activity duration on POD2 and POD3. All patients with CRC in the present study were 61 years on average, and 42% had preoperative comorbidities, which may also be the reason for the low level of postoperative activity. Also, regression analysis showed that the activity duration of patients with colon cancer on POD2 was significantly longer than that of patients with rectal cancer, being 26.42 minutes longer. Patients with colon cancer were 2.28 times more likely than those with rectal cancer to meet the activity standard on POD2. It can be seen that the mobility of patients with colon cancer on POD2 was higher than that of patients with rectal cancer. We speculate that the reason for this finding may be that colon cancer basically does not receive radiotherapy, the proportion of preoperative chemotherapy may be lower than for rectal cancer, and the proportion of temporary ileostomy of colon cancer being much lower than that for rectal cancer. Our finding also suggests the target of early activity recommended by current ERAS guidelines (6 h in POD1) may be too high for these patients. A further increase in activity did not increase the benefit when activity levels reached a certain level. Considering our result of the regression analysis, 215 minutes might be considered as the target for activity time on POD3 to decrease CCI in patients with CRC.

In addition, several studies [43-45] have reported that early postoperative mobilization (within 24 h) was associated with fewer postoperative complications, fewer major complications, faster intestinal function recovery, and shorter hospital stays. The regression analysis in this study found that patients with high levels of physical activity before surgery were about 27 minutes more active the next day compared to patients with moderate levels of physical activity before surgery, and about 37 minutes more active than patients with low levels of physical activity before surgery. Therefore, perhaps by exploring the extent of postoperative activity decline caused by different surgical trauma, the approximate postoperative activity amount of the patient can be inferred and then combined with the preoperative activity level of the patient, a personalized postoperative activity plan can be formulated.

Limitations

This study had some limitations. First, the 216 patients were all admitted to one hospital; therefore, the conclusion still needs to be verified in more hospitals where ERAS guidelines are less practiced. The low proportion of ASA 3 patients in this study may limit our comprehensive assessment of the benefits of patients with high comorbidities from feedback from postoperative activities. Future research may consider including more ASA3 patients to better understand the response of this special population to postoperative activity feedback intervention. Second, since it was not possible to blind the patients in this study, the 2 groups of patients may have influenced each other. Due to practical difficulties such as high patient mobility and limited follow-up time, this study failed to collect information about the views of patients and clinicians with the wearable device and how motivated they felt by wearing this. In the future, the study can be expanded to multiple wards to achieve cluster randomization, which may reduce bias. Future research can also be combined with qualitative items to provide a more comprehensive understanding of patient and clinician experiences and perceptions of wearable devices. Third, as 96% of patients in this study underwent laparoscopic surgery, the results may not be generalized to open surgery or other types of surgery. Although our inclusion criteria were broad, we excluded patients who had difficulty with preoperative activity or were in a poor physical condition. These patients may benefit from the incentive of activity feedback. A final limitation was there may be a difference between self-reported preoperative mobility assessed by IPAQ, which was not designed for hospital patients, but for assessing physical activity in daily life, and actual quantitative mobility.

Conclusions

In the context of ERAS, this randomized controlled trial did not find evidence that activity stimulation based on feedback from ePM/ep pod wearable devices could reduce 30-day postoperative CCI in patients who underwent CRC surgery. However, ePM/ep pod could accurately record daily activity duration, which may be negatively correlated with CCI on POD3.

XSL•FO
RenderX

Acknowledgments

The authors wish to thank Chong Lei and Ziyu Zheng for participating in protocol discussion.

Funding

This research is supported by the University-funded clinical research project (2022L2213) and Hospital-funded military medical research project (JSYXZ12) to HN.

Data Availability

The datasets generated or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

Authors' Contributions

HN and HD designed the study. YM, YW, and HW conducted the study. NY and ZZ performed the data collection and data analysis. FF wrote the original draft of the paper. YM and HN contributed to critical revision. FF and HN reviewed and edited the manuscript and have approved its final version. HN is responsible for the overall content as guarantor.
HD is the co-corresponding author (email: hldong6@hotmail.com; phone: 8613669226699).

Conflicts of Interest

None declared.

Checklist 1
CONSORT-eHEALTH (V 1.6.1) checklist [46].
[PDF File, 11188 KB - mhealth_v14i1e70534_app1.pdf]

References

1. Zygulska AL, Pierzchalski P. Novel diagnostic biomarkers in colorectal cancer. Int J Mol Sci 2022 Jan 13;23(2):852. [doi: 10.3390/ijms23020852] [Medline: 35055034]
2. Baidoun F, Elshiwy K, Elkeraie Y, et al. Colorectal cancer epidemiology: recent trends and impact on outcomes. Curr Drug Targets 2021;22(9):998-1009. [doi: 10.2174/1389450121999201117115717] [Medline: 33208072]
3. van Rooijen S, Carli F, Dalton S, et al. Multimodal prehabilitation in colorectal cancer patients to improve functional capacity and reduce postoperative complications: the first international randomized controlled trial for multimodal prehabilitation. BMC Cancer 2019 Jan 22;19(1):98. [doi: 10.1186/s12885-018-5232-6] [Medline: 30670009]
4. Namba Y, Hirata Y, Mukai S, et al. Clinical indicators for the incidence of postoperative ileus after elective surgery for colorectal cancer. BMC Surg 2021 Feb 11;21(1):80. [doi: 10.1186/s12893-021-01093-7] [Medline: 33573636]
5. Khuri SF, Henderson WG, DePalma RG, et al. Determinants of long-term survival after major surgery and the adverse effect of postoperative complications. Ann Surg 2005 Sep;242(3):326-341. [doi: 10.1097/01.sla.0000179621.33268.83] [Medline: 16135919]
6. Aoyama T, Oba K, Honda M, et al. Impact of postoperative complications on the colorectal cancer survival and recurrence: analyses of pooled individual patients' data from three large phase III randomized trials. Cancer Med 2017 Jul;6(7):1573-1580. [doi: 10.1002/cam4.1126] [Medline: 28639738]
7. Arnarson Ö, Butt-Tuna S, Syk I. Postoperative complications following colonic resection for cancer are associated with impaired long-term survival. Colorectal Dis 2019 Jul;21(7):805-815. [doi: 10.1111/codi.14613] [Medline: 30884061]
8. Schwenk W. Optimized perioperative management (fast-track, ERAS) to enhance postoperative recovery in elective colorectal surgery. GMS Hyg Infect Control 2022;17:Doc10. [doi: 10.3205/dgkh000413] [Medline: 35909653]
9. Portinari M, Ascanelli S, Targa S, et al. Impact of a colorectal enhanced recovery program implementation on clinical outcomes and institutional costs: a prospective cohort study with retrospective control. Int J Surg 2018 May;53:206-213. [doi: 10.1016/j.ijsu.2018.03.005] [Medline: 29548700]
10. Chinese Society of Surgery, Chinese Society of Anesthesiology. Consensus on ERAS and guidelines for pathway management in China (2018). Chin J Pract Surg 2018;38(1):1-20. [doi: 10.19538/j.cjps.issn1005-2208.2018.01.01]
11. Gustafsson UO, Scott MJ, Hubner M, et al. Guidelines for perioperative care in elective colorectal surgery: enhanced recovery after surgery (ERAS®) society recommendations: 2018. World J Surg 2019 Mar;43(3):659-695. [doi: 10.1007/s00268-018-4844-y] [Medline: 30426190]
12. Xiang B, Jiao S, Si Y, Yao Y, Yuan F, Chen R. Risk factors for postoperative pneumonia: a case-control study. Front Public Health 2022;10:913897. [doi: 10.3389/fpubh.2022.913897] [Medline: 35875004]
13. Li Q, Du L, Lu L, et al. Clinical application of enhanced recovery after surgery in perioperative period of laparoscopic colorectal cancer surgery. J Laparoendosc Adv Surg Tech A 2019 Feb;29(2):178-183. [doi: 10.1089/lap.2018.0708] [Medline: 30614769]

XSL•FO
RenderX

14. Gustafsson UO, Oppelstrup H, Thorell A, Nygren J, Ljungqvist O. Adherence to the ERAS protocol is associated with 5-year survival after colorectal cancer surgery: a retrospective cohort study. World J Surg 2016 Jul;40(7):1741-1747. [doi: 10.1007/s00268-016-3460-y] [Medline: 26913728]
15. Fiore JF Jr, Castelino T, Pecorelli N, et al. Ensuring early mobilization within an enhanced recovery program for colorectal surgery: a randomized controlled trial. Ann Surg 2017 Aug;266(2):223-231. [doi: 10.1097/SLA.0000000000002114] [Medline: 27997472]
16. Grass F, Pache B, Martin D, et al. Feasibility of early postoperative mobilisation after colorectal surgery: a retrospective cohort study. Int J Surg 2018 Aug;56:161-166. [doi: 10.1016/j.ijsu.2018.06.024] [Medline: 29935366]
17. Tazreean R, Nelson G, Twomey R. Early mobilization in enhanced recovery after surgery pathways: current evidence and recent advancements. J Comp Eff Res 2022 Feb;11(2):121-129. [doi: 10.2217/cer-2021-0258] [Medline: 35045757]
18. Twomey R, Culos-Reed SN, Daun JT, Ferber R, Dort JC. Wearable activity trackers and mobilization after major head and neck cancer surgery: you can't improve what you don't measure. Int J Surg 2020 Dec;84:120-124. [doi: 10.1016/j.ijsu.2020.10.032] [Medline: 33157275]
19. Wiklund M, Sundqvist E, Fagevik Olsén M. Physical activity in the immediate postoperative phase in patients undergoing Roux-en-Y gastric bypass-a randomized controlled trial. Obes Surg 2015 Dec;25(12):2245-2250. [doi: 10.1007/s11695-015-1690-y] [Medline: 25910983]
20. No JH, Kim K, Kim YB, et al. Effects of an activity tracker with feedback on physical activity in women after midline laparotomy: a randomized controlled trial. J Obstet Gynaecol Res 2021 Jul;47(7):2544-2550. [doi: 10.1111/jog.14807] [Medline: 33899302]
21. Keats MR, Yu X, Sweeney Magee M, et al. Use of wearable activity-monitoring technologies to promote physical activity in cancer survivors: challenges and opportunities for improved cancer care. Int J Environ Res Public Health 2023 Mar 8;20(6):4784. [doi: 10.3390/ijerph20064784] [Medline: 36981693]
22. Singh B, Zopf EM, Howden EJ. Effect and feasibility of wearable physical activity trackers and pedometers for increasing physical activity and improving health outcomes in cancer survivors: a systematic review and meta-analysis. J Sport Health Sci 2022 Mar;11(2):184-193. [doi: 10.1016/j.jshs.2021.07.008] [Medline: 34314878]
23. Ma LN, Nie H, Ji G, et al. Accuracy of wearable device in monitoring postoperative motion and sleep of patients. Chin J Anesthesiol 2020;40(10):1232-1236. [doi: 10.3760/cma.j.cn131073.20200820.01020]
24. The Ethics Committee of the First Affiliated Hospital of the Air Force Military Medical University. Medical ethics committee charter. URL: https://www.fmmu.edu.cn/info/1404/4453.htm [accessed 2026-01-22]
25. Roberts-Lewis SF, White CM, Ashworth M, Rose MR. The validity of the international physical activity questionnaire (IPAQ) for adults with progressive muscle diseases. Disabil Rehabil 2022 Nov;44(23):7312-7320. [doi: 10.1080/09638288.2021.1983042] [Medline: 34606392]
26. Slankamenac K, Graf R, Barkun J, Puhan MA, Clavien PA. The comprehensive complication index: a novel continuous scale to measure surgical morbidity. Ann Surg 2013 Jul;258(1):1-7. [doi: 10.1097/SLA.0b013e318296c732] [Medline: 23728278]
27. Kudo T, Oshikiri T, Goto H, et al. Comprehensive complication index as a prognostic factor in minimally invasive esophagectomy for esophageal squamous cell carcinoma. Esophagus 2022 Jul;19(3):410-416. [doi: 10.1007/s10388-022-00911-y] [Medline: 35220510]
28. Artiles-Armas M, Roque-Castellano C, Conde-Martel A, Marchena-Gómez J. The comprehensive complication index is related to frailty in elderly surgical patients. J Surg Res 2019 Dec;244:218-224. [doi: 10.1016/j.jss.2019.06.011] [Medline: 31301477]
29. Kim SH, Hwang HK, Lee WJ, Kang CM. Comprehensive complication index or Clavien-Dindo classification: which is better for evaluating the severity of postoperative complications following pancreatectomy? World J Surg 2021 Mar;45(3):849-856. [doi: 10.1007/s00268-020-05859-7] [Medline: 33191470]
30. Home of AssesSurgery. AssesSurgery GmbH. URL: https://www.assessurgery.com/ [accessed 2026-01-21]
31. Finet M, Bellicha A, Sage E, et al. Comprehensive assessment of postoperative mobility during the first days after mini-invasive lung surgery: a prospective observational study. J Clin Anesth 2023 Jun;86:111048. [doi: 10.1016/j.jclinane.2022.111048] [Medline: 36716650]
32. Gagliese L, Weizblit N, Ellis W, Chan VWS. The measurement of postoperative pain: a comparison of intensity scales in younger and older surgical patients. Pain 2005 Oct;117(3):412-420. [doi: 10.1016/j.pain.2005.07.004] [Medline: 16153776]
33. Spanjersberg WR, Reurings J, Keus F, van Laarhoven CJ. Fast track surgery versus conventional recovery strategies for colorectal surgery. Cochrane Database Syst Rev 2011 Feb 16(2):CD007635. [doi: 10.1002/14651858.CD007635.pub2] [Medline: 21328298]
34. Lv L, Shao YF, Zhou YB. The enhanced recovery after surgery (ERAS) pathway for patients undergoing colorectal surgery: an update of meta-analysis of randomized controlled trials. Int J Colorectal Dis 2012 Dec;27(12):1549-1554. [doi: 10.1007/s00384-012-1577-5] [Medline: 23001161]
35. Gloor S, Misirlic M, Frei-Lanter C, et al. Prehabilitation in patients undergoing colorectal surgery fails to confer reduction in overall morbidity: results of a single-center, blinded, randomized controlled trial. Langenbecks Arch Surg 2022 May;407(3):897-907. [doi: 10.1007/s00423-022-02449-0] [Medline: 35084526]

XSL•FO
RenderX

36. Chinese Society of Anesthesiology. Expert consensus on postoperative pain management in adults (2017 Edition). J Clin Anesthesiol 2017;33(9):911-917.
37. van Zelm R, Coeckelberghs E, Sermeus W, et al. Variation in care for surgical patients with colorectal cancer: protocol adherence in 12 European hospitals. Int J Colorectal Dis 2017 Oct;32(10):1471-1478. [doi: 10.1007/s00384-017-2863-z] [Medline: 28717841]
38. He Y, Yang J, Jiang L. Analysis of associated factors of early mobilization in colorectal cancer patients who underwent enhanced recovery after surgery pathway. Chin J Bases Clin Gen Surg 2021;28(12):1599-1603. [doi: 10.7507/1007-9424.202106124]
39. Wolk S, Distler M, Müssle B, Söthje S, Weitz J, Welsch T. Adherence to ERAS elements in major visceral surgery-an observational pilot study. Langenbecks Arch Surg 2016 May;401(3):349-356. [doi: 10.1007/s00423-016-1407-2] [Medline: 27013325]
40. Gustafsson UO, Hausel J, Thorell A, et al. Adherence to the enhanced recovery after surgery protocol and outcomes after colorectal cancer surgery. Arch Surg 2011 May;146(5):571-577. [doi: 10.1001/archsurg.2010.309] [Medline: 21242424]
41. Ramírez JM, Blasco JA, Roig JV, et al. Enhanced recovery in colorectal surgery: a multicentre study. BMC Surg 2011 Apr 14;11:9. [doi: 10.1186/1471-2482-11-9] [Medline: 21489315]
42. Reed B, Tabone LE, Tabone JK, Szoka N, Abunnaja S, Bailey K. The use of an activity tracker to objectively measure inpatient activity after bariatric surgery. Surg Obes Relat Dis 2021 Jan;17(1):90-95. [doi: 10.1016/j.soard.2020.08.033] [Medline: 33032917]
43. Twomey R, Matthews TW, Nakoneshny S, et al. Impact of early mobilization on recovery after major head and neck surgery with free flap reconstruction. Cancers (Basel) 2021 Jun 8;13(12):2852. [doi: 10.3390/cancers13122852] [Medline: 34201003]
44. Ding X, Zhang H, Liu H. Early ambulation and postoperative recovery of patients with lung cancer under thoracoscopic surgery-an observational study. J Cardiothorac Surg 2023 Apr 11;18(1):136. [doi: 10.1186/s13019-023-02263-9] [Medline: 37041603]
45. Cook DJ, Thompson JE, Prinsen SK, Dearani JA, Deschamps C. Functional recovery in the elderly after major surgery: assessment of mobility recovery using wireless technology. Ann Thorac Surg 2013 Sep;96(3):1057-1061. [doi: 10.1016/j.athoracsur.2013.05.092] [Medline: 23992697]
46. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

ASA: American Society of Anesthesiologists
CCI: comprehensive complications index
CRC: colorectal cancer
ep pod: electrophysiology pod
ePM: enhanced patient monitoring
ERAS: enhanced recovery after surgery
IPAQ: International Physical Activity Questionnaire
MITT: modified intention-to-treat
NRS: Numeric Rating Scale
POD1: the first day after operation
POD2: the second day after operation
POD3: the third day after operation

Edited by L Buis; submitted 25.Dec.2024; peer-reviewed by A Dosis, G Pisaneschi; revised version received 20.Dec.2025; accepted 28.Dec.2025; published 30.Jan.2026.

Please cite as:
Meng Y, Fan F, Ma Y, Yan N, Wang H, Zhang Z, Wang Y, Dong H, Nie H
Impact of Mobilization Facilitated by Wearable Device Enhanced Patient Monitoring/Electrophysiology Pod–Based Feedback on Postoperative Complications Following Colorectal Cancer Surgery: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e70534
URL: https://mhealth.jmir.org/2026/1/e70534
doi:10.2196/70534

XSL•FO
RenderX

© Yang Meng, Fengyan Fan, Yumei Ma, Nong Yan, Huan Wang, Zhen Zhang, Yiting Wang, Hailong Dong, Huang Nie. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 30.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Real-World Long-Term Engagement With a Mobile App Intervention to Improve Self-Management of Type 2 Diabetes Mellitus in China (SMARTDiabetes): Mixed Methods Study

Xuanchen Tao[1,2], PhD; David Peiris[1], PhD; Yuxia Ma[3,4], PhD; Yaoshen Zhang[5]; Hongyi Song[6], MSc; Limin Mao[7], PhD; Puhong Zhang[1,8], PhD

[1]The George Institute for Global Health, University of New South Wales, Sydney, Australia

[2]Children's Hospital at Westmead Clinical School, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

[3]School of Public Health, Hebei Medical University, Shijiazhuang City, Hebei Province, China

[4]Hebei Key Laboratory of Environment and Human Health, Shijiazhuang City, Hebei Province, China

[5]Luquan Center for Disease Control and Prevention, Shijiazhuang City, Hebei Province, China

[6]The George Institute for Global Health, Beijing, China

[7]Center for Social Research in Health, University of New South Wales, Sydney, Australia

[8]Beijing Physical Examination Center, No 59 Beiweilu, Xicheng District, Beijing, China

Corresponding Author:
Puhong Zhang, PhD
The George Institute for Global Health, University of New South Wales, Sydney, Australia

Abstract

Background: Despite the high prevalence of type 2 diabetes mellitus in China, few patients are achieving adequate management targets. While digital behavior change interventions (DBCIs) are promising solutions, sustained long-term user engagement with these new technologies remains challenging.

Objective: The objectives of this study were to (1) describe the long-term user engagement with the SMARTDiabetes app, (2) explore the associations between user engagement and trial effectiveness, and (3) identify main influencing factors of user engagement.

Methods: A mixed methods process evaluation was conducted alongside the 2-year SMARTDiabetes trial to assess long-term user engagement with the DBCI, investigate its association with trial effectiveness, and explore its influencing factors. Quantitative data included clinical outcomes and app usage data; qualitative data included an expert review of the DBCI and semistructured interviews with key stakeholders. Findings were synthesized using Michie's conceptual framework for engagement with DBCIs.

Results: A total of 1038 participants were included in the quantitative analyses, and 12 in-depth interviews were conducted in June 2020. The mean proportion of monthly active users was 46.4% (SD 7.8%), with a gradual decline from 56.3% at the start (April 2018) to 42.2% at the end of trial (July 2019). The proportion of sustained active users was 32.9% (342/1038), with rural users showing a significantly higher rate than urban users (38.3% vs 27.3%; P<.001). We found a strong link between user engagement and the effectiveness in achieving primary health outcome (risk ratio 1.37, 95% CI 1.21 - 1,55; P<.001) and a clear dose-response relationship between the level of app interactions and degree of improvement in the health outcome. Key drivers of user engagement included 6 behavior change techniques (ie, reward, feedback, reminders, goal setting, self-monitoring, and social support features), 4 implementation facilitators (credibility features, professional support, guidance, and ease of use), and 5 contextual contributors (younger age, better education, greater smartphone literacy, closer patient-doctor relationship, and stronger primary health care facilities commitment).

Conclusions: Long-term, sustained engagement with a mobile app intervention for type 2 diabetes mellitus self-management is achievable in the primary health care setting in China. Higher engagement is associated with better health outcomes. The influencing factors identified in the SMARTDiabetes study provide valuable insights for developing targeted strategies to enhance user engagement with DBCIs and improve overall program effectiveness.

Trial Registration: ClinicalTrials.gov NCT02726100; https://clinicaltrials.gov/study/NCT02726100

(JMIR Mhealth Uhealth 2026;14:e76699) doi:10.2196/76699

KEYWORDS

digital behavior change intervention; mHealth; engagement; type 2 diabetes mellitus; China; primary health care; implementation

XSL•FO
RenderX

Introduction

China has the largest number of patients with type 2 diabetes mellitus (T2DM) globally. The Chinese government has launched a series of health reforms since 2009, aiming to improve the management of diabetes by strengthening primary health care (PHC) [1]. In China, PHC facilities are typically the first point of contact for patients with the national health system, and individuals with T2DM receive most of their health care services at these facilities. PHC facilities are organized into 2 tiers, with grassroots-level facilities managed under the supervision of center-level facilities. In urban areas, community health stations are managed by community health centers, whereas in rural areas, village clinics are managed by township health centers [1]. Despite the reforms, the rates of awareness (36.7%), treatment (32.9%), and control (50.1%) of diabetes remained low [2]. The study projected that the suboptimal management of diabetes in China will lead to a rapid increase in economic burden, with the total costs rising from US $250.2 billion in 2020 to US $460.4 billion by 2030 [3]. These findings underscore the urgent need for better and innovative T2DM management strategies.

To address this challenge, we developed a multifaceted digital behavior change intervention (DBCI) and evaluated it in a 24-month cluster randomized controlled trial in 80 PHC facilities (40 rural villages and 40 urban communities) in Shijiazhuang City, Hebei Province, China [4,5]. The SMARTDiabetes intervention was underpinned by a mobile health (mHealth) system comprising three integrated components: (1) an app for patients and lay family members, also known as family health promoters (FHPs), who were invited to assist patients in managing their health conditions; (2) an app for PHC providers, which provided e-clinical decision support and access to a daily e-log containing data entered by patients; and (3) a web-based dashboard for local health administrators or governors, which generated real-time data-monitoring reports of key performance indicators (Figure S1 in Multimedia Appendix 1). The primary clinical outcome was the difference in proportion of patients achieving at least 2 American Diabetes Association "ABC" management targets ([A] glycated hemoglobin [HbA_{1c}] <7.0%, [B] blood pressure (BP) <140/80 mm Hg, and [C] low-density lipoprotein cholesterol <2.6 mmol/L) at 24 months. Overall, the SMARTDiabetes intervention significantly improved the primary outcome (intervention group: 35.9% vs usual care: 29.9%; *P*=.03). However, we observed significant heterogeneity in the primary outcome between rural and urban settings; the same intervention was significantly more effective in rural areas than in urban areas (rural—intervention group: 42.6% vs usual care: 31.0%; urban—intervention group: 27.9% vs usual care: 28.6% [*P*=.02 for heterogeneity]) [5].

Engagement with DBCIs, defined by (1) the extent (eg, amount, frequency, duration, and depth) of usage and (2) a subjective experience characterized by attention, interest, and affect [6], is logically a precondition for effectiveness [7-11]. Many studies on diabetes self-management apps have shown a positive association between user engagement and intervention effectiveness [12-15]. However, in practice, low clinical uptake, poor user engagement, and sustaining long-term participation are common challenges for these DBCIs [16-18]. The SMARTDiabetes trial is the longest (2 years) mobile app intervention for improving T2DM self-management with a rigorous research design (clustered randomized controlled trial) and a decent sample size (about 2000 patients) in the real-world PHC setting. Therefore, evaluating the long-term user engagement and understanding its influencing factors are crucial for optimizing the DBCIs design and implementation. These insights can help develop targeted strategies to enhance user engagement and improve overall program effectiveness [11,19].

In this study, we adopted a mixed methods multidimensional approach to assess both the subjective and objective measures of user engagement [6,11]. The aims of this study were to (1) describe the long-term user engagement with the SMARTDiabetes app, (2) explore the associations between user engagement and trial effectiveness (ie, the "ABC" biomarkers predetermined as the primary and secondary trial outcomes), and (3) identify main influencing factors of user engagement.

Methods

Study Design

In this mixed methods process evaluation, we conducted a descriptive analysis of user engagement with the SMARTDiabetes app over 2 years and a post-hoc analysis of its association with the trial effectiveness. In addition, to identify main influencing factors with the user engagement and trial effectiveness, we conducted a rapid theory-informed expert review of the DBCI tools, followed by 12 semistructured interviews to key stakeholders (ie, selected PHC providers, patients, and their FHPs in the intervention arm) conducted in June 2020, and concluded with a synthesis of evidence to explain the significant geographic heterogeneity in trial effectiveness observed in our previous publication (see Figure S2 in Multimedia Appendix 1 for timeline for data collection) [5]. The mixed methods study conformed with the GRAMMS (Good Reporting of A Mixed Methods Study) checklist [20].

Setting and Participants

The study recruitment was conducted at 80 PHC facilities (40 rural and 40 urban) in Shijiazhuang city, China, from August 2017 to December 2017. A full description of selection criteria for PHC facilities has previously been published [5]. PHC providers invited participants from the diabetes management list within the routine registration system for screening. This list included all identified patients with diabetes within the service range of the PHC facilities. The recruitment inclusion criteria were adults who (1) were diagnosed with T2DM, (2) were aged 40 years or older, (3) had a HbA_{1c} level of 7.0% or higher, (4) had access to the internet via a smartphone (confirmed verbally), either independently or through a designated family member, and (5) provided informed consent. Patients were excluded if they (1) had severe physical or psychological injury or illness, (2) were unable to attend the site visit or consciously answer questions, (3) were women in the process of or planning for pregnancy or breastfeeding, or (4) participated in any other clinical trial within the previous 6 months.

XSL•FO
RenderX

SMARTDiabetes Intervention

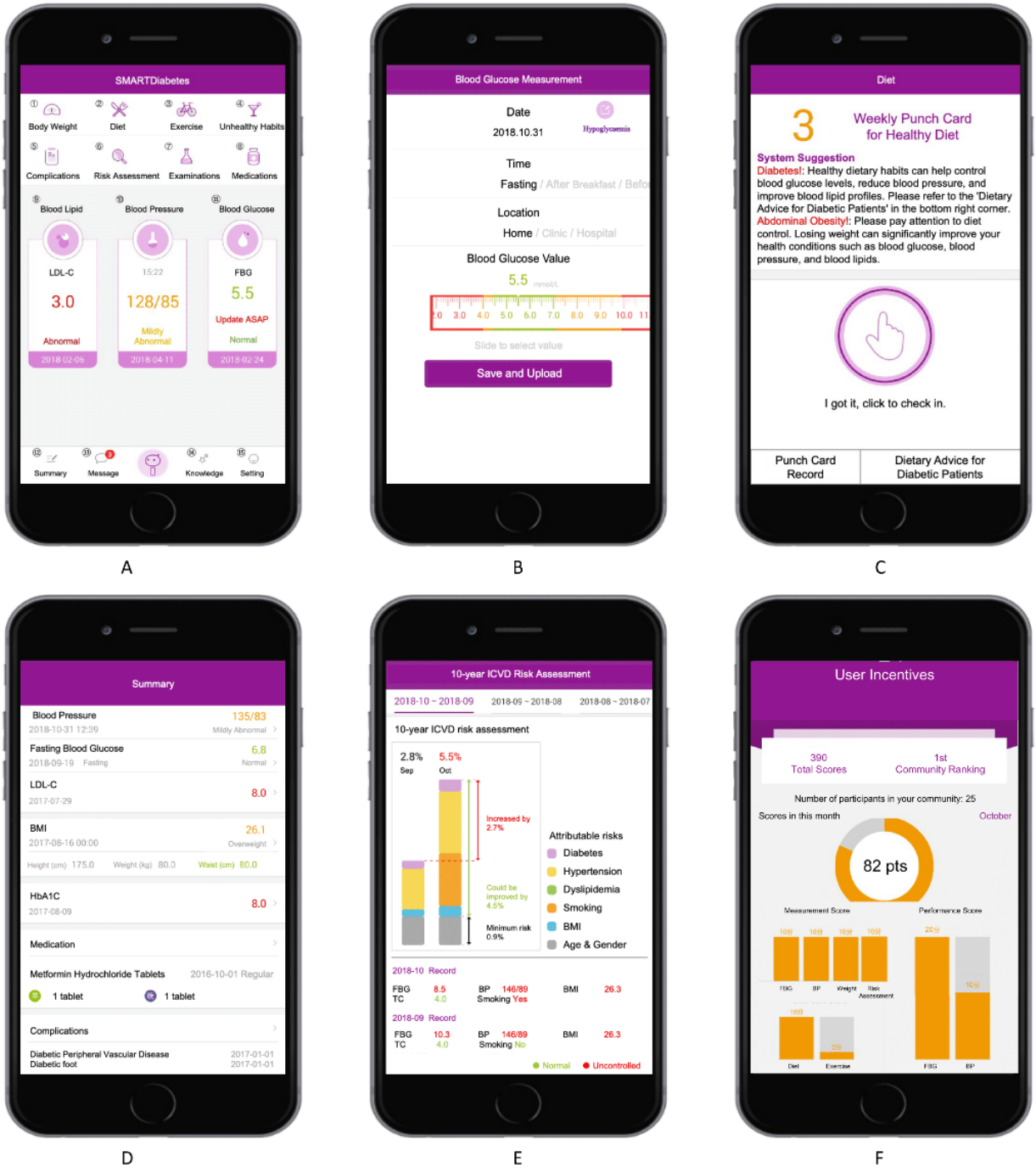
Based on a previous mixed methods study that systematically explored major barriers to and facilitators of PHC management of T2DM in China [21], our team developed the SMARTDiabetes app adopting a user-centered design approach [4,5]. In brief, the SMARTDiabetes app incorporates recommendations from Chinese national and international guidelines to support disease self-monitoring, adoption of healthier lifestyles, better medication adherence, and prevention of diabetes complications [4]. As illustrated in Figure 1, a total of 15 functions are displayed on the main landing page (Figure 1A). Four functions, "Blood glucose," "Blood pressure," "Blood lipid," and "Body weight" (Figure 1B), are for disease self-monitoring, where users can upload their latest data to receive immediate in-app feedback with result interpretation, improvement suggestions (goal setting), and relevant healthy literacy materials. The next 2 functions, "Diet" and "Exercise" (Figure 1C), are assisted by a digital punch card (*Da Ka*—clocking registration) system, which sends reminders for consuming healthier diets and engaging in appropriate physical activities. A further 5 functions, "Unhealthy lifestyles," "Health status summaries," "Physical examinations and lab tests," "Medications," and "Diabetes complications" (Figure 1D), serve as a medical record, enabling the tracking of unhealthy practices, medication adherence, and disease progression. The 12th function—"iCVD risk assessments" (Figure 1E) shows onset probabilities of ischemic cardiovascular diseases (iCVDs) including stroke within the following 10-year period, accompanied by a forecasted, monthly adjusted risk progression curve, on the basis of users' latest self-monitored data entry. This risk forecast model was constructed by the research team, using the risk assessment tools of Wu and colleagues [22]. The final 3 functions include 2 administrative functions ("Settings" and "Messages") and 1 preloaded health literacy function ("Knowledge and skills bank").

The reward program (Figure 1F) allows users to earn and accumulate reward points by interacting with the 6 core app functions (ie, "Blood glucose self-monitoring," "Blood pressure self-monitoring," "Body weight," "Diet punch card," "Exercise punch card," and "iCVD risk assessment quiz"). Users can earn up to 100 points per month (see detailed scoreboard in Table S1 in Multimedia Appendix 1) and redeem their points for small gifts of little monetary value, such as soap, towels, or laundry powder, from their respective PHC providers in the communities.

Before the trial commenced, the research team conducted a comprehensive training session for the 40 PHC providers in the intervention arm, covering downloading, installing, and using both the patient and provider versions of the SMARTDiabetes app. Subsequently, PHC providers organized training sessions for their enrolled patients and the patients' FHPs. To facilitate and standardize these sessions, a PowerPoint slide deck and an instruction brochure were developed (Figure S2 in Multimedia Appendix 1).

XSL•FO
RenderX

Figure 1. Main functions of the SMARTDiabetes app.



User Engagement Metrics

Deidentified app usage data collected between April 2018 and July 2019, including log-in frequencies and app functions click-through counts, were extracted by the service provider (Beihang University) and sent to the research team for secondary data analysis. In addition, a series of autogenerated key performance indicators (eg, BP control rates and fasting blood glucose [FBG] self-monitoring rates), in the form of cluster-level aggregated monthly reports, were presented on the SMARTDiabetes website. The research team documented these key performance indicators at the end of each month.

User engagement was evaluated using the following four metrics. (1) The proportion of monthly active users (MAUs), defined as participants who logged into the app at least once in a certain month, which was calculated by dividing the number of MAUs by the total number of participants in the intervention arm. Sustained active users were defined as participants who maintained MAU status for at least 12 months within the 16-month assessment period (April 2018 to July 2019), while others were categorized as “inactive users.” (2) The proportion of users with FHP involvement, which was calculated by dividing the number of patients who reported receiving FHPs’ assistance in app use by the total number of participants in the

XSL•FO
RenderX

intervention arm. (3) The total click-through counts per month, which was calculated by adding the number of clicks by each function. (4) The proportion of users who adhered to the recommended FBG self-monitoring frequencies per the SMARTDiabetes app's algorithm (see the algorithm in Table S2 in Multimedia Appendix 1).

The first 2 metrics were derived from app usage data, the third metric was obtained from the final trial assessment questionnaire, and the fourth metric was a cluster-level performance indicator retrieved from the SMARTDiabetes website. Descriptive analysis was conducted using SAS software (version 9.4 or above; SAS Institute). Means and SDs were reported for continuous variables with normal distribution; medians and IQRs were reported when the data were skewed. Proportions were reported for categorical variables.

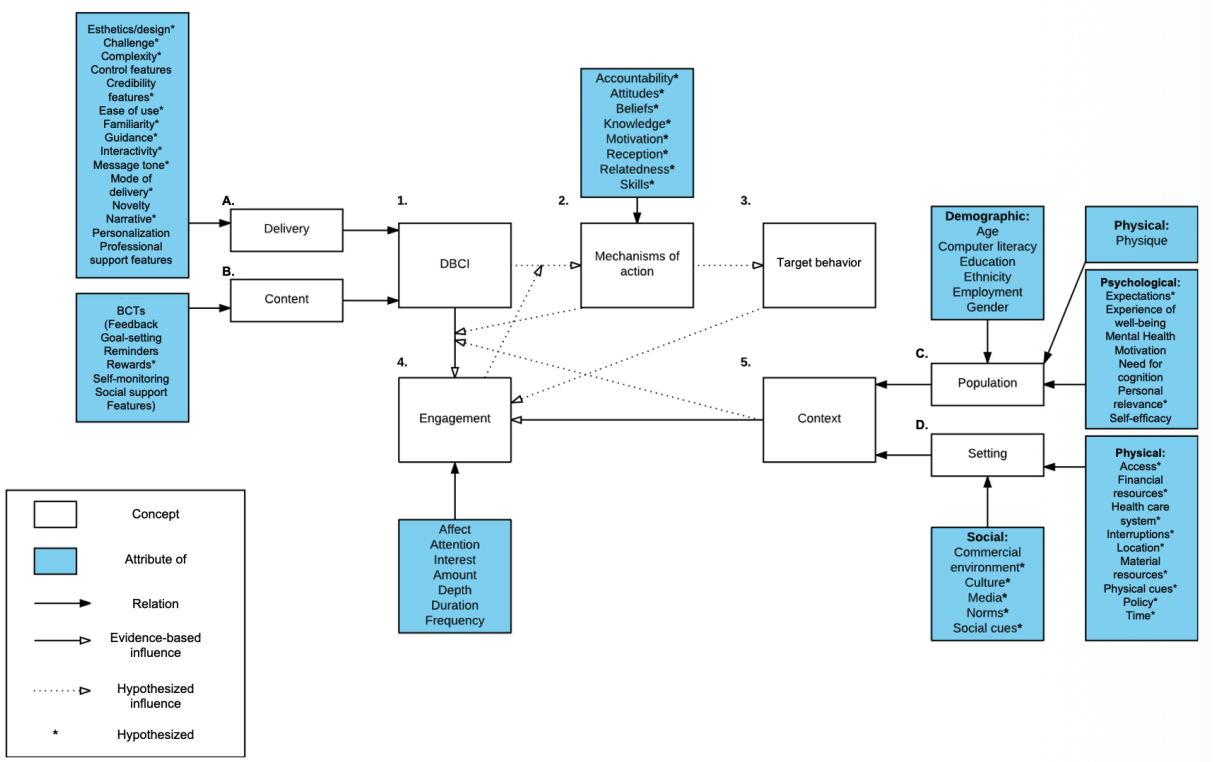
Association Between User Engagement Levels and Trial Effectiveness

In a post hoc analysis, associations with the primary outcome at 24 months were analyzed for the following subgroups: (1) sustained active users versus inactive users, (2) users with versus without FHP involvement in app use assistance, and (3) click-through registration levels, with users grouped into quartiles based on the total number of clicks on app functions during the same period.

An inverse propensity score weighted analysis was conducted to compare the trial effectiveness across subgroups with different engagement levels. Propensity scores were developed to control for the differences in baseline demographics, medical history, and laboratory characteristics across subgroups. Propensity scores were generated by a logit regression model with covariates of age, sex, locality, HBA_{1c}, FBG, systolic BP, diastolic BP, low-density lipoprotein cholesterol, education levels, duration of diabetes category, and diabetic complications. A log-binomial regression model with generalized estimation equation and adjustment for community clustering and the weighted inverse propensity score was then developed. The associations were presented as the relative risk (RR) with its 95% CIs. All analyses were performed with SAS software (version 9.4 or above; SAS Institute).

To better understand the geographic heterogeneity in trial effectiveness observed in our primary results study [5], we conducted additional subgroup analyses by locality to understand their differences in PHC facilities characteristics (Figure 2D), population baseline characteristics (Figure 2C), and engagement patterns (Figures 2-4). Two-tailed *t* tests were conducted for continuous variables with normal distribution, Wilcoxon signed-rank tests were conducted for continuous variables without a normal distribution, and chi-square tests were conducted for categorical variables.

Figure 2. Conceptual framework of direct and indirect influences on engagement with DBCIs. BCTs: behavior change techniques; DBCI: digital behavior change intervention.



XSL•FO
RenderX

Figure 3. The proportion of monthly active users.

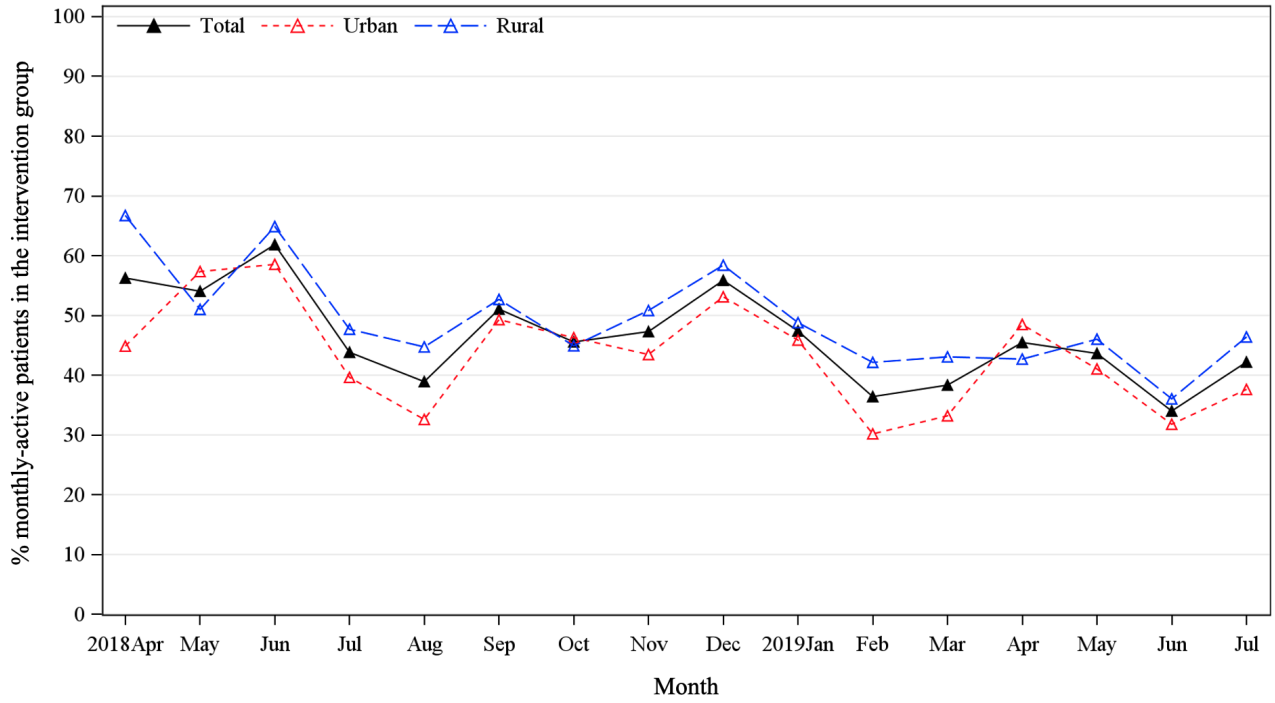
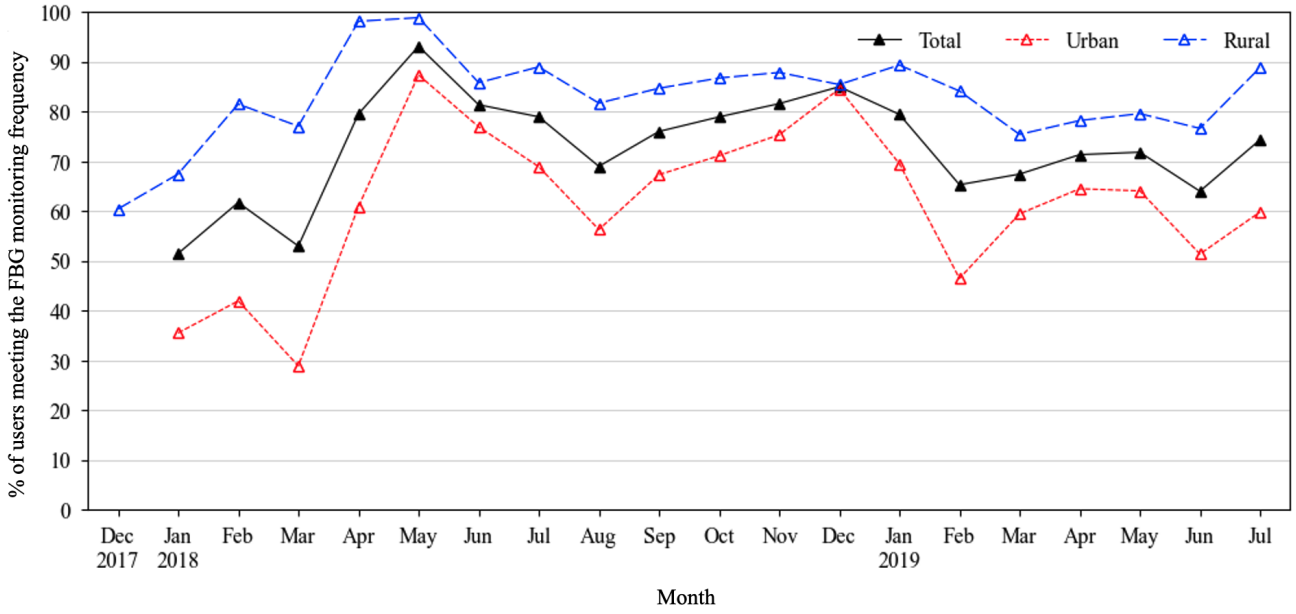


Figure 4. The proportion of users meeting the FBG monitoring frequency. FBG: fasting blood glucose.



Influencing Factors of User Engagement

After the trial, 2 qualitative activities were undertaken to explore the influencing factors for user engagement and trial effectiveness.

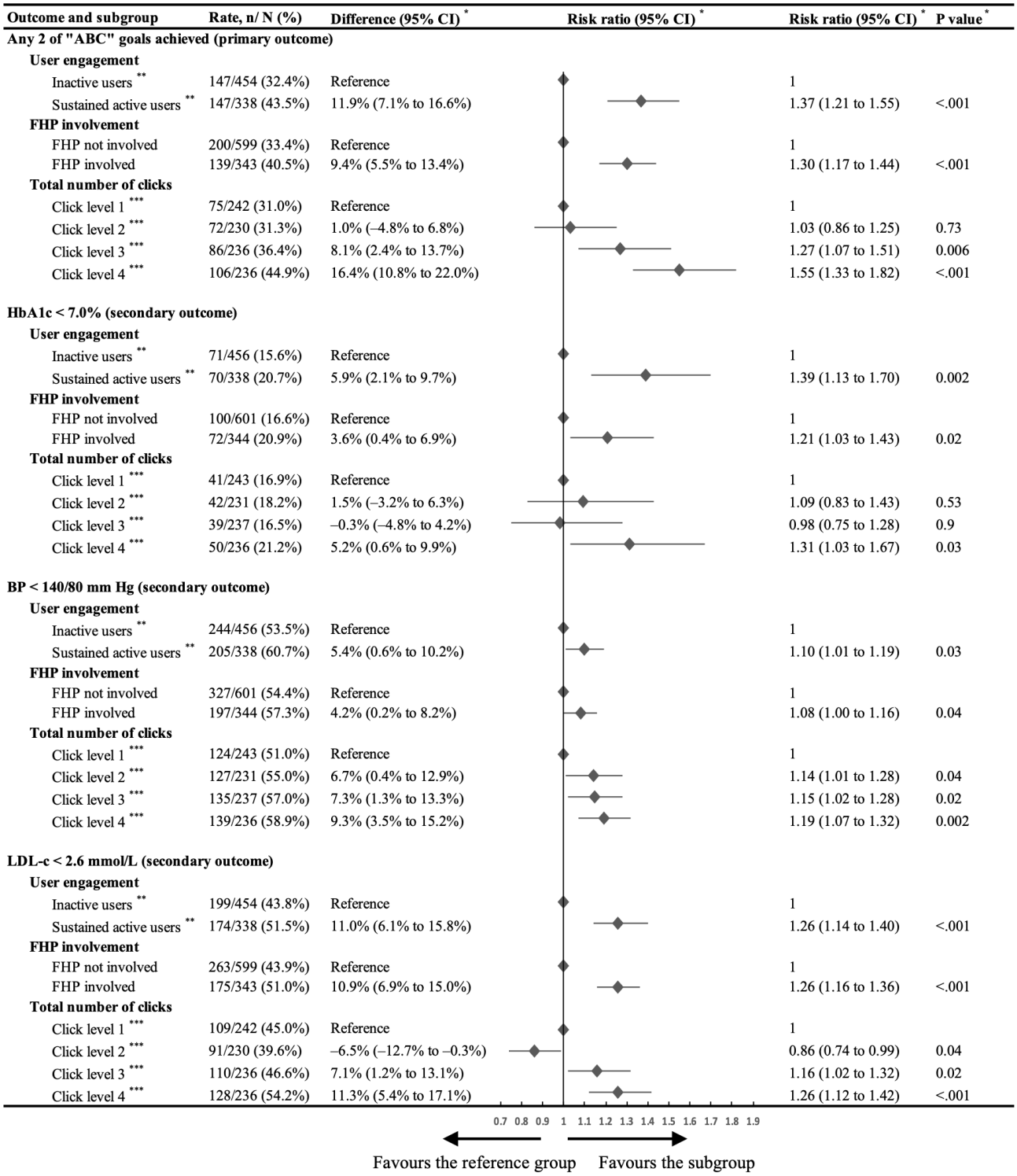
1. An expert review of the DBCI including app features (DBCI content, Figure 2B) and trial implementation activities (DBCI delivery, Figure 2A) based on the conceptual DBCI engagement framework by Michie and colleagues [6]. This framework conceptualizes the potential direct and indirect influences on engagement, as well as the relationships between engagement and the effectiveness of the interventions. The behavior change techniques (BCTs) adopted by our DBCI were identified against the internationally recognized Behavior Change Technique Taxonomy (V1) [23]. Two researchers (XT and PZ) independently conducted the review. With the majority in agreement, a small amount of disagreement was resolved by discussion and a consensus was reached.
2. In-depth interviews. Interviews with key stakeholders, including selected PHC providers, patients, and FHPs in both urban and rural settings, were conducted in June 2020 to deepen our understandings of factors contributing to user engagement with the app. Semistructured, user-type specific interview guides were developed by our multidisciplinary research team, piloted with a few participants, and then revised (see posttrial interview guides in Multimedia Appendix 1). Face-to-face, one-on-one in-depth interviews,

XSL•FO
RenderX

each lasting approximately 1 hour, were conducted in Mandarin Chinese by 2 team members (XT and PZ) in designated private areas. The interviews were audio-recorded and transcribed verbatim. Iterative thematic analysis was conducted by generating codes, forming themes, and further categorizing into 3 domains— DBCI content (Figure 2B), DBCI delivery (Figure 2A), and context (Figures 2-5)—based on the conceptual framework for engagement with DBCIs by Michie and colleagues [6]. During coding, the researchers were blinded to the overall trial outcomes. All analyses were performed with Nvivo12 (QSR International).

XSL•FO
RenderX

Figure 5. Forest plot of the effects of user engagement on main clinical outcomes. BP: blood pressure; FHP: family health promoter; HbA_{1c}: glycated hemoglobin; LDL-c: low-density lipoprotein cholesterol.



*: Inverse PS weighted model: log-binomial model with GEE and with adjustment of cluster with inverse propensity score weighted. Propensity Score was calculated according to differences in baseline sociodemographic characteristics, medical history and lab test results between subgroup users (in the intervention arm). Propensity scores were generated by a logit regression model with covariates of age, sex, locality, HBA1c, FBG, SBP, DBP, LDL-c, education levels, duration of diabetes category, and diabetic complications.

**: Sustained active users are defined as users who logged in the app for at least 12 months (at least once per month) or above, where inactive users are defined as users who logged in the app for less than 12 months.

***: The click levels are divided by quartiles of the total number of clicks of App functions No. 1-15 (Table 1) from July 2019 to April 2018. Level 1 (0–10 clicks, n = 243), Level 2 (11–75 clicks, n=231), Level 3 (76–315 clicks, n=237), and Level 4 (>315 clicks, n=236).

XSL•FO
RenderX

Triangulation and Synthesis of Quantitative and Qualitative Findings

All findings were integrated by matching the quantitative and qualitative results against the attributes and concepts—DBCI content, DBCI delivery, and context—of Michie's conceptual framework (Figure 2) in a joint display for further comparisons and metainferences [6].

Ethical Considerations

The SMARTDiabetes study was approved by the Peking University Health Sciences Institutional Review Board, Beijing, China, (IRB00001052-15062), and The University of Sydney Human Research Ethics Committee, Sydney, Australia (project no. 2016/105). The trial was registered at ClinicalTrials.gov (registration record NCT02726100, March 22, 2016). All participants provided written informed consent to the collection of individual-level trial assessment data and app usage data at the beginning of the trial. All interviewed participants provided verbal consent prior to interviews. All data were deidentified to ensure privacy and confidentiality. Participants were not financially compensated for taking part in the trial; however, all participants received free health examinations at baseline, midterm, and study completion.

Results

User Engagement

A total of 1038 participants from 40 clusters (20 rural villages and 20 urban communities) were assigned to the intervention arm. The mean MAUs proportion during the 16-month assessment period was 46.4% (SD 7.8%), with a gradual decline from 56.3% at the beginning (April 2018) to 42.2% at the end of follow-up (July 2019) (Figure 3). The proportion of sustained active users was 32.9% (342/1038 participants), with rural users showing a significantly higher rate than urban users (rural: 207/541 participants, 38.3% vs urban: 135/497 participants, 27.3%; $P<.001$).

After the initial period of training, installation, and familiarizing with the app, by May 2018, most users had been actively entering their FBG data into the app, with 98.9% (491/495) of rural and 87.4% (473/541) of urban users meeting the app-recommended monitoring frequencies (Figure 4). Throughout the entire trial period, the mean proportion of users meeting the recommended FBG-monitoring frequencies was 72.5% (SD 16.5%), with rural users outperforming their urban counterparts (rural 82.9%, SD 2.0% vs urban 61.6%, SD 3.6%; $P<.001$).

Table 1 summarized the monthly click-through rates per person for each of the 15 functions of the SMARTDiabetes app. The median total clicks per person per month was 3.9, with rural participants using more (rural 4.9 vs urban 3.5; $P=.32$). The most commonly used app functions were the 2 reminders "Diet punch card" and "Exercise punch card," followed by the 2 self-monitoring ones "Blood glucose" and "Blood pressure," all of which were incentivized functions.

XSL•FO
RenderX

Table . Number of clicks per person per month of each app function (1038 participants over 16 months of observation).

Functions	Urban (n=497)		Rural (n=541)		Total (N=1038)	
	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)
Chronic disease self-monitoring						
Blood glucose (self-monitoring, goal setting, and feedback)[a]	1.9 (3.5)	0.6 (0.1- 2.3)	1.2 (1.9)	0.5 (0.1-1.7)	1.6 (2.8)	0.5 (0.1-1.9)
Blood pressure (self-monitoring, goal setting, and feedback)[a]	1.5 (3.0)	0.4 (0.1-1.6)	0.8 (1.5)	0.3 (0.1-0.9)	1.1 (2.3)	0.4 (0.1-1.2)
Body weight (self-monitoring and feedback)[a]	0.5 (1.0)	0.1 (0-0.5)	0.4 (1.0)	0.1 (0-0.6)	0.4 (1.0)	0.1 (0-0.6)
Blood lipid (self-monitoring, goal setting, and feedback)	0.2 (0.7)	0 (0-0.3)	0.3 (1.1)	0 (0-0.3)	0.3 (0.9)	0 (0-0.3)
Healthy lifestyle reminder						
Diet (reminder)[a]	3.0 (8.6)	0.7 (0-2.1)	3.1 (6.2)	0.3 (0-2.7)	3.0 (7.5)	0.5 (0-2.2)
Exercise (reminder)[a]	2.8 (5.8)	0.6 (0-2.0)	2.9 (5.7)	0.3 (0-2.7)	2.8 (5.8)	0.6 (0-2.3)
Health records						
Health status summaries	0.6 (1.8)	0.1 (0-0.4)	1.1 (3.6)	0.1 (0-0.8)	0.9 (2.9)	0.1 (0-0.6)
Healthy lifestyles (self-monitoring and feedback)	0.1 (0.3)	0 (0-0.1)	0.2 (0.5)	0 (0-0.1)	0.1 (0.4)	0 (0-0.1)
Medications (medical record)	0.1 (0.3)	0 (0-0.1)	0.1 (0.2)	0 (0-0.1)	0.1 (0.2)	0 (0-0.1)
Physical examinations and laboratory tests (medical record)	0.1 (0.3)	0 (0-0.1)	0.1 (0.2)	0 (0-0.1)	0.1 (0.3)	0 (0-0.1)
Diabetic complications (medical record)	0.1 (0.3)	0 (0-0.1)	0.1 (0.3)	0 (0-0.1)	0.1 (0.3)	0 (0-0.1)
Risk assessment						
iCVD[b] risk assessments (feedback)[a]	0.4 (0.8)	0.1 (0-0.5)	0.5 (0.8)	0.1 (0-0.8)	0.5 (0.8)	0.1 (0-0.6)
Other						
Settings (system settings and reward system)[a]	1.2 (2.5)	0.3 (0-1.1)	1.0 (1.8)	0.3 (0.1-1.0)	1.1 (2.2)	0.3 (0-1.1)
Messages (communication)	0.7 (1.4)	0.1 (0-0.6)	0.9 (1.5)	0.2 (0-1.1)	0.8 (1.5)	0.1 (0-0.9)
Knowledge and skill bank (healthy literacy)	0.2 (0.5)	0 (0-0.2)	0.3 (0.5)	0.1 (0-0.3)	0.3 (0.5)	0 (0-0.3)
Total	13.4 (22.5)	3.5 (0.5-14.9)	12.9 (17.8)	4.9 (0.6-18.9)	13.2 (20.2)	3.9 (0.6-17.3)

XSL•FO
RenderX

[a]*Directly or indirectly incentivized functions.

[b] iCVD: ischemic cardiovascular diseases.

Associations Between App Engagement and Effectiveness

As shown in Figure 5, sustained active users and users with FHP involvement achieved significantly better results in primary outcome and all 3 secondary outcomes. Users with the fewest clicks (click level 1) had a 31.0% success rate in achieving any 2 of the "ABC" goals. This rate remained unchanged at 31.3% in click level 2 (RR 1.03, 95% CI 0.86-1.25; *P*=.73). However, click level 3 users showed a statistically significant improvement, with a 36.4% success rate (RR 1.27, 95% CI 1.07-1.51; *P*=.006), and those in click level 4 achieved an even higher 44.9% success rate (RR 1.55, 95% CI 1.33-1.82; *P*<.001). This gradient in health outcomes across the different levels of user engagement, as measured by the total number of clicks on the SMARTDiabetes app functions, was also evident in achieving each individual "ABC" goal at 24 months.

Rural-Urban Heterogeneity

Table 2 showed the disparities between rural and urban settings. Rural PHC facilities had fewer health care resources but were 100% (20/20) government-owned compared with 65% (13/20) of urban ones and were more distant from tertiary hospitals (13.3 km vs 1.9 km; *P*<.001). Rural participants were younger (*P*<.001) but had lower educational levels. Despite these differences, rural participants demonstrated higher engagement across all levels: a larger proportion of sustained active users (38.3% vs 27.2%; *P*<.001), greater FHP involvement (48.6% vs 21.5%; *P*<.001), more frequent interaction with the app (4.9 clicks per person per month vs 3.5; *P*=.32), and higher rates of self-monitoring FBG (82.9% vs 61.6%; *P*<.001).

XSL•FO
RenderX

Table . Rural-urban disparities.

Quantitative findings—characteristics	Urban	Rural	*P* value
PHC[a] facilities characteristics			
Number of participating facilities in the intervention arm	20	20	N/A[b]
Government owned facilities, n (%)	13 (65)	20 (100)	.003
Size of the community served, median (IQR)	6366 (5024-7899)	2497.5 (2022-3651)	<.001
Distance to the nearest tertiary hospital (km), mean (SD)	1.9 (1.1)	13.3 (6.7)	<.001
Staff per facility, median (IQR)	6.0 (3.5-12.5)	2.0 (2.0-4.0)	<.001
Registered doctor	2.5 (1.5-5.0)	1.5 (1.0-2.0)	.005
Registered nurse	3.0 (1.5-4.0)	0 (0-0)	<.001
Staff providing NCD[c] management services	3.0 (2.0-4.5)	1.5 (1.0-2.0)	<.001
Users baseline characteristics			
Number of participants in the intervention arm	497	541	N/A
Male, n (%)	217 (43.7)	232 (42.3)	.82
Age (years), mean (SD)	62.7 (6.7)	60.3 (7.4)	<.001
Education level, n (%)			
Primary school or lower	156 (31.4)	238 (44.0)	<.001
Junior high school	171 (34.4)	193 (35.7)	N/A
Senior high school	128 (25.8)	108 (20.0)	N/A
Junior college	28 (5.6)	1 (0.2)	N/A
Bachelor’s degree or higher	14 (2.8)	1 (0.2)	N/A
Engagement metrics			
Proportion of sustained active users, n/ N (%)	135/497 (27.2)	207/541 (38.3)	<.001
Proportion of users with FHP[d] involvement, n/ N (%)	92/427 (21.5)	252/518 (48.6)	<.001
Cohabitation rate of FHPs and patients, n/N (%)	54/92 (58.7)	192/252 (76.2)	<.001
Total clicks per person per month, median (IQR)	3.5 (0.5-14.9)	4.9 (0.6-18.9)	.32
Mean proportion of FBG[e] self-monitoring rate, mean (SD)	61.6% (3.6%)	82.9% (2.0%)	<.001

[a]PHC: primary health care.

[b]N/A: not applicable.

[c]NCD: noncommunicable disease.

[d]FHP: family health promoter.

[e]FBG: fasting blood glucose.

Influencing Factors of User Engagement

The rapid expert review of the SMARTDiabetes app features (DBCI content) suggested that 6 main features, corresponding to previously proven effective BCTs [23], were very likely to have contributed to the success of the trial. These features included (1) the incentive system (Reward), (2) real-time feedback on individual result interpretation and suggestions for improvement (Feedback), (3) in-app reminder messages and healthy lifestyle punch card (Reminders), (4) setting goals and targets (Goal setting), (5) self-monitoring tools (Self-monitoring), and (6) support from FHPs (Social support

XSL•FO
RenderX

features). The review of the SMARTDiabetes trial implementation activities (DBCI delivery) revealed several factors that have been proven effective [6], including partnerships with the local authorities (Credibility features), ongoing professional support from PHC providers (Professional support features), and a step-by-step, illustrative app instruction brochure (Guidance) (see Figure S3 in Multimedia Appendix 1 for the brochure). However, the complex installation process and frequent software updates were identified as major barriers to adoption and sustained engagement (Ease of use).

Twelve stakeholders (2 PHC providers, 7 patients, and 3 FHPs) were interviewed. The thematic analysis confirmed that rewards of nominal gifts, automated reminders, patient involvement in regular chronic disease self-monitoring, and adequate support from both FHPs and PHC providers were the most highly valued features of the SMARTDiabetes app. These features were perceived by the key stakeholders to be most influential in contributing to our app engagement. Furthermore, several key barriers to engagement were also identified in the interviews. First, a few patient users reported reluctance to involve their FHPs as it was perceived to lay an excessive burden to the FHPs, and this was more likely to occur if the FHPs had a busy daily life schedule or they lived far apart from the patient users. Technical difficulties related to the app installation and version updates were cited by several key stakeholders as a major barrier. Also, older participants, those with low literacy, and those inexperienced in smartphone use were the least likely to be active users. Qualitative findings also showed that in rural areas, the patient-doctor relationship is more personal and community-focused, PHC facilities are more committed to government initiatives, and FHPs are actively involved in patient care, whereas in urban areas, patients could bypass PHC facilities in favor of accessible tertiary hospitals.

Synthesis of Findings

Based on the conceptual framework of DBCI engagement by Michie and colleagues [6,23], our team constructed this program logic to suggest potential influences and pathways contributing to user engagement and trial effectiveness, according to (1) DBCI content, (2) DBCI delivery, and (3) context. Table 3 shows the joint display of synthesized mixed methods findings where key quantitative findings were juxtaposed with illustrative interview quotes.

XSL•FO
RenderX

Table . Summary of the quantitative and qualitative findings.

Concept	Attribute	SMARTDiabetes features	Dose/fidelity	Quantitative evidence of effectiveness	Qualitative evidence of effectiveness
DBCI[a] Content	Rewards	• Users accumulate reward points by using certain functions in the app (Table S1 in Multimedia Appendix 1). Each user can earn a maximum of 100 points per month. These points can then be redeemed for nominal gifts (valued 5 - 30 CNY[b]), distributed by their PHC[c] providers. The reward system was launched in April 2018 and had been active throughout the rest of the trial.	• In total, app users generated 982,005 reward points, with 93.2% (915,571/982,005) redeemed. In total, 97.0% (1006/1038) users have received gifts.	• The incentivized app functions were more commonly accessed than those without (Table 1). • After the introduction of the reward system in April 2018, there was a significant increase in the number of active users (Figure S4 in Multimedia Appendix 1). The rate of users meeting the recommended FBG[d]-monitoring frequencies increased from 77% to 99% in rural sites; and from 27% to 87% in urban sites in the following 2 months (Figure S3 in Multimedia Appendix 1).	• "I am delighted to have earned the top-ranked reward points in my group" (Patient user, urban). • "The monetary value of the reward isn't important. The gift serves as motivation for making positive changes" (Patient user, urban). • "The monetary value of the gift doesn't matter. As long as they get something, it should be effective [to motivate patient engagement" (PHC provider, urban). • "Patients will be happy even if they get the smallest gift, but they need to be able to redeem the gifts more frequently" (PHC provider, rural).
DBCI Content	Feedback	• The app provided real-time feedback based on user inputs after users self-recorded their FBG readings.	• The mean proportion of users meeting the recommended FBG self-monitoring and uploading frequencies was 72.5% (SD 16.5%).	N/A[e]	N/A

XSL•FO
RenderX

Concept	Attribute	SMARTDiabetes features	Dose/fidelity	Quantitative evidence of effectiveness	Qualitative evidence of effectiveness
DBCI Content	Reminders	• Users received in-app messages reminders to encourage self-monitoring of FBG and BP[f], along with alerts for potential risks of unhealthy practices. The messages were categorized into 4 genres: urgent, serious, self-management, and system notifications. The "Diet" and "Exercise" functions set reminders to adhere to a healthy diet and daily exercise through a digital register.	• In total, 57.7% (291/504) urgent messages, 14.3% (42,625/298,873) serious messages, 6.4% (35,103/550,912) self-management messages, and 15.1% (1565/10,341) system notifications were read by the users. A total of 58,581 click-throughs to the "Diet" function and 55,126 to the "Exercise" function (Table 1).	N/A	• "The app sends reminders if you forget to measure your FBG or BP. When you receive the warning messages, it serves as a reminder, which is helpful, especially for elderly users who may forget at times. The "Diet" Da Ka (punch card) register advises me daily to consume less flour and starch, which I find beneficial. I'm particularly keen on the "Exercise" feature, where I can record my daily physical activity" (Patient user, urban).
DBCI Content	Goal setting	• On the app's main display screen, the latest FBG, BP, and LDL-c[g] data were shown with a visually memorable color grading scheme: good results were marked in green, moderate ones in yellow, and poor ones in red (Figure 1A). The "ABC" goals were set to achieve FPG[h] level below 7.0 mmol/L; BP readings below 140/80 mm Hg; and LDL-c levels below 100 mg/dL or 2.6 mmol/L.	• Each time at login, users could review their 3 "ABC" goals displayed upfront on main screen.	N/A	N/A
	Self-monitoring	• The app encourages users to self-monitor and upload their FBG and BP readings.	• Throughout the trial, the mean frequency of FBG measurements was 1.50 (SD 0.39) times per month per person and BP levels were 1.45 (SD 0.39) times per month per person.	N/A	

XSL•FO
RenderX

Concept	Attribute	SMARTDiabetes features	Dose/fidelity	Quantitative evidence of effectiveness	Qualitative evidence of effectiveness
					• "I monitored my BP and FBG levels more frequently after using the app" (Patient user, urban). • "I used the BP and FBG functions the most (to upload and record BP and FBG levels)" (Patient user, urban).
DBCI Content	Social support features	• Family members played an active role in assisting patients with diabetes management.	• In the intervention arm, 36.4% (344/945) of participant users reported having a family member who was able to assist them with the app (Table 2—Proportion of users with FHP[i] involvement). • The majority of FHPs were the patients' adult children (78.8%), followed by their spouses (14.2%). Most FHPs were living together with the patient users (71.5%) (Table S3 in Multimedia Appendix 1).	• FHP involvement was associated with better primary clinical outcomes (40.5% vs 33.4%; RR[j] 1.30, 95% CI 1.17-1.44; *P*<.001), as well as better secondary clinical outcomes, including HbA_{1c}[k] (20.9% vs 16.6%; *P*=.02), BP (57.3% vs 54.4%; *P*=.04), and LDL-c (51.0% vs 43.9%, *P*<.001) (Figure 5). • Of the various support and assistance provided by the patient users' FHPs, reminding of medication uptake (66.0%), dietary control (64.8%), and exercise (57.3%) were the most common. This was closely followed by accompanying patient users to seek medical treatment when needed (50.3%) (Table S3 in Multimedia Appendix 1).	• Facilitators: "FHPs supervise the patients. Sometimes they buy medicine for them" (PHC provider, rural). • "My wife supported me in using this app. She learned how to use it first and then taught me. I am not good at using smartphones, and I tend to forget how to use the app" (Patient user, urban). • "I reminded the patient to take the medication" (FHP user, rural). • Barriers: "My children are busy, so I don't want to bother them" (Patient user, urban). • "FHPs are busy" (PHC provider, rural). • "My children come home about once per year" (Patient user, rural).
DBCI Delivery	Credibility features		• At each app login	N/A	N/A

XSL•FO
RenderX

Concept	Attribute	SMARTDiabetes features	Dose/fidelity	Quantitative evidence of effectiveness	Qualitative evidence of effectiveness
		• The app and the brochure (Figure S3 in Multimedia Appendix 1) displayed main local health authority's (our study partners') logos (the Hebei Medical University and Shijiazhuang Center for Disease Control and Prevention).			
	Professional support	• An initial app induction about downloading, installation, and key functions was followed by continuous support from pretrained PHC providers.	• PHC providers in the intervention arm were trained to download, install, update, and skillfully use every function of the app ("train the trainer").	N/A	• "The app made my communication with my doctor more frequent and more convenient. Our relationship became closer" (Patient user, urban). • "I tend to forget how to use the app. I need help from the doctor so that I can continue to use the app" (Patient user, urban). • "I gathered all patients and FHPs in my clinic and talked with them about the app. If they can't make it, I went to organise other teaching chances at another time" (PHC provider, rural).
DBCI Delivery	Guidance	• A detailed, easy-to-understand instruction brochure (app operation manual), accompanied by a set of picture-based illustrative PowerPoint slides, was developed (Figure S1 in Multimedia Appendix 1).	• Written app operation instruction materials were adequately distributed to all participating PHC providers, which further facilitated their patient user training.	N/A	N/A
	Ease of use		• App retraining had to be repeated every time there were major updates.	N/A	

XSL•FO
RenderX

Concept	Attribute	SMARTDiabetes features	Dose/fidelity	Quantitative evidence of effectiveness	Qualitative evidence of effectiveness
		• The app installation and updating process was perceived NOT to be easy to follow. Users were initially instructed to scan a QR code, which then directed them to go through a few steps to gain the system access permission before they could download.			• "Upgrading is challenging for patients as they need to download first and then find the file and obtain permission to access the app...It's too complicated when the upgrades are very frequent" (PHC provider, urban). • "I don't understand the "iCVD risk" function as it is too complicated" (FHP user, rural).
Context	Population	• Older people, those with low education literacy, and those with low smartphone experiences were the least likely to engage with the app.	N/A	• The mean age of the patient users in the intervention arm was 61.4 years (SD 7.1), with 62.8% (652/1038) aged 60 years or older. Close to half (758/1038, 73.0%) of the patient users did not complete the Chinese 9-year compulsory education (ie, not completing senior high school).	• "I feel that older people may struggle to use the app due to lower education levels, making it difficult for them to learn new skills. This is why patients in rural areas often rely more on their FHPs, while urban patients may not face the same challenges" (PHC provider, urban). • "I am old, and it's hard for me to accept new things. Our generation has only recently started using smartphones. We need to take it slowly" (Patient user, urban).
Context	Setting (rural vs urban)		N/A	• Refer to Table 2	

XSL•FO
RenderX

Concept	Attribute	SMARTDiabetes features	Dose/fidelity	Quantitative evidence of effectiveness	Qualitative evidence of effectiveness
		• While the SMARTDiabetes intervention demonstrated significant improvement in the proportion of participants achieving at least 2 American Diabetes Association "ABC" management targets compared with usual care (intervention group: 35.9% vs usual care: 29.9%; *P*=.025), we observed significant heterogeneity in the primary outcome between rural (intervention group: 42.6% vs usual care: 31.0%) and urban (intervention group: 27.9% vs usual care: 28.6%) settings (*P*=.02 for heterogeneity).			

XSL•FO
RenderX

Concept	Attribute	SMARTDiabetes features	Dose/fidelity	Quantitative evidence of effectiveness	Qualitative evidence of effectiveness
					• Closer patient-doctor relationship in rural areas: In rural areas, the patient-doctor relationship often resembles that of close-knit neighborhoods, characterized by familiarity and reciprocity. In contrast, in urban areas, personal contact is often more distant, with the patient-doctor relationship largely shaped by the PHC provider's professional expertise and communication skills. • Stronger commitment of PHC facilities in rural areas: Government-owned or public-funded PHC facilities tend to show greater commitment to government-endorsed initiatives, while private-owned facilities, more common in urban areas, often prioritize profit-driven activities. • Bypassing the PHC facilities in urban areas: In urban areas, where tertiary hospitals are much closer, patients often bypass PHC facilities, opting instead to seek health care directly at nearby hospitals without needing a referral. This practice potentially undermines the effectiveness of PHC providers in delivering continuous care to patients with chronic conditions. • More FHP involvement in the

XSL•FO
RenderX

Concept	Attribute	SMARTDiabetes features	Dose/fidelity	Quantitative evidence of effectiveness	Qualitative evidence of effectiveness
					rural areas: FHPs, primarily patients' children and spouses, played a vital role in patient care. They assisted with app use, including reminding patients to take medications, supporting dietary management, encouraging regular exercise, and often accompanying patients' hospital visits for scheduled medical appointments. While both rural and urban patients reported reluctance to ask for help from FHPs, rural patients generally found it easier due to their closer proximity to family members.

[a]DBCI: digital behavior change intervention.

[b]1 CNY = US $0.1514, based on the average conversion rate in 2018.

[c]PHC: primary health care.

[d]FBG: fasting blood glucose.

[e]N/A: not applicable.

[f]BP: blood pressure.

[g]LDL-c: low-density lipoprotein cholesterol.

[h]FPG: fasting plasma glucose.

[i]FHP: family health promoter.

[j]RR: relative risk.

[k]HbA_{1c}: glycated hemoglobin.

Discussion

Principal Findings

The SMARTDiabetes trial was the first long-term (over 1 year) mobile app intervention for T2DM self-management conducted in a real-world PHC setting. Throughout the whole trial period, 46.4% of participants were MAUs and 32.9% (342/1038) were sustained active users. We observed a dose-response relationship between user engagement and trial effectiveness. Specifically, sustained engagement with the app, involvement of family members, and frequent interaction with app functions were associated with improved clinical outcomes. Key drivers of user engagement and trial effectiveness included 6 known BCTs underpinning the 15 app functions (ie, reward, feedback, reminders, goal setting, self-monitoring, and social support features), 4 implementation facilitators (credibility features, professional support, guidance, and ease of use), and 5 contextual contributors (younger age, better education, greater smartphone literacy, closer patient-doctor relationship, and stronger PHC facilities commitment).

We observed a modest yet effective level of engagement with the SMARTDiabetes app, with half of the participants logging in at least once per month, leading to significant improvement in achieving at least 2 ADA "ABC" management targets [5]. Effective engagement with DBCI is empirically defined as the level of interaction necessary to achieve the intended health outcomes [11]. It involves both "microengagement," which refers to the intensity and consistency of app use, and "macroengagement," which pertains to the alignment with and commitment to the broader intervention goals. Both forms of engagement are closely intertwined, and while each is important, neither alone is sufficient for achieving effective outcomes. Previous DBCIs for diabetes primarily focused on patients and doctors only [12-15]. However, the SMARTDiabetes mHealth

XSL•FO
RenderX

system, consisting of 2 apps and 1 website, fostered stronger connections among patients, family members, PHC providers, and local governors (Figure S1 in Multimedia Appendix 1). It enabled remote family members to stay informed about their loved ones' health and facilitated lay FHPs in delivering evidence-based, guideline-informed care (eg, FHPs helped remind patients to take medications, supported dietary management, encouraged regular exercise, and accompanied patients to medical appointments; Table S3 in Multimedia Appendix 1). The system also encouraged patients to actively engage with their doctors through in-app messages, phone calls, and face-to-face interactions during training and the gift redemption process, addressing a key barrier to diabetes management at the PHC level in China [21]. Additionally, it provided governors with real-time access to validated health statistics and powerful data visualization tools, enabling them to track changes, make informed comparisons, and drive performance-based improvements in the quality of care. These macrolevel improvements in engagement were hard to measure yet indispensable to the trial's overall success.

Comparing engagement levels across DBCI studies is challenging and often inconclusive due to the heterogeneity in intervention types, implementation duration, delivery modes, population demographics, study settings, intended behavioral objectives, and measures of engagement [24]. Standardized checklists to describe engagement with DBCIs will assist with comparing engagement across studies. The observed nonengagement rate (170/1038, 16.4%) in our study compares favorably with previous studies, which reported nonengagement rates of 50.2% (101/201) [14] and 57.2% (16,958/29,643) [25]. The SMARTDiabetes trial's user engagement compared favorably at 6, 9, and 12 months to 3 similar studies conducted in real-world settings with large populations—the Cornerstones4Care study reported a 30.5% engagement rate in 6 months [25], the National Health Service Digital Diabetes Prevention Programme study had a 36.5% engagement rate in 9 months [26], and the myDESMOND study showed a 17.6% engagement rate in 12 months [27].

The decline in app engagement over time has been commonly reported in mHealth interventions [18,28]. We also noticed a gradual decline in engagement in our study, with the proportion of MAUs decreasing by 25.0%, from 56.3% to 42.2% over 16 months, resulting in a retention rate of 75.0%. In comparison, the Cornerstones4Care study reported a 71.4% (9051/12,685) retention rate over 6 months [25], the National Health Service Digital Diabetes Prevention Programme study showed a 54.2% (667/1230) retention rate over 9 months [26], and the myDESMOND study had a 31.3% (1676/5360) retention rate over 12 months [27].

Previous research has demonstrated a strong link between user engagement with DBCIs and their effectiveness in achieving intended health outcomes [9,29]. The dose-response relationship between the total number of app clicks and improvements in the primary outcome suggests that DBCIs require a minimum threshold of engagement—an optimal dose—to achieve intended outcomes [6]. Beyond this threshold, the intensity of engagement has a positive correlation with the extent of outcome improvements. However, due to the complexity of engagement, defining the optimal dose is challenging, as some individuals may require a longer period of interaction with the app than others [30]. It is also plausible that the individuals who are more motivated and successful in achieving behavioral change tend to engage more, or that those who engage more differ systematically from those who are less engaged [29]. Assessing the extent to which engagement influences behavioral and health-related outcomes is challenging, as it may be confounded by factors such as increased motivation and self-regulation skills [11]. Future research should investigate these potential confounders to draw more definitive conclusions.

The design of app features (ie, DBCI content) plays a crucial role in user engagement, and 6 effective BCTs were repeatedly associated with higher engagement levels [6,9]. In our study, we found that incentive features (ie, "rewards") [11] were highly effective both in boosting engagement at the initial adoption phase (ie, "trigger") and in sustaining long-term use. Our findings also suggest that the impact of rewards is believed to rely less on their monetary value and more on their distribution frequencies (instant positive reinforcement). The most frequently used functions were the diet and exercise punch cards, both of which integrated key aspects of gamification and loyalty. Gamification is a promising strategy for enhancing user engagement [31], and the commercial sector has excelled in implementing these concepts. For instance, Duolingo has used game elements such as leaderboard, level system, and badges to motivate users, achieving a staggering 100 million MAUs [32]. Academia can benefit from studying these successful models to improve the design of DBCIs.

One important lesson we learned was that the SMARTDiabetes app was not user-friendly when it came to downloading, installing, and updating. The app was unavailable on public app markets, requiring users to scan a QR code and download it via a third-party website. After downloading, users had to manually grant permissions for the app and complete the installation process. Moreover, with every update, users were forced to repeat the entire procedure. This cumbersome process likely contributed to a significant loss of active users [17]. Future mobile app interventions should focus on integrating the installation and updating processes with commonly used app stores.

Finally, context matters in implementation science research [33]. Our previous study showed that the same mHealth intervention package—using identical intervention tools and delivery strategies—was significantly more effective in rural settings than in urban settings [5]. The quantitative findings of this process evaluation showed that rural participants demonstrated higher engagement with the SMARTDiabetes program, reflected in a higher proportion of sustained active users (38.3% vs 27.2%; $P<.001$), more frequent app interactions (4.9 clicks per person per month vs 3.5; $P=.32$), and a higher rate of FBG self-monitoring (82.9% vs 61.6%; $P<.001$). Our qualitative findings indicated that higher engagement in rural settings was attributed to 3 main factors: closer and more reciprocal patient-doctor relationships, stronger commitment from PHC facilities, and greater involvement from the family members (Table 3). Together, these factors created a more

XSL•FO
RenderX

supportive environment for patients' sustained participation in the SMARTDiabetes program.

Strengths and Limitations

The strength of this study is its mixed methods design to capture multifaceted engagement with DBCIs. Study limitations include the following: First, the study's geographical scope was limited to Northern China, potentially constraining the generalizability of its findings to this specific health system context. However, this region's PHC infrastructure and services, based on the National Essential Public Health Service Programme, are representative of many areas across China. Consequently, we posit that these findings offer valuable insights applicable to other regions. Second, the total number of interviews conducted was less than anticipated, due to the difficulties of face-to-face interaction during the COVID outbreaks. Consequently, the views expressed by limited participants may not be representative of the total study population. Third, since the study was conducted a few years ago, there may have been changes over time. However, the expected impact of such changes is likely minimal, given that the core policies and regulations for diabetes and hypertension management have remained unchanged since 2009.

Conclusions

Long-term, sustained engagement with a mobile app intervention for T2DM self-management is achievable in the PHC setting in China. Our study reveals a dose-response relationship between user engagement with DBCIs and their effectiveness in achieving intended health outcomes. Specifically, DBCIs require a minimum level of engagement to be effective. The factors influencing engagement identified in the SMARTDiabetes study provide valuable insights for developing targeted strategies to enhance user engagement and improve overall program effectiveness.

Acknowledgments

The authors are grateful to Professors Linong Ji and Junfeng Xu for their contributions to the development of their mHealth intervention. The authors thank Xian Li, Fengzhuo Chen, Qiang Li, and Laurent Billot for their assistance in developing the statistical analysis plan and for checking the SAS program. The authors also extend their thanks to the patients, their relatives, and the doctors who participated in this study.

Funding

This study is supported by the Australian National Health and Medical Research Council and Global Alliance for Chronic Diseases (ID 1094712).

Data Availability

Anonymized individual-level data and datasets generated or analyzed during this study, together with the data dictionary and study protocol, are available for researchers who provide a methodologically sound proposal. Proposals should be directed to Professor Puhong Zhang (zpuhong@georgeinstitute.org.cn) and Professor David Peiris (dpeiris@georgeinstitute.org). The data will be available beginning 3 months after publication of this paper, with no end date. Any cross-border transfer of individual data must obtain official approval following legal procedures, in accordance with relevant regulations.

Authors' Contributions

XT, PZ, and DP conceived and designed the study. XT, YM, and YZ were pivotal in project management and trial implementation. XT and PZ were responsible for data collection. XT, HS, and PZ were responsible for quantitative data analyses. XT, LM, and PZ were responsible for qualitative data analyses. XT completed the draft of this paper. PZ, DP, and LM made substantial edits to the draft. YM and YZ provided critical suggestions for the draft. All coauthors approved the submission of this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Intervention design, app algorithms, participant characteristics, and study implementation.
[DOCX File, 1087 KB - mhealth_v14i1e76699_app1.docx]

Checklist 1
GRAMMS (Good Reporting of a Mixed Methods Study) checklist.
[DOCX File, 14 KB - mhealth_v14i1e76699_app2.docx]

References

1. Li X, Lu J, Hu S, et al. The primary health-care system in China. Lancet 2017 Dec 9;390(10112):2584-2594. [doi: 10.1016/S0140-6736(17)33109-4] [Medline: 29231837]

XSL•FO
RenderX

2. Wang L, Peng W, Zhao Z, et al. Prevalence and treatment of diabetes in China, 2013-2018. JAMA 2021 Dec 28;326(24):2498. [doi: 10.1001/jama.2021.22208]
3. Liu J, Liu M, Chai Z, et al. Projected rapid growth in diabetes disease burden and economic burden in China: a spatio-temporal study from 2020 to 2030. Lancet Reg Health West Pac 2023 Apr;33:100700. [doi: 10.1016/j.lanwpc.2023.100700] [Medline: 36817869]
4. Peiris D, Sun L, Patel A, et al. Systematic medical assessment, referral and treatment for diabetes care in China using lay family health promoters: protocol for the SMARTDiabetes cluster randomised controlled trial. Implementation Sci 2015 Dec;11(1):116. [doi: 10.1186/s13012-016-0481-8]
5. Zhang P, Tao X, Ma Y, et al. Improving the management of type 2 diabetes in China using a multifaceted digital health intervention in primary health care: the SMARTDiabetes cluster randomised controlled trial. Lancet Reg Health West Pac 2024 Aug;49:101130. [doi: 10.1016/j.lanwpc.2024.101130] [Medline: 39056088]
6. Michie S, Perski O, Blandford A, West R. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
7. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232. [doi: 10.2196/jmir.7126] [Medline: 28663162]
8. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152. [doi: 10.2196/jmir.2104] [Medline: 23151820]
9. Milne-Ives M, Homer SR, Andrade J, Meinert E. Potential associations between behavior change techniques and engagement with mobile health apps: a systematic review. Front Psychol 2023;14:1227443. [doi: 10.3389/fpsyg.2023.1227443]
10. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567. [doi: 10.2196/14567] [Medline: 31573916]
11. Yardley L, Spring BJ, Riper H, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
12. Lim SL, Tay MHJ, Ong KW, et al. Association between mobile health app engagement and weight loss and glycemic control in adults with type 2 diabetes and prediabetes (D'LITE Study): prospective cohort study. JMIR Diabetes 2022 Sep 30;7(3):e35039. [doi: 10.2196/35039] [Medline: 36178718]
13. Adu MD, Malabu UH, Malau-Aduli AE, Drovandi A, Malau-Aduli BS. User retention and engagement with a mobile app intervention to support self-management in Australians with type 1 or type 2 diabetes (My Care Hub): mixed methods study. JMIR Mhealth Uhealth 2020 Jun 11;8(6):e17802. [doi: 10.2196/17802] [Medline: 32525491]
14. Batch BC, Spratt SE, Blalock DV, et al. General behavioral engagement and changes in clinical and cognitive outcomes of patients with type 2 diabetes using the Time2Focus mobile app for diabetes education: pilot evaluation. J Med Internet Res 2021 Jan 20;23(1):e17537. [doi: 10.2196/17537] [Medline: 33470947]
15. Liu Z, Wang C, Yang D, et al. High engagement in mobile peer support is associated with better glycemic control in type 1 diabetes: a real-world study. J Diabetes Investig 2022 Nov;13(11):1914-1924. [doi: 10.1111/jdi.13870] [Medline: 35708894]
16. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and Validity. Psychiatr Serv 2019 Jul 1;70(7):538-544. [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]
17. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Mental Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891]
18. Van Rhoon L, Byrne M, Morrissey E, Murphy J, McSharry J. A systematic review of the behaviour change techniques and digital features in technology-driven type 2 diabetes prevention interventions. Digit Health 2020;6:2055207620914427. [doi: 10.1177/2055207620914427] [Medline: 32269830]
19. Cole-Lewis H, Ezeanochie N, Turgiss J. Understanding health behavior technology engagement: pathway to measuring digital behavior change interventions. JMIR Form Res 2019 Oct 10;3(4):e14052. [doi: 10.2196/14052] [Medline: 31603427]
20. O'cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy 2008 Apr;13(2):92-98. [doi: 10.1258/jhsrp.2007.007074]
21. Tao X, Mao L, Zhang P, et al. Barriers and facilitators to primary care management of type 2 diabetes in Shijiazhuang City, China: a mixed methods study. BMC Prim Care 2024 Mar 13;25(1):84. [doi: 10.1186/s12875-024-02330-7] [Medline: 38481166]
22. Wu Y, Liu X, Li X, et al. Estimation of 10-year risk of fatal and nonfatal ischemic cardiovascular diseases in Chinese adults. Circulation 2006 Nov 21;114(21):2217-2225. [doi: 10.1161/CIRCULATIONAHA.105.607499] [Medline: 17088464]
23. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]

XSL•FO
RenderX

24. Short CE, DeSmet A, Woods C, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292. [doi: 10.2196/jmir.9397] [Medline: 30446482]
25. Böhm AK, Jensen ML, Sørensen MR, Stargardt T. Real-world evidence of user engagement with mobile health for diabetes management: longitudinal observational study. JMIR Mhealth Uhealth 2020 Nov 6;8(11):e22212. [doi: 10.2196/22212] [Medline: 32975198]
26. Hawkes RE, Miles LM, Ainsworth B, Ross J, Meacock R, French DP. Engagement with a nationally-implemented digital behaviour change intervention: usage patterns over the 9-month duration of the National Health Service Digital Diabetes Prevention Programme. Internet Interv 2023 Sep;33:100647. [doi: 10.1016/j.invent.2023.100647] [Medline: 37502122]
27. Barker MM, Chauhan R, Davies MJ, et al. User retention and engagement in the digital-based diabetes education and self-management for ongoing and newly diagnosed (myDESMOND) program: descriptive longitudinal study. JMIR Diabetes 2023 Jul 21;8(1):e44943. [doi: 10.2196/44943] [Medline: 37477963]
28. Keller R, Hartmann S, Teepe GW, et al. Digital behavior change interventions for the prevention and management of type 2 diabetes: systematic market analysis. J Med Internet Res 2022 Jan 7;24(1):e33348. [doi: 10.2196/33348] [Medline: 34994693]
29. Perski O, Blandford A, Garnett C, Crane D, West R, Michie S. A self-report measure of engagement with digital behavior change interventions (DBCIs): development and psychometric evaluation of the "DBCI Engagement Scale". Transl Behav Med 2020 Feb 3;10(1):267-277. [doi: 10.1093/tbm/ibz039] [Medline: 30927357]
30. Szinay D, Perski O, Jones A, Chadborn T, Brown J, Naughton F. Perceptions of factors influencing engagement with health and well-being apps in the United Kingdom: qualitative interview study. JMIR Mhealth Uhealth 2021 Dec 16;9(12):e29098. [doi: 10.2196/29098] [Medline: 34927597]
31. Bitrián P, Buil I, Catalán S. Enhancing user engagement: the role of gamification in mobile apps. J Bus Res 2021 Aug;132:170-185. [doi: 10.1016/j.jbusres.2021.04.028]
32. Huynh D, Zuo L, Iida H, editors. Analyzing gamification of "Duolingo" with focus on its course structure. Presented at: International Conference on Games and Learning Alliance; Dec 5-7, 2016. [doi: 10.1007/978-3-319-50182-6_24]
33. Nilsen P, Bernhardsson S. Context matters in implementation science: a scoping review of determinant frameworks that describe contextual determinants for implementation outcomes. BMC Health Serv Res 2019 Mar 25;19(1):189. [doi: 10.1186/s12913-019-4015-3] [Medline: 30909897]

Abbreviations

BCT: behavior change technique
BP: blood pressure
DBCI: digital behavior change intervention
FBG: fasting blood glucose
FHP: family health promoter
HbA_{1c}: glycated hemoglobin
iCVD: ischemic cardiovascular disease
MAU: monthly active user
mHealth: mobile health
PHC: primary health care
RR: relative risk
T2DM: type 2 diabetes mellitus

Edited by L Buis; submitted 29.Apr.2025; peer-reviewed by R Jamison, S Parker; revised version received 27.Sep.2025; accepted 09.Feb.2026; published 03.Mar.2026.

Please cite as:
Tao X, Peiris D, Ma Y, Zhang Y, Song H, Mao L, Zhang P
Real-World Long-Term Engagement With a Mobile App Intervention to Improve Self-Management of Type 2 Diabetes Mellitus in China (SMARTDiabetes): Mixed Methods Study
JMIR Mhealth Uhealth 2026;14:e76699
URL: https://mhealth.jmir.org/2026/1/e76699
doi:10.2196/76699

© Xuanchen Tao, David Peiris, Yuxia Ma, Yaoshen Zhang, Hongyi Song, Limin Mao, Puhong Zhang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 3.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use,

XSL•FO
RenderX

distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Cognitive Behavioral Therapy for Chronic Insomnia in South Korea: Cost-Effectiveness Analysis Using Decision Tree and Markov Modeling Based on a Secondary Analysis of a Randomized Clinical Trial

Jung Hyun Kim[1*], PhD; Minjeong Sohn[2*], MSc; Jimin Woo[3], MA; Jaeyong Shin[4], MD, PhD; Euna Han[2], PhD

[1]Division of Tourism and Wellness, Hankuk University of Foreign Studies, Yongin, Republic of Korea

[2]College of Pharmacy, Yonsei Institute of Pharmaceutical Sciences, Yonsei Institute for Digital Health, Yonsei University, Incheon, Republic of Korea

[3]AIMNEXT Inc, Seoul, Republic of Korea

[4]College of Medicine, Yonsei University, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Euna Han, PhD
College of Pharmacy
Yonsei Institute of Pharmaceutical Sciences, Yonsei Institute for Digital Health
Yonsei University
Yonsei International Campus, Veritas Hall D, Rm 212
Incheon, 03722
Republic of Korea
Phone: 82 327494511
Email: eunahan@yonsei.ac.kr

Abstract

Background: Insomnia is a prevalent sleep disorder characterized by difficulty initiating or maintaining sleep and is associated with substantial health and economic burdens. Although cognitive behavioral therapy (CBT) is recommended as the first-line treatment, pharmacotherapy remains widely used despite adverse effects and significant indirect costs related to impaired productivity and workplace safety. Digital therapeutics delivering CBT through mobile platforms have emerged as scalable alternatives to improve access and outcomes. Somzz is a commercially available, domestically developed digital therapeutic that delivers CBT-based interventions for insomnia via a mobile app.

Objective: This study evaluated the cost-effectiveness of Somzz compared with conventional insomnia treatment combining CBT and pharmacotherapy from both health care system and societal perspectives in South Korea.

Methods: A decision-analytic model integrating a short-term decision tree with a Markov model was developed to compare costs and outcomes of digital CBT via Somzz versus conventional care in 2023. The model simulated a 27-week time horizon (three treatment cycles) and applied an annual discount rate of 4.5%. Clinical inputs, including remission probabilities and health utility values, were derived from a published randomized clinical trial comparing digital CBT delivered via Somzz with sleep hygiene education. Additional inputs, including health care resource use and unit costs, were obtained from published literature and national sources. Health outcomes were measured in quality-adjusted life years (QALYs). The cost analysis included direct medical costs and indirect costs related to absenteeism, productivity loss, and workplace accidents attributable to insomnia. Incremental cost-effectiveness ratios (ICERs) were estimated in 2023 South Korean Won (KRW). Deterministic one-way and probabilistic sensitivity analyses were conducted to assess uncertainty.

Results: From a health care system perspective, digital CBT via Somzz resulted in modestly higher costs and improved health outcomes compared with standard care. Over approximately 6.5 months, Somzz generated an additional 0.0092 QALYs per patient at an incremental cost of KRW 79,691 (US $61.87), yielding an ICER of KRW 8,719,727 (US $990,883) per QALY gained. This estimate was well below the Korean willingness-to-pay threshold of KRW 30,000,000 (US $23,192.91) per QALY. From a societal perspective, digital CBT was cost-saving, producing a negative ICER due to reductions in health care utilization, workplace accidents, and productivity losses associated with higher remission rates. Sensitivity analyses identified intervention

XSL•FO
RenderX

costs and remission probabilities as key drivers; however, digital CBT remained cost-effective across all scenarios under willingness-to-pay thresholds of KRW 30,000,000 and KRW 15,000,000 per QALY.

Conclusions: Digital CBT for insomnia offers favorable clinical and economic value in South Korea. Using Korean clinical trial data and locally relevant societal cost inputs, this study provides policy-relevant evidence supporting early integration of digital CBT into routine insomnia care, employer health strategies, and national digital health policy.

Trial Registration: World Health Organization International Clinical Trials Registry Platform KCT0007292; https://trialsearch.who.int/Trial2.aspx?TrialID=KCT0007292

(JMIR Mhealth Uhealth 2026;14:e71750) doi:10.2196/71750

KEYWORDS

insomnia; digital therapeutics; cognitive behavioral therapy; cost-effectiveness analysis; QALYs; quality-adjusted life years

Introduction

Problem

Insomnia is a prevalent sleep disorder among adults, characterized by difficulty initiating or maintaining sleep or early morning awakening despite adequate opportunity for sleep, resulting in poor sleep quality and daytime dysfunction [1-3]. Evidence suggests that the global prevalence of insomnia has increased, particularly during the COVID-19 pandemic, with marked rises among health care workers and individuals infected with COVID-19 [4,5]. Insomnia is associated with numerous adverse health outcomes, including a higher risk of psychiatric disorders and diminished health-related quality of life [6-8]. Furthermore, insomnia may lead to decreased work productivity and a higher risk of motor vehicle accidents [9,10]. Consequently, insomnia imposes substantial health and economic burdens, including direct costs related to health care use and medication use, as well as indirect costs from absenteeism and reduced workplace productivity [6,11].

Several treatment guidelines recommend cognitive behavioral therapy (CBT) as a first-line treatment for chronic insomnia [2,12-14]. Pharmacotherapy-centered management, despite its widespread use, is associated with adverse events, limited long-term effects, and continued health care use, while often failing to restore daily functioning and work performance [15,16]. Nevertheless, pharmacotherapy remains the most commonly used treatment in the United States and other countries, including Korea [17-19]. A survey of Korean physicians reported that pharmacotherapy was chosen by 57.14% of patients, followed by sleep hygiene education (37.11%), whereas only 1.18% received CBT [17]. This discrepancy persists despite strong evidence that CBT provides durable benefits that can be maintained for more than 1 year after treatment completion, in contrast to the short-term symptom control typically observed with pharmacotherapy [20]. As a result, indirect societal costs, such as productivity loss and accident-related costs, frequently exceed direct medical expenditures in insomnia [6,21,22].

Review of Relevant Scholarship

Digital cognitive behavioral therapy for insomnia (dCBT-I) has emerged as a scalable solution that can overcome access barriers through the widespread availability of smartphones, tablets, and computers. Recent randomized controlled trials have demonstrated that dCBT-I not only reduces insomnia symptoms but also improves functional health, psychological well-being, and sleep-related quality of life over time [23]. Digital platforms can allow CBT to reach a wide population [24-26]. Meta-analytic evidence robustly supports the efficacy and broader benefits of digital CBT for insomnia. A 2022 meta-analysis demonstrated large short-term improvements in insomnia severity and moderate reductions in depressive symptoms that persisted long-term [27], while another meta-analysis focusing on occupational outcomes found significant decreases in lost productivity (presenteeism) and work-related rumination among working adults [28].

Building on this clinical evidence, recent economic evaluations have also matured: a 2023 systematic review and meta-analysis reported that dCBT-I consistently achieves favorable cost-utility and cost-effectiveness, particularly when societal costs such as productivity losses are considered [29]. Complementary findings from randomized controlled trials show that guided internet-delivered CBT-I can yield measurable quality-adjusted life years (QALY) gains and, from a societal perspective, has a high probability of dominating control conditions [30]. In addition, simulation modeling using a Markov framework in the United States found that fully automated digital CBT programs (eg, Sleepio) may not only be cost-effective but potentially cost-beneficial compared with no treatment, clinician-delivered CBT, and pharmacotherapy [31]. However, to date, no cost-effectiveness analysis of dCBT-I has been conducted using Korean clinical trial data and Korean-specific cost inputs. Given the strong influence of country-specific health care system structures and current treatment patterns on economic outcomes in insomnia, reliance on evidence generated in Western settings with limited consideration of societal costs may substantially misrepresent the value of digital CBT-I in the Korean context [29-32]. This study addresses an important evidence gap by incorporating Korean randomized clinical trial data and locally relevant cost inputs to evaluate the cost-effectiveness of a domestically developed digital CBT-I from both health care system and societal perspectives.

Hypothesis, Aims, and Objectives

Somzz (AIMNEXT Inc) is the first domestically developed digital therapeutic in Korea designed to deliver CBT for insomnia (CBT-I) [33]. The fully automated mobile app incorporates core CBT-I components, including stimulus control, sleep restriction, sleep hygiene education, relaxation training, cognitive therapy, and relapse prevention, while providing

XSL•FO
RenderX

personalized, real-time feedback based on users' sleep diary data and behavioral patterns. By targeting the psychological, behavioral, and cognitive factors that maintain chronic insomnia, Somzz offers a scalable alternative to clinician-delivered CBT-I.

Based on its demonstrated clinical efficacy and the substantial societal burden of untreated insomnia, we hypothesized that digital CBT-I delivered through Somzz would represent a cost-effective strategy compared with conventional care involving CBT and pharmacotherapy. In particular, we anticipated that Somzz would improve health outcomes while reducing overall costs when broader societal impacts, such as productivity losses and accident-related costs, are taken into account.

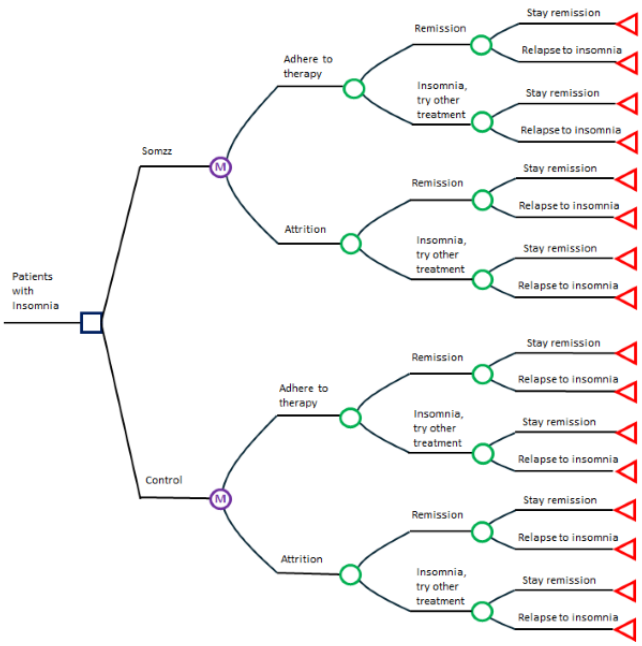
The analysis aimed to evaluate the cost-effectiveness of Somzz compared to standard care from both health care system and societal perspectives by using clinical outcomes from a Korean randomized controlled trial (RCT) and integrating real-world Korean wage, service use, and epidemiological data. We also aimed to examine the robustness of these findings through deterministic and probabilistic sensitivity analyses.

Methods

Conditions and Design

This study is a secondary analysis of a parent single-blind RCT (trial registration: KCT0007292; registered retrospectively on 2022-05-17) to assess the safety and efficacy of a digital therapeutic, Somzz [33]. This study is to compare the costs and outcomes of Somzz versus conventional care in 2023, and thus, we constructed a static decision tree model combined with a Markov model with a 27-week (3 treatment cycles over 9 weeks), which aligns with previous insomnia cost-effectiveness studies [31]. A Markov model represents disease progression by simulating transitions between mutually exclusive health states over discrete time cycles, assuming that future transitions depend only on the current state and not the path taken to reach it [34]. In our model, patients transitioned through health states of adherence, attrition, remission, relapse, and persistence of insomnia (Figure 1).

Figure 1. A basic description of the structure of the decision analytic Markov model.



We adopted the health care system and societal perspectives according to the economic evaluation guidelines in Korea [35]. This dual approach allowed us to assess the economic implications of a health care intervention more thoroughly by capturing the direct financial impact on the health care system and broader economic and societal effects. A discount of 4.5% was applied as per Korean guidelines [35]. We referred to the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) guidelines [36] to conduct the overall cost-effectiveness analysis (Supplementary Table S1 in Multimedia Appendix 1).

Inclusion and Exclusion Criteria

In the aforementioned parent trial, participants were eligible if they were (1) ≥19 years old, (2) had a diagnosis of chronic insomnia, and (3) were able to participate in scheduled visits and assessments. Exclusion criteria in the original trial included: (1) severe psychiatric or neurological conditions, (2) participation in other sleep-related interventions, and (3) medical contraindications to CBT-I or study procedures [33]. This study included only participants from the original trial dataset; no additional inclusion or exclusion criteria were applied.

Participant Characteristics

The parent RCT enrolled 98 adults (Somzz: 49; control: 49). Key characteristics were (1) age: adults 19 years and older (details reported in original trial [33]), (2) employment: 51% engaged in paid work (Somzz: 22/49; control: 28/49), and (3) baseline insomnia severity: clinically significant chronic insomnia. Participants were similar between groups due to randomization [33].

XSL•FO
RenderX

Sampling Procedures

In the aforementioned parental randomized controlled trial, participants were recruited through multiple clinical centers via advertisements and clinician referrals [33]. Sampling followed consecutive enrollment of eligible patients [33]. The sample for this cost-effectiveness study comprises all randomized participants with complete outcome data in the parental trial [33].

Sample Size, Power, and Precision

The sample size was determined in the parent clinical trial based on detecting meaningful differences in insomnia severity (Insomnia Severity Index [ISI] scores) [33]. For the current economic evaluation, all available trial participants were included; therefore, no additional power analysis was required. Precision was improved through probabilistic sensitivity analyses using distributions for all parameters.

Measures and Covariates

Intervention and Comparator

The intervention is the digital CBT, Somzz, for insomnia treatment, and the comparator is CBT or pharmacotherapy without Somzz. As aforementioned, underpinning this study was a multicenter, single-blind, RCT (trial registration: KCT0007292) to assess the safety and efficacy of Somzz [33]. This study evaluates the cost-effectiveness of the Somzz using the collected and published information from the aforementioned parent clinical trial [33].

Outcomes

Outcomes included direct efficacy indicators related to insomnia treatment indicators (probabilities of remission, maintaining remission, attrition) and quality-adjusted-life-years. The direct efficacy values were extracted from the parent clinical trial [33]; (1) remission was defined as an ISI score of 7 or lower, indicating the absence of clinically significant insomnia; (2) the attrition; and (3) the probability of maintaining remission, which were measured during the follow-up at 18 weeks after device withdrawal. The minimum value of the attrition rate was the lower limit of the 95% CI, assuming that the attrition rate followed a normal distribution in the population. The maximum value was determined by applying values obtained from the literature [37], considering the possibility of an underestimated dropout rate in the clinical trial. QALYs were used as a health outcome measure in the analysis. QALYs were derived from the parent clinical trial [33], which measured QALY with the Short Form 6 Dimensions (SF-6D) Health State Short Form (a subset of 6 items from the Short Form Health Survey Version 1) [38]. The SF-6D includes 6 dimensions, each with 2 to 5 levels, and is capable of generating 7500 different health states. Brazier algorithm was used to calculate utility values [39]. The SF-6D was reported to show superior sensitivity to changes in mild-to-moderate physical and mental health conditions compared to the EQ-5D-3L [40,41].

Direct Medical Costs

We defined direct medical costs related to insomnia as expenses incurred from visits to clinics, hospitals, or pharmacies, specifically for insomnia treatment. The base case value for the treatment group was determined as the sum of 2 consultation fees, the cost of CBT software, and medication expenses. The minimum value for the treatment group was calculated as the 2 consultation fees, assuming no costs for CBT software or medication. The upper limit value for the treatment group was calculated by increasing the CBT software cost up to 200,000 KRW (US $155.28) with 2 consultations and related medications. The primary value for the control group included only medication costs, which were adjusted for adherence rates of 14% for prescription medications and 2% for over-the-counter medications. The minimum value of medication costs was calculated as the mean prescription medication cost subtracted from the SD. The maximum value was derived by adding the SD to the mean prescription medication cost. Medical expenses unrelated to insomnia treatment were classified as indirect medical costs.

Indirect Costs

Indirect costs were calculated as the sum of absenteeism, productivity loss, and workplace accident costs. Absenteeism and productivity losses caused by insomnia were monetized as part of absenteeism costs. We converted the number of days absent from work (cost of absenteeism) and the degree of productivity decline (cost of reduced work productivity) due to insomnia into a monetary cost for wage workers participating in the clinical trial [33]. Insomnia increases the likelihood of workplace accidents; therefore, the associated costs were estimated by multiplying the probability of insomnia-related accidents by the average cost per workplace accident. Insomnia can cause workplace accidents; therefore, we calculated the cost of insomnia-related accidents by multiplying the probability of an accident occurring because of insomnia by the average cost of an accident.

Data Collection

All clinical and quality-of-life data were collected prospectively during the parent RCT at 6 scheduled visits (screening, baseline, 1 week, 2 weeks, 4 weeks, and 18 weeks) [33]. Economic variables were derived from Korean national databases, trial-reported work impairment, and published literature. All data were de-identified before secondary analysis.

Quality of Measurements

Clinical outcomes (eg, ISI) were collected in the parent clinical trial by trained personnel following standardized protocols [33]. Economic parameters used validated national datasets. No missing data were reported in the parent clinical trial; no new missingness occurred in the secondary dataset.

Instrumentation

The instrumentation for this study included the Somzz dCBT-I mobile app, which delivers standardized CBT modules for insomnia through a smartphone-based digital platform, and the SF-6D survey instrument, a validated and widely used tool for measuring health-related utility values. All instruments were used in their original, validated formats, and no modifications were made to any components or measures.

XSL•FO
RenderX

Masking

The parent clinical trial used a single-blind design in which outcome assessors were blinded to treatment assignment. Participants were aware of their intervention due to the nature of the digital therapeutic. The economic evaluation was fully blinded to group identifiers [33].

Psychometrics

The SF-6D utilities were derived from SF-36 responses, and the psychometric properties of the Korean SF-36 support its use in Korean populations. In a Korean general population study (N=600), the SF-36 demonstrated acceptable internal consistency and test–retest reliability, with reliability coefficients ranging from 0.54 to 0.80 across domains [38,39,42]. In addition, the Korean Health Insurance Review and Assessment Service economic evaluation guideline allows the use of multi-attribute utility instruments, including SF-6D–derived utilities, in the Korean health technology assessment context [35]. Utility scoring used validated country-neutral algorithms. Attrition and remission measurements followed established ISI-based definitions [33].

Data Diagnostics

We evaluated plausibility ranges for all parameters, checked for outliers in cost distributions, and verified internal model consistency. The attrition rate distribution was modeled using normal assumptions based on trial data. Costs from claims databases were cross-validated with external Korean sources.

Analytic Strategy

Base-Case Analysis

The base-case analysis estimated the incremental cost-effectiveness ratio (ICER) by comparing the costs and health outcomes of digital CBT with those of treatment as usual. Health outcomes were measured in QALYs, and the analysis incorporated both direct medical costs and indirect societal costs to comprehensively assess the economic impact of the intervention.

Sensitivity Analysis

We ensured the robustness of our findings by varying several key assumptions, including the frequency of clinic visits, the need for additional consultations and educational sessions, and incremental health care expenditures related to insomnia. We also varied the financial impact of absenteeism, presenteeism, and workplace accidents resulting from insomnia. For absenteeism and presenteeism, we alternatively used data from the Work Productivity and Activity Impairment, which provides average, maximum, and minimum values for these factors [43]. We established the minimum cost of workplace accidents as zero with a scenario of no accident occurrence and the maximum cost as the highest individual accident cost recorded in the Industrial Accident Insurance Statistics [44] to account for the most severe financial impact observed.

Ethical Considerations

The trial was approved by the institutional review boards (IRBs) of all participating centers and conducted in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines [33]. This study is a secondary analysis using existing data with primary consent, and the original consent or IRB approval covers secondary analysis without additional consent. Furthermore, ethical approval for the cost-effectiveness study was obtained from the Institutional Review Board of Yonsei University (approval number 202111-HR-2636-01). A waiver of informed consent for the use of retrospective data was also granted by the Institutional Review Board of Yonsei University. No financial incentives were provided in the parental clinical trial [33]. All data were deidentified before analysis. This study involved secondary analysis of anonymized clinical trial data already covered under the original IRB approval [33]. No images or supplementary materials in this article contain identifiable information about individual participants or users. All figures and tables present aggregated or schematic data only. Therefore, no additional consent for the use of identifiable images was required. If any materials containing identifiable information were to be included in the future, we would obtain explicit written consent from those individuals and upload the corresponding consent documentation in accordance with journal guidelines.

Results

Participant Flow

Overview

In the parental clinical trial, a total of 98 adults with chronic insomnia were enrolled and randomized in a 1:1 ratio to the Somzz group (n=49) or the control (sleep hygiene education [SHE]) group (n=49). All participants initiated the assigned intervention, and all participants had 6 scheduled visits (screening, baseline, 1 week, 2 weeks, 4 weeks, and 18 weeks as post intervention) over 6-9 months. The attrition rate in the Somzz group was 12% (6/49), whereas the SHE group showed comparable retention (5/49, 10%) over the 4-month follow-up period. Outcome analyses were conducted according to the intention-to-treat principle [33].

Recruitment

Participants were recruited from 3 university-affiliated hospitals through outpatient clinics and study advertisements. Eligible individuals were screened by research staff and invited to participate between baseline and randomization. After informed consent, participants completed baseline assessments and were randomly assigned to the Somzz or SHE condition using a single-blind allocation procedure [33].

Baseline Characteristics

At baseline, the 2 groups were comparable in demographic and clinical characteristics. Participants were adults with chronic insomnia and demonstrated similar insomnia severity, sleep diary metrics, and mental health profiles prior to randomization. No meaningful group differences were observed at baseline, supporting the validity of the randomization process [33].

Primary Outcomes

Outcomes included direct efficacy indicators related to insomnia treatment indicators (probabilities of remission, maintaining remission, and attrition) and quality-adjusted-life-years. The

XSL•FO
RenderX

direct efficacy values were extracted from the parent clinical trial [33]; (1) remission was 51% (22/43) of the digital CBT group and 14% (6/44) of the control group post intervention; (2) the attrition rate was 12% (6/49) for the in the digital CBT group and 10% (5/49) in the control group; (3) the probability of maintaining remission was 35% (15/43) of the treatment group and 11% (5/44) of the control group. For indirect medical costs, we compared medical claims from individuals diagnosed with insomnia to those without an insomnia diagnosis and determined that excess health expenditure caused by insomnia would amount to 553,625 KRW (US $429.83) over 9 weeks (Table 1).

Indirect costs were calculated as the sum of absenteeism, productivity loss, and workplace accident costs. In a parent clinical study, 54% were "individuals employed in paid work," and they reported being absent from work because of insomnia-related issues for 3.11 hours per week on average before starting treatment [33]. Therefore, we calculated the cost of absenteeism due to insomnia per one 9-week treatment cycle as follows: 9 weeks × 3.11h/week × 19,806 KRW (US $15.38) (hourly wage) × 54% (employment rate) ≒ 33,312 KRW (US $25.87). The parent clinical trial participants reported a 55% decline in productivity caused by insomnia [33]. Therefore, the indirect cost of lost productivity for one 9-week treatment cycle was approximated as follows: 45 days (5 days of work per week for the 9-week treatment period) × 55% (productivity loss due to insomnia) × 157,260 KRW (US $122.10)/day (daily wage) × 54% (employment rate) ≒ 2,101,347 KRW (US $1631.49) (Table 1). The probability of an insomnia-related workplace accident was assumed to be 1% [11]. According to the Ministry of Employment and Labor's Industrial Accident Insurance Statistics in Korea, the total amount of industrial accident insurance benefits in 2020 was 6.00 trillion KRW paid across 350,000 beneficiaries, resulting in an average of 17 million KRW per person [44]. Thus, the estimated accident cost per person is 85 million KRW annually and approximately 79,562 KRW (US $61.77) for a 9-week period [45,46] (Table 1).

XSL•FO
RenderX

Table 1. Model input parameters for the cost-effectiveness analysis of digital CBT (Somzz) for chronic insomnia among Korean adults.

Parameters[a]	Value assumed (range)	Distribution	Source
Outcomes			
Remission probability			
Treatment group (Somzz)	0.51 (0.40-0.68)	Beta	[33], ISI[b]<7; (minimum ISI < 6 - maximum ISI < 8)
Control group	0.14 (0.09-0.40)	Beta	[33]
Probability of maintaining remission for 18 weeks			
Treatment group (Somzz)	0.35 (0.18-0.47)	Beta	[33] (minimum ISI<6 - maximum ISI<8)
Control group	0.11 (0.09-0.18)	Beta	[33]
Attrition rate			
Treatment group (Somzz)	0.12 (0.08-0.30)	Beta	[33]; (minimum lower level of 95% CI- maximum [37])
Control group	0.10 (0.06-0.40)	Beta	[33]; (minimum [33] - maximum [31])
Quality-adjusted life years			
Insomnia	0.684 (0.460-1.0)	Beta	[33]
Noninsomnia	0.810 (0.573-1.0)	Beta	[33]
Costs			
Direct medical costs related to insomnia			
Treatment group (Somzz)	KRW 91,194 (29,100-237,194)[c]	Gamma	[47,48]
Control group	KRW 8,094 (7,890-10,068)[d]	Gamma	[47,48]
Direct medical costs unrelated to insomnia			
Treatment group (Somzz)	KRW 553,625 (0-136,074,635)[e]	gamma	[49]
Control group	KRW 437,689 (0-89,272,731)[f]	Gamma	[49]
Indirect costs			
Workplace absenteeism	KRW 33,312 (0-278,497)[g]	Gamma	[33], [45,46]
Workplace presenteeism	KRW 2,101,347 (0-3,827,195)[h]	Gamma	[33], [45,46]
Additional costs from workplace accidents	KRW 79,562 (0-1,590,541)[I]	Gamma	[33], [11,45,46]
Additional costs			
Digital CBT[j]	KRW 91,194 (29,100-237,194)[k]	Gamma	Internal data

[a]Model input parameters for the cost-effectiveness analysis of digital cognitive behavioral therapy (Somzz) for chronic insomnia among Korean adults were derived from clinical trial data and relevant literature. These parameters included remission and attrition probabilities, health-state utilities, direct and indirect costs, and the corresponding probability distributions used for probabilistic modeling.

[b]ISI: Insomnia Severity Index.

[c]US $70.80 (US $22.59-184.16).

[d]US $6.29 (US $6.13-7.82).

[e]US $429.83 (US $0-105648.35).

[f]US $339.82 (US $0-69311.35).

[g]US $25.86 (US $0-216.23).

[h]US $1631.49 (US $0-2971.43).

[i]US $61.77 (US $0-1234.90).

[j]CBT: cognitive behavioral therapy.

[k]US $70.80 (US $22.59-184.16).

XSL•FO
RenderX

Statistics and Data Analysis

Cost-Effectiveness of Digital CBT

Table 2 presents scenarios detailing the simulated health utility (QALYs) and cost outcomes for the baseline model. All scenarios, except Scenario 4, were analyzed by including only direct medical costs. Costs and QALYs increased for patients who implemented the digital CBT solution. The model predicted that the use of digital software to treat insomnia would produce 0.3701 QALYs and 101,729 KRW (US $78.98) of expected costs compared with 0.3609 QALYs and 22,038 KRW generated by treatment as usual without the aid of digital software over 6.5 months. Introducing the digital insomnia solution would increase costs by 79,691 KRW (US $61.87) and QALYs by 0.0091. The ICER associated with insomnia treatment using digital software was 8,719,727 KRW (US $6770.00) per QALY gained, which is well below the hypothetical threshold value in Korea of 30,000,000 KRW (US $23,292)/QALY gained.

Table 2. Cost-effectiveness analysis of digital cognitive behavioral therapy (Somzz) for chronic insomnia among Korean adults relative to the control group by base case and scenario.

Strategy	Cost, KRW (US $)	Effectiveness (QALYs)[a]	Incremental cost, KRW (US $)	Incremental effectiveness (QALYs)	ICER[b], KRW /QALYs (US $/QALYs)
Base case: 2 clinic visits					
Control group	22,038 (17.11)	0.3609	—[c]	—	—
Treatment group (Somzz)	101,729 (78.98)	0.3701	79,691 (61.88)	0.0091	8,719,727 (6,770.87)
Scenario 1: 3 clinic visits					
Control group	22,038 (17.11)	0.3609	—	—	—
Treatment group (Somzz)	113,859 (88.40)	0.3701	91,821 (71.30)	0.0091	10,046,980 (7,801.48)
Scenario 2: 2 clinic visits + educational and consultant fees					
Control group	22,038 (17.11)	0.3609	—	—	—
Treatment group (Somzz)	131,729 (102.27)	0.3701	109,691 (85.18)	0.0091	12,002,298 (9,319.78)
Scenario 3: 3 clinic visits + educational and consultant fees					
Control group	22,038 (17.11)	0.3609	—	—	—
Treatment group (Somzz)	143,859 (111.69)	0.3701	121,821 (94.59)	0.0091	13,329,551 (10,350.40)
Scenario 4: all costs included[d]					
Control group	7,679,657 (596.49)	0.3609	—	—	—
Treatment group (Somzz)	6,777,978 (526.42)	0.3701	–901,678 (–700.15)	0.0091	–98,660,780 (76,610.10)

[a]QALYs: quality-adjusted life years.

[b]ICER: incremental cost-effectiveness ratio.

[c]Not available.

[d]Direct medical costs related to insomnia + (educational and consultation fees) + direct medical costs unrelated to insomnia + indirect costs.

Compared to the control group, the ICER for insomnia treatment using the digital intervention, assuming all possible indirect costs, yielded a lower cost of 901,678 KRW (US $700.06) and higher QALYs of 0.0091. This negative ICER arises from combining the reduced total costs and increased health utility. Digital CBT reduces insomnia-related indirect costs, such as health care expenses and lost workplace productivity, through higher effectiveness and remission rates. These cost reductions surpass the direct costs associated with digital CBT (Table 2).

Sensitivity Analysis

We conducted a one-way deterministic and probabilistic sensitivity analysis to test the robustness of our findings. Figure 2 shows a tornado diagram of the results of this analysis. The direct costs of the digital CBT group, which ranged from 29,100 to 237,194 KRW (US $22.59 to US $184.16), had the largest impact on the ICER, followed by the probability of remission in the control group. These parameters had the greatest impact in all scenarios (Multimedia Appendices 2-4). The Monte Carlo simulation results indicated that digital CBT is the optimal cost-effective strategy with a probability of 67.4%, represented as an ICE scatterplot in Figure 3.

XSL•FO
RenderX

Figure 2. Tornado diagram for incremental cost-effectiveness ratio. ICER: incremental cost-effectiveness ratio.

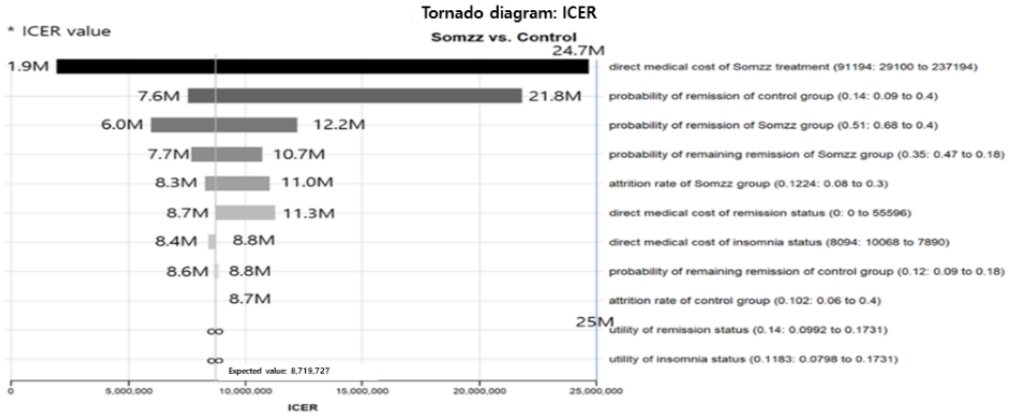
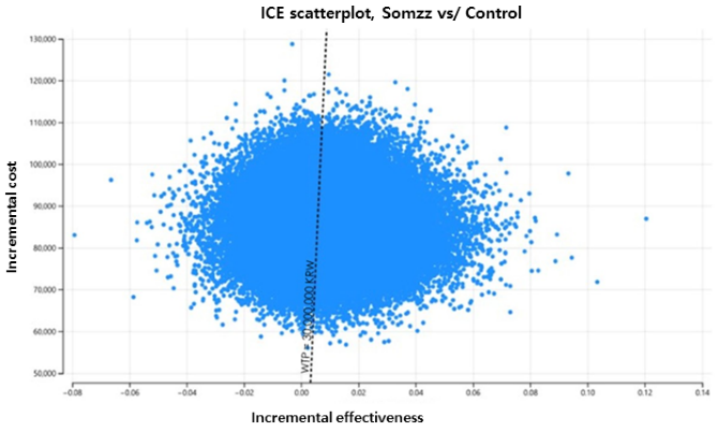


Figure 3. Probabilistic sensitivity analysis results: incremental cost-effectiveness scatterplot (estimates based on the 95% CI).



For Scenarios 1-3, digital CBT was the most cost-effective strategy, with probabilities of 66.9%, 65.1%, and 63.9%, respectively (Multimedia Appendices 5-7). Furthermore, digital CBT was identified as the most cost-effective strategy at a willingness-to-pay threshold of 15,000,000 KRW (US $11,647.5), with probabilities of 58.8%, 57.8%, 53.5%, and 51.3% for the base case and Scenarios 1-3, respectively (Multimedia Appendix 8).

Discussion

Support of Original Hypotheses

We combined a decision tree and the Markov model to assess the cost-effectiveness of digital CBT for adults seeking treatment for insomnia in South Korea. While several prior studies have evaluated the cost-effectiveness of digital CBT for insomnia, most were conducted in Western health care systems [29-32]. In contrast, this study evaluates a regulated, commercially available digital therapeutic using Korean randomized trial data and explicitly incorporates societal costs that are highly relevant to the Korean labor and health care context. From a health care perspective, digital CBT slightly increased health care costs while improving health outcomes compared with standard care. In contrast, from a societal perspective, digital CBT was cost-beneficial, yielding a negative ICER. Accordingly, the digital therapeutic evaluated in this study was highly cost-effective for functional outcomes in insomnia treatment at a willingness-to-pay threshold of 30,000,000 KRW (US $23,295) per QALY gained and remained cost-effective at a lower threshold of 15,000,000 KRW (US $11,647.5) per QALY gained. These findings underscore that incorporating broader societal costs meaningfully changes the economic value of insomnia treatment interventions. Reduced health care use, lower risks of workplace accidents, and improved work productivity associated with insomnia remission were key drivers of cost-effectiveness. Both deterministic and probabilistic sensitivity analyses confirmed the robustness of these results.

Similarity of Results and Interpretation

These results align with those of previous studies that found guided digital CBT to be cost-effective [30-32]. One strength of our study was the inclusion of indirect costs. We considered reductions in health care expenditure as effective insomnia treatment can lead to fewer medical visits, less need for medication, and lower overall health care use [21,22]. Furthermore, we accounted for a decreased risk of workplace accidents. Insomnia impairs cognitive and physical functioning, which increases the likelihood of errors and accidents at work [50,51]. Treating insomnia can mitigate this risk, thereby leading to safer work environments. Our analysis also included improved workplace productivity. Insomnia negatively affects concentration, efficiency, and overall job performance [52,53]. Therefore, effective treatment can enhance productivity by improving employees' sleep quality and daytime functioning. By capturing these indirect costs, our analysis demonstrates the broader economic benefits of insomnia treatment beyond

XSL•FO
RenderX

immediate improvements in individual health. Insomnia is linked to 7.2% of all costly workplace accidents and errors, and accounts for 23.7% of the total costs associated with these incidents [11]. Not including these additional costs could underestimate the total economic burden of insomnia and potential costs saved through the application of effective treatment. In contrast, including costs related to workplace accidents and errors could reveal higher overall costs, and thus, a greater potential for cost savings through insomnia treatment.

The demand for effective insomnia treatment is substantial, with 20%-30% of the population experiencing insomnia symptoms [1-3]. However, several factors hinder traditional clinician-delivered CBT, including the availability of trained therapists and the need for patients to attend scheduled sessions, which many people can find challenging. Patients often face difficulties in taking time off work and securing childcare or older adult care to attend in-person therapy sessions. These logistical challenges limit the reach and effectiveness of clinician-delivered CBT [54].

Digital CBT delivery offers a promising solution to these accessibility issues by enabling instant and unlimited access to CBT resources, thereby allowing patients to engage with therapy at any time and from any location, effectively eliminating waiting times. This flexibility is crucial for patients who may be unable to attend regular appointments because of their busy schedules [54]. Digital platforms can make CBT more widely available, ensuring that more individuals who require treatment can access it promptly and conveniently. This approach not only expands the reach of effective insomnia treatment but also aligns with the high demand for such interventions, ultimately helping bridge the gap between the recommended use of CBT and its actual accessibility and use.

Generalizability

The generalizability of this study's findings should be interpreted in the context of the Korean population and health care system. Participants were digitally literate adults recruited from university hospitals [33], which may not fully represent individuals with limited technology access or older adults who may engage differently with mobile health interventions. Nonetheless, given Korea's high smartphone penetration rate and increasing familiarity with digital health tools, Somzz has strong potential for scalable implementation across the broader adult population.

Economic estimates in this model reflect Korean wage levels, health care use patterns, and societal cost structures. Although absolute cost differences may vary in other countries, the primary mechanisms driving cost-effectiveness, such as higher remission rates, reduced health care use, and improved productivity, are likely applicable to other high-income settings with similar burdens of insomnia and labor market characteristics. Because Somzz was developed within the Korean cultural and linguistic environment, differences in cultural attitudes toward sleep, mental health, and digital therapeutics may influence usability and adherence in other populations. Additionally, real-world engagement with digital CBT-I may vary outside a controlled trial setting, which could influence long-term effectiveness and economic outcomes.

Overall, while these findings are most directly applicable to technologically advanced health care systems with strong digital readiness, they provide valuable insights for other countries considering digital CBT-I adoption. Extrapolation should be undertaken cautiously with attention to cultural, economic, and infrastructural differences.

Implications

This study provides novel evidence by incorporating indirect societal costs, including productivity loss and workplace accidents, into the economic evaluation of digital CBT, dimensions that are rarely captured in prior analyses. Importantly, by focusing on a digital therapeutic that meets regulatory standards and is designed for real-world implementation, this study extends prior digital CBT economic evaluations toward policy-relevant evidence for reimbursable digital therapeutics. By using Korean clinical trial data [33] and locally relevant economic parameters rather than relying on Western-based assumptions, the study offers population-specific insights that more accurately reflect real-world conditions. This comprehensive approach strengthens the relevance of our findings for clinicians, policymakers, and employers by demonstrating the broader economic and workplace impact of digital CBT and informing decision-making for large-scale adoption within the Korean health care and occupational health systems.

One limitation of this study was the relatively short follow-up period. Although a period of 6.5 months is approximately the same or longer than that used in all previous studies on the cost-effectiveness of CBT for insomnia, it may not be sufficient to fully capture the long-term economic benefits of treatment. Simulating costs over a longer period, such as 12 or 24 months, could more effectively demonstrate potential long-term cost reductions for employers. A longer time horizon may allow for a more significant accumulation of treatment benefits, suggesting that our current results may be conservative.

Another limitation is the sample size. Economic evaluations conducted in parallel with clinical trials often face power issues because cost variables typically have a higher variance and require larger sample sizes than clinical outcomes to achieve statistical significance. Considering these challenges, we adopted a probabilistic decision-making approach to account for the uncertainty and variability in cost and utility estimates. This approach helped us perform a more robust analysis despite the limitations of the sample size and follow-up duration in this study.

Finally, the parent clinical trial from which outcome data were extracted [33] was not structured to compare traditional and digital CBT directly, primarily because of practical constraints (eg, despite having insomnia symptoms, only a small proportion of patients actively seek pharmacotherapy in the real world). This approach mimics real-world scenarios in which patients often receive a combination of treatments rather than a single isolated therapy.

This study did not involve direct patient or stakeholder engagement. However, the societal burden and impact of insomnia were incorporated into the analytical framework. No

XSL•FO
RenderX

direct involvement from patients, service recipients, the general public, or other stakeholders influenced the approach or findings of this study.

Conclusions

The results of this study suggest the significant potential of early integration of digital CBT into insomnia treatment to improve health care outcomes and generate substantial cost-effectiveness. These findings underscore the importance of incorporating such technology as a valuable enhancement to routine medical practice, thereby ultimately enhancing overall insomnia care. Beyond demonstrating cost-effectiveness, this study contributes to the growing evidence base supporting digital therapeutics by illustrating how locally grounded economic evaluations can inform reimbursement, employer adoption, and national digital health policy. Given these promising results, future research should focus on expanding access to digital CBT and assessing its effectiveness on a population-wide scale. Understanding the effectiveness of digital CBT in real-world settings, in which factors such as varying levels of access, patient engagement, and implementation challenges can influence outcomes, is crucial.

Acknowledgments

We did not use generative artificial intelligence to write any portion of the manuscript.

Funding

This work was supported by the Ministry of Health and Welfare (RS-2023-00266398) and the Hankuk University of Foreign Studies Research Fund. The funder had no involvement in the study design, data analysis, interpretation of the results, writing of this article, or the decision to submit it for publication.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request. Most of the input data used to derive the results of this study are extracted from published literature, which is all listed in the Reference section with corresponding citations in the text.

Conflicts of Interest

Jimin Woo is affiliated with AIMNEXT Co., Ltd.

Multimedia Appendix 1
Supplementary Table 1.
[PDF File (Adobe PDF File), 65 KB - mhealth_v14i1e71750_app1.pdf]

Multimedia Appendix 2
Deterministic sensitivity analysis results: Scenario 1 – tornado diagram for ICER.
[PNG File , 315 KB - mhealth_v14i1e71750_app2.png]

Multimedia Appendix 3
Deterministic sensitivity analysis results: Scenario 2 – tornado diagram for ICER.
[PNG File , 303 KB - mhealth_v14i1e71750_app3.png]

Multimedia Appendix 4
Probabilistic sensitivity analysis results: incremental cost-effectiveness scatterplot for Scenario 3.
[PNG File , 317 KB - mhealth_v14i1e71750_app4.png]

Multimedia Appendix 5
Probabilistic sensitivity analysis results: incremental cost-effectiveness scatterplot for Scenario 1.
[PNG File , 369 KB - mhealth_v14i1e71750_app5.png]

Multimedia Appendix 6
Probabilistic sensitivity analysis results: incremental cost-effectiveness scatterplot for Scenario 2.
[PNG File , 338 KB - mhealth_v14i1e71750_app6.png]

Multimedia Appendix 7
Probabilistic sensitivity analysis results: incremental cost-effectiveness scatterplot for Scenario 3.

XSL•FO
RenderX

[PNG File , 361 KB - mhealth_v14i1e71750_app7.png]

Multimedia Appendix 8
Probabilistic sensitivity analysis results: incremental cost-effectiveness scatterplot: (A) base case; (B) Scenario 1; (C) Scenario 2; and (D) Scenario 3 (WTP = 15,000,000).
[PNG File , 528 KB - mhealth_v14i1e71750_app8.png]

References

1. Sateia MJ. International classification of sleep disorders-third edition: highlights and modifications. Chest 2014;146(5):1387-1394. [doi: 10.1378/chest.14-0970] [Medline: 25367475]
2. Qaseem A, Kansagara D, Forciea MA, Cooke M, Denberg TD, Clinical Guidelines Committee of the American College of Physicians. Management of chronic insomnia disorder in adults: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2016;165(2):125-133 [FREE Full text] [doi: 10.7326/M15-2175] [Medline: 27136449]
3. Perlis ML, Posner D, Riemann D, Bastien CH, Teel J, Thase M. Insomnia. Lancet 2022;400(10357):1047-1060. [doi: 10.1016/S0140-6736(22)00879-0] [Medline: 36115372]
4. Morin CM, Bjorvatn B, Chung F, Holzinger B, Partinen M, Penzel T, et al. Insomnia, anxiety, and depression during the COVID-19 pandemic: an international collaborative study. Sleep Med 2021;87:38-45 [FREE Full text] [doi: 10.1016/j.sleep.2021.07.035] [Medline: 34508986]
5. Zhang C, Yang L, Liu S, Ma S, Wang Y, Cai Z, et al. Survey of insomnia and related social psychological factors among medical staff involved in the 2019 novel coronavirus disease outbreak. Front Psychiatry 2020;11:306 [FREE Full text] [doi: 10.3389/fpsyt.2020.00306] [Medline: 32346373]
6. Wade AG. The societal costs of insomnia. Neuropsychiatr Dis Treat 2010;7:1-18 [FREE Full text] [doi: 10.2147/NDT.S15123] [Medline: 21326650]
7. Staner L. Comorbidity of insomnia and depression. Sleep Med Rev 2010 Mar;14(1):35-46. [doi: 10.1016/j.smrv.2009.09.003] [Medline: 19939713]
8. Ge L, Guyatt G, Tian J, Pan B, Chang Y, Chen Y, et al. Insomnia and risk of mortality from all-cause, cardiovascular disease, and cancer: systematic review and meta-analysis of prospective cohort studies. Sleep Med Rev 2019;48:101215. [doi: 10.1016/j.smrv.2019.101215] [Medline: 31630016]
9. Bolge SC, Doan JF, Kannan H, Baran RW. Association of insomnia with quality of life, work productivity, and activity impairment. Qual Life Res 2009;18(4):415-422. [doi: 10.1007/s11136-009-9462-6] [Medline: 19288223]
10. Garbarino S, Magnavita N, Guglielmi O, Maestri M, Dini G, Bersi FM, et al. Insomnia is associated with road accidents. Further evidence from a study on truck drivers. PLoS One 2017;12(10):e0187256 [FREE Full text] [doi: 10.1371/journal.pone.0187256] [Medline: 29088276]
11. Shahly V, Berglund PA, Coulouvrat C, Fitzgerald T, Hajak G, Roth T, et al. The associations of insomnia with costly workplace accidents and errors: results from the America insomnia survey. Arch Gen Psychiatry 2012;69(10):1054-1063. [doi: 10.1001/archgenpsychiatry.2011.2188] [Medline: 23026955]
12. Riemann D, Baglioni C, Bassetti C, Bjorvatn B, Dolenc Groselj L, Ellis JG, et al. European guideline for the diagnosis and treatment of insomnia. J Sleep Res 2017;26(6):675-700. [doi: 10.1111/jsr.12594] [Medline: 28875581]
13. Wilson S, Anderson K, Baldwin D, Dijk D, Espie A, Espie C, et al. British Association for Psychopharmacology consensus statement on evidence-based treatment of insomnia, parasomnias and circadian rhythm disorders: an update. J Psychopharmacol 2019;33(8):923-947. [doi: 10.1177/0269881119855343] [Medline: 31271339]
14. Sateia MJ, Buysse DJ, Krystal AD, Neubauer DN, Heald JL. Clinical practice guideline for the pharmacologic treatment of chronic insomnia in adults: an American Academy of Sleep Medicine clinical practice guideline. J Clin Sleep Med 2017;13(2):307-349 [FREE Full text] [doi: 10.5664/jcsm.6470] [Medline: 27998379]
15. Glass J, Lanctôt KL, Herrmann N, Sproule BA, Busto UE. Sedative hypnotics in older people with insomnia: meta-analysis of risks and benefits. BMJ 2005 Nov;331(7526):1169 [FREE Full text] [doi: 10.1136/bmj.38623.768588.47] [Medline: 16284208]
16. Tannenbaum C, Diaby V, Singh D, Perreault S, Luc M, Vasiliadis H. Sedative-hypnotic medicines and falls in community-dwelling older adults: a cost-effectiveness (decision-tree) analysis from a US Medicare perspective. Drugs Aging 2015;32(4):305-314. [doi: 10.1007/s40266-015-0251-3] [Medline: 25825121]
17. Choi Y, Lee MH, Choi J, Kim S, Kim J, Lee YJ. Survey of insomnia treatment status for doctors. Sleep Med Psychophysiol 2016;23(2):77-83. [doi: 10.14401/kasmed.2016.23.2.77]
18. Kaufmann CN, Spira AP, Alexander GC, Rutkow L, Mojtabai R. Trends in prescribing of sedative-hypnotic medications in the USA: 1993-2010. Pharmacoepidemiol Drug Saf 2016;25(6):637-645 [FREE Full text] [doi: 10.1002/pds.3951] [Medline: 26711081]
19. Miller CB, Valenti L, Harrison CM, Bartlett DJ, Glozier N, Cross NE, et al. Time trends in the family physician management of insomnia: the Australian experience (2000-2015). J Clin Sleep Med 2017;13(6):785-790 [FREE Full text] [doi: 10.5664/jcsm.6616] [Medline: 28454597]

XSL•FO
RenderX

20. Ritterband LM, Thorndike FP, Ingersoll KS, Lord HR, Gonder-Frederick L, Frederick C, et al. Effect of a web-based cognitive behavior therapy for insomnia intervention with 1-year follow-up: a randomized clinical trial. JAMA Psychiatry 2017;74(1):68-75. [doi: 10.1001/jamapsychiatry.2016.3249] [Medline: 27902836]
21. Daley M, Morin C, LeBlanc M, Grégoire J, Savard J. The economic burden of insomnia: direct and indirect costs for individuals with insomnia syndrome, insomnia symptoms, and good sleepers. Sleep 2009;32(1):55-64. [doi: 10.5665/sleep/32.1.55]
22. Kessler R, Berglund P, Coulouvrat C, Hajak G, Roth T, Shahly V, et al. Insomnia and the performance of US workers: results from the America insomnia survey. Sleep 2011;34(9):1161-1171 [FREE Full text] [doi: 10.5665/SLEEP.1230] [Medline: 21886353]
23. Sweetman A, Reynolds C, Richardson C. Digital cognitive behavioural therapy for insomnia versus digital sleep education control in an Australian community-based sample: a randomised controlled trial. Intern Med J 2024;54(11):1838-1848. [doi: 10.1111/imj.16521] [Medline: 39257295]
24. Seyffert M, Lagisetty P, Landgraf J, Chopra V, Pfeiffer PN, Conte ML, et al. Internet-delivered cognitive behavioral therapy to treat insomnia: a systematic review and meta-analysis. PLoS One 2016;11(2):e0149139 [FREE Full text] [doi: 10.1371/journal.pone.0149139] [Medline: 26867139]
25. Ye Y, Chen N, Chen J, Liu J, Lin L, Liu Y, et al. Internet-based cognitive-behavioural therapy for insomnia (ICBT-i): a meta-analysis of randomised controlled trials. BMJ Open 2016;6(11):e010707 [FREE Full text] [doi: 10.1136/bmjopen-2015-010707] [Medline: 27903557]
26. Zachariae R, Lyby MS, Ritterband LM, O'Toole MS. Efficacy of internet-delivered cognitive-behavioral therapy for insomnia - a systematic review and meta-analysis of randomized controlled trials. Sleep Med Rev 2016;30:1-10. [doi: 10.1016/j.smrv.2015.10.004] [Medline: 26615572]
27. Lin W, Li N, Yang L, Zhang Y. The efficacy of digital cognitive behavioral therapy for insomnia and depression: a systematic review and meta-analysis of randomized controlled trials. PeerJ 2023;11:e16137 [FREE Full text] [doi: 10.7717/peerj.16137] [Medline: 37927792]
28. Zettor D, Endomba FT, Pierandrei A, Pinoit J, Chauvet-Gelinier J, Forestier N, et al. Effectiveness of digital cognitive behavioral therapy for insomnia on professional activity: a systematic review and meta-analysis of randomized controlled trials. Sleep Med Rev 2025;79:102024. [doi: 10.1016/j.smrv.2024.102024] [Medline: 39571405]
29. Fang L, Lyu Z, Ai S, Du S, Zhou W, Zeng S, et al. Is cognitive behavioral therapy for insomnia more cost-effective? New-perspective on economic evaluations: a systematic review and meta-analysis. Sleep 2024;47(8). [doi: 10.1093/sleep/zsae122] [Medline: 38795362]
30. Buntrock C, Lehr D, Smit F, Horvath H, Berking M, Spiegelhalder K, et al. Guided internet-based cognitive behavioral therapy for insomnia: health-economic evaluation from the societal and public health care perspective alongside a randomized controlled trial. J Med Internet Res 2021;23(5):e25609 [FREE Full text] [doi: 10.2196/25609] [Medline: 34028361]
31. Darden M, Espie CA, Carl JR, Henry AL, Kanady JC, Krystal AD, et al. Cost-effectiveness of digital cognitive behavioral therapy (Sleepio) for insomnia: a Markov simulation model in the United States. Sleep 2021;44(4). [doi: 10.1093/sleep/zsaa223] [Medline: 33151330]
32. Thiart H, Ebert D, Lehr D, Nobis S, Buntrock C, Berking M, et al. Internet-based cognitive behavioral therapy for insomnia: a health economic evaluation. Sleep 2016;39(10):1769-1778 [FREE Full text] [doi: 10.5665/sleep.6152] [Medline: 27450686]
33. Shin J, Kim S, Lee J, Gu H, Ahn J, Park C, et al. Efficacy of mobile app-based cognitive behavioral therapy for insomnia: multicenter, single-blind randomized clinical trial. J Med Internet Res 2024;26:e50555 [FREE Full text] [doi: 10.2196/50555] [Medline: 39058549]
34. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. Med Decis Making 1993;13(4):322-338. [doi: 10.1177/0272989X9301300409] [Medline: 8246705]
35. Guidelines for pharmaceutical economic evaluation [Web page in Korean]. HIRA OAK Repository. 2021. URL: https://repository.hira.or.kr/handle/2019.oak/2541 [accessed 2022-10-10]
36. Husereau D, Drummond M, Augustovski F, de Bekker-Grob E, Briggs AH, Carswell C, et al. Consolidated health economic evaluation reporting standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. Value Health 2022;25(1):3-9 [FREE Full text] [doi: 10.1016/j.jval.2021.11.1351] [Medline: 35031096]
37. Karlin BE, Trockel M, Taylor CB, Gimeno J, Manber R. National dissemination of cognitive behavioral therapy for insomnia in veterans: therapist- and patient-level outcomes. J Consult Clin Psychol 2013;81(5):912-917. [doi: 10.1037/a0032554] [Medline: 23586730]
38. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. Med Care 2004;42(9):851-859. [doi: 10.1097/01.mlr.0000135827.18610.0d] [Medline: 15319610]
39. Mukuria C, Rowen D, Mulhern B, McDool E, Kharroubi S, Bjorner JB, et al. The short form 6 dimensions (SF-6D): development and evolution. Appl Health Econ Health Policy 2025;23(1):19-33 [FREE Full text] [doi: 10.1007/s40258-024-00919-8] [Medline: 39460886]
40. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ 2004;13(9):873-884 [FREE Full text] [doi: 10.1002/hec.866] [Medline: 15362179]

XSL•FO
RenderX

41. Brazier J, Connell J, Papaioannou D, Mukuria C, Mulhern B, Peasgood T, et al. A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. Health Technol Assess 2014;18(34):vii-viii, xiii [FREE Full text] [doi: 10.3310/hta18340] [Medline: 24857402]
42. Kim SH, Jo M, Lee S. Psychometric properties of the Korean short form-36 health survey version 2 for assessing the general population. Asian Nurs Res (Korean Soc Nurs Sci) 2013;7(2):61-66 [FREE Full text] [doi: 10.1016/j.anr.2013.03.001] [Medline: 25029923]
43. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. PharmacoEconomics 1993;4(5):353-365. [doi: 10.2165/00019053-199304050-00006]
44. Industrial Accident Insurance Annual Report. Industrial Accident Insurance Annual Report. 2023. URL: https://www.moel.go.kr/info/publict/publictDataView.do?bbs_seq=20241001451 [accessed 2022-10-10]
45. KOSIS. Insurance Benefit Payment Status (By Year and Business Size). Insurance Benefit Payment Status (By Year and Business Size). URL: https://tinyurl.com/mphycv28 [accessed 2022-10-10]
46. Park C, Lee H. A study on the economic loss costs related to industrial accidents - focusing on the manufacturing industry [Web page in Korean]]. Korea Labor Institute Research Report. 2018. URL: https://www.kli.re.kr/board.es?mid=a10505020000&bid=0007&act=view&list_no=138920&tag=&nPage=12 [accessed 2022-10-10]
47. 2022 Health institution fee schedule and patient co-payment table. HIRA.: Health Insurance Review & Assessment Service URL: https://www.hira.or.kr/bbsDummy.do?pgmid=HIRAA020002000100&brdScnBltNo=4&brdBltNo=9293&pageIndex=1#none [accessed 2022-10-10]
48. Current list of reimbursed drugs and the reimbursement ceiling amount table. Health Insurance Review & Assessment Service. 2022. URL: https://www.hira.or.kr/bbsDummy.do?pgmid=HIRAA030014050000 [accessed 2022-10-10]
49. Sarsour K, Kalsekar A, Swindle R, Foley K, Walsh JK. The association between insomnia severity and healthcare and productivity costs in a health plan sample. Sleep 2011;34(4):443-450 [FREE Full text] [doi: 10.1093/sleep/34.4.443] [Medline: 21461322]
50. Uehli K, Mehta AJ, Miedinger D, Hug K, Schindler C, Holsboer-Trachsler E, et al. Sleep problems and work injuries: a systematic review and meta-analysis. Sleep Med Rev 2014;18(1):61-73 [FREE Full text] [doi: 10.1016/j.smrv.2013.01.004] [Medline: 23702220]
51. Wickwire EM, Shaya FT, Scharf SM. Health economics of insomnia treatments: the return on investment for a good night's sleep. Sleep Med Rev 2016;30:72-82. [doi: 10.1016/j.smrv.2015.11.004] [Medline: 26874067]
52. Hafner M, Stepanek M, Taylor J, Troxel W, van Stolk C. Why sleep matters-the economic costs of insufficient sleep: a cross-country comparative analysis. Rand Health Q 2017;6(4):11 [FREE Full text] [Medline: 28983434]
53. Rosekind M, Gregory K, Mallis M, Brandt S, Seal B, Lerner D. The cost of poor sleep: workplace productivity loss and associated costs. J Occup Environ Med 2010;52(1):91-98. [doi: 10.1097/JOM.0b013e3181c78c30] [Medline: 20042880]
54. Koffel E, Bramoweth AD, Ulmer CS. Increasing access to and utilization of cognitive behavioral therapy for insomnia (CBT-I): a narrative review. J Gen Intern Med 2018;33(6):955-962 [FREE Full text] [doi: 10.1007/s11606-018-4390-1] [Medline: 29619651]

Abbreviations

CBT: cognitive behavioral therapy
CHEERS: Consolidated Health Economic Evaluation Reporting Standards
dCBT-I: digital cognitive behavioral therapy for insomnia
ICER: incremental cost-effectiveness ratio
IRB: institutional review board
ISI: Insomnia Severity Index
QALY: quality-adjusted life years
RCT: randomized controlled trial
SF-6D: Short Form 6 Dimensions

XSL•FO
RenderX

Edited by S Brini; submitted 25.Jan.2025; peer-reviewed by A Baharav, D Xiao; comments to author 10.Jun.2025; revised version received 21.Dec.2025; accepted 25.Dec.2025; published 19.Jan.2026.

Please cite as:
Kim JH, Sohn M, Woo J, Shin J, Han E
Digital Cognitive Behavioral Therapy for Chronic Insomnia in South Korea: Cost-Effectiveness Analysis Using Decision Tree and Markov Modeling Based on a Secondary Analysis of a Randomized Clinical Trial
JMIR Mhealth Uhealth 2026;14:e71750
URL: https://mhealth.jmir.org/2026/1/e71750
doi:10.2196/71750
PMID:

©Jung Hyun Kim, Minjeong Sohn, Jimin Woo, Jaeyong Shin, Euna Han. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 19.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Efficacy and Safety of Mobile App–Based Metamemory Cognitive Training for Mild Cognitive Impairment: Multicenter Randomized Clinical Trial

Sunyoung Kang[1], MD; Jung-In Lim[2], MA; Lukas Stenzel[3], PhD; Keun You Kim[1], MD, PhD; Eosu Kim[1,4,5], MD, PhD; Hong Jun Jeon[6], MD; Doo-Heum Park[6], MD, PhD; Hyun Kook Lim[7,8], MD, PhD; YongSoo Shim[9], MD, PhD; Jae-Won Jang[10], MD, PhD; Yeshin Kim[10], MD, PhD; Sanghoon Lee[11], BS; Kee Hyung Park[12], MD, PhD

[1]Department of Psychiatry & Institute of Behavioral Science in Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

[2]Interdisciplinary Program in Cognitive Science, Seoul National University, Seoul, Republic of Korea

[3]Cogthera GmbH, Munich, Germany

[4]Graduate School of Medical Science, Brain Korea 21 Project, Yonsei University College of Medicine, Seoul, Republic of Korea

[5]Metabolism-Dementia Research Institute, Yonsei University College of Medicine, Seoul, Republic of Korea

[6]Department of Psychiatry, Konkuk University Medical Center, Seoul, Republic of Korea

[7]Department of Psychiatry, Yeouido St. Mary's Hospital, College of Medicine, Catholic University of Korea, Seoul, Republic of Korea

[8]Catholic Medical Center Institute for Basic Medical Science, The Catholic University of Korea Catholic Medical Center, Seoul, Republic of Korea

[9]Department of Neurology, St. Vincent's Hospital, College of Medicine, Catholic University of Korea, Seoul, Republic of Korea

[10]Department of Neurology, Kangwon National University Hospital, College of Medicine, Kangwon National University, Chuncheon, Gangwon-do, Republic of Korea

[11]Emocog Inc., Seoul, Republic of Korea

[12]Department of Neurology, College of Medicine, Gachon University Gil Medical Center, Incheon, Republic of Korea

Corresponding Author:
Kee Hyung Park, MD, PhD
Department of Neurology, College of Medicine
Gachon University Gil Medical Center
21 Namdong-daero 774beon-gil, Namdong-gu
Incheon, 21565
Republic of Korea
Phone: 82 32 460 8683
Email: khpark@gachon.ac.kr

Abstract

Background: Metamemory training (MMT) offers a potential nonpharmacological approach to enhance cognitive function in individuals with mild cognitive impairment (MCI). While digital cognitive training improves accessibility, the effectiveness of mobile app–based MMT has not been evaluated in a randomized clinical trial.

Objective: We aimed to evaluate the efficacy and safety of a mobile app–based MMT program, ET-101 (Cogthera), compared to a sham device control group in individuals with MCI.

Methods: This multicenter, randomized controlled trial enrolled participants with MCI, recruited from 7 medical centers, and randomly assigned them to the ET-101 or control group (1:1 ratio). The intervention lasted 12 weeks, with a 12-week follow-up. The ET-101 group received metamemory-based multimemory strategy training and real-time feedback. Assessments of cognition, the daily activities of living, and the quality of life were conducted at baseline, week 12, and week 24. The primary outcome was the proportion of participants who showed cognitive improvement as assessed by the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog)-14 at weeks 12 and 24. Secondary outcomes included changes in the scores of scales assessing cognition, daily activities, and quality of life. Safety analysis assessed adverse events and their relation to digital therapeutics.

Results: In the full analysis set, 49 participants were included in the ET-101 group and 50 in the control group. At week 24, the proportion of responders who maintained or improved their ADAS-Cog-14 scores was significantly higher in the ET-101 group than in the control group (P=.002). Additionally, the ET-101 group showed a significant improvement in ADAS-Cog-14

XSL•FO
RenderX

scores at week 24 compared to baseline levels (estimates=–2.53; t_{265}=–3.05; Bonferroni-adjusted *P*=.003). A subdomain analysis revealed significant improvements in the memory (estimates=–2.50; t_{264}=–4.03; Bonferroni-adjusted *P*<.001) and language (estimates=–0.807; t_{290}=–3.68; Bonferroni-adjusted *P*<.001) domains at week 24 in the ET-101 group compared to the control group. In the safety analysis, 6 adverse events occurred in the ET-101 group and 4 in the control group, but none were related to the interventions. The attrition rate in the ET-101 group was 22.4% (11/49).

Conclusions: ET-101 significantly improved cognitive function compared to the sham device, with effects observed not only in the memory domain but also in the language domain, indicating a transfer effect. Therefore, ET-101 has the potential to provide effective MMT to a broader population with MCI by overcoming location and personnel limitations through a mobile app–based platform.

Trial Registration: ClinicalTrials.gov NCT05938426; https://clinicaltrials.gov/study/NCT05938426

(JMIR Mhealth Uhealth 2026;14:e73464) doi:10.2196/73464

KEYWORDS

cognitive decline; cognitive training; digital therapeutics; mild cognitive impairment; mobile app–based cognitive training

Introduction

Background

With the global population aging rapidly, the number of people affected by dementia is estimated to reach 153 million by 2050 [1]. Increasing evidence shows that timely diagnosis and treatment can slow the rate of cognitive deterioration, significantly reducing the burden on caregivers and social care systems [1]. As a result, effective interventions for mild cognitive impairment (MCI), which is at high risk of progressing, have become a priority in geriatric care [2].

Notably, anti–amyloid monoclonal antibody treatments, which have recently attracted attention in the management of MCI, may be appropriate only for patients who meet several eligibility criteria. Moreover, this treatment is associated with adverse effects that could limit continued use in some cases [3]. In contrast, cognitive training represents a relatively safe intervention with greater generalizability and applicability in the treatment of MCI, and it can also be used in combination with pharmacotherapy or as a stand-alone treatment option for patients who are unable to undergo or continue pharmacological therapy. According to a meta-analysis, cognitive training significantly improves memory function [4]. Moreover, magnetic resonance imaging–based studies have demonstrated that cognitive training improves cerebral blood flow and brain network connectivity, confirming the biological effects of cognitive interventions [5]. However, while cognitive training, as used in previous studies, improved older adults' performance on practiced tasks, the studies did not consider the participants' ability to self-perceive and regulate cognitive strategies during cognitively demanding tasks [6]. In addition, previous studies indicate that one of the key challenges in cognitive training for MCI is the lack of sustained effects once the training period ends [7,8]. These studies address the need for future interventions that aim to address and overcome this limitation and thereby enhance the real-life applicability of the learned strategies [7,8].

To address the limitations of previous methods of cognitive training, Flavell's concept of "metamemory" can be beneficial [9]. Metamemory refers to one's awareness of one's own memory, including the contents and processes of their memory system [6,9]. Metamemory training (MMT) takes into account the concepts of meta-knowledge, meta-monitoring, and meta-judgement [10]. Meta-knowledge addresses how aging affects memory abilities, enabling older adults to learn strategies to deal with age-related cognitive decline [10]. In the meta-monitoring and meta-judgment components, older adults assess their own performance, enabling them to regulate their memorization ability by independently adjusting effective memory strategies according to the demands of the task at hand [10]. According to previous studies, MMT not only shows positive effects on everyday memory performance in older populations but also improves global cognitive functions while significantly reducing the frequency of memory-related discomfort in daily life [11-14].

Conventionally, the main limitation of cognitive interventions for MCI is their reliance on face-to-face interactions with professionals, which hampers the accessibility and continuity of training [15]. However, digitalization is ideal for facilitating the migration of care from clinics to patients' homes, thereby overcoming constraints related to time and space. According to previous studies that evaluated the feasibility of digital cognitive therapy for MCI, patients adhered more to digital therapy compared to paper-and-pencil training. Additionally, in terms of effectiveness, a review of studies on computerized cognitive training for MCI found that its effect on cognitive function was not inferior to that of conventional cognitive interventions [15].

Objectives

Thus, metamemory-based digital cognitive training can be expected to enhance the benefits of cognitive training in individuals with MCI by addressing the limitations of previous cognitive interventions. Based on this concept, ET-101 (Cogthera)—a mobile app–based MMT program—was developed. In this randomized clinical trial, we aimed to evaluate the efficacy and safety of ET-101 in older individuals with MCI compared to a control group using a sham device. To evaluate both the immediate and sustained effects of ET-101, we assessed its efficacy compared to sham controls at 2 time points that is, at 12 weeks, immediately after completion of the intervention, and 24 weeks, corresponding to a 12-week follow-up after the end of the intervention. This study aimed to evaluate the

XSL•FO
RenderX

following hypotheses: first, for the primary outcome, we compared responder proportions (defined as participants showing maintenance or improvement in Alzheimer's Disease Assessment Scale-cognitive Cognitive Subscale [ADAS-Cog]-14 scores) between the ET-101 and sham control groups, hypothesizing that the responder rate would be higher in the ET-101 group. Second, regarding secondary outcomes, we assessed changes from baseline to 12- and 24-week follow-ups in validated measures of cognitive function, activities of daily living (ADL), and quality of life, hypothesizing that greater improvements would be observed in the ET-101 group versus the sham control group. Finally, exploratory analyses were conducted to determine whether improvements on ADAS-Cog-14 were predominantly observed in the memory, language, or praxis domains to identify which specific cognitive functions were most responsive to ET-101.

Methods

Study Design and Participants

This was a multicenter, 24-week parallel, randomized, sham device–controlled study of ET-101. The participants were recruited through memory clinics of 7 medical centers in the Republic of Korea. The participants were recruited from June to December 2023. Participants who met the following inclusion criteria were included in the study: (1) individuals aged 55-85 years; (2) patients diagnosed with amnestic MCI by trained psychiatrists or neurologists according to Petersen's criteria, characterized by cognitive complaints and objective memory impairment determined by either the Consortium to Establish a Registry for Alzheimer's Disease Neuropsychological Assessment Battery (CERAD-NP) or the Seoul Neuropsychological Screening Battery (SNSB) verbal learning test [16,17]. Memory impairment was defined as a *z* score of –1 or lower on at least one of the following subtests: an immediate word recall test, a delayed word recall test, or a delayed word recognition part of the verbal learning test in CERAD-NP or SNSB, while maintaining essentially normal functional activities; (3) a Mini Mental State Examination (MMSE) score of 27 or lower [18]; (4) not meeting the diagnostic criteria for dementia; (5) a global Clinical Dementia Rating score of 0.5 [19]; (6) taking a stable dose for at least 12 weeks prior to randomization if prescribed acetylcholinesterase inhibitors or memantine; (7) having a study partner who spends more than 8 hours per week with the participant and agrees to assist with the participant's follow-up and clinical evaluations; (8) possession of a personal smartphone and no difficulties using mobile apps; (9) the ability to make phone calls to a study partner independently using the smartphone; (10) no difficulties reading or writing in Korean; and (11) adequate vision and hearing for participating in the clinical trial.

Participants were excluded from the study if any of the following exclusion criteria were met: (1) a history of transient ischemic attack, stroke, or seizure within the past 12 months; (2) a history of severe psychiatric disorders or currently showing unstable psychiatric symptoms; (3) active suicidal intent as assessed by the Columbia-Suicide Severity Rating Scale or a history of treatment for suicidal behavior within the past 5 years [20]; (4) unstable findings upon a physical examination, neurological examination, in vital signs, or the presence of an ongoing unstable physical illness; (5) substance dependence within the past 2 years; (6) the use of prohibited medications as outlined in Table S1 in Multimedia Appendix 1; (7) scheduled to undergo surgery requiring general anesthesia; or (8) participation in any form of cognitive intervention within the past 3 months. Neurologic and psychiatric history, including past substance dependency, current symptoms, physical status, and neurological function, was assessed through clinical interviews conducted by trained psychiatrists and neurologists.

A statistician, independent of the clinical team, carried out the block randomization task, assigning participants to either the study group or the control group at a 1:1 ratio. Throughout the trial, psychiatrists, neurologists, and outcome raters were blinded to the treatment assignment.

The intervention and control groups adhered to an identical visit schedule. Participants completed 5 visits (0-4): screening (visit 0) included screening for study eligibility; week 0 (visit 1, baseline) included baseline demographic data collection and clinical assessments of cognition, function, and quality of life, along with randomization; week 6 (visit 2) investigated compliance with the study protocol and the occurrence of adverse events (AEs); week 12 (visit 3, at the end of the 12-week treatment period) reassessed the same variables from visit 1 and visit 2; and week 24 (visit 4, follow-up assessment 12 weeks after treatment completion) reassessed the same variables from visit 3.

Sample Size Calculation

Sample size was calculated based on a 2-sample proportion test. Delayed recall has been shown to have the highest predictive accuracy (95.2%) in differentiating dementia from MCI [21]. In a previous study, the proportion of participants showing an improvement on delayed recall was 82.8% in the intervention group and 45.16% in the control group [12]. With a 2-sided α of .05 and power of 0.90, and assuming a 1:1 allocation ratio, the minimum sample size per group was determined to be 32 participants. Anticipating a dropout rate of 35%, we intended to enroll 50 participants per group.

Intervention

Study Group: ET-101

Study group participants used ET-101, a mobile app–based MMT program incorporating multimemory strategies. ET-101 is based on an MMT program developed by Youn and colleagues [22] and was produced by Emocog Inc. We previously introduced offline and smart speaker–based versions of this MMT program, and their effectiveness has been well validated [12,14].

Real-time metamemory teaching is a primary feature of ET-101. The training agent of ET-101 interacts with users through voice-based, 2-way communication to enhance metamemory, including meta-knowledge, meta-monitoring, and meta-judgment. Regarding meta-knowledge, on the first day, the agent begins the training by asking users questions about their long-term memory encoding and the memory strategies

XSL•FO
RenderX

they currently use in daily life. Additionally, during the training sessions, the agent introduces multimemory strategies that can enhance long-term memory and explains how these strategies can improve memory performance. This helps users gain objective insight into their cognitive levels and understand how to compensate for cognitive deficits. Regarding meta-monitoring and meta-judgment, the agent explains the daily trends in users' memory spans based on their training results. The agent identifies users' strong and weak memory strategies, helping them recognize their cognitive strengths and weaknesses. Additionally, at the end of the training, the agent provides guidance on how to apply the learned strategies to daily life, facilitating the generalization of memory strategy use. Throughout the metamemory process, the agent provides personalized feedback, adjusts difficulty levels, and modifies training tasks based on users' responses and training results.

ET-101 includes cognitive exercises specifically designed to incorporate memory strategies focusing on attention, imagination, and associations. According to the memory formation process and neuroscience findings, these strategies have been shown to effectively address issues with memory encoding and retrieval commonly observed in older people with cognitive decline [23,24]. Furthermore, attention, imagination, and association processes are known to strengthen related brain regions, including the prefrontal cortex, hippocampus, parahippocampus, precuneus, and the cingulum that connects these areas [23,25]. Attention training includes exercises that enhance trainees' selective focus on stimuli to be memorized, as well as tasks aimed at improving the processing speed and increasing working memory. Imagination training facilitates the transition to long-term memory by linking spatial-temporal memory and visual imagery associated with the objects to be memorized. In association training, 2 or more pieces of information that need to be remembered are connected and assigned mutual meaning, thereby strengthening the semantic cue to aid in long-term memory formation. The detailed content of each activity can be found in Table S2 in Multimedia Appendix 1, and example screenshots are available in Multimedia Appendix 2.

The study group engaged in a 12-week training program using ET-101. The program was structured to involve 2 sessions each day, 7 days a week, lasting approximately 15 minutes per session. The initial session includes 3 representative cognitive exercises, with 1 session each for attention, imagination, and association. In the later session, 4 additional exercises are conducted, selected from 9 available options, excluding the 3 exercises from the initial session, based on a personalized algorithm (Table S3 in Multimedia Appendix 1). A push alarm was provided daily to encourage training.

Control Group: Sham Device

The control group used a sham device over a 12-week period, which featured cognitive assessments and gamification elements but did not include any cognitive training. In the sham device, on the first day, users are asked the same questions as in ET-101 about their long-term memory encoding and the memory strategies they currently use in daily life, presented in a conversational interaction format with the agent. From the second day onward, it offered a gamification element to encourage daily log-ins, in this case displaying a flower that gradually grows. No content related to cognitive training was provided. The guidelines for program usage frequency were identical to those followed by the study group.

Measures

The efficacy of digital therapeutics was evaluated in terms of cognition, ADL, and quality of life. The scales listed below were assessed by a trained psychologist in a face-to-face manner.

Alzheimer's Disease Assessment Scale-Cognitive Subscale-14

ADAS-Cog-14 was used to assess the severity of cognitive dysfunction [26]. ADAS-Cog-14 consists of the following 14 tasks: (1) word recall, (2) commands, (3) constructional praxis, (4) delayed word recall, (5) naming, (6) ideational praxis, (7) orientation, (8) word recognition, (9) maze, (10) number cancellation, (11) remembering instructions, (12) comprehension, (13) word-finding difficulty, and (14) spoken language ability. The total ADAS-Cog-14 score ranges from 0 to 90 points, calculated by summing the number of errors made on each task. A score of 0 indicates no impairment, while a score of 90 reflects the maximum level of impairment. ADAS-Cog subdomains were grouped into 3 domains: memory, language, and praxis [27]. The composite score range for each subdomain is as follows: memory (0-40), language (0-25), and praxis (0-15), with lower scores indicating less impairment.

Korean Mini Mental State Examination, 2nd Edition

MMSE is a brief neuropsychological test that provides an overview of the cognitive function in individuals with cognitive decline [28]. In this study, we used the latest revised version, the Korean Mini Mental State Examination, 2nd Edition (K-MMSE-II), whose validity has already been established [18]. It tests 5 areas of cognitive function: orientation, registration, attention and calculation, recall, and language. The scale ranges from 0 to 30 points, where higher scores indicate less impairment.

Clinical Dementia Rating-Sum of Boxes

Clinical Dementia Rating-Sum of Boxes (CDR-SB) is used to assess the stage of severity in Alzheimer dementia or MCI [29]. The score is determined through interviews with patients and informants, evaluating cognitive functioning across 6 domains: memory, orientation, judgment and problem-solving, community affairs, home and hobbies, and personal care. Each domain is scored from 0 to 3, and the CDR-SB outcome is calculated by summing the scores across all domains, resulting in a total score ranging from 0 to 18. Higher scores indicate greater impairment.

Alzheimer's Disease Composite Score

Alzheimer's Disease Composite Score (ADCOMS) is a tool developed to measure clinical progression and treatment effects in patients with cognitive decline [30]. ADCOMS comprises a total of 12 items, specifically 4 items from ADAS-Cog, 2 items from the MMSE, and all 6 items from CDR-SB. The total score is calculated by applying predetermined weights to each item. Total ADCOMS values range from 0 to 1.97, with higher scores indicating greater impairment.

XSL•FO
RenderX

Digit Symbol Coding

Digit Symbol Coding (DSC) is designed to assess frontal and executive functions [16]. Participants are shown symbols corresponding to numbers from 1 to 9 and are asked to draw the appropriate symbol in 133 blank spaces. Each correctly drawn symbol is awarded 1 point, resulting in a total score ranging from 0 to 133 points. Higher scores indicate better cognitive function.

Clinician Interview–Based Impression of Severity and Clinician Interview–Based Impression of Change Plus Caregiver Input

The Clinician Interview–Based Impression tool is used to assess the severity of cognitive function through semistructured interviews with patients and their caregivers [31]. The Clinician Interview–Based Impression of Severity (CIBIS) is specifically designed to assess baseline severity. Clinicians evaluate responses across 4 domains—general condition, mental or cognitive state, behavior, and ADL—on a scale of 1 to 7 points, where higher scores indicate greater severity. Clinician Interview–Based Impression of Change Plus Caregiver Input (CIBIC-Plus) is used during follow-up to evaluate whether there has been an improvement or decline compared to the baseline level. It assesses 3 aspects: cognition, behavior, and function. Based on these evaluations, the clinician rates the patient's overall condition on a 7-point Likert scale, where 1 indicates very much improved and 7 indicates marked worsening.

Alzheimer's Disease Cooperative Study-Mild Cognitive Impairment-Activities of Daily Living

The Alzheimer's Disease Cooperative Study-Mild Cognitive Impairment-Activities of Daily Living (ADCS-MCI-ADL) assesses the ability of patients to perform ADL through a structured questionnaire administered to the patient and caregiver by a qualified rater [32]. The scale evaluates both basic and instrumental ADL, with the 24-item version used in this study. Scores range from 0 to 69, with higher scores indicating better ability to independently manage daily living activities, while lower scores reflect a higher degree of dependency when performing such tasks.

EQ-5D-3L and EQ-5D-Visual Analogue Scale

EQ-5D is a tool used to measure health-related quality of life, assessing the 5 dimensions of mobility, self-care, usual activities, pain or discomfort, and anxiety or depression [33]. EQ-5D-3L evaluates each of these dimensions on 3 levels: no problems, some problems, and extreme problems. The responses are then converted into an index score by applying weighted values to each item and summing them, resulting in a score ranging from 0 (death) to 1 (perfect health). In addition, EQ-5D-Visual Analogue Scale (EQ-5D-VAS) captures the respondent's overall assessment of their health on a scale ranging from 0 (worst health imaginable) to 100 (best health imaginable). In this study, both EQ-5D-3L and EQ-5D-VAS were administered to assess the quality of life of patients with MCI as well as the quality of life of their study partners who live with them.

Safety Measures

The occurrence of AEs was monitored actively and systematically, following the Consolidated Standards of Reporting Trials (CONSORT) Harms guideline [34]. AEs were assessed weekly by investigators through telephone contact. Additionally, participants were allowed to report AEs spontaneously at any time during the trial. AEs that involve death, life-threatening conditions, hospitalization or prolonged hospitalization, persistent or significant disability or functional impairment, congenital anomalies or birth defects, or other medically important conditions were classified as serious adverse events (SAEs). Furthermore, the relationship between these AEs and the digital therapeutic device used in this trial was evaluated using a 5-level categorical scale ranging from "Definitely not related" to "Definitely related."

Outcomes

For the primary outcome, participants were categorized as responders or nonresponders based on whether their ADAS-Cog-14 scores were maintained or improved, and the proportion of responders was compared between the ET-101 and sham control groups at weeks 12 and 24. For the secondary outcomes, changes from baseline to 12 and 24 weeks were compared between groups using the following assessments: cognitive function (ADAS-Cog-14, K-MMSE-II, CDR-SB, ADCOMS, and DSC), ADL (CIBIS, CIBIC-Plus, and ADCS-MCI-ADL), and quality of life (EQ-5D-3L and EQ-5D-VAS). For the exploratory analysis, group comparisons of the change scores of the ADAS-Cog-14 subdomains were conducted, including the memory, language, and praxis subdomains. For the safety analysis, the proportions of total AEs, SAEs, and events deemed related to the digital therapeutic intervention were compared between groups.

Statistical Analysis

All statistical analyses were conducted on the full analysis set population, which included all randomized participants who received at least 1 session of the assigned study treatment. This approach is consistent with the modified intention-to-treat concept used in many previous clinical trials [35-37]. Baseline characteristics were compared between the intervention and control groups using an independent 2-tailed *t* test for continuous variables and either a chi-square test or Fisher exact test for categorical variables. The primary outcome, the proportion of responders and nonresponders between the ET-101 group and the control group, was compared by means of a chi-square test. In this study, nonresponder imputation (NRI) was applied, classifying trial dropouts as nonresponders. NRI is a method of handling missing data by assigning all missing cases as nonresponders, thereby preventing treatment benefits from being overestimated. This approach is widely recognized and has been extensively adopted in clinical trials that investigate drug efficacy across various areas [38,39]. In this study, based on NRI, responders were defined as participants who completed the study through week 24 and whose ADAS-Cog-14 score was either maintained or decreased compared to the baseline level, indicating preserved or improved cognitive function [40]. In contrast, nonresponders included those who either failed to

XSL•FO
RenderX

complete the study through week 24 or whose ADAS-Cog-14 score increased compared to the baseline level.

Secondary outcomes, in this case, clinical measures of cognition, ADL, and quality of life, were assessed by analyzing the changes from baseline to week 12 and week 24. Missing data in clinical scales due to trial dropouts were replaced using conditional mean imputation via a regression analysis. For the regression analysis used in imputation, the ADAS-Cog-14, K-MMSE-II, and CDR-SB scores at baseline, as well as the visit and visit-specific values of each clinical measure, were considered. Conditional mean imputation through a regression analysis was conducted using Python (version 3.9.20; Python Software Foundation). Independent *t* tests were used to evaluate whether the score changes differed significantly between the 2 groups. Additionally, a linear mixed model analysis was conducted to examine the treatment-time interaction for each clinical assessment. In this model, treatment, time (indicating baseline, week 12, and week 24), treatment-time interaction, and the baseline score of each scale were included as fixed effects. If a significant difference in the treatment-time interaction between the ET-101 and control groups was found in the linear mixed model, a post hoc analysis was performed to identify the specific time points at which the treatment-time interaction was significant. As a sensitivity analysis, to confirm that the imputation of missing data did not affect the results, the same linear mixed model analysis and post hoc analysis were performed without missing value imputation for variables where a significant treatment-time interaction was found in secondary outcomes.

As an exploratory analysis, the linear mixed model was applied to assess which subdomains of ADAS-Cog-14 (memory, language, and praxis) showed significant treatment-time interaction. If a significant treatment-time interaction was present, a post hoc analysis was conducted to determine the specific time points at which this interaction was significant. As a safety analysis, independent *t* tests were performed for continuous variables, while chi-square tests or Fisher exact tests were used for categorical variables.

All statistical tests were 2-tailed, and *P* scores of <.05 were considered statistically significant. For the post hoc analysis of the linear mixed model, Bonferroni-adjusted *P* scores were reported, with values <.05 indicating statistical significance. All statistical analyses were performed using the *lmerTest* package for the linear mixed model in R software (version 4.4.2; The R Foundation for Statistical Computing).

Ethical Considerations

All participants provided written, informed consent to participate, and their study partners were asked to provide separate written, informed consent. The study was conducted following the Declaration of Helsinki and was approved by the institutional review boards (IRBs) at each of the 7 participating medical centers (Gachon University Gil Medical Center IRB: GCIRB2023-048; Konkuk University Medical Center IRB: 2023-02-017; The Catholic University of Korea, Yeouido St. Mary's Hospital IRB: SC23DDDS0007; Severance Hospital IRB: 10-2023-68; The Catholic University of Korea, Eunpyeong St. Mary's Hospital IRB: PC23DDDS0021; Kangwon National University Hospital IRB: 2023-02-001; SMG-SNU Boramae Medical Center IRB: 1-2023-0005). All participants were anonymized by assigning randomly generated identification numbers. All participants received a compensation of KRW 100,000 (US $68.22) per visit, and study partners were provided with KRW 50,000 (US $34.11) per visit. The study consisted of 5 visits for those who completed the entire study. The reporting of this trial follows the CONSORT statement (Multimedia Appendix 3).

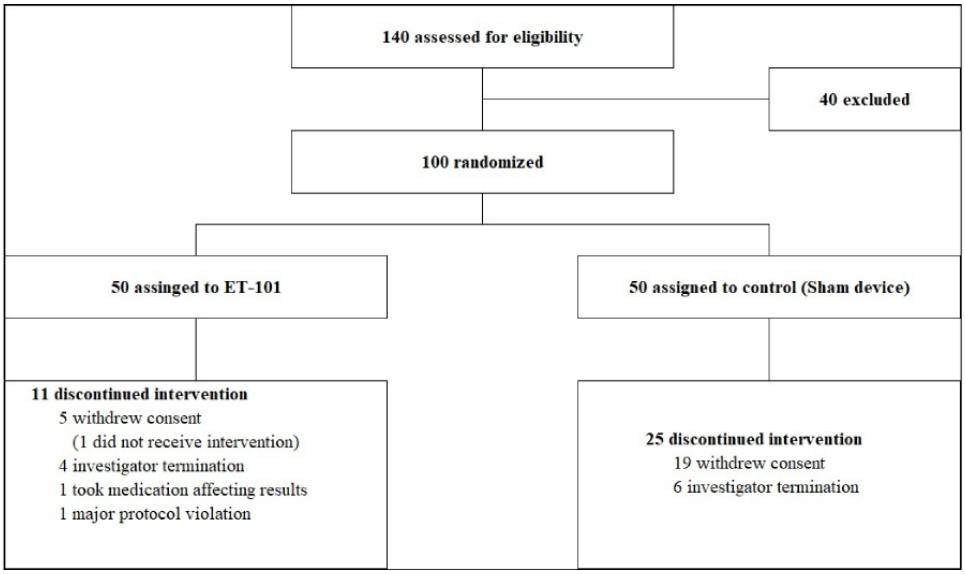
Results

Overview

A total of 140 participants were screened, of whom 40 were excluded based on the inclusion and exclusion criteria (Figure 1). The remaining 100 participants were randomized at a 1:1 ratio. During the clinical trial period, 11 (22.4%) participants in the ET-101 group and 25 (50%) participants in the control group did not complete the study. In the ET-101 group, 1 participant did not use the digital therapeutic device at all after randomization. Therefore, the analysis was conducted on a total of 99 participants in the full analysis set. The dropout rate from the beginning of the training to week 12 was 3 (6%) participants in the ET-101 group and 8 (16%) participants in the sham control group. From week 12 to week 24, the dropout rate was 8 (16%) participants in the ET-101 group and 17 (34%) participants in the sham control group. The adherence rate, defined as the proportion of completed sessions during the 12-week intervention period, was higher in the ET-101 group (mean 83.2%, SD 24.4%) compared to the sham control group (mean 63.8%, SD 31.4%). Daily usage, measured by the number of days participants accessed the program out of the total 84 scheduled training days (12 weeks × 7 days), was also higher in the ET-101 group (mean 69.7, SD 20.2 days) than in the sham control group (mean 53.3, SD 25.8 days). The baseline demographic and clinical characteristics are summarized in Table 1. There were no statistically significant differences in baseline characteristics between the 2 groups.

XSL•FO
RenderX

Figure 1. Flowchart of the study process.



XSL•FO
RenderX

Table 1. Baseline characteristics of study participants.

Characteristics	ET-101 (n=49)	Control (n=50)	*P* value
Age (years), mean (SD)	75 (5.77)	74.8 (5.79)	.84
Sex, n (%)			.91
Male	21 (42.9)	22 (44)	
Female	28 (57.1)	28 (56)	
Education (years), mean (SD)	10.9 (4.47)	10.1 (4.82)	.34
Current medical illness, n (%)			.96
Yes	37 (75.5)	38 (76)	
No	12 (24.5)	12 (24)	
Current use of acetylcholinesterase inhibitors or memantine , n (%)			.90
Yes	18 (36.7)	19 (38)	
No	31 (63.3)	31 (62)	
ADAS-Cog-14[a], mean (SD)[b]	29.2 (6.88)	31 (6.23)	.17
K-MMSE-II[c], mean (SD)	23.9 (2.79)	23 (2.7)	.12
CDR-SB[d], mean (SD)[b]	1.66 (0.856)	2.03 (0.997)	.05
ADCOMS[e], mean (SD)[b]	0.295 (0.146)	0.347 (0.128)	.07
DSC[f], mean (SD)	38.2 (15.4)	34.7 (17.1)	.29
CIBIS[g], mean (SD)[a]	2.67 (0.516)	2.82 (0.523)	.16
ADCS-MCI-ADL[h], mean (SD)	44.3 (6.59)	41.6 (7.06)	.05
EQ-5D-3L patient, mean (SD)	0.929 (0.090)	0.909 (0.087)	.27
EQ-5D-3L study partner, mean (SD)	0.926 (0.092)	0.946 (0.079)	.25
EQ-5D-VAS[i] patient, mean (SD)	78.6 (13)	77.3 (18.7)	.70
EQ-5D-VAS study partner, mean (SD)	76.8 (13.6)	82.2 (14)	.06

[a]ADAS-Cog-14: Alzheimer's Disease Assessment Scale-Cognitive Subscale-14.

[b]Lower scores represent better performance.

[c]K-MMSE-II: Korean Mini Mental State Examination, 2nd Edition.

[d]CDR-SB: Clinical Dementia Rating-Sum of Boxes.

[e]ADCOMS: Alzheimer's Disease Composite Score.

[f]DSC: Digit Symbol Coding.

[g]CIBIS: Clinician Interview–Based Impression of Severity.

[h]ADCS-MCI-ADL: Alzheimer's Disease Cooperative Study-Mild Cognitive Impairment-Activities of Daily Living.

[i]EQ-5D-VAS: EQ-5D-Visual Analogue Scale.

Primary Outcome

As shown in Table 2, the proportion of responders and nonresponders did not differ significantly between the ET-101 group and the control group at week 12 (*P*=.47). However, at visit 4, the proportion of responders was significantly higher in the ET-101 group (n=29, 59.2% responders) compared to the sham device group (n=14, 28% responders; *P*=.002).

XSL•FO
RenderX

Table 2. Comparison of the proportion of responders and nonresponders between the ET-101 and control groups.

	ET-101 (n=49), n (%)	Control (n=50), n (%)	*P* value
12 weeks			.47
Responder	20 (40.8)	24 (48)	
Nonresponder	29 (59.2)	26 (52)	
24 weeks			.002
Responder	29 (59.2)	14 (28)	
Nonresponder	20 (40.8)	36 (72)	

Secondary Outcome

The changes in clinical measures from baseline to weeks 12 and 24 are presented in Table 3. For ADAS-Cog-14, the change in scores from baseline to week 12 was compared between the ET-101 and control groups, showing no statistically significant differences (change difference=1.42; *P*=.13). However, at week 24, the change in scores from baseline showed a statistically significant improvement in the ET-101 group compared to the control group (change difference=–2.55; *P*=.02). These findings remained consistent even when examining the treatment-time interaction effect in a linear mixed model that adjusted for baseline ADAS-Cog-14 scores ($F_{2,\ 194}$=7.45; *P*<.001). In the post hoc analysis conducted to identify the specific time points associated with the significant treatment-time interaction, no significant differences were observed between the 2 groups at baseline (estimates=0.022; t_{265}=0.026; Bonferroni-adjusted *P*=.98) or week 12 (estimates=1.44; t_{265}=1.74; Bonferroni-adjusted *P*=.08). However, at week 24, cognitive improvement as measured by ADAS-Cog-14 was statistically significant in the ET-101 group compared to the control group (estimates=–2.53; t_{265}=–3.05; Bonferroni-adjusted *P*=.003; Figure 2). For the other clinical measures, no statistically significant differences were observed.

XSL•FO
RenderX

Table 3. Comparison of cognition, activities of daily living, and quality of life questionnaires between the ET-101 and control groups.

Questionnaire and week	ET-101 (n=49), mean (SD)	Control (n=50), mean (SD)	Change difference[a] (95% CI)	Linear mixed model (treatment-time interaction)	
				F test (*df*)	*P* value
ADAS-Cog-14[b,c]				7.45 (2, 194)	<.001
Week 12	30.3 (7.81)	30.7 (7.40)	1.42 (–0.44 to 3.29)		
Week 24	25.8 (9.52)	30.1 (8.55)	–2.55 (–4.69 to –0.417)[d]		
K-MMSE-II[e]				0.454 (2, 194)	.64
Week 12	24.1 (3.34)	23.6 (2.82)	–0.377 (–1.38 to 0.631)		
Week 24	23.6 (3.16)	23.1 (2.74)	–0.386 (–1.47 to 0.703)		
CDR-SB[b,f]				0.865 (2, 194)	.42
Week 12	1.68 (1.00)	1.92 (1.04)	0.130 (–0.111 to 0.372)		
Week 24	1.43 (1.01)	1.85 (1.02)	–0.055 (–0.381 to 0.272)		
ADCOMS[b,g]				0.608 (2, 194)	.55
Week 12	0.315 (0.175)	0.343 (0.154)	0.024 (–0.014 to 0.062)		
Week 24	0.267 (0.180)	0.310 (0.161)	0.008 (–0.044 to 0.060)		
DSC[h]				0.781 (2, 194)	.46
Week 12	39.1 (14.9)	34.8 (18.8)	0.798 (–1.34 to 2.94)		
Week 24	39.0 (14.1)	33.9 (16.0)	1.62 (–1.29 to 4.52)		
CIBIC-Plus[b,i]				0.526 (1, 97)	.47
Week 12	3.84 (0.717)	4.02 (0.553)	–0.183 (–0.439 to 0.073)[j]		
Week 24	4.02 (0.433)	4.10 (0.364)	–0.080 (–0.239 to 0.080)[j]		
ADCS-MCI-ADL[k]				1.73 (2, 194)	.18
Week 12	43.3 (5.78)	40.7 (6.41)	–0.101 (–1.73 to 1.53)		
Week 24	43.2 (5.86)	41.9 (6.01)	–1.40 to (–3.16 to 0.352)		
EQ-5D-3L patient				0.963 (2, 194)	.38
Week 12	0.925 (0.095)	0.908 (0.098)	–0.004 (–0.038 to 0.030)		
Week 24	0.920 (0.108)	0.925 (0.085)	–0.024 (–0.060 to 0.012)		
EQ-5D-3L study partner				1.50 (2, 194)	.23
Week 12	0.924 (0.101)	0.937 (0.106)	0.008 (–0.028 to 0.045)		
Week 24	0.926 (0.096)	0.967 (0.056)	–0.021 (–0.050 to 0.009)		
EQ-5D-VAS[l] patient				0.705 (2, 194)	.50
Week 12	77.5 (15.9)	75.2 (15.8)	0.978 (–5.26 to 7.21)		
Week 24	72.3 (15.4)	67.5 (13.8)	3.53 (–2.39 to 9.46)		
EQ-5D-VAS study partner				1.92 (2, 194)	.15
Week 12	77.7 (11.6)	81.6 (13.0)	1.38 (–4.14 to 6.90)		
Week 24	77.0 (10.7)	77.2 (12.0)	5.12 (–0.873 to 11.1)		

[a]The score change from baseline to weeks 12 and 24 in the ET-101 group minus the score change from baseline to weeks 12 and 24 in the control group.

[b]Lower scores represent better performance.

[c]ADAS-Cog-14: Alzheimer's Disease Assessment Scale-Cognitive Subscale-14.

[d]*P*<.05.

[e]K-MMSE-II: Korean Mini Mental State Examination, 2nd Edition.

[f]CDR-SB: Clinical Dementia Rating-Sum of Boxes.

XSL•FO
RenderX

[g]ADCOMS: Alzheimer's Disease Composite Score.

[h]DSC: Digit Symbol Coding.

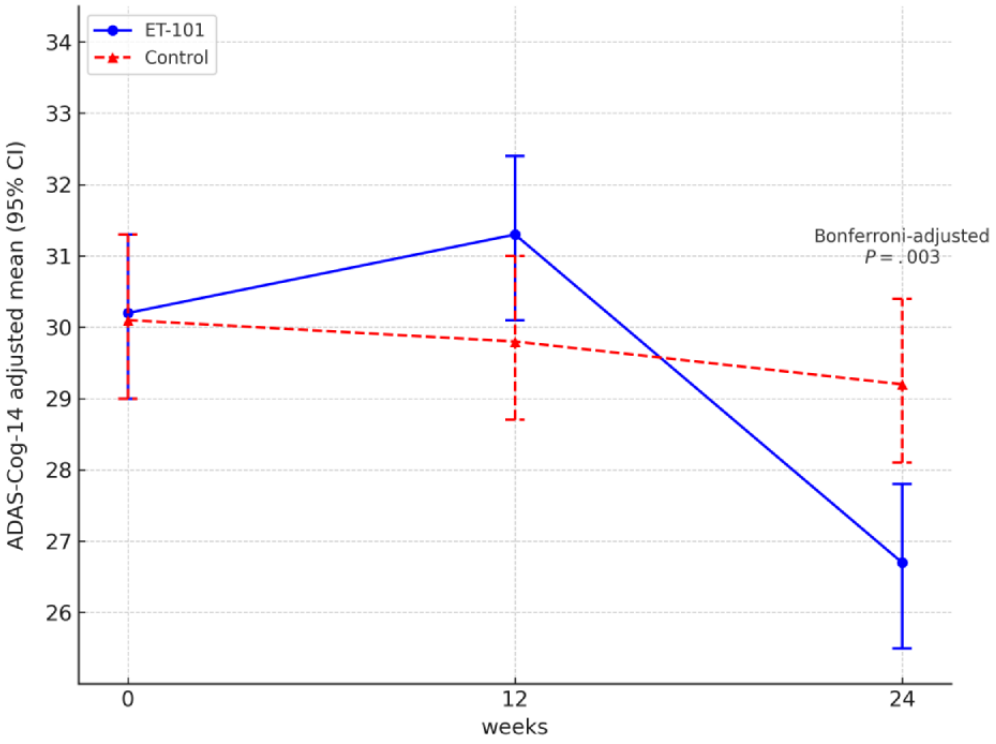
[i]CIBIC-Plus: Clinician Interview–Based Impression of Change Plus Caregiver Input.

[j]Due to the inability to present mean change, the difference in mean change between the groups was replaced with score differences across time points.

[k]ADCS-MCI-ADL: Alzheimer's Disease Cooperative Study-Mild Cognitive Impairment-Activities of Daily Living.

[l]EQ-5D-VAS: EQ-5D-Visual Analogue Scale.

Figure 2. Comparison of adjusted mean change of Alzheimer's Disease Assessment Scale-Cognitive Subscale-14 (ADAS-Cog-14) between ET-101 and the control group using a linear mixed model. Lower scores represent better performance. ADAS-Cog-14 scores adjusted for baseline differences.



Sensitivity Analysis

Even when the linear mixed model was used without applying conditional mean imputation for missing data, the treatment-time interaction remained statistically significant in ADAS-Cog-14 ($F_{2,\ 176}$=3.91; *P*=.02). In the post hoc analysis, no statistically significant differences were observed at baseline (estimates=0.018; t_{232}=0.021; Bonferroni-adjusted *P*=.98) or at week 12 (estimates=1.53; t_{235}=1.69; Bonferroni-adjusted *P*=.09). At week 24, the Bonferroni-adjusted *P* value was .05, indicating a trend toward greater improvement in ADAS-Cog-14 scores in the ET-101 group compared to those of the control group (estimates=–2.13; t_{244}=–1.96).

Exploratory Analysis

The results of the linear mixed model and post hoc analysis for the subdomains of ADAS-Cog-14 are presented in Table S4 in Multimedia Appendix 1. Among the subdomains of ADAS-Cog-14, the treatment-time interaction in the memory domain was statistically significant ($F_{2,\ 194}$=11.2; *P*<.001). In the post hoc analysis, ET-101 showed a statistically significant improvement in the memory domain compared to the control group at week 24 (estimates=–2.50; t_{264}=–4.03; Bonferroni-adjusted *P*<.001).

The treatment-time interaction was also found to be statistically significant for the language domain ($F_{2,\ 194}$=3.99; *P*=.02). In the post hoc analysis, ET-101 demonstrated a statistically significant improvement compared to the control group at week 24 (estimates=–0.807; t_{290}=–3.68; Bonferroni-adjusted *P*<.001). For the praxis subdomain, representing executive function, the treatment-time interaction was not statistically significant ($F_{2,\ 194}$=0.497; *P*=.67).

Safety Analysis

In the ET-101 group, 6 AEs occurred, while 4 AEs were reported in the control group, with no statistically significant difference in AE incidence between the groups (*P*=.53; Table 4). Among the AEs in the ET-101 group, 2 were classified as SAEs due to prolonged hospitalization (non-ST elevation myocardial infarction and herniated disc disease of the lumbar spine); however, the incidence of SAEs also did not differ significantly between the 2 groups (*P*=.47). All reported AEs were determined to be definitely not related to ET-101.

XSL•FO
RenderX

Table 4. Comparison of safety outcomes between ET-101 and the control group.

	ET-101 (n=49), n (%)	Control (n=50), n (%)	*P* value
Serious adverse events			.47
Yes	2 (4.1)	0 (0)	
No	4 (8.2)	4 (8)	
Relevance to digital therapeutics			N/A[a]
Definitely related	0 (0)	0 (0)	
Probably related	0 (0)	0 (0)	
Possibly related	0 (0)	0 (0)	
Probably not related	0 (0)	0 (0)	
Definitely not related	6 (12.2)	4 (8)	
Unknown	0 (0)	0 (0)	
Total	6 (12.2)	4(8)	.52

[a]N/A: not applicable.

Discussion

Principal Findings

This study investigated the efficacy and safety of ET-101 in patients with MCI compared to a sham device control group. We found that the ET-101 group showed significantly greater improvement in the ADAS-Cog-14 score in 24 weeks compared to the control group. Furthermore, no AEs related to the use of ET-101 were observed. Among the various types of digital cognitive training, to the best of our knowledge, this study is the first to investigate the effects of MMT using multimemory strategies in the form of a mobile app.

ET-101, a metamemory-based multistrategic cognitive training program, showed improvement in global cognitive function as well as memory function. In the primary outcome, the responder rate at 24 weeks was significantly higher in the ET-101 group. Additionally, in the secondary outcomes, the ADAS-Cog-14 score showed a statistically significant improvement in the ET-101 group compared to the control group at 24 weeks. According to a meta-analysis that included trials on computerized cognitive training for individuals with MCI, independently delivered digital cognitive training programs, such as ET-101, demonstrated a significant effect on verbal episodic memory, with a standardized mean difference of 0.21 (95% CI 0.04-0.38). This is consistent with the observed memory improvement in this study [41]. ET-101 demonstrated cognitive improvement comparable to that observed with donepezil in terms of ADAS-Cog score changes, suggesting that ET-101 may have a similar cognitive improvement effect [42]. One of the most well-known protective factors against memory decline in MCI is cognitive reserve, which refers to the ability to use cognitive processes and brain networks adaptively to compensate for deterioration [43]. In MCI, there is not only a reduction in the hippocampal volume but also a weakening of the connectivity among key brain regions involved in memory function, such as the frontal, temporal, and parietal lobes [44]. With ET-101, MMT enables individuals to monitor their own memory function objectively and select the most efficient memory strategies tailored to their needs, facilitating the frontal lobe [14]. Additionally, efficient memory strategies, including attention, imagination, and association, enhance the connectivity of brain regions related to cognitive reserve [23,24]. Biological evidence also supports the cognitive reserve–enhancing effects of metamemory. A previous study of MMT found reduced mean diffusivity in the left superior longitudinal fasciculus and corona radiata, which link the frontal, temporal, and parietal lobes [14], suggesting that ET-101 may contribute to enhancing cognitive reserve, which in turn could support improvements in memory function.

Another characteristic of ET-101's cognitive function improvement is that it has a transfer effect on the language domain. In the subdomain analysis of ADAS-Cog-14, the ET-101 group showed a significantly greater improvement in the language domain compared to the control group at 24 weeks. Meta-analyses of several studies of cognitive training in patients with MCI have not shown significant improvements in the language domain [41]. Earlier work has suggested that cognitive training gains typically reflect the training content and that insufficient training on other cognitive domains may therefore make improvements beyond the memory domain less observable [41]. However, MMT-based multistrategic cognitive training, on which ET-101 is based, has consistently demonstrated functional improvements not only in memory but also in the language domain across multiple studies [11,12,14]. One possible explanation for this transfer effect is that the metamemory and memory strategies of ET-101 repeatedly strengthen the connection between episodic memory and semantic memory. MMT helps individuals select compensatory strategies to address their cognitive deficits, and ET-101, in particular, repeatedly trains strategies such as imagination and association, facilitating the connection between episodic and semantic memory [45]. This is especially relevant to lexical production, which is impaired in MCI and requires effortful semantic memory processing, such as in fluency and naming tasks [46]. Notably, naming and semantic fluency have been identified in the literature as strong predictors of the conversion from MCI to Alzheimer dementia, and ET-101 appears to

XSL•FO
RenderX

reinforce cognitive functions that are closely associated with this deterioration [21]. Additional long-term studies are needed to investigate the effects of ET-101 on the rate of conversion from MCI to Alzheimer dementia and to assess its long-term efficacy further.

The observed improvement in the memory and language domains following ET-101 training is closely linked to the memory strategies embedded in the program, which are known to support memory formation. ET-101 primarily incorporates 3 core memory strategies: attention, imagination, and association. First, attention serves as an initial step in memory formation by facilitating selective encoding. Numerous neuroimaging studies have demonstrated that attention enhances memory by activating interactions between the hippocampus and fronto-parietal regions within the dorsal attention network [47-49]. This activation is thought to strengthen neocortical representations, thereby supporting long-term memory storage [50]. Clinically, studies have shown that increased activation of the dorsal attention network is positively correlated with the accuracy and strength of memory recall [51]. Next, imagination shares common neural circuits with episodic memory, involving the hippocampus, medial prefrontal cortex, and anterolateral temporal cortex—regions that are closely associated with memory consolidation [52,53]. Hippocampal replay facilitates the reconstruction of sensory experiences by integrating visual, sensory, and semantic information [52]. This process allows imagination to contribute to memory formation and strengthening, even in the absence of direct sensory input [52]. Regarding association, it has been suggested that newly learned words can facilitate the recall of previously acquired information through retroactive facilitation [54]. Notably, neural evidence indicates that while preexisting semantic relationships are primarily represented within the neocortex, unrelated word pairs can be bound by the hippocampus through associative processes, along with their episodic list context and novelty [54]. Taken together, findings from previous studies suggest that the strategies used in ET-101—attention, imagination, and association—effectively activate the hippocampus and related regions, thereby potentially compensating for the memory deficits observed in individuals with MCI.

However, ET-101 did not show a significant difference compared to the control group on the praxis subdomain, ADL, or quality of life measures. There was no statistically significant difference between the ET-101 and control groups in ADL assessments using CIBIC-Plus and ADCS-MCI-ADL, or in quality-of-life assessments for patients with MCI and their study partners using EQ-5D. Additionally, the praxis domain of ADAS-Cog-14 did not show a significant difference between the groups up to 24 weeks. These findings appear to be related to the characteristics of MCI and the measured scales. MCI involves cognitive decline, but the functional impairment at this stage is not severe enough to warrant a diagnosis of overt dementia [55,56]. Given that those with MCI typically manage daily activities well, ADL and quality-of-life scales may have a ceiling effect, reducing their sensitivity to detect meaningful changes [57]. Therefore, to observe subtle changes in ADL among individuals with MCI, this study used the ADCS-MCI-ADL scale, which is sensitive to detecting changes in ADL among individuals with MCI [32]. In prior pharmacological trials, reported improvements in ADCS-MCI-ADL included participants with both MCI and mild Alzheimer disease [3,58]. However, such improvements in ADCS-MCI-ADL have typically been observed over longer durations, such as 18-month follow-up periods, whereas the 24-week follow-up in our study may not have been long enough to detect meaningful changes [3,58]. This suggests that longer-term follow-up may be necessary to capture significant improvements in ADL or quality of life associated with ET-101. Similarly, in ADAS-Cog-14, the praxis domain has been shown to be less effective in distinguishing different stages of cognitive decline compared to the memory and language domains, and it appears to exhibit a ceiling effect in MCI [26]. Given recent efforts to develop more sensitive scales for detecting changes in ADL, quality of life, and cognitive function in MCI, future studies using these updated tools will be essential for accurately evaluating the effects of ET-101.

The improvement in cognitive function with ET-101 followed a delayed pattern, with significant effects emerging at the 24-week follow-up. While no group differences were observed immediately after the 12-week training period in either the primary or secondary outcomes, the ET-101 group showed significantly greater responder rates and ADAS-Cog-14 improvements at 24 weeks. These findings suggest that the cognitive benefits of ET-101 may not be immediate but instead become evident and persist after the intervention ends. This pattern contrasts with traditional cognitive training, which often shows immediate but short-lived effects. For example, the Advanced Cognitive Training for Independent and Vital Elderly trial reported improvements immediately after memory training, but these effects were not maintained at the 1-year follow-up [7]. In a study of computerized cognitive training that included follow-up data, the initially significant improvements in attention and memory observed immediately after the intervention were no longer present at the 3-month follow-up [8]. Previous meta-analyses have suggested that the effects of memory strategy training tend to be short-lived [59]. Moreover, most studies have not evaluated maintenance effects, resulting in limited evidence on whether training-induced improvements observed during the intervention are sustained in real-life situations [15]. In contrast, ET-101 showed significant gains after the training ended. This delayed onset and sustained improvement align with findings from strategy-based training programs, such as those including memory education, skill practice, and cognitive restructuring of memory-related beliefs [60-62]. Similar to ET-101, these studies did not show immediate cognitive improvements posttraining but demonstrated significant effects during follow-up assessments [60-62]. Such outcomes suggest that benefits emerge gradually as participants apply learned strategies in real-world contexts [61-63]. From a neurobiological perspective, memory consolidation involves synaptic consolidation, which includes short-term molecular processes such as protein synthesis and synaptic potentiation occurring within minutes to hours after memory formation, followed by systems consolidation, during which the interaction between the hippocampus and neocortex is gradually strengthened [64,65]. System consolidation can take several months to years in humans, and sufficient time may

XSL•FO
RenderX

be required to detect changes in memory ensembles and their associated engram networks, which could explain the delayed onset of cognitive improvement observed in this study [65]. However, alternative explanations cannot be excluded. The cognitive improvement observed at the 24-week follow-up may also be partially influenced by practice effects in cognitive testing or by reduced psychological distress after the completion of trial participation, leading to more apparent posttrial improvements. Moreover, the underlying mechanisms of the delayed and lasting cognitive effects remain unclear, and further longitudinal follow-up and functional neuroimaging studies are warranted to elucidate the temporal dynamics of training-induced neural plasticity. Future meta-analyses and systematic reviews should also investigate the characteristics of cognitive training protocols associated with either immediate or delayed effects to clarify the factors that determine these distinct temporal response patterns.

The cognitive improvement effect of ET-101 was observed in the ADAS-Cog-14 case, but not on other cognitive function assessment scales. In the secondary outcomes, the measures K-MMSE-II, CDR-SB, ADCOMS, and DSC did not show a statistically significant treatment-time interaction. These findings may be related to the sensitivity of cognitive function assessment scales. Previous studies comparing ADAS-Cog-14 with other cognitive measures have shown that ADAS-Cog can detect cognitive changes in MCI significantly more sensitively compared to MMSE and CDR-SB [66,67]. Notably, this study used the 14-item version of ADAS-Cog, which has been specifically reported to be highly sensitive for MCI assessment [26]. Therefore, the observed improvement in ADAS-Cog-14, but not on other scales, is likely due to differences in the sensitivity of these assessments in MCI, with ADAS-Cog-14 being more capable of detecting subtle cognitive changes.

Limitations

One limitation of this study is the relatively high number of dropouts during the trial. During the 12-week training period, 3 (6%) participants in the ET-101 group and 8 (16%) participants in the sham control group dropped out, values that fall within or below the typical range of dropout rates (30%-40%) reported in previous digital intervention studies [68,69]. However, between week 12 and week 24—after the intervention had ended—additional dropouts occurred in both groups: 8 (16%) participants in the ET-101 group and 17 (34%) in the sham control group. This increase may reflect reduced motivation to continue participation in the absence of active training. As these attrition rates necessitated an imputation strategy for missing data, the interpretation of the results should be approached with caution. In particular, the NRI method may overpenalize attrition and potentially bias the treatment effect estimates downward, especially when the reasons for dropout differ between groups. Therefore, to minimize the impact of the imputation used in this study, we conducted analyses to verify whether the results remained consistent across different statistical approaches. These included a comparison of the proportion of participants showing maintenance or improvement in cognitive function between the ET-101 and sham control groups based on NRI, as well as analyses of continuous cognitive scores using both independent *t* tests and linear mixed models. Furthermore, to minimize the impact of the imputation as conducted here, we used the same linear mixed model without imputation for missing data. Across these multiple analytic approaches, we consistently observed that ET-101 produced significantly better maintenance or improvement of cognitive function compared to sham controls at the 24-week follow-up.

Additionally, certain clinical variables, in this case CDR-SB and EQ-5D-VAS, had *P* values close to .05 in baseline comparisons, indicating that the 2 groups were not entirely homogeneous in terms of certain variables. To address this, we incorporated each scale's baseline score as a fixed effect in the linear mixed model when analyzing secondary outcomes. This approach allowed us to confirm that the treatment-time interaction remained statistically significant even after adjusting for these differences.

Third, although participants were not explicitly informed of their group assignment, differences in content between the ET-101 and sham control groups may have caused some participants to infer their allocation based on the app experience. This is a common issue in studies evaluating the efficacy of software-based interventions, where a sham condition must be provided without including the active training components. As such, there is a potential risk of performance and expectation bias between groups, which must be taken into consideration when interpreting the results.

Fourth, this study included only participants who were able to own and operate a smartphone, which may have led to the inclusion of a relatively more tech-savvy population. However, this inclusion criterion was necessary because ET-101 is designed to be used independently without additional assistance during the training process, targeting individuals with basic smartphone literacy. In particular, in the context of Korea, epidemiological data from 2023 show that approximately 95% of individuals aged 60 years and older own a personal mobile phone [70]. Therefore, this inclusion criterion does not substantially deviate from the characteristics of the older adult population in Korea.

Conclusions

Metamemory-based multistrategic cognitive training, ET-101, was found to improve cognitive function compared to a sham device control group. This effect was observed in both the targeted domain of memory and in the language domain, indicating a transfer effect. Additionally, the cognitive improvement became pronounced at 24 weeks, suggesting that ET-101 has a sustained effect beyond the training period. These findings indicate that ET-101 has the potential to provide effective MMT to a broader population with MCI, overcoming location and personnel limitations via its mobile app–based platform.

XSL•FO
RenderX

Acknowledgments

This research was financially supported by the Ministry of Trade, Industry, and Energy, Korea, under the “Regional Innovation Cluster Development Program (R&D, P0025292)” supervised by the Korea Institute for Advancement of Technology. Additionally, ChatGPT 4o was used solely for the purpose of grammar checking. No AI tools were used to generate original content or interpret study findings.

Data Availability

The datasets analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

KHP conceived the study, designed the research, and supervised the project. KHP, KYK, EK, HJJ, DHP, HKL, YS, JWJ, and YK collected and organized patient data. SK conducted the data analysis and drafted the manuscript. SK, JIL, and LS interpreted the results. SL was involved in drafting the description of the ET-101. All authors critically reviewed and revised the manuscript. All authors have read and approved the final version for submission.

Conflicts of Interest

Emocog Inc owns the mobile app evaluated in this study and has a commercial interest in its future distribution. SL is employed as a software developer at Emocog Inc. SK was employed in the Research and Development team at Emocog Inc. until August 31, 2025. LS is the Chief Operating Officer of Cogthera GmbH, a subsidiary of Emocog Inc. Cogthera GmbH is responsible for the commercialization and planned distribution of a modified version of this application in Germany. However, SK and LS were not involved in participant enrollment or data collection, which was conducted independently by other authors (KYK, EK, HJJ, DHP, HKL, YS, JWJ, YK, and KHP), who had no conflicts of interest.

Multimedia Appendix 1
Supplementary tables.
[XLSX File (Microsoft Excel File), 16 KB - mhealth_v14i1e73464_app1.xlsx]

Multimedia Appendix 2
Sample exercise from the ET-101 cognitive training program.
[PDF File (Adobe PDF File), 413 KB - mhealth_v14i1e73464_app2.pdf]

Multimedia Appendix 3
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1148 KB - mhealth_v14i1e73464_app3.pdf]

References

1. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. Lancet Public Health 2022;7(2):e105-e125 [FREE Full text] [doi: 10.1016/S2468-2667(21)00249-8] [Medline: 34998485]
2. Xu Z, Sun W, Zhang D, Chung VC, Sit RW, Wong SY. Comparative effectiveness of interventions for global cognition in patients with mild cognitive impairment: a systematic review and network meta-analysis of randomized controlled trials. Front Aging Neurosci 2021;13:653340 [FREE Full text] [doi: 10.3389/fnagi.2021.653340] [Medline: 34220484]
3. van Dyck CH, Swanson CJ, Aisen P, Bateman RJ, Chen C, Gee M, et al. Lecanemab in early alzheimer's disease. N Engl J Med 2023;388(1):9-21. [doi: 10.1056/NEJMoa2212948] [Medline: 36449413]
4. Raimo S, Cropano M, Gaita M, Maggi G, Cavallo ND, Roldan-Tapia MD, et al. The efficacy of cognitive training on neuropsychological outcomes in mild cognitive impairment: a meta-analysis. Brain Sci 2023;13(11):1510 [FREE Full text] [doi: 10.3390/brainsci13111510] [Medline: 38002471]
5. Chapman SB, Aslan S, Spence JS, Hart JJ, Bartz EK, Didehbani N, et al. Neural mechanisms of brain plasticity with complex cognitive training in healthy seniors. Cereb Cortex 2015;25(2):396-405 [FREE Full text] [doi: 10.1093/cercor/bht234] [Medline: 23985135]
6. Hertzog C, Kramer AF, Wilson RS, Lindenberger U. Enrichment effects on adult cognitive development: can the functional capacity of older adults be preserved and enhanced? Psychol Sci Public Interest 2008;9(1):1-65. [doi: 10.1111/j.1539-6053.2009.01034.x] [Medline: 26162004]
7. Sisco SM, Marsiske M, Gross AL, Rebok GW. The influence of cognitive training on older adults' recall for short stories. J Aging Health 2013;25(8 Suppl):230S-248S [FREE Full text] [doi: 10.1177/0898264313501386] [Medline: 24385636]

XSL•FO
RenderX

8. Zelinski EM, Spina LM, Yaffe K, Ruff R, Kennison RF, Mahncke HW, et al. Improvement in memory with plasticity-based adaptive cognitive training: results of the 3-month follow-up. J Am Geriatr Soc 2011;59(2):258-265. [doi: 10.1111/j.1532-5415.2010.03277.x] [Medline: 21314646]
9. Flavell JH. Metacognition and cognitive monitoring: a new area of cognitive–developmental inquiry. Am Psychol 1979;34(10):906-911. [doi: 10.1037/0003-066x.34.10.906]
10. Tarricone P. The Taxonomy of Metacognition. London: Psychology Ppress; 2011.
11. Youn JH, Lee JY, Kim S, Ryu SH. Multistrategic memory training with the metamemory concept in healthy older adults. Psychiatry Investig 2011;8(4):354-361 [FREE Full text] [doi: 10.4306/pi.2011.8.4.354] [Medline: 22216046]
12. Kim J, Shin E, Han K, Park S, Youn JH, Jin G, et al. Efficacy of smart speaker-based metamemory training in older adults: case-control cohort study. J Med Internet Res 2021;23(2):e20177 [FREE Full text] [doi: 10.2196/20177] [Medline: 33591276]
13. Sella E, Carbone E, Vincenzi M, Toffalini E, Borella E. Efficacy of memory training interventions targeting metacognition for older adults: a systematic review and meta-analysis. Aging Ment Health 2023;27(4):674-694. [doi: 10.1080/13607863.2022.2122931] [Medline: 36218025]
14. Youn JH, Ryu SH, Lee JY, Park S, Cho SJ, Kwon H, et al. Brain structural changes after multi-strategic metamemory training in older adults with subjective memory complaints: a randomized controlled trial. Brain Behav 2019;9(5):e01278 [FREE Full text] [doi: 10.1002/brb3.1278] [Medline: 30916450]
15. Chan ATC, Ip RTF, Tran JYS, Chan JYC, Tsoi KKF. Computerized cognitive training for memory functions in mild cognitive impairment or dementia: a systematic review and meta-analysis. NPJ Digit Med 2024;7(1):1 [FREE Full text] [doi: 10.1038/s41746-023-00987-5] [Medline: 38172429]
16. Ryu HJ, Yang DW. The Seoul Neuropsychological Screening Battery (SNSB) for comprehensive neuropsychological assessment. Dement Neurocogn Disord 2023;22(1):1-15 [FREE Full text] [doi: 10.12779/dnd.2023.22.1.1] [Medline: 36814700]
17. Lee JH, Lee KU, Lee DY, Kim KW, Jhoo JH, Kim JH, et al. Development of the Korean version of the Consortium to Establish a Registry for Alzheimer's Disease Assessment Packet (CERAD-K): clinical and neuropsychological assessment batteries. J Gerontol B Psychol Sci Soc Sci 2002;57(1):P47-P53. [doi: 10.1093/geronb/57.1.p47] [Medline: 11773223]
18. Baek MJ, Kim K, Park YH, Kim S. The validity and reliability of the Mini-Mental State Examination-2 for detecting mild cognitive impairment and Alzheimer's disease in a Korean population. PLoS One 2016;11(9):e0163792 [FREE Full text] [doi: 10.1371/journal.pone.0163792] [Medline: 27668883]
19. Choi SH, Na DL, Lee BH, Hahm DS, Jeong JH, Yoon SJ, et al. Estimating the validity of the Korean version of expanded Clinical Dementia Rating (CDR) scale. J Korean Neurol Assoc 2001;19(6):585-591 [FREE Full text]
20. Pai D, Woo JM, Son MH, Lee C. The reliability and validity of the Korean version of Columbia-Suicide Severity Rating Scale in alcohol dependent patients. J Korean Neuropsychiatr Assoc 2015;54(2):222-227 [FREE Full text]
21. Belleville S, Fouquet C, Hudon C, Zomahoun HTV, Croteau J, Consortium for the Early Identification of Alzheimer's disease-Quebec. Neuropsychological measures that predict progression from mild cognitive impairment to Alzheimer's type dementia in older adults: a systematic review and meta-analysis. Neuropsychol Rev 2017;27(4):328-353 [FREE Full text] [doi: 10.1007/s11065-017-9361-5] [Medline: 29019061]
22. Youn J, Park S, Lee J, Cho S, Kim J, Ryu S. Cognitive improvement in older adults with mild cognitive impairment: evidence from a multi-strategic metamemory training. J Clin Med 2020;9(2):362 [FREE Full text] [doi: 10.3390/jcm9020362] [Medline: 32013035]
23. Cabeza R, Grady CL, Nyberg L, McIntosh AR, Tulving E, Kapur S, et al. Age-related differences in neural activity during memory encoding and retrieval: a positron emission tomography study. J Neurosci 1997;17(1):391-400 [FREE Full text] [doi: 10.1523/JNEUROSCI.17-01-00391.1997] [Medline: 8987764]
24. Craik FIM, Rose NS. Memory encoding and aging: a neurocognitive perspective. Neurosci Biobehav Rev 2012;36(7):1729-1739. [doi: 10.1016/j.neubiorev.2011.11.007] [Medline: 22155274]
25. Hampstead BM, Stringer AY, Stilla RF, Sathian K. Mnemonic strategy training increases neocortical activation in healthy older adults and patients with mild cognitive impairment. Int J Psychophysiol 2020;154:27-36 [FREE Full text] [doi: 10.1016/j.ijpsycho.2019.04.011] [Medline: 31067489]
26. Kueper JK, Speechley M, Montero-Odasso M. The Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog): modifications and responsiveness in pre-dementia populations. A narrative review. J Alzheimers Dis 2018;63(2):423-444 [FREE Full text] [doi: 10.3233/JAD-170991] [Medline: 29660938]
27. Verma N, Beretvas SN, Pascual B, Masdeu JC, Markey MK, Alzheimer’s Disease Neuroimaging Initiative. New scoring methodology improves the sensitivity of the Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog) in clinical trials. Alzheimers Res Ther 2015;7(1):64 [FREE Full text] [doi: 10.1186/s13195-015-0151-0] [Medline: 26560146]
28. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
29. Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. Br J Psychiatry 1982;140:566-572. [doi: 10.1192/bjp.140.6.566] [Medline: 7104545]

XSL•FO
RenderX

30. Tahami Monfared AA, Houghton K, Zhang Q, Mauskopf J, Alzheimer's Disease Neuroimaging Initiative. Staging disease severity using the Alzheimer's disease composite score (ADCOMS): a retrospective data analysis. Neurol Ther 2022;11(1):413-434 [FREE Full text] [doi: 10.1007/s40120-022-00326-y] [Medline: 35099758]
31. Knopman DS, Knapp MJ, Gracon SI, Davis CS. The Clinician Interview-Based Impression (CIBI): a clinician's global change rating scale in Alzheimer's disease. Neurology 1994;44(12):2315-2321. [doi: 10.1212/wnl.44.12.2315] [Medline: 7991118]
32. Pedrosa H, De Sa A, Guerreiro M, Maroco J, Simoes MR, Galasko D, et al. Functional evaluation distinguishes MCI patients from healthy elderly people--the ADCS/MCI/ADL scale. J Nutr Health Aging 2010;14(8):703-709 [FREE Full text] [doi: 10.1007/s12603-010-0102-1] [Medline: 20922349]
33. Devlin N, Parkin D, Janssen B. Methods for Analysing and Reporting EQ-5D Data. Cham (CH): Springer; 2020.
34. Junqueira DR, Zorzela L, Golder S, Loke Y, Gagnier JJ, Julious SA, CONSORT Harms Group. CONSORT Harms 2022 statement, explanation, and elaboration: updated guideline for the reporting of harms in randomised trials. BMJ 2023;381:e073725 [FREE Full text] [doi: 10.1136/bmj-2022-073725] [Medline: 37094878]
35. Kahan BC, White IR, Edwards M, Harhay MO. Using modified intention-to-treat as a principal stratum estimator for failure to initiate treatment. Clin Trials 2023;20(3):269-275 [FREE Full text] [doi: 10.1177/17407745231160074] [Medline: 36916466]
36. Abraha I, Montedori A. Modified intention to treat reporting in randomised controlled trials: systematic review. BMJ 2010;340:c2697 [FREE Full text] [doi: 10.1136/bmj.c2697] [Medline: 20547685]
37. Gupta SK. Intention-to-treat concept: a review. Perspect Clin Res 2011;2(3):109-112 [FREE Full text] [doi: 10.4103/2229-3485.83221] [Medline: 21897887]
38. Floden L, Bell ML. Imputation strategies when a continuous outcome is to be dichotomized for responder analysis: a simulation study. BMC Med Res Methodol 2019;19(1):161 [FREE Full text] [doi: 10.1186/s12874-019-0793-x] [Medline: 31345166]
39. McInnes IB, Asahina A, Coates LC, Landewé R, Merola JF, Ritchlin CT, et al. Bimekizumab in patients with psoriatic arthritis, naive to biologic treatment: a randomised, double-blind, placebo-controlled, phase 3 trial (BE OPTIMAL). Lancet 2023;401(10370):25-37 [FREE Full text] [doi: 10.1016/S0140-6736(22)02302-9] [Medline: 36493791]
40. Rockwood K, Howlett SE, Hoffman D, Schindler R, Mitnitski A. Clinical meaningfulness of Alzheimer's Disease Assessment Scale-Cognitive Subscale change in relation to goal attainment in patients on cholinesterase inhibitors. Alzheimers Dement 2017;13(10):1098-1106. [doi: 10.1016/j.jalz.2017.02.005] [Medline: 28341540]
41. Hill NTM, Mowszowski L, Naismith SL, Chadwick VL, Valenzuela M, Lampit A. Computerized cognitive training in older adults with mild cognitive impairment or dementia: a systematic review and meta-analysis. Am J Psychiatry 2017;174(4):329-340. [doi: 10.1176/appi.ajp.2016.16030360] [Medline: 27838936]
42. Doody RS, Ferris SH, Salloway S, Sun Y, Goldman R, Watkins WE, et al. Donepezil treatment of patients with MCI: a 48-week randomized, placebo-controlled trial. Neurology 2009;72(18):1555-1561. [doi: 10.1212/01.wnl.0000344650.95823.03] [Medline: 19176895]
43. Corbo I, Marselli G, Di Ciero V, Casagrande M. The protective role of cognitive reserve in mild cognitive impairment: a systematic review. J Clin Med 2023;12(5):1759 [FREE Full text] [doi: 10.3390/jcm12051759] [Medline: 36902545]
44. Ries ML, Carlsson CM, Rowley HA, Sager MA, Gleason CE, Asthana S, et al. Magnetic resonance imaging characterization of brain structure and function in mild cognitive impairment: a review. J Am Geriatr Soc 2008;56(5):920-934 [FREE Full text] [doi: 10.1111/j.1532-5415.2008.01684.x] [Medline: 18410325]
45. Abraham A, Bubic A. Semantic memory as the root of imagination. Front Psychol 2015;6:325 [FREE Full text] [doi: 10.3389/fpsyg.2015.00325] [Medline: 25852626]
46. Jokel R, Lima BS, Fernandez A, Murphy KJ. Language in amnestic mild cognitive impairment and dementia of Alzheimer's type: quantitatively or qualitatively different? Dementia and geriatric cognitive disorders extra. Dement Geriatr Cogn Disord Extra 2019;9(1):136-151 [FREE Full text]
47. Geng F, Xu W, Riggins T. Interactions between the hippocampus and fronto-parietal regions during memory encoding in early childhood. Hippocampus 2022;32(2):108-120. [doi: 10.1002/hipo.23380] [Medline: 34329507]
48. Buckner RL, Kelley WM, Petersen SE. Frontal cortex contributes to human memory formation. Nat Neurosci 1999;2(4):311-314. [doi: 10.1038/7221] [Medline: 10204536]
49. Chun MM, Turk-Browne NB. Interactions between attention and memory. Curr Opin Neurobiol 2007;17(2):177-184. [doi: 10.1016/j.conb.2007.03.005] [Medline: 17379501]
50. Kim H. Attention-versus significance-driven memory formation: taxonomy, neural substrates, and meta-analyses. Neurosci Biobehav Rev 2022;138:104685. [doi: 10.1016/j.neubiorev.2022.104685] [Medline: 35526692]
51. Das A, Menon V. Hippocampal-parietal cortex causal directed connectivity during human episodic memory formation: replication across three experiments. bioRxiv Preprint posted online on March 21, 2024 [FREE Full text] [doi: 10.1101/2023.11.07.566056] [Medline: 37986855]
52. Spens E, Burgess N. A generative model of memory construction and consolidation. Nat Hum Behav 2024;8(3):526-543 [FREE Full text] [doi: 10.1038/s41562-023-01799-z] [Medline: 38242925]

XSL•FO
RenderX

53. Huang Q, Xiao Z, Yu Q, Luo Y, Xu J, Qu Y, et al. Replay-triggered brain-wide activation in humans. Nat Commun 2024;15(1):7185 [FREE Full text] [doi: 10.1038/s41467-024-51582-5] [Medline: 39169063]
54. Antony JW, Romero A, Vierra AH, Luenser RS, Hawkins RD, Bennion KA. Semantic relatedness retroactively boosts memory and promotes memory interdependence across episodes. Elife 2022;11:e72519 [FREE Full text] [doi: 10.7554/eLife.72519] [Medline: 35704025]
55. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5). Washington, DC: American Psychiatric Association; 2013.
56. Schneider LS, Goldberg TE. Composite cognitive and functional measures for early stage Alzheimer's disease trials. Alzheimers Dement (Amst) 2020;12(1):e12017 [FREE Full text] [doi: 10.1002/dad2.12017] [Medline: 32432155]
57. Gopalakrishnan P, Tiwari S, Nagaraja R, Krishnan G. Quality of life in persons with mild cognitive impairment: a systematic review and meta-analysis. Dement Neuropsychol 2024;18:e20230093 [FREE Full text] [doi: 10.1590/1980-5764-DN-2023-0093] [Medline: 39193465]
58. Budd Haeberlein S, Aisen PS, Barkhof F, Chalkias S, Chen T, Cohen S, et al. Two randomized phase 3 studies of aducanumab in early alzheimer's disease. J Prev Alzheimers Dis 2022;9(2):197-210. [doi: 10.14283/jpad.2022.30] [Medline: 35542991]
59. Hudes R, Rich JB, Troyer AK, Yusupov I, Vandermorris S. The impact of memory-strategy training interventions on participant-reported outcomes in healthy older adults: A systematic review and meta-analysis. Psychol Aging 2019;34(4):587-597. [doi: 10.1037/pag0000340] [Medline: 30896195]
60. Rapp S, Brenes G, Marsh AP. Memory enhancement training for older adults with mild cognitive impairment: a preliminary study. Aging Ment Health 2002;6(1):5-11. [doi: 10.1080/13607860120101077] [Medline: 11827617]
61. Belleville S, Cuesta M, Bier N, Brodeur C, Gauthier S, Gilbert B, et al. Five-year effects of cognitive training in individuals with mild cognitive impairment. Alzheimers Dement (Amst) 2024;16(3):e12626. [doi: 10.1002/dad2.12626] [Medline: 39246830]
62. Yang L. Maintained and delayed benefits of executive function training and low-intensity aerobic exercise over a 3.5-year period in older adults. Front Aging Neurosci 2022;14:905886 [FREE Full text] [doi: 10.3389/fnagi.2022.905886] [Medline: 35847677]
63. Kinsella GJ, Ames D, Storey E, Ong B, Pike KE, Saling MM, et al. Strategies for improving memory: a randomized trial of memory groups for older people, including those with mild cognitive impairment. J Alzheimers Dis 2016;49(1):31-43. [doi: 10.3233/JAD-150378] [Medline: 26444773]
64. Winocur G, Moscovitch M, Sekeres M. Memory consolidation or transformation: context manipulation and hippocampal representations of memory. Nat Neurosci 2007;10(5):555-557. [doi: 10.1038/nn1880] [Medline: 17396121]
65. Quillfeldt JA. Temporal flexibility of systems consolidation and the Synaptic Occupancy/Reset Theory (SORT): cues about the nature of the engram. Front Synaptic Neurosci 2019;11:1 [FREE Full text] [doi: 10.3389/fnsyn.2019.00001] [Medline: 30814946]
66. Balsis S, Benge JF, Lowe DA, Geraci L, Doody RS. How do scores on the ADAS-Cog, MMSE, and CDR-SOB correspond? Clin Neuropsychol 2015;29(7):1002-1009. [doi: 10.1080/13854046.2015.1119312] [Medline: 26617181]
67. Wessels AM, Dowsett SA, Sims JR. Detecting treatment group differences in Alzheimer's disease clinical trials: a comparison of Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog) and the Clinical Dementia Rating-Sum of Boxes (CDR-SB). J Prev Alzheimers Dis 2018;5(1):15-20 [FREE Full text] [doi: 10.14283/jpad.2018.2] [Medline: 29405227]
68. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
69. Iqhrammullah M, Yudhistira Refin R, Fitria Andika F, Amirah S, Fahd Abdurrahman M, Alina M, et al. Dropout rate in clinical trials of smartphone apps for diabetes management: a meta-analysis. Diabetes Res Clin Pract 2024;212:111723. [doi: 10.1016/j.diabres.2024.111723] [Medline: 38830484]
70. 2023 survey on broadcasting media usage behavior. Korea Communications Commission. 2023. URL: https://www.kcc.go.kr/user.do?mode=view&page=A05030000&boardId=1113&boardSeq=58975 [accessed 2025-05-12]

Abbreviations

ADAS-Cog: Alzheimer’s Disease Assessment Scale-Cognitive Subscale
ADCOMS: Alzheimer’s Disease Composite Score
ADCS-MCI-ADL: Alzheimer’s Disease Cooperative Study-Mild Cognitive Impairment-Activities of Daily Living
ADL: activities of daily living
AE: adverse event
CDR-SB: Clinical Dementia Rating-Sum of Boxes
CERAD-NP: Consortium to Establish a Registry for Alzheimer’s Disease Neuropsychological Assessment Battery
CIBIC-Plus: Clinician Interview–Based Impression of Change Plus Caregiver Input
CIBIS: Clinician Interview–Based Impression of Severity
CONSORT: Consolidated Standards of Reporting Trials

XSL•FO
RenderX

DSC: Digit Symbol Coding
EQ-5D-VAS: EQ-5D-Visual Analogue Scale
IRB: institutional review board
K-MMSE-II: Korean Mini Mental State Examination, 2nd Edition
MCI: mild cognitive impairment
MMSE: Mini Mental State Examination
MMT: metamemory training
NRI: nonresponder imputation
SAE: serious adverse event
SNSB: Seoul Neuropsychological Screening Battery

Edited by L Buis; submitted 24.Mar.2025; peer-reviewed by KW Turk, K Yuenyongchaiwat; comments to author 01.May.2025; revised version received 19.Nov.2025; accepted 20.Nov.2025; published 19.Jan.2026.

Please cite as:
Kang S, Lim JI, Stenzel L, Kim KY, Kim E, Jeon HJ, Park DH, Lim HK, Shim Y, Jang JW, Kim Y, Lee S, Park KH
Efficacy and Safety of Mobile App–Based Metamemory Cognitive Training for Mild Cognitive Impairment: Multicenter Randomized Clinical Trial
JMIR Mhealth Uhealth 2026;14:e73464
URL: https://mhealth.jmir.org/2026/1/e73464
doi:10.2196/73464
PMID:

©Sunyoung Kang, Jung-In Lim, Lukas Stenzel, Keun You Kim, Eosu Kim, Hong Jun Jeon, Doo-Heum Park, Hyun Kook Lim, YongSoo Shim, Jae-Won Jang, Yeshin Kim, Sanghoon Lee, Kee Hyung Park. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 19.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Development and Validation of a Novel Lymphedema Monitoring System Using a Bodysuit and a Smartphone: Prospective Comparative Study

Ryo Karakawa[1], MBA, MD, PhD; Hidehiko Yoshimatsu[2], MD, PhD; Katsu Karakawa[3], MD, PhD; Kuniko Utsugi[1], MD; Kengo Ono[4], BA; Eri Kojima[4], BA; Takayasu Yamada[4], BA; Duncan Thomas Eason[5], PhD; Francois Jean Leon Goffinet[5], PhD; Tomoyuki Yano[1], MD, PhD

[1]Cancer Institute Hospital of the Japanese Foundation for Cancer Research, Tokyo, Japan

[2]Department of Plastic and Reconstructive Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, United States

[3]Kawabata Lymphedema Clinic, Kobe, Japan

[4]ZOZO, Inc, Chiba, Japan

[5]ZOZO New Zealand Limited, Auckland, New Zealand

Corresponding Author:
Ryo Karakawa, MBA, MD, PhD
Cancer Institute Hospital of the Japanese Foundation for Cancer Research
3-8-31, Ariake, Koto-ku
Tokyo, 135-8550
Japan
Phone: 81 335200111
Email: ryo.kyara@gmail.com

Abstract

Background: Accurate limb circumference measurement is essential for the diagnosis and management of lymphedema, facilitating the objective evaluation of treatment outcomes and early detection of disease progression. Although manual tape measurements are inexpensive and widely used, they are limited by interobserver variability and low reproducibility. Advanced modalities such as computed tomography and magnetic resonance imaging provide high precision but are costly and impractical for routine or home-based use. To overcome these limitations, we developed and validated a novel 3D measurement system that integrates a marker-based bodysuit (ZOZOSUIT2) with a smartphone app, enabling fast, accurate, and reproducible limb circumference assessment.

Objective: This study aimed to evaluate the accuracy, repeatability, and time efficiency of the 3D measurement bodysuit and smartphone system compared with those of manual tape measurements in patients with upper and lower limb lymphedema.

Methods: This prospective study included 26 patients (n=10 with upper limb lymphedema and n=16 with lower limb lymphedema). Using the ZOZOSUIT2, which contains approximately 20,000 dot markers, 12 full-body photographs were captured with a smartphone, and a 3D model was automatically generated. Circumference values were extracted at 10 standardized points on the upper limb (from the wrist to the top of the upper arm) and 9 on the lower limb (from the ankle to the base of the thigh). Each measurement was repeated 3 times to assess repeatability. Manual tape measurements were used as the reference standard. For each measurement point, the repeatability (intrameasurement variation) and the mean absolute error between the 2 methods were calculated. Measurement time was also recorded for both methods and compared statistically using the Wilcoxon signed-rank test.

Results: Across all measurement points, the median repeatability of the bodysuit smartphone system ranged from 2.2 to 6.4 mm, indicating high reproducibility. When compared with manual measurements, the median of the mean absolute error for upper limb points ranged from 6.7 to 20.5 mm, and for lower limb points ranged from 5.5 to 15.9 mm. Relatively larger discrepancies were observed at the wrist (median 19.8, IQR 15.5-24.1 mm), 5 cm proximal to the wrist (median 20.5, IQR 8.5-28.1 mm), and the base of the thigh (median 15.9, IQR 10.2-25.9 mm). Mean measurement time was significantly shorter with the bodysuit smartphone system (88.2, SD 16.0 seconds) than with manual tape measurement (293.8, SD 58.5 seconds; $P<.001$).

Conclusions: The bodysuit smartphone system enables rapid, precise, and highly reproducible limb circumference assessment in patients with lymphedema. Despite minor differences at anatomical sites with complex contours, such as the wrist and the

XSL•FO
RenderX

thigh base, the overall accuracy and time efficiency were clinically acceptable. This system may serve as a practical and scalable solution for both clinical and home-based lymphedema monitoring, contributing to the objective and standardized assessment of limb volume.

(JMIR Mhealth Uhealth 2026;14:e77935) doi:10.2196/77935

KEYWORDS

measurements; limbs; smartphone; bodysuit; lymphedema; patients

Introduction

Lymphedema is a chronic condition characterized by the abnormal accumulation of lymphatic fluid in the interstitial tissues, often resulting in swelling, discomfort, and functional impairments [1]. Accurate measurement of limb circumference is critical for both the diagnosis and monitoring of lymphedema, as well as for evaluating the effectiveness of conservative therapies or surgical interventions [2]. Early detection through precise measurement is particularly important to mitigate disease progression and optimize therapeutic outcomes [3].

Various methods for limb circumference measurement have been developed, each with distinct advantages and limitations. Manual circumference measurement using a tape measure is widely used because of its simplicity and low cost. However, it is inherently prone to interobserver and intraobserver variability and provides limited data points, restricting its utility in comprehensive assessment [4]. Advanced imaging techniques such as computed tomography and magnetic resonance imaging offer high accuracy and detailed data; however, their high cost, limited accessibility, and invasive nature make them unsuitable for routine or home-based monitoring [5,6]. These limitations underscore the need for a measurement method that is accurate, noninvasive, cost-effective, and user-friendly.

In recent years, technological advancements have facilitated the use of 3D scanning and smartphone-based tools for health monitoring. These innovations have made it possible to obtain detailed and accurate measurements with minimal invasiveness, providing new opportunities for disease management in both clinical and home settings [7,8]. However, there remains a lack of practical solutions that combine accuracy, user-friendliness, and cost-effectiveness, particularly for conditions such as lymphedema that require frequent and precise monitoring. To address this challenge, we have developed a novel approach using a 3D measurement bodysuit (ZOZOSUIT2; ZOZO Inc) and a smartphone [9]. With a standard smartphone camera, participants could generate a 3D virtual representation of their bodies in a few minutes. The primary objective of this study was to validate the accuracy and reliability of this newly developed method for measuring limb circumference in patients with lymphedema.

Methods

Study Design and Participants

This was a prospective, multicenter observational study conducted to validate the accuracy and reliability of limb circumference measurements using a 3D measurement bodysuit (ZOZOSUIT2) and a smartphone. Consecutive patients diagnosed with upper or lower limb lymphedema who visited the Cancer Institute Hospital between November 2021 and February 2023 were screened for eligibility. Patients were identified from the outpatient clinic records and were approached in person during their regular visits. Participants who met the inclusion criteria and provided written informed consent were enrolled. Those who declined participation or were unable to wear the designated bodysuit due to size incompatibility were excluded.

Ethical Considerations

This study was reviewed and approved by the institutional review board of the Japanese Foundation for Cancer Research (2021-GA-1050) on April 5, 2022. All participants provided written informed consent before enrollment after receiving a full explanation of the study objectives and procedures. All collected data were anonymized before analysis to protect participant privacy and confidentiality. No financial or other compensation was provided to participants for their involvement in the study. This study was conducted and reported in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines [10].

Data Collection and Manual Measurement

Patient characteristics and clinical data were obtained, including age, sex, BMI, and clinical stages based on the International Society of Lymphology (ISL) classification. Manual circumference measurements were performed using a standard measuring tape for comparison. For the upper limbs, circumferences were measured at the following 10 predefined points: top of the upper arm, 10 cm proximal to the elbow, midpoint of the upper arm, 5 cm proximal to the elbow, the elbow, 5 cm distal to the elbow, maximum point of the forearm, 10 cm proximal to the wrist, 5 cm proximal to the wrist, and the wrist. For the lower limbs, circumferences were measured at the following 9 predefined points: base of the thigh, 20 cm proximal to the knee, 10 cm proximal to the knee, the knee, just below the knee, 5 cm distal to the knee, maximum point of the lower leg, 10 cm distal to the knee, and the ankle.

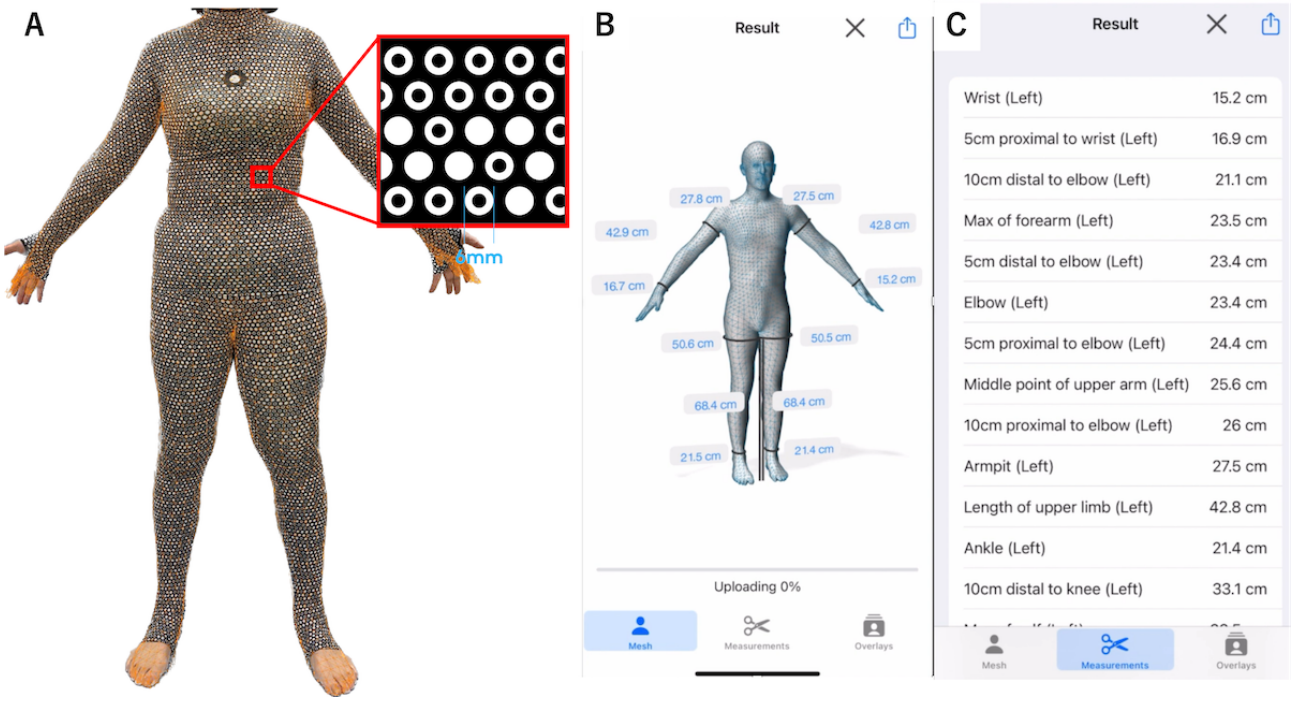
To ensure consistency between methods, manual tape measurements were performed in a standing position, matching the posture used for the bodysuit smartphone measurements. In addition to circumference measurements, the time required to complete manual measurements for both upper and lower limbs was recorded. Measurement time was defined as the total duration from placement of the measuring tape at the first predefined point to completion of the final measurement.

XSL•FO
RenderX

Measurement System and Procedure

The ZOZOSUIT2 is a bodysuit embedded with approximately 20,000 dot markers, enabling high-accuracy 3D measurement through a smartphone camera (Figure 1). Participants wore the ZOZOSUIT2, and 12 photographs were taken from different angles using a smartphone (iPhone 14; Apple Inc) according to the manufacturer's protocol.

Figure 1. Workflow of the 3D measurement system using the ZOZOSUIT2 and smartphone app. (A) The ZOZOSUIT2 is a bodysuit embedded with approximately 20,000 dot markers, enabling high-accuracy 3D measurement through a smartphone camera. (B) The application analyzes the 12 images to extract the 2D position of each visible marker and the participant's silhouette, and the 3D virtual human body model is displayed in the app. (C) The images were then processed by proprietary software to generate a 3D model of the participant's body, from which limb circumferences were automatically extracted at multiple points.



The participants were instructed to adopt a posture with their arms and lower limbs slightly spread out. The smartphone running the fitting app was placed on a phone stand at a height of 70 cm and positioned 1.5 m away from the participant. The app used audio prompts to guide the participant. Participants were instructed to turn clockwise by 30° every 6 seconds. Participants actively rotated their own body position without the use of a revolving platform. The app automatically captured a new image at the end of each interval. At the end of the whole process, the app had acquired 12 images of the participant. The app analyzed the 12 images to extract the 2D position of each visible marker and the participant's silhouette. Finally, a 3D virtual human body model was displayed in the app. These images were processed by proprietary software to generate a 3D model of the participant's body, from which limb circumferences were automatically extracted at multiple points (Figure 1). The measurement time was recorded.

Statistical Analyses

To assess the reproducibility of the bodysuit smartphone system, 3 consecutive measurements were performed for each participant under identical conditions. Repeatability for each measurement point was defined as the mean of the absolute differences between paired measurements (scan 1 vs scan 2, scan 2 vs scan 3, and scan 1 vs scan 3) across all participants. The limb circumference measurements obtained with the ZOZOSUIT2 and smartphone were compared to those obtained through manual measurements using a tape measure. The mean absolute error for each measurement point was calculated to evaluate accuracy.

Measurement times were recorded for both the bodysuit smartphone method and manual measurements. The start time was defined as the point at which the measurement process began (putting on the bodysuit or preparing the tape measure), and the end time was when all measurements were completed. Statistical comparisons of measurement times were performed between the 2 methods.

All statistical analyses were conducted using Google Sheets (Google LLC). Built-in spreadsheet functions were used to calculate descriptive statistics (mean, SD), and paired *t* tests (2-tailed) were performed to compare measurement times between the bodysuit smartphone and manual methods. Statistical significance was set at $P<.05$. The data were cross-checked using Microsoft Excel (version 16.106.2; Microsoft Corp) to verify calculation consistency.

Results

Patient Characteristics

Our study included 26 patients, consisting of 10 with upper limb lymphedema and 16 with lower limb lymphedema. We performed a total of 78 measurements using the bodysuit smartphone method. Table 1 shows the demographics of patients with upper limb lymphedema. The mean age was 56.1 (SD 15.2;

XSL•FO
RenderX

range 30-79) years. All patients were female. The mean BMI was 22.6 (SD 2.3; range 17.9-25.9) kg/m^2. The causative disease of lymphedema was breast cancer in all patients. Of 10 patients, lymphedema severity was stage I in 3 (30%), stage IIa in 2 (20%), and stage IIb in 5 (50%), based on the ISL classification.

Table 1. Demographics of patients with upper limb lymphedema (N=10).

Parameter	Value
Age (years), average (SD; range)	56.1 (15.2; 30-79)
Sex, n (%)	
Female	10 (100)
Physical parameters, average (SD; range)	
Height (cm)	155.2 (5.8; 146-163)
Weight (kg)	54.8 (8.4; 43.0-68.8)
BMI (kg/m^2)	22.6 (2.3; 17.9-25.9)
Cause of disease, n (%)	
Breast cancer	10 (100)
International Society of Lymphology stage, n (%)	
I	3 (30)
IIa	2 (20)
IIb	5 (50)

Table 2 shows the demographics of patients with lower limb lymphedema. The mean age was 60.9 (SD 11.2; range 43-76) years. All patients were female. The mean BMI was 21.8 (SD 2.8; range 16.5-28.1) kg/m^2. Of 16 patients, the causative disease of lymphedema was cervical cancer in 9 (56%), uterine cancer in 5 (31%), and ovarian cancer in 2 (13%). Lymphedema severity was stage I in 2 (13%), stage IIa in 5 (31%), and stage IIb in 9 (56%), based on the ISL classification.

Table 2. Demographics of patients with lower limb lymphedema (N=16).

Parameter	Value
Age (years), average (SD: range)	60.9 (11.2; 43-76)
Sex, n (%)	
Female	10 (100)
Physical parameters, average (SD: range)	
Height (cm)	157.8 (5.4; 148-167)
Weight (kg)	54.2 (6.4; 45.6-65.5)
BMI (kg/m^2)	21.8 (2.8; 16.5-28.1)
Cause of disease, n (%)	
Cervical cancer	9 (56)
Uterine cancer	5 (31)
Ovarian cancer	2 (13)
International Society of Lymphology stage, n (%)	
I	2 (13)
IIa	5 (31)
IIb	9 (56)

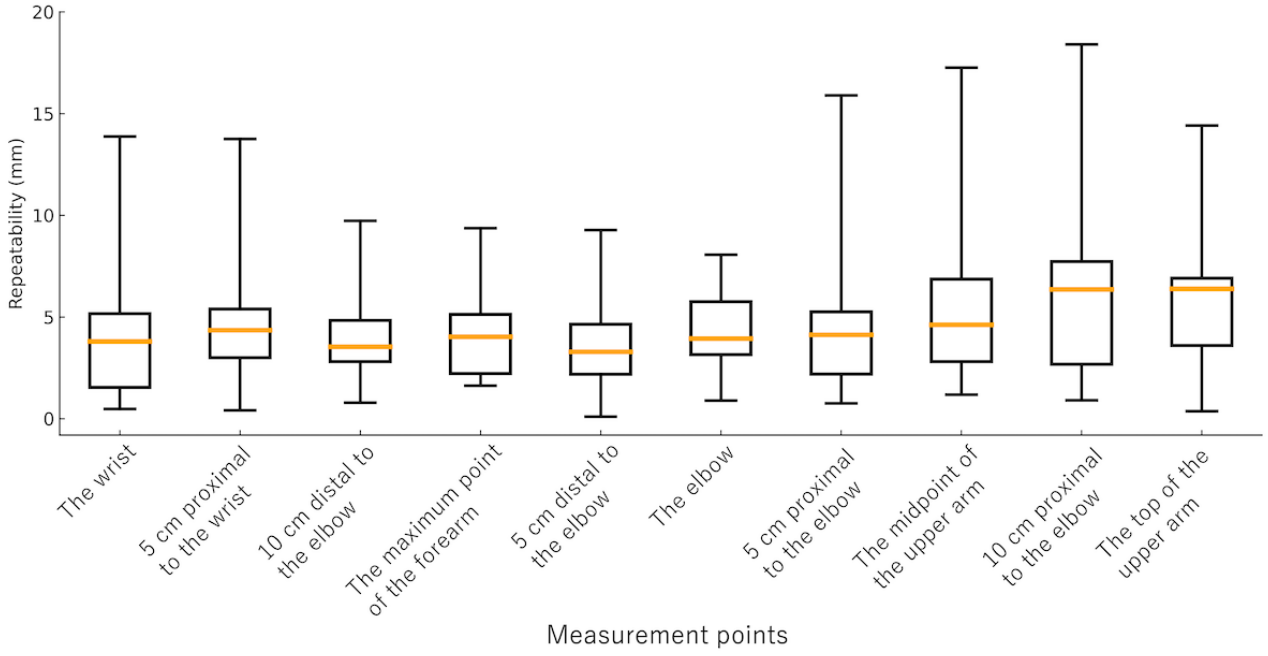
Repeatability of Limb Circumference Measurements

The median repeatability of limb circumference measurements obtained using the bodysuit smartphone system ranged from 2.2 to 6.4 mm across all measurement points for both the upper and lower limbs. Specifically, the median repeatability values for the upper limbs were as follows: the wrist, 3.8 (IQR 1.5-5.2) mm; 5 cm proximal to the wrist, 4.4 (IQR 3.0-5.4) mm; 10 cm distal to the elbow, 3.5 (IQR 2.8-4.8) mm; the maximum point of the forearm, 4.0 (IQR 2.2-5.1) mm; 5 cm distal to the elbow, 3.3 (IQR 2.2-4.6) mm; the elbow, 3.9 (IQR 3.2-5.8) mm; 5 cm

XSL•FO
RenderX

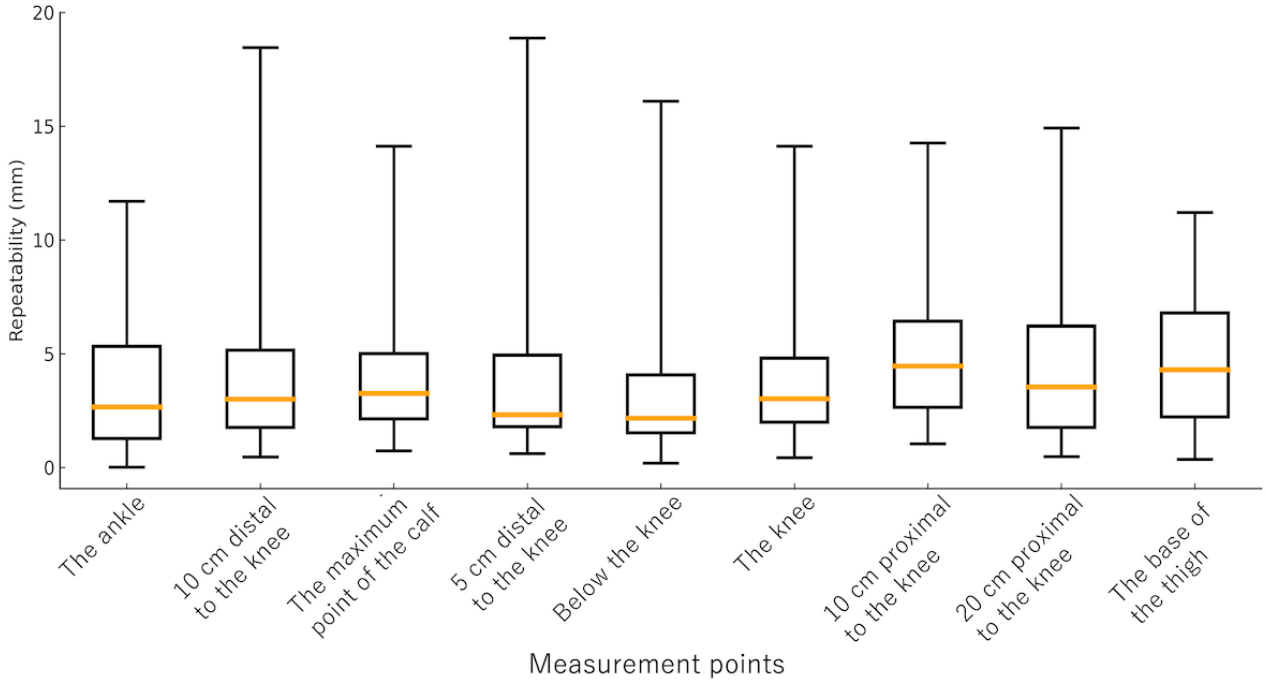
proximal to the elbow, 4.1 (IQR 2.2-5.3) mm; the midpoint of the upper arm, 4.6 (IQR 2.8-6.9) mm; 10 cm proximal to the elbow, 6.4 (IQR 2.7-7.7) mm; and the top of the upper arm, 6.4 (IQR 3.6-6.9) mm (Figure 2). For the lower limbs, the median values were as follows: the ankle, 2.7 (IQR 1.3-5.3) mm; 10 cm distal to the knee, 3.0 (IQR 1.8-5.2) mm; the maximum point of the lower leg, 3.3 (IQR 2.1-5.0) mm; 5 cm distal to the knee, 2.3 (IQR 1.8-4.9) mm; just below the knee, 2.2 (IQR 1.5-4.1) mm; the knee, 3.0 (IQR 2.0-4.8) mm; 10 cm proximal to the knee, 4.5 (IQR 2.6-6.4) mm; 20 cm proximal to the knee, 3.5 (IQR 1.8-6.2) mm; and the base of the thigh, 4.3 (IQR 2.2-6.8) mm (Figure 3). All repeatability values were below 10 mm, indicating a high level of reproducibility and confirming the system's reliability for both upper and lower limb measurements.

Figure 2. Repeatability of measurements at each upper limb point. Box-and-whisker plots show the repeatability of circumference measurements at 10 predefined upper limb points obtained using the 3D measurement bodysuit smartphone system. Each box represents the IQR, the orange line indicates the median, and the whiskers show the full range. Measurement points include the wrist, 5 cm proximal to the wrist, 10 cm distal to the elbow, the maximum point of the forearm, 5 cm distal to the elbow, the elbow, 5 cm proximal to the elbow, the midpoint of the upper arm, 10 cm proximal to the elbow, and the top of the upper arm.



XSL•FO
RenderX

Figure 3. Repeatability of measurements at each lower limb point. Box-and-whisker plots show the repeatability of circumference measurements at 9 predefined lower limb points obtained using the 3D measurement bodysuit smartphone system. Each box represents the IQR, the orange line indicates the median, and the whiskers show the full range. Measurement points include the ankle, 10 cm distal to the knee, the maximum point of the calf, 5 cm distal to the knee, just below the knee, the knee, 10 cm proximal to the knee, 20 cm proximal to the knee, and the base of the thigh.

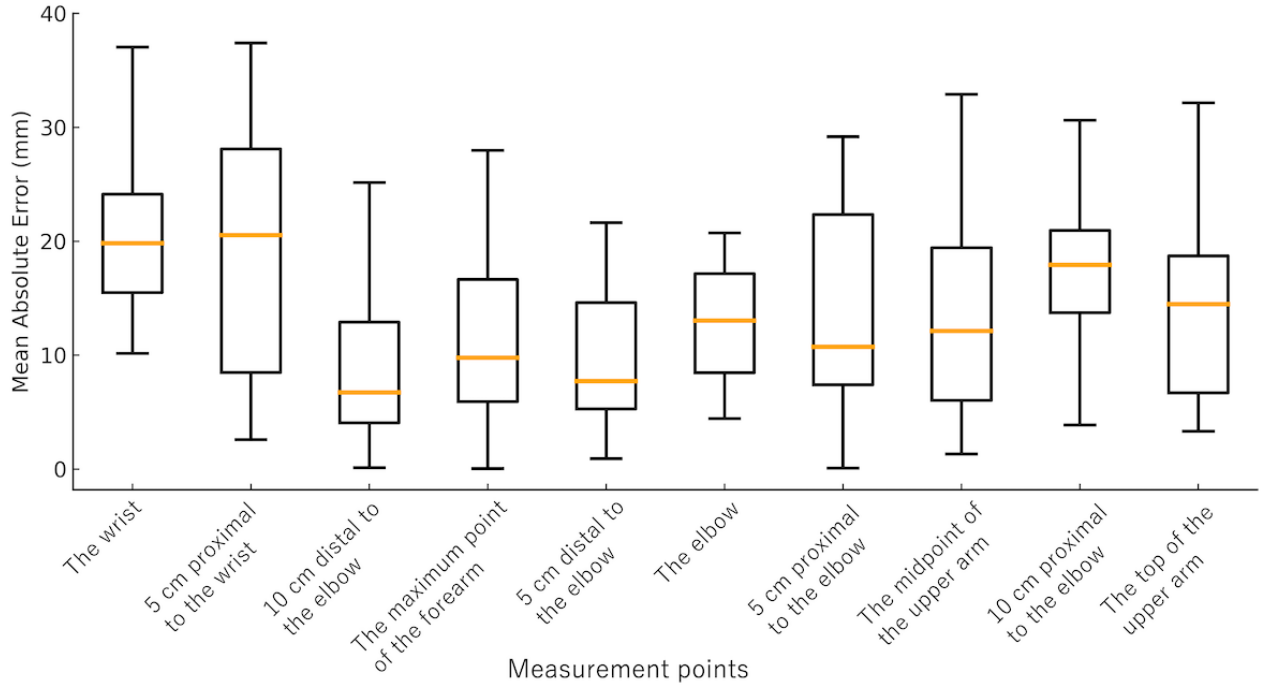


Comparison Between the Bodysuit Smartphone Method and Manual Measurements

The mean absolute error between manual measurements and those obtained using the bodysuit smartphone system for the upper limb is shown in Figure 4. The median of the mean absolute error values across all upper limb measurement points ranged from 6.7 to 20.5 mm. Specifically, the median of the mean absolute error values were as follows: the wrist, 19.8 (IQR 15.5-24.1) mm; 5 cm proximal to the wrist, 20.5 (IQR 8.5-28.1) mm; 10 cm distal to the elbow, 6.7 (IQR 4.1-12.9) mm; the maximum point of the forearm, 9.8 (IQR 5.9-16.7) mm; 5 cm distal to the elbow, 7.7 (IQR 5.3-14.6) mm; the elbow, 13.0 (IQR 8.5-17.2) mm; 5 cm proximal to the elbow, 10.8 (IQR 7.4-22.4) mm; the midpoint of the upper arm, 12.1 (IQR 6.0-19.4) mm; 10 cm proximal to the elbow, 17.9 (IQR 13.7-21.0) mm; and the top of the upper arm, 14.5 (IQR 6.7-18.7) mm.

XSL•FO
RenderX

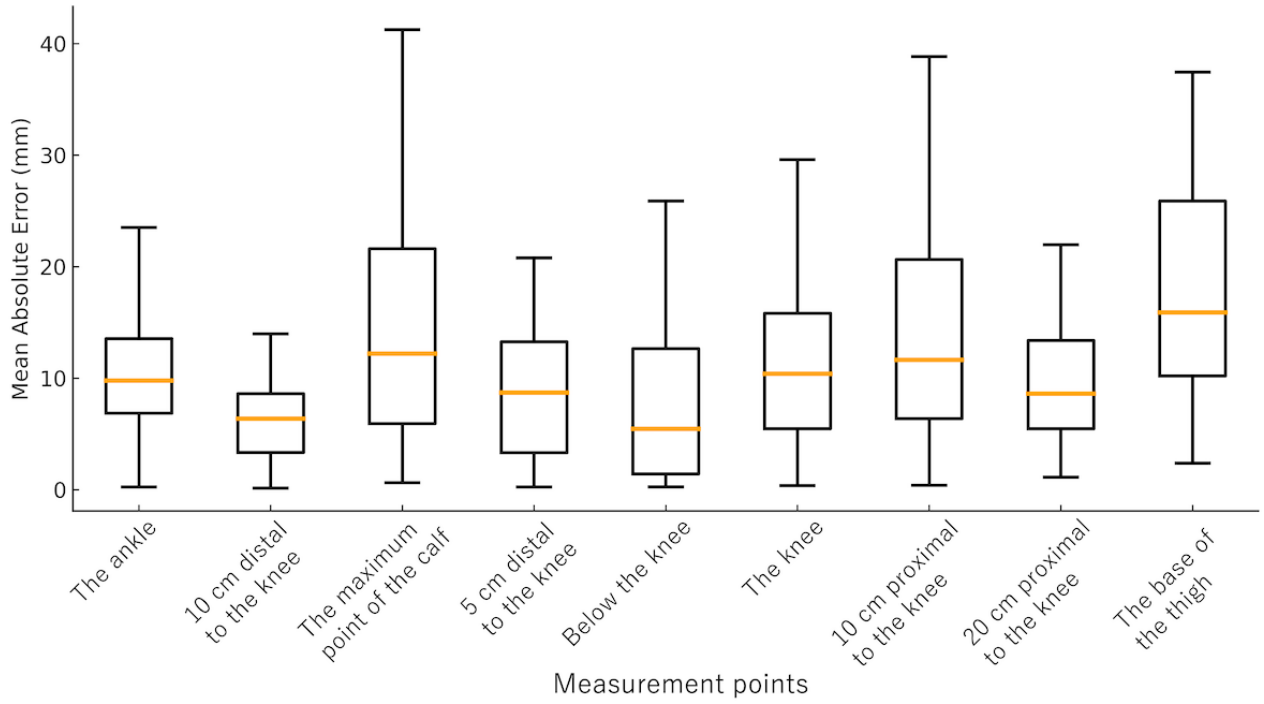
Figure 4. Mean absolute error between manual and bodysuit smartphone measurements at each upper limb point. Box-and-whisker plots illustrate the mean absolute error between manual and digital measurements across 10 measurement points on the upper limb (the wrist, 5 cm proximal to the wrist, 10 cm distal to the elbow, the maximum point of the forearm, 5 cm distal to the elbow, the elbow, 5 cm proximal to the elbow, the midpoint of the upper arm, 10 cm proximal to the elbow, and the top of the upper arm). The boxes represent the IQR, the central line indicates the median, and the whiskers denote the range excluding outliers. The overall median of the mean absolute error ranged from 6.7 to 20.5 mm, with higher variability observed in the proximal regions.



The mean absolute error between manual measurements and those obtained using the bodysuit smartphone system for the lower limb is shown in Figure 5. The median of the mean absolute error values ranged from 5.5 to 15.9 mm across all measurement points. Specifically, the median of the mean absolute error values were as follows: the ankle, 9.8 (IQR 6.9-13.6) mm; 10 cm distal to the knee, 6.4 (IQR 3.3-8.6) mm; the maximum point of the calf, 12.2 (IQR 5.9-21.6) mm; 5 cm distal to the knee, 8.7 (IQR 3.3-13.3) mm; just below the knee, 5.5 (IQR 1.4-12.7) mm; the knee, 10.4 (IQR 5.5-15.8) mm; 10 cm proximal to the knee, 11.6 (IQR 6.4-20.7) mm; 20 cm proximal to the knee, 8.6 (IQR 5.5-13.4) mm; and the base of the thigh, 15.9 (IQR 10.2-25.9) mm.

XSL•FO
RenderX

Figure 5. Mean absolute error between manual and bodysuit smartphone measurements at each lower limb point. Box-and-whisker plots illustrate the mean absolute error between manual and digital measurements across 9 measurement points on the lower limb (the ankle, 10 cm distal to the knee, the maximum point of the calf, 5 cm distal to the knee, just below the knee, the knee, 10 cm proximal to the knee, 20 cm proximal to the knee, and the base of the thigh). The boxes indicate the IQR, the central line shows the median, and the whiskers represent the range. The overall median of the mean absolute error ranged from 5.5 to 15.9 mm, with slightly larger deviations observed at the base of the thigh.



Furthermore, compared to the manual measurement group, the bodysuit smartphone method group demonstrated significantly shorter mean measurement times (88.2, SD 16.0 vs 293.8, SD 58.5 seconds; $P<.001$). The variability in measurement time was also considerably lower in the bodysuit smartphone method group than in the manual measurement group, indicating a more consistent measurement process.

Discussion

This study aimed to validate the accuracy and reliability of a novel limb circumference measurement system using a 3D measurement bodysuit and a smartphone compared to the conventional manual measuring tape method. The findings demonstrate that the bodysuit smartphone system is not only highly accurate and reproducible but also significantly more time-efficient than manual methods, indicating its potential as a valuable tool for lymphedema management.

The median of the mean absolute error across all measurement points for the 3D measurement bodysuit smartphone system was less than 10 mm. This level of accuracy is within clinically acceptable thresholds for lymphedema monitoring, where errors of less than 10 mm are generally considered sufficient to detect changes in limb circumference and disease progression [11]. Furthermore, the reproducibility of this method suggests transformative potential for the way limb circumference measurements are made. Unlike traditional manual measurements, which require the assistance of health care professionals or trained individuals, this method enables patients to perform measurements independently at home. This autonomy could significantly reduce the burden on health care systems by minimizing the need for clinic visits, thereby conserving medical resources and streamlining patient care [7]. The implications of these findings highlight the potential of the 3D measurement bodysuit smartphone system to revolutionize lymphedema management through improved accessibility and efficiency.

When comparing the mean absolute error between manual measurements and those obtained using the 3D measurement bodysuit smartphone system, the errors at most measurement points were less than 20 mm. This level of accuracy is considered clinically acceptable and comparable to manual measurements. However, in the upper limb, relatively larger errors were observed at the wrist and 5 cm proximal to the wrist, with median of the mean absolute errors of 19.8 (IQR 15.5-24.1) mm and 20.5 (IQR 8.5-28.1) mm, respectively.

In the lower limb, the base of the thigh also showed a relatively larger deviation, with median of the mean absolute errors of 15.9 (IQR 10.2-25.9) mm. These differences are likely attributable to the difficulty of obtaining consistent manual measurements at these anatomical sites rather than limitations of the bodysuit smartphone system itself. The thigh base includes a large and deformable soft tissue volume, which makes it challenging to define and reproduce an identical measurement level with a tape measure, particularly in the standing position. At the wrist, bony prominences such as the radial and ulnar styloid processes can interfere with stable tape placement. Considering the high repeatability of the bodysuit based method, it may offer greater reproducibility and reduce variability caused by individual measurers, thus contributing to the standardization of limb circumference measurement. These features highlight the potential of this method to provide consistent and reliable

XSL•FO
RenderX

data for both clinical practice and research settings. In addition, posture may also influence limb circumference measurement because gravity affects lymphatic fluid distribution. In our study, both manual and bodysuit smartphone measurements were performed in a standing position to minimize posture-related discrepancies. This standardized posture helped ensure that any differences between the 2 methods were due to measurement technique rather than variation in limb fluid distribution.

In addition to its accuracy, the 3D measurement bodysuit smartphone system demonstrated significantly shorter measurement times than manual methods. The bodysuit based method required an average of 88.2 (SD 16.0) seconds, whereas manual measurements took 293.8 (SD 58.5) seconds on average. This time efficiency has important implications for clinical workflows, particularly in high-volume lymphedema clinics. By reducing the time needed for circumference measurements, health care providers can focus more on other critical aspects of patient care such as treatment planning and counseling. This efficiency underscores the potential of this system to streamline lymphedema management in both clinical and home settings.

Advanced imaging modalities such as magnetic resonance imaging and computed tomography are capable of providing detailed anatomical and volumetric data, making them highly accurate tools for lymphedema assessment. However, their high cost, limited accessibility, and the need for specialized facilities and personnel render them impractical for routine use or home-based monitoring. The 3D measurement bodysuit smartphone system offers a noninvasive, portable, and cost-effective alternative that bridges the gap between these advanced imaging techniques and traditional manual methods. Additionally, devices such as the Perometer (Pero-System GmbH), which use infrared light to measure limb volume, have been established as reliable tools in clinical settings. While the Perometer is less labor intensive and provides high precision, its reliance on bulky, fixed equipment and higher costs limit its applicability for home use or resource-limited environments [12,13]. In contrast, the bodysuit smartphone system leverages widely available smartphone technology, thereby providing greater accessibility and usability. Its portability and ease of use make it particularly suitable for frequent or remote monitoring, empowering patients and reducing the burden on health care providers. This combination of practicality and accuracy positions it as a promising alternative for routine lymphedema management.

The implications of these findings are significant. Early detection and timely intervention are crucial for preventing lymphedema progression, and the 3D measurement bodysuit smartphone system enables frequent and accurate monitoring, empowering patients to take an active role in their care. By facilitating self-monitoring at home, this method reduces the need for frequent clinic visits, thereby alleviating the burden on both patients and health care providers. Its portability makes it a practical option for diverse clinical settings, including those in remote or underserved areas. Furthermore, the potential to integrate this technology into telemedicine platforms enhances its utility, allowing clinicians to remotely monitor patient progress and provide timely interventions when necessary [8]. These findings align with previous reports emphasizing the need for reproducible, patient-centered measurement systems in chronic disease monitoring and further extend this concept to lymphedema care [14].

While this study has several strengths, including its prospective design and the inclusion of a diverse cohort of patients with upper and lower limb lymphedema, it is not without limitations. The relatively small sample size and the fact that the study was conducted at only 2 centers may limit the generalizability of the findings. Larger, multicenter studies are needed to confirm the results and assess the applicability of the 3D measurement bodysuit smartphone system across different populations and clinical settings. Additionally, the greater measurement errors observed at specific points, such as the base of the thigh and the wrist, suggest potential limitations in the system's ability to accurately assess regions with complex anatomical contours or limited marker visibility. Future improvements to the technology should focus on addressing these challenges to further enhance its accuracy and reliability.

In conclusion, the 3D measurement bodysuit smartphone system represents a significant advancement in the field of lymphedema management. Its high accuracy, excellent reproducibility, and time efficiency make it a promising alternative to conventional manual methods. By providing access to accurate and user-friendly measurement tools, this system has the potential to transform lymphedema care by empowering patients to actively monitor their condition and enabling clinicians to deliver more efficient and personalized care. Future research should focus on addressing the system's limitations, validating its use in broader and more diverse populations, and exploring its integration into telemedicine and advanced data analytics platforms.

Acknowledgments

The authors thank all patients who participated in this study and the clinical staff who supported data collection. The authors also thank ZOZO, Inc, for technical support with the measurement device. Generative artificial intelligence tools (ChatGPT; version 5.0; OpenAI) were used only for language refinement and grammar checking after the scientific content was finalized by the authors. The authors reviewed and verified all outputs for accuracy.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request. Due to patient privacy restrictions, data are not publicly available.

XSL•FO
RenderX

Funding

The authors declared no financial support was received for this work.

Authors' Contributions

RK: conceptualization, methodology, validation, formal analysis, investigation, data curation, writing—original draft, writing—review and editing, visualization, and project administration.
HY: writing—review and editing.
KK and KU: resources.
KO: methodology, resources, and project administration.
EK: methodology and resources.
TY: software, resources, and supervision.
DTE and FJLG: software, formal analysis, and data curation.
TY: supervision.

Conflicts of Interest

RK and TY are inventors of the 3D measurement system described in this study, with the patent rights assigned to the Cancer Institute Hospital of the Japanese Foundation for Cancer Research and jointly held with ZOZO, Inc. KO, EK, TY, DTE, and FJLG are employees of ZOZO, Inc. The other authors declare no conflicts of interest.

References

1. Cheville AL, McGarvey CL, Petrek JA, Russo SA, Taylor ME, Thiadens SR. Lymphedema management. Semin Radiat Oncol 2003 Jul;13(3):290-301. [doi: 10.1016/S1053-4296(03)00035-3] [Medline: 12903017]
2. Decorte T, Van Besien V, Van Calster C, Vanden Bossche L, Randon C, Devoogdt N, et al. Addition of prophylactic compression garments to standard care to prevent irreversible lower limb lymphoedema after gynaeco-oncological therapy (Gynolymph): protocol for a randomised controlled trial embedded within an observation cohort study. BMJ Open 2024 Nov 01;14(10):e088851 [FREE Full text] [doi: 10.1136/bmjopen-2024-088851] [Medline: 39486824]
3. Ridner SH. The psycho-social impact of lymphedema. Lymphat Res Biol 2009;7(2):109-112 [FREE Full text] [doi: 10.1089/lrb.2009.0004] [Medline: 19534633]
4. Sander AP, Hajer NM, Hemenway K, Miller AC. Upper-extremity volume measurements in women with lymphedema: a comparison of measurements obtained via water displacement with geometrically determined volume. Phys Ther 2002 Dec;82(12):1201-1212. [Medline: 12444879]
5. Weiss M, Baumeister RG, Frick A, Wallmichrath J, Bartenstein P, Rominger A. Primary lymphedema of the lower limb: the clinical utility of single photon emission computed tomography/CT. Korean J Radiol 2015;16(1):188-195 [FREE Full text] [doi: 10.3348/kjr.2015.16.1.188] [Medline: 25598689]
6. Case TC, Witte CL, Witte MH, Unger EC, Williams WH. Magnetic resonance imaging in human lymphedema: comparison with lymphangioscintigraphy. Magn Reson Imaging 1992;10(4):549-558. [doi: 10.1016/0730-725x(92)90006-l] [Medline: 1501525]
7. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. Nat Biotechnol 2015 May;33(5):462-463. [doi: 10.1038/nbt.3223] [Medline: 25965751]
8. Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 2016 Jun;41(7):1691-1696 [FREE Full text] [doi: 10.1038/npp.2016.7] [Medline: 26818126]
9. Ito Y, Doi T, Ohtomo N, Fujii Y, Ono K, Kojima E, et al. A novel screening method for scoliosis using a bodysuit and 3-dimensional imaging. Spine (Phila Pa 1976) 2023 Sep 15;48(18):1289-1294. [doi: 10.1097/BRS.0000000000004658] [Medline: 37389986]
10. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Int J Surg 2014 Dec;12(12):1495-1499 [FREE Full text] [doi: 10.1016/j.ijsu.2014.07.013] [Medline: 25046131]
11. Cornish BH, Chapman M, Hirst C, Mirolo B, Bunce IH, Ward LC, et al. Early diagnosis of lymphedema using multiple frequency bioimpedance. Lymphology 2001 Mar;34(1):2-11. [Medline: 11307661]
12. Stanton AW, Northfield JW, Holroyd B, Mortimer PS, Levick JR. Validation of an optoelectronic limb volumeter (Perometer). Lymphology 1997 Jun;30(2):77-97. [Medline: 9215977]
13. Mayrovitz HN, Sims N, Macdonald J. Assessment of limb volume by manual and automated methods in patients with limb edema or lymphedema. Adv Skin Wound Care 2000;13(6):272-276. [Medline: 12669673]
14. Pong C, Tseng RM, Tham YC, Lum E. Current implementation of digital health in chronic disease management: scoping review. J Med Internet Res 2024 Dec 12;26:e53576 [FREE Full text] [doi: 10.2196/53576] [Medline: 39666972]

XSL•FO
RenderX

Abbreviations

ISL: International Society of Lymphology
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by L Buis; submitted 22.May.2025; peer-reviewed by G Tartaglione, P Cha; comments to author 07.Oct.2025; revised version received 14.Oct.2025; accepted 12.Feb.2026; published 06.Mar.2026.

Please cite as:
Karakawa R, Yoshimatsu H, Karakawa K, Utsugi K, Ono K, Kojima E, Yamada T, Eason DT, Goffinet FJL, Yano T
Development and Validation of a Novel Lymphedema Monitoring System Using a Bodysuit and a Smartphone: Prospective Comparative Study
JMIR Mhealth Uhealth 2026;14:e77935
URL: https://mhealth.jmir.org/2026/1/e77935
doi:10.2196/77935
PMID:

©Ryo Karakawa, Hidehiko Yoshimatsu, Katsu Karakawa, Kuniko Utsugi, Kengo Ono, Eri Kojima, Takayasu Yamada, Duncan Thomas Eason, Francois Jean Leon Goffinet, Tomoyuki Yano. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 06.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Associations Between Technology Attitudes, App Use, and Health Outcomes in a Diabetes Self-Management Education and Support Intervention: Secondary Analysis of a Randomized Controlled Trial

King David Oware[1], MPH, MBChB; Gang Han[1,2], PhD; Samuel D Castiglione Towne Jr[1,3,4,5], MPH, PhD; Christi Esquivel[1], MEd, PhD; SangNam Ahn[1,6], MPSA, PhD; Shinduk Lee[7], MSPH, DrPH; Matthew Lee Smith[1,8,9], MPH, PhD; Marcia G Ory[1,3], MPH, PhD

[1]Center for Community Health and Aging, School of Public Health, Texas A&M University, College Station, TX, United States
[2]Department of Epidemiology and Biostatistics, School of Public Health, Texas A&M University, College Station, TX, United States
[3]Department of Environmental and Occupational Health, School of Public Health, Texas A&M University, College Station, TX, United States
[4]School of Global Health Management and Informatics, University of Central Florida, Orlando, FL, United States
[5]Disability, Aging, and Technology Cluster, University of Central Florida, Orlando, FL, United States
[6]Department of Health Management and Policy, Saint Louis University, Saint Louis, MO, United States
[7]College of Nursing, University of Utah, Salt Lake City, UT, United States
[8]Department of Health Behavior, School of Public Health, Texas A&M University, College Station, TX, United States
[9]Center for Health by Design, Indiana University, Bloomington, IN, United States

Corresponding Author:
King David Oware, MPH, MBChB
Center for Community Health and Aging
School of Public Health
Texas A&M University
SPH Administrative Bldg, 3rd Fl.
212 Adriance Lab Rd
College Station, TX, 77843
United States
Phone: 1 979 680 2902
Email: owarekings@tamu.edu

Abstract

Background: Mobile health (mHealth) apps are increasingly incorporated into diabetes self-management education and support (DSMES) programs to improve access to behavioral and clinical support. Although digital DSMES interventions have demonstrated benefits for self-management and glycemic control, considerable variability exists in app use, and less is known about whether preintervention technology attitudes are associated with subsequent app use and related outcomes.

Objective: This study examined whether baseline technology attitudes are associated with app use and whether app use is associated with diabetes self-care and hemoglobin A_{1c} (HbA_{1c}) among adults participating in a digital DSMES intervention.

Methods: We conducted a secondary analysis of the Living Healthier with Diabetes randomized controlled trial (RCT) in Texas (2020-2022). The RCT included 3 study arms: virtual DSMES, smartphone app–based DSMES, and a combined intervention. Analyses were restricted to the 2 intervention arms with app access. Baseline technology attitude was assessed using a 4-item composite scale. App use was measured using 2 indices: a primary app use index derived from days per week, minutes per day, and overall 3-month usage, and a feature-use index summing use across 10 app functions. Outcomes included HbA_{1c} and diabetes self-care. Multivariable linear regression and linear mixed-effects models were used to evaluate associations after adjustment for intervention arm, age, gender, past month financial situation, and baseline outcome values.

Results: Of the 189 participants enrolled in the primary trial, 129 randomized to the intervention arms with smartphone app access were included. Follow-up HbA_{1c} and diabetes self-care data were available for 116 (89.9%) and 114 (88.4%) participants at 3 months and 117 (90.7%) and 116 (89.9%) participants at 6 months, respectively. More positive baseline technology attitudes

XSL•FO
RenderX

were associated with greater app use at both 3 months (β=0.64, 95% CI 0.29 to 0.99; *P*<.001) and 6 months (β=0.78, 95% CI 0.39 to 1.17; *P*<.001) but were not associated with feature use. App use was not associated with HbA_{1c} at 3 months but was associated with lower HbA_{1c} at 6 months (β=−0.04, 95% CI −0.074 to −0.013; *P*=.005) and in longitudinal models (β=−0.04, 95% CI −0.074 to −0.006; *P*=.02). Greater app use was associated with higher self-care at 3 months (β=0.56, 95% CI 0.29 to 0.82; *P*<.001) and 6 months (β=0.38, 95% CI 0.14 to 0.62; *P*=.002) and remained associated with self-care in longitudinal analyses (β=0.44, 95% CI 0.21 to 0.68; *P*<.001). App and feature use did not change significantly between 3 and 6 months (*P*=.14 and *P*=.11, respectively).

Conclusions: These findings suggest that preintervention technology attitudes are associated with greater app use within digital DSMES programs and that greater use of the app may represent an important pathway linking mHealth interventions with behavioral and clinical outcomes in adults with type 2 diabetes.

Trial Registration: ClinicalTrials.gov NCT06370494; http://clinicaltrials.gov/ct2/show/NCT06370494

(JMIR Mhealth Uhealth 2026;14:e105830) doi:10.2196/105830

KEYWORDS

type 2 diabetes; diabetes self-management education and support; mHealth; mobile apps; technology attitudes; app use; self-care; hemoglobin A1c; mobile health; mobile phone

Introduction

The number of people with diabetes (diagnosed and undiagnosed) in the United States surpassed 40 million as of recent estimates [1], with an increase in the age-adjusted prevalence rates (diagnosed and undiagnosed) for adults from 11.2% (2001-2004) to 13.5% (2021-2023). When considering diagnosed diabetes among adults during 2021-2023, the most recent data reveal marked geographic variations [1]. For example, higher rates were identified in nonmetropolitan areas as compared to metropolitan areas (10% vs 8.3%), for those with less than 100% of the federal poverty level (FPL) as compared to those at 500% or more of the FPL (13% vs 5.5%), and for those with less than a high school education vs those with more than high school (13.3 vs 7%) [1]. Beyond its rising prevalence, type 2 diabetes mellitus (T2D) is a complex, chronic illness that, when inadequately managed, confers substantial microvascular and macrovascular complications, which impair health-related quality of life and contribute to excess morbidity and mortality [2]. The economic burden is substantial; recent analyses provide estimates of the growing economic costs for diagnosed diabetes at US $412.9 billion in 2022 alone, of which US $306.6 billion was attributed to direct medical costs while another US $106.3 billion was attributed to indirect costs [3,4]. This represents a major financial burden to the US economy and to the millions of Americans impacted by this major public health issue. As such, there are several opportunities for potential interventions aimed at specific aspects of diabetes-related self-care.

Diabetes self-management education and support (DSMES) is a foundational component of type 2 diabetes care, facilitating the knowledge, skills, and behaviors required for effective self-management [5,6]. Evidence from a network meta-analysis of 108 randomized trials involving 17,735 participants demonstrated that DSMES significantly improves glycemic outcomes, reducing hemoglobin A_{1c} (HbA_{1c}) by an average of 0.61% compared with usual care [7]. Additional benefits include improvements in diabetes knowledge, self-efficacy, and weight-related outcomes, with effects sustained beyond the intervention period in some studies [8]. Despite its established effectiveness, DSMES remains substantially underused [9]. Referral, participation, and completion rates remain low, with disparities observed among individuals with lower income, lower educational attainment, and historically underserved populations [10,11].

Additionally, national evidence suggests that individuals experiencing lower socioeconomic status (SES) or who are Black or Hispanic may experience gaps in access to the internet at home [12]. The authors suggest that this link between lower access and potential for lags in access to health promotion may further worsen existing disparities [12]. Therefore, action must be taken to increase access and ameliorate both existing and potentially further disparities into the future as this cohort ages into adulthood [12,13]. Regardless of age, there is increased recognition of internet access as a social determinant of health [13,14], especially with more services being delivered through telehealth and patient portals that require access to the internet [14]. As such, this is a critical resource that warrants further investigation and widespread recognition to eliminate persistent disparities. These barriers have prompted increasing interest in alternative delivery approaches. In particular, digital and mobile health (mHealth)–based DSMES models have emerged as strategies that extend access to evidence-based DSMES beyond traditional in-person settings [15,16].

In 2018, a global review identified over 1000 apps focused on diabetes management alone [17]. The use of mHealth apps for T2D management has garnered significant research attention in recent years, with studies examining both clinical outcomes and self-management behaviors. Smartphone apps pose an alternative option that can help people with T2D improve HbA_{1c} levels by optimizing insulin management [18]. The proliferation of mHealth apps designed for T2D management has prompted researchers to evaluate their clinical effectiveness and impact on patient self-management. People living with T2D who use diabetes management apps demonstrated modest improvements in glycemic control and enhanced self-care behaviors [19-22]. Key factors linked to better outcomes included sustained app use, app features that support active self-monitoring, and

XSL•FO
RenderX

integration of feedback mechanisms [19-23]. A recent scoping review concluded that mobile apps for people living with diabetes allowed for convenient user experiences and contributed to improved blood glucose levels, with significant HbA_{1c} reductions [24]. Mobile apps have demonstrated potential for clinically meaningful reductions in HbA_{1c} levels, though results vary based on participants, intervention duration, and app use frequency [24].

While mobile apps offer an additional avenue for providing health-related interventions, technology attitudes may be associated with adoption and sustained use of mHealth apps, particularly in chronic disease self-management. A systematic review has demonstrated that attrition rates in mHealth interventions for chronic disease average approximately 43%, with real-world engagement often far lower [25]. Individual orientation toward technology has been described through related constructs such as personal innovativeness in IT and technology readiness. Personal innovativeness refers to an individual's willingness to try new information technologies [26], while technology readiness reflects a person's propensity to embrace and use new technologies [27]. These concepts provide a basis for considering technology attitudes as a person's general orientation toward new technologies. Prior research has linked technology readiness with technology adoption and use [28].

Although diabetes mobile apps and digital DSMES programs have been associated with improvements in self-care and HbA_{1c}, app use remains uneven and is influenced by multiple patient and implementation factors [23,29,30]. Moreover, even among those who initially adopt a diabetes app, a discrepancy has been observed between early uptake and sustained long-term use, with many individuals substantially reducing or discontinuing engagement over time [25,31]. Such variability in app use may help explain differences in behavioral and clinical outcomes observed across participants. Despite evidence that more positive attitudes toward technology are associated with willingness to use digital health tools, less is known about whether baseline technology attitudes predict subsequent app use within DSMES programs [32,33]. Consequently, an important gap remains in understanding whether preintervention technology attitudes are associated with subsequent app use within DSMES programs and whether app use is associated with diabetes self-management behaviors and glycemic outcomes.

To address these gaps, we used data from the Living Healthier with Diabetes trial, a 3-arm randomized controlled DSMES trial conducted in Texas to examine patterns of app use over time and their associations with outcomes [34]. Specifically, we aimed to (1) examine differences in sociodemographic factors between groups with high and low attitudes toward technology at baseline, (2) examine whether baseline technology attitudes were associated with app use at 3 and 6 months, (3) evaluate the association between app use and diabetes self-care and HbA_{1c}, and (4) assess whether app use declined between 3 and 6 months. We hypothesized that more positive baseline technology attitudes would be associated with greater app use at follow-up. We further expected that higher levels of app use would be associated with better self-care behaviors, given the proximal role of behavior change in DSMES, and with lower HbA_{1c}. Finally, given the structured support embedded within the intervention (eg, blood glucose tracking, health coaching), we expected that app use would not decline between 3- and 6-month assessments.

Methods

Study Design and Setting

We conducted a secondary analysis of data from a 3-arm randomized controlled trial (RCT) designed to assess the relative effectiveness of DSMES interventions in adults with T2D in Texas [34]. Participants were recruited and enrolled between November 2020 and March 2022. The three arms of the intervention were (1) a virtual asynchronous DSMES program with individualized follow-up support, called Virtual Making Moves With Diabetes (vMMWD); (2) a technology-based education and support delivered through a smartphone app (TBES); and (3) a combined modality where participants received vMMWD followed by TBES. Participants completed a web-based survey on sociodemographic, behavioral, and health data and HbA_{1c} assessment at 3 time points: baseline, 3 months, and 6 months. Details of the study design, participant recruitment, and primary findings related to HbA_{1c} outcomes have been previously reported [34].

Ory et al [34] evaluated the effectiveness of the 3 intervention approaches and reported significant improvements across all intervention arms. The present secondary analysis focused on technology attitudes, app use, and diabetes-related outcomes among participants with access to the smartphone app. We therefore restricted the sample to participants in the TBES-only and combined arms. Specifically, we examined whether baseline technology attitude was associated with app use at 3- and 6-month follow-up assessments and whether app use was associated with better self-care behaviors and glycemic outcomes. The combined arm provided app use data during the TBES exposure period, which occurred after the vMMWD component was completed.

Participants

The inclusion criteria for the trial were being aged ≥25 years, self-reporting as T2D, having baseline HbA_{1c} ≥7.5%, living in Texas, having a smartphone with internet access, and being able to read and speak English. Pregnant individuals and individuals who previously had conditions known to invalidate HbA_{1c} values (eg, liver failure, kidney dialysis, or organ transplant) were excluded from participation. Participants were recruited through community partnerships, provider referrals, social media, and radio ads [34].

Of the 189 participants enrolled in the trial, those assigned to the vMMWD-only arm (n=60) were excluded because participants in this arm did not have access to the mobile app. Our analytic sample at baseline (N=129) included participants from the TBES-only (n=68) and combined arms (n=61). Analyses were conducted using available case data.

XSL•FO
RenderX

Intervention

Participants randomized to TBES were given access to a smartphone app designed to support diabetes self-management [34]. The intervention was delivered over 3 months; however, participants continued to have access to the app thereafter and used the app at their own pace. The app included features for tracking blood glucose, meals, physical activity, and medications, as well as medication reminders, health coaching, and access to diabetes-related information (Table S1 in Multimedia Appendix 1). Participants could also communicate through the app with a diabetes coach, who provided text-based support several times per week, including encouragement, recipe ideas, and support related to blood glucose and other health tracking. Participants in the combined arm received both interventions sequentially, completing the vMMWD DSMES program before receiving access to the TBES mobile app. Compared to TBES, vMMWD included more structured education and counseling support through 8 hours of online diabetes education and periodic 1:1 virtual education and counseling with a clinical diabetes educator (eg, a registered nurse or registered dietitian nutritionist).

Variables

Primary Exposure: App Use

App use was assessed using an index score based on 3 self-reported items representing the intensity and frequency of app use over the past 3 months. App use was assessed at the 3- and 6-month follow-up surveys. The three items were (1) the number of times participants used the app for diabetes management over the preceding 3 months (10-point scale, 1 [never] to 10 [all the time]), (2) days per week of app use (0 to 7 days per week), and (3) average minutes per day of app use (7-point ordinal scale, 1 through 7 corresponded to 0-10 minutes, 11-20 minutes, 21-30 minutes, 31-40 minutes, 41-50 minutes, 51-60 minutes, and more than 60 minutes per day, respectively). Due to the heterogeneity of these scales, we rescaled each component to a common 0-10 metric using the transformation:



where X is the original observed value, X_{min} is the lowest possible value of the scale, X_{range} denotes the span of possible values of the scale (calculated as the maximum minus the minimum), and n sets the upper limit of the rescaled scale (here, n=10). The app use variable was created by summing the 3 rescaled components. Scores on the app use index ranged from 0 to 30, with higher values indicating more overall use of the app (Table S2 in Multimedia Appendix 1).

Secondary Exposure: Feature Use

Additionally, at 3- and 6-month follow-up assessments, participants rated the frequency with which they used 10 specific app features over the preceding 3 months. Responses were captured using a 10-point scale ranging from 1 (never) to 10 (all the time). The assessed features are summarized in Table S1 in Multimedia Appendix 1. To capture the overall use of the features in the app, a feature use index was calculated as the sum of responses across the 10 features. Possible scores ranged from 10 to 100, where a score of 10 indicates no reported feature use, and higher scores reflect more frequent use of individual features, broader use across multiple features, or both.

Baseline Technology Attitude

In this study, technology attitude refers to a person's general orientation toward new technologies. We assessed baseline technology attitude using 4 items. Study participants indicated their level of agreement with the following items on a 5-point Likert scale ranging from 1 (strongly agree) to 5 (strongly disagree): (1) I usually try new technologies before others do, (2) I generally avoid using new equipment or technology, (3) I enjoy using new equipment or technology, and (4) my friends come to me to learn about new technologies. The 3 positively worded items (1, 3, and 4) were reverse-coded so that higher scores across all 4 items reflected a more positive attitude toward technology (Table S3 in Multimedia Appendix 1). Baseline technology attitude score was computed by summing responses across the 4 items; scores therefore ranged from 4 to 20, with higher scores indicating a more positive attitude toward technology. The 4 items were adapted for this study. Internal consistency of the technology attitude measure was acceptable (Cronbach α=0.798).

Outcomes

The primary clinical outcome was HbA_{1c}. Given the statewide reach of the trial and restrictions on routine health care access during the COVID-19 pandemic, we chose a home-based testing approach using self-administered A_{1c}Now+ kits (PTS Diagnostics). This method has been used in previous diabetes self-management research [35]. Study personnel provided standardized testing instructions to supplement the manufacturer's instructions. Participants assessed their HbA_{1c} at baseline, 3 months, and 6 months and reported their results to the study team. HbA_{1c} was treated as a continuous outcome with possible values ranging from 4% to 14%. Higher scores represent poorer glycemic control.

Diabetes self-care was the secondary outcome and was measured using a 9-item modified Summary of Diabetes Self-Care Activities [36]. Items included assessing diet, physical activity, blood glucose monitoring, and foot care in the past 7 days using a scale ranging from 0 to 7 (Table S4 in Multimedia Appendix 1). The items were harmonized and summed up, creating a possible score ranging from 0 to 63. Higher scores represent greater performance of diabetes self-care activities. Cronbach α at baseline, 3 months, and 6 months were 0.81, 0.75, and 0.79, respectively.

Covariates

Additional variables considered important were selected based on existing literature [34], including intervention arm (TBES-only vs combined), age (years), gender (male or female), and self-reported financial situation over the past month (categorized as having money left over, just enough, or not enough). Race or ethnicity (non-Hispanic White, Hispanic, or other) and marital status (married or partnered, or other) were also collected and are reported to describe the study sample.

XSL•FO
RenderX

Ethical Considerations

The RCT was approved by the Texas A&M Institutional Review Board (IRB; #2019-0804D) and registered at ClinicalTrials.gov (NCT06370494). All participants provided written informed consent before enrollment, and participation was voluntary. Participants could withdraw from the study at any time without a penalty. Study data were deidentified and deposited in the Texas A&M institutional data repository, with access restricted to members of the study team. Participants could receive up to US $375 in compensation based on completion of study assessments. This secondary analysis used deidentified data collected under the original study approval.

Sample Size

The RCT used an a priori sample-size calculation based on its primary outcome, HbA_{1c}. Assuming a moderate effect size (Cohen *d*=0.5), 34 participants with baseline and postintervention measurements were required to provide 80% power at a 2-sided α=.05 [34]. The present study is a secondary analysis restricted to participants in the 2 intervention arms with access to the mobile app, with 129 participants available at baseline. Because this analysis used the available sample from the primary trial, no new prospective sample-size calculation was performed. For transparency, we conducted a sensitivity analysis for the multivariable linear regression models using the analytic sample available for each model (n=106-113). At 80% power and a 2-sided α=.05, the minimum detectable effect size for the focal predictor (Cohen f^2) ranged from 0.074 to 0.080 across the primary regression models. At 90% power, the corresponding minimum detectable effect size ranged from 0.099 to 0.107. These values indicate the analysis was sensitive to small-to-medium standardized effects.

Statistical Analysis

Baseline characteristics of participants were summarized using means with SDs for continuous variables and frequencies with percentages for categorical variables. Median splits of baseline technology attitude and app use at 3-month follow-up were used to separately stratify the participants' characteristics at baseline. Group differences were compared using a 2-sided Welch *t* test for continuous variables and Pearson chi-square test for categorical variables.

Separate multivariable linear regression models were fitted to investigate the association between baseline technology attitude and app use at 3- and 6-month follow-ups. Two outcomes were considered: app use and feature use. All models were adjusted for the intervention arm, age, gender, and the participant's baseline past-month financial situation.

Multivariable linear regression models were then used to investigate the cross-sectional associations between app use and health and behavioral outcomes at each of the follow-up time points. The primary predictor was app use measured at 3 and 6 months, and outcomes included HbA_{1c} and self-care assessed at the respective time points. The models were adjusted for baseline values of the respective outcome variables, intervention arm, age, gender, and past month's financial situation. The same model specifications were subsequently applied using feature use as an alternate predictor. The regression analyses used complete cases for each model. To assess the potential impact of missing data, we conducted sensitivity analyses using multiple imputation by chained equations (50 imputed datasets) with the *mice* package in R (R Foundation). Estimates were pooled using Rubin rules. These analyses yielded comparable estimates for the primary app-use associations.

Linear mixed-effects models with random participant intercepts were used to evaluate the association between app use and outcomes (HbA_{1c} and self-care) over time. In these models, app use at 3 months was included as a predictor of longitudinal outcomes measured at 3 and 6 months. Models were adjusted for baseline values of outcomes, intervention arm, age, gender, and past month's financial situation. The same model specifications were then applied using feature use as an alternate predictor.

To assess whether use of the app was sustained during the 3- to 6-month assessments, paired *t* tests were performed on the app use and on the feature use. Within-subject differences were calculated as 6-month minus 3-month, and the mean changes were reported with a 95% CI.

We assessed model assumptions by visually examining residual plots for linearity, homoscedasticity, and normality. Multicollinearity was assessed using variance inflation factor (VIF) and tolerance values (VIF range 1.00-1.47; tolerance range 0.68-1.00). Influential observations were assessed using Cook distance, with all values below 1 (maximum 0.37). All statistical analyses were conducted in R software, version 4.5.2. A 2-sided α of .05 was used as the threshold for statistical significance.

Results

Participant Characteristics

Figure 1 illustrates the participant selection process and data availability for the secondary analysis. At baseline, the analytic sample included 129 participants with access to the mobile app. Follow-up data were available for 116 (89.9%) participants for HbA_{1c} and 114 (88.4%) for self-care at 3 months and for 117 (90.7%) and 116 (89.9%), respectively, at 6 months.

Table 1 shows baseline characteristics stratified by median split of baseline technology attitude. Participants with an above-median technology attitude were younger (mean 49.2, SD 11.8) compared with those at or below the median (mean 53.5, SD 9.6; *P*=.03). No other significant differences were observed across attitude groups for gender, race and ethnicity, marital status, financial situation, baseline HbA_{1c}, or baseline self-care (all *P*>.05).

Baseline characteristics of participants with available app use data at 3 months (n=116) stratified by median split of 3-month app use are presented in Table S1 in Multimedia Appendix 2. A significantly greater proportion of participants in the combined arm fell into the lower app use group (36/58, 62.1%) compared to the proportion in the app-only arm (*P*=.02). Baseline technology attitude was higher among those with above-median app use (mean 14.6, SD 3.2) compared to those

XSL•FO
RenderX

at or below the median (mean 13.3, SD 3.7; *P*=.04). No other significant differences were observed.

Figure 1. Flow diagram showing participant selection and availability of HbA_{1c} and diabetes self-care data at 3- and 6-month follow-up. HbA_{1c}: hemoglobin A_{1c}; vMMWD: Virtual Making Moves With Diabetes.

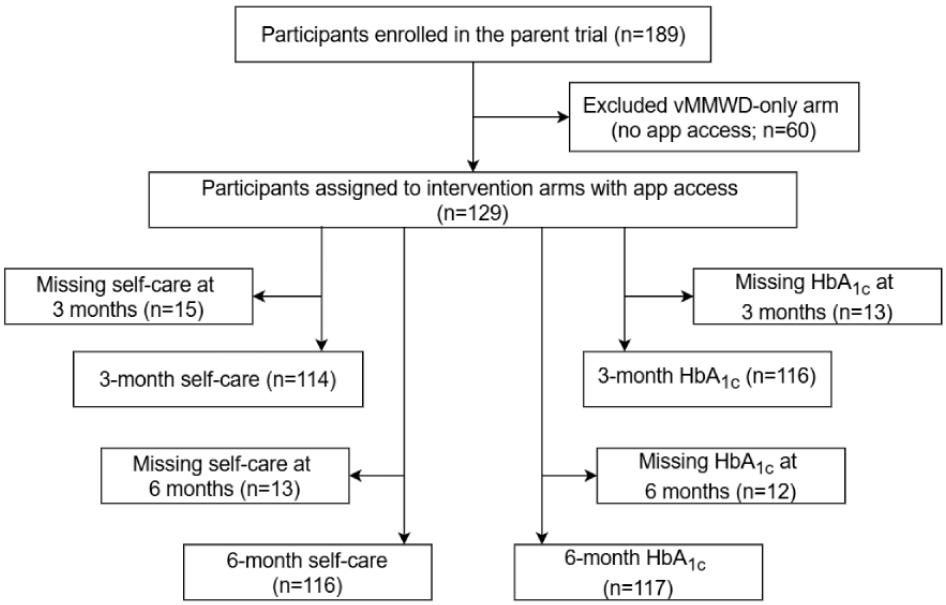


Table 1. Participant characteristics at baseline, stratified by baseline technology attitude (median 14, IQR 12-16).

Participant characteristics and baseline measures	Overall (N=129)	Technology attitude ≤14 (n=79)	Technology attitude >14 (n=50)	*P* value
Intervention arm, n (%)				.50
App only	68 (52.7)	44 (55.7)	24 (48.0)	
Combined	61 (47.3)	35 (44.3)	26 (52.0)	
Age (years), mean (SD)	51.9 (10.7)	53.5 (9.6)	49.2 (11.8)	.03
Gender, n (%)				.63
Female	99 (76.7)	59 (74.7)	40 (80.0)	
Male	30 (23.3)	20 (25.3)	10 (20.0)	
Race or ethnicity, n (%)				.28
Non-Hispanic White	66 (51.2)	39 (49.4)	27 (54.0)	
Hispanic	48 (37.2)	28 (35.4)	20 (40.0)	
Other	15 (11.6)	12 (15.2)	3 (6.0)	
Marital status, n (%)				.87
Married or partnered	85 (65.9)	53 (67.1)	32 (64.0)	
Other marital status	44 (34.1)	26 (32.9)	18 (36.0)	
Financial situation over the past month, n (%)				.18
Money left over	49 (38.9)	30 (39.0)	19 (38.7)	
Just enough	58 (46.0)	32 (41.5)	26 (53.1)	
Not enough	19 (15.1)	15 (19.5)	4 (8.2)	
Baseline HbA_{1c}[a] (%), mean (SD)	9.5 (1.6)	9.4 (1.6)	9.6 (1.7)	.59
Baseline self-care, mean (SD)	27.5 (13.5)	26.3 (13.5)	29.2 (13.4)	.23

[a]HbA_{1c}: hemoglobin A_{1c}.

XSL•FO
RenderX

Association of Baseline Technology Attitude With App Use

Higher baseline technology attitude was significantly associated with greater app use at both 3 months (β=0.64, 95% CI 0.29 to 0.99; *P*<.001) and 6 months (β=0.78, 95% CI 0.39 to 1.17; *P*<.001; Table S2 in Multimedia Appendix 2). Participants randomized to the combined arm had significantly lower app use at 3 months compared with participants randomized to the app-only arm (β=–2.90, 95% CI –5.17 to –0.62; *P*=.01). On the other hand, higher baseline technology attitude was not significantly associated with feature use at 3 months (β=0.72, 95% CI –0.23 to 1.67; *P*=.14) or 6 months (β=0.67, 95% CI –0.35 to 1.68; *P*=.20) after adjustment for covariates (Table S3 in Multimedia Appendix 2).

Association of App Use With HbA1c

The complete-case cross-sectional regression models for HbA_{1c} included 112 participants at 3 months and 113 at 6 months. At 3 months, app use was not associated with HbA_{1c} following adjustment (β=–0.04, 95% CI –0.08 to 0.00; *P*=.08; Table 2). However, at 6-month follow-up, higher app use was associated with lower HbA_{1c} (β=–0.04, 95% CI –0.074 to –0.013; *P*=.005; Table 2). In the longitudinal analysis, app use was associated with lower Hb A_{1c} (β=–0.04, 95% CI –0.074 to –0.006; *P*=.02; Table 3). Time was not significantly associated with HbA_{1c} across follow-up (*P*=.18).

Table 2. Multivariable linear regression analyses examining the association of app use with HbA_{1c}[a]. Analyses were restricted to participants with complete data on all variables included in each regression model.

Variables and factors	Beta value (95% CI)	*P* value
Cross-sectional analysis at the 3-month follow-up (N=112)		
App use at 3 months	–0.04 (–0.08 to 0.00)	.08
Baseline HbA_{1c}	0.39 (0.21 to 0.57)	<.001
Intervention arm (combined; reference: app only)	0.21 (–0.34 to 0.76)	.45
Age (years)	–0.01 (–0.04 to 0.02)	.41
Gender (reference: female)	–0.06 (–0.75 to 0.64)	.87
Financial situation over the past month		
Money left over	—[b]	—
Just enough	–0.16 (–0.78 to 0.46)	.61
Not enough	0.87 (–0.02 to 1.76)	.05
Cross-sectional analysis at the 6-month follow-up (N=113)		
App use at 6 months	–0.04 (–0.074 to –0.013)	.005
Baseline HbA_{1c}	0.38 (0.23 to 0.52)	<.001
Intervention arm (combined; reference: app only)	–0.06 (–0.49 to 0.37)	.79
Age (years)	–0.03 (–0.05 to –0.01)	.01
Gender (reference: female)	–0.50 (–1.06 to 0.05)	.08
Financial situation over the past month		
Money left over	—	—
Just enough	–0.33 (–0.84 to 0.18)	.20
Not enough	0.89 (0.16 to 1.62)	.02

[a]HbA_{1c}: hemoglobin A_{1c}.

[b]Reference group.

XSL•FO
RenderX

Table 3. Linear mixed-effects model examining the longitudinal association of app use with HbA_{1c}[a] across follow-up.

Variables and factors	Beta value (95% CI)	*P* value
Time	–0.17 (–0.41 to 0.08)	.18
App use at 3 months	–0.04 (–0.074 to –0.006)	.02
Baseline HbA_{1c}	0.39 (0.25 to 0.53)	<.001
Intervention arm (combined; reference: app only)	0.10 (–0.32 to 0.53)	.63
Age (years)	–0.02 (–0.04 to 0.00)	.08
Gender (reference: female)	–0.35 (–0.89 to 0.19)	.20
Financial situation over the past month		
Money left over	__[b]	—
Just enough	–0.23 (–0.71 to 0.25)	.34
Not enough	0.83 (0.14 to 1.53)	.02

[a]HbA_{1c}: hemoglobin A_{1c}.

[b]Reference group.

Association of App Use With Diabetes Self-Care

For self-care, the complete-case cross-sectional regression models included 108 participants at 3 months and 111 at 6 months. Greater app use was associated with better self-care at 3 months (β=0.56, 95% CI 0.29 to 0.82; *P*<.001; Table 4). Similarly, at 6 months, app use continued to show an association with self-care (β=0.38, 95% CI 0.14 to 0.62; *P*=.002; Table 4). In longitudinal analysis, app use was associated with higher self-care across follow-up (β=0.44, 95% CI 0.21 to 0.68; *P*<.001; Table 5). The main effect of time was not significantly associated with self-care changes (*P*=.79).

Table 4. Multivariable linear regression analyses examining the association of app use with diabetes self-care. Analyses were restricted to participants with complete data on all variables included in each regression model.

Variables and factors	Beta value (95% CI)	*P* value
Cross-sectional analysis at the 3-month follow-up (N=108)		
App use at 3 months	0.56 (0.29 to 0.82)	<.001
Baseline self-care	0.44 (0.32 to 0.57)	<.001
Intervention arm (combined; reference: app only)	1.14 (–2.24 to 4.52)	.51
Age (years)	0.05 (–0.11 to 0.21)	.53
Gender (reference: female)	2.34 (–1.73 to 6.41)	.26
Financial situation over the past month		
Money left over	__[a]	—
Just enough	2.20 (–1.59 to 5.98)	.25
Not enough	6.81 (1.64 to 11.98)	.01
Cross-sectional analysis at the 6-month follow-up (N=111)		
App use at 6 months	0.38 (0.14 to 0.62)	.002
Baseline self-care	0.54 (0.41 to 0.67)	<.001
Intervention arm (combined; reference: app only)	2.97 (–0.43 to 6.36)	.09
Age (years)	0.16 (–0.01 to 0.33)	.07
Gender (reference: female)	5.16 (0.83 to 9.50)	.02
Financial situation over the past month		
Money left over	—	—
Just enough	4.51 (0.56 to 8.46)	.03
Not enough	6.84 (1.16 to 12.53)	.02

[a]Reference group.

XSL•FO
RenderX

Table 5. Linear mixed-effects model examining the longitudinal association of app use with diabetes self-care across follow-up.

Variables and factors	Beta value (95% CI)	*P* value
Time	–0.24 (–2.01 to 1.54)	.79
App use at 3 months	0.44 (0.21 to 0.68)	<.001
Baseline self-care	0.51 (0.40 to 0.62)	<.001
Intervention arm (combined; reference: app only)	2.18 (–0.78 to 5.14)	.15
Age (years)	0.10 (–0.04 to 0.24)	.18
Gender (reference: female)	3.50 (–0.13 to 7.13)	.06
Financial situation over the past month		
Money left over	—[a]	—
Just enough	3.22 (–0.10 to 6.54)	.06
Not enough	6.02 (1.36 to 10.67)	.01

[a]Reference group.

Association of Feature Use and Outcomes

Feature use was not associated with HbA_{1c} at 3 months (β=0.00, 95% CI –0.02 to 0.02; *P*=.82). At 6 months, feature use was associated with lower HbA_{1c} (β=–0.01, 95% CI –0.03 to 0.00; *P*=.02; Table S4 in Multimedia Appendix 2). In the longitudinal model, the association between feature use at 3 months and HbA_{1c} across follow-up did not reach statistical significance (Table S5 in Multimedia Appendix 2). Turning to self-care, feature use was associated with self-care at both 3 months (β=0.18, 95% CI 0.07 to 0.29; *P*=.002) and 6 months (β=0.16, 95% CI 0.06 to 0.27; *P*=.002; Table S6 in Multimedia Appendix 2). The association remained statistically significant in the longitudinal analysis (β=0.14, 95% CI 0.04 to 0.23; *P*=.005; Table S7 in Multimedia Appendix 2).

Change in App Use Between 3- and 6-Month Follow-Up

Paired *t* tests showed no significant decline in use between the 3- and 6-month assessments. App use was not significantly different between 3 months (mean 9.13, SD 6.48) and 6 months (mean 8.49, SD 7.13), with a mean change of –0.64 (SD 4.52, 95% CI –1.49 to 0.21; *P*=.14). Similarly, feature use was not significantly different between 3 months (mean 24.45, SD 16.49) and 6 months (mean 27, SD 18.77), with a mean change of 2.55 (SD 16.45, 95% CI –0.56 to 5.65; *P*=.11; Tables S8 in Multimedia Appendix 2).

Sensitivity Analysis

Sensitivity analyses using multiple imputation yielded estimates comparable to the complete-case analyses. The association between app use and HbA_{1c} at 6 months remained statistically significant, whereas the association at 3 months remained nonsignificant. Positive associations between app use and self-care were maintained at both follow-up assessments (Table S9 in Multimedia Appendix 2).

Discussion

Principal Results and Interpretations

This study examined the association between baseline technology attitude, app use, and diabetes outcomes within a digital DSMES intervention. Baseline technology attitude was associated with greater app use at both 3 and 6 months, with stronger estimates observed at 6 months. No clear association was observed with feature use. Higher app use, in turn, was associated with better diabetes self-care at both 3 and 6 months. In contrast, the association between app use and HbA_{1c} at 3 months was not statistically significant, but higher app use was associated with lower HbA_{1c} at 6 months. In longitudinal analyses, app use at 3 months remained associated with both self-care and HbA_{1c} across follow-up. Finally, app use did not decline significantly between 3 and 6 months.

In this study, higher baseline technology attitude was associated with greater app use at both 3 and 6 months, independent of age, gender, financial situation, and intervention arm. The larger estimate at 6 months suggests that technology attitudes may remain relevant as participants transition from early exposure to longer-term app use. This finding is consistent with previous research showing that individuals with greater personal innovativeness toward new technologies used a health promotion app more frequently [37]. Similar findings have been reported among adults with T2D, where participants who continued using a diabetes management app had more positive baseline attitudes toward technology, and more positive attitudes were associated with greater frequency and intensity of app use among noninsulin users [38]. In contrast, baseline technology attitude was not associated with feature use. This divergence suggests that general attitudes toward technology may shape willingness to incorporate an app into self-management routines, whereas use of individual features may depend on perceived relevance, usability, and personal need. Similar heterogeneity in module use has been reported in real-world diabetes app studies, where most users used only a single app component, a pattern the authors attributed to the highly heterogeneous self-management needs of patients with diabetes [32].

The pattern of findings for HbA_{1c} suggests that the association between mHealth app use and glycemic control may differ across follow-up assessments. Although app use was not associated with HbA_{1c} at 3 months, higher use was associated with lower HbA_{1c} at 6 months, and 3-month use was associated

XSL•FO
RenderX

with HbA_{1c} in the longitudinal analysis. This temporal pattern is consistent with evidence indicating that clinically meaningful glycemic improvements often require sustained exposure to digital self-management interventions. Studies of diabetes apps and digital therapeutics have reported greater HbA_{1c} reductions among participants with higher levels of app use, including dose-response patterns across engagement groups [39] and stronger clinical effects after 6 months of continued use [40]. However, findings across the mHealth literature are not entirely consistent. Vehi et al [41] observed improvements in glycemic outcomes regardless of use frequency, suggesting that frequency alone may not fully explain clinical benefit. Collectively, these findings suggest that the association between app use and glycemic outcomes may extend beyond frequency of use alone and may involve the quality, consistency, and behavioral context of use.

The clinical significance of these findings warrants consideration. Although the adjusted association between app use and HbA_{1c} at 6 months was modest (β=−0.04 per point), the observed IQR of the app-use index (1.43-14.19) corresponds to an estimated 0.51 percentage-point lower HbA_{1c} at the higher level of app use, based on the fitted regression model. This magnitude is comparable to the pooled HbA_{1c} improvement of approximately 0.2-0.4 percentage points reported across intervention categories in a systematic review and meta-analysis of randomized trials of mHealth-delivered DSMES, which the authors characterized as clinically relevant [16]. However, unlike intervention effects reported in randomized trials, our estimate reflects an observational association between self-reported app use and HbA_{1c}. Thus, the observed association may be clinically relevant across meaningful differences in app use, but its causal interpretation remains uncertain.

Greater app use demonstrated a consistent association with diabetes self-care at both 3 and 6 months and remained significant in longitudinal analyses. This finding is consistent with prior mHealth research showing that diabetes app users report higher levels of self-management behaviors than nonusers [21]. The feature-use findings further support this interpretation, as greater use of app features was also associated with higher self-care scores. Prior research suggests that diabetes apps support self-management through functions such as blood glucose monitoring, dietary recording, medication management and reminders, physical activity tracking, and tailored education or real-time feedback, which are intended to reinforce adherence to self-care regimens [42,43]. Real-world evidence from the myDiabetes app similarly suggests that use of digital education and self-monitoring functions can support structured diabetes education and ongoing self-management activities [44]. Taken together, these findings suggest that digital DSMES apps may support self-care by facilitating ongoing self-monitoring, feedback, and behavioral reinforcement.

App and feature use did not significantly decline between the 3- and 6-month assessments. This finding contrasts with the pattern reported in mHealth interventions, where app use declines commonly after an initial adoption period. Large-scale evaluations of digital diabetes prevention programs have documented progressive reductions in use over time, with many participants disengaging from intervention features within the first several months of enrollment [45]. The stability observed in the present study may reflect the structured nature of DSMES delivery and the continued support provided within the intervention. Similar observations have been reported in studies showing that sustained retention and continued app use are associated with more favorable long-term glycemic outcomes [46].

Significance and Strengths

This study contributes to digital DSMES literature in 3 ways. First, previous research linking technology attitudes to mHealth use has been conducted within acceptance-based frameworks, which primarily model behavioral intention rather than directly measured usage [47,48]. In this study, we assessed technology attitude before exposure to the app and then examined its association with reported app use at 3 and 6 months. This allowed us to evaluate whether baseline disposition toward technology was associated with subsequent app use. Second, many mHealth studies characterize app use using single measures, such as logins or time-based indicators [49,50]. We assessed app use using an app-use index capturing frequency, regularity, and duration, alongside a separate index of use across specific app features. This provided complementary measures of how participants used the app. Third, app use is not often evaluated over time in mHealth research. In a systematic review, only one-third of studies assessed change in engagement over time [50]. By measuring app and feature use at 3 and 6 months within a randomized DSMES trial, our study adds temporal depth to a literature that often relies on a single time point.

Limitations

This study has several limitations. App use was determined using self-report, rather than objective app-use logging data, which, while common in community-based research, may be subject to recall and social desirability biases. Despite this, self-reported app use can capture dimensions of engagement that may not be fully captured by objective usage metrics alone [51]. Additionally, diabetes self-care was self-reported and may therefore be subject to recall and social desirability bias. HbA_{1c} was measured using self-administered home testing kits, with participants reporting their results to the study team. This approach may be less precise than laboratory-based measurement [52] and may introduce reporting errors.

The feature-use index combines frequency of use across 10 app features into a single score. Similar scores could therefore represent different patterns of use, such as frequent use of a small number of features or less frequent use across several features. The index should be interpreted as a measure of overall feature use and does not distinguish the breadth and frequency of feature use. Residual confounding is also possible because some potentially relevant factors were not measured and so were not included in the adjusted models. These include medication regimen or insulin use, digital literacy, and other factors that could be related to both app use and study outcomes. Additionally, participants in the combined arm received structured DSMES before app access. Although the regression models adjusted for the intervention arm, differences in prior structured DSMES exposure, self-management knowledge, and

XSL•FO
RenderX

confidence between the intervention arms could still have contributed to the observed associations.

Restricting the analysis to the 2 intervention arms with app access was necessary because app use was only defined for participants assigned to receive the mobile app. Because app access was determined by randomized treatment assignment, this restriction is less likely to introduce collider bias through self-selection into app use. Nevertheless, trial eligibility required smartphone and internet access, and the analyses relied on available follow-up data. These factors may introduce selection bias and limit the generalizability of the findings to populations with lower digital access or different patterns of follow-up participation. Lastly, recruitment and data collection for the parent trial occurred during the COVID-19 pandemic [34]. The pandemic may have influenced participants' willingness to enroll in a digital DSMES intervention as well as their technology attitudes, access to digital resources, and patterns of app engagement. Consequently, the observed associations may not fully reflect technology use and engagement under usual, nonpandemic conditions.

Research and Policy Implications

Future mHealth research should move beyond evaluating app use through single, isolated metrics such as login counts or session duration [50]. Our parallel analysis of app and feature use indices demonstrated that each may capture distinct dimensions of participant interaction with digital interventions; for example, baseline technology attitude was associated with the app use index but not feature use. Researchers should adopt multidimensional, longitudinal approaches to better understand how technology attitudes influence different patterns of app use and how these patterns relate to behavioral and clinical outcomes over time [50].

From a policy perspective, the absence of a significant decline in app use between 3 and 6 months suggests that engagement with digital DSMES may be sustained over short-term follow-up. However, the present study did not assess costs, cost-effectiveness, health care use, reimbursement, or implementation economics. Future studies should examine the resources required to assess technology-related factors and support digital DSMES uptake, including whether such implementation strategies are cost-effective and whether reimbursement approaches can facilitate equitable access.

Conclusions

This study provides new evidence on the relationship between technology attitudes, app use, and diabetes outcomes within a digital DSMES intervention. More positive baseline technology attitudes were associated with greater app use, while higher app use was associated with better diabetes self-care and lower HbA_{1c} over follow-up. App use did not significantly decline between 3 and 6 months, contrasting with the declining use patterns commonly reported in mHealth interventions. These findings suggest that technology attitudes may influence how individuals interact with digital DSMES programs and highlight the importance of considering technology-related factors when designing, implementing, and evaluating mHealth interventions for adults with type 2 diabetes.

Acknowledgments

The authors sincerely thank Blue Cross and Blue Shield of Texas for their generous support of this research. We express our deepest gratitude to Carrie Byington, MD, and Nancy Dickey, MD, for their vital leadership, guidance, and support at the Texas A&M Health Science Center throughout this joint effort. Additionally, the authors wish to thank Keri Carpenter, MPH, and Wendy Creighton, BSN, for their invaluable contributions to program development and fieldwork coordination.

Funding

This research was supported by a grant from Blue Cross and Blue Shield of Texas to establish the Texas A&M University Health Science Center Rural Health Moonshot Initiative. The funder had no role in the design of the study, data collection, data analysis, interpretation of the results, or preparation of this paper.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to restrictions enforced by Blue Cross Blue Shield of Texas. Requests to access the datasets should be directed to the Texas A&M point of contact for all data (Mr Jim Colson, MS, Texas A&M Associate Vice President, Digital Health, jim.colson@tamu.edu).

Generative AI Use

The authors declare that generative AI was not used in the preparation of this manuscript.

Authors' Contributions

Conceptualization: KDO, GH, MLS, MGO
Formal analysis: KDO
Funding acquisition: MLS, MGO
Project administration: MLS, MGO

XSL•FO
RenderX

Supervision: GH, MLS, MGO
Writing – original draft: KDO, SDCTJ, CE, SA, SL
Writing – review & editing: KDO, GH, SDCTJ, CE, SA, SL, MLS, MGO
MLS and MGO served as senior authors, reflecting their roles as coprincipal investigators and joint scientific directors for this project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definitions and scoring of study measures.
[DOCX File , 19 KB - mhealth_v14i1e105830_app1.docx]

Multimedia Appendix 2
Additional statistical results.
[DOCX File , 41 KB - mhealth_v14i1e105830_app2.docx]

References

1. National Diabetes Statistics Report. Centers for Disease Control and Prevention. 2026. URL: https://usdss.cdc.gov/diabetes/report.html [accessed 2026-11-05]
2. Cannon A, Handelsman Y, Heile M, Shannon M. Burden of illness in type 2 diabetes mellitus. J Manag Care Spec Pharm 2018;24(9-a Suppl):S5-S13 [FREE Full text] [doi: 10.18553/jmcp.2018.24.9-a.s5] [Medline: 30156443]
3. American Diabetes Association. Economic costs of diabetes in the U.S. in 2017. Diabetes Care 2018;41(5):917-928 [FREE Full text] [doi: 10.2337/dci18-0007] [Medline: 29567642]
4. Parker ED, Lin J, Mahoney T, Ume N, Yang G, Gabbay RA, et al. Economic costs of diabetes in the U.S. in 2022. Diabetes Care 2024;47(1):26-43. [doi: 10.2337/dci23-0085] [Medline: 37909353]
5. Davis J, Fischl AH, Beck J, Browning L, Carter A, Condon JE, et al. 2022 national standards for diabetes self-management education and support. Sci Diabetes Self Manage Care 2022;48(1):44-59. [doi: 10.1177/26350106211072203] [Medline: 35049403]
6. Powers MA, Bardsley JK, Cypress M, Funnell MM, Harms D, Hess-Fischl A, et al. Diabetes Self-management Education and Support in Adults With Type 2 Diabetes: A Consensus Report of the American Diabetes Association, the Association of Diabetes Care & Education Specialists, the Academy of Nutrition and Dietetics, the American Academy of Family Physicians, the American Academy of PAs, the American Association of Nurse Practitioners, and the American Pharmacists Association. Diabetes Care 2020;43(7):1636-1649. [doi: 10.2337/dci20-0023] [Medline: 32513817]
7. Romadlon DS, Tu Y, Chen Y, Hasan F, Kurniawan R, Chiu H. Comparative effects of diabetes self-management programs on type 2 diabetes clinical outcomes: a systematic review and network meta-analysis. Diabetes Metab Res Rev 2024;40(6):e3840. [doi: 10.1002/dmrr.3840] [Medline: 39267360]
8. Odgers-Jewell K, Ball LE, Kelly JT, Isenring EA, Reidlinger DP, Thomas R. Effectiveness of group-based self-management education for individuals with type 2 diabetes: a systematic review with meta-analyses and meta-regression. Diabet Med 2017;34(8):1027-1039. [doi: 10.1111/dme.13340] [Medline: 28226200]
9. Beck J, Greenwood DA, Blanton L, Bollinger ST, Butcher MK, Condon JE, 2017 Standards Revision Task Force. 2017 National Standards for Diabetes Self-Management Education and Support. Diabetes Care 2017;40(10):1409-1419. [doi: 10.2337/dci17-0025] [Medline: 28754780]
10. Tharakan A, McPeek Hinz E, Zhu E, Denmeade B, German J, Huang WA, et al. Accessibility of diabetes education in the United States: barriers, policy implications, and the road ahead. Health Aff Sch 2024;2(8):qxae097 [FREE Full text] [doi: 10.1093/haschl/qxae097] [Medline: 39206435]
11. Boakye EA, Varble A, Rojek R, Peavler O, Trainer AK, Osazuwa-Peters N, et al. Sociodemographic factors associated with engagement in diabetes self-management education among people with diabetes in the United States. Public Health Rep 2018;133(6):685-691 [FREE Full text] [doi: 10.1177/0033354918794935] [Medline: 30223759]
12. Dolcini MM, Canchola JA, Catania JA, Mayeda MMS, Dietz EL, Cotto-Negrón C, et al. National-level disparities in internet access among low-income and black and hispanic youth: current population survey. J Med Internet Res 2021;23(10):e27723 [FREE Full text] [doi: 10.2196/27723] [Medline: 34636728]
13. Sieck CJ, Sheon A, Ancker JS, Castek J, Callahan B, Siefer A. Digital inclusion as a social determinant of health. npj Digit Med 2021;4(1):52 [FREE Full text] [doi: 10.1038/s41746-021-00413-8] [Medline: 33731887]
14. Benda NC, Veinot TC, Sieck CJ, Ancker JS. Broadband internet access is a social determinant of health!. Am J Public Health 2020;110(8):1123-1125. [doi: 10.2105/AJPH.2020.305784] [Medline: 32639914]

XSL•FO
RenderX

15. Nkhoma DE, Soko CJ, Bowrin P, Manga YB, Greenfield D, Househ M, et al. Digital interventions self-management education for type 1 and 2 diabetes: a systematic review and meta-analysis. Comput Methods Programs Biomed 2021;210:106370. [doi: 10.1016/j.cmpb.2021.106370] [Medline: 34492544]
16. Versluis A, Boels AM, Huijden MCG, Mijnsbergen MD, Rutten GEHM, Vos RC. Diabetes self-management education and support delivered by mobile health (mHealth) interventions for adults with type 2 diabetes—a systematic review and meta-analysis. Diabet Med 2025;42(5):e70002. [doi: 10.1111/dme.70002] [Medline: 40007461]
17. Huang Z, Soljak M, Boehm BO, Car J. Clinical relevance of smartphone apps for diabetes management: a global overview. Diabetes Metab Res Rev 2018;34(4):e2990 [FREE Full text] [doi: 10.1002/dmrr.2990] [Medline: 29431916]
18. Cai X, Zhang F, Lin C, Zhang X, Wang Z, Xing H, et al. Achieving effective and efficient basal insulin optimal management by using mobile health application (APP) for type 2 diabetes patients in China. Diabetes Metab Syndr Obes 2020;13:1327-1338 [FREE Full text] [doi: 10.2147/DMSO.S244826] [Medline: 32368124]
19. Cui M, Wu X, Mao J, Wang X, Nie M. T2DM self-management via smartphone applications: a systematic review and meta-analysis. PLoS One 2016;11(11):e0166718 [FREE Full text] [doi: 10.1371/journal.pone.0166718] [Medline: 27861583]
20. Bonoto BC, de Araújo VE, Godói IP, de Lemos LLP, Godman B, Bennie M, et al. Efficacy of mobile apps to support the care of patients with diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials. JMIR mHealth uHealth 2017;5(3):e4 [FREE Full text] [doi: 10.2196/mhealth.6309] [Medline: 28249834]
21. Kebede MM, Pischke CR. Popular diabetes apps and the impact of diabetes app use on self-care behaviour: a survey among the digital community of persons with diabetes on social media. Front Endocrinol (Lausanne) 2019;10:135 [FREE Full text] [doi: 10.3389/fendo.2019.00135] [Medline: 30881349]
22. Veazie S, Winchell K, Gilbert J, Paynter R, Ivlev I, Eden KB, et al. Rapid evidence review of mobile applications for self-management of diabetes. J Gen Intern Med 2018;33(7):1167-1176 [FREE Full text] [doi: 10.1007/s11606-018-4410-1] [Medline: 29740786]
23. Hou C, Carter B, Hewitt J, Francisa T, Mayor S. Do mobile phone applications improve glycemic control (HbA1c) in the self-management of diabetes? A systematic review, meta-analysis, and GRADE of 14 randomized trials. Diabetes Care 2016;39(11):2089-2095. [doi: 10.2337/dc16-0346] [Medline: 27926892]
24. Lee JL, Kim Y. Evaluation of mobile applications for patients with diabetes mellitus: a scoping review. Healthcare (Basel) 2024;12(3):368 [FREE Full text] [doi: 10.3390/healthcare12030368] [Medline: 38338253]
25. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
26. Agarwal R, Prasad J. A conceptual and operational definition of personal innovativeness in the domain of information technology. Inform Syst Res 1998;9(2):204-215. [doi: 10.1287/isre.9.2.204]
27. Parasuraman A, Colby CL. An updated and streamlined technology readiness index: TRI 2.0. J Serv Res 2014;18(1):59-74. [doi: 10.1177/1094670514539730]
28. Lam SY, Chiang J, Parasuraman A. The effects of the dimensions of technology readiness on technology acceptance: an empirical analysis. J Interact Mark 2008;22(4):19-39 [FREE Full text] [doi: 10.1002/dir.20119]
29. Bults M, van Leersum CM, Olthuis TJJ, Bekhuis REM, den Ouden MEM. Barriers and drivers regarding the use of mobile health apps among patients with type 2 diabetes mellitus in the Netherlands: explanatory sequential design study. JMIR Diabetes 2022;7(1):e31451 [FREE Full text] [doi: 10.2196/31451] [Medline: 35084357]
30. Alaslawi H, Berrou I, Al Hamid A, Alhuwail D, Aslanpour Z. Diabetes self-management apps: systematic review of adoption determinants and future research agenda. JMIR Diabetes 2022;7(3):e28153 [FREE Full text] [doi: 10.2196/28153] [Medline: 35900826]
31. Böhm AK, Jensen ML, Sørensen MR, Stargardt T. Real-world evidence of user engagement with mobile health for diabetes management: longitudinal observational study. JMIR mHealth uHealth 2020;8(11):e22212 [FREE Full text] [doi: 10.2196/22212] [Medline: 32975198]
32. Faraj S, Ikram RRR, Salahuddin L. Factors influencing patients' adoption and use of mobile health applications for diabetes self-management: a systematic review. Multidiscip Rev 2023;7(2):2024040. [doi: 10.31893/multirev.2024040]
33. Walle AD, Ferede TA, Shibabaw AA, Wubante SM, Guadie HA, Yehula CM, et al. Willingness of diabetes mellitus patients to use mHealth applications and its associated factors for self-care management in a low-income country: an input for digital health implementation. BMJ Health Care Inform 2023;30(1):e100761 [FREE Full text] [doi: 10.1136/bmjhci-2023-100761] [Medline: 37236653]
34. Ory MG, Han G, Nsobundu C, Carpenter K, Towne SD, Smith ML. Comparative effectiveness of diabetes self-management education and support intervention strategies among adults with type 2 diabetes in Texas. Front Public Health 2025;13:1543298 [FREE Full text] [doi: 10.3389/fpubh.2025.1543298] [Medline: 40171438]
35. Peña-Purcell NC, Jay Lee J, Ory MG. Integrating a care supporter in diabetes self-management education to enhance health outcomes: a pilot study. Sci Diabetes Self Manage Care 2025;51(2):156-167. [doi: 10.1177/26350106251320327] [Medline: 40098341]
36. Toobert DJ, Hampson SE, Glasgow RE. The summary of diabetes self-care activities measure: results from 7 studies and a revised scale. Diabetes Care 2000;23(7):943-950. [doi: 10.2337/diacare.23.7.943] [Medline: 10895844]

XSL•FO
RenderX

37. Billmann M, Böhm M, Krcmar H. Use of workplace health promotion apps: analysis of employee log data. Health Policy Technol 2020;9(3):285-293 [FREE Full text] [doi: 10.1016/j.hlpt.2020.06.003]
38. Fisher L, Fortmann AL, Florissi C, Stoner K, Knaebel J, Stuhr A. How frequently and for how long do adults with type 2 diabetes use management apps? The REALL study. J Diabetes Sci Technol 2023;17(2):345-352 [FREE Full text] [doi: 10.1177/19322968211058766] [Medline: 34802281]
39. Berman MA, Guthrie NL, Edwards KL, Appelbaum KJ, Njike VY, Eisenberg DM, et al. Change in glycemic control with use of a digital therapeutic in adults with type 2 diabetes: cohort study. JMIR Diabetes 2018;3(1):e4 [FREE Full text] [doi: 10.2196/diabetes.9591] [Medline: 30291074]
40. Lim SL, Tay MHJ, Ong KW, Johal J, Yap QV, Chan YH, et al. Association between mobile health app engagement and weight loss and glycemic control in adults with type 2 diabetes and prediabetes (D'LITE Study): prospective cohort study. JMIR Diabetes 2022;7(3):e35039 [FREE Full text] [doi: 10.2196/35039] [Medline: 36178718]
41. Vehi J, Isern JR, Parcerisas A, Calm R, Contreras I. Impact of use frequency of a mobile diabetes management app on blood glucose control: evaluation study. JMIR mHealth uHealth 2019;7(3):e11933 [FREE Full text] [doi: 10.2196/11933] [Medline: 30843865]
42. Alenazi HA, Jamal A, Batais MA. Identification of type 2 diabetes management mobile app features and engagement strategies: modified Delphi approach. JMIR mHealth uHealth 2020;8(9):e17083 [FREE Full text] [doi: 10.2196/17083] [Medline: 32678798]
43. Sneha S, Thalla S, Rischie I, Shahriar H. Health internet technology for chronic conditions: review of diabetes management apps. JMIR Diabetes 2021;6(3):e17431 [FREE Full text] [doi: 10.2196/17431] [Medline: 34463627]
44. Blythin A, Elkes J, van Lindholm T, Bhogal A, Wilkinson T, Saville C, et al. Can digital health apps provide patients with support to promote structured diabetes education and ongoing self-management? A real-world evaluation of myDiabetes usage. Digit Health 2023;9:20552076221147109 [FREE Full text] [doi: 10.1177/20552076221147109] [Medline: 36923369]
45. Hawkes RE, Miles LM, Ainsworth B, Ross J, Meacock R, French DP. Engagement with a nationally-implemented digital behaviour change intervention: usage patterns over the 9-month duration of the national health service digital diabetes prevention programme. Internet Interventions 2023;33:100647 [FREE Full text] [doi: 10.1016/j.invent.2023.100647] [Medline: 37502122]
46. Tu YZ, Chang YT, Chiou HY, Lai K. The effects of continuous usage of a diabetes management app on glycemic control in real-world clinical practice: retrospective analysis. J Med Internet Res 2021;23(7):e23227 [FREE Full text] [doi: 10.2196/23227] [Medline: 34264192]
47. Harakeh Z, Van Keulen H, Hogenelst K, Otten W, De Hoogh IM, Van Empelen P. Predictors of the acceptance of an electronic coach targeting self-management of patients with type 2 diabetes: web-based survey. JMIR Form Res 2022;6(8):e34737 [FREE Full text] [doi: 10.2196/34737] [Medline: 35972769]
48. Schretzlmaier P, Hecker A, Ammenwerth E. Extension of the Unified Theory of Acceptance and Use of Technology 2 model for predicting mHealth acceptance using diabetes as an example: a cross-sectional validation study. BMJ Health Care Inform 2022;29(1):e100640. [doi: 10.1136/bmjhci-2022-100640] [Medline: 36379608]
49. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
50. Eaton C, Vallejo N, McDonald X, Wu J, Rodríguez R, Muthusamy N, et al. User engagement with mHealth interventions to promote treatment adherence and self-management in people with chronic health conditions: systematic review. J Med Internet Res 2024;26:e50508 [FREE Full text] [doi: 10.2196/50508] [Medline: 39316431]
51. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
52. Jacobsen LM, Bocchino LE, Lum JW, Kollman C, Barnes-Lomen V, Sulik M, et al. Accuracy of three commercial home-use hemoglobin A1c tests. Diabetes Technol Ther 2022;24(11):789-796. [doi: 10.1089/dia.2022.0187] [Medline: 35763337]

Abbreviations

DSMES: diabetes self-management education and support
FPL: federal poverty level
HbA1c: hemoglobin A1c
IRB: Institutional Review Board
mHealth: mobile health
RCT: randomized controlled trial
SES: socioeconomic status
T2D: type 2 diabetes mellitus
TBES: technology-based education and support delivered through a smartphone app
VIF: variance inflation factor

XSL•FO
RenderX

vMMWD: Virtual Making Moves With Diabetes

Edited by A Stone; submitted 29.Jun.2026; peer-reviewed by A Nguyen, D Dinh; comments to author 15.Sep.2026; revised version received 29.Oct.2026; accepted 29.Oct.2026; published 17.Nov.2026.

Please cite as:
Oware KD, Han G, Towne Jr SDC, Esquivel C, Ahn S, Lee S, Smith ML, Ory MG
Associations Between Technology Attitudes, App Use, and Health Outcomes in a Diabetes Self-Management Education and Support Intervention: Secondary Analysis of a Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e105830
URL: https://mhealth.jmir.org/2026/1/e105830
doi:10.2196/105830
PMID:

©King David Oware, Gang Han, Samuel D Castiglione Towne Jr, Christi Esquivel, SangNam Ahn, Shinduk Lee, Matthew Lee Smith, Marcia G Ory. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 17.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Usage Effects, Effect Modifiers, and Experiences of a Web-Based App for Healthy Habit Formation in Adults: Exploratory Analysis of a Quasi-Experimental Real-World Intervention

Eeva Rantala[1], PhD; Mikko Valtanen[1,2], MSc; Adil Umer[3], MSc; Suvi Parikka[1], PhD; Jussi Pihlajamäki[4,5], Prof Dr Med; Ilona Ruotsalainen[6*], PhD; Jaana Lindström[1,4*], Prof Dr

[1]Department of Public Health, Finnish Institute for Health and Welfare, P.O. Box 30, Helsinki, Finland

[2]Department of Mathematics and Statistics, University of Turku, Turku, Finland

[3]Critical cyber-physical systems team, VTT Technical Research Centre of Finland Ltd., Tampere, Finland

[4]Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland

[5]Department of Medicine, Endocrinology and Clinical Nutrition, Kuopio University Hospital, Kuopio, Finland

[6]Health Data Analytics team, VTT Technical Research Centre of Finland Ltd., Kuopio, Finland

[*]these authors contributed equally

Corresponding Author:
Eeva Rantala, PhD
Department of Public Health, Finnish Institute for Health and Welfare, P.O. Box 30, Helsinki, Finland

Abstract

Background: Digital interventions provide a scalable approach to promote healthy lifestyles, but at-scale evidence of their effects and effect modifiers is limited—even scarcer for habit-based digital interventions, although habits are promising lifestyle intervention targets.

Objective: We conducted a 1-group pre-post effect and effect modification assessment of a web-based app designed to support the formation of health-promoting lifestyle habits by exploring the relationship between objective app use and subjective changes in behavioral outcomes.

Methods: Three-month app access was offered to a subsample of the population-based Healthy Finland survey participants aged 20-74 years via SMS text messaging or mail. The app provided personalized behavioral suggestions that translated evidence-based lifestyle guidelines into simple, repeatable actions (“habits”). Users could browse and select these habits, and report and monitor their performances. The assessment used app log data, questionnaires collected through the app at baseline and at 45 and 90 days, and background information obtained from the national population register and the Healthy Finland Survey. Linear mixed effects models explored the (1) associations between app use and changes in self-reported diet quality (Healthy Diet Index), physical activity (metabolic equivalent of task hours per week), and BMI, and (2) modification of these associations by baseline characteristics related to sociodemographics, health, lifestyle, and e-service use. App use measures comprised the percentage of app use days (ie, days with logins/days of follow-up×100) and the number of reported habit performances (total, diet-related, physical activity–related, and BMI-related).

Results: Of 6975 invitees, 1282 (18.4%) accepted the invitation and 382 (5.5%; mean age 51, SD 15 y; women: n=263, 69%) completed the data collection required for the assessment. Over 90 days, the median percentage of app use days was 5.9% (IQR 3.3% - 11%), and the total number of reported habit performances was 22 (IQR 3 - 68.5). A higher percentage of use days (by 10 percentage points) was associated with a 3.09 (95% CI 0.79 - 5.39) metabolic equivalent of task hours per week greater increase in physical activity and a 1-unit higher logarithmic number of reported performances with a 0.51 (95% CI 0.07 - 0.95) Healthy Diet Index point greater improvement in diet quality. Other associations between app use and outcomes were nonsignificant. Greater physical activity and a more positive attitude to e-services at baseline appeared to enhance the effect of app use on physical activity (*P* values <.05). The app received a median acceptability rating of 2.9 (IQR 2.4 - 3.3, scale: 1 - 4) and a median overall score of 7 (IQR 5 - 8, scale: 0 - 10).

Conclusions: An inexpensive, low-intensity digital support for healthy habits could foster small beneficial lifestyle changes, but effects require user engagement and may depend on the individual. The findings warrant confirmation in more robust study designs and call for further research for identifying those most likely to benefit from such digital support.

(JMIR Mhealth Uhealth 2026;14:e84076) doi:10.2196/84076

XSL•FO
RenderX

KEYWORDS

web-based; app; lifestyle; habit; behavior change; diet; physical activity; BMI; health promotion; prevention

Introduction

Digital lifestyle interventions are considered a relatively inexpensive and scalable way to support the adoption and maintenance of healthy lifestyle behaviors. They enable a delivery independent of time, place, and health professionals and provide a promising avenue to increase the availability and accessibility of health services without additional burden on the health care system. Emerging evidence also suggests that digital lifestyle interventions can save societal and health care costs [1,2]. They thus appear to be an attractive solution for aging societies that face an increasing need for care while under pressure to reduce health care expenses.

While literature reviews have drawn varying conclusions on the overall effectiveness of digital lifestyle interventions, they have demonstrated that diverse digital tools have the potential to improve health behaviors and other health outcomes among various adult populations [3-10]. For example, a recent umbrella review and meta-meta-analysis of 47 systematic reviews covering more than 500 unique randomized controlled trials (RCTs) and more than 200,000 participants found grade A–level evidence that digital lifestyle interventions (mobile apps, web-based, SMS text messaging, or mixed) yield, on average, small to medium favorable effects on dietary outcomes (fruit, vegetable, and energy intake), physical activity (total, moderate to vigorous, and steps per day), and body weight when used predominantly as stand-alone treatments [4].

However, lifestyle change can be promoted in a multitude of ways. Habit-based approaches focus on the repetition of a specific behavior in a consistent context [11,12]. These approaches promote the formation of habitual behaviors that occur automatically as a response to specific contextual cues—even without conscious deliberation, motivation, or awareness [11-13]. Supporting habit formation has the potential to result in lasting lifestyle change, which sustains even when motivation drops [12]. An umbrella review with meta-meta-analyses of behavior change interventions across behavioral domains found habits to be the most potential targets of individual-level interventions, compared with knowledge, beliefs, attitudes, skills, and emotions [14]. For weight loss, nondigital habit-based lifestyle interventions have yielded evidence of short-term effectiveness compared to no treatment or usual care among individuals with overweight or obesity [15].

Nevertheless, the habit-based approach has not yet been widely adopted in digital lifestyle interventions. Furthermore, the few digital solutions that have used a habit-based approach have mainly been studied in specific population groups or in small-scale trials. Examples include the HabitWalk app for increased brisk walking that was tested in a 15-week microrandomized trial among inactive adults [16] and the Top Tips app for weight loss that was piloted in a 3-month RCT among adults with overweight or obesity [17]. Another example was the BitHabit app that was developed for the prevention of type 2 diabetes (T2D) and tested in a 12-month RCT among adults with T2D risk factors [18,19]. While the HabitWalk study demonstrated that the repetition of a target behavior in response to a specific cue is positively associated with habit strength [16], both the Top Tips and BitHabit apps showed the greatest effects among study participants who engaged with the apps more frequently [17,20]. These results indicate that habit-based lifestyle apps may be beneficial as long as they are used and highlight the importance of incorporating use metrics in their effect assessments. Sustained engagement remains a key challenge in digital lifestyle interventions [3,4,16,17,21,22]. The learnings of habit-based and other lifestyle apps suggest that personalization could enhance engagement [17,23,24].

Overall, limited evidence exists of the effects of digital lifestyle interventions—habit-based and other—among the general adult population and thus the potential of these interventions to promote public health at scale. Similarly, evidence of individual factors that may modify the effects of digital lifestyle interventions is limited and inconclusive [3,4,7,23,25]. Studying such effect modifiers is a prerequisite for identifying population groups that could benefit from exclusively digital interventions and groups that would need other types of support, such as in-person counseling.

This study addresses the knowledge gaps mentioned earlier regarding the effects and effect modifiers of digital lifestyle interventions intended to support the formation of health-promoting habits among the general adult population. Simultaneously, the study serves as a real-world example of scaling up a habit-based lifestyle app originally developed for individuals at risk of lifestyle-related chronic noncommunicable diseases to a more heterogeneous adult population. The study used a quasi-experimental pre-post design and deployed the app as a stand-alone support for self-directed habit formation. The overall aim was to conduct an exploratory assessment of the effects of the app use on behavioral outcomes and to identify groups most likely to benefit from using the app. The specific aims were to examine (1) associations between objectively measured app use and changes in self-reported diet quality, physical activity, and BMI over 3 months; (2) the modification of these associations by the baseline characteristics of the participants; and (3) the participants' experiences (ie, acceptability and overall evaluation) of the app.

Methods

Study Design

This study followed a quasi-experimental 1-group pretest-posttest design in which the participants were followed up for 3 months via the app using continuous app log data and questionnaires conducted at baseline, midway through (approximately 45 days since the baseline), and at the end of the follow-up (approximately 90 days since the baseline). The effects and effect modifiers of the app were examined based on an exploratory analysis of the relationship between objectively measured app use and changes in self-reported behavioral

XSL•FO
RenderX

outcomes. Owing to the correlational nature of the assessment, the reporting of the study follows the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guideline for cohort studies [26,27] (for a completed checklist, see Checklist 1).

Ethical Considerations

The research plan of the study was approved by the Institutional Review Board of the Finnish Institute for Health and Welfare (THL/5335/6.02.01/2022). The study participants gave their informed consent to participate. Participation was free of charge, and participants were not compensated.

Participant Recruitment

This study was a substudy of a population-based health survey, Healthy Finland, that targeted a randomly selected and nationally representative sample of adults in Finland [28]. Participants of the Healthy Finland survey were eligible for this study if they (1) spoke Finnish fluently, (2) were aged 20 to 74 years, (3) had completed the questionnaire of the Healthy Finland survey between September and the end of the year 2022, and (4) were not randomly selected for an additional health examination component of the Healthy Finland survey. The latter criterion was to avoid influencing the health examination results and to avoid burdening the participants excessively. Participation in this study also required access to the internet via a smartphone, tablet, or computer.

Among the eligible population (N=10,207), individuals with a known phone number (n=4978, 48.8%) were sent a personal invitation to participate in this study in February 2023 via SMS text message. From the remaining eligible population, 2000 individuals (19.6% of the 10,207 eligible people) were sampled based on 5-year age groups to receive the invitation via letter. The sample size was determined for the assessment of the uptake of the app [29,30] and estimated to be sufficient to detect small effects based on logistic regression with a statistical power of 0.8 and a type I error of 0.05, assuming that 10% to 15% of the invitees would choose to participate. The invitation was valid until the beginning of March 2023. The invitation directed the participant to a website of the Finnish Institute for Health and Welfare, which provided information on the study and the app, a link to the app, and instructions on how to start using the app. Of the 6975 invitees (excluding 3 who asked for their data to be removed), 1282 (18.4%) accepted the invitation, agreed to the terms of the app, and registered for the app with their phone number and a user ID that was featured in the invitation [29,30].

The App

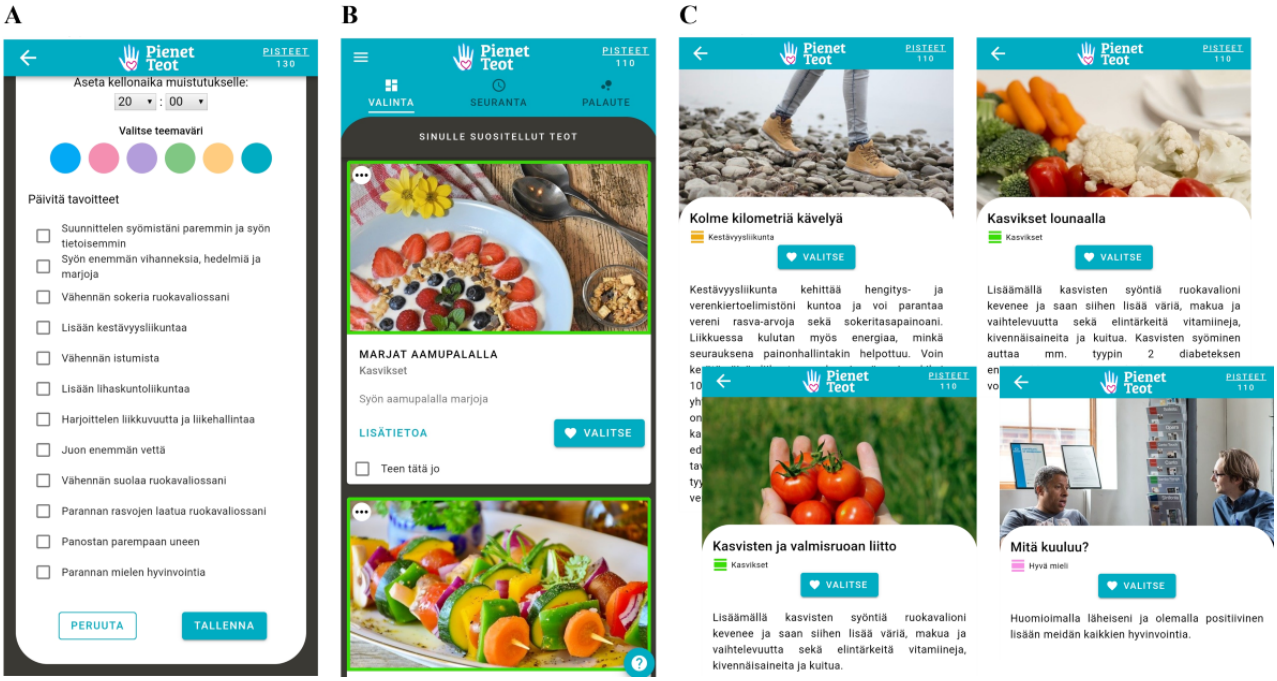
The app used in this study was web-based and closed (ie, not open access) and could be used with a smartphone, tablet, or computer with no need to download the app on the device. The app was based on the BitHabit app that was developed in a T2D prevention study, Stop Diabetes (StopDia) [18,19,31], which involved the authors of this study and/or their affiliated research institutes. The BitHabit app was designed to support the adoption of lifestyle habits that promote health and aid the prevention of T2D and other lifestyle-related chronic diseases [19,20]. The BitHabit app was built on the principles of habit formation [32-34] and self-determination [35,36] theories. Habit formation approaches promote the repetition of a target behavior in the presence of consistent contextual cues until the cues begin to trigger the behavior automatically [32-34]. Self-determination theory emphasizes the role of autonomous motivation, perceived competence, and a sense of community in lifestyle changes and fosters the fulfillment of basic psychological needs—the perception of autonomy, the sense of control or self-efficacy, and the feeling of relatedness with peers or significant others [35,36]. After the StopDia study, the BitHabit app was updated by improving its visual appearance and by adding features for personalization (habit suggestions based on a background questionnaire) and for social support (a researcher-moderated discussion forum). This updated version was used in this study and is referred to as “the app” because the BitHabit brand has been out-licensed.

The app used in this study provided an “online store” or “library” of approximately 300 behavioral suggestions that translated lifestyle guidelines [37-41] into concrete actions. These actions were simple, specific, contextualized, and easily repeatable, and they were presented with a brief description of how they promote health. The actions were tied to and could be triggered by daily life activities, which made them optimal for habit formation [19]. The actions were organized into 19 habit categories related to eating and drinking (10 categories: mindful eating, meal pattern, vegetarian and fish courses, fruits and vegetables, grain products, dietary fats, sugar, salt, alcohol, and drinking water), physical activity (5 categories: aerobic exercise, muscle strengthening exercise, mobility and balance activities, stretching one’s legs, and everyday physical activity), sleep, positive mood, daily life management, and smoking.

On their first login to the app, the participants of this study gave their informed consent and completed a brief questionnaire on personal characteristics (eg, age, sex, and employment or study situation), health behavior (dietary habits, alcohol consumption, physical activity, sleep, and smoking), perceived stress, and well-being goals. On the basis of the participant’s responses to the questionnaire, the app determined the habit categories with the greatest potential for improvement and suggested the categories accordingly. After this, the participant could enter the app to browse the habit library; to inspect, select, and report actions; and to monitor and reflect on the actions performed. In addition, the app allowed setting reminders, enabled modifying the reported well-being goals, provided summary data on actions performed by other users, featured a researcher-moderated forum for discussion and questions, and provided information on health-promoting lifestyles. If the participant did not use the app for over a week, they automatically received up to 3 SMS text message reminders with 1-week intervals. No adjustments were made to the content or functionalities of the app during the study follow-up. Figure 1 illustrates the visual appearance and functionalities of the app.

XSL•FO
RenderX

Figure 1. Views of the app that (A) allowed setting well-being goals, reminders, and a theme color; (B) provided habit suggestions based on the goals set; and (C) described specific habits.



Measures

Participant Characteristics

The participants' age and sex (male or female, as assigned at birth) were retrieved from the Finnish National Population Register. Information on socioeconomic status, health, lifestyles, and the use of e-services was obtained from the questionnaire of the Healthy Finland survey, which the participants had completed a median of 148 (range 47 - 231) days before the baseline of this study.

Information on socioeconomic status comprised education (years of full-time schooling, including primary and comprehensive school), household income (euros per month before taxes) as a 5-category variable, and principal life situation (employed or studying, on family leave or a stay-at-home parent, retired, unemployed, or other). Health-related information comprised perceived health status (poor or fairly poor vs average, fairly good, or good) and perceived cognitive function (average of 3 items: the working of memory, ability to learn new things, and ability to concentrate). Lifestyle-related information comprised getting enough sleep (rarely or hardly ever vs almost always, often, or unsure) and daily smoking or use of other tobacco products. Information on the use of e-services comprised perceived competence (low or no vs moderate, high, or very high); perceived barriers and concerns (average of 6 items related to diverse challenges in using e-services); and perceived benefits (average of 6 items related to diverse benefits of electronic social and health care services). The scoring of the variables reflecting perceived cognitive function, barriers or concerns, and benefits of e-services followed principles used in related research [42,43] and ranged from 1 to 5, with greater values reflecting better cognitive function, greater barriers or concerns, and greater perceived benefits. For original questionnaire items, see Multimedia Appendix 1.

App Use

Data on the use of the app accumulated automatically throughout the study via log files that included time-stamped user interactions with the app. App log data were considered from completion of the baseline questionnaire to the completion of the questionnaire at the end of the follow-up or, if not completed, 90 days since the baseline. The app use metrics used in the effects assessment were the percentage of use days, the total number of reported habit performances, and the number of reported habit performances targeting each primary outcome (diet quality, physical activity, and BMI; Table 1). Similar use metrics that are based on logins and activities performed have been found to be positively associated with favorable changes in behavioral and anthropometric outcomes in digital lifestyle interventions [44], including the StopDia study [45]. Reported performances related to eating or drinking were considered to target diet quality, performances related to physical activity or sedentary behavior were considered to target total physical activity, and performances related to eating, drinking, physical activity, or sedentary behavior were considered to target BMI. In addition, to describe the overall app use over the follow-up, we computed the percentage of active users per week, the number of use days, the number of habit categories with reported performances, and the number of specific habits with reported performances (Table 1).

XSL•FO
RenderX

Table . Use metrics derived from the app log files.

Use metric	Definition
Active users (% per week)	Users with ≥1 login/all users×100
Use days	Number of days with logins
% Use days	Number of days with logins/number of follow-up days×100
N categories with performances	Number of habit categories with reported performances
N habits with performances	Number of specific habits with reported performances
N performances[a]	Number of reported habit performances
Total	Total
Targeting diet quality	Related to eating or drinking
Targeting physical activity	Related to physical activity or sedentary behavior
Targeting BMI	Related to eating, drinking, physical activity, or sedentary behavior

[a]A specific habit could be marked as performed a maximum of once per day.

Primary Outcomes

The primary outcomes comprised self-reported diet quality, total physical activity, and BMI. These data were collected with questionnaires via the app at the start of the follow-up (defined as the completion of the baseline questionnaire), midway through the follow-up (approximately 45 days since the baseline), and at the end of the follow-up (approximately 90 days since the baseline). Total physical activity and BMI were additionally measured in the questionnaire of the Healthy Finland survey that was completed pre-baseline [28], and for comparability, we used the same questionnaire items (Multimedia Appendix 1).

Diet quality was measured with the Healthy Diet Index (HDI) [46], which reflects adherence to the Nordic and Finnish food-based dietary guidelines on a scale ranging from 0 (lowest diet quality) to 100 (highest diet quality). The HDI builds on a food frequency questionnaire that covers 7 dietary domains (meal pattern; grains; fruit and vegetables; fats; fish and meat; dairy; and snacks and treats, including beverages) [46]. Total physical activity covered weekly hours of low-, moderate-, and high-intensity physical exertion. To capture potential changes in both the amount and intensity of total physical activity per week, the time spent at each intensity level was weighted with its approximate metabolic equivalent of task (MET): 2.5 for low-intensity, 5 for moderate-intensity, and 10 for high-intensity exertion. The cutoffs were determined based on the description of each intensity level in the questionnaire items, corresponding MET values reported in the literature [47,48], and consultation with an expert on sports science. The weighted estimates were totaled to reach total MET hours per week. With this weighting, 12.5 MET hours per week met the current recommendation for adults to accumulate at least 2.5 hours of moderate-intensity or 1.25 hours of high-intensity aerobic physical activity per week, or an equivalent combination of both [49,50]. BMI (kg/m^2) was computed based on self-reported height and weight. While self-reported BMI is generally lower than measured BMI due to the underestimation of weight and overestimation of height, self-reported BMI tends to correlate with measured BMI over 90% and is considered a valid measure across a range of sociodemographic groups [51-53].

App Use Experience

User experiences were collected with a questionnaire via the app at the end of the follow-up. Questionnaire items reflected the acceptability and overall evaluation of the app (for original questionnaire items, see Multimedia Appendix 1). The mean acceptance of the app was computed by averaging the ratings (scale: 1 - 4, higher value indicating greater acceptance) the participants gave to 7 questionnaire items that reflected the 7 component constructs of the Theoretical Framework of Acceptability [54]. These constructs comprise affective attitude (how the participant feels about taking part in the intervention), burden (the perceived amount of effort that is required to participate in the intervention), perceived effectiveness (the extent to which the intervention is perceived as likely to achieve its purpose), opportunity costs (the extent to which benefits, profits, or values must be given up to engage in the intervention), intervention coherence (the extent to which the participant understands the intervention and how the intervention works), self-efficacy (the participant's confidence that they can perform the behaviors required to participate in the intervention), and ethicality (the extent to which the intervention has a good fit with the participant's value system) [54].

The overall evaluation of the app was computed by averaging the ratings (scale: 0 - 10, higher value indicating a more positive evaluation) the participants gave to 4 questionnaire items that asked them to evaluate the behavioral suggestions that the app offered, the app as a whole, the likelihood to use the app in the future, and the likelihood to recommend the app to a friend or a colleague. The latter item was also used to compute the net promoter score by categorizing the participants into detractors (rating ≤6), passively satisfied (rating=7 - 8), and promoters (rating ≥9), and then subtracting the percentage of detractors from the percentage of promoters [55].

Statistical Analyses

The effects assessment included participants who completed the baseline questionnaire and at least one questionnaire thereafter; that is, the questionnaire midway through and/or at

XSL•FO
RenderX

the end of the follow-up. We call these participants the "completers" of the study. Linear mixed effects models were used to estimate (1) the associations between app use and the slopes of the primary outcomes (diet quality, total physical activity, and BMI), that is, the interaction of elapsed time and an app use metric; and (2) the modification of these associations by the participants' baseline characteristics. The app use metrics were evaluated at the time of the completion of each questionnaire. The models related to diet quality and BMI were fitted with individual-level random intercepts and slopes. The models related to total physical activity were fitted with individual-level random intercepts only, as the inclusion of random slopes resulted in model convergence issues.

Separate models were fitted for each outcome–app use metric pair. The app use metrics reflecting the number of reported performances were log-transformed. The model coefficients corresponding to the log-transformed numbers of performances thus reflect the expected difference in the outcome corresponding to a 2.7-fold increase in the number of performances. All models were adjusted for the following baseline characteristics potentially influencing the app use and its effects [20,23,25,29]: age, sex, education, perceived health status, BMI, perceived cognitive function, diet quality, and total physical activity, as well as the competence, barriers or concerns, and perceived benefits of using e-services. Effect modification by the baseline characteristics was assessed by estimating the 3-way interaction parameters between each characteristic, app use metric, and elapsed time. The 3-way interaction reflected how the characteristic changed the association between the use metric and the slope of the outcome.

Regarding total physical activity, values >112 hours per week (ie, on average <8 hours per day for sedentary activities and rest) were excluded as implausible. Regarding diet quality and BMI, no implausible values were detected. To assess the sensitivity against large changes in the primary outcomes due to possible misreporting, the main models were fitted also by removing individuals with the largest changes reported between 2 successive measurements. The changes between successive measurements were standardized (ie, by subtracting the mean and dividing by the SD), and the sensitivity analyses were conducted by first removing individuals with the largest standardized change >5, and then those with the largest standardized change >3.

Analyses were performed with the R software (version 4.5.0; R Foundation for Statistical Computing [56]), using the package *lme4* [57] to fit the linear mixed effects models. Statistical significance was assessed with a likelihood ratio test, with a predetermined α level of .05. Data points with missing values were excluded from the analyses, and the data were assumed to be missing at random and that loss to follow-up was noninformative. As the analyses were of an exploratory nature, corrections for an increased type I error rate were not applied.

Results

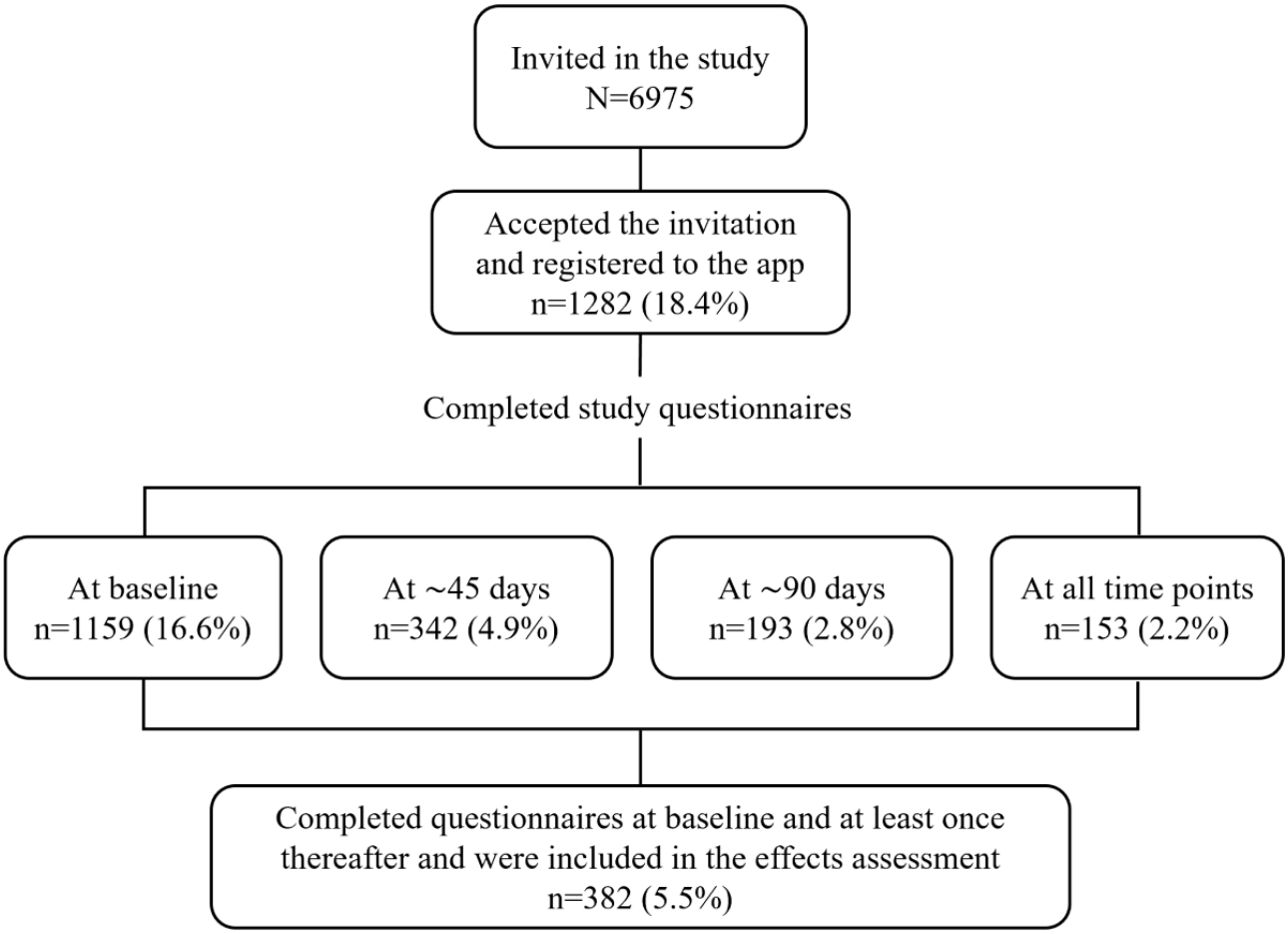
Participant Characteristics

Of the 1282 individuals who accepted the study invitation and registered for the app (18.4% of invitees), 1159 (16.6%) completed the baseline questionnaire and 382 (5.5%) completed at least one questionnaire thereafter ("the completers"; Figure 2). The reasons for noncompletion are not available.

Among the completers, the proportion of men was 31% (119/382), age ranged from 20 to 74 (mean 51.3, SD 15.4) years, and the duration of full-time education ranged from 3 to 25 (mean 14.9, SD 3.4) years (Table 2). Most completers (280/382, 73%) belonged to the 3 higher income categories (household gross income approximately €2916 - €4580, or US $2953-US $4639, per month or more), and the majority were either employed or studying (218/382, 57%) or retired (134/382, 35%). A minority reported poor perceived health status (27/382, 7%), insufficient sleep (80/382, 21%), daily use of tobacco products (25/382, 7%), or poor competence to use e-services (7/382, 2%). On a scale ranging from 1 (poorer) to 5 (greater), the completers perceived their cognitive function to be on average 4.0 (SD 0.7), the extent to which they experienced barriers or concerns related to using e-services was on average 2.2 (SD 0.7), and the extent to which they perceived a benefit from e-services was on average 3.95 (SD 0.7).

XSL•FO
RenderX

Figure 2. Formation of the study sample from invitees to completers (% invitees).



XSL•FO
RenderX

Table . Participant characteristics at baseline.

Characteristic	Completers (n=382)[a]	Noncompleters (n=777)[b]	*P* value[c]
Sociodemographic characteristics			
Age (y), mean (SD)	51.3 (15.4)	49.9 (15.2)	.11
Sex (female), n (%)	263 (69)	516 (66)	.44
Education (y), mean (SD)	14.9 (3.4)	15.5 (3.28)	.004
Household income[d] (per month), n (%)			.67
<€1250 (<US $1266)	30 (8)	55 (7)	
€1251 - €2915 (US $1267-US $2952)	69 (18)	157 (20)	
€2916 - €4580 (US $2953-US $4639)	120 (32)	222 (29)	
€4581 - €6250 (US $4640-US $6330)	83 (22)	162 (21)	
>€6251 (>US $6331)	77 (20)	176 (23)	
Principal life situation, n (%)			.03
Employed or studying	218 (57)	497 (64)	
Family leave or stay-at-home parent	4 (1)	18 (2)	
Retired	134 (35)	206 (27)	
Unemployed	14 (4)	31 (4)	
Other	12 (3)	20 (3)	
Health			
Perceived health status (fairly poor or poor), n (%)	27 (7)	51 (7)	.85
BMI (kg/m^2), mean (SD)	27.6 (5.3)	27.9 (5.7)	.67
Perceived cognitive function (scale: 1 - 5), mean (SD)	4.0 (0.7)	4.0 (0.7)	.55
Lifestyles			
Healthy Diet Index (scale: 0 - 100), mean (SD)	54.4 (11.0)	52.5 (10.7)	.005
Total physical activity (MET[e] hours per week), median (IQR)	30.0 (15.0 - 50.0)	30.0 (16.1 - 52.5)	.55
Enough sleep (rarely or hardly ever), n (%)	80 (21)	166 (21)	.98
Daily use of tobacco products, n (%)	25 (7)	64 (8)	.39
e-Service use			
Competence (low or no), n (%)	7 (2)	26 (3)	.21
Barriers and concerns (scale: 1 - 5), mean (SD)	2.20 (0.73)	2.20 (0.74)	.86
Perceived benefits (scale: 1 - 5), mean (SD)	3.95 (0.74)	3.89 (0.80)	.44

[a]Missing data: 0% - 2%.

[b]Missing data: 0% - 1%.

[c]Chi-square test for categorical and Mann-Whitney *U* test for continuous variables.

[d]Using a conversion rate of €1=US $1.0128, which was the average exchange rate during the data collection between September and December 2022.

[e]MET: metabolic equivalent of task.

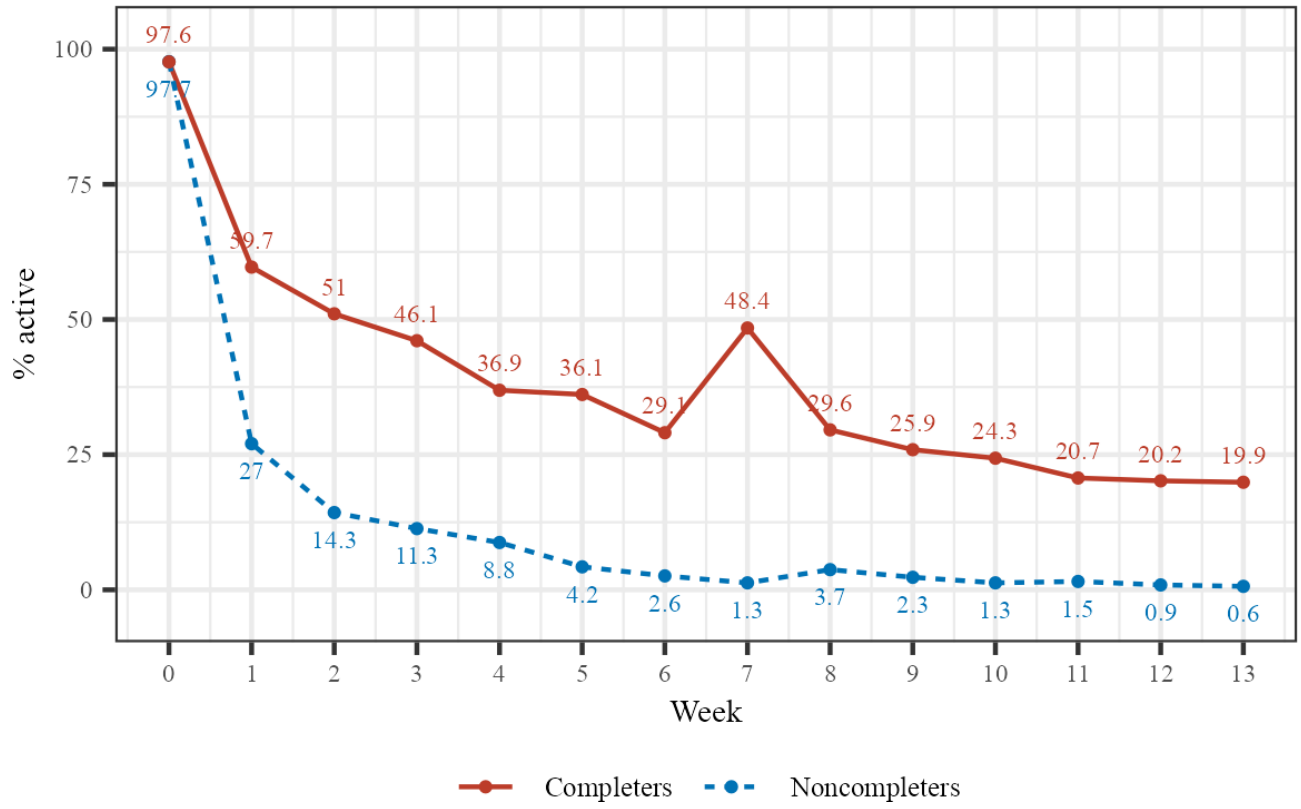
XSL•FO
RenderX

Compared to the noncompleters (n=777), the completers had, on average, fewer years of education (14.9 vs 15.5 years; *P*=.004) and better diet quality (HDI 54.4 vs 52.5; *P*=.005; Table 2). The completers also differed from the noncompleters in their principal life situation (*P*=.03), with a greater proportion being retired (35% vs 27%) and a smaller proportion being employed or studying (57% vs 64%).

App Use

Among the completers, the percentage of active app users dropped from 97.6% (373/382) at the beginning of the follow-up to 29.1% (111/382) by the 6th week of the follow-up and to 19.9% (76/382) by the 13th week of the follow-up (Figure 3).

Figure 3. Percentage of active app users (≥1 login) per week over the 90-day follow-up.



Over the entire follow-up, the completers used the app on a median of 5.5 (IQR 3 - 10) days (Table 3). This equaled a median of 6% (IQR 3% - 11%) of the follow-up days. The completers reported performances in altogether 14 (74%) of the 19 habit categories. These 14 categories were related to physical activity (aerobic exercise, muscle strengthening exercise, mobility and balance activities, and stretching one's legs), eating and drinking (alcohol, mindful eating, drinking water, sugar, fruits and vegetables, dietary fats, salt, and meal pattern), sleep, and positive mood. The median number of habit categories with reported performances was 2 (IQR 1 - 3), while the median number of specific habits with reported performances was 5 (IQR 2 - 7). The median number of total reported performances was 22 (IQR 3 - 68.5). Across the app use metrics examined, the app use was significantly more frequent among the completers than among the noncompleters (*P* values <.001).

XSL•FO
RenderX

Table . App use during the 90-day follow-up.

App use metric	Completers (n=382), median (IQR)	Noncompleters (n=777), median (IQR)	*P* value[a]
Use days	5.5 (3-10)	1 (1-2)	<.001
% Use days	5.9 (3.3 - 11)	1.1 (1.1 - 2.2)	<.001
N categories with performances	2 (1-3)	0 (0 - 1)	<.001
N habits with performances	5 (2-7)	0 (0 - 2)	<.001
N performances			
Total	22 (3 - 68.5)	0 (0 - 2)	<.001
Targeting diet quality	3 (0 - 19)	0 (0 - 0)	<.001
Targeting physical activity	4 (0 - 25.75)	0 (0 - 0)	<.001
Targeting BMI	13 (2 - 49)	0 (0 - 1)	<.001

[a]Mann-Whitney *U* test.

Effect of App Use on Changes in Primary Outcomes

The completers' mean diet quality was 54.4 (SD 11) HDI points out of 100 at baseline, rose by 1.7 points to 56.1 (SD 10.8) by the midpoint of the follow-up, and dropped by 1.5 points to 54.6 (SD 9.8) by the end of the follow-up (Table S1 in Multimedia Appendix 1). The median total physical activity was 30 (IQR 15 - 50) MET hours per week at baseline, decreased by 1.25 units to 28.75 (IQR 17.5 - 50) by the midpoint, and returned to 30 (IQR 17.5 - 48.75) by the end of the follow-up. The mean BMI was 27.6 (SD 5.3) kg/m^2 at baseline, decreased by 0.1 units to 27.5 (SD 5.2) kg/m^2 by the midpoint, and returned to 27.6 (SD 5.7) kg/m^2 by the end of the follow-up.

When estimating the effect of app use on changes in the primary outcomes over the 90-day follow-up, the results differed depending on the app use metric and the outcome. A higher percentage of app use days (by 10 percentage points) was associated with a 3.09 (95% CI 0.79 - 5.39) MET hours per week greater change in total physical activity (Table 4). The percentage of use days was, however, not significantly associated with changes in diet quality or BMI. In contrast, a 1-unit higher logarithmic number of total reported habit performances (ie, 2.7-fold higher on the absolute scale) was associated with a 0.51 (95% CI 0.07-0.95) point greater change in diet quality. The total number of reported performances was, however, not significantly associated with changes in total physical activity or BMI. No significant associations were found between the number of reported performances targeting each outcome and the changes observed in the respective outcomes. For unadjusted effect estimates, see Table S2 in Multimedia Appendix 1.

Table . Effect of app use on changes in primary outcomes.

App use metric	Primary outcomes[a]					
	HDI (n=368)		TPA (n=362)		BMI (n=368)	
	β (95% CI)	*P* value	β (95% CI)	*P* value	β (95% CI)	*P* value
% Use days	.58 (–0.07 to 1.22)	.08	3.09 (0.79 to 5.39)	.01	.00 (–0.09 to 0.08)	.95
N performances[b]						
Total	.51 (0.07 to 0.95)	.02	1.30 (–0.30 to 2.89)	.11	–0.01 (–0.07 to 0.04)	.64
Targeting the outcome	.16 (–0.34 to 0.65)	.54	.62 (–1.02 to 2.2)	.47	–0.03 (–0.08 to 0.03)	.40

[a]Parameter estimates (β), 95% CI, and significance (*P*) of the effect of app use on Healthy Diet Index (HDI), total physical activity (TPA), and BMI. The estimates describe how changes in the outcomes over 90 days differ corresponding to a unit difference in the app use metric.

[b]Log-transformed.

The results remained predominantly robust in sensitivity analyses (Table S3 in Multimedia Appendix 1). The most notable changes were a slight attenuation of the estimated effect of the percentage of use days on physical activity (β=2.33, 95% CI 0.20-4.44; *P*=.03) and a slight strengthening of the estimated effect of the total number of performances on physical activity (β=1.71, 95% CI 0.29-3.10; *P*=.02) when removing individuals with the largest standardized change >3. In addition, the effect estimate for each app use metric on BMI was slightly stronger under the sensitivity analyses.

Effect Modification by Baseline Characteristics

Greater physical activity at baseline appeared to enhance the effect of the app use on total physical activity regardless of the app use metric used: the percentage of app use days (β=.13, 95% CI 0.06-0.2), the total number of reported performances

XSL•FO
RenderX

(β=.05, 95% CI 0.01-0.10), or the number of reported performances targeting physical activity (β=.05, 95% CI 0.00-0.09; Tables S4-S6 in Multimedia Appendix 1). This would mean that, in increasing movement, individuals who are physically more active at baseline benefit from using the app more than individuals who are physically less active.

Additionally, a greater extent to which the participants perceived to benefit from using e-services at baseline appeared to enhance the effect of the percentage of app use days (β=4.34, 95% CI 1.11-7.60) and the number of reported performances targeting physical activity (β=2.13, 95% CI 0.10-4.13) on total physical activity (Tables S4 and S6 in Multimedia Appendix 1). This would mean that, in increasing movement, individuals with more positive attitudes toward e-services at baseline benefit from using the app more than individuals with less positive attitudes. No baseline characteristic was found to modify the effect of the app use on diet quality or BMI (Tables S4-S6 in Multimedia Appendix 1).

App Use Experience

The participants who completed the user experience questionnaire at the end of the study gave the app a median overall acceptability score of 2.9 (IQR 2.4 - 3.3) on a scale ranging from 1 to 4, where a higher value indicated greater acceptance (Table 5). The participants gave both the app and the behaviors the app suggested a median overall score of 7 (IQRs 5 - 8 and 6 - 8, respectively) on a scale ranging from 0 to 10. Both the median likelihood to use the app in the future and to recommend the app to others was 5 (IQR 1 - 8).

Table . App use experiences at the end of the study (n=199).

User experience metric	Values
Acceptability (scale: 1 - 4[a]), median (IQR)	
Affective attitude	3 (2-3)
Burden	3 (2-3)
Perceived effectiveness	3 (2-3)
Opportunity costs	3 (2-4)
Intervention coherence	3 (2 - 3.5)
Self-efficacy	3 (2-3)
Ethicality	3 (3-4)
Mean of all acceptability domains	2.86 (2.43 - 3.29)
Overall evaluation (scale: 0 - 10), median (IQR)	
The behaviors the app suggested	7 (6-8)
The app as a whole	7 (5-8)
Likelihood to use the app in the future	5 (1-8)
Likelihood to recommend the app to a friend or colleague	5 (1-8)
Promoters (rating 9 - 10), n (%)	30 (15)
Passively satisfied (rating 7 - 8), n (%)	43 (22)
Detractors (rating 0 - 6), n (%)	126 (63)
Net promoter score (% promoters−% detractors)	–48

[a]Measures informed by the Theoretical Framework of Acceptability [54].

Discussion

Principal Findings

We assessed the effects and effect modifiers of the use of a habit-based healthy lifestyle app on self-reported diet quality, total physical activity, and BMI in a subsample of population-based health survey participants in Finland. The assessment was exploratory and correlational in nature, building on a quasi-experimental 1-group pre-post intervention. Over 3 months, the study participants used the app on a median of 5.9% (IQR 3%-11%) of days and reported a median of 22 (IQR 3-68.5) total habit performances. The percentage of use days was associated with increased physical activity, and the number of reported performances was associated with improved diet. Some evidence also suggested that greater baseline physical activity and a positive attitude toward e-services could enhance the app use effect on physical activity. Overall, the app was accepted and received moderate ratings. However, the likelihood at which the participants considered using the app in the future and recommending the app to others indicated that there is room to improve the app. The study contributes to the scant evidence base of the effects and factors that influence the effects of digital lifestyle interventions that use a habit-based approach, in a broadly recruited sample of adults. Moreover, the work provides a real-world example of scaling up a digital lifestyle intervention previously studied among individuals at risk of chronic diseases to a more heterogeneous audience.

XSL•FO
RenderX

Comparison With Prior Research

App Use

With a median of 5.5 (IQR 3 - 10) objectively measured app use days (5.9%, IQR 3% - 11% of the follow-up period) and 22 (IQR 3 - 68.5) self-reported habit performances accumulated during the 3-month follow-up, many study participants were unlikely to succeed in forming new habits, particularly because the habit performances were divided into a median of 5 diverse behaviors. The minor changes observed in self-reported diet quality, physical activity, and BMI seem hence credible. While limited evidence exists of the time and number of repetitions needed to reach habit automaticity, a median of 66 (IQR 39-102) days was reported in a pioneering study that tracked the habit formation process among university students who performed a self-selected health-promoting action once per day in response to a stable cue [34]. The time needed for the habit strength to peak varies substantially, however, as the trajectory of habit formation is highly idiographic and often nonlinear [16]. Between-participant variation has been reported to range from 18 to 254 days [34,58]. Despite this variability, prior habit formation interventions have adopted a 3-month duration, considering it a sufficient period for a habit to develop—at least with daily repetition of the target behavior [15,17].

The overall low app use in this study suggests that the app was insufficiently engaging for many participants and may have failed to support them in repeating the selected behaviors. A voluntary habit-based behavior change requires sustained motivation and self-control to repeat the target behavior sufficiently for a habit to form [12]. To achieve this, people should be supported particularly during the early phases of their habit formation attempts because at this point, the emerging habit strengthens most quickly, but people are most likely to give the change attempt up if unsatisfied with initial experiences [12]. Common ways to support the continuation of the habit formation process include emphasizing the importance of the new behavior and training people in the skills needed to act in the chosen setting, for example, by making detailed action plans for the intended behavior and coping plans to overcome unintended behaviors [12]. Ideally, the support would be responsive to initial experiences and, as needed, assist in refining the set goals and methods chosen to pursue them [12].

The percentage of active app users dropped more quickly, and the app use frequency remained lower overall in this study than in the 12-month T2D prevention trial, StopDia, in which the BitHabit app was used by adults at risk of developing T2D [18,19]. During the first 6 months of StopDia, the percentage of active app users remained above 50%, the app was used on a median of 23 (IQR 12 - 42) days, and the users reported a median of 263 (IQR 60 - 703) habit performances [19]. In comparison, the Top Tips habit-based weight loss app reached an average of 25 (SD 44) logins and 10 (SD 21) completed tips during a 3-month pilot trial among individuals with overweight or obesity [17]. These differences in app use might be explained by the different target populations of the studies. While our study targeted the general adult population, the individuals who volunteered to participate tended to be healthier and have healthier lifestyles than those who did not [29,30]. The participants may hence have lacked a need or motivation to change their lifestyles. In contrast, the participants of StopDia were recruited via a T2D risk-screening website based on their risk scores (the Finnish Diabetes Risk Score ≥12 points [59]). The screening may have raised the participants' risk awareness and motivation to manage the risk by improving their lifestyles.

Supporting the interpretation related to risk awareness and motivation for lifestyle modification, digital physical activity interventions have been found to yield greater effects among at-risk and sick individuals than among healthy populations [7]. The interpretation also receives support from the Theoretical Domains Framework, an integrative framework that synthesized 33 theories and 128 theoretical constructs related to behavior change into 14 theoretical domains that facilitate explaining and predicting behavior [60,61]. According to the Theoretical Domains Framework, key determinants of behavior include knowledge, such as the awareness of personal susceptibility to develop a health condition and the severity of the condition, and beliefs about consequences, such as expectations related to the benefits of changing versus the risks of not changing behavior [61]. These determinants can influence other determinants, such as intentions (ie, conscious decisions to act), goals (eg, goal setting and prioritizing), and behavioral regulation (eg, action planning and self-monitoring) [61-64]. All these determinants reflect psychological capability and reflective motivation, which belong to the necessary conditions of behavior according to the COM-B (capability, opportunity, motivation—behavior) system [65].

Despite the generally low intervention engagement observed in this study, the variability in app use was substantial, particularly in the number of reported habit performances. The top quartile reported more than 68.5 total performances, which is a considerable effort. Similar variability was observed in StopDia [19,20], in which 11% of participants used the BitHabit app approximately twice per week and the most active 4% nearly daily throughout the 1-year intervention, reporting a median of 1433 and 4516 total performances, respectively [20]. These observations highlight the value of this study and the need for continued research efforts to identify the characteristics of the population segment that will find inexpensively scalable lifestyle apps intended for self-directed well-being and health promotion engaging and beneficial. Advancements in this research space will facilitate the allocation of more resource-intensive health services to individuals who need them most. As one size does not fit all and no single intervention will solve the public health challenges societies face, we need various tools to serve various preferences and needs, together with knowledge on the types of individuals who benefit from each tool.

Effect of App Use on Changes in Primary Outcomes

The exploratory, quasi-experimental effects assessment was performed with 2 types of app use metrics that reflected participants' engagement with the app: the percentage of app use days and the number of reported habit performances. Each metric had its pros and cons. The percentage of use days was a fully objective measure that was directly based on app logins and reflected the frequency with which the users visited the app. However, it did not reveal how the app was used during these

XSL•FO
RenderX

visits. Habit performances, in contrast, reflected concrete actions taken for habit formation. They, however, were self-reported and could be scheduled in advance or recorded retrospectively, which made them subject to recall bias.

The associations observed between the app use and changes in the primary outcomes varied depending on the app use metric and the outcome. Together, however, the findings support an interpretation that improvements in diet quality and total physical activity were connected with the app use. Diet quality was statistically significantly associated with the total number of performances, and physical activity was associated with the percentage of app use days. In addition, the estimates for the associations of diet quality with the percentage of app use days and physical activity with the total number of performances were, albeit not statistically significant, in line with a favorable app use effect. These results suggest that an active use of a habit-promoting digital tool could facilitate small favorable changes in lifestyle habits. In contrast, no evidence was found for an association between BMI and app use. This might be due to the short duration of the follow-up and the fact that not all dietary or physical activity–related changes in behavior influence energy balance and thus body weight.

Interestingly, we observed no significant associations between changes in the primary outcomes and the number of reported habit performances specifically targeting each outcome. The only exception was a sensitivity analysis that found the performances related to BMI were associated with reduced BMI. While the number of actions that specifically advance an outcome could be considered a probable and precise predictor of improved outcomes, actions that are not directly related to the outcome can also contribute to its improvement. For example, actions that enhance mood or sleep can facilitate the repetition of diet-related habit performances, resulting in improved diet quality. This pathway could explain why we found evidence of an association between diet quality and the total number of habit performances, but not with diet-related performances.

Our results provide support for prior evidence that the effectiveness of a lifestyle app depends on its use. The StopDia trial demonstrated that a higher BitHabit user engagement—measured as at or above the median of 501 self-reported habit performances over the intervention year [18] or the number of monthly use days during intervention months 2 to 12 [20]—was associated with improvements in T2D risk factor levels, including diet quality, physical activity, and BMI. The greater the use was, the greater the benefits observed. While participants with sustained use once or twice per week showed small beneficial changes compared to those who terminated the use, participants with nearly daily use throughout the study showed the most beneficial changes [20].

Similarly, the habit-based Top Tips weight loss app that was piloted in a 3-month trial among adults with overweight or obesity was found most beneficial among active users [17]. Participants with the greatest changes in eating-related self-regulatory skills, weight, and adherence to target behaviors logged into the app and recorded their weight on average 2 to 3 times more and achieved the tips more often than participants with smaller changes in the outcomes [17]. Relatedly, a 4-week feasibility trial of an app that promoted vegetable intake with goal setting, self-monitoring, and feedback found the app use days and the frequency of recording vegetable intake via the app positively associated with increased vegetable intake among young adults [21].

The scientific evidence concerning the relationship between use and outcomes is not entirely consistent, however. Two systematic reviews examined this association in exclusively digital lifestyle interventions among adults [44,66]. One of them found a weak positive overall association between use and physical activity based on a meta-analysis of 11 studies and observed that the direction of the association was consistently positive for use metrics that reflected activities completed or logins [44]. The other review found mixed associations between use and dietary intake based on a narrative synthesis of 5 studies [66]. The inconsistency in the findings might be explained by, inter alia, differences in the digital lifestyle interventions and their target behaviors [66].

Effect Modification by Baseline Characteristics

We found some evidence that, in increasing movement, the use of the app might favor individuals who are physically more active and who have positive attitudes toward e-services. These characteristics also predicted the uptake of the app among the individuals who were invited to participate in this study [29,30]. The odds of uptake, though, were increased by a wide variety of individual characteristics, including middle age, female sex, higher education, higher income, better health, a healthier lifestyle (considering diet, physical activity, sleep, and smoking), as well as higher competence, fewer barriers, and greater perceived benefits of using e-services [29,30]. Hence, the study sample was not representative of the general adult population, which may have influenced the results of the effect modification analysis. In comparison, studies on the BitHabit app have found older age [20,22] and better diet quality [20] to predict greater engagement with the app.

Regarding lifestyle apps in general, a systematic review on factors influencing app adherence also reported that a positive attitude toward technology positively influences adherence to physical activity apps [23]. In contrast, this review found evidence of the positive influence of female gender on adherence to nutrition apps and of healthy weight on adherence to physical activity apps [23]. Another systematic review found overall mixed evidence of factors that influence the effectiveness of digital lifestyle interventions [25]. The review comprised 9 studies that examined inequality in effectiveness among adults with no medical conditions other than overweight or obesity. In these studies, greater effects were observed in participants with younger age (n=2 studies), older age (n=1), female gender (n=3), male gender (n=3), higher education (n=1), professional or managerial or executive job (n=1), and residence in the capital region (n=1) [25].

In sum, our preliminary, exploratory findings on the effect modifiers of the use of a habit-based app targeted at the general adult population—together with the evidence mentioned earlier of other digital lifestyle interventions—demonstrate the multitude of factors that may influence effects. The

XSL•FO
RenderX

inconsistency in existing evidence warrants further efforts to identify groups that are and are not likely to benefit from digital lifestyle interventions. The inconsistency in findings across studies may be due to the heterogeneity of examined interventions, including but not limited to their target behaviors, theoretical underpinnings, approaches to support behavior change, components, content, intended use, target populations, evaluation context, and mode of delivery. Considering these characteristics could hence be meaningful in future studies that compare the effect modifiers of digital lifestyle interventions.

App Use Experience

The participants gave the app and the behaviors it suggested a median rating of 7 out of 10. Only 15% (30/199) were ready to promote the app to their friends or colleagues. The ratings indicate that the app failed to fully satisfy the needs of most participants. This might be because the behaviors included in the app library were rather general or because the app's approach to personalization did not meet the users' preferences. The personalization meant that the app determined and suggested habit categories with the greatest potential for improvement based on the user's self-reported demographics, health behavior, perceived stress, and well-being goals. Digital lifestyle interventions have used varying approaches to personalization, but it remains unclear what type and level of personalization works and for whom [5]. For example, some may prefer a system-driven approach in which the digital tool determines the personalization, thus reducing the cognitive effort required from the user [5]. Some, in contrast, may prefer a user-driven approach that allows the user to control the personalization and thus provides a greater sense of autonomy [5]. For an increased degree of user-driven personalization, the app could, for example, allow and guide users to plan their own target habits and the contextual cues that are supposed to trigger their performance. This way, the pursued habits would better fit the individual's lifestyle and daily routines and could thus strengthen their motivation and commitment to repeat the behaviors until they become automatic. These factors are considered crucial for the formation and maintenance of new habits [12,15,16].

Strengths and Limitations

The strengths of the study include a behavioral theory–based digital lifestyle intervention, a web-based app that had been updated based on the learnings of a prior RCT among adults at risk of T2D [18]. In this study, the app was offered to a large group of adults sampled from the participants of a population-based health survey. As such, the study served as an attempt to scale the intervention up to a broader and more heterogeneous population. The study was rolled out in a realistic setting in which the app was simply made available for self-directed promotion of well-being and health, with no additional professional or peer support provided to boost its use. The assessments of the app use effect and effect modification incorporated participants' objectively measured engagement with the app—a key determinant of effectiveness—and considered a wide variety of participant characteristics potentially influencing the effects. The work thus contributes to the scant evidence base on the effect modifiers of exclusively digital lifestyle interventions.

The results of the study must, however, be interpreted in the context of its limitations, particularly the lack of a control arm that precluded the assessment of causal effects and the reliance on self-reported data, both of which increase the uncertainty of our findings. To reduce the potential bias caused by self-reported data, we used measures that were validated (eg, the HDI), theory-based (the acceptability measures), and/or used in prior research (eg, the physical activity and BMI measurements); removed implausible responses from the data; and conducted sensitivity analyses that excluded the participants who reported the largest changes in the primary outcomes.

Further limitations include a multiplicity of analyses and outcomes, which increases the risk of type I error, and low participation and high attrition rates, which reduce the representativeness of our study sample to the general adult population and the statistical power to detect small effects. While the individuals to whom the app was offered represented a subsample of a population-based health survey that targeted a nationally representative random sample of adults [28], this subsample only included individuals who had chosen to participate in the survey by a certain date. Of these, 18% (1282/6975) accepted the offer and registered for the app. Of those registered, 30% (n=382) completed the measurements required for the effects assessment of this study. These individuals tended to be better off in terms of socioeconomic status, health, lifestyles, and e-service use. Our findings may thus have limited relevance to the broader adult population of Finland. Reaching and engaging more vulnerable population groups would likely require more tailored approaches.

The high attrition rate may cause bias if individuals who benefited from the app were more likely to complete the study questionnaires, leading to overoptimistic effect estimates. We compared the baseline characteristics of the completers and noncompleters of the study. The comparison showed that the completers had, on average, slightly fewer years of education and better diet quality and that they were more often retired than the noncompleters. The low participation and high attrition rate were not unusual, however, as they are widely reported challenges in digital lifestyle interventions [3,4,16,17,21], particularly when conducted among unselected target groups that may lack perceived need and motivation for lifestyle change [7,22]. In this study, the invitees had additionally recently participated in the population-based health survey, which may have reduced their motivation and resources to participate in this study as well. The findings of this study should therefore be considered preliminary, exploratory evidence that encourages further research in more robust study designs among more diverse and representative study populations with sufficient statistical power to detect small effects.

Implications for Future Research

The findings of this study imply that for the development of more widely engaging habit-based digital lifestyle interventions, cocreation with the target group would be crucial. In addition, expert consensus supports a user-centered and iterative development that uses mixed methods and in-depth qualitative

XSL•FO
RenderX

research to progressively refine the intervention to meet users' context and needs [67]. Such work could also deepen our understanding of the relationship between the engagement with and the effects of the app [66], for example, by facilitating the characterization of "effective engagement"—which may not simply mean "more engagement"—to achieve the intended outcomes [67].

Conclusions

This exploratory, correlational study suggests that an inexpensive, low-intensity digital support for healthy habits could foster small beneficial lifestyle changes if it succeeds in sustaining the user's engagement. However, effects may depend on the individual. Identifying those most likely to benefit from such exclusively digital interventions calls for deeper insight into the modifiers of user engagement and effects.

Acknowledgments

We are grateful to Dr Hilkka Liedes (VTT Technical Research Centre of Finland) and Adjunct Professor Jouni Lahti (Finnish Institute for Health and Welfare, THL) for advice on data analysis and to the Healthy Finland research group, particularly Adjunct Professor Annamari Lundqvist (THL) for collaboration.

Funding

This study was funded by the Research Council of Finland (grant 332464, 332465, and 332466). Funding for the article processing fee was provided by the Finnish Institute for Health and Welfare. The funders had no involvement in the study design, data collection, analysis, interpretation, or the writing of the manuscript.

Data Availability

The datasets generated and analyzed in this study are available from the corresponding author upon reasonable request for research collaboration.

Authors' Contributions

Conceptualization: JL, IR
Data curation: MV, AU, IR
Formal analysis: MV
Funding acquisition: JL, JP
Investigation: IR, JL, SP
Methodology: ER, MV, JL, IR
Project administration: IR, JL
Resources: IR, JL, SP
Software: AU
Supervision: JL
Visualization: ER, MV
Writing – original draft: ER, MV
Writing – review & editing: ER, MV, AU, SP, JP, IR, JL
All authors have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Original questionnaire items and additional tables.
[DOCX File, 68 KB - mhealth_v14i1e84076_app1.docx]

Checklist 1
STROBE checklist.
[DOCX File, 43 KB - mhealth_v14i1e84076_app2.docx]

References

1. Agachi E, Mierau JO, van Ittersum K, Bijmolt THA. The impact of a lifestyle behaviour change program on healthcare costs: quasi-experimental real-world evidence from an open-access mobile health app in the Netherlands. Prev Med 2024 Dec;189:108174. [doi: 10.1016/j.ypmed.2024.108174] [Medline: 39532214]

XSL•FO
RenderX

2. Park S, Ballreich J, Ward T, Shi L. Cost-effectiveness analysis of a digital diabetes-prevention programme versus an in-person diabetes-prevention programme in people with prediabetes in the United States. Diabetes Obes Metab 2024 Oct;26(10):4522-4534. [doi: 10.1111/dom.15807] [Medline: 39056211]
3. Gold N, Yau A, Rigby B, Dyke C, Remfry EA, Chadborn T. Effectiveness of digital interventions for reducing behavioral risks of cardiovascular disease in nonclinical adult populations: systematic review of reviews. J Med Internet Res 2021 May 14;23(5):e19688. [doi: 10.2196/19688] [Medline: 33988126]
4. Singh B, Ahmed M, Staiano AE, et al. A systematic umbrella review and meta-meta-analysis of eHealth and mHealth interventions for improving lifestyle behaviours. NPJ Digit Med 2024 Jul 5;7(1):179. [doi: 10.1038/s41746-024-01172-y] [Medline: 38969775]
5. Livingstone KM, Rawstorn JC, Partridge SR, et al. Digital behaviour change interventions to increase vegetable intake in adults: a systematic review. Int J Behav Nutr Phys Act 2023 Mar 27;20(1):36. [doi: 10.1186/s12966-023-01439-9] [Medline: 36973716]
6. Tzelepis F, Mitchell A, Wilson L, et al. The long-term effectiveness of internet-based interventions on multiple health risk behaviors: systematic review and robust variance estimation meta-analysis. J Med Internet Res 2021 Dec 21;23(12):e23513. [doi: 10.2196/23513] [Medline: 34931990]
7. Mönninghoff A, Kramer JN, Hess AJ, et al. Long-term effectiveness of mHealth physical activity interventions: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2021 Apr 30;23(4):e26699. [doi: 10.2196/26699] [Medline: 33811021]
8. Robert C, Erdt M, Lee J, Cao Y, Naharudin NB, Theng YL. Effectiveness of eHealth nutritional interventions for middle-aged and older adults: systematic review and meta-analysis. J Med Internet Res 2021 May 17;23(5):e15649. [doi: 10.2196/15649] [Medline: 33999005]
9. Núñez de Arenas‐Arroyo S, Cavero‐Redondo I, Alvarez‐Bueno C, Sequí‐Domínguez I, Reina‐Gutiérrez S, Martínez‐Vizcaíno V. Effect of eHealth to increase physical activity in healthy adults over 55 years: a systematic review and meta‐analysis. Scandinavian Med Sci Sports 2021 Apr;31(4):776-789. [doi: 10.1111/sms.13903]
10. Shi B, Li G, Wu S, et al. Assessing the effectiveness of eHealth interventions to manage multiple lifestyle risk behaviors among older adults: systematic review and meta-analysis. J Med Internet Res 2024 Jul 31;26:e58174. [doi: 10.2196/58174] [Medline: 39083787]
11. Orbell S, Verplanken B. Changing behavior using habit theory. In: Hagger MS, Cameron LD, Hamilton K, Hankonen N, Lintunen T, editors. The Handbook of Behavior Change, 1st edition: Cambridge University Press; 2020:178-192. [doi: 10.1017/9781108677318.013]
12. Gardner B, Rebar AL, Lally P. Habit interventions. In: Hagger MS, Cameron LD, Hamilton K, Hankonen N, Lintunen T, editors. The Handbook of Behavior Change, 1st edition: Cambridge University Press; 2020:599-616. [doi: 10.1017/9781108677318.041]
13. Gardner B, Rebar AL. Habit Formation and Behavior Change: Oxford Research Encyclopedia of Psychology. Oxford University Press; 2019. [doi: 10.1093/acrefore/9780190236557.013.129]
14. Albarracín D, Fayaz-Farkhad B, Samayoa JAG. Determinants of behaviour and their efficacy as targets of behavioural change interventions. Nat Rev Psychol 2024 Jun;3(6):377-392. [doi: 10.1038/s44159-024-00305-0] [Medline: 40909436]
15. Cleo G, Beller E, Glasziou P, Isenring E, Thomas R. Efficacy of habit-based weight loss interventions: a systematic review and meta-analysis. J Behav Med 2020 Aug;43(4):519-532. [doi: 10.1007/s10865-019-00100-w] [Medline: 31529279]
16. Baretta D, Gillmann N, Edgren R, Inauen J. HabitWalk: a micro‐randomized trial to understand and promote habit formation in physical activity. Applied Psych Health Well 2025 Feb 25;17(1):e12605. [doi: 10.1111/aphw.12605]
17. Kliemann N, Croker H, Johnson F, Beeken RJ. Development of the Top Tips habit-based weight loss app and preliminary indications of its usage, effectiveness, and acceptability: mixed-methods pilot study. JMIR Mhealth Uhealth 2019 May 10;7(5):e12326. [doi: 10.2196/12326] [Medline: 31094352]
18. Lakka TA, Aittola K, Järvelä-Reijonen E, et al. Real-world effectiveness of digital and group-based lifestyle interventions as compared with usual care to reduce type 2 diabetes risk - a stop diabetes pragmatic randomised trial. Lancet Reg Health Eur 2023 Jan;24:100527. [doi: 10.1016/j.lanepe.2022.100527] [Medline: 36620354]
19. Harjumaa M, Absetz P, Ermes M, et al. Internet-based lifestyle intervention to prevent type 2 diabetes through healthy habits: design and 6-month usage results of randomized controlled trial. JMIR Diabetes 2020 Aug 11;5(3):e15219. [doi: 10.2196/15219] [Medline: 32779571]
20. Lavikainen P, Mattila E, Absetz P, et al. Digitally supported lifestyle intervention to prevent type 2 diabetes through healthy habits: secondary analysis of long-term user engagement trajectories in a randomized controlled trial. J Med Internet Res 2022 Feb 24;24(2):e31530. [doi: 10.2196/31530] [Medline: 35200147]
21. Nour M, Chen J, Allman-Farinelli M. Young adults' engagement with a self-monitoring app for vegetable intake and the impact of social media and gamification: feasibility study. JMIR Form Res 2019 May 10;3(2):e13324. [doi: 10.2196/13324] [Medline: 31094322]
22. Rekola H, Tolmunen T, Mattila E, et al. User archetypes of a well-being-promoting mobile app among adults: cross-sectional study and cluster analysis of usage patterns. JMIR Mhealth Uhealth 2025 Aug 18;13:e68982. [doi: 10.2196/68982] [Medline: 40825235]

XSL•FO
RenderX

23. Jakob R, Harperink S, Rudolf AM, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371. [doi: 10.2196/35371] [Medline: 35612886]
24. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572. [doi: 10.2196/17572] [Medline: 32348255]
25. Szinay D, Forbes CC, Busse H, DeSmet A, Smit ES, König LM. Is the uptake, engagement, and effectiveness of exclusively mobile interventions for the promotion of weight-related behaviors equal for all? A systematic review. Obes Rev 2023 Mar;24(3):e13542. [doi: 10.1111/obr.13542] [Medline: 36625062]
26. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med 2007 Oct 16;4(10):e296. [doi: 10.1371/journal.pmed.0040296] [Medline: 17941714]
27. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med 2007 Oct 16;4(10):e297. [doi: 10.1371/journal.pmed.0040297] [Medline: 17941715]
28. Sääksjärvi K, Elonheimo HM, Ikonen J, et al. Cohort profile: healthy Finland survey. Int J Epidemiol 2024 Dec 16;54(1):dyae166. [doi: 10.1093/ije/dyae166] [Medline: 39715488]
29. Ruotsalainen I, Valtanen M, Kärsämä R, et al. Influencing factors in digital health intervention uptake: the interplay of education, lifestyle, and digital literacy. In: Särestöniemi M, Keikhosrokiani P, Singh D, Harjula E, Tiulpin A, Jansson M, et al, editors. Digital Health and Wireless Solutions: Springer Nature Switzerland; 2024:322-332. [doi: 10.1007/978-3-031-59091-7_21]
30. Ruotsalainen I, Valtanen M, Kärsämä R, et al. Unraveling the factors associated with digital health intervention uptake: cross-sectional study. JMIR Mhealth Uhealth 2025 Dec 9;13:e63896. [doi: 10.2196/63896] [Medline: 41364913]
31. Pihlajamäki J, Männikkö R, Tilles-Tirkkonen T, et al. Digitally supported program for type 2 diabetes risk identification and risk reduction in real-world setting: protocol for the StopDia model and randomized controlled trial. BMC Public Health 2019 Mar 1;19(1):255. [doi: 10.1186/s12889-019-6574-y] [Medline: 30823909]
32. Fogg BJ, Hreha J. Behavior wizard: a method for matching target behaviors with solutions. In: Ploug T, Hasle P, Oinas-Kukkonen H, editors. Persuasive Technology: Springer; 2010:117-131. [doi: 10.1007/978-3-642-13226-1_13]
33. Wood W, Neal DT. Healthy through habit: interventions for initiating & maintaining health behavior change. Behav Sci Policy 2016 Apr;2(1):71-83. [doi: 10.1177/237946151600200109]
34. Lally P, van Jaarsveld CHM, Potts HWW, Wardle J. How are habits formed: modelling habit formation in the real world. Euro J Social Psych 2010 Oct;40(6):998-1009. [doi: 10.1002/ejsp.674]
35. Ryan RM, Deci EL. Basic Psychological Needs in Motivation, Development, and Wellness: The Guilford Press; 2017.
36. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037/0003-066X.55.1.68]
37. Nordic Council of Ministers. Nordic Nutrition Recommendations 2012: Integrating Nutrition and Physical Activity: Nordic Council of Ministers; 2014. [doi: 10.26530/OAPEN_483279] [Medline: 25246403]
38. The National Nutrition Council of Finland. The Finnish Nutrition Recommendations 2014: The National Nutrition Council of Finland; 2014.
39. Garber CE, Blissmer B, Deschenes MR, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc 2011 Jul;43(7):1334-1359. [doi: 10.1249/MSS.0b013e318213fefb] [Medline: 21694556]
40. Colberg SR, Sigal RJ, Fernhall B, et al. Exercise and type 2 diabetes: the American College of Sports Medicine and the American Diabetes Association: joint position statement. Diabetes Care 2010 Dec;33(12):e147-e167. [doi: 10.2337/dc10-9990] [Medline: 21115758]
41. Haskell WL, Lee IM, Pate RR, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Med Sci Sports Exerc 2007 Aug;39(8):1423-1434. [doi: 10.1249/mss.0b013e3180616b27] [Medline: 17762377]
42. Kyytsönen M, Aalto AM, Vehko T. Social and health care online service use in 2020–2021: experiences of the population. : Finnish Institute for Health and Welfare (THL); 2021.
43. Heikkinen AL, Hänninen T, Kuikka P, et al. The Cognitive Function at Work Questionnaire (CFWQ): a new scale for measuring cognitive complaints in occupational population. Appl Neuropsychol Adult 2023;30(6):649-660. [doi: 10.1080/23279095.2021.1970553] [Medline: 34482772]
44. Mclaughlin M, Delaney T, Hall A, et al. Associations between digital health intervention engagement, physical activity, and sedentary behavior: systematic review and meta-analysis. J Med Internet Res 2021 Feb 19;23(2):e23180. [doi: 10.2196/23180] [Medline: 33605897]

XSL•FO
RenderX

45. Liedes H, Mattila E, Honka A, et al. Engagement with the bithabit digital lifestyle intervention and changes in type 2 diabetes risk factors: a secondary analysis of a randomized controlled trial (preprint). J Med Internet Res. Preprint posted online on Jun 29, 2023. [doi: 10.2196/preprints.50361]
46. Lindström J, Aittola K, Pölönen A, et al. Formation and validation of the Healthy Diet Index (HDI) for evaluation of diet quality in healthcare. Int J Environ Res Public Health 2021 Feb 28;18(5):2362. [doi: 10.3390/ijerph18052362] [Medline: 33670967]
47. Jetté M, Sidney K, Blümchen G. Metabolic equivalents (METS) in exercise testing, exercise prescription, and evaluation of functional capacity. Clin Cardiol 1990 Aug;13(8):555-565. [doi: 10.1002/clc.4960130809] [Medline: 2204507]
48. Mendes MDA, da Silva I, Ramires V, et al. Metabolic equivalent of task (METs) thresholds as an indicator of physical activity intensity. PLoS One 2018;13(7):e0200701. [doi: 10.1371/journal.pone.0200701] [Medline: 30024953]
49. Weekly physical activity recommendation for 18–64-year-olds. UKK Institute. 2019. URL: https://ukkinstituutti.fi/en/products-services/physical-activity-recommendations/weekly-physical-activity-recommendation-for-18-64-year-olds/ [accessed 2026-08-17]
50. U.S. Department of Health and Human Services. Physical Activity Guidelines for Americans, 2nd edition: U.S. Department of Health and Human Services; 2018. URL: https://odphp.health.gov/our-work/nutrition-physical-activity/physical-activity-guidelines/current-guidelines [accessed 2026-09-02]
51. McAdams MA, Van Dam RM, Hu FB. Comparison of self-reported and measured BMI as correlates of disease markers in US adults. Obesity (Silver Spring) 2007 Jan;15(1):188-196. [doi: 10.1038/oby.2007.504] [Medline: 17228047]
52. Hodge JM, Shah R, McCullough ML, Gapstur SM, Patel AV. Validation of self-reported height and weight in a large, nationwide cohort of U.S. adults. PLoS One 2020;15(4):e0231229. [doi: 10.1371/journal.pone.0231229] [Medline: 32282816]
53. Zhang J, Olsen A, Halkjær J, et al. Self-reported and measured anthropometric variables in association with cardiometabolic markers: a Danish cohort study. PLoS ONE 2023;18(7):e0279795. [doi: 10.1371/journal.pone.0279795]
54. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88. [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
55. Reichheld FF. The one number you need to grow. Harv Bus Rev 2003 Dec;81(12):46-54. [Medline: 14712543]
56. The r project for statistical computing. R Core Team. 2026. URL: https://www.R-project.org [accessed 2026-03-28]
57. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1). [doi: 10.18637/jss.v067.i01]
58. Fournier M, d'Arripe-Longueville F, Rovere C, et al. Effects of circadian cortisol on the development of a health habit. Health Psychol 2017 Nov;36(11):1059-1064. [doi: 10.1037/hea0000510] [Medline: 28650196]
59. Lindstro¨m J, Tuomilehto J. The diabetes risk score. Diabetes Care 2003 Mar 1;26(3):725-731. [doi: 10.2337/diacare.26.3.725]
60. Michie S, Johnston M, Abraham C, et al. Making psychological theory useful for implementing evidence based practice: a consensus approach. Qual Saf Health Care 2005 Feb;14(1):26-33. [doi: 10.1136/qshc.2004.011155] [Medline: 15692000]
61. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012 Apr 24;7:37. [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
62. Orbell S, Zahid H, Henderson CJ. Changing behavior using the health belief model and protection motivation theory. In: Hagger MS, Cameron LD, Hamilton K, Hankonen N, Lintunen T, editors. The Handbook of Behavior Change, 1st edition: Cambridge University Press; 2020:46-59. [doi: 10.1017/9781108677318.004]
63. Schwarzer R, Hamilton K. Changing behavior using the health action process approach. In: Hagger MS, Cameron LD, Hamilton K, Hankonen N, Lintunen T, editors. The Handbook of Behavior Change, 1st edition: Cambridge University Press; 2020:89-103. [doi: 10.1017/9781108677318.007]
64. DiClemente CC, Graydon MM. Changing behavior using the transtheoretical model. In: Hagger MS, Cameron LD, Hamilton K, Hankonen N, Lintunen T, editors. The Handbook of Behavior Change, 1st edition: Cambridge University Press; 2020:136-149. [doi: 10.1017/9781108677318.010]
65. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implementation Sci 2011 Dec;6(1):42. [doi: 10.1186/1748-5908-6-42]
66. Delaney T, Mclaughlin M, Hall A, et al. Associations between digital health intervention engagement and dietary intake: a systematic review. Nutrients 2021 Sep 20;13(9):3281. [doi: 10.3390/nu13093281] [Medline: 34579158]
67. Yardley L, Spring BJ, Riper H, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]

Abbreviations

COM-B: capability, opportunity, motivation—behavior
HDI: Healthy Diet Index
MET: metabolic equivalent of task
RCT : randomized controlled trial
StopDia: Stop Diabetes

XSL•FO
RenderX

STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
T2D: type 2 diabetes

Edited by L Buis; submitted 11.Oct.2025; peer-reviewed by K Saint-Onge, K DiFilippo, X Tao; revised version received 07.Jun.2026; accepted 13.Jul.2026; published 21.Sep.2026.

Please cite as:
Rantala E, Valtanen M, Umer A, Parikka S, Pihlajamäki J, Ruotsalainen I, Lindström J
The Usage Effects, Effect Modifiers, and Experiences of a Web-Based App for Healthy Habit Formation in Adults: Exploratory Analysis of a Quasi-Experimental Real-World Intervention
JMIR Mhealth Uhealth 2026;14:e84076
URL: https://mhealth.jmir.org/2026/1/e84076
doi:10.2196/84076

© Eeva Rantala, Mikko Valtanen, Adil Umer, Suvi Parikka, Jussi Pihlajamäki, Ilona Ruotsalainen, Jaana Lindström. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 21.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Caregiver App Use and Child Behavioral Adherence in Relation to Obesity-Related Outcomes in an mHealth App-Assisted Childhood Obesity Intervention: Secondary Analysis of a Cluster Randomized Trial

Shiyu Yan[1,2], PhD; Jinlang Lyu[2], PhD; Chunjie Yin[2], PhD; Wenhao Li[3], MPH; Yan Li[1*], PhD; Hai-jun Wang[2*], PhD

[1]Clinical Research Center for Children's Health and Diseases, Shandong Provincial Maternal and Child Health Care Hospital Affiliated to Qingdao University, Jinan, China

[2]Department of Maternal and Child Health, School of Public Health, Peking University, National Health Commission Key Laboratory of Reproductive Health, Peking University Health Science Center-Weifang Joint Research Center for Maternal and Child Health, Xueyuan road 38, Beijing, China

[3]Institute of Child Health Care, Jinan Children's Hospital, Jinan, China

[*]these authors contributed equally

Corresponding Author:
Hai-jun Wang, PhD
Department of Maternal and Child Health, School of Public Health, Peking University, National Health Commission Key Laboratory of Reproductive Health, Peking University Health Science Center-Weifang Joint Research Center for Maternal and Child Health, Xueyuan road 38, Beijing, China

Abstract

Background: Childhood obesity remains a major public health challenge. Mobile health (mHealth) interventions offer a scalable approach to support behavior change, but their effectiveness may depend on participant adherence. While both caregiver and child adherence are important, few studies have jointly examined caregiver app-use adherence and child behavioral adherence within digital obesity interventions.

Objective: This study aimed to evaluate the independent and joint associations of caregiver app-use adherence and child behavioral adherence with obesity-related outcomes in a school-based mHealth intervention.

Methods: We conducted a secondary analysis of 684 child-caregiver dyads in the intervention arm of the diet, exercise, and cardiovascular health—children cluster-randomized trial, implemented across 3 socioeconomically diverse regions in China. Caregiver app-use adherence was measured using app-recorded app-use frequency and total usage duration. Child behavioral adherence was assessed through weekly behavior monitoring scores covering diet, screen time, and physical activity. Linear mixed models estimated associations between adherence levels and changes in anthropometric and physical fitness outcomes over 9 months.

Results: Higher caregiver app-use frequency and duration were both significantly associated with reductions in children's obesity-related indicators, such as BMI (β=−0.19; *P*=.005), BMI *z*-score (β=−0.08; *P*=.006), and body fat percentage (BFP) (β=−0.70; *P*=.005) for frequency. Greater child behavioral adherence was associated with lower BFP (β=−0.63; *P*=.01) and better physical fitness, including shorter shuttle run time (β=−3.63; *P*<.001) and longer standing long jump distance (β=2.88; *P*=.005). App-use frequency and duration showed a significant interaction for rope-skipping performance (*P*=.05). In joint adherence analyses, children in the high caregiver-high child adherence group showed the most favorable changes, including lower BMI (β=−0.22; *P*=.02), lower BFP (β=−1.18 percentage points; *P*=.001), and 7.45 additional rope-skipping repetitions (*P*=.003), compared with the low caregiver-low child adherence group.

Conclusions: In this secondary analysis of an mHealth app-assisted childhood obesity intervention, higher caregiver app-use adherence and child behavioral adherence were associated with more favorable changes in weight status and physical fitness. These findings underscore the importance of coengagement in digital health programs and highlight the potential of app-supported, family-engaged strategies for childhood obesity prevention.

Trial Registration: ClinicalTrials.gov. NCT03665857; http://clinicaltrials.gov/study/NCT03665857

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2018-027902

(JMIR Mhealth Uhealth 2026;14:e89548) doi:10.2196/89548

XSL•FO
RenderX

KEYWORDS

child; obesity; mobile health; intervention; adherence

Introduction

Childhood overweight and obesity have become critical global public health challenges. According to the World Health Organization (WHO), the prevalence of overweight and obesity among children and adolescents aged 5‐19 years has risen sharply from 4% in 1975 to 20% in 2022 [1]. In China, the obesity rate among primary and middle school students increased from 0.1% in 1985 to 9.6% in 2019, a 75.6-fold increase [2]. Childhood obesity is associated with adverse physical and psychological health consequences during childhood and elevates the risk of hypertension, type 2 diabetes, and certain cancers in adulthood [3]. Effective prevention in early life is therefore essential to curbing the obesity epidemic and reducing long-term health impacts and economic burdens.

Although numerous intervention studies on childhood obesity have been conducted worldwide, their results have been inconsistent, even when similar strategies were used [4,5]. One key but often overlooked factor related to intervention outcomes is participant adherence, defined as the extent to which participants comply with the study protocol [6]. Poor adherence can reduce the delivered intervention dose and complicate the interpretation of intervention findings. However, adherence remains underreported in obesity intervention research: a systematic review found that only 37% of high-quality studies included adherence data [7]. In addition, when reported, adherence was mostly used to assess feasibility rather than to quantitatively examine its association with intervention outcomes [8]. This gap limits understanding of how adherence is related to intervention outcomes and hinders the identification of scalable, evidence-informed strategies.

The emergence of mobile health (mHealth) technologies offers new opportunities to deliver, monitor, and tailor obesity interventions in real time [9,10]. Smartphone apps enable continuous health education, behavior tracking, and personalized feedback, potentially enhancing engagement and sustainability. This digital approach also helps address a longstanding challenge in intervention research: the repeated and simultaneous measurement of adherence. In childhood obesity interventions, adherence involves both intervention implementers, such as caregivers or teachers, and intervention participants, namely children themselves [11]. However, traditional methods have struggled to capture this process; caregiver adherence is often understudied due to measurement challenges and incomplete data [12]. Likewise, child adherence is frequently reported as a behavioral outcome rather than being examined as a process indicator related to intervention outcomes. Consequently, evidence on how caregiver app-use adherence and child behavioral adherence are jointly associated with obesity-related outcomes remains scarce. The automated data capture capabilities of mHealth platforms now allow caregiver app use and child behavioral adherence to be measured repeatedly, facilitating analysis of their independent and joint associations with health outcomes.

To address these gaps, we leveraged a mobile app-supported, multicomponent childhood obesity intervention. The app automatically recorded caregiver app-use data, while children's behavioral adherence was tracked weekly through structured monitoring. This secondary analysis of the intervention arm of the diet, exercise, and cardiovascular health—children (DECIDE-Children) cluster randomized trial aimed to evaluate how caregiver app-use adherence and child behavioral adherence, individually and jointly, were associated with obesity-related outcomes.

Methods

Study Design and Secondary Analysis Sample

This study was a secondary analysis of the intervention arm of the DECIDE-Children trial conducted in 3 socioeconomically diverse regions of China: Beijing, Changzhi (Shanxi Province), and Urumqi (Xinjiang Province). A total of 24 schools (8 schools per region) with Grade 4 students were randomly assigned (1:1) to the intervention or control group after baseline assessments. The 9-month multifaceted intervention (from September 2018 to June 2019) targeted both children and their environments by promoting a healthy diet and physical activity and by engaging schools and families to support children's behavioral changes, with family involvement strengthened through the "Eat Wisely, Move Happily" smartphone app. Details of the trial protocol have been published [13]. The reporting of this study followed the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines; the completed checklist is provided in Checklist 1.

The primary effectiveness analysis of the DECIDE-Children trial has been published previously [14]. Compared with the control group, the intervention group had a significantly greater reduction in BMI from baseline to the end of the trial, with a mean between-group difference in BMI change of −0.46 (95% CI −0.67 to −0.25; $P<.001$). The prevalence of obesity decreased by 27.0% relative to baseline in the intervention group, compared with 5.6% in the control group. The intervention also improved other adiposity outcomes, dietary, sedentary, and physical activity behaviors, and obesity-related knowledge, although no significant changes were observed in moderate-to-vigorous physical activity, physical fitness, or blood pressure. The present study focused on 684 child-caregiver dyads in the intervention arm to examine how caregiver app use and child behavioral adherence were associated with obesity-related outcomes within the intervention group.

Ethical Considerations

The original trial was approved by the Peking University Institutional Review Board (IRB00001052-18021). Written informed consent was obtained from the parents or legal guardians of all participating children before enrollment. Participation was voluntary, and participants were free to withdraw from the study at any time. The original ethics

XSL•FO
RenderX

approval covered secondary analyses of the collected data; therefore, no additional informed consent was required for the present secondary analysis. To protect participant privacy and confidentiality, the database used for analysis contained no directly identifiable information, and coded identifiers were used instead.

Intervention

The intervention was designed based on a social-ecological framework and implemented at school, family, and individual levels. At the school level, the intervention included establishing school policies to support obesity prevention and providing health education lectures and training for teachers. Teachers then delivered health education sessions to students during regular school health education classes. The core messages focused on healthy eating and physical activity, including avoiding overeating and sugar-sweetened beverages, reducing high-energy foods and sedentary time, and increasing physical activity.

At the family level, caregivers received health education materials and were encouraged to support children's behavioral changes at home, particularly by improving the home food environment and supporting physical activity outside of school. Family involvement was strengthened through a mobile app, which was installed on caregivers' smartphones to avoid increasing children's screen time. Through the app, caregivers could receive information on obesity prevention, record their children's weekly dietary and physical activity behaviors, and view their children's assessment and feedback information.

At the individual student level, the intervention included health education sessions, reinforcement of in-school physical activity, and regular monitoring of weight and height. Physical education teachers encouraged students to engage in moderate-to-vigorous physical activity during school, and trained project staff or trained teachers measured children's weight and height regularly according to standardized procedures. These data were entered into the intervention system and used to generate feedback on children's weight status and weight-management progress.

The mobile app supported the intervention through 4 main modules: information diffusion, behavior monitoring, weight management, and assessment and feedback. In the behavior-monitoring module, caregivers were asked to observe and/or inquire about their child's weekly diet, screen time, and physical activity behaviors and report this information in the app, preferably together with the child. In the assessment and feedback module, the app automatically generated feedback based on children's weight-monitoring data and behavior-monitoring records. Caregivers and students could access these assessment results and feedback through the app, which was intended to help families understand their children's weight-management status and reinforce targeted behavior changes during the 9-month intervention.

Adherence Measurement via Mobile App

Caregiver App-Use Adherence

The mobile app included 4 main modules: information diffusion, behavior monitoring, weight management, and assessment and feedback (Table S1 in Multimedia Appendix 1). Caregiver app use referred to app-recorded use of the "Eat Wisely, Move Happily" mobile app and was used to measure caregiver adherence to the app-based components of the intervention. When caregivers used any function of the app, the background system automatically recorded the student ID, the app-use event, and the usage duration. App-use frequency was defined as the total number of app-use events recorded across all modules, and app-use duration was defined as the cumulative time spent using the app. In the present analysis, these 2 app-recorded indicators were used to quantify caregiver app-use adherence.

Child Behavioral Adherence

Child behavioral adherence was defined as the extent to which children followed the targeted weekly diet, screen time, and physical activity recommendations. It was assessed through the behavior-monitoring module. According to the intervention protocol, caregivers were asked to complete the behavior-monitoring questionnaire once per week, for a total of 35 time points during the intervention. Caregivers were instructed to observe and/or inquire about their child's dietary and physical activity behaviors during the preceding week. The questionnaire was pushed through the app every Sunday at 8:00 PM, when caregivers were generally more available and could review the child's behaviors from the preceding week together with the child.

The behavior score included 7 components aligned with the intervention goals: drinking sugary beverages, eating Western fast food, eating fried food, eating unhealthy snacks, overeating, screen time of more than 1 hour, and at-home exercise. Overeating was explained to caregivers during the health education component as eating beyond satiety or eating noticeably more than the child's usual amount. Scores were assigned according to the weekly frequency of each behavior (Table S2 in Multimedia Appendix 1). For each item, 5 points indicated meeting the recommended standard, and lower scores indicated poorer adherence to the recommended behavior. The maximum weekly score was 35 points, with higher scores indicating better behavioral adherence by the child. Missing behavior-monitoring data accounted for approximately 10% of data each month and were imputed using a 2-step approach, as described in a previously published analysis of the same intervention data [15].

Outcome Measures

Obesity-related outcomes included adiposity indicators and physical fitness measures. Adiposity indicators included changes in BMI, BMI *z*-score, waist circumference (WC), and body fat percentage (BFP) from baseline to the 9-month follow-up. Height was measured using a stadiometer (Huateng GMCS-1; Beijing Xindong Huateng Sports Equipment Co, Ltd), and weight was measured using a lever scale (Wujin RGT-140). BMI was calculated as weight in kilograms divided by height in meters squared. The BMI *z*-score was calculated according to the WHO growth reference using each child's age and sex [16]. WC was measured using a MyoTape, and BFP was assessed using a body composition analyzer (Tanita MC-780 MA). Physical fitness measures included the 50 m × 8 shuttle run, standing long jump, and 1-minute rope skipping, which

XSL•FO
RenderX

were assessed by trained outcome assessors according to standardized procedures. All outcomes were measured at baseline and after the 9-month intervention.

Statistical Analyses

Descriptive Analysis

Baseline characteristics of students and their caregivers were summarized. Continuous variables were presented as means with SDs for normally distributed data, whereas categorical variables were reported as frequencies and percentages.

Association Between Caregivers' and Children's Adherence and Obesity-Related Outcomes

Caregiver app-use adherence was measured using two app-recorded indicators: (1) app-use frequency, defined as the total number of app-use events recorded when caregivers used any function of the app, and (2) app-use duration, defined as the cumulative time spent using the app. Child behavioral adherence was calculated as the average weekly score for diet, screen time, and physical activity behaviors reported through the behavior-monitoring module. Each adherence indicator was classified in two ways: (1) binary classification using the median (low vs high); (2) quartile classification (Q1: ≥ upper quartile, Q2: median to < upper quartile, Q3: lower quartile to < median, Q4: < lower quartile).

Linear mixed-effects models were used to evaluate the associations between adherence levels and changes in obesity-related outcomes, including BMI, BMI *z*-score, WC, BFP, 50 m × 8 shuttle-run time, standing long-jump distance, and 1-minute rope-skipping performance. All models were adjusted for child's age, gender, parental age, parental BMI, parental education level, and the corresponding baseline measure of each outcome. A random intercept for class was included to account for within-class clustering. For quartile-based analyses, trend tests were conducted by treating the quartile variable as an ordinal predictor.

To examine the relative importance of app-use frequency and app-use duration, participants were categorized into 4 subgroups based on median splits of both app-use frequency and app-use duration: high frequency and high duration, high frequency and low duration, low frequency and high duration, and low frequency and low duration. This 2×2 classification allowed us to compare different app-use patterns. A formal interaction term between app-use frequency and app-use duration was included in the model to assess whether their joint association with outcomes differed from their individual associations.

Joint Associations of Caregiver and Child Adherence and Trend Analyses

To investigate joint associations, three joint adherence variables were constructed by combining caregiver and child adherence indicators: (1) app-use duration × child behavioral adherence, (2) app-use frequency × child behavioral adherence, and (3) combined app-use pattern × child behavioral adherence. Each joint variable was categorized into 3 groups: both low adherence, mixed adherence, and both high adherences. The "both low adherence" group served as the reference group. Linear mixed-effects models were used to assess the associations between joint adherence groups and changes in obesity-related outcomes. Covariates and random effects were consistent with the previous models. Trend tests were conducted by treating the joint adherence groups as an ordinal variable. To assess potential variation in associations, interaction terms between caregiver and child adherence were included in the models.

Sensitivity Analyses

To examine the robustness of the findings, sensitivity analyses were conducted by excluding students who were primarily cared for by grandparents, as caregiver type may be related to app-use patterns, behavior monitoring, and response to the intervention.

The results were considered statistically significant at a 2-sided $P<.05$. Statistical analyses were carried out using R software (version 4.1.2; creators: John Chambers and colleagues; location: Jersey City, NJ, USA).

Results

Baseline Characteristics

A total of 684 students and their caregivers were included in the analysis. Baseline characteristics are presented in Table 1. Among the students, 343 were boys and 341 were girls. In addition, 627/683 (91.8%) students were of Han ethnicity. Most students were only children (408/678, 60.2%). The mean (SD) BMI and BMI *z*-score were 18.54 (3.70) and 0.70 (1.44), respectively. At baseline, 158 (23.1%) students had obesity, and 266 (38.9%) students had overweight or obesity.

XSL•FO
RenderX

Table . The baseline characteristics of children and their caregivers.

Characteristics	Values
Students	
Gender, n (%)	
Boy	343 (50.1)
Girl	341 (49.9)
Nation, n (%)[a]	
Han	627 (91.8)
Other	56 (8.2)
Single child, n (%)[a]	
Yes	408 (60.2)
No	270 (39.8)
Age, mean (SD)	9.62 (0.35)
BMI, mean (SD)	18.54 (3.70)
BMI *z* score, mean (SD)	0.70 (1.44)
With obesity, n (%)	
Yes	158 (23.1)
No	526 (76.9)
With overweight or obesity, n (%)	
Yes	266 (38.9)
No	418 (61.1)
Waist circumference (cm), mean (SD)	65.15 (10.23)
Waist-to-height ratio, mean (SD)	0.86 (0.06)
BFP[c], mean (SD)	20.71 (10.43)
Caregivers, mean (SD)	
Mother age (y)	37.87 (4.24)
Father age (y)	39.88 (4.74)
Father BMI	24.20 (4.30)
Mother BMI	21.91 (4.04)
Fathers with overweight or obesity, n (%)[a]	
Yes	364 (57.2)
No	272 (42.8)
Mothers with overweight or obesity, n (%)[a]	
Yes	162 (24.7)
No	495 (75.3)
Fathers with obesity, n (%)[a]	
Yes	70 (11.0)
No	566 (89.0)
Mothers with obesity, n (%)[a]	
Yes	38 (5.8)
No	619 (94.2)
Primary caregiver, n (%)	
Mother	487 (71.2)

XSL•FO
RenderX

Characteristics	Values
Father	163 (23.8)
Nonparents	34 (5.0)
Father's education level, n (%)[a]	
High[b]	362 (54.4)
Low[b]	303 (45.6)
Mother's education level, n (%)[a]	
High[b]	397 (59.8)
Low[b]	267 (40.2)
Mother with the job, n (%)[a]	
Yes	555 (84.2)
No	104 (15.8)

[a]There is a missing value, so the synthesis is not 684, and the missing value appears to be missing at random.

[b]"High" means a college degree or above, and "low" means less than a college degree.

[c]BFP: body fat percentage.

Among caregivers, the mean ages of mothers and fathers were 37.87 (SD 4.24) and 39.88 (SD 4.74) years, respectively. The prevalences of obesity or overweight were 5.8% (n=38) and 24.7% (n=162) among mothers and 11.0% (n=70) and 57.2% (n=364) among fathers. Most primary caregivers were mothers (n=508, 74.3%), whose education levels were slightly higher than those of fathers (higher education: n=397, 59.8% vs n=362, 54.4%). Additionally, 84.2% (n=555) of mothers were employed.

Association Between Caregiver App-Use Adherence and Obesity-Related Outcomes

Caregiver app-use adherence was quantified using 2 app-recorded measures: app-use frequency and app-use duration. For app-use frequency, the corresponding median values were Q1, 32 (IQR 2 - 44) events; Q2, 57 (IQR 45 - 66) events; Q3, 81 (IQR 67 - 103) events; and Q4, 133 (IQR 104 - 529) events. For app-use duration, the quartile ranges were Q1, 17.40 (IQR 1.04 - 24.58) minutes; Q2, 31.10 (IQR 24.59 - 37.64) minutes; Q3, 46.60 (IQR 37.65 - 58.77) minutes; and Q4, 82.20 (IQR 58.78 - 240.04) minutes.

In the dichotomized analysis (Table 2), higher app-use frequency was associated with more favorable changes in BMI (β=−0.19, 95% CI −0.32 to −0.06; *P*=.005), BMI *z*-score (β=−0.08, 95% CI −0.13 to −0.02; *P*=.006), BFP (β=−0.70, 95% CI −1.19 to −0.21; *P*=.005), and rope-skipping performance (β=5.66, 95% CI 2.01-9.30; *P*=.002). Higher app-use duration was associated with lower BMI (β=−0.15, 95% CI −0.28 to −0.01; *P*=.03), BMI *z*-score (β=−0.06, 95% CI −0.12 to −0.01; *P*=.02), BFP (β=−0.51, 95% CI −1.01 to −0.02; *P*=.04), and better rope skipping performance (β=4.99, 95% CI 1.32-8.65; *P*=.008). No significant associations were found for WC, shuttle run time, or standing long jump distance.

XSL•FO
RenderX

Table . Association between binary caregiver and child adherence measures and child health outcomes.

Outcome	App usage frequency		App usage duration		Behavior score	
	β (95% CI)	*P* value	β (95% CI)	*P* value	β (95% CI)	*P* value
△[a]BMI	−0.19 (−0.32 to −0.06)	.005	−0.15 (−0.28 to −0.01)	.03	−0.07 (−0.20 to 0.07)	.34
△BMI *z*-score	−0.08 (−0.13 to −0.02)	.006	−0.06 (−0.12 to −0.01)	.02	−0.02 (−0.07 to 0.04)	.57
△WC[b]	−0.34 (−0.84 to 0.16)	.19	−0.33 (−0.84 to 0.17)	.18	−0.36 (−0.86 to 0.15)	.17
△BFP[c]	−0.70 (−1.19 to −0.21)	.005	−0.51 (−1.01 to −0.02)	.04	−0.63 (−1.13 to −0.14)	.01
△shuttle run time	−0.66 (−2.72 to 1.40)	.53	−0.28 (−2.35 to 1.79)	.79	−3.63 −5.66 to −1.60)	<.001
△ standing long jump	−0.01 (−2.06 to 2.04)	≥.99	0.28 (−1.79 to 2.34)	.79	2.88 (0.85 to 4.91)	.005
△1-minute rope jumping	5.66 (2.01 to 9.30)	.002	4.99 (1.32 to 8.65)	.008	4.06 (0.45 to 7.67)	.03

[a]△ indicates the change from baseline to the end of the intervention, calculated as the postintervention value minus the baseline value.

[b]WC: waist circumference.

[c]BFP: body fat percentage.

Caregiver adherence measures refer to app-recorded app-use frequency and total usage duration. Child adherence refers to the average weekly caregiver-child behavior-monitoring score for diet, screen time, and at-home exercise. *P* values were estimated from linear mixed-effects models adjusted for age, gender, parental age, parental BMI, parental education level, and the baseline value of each outcome, with class as a random effect.

In the quartile-based analysis (Table S3 in Multimedia Appendix 1), a higher app-use frequency was significantly associated with more favorable changes in BMI, BMI *z*-score, WC, BFP, and rope-skipping performance (*P* for trend =.01, .01, .02, .01, and .009, respectively). Higher app-use duration was also significantly associated with more favorable changes in BMI, BMI *z*-score, WC, and rope-skipping performance (*P* for trend=.02, .01, .03, and .004, respectively), whereas the association with BFP was not statistically significant (*P* for trend =.08).

Association Between Child Adherence and Obesity-Related Outcomes

For child behavioral adherence scores, quartile ranges were Q1, 29.0 (IQR 22.25 - 29.95); Q2, 30.80 (IQR 29.96 - 31.54); Q3, 32.40 (IQR 31.55 - 33.14); and Q4, 34.10 (IQR 33.15 - 35). In the dichotomized analysis (Table 2), higher child behavioral adherence was associated with lower BFP (β=−0.63, 95% CI −1.13 to −0.14; *P*=.01), shorter shuttle-run time (β=−3.63, 95% CI −5.66 to −1.60; *P*<.001), longer standing long-jump distance (β=2.88, 95% CI 0.85-4.91; *P*=.005), and greater rope-skipping counts (β=4.06, 95% CI 0.45-7.67; *P*=.03). No statistically significant associations were observed between child behavioral adherence and BMI, BMI *z*-score, or WC.

Further analyses using quartile-based child adherence groups also revealed a dose-response relationship for several outcomes (Table S4 in Multimedia Appendix 1). Compared with children in the lowest quartile, those in the highest quartile of behavioral adherence had more favorable changes in BFP (β=−0.78, 95% CI −1.49 to −0.08; *P* for trend =.009), shuttle run time (β=−4.17, 95% CI −7.05 to −1.28; *P* for trend =.008), standing long jump distance (β=4.07, 95% CI 1.20-6.93; *P* for trend =.004), and rope-skipping counts (β=6.14, 95% CI 1.04-11.25; *P* for trend =.010). No significant trends were observed for BMI, BMI *z*-score, or WC.

Comparison of App-Use Patterns: Frequency Vs Duration

As shown in Table 3, children whose caregivers belonged to group 4 (high frequency and high duration) showed the most favorable changes in several obesity-related outcomes, including BMI (β=−0.21, 95% CI −0.36 to −0.06; *P*<.01), BMI *z*-score (β=−0.09, 95% CI −0.15 to −0.03; *P*<.05), and BFP (β=−0.76, 95% CI −1.32 to −0.21; *P*<.01), compared with the reference group (low frequency and low duration). Favorable changes in physical fitness, particularly rope skipping, were also observed across multiple groups with higher app-use frequency or duration.

XSL•FO
RenderX

Table . Joint associations of app-use frequency and duration with child health outcomes.

Outcome	Group 1[a] (Ref)	Group 2[b], β (95% CI)	Group 3[c], β (95% CI)	Group 4[d], β (95% CI)	*P* value for interaction
△BMI	1.00	−0.11 (−1.16 to 2.53)	0.00 (−0.23 to 0.23)	−0.21 (−0.36 to −0.06)[e]	.55
△BMI z-score	1.00	−0.03 (−0.13 to 0.06)	−0.00 (−0.10 to 0.09)	−0.09 (−0.15 to −0.03)[f]	.43
△WC[g]	1.00	.04 (−0.83 to 0.92)	0.06 (−0.82 to 0.95)	−0.42 (−0.99 to 0.15)	.40
△BFP[h]	1.00	−0.47 (−1.34 to 0.39)	−0.00 (−0.87 to 0.86)	−0.76 (−1.32 to −0.21)[e]	.64
△shuttle run time	1.00	−1.25 (−4.82 to 2.32)	−0.34 (−4.00 to 3.31)	−0.59 (−2.92 to 1.73)	.70
△standing long jump	1.00	1.55 (−2.03 to 5.12)	0.19 (−2.12 to 2.49)	2.25 (−1.34 to 5.84)	.16
△1-minute rope jumping	1.00	6.74 (2.64 to 10.83)[e]	7.06 (0.66 to 13.46)	8.52 (2.26 to 14.77)[e]	.048

[a]Group 1: low frequency and low duration.

[b]Group 2: high frequency and low duration (frequent but short sessions).

[c]Group 3: low frequency and high duration (infrequent but longer sessions).

[d]Group 4: high frequency and high duration.

[e]*P*<.01.

[f]*P*<.05.

[g]WC: waist circumference.

[h]BFP: body fat percentage.

To further compare app-use frequency and duration patterns, we compared group 2 (high frequency and low duration) with Group 3 (low frequency and high duration). No statistically significant differences were found across the health indicators. Although the interaction term between app-use frequency and duration was not significant for most outcomes, it was significant for rope-skipping performance (*P*=.048), suggesting a possible joint association of frequent and prolonged app use with this physical fitness outcome.

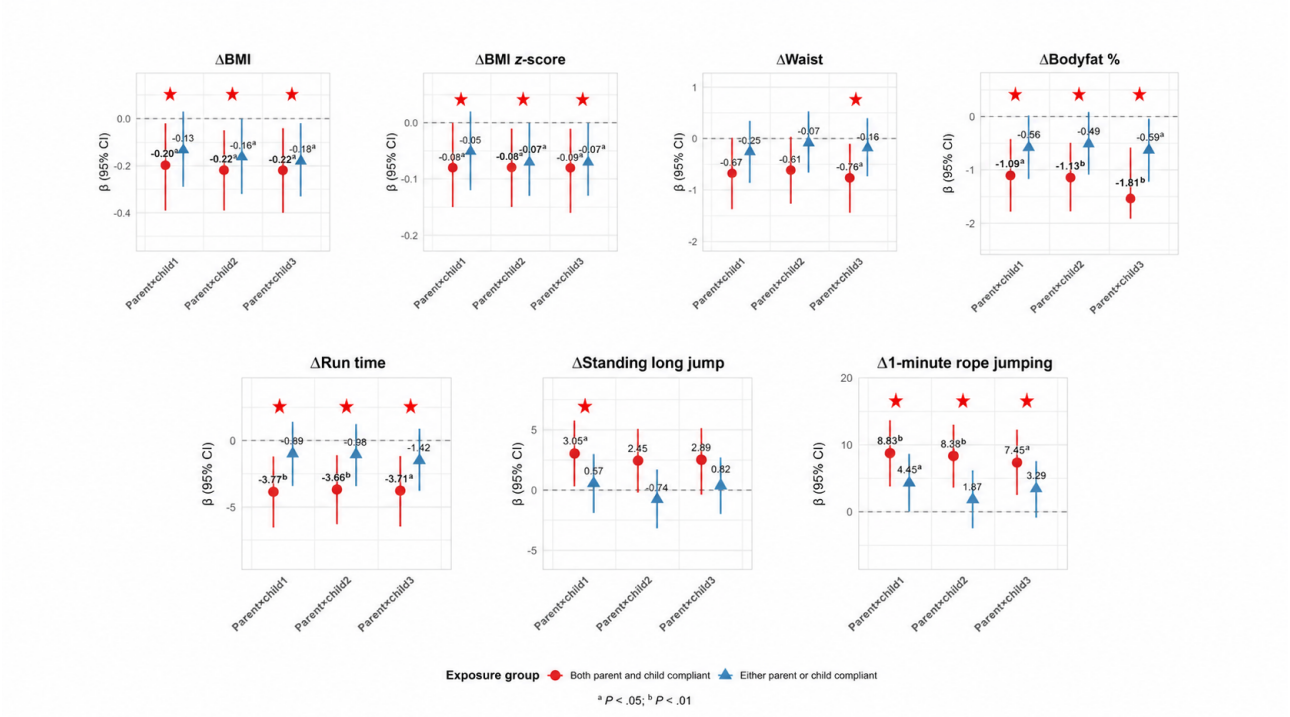
Joint Associations of Caregiver and Child Adherence and Trend Analyses

To explore the combined association of caregiver app-use adherence and child behavioral adherence, we constructed three joint adherence variables: app-use duration × child behavioral adherence, app-use frequency × child behavioral adherence, and combined app-use pattern × child behavioral adherence. Each joint variable was categorized into three groups: both low adherence, mixed adherence, and both high adherences.

As shown in Figure 1 and Table S5 in Multimedia Appendix 1, children in both high-adherence groups showed the most favorable outcomes across several obesity-related and physical-fitness indicators. Compared with the both low-adherence group, the both high-adherence group had more favorable changes in BMI (eg, app-use pattern × child adherence: β=−0.22, 95% CI −0.40 to −0.04; *P*=.02), BMI *z*-score (β=−0.08, 95% CI −0.16 to −0.01; *P*=.03), and BFP (β=−1.18, 95% CI −1.85 to −0.51; *P*=.001). Physical fitness measures also showed favorable changes; for example, the both high adherence group had a shorter shuttle run time (β=−3.71, 95% CI −6.50 to −0.93; *P*=.009) and more rope-skipping repetitions (β=7.45, 95% CI 2.50 to 12.40; *P*=.003) in the app-use pattern × child adherence model.

XSL•FO
RenderX

Figure 1. Joint associations of caregiver app use and child behavioral adherence with obesity-related outcomes.



The figure shows adjusted β coefficients and 95% CIs for the joint associations of caregiver app use and child behavioral adherence with seven obesity-related outcomes. Three caregiver app-use indicators were examined: app-use duration, app-use frequency, and combined high-frequency or high-duration app use. For each indicator, participants were classified into 3 groups: both low adherence (reference), mixed adherence, and both high adherence. Blue triangles indicate mixed adherence, and red circles indicate high adherence. Red stars indicate significant trends across adherence groups (*P* for trend<.05).

Trend analyses demonstrated significant dose–response relationships for most outcomes, indicating that higher levels of joint caregiver–child adherence were associated with progressively greater improvements in both adiposity and physical fitness indicators. Significant interaction effects between caregiver and child adherence were observed for BMI and standing long jump performance, suggesting that the association of one adherence component with these outcomes may vary according to the level of the other adherence component.

Sensitivity Analysis

To assess the robustness of the findings, we conducted sensitivity analyses by excluding participants who were primarily cared for by grandparents. The results, as shown in Table S6 in Multimedia Appendix 1, were largely consistent with the main analyses, supporting the stability of our conclusions.

Discussion

Principal Findings

To our knowledge, this was the first study to simultaneously examine caregiver app-use adherence, child behavioral adherence, and their joint associations with obesity-related outcomes in a multicomponent, app-supported childhood obesity intervention. Caregiver app-use adherence was associated mainly with adiposity-related outcomes: higher app-use frequency and longer app-use duration were consistently associated with more favorable changes in BMI, BMI *z*-score, and BFP, with graded patterns across adherence levels. Child behavioral adherence was more strongly associated with physical fitness indicators. However, given the large number of statistical comparisons performed, these statistically significant associations should be interpreted as exploratory and hypothesis-generating rather than confirmatory. This caution is particularly important for interaction analyses. The statistically significant interaction between app-use frequency and duration for rope-skipping performance suggests that both dimensions of app use may be relevant to this physical fitness outcome, but this finding requires confirmation in future studies. Similarly, the observed interaction terms between caregiver and child adherence suggest that their joint associations may differ from either component alone, although these findings should also be regarded as exploratory. Children with both high caregiver app-use adherence and high child behavioral adherence tended to show more favorable outcomes across several indicators, suggesting the potential importance of family coadherence in digital childhood obesity interventions.

Comparison With Prior Work

Participant adherence has been widely recognized as a key factor related to the effectiveness of obesity interventions. Previous studies have adopted various metrics to assess adherence and have generally reported more favorable outcomes among more adherent participants. One study used caregivers' questionnaire completion as a measure of adherence and found that children of caregivers who complied experienced greater weight loss [17]. Another study grouped participants according to the

XSL•FO
RenderX

number of educational sessions attended and found that participants in the adherent group (attended ≥3 sessions) showed significantly greater improvements in obesity-related indicators compared with those with lower attendance [18]. In interventions using interactive voice response technology, higher engagement, measured by the number of completed calls, was associated with reductions in child weight, BMI, and BMI *z*-scores [19,20]. Our findings provide additional evidence that caregiver app-use adherence and child behavioral adherence were both associated with more favorable obesity-related outcomes. These results support the importance of considering engagement and adherence in the design and implementation of childhood obesity interventions.

Both app-use frequency and app-use duration appear to be important components of digital intervention dose. Our findings suggest that these two dimensions were jointly associated with obesity-related outcomes. Although childhood obesity interventions often require substantial cumulative contact time, such as the suggested 26 hours in the first intervention year [21], the challenge lies in delivering this dose in formats acceptable to families. Evidence indicates that brief but frequent interactions may be more feasible and potentially more effective than longer, infrequent sessions. A dietary education study replaced traditional intensive dietary counseling with short, frequent telephone contacts, an approach that elicited positive participant feedback and meaningful lifestyle improvements [22]. Given increasing time constraints in modern family life, adherence becomes a central barrier. Transitioning from long-format counseling to high-frequency, short-duration digital touchpoints may improve adherence while preserving intervention dose, offering a scalable and family-centered strategy for real-world childhood obesity interventions.

The sensitivity analysis excluding children primarily cared for by grandparents showed generally similar associations, suggesting that the main findings were unlikely to be driven solely by this relatively small caregiver subgroup. One possible explanation is that grandparents caring for primary school-aged children may still have sufficient familiarity with smartphones to participate in app-based intervention activities. Nevertheless, caregiver heterogeneity should be considered in future mHealth interventions. Parents and grandparents may differ in smartphone familiarity, app-use habits, and digital health literacy, which could influence intervention engagement and behavior monitoring. Prior reviews have shown that older adults' use of eHealth or health technologies can be affected by self-efficacy, knowledge, support, technology usability, and socioeconomic factors [23,24]. Future mHealth interventions may therefore need tailored support for caregivers with different levels of technological familiarity.

The joint high-adherence pattern in both caregivers and children was associated with the most favorable changes in obesity-related and physical fitness outcomes, highlighting the importance of engaging both parties in digital health interventions. The different patterns observed for caregiver and child adherence are consistent with the design of our multicomponent intervention. The intervention app was primarily designed for caregivers to receive nutritional guidance. This may explain why caregiver app-use adherence was more strongly associated with weight-related indicators, such as BMI and BFP. By providing actionable dietary advice and tools for home environment modification, the app supported caregivers—who often influence food purchasing, meal preparation, and children's eating opportunities—in implementing healthier eating practices for their children. This interpretation is consistent with evidence that parent or family involvement and parent-targeted dietary support are important components of childhood obesity interventions [25,26]. In contrast, children's behavioral adherence was more strongly associated with physical fitness improvements. This aligns with the student-level component of our intervention, which integrated enhanced, structured physical activity into the school curriculum. Higher child behavioral adherence may reflect more consistent practice of recommended dietary, screen time, and physical activity behaviors [27]. Overall, our results suggest that childhood obesity interventions may benefit from supporting both caregiver involvement and children's adherence to targeted health behaviors. Future interventions should incorporate strategies that help caregivers create supportive home environments while encouraging children to practice and sustain healthy dietary and physical activity behaviors.

Strengths and Limitations

The main strength of this study lies in the richness and objectivity of its process evaluation data and the objective recording of caregiver app use. Caregiver app-use data were automatically recorded by the mobile app, providing objective, time-stamped information on app-use frequency and duration. Children's health behaviors were repeatedly and systematically collected through the caregiver-child behavior-monitoring module, allowing us to examine children's behavioral adherence over the intervention period. Compared with studies relying solely on qualitative process evaluation or retrospective end-of-study self-reports, the repeated app-based data collection in this study provided more detailed quantitative information for adherence analysis [28]. Moreover, we considered adherence from both the perspectives of the intervention implementers (caregivers) and the intervention subjects (children). By simultaneously evaluating caregiver app-use adherence and children's behavioral adherence, we were able to examine the interplay between the 2 and better understand how selected adherence-related processes were associated with intervention outcomes.

Several limitations should be acknowledged. First, the baseline prevalence of obesity and overweight or obesity in our analytical sample was relatively high compared with national estimates for Chinese children of similar age. This should be interpreted cautiously because the present analysis was based on a trial-based regional sample rather than a nationally representative sample. Previous Chinese surveillance studies have reported substantial geographic, urban-rural, sex, and socioeconomic variation in childhood overweight and obesity, with some school populations showing prevalence estimates higher than national averages [29]. Second, although intervention fidelity encompasses multiple components—including adherence, dose, quality of delivery, participant responsiveness, and program differentiation [30], the present analysis focused on the dose of caregiver app-use

XSL•FO
RenderX

and child behavioral adherence. Process evaluation data from the DECIDE-Children trial showed that several school-level fidelity components, such as delivery of health education sessions and implementation of school-based activities, consistently achieved high levels (>80%) with limited variability across schools and participants [14]. Therefore, these indicators were less suitable for explaining individual differences in obesity-related outcomes in the present analysis. Nevertheless, other potentially informative fidelity or engagement indicators were not fully captured, such as teacher implementation quality, children's direct participation in school-based activities, and children's engagement with intervention content. Thus, our findings should be interpreted as reflecting selected adherence-related processes rather than the full implementation fidelity of the intervention. Third, the temporal ordering between caregiver app engagement, child behavioral adherence, and child obesity-related outcomes could not be firmly established in this secondary analysis. Although adherence indicators were assessed during the intervention period and outcomes were evaluated over follow-up, caregiver engagement may both influence and be influenced by children's behavioral changes or weight-related progress. For example, caregivers of children who showed more favorable changes may have been more motivated to continue using the app, whereas caregivers facing greater difficulties may have disengaged over time. Therefore, reverse causality cannot be excluded, and the observed associations should not be interpreted as causal effects of adherence. Fourth, child behavioral adherence was measured using caregiver-child behavior monitoring reported through the app. Although caregivers were instructed to observe and/or inquire about their child's weekly behaviors and preferably complete the report together with the child, this measure remains proxy-reported and may reflect caregiver perceptions. In addition, caregiver app use and caregiver-reported child behavioral adherence were derived from the same respondent, which may introduce common-method bias and inflate the observed associations between caregiver and child adherence indicators. Caregivers who used the app more frequently may also have been more diligent in reporting children's behaviors, which could introduce reporting-related correlation. Future studies should incorporate more direct and multidimensional measures of child engagement, including school-level participation records, direct activity monitoring, and children's own reports where appropriate. Finally, because multiple statistical comparisons were conducted across several adherence indicators and obesity-related outcomes, individual statistically significant findings, particularly interaction terms, should be interpreted as exploratory and hypothesis-generating rather than confirmatory.

Implications and Future Research

Future studies should evaluate adherence to better interpret the true effectiveness of complex interventions. In our study, children with higher caregiver and child adherence levels showed more favorable changes, suggesting that intervention effects may be underestimated when adherence is not considered—an issue aligned with the concept of "type III error," in which poor implementation rather than ineffective intervention leads to null findings [31]. Evidence from prior childhood obesity interventions similarly shows that meaningful benefits emerge when implementation fidelity is moderate to high [32]. Therefore, routine collection of detailed adherence data and integration of implementation evaluation frameworks are essential for accurately assessing intervention impact and informing policy and practice [33].

Conclusion

In this secondary analysis of an mHealth app-assisted childhood obesity intervention, higher caregiver app-use adherence and children's behavioral adherence were associated with more favorable changes in weight status and physical fitness. These findings underscore the importance of coengagement in digital health programs and highlight the potential of app-supported, family-engaged strategies for childhood obesity prevention.

Acknowledgments

We thank all the children and caregivers in the diet, exercise, and cardiovascular health children's study for their participation. We also appreciate the teachers involved in the program and thank all the investigators for their help throughout this study.

Generative AI (ChatGPT, OpenAI) was used during manuscript preparation for language editing, improving clarity of expression, and assisting with responses during the revision and copyediting process. It was not used for data analysis, generation or interpretation of study results, or selection of references. All AI-assisted content was reviewed and verified by the authors, who take full responsibility for the accuracy, originality, and integrity of the manuscript.

Funding

This work was supported by the Shandong Province Traditional Chinese Medicine Science and Technology Project [Z20243204] and the National Key R&D Program of China (2016YFC1300204).

Data Availability

The datasets generated or analyzed during this study are not publicly available due to participant privacy and confidentiality considerations but are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

SY conducted the data analysis and drafted the initial manuscript. JL, CY, and WL contributed to the revision and refinement of the manuscript. They provided supervisory oversight, critical review, and essential resources throughout the study. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
App intervention, adherence measures, and associations of caregiver and child adherence with child health outcomes.
[DOCX File, 36 KB - mhealth_v14i1e89548_app1.docx]

Checklist 1
CONSORT-EHEALTH checklist.
[PDF File, 1174 KB - mhealth_v14i1e89548_app2.pdf]

References

1. Obesity and overweight. World Health Organization (WHO). 2024. URL: https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2026-08-08]
2. Dong YH, Chen L, Liu JY, et al. Epidemiology and prediction of overweight and obesity among children and adolescents aged 7-18 years in China from 1985 to 2019 [Article in Chinese]. Zhonghua Yu Fang Yi Xue Za Zhi 2023 Feb 28;57(4):11-19. [doi: 10.3760/cma.j.cn112150-20220906-00881] [Medline: 36854438]
3. Weihrauch-Blüher S, Schwarz P, Klusmann JH. Childhood obesity: increased risk for cardiometabolic disease and cancer in adulthood. Metabolism 2019 Mar;92:147-152. [doi: 10.1016/j.metabol.2018.12.001] [Medline: 30529454]
4. Nonguierma E, Lesco E, Olak R, et al. Improving obesogenic dietary behaviors among adolescents: a systematic review of randomized controlled trials. Nutrients 2022 Nov 1;14(21):4592. [doi: 10.3390/nu14214592] [Medline: 36364855]
5. Brown T, Moore TH, Hooper L, et al. Interventions for preventing obesity in children. Cochrane Database Syst Rev 2019 Jul 23;7(7):CD001871. [doi: 10.1002/14651858.CD001871.pub4] [Medline: 31332776]
6. Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S. A conceptual framework for implementation fidelity. Implement Sci 2007 Nov 30;2:40. [doi: 10.1186/1748-5908-2-40] [Medline: 18053122]
7. JaKa MM, Haapala JL, Trapl ES, et al. Reporting of treatment fidelity in behavioural paediatric obesity intervention trials: a systematic review. Obes Rev 2016 Dec;17(12):1287-1300. [doi: 10.1111/obr.12464] [Medline: 27612933]
8. Nguyen B, Shrewsbury VA, O'Connor J, et al. A process evaluation of an adolescent weight management intervention: findings and recommendations. Health Promot Int 2015 Jun;30(2):201-212. [doi: 10.1093/heapro/dau110] [Medline: 25550288]
9. Kay M, Santos J, Takane M. mHealth: new horizons for health through mobile technologies: based on the findings of the second global survey on ehealth. : World Health Organization (WHO); 2011 URL: https://www.gfmer.ch/mhealth/coursefiles2013/WHO_mHealthReport_GlobalObservatoryforeHealthseries.pdf [accessed 2026-08-08]
10. Wang Y, Xue H, Huang Y, Huang L, Zhang D. A systematic review of application and effectiveness of mHealth interventions for obesity and diabetes treatment and self-management. Adv Nutr 2017 May;8(3):449-462. [doi: 10.3945/an.116.014100] [Medline: 28507010]
11. Wang H, Yan S. Obesity intervention studies in children and adolescents need compliance evaluation. Chin J Sch Health 2023;44(4):481-484. [doi: 10.16835/j.cnki.1000-9817.2023.04.001]
12. Xu F, Ware RS, Leslie E, et al. Effectiveness of a randomized controlled lifestyle intervention to prevent obesity among Chinese primary school students: CLICK-Obesity Study. PLoS One 2015;10(10):e0141421. [doi: 10.1371/journal.pone.0141421] [Medline: 26510135]
13. Liu Z, Wu Y, Niu WY, et al. A school-based, multi-faceted health promotion programme to prevent obesity among children: protocol of a cluster-randomised controlled trial (the DECIDE-Children study). BMJ Open 2019 Nov 2;9(11):e027902. [doi: 10.1136/bmjopen-2018-027902] [Medline: 31678935]
14. Liu Z, Gao P, Gao AY, et al. Effectiveness of a multifaceted intervention for prevention of obesity in primary school children in China: a cluster randomized clinical trial. JAMA Pediatr 2022 Jan 1;176(1):e214375. [doi: 10.1001/jamapediatrics.2021.4375] [Medline: 34747972]
15. Yan S, Li W, Zheng M, et al. Identifying longitudinal compliance patterns and determinants in a multifaceted childhood obesity intervention using group-based trajectory modeling. Nutrients 2025 May 16;17(10):1701. [doi: 10.3390/nu17101701] [Medline: 40431440]
16. de Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. Bull World Health Organ 2007 Sep;85(9):660-667. [doi: 10.2471/blt.07.043497] [Medline: 18026621]

XSL•FO
RenderX

17. Yackobovitch-Gavan M, Nagelberg N, Phillip M, Ashkenazi-Hoffnung L, Hershkovitz E, Shalitin S. The influence of diet and/or exercise and parental compliance on health-related quality of life in obese children. Nutr Res 2009 Jun;29(6):397-404. [doi: 10.1016/j.nutres.2009.05.007] [Medline: 19628106]
18. Al-Daghri NM, Amer OE, Khattak MNK, et al. Attendance-based adherence and outcomes of obesity management program in Arab adolescents. Children (Basel) 2023 Aug 25;10(9):1449. [doi: 10.3390/children10091449] [Medline: 37761410]
19. Wright JA, Phillips BD, Watson BL, Newby PK, Norman GJ, Adams WG. Randomized trial of a family-based, automated, conversational obesity treatment program for underserved populations. Obesity (Silver Spring) 2013 Sep;21(9):E369-E378. [doi: 10.1002/oby.20388] [Medline: 23512915]
20. Estabrooks PA, Shoup JA, Gattshall M, Dandamudi P, Shetterly S, Xu S. Automated telephone counseling for parents of overweight children: a randomized controlled trial. Am J Prev Med 2009 Jan;36(1):35-42. [doi: 10.1016/j.amepre.2008.09.024] [Medline: 19095163]
21. O'Connor EA, Evans CV, Burda BU, Walsh ES, Eder M, Lozano P. Screening for obesity and intervention for weight management in children and adolescents: evidence report and systematic review for the US Preventive Services Task Force. JAMA 2017 Jun 20;317(23):2427-2444. [doi: 10.1001/jama.2017.0332] [Medline: 28632873]
22. Brekke HK, Sunesson A, Lenner RA. Unannounced telephone interviews: a useful and positively received tool in the reinforcement of lifestyle intervention. Patient Prefer Adherence 2009 Dec 29;3:357-362. [doi: 10.2147/ppa.s6812] [Medline: 20054441]
23. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health 2021 Aug 17;21(1):1556. [doi: 10.1186/s12889-021-11623-w] [Medline: 34399716]
24. Bertolazzi A, Quaglia V, Bongelli R. Barriers and facilitators to health technology adoption by older adults with chronic diseases: an integrative systematic review. BMC Public Health 2024 Feb 16;24(1):506. [doi: 10.1186/s12889-024-18036-5] [Medline: 38365698]
25. Nezami BT, Wasser HM, Tate DF. Parent and child dietary changes in a 6-month mobile-delivered weight loss intervention with tailored messaging for parents. Front Public Health 2022;10:972109. [doi: 10.3389/fpubh.2022.972109] [Medline: 36225761]
26. Kobes A, Kretschmer T, Timmerman G, Schreuder P. Interventions aimed at preventing and reducing overweight/obesity among children and adolescents: a meta-synthesis. Obes Rev 2018 Aug;19(8):1065-1079. [doi: 10.1111/obr.12688] [Medline: 29671938]
27. Nastaskin RS, Fiocco AJ. A survey of diet self-efficacy and food intake in students with high and low perceived stress. Nutr J 2015 Apr 23;14:42. [doi: 10.1186/s12937-015-0026-z] [Medline: 25902797]
28. Grant A, Treweek S, Dreischulte T, Foy R, Guthrie B. Process evaluations for cluster-randomised trials of complex interventions: a proposed framework for design and reporting. Trials 2013 Jan 12;14:15. [doi: 10.1186/1745-6215-14-15] [Medline: 23311722]
29. Song X, Zhou B, Baird S, et al. Trends and inequalities in thinness and obesity among Chinese children and adolescents: evidence from seven national school surveys between 1985 and 2019. Lancet Public Health 2024 Dec;9(12):e1025-e1036. [doi: 10.1016/S2468-2667(24)00211-1] [Medline: 39481418]
30. Dusenbury L, Brannigan R, Falco M, Hansen WB. A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. Health Educ Res 2003 Apr;18(2):237-256. [doi: 10.1093/her/18.2.237] [Medline: 12729182]
31. Dobson D, Cook TJ. Avoiding type III error in program evaluation: results from a field experiment. Eval Program Plann 1980;3(4):269-276. [doi: 10.1016/0149-7189(80)90042-7]
32. Liu Z, Li Q, Maddison R, et al. A school-based comprehensive intervention for childhood obesity in China: a cluster randomized controlled trial. Child Obes 2019;15(2):105-115. [doi: 10.1089/chi.2018.0251] [Medline: 30565955]
33. French C, Dowrick A, Fudge N, Pinnock H, Taylor SJC. What do we want to get out of this? A critical interpretive synthesis of the value of process evaluations, with a practical planning framework. BMC Med Res Methodol 2022 Nov 25;22(1):302. [doi: 10.1186/s12874-022-01767-7] [Medline: 36434520]

Abbreviations

BFP: body fat percentage
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
DECIDE-Children: diet, exercise, and cardiovascular health—children
mHealth: mobile health
WC: waist circumference
WHO: World Health Organization

XSL•FO
RenderX

Edited by G Badicu; submitted 14.Dec.2025; peer-reviewed by M Henström, O Owolo; revised version received 12.Jul.2026; accepted 13.Jul.2026; published 24.Sep.2026.

Please cite as:
Yan S, Lyu J, Yin C, Li W, Li Y, Wang HJ
Caregiver App Use and Child Behavioral Adherence in Relation to Obesity-Related Outcomes in an mHealth App-Assisted Childhood Obesity Intervention: Secondary Analysis of a Cluster Randomized Trial
JMIR Mhealth Uhealth 2026;14:e89548
URL: https://mhealth.jmir.org/2026/1/e89548
doi:10.2196/89548

© Shiyu Yan, Jinlang Lyu, Chunjie Yin, Wenhao Li, Yan Li, Hai-jun Wang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 24.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

How People Integrate Mental Health Apps Into Everyday Self-Care: Qualitative Study

Vít Suchý[1*], MA; Zuzana Scott[1*], PhD; Barbora Boštíková[2], MA; Marek Janoušek[2], MA; Eliška Josková[2], BC; Anna Ševčíková[1*], PhD

[1]Psychology Research Institute, Faculty of Social Studies, Masaryk University, Jostova 10, Brno, Czechia

[2]Department of Psychology, Faculty of Social Studies, Masaryk University, Brno, Czechia

[*]these authors contributed equally

Corresponding Author:
Anna Ševčíková, PhD
Psychology Research Institute, Faculty of Social Studies, Masaryk University, Jostova 10, Brno, Czechia

Abstract

Background: Mobile mental health (MH) apps are increasingly promoted as tools for self-care. However, less is known about how users integrate them into everyday life and how they combine with other forms of support.

Objective: This qualitative study examined what motivates individuals to adopt MH apps and how these tools are used alongside other care practices.

Methods: We conducted semistructured interviews with 31 Czech MH app users aged 18‐56 (23 women), recruited through social media, researchers' networks, and snowball sampling. Eligible participants had used an MH app for at least 2 weeks; individuals currently experiencing a psychological crisis were excluded. Interviews were conducted online or face-to-face and analyzed using reflexive thematic analysis by Braun and Clarke. The analysis was inductive and informed by a critical realist epistemological stance. Initial analytic engagement was conducted independently, followed by collaborative theme development and refinement.

Results: We identified the following three motivational patterns for MH app use: (1) regular use to maintain well-being and support preventive self-care; (2) reactive use during periods of acute distress; and (3) temporary, problem-focused use aimed at understanding or addressing a specific difficulty. Beyond motivations for adoption, our findings revealed the following four distinct trajectories of app engagement and disengagement: (1) emancipation, with apps acting as ongoing support for self-care when therapy is reduced or ended; (2) independence, as techniques learned through apps are transferred into everyday life and used without the app; (3) reliance, where users persist with app-based self-care despite insufficient benefit, occasionally delaying or limiting other forms of help; and (4) episodic use, characterized by intermittent engagement during specific life periods. These findings extend existing research by showing that disengagement is not necessarily a marker of failure but can reflect successful self-care and the transfer of learned skills. We also identified tensions and unintended consequences, including the burden of sustaining engagement, risks of overreliance, emotional costs associated with sustained self-monitoring, and tensions between digital self-care and efforts to reduce screen use. Participants generally viewed self-monitoring as beneficial when guided by clear personal goals and embedded within a broader system of self-care.

Conclusions: Our findings suggest that appropriate disengagement can signal a successful step in self-care. App design and policy should acknowledge episodic use and strengthen connections to human support with clear guidance and accessible pathways to additional care when needed.

(JMIR Mhealth Uhealth 2026;14:e95414) doi:10.2196/95414

KEYWORDS

mental health app; self-care; mental health; adults; qualitative study; mobile applications; self-management; patient engagement; habit

Introduction

Background

A growing number of studies point to a rising prevalence of mental health (MH) problems in the adult population [1-3]. In parallel, broader social processes of individualization in Western countries increasingly frame MH and well-being as an individual responsibility [4]. This orientation is reflected in the individualization of care, whereby individuals are expected to monitor their well-being and manage psychological difficulties largely on their own [5]. Recent technological advances have further reinforced this process through the rapid development

XSL•FO
RenderX

of stand-alone MH smartphone apps designed to improve users' well-being.

These trends matter most where formal care is hard to access. In Czechia, roughly 80% of adults meeting criteria for a current mental disorder receive no treatment, and over half of those seeking help encounter structural barriers [6]. Public stigma also remains high by European standards, with little recent change even as recognition of conditions like depression and anxiety has modestly improved [7]. In such contexts, MH apps may be adopted out of necessity rather than personal preference.

An expanding body of systematic reviews suggests that these digital tools may have moderate positive effects on reducing negative MH symptoms, improving well-being and emotional regulation [8-13], particularly with regard to depression and anxiety-related symptomatology. Despite this emerging evidence, numerous studies highlight persistent challenges related to low user engagement and factors that hinder the sustained use of MH apps [9,14,15]. Recent research therefore suggests that greater attention should be paid to personalization mechanisms within MH apps as promising strategies to enhance use of and perceived benefits from these digital tools [15,16]. However, significant gaps remain in our understanding of when and how apps benefit users and what role they play in broader self-care practices.

Users' Motivations and Experiences of MH App Use

Prior qualitative research has identified several factors motivating individuals to use MH apps. App properties that enable users to substitute emotional support usually provided by social interactions have often been highlighted. Chief among these is their immediate availability, which allows users to address psychological distress independent of the availability of social support. MH apps also provide a discreet space where users can address their MH needs. This privacy is particularly relevant in cases where stigma discourages the disclosure necessary for social support [17-19]. In addition to privacy, MH apps are often valued for their nonjudgmental stance. This provides some users with a sense of being understood and supported, mitigating feelings of loneliness and/or shame, while they remain aware of the app's nonhuman nature [17-19].

Beyond these substitutional properties, users are attracted by the diverse features (eg, breathing exercises and journaling) consolidated within single customizable platforms, which are often economically affordable [17,19-21]. This is an important factor for sustained app usage, particularly because MH apps have been found to be valuable for long-term MH maintenance and for allowing more precise and continuous tracking for mood and daily circumstances, thereby fostering self-awareness of the factors associated with changes in psychosocial functioning [15,18,22]

However, existing research indicates that engagement features, such as streaks and reminders, are often experienced ambivalently; while they may encourage routine for some, the perceived burden of continuous tasks can paradoxically discourage use [15,22]. Furthermore, concerns raised by MH app users relate to potential overreliance on app-based functions, particularly in the absence of complementary forms of support [19,23]. According to some users, this reliance contributes to the emergence of a new norm of independent coping, and failing to conform to this expectation leaves them with feelings of shame [24].

Current research also suggests that sustained, long-term use of MH apps for self-maintenance and health management requires a high level of self-initiative and intrinsic motivation [15]. These demands become compounded when users experience psychological distress, which may make it more difficult for them to familiarize themselves with app content or to effectively navigate available features [15]. In this regard, the involvement of external professional support from counselors and therapists to assist users in engaging with and interpreting MH app content has been identified as a potentially promising approach [19].

A recent qualitative study shifted the research focus from the sustained use of MH apps to macro-engagement, within which off-screen, goal-directed activities associated with behavioral change may continue beyond app use [25]. This study showed that on-screen interactions and off-screen, goal-directed practices often remained disconnected, which limited the extent to which behavioral change was supported in everyday life. The authors therefore argued that greater support for engagement beyond the screen was needed. However, aside from this emerging line of research, we know little about how people integrate MH apps into their broader MH care practices and how these digital tools interact with the other forms of support with which people engage.

Theoretical Framework for Self-Care

To examine the use of MH apps, we draw on the Middle-Range Theory of Self-Care of Chronic Illness (MRTSC) proposed by Riegel et al [26]. Within this framework, self-care is conceptualized as an ongoing process aimed at maintaining health and well-being in both healthy and ill states. The theory distinguishes the following three interrelated aspects: (1) *self-care maintenance*, which involves behaviors intended to preserve health and prevent symptom exacerbation; (2) *self-care monitoring*, which refers to the observation and tracking of bodily or psychological changes; and (3) *self-care management*, which encompasses responses to symptoms based on their detection and interpretation. This characterizes self-care as a decision-making process during which individuals reflect on their symptoms based on experience and contextual factors.

Although MH self-care differs from self-care in chronic illness in that needs and goals may fluctuate over time, the MRTSC framework provides a useful lens for understanding how individuals maintain well-being, monitor psychological states, and respond to perceived changes in their MH. The framework is particularly relevant for studying MH app use, as these apps can facilitate all 3 aspects of the self-care process. They support monitoring and reflection through features such as mood tracking or journaling; they enable self-care maintenance through guided relaxation or mindfulness practices; and they assist in self-care management through actionable coping strategies or crisis resources. Moreover, the MRTSC offers a process-oriented framework that is particularly suited to symptom-driven engagement and decision-making in MH self-care contexts. This perspective allows MH app use to be

XSL•FO
RenderX

understood as part of an ongoing self-care process alongside other forms of professional and informal support. In this respect, understanding the use of MH apps as part of self-care raises important questions about when, how, and for how long individuals integrate these tools into their everyday practices.

This Study

Existing research on MH apps remains largely descriptive and fragmented—motivations, user engagement, and disengagement are often examined in isolation. Moreover, disengagement is often interpreted as a negative outcome. Recent human-computer interaction research [27] offers an alternative perspective by conceptualizing disengagement as a natural cyclical phase of technology use that is driven by user agency and specific goals. Pausing or stopping use can be a deliberate positive decision rather than a breakdown in motivation or usability. Applying this framework may help explain how disengaging from MH apps may reflect successful self-care management.

Furthermore, while recent research has begun to highlight the disconnect between on-screen engagement and off-screen behavioral change [28], less attention has been paid to how MH apps are embedded within broader ecosystems of MH care and support. To address this gap, we draw on the MRTSC to examine how users engage with and disengage from MH apps and how these apps are embedded within broader MH self-care practices and support systems. To achieve this aim, the study addresses the following research questions (RQs):

- RQ1: What motivates individuals to engage with and disengage from the MH apps?
- RQ2: How are MH apps integrated into users' everyday lives and into broader ecosystems of MH care and support?

Methods

Participants and Procedure

We conducted 31 semistructured interviews with Czech users of mobile MH apps (aged 18 - 56, 23 women). Recruitment was initially conducted through a paid social media advertising campaign (Facebook [Meta] and Instagram [Meta]) targeting the general Czech-speaking population in Czechia aged 18 years and older. Owing to low response rates to the advertisement, recruitment was subsequently supplemented through the researchers' personal networks and further through snowball sampling. None of the participants had a prior relationship with the interviewers. We were transparent about the study's purpose, framing the conversation as an exploration of how people use these apps, their benefits, and the boundaries of their usefulness. We excluded individuals who were (1) 17 years old or younger, (2) currently in psychological crisis, or (3) survivors of a traumatic event within the preceding 12 months. The only inclusion criterion was having used a MH app for at least 2 weeks. Demographic data were collected before the interview (Table 1).

XSL•FO
RenderX

Table . Sample characteristics (n=31).

Characteristic	Value
Age (y), mean (SD)	28.81 (9.58)
Interview length (hours:minutes:seconds), mean (SD)	0:57:45 (0:17:05)
Gender, n (%)	
Women	23 (74.2)
Men	8 (25.8)
Nationality, n (%)	
Czech	30 (96.8)
Slovak	1 (3.2)
Size of municipality, n (%)	
>500,000	8 (25.8)
100,000 - 500,000	11 (35.5)
20,000 - 100,000	3 (9.7)
10,000 - 20,000	2 (6.5)
<10,000	7 (22.6)
Education, n (%)	
Primary	2 (6.5)
Secondary	7 (22.6)
Tertiary	22 (71)
Marital status, n (%)	
Single	20 (64.5)
Divorced	2 (6.5)
Widowed	0 (0)
Married	9 (29)
Relationship status, n (%)	
Single	13 (41.9)
In a relationship	18 (58.1)
Number of children, n (%)	
0	30 (96.8)
2	1 (3.2)
Application used, n (%)	
Nepanikař	11 (22.4)
VOS	7 (14.3)
Calm	4 (8.2)
Finch Care	4 (8.2)
Headspace	3 (6.1)
Waking up	2 (4.1)
Calmio	2 (4.1)
Apple Health	2 (4.1)
How We Feel	2 (4.1)
Other	12 (24.5)
Recruited through, n (%)	
Recruitment advert	15 (48.4)

XSL•FO
RenderX

Characteristic	Value
Snowball or personal contact	16 (51.6)
Difficulty in making monthly ends meet (last 12 mo), n (%)	
Each time with difficulty	2 (6.5)
Usually with difficulty	1 (3.2)
Occasionally with difficulty	7 (22.6)
Rarely with difficulty	10 (32.3)
Completely without difficulty	11 (35.5)
Interview format, n (%)	
Face-to-face	11 (35.5)
Online	20 (64.5)

The semistructured interview guide was developed through a multistage, iterative process. Following agreement on the research objectives and questions, all the authors of this study independently proposed thematic domains and interview questions relevant to the investigated phenomenon. Through discussion and consensus, these suggestions were consolidated into 11 thematic areas such as participants' app use, motivations, perceived benefits and drawbacks, comparisons with other forms of MH support, trust, and long-term experiences with MH apps. The preliminary guide was further informed by a screening of MH apps available in major app stores and an analysis of user reviews, which helped identify salient user experiences and recurring concerns. Finally, the guide was pilot tested with three individuals (1 woman aged 26 y and 2 men aged 25 and 27 y) who had experience using MH apps. Based on participant feedback, several questions were reworded, alternative prompts were added, and items perceived as overly broad or insufficiently relevant to participants' experiences were refined. The full interview guide, including translations, is available in Multimedia Appendix 1.

Ethical Considerations

The study received ethical approval from the Masaryk University Research Ethics Committee (0198/2025) prior to data collection. All participants provided written informed consent and consent for the processing of personal data. Informed consent was obtained from all participants in writing and reaffirmed verbally immediately before the interview. Participation was voluntary, and participants could withdraw from the study without penalty before their data were anonymized. Interviews were audio-recorded and transcribed for analysis. Audio recordings were stored securely and deleted following transcription. Transcripts were anonymized prior to analysis, and access to study data was restricted to authorized members of the research team. Participants received CZK 500 (equivalent to US $23.23) as compensation for completing the interview. One participant described past suicidal ideation but indicated during the interview that these thoughts were no longer current. At the end of the interview, the interviewer returned to this disclosure, checked whether the participant wanted information about relevant support services, and offered to provide contact information.

Data Analysis

We conducted reflexive thematic analysis (RTA), following the 6-phase approach outlined by Braun and Clarke [29]. The MRTSC informed the broader conceptual framing of the study and was used as a lens through which the findings were subsequently interpreted. RTA is a flexible method suited to exploring patterned meaning across participant accounts. Our epistemological stance was informed by critical realism [30], which assumes that the participants' accounts provide access to real experiences and underlying mechanisms, while recognizing that these experiences are always mediated through language, context, and interpretation. The analysis was inductive and bottom-up, with themes developed from the data rather than imposed a priori. Consistent with RTA, data saturation was not used as a stopping criterion; instead, recruitment concluded after the project's predefined recruitment target had been reached and the dataset was judged sufficiently rich and varied to address the RQs [31]. Participants provided detailed accounts of their experiences, and the sample encompassed diverse patterns of app use, motivations, perceived benefits, challenges, and trust-related concerns, enabling the development of well-supported themes addressing the study aims. The adequacy of the dataset was assessed throughout the iterative process of data collection and analysis. Toward the end of recruitment, additional interviews continued to provide illustrative examples and nuances of participants' experiences but did not substantially expand the range of issues relevant to the RQs. Given the breadth, depth, and diversity of the dataset, the research team concluded that it was sufficiently rich to address the RQs. Only authors VS, ZS, and AŠ were involved in the analytic process; stages 1 - 3 (ie, data familiarization, initial coding, and early pattern identification) were conducted independently by each author to encourage diverse readings of the data, with regular consultation to discuss developing interpretations, while later stages of theme development, refinement, and definition were collaborative. This iterative and reflexive process allowed us to move between participant accounts and broader explanatory interpretations, attending both to surface meanings and to the underlying conditions that shaped app use and the consequences of app use. Further methodological specifics are reported in the accompanying COREQ checklist (Checklist 1).

XSL•FO
RenderX

Positionality

AŠ and VS approached the analysis from backgrounds in critical health psychology, with particular attention to the interplay between individual self-care and broader health and care contexts. AŠ had limited personal experience with MH apps, while VS had previously used meditation apps before transitioning to unguided practice. ZS approached the research from a critical psychology perspective, with particular sensitivity to the social contexts in which individual responsibility for MH and well-being is constructed. These differing positions informed reflexive discussions of developing interpretations throughout the analytic process. A more detailed description of each author's positionality is available in Multimedia Appendix 2.

Results

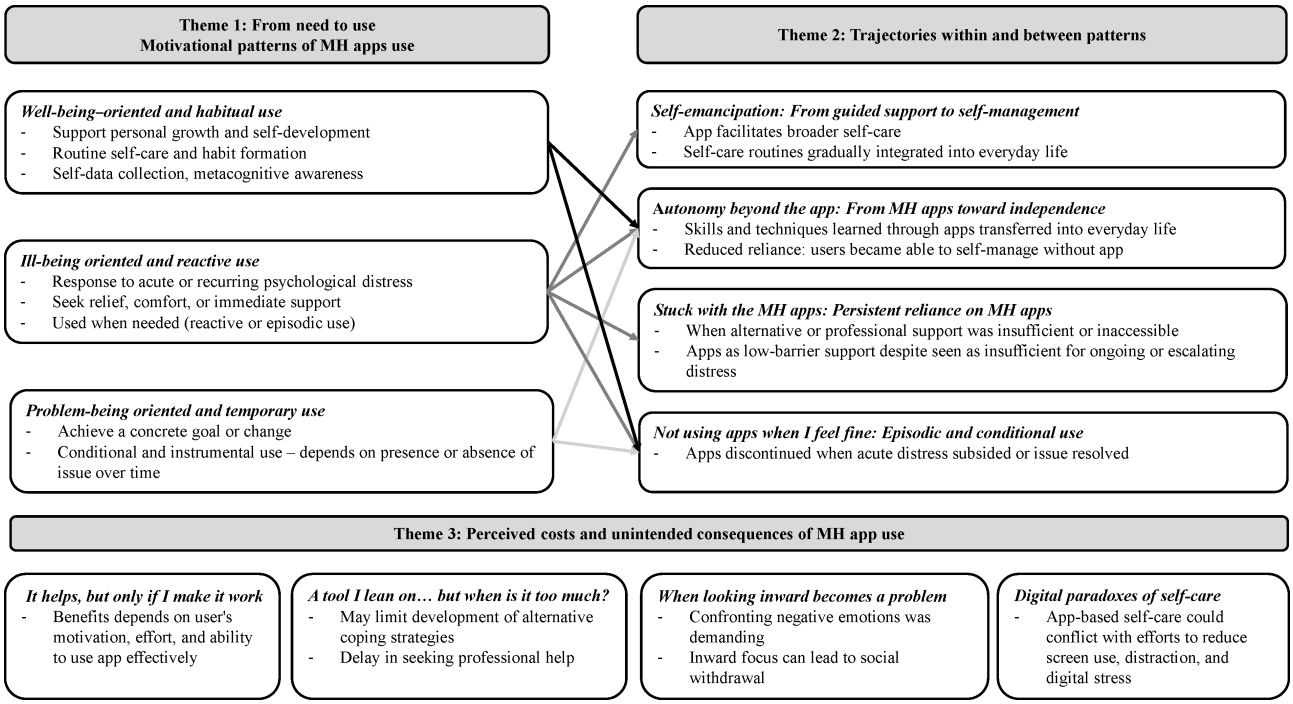
Sample Description

Regarding the duration of use, most of the sample had been using their apps of choice for over a year (n=21), followed by participants who had used them for 6 months to 1 year (n=6), and 2-6 months (n=4). Out of 31 participants, 10 disclosed MH issues that could be classified as symptoms related to a clinical disorder. Specifically, 5 participants reported panic attacks, 4 reported depressive mood, 3 reported anxiety, and 1 reported suicidal ideation. Clinical diagnoses disclosed during the interviews were attention-deficit/hyperactivity disorder (ADHD; n=2), bipolar disorder (n=1), and a history of alcoholism (n=1). For further sample characteristics, refer to Table 1.

Analytical Results

Our analysis generated 3 themes that capture the dynamic nature of MH app use. These themes illustrate how participants initially engaged with apps, how use was reshaped or interrupted over time, and how MH apps were integrated within broader self-care strategies. The themes are analytically distinct but closely intertwined, reflecting self-care decision-making processes (Figure 1).

Figure 1. Motivational patterns, trajectories, and perceived costs of mental health (MH) app use. The figure provides a visual representation of the thematic findings. Arrows illustrate relationships between patterns (theme 1) and trajectories of MH app use (theme 2) identified in participants' accounts. Theme 3 represents perceived costs and unintended consequences that cut across patterns and trajectories.



Theme 1: From Need to Use—Motivational Patterns of MH App Use

Overview

Throughout the interviews, participants described diverse motivations for downloading MH apps. We identified 3 overarching motivational patterns that reflected how participants understood their mental well-being and the role of digital self-care at the onset of MH app use—*well-being–oriented and habitual use*, *ill-being-oriented and reactive use*, and *problem-oriented and temporary use*.

Well-Being–Oriented and Habitual Use

Well-being–oriented use was typically described as proactive and oriented toward maintaining or enhancing psychological well-being over time. Participants who entered app use from this position often framed MH apps as tools for routine self-care, supporting habit formation and reflection, and structuring time devoted to oneself. Most commonly, it was associated with functions that enabled self-data collection (eg, emotion tracking and journaling) or supported the development of metacognitive awareness (through meditation) and general mindfulness.

For several participants, well-being–oriented use also involved tracking functions to facilitate long-term reflection on life events

XSL•FO
RenderX

and experiences. In these cases, participants wanted to document and preserve (mostly positive) personal life experiences over time:

> *...to analyze my states or situations I find myself in... so I'll be able to look back in history and analyze it, even learn from it, from mistakes or from what's repeating.* [P10_10_56M]

Importantly, well-being–oriented engagement did not always originate from a clearly articulated personal need. In some cases, app use began through contextual exposure, such as social media recommendations, employer-provided access, or professional curiosity among health and social care workers who sought to familiarize themselves with MH apps to assess whether to recommend them and which apps to recommend to their clients. What initially started as experimentation evolved into a meaningful self-care practice if participants experienced the app as beneficial.

The perceived value of habitual use often became particularly salient when use was interrupted. Some participants reported that during unintended breaks, most commonly due to lack of time, they felt less emotionally regulated or more scattered:

> *I miss it in the sense that I know I approach life differently and I'm less emotionally stable.* [P32_27M]

These experiences functioned as a feedback loop that reinforced motivation to resume regular self-care practices, with continued support from MH apps.

Ill-Being–Oriented and Reactive Use

By contrast, ill-being–oriented use was characterized by reactive engagement with apps in response to acute or recurrent psychological distress. Participants turned to apps because of anxiety, panic attacks, emotional breakdowns, mood deterioration, or periods of intense stress. In these situations, apps were valued primarily for providing rapid relief in the form of distraction (eg, breathing techniques and simple interactive tasks) or grounding when other forms of support were unavailable. One participant compared having an MH app on her phone to:

> *... [having] a pain killer in your medicine cabinet. One doesn't need it every day, but you know if your head hurts really bad, it's there.* [P29_23F]

A recurring context for reactive use involved situations in which participants felt trapped or unable to leave, with MH apps described as a "lifebuoy" during periods of severe panic because they interrupted escalating anxiety and validated emotional experiences:

> *...when the panic attack started to come on, I wasn't able to think rationally... And I really thought I was going to suffocate. And I knew there were some breathing exercises there, and different exercises to kind of redirect attention, or the app tries to stop the brain from thinking 'I'm dying, help, help...?'* [P29_23F]

Ill-being–oriented use was typically episodic, with participants reporting that they stopped using the app once the acute distress subsided. As one interviewee noted:

> *...the last time [I used the app] was about two weeks ago, so in terms of frequency now... maybe about once every two weeks. Honestly, I'm actually feeling quite well right now. Psychologically, I mean.* [P14_19F]

Although apps were often described as effective in reducing immediate distress, participants generally distinguished this form of use from longer-term psychological work. Apps were framed as first aid or temporary support, rather than substitutes for therapy or deeper interventions.

Problem-Oriented and Temporary Use

The third motivational pattern reflected a problem-oriented approach to MH app use. Participants did not turn to apps during acute episodes but when they sensed that "*something was not right*" and wanted to understand or manage a specific issue over time. Their use of MH apps was framed less as a tool for immediate emotional regulation and more as a resource for understanding the issue. For example, one participant described using a different app for mood tracking to evaluate the emotional impact of her work, explicitly framing the app as a way to *see* and confirm or deny what she had been sensing:

> *I was using [the app] because of work, because I wanted to find out whether that job was really affecting me emotionally so much, and how bad it actually was. ...I wanted to have it on paper and see to what extent it was the job and to what extent it was something else.* [P12_34F]

Engagement with the apps was typically conditional and instrumental. It fluctuated over time, based on the presence or absence of the issue. Participants described adjusting the frequency and intensity of app use in response to changing circumstances or perceived need. Rather than striving for daily use, they deliberately returned to the app during more demanding periods and disengaged when conditions improved. As one participant noted, regular use felt unnecessary outside of more difficult phases:

> *Now I only come back to [the app] when I feel like I need it. When there's a more difficult period, I feel that it's needed more to become more aware again.* [P18_31F]

These motivational patterns captured the distinct ways in which participants entered MH app use. They differed in urgency, temporal framing, and expectations for what the apps could provide. However, our data also showed some overlap between these patterns, and some accounts revealed movement between patterns over time. Rather than discrete categories, the patterns are best understood as orienting distinctions that shaped how participants integrated the apps into their lives and the roles the apps played. How these motivational patterns evolved was the focus of the trajectories described further in this study.

XSL•FO
RenderX

Theme 2: Trajectories Within and Between Patterns

Overview

Across interviews, participants narrated common variations in how they put the apps to use, reflecting shifting needs and life circumstances. We identified four trajectories to describe how participants moved within and between motivational patterns as follows: (1) the *self-emancipatory trajectory*, in which app use supported a shift from distress toward self-management; (2) the *trajectory from MH apps toward independence*, in which managing issues without the app was framed as a positive outcome; (3) the *persistent reliance trajectory*, characterized by continued or long-term use; and (4) the *episodic and conditional trajectory*, in which apps were used selectively, often only during periods of difficulty.

Self-Emancipation: From Guided Support to Self-Management

The first trajectory illustrates how MH app use could support a shift from distress management toward broader self-care routines. Here, autonomy was narrated not primarily as disengagement from the app, but as a gradual movement away from external support and toward more self-directed self-management. Specifically, some participants described transitions from ill-being- to well-being–oriented use, in which MH apps functioned as tools to facilitate the shift. Several participants described installing apps when they decided to reduce or terminate psychotherapy after reaching their initial therapeutic goals:

> *...I think I kind of felt that I didn't need [therapy] that much anymore, so maybe I perceived [the app] as some kind of replacement for therapy, that I still had some kind of certainty that, if something happened, it could maybe help me.* [P06_23F]

This excerpt illustrates how MH apps could function as a psychological safety net, facilitating the transition from psychotherapy to more autonomous self-management.

In this regard, MH apps were often perceived as tools that created a visually appealing and user-friendly space. They legitimized the time devoted to self-care practices and thereby supported the development of self-care management. One interviewee further described how self-care subsequently extended beyond therapy:

> *I'd say that actually [I do] a huge number of things that maybe aren't a priori, that one would normally say are good for mental health[...] So, when a person reflects on it once in a while, or even every day, it at least then leads me to the point where, at a minimum, I'm reminded every day that I should be doing something for myself.* [P18 _31F]

In this trajectory, MH apps supported the development of self-care routines that were subsequently integrated into participants' everyday lives. Participants described a broader reorientation toward caring for their MH on a day-to-day basis, with apps serving as scaffolding during periods of transition.

Autonomy Beyond the App: From MH Apps Toward Independence

Across both ill-being and well-being pathways, some interviewees described a gradual reduction in their reliance on MH apps as they learned and internalized skills and techniques introduced through app use. They subsequently applied them independently in everyday life:

> *[...] so it happens that I just tell myself, "Okay, I'm going to meditate for a few minutes even without the app."* [P16_52M]

For some interviewees, this transfer of skills into everyday life emerged as a positive side effect of MH app use, whereas others, such as the above-mentioned participant, described it as an explicit goal to become more mindful in daily life and to integrate techniques learned through the app beyond the app context.

Trajectories toward independence were characterized by an orientation toward transferring app-introduced skills into everyday life. For well-being–oriented users, this trajectory was motivated by a desire to integrate app-based practices more broadly into existing self-care routines. For those entering through an ill-being–oriented pathway, independence reflected an effort to avoid exclusive reliance on the app and to feel capable of managing situations without app-based guidance. For problem-oriented users, independence meant gaining sufficient understanding of a specific issue, allowing them to scale back or discontinue monitoring.

Stuck With MH Apps: Persistent Reliance on MH Apps

Not all trajectories led toward increased autonomy. For some users, the transition from MH app use as an emergency lifeline to the development of broader self-care practices did not materialize. These participants either maintained reliance on MH app techniques that had previously alleviated acute stress or continued to use MH apps as de facto substitutes for crisis helplines without further developing self-care practices. In both situations, they remained oriented toward MH apps not because the benefits were increasing, but because alternative forms of support were insufficient or perceived as inaccessible.

For 1 participant, escalating stressors eventually rendered app-based coping insufficient:

> *But even then, when I walk outside [after MH app use], the first thought just goes back to the original problem and I'm right where I was before, so basically it doesn't really work anymore. I don't want to bash the app. It's more that I'm currently experiencing much more stress, and in a more long-term way, and it's not just one isolated problem.* [P17_38M]

In this interview, seeking help from health care professionals was not considered, despite the perceived insufficiency of app-based support.

In cases where apps substituted for crisis intervention, especially free versions with limited functionalities, they functioned as low-barrier support for individuals experiencing psychological

XSL•FO
RenderX

distress and limited financial means while simultaneously attempting to access professional care:

> *[...] either calling the crisis hotline, which has happened to me several times when I couldn't get through... I definitely looked for [the app] myself, because I had long known that I needed psychotherapy. I couldn't afford that, so I tried to find free psychotherapy... I always see [MH apps] as a kind of substitute for situations when therapy or another person simply isn't available.* [P15_42F]

Although this interviewee dealt with her mental difficulties largely on her own, she perceived her difficulties as too severe to be addressed through MH apps alone. This perceived limitation was echoed by another participant, who explicitly stated:

> *When... or once it becomes long-term, an app on its own won't really help.* [P23_24F]

Taken together, these accounts suggest that persistent reliance on MH apps may reflect not only individual coping strategies, but also structural constraints that shape access to professional MH care.

Not Using MH Apps When I Feel Fine: Episodic and Conditional Use

Episodic and conditional use was most commonly associated with the ill-being pattern and was characterized by nonsystematic engagement with MH apps. Participants frequently reported that they stopped using the app once acute distress subsided. Unlike trajectories toward independence, episodic and conditional use did not involve internalizing app-based practices. Instead, participants described keeping MH apps available as resources to be reactivated during periods of psychological difficulty, while preferring other self-maintenance strategies during less demanding times.

Some interviewees indicated that their MH app use followed a seasonal pattern, triggered by episodic deteriorations in well-being, typically during autumn or winter. These interviewees expressed interest in seasonal subscriptions with flexible and user-friendly terms, reflecting how their engagement with MH apps was shaped by seasonal need:

> *I wouldn't want to pay for an expensive annual subscription, especially if I know I'm not using it on a daily basis.* [P22_23F]

This participant did not perceive daily MH app use as necessary and consequently was not interested in committing to a long-term subscription.

Theme 3: Perceived Costs and Unintended Consequences of MH App Use

Overview

While most participants described MH apps as helpful (eg, providing structure, calming guidance, and distraction during stress), they also reflected on tensions and unintended consequences arising from their use. We identified 4 interrelated subthemes that captured these perceived costs. First, "*It helps, but only if I make it work*" describes tensions that stem from the perceived responsibility for making app use effective. Second, "*A tool I lean on... but when is it too much?*" captures tensions around reliance on apps, including their use as substitutes for other forms of support and the potential delay of help-seeking. Third, "*Inward focus as a risk*" highlights how self-monitoring and reflective practices, while often stabilizing, can also intensify attention to internal states and become ruminative. Finally, "*Digital paradoxes of self-care*" describes conflicts between app-based self-care and attempts to reduce screen time or disengage from stressful technologies. Across these subthemes, tensions emerged not solely because of functional limitations, but through participants' active efforts to engage with digital self-care and their ongoing evaluations of the costs and benefits of app use.

It Helps, but Only if I Make It Work

Across interviews, participants repeatedly emphasized that, while MH apps offered accessible support, they could not generate the motivation required for consistent or effective use. Apps were commonly framed as passive tools whose benefits depended on users' own effort and willingness to engage:

> *[...] the apps will only remind you that this is important, but they won't do the work for you.* [P04_24F]

Similarly, another participant stressed that change stemmed from her own determination rather than the app itself, which merely provided a starting point:

> *It's not that the app directly forced me to take better care of my mental health. What actually pushed me to do something about it was mainly that [my mental states] were just burdensome.* [P21_24F]

Across accounts, participants recognized a certain level of agency as a prerequisite for benefit. For some, this responsibility was framed as a natural or even empowering aspect of app use; for others, it was experienced as a burden or source of tension.

This ambivalence was particularly visible in relation to reminders and streak-based features. While some participants appreciated the structure these features provided, others described actively protecting their autonomy by resisting notifications or streak mechanisms that induced guilt or created tension between disengaging from the app and losing accumulated rewards. As 1 participant explained, reminders were acceptable only as long as they did not make her feel bad for noncompliance:

> *If it made me feel bad for not doing something, I'd probably be done with this shit.* [P04_25F]

In such cases, perceived pressure was not motivational but instead led to disengagement.

Beyond motivation and effort, some participants emphasized that benefiting from MH apps also presupposed a certain level of competence (eg, understanding and appropriately applying the techniques offered). One participant noted that merely having access to coping strategies was insufficient; through therapy, she learned how to work with techniques she had previously encountered in the app:

XSL•FO
RenderX

> *[...] to get the right effect, you need to have the right techniques and know how to use them... thanks to therapy, I learned how to do that.* [P07_23F]

Again, this reinforced the idea that healthcare providers could play a crucial role in guiding some users to successfully leverage the apps' functions. Taken together, these findings indicate that the apps' effectiveness is not intrinsic, but negotiated through the users' engagement with them.

A Tool I Lean on… but When Is It Too Much?

Participants reported several unintended consequences related to growing reliance on MH apps, often framed as a tension between the benefits of immediate support and concerns about depending too heavily on the app. Apps were often described as psychological crutches that were helpful in moments of distress, yet accompanied by reflections on whether such reliance might limit the development or use of alternative coping strategies. One participant recalled how her partner explicitly raised this concern while observing her use of the app during anxiety-provoking situations on public transport, prompting her to consider what would happen if the app were unavailable:

> *...he came up with this interesting idea... what would I do if I didn't have my cell phone... that I actually rely on it a little bit... and so I become a little dependent on it.* [P21_23F]

Reliance on apps was also linked to reflections on help-seeking beyond the app. Participants frequently emphasized that they themselves were responsible for recognizing when app-based support was no longer sufficient and when professional help was needed. One interviewee articulated particularly clearly how short-term relief provided by apps could intersect with postponing professional help. She first framed this as a general risk:

> *...people often, instead of paying for care, reach for an app and think that it will replace therapeutic care.* [P07_23F]

The interviewer then asked:

> ***Interviewer:*** *Is that also your experience?*
>
> ***P07_23F:*** *Absolutely, yes. I was also that person who reached for the app first [...] but I didn't have any guarantee that it's actually helping me.*

In these accounts, the perceived risk concerned the way reliance on short-term app support could reshape judgments about when additional help was necessary.

When Looking Inward Becomes a Problem

Some participants worried that app use promoted an inward focus, which they found emotionally demanding rather than uniformly supportive, particularly when it involved confronting negative emotions or unresolved difficulties. As one participant described, reflective prompts were not limited to positive experiences but also required deeper engagement:

> *[...] that not only positive questions, but also deeper ones, where you really have to think about it and it hurts. It can hurt.* [P02_20M]

In a similar vein, others questioned whether repeated attention to negative emotional states might reinforce rather than alleviate distress:

> *If I kept feeling like shit and kept saying that I felt sick, I think I would have just continued to feel sick, because I would have kept repeating to myself that I feel sick.* [P04_24F]

In these accounts, self-monitoring was experienced as effortful and potentially counterproductive when not accompanied by sufficient support or resolution.

Others expressed concern that sustained inward engagement could lead to withdrawal from social interaction, noting that excessive self-focus might not be desirable:

> *...if I kept using it, I'd really withdraw into myself, which probably wouldn't be good.* [P12_34F]

Taken together, these accounts point to the double-edged nature of MH apps; while privacy and inward focus are central to digital self-care, they may also become emotionally demanding or socially isolating under certain conditions.

Digital Paradoxes of Self-Care

Finally, some participants reflected on broader digital paradoxes inherent in app-based self-care, highlighting tensions between the goals of mental well-being and the means through which support was delivered. Several respondents described a contradiction between using MH apps and their simultaneous efforts to reduce overall technology or screen use, particularly when digital environments themselves were experienced as stressful or overstimulating. As 1 participant noted:

> *I am trying to distance myself a bit from technology right now because I'm finding it's one of the triggers that stresses me out. Actually, meditating and using an app are perhaps a bit... maybe they go against each other a little. It just shouldn't feel forced, you know? Which technology often does.* [P30_43M]

Others similarly described discomfort with the necessity of keeping the phone close to access support, particularly when opening an app risked drawing them into other, unwanted digital activities:

> *...it bothers me that it forced me... that I have to keep it on me... once it's open, I start dealing with... other things there.* [P25_42F]

In response, some participants reported avoiding app use in certain contexts, such as at work or in overstimulating environments, acknowledging that the device itself could introduce distraction or exacerbate stress.

Discussion

Principal Findings

In our study, we explored why users adopt MH apps and how they integrate these tools into their lives and into broader ecosystems of MH care and support. Our findings highlight that user engagement is a continual negotiation between app functionalities and users' evolving needs. Importantly, we provide further evidence that disengagement can be a

XSL•FO
RenderX

meaningful outcome of MH app use, challenging engagement-centric assumptions that dominate current app design and evaluation.

Our results suggest that sustained MH app use is perceived as a goal only in a limited set of contexts. In other situations, MH apps are understood either as tools for temporary solutions or as instruments intended to foster the development of autonomous self-care practices that eventually become independent of MH apps. Our findings show that for some users, overcoming a difficult period or successfully integrating self-care practices into daily self-care maintenance resulted in what has been previously termed "positive disengagement" [27]. As users' goals were reached, they naturally decreased their app usage, showing that disengagement may follow individuals' own appraisals that the app is no longer needed. These reported experiences challenge the existing body of research on MH apps and the problem of low user engagement [9,14,15] by demonstrating that for some users, long-term sustained engagement is in direct conflict with their goals.

Beyond goal attainment, disengagement also emerged as a protective strategy in what we describe as the *digital self-care paradox*, whereby users intentionally reduce their engagement with digital tools to minimize the negative effects associated with prolonged technology use, such as overreliance on MH apps [19,23] or technology fatigue [15]. Moreover, usage discontinuation was often the result of successful substitution for off-screen self-care behaviors (eg, spending time outside). Therefore, "episodic" use or abandonment can often be a sign of a procedural recalibration of resources toward nondigital self-care, indicating successful self-care management.

This study showed that users varied considerably in their initial motivational patterns, which coincided with different self-care phases described in the MRTSC [26]. These patterns evolved along the following two dimensions: (1) intent, which referred to whether MH app use was oriented towards supporting well-being (emphasizing self-care maintenance), addressing ill-being (driven by self-care management), or understanding a specific problem (through self-care monitoring); and (2) engagement timing, which reflected when and how the apps were used, ranging from habitual and episodic to reactive use. Additionally, our findings showed that these initial patterns evolved over time into distinct trajectories shaped by the users' needs and resources, including enhanced self-care management supported by MH apps (self-emancipatory trajectories), the development of autonomous self-care without continued reliance on MH apps, increased reliance on MH apps as substitutes for human support, and episodic use limited to periods of perceived need.

Our findings provide further evidence of how MH apps enhance self-care maintenance, particularly through well-being–oriented and habitual use aimed at promoting psychological stability. Specifically, MH apps contribute to self-maintenance by legitimizing space for self-care, maintaining self-care routines, or supporting activities that reduce overall stress. The MRTSC suggests that the value of these digital routines lies in their cognitive economics. Ideally, apps offload the cognitive burden of daily maintenance, preserving users' psychological resources. However, our data reveal a critical tension—the digital mechanisms designed to support these habits (eg, notifications or streaks) may backfire. Instead of seamlessly integrating self-care into daily life, these intrusive features paradoxically increased technology fatigue, disrupting the very processes they aimed to facilitate.

Self-monitoring constituted a central component of the problem-oriented pattern of MH app use, although it was also present within the well-being–oriented and habitual patterns. While the MRTSC conceptualizes self-care monitoring as the observation and interpretation of changes, our findings illustrate that in the context of digital MH, this process extends beyond discrete symptom tracking. Some users engaged in detailed journaling and broader self-narrativizing. Here, the app facilitated a highly complex form of appraisal, allowing users to contextualize their psychological states or alternatively record positive experiences to counterbalance negative affect and keep a "life history." Crucially, however, the effectiveness and perception of this meaning-making depended on intentionality. When tied to a specific, personally meaningful goal, monitoring was beneficial. Conversely, when self-monitoring lacked clear intent, it was frequently experienced as burdensome or even depressing, which is consistent with prior research on the adverse outcomes of MH app use [10]. Ultimately, our data expand on this finding, suggesting that the value of digital self-monitoring hinges less on the mechanical act of tracking and more on the user's finding meaning in the process of recording data (ie, on their capacity to articulate why and how monitoring is meaningful for them).

Our study engages with the implications of stand-alone digital technologies designed for private and autonomous use by showing that MH apps were indeed largely managed by users themselves and rarely embedded in shared or professionally supported care practices. One strand of our findings illustrates how MH apps facilitated the processes of self-emancipation, supporting transitions from professionally delivered treatment to self-care. At the same time, our results show that external feedback on MH app use, whether from close others or health care professionals, often enhanced engagement, enabling users to refine their usage patterns or to apply skills acquired through the apps beyond the screen. Such external support proved especially vital for users who reported that in times of acute psychological distress, it was difficult to apply the techniques presented within the app. This finding resonates with the work of Villegas Mejía et al [25], who emphasized the importance of supporting engagement that extends beyond the MH app context.

Consistent with prior research [22], our interviews revealed numerous examples of MH app use enhancing self-care management by fostering skills and self-care habits that enabled some participants to transition from professionally delivered treatment to more autonomous management of their MH. However, this self-enhancement may have unintended consequences. For some users, greater perceived autonomy in self-management was accompanied by a heightened sense of personal responsibility for managing MH difficulties, leading them to hesitate to seek professional help, out of concern that they might burden health care providers or divert resources from

XSL•FO
RenderX

those in greater need. Viewed through the MRTSC, this hesitation is notable, as the theory positions consultation and seeking information as adaptive self-care management responses rather than departures from self-care. In treating help-seeking as a failure of autonomy, these users adopted a framing at odds with the MRTSC, in which turning to others is itself a legitimate self-care behavior. Our findings also resonate with the research of Van der Poll et al [24], who similarly described how MH app users may willingly assume personal responsibility for managing their own MH. Extending their work, our data suggest that this sense of responsibility can be accompanied by increased hesitation to seek professional help.

Moreover, our findings suggest that some users develop reliance on MH app–related practices as a primary source of immediate relief from negative emotional states. Although rapid relief may prevent symptoms from escalating, it may reduce the likelihood that users reflect on the seriousness of their condition and consider seeking professional help, potentially sustaining reactive crisis management. Unlike crisis hotlines, most stand-alone MH apps do not include mechanisms that actively encourage or facilitate seeking professional support.

We consider these findings especially concerning in populations where barriers to MH care are already high. In such contexts, promoting MH apps as a solution risks obscuring the problem in the short term while exacerbating it in the long term. Individuals who might otherwise seek professional help may instead rely on an app as a form of temporary relief, making symptoms more manageable while foregoing appropriate treatment. While this relief may reduce immediate distress, it can also prolong avoidable suffering and leave underlying needs for care unmet. What may appear to be a solution to limited access to MH services may, in fact, function only as a palliative response to a broader structural problem.

Limitations

This study has several limitations. First, the distinction between "well-being" and "ill-being" orientations may partly reflect processes of self-narrativization, whereby participants construct coherent accounts of their MH app use that align with identity work rather than their immediate psychological states. Second, the study relies on retrospective accounts of MH app use. Participants reconstructed the evolution of their engagement over time based on memory and subjective interpretation, which may introduce recall bias. Future research would benefit from longitudinal designs capable of capturing trajectories of MH app use as they unfold. Third, although our analysis focused on MH apps, some participants reported extending their digital self-care practices by consulting generative AI tools (eg, large language models) embedded in, or external to MH apps.

Finally, the sample was relatively homogeneous, consisting predominantly of young, female, Czech, digitally engaged participants, which limits the range of experiences represented and, consequently, the transferability of the findings to other populations and contexts. In addition, individuals currently experiencing a psychological crisis were excluded from participation. Our findings may therefore not fully capture experiences of MH app use during acute psychological distress. Future research should examine whether and how the identified patterns and trajectories manifest across more diverse demographic, cultural, and clinical contexts.

Recommendations

Similar to health care providers, MH app designers may consider "positive disengagement" as a legitimate outcome. Because some app designs discourage this through various mechanisms (eg, daily streaks), incorporating "pause" functions, allowing flexible payment plans, and avoiding punitive gamification mechanics may empower users to successfully integrate coping strategies off-app. MH apps could also be designed to promote connection rather than substitute for human support. Instead of functioning as self-contained solutions, apps might serve as intermediary or boundary objects that facilitate communication with professionals, counselors, or trusted peers. Moreover, apps with emotion tracking or crisis support features may integrate empathic prompts to direct users to low-barrier resources (eg, hotlines and MH professional directories) in response to the detection of sustained distress. Clear guidance to distinguishing between manageable distress and symptoms that warrant clinical attention, alongside normalizing help-seeking as a legitimate and responsible step, may reduce feelings of shame and prevent responsibilization for solving one's own problems. In this way, MH apps can function as bridges to care rather than barriers to professional support.

For clinicians, the practical task is to bring clients' motivations for autonomous app use into the conversation and help educate them about the limits of these tools. Clients prone to reactive use during acute distress warrant particular attention, as these apps rarely offer crisis plans. MH professionals should work with such clients on a crisis plan that does not depend on the app. Mood tracking and journaling should be encouraged only for clients who experience them as beneficial or connected to personally meaningful goals.

Conclusion

This study shows how MH apps are embedded within broader practices of care and support, and under which conditions they may be empowering or problematic. For some users, they foster autonomy, strengthen self-care skills, and enable the transfer of coping strategies beyond digital platforms and even beyond professional treatment. In such cases, discontinuing app use may reflect the successful integration of self-care into everyday life. For others, however, MH apps may reinforce app-based coping and contribute to reliance on digital tools. Overall, providing clearer distinctions between everyday emotional strain and clinically significant symptoms, while framing help-seeking as a responsible and legitimate course of action, may help alleviate shame and counteract the tendency to place excessive responsibility on individuals to manage their own MH.

XSL•FO
RenderX

Acknowledgments

We express our gratitude to the participants of our study for giving us their time and helping us understand digital mental health tools better.

The authors declare the use of generative AI (GenAI) in the writing process. The following tasks were delegated to GenAI tools under full human supervision: text reformulation (refinement of Introduction and Discussion sections of the manuscript), table formatting, proofreading and editing. The GenAI tools used were ChatGPT 5.5, Gemini Pro, and Copilot. All AI-generated suggestions were evaluated and revised by the authors, who take full responsibility for the final content of the manuscript.

Funding

This study comes from the project "On our own: Opportunities and Risks in the Individualization of Society (PRINS) CZ.02.01.01/00/23_025/0008710," which is co-financed by the European Union.

Data Availability

The data that support the findings of this study are available upon request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Authors' Contributions

Conceptualization: VS, AŠ
Data curation: BB, MJ, EJ
Formal analysis: VS, ZS, AŠ
Funding acquisition: AŠ
Investigation: VS, ZS, BB, MJ, EJ, AŠ
Methodology: ZS
Resources: ZS, BB, MJ, EJ
Supervision: AŠ
Writing – original draft: VS, ZS, AŠ
Writing – review editing: VS, ZS, AŠ

Conflicts of Interest

None declared.

Multimedia Appendix 1
English translation of the interview questions.
[DOCX File, 32 KB - mhealth_v14i1e95414_app1.docx]

Multimedia Appendix 2
Positionality.
[PDF File, 65 KB - mhealth_v14i1e95414_app2.pdf]

Checklist 1
COREQ checklist.
[PDF File, 154 KB - mhealth_v14i1e95414_app3.pdf]

References

1. McManus S, Gunnell D, Cooper C, et al. Prevalence of non-suicidal self-harm and service contact in England, 2000-14: repeated cross-sectional surveys of the general population. Lancet Psychiatry 2019 Jul;6(7):573-581. [doi: 10.1016/S2215-0366(19)30188-9] [Medline: 31175059]
2. Richter D, Wall A, Bruen A, Whittington R. Is the global prevalence rate of adult mental illness increasing? Systematic review and meta-analysis. Acta Psychiatr Scand 2019 Nov;140(5):393-407. [doi: 10.1111/acps.13083] [Medline: 31393996]
3. Santomauro DF, Mantilla Herrera AM, Shadid J, et al. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. The Lancet 2021 Nov;398(10312):1700-1712. [doi: 10.1016/S0140-6736(21)02143-7]
4. Cabanas E, Illouz E. Manufacturing Happy Citizens: How the Science and Industry of Happiness Control Our Lives: Polity Press; 2019.
5. Yadavendu VK. Shifting Paradigms in Public Health: From Holism to Individualism: Springer India; 2013.

XSL•FO
RenderX

6. Potočár L, Mladá K, Kučera M, Mohr P, Winkler P, Formánek T. Population mental health, help-seeking and associated barriers following the COVID-19 pandemic: analysis of repeated nationally representative cross-sectional surveys in Czechia. Psychiatry Res 2024 Jan;331:115641. [doi: 10.1016/j.psychres.2023.115641] [Medline: 38042095]
7. Winkler P, Kunc B, Guerrero Z, Mohr P, Schomerus G, Mladá K. Changes in stigma and population mental health literacy before and after the Covid-19 pandemic: analyses of repeated cross-sectional studies. SSM - Mental Health 2024 Dec;6:100369. [doi: 10.1016/j.ssmmh.2024.100369]
8. Eisenstadt M, Liverpool S, Infanti E, Ciuvat RM, Carlsson C. Mobile apps that promote emotion regulation, positive mental health, and well-being in the general population: systematic review and meta-analysis. JMIR Ment Health 2021 Nov 8;8(11):e31170. [doi: 10.2196/31170] [Medline: 34747713]
9. Kulke JK, Fuhrmann LM, Berking M, et al. Efficacy of standalone smartphone apps for mental health: an updated systematic review and meta-analysis. Lancet Digit Health 2025 Nov;7(11):100923. [doi: 10.1016/j.landig.2025.100923] [Medline: 41290454]
10. Linardon J, Fuller-Tyszkiewicz M, Firth J, et al. Systematic review and meta-analysis of adverse events in clinical trials of mental health apps. NPJ Digit Med 2024 Dec 18;7(1):363. [doi: 10.1038/s41746-024-01388-y] [Medline: 39695173]
11. Luo Y, Stice BL, Lenz AS. Mental health apps for depression: a meta‐analysis. Jour of Counseling & Develop 2025 Jan;103(1):25-38. [doi: 10.1002/jcad.12535]
12. Morello K, Schäfer SK, Kunzler AM, Priesterroth LS, Tüscher O, Kubiak T. Cognitive reappraisal in mHealth interventions to foster mental health in adults: a systematic review and meta-analysis. Front Digit Health 2023;5:1253390. [doi: 10.3389/fdgth.2023.1253390] [Medline: 37927578]
13. Valentine L, Hinton JDX, Bajaj K, et al. A meta-analysis of persuasive design, engagement, and efficacy in 92 RCTs of mental health apps. NPJ Digit Med 2025 Apr 29;8(1):229. [doi: 10.1038/s41746-025-01567-5] [Medline: 40301581]
14. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 6;20(6):e199. [doi: 10.2196/jmir.9275] [Medline: 29875089]
15. Jin S, Kim B, Han K. "I don't know why i should use this app": holistic analysis on user engagement challenges in mobile mental health. : ACM; 2025 Apr 26 Presented at: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems Yokohama Japan; Apr 26 to May 1, 2025; Yokohama Japan p. 1-23. [doi: 10.1145/3706598.3713732]
16. Wanniarachchi VU, Greenhalgh C, Choi A, Warren JR. Personalization variables in digital mental health interventions for depression and anxiety in adolescents and youth: a scoping review. Front Digit Health 2025;7:1500220. [doi: 10.3389/fdgth.2025.1500220] [Medline: 40444184]
17. Chawla S, Saha S. Exploring perceptions of psychology students in Delhi-NCR Region towards using mental health apps to promote resilience: a qualitative study. BMC Public Health 2024 Jul 26;24(1):2000. [doi: 10.1186/s12889-024-19565-9] [Medline: 39061041]
18. Saoane Thach K. A qualitative analysis of user reviews on mental health apps: who used it? for what? and why? : IEEE; 2019 Presented at: 2019 IEEE-RIVF International Conference on Computing and Communication Technologies (RIVF); Mar 20-22, 2019; Danang, Vietnam p. 1-4. [doi: 10.1109/RIVF.2019.8713726]
19. Tong F, Lederman R, D'Alfonso S, Berry K, Bucci S. Digital therapeutic alliance with fully automated mental health smartphone apps: a narrative review. Front Psychiatry 2022;13:819623. [doi: 10.3389/fpsyt.2022.819623] [Medline: 35815030]
20. Alqahtani F, Orji R. Insights from user reviews to improve mental health apps. Health Informatics J 2020 Sep;26(3):2042-2066. [doi: 10.1177/1460458219896492] [Medline: 31920160]
21. Nagar R, Quirk HD, Anderson PL. User experiences of college students using mental health applications to improve self-care: implications for improving engagement. Internet Interv 2023 Dec;34:100676. [doi: 10.1016/j.invent.2023.100676] [Medline: 37867616]
22. Simmons N, Goodings L, Tucker I. Experiences of using mental health apps to support psychological health and wellbeing. J Appl Soc Sci (Boulder) 2024 Mar;18(1):32-44. [doi: 10.1177/19367244231196768]
23. Mudau T, Jithoo V, Dietrich J. Exploring the demand, practicality, and acceptability of a mental health application intervention for students at a South African university. South African Journal of Psychology 2024 Jun;54(2):171-184. [doi: 10.1177/00812463241249100]
24. Van Der Poll R, Coetzee B, Bantjes J. Willing and unwilling digital cyborg assemblages: university students talk about mental health apps. Digit Health 2023;9:20552076231210658. [doi: 10.1177/20552076231210658] [Medline: 37915793]
25. Villegas Mejía C, Remmerswaal D, Engels R, Ludden GDS, Boffo M. Macro-engagement in mHealth: exploring user engagement beyond the screen. Digit Health 2024;10:20552076231225591. [doi: 10.1177/20552076231225591] [Medline: 38269371]
26. Riegel B, Jaarsma T, Lee CS, Strömberg A. Integrating symptoms into the middle-range theory of self-care of chronic illness. ANS Adv Nurs Sci 2019;42(3):206-215. [doi: 10.1097/ANS.0000000000000237] [Medline: 30475237]
27. O'Brien HL, Roll I, Kampen A, Davoudi N. Rethinking (dis)engagement in human-computer interaction. Comput Human Behav 2022 Mar;128:107109. [doi: 10.1016/j.chb.2021.107109]

XSL•FO
RenderX

28. Cole-Lewis H, Ezeanochie N, Turgiss J. Understanding health behavior technology engagement: pathway to measuring digital behavior change interventions. JMIR Form Res 2019 Oct 10;3(4):e14052. [doi: 10.2196/14052] [Medline: 31603427]
29. Braun V, Clarke V. Thematic Analysis: A Practical Guide: Sage; 2022.
30. Maxwell JA. A Realist Approach for Qualitative Research: Sage; 2012.
31. Braun V, Clarke V. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. Qualitative Research in Sport, Exercise and Health 2021 Mar 4;13(2):201-216. [doi: 10.1080/2159676X.2019.1704846]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
COREQ: Consolidated Criteria for Reporting Qualitative Research
MH: mental health
MRTSC: Middle-Range Theory of Self-Care of Chronic Illness
RQ: research question
RTA: reflexive thematic analysis

Edited by A Stone; submitted 16.Mar.2026; peer-reviewed by A Teles, K Schroeder; revised version received 29.Aug.2026; accepted 31.Aug.2026; published 09.Oct.2026.

Please cite as:
Suchý V, Scott Z, Boštíková B, Janoušek M, Josková E, Ševčíková A
How People Integrate Mental Health Apps Into Everyday Self-Care: Qualitative Study
JMIR Mhealth Uhealth 2026;14:e95414
URL: https://mhealth.jmir.org/2026/1/e95414
doi:10.2196/95414

© Vít Suchý, Zuzana Scott, Barbora Boštíková, Marek Janoušek, Eliška Josková, Anna Ševčíková. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 9.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Daily Reminders Increase Adherence to an Electronic Headache Diary (DMKG-App): Retrospective Longitudinal Cohort Study

Ruth Ruscheweyh[1], Dr med, PD; Lars Neeb[2], Dr med, PD; Bianca Raffaelli[3,4], Dr med, PD; Torsten Kraya[5,6], Dr med, PD; Tobias Freilinger[7], Dr med, PD; Thomas Dresler[8,9,10], Dipl psych, Dr phil; Gudrun Gossrau[11], Dr med, PD; Tim Patrick Jürgens[12,13], Dr med, PD; Victoria Ruschil[14,15], Dr med; Jörg Scheidt[16], Dr rer nat; Yannic Siebenhaar[16]

[1]Department of Neurology, LMU University Hospital, LMU Munich, Marchioninistr. 15, Munich, Germany

[10]LEAD Graduate School & Research Network, University of Tuebingen, Tuebingen, Germany

[11]Interdisciplinary Pain Center, University Hospital and Faculty of Medicine Carl Gustav Carus, TU Dresden, Dresden, Germany

[12]Department of Neurology, Headache Center North-East, University Medical Center Rostock, Rostock, Germany

[13]Department of Neurology, KMG Klinikum Güstrow, Güstrow, Germany

[14]Center for rare diseases, University Hospital Tuebingen, Tuebingen, Germany

[15]Department of Neurology and Epileptology, Hertie Institute for Clinical Brain Research, University Hospital Tuebingen, Tuebingen, Germany

[16]Institute for Information Systems, University of Applied Sciences Hof, Hof, Germany

[2]Department of Neurology, Brandenburg Medical School, Theodor Fontane, University Hospital Brandenburg/Havel, Brandenburg/Havel, Germany

[3]Department of Neurology, Charité Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

[4]Clinician Scientist Program, Berlin Institute of Health, Berlin, Germany

[5]Department of Neurology, Hospital Sankt Georg Leipzig gGmbH, Leipzig, Germany

[6]Headache Center Halle, Department of Neurology, University Hospital Halle, Halle, Germany

[7]Department of Neurology, Klinikum Passau, Passau, Germany

[8]Department of Psychiatry and Psychotherapy, Tuebingen Center for Mental Health, University Hospital Tuebingen, Tuebingen, Germany

[9]German Center for Mental Health (DZPG), Partner Site Tuebingen, Tuebingen, Germany

Corresponding Author:
Ruth Ruscheweyh, Dr med, PD
Department of Neurology, LMU University Hospital, LMU Munich, Marchioninistr. 15, Munich, Germany

Abstract

Background: The use of electronic headache diaries via smartphone apps has several advantages over paper and pencil diaries. Still, adherence is not always satisfactory. One potential strategy to improve adherence is the implementation of daily reminders, but their effect remains uncertain.

Objective: The DMKG-App (the headache diary app of the German Migraine and Headache Society, DMKG) sends daily reminders to encourage user adherence. Due to a technical problem, reminders failed for a 14-day period in early 2023. This provided a unique opportunity to evaluate the effect of daily reminders on adherence to the headache diary in the form of a retrospective longitudinal observational cohort study.

Methods: The failure period was compared to 14-day prefailure and postfailure periods and to corresponding periods of the previous year. We evaluated effects on the number of daily entries, delay to make the daily entry, and attrition rate, using ANOVA and logistic regression (α<.05). The effect of age, gender, headache frequency, and previous adherence was investigated using interaction analysis.

Results: The DMKG-App had 4236 active users before reminders failed (females 81.6%, 95% CI 80.5 - 82.8), age 39.8 (95% CI 39.4 - 40.2) years. During the failure period, the number of daily entries fell by 11.0% (95% CI 9.2 - 12.8). More specifically, entries per 14-day period significantly declined from 10.2 (95% CI 10.1 - 10.3) prefailure to 9.1 (95% CI 9.0 - 9.2) during the failure period, followed by a rise to 10.2 (95% CI 10.1 - 10.3) during the postfailure period ($F_{2,11498}$=79.2; P<.001). The delay to make the daily entry increased by 46.3% (95% CI 45.8 - 46.8), from 1.90 (95% CI 1.89 - 1.90) days prefailure to 2.78 (95% CI 2.77 - 2.79) days during failure, and returned to 1.85 (95% CI 1.84 - 1.86) days postfailure ($F_{2,39}$=155.4; P<.001). In addition, there was a significant rise in attrition rate by 32.2% (95% CI 8.1 - 61.7) (from 4.0%, 95% CI 3.4 - 4.6, in the prefailure period to 5.2%, 95% CI 4.6 - 5.9, in the failure period, returning to 3.9%, 95% CI 3.4 - 4.5, in the postfailure period, z=3.10; P=.002). Interaction analysis revealed that the effect of reminder failure was larger in males than in females (regarding entry delay), in persons with lower compared to higher headache frequencies (regarding number of daily entries and entry delay), and in persons

XSL•FO
RenderX

with higher vs lower previous adherence rates (regarding number of daily entries, entry delay, and attrition). There was no interaction between age and reminder failure.

Conclusions: This is the first study investigating the effect of daily reminders in a digital headache diary app. Results suggest that reminders are an effective means to improve adherence. Males, persons with low headache frequencies, and those with high adherence during the presence of reminders seem to benefit most. These results are important additions to our knowledge on compliance with digital health apps.

Trial Registration: German Clinical Trials Register DRKS 00021081; https://drks.de/search/en/trial/DRKS00021081/details

(JMIR Mhealth Uhealth 2026;14:e73259) doi:10.2196/73259

KEYWORDS

headache; electronic headache diary; adherence; entry delay; attrition rate

Introduction

Headache diaries are important tools for diagnosis and treatment evaluation in primary headaches [1]. Indeed, it has been shown that self-reported retrospective estimations of headache frequency are highly unreliable. In a study from the Leiden Headache Center, for example, the average difference between retrospectively self-reported and diary-assessed monthly migraine days was 4.7 (SD 5.0) days per month, with both overestimation and underestimation occurring [2]. Paper-and-pencil diaries have been used for a long time and have the advantage of being simple and free from technical requirements. On the other hand, they can get lost, may not be at hand when needed, and patients may sometimes fill in missing days shortly before the appointment, resulting in recall bias [3,4].

While some patients prefer paper diaries, smartphone-based electronic headache diaries have gained popularity in recent years. At a time when most people carry a smartphone, and with the majority of headache patients being relatively young and comfortable with digital technology, smartphone apps have several advantages [4]. The smartphone is always at hand, recall bias can be reduced by allowing retrospective entries only for a defined period, and automatic summaries give a quick overview to be shared with the physician [5,6]. Patients can be guided through data capture, and automated data validation improves data quality [6]. It has been shown that electronic diaries are superior to paper diaries in several respects (better validity, less missing data, and larger patient satisfaction [4,7]). Data can also more easily be used for large-scale research, as recently shown in an analysis of acute medication rating based on more than 270,000 users [8]. In the future, data from headache diary apps might be integrated with other health data, with possible applications, for example, for attack prediction [9].

Still, adherence to the diary is key to headache evaluation, and prompt entries increase accuracy by decreasing recall bias [2]. One attempt to increase adherence to headache apps is to send users daily reminders.

Different behavioral or motivational theory approaches offer explanations of how reminders may improve the target behavior (making the daily entry) [10]. For example, the COM-B model focuses on capability (C), opportunity (O), and motivation (M) as conditions of behavior (B) [11]. Regarding a diary app, the person needs to be proficient in handling a smartphone app (C), must have access to a smartphone and internet connection (O), and needs the motivation to fill in the diary. In this model, reminders might make the original motivation salient again, increasing the frequency of the behavior. The behavioral model developed by Fogg [12] introduces an additional element by assuming that the building blocks of behavior are the combination of motivation, ability, and triggers. A reminder is a prototypical example of a trigger, prompting existing motivation and ability into behavior. There also is the traditional concept of intrinsic vs extrinsic motivation, with the former describing that behavior is motivated "because it is inherently interesting or enjoyable," and the latter behavior that is motivated "because it leads to a separable outcome" [13]. A well-designed reminder might make the daily entry enjoyable. On the other hand, it may help to achieve a goal (headache documentation), so reminders may act on both intrinsic and extrinsic motivation. The self-determination theory is a modern theory of intrinsic motivation [14] and emphasizes that specific psychological needs (ie, autonomy, competence, and relatedness) must be satisfied to motivate human behavior. Reminders may promote a feeling of autonomy, as participants can, but do not have to, act upon them. Reminders may increase the feeling of competence by making it easier to reach the goal. They could also be specifically designed to create a sense of social relatedness. Also, concepts from digital literacy research may be useful for understanding the function of reminders, as this research suggests that both the user's ability and the system's usability must be present for a digital tool to be used effectively. In this framework, reminders increase the usability of a headache diary, as they make the diary easier to use [15].

The electronic headache diary of the German Migraine and Headache Society (DMKG; the smartphone app is called "DMKG-App") uses daily reminders [16]. A technical problem in early 2023 resulted in a 2-week period during which no reminders were sent. This allowed us to analyze if the absence of reminders had an effect on (1) the number of days recorded in the app (daily entries), (2) the delay to record the daily entry (entry delay), or (3) a disproportionately high number of users stopping use of the app altogether (attrition rate). In addition, we investigated whether age, gender, headache frequency, and previous adherence influence the reaction to reminder failure.

XSL•FO
RenderX

Methods

Study Design

This is a retrospective longitudinal cohort study using data from a smartphone headache diary app that sends daily reminders to make the daily entry. For technical reasons, reminders failed for an approximately 2-week period in early 2023. We investigated how adherence was affected by reminder failure by comparing the failure period with 2 periods with intact reminders and with the analogous periods from the previous year.

Setting and Data Sources

The analysis used data from the smartphone headache diary app of the German Migraine and Headache Society (DMKG-App), available in Germany for Android and iOS since June 2020 for users aged at least 16 years. In the DMKG-App, users are asked to record every day either as “no headache” or “headache.” In the case of “headache,” additional headache characteristics are assessed, which were not analyzed here. A reminder to make the daily entry is sent as a smartphone push notification every day around 8 PM (fixed time), reading, “Don’t forget your headache diary. Your DMKG-App is ready for today’s entry.” Entries can be made retrospectively for a maximum of 14 days in the past as a compromise between reducing recall bias and maintaining app usability.

Participants

The DMKG-App is available in Germany for users aged at least 16 years who use the app either for their personal headache documentation or to support their headache treatment. We included participants fulfilling the criteria of being an “active user” during one of the periods of interest (see “Statistical Analysis” section below).

Ethical Considerations

The DMKG-App is part of the larger project “DMKG Headache Registry” [16] that has been approved by the ethics committee of the Ludwig-Maximilians-University Munich (Nr. 20 - 004), complies with European and German Data Protection laws, and is registered with the German Clinical Trials Register (DRKS 00021081). The DMKG-App is freely available for use by persons with headaches in Germany.

Each participant provided informed consent for the use of their data for headache research. As health data constitute a special category of personal data, participants provided explicit consent to the processing and scientific use of their data in accordance with Article 9(2)(a) of the General Data Protection Regulation (GDPR) [17]. The present analysis also includes data from adolescents aged at least 16 years. Article 8(1) of the GDPR provides that, where an information society service is offered directly to a minor, individuals aged at least 16 years may provide consent themselves [18]. Although European Union (EU) Member States may establish a lower age threshold, Germany has not adopted such a lower threshold. The German Data Protection Conference also addresses these requirements in its Short Paper No. 20 [19] on consent under the GDPR, which refers both to the requirement of explicit consent for the processing of special categories of personal data and to the specific conditions under Article 8 of the GDPR for children’s consent in relation to information society services. The consent procedure for DMKG-App use was designed on this legal basis. Accordingly, participants aged 16 or 17 years provided the required informed and explicit consent themselves, and no additional consent from a parent or legal guardian was obtained.

Only anonymized data were used for the present analysis, and only aggregated results are reported. Individual users cannot be identified from any part of this manuscript.

Analyzed Failure and Reference Periods

The failure to send reminders occurred following the release of a new app version (2.0.0) on December 17, 2022, and was corrected with version 2.0.4 released on February 10, 2023. However, after an entry in the app is made, reminders are preplanned for the next 6 (42 d) weeks and then work independently of the app version. This resulted in active app users experiencing failure of reminders only toward the end of January 2023. In addition, not all users updated to the new version immediately. Therefore, although reminders slowly started to fail beginning on December 17, a general failure occurred between January 28 and February 10, 2023. Therefore, we used this 14-day period for analysis. For comparison, we chose 14-day reference periods before December 17 (December 3-16, 2022) and in equal distance after the failure period (March 25 to April 07, 2023). We called these analysis periods the “prefailure,” “failure,” and “postfailure” period, respectively.

To exclude the possibility that circannual variations affected the results, we also analyzed the same time periods in the preceding year (ie, precontrol: December 3-16, 2021; control: January 28 to February 10, 2022; postcontrol: March 25 to April 07, 2022).

Variables

Users can record every day either as “no headache” or “headache,” or make no entry. Both “no headache” or “headache” days counted as recorded days. From this measure, we calculated 3 outcome parameters for each period: the number of recorded days (“daily entries”), the delay to make the daily entry (measured in days; the app allows making entries with a maximum delay of 14 d), and the attrition rate (users who had been active before the period but made no further entry during or after that period).

In addition, we used age, gender, headache frequency, and previous adherence categories to analyze if these variables modulated the effect of reminder failure. For age, 2 groups were defined by median split (younger <37 y and older ≥37 y). For headache frequency, categories were chosen according to migraine frequency categories [20] (low, medium, or high frequency: 0 - 7 d/mo; 8-14 d/mo; ≥15 d/mo, respectively) and determined from the entries in October 2022 (ie, before the prefailure period). Only participants who had made more than 70% of the possible entries in this time frame were included to reduce bias in headache frequency detection related to low adherence. For previous adherence, the number of entries in October 2022 was used. Most participants made daily entries (31 days, which we termed: “high previous adherence”), the

XSL•FO
RenderX

remainder was more or less evenly distributed between 0 - 14 entries ("low previous adherence") and 15 - 30 entries ("medium previous adherence").

There were no missing data, as age and gender are mandatory entries in the DMKG-App, and the fact that some days were recorded and others not was the subject of our analysis.

Statistical Analysis

Statistical analysis was made in R (R Foundation for Statistical Computing; version 4.4.0). 95% CIs are given. $P<.05$ was considered significant. Post hoc tests comparing the 3 periods were corrected for 3 comparisons with Bonferroni Holm correction, and corrected P values (P_c) are reported [21].

As the number of app users changes continuously (new users joining and previous users ceasing to use the app; for the DMKG-App there currently is a net increase in users), we based our analysis on the number of active app users on a specific day, defined as users with at least 2 entries in the app, 1 before the analyzed day and 1 on or after the analyzed day.

To analyze the number of recorded days ("daily entries") during the different periods, we compared the number of entries within the respective 14-day period per active user between periods (prefailure, failure, and postfailure) using ANOVA followed by a Wilcoxon rank sum test with continuity correction. For a day-by-day graphical illustration, the number of entries on every day was set into proportion to active users on this day (Figure 1).

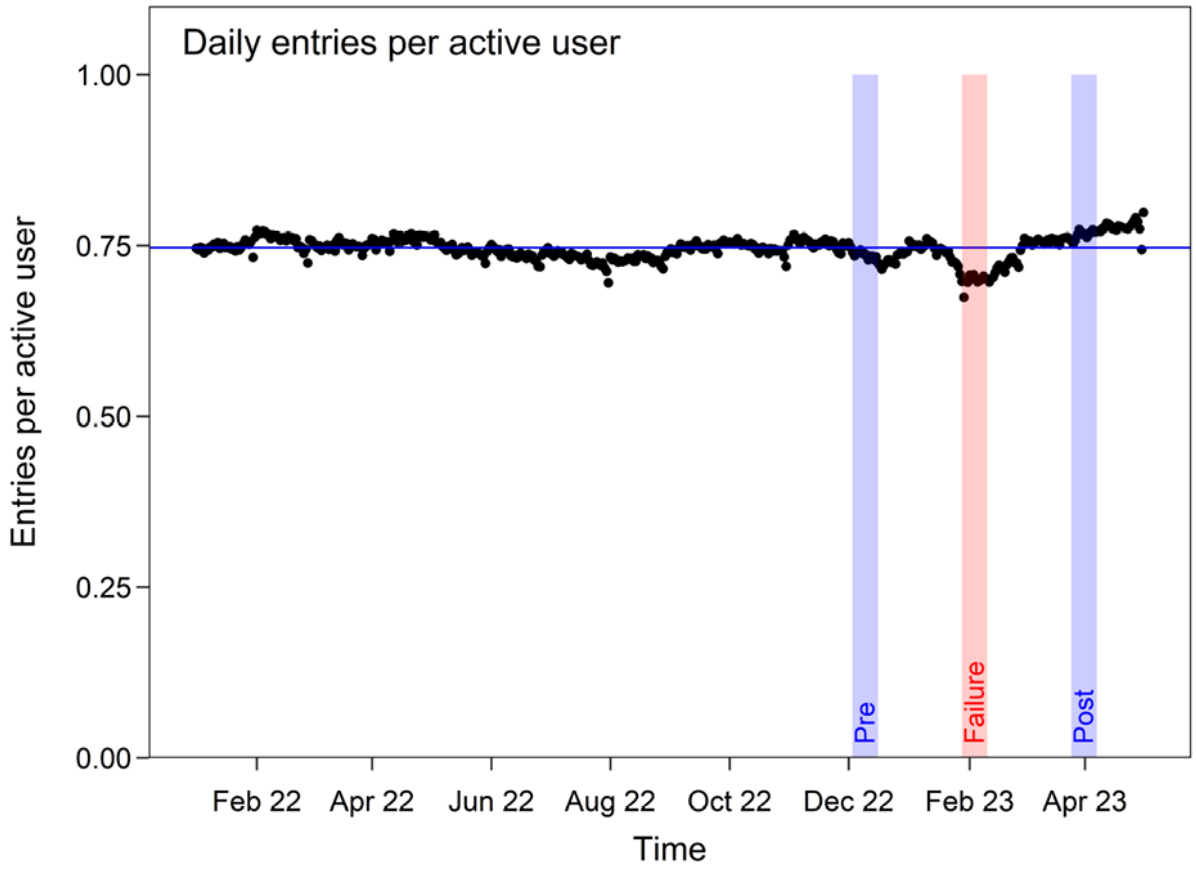
For the delay in recording the daily entry ("entry delay"), the number of days between the recorded day and the day the entry was made was calculated, averaged within periods and compared between periods using ANOVA followed by Welch t tests (2-tailed). If the entry was made on the day itself, the delay was 0. Only days with an entry were considered.

To analyze the proportion of users stopping app use ("attrition rate") during the different periods, all active users on the day before the start of the respective period were considered. For every day during the analyzed period, users that had not made any further entry on that day or any day after this (up to data closure on June 20, 2023) were considered to have stopped app use. Proportions of users having stopped app use at the end of the respective period were compared between periods using logistic regression.

To analyze if age, gender, headache frequency, and previous adherence modulated the effect of reminder failure on daily entries, delay, and attrition rate, categorical variables as defined above were included in the respective ANOVA or logistic regression.

XSL•FO
RenderX

Figure 1. Time course of the number of daily entries with respect to active app users. This was a longitudinal analysis of data from the Deutsche Migräne- und Kopfschmerzgesellschaft e.V. (German Migraine and Headache Society, DMKG) app, a headache diary app freely available in Germany, comparing a period where no daily reminders were sent with periods with reminders. A significant dip in the daily entry rate is seen during the period where no reminders were sent (failure). Each data point represents 1 day. Please note that the red bar denotes the absolute failure period. For some users, failure occurred earlier and/or lasted longer (see "Methods" section), explaining why the reduction of daily entries extended beyond the failure period. Prefailure and postfailure periods are marked by blue bars.



Results

The number of active users at the start of the reminder failure period was 4236. A total of 3460 (81.6%, 95% CI 80.5 - 82.8) were female, and the average age was 39.8 (SD 12.8, 95% CI 39.4 - 40.2) years.

Number of Recorded Days in the Diary (Daily Entries)

Figure 1 shows the course of daily entries in relation to active users, displaying a remarkable dip during the period where no reminders were sent. Active users made 10.2 (SD 4.3, 95% CI 10.1 - 10.3) of 14 possible entries during the prefailure period, 9.1 (SD 4.6, 95% CI 9.0 - 9.2) during the failure period, and 10.2 (SD 4.3, 95% CI 10.1 - 10.3) in the postfailure period. There was a significant difference between the 3 periods ($F_{2,11498}$=79.2; P<.001). Post hoc tests revealed significant differences between the prefailure and failure periods (z=11.9; P_c<.001) and between the failure and postfailure periods (z=12.5; P_c<.001) but not between the prefailure and postfailure periods (z=0.1; P_c=.90).

This was not an effect of time of the year, as no such effect was seen in the respective periods in the previous year (prefailure: mean 10.4 (SD 4.0, 95% CI 10.2 - 10.6) entries; failure: mean 10.5 (SD 4.1, 95% CI 10.3 - 10.7) entries; postfailure: mean 10.5 (SD 42, 95% CI 10.3 - 10.7) entries; $F_{2,5134}$<0.1; P=.91).

Thus, there was a reduction of daily entries per active user of 11.0% (95% CI 9.2 - 12.8, from 10.2 to 9.1 entries per 14-d period) from the prefailure to the failure period.

Delay to Make the Daily Entry

Figure 2 illustrates that the typical average delay in recording was around 2 days, which sharply increased to almost 3 days during the failure period. Specifically, average delays were 1.90 (SD 0.12, 95% CI 1.89 - 1.90) days in the prefailure period, 2.78 (SD 0.20, 95% CI 2.77 - 2.79) in the failure period, and 1.85 (SD 0.14, 95% CI 1.84 - 1.86) days in the postfailure period. This was highly significant between periods ($F_{2,39}$=155.4; P<.001). Post hoc tests revealed significant differences for comparison of prefailure to failure ($t_{20.5}$=−14.0; P_c<.001) and failure to postfailure ($t_{22.7}$=14.1; P_c<.001) but not prefailure vs postfailure ($t_{25.3}$=1.0; P_c=.24).

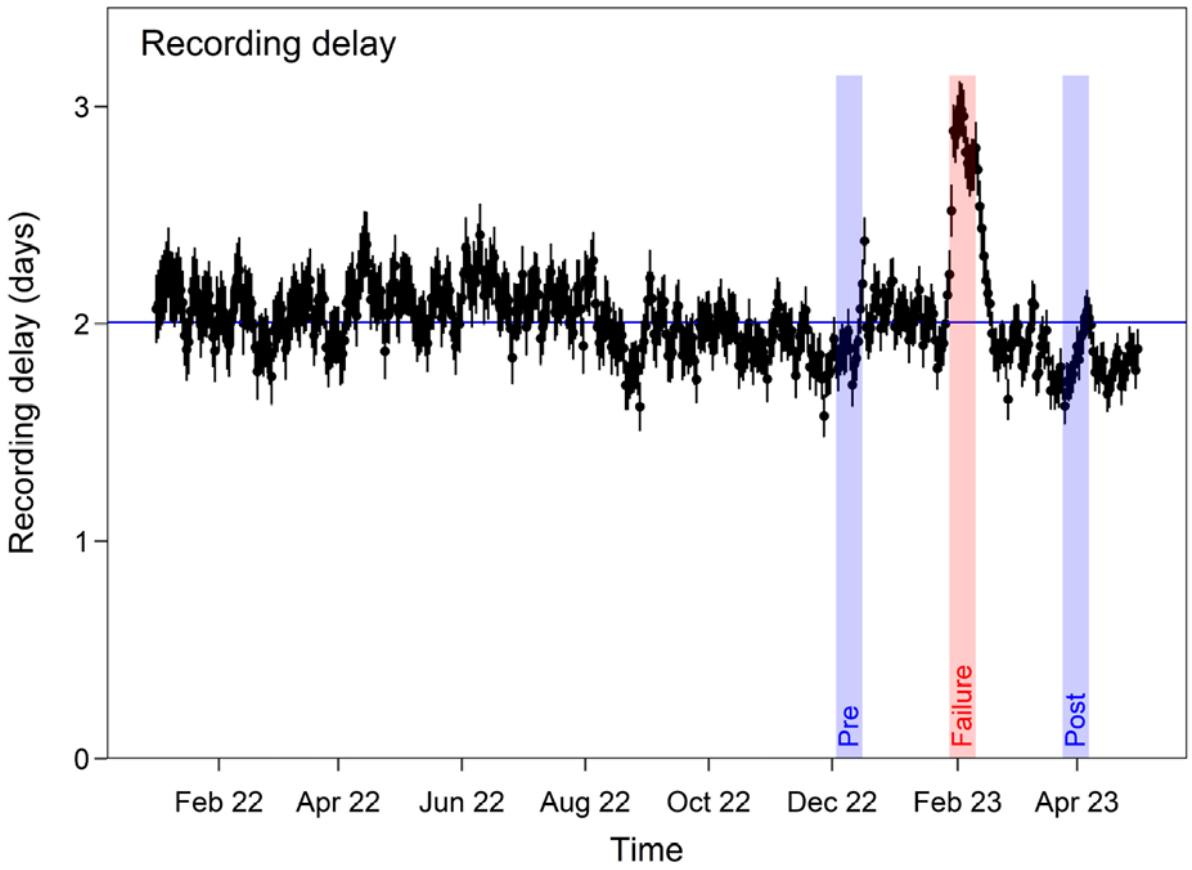
Again, this was not due to circannual variation, as the values for the respective periods of the preceding year were: precontrol

XSL•FO
RenderX

mean 2.08 (SD 0.07, 95% CI 2,08 - 2,08), control mean 2.03 (SD 0.10, 95% CI 2.02 - 2.04) and postcontrol mean 1.97 (SD 0.11, 95% CI 1.96 - 1.98). While ANOVA detected significant differences between periods ($F_{2,39}$=5.9; P=.006), the control period was not significantly different from either precontrol or postcontrol periods (precontrol vs control periods: $t_{23.5}$=1.68; P_c=.11; control vs postcontrol: $t_{25.8}$=1.72; P_c=.19). However, there was a small but significant reduction from precontrol to postcontrol ($t_{22.4}$=3.5; P_c=.002) that was not further explored.

The increase of recording delay from the prefailure period (1.90) to the failure period (2.78) corresponds to a relative increase of 46.3% (95% CI 45.8 - 46.8).

Figure 2. Time course of the delay in recording the daily entry. This was a longitudinal analysis of data from the Deutsche Migräne- und Kopfschmerzgesellschaft e.V. (German Migraine and Headache Society, DMKG) app, a headache diary app freely available in Germany, comparing a period where no daily reminders were sent with periods with reminders. A significant rise in delay is seen during the period where no reminders were sent (failure). Each data point represents 1 day; error bars correspond to SDs. Prefailure and postfailure periods are marked by blue bars.

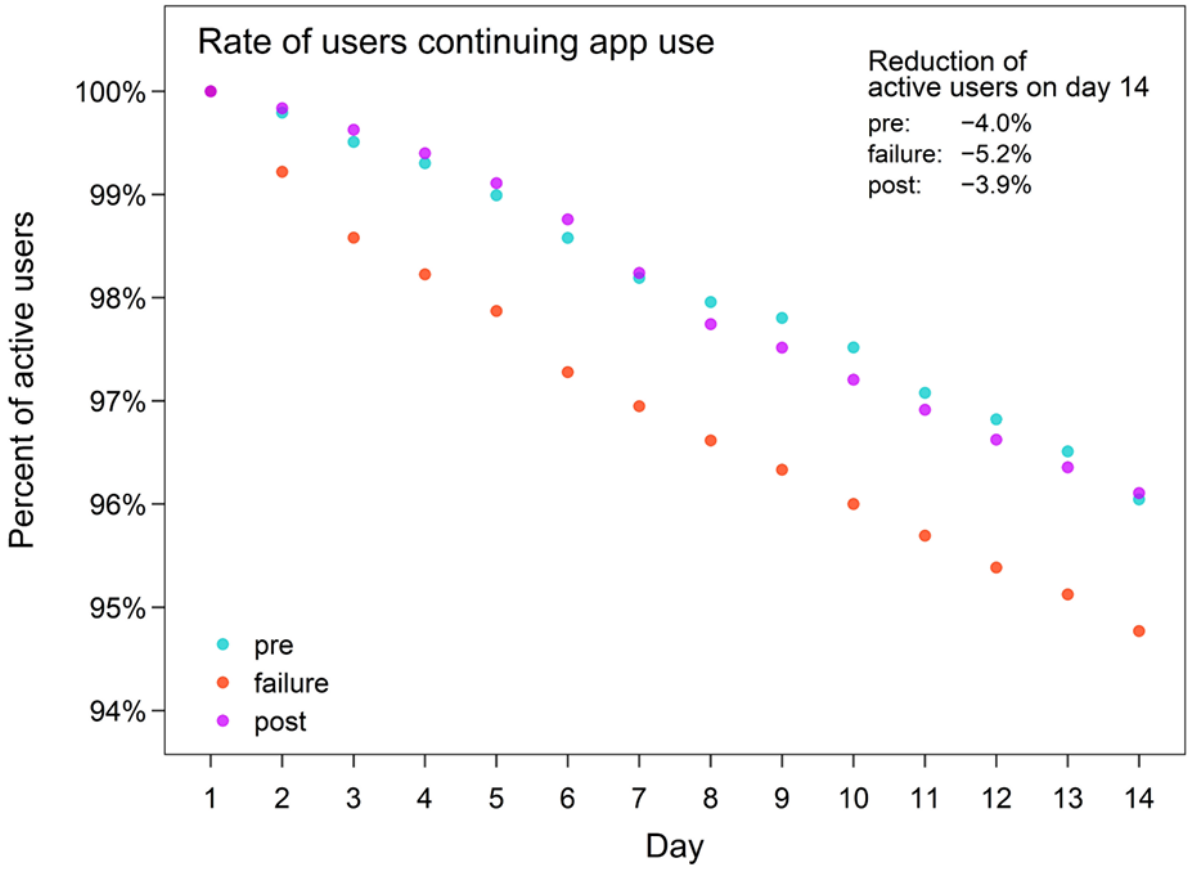


Users Stopping App Use (Attrition Rate)

There is always a certain number of users stopping app use. From all users classified as active before the day where reminders failed, a total of 5.2% (220/4236 users, 95% CI 4.6 - 5.9) stopped using the app for good (ie, made no further entry after that). In comparison, in the prefailure and postfailure periods, 4.0% (155/3886, 95% CI 3.4 - 4.6) and 3.9% (189/4845, 95% CI 3.4 - 4.5) stopped app use (Figure 3). Logistic regression revealed significant differences between periods (prefailure vs failure: z=2.8; P=.006; postfailure vs failure: z=3.10; P=.002). Also, in the previous year, attrition rates were 2.8% (49/1735, 95% CI 2.1 - 3.7), 3.0% (58/1929, 95% CI 2.3 - 3.9) and 2.9% (64/2192, 95% CI 2.3 - 3.7) during the precontrol, control, and postcontrol periods, without significant differences (prefailure vs failure: z=0.1; P=.91; postfailure vs failure: z<0.1; P=.99).

XSL•FO
RenderX

Figure 3. Attrition rates during failure of reminders vs before and after failure of reminders. This was a longitudinal analysis of data from the Deutsche Migräne- und Kopfschmerzgesellschaft e.V. (German Migraine and Headache Society, DMKG) app, a headache diary app freely available in Germany, comparing a period where no daily reminders were sent (failure) with 2 periods with intact reminders (prefailure and postfailure). The rate of previously active users who continued using the app during the failure period compared to the prefailure and postfailure periods.



Thus, there was an excess loss of 1.2% (5.2% in the failure period vs 4.0% in the prefailure period) of active users within the period with notification failure, or, put differently, the loss of active users during the failure period increased by 32.2% (from 4.0% to 5.2%, 95% CI 8.1 - 61.7).

Interaction With Age

There was no significant interaction between age and period, neither for the number of recorded days nor for the delay to make the daily entry, nor for attrition rates (see Table S1 in Multimedia Appendix 1 for details).

Interaction With Gender

The increase of the delay to make the daily entry from the prefailure to the failure period was significantly lower in females than in males (+0.85, 95% CI 0.84 - 0.86) days vs +1.03 (95% CI 1.00 - 1.06) days; interaction between gender and period: $F_{2,78}$=5.0; P=.009). There was no interaction between gender and period regarding the number of recorded days or the attrition. Details are listed in Table S1 in Multimedia Appendix 1. In summary, females were less sensitive than males to reminder failure regarding the delay to make the daily entry.

Interaction With Headache Frequency

Results are summarized in Table S2 in Multimedia Appendix 1. For the number of recorded days, there was an interaction between period and headache frequency ($F_{4,5536}$=3.3; P=.01), with the prefailure to failure reduction being significantly larger in the low frequency group (−1.49, 95% CI −1.10 to −1.76 entries per 14 d) than in the 2 other groups (medium: −0.79, 95% CI −0.42 to −1.16; P=.004; high: −0.88, 95% CI −0.39 to −1.37; P=.04). For the delay to make the daily entry, there was an interaction between period and headache frequency ($F_{4,117}$=18.9; P<.001), with the prefailure to failure increase being larger with lower headache frequencies (low: +1.17, 95% CI 1.15 - 1.19; medium: +0.87, 95% CI 0.85 - 0.89; high: +0.60, 95% CI 0.58 - 0.62 days; all P<.001). For the attrition rate, there was no interaction between period (pre vs failure) and headache frequency (z=−0.64; P=.53).

In summary, patients with lower headache frequencies were more sensitive to reminder failure regarding the number of recorded days and the delay to make the daily entry.

Interaction With Previous Adherence

Results are summarized in Table S2 in Multimedia Appendix 1. For the number of recorded days, there was an interaction between period and previous adherence ($F_{4,6947}$=5.8; P<.001),

XSL•FO
RenderX

with the prefailure to failure decrease being larger in the high compared to the low previous adherence group (−1.21, 95% CI −1.44 to −0.98 d vs +0.13, 95% CI −0.60 to 0.86; P=.002). For the delay, there was an interaction between period and previous adherence ($F_{4,117}$=10.7; P<.001), with prefailure to failure increases of the delay being largest in the high previous adherence group (low: +0.77, 95% CI +0.73 to +0.81; medium: +0.81 (95% CI +0.78 to +0.84); high: +0.98 (95% CI +0.97 to +0.99) days; all 3 comparisons significant, Table S2 in Multimedia Appendix 1). For the attrition rate, there was a significant interaction between period (pre vs failure) and previous adherence (z=−4.22; P<.001). Excess attrition rates (pre vs fail) were highest in the high previous adherence group (low: −1.9% (95% CI −4.4 to +0.5); medium: −0.5% (95% CI −2.8 to +1.7); high: +2.7% (95% CI +1.6 to +3.8); both comparisons with high previous adherence were significant, Table S2 in Multimedia Appendix 1).

In summary, patients with high adherence during the presence of daily reminders were most sensitive to failure of these reminders.

Discussion

Principal Findings

This study demonstrates that daily reminders can significantly contribute to the adherence to electronic headache diaries. Absence of reminders during a 14-day period resulted in a measurable decline in daily entries by 11% and an increase of daily entry delays by 46% and attrition rates by 30%. Moreover, males, participants with lower headache frequencies, and with higher previous adherence showed a larger decrease in adherence in reaction to the reminder failure, while age groups did not differ in reaction to reminder failure.

The present results show that daily reminders indeed improved adherence to the DMKG-App electronic headache diary. From the different behavioral and motivational theories, Fogg's model may be best suited to understand the effect of reminders. In accordance with this model, smartphone reminders serve as external triggers, sometimes also called signals or nudges, to perform the desired behavior and can be seen as elements of persuasive technology [12,22]. Behavioral model developed by Fogg additionally lists ability and motivation as prerequisites for behavior and the digital literacy literature emphasizes that both ability and usability must come together for a digital tool to be used effectively [15]. In the DMKG-App, reminders are implemented as push notifications taking the user directly to the daily entry. Failure of reminders would be supposed to reduce usability [15] by making the daily entry more difficult, as the user has to locate and open the app, and choose the correct day. In addition, in the DMKG-App, the timing of the reminders (8 PM) is chosen to increase the salience of the original motivation [23,24] at a time when the ability of the users to perform the behavior is high (if a headache occurs this day, it will most probably already have occurred; most participants will be back from work but not yet too tired). This effect could be further improved by personalizing the time of the reminder based on the time of previous entries, as personalization is a well-established persuasive design principle [25].

Motivation may also decrease with reminder failure, as it becomes more difficult to reach the desired goal (having a good headache documentation) [22,23]. Reminders offer additional possibilities to increase motivation (not implemented in the DMKG-App), for example, making the response to the daily reminder enjoyable by including elements of gamification (like receiving points or seeing interesting animations). Regarding self-determination theory [14], DMKG-App reminders use language emphasizing user autonomy ("Your DMKG-App is ready for today's entry" instead of "You need to make your entry now" or "Your entry for today is still missing"). The feeling of autonomy could be further increased by allowing the user to choose the timing of the reminders. Up to now, the DMKG-App does not build on social relatedness, which might be an additional possibility to increase the effect of reminders according to self-determination theory [14]. Relatedness could be promoted by including the user's name in the reminders, by including information like "xxx DMKG-App users have already made the entry for today," or by using chatbot elements starting a reminder with "Can I help you to make your entry for today," possibly with personalized reactions like "I am sorry that you had a strong headache today." However, this carries a risk of bias, so that the effect of such an intervention on data quality would need careful evaluation.

While the reduction of daily entries and the excess attrition of users during the failure period may seem modest, the delay to perform the daily entry increased from 1.90 days in the prefailure period to 2.78 days in the failure period. This is a large effect that will likely increase recall bias. The exact clinical significance of a +0.9 day shift in entry delay is difficult to estimate, as most recall bias research in headache compared headache diaries to retrospective ratings after longer periods such as 1 - 3 months [2,26,27]. Results from time-diary research suggest that a 24 hours delay can result in an error of approximately 10% [28]. Our results therefore may have implications for both clinical headache care and scientific use of these data as both rely on data quality. Our data suggest that persons with low headache frequencies and those with high previous adherence are most sensitive to the reminder failure. The first relation seems reasonable as the headache itself might serve as a trigger to make the entry (within theory by Fogg [12]) and motivation (within Fogg or COM-B frameworks [11,12]) may be higher with more frequent headaches as headache-related disability is larger. Indeed, there is some previous evidence that adherence to a headache diary is better on headache days than on headache-free days [29].

On the other hand, in participants with low adherence despite receiving daily reminders, reminders seem to be largely ineffective anyway, possibly explaining why participants with higher adherence in the presence of reminders had a larger drop in adherence when reminders failed. However, there is no literature to support this speculation. In our data, there was also a small effect of males having a larger increase in recording delay with reminder failure than females. There is evidence that women are more socialized to engage in preventive health care and more readily integrate health behavior in their daily routines [30], most likely pointing toward an increased motivation for the behavior. Some (but not all) previous studies on mobile

XSL•FO
RenderX

health apps showed better adherence in females than in males [31], and such a relation has also been reported in a headache diary study [32], but the specific effect of reminders has not been investigated. In addition, there is some evidence that males may respond more to gamification elements than women in the health context (eg, [33]). Reminders may be regarded as a basal level of gamification. Further studies are needed to assess gender differences in adherence and personalized interventions may result from this.

One result from our study needs further comment. The attrition rates were generally lower in the previous year (2.8% - 2.9% in the precontrol and postcontrol periods) than in the year where the failure occurred (3.9% - 4.0% in the prefailure and postfailure periods). This is likely due to the fact that the DMKG-App was made freely available in Germany between these 2 time points. Before that, its use was limited to users being treated at a center participating in the DMKG Headache Registry, likely leading to lower attrition rates.

Comparison to Prior Work

There are not many recent studies investigating adherence to an electronic headache diary, and it is difficult to compare results because of different approaches. An analysis of 1561 users of the Curelator / N1 headache app showed that adherence strongly depended on the setting, being highest for a paid version, intermediate after physician recommendation using a download coupon, and lowest for a free version. In the Curelator / N1 headache analysis, 90-day adherence (proportion of entered days) was 66%, 44% and 34% in the 3 settings, respectively [34]. These values cannot be directly compared to our results showing 70% - 75% adherence, as we included only "active users" in the analysis (see "Methods" section). On the other hand, in a study including 1009 patients of the Leiden Headache Center using the headache center's e-diary for a minimum of 2 weeks and a median of 181 days, adherence was very high with only 10.8% of the daily entries missing [32]. This is likely due to this study being based on patients from 1 tertiary headache center who were explicitly asked to use the e-diary for their clinical follow-up or study participation. A study with adolescents treated at a tertiary headache center found that over 90 days, 68.7% of the entries within an internet-based headache diary were made on the same day. This rate was higher for headache days with acute medication intake [29]. A study using a headache diary app also providing progressive muscle relaxation that recruited participants at a tertiary headache center found an average of 51 days entered during the first 90 days (adherence 57%), but adherence sharply declined over time [35].

It has been emphasized that there are many factors influencing health-related electronic diary use, including app design (usability, content, feedback on entered data, technical reliability) and user factors (age, gender, digital skills, and symptom load) [4]. Regarding the effect of reminders on diary adherence, there are some reports suggesting that reminders increase use of health-related diaries, but there is a lack of formal investigation and no specific data on headache diaries [4]. Early studies on pain (not headache) diaries showed that reminders increased compliance with a paper diary (ie, entry within ±15 min of assigned time, 3 times daily) from 11% to 29% and use of an electronic diary with reminders further increased compliance to 94% over a 21-day study period [3,36].

Daily entries in headache diaries are crucial for obtaining accurate and reliable data, in both clinical practice and research. Consistent daily tracking enables the identification of patterns, triggers, and treatment effects, allowing for more personalized treatment strategies [2]. Regular use of an electronic headache diary was associated with significant improvements in headache and migraine frequency, intensity, and use of acute medication over a 6-month period [37]. These findings underscore the potential of electronic headache diaries to not only enhance data quality but also to actively support standard headache care by raising awareness for the individual headache disorder. The use of reminders likely can further promote adherence to daily entries and more effective treatment adjustments [37].

Strengths and Limitations

Strengths of the present study are the large sample size and the comparison with the same periods in the previous year. There also are some limitations. First, the failure period was relatively short, so it remains open if adherence would have deteriorated further or leveled out during a longer period. Because reminders arrived regularly before the failure period, users may have been relying on reminders for their daily entry, and it cannot be excluded that they would have started regular use of the app without reminders after a while. Indeed, the delay to perform the daily entry was mitigated somewhat after the first few days (Figure 2); however, a return to previous values only occurred after the end of the reminder failure. These questions could only be answered by a dedicated study, for example, comparing different reminder failure periods.

Second, total adherence rates determined within the present study may not be generalizable as we included only active users in the analysis. This was necessary to perform a meaningful analysis of the effect of the reminder failure. However, we have performed an additional analysis of the effect of previous adherence on the reaction to the reminder failure, showing that persons with high previous adherence (during a period with reminders) were more sensitive to reminder failure than persons with lower previous adherence. In future, dedicated studies might investigate the effect of reminders on new users specifically within the first weeks of app use–one could speculate that these initial reminders could help users engage in regular entries. It would also be interesting if sending reminders to users after a period of inactivity can bring them back to regular use.

Third, although the present study showed that daily reminders increase adherence to an electronic headache diary, it remains an open question what type of reminder would have the largest effect on adherence, a question that has also been discussed in other mobile health apps [38]. Future research could further refine understanding of reminders in headache diaries by comparing groups randomly assigned to different conditions, for example, receiving reminders vs no reminders from the beginning of app use vs reminders ceasing after a fixed time of app use with a longer follow-up period without reminders. They could also compare different types of reminders directed at

XSL•FO
RenderX

different effects, for example, increasing ability and/or motivation in addition to serving as a prompt according to behavioral theory by Fogg [12].

Conclusion

This is the first study on the effect of daily reminders in a digital headache diary app. Present results suggest that daily reminders are an effective means to improve adherence to electronic headache diaries, specifically by increasing the number of documented days and by decreasing the delay to make the daily entry. Males, persons with low headache frequencies, and high previous adherence were more sensitive to reminder failure. These results contribute to our knowledge on compliance with digital health apps and point out means to improve adherence and groups especially benefiting from reminders.

Acknowledgments

We wish to thank all DMKG-App users. No generative AI has been used in writing this manuscript.

Funding

The DMKG Headache Registry is funded by the German Migraine and Headache Society and by unrestricted grants or funding of specific analyses from companies with an interest in headache research (up to the time of submission of this work, Lilly, Teva, Novartis, Pfizer, and Lundbeck have made contributions). Supporting companies were given the opportunity to comment on the research protocol, and their comments were considered and judged to be scientifically relevant. Apart from that, supporting companies did not influence the study design.

Data Availability

The datasets analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: RR, JS, YS
Data curation: JS, YS
Formal analysis: JS, YS
Funding acquisition: LN, TK, GG, TPJ
Investigation: RR, BR, TF, GG
Methodology: RR, JS, YS
Project administration: RR
Resources: LN, TK, GG, TPJ
Supervision: RR, LN, GG, TPJ
Validation: RR, BR, TF, TD, VR
Visualization: RR, JS, YS
Writing – original draft: RR
Writing – review & editing: LN, BR, TK, TF, TD, GG, TPJ, VR

Conflicts of Interest

RR has received travel grants and/or honoraria for lectures or advisory boards, or has consulted for AbbVie, Betapharm, Lundbeck, Novartis, Pfizer, and Teva.
LN received honoraria for lectures and advisory boards from Hormosan, Lilly, Novartis, Teva, AbbVie/Allergan.
BR has received research grants from Lundbeck and Novartis, and honoraria for lectures or advisory boards from AbbVie, Lundbeck, Novartis, Organon, Perfood, and Teva. TK has received travel grants and/or honoraria from Allergan/AbbVie, Hormosan, Lilly, Novartis, and Teva.
TK has received travel grants and/or honoraria from Allergan/AbbVie, Hormosan, Lilly, Novartis, and TEVA.
TF received honoraria for advisory boards from Novartis, Teva, and Lundbeck and honoraria for lectures from Novartis, Teva, Lilly, Bayer, and UCB.
TD has received honoraria for consulting and lectures from Hormosan Pharma, Novartis Pharma, and TEVA.
GG has received honoraria from Lilly, Novartis Pharma, Teva, Lundbeck, and research grants from Novartis Pharma during the last 3 years.
TPJ has received honoraria from Allergan/AbbVie, Grünenthal, Hormosan, Lilly, Lundbeck, Novartis, Orion, Perfood, Sanofi, and Teva. He served on Advisory Boards and/or consulted for Allergan/AbbVie, Hormosan, Lilly, Lundbeck, Novartis, and Teva.
VR received the Soyka award as a research grant from Novartis. She has received honoraria from Teva.
JS and YS declare no conflict of interest.

Multimedia Appendix 1

XSL•FO
RenderX

Supplementary tables and interactions with age, gender, headache frequency, and previous adherence rates.
[DOCX File, 20 KB - mhealth_v14i1e73259_app1.docx]

References

1. Tassorelli C, Sances G, Allena M, et al. The usefulness and applicability of a basic headache diary before first consultation: results of a pilot study conducted in two centres. Cephalalgia 2008 Oct;28(10):1023-1030. [doi: 10.1111/j.1468-2982.2008.01639.x] [Medline: 18624804]
2. van Casteren DS, Verhagen IE, de Boer I, et al. E-diary use in clinical headache practice: a prospective observational study. Cephalalgia 2021 Oct;41(11-12):1161-1171. [doi: 10.1177/03331024211010306] [Medline: 33938248]
3. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. Control Clin Trials 2003 Apr;24(2):182-199. [doi: 10.1016/s0197-2456(02)00320-3] [Medline: 12689739]
4. Daniëls NEM, Hochstenbach LMJ, van Zelst C, van Bokhoven MA, Delespaul PAEG, Beurskens AJHM. Factors that influence the use of electronic diaries in health care: scoping review. JMIR Mhealth Uhealth 2021 Jun 1;9(6):e19536. [doi: 10.2196/19536] [Medline: 34061036]
5. Stinson JN, Huguet A, McGrath P, et al. A qualitative review of the psychometric properties and feasibility of electronic headache diaries for children and adults: where we are and where we need to go. Pain Res Manag 2013;18(3):142-152. [doi: 10.1155/2013/369541] [Medline: 23748255]
6. Hundert AS, Huguet A, McGrath PJ, Stinson JN, Wheaton M. Commercially available mobile phone headache diary apps: a systematic review. JMIR Mhealth Uhealth 2014 Aug 19;2(3):e36. [doi: 10.2196/mhealth.3452] [Medline: 25138438]
7. Bandarian-Balooch S, Martin PR, McNally B, Brunelli A, Mackenzie S. Electronic-diary for recording headaches, triggers, and medication use: development and evaluation. Headache 2017 Nov;57(10):1551-1569. [doi: 10.1111/head.13184] [Medline: 28925498]
8. Chiang CC, Fang X, Horvath Z, et al. Simultaneous comparisons of 25 acute migraine medications based on 10 million users' self-reported records from a smartphone application. Neurology 2023 Dec 12;101(24):e2560-e2570. [doi: 10.1212/WNL.0000000000207964] [Medline: 38030397]
9. Petrušić I, Ha WS, Labastida-Ramirez A, et al. Influence of next-generation artificial intelligence on headache research, diagnosis and treatment: the junior editorial board members' vision - part 1. J Headache Pain 2024 Sep 13;25(1):151. [doi: 10.1186/s10194-024-01847-7] [Medline: 39272003]
10. Van Lange PAM, Kruglanski AW, Higgins ET. Handbook of Theories of Social Psychology: SAGE Publications Ltd; 2012.
11. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6(1):42. [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
12. Fogg BJ. A behavior model for persuasive design. Presented at: The 4th International Conference on Persuasive Technology; Apr 26, 2009. [doi: 10.1145/1541948.1541999]
13. Ryan RM, Deci EL. Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp Educ Psychol 2000 Jan;25(1):54-67. [doi: 10.1006/ceps.1999.1020] [Medline: 10620381]
14. Ryan RM, Deci EL. Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness: The Guilford Press; 2017.
15. van Dijk J. The Digital Divide: John Wiley & Sons; 2020.
16. Ruscheweyh R, Klonowski T, Goßrau G, et al. The headache registry of the German Migraine and Headache Society (DMKG): baseline data of the first 1,351 patients. J Headache Pain 2022 Jul 1;23(1):74. [doi: 10.1186/s10194-022-01447-3] [Medline: 35773650]
17. Art. 9 GDPR – processing of special categories of personal data. General Data Protection Regulation (GDPR). URL: https://gdpr-info.eu/art-9-gdpr [accessed 2026-07-28]
18. Art. 8 GDPR – conditions applicable to child's consent in relation to information society services. General Data Protection Regulation (GDPR). URL: https://gdpr-info.eu/art-8-gdpr [accessed 2026-07-28]
19. Kurzpapier nr. 20: einwilligung nach der DS-GVO [Article in German]. : Datenschutzkonferenz (DSK); 2019 URL: http://www.datenschutzkonferenz-online.de/media/kp/dsk_kpnr_20.pdf [accessed 2026-07-27]
20. Buse DC, Reed ML, Fanning KM, Bostic RC, Lipton RB. Demographics, headache features, and comorbidity profiles in relation to headache frequency in people with migraine: results of the American Migraine Prevalence and Prevention (AMPP) study. Headache 2020 Oct 14. [doi: 10.1111/head.13966] [Medline: 33090481]
21. Miller RG jr. Simultaneous Statistical Inference: Springer; 1981.
22. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152. [doi: 10.2196/jmir.2104] [Medline: 23151820]
23. Deci EL, Ryan RM. The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 2000 Oct;11(4):227-268. [doi: 10.1207/S15327965PLI1104_01]
24. Kruglanski AW, Shah JY, Fishbach A, Friedman R, Chun WY, Sleeth-Keppler D. A theory of goal systems. Adv Exp Soc Psychol 2002;34:331-378. [doi: 10.1016/S0065-2601(02)80008-9]

XSL•FO
RenderX

25. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24. [doi: 10.17705/1CAIS.02428]
26. Miller VE, Faurot KR, Palssson OS, et al. Comparing prospective headache diary and retrospective four-week headache questionnaire over 20 weeks: secondary data analysis from a randomized controlled trial. Cephalalgia 2020 Nov;40(13):1523-1531. [doi: 10.1177/0333102420949180] [Medline: 32799667]
27. van den Brink M, Bandell-Hoekstra EN, Abu-Saad HH. The occurrence of recall bias in pediatric headache: a comparison of questionnaire and diary data. Headache 2001 Jan;41(1):11-20. [doi: 10.1046/j.1526-4610.2001.111006011.x] [Medline: 11168599]
28. te Braak P, van Tienoven TP, Minnen J, Glorieux I. Data quality and recall bias in time-diary research: the effects of prolonged recall periods in self-administered online time-use surveys. Sociol Methodol 2023 Feb;53(1):115-138. [doi: 10.1177/00811750221126499]
29. Heyer GL, Rose SC. Which factors affect daily compliance with an internet headache diary among youth with migraine? Clin J Pain 2015 Dec;31(12):1075-1079. [doi: 10.1097/AJP.0000000000000208] [Medline: 25565588]
30. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125. [doi: 10.2196/jmir.5604] [Medline: 28428170]
31. Jakob R, Harperink S, Rudolf AM, et al. Factors influencing adherence to mhealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371. [doi: 10.2196/35371] [Medline: 35612886]
32. van der Arend BWH, Holwerda LJ, Verhagen IE, van Casteren DS, Timmers T, Terwindt GM. Practical experience with the use of electronic headache diaries and video consultations in migraine care from a longitudinal cohort study. Telemed J E Health 2024 Nov;30(11):2696-2703. [doi: 10.1089/tmj.2024.0105] [Medline: 39007201]
33. Forman EM, Manasse SM, Dallal DH, et al. Gender differences in the effect of gamification on weight loss during a daily, neurocognitive training program. Transl Behav Med 2021 Apr 26;11(4):1015-1022. [doi: 10.1093/tbm/ibaa091] [Medline: 32945860]
34. Seng EK, Prieto P, Boucher G, Vives-Mestres M. Anxiety, incentives, and adherence to self-monitoring on a mobile health platform: a naturalistic longitudinal cohort study in people with headache. Headache 2018 Nov;58(10):1541-1555. [doi: 10.1111/head.13422] [Medline: 30334248]
35. Minen MT, Jalloh A, Ortega E, Powers SW, Sevick MA, Lipton RB. User design and experience preferences in a novel smartphone application for migraine management: a think aloud study of the RELAXaHEAD application. Pain Med 2019 Feb 1;20(2):369-377. [doi: 10.1093/pm/pny080] [Medline: 29868895]
36. Broderick JE, Schwartz JE, Shiffman S, Hufford MR, Stone AA. Signaling does not adequately improve diary compliance. Ann Behav Med 2003 Oct;26(2):139-148. [doi: 10.1207/S15324796ABM2602_06] [Medline: 14534031]
37. Raffaelli B, Mecklenburg J, Overeem LH, et al. Determining the evolution of headache among regular users of a daily electronic diary via a smartphone app: observational study. JMIR Mhealth Uhealth 2021 Jul 7;9(7):e26401. [doi: 10.2196/26401] [Medline: 34255716]
38. Lanke V, Trimm K, Habib B, Tamblyn R. Evaluating the effectiveness of mobile apps on medication adherence for chronic conditions: systematic review and meta-analysis. J Med Internet Res 2025 Jul 31;27:e60822. [doi: 10.2196/60822] [Medline: 40743450]

Abbreviations

DMKG: Deutsche Migräne- und Kopfschmerzgesellschaft e.V. (German Migraine and Headache Society)
DMKG-App: the headache diary app of the German Migraine and Headache Society
DRKS: Deutsches Register Klinischer Studien (German Clinical Trials Register)
EU: European Union
GDPR: General Data Protection Regulation

Edited by S Brini; submitted 09.Mar.2025; peer-reviewed by H Kuru, J Payne, M Tatsumoto, P Kropp; revised version received 30.Jul.2026; accepted 03.Aug.2026; published 14.Sep.2026.

Please cite as:
Ruscheweyh R, Neeb L, Raffaelli B, Kraya T, Freilinger T, Dresler T, Gossrau G, Jürgens TP, Ruschil V, Scheidt J, Siebenhaar Y Daily Reminders Increase Adherence to an Electronic Headache Diary (DMKG-App): Retrospective Longitudinal Cohort Study JMIR Mhealth Uhealth 2026;14:e73259
URL: https://mhealth.jmir.org/2026/1/e73259
doi:10.2196/73259

XSL•FO
RenderX

© Ruth Ruscheweyh, Lars Neeb, Bianca Raffaelli, Torsten Kraya, Tobias Freilinger, Thomas Dresler, Gudrun Gossrau, Tim Patrick Jürgens, Victoria Ruschil, Jörg Scheidt, Yannic Siebenhaar. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Weight Loss and Engagement Among Active Users of a Dietitian-Led Digital Nutrition Platform: Retrospective Cohort Study

Frederick G B Goddard[1], PhD; Jessica Garay[2], PhD; Francisco Teixeira Neves[1], PhD; Alexandra Ligeti[1], PhD; Liudmila Zhaunova[1], PhD; Manuela Abreu[1], MA

[1]Healthium - Healthcare Software Solutions, S.A., Rua Andrade Corvo, n.º 242, First Floor, Room 106, Braga, Portugal

[2]Department of Nutrition and Food Studies, Syracuse University, Syracuse, NY, United States

Corresponding Author:
Manuela Abreu, MA
Healthium - Healthcare Software Solutions, S.A., Rua Andrade Corvo, n.º 242, First Floor, Room 106, Braga, Portugal

Abstract

Background: Approximately 2.5 billion adults worldwide were living with overweight or obesity as of 2022, highlighting the need for comprehensive weight management solutions. Digital nutrition platforms have emerged as an accessible solution that can be implemented at scale. Nutrium Care is an employer-sponsored program delivered via the Nutrium platform, facilitating dietitian-led nutrition counseling through consultations and in-app communication, while also enabling users to track food intake, water consumption, and physical activity.

Objective: This study set out to characterize longitudinal weight changes among active users of the Nutrium Care program and estimate associations between program engagement and weight loss.

Methods: This retrospective cohort study analyzed deidentified data from adult users with overweight or obesity enrolled in the Nutrium Care program. Users were included if they were taking action to manage their weight and had at least 3 months of follow-up data available; these users were referred to as active users. The primary outcome was weight loss after 3 and 6 months, estimated using piecewise linear mixed-effects models. To estimate the association between engagement and weight loss, more- and less-engaged users were matched on sociodemographic characteristics. Engagement was defined by the number of dietitian appointments, days of active platform use, and logging of food, water intake, and physical exercise.

Results: The study included 10,288 active users, predominantly female (n=6887, 66.9%) and aged 18 to 49 years (n=9741, 94.7%), who contributed a median of 6 (IQR 4 - 9) months of data to the cohort. In the first 3 months, active users lost 2.0 kg (2.4% of baseline body weight; 95% CI 1.9-2.0 kg) on average, increasing to 2.3 kg (2.8%; 95% CI 2.2-2.4 kg) by 6 months. Higher program engagement was associated with more weight loss, with the strongest association estimated for at least 1 dietitian appointment every 2 months, corresponding to 0.8 kg (95% CI 0.7-1.0 kg) additional weight loss after 3 months.

Conclusions: This retrospective cohort study addresses a critical research gap in hybrid nutrition interventions integrating professional dietitian support with an employer-sponsored digital nutrition program. Active users achieved weight loss, and regular contact with a dietitian was the strongest engagement factor associated with improved outcomes, supporting the potential role of professional support in digital weight management programs.

(JMIR Mhealth Uhealth 2026;14:e94732) doi:10.2196/94732

KEYWORDS

obesity; digital; nutrition; dietitian; weight

Introduction

The global prevalence of adult overweight and obesity has risen in recent decades and now represents one of the most significant public health challenges worldwide [1]. As of 2022, a total of 2.5 billion adults were living with overweight or obesity globally [2]. Obesity substantially increases the risk of chronic diseases, including type 2 diabetes, cardiovascular disease, and certain cancers [3], and contributes to an estimated 4 million deaths annually [4]. The associated economic burden was estimated to reach US $2.5 trillion in 2025 [5,6], highlighting the urgent need for scalable and sustainable solutions to weight management.

Multiple strategies have been used to support weight loss, including bariatric surgery, prescription medication (ie, appetite suppressants), and lifestyle interventions [7]. Recently, pharmacotherapy, particularly glucagon-like peptide-1 (GLP-1) receptor agonists, has gained significant attention due to its demonstrated effectiveness in reducing body weight [8]. However, long-term use of these drugs has not yet been well studied, and there are potential concerns related to loss of muscle mass and other health effects that warrant ongoing monitoring

XSL•FO
RenderX

[9]. Lifestyle interventions that combine dietary modification with physical activity remain a cornerstone of obesity management and are particularly effective when delivered with guidance from health care professionals, such as registered dietitians (RDs) [10,11].

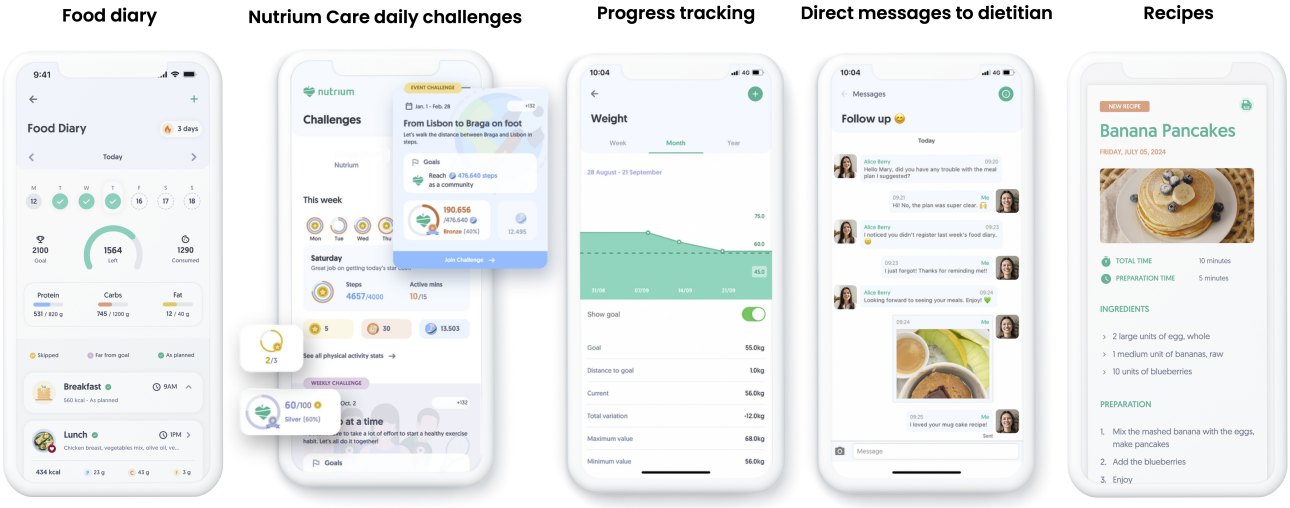
Despite their effectiveness, lifestyle interventions delivered through in-person counseling are difficult to scale to large populations. The emergence of digital health and nutrition platforms offers an alternative scalable approach to supporting weight management. These platforms typically combine individualized nutrition counseling with tools for dietary self-monitoring, educational resources, and behavioral prompts designed to encourage sustained lifestyle change. Although these platforms vary in their specific content and features, initial evidence suggests that mobile and web-based health technologies can effectively support improved diet quality and weight loss among adults with overweight or obesity [12-14].

However, digital nutrition platforms differ substantially in their design, level of professional involvement, and behavioral support mechanisms [15]. Evaluations of digital nutrition platforms consistently identify motivation, education, and support features as core user needs [16,17]. User engagement with specific digital intervention features is an important component of achieving weight loss outcomes. Use of self-regulation techniques has been identified as a predictor of successful weight loss among users of digital nutrition platforms [18,19]. A recent meta-analysis of digital food-focused interventions found positive effects on food consumption, particularly if the intervention used prompts or cues [20]. Positive cues that reinforce healthful behaviors can help to combat the "food noise" (persistent thoughts about food) many individuals report as a barrier to successful weight management [21]. Indeed, many GLP-1 users attribute the reduction in "food noise" to their weight loss [22]. However, the specific types and intensity of engagement associated with optimal weight outcomes remain unclear, particularly in dietitian-led platforms.

Inclusion of tailored feedback or individualized counseling has been demonstrated to enhance participant success in digital interventions focused on changing eating behavior [23-25]. An evaluation of a web platform in Brazil concluded that use of the platform in combination with coaching from a dietitian led to greater weight loss and significant changes in eating behavior (more fruits and vegetables, less ultraprocessed food) compared with users who only accessed the platform on their own [26]. Although limited in scope, these results suggest a high level of effectiveness when digital nutrition platforms include access to a dietitian. While the overall utility of digital nutrition platforms is generally positive, the heterogeneity in features, including the level of dietitian support, underscores the need for platform-specific evaluations. This allows for identification of individual components that best enhance weight management outcomes.

Nutrium is a digital health platform designed to facilitate hybrid nutrition interventions through integrated asynchronous and synchronous communication between users and RDs. Unlike automated or self-guided nutrition applications, which rely on algorithmic feedback [27], this model centers on practitioner-led care. The platform integrates scheduled teleconsultations with continuous in-app messaging, allowing RDs to develop personalized nutritional interventions, monitor real-time dietary intake, and provide iterative feedback throughout the care continuum. By digitizing the dietitian user interface while maintaining professional oversight (Figure 1), the Nutrium platform aligns with behavioral frameworks that prioritize engagement and therapeutic adherence as primary drivers of long-term health outcomes, such as weight management [28].

Figure 1. Examples of the in-app user interface.



Nutrium Care is one delivery model for the Nutrium platform, implemented via employer-sponsored health and wellness programs. Through this model, employees gain access to professional dietetic consultations and sustained nutritional support through the Nutrium platform. This study aimed to provide the first evaluation of the Nutrium Care delivery model by quantifying longitudinal weight loss outcomes among program users and assessing the relationship between weight loss and user engagement with the program. Specifically, the study aimed to (1) characterize absolute weight loss among active Nutrium Care users after 3 and 6 months; (2) describe the proportion of active users achieving weight loss thresholds of 3%, 5%, and 10% of baseline body weight in the action phase; and (3) estimate the association between user engagement with

XSL•FO
RenderX

the program and weight loss across multiple engagement metrics, including dietitian appointments, days of active platform use, and logging of food, water intake, and physical exercise.

Methods

Study Design

This study was a retrospective analysis of secondary observational data from users of the Nutrium Care program, specifically examining user weight outcomes and engagement with the app. The study was completed in 2 parts. First, we examined longitudinal weight data to estimate weight outcomes 3 and 6 months after users joined the Nutrium Care program. Second, we quantified the difference in weight outcomes at 3 months between more and less engaged users.

This evaluation considered users' readiness to engage with behavior change. The transtheoretical model (TTM) conceptualizes behavior change across 5 core stages: precontemplation, contemplation, preparation, action, and maintenance [29]. The action phase reflects individuals who have moved beyond intention into active, committed behavior change. Here, users who attended at least 2 dietitian appointments, with the second occurring within 6 months of baseline, were classified as action phase users [30,31]. This study therefore restricted its evaluation to users whose platform engagement was consistent with active behavior change, rather than incidental or exploratory use, by considering only users in the action phase. This restriction was applied during revision to ground the analysis in the TTM, rather than being prespecified. For the remainder of this manuscript, we refer to users in the action phase as active users.

To estimate differences in weight outcomes by engagement, we compared more and less engaged users matched on sociodemographic characteristics. For each engagement metric, we had 2 groups: one that included users who met the threshold for engagement (more engaged users) and one that included users who did not (less engaged users). Engagement metrics and their thresholds were appointments with dietitians (≥1 appointment at least every 2 months), days of active use (DAU) of the platform (≥10 DAU per month), and logging of food, water intake, and physical exercise (≥1 time per week on average). As no established thresholds exist across the included engagement metrics for digital nutrition platforms, the thresholds applied were exploratory and defined a priori by the research team.

Study Population

The study population included Nutrium Care users aged 18 to 65 years with overweight (BMI 25 - 29.9), obesity (BMI 30 - 39.9), or morbid obesity (BMI ≥40) at baseline. Baseline was defined as the date of the first available weight measurement. Users were excluded if they did not have a valid baseline weight, which we defined as a weight measured within 30 days of their first appointment with a dietitian, had missing data on age or sex, reported using weight loss medications (ie, GLP-1 agonists), or reported a pregnancy. Implausible weight (<30 kg or >300 kg) and BMI (>60) data were filtered out before applying the inclusion and exclusion criteria [32]. Users needed to be in the action phase and have at least 3 months of follow-up data to be included in the primary analysis. Results for an unrestricted sample without action phase and follow-up restrictions are presented alongside the primary analysis.

Data

The data consisted solely of information that was originally generated during ordinary, routine use of the Nutrium Care program. Participants were not prompted to enter data for research purposes. Instead, all information was submitted voluntarily in the normal course of platform operation and service delivery. The data were entered by dietitians as well as self-reported by users. Data extracted from the Nutrium database included sociodemographic variables (age and sex), anthropometrics (weight and height), and engagement with the Nutrium application (including the number of dietitian appointments, exercise, food and water intake logging, DAU, and dietary goals). Users recorded weight by either entering a value manually in the application, reporting it to their dietitian during a consultation for entry into the user's record, or synchronizing a connected smart scale with user permission, which transmitted the weight automatically each time users weighed themselves. All datasets were deidentified prior to analysis. The research team had access only to pseudonymized or anonymized records containing no direct identifiers, and no attempt was made to re-identify individuals.

Outcomes

The primary outcome was weight loss in kg and as a percentage of baseline body weight at 3 and 6 months. Secondary outcomes included the proportion of users achieving any weight loss as well as weight loss thresholds of 3%, 5%, and 10% of baseline body weight at each time point.

Statistical Analysis

Weight Outcomes at 3 and 6 Months

To estimate average weight change at 3 and 6 months compared with baseline, we fit piecewise linear mixed-effects models. Time was represented with piecewise linear terms with break points at 3 and 6 months, allowing for different rates of weight change over time. We chose these time points based on the hypothesis that dietary behavior changes would be most intensive initially (0 to <3 months), continue at a slower rate (3 to <6 months), then stabilize (≥6 months). The model included random effects for the intercept and the overall time slope to account for between-user variation in baseline weight and weight change over time. Weight loss at 3 and 6 months was estimated by computing linear combinations of the fixed effects from the piecewise linear terms. We stratified the analysis by sex, age group, and baseline BMI category.

Weight Loss Thresholds

For the proportion of active users who achieved weight loss thresholds at each time point, we identified the weight measurement closest to each respective time point and compared it with the user's baseline weight. Users with no measurements within 30 days of a given time point were excluded from the primary threshold analysis for that time point. All users recorded a weight at or beyond 3 months, so exclusions at the 3-month

XSL•FO
RenderX

time point reflect measurement timing rather than absent outcome data. At 6 months, approximately half the cohort had no recorded weight at or beyond that point. We therefore report the 6-month thresholds as a range, with the measured-only proportion as the upper bound and a lower bound assuming that every user without a measurement near the time point did not reach the threshold, using the full analyzed cohort as the denominator. Additionally, as a sensitivity analysis, we repeated the threshold analysis using the last nonbaseline weight observation carried forward for excluded users. Users with no eligible prior measurement remained excluded.

Weight Outcomes by Engagement

To quantify the difference in weight change by engagement status, we first created a matched sample for each engagement metric. We used coarsened exact matching (CEM) [33] to balance groups of more engaged users with less engaged users on the following criteria: age group (18-29, 30-49, or 50-65 years), sex, country of residence, and BMI at first appointment (25 - 29.9, 30 - 39.9, or ≥40). Then we fit weighted piecewise linear mixed-effects models to the matched datasets. The model included an interaction term between engagement status and time for each of the 3 time intervals (0 to <3, 3 to <6, and ≥6 months). Because all included users had measurements until at least the 3-month time point, we estimated differential rates of weight change between more and less engaged users at this time point only. The models were weighted by applying the CEM weights to account for any residual imbalance between engagement groups after matching. Separate models were fit for each included engagement metric.

All analyses were conducted independently by the lead author. A coauthor reviewed the data processing pipeline against the inclusion and exclusion criteria for data quality issues. All analyses were completed in R (version 4.4.0; R Foundation for Statistical Computing) [34].

Ethical Considerations

The study received an exempt determination from the Syracuse University Institutional Review Board (protocol number 26 - 018). During account registration, users must review and consent to Nutrium's privacy policy before creating an account or providing any personal data. The policy explained how personal data may be processed in connection with platform use and informed users that data may be deidentified and aggregated for research and analytical purposes. For this study, only deidentified data were used, and no directly identifiable personal data were accessed by the research team. Users did not receive any compensation for participation in this study.

Results

Study Sample

The population considered for this study included 27,689 adults with overweight or obesity using Nutrium Care (Figure 2). The unrestricted sample, after excluding users with no baseline weight measurement (2745/27,689, 9.9%), reported GLP-1 agonist use (1296/24,944, 5.2%), reported pregnancy (611/23,648, 2.6%), or missing data on age or sex (13/23,037, 0.1%), included 23,024 users. The sample analyzed for the primary action phase population included 44.7% (10,288/23,024) of eligible users from the unrestricted population, after excluding 3647 (15.8%) users who were not in the action phase (<2 appointments in the first 6 months), 1544 (6.7%) users who joined within 3 months of the end of the observation window and had no opportunity to accrue 3 months of follow-up, and 7545 (32.8%) users who had fewer than 3 months between their first and last recorded data points.

XSL•FO
RenderX

Figure 2. Study population.

XSL•FO
RenderX

Two-thirds of included users were female (6887/10,288, 66.9%), and most were aged 18 to 29 years (4584/10,288, 44.6%) and 30 to 49 years (5157/10,288, 50.1%; Table 1). The study included 61.2% (6293/10,288) of users with overweight, 34.3% (3531/10,288) of users with obesity, and 4.5% (464/10,288) of users with morbid obesity at baseline. Most users in this sample joined the platform in 2024 or 2025, contributing a median of 6 (IQR 4 - 9) months of data and 5 weight measurements to this study (Table 1). Follow-up duration was the number of months between a user's first and last recorded weight measurements. During the first 6 months, users had a median of 3 (IQR 2-4) dietitian appointments and 25 (IQR 14-44) days of platform app use and logged food, water intake, and exercise a median of 6 (IQR 2-25), 3 (IQR 1-14), and 3 (IQR 0-21) times, respectively (Table 1).

Table . Study population sociodemographic characteristics, data contributed, and engagement metrics in the first 6 months (N=10,288).

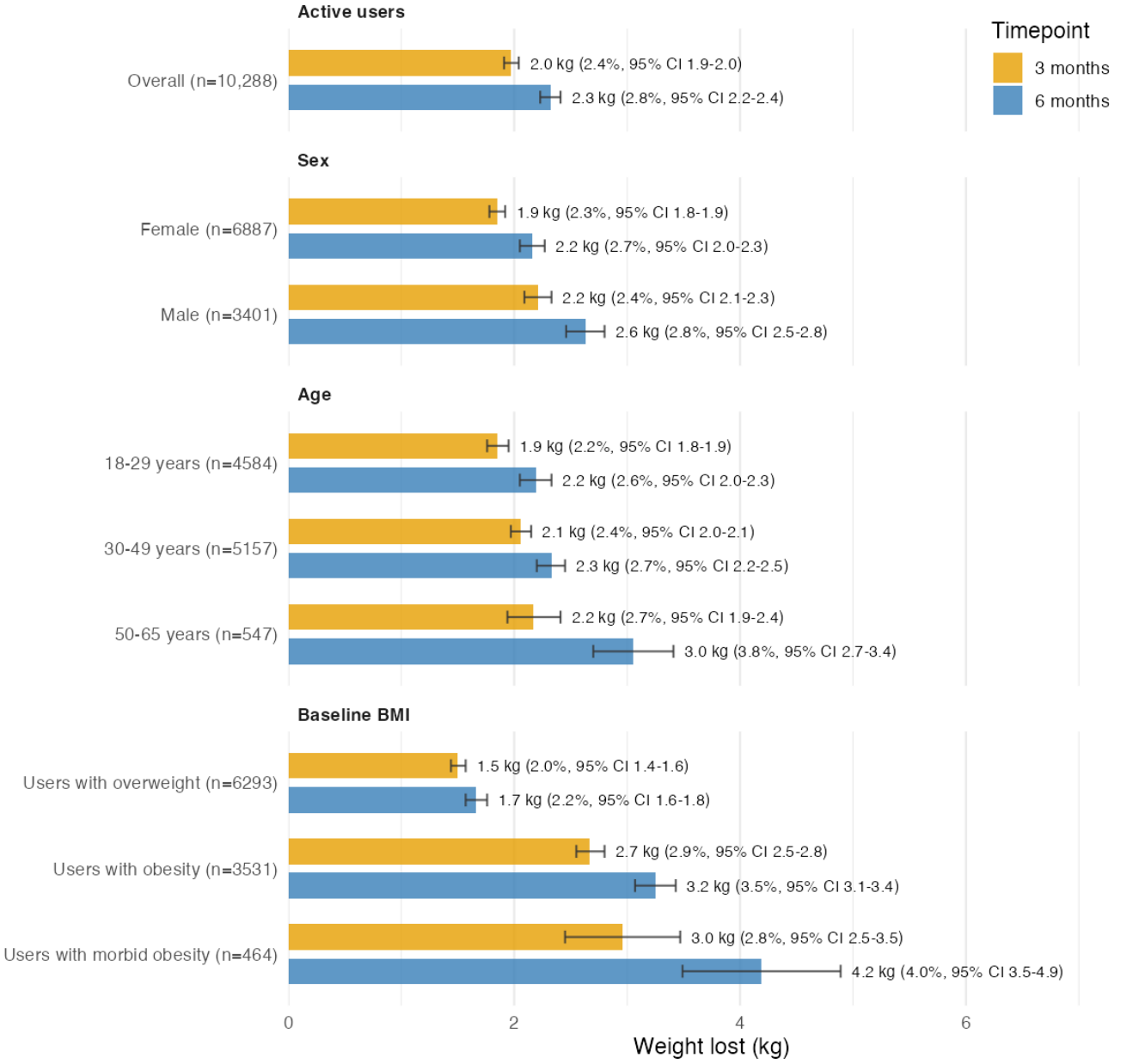
	Values
Sociodemographic characteristics	
Sex, n (%)	
Female	6887 (66.9)
Male	3401 (33.1)
Age (y), n (%)	
18 - 29	4584 (44.6)
30 - 49	5157 (50.1)
50 - 65	547 (5.3)
BMI category, n (%)	
Overweight	6293 (61.2)
Obesity	3531 (34.3)
Morbid obesity	464 (4.5)
Data contributed	
Platform registration year, n (%)	
<2022	8 (0.1)
2022	175 (1.7)
2023	293 (2.8)
2024	4070 (39.6)
2025	5742 (55.8)
Months in the cohort, median (IQR)	6 (4-9)
Users with a minimum of 6 months of follow-up data, n (%)	5343 (51.9)
Weight measurements contributed, median (IQR)	5 (3-7)
Engagement in the first 6 months, median (IQR)	
Number of appointments	3 (2-4)
Days of active use	25 (14 - 44)
Food diary logs	6 (2-25)
Water intake logs	3 (1-14)
Exercise logs	3 (0 - 21)

Weight Outcomes at 3 and 6 Months

Three months after joining Nutrium Care, active users lost 2.0 kg on average (2.4% of their body weight; 95% CI 1.9 - 2.0 kg; Figure 3). By 6 months, average weight loss increased to 2.3 kg (2.8%; 95% CI 2.2 - 2.4 kg). Weight loss was slightly more pronounced among male users and among users aged 50 or older, although this sample of users in this age group was small (n=547, 5.3%), and varied substantially by baseline BMI at both 3 and 6 months. At 6 months, users with overweight lost 1.7 kg on average (2.2%; 95% CI 1.6 - 1.8 kg), whereas users with obesity averaged 3.2 kg of weight loss (3.5%; 95% CI 3.1 - 3.4 kg) and users with morbid obesity averaged 4.2 kg (4%; 95% CI 3.5 - 4.9 kg). In the unrestricted sample, mean weight loss was 2.2 kg (2.6%; 95% CI 2.2 - 2.3 kg) at 3 months and 2.4 kg (2.9%; 95% CI 2.4 - 2.5 kg) at 6 months (Figure S1 in Multimedia Appendix 1).

XSL•FO
RenderX

Figure 3. Absolute weight loss at 3 and 6 months among active users (*P*<.001 for all estimates). A total of 22,043 weight measurements beyond baseline were recorded in the 0-3 month period and 18,641 in the 3-6 month period, contributed by 7941 and 8727 users, respectively.

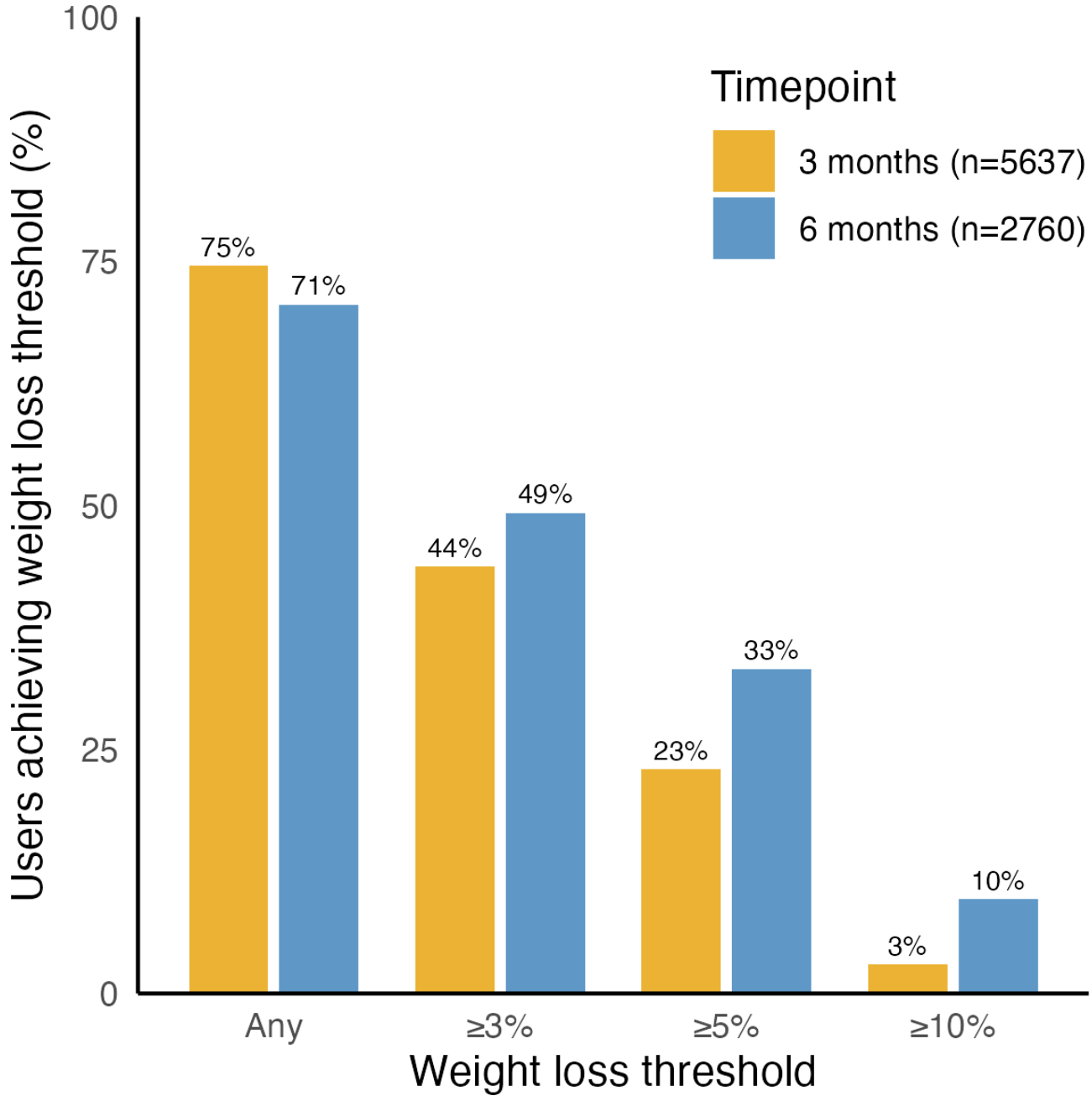


Weight Loss Thresholds

At 3 months, weight was available within 30 days of the time point for 54.8% (5637/10,288) of active users. Among these, most (4200/5637, 74.5%) had lower weight compared with baseline (Figure 4). Nearly half of these active users (2464/5637, 43.7%) achieved at least 3% weight loss, nearly one quarter (1293/5637, 23%) achieved at least 5%, and a smaller proportion (166/5637, 2.9%) at least 10% weight loss at this time point. All users in the action phase cohort recorded a weight at or beyond 3 months, so users excluded at this time point had recorded a weight outside the 30-day window rather than having stopped weighing in.

XSL•FO
RenderX

Figure 4. Threshold weight outcomes at 3 and 6 months among active users. Users without a weight measurement within 30 days of the time point were excluded.



At 6 months, weight was available within 30 days of the time point for 26.8% (2760/10,288) of active users. Among these, a similar proportion of active users (1947/2760, 70.5%) weighed below their baseline weight, almost half (1358/2760, 49.2%) achieved at least 3% weight loss, one third (917/2760, 33.2%) achieved at least 5%, and one in 10 (266/2760, 10%) achieved at least 10% weight loss. Because 52% (5343/10,288) of active users accrued 6 months of follow-up, these proportions are restricted to active users who continued recording weights and are likely to be optimistic. Assuming instead that every active user without a weight near the time point did not reach the threshold gives lower bounds of 18.9% (1947/10,288) for any weight loss, 13.2% (1358/10,288) for 3% weight loss, 8.9% (917/10,288) for 5% weight loss, and 2.6% (266/10,288) for 10% weight loss. In the unrestricted sample, 48.5% (1377/2840) of users lost at least 3% of baseline weight at 6 months, 33% (929/2840) lost at least 5% and 9.4% (268/2840) lost at least 10% (Figure S2 in Multimedia Appendix 1). A sensitivity analysis that carried the last observed weight measurement forward among active users gave intermediate estimates between the primary estimates and lower-bound estimates presented here (Figure S3 in Multimedia Appendix 1).

Weight Outcomes by Engagement

Active users who met the engagement criteria had greater weight loss at 3 months compared with matched users who did not meet these criteria (Table 2). Among the engagement metrics considered, the largest differences were observed for users with more dietitian appointments. Users with at least 1 appointment every 2 months (bimonthly appointments) experienced 0.8 kg (95% CI 0.7 - 1.0) greater weight loss at 3 months compared with matched users without bimonthly appointments.

XSL•FO
RenderX

Table . Difference in weight loss at 3 months by engagement metric (N=10,288).

Engagement metric	Participants, n (%)	Weight loss difference (kg; 95% CI)[a]
Bimonthly appointment	7392 (71.9)	0.8 (0.7 - 1.0)
≥10 days of active use per month	2344 (22.8)	0.7 (0.6 - 0.9)
Weekly food logging	3126 (30.4)	0.4 (0.3 - 0.6)
Weekly water intake logging	2207 (21.5)	0.4 (0.2 - 0.5)
Weekly exercise logging	2799 (27.2)	0.3 (0.2 - 0.5)

[a]All estimates, $P<.001$.

Users with at least 10 DAU per month experienced 0.7 kg (95% CI 0.6 - 0.9) more weight loss. Logging behaviors showed less pronounced associations with weight loss at 3 months. Weekly food diary loggers experienced 0.4 kg (95% CI 0.3 - 0.6) additional weight loss, weekly water intake loggers experienced 0.4 kg (95% CI 0.2 - 0.5), and weekly physical exercise loggers experienced 0.3 kg (95% CI 0.2 - 0.5) additional weight loss compared with matched users who logged less frequently. Differences in weight loss by engagement metric were moderately larger in the cohort before the action phase restriction, with similar ordering across metrics (Table S1 in Multimedia Appendix 1).

Discussion

This study evaluated real-world longitudinal data among more than 10,000 users with overweight and obesity enrolled in the Nutrium Care program who progressed to the action phase, reflecting active commitment to achieving their weight goals. Active users lost, on average, 2.0 kg and 2.3 kg after 3 and 6 months, respectively. Among active users with a weight recorded near the 3-month time point, 74.5% (4200/5637) weighed less than at baseline, 43.7% (2464/5637) lost at least 3% of their baseline weight, 22.9% (1293/5637) lost at least 5% of their baseline weight, and 2.9% (166/5637) lost at least 10% of their baseline weight. Greater program engagement was associated with more weight loss across the included engagement metrics, with bimonthly dietitian appointments showing the strongest association with improved weight loss outcomes.

The findings reported in this study are consistent with the growing evidence supporting digital nutrition platforms [35]. The weight loss observed is comparable to outcomes reported for digital interventions and is consistent with losses reported in meta-analyses of smartphone-based apps [36]. Among active users with a weight recorded near the 3-month time point, 23% (1293/5637) reached at least 5% weight loss. The 5% threshold is the benchmark most commonly used to define clinically significant weight loss [37]. Some evidence indicates benefits for cardiovascular risk factors at more modest losses [38], although this is not the established standard. While current guidelines indicate that individuals with obesity will see the biggest health benefits when weight loss exceeds 10% of body weight, most experts agree that these amounts are more likely to result from surgical or pharmaceutical interventions rather than lifestyle changes [39-41].

Behavior change and lifestyle modification have been identified as integral components of successful weight management, including external support from health care providers such as dietitians [42]. Incorporation of techniques that encourage self-regulation, such as tracking hunger levels, mood, and other common cues associated with eating habits, can support increased awareness of an individual's environmental triggers that either facilitate or derail weight loss attempts [43]. However, simple tracking activities such as recording weight have been shown to be effective, particularly when paired with goal setting within a digital platform [44,45]. In this context, the engagement metrics assessed in the present study may reflect not only platform use but also behaviors associated with sustained participation in weight management.

The association between engagement and weight loss further contextualizes these results within the broader literature [46,47]. Active users attending bimonthly dietitian appointments lost 0.8 kg more than matched users without this level of contact, suggesting a relationship between professional interaction and weight loss [48,49]. These findings are consistent with evidence suggesting that digital interventions that include a component of actual human contact are more effective than fully automated programs [35] and that nutrition coaching, including medical nutrition therapy, delivered via telehealth can be as effective as traditional clinical nutrition care [11,48]. Within the TTM framework, behavioral processes of change such as helping relationships and reinforcement management are most prominent during the action phase, so forms of professional support that provide reinforcement may be particularly relevant for users working to sustain new routines [29]. Dietitian contact showed a stronger association with weight loss than self-directed platform engagement, although the design cannot establish why. However, in the absence of established benchmarks, the thresholds applied to engagement metrics were exploratory, so the estimated differences between more and less engaged users should be limited to hypothesis-generating interpretation. Further, statistical significance does not indicate clinical importance. The differences observed here, ranging from 0.3 kg to 0.8 kg at 3 months, may not be clinically meaningful for individual users.

The findings from this study should be interpreted considering its limitations. First, we relied on secondary observational data with no control group or comparison with other digital nutrition platforms. Therefore, we cannot establish causality in the weight loss results documented here, nor can we directly compare with the effectiveness of other weight loss solutions. Second, both criteria defining the action phase cohort depend on behavior

XSL•FO
RenderX

after baseline. Users met the action phase definition only by attending a second dietitian appointment within 6 months, and follow-up was measured from a user's first to last weight measurement, excluding those who stopped generating data before 3 months. Together, these criteria excluded 48.6% (11,192/23,024) of eligible users, compared with 6.7% (1544/23,024) excluded for joining too late to accrue 3 months of data. The cohort therefore represents users who remained active, and estimates should not be extrapolated to all enrolled users or to employer-sponsored digital nutrition programs more broadly. The action phase criterion is also conceptually adjacent to our strongest engagement exposure, since both are defined by continued dietitian attendance, so the engagement comparison is made within a group already selected on a related behavior. Estimates in the cohort without the action phase restriction were not smaller than our primary estimates, which offers no indication that the restriction inflated our findings, although that cohort includes many users whose values at 3 and 6 months are extrapolated from the fitted model rather than observed, so it cannot serve as a reliable counterfactual. Third, weight measurements were not recorded on a fixed schedule, so the threshold analyses depend on data availability near each time point. At 3 months, this reflects measurement timing rather than absent data, since all users recorded a weight at or beyond 3 months and those excluded had weighed in outside the 30-day window. At 6 months, approximately half the cohort had no recorded weight at or beyond that point. Users who stall or regain weight are more likely to stop recording weights, so the 6-month proportions based on measured weights alone are likely to overstate achievement of each threshold. We therefore report the lower bound of these proportions. Fourth, we were restricted in the length of follow-up. We addressed this by examining 2 time points, 3 and 6 months. However, further evidence on long-term weight loss maintenance, including after discontinuation of platform use, is needed. Fifth, this study was limited to weight loss as an outcome and does not include data on user experiences during the care they received and its impact on well-being. Sixth, the engagement analysis could not control for all possible confounders and potential reverse causality. For example, we did not have data on comorbidities experienced by users. Further, users experiencing difficulty with weight loss may be more engaged with the program, potentially biasing our estimate toward the null. Finally, we relied on data entered both by dietitians and self-reported by users, which may be subject to measurement error and reporting bias.

Our findings support the use of digital nutrition platforms as a meaningful tool to aid in weight management among users motivated and in the action phase of behavior change. Given the reality that primary care providers have limited time to incorporate behavioral change techniques during routine appointments, encouraging users with overweight or obesity to engage with a dietitian via a digital platform will increase access to the services and support needed. Having virtual access to a dietitian increases equity in care and reduces the logistical burden users experience associated with transportation and scheduling challenges that often occur with in-person appointments. Indeed, nutrition interventions using telehealth modalities are more cost-effective overall, with digital (mobile health) rated as the most cost-effective [50]. As employers (and health insurance companies) expand health benefits to include nutrition services from a dietitian, offering a digital nutrition platform provides a scalable yet evidence-based solution that can ultimately drive down medical costs as employees achieve weight management goals.

Nutrium Care's observational findings provide motivation for pragmatic randomized evaluations to more directly test causal effects and durability of outcomes beyond 12 months. This need is increasingly recognized within the digital nutrition field, where planned randomized trials have begun to examine longer-term effectiveness and maintenance of digital programs [51]. Further work is also needed to characterize platform use patterns associated with optimal outcomes, for example, by identifying combinations of engagement behaviors that predict clinically meaningful weight loss and maintenance. Finally, given the scale of obesity-related costs, formal economic evaluation is an important next step. Modeling should consider (1) health care use offsets, (2) productivity impacts, (3) program delivery costs (including clinician time), and (4) subgroup targeting where marginal benefit is highest (eg, higher baseline BMI categories).

Acknowledgments

Claude (Anthropic) was used to assist with language and grammar editing, literature searches during manuscript preparation, and code optimization. All AI-assisted content was reviewed and verified by the authors.

Funding

This study was funded by Healthium – Healthcare Software Solutions, SA.

Data Availability

The datasets analyzed in this study are not publicly available due to privacy and contractual constraints. Deidentified data may be made available upon reasonable request, subject to applicable data protection and governance approvals.

Authors' Contributions

FGBG contributed to the conceptualization of the study, method development, data curation, and led the analysis and writing of the original draft manuscript. JG contributed to the writing of the original draft and investigation. FTN contributed to data curation,

XSL•FO
RenderX

software, and validation. AL contributed to writing of the original draft. LZ contributed to conceptualization of the study, method development, and investigation. MA contributed to conceptualization of the study, investigation, project administration, and supervision. All authors reviewed and approved the final draft of the manuscript.

Conflicts of Interest

AL, FTN, and MA are employed by Healthium – Healthcare Software Solutions, SA and hold equity interests in the company. FGBG, JG, and LZ receive consulting fees from Healthium – Healthcare Software Solutions, SA but are not otherwise employed by or affiliated with the sponsor.

Multimedia Appendix 1
Findings from sensitivity analyses.
[DOCX File, 415 KB - mhealth_v14i1e94732_app1.docx]

References

1. James WP. Obesity: a global public health challenge. Clin Chem 2018 Jan;64(1):24-29. [doi: 10.1373/clinchem.2017.273052] [Medline: 29295834]
2. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in underweight and obesity from 1990 to 2022: a pooled analysis of 3663 population-representative studies with 222 million children, adolescents, and adults. Lancet 2024 Mar;403(10431):1027-1050. [doi: 10.1016/S0140-6736(23)02750-2] [Medline: 38432237]
3. Burra P, Verduci E, Dias JA, et al. The growing burden of obesity: addressing a global public health challenge. United European Gastroenterol J 2025 Sep;13(7):1343-1347. [doi: 10.1002/ueg2.70020] [Medline: 40202908]
4. GBD 2015 Obesity Collaborators. Health Effects of Overweight and Obesity in 195 Countries over 25 Years. N Engl J Med 2017 Jul 6;377(1):13-27. [doi: 10.1056/NEJMoa1614362] [Medline: 28604169]
5. Obesity Atlas 2023. : World Obesity Federation; 2023 URL: https://data.worldobesity.org/publications/?cat=19 [accessed 2026-08-26]
6. Okunogbe A, Nugent R, Spencer G, Powis J, Ralston J, Wilding J. Economic impacts of overweight and obesity: current and future estimates for 161 countries. BMJ Glob Health 2022 Sep;7(9):e009773. [doi: 10.1136/bmjgh-2022-009773] [Medline: 36130777]
7. Kushner RF. Weight loss strategies for treatment of obesity. Prog Cardiovasc Dis 2014;56(4):465-472. [doi: 10.1016/j.pcad.2013.09.005] [Medline: 24438739]
8. Li Z, Han Z, Sun R, Xuan X, Huang C. Long-term efficacy trajectories of GLP-1 receptor agonists: a systematic review and network meta-analysis. Diabetes Metab Syndr Obes 2025;18:3611-3624. [doi: 10.2147/DMSO.S539822] [Medline: 41019499]
9. Gatto A, Liu K, Milan N, Wong S. The effects of GLP-1 agonists on musculoskeletal health and orthopedic care. Curr Rev Musculoskelet Med 2025 Oct;18(10):469-480. [doi: 10.1007/s12178-025-09978-3] [Medline: 40372699]
10. Morgan-Bathke M, Baxter SD, Halliday TM, et al. Weight management interventions provided by a dietitian for adults with overweight or obesity: an evidence analysis center systematic review and meta-analysis. J Acad Nutr Diet 2023 Nov;123(11):1621-1661.e25. [doi: 10.1016/j.jand.2022.03.014] [Medline: 35788061]
11. Raynor HA, Morgan-Bathke M, Baxter SD, et al. Position of the Academy of Nutrition and Dietetics: medical nutrition therapy behavioral interventions provided by dietitians for adults with overweight or obesity, 2024. J Acad Nutr Diet 2024 Mar;124(3):408-415. [doi: 10.1016/j.jand.2023.11.013] [Medline: 38040115]
12. Curtin E, Green R, Brown KA, et al. The effectiveness of mobile app-based interventions in facilitating behaviour change towards healthier and more sustainable diets: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2025 Sep 30;22(1):122. [doi: 10.1186/s12966-025-01823-7] [Medline: 41035009]
13. Tang Y, Guo X, Liu X, Qiu J, Yang J. Impact of mHealth-based lifestyle interventions on weight loss in overweight/obese adults: a systematic meta-analysis. J Gen Intern Med 2026 Jan;41(1):206-216. [doi: 10.1007/s11606-025-09841-8] [Medline: 40986219]
14. Couto FD, de Almeida CP. Mobile and web apps for weight management in overweight and obese adults: an updated umbrella review and meta-analysis. Int J Environ Res Public Health 2025 Jul 21;22(7):1152. [doi: 10.3390/ijerph22071152] [Medline: 40724217]
15. Salwen-Deremer JK, Khan AS, Martin SS, Holloway BM, Coughlin JW. Incorporating health behavior theory into mHealth: an examination of weight loss, dietary, and physical activity interventions. J Technol Behav Sci 2020;5(1):51-60. [doi: 10.1007/s41347-019-00118-6]
16. Pala D, Petrini G, Bosoni P, Larizza C, Quaglini S, Lanzola G. Smartphone applications for nutrition support: a systematic review of the target outcomes and main functionalities. Int J Med Inform 2024 Apr;184:105351. [doi: 10.1016/j.ijmedinf.2024.105351] [Medline: 38295584]
17. König LM, Attig C, Franke T, Renner B. Barriers to and facilitators for using nutrition apps: systematic review and conceptual framework. JMIR Mhealth Uhealth 2021 Apr 1;9(6):e20037. [doi: 10.2196/20037] [Medline: 34254938]

XSL•FO
RenderX

18. Patel ML, Wakayama LN, Bennett GG. Self-monitoring via digital health in weight loss interventions: a systematic review among adults with overweight or obesity. Obesity (Silver Spring) 2021 Mar;29(3):478-499. [doi: 10.1002/oby.23088] [Medline: 33624440]
19. Payne JE, Turk MT, Kalarchian MA, Pellegrini CA. Adherence to mobile-app-based dietary self-monitoring-impact on weight loss in adults. Obes Sci Pract 2021;8(3):279-288. [doi: 10.1002/osp4.566] [Medline: 35664248]
20. Vanwinkelen K, Spruyt B, Smits T. Digital interventions targeting healthy and sustainable eating behavior: systematic review and meta-analysis. J Med Internet Res 2026 Jan 8;28:e80821. [doi: 10.2196/80821] [Medline: 41505650]
21. Hayashi D, Edwards C, Emond JA, et al. What is food noise? A conceptual model of food cue reactivity. Nutrients 2023 Nov 17;15(22):4809. [doi: 10.3390/nu15224809] [Medline: 38004203]
22. Dhurandhar EJ, Maki KC, Dhurandhar NV, et al. Food noise: definition, measurement, and future research directions. Nutr Diabetes 2025 Jul 8;15(1):30. [doi: 10.1038/s41387-025-00382-x] [Medline: 40628707]
23. Chen Y, Perez-Cueto FJ, Giboreau A, Mavridis I, Hartwell H. The promotion of eating behaviour change through digital interventions. Int J Environ Res Public Health 2020;17(20):7488. [doi: 10.3390/ijerph17207488] [Medline: 33076239]
24. Berry R, Kassavou A, Sutton S. Does self-monitoring diet and physical activity behaviors using digital technology support adults with obesity or overweight to lose weight? A systematic literature review with meta-analysis. Obes Rev 2021 Oct;22(10):e13306. [doi: 10.1111/obr.13306] [Medline: 34192411]
25. Lau Y, Wong SH, Chee DG, et al. Technology-delivered personalized nutrition intervention on dietary outcomes among adults with overweight and obesity: a systematic review, meta-analysis, and meta-regression. Obes Rev 2024 May;25(5):e13699. [doi: 10.1111/obr.13699] [Medline: 38296771]
26. Beleigoli A, Andrade AQ, Diniz MD, Ribeiro AL. Personalized web-based weight loss behavior change program with and without dietitian online coaching for adults with overweight and obesity: randomized controlled trial. J Med Internet Res 2020 Nov 5;22(11):e17494. [doi: 10.2196/17494] [Medline: 33151151]
27. Xue J, Zhang B, Zhao Y, et al. Evaluation of the current state of chatbots for digital health: scoping review. J Med Internet Res 2023 Dec 19;25:e47217. [doi: 10.2196/47217] [Medline: 38113097]
28. Chen J, Gemming L, Hanning R, Allman-Farinelli M. Smartphone apps and the nutrition care process: current perspectives and future considerations. Patient Educ Couns 2018 Apr;101(4):750-757. [doi: 10.1016/j.pec.2017.11.011] [Medline: 29169863]
29. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
30. de Freitas PP, de Menezes MC, dos Santos LC, Pimenta AM, Ferreira AV, Lopes AC. The transtheoretical model is an effective weight management intervention: a randomized controlled trial. BMC Public Health 2020 May 11;20(1):652. [doi: 10.1186/s12889-020-08796-1] [Medline: 32393214]
31. Chamarthi VS, Daley SF. Comprehensive behavioral modification and counseling strategies for obesity management. In: StatPearls: StatPearls Publishing; 2025. [Medline: 40991765]
32. Global BMI Mortality Collaboration, Di Angelantonio E, Bhupathiraju S, Wormser D, et al. Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. Lancet 2016 Aug;388(10046):776-786. [doi: 10.1016/S0140-6736(16)30175-1] [Medline: 27423262]
33. Iacus SM, King G, Porro G. Causal inference without balance checking: coarsened exact matching. Political Anal 2012;20(1):1-24. [doi: 10.1093/pan/mpr013]
34. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2024. URL: https://www.R-project.org [accessed 2026-08-26]
35. Kupila SK, Joki A, Suojanen LU, Pietiläinen KH. The effectiveness of eHealth interventions for weight loss and weight loss maintenance in adults with overweight or obesity: a systematic review of systematic reviews. Curr Obes Rep 2023 Sep;12(3):371-394. [doi: 10.1007/s13679-023-00515-2] [Medline: 37354334]
36. Antoun J, Itani H, Alarab N, Elsehmawy A. The effectiveness of combining nonmobile interventions with the use of smartphone apps with various features for weight loss: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022 Apr 8;10(4):e35479. [doi: 10.2196/35479] [Medline: 35394443]
37. Horn DB, Almandoz JP, Look M. What is clinically relevant weight loss for your patients and how can it be achieved? A narrative review. Postgrad Med 2022 May;134(4):359-375. [doi: 10.1080/00325481.2022.2051366] [Medline: 35315311]
38. Zomer E, Gurusamy K, Leach R, et al. Interventions that cause weight loss and the impact on cardiovascular risk factors: a systematic review and meta-analysis. Obes Rev 2016 Oct;17(10):1001-1011. [doi: 10.1111/obr.12433] [Medline: 27324830]
39. Markovic TP, Proietto J, Dixon JB, et al. The Australian obesity management algorithm: a simple tool to guide the management of obesity in primary care. Obes Res Clin Pract 2022;16(5):353-363. [doi: 10.1016/j.orcp.2022.08.003] [Medline: 36050266]
40. Elmaleh-Sachs A, Schwartz JL, Bramante CT, Nicklas JM, Gudzune KA, Jay M. Obesity management in adults: a review. JAMA 2023 Nov 28;330(20):2000-2015. [doi: 10.1001/jama.2023.19897] [Medline: 38015216]
41. Powell-Wiley TM, Poirier P, Burke LE, et al. Obesity and cardiovascular disease: a scientific statement from the American Heart Association. Circulation 2021 May 25;143(21):e984-e1010. [doi: 10.1161/CIR.0000000000000973] [Medline: 33882682]

XSL•FO
RenderX

42. Ryan DH, Kahan S. Guideline recommendations for obesity management. Med Clin North Am 2018 Jan;102(1):49-63. [doi: 10.1016/j.mcna.2017.08.006] [Medline: 29156187]
43. Chew HS, Gao Y, Shabbir A, et al. Personal motivation, self-regulation barriers and strategies for weight loss in people with overweight and obesity: a thematic framework analysis. Public Health Nutr 2022 Sep;25(9):2426-2435. [doi: 10.1017/S136898002200043X] [Medline: 35190011]
44. Frie K, Hartmann-Boyce J, Jebb SA, Aveyard P. Effectiveness of a self-regulation intervention for weight loss: a randomized controlled trial. Br J Health Psychol 2020 Sep;25(3):652-676. [doi: 10.1111/bjhp.12436] [Medline: 32489005]
45. Kooiman TJ, Dijkstra A, Kooy A, Dotinga A, van der Schans CP, de Groot M. The role of self-regulation in the effect of self-tracking of physical activity and weight on BMI. J Technol Behav Sci 2020;5:206-214. [doi: 10.1007/s41347-020-00127-w]
46. Johnson F, Wardle J. The association between weight loss and engagement with a web-based food and exercise diary in a commercial weight loss programme: a retrospective analysis. Int J Behav Nutr Phys Act 2011 Aug 2;8:83. [doi: 10.1186/1479-5868-8-83] [Medline: 21810222]
47. Lin PH, Grambow S, Intille S, et al. The association between engagement and weight loss through personal coaching and cell phone interventions in young adults: randomized controlled trial. JMIR Mhealth Uhealth 2018 Oct 18;6(10):e10471. [doi: 10.2196/10471] [Medline: 30341051]
48. Hu EA, Kelley T, Haryani A. The impact of technology-enabled medical nutrition therapy on weight loss in adults with overweight and obesity: retrospective observational study. JMIR Mhealth Uhealth 2025 May 6;13:e70228. [doi: 10.2196/70228] [Medline: 40327381]
49. Lehmann M, Jones L, Schirmann F. App engagement as a predictor of weight loss in blended-care interventions: retrospective observational study using large-scale real-world data. J Med Internet Res 2024 Jun 7;26:e45469. [doi: 10.2196/45469] [Medline: 38848556]
50. Kelly JT, Law L, De Guzman KR, et al. Cost-effectiveness of telehealth-delivered nutrition interventions: a systematic review of randomized controlled trials. Nutr Rev 2023 Nov 10;81(12):1599-1611. [doi: 10.1093/nutrit/nuad032] [Medline: 37016937]
51. Sysko R, Bibeau J, Boyar A, et al. A 2.5-year weight management program using Noom Health: protocol for a randomized controlled trial. JMIR Res Protoc 2022 Aug 12;11(8):e37541. [doi: 10.2196/37541] [Medline: 35969439]

Abbreviations

CEM: coarsened exact matching
DAU: days of active use
GLP-1: glucagon-like peptide-1
RD: registered dietitian
TTM: transtheoretical model

Edited by A Stone; submitted 25.Mar.2026; peer-reviewed by M Jospe, R Bipat; revised version received 17.Aug.2026; accepted 17.Aug.2026; published 15.Sep.2026.

Please cite as:
Goddard FGB, Garay J, Teixeira Neves F, Ligeti A, Zhaunova L, Abreu M
Weight Loss and Engagement Among Active Users of a Dietitian-Led Digital Nutrition Platform: Retrospective Cohort Study
JMIR Mhealth Uhealth 2026;14:e94732
URL: https://mhealth.jmir.org/2026/1/e94732
doi:10.2196/94732

© Frederick G B Goddard, Jessica Garay, Francisco Teixeira Neves, Alexandra Ligeti, Liudmila Zhaunova, Manuela Abreu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 15.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Feasibility and Preliminary Efficacy of a “Sit Less” Program Leveraging Fitbit Tracking and Tailored Text Messages in Cardiometabolic Disease: Findings From 2 Parallel Randomized Controlled Trials in Coronary Artery Disease and Type 2 Diabetes

Chorong Park[1,2,3], PhD; Mary S Dietrich[4,5], PhD; Britta Larsen[6], PhD; Lindsay S Mayberry[7], PhD; Abigail Doyle[4,8], MSHI; Soojung Ahn[9], PhD; Jason Jean[4], DNP; Kevin Maquiling[10], MD; Mulubrhan F Mogos[4], PhD; James M Muchira[4], PhD; Shelagh Mulvaney[4,11,12], PhD

[1]College of Nursing, Seoul National University, 103 Daehak-ro, Jongno-gu, Main Nursing Building #501, Seoul, Republic of Korea

[10]Vanderbilt University Medical Center, Nashville, TN, United States

[11]Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, United States

[12]College of Connected Computing, Vanderbilt University, Nashville, TN, United States

[2]The Research Institute of Nursing Science, Seoul National University, Seoul, Seoul, Republic of Korea

[3]School of Transdisciplinary Innovations, Seoul National University, Seoul, Republic of Korea

[4]School of Nursing, Vanderbilt University, Nashville, TN, United States

[5]Department of Biostatistics, School of Medicine, Vanderbilt University, Nashville, TN, United States

[6]Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, La Jolla, CA, United States

[7]Division of General Internal Medicine and Public Health, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, United States

[8]Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University School of Medicine, Nashville, TN, United States

[9]William F. Connell School of Nursing, Boston College, Chestnut Hill, MA, United States

Corresponding Author:
Chorong Park, PhD
College of Nursing, Seoul National University, 103 Daehak-ro, Jongno-gu, Main Nursing Building #501, Seoul, Republic of Korea

Abstract

Background: Individuals with cardiometabolic disease typically average 10 to 14 hours of daily sedentary time, which increases the risk of cardiovascular disease.

Objective: This study aimed to evaluate the feasibility, acceptability, and preliminary efficacy of a “Sit Less” program in reducing sedentary behavior in people with cardiometabolic disease.

Methods: Participants with cardiovascular disease or type 2 diabetes underwent separate randomization into the Sit Less intervention or the control group. Sit Less included 1 in-person counseling session, a Fitbit, a smart water bottle, and tailored weekly text messages for 12 weeks. Control group participants received the American Heart Association’s “Answers by Heart” booklet. Sedentary behavior was measured using the activPAL for 7 days at baseline and postintervention. The secondary outcomes included physical activity, cardiometabolic biomarkers, continuous glucose monitoring metrics, and psychological factors. Generalized linear models estimated adjusted between-group differences, controlling for baseline values, cohort, cohort×group interaction, and leisure-time exercise.

Results: Of the 37 randomized participants, 35 (95%) completed the study (Sit Less: n=17, 48.6%; and control: n=18, 51.4%). Of the 35 participants who completed the study, 62% (n=21) were men, the median age was 67 (54-72) years, and their baseline sedentary time was approximately 10.4 hours per day. Sit Less participants demonstrated high adherence, including a text response rate of 81% and a median Fitbit wear time of 15 hours per day on 6.8 days per week, and 15 (88%) of 17 participants reported satisfaction with the program. Compared with the control group, Sit Less participants showed greater reductions in total sedentary time (adjusted mean difference −17.3 min/d, 95% CI −63 to 28; *P*=.45) and prolonged sedentary time of >30 minutes (−39.5 min/d, 95% CI −100 to 21; *P*=.20) and >60 minutes (−49.9 min/d, 95% CI −102 to 2; *P*=.06) and increases in sit-to-stand transitions (4.2 per day, 95% CI −2 to 10; *P*=.13), although none were statistically significant. Hemoglobin A_{1c} was lower in the Sit Less group than in the control group (adjusted mean difference −0.4%, 95% CI −0.8 to −0.1; *P*=.04). Continuous glucose monitoring outcomes, including mean glucose (−6.0 mg/dL, 95% CI −14.9 to 2.8; *P*=.18) and time in range (2.1%, 95% CI −4.8 to 9.1; *P*=.54), were not statistically significant but demonstrated similar directional changes. Other cardiometabolic biomarkers, physical

XSL•FO
RenderX

activity, and psychological outcomes were not significantly different, except for lower self-efficacy for moderate physical activity in the Sit Less group (−0.6, 95% CI −1.1 to −0.2; *P*=.003).

Conclusions: The Sit Less intervention was feasible and acceptable and showed possible improvements in sedentary behavior and glycemic outcomes. Although this pilot study had a small sample size and limited statistical power, with multiple outcomes examined and wide CIs observed, these findings support evaluation in larger, adequately powered trials.

Trial Registration: ClinicalTrials.gov NCT05534256; https://clinicaltrials.gov/study/NCT05534256 and ClinicalTrials.gov NCT05691452; https://clinicaltrials.gov/study/NCT05691452

International Registered Report Identifier (IRRID): RR2-10.1371/journal.pone.0302582

(JMIR Mhealth Uhealth 2026;14:e56497) doi:10.2196/56497

KEYWORDS

sedentary behavior; mobile health; mHealth; digital health; activity tracker; cardiovascular disease; diabetes; text messages

Introduction

Background

Sedentary behavior is a strong modifiable risk factor for cardiometabolic disease [1,2]. It involves a lack of leg muscle contractions and restricted blood flow behind the knee when sitting, which causes reduced insulin sensitivity, vascular dysfunction, activation of low-grade inflammatory responses, and an imbalance in energy expenditure [3]. As a result, greater total sedentary time is related to increased cardiometabolic risk [2], including decreased high-density lipoprotein levels and increased triglycerides, fasting glucose [4], BMI, and waist circumference [5]. Prolonged sedentary time is also associated with a 12% higher risk of incidental cardiovascular disease (CVD) independent of physical activity levels [1]. Emerging evidence suggests that physical activity attenuates cardiometabolic risk but does not eliminate the increased risk associated with high sedentary time [1,6]. Despite the important role of sedentary behavior in CVD and metabolic disease, patients with cardiometabolic diseases, including patients with coronary artery disease or diabetes, spend 70% to 90% of their waking time in sedentary behavior [7-9] and 50% of their total sedentary time is prolonged (>30 min per bout) [7]. The reason for this prolonged sedentary time among these individuals may relate to fear of moderate-to-vigorous levels of physical activity due to their heart condition among those with CVD or the perceived development of hypoglycemia among those with diabetes [10,11]. Targeting sedentary behavior, which occupies most of the waking time of patients with cardiometabolic disease, could be a promising target behavior for secondary prevention in this population.

Frequent standing or walking can reduce total sedentary time and disrupt prolonged sitting, changing physiological pathways and improving cardiometabolic outcomes [12-14]. Others have shown that reducing sedentary time with frequent breaks is safe in other chronic health conditions and applicable to patients with cardiometabolic diseases [15-17]. Mobile technologies, including activity trackers (ie, Fitbit) and smart water bottles, offer a low-cost and sustainable approach to reducing sedentary behavior. The use of wearable technologies may be a promising strategy to reduce total sedentary time and disrupt prolonged sitting patterns through features such as self-monitoring and prompts. In addition, the use of a smart water bottle can naturally break sedentary time by encouraging increased water intake, resulting in more frequent trips to the kitchen and restroom. To date, no wearable technology–based sedentary behavior reduction programs focusing on sedentary breaks for patients with cardiometabolic diseases are available.

Study Objectives

This study aimed to assess the feasibility and acceptability of a multitechnology-based sedentary behavior reduction intervention (Sit Less program) in patients with cardiometabolic disease (CVD) or type 2 diabetes (T2D) and evaluate its preliminary efficacy in changes in sedentary behavior measured by total sedentary time, prolonged sedentary time (sedentary bouts >30 min and 60 min), and number of sit-to-stand transitions.

We hypothesized that participants in the Sit Less program would show greater reductions in total sedentary time and the number of prolonged sedentary bouts compared to those in the control group receiving usual care with an American Heart Association educational booklet. We also explored changes in secondary outcomes, including physical activity (eg, stepping time and standing time), cardiometabolic markers (eg, 24-h glycemic control, hemoglobin A_{1c} [HbA_{1c}], BMI, waist-to-hip ratio, blood pressure, insulin, high-sensitivity C-reactive protein [hs-CRP], and lipids), and psychological factors (eg, confidence in reducing sedentary behavior, confidence in increasing light levels of physical activity and moderate-to-vigorous levels of physical activity, and habit strength for sedentary behavior).

Methods

Study Design

Our study included 2 cohorts of participants with different cardiometabolic diseases (CVD or type 2 diabetes), each completing a randomized (1:1 ratio) controlled trial. The same protocol was implemented in each cohort, with randomization occurring separately. Both studies were approved by the Vanderbilt Institutional Review Board and registered on ClinicalTrials.gov (CVD cohort: NCT05534256, August 24, 2022; T2D cohort: NCT05691452, January 5, 2023). A detailed description of the study protocol has been published elsewhere [18]. The study was conducted in accordance with the CONSORT (Consolidated Standards of Reporting Trials) guidelines (Checklist 1).

XSL•FO
RenderX

Participants

Participants were eligible for inclusion in the trial if they were aged ≥18 years and had at least one of the following conditions: history of heart attack, coronary or carotid artery disease, ischemic heart disease, coronary stent placement, coronary artery bypass surgery, or T2D; self-reported sitting for ≥8 hours per day; ability to stand and walk; and owned a smartphone. Exclusion criteria included current use of an activity tracker, current participation in exercise or cardiac rehabilitation programs, non–English-speaking, unstable health conditions (eg, heart failure and uncontrolled arrhythmia) or kidney disease that limited daily water intake, any other conditions contradictory to standing or walking due to physical or cognitive limitations, or current pregnancy.

This pilot study was designed to assess the feasibility and acceptability of the Sit Less program rather than to conduct a fully powered hypothesis test. According to Viechtbauer et al [19], to detect a feasibility issue that has a 10% likelihood of occurrence, a sample size of 30 participants is required to identify the feasibility issue with a 95% CI.

Ethical Considerations

This study was approved by the Institutional Review Board of Vanderbilt University Medical Center (IRB: 220416 and 221566) and conducted in accordance with the Declaration of Helsinki. All participants provided written informed consent and were informed of their right to withdraw at any time. Data were deidentified to ensure confidentiality. All participants received a possible total of US $150—US $25 for completion of the baseline visit, US $25 for completion of the randomization visit, US $50 for completion of the postintervention visit, and US $50 for device return.

Recruitment

Participants were recruited from September 2022 to September 2023 through Vanderbilt University Medical Center outpatient clinics, the Vanderbilt University Medical Center employee distribution list, and ResearchMatch.org. Clinic patients received an opt-out letter or a message through the My Health at Vanderbilt app after electronic health record screening, followed by staff contact. Additionally, participants recruited via the Vanderbilt University Medical Center employee distribution list or ResearchMatch.org reached out to the study team after being notified by email about the study. Recruitment focused on participants residing near Nashville, Tennessee, due to blood sample collection.

Study Visit Schedule

The study included 3 visits: baseline, randomization, and postintervention. At baseline, eligible participants provided written informed consent, completed surveys, and underwent cardiometabolic assessments, including anthropometrics, blood pressure, and dried blood spot collection. Participants then wore the activPAL3 (PAL Technologies) devices on the thigh and a continuous glucose monitoring (CGM) device (FreeStyle Libre Pro; Abbott) for 7 days, 24 hours per day, and completed a daily sleep diary and a device wear log during this period. After 7 days, participants returned the CGM and activPAL3 devices in person and were randomized into either the Sit Less program or the control group (randomization visit). Following the 12-week intervention or control period, participants completed the same assessments and wore both devices again for 7 days (postintervention visit). Devices were returned by mail, and all participants received their pre- and postintervention data after study completion.

Randomization

Randomization occurred within each disease group. A 1:1 randomization list was created by a statistician and uploaded into the REDCap randomization module. Participants were informed of their group assignment immediately during their randomization visit. Participants and the interventionist were not blinded to the allocation due to the nature of the intervention program. However, outcome assessors and data analysts were blinded to the allocation.

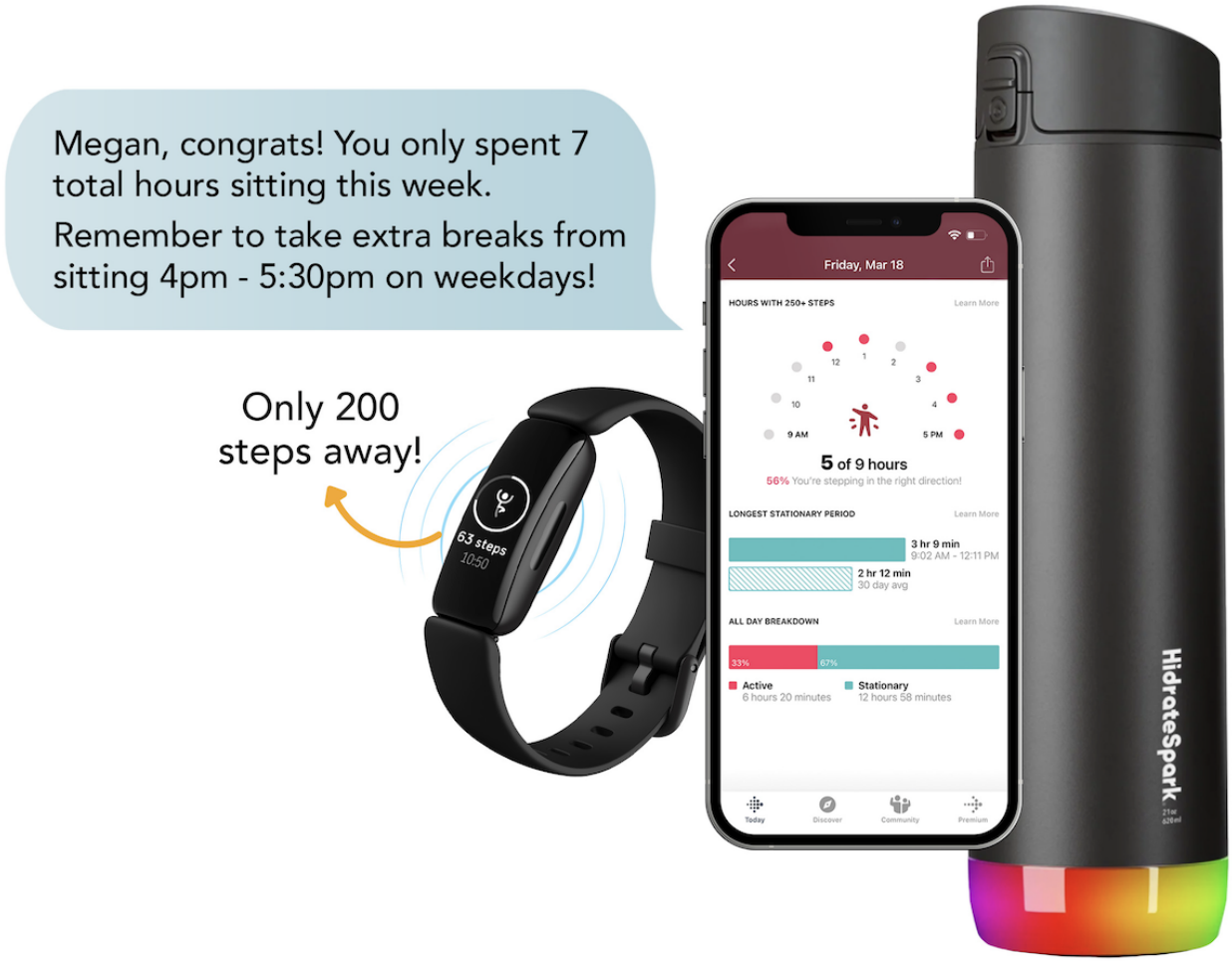
Sit Less Intervention

Overview

The 12-week Sit Less program was developed based on a Habit Formation framework [20] to target sedentary behavior. The Sit Less program included (1) one in-person instructional and goal-setting session at the randomization visit, (2) the use of a wearable device (Fitbit), (3) a smart water bottle (HidrateSpark), and (4) 3 weekly tailored text messages for behavior reinforcement and weekly goal monitoring. Figure 1 describes the Sit Less program. The overall goals of the Sit Less program were (1) to reduce daily sedentary time by 120 minutes, and (2) to achieve a sedentary break (standing or walking) for 5 minutes every 30 minutes by the end of the intervention period. Participants were asked to gradually reduce their sedentary time by at least 10 minutes, up to a maximum of 30 minutes, per week [21]. Detailed information regarding the intervention is described elsewhere [18].

XSL•FO
RenderX

Figure 1. Overview of a multiple technology–based sedentary behavior reduction intervention ("Sit Less" program).



Instructional or Goal-Setting Session

The principal investigator reviewed the activPAL data with Sit Less participants to identify targetable prolonged sedentary bouts. Using a modified Top Ten Tips (TTT) booklet—adapted from Habit Formation Theory—the principal investigator guided participants in developing habits to take frequent sedentary breaks. The original TTT booklet, which included stretching and muscle-strengthening activities [22], was revised to 8 tips focused specifically on reducing sedentary behavior [23]. Sit Less participants completed weekly worksheets to set goals and plan actions to reduce sedentary time based on the selected tips.

Fitbit

Fitbits (Inspire model) were used to support participants as they self-monitor sedentary hours (ie, "total stationary hours" and "longest stationary period") and to provide real-time sedentary break prompts (ie, "Move alert"). The goal was to help participants become aware of their habitual sedentary behavior [24]. The Fitbit "Move alert" alerts participants after 60 minutes of sitting, prompting a 2 - to 3-minute walk (≈250 steps) to reset the alert. Participants were advised to wear the Fitbit during waking hours, optionally at night, and to synchronize weekly. Participants who failed to synchronize or wear the Fitbit for 2 days received a reminder text; continued noncompliance for another 2 days prompted a call from research staff to troubleshoot issues [18].

HidrateSpark Smart Water Bottle

The smart water bottle (HidrateSpark Smart Water Bottle) was used to motivate participants to break their prolonged sitting and stand up and move by frequently visiting the restroom and refilling the bottle. The water bottle used flashing lights and phone notifications to remind participants to drink water if they did not meet personalized hydration targets based on their body weight, age, sex, and activity levels.

Weekly Text Messages

Participants received 3 algorithm-derived, Fitbit-based tailored messages: Thursday (midweek sedentary behavior summary), Sunday (encouragement for achieving the weekly goal), and Monday. The Monday message included feedback on prior week goals (based on Fitbit), a sedentary behavior summary, 1 TTT tip, and a prompt to set new goals for sitting time and the duration and frequency of sitting breaks. These messages supported habit formation and weekly planning to reduce sitting time and increase sitting breaks. The detailed algorithm and contents are described elsewhere [18].

XSL•FO
RenderX

Control Group

The control group received standard care and the American Heart Association's "Answers by Heart" booklet [25]. This booklet guides tracking and managing body weight, blood pressure, and cholesterol levels. The booklet also notes general advice on healthy cooking and physical activity [25]. The control group did not receive a Fitbit, a smart water bottle, or tailored text messages.

Study Measures

Feasibility

Feasibility was evaluated by assessing recruitment or enrollment rates, dropout rates (percentage of participants who withdrew consent among enrolled participants), and adherence to the Sit Less program, defined as (1) ≥75% of participants using the Fitbit for ≥5 valid days per week for 75% of the study duration and (2) engagement with ≥80% of goal setting texts [26].

Acceptability

Acceptability of the Sit Less program was measured through satisfaction surveys by combining items from 2 different questionnaires [27,28]. The Lyons et al [27] questionnaire assessed overall program satisfaction, ease of use, and continued intention for each component and was adapted to evaluate Sit Less technological tools (Fitbit, app, and smart water bottle). The Burner et al [28] questionnaire focused on the text messaging aspect, examining content relevance, motivational degree, frequency, and timing of messages. Responses were rated on a 5-point Likert scale, with 75% agreement indicating acceptability.

Primary Outcome: Sedentary Behavior

Using the activPAL3 devices, we measured sedentary behavior as total daily sitting time, prolonged sitting time (time spent in sedentary bouts ≥30 minutes and time spent in sedentary bouts ≥60 minutes), and the number of sit-to-stand transitions for 7 days at baseline and at postintervention. The activPAL3, worn on the thigh, accurately detects sitting and standing postures and has been validated against direct observation [29]. ActivPAL3 data were used for analysis if the participant wore the device for more than 10 hours on a given day, at least 3 days [30]. Data for analysis were extracted from the device using PAL analysis (activPAL, version 8.11) with the validated CREA algorithm, which closely matches diary data in classifying wake time activities [31].

Secondary Outcomes

Physical Activity

Daily time spent standing and walking (stepping) was measured using the activPAL3 and processed through the CREA algorithm.

Cardiometabolic Markers

Cardiometabolic biomarkers included glycemic outcomes and additional cardiometabolic measures. Glycemic outcomes comprised HbA_{1c} and CGM metrics. Fasting finger-prick blood samples were collected onto dried blood spot cards (ZRT Laboratory) for HbA_{1c} analysis. CGM (Abbott FreeStyle Libre Pro) was conducted over 7 days to assess 24-hour average glucose, glucose management indicator, glycemic variability, and time in range (70 - 180 mg/dL) [32].

Additional cardiometabolic measures included BMI, waist and hip circumferences, waist-to-hip ratio, blood pressure, insulin, hs-CRP, and lipids. Anthropometric and blood pressure measurements were obtained using standardized procedures, and dried blood spot samples were analyzed for insulin, lipids, and hs-CRP [33,34]. Dried blood spot testing has shown a strong correlation with conventional serum tests, making it a reliable and convenient tool for screening for cardiometabolic risk factors [33,34].

Psychological Factors

We measured participant confidence in reducing sedentary behavior using the 12 items from the Self-Efficacy Questionnaire for Physical Activity and Sedentary Behavior (previously reported Cronbach α=0.79 and 0.85) [35]. The items assess the level of confidence for specific sitting behaviors and sedentary breaks. Habit strength for sedentary behavior was assessed using the validated Self-Report Habit Index measure (previously reported Cronbach α=0.91, 0.82, and 0.83) [36]. This 7-item index was adapted to assess the degree to which sedentary breaks (standing or walking) became habitual [36]. Depressive symptoms were measured using the Patient Health Questionnaire-9 (PHQ-9) [37].

Demographic and Clinical Characteristics

Demographics, socioeconomic characteristics, medical history, current medication, tobacco use, and alcohol intake were collected via self-report. Participants were asked about their level of activity associated with their current occupation, and responses ranged from "mostly sedentary" (eg, desk-based job) to "very active and physical" (eg, manual labor). Using the 16-item Rapid Eating and Activity Assessment for Participants Short Version, we determined participants' dietary habits [38].

Statistical Analysis

Statistical analyses were conducted using SPSS Statistics 29 (IBM Corp) and STATA 18 (StataCorp). We summarized nominal and ordinal variables via frequency distributions. Continuous variables were summarized using means and SDs when normally distributed, such as the primary outcome of sedentary behavior and psychological factors, including self-efficacy for physical activity and sedentary behavior and habit strength for sedentary behavior. Medians with IQRs were also used when distributions were skewed or contained extreme outliers. This applied to continuous demographic and clinical characteristics, as well as secondary outcome variables, including physical activity, cardiometabolic and glucose monitoring biomarkers, and PHQ-9 scores. Baseline group differences were assessed using chi-square tests of independence (nominal and ordinal) and Mann-Whitney (continuous) tests.

We evaluated the intervention effects in the following steps. First, we calculated the change in each outcome variable from baseline to postintervention for each participant. Within-group changes were generated using median (for skewed variables) and mean (for normally distributed variables), with bootstrapped 95% CIs. Second, we conducted between-group analyses using

XSL•FO
RenderX

generalized linear models with bias-corrected robust (Huber sandwich estimator, STATA) SEs to evaluate the effects of the intervention on primary and secondary outcomes. Each model included the corresponding baseline value of the outcome, cohort (CVD vs T2D), and the interaction between study group and cohort. Baseline outcome values were included to account for baseline variability, while the cohort and the cohort×group interaction term were included to account for known differences in cardiometabolic responses between the 2 clinical cohorts and to assess whether intervention effects differed by cohort. Leisure-time exercise was additionally included because it was the only baseline variable that differed significantly between groups and was considered clinically relevant to the outcomes. To maintain model parsimony and reduce the risk of overfitting, no additional covariates were included beyond those with clear clinical or methodological justification.

For each model, multicollinearity among the independent variables was assessed first. There was no multicollinearity problem among the variables for all models. We then conducted each model using the Gaussian distribution with the identity link function and evaluated the residuals for indications of extreme influence, lack of normality, and/or heteroscedasticity. When assumptions were violated, models were reestimated using a Gamma distribution with a log link, and the residuals were reevaluated. We also conducted the same analyses using transformed variables and the Gaussian distribution, with similar findings resulting from the models using the Gamma, log link function, enabling us to generate estimated mean differences on the original variable scale of measurement. A few biomarker variables had extreme outliers. Those variables could not be addressed through transformation or alternative model specifications. Given the small sample size, extreme values were winsorized by replacing them with the next highest observed value. Model assumptions were reevaluated following this adjustment. Final generalized linear models with appropriate distributions and link functions were used, and robust SEs were applied to estimate adjusted mean differences with 95% CIs. These estimates were used to evaluate the effect of the study group and to provide measures of effect size. All analyses were conducted using an intention-to-treat approach without data imputation.

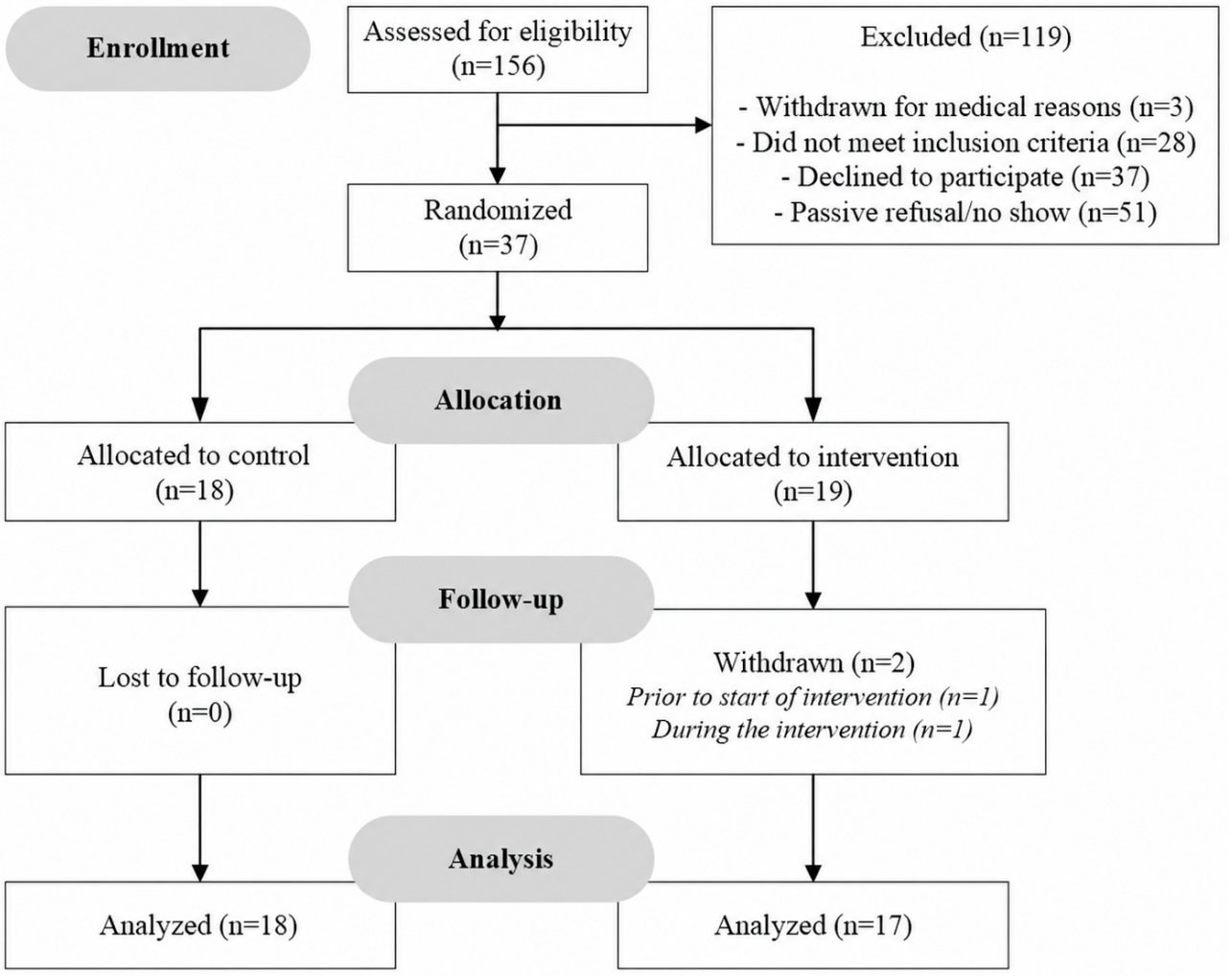
Results

Participants

Of 156 participants who were contacted and assessed for eligibility, 128 (80%) met inclusion criteria, and of those, 37 (30%) completed all baseline assessments and were randomized (Figure 2). Of the 37 randomized participants, 19 (51%) were randomly assigned to the Sit Less group, and 18 (49%) were assigned to the control group. One (5%) participant from the Sit Less group completed baseline assessments but withdrew before the start of the intervention due to privacy concerns about the Fitbit displaying text messages and incoming calls. Another Sit Less group participant withdrew during the intervention due to a Fitbit data recording error. This participant used a walker, and because of the lack of wrist movement while walking with the walker, the Fitbit did not capture any steps or movement. Therefore, 35 (95%) of 37 participants completed the study and were included in the analysis (Sit Less: n=17; and control: n=18).

XSL•FO
RenderX

Figure 2. The CONSORT (Consolidated Standards of Reporting Trials) flow diagram for the Sit Less study with type 2 diabetes and cardiovascular disease cohort.



Baseline Demographics and Clinical Characteristics

The median age of the final sample was 67 (IQR 54 - 72) years. With the exception of leisure time exercise at baseline, no significant differences between groups were found in the demographic or clinical characteristics (Table 1). At baseline, 7 (39%) of 18 participants in the control group and 14 (82%) of 17 participants in the Sit Less group reported leisure-time exercise (*P*=.009).

Both the Sit Less and control groups wore the activPAL continuously (24 h/d), with minimal nonwear time (median 0 min/d, IQR 0 - 0). Participants provided a median of 7 valid wear days at both preintervention (IQR 7 - 8) and postintervention (IQR 6 - 7). On average, the participants spent a median of 624 minutes sitting (IQR 521 - 686), 182 minutes standing (IQR 108 - 230), 72 minutes stepping (IQR 45 - 86), 32 minutes in secondary lying positions (IQR 0 - 83), and 490 minutes sleeping (IQR 450 - 523). No significant between-group differences were found in nonwear time and baseline activities except for stepping time.

XSL•FO
RenderX

Table . Baseline demographic and clinical characteristics.

Sample characteristics	Overall (N=35)	Control (n=18)	Sit Less (n=17)	*P* value[a]
Age (y), median (IQR)	67.0 (54-72)	64.5 (54-73)	68.0 (55-72)	.78
Dietary habits (REAP-S[b]), median (IQR)	28.0 (26-32)	30.0 (25-32)	28.0 (26-31)	.79
Condition cohort, n (%)				.36
Diabetes	13 (37.1)	8 (44.4)	5 (29.4)	
CVD[c]	22 (62.9)	10 (55.6)	12 (70.6)	
Gender, n (%)				.72
Female	13 (38.2)	7 (41.2)	6 (35.3)	
Male	21 (61.8)	10 (58.8)	11 (64.7)	
Married or partnered, n (%)				.72
No	13 (38.3)	6 (35.3)	7 (41.2)	
Yes	21 (61.8)	11 (64.7)	10 (58.8)	
Race or ethnicity, n (%)				.99
Non-Hispanic White	28 (82.4)	14 (82.4)	14 (82.4)	
Non-Hispanic Black	4 (11.8)	2 (11.8)	2 (11.8)	
Asian, Hispanic, or other	2 (5.9)	1 (5.9)	1 (5.9)	
Education level, n (%)				.27
Less than college	11 (32.4)	7 (41.2)	4 (23.5)	
College or higher	23 (67.6)	10 (58.8)	13 (76.5)	
Income sufficiency, n (%)				.55
Poorly or not well	3 (8.8)	2 (11.8)	1 (5.9)	
Well or very well	31 (91.2)	15 (88.2)	16 (94.1)	
Employment, n (%)				.44
Full	14 (41.2)	8 (47.1)	6 (35.3)	
Part	7 (20.6)	2 (11.8)	5 (29.4)	
Unemployed, retired, or other	13 (38.2)	7 (41.2)	6 (35.3)	
If employed, the job (overall, n=21; control, n=10; Sit Less, n=11)				.70
Active	5 (23.8)	2 (20.0)	3 (27.3)	
Sedentary	16 (76.2)	8 (80.0)	8 (72.7)	
Alcohol intake (last 12 mo), n (%)				.48
No	8 (22.9)	5 (27.8)	3 (17.6)	
Yes	27 (77.1)	13 (72.2)	14 (82.4)	
Smoke cigs (last 12 mo), n (%)				.26
No	31 (88.6)	17 (94.4)	14 (82.4)	
Yes	4 (11.4)	1 (5.6)	3 (17.6)	
Leisure time exercise, n (%)				.009
No	14 (40.0)	11 (61.1)	3 (17.6)	
Yes	21 (60.0)	7 (38.9)	14 (82.4)	

XSL•FO
RenderX

Sample characteristics	Overall (N=35)	Control (n=18)	Sit Less (n=17)	*P* value[a]
Health conditions, n (%)				
High blood pressure	19 (54.3)	11 (61.1)	8 (47.1)	.40
High cholesterol	21 (61.8)	11 (64.7)	10 (58.8)	.72
Arthritis	23 (65.7)	13 (72.2)	10 (58.8)	.40
Sleep apnea	15 (42.9)	8 (44.4)	7 (41.2)	.85
Other sleep problems	14 (40.0)	7 (38.9)	7 (41.2)	.89

[a]*P* values were calculated using the Mann-Whitney *U* test for continuous variables with skewed distributions (age and REAP-S score) and the chi-square test for categorical variables.

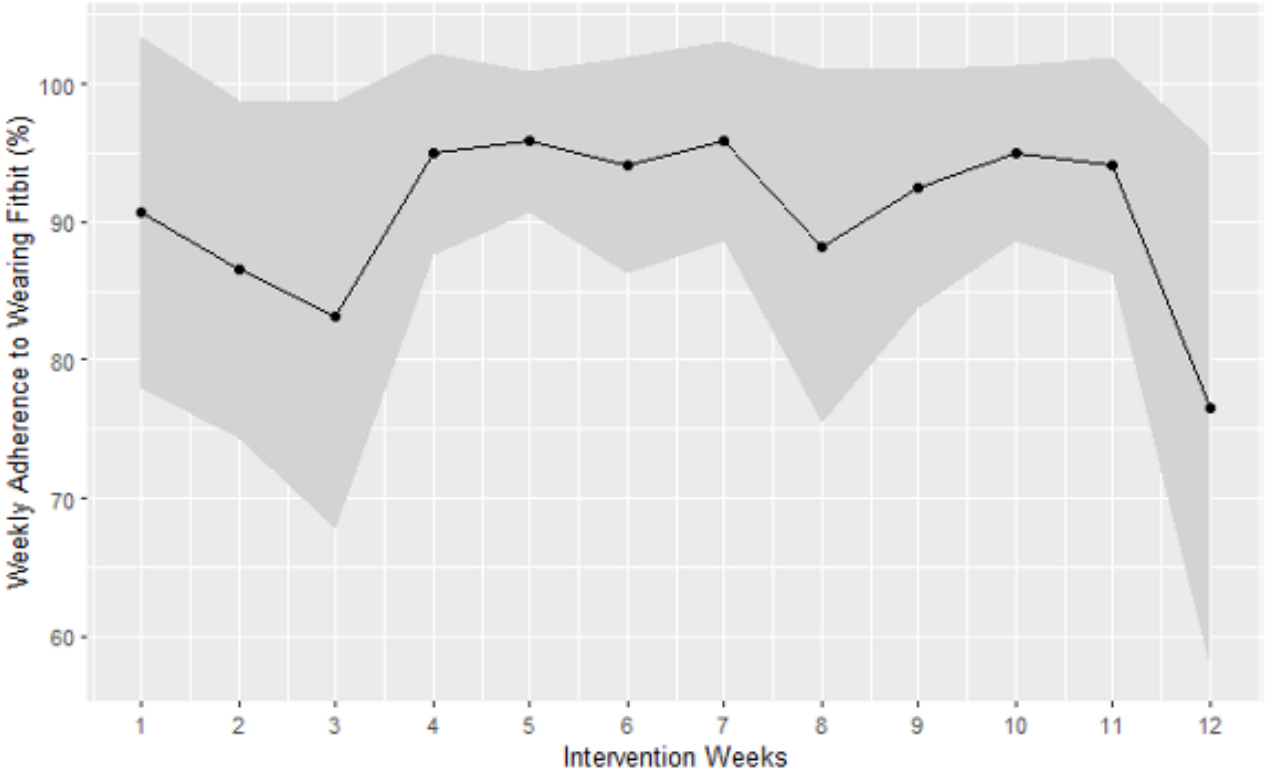
[b]REAP-S: Rapid Eating and Activity Assessment for Participants Short Version.

[c]CVD: cardiovascular disease.

Feasibility

As noted earlier, among the 37 participants enrolled, 35 (95%) completed all study activities. The Sit Less group had a dropout rate of 2 (11%) of 19; the reason for the dropout was privacy concerns about the Fitbit displaying text messages and incoming calls and Fitbit's data recording errors. Adherence to the Sit Less program was high according to our adherence definition (Figure 3). Fifteen (88%) of 17 participants wore the Fitbit for more than 75% of the intervention period. On average, participants wore their Fitbit for 76 of 84 days. Weekly, the Sit Less participants had a median of 6.8 valid wear days per week—where a valid day indicated wearing the Fitbit for more than 10 hours. Daily median Fitbit wear time was 928 minutes per day (IQR 902 - 944). Adherence to Fitbit wear peaked at weeks 5 and 7, with 96% wearing the device, and was maintained throughout the study period, except for week 12, when adherence dropped to 76%.

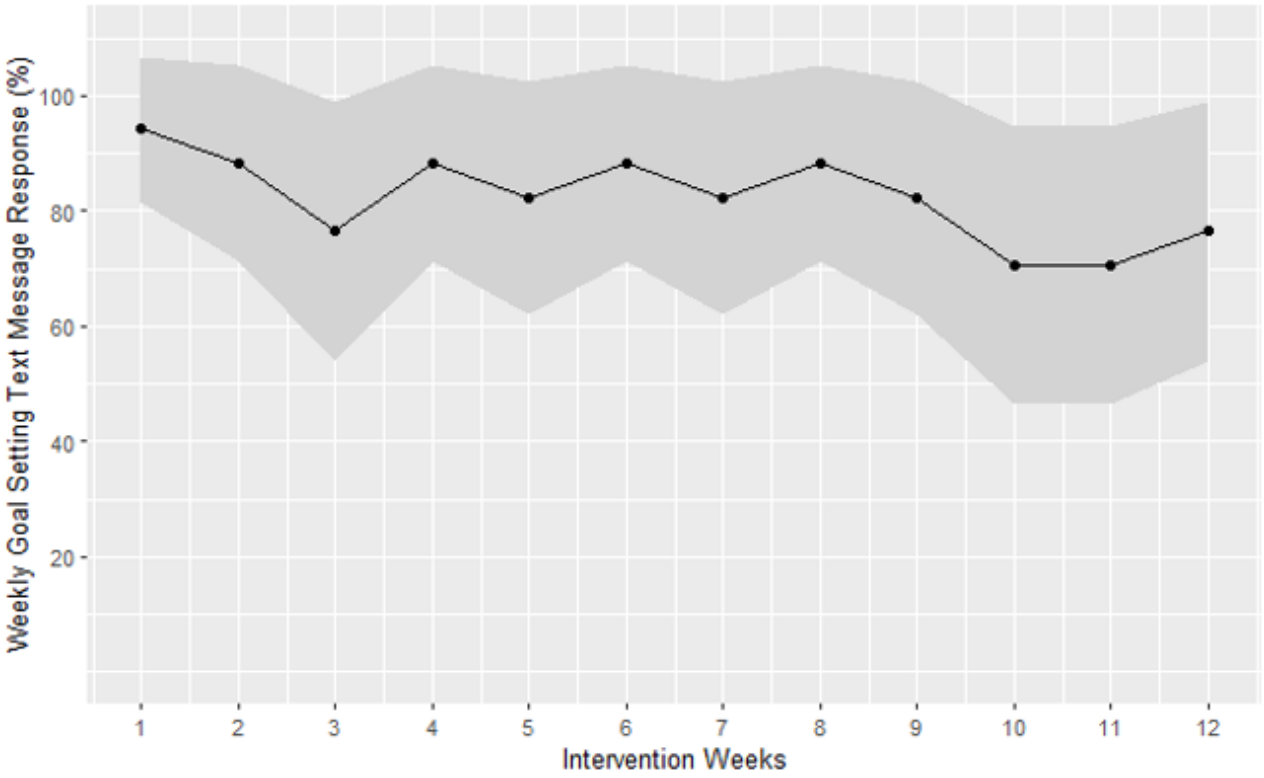
Figure 3. Weekly average adherence to wearing a Fitbit over 12 weeks. The gray area indicates the 95% CI.



For text message responses, participants were diligent in setting weekly goals and replying to us (Figure 4). The average rate of text message response over 12 weeks was 81%. Until week 9, participants consistently responded to goal-setting messages, with adherence rates ranging from 91% to 83%. However, in the last 3 weeks of the program, response rates decreased, ranging between 58% and 66%.

XSL•FO
RenderX

Figure 4. Weekly average adherence to responding to goal-setting text messages over 12 weeks. The gray area indicates the 95% CI.



Acceptability

A total of 17 Sit Less participants completed the satisfaction survey. Fifteen (88%) of 17 participants indicated that they would recommend the program to family and friends. Fifteen (88%) of 17 participants indicated that the program increased awareness of their sedentary behavior. Sixteen (94%) of 17 participants liked wearing the Fitbit monitor; 13 (77%) of 17 participants would continue checking sitting hours on Fitbit, and 11 (65%) of 17 participants indicated that the "Move Alert" from Fitbit was useful. The majority (11/17, 65%) of participants also indicated that they found the smart water bottle convenient to use. Fifteen (88%) of 17 Sit Less participants either agreed (47%) or remained neutral (41%) about whether they would like to receive text messages for a longer period. The majority (13/17, 77%) of participants either agreed or were neutral that the tips and advice felt specific to them, and they were motivated by the text messages. Most participants (16/17, 94%) indicated that the number of text messages was "just right."

Primary Outcome: Sedentary Behavior

Group differences in sedentary behavior at baseline and postintervention are presented in Table 2. Overall, the Sit Less intervention showed consistent directional improvements compared with the control group, although between-group differences were not statistically significant. For total sitting time, the control group increased from 588.3 (SD 100.2) to 593.6 (SD 83.4) minutes per day, whereas the Sit Less group decreased from 643.6 (SD 113.0) to 605.0 (SD 111.1) minutes per day. After adjustment for baseline values, cohort, cohort×group interaction, and leisure-time exercise, the Sit Less group showed a nonsignificant reduction of 17.3 minutes per day in total sitting time compared with the control group at postintervention (95% CI −63 to 28; *P*=.45).

XSL•FO
RenderX

Table . Changes in sedentary behavior outcomes from baseline to postintervention by study group (N=35; control: n=18; Sit Less: n=17; all baseline between-group differences were *P*>.05).

Outcomes	Baseline, mean (SD)	Postintervention, mean (SD)	Mean change (95% CI)	Estimated adjusted mean difference (95% CI)[a]	*P* value[b]
Sitting time (min/d)				−17.3 (−63 to 28)	.45
Control	588.3 (100.2)	593.6 (83.4)	5.3 (−38 to 50)		
Sit Less	643.6 (113.0)	605.0 (111.1)	−38.6 (−71 to 4)		
Sitting and secondary lying time (min/d)				−21.7 (−76 to 32)	.43
Control	660.2 (112.0)	669.4 (83.5)	9.2 (−47 to 70)		
Sit Less	679.4 (129.4)	642.0 (113.5)	−37.4 (−80 to 11)		
Time spent sitting bouts >30 min (min/d)				−39.5 (−100 to 21)	.20
Control	345.0 (144.0)	355.6 (106.4)	10.7 (−41 to 65)		
Sit Less	355.5 (135.7)	304.8 (140.5)	−50.7 (−115 to 8)		
Time spent sitting bouts >60 min (min/d)				−49.9 (−102 to 2)	.06
Control	203.1 (144.8)	205.0 (116.6)	1.9 (−47 to 44)		
Sit Less	194.4 (113.0)	142.4 (112.8)	−52.0 (−111 to 4)		
Sit-to-stand transitions (n per day)				4.2 (−2 to 10)	.13
Control	36.2 (13.7)	36.8 (12.0)	0.6 (−2 to 3)		
Sit Less	43.7 (9.9)	46.6 (8.9)	2.9 (−3 to 9)		

[a]Test of the difference between adjusted marginal means using a generalized linear model with a Gaussian distribution and identity link, with robust SEs (χ^2 statistic, df=1).

[b]Adjusted for study cohort, cohort×group interaction, baseline value, and leisure-time exercise; estimates generated using robust SEs.

Time spent in sedentary bouts of >30 minutes increased in the control group and decreased in the Sit Less group (Table 2), with a nonsignificant adjusted between-group difference of −39.5 minutes per day (95% CI −100 to 21; *P*=.20). For sedentary bouts of >60 minutes, the control group showed minimal change (1.9 min/d; 95% CI −47 to 44), whereas the Sit Less group decreased by 52.0 minutes per day (95% CI −111 to 4), with an adjusted difference of −49.9 minutes per day (95% CI −102 to 2; *P*=.06).

Sit-to-stand transitions were unchanged in the control group (36.2, SD 13.7 vs 36.8, SD 12.0) but increased in the Sit Less group (43.7, SD 9.9 vs 46.6, SD 8.9), with a nonsignificant adjusted mean difference of 4.2 transitions per day (95% CI −2 to 10; *P*=.13).

Secondary Outcome: Physical Activity

Group differences in physical activity at baseline and postintervention are presented in Table 3. Although not statistically significant, participants in the Sit Less group increased both standing and stepping time by approximately 10 minutes more than those in the control group, despite variability in change scores. For standing time, the control group decreased from 170.5 (IQR 99 - 255) to 158.1 (IQR 115 - 256) minutes per day, whereas the Sit Less group increased from 182.9 (IQR 144 - 228) to 193.4 (IQR 144 - 239). The adjusted between-group difference was 9.7 minutes per day (95% CI −52 to 71; *P*=.76), but it was not significant.

For stepping time, baseline values differed between groups (control: median 49.7 min/d, IQR 37 - 75; Sit Less: median 80.8 min/d, IQR 52 - 108; *P*=.01). Following the intervention, the control group and the Sit Less group both increased stepping time, but there was no statistically significant between-group difference following the intervention (adjusted mean difference 9.6 min/d, 95% CI −15 to 34; *P*=.44).

Step counts also increased in both groups; however, there was no statistically significant between-group difference following the intervention (adjusted mean difference 711.4 steps per day, 95% CI −983 to 2406; *P*=.41).

XSL•FO
RenderX

Table . Changes in physical activity outcomes from baseline to postintervention by study group (N=35; control: n=18; Sit Less: n=17)[a].

Outcomes	Baseline, median (IQR)	Postintervention, median (IQR)	Median change (95% CI)	Estimated adjusted mean difference[b] (95% CI)	*P* value[c]
Standing time (min/d)				9.7 (−52 to 71)	.76
Control	170.5. (99 - 255)	158.1 (115 - 256)	26.1 (−39 to 59)		
Sit Less	182.9 (144 - 228)	193.4 (144 - 239)	−0.5 (−30 to 44)		
Stepping time (min/d)				9.6 (−15 to 34)	.44
Control	49.7 (37-75)[d]	58.8 (45 - 80)	11.7 (−1 to 19)		
Sit Less	80.8 (52 - 108)	89.5 (53 - 107)	−0.8 (−11 to 22)		
Step counts (steps per day)				711.4 (−983 to 2406)	.41
Control	3682 (2680 - 5248)	4085 (3518 - 6274)	991 (−145 to 1340)		
Sit Less	5759 (4254 - 8337)	6299 (4216 - 8594)	36 (−447 to 1427)		

[a]Due to moderate skewness, median (IQR) and median change used to describe the observed distributions.

[b]Adjusted for the main effect of study cohort, the interaction of study cohort with study group, exercise during leisure time prior to study, and baseline values; estimated mean differences generated using robust SEs.

[c]Test of the difference between adjusted marginal means using a generalized linear model with a Gaussian distribution and identity link, with robust standard errors (χ^2 statistic, df=1).

[d]Baseline between-group difference, Mann-Whitney *U* test, *P*<.05. All other baseline between-group differences were *P*>.05.

Secondary Outcomes: Cardiometabolic Markers

Glycemic outcomes at baseline and postintervention are summarized in Table 4. HbA_{1c} levels were significantly lower in the Sit Less group compared with the control group at postintervention (adjusted mean difference −0.4%, 95% CI −0.8 to −0.1; *P*=.04), with a decrease observed in the Sit Less group (median change −0.4%, 95% CI −0.6 to −0.2) and minimal change in the control group (median change 0.1%, 95% CI −0.1 to 0.5). Although not statistically significant, CGM outcomes were directionally consistent with HbA_{1c}. Adjusted mean differences indicated lower glucose management index (−0.1%, 95% CI −0.4 to 0.1; *P*=.16) and average glucose level (−6.0 mg/dL, 95% CI −14.9 to 2.8; *P*=.18) in the Sit Less group compared with the control group at postintervention (Table 4). Participants wore the CGM device for a median of 9 days in both groups, with no between-group differences. Sensitivity analyses excluding extreme outliers were conducted for HbA_{1c} (n=1 control), CGM-based glucose management index (n=2 controls), and average glucose level (same 2 controls as GMI). While the estimated mean differences and *P* values changed slightly, interpretations of the results remained the same. Those summaries and findings are provided in Multimedia Appendix 1.

XSL•FO
RenderX

Table . Changes in glycemic outcomes from baseline to postintervention by study group (N=35)[a].

Outcomes	Baseline, median (IQR)	Postintervention, median (IQR)	Median change (95% CI)	Estimated adjusted mean difference[b] (95% CI)	*P* value
HbA_{1c}[c], % (control: n=18 and Sit Less: n=16)				−0.4 (−0.8 to −0.1)	.04[d]
Control	5.5 (4.3 - 5.7)	5.3 (4.8 - 5.9)	0.1 (−0.1 to 0.5)		
Sit Less	5.0 (4.7 - 5.9)	4.9 (4.4 - 5.5)	−0.4 (−0.6 to −0.2)[e]		
Glucose management index, % (control: n=15 and Sit Less: n=15)				−0.1 (−0.4 to 0.1)	.16[f]
Control	5.7 (5.6 - 6.6)	6.0 (5.4 - 6.5)	−0.1 (−0.5 to 0.1)		
Sit Less	6.0 (5.5 - 6.4)	5.7 (5.5 - 6.2)	−0.1 (−0.2 to 0.1)		
Average glucose level, mg/dL (control: n=15 and Sit Less: n=15)				−6.0 (−14.9 to 2.8)	.18[f]
Control	103.0 (96 - 136)	113.0 (90 - 133)	−2.0 (−18 to 5)		
Sit Less	113.0 (94 - 127)	100.0 (93 - 120)	−3.0 (−8 to 1)		
Glycemic variability (coefficient of variation), % (control: n=15 and Sit Less: n=15)				0.6 (−3.5 to 4.6)	.78[g]
Control	21.7 (18 - 32)	22.2 (19 - 30)	−0.1 (−2.8 to 1.8)		
Sit Less	20.3 (16 - 34)	20.9 (17 - 27)	1.3 (0.3 to 2.0)[e]		
Time in range (70 - 180 mg/dL), % (control: n=15 and Sit Less: n=15)				2.1 (−4.8 to 9.1)	.54[g]
Control	91.0 (75 - 94)	94.0 (71 - 96)	0.0 (−5 to 5)		
Sit Less	94.0 (85 - 98)	94.0 (86 - 98)	1.0 (0 to 1)		

[a]Due to extreme skewness, median (IQR) and median change used to describe the observed distributions; all baseline between-group differences were *P*>.05.

[b]Adjusted for study cohort, cohort×group interaction, baseline value, and leisure-time exercise; estimates generated using robust SEs.

[c]$HbA_{1c:}$ hemoglobin $_{A1c}$.

[d]Test of the difference between adjusted marginal means using a generalized linear model with a Gamma distribution and log link function, with robust SEs (chi-square statistic, df=1). One extreme outlier in the control group at postintervention was winsorized prior to analysis.

[e]Within-group change statistically significant based on bootstrapped 95% CI.

[f]Test of the difference between adjusted marginal means using a generalized linear model with a Gamma distribution and log link function, with robust SEs (chi-square statistic, df=1). Two extreme outliers in the control group, one at baseline and another at postintervention were winsorized prior to analysis.

[g]Test of the difference between adjusted marginal means using a generalized linear model with a Gaussian distribution and identity link, with robust SEs (chi-square statistic, df=1).

No statistically significant between-group differences were observed for other cardiometabolic outcomes, including weight, BMI, waist circumference, insulin, or blood pressure. Total cholesterol and low-density lipoprotein decreased and high-density lipoprotein increased, whereas triglycerides increased; however, none of these adjusted between-group differences were statistically significant (Multimedia Appendix 2).

Secondary Outcomes: Psychological Outcomes

Psychological outcomes at baseline and postintervention are summarized in Table 5. No adjusted between-group differences were observed for sedentary behavior habit strength (adjusted mean difference 0.0; 95% CI −0.7 to 0.8; *P*=.89). Self-efficacy for increasing physical activity increased in the control group but decreased in the Sit Less group, with a significant between-group difference observed for moderate activity (adjusted mean difference −0.6, 95% CI −1.1 to −0.2; *P*=.003;

XSL•FO
RenderX

Table 5). For depressive symptoms (PHQ-9), the adjusted between-group difference at postintervention was not statistically significant.

Table . Changes in psychological outcomes from baseline to postintervention by study group (N=35)[a].

Outcomes	Baseline, mean (SD)	Postintervention, mean (SD)	Mean change (95% CI)	Estimated adjusted mean difference[b] (95% CI)	*P* value[c]
Sedentary behavior habit strength (minimum=1 and maximum=7)				0.0 (−0.7 to 0.8)	.89
Control	6.0 (0.7)	5.8 (0.9)	−0.2 (−0.6 to 0.3)		
Sit Less	5.9 (1.1)	5.8 (1.0)	−0.1 (−0.7 to 0.4)		
Self-efficacy in reducing sitting (minimum=1 and maximum=5)				−0.4 (−0.9 to 0.2)	.15
Control	3.6 (0.9)	3.3 (0.9)	−0.3 (−0.8 to 0.2)		
Sit Less	3.5 (0.6)	3.1 (1.0)	−0.4 (−0.8 to 0.1)		
Self-efficacy in increasing light physical activity (minimum=1 and maximum=5)				−0.4 (−0.9 to 0.1)	.07
Control	3.6 (0.9)	3.6 (0.7)	0.1 (−0.3 to 0.5)		
Sit Less	3.6 (0.6)	3.5 (0.8)	−0.1 (−0.5 to 0.2)		
Self-efficacy in increasing moderate physical activity (minimum=1 and maximum=5)				−0.6 (−1.1 to −0.2)	.003
Control	3.3 (1.2)	3.7 (1.2)	0.4 (0.0 to 0.9)		
Sit Less	3.8 (1.2)	3.8 (1.1)	−0.1 (−0.5 to 0.3)		
PHQ-9[d] (minimum=0 and maximum=27), median (IQR)				−0.2 (−1.4 to 1.1)	.76
Control	4.0 (1-10)	3.0 (1-7)	−1.0 (−2 to 0)		
Sit Less	4.0 (2-5)	2.0 (1-5)	−1.0 (−2 to 0)		

[a]Control: n=18; Sit Less: n=17; all baseline between-group differences were *P*>.05.

[b]Adjusted for the main effect of study cohort, the interaction of study cohort with study group, exercise during leisure time prior to study, and baseline values; estimated mean differences generated using with robust SEs.

[c]Test of the difference between adjusted marginal means using a generalized linear model with a Gaussian distribution and identity link, with robust SEs (chi-square statistic, df=1).

[d]Patient Health Questionnaire-9 (PHQ-9) scores are reported as median (IQR) at baseline, postintervention, and for median change due to a skewed distribution. All other psychological values are presented as mean (SD) unless otherwise specified.

Discussion

Principal Findings

Frequent, short breaks (walking 5 min for every hour sitting) are practical, low-burden alternatives to exercise, proven to improve glucose and insulin levels [7,39]. However, intervention studies on sedentary breaks in T2D or CVD are limited and mostly short-term laboratory studies [7,39]. This is the first study integrating Fitbit devices and smart water bottle technologies and algorithm-driven text messages to encourage frequent sedentary breaks. Focusing on sedentary breaks addresses barriers associated with traditional exercise programs—such as fear of hypoglycemia, cardiac discomfort, and time constraints [1,6,11]—while concurrently reducing overall sedentary time and increasing standing and walking duration. The intervention demonstrated high feasibility, acceptability, and adherence within the patient with cardiometabolic diseases groups over a 12-week period, characterized by high engagement with text messaging and consistent use of Fitbits. Participants in the Sit Less group also reported that wearable technology and smart water bottles were instrumental in breaking up their prolonged sitting. Notably, these positive outcomes were observed in a participant group

XSL•FO
RenderX

with a median age of 67 years, indicating the potential of our multitechnology-based Sit Less program to reduce sedentary time and foster more active lifestyles among older adults with cardiometabolic conditions.

In this randomized controlled pilot trial, we were underpowered to detect intervention effects on outcomes, and CIs are wide; however, we still detected some directional changes in outcomes of interest in the Sit Less group relative to the control group. The Sit Less program was associated with reductions in total daily sedentary time compared with the control group, although between-group differences were not statistically significant. Over 12 weeks, the control group increased sitting time by 5.3 minutes per day, whereas the Sit Less group reduced sitting time by 38.6 minutes per day, corresponding to an adjusted difference of −17.3 minutes per day. Considering that reducing and replacing 30 minutes of sedentary time yields significant cardiometabolic benefits and reduces the risk of developing CVD and mortality [21], the observed reductions in the Sit Less group may have clinical relevance for individuals with diabetes and/or CVD.

Consistent patterns were observed for prolonged sedentary behavior, with decreases in time spent in bouts longer than 30 and 60 minutes in the Sit Less group compared with the control group. Prolonged, uninterrupted sitting is more strongly associated with adverse cardiometabolic outcomes than interrupted sitting [40,41]. A recent meta-analysis, while accounting for total sedentary time and moderate-to-vigorous physical activity time, revealed that even a single break per hour can modestly improve BMI, waist circumference, inflammation, and blood pressure [12,13]. Experimental studies confirm that sedentary breaks can improve glucose levels by 2% to 17% and fasting insulin levels by 15% [12,14]. In line with this, Sit Less participants had more sedentary breaks, shorter prolonged sitting, and lower HbA_{1c} (via dried blood spot) than controls postintervention. Our findings suggest that the intervention was feasible for targeting prolonged sedentary time and provide preliminary suggestions for further investigation in adequately powered trials. However, given the pilot nature of the study, small sample size, and multiple outcomes examined, the HbA_{1c} finding should be interpreted cautiously. Future studies powered to detect intervention effects and mechanisms of effects are needed.

With regard to physical activity outcomes, the Sit Less group spent 10 minutes more on stepping activities and standing compared to the control group postintervention, after adjustments. This level of change may have been insufficient to improve some cardiometabolic outcomes, including blood pressure and anthropometric measures. Considering that replacing sedentary time with higher-intensity physical activity leads to more substantial improvements in cardiometabolic markers, future interventions should encourage not only standing but also engaging in higher-intensity stepping for greater cardiometabolic benefits.

Despite these improvements in objective outcomes, psychological outcomes (eg, self-efficacy for reducing sedentary behavior and habit strength) did not change. Moreover, self-efficacy for increasing moderate physical activity was significantly higher in the control group compared to the Sit Less group at postintervention after adjustment. One possible explanation is that participants in the Sit Less group may have become more aware of the challenges of replacing sedentary behavior with moderate levels of physical activity, a phenomenon known as response shift, which is commonly observed in self-reported outcomes in behavioral interventions [42]. However, it is also possible that the measures used were not sufficiently sensitive to detect short-term changes in psychological constructs such as habit formation or that the pilot sample size was underpowered to detect modest psychological effects. Future research could incorporate approaches such as postintervention pretests, in which participants retrospectively assess their baseline perceptions, to better capture these changes. Future research could use a postintervention pretest in which participants report after the intervention what they would have reported at pretest, given what they have learned through the intervention.

Limitations

The goals of this pilot study were to assess the acceptability of the intervention and the feasibility of the study design. The small sample size aligns with these goals but limits our ability to make conclusions about the efficacy of the Sit Less intervention. The study was not powered to formally test hypotheses regarding intervention efficacy, and CIs were wide. Those findings should therefore be interpreted as a preliminary suggestion for testing the Sit Less intervention in a future fully powered trial. To maintain model parsimony and reduce the risk of overfitting, no additional exploratory covariates were included in the final models beyond those considered clinically or methodologically essential. Accordingly, the significant finding for HbA_{1c} should be interpreted with caution in light of the limited sample size and potential model instability.

In addition, several features of this study limit our generalizability. For instance, participants were recruited from a single hospital in Nashville, USA, and were predominantly White and highly educated. Although feasibility and acceptability were high in this sample, challenges such as varying comfort with technology and device usability may be more pronounced in other populations.

Moreover, baseline HbA_{1c} levels (median, approximately 5.0% - 5.5%) were within the nondiabetic or well-controlled range for many participants, and no participants had baseline HbA_{1c} levels >7.0%, further limiting generalizability. In addition, given the very small sample size within each cohort, the study is not designed to estimate differential treatment effects between CVD and T2D.

Finally, we are unable to determine the mechanisms driving change in objective outcomes. For instance, we did not collect objective data on smart water bottle use (eg, refill frequency), limiting our ability to assess its independent contribution to behavioral changes. Future larger studies, including those using multiple intervention conditions or fractional factorial designs that vary components, could examine how different components impact outcomes to understand mechanisms of change and seek

XSL•FO
RenderX

to evaluate Sit Less in more diverse and clinically complex populations.

Clinical Implications

Breaking up prolonged sitting should be a priority in the clinical management of patients with cardiometabolic diseases, as even short, frequent activity breaks can improve glycemic control and reduce sedentary time. The Sit Less program achieved these outcomes by integrating simple standing and walking breaks into daily routines, effectively overcoming common exercise barriers such as fear of hypoglycemia, cardiac discomfort, and time constraints. Its multitechnology was well received by older adults, reinforcing its suitability for patients with cardiometabolic diseases. In addition, the dropout of 1 participant due to inaccurate Fitbit data associated with walker use highlights a limitation of wrist-worn activity monitors in this population. Such devices may not accurately capture activity in individuals who rely on assistive devices, as arm movement is restricted. This should be considered in future studies and implementation efforts involving older or mobility-impaired cardiometabolic populations, where alternative monitoring approaches such as hip monitoring may be more appropriate.

Conclusions

In conclusion, the Sit Less intervention effectively integrated multiple technologies to reduce sedentary behavior in patients with cardiometabolic diseases. By targeting prolonged sedentary time, the program has shown a trend to decrease in total sitting time, increase in sedentary breaks, and possible improvements in glycemic outcomes. Our approach, which minimizes typical exercise barriers and promotes short active breaks, has demonstrated high feasibility and acceptance, suggesting its promise for future large-scale interventions. While further research is necessary to confirm these findings and assess their impact on a range of cardiometabolic markers, the initial results advocate for health care strategies that promote sedentary breaks to manage cardiometabolic disease.

Funding

This study was funded by the Moore Foundation and NIH/NIDDK (NIHDK020593). The funder played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript. Dr. Park is supported, in part, by New Faculty Startup Fund from Seoul National University the under Award Number 810-20240026. This research was conducted while Dr. Chorong Park was affiliated with Vanderbilt University School of Nursing.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request and with appropriate approvals in place.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sensitivity analyses of changes in glycemic outcomes from baseline to postintervention by study group.
[DOCX File, 17 KB - mhealth_v14i1e56497_app1.docx]

Multimedia Appendix 2
Changes in additional cardiometabolic outcomes from baseline to postintervention by study group.
[DOCX File, 20 KB - mhealth_v14i1e56497_app2.docx]

Checklist 1
CONSORT-EHEALTH (version 1.6.1)—Submission_Publication Form.
[PDF File, 1144 KB - mhealth_v14i1e56497_app3.pdf]

References

1. Bellettiere J, LaMonte MJ, Evenson KR, et al. Sedentary behavior and cardiovascular disease in older women. Circulation 2019 Feb 19;139(8):1036-1046. [doi: 10.1161/CIRCULATIONAHA.118.035312]
2. Young DR, Hivert MF, Alhassan S, et al. Sedentary behavior and cardiovascular morbidity and mortality: a science advisory from the American Heart Association. Circulation 2016 Sep 27;134(13):e262-e279. [doi: 10.1161/CIR.0000000000000440] [Medline: 27528691]
3. Bergouignan A, Latouche C, Heywood S, et al. Frequent interruptions of sedentary time modulates contraction- and insulin-stimulated glucose uptake pathways in muscle: ancillary analysis from randomized clinical trials. Sci Rep 2016 Aug 24;6:32044. [doi: 10.1038/srep32044] [Medline: 27554943]

XSL•FO
RenderX

4. Powell C, Herring MP, Dowd KP, Donnelly AE, Carson BP. The cross-sectional associations between objectively measured sedentary time and cardiometabolic health markers in adults - a systematic review with meta-analysis component. Obes Rev 2018 Mar;19(3):381-395. [doi: 10.1111/obr.12642] [Medline: 29178252]
5. Prince SA, Blanchard CM, Grace SL, Reid RD. Objectively-measured sedentary time and its association with markers of cardiometabolic health and fitness among cardiac rehabilitation graduates. Eur J Prev Cardiol 2016 May;23(8):818-825. [doi: 10.1177/2047487315617101] [Medline: 26607698]
6. Ekelund U, Steene-Johannessen J, Brown WJ, et al. Does physical activity attenuate, or even eliminate, the detrimental association of sitting time with mortality? A harmonised meta-analysis of data from more than 1 million men and women. Lancet 2016 Sep 24;388(10051):1302-1310. [doi: 10.1016/S0140-6736(16)30370-1] [Medline: 27475271]
7. Prince SA, Reed JL, Cotie LM, Harris J, Pipe AL, Reid RD. Results of the Sedentary Intervention Trial in Cardiac Rehabilitation (SIT-CR Study): a pilot randomized controlled trial. Int J Cardiol 2018 Oct 15;269:317-324. [doi: 10.1016/j.ijcard.2018.07.082] [Medline: 30072156]
8. Evenson KR, Butler EN, Rosamond WD. Prevalence of physical activity and sedentary behavior among adults with cardiovascular disease in the United States. J Cardiopulm Rehabil Prev 2014;34(6):406-419. [doi: 10.1097/HCR.0000000000000064] [Medline: 25068393]
9. Biswas A, Oh PI, Faulkner GE, Alter DA. A prospective study examining the influence of cardiac rehabilitation on the sedentary time of highly sedentary, physically inactive patients. Ann Phys Rehabil Med 2018 Jul;61(4):207-214. [doi: 10.1016/j.rehab.2017.06.003] [Medline: 28923367]
10. Bouchard C, Blair SN, Katzmarzyk PT. Less sitting, more physical activity, or higher fitness? Mayo Clin Proc 2015 Nov;90(11):1533-1540. [doi: 10.1016/j.mayocp.2015.08.005] [Medline: 26422244]
11. Brazeau AS, Rabasa-Lhoret R, Strychar I, Mircescu H. Barriers to physical activity among patients with type 1 diabetes. Diabetes Care 2008 Nov;31(11):2108-2109. [doi: 10.2337/dc08-0720] [Medline: 18689694]
12. Chastin SFM, Egerton T, Leask C, Stamatakis E. Meta-analysis of the relationship between breaks in sedentary behavior and cardiometabolic health. Obesity (Silver Spring) 2015 Sep;23(9):1800-1810. [doi: 10.1002/oby.21180] [Medline: 26308477]
13. Dempsey PC, Larsen RN, Dunstan DW, Owen N, Kingwell BA. Sitting less and moving more: implications for hypertension. Hypertension 2018 Nov;72(5):1037-1046. [doi: 10.1161/HYPERTENSIONAHA.118.11190] [Medline: 30354827]
14. Dunstan DW, Kingwell BA, Larsen R, et al. Breaking up prolonged sitting reduces postprandial glucose and insulin responses. Diabetes Care 2012 May;35(5):976-983. [doi: 10.2337/dc11-1931] [Medline: 22374636]
15. Ezeugwu VE, Manns PJ. The feasibility and longitudinal effects of a home-based sedentary behavior change intervention after stroke. Arch Phys Med Rehabil 2018 Dec;99(12):2540-2547. [doi: 10.1016/j.apmr.2018.06.014] [Medline: 29981314]
16. Henson J, Davies MJ, Bodicoat DH, et al. Breaking up prolonged sitting with standing or walking attenuates the postprandial metabolic response in postmenopausal women: a randomized acute study. Diabetes Care 2016 Jan;39(1):130-138. [doi: 10.2337/dc15-1240] [Medline: 26628415]
17. Lynch BM, Nguyen NH, Moore MM, et al. A randomized controlled trial of a wearable technology-based intervention for increasing moderate to vigorous physical activity and reducing sedentary behavior in breast cancer survivors: the ACTIVATE trial. Cancer 2019 Aug 15;125(16):2846-2855. [doi: 10.1002/cncr.32143] [Medline: 31012970]
18. Park C, Larsen B, Mogos M, et al. A multiple technology-based and individually-tailored Sit Less program for people with cardiovascular disease: a randomized controlled trial study protocol. PLoS ONE 2024;19(5):e0302582. [doi: 10.1371/journal.pone.0302582] [Medline: 38722831]
19. Viechtbauer W, Smits L, Kotz D, et al. A simple formula for the calculation of sample size in pilot studies. J Clin Epidemiol 2015 Nov;68(11):1375-1379. [doi: 10.1016/j.jclinepi.2015.04.014] [Medline: 26146089]
20. Lally P, van Jaarsveld CHM, Potts HWW, Wardle J. How are habits formed: modelling habit formation in the real world. Euro J Social Psych 2010 Oct;40(6):998-1009. [doi: 10.1002/ejsp.674]
21. Yamanouchi K, Shinozaki T, Chikada K, et al. Daily walking combined with diet therapy is a useful means for obese NIDDM patients not only to reduce body weight but also to improve insulin sensitivity. Diabetes Care 1995 Jun;18(6):775-778. [doi: 10.2337/diacare.18.6.775] [Medline: 7555502]
22. Matei R, Thuné-Boyle I, Hamer M, et al. Acceptability of a theory-based sedentary behaviour reduction intervention for older adults ('On Your Feet to Earn Your Seat'). BMC Public Health 2015 Jul 2;15(1):606. [doi: 10.1186/s12889-015-1921-0] [Medline: 26135402]
23. White I, Smith L, Aggio D, et al. On Your Feet to Earn Your Seat: pilot RCT of a theory-based sedentary behaviour reduction intervention for older adults. Pilot Feasibility Stud 2017;3(1):23. [doi: 10.1186/s40814-017-0139-6] [Medline: 28491459]
24. Gardner B, Smith L, Lorencatto F, Hamer M, Biddle SJH. How to reduce sitting time? A review of behaviour change strategies used in sedentary behaviour reduction interventions among adults. Health Psychol Rev 2016;10(1):89-112. [doi: 10.1080/17437199.2015.1082146] [Medline: 26315814]
25. Answers by heart fact sheets: lifestyle and risk reduction. American Heart Association. 2020. URL: https://www.heart.org/en/health-topics/consumer-healthcare/answers-by-heart-fact-sheets/answers-by-heart-fact-sheets-lifestyle-and-risk-reduction [accessed 2022-08-01]

XSL•FO
RenderX

26. Brickwood KJ, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819. [doi: 10.2196/11819] [Medline: 30977740]
27. Lyons EJ, Swartz MC, Lewis ZH, Martinez E, Jennings K. Feasibility and acceptability of a wearable technology physical activity intervention with telephone counseling for mid-aged and older adults: a randomized controlled pilot trial. JMIR Mhealth Uhealth 2017 Mar 6;5(3):e28. [doi: 10.2196/mhealth.6967] [Medline: 28264796]
28. Burner E, Zhang M, Terp S, et al. Feasibility and acceptability of a text message-based intervention to reduce overuse of alcohol in emergency department patients: controlled proof-of-concept trial. JMIR Mhealth Uhealth 2020 Jun 4;8(6):e17557. [doi: 10.2196/17557] [Medline: 32496203]
29. O'Brien MW, Wu Y, Petterson JL, Bray NW, Kimmerly DS. Validity of the ActivPAL monitor to distinguish postures: a systematic review. Gait Posture 2022 May;94:107-113. [doi: 10.1016/j.gaitpost.2022.03.002]
30. Edwardson CL, Winkler EAH, Bodicoat DH, et al. Considerations when using the activPAL monitor in field-based research with adult populations. J Sport Health Sci 2017 Jun;6(2):162-178. [doi: 10.1016/j.jshs.2016.02.002] [Medline: 30356601]
31. Carlson JA, Tuz-Zahra F, Bellettiere J, et al. Validity of two awake wear-time classification algorithms for activPAL in youth, adults, and older adults. J Meas Phys Behav 2021;4(2):151-162. [doi: 10.1123/jmpb.2020-0045] [Medline: 34447927]
32. American Diabetes Association Professional Practice Committee. 6. Glycemic targets: Standards of Medical Care in Diabetes—2022. Diabetes Care 2022 Jan 1;45(Supplement_1):S83-S96. [doi: 10.2337/dc22-S006]
33. Kapur S, Kapur S, Zava D. Cardiometabolic risk factors assessed by a finger stick dried blood spot method. J Diabetes Sci Technol 2008 Mar;2(2):236-241. [doi: 10.1177/193229680800200210] [Medline: 19885348]
34. Kapur S, Groves MN, Zava DT, Kapur S. Postprandial insulin and triglycerides after different breakfast meal challenges: use of finger stick capillary dried blood spots to study postprandial dysmetabolism. J Diabetes Sci Technol 2010 Mar 1;4(2):236-243. [doi: 10.1177/193229681000400202] [Medline: 20307382]
35. Adams MM. On our feet: feasibility trial of an intervention to reduce sedentary behavior and increase physical activity. : The University of North Carolina at Greensboro; 2012.
36. Verplanken B, Orbell S. Reflections on past behavior: a self‐report index of habit strength 1. J Applied Social Pyschol 2003 Jun;33(6):1313-1330. [doi: 10.1111/j.1559-1816.2003.tb01951.x]
37. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
38. Segal-Isaacson CJ, Wylie-Rosett J, Gans KM. Validation of a short dietary assessment questionnaire: the Rapid Eating and Activity Assessment for Participants short version (REAP-S). Diabetes Educ 2004;30(5):774-781. [doi: 10.1177/014572170403000512] [Medline: 15510530]
39. Freene N, McManus M, Mair T, Tan R, Davey R. High sedentary behaviour and low physical activity levels at 12 months after cardiac rehabilitation: a prospective cohort study. Ann Phys Rehabil Med 2020 Jan;63(1):53-58. [doi: 10.1016/j.rehab.2019.07.008] [Medline: 31465863]
40. Diaz KM, Howard VJ, Hutto B, et al. Patterns of sedentary behavior in US middle-age and older adults: the REGARDS study. Med Sci Sports Exerc 2016 Mar;48(3):430-438. [doi: 10.1249/MSS.0000000000000792] [Medline: 26460633]
41. Bellettiere J, Winkler EAH, Chastin SFM, et al. Associations of sitting accumulation patterns with cardio-metabolic risk biomarkers in Australian adults. PLoS ONE 2017;12(6):e0180119. [doi: 10.1371/journal.pone.0180119] [Medline: 28662164]
42. Ortega-Gómez E, Vicente-Galindo P, Martín-Rodero H, Galindo-Villardón P. Detection of response shift in health-related quality of life studies: a systematic review. Health Qual Life Outcomes 2022 Feb 5;20(1):20. [doi: 10.1186/s12955-022-01926-w] [Medline: 35123496]

Abbreviations

CGM: contiguous glucose monitor
CONSORT: Consolidated Standards of Reporting Trials
CVD: cardiovascular disease
HbA_{1c}: hemoglobin A_{1c}
Hs-CRP: high-sensitivity C-reactive protein
PHQ-9: Patient Health Questionnaire-9
T2D: type 2 diabetes
TTT: Ten Top Tips

XSL•FO
RenderX

Edited by Z Ni; submitted 07.Nov.2025; peer-reviewed by J Zhang, R Chen; revised version received 27.May.2026; accepted 10.Jun.2026; published 25.Aug.2026.

Please cite as:
Park C, Dietrich MS, Larsen B, Mayberry LS, Doyle A, Ahn S, Jean J, Maquiling K, Mogos MF, Muchira JM, Mulvaney S
Feasibility and Preliminary Efficacy of a "Sit Less" Program Leveraging Fitbit Tracking and Tailored Text Messages in Cardiometabolic Disease: Findings From 2 Parallel Randomized Controlled Trials in Coronary Artery Disease and Type 2 Diabetes
JMIR Mhealth Uhealth 2026;14:e56497
URL: https://mhealth.jmir.org/2026/1/e56497
doi:10.2196/56497

© Chorong Park, Mary S Dietrich, Britta Larsen, Lindsay S Mayberry, Abigail Doyle, Soojung Ahn, Jason Jean, Kevin Maquiling, Mulubrhan F Mogos, James M Muchira, Shelagh Mulvaney. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 25.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Immediate and Sustained Improvements in Mood and Stress Associated With Yuna, an AI-Powered Digital Mental Health Intervention: Real-World Retrospective Study

Kelsey McAlister[1*], PhD; Courtney Jewell[1*], PhD; Chad Stecher[2], PhD; Jennifer Huberty[1], PhD

[1]Fit Minded, Inc., 2901 E Greenway Rd PO Box 30271, Phoenix, AZ, United States

[2]College of Health Solutions, Arizona State University, Phoenix, AZ, United States

[*]these authors contributed equally

Corresponding Author:
Kelsey McAlister, PhD
Fit Minded, Inc., 2901 E Greenway Rd PO Box 30271, Phoenix, AZ, United States

Abstract

Background: AI-powered digital mental health interventions (DMHIs) are a promising approach to address barriers to traditional mental health care. However, real-world evidence of their immediate and sustained benefits remains limited.

Objective: The purpose of this real-world, retrospective study is to explore patterns of perceived mood and stress change associated with the use of Yuna, an AI-powered DMHI. We aimed to (1) describe user demographics and session characteristics, (2) quantify the magnitude of mood and stress change within sessions and over time with continued Yuna use, and (3) identify session-level factors associated with changes in mood and stress.

Methods: Adult Yuna users (aged ≥18 y) who initiated at least one session with the Yuna app were included in this study. Users self-reported mood and stress on Visual Analog Scales (VASs; range 0 - 1) before and after sessions. Linear mixed effects models were used to explore the immediate, within-session improvements in mood and stress, and the sustained, between-session changes in symptoms of mood and stress. Linear mixed effects models were also used to examine session-level predictors of within-session improvements, including baseline symptom severity, session duration, total number of unique therapeutic approaches used, safety guardrail activation, and gender.

Results: A total of 5549 real-world users were included (2901/5549, 52.3% female; mean sessions 3.44, SD 9.83). Users demonstrated significant, immediate within-session improvements in both mood (d=0.55) and stress (d=0.56), with sensitivity analyses yielding consistent results. Between-session analyses revealed gradual improvements in baseline mood (d=−0.009) and stress (d=−0.011). Baseline symptom severity was the strongest predictor of immediate, within-session change (mood: β=.118, SE 0.003; P<.001; stress: β=.131, SE 0.004; P<.001), followed by session duration (mood: β=.027, SE 0.003; P=.003; stress: β=.029, SE 0.003; P<.001). A greater number of unique therapeutic approaches used was associated with smaller improvements in both outcomes (mood: β=−0.007, SE 0.003; P=.043; stress: β=-0.011, SE 0.003; P=.002).

Conclusions: Use of Yuna, an AI-powered DMHI, was associated with perceived within-session improvements in mood and stress, with preliminary evidence of gradual improvements in mood and stress across repeated sessions. However, the absence of a control group and potential selection bias preclude causal conclusions. These findings offer promising, real-world evidence for AI-powered DMHIs as accessible, on-demand support tools. Prospective, controlled designs are needed to establish causal effects and evaluate the sustainability of observed improvements.

(JMIR Mhealth Uhealth 2026;14:e94070) doi:10.2196/94070

KEYWORDS

mental health; digital mental health intervention; real-world evidence; conversational AI; mental health app

Introduction

Low mood (ie, persistent sadness or reduced emotional well-being) and chronic stress represent a major and growing public health challenge worldwide. Depression, a common form of persistent low mood, affects hundreds of millions of individuals globally [1], and emotional stress has steadily worsened across countries and employment groups over the past decade [2]. Together, mood and stress-related concerns represent a major source of disability worldwide, underscoring their substantial individual and societal burden [3]. These concerns are particularly salient in the workplace, where elevated stress, low mood, and emotional exhaustion are linked to diminished productivity, increased absenteeism, and impaired well-being [4-6]. As employers increasingly recognize the

XSL•FO
RenderX

impact of both mood and stress on workforce functioning, there is a growing demand for scalable, effective approaches to support employee mental health.

Despite heightened attention, access to timely and effective mental health care remains limited for many working adults. Traditional treatments such as psychotherapy and pharmacological care are often constrained by cost, provider shortages, scheduling barriers, and stigma [7,8]. Employee Assistance Programs (EAPs) are commonly offered as an initial workplace resource, yet usage remains persistently low, in part due to limited awareness, perceived stigma, and short-term or reactive models of care [9,10]. As a result, many employees who experience significant stress or mood symptoms do not receive sustained support through existing workplace systems, highlighting the need for interventions that are both accessible and engaging.

In response to barriers in traditional workplace mental health care, digital mental health interventions (DMHIs) have been increasingly adopted by employers as scalable tools to support employee well-being. These programs typically offer app-based cognitive behavioral strategies, mindfulness content, and stress-management resources [11]. While workplace DMHIs can produce modest improvements in mental health outcomes, real-world effectiveness is often limited by low engagement and high dropout [12-14]. Systematic reviews suggest that DMHIs achieve stronger outcomes and higher adherence when programs include guidance or reminders, raising concerns about the long-term sufficiency of purely self-guided formats and highlighting the potential value of more interactive support beyond self-help tools alone [13,15]. At the same time, stigma, privacy concerns, resource constraints, and scheduling barriers continue to limit uptake of human-delivered services in workplace settings [16-18]. Together, these limitations highlight the need for workplace mental health tools that provide timely, personalized support without requiring traditional appointments.

AI-powered DMHIs represent an emerging approach that may help address these challenges by offering conversational, adaptive support that bridges the gap between unguided self-help apps and traditional human care [19,20]. Conversational AI systems can provide real-time emotional check-ins, coping guidance, and reflective coaching in a format that may feel more engaging and responsive than static digital content, while also reducing barriers related to stigma, scheduling, and discomfort with human interaction [21,22]. One example is Yuna, an AI-powered DMHI, which provides employees with 24/7 access to an AI mental health coach through personalized, on-demand conversations. Yuna has been used by thousands of individuals through workplace partnerships, illustrating its potential as a scalable model for delivering evidence-based support. Despite rapid growth in the availability and adoption of AI-driven mental health solutions, empirical evidence remains limited regarding how these platforms are used in real-world workplace populations and whether employees experience meaningful improvements in mood and stress over repeated engagement.

The purpose of this real-world, retrospective study is to explore the association between use of Yuna, an AI-powered DMHI, and perceived mood and stress change. We aimed to (1) describe user demographics and session characteristics, (2) quantify the magnitude of mood and stress change within sessions and over time with continued Yuna use, and (3) identify session-level factors associated with changes in mood and stress. These findings provide real-world evidence on conversational AI mental health support in workplace settings and help inform the development of scalable AI-powered DMHIs for employees.

Methods

Study Design and Participants

This retrospective study analyzed secondary data from Yuna, an AI-powered DMHI. The dataset included self-reported mood and stress and app usage metrics from real-world Yuna users, collected as part of routine product evaluation activities in the Yuna platform. Users were included if they: (1) were adults (aged 18+ y), (2) were located in a country where Yuna is available, and (3) initiated at least one session with the Yuna platform. Eligible participants were identified through a retrospective cohort sampling approach from backend app usage records, spanning from December 29, 2024, to January 12, 2026 (ie, the date the data was extracted). As this study is a retrospective study, no direct recruitment took place. A formal power analysis was not conducted, as the retrospective nature of this study precluded prospective sample size determination. Instead, the analytic sample reflected the full population of eligible users available within the observation window. This sample size was sufficient to support accurate and stable estimation of multilevel model parameters [23]. This study was conducted and reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for observational cohort studies (Checklist 1).

Ethical Considerations

The study protocol was reviewed and approved by the Solutions Institutional Review Board (IRB; study ID: 1001) under international expedited review procedures (ie, the study was not deemed exempt) and was conducted in accordance with US human subjects research regulations. The study involved a retrospective analysis of preexisting, deidentified data, with no direct interaction or intervention with users. Investigators did not have access to identifiable private information, and informed consent was waived by the IRB due to minimal risk. User data were collected during routine app use, and users agreed to the platform's terms and privacy policy, which permit the use of deidentified and/or aggregated data for service improvement and research purposes. The study used data from a global user population, processed in accordance with the platform's privacy policy. Participants were not compensated in any way for using the app or completing in-app questions or surveys. All data were deidentified prior to analysis.

App Description and Measures: Yuna App

Yuna is a commercially available mental health and wellness platform offered through workplace partnerships. Yuna's primary feature is a clinician-informed and evidence-based AI-powered mental health coach that provides users with unlimited, 24/7 support to improve well-being. Yuna

XSL•FO
RenderX

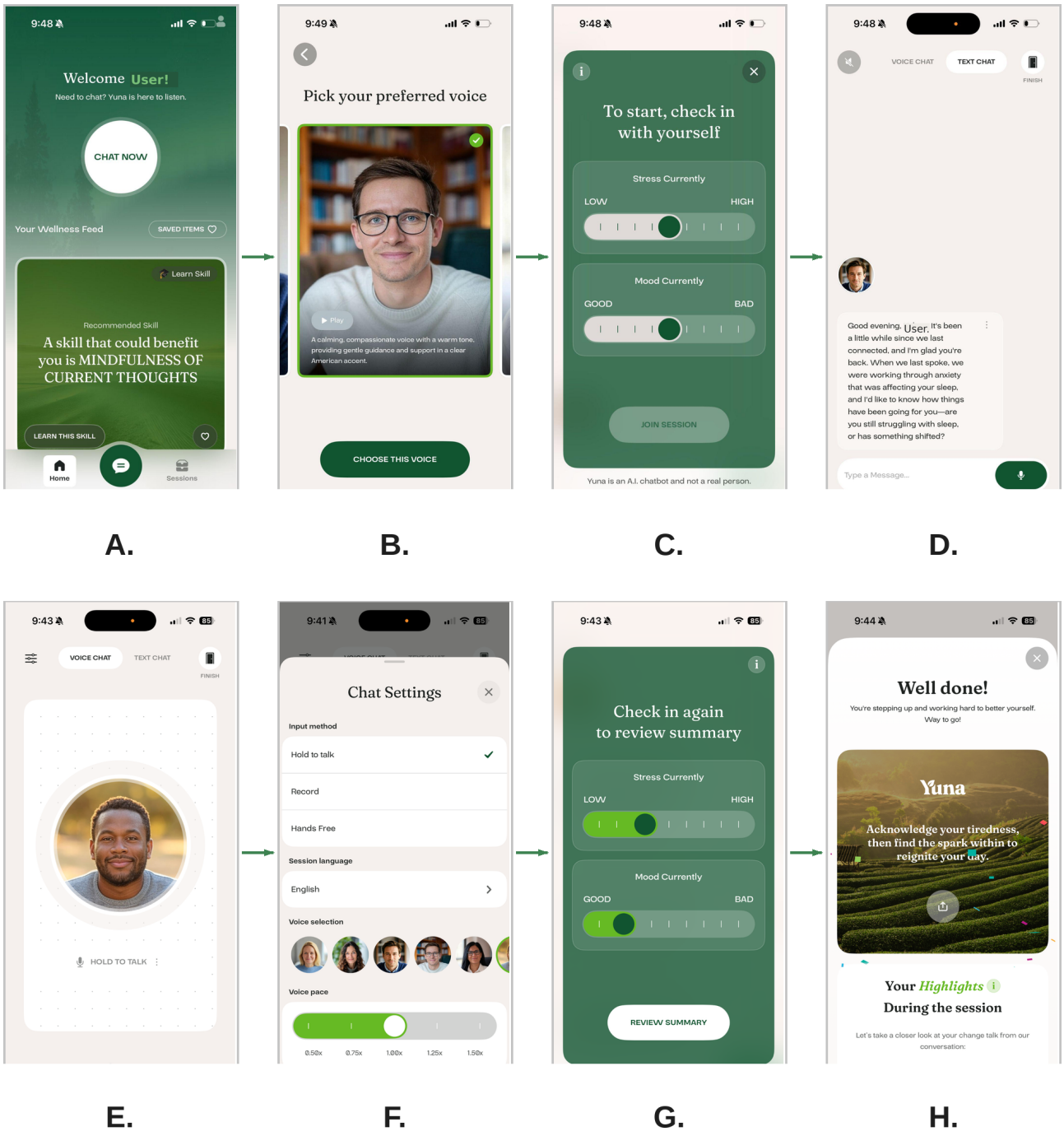
incorporates personalization features, such as selecting their preferred voice for the AI mental health coach, to foster a comfortable and individualized interaction. Additional features of Yuna include a library of wellness content, guided tools, and self-help resources. Together, Yuna is an emerging, scalable solution that offers more immediate and accessible support than traditional therapy services.

Yuna's AI mental health coach draws upon evidence-backed therapeutic frameworks, such as dialectical behavior therapy (DBT), cognitive behavioral therapy (CBT), and acceptance and commitment therapy (ACT). During conversations, the system dynamically applies techniques from these frameworks based on the user's expressed needs. These techniques are implemented as "therapeutic approaches," defined as discrete intervention strategies (eg, mindfulness exercises, emotion regulation skills, and cognitive reframing). Yuna leverages over 20 such therapeutic approaches, which are informed by clinician-developed prompts. Therapeutic approach selection follows a dual-layer logic. A central controller model analyzes user input, conversational context, and emotional tone to determine which approach is most appropriate for the user's presenting needs. Once activated, intraapproach logic guides the delivery of the specific approach sequence to ensure protocol fidelity. All therapeutic approach prompts and clinical protocols were developed by a licensed clinician, and the system undergoes iterative refinement through ongoing clinician review.

High-priority safety therapeutic approaches, including suicide risk-specific strategies and safety guardrails, continuously monitor conversations for high-risk content. Upon activation, standard coaching is suspended and users are provided with crisis deescalation support and emergency resources (eg, 988). To mitigate potential bias in AI-generated responses, Yuna's clinical team periodically audits deidentified session transcripts to identify and address inconsistencies in approach behavior, inappropriate responses, or gaps in protocol adherence. System performance is additionally monitored through automated quality metrics and a human-in-the-loop review process. Figure 1 depicts a standard Yuna user flow.

XSL•FO
RenderX

Figure 1. Yuna app user flow across a typical session. Note. (A) home screen with "chat now" entry point. (B) AI coach voice selection during onboarding. (C) Presession mood and stress ratings via sliding scales. (D) In-session text chat interface. (E) In-session voice interaction interface. (F) Chat settings (eg, input mode, language, voice, and pace). (G) Postsession mood and stress ratings. (H) Postsession summary with reflection and highlights.



Study Measures

Overview

When a user first signs up for Yuna, they are able to provide their age and gender within their account profile. Gender was assessed with the item "What's your gender?" with response options of Male, Female, or Other. User country is derived from backend data.

Mood and Stress Levels

In the Yuna app, users are prompted to self-report their current mood and stress levels on a Visual Analog Scale (VAS) immediately before and after a session, defined as a discrete interaction with the AI mental health coach that begins when a user initiates a conversation and ends when the user selects "finish" in the conversation. For stress, users move a slider along a continuum from "Low" to "High," whereas mood is rated by sliding along a scale from "Good" to "Bad." The slider-based scale is continuous and does not display numerical values to users. Responses are recorded on a 0 - 1 continuous scale to the hundredth decimal place, with 10 tick marks serving as visual anchors to facilitate user responding. This measurement approach allows users to indicate their subjective experiences without being restricted to specific numerical values. Mood and

XSL•FO
RenderX

stress ratings before sessions were required, while postsession ratings were optional. This prepost assessment approach is consistent with measurement strategies used in other DMHIs [24].

Single-item VASs minimize user burden and support repeated, in-the-moment assessments within a real-world digital intervention context, and have been used widely in digital health research [25]. However, the measures examined in this study have not been psychometrically validated. These metrics are intended to capture subjective, perceived changes in mood and stress rather than capture clinically validated outcomes.

App Engagement

Engagement with the Yuna app was described by users' time with Yuna, total sessions, and session length. Time with Yuna was defined as the total number of days from a user's first initiated Yuna session to the last recorded Yuna session, representing the user's total period of possible exposure to the app, whether or not they were actively engaged daily or monthly. Sessions per user were defined as the total number of sessions each user had completed at the time of data extraction. Session length was defined as the average total minutes each session lasted.

Statistical Analyses

User demographics and session characteristics were characterized using descriptives, including mean and SD for continuous variables and total count and percentages for categorical variables. Mood and stress scores were scored continuously on a scale from 0 to 1, with higher values representing worse mood or higher stress. Data availability differed by research aim, resulting in separate analytic samples and models for within-session and between-session analyses. Within-session analyses required sessions with complete pre- and postsession ratings of mood and stress. Between-session analyses required users with repeated sessions with presession ratings of mood and stress. Because of this, sample sizes vary across models and are reported with each analysis. This approach aligns with best practices in observational, real-world studies, where complete data across all sessions cannot be assumed [26,27]. To characterize missingness and potential selection effects, missingness was quantified for key variables, and engagement was described using median and IQR for sessions per user. Users included in within-session analyses (ie, users with ≥1 session complete with pre-postsession data) were compared to excluded users. Sessions with complete postsession data were compared to sessions without postsession data. To formally evaluate missingness, a logistic regression model was conducted to predict postsession rating completion from presession mood, stress, gender, repeat-user status, and session number.

To examine within-session improvements in mood and stress, we fit separate linear mixed-effects models (LMMs) predicting mood improvement and stress improvement with a random intercept for the user to account for within-user correlation across repeated sessions. The outcome for the within-session models was a prepost change score computed within each session (ie, presession rating minus postsession rating). From this coding, positive values indicated improvement in mood or stress, whereas negative values reflect worsening mood or stress. To evaluate the robustness of within-session findings, two sensitivity analyses were conducted using the same LMM specifications as the primary model. One sensitivity model restricted the analytic sample to each user's first completed session to assess whether results were driven by repeat users. The second sensitivity model restricted the analytic sample to users with at least two sessions to evaluate whether findings were driven by one-time or minimally engaged users. To examine between-session improvements in mood and stress, separate LMMs were fit to model presession mood and presession stress over time among users with more than one session. To test whether users began subsequent sessions with improved mood or reduced stress, the outcome was baseline ratings at the start of each session. The fixed effect was the session number, defined as the within-user order of sessions, and the random intercept was the user. In these models, negative coefficients for the session number indicated an improvement, meaning that baseline mood or stress was improved at later sessions relative to earlier sessions. Effect sizes for within-session and between-session changes were calculated as standardized mean differences.

To identify session-level factors that were associated with within-session improvements in mood and stress, we fit fully adjusted LMMs with fixed effects for presession mood or stress, session length, safety guardrail activation, total number of unique therapeutic approaches used (ie, count of distinct therapeutic approaches deployed), and gender, and a random intercept for user to account for correlation across repeated sessions within individuals. Continuous predictors were standardized (z-scored) before modeling to facilitate comparability of coefficient sizes. To contextualize findings related to therapeutic approach use, we conducted an exploratory analysis of the three most frequently used therapeutic approaches: the default module (ie, a general supportive agent that guides the conversation and transitions to specialized agents when deeper support is needed), cognitive defusion (ie, an ACT-based technique that helps users disengage from distressing thoughts), and interpersonal effectiveness (eg, a DBT-based approach focused on assertive communication and boundary setting). Session-level indicators were created for each of these approaches and were included as fixed effects in LMMs predicting whether session improvements in mood and stress.

All models were estimated using restricted maximum likelihood (REML) and were adjusted for gender. *P* values were obtained using Satterthwaite approximation. Given the large number of sessions and users available across all analytic samples, the study was considered adequately powered to detect associations of the magnitude observed. All analyses were conducted in R Version 2024.12.0+4.67 (R Core Team) [28].

Bias Mitigation

Fit Minded, Inc, uses a standardized bias mitigation protocol across all industry-partnered research, applied prospectively and independent of study outcomes. For this study, the following safeguards were implemented. First, an independent academic researcher (CS), external to the commercial relationship with

XSL•FO
RenderX

Yuna, conducted a manuscript-level review encompassing the analysis plan, statistical approach, reported results, and interpretation. This review did not include direct access to the raw dataset or independent reexecution of the statistical models. CS received no financial compensation from Fit Minded, Inc. or Yuna in connection with this study. Second, missingness due to optional pre- and postsession mood and stress ratings was addressed by restricting each analytic model to sessions with complete data for the relevant outcomes and predictors, and additional sensitivity analyses were used to explore how different subsamples may be driving the estimated within-session changes in mood and stress.

Results

User Demographics and Session Characteristics

The full data set included 5549 Yuna users (Table 1). Users were predominantly female and primarily located in the United States, with additional representation from more than 100 countries. Users completed 3.44 sessions on average, though this distribution was highly right-skewed (median 1, IQR 1 - 2). The mean span between first and last session (ie, time with Yuna) was 29.6 days (median 0 d, IQR 0 - 15.9). Approximately 62% (3450/5549) of users completed only one session. Yuna users with ≥2 sessions presented with slightly lower baseline mood and stress scores, on average, compared to single-session users (Multimedia Appendix 1).

Postsession mood and stress ratings were missing for 75.3% (11,985/15,916) of sessions, and more engaged users completed more sessions than users who did not complete postsession ratings (mean 6.66, SD 15.87 vs mean 1.85, SD 3.51, respectively). Full missingness data and comparisons between users and sessions with and without complete data are reported in Multimedia Appendix 1.

XSL•FO
RenderX

Table . User-level demographics and session descriptives (N=5549 users).

Demographics	Values
Age (years), n (%)	
18 - 25	117 (2.1)
26 - 35	105 (1.9)
36 - 45	44 (0.8)
46 - 55	22 (0.4)
55+	10 (0.2)
Missing	5251 (94.6)
Sex	
Female	2901 (52.3)
Male	1996 (36.0)
Missing	652 (11.7)
Country	
United States	3386 (61.0)
India	309 (5.6)
United Kingdom	280 (5.0)
Canada	204 (3.7)
Germany	156 (2.8)
Australia	133 (2.4)
Other[a]	1021 (18.4)
Missing	60 (1.1)
Sessions per user, mean (SD)	3.44 (9.83)
Time with Yuna (days), mean (SD)	29.55 (65.15)
Session length (minutes), mean (SD)	20.38 (21.48)
Presession mood, mean (SD)	0.54 (0.22)
Postsession mood, mean (SD)	0.45 (0.23)
Presession stress, mean (SD)	0.54 (0.24)
Postsession stress, mean (SD)	0.45 (0.24)

[a]Users from 107 additional countries were collapsed into “Other.” Higher mood and stress values indicate worse mood or higher stress. Lower values indicate better mood or lower stress.

Within-Session Improvement of Mood and Stress

LMMs were used to assess the within-session change of mood and stress across participants, adjusted for gender. Yuna users demonstrated statistically significant within-session level improvements in both reported mood (d=0.55) and stress (d=0.56; both $P<.001$). Male users presented with a smaller magnitude of improvements compared to female users, though improvements were positive across groups (Table 2 and Figure 2).

To evaluate the robustness of the within-session findings, two sensitivity analyses were conducted. The first restricted the analytic sample to each user’s first completed session to assess whether primary results were driven by repeat users. The second restricted the samples to users with at least two sessions to assess whether findings were driven by one-time users. Results were consistent with the primary findings, indicating that primary within-session estimates were not driven by the subset of highly engaged users (Multimedia Appendix 1).

XSL•FO
RenderX

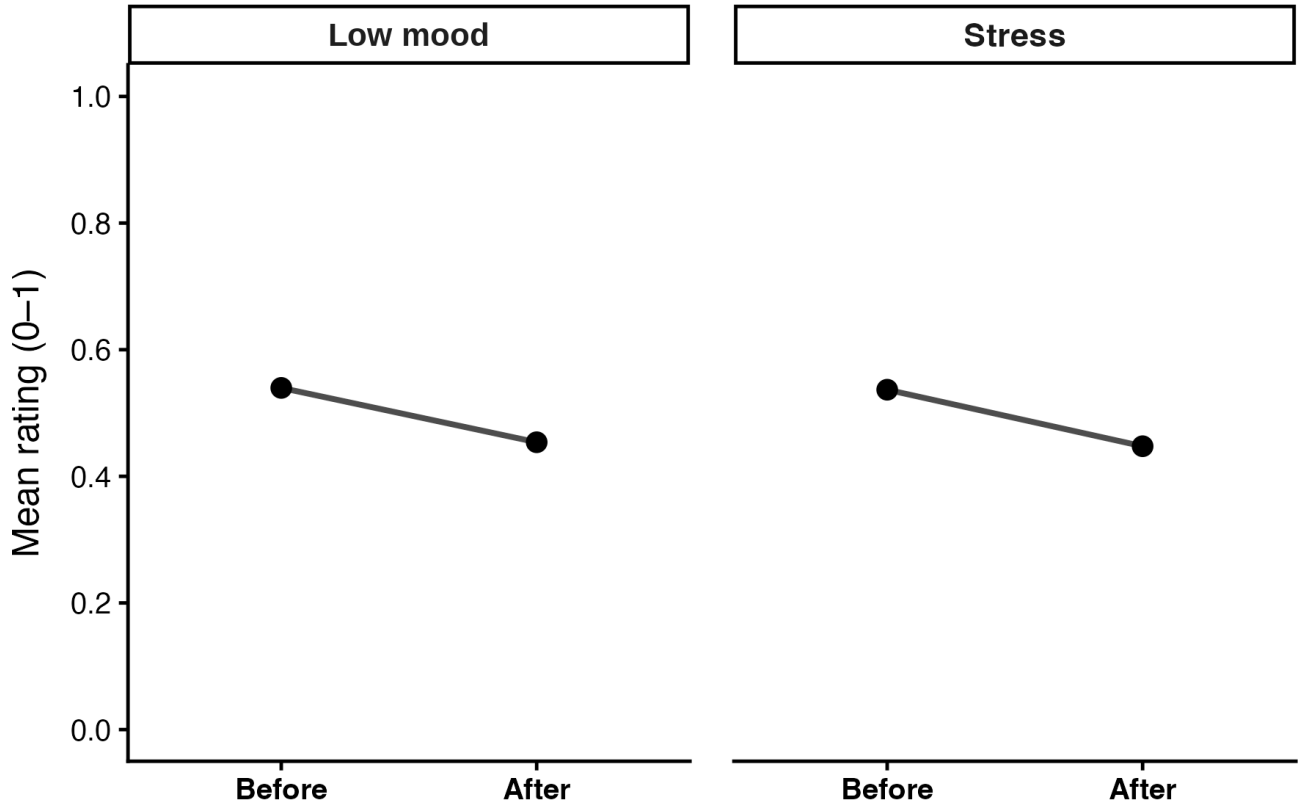
Table . Mixed-effects models of within-session improvement in mood and stress (n=3931 sessions; n=1570 users)[a].

	Mood				Stress			
	Estimate	SE	95% CI	*P* value	Estimate	SE	95% CI	*P* value
Fixed effects								
Intercept	0.100	0.006	0.09 to 0.11	<.001	0.100	0.007	0.09 to 0.11	<.001
Gender (male)	–0.040	0.009	–0.06 to –0.02	<.001	–0.030	0.010	–0.05 to –0.01	.006
Random effects								
User (intercept), SD	0.110	—[b]	—	—	0.140	—	—	—
Residual, SD	0.180	—	—	—	0.180	—	—	—

[a]Outcomes represented by within-session level improvement. Positive coefficients indicate improvement in mood or stress. Both models were adjusted for gender. Model diagnostics indicated no singular fits and low multicollinearity. Residual diagnostics suggested some nonnormality and heteroscedasticity, which is common in large observational datasets. For the mood model, adjusted intraclass correlation coefficient (ICC)=0.296, marginal R^2=0.009, and conditional R^2=0.302. For the stress model, adjusted ICC=0.357, marginal R^2=0.004, and conditional R^2=0.359, indicates not applicable.

[b]Not applicable.

Figure 2. Within-session changes in mood and stress. Mood and stress were assessed using a single-item visual analog scale ranging from 0 to 1, with higher values indicating lower mood and higher stress. Values represent mean prepost session ratings of mood and stress.



Between-Session Changes in Mood and Stress

LMMs were used to explore the changes in mood and stress over time. Baseline mood (β=−.00020, SE 0.00008; *P*=.011) and baseline stress (β=−.00029, SE 0.00008; *P*<.001) demonstrated statistically significant decreases across sessions (Table 3 and Figure 3), reflecting gradual improvements in mood (*d*=−0.009) and stress (*d*=−0.011). Gender was a significant predictor of baseline mood, with male users reporting lower presession mood compared to female users (β=−.030, SE 0.008; *P*<.001).

In Figure 3, baseline mood and stress were assessed using a single-item VAS ranging from 0 to 1, with higher values indicating lower mood and higher stress. Predicted baseline mood and stress ratings were estimated from LMMs. Shaded regions represent 95% CIs. The y-axis is truncated to improve the interpretability of between-session trajectories. The full scale ranges from 0 to 1. Truncation does not alter the statistical findings. The small magnitude of observed changes is acknowledged in the discussion.

XSL•FO
RenderX

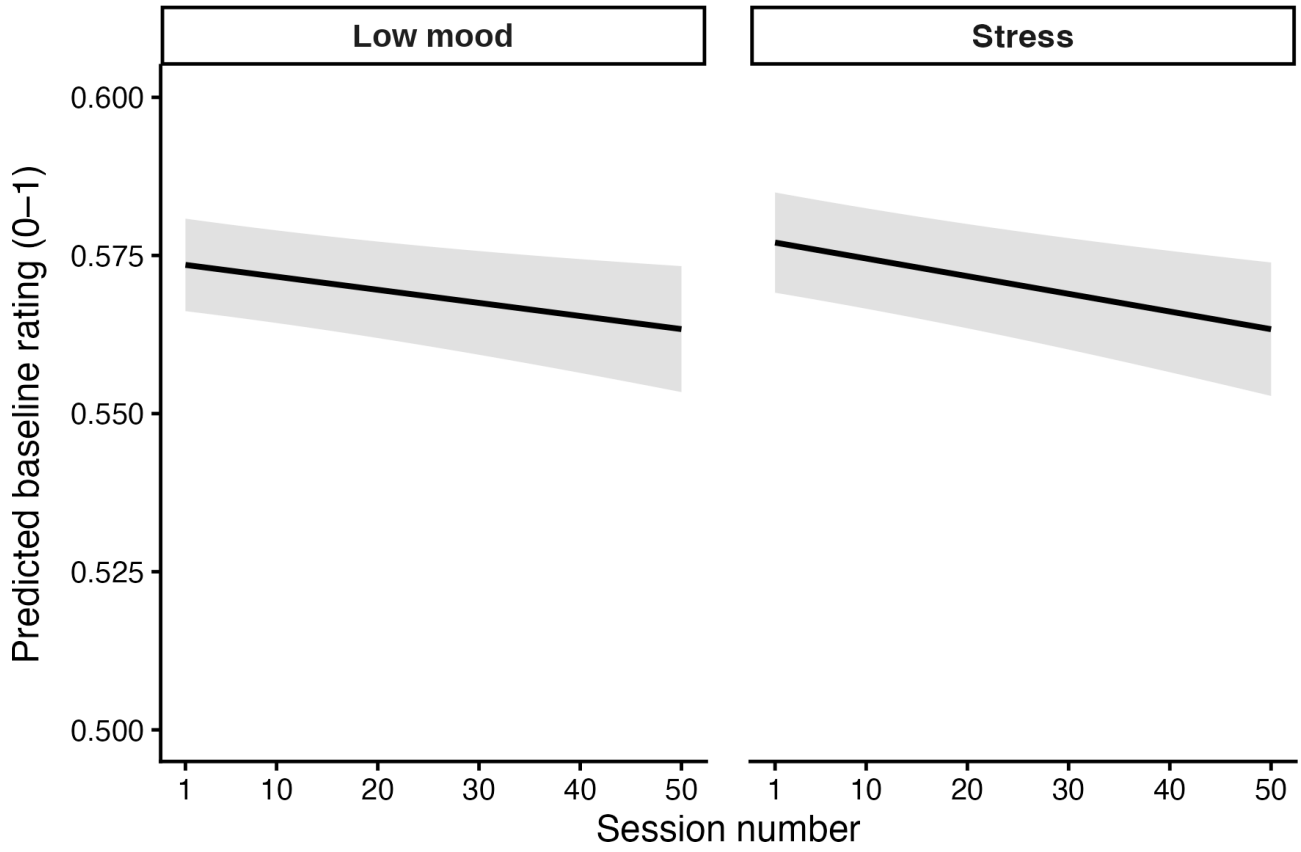
Table . Mixed-effects models of between-session change in baseline mood and stress (n=13,877 sessions; n=1838 users)[a].

	Baseline mood				Baseline stress			
	Estimate	SE	95% CI	*P* value	Estimate	SE	95% CI	*P* value
Fixed effects								
Intercept	0.590	0.005	0.58 to 0.60	<.001	0.590	0.006	0.58 to 0.60	<.001
Session number	–0.00020	0.00008	0.00 to 0.00	.01	–0.00029	0.00008	0.00 to 0.00	<.001
Gender (Male)	–0.030	0.008	–0.04 to –0.01	<.001	–0.020	0.009	–0.04 to 0.00	.03
Random effects								
User (intercept), SD	0.140	—[b]	—	—	0.160	—	—	—
Residual, SD	0.180	—	—	—	0.190	—	—	—

[a]Baseline mood and stress refer to presession ratings. Session number represents the order of sessions within each user. Negative coefficients indicate lower baseline distress at later sessions. Both models were adjusted for gender. Model diagnostics indicated no singular fits and low multicollinearity. Residual diagnostics suggested some nonnormality and heteroscedasticity. For the mood model, adjusted intraclass correlation coefficient (ICC)=0.378, marginal R^2=0.004, and conditional R^2=0.381. For the stress model, adjusted ICC=0.420, marginal R^2=0.003, and conditional R^2=0.422.

[b]Not applicable.

Figure 3. Baseline mood and stress across repeat sessions.



Session-Level Predictors of Mood and Stress

LMMs were fit to identify session-level factors that were associated with within-session improvements in mood and stress (Table 4). Higher baseline mood (β=.118, SE 0.003) and baseline stress (β=.131, SE 0.004) at the beginning of the session were significantly associated with greater within-session improvement for both outcomes. Longer sessions were significantly associated with greater improvements in mood (β=.027, SE 0.003) and stress (β=.029, SE 0.003), though the magnitude of these coefficients was smaller than the magnitude of baseline stress.

A greater number of unique therapeutic approaches used was associated with smaller within-session improvements in mood (β=−.007, SE 0.003; *P*=.043) and stress (β=−.011, SE 0.003; *P*=.002). Safety guardrail activation showed no significant

XSL•FO
RenderX

association with mood or stress improvements. Gender was not significantly associated with within-session improvement after controlling for other session-level factors.

To contextualize findings that a greater number of unique therapeutic approaches were associated with smaller within-session improvements, an exploratory analysis examined outcomes for sessions that included the three most frequently used therapeutic approaches: the default module (17909/42742, 41.9% of sessions), cognitive defusion (3826/42742, 9.0%), and interpersonal effectiveness (3288/42742, 7.7%). In adjusted mixed-effects models including indicators for these approaches, sessions that included the default module were associated with greater improvements in mood (β=.144, SE 0.040; *P*<.001). Cognitive defusion (β=.003, SE 0.007; *P*=.726) and interpersonal effectiveness (β=.008, SE 0.008; *P*=.310) did not reveal significant associations with mood. No therapeutic approach was significantly associated with stress improvements. (Multimedia Appendix 1).

Table . Session-level predictors of within-session improvements in mood and stress (n=2955 sessions; n=1087 users).[a]

	Mood				Stress			
	Estimate	SE	95% CI	*P* value	Estimate	SE	95% CI	*P* value
Fixed effects								
Intercept	0.106	0.006	0.09 to 0.11	.006	0.108	0.007	0.09 to 0.12	<.001
Baseline symptom severity	0.118	0.003	0.11 to 0.12	<.001	0.131	0.004	0.12 to 0.14	<.001
Session length	0.027	0.003	0.02 to 0.03	.003	0.029	0.003	0.02 to 0.03	<.001
Safety guardrail activation	0.018	0.018	–0.01 to 0.06	.305	0.032	0.018	0.00 to 0.07	.075
Total number of unique therapeutic approaches used	–0.007	0.003	–0.01 to –0.001	.043	–0.011	0.003	–0.02 to –0.01	.002
Gender (male)	–0.018	0.010	–0.04 to 0.00	.083	–0.008	0.011	–0.03 to 0.02	.470
Random effects								
User (intercept), SD	0.119	—[b]	—	—	0.136	—	—	—
Residual, SD	0.140	—	—	—	0.140	—	—	—

[a]Outcomes represented by within-session level improvement. Positive coefficients indicate improvement in mood or stress. Continuous predictors were standardized (z-scored). Both models were adjusted for gender. Model diagnostics indicated no singular fits and low multicollinearity (variance inflation factors approximately 1.0–2.0). Residual diagnostics suggested some nonnormality and heteroscedasticity, which is common in large observational datasets. For the mood model, adjusted intraclass correlation coefficient (ICC)=0.419, marginal R^2=0.302, and conditional R^2=0.594. For the stress model, adjusted ICC=0.483, marginal R^2=0.323, and conditional R^2=0.650. - Indicates not applicable.

[b]Not applicable.

Discussion

Principal Findings

The purpose of this real-world, retrospective study is to explore the association between use of Yuna, an AI-powered DMHI, and changes in perceived mood and stress. We aimed to: (1) describe Yuna user demographics and session characteristics, (2) quantify the magnitude of mood and stress change within sessions and over time with continued Yuna use, and (3) identify session-level factors associated with changes in mood and stress. Yuna users were predominantly from the United States (3386/5549, 61.0%) and female (2901/5549, 52.3%). On average, Yuna users spent approximately 30 days with Yuna, with an average of three total sessions. Yuna users showed statistically significant within-session improvements in both mood and stress. Within-session improvements were observed across gender groups, but male users demonstrated slightly smaller improvements compared to female users. Users also demonstrated gradual, between-session improvements associated with Yuna use, entering subsequent sessions with higher baseline mood and lower baseline stress compared to previous sessions. Baseline symptom severity of mood or stress and session duration were associated with greater within-session improvements in mood and stress, whereas the usage of more unique therapeutic approaches (ie, a greater number of distinct therapeutic approach types) was associated with smaller improvements. Safety guardrail activation did not negatively impact improvements in stress or mood, suggesting that safety mechanisms can operate without compromising perceived benefits.

XSL•FO
RenderX

Despite over half of users being from the United States (3386/5549, 61.0%), Yuna demonstrates notable international reach, with users spanning over 110 countries and six continents. This finding highlights the potential of AI-powered DMHIs (like Yuna) to expand access in low- and middle-income countries, where over 75% of individuals do not receive mental health treatment due to limited infrastructure [29]. On average, users engaged with the Yuna platform (ie, time with Yuna) for approximately 30 days, with an average of 3 sessions completed, but demonstrated considerable heterogeneity in their engagement (d with Yuna: mean 29.55, SD 65.15; total number of sessions: mean 3.44, SD 9.83). This suggests that Yuna may be suited to diverse user needs and engagement patterns. Unlike traditional interventions with fixed durations, Yuna's flexible format aligns with stepped care frameworks that emphasize as-needed support with the capability to modify intensity based on needs and preferences [30,31]. Future research is needed to explore how user engagement and reported outcomes may vary across diverse demographic profiles, especially given some of these tools have international reach [32].

Yuna was associated with within-session mood and stress improvements. Users demonstrated significant improvements in mood and stress following a single, 20-minute session (on average). Within-session effect sizes (d=0.55 for mood; d=0.56 for stress) are comparable in magnitude to those documented for single-session interventions delivered in controlled therapeutic environments (g=0.56 for anxiety; g=0.32 overall [33,34]). However, this comparison is illustrative of scale only and does not imply methodological equivalence. Effect sizes from unvalidated, single-item VAS scales cannot be directly compared to standardized mean differences from controlled trials using validated, normed instruments, where regression to the mean, demand characteristics, and selection bias are controlled. Specifically, widely used measures such as the Patient Health Questionnaire-9 and Generalized Anxiety Disorder-7 carry established psychometric properties and clinically defined thresholds for meaningful change that the single-item VAS measures used here do not. The observed within-session improvements demonstrate that Yuna may be associated with immediate benefits through brief, self-initiated conversations with an AI mental health coach, positioning AI-powered DMHIs as a promising option for mood and stress management. A couple of Yuna's features may explain the observed within-session changes. First, Yuna is an accessible, on-demand platform that connects users with support at moments of high distress and when a sense of support may be most salient [35,36]. Additionally, the judgment-free, user-directed nature of AI-powered DMHIs may influence deeper engagement with the platform's content. Although male users demonstrated smaller within-session changes compared to female users, they still presented with meaningful improvements. Future research should explore how stigma and gender norms may influence user engagement and outcomes in AI-powered DMHIs. Importantly, whether the within-session improvements observed here reflect transient distress relief or more durable benefit remains an open question. Users may have self-initiated sessions at moments of heightened distress, and the observed improvements may partly reflect natural mood fluctuation following peak distress rather than a lasting change. Future research incorporating follow-up assessments and validated outcome measures would help clarify whether session-level changes translate into meaningful improvements in real-world functioning such as productivity, sleep, or interpersonal relationships.

In addition to immediate symptom relief, AI-powered DMHIs like Yuna may also be associated with sustained, between-session benefits. Users who returned for more than one session showed gradual reductions in baseline stress and mood (ie, lower scores indicating better functioning) across subsequent sessions. However, this finding must be interpreted with considerable caution. The between-session sample represents 33% (1838/5549) of the full analytic sample, and the majority of users completed only a single session and are absent from this analysis. Users who returned are, by definition, a self-selected group most likely to have responded positively to initial use. The observed improvement trend is precisely what survivorship bias would produce, independent of any therapeutic mechanism. These gradual improvements may reflect, but do not confirm, Yuna's application of therapeutic approaches across sessions. Unlike many AI-generated mental health chatbots that focus on supportive listening or psychoeducation, Yuna leverages therapeutic approaches that are designed to teach evidence-backed skills (eg, cognitive reframing and emotion regulation) and transferable strategies to real-world contexts [37]. If these mechanisms are operative, AI-powered DMHIs may build on previous recommendations and reinforce skills over time, enabling a scaffolded learning experience for users, an approach consistent with prior work indicating that progressive skill-building and reinforcement can enhance skill acquisition [38,39]. However, these interpretations are speculative and are not directly supported by the current data, especially considering the VAS measures used are not validated, so the clinical significance of the observed reductions cannot be confirmed. Lastly, though the between-session coefficients reached significance, their magnitude was small, and the practical significance of these reductions remains uncertain. Future research should incorporate validated outcome measures into routine practice to expand upon these findings and allow for more robust comparisons with traditional intervention modalities.

Several session-level factors were associated with within-session improvements in mood and stress. Baseline symptom severity was the strongest predictor of within-session change: users who entered a session with high baseline stress or lower baseline mood showed significantly greater within-session improvements in both outcomes, suggesting that the association between Yuna use and symptom change is more pronounced when users are experiencing higher levels of distress. This finding aligns with common patterns observed across both traditional interventions and DMHIs [40,41]. However, it is important to acknowledge that baseline symptom severity as the strongest predictor of improvement is also consistent with regression to the mean (eg, [42]). Users may be more likely to initiate sessions during periods of heightened distress, and the observed within-session improvements may partly reflect natural fluctuation in mood and stress rather than a therapeutic response. This alternative

XSL•FO
RenderX

explanation cannot be ruled out in the absence of an appropriate control condition.

Longer session durations were positively associated with improvements in both mood and stress, consistent with prior work indicating that increased engagement time in AI-powered DMHIs can support symptom improvement [14]. The number of unique therapeutic approaches activated (ie, clinically-informed strategies and intervention techniques) negatively predicted within-session improvements. In alignment with prior work on intervention complexity, this finding suggests that leveraging a smaller set of focused therapeutic approaches may be more effective in improving baseline mood than integrating numerous therapeutic approaches that may reduce the clarity or impact of individual approaches [43]. Future research should explore and validate how AI-powered DMHIs, such as Yuna, combine, sequence, and select therapeutic approaches during sessions, and how these selections can influence symptom improvement.

Safety guardrail activation was not significantly associated with within-session improvements in mood or stress. This finding suggests that the activation of safety protocols does not detract from the perceived benefits of a session, addressing concerns in the literature about the potential trade-offs between safety mechanisms and engagement with AI-powered DMHIs [44,45]. Well-designed safety guardrails, such as those implemented by Yuna, may not be associated with diminished symptom improvement and may co-occur with stress reduction when discussing a sensitive topic. An important next step will be to qualitatively examine users' experiences of safety, trust, and disclosure in AI-powered DMHIs to understand how safety guardrail design may influence engagement and symptom improvement.

Strengths and Limitations

This study has multiple strengths. First, the findings were a combination of self-reported mood and stress measures and app usage metrics from real-world users of Yuna, a commercially available AI-powered DMHI. This approach captures naturalistic user experiences in real-world settings, offering critical insights into outcomes associated with AI-powered DMHI use beyond the controlled conditions of standard efficacy trials [46]. Second, analyses at both the single session and across session levels offered a comprehensive view of the immediate and sustained changes observed across Yuna use. These strengths support the value of real-world evidence for understanding the clinical impact of AI-powered DMHIs, like Yuna, and how these tools function across users and needs.

The study is not without limitations. First, this study does not include a control group, precluding the ability to make causal inferences about the impact of Yuna on mood and stress. Second, demographic data were incomplete, with only 5.4% (298/5549) of users reporting their age range and no collection of other critical variables that may influence response to AI-powered DMHIs, such as education level, socioeconomic status, race, or clinical history. In addition, although gender was assessed directly, the response options offered (Male, Female, and Other) did not follow current best practices for inclusive gender measurement, such as including specific nonbinary or transgender categories or using a two-step gender assessment; this may limit both the precision of gender-based analyses and comparability with studies using more inclusive measures. Though common to retrospective, real-world studies [47], the lack of complete demographic data has direct implications across three domains: (1) without demographic data, we cannot assess whether the analytic sample is representative of the broader population of working adults who might use AI-powered DMHIs; (2) this limits the generalizability of findings to specific age groups, cultural contexts, or clinical populations; and (3) systematic differences between users who provided demographic information and those who did not may introduce bias into observed associations; for example, users who complete profile fields may differ in motivation, digital literacy, or baseline symptom severity. Third, this study included a self-selected sample of Yuna users. Those who are voluntarily engaging in AI-powered DMHIs such as Yuna may differ systematically from the broader population due to differences in motivation or symptom severity. In addition, due to the voluntary nature of app use, data from nonusers for Yuna were not available, and direct comparison between Yuna users and nonusers could not be conducted. As a result, findings should be interpreted with caution and may not generalize to broader populations. Future research should include a prospective study with a larger, more representative sample to understand how effectiveness may vary across demographic groups and profiles, as previous work has reported differences in engagement patterns, therapeutic alliance, and trust with AI-powered DMHIs and AI mental health coaches (like Yuna [32,48]). Fourth, missing data represent a key limitation. Users included in the within-session analyses were more engaged, on average, than those excluded. This means that findings may not generalize to one-time users. While sensitivity analyses among subsamples with different engagement patterns yielded similar results, selection bias may still have influenced the findings. Fifth, the present study relied on a single-item VAS rather than validated outcome measures to assess mood and stress. Although this may influence the reliability and validity of the reported outcomes, it is important to note that these single-item measures are typically used in industry because they enable rapid, low-burden assessments and are more suited for repeated designs [49-51]. Because these single-item scales have not been psychometrically validated, and their reliability and validity for measuring mood and stress have not been formally established, it is not possible to determine whether the observed changes correspond to clinically meaningful improvements. For context, validated instruments such as the Patient Health Questionnaire-9 (PHQ-9) and Generalized Anxiety Disorder-7 Scale (GAD-7) use established thresholds for minimal clinically important differences (eg, a 5-point change on the PHQ-9). The single-item scales used in this study do not permit such comparisons, and findings should be interpreted as reflecting subjective, within-person perceived changes in mood and stress rather than clinically defined symptom change. Future research should integrate repeated validated outcome measures to more accurately capture perceptions of and changes in mood and stress. Future research would benefit from controlled designs to establish causal effects between AI-powered DMHIs relative to standard care. Lastly, the external review conducted by CS constituted a

XSL•FO
RenderX

manuscript-level evaluation rather than an independent data audit; the absence of independent verification of the raw data and statistical models is a limitation of the current bias mitigation protocol, and future industry-partnered research should seek to establish more structurally independent oversight at the level of data access and analysis.

Conclusion

This study found that use of Yuna, an AI-powered DMHI, was associated with both immediate, within-session improvements and gradual, sustained changes in self-reported stress and mood across repeated sessions. Baseline symptom severity was the strongest predictor of within-session improvement, with users entering sessions with higher stress or lower mood presenting with the greatest observed changes, though regression to the mean cannot be ruled out as a contributing explanation. Session duration and the total number of unique therapeutic approaches implemented during the session were also associated with within-session improvements. Safety guardrail activations were not associated with diminished perceived session benefits, suggesting that safety mechanisms can operate without compromising user experience. These findings offer preliminary, real-world evidence that AI-powered DMHIs may represent an accessible and scalable option for mood and stress support. Future research should prioritize prospective, controlled designs to establish causal effects, incorporate validated outcome measures, and explore design features that maximize positive outcomes and user safety.

Acknowledgments

Generative AI tools were used in a limited capacity during the preparation of this manuscript, including to review and confirm editorial changes and language revisions. Generative AI was not used to generate, draft, or substantively write any portion of the manuscript. All content was written and approved by the authors, who take full responsibility for the integrity and accuracy of the content presented.

Funding

This study was not externally funded. Retrospective, de-identified data were provided by Yuna as part of an ongoing scientific partnership with Fit Minded, Inc. Fit Minded, Inc. received compensation for services rendered as the embedded scientific team for Yuna, independent of the conduct or outcomes of this research.

Data Availability

The dataset used for this retrospective analysis is not publicly available given Yuna's privacy policy for patient data. However, aggregated and anonymized data may be shared with interested parties upon reasonable request.

Authors' Contributions

All authors contributed substantially to the manuscript. KLM and CJ: Conceptualization, Methodology, Formal Analysis, Writing – Original Draft, Writing – Review & Editing, Final Approval. CS: Methodology, Writing – Review & Editing, Final Approval. JLH: Conceptualization, Methodology, Supervision, Writing – Original Draft, Writing – Review & Editing, Final Approval.

Conflicts of Interest

Fit Minded, Inc, served as the embedded scientific team for Yuna and received compensation for those services. JH is the Founder and CEO of Fit Minded, Inc, and KM and CJ are employees of Fit Minded, Inc. CS is an independent academic researcher affiliated with Arizona State University who received no financial compensation from Fit Minded, Inc, or Yuna in connection with this study. No author's employment status or compensation is contingent upon the direction or outcome of the findings.

Multimedia Appendix 1
Additional tables and data.
[DOCX File, 2751 KB - mhealth_v14i1e94070_app1.docx]

Checklist 1
STROBE checklist.
[DOCX File, 20 KB - mhealth_v14i1e94070_app2.docx]

References

1. Cuijpers P, Javed A, Bhui K. The WHO World Mental Health Report: a call for action. Br J Psychiatry 2023 Jun;222(6):227-229. [doi: 10.1192/bjp.2023.9] [Medline: 36794529]
2. Piao X, Xie J, Managi S. Continuous worsening of population emotional stress globally: universality and variations. BMC Public Health 2024 Dec 23;24(1):3576. [doi: 10.1186/s12889-024-20961-4] [Medline: 39716139]

XSL•FO
RenderX

3. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. Lancet Psychiatry 2022 Feb;9(2):137-150. [doi: 10.1016/S2215-0366(21)00395-3]
4. Workplace mental health & well-being. US Department of Health and Human Services. URL: https://www.hhs.gov/surgeongeneral/reports-and-publications/workplace-well-being/index.html [accessed 2026-03-23]
5. Work and well-being survey. American Psychological Association. URL: https://www.apa.org/pubs/reports/work-well-being [accessed 2026-02-04]
6. de Oliveira C, Saka M, Bone L, Jacobs R. The role of mental health on workplace productivity: a critical review of the literature. Appl Health Econ Health Policy 2023 Mar;21(2):167-193. [doi: 10.1007/s40258-022-00761-w] [Medline: 36376610]
7. Kazdin AE. Treatment gap: barriers to providing and receiving services. In: Innovations in Psychosocial Interventions and Their Delivery: Leveraging Cutting-Edge Science to Improve the World's Mental Health: Oxford University Press; 2018. [doi: 10.1093/med-psych/9780190463281.003.0003]
8. Ahad AA, Sanchez-Gonzalez M, Junquera P. Understanding and addressing mental health stigma across cultures for improving psychiatric care: a narrative review. Cureus 2023 May;15(5):e39549. [doi: 10.7759/cureus.39549] [Medline: 37250612]
9. Long T. Why are employee assistance programmes under‐utilised and marginalised and how to address it? a critical review and a labour process analysis. Hum Resour Manag J 2024 Nov;34(4):1134-1153. [doi: 10.1111/1748-8583.12547]
10. Moore JT, Wigington C, Green J, et al. Understanding low utilization of employee assistance programs and time off by US public health workers during the COVID-19 pandemic. Public Health Rep 2023;138(4):645-654. [doi: 10.1177/00333549231165287] [Medline: 37052332]
11. Cameron G, Mulvenna M, Ennis E, et al. Effectiveness of digital mental health interventions in the workplace: umbrella review of systematic reviews. JMIR Ment Health 2025 Jan 24;12:e67785. [doi: 10.2196/67785] [Medline: 39854722]
12. Phillips EA, Gordeev VS, Schreyögg J. Effectiveness of occupational e-mental health interventions: a systematic review and meta-analysis of randomized controlled trials. Scand J Work Environ Health 2019 Nov 1;45(6):560-576. [doi: 10.5271/sjweh.3839] [Medline: 31184758]
13. de Miquel C, Haro JM, van der Feltz-Cornelis CM, et al. Differential attrition in randomized controlled trials of digital mental health interventions in the workplace: a systematic review and meta-analysis (EMPOWER H2020 project). Eur Psychiatry 2024 Apr;67(S1):S151-S152. [doi: 10.1192/j.eurpsy.2024.338]
14. Boucher EM, Raiker JS. Engagement and retention in digital mental health interventions: a narrative review. BMC Digit Health 2024 Aug;2(1):52. [doi: 10.1186/s44247-024-00105-9]
15. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325-336. [doi: 10.1002/wps.20673] [Medline: 31496095]
16. Drew JM, Martin S. A national study of police mental health in the USA: stigma, mental health and help-seeking behaviors. J Police Crim Psychol 2021 Jun;36(2):295-306. [doi: 10.1007/s11896-020-09424-9]
17. Stigma towards injured or ill workers. : Griffith University; 2026 URL: https://www.safeworkaustralia.gov.au/sites/default/files/2021-10/Research%20on%20the%20causes%20and%20impact%20of%20stigma%20in%20workplaces_0.pdf [accessed 2026-08-07]
18. Casey TW, Hu X, Carden C, Lee QY. Sticky interventions for a sticky problem: a systematic review of recent workplace mental health stigma reduction interventions with implications for training transfer. J Safety Res 2025 Feb;92:448-458. [doi: 10.1016/j.jsr.2025.01.003] [Medline: 39986863]
19. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. J Med Internet Res 2020 Jul 13;22(7):e16021. [doi: 10.2196/16021] [Medline: 32673216]
20. Torous J, Bucci S, Bell IH, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct;20(3):318-335. [doi: 10.1002/wps.20883] [Medline: 34505369]
21. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464. [doi: 10.1177/0706743719828977] [Medline: 30897957]
22. Gutierrez G, Stephenson C, Eadie J, Asadpour K, Alavi N. Examining the role of AI technology in online mental healthcare: opportunities, challenges, and implications, a mixed-methods review. Front Psychiatry 2024;15:1356773. [doi: 10.3389/fpsyt.2024.1356773] [Medline: 38774435]
23. Maas CJM, Hox JJ. Sufficient sample sizes for multilevel modeling. Methodology 2005;1(3):85-91. [doi: 10.1027/1614-2241.1.3.85]
24. Tran N, Saperia C, Neri E, et al. Effects of app delivered self hypnosis on stress management. NPJ Digit Med 2025 Dec 18;9(1):11. [doi: 10.1038/s41746-025-02182-0] [Medline: 41413254]

XSL•FO
RenderX

25. Haslbeck JMB, Martínez AJ, Roefs AJ, et al. Comparing Likert and visual analogue scales in ecological momentary assessment. Behav Res Methods 2025 Jul 2;57(8):217. [doi: 10.3758/s13428-025-02706-2] [Medline: 40603809]
26. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S. Mobile sensing and support for people with depression: a pilot trial in the wild. JMIR Mhealth Uhealth 2016 Sep 21;4(3):e111. [doi: 10.2196/mhealth.5960] [Medline: 27655245]
27. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
28. The R project for statistical computing. The R Foundation. URL: https://www.r-project.org/ [accessed 2025-03-26]
29. World mental health report: transforming mental health for all. World Health Organization. URL: https://www.who.int/publications/i/item/9789240049338 [accessed 2026-02-04]
30. van Straten A, Hill J, Richards DA, Cuijpers P. Stepped care treatment delivery for depression: a systematic review and meta-analysis. Psychol Med 2015 Jan;45(2):231-246. [doi: 10.1017/S0033291714000701] [Medline: 25065653]
31. Jeitani A, Fahey PP, Gascoigne M, Darnal A, Lim D. Effectiveness of stepped care for mental health disorders: an umbrella review of meta-analyses. Pers Med Psychiatry 2024 Nov;47-48:100140. [doi: 10.1016/j.pmip.2024.100140]
32. Yeo YH, Peng Y, Mehra M, et al. Evaluating for evidence of sociodemographic bias in conversational AI for mental health support. Cyberpsychol Behav Soc Netw 2025 Jan;28(1):44-51. [doi: 10.1089/cyber.2024.0199] [Medline: 39446671]
33. Schleider JL, Weisz JR. Little treatments, promising effects? meta-analysis of single-session interventions for youth psychiatric problems. J Am Acad Child Adolesc Psychiatry 2017 Feb;56(2):107-115. [doi: 10.1016/j.jaac.2016.11.007] [Medline: 28117056]
34. Kim J, Ryu N, Chibanda D. Effectiveness of single-session therapy for adult common mental disorders: a systematic review. BMC Psychol 2023 Nov 7;11(1):373. [doi: 10.1186/s40359-023-01410-0] [Medline: 37936186]
35. Hipgrave L, Goldie J, Dennis S, Coleman A. Balancing risks and benefits: clinicians' perspectives on the use of generative AI chatbots in mental healthcare. Front Digit Health 2025;7:1606291. [doi: 10.3389/fdgth.2025.1606291] [Medline: 40510413]
36. Haque MDR, Rubya S. An overview of chatbot-based mobile mental health apps: insights from app description and user reviews. JMIR Mhealth Uhealth 2023 May 22;11:e44838. [doi: 10.2196/44838] [Medline: 37213181]
37. Jadhakhan F, Blake H, Hett D, Marwaha S. Efficacy of digital technologies aimed at enhancing emotion regulation skills: literature review. Front Psychiatry 2022;13:809332. [doi: 10.3389/fpsyt.2022.809332] [Medline: 36159937]
38. Cividini-Motta C, Livingston C, Efaw H. Systematic review of differential reinforcement in skill acquisition. Behav Anal Pract 2024 Jun;17(2):401-416. [doi: 10.1007/s40617-023-00903-z] [Medline: 38966267]
39. Jackson HM, Batterham PJ, Ohan JL, Calear AL, Farrer LM. Skill enactment and knowledge acquisition among community users of digital mental health interventions: qualitative study with thematic analysis. BMC Psychiatry 2024 Aug 1;24(1):545. [doi: 10.1186/s12888-024-05953-3] [Medline: 39090611]
40. Scholten W, Seldenrijk A, Hoogendoorn A, et al. Baseline severity as a moderator of the waiting list–controlled association of cognitive behavioral therapy with symptom change in social anxiety disorder. JAMA Psychiatry 2023 Aug 1;80(8):822 [FREE Full text] [doi: 10.1001/jamapsychiatry.2023.1291]
41. Sextl-Plötz T, Steinhoff M, Baumeister H, Cuijpers P, Ebert DD, Zarski AC. A systematic review of predictors and moderators of treatment outcomes in internet- and mobile-based interventions for depression. Internet Interv 2024 Sep;37:100760. [doi: 10.1016/j.invent.2024.100760] [Medline: 39139716]
42. Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. Int J Epidemiol 2005 Feb;34(1):215-220. [doi: 10.1093/ije/dyh299] [Medline: 15333621]
43. Chapter 17: intervention complexity. Cochrane. URL: https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-17 [accessed 2026-02-04]
44. Coghlan S, Leins K, Sheldrick S, Cheong M, Gooding P, D'Alfonso S. To chat or bot to chat: ethical issues with using chatbots in mental health. Digit Health 2023;9:20552076231183542. [doi: 10.1177/20552076231183542] [Medline: 37377565]
45. Rahsepar Meadi M, Sillekens T, Metselaar S, van Balkom A, Bernstein J, Batelaan N. Exploring the ethical challenges of conversational AI in mental health care: scoping review. JMIR Ment Health 2025 Feb 21;12(1):e60432. [doi: 10.2196/60432] [Medline: 39983102]
46. Lipschitz JM, Pike CK, Hogan TP, Murphy SA, Burdick KE. The engagement problem: a review of engagement with digital mental health interventions and recommendations for a path forward. Curr Treat Options Psychiatry 2023 Sep;10(3):119-135. [doi: 10.1007/s40501-023-00297-3] [Medline: 38390026]
47. Liu F, Panagiotakos D. Real-world data: a brief review of the methods, applications, challenges and opportunities. BMC Med Res Methodol 2022 Nov 5;22(1):287. [doi: 10.1186/s12874-022-01768-6] [Medline: 36335315]
48. Vanhoffelen G, Vandenbosch L, Schreurs L. Teens, tech, and talk: adolescents' use of and emotional reactions to Snapchat's My AI Chatbot. Behav Sci (Basel) 2025 Jul 30;15(8):1037. [doi: 10.3390/bs15081037] [Medline: 40867394]
49. Mengelkoch S, Moriarity DP, Novak AM, Snyder MP, Slavich GM, Lev-Ari S. Using ecological momentary assessments to study how daily fluctuations in psychological states impact stress, well-being, and health. J Clin Med 2023 Dec 19;13(1):24. [doi: 10.3390/jcm13010024] [Medline: 38202031]

XSL•FO
RenderX

50. Song J, Howe E, Oltmanns JR, Fisher AJ. Examining the concurrent and predictive validity of single items in ecological momentary assessments. Assessment 2023 Jul;30(5):1662-1671. [doi: 10.1177/10731911221113563] [Medline: 36004406]
51. Sundquist K, Schwartz JE, Burg MM, Davidson KW, Diaz KM. Use of a single-item ecological momentary assessment to measure daily exercise: agreement with accelerometer-measured exercise. Sensors (Basel) 2024 Feb 1;24(3):946. [doi: 10.3390/s24030946] [Medline: 38339663]

Abbreviations

ACT: Acceptance and Commitment Therapy
CBT: Cognitive Behavioral Therapy
CI: Confidence Interval
DBT: Dialectical Behavior Therapy
DMHI: Digital Mental Health Intervention
EAP: Employee Assistance Program
GAD-7: Generalized Anxiety Disorder-7
ICC: Intraclass Correlation Coefficient
IQR: Interquartile Range
IRB: Institutional Review Board
LMM: Linear Mixed Model
PHQ-9: Patient Health Questionnaire-9
REML: Restricted Maximum Likelihood
SD: Standard Deviation
SE: Standard Error
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
VAS: Visual Analog Scale

Edited by A Stone, LR Buis; submitted 24.Feb.2026; peer-reviewed by A Sagkriotis, D Shin, U Pasupulety, VL Forman-Hoffman; revised version received 27.Jul.2026; accepted 28.Jul.2026; published 25.Aug.2026.

Please cite as:
McAlister K, Jewell C, Stecher C, Huberty J
Immediate and Sustained Improvements in Mood and Stress Associated With Yuna, an AI-Powered Digital Mental Health Intervention: Real-World Retrospective Study
JMIR Mhealth Uhealth 2026;14:e94070
URL: https://mhealth.jmir.org/2026/1/e94070
doi:10.2196/94070

© Kelsey McAlister, Courtney Jewell, Chad Stecher, Jennifer Huberty. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 25.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Self-Monitoring of Weight Loss Over Time: Secondary Data Analysis of a Randomized Controlled Trial

Renata Savian Colvero de Oliveira[1], MD, MSc; Sharon Nabwire[1], MSc; Heta Merikallio[2], PhD; Markku J Savolainen[2], MD, PhD; Janne Hukkanen[2], MD, PhD; Harri Oinas-Kukkonen[1], PhD

[1]Software Engineering and Information Systems Unit, Oulu Advanced Research on Service and Information Systems, University of Oulu, Linnanmaa campus, Pentti Kaiteran katu 1, Oulu, Finland

[2]Medical Research Center Oulu, Research Unit of Biomedicine and Internal Medicine, University of Oulu and Oulu University Hospital, Oulu, Finland

Corresponding Author:
Renata Savian Colvero de Oliveira, MD, MSc
Software Engineering and Information Systems Unit, Oulu Advanced Research on Service and Information Systems, University of Oulu, Linnanmaa campus, Pentti Kaiteran katu 1, Oulu, Finland

Abstract

Background: Behavior change support systems aim to shape, modify, or strengthen attitudes or behaviors without using coercion or deception. One of the main software features of persuasive system design is self-monitoring, which provides the means for users to continuously track their own performance or status, thereby facilitating goal attainment.

Objective: The aim of this study is to examine whether self-input of weight (self-monitoring frequency) and its interaction with time influence weight loss in adults using a mobile health behavior change support system (mHBCSS). We hypothesized that higher self-monitoring frequency would be associated with greater weight loss, with effects varying across the intervention period.

Methods: This secondary analysis used data from the intervention group of a randomized, open, waitlist-controlled trial in adults with obesity (BMI 30 - 40 kg/m²). Participants used the mHBCSS for 12 months, and analyses included only participants who maintained self-monitoring for at least 6 months (N=75). Weight changes were analyzed across 9 time periods. Quantile regression (QR) was applied to examine effects on the 25th, 50th (median), and 75th percentiles of weight loss. The models included self-monitoring frequency, time periods, and their interaction. A sensitivity analysis using a multiple imputation procedure was performed to assess the robustness of the QR.

Results: The time period variable was significant at the 25th weight loss percentile (QR coefficient: −1.164, 95% CI −1.453 to −0.756) and at the 50th weight loss percentile (QR coefficient: −0.603, 95% CI −0.733 to −0.493). The interaction variable was significant at the 50th (QR coefficient: −0.018, 95% CI −0.047 to −0.003) and 75th (QR coefficient: −0.036, 95% CI −0.051 to −0.027) weight loss percentiles. Self-monitoring frequency alone was not statistically significant.

Conclusions: The study demonstrates that the effect of self-monitoring on weight loss is time-dependent. While a higher frequency of self-monitoring is associated with greater weight loss early in the intervention, its influence decreases as time progresses. These findings emphasize the importance of sustained engagement with self-monitoring rather than focusing solely on frequency, suggesting that interventions should incorporate strategies to maintain consistent self-monitoring use throughout the behavior change process.

Trial Registration: ClinicalTrials.gov NCT04558801; https://clinicaltrials.gov/study/NCT04558801

(JMIR Mhealth Uhealth 2026;14:e71604) doi:10.2196/71604

KEYWORDS

health behavior; behavior change; persuasive systems design; self-tracking; self-monitoring; weight loss

Introduction

Self-monitoring in Digital Health

Persuasive systems have been widely used to motivate users and enhance user experiences, especially in the field of health and well-being [1]. Nevertheless, for these interventions to be practically effective, they must lead to significant modifications in behavior, resulting in better health outcomes. The concept of behavior change support systems (BCSSs) was developed to fulfill this objective. A BCSS is defined as a sociotechnical information system that aims to shape, modify, or strengthen attitudes, behaviors, or compliance without coercion or deception [2].

The implementation of BCSS is approached through the persuasive systems design (PSD) model. This model establishes the postulates (psychological and software design principles) that underlie the design of persuasive systems and emphasizes the importance of considering the context of use, the user, the

XSL•FO
RenderX

technology context, and the specific features implemented in persuasive systems. Within the 4 main categories, primary task support specifically refers to design aspects meant to assist users in performing their primary objective [3].

Self-monitoring, a feature of this category, enables individuals to continuously track their own performance or status, facilitating their goal attainment [3]. It may also happen unknowingly through the system, for example, through wearable sensors [4]. To elaborate, self-monitoring is to a great extent a dynamic self-management skill that involves observing and recording one's own behavior. It is a multicomponent procedure that involves identifying a specific behavior that needs improvement, being aware of when the behavior occurs, documenting how often and how long it happens, comparing it to a standard, and then creating a plan to change it [5]. Examples of usage include self-recognition of symptoms (such as difficulty breathing in asthma), manual recording of blood pressure, and the use of self-maintained electronic databases to monitor blood glucose levels in the management of diabetes [6].

According to the literature, the self-monitoring feature is the most widely used strategy in interventions aimed at promoting health and wellness [7-9]. For instance, a meta-analysis comprising 19 studies with a total of 2800 participants showed that interventions using self-monitoring as a behavior change strategy have the potential to decrease sedentary behavior in adults [10].

Self-monitoring is also widely used in cognitive behavioral therapy (CBT)–based interventions, where it supports the assessment, change, and maintenance of behavioral patterns. Prior qualitative evidence from mobile smoking cessation interventions indicates that CBT-based app designs are favorably perceived in terms of features, design, information quality, and engagement and are associated with changes in users' perceptions and ways of thinking about the target behavior [11], thereby highlighting the relevance of such mechanisms beyond any single application domain. One idea within CBT's mechanisms posits that mere cognitive awareness of one's actions might induce modifications. Another interpretation suggests that the act of monitoring itself triggers reactivity, with individuals internally generating reinforcement or punishment. Yet another view sees monitoring as functioning like prompts or cues that influence behavior. Those explanations show that both cognitive and behavioral processes contribute to the effectiveness of self-monitoring techniques [12]. Recently, a systematic review and meta-analysis of randomized controlled trials (RCTs) evaluating mobile health (mHealth) app self-management interventions for blood pressure management found that self-monitoring was among the most commonly used behavior change techniques. These interventions led to significant reductions in both systolic and diastolic blood pressure compared to usual care, emphasizing the central role of self-monitoring across mHealth applications [13]. In the context of weight change, digital interventions have also made extensive use of self-monitoring alongside established behavior change theories. Most programs combined behavior change techniques and PSD principles, particularly self-monitoring, feedback, goal setting, and tailoring, to guide behavior [14]. Taken together, these findings highlight self-monitoring not only as a frequently applied technique but also as a core functionality embedded across digital health interventions.

Importantly, while self-monitoring is one of the most ubiquitous features in digital BCSS, it is often treated as a generic component rather than being examined as a central mechanism of change whose usage patterns and temporal dynamics may critically shape intervention effectiveness. As a result, the theoretical and empirical understanding of how self-monitoring exerts its effects in real-world digital interventions remains limited.

Another important aspect of this function is the interaction between the frequency of self-monitoring usage and time. A study that investigated how the usage patterns of mHealth app functions affect user retention found that the consistent use of the self-monitoring function helped slow the decline in app usage over time. This suggests that self-monitoring has a beneficial long-term influence on user engagement [15].

Building on this evidence, a large-scale digital weight loss study demonstrated that higher weight self-monitoring frequency was significantly associated with greater weight loss over 6 months. Clinically meaningful weight loss (≥5%) was linked to at least 3 weight SMOs per week, and analysis of outcome-based subgroups further confirmed that individuals with ≥10% weight loss engaged in weight self-monitoring more frequently than those with smaller or no weight changes. These findings reinforce the importance of self-monitoring frequency for weight management while also suggesting that different patterns of usage may correspond to different levels of success [16].

Despite substantial evidence supporting the benefits of self-monitoring for health behavior change, existing research has primarily treated self-monitoring as a static behavior change technique, offering limited insight into how its intensity and temporal dynamics function as mechanisms of change. In particular, it remains unclear whether self-monitoring frequency alone, or its interaction with time, differentially shapes outcomes such as weight loss in digitally delivered BCSSs. For example, prior research has shown that greater consistency of self-monitoring across weight, food, and exercise behaviors predicts weight loss, although adherence to self-monitoring tends to decline over time [17]. This suggests that both how often and how consistently individuals engage in self-monitoring may be critical to understanding its effectiveness.

Therefore, the objective of this study was to examine whether self-monitoring frequency and its interaction with time affect weight loss in adults using an mHBCSS. By focusing on self-monitoring as a core persuasive mechanism and modeling its temporal dynamics, this study aims to advance a more fine-grained understanding of how persuasive system features operate over time, thereby contributing to both BCSS theory and the design of more effective digital weight management interventions. We hypothesized that higher self-monitoring frequency would be associated with greater weight loss and that this association might vary over the course of the intervention, reflecting the dynamic nature of behavior change processes.

XSL•FO
RenderX

Theoretical Background

Analyzing the persuasion context, choosing persuasive design principles, specifying software requirements, and then implementing the software are among the key processes that Oinas-Kukkonen and Harjumaa [3] suggested for creating persuasive systems for behavioral modification. By considering these processes, the aim is to place greater emphasis on improving the entire user experience and increasing user involvement over the long term [2].

The PSD model comprises 4 feature categories: primary task support, dialog support, social support, and credibility support [3]. This study focuses on primary task support (supporting weight tracking and progress monitoring) and dialog support (reminders and suggestions that encourage sustained self-monitoring). Credibility support was present in the app (verifiable expert content and links to external sources), but its analysis is beyond the scope of this paper.

The PSD framework is guided by postulates emphasizing that persuasive systems are value-laden rather than neutral, that persuasion often unfolds incrementally, that both direct and indirect routes of influence are important, and that persuasive systems should be open, unobtrusive, useful, and easy to use [3]. Grounding this study in PSD allows self-monitoring to be examined not as an isolated feature but as part of a broader persuasive system, highlighting the roles of context, temporal dynamics, and sustained engagement in supporting weight loss.

Methods

Sample and Mobile Health Behavior Change Support System

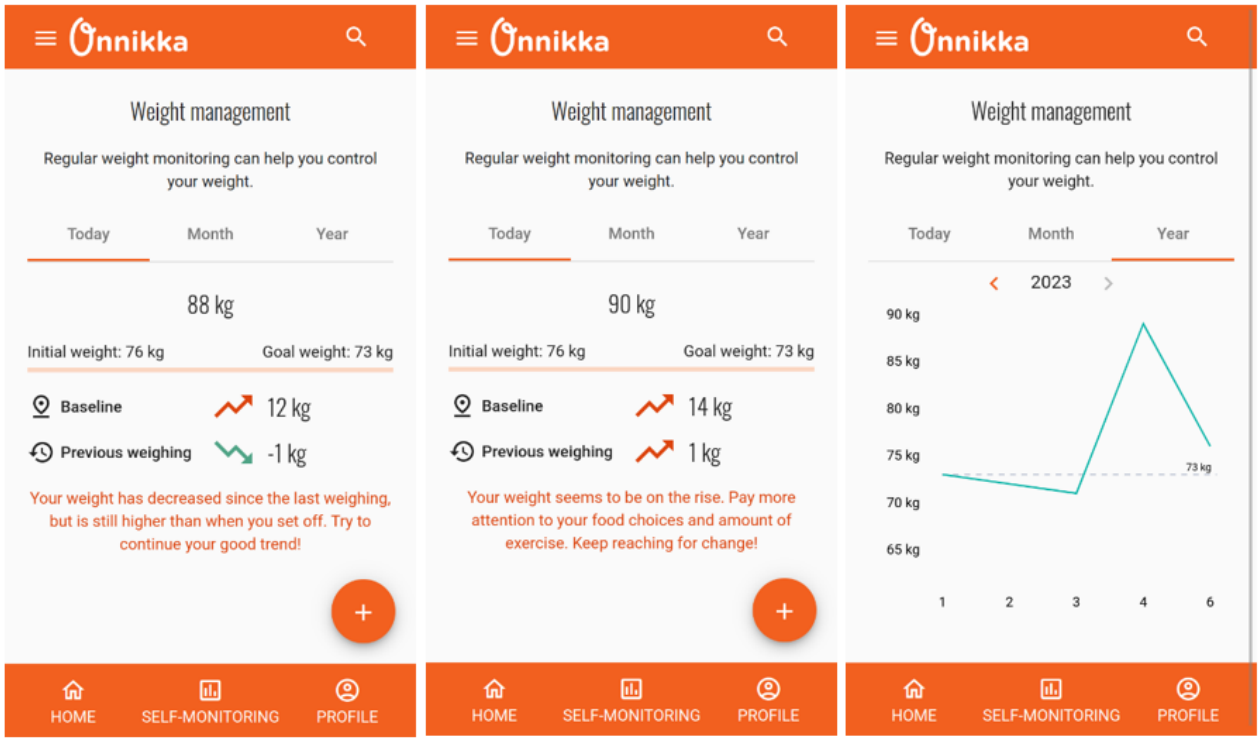
This study is a secondary data analysis of a randomized, open, waitlist-controlled, 2-arm trial. The trial's design was approved by the Ethics Committee of the Northern Ostrobothnia Hospital District and the Finnish Medicines Agency, which classified the mobile health behavior change support system (mHBCSS) as an investigational medical device. The trial investigated whether an mHBCSS could help participants lose weight. It started in October 2020, and the clinical phase ended in September 2022. Participants were randomly assigned to one of two groups: (1) the intervention group (n=100), who received immediate access to the mHBCSS, and (2) the waitlist control group (n=100), who did not receive the intervention during the initial 6 months of the study period. In total, 200 adults (18 - 65 years) with a BMI of 30 to 40 kg/m² were eligible to participate in the study. The primary outcome of the main trial was weight loss from baseline to the 6-month visit. Detailed information, including inclusion and exclusion criteria, was published previously [18,19]. Our study sample comprises the intervention group, with a subsample of "self-monitoring users," that is, individuals who continued self-monitoring their weight from 6 months until the 12-month period, when access to the application was discontinued (n/N=75/100). Two participants were excluded because pregnancy and cortisone use could have influenced weight changes, introducing potential bias.

The app named "Onnikka," (Onnikka Health Ltd) translated as "bus" in Finnish, provided self-monitoring features, enabling participants to monitor their own progress as shown in Figure 1. These functionalities included weight monitoring, food and exercise diaries, and a mood log where users could register their feelings and motivation during the behavior change process. After recording their weight in the application, participants also received an automated, system–generated comment about their weight loss progress (Figure 1). The gradual process of behavior change was metaphorized as a bus journey with symbolic "stops" where new content (expert-created material, videos, external links along with open-ended, multiple-choice, and slider-based interactive tasks) was delivered, aimed at guiding participants on the weight loss journey. The theoretical content was developed based on the PSD principles and the BCSS framework to support weight management among adults with obesity.

The mobile intervention applied the 3 PSD categories: primary task support (self-monitoring–related tools), dialog support (praise for progress, suggestions, and reminders for tasks), and credibility support (expertise regarding the content, verifiability by providing links to external evidence–based resources, and transparency in the privacy policy). Each Monday morning, participants received a reminder to weigh themselves. This approach aligns with the PSD principle of unobtrusiveness, providing scheduled prompts at consistent times to encourage predictability, minimize required interactions, and reduce perceived intrusiveness.

XSL•FO
RenderX

Figure 1. Onnikka screenshots representing the self-monitoring feature.



Quantile Regression

After assessing the normality of the data using the Shapiro-Wilk test in R 4.4.0 software (R Foundation for Statistical Computing) [20], we used a nonparametric approach, quantile regression (QR) [21], to examine the influence of self-monitoring frequency on weight loss. QR allows the modeling of different quantiles of the conditional distribution of the response variable, providing a more detailed view of the effects of the explanatory variables at different points in the weight loss distribution [22]. Modeling multiple quantiles enables a more comprehensive understanding of the influence of predictors on the response distribution [23]. The rationale for using this method is that it offers greater versatility than other regression techniques in identifying associations at different points of the dependent-variable distribution [24]. Moreover, quantiles, such as the median (p=50%, where p denotes the quantile being estimated), are robust to outliers due to their emphasis on the proportion of data within a specific range, rather than on distance [25]. The QR parameter estimates the change in a specified quantile of the response variable produced by a 1-unit change in the predictor variable.

QR models were fitted to the 0.25, 0.50 (median), and 0.75 quantiles using the rq function of the quantreg package in R [26]. The model included an interaction term between self-monitoring frequency at each time period (TP) and the TPs themselves, aiming to capture how the effect of self-monitoring on weight change varied over time. Multicollinearity was detected using the variance inflation factor, with all values exceeding 5 [27] (Multimedia Appendix 1).

Variables

Dependent Variable

The dependent variable was the average weight loss percentage (self-reported data). This variable represents the average percentage change in weight relative to the initial weight recorded by the user for each TP (1 - 41 d, 42 - 83 d, 84 - 125 d, 126 - 167 d, 168 - 209 d, 210 - 251 d, 252 - 293 d, 294 - 335 d, and 336 - 377 d). Negative values indicate weight loss, while positive values indicate weight gain.

It was divided into 9 TPs in this analysis to balance temporal resolution and interpretability. Given that users were required to weigh themselves weekly, the intervention was organized into 6-week windows, with the exception of the first period (40 d) to ensure symmetry. This structure allowed each interval to include sufficient observations for stable QR estimates while capturing meaningful variation over time.

Independent Variables

The independent variables consisted of the self-monitoring frequency at each time point, the 9 TPs, and the interaction between self-monitoring and the TP (TP interaction).

Sensitivity Analysis

To assess the robustness of the results to missing data, a sensitivity analysis was performed using multiple imputation [28,29]. Missing data were imputed using the multiple imputation method implemented in the mice package in R, which generates multiple imputed datasets to reflect the uncertainty associated with imputation [30]. Finally, the QR was applied to each imputed dataset separately.

XSL•FO
RenderX

Ethical Considerations

The study was approved by the Ethics Committee of the Northern Ostrobothnia Hospital District (approval: 138/2020) and the Finnish Medicines Agency, with the mHBCSS classified as an investigational medical device. The trial was registered at ClinicalTrials.gov (identifier: NCT04558801). During the conduct of the trial, the participants received oral and written information about the trial, and written informed consent was obtained.

Results

Sample Characteristics

In the analyzed dataset, a total of 75 participants were included, consisting of 11 males (14.7%) and 64 females (85.3%). The mean age of the participants was 48.4 (SD 9.90) years, with a median age of 49 years (IQR 41-58). Participants exhibited a mean total self-monitoring frequency of 34.6 (SD 22.4), with a median of 30 (IQR 24-40), representing the cumulative number of self-weighing inputs over the full 12-month study period. Table 1 summarizes descriptive statistics for the participants. The individuals' detailed demographic information for the participants is shown in Multimedia Appendix 2.

Table . Descriptive statistics of self-monitoring frequency and weight by age group[a].

Age range (y)	Frequency, n (%)	Mean self-monitoring frequency (SD)	Median self-monitoring frequency (IQR)	Mean weight change (SD), (%)	Median weight change (IQR), (%)	Mean weight at baseline (SD), (kg)	Median weight at baseline (IQR), (kg)
24 - 39	14 (18.7)	26.3 (9.77)	28.5 (19.8-32.2)	−2.77 (2.77)	−3.03 (−4.95 −0.91)	97.2 (13.8)	92.6 (85.20-109.20)
40 - 49	24 (32)	36.1 (28.7)	29 (24.8-40)	−3.13 (5.03)	−1.55 (−4.44 −0.14)	96.2 (12.8)	97.7 (86-102.45)
50 - 59	27 (36)	39.7 (21.9)	37 (27-48)	−3.88 (3.83)	−2.72 (−5.77 −1.14)	94 (13)	91 (87.60-99.10)
60 - 65	10 (13.3)	28.6 (10.7)	26 (19.2-38.8)	−2.27 (2.91)	−1.57 (−4.04 −0.39)	97.4 (13.8)	93.9 (86.10-103.10)

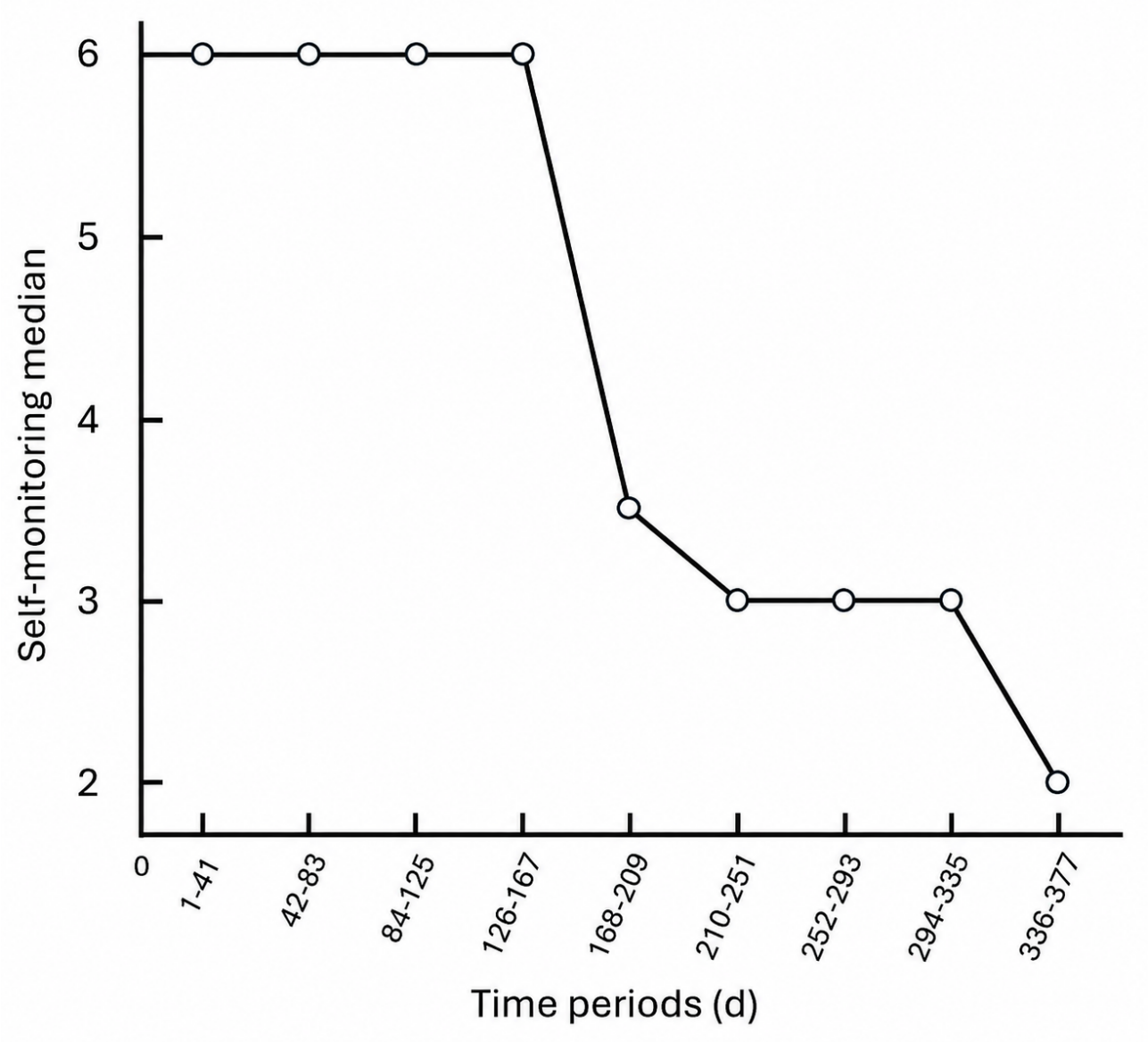
[a]Weight change is expressed as percentage change over the full 12-month study period, with negative values indicating weight loss. Self-monitoring frequency refers to the total number of self-weighing inputs over the 12-month period.

The mean weight change relative to baseline across all 9 TPs was −3.24%, which was lower than the median decrease of 2.22%, suggesting a left-skewed distribution. Moreover, the 25th percentile of weight change was −5.11%, meaning that 25% of participants lost 5.11% of their weight or more. The median weight change was −2.22%, indicating that 50% of participants lost at least 2.22% of their weight.

Figure 2 shows that during the first 4 TPs (1 - 41 d, 42 - 83 d, 84 - 125 d, and 126 - 167 d), the median self-monitoring frequency remained consistent at 6 times. In contrast, the median began to decrease in the fifth TP, reaching 3.5 times and stabilizing at 3 times during TPs 6 (210 - 251 d) through 8 (294 - 335 d). By TP 9 (336 - 377 d), the median dropped further to 2 times.

XSL•FO
RenderX

Figure 2. The median number of self-monitored frequency for the 9 time periods.



Percentile Analysis

In this study, the 25th, 50th, and 75th percentiles refer to participants experiencing low, median, and high percentage weight loss, respectively, across the study period.

Results of the percentile analysis in Table 2 show that for the 25th percentile (participants with relatively low percentage weight loss), the coefficient for TP was −1.164 (95% CI −1.453 to −0.756), with statistical significance. This indicates that, across TPs, participants in this group lost more weight in earlier periods, and the rate of weight loss decreased over subsequent periods.

XSL•FO
RenderX

Table . Quantile regression coefficients and 95% CIs for weight loss by percentile[a].

Variables	Quantile regression coefficients (95% CI)		
	25th percentile	50th percentile	75th percentile
Intercept	0.53 (−1.24 to 1.64)	−0.16 (−0.78 to 0.93)	−0.10 (−0.45 to 0.42)
SMO frequency	−0.20 (−0.37 to 0.08)	0.01 (−0.17 to 0.08)	0.05 (−0.03 to 0.10)
TP[b]	−1.16 (−1.45 to −0.76)[c]	−0.60 (−0.73 to −0.49)[c]	−0.12 (−0.28 to 0.02)
TP interaction	0.02 (−0.09 to 0.05)	−0.02 (−0.05 to −0.00)[c]	−0.04 (−0.05 to −0.03)[c]

[a]Values are rounded to decimal places.

[b]TP: time period.

[c]Significant coefficients at the 95% CI. The 25th, 50th, and 75th percentiles refer to participants with relatively low, median, and high percentage weight loss, respectively, across all TPs.

In the 50th percentile (participants with median percentage weight loss), self-monitoring frequency alone was not significantly associated with weight change. However, both the TP variable (QR coefficient: −0.603, 95% CI −0.733to −0.493) and the interaction between self-monitoring frequency and TPs (QR coefficient: −0.018, 95% CI −0.047 to −0.003) were statistically significant. This suggests that weight loss per TP tends to be smaller in later periods at the median, and the influence of self-monitoring frequency on weight loss decreases as time progresses.

Finally, for the 75th percentile (participants with high percentage weight loss), only the interaction (QR coefficient: −0.036, 95% CI −0.051 to −0.027) was significant, suggesting that the effect of self-monitoring frequency on weight loss also diminishes over time for this group.

Overall, the influence of self-monitoring frequency on weight loss diminishes over time, with this effect being more pronounced at higher percentiles of weight loss.

The 3 intercepts (the expected initial weight loss when both the TP and self-monitoring frequency are zero) were not statistically significant.

Multiple Imputation Procedure

To account for missing data, a multiple imputation procedure was applied, generating 5 imputed datasets. QR was performed on each imputed dataset, and the results were consistent across all imputations. Detailed results are presented in Tables 1 to 3 of Multimedia Appendix 3.

The effect of time was significant and negative, indicating that weight loss was greater in the earlier periods and declined over subsequent periods. The interaction between self-monitoring frequency and time was significant in most models, showing that the effect of self-monitoring frequency on weight loss changes over time. In contrast, self-monitoring frequency alone was not significant, highlighting that its impact can only be understood in combination with time.

Discussion

Principal Findings

The main finding of this study is that self-monitoring does not exert a stable effect on weight loss over time; rather, its influence is time-dependent. Specifically, the interaction between self-monitoring frequency and time was significant at the median (50th percentile) and higher (75th percentile) levels of weight loss. More frequent self-monitoring was associated with greater weight loss in the early stages of the intervention, but this effect diminished over time. Self-monitoring frequency alone did not have a significant independent effect, indicating that its impact depends on when it occurs during the behavior change process.

These results suggest that *when* and *how consistently* individuals engage in self-monitoring may matter more than how frequently they do so at a single point in time. The added value of frequent self-monitoring appears to be concentrated in the early phases of the intervention, whereas sustained engagement over time may be more relevant for maintaining weight loss, particularly among individuals achieving higher levels of success.

In contrast to prior work that has primarily examined self-monitoring as a static behavior change technique [8-10] or focused on its overall frequency [16], this study extends existing research by demonstrating that the effectiveness of self-monitoring is time-dependent and varies across different outcome levels of weight loss. By applying QR, we show that self-monitoring operates differently for individuals achieving median versus higher levels of weight loss, thereby offering a more nuanced, distribution-sensitive understanding of its role in digital weight management interventions.

The results of this study also add to the findings of Karppinen et al [31]. Based on their qualitative results, self-monitoring was perceived by users as a beneficial persuasive feature among those who achieved 5% weight loss. The interaction with time reflects and explains how self-monitoring integrates into an ongoing process of adaptation, learning, and sustained behavior change, emphasizing that its power lies in commitment rather than immediacy. This idea is supported by previous research on BCSS [32], which highlighted the importance of encouraging consistency by helping users commit to specific goals over time. Systems that reinforce user commitments are more successful in driving behavior change. That research also identified a key challenge: many individuals struggle to follow through on health-improvement goals due to a lack of self-control or willpower. Consequently, people often choose simpler solutions that fail to produce significant, long-term results because they lack strong commitments. Finally, Oduor and Oinas-Kukkonen

XSL•FO
RenderX

[32] also found that primary task support, a main category of self-monitoring features, had the greatest impact on perceived competence and users' intentions to continue, suggesting that systems encouraging goal commitment are more likely to drive lasting behavior change.

In this study, the frequency of self-monitoring alone was not significant. This is consistent with the findings of a secondary data analysis of an RCT in which individuals who used a smartphone-based system for behavioral weight loss did not experience increased weight loss despite engaging in more self-monitoring [33]. Similar findings in a prospective study showed that self-monitoring of blood glucose levels had no influence on glycemic management in individuals newly diagnosed with type 2 diabetes [34]. This evidence emphasizes our findings, indicating the importance of analyzing the interaction with time to understand the impact of self-monitoring.

Since the diminishing effect is significant at the 50th and 75th weight loss percentiles, these insights are particularly relevant for individuals in the median-to-upper ranges of weight loss. For individuals in this stage, sustained engagement with self-monitoring may demand additional support or motivational, personalized strategies. A previous study has shown that users who proactively set up a reminder have significantly higher long-term adherence to an mHealth app to improve nutrition literacy and support healthy dietary behaviors. Moreover, the relevance of the social role feature was further demonstrated by the fact that users who selected a female conversational agent had a higher likelihood of being long-term adherents [35]. A recent systematic literature review supports this evidence by showing that the most persuasive patient engagement features of mHealth apps come from dialog support (praise, rewards, reminders, suggestions, and social role) [36]. In this way, PSD features within dialog support, such as reminders, suggestions, and social role, could be effective strategies to sustain the user's engagement. In this study, participants received 26 weekly reminders to weigh themselves during the active intervention period (the first 6 months). This structured approach ensured consistent self-monitoring during the initial phase of the study. However, after the 6-month period, these reminders ceased, potentially contributing to a decline in self-monitoring. The lack of continuous reminders may have affected participants' adherence to routine self-weighing, leading to a decline in their capacity to maintain consistent self-monitoring behavior over time.

Moreover, self-expansion, explained as increasing positive self-content by engaging in novel and rewarding activities [37], is associated with better behavioral weight loss outcomes, including a greater likelihood of achieving a clinically significant (5%) weight loss through an internet-based behavioral weight loss intervention [38]. Self-expanding activities and/or experiences, such as learning new things about weight loss, behavioral challenges, and social interaction, might have offered alternatives to food rewards [38]. Those interpretations could lead us to infer that adding the persuasive feature "reward," which also belongs to dialog support, could be another good strategy to be combined with self-monitoring. There is a neurophysiological basis behind this claim, since food triggers the release of dopamine in the brain, generating feelings of pleasure and reinforcement [39].

In this same direction, the reward feature was shown to be linked to the emotional state of happiness in a large-scale empirical study (n=660) aimed at investigating if and how people respond emotionally to persuasive features. Reward was the only feature that was not connected with any negative feelings (eg, anger, disgust, fear, and sadness) [40]. This is particularly relevant because people with obesity or who are overweight often turn to food as a source of reward, which can perpetuate a cycle of addiction [39]. Introducing digital rewards (eg, reward features), therefore, could represent an alternative form of stimulation, helping to replace the search for pleasure through food and facilitating the adoption of new healthy behaviors.

In practice, our study suggests that when self-monitoring is decreasing over time, it should be interpreted as a signal that the intervention needs to be strengthened or adjusted. While self-monitoring remains a valuable feature in itself, additional supportive strategies are needed to maintain or enhance its long-term effectiveness. Interventions might need to evolve to sustain engagement and impact, perhaps by introducing new elements such as social support features. Karppinen et al [31] found that the need for these features increased over the duration of an earlier web version of this same intervention. Additionally, the absence of social support was linked to lower self-monitoring, according to a qualitative study that examined the experiences and emotions related to self-monitoring in eating behaviors among participants who had completed an 18-month behavior weight reduction study [41].

Furthermore, prior research revealed that self-monitoring encourages healthy habits by inspiring reflective thinking about an individual's behavior. As individuals tend to strive to surpass their previous main task-related records, this tendency also leads to internal competition between their previous and current attitudes [7]. However, reflection alone may not be sufficient; combining self-monitoring with other software features (such as dialog and primary task support) may help transform the need for reflection into actionable insights [42], thereby supporting long-term behavior change.

Another aspect of reducing or stopping users' self-monitoring should be considered. Based on the results of an international survey involving users of self-monitoring technologies, Ajana [43] found that, although this feature increases motivation and encourages a healthier lifestyle, its quantitative aspects can sometimes lead to excessive behavior, which can increase pressure on the user, resulting in feelings of inadequacy, for example. Self-monitoring could also lead to health disorders (eg, in users with anorexic tendencies) and depression [7]. Other aspects on the negative side of self-monitoring are the tediousness and lack of enjoyment of the activity, reinforcing that it should be implemented alongside other strategies [7]. In the mHBCSS under investigation here, users received a brief comment about their weight loss progress after recording their weight, potentially reducing the likelihood that these last 2 factors would cause a decrease in self-monitoring. However, to overcome potential adverse effects of self-monitoring, the designers probably should adopt a comprehensive approach to

XSL•FO
RenderX

promoting health and well-being, rather than monitoring just one marker of health behavior [7].

Finally, our results extend PSD theory by highlighting the temporal dimension of engagement: although self-monitoring is most impactful in the early stages, its influence diminishes over time, suggesting that persuasive systems should be designed not only to encourage initial adoption but also to sustain long-term engagement. This suggests that while PSD principles such as reminders (dialog support) can initially sustain engagement, additional strategies may be required to maintain long-term adherence. Together, these findings refine the application of PSD by emphasizing that its principles need to be considered in a temporal context. Specifically, persuasive features may be most effective in the early phases of behavior change, whereas later phases may require adaptive or complementary strategies.

Limitations and Future Research

This study has several limitations. First, the sample size (N=75) was relatively small, which limits statistical power and may increase the risk of type II errors. Second, the majority of the sample was female (64/75, 85.3%), and the study was conducted in a single cultural and geographic context (Finland), which may limit the applicability of the results to other populations or settings.

Third, while QR allowed us to examine effects across different percentiles of weight loss, the findings are specific to the quantiles analyzed (25th, 50th, and 75th) and cannot be directly generalized to the entire population. Fourth, self-monitoring frequency was measured as the total number of self-weighing entries recorded, but this metric does not capture qualitative aspects of engagement, such as adherence to recommended behaviors, motivation, or the accuracy of self-reported data. Finally, the study design was observational, limiting causal inference between self-monitoring and weight loss outcomes.

Building on the time-dependent effects observed in this study, future longitudinal research could examine how different trajectories of self-monitoring engagement relate to distinct weight loss trajectories and how adaptive intervention strategies might sustain engagement in later phases. Given the observational design of this study, experimental or adaptive trial designs (eg, just-in-time adaptive interventions) would be valuable to establish causal relationships between self-monitoring patterns and outcomes. It might also be valuable to explore the mental and emotional impact of self-monitoring and adapt this strategy, avoiding possible negative effects and improving adherence. Additionally, the effects of age, gender, and other characteristics should be studied to personalize the interventions. Finally, future studies could examine how the framing of system-generated feedback, such as positive versus negative comments, affects motivation, consistency, and commitment to self-monitoring, providing insights to further optimize digital behavior change interventions.

Conclusions

In this research, self-monitoring, a core functionality of the presented mHBCSS, was studied over time to understand its impact on weight loss. While self-monitoring frequency alone did not show a statistically significant independent effect, the interaction between self-monitoring frequency and time was significant at the 50th and 75th weight loss percentiles. This suggests that the potential influence of self-monitoring on weight loss may vary across the intervention period, with a stronger association observed in earlier phases that diminishes over time. These findings highlight the importance of considering the timing of behavioral engagement, suggesting that sustained self-monitoring may be more valuable for long-term effectiveness than intensity alone. Future interventions may benefit from designing strategies that encourage consistent engagement with self-monitoring tools throughout the behavior change process to maximize weight loss outcomes.

Acknowledgments

The authors thank the research staff of the 2 research units as well as the study participants for their dedication and contributions to this study.

Funding

This study has received funding from the Research Council of Finland under decision 351670, with the theme “Persuasive digital health interventions: Software features as key predictors for successful prevention and treatment of overweight, obesity, metabolic syndrome and cardiovascular diseases.” It has also received funding from the Business Finland under decision 6704/31/2019. The trial was also supported by the Novo Nordisk Foundation (NNF21OC0072621), the Suorsa Healthcare Foundation, and the Finnish Foundation for Cardiovascular Research.

Data Availability

The data that supports the findings of this study may be available on reasonable request from the corresponding author. Personal data is not publicly available due to privacy and ethical restrictions.

Authors' Contributions

Conceptualization: RSCdO, HO-K
Data curation: HM
Formal analysis: RSCdO, SN

XSL•FO
RenderX

Funding acquisition: MJS, JH, HO-K
Investigation: HM, MJS, JH, HO-K
Methodology: RSCdO
Supervision: HO-K
Validation: RSCdO, HO-K
Visualization: RSCdO
Writing – original draft: RSCdO, HO-K
Writing – review & editing: RSCdO, SN, HM, MJS, JH, HO-K

Conflicts of Interest

The system described in this study and the data collected were obtained in an academic setting. Later, a spin-off was established. HO-K and MJS are the founders and shareholders of Onnikka Health Oy, which owns the intellectual property rights. All other authors have reported that they have no relationships relevant to the content of this paper. The authors declare they have no financial interests related to this study.

Multimedia Appendix 1
Variance inflation factor.
[DOCX File, 16 KB - mhealth_v14i1e71604_app1.docx]

Multimedia Appendix 2
Demographic information.
[DOCX File, 18 KB - mhealth_v14i1e71604_app2.docx]

Multimedia Appendix 3
Quantile regression analysis with multiple imputation.
[DOCX File, 18 KB - mhealth_v14i1e71604_app3.docx]

References

1. Orji R, Moffatt K. Persuasive technology for health and wellness: state-of-the-art and emerging trends. Health Informatics J 2018 Mar;24(1):66-91. [doi: 10.1177/1460458216650979] [Medline: 27245673]
2. Oinas-Kukkonen H. A foundation for the study of behavior change support systems. Pers Ubiquit Comput 2013 Aug;17(6):1223-1235. [doi: 10.1007/s00779-012-0591-5]
3. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24:24. [doi: 10.17705/1CAIS.02428]
4. Khoshmanesh F, Thurgood P, Pirogova E, Nahavandi S, Baratchi S. Wearable sensors: at the frontier of personalised health monitoring, smart prosthetics and assistive technologies. Biosens Bioelectron 2021 Mar 15;176:112946. [doi: 10.1016/j.bios.2020.112946] [Medline: 33412429]
5. Chan PE, Graham-Day KJ, Ressa VA, Peters MT, Konrad M. Beyond involvement: promoting student ownership of learning in classrooms. Interv Sch Clin 2014;50(2):105-113. [doi: 10.1177/1053451214536039]
6. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016;11(5):e0156164. [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
7. Orji R, Lomotey R, Oyibo K, Orji F, Blustein J, Shahid S. Tracking feels oppressive and "punishy": exploring the costs and benefits of self-monitoring for health and wellness. Digit Health 2018;4:2055207618797554. [doi: 10.1177/2055207618797554] [Medline: 30202544]
8. Pala D, Petrini G, Bosoni P, Larizza C, Quaglini S, Lanzola G. Smartphone applications for nutrition support: a systematic review of the target outcomes and main functionalities. Int J Med Inform 2024 Apr;184:105351. [doi: 10.1016/j.ijmedinf.2024.105351] [Medline: 38295584]
9. Zhu Y, Long Y, Wang H, Lee KP, Zhang L, Wang SJ. Digital behavior change intervention designs for habit formation: systematic review. J Med Internet Res 2024 May 24;26:e54375. [doi: 10.2196/54375] [Medline: 38787601]
10. Compernolle S, DeSmet A, Poppe L, et al. Effectiveness of interventions using self-monitoring to reduce sedentary behavior in adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2019 Aug 13;16(1):63. [doi: 10.1186/s12966-019-0824-3] [Medline: 31409357]
11. Tudor-Sfetea C, Rabee R, Najim M, et al. Evaluation of two mobile health apps in the context of smoking cessation: qualitative study of cognitive behavioral therapy (CBT) versus non-CBT-based digital solutions. JMIR Mhealth Uhealth 2018 Apr 18;6(4):e98. [doi: 10.2196/mhealth.9405] [Medline: 29669708]
12. Hufstedler A, Varghese FP. Self-monitoring behavioral assessment techniques. In: Carducci BJ, Nave CS, Mio JS, Riggio RE, editors. The Wiley Encyclopedia of Personality and Individual Differences: Measurement and Assessment: John Wiley & Sons, Ltd; 2020:439-445. [doi: 10.1002/9781119547167]

XSL•FO
RenderX

13. Zhou Y, Li SJ, Huang RQ, et al. Behavior change techniques used in self-management interventions based on mhealth apps for adults with hypertension: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2024 Oct 22;26:e54978. [doi: 10.2196/54978] [Medline: 39437388]
14. Asbjørnsen RA, Smedsrød ML, Solberg Nes L, et al. Persuasive system design principles and behavior change techniques to stimulate motivation and adherence in electronic health interventions to support weight loss maintenance: scoping review. J Med Internet Res 2019 Jun 21;21(6):e14265. [doi: 10.2196/14265] [Medline: 31228174]
15. Lee K, Kwon H, Lee B, et al. Effect of self-monitoring on long-term patient engagement with mobile health applications. PLoS One 2018;13(7):e0201166. [doi: 10.1371/journal.pone.0201166] [Medline: 30048546]
16. Painter SL, Ahmed R, Hill JO, et al. What matters in weight loss? An in-depth analysis of self-monitoring. J Med Internet Res 2017 May 12;19(5):e160. [doi: 10.2196/jmir.7457] [Medline: 28500022]
17. Hutchesson MJ, Tan CY, Morgan P, Callister R, Collins C. Enhancement of self-monitoring in a web-based weight loss program by extra individualized feedback and reminders: randomized trial. J Med Internet Res 2016 Apr 12;18(4):e82. [doi: 10.2196/jmir.4100] [Medline: 27072817]
18. Markkanen JO, Oikarinen N, Savolainen MJ, et al. Mobile health behaviour change support system as independent treatment tool for obesity: a randomized controlled trial. Int J Obes (Lond) 2024 Mar;48(3):376-383. [doi: 10.1038/s41366-023-01426-x] [Medline: 38062218]
19. de Oliveira RSC, Nabwire S, Merikallio H, Savolainen M, Hukkanen J, Oinas-Kukkonen H. Behind the software: the impact of unobtrusiveness, goal setting and persuasive features on BMI. Int J Med Inform 2025 Apr;196:105795. [doi: 10.1016/j.ijmedinf.2025.105795] [Medline: 39862566]
20. RStudio. posit. 2023. URL: https://posit.co/products/open-source/rstudio [accessed 2026-08-22]
21. Koenker R, Bassett Jr G. Regression quantiles. Econometrica 1978 Jan;46(1):33. [doi: 10.2307/1913643]
22. Koenker R, Chernozhukov V, He X, Peng L, editors. Handbook of Quantile Regression, 1st edition: Chapman and Hall/CRC; 2017. [doi: 10.1201/9781315120256]
23. Hao L, Naiman DQ. Quantile Regression: Sage Publications; 2007.
24. Lê Cook B, Manning WG. Thinking beyond the mean: a practical guide for using quantile regression methods for health services research. Shanghai Arch Psychiatry 2013 Feb;25(1):55-59. [doi: 10.3969/j.issn.1002-0829.2013.01.011] [Medline: 24948867]
25. Das K, Krzywinski M, Altman N. Quantile regression. Nat Methods 2019 Jun;16(6):451-452. [doi: 10.1038/s41592-019-0406-y] [Medline: 31147637]
26. Koenker R. quantreg: quantile regression. Comprehensive R Archive Network (CRAN). 2020. URL: https://CRAN.R-project.org/package=quantreg [accessed 2026-08-15]
4. O'brien RM. A caution regarding rules of thumb for variance inflation factors. Qual Quant 2007 Sep 11;41(5):673-690. [doi: 10.1007/s11135-006-9018-6]
28. Rubin DB. Multiple Imputation for Nonresponse in Surveys: John Wiley & Sons; 1987. [doi: 10.1002/9780470316696]
29. Rubin DB. Multiple imputation after 18+ years. J Am Stat Assoc 1996 Jun;91(434):473-489. [doi: 10.1080/01621459.1996.10476908]
3. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J Stat Soft 2011;45(3). [doi: 10.18637/jss.v045.i03]
31. Karppinen P, Oinas-Kukkonen H, Alahäivälä T, et al. Persuasive user experiences of a health Behavior Change Support System: a 12-month study for prevention of metabolic syndrome. Int J Med Inform 2016 Dec;96:51-61. [doi: 10.1016/j.ijmedinf.2016.02.005] [Medline: 26992482]
32. Oduor M, Oinas-Kukkonen H. Committing to change: a persuasive systems design analysis of user commitments for a behaviour change support system. Behav Inf Technol 2021 Jan 2;40(1):20-38. [doi: 10.1080/0144929X.2019.1598495]
33. Goldstein SP, Goldstein CM, Bond DS, Raynor HA, Wing RR, Thomas JG. Associations between self-monitoring and weight change in behavioral weight loss interventions. Health Psychol 2019 Dec;38(12):1128-1136. [doi: 10.1037/hea0000800] [Medline: 31556659]
34. O'Kane MJ, Bunting B, Copeland M, Coates VE, ESMON study group. Efficacy of self monitoring of blood glucose in patients with newly diagnosed type 2 diabetes (ESMON study): randomised controlled trial. BMJ 2008 May 24;336(7654):1174-1177. [doi: 10.1136/bmj.39534.571644.BE] [Medline: 18420662]
35. Jakob R, Narauskas J, Fleisch E, König LM, Kowatsch T. Factors associated with adherence to a public mobile nutritional health intervention: retrospective cohort study. Comput Hum Behav Rep 2024 Aug;15:100445. [doi: 10.1016/j.chbr.2024.100445]
36. Almutairi N, Vlahu-Gjorgievska E, Win KT. Persuasive features for patient engagement through mHealth applications in managing chronic conditions: a systematic literature review and meta-analysis. Inform Health Soc Care 2023 Jul 3;48(3):267-291. [doi: 10.1080/17538157.2023.2165083] [Medline: 36650714]
37. Aron A, Lewandowski GW Jr, Mashek D, Aron EN. The self-expansion model of motivation and cognition in close relationships. In: Simpson JA, Campbell L, editors. The Oxford Handbook of Close Relationships: Oxford University Press; 2013:90-115 URL: https://psycnet.apa.org/record/2012-32735-005 [accessed 2026-08-15]

XSL•FO
RenderX

38. Xu X, Leahey TM, Boguszewski K, Krupel K, Mailloux KA, Wing RR. Self-expansion is associated with better adherence and obesity treatment outcomes in adults. Ann Behav Med 2017 Feb;51(1):13-17. [doi: 10.1007/s12160-016-9823-7] [Medline: 27436226]
39. Volkow ND, Wang GJ, Tomasi D, Baler RD. Obesity and addiction: neurobiological overlaps. Obes Rev 2013 Jan;14(1):2-18. [doi: 10.1111/j.1467-789X.2012.01031.x] [Medline: 23016694]
40. Oyebode O, Steeves D, Orji R. Persuasive strategies and emotional states: towards designing personalized and emotion-adaptive persuasive systems. User Model User-Adap Inter 2024 Sep;34(4):1175-1225. [doi: 10.1007/s11257-023-09390-x]
41. Burke LE, Swigart V, Warziski Turk M, Derro N, Ewing LJ. Experiences of self-monitoring: successes and struggles during treatment for weight loss. Qual Health Res 2009 Jun;19(6):815-828. [doi: 10.1177/1049732309335395] [Medline: 19365099]
42. Halttu K, Oinas-Kukkonen H. Persuading to reflect: role of reflection and insight in persuasive systems design for physical health. Hum Comput Interact 2017 Nov 2;32(5-6):381-412. [doi: 10.1080/07370024.2017.1283227]
43. Ajana B. Personal metrics: users' experiences and perceptions of self-tracking practices and data. Soc Sci Inf 2020 Dec;59(4):654-678. [doi: 10.1177/0539018420959522]

Abbreviations

BCSS: behavior change support system
CBT: cognitive behavioral therapy
mHBCSS: mobile health behavior change support system
mHealth: mobile health
PSD: persuasive systems design
QR: quantile regression
RCT: randomized controlled trial
TP: time period

Edited by L Buis; submitted 22.Jan.2025; peer-reviewed by A Eisingerich, CJ Hagerman, L Bai; revised version received 20.Feb.2026; accepted 10.Jun.2026; published 31.Aug.2026.

Please cite as:
de Oliveira RSC, Nabwire S, Merikallio H, J Savolainen M, Hukkanen J, Oinas-Kukkonen H
Self-Monitoring of Weight Loss Over Time: Secondary Data Analysis of a Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e71604
URL: https://mhealth.jmir.org/2026/1/e71604
doi:10.2196/71604

© Renata Savian Colvero de Oliveira, Sharon Nabwire, Heta Merikallio, Markku J Savolainen, Janne Hukkanen, Harri Oinas-Kukkonen. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 31.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Behavioral Mechanisms of a Digital Health Intervention for Self-Management in Type 2 Diabetes Mellitus: Prospective Longitudinal Cohort Study

Yibo Wu[1*], PhD; Yang Ni[2*], MSN; Yang Jiang[3*], BSMed; Zijie Xu[4], MPH; Hewei Min[1], PhD; Ping Chen[1], MPharm; Xinbao Gu[5], BM; Bingyang Kong[5], BM; Yadi Gan[6]; Pei Li[7], MMed; Mingzi Li[8], PhD; Xiaohui Guo[9], PhD; Xuxi Zhang[10], PhD; Aijuan Ma[6], PhD; Xinying Sun[1], PhD

[1]School of Public Health, Peking University, 38 Xueyuan Road, Haidian District, Beijing, China

[10]Department of Social Medicine and Health Education, School of Public Health, Peking University, Beijing, China

[2]Wuhu Coach Hospital, Wuhu, China

[3]North China University of Science and Technology, Tangshan, China

[4]City University of Hong Kong, Hongkong, China

[5]Capital Medical University, Beijing, China

[6]Chinese Center for Disease Control and Prevention, Beijing, China

[7]PopMed Technology Incorporate, Beijing, China

[8]School of Nursing, Peking University, Beijing, China

[9]Department of Endocrinology, Peking University First Hospital, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Xinying Sun, PhD
School of Public Health, Peking University, 38 Xueyuan Road, Haidian District, Beijing, China

Abstract

Background: Sustaining self-management is critical for optimizing clinical outcomes in individuals with type 2 diabetes mellitus (T2DM). Although digital health interventions (DHIs) have shown benefits for glycemic control and self-care, much of this evidence has focused on efficacy, and the behavioral mechanisms through which DHIs produce sustained effects remain insufficiently understood. Clarifying these mechanisms could inform the development of theory-driven interventions.

Objective: This study examined the longitudinal behavioral pathways through which Artificial Intelligence-based Health Education Accurately Linking System, a WeChat (Tencent)–based digital health program, influences T2DM self-management, using the Extended Multi-Theory Model (MTM) of health behavior change.

Methods: An explanatory sequential mixed methods prospective longitudinal cohort study was conducted among adults with T2DM (aged ≥18 y and proficient in WeChat use), recruited from 45 primary health care institutions in Beijing, China, between July 2023 and July 2024. Self-management behavior was assessed as the primary outcome using the Summary of Diabetes Self-Care Activities, and psychosocial determinants using the Extended MTM Scale and the Diabetes-related Skills Scale. Exploratory and confirmatory factor analyses assessed the construct validity of the Extended MTM Scale. Structural equation modeling examined longitudinal pathways among Artificial Intelligence-based Health Education Accurately Linking System users across baseline and 3, 6, and 12 months. For the qualitative phase, a purposive subsample was selected through maximum variation sampling based on baseline glycated hemoglobin; interviews were analyzed thematically until thematic saturation, and integrated with quantitative findings using a joint display.

Results: Of the 406 enrolled participants, 391 completed baseline assessments. The Extended MTM Scale demonstrated a 6-factor, 22-item structure with excellent internal consistency (Cronbach α=0.928) and satisfactory construct validity. The structural equation modeling showed satisfactory fit (CFI=0.984, RMSEA=0.036). Changes in the social environment (β=0.23, 95% CI 0.07 - 0.38; *P*=.003) and physical environment (β=0.25, 95% CI 0.10 - 0.40; *P*=.001) at baseline, and diabetes-related skills at month 6 (β=0.16, 95% CI 0.03 - 0.29; *P*=.01), were directly associated with self-management behavior at month 12, whereas behavioral confidence and emotional transformation showed no significant direct effects. Social environment changes were indirectly associated with behavioral confidence through participatory dialogue at month 3 (β=0.20, 95% CI 0.04 - 0.36; *P*=.01; β=0.66, 95% CI 0.57 - 0.74; *P*<.001). Thematic analysis of 17 interviews identified 3 domains: environmental context, cognitive processes, and attitudes and skills. Environmental factors converged across both data strands, while qualitative data expanded on the cognitive and attitudinal processes underlying sustained self-management.

XSL•FO
RenderX

Conclusions: This study is among the first to apply the Extended MTM framework to DHI-supported T2DM self-management, with environmental factors emerging as key drivers alongside selected cognitive, attitudinal, and skills-related processes. Complementing efficacy-focused research, it illuminates the psychosocial pathways underlying sustained self-management and refines the Extended MTM in digital health contexts. These insights can inform the design of theory-driven DHIs in primary care.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2300068952; https://www.chictr.org.cn/showprojEN.html?proj=170361

(JMIR Mhealth Uhealth 2026;14:e79081) doi:10.2196/79081

KEYWORDS

artificial intelligence; self-management; self care; type 2 diabetes mellitus; structural equation modeling

Introduction

Background

The prevalence of type 2 diabetes mellitus (T2DM) continues to rise and has emerged as a major global public health challenge [1]. This chronic condition not only imposes a substantial burden on individual health, marked by increased risk of disability, reduced quality of life, and shortened life expectancy [2], but also results in significant medical and socioeconomic costs [3]. According to the International Diabetes Federation, China has the highest number of people with diabetes and diabetes-related deaths worldwide, and the age of onset is trending younger [4].

As a lifelong condition, effective management of T2DM largely depends on patients' self-management capabilities [5], which encompass dietary regulation, regular physical activity, medication adherence, blood glucose monitoring, foot care, and routine eye examinations. However, awareness, treatment, and glycemic control rates in China remain suboptimal [6], highlighting a substantial gap in diabetes self-management practices. Therefore, the implementation of professionalized, standardized, and evidence-based self-management strategies is critically important for individuals living with T2DM.

In recent years, digital health interventions (DHIs) have shown great promise in supporting diabetes self-management [7,8]. By leveraging smartphones and other mobile devices, DHIs can overcome the time and space limitations of traditional community-based care, promote behavioral changes, and enhance the accessibility and equity of health services. To this end, our research team developed the Artificial Intelligence-based Health Education Accurately Linking System (AI-HEALS), implemented on the WeChat (Tencent) platform under the name Peking Diabetes Butler. This mobile application integrates knowledge graphs and natural language processing technologies and comprises an AI chatbot, tailored health education delivery, and blood glucose monitoring reminders [9]. It aims to provide patients with personalized and continuous support for diabetes self-management.

Theoretical Framework

Various theoretical models have been used in diabetes self-management interventions, including the Health Belief Model, Social Cognitive Theory, Self-Efficacy Theory, and social support frameworks [10]. These models emphasize strengthening individuals' cognitive, psychological, and social resources to address barriers to behavior change. However, most conventional models primarily focus on behavior initiation and provide more limited insight into the processes that support long-term maintenance. The Multi-Theory Model (MTM) of health behavior change has emerged as an integrative framework that draws on key elements from several traditional theories. MTM is particularly relevant to diabetes self-management, which involves multiple sustained behaviors such as dietary regulation, physical activity, blood glucose monitoring, and medication adherence. The model conceptualizes behavior change in 2 phases: behavioral initiation, including participatory dialogue, behavioral confidence, and changes in the physical environment, and behavioral sustenance, including emotional transformation, practice for change, and changes in the social environment [11,12].

To date, MTM has been applied in interventions targeting smoking cessation [13], dietary behavior [14], and physical activity [15], with evidence of benefit for quality of life and disease-related psychological outcomes [16,17]. However, most applications have been conducted in offline settings, and empirical evidence supporting MTM in DHIs remains limited. In addition, previous studies suggest that the explanatory performance of MTM varies across behavioral contexts and populations [11], indicating a need to examine its applicability in specific settings. In diabetes self-management, sustained behavior depends not only on motivation and environmental support but also on patients' ability to carry out routine self-care tasks, such as blood glucose monitoring, dietary adjustment, and medication management [18]. Therefore, the present study used MTM as the core theoretical framework and incorporated diabetes-related skills as an additional construct to better capture the practical demands of diabetes self-management in a digital intervention context.

Study Objective

Therefore, this mixed methods study aimed to examine the behavioral mechanisms through which AI-HEALS influences diabetes self-management using an Extended MTM framework and to integrate quantitative and qualitative evidence regarding these mechanisms.

XSL•FO
RenderX

Methods

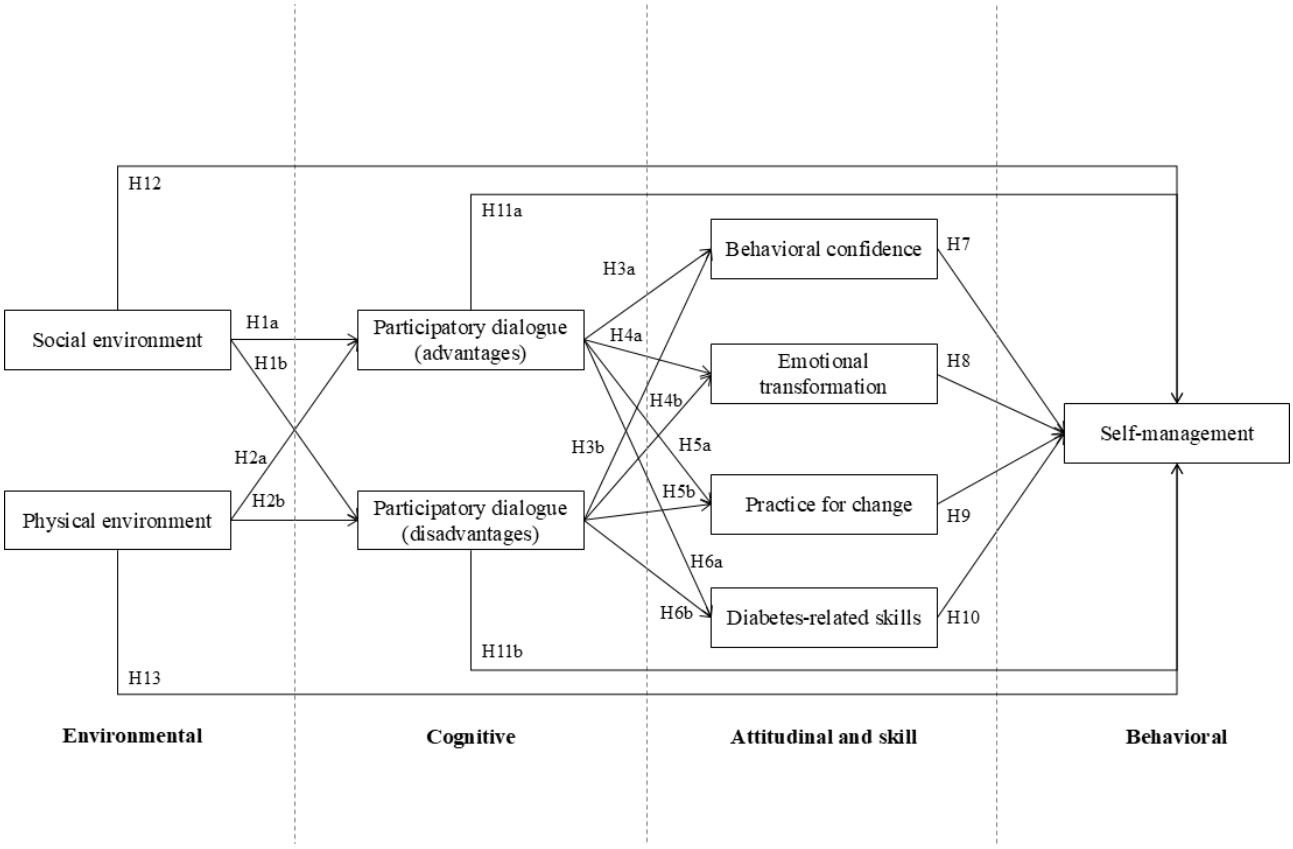
Study Design

This study was a mixed methods prospective longitudinal cohort substudy within the AI-HEALS program (Chinese Clinical Trial Registry: ChiCTR2300068952; registered March 2023), examining the behavioral mechanisms underlying diabetes self-management. An explanatory sequential mixed methods design was used, in which data were collected prospectively. Reporting followed the Good Reporting of A Mixed Methods Study guideline [19]. The completed checklist is provided in Checklist 1.

The quantitative phase comprised 2 components. First, the psychometric properties and construct validity of the Extended MTM Scale were assessed using baseline data through exploratory and confirmatory factor analyses. Second, to examine the hypothesized behavioral pathways, AI-HEALS users were followed over 12 months, with self-report assessments at 4 time points (baseline and 3, 6, and 12 months), and structural equation modeling was applied to these repeated measures. In the subsequent qualitative phase, a purposive subsample of the AI-HEALS users was interviewed to interpret and contextualize the quantitative findings.

The Extended MTM framework was organized into 4 domains—environmental, cognitive, attitudinal and skills, and behavioral—reflecting the hypothesized progression from contextual conditions to proximal individual-level processes and, ultimately, to self-management behavior. Figure 1 presents the hypothesized model, comprising 13 hypothesized pathways (H1–H13) linking environmental factors to cognitive processes, cognitive processes to attitudinal and skill-related constructs, and all preceding domains to self-management behavior. This framework guided both the quantitative path analysis and the qualitative inquiry.

Figure 1. Hypothesized model of the extended Multi-Theory Model of health behavior change for diabetes self-management behavior. The hypothesized model comprises 13 pathways (H1–H13) linking environmental, cognitive, attitudinal and skill, and behavioral domains among adults with type 2 diabetes mellitus. H1–H2 represent pathways from environmental factors to participatory dialogue; H3–H6 represent pathways from participatory dialogue to attitudinal and skill-related constructs; H7–H10 represent pathways from attitudinal and skill-related constructs to self-management behavior; H11–H13 represent hypothesized direct pathways from environmental and cognitive constructs to self-management behavior.



Intervention

Participants received standard diabetes primary care supplemented by AI-HEALS, a personalized DHI delivered via WeChat. AI-HEALS comprised an interactive AI chatbot, tailored health education delivery, and individualized blood glucose monitoring reminders. During the first 3 months, participants received at least one tailored self-management message per day and were provided with an operational guide in the first week. After the initial 3 months, the program shifted from proactive daily message delivery to a maintenance phase in which participants retained access to the chatbot, educational resources, and reminder functions.

Setting and Participants

A total of 406 adults with T2DM were enrolled from 45 primary health care institutions in Daxing and Shunyi districts of Beijing. Eligibility criteria included age≥18 years, diagnosis of T2DM,

XSL•FO
RenderX

residency in Beijing with less than one month of annual travel outside the city, smartphone ownership and proficiency in WeChat use, no use of psychiatric medications, and the ability to independently complete questionnaires.

Baseline data from participants with complete assessments were randomly split into 2 subsets: one for item selection via exploratory factor analysis (EFA), and the other for scale validation via confirmatory factor analysis (CFA). Longitudinal structural equation modeling (SEM) was conducted among participants who used AI-HEALS, examined across baseline, 3, 6, and 12 months.

For the interviews, participants were purposively sampled from among AI-HEALS users who had completed the quantitative assessments. A maximum variation sampling strategy based on glycated hemoglobin (HbA_{1c}) levels measured at baseline was applied to capture diverse diabetes management profiles. Recruitment continued until data sufficiency was achieved across the range of sampling variation.

Quantitative Data Collection

Quantitative data were collected using coded questionnaires between July 2023 and July 2024. At baseline, all eligible participants completed the Extended MTM scale for construct validity assessment. Factor analyses were conducted to assess the construct validity of the Extended MTM Scale. Participants who used AI-HEALS completed follow-up assessments at 3, 6, and 12 months.

Qualitative Data Collection

Qualitative data were collected through one-on-one telephone interviews conducted by the author at the 6-month follow-up (January 8 and February 20, 2024). After obtaining verbal informed consent, interviews were conducted privately with only the interviewer and participant present. Interviews continued until thematic saturation was reached, defined as the point at which no new themes or codes emerged from successive interviews. A semistructured interview guide was used to explore participants' experiences with AI-HEALS, covering usage behaviors, perceived usefulness and accessibility, perceived impact on self-management (including diet, physical activity, blood glucose monitoring, and medication adherence), and suggestions for improvement. The full interview guide is provided in Multimedia Appendix 1.

Ethical Considerations

The study was approved by the Peking University Biomedical Ethics Committee (IRB00001052-22058). Informed consent was obtained from all participants prior to participation in any component of the study. Written consent was obtained before the completion of the quantitative surveys, and verbal consent was obtained prior to the telephone interviews. Prior to analysis, all data were deidentified by replacing personal identifiers with coded study identifiers, with the linkage key stored separately; no images or materials in the manuscript or supplementary files contain information that could identify individual participants. Participants who completed all follow-up assessments in the quantitative phase received a gift valued at approximately 40 Chinese Yuan (CNY; approximately US $5 - 6 [1]), and those who participated in the qualitative interviews received a gift valued at approximately 100 CNY (approximately US $13 - 14) upon completion.

Measurements

Baseline data included sociodemographic and behavioral variables (eg, sex, age, education, marital and employment status, household income, smoking, and alcohol consumption). The following validated self-report instruments were administered: (1) Chinese Version of the Extended MTM Scale (Developed specifically for this study based on the MTM [12]), this 5-point Likert scale (1=strongly disagree, 5=strongly agree) assesses cognitive and behavioral constructs associated with diabetes self-management. Scale development and psychometric evaluation, including reliability and validity testing, are described in the Results section. (2) Diabetes-related Skills Scale: This 12-item scale evaluates patients' competencies in dietary planning, physical activity adjustment, blood glucose monitoring, and hypoglycemia prevention [20]. Responses are scored on a 5-point scale, with higher scores indicating better skills. In the current study, the scale showed excellent internal consistency (Cronbach α=0.912). (3) Summary of Diabetes Self-Care Activities: This 10-item instrument assesses the frequency of diabetes self-management behaviors over the past week, across 5 domains: general diet, specific diet, exercise, blood glucose testing, and foot care [21]. Responses range from 0 to 7 days, with total scores from 0 to 77; higher scores reflect better self-care. Internal consistency in this sample was acceptable (Cronbach α=0.736).

Statistical Analyses

To validate the Extended MTM scale, baseline data from participants with complete assessments were randomly split into 2 subsets for item screening and confirmatory evaluation. Item selection was based on EFA, item-total correlations, and internal consistency (Cronbach α). Construct validity was further assessed using CFA.

Sample size for the path analysis was determined based on empirically derived SEM criteria, including a ratio of approximately 10 - 20 observations per estimated parameter and a minimum sample size of 150 - 200 [22,23]. The Extended MTM model included 9 - 12 free parameters in the structural component; therefore, a minimum sample size of approximately 180 was required. Path analysis using maximum likelihood estimation was conducted on longitudinal data from AI-HEALS users to estimate direct and indirect effects among MTM constructs, diabetes-related skills, and self-management behavior. Little Missing Completely at Random (MCAR) test was used to examine the missing data mechanism, and missing data were handled using full information maximum likelihood estimation. Pearson correlation coefficients were calculated to examine associations between variables. Model fit was evaluated using standard indices, including the comparative fit index (CFI), Tucker–Lewis index (TLI), and root mean square error of approximation (RMSEA), with acceptable thresholds defined as CFI and TLI>0.90 and RMSEA<0.08. Statistical significance was set at P<.05. All analyses were performed using IBM SPSS Statistics 27.0 (IBM Corp), IBM SPSS Amos 23.0 (IBM Corp), and Mplus 8.3 (Muthén & Muthén).

XSL•FO
RenderX

Qualitative data were analyzed using thematic analysis following the approach described by Braun and Clarke [24], adopting a primarily inductive coding strategy within a framework-informed analytic approach. Audio recordings were transcribed verbatim within 24 hours of each interview, with nonverbal cues (eg, tone, pauses, laughter) noted to preserve contextual meaning. Where local dialect terms were encountered, clarification was sought from native speakers to ensure accurate interpretation. A large language model (Doubao, version 7.3.0) was used solely for minor grammatical refinement of transcripts without altering meaning. Two researchers (YJ and BK) independently verified all transcripts against the recordings for accuracy. Transcripts were imported into NVivo 12.0 (QSR International) for analysis. Researchers first familiarized themselves with the data through repeated reading. Initial codes were generated inductively and recorded in a coding log to ensure auditability and traceability of analytic decisions. The coding log was iteratively consolidated into subthemes and higher-order themes through constant comparison within and across cases, with regular team discussions held to resolve discrepancies and reach consensus. The Extended MTM domains were used as an interpretive lens to organize emergent themes and facilitate integration with quantitative findings.

Integration of quantitative and qualitative findings was conducted using a joint display approach, mapping MTM constructs to qualitative themes to examine areas of confirmation and expansion [25,26].

Results

Participant Characteristics

Of 406 enrolled participants, 15 were lost to follow-up over 12 months (attrition rate 3.7%), resulting in 391 participants with complete baseline data. These baseline data were randomly split for item selection via exploratory factor analysis (EFA; n=202) and scale validation via confirmatory factor analysis (CFA; n=189). The longitudinal structural equation model was subsequently conducted among the 202 AI-HEALS users across baseline and 3, 6, and 12 months. The mean age was 56.6 (SD 11.2) years, 207/391 (53%) participants were female (Table 1), and the mean HbA_{1c} was 7.29% (SD 1.61%).

XSL•FO
RenderX

Table . Demographics and baseline characteristics of study participants[a] (N=202). Data collected at baseline (July 2023). .

Variables	Values
Sex, n (%)	
Male	95 (47)
Female	107 (53)
Age (years), mean (SD)	56.6 (11.2)
HbA_{1c},[b] (%), mean (SD)	7.29 (1.61)
Disease duration (years), n (%)	
≤5	57 (28.2)
6 - 10	70 (34.7)
11 - 20	58 (28.7)
≥21	17 (8.4)
Diabetes management, n (%)	
Diet	11 (5.4)
OADs[c]	142 (70.3)
Injectable insulin	4 (2)
OADs + injectable insulin	45 (22.3)
BMI, kg/m², n (%)	
Underweight (<18.5)	1 (0.5)
Normal weight (18.5 - 23.9)	60 (29.7)
Overweight (24.0 - 27.9)	75 (37.1)
Obese (≥28.0)	66 (32.7)
Waist (cm), n (%)	
Normalcy (<80)	11 (5.4)
Exceeding standard (80-94)	96 (47.5)
Obese (>94)	95 (47)
Per capita monthly income, CNY[d,e], n (%)	
<3000	58 (28.7)
3000 - 9000	114 (56.4)
>9000	30 (14.9)
Other diseases, n (%)	
0	44 (21.8)
1	49 (24.3)
≥2	109 (54)
Family history of diabetes, n (%)	
None	93 (46)
Yes	109 (54)
Smoking status, n (%)	
Non-smoker	139 (68.3)
Former smoker (quit)	10 (5)
Current smoker	54 (26.7)
Alcohol consumption status, n (%)	
Non-drinker	117 (57.9)

XSL•FO
RenderX

Variables	Values
Former drinker (quit)	18 (8.9)
Current drinker	67 (33.1)

[a]Characteristics pertain to participants with type 2 diabetes mellitus (T2DM) who used Artificial Intelligence-based Health Education Accurately Linking System (AI-HEALS), recruited from community health centers and outpatient clinics in Beijing, China.

[b]HbA_{1c}: glycated hemoglobin.

[c]OAD: oral antidiabetic drug.

[d]CNY=Chinese Yuan.

[e]Monthly income ≤3000 Chinese Yuan (CNY) (approximately ≤ US $400); 3000 - 9000 CNY (approximately US $400 - 1200); ≥9000 CNY (approximately ≥ US $1200)

For the qualitative phase, thematic saturation was achieved after 17 interviews, with no new themes emerging in the final interviews. Among them, 10 of 17 (58.8%) were women and 7 of 17 (41.2%) were men. Ten of 17 (58.8%) were aged <60 years and 7 of 17 (41.2%) were aged ≥60 years. HbA1c was <7% in 8 of 17 (47.1%) and ≥7% in 9 of 17 (52.9%). Monthly income ranged from ¥3000-¥9000 (approximately US $400 - 1,200) for 10 of 17 (58.8%) participants, and 10 of 17 (58.8%) had completed high school or vocational education. Interview duration ranged from 12 - 45 minutes.

Scale Development and Validation of the Extended MTM Scale

Baseline data were randomly split for item selection via EFA and scale validation via CFA. All items demonstrated satisfactory psychometric properties, with coefficient of variation values exceeding 0.3 and item-total correlation coefficients above 0.4. EFA with varimax rotation was conducted based on the original 7-dimension MTM framework. As practice for change and emotional transformation loaded onto the same factor, expert consultation and theoretical review were undertaken, resulting in the exclusion of practice for change due to conceptual overlap with emotional transformation. The final 6-factor structure encompassed changes in the physical environment, participatory dialogue (advantages and disadvantages), behavioral confidence, emotional transformation, and changes in the social environment, with a cumulative variance explained of 86.34%, well above the recommended 50% threshold. The retained 22-item scale demonstrated excellent internal consistency (Cronbach α=0.928; greatest lower bound =0.967). CFA confirmed satisfactory structural validity (CFI=0.979, RMSEA=0.052; Multimedia Appendix 2). It should be noted that the exclusion of practice for change limited the availability of direct quantitative evidence for this dimension.

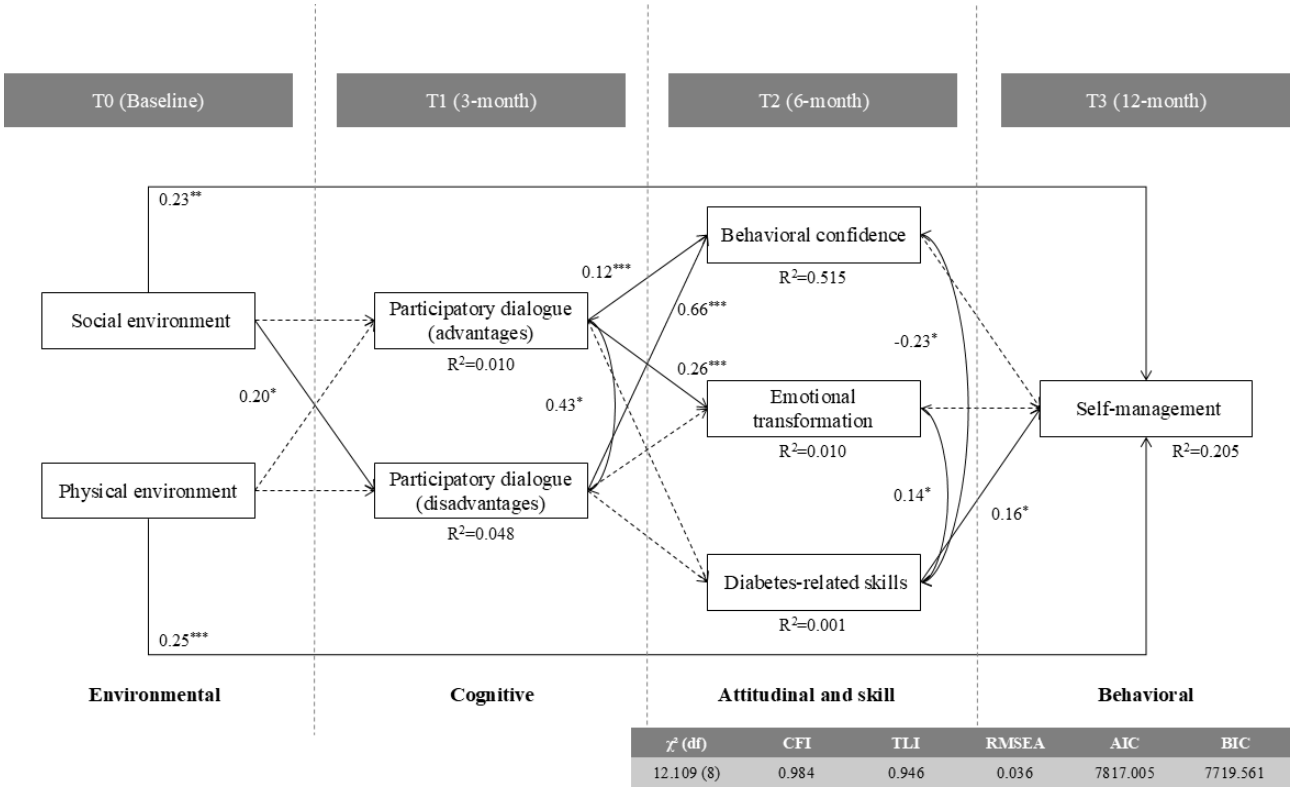
Quantitative Findings

Descriptive statistics and correlations among Extended MTM constructs, diabetes-related skills, and self-management behaviors are summarized in Multimedia Appendix 2. The Extended MTM framework includes 4 levels: environmental (changes in the physical and social environment), cognitive (participatory dialogue advantages and disadvantages), attitudinal and skill (behavioral confidence, emotional transformation, diabetes-related skills), and behavioral (self-management). At baseline, the mean scores for changes in the physical and social environment were 9.64 (SD 3.57) and 12.35 (SD 3.03), respectively. At month 6, emotional transformation averaged 12.27 (SD 2.91), diabetes-related skills 37.66 (SD 10.03), and behavioral confidence 12.75 (SD 2.64). Self-management behaviors at month 12 had a mean score of 49.70 (SD 17.16). Emotional transformation showed strong correlation with diabetes-related skills (r=0.744; P=.008) and moderate correlation with behavioral confidence (r=0.474; P=.005). Behavioral confidence was also positively correlated with diabetes-related skills (r=0.471; P=.004).

Among the 202 AI-HEALS users included in the SEM, complete data were available for 173 (85.6%) at month 12. Little MCAR test indicated that the data were missing completely at random (χ^2_{272}=270.54; P=.51); missing data were therefore handled using full information maximum likelihood estimation. As shown in Figure 2, changes in the social environment (β=0.23, 95% CI 0.07 - 0.38; P=.003) and physical environment (β=0.25, 95% CI 0.10 - 0.40; P=.001) at baseline, and diabetes-related skills at month 6 (β=0.16, 95% CI 0.03 - 0.29; P=.01) were directly associated with self-management behaviors at month 12, supporting H12, H13, and H10. In contrast, behavioral confidence and emotional transformation showed no significant direct effects on self-management behaviors, leaving H7 and H8 unsupported. Changes in the social environment at baseline were significantly associated with participatory dialogue (disadvantages) at Month 3 (β=0.20, 95% CI 0.04 - 0.36; P=.01; H1b), and participatory dialogue (disadvantages) was in turn significantly associated with behavioral confidence at month 6 (β=0.66, 95% CI 0.57 - 0.74; P<.001; H3b). As participatory dialogue (disadvantages) was reverse-coded, these positive coefficients reflect negative underlying effects. Participatory dialogue (advantages) at month 3 was additionally associated with behavioral confidence (β=0.12, 95% CI 0.02 - 0.23; P=.02; H3a) and emotional transformation (β=0.26, 95% CI 0.11 - 0.40; P=.001; H4a) at month 6. All remaining hypothesized pathways were non-significant (Multimedia Appendix 2).

XSL•FO
RenderX

Figure 2. Longitudinal structural equation model of diabetes self-management behavior based on the extended Multi-Theory Model (MTM) of health behavior change. The model was estimated among adults with type 2 diabetes mellitus who used AI-HEALS (n=202), recruited from community health centers and outpatient clinics in Beijing, China, across 4 time points: baseline (T0), month 3 (T1), month 6 (T2), and month 12 (T3). Standardized path coefficients (β) are shown. Solid lines indicate significant paths; dashed lines indicate non-significant paths. R^2 values represent explained variance. *P<.05, **P<.01, ***P<.001. AIC: Akaike information criterion; AI-HEALS: Artificial Intelligence-based Health Education Accurately Linking System; BIC: Bayesian information criterion CFI: comparative fit index; MTM: Multi-Theory Model; RMSEA: root mean square error of approximation; TLI: Tucker–Lewis index;



χ^2 (df)	CFI	TLI	RMSEA	AIC	BIC
12.109 (8)	0.984	0.946	0.036	7817.005	7719.561

The model demonstrated satisfactory fit (χ^2_8=12.109, CFI=0.984, TLI=0.946, RMSEA=0.036), supporting the hypothesized structure of the Extended MTM framework.

Qualitative Findings

Based on the Extended MTM, qualitative findings were organized into 3 hierarchical domains: environmental, cognitive, and attitudinal and skills, encompassing 7 themes and 18 subthemes (Multimedia Appendix 1). Below is a concise summary of each domain.

Environmental Domain

Changes in both physical and social environments influenced participants' engagement in diabetes self-management. At the level of the physical environment, accessibility and usability of digital tools were key determinants of engagement. Participants were more likely to apply health information when content was perceived as relevant, clear, and easy to use. For example, one participant described using dietary information to guide daily decisions:

> *If I want to eat something, I'll check whether it's suitable for people with diabetes. If the sugar content is high, I won't buy it.* [T-BJ-360, female, 62 years]

However, limitations in information reliability and clarity were also reported. Some participants expressed uncertainty about the trustworthiness of certain content, which could reduce confidence in using the platform. As one participant noted,

> *Some medication info seems unreliable.* [T-BJ-033, male, 58 years]

This suggests that inconsistencies or lack of transparency in information presentation may hinder sustained engagement.

At the level of the social environment, family involvement and broader social support played a reinforcing role. Emotional and practical support from family members enhanced motivation and facilitated adherence to self-management behaviors. For instance, one participant stated,

> *My daughter and son both really care about this.* [T-BJ-194, female, 72 years]

Participants also highlighted the importance of wider societal support for health promotion:

> *I truly hope the government can promote health knowledge among young people.* [T-BJ-194, female, 72 years]

Cognitive Domain

Interactions with the digital platform, together with input from health care providers and social networks, supported participants' understanding and decision-making processes. Participants described how exposure to structured information helped them reassess prior beliefs and make more informed choices. For example, one participant noted,

XSL•FO
RenderX

> *I've read a lot of the content. I used to think certain foods were low in sugar or sugar-free, but now I know they're not.* [T-BJ-182, female, 51 years]

This reflects a process of cognitive correction and improved health literacy.

In addition, the combination of digital information and offline guidance appeared to enhance perceived usefulness and credibility. As one participant stated,

> *This content, combined with the community doctor's talks, fully meets my needs.* [T-BJ-173, female, 67 years]

Another participant emphasized the practical value of the platform:

> *I've learned so many things through this. I'm sure it will help me solve a lot of problems going forward.* [T-BJ-035, male, 61 years]

Attitudinal and Skills Domain

This domain encompassed behavioral confidence, emotional transformation, practice for change, and diabetes-related skills. Behavioral confidence emerged as a central factor influencing proactive engagement. Participants who perceived themselves as capable of managing their condition were more likely to adopt and maintain behavioral changes. For example, one participant stated,

> *If you want to live long and live well, you must have confidence.* [T-BJ-194, female, 72 years]

Trust in the platform also contributed to this confidence, as reflected in the following statement:

> *The materials you provided feel really trustworthy…* [T-BJ-407, male, 49 years]

Emotional transformation was another important process, whereby participants shifted from anxiety or passive coping to more proactive self-management. Structured information and repeated engagement appeared to support this transition. As one participant noted,

> *There's no use being anxious. Since I've got the disease, I might as well manage it properly...* [T-BJ-182, female, 51]

Participants also described practice for change through reminders, self-monitoring, and the gradual incorporation of self-management behaviors into daily routines. For instance, one participant reported,

> *Just last night, I forgot to check, and the Peking Diabetes Butler reminded me.* [T-BJ-184, male, 56 years]

In parallel, participants described improvements in diabetes-related skills, particularly in blood glucose monitoring, diet, medication use, and physical activity. As one participant explained,

> *My workouts are regular now... I used to exercise blindly, but now I know what I'm doing.* [T-BJ-182, female, 51 years]

Nonetheless, some participants identified limitations in content presentation, particularly regarding clarity and format. For example, one participant suggested,

> *Some of the exercise illustrations aren't clear enough. Videos would be much better...* [T-BJ-182, female, 51 years]

This indicates that insufficiently intuitive or engaging content may reduce usability and limit the effectiveness of skill acquisition.

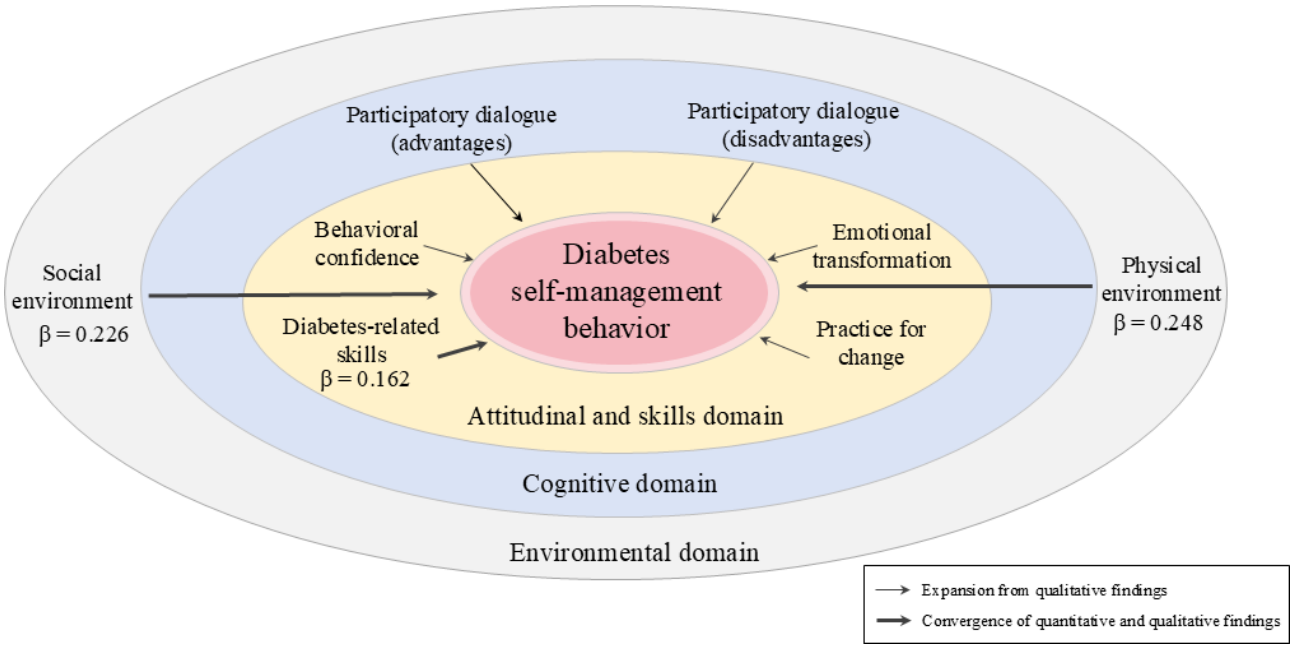
Building on the subtheme analysis, we assessed whether overarching themes differed by participants' HbA_{1c} levels. No clear differences in the overarching themes were observed across interview participants with different HbA_{1c} levels, although some variation in emphasis was noted in individual accounts.

Integration of Quantitative and Qualitative Findings

As illustrated in Figure 3 and Multimedia Appendix 3, integration of quantitative and qualitative data revealed both convergent and complementary patterns across the Extended MTM domains.

XSL•FO
RenderX

Figure 3. Integrated model of quantitative and qualitative findings on diabetes self-management behavior based on the extended Multi-Theory Model of health behavior change. Based on an explanatory sequential mixed methods study among adults with type 2 diabetes mellitus who used Artificial Intelligence-based Health Education Accurately Linking System, recruited in Beijing, China (n=202). Thin arrows indicate pathways derived from qualitative findings, whereas thick arrows indicate pathways supported by both quantitative and qualitative evidence. Standardized path coefficients (β) are shown for direct associations with diabetes self-management behavior identified in the quantitative model.



Environmental factors showed the strongest convergence. SEM confirmed direct associations of physical and social environment changes with self-management behavior at month 12 (β=0.25, 95% CI 0.10-0.40 and β=0.23, 95% CI 0.07-0.38), and qualitative accounts described how access to digital tools, self-management resources, and family and community support facilitated behavioral engagement. Diabetes-related skills similarly showed convergence, with a direct quantitative effect (β=0.16, 95% CI 0.03-0.29) corroborated by participants' descriptions of applying glucose monitoring, dietary, and medication management skills in daily practice.

Qualitative findings expanded on constructs showing limited direct effects in the SEM. Behavioral confidence and emotional transformation were not significant direct predictors of self-management behavior, yet participants described both as present and influential in their self-management experience. Practice for change, which was excluded from the quantitative model during scale revision, was nonetheless frequently referenced by participants as part of their behavioral routines.

Participatory dialogue showed complementary patterns across both strands. Quantitative pathways linked social environment changes to behavioral confidence and emotional transformation via participatory dialogue at month 3 and month 6. Qualitative accounts described interactions with the platform, health care providers, and peers as sources of knowledge, motivation, and decision-making support.

Discussion

Principal Findings

This study used an explanatory sequential mixed methods design. The integrated findings provided partial confirmation and substantial expansion of the Extended MTM framework in the context of DHIs for diabetes self-management. SEM confirmed that physical and social environmental factors and diabetes-related skills had direct positive effects on self-management behavior, with environmental factors showing particularly prominent influence. Qualitative findings expanded these results by elucidating how participatory dialogue, behavioral confidence, emotional transformation, and practice for change operated alongside environmental support to facilitate sustained behavior change. Together, these findings underscore the prominent role of environmental factors, complemented by cognitive, attitudinal, and skills-related processes, in shaping effective self-management. This study represents the first application of the MTM framework to diabetes self-management and extends it based on the behavioral characteristics of this context.

Results from the SEM showed that changes in the physical and social environment had direct positive associations with diabetes self-management behaviors [27,28]. Environmental constructs showed clearer direct associations than the psychological constructs included in the model, whose pathways were not statistically significant. This pattern may be interpreted from a socioecological perspective, which views self-management behavior as embedded within broader physical and social environments rather than determined solely by individual-level psychological processes [29]. In this model, physical and social environmental changes were therefore positioned as antecedent contextual conditions for self-management, whereas psychological and skills-related constructs were modeled as downstream individual-level processes that develop within, and partly in response to, these conditions. In AI-HEALS, accessible information, reminders, and ongoing support may have strengthened the environmental conditions under which self-management occurs, which may help explain the

XSL•FO
RenderX

prominence of environmental constructs in the final model. This suggests that DHIs may not act solely by strengthening internal determinants such as confidence or emotion-related processes, but may also increase the salience of contextual determinants, consistent with evidence that context-aware digital behavior change interventions dynamically tailor support based on real-world factors [30]. From an intervention perspective, strengthening supportive environmental conditions may therefore be an important entry point for promoting sustained behavioral engagement [31].

Diabetes-related skills were also found to have a direct positive influence on self-management behaviors, aligning with previous studies [32]. Diabetes self-management education should integrate behavior change techniques, such as goal setting and self-monitoring, with skills training to enhance patients' practical capabilities [33]. Evidence suggests that skill internalization interventions can support behavioral persistence [34]; patients who acquire blood glucose monitoring skills early demonstrate significantly better self-management awareness than those without systematic training [35]. Neuroscientific studies suggest that skilled individuals show reduced activation in decision-related brain regions like the prefrontal cortex [36]. Skill internalization may shorten the neural pathway between cognitive appraisal and behavioral execution, thereby improving the efficiency and automatization of behavioral performance [37]. These findings suggest that practical skill acquisition may represent a more proximal determinant of behavioral outcomes than attitudinal processes alone, with implications for the design of DHIs that prioritize applied skills training alongside informational content.

This study found that changes in the social environment indirectly enhanced behavioral confidence through participatory dialogue (disadvantages), while participatory dialogue (advantages) directly improved behavioral confidence and emotional transformation. These findings align with previous research. In T2DM management, participatory strategies such as interactive education and motivational interviewing have been shown to strengthen behavioral confidence and emotional resonance, thereby improving patients' motivation for self-management [38,39]. Encouraging a focus on positive emotions, such as hope and achievement, can reduce fear and helplessness and promote behavioral engagement [40]. Experience sharing and emotional support in online communities may further enhance self-efficacy as an extension of social support [41]. However, neither behavioral confidence nor emotional transformation demonstrated a significant direct effect on self-management behavior in the SEM, suggesting that their influence may operate indirectly through skills development and routine consolidation rather than directly driving behavioral outcomes. Prior meta-analyses have indicated that knowledge-based interventions exert limited behavioral impact in the absence of accompanying skills training [42], which may partly account for this pattern. This interpretation was further supported by qualitative accounts, in which participants described confidence and emotional regulation as enabling conditions for sustained engagement rather than immediate behavioral triggers.

Qualitative findings further suggest that improvements in diabetes self-management are shaped by the interplay between environmental accessibility, cognitive support, and the development of attitudes and practical skills. At the environmental level, the accessibility and usability of the AI-HEALS platform appear to provide a foundation for behavioral engagement, facilitating access to structured health information and ongoing support. In addition to these system-level features, participants frequently described the importance of family involvement and professional guidance in sustaining behavioral continuity, indicating that broader social support contexts may complement DHIs in reducing barriers to long-term engagement. These observations are consistent with prior evidence indicating that supportive environments can facilitate sustained self-management in patients with T2DM [31,43]. From a cognitive perspective, interaction with the platform appeared to support processes of understanding, appraisal, and decision-making. The delivery of timely and structured information, together with interactive feedback, may help reduce informational asymmetry and enhance users' confidence in interpreting and applying health guidance. Rather than functioning as a standalone driver, this cognitive support seems to operate in conjunction with external reinforcement from health care providers and personal networks. The use of interactive digital platforms and conversational technologies has been shown to enhance user trust and engagement in health communication [44,45]. At the level of attitudes and skills, behavioral confidence and emotional adjustment emerged as closely linked processes that support sustained self-management. Participants described how increased confidence enabled them to translate concerns related to glycemic control into actionable strategies, while repeated engagement with the platform facilitated the gradual internalization of self-management skills. Over time, this process contributed to the establishment of routine practices across key domains, including glucose monitoring, dietary regulation, physical activity, and medication adherence, reflecting a transition towards more autonomous self-regulation. This interpretation is supported by prior research highlighting the role of emotional regulation and self-efficacy in sustaining behavioral change [46,47]; as well as behavior change techniques embedded in DHIs [48]. Notably, practice for change—though excluded from the quantitative model due to scale revision—was consistently described by participants as a mechanism linking knowledge acquisition to behavioral maintenance, suggesting its continued theoretical relevance within the Extended MTM framework. The absence of quantitative evidence for this dimension reflects a methodological limitation of the scale revision process rather than a theoretical dismissal of its role. Future studies should consider retaining or separately operationalizing practice for change to enable fuller quantitative testing of the Extended MTM framework.

Building on these observations, variation in participants' accounts suggested that behavioral processes are shaped by the relative prominence of barriers and self-regulatory capacity. Participants with comparatively higher HbA_{1c} at month 6 more often emphasized constraints related to symptom burden and practical barriers, whereas those with lower HbA_{1c} levels tended

XSL•FO
RenderX

to highlight established routines and greater behavioral confidence [49]. These differences illustrate how behavioral priorities and self-regulatory strategies vary across individuals, highlighting the dynamic nature of diabetes self-management processes. In this context, DHIs may benefit from adapting to these differing needs, particularly by strengthening external support and providing more tailored guidance for individuals facing greater behavioral constraints [50,51]. These observations also highlight areas for further optimization of the AI-HEALS system, as qualitative findings indicated concerns regarding the reliability of certain medication information and the clarity of visual content, alongside limited adaptability to individual needs, particularly among older adults and those with lower digital literacy. Addressing these issues may enhance system effectiveness through improved interface design, incorporation of multimodal content such as video-based guidance, and greater adaptability to diverse user needs, while closer integration with broader care systems and real-world support structures, including community-based platforms [52,53], may further support sustained engagement in diabetes self-management.

Grounded in the Extended MTM, this study demonstrates that AI-HEALS supports diabetes self-management through a pathway of environmental facilitation, cognitive reinforcement, and attitudinal and skill consolidation. Across both data strands, environmental factors emerged as the most consistently supported drivers of behavior change, while cognitive and attitudinal processes showed complementary roles that were more fully captured through qualitative inquiry. The convergence of findings across these domains supports the applicability of the Extended MTM as an explanatory framework for DHI-supported behavior change in chronic disease management.

Limitations

There are several limitations that should be recognized. First, participants were recruited from community health centers in Beijing, which may limit the generalizability of findings to other geographic, cultural, or health care contexts. Second, the sample may have been subject to selection bias, as less cooperative or less digitally proficient patients may have been systematically excluded. Third, reliance on self-reported outcome measures may introduce recall bias, potentially affecting the accuracy of self-management behavior assessments. Fourth, although the longitudinal design and SEM allowed examination of temporal pathways, SEM identifies associations consistent with hypothesized causal structures but cannot establish causality definitively. Fifth, the findings are based on AI-HEALS, a single program delivered via WeChat, and may not generalize to other digital health programs with different technological platforms, user interfaces, or engagement strategies. Sixth, the exclusion of the practice for change dimension during scale revision precluded its quantitative examination, limiting the completeness of the Extended MTM framework testing.

Conclusions

This study is among the first to apply the Extended MTM framework to diabetes self-management, demonstrating that environmental factors, alongside selected cognitive, attitudinal, and skills-related processes, play an important role in sustaining behavior change among AI-HEALS users. While environmental factors and diabetes-related skills showed direct associations with self-management behavior, behavioral confidence and emotional transformation appeared to operate indirectly, illuminating psychosocial mechanisms that efficacy-focused evaluations alone cannot capture. These findings contribute to the theoretical refinement of the Extended MTM in digital health contexts and suggest that prioritizing environmental support and skill-building in the design of digital health programs may help sustain patient engagement and self-management in primary care settings.

Acknowledgments

We thank the trial steering experts, participating patients, and all staff involved in the development, implementation, and maintenance of AI-HEALS for their support of this study. The authors declare the use of generative AI in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to generative AI (GenAI) tools under full human supervision: proofreading and editing. The GenAI tool used was Doubao (version 7.3.0). Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes. The GenAI tool Doubao (version 7.3.0) was used to assist with language refinement of qualitative interview transcripts during data processing. The AI was applied solely to improve clarity and readability and did not generate, modify, or analyze study data. All AI-assisted outputs were reviewed and verified by the research team against the original audio recordings to ensure accuracy. No AI was used for data analysis, interpretation, or independent decision-making.

Funding

This study was supported by Capital's Funds for Health Improvement and Research (grants 2022-1G-4252 and 2026-2R-4403). The funder had no role in the design of the study; collection, analysis, and interpretation of data; writing of the manuscript; or the decision to submit the manuscript for publication.

Data Availability

The data sets generated and analyzed during this study are not publicly available because participants were assured that the raw data would remain confidential. Deidentified data may be made available from the corresponding author on reasonable request, subject to institutional review and data governance requirements.

XSL•FO
RenderX

Authors' Contributions

Writing–review & editing, writing–original draft, supervision, project administration, methodology, investigation, formal analysis, data curation, conceptualization: YW
Writing – review & editing, Writing – original draft, methodology, formal analysis, data curation: YN
Writing – review & editing, methodology, formal analysis, data curation: YJ
Writing – review & editing, supervision, investigation: ZX
Writing – review & editing, supervision, methodology, investigation, conceptualization: HM
Writing – review & editing, supervision, investigation: PC
Supervision, investigation, data curation: XG, BK, GY
Supervision, conceptualization: PL, XG
Conceptualization, methodology, supervision, writing – review & editing: AM
Writing review & editing, supervision, project administration, methodology, investigation, funding acquisition, formal analysis, data curation, conceptualization: XS

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview material.
[DOCX File, 31 KB - mhealth_v14i1e79081_app1.docx]

Multimedia Appendix 2
Supplementary quantitative data.
[DOCX File, 136 KB - mhealth_v14i1e79081_app2.docx]

Multimedia Appendix 3
Joint display of quantitative and qualitative findings.
[DOCX File, 19 KB - mhealth_v14i1e79081_app3.docx]

Checklist 1
GRAMMS checklist.
[DOCX File, 16 KB - mhealth_v14i1e79081_app4.docx]

References

1. Ahmad E, Lim S, Lamptey R, Webb DR, Davies MJ. Type 2 diabetes. Lancet 2022 Nov 19;400(10365):1803-1820. [doi: 10.1016/S0140-6736(22)01655-5] [Medline: 36332637]
2. Salehidoost R, Mansouri A, Amini M, Aminorroaya Yamini S, Aminorroaya A. Diabetes and all-cause mortality, a 18-year follow-up study. Sci Rep 2020 Feb 21;10(1):3183. [doi: 10.1038/s41598-020-60142-y] [Medline: 32081921]
3. Bommer C, Sagalova V, Heesemann E, et al. Global economic burden of diabetes in adults: projections from 2015 to 2030. Diabetes Care 2018 May;41(5):963-970. [doi: 10.2337/dc17-1962] [Medline: 29475843]
4. Ogurtsova K, Guariguata L, Barengo NC, et al. IDF diabetes Atlas: global estimates of undiagnosed diabetes in adults for 2021. Diabetes Res Clin Pract 2022 Jan;183:109118. [doi: 10.1016/j.diabres.2021.109118] [Medline: 34883189]
5. Vas A, Devi ES, Vidyasagar S, et al. Effectiveness of self-management programmes in diabetes management: a systematic review. Int J Nurs Pract 2017 Oct;23(5). [doi: 10.1111/ijn.12571] [Medline: 28758701]
6. Chinese Diabetes Society, Dalong Z, Lixin G. Guideline for the prevention and treatment of diabetes mellitus in China (2024 edition) [Article in Chinese]. Chin J Diabetes Mellitus 2025;17(1). [doi: 10.3760/cma.j.cn115791-20241203-00705]
7. Lee JJN, Abdul Aziz A, Chan ST, et al. Effects of mobile health interventions on health-related outcomes in older adults with type 2 diabetes: a systematic review and meta-analysis. J Diabetes 2023 Jan;15(1):47-57. [doi: 10.1111/1753-0407.13346] [Medline: 36649940]
8. Nkhoma DE, Soko CJ, Bowrin P, et al. Digital interventions self-management education for type 1 and 2 diabetes: a systematic review and meta-analysis. Comput Methods Programs Biomed 2021 Oct;210:106370. [doi: 10.1016/j.cmpb.2021.106370] [Medline: 34492544]
9. Wu Y, Min H, Li M, et al. Effect of artificial intelligence-based health education accurately linking system (AI-HEALS) for type 2 diabetes self-management: protocol for a mixed-methods study. BMC Public Health 2023 Jul 11;23(1):1325. [doi: 10.1186/s12889-023-16066-z] [Medline: 37434126]
10. Zare S, Ostovarfar J, Kaveh MH, Vali M. Effectiveness of theory-based diabetes self-care training interventions; a systematic review. Diabetes Metab Syndr 2020;14(4):423-433. [doi: 10.1016/j.dsx.2020.04.008] [Medline: 32361532]

XSL•FO
RenderX

11. Sharma M, Awan A, Kapukotuwa S. Mini review: possible role of the multi-theory model of health behavior change in designing substance use prevention and treatment interventions. Front Public Health 2024 Mar;12:1298614. [doi: 10.3389/fpubh.2024.1298614] [Medline: 38496384]
12. Kapukotuwa S, Nerida T, Batra K, Sharma M. Utilization of the multi-theory model (MTM) of health behavior change to explain health behaviors: a systematic review. Health Promot Perspect 2024;14(2):121-135. [doi: 10.34172/hpp.42887] [Medline: 39291044]
13. Kumar V, Sabbarwal B, Jaggi A, Taneja P. Effectiveness of tobacco cessation counselling and behavioural changes using multi theory model (MTM): a follow-up study. Indian J Dent Res 2021;32(1):56-60. [doi: 10.4103/ijdr.IJDR_904_19] [Medline: 34269238]
14. Williams JL, Sharma M, Mendy VL, Leggett S, Akil L, Perkins S. Using multi theory model (MTM) of health behavior change to explain intention for initiation and sustenance of the consumption of fruits and vegetables among African American men from barbershops in Mississippi. Health Promot Perspect 2020 Jul 12;10(3):200-206. [doi: 10.34172/hpp.2020.33] [Medline: 32802756]
15. Hayes T, Sharma M, Shahbazi M, Sung JH, Bennett R, Reese-Smith J. The evaluation of a fourth-generation multi-theory model (MTM) based intervention to initiate and sustain physical activity. Health Promot Perspect 2019 Jan;9(1):13-23. [doi: 10.15171/hpp.2019.02] [Medline: 30788263]
16. Yoshany N, Morowatisharifabad MA, Bahri N, Jambarsang S, Sharma M, Hanna F. Effect of the fourth generation multi-theory model intervention on the quality of life in Iranian postmenopausal women: a randomized controlled trial. Post Reprod Health 2021 Dec;27(4):189-197. [doi: 10.1177/20533691211042164] [Medline: 34792396]
17. Sun X, Jiang Y, Wang J, et al. Effects of a mobile health intervention based on a multitheoretical model of health behavior change on anxiety and depression, fear of cancer progression, and quality of life in patients with differentiated thyroid cancer: a randomized controlled trial. BMC Med 2024 Oct 15;22(1):466. [doi: 10.1186/s12916-024-03652-0] [Medline: 39407174]
18. ElSayed NA, McCoy RG, Aleppo G. 5. Facilitating positive health behaviors and well-being to improve health outcomes: standards of care in diabetes—2025. Diabetes Care 2025 Jan 1;48(Supplement_1):S86-S127. [doi: 10.2337/dc25-S005]
19. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy 2008 Apr;13(2):92-98. [doi: 10.1258/jhsrp.2007.007074] [Medline: 18416914]
20. Wang FB, Gao M, Chen XY, et al. Study on the correlation between family support and dietary behavior in patients with type 2 diabetes mellitus in community. Chin J Health Educ 2020;36(4):300-304 [FREE Full text]
21. Toobert DJ, Hampson SE, Glasgow RE. The summary of diabetes self-care activities measure: results from 7 studies and a revised scale. Diabetes Care 2000 Jul;23(7):943-950. [doi: 10.2337/diacare.23.7.943] [Medline: 10895844]
22. Ding L, Velicer WF, Harlow LL. Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. Struct Equ Modeling 1995 Jan;2(2):119-143. [doi: 10.1080/10705519509540000]
23. Mitchell M. Situational interest: its multifaceted structure in the secondary school mathematics classroom. J Educ Psychol 1993 Sep;85(3):424-436. [doi: 10.1037/0022-0663.85.3.424]
24. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
25. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. Health Serv Res 2013 Dec;48(6 Pt 2):2134-2156. [doi: 10.1111/1475-6773.12117] [Medline: 24279835]
26. Guetterman TC, Fetters MD, Creswell JW. Integrating quantitative and qualitative results in health science mixed methods research through joint displays. Ann Fam Med 2015 Nov;13(6):554-561. [doi: 10.1370/afm.1865] [Medline: 26553895]
27. Smalls BL, Gregory CM, Zoller JS, Egede LE. Effect of neighborhood factors on diabetes self-care behaviors in adults with type 2 diabetes. Diabetes Res Clin Pract 2014 Dec;106(3):435-442. [doi: 10.1016/j.diabres.2014.09.029] [Medline: 25451904]
28. Mosley-Johnson E, Walker R, Hawks L, et al. Pathways between neighbourhood factors, stress and glycaemic control in individuals with type 2 diabetes in Southeastern United States: a cross-sectional pathway analysis. BMJ Open 2022 Oct 25;12(10):e060263. [doi: 10.1136/bmjopen-2021-060263] [Medline: 36283754]
29. Hill JO, Galloway JM, Goley A, et al. Scientific statement: socioecological determinants of prediabetes and type 2 diabetes. Diabetes Care 2013 Aug;36(8):2430-2439. [doi: 10.2337/dc13-1161] [Medline: 23788649]
30. Thomas Craig KJ, Morgan LC, Chen CH, et al. Systematic review of context-aware digital behavior change interventions to improve health. Transl Behav Med 2021 May 25;11(5):1037-1048. [doi: 10.1093/tbm/ibaa099] [Medline: 33085767]
31. Cradock KA, ÓLaighin G, Finucane FM, et al. Diet behavior change techniques in type 2 diabetes: a systematic review and meta-analysis. Diabetes Care 2017 Dec;40(12):1800-1810. [doi: 10.2337/dc17-0462] [Medline: 29162585]
32. Pai LW, Hung CT, Chen LL, Lin RL, Lockwood C. Efficacy of a health education technology program in improving adherence to self-management behaviors and quality of life among adults with type 2 diabetes: a randomized controlled trial. Prim Care Diabetes 2024 Oct;18(5):479-485. [doi: 10.1016/j.pcd.2024.08.003] [Medline: 39223060]
33. DPaCCoCPM A, Diabetes Prevention and Control Committee of Chinese Preventive Medicine Association. Chinese diabetes behaviour and lifestyle intervention guidelines (2024)[Article in Chinese]. Chin Gen Pract 2025;28(7):777-796. [doi: 10.12114/j.issn.1007-9572.2024.0548]

XSL•FO
RenderX

34. Yangshuang Z, Wen L, Yangdi L, Xujuan L, Wei W. Lasting effects of diabetes conversation mapping on health beliefs and health behaviors and psychosocial interventions [Article in Chinese]. J Int Psychiatry 2025;52(1):268-272. [doi: 10.13479/j.cnki.jip.2025.01.034]
35. Captieux M, Pearce G, Parke HL, et al. Supported self-management for people with type 2 diabetes: a meta-review of quantitative systematic reviews. BMJ Open 2018 Dec 14;8(12):e024262. [doi: 10.1136/bmjopen-2018-024262] [Medline: 30552277]
36. Lacourse MG, Orr ELR, Cramer SC, Cohen MJ. Brain activation during execution and motor imagery of novel and skilled sequential hand movements. Neuroimage 2005 Sep;27(3):505-519. [doi: 10.1016/j.neuroimage.2005.04.025] [Medline: 16046149]
37. Luu P, Tucker DM, Stripling R. Neural mechanisms for learning actions in context. Brain Res 2007 Nov 7;1179:89-105. [doi: 10.1016/j.brainres.2007.03.092] [Medline: 17936726]
38. Motevalli S, Salahshour HM, Bailey RP. The mediating role of cognitive flexibility in the relationship between cognitive emotion regulation strategies and mindfulness in patients with type 2 diabetes. J Affect Disord 2023 Oct 15;339:676-682. [doi: 10.1016/j.jad.2023.07.043] [Medline: 37479040]
39. Ghoreishi MS, Vahedian-Shahroodi M, Jafari A, Tehranid H. Self-care behaviors in patients with type 2 diabetes: education intervention base on social cognitive theory. Diabetes Metab Syndr 2019;13(3):2049-2056. [doi: 10.1016/j.dsx.2019.04.045] [Medline: 31235135]
40. Zambrano J, Celano CM, Chung WJ, et al. Exploring the feasibility and impact of positive psychology-motivational interviewing interventions to promote positive affect and physical activity in type 2 diabetes: design and methods from the BEHOLD-8 and BEHOLD-16 clinical trials. Health Psychol Behav Med 2020;8(1):398-422. [doi: 10.1080/21642850.2020.1815538] [Medline: 33763296]
41. Bujanda-Sainz de Murieta A, Soto-Ruiz N, García-Vivar C, San Martín-Rodríguez L, Escalada-Hernández P. Use of online communities among people with type 2 diabetes: a scoping review. Curr Diab Rep 2024 May;24(5):96-107. [doi: 10.1007/s11892-024-01538-2] [Medline: 38457015]
42. Haijun J, Bingmei G, Congcong L, Yun Z. Effects of different forms of self-management education on glycosylated hemoglobin in diabetic patients: a meta-analysis [Article in Chinese]. Chin J Mod Nurs 2019;25(9):1139-1144. [doi: 10.3760/cma.j.issn.1674-2907.2019.09.018]
43. Li an DE. Personality and adaptive coping in the digital age: psychological pathways to emerging technology adoption, with a focus on artificial intelligence. Health Psychol Res 2026;14(1):e81240057. [doi: 10.14440/hpr.0430]
44. Clay TJ, Da Custodia Steel ZJ, Jacobs C. Human-computer interaction: a literature review of artificial intelligence and communication in healthcare. Cureus 2024 Nov;16(11):e73763. [doi: 10.7759/cureus.73763] [Medline: 39677224]
45. Will ChatGPT transform healthcare? Nat Med 2023 Mar;29(3):505-506. [doi: 10.1038/s41591-023-02289-5] [Medline: 36918736]
46. Deo PS, Barber TM, Gotts C, et al. A pilot study to assess the feasibility and uptake of MindDEC, a mindfulness-based self-management course for type 2 diabetes. Complement Ther Med 2024 Aug;83:103059. [doi: 10.1016/j.ctim.2024.103059] [Medline: 38821377]
47. Fall E, Roche B, Izaute M, Batisse M, Tauveron I, Chakroun N. A brief psychological intervention to improve adherence in type 2 diabetes. Diabetes Metab 2013 Oct;39(5):432-438. [doi: 10.1016/j.diabet.2013.05.003] [Medline: 24094567]
48. Benton JS, Cotterill S, Hawkes RE, Miles LM, French DP. Changes in a digital type 2 diabetes self-management intervention during national rollout: mixed methods study of fidelity. J Med Internet Res 2022 Dec 7;24(12):e39483. [doi: 10.2196/39483] [Medline: 36476723]
49. Veluvali A, Dehghani Zahedani A, Hosseinian A, et al. Impact of digital health interventions on glycemic control and weight management. NPJ Digit Med 2025 Jan 9;8(1):20. [doi: 10.1038/s41746-025-01430-7] [Medline: 39789102]
50. Kerr D, Ahn D, Waki K, Wang J, Breznen B, Klonoff DC. Digital interventions for self-management of type 2 diabetes mellitus: systematic literature review and meta-analysis. J Med Internet Res 2024 Jul 22;26:e55757. [doi: 10.2196/55757] [Medline: 39037772]
51. Ruissen MM, Torres-Peña JD, Uitbeijerse BS, et al. Clinical impact of an integrated e-health system for diabetes self-management support and shared decision making (POWER2DM): a randomised controlled trial. Diabetologia 2023 Dec;66(12):2213-2225. [doi: 10.1007/s00125-023-06006-2] [Medline: 37775611]
52. Lee YB, Kim G, Jun JE, et al. An integrated digital health care platform for diabetes management with AI-based dietary management: 48-week results from a randomized controlled trial. Diabetes Care 2023 May 1;46(5):959-966. [doi: 10.2337/dc22-1929] [Medline: 36821833]
53. Patel MR, Zhang G, Heisler M, et al. A randomized controlled trial to improve unmet social needs and clinical outcomes among adults with diabetes. J Gen Intern Med 2024 Oct;39(13):2415-2424. [doi: 10.1007/s11606-024-08708-8] [Medline: 38467918]

Abbreviations

AI-HEALS: Artificial Intelligence-based Health Education Accurately Linking System

XSL•FO
RenderX

CFA: confirmatory factor analysis
CNY: Chinese Yuan
DHIs: digital health interventions
EFA: exploratory factor analysis
HbA_{1c}: glycated hemoglobin
MCAR: Missing completely at random
MTM: Multi-Theory Model
RMSEA: root mean square error of approximation
SEM: structural equation modeling
T2DM: type 2 diabetes mellitus
TLI: Tucker–Lewis index

Edited by S Brini; submitted 15.Jun.2025; peer-reviewed by JLC Lee, V Raj; revised version received 08.Jul.2026; accepted 09.Jul.2026; published 10.Aug.2026.

Please cite as:
Wu Y, Ni Y, Jiang Y, Xu Z, Min H, Chen P, Gu X, Kong B, Gan Y, Li P, Li M, Guo X, Zhang X, Ma A, Sun X
Behavioral Mechanisms of a Digital Health Intervention for Self-Management in Type 2 Diabetes Mellitus: Prospective Longitudinal Cohort Study
JMIR Mhealth Uhealth 2026;14:e79081
URL: https://mhealth.jmir.org/2026/1/e79081
doi:10.2196/79081

© Yibo Wu, Yang Ni, Yang Jiang, Zijie Xu, Hewei Min, Ping Chen, Xinbao Gu, Bingyang Kong, Yadi Gan, Pei Li, Mingzi Li, Xiaohui Guo, Xuxi Zhang, Aijuan Ma, Xinying Sun. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Feasibility, Usability, and Preliminary Effectiveness of an mHealth App to Promote Screening Behaviors Among High-Risk Populations for Breast Cancer: Randomized Controlled Pilot Study

Linping Zhang[1], MSN, RN; Mengjiao Xu[1], MSN, RN; Xian Zhang[2], BN, RN; Xiaoxu Li[2], BN, RN; Shuo Wang[1], RN, PhD; Fang Zhou[1], RN, PhD; Jing Han[1], RN, PhD

[1]School of Nursing, Xuzhou Medical University, No 209, Tongshan Road, Huangshan Street, Yunlong District, Xuzhou, China

[2]Department of Nursing, Affiliated Hospital of Xuzhou Medical University, Xuzhou, China

Corresponding Author:
Jing Han, RN, PhD
School of Nursing, Xuzhou Medical University, No 209, Tongshan Road, Huangshan Street, Yunlong District, Xuzhou, China

Abstract

Background: Breast cancer is a major public health challenge worldwide. Women at high risk for breast cancer are more likely to develop the disease; yet, screening participation remains. Mobile health interventions may improve breast health awareness and screening behaviors, but evidence in high-risk populations for breast cancer remains limited.

Objective: This study aimed to evaluate the feasibility, usability, and preliminary signals of potential effectiveness of the "Ruaikang" mobile health app, designed to support breast health awareness and engagement with early screening among high-risk populations for breast cancer.

Methods: In this pilot randomized controlled trial, eligible high-risk populations for breast cancer were randomly assigned (1:1) to a 4-week intervention using the "Ruaikang" app or a control group receiving standard breast cancer screening guidelines. Feasibility was assessed via recruitment and completion rates. Usability was measured using the System Usability Scale. Use data were collected through the app's back-end system. Preliminary effectiveness was assessed through changes in willingness to participate in breast cancer screening, screening knowledge, health beliefs, and self-efficacy at both the fourth (T1) and eighth (T2) weeks. Semistructured interviews were conducted with a purposive subsample of intervention participants to explore user experience and suggestions for improvement.

Results: Of 62 eligible women approached, 50 enrolled (recruitment rate 80.6%) and were randomized to intervention (n=25) and control (n=25) groups. In total, 19 participants in the intervention group completed the study (completion rate 76%), reflecting good feasibility. The mean usability score was 88.82 (SD 10.29), indicating high usability. Back-end use data showed that participants actively engaged during the intervention period. At T1, the intervention group showed preliminary signals of improvement in willingness to participate in breast cancer screening, screening knowledge, health beliefs, and self-efficacy, although these changes were less pronounced at T2. Qualitative interviews with 8 participants identified 2 core themes: "perceived value and acceptability of the app" and "system limitations and user-centered improvement needs." Participants highlighted that the app was easy to use and content-rich and provided actionable feedback for optimizing interactive features, navigation, and content depth.

Conclusions: The "Ruaikang" mobile health app demonstrated good feasibility and usability and provided preliminary signals of potential effectiveness in improving willingness to participate in breast cancer screening, screening knowledge, health beliefs, and self-efficacy among high-risk populations for breast cancer. Participants' feedback indicates that the app aligns with users' core needs and supports engagement with breast health information. Future work should refine the app based on user feedback and conduct larger-scale, longer-term randomized controlled trials to confirm and further validate the preliminary effectiveness observed.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2500108593; https://tinyurl.com/f99mfjwp

(JMIR Mhealth Uhealth 2026;14:e86429) doi:10.2196/86429

KEYWORDS

breast cancer; cancer screening; high-risk populations; mobile health app; feasibility; usability

XSL•FO
RenderX

Introduction

Breast cancer presents a significant public health challenge for women globally. According to the International Agency for Research on Cancer [1], there were approximately 2.3 million new breast cancer cases and 670,000 deaths worldwide in 2022, making it the leading cause of cancer death among women after lung cancer. In China, approximately 350,000 new cases and 75,000 deaths were estimated in 2022, ranking it as the second leading cause of cancer-related mortality among Chinese women [2]. Cancer screening is an effective strategy for the early detection of breast cancer [3]. It involves the application of effective, simple, and economical breast examination to identify asymptomatic individuals at an early stage. As the most effective approach to enhancing early diagnosis rates, survival, and quality of life for patients, breast cancer screening is particularly vital for high-risk populations for breast cancer [4]. High-risk populations for breast cancer refer to women who carry established risk factors, such as a family history of breast cancer, relevant gene mutations, a history of benign breast disease, unhealthy lifestyle habits, smoking, and alcohol abuse [5]. Studies show that such individuals face a significantly elevated risk of developing breast cancer, which can be up to 10 times that of the general population [6].

Since the beginning of the 21st century, European Union countries such as the United Kingdom, Ireland, and Italy have actively promoted breast cancer screening programs, recommending that eligible women be screened every 2 to 3 years [7]. In 2009, the Chinese government launched a free "two cancers" (cervical cancer and breast cancer) screening program for eligible women and incorporated breast cancer screening into the national public health service in 2019 to encourage broader participation [8]. Despite these clear policy measures, participation in breast cancer screening remains low, particularly among high-risk populations for breast cancer. A study by Kırca et al [9] among first-degree relatives of patients with breast cancer in the largest hospital in southwestern Turkey revealed that fewer than 20% of participants had undergone screening. Similarly, a survey by Zhang et al [10] of 15,550 high-risk populations for breast cancer in Beijing showed that less than 50% were willing to participate in screening. Further findings from Zhao et al [11] highlighted low screening participation among first-degree relatives of patients with breast cancer, with only 6.7% of respondents performing breast self-examination once a month and 22.8% undergoing breast ultrasound or mammography once a year. According to the National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines [12], individuals identified as being at high risk for breast cancer include those with a lifetime risk of 20% or greater based on validated risk assessment models, a history of chest radiation therapy or breast cancer, the presence of high-risk breast lesions, or a family history suggestive of hereditary breast cancer. For this population, NCCN recommends earlier and more intensive screening, including annual mammography and breast magnetic resonance imaging, as well as clinical breast examinations every 6 to 12 months. The Chinese breast cancer screening guidelines [3] provide definitions and recommendations for high-risk populations for breast cancer that align closely with those of the NCCN. Therefore, it is imperative to develop and implement targeted interventions for high-risk populations for breast cancer to improve breast cancer screening participation, thereby alleviating the public health burden.

Previous studies suggested that factors such as breast cancer awareness, health beliefs, social support, and self-efficacy play significant roles in influencing women's screening behaviors [9]. Researchers have developed various interventions, such as health education, peer support, patient navigation services, and decision assistance tools, to improve participation in breast cancer screening, and these studies have shown a positive promotion effect on screening participation [13-16]. Despite the demonstrated benefits of existing interventions, their implementation often depends on face-to-face delivery or printed materials, which require substantial personnel, time, and financial resources, thereby limiting scalability, flexibility, and long-term sustainability [17,18]. These challenges are particularly pronounced for women with demanding schedules, limited mobility, or those living in geographically dispersed areas, who may face difficulties in accessing such support [19]. In addition, most interventions focus on a single behavioral determinant, whereas women's breast cancer screening behaviors are influenced by multiple interacting psychosocial factors [20]. Consequently, there is a need to develop more comprehensive, multicomponent intervention strategies that leverage digital platforms to improve accessibility, deliver timely support, and promote sustained participation in breast cancer screening.

With the global popularization of the internet, mobile health has become an integral component of contemporary health care [21]. As a key medium of mobile health, mobile health apps have been widely applied in multiple fields, including health monitoring, disease prevention, and chronic disease management [22-24]. To date, a considerable number of studies have been devoted to developing mobile health apps to promote breast cancer screening. For instance, Yusuf et al [25] developed a health education app tailored for women in northeastern Malaysia. By integrating content on breast anatomy, breast cancer knowledge, support group information, screening reminders, and culturally adapted educational materials, the app successfully enhanced screening confidence and willingness among the target population. Similarly, Roh et al [26] developed an app to encourage breast cancer screening among Native American people. This app offers breast cancer education, information on screening clinics, and regular text-message reminders and has shown promising results in real-world applications. However, existing apps primarily target the general population of healthy women, with limited attention to those high-risk populations for breast cancer. Compared to the general population, high-risk populations for breast cancer require more comprehensive breast health knowledge and more proactive health behaviors.

The research team developed a mobile health app named "Ruaikang" to provide breast cancer screening support resources for high-risk populations for breast cancer, enhance their breast health awareness, and promote participation in early screening. This pilot randomized controlled study aimed to evaluate the

XSL•FO
RenderX

feasibility and usability of the "Ruaikang" mobile health app among high-risk populations for breast cancer and to explore its potential impact on promoting breast cancer screening behaviors.

Methods

Study Design

This pilot randomized controlled study was conducted from January to October 2025. The trial was registered in the Chinese Clinical Trial Registry (registration: ChiCTR2500108593) and reported in accordance with the CONSORT (Consolidated Standards of Reporting Trials) extension for pilot and feasibility trials (Checklist 1) [27].

Participants and Setting

Participants were from a grade A tertiary hospital in eastern China. Convenience sampling was adopted to recruit the high-risk populations for breast cancer who visited the hospital during the study period. The inclusion criteria for the participants were as follows: (1) meeting the definition of high-risk populations for breast cancer per Chinese women's breast cancer screening guidelines [3], including those with breast cancer susceptibility gene, family history of breast cancer (parents, children, and siblings), personal history of breast cancer, history of chest radiotherapy, and diagnosed with lobular carcinoma in situ, atypical ductal hyperplasia, or atypical lobular hyperplasia before 40 years; (2) women; (3) age ≥18 years; (4) possession of an Android smartphone and proficiency in its operation; and (5) those who voluntarily participated in this study and signed an informed consent form. The exclusion criteria were as follows: (1) physician-diagnosed confusion or cognitive impairment and (2) previous history of mental illness.

Sample Size

In pilot studies, a minimum of 12 participants per group has been recommended to provide reasonable estimates of key parameters, such as the mean and variance, for planning future definitive trials [28]. In accordance with this guideline, the quantitative component of this study established a minimum sample size of 12 participants in each group. To account for a potential dropout rate of 20%, the final sample size was adjusted to at least 15 participants per group. For the qualitative component, participants were purposively sampled from the intervention group based on age, educational level, and app use patterns to ensure diverse and information-rich perspectives. The sample size was guided by the concept of information power [29], and data collection continued until sufficient information power was achieved to adequately address the study objectives.

Procedures

The research team comprised 3 research assistants, 1 breast cancer care specialist, and 1 software engineer. The research assistants were primarily responsible for participant recruitment, intervention implementation, and data collection. The breast cancer care specialist provided clinical expertise and addressed participants' medical questions. The software engineer was responsible for app maintenance and ensuring stable technical operation.

Research assistants screened potential participants in hospital wards and outpatient departments against the inclusion criteria. Eligible individuals received an explanation of the study's background, purpose, and procedures. After providing informed consent, participants completed the baseline (T0) questionnaire and were randomly assigned to either the intervention group or the control group. The intervention group received a 4-week intervention via a mobile health app named "Ruaikang," while the control group received the printed "Breast Cancer Screening Guidelines" manual [3]. Participants in the intervention group installed and registered the "Ruaikang" with the research assistants' guidance. The assistants helped them log in, complete their personal profiles, and demonstrate the functionality of each module. Follow-up assessments were conducted with all participants at the fourth (T1) and eighth (T2) weeks. At T1, participants in the intervention group were also invited to participate in a semistructured interview about user experience and improvement suggestions for the app.

Randomization and Blinding

A computer-generated random sequence was created by an independent statistician, who was not involved in participant recruitment or intervention delivery. Eligible participants were allocated in a 1:1 ratio to the intervention or control group. To ensure allocation concealment, the statistician prepared sequentially numbered, opaque, sealed envelopes, which were stored and managed by a research coordinator independent of participant enrollment and outcome assessment. Envelopes were opened sequentially only after participant enrollment and baseline data collection. Participants could not be blinded due to the nature of the intervention. Outcome assessors and research staff involved in follow-up and data collection remained blinded, and group assignment was not disclosed to assessors during follow-up to minimize detection bias.

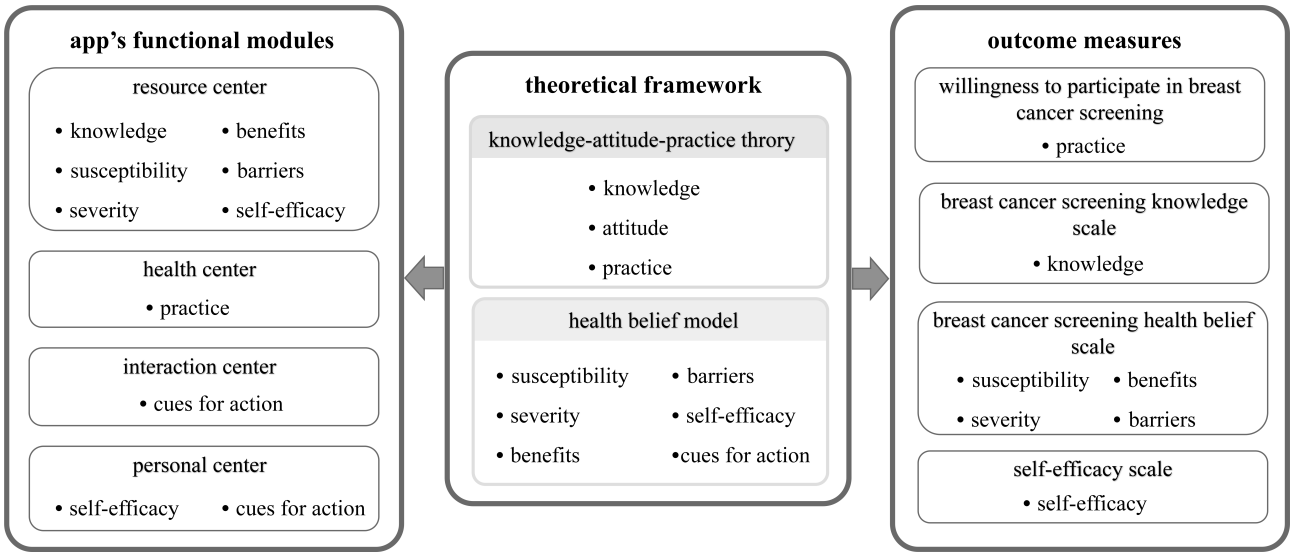
Intervention

The "Ruaikang" mobile health app was developed based on the knowledge-attitude-practice theory [30] and the health belief model (HBM) [31]. The knowledge-attitude-practice model outlines 3 stages for behavioral change: knowledge acquisition, attitude formation, and practice adoption. Knowledge enables attitude shifts, which drive behavioral change for health goals. HBM explains health behaviors through 6 belief dimensions: perceived susceptibility, perceived severity, perceived benefits, perceived barriers, self-efficacy, and cues for action. In this study, "knowledge" refers to understanding of breast cancer pathology and early screening protocols. "Attitude" refers to the cultivation of positive beliefs and attitudes toward early screening participation among high-risk populations for breast cancer, with conceptual alignment to core HBM dimensions. "Practice" refers to adoption of regular breast cancer screening behaviors by high-risk populations for breast cancer, informed by acquired knowledge and reinforced attitudes. This theoretical framework informed both the design of the app's functional modules and the selection of outcome measures, as illustrated in Figure 1. The app comprises 2 core systems: the user side and the back-end management system. The user side is designed to deliver breast health education, early screening guidance, health behavior promotion, and psychosocial support for

XSL•FO
RenderX

high-risk populations for breast cancer, including 4 core functional modules: the resource center, health center, interaction center, and personal center. The back-end management system is designed to oversee and control the overall operation of the app. The detailed information on the app's development process, user side features, and back-end management functions is shown in Multimedia Appendix 1.

Figure 1. Conceptual framework.



Outcome Measures

The Feasibility and Usability of the App

The feasibility and usability of the app are as follows:

- Feasibility: The feasibility of the app was assessed using 2 key indicators: recruitment rate and completion rate. The recruitment rate was defined as the percentage of volunteers who participated in the study among all individuals who met the eligibility criteria. The completion rate referred to the percentage of participants who finished the entire intervention out of all those who agreed to participate. Based on previous studies [32], the feasibility threshold was set as follows: recruitment rate ≥80% and completion rate ≥75%.
- Usability: The usability of the app was assessed using the System Usability Scale [33]. The scale comprises 10 items; responses were scored on a 5-point Likert scale, and each item yielded a value between 0 and 4. The sum of all item scores is multiplied by 2.5 to produce a total score ranging from 0 to 100. If the total score is ≥70 [34], the usability of the app is considered to meet the standard. The Cronbach α value of the scale was 0.84, and in this study, it was 0.87.
- Use: Researchers collected users' use duration, the number of resource page visits, comments posted, and health consultations through the app's back-end management system.

The Preliminary Effectiveness of the App

Primary Outcome

Willingness to participate in breast cancer screening was assessed using items embedded within the general information questionnaire.

Secondary Outcomes

Breast Cancer Screening Knowledge

The Breast Cancer Screening Knowledge Scale [35] was used to assess participants' knowledge of breast cancer and screening. This instrument was developed based on the Chinese breast cancer screening guidelines and authoritative educational materials and has been publicly used in China [36,37]. This questionnaire comprises 17 items assessing knowledge across 3 domains: high-risk factors for breast cancer, clinical manifestations, and early screening methods. All items are presented as true or false statements. Each correct response earns 1 point, with no points awarded for incorrect answers. The total score ranges from 0 to 17 points, with higher scores indicating a higher level of knowledge about breast cancer and screening. The Cronbach α value of the scale was 0.72 [35], and the Cronbach α in this study was 0.82.

Breast Cancer Screening Health Belief

The Breast Cancer Screening Health Belief Scale [38], which was developed based on the HBM framework, was used to assess women's beliefs and attitudes toward breast cancer screening. The scale consists of 12 items, each scored on a 5-point Likert scale (ranging from 1 to 5). The scale comprised 4 dimensions: perceived susceptibility, perceived severity, perceived benefits, and perceived barriers. Total scores for each dimension ranged from 2 to 10 for perceived susceptibility, 3 to 15 for both perceived severity and perceived benefits, and 4 to 20 for perceived barriers. Higher scores indicated stronger health beliefs regarding breast cancer screening. The scale has been publicly used [39]. The Cronbach α value of the scale was 0.89 [38], and in this study, it was 0.70.

Self-Efficacy

The General Self-Efficacy Scale [40] was used to assess women's confidence in their ability to complete breast cancer

XSL•FO
RenderX

screening when faced with various challenges. The General Self-Efficacy Scale has been applied in patients with cancer and has demonstrated good reliability and validity [41]. The scale was translated and culturally adapted into Chinese by Wang [42]. It consists of 10 items, and responses to the questions were scored on a 4-point Likert scale. The total score ranges from 10 to 40 points, with higher scores indicating higher levels of self-efficacy. The Cronbach α value of the scale was 0.87, and in this study, it was 0.95.

User Experience and Improvement Suggestions for the App

The interview guide was developed based on the unified theory of acceptance and use of technology framework [43], prior usability and user experience studies of mobile health apps [44], and the specific objectives of this study. It was further refined through discussion among the research team and consultation with oncology and mobile health experts to ensure that the questions were clear, relevant, and aligned with the study goals. The interview guide included the following questions: What is your overall experience with the app? What benefits did you gain from using the app? Which aspects of the app do you believe need improvement? Would you recommend the app to others? Why?

Data Collection

After receiving standardized training, the research team members conducted data collection following unified instructions. Participants in both groups completed questionnaires at T0, T1, and T2, with each questionnaire taking approximately 10 - 15 minutes to complete. Data at T0 were collected through face-to-face paper questionnaires, while data at T1 and T2 were collected via telephone, electronic questionnaires, or face-to-face administration. After retrieving the questionnaires, researchers promptly checked the completeness and validity of the responses and conducted supplementary investigations when necessary. All collected data were anonymized, securely stored, and used solely for the purposes of this study. The research process strictly adhered to ethical principles, ensuring participants' voluntariness, confidentiality, and data security, while minimizing potential risks.

Feedback regarding the experience and improvement suggestions for the app were collected from participants in the intervention group through face-to-face or telephone interviews. The interviews were conducted by 2 researchers (LZ and MX) with 3 years of clinical experience in oncology and training in qualitative research. The data collection process was as follows: (1) the purpose and significance of the interview were explained to participants who had completed the intervention, written informed consent was obtained, and permission to record the interview was granted; and (2) the researchers conducted one-to-one, semistructured interviews while audio-recording each session. Each interview lasted approximately 20 - 30 minutes.

Data Analysis

Quantitative data were analyzed using SPSS software (version 27; IBM Corp). Descriptive statistics were used to summarize participants' sociodemographic characteristics, as well as the feasibility, usability, and use of the app, and the outcome variables of effectiveness at different time points. Continuous variables were presented as mean and SD, while categorical variables were presented as frequency and percentage. Differences in baseline characteristics between groups were examined using 2-tailed independent-samples *t* tests and chi-square statistics. In accordance with the intention-to-treat principle, generalized estimating equations (GEE) were used to compare changes in outcome variables between the control and intervention groups across time points. In the GEE analysis, missing data were handled using multiple imputation, and a sensitivity analysis was conducted to assess the robustness of results by comparing outcomes before and after handling missing data. The level of statistical significance was set at $P<.05$.

Qualitative data analysis was initiated following the first interview. Upon completion of each session, the interviewing researchers reviewed the audio recordings, performed verbatim transcription, and imported the transcripts into NVivo (version 12.0; QSR International) within 24 hours. Analysis was conducted following a coding-reliability thematic analysis approach to ensure transparency, objectivity, and reproducibility of the findings [45]. The research team repeatedly read all transcripts to become thoroughly familiar with the data and held meetings to discuss preliminary impressions, establishing a shared understanding to guide systematic coding. Based on the research questions, a structured initial codebook was developed, with each code assigned a clear name, operational definition, and inclusion or exclusion criteria. Two researchers (LZ and MX) independently applied this codebook to a randomly selected subset of transcripts for pilot coding. Intercoder reliability was assessed using both percentage agreement and Cohen κ (initial agreement=85%; κ=0.78). Discrepancies were discussed and resolved by referring to the original transcripts, and the codebook was revised and finalized for use in coding the full dataset. The finalized codebook was then applied independently to the remaining transcripts. Any novel or marginal data encountered during coding were documented for periodic review but did not modify the established coding framework. Once coding was complete, all coded data were exported, and the researchers examined the codes in relation to their frequency, patterns of co-occurring codes, and coverage within the dataset to identify potential themes. Each theme was then reviewed against the original transcripts to ensure that it accurately represented the data and was supported by sufficient evidence. Final themes were refined and clearly defined, with descriptive labels that accurately reflected the content of the data while avoiding overinterpretation. These descriptive themes directly addressed the research questions and provided reliable evidence. Data collection was concluded once the researchers determined that sufficient information power had been achieved. Additionally, both researchers selected representative quotations from the dataset and translated them into English. In cases of ambiguity during translation, the team consulted professional translation experts to ensure semantic accuracy and preserve the original intent of the participants' statements.

XSL•FO
RenderX

Ethical Considerations

This study received approval from the ethics committee of Xuzhou Mining Group General Hospital (approval 2025082501). All procedures performed were in accordance with the ethical principles of the Declaration of Helsinki. Written informed consent was obtained from all participants before randomization. Prior to the study, the research objectives and procedures were thoroughly explained to participants, and all inquiries were addressed. Participants were assured of the strict confidentiality of their personal information and were informed of their right to withdraw unconditionally at any time. As compensation for their time and contribution, each participant received a gift worth US $7.35 upon completion of data collection.

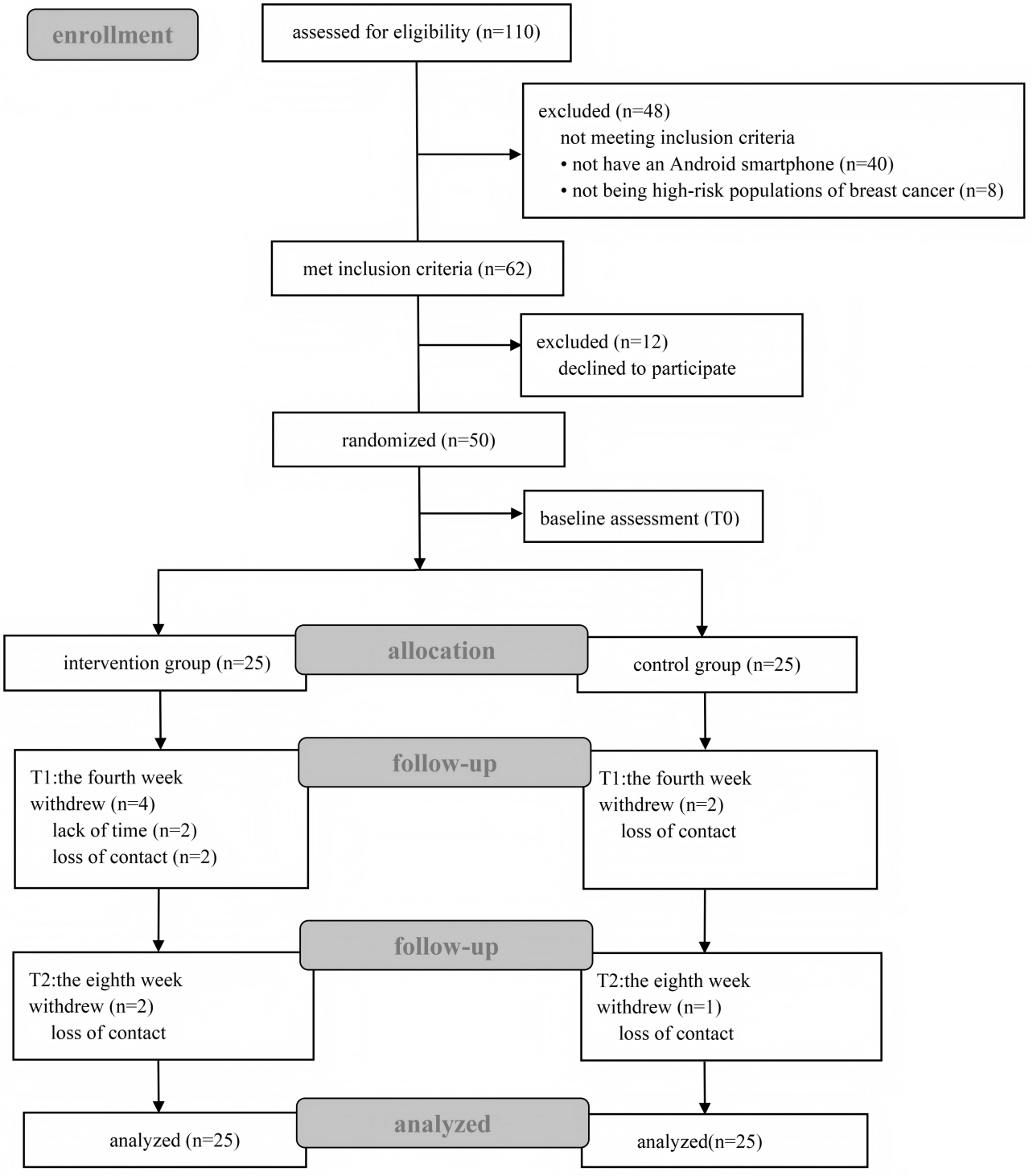
Results

Recruitment and Characteristics of the Participants

A total of 110 women were assessed for eligibility, with 62 meeting the inclusion criteria and 48 excluded due to not having an Android smartphone (n=40) or not being high-risk populations for breast cancer (n=8). Among the eligible participants, 12 declined to participate, resulting in 50 participants who were successfully enrolled and randomly assigned to either the intervention or control group. During follow-up, 6 participants in the intervention group withdrew due to lack of time or loss of contact, while 3 participants in the control group withdrew due to loss of contact. The recruitment flow and reasons for dropout are detailed in Figure 2. Baseline characteristics of participants in both groups are shown in Table 1, with no significant differences across measures, indicating comparability between groups.

XSL•FO
RenderX

Figure 2. CONSORT flow diagram. CONSORT: Consolidated Standards of Reporting Trials.



XSL•FO
RenderX

Table . Baseline characteristics of the participants (N=50).

Variables	Intervention group (n=25)	Control group (n=25)	Statistics	*P* value
Age (years), mean (SD)	43.48 (8.90)	41.20 (8.97)	t_{48}=−0.90	.37
Ethnicity, n (%)			χ^2_1=0.2	>.99
Han	23 (92)	22 (88)		
Other	2 (8)	3 (12)		
Educational attainment, n (%)			χ^2_2=1.8	.44
Junior high school and below	6 (24)	10 (40)		
High school or junior college	12 (48)	8 (32)		
Bachelor's degree and above	7 (28)	7 (28)		
Employment status, n (%)			χ^2_2=0.6	.84
Employed	16 (64)	17 (68)		
Freelancer	4 (16)	5 (20)		
Retired	5 (20)	3 (12)		
Marital status, n (%)			χ^2_1=0.4	>.99
Married	24 (96)	23 (92)		
Unmarried	1 (4)	2 (8)		
Medical insurance, n (%)			χ^2_1=0.0	>.99
Yes	24 (96)	24 (96)		
No	1 (4)	1 (4)		
Family history of breast cancer, n (%)			χ^2_1=0.4	.54
Yes	16 (64)	18 (72)		
No	9 (36)	7 (28)		
Personal history of breast cancer, n (%)			χ^2_1=0.6	.44
Yes	3 (12)	5 (20)		
No	22 (88)	20 (80)		
Diagnosed with LCIS[a], ADH[b], or ALH[c] before 40 years, n (%)			χ^2_1=0.9	.33
Yes	8 (32)	5 (20)		
No	17 (68)	20 (80)		
History of breast cancer screening, n (%)			χ^2_1=0.1	>.99
Yes	7 (28)	6 (24)		
No	18 (72)	19 (76)		
Willingness to participate in breast cancer screening, n (%)			χ^2_1=0.8	>.99
Yes	13 (52)	12 (48)		
No	12 (48)	13 (52)		
Breast cancer screening knowledge, mean (SD)	8.12 (2.80)	8.48 (2.55)	t_{48}=0.48	.64
Perceived susceptibility, mean (SD)	6.20 (0.91)	6.72 (1.06)	t_{48}=1.86	.07
Perceived severity, mean (SD)	7.40 (1.85)	8.12 (1.56)	t_{48}=1.49	.14

XSL•FO
RenderX

Variables	Intervention group (n=25)	Control group (n=25)	Statistics	*P* value
Perceived benefits, mean (SD)	13.04 (1.17)	12.72 (1.79)	t_{48}=−0.75	.46
Perceived barriers, mean (SD)	13.64 (2.43)	12.52 (2.37)	t_{48}=−1.65	.11
Self-efficacy, mean (SD)	22.36 (5.25)	23.24 (5.39)	t_{48}=0.59	.56

[a]LCIS: lobular carcinoma in situ.

[b]ADH: atypical ductal hyperplasia.

[c]ALH: atypical lobular hyperplasia.

Feasibility and Usability of the App

The feasibility of the app was illustrated by the CONSORT flowchart (Figure 2). The recruitment rate was 80.6% (50/62), and the completion rate was 76% (19/25).

The average total score on the System Usability Scale was 88.82 (SD 10.29), and the score of each item is shown in Table 2. Among the odd-numbered items on the scale, the highest score was for the item “I think that I would like to use this system frequently” (mean 3.89, SD 0.32), indicating the strongest agreement among participants with this statement. Among the even-numbered items, the lowest score was for the item “I think that I would need the support of a technical person to be able to use this system” (mean 3.05, SD 1.03), reflecting the highest level of disagreement among participants.

Table . The average score of the System Usability Scale (n=19).

Items	Mean (SD)
I thought that I would like to use this system frequently	3.89 (0.32)
I found the system unnecessarily complex	3.58 (0.61)
I thought the system was easy to use	3.74 (0.45)
I thought that I would need the support of a technical person to be able to use this system	3.05 (1.03)
I found the various functions in this system were well integrated	3.74 (0.45)
I thought there was too much inconsistency in this system	3.42 (0.51)
I would imagine that most people would learn to use this system very quickly	3.68 (0.58)
I found the system very cumbersome to use	3.47 (0.51)
I felt very confident using the system	3.74 (0.45)
I needed to learn a lot of things before I could get going with this system	3.21 (0.86)

During the intervention period, the app use of participants in the intervention group is shown in Table 3. The average app use duration was 2.03(SD 0.77) hours, the average number of resource page visits was 30.42 (SD 10.93), the average number of comments posted was 1.05 (SD 1.27), and the average number of health consultations was 0.89 (SD 0.99).

XSL•FO
RenderX

Table . The use of the app during the intervention period (n=19)[a].

Participants[b]	Use duration (hours)	Resource page visits, n	Comments posted, n	Health consultations, n
S1	3	46	3	2
S2	2	30	2	1
S3	2.5	37	3	2
S4	3.5	48	3	2
S5	3	45	2	1
S6	1	31	0	0
S7	2	29	0	0
S8	2.5	38	1	3
S9	1	16	0	0
S10	1.5	23	0	0
S11	2	16	0	0
S12	2	30	1	1
S13	1.5	23	0	0
S14	1	16	0	0
S15	2.5	37	2	2
S16	2	30	0	1
S17	3	45	3	2
S18	1.5	23	0	0
S19	1	15	0	0

[a]Use duration: mean 2.03, SD 0.77; resource page visits: mean 30.42, SD 10.93; comments posted: mean 1.05, SD 1.27; health consultations: mean 0.89, SD 0.99.

[b]S1 to S19 refer to the participants who completed the study.

Preliminary Signals of Effectiveness

As shown in Table 4, the proportion of participants willing to participate in breast cancer screening increased from T0 to T1 in both groups, with a more pronounced increase in the intervention group. However, this proportion declined in the intervention group from T1 to T2. As shown in Table 5, GEE analysis indicated that at T1, participants in the intervention group tended to report higher willingness compared to those in the control group (odds ratio 4.889, 95% CI 1.343-17.801), while by T2, this difference was less pronounced (odds ratio 2.667, 95% CI 0.717-9.912).

XSL•FO
RenderX

Table . Descriptive statistics for outcome measures at each time point for the intervention and control groups (number of each group=25).

Variables	T0[a]		T1[b]		T2[c]	
	Intervention group	Control group	Intervention group	Control group	Intervention group	Control group
Willingness to participate in breast cancer screening, n (%)						
Yes	13 (52)	12 (48)	22 (88)	15 (60)	20 (80)	15 (60)
No	12 (48)	13 (52)	3 (12)	10 (40)	5 (20)	10 (40)
Breast cancer screening knowledge, mean (SD)	8.12 (2.80)	8.48 (2.55)	16.45 (1.74)	9.25 (2.13)	15.20 (2.24)	9.75 (1.88)
Perceived susceptibility, mean (SD)	6.20 (0.91)	6.72 (1.06)	8.06 (1.09)	6.74 (1.00)	7.88 (1.04)	6.96 (0.84)
Perceived severity, mean (SD)	7.40 (1.85)	8.12 (1.56)	10.43 (1.19)	8.52 (1.20)	10.29 (1.49)	9.11 (1.23)
Perceived benefits, mean (SD)	13.04 (1.17)	12.72 (1.79)	14.76 (0.95)	13.20 (1.27)	14.46 (1.09)	13.30 (1.13)
Perceived barriers, mean (SD)	13.64 (2.43)	12.52 (2.37)	10.30 (1.60)	12.53 (2.07)	10.65 (1.80)	12.20 (1.41)
Self-efficacy, mean (SD)	22.36 (5.25)	23.24 (5.39)	33.79 (4.61)	23.45 (4.88)	31.21 (4.87)	23.89 (4.47)

[a]T0: baseline.

[b]T1: the fourth week.

[c]T2: the eighth week.

XSL•FO
RenderX

Table . Generalized estimating equation models for the comparison of outcomes across time between the 2 groups (number of each group=25).

Variables and parameters	Statistics	*P* value
Willingness to participate in breast cancer screening, OR[a] (95% CI)		
Group	1.000 (0.330 to 3.033)	>.99
T1[b]	1.625 (0.966 to 2.735)	.07
T2[c]	1.625 (0.966 to 2.735)	.07
Group*T1	4.889 (1.343 to 17.801)	.02
Group*T2	2.667 (0.717 to 9.912)	.14
Breast cancer screening knowledge, β[d] (95% CI)		
Group	−0.360 (−1.816 to 1.096)	.63
T1	0.773 (0.055 to 1.491)	.03
T2	1.268 (0.168 to 2.368)	.02
Group*T1	7.557 (6.031 to 9.083)	<.001
Group*T2	5.813 (4.055 to 7.570)	<.001
Perceived susceptibility, β (95% CI)		
Group	−0.520 (−1.058 to 0.018)	.06
T1	0.022 (−0.219 to 0.264)	.86
T2	0.238 (−0.011 to 0.487)	.06
Group*T1	1.835 (1.272 to 2.398)	<.001
Group*T2	1.440 (0.848 to 2.032)	<.001
Perceived severity, β (95% CI)		
Group	−0.720 (−1.650 to 0.210)	.13
T1	0.396 (−0.165 to 0.957)	.17
T2	0.995 (0.289 to 1.700)	.01
Group*T1	2.634 (1.721 to 3.547)	<.001
Group*T2	1.891 (0.713 to 3.070)	<.01
Perceived benefits, β (95% CI)		
Group	0.320 (−0.502 to 1.142)	.45
T1	0.479 (0.061 to 0.896)	.03
T2	0.576 (0.017 to 1.135)	.04
Group*T1	1.236 (0.581 to 1.891)	<.001
Group*T2	0.840 (0.055 to 1.626)	.04
Perceived barriers, β (95% CI)		
Group	1.120 (−0.182 to 2.422)	.09
T1	0.007 (−0.447 to 0.462)	.98
T2	−0.320 (−1.067 to 0.427)	.40
Group*T1	−3.351 (−4.502 to −2.199)	<.001
Group*T2	−2.670 (−4.007 to −1.262)	<.001
Self-efficacy, β (95% CI)		
Group	−0.880 (−3.769 to 2.009)	.55
T1	0.291 (−0.528 to 0.949)	.58
T2	0.647 (−1.006 to 2.299)	.44
Group*T1	11.224 (8.952 to 13.496)	<.001

XSL•FO
RenderX

Variables and parameters	Statistics	*P* value
Group*T2	8.206 (5.598 to 10.813)	<.001

[a]OR: odds ratio.

[b]T1: the fourth week.

[c]T2: the eighth week.

[d]β: regression coefficient.

Breast cancer screening knowledge scores increased from T0 to T1 in both groups, with a greater improvement observed in the intervention group. Although the intervention group's scores declined from T1 to T2, they remained above the baseline (Table 4). As shown in Table 5, GEE analysis indicated that knowledge scores in the intervention group tended to be higher than those in the control group at T1 (β=7.557, 95% CI 6.031-9.083) and T2 (β=5.813, 95% CI 4.055-7.570).

The intervention group showed increases from T0 to T1 in scores for perceived susceptibility, severity, and benefits. Although a slight decline occurred by T2, scores remained above baseline. Conversely, perceived barriers decreased at T1 before a slight rebound at T2. Throughout the study, scores in the control group showed minimal fluctuation (Table 4). Analysis of GEE indicated that changes in health belief scores across all dimensions tended to be greater in the intervention group compared with the control group (Table 5).

Regarding self-efficacy, scores in the intervention group increased from T0 to T1. Although a slight decline was observed at T2, they remained above baseline. Changes in the intervention group tended to be greater than those in the control group, which showed smaller fluctuations (Table 4). According to GEE analysis (Table 5), the improvement in self-efficacy in the intervention group was consistently greater than in the control group at both T1 (β=11.224, 95% CI 8.952-13.496) and T2 (β=8.206, 95% CI 5.598-10.813). Given the pilot and formative nature of this study, these findings should be interpreted as exploratory trends rather than conclusive evidence of effectiveness.

To evaluate the robustness of the findings to missing data, a sensitivity analysis was conducted by comparing the GEE results from the multiple imputed dataset with those obtained from the original dataset. Overall, the direction of the intervention effects was largely consistent across analyses, indicating that the findings were stable regardless of the method used to handle missing data. The detailed results of the sensitivity analysis are presented in Multimedia Appendix 2.

User Experience and Improvement Suggestions for the App

Overview

In total, 8 participants were interviewed. The participants ranged in age from 35 to 57 years (mean 45.38 years, SD 6.99), used the app for a total of 1 to 4 hours (mean 2.25 hours, SD 1.04), and interview duration ranged from 21 to 27 minutes (mean 23.75 minutes, SD 2.43). Detailed demographic and use characteristics are presented in Multimedia Appendix 3. Analysis of the interview transcripts initially generated 20 codes. Based on recurring patterns and relationships among the coded data, these codes were grouped into 4 subthemes, which were subsequently refined into 2 core themes: "perceived value and acceptability of the app" and "system limitations and user-centered improvement needs." The details are shown in Multimedia Appendix 4.

Theme 1: Perceived Value and Acceptability of the App

Participants provided positive feedback on the "Ruaikang" mobile health app, with its strengths and value summarized into 2 subthemes: hybrid approach to promote effectiveness and long-term or sustained use and significant health promotion effectiveness.

Hybrid Approach to Promote Effectiveness and Long-Term or Sustained Use

Participants reported that the app's content was professionally rigorous, clearly illustrated, and highly practical. Its information architecture was logically structured to cover the entire spectrum of breast health management, including disease awareness, prevention strategies, and daily self-management. They also reported that the app was simple and easy to use. The combination of high-quality content and user-friendly design enhanced participants' engagement with the app and fostered a strong willingness to continue using it. Furthermore, participants expressed a willingness to actively recommend the app to others, citing its professional content, practical utility, and ongoing updates. They viewed it as a valuable tool for educating friends and family about breast health and reducing anxiety, which further reinforced its potential for sustained use and wider dissemination.

> *The application provides comprehensive content, from disease awareness and prevention strategies to daily self-management strategies, eliminating the need to consult disparate online sources. This integrated approach offers considerable convenience. The detailed self-examination video facilitates effective self-screening at home, promoting earlier detection of potential issues. Furthermore, the user experience is very comfortable, with no unnecessary clutter, making all features readily accessible.* [S4]

> *People are paying more attention to self-care these days, and knowing about breast health is important for protection. The application provides professional, readily accessible information. I plan to use it often if it stays updated ... It is comprehensive and practical, useful for helping my mother and friends gain knowledge and reduce anxiety. I'll definitely recommend it.* [S2]

XSL•FO
RenderX

Significant Health Promotion Effectiveness

By providing professional content and practical features, the app contributed significantly to participants' breast health knowledge and awareness, while also supporting the development of healthier lifestyle habits.

> *The application taught me the warning signs for breast cancer and how to prevent it. It clearly explained the screening process, including when and how to get checked, which was so valuable. Thanks to the self-examination guide, I now check myself every week. It makes me feel much more secure.* [S5]

> *Now I make more mindful, healthier choices, like buying more of the recommended vegetables when shopping for groceries, controlling my dinner portions, and sticking to my daily health tracking. It's all helping me build solid habits.* [S8]

Theme 2: System Limitations and User-Centered Improvement Needs

Participants identified limitations in the "Ruaikang" mobile health app pertaining to its functionality, content, and user interaction and proposed constructive optimizations. This theme can be summarized into 2 subthemes: lack of active engagement and interactions and functional and content-related limitations.

Lack of Active Engagement and Interactions

Participants identified several issues related to active engagement and interaction within the app. They proposed adding an AI-powered question-and-answer system or chatbot to facilitate timely communication and support. Participants also recommended incorporating a check-in incentive mechanism to enhance user engagement. In addition, they highlighted the absence of a platform for user-to-user communication and suggested creating a community forum to facilitate discussion and experience sharing.

> *I would like to see AI-powered features added, such as an instant Q&A system or a chatbot ... Additionally, incorporating incentives into the Health Center, for example, rewarding users with a free consultation upon completing a health plan, would further boost engagement and support self-management.* [S3]

> *The interactive center lacks a section for users to discuss issues, and users cannot view others' shared experiences or respond to questions. I hope to see an interactive space where people can talk about problems they encounter and learn how others have solved them.* [S6]

Functional and Content-Related Limitations

Participants highlighted several limitations related to the app's interface and content. They noted ambiguous section boundaries, difficulty locating desired information, a lack of direct access to frequently used content, and highly repetitive visual elements, which reduced usability and visual comfort. Participants recommended optimizing page layouts, improving content organization, providing quicker access to commonly used information, and enhancing design diversity to improve the user experience. Regarding content presentation, participants reported that locating key information within videos was time-consuming and suggested providing concise summaries, such as text descriptions, slide decks, or flowcharts, to facilitate information retrieval and comprehension. Participants also perceived limitations in the existing content, particularly within the health center, and expressed a need for more detailed explanations of specialized topics, such as breast cancer staging, symptoms, and policy-related information. In addition, they emphasized the importance of regularly updating content to ensure that information remains current and relevant.

> *The dietary recommendations are buried within a broad category and lack a direct access point. Creating a more visible shortcut for such frequently used content would save significant time ... It would be helpful to summarize the key points from the video into a separate format, like a flowchart or a slide deck, to distill the core content for quick comprehension.* [S5]

> *Regarding breast cancer staging, I hope to see clearer explanations for each stage's definition, symptoms, and management approaches. The current symptom illustrations are only in cartoon form, I recommend adding realistic image options to make the content more comprehensive and meet diverse needs.* [S4.]

Discussion

Principal Findings

"Ruaikang" as a pioneering mobile health app designed specifically to promote early screening behaviors among high-risk populations for breast cancer demonstrated good feasibility and usability in this study. Preliminary findings suggested trends toward improvements in participants' willingness to participate in breast cancer screening, screening knowledge, health beliefs, and self-efficacy, although these results should be interpreted cautiously, given the exploratory, formative nature of the pilot study. Qualitative analysis further suggested that participants perceived its strengths in content expertise, user-friendly operation, practical functionality, and health promotion impact. However, it also identified areas for improvement in functional optimization, user interaction, and content presentation, which can inform subsequent refinement and development.

In this study, the recruitment rate reached 80.6% (50/62), and the completion rate was 76% (19/25), exceeding the feasibility thresholds for mobile health apps (recruitment rate ≥80% and completion rate ≥75%) [32], indicating that "Ruaikang" demonstrates strong feasibility among high-risk populations for breast cancer. High engagement levels not only confirm that the app's core functions effectively align with the health needs of the target group [46] but also reflect the potential of mobile health tools in early breast cancer prevention [47]. Moreover, qualitative findings corroborate these results, as participants reported the app's content to be practical and its features relevant to their needs. When conducting large-scale randomized controlled trials in the future, participation and completion rates can be further enhanced through optimizing recruitment

XSL•FO
RenderX

strategies [32], establishing strategies to support sustained engagement [48], and others.

The usability score of "Ruaikang" was 88.82 (SD 10.29), which is similar to the score (83.20) reported for the mobile health app developed by Rezaee et al [49], indicating that "Ruaikang" demonstrates good usability. Participants reported that the app was easy to operate, intuitive, and usable without professional guidance. These findings were further supported by use data and qualitative feedback, which together indicated active engagement and ease of navigation, reflecting strong learnability and user-friendliness. Several factors may explain this high level of usability. First, "Ruaikang" was specifically developed for high-risk populations for breast cancer, allowing the content to closely align with their unique needs, concerns, and risk perceptions. Prior studies have emphasized that tailoring health information to a defined target group is a key determinant of usability and engagement, as users are more likely to interact with information they perceive as personally relevant [50]. Second, the structured and comprehensive design of the app integrates educational content, self-management guidance, and practical tools in a logical and accessible manner, reducing the need to seek information from multiple sources and thereby enhancing convenience. Systematic reviews of mobile health usability note that clear organization and efficient information flow are central to perceived usability, as they reduce cognitive load and support efficient task completion [51]. Finally, the inclusion of multimedia resources, such as videos and images, improves clarity and engagement, providing a more interactive and motivating learning experience than text-based apps. Empirical studies suggest that multimedia elements can enhance understanding and make complex health information easier to process, which in turn supports sustained interaction with mobile health tools [52]. These features likely contributed to participants' strong engagement and willingness to use the app frequently, suggesting that tailored content, structured design, and interactive elements can significantly enhance the usability of mobile health interventions targeting high-risk populations for breast cancer. Nevertheless, the relatively high feasibility and usability observed in this study may be partly attributable to the recruitment of participants from a single tertiary hospital, all of whom owned and regularly used smartphones. Such participants likely possessed higher digital literacy and greater familiarity with mobile technologies, which facilitated their engagement with the app and led to more positive evaluations [53].

This study provides preliminary signals of improvement in willingness to participate in breast cancer screening, screening knowledge, health beliefs, and self-efficacy among high-risk populations for breast cancer, as supported by both quantitative and qualitative findings. These results are consistent with a systematic review of Salman et al [47], which evaluated mobile health apps in cancer screening and reported potential effects on screening participation and health outcomes. Importantly, several key outcome scores declined at the follow-up (T2) assessment, suggesting that the initial gains were not fully sustained. This pattern is consistent with prior research on mobile health interventions, which has shown that benefits often peak immediately after the intervention and may diminish over time without ongoing support, highlighting the potential importance of sustained engagement for maintaining behavior change [54]. In addition, the relatively wide CIs further limit the precision of these estimates and should be considered when interpreting the observed trends. Qualitative insights suggest that the decline in engagement and outcome scores at T2 may be interpreted as reflecting participants' need for more interactive and dynamically updated content, as well as motivational elements. These findings may highlight the need for strategies that extend beyond short-term initiation to promote continued engagement and behavior reinforcement. Future iterations of the app could incorporate more interactive and updated content, gamification mechanisms, personalized reminders, and theory-driven features to strengthen long-term engagement and maintenance of screening-related behaviors. Importantly, the primary outcome of this study was willingness to participate in screening, which reflects behavioral intention rather than actual screening uptake. However, these findings should be interpreted as hypothesis-generating rather than confirmatory, given the pilot nature of the trial. Future studies with larger samples and longer follow-up periods using fully powered randomized controlled trial designs are warranted to confirm these findings.

In addition to recognizing the app's strengths, participants also identified several areas for improvement, including functional and interaction features, interface and content presentation, and the depth and frequency of content updates. These comments do not negate that the app meets the core needs of high-risk populations for breast cancer and instead reflect users' expectations for greater convenience, enhanced interactivity, and content relevance in real-world use. The low frequency of interactive feature use may be related to inconspicuous entry points, delayed feedback, and limited opportunities for peer interaction. Consistent with research on mobile health interventions, users prefer features that facilitate active engagement and meaningful interaction, which are critical for sustained use [44]. Navigation and content presentation issues suggest a misalignment between the app's linear, chapter-based knowledge structure and the fragmented, problem-focused use patterns of high-risk populations for breast cancer. Research on user-centered design for breast cancer-related apps has highlighted that clear navigation tools and well-organized content are key factors influencing usability and engagement [55]. Feedback regarding content depth and update frequency further indicates that users expect information to be directly relevant to their health status and screening decisions, as well as updated promptly to reflect the latest evidence and guidelines. Overall, these improvement suggestions emphasize the importance of usability, interactivity, and content relevance for high-risk populations for breast cancer, pointing to future optimization strategies such as enhancing interactive features, reorganizing information architecture for rapid access, and increasing content update frequency, thereby strengthening user engagement and the overall value of the app.

Limitations

This study has several limitations. First, the small sample size from one hospital inherent in this pilot study limits the ability to draw definitive conclusions, and the preliminary effectiveness

XSL•FO
RenderX

findings should therefore be interpreted as hypothesis-generating rather than confirmatory. Second, the enrolled high-risk populations for breast cancer in this study did not encompass individuals with genetic susceptibility or a history of chest radiotherapy. Subsequent studies should incorporate these groups to improve the generalizability of the results. Third, willingness to participate in screening was selected as the primary outcome, which represents behavioral intention rather than actual screening uptake. Future multicenter randomized controlled trials with longer follow-up periods are warranted to further evaluate the preliminary findings and to determine whether improvements in screening willingness translate into actual screening participation. Finally, the app was developed only for Android devices, which may limit the generalizability of feasibility and usability findings to iPhone users.

Implications and Future Work

This pilot study demonstrates that the "Ruaikang" mobile health app exhibits good feasibility and usability among high-risk populations for breast cancer, providing preliminary signals of potential effectiveness in supporting health-related beliefs. The app is well-aligned with users' core needs and may contribute to improvements in screening intention, although these effects should be interpreted cautiously, given the exploratory nature of the study. "Ruaikang" could serve as a key bridge connecting health care resources with high-risk populations for breast cancer. Through continuous refinement of health intervention strategies, it may facilitate broader adoption of early screening practices, alleviate the public health burden of breast cancer among Chinese women, and enhance their sense of health empowerment and overall well-being.

In future work, the research team will systematically refine the app's core features, user interaction, content depth, and presentation based on feedback collected during the pilot phase. A larger-scale, multicenter randomized controlled trial with an expanded sample covering diverse geographic regions will be conducted to further validate these preliminary findings. In addition, a long-term follow-up mechanism will be established to monitor the sustainability of screening behaviors and changes in early detection rates, thereby generating longitudinal evidence to support the app's continued development and broader implementation.

Conclusions

This study demonstrated that the "Ruaikang" mobile health app is feasible and usable and provides preliminary signals of potential effectiveness for high-risk populations for breast cancer. The app may contribute to enhancing awareness and supporting willingness to participate in breast cancer screening among this population. Future research will optimize the app based on participant feedback and conduct studies with larger samples and longer follow-up periods to further examine the sustainability and potential impact on screening behaviors among high-risk populations for breast cancer.

Acknowledgments

The authors appreciate all women who agreed to participate in the study. The authors confirm that no generative AI or AI-assisted tools were used in any part of the manuscript's writing, analysis, or figure preparation.

Funding

This study was supported by the National Science and Technology Innovation 2030, Noncommunicable Chronic Diseases-National Science and Technology Major Project (grants 2024ZD0524300 and 2024ZD0524305), and the National Natural Science Foundation of China (grant 72204209).

Data Availability

The datasets used and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: LZ, FZ, JH
Methodology: LZ, XL
Software: LZ, MX
Formal analysis: LZ, MX
Data curation: LZ, XL
Resources: XZ, XL
Validation: XZ, SW, JH
Investigation: XZ
Project administration: FZ, JH
Funding acquisition: FZ, JH
Writing—original draft: LZ, MX
Writing—review and editing: SW, FZ, JH

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
The detailed information on the app's development process, user side features, and back-end management functions.
[DOCX File, 15748 KB - mhealth_v14i1e86429_app1.docx]

Multimedia Appendix 2
Sensitivity analysis of generalized estimating equations models before and after multiple imputation.
[DOCX File, 26 KB - mhealth_v14i1e86429_app2.docx]

Multimedia Appendix 3
Demographic and use characteristics of the participants in the qualitative study.
[DOCX File, 16 KB - mhealth_v14i1e86429_app3.docx]

Multimedia Appendix 4
The results of the qualitative study.
[DOCX File, 30 KB - mhealth_v14i1e86429_app4.docx]

Checklist 1
CONSORT checklist.
[PDF File, 41 KB - mhealth_v14i1e86429_app5.pdf]

References

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024;74(3):229-263. [doi: 10.3322/caac.21834] [Medline: 38572751]
2. Wu Q, Fan BM, Li Y. Analysis and interpretation of the 2022 Global Cancer Statistics Report: China and the world's cancer disease burden and epidemiological trends. Theory Pract Diagn 2025;24(2):135-145. [doi: 10.16150/j.1671-2870.2025.02.004]
3. Shen SJ, Sun Q, Huang X, et al. Chinese guidelines for breast cancer screening in women. J Chin Res Hosp 2022;9(2):1-6. [doi: 10.19450/j.cnki.jcrh.2022.02.003]
4. Pilewskie M, Eroglu I, Sevilimedu V, Le T, Mangino D, Morrow M. Participation in a high-risk program is associated with a diagnosis of earlier-stage disease among women at increased risk for breast cancer development. Ann Surg Oncol 2024 Oct;31(10):6764-6773. [doi: 10.1245/s10434-024-15633-x] [Medline: 38949720]
5. Cohen SY, Stoll CR, Anandarajah A, Doering M, Colditz GA. Modifiable risk factors in women at high risk of breast cancer: a systematic review. Breast Cancer Res 2023 Apr 24;25(1):45. [doi: 10.1186/s13058-023-01636-1] [Medline: 37095519]
6. Shi Q, Gao J, Yu H, et al. Characterization and functional analysis of BRCA1 and BRCA2 variants in a cohort of 100 unselected patients undergoing germline screening. Transl Oncol 2025 Sep;59:102471. [doi: 10.1016/j.tranon.2025.102471] [Medline: 40664060]
7. Cardoso R, Hoffmeister M, Brenner H. Breast cancer screening programmes and self-reported mammography use in European countries. Int J Cancer 2023 Jun 15;152(12):2512-2527. [doi: 10.1002/ijc.34494] [Medline: 36883419]
8. Healthy China Initiative (2019–2030). National Health Commission of the People's Republic of China. 2019. URL: https://www.nhc.gov.cn/guihuaxxs/c100133/201907/2a6ed52f1c264203b5351bdbbadd2da8.shtml [accessed 2025-11-29]
9. Kırca N, Tuzcu A, Gözüm S. Breast cancer screening behaviors of first degree relatives of women receiving breast cancer treatment and the affecting factors. Eur J Breast Health 2018 Jan;14(1):23-28. [doi: 10.5152/ejbh.2017.3272] [Medline: 29322115]
10. Zhang X, Yang L, Liu S, et al. Evaluation of different breast cancer screening strategies for high-risk women in Beijing, China: a real-world population-based study. Front Oncol 2021;11:776848. [doi: 10.3389/fonc.2021.776848]
11. Zhao X, Meng HY, Sun GL, et al. Knowledge, attitude and practice on early screening for breast cancer of female first-degree relatives of breast cancer patients. Chin J Modern Nurs 2019;5(20):2581-2585 [FREE Full text]
12. Bevers TB, Niell BL, Baker JL, et al. NCCN Guidelines® insights: Breast Cancer Screening and Diagnosis, Version 1.2023. J Natl Compr Canc Netw 2023 Sep;21(9):900-909. [doi: 10.6004/jnccn.2023.0046] [Medline: 37673117]
13. Noman S, Elarusy NME, Rahman HA, et al. Investigating the effect of the educational intervention based on the Health Belief Model on the knowledge and beliefs of Yemeni teachers in the use of breast cancer screening: a randomized controlled trial study. BMC Cancer 2024 Dec 6;24(1):1506. [doi: 10.1186/s12885-024-13214-5] [Medline: 39643866]
14. Kowitt SD, Ellis KR, Carlisle V, et al. Peer support opportunities across the cancer care continuum: a systematic scoping review of recent peer-reviewed literature. Support Care Cancer 2019 Jan;27(1):97-108. [doi: 10.1007/s00520-018-4479-4] [Medline: 30293093]

XSL•FO
RenderX

15. Molokwu JC, Dwivedi A, Alomari A, Shokar N. Effectiveness of a Breast Cancer Education Screening and NavigaTion (BEST) intervention among Hispanic women. Health Promot Pract 2024 Nov;25(6):1070-1081. [doi: 10.1177/15248399221135762] [Medline: 36635866]
16. Hersch J, Barratt A, Jansen J, et al. Use of a decision aid including information on overdetection to support informed choice about breast cancer screening: a randomised controlled trial. Lancet 2015 Apr 25;385(9978):1642-1652. [doi: 10.1016/S0140-6736(15)60123-4] [Medline: 25701273]
17. Baron RC, Rimer BK, Breslow RA, et al. Client-directed interventions to increase community demand for breast, cervical, and colorectal cancer screening a systematic review. Am J Prev Med 2008 Jul;35(1 Suppl):S34-S55. [doi: 10.1016/j.amepre.2008.04.002] [Medline: 18541187]
18. Aguiar-Ibáñez R, Mbous YPV, Sharma S, Chakali R, Chawla E. Barriers to cancer screening uptake and approaches to overcome them: a systematic literature review. Front Oncol 2025;15:1575820. [doi: 10.3389/fonc.2025.1575820]
19. Sprague BL, Ahern TP, Herschorn SD, Sowden M, Weaver DL, Wood ME. Identifying key barriers to effective breast cancer control in rural settings. Prev Med 2021 Nov;152(Pt 2):106741. [doi: 10.1016/j.ypmed.2021.106741] [Medline: 34302837]
20. Tavakoli B, Feizi A, Zamani-Alavijeh F, Shahnazi H. Factors influencing breast cancer screening practices among women worldwide: a systematic review of observational and qualitative studies. BMC Womens Health 2024 Apr 27;24(1):268. [doi: 10.1186/s12905-024-03096-x] [Medline: 38678224]
21. Global strategy on digital health 2020-2025. World Health Organization. URL: https://iris.who.int/handle/10665/344249 [accessed 2025-11-29]
22. Baniasadi T, Hassaniazad M, Rostam Niakan Kalhori S, Shahi M, Ghazisaeedi M. Developing a mobile health application for wound telemonitoring: a pilot study on abdominal surgeries post-discharge care. BMC Med Inform Decis Mak 2023 Jun 2;23(1):103. [doi: 10.1186/s12911-023-02199-z] [Medline: 37268995]
23. Li B, Heydari K, Enichen EJ, Kvedar JC. A mobile health application that supports a patient centered approach to cardiovascular risk management. NPJ Digit Med 2025;8(1):150. [doi: 10.1038/s41746-025-01549-7]
24. Mugabirwe B, Flickinger T, Cox L, Ariho P, Dillingham R, Okello S. Acceptability and feasibility of a mobile health application for blood pressure monitoring in rural Uganda. JAMIA Open 2021 Jul;4(3):ooaa068. [doi: 10.1093/jamiaopen/ooaa068] [Medline: 34514350]
25. Yusuf A, P Iskandar YH, Ab Hadi IS, Nasution A, Lean Keng S. Breast awareness mobile apps for health education and promotion for breast cancer. Front Public Health 2022;10:951641. [doi: 10.3389/fpubh.2022.951641] [Medline: 36324460]
26. Roh S, Lee YS, Kenyon DB, et al. Mobile web app intervention to promote breast cancer screening among American Indian women in the Northern Plains: feasibility and efficacy study. JMIR Form Res 2023 Jul 20;7:e47851. [doi: 10.2196/47851] [Medline: 37471115]
27. Eldridge SM, Chan CL, Campbell MJ, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Pilot Feasibility Stud 2016;2:64. [doi: 10.1186/s40814-016-0105-8] [Medline: 27965879]
28. Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharm Stat 2005 Oct;4(4):287-291. [doi: 10.1002/pst.185]
29. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
30. Li XX, Du XW, Song W, Lu C, Hao WN. Effect of continuous nursing care based on the IKAP theory on the quality of life of patients with chronic obstructive pulmonary disease. Medicine (Baltimore) 2020 Mar;99(11):e19543. [doi: 10.1097/MD.0000000000019543]
31. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. Health Educ Q 1988;15(2):175-183. [doi: 10.1177/109019818801500203] [Medline: 3378902]
32. Josefsson C, Liljeroos T, Hellgren M, Pöder U, Hedström M, Olsson EMG. The Sukaribit smartphone app for better self-management of type 2 diabetes: randomized controlled feasibility study. JMIR Form Res 2024 Jan 10;8:e46222. [doi: 10.2196/46222] [Medline: 38198200]
33. Wang Y, Lei T, Liu X. Chinese System Usability Scale: translation, revision, psychological measurement. Int J Hum Comput Interact 2020 Jun 14;36(10):953-963. [doi: 10.1080/10447318.2019.1700644]
34. Bolz J, Löscher A, Muhl R, Badke A, Predel HG, Perret C. Feasibility, usability, and safety of ParaGym, an intelligent mobile exercise app for individuals with paraplegia: protocol for a pilot block-randomized controlled trial. JMIR Res Protoc 2023 May 19;12:e45652. [doi: 10.2196/45652] [Medline: 37204855]
35. Chen Y, Xia HO, Oakley D, et al. A survey on women's awareness of breast cancer and its early screening in Shanghai. Shanghai Nurs 2007;7(3):13-16 [FREE Full text]
36. Lu J, Ren H, Liu Y, et al. Knowledge, attitude, and willingness toward breast magnetic resonance imaging screening among women at high risk of breast cancer in Beijing, China. BMC Public Health 2024 Oct 21;24(1):2909. [doi: 10.1186/s12889-024-20370-7] [Medline: 39434008]
37. Ye X, Xu T, Cao J, et al. Knowledge, attitude, and practice toward genetic testing in breast cancer patients in China. PLoS One 2025;20(5):e0322526. [doi: 10.1371/journal.pone.0322526]

XSL•FO
RenderX

38. Zhang X. Study on Influencing Factors of Breast Cancer Screening Behavior Among Women Based on the Health Ecology Model [Master's thesis]. : Jilin University; 2021 URL: https://tinyurl.com/2jdkmfza [accessed 2026-07-08]
39. Liao Y, Mohd Hairon S, Majdi Yaacob N, Alina Tengku Ismail T, Luo L. Psychometric validation of a culturally adapted health belief model scale for breast cancer screening in Chinese women. PLoS One 2025;20(9):e0331279. [doi: 10.1371/journal.pone.0331279] [Medline: 40901788]
40. Schwarzer R, Jerusalem M. Generalized self-efficacy scale. In: Weinman J, Wright S, editors. Measures in Health Psychology: A User's Portfolio Causal and Control Beliefs: NFER-NELSON; 1995:35-37.
41. Mystakidou K, Parpa E, Tsilika E, Galanos A, Vlahos L. General perceived self-efficacy: validation analysis in Greek cancer patients. Support Care Cancer 2008 Dec;16(12):1317-1322. [doi: 10.1007/s00520-008-0443-z] [Medline: 18392861]
42. Wang C. General Self-Efficacy Scale: Chinese version and psychometric evaluation. Chin J Clin Psychol 2001;9(4):123-125. [doi: 10.3969/j.issn.1006-6020.2001.01.007]
43. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view1. MIS Q 2003 Sep 1;27(3):425-478. [doi: 10.2307/30036540]
44. Anders C, Moorthy P, Svensson L, et al. Usability and user experience of an mHealth app for therapy support of patients with breast cancer: mixed methods study using eye tracking. JMIR Hum Factors 2024 Mar 5;11:e50926. [doi: 10.2196/50926] [Medline: 38441959]
45. Willig C, Rogers WS. The SAGE Handbook of Qualitative Research in Psychology: SAGE Publications Ltd; 2017.
46. Hommel KA, Noser AE, Plevinsky J, Gamwell K, Denson LA. Pilot and feasibility of the SMART IBD mobile app to improve self-management in pediatric inflammatory bowel disease. J Pediatr Gastroenterol Nutr 2024 Jun;78(6):1273-1278. [doi: 10.1002/jpn3.12229] [Medline: 38682397]
47. Salmani H, Ahmadi M, Shahrokhi N. The impact of mobile health on cancer screening: a systematic review. Cancer Inform 2020;19:1176935120954191. [doi: 10.1177/1176935120954191] [Medline: 33116352]
48. High J, Grant K, Hope A, et al. Effects of an increased financial incentive on follow-up in an online, automated smoking cessation trial: a randomized controlled study within a trial. Nicotine Tob Res 2024 Aug 22;26(9):1259-1263. [doi: 10.1093/ntr/ntae068] [Medline: 38513087]
49. Rezaee R, Asadi S, Yazdani A, Rezvani A, Kazeroon AM. Development, usability and quality evaluation of the resilient mobile application for women with breast cancer. Health Sci Rep 2022 Jul;5(4):e708. [doi: 10.1002/hsr2.708] [Medline: 35782301]
50. Lustria MLA, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
51. Galavi Z, Montazeri M, Khajouei R. Which criteria are important in usability evaluation of mHealth applications: an umbrella review. BMC Med Inform Decis Mak 2024 Nov 29;24(1):365. [doi: 10.1186/s12911-024-02738-2] [Medline: 39614283]
52. Galmarini E, Marciano L, Schulz PJ. The effectiveness of visual-based interventions on health literacy in health care: a systematic review and meta-analysis. BMC Health Serv Res 2024 Jun 11;24(1):718. [doi: 10.1186/s12913-024-11138-1] [Medline: 38862966]
53. Durmuş A. The influence of digital literacy on mHealth app usability: the mediating role of patient expertise. Digit Health 2024;10:20552076241299061. [doi: 10.1177/20552076241299061] [Medline: 39600388]
54. Oakley-Girvan I, Yunis R, Fonda SJ, et al. Usability evaluation of mobile phone technologies for capturing cancer patient-reported outcomes and physical functions. Digit Health 2023;9. [doi: 10.1177/20552076231186515] [Medline: 37456127]
55. Nuseibeh BZ, Johns SA, Shih PC, Lewis GF, Gowan TM, Jordan EJ. Co-designing the MOSAIC mHealth app with breast cancer survivors: user-centered design approach. JMIR Form Res 2024 Dec 9;8:e59426. [doi: 10.2196/59426] [Medline: 39652855]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
GEE: generalized estimating equation
HBM: health belief model
NCCN: National Comprehensive Cancer Network

XSL•FO
RenderX

Edited by L Buis; submitted 24.Oct.2025; peer-reviewed by F Liu, J Zhu; revised version received 15.Jun.2026; accepted 15.Jun.2026; published 14.Jul.2026.

Please cite as:
Zhang L, Xu M, Zhang X, Li X, Wang S, Zhou F, Han J
Feasibility, Usability, and Preliminary Effectiveness of an mHealth App to Promote Screening Behaviors Among High-Risk Populations for Breast Cancer: Randomized Controlled Pilot Study
JMIR Mhealth Uhealth 2026;14:e86429
URL: https://mhealth.jmir.org/2026/1/e86429
doi:10.2196/86429

© Linping Zhang, Mengjiao Xu, Xian Zhang, Xiaoxu Li, Shuo Wang, Fang Zhou, Jing Han. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

The ManageHF Just-in-Time Adaptive Mobile App Interventions to Promote Self-Management and Improve Outcomes in Heart Failure: Randomized Controlled Trial

Michael P Dorsch[1,2], MS, PharmD; Mohamed S Ali[1], MS, PharmD; Amy Krambrink[3], MS; Giselle Kolenic[3], MA; Sabah Ganai[1], MPH; Juan Arzac[1], MA; Xutong Zhang[1], MS; Kaitlyn M Greer[1], PhD; Amit J Shah[4], MSCR, MD; Jennifer A Cowger[5], MS, MD; Gregory Ewald[6], MD; Jo Ellen Rodgers[7], PharmD; Dave L Dixon[8], PharmD; Todd M Koelling[2,9], MD; Scott L Hummel[2,9,10], MS, MD

[1]College of Pharmacy, University of Michigan–Ann Arbor, 248 Church Street, Ann Arbor, United States

[10]VA Ann Arbor Health System, Ann Arbor, MI, United States

[2]Frankel Cardiovascular Center, University of Michigan, Ann Arbor, MI, United States

[3]Biostatistics, School of Public Health, University of Michigan–Ann Arbor, Ann Arbor, MI, United States

[4]Department of Epidemiology, Rollins School of Public Health, Emory Univeristy, Atlanta, GA, United States

[5]Henry Ford Health System, Detroit, MI, United States

[6]School of Medicine, Washington University, St Louis, MO, United States

[7]Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[8]School of Pharmacy, Virginia Commonwealth University, Richmond, VA, United States

[9]Medical School, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Michael P Dorsch, MS, PharmD
College of Pharmacy, University of Michigan–Ann Arbor, 248 Church Street, Ann Arbor, United States

Abstract

Background: Heart failure (HF) is a major health care challenge in the United States, with approximately 900,000 older adults hospitalized annually. Gaps in self-management, including unrecognized worsening symptoms and failure to adhere to dietary sodium restriction, can reduce quality of life and precipitate hospital admissions. Existing mobile health approaches to HF self-management have produced mixed results, highlighting the need for innovative strategies to improve postdischarge outcomes in at-risk patients.

Objective: The ManageHF trial aimed to evaluate the effectiveness of 2 just-in-time adaptive interventions delivered via a mobile app to enhance HF self-management. The interventions focused on symptom recognition and lower dietary sodium restriction, with the goal of reducing readmissions and improving health-related quality of life (HRQOL) over a 12-week period.

Methods: The trial was a 2×2 factorial, double-blind, randomized controlled study conducted across several US institutions. Participants recently hospitalized for acute HF were randomized into 4 groups: both interventions, either intervention alone, or an active control. The primary outcome was a composite measure assessing time to all-cause death, time to first HF readmission, and HRQOL changes, using the Minnesota Living with HF Questionnaire.

Results: Recruitment was hindered by the COVID-19 pandemic, leading to the early discontinuation of the trial. Of the 62 participants enrolled, 43 completed the study. Participants were diverse, with a mean age of 55 (SD 14.4) years, 32% (20/62) were female, and 55% (34/62) identified as Black or African American. Most had HF with reduced ejection fraction. However, due to the early termination and small sample size, the ability to detect statistically significant differences was limited.

Conclusions: The ManageHF trial highlighted the potential of mobile health technology to support HF management, particularly in enhancing HRQOL. Future studies using more effective recruitment and retention strategies are crucial for establishing the efficacy of these interventions with greater certainty.

Trial Registration: ClinicalTrials.gov NCT04755816; https://www.clinicaltrials.gov/study/NCT04755816

(JMIR Mhealth Uhealth 2026;14:e74121) doi:10.2196/74121

KEYWORDS

mobile health; mHealth; just-in-time adaptive interventions; self-management; heart failure; dietary sodium

XSL•FO
RenderX

Introduction

Heart failure (HF) is the leading hospital discharge diagnosis among older adults in the United States, affecting 1.2 million patients in 2021 [1]. Roughly 20% of these patients are readmitted within 30 days, and half are readmitted within 6 months [2-8]. HF readmissions represent significant and potentially avoidable costs, with hospitalizations accounting for nearly 70% of annual HF costs [4]. Approximately one-third of patients who survive an HF hospitalization die within a year [9]. Many US hospitals have established quality improvement teams focused on 4 key domains: enhancing patient education, increasing the use of evidence-based therapies, ensuring timely postdischarge follow-up, and improving communication between inpatient and outpatient care providers [10,11]. However, despite these initiatives, results have been mixed, with recent improvements in readmission rates possibly attributable to changes in risk coding rather than actual outcomes [12,13]. Both crude and risk-adjusted 30-day mortality rates for HF have increased, indicating a need for new approaches to improve postdischarge outcomes [14,15].

Challenges with HF self-management, including failure to recognize clinical worsening and nonadherence to dietary sodium restrictions [2-6], are primary drivers of readmissions. Studies suggested that better dietary adherence, early help-seeking, and adequate medical support could prevent 23% to 31% of readmissions [3], reducing the 30-day readmission rate from approximately 20% to roughly 13.8% to 15.4%.

Self-monitoring has been shown to improve health-related quality of life (HRQOL) in patients with HF [16,17]. In a prospective single-center study, a web application designed to promote self-regulation was tested over 12 weeks, yielding positive outcomes for chronic HF symptoms [18]. Participants, predominantly older adults with New York Heart Association class II or III symptoms, used the application daily. Improvements were observed in the New York Heart Association classification and the Minnesota Living with HF Questionnaire (MLHFQ) scores, with notable numerical improvements in weight, exercise frequency, walking distance, jugular venous distension, and peripheral edema.

Dietary nonadherence, particularly excess dietary sodium intake, is a frequent precursor to HF decompensation and hospitalization [2-6]. Chronic neurohormonal activation and reduced renal perfusion in patients with HF promote sodium and fluid retention. Historically, guidelines have advocated a dietary sodium restriction of 2000 to 3000 mg/d [19,20], and current HF guidelines recommend avoiding “excessive” sodium intake [21]. In the Study of Dietary Intervention under 100 mmol in HF trial, aggressive sodium restriction did not reduce clinical outcomes. Still, it was associated with improved HRQOL [22] in stable outpatients with HF. The baseline sodium intake in the Study of Dietary Intervention under 100 mmol in HF was approximately 30% less than in US patients hospitalized for HF [23], who could derive more significant benefits from low-sodium dietary interventions [24].

The ManageHF trial aimed to evaluate the effectiveness of 2 just-in-time adaptive interventions (JITAIs) within a mobile app. These interventions targeted the two leading causes of hospital readmissions: (1) failure to recognize clinical worsening and (2) dietary nonadherence to lower-sodium food options. The study assessed the impact of JITAIs directed at HF symptoms and diet on mortality, hospital readmissions, and HRQOL over 12 weeks, hypothesizing that each intervention would yield reductions in mortality and readmissions, as well as improvements in HRQOL, as measured by the MLHFQ.

Methods

General Design

ManageHF was a 2 × 2 factorial, double-blind, randomized, controlled, multicenter clinical trial to determine the efficacy of JITAIs using a mobile app that focused on self-management and dietary support, either together or separately, compared with a control. Participants and study personnel were blinded to participant group assignment throughout the study. These interventions build upon previous work done by the research team [18,25,26]. The University of Michigan was the lead site, along with investigators at Henry Ford Health System, Virginia Commonwealth University, University of North Carolina, Washington University, University of Alabama at Birmingham, and Emory University. Eligible participants were randomized in a 1:1:1:1 manner to 4 conditions and stratified by site, gender, and HF type (HF with reduced ejection fraction vs HF with preserved ejection fraction). The four conditions were (1) diet and clinical worsening interventions (both), (2) diet intervention only, (3) clinical worsening intervention only, and (4) active control (neither intervention). All enrolled participants were asked to self-monitor with the mobile app, which included a daily survey, a blood pressure cuff, a scale, and a wearable activity monitor (3 devices provided to participants by Withings). The study protocol was updated throughout the study to reflect changes in the study workflow and updates to participant eligibility criteria. A full description of all protocol changes can be found in Multimedia Appendix 1. Advarra acted as the single institutional review board and approved this study. The trial was registered at ClinicalTrials.gov (NCT04755816; March 19, 2021) and was funded by the National Institute on Aging (R01AG062582).

Study Participants

To participate in the study, individuals had to provide informed consent and meet the following criteria: be aged 18 years or older, hospitalized or recently hospitalized (within 14 d) for acute or acute-on-chronic decompensated HF, and fulfill specific heart function requirements, either left ventricular ejection fraction less than or equal to 40% or left ventricular ejection fraction greater than 40% with elevated B-type natriuretic peptide or N-terminal pro-B-type natriuretic peptide levels adjusted for BMI. Additional criteria included having a personal physician for follow-up, a compatible smartphone for the ManageHF app, a valid email address, being fluent in English, and providing signed informed consent. Participants were excluded if there was a contraindication to recommending a sodium-restricted diet; planned intervention for primary valvular heart disease during the study period; cardiac resynchronization therapy; on dialysis; previous cardiac transplantation or

XSL•FO
RenderX

implantation of a ventricular assistance device or implantation expected within 3 months after randomization; cardiac or noncardiac illness with expected survival less than 3 months based on clinical judgment; listed status 1, 2, or 3 for a heart transplant; discharged to a setting other than their home; requirement for chronic inotropic therapy; inability to use Withings devices (weight scale, blood pressure monitor, and watch) due to equipment limitations or contraindications; or were currently pregnant or intended to become pregnant during the study period.

Potential participants were identified from current HF admissions at the enrollment site. The study teams were responsible for identifying eligible participants using inclusion and exclusion criteria. A potential participant was approached by a study team member to obtain consent before enrollment. Once enrolled, participants completed baseline questionnaires and surveys, created accounts with both Withings and ManageHF, and downloaded the ManageHF app.

Intervention

The University of Michigan developed a mobile app named ManageHF, which was studied in this clinical trial. The app featured 2 novel JITAIs activated based on randomization. The mobile app aimed to facilitate self-management and behavioral change by providing participants with real-time, contextual feedback and support. This approach aimed to leverage adaptive technologies and personalized interventions to improve participants' health outcomes and dietary habits.

The clinical worsening intervention focused on enhancing self-monitoring and self-management skills [27]. It used a health status indicator that adapted to users' reported symptoms and provided feedback. The health status indicator was presented in a speedometer-like graphic that ranges from green (good health) to red (poor health) based on the answers to daily app questionnaires. Participants received 2 push notifications per day if they were in the yellow zone (fair health) and 3 per day if they were in the red zone, until they returned to the green zone. These notifications are linked to in-app educational resources to promote self-management behaviors. Personalization involved a baseline survey adapted from the Self-Care in HF Index [28] that gauged the user's confidence and past behaviors in managing symptoms. These data helped tailor the messages for each user, using their name, emojis, and appropriate times of day to engage them effectively.

The dietary sodium intervention was designed to help users make lower-sodium choices by sending contextual messages linked to tailored in-app content [26]. The intervention notified the user when they were in a grocery store or restaurant. The Universal Product Code (UPC) in grocery stores helps to view tailored dietary information from the store, helping them identify and select lower-sodium options. Personalization was based on a confidence survey and a sodium intake screener, which asked users to identify their high-sodium foods and lower-sodium alternatives. Push notifications were purposely framed to align with the user's confidence and dietary habits.

Both JITAIs relied on NumberEight, an internet-based service that uses various sensors to predict the user's location. The NumberEight algorithm ran on the device, and data from these sensors were anonymized to ensure participant privacy and stored securely when the app was active. The study staff did not have access to participants' location data.

Outcomes

The primary outcome was a hierarchical composite of time to all-cause death, time to first HF readmission, and change from baseline to week 12 in the MLHFQ using the win ratio [29]. Death from any cause occurring during the study period was reported by each site. HF readmission was defined as admission to an inpatient unit or a stay in the emergency department following discharge after the index hospitalization, resulting in at least a 12-hour stay [30]. Only admissions associated with events that occur on an emergency or unplanned basis were considered potential end points and adjudicated. An HF readmission must have clinical manifestation of new dyspnea, orthopnea, paroxysmal nocturnal dyspnea, edema, pulmonary basilar rales or crackles, jugular venous distension, renal hypoperfusion with no other apparent cause, or radiological evidence of acute HF, and additional increased treatment with either intravenous HF medications (diuretics, inotropes, and vasodilators) or mechanical or surgical intervention directed explicitly at the treatment of HF. Only readmissions during the study period were assessed.

The MLHFQ tool comprises 21 questions that assess patients' perceptions of the effects of HF on their daily lives [31]. The questions produce a total score ranging from 0 to 105, with higher scores indicating poorer HRQOL. The MLHFQ was collected at baseline, 6 weeks, and 12 weeks. A score change of 5 or more was considered a clinically significant difference for the hierarchical composite end point.

Nutritional questionnaires were used to assess the participants' nutritional status and sodium intake. The Food Frequency Questionnaire (FFQ), collected at baseline, 6 weeks, and 12 weeks, was used to measure sodium intake. The 110-item Block FFQ (NutritionQuest, LLC) was electronically self-administered to record commonly consumed foods and estimate nutrient and energy intake.

Statistical Approach

The primary outcome was a composite of death, HF readmission, and change in quality of life using the win ratio method, which incorporates a hierarchy of clinical prioritization on event occurrence. The win ratio was chosen to prioritize outcomes in a patient-focused fashion using a composite end point. Following the Finkelstein and Schoenfeld unmatched pairs win ratio method [32], each participant receiving the intervention was compared to each who did not receive the intervention, resulting in pairwise comparisons [29,33]. Within each pairing, the participant with the longest survival time is counted as the "winner" of that pairing. The win ratio is the number of winners divided by the number of losers associated with the intervention. Ties were excluded. The estimate of the win ratio and its 95% CI were calculated using bootstrap methods. The test for intervention effect is a nonparametric rank sum test. A win ratio of 1 indicates that both comparison groups have equivalent outcomes in terms of events and quality of life. In contrast, a

XSL•FO
RenderX

win ratio greater than 1 indicates that the intervention group has a favorable profile compared to the control group, while a ratio less than 1 indicates an unfavorable profile. Using the intention-to-treat approach, the primary outcome was compared between each intervention and control. Based on standard 2 × 2 factorial designs, the clinical worsening intervention included those randomized to both the interventions and the clinical worsening-only conditions. The dietary sodium intervention included those randomized to both interventions and the dietary sodium–only conditions. The control for each intervention was the 2 remaining groups of the 4 randomized conditions.

The win ratio for the hierarchical end-point was estimated from our pilot study, where the win ratio was 1.33 in favor of the intervention group [27], meaning that people who received the intervention would be 33% more likely to experience favorable outcomes compared to the group not given the intervention. Assuming 10% dropout, 500 randomized participants (250 per main intervention arm) would provide 80% power to show a win ratio of 1.30 (56.6% winners in the intervention group excluding ties), assuming a 2-sided 5% type I error rate [29,32,33]. After accounting for dropout, 225 participants per arm would produce 50,400 (225 × 224) permutations of paired comparisons. The same reasoning was used to determine the power and sample size necessary to analyze the dietary sodium intervention. We did not design the study for sufficient power to detect an interaction effect of both interventions relative to the control.

For secondary end-points, each intervention versus control was examined in terms of individual components of the composite outcome using survival methods for time-to-event outcomes and linear mixed-effects models or piecewise linear regression for longitudinal continuous outcomes (eg, MLHFQ). Days alive and out of hospital (DAOH) were analyzed using Poisson regression. SAS (Statistical Analysis System; version 9.4; SAS Institute Inc) was used for all statistical analyses. Data are also presented in mean and SD for continuous variables normally distributed and median and IQR for those not normally distributed. Categorical variables are presented as a percentage of the total in that group. For completeness, outcome variables are presented across all 4 randomized conditions.

Ethical Considerations

Advarra acted as the single institutional review board and approved this study (Pro00046349). Participants signed an informed consent document before participation in the clinical trial. After the study was completed, the data were deidentified for analysis and eventual release into a repository. Participants were compensated with US $200 for their time spent completing study-related tasks, and all participants were allowed to keep the devices provided during enrollment in the study. The reporting of the trial follows the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines (Checklist 1).

Results

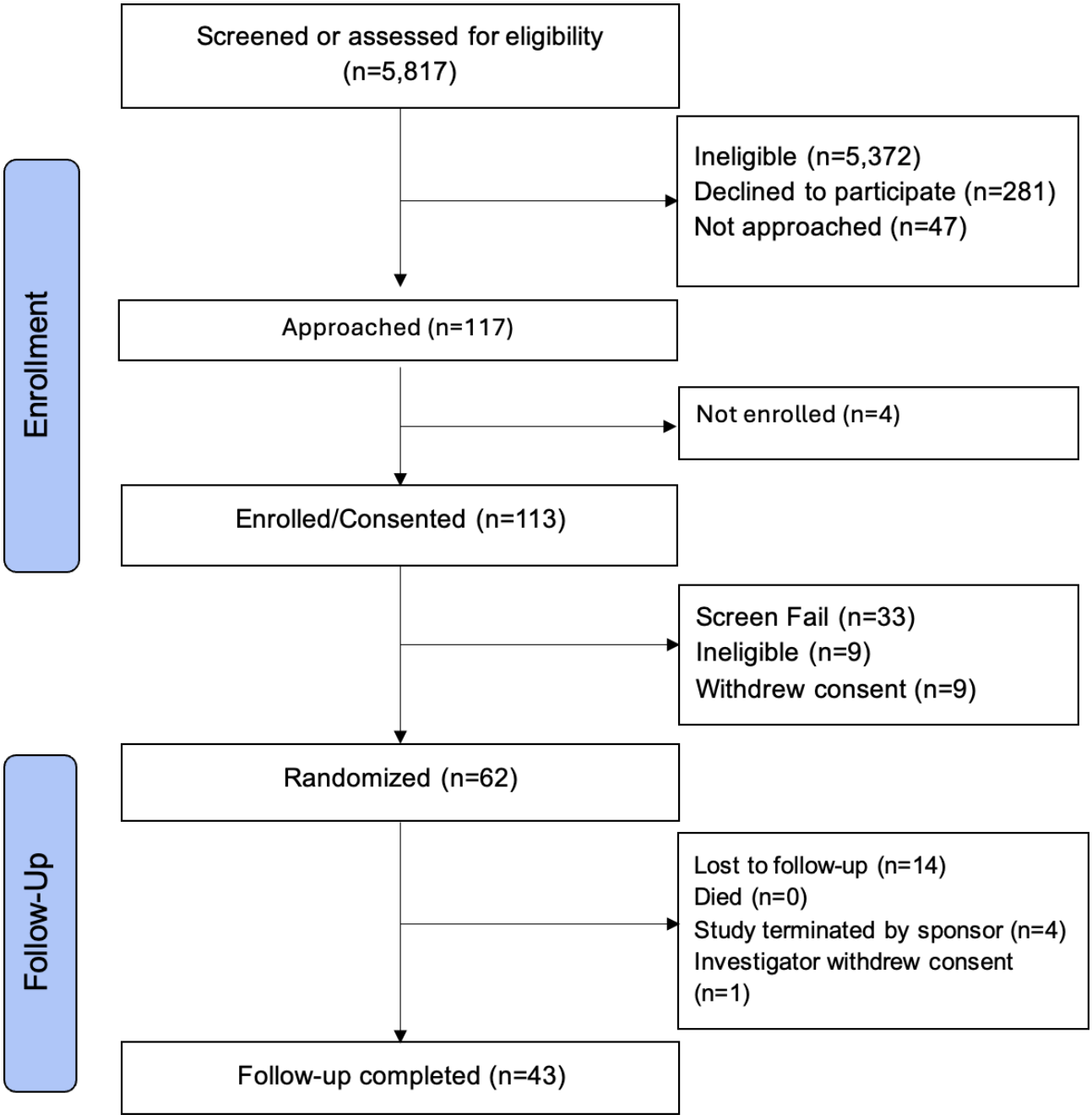
Overview

The clinical trial was funded in April 2020, and enrollment began in June 2021. In February 2023, the Data and Safety Monitoring Board recommended discontinuing the clinical trial due to concerns regarding recruitment and retention. A total of 62 participants were randomized, but 43 completed follow-up. As requested by the funding agency, the National Institute on Aging, 19 participants were contacted in February 2023 to discontinue their participation in the study. This paper reports the trial's results up until that point. Figure 1 shows the CONSORT (Consolidated Standards of Reporting Trials) diagram for this trial.

The mean age of participants was 55 (SD 14.4) years. Among the 62 participants, 32% (n=20) were female and 55% (n=34) identified as Black or African American. Most participants (n=48, 77%) were diagnosed with HF with reduced ejection fraction, and 40% (n=25) expressed class III symptoms. They had an average time from HF diagnosis of 2.9 (SD 4.56) years. Upon enrollment, 73% (n=45) were taking an aldosterone antagonist, and 48% (n=30) were taking an SGLT2 inhibitor. Table 1 demonstrates the baseline demographics overall and for each study group.

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram for the ManageHF trial.



XSL•FO
RenderX

Table . Demographics of participants.

Characteristic or category	Overall (n=62)	Clinical worsening intervention group (n=30)	Clinical worsening control group (n=32)	Dietary sodium intervention group (n=38)	Dietary sodium control group (n=24)
Age (y), mean (SD)	54.8 (14.4)	54.5 (14.6)	55.1 (14.5)	54.4 (15.9)	55.4 (12.1)
Female, n (%)	20 (32)	10 (33)	10 (31)	13 (34)	7 (29)
Hispanic or Latino, n (%)	2 (3)	0 (0)	2 (6)	1 (3)	1 (4)
Race, n (%)					
Black or African American	34 (55)	14 (47)	20 (63)	20 (53)	14 (58)
White	25 (40)	15 (50)	10 (31)	16 (42)	9 (38)
Other or unknown	3 (5)	1 (3)	2 (6)	2 (5)	1 (4)
Ischemic etiology of HF[a], n (%)	14 (23)	6 (20)	8 (25)	10 (26)	4 (17)
HF type, n (%)					
Reduced ejection fraction	48 (77)	22 (73)	26 (81)	28 (74)	20 (83)
Preserved ejection fraction	14 (23)	8 (27)	6 (19)	10 (26)	4 (17)
NYHA[b] class, n (%)					
I	6 (10)	1 (3)	5 (16)	3 (8)	3 (13)
II	17 (27)	7 (23)	10 (31)	12 (32)	5 (21)
III	25 (40)	15 (50)	10 (31)	13 (34)	12 (50)
IV	3 (5)	2 (7)	1 (3)	3 (8)	0 (0)
Missing	11 (18)	5 (17)	6 (19)	7 (18)	4 (17)
Number of hospitalizations due to HF within the prior 12 months, n (%)					
0	20 (32)	11 (37)	9 (28)	15 (39)	5 (21)
1	22 (35)	10 (33)	12 (38)	11 (29)	11 (46)
2+	20 (32)	9 (30)	11 (34)	12 (32)	8 (33)
Previous MI[c], n (%)	15 (24)	7 (23)	8 (25)	7 (18)	8 (33)
History of hypertension, n (%)	47 (76)	21 (70)	26 (81)	30 (79)	17 (71)
Serum creatinine, mean (SD)	1.4 (0.9)	1.3 (0.5)	1.5 (1.2)	1.5 (1.1)	1.3 (0.5)
Heart rate, mean (SD)	81.8 (15)	81.2 (14)	82.4 (15)	84.3 (12)	77.8 (18)
Systolic blood pressure, mean (SD)	119.5 (19)	115.6 (19)	123.2 (18)	118.1 (21)	121.7 (14)
BNP[d]					
n (%)	44 (71)	21 (70)	23 (72)	29 (76)	15 (63)
Mean (SD)	788.1 (846)	824.8 (640)	754.7 (1012)	738.7 (868)	883.8 (822)
NT-proBNP[e]					
n (%)	13 (21)	7 (23)	6 (19)	6 (16)	7 (29)
Mean (SD)	5562.3 (7402)	7006.7 (10040)	3877.2 (2044)	6937 (11092)	4384 (1877)
ACE[f] or ARB[g] or ARNI[h], n (%)	45 (73)	23 (77)	22 (69)	27 (71)	18 (75)
Beta-blocker, n (%)	48 (77)	22 (73)	26 (81)	27 (71)	21 (88)

XSL•FO
RenderX

Characteristic or category	Overall (n=62)	Clinical worsening intervention group (n=30)	Clinical worsening control group (n=32)	Dietary sodium intervention group (n=38)	Dietary sodium control group (n=24)
Aldosterone antagonist, n (%)	35 (56)	14 (47)	21 (66)	19 (50)	16 (67)
SGLT2i[i], n (%)	30 (48)	15 (50)	15 (47)	21 (55)	9 (38)
MLHFQ[j], mean (SD)	67.9 (23.65)	74.7 (20.56)	61.5 (24.85)	70.7 (21.43)	63.4 (26.65)
Sodium screener (mg/d), mean (SD)	3873.1 (1907)	4133.3 (2021)	3629.6 (1793)	3818.5 (2118)	3960.9 (1550)
Sodium FFQ[k] (mg/d), mean (SD)	4064.9 (2642)	4170.9 (2758)	3962.4 (2567)	3998.7 (2760)	4176.1 (2489)

[a]HF: heart failure.

[b]NYHA: New York Heart Association.

[c]MI: myocardial infarction.

[d]BNP: B-type natriuretic peptide.

[e]NT-proBNP: N-Terminal pro-B-type natriuretic peptide.

[f]ACE: angiotensin-converting enzyme.

[g]ARB: angiotensin receptor blocker.

[h]ARNI: angiotensin receptor neprilysin inhibitor.

[i]SGLT2i: sodium-glucose cotransporter-2 inhibitor.

[j]MLHFQ: Minnesota Living with Heart Failure Questionnaire.

[k]FFQ: Food Frequency Questionnaire.

Participant Engagement

Regardless of subgroup, all participants were asked to engage with the ManageHF app survey and use remote device monitoring on a daily basis for the 84-day study. Of the 62 participants enrolled in the clinical trial, 66.1% (41/62) completed at least one app survey. Among the 41 participants who completed at least 1 survey, the median was 22 (IQR 2-47) days. The wearable device was used by 78.5% (47/62) of participants. The median number of days using the device was 74 (IQR 39-84). Blood pressure was measured by 78.5% (47/62) of participants. Participants measured blood pressure on a median of 35 (IQR 15-51) days.

Clinical Worsening Intervention

The hierarchical analysis of the overall primary end point showed a numerically larger number of wins in the clinical worsening intervention compared to the control group (346/960, 36% vs 200/960, 20.8%; win ratio 1.73, 95% CI 0.81 - 4.44; *P*=.18). Both individual outcomes of time to HF readmission and change in MLHFQ demonstrated a larger number of wins in the clinical worsening intervention compared to the control group (HF readmissions: 201/960, 20.9% vs 109/960, 11.4%; MLHFQ: 145/960, 15.1% vs 91/960, 9.5%). Table 2 summarizes the hierarchical analysis of the primary end point in participants randomized to the clinical worsening intervention versus control. There were no deaths in the clinical trial. HF readmission was 13.3% (4/30) in the clinical worsening intervention and 21.8% (7/32) in the control group (hazard ratio 0.56, 95% CI 0.16-1.90; *P*=.35). From baseline to 12 weeks, the MLHFQ improved by a mean of 42 (SD 28) points in the clinical worsening intervention and by a mean of 29 (SD 28) points in the control group (*P*=.21). The mean number of DAOH was 83 (SD 22.45) in the clinical worsening group and 84 (SD 22.65) in the control group (*P*=.7).

Out of the 30 participants in the clinical worsening intervention group, 16 received 1066 push notifications during the study period, classified as yellow or red. Throughout the study, these participants received a median of 24 (IQR 6-70) push notifications.

XSL•FO
RenderX

Table . Hierarchical end point and win ratio for both interventions[ab].

Outcome	Clinical worsening intervention wins, n (%)	Clinical worsening control wins, n (%)	Ties, n (%)	Dietary sodium intervention wins, n (%)	Dietary sodium control wins, n (%)	Ties, n (%)
Time to all-cause death	0 (0)	0 (0)	—[c]	0 (0)	0 (0)	—
Time to heart failure readmission	201 (20.9)	109 (11.4)	—	136 (14.9)	156 (17.1)	—
Change in MLHFQ[d] from baseline to 12 week	145 (15.1)	91 (9.5)	—	147 (16.1)	61 (6.7)	—
Overall (across 3 outcomes)	346 (36)	200 (20.8)	414 (43.1)	283 (31)	217 (23.8)	412 (45.2)

[a]Clinical worsening intervention versus control: win ratio 1.73 (95% CI 0.81–4.44); *P*=.18.

[b]Dietary sodium intervention vs control: win ratio 1.30 (95% CI 0.57–3.26); *P*=.53.

[c]—: not applicable.

[d]MLHFQ: Minnesota Living with Heart Failure Questionnaire.

Dietary Sodium Intervention

The hierarchical analysis of the primary end point, as shown in Table 2, demonstrated a numerically larger number of wins in the dietary sodium intervention compared with the control group (283/912, 31% vs 217/912, 23.8%; win ratio 1.30, 95% CI 0.57 - 3.26; *P*=.53). The change in MLHFQ showed a larger number of wins in the dietary sodium intervention compared to the control group (147/912, 16.1% vs 61/912, 6.7%). From baseline to 12 weeks, the MLHFQ improved by 39 (SD 34) points in the dietary sodium intervention and by 24 (SD 25) points in the control group (*P*=.17). In contrast, the dietary sodium intervention group had a numerically smaller number of wins than the control group in terms of the outcome time to HF readmission (136/912, 14.9% vs 156/912, 17.1%). HF readmission was 18.4% (7/38) in the dietary sodium intervention and 16.7% (4/24) in the control group (hazard ratio 1.10, 95% CI 0.32-3.74; *P*=.88). By FFQ at 12 weeks, sodium intake was 2359 (SD 2770) mg in the intervention and 2142 (SD 1671) mg in the control group (n=32; *P*=.23). By sodium screener at 12 weeks, sodium intake was 2122 (SD 1280) mg in the intervention and 2390 (SD 796) mg in the control group (n=30; *P*=.89). The mean number of DAOH was 86 (SD 18.49) in the dietary sodium intervention and 82 (SD 21.96) in the control group (*P*=.16).

Of the 38 participants in the dietary intervention group, 23 received 786 notifications at grocery stores or restaurants. Throughout the study, these participants received a median of 10 (IQR 2-41) push notifications.

Both Interventions

Across all 4 conditions, 18 participants were randomized to both interventions, 12 to the clinical worsening intervention only, 20 to the dietary sodium intervention only, and 12 to neither intervention. HF readmission was 11.1% (2/18) for those randomized to both interventions, 16.7% (2/12) in the clinical worsening only group, 25% (5/20) in the dietary sodium only group, and 16.7% (2/12) for those in neither intervention. The change in MLHFQ showed a 47 (SD 28) point improvement for those in both interventions, a 31 (SD 29) for those in the clinical worsening only group, a 34 (SD 38) point improvement in the dietary sodium only group, and a 19 (SD 22) point improvement for those assigned to neither intervention. There were no statistically significant differences between these 4 conditions.

Discussion

This study was a novel trial of 2 JITAIs delivered via a mobile app, using push notifications and in-app content, at 2 critical time points in HF self-care: when a patient feels worse and when dietary sodium support is needed when entering grocery stores and restaurants. Although the primary end-point showed a trend in both intervention groups relative to control, it did not reach statistical significance and was substantially underpowered due to early stopping. The early discontinuation of the trial resulted in a smaller-than-planned sample size (62 participants, of whom only 43 completed follow-up). This limited the statistical power to detect significant differences between the intervention and control groups. The reduced number of participants limits the generalizability of the findings and is an important limitation of this study. This trial demonstrates innovative methods that, unfortunately, were implemented during a global pandemic that significantly hindered clinical research.

Several mobile app–based interventions for HF [34-36] and sodium intake [37-39] have been studied. While several of these studies implement self-monitoring to drive self-management in HF, none provide a JITAI using several factors to personalize the intervention for success. The ManageHF app provides self-care push notifications and in-app educational content when patients are experiencing worsening symptoms. The self-care push notifications were personalized to the self-care methods the participant had found effective in the past. A bank of over 50 push notifications paired with detailed in-app educational content was used for this intervention. Push notifications and in-app content did include information about guideline-directed medical therapy, medication adherence, and loop diuretic efficacy. The ManageHF app did not directly provide

XSL•FO
RenderX

information about adding or titrating guideline-directed medical therapy. The lack of medication optimization is one of the most critical interventions in HF to reduce hospitalizations, and digital health interventions in this area are essential [40].

In a recent systematic review of mobile health interventions aimed at reducing sodium intake, only 6 were randomized controlled clinical trials [38]. Of those trials, only 2 were mobile app interventions published in English. One of those apps is SaltSwitch, which originates from New Zealand [37]. SaltSwitch was designed to help users locate food products with low sodium levels at grocery stores. However, the SaltSwitch app does not personalize recommendations based on the user's specific requirements for high-sodium foods. Additionally, it lacks features such as push notifications to remind users to scan items and does not provide contextual food information for restaurants. SaltSwitch has not been studied in HF to date. Two quasi-experimental dietary intervention studies were conducted in patients with HF, and neither involved mobile apps. One was a game-based interactive tool for self-management [41], and another was an SMS text messaging intervention [42]. Neither study assessed clinical outcomes.

The National Institute on Aging funded this trial in April 2020, 1 month after the United States declared a nationwide public health emergency, and the same month, more than 1 million cases of COVID-19 were confirmed worldwide. Multiple funding agencies issued statements to researchers regarding collaboration during this challenging period to facilitate research [43,44]. Due to the COVID-19 pandemic, significant challenges arose in all research areas [45]. Many of our sites saw unprecedented issues with access to their health care facilities, and resources for research were impacted. All study sites experienced research closures, which were site-specific, meaning that at any given time, only some of our research sites were recruiting participants. There were also participant-level challenges in recruiting patients with HF. During the COVID-19 pandemic, HF hospitalizations declined significantly [45-47], resulting in [46-48] a smaller patient pool to recruit from and a lower-than-expected HF readmission rate for our clinical trial. Potential participants may have been less likely to enroll due to societal or individual pandemic-related circumstances. Additionally, participants who did enroll were less likely to visit restaurants and grocery stores during the pandemic, which is a limitation of this trial.

Demographic diversity is a crucial discussion point in clinical trials [49], particularly in those discontinued by the Data and Safety Monitoring Board due to low recruitment. Black participants represented 55% (34/62) of the enrolled population in this study, compared to less than 10% in other recent randomized trials enrolling patients with HF [50-53]. Recruiting and enrolling diverse patients in clinical trials is vital, but it requires additional staff resources and time. We have shown that enrolling diverse participants from a federally qualified health center required 1.4 times the full-time equivalent staff compared to a less diverse university hospital setting [54]. This is even more important in mobile health–related clinical trials, where interventions can exacerbate inequities [55].

An important aspect of this clinical trial is the inclusion of an active control. All participants received a blood pressure cuff, a scale, and a wearable activity monitor and were asked to self-monitor using the mobile app, which included a daily survey. This was done to fulfill the third aim of the research grant, which was to determine if passive preclinical measurements from these devices could predict symptoms of worsening HF. Using an active control can reduce the chance of detecting a statistically significant outcome. An active control can also add rigor to the trial because it biases the results toward showing no difference between the groups. This should be accounted for in the clinical trial statistical power assessment and is a limitation of this study.

Despite pandemic-related recruitment challenges, the ManageHF study successfully engaged a diverse cohort of high-risk patients with HF in self-management activities. Future studies using more effective recruitment and retention strategies are crucial for establishing the efficacy of these interventions with greater certainty.

Acknowledgments

Microsoft Word and Grammarly were used to edit and correct the text to improve the manuscript's clarity.

Funding

Funding for this trial was provided by the National Institute on Aging (R01AG062582). The funder had no involvement in the study design, data collection, analysis, interpretation, or manuscript writing.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: MPD, TMK, SLH
Methodology: MPD, TMK, SLH
Formal analysis: AK, GK
Project administration: SG

XSL•FO
RenderX

Software: JA, XZ
Investigation: AJS, JAC, GE, JER, DLD
Supervision: MPD
Writing – original draft: MPD
Writing – review & editing: MSA, SG, KMG, AJS, JAC, GE, JER, DLD, TMK, SLH

Conflicts of Interest

None declared.

Multimedia Appendix 1
Protocol changes or updates.
[PDF File, 51 KB - mhealth_v14i1e74121_app1.pdf]

Checklist 1
CONSORT-EHEALTH checklist.
[PDF File, 1157 KB - mhealth_v14i1e74121_app2.pdf]

References

1. Martin SS, Aday AW, Allen NB, et al. 2025 Heart disease and stroke statistics: a report of US and global data from the American Heart Association. Circulation 2025 Feb 25;151(8):e41-e660. [doi: 10.1161/CIR.0000000000001303] [Medline: 39866113]
2. Retrum JH, Boggs J, Hersh A, et al. Patient-identified factors related to heart failure readmissions. Circ Cardiovasc Qual Outcomes 2013 Mar 1;6(2):171-177. [doi: 10.1161/CIRCOUTCOMES.112.967356] [Medline: 23386663]
3. Gilotra NA, Shpigel A, Okwuosa IS, Tamrat R, Flowers D, Russell SD. Patients commonly believe their heart failure hospitalizations are preventable and identify worsening heart failure, nonadherence, and a knowledge gap as reasons for admission. J Card Fail 2017 Mar;23(3):252-256. [doi: 10.1016/j.cardfail.2016.09.024] [Medline: 27742454]
4. Annema C, Luttik ML, Jaarsma T. Reasons for readmission in heart failure: perspectives of patients, caregivers, cardiologists, and heart failure nurses. Heart Lung 2009;38(5):427-434. [doi: 10.1016/j.hrtlng.2008.12.002] [Medline: 19755193]
5. Michalsen A, König G, Thimme W. Preventable causative factors leading to hospital admission with decompensated heart failure. Heart 1998 Nov;80(5):437-441. [doi: 10.1136/hrt.80.5.437] [Medline: 9930040]
6. Dharmarajan K, Hsieh AF, Lin Z, et al. Diagnoses and timing of 30-day readmissions after hospitalization for heart failure, acute myocardial infarction, or pneumonia. JAMA 2013 Jan 23;309(4):355-363. [doi: 10.1001/jama.2012.216476] [Medline: 23340637]
7. Dharmarajan K, Wang Y, Lin Z, et al. Association of changing hospital readmission rates with mortality rates after hospital discharge. JAMA 2017 Jul 18;318(3):270-278. [doi: 10.1001/jama.2017.8444] [Medline: 28719692]
8. Van Spall HGC, DeFilippis EM, Lee SF, et al. Sex-specific clinical outcomes of the PACT-HF randomized trial. Circ Heart Fail 2021 Nov;14(11):e008548. [doi: 10.1161/CIRCHEARTFAILURE.121.008548] [Medline: 34711072]
9. Kommuri NVA, Koelling TM, Hummel SL. The impact of prior heart failure hospitalizations on long-term mortality differs by baseline risk of death. Am J Med 2012 Feb;125(2):209.e15. [doi: 10.1016/j.amjmed.2011.06.014] [Medline: 22269626]
10. Bradley EH, Curry L, Horwitz LI, et al. Contemporary evidence about hospital strategies for reducing 30-day readmissions: a national study. J Am Coll Cardiol 2012 Aug 14;60(7):607-614. [doi: 10.1016/j.jacc.2012.03.067] [Medline: 22818070]
11. Kociol RD, Peterson ED, Hammill BG, et al. National survey of hospital strategies to reduce heart failure readmissions: findings from the Get With the Guidelines-Heart Failure registry. Circ Heart Fail 2012 Nov;5(6):680-687. [doi: 10.1161/CIRCHEARTFAILURE.112.967406] [Medline: 22933525]
12. Baker H, Oliver-McNeil S, Deng L, Hummel SL. Regional hospital collaboration and outcomes in Medicare heart failure patients: see you in 7. JACC Heart Fail 2015 Oct;3(10):765-773. [doi: 10.1016/j.jchf.2015.06.007] [Medline: 26364256]
13. Ibrahim AM, Dimick JB, Sinha SS, Hollingsworth JM, Nuliyalu U, Ryan AM. Association of coded severity with readmission reduction after the hospital readmissions reduction program. JAMA Intern Med 2018 Feb 1;178(2):290-292. [doi: 10.1001/jamainternmed.2017.6148] [Medline: 29131896]
14. Lee DS, Austin PC, Stukel TA, et al. "Dose-dependent" impact of recurrent cardiac events on mortality in patients with heart failure. Am J Med 2009 Feb;122(2):162-169.e1. [doi: 10.1016/j.amjmed.2008.08.026] [Medline: 19100960]
15. Abdul-Aziz AA, Hayward RA, Aaronson KD, Hummel SL. Association between Medicare hospital readmission penalties and 30-day combined excess readmission and mortality. JAMA Cardiol 2017 Feb 1;2(2):200-203. [doi: 10.1001/jamacardio.2016.3704] [Medline: 27784054]
16. Shively M, Kodiath M, Smith TL, et al. Effect of behavioral management on quality of life in mild heart failure: a randomized controlled trial. Patient Educ Couns 2005 Jul;58(1):27-34. [doi: 10.1016/j.pec.2004.06.007] [Medline: 15950833]

XSL•FO
RenderX

17. Varma S, McElnay JC, Hughes CM, Passmore AP, Varma M. Pharmaceutical care of patients with congestive heart failure: interventions and outcomes. Pharmacotherapy 1999 Jul;19(7):860-869. [doi: 10.1592/phco.19.10.860.31565] [Medline: 10417035]
18. Dorsch MP, Farris KB, Bleske BE, Koelling TM. A web application for self-monitoring improves symptoms in chronic systolic heart failure. Telemed J E Health 2015 Apr;21(4):267-270. [doi: 10.1089/tmj.2014.0095] [Medline: 25654615]
19. WRITING COMMITTEE MEMBERS, Yancy CW, Jessup M, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines. Circulation 2013 Oct 15;128(16):e240-e327. [doi: 10.1161/CIR.0b013e31829e8776] [Medline: 23741058]
20. Heart Failure Society of America, Lindenfeld J, Albert NM, et al. HFSA 2010 comprehensive heart failure practice guideline. J Card Fail 2010 Jun;16(6):e1-194. [doi: 10.1016/j.cardfail.2010.04.004] [Medline: 20610207]
21. Heidenreich PA, Bozkurt B, Aguilar D, et al. 2022 AHA/ACC/HFSA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Joint Committee on clinical practice guidelines. Circulation 2022 May 3;145(18):e895-e1032. [doi: 10.1161/CIR.0000000000001063] [Medline: 35363499]
22. Ezekowitz JA, Colin-Ramirez E, Ross H, et al. Reduction of dietary SODIUM to less than 100 mmol in heart failure (SODIUM-HF): an international, open-label, randomised, controlled trial. Lancet 2022 Apr 9;399(10333):1391-1400. [doi: 10.1016/S0140-6736(22)00369-5] [Medline: 35381194]
23. Doukky R, Avery E, Mangla A, et al. Impact of dietary sodium restriction on heart failure outcomes. JACC Heart Fail 2016 Jan;4(1):24-35. [doi: 10.1016/j.jchf.2015.08.007] [Medline: 26738949]
24. Hummel SL, Karmally W, Gillespie BW, et al. Home-delivered meals postdischarge from heart failure hospitalization. Circ Heart Fail 2018 Aug;11(8):e004886. [doi: 10.1161/CIRCHEARTFAILURE.117.004886] [Medline: 30354562]
25. Dorsch MP, Farris KB, Hummel SL, Koelling TM. A patient-centered mobile intervention to promote self-management and improve patient outcomes in chronic heart failure: the ManageHF trial. J Card Fail 2019 Aug;25(8):S104. [doi: 10.1016/j.cardfail.2019.07.296]
26. Dorsch MP, Cornellier ML, Poggi AD, et al. Effects of a novel contextual just-in-time mobile app intervention (LowSalt4Life) on sodium intake in adults with hypertension: pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Aug 10;8(8):e16696. [doi: 10.2196/16696] [Medline: 32663139]
27. Dorsch MP, Farris KB, Rowell BE, Hummel SL, Koelling TM. The effects of the ManageHF4Life mobile app on patients with chronic heart failure: randomized controlled trial. JMIR Mhealth Uhealth 2021 Dec 7;9(12):e26185. [doi: 10.2196/26185] [Medline: 34878990]
28. Riegel B, Carlson B, Moser DK, Sebern M, Hicks FD, Roland V. Psychometric testing of the self-care of heart failure index. J Card Fail 2004 Aug;10(4):350-360. [doi: 10.1016/j.cardfail.2003.12.001] [Medline: 15309704]
29. Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. Eur Heart J 2012 Jan;33(2):176-182. [doi: 10.1093/eurheartj/ehr352] [Medline: 21900289]
30. Hicks KA, Mahaffey KW, Mehran R, et al. 2017 Cardiovascular and stroke endpoint definitions for clinical trials. Circulation 2018 Feb 27;137(9):961-972. [doi: 10.1161/CIRCULATIONAHA.117.033502] [Medline: 29483172]
31. Heo S, Moser DK, Lennie TA, Payne-Emerson H, Welch JL, Weaver M. Development and testing of the feasibility and acceptability of a tailored dietary intervention in patients with heart failure. J Cardiovasc Nurs 2015;30(3):213-221. [doi: 10.1097/JCN.0000000000000148] [Medline: 24695076]
32. Finkelstein DM, Schoenfeld DA. Combining mortality and longitudinal measures in clinical trials. Stat Med 1999 Jun 15;18(11):1341-1354. [doi: 10.1002/(sici)1097-0258(19990615)18:11<1341::aid-sim129>3.0.co;2-7] [Medline: 10399200]
33. Kotalik A, Eaton A, Lian Q, Serrano C, Connett J, Neaton JD. A win ratio approach to the re-analysis of Multiple Risk Factor Intervention Trial. Clin Trials 2019 Dec;16(6):626-634. [doi: 10.1177/1740774519868233] [Medline: 31389723]
34. Athilingam P, Jenkins B, Johansson M, Labrador M. A mobile health intervention to improve self-care in patients with heart failure: pilot randomized control trial. JMIR Cardio 2017 Aug 11;1(2):e3. [doi: 10.2196/cardio.7848] [Medline: 31758759]
35. Masterson Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74. [doi: 10.2196/mhealth.5882] [Medline: 27302310]
36. Athilingam P, Jenkins B. Mobile phone apps to support heart failure self-care management: integrative review. JMIR Cardio 2018 May 2;2(1):e10057. [doi: 10.2196/10057] [Medline: 31758762]
37. Eyles H, McLean R, Neal B, et al. A salt-reduction smartphone app supports lower-salt food purchases for people with cardiovascular disease: findings from the SaltSwitch randomised controlled trial. Eur J Prev Cardiol 2017 Sep;24(13):1435-1444. [doi: 10.1177/2047487317715713] [Medline: 28631933]
38. Ali SH, Luo R, Li Y, Liu X, Tang C, Zhang P. Application of mobile health technologies aimed at salt reduction: systematic review. JMIR Mhealth Uhealth 2019 Apr 17;7(4):e13250. [doi: 10.2196/13250] [Medline: 30994467]
39. Ipjian ML, Johnston CS. Smartphone technology facilitates dietary change in healthy adults. Nutrition 2017 Jan;33:343-347. [doi: 10.1016/j.nut.2016.08.003] [Medline: 27742102]

XSL•FO
RenderX

40. Azizi Z, Golbus JR, Spaulding EM, et al. Challenge of optimizing medical therapy in heart failure: unlocking the potential of digital health and patient engagement. J Am Heart Assoc 2024 Jan 16;13(2):e030952. [doi: 10.1161/JAHA.123.030952] [Medline: 38226520]
41. Radhakrishnan K, Toprac P, O'Hair M, et al. Interactive digital e-Health game for heart failure self-management: a feasibility study. Games Health J 2016 Dec;5(6):366-374. [doi: 10.1089/g4h.2016.0038] [Medline: 27976955]
42. Nundy S, Razi RR, Dick JJ, et al. A text messaging intervention to improve heart failure self-management after hospital discharge in a largely African-American population: before-after study. J Med Internet Res 2013 Mar 11;15(3):e53. [doi: 10.2196/jmir.2317] [Medline: 23478028]
43. Applicant and awardee faqs related to COVID-19 or other unexpected public health emergencies. Patient-Centered Outcomes Research Institute. 2020. URL: https://tinyurl.com/yw7wxbrh [accessed 2026-01-07]
44. NIH implementation of OMB memorandum m-20-26 "extension of administrative relief for recipients and applicants of federal financial assistance directly impacted by the novel coronavirus (COVID-19) due to loss of operations" dated june 18, 2020. National Institutes of Health. URL: https://grants.nih.gov/policy-and-compliance/policy-topics/natural-disasters/corona-virus/nih-omb-memo [accessed 2026-01-07]
45. Nomali M, Mehrdad N, Heidari ME, et al. Challenges and solutions in clinical research during the COVID-19 pandemic: a narrative review. Health Sci Rep 2023 Aug;6(8):e1482. [doi: 10.1002/hsr2.1482] [Medline: 37554954]
46. Hall ME, Vaduganathan M, Khan MS, et al. Reductions in heart failure hospitalizations during the COVID-19 pandemic. J Card Fail 2020 Jun;26(6):462-463. [doi: 10.1016/j.cardfail.2020.05.005] [Medline: 32405232]
47. Cannatà A, Bromage DI, Rind IA, et al. Temporal trends in decompensated heart failure and outcomes during COVID-19: a multisite report from heart failure referral centres in London. Eur J Heart Fail 2020 Dec;22(12):2219-2224. [doi: 10.1002/ejhf.1986] [Medline: 32809274]
48. Cowie MR, Mourilhe-Rocha R, Chang HY, et al. The impact of the COVID-19 pandemic on heart failure management: global experience of the OPTIMIZE Heart Failure Care network. Int J Cardiol 2022 Sep 15;363:240-246. [doi: 10.1016/j.ijcard.2022.06.022] [Medline: 35750302]
49. Schwartz AL, Alsan M, Morris AA, Halpern SD. Why diverse clinical trial participation matters. N Engl J Med 2023 Apr 6;388(14):1252-1254. [doi: 10.1056/NEJMp2215609] [Medline: 37017480]
50. Allen LA, Venechuk G, McIlvennan CK, et al. An electronically delivered patient-activation tool for intensification of medications for chronic heart failure with reduced ejection fraction: the EPIC-HF trial. Circulation 2021 Feb 2;143(5):427-437. [doi: 10.1161/CIRCULATIONAHA.120.051863] [Medline: 33201741]
51. McMurray JJV, Solomon SD, Inzucchi SE, et al. Dapagliflozin in patients with heart failure and reduced ejection fraction. N Engl J Med 2019 Nov 21;381(21):1995-2008. [doi: 10.1056/NEJMoa1911303] [Medline: 31535829]
52. Anker SD, Butler J, Filippatos G, et al. Empagliflozin in heart failure with a preserved ejection fraction. N Engl J Med 2021 Oct 14;385(16):1451-1461. [doi: 10.1056/NEJMoa2107038] [Medline: 34449189]
53. Packer M, Anker SD, Butler J, et al. Cardiovascular and renal outcomes with empagliflozin in heart failure. N Engl J Med 2020 Oct 8;383(15):1413-1424. [doi: 10.1056/NEJMoa2022190] [Medline: 32865377]
54. Skolarus LE, Farhan Z, Mishra SR, et al. Resource requirements for participant enrollment from a university health system and a federally qualified health center network in a mobile health study: the myBPmyLife trial. J Am Heart Assoc 2024 Jan 16;13(2):e030825. [doi: 10.1161/JAHA.123.030825] [Medline: 38226521]
55. Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc 2018 Aug 1;25(8):1080-1088. [doi: 10.1093/jamia/ocy052] [Medline: 29788380]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
DAOH: days alive and out of hospital
FFQ: Food Frequency Questionnaire
HF: heart failure
HRQOL: health-related quality of life
JITAI: Just-in-time Adaptive Intervention
MLHFQ: Minnesota Living with Heart Failure Questionnaire
SAS: Statistical Analysis System
UPC: Universal Product Code

XSL•FO
RenderX

Edited by A Stone; submitted 18.Mar.2025; peer-reviewed by J Wang, Y Li; accepted 19.Dec.2025; published 28.Jul.2026.

Please cite as:
Dorsch MP, Ali MS, Krambrink A, Kolenic G, Ganai S, Arzac J, Zhang X, Greer KM, Shah AJ, Cowger JA, Ewald G, Rodgers JE, Dixon DL, Koelling TM, Hummel SL
The ManageHF Just-in-Time Adaptive Mobile App Interventions to Promote Self-Management and Improve Outcomes in Heart Failure: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e74121
URL: https://mhealth.jmir.org/2026/1/e74121
doi:10.2196/74121

© Michael P Dorsch, Mohamed S Ali, Amy Krambrink, Giselle Kolenic, Sabah Ganai, Juan Arzac, Xutong Zhang, Kaitlyn M Greer, Amit J Shah, Jennifer A Cowger, Gregory Ewald, Jo Ellen Rodgers, Dave L Dixon, Todd M Koelling, Scott L Hummel. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 28.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Adoption of App-Based Peer Support by Canadian Health Care Providers: Mixed Methods Organizational Implementation Study

Sandra Moll[1], MSc, PhD; Melissa Parker[1], MSc; Peter Smith[2], PhD; Edward Sykes[3], PhD

[1]Department of Occupational Therapy, Faculty of Health Sciences, McMaster University, 1400 Main St. West, Hamilton, ON, Canada

[2]Institute for Work and Health, Toronto, ON, Canada

[3]School of Computer Science, College of Computational, Mathematical and Physical Sciences, University of Guelph, Guelph, ON, Canada

Corresponding Author:
Sandra Moll, MSc, PhD
Department of Occupational Therapy, Faculty of Health Sciences, McMaster University, 1400 Main St. West, Hamilton, ON, Canada

Abstract

Background: There is an urgent and critical need to support the mental health of health care providers, given high rates of stress and burnout. Although the issues are complex, digital access to information and support can help address the needs, as technology can facilitate on-demand links to private, customized resources, including peer support. Beyond Silence (McMaster University) is an evidence-informed mobile health platform co-designed with health care workers and grounded in prior evidence that mental health literacy and peer support can reduce stigma and facilitate earlier help-seeking.

Objective: This study aimed to (1) explore how health care workers across diverse health care settings use the app and (2) identify opportunities and barriers to implementation.

Methods: A multiple-case study framework, informed by the Consolidated Framework for Implementation Research (CFIR), was applied to capture 4 months of implementation across a purposive sample of 7 diverse Canadian health care organizations. Implementation within each organization was led by designated organizational champions who leveraged existing communication channels and standardized promotional materials to invite employees to voluntarily download and use the app. Implementation outcomes were assessed using app analytics (downloads and feature use) and semistructured baseline and follow-up interviews with organizational champions to explore contextual influences on uptake.

Results: Approximately 1066 employees downloaded the app over the 4-month period, ranging from <2% to >45% of employees across the 7 organizations. Interviews with 28 organizational champions noted that there was good leadership support for the technology, aligning with their mission to address employee mental health. Barriers to use, however, included workplace culture surrounding mental health and help-seeking, lack of awareness about when and how to use the app, and infrastructure-related challenges, such as limited time and a lack of private spaces to download and use the technology.

Conclusions: Effective implementation is a precondition for positive outcomes; therefore, strategies are needed to optimize technology implementation. Recommendations include evaluating organizational readiness, building mental health literacy, creating a multimodal communication and implementation plan, addressing technology requirements, and embedding the technology into organizational policies and practices. This study highlights key challenges in the implementation of the Beyond Silence peer support platform for health care workers, including slow adoption linked to mental health stigma, competing demands, and limited frontline engagement. Addressing these barriers will require innovative, trust-building strategies to support meaningful uptake and sustained use.

Trial Registration: ClinicalTrials.gov NCT05514093; https://clinicaltrials.gov/study/NCT05514093

(JMIR Mhealth Uhealth 2026;14:e85658) doi:10.2196/85658

KEYWORDS

implementation science; mental health; technology; workplace; mhealth; mobile health

Introduction

The mental health of health care providers demands immediate and sustained support. Health care workers experience elevated rates of stress and burnout, which have been exacerbated by the COVID-19 pandemic [1]. Sustained psychological burden and high workloads have led to a high rate of workforce attrition, leading to a human resource crisis that may take years to resolve [2]. Issues facing health care workers are complex and can range from high workloads to exposure to workplace violence and/or potentially traumatic situations [1]. For example, many clinicians feel isolated and unsupported by their workplace [3,4].

XSL•FO
RenderX

A growing body of literature underlines the value of peer support for health care workers in mitigating the impact of stress and posttraumatic stress disorders (PTSD) [5,6]. Peers understand the unique context of health care work and can therefore be a trusted source of support. Access to peer support, however, can be challenging in some workplace contexts, particularly if there is a negative workplace culture or resource and time pressures that make it difficult to reach out. E–mental health interventions are gaining recognition as a strategy to enable mental health support across different contexts, including the workplace [7]. Mobile technology can provide accessible, on-demand information and links to support and has the potential to reach a wide range of users, including those in underresourced settings [7,8]. Access to peer support via mobile technology can also be a cost-effective first-line approach and conduit for professional support within a stepped model of care [9].

Evidence for the effectiveness of mobile health (mHealth) apps has been noted through systematic reviews and meta-analyses, including evidence for reducing symptoms related to common mental health disorders such as anxiety and depression both in the general population and clinical populations [8,10]. In addition to studies in the general population, there are a growing number of studies examining the impact of digital mental health interventions in workplace settings [11]. One systematic review and meta-analysis of mHealth apps implemented in workplace settings found mixed evidence for improving mental health symptoms with poor sustainability of the effects over time [12]. An umbrella review was published in 2025 of 14 systematic literature reviews of the effectiveness of digital mental health interventions in the workplace (9 systematic reviews and 5 meta-analyses) [13]. The review noted growing evidence of the small to moderate impact of digital tools on reducing depression, anxiety, and perceived stress.

Although these studies highlight the potential of mHealth apps, outcomes remain variable across and within studies, and gaps persist in our understanding of implementation over time and across contexts [13]. A review of eHealth interventions for health care professionals, for example, highlighted inconsistencies in outcomes and limited implementation data [14]. Implementation science frameworks such as the Nonadoption, Abandonment, Scale-up, Spread, and Sustainability (NASSS) model of health and care technology highlight the challenges associated with consistent and sustained engagement with mHealth technology, grounded in an appreciation of the various complexities of implementation [15]. The Consolidated Framework for Implementation Research (CFIR) also notes the importance of understanding different domains of implementation, considering how the context and process of implementation shape adoption [16]. There remains a need not only to assess the impact of technologies but also to understand the factors influencing uptake and implementation, and how these evolve over time. It is not enough to understand whether an app is designed well; it is important to systematically explore how it is adopted in a real-world context [7,8,12]. This study aimed to (1) examine the adoption of Beyond Silence (McMaster University), an mHealth peer support platform designed by and for health care employees across 7 diverse Canadian health care organizations, and (2) identify and describe perceived organizational facilitators and barriers to its implementation both within and across these settings.

Methods

Study Design

A mixed methods implementation case study design [17] was adopted to explore implementation and uptake of the Beyond Silence platform across 7 health care organizations. A case study approach provides an effective way to study contemporary phenomena within a real-life context, particularly when it is difficult to separate the case from the context and can lead to new explanations that consider the complexities of place and time [18]. Exploring multiple case study sites enabled cross-case comparison across different workplace contexts. A multistage mixed methods framework was adopted, starting with qualitative interviews to explore the perceptions of organizational champions regarding the context for app implementation [19]. This was followed by the collection of quantitative app usage data. Finally, follow-up interviews were conducted with the organizational champions, during which they were asked about the individual and contextual forces that shaped trends in app usage within the workplace context [20]. Data collection and analysis were informed by CFIR, one of the most commonly used frameworks to assess contextual factors affecting real-world implementation [21].

Ethical Considerations

Ethics approval was obtained through the Hamilton Integrated Research Ethics Board (HiREB number 14740), and all participants provided written consent to participate. In order to maintain privacy and confidentiality, the names of participating organizations and organizational champions are disguised. All app usage data are reported in aggregate; app users were anonymous. Organizational champions received a CAD $100 (US $70.95 as of July 24, 2026) gift card honorarium for participating in each champion interview. Peer supporters in each organization were not paid for their role; however, external peer support providers were provided a stipend of $55 per 3-hour shift. Peer support providers who participated in the focus group discussions were invited to enter a draw for one of 2 CAD $50 (US $36.50) gift cards of their choice.

An advisory team provided input throughout the project, including 4 academic leaders in workplace mental health, mHealth technology, and trauma-informed care, and 6 knowledge user partners representing national associations of nurses, personal support workers (PSWs), social workers, and long-term care workers, a national mental health service provider, and a provincial health and safety association. The team met 3 - 4 times a year to provide input at key stages in the project and to facilitate communication within their extensive networks.

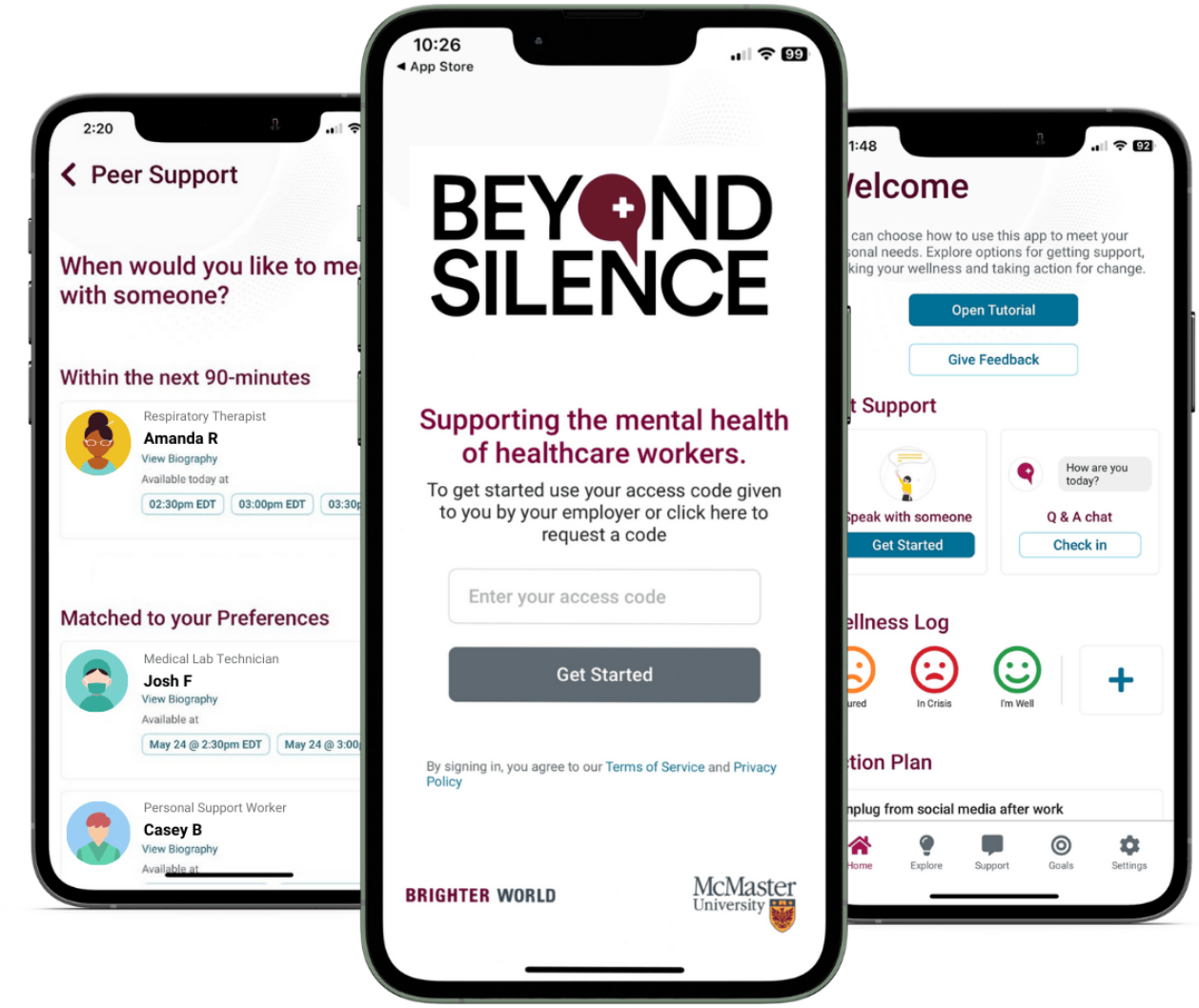
Intervention

Beyond Silence (see Figure 1) is a digital platform (including a mobile app for frontline health care workers and an organizational portal for peer supporters and organizational champions) designed to provide accessible, anonymous, and peer-led mental health support to Canadian health care workers.

XSL•FO
RenderX

It builds on a decade of successful research collaborations with health care stakeholders to identify and address their unique mental health and support needs [19,20]. Development was informed by findings from a randomized controlled trial of an in-person workplace mental health training program with over 200 health care workers [21]. The in-person training was effective but resource-intensive and not accessible to all workers; therefore, an mHealth platform was explored as an alternative approach.

Figure 1. Screenshots of the Beyond Silence mobile app for frontline workers.



Development of the Beyond Silence platform (2017 - 2019) was supported by a Canadian Institute of Health Research (CIHR) development grant, and informed by the Knowledge-to-Action Framework, a structured approach for knowledge creation and a 7-phase cycle that moves knowledge into practice [22]. In addition, best practice principles of mHealth design were considered, including engagement (customization and interactivity), functionality (ease of use), aesthetics, information quality, and data security [23]. Additional funding (2022 - 2024) from the Public Health Agency of Canada (PHAC) allowed for additional app enhancements. Core functionalities of the app for frontline workers include (1) access to evidence-informed information customized for health care workers (articles, infographics, video, and audio clips) on a range of topics from mindfulness to moral injury; (2) links to self-management tools, including a wellness toolbox, wellness check-ins, and goal-setting feature; and (3) private, secure links via text or phone to a range of health care providers trained to offer peer support. App users can browse the biographies of the available peer supporters within and outside their organization, choose how they want to connect (phone or text), click on who they want to connect, and the time at which they want to connect (either within the next 90 minutes or at a scheduled time that their preferred peer is available). A notification is then sent to the identified peer regarding the anonymous request for support and a secure, private connection is made through the app for the health care worker seeking support.

In addition to the app for frontline health care workers, the platform includes an interconnected web-based portal for trained peer support providers to respond to requests for support. The online portal provides space for peer supporters to upload their pictures and short biographies (visible to the app users), as well as sign up for 3-hour shifts, respond to peer support requests (by text or phone), record digital notes on their encounters, access resources to support their work (eg, crisis protocols and contact details for relevant resources), and connect with a clinical supervisor and/or other peer supporters. Although the

XSL•FO
RenderX

peer support providers were visible to frontline app users, the person accessing support was anonymous (digital ID number). A total of 156 health care professionals were trained to offer peer support in the app over the course of the study, including 88 from participating organizations and 68 central peer supporters who were external to the participating organizations and accessible to all. Details of the peer support training are reported in a separate paper [24].

The Beyond Silence platform also included an administrative portal where organizational champions could upload local resources, manage their internal peer support provider list, send notifications, and track overall patterns of app use. It is important to note that the app was available through app stores, but an organizational code was required to download the app, since the app is customized for each organization (including local resources and peer support access). As noted above, there was an onboarding process for the participating organizations and peer supporters, which included training on risk management and data security protocols.

Organization Recruitment

Health care organizations were recruited through established networks of the project advisory team. Interested organizations were asked to complete an application form to assist in screening for diversity across size, geographic location, workplace setting, and type of health care service. To be eligible to participate, organizations required: (1) a project agreement letter signed by organizational leaders, (2) at least one organizational champion designated to liaise with the research team, and (3) at least 2 - 3 internal staff members who agreed to be trained peer support providers (numbers varied based on size and resources available within the organization).

The goal was to recruit 6 - 8 organizations to explore how implementation varied across workplace contexts. Purposive sampling was conducted to ensure diversity in size (small/medium/large), geographical location, setting (rural/remote and urban), type of service (long-term care, acute care, tertiary care, and primary care), and at least 1 organization where employees spoke French (since the app is available in both French and English). A total of 22 organizations expressed interest in the study; however, 14 of these organizations did not have the capacity to participate. For organizations that declined to participate, the primary reason for declining was documented (typically due to limited resources to support the app implementation and evaluation).

Implementation setting and scope A champion-led, organization-embedded implementation strategy was used to introduce Beyond Silence within each participating organization. Champions were individuals representing participating organizations who approached us to be involved in the study. They were responsible for liaising with the research team, communicating about the app to employees, and acting as the contact point for any questions or concerns among staff. Each site designated one or more organizational champions to serve as local implementation leads and primary points of contact.

Champions represented a range of organizational roles, including senior leaders, health and safety coordinators, and frontline clinicians, including 20 women and 6 men in both clinical (n=13) and nonclinical (n=13) positions. As noted in Table 1, there were typically 1 - 3 champions in each organization who agreed to communicate with employees about the app and to be interviewed about their insights at the outset and conclusion of the project. One large multisite organization had 11 champions volunteer to participate in the study.

XSL•FO
RenderX

Table . Participating health care organizations.

Organization	Employees, n	Dates	Champions, n	Trained providers, n
1. Urban community care (nurses)	600	November 2022-March 2023	1	3
2. National professional association	13,000		4	12
2A. Department in urban hospital		November 2022-March 2023		
2B. Nonclinical workers		November 2022-March 2023		
2C. National association		August 2023-December 2023		
3. Northern Ontario Health Centre	769	November 2022-March 2023	2	9
4. Urban community care (PSWs[a])	866	November 2023-March 2024	2	6
5. Rural community care	55	November 2023-March 2024	2	3
6. Large urban hospital	12,463	November 2023-March 2024	12	19
7. National health service	5000	February 2024-June 2024	4	36
Total (all organizations)	32,753	November 2022-June 2024	27	88

[a]PSW: personal support worker.

The target population for implementation was determined by each organization based on local capacity and readiness. Most included all employees, but 3 organizations chose a specific subgroup (Organization 1–nurses, Organization 2A–laboratory technicians, and Organization 4–PSWs). Recruitment relied on existing organizational communication channels and included distribution of standardized promotional materials (posters, email scripts, brochures, and branded materials) containing information about the app, a QR code for download, and a unique organizational access code. Employees were invited to voluntarily download and use the app for the implementation period.

Implementation included organizational onboarding, app deployment to frontline staff, and integration of trained peer support providers accessible through the platform. Uptake, engagement, and variability were tracked across organizational contexts over a 4-month period. Guided by an implementation framework (CFIR), contextual factors, resource requirements, and implementation processes were examined to inform scalability and sustainability.

Data Collection and Outcomes

Data collection focused on prespecified implementation outcomes and used a mixed methods approach combining app usage analytics and qualitative interviews with organizational champions. Evaluation focused on reach, adoption, implementation, and patterns of use rather than effectiveness or health impact.

Quantitative implementation metrics were assessed using real-time, aggregate app analytics captured through the administrative portal. App users remained anonymous and were identified only through a unique user ID. Metrics were selected to characterize adoption and engagement and included the number and proportion of employees who downloaded the app, the number of times the app was opened, mean number of app opens per user, and frequency of use of key features, including content views, goals set, wellness check-ins, and peer support appointments booked. Additional descriptive data captured peer support outreach modality (text or call) and choice of peer supporter. App use data were examined monthly within each organization to describe trends over the 4-month implementation period and to support cross-case comparison; inferential analyses were not conducted.

An update to the app occurred during which onboarding questions were added, asking app users to respond to optional demographic questions regarding age, gender, years of professional experience, and health care role (clinical/nonclinical). Given the timing of this change, only 5 of the 7 organizations included this demographic data. As such, demographic data were not collected or reported for organizations 1, 2A, 2B, and 3.

The study is registered on ClinicalTrials.gov (NCT05514093) as an implementation and evaluation study of a mHealth intervention designed to support the mental health of health care workers in Canada. However, while multiple mental health outcome measures (including standardized pre-post surveys of depression, anxiety, PTSD, burnout, and help-seeking behavior) were included in the study protocol, survey response rates at follow-up were too low to permit meaningful quantitative analysis. As a result, the mental health survey data are not reported, and no conclusions can be drawn regarding changes in mental health outcomes based on these measures.

XSL•FO
RenderX

Qualitative Implementation Data

Qualitative data were collected to explore contextual factors influencing implementation. Interviews with the organizational champions included baseline interviews conducted prior to app implementation (n=19 interviews with 27 individuals) and follow-up interviews conducted at least 3 - 4 months following implementation (n=18 interviews with 26 individuals). Interviews were conducted virtually, audio-recorded, and transcribed verbatim, with durations of approximately 60 - 90 minutes. Some champions were interviewed individually, but others were interviewed in pairs or groups of 3. Interview guides with the champions were informed by the five dimensions of the CFIR, including questions exploring (1) perceptions of the intervention (characteristics of the app), (2) the outer setting (eg, societal context and policies), (3) the inner setting (eg, organizational culture and access to resources), (4) individuals (role of leaders, peer supporters, and characteristics of frontline workers), and (5) the implementation process [16]. Specific questions within each domain were developed using the CFIR interview guide tool. Since organizational champions were embedded within the organization, their insights enabled identification of critical factors that shaped uptake of the technology.

No predetermined numerical targets for adoption or engagement were specified. Given the exploratory nature of the implementation and the lack of established benchmarks for uptake of this type of workplace mental health and peer support app in comparable health care settings, the study aimed to descriptively characterize implementation outcomes and variability across organizations rather than assess performance against predefined thresholds.

Analysis

Descriptive data regarding app use included monthly reporting within each organization to examine trends in downloads and use of the different features. Cross-case comparisons were also conducted to review similarities and differences in usage patterns across organizations, but inferential statistics were not calculated.

Organizational champion interview data were transcribed and coded, with initial coding categories informed by the CFIR. Coding was an iterative process that began with reviewing transcripts, categorizing them by the 5 CFIR domains (deductively), and then inductively identifying the key issues within each domain. At least 2 team members coded each transcript, initially coding 2 - 3 transcripts independently, then meeting together to compare codes and refine the codebook to ensure consistency. A series of team meetings was held to discuss emerging analytic ideas and revise the codebook to capture key facilitators or barriers to implementation across organizational contexts. Strength and valence ratings of the key forces within each organization were conducted in order to facilitate cross-case comparison [24]. Identification of key forces represented comments across multiple organizations, with quotes selected to illustrate, considering representation across all participants. Dedoose (SocioCultural Research Consultants, LLC) software [25] was used to manage the qualitative interview data. The qualitative data were used to explain the app use trends and facilitate actionable insights to optimize implementation.

Strategies to ensure methodological rigor included triangulation of data sources (app use metrics and interviews with multiple champions over different time points in each organization). Triangulation strengthened in-depth understanding of the organizational context and implementation forces within and across organizations that shaped app use, leading to data saturation across categories and understanding of the contextual forces shaping objective patterns of app use. In order to ensure coding consistency, biweekly meetings were held with the research team to review the coding and to engage in peer debriefing. Meetings between coders were scheduled to calibrate approaches and refine definitions in the codebook until there was at least 85% coding consistency. Reflexive analysis was strengthened through analytic memos that were discussed in the peer debriefing meetings. CFIR was used as a sensitizing framework, but the iterative, inductive analysis process led to nuanced understanding of the key issues shared by participants. Member checking was conducted through sharing preliminary analyses with champions via draft organizational reports, then exploring their reflections on the analysis in follow-up interviews.

Results

Overview

The study findings are presented in three parts: (1) an overview of the participating organizations and app users, (2) descriptive metrics regarding patterns of app use, and (3) qualitative analysis of champion interview data regarding factors shaping app use (informed by the CFIR framework), considering the social and organizational context (outer and inner setting), response to the app innovation, influence of organizational stakeholders, and impact of the implementation strategies.

Participating Organizations

A total of 7 diverse health care organizations participated in the study, joining at different time points. As outlined in Table 1, 2 of the organizations were large (over 10,000 employees), and one was small (n=55 employees). Organization 2 was a large professional association that engaged in the study at different time points, initially with a subgroup of workers (2A/2B) and then later as the full organization (2C). Their participants were members of the association (paying membership dues), rather than employees, and the app was offered as a member benefit. Organizations 1 and 4 chose to focus on a subgroup of health care providers within the organization (nurses and PSWs, respectively), but other organizations provided access to all employees (clinical and nonclinical). Organizations 3, 6, and 7 were multisite facilities, and organizations 1 and 4 provided community care with no on-site services. The number of champions and peer supporters varied across sites. As such, there were various implementation contexts for implementation. The app for frontline workers was made available to a total of 32,753 health care workers across these organizations.

Demographic data were collected from app users in organizations 2C, 4, 5, 6, and 7 via onboarding demographic

XSL•FO
RenderX

questions that were added in June 2023. As noted in Table 2, app users in these organizations (n=840) included 77% (n=650) who identified as women, 21% (n=173) as men, and 2% (n=17) as nonbinary or chose not to disclose. Just over half of participants were between the ages of 30 and 49 (n=458, 55%) years. In terms of professional experience, 41% (n=348) had ≥16 years, followed by 15% (n=129) with 6 - 10 years. Most participants held clinical roles (n=505, 60%), while 24% (n=205) were nonclinical.

Table . Characteristics of app users (in organizations 2C, 4, 5, 6, and 7).

Characteristic	Participants
Age (years), n (%)	
18 - 29	103 (12)
30 - 39	217 (26)
40 - 49	241 (29)
50 - 59	204 (24)
60 - 69	47 (6)
Undisclosed	28 (3)
Gender, n (%)	
Man	173 (21)
Woman	650 (77)
Nonbinary or undisclosed	17 (2)
Experience (years), n (%)	
<2	101 (12)
2 - 5	108 (13)
6 - 10	219 (15)
11 - 15	114 (14)
≥16	348 (41)
Undisclosed	40 (5)
Role, n (%)	
Clinical	505 (60)
Nonclinical	205 (24)
Undisclosed	130 (16)

App Engagement Metrics

A total of 1066 employees downloaded the app across all organizations over the 4-month implementation trial. As noted in Table 3, the number and proportion of downloads within each organization ranged from <2% (n=108) in the largest organization to >45% (n=25) in the smallest organization. The mean download rate across organizations was approximately 11.6% (SD 15.2) of employees, but the overall downloads represented only 3% of the total number of eligible participants.

XSL•FO
RenderX

Table . App user downloads and feature use by organization.

Organization	Downloads, n (%)	App opens	Mean opens/ user	Content views, n (%)	Goals set, n (%)	Wellness check-ins, n (%)	Appointments booked (n)
1	38 (6)	127	3.34	32 (25)	2 (2)	13 (10)	2
2	112 (2)	457	4.89	154 (34)	38 (8)	79 (17)	49
3	84 (11)	273	3.25	38 (14)	24 (9)	43 (16)	4
4	73 (8)	157	2.15	29 (18)	6 (4)	23 (15)	4
5	25 (45)	85	3.40	63 (74)	19 (22)	19 (22)	0
6	468 (4)	1324	2.83	323 (24)	73 (6)	220 (17)	16
7	266 (5)	704	2.65	140 (20)	39 (6)	95 (13)	21
Total	1066 (3.24)	3127	—[a]	779	201	492	94
Average, mean (SD)	11.6 (15.2)	—	3.22	111	29	70	10

[a]Not applicable.

User engagement, measured by app opens and feature use, also varied across organizations. As noted in Table 3, the app was opened 3127 times over the course of the study, ranging from a total of 85 opens (45% of employees) in the smallest organization (Organization 5) to 1324 opens (4% of employees) in the second largest organization (Organization 6). Each downloaded app was opened a mean of 3.22 times (range from 2.1 to 4.9).

Given the sample size, normal approximation methods were used to calculate 95% CIs for proportions. A total of 309 app users viewed at least one content item, representing 30% of users (95% CI 27.2% - 32.9%). When viewing the content, users could provide a rating from 1 to 3 stars. There were 105 user-submitted content ratings, with a mean rating of 2.58 (SD 0.68) out of 3, corresponding to 86% of the maximum possible score. The most frequently accessed content by app users was an article titled “6 Mindfulness Exercises,” which received more than twice as many views (n=92) as the next most accessed item (n=35).

Engagement with other app features included (1) wellness check-ins initiated by 25% of app users (95% CI 22.3% - 27.7%), representing between 10% and 22% of the app openings; (2) goal setting initiated by 13% of app users (95% CI 11.0% - 15.2%); and (3) peer support outreach was accessed 94 times (56 calls and 38 texts) by 40 unique app users, with a wide range from 0 (in the smallest organization) to 49 (in the largest organization). Appointments were booked with 38 different peer support providers.

Analysis of Forces Shaping App Engagement

Overview

Data from 19 baseline and 18 follow-up interviews with organizational champions (n=37 in total) informed the analysis of factors shaping patterns of app use. Informed by dimensions of the CFIR framework (outer/external context, inner/organizational context, the app innovation, the individuals’ role/characteristics, and the implementation process), we will highlight key facilitators and barriers to engagement noted by the organizational champions, followed by an in-depth analysis of implementation strategies within and across organizations. See Table 4 for a summary of the findings organized by the CFIR dimensions. Participant quotations to illustrate each issue are identified by their organizational affiliation.

XSL•FO
RenderX

Table . Summary of facilitators and barriers to app use.

CFIR dimension	Facilitator	Barrier
Outer context	• High mental health needs as a result of system/pandemic pressures	• High fatigue, burnout, and mistrust of new initiatives as a result of system/pandemic pressures
Inner context	• Alignment with organizational mission to support staff • Complements other wellness initiatives	• Stigma associated with the need for mental health support • Culture of skepticism/mistrust • Not able to use phones at work • Unstable internet • No time or private space to call
Innovation characteristics	• Easy to use, accessible, relevant content, choice re: when/how to access support, anonymous and private connection to support	• Time needed to download
Individual roles and characteristics	• Commitment from organization champions and leaders • Trained peer supporters (local and external)	• Health care worker reluctance to reach out for support • Limited digital literacy
Implementation process	• In-person connections • Leadership endorsement • Messaging from trusted organization champions • Ongoing, strategic communication	• Passive promotion (email and posters) not engaging • In-person events don't reach everyone (shift work and offsite) • Employer-endorsed approach can be met with suspicion • Lack of proactive follow-up • Lack of integration with other wellness initiatives

Opportunities and Facilitators to Uptake

One consistent finding was that the external context of health care has created the need for mental health support. Many champions talked about the unique challenges and pressures faced by frontline health care workers. Even though the work context ranged from acute hospitals to community rehabilitation, and from large urban centers to small rural settings, almost all champions described challenges related to workload that were exacerbated by the COVID-19 pandemic, and how this elevated the level of fatigue and staff burnout. As one champion explained, "right now, we're seeing levels of burnout and a bit of jadedness that's come from being in this grind of the pandemic for quite a long time" (Organization 2). Similarly, another champion noted, "Health care is still reeling [from] the effects of COVID…there are staffing pressures and fatigue," underlining the urgent need for support, explaining that "the building is burning as we're trying to build the new fire station" (Organization 6). As a result, organizations were acutely aware of the significant pressures on staff members and the urgent need to support their frontline workers.

Introducing a tool like the Beyond Silence platform also aligned with the inner context or mission of most participating organizations to support employee mental health. Organizational planning in all organizations included a key focus on employee mental health and included support from the organizational leadership team. Champions explained that the app for frontline workers complemented existing wellness strategies, noting that peer support and digital tools were complementary or even natural extensions of their ongoing psychological health and mental wellness efforts. As noted below, one champion explained that the app was a catalyst for innovation and expansion of their wellness services:

> *It came at a very good time because our psychological health and wellness initiatives were occurring. And actually, having the app helped us broaden the people involved in our peer support efforts…. it also gave us momentum to do some creative things* [Organization 3]

There were also many characteristics of the app that were valued. Participants valued the relevance of the content, with customization to the unique context of health care. It was noted that the app was "easy to use... helpful, interesting articles, appropriate… relevant" (Organization 1). One of the participants explained:

> *the fact that it's geared specifically for health care providers, I think was really important for us. Anytime we're looking at providing a new product or service to our members, making sure that… it's very specific to their role. And so that was really appealing* [Organization 2]

Participants talked about the value of the portability of the app since employees could carry their phones with them as a "one-stop shop" (Organization 7) for information and support that they can access at any time. In terms of features, some talked about how they preferred to access the online resources; others liked the wellness check as a quick and easy tool to track

XSL•FO
RenderX

change over time. It was noted that the opportunity to text, as opposed to phoning someone, was important.

> *I definitely can see why the chat function would probably be-- it's easy, accessible. People chat a lot now as opposed to picking up a phone. Like, we're not as much a phone culture as we used to be.* [Organization 6]

Another valued feature was the opportunity to access anonymous, private connections to information and peer support.

> *There's anonymity there. You can go and talk to someone, look at resources, set goals for yourself that nobody else is privy to. You can listen to what you want to help build your pocket of resilience* [Organization 7]

In addition, several participants talked about the value of being able to browse the biographies of the peer support providers and choose whom and when to reach out. Overall, the opportunity to choose when and how to use the app features was noted as a benefit.

The final key facilitator relates to the characteristics and roles of individuals in the organization (champions, leaders, and peer supporters). In particular, many of the organizational champions played an active role in reaching out to their colleagues to promote the app, adopting a range of contextually relevant communication approaches. Several champions also noted the important role of supportive leaders:

> *I think those types of [supportive] messages coming from leaders carry a lot of weight, you know, and especially around sensitive topics, such as mental health and self-care.* [Organization 1]

In addition to the leaders, access to a network of trained peer support providers both within and outside the organization was noted to play a key role in facilitating engagement. This was particularly relevant for employees who did not have access to other supports. One organizational champion explained that not everyone has access to employee assistance programs (EAPs) and not all of them feel safe going to EAP. "And some of our lowest-paid workers, because they're part-time and casual, don't have access to EAP. The very people who need it the most" (Organization 4). The app for frontline workers therefore provided an opportunity to link with a trusted peer who had an understanding of their unique situation, filling an important gap in workplace supports.

Challenges or Barriers to Engagement

Despite the recognized need for mental health support and the potential for the Beyond Silence platform to meet this need, we heard about a number of barriers to adoption of the technology, many of which were related to the inner organizational context. These included cultural factors, such as workplace attitudes toward mental health and help-seeking, and structural challenges, including availability of time and space to download and use the app.

The workplace culture was identified as a challenge in many organizations in terms of trust and openness to seeking support. In some cases, we heard that there was a culture of stoicism and an expectation to manage the stress of the work.

> *Based on some preliminary conversations that I've had with our nursing group is that they … are more inclined to, kind of assume sort of a stoicism, like, you know, "This is sort of part of the job, part of the role. … and it is the responsibility of those that are in this role to be able to deal with that.* [Organization 1]

As outlined, many health care workers internalize the belief that their role is to provide support rather than seek it. This self-perception can hinder both the recognition and acknowledgment of personal mental health needs. Participants described a tendency among staff to often "avoid or ignore symptoms" (Organization 2A) and may not recognize a situation when the app might be valuable. Persistent stigma surrounding mental health further compounded this challenge, serving as a barrier to reaching out.

In addition to stigma, participants talked about skepticism and lack of trust, which interfered with uptake of any initiative or reaching out to others. Some explained how trust has been broken in the aftermath of COVID-19, and/or how people are tired and don't see how things will change. One participant talked about the skepticism, with employees asking, "How is this gonna help me when the situation is not changing?" (Organization 6). They noted high levels of fatigue and questioned the value of any new initiative. Another participant talked about the overall persona of a subgroup of laboratory workers who "trust no one." They explained that

> *they're highly skeptical. It makes them awesome at what they do. It also makes them sometimes not the easiest to work with because we trust nobody. So sharing things might be too much vulnerability for them* [Organization 2]

It was recognized that some employees may be more receptive than others to technology, or any new initiative. One participant explained that some are "just against whatever new thing that was going to happen" (Organization 3), and another explained that,

> *there's people who love technology and people who don't. There's people who really bring their issues to the team, and then there's people who are extremely private and, you know, won't want to be engaged in this at all* [Organization 5]

Although some may be more receptive than others, overcoming the reluctance of frontline workers was noted to be a significant challenge both within and across organizations.

Another barrier noted by many participants was linked to the infrastructure of work. Some could not use their phones at work, had unstable internet connections, or did not have the time or private space to make a call. Given that many are short-staffed at work, they explained that there is no time even to download the app, let alone reach out for support. As one participant noted, "If it takes more than three clicks to install, then forget it. … They have so many other priorities" (Organization 5). There were many comments about the lack of time to access the app

XSL•FO
RenderX

during work hours, explaining that they think "I don't have time now, I'll do it later," but then later never comes (Organization 6).

Overall, analysis of the forces shaping app engagement revealed that despite perceiving the app as a valued tool that complements existing services and supports, there was a disconnect in terms of actual uptake. As one champion noted,

> *"They see it. They value it. They say they like it. They recognize they need it moving forward. However… when they're asked to opt into those types of services, uptake is almost always super low, lower than what we would expect given how much they say they value it, how much they say they like it"* [Organization 2]

This sentiment was echoed by another participant,

> *Some people said they weren't sure why they would use it. … [others said] 'Oh, this is great. This is a great idea.' But they weren't at the stage where they needed it yet. It was just at the curiosity stage* [Organization 6]

These comments underscore the challenges of engagement and suggest that transitioning from initial interest to adoption may take time. They also reflect a perceived lack of an immediate or urgent need to use the new technology.

Implementation Strategies to Promote Engagement

The implementation process was an important factor that shaped uptake of the technology. Implementation included communication about the app to facilitate awareness and inspire employee engagement. This process looked a little different in each organization given their established communication channels and resources. Common engagement methods included targeted email campaigns, in-person launch events, and dissemination of printed materials such as posters, wallet cards, and flyers, and distribution of promotional materials (pens and water bottles) to enhance visibility and encourage uptake.

In all organizations, champions were asked to send email messages to employees inviting them to participate in the research study, with a brief description of the project and features of the app, a poster that provided a QR code to learn more about and download the app, and a unique organizational code to open the app. In 4 of the 7 organizations, information was posted in monthly or bimonthly staff newsletters and/or on the staff wellness section of their intranet. In general, more passive strategies (eg, email and newsletters) were seen as more efficient but not as effective in promoting engagement. As one champion noted, "I don't think blasting emails works; our staff are always in transit" (Organization 4). Another champion talked about information overload, "[I] suspect people are just inundated with information hanging on the bulletin boards. So, it probably, unfortunately, got overlooked" (Organization 5).

In contrast, in-person launch events were reported to be much more effective. In six of the organizations, the materials were distributed at on-site launch events where staff could get assistance downloading the app (if needed) and ask questions about the project and the technology. The highest download rate (45%) was in a small organization where employees in a staff meeting were engaged in a structured engagement activity (20 - 30 minutes), to not only download the app, but to engage with different features. In 2 of the multisite hospitals, a brief visit was scheduled at multiple sites where staff were invited to meet the team and engage with the technology. Although it was noted that in-person promotion was effective, it can also be difficult to achieve, "the challenge... is trying to just find people... and get in front of them for, for five minutes in, in a face-to-face" (Organization 1). In some cases, staff worked shifts or across different sites, so it was difficult to find a time and place to reach everyone. Not everyone was able to engage in the on-site launches due to the demands of their work; therefore, the in-person sessions only reached a small fraction of employees, particularly in larger sites. In other cases, staff worked primarily offsite so it was not possible to do an on-site launch.

In addition to the format for information sharing, the people providing the messaging had an important impact on uptake of the technology. Champions in this study came from a range of departments and levels within their respective organizations, varying from frontline supervisors to senior directors to health and wellness leads. Implementation success was often tied to a champion's ability to leverage established networks and tailor outreach in ways that aligned with the organization's structure. Endorsement from leadership was noted to be important, "leadership involvement made all the difference; when supervisors talked about the app at team meetings, it sent a message that this was important" (Organization 6). While top-down endorsement was valuable, champions also reflected on the potential drawbacks of delivery through formal leadership structures, particularly if staff did not trust management. As one champion noted:

> *So, I think we were very well-intentioned to have our... delivery of this app come from management so that staff understood that it was valued and important. However, the double-edge of that is then perhaps the buy-in isn't as great if folks are struggling with anything to do with the agency. So, they may see it as, 'Here's yet another thing my employer is trying to do, and I'm already frustrated with them. Why would I engage?'* [Organization 5]

The same participant described the importance of trusted peers as an alternative to messaging from leaders.

> *It doesn't necessarily have to be the most extroverted person. It doesn't even have to be anybody in any kind of formal leadership position. It just has to be somebody whose staff are willing to look to and hear them when they speak. And I think that would create better buy-in.* [Organization 5]

The person delivering the information was therefore considered to be just as important as the nature of the message.

The other important lesson in implementation was the importance of ongoing, strategic communication to sustain interest and engagement over time. The organization with the lowest level of engagement sent an initial message but did not engage in proactive follow-up. Regular updates, transparent communication, and collaborative decision-making processes

XSL•FO
RenderX

were important. Sending "push notifications" through the app itself regarding new content, or reminders to reach out was noted to be a valuable feature of the app. It was noted that "every time we push something out, you get a little [spark] in behavior" (Organization 2). One of the most impactful strategies seemed to be personal follow-up. As one champion noted:

> *it's really me walking around and doing check-ins. And I think it's the champions keeping it front and center … us having it on our iPads and showing them [in peer support drop-in sessions]* [Organization 3]

Champions talked about natural time points when a reminder about the app was effective, explaining that a critical incident in the organization would be an ideal time to remind staff about the resources available to them.

Discussion

Principal Findings

The study findings point to an interesting paradox regarding adoption of the Beyond Silence platform. On the one hand, we heard many positive comments about the need to support frontline health care workers and how the app met many needs for a customized, confidential tool to increase access to relevant information and peer support. On the other hand, the overall app use was relatively low (download rate of <10% in 5 of 7 organizations; average of just over 3 opens per user), which was attributed to a range of barriers to engagement.

Theories about technology adoption can shed light on the findings related to app engagement. The technology acceptance model (TAM), for example, focuses on perceived usefulness and perceived ease of use, and how these impact intention to use the technology [23]. In our study, organizational champions noted that the app was easy to use and that it was very useful in enhancing their current employee support initiatives. These beliefs, however, did not necessarily translate to frontline workers. Ease of use for frontline workers was complicated by poor internet connections in some organizations or inability to use their phones at work, and by lack of time and private space to download and use the app. Furthermore, frontline health care providers may not see the usefulness of the app given that they often ignore their own support needs in their effort to help others, and by their skepticism, lack of energy, and time to invest in new initiatives. As such, they may be more reluctant to use a digital platform to access formal peer support given that it represents a departure from simple, informal interactions with peers and requires time and effort to learn something new.

The unified theory of acceptance and use of technology (UTAUT) is another model, which builds on the factors outlined in the TAM model to expand understanding of forces influencing adoption. In addition to the focus on effort expectancy (perceptions re: ease of use) and performance expectancy (the belief that using the app will increase personal productivity), this model sheds light on the important role of social influence (beliefs that key people expect them to use the app) and facilitating conditions (belief that the necessary organizational and technical infrastructure exists to support usage) on intention to use and actual use of technology [24]. In our study, there were differences between organizations regarding both social influence and facilitating conditions. In the smallest organization, which had the highest proportion of downloads (45% in Organization 5), the organizational champion had many personal connections with their colleagues (social influence), and they set aside time during a staff meeting for employees to learn about the app and try out the different features (facilitating conditions), thereby increasing initial employee engagement. In contrast, the large professional organization with the lowest proportion of downloads (2% in Organization 2) had limited personal connection with members since members worked in diverse organizations across the country. Furthermore, this organization did not have the infrastructure to actively invest in supporting app usage, using largely passive means of communication about the opportunity to participate (initial but no follow-up emails and occasional posts about the app in the electronic membership newsletter). These implementation conditions at an organizational level with limited contact and few active engagement opportunities clearly had an impact on employee engagement.

Finally, the framework outlining NASSS of health and care technology provides additional insights to consider given that the Beyond Silence platform represents a complex intervention that is implemented at an organizational level [14]. According to the NASSS framework, there are 6 main sources of complexity that need to be considered in understanding technology implementation and uptake, namely the condition, the technology, the value proposition, the adopter system (eg, staff and patients), organization implementation, and the wider (institutional and societal) context. In the case of the Beyond Silence platform, mental health "conditions" can be considered complex in that they are often insidious, unpredictable, and stigmatized, and therefore can be difficult to identify and manage. Health care workers may fail to perceive or acknowledge their own psychological distress and often avoid seeking support due to the stigma associated with mental health issues [21,26]. The technology itself was noted to be valuable, but it is available on an "as needed" basis with no requirement for initial or ongoing use. As noted in a systematic review of eHealth interventions with health care professionals, smartphone apps have the advantage of convenience, but they may be either too flexible or too subtle to prompt regular engagement [14]. Another area of complexity relates to the value proposition for frontline workers. Although there was general buy-in from leaders, the value was not necessarily evident for frontline workers given the time, energy, and trust required to adopt something new. Organization implementation was also complex given the challenges in connecting with workers who may work different shifts, across different sites, and have little time available to download or use new technology. Understanding and addressing these dimensions of complexity is critical to creating an effective implementation plan.

The insights provided by these technology implementation theories reveal that low adoption rates may not be a failure of the technology itself, but rather a failure of implementation. There is a growing body of literature on e-mental health interventions in the workplace, including several systematic reviews [27,28], and one umbrella review of 14 review articles

XSL•FO
RenderX

published between 2014 and 2023 [13]. The review studies explore a range of web-based and app-based tools, with many rooted in principles of cognitive behavioral therapy (CBT) and mindfulness, including studies that focus specifically on health care workers. Overall, there appears to be a positive impact of these tools on reducing mental distress and improving employee mental health, with a clear value proposition related to accessible, customizable tools that can expand the reach of workplace mental health initiatives [6,13,28]. These studies underscore the value of the technology itself in the context of the workplace.

A notable gap in the literature, however, is in examining implementation processes that shape uptake and impact. Studies of digital mental health interventions note variability in engagement across studies, with a paucity of consistent data to track adoption and engagement, making it difficult to conduct comparisons across different types of technology innovations and workplace contexts [29,30]. As such, this study contributes important data regarding implementation of an app-based peer support intervention, comparing implementation across a range of organizational contexts in the health care sector. The focus on an upstream mHealth intervention, rooted in self-management and peer support, is also unique, given that many studies focus on therapeutic apps (cognitive behavioral therapy and mindfulness).

Recommendations

Based on analysis of the key facilitators and barriers noted in this study, as well as principles of implementation science theory and practice noted above, there are several recommendations to consider in order to optimize future uptake of workplace mental health apps such as the Beyond Silence platform within the context of health care workplaces. The recommendations are also based on a review of several implementation checklists and guidelines regarding strategies to foster implementation [31], including the Expert Recommendations for Implementing Change (ERIC) [32] and the European Platform to Promote Wellbeing and Health in the workplace (EMPOWER) checklist, which specifically examines strategies to foster adoption of occupational e-mental health interventions [33].

Evaluate Organizational Readiness for Implementation

The concept of organizational readiness to change, as theorized by Weiner [34], involves a shared resolve among organizational members (health care workers in this case) to implement a change (adoption of the Beyond Silence platform), and a shared belief in their individual and collective capability to do so. As noted in the TAM and UTAUT models, shared beliefs about the value of the app and how easy it is to use are important prerequisites to acceptance and actual use [35,36]. Readiness can be different at an individual, group, program, or organizational level [34]. The findings of our study, for example, revealed that organizational leaders appeared to be more committed and confident in adopting the Beyond Silence technology when compared to frontline workers. The degree to which they were committed to building a psychologically safe work environment beyond technology implementation; however, was not evaluated. Leadership commitment is consistently noted to be an important determinant of implementation success; trusted frontline managers can be influential, underscoring the importance of relational trust in driving intervention uptake [37]. Less is known, however, about the commitment of frontline workers. Even some programs within multisite organizations in our study were more engaged than others. Taking time to note dimensions of readiness and as well as potential sources of resistance is important in developing an implementation plan. Establishing the infrastructure to support implementation (eg, champions for change, clear, and sustained communication strategies) as well as opportunities to build collective commitment are critical to success.

Address Mental Health Literacy and Stigma

Beyond Silence is a platform for accessing information and support, but if people don't recognize when and how to reach out, and/or have concerns about the stigma associated with seeking help, they may not see the need and be reluctant to use the technology. It is therefore important to build awareness such as education on burnout, moral injury, and signs of anxiety or depression, as well as access to self-screening tools (like the ones available in the app). Personal stories about seeking help from respected colleagues can be effective to reduce stigma and build a culture of proactive outreach [21,38], as is encouragement from managers and peers regarding the value of peer and professional support. We have started to develop "use case" examples for when and how to use the app, with strategies to emphasize access, noting that "help is just a click away." To address stigma, it is also important to emphasize the privacy and anonymity features of the app and the peer support outreach.

Generate a Communication and Implementation Plan That Includes Multimodal Strategies to Reach a Range of Employees

Identifying who might be credible, trusted champions for change, including both leaders and frontline influencers, is a critical step since a range of voices at different levels of the organization is needed to inspire change [33]. In terms of messaging, passive strategies such as email and posters are an efficient way to reach people but may be overlooked given the many competing demands for attention. The most effective strategies involved setting aside time and space to download the app, try out features, and ask questions. Time has been noted as an influential determinant of app use in the implementation literature and one that can compromise adoption and sustainability [39]. Addressing time barriers therefore requires selecting or tailoring strategies that reduce the time burden and align with existing workflows. Implementation efforts should therefore incorporate structured assessments of time availability early in the planning process and prioritize low-effort, high-impact adaptations that facilitate integration into routine practice. Consider introducing the app at key moments (eg, new staff orientation, morning huddles, or following difficult shifts). Since staff turnover can be high in some health care settings, introducing the app as part of new employee orientation can be an effective way to embed it into organizational communication strategies. There is considerable evidence that initial and ongoing communication is also needed to sustain engagement over time [33,39]. Consider how and when to promote during

XSL•FO
RenderX

the year (eg, mental health week or prior to the holiday season, which can be difficult and isolating for some employees). Without ongoing reinforcement and integration into regular workflows, engagement may decline over time [31].

Consider Technology Requirements, Including mHealth Access and Digital Literacy

Digital e-mental health interventions can be valuable tools to increase access to information and support, but varying levels of comfort with technology can impact adoption [33,39]. Champions in our study noted generational differences in digital literacy; therefore, it may be important to host hands-on/experiential practice sessions with employees to not only download the app but also to explore the different features. Virtual onboarding sessions, question and answer (Q&A) sessions, tutorials, frequently asked questions (FAQs; available in the app), and a contact for technology troubleshooting are important strategies to address questions that may arise over time [33]. It should be noted that some of the technology issues noted in our study were related more to the infrastructure of work, from unreliable internet to a lack of time or private space to access support. This may require organization-level interventions to build time and space for support.

Embed the App Platform Into Existing Organizational Policies and Practices

It takes time to introduce new ways of thinking about peer support and to build awareness, trust, and new habits regarding technology as a tool to access information and support [39]. There are many reports of the importance of social support in the context of health care [3,40], but this is often informal, and the idea of reaching out through an app may take time to build awareness and comfort with what peer support is and how it can help. Integrating the app platform into the workplace requires clear communication on how it fits within an overall model of care [33,41]. There is no “one size fits all” approach since every person is different in terms of their needs and preferences for accessing support; there may be generational differences in digital literacy and openness to accessing support, as well as different levels of need for peer support. The Beyond Silence platform was designed to promote early intervention and support but could also be another tool in the toolbox to support new employees, to facilitate connection with professional supports, or to facilitate reintegration following a workplace absence. There is currently no protocol for using the app; it is available on an “as needed” basis, but the platform is designed to easily add new content, depending upon what might be relevant to the organization. Co-developing a vision for the peer support technology within existing initiatives and working with intended users to co-create the work routines that the technology is intended to support are important strategies to optimize uptake [41]. Depending on the goal of the organization, it may not be realistic to expect all employees to use the app on a regular basis; in some cases, simply ensuring it is available when needed may be sufficient.

Study Strengths, Limitations, and Future Work

This study provides an early examination of mental health app implementation across a diverse range of health care organizations, informed by the CFIR framework and consideration of a range of technology adoption models. There are, however, several limitations, which constrain the strength and generalizability of the findings. A 4-month implementation period provided an opportunity to observe initial short-term patterns, but may not have been sufficient to assess longer-term integration into organizational culture. The follow-up period revealed the challenges associated with maintaining long-term engagement, highlighting the importance of iterative refinement and continued evaluation. Studies of EAP use, for example, report low annual usage rates but significantly higher uptake over a 5-year period, reflecting the time required for integration into organizational culture and workflows [42]. These findings underscore the importance of examining cumulative use over time, with at least a 6 - 18 month implementation period. Future research should explore extended longitudinal studies to assess delayed impacts and long-term behavioral changes. As such, the present findings should be interpreted as preliminary, and future studies should use longer observation periods to capture delayed adoption and sustained behavioral change.

The sample of 7 participating organizations also limits generalizability. Although they represented diversity in size, function, and location, all were situated within the Canadian health care system and represented a time when formal peer support and technology to support workplace health were relatively new and a time when the health care system was still recovering from the impact of the pandemic. International contexts may vary in terms of policy environments, workplace cultures, and the maturity of peer support infrastructures.

Data were only collected from organizational champions regarding why they thought that employees downloaded and engaged with (or did not engage with) the app. As nonadopters represented a significant proportion of eligible staff, their perspectives, barriers, and decision-making processes are unknown. This limits understanding of overall receptivity, organizational fit, and determinants of uptake. In addition, the absence of a comparison group restricts the ability to infer causal relationships between implementation strategies and observed outcomes.

The focus of the study was on implementation but did not track mental health or workplace outcomes of the intervention. Future research should incorporate outcome evaluation alongside implementation assessment, include comparison groups, extend follow-up periods, and gather data from both adopters and nonadopters to produce more conclusive evidence regarding effectiveness, reach, and generalizability. There is also a need to track the cost-effectiveness of this intervention approach to help employers to reflect on the value of this type of digital support intervention in their workplace context.

Conclusion

This study provides important insights regarding implementation of the Beyond Silence peer support platform for health care workers. The findings point to a relatively slow adoption of the technology, attributed in part to the stigma and silence surrounding mental health issues, to competing demands on attention and time, and challenges with initial and ongoing engagement of frontline workers. At the same time, the findings

XSL•FO
RenderX

also point to strong mission alignment for organizations that are looking for tools to support their frontline workers, and strong support for the opportunity to access high-quality customized information and resources for health care providers and to access tools to facilitate anonymous connections to trained peer supporters. We argue that understanding and addressing issues related to implementation is a necessary precondition to optimize impacts on mental health and workplace outcomes. Inspiring frontline workers to take the time and energy to access resources and peer support via the app requires reflection on potential barriers to change, and innovative approaches to build trust and inspire new ways of responding to mental health issues. Longitudinal research is needed since time is needed to study integration of the technology into workplace policies and practices.

Acknowledgments

We are grateful to trainees and research staff, including Sheila Addanki, Saleha Ansari, Maria Bargeman, Lindsay Crocco, Esther Chow, Vivian Lee, and Lisa Wolfe who played an instrumental role in supporting various stages of the app development, project implementation and evaluation. Thanks also to Henrietta Van Hulle who was an integral member of the advisory team and to the many health care leaders and champions who contributed their valuable time and insights.

Generative AI was not used in any portion of the manuscript writing.

Funding

The research was supported through funding from the Public Health Agency of Canada (PHAC) as part of their Addressing Posttraumatic Stress Disorder (PTSD) and Trauma in Those Most Affected by COVID – 19 initiative [Grant: 2122-HQ-000408]. The funder had no involvement in the study design, data collection, analysis, interpretation, or the writing of the manuscript.

Data Availability

Data specific to each organization and app user are considered confidential, so they are not available in a public repository. Requests for deidentified data files can be made directly to the corresponding author.

Authors' Contributions

SM led conceptualization of the study and funding acquisition, and was primarily responsible for the overall investigation, methodology, formal analysis, supervision of the study team, and writing the initial and revised drafts of the paper. MP was responsible for project administration and managing resources, and contributed to data curation, formal analysis, data visualization, and writing the initial draft, as well as review and editing. PS contributed to the conceptualization of the study, formal analysis of the study data, and to the review and editing of the final paper. ES contributed to computer software resources and support, and to writing, review, and editing.

Conflicts of Interest

The first author (SM) was the developer of the Beyond Silence technology, owned by McMaster University. A partnership with Stepped Care Solutions (a not-for-profit organization) was developed following completion of the research in 2024 in order to lead commercialization on a cost-recovery basis. The agreement included a clause for ongoing access for research, and the potential for a small royalty from commercialization that would be reinvested into research.

References

1. Singh J, Poon DEO, Alvarez E, et al. Burnout among public health workers in Canada: a cross-sectional study. BMC Public Health 2024 Jan 2;24(1):48. [doi: 10.1186/s12889-023-17572-w] [Medline: 38166742]
2. Shanafelt T, Kuriakose C. Widespread clinician shortages create a crisis that will take years to resolve. NEJM Catalyst 2023 Feb 15;4(3). [doi: 10.1056/CAT.23.0044]
3. Carbone R, Ferrari S, Callegarin S, et al. Peer support between healthcare workers in hospital and out-of-hospital settings: a scoping review. Acta Biomed 2022 Oct 26;93(5):e2022308. [doi: 10.23750/abm.v93i5.13729] [Medline: 36300208]
4. Simms L, Ottman KE, Griffith JL, et al. Psychosocial peer support to address mental health and burnout of health care workers affected by COVID-19: a qualitative evaluation. Int J Environ Res Public Health 2023 Mar 3;20(5):4536. [doi: 10.3390/ijerph20054536] [Medline: 36901545]
5. Albott CS, Wozniak JR, McGlinch BP, Wall MH, Gold BS, Vinogradov S. Battle buddies: rapid deployment of a psychological resilience intervention for health care workers during the COVID-19 pandemic. Anesth Analg 2020 Jul;131(1):43-54. [doi: 10.1213/ANE.0000000000004912] [Medline: 32345861]
6. Anderson GS, Di Nota PM, Groll D, Carleton RN. Peer support and crisis-focused psychological interventions designed to mitigate post-traumatic stress injuries among public safety and frontline healthcare personnel: a systematic review. Int J Environ Res Public Health 2020 Oct 20;17(20):7645. [doi: 10.3390/ijerph17207645] [Medline: 33092146]

XSL•FO
RenderX

7. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. J Psychiatr Res 2018 Dec;107:73-78. [doi: 10.1016/j.jpsychires.2018.10.006] [Medline: 30347316]
8. Stratton E, Jones N, Peters SE, Torous J, Glozier N. Digital mHealth interventions for employees: systematic review and meta-analysis of their effects on workplace outcomes. J Occup Environ Med 2021 Aug 1;63(8):e512-e525. [doi: 10.1097/JOM.0000000000002267] [Medline: 34050095]
9. Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency. Narrative literature review. Br J Psychiatry 2005 Jan;186(1):11-17. [doi: 10.1192/bjp.186.1.11] [Medline: 15630118]
10. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 1;3(1):e7. [doi: 10.2196/mental.4984] [Medline: 26932350]
11. Addanki S, Macedo L, MacDermid J, Moll S. Mental health apps implemented in the workplace: scoping review of trends and gaps in evaluation research. JMIR Mhealth Uhealth 2026 Mar 31;14:e57046. [doi: 10.2196/57046] [Medline: 41915896]
12. Muuraiskangas S, Harjumaa M, Kaipainen K, Ermes M. Process and effects evaluation of a digital mental health intervention targeted at improving occupational well-being: lessons from an intervention study with failed adoption. JMIR Ment Health 2016 May 11;3(2):e13. [doi: 10.2196/mental.4465] [Medline: 27170553]
13. Cameron G, Mulvenna M, Ennis E, et al. Effectiveness of digital mental health interventions in the workplace: umbrella review of systematic reviews. JMIR Ment Health 2025 Jan 24;12(1):e67785. [doi: 10.2196/67785] [Medline: 39854722]
14. López-Del-Hoyo Y, Fernández-Martínez S, Pérez-Aranda A, et al. Effects of eHealth interventions on stress reduction and mental health promotion in healthcare professionals: a systematic review. J Clin Nurs 2023 Sep;32(17-18):5514-5533. [doi: 10.1111/jocn.16634] [Medline: 36703266]
15. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367. [doi: 10.2196/jmir.8775] [Medline: 29092808]
16. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated consolidated framework for implementation research based on user feedback. Implement Sci 2022 Oct 29;17(1):75. [doi: 10.1186/s13012-022-01245-0] [Medline: 36309746]
17. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. Health Serv Res 2013 Dec;48(6 Pt 2):2134-2156. [doi: 10.1111/1475-6773.12117] [Medline: 24279835]
18. Yin RK. Case Study Research: Design and Methods, 5th edition: Sage Publications; 2014.
19. Moll SE. The web of silence: a qualitative case study of early intervention and support for healthcare workers with mental ill-health. BMC Public Health 2014 Feb 8;14(1):138. [doi: 10.1186/1471-2458-14-138] [Medline: 24507543]
20. Moll SE, VandenBussche J, Brooks K, et al. Workplace Mental Health Training in Health Care: Key Ingredients of Implementation. Can J Psychiatry 2018 Dec;63(12):834-841. [doi: 10.1177/0706743718762100] [Medline: 29482357]
21. Graham ID, Logan J, Harrison MB, et al. Lost in knowledge translation: time for a map? J Contin Educ Health Prof 2006;26(1):13-24. [doi: 10.1002/chp.47] [Medline: 16557505]
22. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
23. Moll SE, Patten S, Stuart H, MacDermid JC, Kirsh B. Beyond silence: a randomized, parallel-group trial exploring the impact of workplace mental health literacy training with healthcare employees. Can J Psychiatry 2018 Dec;63(12):826-833. [doi: 10.1177/0706743718766051] [Medline: 29673271]
24. Reardon CM, Damschroder LJ, Ashcraft LE, et al. The Consolidated Framework for Implementation Research (CFIR) user guide: a five-step guide for conducting implementation research using the framework. Implement Sci 2025 Aug 16;20(1):39. [doi: 10.1186/s13012-025-01450-7] [Medline: 40819067]
25. Dedoose version 90107: cloud application for managing, analyzing, and presenting qualitative and mixed method research data. Dedoose.: SocioCultural Research Consultants, LLC; 2023. URL: https://www.dedoose.com [accessed 2026-07-02]
26. Liu S, Hu Y, Pfaff H, Lei X, Qi Z, Feng X. Barriers and facilitators to seeking psychological support among healthcare professionals: a qualitative study using the theoretical domains framework. BMC Public Health 2025;25(1):848. [doi: 10.1186/s12889-025-21912-3]
27. Stratton E, Lampit A, Choi I, Calvo RA, Harvey SB, Glozier N. Effectiveness of eHealth interventions for reducing mental health conditions in employees: a systematic review and meta-analysis. In: Reed P, editor. PLOS ONE 2017;12(12):e0189904. [doi: 10.1371/journal.pone.0189904] [Medline: 29267334]
28. Carolan S, Harris PR, Cavanagh K. Improving employee well-being and effectiveness: systematic review and meta-analysis of web-based psychological interventions delivered in the workplace. J Med Internet Res 2017 Jul 26;19(7):e271. [doi: 10.2196/jmir.7583] [Medline: 28747293]
29. Addanki S, Macedo L, MacDermid J, Moll S. Mental health apps implemented in the workplace: a scoping review of trends and gaps in evaluation research. JMIR Mhealth Uhealth ;14:e57046-e57046. [doi: 10.2196/57046] [Medline: 41915896]

XSL•FO
RenderX

30. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
31. Proctor E, Silmere H, Raghavan R, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76. [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
32. Powell BJ, Waltz TJ, Chinman MJ, et al. A refined compilation of implementation strategies: results from the expert recommendations for implementing change (ERIC) project. Implement Sci 2015 Feb 12;10(1):21. [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
33. Raggi A, Bernard RM, Toppo C, et al. The EMPOWER occupational e-mental health intervention implementation checklist to foster e-mental health interventions in the workplace: development study. J Med Internet Res 2024 Mar 15;26:e48504. [doi: 10.2196/48504] [Medline: 38488846]
34. Weiner BJ. A theory of organizational readiness for change. Implement Sci 2009 Oct 19;4(1):67. [doi: 10.1186/1748-5908-4-67] [Medline: 19840381]
35. Mugo D, Njagi K, Chemwei B, Motanya J. The technology acceptance model (TAM) and its application to the utilization of mobile learning technologies. Br J Math Comput Sci 2017;20(4):1-8. [doi: 10.9734/BJMCS/2017/29015]
36. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view1. MIS Q 2003 Sep 1;27(3):425-478. [doi: 10.2307/30036540]
37. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629. [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
38. Fernandez ME, Ten Hoor GA, van Lieshout S, et al. Implementation mapping: using intervention mapping to develop implementation strategies. Front Public Health 2019;7:158. [doi: 10.3389/fpubh.2019.00158] [Medline: 31275915]
39. Bernard RM, Toppo C, Raggi A, et al. Strategies for implementing occupational eMental health interventions: scoping review. J Med Internet Res 2022 Jun 1;24(6):e34479. [doi: 10.2196/34479] [Medline: 35648457]
40. Russell B, Ahmadzadeh A, Haris K, et al. Peer support programs for physicians and health care providers: a scoping review. Jt Comm J Qual Patient Saf 2025 Sep;51(9):589-600. [doi: 10.1016/j.jcjq.2025.05.003] [Medline: 40683815]
41. Greenhalgh T. How to improve success of technology projects in health and social care. Public Health Res Pract 2018 Sep 27;28(3):2831815. [doi: 10.17061/phrp2831815] [Medline: 30406256]
42. Masi DA, Frey JJ, Harting J, Spearing M. Data Game Changer: Current Utilization Figures Inaccurate. Employee Assistance 2022;52:18-21 [FREE Full text]

Abbreviations

CBT: cognitive behavioral therapy
CFIR: Consolidated Framework for Implementation Research
CIHR: Canadian Institute of Health Research
EAP: employee assistance program
EMPOWER: European Platform to Promote Wellbeing and Health in the Workplace
ERIC: Expert Recommendations for Implementing Change
FAQ: frequently asked questions
HiREB: Hamilton Integrated Research Ethics Board
mHealth: mobile health
NASSS: Nonadoption, Abandonment, Scale-up, Spread, and Sustainability
PHAC : Public Health Agency of Canada
PSW: personal support worker
PTSD: posttraumatic stress disorder
Q&A: questions and answers
TAM: technology acceptance model
UTAUT: unified theory of acceptance and use of technology

XSL•FO
RenderX

Edited by A Stone, L Buis; submitted 13.Oct.2025; peer-reviewed by C Anyaegbuna, M Pulier; revised version received 17.Jun.2026; accepted 22.Jun.2026; published 05.Aug.2026.

Please cite as:
Moll S, Parker M, Smith P, Sykes E
Adoption of App-Based Peer Support by Canadian Health Care Providers: Mixed Methods Organizational Implementation Study
JMIR Mhealth Uhealth 2026;14:e85658
URL: https://mhealth.jmir.org/2026/1/e85658
doi:10.2196/85658

© Sandra Moll, Melissa Parker, Peter Smith, Edward Sykes. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 5.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Action Planning for Reducing Sugar-Sweetened Beverage Intake in Appalachian Adults: Longitudinal Process Evaluation of a Digital Behavioral Health Intervention

Annie Reid[1], MPH, CHES, RDN; Lee Ritterband[2], PhD; Kathleen Porter[1], PhD, RDN; Donna-Jean Brock[1], MS; Christina Frederick[2], BS; Jamie Zoellner[1], PhD, RDN

[1]Department of Public Health Sciences, School of Medicine, University of Virginia, 16 E Main St, Christiansburg, VA, United States

[2]Department of Psychiatry and Neurobehavioral Sciences, School of Medicine, University of Virginia, Charlottesville, VA, United States

Corresponding Author:
Jamie Zoellner, PhD, RDN
Department of Public Health Sciences, School of Medicine, University of Virginia, 16 E Main St, Christiansburg, VA, United States

Abstract

Background: Digital health interventions show promise for promoting behavior change, but how they incorporate action planning strategies is underreported. This oversight limits understanding of how to implement behavior change techniques. iSIPsmarter is a digital health intervention aimed at reducing sugar-sweetened beverage (SSB) consumption among Appalachian adults.

Objective: This study aimed to examine the digital action planning process in the iSIPsmarter intervention, specifically by (1) assessing the frequency of action plan engagement, (2) evaluating participants' perceived difficulty implementing their action plans and examining progress toward achieving SSB reduction and weight goals, and (3) exploring the selection of barriers and strategies.

Methods: The digital action planning process is embedded within 5 of iSIPsmarter's 6 behavioral content modules (Cores) and paired with self-monitoring of SSB intake via SMS text messaging and weight via a cellular-enabled scale. Participants first self-select program goals for SSB intake (in ounces) and weight (loss or maintenance). Then, in Cores 2 - 6, they complete action plans using personalized tracking feedback, recommendations, and goal progress updates. Participants identify barriers and strategies using preprogrammed or write-in responses. Summary statistics described the aims.

Results: Participants (n=119) were predominately White (Caucasian), female, aged between 18 and 44 years, college-educated, and from rural counties. On average, participants completed 4.5 (SD 1.1) of 5 possible SSB action plans, with 80% (95/119) completing all 5. Across all Cores, perceived difficulty implementing action plans and achieving goals remained relatively stable, with an average of 48% rating the tasks as impossible or hard, 29% as neither hard nor easy, and 24% as easy or very easy. Nearly half achieved their self-selected weekly SSB goals, and one-third made progress toward them. At Core 6, 57% (54/95) of participants met their self-selected SSB program goal, while 46% (44/95) met the recommended SSB intake of less than 8 ounces per day. Of 119 participants, 53 (45%) modified their SSB barriers, and 63 (53%) selected new strategies during action planning. Top reported SSB barriers included (1) caffeine, (2) taste, and (3) habit. Among those with a program weight loss goal (n=94), the average weight loss was −1.3% (SD 2.6) at the 9-week follow-up, with 61% (57/94) achieving their goal. By 6 months, weight loss increased to −2.1% (SD 5.6), with 54% (49/90) achieving their goal. Of those completing weight action plans, 62% (69/112) modified their barriers. Top weight barriers included (1) sweets, (2) portion sizes, and (3) eating healthy foods.

Conclusions: Findings underscore the value of digital action planning as a central behavior change technique within a nutrition-focused digital intervention. High action plan completion and consistent strategy adaptation suggest that structured, digitally personalized goal setting and action planning processes can effectively support behavior change, particularly among underserved populations with limited access to preventative care.

Trial Registration: ClinicalTrials.gov NCT05030753; https://clinicaltrials.gov/study/NCT05030753

International Registered Report Identifier (IRRID): RR2-10.1016/j.cct.2021.106566

(JMIR Mhealth Uhealth 2026;14:e71241) doi:10.2196/71241

KEYWORDS

digital health intervention; action planning; sugar-sweetened beverages; behavior change; process evaluation

XSL•FO
RenderX

Introduction

Background

Digital technologies, which include web programs, mobile apps, and SMS text messaging, are increasingly being used within behavior change interventions. Their widespread availability marks a promising shift in the field, offering improvements in accessibility and personalized support. This is especially important for hard-to-reach populations, such as those in rural or underserved areas, who often face barriers to traditional in-person interventions [1]. By leveraging digital platforms, these interventions can extend their reach and improve health outcomes in populations with limited access to preventive care [2].

Behavioral interventions often use behavior change techniques (BCTs), which provide structured strategies targeting specific behavioral mechanisms [3]. Michie and colleagues' [4] well-recognized taxonomy of 93 distinct BCTs offers a comprehensive set of evidence-based strategies to facilitate behavior change. While BCTs are often used in behavior change interventions, there is limited understanding of their application and impact within digital health interventions. Similarly, although action planning is a key BCT associated with improved outcomes in behavioral interventions [5,6], its use and impact within digital health interventions remain underexplored and underutilized.

Action planning involves creating a detailed and concrete plan for how and when a specific behavior will be performed, including details on the context, frequency, duration, and intensity of the behavior [4]. Among BCTs, action planning is particularly vital because it helps individuals translate behavior change intentions into actionable steps, effectively bridging the gap between motivation and sustained behavior change. Likewise, action planning functions as the central BCT that integrates self-monitoring, goal setting, and problem-solving by providing a structured technique for translating participants' goals into concrete plans through barrier identification and strategy selection [4,7]. Given its central role in facilitating the implementation of behavior change intentions, action planning has been consistently associated with improved outcomes across a range of health behaviors, making it a key focus for behavior change intervention design and evaluation [7,8].

A recent umbrella review of systematic reviews and meta-analyses identified the most effective BCTs at improving health outcomes among adults within digital health interventions targeting the prevention and management of noncommunicable disease. The review concluded that action planning, goal setting, feedback and monitoring, and personalization should be prioritized for inclusion in digital health interventions [9]. Yet, only 15% (13/85) of systematic reviews identified digital health intervention studies using action planning, and the included studies rarely provided sufficient detail on how action planning was implemented. This lack of use and limited reporting create a significant gap in understanding how to design and use action planning components effectively in digital health contexts. Addressing this gap in both implementation and reporting of digital action planning strategies is essential to improving the design of digital health interventions and ensuring that high-impact BCTs are used effectively to support behavior change and improve health outcomes.

Digital Health Interventions and Sugar-Sweetened Beverage Context in Rural Appalachia (United States)

As the digital divide—the gap between those with and with no reliable access to internet-connected technology [10]—continues to narrow, digital health interventions have become increasingly vital for rural communities like those in the Appalachian region of the United States [11,12]. Appalachia faces numerous barriers to accessing behavioral health interventions, including transportation challenges, limited infrastructure for health care, and socioeconomic constraints [13]. Compounding these issues are the high prevalence of sugar-sweetened beverage (SSB; eg, soda, sweet tea, sports and energy drinks, and fruit drinks) consumption and SSB-related health outcomes, including obesity and weight-related conditions [13,14]. In southwest Virginia, a part of Appalachia and the focus region of this study, SSB intake is more than double the national average and more than 4 times the recommended daily amount of 8 ounces or less per day [15,16]. Importantly, in recent years, strides have been made in bridging the digital divide as access to broadband and greater device ownership throughout Appalachia has increased [17,18]. Thus, given the relatively low cost and high scalability of digital health interventions, they offer promise in overcoming barriers to accessing behavioral interventions and improving nutrition-related health outcomes in the underserved Appalachian region. Although a recent review identified several digital behavioral interventions that incorporated action planning, none specifically targeted nutrition-related behaviors [9]. Notably, one promising feasibility study has begun to address this gap, highlighting the need for further research in this area [19].

iSIPsmarter, the program of focus of this paper, is a theory-based, digital behavioral health intervention designed to reduce SSB intake among Appalachian adults [20]. Using a randomized controlled trial (RCT), iSIPsmarter was shown to be effective in reducing SSB ounces and in improving weight compared with a static participant education comparison group [21]. Also, intervention engagement was high, and no significant relationship was found between engagement levels and outcomes. However, knowing that action planning is a vital self-regulatory BCT, providing detail and examining its role explicitly will help not only in better understanding the success of iSIPsmarter but also for others in creating digital health interventions [9].

Therefore, this study reports findings from a personalized digital action planning process, focused on SSB and weight, that is embedded within the iSIPsmarter digital behavioral health intervention. The study aims are to (1) assess the frequency of action plan engagement, (2) evaluate participants' perceived difficulty implementing their action plan and examine their progress toward achieving SSB reduction and weight goals, and (3) examine the selection of SSB- and weight-related barriers and strategies. This process evaluation corresponds to a prespecified secondary aim of the larger RCT focused on engagement with intervention components [20].

XSL•FO
RenderX

A recently published secondary analysis from the larger RCT examined the demand and implementation costs associated with iSIPsmarter's stepped care engagement approach to promoting participant engagement in the digital intervention [22]. In contrast, this manuscript focuses specifically on engagement with the intervention's action planning component. Findings are used to offer considerations for future integration of digital action planning components into behavioral interventions.

Methods

Study Design

Overview

This paper is an exploratory, longitudinal process evaluation of digital action planning embedded within the iSIPsmarter intervention. A 2-group RCT examined the efficacy of 2 internet-based programs aimed at decreasing SSB consumption [20]. Participants were randomized to receive one of two conditions: (1) iSIPsmarter (n=127), in which participants received 6 modules (ie, "Cores") with personalized action planning and behavioral tracking of SSB and weight via SMS and cellular-enabled scales, or (2) a patient education website (n=122), in which participants had access to a static education site with no action planning and optional tracking of SSB and weight via paper diaries. Participants were enrolled in this study from July 2021 to December 2022. This paper focuses on process data from iSIPsmarter participants. While most digital action planning occurred in the first 9 weeks, extending the analysis through the 6-month assessment allowed for exploration of longer-term weight-related goal achievement.

Participants

The larger 2-group RCT recruited participants from Southwest Virginia and surrounding Appalachian counties using community-engaged, multimodal recruitment strategies. Recruitment efforts were conducted in partnership with a wide range of community stakeholders, including federally qualified health centers, local health districts, hospitals, local government agencies, and higher education institutions [23]. Recruitment strategies included both passive methods (eg, print materials, emails to listservs, web page postings, and press releases) and active methods (eg, community presentations and outreach events). Interested individuals completed an online eligibility screener, and eligible participants were subsequently contacted by study staff to complete informed consent and baseline assessment procedures. To be eligible for the study, participants were required to (1) be English-speaking, (2) be aged at least 18 years or older, (3) reside in Southwest Virginia or surrounding Appalachian counties, (4) consume more than 200 calories of SSB per day as assessed by a validated beverage intake questionnaire [24], (5) not participate in another program focused on energy balance–related behaviors, and (6) be willing to access an internet-enabled computer or tablet at least 1 time per week and receive SMS text messaging–based reminders. Participants included in this analysis were those who were randomized to receive the iSIPsmarter intervention and who had completed at least 1 iSIPsmarter Action Plan (ie, 2 Cores) (119/127, 94%).

iSIPsmarter Intervention Overview

In brief, iSIPsmarter was adapted from the original evidence-based SIP*smart*ER SSB intervention, which was informed by several phases of formative research and included three 90-minute in-person small-group classes and interactive voice response calls for behavioral self-monitoring and action planning [25,26]. Furthermore, guided by the Model for Internet Interventions [27] and best practices in human-centered design [28], a robust human-centered design process was used to adapt iSIPsmarter's digital intervention structure, including the Cores, behavioral self-monitoring, and action planning [29]. The comprehensive 3-phase human-centered design process (contextual inquiry, prototype testing, and end user testing) spanned approximately 18 months. It involved a series of 13 semistructured, one-on-one interviews with 7 advisory team members from the targeted user group. Complete details of the intervention's development and content are described elsewhere [20,29].

iSIPsmarter is a self-guided digital health intervention delivered via a secure web-based platform. Participants log in to iSIPsmarter with their email address and password to access 6 behavioral content Cores. Each Core is metered out to the user, with a new Core unlocked 7 days after completion of the previous Core. Therefore, while it is possible for Cores to unlock on a weekly schedule, the Cores do not always correspond to calendar weeks as participants may take longer than 1 week to complete a Core. Core content includes various media formats of text, audio, video, interactions, animations, and vignettes focused on SSB-related topics and skill building. Core content highlights the connection between SSB and weight outcomes through personalized interactions and messaging, which helps participants recognize how SSB reduction supports weight management and overall health. The Cores also provide tailored calorie and added sugar recommendations aligned with each participant's weight-related goal. A complete list of each Core's learning objectives is described elsewhere [20]. The program integrates digital behavioral tracking through SMS prompts and a cellular-enabled scale to support daily self-monitoring of SSB intake and weight. A personalized dashboard allows participants to view and track their progress, while the digital action planning process—embedded within 5 of the 6 Cores (ie, Cores 2 through 6)—guides participants through goal setting and problem-solving or coping planning (ie, barrier identification and identification of strategies to overcome barriers) with personalized feedback.

Behavioral Self-Monitoring

Throughout the intervention, participants engage in daily behavioral self-monitoring of both SSB and weight through an integrated system. iSIPsmarter delivers daily prompts for participants to report their total SSB intake in ounces for the previous day, offering the flexibility to receive these reminders via SMS text messaging, email, or both. Two-way SMS text messaging queries allow participants to reply directly with their SSB intake, which is automatically uploaded to their online iSIPsmarter dashboard. Alternatively, participants can manually enter their intake by signing into the program online. In addition, the daily email prompt encourages participants to weigh

XSL•FO
RenderX

themselves using their cellular-enabled electronic BodyTrace scale and monitor their progress toward their weight goals. The BodyTrace scale automatically transmits weight data to the dashboard when a cellular connection is available, with the option for participants to manually log their weight online, if necessary.

Program Goals

Participants set a program goal in ounces for SSB in Core 2. Additionally, at the start of the intervention, participants select a weight-related reason for joining iSIPsmarter to either "lose weight" or "maintain weight."

Action Planning With Personalized Feedback

Throughout the program, participants can complete up to 5 personalized action plans. Action plans are embedded at the end of Cores 2 through 6. Thus, participants must complete an action plan in order to complete the Core. The iSIPsmarter action planning process was designed to provide automated, tailored behavioral support to participants to self-monitor and reduce their SSB intake and manage weight behaviors. During this structured process, participants set personal goals and identify barriers to making behavioral changes and strategies to overcome the barriers. Figure 1 provides a step-by-step overview of the action planning process.

As shown in Figure 1, the first step in the action planning process includes feedback on participants' SSB behaviors and weight through a summary of tracking completed during the previous 7 days (step 1). Participants then receive personalized feedback on SSB and weight data (step 2), including progress toward SSB and weight goals, total SSB and weight change, comparisons to SSB and weight recommendations, and total number of days tracked. Participants are required to have at least 3 days of data to receive personalized feedback. Those with less than 3 days of data are prompted to manually enter data for up to the previous 7 days. Next, participants are encouraged to reflect on factors influencing SSB behaviors (ie, select a top reason for their highest and lowest SSB days), before receiving personalized feedback on their SSB patterns (step 3). Then, participants are encouraged to set a weekly goal to decrease their SSB in ounces by 20% - 50% from their previous week's average (step 4); however, participants can choose a goal outside of the recommended range. This reduction target is repeated in subsequent action plans until the participant meets the recommendation of less than 8 ounces of SSB per day.

Participants' achievement of their weekly SSB goal is assessed using their average intake during the 7 days preceding each Core. A participant's goal is met if the average SSB ounces are less than or equal to the goal amount set in the prior Core. Some progress is achieved if the average SSB ounces are greater than the goal amount set in the prior Core but less than the average SSB ounces from the prior Core. No progress is achieved if the average SSB ounces are greater than the goal amount and greater than or equal to the average SSB ounces from the prior Core. If participants have less than 3 days of SSB diary data, they receive feedback that there are not enough data to determine progress toward their goal.

After setting their goals, participants are guided through a structured problem-solving process (step 5). They begin by selecting 2 SSB-specific barriers from a prepopulated list, followed by identifying 2 strategies per barrier to help overcome these challenges. To promote behavior change intentions, participants are prompted to complete the statement, "To help me reach my goal this week, I will..." and choose from a list of specific strategies mapped to each barrier.

The prepopulated barriers and strategies were intentionally designed to be specific and actionable, with many strategies incorporating key BCTs such as context cues and habit formation. The barrier and strategy lists were developed and tested through formative research to inform the original SIPsmartER intervention [26] and iterative end user testing to ensure that they were both culturally relevant and feasible for the intended population [29]. Barriers and strategies are prepopulated in drop-down menus to streamline decision-making and reduce participant burden, a critical consideration in internet-delivered interventions.

Participants also have the option to enter their own strategy if none of the listed options align with their needs, which prompts a free-text response field for personalized entry. In subsequent action plans, participants are shown their previously selected barriers and strategies and can choose to either maintain or revise them to reflect evolving challenges or preferences.

XSL•FO
RenderX

Figure 1. iSIPsmarter digital action planning process. SSB: sugar-sweetened beverage.

XSL•FO
RenderX

After participants select SSB-related barriers and strategies, they are offered the choice to select weight-related barriers and strategies or skip them if they want to focus exclusively on SSB. The weight action plan mirrors the SSB action planning process, in that participants identify 2 weight-related barriers and 2 corresponding strategies per barrier (step 6). Either skipping or selecting weight barriers and strategies concludes the action planning process for the Core. Thus, participants' final action plan includes up to 4 self-selected barriers and up to 8 strategies to work on for the upcoming week. Once complete, the action plan is locked and cannot be changed (step 7). Participants can access their completed action plans from their website dashboard, and they can be printed as a PDF document.

In subsequent Cores, participants are prompted to rate how easy or hard it was for them to stick to the action plan they set in the previous Core. Tailored feedback is provided based on their response. For example, participants who report that their action plan was impossible or hard to follow are encouraged to choose smaller weekly goals and to select different strategies to support their success.

Measures

Overview

Demographic data were self-reported at baseline via an online questionnaire. Participants' zip codes and the rural-urban continuum codes (RUCC) were used to determine rurality status. The codes range from 1 to 9, with higher values indicating more rural areas [30]. Process data for each participant, including action planning goals and goal achievement, perceived action planning difficulty, barrier and strategy selections, SMS text messaging–based and manual SSB tracking entries, and weight data transmitted through cellular-enabled scales or manually entered online, were automatically captured through the iSIPsmarter web-based administrative platform. Baseline and 6-month follow-up data were exported from the administrative platform for analysis.

SSB Action Planning

Participants' action planning data included self-selected personal reasons to change, summaries of SSB behavioral self-monitoring, a program SSB goal, and weekly SSB goals (all SSB summaries and goals were in ounces).

Participants' Perceived Difficulty Implementing Their Action Plan

Participants' perceived difficulty implementing their action plan was assessed at the start of Cores 3 - 6. Perceived difficulty was not assessed in Core 2 because participants did not complete an action plan in Core 1. The question, "How hard or easy was it to stick to your action plan this past week?" was rated on a 5-point Likert scale ranging from (0) "impossible" to (4) "very easy."

SSB Goal Achievement

Participants received feedback on their weekly goals, which included (1) goal achieved, (2) some progress toward the goal, (3) no progress toward the goal, and (4) not enough diary data to provide feedback.

SSB-Related Barriers

Data included participants' self-selected SSB barriers for each completed action plan. For each action plan, participants selected 2 barriers from a list of 14 SSB-related barriers. See the Results section for the complete list.

SSB-Related Strategies

Within each SSB barrier, participants selected 2 strategies. Strategies were mapped to barriers and ranged from a total of 4 - 9 per barrier, with a total of 80 possible strategies.

Weight Action Planning

Participant weight was measured via a cellular-enabled BodyTrace scale. Participants' weight action planning data included a program weight goal (ie, lose or maintain weight), which was set in Core 1.

Weight Goal Achievement

Participants' percent weight change was assessed from baseline to 9-week and baseline to 6-month follow-up assessments. The program weight loss goal was achieved if the participants lost at least 1 pound. The program weight maintenance goal was achieved if the participants' weight change was less than 1 pound [31].

Weight-Related Barriers

Participants selected 2 weight-related barriers from a list of 13. See the Results section for the complete list.

Weight-Related Strategies

Participants selected 2 weight-related strategies per barrier. Possible strategies to choose from for each barrier ranged from 8 to 13, with a total of 115 possible strategies.

Data Analysis

Data analysis was conducted using SPSS (version 29.0; IBM Corp). Descriptive statistics were used to calculate frequencies and percentages at each Core for action plan completion, perceived difficulty implementing action plans, and SSB and weight goal achievement. SSB and weight barrier and strategy frequencies were aggregated across all completed action plans.

Ethical Considerations

This study was approved by the University of Virginia Institutional Review Board (IRB-HSR 22130) and was also registered with the ClinicalTrials.gov registry: NCT05030753. All participants provided written informed consent prior to enrollment. Participants were allowed to withdraw from the study at any time during the research process. All data used in this analysis were deidentified to ensure participant privacy and confidentiality. Participants received compensation in the form of a cellular-enabled scale at baseline and gift cards (US $50) at the follow-up assessments.

Results

Sample and Demographics

Of the 127 enrolled iSIPsmarter participants, 119 (94%) participants completed at least 1 action plan and are included in this analysis. As shown in Table 1, most participants were

XSL•FO
RenderX

Caucasian, female, and from more rural counties (ie, RUCC 3=54, 45%; RUCC 4 - 9=52, 44%), with an average age of 42 (SD 11.9) years. Just over half were college educated and nearly half had household incomes less than US $55,000 per year. SSB intake at baseline averaged 43.2 (SD 23.6) fluid ounces. Notably, 27 (23%) of participants were overweight and 74 (62%) had obesity.

Table . Baseline characteristics of iSIPsmarter intervention participants (N=119).

Baseline characteristics	Participants
Race, n (%)	
White (Caucasian)	106 (89.1)
African American	8 (6.7)
>1 race	2 (1.7)
Other	3 (3)
Ethnicity, n (%)	
Hispanic/Latino and Latina	3 (2)
Sex, n (%)	
Male	23 (19.3)
Female	96 (80.7)
Age (years), mean (SD)	42 (11.9)
Age (years), n (%)	
18 - 24	7 (5.9)
25 - 44	56 (47.1)
45 - 64	53 (44.5)
≥65	2 (1.7)
Education, n (%)	
High school graduate or less	13 (10.9)
Some college	40 (33.6)
College graduate	27 (22.7)
Graduate school	39 (32.8)
Annual household income (US $), n (%)	
14,999 or less	8 (6.8)
15,000 - $34,999	24 (20.5)
35,000 - $54,999	25 (21.4)
≥55,000	60 (51.2)
Rural-Urban Continuum Code, n (%)	
1 - 2 (medium-large metro)	13 (10.9)
3 (small metro)	54 (45.4)
4 - 9 (nonmetro)	52 (43.7)
SSB[a], mean (SD)	
SSB intake, fluid ounces	43.2 (23.6)
BMI, n (%)	
Underweight	1 (0.8)
Normal	17 (14.3)
Overweight	27 (22.7)
Obesity	74 (62.2)

[a]SSB: sugar-sweetened beverage.

XSL•FO
RenderX

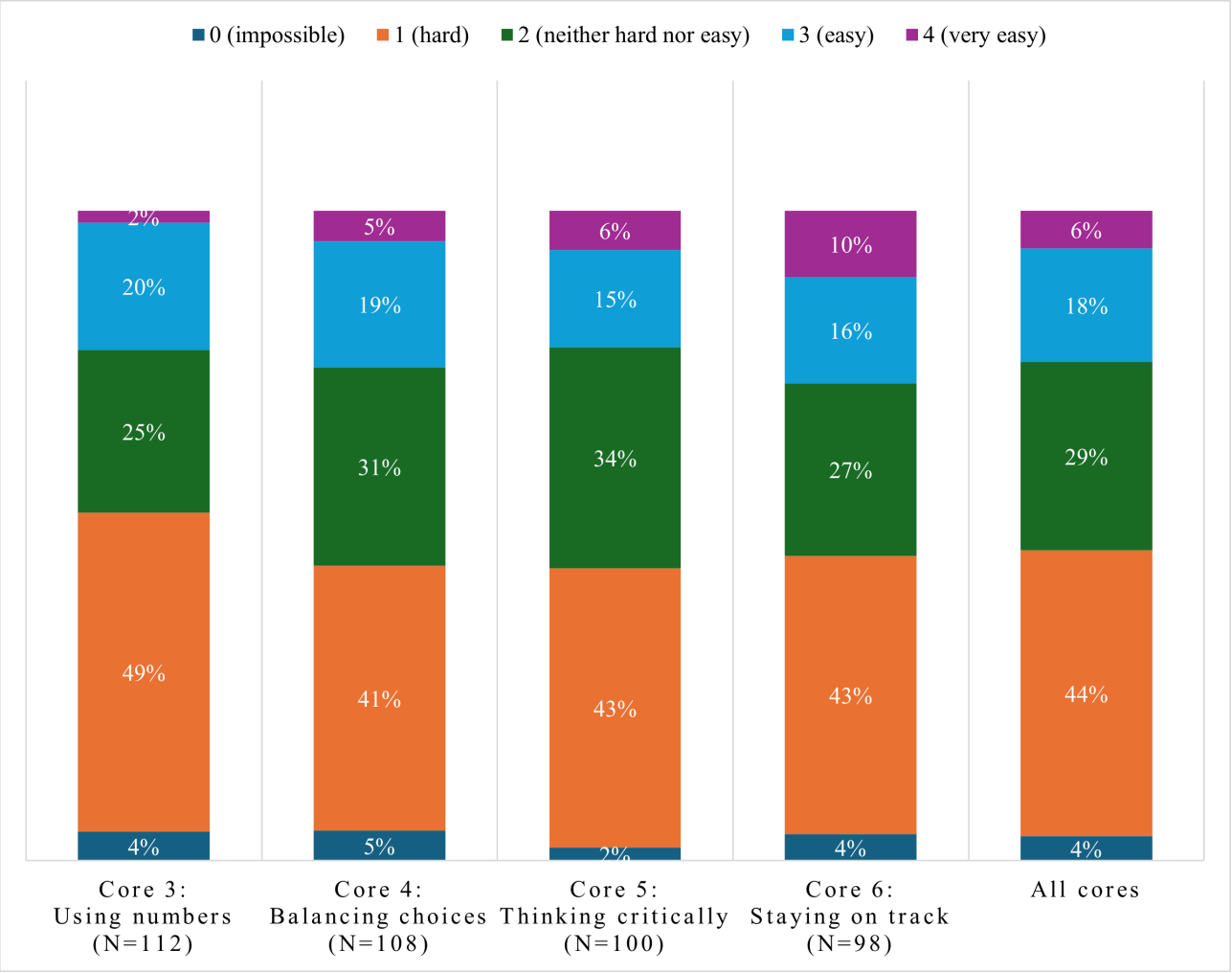
SSB Action Plan Engagement

Participants completed on average 4.5 (SD 1.1) out of 5 SSB action plans. Furthermore, of the 119 participants who completed at least 1 action plan (Core 2), 95 (80%) completed all 5 (Cores 2 - 6). Action plan completion declined slightly over time, with 113 (95%) participants completing Core 3, 106 (89%) completing Core 4, and 101 (85%) completing Core 5. This pattern mirrors the overall decline in Core completion across the intervention.

Participants' Perceived Difficulty Implementing Their Action Plan

As shown in Figure 2, across all Cores, participants' perceived difficulty implementing their action plan remained relatively stable. On average, 48% rated the tasks as impossible or hard, 29% as neither hard nor easy, and 24% as easy or very easy.

Figure 2. Participants' perceived difficulty implementing their action plan.

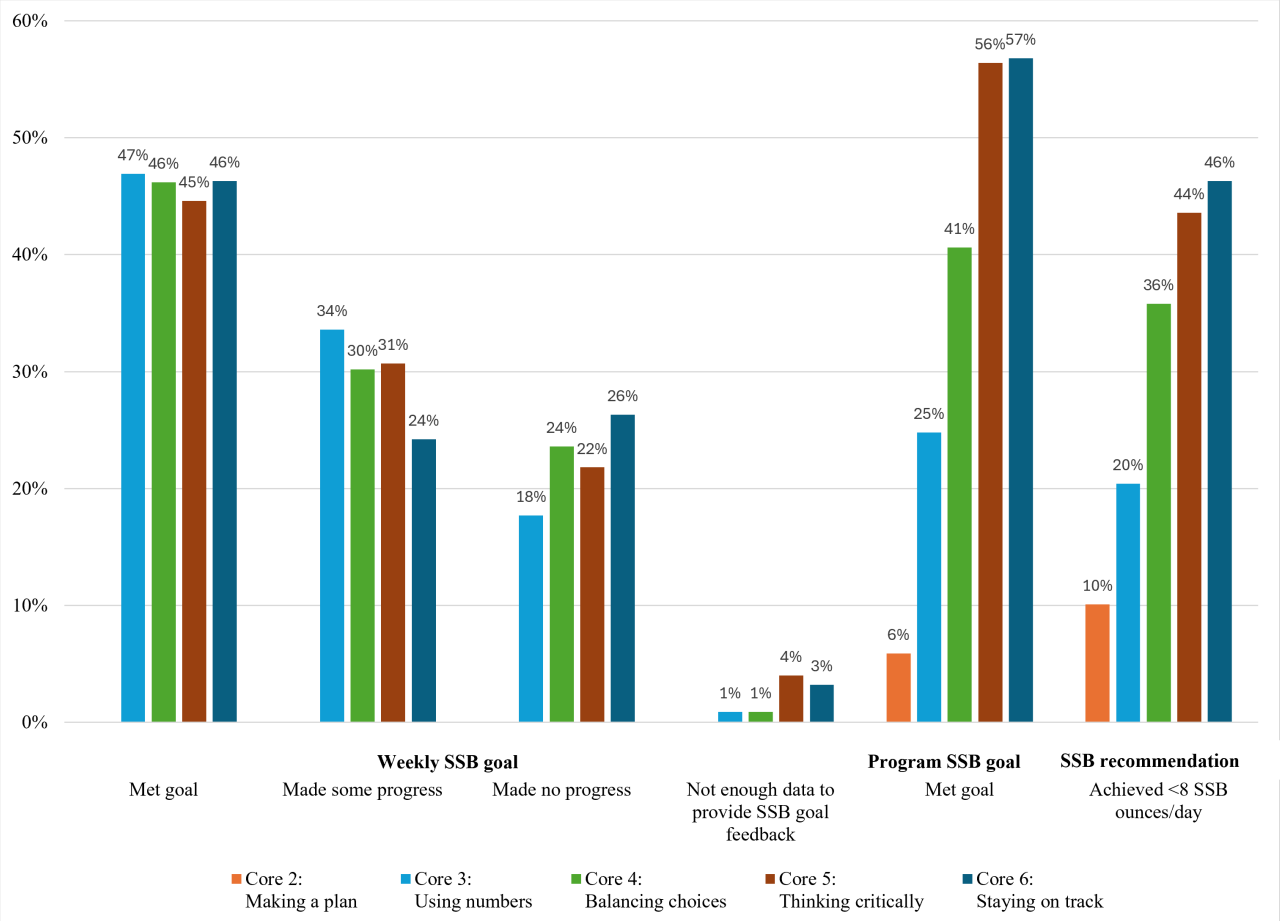


SSB Goal Achievement

Across each of the Cores, Figure 3 illustrates participants' progress toward three SSB goals: (1) their weekly goal, (2) their program goal, and (3) achievement of <8 ounces of SSB per day recommendation. Regarding weekly SSB goals, the percentage of participants meeting their goals remained relatively stable across the Cores, averaging about 45%. Also, about one-third of participants made some progress toward their weekly goals at each Core. Meanwhile, there was a steady increase in the proportion of participants who reduced their daily SSB intake to achieve the recommendation of <8 ounces per day. At the end of Core 6, 57% (54/95) of participants met their program goal and 46% (44/95) achieved the recommendation of <8 ounces of SSB per day.

XSL•FO
RenderX

Figure 3. Participants' progress in achieving SSB goals and recommendations. SSB: sugar-sweetened beverage.



SSB Barrier and Strategy Selection

As shown in Table 2, 45% (53/119) of participants selected different SSB barriers at least once during the action planning process, and 53% (63/119) of participants revised their chosen strategies at least once. On average, participants selected 2.8 (SD 1.1) unique barriers and 6.4 (SD 2.9) unique strategies. As shown in Table 3, the most common reported barriers to reducing SSB were (1) needing caffeine from SSB, (2) not wanting to give up the taste of SSB, (3) SSB being a habit, (4) not liking the taste of no-calorie sugar substitutes, and (5) needing energy from SSB. The top five reported strategies to reducing SSB were as follows: (1) reduce my caffeinated SSB slowly to avoid caffeine withdrawal, (2) replace sugary caffeinated drinks with nonsugary caffeinated drinks, (3) get more sleep to feel less tired and dependent on caffeinated SSB, (4) slowly reduce SSB to tame my sweet tooth, and (5) use the iSIPsmarter Diary to track SSB and learn my habit.

Table . Summary of SSB[a]-related and weight-related barriers and strategies.

	SSB-related barriers and strategies (n=119)	Weight-related barriers and strategies (n=112)
Participants who changed barriers, n (%)	53 (45)	69 (62)
Number of unique barriers selected, mean (SD) (possible range 2 - 10)	2.8 (1.1)	3.0 (1.0)
Participants who changed strategies, n (%)	63 (53)	82 (73)
Number of strategies selected, mean (SD) (possible range 4 - 20)	6.4 (2.9)	6.6 (2.7)

[a]SSB: sugar-sweetened beverage.

XSL•FO
RenderX

Table . Frequency of iSIPsmarter SSB[a]-related barriers selected across all completed iSIPsmarter action plans and participants.

Barrier	Frequency
Need the caffeine from SSB	212
Do not want to give up the taste of SSB	193
SSBs are a habit	139
Do not like the taste of no-calorie sugar substitutes	98
Need the energy from SSB	97
Drink SSB with family or friends	88
Do not like the taste of water	88
Have too many SSBs around	45
Concerns about no-calorie sugar substitutes	44
Hard to choose smaller sizes	43
Buy SSB because they are low-cost	14
Concerns about water safety	4
Persuaded by advertising	3
Confused by the nutrition facts label	1
Total	1069

[a]SSB: sugar-sweetened beverage.

Weight Action Plan Engagement

Participants completed on average 3.3 (SD 1.5) weight action plans, with 94% (112/119) of participants completing at least 1 weight action plan. Engagement in weight action planning varied, as 6% (7/119) of participants skipped weight action planning, 7% (8/119) completed 1 weight action plan, 24% (29/119) completed 2 weight action plans, 7% (8/119) completed 3 weight action plans, 29% (34/119) completed 4 weight action plans, and 28% (33/119) completed all 5 weight action plans.

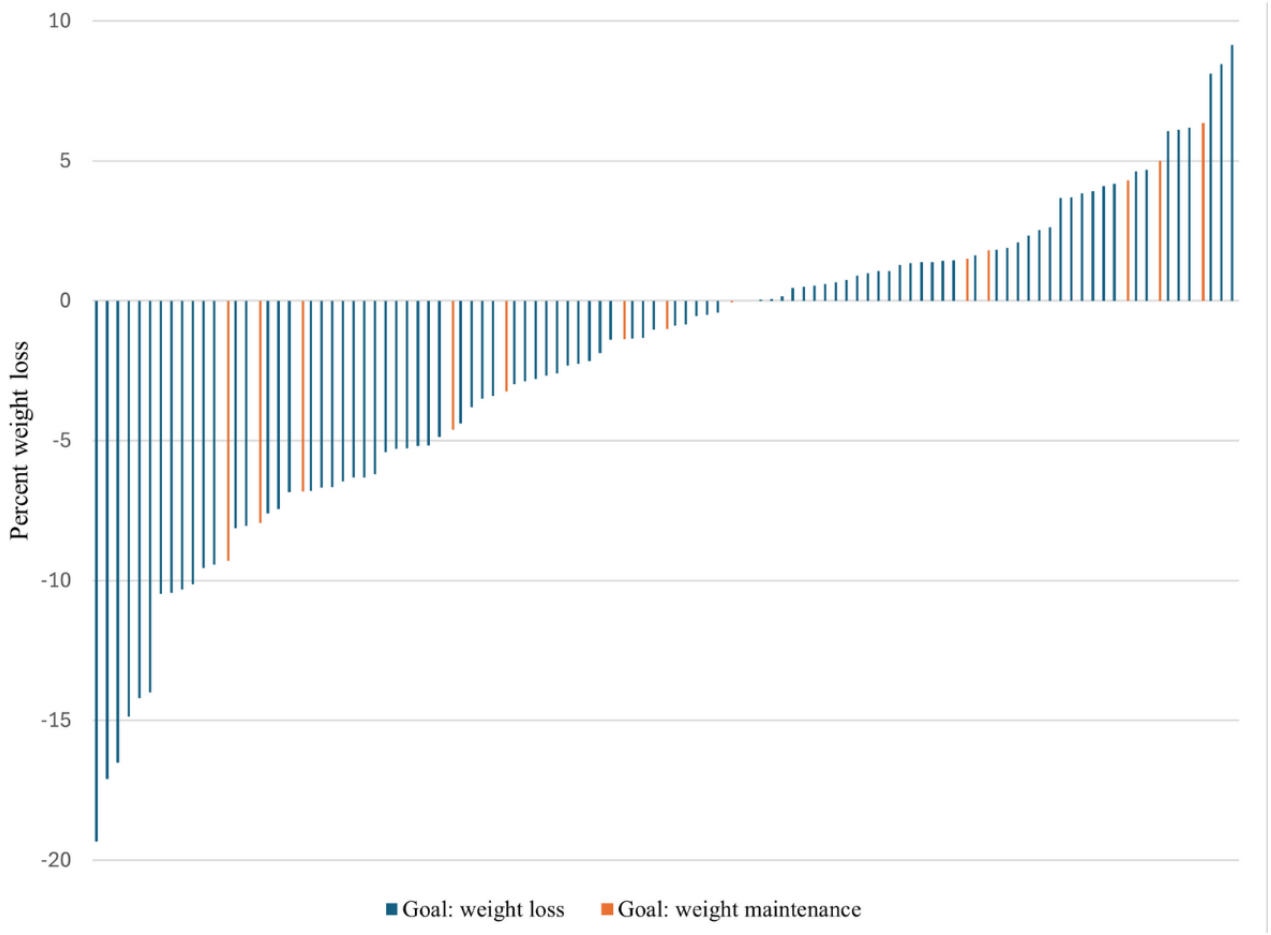
Weight Goal Achievement

At the beginning of the intervention, 86.6% (103/119) of participants set a weight goal to lose weight. Among these participants with follow-up weight data, the average weight loss was −1.3% (SD 2.6) at the 9-week follow-up, with 61% (57/94) achieving their goal. By 6 months, weight loss increased to −2.1% (SD 5.6), with 54% (49/90) achieving their goal (Figure 4).

The other participants (16/119, 13.4%) set a goal to maintain weight during the intervention. Among these participants with follow-up weight data, 7% (1/14) achieved their goal at the 9-week follow-up, and 8% (1/13) achieved their goal at the 6-month follow-up. At 6 months, the average weight change was −1.2% (SD 5), with most participants losing, rather than maintaining weight (Figure 4).

XSL•FO
RenderX

Figure 4. Participants' progress in achieving program weight goal at 6 months.



Weight Barrier and Strategy Selection

As illustrated in Table 2, 62% (69/112) of participants revised their weight barriers throughout their action planning process, and 73% (82/112) of participants updated their weight-related strategies. As shown in Table 4, the most common reported weight-related barriers were (1) enjoying sweets, (2) struggling with portion sizes, (3) being too busy or tired to eat healthy foods, (4) having difficulty staying motivated, and (5) snacking throughout the day. The top five reported weight-related strategies were as follows: (1) limit or stop buying sweets at the grocery or convenience store; (2) preportion my serving when I have sweets, and not eat straight from the package; (3) keep sweets out of sight or in a hard to reach area at home and at work; (4) replace sweet treats with fruit, such as berries, melon, or apples; and (5) ask myself if I really need a second helping and wait 20 minutes before deciding.

XSL•FO
RenderX

Table . Frequency of iSIPsmarter weight-related barriers selected across all completed iSIPsmarter action plans and participants.

Barrier	Frequency
Enjoy sweets	198
Struggle with portion sizes	101
Too busy or tired to prepare healthy foods	87
Eat out often	86
Hard time staying motivated	61
Snack throughout day	59
Enjoy high-fat dairy	51
Enjoy high-fat meats	43
Enjoy salty snacks	36
Healthy foods cost a lot of money	27
Hard time eating fruits and vegetables	23
Drink calories through alcohol	9
Do not like the taste of healthy foods	1
Total	782

Discussion

Overview

This is one of the first studies to thoroughly explore a digital action planning process and to conduct a comprehensive process evaluation of a nutrition-focused digital behavioral health intervention [3,9,26,32]. In contrast to existing digital interventions that mention action planning superficially or as a single feature, the iSIPsmarter intervention incorporates it as a multistep, iterative process aligned with behavioral theory and delivered through personalized, user-guided digital Cores. This level of integration and process evaluation contributes to the much-needed understanding of how key BCTs are operationalized in digital formats. Findings from our study contribute to the growing body of literature on digital behavioral health interventions, emphasizing the value of structured and personalized goal-setting and action planning strategies in supporting self-regulatory behaviors, such as SSB intake and weight management.

Principal Results

SSB Action Planning

Findings from this study highlight the usefulness of embedding a low-cost, scalable action planning process within a digital behavioral health intervention. The iSIPsmarter action planning component provided automated, tailored support to help participants monitor their SSB intake, set goals, identify barriers, and select personalized strategies. Engagement was high, with 80% (95/119) of participants completing all 5 action plans. iSIPsmarter's action planning process allowed the user to create new action plans or repeat previously made plans to reinforce them. High engagement across all 5 action plans aligns with recent research indicating that the number of action plans is positively associated with improved health behaviors, with the strongest associations observed when individuals create 5 plans [7]. Furthermore, high engagement with iSIPsmarter's action planning process underscores the program's usability and key design features, which were adapted and developed from our prior efficacious SSB intervention [26] and using best practices in human-centered design [28,29]. Findings from this study are especially meaningful, given its rural context, in which rural populations often encounter persistent barriers to accessing face-to-face behavioral health interventions. This suggests that well-designed, digitally delivered behavior change interventions can extend the reach of behavioral programs to underserved communities.

Participants' Perceived Difficulty Implementing Their Action Plan

Reducing SSB intake is a complex and difficult behavior to change. It is often embedded in daily routines and driven by habit, preference, and societal norms. Within this study, participants' perceived difficulty in implementing their weekly action plans reflected this challenge. Across much of the program (Cores 3 - 6), participants rated implementing their action plan "hard" most often, with only a small number of participants rating their plans "impossible" or "very easy." These difficulty ratings closely mirrored goal achievement patterns. Importantly, those who rated their plans as difficult continued to engage in the action planning process. This persistence speaks to the value participants placed on the intervention and its relevance to their goals. This suggests that well-designed digital behavior change interventions can sustain engagement and support progress, even when the target behavior is difficult. Future research should use data capturing participants' perceived difficulty to provide real-time tailoring, with added support for those consistently reporting difficulty sticking to their action plan.

SSB Goal Achievement

Our study findings illustrate an encouraging trend in participants' progress toward reducing SSB intake. While the proportion of participants meeting their personalized weekly

XSL•FO
RenderX

SSB goal was stable at about 46% across the action planning process, an additional 33% made some progress toward their weekly goals. Together, 79% of participants made progress each week, which reflects a high level of progress, especially given the complexities of changing SSB behaviors. Goal attainment rates align with other behavioral nutrition interventions, where about half of the participants meet their goals [33].

Furthermore, there was a notable, steady increase in those reducing their SSB intake to the recommended less than 8 ounces. By the final action plan, nearly half of the participants had met this recommendation. Although nationally representative data typically report SSB intake by frequency or total added sugars intake, available estimates suggest that on a given day, US adults consume approximately 145 calories from SSB [34], which is approximately one 12-ounce serving of soda. Thus, very few US adults meet the less than 8-ounce recommendation. The achievement of this recommendation among iSIPsmarter participants is especially encouraging, given that participants' average SSB intake at the start of the intervention was approximately 43 ounces per day, equivalent to more than 3½ cans of soda.

SSB-Related Barriers and Strategies

Over half of the participants maintained the same 2 SSB-related barriers throughout the action planning process, indicating consistency in perceived challenges. The most frequently cited barriers (ie, the need for caffeine, reluctance to give up the taste of SSBs, and the habitual nature of SSB consumption) are consistent with findings from prior SSB-related literature across diverse populations [35,36]. However, over half of the participants modified their strategies associated with these goals at least once during the action planning process. This adjustment in strategy selection may reflect participants' efforts to identify the most effective approaches for overcoming barriers, which are likely to change over time. This finding highlights the importance of flexibility in goal setting and troubleshooting barriers through repeated action planning. Moreover, research has shown that counterhabitual actions are important for intentions to be effective in breaking strong habits [7]. In the case of iSIPsmarter, strategies were tailored to barriers and were structured so that they could facilitate the deconstruction of ingrained SSB consumption barriers.

The most frequently selected SSB strategies–such as gradually reducing caffeine intake, substituting nonsugary caffeinated drinks, improving sleep to reduce SSB caffeine dependency, and using tracking tools such as the iSIPsmarter Diary—highlight the importance of providing users with multiple strategies to reduce SSB consumption. These strategies target both the physical drivers of SSB consumption, such as caffeine dependence and taste preferences, and behavioral factors, such as habit formation and building self-awareness through behavioral self-monitoring. These strategies have been shown to be key BCTs in behavioral health interventions focused on improving dietary behaviors [37].

While some research has emphasized the potential benefits of self-constructed plans, this approach can introduce participant burden, especially in digital settings without human-delivered support. To balance personalization with usability, iSIPsmarter offered an evidence-informed list of barriers and strategies developed through prior evidence-based research and rigorously tested using a user-centered design approach [26,29]. Furthermore, a free-text "use my own strategy" option was available but was used infrequently, suggesting that most participants preferred and were able to identify their strategies using the structured list. This design may reduce cognitive load while still allowing for personalization, particularly among populations with limited health literacy or digital experience. Developing and testing intervention components, such as barriers and strategies, from the end user's perspective is recognized as a critical element in the design and implementation of digital behavioral interventions [38].

Weight Action Planning

The varying levels of participant engagement with the weight action planning component provide insights into how optional aspects of an intervention may influence participation. While the majority (112/119, 94%) of the participants completed at least 1 weight action plan, fewer (33/119, 28%) completed all 5 weight action plans. Although 87% (103/119) of the participants identified weight loss as an initial goal, engagement with weight action planning diminished over time. Reasons for this decline in weight action planning are not entirely clear, yet it is plausible that some participants may have experienced weight-related information avoidance or negative emotions with weight self-monitoring and action planning [39,40]. Furthermore, the weight action planning patterns may suggest that participants were selective, possibly prioritizing the SSB-focused components of the intervention over the weight-specific ones. The ability to choose between SSB-related and weight-related barriers likely allowed for a more customized approach but may have also led participants to limit engagement with optional elements when those did not align as closely with their perceived barriers. This selective engagement reflects the need for flexibility in interventions, as well as for further examination of how weight self-monitoring can best be structured to support engagement over time. Given that 85% (101/119) of the enrolled iSIPsmarter participants were with overweight or obesity, compared with the estimated national prevalence of 74%, the need for future research efforts to promote engagement in weight action planning is especially critical in Appalachia and other regions facing similar weight disparities [41].

Weight Goal Achievement

Participants' weight-related goal achievement underscores iSIPsmarter's success in supporting weight loss goals. Over half of the participants who set a weight loss goal successfully met it, demonstrating the intervention's potential to support meaningful behavior change. By 6 months, the average percent weight loss was nearly double among those with a weight loss goal compared with those with a weight maintenance goal. Additional opportunities for future research are to explore the relationship between engagement with weight action planning components and the achievement of weight-related goals.

XSL•FO
RenderX

Weight-Related Barriers and Strategies

The high proportion of participants who changed their weight-related barriers and strategies illustrates a dynamic process of problem-solving and coping planning. In contrast to SSB-related barriers, participants appeared to face more variable challenges related to weight. The most common weight-related barriers—enjoying sweets, struggling with portion sizes, and being too busy or tired to eat healthy foods—reflect common and complex challenges in weight management [42,43]. The top barrier of sweets reveals participants' affinity for all sugar and should be considered for future SSB interventions incorporating nutrition-related content. Frequent updates to weight-related strategies suggest that participants engaged in action planning as an ongoing, iterative process rather than a one-time exercise. Notably, most participants selected new strategies at least once, indicating that repeated action planning may promote more flexible and personalized approaches to overcoming barriers. This pattern of strategy modification may reflect participants' efforts to fine-tune their weight-related behaviors through trial and error, adapting their approaches in response to evolving challenges and contexts, as well as instances where previous strategies were unsuccessful. The most reported strategies, such as limiting or stopping the purchase of sweets, preportioning servings, and replacing sweet treats with fruit, are practical and actionable steps that participants can take to modify their eating habits. The emphasis on portion control and environmental modifications (eg, keeping sweets out of sight) aligns with behavioral strategies to support sustainable weight management [44].

Comparisons With Prior Work

While prior research has established the importance of action planning as a BCT, detailed reporting on how action planning is operationalized and reported across digital interventions remains limited. Many studies reference action planning without describing how plans are structured, how frequently they are completed, or the extent to which plans are personalized and adapted over time [3,32,45]. This limits the ability to compare digital health interventions and understand which aspects of digital action planning best support behavior change.

A recent umbrella review underscores this gap in reporting. Although a small subset (ie, 41/388, 11%) of digital health intervention studies used action planning, descriptions of action plan completion varied widely, and details regarding the number of action plans created were often overlooked [9]. Among those who did report frequency, more than half relied on a single action plan throughout the intervention period, despite evidence that goal setting and action planning are most effective when implemented at multiple time points and reevaluated with personalized feedback [46].

Our study helps address these gaps by providing detailed process data on a multitimepoint, digitally delivered action planning process within a nutrition-focused intervention. Without this level of detail in the literature, it is difficult to replicate successful digital action planning strategies across health domains and populations. Moreover, although evidence supports the effectiveness of digital action planning across multiple health domains, there is a dearth of literature on digital action planning in nutrition-related behavioral interventions [9].

Limitations

Several limitations should be considered when interpreting study findings. First, generalizability may be limited by the study population and targeted rural Appalachian region. Second, reliance on self-reported SSB intake data may introduce recall bias, potentially affecting the accuracy of reported consumption levels. Third, due to the exploratory nature of this paper and the absence of hypothesis testing, only descriptive statistics were used, which may limit interpretations. Although there may be interesting questions regarding associations among the reported descriptive variables, this study was not designed or powered to test inferential relationships, and attempting to do so could introduce analytical error and interpretations. Fourth, while participants rated their perceived difficulty sticking to their action plans, they did not explicitly rate whether they executed their action plan. This is an important area of future research. Finally, while the study reports on engagement with action planning, it does not explore user perceptions or qualitative feedback on the digital action planning experience. Perceptions of participants' experience with the iSIPsmarter intervention, including barriers to adherence, usefulness, and perceived impact, are explored elsewhere [47]. These limitations should be considered in the context of the study's strengths, which include (1) leveraging a digital health intervention to gather rich, longitudinal data on participants' engagement with digital action planning, (2) using digital technologies to collect SSB and weight data for self-monitoring and personalized feedback, and (3) the inclusion of a comprehensive set of prepopulated barriers and strategies informed by prior research.

Future Research

Our study highlights several important areas for future research. First, future studies should capture adherence to selected action plan strategies, perhaps through ecological momentary assessment methods [48]. Ecological momentary assessment could provide valuable real-time insights into when and why action plans are or are not enacted and provide opportunities to support implementation intentions. Second, future research should examine the optimal number of action plans, barriers, and strategies, as well as the impact of prepopulated versus open-ended (write-in) options on participant engagement and behavior change. These studies could help determine the best balance between minimizing participant burden and maximizing intervention effectiveness. Third, researchers should investigate the long-term effects of digital action planning on maintaining behavior changes beyond the intervention period. Fourth, future research should evaluate the scalability and adaptability of digital action planning strategies in diverse populations, including those with varying socioeconomic, cultural, and geographic contexts. Such efforts will be critical for optimizing digital action planning strategies and maximizing their potential to support sustainable health behavior change across diverse settings. Additional considerations for digital action planning in behavioral health interventions are provided in Textbox 1.

XSL•FO
RenderX

Textbox 1. Considerations for digital action planning in behavioral health interventions.

Although action planning has been identified as a key behavior change technique, there are a limited number of studies using digital action planning in digital health interventions. Moreover, few studies report on key design features or key process data from action planning. As such, we offer four considerations for developing digital action plans for digital health interventions:

1. *Leverage formative data and human-centered design processes to inform the development of digital action planning components*: Integrating formative data (eg, qualitative interviews or focus groups with intended users) with human-centered design processes can support the development of structured, preprogrammed action planning components that are relevant, realistic, and accessible to the intended users. In digital interventions, this approach helps balance personalization with usability and scalability, reduces participant burden, and supports sustained engagement by ensuring that action planning processes are culturally relevant and responsive to user needs.
2. *Capture and transparently report process-level engagement data*: Descriptive data on action plan utilization offer valuable context for interpreting participant engagement with an intervention. Furthermore, reporting participants' perceived difficulty, goal progress, and strategy modification can help researchers understand how individuals interact with action planning components over time and support replication of successful digital action planning processes.
3. *Digital action planning strategies should be personalized and incorporate feedback on self-monitoring behaviors*: Personalized feedback on self-monitoring behaviors provides a tailored approach to supporting users in achieving and sustaining behavior change goals. This approach can enhance motivation and engagement and reduce barriers to making changes. Digital action plans that tailor feedback to the user provide personalized insights into strategy modifications needed to maintain or improve behaviors, ultimately leading to more successful and sustainable behavior change outcomes.
4. *Optimize data collection to align with the intervention's budget and constraints*: Digital action planning offers unique opportunities to not only collect and analyze individual-level process data but also structure interventions in novel ways, such as adaptive interventions and those with personalized, tailored feedback. Ultimately, the budget, scope, and technical capabilities will inform how digital action planning can be structured and how much data can be collected. Researchers and interventionists should consider the cost-benefit of developing action planning systems that can be replicated across multiple health behaviors and behavioral interventions.

Conclusions

This research provides a detailed process evaluation of digital action planning as a core BCT within a nutrition-focused digital behavioral health intervention. Specifically, the findings highlight the usefulness and strong participant engagement with a structured, digitally personalized goal setting and action planning process. Participants' high utilization of action plans, frequent strategy updates, and steady progress toward SSB and weight goals suggest that action planning was useful. This is especially relevant for underserved populations where access to preventive health care is limited. By embedding action planning alongside other self-regulatory BCTs—such as goal setting, self-monitoring, and personalized feedback—iSIPsmarter created a cohesive digital behavior change framework that allowed participants to adapt their SSB and weight-related strategies over time. These findings reinforce the importance of designing digital action planning components that are user-centered and responsive to individual needs and contexts.

As digital behavioral health interventions continue to expand, this study underscores the value of elevating action planning from a passive or single-use BCT to a central, iterative component that drives engagement and behavior change. Researchers and interventionists interested in digital action planning can use findings from this study and considerations for digital action planning to integrate effective action planning components into their digital behavioral health interventions. Although focused on a nutrition behavior, it may provide insights into digital action planning strategies that can be applied across various health domains and behaviors. Overall, this study contributes to the growing body of literature on digital action planning as a key BCT and highlights its potential for promoting health behavior change in underserved populations.

Acknowledgments

The authors thank the community partners who supported recruitment efforts for this study. The authors are also grateful to their colleagues at the University of Virginia's Community Cancer Control and Obesity Research Center for their assistance with participant management, and to the Center for Behavioral Health and Technology for their program development and technical support. Finally, the authors sincerely thank the participants whose time and involvement made this study possible. No generative artificial intelligence tools were used in the manuscript writing process.

Funding

This study was funded by the National Institutes of Health, National Institute on Minority Health and Health Disparities (R01MD015033). The National Institutes of Health was not involved in the design of this study or writing of this manuscript.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

JZ, LR, and KP were responsible for the acquisition of funding. AR and JZ prepared the first draft of the manuscript. All authors provided critical review, editing, and approval of the manuscript.

Conflicts of Interest

None declared.

References

1. Dugas M, Gao GG, Agarwal R. Unpacking mHealth interventions: a systematic review of behavior change techniques used in randomized controlled trials assessing mHealth effectiveness. Digit Health 2020;6:2055207620905411. [doi: 10.1177/2055207620905411] [Medline: 32128233]
2. Thomas Craig KJ, Morgan LC, Chen CH, et al. Systematic review of context-aware digital behavior change interventions to improve health. Transl Behav Med 2020 Oct 21;11(5):1037-1048. [doi: 10.1093/tbm/ibaa099]
3. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4. [doi: 10.2196/jmir.1376] [Medline: 20164043]
4. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Mar 20;46(1):81-95. [doi: 10.1007/s12160-013-9486-6]
5. Rhodes RE, Grant S, de Bruijn GJ. Planning and implementation intention interventions. In: Hagger MS, Cameron LD, Hamilton K, Hankonen N, Lintunen T, editors. The Handbook of Behavior Change: Cambridge University Press; 2020:572-585. [doi: 10.1017/9781108677318.039]
6. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
7. Hagger MS, Luszczynska A. Implementation intention and action planning interventions in health contexts: state of the research and proposals for the way forward. Appl Psychol 2013 Oct 8;6(1):1-47. [doi: 10.1111/aphw.12017]
8. Lin H, Xu D, Yang M, et al. Behaviour change techniques that constitute effective planning interventions to improve physical activity and diet behaviour for people with chronic conditions: a systematic review. BMJ Open 2022 Aug;12(8):e058229. [doi: 10.1136/bmjopen-2021-058229]
9. Mair JL, Salamanca-Sanabria A, Augsburger M, et al. Effective behavior change techniques in digital health interventions for the prevention or management of noncommunicable diseases: an umbrella review. Ann Behav Med 2023 Sep 13;57(10):817-835. [doi: 10.1093/abm/kaad041] [Medline: 37625030]
10. Vogels E. Some digital divides persist between rural, urban and suburban America. Pew Research Center. 2021. URL: https://www.pewresearch.org/short-reads/2021/08/19/some-digital-divides-persist-between-rural-urban-and-suburban-america/ [accessed 2026-06-01]
11. Maita KC, Maniaci MJ, Haider CR, et al. The impact of digital health solutions on bridging the health care gap in rural areas: a scoping review. Perm J 2024 Sep 16;28(3):130-143. [doi: 10.7812/TPP/23.134] [Medline: 39135461]
12. Woods L, Eden R, Macklin S, et al. Strengthening rural healthcare outcomes through digital health: qualitative multi-site case study. BMC Health Serv Res 2024 Sep 19;24(1):1096. [doi: 10.1186/s12913-024-11402-4]
13. Lt JLM, Lane NM, Holmes GM, et al. Health disparities in Appalachia. Appalachian Regional Commission. 2017. URL: https://www.arc.gov/report/health-disparities-in-appalachia/ [accessed 2026-06-01]
14. Yuhas M, Hedrick V, Zoellner J. Consumption of added sugars by rural residents of Southwest Virginia. J Appalach Health 2020;2(3):53-68. [doi: 10.13023/jah.0203.07] [Medline: 35770211]
15. Hedrick VE, Davy BM, You W, Porter KJ, Estabrooks PA, Zoellner JM. Dietary quality changes in response to a sugar-sweetened beverage-reduction intervention: results from the Talking Health randomized controlled clinical trial. Am J Clin Nutr 2017 Apr;105(4):824-833. [doi: 10.3945/ajcn.116.144543] [Medline: 28251935]
16. Kit BK, Fakhouri THI, Park S, Nielsen SJ, Ogden CL. Trends in sugar-sweetened beverage consumption among youth and adults in the United States: 1999-2010. Am J Clin Nutr 2013 Jul;98(1):180-188. [doi: 10.3945/ajcn.112.057943] [Medline: 23676424]
17. Khairunnisa N, Srygley S. Digital access improves in Appalachia, bridging rural-urban divide. Population Reference Bureau. 2024. URL: https://www.prb.org/articles/digital-access-improves-in-appalachia-bridging-rural-urban-divide [accessed 2026-06-01]
18. Srygley S, Khairunnisa N, Elliott D. The Appalachian region: a data overview from the 2018-2022 American Community Survey Chartbook. Appalachian Regional Commission. 2024. URL: https://www.arc.gov/wp-content/uploads/2024/06/PRB_ARC_Chartbook_ACS_2018-2022_FINAL_2024-06.pdf [accessed 2026-06-01]
19. Brittain M, Consedine N, Bagot KL, Booth N, Rodda SN. Sugar Habit Hacker: initial evidence that a planning intervention reduces sugar intake. J Behav Addict 2021 Sep 22;10(3):471-481. [doi: 10.1556/2006.2021.00054] [Medline: 34550904]

XSL•FO
RenderX

20. Zoellner JM, Porter KJ, You W, et al. Study protocol for iSIPsmarter: a randomized-controlled trial to evaluate the efficacy, reach, and engagement of a technology-based behavioral intervention to reduce sugary beverages among rural Appalachian adults. Contemp Clin Trials 2021 Nov;110:106566. [doi: 10.1016/j.cct.2021.106566] [Medline: 34492306]
21. Zoellner JM, You W, Porter K, et al. A digital behavioral intervention to reduce sugar-sweetened beverage consumption: a randomized, controlled trial. Am J Clin Nutr 2025 Aug;122(2):544-555. [doi: 10.1016/j.ajcnut.2025.06.010]
22. Dwibedi E, You W, Ritterband LM, et al. A stepped care approach to promoting engagement in a digital health intervention: secondary analysis of a randomized controlled trial. J Med Internet Res 2026 May 4;28:e80921. [doi: 10.2196/80921]
23. Brock DP, Ritterband LM, You W, et al. An exploratory study to understand how rurality status and demographic characteristics are associated with enrollment, engagement, and retention in a digital health intervention targeting the Appalachian region. J Rural Health 2025 Mar;41(2):e70042. [doi: 10.1111/jrh.70042]
24. Hedrick VE, Savla J, Comber DL, et al. Development of a brief questionnaire to assess habitual beverage intake (BEVQ-15): sugar-sweetened beverages and total beverage energy intake. J Acad Nutr Diet 2012 Jun;112(6):840-849. [doi: 10.1016/j.jand.2012.01.023] [Medline: 22709811]
25. Zoellner J, Cook E, Chen Y, You W, Davy B, Estabrooks P. Mixed methods evaluation of a randomized control pilot trial targeting sugar-sweetened beverage behaviors. Open J Prev Med 2013 Feb 1;3(1):51-57. [doi: 10.4236/ojpm.2013.31007] [Medline: 23997992]
26. Zoellner J, Krzeski E, Harden S, Cook E, Allen K, Estabrooks PA. Qualitative application of the theory of planned behavior to understand beverage consumption behaviors among adults. J Acad Nutr Diet 2012 Nov;112(11):1774-1784. [doi: 10.1016/j.jand.2012.06.368] [Medline: 23102176]
27. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27. [doi: 10.1007/s12160-009-9133-4]
28. Kinzie MB. A user-centered model for web site design: needs assessment, user interface design, and rapid prototyping. J Am Med Informatics Assoc 2002 Jul 1;9(4):320-330. [doi: 10.1197/jamia.m0822]
29. Zoellner J, Reid A, Porter K, Frederick C, Hilgart M, Ritterband L. Development of a digital behavioral intervention to reduce the consumption of sugar-sweetened beverages among rural Appalachian adults: multiphased, human-centered design approach. JMIR Hum Factors 2023 Feb 1;10:e41262. [doi: 10.2196/41262] [Medline: 36724036]
30. Rural-urban continuum codes. Economic Research Services US Department of Agriculture. URL: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes [accessed 2025-06-24]
31. Wing RR, Tate DF, Espeland MA, et al. Innovative self-regulation strategies to reduce weight gain in young adults. JAMA Intern Med 2016 Jun 1;176(6):755-762. [doi: 10.1001/jamainternmed.2016.1236]
32. McCrabb S, Baker AL, Attia J, et al. Internet-based programs incorporating behavior change techniques are associated with increased smoking cessation in the general population: a systematic review and meta-analysis. Ann Behav Med 2019 Feb 1;53(2):180-195. [doi: 10.1093/abm/kay026] [Medline: 29750240]
33. Cullen KW, Thompson DI, Scott AR, Lara-Smalling A, Watson KB, Konzelmann K. The impact of goal attainment on behavioral and mediating variables among low income women participating in an Expanded Food and Nutrition Education Program intervention study. Appetite 2010 Oct;55(2):305-310. [doi: 10.1016/j.appet.2010.06.017]
34. Rosinger A, Herrick K, Gahche J, Park S. Sugar-sweetened beverage consumption among U.S. adults, 2011-2014. NCHS Data Brief 2017 Jan(270):1-8. [Medline: 28135185]
35. Cuy Castellanos D, Miller ME. Exploration of sugar-sweetened beverage intake in a Latinx population. J Nutr Educ Behav 2020 Mar;52(3):281-289. [doi: 10.1016/j.jneb.2019.08.013]
36. Al-Alawy K, Gaafar R, Moonesar IA, Muhsineh S. The nexus of sugar-sweetened beverages: insights on demographics, enablers and barriers. Public Health Pract 2021 Nov;2:100189. [doi: 10.1016/j.puhip.2021.100189]
37. Teasdale N, Elhussein A, Butcher F, et al. Systematic review and meta-analysis of remotely delivered interventions using self-monitoring or tailored feedback to change dietary behavior. Am J Clin Nutr 2018 Feb;107(2):247-256. [doi: 10.1093/ajcn/nqx048]
38. Marcu G, Ondersma SJ, Spiller AN, Broderick BM, Kadri R, Buis LR. Barriers and considerations in the design and implementation of digital behavioral interventions: qualitative analysis. J Med Internet Res 2022 Mar 30;24(3):e34301. [doi: 10.2196/34301] [Medline: 35353043]
39. Schumacher LM, Martinelli MK, Convertino AD, Forman EM, Butryn ML. Weight-related information avoidance prospectively predicts poorer self-monitoring and engagement in a behavioral weight loss intervention. Ann Behav Med 2021 Mar 16;55(2):103-111. [doi: 10.1093/abm/kaaa034] [Medline: 32491152]
40. Crane N, Hagerman C, Horgan O, Butryn M. Patterns and predictors of engagement with digital self-monitoring during the maintenance phase of a behavioral weight loss program: quantitative study. JMIR MHealth UHealth 2023 Jul 18;11:e45057. [doi: 10.2196/45057] [Medline: 37463017]
41. Fryar C, Carrol M, Afful J. Prevalence of overweight, obesity, and severe obesity among adults aged 20 and over: United States, 1960-1962 through 2017-2018. : Centers for Disease Control and Prevention; 2020 URL: https://www.cdc.gov/nchs/data/hestat/obesity-adult-17-18/obesity-adult.htm [accessed 2025-06-23]
42. Pinho MGM, Mackenbach JD, Charreire H, et al. Exploring the relationship between perceived barriers to healthy eating and dietary behaviours in European adults. Eur J Nutr 2017 Apr 26;57(5):1761-1770. [doi: 10.1007/s00394-017-1458-3]

XSL•FO
RenderX

43. De Leon A, Roemmich JN, Casperson SL. Identification of barriers to adherence to a weight loss diet in women using the nominal group technique. Nutrients 2020 Dec 6;12(12). [doi: 10.3390/nu12123750] [Medline: 33291249]
44. Kelley CP, Sbrocco G, Sbrocco T. Behavioral modification for the management of obesity. Prim Care 2016 Mar;43(1):159-175. [doi: 10.1016/j.pop.2015.10.004] [Medline: 26896208]
45. Ester M, Eisele M, Wurz A, McDonough MH, McNeely M, Culos-Reed SN. Current evidence and directions for future research in eHealth physical activity interventions for adults affected by cancer: systematic review. JMIR Cancer 2021 Sep 20;7(3):e28852. [doi: 10.2196/28852] [Medline: 34542415]
46. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation: a 35-year odyssey. Am Psychol 2002 Sep;57(9):705-717. [doi: 10.1037/0003-066x.57.9.705]
47. Porter KJ, Brock DJP, Reid AL, et al. Participants' experiences with and perceptions of a digital health intervention to reduce sugar-sweetened beverage intake: A mixed methods summative evaluation of iSIPsmarter". Internet Interv 2026 Mar;43:100921. [doi: 10.1016/j.invent.2026.100921] [Medline: 41816219]
48. Burke LE, Shiffman S, Music E, et al. Ecological momentary assessment in behavioral research: addressing technological and human participant challenges. J Med Internet Res 2017 Mar 15;19(3):e77. [doi: 10.2196/jmir.7138] [Medline: 28298264]

Abbreviations

BCT: behavior change technique
RCT: randomized controlled trial
RUCC: rural-urban continuum code
SSB: sugar-sweetened beverage

Edited by A Stone, L Buis; submitted 17.Jan.2025; peer-reviewed by CJ Hagerman, S Rodda; revised version received 14.May.2026; accepted 15.May.2026; published 12.Jun.2026.

Please cite as:
Reid A, Ritterband L, Porter K, Brock DJ, Frederick C, Zoellner J
Action Planning for Reducing Sugar-Sweetened Beverage Intake in Appalachian Adults: Longitudinal Process Evaluation of a Digital Behavioral Health Intervention
JMIR Mhealth Uhealth 2026;14:e71241
URL: https://mhealth.jmir.org/2026/1/e71241
doi:10.2196/71241

© Annie Reid, Lee Ritterband, Kathleen Porter, Donna-Jean Brock, Christina Frederick, Jamie Zoellner. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 12.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Metaverse Program to Prevent E-Cigarette Use Among Adolescents: Pilot Mixed Methods Study

Hyunmi Son, PhD; Gyumin Han, PhD

College of Nursing, Research Institute of Nursing Science, Pusan National University, 49, Busandaehak-ro, Mulgeum-eup, Yangsan-si, Republic of Korea

Corresponding Author:
Gyumin Han, PhD
College of Nursing, Research Institute of Nursing Science, Pusan National University, 49, Busandaehak-ro, Mulgeum-eup, Yangsan-si, Republic of Korea

Abstract

Background: E-cigarette use among adolescents presents challenges distinct from traditional smoking issues. The perception that e-cigarettes are less harmful and their distinctive appeal as a means of sensation seeking among adolescents suggest that traditional smoking prevention programs may not guarantee effectiveness in preventing e-cigarette use. Therefore, there is an urgent need for educational interventions to prevent adolescents from using e-cigarettes.

Objective: This pilot study aimed to assess the potential of the Meta Nodam program, a novel metaverse-based educational initiative to enhance adolescents' resistance to e-cigarette use and increase their engagement in preventive education.

Methods: The Meta Nodam program was developed on the Roblox platform, designed to promote autonomous engagement and health-related decision-making among adolescents. The study included 120 participants in a 2-session intervention over 2 weeks, each session lasting 45 minutes. Changes in key cognitive outcomes (e-cigarette use intentions, attitudes, norms, self-efficacy, and interest development) were examined using pre- and postintervention surveys. In addition, focus group interviews with 10 participants were conducted to gain deeper insights into participants' experiences.

Results: Quantitative analysis showed a significant within-group increase in e-cigarette–related self-efficacy (t_{119}=–3.171; P=.002), whereas no statistically significant pretest-posttest differences were observed in intentions (t_{119}=–0.846; P=.40) or attitudes (t_{119}=−0.080; P=.94). Qualitative findings explained and expanded these results by indicating that immersive metaverse experiences made e-cigarette–related risks more tangible, encouraged peer discussion, and promoted reflection on e-cigarette–related decision-making. Participants also described the program as engaging and reported interest in further exploring information about the health risks of e-cigarette use.

Conclusions: Owing to the short duration of the metaverse experience, limited to 2 sessions over 2 weeks, the program did not measurably alter adolescents' attitudes or intentions regarding e-cigarette use. However, it was associated with enhanced e-cigarette–related self-efficacy and demonstrated the applicability, usefulness, and attractiveness of the metaverse program. This study highlights the potential of digital platforms such as the metaverse for health education, emphasizing user engagement, interaction, and experiential learning in preventive education. Future research should consider broader participant demographics and more robust experimental designs to more rigorously evaluate the potential of such interventions.

(JMIR Mhealth Uhealth 2026;14:e65817) doi:10.2196/65817

KEYWORDS

e-cigarette; adolescent; metaverse; pilot study; smoking prevention; focus group; health education; user engagement

Introduction

Recent global statistics indicate that e-cigarettes have the highest use rate among tobacco-related products used by adolescents [1]. E-cigarettes are linked to health risks such as myocardial infarction, upper respiratory irritation, immune system disruption, and reduced fertility [1,2]. The use of e-cigarettes by adolescents demonstrates a "gateway effect" that leads to other health issues [3]. Notably, the inclusion of nicotine not only induces addiction but also impairs brain development in adolescents [4], and the capability of users to adjust the quantity of nicotine solutions puts them at high risk of nicotine exposure. In South Korea, national data also show that adolescent e-cigarette use remains a pressing concern. Although the prevalence of conventional cigarette smoking among Korean adolescents continues to decline, e-cigarette use shows an opposite trend. According to the 2023 Korea Youth Risk Behavior Web-based Survey, the rate of e-cigarette use among adolescents steadily increased from 1.9% in 2020 to 2.9% in 2021, reaching 3.3% in 2022 [5]. These consistently rising rates

XSL•FO
RenderX

underscore the necessity for targeted prevention efforts within this population.

Existing tobacco-related youth prevention programs are focused primarily on conventional tobacco products [6]. Although some studies have included e-cigarettes in broader adolescent behavioral interventions [7,8], there has been no focus on e-cigarettes. A systematic review of interventions for the prevention of youth tobacco and nicotine use revealed that no studies targeted e-cigarettes [9]. This underscores a critical oversight as most existing programs encompass general tobacco use rather than focusing on e-cigarette use. Although recent strategies targeting e-cigarette use have evolved, they often repurpose traditional smoking prevention techniques [10]. Unlike traditional cigarettes, e-cigarettes foster a positive social norm around their use, largely due to perceptions of reduced harm and the customizable nature of flavors and scents, appealing to youth's desire for personalization and variety. Additionally, the sleek, technology-oriented design of e-cigarettes, coupled with their discreet use and portrayal on social media, enhances their attractiveness and social acceptability among peers. This unique appeal, combined with the communal aspect of vaping, underscores the need for specifically tailored interventions that address the nuances of e-cigarette use among youth, challenge positive norms, and highlight the health risks involved.

Given adolescents' growing perception of e-cigarettes as appealing [11], prevention efforts must address the key cognitive determinants of decision-making—specifically, attitudes, perceived social norms, and refusal self-efficacy—within the social contexts where vaping occurs. These cognitive factors are not isolated; they are shaped by multilevel influences, including peer behavior, school and family norms, retail access, and media exposure. Refusal self-efficacy, a concept rooted in the social cognitive theory by Bandura [12,13], refers to the belief in one's capability to successfully decline e-cigarettes despite peer pressure or temptation.

To reinforce this efficacy, repeated mastery experiences in diverse contexts are essential. Metaverse environments are particularly well suited for this purpose, offering immersive, realistic, yet safe simulations of high-risk situations that are difficult to replicate in real life [14-16]. Within these virtual spaces, adolescents can practice refusal skills and observe how peer reactions and environmental cues influence their attitudes and norms through collaborative gameplay and problem-based scenarios. Therefore, developing metaverse-based educational programs that incorporate these multilevel influences into their design to specifically target changes in adolescents' attitudes, norms, and self-efficacy represents a promising strategy for e-cigarette use prevention.

An examination of existing adolescent educational programs using the metaverse reveals that traditional lecture-based content is frequently transferred directly into virtual realms. However, this approach has been criticized for not fully exploiting the potential of metaverse environments, which offer significant freedom and diverse experiential possibilities [14,17]. Such programs often limit opportunities for real-life experiential learning and fail to leverage the extensive capabilities of the metaverse, especially in contexts where practically any scenario can be simulated [18]. Additionally, research on the actual experiences of adolescents within these highly interactive and immersive metaverse environments remains scarce [19]. To address this gap, a pilot study was conducted to investigate the potential of the metaverse for raising awareness and delivering preventive health education to adolescents. The immersive and interactive nature of metaverse environments uniquely supports experiential learning and the strengthening of refusal self-efficacy, crucial components for effective health education. This pilot study aimed to evaluate the feasibility and user engagement of an interactive metaverse program and examine preliminary within-group changes in key cognitive outcomes (self-efficacy, attitudes, norms, and intentions) related to e-cigarette use among adolescents.

Methods

Research Design

This study used an explanatory sequential mixed methods pilot design in which the quantitative phase was conducted first followed by a qualitative phase to provide a more in-depth understanding of the quantitative findings. In the quantitative phase, a single-group pretest-posttest design was used with 120 participants to evaluate the feasibility and acceptability of the intervention and assess preliminary changes in key cognitive outcomes. After the intervention and postintervention survey, qualitative data were obtained from 10 participants through focus group interviews to explore participants' experiences within the metaverse and provide contextual insights into how the program influenced their perceptions of and attitudes toward e-cigarettes. The qualitative themes were then used to explain and expand upon the quantitative results. Integration occurred during the interpretation and presentation of the findings by linking quantitative outcomes with related qualitative themes.

Development of Meta Nodam

The Meta Nodam program was developed to leverage the immersive capabilities of the metaverse, allowing adolescents to safely experience and navigate the multilevel factors influencing e-cigarette use within a simulated environment. The program was developed and implemented on the Roblox platform. The name Meta Nodam combines "Meta," referencing its metaverse-based platform, with "Nodam," a Korean term meaning "no tobacco," reflecting the program's purpose of promoting healthy decision-making regarding e-cigarette use.

The development of Meta Nodam was grounded in the multilevel determinants of adolescent e-cigarette use identified by Han and Son [20]. Recognizing that e-cigarette use is driven by a complex interplay of individual, interpersonal, and environmental factors, the program used the metaverse to simulate these multilevel influences within a safe, virtual space. While the intervention environment mirrored these broader contextual factors, the educational strategies were specifically designed to enhance individual-level cognitive determinants—attitudes, norms, and self-efficacy—as the proximal mechanisms for driving behavior change.

XSL•FO
RenderX

The Meta Nodam program incorporated its learning objectives into a set of interconnected metaverse spaces that adolescents could explore. These included Fashion Square (Figure 1C), Play Zone (Figure 1A), Living Zone (Figure 1B), Forest Zone (Figure 1E), and Treatment Zone (Figure 1D). Each zone provided a distinct contextual experience relevant to how adolescents encounter e-cigarette–related cues in their daily lives. Fashion Square served as the central gathering space where participants interacted with peers' avatars and expressed identity through customization, offering natural opportunities to observe social dynamics. The Play Zone and Living Zone reflected familiar social environments in which adolescents often encounter peer influence and normative cues. These spaces allowed adolescents to observe others' behaviors and reflect on how social expectations shape their own decisions.

Figure 1. Screenshots of the Meta Nodam metaverse environment. The labeled panels show the interconnected virtual spaces used in the program: (A) Play Zone, (B) Living Zone, (C) Fashion Square, (D) Treatment Zone, and (E) Forest Zone.



The Forest Zone incorporated a symbolic representation of the negative consequences of sustained e-cigarette use through encounters with zombies. This design provided an experiential but age-appropriate way to illustrate potential harms associated with nicotine use. Virtual e-cigarette vending machines were strategically placed throughout the Play, Living, and Forest Zones to create realistic decision-making points. When participants approached a machine, they were required to make a choice—interact with it or not. If they chose to interact, they were transported to the Treatment Zone, where their activities were temporarily restricted and they were informed about the consequences of the action. This mechanism encouraged participants to reflect on their decision and reinforced the link between choices and outcomes.

Nonplayer characters (NPCs) were integrated throughout the environment to deliver informational messages about health risks, model social reactions, and simulate peer influence. Some NPCs provided direct health-related information intended to support shifts in attitudes toward e-cigarettes. Others expressed approving or disapproving views of vaping, enabling participants to consider how social expectations and peer cues can shape normative beliefs. NPC conversations and scripted scenarios were carefully designed to resemble situations that adolescents may encounter in their everyday lives.

The program's interactive components were strategically designed to function as pedagogical tools, transforming abstract health concepts into concrete virtual experiences. The combination of informational messages, symbolic representations, peer interactions, situational cues, and

XSL•FO
RenderX

decision-making moments created an integrated experiential learning environment that supported the targeted cognitive constructs. Opportunities to make choices in ambiguous or socially pressured situations were included to strengthen self-efficacy by allowing adolescents to practice navigating tempting or uncomfortable scenarios within a safe virtual space. The structured yet exploratory nature of the metaverse world enabled repeated engagement with these learning opportunities while maintaining a sense of autonomy.

The program was delivered over 2 structured 45-minute sessions conducted in person in school classrooms. Students used laptops with stable internet connections to access the Meta Nodam environment simultaneously, allowing for real-time interactions with peers' avatars. The first session focused on orientation, initial exposure, and exploration of the metaverse environment. It included a 20-minute prebriefing via Zoom (Zoom Video Communications) explaining the program's objectives and instructions, followed by 20 minutes of exploration within the metaverse and a short debriefing. The second session was designed to provide further guided and unguided exploration, peer interaction, reflection on e-cigarette–related decision-making, participation in a fashion contest, and a concluding debriefing along with the postsurvey. Teachers monitored students' activities through their own computers, providing technical support and ensuring safety while minimizing disruptions to the immersive experience.

Pilot Test of Meta Nodam

The pilot test of the Meta Nodam program was conducted as a preliminary investigation among 120 girls' middle school students. This sample size was determined using the G*Power program (version 3.1.9.7) based on a 2-tailed paired *t* test with an 80% power level and a significance level of .05. The effect size, set at 0.257, was based on the impact of a virtual reality game program aimed at preventing adolescent e-cigarette use [21]. To account for a potential dropout rate of approximately 20%, as observed in similar school-based smoking prevention programs [22], 120 participants were initially recruited.

The pilot included pre- and postintervention surveys and focus group interviews with 10 highly engaged students to gain qualitative insights into their experiences. These interviews were conducted in 2 groups, each consisting of 5 highly engaged students. Highly engaged students were selected because the qualitative phase of this pilot study aimed to obtain detailed feedback on program experience, usability, engagement, and areas for improvement. Students who had actively explored the metaverse environment were expected to provide more specific descriptions of program functions, immersive elements, difficulties, and suggestions for refinement. The 2-session, 90-minute format was selected to align with the regular class period structure of the participating middle school and examine the feasibility of implementing the program in a real school setting. As this was a pilot study, the program was intended to assess feasibility, acceptability, engagement, and preliminary cognitive responses rather than long-term cognitive or behavior change. Surveys were administered before the first session and after the second session, each taking approximately 6 to 9 minutes to complete.

Measurement and Data Collection

The first section of the questionnaire assessed self-efficacy in resisting e-cigarette use, a key factor in understanding participants' confidence in handling temptations related to e-cigarette use. This was measured using a tool adapted from Condiotte and Lichtenstein [23]. This scale demonstrated good internal consistency, with Cronbach α values of 0.970 (pretest) and 0.973 (posttest). The next section focused on e-cigarette–related characteristics, particularly the influence of peers' e-cigarette use, which is a critical determinant in adolescent behavior. The aim of this section was to capture the social environment of the participants regarding e-cigarette use. To further categorize the participants, e-cigarette use patterns were assessed by inquiring about both lifetime and recent e-cigarette use, thereby classifying participants as lifetime or current e-cigarette users.

Participants' intentions toward future e-cigarette use were evaluated using modified questions from Pierce et al [24-26] to gauge their susceptibility to e-cigarette use. The Cronbach α for this scale was 0.748 (pretest) and 0.840 (posttest).

Attitudes toward e-cigarettes were measured using an instrument adapted from de Vries et al [27] and modified by Kim and Choi [28]. This section of the questionnaire explored the participants' perceptions of the benefits and risks associated with e-cigarette use. The Cronbach α for this scale was 0.921 (pretest) and 0.842 (posttest). Additionally, e-cigarette–related norms were assessed by adapting a tool from Primack et al [29] to evaluate the perceived social acceptance and prevalence of e-cigarette use within the participants' social circles. The Cronbach α for this measure was 0.617 (pretest) and 0.665 (posttest).

After the intervention, the survey included additional items to evaluate the impact on participants' interest development using an instrument adapted from Wang and Adesope [30]. The subdomains of this instrument assessed various facets of interest development, including triggered situational interest, maintained situational interest, emerging individual interest, and well-developed individual interest, each rated on a scale from 1 to 7, where 1 indicated the lowest level of interest and 7 indicated the highest. The Cronbach α for this scale was 0.941 (posttest).

To complement the quantitative data, qualitative insights were gathered through focus group interviews with participants who showed high engagement in the program and were willing to discuss their experiences extensively. This approach was used to obtain rich, experience-based feedback on program usability, immersion, perceived educational value, and areas for refinement. These interviews began with the following broad question: "How was your experience with the Meta Nodam program?" This open-ended approach allowed participants to share their detailed experiences freely, providing a deeper understanding of the program's impact from their perspective. The researchers facilitated natural discussions among participants, asking clarifying questions only when necessary to confirm or explore specific points in greater depth.

XSL•FO
RenderX

Ethical Considerations

This study was approved by the institutional review board of Pusan National University (PNU IRB/2022_165_HR). Participants who voluntarily expressed interest were thoroughly informed about the study's purpose, content, procedures for withdrawal, privacy protections, and compensation measures in case of research-related harm. As all participants were under 18 years of age, written informed consent was obtained from each adolescent participant as well as their legal guardians through a detailed explanatory letter and consent form. Participants whose guardians declined consent or who chose not to take part were allowed to opt out without any consequences; however, apart from students who were absent from school, there were no cases of withheld consent in this study. Participants received a snack voucher worth KRW 1500, approximately US $1.10 (KRW 1=US $0.00073 as of June 11, 2026), after completing each of the pre- and postintervention surveys as a token of appreciation for their time. This compensation was not linked to their survey responses. The consent procedure ensured informed and voluntary participation, adhering to the highest standards of research ethics and safeguarding the rights and well-being of adolescent participants and their guardians.

Data Analysis

For quantitative data, statistical analysis was performed using SPSS (version 23; IBM Corp). Descriptive statistics were used to assess demographic details, and paired *t* tests were used to compare pre- and postintervention measures of e-cigarette–related cognitive outcomes, including self-efficacy, intentions, attitudes, and normative beliefs, with the significance threshold set at a *P* value of less than .05.

Content analysis was carried out on qualitative data from the focus group interviews to extract themes and patterns. All discussions were audio recorded, transcribed verbatim, and subjected to an inductive content analysis process. Two researchers independently performed line-by-line coding, compared their codebooks, and resolved discrepancies through discussion until full agreement was reached. Codes were then grouped into subcategories and distilled into overarching themes that reflected common perceptions of program usefulness, realism, and perceived changes in refusal confidence. Data saturation was reached after conducting 2 focus group interviews (totaling 10 participants) as no new themes emerged. An audit trail documenting analytic decisions and periodic peer debriefings enhanced dependability, whereas member checking with 30% (3/10) of the participants confirmed that the derived themes accurately represented their views.

The quantitative and qualitative findings were then integrated at the interpretation stage. Specifically, quantitative results were first reviewed for each outcome domain, and related qualitative themes were subsequently linked to these outcomes to determine whether the qualitative findings explained, complemented, or expanded the quantitative findings. This integrated interpretation was used to organize the Results section around preliminary cognitive responses and program experiences.

Results

Participant Characteristics

This study involved female middle school students, none of whom had ever used or were currently using e-cigarettes. Regarding friends' e-cigarette use, 75% (90/120) reported no use among their peers, whereas 25% (30/120) acknowledged that some friends used e-cigarettes. Among the 10 students who participated in the focus group interviews, none had personal experience using e-cigarettes, although 3 (30%) mentioned having friends who used e-cigarettes.

Engagement and Participation Experience in the Metaverse Program

The overall postintervention interest development score averaged 5.59 (SD 0.97), suggesting that participants were engaged with the program content. Among interest subdomains, triggered situational interest was highest (mean 6.11, SD 1.07), followed by maintained situational interest (mean 5.86, SD 1.23), well-developed individual interest (mean 5.55, SD 0.85), and emerging individual interest (mean 5.33, SD 1.10; Table 1). These findings should be interpreted as exploratory, reflecting short-term responses observed within the context of a brief, 2-session pilot program.

XSL•FO
RenderX

Table . Pretest-posttest changes in e-cigarette–related cognitive outcomes and interest development from the Meta Nodam program.

Variable and subcategory	Pretest, mean (SD)	Posttest, mean (SD)	*t* test (*df*)	*P* value
Self-efficacy (0-10)	9.65 (0.91)	9.86 (0.58)	−3.171 (119)	.002
Intention (4-16)	15.57 (1.06)	15.64 (1.07)	−0.846 (119)	.40
Attitudes (12-60)	54.68 (7.43)	54.74 (6.55)	−0.080 (119)	.94
Normative belief				
Perceived prevalence (0-300)	125.92 (50.03)	120.25 (52.64)	1.695 (119)	.09
Approval by parents or peers (3-12)	4.52 (2.34)	4.78 (2.97)	−0.922 (119)	.36
Popularity among the successful or elite (4-16)	12.78 (1.95)	12.77 (2.45)	0.086 (119)	.93
Interest development (1-7)		5.59 (0.97)		
Triggered situational interest	—[a]	6.11 (1.07)	—	—
Maintained situational interest	—	5.86 (1.23)	—	—
Emerging individual interest	—	5.33 (1.10)	—	—
Well-developed individual interest	—	5.55 (0.85)	—	—

[a]Not applicable.

The focus group interviews also showed that the metaverse environment enhanced engagement and learning compared with traditional methods. Participants described the program as enjoyable, gamelike, and immersive, which helped maintain their interest in e-cigarette use prevention education.

Participants noted that, unlike traditional smoking prevention education, which they found boring, the metaverse program was enjoyable and encouraged active participation. The program's playful approach was key to maintaining interest:

> *We usually play in the metaverse for fun or as a hobby. But this time, we first encountered it through an educational program. At first, I thought it would be boring like the typical education we received in elementary school, where they just keep telling us facts and why smoking is bad. But this program turned education into a game, which made it much more interesting and engaging for us.* [Participant 2]

Participants also reported sustained engagement and broadened interest beyond the assigned sessions. The program was so engaging that some participants continued to interact with it even after the sessions ended, sparking broader educational interests beyond e-cigarettes:

> *After just one session, I found the storyline to be really extensive and interesting. So, I kept playing two or three more times over the week. The game had a lot of addicts, and shooting them with a gun to deal with them was surprisingly fun. It made me curious about why they became addicted in the first place, which encouraged me to look into it more.* [Participant 3]

E-Cigarette Refusal Self-Efficacy

Analysis of the quantitative data indicated a statistically significant within-group increase in self-efficacy related to making decisions about e-cigarette use, with scores rising from 9.65 (SD 0.91) before participation to 9.86 (SD 0.58) after participation (t=–3.171; P=.002; Table 1).

The qualitative findings provided contextual insights into this increase by showing that participants did not experience e-cigarette–related situations as abstract health messages but as personally relevant scenarios through their avatars. In particular, the avatar-based experience appeared to help participants place themselves in virtual situations involving e-cigarette use and consider how they might respond to similar temptations or peer-related situations in real life:

> *Experiencing the effects of smoking in the metaverse felt more real than just hearing about it. The metaverse experience was more impactful because my avatar in the game represented me. It felt like I was the one smoking and being taken to a confined space, which made the experience very novel and engaging.* [Participant 6]

The postintervention survey also appeared to provide an additional opportunity for participants to reflect on how they would respond to e-cigarette offers from peers. Rather than only recalling information about e-cigarette risks, participants described reconsidering their own behavioral responses in peer-related situations, including refusing use and discouraging peers from using e-cigarettes:

> *When I was filling out the survey, there were many question items like "I won't smoke even if a friend*

XSL•FO
RenderX

offers." Seeing these question items made me think about my actions. Now, if a friend smokes, I'll stop them and make sure I don't join in. [Participant 1]

Attitudes and Intentions Toward E-Cigarettes

No statistically significant pretest-posttest differences were observed in intentions, attitudes, or normative beliefs (Table 1). However, the qualitative findings suggested that participants developed greater awareness of e-cigarette–related risks and reflected on how e-cigarette use could affect them and their peers.

Participants described the metaverse scenarios as making the risks and consequences of e-cigarette use more vivid. In particular, the symbolic portrayal of addiction appeared to help participants recognize the difficulty of escaping e-cigarette dependence:

I remember the zombies saying things like "help me," which made me realize how hard it must be to escape e-cigarette addiction. It really showed how difficult it can be to break free. [Participant 8]

Participants also reported that the program created opportunities for more open conversations with peers about the risks of e-cigarettes. These discussions appeared to extend participants' risk awareness beyond individual reflection to peer-related communication and prevention-oriented dialogue, although such changes were not captured as statistically significant pretest-posttest changes in attitudes, intentions, or normative beliefs:

First of all, I think the biggest change was in our awareness. In the metaverse, when we finish an activity, we talk to our friends about it. For example, we might say, "Did you know about the dangers of e-cigarette usage?" This creates an atmosphere where we can comfortably talk about these issues. If someone jokes about a friend who might be an e-cigarette user, we can now openly say, "Don't use it; it's not good for you." Before, everyone would whisper and talk secretly, but now, since we all learned together and understand how bad it is, it's more effective when several friends talk about it together. This has been the biggest real-world change. [Participant 2]

Metaverse Program Usability and Refinement

Focus group interviews provided feedback on the usability of the metaverse program and areas for refinement. Participants identified the need for clearer guidance, more realistic and direct educational content, increased interaction, and longer session duration.

Participants suggested the addition of a guide or tutorial for the metaverse environment, including a mini map, to enhance their experience and ease of use:

At first, I didn't know what was going on at all. It would be helpful to have some kind of tutorial or help guide. Also, having a mini-map on the screen would be great to help navigate the structure, maybe placed in the corner. [Participant 4]

Participants recommended making the virtual environment more realistic and incorporating more direct educational messages about e-cigarettes to increase the impact of the program:

I think it would be better if there was more content about e-cigarettes. Like, when you get stuck in the treatment room and have to talk to the NPC, it would be cool if you got more info about e-cigarettes at that point. Maybe even add a quiz or something. [Participant 1]

When I go to the convenience store near school with my friends, there are so many e-cigarette ads that really stand out. It's like They are made for students like us to see. If you could show scenes like this in the metaverse, I think it would help us realize the dangers of e-cigarettes. [Participant 3]

Participants also expressed a desire for increased interaction with NPCs and students from other schools, as well as a longer program duration to fully explore the metaverse:

It would be great if we could interact with the NPCs. Besides that, I think it would be beneficial if this kind of game-based smoking prevention program could be implemented in other schools as well, not just ours. [Participant 5]

The teacher explained everything well, but I felt like two sessions were not enough. It would have been better if the program had been longer. [Participant 2]

Discussion

Principal Findings

This pilot study of the Meta Nodam program provided preliminary insights into how a metaverse-based intervention may support adolescents' engagement, awareness, and cognitive readiness to resist e-cigarette use. The most notable quantitative finding was a statistically significant within-group increase in e-cigarette refusal self-efficacy. This finding was contextualized by the qualitative results, which showed that participants experienced e-cigarette–related situations as personally relevant through their avatars and reflected on how they might respond to peer offers or temptations in real life.

This improvement in self-efficacy may be related to the pedagogical strategies embedded within the metaverse environment. According to self-efficacy theory [12,13], mastery experiences are an important source of confidence. Meta Nodam was designed to provide simulated decision-making experiences in realistic but safe scenarios, allowing students to encounter e-cigarette–related cues and consequences without real-world risks. Qualitative findings suggested that the avatar-based experience made these situations feel personally meaningful and prompted participants to consider their own responses to similar situations. These findings suggest a possible pathway through which virtual rehearsal and reflection may support e-cigarette refusal self-efficacy.

The observed pattern aligns with earlier studies showing that virtual and immersive environments can effectively enhance

XSL•FO
RenderX

adolescents' knowledge and confidence [21]. Unlike traditional didactic education, which often relies on passive information transfer, the metaverse allowed for situated learning, where students actively navigated peer influence and environmental cues. Accumulating evidence supports this approach, suggesting that immersive or virtual reality–based lessons elicit higher engagement and learning outcomes than traditional methods among secondary school learners [14,16]. In this respect, Meta Nodam offers a promising strategy to supplement conventional prevention approaches by providing digital native youth with a safe, interactive space to build behavioral competence.

The absence of statistically significant changes in attitudes and intentions regarding e-cigarette use contrasts with findings from some earlier prevention programs [7,28]. However, several contextual factors may help explain this pattern. Many participants in this study already reported low intention to use e-cigarettes and generally negative attitudes toward vaping before the intervention, suggesting limited room for measurable shifts. Moreover, the program consisted of only 2 short sessions, which may not have been sufficient to influence more stable cognitive constructs. Nonetheless, qualitative data revealed that many participants reflected more deeply on vaping risks, expressed heightened awareness, and reported increased comfort discussing e-cigarette issues with peers following the program. These qualitative impressions may reflect subtle early shifts that were not detectable in the quantitative measures over such a brief time frame. As a pilot study, these findings highlight the need for further research with longer or more intensive implementations and more diverse samples to more fully examine the potential for cognitive and attitudinal shifts.

The evaluation of interest development scores indicated strong engagement and intrinsic motivation among participants, consistent with outcomes observed in prior metaverse-based educational programs [31]. Triggered situational interest showed the highest postintervention levels, suggesting that the novelty and interactive nature of the metaverse environment successfully captured students' attention. This finding aligns with those of research demonstrating the motivational benefits of gamified or interactive learning environments [18,32]. Participants' qualitative responses also supported this as several reported exploring the program beyond assigned sessions and described becoming curious about issues related to e-cigarette use. The program's autonomy-enhancing features, symbolic representations, and opportunities for experiential learning may have contributed to these patterns. Additionally, the implementation of scenario-based role-play allowed students to actively rehearse refusal skills within contexts that closely mirrored real-life situations [21,32]. This practical application likely contributed to the students' perceived increase in confidence in managing vaping temptations.

Regarding the program design, participants differentiated Meta Nodam from existing metaverse educational platforms that often replicate traditional classroom environments, such as those on Gather Town [33]. In contrast, this program was appreciated for its experiential and narrative-driven elements that transcended simple knowledge transmission. However, feedback indicating a desire for more explicit educational content suggests that future iterations should aim for a refined balance between open-ended exploration and structured instructional density. This aligns with user-centered design principles [34], which emphasize that harmonizing learner autonomy with clear educational goals is essential for maximizing both engagement and learning efficacy.

Successful implementation also relied heavily on the educator's role. Teacher-led prebriefing sessions were identified as a critical component for orienting students and framing the experience as an educational intervention rather than purely recreational. As real-time interventions during the simulation can disrupt immersion [14], providing clear guidance beforehand proved to be the most effective strategy. However, adopting such high-technology tools imposes practical demands on educators, including the need for technical proficiency and additional preparation time [14,35], highlighting the necessity for adequate infrastructure and teacher support systems.

This pilot study offers preliminary indications that metaverse-based approaches may be promising for supporting adolescents' awareness and confidence regarding e-cigarette use. The immersive and interactive nature of the program may help create opportunities for reflective learning and peer dialogue. However, several limitations must be considered. First, as this study was designed as a pilot trial to evaluate feasibility and immediate preliminary responses, the findings reflect short-term changes only. Therefore, these results should be interpreted as exploratory evidence rather than definitive proof of the program's long-term effectiveness. The brief 2-session, 90-minute intervention exposure was an important limitation and may have been insufficient to produce measurable changes in relatively stable cognitive constructs such as attitudes and intentions. The single-gender, single-school sample also restricts generalizability to more diverse populations. Technical requirements, teacher workload, and system stability pose additional considerations for broader implementation. Furthermore, initial program deployment revealed operational challenges such as software bugs and the breadth of the virtual world relative to available time, suggesting opportunities to streamline content in future iterations. Addressing these limitations—through longer interventions, additional or booster sessions, repeated role-play or refusal practice, stronger integration of structured educational content, improved technical support, follow-up assessments, and multisite evaluation—will be essential to fully understand the potential role of metaverse-based programs in adolescent tobacco use prevention.

Conclusions

This pilot study suggests that the Meta Nodam program may offer an engaging and interactive means of supporting adolescents' awareness and confidence related to e-cigarette use. By situating educational content within a metaverse environment, the program provided opportunities for students to explore realistic social situations, interact with peers, and reflect on vaping-related decisions in a safe and immersive setting. While no measurable changes were observed in attitudes or intentions during this short 2-session format, participants demonstrated increased self-efficacy and described positive engagement with the program's activities.

XSL•FO
RenderX

Given its exploratory nature, the findings should be interpreted cautiously. Nonetheless, the results indicate that metaverse-based environments may hold promise as a complementary tool within school-based prevention efforts, particularly for digital native adolescents who may benefit from interactive and experience-driven learning. Further refinement and evaluation—including longer program duration, expanded educational content, and more diverse samples—are needed to more fully understand how metaverse-based programs can support adolescents' decision-making and contribute to broader strategies aimed at preventing e-cigarette use among youth.

Acknowledgments

The authors express their sincere gratitude to Mi-Young Jung, a school nurse at Hanseong Girls' Middle School, for her invaluable assistance in the development and implementation of the program. They also extend their thanks to Chae-won Lim and the members of an extracurricular club program for high school students affiliated with the Korea Advanced Institute of Science and Technology (KAIST), whose support was instrumental in creating the program. During the revision process, generative artificial intelligence was used to assist with language refinement and organization of the manuscript. The authors reviewed, verified, and edited all artificial intelligence–assisted text and take full responsibility for the accuracy, integrity, and final content of the manuscript.

Funding

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (RS-2023-00249649).

Authors' Contributions

GH contributed to conceptualization, methodology, investigation, formal analysis, data curation, writing—original draft, and writing—review and editing. HS contributed to conceptualization, methodology, supervision, validation, and writing—review and editing. Both authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Stratton KR, Kwan LY, Eaton DL, et al. Public Health Consequences of E-Cigarettes: National Academies Press; 2018. [doi: 10.17226/24952]
2. E-cigarette use among youth and young adults: a report of the Surgeon General. US Department of Health and Human Services. 2016. URL: https://cdn.who.int/media/docs/default-source/searo/india/tobacoo/2016-sgr-full-report-non-508.pdf?sfvrsn=c1f4ad3b_2 [accessed 2026-06-05]
3. Etter JF. Gateway effects and electronic cigarettes. Addiction 2018 Oct;113(10):1776-1783. [doi: 10.1111/add.13924] [Medline: 28786147]
4. Yuan M, Cross SJ, Loughlin SE, Leslie FM. Nicotine and the adolescent brain. J Physiol 2015 Aug 15;593(16):3397-3412. [doi: 10.1113/JP270492] [Medline: 26018031]
5. Korea youth health behavior survey. Korea Disease Control and Prevention Agency. URL: https://kdca.go.kr/yhs/yhs/main.do [accessed 2024-12-23]
6. Selph S, Patnode C, Bailey SR, Pappas M, Stoner R, Chou R. Primary care-relevant interventions for tobacco and nicotine use prevention and cessation in children and adolescents: updated evidence report and systematic review for the US Preventive Services Task Force. JAMA 2020 Apr 28;323(16):1599-1608. [doi: 10.1001/jama.2020.3332] [Medline: 32343335]
7. Kelder SH, Mantey DS, Van Dusen D, Case K, Haas A, Springer AE. A middle school program to prevent e-cigarette use: a pilot study of "CATCH My Breath". Public Health Rep 2020;135(2):220-229. [doi: 10.1177/0033354919900887] [Medline: 31968177]
8. Gaiha SM, Duemler A, Silverwood L, Razo A, Halpern-Felsher B, Walley SC. School-based e-cigarette education in Alabama: impact on knowledge of e-cigarettes, perceptions and intent to try. Addict Behav 2021 Jan;112:106519. [doi: 10.1016/j.addbeh.2020.106519] [Medline: 32890911]
9. Gardner LA, Rowe AL, Newton NC, et al. A systematic review and meta-analysis of school-based preventive interventions targeting e-cigarette use among adolescents. Prev Sci 2024 Oct;25(7):1104-1121. [doi: 10.1007/s11121-024-01730-6] [Medline: 39325296]
10. Goldenson NI, Leventhal AM, Simpson KA, Barrington-Trimis JL. A review of the use and appeal of flavored electronic cigarettes. Curr Addict Rep 2019 Jun;6(2):98-113. [doi: 10.1007/s40429-019-00244-4] [Medline: 31453046]

XSL•FO
RenderX

11. Sharma A, McCausland K, Jancey J. Adolescent's health perceptions of e-cigarettes: a systematic review. Am J Prev Med 2021 May;60(5):716-725. [doi: 10.1016/j.amepre.2020.12.013] [Medline: 33775514]
12. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
13. Bandura A, Adams NE. Analysis of self-efficacy theory of behavioral change. Cogn Ther Res 1977;1:287-310. [doi: 10.1007/BF01663995]
14. Kye B, Han N, Kim E, Park Y, Jo S. Educational applications of metaverse: possibilities and limitations. J Educ Eval Health Prof 2021;18:32. [doi: 10.3352/jeehp.2021.18.32] [Medline: 34897242]
15. Park SM, Kim YG. A metaverse: taxonomy, components, applications, and open challenges. IEEE Access 2022;10:4209-4251. [doi: 10.1109/ACCESS.2021.3140175]
16. Tlili A, Huang R, Shehata B, et al. Is metaverse in education a blessing or a curse: a combined content and bibliometric analysis. Smart Learn Environ 2022 Dec;9(1):24. [doi: 10.1186/s40561-022-00205-x]
17. Jagatheesaperumal SK, Ahmad K, Al-Fuqaha A, Qadir J. Advancing education through extended reality and internet of everything enabled metaverses: applications, challenges, and open issues. IEEE Trans Learn Technol 2024;17:1120-1139. [doi: 10.1109/TLT.2024.3358859]
18. Jovanović A, Milosavljević A. VoRtex metaverse platform for gamified collaborative learning. Electronics 2022;11(3):317. [doi: 10.3390/electronics11030317]
19. Agac G, Sevim F, Celik O, Bostan S, Erdem R, Yalcin YI. Research hotspots, trends and opportunities on the metaverse in health education: a bibliometric analysis. Library Hi Tech 2025 Feb 4;43(1):1-35. [doi: 10.1108/LHT-04-2023-0168]
20. Zhang Q, Lee J, Jiang B, Kim G. Revitalization of the waterfront park based on industrial heritage using post-occupancy evaluation-a case study of Shanghai (China). Int J Environ Res Public Health 2022;19(15):9107. [doi: 10.3390/ijerph19159107]
21. Weser VU, Duncan LR, Pendergrass TM, Fernandes CS, Fiellin LE, Hieftje KD. A quasi-experimental test of a virtual reality game prototype for adolescent e-cigarette prevention. Addict Behav 2021 Jan;112:106639. [doi: 10.1016/j.addbeh.2020.106639] [Medline: 32920455]
22. Scheffers-van Schayck T, Mujcic A, Otten R, Engels R, Kleinjan M. The effectiveness of smoking-cessation interventions tailored to smoking parents of children aged 0-18 years: a meta-analysis. Eur Addict Res 2021;27(4):278-293. [doi: 10.1159/000511145] [Medline: 33311028]
23. Condiotte MM, Lichtenstein E. Self-efficacy and relapse in smoking cessation programs. J Consult Clin Psychol 1981 Oct;49(5):648-658. [doi: 10.1037//0022-006x.49.5.648] [Medline: 7287974]
24. Pierce JP, Choi WS, Gilpin EA, Farkas AJ, Merritt RK. Validation of susceptibility as a predictor of which adolescents take up smoking in the United States. Health Psychol 1996 Sep;15(5):355-361. [doi: 10.1037//0278-6133.15.5.355] [Medline: 8891714]
25. Pierce JP, Choi WS, Gilpin EA, Farkas AJ, Berry CC. Tobacco industry promotion of cigarettes and adolescent smoking. JAMA 1998 Feb 18;279(7):511-515. [doi: 10.1001/jama.279.7.511] [Medline: 9480360]
26. Pierce JP, Farkas AJ, Evans N, Gilpin E. An improved surveillance measure for adolescent smoking? Tob Control 1995;4(Suppl 1):S47-S56. [doi: 10.1136/tc.4.suppl1.S47]
27. de Vries H, Mudde A, Leijs I, et al. The European Smoking Prevention Framework Approach (EFSA): an example of integral prevention. Health Educ Res 2003 Oct;18(5):611-626. [doi: 10.1093/her/cyg031] [Medline: 14572020]
28. Kim JH, Choi YH. Effects of a strength based I-change smoking cessation program for smoking middle school boys [Article in Korean]. J Korean Acad Nurs 2017 Apr;47(2):164-177. [doi: 10.4040/jkan.2017.47.2.164] [Medline: 28470154]
29. Primack BA, Switzer GE, Dalton MA. Improving measurement of normative beliefs involving smoking among adolescents. Arch Pediatr Adolesc Med 2007 May;161(5):434-439. [doi: 10.1001/archpedi.161.5.434] [Medline: 17485617]
30. Wang Z, Adesope O. Exploring the effects of seductive details with the 4-phase model of interest. Learn Motiv 2016 Aug;55:65-77. [doi: 10.1016/j.lmot.2016.06.003]
31. Lim T, Ryu J, Jeong Y. The effects of emotional interaction by avatar on presence and interest development in the metaverse learning environment [Article in Korean]. Korea Educ Rev 2022 Mar;28(1):167-189. [doi: 10.29318/KER.28.1.7]
32. Rheu MM, Jang Y, Peng W. Enhancing healthy behaviors through virtual self: a systematic review of health interventions using avatars. Games Health J 2020 Apr;9(2):85-94. [doi: 10.1089/g4h.2018.0134] [Medline: 31724888]
33. Lo CK, Song Y. A scoping review of empirical studies in Gather.town. In: 2023 11th International Conference on Information and Education Technology (ICIET): IEEE; 2023:1-5. [doi: 10.1109/ICIET56899.2023.10111430]
34. Eldredge LK, Markham CM, Ruiter RA, Fernández ME, Kok G, Parcel GS. Planning Health Promotion Programs: An Intervention Mapping Approach, 4th edition: John Wiley & Sons; 2016. URL: https://books.google.co.in/books/about/Planning_Health_Promotion_Programs.html?id=UyrdCQAAQBAJ&redir_esc=y [accessed 2026-06-15]
35. Prakash A, Haque A, Islam F, Sonal D. Exploring the potential of metaverse for higher education: opportunities, challenges, and implications. Metaverse Basic Appl Res 2023;2:40. [doi: 10.56294/mr202340]

XSL•FO
RenderX

Abbreviations

NPC: nonplayer character

Edited by A Stone, L Buis; submitted 01.Sep.2024; peer-reviewed by F Yang, JY Liao; revised version received 12.May.2026; accepted 13.May.2026; published 24.Jun.2026.

Please cite as:
Son H, Han G
A Metaverse Program to Prevent E-Cigarette Use Among Adolescents: Pilot Mixed Methods Study
JMIR Mhealth Uhealth 2026;14:e65817
URL: https://mhealth.jmir.org/2026/1/e65817
doi:10.2196/65817

© Hyunmi Son, Gyumin Han. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 24.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Enhancing Cognitive Functions of Older Adults With Software Robot: Longitudinal Exploratory Field Study

Byunghun Yun[1,2*], MS; Bohee Kim[3*], MS; Hyunjeong Ko[2], MS; Jinsung Kim[4], MS; Bori R Kim[5,6], PhD; Whani Kim[2], MA; Soyoon Park[1,2], BS; Jinwoo Kim[1,2], PhD; Jee Hang Lee[4,7], PhD; Geon Ha Kim[6], MD, PhD

[1]HCI Lab, Department of Cognitive Science, Yonsei University, Seoul, Republic of Korea

[2]Digital Therapeutics Research & Development, HAII Inc, Seoul, Republic of Korea

[3]Cloud Service Team, LG Electronics, Seoul, Republic of Korea

[4]Department of Human-Centered AI, Sangmyung University, Seoul, Republic of Korea

[5]Ewha Medical Research Institute, Ewha Womans University College of Medicine, Seoul, Republic of Korea

[6]Department of Neurology, Ewha Womans University College of Medicine, Ewha Womans University Mokdong Hospital 1071, Anyangcheon-ro, Yangcheon-gu, Seoul, Republic of Korea

[7]Institute for Advanced Intelligence Study, Daejeon, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Geon Ha Kim, MD, PhD
Department of Neurology, Ewha Womans University College of Medicine, Ewha Womans University Mokdong Hospital 1071, Anyangcheon-ro, Yangcheon-gu, Seoul, Republic of Korea

Abstract

Background: The global prevalence of dementia continues to rise and demands scalable, nonpharmacological interventions. Digital cognitive training has expanded, but many older adults, who have limited digital literacy, struggle to sustain use. We designed an intervention that integrates social interaction, reward-based engagement, and an artificial intelligence (AI) conversational agent, which aims to reduce these barriers and support continuous participation.

Objective: This study examined whether a 12-week digital cognitive training program improves cognitive function in older adults. It also tested whether a group chat service, which enables interaction among participants and with an AI agent, increases engagement and social support.

Methods: We recruited 133 participants (mean age 64.75, SD 6.76; range 55-75 years) who had no diagnosis of dementia. All participants used the *Care & Cure* program for 12 weeks. The program includes an AI chatbot (*Sammy Talk*) and a group chat service (*Our Town*), which supports peer interaction. We measured cognitive function using the Korean Mini-Mental State Examination—Version 2 (K-MMSE-2). We also assessed degrees of social support (Medical Outcomes Study—Social Support Survey), depression (Short Form Geriatric Depression Scale—Korean Version), and engagement (Twente Engagement with eHealth Technologies Scale), and analyzed usage logs to examine participation patterns.

Results: Participants showed improved cognitive function after the intervention (Hedges g=0.350, P<.001). Active users (n=67), who engaged more frequently with the program, showed greater improvement than nonactive users (n=66), especially among those who had lower baseline cognitive scores (Hedges g=0.523, P<.001). Social support increased, particularly emotional and informational support (t_{132}=−6.509, P<.001). Participants reported higher engagement (t_{132}=2.008, P<.05) and lower depression scores (t_{132}=3.093, P<.01). Regression results showed that group chat participation, which promotes interaction with the AI agent, increased engagement in cognitive training (t_{131}=12.395, P<.001). Increased engagement was associated with higher social support (t_{131}=4.102, P<.001) and improved cognitive function (t_{131}=2.467, P<.05). Cognitive training alone did not produce a significant effect. Participants showed low adherence, which indicates a need for strategies that sustain long-term use.

Conclusions: The *Care & Cure* program improved cognitive function and strengthened social support in older adults. Social interaction, which increases engagement, played a central role. These findings suggest that digital cognitive interventions should incorporate social mechanisms to achieve meaningful effects.

(JMIR Mhealth Uhealth 2026;14:e75308) doi:10.2196/75308

KEYWORDS

mobile phone; mild cognitive impairment; chatbot; aging; dementia

XSL•FO
RenderX

Introduction

Background

The global prevalence of dementia is rising, with recent estimates indicating that approximately 50 million people currently live with dementia. It is estimated that the number will increase to 152 million by 2050 [1]. Dementia has a significant impact on personal, family, and economic levels, costing the world about 1.3 trillion dollars [2]. Prevention is essential for dementia management. Among nonpharmaceutical interventions, digital cognitive training has become more popular because prevention is essential for dementia management. Among nonpharmaceutical interventions, digital cognitive training has gained increasing popularity as long-term clinical studies demonstrate that older adults who engage in regular digital cognitive training could delay dementia onset by preserving memory, attention, calculation, language, visuospatial, and executive functions [3-7]. This preservation helps build cognitive reserve that protects against age-related cognitive decline [8,9]. In practice, traditional cognitive training programs are constrained by several limitations. These include (1) limited accessibility for older adults who have mobility issues, (2) heavy reliance on social workers, and (3) substantial financial costs, which are incurred by individuals through personal expenses for participation and by society through health care and social services for people living with dementia [6,10,11].

Many studies suggest that digital cognitive training programs with remote access can overcome these barriers [3-7]. These programs demonstrate greater scalability since they can reach a larger number of users simultaneously without requiring additional resources or staff. For example, while traditional cognitive training requires more therapists or social workers for each new group of users, digital cognitive training programs can be used by many more users with minimal additional cost. This efficiency in scaling makes them more cost-effective, creating a more sustainable and widely accessible option [12-14].

However, current digital training programs have limited accessibility from home mainly because of inadequate consideration of older adult users' digital literacy [15,16]. Low digital proficiency among older adults restricts their access to digital training programs [17]. Even older users with the ability to use digital training programs at home and with the necessary technological infrastructure often struggle with prolonged engagement [18]. Older adults face many challenges, including physical limitations, difficulty processing information, and digital service interfaces that are not easy to use [19]. These barriers worsen clinical results as the disease progresses, increase medical costs, and impose a greater patient burden [20].

Several solutions address these challenges, including simplified user interfaces that reduce cognitive load, gamification elements designed to promote user engagement, personalized digital literacy support, sending push notifications, using chatbots, and using tools to help people connect with others. These features (1) enhance usability, (2) reduce cognitive demand and screen fatigue, and (3) ultimately improve compliance and clinical outcomes [21]. Among these solutions, chatbots and group-based interventions have shown relatively strong potential for enhancing user engagement and effectiveness in digital cognitive training programs demonstrated by numerous studies [21-23].

Conversational agents, also known as chatbots, have the potential to resolve these barriers [24]. A substantial number of studies have shown that conversational agents used for health care interventions elicit moderate to high levels of participant satisfaction and acceptance, largely due to social support features [25-29]. These agents show promise to advance patient-centered care by supporting users in managing their own health [24,25]. Additionally, group-based interventions enhance program engagement by fostering interaction among members [27]. It is widely accepted that conversational agents provide effective social support when integrated with group interventions [28,29]. Two recent studies have validated the effectiveness of a fully automated mobile conversational agent in providing social support [28,29]. In health care settings, the Shared Medical Appointments model, which integrates group-based interventions, continues to gain wider acceptance as a way to improve patient engagement in treatment programs [30-33].

Under this circumstance, we propose *Care & Cure*, a chatbot-based service. The *Care & Cure* service offers 15 different cognitive training programs through *Sammy Talk*, along with a group chat feature called *Our Town*, where participants can interact with one another. This group interaction is intended to improve user engagement with *Sammy Talk*. At the core of this service is *Sammy*, a software robot that delivers personalized cognitive training and fosters social support through conversational interactions. By integrating these features, *Care & Cure* aims to meet both cognitive training and social needs of older adults.

Objectives

The primary objective of this study is to evaluate whether the *Care & Cure* program enhances cognitive function in older adults. The secondary objective is 2-fold: (1) to assess how social support impacts cognitive function and (2) to determine whether increased engagement with *Our Town* encourages greater usage of *Sammy Talk* and interaction with the artificial intelligence (AI) agent. To achieve these objectives, we analyzed the utilization patterns of both *Our Town* and *Sammy Talk* to assess their effects on cognitive function and social support. As an additional evaluation beyond the primary focus on cognitive functions, we examined the effect of social support functionality embedded in both *Sammy Talk* and *Our Town* on emotional well-being, specifically depressive symptoms, which emerged as a complementary benefit to the cognitive improvements.

Methods

Study Design

This was a single-arm, prospective, exploratory study. The protocol and all related materials were approved by the Yonsei University Institutional Review Board (IRB; 7001988 - 202401-HR-1397-06) prior to the study's initiation.

XSL•FO
RenderX

Study Recruitment

We recruited participants from Suncheon, a midsized city in South Jeolla Province, South Korea, with a population of approximately 280,000. The city's older adult population reflects South Korea's aging demographics, with about 20% of residents aged 65 years or older [34]. Suncheon is recognized for facilities that support senior citizens through welfare centers, nursing homes, and silver towns that provide housing and health care services for older adults [35] (see Section 1 in Multimedia Appendix 1 for more details).

Suncheon Nonghyup, a key institution in the city, provides substantial support to the older adult community. Its well-established network facilitated the recruitment process (see Multimedia Appendix 1, Section II for more details). The central office of the farmers' cooperatives distributed formal recruitment documents to local branches. Through this network, we recruited participants who met the inclusion criteria: adults aged 55 and 75 years and proficient in using a smartphone. All participants provided written informed consent prior to participating in the study.

Inclusion and Exclusion Criteria

The inclusion criteria were adults (1) aged 55-75 years who could read and write Korean (for more rationale, see Multimedia Appendix 1, Section III) and (2) owned a smartphone with text typing proficiency. We excluded adults who (1) lacked formal education or were illiterate, (2) had severe visual or auditory impairments, and (3) experienced difficulty using a smartphone.

Prior to recruitment, we evaluated these criteria through a preliminary screening survey. Participants provided demographic information, including age, educational background, and literacy levels. They also demonstrated smartphone proficiency by performing basic tasks such as navigating the interface and sending messages. Additionally, the survey assessed visual and auditory impairments. This screening process identified participants who met all study requirements.

Intervention

Our intervention delivered cognitive enhancement games specifically designed by neurologists through 2 key components on the *KakaoTalk* platform: *Sammy Talk*, a conversational agent providing cognitive games, and *Our Town*, a group chat feature. We chose *KakaoTalk* because more than 70% of South Korea's senior population use it [36]. The *Care & Cure* program, designed to prevent dementia, integrates both components into this widely used messaging platform.

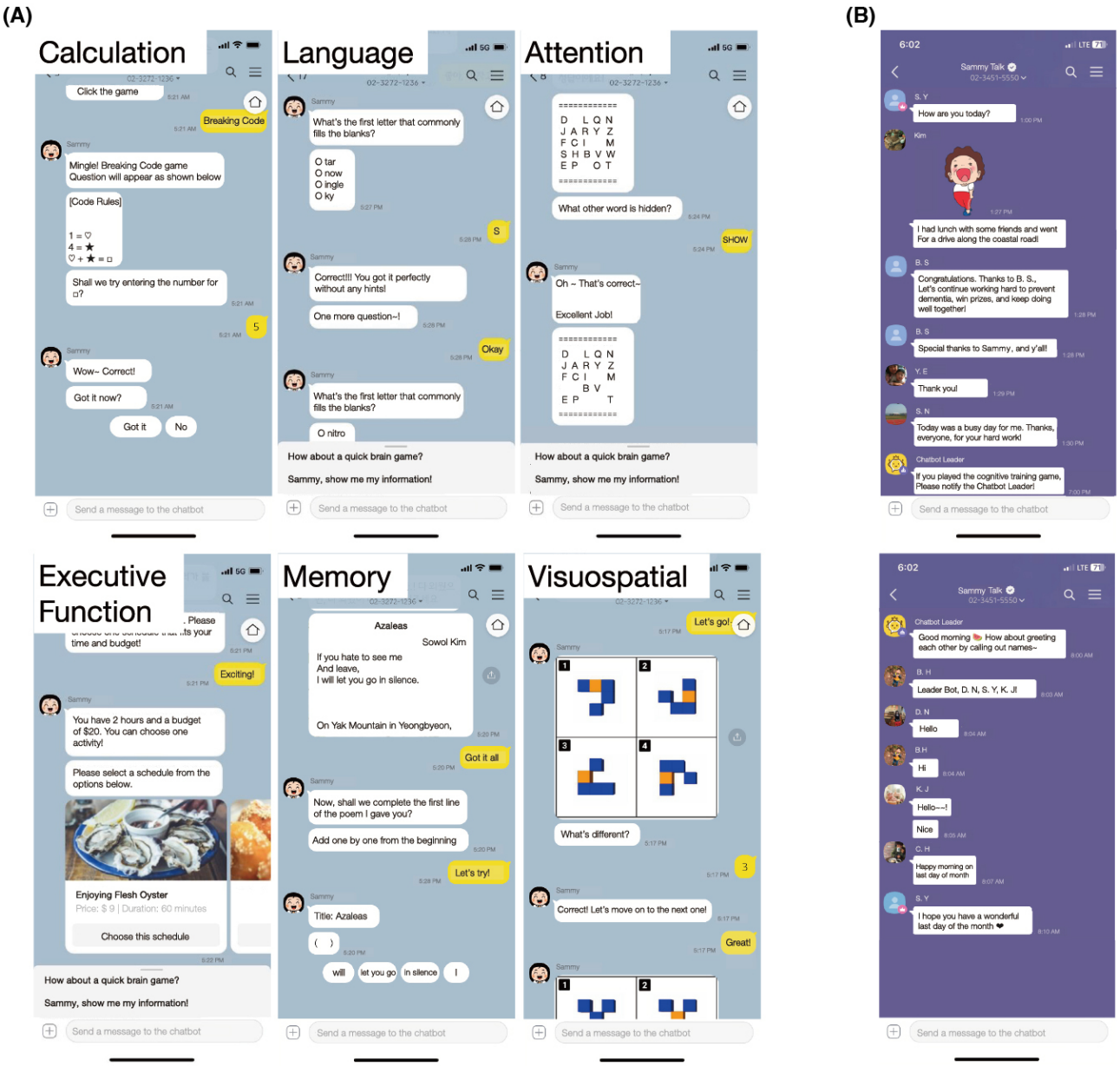
- A software robot, *Sammy*, functions as a conversational agent providing cognitive training through chatbot interactions. Conversational agents, which have demonstrated effectiveness in remote treatment delivery [37], support participants with communication difficulties through verbal or textual interactions [38,39] and are widely used for purposes such as memory enhancement [40] and CBT-based interactive tools for alleviating depression and anxiety [28]. For this reason, we developed *Sammy* as a conversational agent to offer 15 training games targeting 6 cognitive domains: calculation, language, attention, executive function, memory, and visuospatial processing [41-52] (Figure 1A; see Multimedia Appendix 1, Section IV). The game difficulty adapts across 5 levels as participants provide correct answers in *Sammy Talk*. Participants receive 4 games for cognitive improvement at 8 AM. A detailed description is available in Multimedia Appendix 1, Section IV.
- *Our Town* provides a virtual space where participants interact through group chat, promoting regular engagement with other members. Each group includes more than 5 members, and they receive notifications, and the members are expected to interact 3 times a day. An AI chatbot integrated into *Our Town* facilitates these interactions by moderating the group, encouraging participants to engage in *Sammy Talk*, and inquiring about their daily gaming activities and general well-being. Participants receive feedback and rewards based on their daily, weekly, or monthly game completion rates (Figure 1B).

The *Sammy Talk* chatbot functions as an autonomous digital facilitator, delivering morning greetings and encouraging cognitive training activities. The chatbot sends automated prompts 3 times a day (8 AM, 1 PM, and 7 PM) in *Our Town* to promote user engagement with digital cognitive games and ensure task completion. To complement *Sammy Talk*'s limited autonomous functions, a human facilitator motivates less active members, monitors individual participation, and provides personalized feedback with external rewards. The facilitator tracks participant messages in *Our Town* and monitors game performance through an administrative website.

For rewards, the group with the highest weekly participation rate receives gift cards worth KRW 5000 (approximately US $3.70, based on the exchange rate at the time of the study) for each participant and a group certificate. Additionally, when all group members complete their required 4 daily games, they receive restaurant gift cards worth KRW 10,000 per participant (approximately US $7.40, based on the exchange rate at the time of the study) for offline meetups. This integrated approach, combining an autonomous chatbot and human facilitation, fosters an environment that actively encourages engagement and consistent participation in cognitive training.

XSL•FO
RenderX

Figure 1. Overview on the proposed application. (A) Snapshots of *Sammy Talk* represent different cognitive domains in the following order: Calculation, language, attention, executive function, memory, and visuospatial function. (B) Snapshots of *Our Town*.



The interactions in *Our Town* exhibit straightforward structures with moderate turn-taking patterns. A daily interaction usually begins when the chatbot sends an automated morning greeting, which prompts individual responses from group members. These responses emerge sequentially, as members choose either personalized greetings or emoji-based communication. The exchanges primarily focus on 3 elements: mutual well-being inquiries, weather-related small talk, and positive message sharing. The structural simplicity of these interactions, which consist of brief greetings and basic phrases, ensures comfortable participation. Members demonstrate varied communication preferences: some address each participant individually, while others prefer concise responses or emojis. Each interaction typically consists of 5-10 turns, which reflects varying engagement levels. The uncomplicated, conversational nature of *Our Town* creates an accessible environment that accommodates diverse schedules and active communication styles.

A 2-day educational session took place at the *Suncheon Nonghyup* headquarters to establish baseline training. During this session, participants received an introduction to both the *Sammy Talk* program and *Our Town* group chat, alongside a comprehensive guidebook. The research team provided telephone instructions for those unable to attend in person. Subsequently, all participants received a text message with the program access link. The team installed *Sammy Talk* and *Our Town* directly onto the smartphones of session attendees.

Outcome Assessment

The evaluation included 4 areas: cognitive function, emotional well-being, participant engagement, and application usage. We used 2 cognitive assessment scales: the Korean-Mini Mental State Examination—Version 2 (K-MMSE-2) [53] (see Multimedia Appendix 1, Section V for more details). These assessments took place at baseline and within 7 days after the 12-week intervention. The K-MMSE-2 uses a 30-point scale,

XSL•FO
RenderX

where scores of 24 or above indicate normal cognition, 18 - 23 suggest mild cognitive impairment, and below 17 indicate severe cognitive impairment. For emotional function assessment as a supplementary outcome, we used the Short Form Geriatric Depression Scale—Korean version (SGDS-K) [54] and the Medical Outcomes Study–Social Support Survey (MOS-SSS) [55] (see Multimedia Appendix 1, Section V for more details). The SGDS-K uses a 15-point scale, where scores of 5 or below indicate normal mood, scores of 6 - 9 suggest moderate depression, and scores of 10 or above indicate severe depression. The MOS-SSS ranges from 19 to 95 points, with higher scores indicating stronger perceived social support. Changes in depression symptoms and social support served as secondary outcomes, with assessments occurring pre- and postintervention.

In addition to these, we measured participation willingness using the Twente Engagement with eHealth Technologies Scale (TWEETS) [55] (see Multimedia Appendix 1, Section V for more details), which uses a 9-point scale where higher scores indicate greater engagement willingness. The analysis on app engagement relied on log data acquired from the administrative websites of both *Sammy Talk* and *Our Town* (see Multimedia Appendix 1, Section V for more details). For *Sammy Talk*, engagement assessment included play frequency and duration across daily, weekly, and monthly intervals. For *Our Town*, engagement metrics comprised the number of chat exchanges and user-agent interaction duration across identical time intervals. The overall usage metrics included total *Sammy Talk* hours and *Our Town* chat bubble counts throughout the intervention period (Table 1).

Table . Measurement tool.

Category	Measurement		Standard
Cognition	Cognitive function	K-MMSE-2[a]	30
Emotion	Depression	SGDS-K[b]	15
Emotion	Social support	MOS-SSS[c]	19
Willingness to participation	Participation	TWEETS[d]	9
Log data	User engagement	Amount of time of cognitive enhancement game	Minute
Log data	User engagement	Number of interaction in group chat	Number

[a]K-MMSE-2: Korean-Mini Mental State Examination—Version 2.

[b]SGDS-K: Short Form Geriatric Depression Scale—Korean version.

[c]MOS-SSS: Medical Outcomes Study–Social Support Survey.

[d]TWEETS: Twente Engagement with eHealth Technologies Scale.

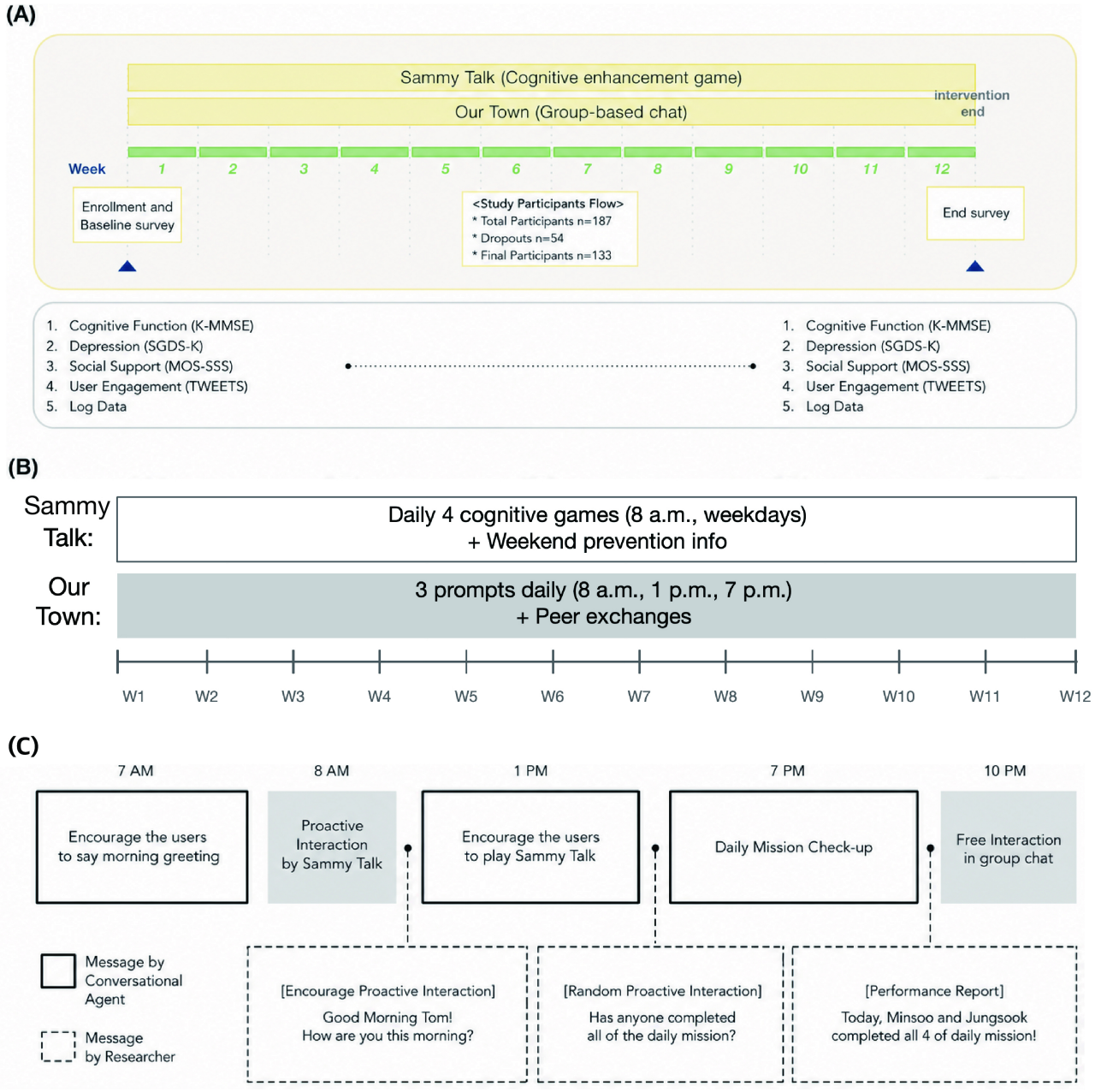
We arranged in-person meetings for pre- and postsurveys based on practical and methodological considerations: participant age preferences, assessment administration requirements, facility availability, and geographical accessibility. The target population, aged 55 - 75 years, typically preferred face-to-face surveys over online formats. The *Suncheon Nonghyup* branch offered appropriate facilities, including halls and meeting rooms, for administering the K-MMSE-2, which required direct facilitator-participant interaction. The branch's proximity to participants' residences enhanced accessibility. Facilitators conducted presurveys during *Care & Cure* program introduction sessions, while postsurveys coincided with reward collection visits, ensuring effective data gathering.

Study Procedures

We conducted the human subjects experiment for 12 weeks (Figure 2A). During this period, 133 participants (1) engaged in daily message exchanges through *Our Town* and (2) completed 4 daily sessions of *Sammy Talk*. Each day at 8 AM, participants received their daily activity instructions. Weekday activities consisted of 4 preselected games, whereas during weekends and holidays, the system delivered dementia prevention information and notifications at 8 AM. We conducted the 12-week intervention in a structured sequence of activities. The flowchart illustrates how customized cognitive training in *Sammy Talk* and group-based interactions in *Our Town* progressed in parallel and reinforced each other (Figure 2B). Participants maintained *Sammy Talk* engagement throughout the 12-week period. All participants completed these activities in their homes (Figure 2C).

XSL•FO
RenderX

Figure 2. Overview of the study procedures. (A) Measurement tool according to the experimental procedure. (B) Timeline of participant engagement across *Sammy Talk* and *Our Town* during the 12-week intervention (C) Facilitator interactions in *Our Town*. CIST: Cognitive Impairment Screening Test; K-MMSE-2: Korean-Mini Mental State Examination—Version 2; MOS-SSS: Medical Outcomes Study–Social Support Survey; SGDS-K: Short Form Geriatric Depression Scale—Korean version; TWEETS: Twente Engagement with eHealth Technologies Scale.



Statistical Analysis

We analyzed data from 133 participants using SPSS (version 28.0; IBM SPSS Inc). All participants completed the 12-week intervention, which included cognitive training games on *Sammy Talk* and the group chat platform *Our Town*. The study divided participants into 2 groups based on their engagement levels: active users (n=66) and nonactive users (n=67). Total usage included both *Sammy Talk* (which counted the number of games played) and *Our Town* (which counted the number of group chat posts). The median value of 65 sessions served as the dividing point: participants who recorded more than 65 sessions were classified as the active group, and those who recorded 65 sessions or fewer were classified as the nonactive group. We applied a “50% of the time” threshold, which ensured an even distribution between the 2 groups. This approach aligns with the methodologies used by Horwitz et al [56], who categorized participants into top and bottom 50% engagement groups in digital mental health interventions. This approach was also supported by Yardley et al [57] and Bucci et al [58], who identified that higher engagement correlated with improved clinical outcomes.

For comprehensive understanding of the impact of the training regarding the K-MMSE-2 scores, we additionally conducted a sensitivity analysis that categorized participants into lower (n=66) and higher (n=67) K-MMSE-2 groups, using baseline K-MMSE-2 median and SD. Participants with baseline K-MMSE-2 scores ≤27, which represented the median score, were categorized as the lower K-MMSE-2 group, while those who scored >27 were categorized as the higher K-MMSE-2

XSL•FO
RenderX

group. We also applied an alternative classification: participants whose baseline scores fell more than 1 SD below the mean were classified as having lower K-MMSE-2 status.

The analysis consisted of several components. Descriptive statistical analyses provided demographic information. Paired-sample *t* tests identified significant differences in cognitive function, emotional status, and participation between pre- and postintervention periods. Additional paired *t* tests compared outcomes between lower- and higher K-MMSE-2 groups and between high- and low engagement groups. Path analysis, through simple regression analyses, examined how group interventions affected *Sammy Talk* engagement, social support improvement, and cognitive function. Repeated-measures ANOVA assessed the mean and variability of daily *Sammy Talk* and *Our Town* usage across 3 months. Descriptive statistics supplemented this assessment for detailed data interpretation.

Ethical Considerations

This study was approved by the Yonsei University IRB (IRB no. 7001988 - 202401-HR-1397-06). All participants provided written informed consent prior to participation. The collected data were fully anonymized and securely stored, with access restricted to the research team. Participants received small incentives, such as gift cards worth KRW 5000 (approximately US $3.70) and restaurant vouchers worth KRW 10,000 (approximately US $7.40), based on the exchange rate at the time of the study, as compensation for their participation.

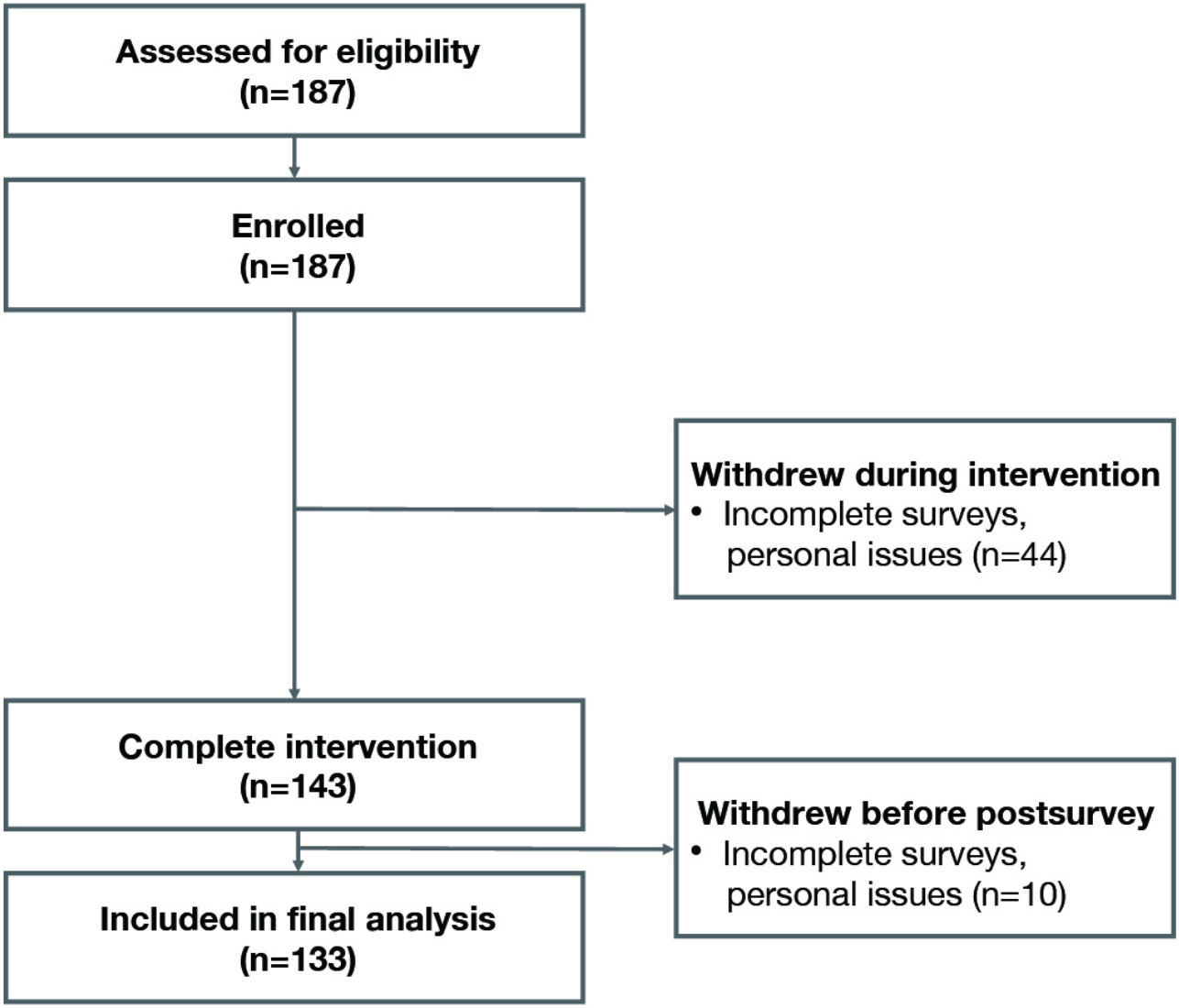
Results

Baseline Characteristics of the Participants

Initially, we recruited 187 participants for the study. Of these, 44 participants withdrew during the intervention period from the primary measure because they did not complete the survey or due to personal circumstances, leaving 143 participants. After the 12-week intervention, 10 more participants withdrew for similar reasons before completing the second measure, resulting in 133 participants available for final statistical analysis. Participants who did not complete the postintervention survey (n=54) were excluded from the final analysis. No statistical adjustments were applied for missing data from participants who withdrew before completion. Therefore, all statistical analyses were conducted using complete cases only. Figure 3 illustrates the flow of participant recruitment, retention, and exclusion throughout the study.

XSL•FO
RenderX

Figure 3. Flow diagram of study participants.



These 133 participants ranged in age from 55 to 75 years (mean 64.76, SD 6.76 years). Based on previous studies reporting medium effect sizes (Hedges *g*=0.5) and using G*Power (α=.05, power=0.80), we estimated that a minimum sample size of 128 would be required. This estimate is also consistent with the numbers of participants actually analyzed in similar studies involving older adults (100, 103, 145, 132, and 180 participants, respectively) [59-63]. To account for a potential dropout rate of approximately 30%, we aimed to recruit at least 180 participants. Ultimately, 187 were enrolled (Table 2).

Table . Summary of our study and comparable digital cognitive intervention studies.

	Our study	Research 1 (Ahn and Park, 2016 [59])	Research 2 (Campbell et al, 2020 [60])	Research 3 (Chapman et al, 2021 [61])	Research 4 (Nakada et al, 2024 [62])	Research 5 (Laane et al, 2023 [63])
Study design	Single-arm	Single-arm	Single-arm	Single-arm	Single-arm	Single-arm
Sample size	187	100	103	145	132 (pilot)	180
Effect size	Hedges *g*= 0.5	N/A[a] (moderate improvement)	N/A (focus on correlations)	N/A (improvement reported)	N/A (improvement reported)	N/A (improvement reported)
P value	<.05	<.05	<.05	<.05	<.05	<.05

[a]N/A: not applicable.

The demographic profile showed that out of 133 participants, 69 (51.9%) were female, 119 (89.5%) had been married, and 51 (38.3%) held elementary school diplomas. Among them, out of 133 participants, 91 (68.4%) reported living with a spouse (Table 3). No missing data were observed for the baseline demographic variables, including cognitive function score (K-MMSE-2), depression score (SGDS-K), social support measure (MOS-SSS), or participation measure (TWEETS),

XSL•FO
RenderX

among the participants who were included in the final analysis. Participants who had incomplete outcome data (n=54) were excluded from the analysis.

Table . Demographic characteristics of the participants (N=133).

Demographic characteristics	Values
Age (years)	
Mean (SD)	64.76 (6.76)
50 - 59, n (%)	32 (24.1)
60 - 69, n (%)	67 (50.4)
70 - 79, n (%)	34 (24.6)
Sex, n (%)	
Male	64 (48.1)
Female	69 (51.9)
Marital status, n (%)	
Married	119 (89.5)
Unmarried	14 (10.5)
Education, n (%)	
No education	2 (1.5)
Elementary school diploma	51 (38.3)
Middle school diploma	33 (24.8)
High school diploma	30 (22.6)
Community college diploma	3 (2.3)
College diploma and professional degree	14 (13.4)
Cohabitation type, n (%)	
Living alone	5 (3.8)
Married couple	91 (68.4)
Married couple with children	31 (23.3)
Living with relatives other than spouse/children (eg, siblings) or with nonrelatives	6 (4.5)

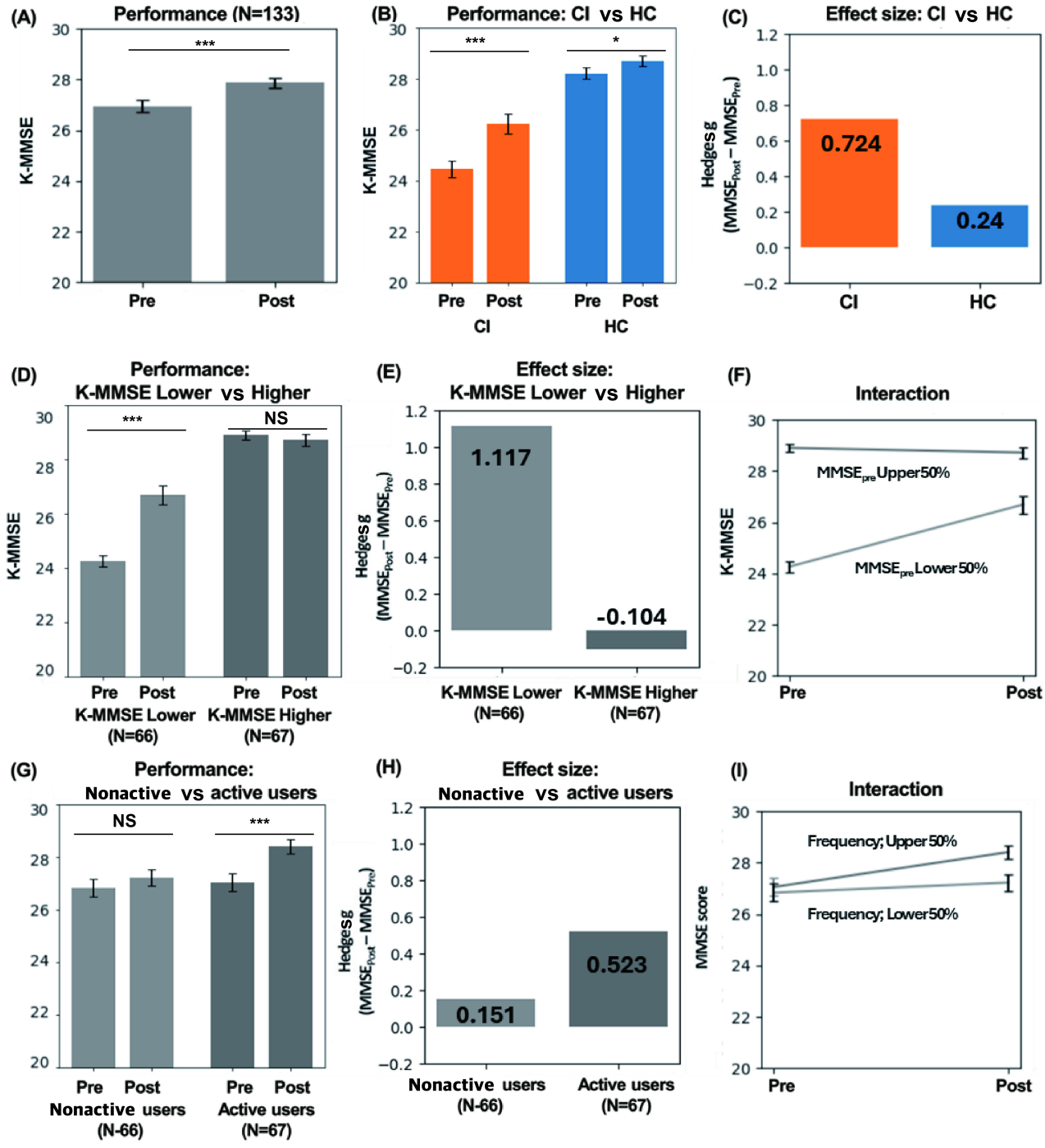
Primary Result

Overall Change in Cognitive Function

For all 133 participants, K-MMSE-2 scores showed significant improvement after using the *Care & Cure* program (Figure 4A; pretrial: mean 26.95, SD 2.77; posttrial: mean 27.87, SD 2.46; t_{132}=−4.106, *P*<.001) with a weak effect size (Hedges *g*=0.350). The cognitive impairment risk group included 15 participants with baseline K-MMSE-2 scores ranging from 19 to 23, corresponding to the K-MMSE-2 normative range for mild cognitive impairment (18-23). The healthy control (HC) group included 118 participants with baseline scores of 24 or higher, corresponding to the cognitively normal range in the K-MMSE-2 norms. The cognitive impairment group showed greater improvement in cognitive function than the HC group (Figure 4B; cognitive impairment: *P*<.001; HC: *P*<.05). Effect sizes for each group are presented in Figure 4C. The cognitive impairment group demonstrated a strong effect size (Hedges *g*=0.724), whereas the HC group showed a weak effect (Hedges *g*=0.240).

XSL•FO
RenderX

Figure 4. Overall results. (A) Pre- and postintervention comparison of cognitive function. (B) Performance enhancement comparing participants in the cognitive impairment (CI) risk group (K-MMSE-2 scores 19 - 23) versus those in the HC group (K-MMSE-2 scores ≥24). (C) Effect size. (D) Performance comparison between participants below and above baseline K-MMSE-2 median scores. (E) Effect size. (F) Interaction effect. (G) Performance comparison between low (nonactive) and high (active) app usage frequency groups (based on end-of-experiment user logs). (H) Effect size. (I) Interaction effect. We note that "K-MMSE" in the plot refers to K-MMSE-2 (independent samples *t* test; **P*<.05,***P*<.01,****P*<.001; note: K-MMSE-2 baseline values were adjusted from 0 to 23 - 26 for readability; error bars indicate standard error of the mean). Orange bars indicate the CI group, blue bars indicate the HC group, and gray bars represent overall or subgroup comparisons. CI: cognitive impairment; HC: healthy control; K-MMSE-2: Korean-Mini Mental State Examination—Version 2; MMSE: Mini Mental State Examination; NS: not significant.



Performance Changes in K-MMSE-2 Subject to Baseline Cognitive Function

Similar patterns emerged between groups categorized by baseline K-MMSE-2 median scores. For sensitivity analysis, we divided participants into 2 groups based on their baseline K-MMSE-2 scores: the lower K-MMSE-2 group (n=66) and the higher K-MMSE-2 group (n=67). Participants whose K-MMSE-2 scores fell below the baseline median formed a lower group, while the remaining participants were assigned to the higher group.

Independent samples *t* tests examined cognitive improvement differences between groups. The analysis showed that the K-MMSE-2 lower group demonstrated significant score increases posttreatment (independent *t* test; *P*<.001), whereas the K-MMSE-2 higher group showed no statistically significant

XSL•FO
RenderX

changes (independent *t* test; *P*<.241) (Figure 4D). Effect size analysis showed substantial impact in the K-MMSE-2 lower group (Hedges *g*=1.117), while the higher group displayed an ignorable negative effect (Hedges *g*=−0.104; Figure 4E). The interaction effect between baseline cognitive level and time is presented in Figure 4F.

Performance Changes in K-MMSE-2 Based on the Application Usage Frequency

The active user group (n=67) showed significant increases in K-MMSE-2 scores after the intervention, whereas the nonactive user group (n=66) did not show statistically significant changes in K-MMSE-2 scores (independent *t* test; *P*<.001) (Figure 4G). The lower frequency group (nonactive group) showed a weak effect (Hedges *g*=0.151), while the higher frequency group demonstrated a moderate effect (Hedges *g*=0.523; Figure 4H). The interaction effect between usage frequency and time is presented in Figure 4I.

Performance Differences in K-MMSE-2 According to Age

To investigate the influence of age, we again divided participants once more into 2 age groups based on the median age (66 years): the lower age group (age ≤66 years; n=70) and the higher age group (age>66 years; n=63). Independent *t* tests examined cognitive improvement differences using K-MMSE-2 Delta scores. Both age groups showed increased K-MMSE-2 Delta scores, although these changes lacked statistical significance (lower age group: mean 0.96, SD 2.27 years; higher age group: mean 0.87, SD 2.90 years; t_{131}=−0.187, *P* value not applicable). Effect size analysis revealed negligible differences between groups (Hedges *g*=0.03). See Figure S2 in Multimedia Appendix 1 for more details.

Secondary Result

Impact of Group Intervention on Engagement, Social Support, and Cognitive Function

To investigate the impact of group chat usage and cognitive training on cognitive function, simple regression analyses were conducted on 2 separate occasions. In the initial regression model, usage of the *Our Town* (group chat), measured as the number of interactions, significantly influenced usage of cognitive training ($F_{1,131}$=153.646, *P*<.001). The explanatory power of the regression model was approximately 54.0% (R^2=0.540). The Durbin-Watson statistic was 1.502, which is close to 2, indicating no significant concerns regarding the assumption of independence of residuals. Regarding the significance of the regression coefficients, engagement in *Our Town* (group chat) demonstrated a statistically significant positive effect on the usage of cognitive training (β=.735, *P*<.001). This finding suggests a positive and significant relationship between increased engagement in the group chat and greater usage of cognitive training tools.

In the second regression model, usage of the cognitive training tools, measured as total log time, did not significantly influence improvement in cognitive function, as assessed by the K-MMSE-2 Delta scores ($F_{1,131}$=0.158, *P*=.692). The explanatory power of the second regression model was minimal, accounting for only 0.1% (R^2=0.001). The Durbin-Watson statistic was 2.050, which is close to 2, indicating no significant concerns regarding the assumption of independence of residuals. Regarding the significance of the regression coefficients, cognitive training usage did not statistically significantly influence cognitive function improvement (β=−.035, *P*=.692). This finding suggests that cognitive training usage did not contribute to cognitive function enhancement (Figure 5 and Table 4).

Figure 5. Impact of group intervention on engagement and cognitive function. K-MMSE-2: Korean-Mini Mental State Examination—Version 2; NS: not significant.



XSL•FO
RenderX

Table . Summary of regression models examining group chat usage, cognitive training, and cognitive function.

	Model	*B*	SE	β	*t* test (*df*)	*P* value	*F* test (*df*); *P* value	R^2	Durbin-Watson (95% CI)
1	Group chat usage → cognitive training usage	1.138	0.092	.735	12.395 (131)	<.001	153.646 (1,131); <.001	0.54	1.138(0.956 to 1.320)
2	Cognitive training usage → cognitive function improvement	−0.001	0.003	−.035	−0.397 (131)	.69	0.158 (1,131); .69	0.001	-0.001(-0.007 to 0.005)

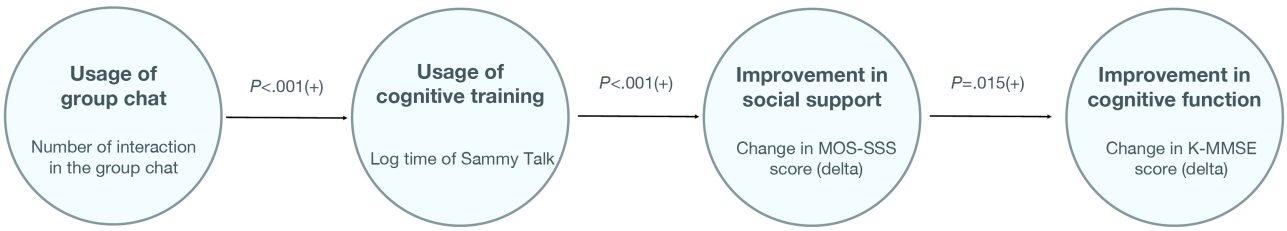
We conducted additional regression analyses to examine how social support mediated the relationships between group chat usage, cognitive training, and cognitive function improvement. In the initial regression model, *Our Town* usage (measured by the number of interactions) significantly influenced cognitive training usage ($F_{1,131}$=153.646, *P*<.001). The explanatory power of the regression model was approximately 54.0% (R^2=0.540). The Durbin-Watson statistic was 1.502, which is close to 2, indicating no significant concerns regarding the assumption of independence of residuals. Regarding the significance of the regression coefficients, engagement in *Our Town* (group chat) demonstrated a statistically significant positive effect on the usage of cognitive training (β=.735, *P*<.001). This finding suggests a clear positive relationship between increased engagement in the group chat and greater usage of cognitive training tools.

In the second regression model, frequent cognitive training usage, defined as total interaction time with the conversational agent, significantly influenced social support improvement, as indicated by social support scores ($F_{1,131}$=16.830, *P*<.001). The explanatory power of the second regression model was 11.4% (R^2=0.114). The Durbin-Watson statistic was 1.946, which is close to 2, indicating no significant concerns regarding the assumption of independence of residuals. Regarding the significance of the regression coefficients, frequent engagement with the conversational agent demonstrated a statistically significant positive effect on social support improvement (β=.337, *P*<.001). This finding suggests that participants who interacted more frequently with the conversational agent reported greater perceived social support.

In the third regression model, improvements in social support had a statistically significant effect on cognitive function enhancement, as measured by the K-MMSE-2 Delta scores ($F_{1,131}$=6.830, *P*<.05). The explanatory power of the regression model was 4.4% (R^2=0.044). The Durbin-Watson statistic was 2.084, which is close to 2, indicating no significant concerns regarding the assumption of independence of residuals. Regarding the significance of the regression coefficients, higher social support scores were associated with greater improvements in cognitive function (β=.211, *P*<.05). This finding suggests that participants with strong social support experienced greater gains in cognitive function.

Conversely, cognitive training usage did not have a statistically significant impact on cognitive function enhancement (*P*=.692). These findings suggest that cognitive function does not improve when individuals use cognitive games as a stand-alone intervention. However, combining cognitive games with group interventions that increase social support and motivation leads to cognitive function improvement (Figure 6 and Table 5).

Figure 6. Impact of group intervention on engagement, social support, and cognitive function (as a part of path analysis). K-MMSE: Korean-Mini Mental State Examination; MMSE: Mini Mental State Examination.



XSL•FO
RenderX

Table . Summary of regression analyses examining the mediating role of social support.

	Model	*B*	SE	β	*t* test (*df*)	*P* value	*F* test (*df*); *P* value	R^2	Durbin-Watson (95% CI)
1	Group chat usage → cognitive training usage	1.138	0.092	.735	12.395 (131)	<.001	153.646 (1,131); <.001	0.540	1.138(0.956 to 1.320)
2	Cognitive training usage → social support improvement	0.061	0.015	.337	4.102 (131)	<.001	16.830 (1,131); <.001	0.114	0.061 (0.031 to 0.091)
3	Social support improvement → cognitive function improvement	0.034	0.014	.211	2.467 (131)	<.05	6.084 (1,131); <.05	0.044	0.034(0.006 to 0.062)

Change in Social Support, Willingness to Participation, and Depression

Social support was divided into four categories: (1) emotional or informational, (2) material, (3) affectional, and (4) positive social interaction support. Emotional or informational support showed the highest pre-post difference with statistical significance (t_{132}=−6.509, *P*<.001) and a moderately strong effect size (Hedges *g*=0.70). Willingness to participate was assessed across 3 dimensions: affect, behavior, and cognition. Affect captured participants' emotional attitudes toward their involvement, such as expressing enjoyment when using *Sammy Talk* for cognitive training or showing enthusiasm about interacting with other members in *Our Town*. Next, behavior referred to observable participation such as attendance or task engagement, including frequency of playing arithmetic and memory games in *Sammy Talk* or actively posting messages and responding to others in *Our Town* social platform. Finally, cognition reflected participants' beliefs and understanding of the program's value, such as recognizing the *Sammy Talk*'s cognitive exercises could improve their mental abilities or believing that social interactions in *Our Town* could enhance their well-being and reduce isolation. Willingness to participate demonstrated significant differences in cognitive factors (t_{132}=2.159, *P*<.05) with a nearly moderate effect size (Hedges *g*=0.43). Within the subcategories of willingness to participate, affect showed significant pre-post differences (t_{132}=2.008, *P*<.05) with a moderate effect size (Hedges *g*=0.51), whereas behavioral measures showed no significant differences (t_{132}=0.244, *P*=.81). Depression levels exhibited significant pre-post differences (t_{132}=3.093, *P*<.01) with a weak effect size (Hedges *g*=0.21). Table 6 presents these results in summary.

Table . Comparison of social support, emotional relief with program use before and after.

Parameters	Week 0 (n=133), mean (SD)	Week 12 (n=133), mean (SD)	Hedges *g*	*t* test (*df*)	*P* value
Social support					
Emotional/informational support	30.19 (7.04)	34.10 (5.43)	0.70	−6.509 (132)	<.001
Material support	15.18 (3.78)	16.25 (3.06)	0.37	−3.345 (132)	<.01
Affectionate support	10.81 (2.86)	11.87 (2.56)	0.29	−4.256 (132)	<.001
Positive social interaction support	15.23 (3.76)	16.43 (3.00)	0.37	−3.803 (132)	<.001
Willingness to participation					
Behavior	10.45 (3.60)	10.35 (3.77)	0.47	0.244 (132)	.81
Cognition	12.08 (3.34)	11.27 (3.58)	0.43	2.159 (132)	<.05
Affect	10.78 (3.90)	9.90 (4.02)	0.51	2.008 (132)	<.05
Depression					
SGDS-K[a]	2.77 (2.59)	2.23 (2.13)	0.21	3.093 (132)	<.01

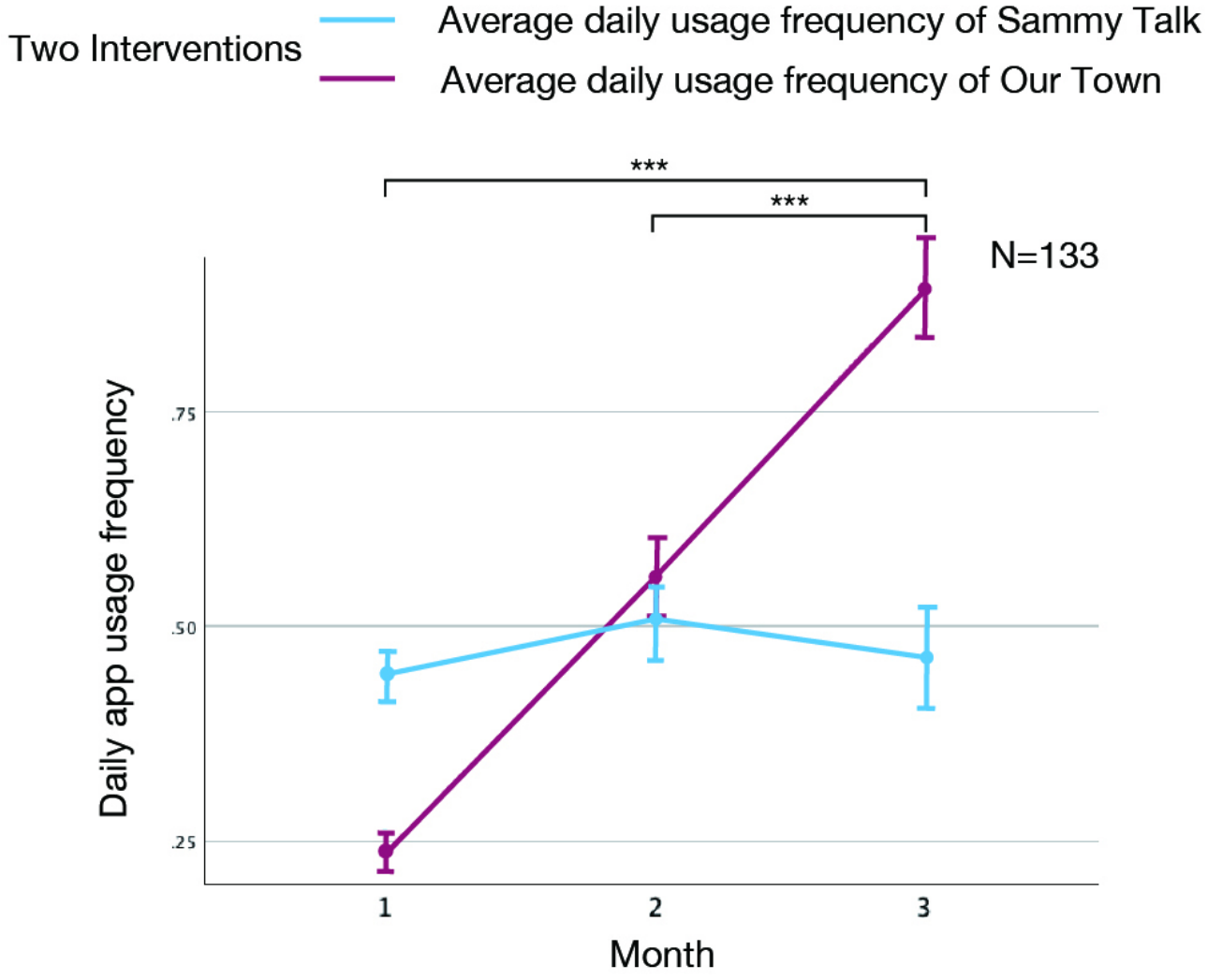
[a]SGDS-K: Short Form Geriatric Depression Scale—Korean version.

XSL•FO
RenderX

Engagement Analysis: Sammy Talk and Our Town

To examine the engagement patterns, we conducted a repeated-measures ANOVA on the average daily usage frequency of *Sammy Talk* (conversation-based intervention) and *Our Town* (community-based intervention) over 3 months (Figure 7). The analysis showed significant differences in average daily usage frequency between *Sammy Talk* and *Our Town* during the third month (P<.001). Results indicated a significant main effect of usage duration ($F_{2,264}$=20.282, P<.001) and a significant interaction between intervention types (eg, *Sammy Talk* and *Our Town*) and usage duration ($F_{2,264}$=17.859, P<.001). *Our Town* demonstrated steadily increasing average daily usage, which reached the highest in month 3 (mean 0.90, SD 0.11), followed by month 2 (mean 0.56, SD0.08) and month 1 (mean 0.24, SD 0.05). In contrast, *Sammy Talk* usage frequency varied, with highest engagement observed in month 2 (mean 0.52, SD 0.08), followed by month 3 (mean 0.47, SD0.11) and month 1 (mean 0.42, SD 0.05).

Figure 7. Changes in participant *Sammy Talk* and *Our Town* usage frequency following intervention implementation. *P<.05,**P<.01,***P<.001.



Discussion

Principal Findings

The *Care & Cure* program demonstrated effectiveness in improving cognitive function across all 133 participants, with particularly strong enhancement among individuals with lower K-MMSE-2 scores. This aligns with the program's evidence-based development, which incorporated clinical expertise from neurologists at Mokdong Ewha Womans University Hospital.

The *Sammy Talk* content followed validated clinical frameworks across multiple cognitive domains. The calculation games were derived from the *WRAT-3 Arithmetic subset* and *Digit Symbol Substitution Test* [41,42]. Visuospatial tasks incorporated the *Mental Rotation Test* [43], whereas attention-focused games used crossword puzzles and the *Montreal Cognitive Assessment letter—a tapping test* [44,45]. Executive function games were adapted from the *University of California, San Diego Performance-Based Skills Assessment* (*UPSA-3*) and the *Trail Making Test* [46,47]. Memory tasks integrated the *Korean Naming Test*, *Chinese idioms*, and *Rey-Osterrieth Complex Figure Test* [48-51]. Language games incorporated the *Montreal Cognitive Assessment* initial letter fluency test and *Scrambled Sentences Test* [44,52]. However, the specific cognitive impact of individual game components remains to be clarified in future research.

Participants with lower baseline K-MMSE-2 scores (hereafter referred to as the lower group) showed larger effect sizes, as

XSL•FO
RenderX

they had greater potential for measurable improvement. By contrast, those with higher baseline scores (hereafter referred to as the higher group) showed little change, having already approached their cognitive performance ceiling. This difference reflects a well-known ceiling effect among cognitively normal participants. Previous studies examining short-term interventions have consistently reported that people with cognitively normal conditions achieve only limited gains, whereas those with cognitive impairment show greater improvement [64-66]. Moreover, participants in the lower group faced a higher risk of decline and therefore benefited more from the clinically validated content designed by neurologists to target multiple cognitive domains, including calculation, attention, language, memory, visuospatial ability, and executive function. This finding is consistent with previous research using validated clinical tasks, which confirmed that participants in the lower group achieved greater improvements than those in the higher group [64,65].

Higher usage frequency of *Care & Cure* corresponded with enhanced cognitive function. *Our Town* appeared to moderate this effect by increasing *Sammy Talk* engagement. Participants who engaged more often in *Our Town* engagement dedicated more time to *Sammy Talk* and achieved superior cognitive function scores. The group chat format likely enhanced engagement through shared performance goals, achievement reports, and peer encouragement. These findings align with previous research on group chat–based cognitive enhancement programs [31,32].

Nonetheless, *Our Town* did not directly enhance cognitive function; rather, its impact emerged through social support mechanisms. The integration of *Our Town* and *Sammy Talk* (a software robot) in the *Care & Cure* program enhanced social support, which in turn improved cognitive function. While *Our Town* did not bring about direct cognitive benefits, it served as an indirect motivational means by sending consistent reminder messages to participants. Participants exchanged greetings and shared daily experiences in *Our Town* while learning practical knowledge and conversation-related vocabulary through *Sammy Talk*. This encouraged more frequent communication both online and offline. These repeated interactions in turn created a social support network through daily greetings, experience sharing, and conversational practice. The AI chatbot provided additional support by offering affective encouragement and fostering positive exchanges, leading participants to report higher social support scores. This finding aligns well with previous studies demonstrating that AI chatbots deliver social support and reduce depressive symptoms among older adults [67,68]. Social support and emotional stability function as preventive factors that preserve cognitive function by reducing loneliness and chronic stress, both recognized contributors to memory decline and executive dysfunction [67-71]. Conversations and social interactions demand memory recall, language use, and problem-solving, repeatedly stimulating cognitive processes and strengthening cognitive reserve [72-74].

On a secondary issue, the *Care & Cure* program effectively reduced emotional distress, with participants experiencing emotional changes through enhanced social support. Program engagement, measured through *Our Town* chat interactions, showed increasing participant responses to the chatbot-based software agent, reaching 3.8 times the initial level by month 3 (Figure 6). This increased engagement suggested more frequent conversations with both the software agent and other participants, potentially facilitating positive social exchanges and providing emotional, affectionate, and informational support. However, chat interaction frequency alone provides limited insight into intervention adherence. Future research should examine more precise adherence measurements. The study suggests that increased social support influenced participants' depressive symptoms [75-78].

Postanalysis on Dropout

Nearly 3 out of 10 participants (28.8%) dropped out of the study, with education emerging as the primary differentiating factor. Education, rather than age or sex, accounted for this difference, as people who completed the intervention had significantly more years of education than those who dropped out (t_{131}=2.43, P=.016). These results show that education level, which reflects accumulated social and cognitive resources, was associated with dropout, whereas age and sex were not. Participants with limited education and digital literacy often struggled to use the digital cognitive training program. Lower educational attainment, which often limits digital literacy, reduced participants' ability to use messenger applications and type messages, preventing them from joining group chats or completing training tasks [18,20,21]. Those who lacked primary school education and could not perform basic arithmetic operations were unable to complete calculations tasks and eventually dropped out. Participants with very low education levels may therefore be more likely to drop out of digital cognitive training programs [18,20,21]. Although not statistically significant, males were likely to drop out more frequently than females, suggesting that future studies should consider strategies to improve male retention (see Multimedia Appendix 1, Section VI for more details).

Limitations

Despite the significant improvement in participants' cognitive capacity through the use of our proposal in general, the research has some limitations to address. First, the single-arm design without a control group reduces confidence in attributing changes to the intervention. Without a comparison group, it would be very difficult to establish that observed improvements resulted solely from the intervention, as changes might be able to reflect natural progression, practice effects from repeated testing, or placebo effects. In addition, it would be challenging to determine whether improvements are clinically meaningful compared with standard care or alternative interventions. Second, participant interaction with the software robot in *Our Town* (the group chat component) was lower than anticipated, with its primary effect manifesting through increased usage of *Sammy Talk*. Third, the sample was drawn from a single midsize Korean city with relatively homogeneous demographic characteristics, limiting the generalizability of findings to other regions or populations. Fourth, the program's reliance on *KakaoTalk* as the delivery platform might restrict applicability beyond the Korean context. These methodological considerations highlight areas for future investigation, including randomized controlled trials with appropriate control groups,

XSL•FO
RenderX

multisite studies across diverse populations, and head-to-head comparisons with standard assessment tools to establish clinical superiority.

Conclusions and Future Research Direction

The *Care & Cure* program demonstrated effectiveness for cognitive enhancement, with more frequent participation corresponding to greater improvement. The program showed particular efficacy among cognitively impaired participants. Additional benefits included reduced depressive symptoms, enhanced social support, and increased participation willingness.

Our hypothesis proposed that increased *Our Town* usage would enhance *Sammy Talk* engagement, subsequently improving cognitive function. The data confirmed that higher *Our Town* usage directly increased *Sammy Talk* participation. However, increased *Sammy Talk* usage alone did not yield statistically significant cognitive improvements. When examining social support as a mediator between *Sammy Talk* usage and cognitive enhancement, results indicated that social support directly influenced cognitive improvement. This relationship requires further investigation to understand underlying mechanisms.

While the top 50% of program participants showed greater K-MMSE-2 improvements, the causal relationship between *Sammy Talk* usage and cognitive enhancement lacked statistical significance. Future research should examine the relationship between *Our Town* and *Sammy Talk* usage patterns while addressing the methodological issues raised in this study. Controlled trials with comparison groups are necessary to establish causal effects. For example, a potential future study could compare an intervention group using both *Sammy Talk* and *Our Town*, with a control group attending a traditional face-to-face dementia prevention program. Researchers should also test strategies to increase participation in group chat functions, which serve as sources of social support [79-81]. Gamification approaches, such as small financial rewards, goal setting, or competition elements, may enhance motivation to engage [82,83]. Additionally, context-based reminders drawing on Ecological Momentary Interventions or Ecological Momentary Assessments could encourage more consistent use of both *Sammy Talk* and *Our Town* [84-86].

Next, future studies had better recruit participants from varied regions and sociodemographic groups to extend findings beyond a single Korean midsized city. For instance, researchers could target English-speaking populations from middle-class and low-class households or urban groups with higher digital literacy and educational attainment [87,88]. Moreover, researchers should conduct longitudinal follow-up studies that track participants for at least 1 year to examine whether cognitive benefits persist and to determine whether digital training combined with social support prevents long-term cognitive decline [89,90]. For example, participants could use the intervention for 1 year or more, with outcomes measured at regular intervals to evaluate intervention effectiveness over time. Finally, since this program relied on *KakaoTalk* as its delivery platform, which limits generalizability beyond Korea, future studies should better adapt the same model to other messaging services that provide chatbot Application Programming Interfaces, such as *WeChat*, *WhatsApp*, or *Facebook Messenger*, to test how the intervention performs across different cultural and technological contexts [91,92].

Finally, future works should also take into account education as a covariate, which requires stratified randomization and sensitivity analyses to ensure the robustness of findings. Researchers should also design tasks that assess cognition without relying on arithmetic, so that individuals without formal schooling can participate fully in the longitudinal intervention. Finally, incorporating competitive gamification elements, which may appeal particularly to males who often demonstrate stronger achievement motivation than females, could enhance digital program retention by using leaderboards or quizzes that provide modest rewards and recognize performance [93-95]. By pursuing these directions, future research would be able to clarify how digital cognitive training and social interaction interplay with each other, and identify practical strategies that strengthen engagement and broaden applicability across populations.

Acknowledgments

The manuscript was edited for the English language by Editage, and corrected grammar and typographical errors using DeepL and ChatGPT 4o. All intellectual content, data interpretation, and final approval of the manuscript were conducted by the authors.

Funding

This research was supported by (1) the Basic Science Research Program through the *National Research Foundation of Korea* (*NRF*) funded by the *Ministry of Education* (NRF-2016R1D1A1B02015987), and (2) by the *National Research Foundation of Korea* (*NRF*) grant funded by the *Korea government* (*MSIT*) (RS-2025-23963523).

Data Availability

The datasets generated and analyzed during this study are not publicly available because participant agreements on privacy and confidentiality restrict open access. However, the corresponding author can provide deidentified data to qualified researchers who submit a reasonable request and obtain the necessary institutional approvals.

Authors' Contributions

Conceptualization: BK, JHL, GHK

XSL•FO
RenderX

Methodology: BY, BK, JHL
Data curation: BY, BK, JHL
Formal analysis: BY, JK, JHL
Investigation: BK, WK, SP, HK
Project administration: BK, WK, SP, HK
Supervision: JK, JHL, GHK, BRK
Funding acquisition: JK
Writing – original draft: BY, BK, JK
Writing – review & editing: JK, JHL, GHK
All authors had full access to the study data and took responsibility for the integrity and accuracy of the analysis. BY and BK contributed equally to this work and share first authorship. JHL and GHK contributed equally as co-corresponding authors.

Conflicts of Interest

JK holds the intellectual property rights for the applications used in this study. Additionally, he has a direct financial interest in HAII Inc, where he serves as chief executive officer and shareholder. He has also been a participant in the Suncheon Care & Cure project outside of grant support. The financial relationship between JK and HAII Inc was reviewed by Yonsei University, and his involvement in this research project was approved in accordance with its conflict of interest policies. Authors BY, HK, W. Kim, and SP are employed by HAII Inc. Author BRK is employed by LG Electronics. All authors declare no other competing interests.

Multimedia Appendix 1
Cognitive functions software robot study.
[DOCX File, 611 KB - mhealth_v14i1e75308_app1.docx]

References

1. Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet 2020 Aug 8;396(10248):413-446. [doi: 10.1016/S0140-6736(20)30367-6] [Medline: 32738937]
2. The Lancet Public Health. Reinvigorating the public health response to dementia. Lancet Public Health 2021 Oct;6(10):e696. [doi: 10.1016/S2468-2667(21)00215-2] [Medline: 34563278]
3. Kivipelto M, Mangialasche F, Ngandu T. Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease. Nat Rev Neurol 2018 Nov;14(11):653-666. [doi: 10.1038/s41582-018-0070-3] [Medline: 30291317]
4. Kane R, Fink HA, Brasure M, et al. Interventions to prevent age-related cognitive decline, mild cognitive impairment, and clinical Alzheimer's-type dementia. : Agency for Healthcare Research and Quality; 2017. [Medline: 28759193]
5. Butler M, McCreedy E, Nelson VA, et al. Does cognitive training prevent cognitive decline?: a systematic review. Ann Intern Med 2018 Jan 2;168(1):63-68. [doi: 10.7326/M17-1531] [Medline: 29255842]
6. Gates NJ, Rutjes AW, Di Nisio M, et al. Computerised cognitive training for maintaining cognitive function in cognitively healthy people in midlife. Cochrane Database Syst Rev 2019 Mar 13;3(3):CD012278. [doi: 10.1002/14651858.CD012278.pub2] [Medline: 30864746]
7. Moon SY, Hong CH, Jeong JH, et al. Facility-based and home-based multidomain interventions including cognitive training, exercise, diet, vascular risk management, and motivation for older adults: a randomized controlled feasibility trial. Aging (Albany NY) 2021 Jun 18;13(12):15898-15916. [doi: 10.18632/aging.203213] [Medline: 34148030]
8. Stern Y. Cognitive reserve in ageing and Alzheimer's disease. Lancet Neurol 2012 Nov;11(11):1006-1012. [doi: 10.1016/S1474-4422(12)70191-6] [Medline: 23079557]
9. Stern Y, Barnes CA, Grady C, Jones RN, Raz N. Brain reserve, cognitive reserve, and brain maintenance: issues and prospects. Neurobiol Aging 2019 Feb;83:124-129. [doi: 10.1016/j.neurobiolaging.2019.03.022] [Medline: 31732015]
10. Wimo A, Jönsson L, Bond J, Prince M, Winblad B, Alzheimer Disease International. The worldwide economic impact of dementia 2010. Alzheimers Dement 2013 Jan;9(1):1-11. [doi: 10.1016/j.jalz.2012.11.006] [Medline: 23305821]
11. Jutkowitz E, Kane RL, Gaugler JE, MacLehose RF, Dowd B, Kuntz KM. Societal and family lifetime cost of dementia: Implications for policy. J Am Geriatr Soc 2017 Oct;65(10):2169-2175. [doi: 10.1111/jgs.15043] [Medline: 28815557]
12. Quail Z, Bolton L, Massey K. Digital delivery of non-pharmacological intervention programmes for people living with dementia during the COVID-19 pandemic. BMJ Case Rep 2021 Jun 17;14(6):e242550. [doi: 10.1136/bcr-2021-242550] [Medline: 34140328]
13. Czaja SJ. The potential role of technology in supporting older adults. Public Policy Aging Rep 2017;27(2):44-48. [doi: 10.1093/ppar/prx006]
14. Lampit A, Hallock H, Valenzuela M. Computerized cognitive training in older adults: A meta-analysis. PLoS Med 2014 Jan;11(1). [doi: 10.1371/journal.pmed.1001756] [Medline: 25405755]
15. Friemel TN. The digital divide has grown old: determinants of a digital divide among seniors. New Media Soc 2016 Feb;18(2):313-331. [doi: 10.1177/1461444814538648]

XSL•FO
RenderX

16. Eynon R, Helsper E. Adults learning online: digital choice and/or digital exclusion? New Media Soc 2011 Jun;13(4):534-551. [doi: 10.1177/1461444810374789]
17. D'Onofrio G, Sancarlo D, Ricciardi F, et al. Information and communication technologies for the activities of daily living in older patients with dementia: a systematic review. J Alzheimers Dis 2017;57(3):927-935. [doi: 10.3233/JAD-161145] [Medline: 28304297]
18. Contreras-Somoza LM, Irazoki E, Castilla D, et al. Study on the acceptability of an ICT platform for older adults with mild cognitive impairment. J Med Syst 2020 May 25;44(7):120. [doi: 10.1007/s10916-020-01566-x] [Medline: 32451740]
19. Chen K, Chan AHS. Use or non-use of gerontechnology—a qualitative Study. Int J Environ Res Public Health 2014 Aug;10(10):4645-4666. [doi: 10.3390/ijerph10104645]
20. Callisaya ML, Jayakody O, Vaidya A, Srikanth V, Farrow M, Delbaere K. A novel cognitive-motor exercise program delivered via a tablet to improve mobility in older people with cognitive impairment—StandingTall Cognition and Mobility. Exp Gerontol 2021 Sep;152:111434. [doi: 10.1016/j.exger.2021.111434] [Medline: 34098009]
21. Tsai YIP, Beh J, Ganderton C, Pranata A. Digital interventions for healthy ageing and cognitive health in older adults: a systematic review of mixed method studies and meta-analysis. BMC Geriatr 2024 Mar 4;24(1):217. [doi: 10.1186/s12877-023-04617-3] [Medline: 38438870]
22. Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 1;25(9):1248-1258. [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
23. Rotenberg S, Anderson ND, Binns MA, et al. Effectiveness of a meta-cognitive group intervention for older adults with subjective cognitive decline or mild cognitive impairment: the ASPIRE randomized controlled trial. J Prev Alzheimers Dis 2024;11(6):1534-1548. [doi: 10.14283/jpad.2024.166] [Medline: 39559867]
24. Bickmore TW, Pfeifer LM, Byron D, et al. Usability of conversational agents by patients with inadequate health literacy: evidence from two clinical trials. J Health Commun 2010;15 Suppl 2(Suppl 2):197-210. [doi: 10.1080/10810730.2010.499991] [Medline: 20845204]
25. Li H, Zhang R, Lee YC, Kraut RE, Mohr DC. Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. NPJ Digit Med 2023 Dec 19;6(1):236. [doi: 10.1038/s41746-023-00979-5] [Medline: 38114588]
26. Zhang Z, Bickmore T. Medical shared decision making with a virtual agent. Presented at: Proceedings of the 18th International Conference on Intelligent Virtual Agents; IVA'18; Nov 5-8, 2018; Sydney, NSW, Australia p. 113-118. [doi: 10.1145/3267851.3267883]
27. Kim H, Ray CD, Veluscek AM. Complementary support from facilitators and peers for promoting mHealth engagement and weight loss. J Health Commun 2017 Nov;22(11):905-912. [doi: 10.1080/10810730.2017.1373876] [Medline: 29125393]
28. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 6;4(2):e19. [doi: 10.2196/mental.7785] [Medline: 28588005]
29. Ly KH, Ly AM, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. Internet Interv 2017 Dec;10:39-46. [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]
30. Edelman D, et al. Shared Medical Appointments for Chronic Medical Conditions: A Systematic Review: Department of Veterans Affairs (US) Office of Research and Development; 2012. [Medline: 22916369]
31. Kirsh SR, Aron DC, Johnson KD, et al. A realist review of shared medical appointments: how, for whom, and under what circumstances do they work? BMC Health Serv Res 2017 Feb 4;17(1):113. [doi: 10.1186/s12913-017-2064-z] [Medline: 28160771]
32. Drake C, Kirk JK, Buse JB, et al. Characteristics and delivery of diabetes shared medical appointments in North Carolina. N C Med J 2019;80(5):261-268. [doi: 10.18043/ncm.80.5.261] [Medline: 31471505]
33. Introducing a shared consultation system to revolutionize healthcare delivery. Korea Health Industry Promotion Agency. 2019. URL: https://www.khidi.or.kr [accessed 2026-02-19]
34. Suncheon demographics and city profile. Suncheon City Government. 2023. URL: https://www.suncheon.go.kr [accessed 2026-02-19]
35. Aging population in South Jeolla province. Korea Statistical Information Service (KOSIS). 2022. URL: https://kosis.kr [accessed 2026-02-19]
36. 2014 internet usage survey. Korea Internet & Security Agency. 2014. URL: http://www.kisa.or.kr [accessed 2026-04-30]
37. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848. [doi: 10.2196/18848] [Medline: 32213476]
38. Nishida T, Nakazawa A, Ohmoto Y, Mohammad Y. Conversational Informatics: A Data-Intensive Approach with Emphasis on Nonverbal Communication: Springer Japan; 2014:344. [doi: 10.1007/978-4-431-55040-2_1]
39. Kocaballi AB, Quiroz JC, Laranjo L, et al. Conversational agents for health and wellbeing. Presented at: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems; Apr 25-30, 2020; Honolulu, HI p. 1-8. [doi: 10.1145/3334480.3375154]

XSL•FO
RenderX

40. Richter KM, Mödden C, Eling P, Hildebrandt H. Working memory training and semantic structuring improves remembering future events, not past events. Neurorehabil Neural Repair 2015 Jan;29(1):33-40. [doi: 10.1177/1545968314527352] [Medline: 24699430]
41. Martin RC, Annis SM, Darling LZ, Wadley V, Harrell L, Marson DC. Loss of calculation abilities in patients with mild and moderate Alzheimer disease. Arch Neurol 2003 Nov;60(11):1585-1589. [doi: 10.1001/archneur.60.11.1585] [Medline: 14623731]
42. Wechsler D. WAIS-III Administration and Scoring Manual: Psychological Corporation; 1997.
43. Shepard RN, Metzler J. Mental rotation of three-dimensional objects. Science 1971 Feb 19;171(3972):701-703. [doi: 10.1126/science.171.3972.701] [Medline: 5540314]
44. Mayas J, Parmentier FBR, Andrés P, Ballesteros S. Plasticity of attentional functions in older adults after non-action video game training: a randomized controlled trial. PLoS One 2014;9(3):e92269. [doi: 10.1371/journal.pone.0092269] [Medline: 24647551]
45. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
46. Goldberg TE, Harvey PD, Devanand DP, Keefe RSE, Gomar JJ. Development of an UPSA short form for use in longitudinal studies in the early Alzheimer's disease spectrum. J Prev Alzheimers Dis 2020;7(3):179-183. [doi: 10.14283/jpad.2019.51] [Medline: 32463071]
47. Bowie CR, Harvey PD. Administration and interpretation of the Trail Making Test. Nat Protoc 2006;1(5):2277-2281. [doi: 10.1038/nprot.2006.390] [Medline: 17406468]
48. Galante E, Venturini G, Fiaccadori C. Computer-based cognitive intervention for dementia: preliminary results of a randomized clinical trial. G Ital Med Lav Ergon 2007;29(3 Suppl B):B26-B32. [Medline: 18575355]
49. Kim TY, Kim SY, Choi SH, et al. Development of the Korean naming test and estimation of its validity and reliability. Dement Neurocogn Disord 2004;3(2):117-123 [FREE Full text]
50. Kwok T, Wong A, Chan G, et al. Effectiveness of cognitive training for Chinese elderly in Hong Kong. Clin Interv Aging 2013;8:213-219. [doi: 10.2147/CIA.S38070] [Medline: 23440076]
51. Schreiber M. Potential of an interactive computer-based training in the rehabilitation of dementia: an initial study. Neuropsychol Rehabil 1999 Apr;9(2):155-167. [doi: 10.1080/713755596]
52. Bernini S, Alloni A, Panzarasa S, et al. A computer-based cognitive training in mild cognitive impairment in Parkinson's disease. Neurorehabilitation 2019;44(4):555-567. [doi: 10.3233/NRE-192714] [Medline: 31256092]
53. Kwon YC, Park JH. Korean version of Mini-Mental State Examination (MMSE-K): part I. Development of the test for the elderly. J Korean Neuropsychiatr Assoc 1989;28(1):125-135 [FREE Full text]
54. Cho MJ, Bae JN, Suh GH, et al. Validation of geriatric depression scale, Korean version (GDS) in the assessment of DSM-III-R major depression. J Korean Neuropsychiatr Assoc 1999;38(1):48-63 [FREE Full text]
55. Lim MK, Kim MH, Kim SS, Kang S, Lee SY. Social support and self-rated health status in a low income neighborhood of Seoul, Korea. Korean J Prev Med Public Health 2003;36(1):54-62 [FREE Full text]
56. Horwitz AG, Mills ED, Sen S, Bohnert ASB. Comparative effectiveness of three digital interventions for adults seeking psychiatric services: a randomized clinical trial. JAMA Netw Open 2024 Jul 1;7(7):e2422115. [doi: 10.1001/jamanetworkopen.2024.22115] [Medline: 39023893]
57. Yardley L, Spring BJ, Riper H, et al. Understanding and promoting effective engagement With digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
58. Bucci S, Barrowclough C, Ainsworth J, et al. Actissist: proof-of-concept trial of a theory-driven digital intervention for psychosis. Schizophr Bull 2018 Aug 20;44(5):1070-1080. [doi: 10.1093/schbul/sby032] [Medline: 29566206]
59. Ahn J, Park SJ. The effects of serious game for elderly people—focused on the case of development of 'Rejuvenesce Village'. J Korean Inst Inf Technol 2016 Mar 31;14(3):195. [doi: 10.14801/jkiit.2016.14.3.195]
60. Campbell LM, Paolillo EW, Heaton A, et al. Daily activities related to mobile cognitive performance in middle-aged and older adults: an ecological momentary cognitive assessment study. JMIR Mhealth Uhealth 2020 Sep 24;8(9):e19579. [doi: 10.2196/19579] [Medline: 32969829]
61. Chapman SB, Fratantoni JM, Robertson IH, et al. A novel BrainHealth index prototype improved by telehealth-delivered training during COVID-19. Front Public Health 2021;9:641754. [doi: 10.3389/fpubh.2021.641754] [Medline: 33796498]
62. Nakada T, Kurotani K, Kozawa T, et al. A real-time web-based intervention with a multicomponent group-based program for older adults: single-arm feasibility study. Healthcare (Basel) 2024 Nov 26;12(23):2365. [doi: 10.3390/healthcare12232365] [Medline: 39684987]
63. Laane SA, Cook LG, Spence JS, Harris MN, Chapman SB. Effects of online brain training on self-reported mental health symptoms for generally healthy adults during the Covid-19 pandemic. Brain Behav 2023 Jan;13(1):e2853. [doi: 10.1002/brb3.2853] [Medline: 36542528]
64. Kim DR, Moon E, Shin MJ, Yang YA, Park JH. Effect of individual virtual reality cognitive training programs on cognitive function and depression in middle-aged women: randomized controlled trial. JMIR Ment Health 2023 Oct 25;10:e48912. [doi: 10.2196/48912] [Medline: 37878378]

XSL•FO
RenderX

65. Kang JM, Kim N, Lee SY, et al. Effect of cognitive training in fully immersive virtual reality on visuospatial function and frontal-occipital functional connectivity in predementia: randomized controlled trial. J Med Internet Res 2021 May 6;23(5):e24526. [doi: 10.2196/24526] [Medline: 33955835]
66. Li R, Geng J, Yang R, Ge Y, Hesketh T. Effectiveness of computerized cognitive training in delaying cognitive function decline in people with mild cognitive impairment: systematic review and meta-analysis. J Med Internet Res 2022 Oct 27;24(10):e38624. [doi: 10.2196/38624] [Medline: 36301590]
67. Ryu H, Kim S, Kim D, Han S, Lee K, Kang Y. Simple and steady interactions win the healthy mentality: designing a chatbot service for the elderly. Proc ACM Hum Comput Interact 2020;4(CSCW2):1-25. [doi: 10.1145/3415223]
68. Dhimolea TK, Kaplan-Rakowski R, Lin L. Supporting social and emotional well-being with artificial intelligence. In: Bridging Human Intelligence and Artificial Intelligence: Springer International Publishing; 2022:125-138. [doi: 10.1007/978-3-030-84729-6_8]
69. Yoo SS, Shin D, Kim GM, et al. The association between functional social support and cognitive change in older adults: a 3-year longitudinal study. Arch Gerontol Geriatr 2023;107:104884. [doi: 10.1016/j.archger.2023.105076] [Medline: 37245489]
70. Kang JE, Martire LM, Graham-Engeland JE, Almeida DE, Sliwinski MJ. Chronic loneliness and longitudinal changes in cognitive functioning. BMC Public Health 2025 Mar 29;25(1):1190. [doi: 10.1186/s12889-025-22313-2] [Medline: 40155901]
71. Guarnera J, Yuen E, Macpherson H. The impact of loneliness and social isolation on cognitive aging: a narrative review. J Alzheimers Dis Rep 2023;7(1):699-714. [doi: 10.3233/ADR-230011] [Medline: 37483321]
72. Windsor TD, Ghisletta P, Gerstorf D. Social resources as compensatory cognitive reserve? interactions of social resources with education in predicting late-life cognition. J Gerontol B Psychol Sci Soc Sci 2020 Aug 13;75(7):1451-1461. [doi: 10.1093/geronb/gby143] [Medline: 30590858]
73. Woods B, Rai HK, Elliott E, Aguirre E, Orrell M, Spector A. Cognitive stimulation to improve cognitive functioning in people with dementia. Cochrane Database Syst Rev 2023 Jan 31;1(1):CD005562. [doi: 10.1002/14651858.CD005562.pub3] [Medline: 39804128]
74. Wang R, Liu M, Cheng X, Wu Y, Hildebrandt A, Zhou C. Segregation, integration, and balance of large-scale resting brain networks configure different cognitive abilities. Proc Natl Acad Sci USA 2021 Jun 8;118(23). [doi: 10.1073/pnas.2022288118]
75. Ahn S, Kim S, Zhang H. Changes in depressive symptoms among older adults with multiple chronic conditions: role of positive and negative social support. Int J Environ Res Public Health 2017;14(1):16. [doi: 10.3390/ijerph14010016]
76. Oddone CG, Hybels CF, McQuoid DR, Steffens DC. Social support modifies the relationship between personality and depressive symptoms in older adults. Am J Geriatr Psychiatry 2011 Feb;19(2):123-131. [doi: 10.1097/jgp.0b013e3181f7d89a] [Medline: 21328795]
77. Wang J, Xue J, Jiang Y, Zhu T, Chen S. Mediating effects of depressive symptoms on social support and quality of life among rural older Chinese. Health Qual Life Outcomes 2020 Jul 20;18(1):242. [doi: 10.1186/s12955-020-01490-1] [Medline: 32690060]
78. Kong LN, Zhu WF, He S, Yao Y, Yang L. Relationships among social support, coping strategy, and depressive symptoms in older adults with diabetes. J Gerontol Nurs 2019 Apr;45(4):40-46. [doi: 10.3928/00989134-20190305-03]
79. Zhu D, Al Mahmud A, Liu W. Digital storytelling intervention for enhancing the social participation of people with mild cognitive impairment: co-design and usability study. JMIR Aging 2024 Jan 17;7:e54138. [doi: 10.2196/54138] [Medline: 38231541]
80. Chau PH, Cheung DST, Kwok JYY, Chan WC, Yu DSF. Online group-based dual-task training to improve cognitive function of community-dwelling older adults: randomized controlled feasibility study. JMIR Aging 2025 May 16;8:e67267. [doi: 10.2196/67267] [Medline: 40378407]
81. Fisher M, Etter K, Murray A, et al. The effects of remote cognitive training combined with a mobile app intervention on psychosis: double-blind randomized controlled trial. J Med Internet Res 2023 Nov 13;25:e48634. [doi: 10.2196/48634] [Medline: 37955951]
82. Khaleghi A, Aghaei Z, Mahdavi MA. A gamification framework for cognitive assessment and cognitive training: qualitative study. JMIR Serious Games 2021 May 18;9(2):e21900. [doi: 10.2196/21900] [Medline: 33819164]
83. Vermeir JF, White MJ, Johnson D, Crombez G, Van Ryckeghem DML. The effects of gamification on computerized cognitive training: systematic review and meta-analysis. JMIR Serious Games 2020 Aug 10;8(3):e18644. [doi: 10.2196/18644] [Medline: 32773374]
84. Morrow EL, Nelson LA, Duff MC, Mayberry LS. An ecological momentary assessment and intervention tool for memory in chronic traumatic brain injury: development and usability of memory ecological momentary intervention. JMIR Rehabil Assist Technol 2024 Nov 26;11:e59630. [doi: 10.2196/59630] [Medline: 39591513]
85. Singh A, Chakraborty S, He Z, et al. Predicting adherence to computer-based cognitive training programs among older adults: study of domain adaptation and deep learning. JMIR Aging 2024 Sep 16;7:e53793. [doi: 10.2196/53793] [Medline: 39283346]

XSL•FO
RenderX

86. Bälter O, Jemstedt A, Javan Abraham F, Persson Osowski C, Mugisha R, Bälter K. Effect of personalized email-based reminders on participants' timeliness in an online education program: randomized controlled trial. JMIR Form Res 2023 Oct 13;7:e43977. [doi: 10.2196/43977] [Medline: 37831487]
87. Lu SY, Yoon S, Yee WQ, Heng Wen Ngiam N, Ng KYY, Low LL. Experiences of a community-based digital intervention among older people living in a low-income neighborhood: qualitative study. JMIR Aging 2024 Apr 25;7:e52292. [doi: 10.2196/52292] [Medline: 38662423]
88. Benham S, Otchet K, Senft D, Potter AM. A feasibility study of two cognitive training programs for urban community-dwelling older adults. JAL 2022;2(2):74-84. [doi: 10.3390/jal2020007]
89. Vance DE, Fazeli PL, Azuero A, et al. A 2-Year longitudinal randomized control trial of speed of processing cognitive training in aging adults with HIV-associated neurocognitive disorder: results of the think fast study. AIDS Behav 2024 Oct;28(10):3300-3314. [doi: 10.1007/s10461-024-04409-9] [Medline: 38954173]
90. Jeon S, Charles ST. Internet-based social activities and cognitive functioning 2 years later among middle-aged and older adults: prospective cohort study. JMIR Aging 2024 Dec 10;7:e63907. [doi: 10.2196/63907] [Medline: 39656544]
91. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence-based chatbots for promoting health behavioral changes: systematic review. J Med Internet Res 2023 Feb 24;25:e40789. [doi: 10.2196/40789] [Medline: 36826990]
92. Chen J, Hu RZ, Zhuang YX, et al. Natural language processing chatbot-based interventions for improvement of diet, physical activity, and tobacco smoking behaviors: systematic review. JMIR Mhealth Uhealth 2025 Jun 11;13:e66403. [doi: 10.2196/66403] [Medline: 40503914]
93. Forman EM, et al. Gender differences in the effect of gamification on weight loss outcomes. Transl Behav Med 2021;11(4):1015-1022. [doi: 10.1093/tbm/ibaa091] [Medline: 32945860]
94. Piquer-Martinez C, Valverde-Merino MI, Gomez-Guzman M, Zarzuelo MJ. Gender-based differences in gamification and mobile learning. Acta Physiol (Oxf) 2024 Sep;240(9):e14206. [doi: 10.1111/apha.14206] [Medline: 38990085]
95. Almusharraf N, Aljasser M, Dalbani H, Alsheikh D. Gender differences in utilizing a game-based approach within the EFL online classrooms. Heliyon 2023 Feb;9(2):e13136. [doi: 10.1016/j.heliyon.2023.e13136] [Medline: 36793946]

Abbreviations

AI: artificial intelligence
HC: healthy control
IRB: institutional review board
K-MMSE-2: Korean-Mini Mental State Examination—Version 2
MOS-SSS: Medical Outcomes Study–Social Support Survey
SGDS-K: Short Form Geriatric Depression Scale—Korean Version
TWEETS: Twente Engagement with eHealth Technologies Scale
UPSA-3: University of California, San Diego Performance-Based Skills Assessment

Edited by L Buis; submitted 01.Apr.2025; peer-reviewed by M Schure, M Adnan; revised version received 26.Oct.2025; accepted 04.Feb.2026; published 24.Jun.2026.

Please cite as:
Yun B, Kim B, Ko H, Kim J, Kim BR, Kim W, Park S, Kim J, Lee JH, Kim GH
Enhancing Cognitive Functions of Older Adults With Software Robot: Longitudinal Exploratory Field Study
JMIR Mhealth Uhealth 2026;14:e75308
URL: https://mhealth.jmir.org/2026/1/e75308
doi:10.2196/75308

© Byunghun Yun, Bohee Kim, Hyunjeong Ko, Jinsung Kim, Bori R Kim, Whani Kim, Soyoon Park, Jinwoo Kim, Jee Hang Lee, Geon Ha Kim. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 24.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Long-Term Effectiveness of Unguided Internet-Based Cognitive Behavioral Therapy on Major Depressive Disorder in Chinese Adults: Randomized Controlled Trial With a 12-Month Follow-Up

Wenjing Zhou[1,2*], PhD; Huimin Zhang[3*], PhD; Yunbin Jiang[1,2], MPH; Yanzhi Li[1,2], PhD; Guangduoji Shi[1,2], PhD; Hao Zhao[1,2], PhD; Wanxin Wang[1,2], PhD; Yuhua Liao[3], PhD; Yifeng Liu[3], MS; Jiejing Hao[1,2,3], PhD; Roger S McIntyre[4], MD; Beifang Fan[3], MM; Ciyong Lu[1], PhD

[1]Department of Medical Statistics and Epidemiology, School of Public Health, Sun Yat-sen University, 74 Zhongshan 2nd, Guangzhou, China

[2]Guangdong Provincial Key Laboratory of Food, Nutrition and Health, School of Public Health, Sun Yat-sen University, Guangzhou, China

[3]Department of Psychiatry, Shenzhen Nanshan Center for Chronic Disease Control, Shenzhen, China

[4]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Ciyong Lu, PhD
Department of Medical Statistics and Epidemiology, School of Public Health, Sun Yat-sen University, 74 Zhongshan 2nd, Guangzhou, China

Abstract

Background: Unguided internet-based cognitive behavioral therapy (ICBT) is a low-cost and scalable treatment for major depressive disorder (MDD), but its long-term effects in Chinese populations remain unclear.

Objective: This study aimed (1) to explore the short- and long-term effectiveness of unguided ICBT in treating adults with MDD; (2) to investigate the short- and long-term effects on disease-related symptoms, individual and social functioning, and quality of life; and (3) to assess the acceptability and satisfaction with the ICBT.

Methods: An 8-week randomized controlled trial (ChiCTR2100046425) was conducted between August 2021 and June 2023 in Shenzhen, China, with 159 participants in the immediate ICBT group (7-module ICBT course plus usual care) and 158 in the waitlist control (WLC) group (usual care). The WLC group later completed the same ICBT course and follow-up assessments. Outcome measures (depressive and anxiety symptoms, psychological distress, social functioning, self-efficacy, quality of life, and stigma) were assessed before and after treatment and at 3-, 6-, and 12-month follow-ups for ICBT participants. Remission and response, adherence, and satisfaction were evaluated by predefined standards.

Results: Among 300 participants analyzed (mean age 28.49, SD 7.0 years; female: n=225, 75%), dropout rates were 22.4% (34/152) in the immediate ICBT group versus 6.3% (10/158) in the WLC group. At posttreatment, the immediate ICBT group showed greater reduction in depressive symptoms versus WLC (mean difference −3.65, SE 0.60; P<.001; d=0.50), with higher remission (80/121, 66.1% vs 58/148, 39.2%; P<.001) and response rates (50/121, 41.3% vs 27/148, 18.2%; P<.001). At 12-month follow-up, the depressive symptoms were improved compared with that at pretreatment (mean difference −3.90, SE 0.32; P<.001; d=0.70), and no significant change was observed in comparison with the outcomes at posttreatment (mean difference −0.81, SE 0.33; P=.33; d=−0.15). ICBT treatment also exhibited similar short- and long-term effects on secondary outcomes, with significant improvement of disease-related symptoms, individual and social functioning, and quality of life. Moreover, the majority of the participants treated with ICBT reported high acceptability of and satisfaction with the ICBT course.

Conclusions: Unguided ICBT effectively reduces depressive symptoms and enhances functioning in Chinese patients with MDD, with sustained benefits over 12 months. Its scalability and low-cost nature make it a promising option for resource-limited settings.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2100046425; https://tinyurl.com/bdcrj4zv

(JMIR Mhealth Uhealth 2026;14:e68394) doi:10.2196/68394

KEYWORDS

internet-based cognitive behavioral therapy; major depressive disorder; depression; China; randomized controlled trial; long-term effectiveness

XSL•FO
RenderX

Introduction

Background

Major depressive disorder (MDD) affects approximately 332 million people globally in 2021 [1]. As the second leading cause of nonfatal disability [1], MDD is associated with severe emotional symptoms, low self-efficacy, impaired social functioning, and diminished quality of life [2]. Additionally, MDD has imposed a significant economic burden on individuals and societies, which is predominantly from productivity loss [3,4]. Although there are efficacious treatments for MDD with robust clinical evidence (ie, psychotherapies and pharmacotherapies) [5], the treatment gap is still substantial. In low- and middle-income countries, more than 70% of patients remain untreated [6,7]. Social discrimination and personal stigma, high treatment expenditure, side effects of antidepressants, and the inconvenience of traditional in-person psychotherapies have prevented people from accessing these conventional forms of treatment [8].

Internet-based cognitive behavioral therapy (ICBT), a scalable, confidential, low-cost, and flexible modality of cognitive behavioral therapy (CBT) delivered through the internet with guided and unguided forms, has been introduced to the treatment of MDD to overcome the treatment barriers mentioned earlier [9]. Although CBT has been recommended as a first-line treatment for MDD by many treatment guidelines [10-12], the levels of evidence and applicable scenarios of ICBT vary across these guidelines. In China, mental health resources are severely limited, with per capita mental health investment standing at US $1.07 and only 2.19 certified psychiatrists per 100,000 people, far below the high-income country averages of US $35.06 and 13.06 psychiatrists [13]. The structural imbalance in the distribution further compounds the shortage, with hospitals and professionals disproportionately clustered in provincial capitals and the more developed eastern regions [14]. Moreover, despite the high prevalence of MDD in China (6.8%), the service coverage rate remains strikingly low at only 9.5% [15]. The unguided form, defined as self-help ICBT with different levels of technical support, offers greater potential for scalability compared to guided ICBT. Consequently, the earlier-mentioned mental health realities in China provided the impetus for our team to develop an unguided ICBT course for MDD called Morning Mood on WeChat Mini Program.

There is a growing body of evidence that unguided ICBT showed superior outcomes in alleviating depressive symptoms compared to control treatments (ie, treatment-as-usual and waitlist control [WLC]) in short-term clinical end points (ie, posttreatment) [16-31]. However, a large variance is observed in reported short-term effect sizes, ranging from no effect [23] to large effect sizes [22,28,29]. Moreover, the long-term efficacy of unguided ICBT remains understudied, with existing studies yielding inconsistent conclusions regarding whether sustained improvements occur beyond the intervention period [23,26,32-35]. Newby et al [26] found that the depressive outcomes at the posttreatment stage were favored over the 3-month follow-up stage, with a within-group effect size of −0.21, while Gilbody et al [23] reported that the within-group effect size between the posttreatment and 12-month follow-up stage was 0.35. Regarding binary outcomes, both response rates (defined as ≥50% reduction in symptom severity) and remission rates (defined as symptom levels below clinical thresholds) exhibited considerable variability across studies. Posttreatment analyses revealed that unguided ICBT achieved response rates of 37% - 58.6% and remission rates of 30.5% - 64%, representing more improvements over control conditions [18,20,25,27,32,34,36,37]. Only 1 study reported a response rate at 12-month follow-up of 51.4% [32]. In addition, existing studies focusing on unguided ICBT for MDD are largely represented by Western studies, thereby limiting the ability to generalize the results to patients from different races, ethnicities, and cultural backgrounds [38]. In China, only 1 randomized controlled trial (RCT) has verified the short-term effectiveness of clinician-supported unguided ICBT (the Chinese-translated version of MoodGYM) in patients with MDD [31]. However, previous research has explored the effect of ICBT for depressive symptoms or subthreshold depression. During the COVID-19 pandemic, unguided ICBT has been proven effective for patients with depressive symptoms and COVID-19 [39]. Another study demonstrated that unguided ICBT significantly reduced depressive severity among undergraduate students with mild to severe symptoms of depression [40]. Beyond unguided form, guided ICBT has also shown a moderate antidepressant effect among individuals with subthreshold depression, regardless of age [41,42]. Furthermore, a trial in Hong Kong confirmed that both web- and app-based guided ICBT led to significant improvements in depressive symptoms [43]. Furthermore, the aforementioned studies have primarily focused on depressive symptomatology outcomes, limiting possible insights into other relevant depression domains (eg, anxiety symptoms, self-efficacy, quality of life, and functional outcomes) [44]. Given the earlier-mentioned limitations, there is a need to evaluate the short- and long-term effectiveness of unguided ICBT for MDD, especially in China.

Objectives

Herein, this study aimed (1) to evaluate the short- and long-term effectiveness of unguided ICBT in treating adults with MDD, (2) to further evaluate the short- and long-term effects on specific depressive symptoms as well as individual and social functioning, as well as quality of life, and (3) to assess the acceptability and satisfaction with the ICBT.

Methods

Study Design

This study was an 8-week nonblinded pragmatic RCT with 12-month follow-up, investigating whether an unguided ICBT course as an adjunct to usual care could reduce depression symptoms in adult patients with MDD and its long-term effectiveness. Eligible participants were randomized to either (1) the immediate ICBT group (ICBT group I), receiving the ICBT intervention with nonspecialists' support plus usual care, or (2) the WLC group, receiving usual care only. After the 8-week control period, WLC participants crossed over to receive the ICBT intervention (ICBT group II). Both ICBT groups (I and II) were assessed at baseline, posttreatment, and at 3, 6, and

XSL•FO
RenderX

12 months after the intervention to examine immediate and long-term effects.

Ethical Considerations

This trial was registered with the Chinese Clinical Trial Registry (ChiCTR2100046425). The study protocol, patient information sheets, and informed consent forms were approved by the Medical Ethics Committee of Shenzhen Nanshan Center for Chronic Disease Control, China (ethics approval ll20210012). The CONSORT (Consolidated Standards of Reporting Trials) with the eHealth extension was adopted to guide and report the trial (Checklist 1) [45]. All participants provided informed consent. Deidentified data were used for the whole analysis. Financial incentives (US $6.82 electronic cash) were distributed to those who completed the trial.

Participants and Procedure

All participants were recruited from the Department of Depressive Disorder at Shenzhen Kangning Hospital and the Department of Psychiatry at Shenzhen Nanshan Center for Chronic Disease Control. MDD diagnosis was determined by using the Mini-International Neuropsychiatric Interview (M.I.N.I.), which was administered according to the *Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition* (DSM-4) criteria [46,47]. The enrollment period spanned from August 2021 to June 2023. The key inclusion criteria were (1) Chinese residents, (2) aged between 18 and 60 years, (3) met criteria for MDD in DSM-4, (4) had access to the internet, (5) a score of Patient Health Questionnaire-9 (PHQ-9) more than 5, (6) no severe suicidal risk according to section C of the M.I.N.I., and (7) in a stable condition (if taking antidepressants, dosage had to be stable for at least 30 days before enrollment). The key exclusion criteria were (1) met DSM-4 criteria for bipolar disorder or other psychotic disorders, (2) had an ongoing alcohol or substance abuse disorder, (3) was pregnant or lactating, (4) underwent physiotherapy or psychotherapy treatments at the time of enrollment, and (5) any other condition deemed by the psychiatrists to interfere with ICBT (eg, cognitive impairment).

A 2-stage procedure was adopted for the selection of eligible participants. Outpatients who visited the department of depressive disorder and psychiatry were first screened by psychiatrists to identify potential study participants. Trained researchers (YJ and H Zhao) then conducted structured diagnostic interviews (ie, M.I.N.I.) to make the MDD diagnosis. Both researchers hold master's degrees and have medical backgrounds. They underwent standardized M.I.N.I. interview training organized by the National Center for Mental Disorders, which comprised theoretical instruction, simulated interviews, and a reliability test. Furthermore, both researchers completed a 1-year clinical practicum at the Department of Psychiatry, Nanshan Center for Chronic Disease Control (ie, Nanshan Mental Health Center). Throughout the data collection period, they received regular supervision from senior psychiatrists to ensure diagnostic consistency. Randomization was performed using a computer-generated sequence with randomly permuted block sizes (4 and 6). This sequence was managed centrally and applied sequentially to participants in biweekly batches throughout the recruitment period to ensure ongoing group balance. Due to the nature of the interventions, the blind method was hard to carry out.

Eligible participants were then instructed to complete baseline assessments through online questionnaires. Sociodemographic information (eg, age, sex, and ethnicity), lifestyle features (eg, exercise, drinking, and smoking), clinical characteristics (eg, antidepressant use, first of onset, and age of MDD onset), and baseline symptoms (eg, somatic pain, resilience, and insomnia severity) were collected.

Interventions

The ICBT intervention in this trial was tailored for the Chinese population, aiming at teaching practical CBT skills for depression. A total of four phases were carried out for intervention development and cultural adaptation: (1) literature review on the ICBT intervention and cultural adaptation framework, (2) preliminary content setting, (3) face-to-face group discussion and content adaptation, and (4) expert consultations and finalized adaptation.

Guided by the ecological validity model, cultural adaptations were made across 8 dimensions (ie, language, persons, metaphors, content, concepts, goals, methods, and contexts) to ensure relevance, acceptability, and comprehensibility [48]. In terms of language, the intervention was delivered in Mandarin, using idioms and communication styles that align with local habits to prevent misinterpretation. For persons, all illustrative characters designed with names, appearances, and attire were in line with Chinese people's characteristics, and video cases featured individuals from diverse age groups to enhance relatability. Metaphors were replaced with localized, everyday analogies to improve understanding. The content, including scenarios and activities, was tailored to align with common daily practices, social customs, and economic realities. Concepts of stressors and depressive symptoms were refined to resonate with the population's lived experiences. To mitigate mental health stigma, the intervention's goals were framed around "mood management" rather than depression treatment. The core therapeutic methods, such as behavioral activation, were preserved, but the suggested activities were customized according to culturally relevant options. Finally, the context of delivery was adapted by designing the intervention as a WeChat Mini Program, leveraging a ubiquitous platform in China to maximize accessibility and minimize barriers to engagement.

The ICBT course consists of 7 lessons, covering psychoeducation, behavioral activation, cognitive restructuring, affect regulation, and relapse prevention. The detailed content of each lesson and the corresponding screenshots are displayed in Multimedia Appendix 1. Participants in the ICBT group were required to complete 1 lesson per week; one week after completing the prior lesson, the subsequent lesson would be open automatically. This setting was used to prevent participants from completing multiple lessons in a short period, which might potentially impact the effectiveness. General support involving technical support and daily reminders provided by nonspecialists (eg, nurses, lay health workers, and social workers) via telephone or WeChat was applied to prompt engagement. The active treatment phase comprised 7 weeks, while the total

XSL•FO
RenderX

intervention period, which included a final week for outcome consolidation, was defined as 8 weeks.

Participants in the WLC group were treated with usual care, and no additional care was offered during the 8-week study period. After completing the posttreatment assessment, participants in the WLC group were asked whether they were willing to receive ICBT intervention at no cost. The login accounts and passwords for Morning Mood were then provided to the intended individuals. All participants with antidepressant prescriptions were instructed to be on a stable dose for the past 4 weeks before enrollment and maintain the antidepressant regimen (unchanged drug type and dosage) during the intervention phase. Only necessary adjustments due to severe adverse reactions are permitted.

Outcome Measures

The primary outcome was depressive symptoms measured by the Chinese version of PHQ-9, which was with good reliability and validity [49]. Its Cronbach α was 0.91 in this study. Remission was defined as a PHQ-9 score less than 10, where the threshold was set and adopted by the Improving Access to Psychological Therapies of the United Kingdom to distinguish clinical cases and noncases [50,51]. Response was defined as symptom improvement (ie, the reduction of PHQ-9 score of 50% or more) [34].

The secondary outcome measures included anxiety symptoms, nonspecific psychological distress, social function impairment, general self-efficacy, quality of life, and depression stigma, which were assessed by the Chinese version of the Generalized Anxiety Disorder-7 (GAD-7), the Kessler 10-Item Psychological Distress Scale (K-10), the Sheehan Disability Scale (SDS), the General Self-Efficacy Scale (GSES), the Short Form Six-Dimension (SF-6D), and Depression Stigma Scale (DSS), respectively. These measurement tools have good reliability and validity [52-57]. In this study, the corresponding Cronbach α was 0.94, 0.96, 0.94, 0.71, 0.95, and 0.84, respectively. All outcomes were evaluated at pre- and posttreatment, with further follow-up at 3, 6, and 12 months after ICBT intervention.

Participants' satisfaction and acceptability for ICBT treatment were investigated after the treatment by the following three questions: (1) "Do you think the courses in this project have been helpful for you?" (2) "Are you satisfied with the courses in this project?" and (3) "Would you be willing to recommend this project to people in need?"

Statistical Analysis

The sample size was determined by a priori power analysis. A total of 86 participants per group (n=172) was thought to be sufficient to detect a moderate between-group effect size (Cohen *d*=0.50, defined as the standardized difference between groups obtained by calculating the mean difference and dividing by the pooled SD) on repeated measures for PHQ-9, with a power of 90% and 2-tailed α of .05. The expected sample size was calculated based on a previous meta-analysis of individual patient data that investigated the effectiveness of unguided ICBT versus inactive controls for depressive symptoms [34]. Taking the attrition into consideration, the target sample size should be 228 patients when the anticipated dropout rate is set at 25% [39].

Participants who did not start lesson 1 in the immediate ICBT group and who could not be contacted in the WLC group after randomization were excluded from the main analysis, which followed the modified intention-to-treat (mITT) principle [58-60]. Sensitivity analyses were conducted in the intention-to-treat (ITT) sample (ie, all randomized participants) and the per-protocol set (PPS; ie, participants who completed all modules in the immediate ICBT group and who completed the posttreatment assessment in the WLC group). According to the statistical principles for clinical trials, the primary analysis in a clinical trial should adhere to the ITT principle by including all randomized participants. In practice, the mITT approach is often used as a refinement for ITT. This approach justifiably excludes a limited number of participants, such as those who violated major eligibility criteria, never received any dose of the study treatment, or provided no data after randomization [61]. Excluding patients who received no treatment is methodologically sound, as such analyses can serve as an unbiased estimator for the principal stratum effect (ie, the effect in the population of patients who would initiate treatment) [62]. Furthermore, the mITT analysis complements the ITT findings by providing a more precise estimate of the treatment effect among patients who actually started the intervention [58]. Missing data of posttreatment assessment in the ITT sample were addressed via multiple imputation using chained equations (*MICE* package in R [53]), with pooled estimates derived from Rubin's rules [63,64]. Baseline characteristics and outcome measures were described as mean (SD) for continuous variables and frequencies (percentages) for categorical variables. The differences between the 2 groups were compared using 2-tailed *t* tests, chi-square tests, and Mann-Whitney tests.

The 8-week model (ie, random intercept linear mixed models [LMMs]) was applied to evaluate the effectiveness of the ICBT by estimating the mean between-group differences (the ICBT group I: n=152 and the WLC group: n=148) in 9 outcomes at the pre- and posttreatment time points (8 weeks). In this model, treatment groups, the time point of measurement, and the treatment by time point interaction were treated as fixed effects, and the individual patient effect was set as a random effect with a normal distribution and expected value 0. Parameters were estimated using maximum likelihood. In addition, the follow-up model (ie, marginal LMM with random intercept) was used to estimate the outcome changes in the ICBT group (ICBT group I: n=152 and ICBT group II: n=108). The fixed effect of time was used to estimate the marginal means of each outcome at 8 weeks, 3, 6, and 12-month follow-ups. Post-hoc pairwise comparisons of these marginal means across time points were conducted with Bonferroni correction. There were 18 models, one up to 8 weeks across arms ("8-week model") and one including combined intervention group data ("follow-up model") per outcome variable (depressive symptoms, anxiety symptoms, psychological distress, social function impairment, health-related quality of life, general self-efficacy, depression stigma, personal depression stigma, and perceived depression stigma). Model fit was evaluated using Schwarz's Bayesian criterion. Within-group and between-group effect sizes (Cohen *d*) were calculated based

XSL•FO
RenderX

on the method suggested for LMM analysis [65,66]. It was reported that the effect be categorized as small (>0.20), moderate (>0.50), and large (>0.80) according to Cohen *d* [67]. Sensitivity analyses for long-term effectiveness evaluation were conducted, respectively, among participants treated with ICBT in the ICBT group and the WLC group.

The remission and response rates in the ICBT and WLC groups were also estimated by treatment arms (ICBT vs WLC) and time points (posttreatment, 3-, 6-, and 12-month follow-ups). Chi-square tests were performed to assess the frequency differences in remission and response between the 2 groups. All analyses were performed using R language (version 4.2.2; R Foundation for Statistical Computing), and a *P* value less than .05 was considered statistically significant.

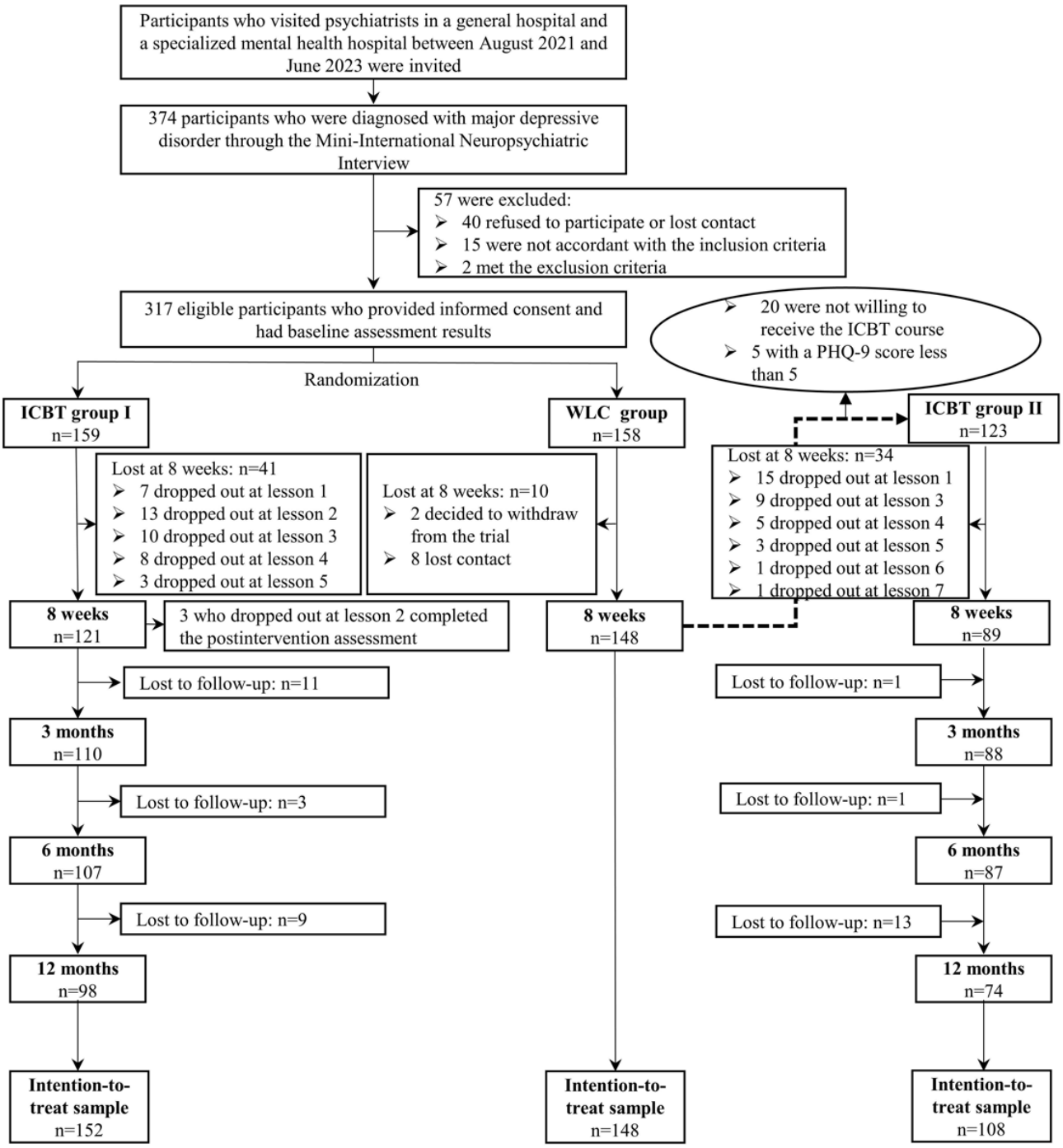
Results

Overview

A total of 374 participants with MDD were recruited for this trial between August 2021 and June 2023, of whom 317 cases entered the trial and were randomized (ICBT group I: n=159; WLC group: n=158). In the ICBT group I, 7 participants did not start lesson 1 and thus were excluded. According to the criteria for completion and attrition defined in Multimedia Appendix 2, the treatment completion rate was 77.6% (118/152), and the attrition rate at 12-month follow-up was 35.5% (54/152). The average number of ICBT lessons completed was 5.89 (SD 2.13). In the WLC group, the dropout rate was 6.3% (10/158). After the 8-week waitlist phase, participants in the WLC group were provided access to the ICBT intervention (ICBT group II). In total, there were 108 (73%) who initiated the intervention. The remaining participants were excluded for the following reasons: 20 (13.5%) declined to receive treatment, 5 (3.4%) no longer met the inclusion criteria (PHQ-9 score<5), and 15 (10.1%) only registered at Morning Mood and did not start lesson 1. In this group, the treatment completion rate was 82.4% (89/108), and the attrition rate at 12-month follow-up was 31.5% (34/108). The detailed completion and dropout information are presented in Figure 1. The mITT sample used for the short-term effectiveness analyses consisted of 152 patients in the ICBT group I and 148 patients in the WLC group. For the long-term effectiveness evaluation, a total of 260 participants receiving ICBT intervention in the 2 ICBT groups were included, with 152 from group I and 108 from group II.

XSL•FO
RenderX

Figure 1. Study flowchart. ICBT: internet-based cognitive behavioral therapy; PHQ-9: Patient Health Questionnaire-9; WLC: waitlist control.



Baseline Characteristics

Table 1 outlines the baseline characteristics of the mITT sample. The mean age was 28.49 (SD 7.00) years, and the female-to-male ratio was 3:1. Most participants were Han ethnic (283/298, 95%), highly educated (247/299, 82.6%), employed (192/297, 64.6%), and nonmarried (201/285, 70.5%). Participants whose monthly household income was below 10,000 Chinese Yuan (1 Chinese Yuan=US $0.1419 in 2023) accounted for the largest part (96/281, 34.2%). In modifiable lifestyle behaviors, the majority of the participants did not smoke cigarettes (186/299, 62.2%), lived with other people (210/283, 74.2%), drank alcohol (249/299, 83.3%), and had no exercise habits (218/299, 72.9%). In terms of clinical features, more than half of the participants were in their first episodes of MDD (175/292, 59.9%), did not take antidepressants (163/300, 54.3%), or did not have comorbidities at baseline (190/298, 63.8%). The mean (SD) of baseline outcome measures including PHQ-9, GAD-7, K-10, SDS, SF-6D, GSES, and DSS were 13.40 (4.83), 10.25 (4.60), 29.73 (7.48), 13.81 (6.81), 0.56 (0.21), 19.61 (5.71), and 52.08 (9.14), respectively. Baseline characteristics were comparable between the ICBT and WLC groups, suggesting that the randomization was successful. Table S1 in Multimedia Appendix 3 and Table S1 in Multimedia Appendix 4 display the baseline characteristics of the ITT sample and the PPS.

XSL•FO
RenderX

Table . Baseline characteristics of the participants included in the modified intention-to-treat sample.

Variable	Overall (N=300)	ICBT[a] (n=152)	WLC[b] (n=148)	*P* value[c]
Age (years), mean (SD)	28.49 (7.00)	29.05 (6.56)	27.92 (7.41)	.16
Sex, n (%)				.69
Male	75 (25)	40 (26.3)	35 (23.6)	
Female	225 (75)	112 (73.7)	113 (76.4)	
Nationality, n (%)				.98
Han	283 (95)	143 (95.3)	140 (94.6)	
Others	15 (5)	7 (4.7)	8 (5.4)	
Not available	2 (—[d])	2 (—)	0 (—)	
Educational levels, n (%)				.63
High school or below	52 (17.4)	28 (18.4)	24 (16.3)	
Undergraduate	206 (68.9)	101 (66.4)	105 (71.4)	
Master degree or above	41 (13.7)	23 (15.1)	18 (12.2)	
Not available	1 (—)	0 (—)	1 (—)	
Employment status, n (%)				.06
Employed	192 (64.6)	106 (70.2)	86 (58.9)	
Unemployed	105 (35.4)	45 (29.8)	60 (41.1)	
Not available	3 (—)	1 (—)	2 (—)	
Marital status, n (%)				.43
Married	84 (29.5)	46 (31.9)	38 (27)	
Unmarried or divorced or widowed	201 (70.5)	98 (68.1)	103 (73)	
Not available	15 (—)	8 (—)	7 (—)	
Monthly household income, n (%)				.10
No fixed income	35 (12.5)	12 (8.3)	23 (16.8)	
Below 10,000 CNY[e]	96 (34.2)	52 (36.1)	44 (32.1)	
10,000 - 20,000 CNY	77 (27.4)	45 (31.2)	32 (23.4)	
Above 20,000 CNY	73 (26)	35 (24.3)	38 (27.7)	
Not available	19 (—)	8 (—)	11 (—)	
Exercise, n (%)				.88
Yes	81 (27.1)	42 (27.8)	39 (26.4)	
No	218 (72.9)	109 (72.2)	109 (73.6)	
Not available	1 (—)	1 (—)	0 (—)	
Current drinking status, n (%)				.37
Yes	249 (83.3)	130 (85.5)	119 (81)	
No	50 (16.7)	22 (14.5)	28 (19)	
Not available	1 (—)	0 (—)	1 (—)	
Current smoking status, n (%)				.35
Yes	113 (37.8)	53 (34.9)	60 (40.8)	
No	186 (62.2)	99 (65.1)	87 (59.2)	
Not available	1 (—)	0 (—)	1 (—)	
Living alone, n (%)				.24
Yes	73 (25.8)	42 (29.2)	31 (22.3)	

XSL•FO
RenderX

Variable	Overall (N=300)	ICBT[a] (n=152)	WLC[b] (n=148)	*P* value[c]
No	210 (74.2)	102 (70.8)	108 (77.7)	
Not available	17 (—)	8 (—)	9 (—)	
Antidepressant use, n (%)				.36
Yes	137 (45.7)	65 (42.8)	72 (48.6)	
No	163 (54.3)	87 (57.2)	76 (51.4)	
First episode, n (%)				.92
Yes	175 (59.9)	89 (60.5)	86 (59.3)	
No	117 (40.1)	58 (39.5)	59 (40.7)	
Not available	8 (—)	5 (—)	3 (—)	
Age of onset (years), mean (SD)	24.51 (7.33)	25.05 (7.11)	23.94 (7.53)	.20
Comorbidity, n (%)				.53
Yes	108 (36.2)	52 (34.2)	56 (38.4)	
No	190 (63.8)	100 (65.8)	90 (61.6)	
Not available	2 (—)	0 (—)	2 (—)	
Number of SLE[f], mean (SD)	1.85 (1.87)	1.91 (1.90)	1.79 (1.85)	.58
CTQ[g], mean (SD)	49.00 (13.61)	49.99 (13.65)	47.98 (13.55)	.23
Emotional abuse scores, mean (SD)	10.60 (4.93)	10.56 (5.06)	10.64 (4.82)	.90
Physical abuse scores, mean (SD)	6.93 (3.15)	6.94 (3.24)	6.93 (3.06)	.97
Sexual abuse scores, mean (SD)	5.80 (1.94)	5.88 (2.12)	5.72 (1.75)	.48
Emotional neglect scores, mean (SD)	15.87 (5.22)	16.30 (5.04)	15.45 (5.38)	.18
Physical neglect scores, mean (SD)	9.94 (3.82)	10.30 (4.21)	9.58 (3.38)	.12
SSI[h] scores, mean (SD)	63.29 (22.01)	63.34 (20.39)	63.24 (23.62)	.97
CD-RISC[i] scores, mean (SD)	36.50 (14.87)	36.38 (14.60)	36.62 (15.19)	.90
RRS[j] scores, mean (SD)	53.70 (10.17)	53.58 (10.35)	53.83 (10.03)	.83
ISI[k] scores, mean (SD)	15.30 (6.61)	15.05 (6.30)	15.57 (6.93)	.50
PHQ-9[l] scores, mean (SD)	13.40 (4.83)	13.91 (5.12)	12.87 (4.47)	.06
GAD-7[m] scores, mean (SD)	10.25 (4.60)	10.63 (4.75)	9.86 (4.42)	.15
K-10[n] scores, mean (SD)	29.73 (7.48)	30.28 (7.36)	29.18 (7.59)	.20
SDS[o] scores, mean (SD)	13.81 (6.81)	15.19 (7.08)	12.40 (6.23)	<.001
GSES[p] scores, mean (SD)	19.61 (5.71)	19.53 (5.34)	19.70 (6.08)	.80
SF-6D[q] scores, mean (SD)	0.56 (0.21)	0.55 (0.21)	0.57 (0.20)	.26
DSS[r] scores, mean (SD)	52.08 (9.14)	52.70 (8.98)	51.46 (9.28)	.25
Personal DSS scores, mean (SD)	22.65 (5.25)	22.97 (5.21)	22.33 (5.28)	.29

XSL•FO
RenderX

Variable	Overall (N=300)	ICBT[a] (n=152)	WLC[b] (n=148)	*P* value[c]
Perceived DSS scores, mean (SD)	29.43 (6.72)	29.73 (6.47)	29.14 (6.97)	.45

[a]ICBT: internet-based cognitive behavioral therapy.

[b]WLC: waitlist control.

[c]Baseline characteristics were compared between the 2 groups using 2 independent-sample 2-tailed *t* tests for continuous variables and chi-square tests or Fisher exact probabilities for categorical variables.

[d]Not applicable.

[e]CNY: Chinese Yuan (1 CNY=US $0.1419 in 2023).

[f]SLE: stressful life event.

[g]CTQ: Childhood Trauma Questionnaire.

[h]SSI: Somatic Symptom Inventory.

[i]CD-RISC: Connor-Davidson Resilience Scale.

[j]RRS: Ruminative Responses Scale.

[k]ISI: Insomnia Severe Index.

[l]PHQ-9: Patient Health Questionnaire-9.

[m]GAD-7: General Anxiety Disorder-7.

[n]K-10: Kessler 10-Item Psychological Distress Scale.

[o]SDS: Sheehan Disability Scale.

[p]GSES: General Self-Efficacy Scale.

[q]SF-6D: Short Form Six-Dimension.

[r]DSS: Depression Stigma Scale.

Outcome Measures at Posttreatment

Table 2 depicts observed and estimated means of outcomes by treatment arms and time points, and the mean differences between groups calculated by 8-week models. The 8-week model suggested significant interaction effects of time-by-intervention-arm for PHQ-9 (β=−3.65; *P*<.001), GAD-7 (β=−2.85; *P*<.001), K-10 (β=−4.87; *P*<.001), SDS (β=−5.57; *P*<.001), GSES (β=1.42; *P*=.01), DSS (β=−2.93; *P*=.03), and personal DSS (β=−1.80; *P*=.008), except for SF-6D (β=.01; *P*=.50) and perceived DSS (β=−1.21; *P*=.21; Table S1 in Multimedia Appendix 5).

Figure 2 illustrates the change in estimated outcome measures across time by treatment arms. In both treatments, the scores of PHQ-9, GAD-7, and K-10 decreased over time, while the scores of SF-6D, GSES, and perceived DSS increased after treatment. Compared with the WLC group, larger improvements were obtained in these outcomes for those who received ICBT treatment. The changes in SDS, DSS, and personal DSS of the 2 groups moved in opposite directions, with a reduction in social function impairment, depression stigma, and personal stigma showing in the ICBT group.

XSL•FO
RenderX

Table . Outcomes for the modified intention-to-treat sample at pre- and posttreatment time points by treatment groups.

Outcome	Pretreatment						Posttreatment						Mean difference (SE)[c]	*P* value
	ICBT[a]			WLC[b]			ICBT			WLC				
	Values, n	Observed mean (SD)	Estimated mean (SE)[d]	Values, n	Observed mean (SD)	Estimated mean (SE)[d]	Values, n	Observed mean (SD)	Estimated mean (SE)[d]	Values, n	Observed mean (SD)	Estimated mean (SE)[d]		
PHQ-9[e] scores	152	13.91 (5.12)	13.91 (0.42)	148	12.87 (4.47)	12.87 (0.42)	121	8.42 (5.70)	8.58 (0.46)	148	11.18 (5.37)	11.18 (0.42)	−3.65 (0.63)	<.001
GAD-7[f] scores	152	10.63 (4.75)	10.63 (0.38)	148	9.86 (4.42)	9.86 (0.39)	121	6.36 (4.92)	6.33 (0.42)	148	8.42 (4.78)	8.42 (0.39)	−2.85 (0.56)	<.001
K-10[g] scores	152	30.28 (7.36)	30.28 (0.68)	148	29.18 (7.59)	29.18 (0.69)	121	22.93 (9.72)	23.03 (0.74)	148	26.80 (8.99)	26.80 (0.69)	−4.87 (0.96)	<.001
SDS[h] scores	152	15.19 (7.08)	15.19 (0.55)	148	12.40 (6.23)	12.40 (0.56)	119	11.03 (7.28)	11.18 (0.60)	148	13.94 (6.87)	13.94 (0.56)	−5.57 (0.72)	<.001
GSES[i] scores	152	19.53 (5.34)	19.66 (0.50)	148	19.70 (6.08)	20.03 (0.51)	116	21.69 (5.84)	21.90 (0.53)	148	20.52 (6.22)	20.85 (0.51)	1.42 (0.56)	.01
SF-6D[j] scores	152	0.55 (0.21)	0.55 (0.02)	148	0.57 (0.20)	0.57 (0.02)	116	0.61 (0.23)	0.61 (0.02)	148	0.62 (0.20)	0.62 (0.02)	0.01 (0.02)	.50
DSS[k] scores	148	52.70 (8.98)	52.72 (0.77)	148	51.46 (9.28)	51.46 (0.77)	120	52.57 (10.45)	52.40 (0.85)	148	54.07 (8.89)	54.07 (0.77)	−2.93 (1.33)	.03
Personal DSS scores	148	22.97 (5.21)	23.02 (0.45)	148	22.33 (5.28)	22.31 (0.45)	120	22.18 (5.67)	22.07 (0.48)	148	23.15 (5.61)	23.15 (0.45)	−1.80 (0.68)	.008
Perceived DSS scores	148	29.73 (6.47)	29.71 (0.57)	148	29.14 (6.97)	29.15 (0.57)	120	30.39 (7.32)	30.28 (0.63)	148	30.93 (7.18)	30.93 (0.57)	−1.21 (0.95)	.21

[a]ICBT: internet-based cognitive behavioral therapy.

[b]WLC: waitlist control.

[c]The differences posttreatment to pretreatment outcomes between the 2 treatment groups are based on the difference of least square means of linear mixed models for repeated measures.

[d]Estimated means are based on linear mixed models.

[e]PHQ-9: Patient Health Questionnaire-9.

[f]GAD-7: General Anxiety Disorder-7.

[g]K-10: Kessler 10-Item Psychological Distress Scale.

[h]SDS: Sheehan Disability Scale.

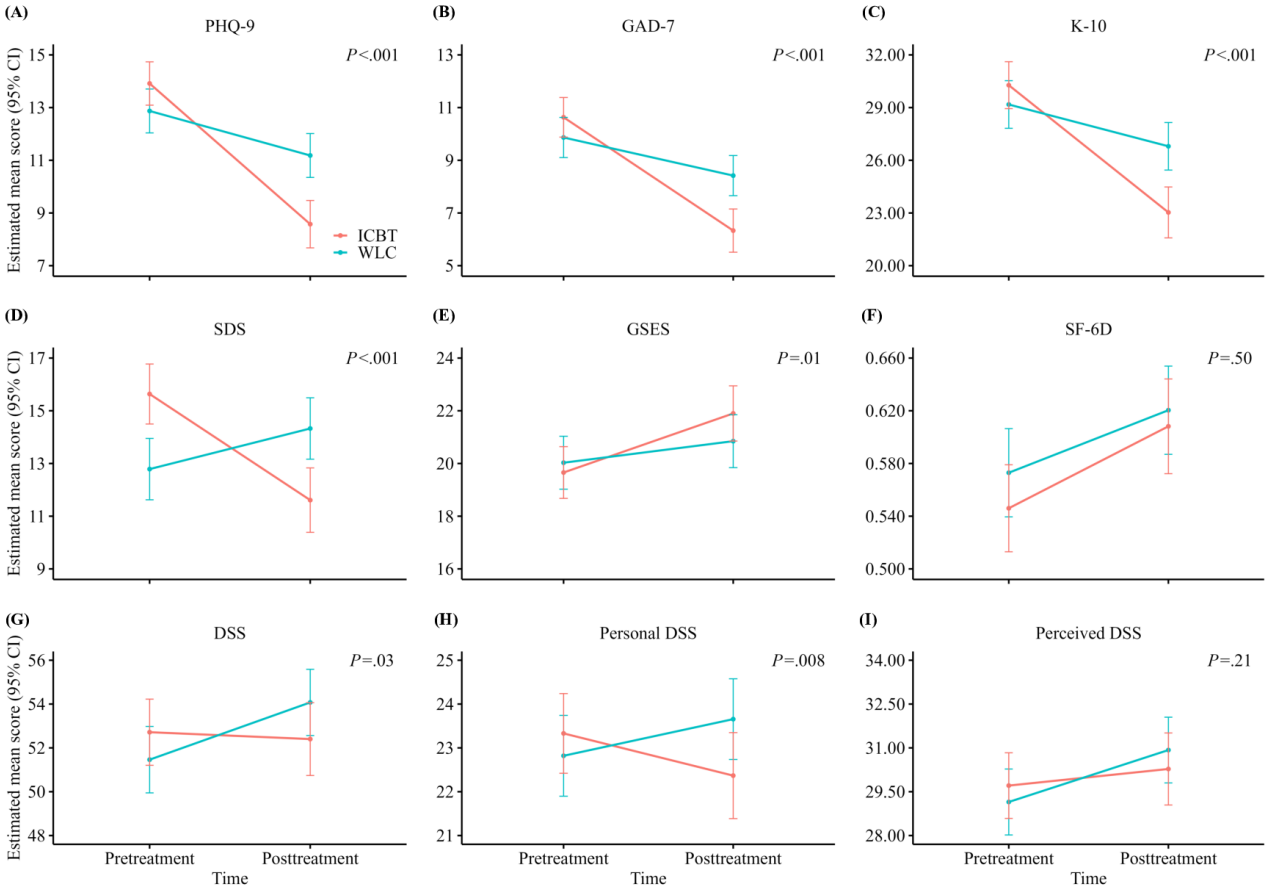
[i]GSES: General Self-Efficacy Scale.

[j]SF-6D: Short Form Six-Dimension.

[k]DSS: Depression Stigma Scale.

XSL•FO
RenderX

Figure 2. Estimated means and 95% CIs for time by treatment group interaction effects at 8 weeks. Linear mixed models with random intercept, including treatment groups (ie, internet-based cognitive behavioral therapy and waitlist control), time points of measurement (ie, pre- and posttreatment), and the treatment by time point interaction as fixed effects, are applied to calculate estimated means, mean differences in pre- and posttreatment changes between the 2 groups, and their *P* values. Patient-specific effects enter the model as a random effect with a normal distribution and an expected value of 0. DSS: Depression Stigma Scale; GAD-7: General Anxiety Disorder-7; GSES: General Self-Efficacy Scale; K-10: Kessler 10-Item Psychological Distress Scale; PHQ-9: Patient Health Questionnaire-9; SDS: Sheehan Disability Scale; SF-6D: Short Form Six-Dimension.



The within- and between-group effect sizes measured by Cohen *d* were calculated based on the estimated means (Table S2 in Multimedia Appendix 5). For depressive symptoms, the between-group effect size was moderate (*d*=0.50; 95% CI 0.26 - 0.75). The within-group effect size was large for the ICBT group (*d*=1.02; 95% CI 0.76 - 1.27), whereas in the WLC group, the effect size was small (*d*=0.34; 95% CI 0.11 - 0.57). For secondary outcomes, the between-group effect sizes were small (*d* range=0.07 - 0.43). The within-group effect sizes in the ICBT group ranged from small to large, with GAD-7 (*d*=0.89; 95% CI 0.63-1.14) and K-10 (*d*=0.87; 95% CI 0.61-1.12) exhibiting large sizes, SDS (*d*=0.58; 95% CI 0.33-0.83) showing moderate size, and the left displaying small sizes (*d* range=0.01 - 0.39). The within-group effect sizes in the WLC group were small, with a *d* range of 0.10-0.31.

Sensitivity analyses of short-term effectiveness outcomes using the ITT sample (Tables S2-S4 and Figure S1 in Multimedia Appendix 3) and the PPS population (Tables S2-S4 and Figure S1 in Multimedia Appendix 4) yielded results concordant with those of the primary analysis.

Outcome Measures at Follow-Up

Table 3 presents the observed and estimated means of outcomes among all participants receiving ICBT treatment in the ICBT I and II groups by time points. Figure 3 demonstrates the change course of all outcome measures during the intervention and follow-up period. Follow-up models demonstrated that improvements in outcomes were maintained across 12-month follow-ups except for DSS and its subscales. The results of paired comparisons are displayed in Table S3 in Multimedia Appendix 5. Paired comparisons confirmed no significant change from posttreatment to 12-month follow-up on PHQ-9 (Bonferroni adjusted estimated mean difference −0.81; SE=0.33; *P*=.33), GAD-7 (Bonferroni adjusted estimated mean difference −0.63; SE=0.31; *P*=.63), K-10 (Bonferroni adjusted estimated mean difference −1.11; SE=0.55; *P*=.64), SDS (Bonferroni adjusted estimated mean difference 0.88; SE=0.47; *P*=.33), GSES (Bonferroni adjusted estimated mean difference 0.12; SE=0.38; *P*>.99), DSS (Bonferroni adjusted estimated mean difference −0.49; SE=0.69; *P*=.95), and perceived DSS (Bonferroni adjusted estimated mean difference 0.58; SE=0.47; *P*=.74), while significant increases in SF-6D (Bonferroni adjusted estimated mean difference −0.04; SE=0.01; *P*=.04) and personal DSS (Bonferroni adjusted estimated mean difference −1.06; SE=0.36; *P*=.03; Table 3).

XSL•FO
RenderX

Table . Observed and estimated means across time points in participants receiving ICBT treatment in the immediate ICBT[a] and waitlist control group.

Outcomes	Pretreatment, T0[b]	Posttreatment, T1[c]	3 months, T2[d]	6 months, T3[e]	12 months, T4[f]	*F* test	*P* value[g]	*P* value (T0 vs T1)[h]	*P* value (T1 vs T4)[h]
Observed, mean (SD)						—[i]	—	—	—
PHQ-9[j] scores	13.56 (4.96)	8.63 (5.56)	9.26 (5.56)	9.64 (5.95)	9.47 (5.90)				
GAD-7[k] scores	10.02 (4.74)	6.46 (4.89)	7.24 (5.39)	7.30 (5.39)	7.08 (5.50)				
K-10[l] scores	29.32 (7.85)	23.24 (9.26)	23.78 (9.64)	24.59 (10.05)	24.52 (9.94)				
SDS[m] scores	14.56 (7.06)	10.65 (6.91)	9.96 (7.75)	9.72 (7.52)	9.58 (7.54)				
GSES[n] scores	19.80 (5.54)	21.32 (5.74)	21.07 (6.24)	21.26 (6.64)	20.98 (6.92)				
SF-6D[o] scores	0.57 (0.20)	0.61 (0.23)	0.62 (0.23)	0.65 (0.22)	0.64 (0.23)				
DSS[p] scores	53.53 (8.92)	52.63 (10.16)	53.81 (9.50)	54.06 (9.92)	53.31 (10.13)				
Personal DSS scores	23.08 (5.29)	22.21 (5.63)	22.93 (5.73)	23.24 (6.14)	23.28 (6.41)				
Perceived DSS scores	30.45 (6.52)	30.42 (7.32)	30.87 (7.05)	30.82 (7.48)	30.03 (6.99)				
Estimated, mean (SE)[q]									
PHQ-9 scores	13.56 (0.35)	8.85 (0.37)	9.46 (0.37)	9.83 (0.38)	9.66 (0.39)	24.8(1181)	<.001	<.001	.33
GAD-7 scores	10.02 (0.32)	6.51 (0.34)	7.28 (0.34)	7.39 (0.34)	7.14 (0.35)	21.2 (1180)	<.001	<.001	.63
K-10 scores	29.32 (0.59)	23.37 (0.62)	23.84 (0.62)	24.63 (0.63)	24.47 (0.64)	21.8 (1179)	<.001	<.001	.64
SDS scores	14.56 (0.45)	10.84 (0.49)	10.22 (0.49)	9.94 (0.49)	9.96 (0.51)	22.2 (448)	<.001	<.001	.33
GSES scores	19.80 (0.38)	21.34 (0.41)	21.25 (0.41)	21.33 (0.42)	21.22 (0.43)	7.39 (449)	<.001	<.001	>.99
SF-6D scores	0.57 (0.01)	0.61 (0.01)	0.62 (0.01)	0.65 (0.01)	0.65 (0.02)	9.11 (450)	<.001	.02	.04
DSS scores	53.53 (0.60)	52.44 (0.64)	53.43 (0.66)	53.66 (0.66)	52.93 (0.69)	.35 (457)	.84	.42	.95
Personal DSS scores	23.11 (0.36)	22.18 (0.38)	22.80 (0.39)	23.11 (0.39)	23.24 (0.40)	2.11 (446)	.08	.046	.03
Perceived DSS scores	30.42 (0.43)	30.24 (0.47)	30.60 (0.47)	30.52 (0.48)	29.66 (0.49)	.99 (456)	.41	.99	.74

[a]ICBT: internet-based cognitive behavioral therapy.

[b]T0: pretreatment.

[c]T1: posttreatment.

[d]T2: 3-month follow-up.

[e]T3: 6-month follow-up.

[f]T4: 12-month follow-up.

[g]*P* values were associated with the test (*F* test, fixed effects) of the overall time period difference.

[h]*P* values adjusted by the Bonferroni method are shown.

[i]Not applicable.

[j]PHQ-9: Patient Health Questionnaire-9.

XSL•FO
RenderX

[k]GAD-7: General Anxiety Disorder-7.

[l]K-10: Kessler 10-Item Psychological Distress Scale.

[m]SDS: Sheehan Disability Scale.

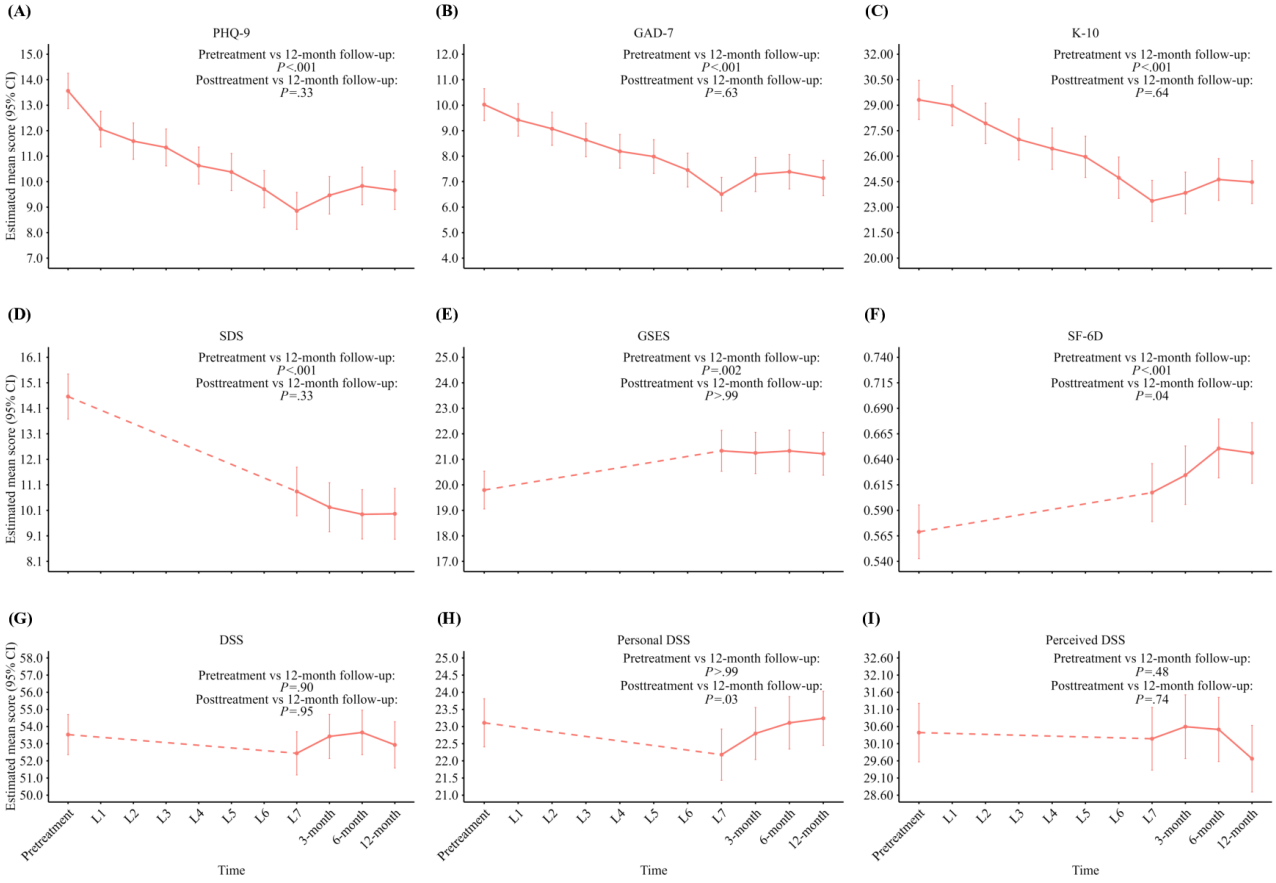
[n]GSES: General Self-Efficacy Scale.

[o]SF-6D: Short Form Six-Dimension.

[p]DSS: Depression Stigma Scale.

[q]Estimated means are based on the follow-up models.

Figure 3. Estimated marginal means and 95% CIs across all time points for participants receiving ICBT intervention in both ICBT and waitlist control groups. Marginal linear models are applied to calculate estimated marginal means, mean differences, and their *P* values. Red dashed lines mean that no available data can be used for the intervention period. DSS: Depression Stigma Scale; GAD-7: General Anxiety Disorder-7; GSES: General Self-Efficacy Scale; ICBT: internet-based cognitive behavioral therapy; K-10: Kessler 10-Item Psychological Distress Scale; PHQ-9: Patient Health Questionnaire-9; SDS: Sheehan Disability Scale; SF-6D: Short Form Six-Dimension.



The within-group effect sizes for different outcome measures across time points are presented in Table S4 in Multimedia Appendix 5. During the follow-up period, the effect sizes for PHQ-9 were maintained at moderate levels ($d \geq 0.50$) compared with that of the pretreatment stage. For GAD-7, K-10, and SDS, moderate effect sizes were also observed over 12 months (d range=0.52 - 0.73). The remaining secondary outcomes (eg, SF-6D and GSES) had shown small effect sizes both at posttreatment and follow-ups. Compared with the posttreatment stage, the effect sizes for all outcomes at follow-up were small, ranging from 0.01 to 0.18.

The results of the 2 sensitivity analyses for the long-term effectiveness evaluation were almost consistent with those of the main analysis (Tables S5-S8 and Figures S1 and S2 in Multimedia Appendix 5). For personal DSS, both the main analysis and sensitivity analyses showed down first and then up, with only the main analysis reporting statistical significance.

Remission and Response

Table S9 in Multimedia Appendix 5 displays the numbers and rates of participants reaching remission from depression and responding to treatment by trial arms and across time points. At posttreatment, a total of 80 complete remissions and 50 responders were observed in 152 patients who received ICBT treatment. The remission and response rates reached 66.1% (80/121) and 41.3% (50/121), respectively. The clinical symptomatic remissions in the WLC group were significantly lower than those of the ICBT group, with complete remissions in 58 (39.2%) cases (P<.001) and response in 27 of the 148 (18.2%) patients (P<.001). During the follow-up period, the remission and response rates in the ICBT group remained high, with rates of 58.2% and 32.7%.

XSL•FO
RenderX

Treatment Acceptability and Satisfaction

A total of 119 participants completed the course evaluation regarding perceived helpfulness, satisfaction, and recommended intention. In total, 98 (82.4%) reported that the course was helpful, 117 (98.3%) reported feeling satisfied with the course, and 111 (93.2%) participants were willing to recommend the course to others around them.

Discussion

Principal Findings

This study assessed the short- and long-term effectiveness of an unguided ICBT course in Chinese patients with MDD and evaluated participants' acceptability and satisfaction with ICBT. Compared to WLC, depressive symptoms, disease-related symptoms, individual and social functioning, and quality of life in the ICBT group were significantly improved at the posttreatment stage, with the between-group effect size of PHQ-9 reaching 0.50. Remission and response rates of the ICBT group were statistically higher than those of the WLC group (80/121, 66.1% vs 58/148, 39.2%; P<.001; 50/121, 41.3% vs 27/148, 18.2%; P<.001). Moreover, these positive results in the ICBT group were maintained after the intervention. At 12-month follow-up, the depressive symptoms were improved compared with that at pretreatment (mean difference 3.90, SE 0.32; P<.001; d=0.70), and no significant change was observed in comparison with the outcomes at posttreatment (mean difference −0.81, SE 0.33; P=.33; d=−0.15). Patients treated with ICBT reported high acceptability with an engagement rate of 77.6%. Above 90% of participants in the ICBT group reported high treatment satisfaction. To our knowledge, this was the first study that investigated both the short- and long-term effectiveness of unguided ICBT supported by nonspecialists for MDD in China.

Comparisons With Previous Work

The result that unguided ICBT as an additional treatment to usual care was more effective than usual care alone at posttreatment was well supported by the findings of most previous meta-analyses and RCTs on unguided ICBT [16,17,30,31,34]. The demonstrated effectiveness of the Morning Mood for Chinese clients with MDD can be attributed to its strict adherence to CBT principles and thoughtful cultural adaptation. First, the program successfully delivers core CBT components in a structured, sequential manner. This allows users to systematically learn and practice skills to break the cycle of negative thoughts and behaviors that maintain depression, even without a therapist guiding each step. Second, the program was delivered in the participants' native language, ensuring clarity and facilitating a deeper understanding of complex psychoeducational content. Third, culturally resonant materials were used, such as incorporating familiar idioms and real-life scenarios that reflect the typical stressors and automatic thoughts experienced by the mainland Chinese population. This likely fostered a stronger connection with the content and improved the self-application of CBT skills. Fourth, the program's design aligns with modern digital consumption habits, prioritizing video-based demonstrations and interactive exercises over text-heavy narratives, which suits contemporary learning preferences. Finally, the integration of mindfulness and relaxation exercises resonated with the traditional Chinese philosophy of mind-body harmony, thereby complementing the Western CBT framework and making the therapeutic concepts more intuitively understandable and personally meaningful to the target audience.

The ICBT group exhibited a statistically significant improvement in depressive symptoms than the WLC group, with a medium between-group size of 0.50. However, the effect size calculated in this study was lower than that of a meta-analysis (0.65) [17]. The relatively poor performance could be attributed to the supporting provider and implementing environment, which led to different treatment effects [16]. Previous investigations usually provided support by clinicians and were implemented in clinical settings [68], while this study differed by providing support through nonspecialists and being conducted in a more practical setting. It is acknowledged that unguided ICBT with clinician support leads to better treatment compliance due to the authority of doctors [23,42]. Moreover, the clinical setting has contextual factors, favoring the deployment and uptake of the intervention [51]. Therefore, higher effect sizes were reported in previous research. However, given the shortage of mental health specialists in China, unguided ICBT supported by nonspecialists might be more pragmatic and workable, with the potential of increasing accessibility, availability, and affordability of MDD treatment. More measures should be taken to enhance treatment outcomes when putting Morning Mood into large-scale use.

This study has also proven the sustained long-term effectiveness within the ICBT treatment conditions, with no significant difference in depressive symptoms between the posttreatment and follow-up stages. High remission and response rates were also observed both at posttreatment and follow-up under the ICBT intervention. This finding was consistent with the existing research [18-20,22,25,26,28,29]. However, 1 meta-analysis [35] and 2 large RCTs [21,23] reported a significantly large magnitude of effect at follow-up compared to the posttreatment outcome despite the varied within-group effect size. This could be partially explained by the study characteristics. In our study, patients were diagnosed with MDD, and more than half of them had antidepressant treatments, indicating a more severe disease condition. It is mentioned that the short- and long-term treatment effects of unguided ICBT are better in individuals with mild to moderate depressive symptoms [34]. The long-term effect of ICBT on depressive symptoms and absolute treatment outcomes suggested that intensified measures might be needed to consolidate the desired therapeutic effects for the long term in our future application.

Our results showed that at posttreatment, participants in the ICBT group experienced a significantly larger reduction of anxiety symptoms, nonspecific psychological distress, and depression stigma than individuals in the WLC group, while a greater improvement of self-efficacy and social functioning. Existing research on ICBT for MDD reported similarly favorable effects except for depression stigma [69-73]. These results provided new supportive evidence for the addition of unguided ICBT to usual care in developing countries.

XSL•FO
RenderX

We also found that ICBT produced long-term positive effects on anxiety symptoms, nonspecific distress, social functioning, general self-efficacy, and quality of life. However, few studies have comprehensively investigated these outcomes in the long run. As comorbidity depression with anxiety and other psychiatric disorders are common [74,75], ICBT interventions for depression have been proposed to be potentially effective in improving general mental health, including anxiety symptoms and nonspecific distress, which aligns with the findings in previous research [20,30,32]. General self-efficacy, which essentially reflects the subjective thoughts of participants, was considered to be an indicator of cognitive and behavioral changes in patients with MDD. It was found that ICBT intervention has significantly increased the individuals' self-efficacy even after the acute treatment phase. The ultimate aim of the ICBT intervention is to develop the psychological and emotional skills to restore social functioning and improve quality of life. This self-developed ICBT course has finally reached its goals, which agrees with the results of prior research [23,32,35]. Together, these results indicate that ICBT for MDD has widespread favorable outcomes.

The majority of ICBT completers were satisfied with the course, and the engagement rate of the ICBT intervention in this study was high, with 77.6% (118/152) of individuals completing the course. A previous study reported that the average completion rate of a self-guided intervention with general support was 65% [16]. Several factors could account for the notably higher rate in our study. Compared with clinicians, nonspecialists have more time and energy to dedicate to the provision of general support to prompt engagement. Furthermore, the ICBT course in our study was designed for Chinese people with good cultural adaptability and was embedded on the WeChat app with good user convenience. Taken together, these findings showed that unguided ICBT supported by nonspecialists could enhance the treatment effect in multiple dimensions.

Implications

This trial provides evidence on the short- and long-term effectiveness of the unguided ICBT intervention (ie, Morning Mood) for MDD in China across multiple dimensions: depressive symptoms, disease-related symptoms, psychosocial functioning, and quality of life. The high adherence rate (77.6%) in our study suggests that the program's delivery format and support intensity align well with the needs of individuals with MDD in China. The unguided ICBT supported by nonspecialists shows a positive effect for MDD in conditions that require fewer resources, with the cost-effectiveness being proven in a previous study [76]. Therefore, it is indicated that integrating the unguided ICBT into the primary community health centers or other care pathways is a good choice in China, where trained therapists are limited, social discrimination is serious, and the accessibility of mental health care services is poor [77]. Second, the small to moderate between-group effect sizes in short-term outcomes emphasize the necessity of optimizing the efficacy by recognizing those who respond well to the intervention and those who show no response to treatment, which could further increase the cost-effectiveness [78]. Third, the suboptimal enduring remission rate (58.2%) underscores proactive monitoring in real-world applications to take timely intervention measures and prevent further deterioration.

Strengths and Limitations

This study has several strengths. First, the ICBT program in this study is developed based on Chinese culture and habits of Chinese expressions, with examples demonstrated in the course being selected from real life; thus, the effect of cultural and linguistic incompatibility was excluded. Moreover, the ICBT intervention is delivered through the WeChat Mini Program, which makes the implementation more convenient than web- or app-based forms. Participants do not need to download an additional app or search on the Internet. To some extent, this study has provided more robust evidence of ICBT for depression in China. Second, in our study, the unguided ICBT intervention is supported by nonspecialist health workers (ie, any type of health worker, like a nurse, lay health worker, or medical social worker, who is not a specialist in mental health). This form of ICBT has no requirement for qualified mental health specialists and thus is easier to implement. The results of our study provide a reference for solving the problem of the treatment gap in resource-limited countries like China. Third, measurement tools with good reliability and validity have been used to evaluate the short- and long-term effectiveness of ICBT from multiple dimensions, including depressive symptoms, disease-related symptoms, individual and social functioning, and quality of life. Few previous studies have comprehensively explored these perspectives.

There are also some limitations in this study. First, due to ethical considerations, participants in the WLC group are arranged to receive ICBT treatment after an 8-week intervention period. That is, no control exists during the follow-up phase. As a result, it is difficult to distinguish the effect of ICBT from usual care and the natural course. A further RCT with long-term follow-ups in all treatment arms should be conducted to solve this problem. Second, the participants in our study were recruited from mental health clinics in Shenzhen, China, and were diagnosed with MDD. The majority of them were female and received high levels of education. Therefore, the generalization of the present findings may be limited. Third, due to data qualification of the sample, we could only compare the proportions of participants taking antidepressants in the ICBT and WLC groups. It is thought that drug types, dosage, and duration influence the treatment outcomes.

Conclusions

This study has demonstrated the short- and long-term effectiveness of unguided ICBT with general support delivered by nonspecialists for patients with MDD from multiple dimensions, including depressive symptoms, anxiety symptoms, nonspecific psychological distress, social functioning, quality of life, and general self-efficacy. Participants report high acceptability of and satisfaction with the ICBT course. Taken together, these results add to existing literature consolidating unguided ICBT supported by nonspecialists as an effective treatment to solve the substantial treatment discrepancies in low-resource settings. Additional studies with larger samples are needed to examine potential factors that could enhance the long-term effectiveness of ICBT.

XSL•FO
RenderX

Acknowledgments

The authors thank all the participants of the study. This manuscript was prepared without the use of generative artificial intelligence.

Funding

This work was supported by the National Natural Science Foundation of China (81761128030) and Sanming Project of Medicine in Shenzhen Nanshan (grant 11).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

All authors conceptualized and designed the trial. WZ, H Zhang, YJ, Y Liao, Y Liu, and JH conducted the trial and follow-up. WZ, Y Li, GS, and H Zhao collected and prepared the data. WZ and H Zhang analyzed and interpreted the patient data and were major contributors in writing the manuscript. WW and RSM polished the manuscript and revised the paper. BF and CL offered administrative, technical, and material support for the implementation of the trial. BF and CL provided supervision for the whole trial. CL obtained the funding for the trial. All authors read and approved the final manuscript.

Conflicts of Interest

RSM has received research grant support from the Canadian Institutes of Health Research, the Global Alliance for Chronic Diseases, the National Natural Science Foundation of China, and the Milken Institute. RSM also received speaker and consultation fees from Lundbeck, Janssen, Alkermes, Neumora Therapeutics, Boehringer Ingelheim, Sage, Biogen, Mitsubishi Tanabe, Purdue, Pfizer, Otsuka, Takeda, Neurocrine, Neurawell, Sunovion, Bausch Health, Axsome, Novo Nordisk, Kris, Sanofi, Eisai, Intra-Cellular, NewBridge Pharmaceuticals, Viatris, AbbVie, Bristol Myers Squibb (BMS), Teva, Adhere Tech, GH Research, Autobahn Therapeutics, and Atai Life Sciences. RSM is the chief executive officer of Braxia Scientific Corp. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Outline of the treatment content, homework, and corresponding screenshots of the internet-based cognitive behavioral therapy program.
[DOCX File, 433 KB - mhealth_v14i1e68394_app1.docx]

Multimedia Appendix 2
Criteria for defining treatment completion and participant attrition.
[DOCX File, 14 KB - mhealth_v14i1e68394_app2.docx]

Multimedia Appendix 3
Sensitivity analysis in the intention-to-treat sample.
[DOCX File, 244 KB - mhealth_v14i1e68394_app3.docx]

Multimedia Appendix 4
Sensitivity analysis in the per-protocol set.
[DOCX File, 267 KB - mhealth_v14i1e68394_app4.docx]

Multimedia Appendix 5
Supplementary results of the main analysis.
[DOCX File, 527 KB - mhealth_v14i1e68394_app5.docx]

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 1359 KB - mhealth_v14i1e68394_app6.pdf]

References

1. Ferrari AJ, Santomauro DF, Aali A, Abate YH, Abbafati C, Abbastabar H. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries

XSL•FO
RenderX

in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2024 May 18;403(10440):2133-2161. [doi: 10.1016/S0140-6736(24)00757-8] [Medline: 38642570]
2. Marx W, Penninx B, Solmi M, et al. Major depressive disorder. Nat Rev Dis Primers 2023 Aug 24;9(1):44. [doi: 10.1038/s41572-023-00454-1] [Medline: 37620370]
3. Lundberg J, Cars T, Lööv S, et al. Clinical and societal burden of incident major depressive disorder: a population-wide cohort study in Stockholm. Acta Psychiatr Scand 2022 Jul;146(1):51-63. [doi: 10.1111/acps.13414] [Medline: 35165894]
4. Mental health, brain health and substance use. World Health Organization. 2017. URL: https://www.who.int/teams/mental-health-and-substance-use/promotion-prevention/mental-health-in-the-workplace [accessed 2023-03-29]
5. Cuijpers P, Noma H, Karyotaki E, Vinkers CH, Cipriani A, Furukawa TA. A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. World Psychiatry 2020 Feb;19(1):92-107. [doi: 10.1002/wps.20701] [Medline: 31922679]
6. Thornicroft G, Chatterji S, Evans-Lacko S, et al. Undertreatment of people with major depressive disorder in 21 countries. Br J Psychiatry 2017 Feb;210(2):119-124. [doi: 10.1192/bjp.bp.116.188078] [Medline: 27908899]
7. Evans-Lacko S, Aguilar-Gaxiola S, Al-Hamzawi A, et al. Socio-economic variations in the mental health treatment gap for people with anxiety, mood, and substance use disorders: results from the WHO World Mental Health (WMH) surveys. Psychol Med 2018 Jul;48(9):1560-1571. [doi: 10.1017/S0033291717003336] [Medline: 29173244]
8. Gao Y, Burns R, Leach L, Chilver MR, Butterworth P. Examining the mental health services among people with mental disorders: a literature review. BMC Psychiatry 2024 Aug 20;24(1). [doi: 10.1186/s12888-024-05965-z]
9. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: from innovation to implementation. World Psychiatry 2019 Feb;18(1):20-28. [doi: 10.1002/wps.20610] [Medline: 30600624]
10. Etxeandia-Ikobaltzeta I, Crandall CJ, Yost J, Owens DK, Qaseem A. Nonpharmacologic and pharmacologic treatments of adults in the acute phase of major depressive disorder: a living clinical guideline from the American College of Physicians. Ann Intern Med 2024 Dec;177(12):1742. [doi: 10.7326/ANNALS-24-02180] [Medline: 39680950]
11. Depression in Adults: Treatment and Management: National Institute for Health and Care Excellence (NICE); 2022. [Medline: 35977056]
12. Feng Y, Xiao L, Wang WW, et al. Guidelines for the diagnosis and treatment of depressive disorders in China: the second edition. J Affect Disord 2019 Jun 15;253:352-356. [doi: 10.1016/j.jad.2019.04.104] [Medline: 31078835]
13. Que J, Lu L, Shi L. Development and challenges of mental health in China. Gen Psychiatry 2019 Jan;32(1). [doi: 10.1136/gpsych-2019-100053]
14. Sun M, Zhou H, Li Y, et al. Professional characteristics, numbers, distribution and training of China's mental health workforce from 2000 to 2020: a scoping review. Lancet Reg Health West Pac 2024 Apr;45:100992. [doi: 10.1016/j.lanwpc.2023.100992] [Medline: 38699293]
15. Lu J, Xu X, Huang Y, et al. Prevalence of depressive disorders and treatment in China: a cross-sectional epidemiological study. Lancet Psychiatry 2021 Nov;8(11):981-990. [doi: 10.1016/S2215-0366(21)00251-0] [Medline: 34559991]
16. Karyotaki E, Riper H, Twisk J, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 1;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
17. Mamukashvili-Delau M, Koburger N, Dietrich S, Rummel-Kluge C. Efficacy of computer- and/or internet-based cognitive-behavioral guided self-management for depression in adults: a systematic review and meta-analysis of randomized controlled trials. BMC Psychiatry 2022 Nov 24;22(1):730. [doi: 10.1186/s12888-022-04325-z] [Medline: 36424570]
18. Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K. Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010 Jun 8;5(6):e10939. [doi: 10.1371/journal.pone.0010939] [Medline: 20544030]
19. Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. Cogn Behav Ther 2011;40(4):251-266. [doi: 10.1080/16506073.2011.616531] [Medline: 22060248]
20. Choi I, Zou J, Titov N, et al. Culturally attuned Internet treatment for depression amongst Chinese Australians: a randomised controlled trial. J Affect Disord 2012 Feb;136(3):459-468. [doi: 10.1016/j.jad.2011.11.003] [Medline: 22177742]
21. Mohr DC, Duffecy J, Ho J, et al. A randomized controlled trial evaluating a manualized TeleCoaching protocol for improving adherence to a web-based intervention for the treatment of depression. PLoS One 2013;8(8):e70086. [doi: 10.1371/journal.pone.0070086] [Medline: 23990896]
22. Newby JM, Mackenzie A, Williams AD, et al. Internet cognitive behavioural therapy for mixed anxiety and depression: a randomized controlled trial and evidence of effectiveness in primary care. Psychol Med 2013 Dec;43(12):2635-2648. [doi: 10.1017/S0033291713000111] [Medline: 23419552]
23. Gilbody S, Littlewood E, Hewitt C, et al. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. BMJ 2015 Nov 11;351:h5627. [doi: 10.1136/bmj.h5627] [Medline: 26559241]

XSL•FO
RenderX

24. Montero-Marín J, Araya R, Pérez-Yus MC, et al. An internet-based intervention for depression in primary care in Spain: a randomized controlled trial. J Med Internet Res 2016 Aug 26;18(8):e231. [doi: 10.2196/jmir.5695] [Medline: 27565118]
25. Mantani A, Kato T, Furukawa TA, et al. Smartphone cognitive behavioral therapy as an adjunct to pharmacotherapy for refractory depression: randomized controlled trial. J Med Internet Res 2017 Nov 3;19(11):e373. [doi: 10.2196/jmir.8602] [Medline: 29101095]
26. Newby J, Robins L, Wilhelm K, et al. Web-based cognitive behavior therapy for depression in people with diabetes mellitus: a randomized controlled trial. J Med Internet Res 2017 May 15;19(5):e157. [doi: 10.2196/jmir.7274] [Medline: 28506956]
27. Rosso IM, Killgore WDS, Olson EA, et al. Internet-based cognitive behavior therapy for major depressive disorder: a randomized controlled trial. Depress Anxiety 2017 Mar;34(3):236-245. [doi: 10.1002/da.22590] [Medline: 28009467]
28. Smith J, Newby JM, Burston N, et al. Help from home for depression: a randomised controlled trial comparing internet-delivered cognitive behaviour therapy with bibliotherapy for depression. Internet Interv 2017 Sep;9:25-37. [doi: 10.1016/j.invent.2017.05.001] [Medline: 30135834]
29. Stiles-Shields C, Montague E, Kwasny MJ, Mohr DC. Behavioral and cognitive intervention strategies delivered via coached apps for depression: pilot trial. Psychol Serv 2019 May;16(2):233-238. [doi: 10.1037/ser0000261] [Medline: 30407055]
30. Ghosh A, Cherian RJ, Wagle S, et al. An unguided, computerized cognitive behavioral therapy intervention (TreadWill) in a lower middle-income country: pragmatic randomized controlled trial. J Med Internet Res 2023 Apr 26;25:e41005. [doi: 10.2196/41005] [Medline: 37099376]
31. Yeung A, Wang F, Feng F, et al. Outcomes of an online computerized cognitive behavioral treatment program for treating Chinese patients with depression: a pilot study. Asian J Psychiatr 2018 Dec;38:102-107. [doi: 10.1016/j.ajp.2017.11.007] [Medline: 29146042]
32. Schure M, McCrory B, Tuchscherer Franklin K, Greist J, Weissman RS. Twelve-month follow-up to a fully automated internet-based cognitive behavior therapy intervention for rural adults with depression symptoms: single-arm longitudinal study. J Med Internet Res 2020 Oct 2;22(10):e21336. [doi: 10.2196/21336] [Medline: 33006561]
33. Zainal NH, Soh CP, Van Doren N, Benjet C. Do the effects of internet-delivered cognitive-behavioral therapy (i-CBT) last after a year and beyond? A meta-analysis of 154 randomized controlled trials (RCTs). Clin Psychol Rev 2024 Dec;114:102518. [doi: 10.1016/j.cpr.2024.102518] [Medline: 39579466]
34. Karyotaki E, Efthimiou O, Miguel C, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry 2021 Apr 1;78(4):361-371. [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
35. Mamukashvili-Delau M, Koburger N, Dietrich S, Rummel-Kluge C. Long-term efficacy of internet-based cognitive behavioral therapy self-help programs for adults with depression: systematic review and meta-analysis of randomized controlled trials. JMIR Ment Health 2023 Aug 22;10:e46925. [doi: 10.2196/46925] [Medline: 37606990]
36. Beevers CG, Pearson R, Hoffman JS, Foulser AA, Shumake J, Meyer B. Effectiveness of an internet intervention (Deprexis) for depression in a United States adult sample: a parallel-group pragmatic randomized controlled trial. J Consult Clin Psychol 2017 Apr;85(4):367-380. [doi: 10.1037/ccp0000171] [Medline: 28230390]
37. Lin Z, Cheng L, Han X, et al. The effect of internet-based cognitive behavioral therapy on major depressive disorder: randomized controlled trial. J Med Internet Res 2023 Sep 22;25:e42786. [doi: 10.2196/42786] [Medline: 37738092]
38. De Jesús-Romero R, Holder-Dixon AR, Buss JF, Lorenzo-Luaces L. Race, ethnicity, and other cultural background factors in trials of internet-based cognitive behavioral therapy for depression: systematic review. J Med Internet Res 2024 Feb 1;26:e50780. [doi: 10.2196/50780] [Medline: 38300699]
39. Liu Z, Qiao D, Xu Y, et al. The efficacy of computerized cognitive behavioral therapy for depressive and anxiety symptoms in patients with COVID-19: randomized controlled trial. J Med Internet Res 2021 May 14;23(5):e26883. [doi: 10.2196/26883] [Medline: 33900931]
40. Zhao C, Wampold BE, Ren Z, Zhang L, Jiang G. The efficacy and optimal matching of an Internet-based acceptance and commitment therapy intervention for depressive symptoms among university students: a randomized controlled trial in China. J Clin Psychol 2022 Jul;78(7):1354-1375. [doi: 10.1002/jclp.23329] [Medline: 35191525]
41. Ying Y, Ji Y, Kong F, et al. Efficacy of an internet-based cognitive behavioral therapy for subthreshold depression among Chinese adults: a randomized controlled trial. Psychol Med 2023 Jul;53(9):3932-3942. [doi: 10.1017/S0033291722000599] [Medline: 35388776]
42. Kong F, Yu L, Hou Y, et al. Efficacy of internet-based cognitive behavioral therapy for subthreshold depression among older adults in institutional long-term care settings: pragmatic randomized controlled trial. J Med Internet Res 2024 Mar 1;26:e40187. [doi: 10.2196/40187] [Medline: 38427424]
43. Pan JY, Rafi J. Culturally adapted guided internet-based cognitive behavioral therapy for Hong Kong people with depressive symptoms: randomized controlled trial. J Med Internet Res 2025 Feb 25;27:e64303. [doi: 10.2196/64303] [Medline: 39998865]
44. Yang H, Gao S, Li J, et al. Remission of symptoms is not equal to functional recovery: Psychosocial functioning impairment in major depression. Front Psychiatry 2022;13:915689. [doi: 10.3389/fpsyt.2022.915689]

XSL•FO
RenderX

45. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
46. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33. [Medline: 9881538]
47. Guze SB. Diagnostic and Statistical Manual of Mental Disorders, 4th edition: AJP American Psychiatric Publishing; 1995, Vol. 152:1228-1228. [doi: 10.1176/ajp.152.8.1228]
48. Barrera M, Castro FG. A heuristic framework for the cultural adaptation of interventions. Clin Psychol Sci Pract 2006 Dec;13(4):311-316. [doi: 10.1111/j.1468-2850.2006.00043.x]
49. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 1999 Nov 10;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
50. Clark PD, Oates M. Measuring improvement and recovery adult services version 2. Improving Access to Psychological Therapies. 2014. URL: https://www.healthinnovationoxford.org/wp-content/uploads/2015/11/measuring-recovery-2014.pdf [accessed 2026-05-20]
51. Richards D, Enrique A, Eilert N, et al. A pragmatic randomized waitlist-controlled effectiveness and cost-effectiveness trial of digital interventions for depression and anxiety. NPJ Digit Med 2020;3(85):85. [doi: 10.1038/s41746-020-0293-8] [Medline: 32566763]
52. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
53. Kessler RC, Andrews G, Colpe LJ, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002 Aug;32(6):959-976. [doi: 10.1017/s0033291702006074] [Medline: 12214795]
54. Sheehan DV, Harnett-Sheehan K, Raj BA. The measurement of disability. Int Clin Psychopharmacol 1996 Jun;11 Suppl 3:89-95. [doi: 10.1097/00004850-199606003-00015] [Medline: 8923116]
55. Brazier JE, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002 Mar;21(2):271-292. [doi: 10.1016/S0167-6296(01)00130-8]
56. Schwarzer R, Born A. Optimistic self-beliefs: assessment of general perceived self-efficacy in thirteen cultures. World Psychol 1997;3:177-190 [FREE Full text]
57. Griffiths KM, Christensen H, Jorm AF. Predictors of depression stigma. BMC Psychiatry 2008 Apr 18;8:25. [doi: 10.1186/1471-244X-8-25] [Medline: 18423003]
58. Dossing A, Tarp S, Furst DE, et al. Modified intention-to-treat analysis did not bias trial results. J Clin Epidemiol 2016 Apr;72:66-74. [doi: 10.1016/j.jclinepi.2015.11.003] [Medline: 26562052]
59. Abraha I, Cherubini A, Cozzolino F, et al. Deviation from intention to treat analysis in randomised trials and treatment effect estimates: meta-epidemiological study. BMJ 2015 May 27;350:h2445. [doi: 10.1136/bmj.h2445] [Medline: 26016488]
60. Gold SM, Friede T, Meyer B, et al. Internet-delivered cognitive behavioural therapy programme to reduce depressive symptoms in patients with multiple sclerosis: a multicentre, randomised, controlled, phase 3 trial. Lancet Digit Health 2023 Oct;5(10):e668-e678. [doi: 10.1016/S2589-7500(23)00109-7] [Medline: 37775187]
61. International Conference on Harmonisation. E9: statistical principles for clinical trials. US Food & Drug Administration. 1998. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e9-statistical-principles-clinical-trials [accessed 2025-11-17]
62. Kahan BC, White IR, Edwards M, Harhay MO. Using modified intention-to-treat as a principal stratum estimator for failure to initiate treatment. Clin Trials 2023 Jun;20(3):269-275. [doi: 10.1177/17407745231160074] [Medline: 36916466]
63. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res 2011 Mar;20(1):40-49. [doi: 10.1002/mpr.329] [Medline: 21499542]
64. Rubin DB. Statistical matching using file concatenation with adjusted weights and multiple imputations. J Bus Econ Stat 1986 Jan;4(1):87-94. [doi: 10.1080/07350015.1986.10509497]
65. Feingold A. Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. Psychol Methods 2009 Mar;14(1):43-53. [doi: 10.1037/a0014699] [Medline: 19271847]
66. Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychol Methods 2002 Mar;7(1):105-125. [doi: 10.1037/1082-989x.7.1.105] [Medline: 11928886]
67. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
68. Rosenström TH, Saarni SE, Saarni SI, Tammilehto J, Stenberg JH. Efficacy and effectiveness of therapist-guided internet versus face-to-face cognitive behavioural therapy for depression via counterfactual inference using naturalistic registers and machine learning in Finland: a retrospective cohort study. Lancet Psychiatry 2025 Mar;12(3):189-197. [doi: 10.1016/S2215-0366(24)00404-8] [Medline: 39954684]
69. Morgan C, Mason E, Newby JM, et al. The effectiveness of unguided internet cognitive behavioural therapy for mixed anxiety and depression. Internet Interv 2017 Dec;10:47-53. [doi: 10.1016/j.invent.2017.10.003]

XSL•FO
RenderX

70. Kladnitski N, Smith J, Uppal S, et al. Transdiagnostic internet-delivered CBT and mindfulness-based treatment for depression and anxiety: a randomised controlled trial. Internet Interv 2020 Apr;20:100310. [doi: 10.1016/j.invent.2020.100310] [Medline: 32140427]
71. Earley C, Joyce C, McElvaney J, Richards D, Timulak L. Preventing depression: qualitatively examining the benefits of depression-focused iCBT for participants who do not meet clinical thresholds. Internet Interv 2017 Sep;9:82-87. [doi: 10.1016/j.invent.2017.07.003] [Medline: 30135841]
72. Lindegaard T, Kashoush F, Holm S, Halaj A, Berg M, Andersson G. Experiences of internet-based cognitive behavioural therapy for depression and anxiety among Arabic-speaking individuals in Sweden: a qualitative study. BMC Psychiatry 2021 Jun 3;21(1):288. [doi: 10.1186/s12888-021-03297-w] [Medline: 34082745]
73. Maj A, Michalak N, Graczykowska A, Andersson G. The effect of internet-delivered cognitive behavioral therapy for depression and anxiety on quality of life: a meta-analysis of randomized controlled trials. Internet Interv 2023 Sep;33:100654. [doi: 10.1016/j.invent.2023.100654] [Medline: 37555075]
74. Lamers F, van Oppen P, Comijs HC, et al. Comorbidity patterns of anxiety and depressive disorders in a large cohort study: the Netherlands Study of Depression and Anxiety (NESDA). J Clin Psychiatry 2011 Mar;72(3):341-348. [doi: 10.4088/JCP.10m06176blu] [Medline: 21294994]
75. Kessler RC, Sampson NA, Berglund P, et al. Anxious and non-anxious major depressive disorder in the World Health Organization World Mental Health Surveys. Epidemiol Psychiatr Sci 2015 Jun;24(3):210-226. [doi: 10.1017/S2045796015000189] [Medline: 25720357]
76. Zhou W, Chen Y, Wu H, et al. Cost utility analysis of internet-based cognitive behavioral therapy for major depressive disorder: randomized controlled trial. J Med Internet Res 2025 Feb 19;27:e67567. [doi: 10.2196/67567] [Medline: 39970426]
77. Zhu Y, Sun L, Guan S, et al. Evolving role of the internet in China as a medium for promoting psychological well-being and mental health: a scoping review. Lancet Reg Health West Pac 2025 Mar;56:101515. [doi: 10.1016/j.lanwpc.2025.101515] [Medline: 40226776]
78. Cumpanasoiu DC, Enrique A, Palacios JE, Duffy D, McNamara S, Richards D. Trajectories of symptoms in digital interventions for depression and anxiety using routine outcome monitoring data: secondary analysis study. JMIR Mhealth Uhealth 2023 Jul 12;11:e41815. [doi: 10.2196/41815] [Medline: 37436812]

Abbreviations

CBT: cognitive behavioral therapy
CONSORT: Consolidated standards of reporting trials
DSM-4: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition
DSS: Depression Stigma Scale
GAD-7: Generalized Anxiety Disorder-7
GSES: General Self-Efficacy Scale
ICBT: internet-based cognitive behavioral therapy
ITT: intention-to-treat
K-10: Kessler 10-Item Psychological Distress Scale
LMM: linear mixed model
M.I.N.I.: Mini-International Neuropsychiatric Interview
MDD: major depressive disorder
mITT: modified intention-to-treat
PHQ-9: Patient Health Questionnaire-9
PPS: per-protocol set
RCT: randomized controlled trial
SDS: Sheehan Disability Scale
SF-6D: Short Form Six-Dimension
WLC: waitlist control

Edited by L Buis; submitted 04.Nov.2024; peer-reviewed by D Sato, JY Pan; revised version received 23.Nov.2025; accepted 02.Apr.2026; published 24.Jun.2026.

Please cite as:
Zhou W, Zhang H, Jiang Y, Li Y, Shi G, Zhao H, Wang W, Liao Y, Liu Y, Hao J, McIntyre RS, Fan B, Lu C
Long-Term Effectiveness of Unguided Internet-Based Cognitive Behavioral Therapy on Major Depressive Disorder in Chinese Adults: Randomized Controlled Trial With a 12-Month Follow-Up
JMIR Mhealth Uhealth 2026;14:e68394
URL: https://mhealth.jmir.org/2026/1/e68394
doi:10.2196/68394

XSL•FO
RenderX

© Wenjing Zhou, Huimin Zhang, Yunbin Jiang, Yanzhi Li, Guangduoji Shi, Hao Zhao, Wanxin Wang, Yuhua Liao, Yifeng Liu, Jiejing Hao, Roger S McIntyre, Beifang Fan, Ciyong Lu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 24.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Efficacy of a Cognitive Behavioral Therapy–Based Online Self-Help Group for Depression and Suicide Ideation: Randomized Controlled Trial

Minkyung Yim[1*], PhD; Haeun Kim[2*], MA; Soo-Eun Lee[3], MA; Eunjin Jo[4], MA; Ji-Won Hur[1], Prof Dr

[1]School of Psychology, Korea University, #115, College of Law Annex (Old Law Bldg), 145 Anam-ro, Seoul, Republic of Korea

[2]Department of Psychiatry, Sungkyunkwan University College of Medicine, Samsung Medical Center, Seoul, Republic of Korea

[3]Department of Psychology, Virginia Commonwealth University, Richmond, Virginia, United States

[4]Department of Psychiatry, Kangbuk Samsung Hospital, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Ji-Won Hur, Prof Dr
School of Psychology, Korea University, #115, College of Law Annex (Old Law Bldg), 145 Anam-ro, Seoul, Republic of Korea

Abstract

Background: Despite the high prevalence of depressive disorders, access to effective treatment remains limited due to financial, geographic, and social barriers. Online self-help groups offer a promising and scalable form of peer-based support beyond traditional clinical settings. Integrating cognitive behavioral therapy (CBT) techniques such as cognitive restructuring and behavioral activation into self-help groups may enhance their effectiveness.

Objective: This study evaluated the efficacy of a cognitive behavioral therapy–based online self-help group (COS) that integrates structured CBT techniques with peer-led group support as a low-intensity intervention for individuals with depressive symptoms. A randomized controlled trial (RCT) comparing COS with a CBT-based mobile application was conducted. Additionally, a separately recruited waitlist control group was included as a supplementary comparison condition.

Methods: Participants were recruited online. After eligibility screening via a structured clinical interview, participants were randomly assigned to a COS group (n=79) or a CBT-based mobile application group (n=39). An additional waitlist control group (n=48) was recruited separately during the second phase of the study. The COS intervention involved 7 videoconferencing sessions that incorporated peer-led group discussions, sharing lived experiences, and core CBT techniques such as cognitive restructuring. The primary outcome measure was depressive symptoms, assessed using the Beck Depression Inventory-II, and the secondary outcome was suicidal ideation, estimated using the Beck Scale for Suicide Ideation, measured at baseline, postintervention, and 3-month follow-up. Linear mixed models were used to evaluate group × time interaction effects. Reliable change indices were also calculated to assess clinical significance. All statistical tests were 2-tailed.

Results: Among participants assigned to the COS group, 61% (48/79) completed all 7 sessions, and 84% (66/79) attended 5 or more sessions. A significant time × group interaction was observed for depressive symptoms ($F_{4,288.47}$=7.23, P<.001). The COS group exhibited a substantial reduction in depressive symptoms from baseline to postintervention ($t_{285.76}$=10.77, two-tailed; P<.001), with a large within-group effect size (d=1.38); this improvement was maintained at the 3-month follow-up. Suicidal ideation also significantly decreased in the COS group ($t_{277.11}$=4.55, two-tailed; P<.001), with sustained effects at follow-up. Clinically meaningful improvement in depressive symptoms, as defined by the reliable change index, was observed in 75% (56/75) of COS participants. While both the COS and app-based CBT groups achieved comparable reductions in depressive symptoms, only the COS group demonstrated a significant reduction in suicidal ideation.

Conclusions: This RCT provides evidence that a structured, CBT-informed online self-help group can reduce depressive symptoms and suicidal ideation. The COS program offers a scalable, accessible alternative to traditional therapy, particularly in settings with limited access to mental health professionals.

Trial Registration: Clinical Research Information Service (CRiS) KCT0007673; https://tinyurl.com/328madxe

(JMIR Mhealth Uhealth 2026;14:e76028) doi:10.2196/76028

KEYWORDS

depression; self-help group; suicidal ideation; randomized controlled trials; cognitive behavioral therapy

XSL•FO
RenderX

Introduction

Depressive disorders have far-reaching public health consequences, contributing to work absenteeism, increased economic burden, elevated suicide risk, and higher all-cause mortality [1-3]. The duration of untreated depression is a critical determinant of clinical outcomes; shorter durations are associated with faster remission and better treatment response [4]. Thus, timely access to effective intervention is essential for mitigating the course and consequences of depressive illness [5,6].

Despite this urgency, approximately 30% of individuals diagnosed with depression receive no treatment at all [7]. Common barriers include financial hardship, geographic inaccessibility, stigma, and a shortage of qualified mental health professionals [8,9]. These systemic barriers underscore the need for alternative models of care that are both accessible and scalable. Digital health technologies, including videoconferencing and mobile applications, have expanded access to mental health care by providing flexible, cost-effective, and evidence-based interventions for underserved populations [10-13]. Among these, online self-help groups are promising platforms for delivering support outside traditional clinical settings [14].

Traditional self-help groups are typically peer-led and rooted in shared lived experiences [15], fostering mutual support among individuals facing common challenges [14,16,17]. Participation in these groups has been associated with reduced depressive symptoms [18,19], improvements in quality of life [20], decreased feelings of loneliness [21], and overall reductions in mental health problems [22,23]. These benefits, supported by meta-analyses and systematic reviews [18,20,22,24], underscore the potential value of self-help groups as scalable, low-barrier interventions.

However, the utility of traditional in-person self-help groups is often curtailed by significant structural impediments. Community-based local groups depend on physical proximity, regular attendance, and a sufficient number of members to maintain operations, which often serve as barriers to entry and sustainability [25]. Individuals residing in rural areas, or those burdened by mobility limitations, physical illness, caregiving responsibilities, or irregular working hours, often find in-person attendance untenable [25-27]. Furthermore, in smaller communities, concerns regarding anonymity or privacy may deter participation, while low population density may preclude the formation of groups entirely. These persistent barriers underscore the urgent need for delivery formats capable of extending the reach of self-help interventions beyond local constraints.

The application of digital health technologies, particularly real-time videoconferencing, offers a robust solution to these challenges. By enabling participation irrespective of location, reducing the burden of travel, and facilitating flexible scheduling, these technologies effectively mitigate traditional barriers while enhancing anonymity [28,29]. Crucially, synchronous online formats preserve key therapeutic elements of self-help groups: mutual support and real-time emotional exchange among individuals with lived experiences. Consequently, online self-help groups represent a vital mobile health innovation, leveraging the technology to connect participants nationwide and expanding access to structured psychological support where traditional models fall short.

The term “online self-help group” encompasses a diverse range of formats, ranging from large, asynchronous text-based discussion forums to synchronous, small group-based mutual support sessions. This study focuses specifically on the latter—synchronous, videoconferencing-based groups designed to foster real-time interaction and support. Despite the promise of this format, there is a notable gap in the literature [30]: to date, few randomized controlled trials (RCTs) have rigorously evaluated the clinical efficacy of synchronous online self-help groups in treating depression. Moreover, a key limitation of typical self-help groups is their lack of structured therapeutic content, resulting in wide variability in group facilitation, composition, and outcomes [31]. This inconsistency poses challenges to their reliability and scalability as interventions for individuals with clinically significant depressive symptoms. In this study, we developed a semistructured manual for a cognitive behavioral therapy–based online group self-help program (COS). We then examined its efficacy among individuals experiencing depressive symptoms through an RCT. The COS program integrated core cognitive behavioral therapy (CBT) components, including cognitive restructuring and coping skills, to enhance therapeutic consistency and effectiveness within a peer-led group format. This hybrid model builds on prior work, such as SMART Recovery [32], showing that structured CBT elements can improve psychological outcomes in mutual aid settings.

A noninteractive, CBT-based mobile application served as the active comparator. While this application has demonstrated efficacy in reducing depressive symptoms through interventions targeting dysfunctional thoughts [33], it is designed for individual use and does not offer the social support and accountability that characterize group-based interventions such as the COS program. In addition, a separately recruited waitlist control group was included to account for potential symptom improvement attributable to the natural course of depression.

Consequently, we hypothesized that the COS program would be at least as efficacious as the CBT-based mobile application intervention for general CBT-based outcomes, such as depressive symptoms, stress, and dysfunctional attitudes, and that the peer-based format of COS might offer additional benefits for outcomes involving emotional or interpersonal processes. Furthermore, we expected that a higher proportion of individuals in the COS group would demonstrate clinically meaningful improvement.

To our knowledge, few RCTs have evaluated online self-help groups for depression, and prior studies have typically examined only unstructured peer-support interventions. This study contributes to this literature by examining a semistructured, CBT-informed online self-help group and by including an active digital CBT comparator. We anticipate that the findings will highlight the potential of online self-help groups as a scalable, low-barrier intervention that can help overcome the logistical

XSL•FO
RenderX

and geographic challenges often related to accessing traditional, evidence-based practices.

Methods

Study Design

This study primarily used a 2-arm RCT in which participants were randomly assigned to either the COS or the cognitive behavioral therapy–based mobile application (APP) group at a 2:1 ratio, using block randomization. Participants of the RCT were recruited between May and November 2021. Interventions were delivered between August and December 2021, and follow-up assessments were completed in March 2022.

To complement the primary RCT design, a nonrandomized waitlist control group was recruited during the second phase of the study from December 2021 to April 2022. This waitlist control group provided a baseline for the natural course of depression. While the active comparator (APP) has demonstrated efficacy in reducing depressive symptoms [33], relying exclusively on this digital intervention prevents a direct assessment of absolute efficacy relative to a no-treatment control. Therefore, following the completion of enrollment for the randomized arms, a waitlist control group was recruited to distinguish the intervention effect from spontaneous remission.

Participant enrollment proceeded in two distinct phases due to operational and timeline constraints. During the initial recruitment period, the research team prioritized the randomized comparison between the COS and APP groups, as this constituted the primary objective of the trial. To minimize potential confounding, identical inclusion and exclusion criteria were applied across phases, and sociodemographic characteristics were monitored rigorously.

Recruitment and Screening Procedures

Participants for the RCT were recruited between May and November 2021 through online advertisements disseminated via social media platforms (eg, X and Instagram) and university-based online community platforms in Korea. Interested individuals were invited to contact the research team via direct message or email to receive detailed study information and to schedule an initial telephone screening.

Inclusion criteria were as follows: (1) aged between 19 and 55 years, (2) diagnosed with any depressive disorder based on the Structured Clinical Interview for DSM-5 – Clinician Version (SCID-5-CV) or having a PHQ-9 (Patient Health Questionnaire-9) score of ≥5 indicating mild depression [34], (3) fluent in Korean, and (4) able to use a real-time videoconferencing program. Participants undergoing pharmacotherapy were eligible only if their dosage had been stable for at least 1 month.

Exclusion criteria included current psychotic symptoms (ie, hallucinations and delusions), concurrent psychological interventions, or high suicide risk. Individuals with high suicide risk, defined as those reporting suicide ideation with planning during the SCID-5-CV, were excluded and referred to appropriate services. No participants were excluded on this basis.

Recruitment followed a 2-step screening procedure. Interested individuals first completed a brief telephone screening that assessed medication stability, current psychological intervention, and commitment to the study schedule. Eligible participants received a brief explanation of the study after eligibility screening (SCID-5-CV) and provided informed consent via an online signature. Of 186 individuals initially screened, 141 underwent online SCID-5-CV interviews and initial assessments. SCID-5-CV interviews were conducted by graduate students in clinical and counseling psychology or licensed clinical psychologists, all of whom received more than 2 weeks of SCID-5-CV administration training. A total of 127 participants were randomized to the COS or APP group.

For the recruitment and screening of the waitlist control group, the same inclusion and exclusion criteria used for the RCT were applied to the waitlist control, except that the requirement for real-time videoconferencing ability was waived. During this phase, 74 individuals were screened, 55 completed eligibility procedures, and 48 were enrolled. The study adhered to the CONSORT (Consolidated Standards of Reporting Trials) 2010 guidelines [35], and the CONSORT checklist is provided as Checklist 1.

Sample Size Calculation

The sample size was determined based on power estimates and feasibility considerations. We adopted a statistical power of 0.80 and a significance level of α=.05. The assumed effect size was set to f=0.25, a conservative estimate informed by prior meta-analyses of digital psychological interventions and peer support–based approaches, both of which generally report small-to-moderate improvements in depressive symptoms [36,37]. Using f=0.25 for detecting a group × time interaction in a repeated-measures design, a total of approximately 66 participants were required. Given the 2:1 randomization ratio between the COS and APP groups, we aimed to recruit at least 99 participants to ensure adequate power.

Randomization

Block randomization was performed using a computer-generated list, with a fixed block size of 3. Participants were randomly assigned to the COS or APP group at a 2:1 ratio.

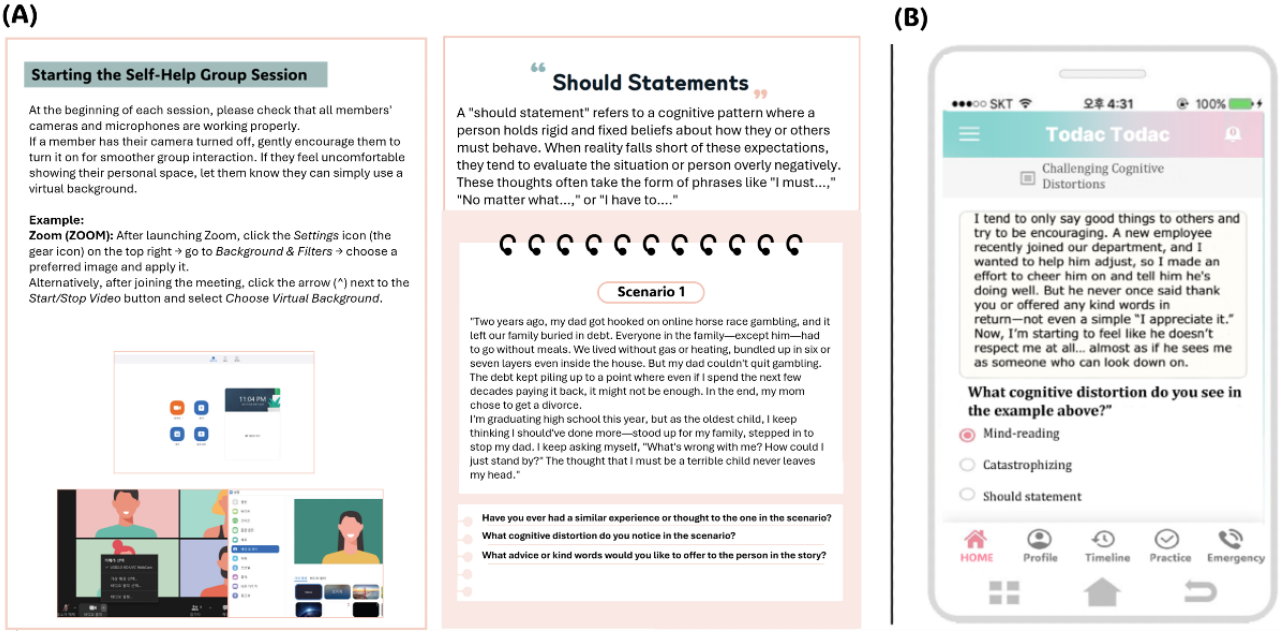
Interventions

The COS Program

The COS program is a semistructured, 7-session, CBT-based online self-help group protocol designed to alleviate depressive symptoms. The COS manual and workbooks were developed and reviewed by licensed clinical psychologists to ensure fidelity to the self-help group format and CBT principles and accuracy of therapeutic content. Figure 1A shows a sample page from the participant workbook provided in the COS program.

XSL•FO
RenderX

Figure 1. (A) Sample page from the participant workbook provided in the COS program. (B) Screenshot of the cognitive distortion exercise featured in the cognitive behavioral therapy–based mobile application. These materials were originally developed in Korean and translated into English. COS: cognitive behavioral therapy–based online self-help group.



Each group typically consisted of 4 - 6 participants and 1 facilitator, with each session lasting approximately 50 minutes. Sessions were held twice weekly via videoconferencing platforms such as Zoom. A face-to-face pilot version of the COS program was conducted in 2020 to refine content and structure prior to web-based adaptation [38]. The first session of COS focused on rapport-building, sharing lived experiences with depression, and psychoeducation. Subsequent sessions (2-7) focused on introducing various cognitive distortions (eg, catastrophizing, mind reading, overgeneralization, "should" statements, tunnel vision, and black-and-white thinking). In each session, participants read a series of scenarios written in the manual of the COS program, reflecting on each cognitive distortion and suggesting possible alternatives for the protagonist. After reading the given scenario, the participants were encouraged to share personal experiences related to these negative and distressing emotions related to each cognitive distortion. These activities were designed to provide participants insight into their negative automatic thoughts and emotional support from others with similar experiences.

Although the COS program was fundamentally designed to permit participants to assume the facilitator role, a designated facilitator who was either a graduate psychology student or a licensed clinical psychologist was assigned in this study. This was implemented to ensure strict adherence to protocols that monitor participant safety, in accordance with research ethical standards. To facilitate the intended peer-led model, the program includes a publicly available participant manual and a step-by-step facilitator guidebook, alongside a brief introductory video provided through the Korean National Center for Mental Health. Crucially, substantive discussions, including the sharing of lived experiences and the restructuring of cognitive distortions, remained participant-led. All facilitators received 30 to 40 minutes of training on COS procedures before implementation. Their role was strictly limited to logistical management, such as opening the videoconferencing sessions, introducing session activities, and timekeeping, alongside the monitoring of participant safety. Facilitators were trained to monitor participant safety and to notify the research team immediately if a participant reported active suicidal thoughts. In such cases, they were instructed to administer the P4 screener [39] to assess suicide risk and to contact the research team for a follow-up safety evaluation. No instances of active suicidal ideation were reported during the study.

The CBT-Based Mobile Application

We used a CBT-based mobile application as an active comparator. The efficacy of the app was confirmed in an RCT [33]. Similar to the COS program, the app presented short stories illustrating frustrating situations that highlighted various cognitive distortions, followed by quizzes designed to identify these distortions and reduce related negative emotions (Figure 1B). Participants in the APP group were instructed to use the app for approximately 270 to 360 minutes (5 minutes per session × 3 sessions daily × 24 days), ensuring comparable total usage time to the COS program. Weekly reminders were sent to encourage adherence to the app usage schedule. Participants were considered to have completed the APP program if they used the app for at least 18 days (270 minutes), completing 3 or more sessions per day.

Outcome Measures

The primary outcome was depressive symptoms, assessed using the Beck Depression Inventory-II (BDI-II), a widely used measure of depression [40]. Scores from 0 to 13 indicate no or minimal depressive symptoms, 14 to 19 mild, 20 to 28 moderate depression, and 29 to 63 severe depressive symptoms.

The secondary outcome was suicidal ideation, measured using the Beck Scale for Suicide Ideation (BSS). The BSS consists of 21 items that assess the severity of suicidal symptoms [41].

XSL•FO
RenderX

Exploratory outcomes included cognitive and psychosocial variables. Dysfunctional attitudes, a predisposing cognitive structure underlying depression, were measured using the Dysfunctional Attitude Scale (Form A) Revised (DAS-17) [42,43]. Daily life stressors were measured using the Perceived Stress Scale (PSS) [44]. Loneliness and perceptions of social support were assessed using the UCLA Loneliness Scale version 3 (UCLA-LS) [45,46] and the Perceived Available Support (PAS) scale of the Berlin Social Support Scale (BSSS) [47], respectively. Protective factors against suicide risk were assessed using the Reasons for Living Inventory for Young Adults (RFL-YA) [48]. Finally, general self-efficacy was measured using the General Self-Efficacy Scale (GSES) [49], given its known association with depression. All outcome measures were administered via Qualtrics as online self-report questionnaires.

Statistical Analysis

An intent-to-treat analysis was conducted to compare the immediate and 3-month efficacy of the programs using a linear mixed model (LMM). Prior to the main analysis, the appropriateness of the LMM for the repeated-measures design was confirmed by evaluating within-participant dependence using unconditional mixed-effects models. Intraclass correlation coefficients (ICCs) were calculated for this preliminary assessment. A significant (P<.05) interaction between time and condition in the hypothesized directions indicated that one group showed greater improvement.

Prior to the main analysis, we examined the missingness mechanism for the primary outcome, BDI-II. Little's missing completely at random (MCAR) test was insignificant (χ^2_3=0.96, P=.62), and baseline BDI-II scores did not predict missingness (P=.24), supporting a missing at random (MAR) mechanism. Missing data were not imputed because the mixed-model approach provided equal or better power without imputing missing data for the intent-to-treat analysis under MAR assumptions [50]. To test the efficacy of the intervention, we evaluated the differences in scores between groups at each time point (T_1, T_2, and T_3). In the LMM, all the analyses included concurrent medication and social contact frequency as covariates. For post hoc analyses, both uncorrected P values and Holm-Bonferroni corrected P values (P_{holm}) were reported. Effect sizes were calculated using Cohen d. Within-group effect sizes for within-group changes were adjusted for correlations between repeated measurements to account for the dependency between time points [51-53].

To determine the statistical significance of the changes in the primary outcome measures between baseline and postintervention, we calculated a reliable change index (RCI) [54,55]. The RCI was computed using the RCI calculator [56]. A significant RCI indicates that the observed change in an individual participant's score over time is statistically meaningful. We used chi-square analysis to compare the proportions of individuals in each group demonstrating meaningful improvements or deterioration. All statistical tests were 2-tailed.

Ethical Considerations

The study protocol was reviewed and approved by the Institutional Review Board (IRB) of Korea University (KUIRB-2021-0157-06), and the trial was retrospectively registered with the Clinical Research Information Service (KCT0007673), a primary registry of the WHO International Clinical Trials Registry Platform.

All participants provided written informed consent after being informed about the study's objectives, procedures, and potential risks and benefits. Participants were explicitly informed that their involvement was voluntary and that they retained the right to withdraw from the study at any time without penalty. To ensure privacy, all participant data were anonymized prior to analysis, with identifiable information removed. Data were stored securely on password-protected servers, with access restricted strictly to authorized research personnel. Compensation was provided for each component of the study, including the screening interview, intervention sessions, and assessments at baseline, postintervention, and the 3-month follow-up. Participants who completed all study components received a maximum total compensation of approximately KRW 115,000 (about US $78 as of December 5, 2025).

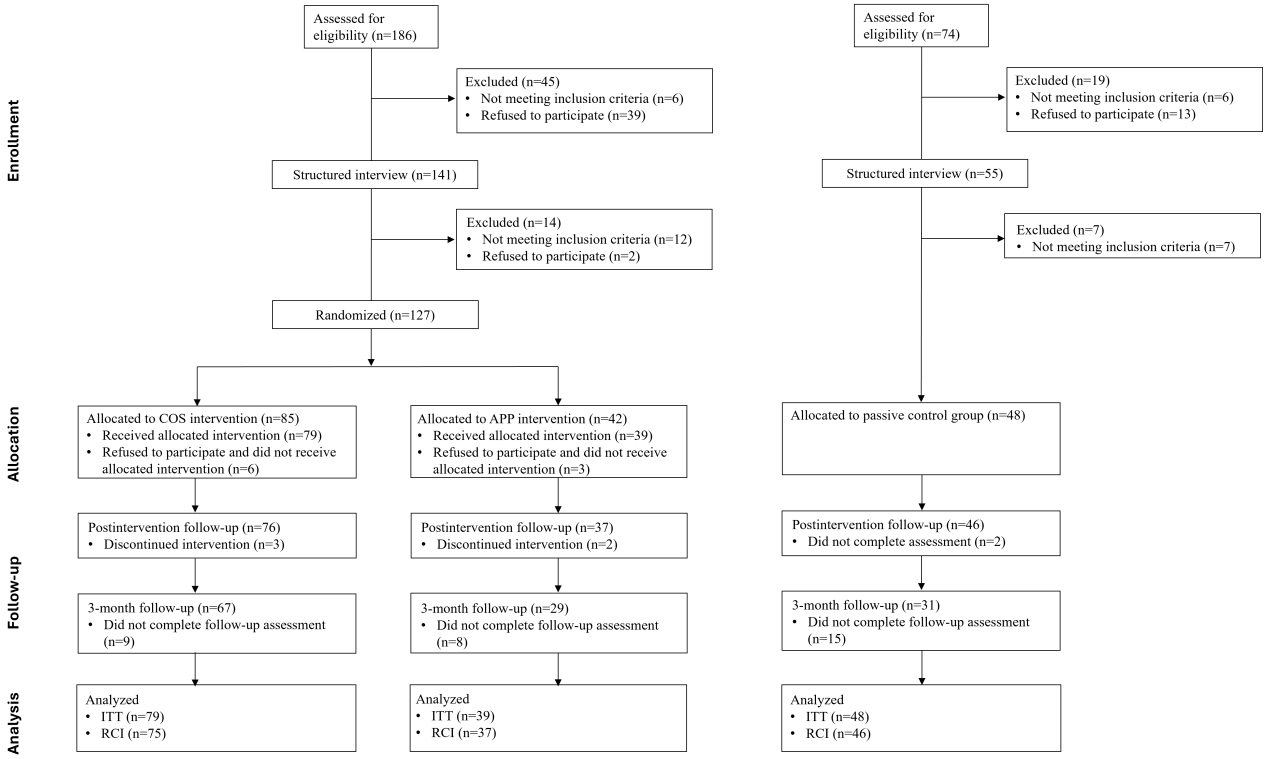
Results

Adherence and Baseline Sociodemographic and Clinical Characteristics

Figure 2 presents the CONSORT participant flowchart. Of the 127 individuals randomized, 6 in the COS group and 3 in the APP group declined to participate prior to the start of the trial, resulting in 79 participants in the COS group and 39 in the APP group completing the baseline assessment (Table 1). In terms of adherence, 84% (n=66) of participants in the COS group attended 5 or more sessions of the program, and 82% (n=32) of those in the APP group used the app for at least 18 days.

XSL•FO
RenderX

Figure 2. The CONSORT flow diagram of the randomized controlled trials (left) and waitlist control group (right). APP: cognitive behavioral therapy–based mobile application group; COS: cognitive behavioral therapy–based online self-help group; ITT: intent-to-treat analysis; RCI: reliable change index analysis.



XSL•FO
RenderX

Table . Baseline sociodemographic characteristics of the participants.

Characteristics	COS[a] (n=79)	APP[b] (n=39)	Waitlist control (n=48)	Chi-square (*df*) or *F* test (*df*)	*P* value
Age (years), mean (SD)	27.01 (6.05)	27.69 (6.82)	26.83 (7.47)	0.18 (2, 84.82)	.83
Sex, n (%)				1.32 (4)	.86
Female	71 (90)	34 (87)	44 (92)		
Male	7 (9)	4 (10)	4 (8)		
Prefer not to answer	1 (1)	1 (3)	0 (0)		
Sexual orientation, n (%)				3.07 (8)	.93
Heterosexual	55 (70)	29 (74)	38 (79)		
Gay or lesbian	1 (1)	1 (3)	1 (2)		
Bi- or pansexual	14 (18)	4 (10)	5 (10)		
Others	4 (5)	3 (8)	2 (4)		
Prefer not to answer	5 (6)	2 (5)	2 (4)		
Socioeconomic status, n (%)				11.27 (8)	.19
Lower	9 (11)	1 (3)	7 (15)		
Upper lower	16 (20)	5 (13)	12 (25)		
Middle	29 (37)	18 (46)	22 (46)		
Upper middle	24 (30)	14 (36)	6 (13)		
High	1 (1)	1 (3)	1 (2)		
Educational attainment, n (%)				2.25 (4)	.69
High school	3 (4)	2 (5)	2 (4)		
College	64 (81)	27 (69)	38 (79)		
Graduate school	12 (15)	10 (26)	8 (17)		
BDI-II[c], mean (SD)	27.03 (10.01)	26.72 (11.31)	26.48 (9.93)	0.05 (2, 87.31)	.96
BSS[d], mean (SD)	9.78 (7.11)	8.31 (8.36)	7.54 (7.97)	1.36 (2, 85.44)	.26
Concurrent pharmacotherapy, n (%)				1.2 (2)	.55
Positive	19 (24)	7 (18)	8 (17)		
Negative	60 (76)	32 (82)	40 (83)		

[a]COS: cognitive behavioral therapy–based online self-help group.

[b]APP: cognitive behavioral therapy–based mobile application group.

[c]BDI-II: Beck Depression Inventory-II.

[d]BSS: Beck Scale for Suicide Ideation.

There were no significant differences across groups in age, sex, sexual orientation, socioeconomic status, or educational level (all *P*>.05). Individuals receiving concurrent pharmacotherapy accounted for 20% of all participants. Chi-square analysis revealed no significant differences between the groups (*P*=.55). Baseline depressive symptoms (as measured by the BDI-II) and suicidal ideation (as measured by the BSS) also did not differ significantly across groups (both *P*>.05).

Primary Outcomes

Initial unconditional mixed-effects models indicated substantial within-subject clustering in depressive symptoms (ICC=0.59), supporting the use of LMM. Consistent with the study design, the primary comparison focused on the 2 randomized groups, COS and APP, while the waitlist group served as a nonrandomized comparison group to assess natural symptom fluctuation.

The LMM analysis revealed significant main effects of time ($F_{2,289.99}$=86.76, *P*<.001), group ($F_{2,163.68}$=3.32, *P*=.04), and

XSL•FO
RenderX

time × group interaction ($F_{4,288.47}$=7.23, P<.001) for depressive symptoms, as measured by the BDI-II (Table 2). Figure 3A illustrates changes in depressive symptoms across T_1, T_2, and T_3 for each group.

XSL•FO
RenderX

Table . Fixed effect omnibus test results for group and time.

Variables and effect	*F* test (*df*)	*P* value
BDI-II[a]		
Time	86.76 (2, 289.99)	<.001
Group	3.32 (2, 163.68)	.04
Time × group	7.23 (4, 288.47)	<.001
BSS[b]		
Time	7.66 (2, 277.57)	<.001
Group	0.23 (2, 161.83)	.80
Time × group	2.05 (4, 276.40)	.09
DAS-17[c]		
Time	21.2 (2, 288.82)	<.001
Group	1.67 (2, 164.81)	.19
Time × group	5.90 (4, 287.54)	<.001
PSS[d]		
Time	34.41 (2, 295.20)	<.001
Group	2.33 (2, 166.30)	.10
Time × group	2.72 (4, 293.49)	.03
GSES[e]		
Time	24.43 (2, 285.41)	<.001
Group	1.56 (2, 162.28)	.21
Time × group	4.21 (4, 284.24)	.003
UCLA-LS[f]		
Time	22.77 (2, 282.51)	<.001
Group	1.15 (2, 158.60)	.32
Time × group	3.21 (4, 281.24)	.01
PAS[g]		
Time	11.35 (2, 285.10)	<.001
Group	0.50 (2, 158.70)	.61
Time × group	0.49 (4, 283.46)	.75
RFL-YA[h]		
Time	14.26 (2, 281.49)	<.001
Group	1.19 (2, 159.51)	.31
Time × group	3.53 (4, 280.28)	.008

[a]BDI-II: Beck Depression Inventory-II.

[b]BSS: Beck Scale for Suicide Ideation.

[c]DAS-17: Dysfunctional Attitude Scale (Form A) Revised.

[d]PSS: Perceived Stress Scale.

[e]GSES: General Self-Efficacy Scale.

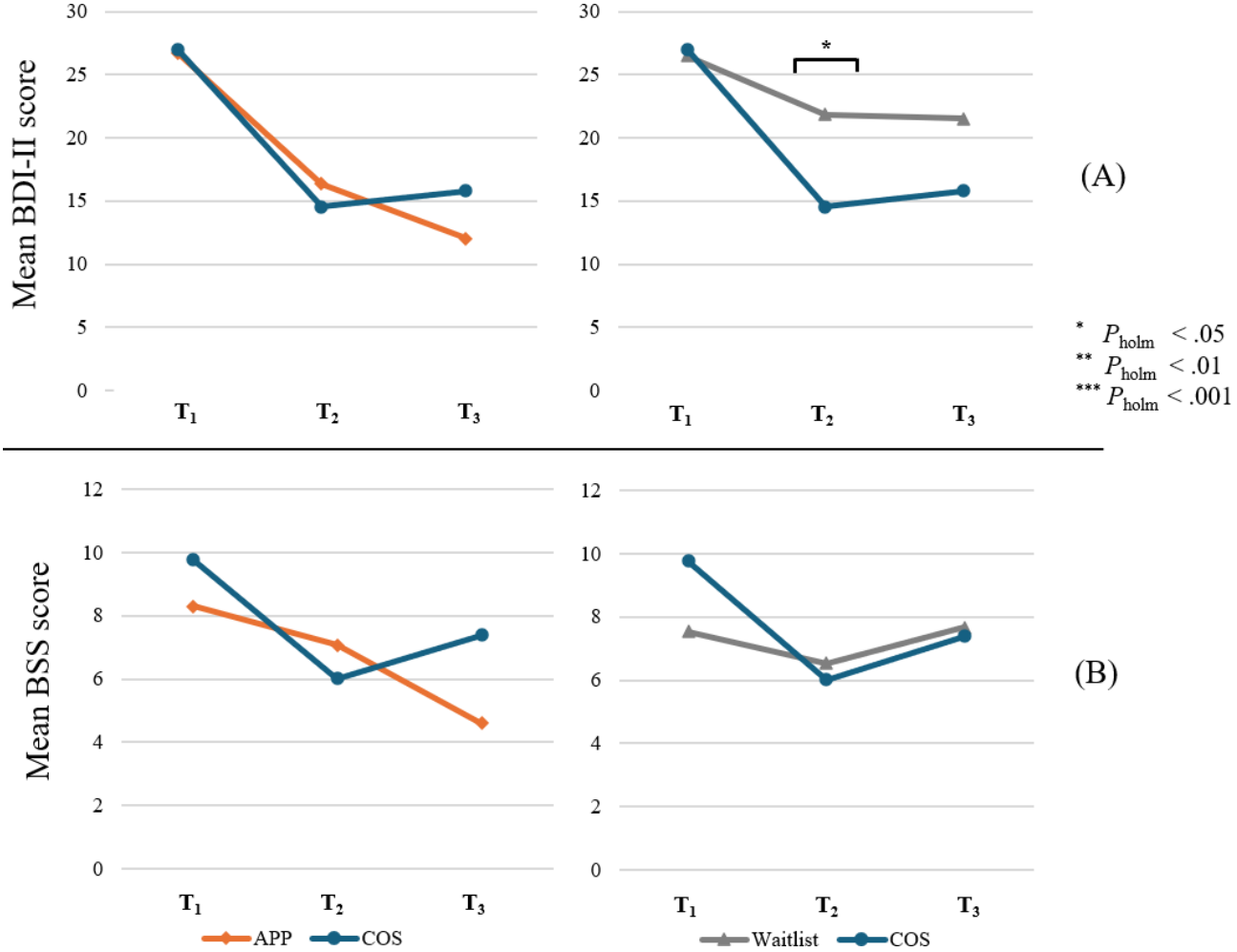
[f]UCLA-LS: UCLA Loneliness Scale version 3.

[g]PAS: Perceived Available Support.

[h]RFL-YA: Reasons for Living Inventory for Young Adults.

XSL•FO
RenderX

Figure 3. (A) Estimated mean BDI-II scores over time. (B) Estimated mean BSS scores over time. APP: cognitive behavioral therapy–based mobile application group; BDI-II: Beck Depression Inventory-II; BSS: Beck Scale for Suicide Ideation; COS: cognitive behavioral therapy–based online self-help group; P_{holm}: Holm-Bonferroni corrected *P* values; Waitlist: waitlist control group.



After each program (T_2), both the COS and APP groups showed significant reductions in depression from baseline (COS: difference=11.84, $t_{285.76}$=10.77, two-tailed, P_{holm}<.001; APP: difference=10.26, $t_{283.29}$=6.6, two-tailed, P_{holm}<.001) (Table 3). Mean BDI-II scores decreased from the moderate range at baseline (COS: T_1 mean 27.03, SD 10.01; APP: T_1 mean 26.72, SD 11.31) to the mild range at postintervention (COS: T_2 mean 14.57, SD 10.99; APP: T_2 mean 16.35, SD 10.87) after the program in both groups (Table 4). Within-group comparisons indicated large effect sizes in both groups (*d*=1.38 for COS; *d*=0.93 for APP).

At the 3-month follow-up (T_3), depressive symptoms in the COS group did not show a statistically significant change compared to T_2 (difference=−0.72; $t_{283.76}$=−0.62, two-tailed; *P*=.53). In the APP group, depressive symptoms decreased from T_2 to T_3 (difference=3.85; $t_{288.53}$=2.25, two-tailed; *P*=.03), although this reduction did not remain significant after Holm-Bonferroni correction (P_{holm}=.30). The difference between the COS and the APP group at T_3 was not significant (difference=3.18; $t_{168.7}$=1.43, two-tailed; *P*=.15), indicating comparable longer-term outcomes between the two randomized groups.

As secondary evidence, both the COS and the APP groups were compared with the waitlist control group. At T_2, the COS group showed a significantly greater reduction in depressive symptoms compared with the waitlist control group (difference=−6.9; $t_{279.71}$=−3.51, two-tailed; *P*<.001, P_{holm}=.01, *d*=0.68), indicating a medium effect size. The APP group also showed lower depressive symptoms than the waitlist controls at T_2, but this did not meet the Holm-Bonferroni correction (difference=−5.51; $t_{279.9}$=−2.39, two-tailed; *P*=.02, P_{holm}=.25).

At T_3, the COS group showed lower depressive symptoms than the waitlist control group (difference=−5.43; $t_{330.55}$=−2.56, two-tailed; *P*=.01), but this difference did not remain significant after Holm-Bonferroni correction (P_{holm}=.17). The APP group showed significantly lower depressive symptoms than the waitlist control group at T_3 (difference=−8.6; $t_{338.28}$=−3.42, two-tailed; *P*<.001, P_{holm}=.02).

XSL•FO
RenderX

Table . Time × group comparison for the primary and secondary measures.

	EMM[a] (SE)	EMM[a] (SE)	Difference	SE	*t* test (*df*)[b]	*P* value	P^{c}_{holm}
BDI-II[d], comparison by time points							
T_1 COS[e] vs T_2 COS	27.00 (1.26)	15.16 (1.29)	11.84	1.10	10.77 (285.76)	<.001	<.001
T_1 COS vs T_3 COS	27.00 (1.26)	15.88 (1.33)	11.12	1.15	9.71 (287.52)	<.001	<.001
T_2 COS vs T_3 COS	15.16 (1.29)	15.88 (1.33)	−0.72	1.15	−0.62 (283.76)	.53	≥.99
T_1 APP[f] vs T_2 APP	26.81 (1.77)	16.55 (1.80)	10.26	1.56	6.6 (283.29)	<.001	<.001
T_1 APP vs T_3 APP	26.81 (1.77)	12.70 (1.92)	14.11	1.70	8.28 (292.08)	<.001	<.001
T_2 APP vs T_3 APP	16.55 (1.80)	12.70 (1.92)	3.85	1.71	2.25 (288.53)	.03	.30
T_1 Waitlist[g] vs T_2 Waitlist	26.08 (1.64)	22.06 (1.66)	4.02	1.42	2.83 (284.55)	.005	.09
T_1 Waitlist vs T_3 Waitlist	26.08 (1.64)	21.30 (1.80)	4.78	1.58	3.02 (293.78)	.003	.06
T_2 Waitlist vs T_3 Waitlist	22.06 (1.66)	21.30 (1.80)	0.76	1.59	0.48 (288.52)	.63	≥.99
BDI-II, comparison by groups							
T_2 COS vs T_2 APP	15.16 (1.29)	16.55 (1.80)	−1.39	2.10	−0.66 (139.49)	.51	≥.99
T_2 COS vs T_2 Waitlist	15.16 (1.29)	22.06 (1.66)	−6.9	1.97	−3.51 (279.71)	<.001	.01
T_2 APP vs T_2 Waitlist	16.55 (1.80)	22.06 (1.66)	−5.51	2.31	−2.39 (279.9)	.02	.25
T_3 COS vs T_3 APP	15.88 (1.33)	12.70 (1.92)	3.18	2.24	1.43 (168.7)	.15	≥.99
T_3 COS vs T_3 Waitlist	15.88 (1.33)	21.30 (1.80)	−5.43	2.12	−2.56 (330.55)	.01	.17
T_3 APP vs T_3 Waitlist	12.70 (1.92)	21.30 (1.80)	−8.6	2.51	−3.42 (338.28)	<.001	.02
BSS[h], comparison by time points							
T_1 COS vs T_2 COS	10.28 (0.95)	7.15 (0.96)	3.13	0.69	4.55 (277.11)	<.001	<.001
T_1 COS vs T_3 COS	10.28 (0.95)	8.22 (0.98)	2.06	0.71	2.91 (275.76)	.004	.14
T_2 COS vs T_3 COS	7.15 (0.96)	8.22 (0.98)	−1.07	0.71	−1.5 (274.21)	.13	≥.99
T_1 APP vs T_2 APP	9.08 (1.31)	7.88 (1.33)	1.2	0.94	1.27 (272.63)	.21	≥.99
T_1 APP vs T_3 APP	9.08 (1.31)	6.45 (1.41)	2.63	1.05	2.53 (278.73)	.01	.44
T_2 APP vs T_3 APP	7.88 (1.33)	6.45 (1.41)	1.43	1.05	1.36 (276.39)	.18	≥.99

XSL•FO
RenderX

	EMM[a] (SE)	EMM[a] (SE)	Difference	SE	*t* test (*df*)[b]	*P* value	P^{c}_{holm}
T_1 Waitlist vs T_2 Waitlist	8.21 (1.22)	7.52 (1.23)	0.69	0.86	0.8 (273.74)	.43	≥.99
T_1 Waitlist vs T_3 Waitlist	8.21 (1.22)	7.68 (1.32)	0.53	0.99	0.54 (280.53)	.59	≥.99
T_2 Waitlist vs T_3 Waitlist	7.52 (1.23)	7.68 (1.32)	−0.16	0.99	−0.16 (276.51)	.87	≥.99
BSS, comparison by groups							
T_2 COS vs T_2 APP	7.15 (0.96)	7.88 (1.33)	−0.73	1.54	−0.47 (234.85)	.64	≥.99
T_2 COS vs T_2 Waitlist	7.15 (0.96)	7.52 (1.23)	−0.37	1.45	−0.25 (234.39)	.80	≥.99
T_2 APP vs T_2 Waitlist	7.88 (1.33)	7.52 (1.23)	0.36	1.70	0.21 (232.73)	.83	≥.99
T_3 COS vs T_3 APP	8.22 (0.98)	6.45 (1.41)	1.77	1.62	1.1 (270.36)	.27	≥.99
T_3 COS vs T_3 Waitlist	8.22 (0.98)	7.68 (1.32)	0.54	1.54	0.35 (278.59)	.72	≥.99
T_3 APP vs T_3 Waitlist	6.45 (1.41)	7.68 (1.32)	−1.23	1.82	−0.67 (286.75)	.50	≥.99

[a]The two EMM columns represent the first and second conditions being compared in each row, respectively. EMM: estimated marginal mean.

[b]All *t* tests were 2-tailed.

[c]P_{holm}: Holm-Bonferroni corrected *P* values.

[d]BDI-II: Beck Depression Inventory-II.

[e]COS: cognitive behavioral therapy–based online self-help group.

[f]APP: cognitive behavioral therapy–based mobile application group.

[g]Waitlist: waitlist control group.

[h]BSS: Beck Scale for Suicide Ideation.

XSL•FO
RenderX

Table . Mean (SD) and effect sizes for primary and secondary outcomes.

Outcome, time points, and group	Mean (SD)	Effect size (Cohen *d*)	
		Within-group	Between-group
BDI-II[a]			
T_1			
COS[b]	27.03 (10.01)	—[c]	—
APP[d]	26.72 (11.31)	—	—
Waitlist[e]	26.48 (9.93)	—	—
T_2			
COS	14.57 (10.99)	1.38	—
APP	16.35 (10.87)	0.93	COS vs APP: 0.16
Waitlist	21.87 (10.42)	0.64	COS vs Waitlist: 0.68
T_3			
COS	15.82 (12.86)	1.12[f]	—
APP	12.03 (10.20)	1.35[f]	COS vs APP: −0.33
Waitlist	21.53 (11.46)	0.48[f]	COS vs Waitlist: 0.47
BSS[g]			
T_1			
COS	9.78 (7.11)	—	—
APP	8.31 (8.36)	—	—
Waitlist	7.54 (7.97)	—	—
T_2			
COS	6.01 (7.08)	0.71	—
APP	7.08 (8.94)	0.23	COS vs APP: 0.13
Waitlist	6.54 (7.71)	025	COS vs Waitlist: 0.07
T_3			
COS	7.40 (8.74)	0.33[f]	—
APP	4.61 (6.04)	0.59[f]	COS vs APP: –0.37
Waitlist	7.68 (8.07)	−0.02[f]	COS vs Waitlist: 0.03

[a]BDI-II: Beck Depression Inventory-II.

[b]COS: cognitive behavioral therapy–based online self-help group.

[c]Not applicable.

[d]APP: cognitive behavioral therapy–based mobile application group.

[e]Waitlist: waitlist control group.

[f]Within-group effect sizes for T_3 were calculated relative to the baseline (T_1) score.

[g]BSS: Beck Scale for Suicide Ideation.

RCI Analysis

The RCI analysis included participants who completed both the T_1 and T_2 assessments. In the COS group, 76 participants completed the T_2 assessment; however, one individual did not complete the baseline assessment and was therefore excluded, resulting in a sample of 75 participants. The APP and waitlist control groups had 37 and 46 participants, respectively. In the COS group, 75% (56/75) showed a significant change from pre- to postinterventions. About 57% (21/37) of participants in the APP group and 35% (16/46) in the waitlist control group demonstrated significant changes over the same period. The difference in the proportion of participants with reliable change between the COS and APP groups was marginally significant (P=.054).

XSL•FO
RenderX

Secondary Outcomes

Within-subject dependence in suicidal ideation was identified (ICC=0.72), supporting the use of LMM. For suicidal ideation (as measured by the BSS), the LMM analysis revealed a significant main effect of time (P<.001) and a trend toward a time × group interaction (P=.09). The main effect of the group was not statistically significant (P=.80).

Figure 3B presents changes in suicidal ideation over time (T_1 – T_3) by group. Post hoc within-group comparisons indicated a reduction in suicidal ideation from T_1 to T_2 in the COS group (difference=3.13; $t_{277.11}$=4.55, two-tailed; P<.001, P_{holm}<.001; d=0.71), with no significant change between T_2 and T_3 (difference=−1.07; $t_{274.21}$=−1.5, two-tailed; P=.13). In the APP group, the change from T_1 to T_2 was not statistically significant (difference=1.2; $t_{272.63}$=1.27, two-tailed; P=.21). The comparison between T_1 and T_3 indicated a reduction, but did not reach the threshold for significance after Holm-Bonferroni correction (difference=2.63; $t_{278.73}$=2.53, two-tailed; P=.01, P_{holm}=.44). As secondary evidence, the waitlist control group showed minimal change in suicidal ideation over time. Neither the comparison between T_1 and T_2 (difference=0.69; $t_{273.74}$=0.80, two-tailed; P=.43) nor the comparison between T_2 and T_3 (difference=−0.16; $t_{276.51}$=−0.16, two-tailed; P=.87) reached statistical significance.

Exploratory Outcomes

Intraclass correlations for the exploratory outcomes also demonstrated notable within-participant dependence, with ICCs ranging from 0.48 to 0.73.

There were significant main effects of time and time × group interaction for all exploratory outcomes (all P<.05), with the exception of perceived available support (as measured by PAS) (P=.75) (Multimedia Appendix 1).

At postintervention, both the COS and APP groups showed significant improvements in multiple domains: reduced perceived stress (PSS), dysfunctional attitudes (DAS-17), and loneliness (UCLA-LS), as well as increased self-efficacy (GSES), perceived social support (PAS), and reasons for living (RFL-YA) (all P<.05). In contrast, the waitlist control group showed no significant change across any of these outcomes between T_1 and T_2 (all P>.05).

Most improvements observed at T_2 were sustained at the 3-month follow-up, with no significant changes from T_2 for any outcomes except DAS-17 and UCLA-LS in the COS group. Specifically, DAS-17 scores increased (P=.02), and loneliness scores decreased (P=.04) from T_2 to T_3; however, neither result remained significant after applying the Holm-Bonferroni correction (P_{holm}=.56 and P_{holm}=.98, respectively).

At T_3, the APP group showed higher GSES and RFL-YA scores than the waitlist control group (P=.006 and P=.03, respectively); however, these differences did not reach statistical significance after correction (P_{holm}=.16 and P_{holm}=.95, respectively). No significant between-group differences were observed for other variables at follow-up.

Discussion

Principal Findings

The primary objective of this RCT study was to evaluate the efficacy of a CBT-based online self-help group program (COS) for individuals experiencing depressive symptoms. Findings indicated that participants in the COS group showed a significant reduction in both depressive symptoms and suicide ideation, with these effects largely maintained at the 3-month follow-up. Notably, 75% (56/75) of those in the COS group exhibited clinically meaningful improvement in depressive symptoms, as determined by the RCI. At baseline, depressive symptoms in the COS group were of moderate severity, but following the intervention, their average symptom levels fell within the mild range. This level of improvement is comparable to outcomes typically achieved through standard, therapist-delivered CBT [57].

Although we expected the COS program to be at least as efficacious as the APP intervention for general CBT-based outcomes and potentially superior for outcomes influenced by peer interaction, the results revealed that the app-based program yielded similar improvement in reducing depressive symptoms, perceived stress, dysfunctional attitudes, and loneliness. This parity in outcomes is likely attributable to the shared therapeutic components across the COS and APP conditions. In both programs, participants engaged in cognitive restructuring exercises, including the identification of cognitive distortions within hypothetical scenarios and the generation of more adaptive, alternative thoughts. These fundamental CBT techniques were consistently applied in both conditions and may have contributed to the comparable levels of improvement observed.

Beyond the shared therapeutic elements, several intervention-related features may help explain why COS and APP yielded largely comparable outcomes. The interpersonal features unique to COS, such as validation, reciprocal feedback, and collaborative cognitive restructuring, may require a longer intervention period or greater group contact for their unique benefits to emerge fully. In this study, the COS program consisted of 7 sessions delivered over 3 weeks, which is considerably shorter than most peer support or group-based depression interventions that typically span 10 to 12 sessions [36]. The relatively brief structure may have limited the development of group processes that could otherwise differentiate COS from APP.

Implementation fidelity is another relevant consideration [58]. The APP condition delivered identical content with consistent pacing for every participant, ensuring high consistency. Conversely, the COS format, although manualized, naturally introduced variability in discussion depth, interaction patterns, and pacing across groups. While this variability may have supported peer engagement, it may also reduce the consistency with which CBT skills were practiced compared to the standardized app.

Despite the similarity in outcomes, the COS group demonstrated a marginally higher proportion of participants achieving

XSL•FO
RenderX

clinically significant change, as indicated by the RCI, compared to the APP group. Furthermore, the analysis for suicide ideation revealed a trend-level time × condition interaction, with a significant reduction emerging only within the COS group. Post hoc comparisons indicated a statistically significant decrease in suicidal ideation from T_1 to T_2 only in the COS condition, even though the intervention did not explicitly target suicidality; in contrast, no significant within-group changes were detected in the APP or waitlist control groups. This finding is particularly salient, given the elevated suicide risk commonly associated with depressive disorders [59,60]. The reduction in suicidal ideation may be attributable to the relational dynamics of the self-help group format. One possible explanation is the supportive and interactive nature of the group setting, which may have provided emotional validation, normalization, and interpersonal connection—factors that have been identified as protective in the context of suicide risk [61]. Previous research has demonstrated that group-based interventions can be effective in reducing suicidal thoughts, with social support and a sense of belonging identified as key mechanisms [62,63]. Recently, peer-led interventions have gained increasing attention in suicide prevention efforts and have been recommended for their potential to meaningfully integrate lived experience into mental health care [64]. However, their adoption in practice remains limited. The COS program may offer a scalable and accessible intervention that not only aligns with these emerging recommendations but also addresses the current gaps in service delivery.

The COS program was specifically designed to integrate structured CBT techniques within a peer-led, self-help group framework. This hybrid model aimed to combine the clinical efficacy of CBT with the emotional resonance and empowerment afforded by lived-experience sharing, which is characteristic of self-help groups [65]. Importantly, since this low-intensity psychological intervention was manualized, brief in duration, and minimally dependent on professional facilitation [66], these features render the COS program particularly amenable to digital dissemination, offering a scalable and resource-efficient option for addressing the treatment gap in depression care.

Limitations

This study has several limitations that must be acknowledged. First, although the COS program was designed for peer-led delivery, trained facilitators with psychological knowledge were present during sessions to ensure participant safety and adherence to the intervention protocol. While this approach helped maintain fidelity and manage potential risks, it may have reduced the ecological validity of a fully peer-led mode. We envision that future implementation will use a stepped-care approach, in which professionals initially lead the groups before transitioning leadership to peers. Consequently, future studies should evaluate the comparative efficacy of the COS program when delivered exclusively by peers versus professionals. This will help establish the viability of this scalable model and its clinical effectiveness in real-world settings. Second, the waitlist control group was recruited independently from the randomized intervention groups. Although efforts were made to match key sociodemographic characteristics, the absence of full randomization may have introduced unmeasured confounding variables, including differences in participants' motivation or expectations, as well as potential time-related confounds that limit the strength of causal inferences. Third, all outcomes were measured using self-report instruments. Although standardized instruments with strong psychometric support were used, the exclusive reliance on self-report measures may introduce reporting bias. Future studies may benefit from incorporating clinician-administered assessments or objective indicators, such as behavioral or physiological measures. Fourth, blinding procedures could not be implemented due to the behavioral nature of the intervention and its peer-supported, online-delivery format. Although outcome assessments were completed directly by participants, thereby limiting the risk of assessor-rated bias, the absence of participant blinding remains a methodological limitation that should be considered when interpreting the results. Fifth, the participants were predominantly female and consisted largely of young adults, with an average age in the late twenties. This consideration is particularly pertinent for outcomes such as suicidal ideation, where gender differences in prevalence and expression are well documented. Although the inclusion criteria permitted participation up to age 55 years, the reliance on online recruitment methods likely contributed to this skew. Consequently, this demographic profile may limit the generalizability of the findings to older adults or populations with different digital usage patterns. Finally, the study did not include long-term follow-up beyond 3 months. Future studies should assess whether the observed improvements are sustained over extended periods, including follow-ups of at least 12 months, and further evaluate the real-world applicability of fully peer-led implementations.

Conclusions

This study contributes valuable evidence supporting digital approaches to depression and suicidality. The continued use of remote, group-based psychological interventions may help address persistent barriers to accessing conventional services. Within this context, the development and evaluation of scalable, evidence-based digital programs remain an important public health objective, particularly for populations with limited access to mental health professionals. Notably, the findings highlight the potential of structured self-help group interventions that can be delivered with minimal professional involvement, offering a practical and accessible option for expanding psychological support.

XSL•FO
RenderX

Acknowledgments

Both the cognitive behavioral therapy–based online self-help program and the cognitive behavioral therapy–based mobile application used in this study were developed by the authors, with support from the National Center for Mental Health (Republic of Korea) for public use. Neither program is commercially owned or distributed.

The authors declare that generative AI tools (Google Gemini and ChatGPT) were used exclusively for proofreading and language editing under full human supervision. All authors confirm that no generative AI tools were used to generate scientific content, analyze data, or interpret study findings.

Funding

This research was supported by a grant of the Korea Mental Health R&D Project, funded by the Ministry of Health & Welfare, Republic of Korea (grant HL19C0048) and by Information and Communications Technology (ICT) and Future Planning for Convergent Research in the Development Program for R&D Convergence over Science and Technology Liberal Arts (NRF-2022M3C1B6080866). The funders had no role in the study design, data collection, analysis, interpretation of data, writing of the manuscript, or the decision to submit the article for publication.

Data Availability

The datasets used in this study are not publicly available because the institutional review board–approved informed consent did not include permission for data sharing. Deidentified data may be made available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: MY, HK, SEL, EJ, JWH
Data curation: MY, HK, SEL, EJ
Formal analysis: MY, HK
Funding acquisition: JWH
Investigation: MY, HK, SEL, EJ, JWH
Methodology: MY, HK
Project administration: MY, HK, SEL, EJ
Supervision: JWH
Validation: MY, HK, JWH
Visualization: MY
Writing – original draft: MY, HK, JWH
Writing – review & editing: MY, HK, SEL, JWH

Conflicts of Interest

None declared.

Editorial Notice

This randomized study was retrospectively registered. The authors noted that the registration was conducted retrospectively due to an administrative oversight during the initial stages of the study. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials because the risk of bias appears low, and the study information registered with the authors' ethics committee matches the trial reporting. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
Exploratory outcome measures.
[DOCX File, 75 KB - mhealth_v14i1e76028_app1.docx]

Checklist 1
CONSORT-eHEALTH (V1.6.1) checklist.
[PDF File, 49471 KB - mhealth_v14i1e76028_app2.pdf]

References

XSL•FO
RenderX

1. Greenberg PE, Fournier AA, Sisitsky T, et al. The economic burden of adults with major depressive disorder in the United States (2010 and 2018). Pharmacoeconomics 2021 Jun;39(6):653-665. [doi: 10.1007/s40273-021-01019-4] [Medline: 33950419]
2. Lundberg J, Cars T, Lööv S, et al. Clinical and societal burden of incident major depressive disorder: a population-wide cohort study in Stockholm. Acta Psychiatr Scand 2022 Jul;146(1):51-63. [doi: 10.1111/acps.13414] [Medline: 35165894]
3. Patel V, Chisholm D, Parikh R, et al. Addressing the burden of mental, neurological, and substance use disorders: key messages from Disease Control Priorities, 3rd edition. Lancet 2016 Apr 16;387(10028):1672-1685. [doi: 10.1016/S0140-6736(15)00390-6] [Medline: 26454360]
4. Ghio L, Gotelli S, Cervetti A, et al. Duration of untreated depression influences clinical outcomes and disability. J Affect Disord 2015 Apr 1;175:224-228. [doi: 10.1016/j.jad.2015.01.014] [Medline: 25658495]
5. Mekonen T, Ford S, Chan GCK, Hides L, Connor JP, Leung J. What is the short-term remission rate for people with untreated depression? A systematic review and meta-analysis. J Affect Disord 2022 Jan 1;296:17-25. [doi: 10.1016/j.jad.2021.09.046] [Medline: 34583099]
6. Whiteford HA, Harris MG, McKeon G, et al. Estimating remission from untreated major depression: a systematic review and meta-analysis. Psychol Med 2013 Aug;43(8):1569-1585. [doi: 10.1017/S0033291712001717] [Medline: 22883473]
7. Thornicroft G, Chatterji S, Evans-Lacko S, et al. Undertreatment of people with major depressive disorder in 21 countries. Br J Psychiatry 2017 Feb;210(2):119-124. [doi: 10.1192/bjp.bp.116.188078] [Medline: 27908899]
8. Chekroud AM, Foster D, Zheutlin AB, et al. Predicting barriers to treatment for depression in a U.S. national sample: a cross-sectional, proof-of-concept study. Psychiatr Serv 2018 Aug 1;69(8):927-934. [doi: 10.1176/appi.ps.201800094] [Medline: 29962307]
9. Shapiro DA, Cavanagh K, Lomas H. Geographic inequity in the availability of cognitive behavioural therapy in England and Wales. Behav Cogn Psychother 2003 Jan;31(2):185-192. [doi: 10.1017/S1352465803002066]
10. Busch AB, Sugarman DE, Horvitz LE, Greenfield SF. Telemedicine for treating mental health and substance use disorders: reflections since the pandemic. Neuropsychopharmacology 2021 May;46(6):1068-1070. [doi: 10.1038/s41386-021-00960-4] [Medline: 33479513]
11. Ma L, Huang C, Tao R, Cui Z, Schluter P. Meta-analytic review of online guided self-help interventions for depressive symptoms among college students. Internet Interv 2021 Sep;25:100427. [doi: 10.1016/j.invent.2021.100427] [Medline: 34401386]
12. Moshe I, Terhorst Y, Philippi P, et al. Digital interventions for the treatment of depression: a meta-analytic review. Psychol Bull 2021 Aug;147(8):749-786. [doi: 10.1037/bul0000334] [Medline: 34898233]
13. Sriati A, Kurniawan K, Senjaya S, et al. The effectiveness of digital-based psychotherapy in overcoming psychological problems in college students during the COVID-19 pandemic: a scoping review. J Holist Nurs 2024 Jun;42(2_suppl):S26-S39. [doi: 10.1177/08980101231162990] [Medline: 37097892]
14. Lainas S. Self-help/mutual aid groups for health and psychosocial problems: key features and their perspectives in the 21st century. Am J Community Psychol 2023 Dec;72(3-4):271-287. [doi: 10.1002/ajcp.12718] [Medline: 37931178]
15. Munn-Giddings CD, Borkman T. Dialogic sharing of lived experience in different self-help/mutual aid groups. Groupwork 2017;27(3):26-46. [doi: 10.1921/gpwk.v27i3.1152]
16. Borkman T. Experiential knowledge: a new concept for the analysis of self-help groups. Social Service Review 1976 Sep;50(3):445-456. [doi: 10.1086/643401]
17. Chinman M, Kloos B, O'Connell M, Davidson L. Service providers' views of psychiatric mutual support groups. J Community Psychol 2002 Jul;30(4):349-366. [doi: 10.1002/jcop.10010]
18. Banbury A, Nancarrow S, Dart J, Gray L, Parkinson L. Telehealth interventions delivering home-based support group videoconferencing: systematic review. J Med Internet Res 2018 Feb 2;20(2):e25. [doi: 10.2196/jmir.8090] [Medline: 29396387]
19. Brown LD, Tang X, Hollman RL. The structure of social exchange in self-help support groups: development of a measure. Am J Community Psychol 2014 Mar;53(1-2):83-95. [doi: 10.1007/s10464-013-9621-3] [Medline: 24398622]
20. Worrall H, Schweizer R, Marks E, Yuan L, Lloyd C, Ramjan R. The effectiveness of support groups: a literature review. Ment Health Soc Incl 2018 Apr 9;22(2):85-93. [doi: 10.1108/MHSI-12-2017-0055]
21. Agudelo-Hernández F, Guapacha-Montoya M, Rojas-Andrade R. Mutual aid groups for loneliness, psychosocial disability, and continuity of care. Community Ment Health J 2024 Apr;60(3):608-619. [doi: 10.1007/s10597-023-01216-9] [Medline: 38194119]
22. Pistrang N, Barker C, Humphreys K. Mutual help groups for mental health problems: a review of effectiveness studies. Am J Community Psychol 2008 Sep;42(1-2):110-121. [doi: 10.1007/s10464-008-9181-0] [Medline: 18679792]
23. Lloyd-Evans B, Mayo-Wilson E, Harrison B, et al. A systematic review and meta-analysis of randomised controlled trials of peer support for people with severe mental illness. BMC Psychiatry 2014 Feb 14;14(1):1-12. [doi: 10.1186/1471-244X-14-39] [Medline: 24528545]
24. Bryan AEB, Arkowitz H. Meta-analysis of the effects of peer-administered psychosocial interventions on symptoms of depression. Am J Community Psychol 2015 Jun;55(3-4):455-471. [doi: 10.1007/s10464-015-9718-y] [Medline: 25861883]

XSL•FO
RenderX

25. Gram L, Fitchett A, Ashraf A, Daruwalla N, Osrin D. Promoting women's and children's health through community groups in low-income and middle-income countries: a mixed-methods systematic review of mechanisms, enablers and barriers. BMJ Glob Health 2019;4(6):e001972. [doi: 10.1136/bmjgh-2019-001972] [Medline: 31908874]
26. van de Ven K, Deane FP, Kelly PJ, Larance B, Beck AK. Online SMART Recovery mutual-help groups: a comparison of the characteristics and experiences of men living in rural and urban regions of Australia. Rural Remote Health 2024 Sep;24(3):1-8. [doi: 10.22605/RRH8861] [Medline: 39307541]
27. Taylor F, Gutteridge R, Willis C. Peer support for CKD patients and carers: overcoming barriers and facilitating access. Health Expect 2016 Jun;19(3):617-630. [doi: 10.1111/hex.12348] [Medline: 25649115]
28. Fortuna KL, Naslund JA, LaCroix JM, et al. Digital peer support mental health interventions for people with a lived experience of a serious mental illness: systematic review. JMIR Ment Health 2020 Apr 3;7(4):e16460. [doi: 10.2196/16460] [Medline: 32243256]
29. Beck AK, Waks S, Argent A, et al. The benefits and challenges of virtual SMART recovery mutual-help groups: participant and facilitator perspectives. Int J Drug Policy 2023 Oct;120:104174. [doi: 10.1016/j.drugpo.2023.104174] [Medline: 37659377]
30. Ali K, Farrer L, Gulliver A, Griffiths KM. Online peer-to-peer support for young people with mental health problems: a systematic review. JMIR Ment Health 2015;2(2):e19. [doi: 10.2196/mental.4418] [Medline: 26543923]
31. Fortuna KL, Solomon P, Rivera J. An update of peer support/peer provided services underlying processes, benefits, and critical ingredients. Psychiatr Q 2022 Jun;93(2):571-586. [doi: 10.1007/s11126-022-09971-w] [Medline: 35179660]
32. Penn PE, Brooks AJ. Five years, twelve steps, and REBT in the treatment of dual diagnosis. J Ration Emot Cogn Behav Ther 2000 Dec;18(4):197-208. [doi: 10.1023/A:1007883021936]
33. Hur JW, Kim B, Park D, Choi SW. A scenario-based cognitive behavioral therapy mobile app to reduce dysfunctional beliefs in individuals with depression: a randomized controlled trial. Telemed J E Health 2018 Sep;24(9):710-716. [doi: 10.1089/tmj.2017.0214] [Medline: 29323626]
34. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x]
35. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. J Pharmacol Pharmacother 2010 Jul;1(2):100-107. [doi: 10.4103/0976-500X.72352] [Medline: 21350618]
36. Pfeiffer PN, Heisler M, Piette JD, Rogers MAM, Valenstein M. Efficacy of peer support interventions for depression: a meta-analysis. Gen Hosp Psychiatry 2011;33(1):29-36. [doi: 10.1016/j.genhosppsych.2010.10.002] [Medline: 21353125]
37. Kambeitz-Ilankovic L, Rzayeva U, Völkel L, et al. A systematic review of digital and face-to-face cognitive behavioral therapy for depression. NPJ Digit Med 2022 Sep 15;5(1):144. [doi: 10.1038/s41746-022-00677-8] [Medline: 36109583]
38. Todak todak app optimization and self-help group guidebook effectiveness study. National Mental Health Center. 2020. URL: https://www.ncmh.go.kr/research/board/boardView.do?no=3482&fno=84&bn=SNMH_COMMON_BOARD&menu_cd=03_02_00_01&bno=&pageIndex=1&search_item=&search_content= [accessed 2026-06-11]
39. Dube P, Kurt K, Bair MJ, Theobald D, Williams LS. The P4 screener: evaluation of a brief measure for assessing potential suicide risk in 2 randomized effectiveness trials of primary care and oncology patients. Prim Care Companion J Clin Psychiatry 2010;12(6):e1-e8. [doi: 10.4088/PCC.10m00978blu] [Medline: 21494337]
40. Beck AT, Steer RA, Brown GK. Manual for the Beck Depression Inventory-II: Psychological Corporation; 1996. [doi: 10.1037/t00742-000]
41. Beck AT, Steer RA. Manual for the Beck Scale for Suicide Ideation: Psychological Corporation; 1991.
42. Weissman AN, Beck AT. Development and validation of the dysfunctional attitude scale: a preliminary investigation. Presented at: Meeting of the American Educational Research Association; Mar 27-31, 1978 URL: https://eric.ed.gov/?id=ED167619 [accessed 2026-06-22]
43. de Graaf LE, Roelofs J, Huibers MJH. Measuring dysfunctional attitudes in the general population: the Dysfunctional Attitude Scale (Form A) Revised. Cognit Ther Res 2009 Aug;33(4):345-355. [doi: 10.1007/s10608-009-9229-y] [Medline: 19623267]
44. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [doi: 10.2307/2136404] [Medline: 6668417]
45. Russell DW. UCLA Loneliness Scale (Version 3): reliability, validity, and factor structure. J Pers Assess 1996 Feb;66(1):20-40. [doi: 10.1207/s15327752jpa6601_2] [Medline: 8576833]
46. Russell D, Peplau LA, Cutrona CE. The revised UCLA Loneliness Scale: concurrent and discriminant validity evidence. J Pers Soc Psychol 1980;39(3):472-480. [doi: 10.1037/0022-3514.39.3.472]
47. Schulz U, Schwarzer R. Soziale unterstützung bei der krankheitsbewältigung: die Berliner Social Support Skalen (BSSS). Diagnostica 2003 Apr;49(2):73-82. [doi: 10.1026//0012-1924.49.2.73]
48. Gutierrez PM, Osman A, Barrios FX, Kopper BA, Baker MT, Haraburda CM. Development of the Reasons for Living Inventory for Young Adults. J Clin Psychol 2002 Apr;58(4):339-357. [doi: 10.1002/jclp.1147] [Medline: 11920689]
49. Sherer M, Maddux JE, Mercandante B, Prentice-Dunn S, Jacobs B, Rogers RW. The self-efficacy scale: construction and validation. Psychol Rep 1982 Oct;51(2):663-671. [doi: 10.2466/pr0.1982.51.2.663]

XSL•FO
RenderX

50. Chakraborty H, Gu H. A Mixed Model Approach for Intent-to-Treat Analysis in Longitudinal Clinical Trials With Missing Values: RTI Press; 2009. [doi: 10.3768/rtipress.2009.mr.0009.0903]
51. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition: Lawrence Erlbaum Associates; 1988.
52. Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychol Methods 2002 Mar;7(1):105-125. [doi: 10.1037/1082-989x.7.1.105] [Medline: 11928886]
53. Goulet-Pelletier JC, Cousineau D. A review of effect sizes and their confidence intervals, Part I: The Cohen's d family. Quant Meth Psychol 2018;14(4):242-265. [doi: 10.20982/tqmp.14.4.p242]
54. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991;59(1):12-19. [doi: 10.1037/0022-006X.59.1.12]
55. Blampied NM. Reliable change and the reliable change index: still useful after all these years? Cogn Behav Ther 2022;15:e50. [doi: 10.1017/S1754470X22000484]
56. Evans C. Compute RCI from SD and reliability. PSYCTC.org. 2023. URL: https://shiny.psyctc.org/apps/RCI1/ [accessed 2026-06-22]
57. Malkomsen A, Wilberg T, Bull-Hansen B, et al. Comparative effectiveness of short-term psychodynamic psychotherapy and cognitive behavioral therapy for major depression in psychiatric outpatient clinics: a randomized controlled trial. BMC Psychiatry 2025 Feb 11;25(1):113. [doi: 10.1186/s12888-025-06544-6] [Medline: 39934737]
58. Durlak JA, DuPre EP. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. Am J Community Psychol 2008 Jun;41(3-4):327-350. [doi: 10.1007/s10464-008-9165-0] [Medline: 18322790]
59. Moitra M, Santomauro D, Degenhardt L, et al. Estimating the risk of suicide associated with mental disorders: a systematic review and meta-regression analysis. J Psychiatr Res 2021 May;137:242-249. [doi: 10.1016/j.jpsychires.2021.02.053] [Medline: 33714076]
60. Arnone D, Karmegam SR, Östlundh L, et al. Risk of suicidal behavior in patients with major depression and bipolar disorder – a systematic review and meta-analysis of registry-based studies. Neurosci Biobehav Rev 2024 Apr;159:105594. [doi: 10.1016/j.neubiorev.2024.105594]
61. Hussain-Shamsy N, Shah A, Wasserman L, et al. Virtual connection and real community: the qualitative experience of participating in a videoconferencing-based psychotherapy group for postpartum depression and anxiety. BMC Health Serv Res 2024 Nov 1;24(1):1329. [doi: 10.1186/s12913-024-11753-y] [Medline: 39482679]
62. Oyama H, Watanabe N, Ono Y, et al. Community-based suicide prevention through group activity for the elderly successfully reduced the high suicide rate for females. Psychiatry Clin Neurosci 2005 Jun;59(3):337-344. [doi: 10.1111/j.1440-1819.2005.01379.x] [Medline: 15896228]
63. Wyman PA, Pisani AR, Brown CH, et al. Effect of the Wingman-Connect Upstream Suicide Prevention Program for Air Force personnel in training: a cluster randomized clinical trial. JAMA Netw Open 2020 Oct 1;3(10):e2022532. [doi: 10.1001/jamanetworkopen.2020.22532] [Medline: 33084901]
64. Moran P, Chandler A, Dudgeon P, et al. The Lancet Commission on self-harm. Lancet 2024 Oct 12;404(10461):1445-1492. [doi: 10.1016/S0140-6736(24)01121-8] [Medline: 39395434]
65. Seebohm P, Chaudhary S, Boyce M, Elkan R, Avis M, Munn-Giddings C. The contribution of self-help/mutual aid groups to mental well-being. Health Soc Care Community 2013 Jul;21(4):391-401. [doi: 10.1111/hsc.12021] [Medline: 23445336]
66. Bennett-Levy J, Richards DA, Farrand P, et al. Low intensity CBT interventions: a revolution in mental health care. In: Bennett-Levy J, Richards DA, Farrand P, editors. Oxford Guide to Low Intensity CBT Interventions: Oxford University Press; 2010:3-18. [doi: 10.1093/med:psych/9780199590117.003.0001]

Abbreviations

APP: cognitive behavioral therapy–based mobile application group
BDI-II: Beck Depression Inventory-II
BSS: Beck Scale for Suicide Ideation
BSSS: Berlin Social Support Scale
CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
COS: cognitive behavioral therapy–based online self-help group
DAS-17: Dysfunctional Attitude Scale (Form A) Revised
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th Edition
GSES: General Self-Efficacy Scale
ICC: intraclass correlation coefficient
LMM: linear mixed model
MAR: missing at random
PAS: Perceived Available Support
PHQ-9: Patient Health Questionnaire-9

XSL•FO
RenderX

PSS: Perceived Stress Scale
RCI: reliable change index
RCT: randomized controlled trial
RFL-YA: Reasons for Living Inventory for Young Adults
SCID-5-CV: Structured Clinical Interview for DSM-5 – Clinician Version
UCLA: University of California, Los Angeles
UCLA-LS: UCLA Loneliness Scale version 3

Edited by L Buis; submitted 16.Apr.2025; peer-reviewed by A Chen, D Jung, S Choi; revised version received 07.Dec.2025; accepted 27.Apr.2026; published 03.Jul.2026.

Please cite as:
Yim M, Kim H, Lee SE, Jo E, Hur JW
Efficacy of a Cognitive Behavioral Therapy–Based Online Self-Help Group for Depression and Suicide Ideation: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e76028
URL: https://mhealth.jmir.org/2026/1/e76028
doi:10.2196/76028

© Minkyung Yim, Haeun Kim, Soo-Eun Lee, Eunjin Jo, Ji-Won Hur. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 3.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Alleviating Nurse Burnout With an Artificial Intelligence–Selected Mobile Cognitive Behavioral Therapy–Based Intervention: Mixed Methods Randomized Controlled Trial

Yeongeun Kim[1,2], MSN; Chiyoung Cha[2,3], PhD; Gumhee Baek[2], PhD

[1]Dongnam Institute of Radiological and Medical Sciences, Busan, Republic of Korea

[2]College of Nursing, Ewha Womans University, Helen Hall 112, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul, Republic of Korea

[3]Ewha Research Institute of Nursing Science, Ewha Womans University, Seoul, Republic of Korea

Corresponding Author:
Gumhee Baek, PhD
College of Nursing, Ewha Womans University, Helen Hall 112, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul, Republic of Korea

Abstract

Background: Nurse burnout is a pervasive global problem. Cognitive behavioral therapy (CBT) has been shown to reduce burnout; however, most digital CBT programs use standardized approaches that overlook individual differences in burnout profiles. With advances in artificial intelligence (AI), algorithm-based recommendation systems now enable personalized intervention delivery by matching specific CBT modules to users.

Objective: This study aimed to test the effects of an AI-selected mobile CBT-based intervention on nurse burnout and to describe participants' experiences with the intervention. Specifically, it evaluated whether an AI-selected CBT-based intervention differentially reduced burnout subdomains compared with an information-only control group and explored how nurses perceived and engaged with the AI-selected program.

Methods: This study adopted a mixed methods design, integrating a 2-group randomized controlled trial and qualitative content analysis exploring participants' experiences. For this randomized controlled trial, a total of 125 nurses were enrolled and randomly assigned to either the experimental group (n=62) or the control group (n=63) between October 2024 and December 2024. The experimental group received an AI-selected mobile CBT-based intervention, in which an AI algorithm assigned CBT modules based on participants' burnout profiles (client-related, personal, and work-related), job stress, and coping characteristics. The control group received information related to burnout management. Primary outcomes, client-related, personal, and work-related burnout, were assessed at baseline, 2 weeks, and 4 weeks. Secondary outcomes, including coping strategies, job stress, and stress response, were assessed at baseline and 4 weeks. Between-group differences in burnout over time were examined using repeated measures analysis of variance, with adjustment for job stress and stress response. Within-group changes and postintervention group differences were analyzed using *t* tests. Open-ended survey responses and follow-up interviews (n=5 in the experimental group) were analyzed using thematic content analysis.

Results: Follow-up completion rates were 84.6% (137/162) at both 2 and 4 weeks. The experimental group showed a greater reduction in client-related ($F_{1,121}$=7.548; P=.007), personal ($F_{1,121}$=6.533; P=.01), and work-related burnout ($F_{1,121}$=38.194; P<.001) than the control group, reflecting more pronounced within-group improvements over time. No significant between-group differences were observed for coping strategies, job stress, or stress response. Qualitative findings suggested that some participants were receptive to the AI-selected CBT-based intervention and reported increased self-awareness and reflective engagement with coping strategies that they might not have selected independently.

Conclusions: The findings suggest that participants were receptive to AI-selected CBT-based interventions, suggesting the potential of such interventions as a supportive approach for alleviating nurse burnout. Future research should explore the sustainability of these effects and optimize the intervention duration to enhance engagement and impact.

Trial Registration: Clinical Research Information Service (CRIS) KCT0009853; https://tinyurl.com/3b5pesd7

(JMIR Mhealth Uhealth 2026;14:e85986) doi:10.2196/85986

KEYWORDS

artificial intelligence; burnout; cognitive behavioral therapy; mHealth; mobile health; nurse; randomized controlled trial

XSL•FO
RenderX

Introduction

Nurses play an essential role in delivering health care services and represent nearly half of the global human resources in the health sector [1]. However, many nurses are increasingly vulnerable to burnout due to prolonged exposure to physical and psychological stressors in clinical settings [2]. Several meta-analyses have estimated that the prevalence of burnout among nurses ranges from 11% to 56% [3-5], and in Korea, 55.9% of frontline nurses experience burnout [6].

Nurse burnout is multifaceted and often categorized into 3 subdomains: client-related burnout (exhaustion related to patient care), personal burnout (emotional and physical exhaustion), and work-related burnout (stress stemming from job demands) [7]. Given the ongoing nursing shortage [8] and rising health care demands [9], accessible and effective interventions are required to address burnout. These interventions should consider not only general experiences of burnout but also its distinct subdomains, which may vary according to individual roles and work contexts.

Cognitive behavioral therapy (CBT) is a structured, problem-focused intervention that aims to modify maladaptive and irrational cognitions underlying emotional and behavioral issues [10]. Beyond traditional protocol-based CBT, an expanding body of research has applied CBT-based or CBT-informed psychological interventions, including mindfulness-based approaches [11], laughter therapy [12], storytelling [13], and reflective writing [14], to address burnout-related symptoms. CBT has been widely recognized as an effective intervention for reducing burnout symptoms among nurses by helping them restructure their dysfunctional thoughts and behaviors across various clinical settings. Meta-analytic evidence suggests that CBT is particularly effective in alleviating emotional exhaustion [15] and depersonalization [16], which are 2 core subdomains of burnout. These findings highlight the potential of CBT-based interventions to effectively address key aspects of nurse burnout.

With the rapid advancement of mobile technology, mobile health (mHealth) interventions have become increasingly common in supporting mental health [17]. These platforms offer scalable, flexible, and cost-effective alternatives to traditional face-to-face therapy, making them particularly useful for health care professionals, such as nurses, who often face temporal and spatial barriers to accessing psychological support [18]. Application-based CBT programs, such as Wysa [19] and Woebot [20], have shown effectiveness in reducing symptoms of depression and stress. Artificial intelligence (AI)–based recommendation systems are increasingly being integrated into mHealth platforms to enable the delivery of personalized interventions [21]. Originally developed and widely adopted in industries such as retail and entertainment, collaborative filtering and content-based recommendation algorithms are currently being utilized in health care settings to tailor interventions based on user profiles, symptom severity, and behavioral patterns [22]. These intelligent systems have the potential to enhance user engagement and therapeutic outcomes by offering interventions that optimally match individual needs [23].

Despite growing evidence supporting the effectiveness of CBT-based interventions in reducing burnout among nurses, most studies have focused on evaluating the effects of a single standardized program [16]. These approaches often adopt a one-size-fits-all model, overlooking the multidimensional nature of burnout and the diverse needs of individual users. Few interventions have addressed burnout subdomains as distinct targets for tailored support. Moreover, there is a notable lack of research that applies user-based recommendation systems to suggest interventions based on nurses' individual characteristics, such as their burnout profiles or professional backgrounds. This gap highlights the need for an adaptive and tailored system that can enhance the effectiveness and relevance of digital mental health interventions in the nursing workforce.

This study evaluated the effectiveness of an AI-selected mobile CBT-based intervention designed to reduce burnout among nurses. The intervention used a user-based recommendation system to enhance its impact by suggesting 1 of 4 CBT-based modules tailored to the user's burnout subdomain, as well as demographic and work-related factors. Previously, we conducted a 3-group randomized trial using this system [24], which demonstrated significant improvements in some burnout subdomains.

Since then, the algorithm has been further optimized, but it remains unclear how participants perceive and engage with programs selected by AI. To address both issues, we designed this study to evaluate the intervention using quantitative and qualitative approaches, thereby providing a more comprehensive understanding of its effectiveness and user acceptance.

Methods

Study Design

We adopted a mixed methods design, integrating a 2-group randomized controlled trial (single-anonymized) and qualitative content analysis exploring participants' experiences. The study was conducted in accordance with the CONSORT (Consolidated Standards of Reporting Trials) guidelines for reporting randomized controlled trials [25] (Checklist 1). In addition, the qualitative component of this study was reported in accordance with the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist [26].

Randomization

Randomization was conducted using an online tool (randomizer.org, a service provided by the Social Psychology Network), ensuring an equal 1:1 allocation between the experimental and control groups. Participants were not informed about the randomization before enrollment to reduce selection bias and preserve objectivity during the intervention process.

Participants and Sample Size

Participants were recruited through 2 large South Korean online nurse communities (1500 registered nurse-verified members). Recruitment notices containing the study purpose, procedures, and inclusion criteria were posted as online announcements. Nurses who wished to participate accessed a Google Forms survey link, reviewed the study information, and provided

XSL•FO
RenderX

electronic informed consent. Inclusion criteria were nurses who had completed preceptorship and worked independently for at least 1 month; those not providing direct care or with prior exposure to mobile burnout interventions were excluded. Additionally, participants were allowed to withdraw from the study after enrollment if (1) personal- or health-related circumstances made continued participation impossible, (2) they chose not to complete the study, or (3) they had difficulty participating in the study due to safety or ethical reasons.

We estimated the required sample size using G*Power (version 3.1.9.7) [27], based on a previous study [28], with an effect size of 0.13, α=.05, power=0.95, and 3 repeated measures, resulting in 130 participants. Allowing for a 20% attrition rate [29], the recruitment target was set at 162 participants (81 per group). To ensure data quality, baseline variables were screened for univariate outliers. Observations more than 1.5 to 3 IQRs from the median were removed, resulting in the exclusion of 12 cases (experimental=4; control=8). Of the 162 eligible participants randomized, 137 (84.6%) completed both the 2-week (posttest 1) and 4-week (posttest 2) follow-up assessments. For the final quantitative analyses, 125 participants (experimental=62, control=63) were included after the data-cleaning process for missing values and outliers, indicating a 77.2% retention rate for a fully remote intervention.

Interventions

Experimental Group

The intervention designed by the research team was delivered via a mobile app called Nurse Healing Space. This app comprises 4 distinct modules that target burnout reduction: (1) mindfulness meditation, (2) storytelling and reflective writing, (3) laughter therapy, and (4) acceptance and commitment therapy (Table 1). The experimental group received AI-selected CBT-based intervention modules at 2-week intervals. Over the 4-week intervention period, the AI algorithm sequentially selected 2 modules from the 4 available CBT-based interventions. Participants engaged in 3 sessions per week, with each session lasting approximately 10 to 15 minutes. Animated female characters appeared in the videos to guide participants through the activities. For the mindfulness meditation, users followed guided instructions regarding body movements, breathing techniques, and meditative practices. Storytelling and reflective writing involved watching narratives shared by fellow nurses, followed by reflective writing about their personal experiences. In laughter therapy, participants were led through songs and dance routines designed to elicit positive affect and laughter. The acceptance and commitment therapy component focused on emotional acceptance and thought regulation, as modeled in the videos. All program content was developed with input from domain experts to ensure both clinical relevance and psychological effectiveness.

Table . Program components of intervention.

Week and session	Programs			
	Mindfulness meditation	Storytelling and reflective writing	Laughter therapy	Acceptance and commitment therapy
Week 1				
Session 1	Body scan (upper body), self-compassion meditation	Work is piling up	Look like my mom, bee dance	Hello my heart (introduction)
Session 2	Body scan (lower body), yoga meditation	I am a person too	Don’t make a wry face, bee dance	Thinking is just thinking (cognitive diffusion)
Session 3	Sitting up straight meditation, relaxation meditation	There aren’t enough hours in the day	When I go to Los Angeles, creative movement	This is the moment (being present in the moment)
Week 2				
Session 4	Body scan (upper body), yoga meditation	Am I an emotional outlet?	Look like my mom, open chest 1, 2, 3	Look back at me (self as context)
Session 5	Body scan (lower body), sitting up straight meditation	The working environment is too poor	Don’t make a wry face, free movement	Worthy life (the best 3 moments of my life)
Session 6	Self-compassion meditation, relaxation meditation	There is a shortage of nurses	When I go to Los Angeles, creative movement	Let’s do it together (committed action and willingness)

The algorithm was developed to provide personalized burnout interventions by identifying the most similar prior participant using both profile similarity scores and burnout patterns. The development, operation, optimization, and intervention content of the AI algorithm are discussed in greater detail in a separate study [30].

After completing Program 1, participants’ burnout levels were reassessed. If the most elevated burnout dimension showed a reduction of 50 points or more, the data were considered valid and incorporated into the algorithm. Otherwise, the data were considered invalid and excluded from AI model training, although the participants continued to receive the subsequent intervention modules without restriction. As data accumulated,

XSL•FO
RenderX

the AI continually updated its artificial neural network through iterative learning. Based on the reassessed burnout dimension scores after Program 1, the AI algorithm recommended a second 2-week program (Program 2) using the same algorithm. The variables were reassessed after Program 2. Following this, the AI algorithm updated its artificial neural network upon recognizing valid data. Participants in the experimental group completed a pretest survey (107 items: demographic and work-related characteristics and study variables), followed by Program 1 for 2 weeks, posttest 1 (19 items: burnout), Program 2 for 2 weeks, and posttest 2 (97 items: study variables).

Control Group

Participants in the control group were provided with online informational materials covering all 4 intervention programs (mindfulness meditation, storytelling and reflective writing, laughter therapy, and acceptance and commitment therapy) through a blog platform. We sent the URL to participants after they had completed the pretest survey. The control group also completed the same sequence of assessments as the experimental group: a pretest survey, posttest 1 at week 2, and posttest 2 at week 4. After completing the final assessment, participants in the control group were given access to *Nurse Healing Space*.

Measurement

Demographic and Work-Related Characteristics

Data included age, sex, marital status, job title, hospital size, clinical experience, department, overtime, shift type, and turnover intention (0 - 10 scale).

Primary Outcome

We assessed burnout using the Korean-adapted Copenhagen Burnout Inventory [7,31]. This 19-item instrument measures client-related, personal, and work-related burnout on a 5-point Likert scale. Each item is transformed to a 0 to 100 score, and subscale scores are calculated as the mean of the corresponding items, with higher scores indicating greater burnout. The possible score range for each subscale is 0 to 100. Cronbach α in this study was 0.936.

Secondary Outcomes

We measured coping using the Korean version of the Coping Strategy Indicator [32,33]. The 33-item scale includes social support-seeking, problem-focused, and avoidance-focused coping, rated on a 3-point Likert scale. Subscale scores are calculated by summing the corresponding item scores (possible range: 11 - 33), with higher scores indicating greater use of each coping strategy. Cronbach α in this study was 0.888. We measured job stress with a 23-item scale developed for Korean nurses [34,35]. It covers 6 subdomains: workload overload, role conflict as a professional, lack of professional knowledge and skills, interpersonal relationship problems, inappropriate treatment and compensation, and night shift. These are rated on a 5-point Likert scale. Item scores range from 1 to 5, and the overall job stress score is calculated as the mean of the item scores, with higher scores indicating greater job stress. The possible score range is 1 to 5. Cronbach α in this study was 0.881. We measured stress response with the 22-item Stress Response Inventory-Modified Form [36,37], covering somatization, depression, and anger, rated on a 5-point Likert scale (0 - 4). Total scores range from 0 to 88, with higher scores indicating greater stress response. Cronbach α in this study was 0.936.

Data Collection

Data were collected between October 8 and December 31, 2024. Quantitative data were assessed at baseline, 2 weeks, and 4 weeks. Participants advanced to the next program only after completing the previous one. Although each module was designed to last 2 weeks, actual usage varied: Program 1 ranged from 14 to 20 days (mean 15.80, SD 1.30 d) and Program 2 ranged from 14 to 18 days (mean 15.55, SD 1.0 d). The integrated management platform enabled the research team to monitor participants' study progress and completion status at each stage of the study. To support continued participation, weekly reminder messages were sent during the intervention period, and participants received compensation after completion of each survey assessment point.

Qualitative data were collected to explore participants' experiences with the AI-recommended interventions. As part of the survey, all experimental group participants completed a short open-ended questionnaire. Following the completion of the intervention, participants in the experimental group were contacted via text messages to assess their willingness to participate in one-on-one follow-up Zoom interviews, and 5 agreed to participate. Each interview lasted approximately 30 minutes (Textbox 1). The semistructured interview guide was collaboratively developed by the research team and reviewed by an experienced qualitative researcher. The follow-up interviews were conducted by the first author, a female clinical nurse and doctoral student with approximately 10 years of inpatient experience and prior qualitative research experience. No prior relationship was established between the interviewer and participants before the study, and participants were informed of the interviewer's role as a researcher and the purpose of the study prior to participation. To minimize potential interviewer bias, the interviewer maintained reflexive memos during data collection, adhered to the semistructured interview guide, and engaged in peer debriefing with an experienced qualitative researcher. These procedures helped enhance the credibility and trustworthiness of the qualitative findings. With participants' consent, all Zoom interviews were audio-recorded for transcription and analysis purposes. No video recordings were made, and participants were informed that they could freely disable their cameras during the interview.

XSL•FO
RenderX

Textbox 1. Interview questions.

1. If you were to describe the program you completed to a friend, what would you say?
2. In what ways did the program help reduce your burnout?
3. What positive experiences did you have while participating in the recommended program?
4. What negative experiences did you have while participating in the recommended program?
5. What was the most significant change you noticed before and after completing the program?
6. What are your thoughts on the use of artificial intelligence (AI) in health programs?
7. During the intervention period, were you aware that the program you received had been selected by an AI algorithm?
8. How do you feel about receiving and following a program recommended by an AI algorithm?
9. How do you think your experience would differ if you chose a program yourself compared with following one recommended by an AI algorithm?

Data Analysis

We analyzed data using SPSS Statistics (version 27; IBM Corp). Descriptive statistics summarized participants' characteristics and study variables. We tested baseline homogeneity between groups using chi-square tests and *t* tests. Group differences in burnout over time were examined using repeated-measures ANOVA, with job stress and stress response as covariates due to their significant correlations with burnout. We assessed within-group changes in coping strategies, job stress, and stress response using independent *t* tests. Within-group changes in burnout from baseline to posttest 2 were further examined using paired-sample *t* tests, and effect sizes were calculated using Cohen *dz*. An intention-to-treat (ITT) analysis was not conducted. This study used a sequential AI-based intervention in which outcome data at each assessment point were required to inform subsequent algorithm-selected module recommendations and to ensure the validity of adaptive intervention delivery. Accordingly, the primary quantitative analyses were conducted using participants with complete outcome data across assessment points. The absence of an ITT analysis is acknowledged as a methodological limitation of this study.

Qualitative data were analyzed using content analysis to identify participants' experiences with the AI-based CBT intervention. For this qualitative component, an inductive content analysis approach was used. The first author repeatedly read the interview transcripts and open-ended responses to become familiar with the data and generated initial open codes. These codes were then organized into broader categories through an iterative comparison process. To enhance the credibility of the findings, a peer debriefing procedure was conducted, in which an experienced qualitative researcher reviewed the coding structure and provided feedback. Throughout the analysis, reflexive memos were maintained to document analytic decisions and reduce potential researcher bias. These steps helped strengthen the dependability and confirmability of the qualitative findings, which served to complement and contextualize the quantitative results.

Ethical Considerations

This study was approved by the institutional review board of the principal investigator's institution (institutional review board number ewha-202407-0025-01) and registered in the Clinical Research Information Service (KCT0009853; October 16, 2024). All participants received detailed information about the study's purpose, procedures, risks, and benefits through an online study information page, and electronic informed consent was obtained prior to enrollment. Participants were informed of their right to withdraw from the study at any time without penalty. To ensure privacy and confidentiality, all collected data were deidentified and stored in encrypted, password-protected files. Contact information used for randomization and compensation delivery was stored separately from outcome data, and only authorized research personnel had access to the deidentified dataset. All participants received a token of appreciation valued at KRW 5000 (approximately US $4) for each survey completed (baseline, posttest 1, and posttest 2). In addition, experimental group participants who took part in the optional follow-up interview received an additional KRW 20,000 (approximately US $15) gift card as compensation.

Results

Homogeneity of Demographic Characteristics and Study Variables Between Groups

Baseline comparisons indicated no significant differences in demographic characteristics between groups (Table 2). However, job stress (t_{123}=3.80, P<.001) and stress response (t_{123}=3.77, P<.001) differed significantly and were therefore included as covariates in subsequent analyses to adjust for potential confounding effects (Figure 1).

XSL•FO
RenderX

Table . Homogeneity of demographic and study variables between the groups (N=125).

Variables and categories	Exp.[a] (n=62)	Cont.[b] (n=63)	Chi-square (*df*) or *t* (*df*)	*P* value
Age (y), mean (SD)	28.18 (5.49)	27.81 (3.76)	0.44 (124)	.66
Age (y), n (%)			5.36 (1)	.15
<30	43 (69.4)	41 (65.1)		
≥30	19 (30.6)	22 (34.9)		
Sex, n (%)			1.31 (1)	.44
Female	60 (96.8)	58 (92.1)		
Male	2 (3.2)	5 (7.9)		
Marital status, n (%)			3.01 (1)	.22
Single	54 (87.1)	58 (92.1)		
Married	8 (12.9)	5 (7.9)		
Job position, n (%)			1.53 (1)	.47
Staff nurse	55 (88.7)	59 (93.7)		
Charge/head nurse	7 (11.3)	4 (6.3)		
Hospital size (beds), n (%)			9.29 (2)	.06
≤100	7 (11.3)	9 (14.2)		
101-500	36 (58.1)	26 (41.4)		
≥500	19 (30.6)	28 (44.4)		
Clinical experience (y), mean (SD)	4.21 (4.43)	3.71 (3.03)	0.73 (124)	.47
Clinical experience (y), n (%)			1.50 (1)	.22
≤5	40 (64.5)	47 (74.6)		
>5	22 (35.5)	16 (25.4)		
Working department, n (%)			7.26 (2)	.30
General ward	43 (69.4)	37 (58.7)		
ICU[c]/ER[d]	14 (22.6)	17 (27.0)		
Others[e]	5 (8.0)	9 (14.3)		
Overtime on average during the past month (h), n (%)			1.35 (4)	.93
≤0.5	29 (46.7)	30 (47.6)		
≤1	17 (27.4)	19 (30.2)		
≤1.5	8 (12.9)	5 (7.9)		
≤2	4 (6.5)	5 (7.9)		
>2	4 (6.5)	4 (6.3)		
Type of shift, n (%)			5.83 (2)	.12
8-h shift	53 (85.5)	49 (77.8)		
12-h shift	0 (0.0)	1 (1.6)		
Fixed	9 (14.5)	13 (20.6)		
Turnover intention, mean (SD)	5.61 (2.29)	5.37 (2.29)	0.61 (124)	.55
Burnout, mean (SD)				
Total score	63.01 (17.57)	61.49 (18.54)	0.47 (124)	.64
Client-related burnout	59.88 (24.17)	53.90 (22.44)	1.43 (124)	.15

XSL•FO
RenderX

Variables and categories	Exp.[a] (n=62)	Cont.[b] (n=63)	Chi-square (*df*) or *t* (*df*)	*P* value
Personal burnout	71.10 (18.57)	66.73 (19.72)	1.28 (124)	.21
Work-related burnout	58.76 (16.56)	63.49 (19.25)	−1.47 (124)	.14
Coping strategy, mean (SD)	2.08 (0.28)	2.10 (0.26)	−0.35 (124)	.72
Job stress, mean (SD)	3.95 (0.44)	3.65 (0.44)	3.80 (124)	<.001
Stress response, mean (SD)	2.05 (0.66)	1.55 (0.80)	3.77 (124)	<.001

[a]Exp.: experimental group.

[b]Cont.: control group.

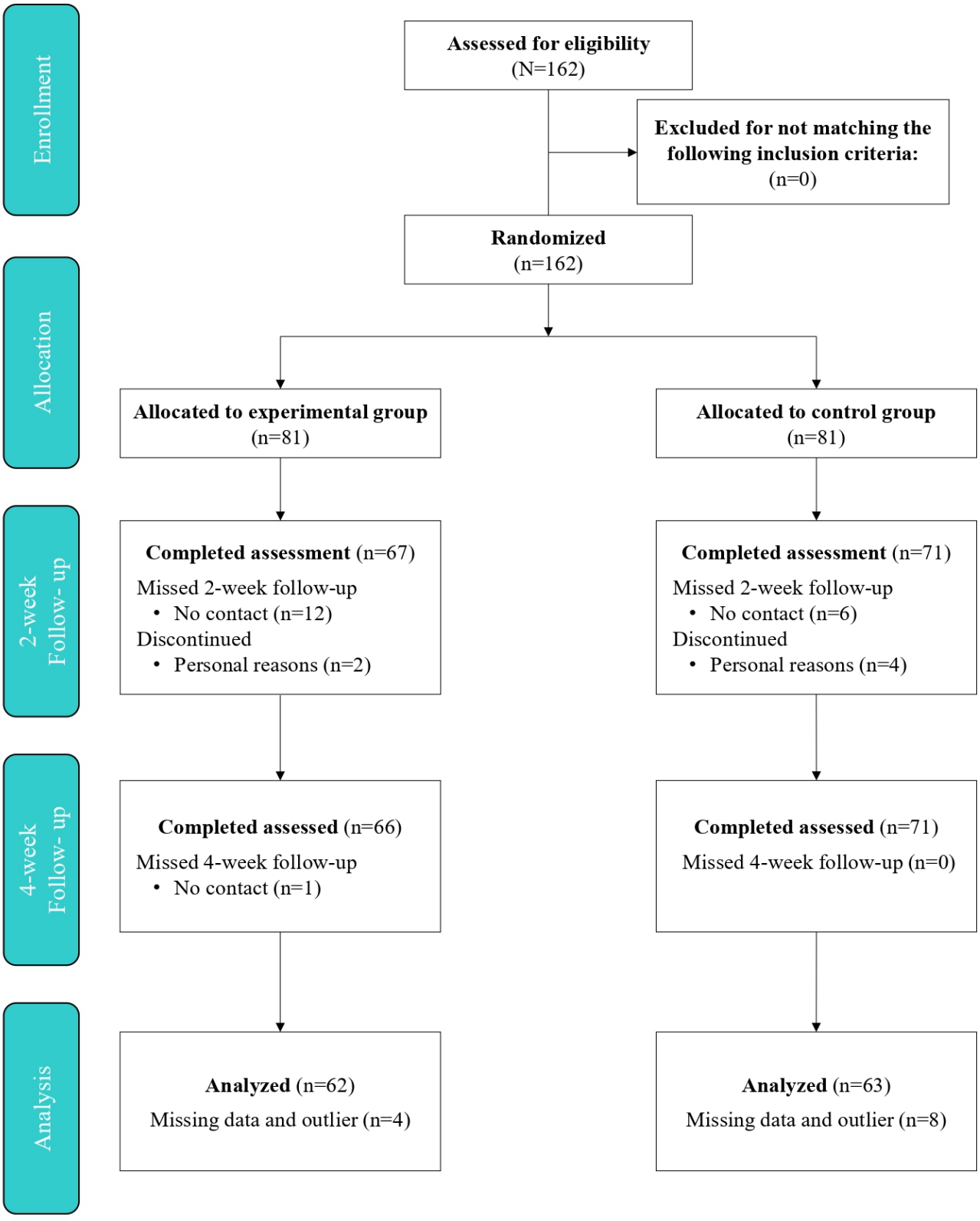
[c]ICU: intensive care unit.

[d]ER: emergency room.

[e]Others: operating room, recovery room, delivery room, outpatient, administration.

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram profile flow.

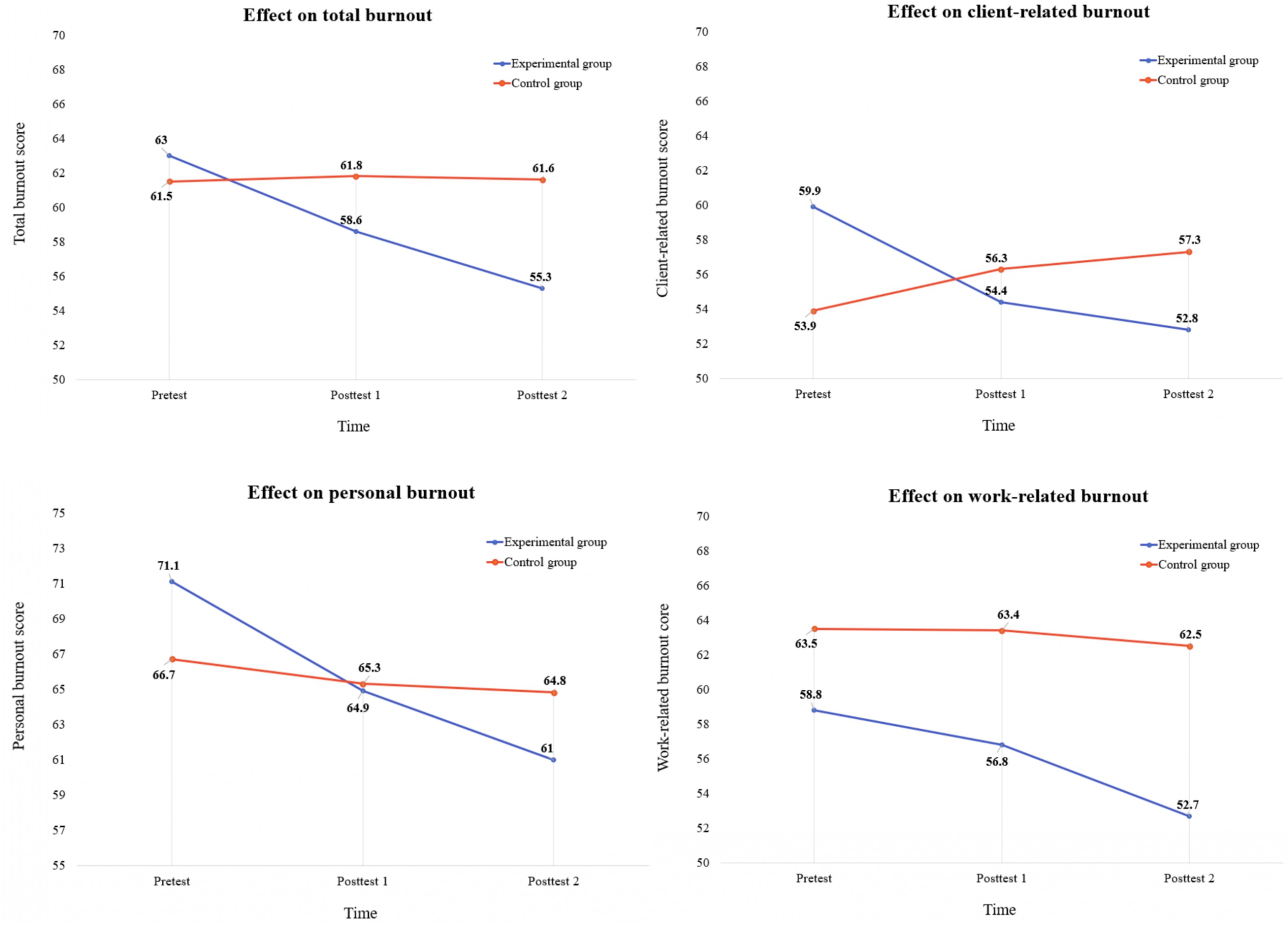


Effects of the Intervention on Nurse Burnout

Figure 2 and Table 3 present the repeated-measures ANOVA results with covariates. The experimental group showed significantly greater reductions than the control group in client-related burnout ($F_{1,121}$=7.548, P=.007), personal burnout ($F_{1,121}$=6.533, P=.01), and work-related burnout ($F_{1,121}$=38.194, P<.001), reflecting significant group effects. No significant time or interaction effects were observed. In addition, paired-sample t tests indicated significant within-group reductions from baseline to posttest 2 in the experimental group across total burnout and all burnout subdomains, with small-to-moderate effect sizes. However, no significant within-group changes were observed in the control group.

XSL•FO
RenderX

Figure 2. Trends in burnout scores over time, based on repeated-measures ANOVA with covariates.



XSL•FO
RenderX

Table . Paired-sample *t* tests and repeated-measures ANOVA with covariates (N=125).

Study variables and group	Pretest, mean (SD)	Posttest 1, mean (SD)	Posttest 2, mean (SD)	Difference[a], mean (SD)	Paired *t* test (*P* value)	Cohen *d*	Repeated-measures ANOVA, *F* (*df*=1,121) (*P* value)
Total burnout score							2.943 (.06)[b]
Exp.[c] (n=62)	63.01 (17.57)	58.62 (17.25)	55.30 (23.11)	7.70 (19.82)	3.06 (.003)	0.39	16.960 (<.001)[d]
Cont.[e] (n=63)	61.49 (18.54)	61.77 (18.31)	61.59 (21.60)	−0.10 (14.13)	−0.06 (.95)	−0.01	0.718 (.47)[f]
Client-related burnout							2.923 (.06)[b]
Exp. (n=62)	59.88 (24.17)	54.44 (21.99)	52.76 (25.02)	7.64 (24.93)	2.34 (.02)	0.30	7.548 (.007)[d]
Cont. (n=63)	53.90 (22.44)	56.28 (22.30)	57.34 (24.10)	−4.05 (19.13)	−1.40 (.17)	−0.18	0.384 (.66)[f]
Personal burnout							2.596 (.08)[b]
Exp. (n=62)	71.10 (18.57)	64.92 (19.39)	60.95 (26.87)	10.15 (20.67)	3.87 (<.001)	0.49	6.533 (.01)[d]
Cont. (n=63)	66.73 (19.72)	65.34 (19.05)	64.81 (23.24)	1.92 (15.77)	0.97 (.34)	0.12	0.283 (.74)[f]
Work-related burnout							1.669 (.19)[b]
Exp. (n=62)	58.76 (16.56)	56.80 (15.00)	52.65 (21.69)	6.11 (20.90)	2.30 (.03)	0.29	38.194 (<.001)[d]
Cont. (n=63)	63.49 (19.25)	63.44 (19.48)	62.47 (21.62)	1.02 (14.21)	0.57 (.57)	0.07	1.854 (.16)[f]

[a]Differences: pretest score—posttest 2 score.

[b]Source: Interaction effects.

[c]Exp.: experimental group.

[d]Source: Group effects.

[e]Cont.: control group.

[f]Source: Time effects.

Effects of the Intervention on Coping Strategy, Job Stress, and Stress Response

Table 4 presents between-group comparisons of coping strategy, job stress, and stress response from pretest to posttest 2. No statistically significant between-group differences were observed for these secondary outcomes. Although the experimental group showed numerically greater reductions in job stress and stress response compared with the control group, the between-group differences effect sizes were small to modest in magnitude.

Table . Independent *t* test: group differences in job stress, stress response, and coping strategy (N=125).

Study variables and group	Pretest, mean (SD)	Posttest 2, mean (SD)	Differences[a], mean (SD)	*t* (*df*=123) (*P* value)	Cohen *d*
Coping strategy				−1.358 (.18)	−0.24
Exp.[b] (n=62)	2.08 (0.28)	2.15 (0.35)	−0.07 (0.35)		
Cont.[c] (n=63)	2.10 (0.26)	2.09 (0.33)	0.01 (0.31)		
Job stress				1.848 (.07)	0.33
Exp. (n=62)	3.95 (0.44)	3.64 (0.81)	0.31 (0.66)		
Cont. (n=63)	3.65 (0.44)	3.56 (0.67)	0.09 (0.66)		
Stress response				1.237 (.22)	0.22
Exp. (n=62)	2.05 (0.66)	1.77 (0.83)	0.28 (0.79)		
Cont. (n=63)	1.55 (0.80)	1.44 (0.88)	0.11 (0.76)		

[a]Differences: pretest score—posttest 2 score.

[b]Exp.: experimental group.

[c]Cont.: control group.

XSL•FO
RenderX

Participant Experiences With the Intervention

Overview

We identified 2 overarching themes: experiences with the AI algorithm and perceptions of the burnout reduction program. Subthemes within each theme reflected participants' emotional, cognitive, and behavioral responses during the intervention.

Experiences With the Burnout Reduction Program

Participants described the program as an opportunity for self-reflection, emotional relief, and the development of coping strategies applicable to daily life.

Reflecting on Oneself

Several participants noted that guided activities, such as reflective writing and mindfulness meditation, enabled them to reassess their emotional state and recognize personal growth. These activities facilitated intentional self-observation and reflection on past experiences. For example, 1 participant explained:

> *During reflective writing, I looked back on past memories. It was tough then, but now I could see how much I have grown.*

Another participant stated:

> *The meditation gave me space to empty my mind and reflect. Having time just for myself was the biggest change.*

Feeling Calmer Overall

Participants reported feeling more relaxed and emotionally composed through the program's cognitive and physical relaxation components, which, in some cases, carried over into their clinical practice. One participant noted:

> *I do not know if it helped with burnout directly, but I felt more at ease. I felt calmer.*

Another participant shared:

> *When dealing with patients, I responded more gently instead of reacting harshly. I could keep calm even in busy environments.*

Creating a Mental Space for Recovery

Participants described how emotional residue from clinical work often lingered after shifts, exacerbating burnout. The program provided temporary psychological detachment from work-related stressors and created a mental space for recovery. One participant explained:

> *I used to just eat or sleep after work, but even after sleeping, the stress did not go away. This program helped me forget about it for a while.*

Another participant noted:

> *Focusing on the video and sound helped me forget things, even if just briefly.*

Strengthening Self-Efficacy Through Personalized Coping

Participants reported increased confidence in managing stress and preventing burnout. By identifying coping techniques suited to their needs and integrating them into daily life, they strengthened their sense of self-efficacy. One participant shared:

> *Stretching helped me regulate my emotions, and now I stretch as a daily habit.*

Another participant explained:

> *When I am struggling, I try taking deep breaths or moving my body—my own way of coping.*

Experiences With the AI Algorithm

Participants were often introduced to interventions they had not previously considered. They also expressed a strong sense of trust in the AI-based recommendation process.

Exposure to New Programs Beyond Preexisting Preferences

Participants noted that the AI algorithm encouraged them to engage in programs they would not have selected independently, thereby reducing reliance on habitual or preference-driven choices. One participant reflected:

> *As people, we tend to choose what we already know. But because the AI recommended something different, I could experience a variety of options.*

Another participant added:

> *It was a refreshing experience. I could try methods I would not have chosen myself, and they turned out to work surprisingly well for me.*

Trust in AI-Based Recommendations

Participants expressed confidence in the AI-generated recommendations, emphasizing that the suggestions felt objective and tailored to their personal condition. These comments suggest that some participants were receptive to the AI-recommended programs, even when the content differed from what they would have personally preferred. One participant noted:

> *When friends recommend something, I sometimes dismiss it because they do not really know me. However, with AI, it feels like it is based on a thorough understanding of my condition.*

Another participant added:

> *If I chose a program myself, I would have picked something I liked, not necessarily what I needed. The AI seemed to know what was appropriate for me at the moment.*

Discussion

Principal Findings

In this study, a greater reduction was observed in all 3 burnout subdomains in the experimental group compared with the control group. However, these results should be interpreted in light of the absence of significant time and group interaction effects. These findings suggest a potential role for tailored interventions in supporting burnout reduction, although confirmatory evidence demonstrating significant interaction effects is still needed. Previous studies have suggested the role of CBT in mitigating

XSL•FO
RenderX

burnout [15,16] and extending it by integrating AI-based personalization into mHealth delivery. Previous studies applied standardized programs to all participants without accounting for individual differences in burnout profiles or coping strategies [38,39]. A recent meta-analysis also confirmed that mobile CBT-based interventions generally reduce stress and burnout among nurses but noted heterogeneity in effects due to intervention type and user engagement [40]. This study assists in addressing these gaps by using an AI algorithm that tailors CBT-based interventions to users' demographic, work-related characteristics, and psychological profiles.

Although several mHealth interventions, such as Wysa and Woebot, have demonstrated effectiveness in reducing psychological symptoms through chatbot-delivered CBT, these platforms typically tailor content based on brief, momentary self-reported mood or symptom check-ins [19,20]. By contrast, this study used a data-driven recommendation algorithm informed by multidimensional baseline self-report data, including burnout subdomains, job stress levels, and coping styles, to guide the selection of CBT-based modules. This approach was intended to support a relatively more condition-specific and theoretically grounded personalization strategy rather than real-time mood-based content adaptation. The theoretical grounding of this approach is based on a multidimensional conceptualization of burnout, which assumes that different burnout subtypes are associated with distinct cognitive and behavioral mechanisms [7]. Accordingly, the recommendation algorithm was designed to align users' burnout profiles and coping characteristics with CBT-based modules targeting specific therapeutic mechanisms, such as cognitive restructuring, behavioral activation, and emotion regulation [15,16]. Moreover, while many digital mental health interventions primarily rely on quantitative outcome metrics, this study additionally incorporated a qualitative component to capture participants' lived experiences with the AI-assisted program. The thematic analysis highlighted users' reflections on increased self-awareness [41], trust in AI recommendations, and emotional relief, thereby providing contextual insight into the observed quantitative results.

Despite the generally favorable trends observed, the magnitude of the observed changes in burnout was modest, which necessitates a cautious interpretation of the direct effectiveness of the AI-based mobile CBT-based intervention. This pattern is consistent with findings from a recent systematic review of person-directed psychoeducational interventions for nurses, which reported that CBT-based approaches generally contribute to reductions in burnout while also demonstrating considerable variability in effect sizes and limited evidence regarding the long-term sustainability of effects [42]. Furthermore, the fully remote and self-guided nature of the intervention may have contributed to variability in how participants engaged with the program, which could partially account for the modest intervention effects observed in this study. Future studies should incorporate more comprehensive adherence-reporting methods, including objective indicators of behavioral engagement, such as module completion rates, frequency of session participation, usage duration, and interaction logs. Such approaches may help clarify the relationship between intervention engagement and treatment outcomes in fully remote CBT-based interventions. Despite these limitations, the qualitative results of this study provide important contextual insight into how such interventions may exert their influence. Rather than indicating immediate or substantial symptom reduction, participants described gradual changes in self-awareness, coping orientation, and engagement with stress-related situations. These experiential changes suggest that AI-selected mobile CBT-based interventions may operate by facilitating early-stage cognitive and behavioral adjustments, which may precede more pronounced reductions in burnout over time. Integrating these qualitative insights with the quantitative outcomes allows for a more nuanced interpretation of the intervention effects, particularly in the context of digital and mobile-based burnout interventions for nurses.

Although mobile and AI-assisted interventions are often assumed to reduce dropout by improving accessibility and flexibility, the attrition rate in this study approached 20% [11], which is comparable to rates reported in face-to-face CBT-based interventions. This finding suggests that the mobile format alone may not be sufficient to substantially reduce dropout among nurses. Factors such as high baseline occupational stress, shift work, and limited cognitive and emotional resources may have constrained sustained engagement, even in a self-guided and flexible intervention context [2]. These findings indicate that although mobile delivery may lower structural barriers, additional strategies, such as adaptive reminders, enhanced human support, or shorter intervention modules, may be necessary to further improve retention in future digital burnout interventions for nurses.

A particularly noteworthy finding of this study is that participants reported engaging in coping strategies they might not have selected on their own, specifically because they were recommended by the AI algorithm. From the interviews, participants reported that AI-generated recommendations encouraged engagement with unfamiliar coping strategies that were later perceived to be helpful. Such accounts indicate that the AI recommendations may have contributed to prompting participants to move beyond their habitual coping tendencies and consider alternative strategies. Prior research suggests that individuals experiencing burnout tend to exhibit reduced cognitive flexibility due to emotional exhaustion, often resorting to avoidant and passive coping strategies, such as denial, self-blame, and behavioral disengagement [43,44]. Although these strategies may temporarily conserve psychological energy, they ultimately fail to address the root causes of stress and can perpetuate a cycle of worsening burnout. Within this context, our findings suggest that the AI-selected intervention did more than merely provide additional options. By tailoring recommendations to participants' burnout subdomains and psychological profiles, it may have helped reduce preference-driven selection and encouraged consideration of strategies that participants might otherwise have overlooked. In this sense, the AI algorithm functioned not only as a personalization tool but also as a facilitator that broadened participants' coping repertoires and supported openness to new and potentially more adaptive approaches.

Moreover, the participants who were interviewed expressed a high level of trust in the AI-generated recommendations. This

XSL•FO
RenderX

trust appeared to stem from their perception that the recommendations were objective and data-driven; however, this interpretation should be viewed cautiously, as the study did not include a comparison condition involving recommendations from human professionals. Prior research has shown that some individuals tend to attribute greater objectivity to algorithmic guidance [45], and trust is known to be a key factor influencing engagement with and continued use of digital mental health interventions [46]. In this study, participants' trust in the AI-assisted program may have supported openness toward continued participation, although the relative advantage of AI-generated recommendations over human-delivered recommendations cannot be inferred from the current study design. Furthermore, this study suggests that AI-based mental health interventions may represent a promising complementary approach to conventional group-based or offline burnout programs, which often face structural limitations. Traditional interventions rely on fixed formats, predetermined schedules, and in-person settings, making them less accessible or feasible for professionals with demanding schedules, such as shift-working nurses. A systematic review reported that many burnout interventions were group-based and conducted offline, and they consistently noted barriers, such as scheduling difficulties, participant dropout, and lack of long-term effectiveness [42]. By contrast, the AI algorithm applied in this study delivered tailored CBT-based modules by automatically analyzing multidimensional user data, including burnout subtypes, stress levels, and coping styles. This approach enabled precise and flexible intervention without time or location constraints.

Such technological adaptability is not limited to specific professions but offers the scalability and inclusiveness needed to support high-risk occupational groups (eg, teachers, firefighters, and social workers) as well as vulnerable populations (eg, adolescents, older adults, and informal caregivers). As data from more diverse groups accumulate and algorithmic learning advances, AI interventions may become increasingly capable of addressing a broader range of mental health conditions, such as anxiety, depression, sleep disorders, and posttraumatic stress disorder. In this regard, AI-powered digital interventions may offer a scalable and sustainable approach that simultaneously enhances accessibility and personalization—2 essential elements in the future of mental health services.

Limitations

This study has several limitations that should be considered when interpreting the findings. First, all outcome measures were self-reported, thereby raising concerns about potential reporting bias, including social desirability. Second, participants were recruited without a minimum screening threshold for burnout symptoms. Although the mean baseline burnout scores exceeded 50, indicating generally elevated burnout levels among participants, the absence of a formal screening process means that participants were not explicitly selected based on a confirmed need for intervention. Although this design allows for the examination of the AI-selected mobile CBT-based intervention in real-world nurses who are vulnerable to burnout, it limits the ability to draw conclusions about intervention effects among individuals with suspected or clinically significant burnout. Future studies should consider targeting participants with elevated burnout levels to more directly test intervention efficacy. Third, although participants' access to the intervention platform and participation status were monitored digitally, this study did not include objective adherence indicators reflecting the depth of engagement with the intervention, such as total usage duration, module completion rates, session completion frequency, or detailed interaction logs. As a result, it was not possible to determine the extent to which participants actively engaged with the intervention content throughout the study period. Because this study used a fully remote, self-guided intervention format, the absence of more robust adherence data limits the interpretation of the observed intervention effects and may partially explain the relatively modest or inconsistent effects observed across several outcomes. Future studies should incorporate more comprehensive adherence-reporting methods, including objective behavioral engagement indicators, to better evaluate the relationship between intervention engagement and treatment outcomes. Fourth, although this study was conducted as a single-blind randomized controlled trial, complete participant blinding could not be fully guaranteed. Because the control group received only informational materials while the experimental group engaged with interactive AI-selected modules, participants may have inferred their group assignment. Such potential unblinding may have influenced their expectations or engagement levels, thereby introducing performance bias. Fifth, this study did not include a comparison group receiving recommendations from human professionals. As a result, interpretations regarding the relative superiority or added benefit of AI-generated recommendations should be made with caution. Sixth, baseline levels of job stress and stress response were significantly higher in the experimental group than in the control group, and these variables were statistically adjusted for using analysis of covariance. However, such baseline heterogeneity may have influenced responsiveness to the intervention, thereby complicating interpretation of the observed group differences. Participants with higher baseline psychological distress may have responded differently to the intervention compared with less distressed participants. Future studies should consider stratified randomization or subgroup analyses to better account for baseline psychological differences. Seventh, this study did not use an ITT analysis. Excluding participants with missing data or outliers may have reduced the benefits of randomization and introduced potential selection bias. Future studies should consider applying ITT approaches, such as mixed-effects modeling or multiple imputation, to better preserve randomization and enhance the robustness of the findings.

Conclusions

This study evaluated the effectiveness of an AI-selected, mobile, CBT-based intervention in reducing nurse burnout and explored participants' experiences with how the AI-selected program supported them during the intervention period. By tailoring program selection based on individual characteristics and burnout-related factors, the intervention was designed to address some limitations of standardized approaches. Nurses who received the AI-selected intervention showed reductions in

XSL•FO
RenderX

burnout, suggesting the potential use of personalized digital mental health interventions as a supportive approach for nurse burnout management. These findings highlight the potential of AI technology to enhance individualized care in occupational mental health and support further studies to examine its long-term effects and applicability across diverse health care settings.

Acknowledgments

The authors sincerely thank all the nurses who participated in this study for their time and valuable contributions. During the preparation of this manuscript, generative artificial intelligence (AI) tool ChatGPT (GPT-5.5; OpenAI) was used to assist with language polishing and improving sentence clarity. The AI tool was not used for idea generation, interpretation of results, or reference creation. All content generated with AI assistance was carefully reviewed, edited, and verified by the authors. The authors take full responsibility for the accuracy and integrity of the manuscript.

Funding

This work was supported by the National Research Foundation of Korea grant funded by the South Korean government (Ministry of Science and Information and Communication Technology) (RS-2024-00351194).

Data Availability

The datasets analyzed during this study are not publicly available due to institutional review board restrictions; however, they are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: YK, CC
Data curation: YK, GB
Formal analysis: GB
Funding acquisition: CC
Investigation: YK, GB
Methodology: YK, GB, CC
Project administration: YK
Supervision: CC
Visualization: YK
Writing – original draft: GB
Writing – review & editing: YK, GB, CC

Conflicts of Interest

None declared.

Checklist 1
CONSORT checklist.
[PDF File, 165 KB - mhealth_v14i1e85986_app1.pdf]

References

1. The Lancet. 2020: unleashing the full potential of nursing. Lancet 2019 Nov 23;394(10212):1879. [doi: 10.1016/S0140-6736(19)32794-1] [Medline: 31777376]
2. Khatatbeh H, Pakai A, Al-Dwaikat T, et al. Nurses' burnout and quality of life: a systematic review and critical analysis of measures used. Nurs Open 2022 May;9(3):1564-1574. [doi: 10.1002/nop2.936] [Medline: 33991408]
3. Zhang YY, Han WL, Qin W, et al. Extent of compassion satisfaction, compassion fatigue and burnout in nursing: a meta-analysis. J Nurs Manag 2018 Oct;26(7):810-819. [doi: 10.1111/jonm.12589] [Medline: 30129106]
4. Woo T, Ho R, Tang A, Tam W. Global prevalence of burnout symptoms among nurses: a systematic review and meta-analysis. J Psychiatr Res 2020 Apr;123:9-20. [doi: 10.1016/j.jpsychires.2019.12.015] [Medline: 32007680]
5. Li LZ, Yang P, Singer SJ, Pfeffer J, Mathur MB, Shanafelt T. Nurse burnout and patient safety, satisfaction, and quality of care: a systematic review and meta-analysis. JAMA Netw Open 2024 Nov 4;7(11):e2443059. [doi: 10.1001/jamanetworkopen.2024.43059] [Medline: 39499515]
6. Noh EY, Park YH, Chai YJ, Kim HJ, Kim E. Frontline nurses' burnout and its associated factors during the COVID-19 pandemic in South Korea. Appl Nurs Res 2022 Oct;67:151622. [doi: 10.1016/j.apnr.2022.151622] [Medline: 36116862]
7. Kristensen TS, Borritz M, Villadsen E, Christensen KB. The Copenhagen Burnout Inventory: a new tool for the assessment of burnout. Work Stress 2005 Jul;19(3):192-207. [doi: 10.1080/02678370500297720]

XSL•FO
RenderX

8. Peters M. Time to solve persistent, pernicious and widespread nursing workforce shortages. Int Nurs Rev 2023 Jun;70(2):247-253. [doi: 10.1111/inr.12837] [Medline: 37000668]
9. Ogugua JO, Muonde M, Maduka CP, Olorunsogo TO, Omotayo O. Demographic shifts and healthcare: a review of aging populations and systemic challenges. Int J Sci Res Arch 2024;11(1):383-395. [doi: 10.30574/ijsra.2024.11.1.0067]
10. Nehra DK, Sharma NR, Kumar P, Nehra S, et al. Cognitive behaviour therapy: an overview. In: Hooda D, Sharma NR, editors. Mental Health: Risks and Resources: Global Vision Publishing House; 2013:272-303.
11. Han JH, Lee M, Cha C, Baek G. Effects of third-wave cognitive behavioral therapy for healthcare professionals' burnout: a systematic review and meta-analysis. Healthcare (Basel) 2025 Dec 11;13(24):3253. [doi: 10.3390/healthcare13243253] [Medline: 41464322]
12. Dionigi A, Canestrari C. The role of laughter in cognitive-behavioral therapy: case studies. Discourse Stud 2018;20(3):323-339. [doi: 10.1177/1461445618754426]
13. Davis CH, Donahue ML, Gaudiano BA, Uebelacker LA, Twohig MP, Levin ME. Adding online storytelling-based acceptance and commitment therapy to antidepressant treatment for primary care patients: a randomized clinical trial. Cogn Behav Ther 2024 Jan;53(1):48-69. [doi: 10.1080/16506073.2023.2265560] [Medline: 37855277]
14. Prasko J, Mozny P, Novotny M, Slepecky M, Vyskocilova J. Self-reflection in cognitive behavioural therapy and supervision. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub 2012 Dec;156(4):377-384. [doi: 10.5507/bp.2012.027] [Medline: 23073517]
15. Musker M, Othman S. Effective interventions to reduce burnout in nurses: a meta-analysis. Complement Ther Clin Pract 2024 Feb;54:101827. [doi: 10.1016/j.ctcp.2023.101827] [Medline: 38181650]
16. Lee M, Cha C. Interventions to reduce burnout among clinical nurses: systematic review and meta-analysis. Sci Rep 2023 Jul 6;13(1):10971. [doi: 10.1038/s41598-023-38169-8] [Medline: 37414811]
17. Singh B, Ahmed M, Staiano AE, et al. A systematic umbrella review and meta-meta-analysis of eHealth and mHealth interventions for improving lifestyle behaviours. NPJ Digit Med 2024 Jul 5;7(1):179. [doi: 10.1038/s41746-024-01172-y] [Medline: 38969775]
18. Paterson C, Leduc C, Maxwell M, et al. Barriers and facilitators to implementing workplace interventions to promote mental health: qualitative evidence synthesis. Syst Rev 2024 Jun 7;13(1):152. [doi: 10.1186/s13643-024-02569-2] [Medline: 38849924]
19. Meheli S, Sinha C, Kadaba M. Understanding people with chronic pain who use a cognitive behavioral therapy-based artificial intelligence mental health app (Wysa): mixed methods retrospective observational study. JMIR Hum Factors 2022 Apr 27;9(2):e35671. [doi: 10.2196/35671] [Medline: 35314422]
20. Yeh PL, Kuo WC, Tseng BL, Sung YH. Does the AI-driven chatbot work? Effectiveness of the Woebot app in reducing anxiety and depression in group counseling courses and student acceptance of technological aids. Curr Psychol 2025 May;44(9):8133-8145. [doi: 10.1007/s12144-025-07359-0]
21. Chaturvedi U, Chauhan SB, Singh I. The impact of artificial intelligence on remote healthcare: enhancing patient engagement, connectivity, and overcoming challenges. Intell Pharm 2025 Oct;3(5):323-329. [doi: 10.1016/j.ipha.2024.12.003]
22. Chinnasamy P, Wong WK, Raja AA, Khalaf OI, Kiran A, Babu JC. Health recommendation system using deep learning-based collaborative filtering. Heliyon 2023;9(12):e22844. [doi: 10.1016/j.heliyon.2023.e22844] [Medline: 38144343]
23. Farzan M, Ebrahimi H, Pourali M, Sabeti F. Artificial intelligence-powered cognitive behavioral therapy chatbots, a systematic review. Iran J Psychiatry 2025 Jan;20(1):102-110. [doi: 10.18502/ijps.v20i1.17395] [Medline: 40093525]
24. Baek G, Cha C. AI-assisted tailored intervention for nurse burnout: a three-group randomized controlled trial. Worldviews Evid Based Nurs 2025 Feb;22(1):e70003. [doi: 10.1111/wvn.70003] [Medline: 39981583]
25. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. BMJ 2020 Sep 9;370:m3164. [doi: 10.1136/bmj.m3164] [Medline: 32909959]
26. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
27. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition: Lawrence Erlbaum Associates; 1988.
28. Luangapichart P, Saisavoey N, Viravan N. Efficacy and feasibility of the minimal therapist-guided four-week online audio-based mindfulness program 'Mindful Senses' for burnout and stress reduction in medical personnel: a randomized controlled trial. Healthcare (Basel) 2022 Dec 14;10(12):2532. [doi: 10.3390/healthcare10122532] [Medline: 36554056]
29. Schmidt ID, Forand NR, Strunk DR. Predictors of dropout in internet-based cognitive behavioral therapy for depression. Cognit Ther Res 2019 Jun;43(3):620-630. [doi: 10.1007/s10608-018-9979-5] [Medline: 32879540]
30. Cho A, Cha C, Baek G. Development of an artificial intelligence-based tailored mobile intervention for nurse burnout: single-arm trial. J Med Internet Res 2024 Jun 21;26:e54029. [doi: 10.2196/54029] [Medline: 38905631]
31. Ham MJ. A path analysis study on nurses' emotional labor and burnout [Master's Thesis]. : Gyeongsang National University; 2011 URL: https://dcollection.gnu.ac.kr/public_resource/pdf/000000014045_20260614085312.pdf [accessed 2026-06-15]
32. Amirkhan JH. A factor analytically derived measure of coping: the Coping Strategy Indicator. J Pers Soc Psychol 1999;59(5):1066-1074. [doi: 10.1037/0022-3514.59.5.1066]

XSL•FO
RenderX

33. Shin HJ, Kim CD. A validation study of Coping Strategy Indicator (CSI). Korean J Couns Psychother 2002;14(4):919-935 [FREE Full text]
34. Gu MO, Kim MJ. An analytical study of work stress among clinical nurses. J Nurs Acad Soc 1985;15(3):39-49. [doi: 10.4040/jnas.1985.15.3.39]
35. Ahn MK. The effect of nurse's job stress on health practice behavior [Master's Thesis]. : Yonsei University; 2003.
36. Koh KB, Park JK, Kim CH. Development of the Stress Response Inventory. J Korean Neuropsychiatr Assoc 2000;39(4):707-719 [FREE Full text]
37. Choi S, Kang T, Woo J. Development and validation of a modified form of the Stress Response Inventory for workers. J Korean Neuropsychiatr Assoc 2006;45(6):541-553 [FREE Full text]
38. Fiol-DeRoque MA, Serrano-Ripoll MJ, Jiménez R, et al. A mobile phone-based intervention to reduce mental health problems in health care workers during the COVID-19 pandemic (PsyCovidApp): randomized controlled trial. JMIR mHealth uHealth 2021 May 18;9(5):e27039. [doi: 10.2196/27039] [Medline: 33909587]
39. Lu Y, Li Y, Huang Y, et al. Effects and mechanisms of a web- and mobile-based acceptance and commitment therapy intervention for anxiety and depression symptoms in nurses: fully decentralized randomized controlled trial. J Med Internet Res 2023 Nov 27;25:e51549. [doi: 10.2196/51549] [Medline: 38010787]
40. Chakraborty B, Trivedi K, Asha Kiran K, Chatterjee P, Kujur M, Sarkar B. Impact of using mobile applications on the psychological wellbeing of nurses to reduce job stress and burnout: systematic review & meta-analysis. Malays J Nurs 2023;15(2):143-153. [doi: 10.31674/mjn.2023.v15i02.016]
41. Clarke V, Braun V. Thematic analysis. In: Teo T, editor. Encyclopedia of Critical Psychology: Springer; 2014:1947-1952. [doi: 10.1007/978-1-4614-5583-7_311]
42. Abdullah Sharin IB, Jinah NB, Bakit PA, et al. Person-directed burnout intervention for nurses: a systematic review of psychoeducational approaches. PLoS One 2025;20(5):e0322282. [doi: 10.1371/journal.pone.0322282] [Medline: 40343910]
43. Montero-Marin J, Prado-Abril J, Piva Demarzo MM, Gascon S, García-Campayo J. Coping with stress and types of burnout: explanatory power of different coping strategies. PLoS One 2014;9(2):e89090. [doi: 10.1371/journal.pone.0089090] [Medline: 24551223]
44. Xavier GMV, Dos Santos Júnior CJ, Ribeiro MC, Teixeira GM. Burnout syndrome and coping strategies among professors in the health area. Rev Bras Med Trab 2024;22(3):e20241175. [doi: 10.47626/1679-4435-2024-1175] [Medline: 39606754]
45. Logg JM, Minson JA, Moore DA. Algorithm appreciation: people prefer algorithmic to human judgment. Organ Behav Hum Decis Process 2019 Mar;151:90-103. [doi: 10.1016/j.obhdp.2018.12.005]
46. Kauttonen J, Rousi R, Alamäki A. Trust and acceptance challenges in the adoption of AI applications in health care: quantitative survey analysis. J Med Internet Res 2025 Mar 21;27:e65567. [doi: 10.2196/65567] [Medline: 40116853]

Abbreviations

AI: artificial intelligence
CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
COREQ: Consolidated Criteria for Reporting Qualitative Research
ITT: intention-to-treat
mHealth: mobile health

Edited by L Buis; submitted 16.Oct.2025; peer-reviewed by MA Bushey, SI Yoon; revised version received 11.May.2026; accepted 04.Jun.2026; published 03.Jul.2026.

Please cite as:
Kim Y, Cha C, Baek G
Alleviating Nurse Burnout With an Artificial Intelligence–Selected Mobile Cognitive Behavioral Therapy–Based Intervention: Mixed Methods Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e85986
URL: https://mhealth.jmir.org/2026/1/e85986
doi:10.2196/85986

© Yeongeun Kim, Chiyoung Cha, Gumhee Baek. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 3.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effects, Acceptability, and Use of a Dynamically Tailored Mobile What Do You Drink Intervention to Reduce Excessive Drinking Among Adolescents and Young Adults in the Netherlands: Randomized Controlled Trial

Hilde van Keulen[1], PhD; Carmen Voogt[2], PhD; Marloes Kleinjan[3], PhD; Arjan Huizing[1], MSc; Rosa Andree[4], MSc; Pepijn van Empelen[1,5], PhD

[1]Department of Child Health, Netherlands Organisation for Applied Scientific Research, Sylviusweg 71, Leiden, The Netherlands

[2]Eating Disorder Experts Network, Emergis and Parnassia Groep, The Hague, The Netherlands

[3]HAN (Hogeschool Arnhem Nijmegen) University of Applied Sciences, Research Group Organisation of Care and Service Delivery, Arnhem, The Netherlands

[4]Drug Monitoring and Policy, Trimbos Institute, Utrecht, The Netherlands

[5]Department of Health, Medical and Neuropsychology, University of Leiden, Leiden, The Netherlands

Corresponding Author:
Hilde van Keulen, PhD
Department of Child Health, Netherlands Organisation for Applied Scientific Research, Sylviusweg 71, Leiden, The Netherlands

Abstract

Background: Excessive alcohol consumption among adolescents and young adults is a serious health problem. Dynamically tailored interventions could reduce their excessive drinking. We therefore developed “What Do You Drink” (WDYD), a 17-week dynamically tailored mHealth (mobile health) intervention providing personalized support on alcohol consumption.

Objective: We aim to evaluate the effectiveness, acceptability, and use of WDYD in reducing alcohol consumption of adolescents and young adults at risk.

Methods: We conducted a 2-arm, parallel-group randomized controlled trial using ecological momentary assessments. Recruitment was via an educational alcohol program, an online lifestyle monitor, social media advertisements, or news items on websites. Participants downloaded the standalone WDYD app, and when having given active informed consent, were randomized to the intervention or control group. Participants in the intervention group received dynamically tailored feedback sessions on alcohol consumption (wk 0 - 5, 7, 9, 13, and 17) and goal-monitoring reminders. Both groups completed an online baseline survey, 2 follow-up surveys (wk 9 and 33), and various ecological momentary assessments (7 daily assessments during wk 1, 7, 13, 19, 25, 31, and 33). Participants provided consent before randomization, in which they were informed that 2 study groups existed. After randomization, no disclosure of group assignment was provided, although participants could potentially infer it from receiving tailored sessions vs no tailored sessions. Primary outcomes were excessive drinking, binge drinking, and weekly alcohol consumption. Secondary outcomes were intrinsic motivation, self-confidence, and mood. Acceptability of WDYD was measured by survey questions; use was tracked via app data logs.

Results: Analyses were based on data from 1767 participants; 720 in the intervention group and 1047 participants in the control group. Almost half of them were female (2276/4795, 47.5%), and most (3471/4595, 72.4%) participants were aged 18 - 24 (median 19.40, IQR 2.92) years. The dropout rate was high, up to 96% (4603/4795) in the final 33rd week. No significant effect of WDYD was found on primary outcomes and mood, except for week 1 (excessive drinking: standardized β=−0.35, SE 0.15; 95% CI −0.64 to −0.05; binge drinking: standardized β=−0.36, SE 0.16; 95% CI −0.68 to −0.04; mood: standardized β=0.20, SE 0.06, 95% CI 0.08 to 0.32). Both groups reduced their alcohol consumption. Significant positive effects were found for intrinsic motivation and self-confidence up to 25 weeks (wk 25: standardized β=0.54, SE 0.24; 95% CI 0.06 to 1.02 for motivation; standardized β=0.72, SE 0.26; 95% CI 0.22 to 1.23 for self-confidence). Participants evaluated WDYD as acceptable and usable.

Conclusions: WDYD did not significantly reduce excessive drinking compared to control, but improved motivation and self-confidence. High dropout rates highlight challenges in sustaining engagement in long-term mHealth interventions. Future research should explore strategies to enhance retention and optimize dynamic tailoring.

Trial Registration: ICTRP NL-OMON28135; https://trialsearch.who.int/Trial2.aspx?TrialID=NL-OMON28135

(JMIR Mhealth Uhealth 2026;14:e68468) doi:10.2196/68468

XSL•FO
RenderX

KEYWORDS

alcohol consumption; excessive drinking; at-risk population; adolescents; young adults; dynamic tailoring; mHealth intervention; randomized controlled trial; health promotion

Introduction

Background

Excessive alcohol drinking among adolescents and young adults is a large public health concern in Western countries [1]; it is negatively associated with short-term and longer-term negative outcomes, with regard to health, academic performance, social relationships, and finishing school [1]. Specifically among adolescents and young adults, alcohol consumption occurs in groups and is strongly influenced by the group [2-4]. In the Netherlands, various studies show that adolescents and young adults engage in excessive drinking [5-7]. In accordance with the guidelines, excessive drinking is defined as heavy drinking (ie, drinking at least 8 [women] or 15 [men] alcoholic beverages per week) and/or binge drinking (ie, drinking at least 4 [women] or 6 [men] drinks at 1 occasion [8,9]). Preventing excessive drinking is important to improve the health of individuals and to reduce societal costs.

Potential for mHealth Interventions

mHealth (mobile health) interventions can be a useful means to reach young people and to prevent them from drinking excessively [10]. mHealth interventions use mobile phones to deliver interventions and improve health, which are generally prevalent among young people (in the Netherlands, 99.1% uses a mobile phone for internet [11]). They are also widely accessible and can safeguard the users' anonymity [12]. In addition, mHealth enables the effective combination of information and communication technology with behavior change methods [13]. Although mHealth can be useful, usable, and effective, currently available alcohol apps generally contain few behavior change strategies [14]. Empirical evidence is lacking for alcohol prevention mHealth interventions among adolescents and young adults specifically [10]. Research emphasized the need for tailoring [15], personalization [16], and intervening in real-life contexts [10] for this target group.

WDYD Intervention

In this study, we aimed to evaluate an evidence-based 17-week dynamically tailored mHealth intervention, "What Do You Drink" (WDYD; Trial Registration: onderzoekmetmensen.nl Trial 28135 [17,18]). Dynamic tailored interventions, also called just-in-time adaptive interventions [19-21], are interventions where the information about the individual is used to determine when and how to intervene; that is, feedback and support are provided based on prior assessments of behavioral and psychological factors. In other words, to make these intervention-related decisions, not only static information (eg, baseline characteristics), but also time-varying information (eg, mood or response to intervention) is used [20]. Thus, dynamically tailored interventions are adapted over time to the dynamic changes of individuals, based on their varying needs. A meta-analytic review has shown that dynamic, tailored, just-in-time adaptive interventions are more effective than nondynamic, non–just-in-time adaptive interventions [19,21]. Moreover, dynamically tailored interventions should reduce dropout, a general problem of digital health interventions [10].

In addition, WDYD is based on behavior change techniques that have shown to be effective in changing health behavior and reducing alcohol consumption [22-24]. Examples of general methods used in WDYD are goal setting, action planning, behavioral feedback, and self-monitoring. In addition, based on daily assessments of behavioral (alcohol consumption) and psychological factors (motivation, self-confidence, and mood), users were provided with weekly tailored support based on a diversity of behavioral change techniques (eg, planning and problem solving, behavior substitution, modeling, and motivational interviewing techniques [18]).

Objectives

The primary objective was to evaluate the effectiveness of WDYD in reducing excessive alcohol consumption among adolescents and young adults (≥16 y) at risk. Secondary objectives were to assess changes in determinants of drinking behavior (intrinsic motivation, self-confidence, and mood) and to examine the acceptability and use of WDYD.

Hypotheses

We hypothesized that, compared to the control group (no intervention), participants receiving WDYD would (1) show greater reductions in excessive alcohol use, binge drinking, and weekly alcohol consumption (primary outcomes); (2) demonstrate positive changes in determinants of drinking behavior (secondary outcomes); and (3) evaluate the intervention as acceptable and usable, with measurable engagement reflected in app usage logs.

Methods

Ethical Considerations

Human Subjects Ethics Review

This study was approved by the Ethical Committee of the Faculty of Social Sciences of Radboud University Nijmegen (approval number ECSW2016-1403-390).

Informed Consent

Participants provided active informed consent within the app before study enrollment. Individuals younger than 16 years of age were excluded from participation; consent included confirmation that participants were aged 16 years or older (see Multimedia Appendix 1 for consent documents). Participants received a summary of this study's purpose (ie, personalized education about alcohol intake), duration (ie, approximately 6 months), and procedures (including participation via an app, questionnaires, and ecological momentary assessments [EMAs], information about the study groups, and possible interventions). They were also informed that participation was voluntary and that they could withdraw at any time without providing a reason.

XSL•FO
RenderX

Privacy and Confidentiality

Participants were quasi-anonymous. Registration required an email address, which was linked to a unique participant code and stored separately. The Information and Communication Technology provider managing the app had access to both datasets; researchers only accessed deidentified data. Email addresses were shared with researchers solely for reward distribution and could not be linked to participant codes. Data protection measures were outlined in the terms of use and privacy statement included in the consent process (see Multimedia Appendix 1). No cookies or third-party tracking tools were used to prevent multiple registrations, and no additional identity verification measures were implemented.

Compensation

Participants were compensated for their participation. Those who completed both follow-up surveys received a minimum of €25 (US $29.15; €12.50 [US $14.63] per completed survey) and up to €80 (US $93.92) in bol.com vouchers, depending on overall study completion rates. Compensation procedures were communicated during the consent process (see Multimedia Appendix 1).

Design

This study was a 2-arm, parallel-group, online-only randomized controlled trial conducted in the Netherlands between September 2018 and October 2019. The target population was adolescents and young adults aged ≥16 years who drank excessively. Participants could enroll between September 2018 until the end of January 2019 by downloading the stand-alone WDYD app from the iOS App Store (Apple Inc)or Google Play Store (Google LLC). Within the WDYD app, participants were randomly assigned (1:1) to the experimental group (ie, WDYD intervention) or control group (no intervention), using a preprogrammed algorithm embedded in WDYD (see reference for the language code [25]). During consent, participants were informed that 2 study groups existed. Individual allocation was unknown at the time of consent, because randomization occurred after participants provided consent. After randomization, no allocation disclosure was provided. However, participants could infer their group assignment from the presence (intervention) or absence (control) of tailored sessions. The trial lasted 33 weeks and consisted of an online survey at baseline, followed by additional online surveys at 9 and 33 weeks. To account for the fluctuating nature of alcohol drinking and contextual influences [26,27], participants were also asked to report their alcohol intake each day using EMA over 7 weeks (wk 1, 7, 13, 19, 25, 31, and 33), each consisting of 7 consecutive daily assessments. Surveys and EMAs were administered via the WDYD app, with push notifications as reminders.

The trial was conducted with version 2.0.0 of the WDYD app (d.d. August 27, 2018). Two updates were made to WDYD during the trial: (1) small textual changes (September 12, 2018), and (2) a bug fix and removal of text about participation rewards at the end of the enrollment period (January 30, 2019).

Participants

The eligibility criteria were based on the Dutch Health Council guidelines for excessive drinking [9] and being 16 years or older. Adolescents (aged 16 - 18 years) indicating any amount of routine alcohol consumption were considered excessive drinkers. For adults, excessive drinking was defined as:

- Men: drinking more than 2 drinks per day (heavy drinking) or 6 or more glasses in 1 day (binge drinking).
- Women: drinking more than 1 glass per day (heavy drinking) or 4 or more glasses in 1 day (binge drinking).

Participants also had to be 16 years of age or older at the time they signed up.

Initially, WDYD was developed as an intervention with a primary focus on students in secondary vocational education who drank excessively (aged 16 to 24 years). The intervention was developed in cocreation with these students using a user-centered approach [18]. Due to low enrollment in a prior pilot (n=253; 19.3% of the 1310 participants needed), recruitment was expanded to all Dutch adolescents and young adults who drank excessively.

Multiple recruitment channels were used:

1. Students in secondary vocational education were recruited via an online lifestyle monitor (“TestYourLifestyle”; *Testjeleefstijl* in Dutch) specifically designed for schools and students in secondary vocational education [28]. The monitor is used annually by multiple schools as part of their health-education curriculum. The monitor consists of various lifestyle modules (eg, physical activity, nutrition behavior, sexual behavior, or smoking), including alcohol consumption. Within the monitor, students can create a user account, receive module-specific questions followed by brief tailored normative feedback (ie, comparison of their behavior to the Dutch guidelines), behavior-related risk information, and information about websites or programs related to the lifestyle topic. To recruit students in secondary vocational education for this study, students who drank alcohol excessively according to the monitor received an invitation to participate with a link to further information about this study and the possibility to download the WDYD app to subscribe and provide informed consent for participation. During the recruitment period (September 2018 to January 2019), the alcohol module of the lifestyle monitor was completed by 12,004 students. Of these students, 55.1% (6619/12004) exceeded the heavy-drinking guideline, and 52.4% (6290/12004) reported binge drinking, making them eligible for participation in WDYD. The lifestyle monitor is implemented at the school level, but detailed information on the number of participating schools and class-level participation rates was not available.
2. Students in secondary vocational education were also recruited via teachers who implemented an educational program on alcohol consumption. The Secondary Vocational Education Council distributed an email invitation and published a news item on its website aimed at teachers responsible for citizenship education, a compulsory subject in secondary vocational education. Interested teachers could sign up via email to deliver the educational program in their classes. Unfortunately, the number of teachers and schools that received the initial invitation is not known. After completing the educational

XSL•FO
RenderX

program, teachers were asked to invite their students to participate in the WDYD study by providing access to this study's information page and the WDYD app.

3. Via a social media campaign (Instagram or Facebook) targeting people aged 16 - 24 years in the Netherlands.
4. Via news items and flyers by the Addiction Prevention and Care Institute [29].
5. As the eligibility of participants who drank excessively could not be established via the recruitment channels before enrollment, we intentionally oversampled and subsequently excluded individuals who did not meet the inclusion criteria.

Intervention

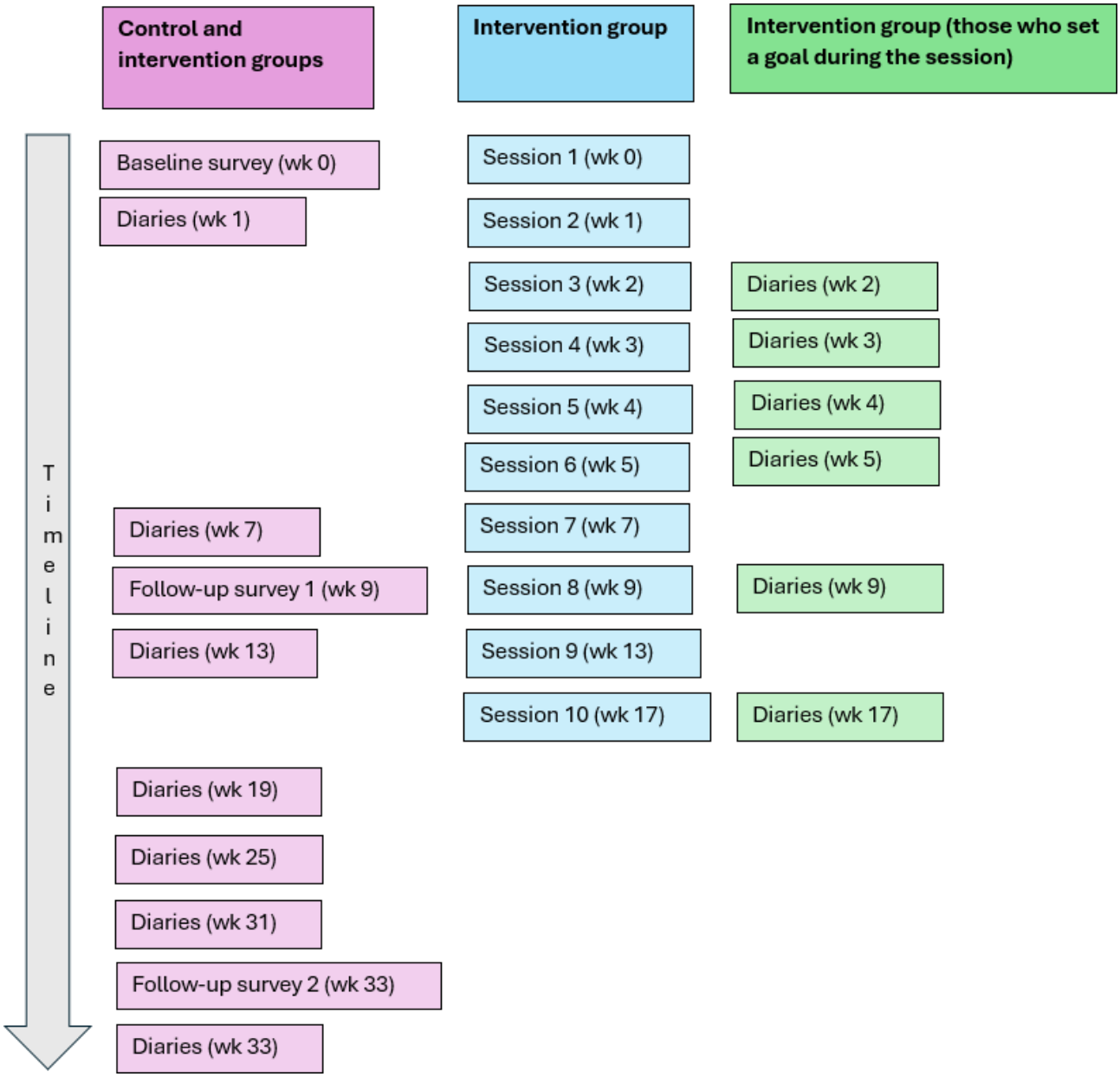
WDYD was developed systematically using the Intervention Mapping planning protocol [30] and an iterative, user-centered design process [31]. The development is described elsewhere [18]. WDYD (version 2.0.0) delivered a 17-week dynamically tailored mHealth program.

Regarding content and access to the intervention, WDYD offered a total of 39 different exercises and 22 videos targeting five key goals: (1) motivation, (2) self-confidence, (3) mood, (4) planning, and (5) relapse prevention, using evidence-based techniques (eg, motivational interviewing [32], behavioral modeling [33], goal setting [34], self-monitoring [35], implementation intentions [36], cognitive-behavior therapy [37], and provision of rewards [38]) and engagement strategies [22,23,39]. In addition, participants could monitor their goal progress in WDYD, in which historic daily alcohol intake was registered, as well as whether they had reached their self-set goal. Screenshots of the intervention are shown in Multimedia Appendix 2.

Exercises were tailored based on EMA data. Participants who set a goal were asked to daily monitor their alcohol drinking behavior, mood, motivation, and self-confidence. They received push notifications to remind them to provide a daily report. The intervention group received tailored weekly sessions and reminders; this started with 6 weekly sessions (wk 0 - 5), after which the frequency of sessions reduced to 2 biweekly sessions (wk 7 and 9), and 2 monthly sessions (wk 13 and 17). See Figure 1 for an overview of the intervention content and timing.

XSL•FO
RenderX

Figure 1. Flow of the WDYD intervention content and timing. WDYD: What Do You Drink.



Control Group

The control group completed surveys (baseline, wk 9, and wk 33) and EMAs only (7 daily assessments during wk 1, 7, 13, 19, 25, 31, and 33), without receiving any feedback or support. They reported their day-to-day alcohol consumption and mood in the EMAs. Motivation and self-confidence to drink less were only asked on the first day of each EMA week.

The control group was also invited to fill in the online baseline survey, and the first and second online follow-up surveys (wk 9 and 33).

Procedure

Participants were invited to participate in this study via various recruitment strategies (see Participants section) and invited to download the WDYD app; access was open and free of charge. Participants who downloaded the app could sign up and provide informed consent in the app. After consent, participants completed a baseline survey measuring sociodemographic variables, and primary and secondary outcomes. Participants in the experimental group received tailored feedback sessions. Invitations for follow-up surveys at weeks 9 and 33, with an estimated completion time of, respectively, 10 and 5 minutes, were sent via push notifications at 7 PM, followed by up to 3 reminders (ie, at d 1, 4, and 7). At the start of a survey, participants were informed about the expected duration of each survey. Participants were given 30 days to complete the follow-up surveys.

Outcome Variables

Outcomes were measured via closed online surveys at baseline, 9, and 33 weeks, and via EMAs consisting of brief daily assessments for 7 consecutive days (wk 1, 7, 13, 19, 25, 31, and 33). All measurements were administered within the WDYD app and pretested for usability and technical functionality.

XSL•FO
RenderX

Primary Outcomes

Weekly alcohol consumption was measured as the number of standard drinks per week, assessed at baseline using daily intake questions [40,41] and during EMA bursts using the question: "How many glasses of alcoholic drinks did you drink yesterday?" (range: 0 - 100) [40]. Weekly alcohol intake was calculated based on a minimum of 4 EMAs in a given week.

Binge drinking was defined as drinking more than 3 (women) or more than 5 (men) standard alcohol units on at least 1 day of the week; the day with the highest reported number of glasses consumed was used to determine whether binge drinking occurred.

Excessive drinking was defined as binge drinking (ie, drinking 4 [women] or 6 [men] or more glasses in 1 day) and/or heavy drinking (ie, drinking any amount of alcohol when under the legal drinking age of 18 years, and for adults to drink a maximum of 1 [women] or 2 [men] glasses of standard alcohol units on an average day [9]). Heavy drinking was determined by comparing the extrapolated number of glasses of alcoholic drinks in a week (more than 7 glasses for women, and more than 14 glasses for men).

Secondary Outcomes

Intrinsic motivation was measured by the question: "How important is it for you to drink less alcoholic drinks" (0=not at all important to 10=very important) [32].

Self-confidence was assessed with the question: "How confident are you that you can drink less alcoholic drinks" (0=not at all confident to 10=very confident) [41,42].

Mood was assessed with the question: "How do you feel?" (0=very bad to 10=very good [41,43,44]. These were measured at baseline (except for mood), week 1 (except for motivation and confidence), and weeks 7, 13, 19, 25, 31, and 33. For mood, the weekly average was calculated if participants filled out a minimum of 4 daily assessments.

Motivation and self-confidence were measured on the first day of each EMA week. Mood was measured daily in an EMA week, and an average weekly mood index was created based on a minimum of 4 EMA entries.

Sociodemographic Variables

Sociodemographic variables were measured at baseline. These variables were age, gender, migrant background, ongoing educational level, and highest completed educational level. Both education variables were rescored to low (less than secondary or vocational education), intermediate (secondary through preuniversity education) or high (professional or university education). Specifically for students in secondary vocational education, we assessed their qualification level (1-4), school year (1-4), and pathway (practical, vocational training, or combination).

Program Evaluation and Use

Program Evaluation

Participants in the intervention group were asked to evaluate WDYD in the first follow-up survey (week 9).

WDYD usability was assessed by 5 items, for example, "How do you perceive the user-friendliness of WDYD?" on a 10-point scale (eg, 1=not at all user-friendly to 10=very user-friendly),

Information quality was also assessed by 5 items (eg, "How do you perceive the credibility of the information in WDYD?" with 10-point scales (eg, 1=not at all credible to 10=very credible).

Personalization was assessed by 8 items (eg, "WDYD took into account my personal preferences," scales 1=totally agree to 5=totally disagree).

Finally, participants were asked to grade the overall acceptability of WDYD on a scale from 1 (very bad) to 10 (excellent) [45].

Program Use

For participants in the intervention group, use of WDYD was logged in the app. Session completion was calculated as the average number of sessions completed across all participants in the intervention group.

Exercises completed were calculated by dividing the number of completed exercises within a given category by the total number of exercises participants had chosen.

Power Calculation

The sample size was based on a priori power calculation for detecting a reduction in excessive drinking in the intervention group compared to the control group. We anticipated a small effect size (Cohen d=0.20) based on a previous study of an earlier version of WDYD [27] and a just-in-time adaptive alcohol intervention [39]. In the previous study, the EMA procedure, including the rewarding of participants, resulted in a 20% dropout [27]. When anticipating a conservative dropout rate of 40% [46], a small effect size (d=0.20), a power of 0.80, and a 2-sided α of .05, a total of 655 participants were required per group at baseline.

Statistical Analyses

Data were analyzed using R (R Foundation) [47]. Logistic regression analyses were also used to analyze the dropout probability at the start (ie, within the first week), and throughout this study (ie, within 33 wk), with primary, secondary, and sociodemographic variables at baseline as predictors. Descriptive statistics (%, n, mean, SD, median, or IQR) were used to analyze baseline characteristics of the sample (ie, sociodemographic and outcome variables) and to evaluate usability and acceptability of WDYD. A logistic regression analysis was used to evaluate the randomization by analyzing the probability of intervention group assignment using the primary, secondary, and sociodemographic variables observed at baseline. Baseline logistic regression analyses were used to examine differences between recruitment strategies in drinking behavior at baseline.

Effects of WDYD on primary and secondary outcomes were analyzed using multilevel generalized linear regression models using the R package *lme4* [48]. A multilevel logistic regression model was used for dichotomous outcomes (ie, excessive drinking and binge drinking), a multilevel negative binomial regression model for count outcomes (ie, number of glasses), and a linear multilevel regression model for continuous

XSL•FO
RenderX

outcomes (ie, mood, motivation, and self-confidence). In these models, time (categorical: a dummy for every diary week with baseline as reference category), group (intervention vs control), and the interaction between time and group were used as predictors. In addition, a random intercept was included in each model to account for within-person clustering over time.

Missing values were imputed using multiple imputation with multilevel predictive mean matching using 100 imputations and 25 iterations, using the R package *mice* [49]. The imputation model contained the primary and secondary outcomes (alcohol consumption, mood, intrinsic motivation, and self-confidence) from the daily assessment data, group, enrolled in secondary vocational education, time, number of sessions attended (0 - 10), gender, and intention to reduce excessive drinking at baseline and week 9. We performed additional complete case analyses on both primary and secondary outcomes to check for differences between analyses based on multiple imputation and those based on complete cases only.

Results

Response Rates and Dropout

The flow of participants through this study is shown in Figure 2. A total of 6146 people downloaded the WDYD app between September 1, 2018, and January 31, 2019, and agreed to participate in this study. Most participants (4168/6146, 68.1%) indicated having been recruited via Facebook (Meta) or Instagram (Meta). Other recruitment sources included referrals from friends or relatives (637/6146, 10.4%), nonsocial media websites (465/6146, 7.6%), the online lifestyle monitor (327/6146, 5.3%), and through their teacher (272/6146, 4.4%). Finally, 4.6% (280/6146) participants found this study through another channel.

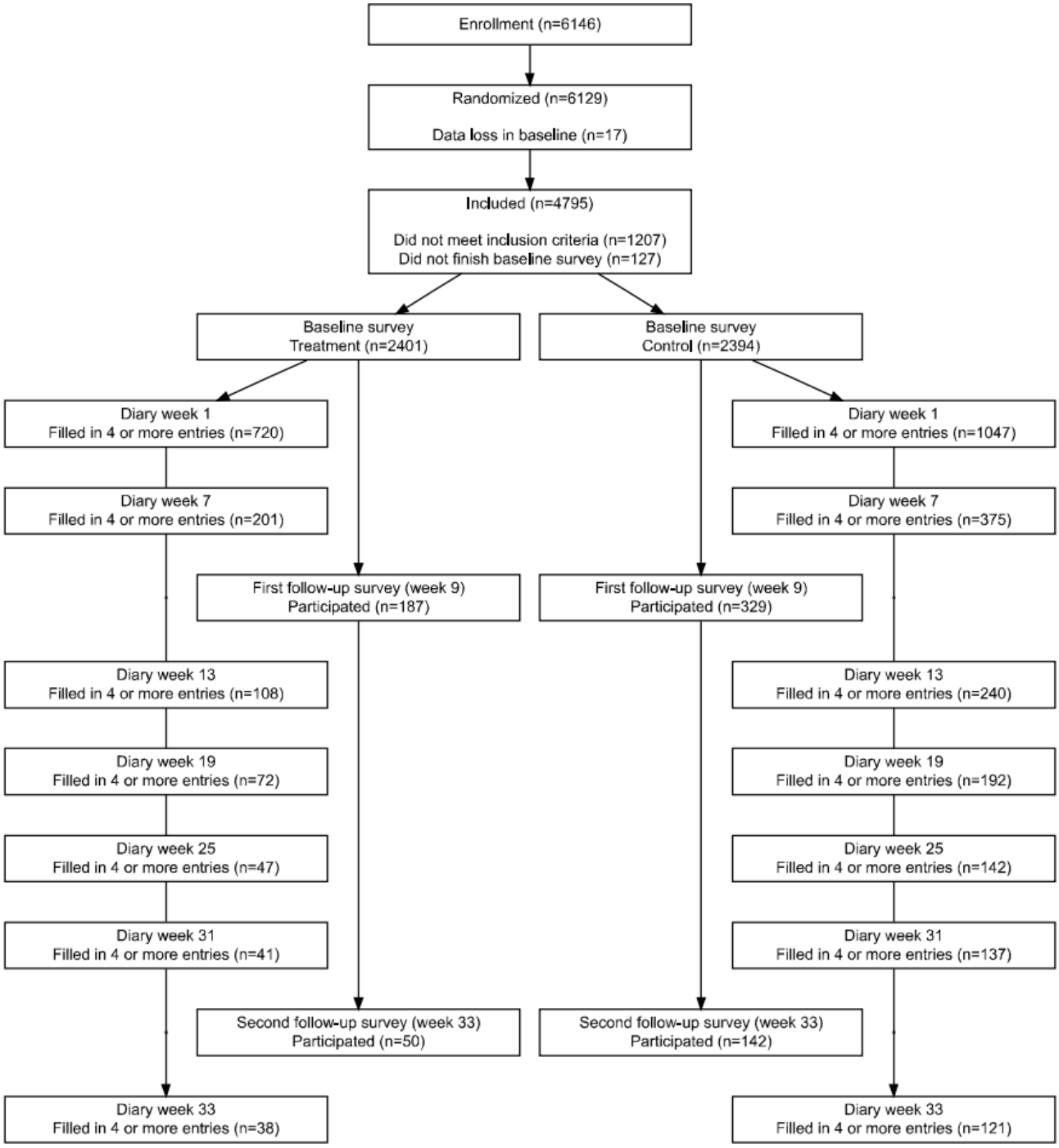
We were unable to determine the group allocation of 17 participants due to data loss of group allocation at baseline. The remaining 6129 participants were randomized into the intervention (n=3088) or control group (n=3041). After filling in the baseline survey, 1334 participants did not meet the inclusion criteria and were excluded from further analysis. Of the participants who met the inclusion criteria, 36.9% (1767/4795) completed the first week of daily assessments, 10.8% (516/4795) completed the first follow-up survey (wk 9), and 4% (192/4795) completed the second follow-up survey (wk 33). Participants who were enrolled in secondary vocational education (exp(β)=2.034, *P*<.001) and had a higher peak daily alcohol consumption (exp(β)=1.040, *P*<.001) were more likely to drop out within the first week of this study. Over the entire 33 weeks of this study, participants assigned to the intervention group (exp(β)=2.96, *P*<.001), enrolled in secondary vocational education (exp(β)=3.59, *P*<.001), and had a higher mean weekly alcohol consumption (exp(β)=1.034, *P*<.001) were more likely to drop out.

Participants were excluded from all analyses when they did not meet the inclusion criteria (ie, excessive drinking and ≥16 y of age; see Participants section; n=1207) or if they did not complete the baseline questionnaire (n=127). The number of participants that dropped out of this study was large, especially in the first diary week (n=3028). We excluded participants who dropped out within the first diary week from the main analysis, as not enough information was observed to allow for imputation. In addition, the number of participants who completed the measurements during weeks 31 and 33 was insufficient for reliable imputation. As a result, our imputation model was unable to produce stable imputed outcomes. For this reason, we restricted our analysis to data collected up to week 25.

The final sample for the main analyses consisted of 1767 participants ($n_{intervention}$=720, $n_{control}$=1047). Missing values of these participants were imputed for analyses. Trace plots showed that the imputed datasets had converged, and density and scatterplots between observed and imputed data showed plausible imputed values.

XSL•FO
RenderX

Figure 2. Flow of participants in the WDYD study. WDYD: What Do You Drink.



Baseline Characteristics

Baseline characteristics of participants are depicted in Table 1. There were no significant differences between participants in the intervention and control groups regarding sociodemographic characteristics, primary and secondary outcomes (data not shown). Small differences were found in drinking behavior by recruitment strategy. Participants reporting higher weekly alcohol consumption were more likely to be recruited via social media (exp(β)=1.01, *P*<.001) and less likely via teachers (exp(β)=0.95, *P*<.001) or the lifestyle monitor (exp(β)=0.92, *P*<.001). No other significant baseline differences were observed in drinking behavior, and no interaction effects of recruitment strategy with group assignment were found. Slightly more than half of the participants were male (2519/4795, 52.5%), 72.4% (3471/4795) participants were aged 18 to 24 years, most of them attended or had completed a higher educational level (2673/4795, 56.3%), and 34.6% (1659/4795) of them attended secondary vocational education.

XSL•FO
RenderX

Table . Baseline characteristics of WDYD[a] study participants.

	Control group (n=2394)	Intervention group (n=2401)	Total (n=4795)
Sociodemographic background			
Sex, n (%)			
Male	1248 (52.1)	1271 (52.9)	2519 (52.5)
Female	1146 (47.9)	1130 (47.1)	2276 (47.5)
Age (years), median (IQR)	19.36 (2.95)	19.42 (2.90)	19.40 (2.92)
<18, n (%)	574 (24)	563 (23.5)	1137 (23.7)
18 - 24, n (%)	1732 (72.4)	1739 (72.4)	3471 (72.4)
≥24, n (%)	88 (3.7)	99 (4.12)	187 (3.9)
Education, n (%)			
Low	158 (6.67)	126 (5.3)	284 (6)
Medium	880 (37.2)	909 (38.2)	1789 (37.7)
High	1330 (56.2)	1343 (56.5)	2673 (56.3)
Student in secondary vocational education, n (%)	824 (34.4)	835 (34.8)	1659 (34.6)
Secondary vocational education school year, n (%)			
1	322 (39.1)	339 (40.6)	661 (39.8)
2	273 (33.1)	275 (32.9)	548 (33)
3	175 (21.2)	165 (19.8)	340 (20.5)
4	54 (6.55)	56 (6.71)	110 (6.6)
Secondary vocational education qualification level, n (%)			
1	8 (0.97)	3 (0.44)	11 (0.66)
2	83 (10.07)	82 (9.82)	165 (10)
3	169 (20.5)	184 (22)	353 (21.3)
4	564 (68.5)	566 (67.8)	1130 (68.1)
Secondary vocational education pathway, n (%)			
Practical training	229 (27.8)	225 (27)	454 (27.4)
Vocational training	571 (69.3)	580 (69.5)	1151 (69.38)
Combination	24 (2.91)	30 (3.59)	54 (3.25)
Migrant background, n (%)			
No	2192 (91.6)	2189 (91.2)	1615 (91.4)
Yes	202 (8.44)	212 (8.83)	414 (8.63)
Primary outcomes			
Excessive drinking, n (%)[b]			
No	0 (0)	0 (0)	0 (0)
Yes	2394 (100)	2401 (100)	4795 (100)
Number of glasses of alcoholic drinks consumed in a week, median (IQR)	16.00 (16.00)	16.00 (15.00)	16.00 (16.00)
Binge drinking, n (%)			
No	655 (27.4)	680 (28.3)	1335 (27.8)
Yes	1739 (72.6)	1721 (71.7)	3460 (72.2)

XSL•FO
RenderX

	Control group (n=2394)	Intervention group (n=2401)	Total (n=4795)
Binge drinking frequency per week, mean (SD)	1.67 (1.11)	1.66 (1.12)	1.67 (1.12)
Peak daily consumption, median (IQR)	6.00 (3.00)	6.00 (3.00)	6.00 (3.00)
Secondary outcomes			
Motivation to drink less (0 - 10), median (IQR)	3.00 (5.00)	3.00 (5.00)	3.00 (5.00)
Self-confidence to drink less (0 - 10), mean (SD)	4.52 (2.86)	4.51 (2.94)	4.52 (2.90)

[a]WDYD: What Do You Drink.

[b]This variable was the inclusion criterion.

Intervention Effects

Effects of WDYD on primary and secondary outcomes based on multiple imputation are depicted in Table 2 and Figure 3. In both groups, binge drinking and mean weekly alcohol consumption significantly increased from baseline to week 1. After week 1, both groups decreased their alcohol consumption, but this did not significantly differ from baseline. There were no significant differences between groups on the primary outcomes at different time points (wk 1, 7, 13, 19, and 25), with the exception of a positive effect in favor of the intervention group for binge drinking and excessive drinking at week 1. Participants in the intervention group were less likely to drink excessively or binge compared to the control group in week 1. At later time points (wk 7, 13, 19, and 25), we do not have enough evidence to suggest that both groups differ. Figure 3 shows that there is a large amount of uncertainty in our model parameters, especially at later points in time when less data was available.

With regard to the secondary outcomes, motivation and self-confidence increased in both groups from weeks 7 to 25. Positive effects of WDYD were found for motivation and self-confidence at weeks 7, 13, 19, and 25; participants in the intervention group had significantly increased their motivation and self-confidence to consume less alcohol than those in the control group. As for mood, participants in the intervention group reported a more positive outcome in week 1 than participants in the control group; this difference was small but significant. After week 1, the difference between groups was no longer significant. Week 3 shows similar uncertainty in estimation for motivation, self-confidence, and mood.

We did conduct additional moderation analyses and found no evidence of educational attainment moderating the treatment effect (data not shown).

Effects on the primary and secondary outcomes were additionally evaluated using complete case analysis; results (data not shown) were similar to those based on multiple imputation of missing values.

XSL•FO
RenderX

Table . Effects of WDYD[a] on primary and secondary outcomes based on multiple imputation with sample statistics based on unimputed data.

Time	Sample statistics						Model coefficients (group×time)		Effect sizes
	$N_{control}$	n (%)	Mean (SD)	$N_{intervention}$	n (%)	Mean (SD)	$B_{intervention}$ (SE)	95% CI	OR[b]/IRR[c]/ StdB[d]
Excessive drinking[e]									
Baseline	N/A[f]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Week 1	1047	877 (83.7)	N/A[g]	720	568 (78.9)	N/A[g]	−0.35 (0.15)	−0.64 to −0.05[e]	0.70[b]
Week 7	297	220 (74.1)	N/A[g]	198	146 (73.7)	N/A[g]	0.34 (0.24)	−0.13 to 0.80	0.99[b]
Week 13	192	135 (70.3)	N/A[g]	106	76 (71.7)	N/A[g]	0.43 (0.24)	−0.05 to 0.90	1.08[b]
Week 19	157	111 (70.7)	N/A[g]	70	51 (72.86)	N/A[g]	0.45 (0.26)	−0.06 to 0.96	1.11[b]
Week 25	106	81 (76.42)	N/A[g]	47	32 (68.09)	N/A[g]	0.32 (0.26)	−0.20 to 0.83	0.97[b]
Binge drinking[e]									
Baseline	1047	734 (70.1)	N/A[g]	720	510 (70.8)	N/A[g]	0.04 (0.11)	−0.18 to 0.25	1.04[b]
Week 1	1047	863 (82.4)	N/A[g]	720	558 (77.5)	N/A[g]	−0.36 (0.16)	−0.68 to −0.04[e]	0.73[b]
Week 7	297	214 (72.1)	N/A[g]	198	144 (72.7)	N/A[g]	−0.01 (0.19)	−0.38 to 0.35	1.03[b]
Week 13	192	133 (69.3)	N/A[g]	106	76 (71.7)	N/A[g]	0.00 (0.19)	−0.38 to 0.38	1.04[b]
Week 19	157	110 (70.1)	N/A[g]	70	49 (70)	N/A[g]	−0.02 (0.19)	−0.39 to 0.36	1.02[b]
Week 25	106	78 (73.58)	N/A[g]	47	32 (68.09)	N/A[g]	−0.02 (0.20)	−0.42 to 0.38	1.01[b]
Number of glasses of alcoholic drinks per week[e]									
Baseline	1047	N/A[g]	16 (15)	720	N/A[g]	17 (16)	0.00 (0.05)	−0.09 to 0.09	1.00[c]
Week 1	1047	N/A[g]	17 (22.5)	720	N/A[g]	16 (22)	−0.07 (0.05)	−0.16 to 0.02	0.93[c]
Week 7	297	N/A[g]	15	198	N/A[g]	12.5 (20)	−0.05 (0.05)	−0.16 to 0.06	0.95[c]
Week 13	192	N/A[g]	11.5 (24)	106	N/A[g]	13 (19.75)	−0.04 (0.05)	−0.14 to 0.07	0.96[c]
Week 19	157	N/A[g]	11 (22)	70	N/A[g]	10 (15)	−0.03 (0.05)	−0.13 to 0.08	0.97[c]
Week 25	106	N/A[g]	13.5 (19.5)	47	N/A[g]	10 (16.5)	−0.03 (0.06)	−0.15 to 0.08	0.96[c]
Mood[e]									
Baseline[h]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Week 1	973	N/A[g]	6.59 (1.27)	691	N/A[g]	6.81 (1.34)	0.20 (0.06)	0.08 to 0.32[e]	0.19[d]
Week 7	309	N/A[g]	6.63 (1.11)	193	N/A[g]	6.94 (1.36)	−0.12 (0.10)	−0.32 to 0.08	0.08[d]

XSL•FO
RenderX

Time	Sample statistics						Model coefficients (group×time)		Effect sizes
	$N_{control}$	n (%)	Mean (SD)	$N_{intervention}$	n (%)	Mean (SD)	$B_{intervention}$ (SE)	95% CI	OR[b]/IRR[c]/ StdB[d]
Week 13	196	N/A[g]	6.68 (1.11)	105	N/A[g]	6.97 (1.27)	−0.15 (0.10)	−0.35 to 0.05	0.05[d]
Week 19	158	N/A[g]	6.62 (1.22)	71	N/A[g]	7.01 (1.37)	−0.15 (0.11)	−0.36 to 0.06	0.04[d]
Week 25	112	N/A[g]	6.65 (1.14)	46	N/A[g]	6.80 (1.44)	−0.18 (0.1)	−0.38 to 0.02	0.02[d]
Self-confidence[i]									
Baseline	1046	N/A[g]	4.76 (2.80)	719	N/A[g]	4.62 (2.90)	−0.13 (0.15)	−0.42 to 0.15	-0.05[d]
Week 1	N/A[j]	N/A[j]	N/A[j]	324	N/A[g]	7.65 (2.96)	N/A[j]	N/A[j]	N/A[j]
Week 7	164	N/A[g]	5.23 (2.86)	185	N/A[g]	6.95 (2.93)	0.91 (0.26)	0.40 to 1.41[e]	0.22[d]
Week 13	111	N/A[g]	5.18 (2.90)	91	N/A[g]	7.02 (2.96)	0.75 (0.26)	0.25 to 1.25[e]	0.18[d]
Week 19	75	N/A[g]	5.04 (3.00)	73	N/A[g]	6.88 (3.20)	0.72 (0.27)	0.18 to 1.25[e]	0.16[d]
Week 25	50	N/A[g]	5.02 (3.06)	44	N/A[g]	7.32 (3.00)	0.72 (0.26)	0.22 to 1.23[e]	0.15[d]
Motivation[i]									
Baseline	1046	N/A[g]	3.78 (2.74)	718	N/A[g]	3.79 (2.85)	0.01 (0.14)	−0.27 to 0.3	0.00[d]
Week 1	N/A[j]	N/A[j]	N/A[j]	324	N/A[g]	6.63 (2.93)	N/A[j]	N/A[j]	N/A[j]
Week 7	166	N/A[g]	4.40 (2.74)	185	N/A[g]	6.38 (2.87)	0.81 (0.25)	0.31 to 1.31[e]	0.22[d]
Week 13	114	N/A[g]	4.51 (3.02)	92	N/A[g]	6.66 (3.00)	0.60 (0.24)	0.13 to 1.07[e]	0.18[d]
Week 19	76	N/A[g]	4.30 (2.70)	73	N/A[g]	6.58 (2.89)	0.56 (0.24)	0.08 to 1.04[e]	0.17[d]
Week 25	50	N/A[g]	3.96 (2.94)	44	N/A[g]	6.91 (2.84)	0.54 (0.24)	0.06 to 1.02[e]	0.15[d]

[a]WDYD: What Do You Drink.

[b]OR: odds ratio for excessive drinking and binge drinking.

[c]IRR: incidence rate ratio for the number of glasses per week.

[d]StdB: standardized β for mood, self-confidence, and motivation.

[e]Significant difference between intervention and control condition (P<.05).

[f]N/A: not applicable.

[g]Mean (SD) was not applicable for the dichotomous outcomes excessive drinking and binge drinking, and n (%) was not applicable for the continuous outcomes of number of glasses of alcoholic drinks per week, mood, self-confidence, and motivation.

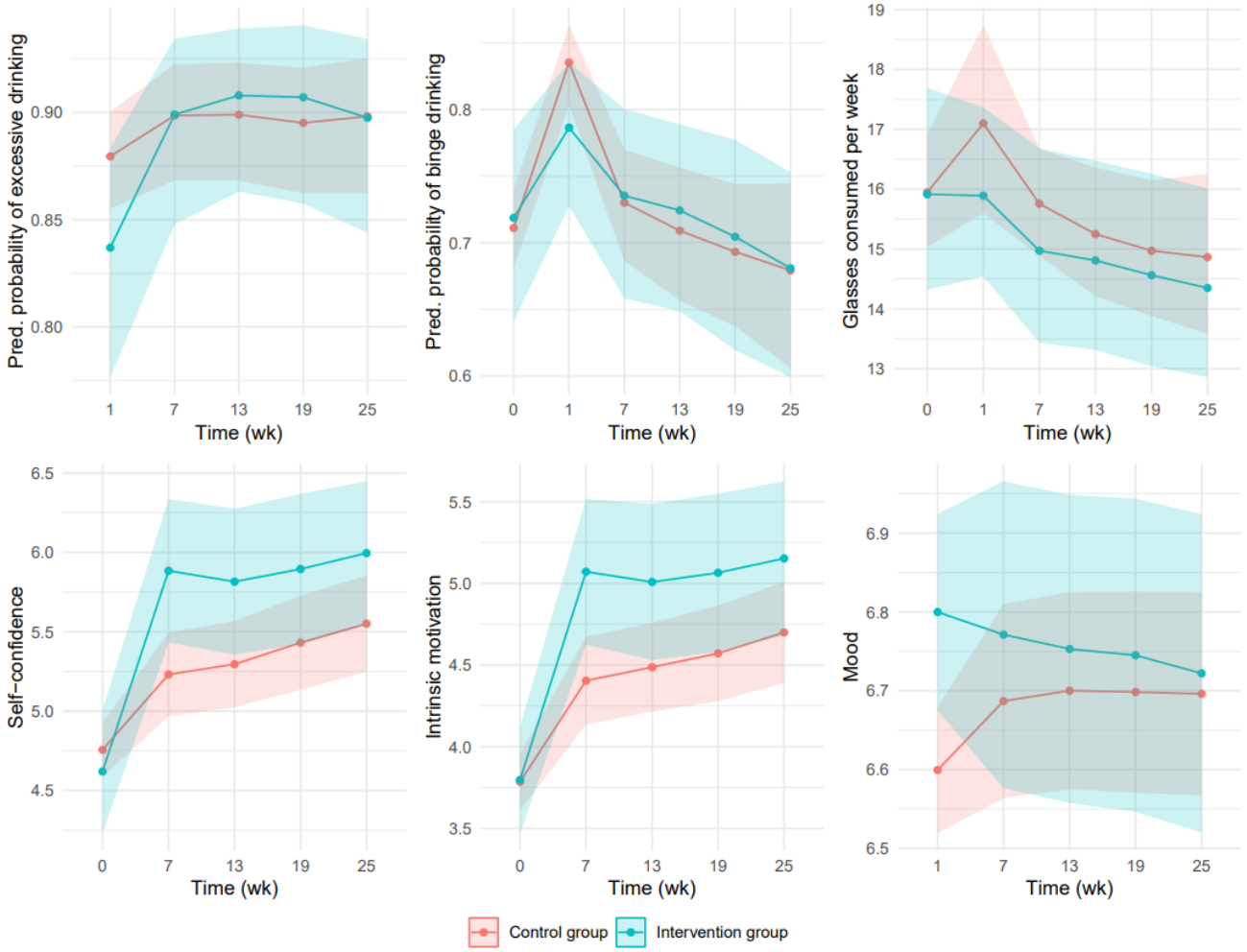
[h]Baseline was not included in the analyses because all participants engaged in excessive drinking at baseline.

[i]Mood was not measured at baseline, and had week 1 as the reference category for time.

[j]Self-confidence and motivation were not measured at week 1 in the control group.

XSL•FO
RenderX

Figure 3. Predicted outcomes of WDYD for the control and intervention groups according to the multilevel model based on multiple imputation. Bands indicate the 95% pointwise CIs for the control and intervention groups. The CI for the intervention group has the coefficient of time and group fixed at its mean predictor values. Pred: predicted; WDYD: What Do You Drink.



Program Evaluation and Use

The program evaluation and use of WDYD among intervention participants are depicted in Table 3. Participants were very positive about WDYD; they evaluated the overall acceptability of WDYD with a mean grade of 7.17 (SD 1.07; 0=very bad to 10=excellent). The usability of WDYD was evaluated as good (with mean grades of 6.78, SD 1.73, or higher at a 10-point scale). The informativity and usefulness of WDYD were evaluated as satisfactory (means 5.69, SD 2.04, and 5.46, SD 2.05, respectively), and the credibility and understandability as good (means 7.25, SD 1.41, and 7.97, SD 1.44, respectively). The amount of information was perceived as adequate (mean 5.78, SD on a scale from 0=too little to 10=too much). In addition, participants positively evaluated the design principles of WDYD (eg, acceptation and freedom of choice).

The average number of sessions followed reflects the high dropout rate (mean 1.67, SD 2.12). Within the intervention sessions, most participants received a motivation exercise (1475/4298, 34.3%) or a planning exercise (1465/4298, 34.1%).

XSL•FO
RenderX

Table . Program evaluation and use of WDYD[a]. A higher value is always better unless indicated otherwise.

	Values
Program evaluation (n=187)	
Overall acceptability (0=very bad to 10=excellent), mean (SD)	7.17 (1.07)
Usability (0=not at all or very bad to 10=excellent), mean (SD)	
User-friendliness	7.67 (1.59)
Fastness	8.51 (1.46)
Navigation ease	7.99 (1.55)
Design	7.43 (1.48)
Fun of use	6.78 (1.73)
Information in WDYD (0=not at all/very bad to 10=excellent), mean (SD)	
Informative	5.69 (2.04)
Usefulness	5.46 (2.05)
Credibility	7.25 (1.41)
Understandability	7.97 (1.44)
Amount of information[b]	5.78 (1.77)
Design principles (1=totally disagree to 5=totally agree), mean (SD)	
Free to choose my alcohol consumption	4.30 (0.75)
Learned to set goals to reduce drinking	3.33 (0.95)
Assumed I knew what was best for myself	3.74 (0.91)
Discussed topics I find important	3.63 (0.87)
Addressed in a pleasant way	4.17 (0.66)
Accounted for my personal preferences	3.71 (0.90)
Accepted me as I am	4.01 (0.80)
Program use (n=2401)	
Average number of sessions followed, mean (SD)	1.67 (2.12)
Exercises received (n and % of exercises)	
Mood exercise	807 (18.8)
Motivation exercise	1475 (34.3)
Self-confidence exercise	530 (12.3)
Planning exercise	1465 (34.1)
Relapse exercise	0.49 (21)

[a]WDYD: What Do You Drink.

[b]0=too little, 10=too much, with 5 as the center.

Discussion

Principal Results

This study evaluated the effect, use, and acceptability of a dynamically tailored mobile intervention, WDYD, to reduce excessive alcohol drinking among at-risk adolescents and young adults. Regarding the primary outcomes, we found no support for our hypothesis that WDYD was more effective than the control group in reducing alcohol consumption. No effect of WDYD was found on mean weekly alcohol consumption. Positive effects of WDYD were found only in the first week for excessive drinking. In addition, there was an increase in binge drinking in the first week in both groups, but this increase was significantly smaller in the intervention group. After this first week, alcohol consumption seemed to decrease over time in both groups, but differences between groups were nonsignificant at all time points. The increase in binge drinking in the first week was possibly due to the measurement difference between baseline and the following weeks. At baseline, alcohol consumption was based on a participant's self-reported alcohol intake during a typical week, whereas at the follow-up measurements, participants were asked to report their consumption of the day before (1-day recall) for 7 consecutive days. A 1-day recall leads to more accurate estimates than a

XSL•FO
RenderX

7-day recall because it is less susceptible to recall bias [50], and this difference is even more pronounced with binge drinking [40].

The reduction in alcohol consumption over time in both groups in this study is in contrast with the study about a previous web-based version of WDYD [27], which showed that both groups increased their alcohol consumption over time, with the intervention group showing a significantly lower increase than the control group. Differences between this study and the previous study may partly explain the variation of observed effects. The previous study used a web-based brief, single-session tailored feedback intervention. Effects were measured during 30 weekly 7-day recall EMAs over 6 months among a highly educated sample of young adults (ie, students in higher professional education or university) recruited via flyers and included when they were ready to change their alcohol consumption. In this study, the reduction of alcohol consumption in both groups over time, as well as the absence of a significant intervention effect on alcohol consumption, could be explained by the use of daily 1-day recall EMAs; the study's participants, including those in the control group, self-monitored their alcohol drinking behavior daily for 7 weeks during this study. Self-monitoring is shown to be an important behavior change technique in predicting positive outcomes regarding alcohol reduction research [14,23]. Therefore, it is highly likely that the difference in outcomes between groups would have been greater if the control group had engaged in less frequent self-monitoring. Moreover, a difference between the studies lies in the inclusion criteria. First, the 2 studies targeted different populations. The previous study focused on highly educated young adults (ie, students in higher professional education and university), whereas this study primarily targeted adolescents and young adults from secondary vocational education and expanded recruitment to a broader group of Dutch adolescents and young adults. These populations differ in educational level, lifestyle context, and potentially in digital literacy, which may influence engagement and responsiveness to tailored feedback. Second, the previous study included participants who were already ready to change their alcohol consumption [27], a factor known to predict intervention responsiveness and behavior change outcomes [51,52]. In contrast, this study did not preselect participants based on their readiness to change, resulting in a more heterogeneous sample in terms of readiness. This combination of a different target group and broader variation in readiness may have reduced intervention effects compared to the earlier trial.

This study shows that there was a large variation between individuals in alcohol use and its determinants over time. This is in line with earlier research [27,53] and supports the dynamically tailored nature of WDYD. At the same time, it emphasizes a need for additional data analyses besides the evaluation of group-level effects. Specifically, an analysis method that accounts for differences between and within individuals over time to learn why and for whom the intervention is effective and to improve future dynamically tailored interventions [54].

An important flaw of this study was the high dropout rates, with 63.1% (3028/4795) within the first week, rising to 96% (4603/4795) at the final week 33, even though we applied several strategies to prevent dropout (eg, rewards, therapeutic alliance, and dynamic tailoring [10,21]). Dropout was higher among students in secondary vocational education and those who consumed more glasses of alcohol weekly. High dropout is a common problem in mHealth interventions, especially in the first 2 weeks of the study and among those most at-risk, and this needs attention (eg, [10,55]). EMA is a promising means to provide dynamic tailoring; however, it also places a cognitive burden upon participants and presents difficulties in keeping participants engaged [21]. Although sensor technology may be a promising strategy to reduce this burden, this is not yet available, reliable, or practical for the outcomes measured in this study [56]. WDYD was a stand-alone intervention and did not use in-person contact. Both in-person contact (eg, with a counselor) or a combined or stepped-care intervention (eg, mHealth combined with an educational program at school) could be promising ways to reduce dropout while also improving effectiveness [39,57,58]. In addition, although WDYD included pictures, videos, and a design that was tailored to accommodate low literacy levels, it relied on text to provide tailored feedback. As health literacy is related to dropout, future mHealth interventions could rely more on audio and video feedback [57]. Additionally, providing tailored information by a conversational agent using generative artificial intelligence (AI) could be a promising way to reduce dropout and increase intervention effects [59,60]. Possibly, the WDYD intervention was too static, thereby limiting engagement. Emerging developments in generative AI offer additional opportunities to personalize interventions at scale. AI-driven conversational agents can deliver dynamic, context-aware feedback, simulate motivational interviewing, and adapt content based on real-time user data. These technologies could enhance engagement, reduce attrition, and provide more nuanced support for behavior change compared to static tailoring. However, ethical considerations such as privacy, transparency, and bias mitigation must be addressed before widespread implementation in mHealth interventions.

Despite not finding significant effects on the primary outcomes, and dropout being a serious problem, WDYD could be seen as a promising intervention. This is especially true when taking into account that young and lower-educated adolescents and young adults are a difficult group to reach, as the majority (90%) lack motivation to reduce their alcohol consumption [27]. First, the combination of recruitment strategies used in this study showed that we are able to reach young adults who drink excessively; the use of social media advertisements resulted in the largest number of participants. Second, we found positive effects of WDYD on the secondary outcomes, intrinsic motivation, and self-confidence toward drinking less alcohol. Both outcomes significantly increased in the intervention group compared to the control group, and this effect lasted over 25 weeks. Third, additional analyses revealed that increases in intrinsic motivation and self-confidence among active users of WDYD were significantly associated with reductions in alcohol consumption (data not shown). WDYD can thus be an effective means to improve motivation and self-confidence toward reducing alcohol consumption, and this increase seems to have promising effects on alcohol consumption itself. In addition,

XSL•FO
RenderX

the user-based assessments revealed that the intervention was evaluated very positively. WDYD was evaluated as usable (eg, user-friendly and easy in navigation), credible, understandable, and provided an adequate amount of information. The informativity and usefulness were evaluated as sufficient. Dynamic tailoring was successfully implemented in the intervention design, as participants indicated that WDYD accounted for their personal preferences, discussed topics they found important, and provided freedom in choosing their alcohol consumption. This was also the case for motivational interviewing because participants felt accepted for who they are and felt addressed pleasantly.

A small proportion of participants were recruited via channels that may have provided prior exposure to alcohol-related information, such as the online lifestyle monitor (327/6146, 5.3%) or educational programs delivered by teachers (272/6146, 4.4%). Although this may have reduced contrast between intervention and control conditions, moderation analyses showed that the recruitment method did not significantly influence group effects (data not shown).

Limitations

This study had multiple limitations. First and foremost, dropout rates were high, particularly in the intervention group. This selective attrition may have introduced bias. To mitigate this, we applied multiple imputation [61]. However, due to difficulties imputing data from weeks 31 and 33 after baseline, these time points were omitted from the imputation model, and analyses were limited to data up to 25 weeks. Additionally, we did not examine reasons for dropout, which could have provided important information to reduce dropout in the future [57].

Second, although the intervention was originally developed for students in secondary vocational education, recruitment challenges led us to broaden our inclusion criteria to all Dutch adolescents and young adults engaging in excessive drinking. As a result, only 34.6% (1659/4795) of the sample consisted of students in secondary vocational education, and 6% (284/4795) had a low educational level. While moderation analyses showed no significant interaction between educational level and treatment effects (data not shown), caution is warranted when generalizing results to students in secondary vocational education or students from lower-educated populations. Sampling from a more diverse population may have introduced more variability in drinking behavior and decreased receptiveness to our intervention, which could affect the interpretation of our outcomes.

Third, a small proportion of participants (5.3%) were recruited via an online lifestyle monitor or via teachers delivering alcohol education (4.4%). These participants may have already been exposed to normative feedback or alcohol-related risk information before this study. This prior exposure could have reduced the contrast between the intervention and control conditions, potentially reducing observed effects.

Fourth, we used single-item measures to assess the primary and secondary outcomes. Although these items were selected based on prior studies and some of these items showed comparable or adequate performance compared to multi-item scales [42-44], single-item measures are generally more susceptible to measurement error and have lower psychometric quality [62]. In addition, the items used were not validated before our study, which may limit the reliability and generalizability of our findings. However, in studies using EMA, multiple-item scales may enhance reactivity and participant burden [63], justifying our choice for brevity.

Finally, there was a potential risk of contamination between conditions. Although participants were individually randomized into the control or experimental condition within the WDYD app, it is possible that participants in the control group were acquainted with those in the intervention group and discussed intervention content. To assess this, we asked participants in the control group 12 questions about the intervention at the first follow-up (9 weeks after baseline). The answers provided no strong indication of contamination. The distribution of correct answers resembled random guessing, and no participant answered more than 8 of 12 items correctly (data not shown).

Strengths

Strong points of this study are the participatory design, use of intervention mapping as a framework for systematic development, the focus on a high-risk group for whom hardly any interventions are available, the use of EMAs to account for the fluctuating nature of alcohol consumption and to account for recall bias [26,50], and the use of multilevel multiple imputation analyses to account for dropout [61]. The current study adhered to the CONSORT (Consolidated Standards of Reporting Trials) statement [64,65], thereby improving the quality of reporting.

Conclusions

The dynamically tailored mHealth intervention WDYD targeting adolescents and young adults with excessive alcohol consumption did not show favorable changes in alcohol intake, and dropout rates were high. Nevertheless, WDYD might be useful for adolescents and young adults to improve intrinsic motivation and self-confidence to reduce their alcohol consumption as important first steps toward behavior change. As such, the intervention has been implemented in the Netherlands as part of a larger campaign. To reduce dropout rates in mHealth interventions, future research should explore underlying reasons for dropout, as well as using in-person contact, a combined intervention (eg, mHealth with educational program), and audio- or visual-feedback to improve user retention.

Acknowledgments

We would like to thank all the participants and Nova College and Deltion College for their participation in the development of WDYD. The authors used Microsoft Copilot (2025 version; Microsoft Corp) to provide English translations of the references

XSL•FO
RenderX

and study information in Multimedia Appendices 1,2,3, and to make the Method section more compliant with the CONSORT (Consolidated Standards of Reporting Trials) guidelines.

Funding

This study is funded by the Netherlands Organisation for Health Research and Development (ZonMw #531001105). The funder had no involvement in this study's design, data collection, analysis, interpretation or the writing of this paper.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: HvK, CV, MK, PvE
Data curation: AH
Formal analysis: AH
Funding acquisition: HvK, CV, MK, PvE
Investigation: HvK, CV, MK, PvE
Methodology: HvK, CV, MK, PvE
Project administration: HvK, PvE
Resources: HvK, AH, PvE
Supervision: HvK, PvE
Validation: HvK, PvE
Visualization: HvK, AH, PvE
Writing – original draft: HvK, RA, AH, PvE
Writing – review & editing: CV, MK
AH worked on the software.
All authors read and approved this paper for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Research information.
[DOCX File, 1098 KB - mhealth_v14i1e68468_app1.docx]

Multimedia Appendix 2
Screenshots of the WDYD intervention. WDYD: What Do You Drink.
[DOCX File, 1040 KB - mhealth_v14i1e68468_app2.docx]

Multimedia Appendix 3
Recruitment information.
[DOCX File, 571 KB - mhealth_v14i1e68468_app3.docx]

Checklist 1
CONSORT-eHealth checklist WDYD intervention, October 23, 2025. WDYD: What Do You Drink.
[PDF File, 12236 KB - mhealth_v14i1e68468_app4.pdf]

References

1. Global status report on alcohol and health 2018. : World Health Organization; 2019 URL: https://www.who.int/publications/i/item/9789241565639 [accessed 2026-04-29]
2. Overbeek G, Bot SM, Meeus WHJ, Sentse M, Knibbe RA, Engels R. Where it's at! The role of best friends and peer group members in young adults' alcohol use. J Res Adolesc 2011 Sep;21(3):631-638. [doi: 10.1111/j.1532-7795.2010.00696.x]
3. Smit K, Groefsema M, Luijten M, Engels R, Kuntsche E. Drinking motives moderate the effect of the social environment on alcohol use: an event-level study among young adults. J Stud Alcohol Drugs 2015 Nov;76(6):971-980. [doi: 10.15288/jsad.2015.76.971] [Medline: 26562607]

XSL•FO
RenderX

4. Groefsema M, Engels R, Kuntsche E, Smit K, Luijten M. Cognitive biases for social alcohol-related pictures and alcohol use in specific social settings: an event-level study. Alcohol Clin Exp Res 2016 Sep;40(9):2001-2010. [doi: 10.1111/acer.13165] [Medline: 27511292]
5. Dorsselaer S, de Beurs D, Monshouwer K, Middelengebruik onder studenten van 16-18 jaar op het MBO en HBO 2019. Substance use among students aged 16-18 years in secondary vocational education and higher professional education in 2019. : Trimbos-instituut; 2020 URL: https://www.trimbos.nl/docs/b9a089b6-75bb-4864-9274-1ffc29ce1c79.pdf [accessed 2026-05-06]
6. Monshouwer K, v MC, van Beek R, den Hollander W, Schouten F, Blankers M. Uitgaanspatronen, middelengebruik, gezondheid en intentie tot stoppen of minderen onder uitgaande jongeren en jongvolwassenen [Article in Dutch]. : Trimbos-instituut; 2020 URL: https://www.trimbos.nl/kennisbank/rapport-het-grote-uitgaansonderzoek-2023/ [accessed 2026-04-29]
7. Alcoholgebruik onder volwassenen in nederland kerncijfers 2018 [Article in Dutch]. : Trimbos instituut; 2020 URL: https://www.trimbos.nl/wp-content/uploads/sites/31/2021/09/pm0396-alcoholgebruik-in-nederland-kerncijfers-2018.pdf [accessed 2026-04-29]
8. Data on excessive alcohol use. CDC. 2024 Aug 6. URL: https://www.cdc.gov/alcohol/excessive-drinking-data/index.html [accessed 2026-04-29]
9. Richtlijnen goede voeding 2006 [Article in Dutch]. : Gezondheidsraad; 2006 Dec 12 URL: https://www.gezondheidsraad.nl/documenten/2006/12/18/richtlijnen-goede-voeding-2006 [accessed 2026-04-29]
10. Kazemi DM, Borsari B, Levine MJ, Li S, Lamberson KA, Matta LA. A systematic review of the mHealth interventions to prevent alcohol and substance abuse. J Health Commun 2017 May;22(5):413-432. [doi: 10.1080/10810730.2017.1303556] [Medline: 28394729]
11. Internet; toegang, gebruik en faciliteiten; 2012-2019. : Centraal Bureau voor de Statistiek; 2020 Sep 18 URL: https://www.cbs.nl/nl-nl/cijfers/detail/83429NED?dl=35852 [accessed 2026-05-06]
12. Dugas M, Gao GG, Agarwal R. Unpacking mHealth interventions: a systematic review of behavior change techniques used in randomized controlled trials assessing mHealth effectiveness. Digit Health 2020;6:2055207620905411. [doi: 10.1177/2055207620905411] [Medline: 32128233]
13. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 5;16(6):e146. [doi: 10.2196/jmir.3077] [Medline: 24905070]
14. Crane D, Garnett C, Brown J, West R, Michie S. Behavior change techniques in popular alcohol reduction apps: content analysis. J Med Internet Res 2015 May 14;17(5):e118. [doi: 10.2196/jmir.4060] [Medline: 25977135]
15. Gajecki M, Berman AH, Sinadinovic K, Rosendahl I, Andersson C. Mobile phone brief intervention applications for risky alcohol use among university students: a randomized controlled study. Addict Sci Clin Pract 2014 Dec;9(1):11. [doi: 10.1186/1940-0640-9-11]
16. Suffoletto B, Kristan J, Callaway C, et al. A text message alcohol intervention for young adult emergency department patients: a randomized clinical trial. Ann Emerg Med 2014 Dec;64(6):664-672. [doi: 10.1016/j.annemergmed.2014.06.010] [Medline: 25017822]
17. van Keulen H. Trial NL6441 (NTR6619) what do you drink (WDYD)+. World Health Organization. 2017 Aug 16. URL: https://trialsearch.who.int/Trial2.aspx?TrialID=NL-OMON28135 [accessed 2026-05-06]
18. van Keulen H, Voogt C, Kleinjan M, Kramer J, Andree R, van Empelen P. Development of a dynamically tailored mHealth intervention (What Do You Drink) to reduce excessive drinking among Dutch lower-educated students: user-centered design approach. JMIR Form Res 2022 Aug 11;6(8):e36969. [doi: 10.2196/36969] [Medline: 35969428]
19. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010 Sep;51(3-4):214-221. [doi: 10.1016/j.ypmed.2010.06.004]
20. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: a pragmatic framework. Health Psychol 2015 Dec;34S:1209-1219. [doi: 10.1037/hea0000306] [Medline: 26651462]
21. Wang L, Miller LC. Just-in-the-moment adaptive interventions (JITAI): a meta-analytical review. Health Commun 2020 Nov;35(12):1531-1544. [doi: 10.1080/10410236.2019.1652388] [Medline: 31488002]
22. Garnett CV, Crane D, Brown J, et al. Behavior change techniques used in digital behavior change interventions to reduce excessive alcohol consumption: a meta-regression. Ann Behav Med 2018 May 18;52(6):530-543. [doi: 10.1093/abm/kax029] [Medline: 29788261]
23. Michie S, Whittington C, Hamoudi Z, Zarnani F, Tober G, West R. Identification of behaviour change techniques to reduce excessive alcohol consumption. Addiction 2012 Aug;107(8):1431-1440. [doi: 10.1111/j.1360-0443.2012.03845.x] [Medline: 22340523]
24. Kaner EF, Beyer FR, Garnett C, et al. Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. Cochrane Database Syst Rev 2017 Sep 25;9(9):CD011479. [doi: 10.1002/14651858.CD011479.pub2] [Medline: 28944453]

XSL•FO
RenderX

25. RandRange. Adobe. 2025 Mar 25. URL: https://helpx.adobe.com/coldfusion/cfml-reference/coldfusion-functions/functions-m-r/randrange.html [accessed 2026-04-29]
26. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4(1):1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
27. Voogt CV, Kuntsche E, Kleinjan M, Poelen EAP, Lemmers L, Engels R. Using ecological momentary assessment in testing the effectiveness of an alcohol intervention: a two-arm parallel group randomized controlled trial. PLoS ONE 2013;8(11):e78436. [doi: 10.1371/journal.pone.0078436] [Medline: 24223806]
28. Lifestyle reality check [Article in Dutch]. Test Je Leef Stijl. URL: https://www.testjeleefstijl.nl [accessed 2026-04-29]
29. Onderzoek drinkgedrag jongeren via app [Article in Dutch]. Trimbos instituut. 2018 Sep 3. URL: https://www.trimbos.nl/actueel/nieuws/onderzoek-drinkgedrag-jongeren-via-app [accessed 2026-04-29]
30. Bartholomew LK, Markham CM, Ruiter RAC, Fernandez ME, Kok G, Parcel GS. Planning Health Promotion Programs: An Intervention Mapping Approach: John Wiley & Sons; 2016.
31. Kristensson P, Matthing J, Johansson N. Key strategies for the successful involvement of customers in the co‐creation of new technology‐based services. Int J Serv Ind Manag 2008 Aug 8;19(4):474-491. [doi: 10.1108/09564230810891914]
32. Miller WR, Rollnick S. Motivational Interviewing: Helping People Change: Guilford Press; 2013.
33. Kelder SH, Hoelscher D, Perry CL. How individuals, environments, and health behaviors interact. In: Health Behavior: Theory, Research, and Practice: Jossey-Bass/Wiley; 2015:159-181.
34. Latham GP, Locke EA. New developments in and directions for goal-setting research. Eur Psychol 2007 Jan;12(4):290-300. [doi: 10.1027/1016-9040.12.4.290]
35. Harkin B, Webb TL, Chang BPI, et al. Does monitoring goal progress promote goal attainment? A meta-analysis of the experimental evidence. Psychol Bull 2016 Feb;142(2):198-229. [doi: 10.1037/bul0000025] [Medline: 26479070]
36. Gollwitzer PM, Sheeran P. Implementation intentions and goal achievement: a meta‐analysis of effects and processes. Adv Exp Soc Psychol 2006;38:69-119. [doi: 10.1016/S0065-2601(06)38002-1]
37. Beck JS. Cognitive Therapy: Basics and Beyond: Guilford Press; 1995.
38. Bandura A, National Institute of Mental Health. Social Foundations of Thought and Action: A Social Cognitive Theory: Prentice-Hall, Inc; 1986.
39. Brendryen H, Lund IO, Johansen AB, Riksheim M, Nesvåg S, Duckert F. Balance--a pragmatic randomized controlled trial of an online intensive self-help alcohol intervention. Addiction 2014 Feb;109(2):218-226. [doi: 10.1111/add.12383] [Medline: 24134709]
40. Stockwell T, Donath S, Cooper-Stanbury M, Chikritzhs T, Catalano P, Mateo C. Under-reporting of alcohol consumption in household surveys: a comparison of quantity-frequency, graduated-frequency and recent recall. Addiction 2004 Aug;99(8):1024-1033. [doi: 10.1111/j.1360-0443.2004.00815.x] [Medline: 15265099]
41. Morgenstern J, Kuerbis A, Houser J, Muench FJ, Shao S, Treloar H. Within-person associations between daily motivation and self-efficacy and drinking among problem drinkers in treatment. Psychol Addict Behav 2016 Sep;30(6):630-638. [doi: 10.1037/adb0000204] [Medline: 27560995]
42. Hoeppner BB, Kelly JF, Urbanoski KA, Slaymaker V. Comparative utility of a single-item versus multiple-item measure of self-efficacy in predicting relapse among young adults. J Subst Abuse Treat 2011 Oct;41(3):305-312. [doi: 10.1016/j.jsat.2011.04.005] [Medline: 21700411]
43. van Rijsbergen GD, Bockting CLH, Berking M, Koeter MWJ, Schene AH. Can a one-item mood scale do the trick? Predicting relapse over 5.5-years in recurrent depression. PLoS ONE 2012;7(10):e46796. [doi: 10.1371/journal.pone.0046796] [Medline: 23056456]
44. van Rijsbergen GD, Burger H, Hollon SD, et al. How do you feel? Detection of recurrent major depressive disorder using a single-item screening tool. Psychiatry Res 2014 Dec 15;220(1-2):287-293. [doi: 10.1016/j.psychres.2014.06.052] [Medline: 25070177]
45. Pot M, Paulussen TG, Ruiter RA, Mollema L, Hofstra M, Van Keulen HM. Dose-response relationship of a web-based tailored intervention promoting human papillomavirus vaccination: process evaluation of a randomized controlled trial. J Med Internet Res 2020 Jul 17;22(7):e14822. [doi: 10.2196/14822] [Medline: 32520718]
46. Pyatak EA, Blanche EI, Garber SL, et al. Conducting intervention research among underserved populations: lessons learned and recommendations for researchers. Arch Phys Med Rehabil 2013 Jun;94(6):1190-1198. [doi: 10.1016/j.apmr.2012.12.009] [Medline: 23262157]
47. The R Project for Statistical Computing. URL: https://www.r-project.org [accessed 2026-04-29]
48. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Software 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
49. Van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. J Stat Softw 2011;45:1-67. [doi: 10.18637/jss.v045.i03]
50. Stone AA, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med 2002;24(3):236-243. [doi: 10.1207/S15324796ABM2403_09] [Medline: 12173681]

XSL•FO
RenderX

51. Li K, Jolly T, Dziura J, Vaca FE. Greater readiness to cut down on drinking behaviors increases effects of a digital behavioral health intervention tool in reducing unhealthy alcohol use among Latino emergency department patients. Alcohol Alcohol 2025 Sep 16;60(6):agaf062. [doi: 10.1093/alcalc/agaf062] [Medline: 41020495]
52. Gaume J, Bertholet N, Daeppen JB. Readiness to change predicts drinking: findings from 12-month follow-up of alcohol use disorder outpatients. Alcohol Alcohol 2017 Jan;52(1):65-71. [doi: 10.1093/alcalc/agw047] [Medline: 27469491]
53. Maggs JL, Williams LR, Lee CM. Ups and downs of alcohol use among first-year college students: number of drinks, heavy drinking, and stumble and pass out drinking days. Addict Behav 2011 Mar;36(3):197-202. [doi: 10.1016/j.addbeh.2010.10.005] [Medline: 21106298]
54. Mohr DC, Cheung K, Schueller SM, Hendricks Brown C, Duan N. Continuous evaluation of evolving behavioral intervention technologies. Am J Prev Med 2013 Oct;45(4):517-523. [doi: 10.1016/j.amepre.2013.06.006] [Medline: 24050429]
55. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4. [doi: 10.2196/jmir.1376] [Medline: 20164043]
56. Malhi GS, Hamilton A, Morris G, Mannie Z, Das P, Outhred T. The promise of digital mood tracking technologies: are we heading on the right track? Evid Based Ment Health 2017 Nov;20(4):102-107. [doi: 10.1136/eb-2017-102757] [Medline: 28855245]
57. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283. [doi: 10.2196/20283] [Medline: 32990635]
58. Riper H, Hoogendoorn A, Cuijpers P, et al. Effectiveness and treatment moderators of internet interventions for adult problem drinking: an individual patient data meta-analysis of 19 randomised controlled trials. PLoS Med 2018 Dec;15(12):e1002714. [doi: 10.1371/journal.pmed.1002714] [Medline: 30562347]
59. Haug S, Boumparis N, Wenger A, Paz Castro R, Schaub MP. Mobile app-based coaching for alcohol prevention among adolescents: pre-post study on the acceptance and effectiveness of the program “MobileCoach Alcohol”. Int J Environ Res Public Health 2023 Feb 13;20(4):3263. [doi: 10.3390/ijerph20043263] [Medline: 36833958]
60. Suffoletto B, Clark DB, Lee C, et al. Development and preliminary testing of a secure large language model-based chatbot for brief alcohol counseling in young adults. Drug Alcohol Depend 2025 Jul 1;272:112697. [doi: 10.1016/j.drugalcdep.2025.112697] [Medline: 40334327]
61. van Buuren S. Flexible Imputation of Missing Data: CRC Press; 2012.
62. Nunnally JC. Psychometric Theory, 2nd edition: McGraw-Hill; 1978.
63. Waters AJ, Schoenmakers TM, Snelleman M, et al. Affect, motivation, temptation, and drinking among alcohol-dependent outpatients trying to maintain abstinence: an ecological momentary assessment study. Drug Alcohol Depend 2020 Jan 1;206:107626. [doi: 10.1016/j.drugalcdep.2019.107626] [Medline: 31786398]
64. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
65. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c869. [doi: 10.1136/bmj.c869] [Medline: 20332511]

Abbreviations

AI: artificial intelligence
CONSORT: Consolidated Standards of Reporting Trials
EMA: ecological momentary assessment
mHealth: mobile health
WDYD: What Do You Drink

Edited by L Buis; submitted 04.Dec.2024; peer-reviewed by D Gürtler, PY Lim; revised version received 05.Mar.2026; accepted 07.Mar.2026; published 26.May.2026.

Please cite as:
van Keulen H, Voogt C, Kleinjan M, Huizing A, Andree R, van Empelen P
Effects, Acceptability, and Use of a Dynamically Tailored Mobile What Do You Drink Intervention to Reduce Excessive Drinking Among Adolescents and Young Adults in the Netherlands: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e68468
URL: https://mhealth.jmir.org/2026/1/e68468
doi:10.2196/68468

XSL•FO
RenderX

© Hilde van Keulen, Carmen Voogt, Marloes Kleinjan, Arjan Huizing, Rosa Andree, Pepijn van Empelen. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 26.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Adherence to and Engagement With an mHealth Physical Activity Intervention After Mild Stroke or Transient Ischemic Attack: Secondary Analysis of a Feasibility Randomized Controlled Trial

Hanna Lagerlund[1], MSc; Lucian Bezuidenhout[1], PhD; Sophia Humphries[1], PhD; Lisa Holmlund[2], PhD; Lydia Kwak[3], PhD; Charlotte K Häger[4,5], PhD; David Moulaee Conradsson[1,6], PhD

[1]Division of Physiotherapy, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Alfred Nobels Allé 23, 23100, Stockholm, Sweden

[2]Division of Occupational Therapy, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden

[3]Unit of Intervention and Implementation Research for Worker Health, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

[4]Department of Community Medicine and Rehabilitation, Umeå University, Umeå, Sweden

[5]Department of Diagnostics and Intervention, Umeå University, Umeå, Sweden

[6]Medical Unit Allied Health Professionals, Women's Health and Allied Health Professionals Theme, Karolinska University Hospital, Stockholm, Sweden

Corresponding Author:
Hanna Lagerlund, MSc
Division of Physiotherapy, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Alfred Nobels Allé 23, 23100, Stockholm, Sweden

Abstract

Background: Regular physical activity is a crucial and an important modifiable lifestyle factor reducing the risk of recurrent incidents after stroke or transient ischemic attack (TIA). Mobile health (mHealth) has emerged as a promising approach for providing long-term support for physical activity. However, little is known about how individuals poststroke or TIA adhere to and engage with mHealth interventions.

Objective: This study aimed to (1) describe adherence to supervised sessions in an mHealth intervention targeting physical activity, (2) describe engagement with self-managed mHealth support for physical activity during and after the intervention, (3) compare characteristics of participants with high and low adherence and app engagement, and (4) examine whether high adherence and app engagement were associated with maintained physical activity after having completed the intervention and at a 12-month follow-up.

Methods: In this study, a secondary analysis of data from the experimental arm of a feasibility randomized controlled trial was conducted. The experimental group received a 6-month mHealth version of the i-REBOUND intervention, which included supervised mHealth support for physical activity and behavior change, followed by a 6-month postintervention period with access to self-managed mHealth support. The control group received mHealth consultations via video conferencing. Adherence measures included attendance at supervised exercise and counseling sessions, while app engagement was measured by weekly interactions with self-managed mHealth support during and after the intervention. Participants' level of physical activity (steps per day) was measured using accelerometers at baseline, and at 6- and 12-month postbaseline. Logistic regression analysis examined the associations between high adherence and app engagement during the intervention and postintervention period and maintained physical activity (ie, >7000 steps/day) across the 12-month study period.

Results: Of the 57 participants enrolled, 51 (89%) completed the intervention; the average age was 71 years, 34/51 (67%) were female, and 47/51 (92%) had mild stroke symptoms. Adherence to supervised mHealth support was high (supervised exercise sessions: 79%, counseling sessions: 98%), while engagement with self-managed mHealth support was high during the intervention (83%) but declined postintervention (38%). A larger proportion of females (24/31, 77%) demonstrated high adherence to the intervention compared to males (7/31, 23%, χ^2_1=4.1; *P*=.04). High adherence (≥80%) during the intervention was associated with maintained physical activity between baseline and the 6-month follow-up (OR 12.07, 95% CI 2 - 72.76; *P*=.01), while high app engagement (≥80%) during postintervention was associated with maintained physical activity between the 6- and 12-month follow-up (OR 5.10, 95% CI 1.02 - 25.52; *P*=.05).

Conclusions: Supervised mHealth support was well received with high adherence, while modules for self-management of physical activity faced challenges in engaging the participants. Future studies could benefit from qualitative and cocreative approaches to better understand and refine self-managed mHealth support for individuals poststroke or TIA.

XSL•FO
RenderX

Trial Registration: ClinicalTrials.gov NCT0511195; http://clinicaltrials.gov/ct2/show/NCT0511195

(JMIR Mhealth Uhealth 2026;14:e75662) doi:10.2196/75662

KEYWORDS

digital interventions; digital health; mobile health; mobile applications; secondary prevention; exercise; engagement; adherence

Introduction

Stroke affects 15 million people annually and is the third leading cause of death and disability worldwide [1]. A previous stroke or transient ischemic attack (TIA) increases the risk of a recurrent stroke, often with more severe health consequences [2]. Physical activity is an important modifiable lifestyle factor that contributes to a risk reduction of recurrent stroke and long-term disability for up to five years after a mild stroke or TIA [3]. Despite the crucial role of physical activity in secondary prevention [2], access to long-term support for physical activity after a stroke or TIA remains limited [4,5]. Many individuals poststroke fail to meet recommended physical activity levels, spend extended periods sedentary [6], and often require therapist-led support to maintain long-term engagement in physical activity.

Interventions combining onsite supervised physical exercise with support for physical activity through behavior change techniques appear to be particularly effective in improving cardiovascular health (eg, systolic blood pressure) [7]. However, barriers such as lack of access to professional support, long travel times, and inadequate public transportation limit people poststroke or TIA from attending on-site physical activity services [8,9]. Mobile health (mHealth), that is, health care delivered via mobile devices, is a promising intervention modality for promoting physical activity poststroke [10], offering a flexible solution with the potential to support long-term physical activity. Yet, individuals poststroke may face challenges in using mHealth technology due to limited digital literacy and stroke-related impairments such as fatigue, lack of motor skills, or cognitive difficulties, which can hinder their ability to effectively interact with mobile devices [11]. Despite these challenges, individuals poststroke are generally satisfied with digital services and show a willingness to engage with technology-based support, indicating that mHealth may be a feasible and acceptable option for this population [12]. In the context of secondary stroke prevention, mHealth provides an underutilized opportunity to enhance access to physical activity support and reduce the risk of secondary complications after stroke or TIA [10,13].

The health outcomes of digital interventions, such as mHealth, are influenced by the individuals' adherence to (ie, the extent to which the participant followed the intervention as designed) and engagement with the intervention [14-16]. Engagement is a multifaceted concept involving both *what* users are engaged with (ie, the targeted health behavior and the digital intervention) and *how* they engage with the digital intervention [17,18]. In terms of *how*, Perski et al [19] conceptualize engagement as having both behavioral and experiential dimensions. The behavioral dimension refers to the extent of active interaction with the intervention, while the experiential dimension encompasses subjective experiences such as feelings and interest [19]. A recent scoping review suggested higher adherence to mHealth interventions among males, older individuals, and those with secondary education [15]. Limited research has explored adherence to and engagement with mHealth interventions in individuals poststroke or TIA [10], and the factors influencing these behaviors remain largely unknown. A better understanding of adherence and app engagement can inform future development of mHealth interventions to promote and maintain physical activity poststroke or TIA.

This study builds on "i-REBOUND—let's get moving," a telehealth intervention developed in Australia to support home-based exercise and the promotion and maintenance of physical activity in people poststroke or TIA [20]. To improve reach, i-REBOUND was developed into a fully digital mHealth intervention [21], which provides 6 months of supervised support for physical activity and behavior change, followed by a 6-month postintervention period with access to self-managed mHealth support. Effective mHealth interventions depend on individuals' adherence and app engagement to ensure sufficient exposure to the intervention and, ultimately, to achieving the health outcomes. Thus, optimizing and tailoring support for future mHealth interventions requires a deeper understanding of how individuals adhere to and engage with these technologies. While a previous feasibility trial demonstrated that the i-REBOUND intervention was safe, acceptable, and accessible for people poststroke or TIA across Sweden [22], it did not assess participant adherence or engagement over time, nor the association with physical activity outcomes. Therefore, the present secondary analysis examines patterns of adherence to and engagement with the mHealth intervention and the association with maintained physical activity after the intervention. The specific aims were to (1) describe adherence to supervised exercise and individual counseling sessions during the intervention, (2) describe engagement with the self-managed mHealth support for physical activity during intervention and postintervention, (3) compare the characteristics of participants with high and low levels of adherence and engagement in the mHealth intervention, and (4) examine whether high adherence and app engagement were associated with maintained physical activity between baseline and the 6-month follow-up, and between the 6- and 12-month follow-up.

Methods

Study Design

The present study was a preplanned secondary analysis of data from the experimental arm of a feasibility randomized controlled trial (ClinicalTrials.gov: NCT0511195). The trial included individuals poststroke or TIA randomized into an experimental group receiving the mHealth version of the i-REBOUND intervention or a control group receiving mHealth consultations

XSL•FO
RenderX

via videoconferencing [21,22]. The analysis focused on participants' adherence to and engagement with the mHealth intervention and was therefore limited to the experimental arm. The data were collected between September 2021 and December 2023. This study was reported in accordance with the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (Checklist 1) [23].

Ethical Considerations

The study was approved by the Swedish Ethical Review Authority (dnr 2020 - 05062 and 2021 - 03622). Participation in the study was voluntary, and participants could withdraw at any time without consequence. No compensation was provided. Individuals who expressed interest received both verbal and written information before providing written informed consent. This consent also included approval for the secondary analysis conducted in the present study. A detailed description of the inclusion procedure and ethical considerations is available in the study protocol [21]. All data were anonymized to ensure participant confidentiality.

Participants

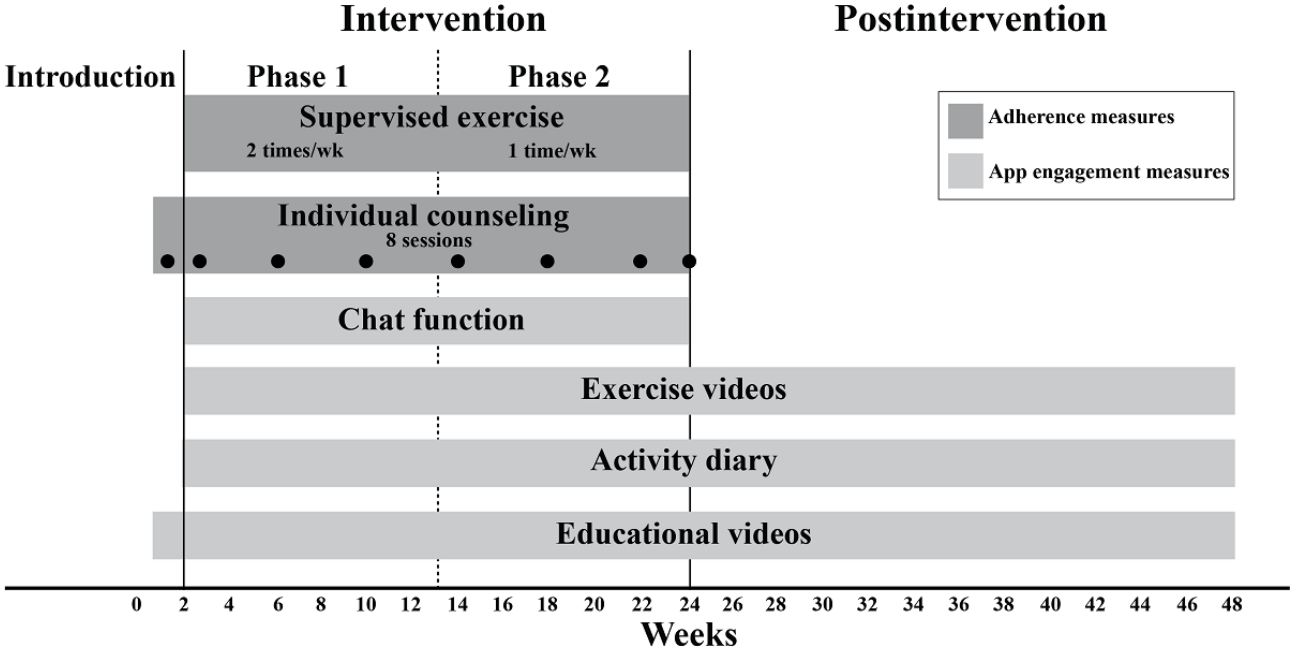
Participants were recruited through advertisements at collaborating clinics, social media, patient organizations, and the webpage of Karolinska Institutet. Inclusion criteria were (1) clinical diagnosis of stroke or TIA between 3 months and 10 years prior to study enrollment, confirmed by the participants' physician, (2) living at home, (3) being able to walk a short distance indoors with or without a walking device, (4) being able to use a smartphone including e-signature identification with or without the help of a relative or carer, and (5) having access to a stable internet connection. Exclusion criteria were (1) already meeting the recommended physical activity levels of at least 150 minutes per week of moderate physical activity or at least 75 minutes per week of vigorous-intensity physical activity, (2) severe health conditions compromising engagement in the intervention, or (3) enrolled in a concomitant clinical trial or participating in rehabilitation (eg, aerobic exercises) at the time point of recruitment.

Intervention

The mHealth Version of the i-REBOUND Intervention

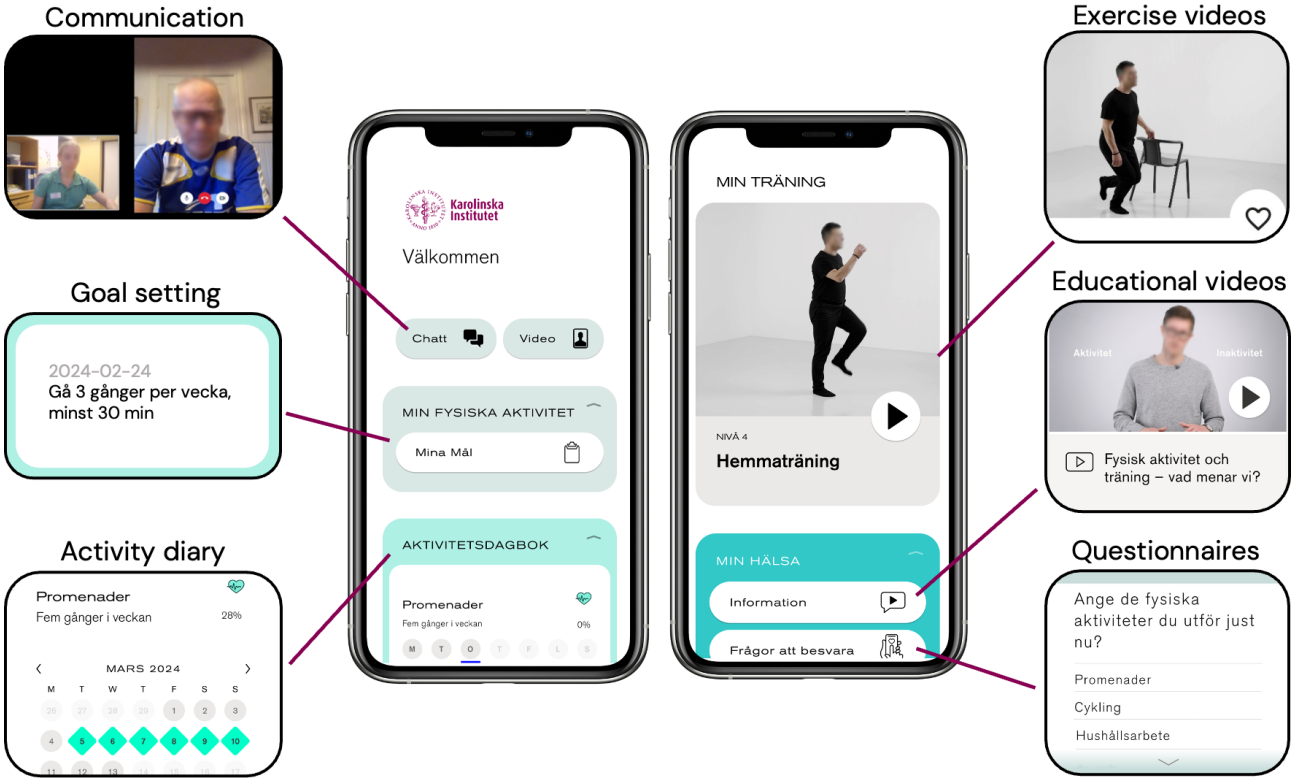
The mHealth intervention provided support for physical activity and behavior change to individuals poststroke or TIA across 48 weeks, divided into an intervention and postintervention period (Figure 1). Full intervention details are outlined in the study protocol [21]. The intervention was delivered via the STAAR (Stroke Treatment Through Active and Accessible Rehabilitation) app, which was managed by the medical technology company Empowered Health and available on iOS and Android devices. Participants accessed all intervention components through the app, including an overview of their individual goals, educational videos, an activity diary for self-monitoring, and prerecorded exercise videos (Figure 2). Supervised sessions were delivered via video calls within the app, and written communication occurred through a chat function. Two physical therapists with at least 5 years of experience in stroke rehabilitation delivered the intervention through a web-based digital clinic connected to the app. Baseline assessments were conducted using digital questionnaires administered via the app.

Figure 1. Illustration of the mobile health intervention and postintervention across the 48-week study period.



XSL•FO
RenderX

Figure 2. Overview of the mobile health intervention components and the Stroke Treatment Through Active and Accessible Rehabilitation (STAAR) app's interface.



Supervised mHealth Support

During the first week, the mHealth intervention was introduced in 2 sessions with the physical therapist, including a medical history review, discussions about resources for physical activity and home exercises, and a practice exercise session. Phase 1 of the intervention (wk 2 - 12) offered 2 supervised exercise sessions per week, and phase 2 (wk 13 - 24) offered 1 session. The supervised 30-minute exercise sessions aimed to reach physical activity at moderate intensity (ie, Borg Rating of Perceived Exercise Exertion Scale >12 [24]) and were delivered individually or in groups based on participants' physical capacity.

Support for behavior change consisted of individual monthly counseling sessions with the physical therapist. These sessions included discussions on goals, self-management strategies, and barriers and facilitators to physical activity. In addition to the supervised exercise sessions, participants were encouraged to engage in any physical activity of their choice. The final counseling session during the last week of the intervention period focused on developing an individualized self-management plan and setting goals for the postintervention period, supporting maintenance of exercise routines and engagement in physical activity.

Self-Managed mHealth Support

Educational videos on physical activity and exercise recommendations, individually prescribed exercise videos, and an activity diary were available throughout the intervention and postintervention period as behavioral change techniques to support the promotion of physical activity in analogy with the behavior change technique taxonomy (v1) by Michie et al [25] (see Multimedia Appendix 1). Participants were encouraged to engage with the self-managed mHealth support across the intervention and postintervention periods, although no specific instructions for intended usage were provided. Individual exercises, with prerecorded instructional videos for strength and/or aerobic training, were prescribed by the physical therapist based on the needs and preferences of each participant. The digital activity diary encouraged self-management by enabling participants to self-monitor their physical activity. Participants selected activities to monitor, such as outdoor walking, and designated days to perform them, marking each activity as "performed" or "not performed" in the diary. Participants had the option to receive mobile phone reminders for the selected activities through the activity diary. During postintervention, the study participants had access to self-managed mHealth support, but no physical therapist contact was offered.

Measures

Baseline Measures

Baseline data collection included demographics (age, sex, level of education, employment status, stroke/TIA diagnosis, and living situation) and assessment of level of stroke-disability according to Modified Rankin Scale; ranging from 0 to 5, where 0 indicates no disability and 5 indicates severe disability [26].

Digital questionnaires were administered via the STAAR app to assess functioning. Self-efficacy for exercise was assessed by the 9-item Exercise Self-Efficacy Scale [27]. For each item, participants indicate their confidence to execute the behavior on a 100-point percentage scale divided into 10-point increments, ranging from 0% (not at all confident) to 100% (highly confident). Fatigue was assessed by the Fatigue Severity

XSL•FO
RenderX

Scale 9-item version scored on a 7-point Likert scale, ranging from 1 (“disagree”) to 7 (“fully agree”), with the median score of the 9 items used for analysis [28]. Levels of depression, anxiety, and stress were assessed by the Depression, Anxiety, and Stress Scale 21 [29]. Each of the 21 items is rated on a 4-point Likert scale ranging from 0 (“did not apply to me at all”) to 3 (“applied to me very much, or most of the time”), 7 items each pertaining to depression, anxiety, and stress. Scores for the subscales were calculated by summing the scores of each domain. Self-perceived impact of stroke was assessed using the 8 subscales (strength, memory, emotion, communication, activities of daily living/instrumental activities of daily living, mobility, hand function, social and participation) of the Stroke Impact Scale [30]. The score ranges from 0 to 100; higher scores indicate a lower perceived impact. Recovery after stroke was assessed with a visual analog scale, in which participants were asked to score their global perceived stroke recovery from 0 (“no recovery”) to 100 (“complete recovery”) [30].

Measures of Adherence to Supervised Exercise and Individual Counseling Sessions

Adherence to supervised exercise and individual counseling sessions during the intervention was documented in the digital clinic by the physical therapist. The intended adherence was 2 supervised exercise sessions per week in phase 1 and one exercise session per week in phase 2, along with eight counseling sessions across phases 1 and 2. Individual adherence rates in percentage were calculated by dividing exercise sessions attended by the total sessions offered in phase 1 (n=22) and phase 2 (n=12). The same approach was applied to calculate adherence to the counseling sessions offered (n=8). Based on their adherence to supervised exercise, participants were categorized into high adherence (>80%) and low adherence (<80%) groups [22].

App Engagement Measures

Participants’ weekly app engagement was tracked via user logs and assessed during the intervention and postintervention, with an “active week” defined as at least one interaction with the chat function (only intervention period), prescribed exercises, or activity diary. Overall app engagement was then calculated as the percentage of active weeks over the 48-week study period, excluding the introductory week. Based on this metric, participants were categorized into high app engagement (≥80%) and low app engagement (<80%) groups. Disengagement was defined as the point at which participants ceased interacting with self-managed mHealth support (ie, the activity diary and the prescribed exercise videos) for the remainder of the study until completion. Time to disengagement was calculated as the number of weeks from the start of the postintervention period to the participant’s last recorded engagement.

Participants were categorized by activity diary use as low (<19 wk), sporadic (20 - 39 wk), or frequent (>40 wk) engagers and by engagement with prescribed exercises as limited (≤5 sessions), sporadic (6 - 19 sessions), or frequent (>20 sessions) engagers. Participant categorization by engagement with activity diary and prescribed exercises was guided by a pragmatic approach, with thresholds selected based on the observed distribution of data. Engagement with educational videos across the entire study period was calculated as the number of viewed videos divided by the total number available (n=8).

Physical Activity Measures

To explore the relationship between intervention adherence, app engagement, and maintenance of physical activity in daily life, physical activity was measured objectively using the activPAL accelerometer (PAL Technologies Ltd) at baseline, and at 6- and 12-month postbaseline [31]. The activPAL device demonstrates reliability and validity in measuring step count [32] and is recommended for assessing physical activity after a stroke [33]. Participants received the monitor by mail with instructions to wear it on their nonaffected or dominant leg (as applicable) for 7 consecutive days (24 h/d) before returning it via prepaid post. Participants with at least 3 days of physical activity data, each containing a minimum of 10 hours of data, were included in the analysis [34]. The outcome was the number of steps per day. As no minimal clinically important difference in steps has been established for the stroke population, we used at least 7000 steps/day as a criterion for optimal cardiovascular risk reduction [35]. This cut-off was also deemed a reasonable and achievable target at the 6- and 12-month follow-ups, given that the baseline physical activity level in our sample was approximately 6000 steps per day. Changes in physical activity between baseline and the 6-month follow-up, and between the 6- and 12-month follow-up, were grouped into 4 categories:

1. Unchanged—low physical activity (<7000 steps/d at baseline and 6-mo follow-up, <7000 steps/d at 6-mo and 12-mo follow-up)
2. Declined physical activity (≥7000 steps/d at baseline and <7000 at 6-mo follow-up, >7000 steps/d at 6-mo follow-up and <7000 at 12-mo follow-up)
3. Unchanged—active (≥7000 steps/d at baseline and 6-mo follow-up, ≥7000 steps/d at 6-mo and 12-mo follow-up)
4. Improved physical activity (<7000 steps/d at baseline and ≥7000 at 6 mo follow-up, <7000 steps/d at 6 mo follow-up and ≥7000 at 12 mo follow-up).

Participants in categories 3 and 4 were classified as having maintained their level of physical activity.

Data Analysis

IBM SPSS Statistics software (version 24.0; IBM Corp) was used for statistical analyses. Baseline characteristics and descriptive statistics for aims 1 to 3 were calculated and presented as mean (SD), median (IQR), and frequencies (%).

For the first and second aim, descriptive statistics described the adherence to and engagement in the mHealth intervention across the study period. Participants who completed the introductory week and began intervention phase 1 were included in the analysis. Those who discontinued the intervention were considered lost to follow-up, with data analyzed up to the week of discontinuation.

For the third aim, between-group differences regarding demographic characteristics between the sub-groups with high and low adherence, and high and low app engagement, were analyzed using Mann-Whitney *U* test for continuous variables and chi-square test for binary variables.

XSL•FO
RenderX

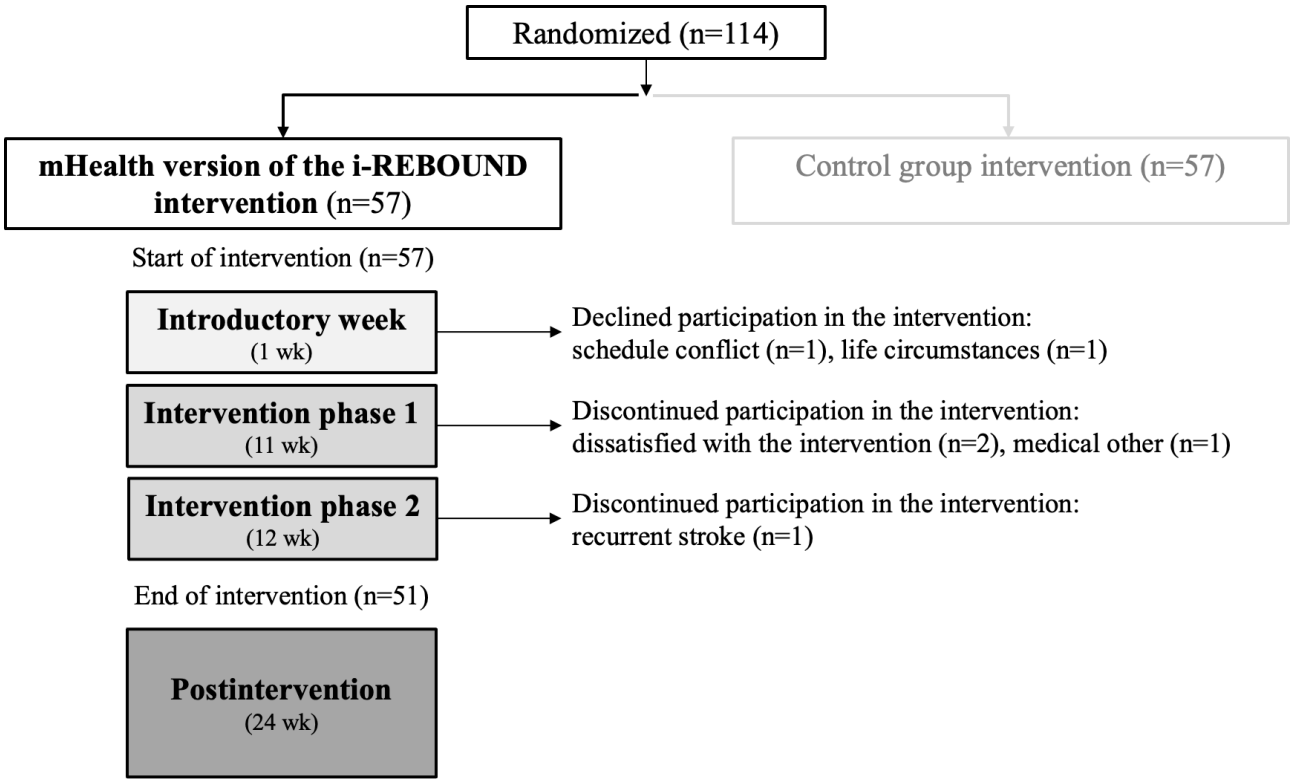
For the fourth aim, logistic regression analysis was used to examine the associations between high adherence during the intervention, high app engagement during the intervention and postintervention periods, and maintained physical activity levels (categories 3 and 4, as described earlier) from baseline to the 6-month follow-up and from the 6-month to the 12-month follow-up. These analyses included age (years) and use of a walking aid as covariates, as they have been shown to be associated with the level of physical activity in people poststroke [36]. The α level was set at .05 for all analysis.

Results

Characteristics of Participants

Of the 57 participants enrolled, 51 (89%) completed the intervention (Figure 3). The mean age of the participants completing the intervention was 71 (SD 8) years, and the median time since stroke or TIA was 2.4 (IQR 1.2-5) years. Of the 51 participants who completed the intervention, 34 (67%) were female, 27 (53%) had a university degree, 8 (16%) were employed, and 47 (93%) reported no or mild poststroke or TIA symptoms (ie, Modified Rankin Scale score 0 - 1) (Table 1).

Figure 3. Participant flow chart.



XSL•FO
RenderX

Table . Baseline demographic and clinical characteristics of participants who completed the intervention versus those who discontinued, and both groups pooled.

	Completed intervention (n=51)	Discontinued (n=6)	All (n=57)
Age, mean (SD)	71 (8)	70 (12)	71 (9)
Sex, female, n (%)	34 (67)	4 (67)	38 (67)
University degree, n (%)	27 (53)	4 (67)	31 (54)
Employed, n (%)	8 (16)	1 (17)	9 (16)
TIA[a], n (%)	15 (29)	2 (33)	17 (30)
Stroke, n (%)	36 (71)	4 (67)	40 (70)
Years since stroke, median (IQR)	2.4 (1.2-5)	2.4 (1.1-4.6)	2.4 (1.2-4.9)
Use walking aid, n (%)	16 (31)	2 (33)	18 (32)
Living with someone, n (%)	34 (67)	2 (33)	36 (63)
Modified Rankin Scale, n (%)			
No symptoms	10 (20)	—[b]	10 (18)
No significant symptoms	37 (73)	6 (100)	43 (75)
Slight disability	3 (6)	—	3 (5)
Moderate disability	1 (2)	—	1 (2)
Depression, Anxiety, and Stress Scale (0 - 42), median (IQR)			
Depression	4 (2-10)	5 (4-14)	4 (2-10)
Anxiety	2 (0-6)	3 (2-8)	2 (0-6)
Stress	6 (4-10)	13 (0-19)	8 (3-10)
Stroke Impact Scale (0 - 100), median (IQR)			
Strength	75 (63-94)	81 (75-95)	81 (63-94)
Memory	91 (81-94)	89 (66-98)	91 (80-95)
Emotion	53 (47-56)	53 (47-63)	53 (47-56)
Communication	96 (89-100)	96 (85-97)	96 (89-100)
ADL/iADL[c]	94 (83-98)	93 (85-97)	94 (83-98)
Mobility	88 (76-98)	89 (82-94)	88 (76-98)
Hand function	95 (65-100)	85 (78-100)	90 (70-100)
Social participation	81 (61-97)	72 (59-81)	78 (61-94)
Recovery	77 (60-90)	58 (44-83)	77 (55-90)
Self-efficacy for exercise (0 - 90), median (IQR)	68 (54-79)	58 (26-77)	66 (53-79)
Fatigue Severity Scale (1-7), median (IQR)	4 (3-5)	4 (3-6)	4 (3-5)
Physical activity ≥7000 steps/d, n (%)	24 (47)	1 (17)	25 (44)

[a]TIA: transient ischemic attack.

[b]Not applicable.

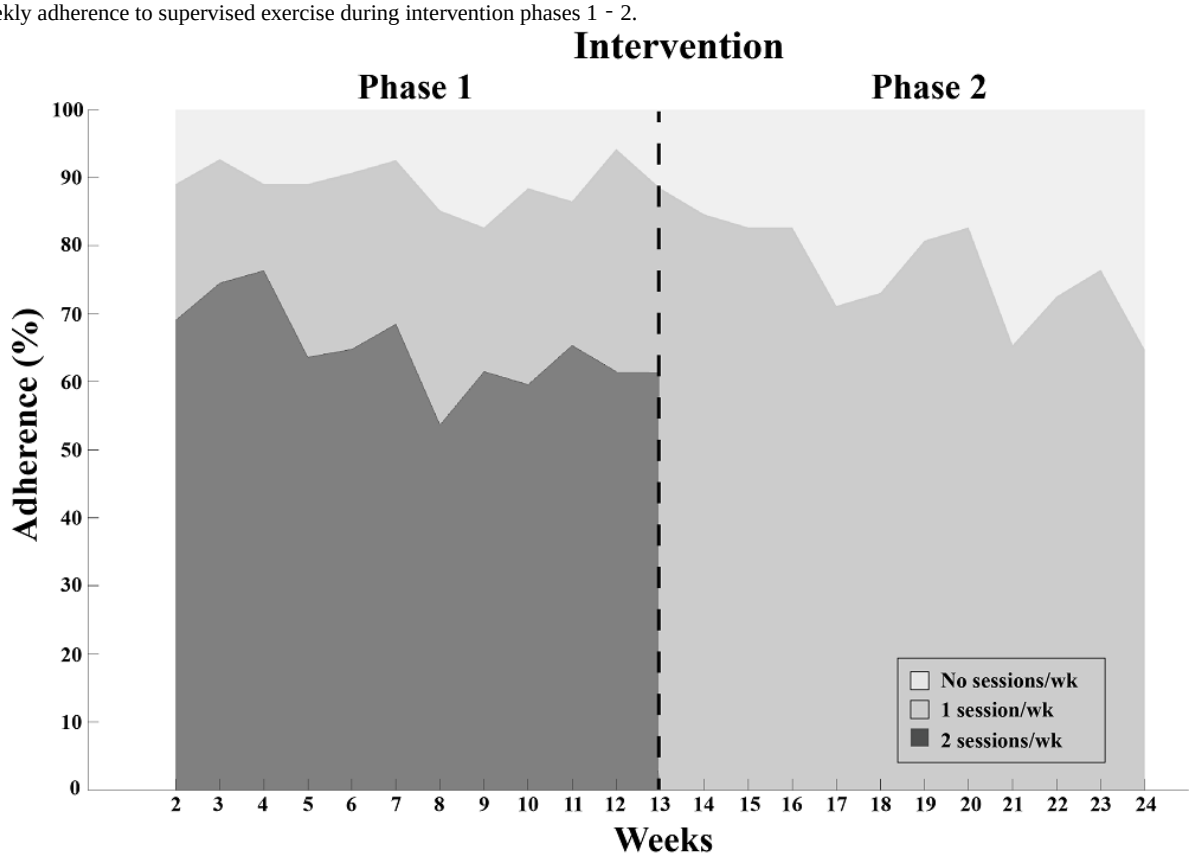
[c]ADL/iADL: activities of daily living/instrumental activities of daily living.

Adherence to Supervised Exercise and Individual Counseling Sessions

The mean adherence to supervised exercise across the intervention period was 79% (min-max: 29% - 100%) (Figure 4), with 911 of 1151 sessions attended in phase 1 and 489 of 618 sessions attended in phase 2. The mean adherence rate to individual counseling sessions across the intervention was 97% (min-max: 50% - 100%) and among those who completed the intervention, 82% (42/51) attended all counseling sessions.

XSL•FO
RenderX

Figure 4. Weekly adherence to supervised exercise during intervention phases 1 - 2.

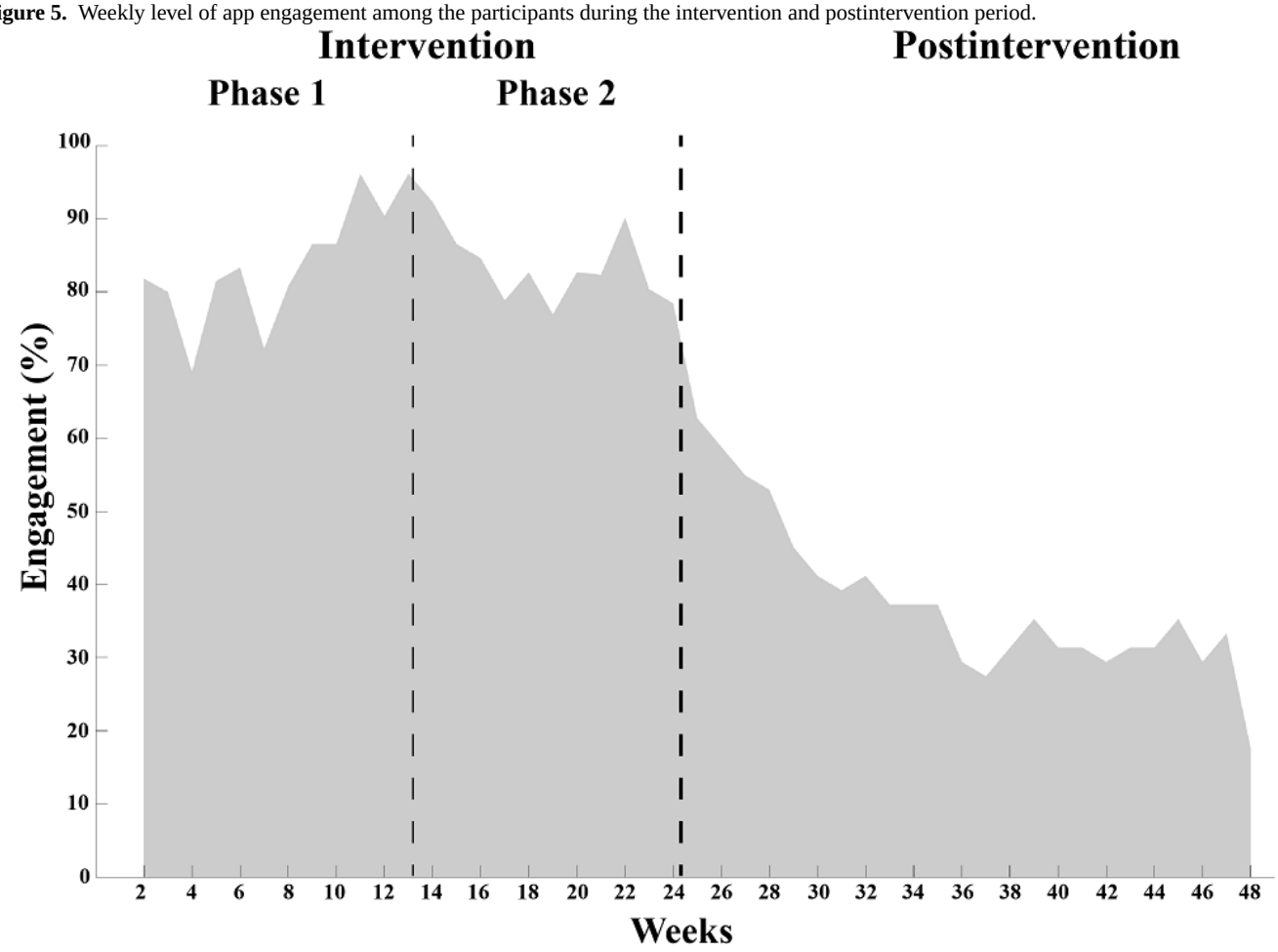


Engagement With Self-Managed mHealth Support

App engagement was higher during the intervention phase, with an average of 84% of participants using the app weekly (min-max: 69% - 94%), compared to the postintervention period (average 38%; min-max: 18% - 63%) (Figure 5). The majority of participants (42/51, 82%) disengaged during the postintervention period, with an average time to disengagement of 8 weeks (min-max, 0 - 23).

XSL•FO
RenderX

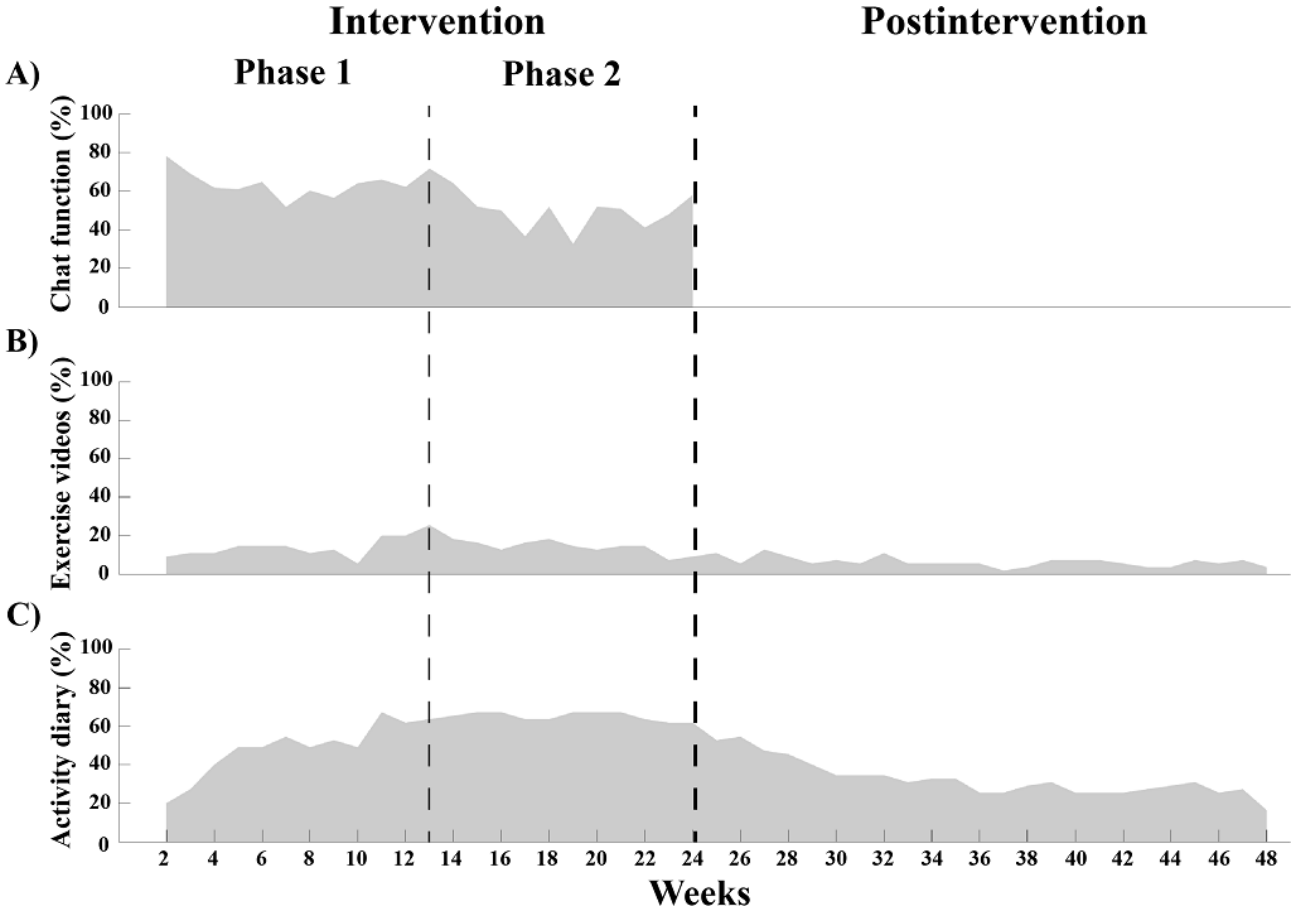
Figure 5. Weekly level of app engagement among the participants during the intervention and postintervention period.



Engagement with the chat function and with the activity diary were the most common types of app engagement (Figure 6). The average weekly engagement with the chat function was 63% in phase 1, decreasing to 50% in phase 2 (Figure 6A). On average, participants sent 1.6 weekly chat messages during the intervention. Forty-three of 55 participants (78%) used the activity diary to some extent across the study period (Figure 6C). In total, 103 activities were monitored, and the most frequently monitored activities were outdoor walking (n=29, 28%), aerobic exercise (n=13, 13%), and unspecified physical exercise (n=12, 12%). Among the 43 participants using the activity diary, 14 (33%) demonstrated low usage (<19 wk), 17 (40%) sporadic usage (20 - 39 wk), and 12 (28%) frequent usage (>40 wk). Fifty of 55 participants (91%) were prescribed individual exercises, and among those, 33 (66%) had limited engagement (<5 sessions), 12 (24%) sporadic engagement (6 - 19 sessions), and 5 (10%) frequent engagement (>20 sessions). The mean engagement with educational videos was 67% (min-max: 0% - 100%), with each participant viewing 5 of 8 videos on average, and 53% (29/55) participants watching all content.

XSL•FO
RenderX

Figure 6. Weekly engagement in the chat function (A), exercise videos (B), and activity diary (C) during the intervention and postintervention periods.



Comparison of Characteristics of Users With Different Levels of Adherence and App Engagement

Of the 51 participants completing the intervention, 31 (61%) had high adherence (ie, ≥80%) and 15 (29%) had high app engagement (ie, ≥80%). Most personal and contextual factors, as well as clinical and functional measures, showed no significant differences between participants with high versus low adherence or high versus low app engagement in the mHealth intervention (Table 2). A higher proportion of females (24/31, 77%) demonstrated high adherence to the intervention compared to males (7/31, 23%, χ^2_1=4.1; *P*=.04).

XSL•FO
RenderX

Table . Comparison between 4 user groups.

Variable	High adherence (n=31)	Low adherence (n=20)	*P* value	High app engagement (n=15)	Low app engagement (n=36)	*P* value
Age (years), median (IQR)	72 (68-78)	71 (64-77)	.38	72 (63-76)	72 (68-78)	.76
Sex, female, n (%)	24 (77)	10 (50)	.04	9 (60)	25 (69)	.51
University degree, n (%)	16 (52)	11 (55)	.81	7 (47)	20 (55)	.56
Employed, n (%)	5 (16)	3 (16)	.91	1 (7)	7 (19)	.25
Stroke diagnosis, n (%)	22 (71)	14 (70)	.94	9 (60)	27 (75)	.28
Years since stroke, median (IQR)	2 (1.3-5)	2.5 (0.9-4.9)	.99	3 (1.3-7.7)	2.4 (1-4.6)	.53
Use walking aid, n (%)	9 (29)	7 (35)	.65	5 (33)	11 (31)	.85
Depression, Anxiety, and Stress Scale (0 - 42), median (IQR)						
Depression	4 (2-10)	4 (2-12)	.39	2 (0-10)	4 (2-10)	.10
Anxiety	2 (2-6)	3 (0-6)	.21	2 (0-6)	4 (2-6)	.83
Stress	8 (2-10)	6 (4-10)	.89	4 (2-10)	8 (4-10)	.73
Stroke Impact Scale, (0 - 100), median (IQR)						
Strength	75 (63-94)	81 (58-94)	.95	88 (75-100)	75 (58-92)	.14
Memory	91 (81-94)	91 (79-96)	.73	94 (78-100)	88 (81-94)	.15
Emotion	53 (47-56)	51 (47-56)	.50	56 (53-56)	51 (47-56)	.14
Communication	96 (89-100)	95 (87-99)	.54	96 (89-100)	95 (89-99)	.62
ADL/iADL[a]	94 (83-100)	96 (80-98)	.52	98 (83-100)	93 (77-95)	.11
Mobility	85 (78-98)	89 (85-99)	.70	98 (78-100)	85 (78-95)	.15
Hand function	95 (70-100)	90 (50-100)	.60	95 (70-100)	90 (61-100)	.39
Social participation	81 (69-97)	81 (51-97)	.49	86 (67-100)	76 (59-94)	.24
Recovery	79 (65-91)	75 (51-90)	.27	81 (50-99)	76 (61-90)	.23
Self-efficacy for exercise (0 - 90), median (IQR)	68 (54-81)	65 (53-79)	.69	68 (51-79)	68 (56-81)	.80
Fatigue severity scale (1-7), median (IQR)	4 (3-5)	4 (3-5)	.91	3 (2-5)	4 (3-5)	.19

[a]ADL/iADL: activities of daily living/instrumental activities of daily living.

Factors Associated With Maintained Physical Activity

Valid physical activity data were available for 45 participants (88%) at 6 months and 42 participants (82%) at 12 months among those who completed the intervention. Twenty-four participants (53%) maintained their physical activity between baseline and the 6-month follow-up, whereas 19 (45%) maintained their physical activity between the 6- and 12-month follow-ups (Table 3). High adherence during the intervention was associated with maintained physical activity between baseline and the 6 months follow-up (Odds ratio [OR] 12.07, 95% CI 2-72.76; *P*=.01), and high app engagement during the postintervention was associated with maintained physical activity between 6 and 12 months (OR 5.10, 95% CI 1.02-25.52; *P*=.05). High app engagement during the intervention was not associated with maintained physical activity between baseline and the 6-month follow-up (OR 2.80, 95% CI 0.55-14.28; *P*=.22), nor was high adherence during the intervention associated with maintained physical activity between the 6- and 12-month follow-ups (OR 3.53, 95% CI 0.68-18.35; *P*=.13).

XSL•FO
RenderX

Table . Study participants demonstrating ≥7000 steps/day at 6- and 12-month follow-up, as well as a classification of the change between baseline and the follow-ups.

	Classification of change[a]	
	Baseline to 6-month follow-up (n=45), n (%)	6- to 12-month follow-up (n=42), n (%)
Unchanged—low physical activity	16 (36)	16 (38)
Declined physical activity	5 (11)	7 (17)
Unchanged—physically active	17 (38)	17 (40)
Improved physical activity	7 (15)	2 (5)

[a]Unchanged—low physical activity (ie, <7000 steps/d at baseline and 6-mo follow-up, <7000 steps/day at 6-mo and 12-mo follow-up); declined physical activity (ie, >7000 steps/day at baseline and <7000 at 6-mo follow-up, >7000 steps/d at 6-mo follow-up and <7000 at 12-mo follow-up); unchanged—active (ie, ≥7000 steps/d at baseline and 6-mo follow-up, ≥7000 steps/d at 6-mo and 12-mo follow-up); and improved physical activity (ie, <7000 steps/d at baseline and ≥7000 at 6 mo follow-up, <7000 steps/d at 6 mo follow-up and ≥7000 at 12 mo follow-up).

Discussion

Principal Findings

This study investigated adherence and app engagement during and 6 months after an mHealth intervention for people poststroke or TIA. Adherence to supervised sessions remained high (>75%) across the intervention, and more females adhered to the supervised exercise than males. Engagement with self-managed mHealth support was initially high but decreased postintervention. High adherence during the intervention was associated with maintained physical activity between baseline and the 6-month follow-up, while high postintervention engagement was associated with maintained physical activity between the 6- and 12-month follow-ups.

Findings in Context

Adherence to supervised exercise during the intervention averaged 79%, which is comparable to adherence rates in nondigital [37] and digital interventions for people poststroke or TIA [20,38]. Adherence to counseling sessions (97%) was higher compared to a previous study showing decreasing adherence to long-term nondigital physical activity support in people poststroke [39]. The current study's digital format provided an accessible and flexible solution that may have contributed to the high adherence rates. High adherence to supervised phases of exercise programs is common [40]. In this study, weekly interactions with a physical therapist as part of the mHealth intervention likely contributed positively to adherence. Additionally, most participants attended group exercise sessions, which may have further promoted adherence by creating a sense of accountability and commitment to the group [12,38]. Low app engagement within physical activity apps is a well-known challenge in mHealth interventions [41], yet our findings showed high engagement (84%) with self-managed support modules during the intervention, particularly the chat function and activity diary. However, app engagement postintervention decreased to 38%. As with adherence, app engagement during the intervention was likely promoted by regular interaction with a physical therapist, whereas the lack of such contact postintervention may have contributed to the decreasing patterns of engagement [19,42]. Similar patterns of declining engagement have been observed in unsupervised mHealth interventions for people with chronic conditions [43]. While the supervised sessions followed a structured schedule, engagement with the self-managed mHealth support was optional and guided by the participant's preferences. A recent review suggests that clear guidance on the frequency and type of mHealth engagement can promote maintained participant engagement [43]. Thus, the lack of explicitly defined goals of app engagement in the current mHealth intervention might further explain the decreasing app engagement postintervention.

In line with a meta-analysis of digital physical activity interventions in adults [16], our results suggest that adherence during the intervention and app engagement postintervention is important for maintaining physical activity after the intervention period. This study found no significant differences in high adherence or app engagement between sub-groups, such as age or education level, which has previously been shown to influence mHealth intervention adherence [15]. However, our results suggest that females may have adhered to supervised sessions more than males, potentially because females tend to be more proactive and engaged in health-related issues [44]. The results from this study support that adherence to and engagement with an mHealth intervention might be important factors influencing the intervention's impact on maintained engagement in physical activity.

Several aspects of the self-managed mHealth support could be improved to enhance the app engagement in the current mHealth intervention, as well as long-term engagement in physical activity beyond the intervention period. Personalized app content commonly benefits app engagement [41]. However, despite access to an activity diary with self-selected activities and individually prescribed exercises, only 28% (12/43) and 10% (5/50), respectively, frequently engaged with these modules in the present study. Engaging mHealth features, such as prompts, gamification, and regular content updates, have been shown to significantly enhance app engagement [19,41,42]. These elements were not incorporated into the current intervention; however, integrating them in future development could be a valuable strategy to encourage maintained app engagement among individuals poststroke or TIA. Another promising approach to enhance engagement in people poststroke or TIA, particularly those with lower digital health literacy, is to develop strategies that actively involve caregivers in participating in the mHealth intervention and mediating treatment delivery. This approach has been tested in on-site interventions [45] but has

XSL•FO
RenderX

been rarely implemented in digital formats in stroke trials. Furthermore, a potential refinement of the intervention design could be development of distinct phases, each with specific goals for supporting behavior change. This might include an initiation phase focusing on establishing routines for exercise and physical activity (like the present intervention), followed by a maintenance phase in which therapist support occurs less frequently. This design could be especially important for people poststroke or TIA, who often require long-term support to establish and maintain behavior change [39]. Advancing toward more independent self-management of physical activity would also require extending the intervention's focus to support participants in developing strategies for sustaining behavior change after the intervention, which should be embedded within the individual counseling and educational materials. We believe that refinement of the mHealth version of the i-REBOUND intervention should be cocreated with people poststroke or TIA and target multiple domains, including technology, intervention design, and the establishment of clear expectations for adherence and engagement [46].

Limitations

The findings should be interpreted cautiously given the relatively small sample size. In addition, the participants were mostly female, highly educated, and with mild stroke symptoms, which limits the generalizability of the findings to other stroke or TIA populations [47]. Furthermore, the 7000 steps/day cut-off for physical activity, while clinically relevant for optimal cardiovascular risk reduction [35], is a crude measure for detecting change over time. Moreover, participants in the sample exhibited a notably high level of physical activity at baseline, exceeding the average daily step count of 4078 for individuals with stroke [6]. User logs yielded an in-depth examination of the participants' mHealth engagement behavior but could not sufficiently capture the multifaceted nature of engagement. For a more comprehensive understanding of engagement, exploration of participants' subjective experiences is required, which we acknowledge as a limitation of the present study. A strength was the 48-week study period, which allowed for a thorough investigation of changes in adherence and engagement patterns over time. However, adherence and engagement levels were classified solely as 'low' (<80%) or 'high' (>80%). For more substantiated levels, future studies should determine adequate levels for effective mHealth engagement and adherence and set predefined benchmarks [43].

Conclusion

Supervised mHealth support was well received, with high adherence. In contrast, the self-management modules for physical activity faced challenges in engaging participants. High adherence during the intervention was associated with physical activity maintenance immediately afterward, while high app engagement postintervention was linked to maintenance at the 12-month follow-up. These exploratory findings may be hypothesis-generating and inform future research. To enhance engagement with mHealth interventions and support maintained physical activity after the supervised phase, future studies could benefit from qualitative and cocreative approaches to better understand and refine self-managed mHealth support for individuals poststroke or TIA.

Acknowledgments

We thank all participants in the study and the physical therapists who delivered the intervention. To assist with language refinement and editorial improvements during manuscript writing, a generative artificial intelligence tool (Microsoft Copilot) was used. All intellectual content, including the study design, interpretation, and conclusions, is entirely the authors' own.

Funding

This work was supported by grants from the Swedish Research Council (2022-01403), Swedish Research Council for Health, Working Life and Welfare (2021-01018), Center for Innovative Medicine, Region Stockholm (FoUI-960631), Sweden's Innovation Agency (2021-01726), Swedish Stroke Association, Strategic Research Area Health Care Sciences at Karolinska Institutet, and Research School of Health Science at Karolinska Institutet.

Data Availability

The dataset used in this article is not publicly available due to the risk of compromising research participants' privacy in accordance with Swedish legislation. Requests to access data should be directed to the Research Data Office at Karolinska Institutet (rdo@ki.se).

Authors' Contributions

Conceptualization: HL, LB, SH, LH, LK, CKH, DMC
Data curation: LB, HL
Formal analysis: HL, DMC
Funding acquisition: DMC
Investigation: DMC, SH, LB, HL
Methodology: HL, DMC
Resources: DMC
Project administration: DMC, HL, SH
Visualization: LB, HL

XSL•FO
RenderX

Writing – original draft: HL, DMC
Writing – review & editing: HL, LB, SH, LH, LK, CKH, DMC

Conflicts of Interest

None declared.

Multimedia Appendix 1
Intervention components and their alignment to the behavior change technique taxonomy (v1) by Michie et al [25].
[DOCX File, 22 KB - mhealth_v14i1e75662_app1.docx]

Checklist 1
CONSORT-EHEALTH checklist (version 1.6.1).
[PDF File, 1200 KB - mhealth_v14i1e75662_app2.pdf]

References

1. Feigin VL, Stark BA, Johnson CO. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Neurol 2021 Oct;20(10):795-820. [doi: 10.1016/S1474-4422(21)00252-0] [Medline: 34487721]
2. Hankey GJ. Secondary stroke prevention. Lancet Neurol 2014 Feb;13(2):178-194. [doi: 10.1016/S1474-4422(13)70255-2] [Medline: 24361114]
3. Hobeanu C, Lavallée PC, Charles H, et al. Risk of subsequent disabling or fatal stroke in patients with transient ischaemic attack or minor ischaemic stroke: an international, prospective cohort study. Lancet Neurol 2022 Oct;21(10):889-898. [doi: 10.1016/S1474-4422(22)00302-7] [Medline: 36115361]
4. Cuccurullo SJ, Fleming TK, Petrosyan H, Hanley DF, Raghavan P. Mechanisms and benefits of cardiac rehabilitation in individuals with stroke: emerging role of its impact on improving cardiovascular and neurovascular health. Front Cardiovasc Med 2024;11:1376616. [doi: 10.3389/fcvm.2024.1376616] [Medline: 38756753]
5. Norrving B, Barrick J, Davalos A, et al. Action plan for stroke in Europe 2018-2030. Eur Stroke J 2018 Dec;3(4):309-336. [doi: 10.1177/2396987318808719] [Medline: 31236480]
6. Fini NA, Holland AE, Keating J, Simek J, Bernhardt J. How physically active are people following stroke? Systematic review and quantitative synthesis. Phys Ther 2017 Jul 1;97(7):707-717. [doi: 10.1093/ptj/pzx038] [Medline: 28444348]
7. Deijle IA, Van Schaik SM, Van Wegen EEH, Weinstein HC, Kwakkel G, Van den Berg-Vos RM. Lifestyle interventions to prevent cardiovascular events after stroke and transient ischemic attack: systematic review and meta-analysis. Stroke 2017 Jan;48(1):174-179. [doi: 10.1161/STROKEAHA.116.013794] [Medline: 27924055]
8. Sammut M, Haracz K, Shakespeare D, et al. Physical activity after transient ischemic attack or mild stroke is business as usual. J Neurol Phys Ther 2022 Jul 1;46(3):189-197. [doi: 10.1097/NPT.0000000000000395] [Medline: 35727994]
9. Nicholson SL, Donaghy M, Johnston M, et al. A qualitative theory guided analysis of stroke survivors' perceived barriers and facilitators to physical activity. Disabil Rehabil 2014;36(22):1857-1868. [doi: 10.3109/09638288.2013.874506] [Medline: 24377329]
10. Rintala A, Kossi O, Bonnechère B, Evers L, Printemps E, Feys P. Mobile health applications for improving physical function, physical activity, and quality of life in stroke survivors: a systematic review. Disabil Rehabil 2023 Dec;45(24):4001-4015. [doi: 10.1080/09638288.2022.2140844] [Medline: 36325613]
11. Niyomyart A, Ruksakulpiwat S, Benjasirisan C, et al. Current status of barriers to mHealth access among patients with stroke and steps toward the digital health era: systematic review. JMIR Mhealth Uhealth 2024 Aug 22;12:e54511. [doi: 10.2196/54511] [Medline: 39173152]
12. Caughlin S, Mehta S, Corriveau H, et al. Implementing telerehabilitation after stroke: lessons learned from Canadian trials. Telemed J E Health 2020 Jun;26(6):710-719. [doi: 10.1089/tmj.2019.0097] [Medline: 31633454]
13. English C, Ceravolo MG, Dorsch S, et al. Telehealth for rehabilitation and recovery after stroke: state of the evidence and future directions. Int J Stroke 2022 Jun;17(5):487-493. [doi: 10.1177/17474930211062480] [Medline: 34983266]
14. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 5;13(3):e52. [doi: 10.2196/jmir.1772] [Medline: 21821503]
15. Yang Y, Boulton E, Todd C. Measurement of adherence to mHealth physical activity interventions and exploration of the factors that affect the adherence: scoping review and proposed framework. J Med Internet Res 2022 Jun 8;24(6):e30817. [doi: 10.2196/30817] [Medline: 35675111]
16. Mclaughlin M, Delaney T, Hall A, et al. Associations between digital health intervention engagement, physical activity, and sedentary behavior: systematic review and meta-analysis. J Med Internet Res 2021 Feb 19;23(2):e23180. [doi: 10.2196/23180] [Medline: 33605897]
17. Milne-Ives M, Homer S, Andrade J, Meinert E. The conceptualisation and measurement of engagement in digital health. Internet Interv 2024 Jun;36:100735. [doi: 10.1016/j.invent.2024.100735] [Medline: 38558760]

XSL•FO
RenderX

18. Cole-Lewis H, Ezeanochie N, Turgiss J. Understanding health behavior technology engagement: pathway to measuring digital behavior change interventions. JMIR Form Res 2019 Oct 10;3(4):e14052. [doi: 10.2196/14052] [Medline: 31603427]
19. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
20. English C, Ramage ER, Attia J, et al. Secondary prevention of stroke. a telehealth-delivered physical activity and diet pilot randomized trial (ENAbLE-pilot). Int J Stroke 2024 Feb;19(2):199-208. [doi: 10.1177/17474930231201360] [Medline: 37658738]
21. Thurston C, Bezuidenhout L, Humphries S, et al. Mobile health to promote physical activity in people post stroke or transient ischemic attack - study protocol for a feasibility randomised controlled trial. BMC Neurol 2023 Mar 28;23(1):124. [doi: 10.1186/s12883-023-03163-0] [Medline: 36978045]
22. Thurston C, Humphries S, Bezuidenhout L, et al. Mobile health delivered physical activity after mild stroke or transient ischemic attack: is it feasible and acceptable? Int J Stroke 2025 Aug;20(7):801-811. [doi: 10.1177/17474930251315628] [Medline: 39800974]
23. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
24. Sage M, Middleton LE, Tang A, Sibley KM, Brooks D, McIlroy W. Validity of rating of perceived exertion ranges in individuals in the subacute stage of stroke recovery. Top Stroke Rehabil 2013;20(6):519-527. [doi: 10.1310/tsr2006-519] [Medline: 24273299]
25. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
26. Banks JL, Marotta CA. Outcomes validity and reliability of the modified Rankin scale: implications for stroke clinical trials: a literature review and synthesis. Stroke 2007 Mar;38(3):1091-1096. [doi: 10.1161/01.STR.0000258355.23810.c6] [Medline: 17272767]
27. Ahlström I, Hellström K, Emtner M, Anens E. Reliability of the Swedish version of the Exercise Self-Efficacy Scale (S-ESES): a test-retest study in adults with neurological disease. Physiother Theory Pract 2015 Mar;31(3):194-199. [doi: 10.3109/09593985.2014.982776] [Medline: 25418018]
28. Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. Arch Neurol 1989 Oct;46(10):1121-1123. [doi: 10.1001/archneur.1989.00520460115022] [Medline: 2803071]
29. Brown TA, Chorpita BF, Korotitsch W, Barlow DH. Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. Behav Res Ther 1997 Jan;35(1):79-89. [doi: 10.1016/s0005-7967(96)00068-x] [Medline: 9009048]
30. Lin KC, Fu T, Wu CY, Hsieh YW, Chen CL, Lee PC. Psychometric comparisons of the Stroke Impact Scale 3.0 and Stroke-Specific Quality of Life Scale. Qual Life Res 2010 Apr;19(3):435-443. [doi: 10.1007/s11136-010-9597-5] [Medline: 20127418]
31. Blackwood J, Suzuki R, Webster N, Karczewski H, Ziccardi T, Shah S. Use of activPAL to measure physical activity in community-dwelling older adults: a systematic review. Arch Rehabil Res Clin Transl 2022 Jun;4(2):100190. [doi: 10.1016/j.arrct.2022.100190] [Medline: 35756981]
32. Ryan CG, Grant PM, Tigbe WW, Granat MH. The validity and reliability of a novel activity monitor as a measure of walking. Br J Sports Med 2006 Sep;40(9):779-784. [doi: 10.1136/bjsm.2006.027276] [Medline: 16825270]
33. Fini NA, Simpson D, Moore SA, et al. How should we measure physical activity after stroke? An international consensus. Int J Stroke 2023 Oct;18(9):1132-1142. [doi: 10.1177/17474930231184108] [Medline: 37300499]
34. Fini NA, Bernhardt J, Holland AE. Low gait speed is associated with low physical activity and high sedentary time following stroke. Disabil Rehabil 2021 Jul;43(14):2001-2008. [doi: 10.1080/09638288.2019.1691273] [Medline: 31755311]
35. Stens NA, Bakker EA, Mañas A, et al. Relationship of daily step counts to all-cause mortality and cardiovascular events. J Am Coll Cardiol 2023 Oct 10;82(15):1483-1494. [doi: 10.1016/j.jacc.2023.07.029] [Medline: 37676198]
36. Thilarajah S, Mentiplay BF, Bower KJ, et al. Factors associated with post-stroke physical activity: a systematic review and meta-analysis. Arch Phys Med Rehabil 2018 Sep;99(9):1876-1889. [doi: 10.1016/j.apmr.2017.09.117] [Medline: 29056502]
37. Saunders DH, Sanderson M, Hayes S, et al. Physical fitness training for stroke patients. Cochrane Database Syst Rev 2020 Mar 20;3(3):CD003316. [doi: 10.1002/14651858.CD003316.pub7] [Medline: 32196635]
38. Gagnon MA, Batcho CS, Bird ML, Labbé B, Best KL. Feasibility of a remotely supervised home-based group eHealth fitness and mobility exercise program for stroke: French-Canadian version preliminary study. Top Stroke Rehabil 2023 Mar;30(2):169-179. [doi: 10.1080/10749357.2021.2012008] [Medline: 34994303]
39. Gunnes M, Langhammer B, Aamot IL, et al. Adherence to a long-term physical activity and exercise program after stroke applied in a randomized controlled trial. Phys Ther 2019 Jan 1;99(1):74-85. [doi: 10.1093/ptj/pzy126] [Medline: 30329136]

XSL•FO
RenderX

40. Picorelli AMA, Pereira LSM, Pereira DS, Felício D, Sherrington C. Adherence to exercise programs for older people is influenced by program characteristics and personal factors: a systematic review. J Physiother 2014 Sep;60(3):151-156. [doi: 10.1016/j.jphys.2014.06.012] [Medline: 25092418]
41. Yang X, Ma L, Zhao X, Kankanhalli A. Factors influencing user's adherence to physical activity applications: a scoping literature review and future directions. Int J Med Inform 2020 Feb;134:104039. [doi: 10.1016/j.ijmedinf.2019.104039] [Medline: 31865054]
42. Jakob R, Harperink S, Rudolf AM, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371. [doi: 10.2196/35371] [Medline: 35612886]
43. Eaton C, Vallejo N, McDonald X, et al. User engagement with mHealth interventions to promote treatment adherence and self-management in people with chronic health conditions: systematic review. J Med Internet Res 2024 Sep 24;26:e50508. [doi: 10.2196/50508] [Medline: 39316431]
44. Ek S. Gender differences in health information behaviour: a Finnish population-based survey. Health Promot Int 2015 Sep;30(3):736-745. [doi: 10.1093/heapro/dat063] [Medline: 23985248]
45. Wang TC, Tsai AC, Wang JY, et al. Caregiver-mediated intervention can improve physical functional recovery of patients with chronic stroke: a randomized controlled trial. Neurorehabil Neural Repair 2015 Jan;29(1):3-12. [doi: 10.1177/1545968314532030] [Medline: 24788580]
46. Voorberg WH, Bekkers VJJM, Tummers LG. A systematic review of co-creation and co-production: embarking on the social innovation journey. Public Manag Rev 2015 Oct 21;17(9):1333-1357. [doi: 10.1080/14719037.2014.930505]
47. RIKSSTROKE – quality of the Swedish stroke care: a brief summary of data for the full year 2024. : Riksstroke – The Swedish Stroke Register; 2025 URL: https://www.riksstroke.org/wp-content/uploads/2025/06/Arsrapport-2024-engelsk-sammanfattning.pdf [accessed 2026-02-13]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
mHealth: mobile health
OR: odds ratio
STAAR: Stroke Treatment Through Active and Accessible Rehabilitation
TIA: transient ischemic attack

Edited by L Buis; submitted 17.Apr.2025; peer-reviewed by C Madu, D Lee, L Hurley; revised version received 30.Oct.2025; accepted 21.Jan.2026; published 17.Mar.2026.

Please cite as:
Lagerlund H, Bezuidenhout L, Humphries S, Holmlund L, Kwak L, Häger CK, Moulaee Conradsson D
Adherence to and Engagement With an mHealth Physical Activity Intervention After Mild Stroke or Transient Ischemic Attack: Secondary Analysis of a Feasibility Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e75662
URL: https://mhealth.jmir.org/2026/1/e75662
doi:10.2196/75662

© Hanna Lagerlund, Lucian Bezuidenhout, Sophia Humphries, Lisa Holmlund, Lydia Kwak, Charlotte K Häger, David Moulaee Conradsson. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 17.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Comparison Between Browser- and App-Based Versions of a Program for Self-Management of Mild to Moderate Depression: Log Data Analysis of a Convenience Sample

Katharina Scholze[1,2], MSc; Caroline Allenhof[1], Dr rer nat; Ulrich Hegerl[1,2,3], Prof Dr

[1]Depression Research Centre of German Foundation for Depression and Suicide Prevention, Leipzig, Germany

[2]Depression Research Centre of German Foundation for Depression and Suicide Prevention, Department of Psychiatry, Psychosomatics, and Psychotherapy, Goethe University Frankfurt, Frankfurt, Germany

[3]Goethe University Frankfurt, Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, Goethe University Frankfurt, Frankfurt am Main, Germany

Corresponding Author:
Katharina Scholze, MSc
Depression Research Centre of German Foundation for Depression and Suicide Prevention, Leipzig, Germany

Abstract

Background: Internet-based cognitive behavioral therapy (iCBT) for the treatment of depression has proven to be an effective and accessible option. However, iCBTs tend to have low adherence rates, which may negatively impact their effectiveness. One such iCBT program is the browser-based, proven-effective iFightDepression (iFD) tool. An app-based version of the iFD tool is the iFD app, which was developed to improve usability with a smartphone. The iFD app provides enhanced usability and a more optimized user experience on mobile devices. Additionally, it offers more comfortable interaction with worksheets, reduced text with added videos, and quicker access to the program via the smartphone icon. These improved usability on smartphones and could have an impact on adherence.

Objective: This study investigated (1) whether adherence parameters, that is, the *number of worksheets*, the *number of sessions*, and the *number of workshops*, significantly differed between users of the iFD app and the iFD tool and (2) exploratorily whether symptom reduction (delta Patient Health Questionnaire-9 [PHQ-9] scores) differed after 5 to 9 weeks between the iFD app and the iFD tool, after controlling for covariates.

Methods: We used *t* tests to compare data from 56 participants using the iFD app for 8 weeks with data from 172 participants using the iFD tool in a previous 6-week study. Exploratively, symptom reduction was compared between formats. A multiple regression model was calculated with the delta PHQ-9 score as the dependent variable and format, baseline PHQ-9 score, adherence, current psychotherapy, antidepressants, age, and sex as independent variables.

Results: There was no significant difference between the iFD tool and the iFD app in terms of the *number of sessions per week* ($t_{67.393}$=0.920; *P*=.36; corrected *P*=.36), the *number of workshops* ($t_{76.368}$=–1.217; *P*=.30; corrected *P*=.36), and the *number of worksheets per week* (t_{74}=0.984; *P*=.33; corrected *P*=.36). We found no difference in delta PHQ-9 scores between the iFD app and the iFD tool, and baseline PHQ-9 scores were the only significant predictor (*b*=–0.61; *P*<.001).

Conclusions: Despite the improved availability of the app version in daily life, there were no significant differences in the use parameters we analyzed and no differences in symptom reduction. This study provides the first evidence that adherence to iCBT content is comparable for browser- and app-based interventions and that symptom reduction is similar for both formats. However, this study used a convenience sample, and therefore, the results must be interpreted with caution. Notably, in the study of the iFD tool, guidance was provided by the study assistants in a standardized manner. In the pilot study on the iFD app, the amount of guidance varied substantially, as it was provided by the participants' health care practitioners. These differences in guidance could also have an influence on adherence.

(JMIR Mhealth Uhealth 2026;14:e58835) doi:10.2196/58835

KEYWORDS

iCBT; internet-based cognitive behavioral therapy; internet-based treatment; internet- and mobile-based intervention; mobile app; depression; cognitive behavioral therapy; smartphone app; app; usability; data analysis; regression model

XSL•FO
RenderX

Introduction

Internet-based cognitive behavioral therapy (iCBT) for the treatment of depression has proven to be an effective and accessible option [1]. However, adherence rates tend to be low [2,3], which may negatively impact effectiveness [4]. One such iCBT is the browser-based, proven-effective [5] iFightDepression (iFD) tool, which has been available since 2014. To improve its usability on smartphones and thus potentially increase adherence, an app-based version of the iFD tool, the iFD app, was developed based on the content of the iFD tool. The iFD tool is accessible via mobile web browsers on smartphones; however, the iFD app provides enhanced usability and a more optimized user experience on mobile devices. Further advantages are that the worksheets can be filled in more conveniently, the amount of text has been reduced in favor of new information videos, and the access is faster through the icon on the smartphone.

Smartphone apps are particularly attractive for health-related applications because, unlike desktop computers or laptops, people tend to carry their smartphones everywhere and to keep them on at all times. In a survey, participants were divided into focus groups and interviewed using open-ended questions about their experiences with apps designed to support a healthy lifestyle [6]. The focus groups were recorded, transcribed, and analyzed to identify frequent topics. One topic that frequently emerged was that participants appreciated health apps on smartphones because they provide a "quick and efficient link to a wealth of information 'on the go'."

This raises the question of whether the improved availability and usability of a program on a smartphone has a positive effect on adherence and thus potentially on symptom reduction. Although the dose-response relationship remains a subject of debate due to mixed findings [4], some studies have found no significant impact of adherence on efficacy [7-9], while many others indicate a positive correlation between adherence and treatment effectiveness [10-12]. Nevertheless, it is certainly worthwhile to investigate potential factors that could improve adherence. Outside the field of depression, the evidence appears to support the assumption that a program delivered as a smartphone app may enhance treatment adherence. A study comparing an app to a browser-based and paper diary–based weight loss program showed that the app group had a higher frequency of program use than the browser-based and the paper diary–based program [13]. However, in this case, the content of the 2 programs was similar but not identical. The study was not statistically powered to detect a change in weight loss; however, the results were displayed to give an idea of the effect size. After 6 months, participants in the smartphone group demonstrated a statistically significantly greater reduction in body weight compared with those in the website group. However, no significant difference was observed when comparing the smartphone group to the paper diary group.

To the best of our knowledge, there is only 1 study regarding digital interventions on depression comparing app-based vs browser-based access to identical programs. Watts et al [14] compared the adherence of a mobile app group with a computer group in a program for depression. However, in this study, only the *number of completed lessons* was analyzed and found not to be significantly different based on the format. In this study, we examine various use parameters beyond just the *number of completed lessons*. The frequency of log-ins is also a noteworthy factor, as frequent use may indicate a deeper integration of these programs into daily routines. The number of completed worksheets is also a significant parameter, as a higher completion rate may suggest more active engagement with the content. In combination, the various use parameters provide a more comprehensive insight into how engaged and active users are. This allows us to provide a more informed comparison of adherence between the two formats. The aim of our study is to discover whether the app-based format can achieve higher adherence rates than the browser-based format. The following questions were investigated in this study:

1. Do adherence parameters, that is, the *number of completed worksheets*, the *number of sessions*, and the *number of workshops completed*, significantly differ between users of the iFD app and the iFD tool?
2. Does symptom reduction (delta Patient Health Questionnaire-9 [PHQ-9] score) differ between users of the iFD app and the iFD tool after 5 to 9 weeks, after controlling for covariates such as, age, sex, and adherence (this was an exploratory question)?

Methods

Study Design

This study is a secondary analysis of data from 2 different studies. We analyzed data from 56 participants from a pilot study (authors KS, CA, and UH; unpublished data, April 2023) of the newly developed iFD app, along with data from 172 participants from the intervention group of a randomized controlled trial (RCT) using the iFD tool published in 2020 [5]. Participants of the pilot study were instructed to use the iFD app for 8 weeks in addition to treatment as usual (TAU). TAU could include psychotherapy, antidepressants, or watchful waiting. The pilot study was conducted from June 2022 to January 2023. Participants in the intervention group of the RCT (who used the iFD tool) were instructed to use the iFD tool for 6 weeks. All participants were treated by either a physician or a psychotherapist. The RCT was conducted between June 2016 and August 2018.

Sample and Recruitment

The participants of the pilot study of the iFD app were referred by practitioners experienced with the iFD tool who agreed to test the app with their patients. Participants of the pilot study were aged 18 years or above and received a diagnosis of mild to moderate depression (F32.0, F32.1, F32.9, F33.0, F33.1, and F33.9) or dysthymia (F34.1) from a physician or psychotherapist. In Germany, both licensed psychological psychotherapists and physicians are authorized to diagnose depressive disorders. The practitioners provided the patients' diagnosis to the study team after the patients had given their written informed consent. At the beginning of the study, participants had a PHQ-9 [15,16] score between 5 and 19.

XSL•FO
RenderX

Recruitment for the RCT of the iFD tool took place throughout Germany via the online channels of the German Foundation of Depression and Suicide Prevention. The participants of the intervention group from the RCT were aged 18 years or above and had a current diagnosis of mild or moderate depressive disorder (F32.0, F32.1, F33.0, and F33.1) or dysthymia (F34.1), according to the Mini International Neuropsychiatric Interview and PHQ-9 (score 5‐14), indicating mild to moderate symptoms.

Sociodemographic data for users of the iFD app and the iFD tool are summarized in Table 1.

Table . Overview of sociodemographic characteristics.

Variables	Study iFightDepression app (n=56)	Study iFightDepression tool (n=172)	Comparison, *P* value (BH[a]-adjusted *P* value)
Age (years)			.97 (.97)
Mean (SD)	42.79 (14.39)	42.86 (12.36)	
Median (IQR)	44.5 (29.0-57.0)	43.5 (33.8-52.0)	
Female, n (%)	34 (60.7)	137 (79.6)	<.001 (<.001)[b]
Baseline PHQ-9[c] score, mean (SD)	11.84 (4.12)	9.81 (3.51)	.002 (.002)[d]
Current psychotherapy use, n (%)	40 (71.4)	97 (56.4)	.07 (.08)
Current antidepressant use, n (%)	27 (48.2)	114 (66.3)	<.001 (<.001)[b]

[a]BH: Benjamini-Hochberg.

[b]These results are significant (*P*<.001).

[c]PHQ-9: Patient Health Questionnaire-9.

[d]This result is significant (*P*<.01).

Ethical Considerations

The study protocol for the RCT of the iFD tool was reviewed and approved by the Ethics Committee at the Faculty of Medicine, University of Leipzig, on February 11, 2015 (case number 080-15-09032015). The protocol of the pilot study of the iFD app was reviewed and approved by the Ethics Committee at the Faculty of Medicine, Goethe University Frankfurt, on June 11, 2021 (reference number 2021-53). Furthermore, we received confirmation from the Ethics Committee at the Faculty of Medicine, Goethe University Frankfurt, on June 4, 2025 (reference number 2025-2467-Waiver), that no further approval from the ethics committee was required for the secondary analysis. The original informed consent obtained from participants for both studies covered the secondary analysis without the need for additional consent. Data were used in a pseudonymized form for the statistical analysis. The interventions were offered free of charge, and no reimbursement was offered to participants.

Guidance

Guidance for the iFD app was provided by the practitioners managing the participants' TAU. The study team asked the practitioners to check in with participants at each appointment during TAU about their experience with the iFD app and to discuss how their work with it was progressing. The practitioners were qualified iFD guides, that is, they had completed a 70-minute webinar and actively used the iFD tool. Few practitioners were using the iFD tool for years and therefore completed the webinar several years ago. Among other things, the iFD workshops were explained in detail in the webinar. Although the webinar referred to the iFD tool, the content of the workshops in the iFD app has not changed. Shortly before the start of the study, the practitioners were given a brief introduction to the iFD app, where differences from the iFD tool were explained. Furthermore, calls at 3 time points were scheduled by the study assistants to collect feedback on the app and discuss any changes in treatment. The study assistants also asked how often and for how many minutes they had spoken to the practitioners about the iFD app to estimate the amount of guidance provided by the practitioner.

The guidance for the iFD tool was provided by the study assistants (psychologists and psychotherapists). The study assistants guided the participants in 5 telephone calls, which were supervised by a senior psychiatrist who was involved in the development of the iFD tool. All study assistants were qualified through the standard webinar process and used a guide for the calls based on the webinar content. The focus of the guideline calls was to motivate the participants rather than discuss the content. The study assistant then noted the duration of the phone calls to determine the amount of guidance.

Measures

The PHQ-9 completion was mandatory at first log-in in both studies and could be completed again at any time during the intervention. Once a week, patients were asked to complete the questionnaire to monitor their symptoms. The instructions in the iFD tool were adjusted so that the measurement now referred to the period starting from the last time the PHQ-9 was completed. In the iFD app, the original instruction—referring to the past 2 weeks—was retained. The sum scores of the baseline PHQ-9 and the last PHQ-9 completed within the time span of 5 to 9 weeks after registration were used in the current analysis. This time span was chosen based on the time span of the 2 studies (6-week RCT iFD tool and 8-week pilot study iFD app, with a buffer of 1 week).

XSL•FO
RenderX

Anonymized log files were used to examine user behavior. These log files contained time-stamped logs of all activity within the iFD tool or app. From these, composite measures of use were created. We determined the *number of sessions per week*, *the number of workshops completed* (at least 70.0% of the material was seen or read), and *the number of completed worksheets per week* to measure and compare adherence.

Statistical Analyses

For the comparison of the parameters of adherence between the iFD tool and the iFD app, we used the Welch 2-sample *t* test, chosen for its robustness with unequal variances [17]. *P* values were corrected using the Benjamini-Hochberg method [18].

To compare symptom reduction, a multiple regression analysis was performed with *age*, sex, *baseline PHQ-9*, *format* (app vs browser), *workshops completed*, *sessions per week*, *completed worksheets per week*, *amount of guidance* (min), *current psychotherapy* (yes or no), and *current antidepressants* (yes or no) as predictor variables for the delta PHQ-9. Delta PHQ-9 refers to the mean difference between PHQ-9 scores at registration and the last PHQ-9 filled 5 to 9 weeks after registration. It was used as a measure of symptom change. Negative values indicate a reduction in symptoms, while positive values reflect an increase. Only participants who completed the PHQ-9 questionnaire at least once within 5 to 9 weeks were included in the analysis.

The assumptions of multiple regression were assessed and tested. Normality of the residuals was evaluated using graphical methods, specifically a histogram and a Q-Q plot. The histogram displayed a bell-shaped curve characteristic of a Gaussian distribution, indicating approximate normality. In the Q-Q plot, the majority of data points aligned closely with the reference line, with several lying directly on it. However, the Shapiro-Wilk test indicated a significant deviation from normality (P=.01). Given the known high statistical power and resulting high sensitivity of the Shapiro-Wilk test in larger samples (N>50) to even the slightest deviations from normality, the graphical results were considered sufficient to justify the use of linear regression [19-22]. The nonsignificant Breusch-Pagan test (P=.70) indicated that homoscedasticity was present [23]. All predictors were correlated with each other in pairs to test for multicollinearity. The highest correlation (0.77) was observed between the *number of sessions* and the *number of completed worksheets*. This indicated that multicollinearity was not present, as there is an assumption that correlation values above 0.8 are a sign of multicollinearity [24]. However, the variance inflation factor (VIF) in the full model with the interactions indicates multicollinearity for the predictors' *format*, *completed worksheets*, and *number of sessions*, as well as the interactions (VIF>10) [25]. Accordingly, we conducted a stepwise hierarchical regression analysis to localize multicollinearity and compare the models. In the first step (model 1), only the covariates (age, sex, amount of guidance, current psychotherapy, and current antidepressant use) were included as predictors in the model. In the second step (model 2), the format was added. In step 3 (model 3), the use parameters were included. Models 1 to 3 indicated no critical multicollinearity (VIF <2.6). In step 4 (model 4), interaction terms between format and use parameter were included. The interactions did not significantly improve the model and led to increased multicollinearity (VIF >10) [25]. Therefore, a regression model without interactions was used (Table 2). The full model including interaction terms is available in Table S2 in Multimedia Appendix 1.

Table . Summary of the multiple regression model predicting delta Patient Health Questionnaire-9 scores (negative values indicate a reduction in symptoms; N=167; R^2[a]=0.26; adjusted R^2=0.21).

	B[b]	*P* value
Constant	3.71	.10
Age (years)	–0.01	.66
Sex (male)	–0.39	.59
Antidepressants (yes)	0.44	.48
Psychotherapy (yes)	0.46	.47
Baseline Patient Health Questionnaire-9	–0.61	<.001[c]
Guidance	0.02	.24
Format (browser)	–1.48	.14
Worksheets completed	–0.05	.43
Number of sessions	0.04	.78
Workshops completed	0.06	.83

[a]R^2: explained variance of the model.

[b]B: unstandardized coefficients.

[c]This result is significant (P<.001).

Statistical analyses were performed using R (R Foundation for Statistical Computing) statistics [26]. The significance level was set at α=.05.

XSL•FO
RenderX

Results

Comparison of Adherence Parameters

We found no significant differences between the iFD tool and the iFD app in terms of the *number of sessions per week* ($t_{67.393}$=0.920; *P*=.36; corrected *P*=.36), the *number of workshops completed* ($t_{76.368}$=–1.217; *P*=.30; corrected *P*=.36), and the *number of completed worksheets per week* ($t_{74.099}$=0.984; *P*=.33; corrected *P*=.36).

Comparison of Symptom Reduction

Overall, 37 (n=56, 66.1%) participants from the iFD app pilot study and 130 (n=172, 75.6%) participants from the RCT of the iFD tool completed the PHQ-9 at least once within 5 to 9 weeks. The sociodemographic data of the completers are summarized in Table S1 in Multimedia Appendix 1. Table 3 shows the PHQ-9 scores in both groups before and after the intervention. Table 4 presents a comparison of sociodemographic characteristics between participants who completed at least 1 PHQ-9 assessment within 5 to 9 weeks (completers) and those who did not and were excluded from further analysis (noncompleters).

Table . Patient Health Questionnaire-9 (PHQ-9) scores at the start of the intervention and after the participants had completed the PHQ-9 at least once within 5 to 9 weeks[a].

Group	PHQ-9: intervention start, mean (SD)	Last PHQ-9 after 5-9 weeks, mean (SD)	Within-group effect size[b] (95% CI)
Study iFD app (n=37)	12.62 (4.11)	8.84 (4.51)	0.88 (0.39 to 1.36)
Study iFD tool (n=130)	9.52 (3.37)	6.98 (3.89)	0.70 (0.45 to 0.95)

[a]The between-group effect size comparing the iFD app and iFD tool groups was 0.27 (95% CI –0.09 to 0.64).

[b]Effect size is calculated as Cohen *d*.

Table . Comparison of sociodemographic characteristics between Patient Health Questionnaire-9 (PHQ-9) completers (at least 1 PHQ-9 completion within 5-9 weeks) and noncompleters.

Variables	Completers (n=167)	Noncompleters (n=61)	Comparison, *P* value (BH[a]-adjusted *P* value)
Age (years), mean (SD); median (IQR)	43.45 (12.98); 44.0 (33.5-53.0)	41.18 (12.45); 41 (31.0-50.0)	.23 (.55)
Female, n (%)	126 (75.4)	45 (73.8)	.93 (.93)
Baseline PHQ-9 score, mean (SD)	10.2 (3.76)	10.8 (3.70)[b]	.33 (.55)
Current psychotherapy, n (%)	96 (57.5)	41 (67.2)	.24 (.55)
Current antidepressants, n (%)	102 (61.1)	39 (63.9)	.81 (.93)

[a]BH: Benjamini-Hochberg.

[b]n=58.

The multiple regression analysis revealed no significant difference in symptom reduction between the iFD app and the iFD tool (Table 4). The predictors accounted for 25.8% (adjusted R^2=0.21) of the variance in delta PHQ-9. The baseline PHQ-9 score was the only significant predictor of change in delta PHQ-9 (*P*<.001; *b*=–0.61). For a comprehensive overview of the full model results, including all covariates prior to the stepwise regression, see Table S2 in Multimedia Appendix 1.

Discussion

Principal Findings

In this study, we compared adherence based on various use parameters between the app version and the browser-based version of an iCBT for the treatment of depression. Although the app version was expected to offer improved accessibility through enhanced smartphone usability and therefore an improved possibility of using the program everywhere and “on the go,” no significant differences in use parameters were observed. Watts et al [14] similarly reported no significant differences in the *number of sessions* between the app and browser versions. Although our study included a broader set of use parameters to gain a more comprehensive insight into user engagement and activity, we arrived at comparable findings. This study also found no significant differences in symptom reduction between the iFD app and the iFD tool, which is not surprising given the similar adherence. Although in a survey many participants stated that they value health apps on their smartphone because they provide a quick and efficient connection to a wealth of information “on the go” [6], this does not seem to contribute to their being used more actively and more frequently than browser-based programs. It is possible that the iFD tool was frequently accessed via a smartphone, which allowed the use of the program “on the go” too, despite the tool’s limited usability on smartphones compared with the iFD app. However, it is notable that in the RCT of the iFD tool, guidance was provided by the study assistants in a standardized manner. In the pilot study of the iFD app, guidance for participants was provided by their practitioners, and the extent of guidance varied considerably; the guidance often being

XSL•FO
RenderX

considerably lower than that for the browser-based version, as time is very limited due to a high volume of patients in routine care, a factor that could contribute to reduced adherence [27,28]. In the pilot study, participants reported spending an average of 13.8 minutes over 8 weeks discussing the app with their practitioner. In contrast, with the iFD tool, study assistants spent an average of 29.76 minutes speaking with the participants about the tool over the course of 6 weeks. Therefore, achieving a similar level of adherence nonetheless could still be an indicator of good usability of the iFD app.

Limitations

The results from this study should be interpreted with caution as they are based on convenience samples with differing recruitment strategies, and the sample size of the pilot study was limited. Additionally, several limitations should be considered. The amount of guidance was included as a control variable; however, in the iFD app study, it was only estimated by participants, whereas in the iFD tool study, it was objectively measured. This limits comparability but still allows for controlling for broad differences. Furthermore, the PHQ-9 instructions differed between the studies: the iFD app participants were asked to rate their symptoms for the past 2 weeks (standard instruction), while participants in the iFD tool study were asked to rate symptoms since their last PHQ-9 completion. In addition, the period between completion of the baseline PHQ-9 and the last PHQ-9 varied between weeks 5 and 9, as the last completed PHQ-9 was included in the analysis and the study duration differed between the studies (iFD app: 8 weeks; iFD tool: 6 weeks). This combination of different instructions and varying periods may limit the comparability of the PHQ-9 changes between the studies. Although no significant differences in symptom reduction were found, these methodological differences remain a limitation. In addition, the iFD tool offers the option of working with self-printed worksheets that could not be measured in the study, which means that the comparability of the *number of worksheets* could also be limited if this option was used frequently. Furthermore, it was not feasible to monitor whether the iFD tool was used on mobile phones—a possibility that exists but does not offer as good a usability experience compared with the iFD app. Additionally, the 2 studies were conducted 4 years apart (iFD tool: 2016 - 2018; iFD app: 2022 - 2023), during which digital health use and awareness increased [29]. Legislative changes (eg, the Digitale-Versorgungs-Gesetz in 2019, which enables practitioners to prescribe digital interventions as part of routine medical care [30]) and the COVID-19 pandemic (in-person appointments were often not possible) may have contributed to a greater acceptance and use of digital interventions during the period in which the pilot study was conducted.

Conclusions

This study provides initial evidence that adherence to iCBT content is comparable for browser- and app-based interventions and that symptom reduction is similar for both formats. Despite the limitations of this secondary analysis, it nonetheless offers a rare opportunity to directly compare the same intervention delivered via a mobile app and a browser-based format. The superiority of apps in terms of adherence seems plausible, but this could not be confirmed in this study. Future studies are needed to establish an evidence base for selecting the optimal format.

Acknowledgments

The authors would like to thank the European Alliance Against Depression (EAAD) for their support of the iFightDepression tool. Furthermore, they would like to express their gratitude to all the participants for taking part in the studies.

Funding

The randomized controlled trial in this secondary analysis was funded within the framework of cooperation between the German Foundation for Depression and Suicide Prevention and the Deutsche Bahn Stiftung GmbH. For the pilot study and the secondary analysis itself, no external funding or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this paper.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

KS, CA, and UH were involved in conceptualizing the secondary analysis. CA and KS contributed to data collection for the primary studies used in the secondary analysis. KS conducted the analyses and wrote the original draft under the close supervision of CA. Visualizations were created by KS. UH supervised the development of the original draft. Review and editing were carried out by KS under the supervision of UH and CA.

Conflicts of Interest

The iFightDepression tool was developed by the European Alliance Against Depression (EAAD), where UH is the chairman of the board. In Germany, the iFightDepression tool is operated, continuously developed, and evaluated by the German Foundation for Depression and Suicide Prevention, a member organization of the EAAD. KS and CA are employed by the German Foundation for Depression and Suicide Prevention. The German Foundation for Depression and Suicide Prevention is an independent nonprofit

XSL•FO
RenderX

foundation under civil law, financed primarily by donations, endowments, grants, and third-party funding for projects and research. It also receives income from its business operations but works independently of the pharmacological industry.

Multimedia Appendix 1
Sociodemographic characteristics of Patient Health Questionnaire-9 completers and full regression model–inclusive interaction terms.
[DOCX File, 23 KB - mhealth_v14i1e58835_app1.docx]

References

1. Cuijpers P, Noma H, Karyotaki E, Cipriani A, Furukawa TA. Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: a network meta-analysis. JAMA Psychiatry 2019 Jul 1;76(7):700-707. [doi: 10.1001/jamapsychiatry.2019.0268] [Medline: 30994877]
2. Christensen H, Griffiths KM, Korten AE, Brittliffe K, Groves C. A comparison of changes in anxiety and depression symptoms of spontaneous users and trial participants of a cognitive behavior therapy website. J Med Internet Res 2004 Dec 22;6(4):e46. [doi: 10.2196/jmir.6.4.e46] [Medline: 15631970]
3. Farrer L, Christensen H, Griffiths KM, Mackinnon A. Internet-based CBT for depression with and without telephone tracking in a national helpline: randomised controlled trial. PLoS ONE 2011;6(11):e28099. [doi: 10.1371/journal.pone.0028099] [Medline: 22140514]
4. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 5;13(3):e52. [doi: 10.2196/jmir.1772] [Medline: 21821503]
5. Oehler C, Görges F, Rogalla M, Rummel-Kluge C, Hegerl U. Efficacy of a guided web-based self-management intervention for depression or dysthymia: randomized controlled trial with a 12-month follow-up using an active control condition. J Med Internet Res 2020 Jul 14;22(7):e15361. [doi: 10.2196/15361] [Medline: 32673233]
6. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86. [doi: 10.2196/jmir.2583] [Medline: 23598614]
7. Clarke G, Eubanks D, Reid E, et al. Overcoming Depression on the Internet (ODIN) (2): a randomized trial of a self-help depression skills program with reminders. J Med Internet Res 2005 Jun 21;7(2):e16. [doi: 10.2196/jmir.7.2.e16] [Medline: 15998607]
8. Clarke G, Kelleher C, Hornbrook M, Debar L, Dickerson J, Gullion C. Randomized effectiveness trial of an Internet, pure self-help, cognitive behavioral intervention for depressive symptoms in young adults. Cogn Behav Ther 2009;38(4):222-234. [doi: 10.1080/16506070802675353] [Medline: 19440896]
9. Spek V, Cuijpers P, Nyklícek I, et al. One-year follow-up results of a randomized controlled clinical trial on internet-based cognitive behavioural therapy for subthreshold depression in people over 50 years. Psychol Med 2008 May;38(5):635-639. [doi: 10.1017/S0033291707002590] [Medline: 18205965]
10. Andersson G, Bergström J, Holländare F, Carlbring P, Kaldo V, Ekselius L. Internet-based self-help for depression: randomised controlled trial. Br J Psychiatry 2005 Nov;187:456-461. [doi: 10.1192/bjp.187.5.456] [Medline: 16260822]
11. Meyer B, Berger T, Caspar F, Beevers CG, Andersson G, Weiss M. Effectiveness of a novel integrative online treatment for depression (Deprexis): randomized controlled trial. J Med Internet Res 2009 May 11;11(2):e15. [doi: 10.2196/jmir.1151] [Medline: 19632969]
12. van Straten A, Cuijpers P, Smits N. Effectiveness of a web-based self-help intervention for symptoms of depression, anxiety, and stress: randomized controlled trial. J Med Internet Res 2008 Mar 25;10(1):e7. [doi: 10.2196/jmir.954] [Medline: 18364344]
13. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013 Apr 15;15(4):e32. [doi: 10.2196/jmir.2283] [Medline: 23587561]
14. Watts S, Mackenzie A, Thomas C, et al. CBT for depression: a pilot RCT comparing mobile phone vs. computer. BMC Psychiatry 2013 Feb 7;13:49. [doi: 10.1186/1471-244X-13-49] [Medline: 23391304]
15. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
16. Löwe B, Spitzer RL, Zipfel S, Herzog W. Patient Health Questionnaire (PHQ) — German Version Manual and Materials, 2nd edition: Pfizer; 2002. [doi: 10.3233/ZMP-2003-12_2_08]
17. Rasch D, Kubinger KD, Moder K. The two-sample t test: pre-testing its assumptions does not pay off. Stat Papers 2011 Feb;52:219-231. [doi: 10.1007/s00362-009-0224-x]
18. Félix VB, Menezes AF. Comparisons of ten corrections methods for t-test in multiple comparisons via Monte Carlo study. Electron J Appl Stat Anal 2018;11(1):74-91. [doi: 10.1285/i20705948v11n1p74]
19. Razali NM, Wah YB. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. J Stat Model Anal 2011;2(1):21-33 [FREE Full text]

XSL•FO
RenderX

20. Royston P. Approximating the Shapiro-Wilk W-test for non-normality. Stat Comput 1992 Sep;2:117-119. [doi: 10.1007/BF01891203]
21. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. Int J Endocrinol Metab 2012;10(2):486-489. [doi: 10.5812/ijem.3505] [Medline: 23843808]
22. Field A. Discovering Statistics Using IBM SPSS Statistics, 5th edition: SAGE Publications; 2018.
23. Breusch TS, Pagan AR. A simple test for heteroscedasticity and random coefficient variation. Econometrica 1979 Sep;47(5):1287-1294. [doi: 10.2307/1911963]
24. Judge GG, Griffiths WE, Hill RC, Lütkepohl H, Lee TC. The Theory and Practice of Econometrics: John Wiley & Sons; 1991.
25. Dormann CF, Elith J, Bacher S, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 2013 Apr;36(1):27-46. [doi: 10.1111/j.1600-0587.2012.07348.x]
26. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2013. URL: https://www.r-project.org/ [accessed 2026-02-24]
27. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
28. Castro A, López-Del-Hoyo Y, Peake C, et al. Adherence predictors in an internet-based intervention program for depression. Cogn Behav Ther 2018 May;47(3):246-261. [doi: 10.1080/16506073.2017.1366546] [Medline: 28871896]
29. Adler J, Van Brunt D. It is time to realize the promise of the digital mental health transformation: application for population mental health. J Med Internet Res 2025 Jun 19;27:e63791. [doi: 10.2196/63791] [Medline: 40536919]
30. Digitale-Versorgung-Gesetz (DVG). Bundesministerium für Gesundheit. 2019. URL: https://www.bundesgesundheitsministerium.de/service/gesetze-und-verordnungen/detail/digitale-versorgung-gesetz-dvg.html [accessed 2025-07-22]

Abbreviations

iCBT: internet-based cognitive behavioral therapy
iFD: iFightDepression
PHQ-9: Patient Health Questionaire-9
RCT: randomized controlled trial
TAU: treatment as usual
VIF: variance inflation factor

Edited by L Buis; submitted 27.Mar.2024; peer-reviewed by A Baumeister, P Toczyski; accepted 03.Feb.2026; published 20.Mar.2026.

Please cite as:
Scholze K, Allenhof C, Hegerl U
Comparison Between Browser- and App-Based Versions of a Program for Self-Management of Mild to Moderate Depression: Log Data Analysis of a Convenience Sample
JMIR Mhealth Uhealth 2026;14:e58835
URL: https://mhealth.jmir.org/2026/1/e58835
doi:10.2196/58835

© Katharina Scholze, Caroline Allenhof, Ulrich Hegerl. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 20.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Promoting Self-Regulated Social Media Use on Smartphones With a Mobile Intervention App (Wellspent): Randomized Controlled Trial

Lea Mertens[1], MSc; Lina Christin Brockmeier[2,3], MSc; Christina Roitzheim[4], MSc; Theda Radtke[5], Prof Dr; Tilman Dingler[6], Dr rer nat; Jan Keller[2,7], Prof Dr

[1]Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

[2]Department of Education and Psychology, Freie Universität Berlin, Habelschwerdter Allee 45, Berlin, Germany

[3]School of Computing and Information Systems, University of Melbourne, Melbourne, Australia

[4]Wellspent GmbH, Berlin, Germany

[5]Health Psychology and Applied Diagnostics, University of Wuppertal, Wuppertal, Germany

[6]Department of Sustainable Design Engineering, Delft University of Technology, Delft, The Netherlands

[7]Department of Psychology, Heidelberg University, Heidelberg, Germany

Corresponding Author:
Jan Keller, Prof Dr
Department of Education and Psychology, Freie Universität Berlin, Habelschwerdter Allee 45, Berlin, Germany

Abstract

Background: Problematic social media use has been linked to reduced well-being and impulse control difficulties. While digital self-control apps show potential for reducing general app usage, they often lack customization, leading to limited effectiveness and increased user resistance. Their impact on problematic social media use remains uncertain.

Objective: This study evaluates the effectiveness of the Wellspent app, a customizable mobile intervention app designed to promote self-regulated social media use by targeting user-defined problematic app use and offering tailored behavioral nudges.

Methods: In a 3-week randomized controlled trial, 70 iPhone users (mean age 26.2, SD 5.6 years; 47/70, 67% female), regularly using at least 1 social media app, were randomly assigned to an intervention (n=35) or control group (n=35). The intervention group received personalized full-screen reminders with the option to quit or continue social media app use whenever an app session exceeded a self-defined time limit. Participants completed weekly online surveys measuring problematic social media use, problematic smartphone use, self-efficacy, and daily screen time on their most problematic app. Linear mixed models tested intervention effects.

Results: While no significant reduction in problematic social media use or increase in self-efficacy was observed, the intervention group showed a significant reduction in daily screen time on their most problematic app by approximately 29 minutes (estimate=−29.35, SE 6.84, 95% CI −42.79 to –15.99; *P*<.001), and a significant decrease in perceived problematic smartphone use (estimate=−0.46, SE 0.18, 95% CI −0.80 to –0.11; *P*=.01).

Conclusions: The Wellspent app demonstrated short-term efficacy in reducing problematic smartphone use. By allowing users to tailor interventions to their personal goals, the app shows promise as a self-directed tool to support healthier digital habits. Further research should explore long-term effects and feature-specific impacts.

Trial Registration: German Clinical Trials Register DRKS00031767; https://drks.de/search/en/trial/DRKS00031767/details

(JMIR Mhealth Uhealth 2026;14:e56824) doi:10.2196/56824

KEYWORDS

app usage; computer use; digital self-control tool; intervention app; mixed model; nudging; online survey; problematic social media use; randomized controlled trial; screen time; self-efficacy; smartphone use; social media; technology use; well being

Introduction

Overview

In 2022, nearly 60% of the world's population spent an average of 2 hours and 27 minutes per day on social media platforms [1]. While these platforms offer various positive benefits, like providing a sense of closeness or inspiration [2], they also possess [1] features that can foster habitual engagement and contribute to cycles of instant gratification, thereby potentially impeding the pursuit of long-term goals [3].

XSL•FO
RenderX

This duality has spurred an expanding body of research examining the complex relationship between social media use and well-being, highlighting the importance of individual, contextual, and content-related factors in shaping its impact [4-9]. Although findings on the relationship between total time spent on social media and well-being remain inconsistent [7,10], problematic social media use, characterized by excessive, compulsive, or poorly regulated engagement that disrupts daily functioning, has been consistently associated with reduced well-being [7,11,12]. In particular, problematic social media use has been associated with adverse health outcomes, including reduced life satisfaction, increased symptoms of depression, loneliness, and stress [11], as well as lower levels of physical activity [9] and decreased productivity [13]. A meta-analysis by Cheng et al [14] further illustrates the global relevance of this issue, estimating that the prevalence of problematic social media use among adults and adolescents reaches up to 31% in the most affected countries.

Healthy and Problematic Social Media Use

A healthy balance between connection and disconnection is a key aspect of "digital well-being," which has been defined as an "optimal balance between the benefits and drawbacks obtained from mobile connectivity" that is achieved "when experiencing maximal controlled pleasure and functional support, together with minimal loss of control and functional impairment" [15]. Problematic social media use, on the other hand, is characterized by a compulsive pattern of social media use that interferes with daily functioning [16]. It displays elements of addictive behaviors, including cognitive salience, mood modification, tolerance, withdrawal, conflict, and relapse [17,18].

The Integrative Model of Mobile Media Use and Need Experiences (IM³UNE model) highlights self-control—the ability to override or inhibit immediate desires in favor of goal-directed behavior [19]—as an important factor in the relationship between mobile connectivity and digital well-being. High levels of self-control protect users from excessive or habitual social media use [20], while deficits in self-control are closely linked to problematic social media use [21].

To better understand self-control in both digital and nondigital environments, research has drawn on the dual-process model. This framework posits 2 modes of information processing: an automatic, impulsive, and unconscious process (System 1), and a reflective, deliberate, and conscious process (System 2) [22]. In the context of problematic social media use, System 1 reflects the habitual, cue-driven use of social media, which is frequently associated with adverse well-being outcomes [23,24].

Specific design features of social media platforms, often referred to as "dark patterns," are deliberately constructed to exploit System 1 processing and maximize user engagement [25]. For instance, infinite scrolling mechanisms create an illusion of endless content and contribute to compulsive engagement by removing natural stopping cues [26]. Once immersed in these apps, disengaging becomes increasingly difficult, as System 2, despite recognizing the negative consequences of such use, often fails to override automatic behavior. This interplay between automatic engagement and diminished self-regulation underscores the need for interventions that facilitate reflective, intentional technology use. As a result, digital self-control tools have emerged to help individuals regulate their social media use and regain control over their behavior [27].

Digital Self-Control Tools

Digital self-control tools aim to prompt System 2–driven behavior to support users in regaining self-control over their smartphone use. These tools, often informed by interventionist theories like nudge theory [28], guide users' choices and behaviors in the digital environment through behavioral interventions. Many digital self-control tools operate as self-nudges, enabling users to set up choice architectures that protect their goals from competing temptations [29]. Most of these interventions often use a "just-in-time" approach to intervene precisely when overuse is likely to occur [30].

Several apps have been proposed in recent years with varying techniques and degrees of enforcement [31-36]. For instance, the app MyTime offers real-time feedback on time spent across apps and prompts goal reflection, which led to reduced daily use and increased perceived self-regulation in a randomized trial with college students [31]. MindPhone, a smartphone app that interrupts habitual unlocking with intention prompts, was found to increase awareness and reduce habitual use over 14 days, especially among users with initially higher problematic smartphone use scores [36]. One Sec delays app access with a short breathing pause and prompt. A within-subject field study showed that this design significantly reduced Instagram use and increased perceived control without notable user frustration [35].

Despite promising developments, existing digital self-control tools continue to face critical limitations regarding long-term effectiveness and sustained user engagement [37-39]. Many tools rely on overly restrictive mechanisms, such as app blocking, which can undermine users' sense of autonomy and trigger psychological reactance—ultimately reducing continued use. Others adopt low-friction approaches, such as passive notifications, which are often ignored at the moment of temptation and fail to interrupt habitual use [37,40-42].

A further limitation lies in the lack of customization: most tools do not allow users to align the intervention with their personal goals, definitions of distraction, or preferred self-regulation strategies. A systematic review of 334 digital self-control tools found that few support the nuanced, user-defined boundaries essential for meaningful and sustainable behavior change [38].

Moreover, evidence on the effectiveness of many digital self-control tools remains limited, particularly with regard to addressing the underlying psychological mechanisms of problematic social media use, such as impulsivity, habitual behavior, and deficient self-control [30,39]. Although a wide variety of tools exists, few interventions integrate empirically grounded behavior change strategies with autonomy-supportive design principles. This presents a critical gap in the literature: the need for digital self-control tools that combine customization, user agency, and theory-based mechanisms of change. Studies indicate that users are more likely to engage with interventions that allow them to set individualized goals,

XSL•FO
RenderX

determine when and how reminders are delivered, and track their progress toward greater self-regulation [43,44].

This study addresses this gap by evaluating the Wellspent app, a customizable, autonomy-supportive intervention designed to promote self-regulated social media use. By combining evidence-based behavior change techniques (BCTs) [45] with flexible, user-driven customization, Wellspent aims to effectively support users in regaining control over their usage patterns while preserving users' sense of autonomy. Through a preregistered randomized controlled trial, this study offers one of the first rigorous evaluations of a digital self-control tool specifically designed to target problematic social media use in a personalized and theory-informed manner.

Aims and Research Questions

The increasing need for effective tools to support healthier and more intentional social media use has led to the development of the Wellspent app, a theory-informed digital self-control tool. The app represents a novel approach to digital self-regulation by combining evidence-based BCTs [45], such as goal setting, behavioral feedback, self-monitoring, and intention-based nudges, with a high degree of user autonomy and customization. Unlike many existing tools that rely on rigid blocking mechanisms or generic notifications [38], the Wellspent app allows users to define their own usage goals, select problematic apps, and personalize reminder timing, tone, and content.

A core differentiator of the Wellspent app lies in its autonomy-supportive design. Rather than enforcing restrictions, the app preserves users' freedom to choose whether to quit or continue using an app after each prompt. This design is grounded in the dual-process model of behavior regulation [22], aiming to interrupt automatic, habitual engagement (System 1) and foster more reflective, goal-directed decision-making (System 2). Prior research suggests that such personalized and autonomy-supportive interventions may enhance message resonance, reduce psychological reactance, and improve long-term adherence [46]. With its emphasis on autonomy, customization, and theory-driven behavior change, this study addresses a critical gap in the digital well-being literature and contributes to the development of more effective, user-centered self-regulation tools.

The significance of this study lies in its methodological rigor. To date, few interventions in the field of digital self-regulation have undergone a rigorous empirical evaluation using gold-standard methods [47,48]. This study presents a preregistered randomized controlled trial with 70 participants, comparing the effects of the Wellspent app with a no-treatment control group. The design ensures high internal validity and allows for causal inferences regarding the app's impact on behavior.

An earlier publication from the overarching study project on intervention effects examining separate outcomes found no significant effects on overall well-being, affect, or stress but observed reduced stress among users with initially high levels of problematic smartphone use in the intervention group [47].

The primary objective of this study is to investigate whether participants in the intervention group receiving the Wellspent app, compared to a control group, demonstrate reductions in problematic social media use, problematic smartphone use, and time spent on the most problematic social media app, as well as an increase in self-efficacy for self-regulating social media app use.

Methods

Study Design and Participants

This study presents findings from a randomized controlled trial designed to evaluate the effectiveness of the Wellspent app, a digital self-control tool aimed at promoting self-regulated social media use on smartphones. The trial was preregistered at the German Clinical Trial Register (DRKS00031767) on May 8, 2023. The first participant was enrolled on May 22, 2023, and a sample size of 108 participants was targeted. The study was conducted between May and June 2023.

Participants were recruited via private social media channels (Facebook, LinkedIn, WhatsApp, and Instagram) and email lists for study opportunities at the Freie Universität Berlin, Germany. Prospective participants enrolled through an online registration page, where they received information regarding the study's purpose, data protection guidelines, and the voluntary nature of participation. Informed consent was obtained from all participants prior to their inclusion in the study.

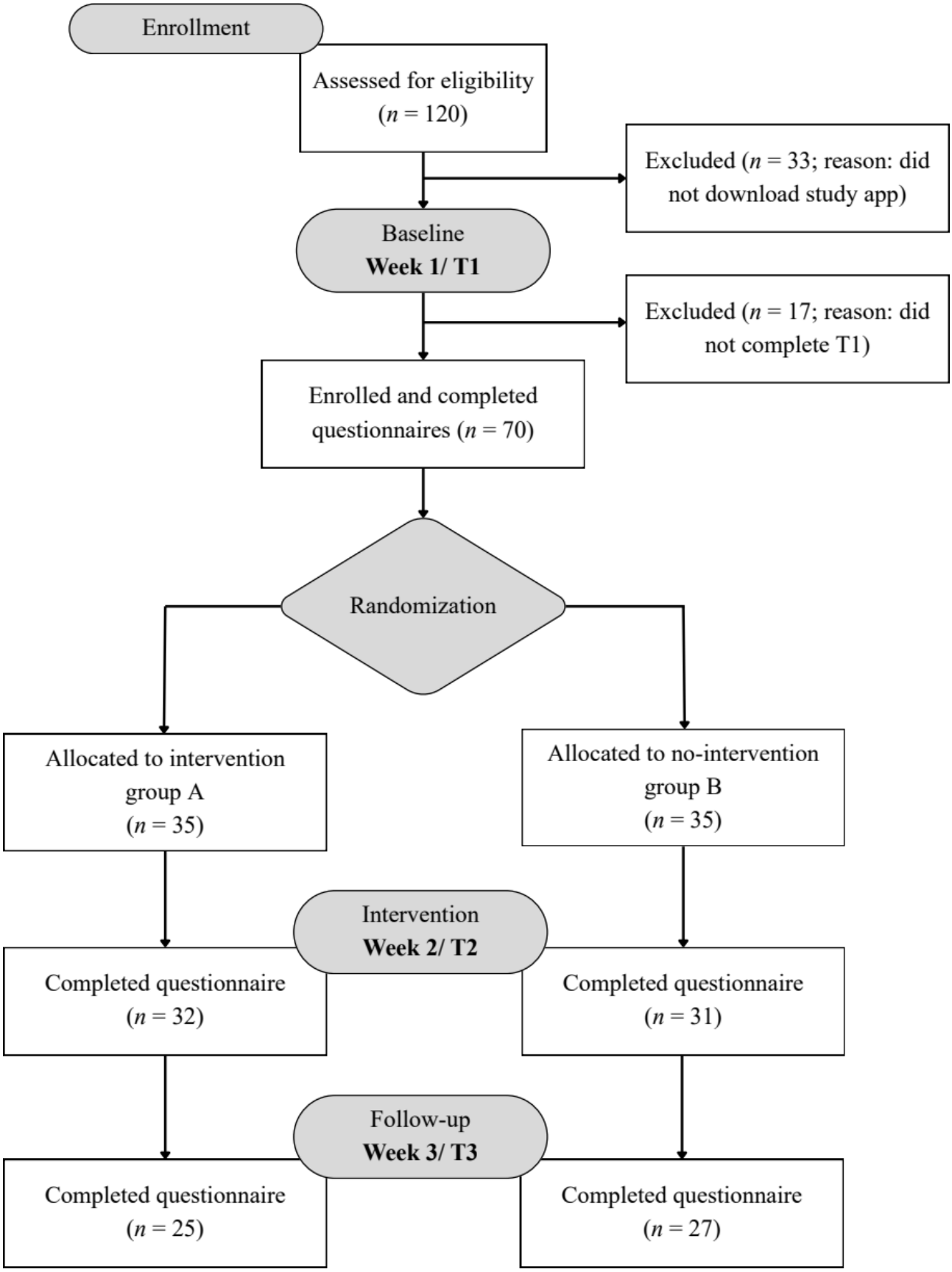
Inclusion criteria required participants to be at least 18 years of age, proficient in English, own an iPhone with iOS 16 (as the Wellspent app was exclusively available for this operating system), and regularly use social media. Social media was broadly defined as platforms enabling users to engage with content, share updates, or interact with others online (eg, Instagram, Facebook, and TikTok).

The final study sample consisted of 70 participants (47/70, 67% female; mean age 26.2, SD 5.6 years), with 35 participants allocated to each group. The remaining 70 participants were randomly assigned (1:1) by the author team to either the intervention group (n=35) or the control group (n=35) using a simple randomization procedure via Research Randomizer [49]. No blocking or stratification was used. Formal blinding was not implemented, as participants were aware of their group assignment.

Participants assigned to the intervention group received the app's intervention features during week 2. The control group received no intervention and continued using their phones as usual. Use of the intervention app in week 3 was voluntary for the intervention group. Both groups completed follow-up assessments at the end of week 2 (T2) and week 3 (T3). At each time point, participants were prompted via email to complete the respective questionnaires, with follow-up reminders sent 24 hours later. Demographic data (including age, sex, level of education, and occupation) were collected at baseline (T1; Figure 1).

XSL•FO
RenderX

Figure 1. Flow diagram outlining participant allocation into the intervention or control group.



Ethical Considerations

This study was approved by the Ethics Committee of the Department of Education and Psychology at Freie Universität Berlin, Germany (proposal number 009.2023; approval date May 5, 2023). Prior to enrollment, all participants provided informed consent, having received detailed information about the study's purpose, data processing procedures, and their right

XSL•FO
RenderX

to withdraw at any time. To ensure privacy and confidentiality, all data were pseudonymized at the point of data collection, and no personally identifiable information was retained. All collected data were stored on secure servers in accordance with the General Data Protection Regulations regulations. Participants were compensated with academic course credits (when they were students) and received a 1-year premium account for the Wellspent app, provided by the app's developer; no other financial incentives were offered.

Intervention

The Wellspent app is a theory-informed, customizable mobile intervention developed to support self-regulated social media use. It targets user-identified problematic apps and delivers personalized behavioral nudges. The intervention draws on BCTs from the BCT Taxonomy v1 [45], addressing psychological mechanisms associated with problematic social media use [50]. Specifically, the intervention incorporates the following BCTs [51]:

1. Goal setting (behavior; Behavior Change Intervention Ontology [BCIO: 007004]): users define daily time limits for selected social media apps.
2. Prompts/cues (BCIO: 050578): full-screen reminders are displayed when usage exceeds a user-defined threshold.
3. Feedback on behavior (BCIO: 007022): real-time messages show time spent on selected apps.
4. Self-monitoring of behavior (BCIO: 007024): visual summaries allow users to track progress toward their individual goals.
5. Substitution of behavior (BCIO: 007095): users specify alternative behaviors to engage in when prompted to take a break from social media.

The intervention was designed according to self-nudging principles [29], enabling users to customize the app's behavior to align with their habits, goals, and preferences. This approach preserves autonomy and supports reflective (System 2) rather than habitual (System 1) decision-making [22]. Customization features and exemplary reminder nudges are detailed in Multimedia Appendix 1.

All users in the intervention group completed the same standardized setup process:

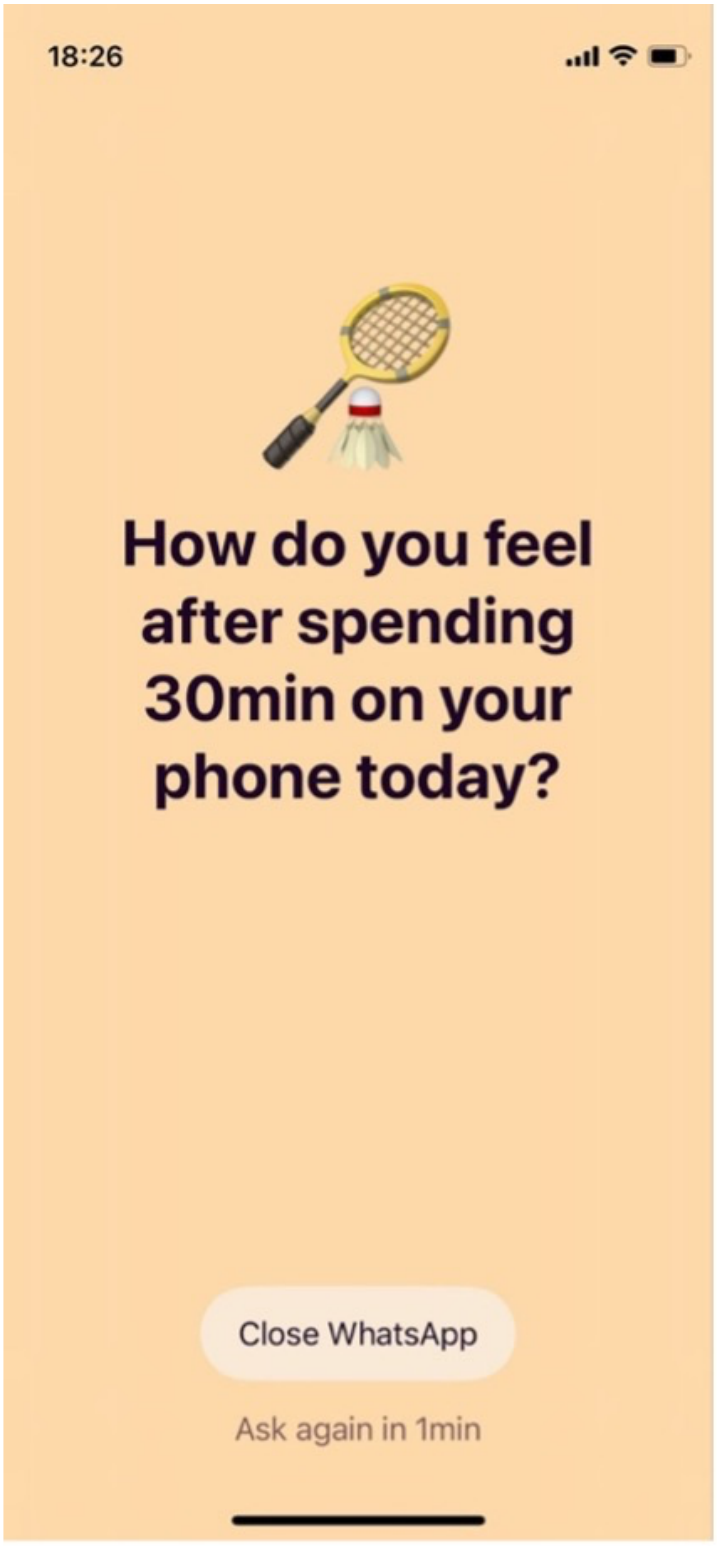
1. Installation and onboarding: participants installed the Wellspent app on their iPhones (iOS 16+).
2. Selection of problematic apps: participants identified one or more social media apps they personally considered problematic (eg, Instagram and TikTok).
3. Goal setting: for each selected app, participants defined a daily usage limit (eg, 45 min) and a nudge interval (eg, after 10 min of continuous use).
4. Customization of reminder prompts: participants tailored full-screen reminders to their preferences:

- Reminder frequency: how often reminders should appear once the nudge interval is exceeded (eg, every 3 min).
- Alternative activity: users selected an activity to engage in instead of scrolling (eg, reading and exercising).
- Tone of voice: participants chose the communicative style of the reminder (eg, motivational or reflective). See Figure 2 for an example.
- "Danger Zones" (optional): users could define vulnerable time windows (eg, before bed), during which reminders became more salient.

- Triggering of full-screen reminders: once the self-defined nudge interval was exceeded during app use, a full-screen reminder appeared. These reminders required active dismissal but did not block app access, allowing users to decide whether to continue or take a break.
- Timing within the study: the intervention began in week 2. In week 3, continued app use remained voluntary for participants in the intervention group.

XSL•FO
RenderX

Figure 2. A screenshot of an exemplary reminder message from the Wellspent app.



Measures

Weekly self-report measures were collected between baseline (T1) and follow-up (T3) to assess the following outcomes: problematic social media app use, problematic smartphone use, self-efficacy to self-regulate social media app use, and participants' screen time for their most problematic social media app.

Problematic Social Media App Use

The primary outcome, problematic social media app use, was measured using an adaptation of the Bergen Social Networking Addiction Scale (BSNAS) [52]. The BSNAS comprises 6 items, such as "How often during the last 7 days have you used social media so much that it has had a negative impact on your job/studies?" rated on a scale from 1 ("very rarely") to 5 ("very often"). The internal consistency of the baseline problematic

XSL•FO
RenderX

social media app use scale demonstrated acceptable reliability (α=.77).

Problematic Smartphone Use

Self-reported problematic smartphone use was included as an additional measure beyond preregistered outcomes to examine effects on overall smartphone use and its known correlation with various health outcomes [53]. Problematic smartphone use was measured with a single item: "How often during the 7 last days have you found your smartphone use problematic?" rated on a scale from 1 ("very rarely") to 5 ("very often"). Prior research indicates a high correlation between this 1-item measure and the established scale from Foerster et al [54] in an earlier study [48].

Self-Efficacy to Self-Regulate Social Media App Use

Three items adapted from Keller et al [48] measured self-efficacy to self-regulate social media app use. Participants responded to statements commencing with "During the last 7 days have you felt that you can..." (eg, "...use social media apps consciously even if I first have to find a way to integrate this into my daily routine."), rated on a scale from 1 ("not at all true") to 6 ("exactly true"). The internal consistency at baseline (T1) indicated acceptable reliability (α=.63).

Social Media App Use

Participants' daily minutes spent on their most problematic social media app were derived from objectively measured screen time app records. This data was input by participants in hours and minutes for their most problematic social media app when completing the questionnaires for the respective study weeks (week 1, week 2, and week 3). Weekly scores were calculated by averaging the daily minutes of social media app use over the past 7 days. Univariate outliers (z>3.29) were winsorized to 1 unit higher than the highest valid value, with z<3.29 in the distribution [55].

User Acceptance

User acceptance of the intervention was defined as the user refraining from dismissing a reminder (and thereby continuing their app use) within 1 minute of receiving the reminder nudge. The acceptance rate of the Wellspent app's reminders was computed by dividing the instances of accepted interventions by the total number of occurrences.

Qualitative Information

Qualitative data were collected through an online survey conducted via Qualtrics at the end of the intervention period. Participants in the intervention group answered a series of 10 open-ended questions aimed at exploring their experiences with the Wellspent app (Table S1 in Multimedia Appendix 2). These questions included prompts such as "How would you describe the Wellspent app to a friend?" and "What does the Wellspent app help you with?" The questions were designed to gather insights into users' perceptions of the app's functionality, usability, and effectiveness, as well as to identify potential areas for improvement. A thematic analysis approach was used informally to identify key themes from participants' responses [56].

Covariates

The list of covariates included participants' age, sex (women=1; male=0), education (1=university degree; 0=no university degree), positive affect, and screen time in minutes at baseline. Positive affect was measured with an adaptation of the Positive and Negative Affect Schedule with items such as "Please indicate the extent you have felt this way when thinking about your social media app use over the last 7 days" (1="very slightly or not at all"; 6="extremely") [47].

Data Preparation and Preliminary Analyses

Prior to hypothesis testing, data were screened and cleaned. Univariate outliers in screen time (z>3.29) were winsorized to 1 unit above the highest valid score with z<3.29, following recommendations by Tabachnick and Fidell [55]. To assess baseline comparability between the intervention and control groups, t tests and chi-square tests were conducted for key demographic and outcome variables. These tests also informed attrition analyses comparing retained and dropout participants.

Statistical Analysis

Power

A power analysis was conducted using G*Power (version 3.1.9.7; Heinrich-Heine-Universität Düsseldorf) [57] to determine the necessary sample size for a repeated measures ANOVA, for 2 groups across 2 measurement points in time (ie, from baseline to postintervention and baseline to follow-up) [58]. It indicated that a minimum of 54 participants (27 per group) were required to identify a significant group × time interaction effect. This calculation assumed a medium effect size of 0.25 (Cohen medium effect size) [35,48,58], with a significance level (α) of .05, and a power value of 1–β=0.95 [59]. Anticipating a 50% dropout rate, as observed in a similar intervention study by Keller and colleagues [48], our target recruitment was set at a minimum of 108 participants.

Main Analysis

Data analyses were conducted according to the intention-to-treat approach using both SPSS (version 29; IBM Corp) and RStudio (version 4.3.2; R Core Team) [60]. While SPSS was primarily used for data cleaning purposes, RStudio with its *lme4* package [61] was used for multilevel modeling. We acknowledge that this data analysis deviated from the ANOVA-based power analysis; however, multilevel modeling accounts for the nested data structure and the within-group and between-group variability. Group differences in problematic social media use, problematic smartphone use, self-efficacy, and screen time were assessed using linear mixed models. These models were structured with 3 time points (T1, T2, and T3; within level) nested within participants (between level), and to apply a restricted maximum likelihood procedure in order to account for missing data [61]. To determine intervention effects on the outcome measure, time (0=T1, 1=T2, and 2=T3) × experimental condition (0=control group and 1=intervention group) interactions were estimated as predictors. In subsequent sensitivity analyses, age, sex, positive affect, education level, and screen time in minutes were included as covariates in models 1 - 3. Model 4 included age, sex, positive affect, and education level as covariates. While we are unable to share

XSL•FO
RenderX

participant data due to privacy regulations, the R code for the data analyses is available in the project's Open Science Framework repository [62].

Results

Sample Characteristics, Randomization, and Attrition Check

Of the 120 persons screened for eligibility, 33 of them did not install the Wellspent app, and 17 persons did not complete the baseline questionnaire. The randomized sample consisted of 70 individuals (47 women and 23 men) with a mean age of 26.19 (SD 5.63; range 18 - 45) years. Figure 1 shows a CONSORT (Consolidated Standards of Reporting Trials) flow diagram outlining the study's design and participant allocation across the randomized controlled trial with 3 measurement points. Most participants were students or young professionals. Participants' most chosen problematic social media app was Instagram (29/35, 82%), followed by TikTok (4/35, 10%) and Twitter (1/35, 4%). Detailed demographic data are presented in Table 1, and descriptive statistics for all outcome measures across time and condition are provided in Table 2.

Table . Demographic characteristics of the baseline sample and per study condition. Table reproduced from [47].

Variables	Total sample (n=70)	Intervention group (n=35)	Control group (n=35)
Sex, n (%)			
Female	47 (67)	22 (63)	25 (71)
Male	23 (33)	13 (37)	10 (29)
Age (years), mean (SD)	26.19 (5.63)	26.77 (6.56)	25.6 (4.53)
Nationality, n (%)			
German	56 (80)	28 (80)	28 (80)
American	3 (4)	2 (6)	1 (3)
Other nationalities	11 (16)	5 (14)	6 (17)
Education, n (%)			
Doctorate degree	3 (4)	1 (3)	2 (6)
Masters degree	13 (19)	8 (23)	5 (14)
Bachelor degree	23 (33)	10 (29)	13 (37)
Professional degree	4 (6)	2 (6)	2 (6)
Diploma	1 (1)	1 (3)	0 (0)
Trade	1 (1)	0 (0)	1 (3)
High school graduate	25 (36)	13 (37)	12 (34)
Employment, n (%)			
Student	39 (56)	21 (60)	18 (51)
Full-time	20 (29)	10 (29)	10 (29)
Part-time	8 (11)	2 (6)	6 (17)
Unemployed	1 (1)	1 (3)	0 (0)
Other	2 (3)	1 (3)	1 (3)

XSL•FO
RenderX

Table . Descriptive statistics of study outcomes for both experimental conditions. Reduced sample size for social media app use is due to missing data.

Variable and time point	n	Intervention group, mean (SD)	n	Control group, mean (SD)
Problematic social media app use (rating scale: 1-5)				
T1[a]	35	2.58 (0.76)	35	2.79 (0.71)
T2[b]	32	2.47 (0.61)	31	2.69 (0.66)
T3[c]	25	2.32 (0.72)	27	2.77 (0.68)
Problematic smartphone use (rating scale: 1-5)				
T1	35	3.00 (0.94)	35	3.03 (0.99)
T2	32	2.50 (0.95)	31	3.16 (0.90)
T3	25	2.68 (0.85)	27	3.37 (0.84)
Self-efficacy to self-regulate smartphone app use (rating scale: 1-6)				
T1	35	3.30 (0.70)	35	3.24 (0.70)
T2	32	4.03 (0.73)	31	3.80 (0.76)
T3	25	3.95 (0.90)	27	4.12 (0.72)
Social media app use (min/d)				
T1	21	77 (62)	24	59 (51)
T2	21	54 (46)	24	59 (44)
T3	21	46 (34)	24	63 (48)

[a]T1: baseline.

[b]T2: postintervention.

[c]T3: follow-up.

Randomization confirmed no significant differences across baseline variables between groups. In the attrition analysis, we found that participants who dropped out from the study before T3 (n=18) showed higher levels of positive affect than participants retained in the longitudinal sample (n=52; nonretainers: mean 2.59, SD 0.53 vs retainers: mean 2.20, SD 0.67; *P*=.03). Among the total sample, a subsample of 52 participants (representing a dropout rate of 26%) completed the study by week 3, comprising 25 participants from the intervention group and 27 participants from the control group. These numbers are almost in accordance with the power analysis, which indicated a targeted longitudinal sample size of 27 persons per group. The attrition rate of the present study aligns with observations from other online-based studies [63]. Table 1 provides an overview of key demographic characteristics of the baseline sample.

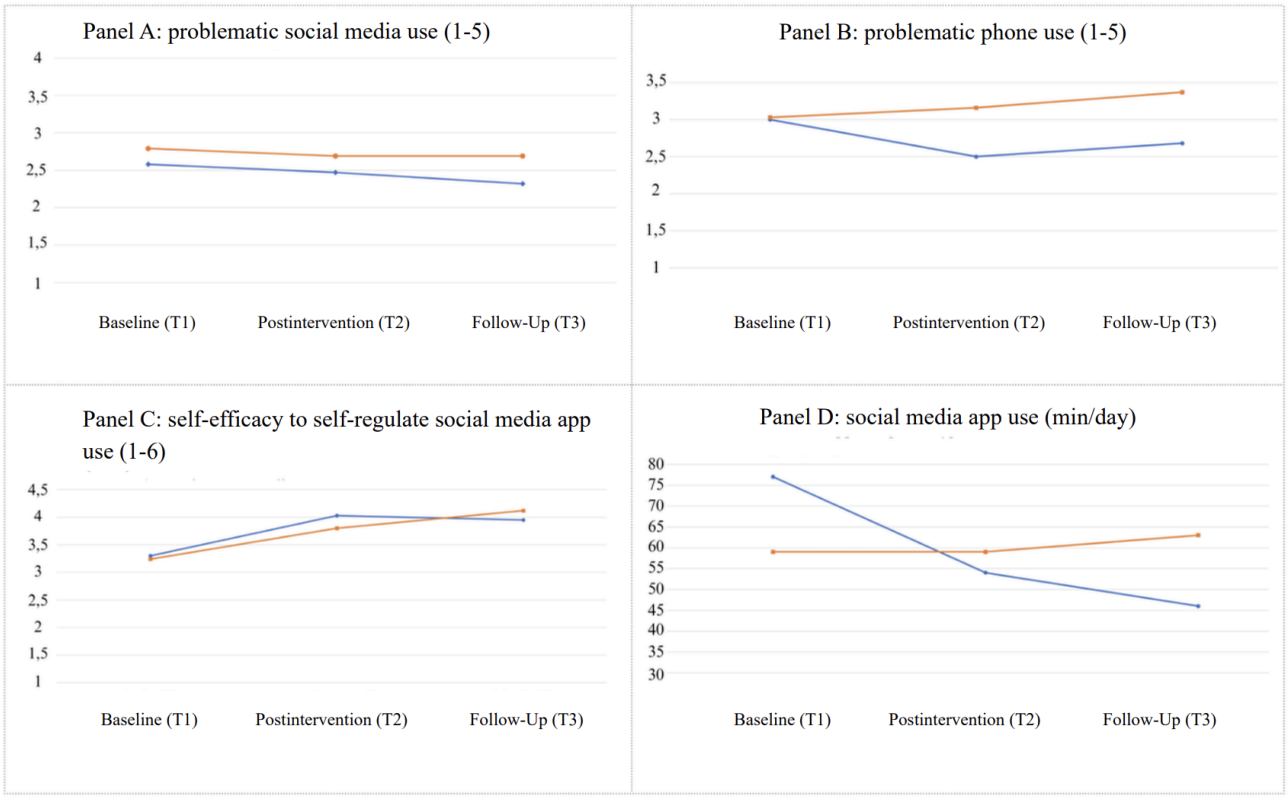
Changes in Study Outcomes Over Time

Overview

Mean levels of problematic social media app use, problematic smartphone use, social media app use, and self-efficacy to self-regulate social media app use over time and across both conditions were examined and presented in Table 2 and visualized in Figure 3. Within the intervention group, 33% (6430/19,522) of reminder events in week 2 led to participants refraining from further app use within 1 minute after receiving the reminder. This rate increased slightly to 34% (5394/15,996) in week 3. Notably, within the intervention group, 25 (78%) participants voluntarily continued their use of the Wellspent app in week 3.

XSL•FO
RenderX

Figure 3. Mean levels over time of outcome variable in the intervention (blue) and control (orange) group. Only parts of the response scales are displayed in panels A, B, and C.



Primary Outcome: Problematic Social Media Use

For the primary outcome, problematic social media app use, the multilevel model revealed a nonsignificant time × group interaction effect (model 1: estimate [Est]=−0.27, SE 0.13, 95% CI −0.51 to –0.01; *P*=.05; intraclass correlation [ICC]=0.69).

Secondary Outcome: Problematic Smartphone Use

Problematic smartphone use showed a significant time × group interaction effect (model 2: Est=−0.46, SE 0.18, 95% CI –0.80 to –0.11; *P*=.01; ICC=0.52), indicating improvements in the intervention group when compared to the control group (Figure 3, Panel B).

Secondary Outcome: Self-Efficacy to Self-Regulate Social Media Use

Self-efficacy to self-regulate social media use showed a similar increase over time in both the intervention and control group (model 3; time effect: Est=0.57, SE 0.11, 95% CI 0.34-0.78; *P*<.001; ICC=0.34), but no significant time × group effect emerged (model 3: Est=−0.10, SE 0.17, 95% CI −0.44 to 0.22; *P*=.55).

Secondary Outcome: Social Media App Use

Regarding time of social media app use, a significant time × group effect was observed (model 4: Est=−29.35, SE 6.84, 95% CI −42.79 to −15.99; *P*<.001). This indicates that social media app use significantly decreased in the intervention group (vs the control group), with an average reduction of approximately 29 minutes per day from week 1 to week 3 (Figure 3, Panel D). The overall pattern of results remained consistent when covariates age, sex, positive affect, and education level were added in sensitivity analyses (reported in Table 3). For problematic social media use, the pattern differed, with the effect reaching statistical significance only in the adjusted model. This suggests that the finding may not represent a robust effect.

XSL•FO
RenderX

Table . Multilevel model estimates of outcome variables up to a 3-week follow-up. SEs are reported for fixed effects. Random effects are summarized as SDs with 95% CIs.

	Model 1: problematic social media app use			Model 2: problematic smartphone use			Model 3: self-efficacy to self-regulate social media app use			Model 4: time of social media app use		
Fixed effects	Est[a] (SE)	*P* value	95% CI	Est (SE)	*P* value	95% CI	Est (SE)	*P* value	95% CI	Est (SE)	*P* value	95% CI
Intercept at baseline	2.43 (0.21)	<.001	2.04 to 2.82	2.47 (0.30)	<.001	1.91 to 3.03	3.47 (0.25)	<.001	2.98 to 3.97	78.27 (14.51)	<.001	49.37 to 106.26
Time	0.02 (0.09)	.80	–0.15 to 0.19	0.13 (0.12)	.27	–0.10 to 0.37	0.57 (0.11)	<.001	0.34 to 0.78	3.70 (4.53)	.40	–5.14 to 12.60
Group	–0.08 (0.18)	.67	–0.41 to 0.26	–0.10 (0.26)	.71	–0.58 to 0.39	0.12 (0.20)	.57	–0.30 to 0.50	22.79 (13.86)	.10	–3.72 to 49.33
Group by time	–0.27 (0.13)	.05	–0.51 to –0.01	–0.46 (0.18)	.01	–0.80 to –0.11	–0.10 (0.17)	.55	–0.44 to 0.22	–29.35 (6.84)	<.001	–42.79 to –15.99
Age	–0.09 (0.02)	<.001	–0.13 to 0.04	–0.10 (0.03)	.01	–0.17 to –0.04	0.05 (0.03)	.09	–0.01 to 0.11	–2.30 (1.60)	.16	–5.30 to 0.69
Sex[b]	0.21 (0.16)	.20	0.09 to 0.51	0.35 (0.26)	.07	–0.001 to 0.86	0.15 (0.20)	.47	–0.28 to 0.55	1.50 (10.73)	.89	–19.47 to 22.40
Education[c]	0.03 (0.18)	.86	–0.29 to 0.36	0.35 (0.26)	.18	–0.12 to 0.82	–0.34 (0.22)	.13	–0.78 to 0.08	–37.61 (11.79)	<.001	–60.63 to –14.65
Phone usage	–0.00 (0.00)	.62	–.001 to .001	–0.00 (0.00)	.69	–0.001 to 0.001	–0.00 (0.00)	.09	–0.001 to 0.00	—[d]	—	—
Positive affect	0.10 (0.12)	.40	–0.12 to 0.32	–0.15 (0.17)	.37	–0.46 to 0.16	–0.28 (0.14)	.06	–0.56 to - 0.03	–3.69 (7.77)	.64	–19.15 to 11.82
Random effects (SD)												
Intercept	0.50	—	0.31 to 0.62	0.64	—	0.32 to 0.83	0.40	—	0.18 to 0.64	42.77	—	31.13 to 52.03
Time	0.29	—	0.00 to 0.42	0.19	—	0.00 to 0.49	0.14	—	0.01 to 0.46	16.42	—	6.18 to 23.21
Residual variance	0.34	—	0.28 to 0.42	0.61	—	0.50 to 0.72	0.59	—	0.46 to 0.67	16.61	—	13.61 to 20.89
ICC[e]	0.69	—	—	0.52	—	—	0.34	—	—	0.87	—	—
Pseudo-R^2	0.78	—	—	0.62	—	—	0.57	—	—	0.88	—	—

[a]Est: estimate.

[b]Sex coded as 0=men and 1=women.

[c]Education coded as 0=no tertiary education and 1=tertiary education.

[d]Not applicable.

[e]ICC: intraclass correlation.

Qualitative Results

Qualitative feedback was explored informally by reviewing the open-ended responses from the intervention group, collected via Qualtrics at the end of the intervention period. Responses were grouped into recurring patterns based on the researcher's judgment, without following a formal coding or analysis protocol. Emerging themes included the following.

App Effectiveness and Positive Feedback

Participants provided positive feedback on the app's effectiveness in managing screen time, frequently praising its simplicity and user-friendly interface. One participant remarked, "It's a very easy-to-use and straightforward app that allows you to monitor your social media use through reminders sent during self-chosen time intervals." (P26). The app's reminders were particularly appreciated for interrupting mindless scrolling, serving as timely prompts about time spent on social media, and encouraging users to put their phones down. For instance,

XSL•FO
RenderX

a participant noted, “Great app that keeps you alarmed that there’s life outside of the social media apps.” (P32). The ability to set individual budgets for different apps was also considered valuable for promoting conscious social media use: “The budgets are very useful because different apps need different kinds of monitoring.” (P26).

Concerns and Areas for Improvement

The main concerns raised by the intervention group primarily revolved around technical bugs, such as repeated reminders after dismissal, inaccuracies in screen time tracking, and restrictions on selected apps. These issues were seen as areas for improvement, as they impacted the overall user experience.

Integration Into Daily Life and User Acceptance

Participants expressed varying views on integrating the Wellspent app into their daily social media use. Some highlighted the app’s value in enhancing awareness and reducing usage time, describing it as “helpful but not intrusive” (P21). Others underscored its assistance in enabling alternative activities: “It helps me to do other things instead.” (P10). However, instances emerged where user acceptance was hindered, particularly when the app’s reminders interrupted enjoyable social media experiences: “It can be annoying when you are enjoying being on social media, and have done so in moderation, to get constant notifications. I ended up moving the nudge interval when I wanted to use social media for longer periods of time.” (P3).

Overall, while participants generally acknowledged the app’s efficacy in monitoring screen time, some expressed a need for improvements in certain features to fully embrace its integration into daily life. Suggested enhancements included additional statistics tracking (eg, weekly screen time reduction), enhanced friction when exceeding limits (eg, app lock with a code), different modes (eg, study, work, or intentional scrolling mode), and adaptive nudge intervals (eg, prolonged nudge intervals after extended periods of nonuse).

Discussion

Principal Results

This randomized controlled trial investigated the effectiveness of the customizable intervention app Wellspent in supporting self-regulated social media use. The intervention produced significant reductions in daily time spent on the most problematic app (approximately 29 minutes) and reduced perceived problematic smartphone use. In contrast, no significant effects were found for problematic social media use or self-efficacy to self-regulate social media use. These findings suggest that behavioral change, in the form of reduced usage time, can occur relatively quickly through moderate, autonomy-supportive interventions, whereas cognitive and self-perception outcomes may require longer, more reflective processes.

Effects on Problematic Social Media Use

The intervention did not significantly decrease problematic social media use. This aligns with research showing that problematic use arises from deeper, multifaceted patterns, such as emotional health [16], habit strength [64], and personality characteristics [65], making it difficult to change quickly. The intervention app may not have supplied the sustained cognitive scaffolding needed to alter participants’ perceived problematic use within 2 weeks.

Effects on Problematic Smartphone Use

The intervention group reported lower levels of problematic smartphone use compared to the control group. This suggests that behavioral reductions may have translated into a more general feeling of control over one’s device. Notably, perceived problematic smartphone use may be easier to shift in the short term than problematic social media use because it reflects global awareness (“I use my phone too much”) rather than app-specific patterns linked to social, emotional, or identity-related motives.

Effects on Self-Efficacy to Self-Regulate Social Media App Use

Self-efficacy significantly increased in both groups, but participants from the intervention group did not exceed those from the control group. Self-efficacy typically grows through repeated mastery experiences and feedback on successful behavior regulation [66]. As the intervention app focused on in-the-moment friction rather than visible progress tracking, participants may have lacked the experience of sustained success and explicit feedback loops. Similar interventions have also failed to find strong between-group intervention effects on self-efficacy [19], pointing to the need for longer interventions, richer in feedback and reflection.

Effects on Social Media App Use

The intervention produced a reduction in time spent using social media apps of approximately 29 minutes per day. This finding aligns with prior research highlighting the potential of digital self-control tools to reduce usage time when users are nudged at critical moments [35,40]. Compared to digital self-control tools with generic interventions or one-size-fits-all designs [38], Wellspent offered a tailored and user-centric alternative. The full-screen reminder appearing after a personally defined nudge interval (eg, 10 min on Instagram) created a structured moment of pause, allowing users to shift from automatic to deliberate decision-making [22]. Previous literature emphasizes that user autonomy and personalization increase intervention acceptance and long-term engagement [46]. Our results support this perspective, particularly the positive user feedback from intervention participants regarding flexibility and customization. By integrating customizable, goal-aligned nudges into real-time usage patterns, the app sets a precedent for user-centered intervention strategies in digital self-control.

Limitations and Future Research

Together, the findings support emerging evidence that autonomy-preserving tools can meaningfully influence digital behavior within short periods. These findings highlight three key considerations for practitioners and designers of digital self-control tools:

1. Moderate enforcement can be effective: interventions do not need to be coercive to reduce screen time. Full-screen reminders that offer users a choice, rather than restricting

XSL•FO
RenderX

access, can facilitate reflective engagement without triggering reactance.
2. Customization enhances acceptance: giving users the ability to select their target apps, usage limits, and reminder style appears to foster a sense of ownership and autonomy, which in turn increases voluntary engagement and retention. Future digital self-control tools should prioritize configurability to accommodate individual usage patterns and values.
3. Behavioral change may require more than just friction: while nudging can reduce immediate app use, deeper cognitive outcomes such as self-efficacy or problematic social media use may require additional components, for example, reflective journaling, feedback on progress, or social accountability mechanisms.

This study has several limitations. First, the sample size was relatively small, and self-report measures may be subject to bias, particularly regarding subjective app usage and problematic use. The target sample size of 54 participants, as determined by the power analysis, was not reached. The final analysis sample for the multilevel models on the primary outcome consisted of 46 participants, which may have limited the statistical power to detect medium effects. Second, the short intervention duration may have limited the app's impact on self-efficacy and habitual behavior. Third, although self-efficacy to self-regulate social media use was assessed, a direct measure of self-regulation (eg, action control) was not included in the study. This omission may limit the ability to fully capture the psychological mechanisms underlying digital behavior change. Future studies should incorporate validated measures of self-regulation to provide a more comprehensive understanding of intervention effects. Fourth, the study was limited to iOS users, excluding other relevant target groups.

Future research should explore the long-term effects of customizable digital self-control tools and investigate how additional features, such as social comparison, adaptive feedback, or intention tracking, could enhance both behavioral and psychological outcomes. Mixed method designs integrating behavioral data, qualitative feedback, and longitudinal tracking could provide deeper insights into how users engage with interventions over time.

Conclusions

This study provides initial evidence for the effectiveness of a customizable, moderately enforced digital self-control tool in reducing social media screen time and perceived problematic smartphone use. By allowing users to define when, where, and how interventions occur, the Wellspent app demonstrates a user-centered approach to digital self-regulation. These findings support the growing consensus that flexible and autonomy-supportive digital self-control tools offer a promising path toward sustainable behavior change in digital environments.

Acknowledgments

The authors are thankful for the contributions of Marius Rackwitz (development of the study app).

All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

Funding

The authors declared no financial support was received for this work.

Authors' Contributions

LM developed the study concept and conducted the data collection under JK's supervision. LCB conducted the data analysis. LM drafted the initial manuscript, and LCB, CR, TD, TR, and JK collaboratively provided feedback. The final manuscript was approved by all authors for submission.

Conflicts of Interest

LM was employed as a working student at Wellspent GmbH during the time of the intervention period. CR is a cofounder of Wellspent GmbH. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Screenshots from the intervention app.
[PDF File, 2477 KB - mhealth_v14i1e56824_app1.pdf]

Multimedia Appendix 2
Qualitative questions at postintervention and follow-up.
[DOCX File, 15 KB - mhealth_v14i1e56824_app2.docx]

Checklist 1

XSL•FO
RenderX

CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 1277 KB - mhealth_v14i1e56824_app3.pdf]

References

1. DataReportal. Digital 2022: global overview report. : DataReportal – Global Digital Insights URL: https://datareportal.com/reports/digital-2022-global-overview-report [accessed 2023-03-26]
2. Weinstein E. The social media see-saw: positive and negative influences on adolescents' affective well-being. New Media Soc 2018;20(10):3597-3623. [doi: 10.1177/1461444818755634] [Medline: 30298429]
3. Wartberg L, Thomasius R, Paschke K. The relevance of emotion regulation, procrastination, and perceived stress for problematic social media use in a representative sample of children and adolescents. Comput Human Behav 2021;121:106788. [doi: 10.1016/j.chb.2021.106788] [Medline: 34019838]
4. Beyens I, Pouwels JL, van Driel II, et al. The effect of social media on wellbeing differs from adolescent to adolescent. Sci Rep 2020;10(1):10763. [doi: 10.1038/s41598020677277] [Medline: 32612108]
5. Kross E, Verduyn P, Sheppes G, Costello CK, Jonides J, Ybarra O. Social media and well-being: pitfalls, progress, and next steps. Trends Cogn Sci 2021;25(1):55-66. [doi: 10.1016/j.tics.2020.10.005] [Medline: 33187873]
6. Lukoff K, Munson S. Digital wellbeing is way more than just reducing screen time: UX Collective; 2019. URL: https://uxdesign.cc/digital-wellbeing-more-than-just-reducing-screen-time-46223db9f057?gi=088e16387e93 [accessed 2026-03-18]
7. Valkenburg PM. Social media use and well-being: what we know and what we need to know. Curr Opin Psychol 2022;45:101294. [doi: 10.1016/j.copsyc.2021.12.006] [Medline: 35016087]
8. Verduyn P, Gugushvili N, Kross E. The impact of social network sites on mental health: distinguishing active from passive use. World Psychiatry 2021;20(1):133-134. [doi: 10.1002/wps.20820] [Medline: 33432746]
9. Zhang J, Marino C, Canale N, et al. The effect of problematic social media use on happiness among adolescents: the mediating role of lifestyle habits. IJERPH 2022;19(5):2576. [doi: 10.3390/ijerph19052576] [Medline: 35204799]
10. Odgers CL, Schueller SM, Ito M. Screen time, social media use, and adolescent development. Annu Rev Dev Psychol 2020 Dec 15;2(1):485-502. [doi: 10.1146/annurev-devpsych-121318-084815]
11. Huang C. A meta-analysis of the problematic social media use and mental health. Int J Soc Psychiatry 2022;68(1):12-33. [doi: 10.1177/00207640209784344] [Medline: 33295241]
12. Shannon H, Bush K, Villeneuve PJ, Hellemans KG, Guimond S. Problematic social media use in adolescents and young adults: systematic review and meta-analysis. JMIR Ment Health 2022;9(4):e33450. [doi: 10.2196/33450] [Medline: 35436240]
13. Rozgonjuk D, Sindermann C, Elhai JD, Montag C. Fear of missing out (FoMO) and social media's impact on daily-life and productivity at work: do WhatsApp, Facebook, Instagram, and Snapchat use disorders mediate that association? Addict Behav 2020;110:106487. [doi: 10.1016/j.addbeh.2020.106487] [Medline: 32674020]
14. Cheng C, Lau YC, Chan L, Luk JW. Prevalence of social media addiction across 32 nations: Meta-analysis with subgroup analysis of classification schemes and cultural values. Addict Behav 2021 Jun;117:106845. [doi: 10.1016/j.addbeh.2021.106845] [Medline: 33550200]
15. Vanden Abeele MMP. Digital wellbeing as a dynamic construct. Commun Theory 2021;31(4):932-955. [doi: 10.1093/ct/qtaa024]
16. Busch PA, McCarthy S. Antecedents and consequences of problematic smartphone use: a systematic literature review of an emerging research area. Comput Human Behav 2021;114:106414. [doi: 10.1016/j.chb.2020.106414]
17. Billieux J, Maurage P, Lopez-Fernandez O, Kuss DJ, Griffiths MD. Can disordered mobile phone use be considered a behavioral addiction? An update on current evidence and a comprehensive model for future research. Curr Addict Rep 2015;2(2):156-162. [doi: 10.1007/s40429-015-0054-y]
18. Griffiths M. A 'components' model of addiction within a biopsychosocial framework. J Subst Use 2005 Jan;10(4):191-197. [doi: 10.1080/14659890500114359]
19. Schneider FM, Lutz S, Halfmann A, Meier A, Reinecke L. How and when do mobile media demands impact well-being? Explicating the Integrative Model of Mobile Media Use and Need Experiences (IM 3 UNE). Mob Media Commun 2022 May;10(2):251-271. [doi: 10.1177/20501579211054928]
20. Brevers D, Turel O. Strategies for self-controlling social media use: Classification and role in preventing social media addiction symptoms. J Behav Addict 2019 Sep 1;8(3):554-563. [doi: 10.1556/2006.8.2019.49] [Medline: 31545100]
21. Zahrai K, Veer E, Ballantine PW, Peter de Vries H. Conceptualizing self-control on problematic social media use. Australas Mark J 2022 Feb;30(1):74-89. [doi: 10.1177/1839334921998866]
22. Kahneman D. Thinking, Fast and Slow: Macmillan; 2011.
23. Kuss DJ, Griffiths MD. Social networking sites and addiction: ten lessons learned. Int J Environ Res Public Health 2017;14(3):311. [doi: 10.3390/ijerph14030311] [Medline: 28304359]
24. Radovic A, Gmelin T, Stein BD, Miller E. Depressed adolescents' positive and negative use of social media. J Adolesc 2017;55:5-15. [doi: 10.1016/j.adolescence.2016.12.002] [Medline: 27997851]
25. Monge Roffarello A, De Russis L. Towards understanding the dark patterns that steal our attention. Presented at: ACM CHI Conference on Human Factors in Computing Systems; Apr 30 to May 5, 2022. [doi: 10.1145/3491101.3519829]

XSL•FO
RenderX

26. Mildner T, Savino GL. How social are social media? The dark patterns in facebook's interface. arXiv. Preprint posted online on Mar 19, 2021. [doi: 10.48550/arXiv.2103.10725]
27. Pinder C, Vermeulen J, Cowan BR, Beale R. Digital behaviour change interventions to break and form habits. ACM Trans Comput-Hum Interact 2018;25(3):1-66. [doi: 10.1145/3196830]
28. Thaler RH, Sunstein CR. Nudge: The Final Edition: Yale University Press; 2021.
29. Reijula S, Hertwig R. Self-nudging and the citizen choice architect. Behav Public Policy 2022;6(1):119-149. [doi: 10.1017/bpp.2020.5]
30. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
31. Hiniker A, Hong S, Kohno T, Kientz JA. MyTime: designing and evaluating an intervention for smartphone nonuse. Presented at: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems; May 7-12, 2016 p. 4746-4757. [doi: 10.1145/2858036.2858403]
32. Kim I, Jung G, Jung H, Ko M, Lee U. Let's FOCUS: mitigating mobile phone use in college classrooms. Proceedings of the ACM Interact Mob Wearable Ubiquitous Technol 2017;1(3):1-29. [doi: 10.1145/3130928]
33. Ko M, Yang S, Lee J, et al. NUGU: a groupbased intervention app for improving selfregulation of limiting smartphone use. Presented at: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing; Feb 28 to Mar 4, 2015 p. 1235-1245. [doi: 10.1145/2675133.2675244]
34. Feng CS, Hu TW, Chen YR, Tsai CY. Research on improving the user experience and usability evaluation of tomato work method app—using forest app. Proceedings of the 2019 7th International Conference on Information Technology: IoT and Smart City 2019 Dec 20:147-151. [doi: 10.1145/3377170.3377172]
35. Grüning DJ, Riedel F, Lorenz-Spreen P. Directing smartphone use through the self-nudge app one sec. Proc Natl Acad Sci U S A 2023 Feb 21;120(8):e2213114120. [doi: 10.1073/pnas.2213114120] [Medline: 36795756]
36. Terzimehić N, Haliburton L, Greiner P, Schmidt A, Hussmann H, MindPhone MV. Mindful reflection at unlock can reduce absentminded smartphone use. Presented at: Proceedings of the 2022 Designing Interactive Systems Conference; Jun 13-17, 2022 p. 1818-1830. [doi: 10.1145/3532106.3533575]
37. Kovacs G, Wu Z, Bernstein MS. Rotating online behavior change interventions increases effectiveness but also increases attrition. Proceedings of the ACM Hum-Comput Interact 2018 Nov;2(CSCW):1-25. [doi: 10.1145/3274364]
38. Lyngs U, Lukoff K, Csuka L, Slovák P, Van Kleek M, Shadbolt N. The Goldilocks level of support: using user reviews, ratings, and installation numbers to investigate digital self-control tools. Int J Hum Comput Stud 2022;166:102869. [doi: 10.1016/j.ijhcs.2022.102869]
39. Schwartz RX, Monge Roffarello A, Russis L, Apostolellis P. Reducing risk in digital selfcontrol tools: design patterns and prototype. Presented at: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems; May 8-13, 2021 p. 1-7. [doi: 10.1145/3411763.3451843]
40. Kim J, Jung H, Ko M, Lee U. GoalKeeper: exploring interaction lockout mechanisms for regulating smartphone use. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019;3(1):1-29. [doi: 10.1145/3314403]
41. Mustafa M, Zaman KT, Ahmad T, Batool A, Ghazali M, Ahmed N. Religion and women's intimate health: towards an inclusive approach to healthcare. 2021 May 6 Presented at: CHI '21. [doi: 10.1145/3411764.3445605]
42. Radtke T, Apel T, Schenkel K, Keller J, von Lindern E. Digital detox: an effective solution in the smartphone era? A systematic literature review. Mob Media Commun 2022;10(2):190-215. [doi: 10.1177/20501579211028647]
43. Deci EL, Ryan RM. The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 2000;11(4):227-268. [doi: 10.1207/S15327965PLI1104_01]
44. McAlaney J, Aldhayan M, Almourad MB, Cham S, Ali R. On the need for cultural sensitivity in digital wellbeing tools and messages: A UK-China comparison. In: Trends and Innovations in Information Systems and Technologies: Springer International Publishing; 2020, Vol. 2:723-733. [doi: 10.1007/978-3-030-45691-7_68]
45. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implementation Sci 2011;6(1):42. [doi: 10.1186/1748-5908-6-42]
46. Lukoff K, Lyngs U, Alberts L. Designing to support autonomy and reduce psychological reactance in digital self-control tools. In: Position Papers for the Workshop "Self-Determination Theory in HCI: Shaping a Research Agenda" at CHI'22 2022:5.
47. Brockmeier LC, Mertens L, Roitzheim C, Radtke T, Dingler T, Keller J. Effects of an intervention targeting social media app use on well‐being outcomes: a randomized controlled trial. Applied Psych Health Well 2025;17(1):e12646. [doi: 10.1111/aphw.12646]
48. Keller J, Roitzheim C, Radtke T, Schenkel K, Schwarzer R. A mobile intervention for self-efficacious and goal-directed smartphone use in the general population: randomized controlled trial. JMIR Mhealth Uhealth 2021;9(11):e26397. [doi: 10.2196/26397] [Medline: 34817388]
49. Research Randomizer. URL: https://randomizer.org [accessed 2026-03-18]

XSL•FO
RenderX

50. Simeon R, Dewidar O, Trawin J, et al. Behavior change techniques included in reports of social media interventions for promoting health behaviors in adults: content analysis within a systematic review. J Med Internet Res 2020;22(6):e16002. [doi: 10.2196/16002] [Medline: 32525482]
51. Marques MM, Wright AJ, Corker E, et al. The Behaviour Change Technique Ontology: Transforming the Behaviour Change Technique Taxonomy v1. Wellcome Open Res 2023;8:308. [doi: 10.12688/wellcomeopenres.19363.2] [Medline: 37593567]
52. Schou Andreassen C, Billieux J, Griffiths MD, et al. The relationship between addictive use of social media and video games and symptoms of psychiatric disorders: a large-scale cross-sectional study. Psychol Addict Behav 2016;30(2):252-262. [doi: 10.1037/adb0000160] [Medline: 26999354]
53. Elhai JD, Dvorak RD, Levine JC, Hall BJ. Problematic smartphone use: a conceptual overview and systematic review of relations with anxiety and depression psychopathology. J Affect Disord 2017;207:251-259. [doi: 10.1016/j.jad.2016.08.030] [Medline: 27736736]
54. Foerster M, Roser K, Schoeni A, Röösli M. Problematic mobile phone use in adolescents: derivation of a short scale MPPUS-10. Int J Public Health 2015 Feb;60(2):277-286. [doi: 10.1007/s00038-015-0660-4] [Medline: 25645102]
55. Tabachnick BG, Fidell LS. Experimental Designs Using ANOVA: Thomson/Brooks/Cole; 2007.
56. Lochmiller CR. Conducting thematic analysis with qualitative data. TQR 2021;26(6):2029-2044. [doi: 10.46743/2160-3715/2021.5008]
57. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
58. Barnett I, Torous J, Reeder HT, Baker J, Onnela JP. Determining sample size and length of follow-up for smartphone-based digital phenotyping studies. J Am Med Inform Assoc 2020 Dec 9;27(12):1844-1849. [doi: 10.1093/jamia/ocaa201] [Medline: 33043370]
59. Park J, Lee H, Park S, Chung KM, Lee U. GoldenTime: exploring system-driven timeboxing and micro-financial incentives for self-regulated phone use. : Association for Computing Machinery; 2021 May 6 Presented at: CHI '21. [doi: 10.1145/3411764.3445489]
60. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2023. URL: https://www.r-project.org/ [accessed 2025-07-19]
61. Bates D, Kliegl R, Vasishth S, Baayen H. Parsimonious mixed models. arXiv. Preprint posted online on Jun 16, 2015. [doi: 10.48550/arXiv.1506.04967]
62. Brockmeier LC. Wellspent [r code]. : OSF; 2026. [doi: 10.17605/OSF.IO/J793V]
63. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11. [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
64. Oulasvirta A, Rattenbury T, Ma L, Raita E. Habits make smartphone use more pervasive. Pers Ubiquit Comput 2012;16(1):105-114. [doi: 10.1007/s00779-011-0412-2]
65. Horwood S, Anglim J. Personality and problematic smartphone use: a facet-level analysis using the five factor model and HEXACO frameworks. Comput Human Behav 2018;85:349-359. [doi: 10.1016/j.chb.2018.04.013]
66. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]

Abbreviations

BCIO: Behavior Change Intervention Ontology
BCT: behavior change technique
BSNAS: Bergen Social Networking Addiction Scale
CONSORT: Consolidated Standards of Reporting Trials
Est: estimate
ICC: intraclass correlation
IM³UNE model: Integrative Model of Mobile Media Use and Need Experiences

Edited by L Buis; submitted 29.Jan.2024; peer-reviewed by AM Roffarello, J Lambert, O El-Gayar; revised version received 05.Dec.2025; accepted 07.Dec.2025; published 08.Apr.2026.

Please cite as:
Mertens L, Brockmeier LC, Roitzheim C, Radtke T, Dingler T, Keller J
Promoting Self-Regulated Social Media Use on Smartphones With a Mobile Intervention App (Wellspent): Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e56824
URL: https://mhealth.jmir.org/2026/1/e56824
doi:10.2196/56824

XSL•FO
RenderX

© Lea Mertens, Lina Christin Brockmeier, Christina Roitzheim, Theda Radtke, Tilman Dingler, Jan Keller. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 8.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Short-Term Effects of an mHealth Intervention on Healthy Behaviors and Cardiometabolic Health in Sedentary Employees: Quasi-Experimental Study

Yun-Ping Lin[1,2], PhD; Shu-Hua Lu[1,2], PhD; Kwo-Chen Lee[1,2], PhD; Wei-Fen Ma[1,2], PhD; Ya-Fang Ho[1,2], PhD; Wen-Chun Liao[1,2], PhD; Hui-Ting Yang[3], PhD; OiSaeng Hong[4], PhD

[1]School of Nursing, China Medical University, 100, Sec 1, Jingmao Rd, Taichung, Taiwan

[2]Department of Nursing, China Medical University Hospital, Taichung, Taiwan

[3]School of Food Safety, Taipei Medical University, Taipei, Taiwan

[4]School of Nursing, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Yun-Ping Lin, PhD
School of Nursing, China Medical University, 100, Sec 1, Jingmao Rd, Taichung, Taiwan

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2026/1/e106593

Abstract

Background: Sedentary employees face increased chronic health risks due to physical inactivity, immobility, and unhealthy eating behavior. Although mobile health (mHealth) interventions show promise in improving lifestyle behaviors, their effectiveness in occupational settings remains underexplored. Building on previous workplace interventions, this study developed and evaluated a mobile-enabled web app, SIMPLE HEALTH, developed with Din-J Design Co, Ltd, integrating activity tracking, healthy eating, and behavioral support for sedentary employees.

Objective: This study evaluated the short-term effects of a 12-week mHealth intervention on physical activity, sedentary behavior, dietary habits, and cardiometabolic health indicators among sedentary employees in Taiwan.

Methods: A 2-arm quasi-experimental study was conducted at 2 aerospace industrial workplaces. A total of 101 sedentary employees (mean age 46.9, SD 12.2 years; 52/101, 51.5% female) were enrolled from 2 worksites that were assigned by coin toss to either the intervention condition (n=50) or the control condition (n=51). The intervention group participated in the SIMPLE HEALTH program, an mHealth intervention grounded in Social Cognitive Theory and the Ecological Model, consisting of 8 components: activity tracking, goal setting, behavior logging, reminders, personalized advice, educational and motivational electronic booklets, and individual and team challenges. The control group received 6 print educational booklets. Cardiometabolic biomarkers, objectively measured physical activity (Fitbit Charge 3; Fitbit Inc), occupational sitting (occupational sitting and physical activity questionnaire), and dietary behavior (3-day photographic food records and the healthy eating behavior inventory) were assessed at baseline and 12 weeks. Data were analyzed using generalized estimating equations following the intention-to-treat principle.

Results: At 12 weeks, the intervention group showed a significant increase in step counts (adjusted mean difference, MD 1227.13, 95% CI 2.90-2451.36; *P*=.049), a more favorable between-group change in moderate physical activity (adjusted MD 0.17, 95% CI 0.01-0.33; *P*=.04), and favorable dietary behaviors, including reduced intake of calories (adjusted MD −144.59, 95% CI −276.57 to −12.60; *P*=.03), carbohydrates (adjusted MD −19.88, 95% CI −37.99 to −1.78; *P*=.03), fats (adjusted MD −6.99, 95% CI −13.69 to −0.29; *P*=.04), and grains (adjusted MD −1.46, 95% CI −2.43 to −0.50; *P*=.003), and increased vegetable intake (adjusted MD 0.47, 95% CI 0.06-0.88; *P*=.02), compared to the control group. Favorable trends were noted in diastolic blood pressure (adjusted MD −2.38, 95% CI −4.99 to 0.22; *P*=.07) and soft lean mass (adjusted MD 0.34, 95% CI −0.06 to 0.75; *P*=.10). Both groups showed significant within-group improvements in low-density lipoprotein cholesterol (intervention: *P*=.01; control: *P*=.03), body fat percentage (intervention: *P*<.001; control: *P*=.01), waist circumference (intervention: *P*=.001; control: *P*=.002), and occupational sitting (intervention: *P*<.001; control: *P*=.03), and occupational walking (intervention: *P*=.01; control: *P*=.046), but between-group differences were nonsignificant.

Conclusions: The 12-week mHealth intervention improved physical activity and dietary behaviors and showed favorable trends in cardiometabolic indicators among sedentary employees. These findings support integrating mHealth programs into employee

XSL•FO
RenderX

wellness initiatives to promote healthy behaviors, mitigate productivity loss, and reduce chronic disease burden. Further research should assess long-term sustainability, scalability, and cost-effectiveness in diverse occupational settings.

Trial Registration: ClinicalTrials.gov NCT04347096; https://clinicaltrials.gov/study/NCT04347096

(JMIR Mhealth Uhealth 2026;14:e70074) doi:10.2196/70074

KEYWORDS

sedentary work; physical activity; healthy diet; mobile health intervention; cardiometabolic health; employee

Introduction

Sedentary employees face an elevated risk of developing chronic health conditions due to physical inactivity, prolonged sitting, and poor dietary habits. In modern economies, the demand for physical labor has declined, resulting in more desk-based occupations that require employees to remain seated for extended periods. This occupational shift, compounded by the pervasive use of digital technologies, has contributed to widespread sedentary behavior in the workplace [1,2].

Furthermore, sedentary habits often extend beyond work hours, with employees spending their leisure time engaging in passive activities such as watching television, browsing social media, or using computers [3]. Globally, approximately 27.5% of adults fail to meet the World Health Organization (WHO) recommendations for physical activity [2]. In Taiwan, this issue is more pronounced—43.3% of adults do not achieve the recommended levels of physical activity [4], and 34.6% of workers report not engaging in any exercise in the past month [5]. Without effective intervention, it is projected that nearly 500 million new cases of preventable noncommunicable diseases will occur between 2020 and 2030, resulting in an annual global health care cost of approximately US $27 billion [2].

Unhealthy dietary patterns further exacerbate these risks. Office environments often promote overconsumption of calorie-dense and nutrient-poor foods, such as snacks and meals high in fat, sugar, and sodium [6,7]. In Taiwan, only 16% and 15.5% of adults meet daily vegetable and fruit intake recommendations, respectively, while average salt intake—9 g for men and 7.2 g for women—exceeds the WHO's recommended limit of less than 5 g per day [4]. Inadequate fruit and vegetable intake is estimated to contribute to 19% of gastrointestinal cancers, 31% of ischemic heart disease cases, and 11% of strokes, with diet-related cardiovascular diseases and cancers comprising approximately 85% and 15% of the total disease burden, respectively [8]. Moreover, global analyses attribute 2.4 million deaths annually to insufficient fruit and vegetable consumption and 1.8 million deaths to excess sodium intake [9].

Considering these challenges, the adoption of healthy lifestyles is crucial in preventing chronic diseases, premature death, and mitigating social and economic burdens. A large prospective cohort study [10] in Iran with more than 50,000 participants has demonstrated that combining healthy eating with physical activity significantly reduced the risk of mortality from all causes and cardiovascular diseases. Meta-analyses have also shown that a prudent or healthy diet is linked to reduced all-cause and cardiovascular disease mortality and a lower prevalence of metabolic syndrome [11,12]. A high-quality diet could lower the risk of atherothrombotic ischemic stroke, regardless of physical activity levels [13]. Dietary fiber intake is associated with a lower risk of metabolic syndrome [14], while high carbohydrate intake is linked to a higher risk [15].

Workplace interventions promoting good nutrition and physical activity have been shown to improve work-related and health-related outcomes. Grimani et al [16] conducted a review of 14 studies on nutrition and physical activity interventions in the workplace, revealing significant improvements in absenteeism, workability, work performance, and productivity. These benefits were particularly evident when interventions were designed to target the physical work environment and organizational structure. Similarly, Glympi et al [17] explored workplace dietary interventions among office workers, categorizing them into 4 main types: web-based, food-based, information-based, and multicomponent interventions. Among these, information- and web-based interventions emerged as the most prevalent, attributed to their cost-effectiveness. In this study, interventions refer to structured, evidence-based behavioral programs designed to improve health outcomes, while approaches refer to the delivery frameworks—such as eHealth and mobile health (mHealth)—used to implement them.

The growing availability of mobile devices and smartphones has made mHealth interventions increasingly attractive for promoting healthy behaviors in workplace settings. These interventions, which may include mobile apps, web-based platforms, text messaging, wearable devices, or sensor-assisted feedback, have demonstrated effectiveness in improving physical activity, diet quality, and health outcomes in various populations [18-20]. However, relatively few have been specifically designed for or tested among sedentary workers, who face unique environmental and behavioral challenges.

Building on our previous research [21,22], we aimed to extend traditional workplace interventions to a mobile platform. Our earlier "Sit Less, Walk More" studies demonstrated that modifying the physical environment and providing printed educational materials could increase physical activity, reduce sitting time, and improve cardiometabolic health and work productivity among office employees [21,22]. Nevertheless, postintervention feedback suggested that participants sought more flexibility, personalization, and accessibility—particularly outside the workplace. These findings guided the development of an evidence-based mHealth intervention, SIMPLE HEALTH, designed to integrate physical activity, sedentary behavior reduction, and healthy eating through a mobile-enabled web app.

The SIMPLE HEALTH program builds directly upon the theoretical and practical foundation established by our prior

XSL•FO
RenderX

work. It applies Social Cognitive Theory [23], Ecological Model [24], and incorporates evidence-based behavior change techniques, such as self-monitoring, goal setting, feedback, and social support. Unlike our previous environmental interventions, which relied on printed materials and workplace modifications, SIMPLE HEALTH was delivered digitally through a web app—a platform optimized for smartphones but accessible through any internet browser. The term mHealth intervention here denotes the broader behavioral program delivered via mobile technology, whereas the web app refers to the digital interface through which users accessed the program. Although a web app can technically be opened on a desktop, SIMPLE HEALTH was specifically designed for smartphone use, allowing real-time interaction, behavior tracking, and personalized reminders through LINE (LINE Corporation), a widely used instant messaging app in Asia.

While long-term behavior change is the ultimate goal of health promotion, evaluating short-term effects is a necessary first step to assess the initial efficacy, feasibility, and user engagement of newly developed digital interventions. Consistent with the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) framework, short-term evaluations provide critical feedback to refine program design and optimize future long-term trials. Therefore, this study focuses on the 12-week short-term outcomes of the SIMPLE HEALTH mHealth intervention, assessing its effects on physical activity, sedentary behavior, dietary habits, and cardiometabolic health among sedentary employees. We hypothesized that, compared with a control group receiving only print educational booklets, participants in the 12-week mHealth intervention would demonstrate significant improvements in physical activity, reductions in sitting time, healthier dietary behaviors, and favorable changes in cardiometabolic biomarkers.

Methods

Study Design and Setting

This study used a 2-arm, parallel, quasi-experimental design to examine the short-term effects of a mHealth intervention on healthy behaviors and cardiometabolic indicators among sedentary employees. The quasi-experimental design was selected to minimize contamination between groups while preserving the integrity of intervention delivery [25]. Recruitment and baseline assessments were conducted between April and May 2020 in 2 aerospace industrial workplaces in Taiwan.

Participants

Inclusion criteria were full-time employment in sedentary roles (sitting ≥6 hours per workday), age ≥20 years, absence of physical limitations restricting physical activity, and access to a smartphone with the LINE messaging app and an internet connection. Exclusion criteria were part-time employment, anticipated absence from work >2 weeks, relocation within 3 months, or pregnancy or planned pregnancy during the study period. A priori power analysis [26] (effect size=0.5; power=80%; correlation=0.5) indicated 48 participants per group were needed, accounting for a 20% dropout rate as suggested by prior work [21].

Recruitment and Allocation

Approval to conduct the study was obtained from workplace senior management. Study invitations were disseminated via company websites, email, word-of-mouth, and on-site information sessions. Occupational health nurses introduced the study at both sites, and the nurse at the intervention site was trained as a research liaison using a standardized handbook.

Participants were enrolled by occupational health nurses after obtaining site-level consent from senior management. Worksite allocation to the intervention or control condition was determined by a coin toss conducted by an independent research assistant. Group assignments were concealed until completion of baseline assessments to minimize bias. Although the quasi-experimental design limited full allocation concealment, procedures were implemented postbaseline to reduce selection bias.

Participants and the site nurse were not blinded due to the nature of the mHealth intervention. However, all data collectors (except for 1 trained research assistant) and the dietitian analyzing dietary records were blinded. Statistical analyses were conducted on deidentified data to ensure analytical blinding.

A total of 101 employees were enrolled and assigned to the intervention group (n=50) or the control group (n=51). Intervention participants also received a Mi Smart Band 4 (Anhui Huami Information Technology Co) for daily activity tracking and a participant handbook.

Procedures

Assessments were performed at baseline and 12 weeks during paid work hours. At the control site, an occupational health nurse distributed monthly educational booklets. At the intervention site, participants engaged with the SIMPLE HEALTH web app for 12 weeks. Both groups received comparable educational content, though the delivery format differed (print vs digital).

Intervention

The mHealth intervention, SIMPLE HEALTH, was grounded in Social Cognitive Theory [23] and the Ecological Model [24], emphasizing self-regulation, self-efficacy, and multilevel behavioral influences. Program design was guided by dietary and physical activity guidelines [27-29], expert consultation, and prior participant feedback [21]. To facilitate sustainable behavior change, evidence-based behavior change techniques were systematically incorporated [30].

SIMPLE HEALTH was a smartphone-optimized web app developed through iterative testing. It included eight interactive components: (1) activity tracking, (2) goal setting, (3) behavior logging, (4) reminders, (5) personalized advice, (6) electronic educational booklets, (7) electronic motivational booklets, and (8) individual and team challenges. A summary of the intervention components is presented in Multimedia Appendix 1, with screenshots in Multimedia Appendix 2 and detailed development and usability findings reported elsewhere [31].

XSL•FO
RenderX

Intervention participants accessed educational content equivalent to that of the control group but in a digital format. They were instructed to wear the activity tracker daily, set biweekly goals, log diet and activity behaviors, and use app-based reminders. Notifications were delivered through LINE. For data security, participants created personal login credentials and linked their accounts to the study's LINE Official Account using encrypted connections.

Comparator

Control participants received 6 colorful print booklets over 3 months covering 2 themes—"sit less, move more" and "healthy eating." These materials aimed to enhance knowledge, awareness, and practical skills for healthier behaviors. Control participants did not receive access to the app or activity trackers.

Measures

To assess the short-term effects of the intervention, we measured biomarkers of glucose metabolism, lipid metabolism, blood pressure, body composition, and physical activity and dietary behavior at baseline and 12 weeks. Data collection included an on-site assessment, activity monitoring, and a self-administered questionnaire.

Cardiometabolic Biomarkers

Morning on-site assessments involved collecting fasting blood samples (≥8 hours) by a trained medical technologist and a nurse. Samples were analyzed at an accredited medical laboratory for glucose, insulin, triglycerides, total cholesterol, high-density lipoprotein (HDL) cholesterol, and low-density lipoprotein (LDL) cholesterol. Blood pressure was measured twice using an automated sphygmomanometer (OMRON HEM-7320; Omron Healthcare) on the right arm, following a 3 - 5 minute rest. If readings differed by >10 mm Hg for systolic or >6 mm Hg for diastolic pressure, a third measurement was taken. The average of the 2 closest readings was used for analysis. Weight, BMI, waist circumference, percent body fat, and soft lean mass were measured using a body composition analyzer (model: ioi 353; JAWON MEDICAL). This device is certified by MDD/CE (93/42/EEC) CE 0197, meeting EN ISO 13485 standards for manufacture and quality. We adhered to guidelines to ensure accuracy and participant safety.

Physical activity was objectively measured using a triaxial accelerometer (Fitbit Charge 3; Fitbit Inc). Fitbit devices have been shown in systematic reviews to provide accurate step counts in free-living conditions [32,33]. Participants in both study groups were instructed to wear the Fitbit on the nondominant wrist during all waking hours, except while bathing or swimming, for 5 consecutive days (3 workdays and 2 nonworkdays).

Activity intensity was categorized into 5 levels based on established cut points derived from metabolic equivalents (METs) [34]: (1) sedentary behavior: <1.5 METs (eg, sitting, reclining, or lying down). (2) Light physical activity (lightly active): 1.5 to <3.0 METs (eg, slow walking and household chores). (3) Moderate physical activity (fairly active): 3.0 to <6.0 METs (eg, brisk walking). (4) Vigorous physical activity (very active): ≥6.0 METs (eg, running and aerobic exercise). (5) Moderate-to-vigorous physical activity: combined time spent in moderate and vigorous activities.

The Fitbit algorithm automatically recorded time in each intensity category, total steps, and daily energy expenditure. To complement device data, participants maintained a daily activity log to document their wake and sleep times, work hours, and any instances of device removal >15 minutes. They also recorded non-primary workplace work times.

Occupational sitting time was assessed by a reliable and valid instrument, the Chinese version of the Occupational Sitting and Physical Activity Questionnaire (OSPAQ) [22,35]. Participants estimated the proportion of time spent sitting, standing, walking, and performing physically demanding tasks during a typical workday in the past 7 days. They also reported the number of hours worked and days worked in the last 7 days. Time spent in each activity (min/d) was calculated by multiplying the percent time by hours worked per day.

Dietary behavior was objectively measured using food photography. Participants in both groups were asked to take pictures of their meals for 3 consecutive days (including 2 workdays and 1 nonworkday). They uploaded the images and descriptions of their meals to LINE, where a trained research assistant collected the data. A dietitian then analyzed each meal to determine calorie content, macronutrients (carbohydrates, protein, and fat), and daily food group intakes (grains, protein foods, vegetables, and fruits). Healthy eating behavior was assessed using the 12-item Healthy Eating Behavior Inventory (HEBI) [36], which was translated into Mandarin Chinese for this study. The HEBI-C demonstrated good reliability with a Cronbach α coefficient of 0.85. Each item was rated on a 5-point Likert scale (1=never, 5=very often), with higher scores indicating healthier eating behavior.

Demographic characteristics were assessed using single-item questions covering age, sex, education, marital status, working hours, and perceived health status. Perceived health status was measured with a question from the general health domain of the 36-Item Short Form Health Survey (SF-36) Questionnaire in Mandarin Chinese: "In general, would you say your health is excellent, very good, good, fair, or poor?" [37]. Additionally, participants were asked about their history of 18 common health conditions with the question, "Have you been diagnosed by a physician with any health conditions? (Response options: no, had in the past, currently have, or under medication control)." Smoking status was assessed with two questions: (1) "How would you describe your smoking habit? (Response options: never smoked, used to smoke, or currently smoke)" and (2) "If you currently smoke, how many cigarettes do you smoke on an average day? "

Statistical Analysis

The intention-to-treat principle was followed, where all participants were analyzed in the groups to which they were initially assigned, regardless of whether they completed the intervention or adhered to the protocol. This approach helps to minimize bias and provides a more realistic estimate of the intervention's effect by reflecting real-world scenarios where not all participants fully comply with the study procedures [38].

XSL•FO
RenderX

Descriptive statistics summarized the characteristics of the participants. The equivalence of the 2 groups was assessed using chi-square tests for categorical variables and *t* tests for interval and continuous variables. To examine the effects of the mHealth intervention on healthy behaviors and cardiometabolic health indicators, a generalized estimating equation (GEE) model was used. GEE was chosen because it accounts for intracorrelation between repeated measurements and provides unbiased estimates despite missing data [39]. GEE models with robust estimators and unstructured correlation structures were used, adjusting for baseline potentially confounding variables. For consistency across the text and tables, the group × time interaction effect from the GEE model is reported as the adjusted mean difference (adjusted MD), representing the between-group difference in change from baseline to 12 weeks. A 2-tailed significance level was set at .05. All data analyses were conducted using SPSS version 26.0 (IBM Corp).

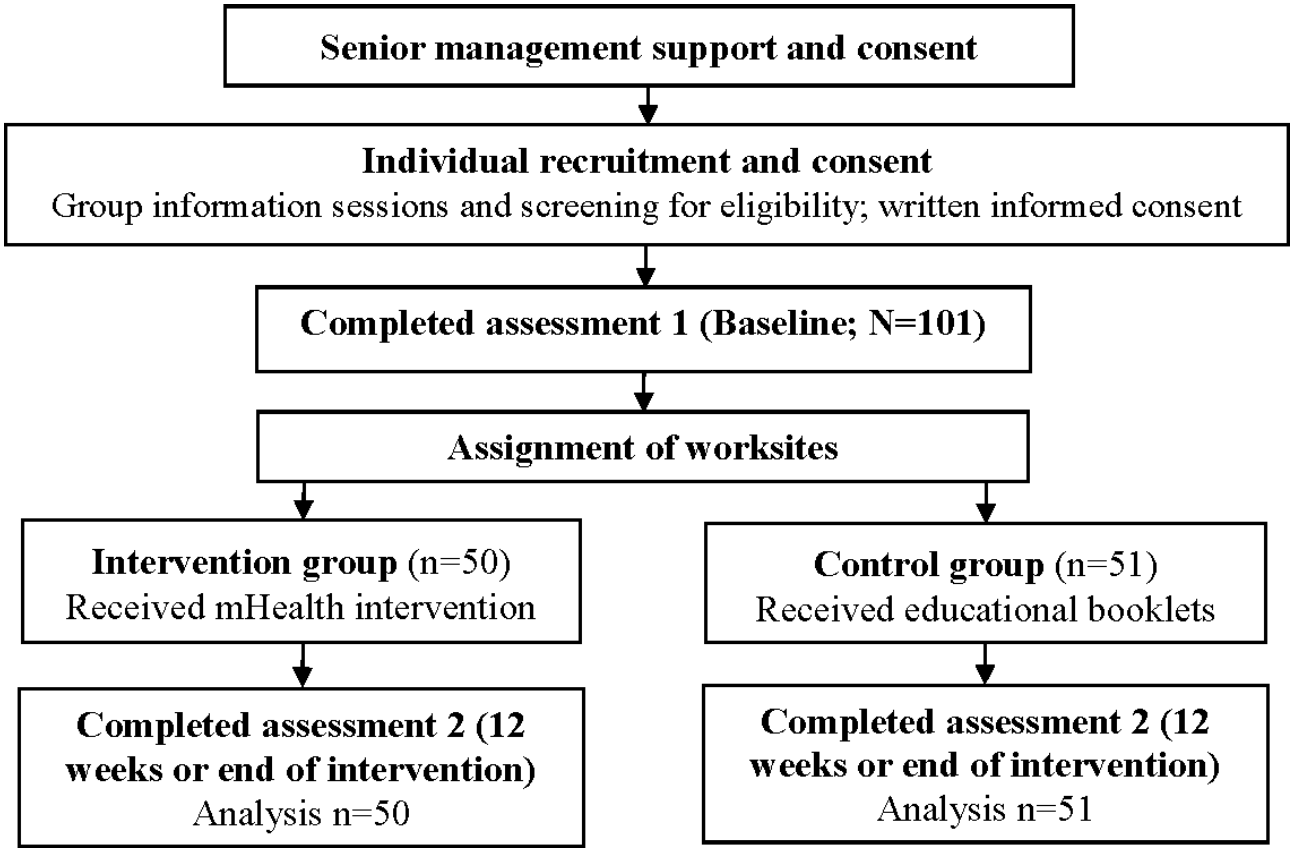
Ethical Considerations

The study protocol was approved by the China Medical University and Hospital Research Ethics Center (CRREC-109 - 017). Written informed consent was obtained from all participants before enrollment, detailing study objectives, procedures, potential risks and benefits, confidentiality, and voluntary participation rights. All data were anonymized using coded identifiers and stored on encrypted servers. Participants received gift vouchers valued at NT $100 (approximately US $3) after each assessment as recognition of their time. The trial was registered at ClinicalTrials.gov (NCT04347096; registered April 13, 2020).

Results

The study flow, including baseline and 12-week assessment participant numbers, is detailed in Figure 1. Demographic characteristics of the 2 groups are presented in Table 1. Both groups were similar at baseline, except for differences in age and marital status. The participants were predominantly middle-aged (mean age 46.9, SD 12.2 years), with a slight majority of females (52/101, 51.5%). Most participants were married or living together (69/101, 68.3%), had experienced common health conditions (79/101, 78.2%), had completed college or university education (47/101, 46.5%), perceived their health as fair or good (77/101, 76.2%), and were nonsmokers (88/101, 87.1%). The average weekly working hours were 46.0 (SD 2.16), and the average BMI was 24.5 (SD 4.07) kg/m².

Figure 1. Flow diagram of participant recruitment and retention. mHealth: mobile health.



XSL•FO
RenderX

Table . Comparison of the demographic characteristics of the participants in the intervention and control groups.

Demographic characteristics	Intervention (n=50)	Control (n=51)	*P* value[a]
Age (years), mean (SD)	50.0 (9.7)	43.9 (13.6)	.01
Sex, n (%)			.09
Female	30 (60)	22 (43)	
Married or living together, n (%)	40 (80)	29 (57)	.01
Hours worked per week, mean (SD)	45.7 (1.6)	46.4 (2.6)	.11
BMI (kg/m^2), mean (SD)	23.8 (2.82)	25.2 (4.93)	.08
Comorbid conditions[b], n (%)			.06
Yes	43 (86)	36 (71)	
Education, n (%)			.09
Junior college or lower	16 (32)	9 (18)	
College or university	18 (36)	29 (57)	
Graduate school and above	16 (32)	13 (26)	
Perceived health, n (%)			.35
Poor	2 (4)	4 (8)	
Fair	28 (56)	23 (45)	
Good	14 (28)	12 (24)	
Very good	6 (12)	12 (24)	
Excellent	0 (0)	0 (0)	
Smoking status, n (%)			.70
Never	44 (88)	44 (86)	
Ex-smoker	4 (8)	6 (12)	
Current smoker	2 (4)	1 (2)	

[a]*t* test or chi-square test between the intervention and control groups.

[b]Presence of at least 1 of 18 common health problems.

Table 2 summarizes the intervention effects on physical activity, sitting, and dietary behavior outcomes. The intervention group showed a significant increase in steps (adjusted MD 1227.13, 95% CI 2.90-2451.36; *P*=.049), a more favorable between-group change in moderate physical activity (adjusted MD 0.17, 95% CI 0.01-0.33; *P*=.04), and intake of calories (adjusted MD −144.59, 95% CI −276.57 to −12.60; *P*=.03), carbohydrates (adjusted MD −19.88, 95% CI −37.99 to −1.78; *P*=.03), fats (adjusted MD −6.99, 95% CI −13.69 to −0.29; *P*=.04), grains (adjusted MD −1.46, 95% CI −2.43 to −0.50; *P*=.003), and vegetables (adjusted MD 0.47, 95% CI 0.06-0.88; *P*=.02) compared to the control group (all *P*<.05). Although no treatment effects were seen for occupational sitting (intervention: baseline mean 7.90, SD 0.81 vs 12-week mean 7.32, SD 0.96; control: baseline mean 7.30, SD 1.23 vs 12-week mean 6.89, SD 1.36; mean difference −0.58, 95% CI −0.81 to −0.35; *P*<.001 vs mean difference −0.41, 95% CI −0.77 to −0.05; *P*=.03) and occupational walking (intervention: baseline mean 0.73, SD 0.66 vs 12-week mean 1.05, SD 0.76; control: baseline mean 0.86, SD 0.49 vs 12-week mean 1.07, SD 0.79; mean difference 0.32, 95% CI 0.08-0.57; *P*=.01 vs mean difference 0.21, 95% CI 0.004-0.415; *P*=.046), significant within-group improvements were observed in both groups.

XSL•FO
RenderX

Table . Effects of the intervention on physical activity, sitting, and dietary behavior outcomes.

Study outcomes	Intervention (n=50), mean (SD)	Control (n=51), mean (SD)	Group × time interaction effects in the GEE[a] models[b], adjusted MD[c] (95% CI)	*P* value
Fitbit				
Steps daily				
Baseline	9975.66 (3027.42)	10530.47 (4688.18)	—[d]	—
12-week	11200.20 (3761.47)	10527.89 (4728.03)	1227.13 (2.90 to 2451.36)	.049
Lightly active (h/d)				
Baseline	4.18 (0.82)	3.95 (1.03)	—	—
12-week	4.09 (1.16)	3.93 (1.00)	–0.06 (–0.41 to 0.30)	.76
Fairly active (h/d)				
Baseline	0.39 (0.25)	0.49 (0.44)	—	—
12-week	0.36 (0.30)	0.30 (0.26)	0.17 (0.01 to 0.33)	.04
Very active (h/d)				
Baseline	0.57 (0.64)	0.56 (0.56)	—	—
12-week	0.60 (0.48)	0.46 (0.47)	0.13 (–0.11 to 0.37)	.30
MVPA[e] (≥3.0 METs[f]; h/d)				
Baseline	0.96 (0.83)	1.05 (0.88)	—	—
12-week	0.96 (0.73)	0.76 (0.69)	0.30 (–0.07 to 0.66)	.11
Sitting time (h/d)				
Baseline	11.40 (1.46)	11.25 (1.52)	—	—
12-week	11.48 (1.72)	11.69 (1.38)	–0.40 (–0.97 to 0.17)	.17
OSPAQ[g]				
Sitting (h/d)				
Baseline	7.90 (0.81)	7.30 (1.23)	—	—
12-week	7.32 (0.96)	6.89 (1.36)	–0.17 (–0.60 to 0.26)	.44
Standing (h/d)				
Baseline	0.57 (0.64)	0.86 (0.74)	—	—
12-week	0.72 (0.44)	0.94 (0.79)	0.07 (–0.17 to 0.30)	.56
Walking (h/d)				
Baseline	0.73 (0.66)	0.86 (0.49)	—	—
12-week	1.05 (0.76)	1.07 (0.79)	0.11 (–0.21 to 0.43)	.49
Doing heavy labor tasks (h/d)				
Baseline	0.10 (0.27)	0.19 (0.41)	—	—
12-week	0.10 (0.26)	0.26 (0.70)	–0.07 (–0.27 to 0.14)	.54
HEBI[h]				
Healthy eating behavior				
Baseline	3.56 (0.44)	3.30 (0.51)	—	—
12-week	3.63 (0.42)	3.49 (0.48)	–0.12 (–0.28 to 0.04)	.14
3-day photographic food record				
Calories				
Baseline	1783.78 (245.22)	1713.15 (361.80)	—	—

XSL•FO
RenderX

Study outcomes	Intervention (n=50), mean (SD)	Control (n=51), mean (SD)	Group × time interaction effects in the GEE[a] models[b], adjusted MD[c] (95% CI)	*P* value
12-week	1701.15 (324.95)	1775.10 (318.71)	–144.59 (–276.57 to –12.60)	.03
Macronutrients				
CHO[i] (g/day)				
Baseline	215.70 (43.17)	204.26 (56.40)	—	—
12-week	207.91 (51.81)	216.35 (53.28)	–19.88 (–37.99 to –1.78)	.03
Protein (g/day)				
Baseline	65.93 (8.85)	63.89 (16.09)	—	—
12-week	64.58 (14.00)	62.37 (11.41)	0.16 (–5.73 to 6.06)	.96
Fats (g/day)				
Baseline	69.68 (10.63)	67.91 (15.98)	—	—
12-week	64.66 (14.17)	69.88 (14.27)	–6.99 (–13.69 to –0.29)	.04
Daily food group intakes				
Grains (portion/day)				
Baseline	10.59 (2.54)	10.25 (3.08)	—	—
12-week	10.31 (3.13)	11.43 (3.34)	–1.46 (–2.43 to –0.50)	.003
Protein foods (portion/day)				
Baseline	6.10 (1.18)	5.85 (1.96)	—	—
12-week	5.90 (1.71)	5.33 (1.44)	0.32 (–0.41 to 1.04)	.39
Vegetables (portion/day)				
Baseline	2.19 (0.97)	2.44 (1.14)	—	—
12-week	2.39 (1.19)	2.40 (1.27)	0.47 (0.06 to 0.88)	.02
Fruits (portion/day)				
Baseline	1.61 (0.95)	1.00 (1.23)	—	—
12-week	1.65 (1.14)	0.95 (1.03)	0.10 (–0.33 to 0.52)	.66

[a]GEE: generalized estimating equation

[b]Adjusted generalized estimating equations model after controlling for baseline age and sex.

[c]MD: mean difference.

[d]Not available.

[e]MVPA: moderate-to-vigorous physical activity.

[f]MET: metabolic equivalent.

[g]OSPAQ: Occupational Sitting and Physical Activity Questionnaire.

[h]HEBI: Healthy Eating Behavior Inventory.

[i]CHO: carbohydrate.

Table 3 summarizes the effects of the intervention on cardiometabolic biomarkers. The direction of the effects for diastolic blood pressure (adjusted MD −2.38, 95% CI −4.99 to 0.22; *P*=.07) and soft lean mass (adjusted MD 0.34, 95% CI −0.06 to 0.75; *P*=.10) favored the intervention. However, unfavorable changes were observed in high-density lipoprotein cholesterol (adjusted MD −3.99, 95% CI −5.94 to −2.06; *P*<.001) and triglyceride (adjusted MD 15.66, 95% CI 0.87-30.46; *P*=.04) levels. Within-group analysis showed that the intervention group experienced a favorable change in fasting glucose at 12 weeks (baseline mean 96.24, SD 13.66 vs mean 93.82, SD 8.70; mean difference −2.42, 95% CI −4.89 to 0.05; *P*=.06). While no treatment effects were seen for low-density lipoprotein cholesterol (intervention: baseline mean 118.76, SD 34.11 vs mean 112.97, SD 32.07; control: baseline mean 111.73, SD 24.97 vs mean 107.63, SD 22.45; mean difference −5.79, 95% CI −10.14 to −1.45; *P*<.01 vs mean difference −4.10, 95% CI −7.88 to −0.32; *P*=.03), body fat percentage (intervention: baseline mean 26.27, SD 6.03 vs mean 25.59, SD 6.00; control: baseline mean 26.02, SD 5.33 vs mean 25.55, SD 5.19; mean difference −0.68, 95% CI −1.04 to −0.33; *P*<.001 vs mean difference −0.47, 95% CI −0.85 to −0.10; *P*=.01), and waist circumference (intervention: baseline mean 80.32, SD 7.00 vs mean 79.62, SD 6.95; control: baseline mean 85.25, SD 13.77

XSL•FO
RenderX

vs mean 84.57, SD 13.51; mean difference −0.70, 95% CI −1.12 to −0.28; *P*=.001 vs mean difference −0.68, 95% CI −1.11 to −0.25; *P*=.002), significant within-group improvements were noted in both groups.

XSL•FO
RenderX

Table . Effects of the intervention on cardiometabolic biomarker outcomes.

Study outcomes	Intervention (n=50), mean (SD)	Control (n=51), mean (SD)	Group × time interaction effects in the GEE[a] models[b], adjusted MD[c] (95% CI)	*P* value
Weight (kg)				
Baseline	63.56 (10.34)	69.78 (18.04)	—[d]	—
12-week	63.12 (10.42)	69.03 (17.83)	0.319 (−0.256 to 0.894)	.28
BMI (kg/m^2)				
Baseline	23.83 (2.82)	25.25 (4.93)	—	—
12-week	23.67 (2.80)	24.98 (4.91)	0.103 (−0.120 to 0.325)	.37
WC[e] (cm)				
Baseline	80.32 (7.00)	85.25 (13.77)	—	—
12-week	79.62 (6.95)	84.57 (13.51)	−0.024 (−0.623 to 0.576)	.94
Body fat (%)				
Baseline	26.27 (6.03)	26.02 (5.33)	—	—
12-week	25.59 (6.00)	25.55 (5.19)	−0.207 (−0.725 to 0.310)	.43
Soft lean mass (kg)				
Baseline	43.28 (8.57)	47.22 (10.79)	—	—
12-week	43.43 (8.84)	47.03 (10.71)	0.342 (−0.061 to 0.746)	.10
WHR[f]				
Baseline	0.854 (0.07)	0.863 (0.09)	—	—
12-week	0.845 (0.07)	0.858 (0.09)	−.003 (−0.010 to 0.003)	.31
SBP[g] (mm Hg)				
Baseline	120.08 (17.56)	122.06 (14.86)	—	—
12-week	120.56 (18.98)	119.84 (13.28)	2.696 (−0.926 to 6.317)	.15
DBP[h] (mm Hg)				
Baseline	79.54 (9.74)	82.57 (11.28)	—	—
12-week	75.98 (11.19)	81.39 (10.48)	−2.384 (−4.986 to 0.219)	.07
Glucose (mg/dL)				
Baseline	96.24 (13.66)	93.98 (12.21)	—	—
12-week	93.82 (8.70)	93.53 (16.80)	−1.969 (−5.729 to 1.791)	.31
Insulin (uIU/mL)				
Baseline	11.45 (9.03)	11.00 (7.69)	—	—
12-week	10.76 (6.90)	10.75 (6.36)	−0.428 (−2.587 to 1.731)	.70
Triglycerides (mg/dL)				
Baseline	118.88 (60.78)	117.59 (58.37)	—	—
12-week	125.72 (60.89)	108.76 (54.36)	15.664 (0.871 to 30.456)	.04
Total cholesterol (mg/dL)				
Baseline	198.96 (39.38)	187.00 (31.34)	—	—
12-week	198.56 (39.47)	186.65 (29.91)	−0.047 (−6.876 to 6.782)	.99
HDL-C[i] (mg/dL)				
Baseline	59.66 (16.94)	51.80 (11.36)	—	—
12-week	55.74 (15.31)	51.88 (10.80)	−3.998 (−5.936 to –2.061)	<.001

XSL•FO
RenderX

Study outcomes	Intervention (n=50), mean (SD)	Control (n=51), mean (SD)	Group × time interaction effects in the GEE[a] models[b], adjusted MD[c] (95% CI)	*P* value
LDL-C[j] (mg/dL)				
Baseline	118.76 (34.11)	111.73 (24.97)	—	—
12-week	112.97 (32.07)	107.63 (22.45)	−1.694 (−7.451 to 4.063)	.56

[a]GEE: generalized estimating equation.

[b]Adjusted generalized estimating equations model after controlling for baseline age and sex.

[c]MD: mean difference.

[d]Not available.

[e]WC: waist circumference.

[f]WHR: waist-hip ratio.

[g]SBP: systolic blood pressure.

[h]DBP: diastolic blood pressure.

[i]HDL-C: high-density lipoprotein cholesterol.

[j]LDL-C: low-density lipoprotein cholesterol.

Discussion

Principal Results

This study evaluated the short-term effects of an mHealth program on physical activity, sitting time, dietary behavior, and cardiometabolic biomarker outcomes among sedentary employees. The intervention led to significant improvements in steps, moderate physical activity, and intake of calories, carbohydrates, fats, grains, and vegetables. Both the intervention and control groups demonstrated significant improvements in LDL cholesterol, body fat percentage, waist circumference, and occupational sitting and walking. However, the intervention group experienced unfavorable changes in triglyceride levels and HDL cholesterol. These findings contribute to the limited evidence on the benefits of mHealth programs targeting physical activity, sedentary behavior, and dietary habits, suggesting that such interventions may offer some cardiometabolic health benefits for sedentary employees.

Comparison With Prior Work

The significant improvements in physical activity (steps and moderate physical activity) and dietary behavior (intake of calories, carbohydrates, fats, grains, and vegetables) observed in our study align with previous research on mHealth interventions and web-based health promotion programs targeting multiple health behaviors, including physical activity and diet [40-43]. Cook et al [40] evaluated the effectiveness of a web-based multimedia health promotion program aimed at improving dietary practices, increasing physical activity, and reducing stress in employees of a human resources company. The intervention group showed significant improvement in attitudes toward a healthy diet and dietary stage of change. Another study by Cook et al [41] examined the effectiveness of a web-based health promotion program, HealthyPast50, on diet, physical activity, stress, and tobacco use in older employees of an information technology company. The intervention group significantly improved diet behavior change self-efficacy, planning healthy eating, and mild exercise, but not physical activity. Deitz et al [42] tested the effectiveness of a web-based cardiovascular health promotion program targeting physical activity, diet, stress, mood, and smoking in hospital employees. The intervention group significantly improved dietary attitudes, intentions, self-efficacy, exercise self-efficacy, and overall level of physical activities, particularly strenuous physical activity. Finally, Oftedal et al [43] examined the preliminary efficacy of a mHealth intervention targeting shift workers' physical activity, diet, and sleep quality. The intervention group significantly improved diet quality, though there were no significant changes in physical activity or sleep quality.

In this study, we observed inconsistent findings on the impact of increasing physical activity, reducing sitting time, and improving dietary behavior on cardiometabolic biomarkers, reflecting similar patterns found in previous research. Positive trends were noted in diastolic blood pressure, soft lean mass, and fasting glucose, though these were not statistically significant. In both intervention and control groups, there were significant improvements in LDL cholesterol, body fat percentage, and waist circumference. However, the intervention group showed unfavorable changes in triglycerides and HDL cholesterol.

The unfavorable changes in triglycerides and HDL cholesterol in the intervention group, despite promoting a low-carbohydrate diet that is high in healthy fats, may be due to several factors. Even small increases in carbohydrate intake can significantly impact lipid metabolism, potentially raising triglycerides and lowering HDL cholesterol, as the body's lipid response is highly sensitive to carbohydrate changes [44]. The type of carbohydrates and fats consumed also matters; diets high in refined carbs or low-quality fats can negatively affect lipid profiles, even with balanced macronutrient ratios [45]. Additionally, individual metabolic responses and genetic factors can lead to varied lipid responses, further complicating outcomes [44]. In brief, these unfavorable changes likely result from subtle shifts in macronutrient intake and the types of foods consumed, underscoring the complexity of dietary impacts on lipid metabolism.

XSL•FO
RenderX

Previous research on the impact of mHealth interventions on cardiometabolic biomarkers has yielded mixed results. For example, Deitz et al [42] found no significant effects of a web-based program on waist circumference and blood pressure. In contrast, Fukuoka et al [46] evaluated a diabetes prevention intervention that combined a mobile phone app and a pedometer targeting physical activity and dietary intake in overweight adults at risk for type 2 diabetes. This intervention group showed significant reductions in hip circumference and blood pressure but no significant effects on fasting lipid levels (triglycerides, total cholesterol, LDL, and HDL) or glucose levels (hemoglobin A_{1c} and glucose).

A recent review by Jung et al [47] provided a comprehensive analysis of the effectiveness of mHealth interventions in promoting physical activity and weight loss among workers. While the mHealth intervention group significantly improved physical activity, they did not show a significant difference in weight loss. Another review by Buckingham et al [48] assessed the effectiveness of mHealth technology, including wearable activity monitors and smartphone apps, for promoting physical activity and reducing sedentary behavior in workplace settings. The review found that 69% of studies reported significant impacts on reductions in weight, systolic blood pressure, and total cholesterol.

These mixed findings highlight the complexity of mHealth interventions and their varied effects on cardiometabolic health indicators. While some studies demonstrate significant benefits in specific biomarkers, others show limited or no impact. This underscores the need for further research to identify the most effective components and contexts for mHealth interventions. This includes examining the specific features that contribute to behavioral changes and improvements in health indicators. Understanding how these interventions influence behavior and health at a physiological level can help in designing more targeted and effective programs. Furthermore, it is crucial to conduct longer-term studies to assess the sustainability of behavior changes and the long-term impact on cardiometabolic health.

Of particular interest was the finding that both the intervention and control groups showed significant improvements in LDL cholesterol, body fat percentage, waist circumference, and occupational sitting and walking from baseline to 12 weeks after the intervention. This suggests either (1) both interventions effectively improved healthy behaviors and cardiometabolic health indicators or (2) most participants in both groups were highly motivated to improve their health, contributing to the observed changes. Initially, we hypothesized that the mHealth intervention group would be more effective than the control group with print educational booklets. Since this study did not have a no-treatment control group, a cautious interpretation of the intervention effect is needed. The improvements could be attributed to assessment effects for both groups. For example, after the baseline assessment, both groups received feedback on biomarkers related to glucose metabolism, lipid metabolism, blood pressure, and body composition, as well as dietary assessment along with recommendations for calorie intake and portion sizes for the 4 food groups (whole grains, protein, vegetables, and fruits). This feedback likely served as a motivating factor and benefited the control group in addressing their health and behavioral issues. Thus, this study's lack of a no-treatment control group limits our ability to definitively attribute the improvements to the interventions rather than other motivating factors or external influences. Future research should consider including a no-treatment control group to isolate the interventions' effects better.

Although the control group receiving colorful print educational booklets was originally intended as a control, the scope and quality of these materials likely played a significant role in promoting behavioral regulation and self-efficacy. As a result, the print materials effectively functioned as interventions, posing a substantial rival to the mHealth program. The effectiveness of print materials in changing health behaviors and outcomes is well-supported by research [40].

Compared to the print materials, participants in the mHealth program received electronic versions of the 6 educational booklets and additional intervention components. However, the print materials offered distinct advantages: users had all the material in hand from the start without needing to access a web app and could read the booklets easily. This is particularly advantageous for middle-aged people with presbyopia, who may prefer print materials over smartphones due to the small screen size. The apparent efficacy of the print materials suggests that providing them to the mHealth intervention group could further enhance the program's effectiveness. This finding highlights the importance of considering the format and accessibility of educational materials in health interventions, especially for populations that may have preferences or limitations related to technology use.

Implications for Occupational Health Practice

Integrating mHealth interventions into worksite health promotion programs can benefit employee well-being and productivity. The findings from this study suggest several practical implications for occupational health practitioners:

1. mHealth technologies can potentially enhance employee engagement in health promotion activities. The convenience and accessibility of web- and smartphone-based interventions can encourage more consistent participation than traditional methods. Additionally, mHealth technologies provide real-time monitoring and feedback, helping employees set their personal goals to achieve, track their progress, and stay motivated. Personalized feedback mechanisms are essential to maximize their impact.
2. Creating a supportive environment that promotes healthy behaviors is crucial for enhancing the effectiveness of health interventions. This includes providing access to mHealth tools and resources and fostering a workplace culture that values and encourages physical activity, healthy eating, and regular health monitoring.
3. Developing educational programs that leverage both digital and print materials can cater to different employee preferences. Our study showed that print materials were effective and could complement digital interventions, particularly for employees who prefer nondigital formats.

XSL•FO
RenderX

Offering a variety of educational materials ensures broader engagement.

4. Adopting a holistic approach to occupational health is essential to address multiple health behaviors simultaneously. mHealth interventions can target physical activity, sedentary behavior, diet, and other relevant health behaviors, offering a comprehensive solution for improving overall health. This approach ensures that various aspects of health are addressed together, leading to better outcomes.
5. Advocating for the integration of mHealth interventions into workplace health policies and practices is essential. Employers should be encouraged to invest in these technologies as part of their health promotion and disease prevention strategies. Integrating mHealth into the workplace can potentially sustain health improvements and cost savings in the long term.

Strengths and Limitations

This study has several notable strengths. First, it is the first mHealth intervention study to target physical activity, sedentary behavior, and dietary behaviors in a workplace context. This comprehensive approach provides valuable insights into how these interventions can be integrated into employee wellness programs. Second, the study achieved good participant retention, enhancing the reliability of the findings. Third, using objective measures to assess most outcomes, including cardiometabolic biomarkers, physical activity, and dietary behavior, adds robustness to the results.

However, the study also has limitations. First, the relatively short duration of the intervention may not capture the long-term effects on cardiometabolic health. To mitigate this, we recommend future studies with extended follow-up periods to understand the observed improvements' sustainability better. Second, the self-reported nature of the healthy eating behavior index and occupational sitting and physical activity questionnaire may introduce recall or social desirability bias. As we also included objective measures, the concern is minimal. Third, participant allocation was determined using a coin toss, which, although simple and transparent, is not an entirely unbiased method. Prior research has demonstrated that coin toss outcomes can be subtly manipulated or deviate from a true 50:50 probability due to human control or physical factors of the coin [49]. This limitation may have introduced a small but possible allocation bias, and future studies should use computer-generated randomization to minimize this risk. Finally, the generalizability of the results is limited as the study involved only 2 workplaces, and participants were not randomly selected. To enhance external validity, more studies with larger samples from diverse occupational settings and randomized sampling methods are recommended.

Conclusions

The mHealth intervention effectively improved physical activity, dietary behaviors, and certain cardiometabolic health indicators among sedentary employees. These findings suggest that integrating mHealth tools into workplace wellness programs can be beneficial. Future studies should explore long-term effects and broader applications across diverse occupational settings to confirm and extend these results.

Acknowledgments

We sincerely thank all participants and the collaborating workplaces for their invaluable contributions to this research. YPL acknowledges the financial support from the National Science and Technology Council (MOST108-2314-B-039 - 058-MY2) and China Medical University (CMU108-MF-36 and CMU109-MF-72), which made this study possible. We thank Dr. Jack Chen from ESTAT Statistical Consulting Co. Ltd. for his expert statistical support.

Funding

The financial support was from the National Science and Technology Council (MOST108-2314-B-039-058-MY2) and China Medical University (CMU108-MF-36 and CMU109-MF-72).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: YPL (lead), OSH (supporting)
Formal analysis: WCL (lead), YPL (supporting)
Funding acquisition: YPL
Investigation: YFH, HTY, YPL
Methodology: YPL, SHL, KCL, WFM, WCL, HTY, OSH
Project administration: YPL
Supervision: YPL
Validation: YPL (lead), OSH, SHL (supporting)
Writing – original draft: YPL, SHL (supporting)
Writing – review & editing: YPL (lead), SHL, OSH, KCL, WFM, WCL, YFH, HTY

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mobile health intervention components and the behavior change techniques applied.
[DOCX File, 30 KB - mhealth_v14i1e70074_app1.docx]

Multimedia Appendix 2
Screenshot of the homepage.
[PNG File, 598 KB - mhealth_v14i1e70074_app2.png]

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 859 KB - mhealth_v14i1e70074_app3.pdf]

References

1. Preventing disease through a healthier and safer workplace. World Health Organization. URL: https://www.who.int/publications/i/item/9789241513777 [accessed 2024-12-01]
2. Global status report on physical activity 2022. World Health Organization. URL: https://www.who.int/teams/health-promotion/physical-activity/global-status-report-on-physical-activity-2022 [accessed 2024-12-01]
3. Jans MP, Proper KI, Hildebrandt VH. Sedentary behavior in Dutch workers: differences between occupations and business sectors. Am J Prev Med 2007 Dec;33(6):450-454. [doi: 10.1016/j.amepre.2007.07.033] [Medline: 18022060]
4. 2024 health promotion administration annual report. Health Promotion Administration, Ministry of Health and Welfare. 2024. URL: https://www.hpa.gov.tw/EngPages/Detail.aspx?nodeid=1070&pid=18693 [accessed 2026-04-10]
5. Lee CH, Peng CL, Lin KJ. Survey of perceptions of safety and health in the work environment 2022. : New Taipei City Institute of Labor, Occupational Safety and Health, Ministry of Labor; 2023 URL: https://results.ilosh.gov.tw/iLosh/wSite/ct?xItem=40066&ctNode=322&mp=3 [accessed 2026-04-10]
6. Tanaka R, Tsuji M, Tsuchiya T, Kawamoto T. Association between work-related factors and diet: a review of the literature. Workplace Health Saf 2019 Mar;67(3):137-145. [doi: 10.1177/2165079918812481] [Medline: 30602349]
7. Toothaker R, Chikotas N. Comprehensive review of nutritional components for occupational health nurses-Part 1. Workplace Health Saf 2018 May;66(5):252-259. [doi: 10.1177/2165079917734881] [Medline: 29717659]
8. The world health report 2002: reducing risks, promoting healthy life. World Health Organization. URL: https://www.who.int/publications/i/item/9241562072 [accessed 2026-02-18]
9. GBD 2017 Diet Collaborators. Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019 May 11;393(10184):1958-1972. [doi: 10.1016/S0140-6736(19)30041-8] [Medline: 30954305]
10. Kazemi A, Sasani N, Mokhtari Z, et al. Comparing the risk of cardiovascular diseases and all-cause mortality in four lifestyles with a combination of high/low physical activity and healthy/unhealthy diet: a prospective cohort study. Int J Behav Nutr Phys Act 2022 Nov 16;19(1):138. [doi: 10.1186/s12966-022-01374-1] [Medline: 36384713]
11. Li F, Hou LN, Chen W, et al. Associations of dietary patterns with the risk of all-cause, CVD and stroke mortality: a meta-analysis of prospective cohort studies. Br J Nutr 2015 Jan 14;113(1):16-24. [doi: 10.1017/S000711451400289X] [Medline: 25430485]
12. Rodríguez-Monforte M, Sánchez E, Barrio F, Costa B, Flores-Mateo G. Metabolic syndrome and dietary patterns: a systematic review and meta-analysis of observational studies. Eur J Nutr 2017 Apr;56(3):925-947. [doi: 10.1007/s00394-016-1305-y] [Medline: 27605002]
13. Johansson A, Acosta S, Mutie PM, Sonestedt E, Engström G, Drake I. Components of a healthy diet and different types of physical activity and risk of atherothrombotic ischemic stroke: a prospective cohort study. Front Cardiovasc Med 2022;9:993112. [doi: 10.3389/fcvm.2022.993112] [Medline: 36312237]
14. Wei B, Liu Y, Lin X, Fang Y, Cui J, Wan J. Dietary fiber intake and risk of metabolic syndrome: a meta-analysis of observational studies. Clin Nutr 2018 Dec;37(6 Pt A):1935-1942. [doi: 10.1016/j.clnu.2017.10.019] [Medline: 29137803]
15. Liu YS, Wu QJ, Xia Y, et al. Carbohydrate intake and risk of metabolic syndrome: a dose-response meta-analysis of observational studies. Nutr Metab Cardiovasc Dis 2019 Dec;29(12):1288-1298. [doi: 10.1016/j.numecd.2019.09.003] [Medline: 31653521]
16. Grimani A, Aboagye E, Kwak L. The effectiveness of workplace nutrition and physical activity interventions in improving productivity, work performance and workability: a systematic review. BMC Public Health 2019 Dec 12;19(1):1676. [doi: 10.1186/s12889-019-8033-1] [Medline: 31830955]
17. Glympi A, Chasioti A, Bälter K. Dietary interventions to promote healthy eating among office workers: a literature review. Nutrients 2020 Dec 7;12(12):3754. [doi: 10.3390/nu12123754] [Medline: 33297328]

XSL•FO
RenderX

18. Hansen WB, Scheier LM. Specialized smartphone intervention apps: review of 2014 to 2018 NIH funded grants. JMIR Mhealth Uhealth 2019 Jul 29;7(7):e14655. [doi: 10.2196/14655] [Medline: 31359866]
19. Hingle M, Patrick H. There are thousands of apps for that: navigating mobile technology for nutrition education and behavior. J Nutr Educ Behav 2016 Mar;48(3):213-218. [doi: 10.1016/j.jneb.2015.12.009] [Medline: 26965099]
20. Lee M, Lee H, Kim Y, et al. Mobile app-based health promotion programs: a systematic review of the literature. Int J Environ Res Public Health 2018 Dec 13;15(12):2838. [doi: 10.3390/ijerph15122838] [Medline: 30551555]
21. Lin YP, Hong O, Lin CC, Lu SH, Chen MM, Lee KC. A “Sit Less, Walk More” workplace intervention for office workers: long-term efficacy of a quasi-experimental study. J Occup Environ Med 2018 Jun;60(6):e290-e299. [doi: 10.1097/JOM.0000000000001299] [Medline: 29438155]
22. Lin YP, Lin CC, Chen MM, Lee KC. Short-term efficacy of a “Sit Less, Walk More” workplace intervention on improving cardiometabolic health and work productivity in office workers. J Occup Environ Med 2017 Mar;59(3):327-334. [doi: 10.1097/JOM.0000000000000955] [Medline: 28267104]
23. Bandura A, Cervone D. Social Cognitive Theory: An Agentic Perspective on Human Nature, 1st edition: John Wiley & Sons; 2023.
24. McLeroy KR, Bibeau D, Steckler A, Glanz K. An ecological perspective on health promotion programs. Health Educ Q 1988;15(4):351-377. [doi: 10.1177/109019818801500401] [Medline: 3068205]
25. Pinnock H, Epiphaniou E, Taylor SJC. Phase IV implementation studies. The forgotten finale to the complex intervention methodology framework. Ann Am Thorac Soc 2014 Feb;11 Suppl 2:S118-S122. [doi: 10.1513/AnnalsATS.201308-259RM] [Medline: 24559024]
26. Elashoff JD. NQuery advisor® version 7.0 user’s guide. : Statistical Solutions Ltd; 2007 URL: http://www.statsols.com/wp-content/uploads/2013/10/nQ70_version2_manual.pdf [accessed 2026-02-18]
27. Healthy Eating Plate. Harvard TH Chan School of Public Health. 2011. URL: https://nutritionsource.hsph.harvard.edu/healthy-eating-plate/ [accessed 2024-12-01]
28. WHO Guidelines on physical activity and sedentary behaviour. World Health Organization. 2020. URL: https://iris.who.int/bitstream/handle/10665/336656/9789240015128-eng.pdf?sequence=1&isAllowed=y [accessed 2024-08-01]
29. Dietary Guidelines for Americans, 2020-2025, 9th edition: U.S. Department of Health and Human Services; 2020. URL: https://www.dietaryguidelines.gov [accessed 2021-12-01]
30. Carey RN, Connell LE, Johnston M, et al. Behavior change techniques and their mechanisms of action: a synthesis of links described in published intervention literature. Ann Behav Med 2019 Jul 17;53(8):693-707. [doi: 10.1093/abm/kay078] [Medline: 30304386]
31. Lin YP, Lee KC, Ma WF, et al. A mobile technology-based tailored health promotion program for sedentary employees: development and usability study. BMC Public Health 2025 Apr 17;25(1):1452. [doi: 10.1186/s12889-025-22401-3] [Medline: 40247254]
32. Fuller D, Colwell E, Low J, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 8;8(9):e18694. [doi: 10.2196/18694] [Medline: 32897239]
33. Germini F, Noronha N, Borg Debono V, et al. Accuracy and acceptability of wrist-wearable activity-tracking devices: systematic review of the literature. J Med Internet Res 2022 Jan 21;24(1):e30791. [doi: 10.2196/30791] [Medline: 35060915]
34. An KH, Han KA, Sohn TS, et al. Body fat is related to sedentary behavior and light physical activity but not to moderate-vigorous physical activity in type 2 diabetes mellitus. Diabetes Metab J 2020 Apr;44(2):316-325. [doi: 10.4093/dmj.2019.0029] [Medline: 31769237]
35. Chau JY, Van Der Ploeg HP, Dunn S, Kurko J, Bauman AE. Validity of the occupational sitting and physical activity questionnaire. Med Sci Sports Exerc 2012 Jan;44(1):118-125. [doi: 10.1249/MSS.0b013e3182251060] [Medline: 21659903]
36. Shimazaki T, Iio M, Lee YH, et al. Construction of a short form of the healthy eating behaviour inventory for the Japanese population. Obes Res Clin Pract 2016 Sep;10 Suppl 1:S96-S102. [doi: 10.1016/j.orcp.2015.08.011] [Medline: 26359307]
37. Lu JF, Tseng HM, Tsai YJ. Assessment of health-related quality of life in Taiwan (I): development and psychometric testing of SF-36 Taiwan version. Taiwan J Public Health 2003;22(6):501-511. [doi: 10.6288/TJPH2003-22-06-09]
38. Gupta SK. Intention-to-treat concept: a review. Perspect Clin Res 2011 Jul;2(3):109-112. [doi: 10.4103/2229-3485.83221] [Medline: 21897887]
39. Ma Y, Mazumdar M, Memtsoudis SG. Beyond repeated-measures analysis of variance: advanced statistical methods for the analysis of longitudinal data in anesthesia research. Reg Anesth Pain Med 2012;37(1):99-105. [doi: 10.1097/AAP.0b013e31823ebc74] [Medline: 22189576]
40. Cook RF, Billings DW, Hersch RK, Back AS, Hendrickson A. A field test of a web-based workplace health promotion program to improve dietary practices, reduce stress, and increase physical activity: randomized controlled trial. J Med Internet Res 2007 Jun 19;9(2):e17. [doi: 10.2196/jmir.9.2.e17] [Medline: 17581811]
41. Cook RF, Hersch RK, Schlossberg D, Leaf SL. A web-based health promotion program for older workers: randomized controlled trial. J Med Internet Res 2015 Mar 25;17(3):e82. [doi: 10.2196/jmir.3399] [Medline: 25830503]
42. Deitz D, Cook RF, Hersch RK, Leaf S. Heart healthy online: an innovative approach to risk reduction in the workplace. J Occup Environ Med 2014 May;56(5):547-553. [doi: 10.1097/JOM.0000000000000148] [Medline: 24806568]

XSL•FO
RenderX

43. Oftedal S, Burrows T, Fenton S, Murawski B, Rayward AB, Duncan MJ. Feasibility and preliminary efficacy of an m-Health intervention targeting physical activity, diet, and sleep quality in shift-workers. Int J Environ Res Public Health 2019 Oct 10;16(20):3810. [doi: 10.3390/ijerph16203810] [Medline: 31658624]
44. Rock CL, Flatt SW, Thomson CA, et al. Plasma triacylglycerol and HDL cholesterol concentrations confirm self-reported changes in carbohydrate and fat intakes in women in a diet intervention trial. J Nutr 2004 Feb;134(2):342-347. [doi: 10.1093/jn/134.2.342] [Medline: 14747670]
45. Parks EJ, Hellerstein MK. Carbohydrate-induced hypertriacylglycerolemia: historical perspective and review of biological mechanisms. Am J Clin Nutr 2000 Feb;71(2):412-433. [doi: 10.1093/ajcn/71.2.412] [Medline: 10648253]
46. Fukuoka Y, Gay CL, Joiner KL, Vittinghoff E. A novel diabetes prevention intervention using a mobile app: a randomized controlled trial with overweight adults at risk. Am J Prev Med 2015 Aug;49(2):223-237. [doi: 10.1016/j.amepre.2015.01.003] [Medline: 26033349]
47. Jung J, Cho I. Promoting physical activity and weight loss with mHealth interventions among workers: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2022 Jan 21;10(1):e30682. [doi: 10.2196/30682] [Medline: 35060913]
48. Buckingham SA, Williams AJ, Morrissey K, Price L, Harrison J. Mobile health interventions to promote physical activity and reduce sedentary behaviour in the workplace: a systematic review. Digit Health 2019;5:2055207619839883. [doi: 10.1177/2055207619839883] [Medline: 30944728]
49. Clark MPA, Westerberg BD. Holiday review. How random is the toss of a coin? CMAJ 2009 Dec 8;181(12):E306-E308. [doi: 10.1503/cmaj.091733] [Medline: 19969575]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
GEE: generalized estimating equation
HDL: high-density lipoprotein
HEBI: Healthy Eating Behavior Inventory
LDL: low-density lipoprotein
MET: metabolic equivalent
mHealth: mobile health
OSPAQ: Occupational Sitting and Physical Activity Questionnaire
SF-36: 36-Item Short Form Health Survey
WHO: World Health Organization

Edited by L Buis; submitted 15.Dec.2024; peer-reviewed by A Brzek, BC Mazzolani, F Mohsenzadeh-Ledari, RN Al-Jabali; accepted 22.Jan.2026; published 27.Apr.2026.

Please cite as:
Lin YP, Lu SH, Lee KC, Ma WF, Ho YF, Liao WC, Yang HT, Hong O
Short-Term Effects of an mHealth Intervention on Healthy Behaviors and Cardiometabolic Health in Sedentary Employees: Quasi-Experimental Study
JMIR Mhealth Uhealth 2026;14:e70074
URL: https://mhealth.jmir.org/2026/1/e70074
doi:10.2196/70074

© Yun-Ping Lin, Shu-Hua Lu, Kwo-Chen Lee, Wei-Fen Ma, Ya-Fang Ho, Wen-Chun Liao, Hui-Ting Yang, OiSaeng Hong. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 27.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Alcohol Use Trajectories During the First 72 Weeks of WHOOP Wearable Platform Membership: Observational Cohort Study

Gregory J Grosicki[1], PhD; William von Hippel[1], PhD; Finnbarr Fielding[1], BS; Christopher J Chapman[1], MS; David M Presby[2,3,4], PhD; Josh Leota[2,5,6], PhD; Kristen E Holmes[1], PhD

[1]Department of Performance Science and Health Outcomes, WHOOP, Inc., One Kenmore Square, Boston, MA, United States

[2]Department of Data Science, WHOOP, Inc., Boston, MA, United States

[3]SIB Swiss Institute of Bioinformatics, Lausanne, Vaud, Switzerland

[4]Department of Computational Biology, University of Lausanne, Lausanne, Vaud, Switzerland

[5]Division of Sleep and Circadian Disorders, Department of Medicine, Brigham and Women's Hospital, Boston, MA, United States

[6]School of Psychological Sciences, Monash University, Victoria, Australia

Corresponding Author:
Gregory J Grosicki, PhD
Department of Performance Science and Health Outcomes, WHOOP, Inc., One Kenmore Square, Boston, MA, United States

Abstract

Among 30,000 new wearable device users (11.6 million person-days), self-reported alcohol consumption significantly declined over 72 weeks, with the daily probability of drinking decreasing by 5.8 percentage points (P<.001) and reductions across age and biological sex.

(JMIR Mhealth Uhealth 2026;14:e91288) doi:10.2196/91288

KEYWORDS

behavior change; digital health; alcohol consumption; wearable technology; real-world data; health behavior; self-monitoring

Introduction

Alcohol use is a major contributor to global morbidity and mortality [1], and even low levels of consumption confer measurable health risk [2]. Despite growing awareness of this risk, alcohol remains among the most widely used drugs worldwide [3]. Although behavioral interventions can reduce intake, effects are often modest and short-lived [4], highlighting the need for scalable strategies supporting sustained, population-level reductions in alcohol use.

Wearable devices have become increasingly popular for continuous health monitoring, enabling real-time feedback on physiology and behavior. However, evidence linking wearable use to alcohol-related behaviors remains limited. A recent randomized controlled trial found that a wearable-supported behavioral program was associated with clinically meaningful reductions in alcohol use over 12 weeks [5]. To evaluate whether similar patterns emerge at scale and persist over longer durations, we examined longitudinal trajectories of self-reported alcohol consumption during the first 72 weeks of wearable membership in a large, real-world cohort.

Methods

Study Population

We studied 30,000 international adults (50.0% females) aged 18 - 79 years who joined the WHOOP platform (WHOOP, Inc.) in 2023. Participants enrolled throughout the year and contributed up to 72 weeks of data, structured into consecutive seven-day weeks anchored to each participant's join date. Quarters (Q1−Q6) were defined as consecutive 12-week intervals. Eligibility required at least 18 seven-day weeks with complete alcohol use reporting (≥25% of the observation window) [6], with additional weeks containing ≥3 entries also retained (Multimedia Appendix 1). Stratified random sampling by sex ensured balanced representation.

Self-Reported Alcohol Use

Daily alcohol use was self-reported through the WHOOP smartphone application via a customizable morning journal prompt asking users to confirm behaviors from the prior day (Figure S1 in Multimedia Appendix 1). Users indicated whether alcohol was consumed and could optionally report the number of drinks. Alcohol logging was not required for device use. Only days marked explicitly as "yes" or "no" were analyzed.

Members could also set personalized alcohol-related goals within the application, which could range from abstaining entirely to limiting drinking on specific days. Due to

XSL•FO
RenderX

heterogeneity in goal types, goal setting was analyzed as a binary indicator (any goal vs none).

Statistical Analysis

Drinking frequency was modeled as the daily probability of alcohol use within each week using generalized linear mixed-effects models (binomial, logit link) with fixed effects for quarter, age, sex, season, total weeks contributed, and weekend proportion, with participant-level random intercepts. Estimated marginal means were computed by quarter and compared with Q1 using Dunnett-adjusted contrasts. Sensitivity analyses and modeling details are described in the Multimedia Appendix 1. All analyses were performed in R (version 4.2.2; R Foundation for Statistical Computing) with statistical significance set at α=.05.

Ethical Considerations

Participants consented to their anonymized data being used for research purposes. The study protocol was approved by an independent institutional review board (Salus IRB #251121, Austin, TX). Participants did not receive compensation for their participation in this study.

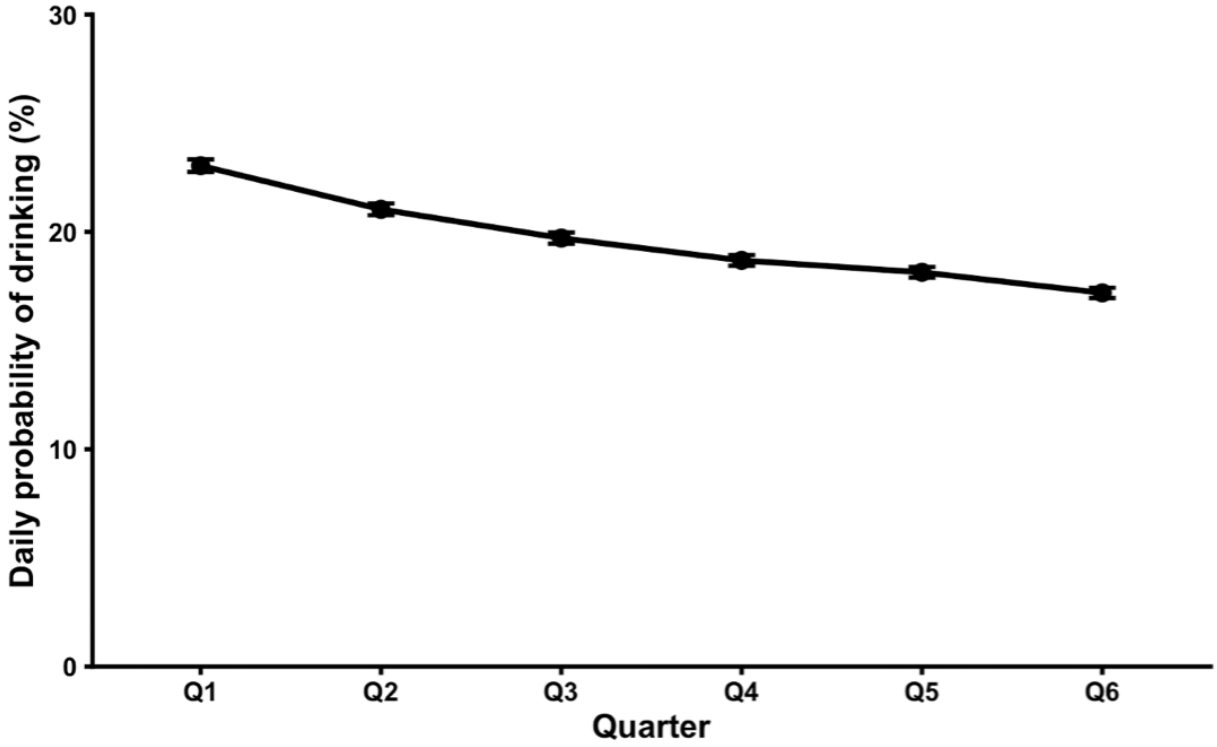
Results

The final dataset included 11,651,827 person-days (1,776,673 person-wk). Mean age was 34.4 (SD 10.5) yrs, BMI was 25.1 (SD 4.0) kg/m^2, and participants contributed a mean of 59.2 (SD 13.5) weeks (median 65 wk, IQR 18 - 72 weeks), with an average of 6.6 (SD 0.9) alcohol entries per week (Table S1 in Multimedia Appendix 1).

The daily probability of alcohol use declined over time (Figure 1), from 23.0% (95% CI 22.7 to 23.3) in Q1 to 17.2% (95% CI 16.9 to 17.4) in Q6, an absolute reduction of 5.8 percentage points (95% CI −6.0 to −5.7). Results were robust in sensitivity analyses, including restriction to complete seven-day drink number reporting, where drinking volume decreased by 1.11 drinks per week (95% CI −1.15 to −1.07; Figure S2 in Multimedia Appendix 1).

XSL•FO
RenderX

Figure 1. Adjusted daily probability of alcohol use during the first 72 weeks of membership. Marginal estimates from a binomial generalized linear mixed-effects model are shown for six sequential 12-week quarters (Q1–Q6). Points represent model-based predicted probabilities that a logged day was a drinking day, and error bars indicate 95% CIs. Absolute changes versus Q1 (percentage points) and corresponding Dunnett-adjusted *P* values are derived from model-based contrasts. The model included fixed effects for quarter, age, biological sex, season, total number of weeks contributed, and the proportion of days in each user-week that fell on a weekend, with a random intercept for participant.

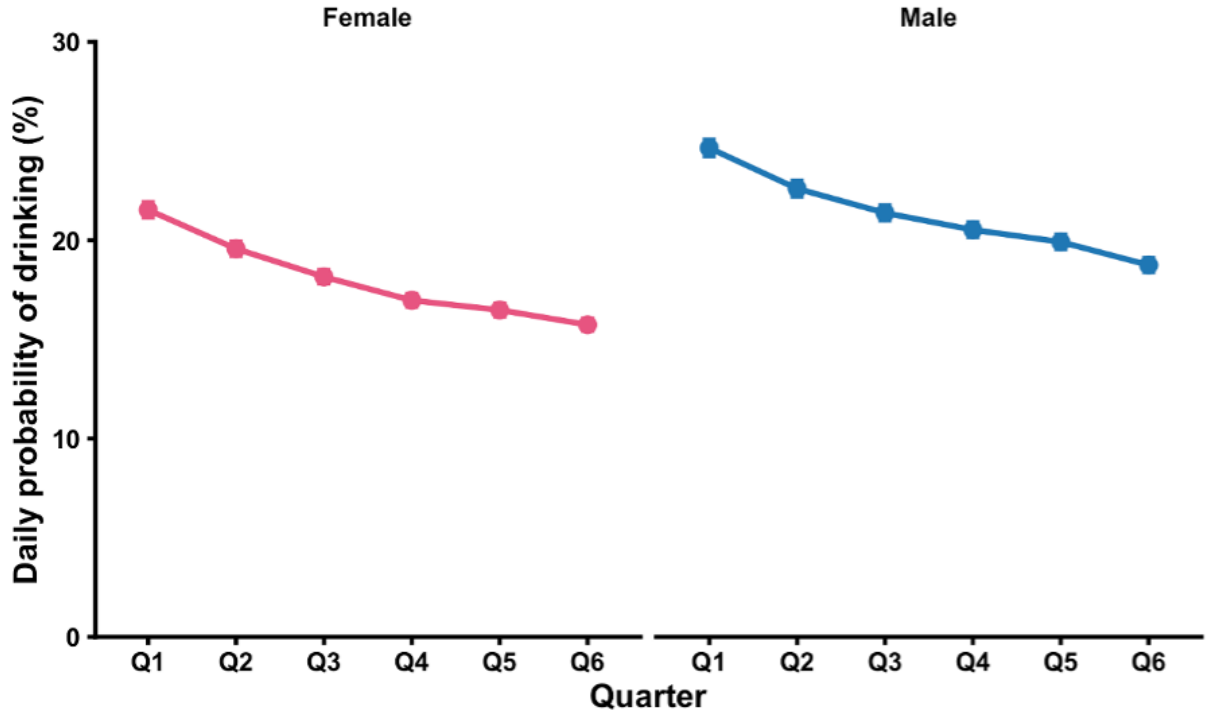


Quarter	Daily Probability (%) (95% CI)	Absolute Change vs Q1 (pp, 95% CI)	*P*-value
Q1	23.0 (22.7–23.3)	—	—
Q2	21.0 (20.7–21.3)	−2.0 (−2.1 to −1.9)	<.001
Q3	19.7 (19.4–20.0)	−3.3 (−3.4 to −3.2)	<.001
Q4	18.7 (18.4–18.9)	−4.3 (−4.5 to −4.2)	<.001
Q5	18.1 (17.9–18.4)	−4.9 (−5.0 to −4.8)	<.001
Q6	17.2 (16.9–17.4)	−5.8 (−6.0 to −5.7)	<.001

The baseline daily drinking probability was higher among males than females (*P*<.001), but the relative decline from Q1 to Q6 was greater in females (*P*<.001; Figure 2). Baseline daily drinking probability differed across age groups (*P*s <.001). Progressively larger relative and absolute declines in drinking were observed with increasing age (*P*s <.001; Figure 3).

XSL•FO
RenderX

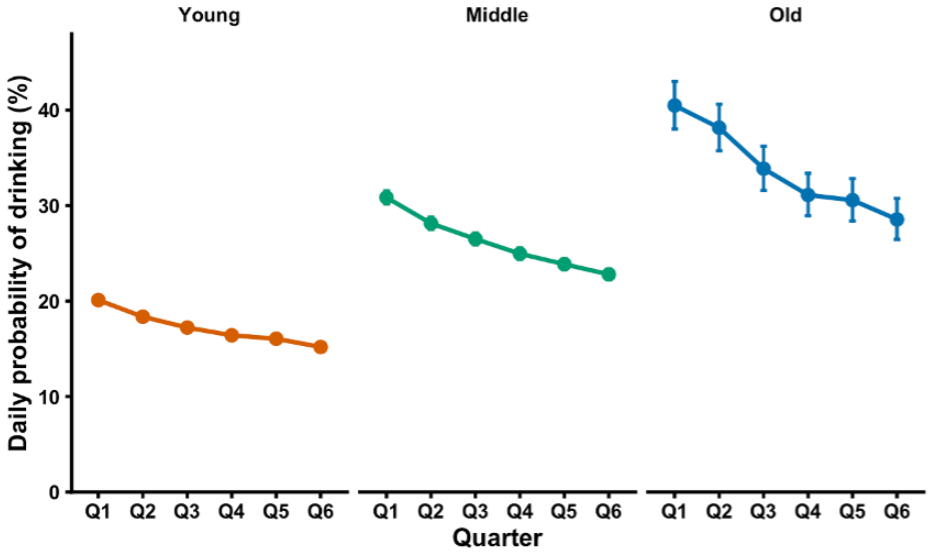
Figure 2. Adjusted daily probability of alcohol use during the first 72 weeks of membership, stratified by biological sex. Estimated marginal means from a binomial generalized linear mixed-effects model are shown separately for females and males across six sequential 12-week quarters (Q1–Q6). Points represent model-based predicted probabilities that a logged day was a drinking day, and error bars indicate 95% CIs. Absolute percentage-point changes relative to Q1 and associated Dunnett-adjusted *P* values were derived from model-based contrasts within each sex. Models adjusted for age, season, total number of weeks contributed, and the proportion of days in each user-week that fell on a weekend, with a participant-level random intercept.



Quarter	Females			Males		
	Daily Probability (%) (95% CI)	Absolute Change vs Q1 (pp, 95% CI)	*P*-value	Daily Probability (%) (95% CI)	Absolute Change vs Q1 (pp, 95% CI)	*P*-value
Q1	21.5 (21.1–21.9)	—	—	24.6 (24.2–25.0)	—	—
Q2	19.5 (19.2–19.9)	−1.9 (−2.1 to −1.8)	<.001	22.6 (22.2–23.0)	−2.0 (−2.2 to −1.9)	<.001
Q3	18.1 (17.8–18.5)	−3.4 (−3.5 to −3.2)	<.001	21.4 (21.0–21.7)	−3.3 (−3.4 to −3.1)	<.001
Q4	17.0 (16.6–17.3)	−4.5 (−4.7 to −4.4)	<.001	20.5 (20.1–20.9)	−4.1 (−4.3 to −3.9)	<.001
Q5	16.5 (16.1–16.8)	−5.0 (−5.2 to −4.9)	<.001	19.9 (19.5–20.3)	−4.7 (−4.9 to −4.5)	<.001
Q6	15.7 (15.4–16.0)	−5.8 (−6.0 to −5.6)	<.001	18.7 (18.4–19.1)	−5.9 (−6.1 to −5.7)	<.001

XSL•FO
RenderX

Figure 3. Adjusted daily probability of alcohol use during the first 72 weeks of membership, stratified by age group. Estimated marginal means from a binomial generalized linear mixed-effects model are shown separately for participants aged 18–39 years (n = 21,670), 40–59 years (n = 7,610), and 60–79 years (n = 720) across six sequential 12-week quarters (Q1–Q6). Points represent model-based predicted probabilities that a logged day was a drinking day, and error bars indicate 95% CIs. Absolute percentage-point changes relative to Q1 and corresponding Dunnett-adjusted *P* values were derived from model-based contrasts within each age group. Models adjusted for biological sex, season, total number of weeks contributed, and the proportion of days in each user-week that fell on a weekend, with a participant-level random intercept.



Quarter	Young			Middle			Old		
	Daily Probability (%) (95% CI)	Absolute Change vs Q1 (pp, 95% CI)	*P*-value	Daily Probability (%) (95% CI)	Absolute Change vs Q1 (pp, 95% CI)	*P*-value	Daily Probability (%) (95% CI)	Absolute Change vs Q1 (pp, 95% CI)	*P*-value
Q1	20.1 (19.8–20.4)	—	—	30.8 (30.1–31.5)	—	—	40.4 (37.9–42.9)	—	—
Q2	18.3 (18.1–18.6)	−1.7 (−1.9 to −1.6)	<.001	28.1 (27.4–28.7)	−2.7 (−2.9 to −2.4)	<.001	38.1 (35.6–40.5)	−2.3 (−3.3 to −1.4)	<.001
Q3	17.2 (16.9–17.5)	−2.9 (−3.0 to −2.7)	<.001	26.5 (25.8–27.1)	−4.3 (−4.6 to −4.1)	<.001	33.8 (31.5–36.2)	−6.6 (−7.5 to −5.6)	<.001
Q4	16.4 (16.1–16.7)	−3.7 (−3.8 to −3.5)	<.001	24.9 (24.3–25.5)	−5.9 (−6.1 to −5.6)	<.001	31.0 (28.9–33.3)	−9.3 (−10.3 to −8.3)	<.001
Q5	16.0 (15.8–16.3)	−4.0 (−4.2 to −3.9)	<.001	23.8 (23.3–24.4)	−6.9 (−7.3 to −6.7)	<.001	30.5 (28.3–32.7)	−9.9 (−11.0 to −8.9)	<.001
Q6	15.2 (14.9–15.4)	−4.9 (−5.0 to −4.7)	<.001	22.7 (22.2–23.3)	−8.1 (−8.3 to −7.7)	<.001	28.5 (26.4–30.7)	−11.9 (−13.0 to −10.8)	<.001

Both goal setters (n=13,298; Table S2 in Multimedia Appendix 1) and nongoal setters showed significant declines, though the relative decline was greater among non-goal setters (*P*<.001; Figure S5 in Multimedia Appendix 1).

Discussion

In this large real-world cohort of 30,000 WHOOP members, self-reported alcohol use declined over the first 72 weeks of membership, with both drinking frequency and volume decreasing across age and sex. These findings extend recent experimental results [5] by describing sustained reductions in alcohol use over 72 weeks in a large real-world cohort.

The observed declines may reflect increased awareness of health behaviors, heightened sensitivity to physiological feedback, or broader lifestyle changes after wearable enrollment [7]. Importantly, participants engaged in active self-monitoring through journal entries, a behavior shown to promote reductions in drinking [8]. Additionally, large-scale wearable-based analyses demonstrate that even moderate alcohol consumption disrupts cardiac autonomic regulation, sleep, and next-day physical activity [9]. Awareness of these consequences may reinforce behavior change. Notably, declines in drinking frequency were accompanied by reductions in volume, suggesting decreased frequency was not offset by increased intake on drinking occasions. Although causal effects cannot be inferred, the persistence of these changes over more than one year highlights the potential for digital health platforms to support sustained alcohol-related behavior change at scale.

Several limitations should be considered. Alcohol use was self-reported without standardized drink definitions, and participants were subscription-based wearable users who may be more health-conscious than the general population. The requirement for sustained reporting engagement (≥18 wk) may further limit generalizability. Although reporting frequency remained relatively high across quarters and sensitivity analyses restricted to complete reporting weeks and members with full 72-week follow-up yielded nearly identical results, informative missingness cannot be excluded. Without a control group, observed changes cannot be causally attributed to wearable enrollment or distinguished from secular trends [10]. Although statistically significant, absolute changes were modest, and clinical significance at the individual level cannot be determined. However, dose-response evidence indicates that alcohol-related

XSL•FO
RenderX

health risks increase continuously with consumption [11], with no established safe threshold [12], suggesting even modest population-level reductions may carry public health relevance, particularly when sustained over time and observed across demographic subgroups. Longer follow-up is needed to evaluate persistence beyond the study window. Despite these limitations, this study provides one of the largest longitudinal assessments of self-reported alcohol use to date, describing long-term trajectories of drinking frequency and volume over 72 weeks among new WHOOP members.

Acknowledgments

The authors wish to acknowledge the technical assistance of the WHOOP Data Science and Data Engineering teams for their contributions to feature development and data infrastructure. We also thank the broader WHOOP employee community for their continued support of the company's research initiatives. The authors used the generative AI tool ChatGPT (OpenAI) for language editing.

Funding

This work was supported by WHOOP, Inc. through salary support provided to authors GJG, WvH, FF, CC, DMP, JL, and KEH. No specific grant numbers are associated with this work. WHOOP, Inc. (https://whoop.com/) provided general institutional support but had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

The data and code that support the findings of this study are not publicly available due to intellectual property concerns of WHOOP, Inc. Deidentified participant data and a data dictionary may be made available upon reasonable request to WHOOP (via research@whoop.com). Access will require submission of a methodologically sound proposal, approved by WHOOP's research team, and a signed data use agreement. The timeframe for response to requests will be 4 weeks.

Authors' Contributions

Conceptualization: GJG, KEH, WvH, FF, CJC, DMP, JL
Data Curation: GJG
Formal Analysis: GJG (lead), WvH and JL (supporting)
Project Administration: KEH
Visualization: GJG
Writing - original draft: GJG
Writing - reviewing & editing: GJG, KEH, WvH, FF, CJC, DMP, JL

Conflicts of Interest

The study was funded by WHOOP, Inc., GJG, WvH, FF, CC, DMP, JL, and KEH are employees of WHOOP, Inc., and GJG, FF, CC, DMP, JL, and KEH hold stock options as part of their employee compensation, which represents a potential financial interest. All analysis was performed solely by the research and data science teams. Marketing, enterprise, and other non-research groups at WHOOP, Inc. had no role in study approval, data interpretation, manuscript preparation, or decisions regarding submission or publication.

Multimedia Appendix 1
Extended methods, sensitivity analyses, and subgroup results with supporting tables and figures.
[DOCX File, 1489 KB - mhealth_v14i1e91288_app1.docx]

References

1. Ortolá R, Sotos-Prieto M, García-Esquinas E, Galán I, Rodríguez-Artalejo F. Alcohol consumption patterns and mortality among older adults with health-related or socioeconomic risk factors. JAMA Netw Open 2024 Aug 1;7(8):e2424495. [doi: 10.1001/jamanetworkopen.2024.24495] [Medline: 39133491]
2. Griswold MG, Fullman N, Hawley C, et al. Alcohol use and burden for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet 2018 Sep;392(10152):1015-1035. [doi: 10.1016/S0140-6736(18)31310-2]
3. Global status report on alcohol and health and treatment of substance use disorders. : World Health Organization; 2024 URL: https://www.who.int/publications/i/item/9789240096745 [accessed 2025-12-16]
4. O'Logbon J, Wickersham A, Williamson C, Leightley D. The effectiveness of digital health technologies for reducing substance use among young people: a systematic review & meta-analysis. J Ment Health 2024 Oct;33(5):645-673. [doi: 10.1080/09638237.2023.2245902] [Medline: 37664884]

XSL•FO
RenderX

5. Fucito LM, Ash GI, Wu R, et al. Wearable intervention for alcohol use risk and sleep in young adults: a randomized clinical trial. JAMA Netw Open 2025 May 1;8(5):e2513167. [doi: 10.1001/jamanetworkopen.2025.13167] [Medline: 40445615]
6. Turner-McGrievy GM, Dunn CG, Wilcox S, et al. Defining adherence to mobile dietary self-monitoring and assessing tracking over time: tracking at least two eating occasions per day is best marker of adherence within two different mobile health randomized weight loss interventions. J Acad Nutr Diet 2019 Sep;119(9):1516-1524. [doi: 10.1016/j.jand.2019.03.012] [Medline: 31155473]
7. Grosicki GJ, Fielding F, Kim J, et al. Wearing WHOOP more frequently is associated with better biometrics and healthier sleep and activity patterns. Sensors (Basel) 2025 Apr 12;25(8):2437. [doi: 10.3390/s25082437] [Medline: 40285124]
8. Carpenter SM, Yap J, Patrick ME, et al. Self-relevant appeals to engage in self-monitoring of alcohol use: a microrandomized trial. Psychol Addict Behav 2023 May;37(3):434-446. [doi: 10.1037/adb0000855] [Medline: 35834200]
9. Grosicki GJ, Robinson AT, Joyner MJ, et al. Real-world effects of alcohol on heart rate, sleep, and physical activity by age and sex. PLOS Digit Health 2026 Mar;5(3):e0001284. [doi: 10.1371/journal.pdig.0001284] [Medline: 41801993]
10. Kerr WC, Lui CK, Ye Y, et al. Long-term trends in beverage-specific drinking in the National Alcohol Surveys: Differences by sex, age, and race and ethnicity. Alcohol Clin Exp Res (Hoboken) 2024 Jul;48(7):1322-1335. [doi: 10.1111/acer.15335] [Medline: 38658368]
11. Rehm J, Rovira P, Llamosas-Falcón L, Shield KD. Dose-response relationships between levels of alcohol use and risks of mortality or disease, for all people, by age, sex, and specific risk factors. Nutrients 2021 Jul 30;13(8):2652. [doi: 10.3390/nu13082652] [Medline: 34444809]
12. Morrow L, Greenwald B. Spread the word: no amount of alcohol is safe!. Gastroenterol Nurs 2024;47(4):260-264. [doi: 10.1097/SGA.0000000000000818] [Medline: 39087991]

Abbreviations

IRB: institutional review board
Q1–Q6: Quarter 1 through Quarter 6

Edited by A Stone; submitted 16.Jan.2026; peer-reviewed by KJ Kim, L Fucito, R Davis-Martin; revised version received 10.Apr.2026; accepted 10.Apr.2026; published 27.Apr.2026.

Please cite as:
Grosicki GJ, Hippel WV, Fielding F, Chapman CJ, Presby DM, Leota J, Holmes KE
Alcohol Use Trajectories During the First 72 Weeks of WHOOP Wearable Platform Membership: Observational Cohort Study
JMIR Mhealth Uhealth 2026;14:e91288
URL: https://mhealth.jmir.org/2026/1/e91288
doi:10.2196/91288

© Gregory J Grosicki, William von Hippel, Finnbarr Fielding, Christopher J Chapman, David M Presby, Josh Leota, Kristen E Holmes. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 27.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

One-Year Trajectory of Step Counts and Weight Loss in Adults With Overweight/Obesity: Retrospective Cohort Study

Kenshiro Taguchi[1,2,3*], MS; Asuka Oyama[2,4*], PhD; Jun'ichi Kotoku[1,2,5], PhD; Hiroshi Toki[2], PhD; Ryohei Yamamoto[2,6,7,8], MD, PhD

[1]Graduate Degree Program of Health Data Science, Teikyo University, 2-11-1 Kaga, Itabashi, Japan

[2]Health and Counseling Center, The University of Osaka, Toyonaka, Japan

[3]Fundamental Technology Research Department, Biochemical Research Laboratory, Eiken Chemical Co, Ltd, Nogi, Japan

[4]Epidemiology Section, Division of Public Health, Osaka Institute of Public Health, Osaka, Japan

[5]Graduate School of Medical Care and Technology, Teikyo University, Itabashi, Japan

[6]Department of Nephrology, Graduate School of Medicine, The University of Osaka, Suita, Japan

[7]Laboratory of Behavioral Health Promotion, Department of Health Promotion, Graduate School of Medicine, The University of Osaka, Suita, Japan

[8]Institute for Sports and Global Health, The University of Osaka, Suita, Japan

[*]these authors contributed equally

Corresponding Author:
Jun'ichi Kotoku, PhD
Graduate Degree Program of Health Data Science, Teikyo University, 2-11-1 Kaga, Itabashi, Japan

Abstract

Background: Being overweight and obese are major health concerns worldwide, contributing to lifestyle-related diseases such as hypertension, dyslipidemia, type 2 diabetes, and cardiovascular disease. Increasing physical activity is an effective strategy for weight management. However, earlier step count studies have remained limited to small populations, short-term measurements of 1 - 2 weeks, and mainly cross-sectional comparisons of average step counts. The effects of long-term step count changes on weight loss remain unclear.

Objective: This study was conducted to assess the effects of long-term patterns of step counts on weight loss using data from the “Asmile” mobile health app in Japan. We hypothesized that participants with continuously increasing step counts over time would have a higher likelihood of significant weight reduction than participants who show steady or fluctuating patterns, even if their average step counts were similar.

Methods: We analyzed data of 2778 Asmile users aged 40 - 74 years with BMI ≥25 kg/m² who underwent a specific health checkup during fiscal years 2019 - 2023 and who had valid step count records for 10 - 14 months. Step count trajectories, reflecting long-term trends in physical activity, were classified using a latent class mixed model into four patterns: *UP* (increasing), *FLAT* (steady), *DOWN* (decreasing), and *UP/DOWN* (increasing then decreasing). Logistic regression was applied to estimate odds ratios for achieving ≥3% weight loss, with step trajectory as the explanatory variable and weight loss as the outcome.

Results: Among participants, 1601 (57.6%) were men and 1177 (42.4%) were women, with respective mean ages of 65.8 (SD 7.9) and 64 (SD 8.2) years. Step count trajectories were distributed as 28.5% *UP*, 36.2% *FLAT*, 20.1% *DOWN*, and 15.2% *UP/DOWN*. Compared with the *FLAT* group, participants in the *UP* group had a significantly higher likelihood of achieving ≥3% weight loss (adjusted odds ratio 2.45, 95% CI 1.78 - 3.38).

Conclusions: Long-term tracking of step counts using the Asmile app revealed distinct activity patterns. Continuous increases in step counts were associated with the greatest likelihood of weight loss, emphasizing the importance of sustained physical activity. These findings support the use of long-term step monitoring to guide interventions for obesity and lifestyle-related disease prevention.

(JMIR Mhealth Uhealth 2026;14:e80339) doi:10.2196/80339

KEYWORDS

latent class mixed models; mHealth app; step counts; trajectory analysis; weight loss; physical activity; mobile health; smartphone; mobile phone; exercise; app; mHealth; observational study; Osaka; Japan; healthy lifestyle; lifestyle-related disease

XSL•FO
RenderX

Introduction

Being overweight and obese have come to pose medical and social problems in economically advanced countries, including Japan, and in some low- and middle-income countries [1-4]. According to the World Health Organization, which provides widely accepted global guidelines, people with a BMI of 25 kg/m² or greater are classified as overweight, whereas those with a BMI of 30 kg/m² or greater are classified as obese [4]. Obesity is associated with lifestyle-related diseases such as hypertension [5], dyslipidemia [6], type 2 diabetes [7], and arteriosclerotic cardiovascular disease (CVD) [8]. In Japan, the Japan Society for the Study of Obesity in 2000 defined obesity as having a BMI greater than 25 kg/m^2 [9]. To avoid confusion caused by different definitions among countries, we designate both being overweight and being obese as "obesity" throughout this paper. For 3480 Japanese people with obesity or metabolic syndrome, Muramoto et al [10] reported that achievement of 3% weight loss improves risk factors associated with obesity. In obese populations, promoting weight loss is useful and important to reduce the risks of developing lifestyle-related diseases. Therefore, weight loss is often necessary to maintain health.

Walking is a daily physical activity that promotes weight loss and improves lifestyle-related diseases. Pedometers have been used for objective measurement of an individual's per-day step count. Earlier cross-sectional studies comparing BMI and average daily step counts have shown that individuals with higher step counts tend to have lower body weight than those with fewer steps [11-13]. A negative correlation was found between short-term physical activity and physical status [14]. However, earlier studies using pedometers have been burdened by several limitations: small sample sizes and short-term monitoring periods, typically 1 - 2 weeks. Moreover, although a negative correlation has been found, those results derive only from cross-sectional analyses. They do not reflect changes in step counts and weight loss induced by these changes. Therefore, relations between long-term and longitudinal step count changes and weight loss remain unclear.

For this study, we applied trajectory analysis to assess long-term changes in step counts. Latent class mixed models (LCMMs), a trajectory analysis approach, can cluster heterogeneous populations into latent and more homogeneous trajectories [15]. They are then classifiable by group, reflecting different long-term changes. Such models have been applied to the identification of long-term clinical phenotypic changes and elucidation of their group characteristics because skin-thickening trajectories are associated with organ involvement and survival [16]. The trajectories of systemic lupus erythematosus support the feasibility of performing adaptive trial designs [17]. Therefore, by analyzing the trajectories as variables reflecting homogeneous changes over a long period, one can better understand the near-causal effects on outcomes from those trajectories [18].

Therefore, this study was conducted to elucidate the effects of long-term fluctuations in step counts on weight reduction. From this observational study, we obtained approximately 1 year of step count data from 2778 users of the "Asmile" mobile health care (mHealth) app. Users were classified based on their step count trajectories inferred using LCMM. We hypothesized that, rather than simply comparing average step counts, long-term walking patterns would influence subsequent weight loss: specifically, participants with a continuous increase in step counts, rather than a temporary increase, would be more likely to achieve significant weight reduction.

Methods

Mobile Health Care App "Asmile"

The Asmile mHealth app, available on iOS and Android, was released in 2019. It is provided by the Osaka Prefectural Government [19]. As of May 2025, more than 450,000 Asmile users have adopted this app to promote their health by recording daily activities and facilitating self-monitoring. One Asmile feature allows a user to record the walking step count automatically as a daily activity by integration with standard smartphone health care apps. Furthermore, daily user information (weight, body temperature, etc) and lifestyle habits (sleeping hours, breakfast intake, tooth brushing, etc) can be recorded manually in the Asmile app. Users with healthy lifestyles, such as those who walk a lot, are awarded points, which they can use to enter drawings for prizes such as electronic money. The granting of such digital incentives encourages more walking [20,21].

Another feature is that results of specific health checkups (SHCs) for National Health Insurance subscribers can be synchronized automatically with the Asmile app. This SHC is aimed at primary prevention and early detection of obesity and lifestyle-related diseases for all insured persons aged 40 - 74 years [22-24]. SHC data such as height, weight, BMI, blood test data, and urinalysis data are recorded. Therefore, the Asmile app allows users to check daily activity records over the long term and to check their health status by fiscal year. Using the Asmile app promotes healthy daily activities and increases awareness of one's health, which is expected to prevent lifestyle-related diseases.

Furthermore, recent usage of the Asmile app has revealed effects of the declaration of a state of emergency for COVID-19 on smoking behavior and the relation between oral frailty and falls [25,26], in addition to a causal effect on increased step counts [27]. Utilization promotes healthy daily activities and increases awareness of one's health, both of which are expected to prevent lifestyle-related diseases.

Participants

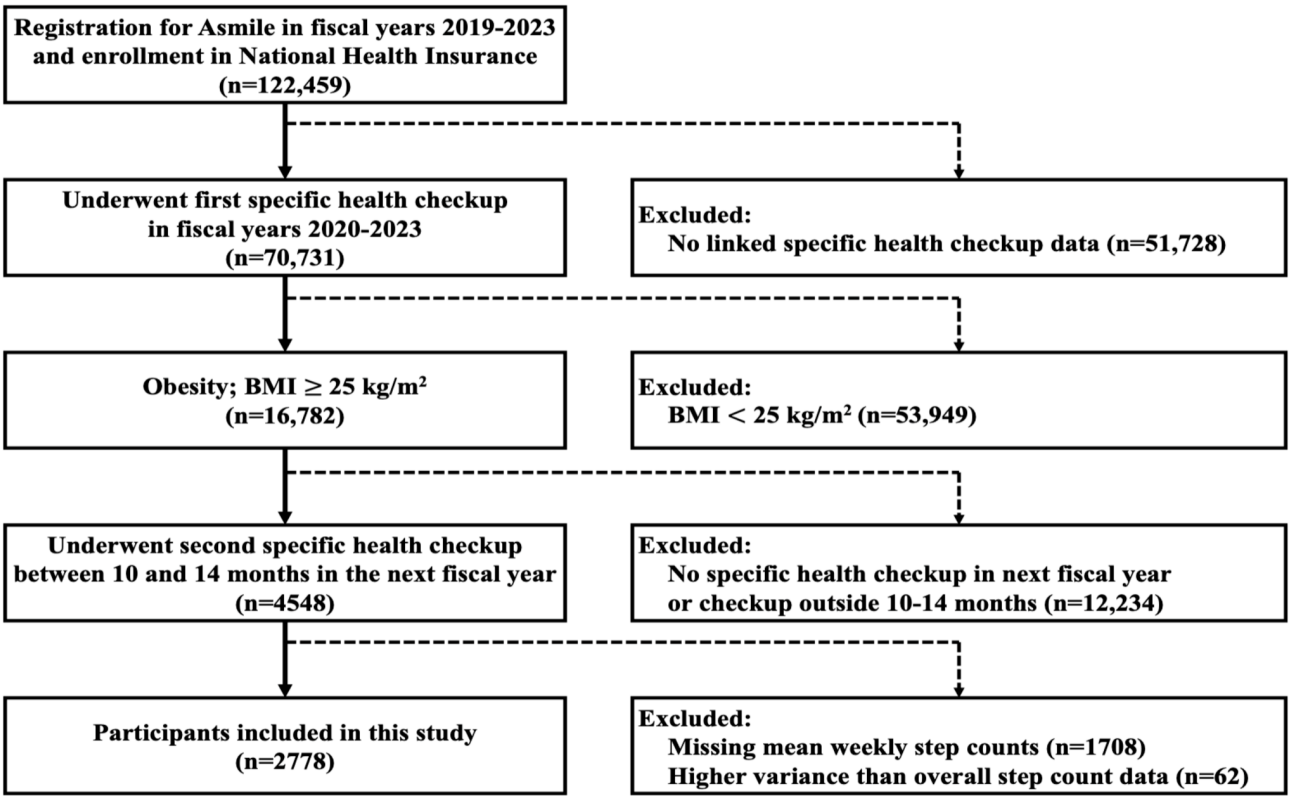
This study included participants who registered for the Asmile app during fiscal years 2019 - 2023 and who underwent their first SHC after registration and before the end of fiscal year 2023. We excluded participants with a BMI of less than 25 kg/m². In addition, because this study specifically addressed changes in step count and weight over a period of approximately 1 year, we excluded users who did not undergo a second SHC 10 - 14 months after the first SHC. For participants with more than one eligible SHC pair, we selected the SHC pair for which

XSL•FO
RenderX

the interval between the first and second SHC was closest to 12 months.

In total, data from 122,459 users enrolled in the National Health Insurance and newly registered for the Asmile app were screened. Among those users, 70,731 underwent the first SHC during the 2019 - 2023 fiscal years. After application of the exclusion criteria, 2778 participants with complete step count records and SHC links were included in the analysis (Figure 1).

Figure 1. Flowchart of the included participants' selection for analysis in this study.



Step Count Data

Step count data were recorded in the Asmile app by automatic integration with standard smartphone health care apps. To provide additional detail, the daily step count for the last 42 days was transferred each time the user opened the Asmile app. Considering misplaced or forgotten smartphone or system inadequacies, inappropriate step count data of fewer than 200 steps and greater than 50,000 steps per day were excluded. Those cutoff values were inferred based on the step count distribution of Asmile users [27]. We used step count data with larger steps in a day, for which the data were recorded using both iOS and Android.

When preprocessing the step count data, we first excluded participants without step count data for at least 1 day per week, and those with a higher variance than all step count data. Then, we calculated the weekly average step count for each week from the first SHC until the second SHC. In all, 2778 participants were included in the analysis, with no missing values.

Latent Class Mixed Models

For this study, we used an LCMM [15] to identify long-term step trajectories of Asmile users. This approach, assuming that there are *G* latent classes in a heterogeneous population, can classify them into *G* trajectories. This method accommodates the identification of long-term changes in physical activity from longitudinal step count data.

Letting *N* represent the number of individuals i∈1,…,N, with step count observations Y_{ij} at occasion *j*, we aim to classify these individuals into latent classes g∈1,…,G, each representing characteristic patterns such as increasing, decreasing, flat, or non-monotonic trajectories over time. The step count model is

(1)Yij|g=β0g+β1gtij+β2gtij2+uig+ ij

In the above equation, t_{ij} stands for the elapsed weeks at occasion *j*, β represents the fixed effects of latent class *g*, and u_{ig} denotes the random effects for participant *i* of latent class *g* that is independent of time. In addition, measurement error ij follows a Gaussian distribution of mean 0 and variance σ 2. The LCMM approach estimates model parameters by maximizing the marginal log-likelihood [28] as

(2)l(θ)=∑i=1Nlog (∑g=1GπgLg(Yi|ti, θ))

In the equation, θ denotes the set of parameters to be estimated; πg represents the prior probability of latent class *g* as

(3)πg=eγg∑l=1Geγl

where γg is a class-specific intercept-like parameter. Because the step counts are typically approximately 5000, we assume a multivariate normal distribution for the likelihood function LgYiti,θ. The posterior probability that individual *i* belongs to class *g* is given as

(4)π^ig=πg Lg (yi ∣ ti,θ)∑l=1Gπl Ll (yi ∣ ti,θ)

XSL•FO
RenderX

In practice, the log-likelihood is then maximized iteratively using a modified Marquardt iterative algorithm and the Newton-Raphson method. Each individual is assigned to the latent class *g* for which the posterior probability π^ig is the highest.

Next, as one condition for determining the number of latent classes, we obtained π-g: the mean of the posterior probabilities of belonging to the latent class among the participants. We determined *G* (the number of latent classes) under the following three conditions. First, a participant *i* belongs to latent class *g* that has the largest π^ig. At this point, we ensured that the number of participants belonging to each latent class constituted at least 5% of the total number of participants analyzed. Second, the average posterior probability π−g in each latent class was set as approximately 90% or higher. Third, we selected the optimal *G* (number of latent classes) with the minimum Bayesian information criterion under conditions 1 and 2.

We used the lcmm package [29] with R software to identify step trajectories. Earlier reports [15,30] provide additional details related to the LCMM algorithm.

Statistical Analysis

Categorical variables were expressed as numbers and proportions, whereas continuous variables were expressed as the mean (SD). The means of step counts were calculated for three periods: from undergoing the first SHC to undergoing the second SHC, from undergoing the first SHC to 28 elapsed weeks, and from 28 elapsed weeks to undergoing the second SHC (maximum 56 elapsed weeks). Differences in these periods' associated variables were analyzed using *t* tests for men and women, analysis of variance between latent classes, and the chi-square test for categorical variables.

Long-term step counts of Asmile users were classified as step trajectories using LCMM. All Asmile users belonged to one of the *g* latent classes. We performed logistic regression using the *g* latent classes classified by LCMM as explanatory variables. The response variable was the presence or absence of weight loss of 3% or more at the second SHC, around one year after the start of step count recording, because achievement of 3% weight loss improves risk factors associated with obesity [10]. Adjustment variables included age, sex, mean step counts, smoking status, and the presence or absence of drug history. Here, drug history refers to whether a participant reported, at the time of the health checkup, taking medication for one or more of the following: diabetes, hypertension, or dyslipidemia. Specific drug names or classes, such as GLP-1 RAs or SGLT2 inhibitors, were not available. The odds ratios for each step trajectory were estimated. For sensitivity analysis, odds ratios were estimated for 1006 Asmile users who were not taking medication, to exclude the possibility of hospital attendance. The *P* values of the estimated partial regression coefficients are based on the Wald statistic.

Here, $P<.05$ was inferred as significant. All statistical analyses were conducted using R (version 4.3.1) with the lcmm package (version 2.1.0). This study is reported in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines.

Ethical Considerations

The study protocol was approved by the ethics committee of the Health and Counseling Center of The University of Osaka (institutional review board approval number 8 in 2024). All procedures involving human participants were conducted according to the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. At the time of app registration, the Asmile users consented to the use of nonidentifiable information in accordance with the terms of service related to the privacy policy. Informed consent to this study from the participants was waived because all data were anonymized according to the Japanese Ethical Guidelines for Medical and Health Research Involving Human Subjects enacted by Japan's Ministry of Health, Labour and Welfare. Anonymized data were provided by the Osaka Prefectural Government. No compensation was provided to participants.

Results

Characteristics

Table 1 presents baseline characteristics of the 2778 users. For weight and BMI, the first SHC was defined as preweight and pre-BMI. The second SHC was defined as postweight and post-BMI. Of the participants, 1601 (57.6%) were men and 1177 (42.4%) were women, with respective mean ages of 65.8 (SD 7.9) and 64 (SD 8.2) years. The mean step counts were significantly different between men and women, with men averaging 6461 (SD 3481) steps/day and women averaging 4953 (SD 2483) steps/day ($P<.001$). In addition, men and women had different weights ($P<.001$). Nevertheless, no difference was found for either BMI (pre-BMI: $P=.18$; post-BMI: $P=.84$) or weight difference after about 1 year ($P=.41$).

XSL•FO
RenderX

Table . Baseline characteristics.

Characteristics	Overall (n=2778)	Men (n=1601)	Women (n=1177)	*P* value
Age (years), mean (SD)	65.0 (8.1)	65.8 (7.9)	64.0 (8.2)	<.001
Elapsed weeks, mean (SD)	52.1 (3.4)	52.1 (3.4)	52.1 (3.4)	.76
Step counts, mean (SD)	5822.1 (3185.6)	6460.9 (3481.2)	4953.2 (2482.9)	<.001
Preweight[a], mean (SD)	72.1 (9.8)	76.9 (8.7)	65.6 (7.2)	<.001
Postweight[a], mean (SD)	71.3 (10.1)	76.1 (8.9)	64.7 (7.6)	<.001
Weight difference, mean (SD)	−0.8 (3.0)	−0.8 (2.9)	−0.9 (3.1)	.41
Weight loss rate, mean (SD)	−1.2 (4.0)	−1.0 (3.6)	−1.3 (4.5)	.04
Pre-BMI[a], mean (SD)	27.2 (2.3)	27.2 (2.3)	27.3 (2.4)	.18
Post-BMI[a], mean (SD)	26.9 (2.5)	26.9 (2.4)	26.9 (2.6)	.84
Smoking status, n (%)				<.001
No	2569 (92.5)	1427 (89.1)	1142 (97.0)	
Yes	209 (7.5)	174 (10.9)	35 (3.0)	
Medication status, n (%)				<.001
No	1006 (36.2)	529 (33.0)	477 (40.5)	
Yes	1772 (63.8)	1072 (67.0)	700 (59.5)	

[a]Preweight and pre-BMI were defined as measurements obtained at the first specific health checkup (SHC). Postweight and post-BMI as those obtained at the second SHC.

Classification of Step Count Trajectory

Figure 2 shows each trajectory of the 2778 participants and notable heterogeneity among participants. Models with 1 - 5 latent classes were built sequentially. Under conditions in which at least 5% of 2778 participants belonged to each latent class with at least a 90% probability, the 4-class model had the lowest Bayesian information criterion (Multimedia Appendices 1-3). That finding confirmed that assuming 4 classes provided the best explanatory power. Actually, the 4 classified step trajectories reflected walking styles over the long term. We labeled these trajectories as *UP*, *DOWN*, *UP/DOWN*, and *FLAT*.

Figure 2. Results of step trajectories. Horizontal axis: number of weeks elapsed from undergoing the first SHC to the second SHC. Vertical axis: average steps per day at each elapsed week. (A) All individual trajectories and the average trend using B-splines. Gray solid line: step trajectories for each of the 2778 individuals. Black dotted line: average step trajectory of 2778 participants. (B) Results of the latent 4 classes as step trajectories classified by the step count model using latent class mixed models. SHC: specific health checkup.

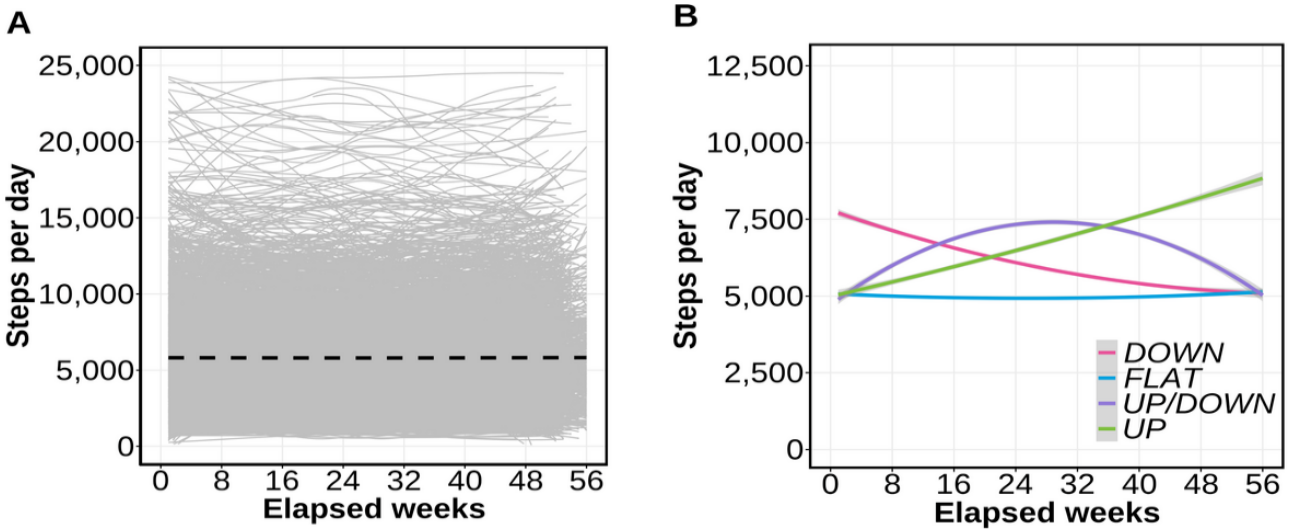


Table 2 presents characteristics of the 4 latent classes (*UP*, *DOWN*, *UP/DOWN*, and *FLAT*). Among the classes, age (*P*=.82) and the duration until undergoing the next checkup (*P*=.07) were found to have no significant difference. In addition, no significant difference was found for any class in pre-BMI (*P*=.39), but a reduction in BMI for the *UP* class was found in comparison to the other classes for post-BMI (*P*=.01). The average step count of the *FLAT* class was low compared to the

XSL•FO
RenderX

other classes (*P*<.001). More notably, irrespective of the long-term step count data, the average physical activities of the *UP*, *DOWN*, and *UP/DOWN* classes were almost identical when compared simply: 7592 (SD 3302) steps/day, 7223 (SD 3852) steps/day, and 7541 (SD 3282) steps/day.

Table . Characteristics of the 4 classes as step trajectories.

Characteristic	Overall (n=2778)	DOWN (n=355)	UP/DOWN (n=193)	FLAT (n=2045)	UP (n=185)	*P* value
Gender, n (%)						<.001
Men	1601 (57.6)	223 (62.8)	134 (69.4)	1125 (55.0)	119 (64.3)	
Women	1177 (42.4)	132 (37.2)	59 (30.6)	920 (45.0)	66 (35.7)	
Age (year), mean (SD)	65.0 (8.1)	64.6 (8.2)	65.1 (8.7)	65.1 (8.0)	65.1 (8.3)	.82
Elapsed weeks, mean (SD)	52.1 (3.4)	51.2 (3.4)	52.62 (3.3)	52.0 (3.4)	52.3 (3.5)	.07
Step counts, mean (SD)	5822.1 (3185.6)	7223.4 (3851.8)	7541.0 (3282.3)	5256.5 (2828.6)	7591.5 (3301.8)	<.001
Steps up to 28 weeks, mean (SD)	5826.3 (3227.1)	8023.8 (3951.5)	7347.1 (3304.2)	5243.2 (2828.2)	6469.0 (3309.3)	<.001
Steps after 29 weeks, mean (SD)	5813.5 (3248.1)	6278.0 (3801.5)	7763.0 (3345.3)	5270.2 (2863.5)	8894.1 (3381.3)	<.001
Preweight[a], mean (SD)	72.1 (9.8)	72.5 (10.0)	74.5 (10.0)	71.9 (9.8)	72.2 (9.0)	.004
Postweight[a], mean (SD)	71.3 (10.1)	72.0 (10.6)	73.6 (10.5)	71.1 (10.1)	70.1 (9.1)	.002
Weight difference, mean (SD)	−0.8 (3.0)	−0.5 (2.7)	−0.9 (3.5)	−0.8 (2.9)	−2.1 (3.4)	<.001
Weight loss rate, mean (SD)	−1.2 (4.0)	−0.7 (3.8)	−1.3 (4.6)	−1.1 (3.9)	−2.8 (4.7)	<.001
Pre-BMI[a], mean (SD)	27.2 (2.3)	27.1 (2.1)	27.4 (2.3)	27.2 (2.4)	27.1 (2.0)	.39
Post-BMI[a], mean (SD)	26.9 (2.5)	26.9 (2.4)	27.1 (2.6)	27.0 (2.5)	26.3 (2.3)	.010
Smoking status, n (%)						.29
No	2569 (92.5)	337 (94.9)	177 (91.7)	1883 (92.1)	172 (93.0)	
Yes	209 (7.5)	18 (5.1)	16 (8.3)	162 (7.9)	13 (7.0)	
Medication status, n (%)						.11
No	1006 (36.2)	145 (40.8)	78 (40.4)	719 (35.2)	64 (34.6)	
Yes	1772 (63.8)	210 (59.2)	115 (59.6)	1326 (64.8)	121 (65.4)	
Outcome, n (%)						<.001
Weight loss <3%	2123 (76.4)	280 (78.9)	146 (75.6)	1588 (77.7)	109 (58.9)	
Weight loss ≥3%	655 (23.6)	75 (21.1)	47 (24.4)	457 (22.3)	76 (41.1)	

[a]Preweight and pre-BMI were defined as the measurements obtained at the first specific health checkup (SHC). Postweight and post-BMI were defined as those obtained at the second SHC.

Figure 3 shows trajectories of the 4 latent classes (*UP*, *DOWN*, *UP/DOWN*, and *FLAT*) and the weight changes in the respective classes. *UP*, *DOWN*, *UP/DOWN*, and *FLAT* exhibited long-term step changes with the respective trends of increasing, decreasing, increasing and decreasing, and steady. It is noteworthy that the *UP* class included the most cases of more than 3% weight loss. The distribution of weight change was symmetrical in *FLAT*. In *DOWN*, compared to *FLAT*, the proportion with weight loss of around 3% was lower, whereas the proportion with weight gain of 5% or more was higher. The distribution in the *UP*/*DOWN* class was polarized. Among all classes, the *UP* class

XSL•FO
RenderX

showed the highest proportion of users with weight loss of 3% or more.

The *UP* class was characterized by a long-term increase in step counts: the mean of step counts for 1 - 28 elapsed weeks was 6469 (SD 3309) steps/day, but after 29 elapsed weeks, it was 8894 (SD 3381) steps/day. This class included 185 participants, 76 (41.1%) of whom were found to have greater than 3% weight loss. The absolute weight change was −2.1 (SD 3.4) kg. The relative weight change was −2.8% (SD 4.7%). These were the greatest negative values among all classes.

The *DOWN* class was characterized by a long-term decrease in step counts: the mean step count for 1 - 28 elapsed weeks was 8024 (SD 3952) steps/day; that after 29 elapsed weeks was 6287 (SD 3802) steps/day. This class included 355 participants, 75 (21.1%) of whom achieved greater than 3% weight loss. The preweight and postweight differences and weight loss were, respectively, −0.5 (SD 2.7) kg and −0.7% (SD 3.8%): the largest values among all classes.

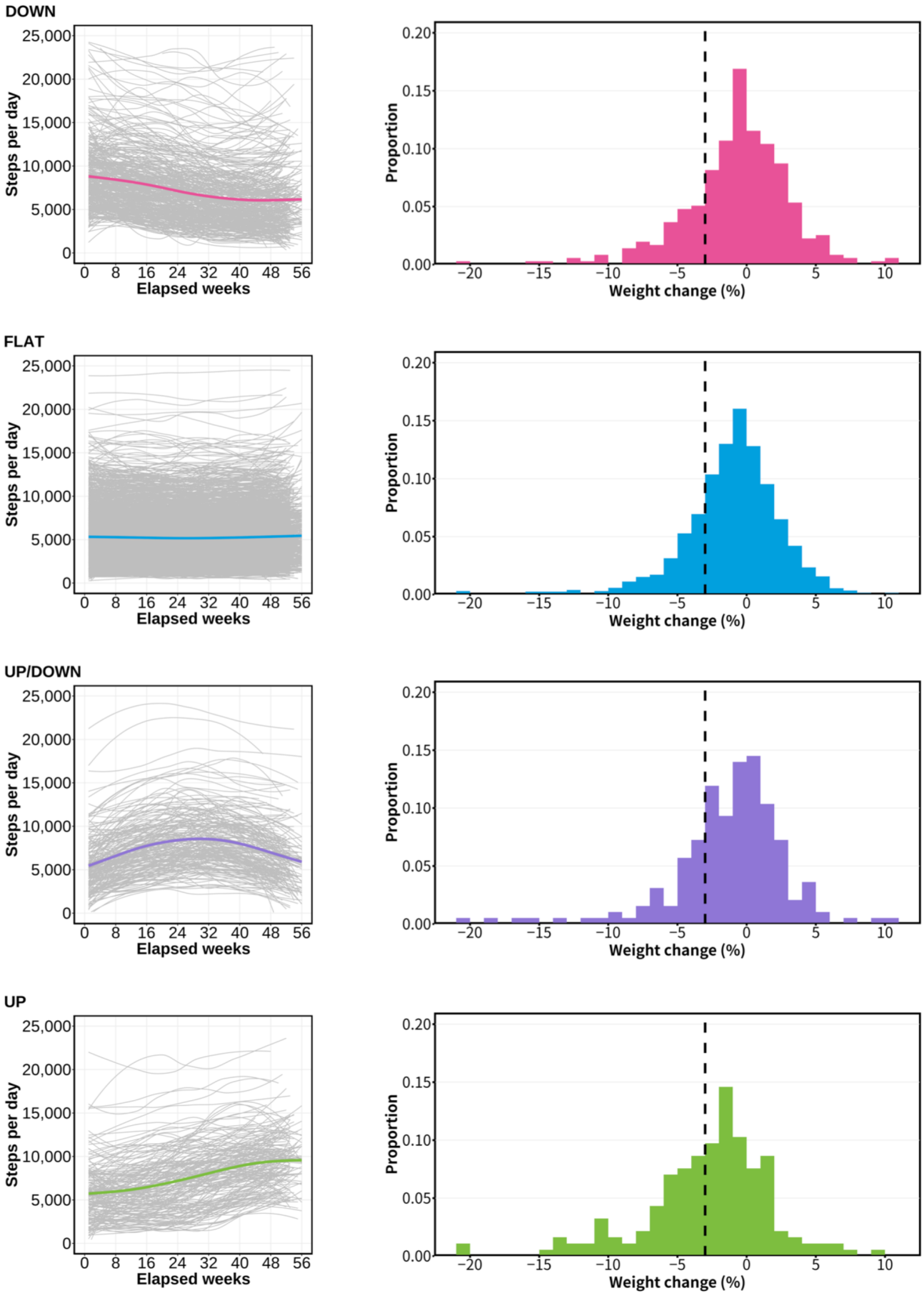
The *UP/DOWN* class was characterized by a long-term increase and decrease in step counts: the mean step count for 1 - 28 elapsed weeks was 7347 (SD 3304) steps/day; after 29 elapsed weeks, it was 7763 (SD 3345) steps/day. This class included 193 participants whose mean step counts of 7541 (SD 3282) steps/day were higher than the overall mean step counts of 5822 (SD 3186) steps/day and the *FLAT* group's mean step counts of 5257 (SD 2829) steps/day.

The *FLAT* class was characterized by long-term stay-in step counts: the mean step count for 1 - 28 elapsed weeks was 5243 (SD 2828) steps/day; that after 29 elapsed weeks was 5270 (SD 2864) steps/day. This class, which included 2045 participants, had mean step counts of 5257 (SD 2829) steps/day, which were closest to the overall mean step counts of 5822 (SD 3186) steps/day.

XSL•FO
RenderX

Figure 3. Comparison of step trajectory and weight changes among the 4 latent classes. (Left panels) Individual step trajectories and the average trend using B-splines in each latent class. Gray solid line: walking trends for individuals in each latent class. Each color line: average step trajectory in each latent class. (Right panels) Distribution by weight changes about 1 year in each latent class. Horizontal axis: weight change in increments of 1%, rounded below −20% and above 10%. Vertical axis: percentage of people in the latent class who have achieved a certain weight change after about 1 year. Black dotted line: 3% weight loss target cutoff used for this study. Latent classes from (top panels) to (bottom panels): *DOWN*, *FLAT*, *UP/DOWN*, and *UP*.

XSL•FO
RenderX

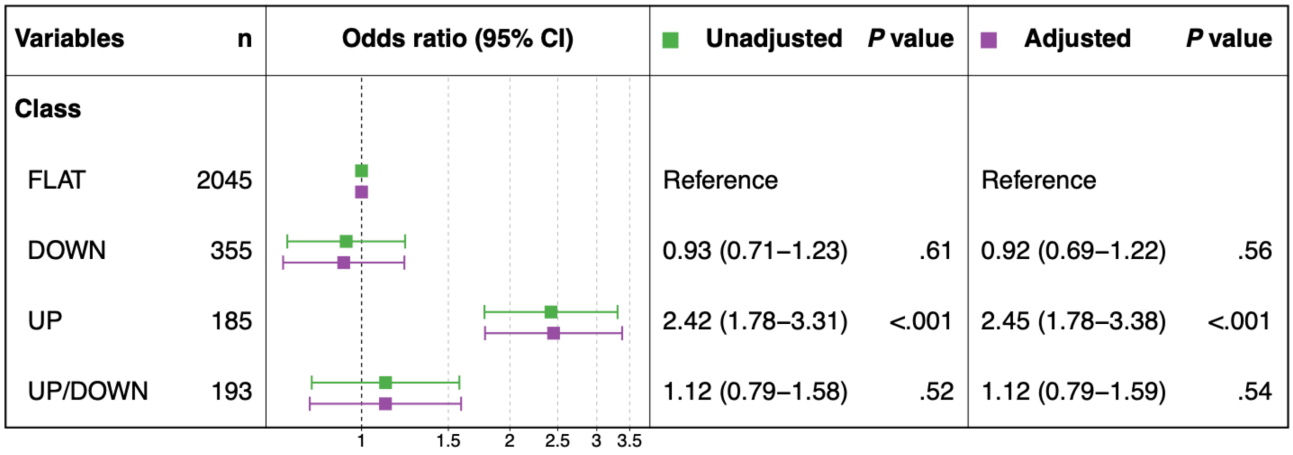


XSL•FO
RenderX

Step Count Trajectory Effects on Weight Loss

Figure 4 presents the estimated odds ratios of each trajectory for 3% weight loss. Performing multivariable logistic regression analysis with 4 latent classes (*UP*, *DOWN*, *UP/DOWN*, and *FLAT*), we assessed effects on weight loss from *UP*, *DOWN*, and *UP/DOWN* compared with *FLAT* (reference). The adjusted odds ratios of *UP*, *DOWN*, and *UP/DOWN* were, respectively, 2.45 (95% CI 1.78 - 3.38), 0.92 (95% CI 0.69 - 1.22), and 1.12 (95% CI 0.79 - 1.59). Among these, a significant association was found only for the *UP* class. By contrast, no significant difference was found for any *DOWN* or *UP/DOWN* class relative to the *FLAT* class.

Figure 4. Forest plot of odds ratios of the respective trajectories for weight loss.



Sensitivity Analysis

The effects of those step trajectories on weight loss were estimated as odds ratios for 1006 Asmile users who were taking no medication. These adjusted odds ratios of *UP*, *DOWN*, and *UP/DOWN* were, respectively, 2.42 (95% CI 1.41 - 4.13), 1.07 (95% CI 0.71 - 1.63), and 1.30 (95% CI 0.77 - 2.19). Sensitivity analysis estimated the effects on weight loss for those who did not take medications. Results demonstrated that the effects of clusters of step counts on weight loss remained (Multimedia Appendix 4).

Discussion

Principal Results

This trajectory analysis of 2778 Asmile members linked to SHC confirmed the usefulness of classification into 4 latent classes of *UP*, *DOWN*, *UP/DOWN*, and *FLAT*, which respectively represented increased, decreased, increased/decreased, and unchanged step counts for 10 - 14 months. Logistic regression analysis with these 4 latent classes suggested an association of a long-term increase in step counts with achieving greater than 3% weight loss, whereas a decrease in step counts was not associated significantly with weight loss about 1 year later.

Results indicated that subgroups, as latent classes, exist within the long-term step trajectory (Figure 2). We also characterized those subgroups (Table 2). These subgroups were classified as 4 heterogeneities of *UP*, *DOWN*, *UP/DOWN*, and *FLAT*, respectively, defined as described above (Figure 3).

For *UP*, the highest weight loss and the lowest post-BMI were observed. In earlier studies, a correlation between physical activity and BMI was proposed. Particularly, Krumm et al [11] and Thompson et al [12] reported that women with levels of activity such as approximately 10,000 steps/day typically have a BMI of less than 25 kg/m^2. Dwyer et al [31] reported that, for people with low daily energy expenditure, such as <10,000 steps/day, increasing daily step counts contributes to the reduction of obesity. These results are cross-sectional analyses, particularly addressing average steps, but they reveal a negative correlation between the average steps and BMI. In contrast, our results are longitudinal analyses, particularly addressing long-term step changes. In addition, the *UP* results support their association because the increased step counts show the greatest weight loss over the long term. It is noteworthy that the results indicate long-term increased physical activity as one factor affecting weight loss.

For *DOWN*, we observed the lowest weight loss, along with higher post-BMI than *UP*. Wyatt et al [13] reported that people with a BMI greater than or equal to 30 kg/m² take 2400 steps/day fewer than those with a BMI of less than 25 kg/m². They report a negative correlation between the average steps and BMI in cross-sectional analyses, particularly addressing average steps. Additionally, for some patients with advanced cancer undergoing chemotherapy, Manz et al [32] reported that a 1000-step/day decrease is associated with a 16% higher risk of hospitalization or death. Their results revealed risks attributable to a decrease in daily steps. The *DOWN* results can support their finding of negative correlation because intermittent decreased step counts over a long term show higher post-BMI than *UP*. In fact, for almost identical average step counts reported for *UP* and *DOWN*, findings indicate that decreased physical activity over the long term might achieve only insufficient weight loss. Therefore, to reduce risks associated with the decrease in step counts, a person should strive to increase and maintain step counts over a long period.

For *UP/DOWN*, we observed higher mean step counts than for *FLAT*, indicating greater weight loss than for *FLAT*. In addition, even though the mean step counts for *UP/DOWN* were closer to those of *UP* and *DOWN*, they showed less weight loss than that achieved by *UP* and greater weight loss than that achieved by *DOWN*. This finding, which was not obtained from a simple

XSL•FO
RenderX

comparison of mean step counts, suggests that increased intensity of physical activity over the long term is strongly reflected in weight loss.

For *FLAT*, which included the most Asmile users, baseline characteristics including age, sex, mean step counts, and BMI were similar to those of the overall population. Because *FLAT* reflects the distribution of the overall population in this study, we used *FLAT* as a reference when estimating weight loss effects.

To estimate the odds ratios of step trajectories for greater than 3% weight loss, we performed logistic regression analysis using the 4 latent classes: *UP*, *DOWN*, *UP/DOWN*, and *FLAT*. Actually, *UP* showed a higher adjusted odds ratio of 2.45 when using *FLAT* as a reference, including significantly more cases of weight loss greater than 3%. The results estimated a significant effect of increased step counts, reflecting long-term fluctuations of weight loss, because explanatory variables classified long-term fluctuations into several step trajectories as a latent class with LCMM. It is noteworthy that the average amounts of physical activity were similar for *UP*, *DOWN*, and *UP/DOWN*, but the *UP* group, which reflected increased and maintained step counts over the long term, was more likely to achieve weight loss than either the *DOWN* or *UP/DOWN* class.

Furthermore, most earlier studies using step counts have compared mean step counts. They have not considered long-term fluctuations in step counts, particularly capturing a decrease or a pattern of increase and decrease. From this study, *DOWN* and *UP/DOWN* were found to have no significant weight loss for *FLAT* as a reference, with respective odds ratios of 0.92 and 1.12, which are close to 1. As a notable finding, results suggest that a sustained walking style was associated more strongly with achievement of significant weight loss because the odds ratio for the overall mean step count was 1.01 per 1000-step increase, suggesting a possible, though nonsignificant, association between daily step count and weight loss. However, a higher odds ratio was observed for *UP*, which reflects a long-term walking style (Multimedia Appendix 5). These findings suggest that long-term increase and maintenance of physical activity such as walking, rather than temporary increases in physical activity, more strongly affect weight loss. In turn, this will engender a reduction in the risk of developing diseases caused by obesity.

Comparison With Earlier Work

Several earlier studies have evaluated mHealth app effects on physical activity, but most have had limited sample sizes and comparisons of simple step counts. In their meta-analysis, Flores Mateo et al [33] and Islam et al [34] reported that interventions using the mHealth app showed greater weight loss than the control. However, no significant difference in physical activity was found. Moreover, an important limitation is that these physical activities were recorded and assessed using a questionnaire and self-reporting. The data were not recorded automatically. In a randomized controlled trial of a mobile app intervention during a longer-term 32 weeks, Yoshimura et al [35] recorded and assessed physical activities using a step-count-specific app during weight loss. The intervention group was also instructed to use the app to check their daily step counts and ranks. Particularly, they revealed effective step increases on weekends. Nevertheless, no effect on weight loss was observed. They suggested the need to develop specialized tools to enhance weight loss effects. Their study represents an excellent analysis of the long-term effects of physical activity using a step count-specific app, but it falls short as an evaluation of weight loss when considering long-term step changes.

Painter et al [36] reported self-monitoring, such as weight, step counts, and food, as significant predictors of weight loss during a 6-month intervention. Their analysis showed that the weight loss ≥10%, 5% - 10%, and <5% groups had 6-month mean step counts, respectively, of 8078, 6657, and 5277 steps/day. A significant difference was also found among those means of step counts. Their study simply compared mean step counts and weight loss, confirming their mutual relation. By contrast, our observational study revealed a relation between long-term step-count changes and weight loss. Mean step counts were nearly equivalent among the *UP*, *DOWN*, and *UP/DOWN* classes, but significant weight loss was observed only for *UP*. These results clarify the relation between step count and weight loss from different perspectives. Therefore, considering that the *UP* class, the group with increased step counts, is more likely to achieve weight loss, the results reported by Painter et al [36] are supported: groups with greater weight loss had higher mean step counts. The findings reported herein further suggest that walking style is important for achieving long-term weight loss and suggest that increasing and maintaining physical activity levels is crucially important.

Benefits of Increasing and Maintaining Step Counts Long-Term

Oyama et al [27] reported a causal effect of using Asmile on increased step counts by approximately 400 steps/day and approximately 10,000 steps during 4 weeks in a large observational study of 80,689 Asmile users in Osaka prefecture.

The benefits of increased step counts for aiding weight loss and reducing CVD risk have been reported. In a randomized controlled trial including young adults, Rogers et al [37] analyzed step count effects on postprandial metabolism and showed that CVD risk was reduced by 10,000 steps/day, which significantly reduced postprandial lipemia, an independent predictor of CVD, compared with 2000 steps/day activity. Therefore, considering the *UP* effects of weight loss found from our study, continuous health promotion in mHealth, including Asmile, supports improved health status and prevention of lifestyle-related diseases.

Limitations

This study has several strengths, such as its large sample size linked to continuous daily activity (long-term step counts) and health status (medical checkup for each year), the classification of individuals' heterogeneous step count trajectories into latent classes, and the odds ratio estimation as identifying effects of long-term step count trajectories on weight loss.

However, the study also has some limitations. First, we were able to track the medication use of Asmile users (ie, whether they reported taking medications for diabetes, hypertension, or dyslipidemia), but their actual hospital attendance was unknown.

XSL•FO
RenderX

These are fundamentally important unobserved confounders because they might influence physical activity and weight change. To address potential confounding by medication use, we conducted a sensitivity analysis including only participants who were not taking any medication. The results were consistent with those of the main analysis.

Second, the Asmile users' occupational characteristics are not considered. Changes in physical activities might occur because of occupational changes; for example, white-collar workers spend more time sitting than blue-collar workers during daily life [38]. Whereas step counts provided detailed information related to daily movement, participants with 1 or more weeks in which step count data were missing completely were excluded from the analysis. In the Asmile app, daily step counts for the last 42 days are transferred each time the user opens the app. Therefore, missing data generally occur because of participants' app usage behavior rather than because of random loss. For that reason, excluding these participants might introduce selection bias. Nevertheless, we regarded this approach as appropriate for a more conservative analysis because it allowed us to examine high-quality longitudinal step count data specifically and address an important knowledge gap related to the effects of physical activity over time on weight change.

Third, Asmile users are likely to be more health conscious than people who are not using the mHealth app, including Asmile. Therefore, caution must be exercised when roughly generalizing and interpreting the results of this analysis. Moreover, because the study population consisted mainly of active individuals with a BMI ≥25 kg/m² who participated in the app-based program voluntarily, the findings might not be generalizable to individuals with obesity or less-active populations.

In addition, dietary habits were not available in this dataset, although they are important determinants of weight loss and maintenance [39,40]. Many other factors, such as metabolic-related variables, general health conditions, and genetic predisposition, can also influence weight changes [41-43]. Comorbidities common in older adults, such as cancer, dementia, or other chronic conditions, were not available in the dataset, but they might also influence physical activity levels and weight change [44,45].

Finally, step count data capture the amount of movement but might not fully reflect the intensity or type of physical activity (eg, step-based measures might underestimate energy expenditure under certain stride frequencies or fail to encompass nonwalking activities such as cycling, running, or standing) [46-49]. Therefore, physical activity effects on weight might be underestimated in this study.

Maximizing ecological validity by linking and considering additional user characteristics, including dietary, metabolic, genetic, comorbidity, and physical activity intensity data, will be necessary for future studies.

Conclusions

This study revealed effects on weight loss of the trajectories of long-term step counts, as monitored using the "Asmile" mHealth app. Particularly, the long-term trajectory of increased step counts significantly affected 3% weight loss. These findings suggest that sustained engagement in physical activity might play an important role in long-term health management.

Acknowledgments

The authors are grateful to R Sasaki, Y Yamazaki, and all members of the National Health Insurance Division, Department of Public Health and Medical Affairs, Osaka Prefectural Government, for providing Asmile data. Generative artificial intelligence tools, including ChatGPT (OpenAI), were used to assist in improving the language and clarity of the manuscript under the supervision of the authors. All intellectual content and interpretations are the authors' own.

Funding

This work was partly supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI grants (24K10918, 23K17228, and 22H05108) and by the Japan Science and Technology Agency (JST) ERATO grant (JPMJER2102).

Data Availability

Data cannot be shared publicly because local governments own medical checkup data. Data are available from the Health and Counseling Center, The University of Osaka (contact via campuslifekenkou-syomu@office.osaka-u.ac.jp) for researchers who meet the criteria for access to confidential data.

Authors' Contributions

Conceptualization: KT, AO, JK
Data curation: KT, AO
Discussion: HT, RY
Formal analysis: KT, AO
Methodology: KT, AO
Project administration: JK
Software: KT, AO
Supervision: JK

XSL•FO
RenderX

Validation: KT, AO
Visualization: KT, AO
Writing – original draft: KT, AO
Writing – review & editing: KT, AO, JK, HT, RY

Conflicts of Interest

KT is an employee of Eiken Chemical Co, Ltd, which manufactures diagnostic reagents. The company had no role in the study design, data collection, data analysis, decision to publish, or preparation of the manuscript. The authors have no other conflict of interest to declare.

Multimedia Appendix 1
Percentages of participants belonging to latent classes in each latent class mixed model.
[DOCX File, 31 KB - mhealth_v14i1e80339_app1.docx]

Multimedia Appendix 2
Means of posterior probabilities belonging to a latent class in each latent class mixed model.
[DOCX File, 31 KB - mhealth_v14i1e80339_app2.docx]

Multimedia Appendix 3
Bayesian information criterion at each latent class mixed model.
[DOCX File, 31 KB - mhealth_v14i1e80339_app3.docx]

Multimedia Appendix 4
Forest plot of odds ratios of each trajectory for weight loss in sensitivity analysis.
[DOCX File, 280 KB - mhealth_v14i1e80339_app4.docx]

Multimedia Appendix 5
List of odds ratios including adjustment variables.
[DOCX File, 31 KB - mhealth_v14i1e80339_app5.docx]

Checklist 1
STROBE checklist.
[DOCX File, 27 KB - mhealth_v14i1e80339_app6.docx]

References

1. McLaren L. Socioeconomic status and obesity. Epidemiol Rev 2007 Oct;29(1):29-48. [doi: 10.1093/epirev/mxm001] [Medline: 17478442]
2. Malik VS, Willett WC, Hu FB. Global obesity: trends, risk factors and policy implications. Nat Rev Endocrinol 2013 Jan;9(1):13-27. [doi: 10.1038/nrendo.2012.199]
3. Kopelman PG. Obesity as a medical problem. Nature New Biol 2000 Apr;404(6778):635-643. [doi: 10.1038/35007508]
4. Obesity and overweight. World Health Organization. 2025 Dec 8. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2026-04-14]
5. Dorresteijn JAN, Visseren FLJ, Spiering W. Mechanisms linking obesity to hypertension. Obes Rev 2012 Jan;13(1):17-26. [doi: 10.1111/j.1467-789X.2011.00914.x] [Medline: 21831233]
6. Vekic J, Zeljkovic A, Stefanovic A, Jelic-Ivanovic Z, Spasojevic-Kalimanovska V. Obesity and dyslipidemia. Metab Clin Exp 2019 Mar;92:71-81. [doi: 10.1016/j.metabol.2018.11.005]
7. Kahn SE, Hull RL, Utzschneider KM. Mechanisms linking obesity to insulin resistance and type 2 diabetes. Nature New Biol 2006 Dec;444(7121):840-846. [doi: 10.1038/nature05482]
8. Powell-Wiley TM, Poirier P, Burke LE, et al. Obesity and cardiovascular disease: a scientific statement from the American Heart Association. Circulation 2021 May 25;143(21). [doi: 10.1161/CIR.0000000000000973]
9. The Examination Committee of Criteria for 'Obesity Disease' in Japan, Japan Society for the Study of Obesity. New criteria for `obesity disease' in Japan. Circ J 2002 Nov;66(11):987-992. [doi: 10.1253/circj.66.987]
10. Muramoto A, Matsushita M, Kato A, et al. Three percent weight reduction is the minimum requirement to improve health hazards in obese and overweight people in Japan. Obes Res Clin Pract 2014;8(5):e466-e475. [doi: 10.1016/j.orcp.2013.10.003] [Medline: 25263836]
11. Krumm EM, Dessieux OL, Andrews P, Thompson DL. The relationship between daily steps and body composition in postmenopausal women. J Womens Health (Larchmt) 2006 Mar;15(2):202-210. [doi: 10.1089/jwh.2006.15.202]

XSL•FO
RenderX

12. Thompson DL, Rakow J, Perdue SM. Relationship between accumulated walking and body composition in middle-aged women. Med Sci Sports Exerc 2004 May;36(5):911-914. [doi: 10.1249/01.mss.0000126787.14165.b3] [Medline: 15126729]
13. Wyatt HR, Peters JC, Reed GW, Barry M, Hill JO. A Colorado statewide survey of walking and its relation to excessive weight. Med Sci Sports Exerc 2005 May;37(5):724-730. [doi: 10.1249/01.mss.0000161750.84096.d4] [Medline: 15870624]
14. Ferrari G, Marques A, Barreira T, et al. Accelerometer-measured daily step counts and adiposity indicators among Latin American adults: a multi-country study. IJERPH ;18(9):4641. [doi: 10.3390/ijerph18094641]
15. Proust-Lima C, Philipps V, Liquet B. Estimation of extended mixed models using latent classes and latent processes: the R Package lcmm. J Stat Softw 2017;78(2):1-56. [doi: 10.18637/jss.v078.i02]
16. Ledoult E, Launay D, Béhal H, et al. Early trajectories of skin thickening are associated with severity and mortality in systemic sclerosis. Arthritis Res Ther 2020 Feb 18;22(1):30. [doi: 10.1186/s13075-020-2113-6] [Medline: 32070422]
17. Reynolds JA, Prattley J, Geifman N, et al. Distinct patterns of disease activity over time in patients with active SLE revealed using latent class trajectory models. Arthritis Res Ther 2021 Jul 29;23(1):203. [doi: 10.1186/s13075-021-02584-x] [Medline: 34321096]
18. Lennon H, Kelly S, Sperrin M, et al. Framework to construct and interpret latent class trajectory modelling. BMJ Open 2018 Jul 7;8(7):e020683. [doi: 10.1136/bmjopen-2017-020683] [Medline: 29982203]
19. Asmile. Osaka Prefecture's Health Application. 2019. URL: https://www.asmile.pref.osaka.jp/ [accessed 2026-4-14]
20. Matsuoka Y, Yoshida H, Hanazato M. A smartphone-based shopping mall walking program and daily walking steps. JAMA Netw Open 2024 Jan 2;7(1):e2353957. [doi: 10.1001/jamanetworkopen.2023.53957] [Medline: 38289599]
21. Tanji F, Tomata Y, Abe S, et al. Effect of a financial incentive (shopping point) on increasing the number of daily walking steps among community-dwelling adults in Japan: a randomised controlled trial. BMJ Open 2020 Nov;10(11):e037303. [doi: 10.1136/bmjopen-2020-037303]
22. Tsushita K, Hosler AS, Miura K, et al. Rationale and descriptive analysis of specific health guidance: the nationwide lifestyle intervention program targeting metabolic syndrome in Japan. J Atheroscler Thromb 2018;25(4):308-322. [doi: 10.5551/jat.42010]
23. Sonoda N, Koh C, Yasumoto R, et al. Association between recommendations from public health nurses, medical professionals, and family members and participation in health checkups among middle-aged community residents with National Health Insurance. JMA J 2022;5(2):199-206. [doi: 10.31662/jmaj.2021-0225]
24. Takeuchi M, Shinozaki T, Kawakami K. Effectiveness of specific health check-ups in Japan for the primary prevention of obesity-related diseases: a protocol for a target trial emulation. BMJ Open 2023 Jul 30;13(7):e070417. [doi: 10.1136/bmjopen-2022-070417] [Medline: 37518087]
25. Koyama S, Tabuchi T, Okawa S, et al. Changes in smoking behavior since the declaration of the COVID-19 state of emergency in Japan: a cross-sectional study from the Osaka Health app. J Epidemiol 2021;31(6):378-386. [doi: 10.2188/jea.JE20200533]
26. Yokoyama H, Kitano Y. Oral frailty as a risk factor for fall incidents among community-dwelling people. Geriatrics (Basel) 2024 Apr 22;9(2):54. [doi: 10.3390/geriatrics9020054] [Medline: 38667521]
27. Oyama A, Taguchi K, Seto H, et al. Effects of mobile health care app "Asmile" on physical activity of 80,689 users in Osaka Prefecture, Japan: longitudinal observational study. J Med Internet Res 2025 May 21;27:e65943. [doi: 10.2196/65943]
28. Proust C, Jacqmin-Gadda H. Estimation of linear mixed models with a mixture of distribution for the random effects. Comput Methods Programs Biomed 2005 May;78(2):165-173. [doi: 10.1016/j.cmpb.2004.12.004]
29. Extended mixed models using latent classes and latent processes. GitHub. URL: https://cecileproust-lima.github.io/lcmm/ [accessed 2026-04-14]
30. Proust C, Jacqmin-Gadda H, Taylor JMG, Ganiayre J, Commenges D. A nonlinear model with latent process for cognitive evolution using multivariate longitudinal data. Biometrics 2006 Dec;62(4):1014-1024. [doi: 10.1111/j.1541-0420.2006.00573.x] [Medline: 17156275]
31. Dwyer T, Hosmer D, Hosmer T, et al. The inverse relationship between number of steps per day and obesity in a population-based sample – the AusDiab study. Int J Obes 2007 May;31(5):797-804. [doi: 10.1038/sj.ijo.0803472]
32. Manz CR, Schriver E, Ferrell WJ, et al. Association of remote patient-reported outcomes and step counts with hospitalization or death among patients with advanced cancer undergoing chemotherapy: secondary analysis of the PROStep randomized trial. J Med Internet Res 2024 May 17;26(1):e51059. [doi: 10.2196/51059]
33. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 2015 Nov 10;17(11):e253. [doi: 10.2196/jmir.4836]
34. Islam MM, Poly TN, Walther BA, (Jack) Li YC. Use of mobile phone app interventions to promote weight loss: meta-analysis. JMIR Mhealth Uhealth 2020 Jul 22;8(7):e17039. [doi: 10.2196/17039]
35. Yoshimura E, Tajiri E, Michiwaki R, Matsumoto N, Hatamoto Y, Tanaka S. Long-term effects of the use of a step count–specific smartphone app on physical activity and weight loss: randomized controlled clinical trial. JMIR Mhealth Uhealth 2022 Oct 24;10(10):e35628. [doi: 10.2196/35628]
36. Painter SL, Ahmed R, Hill JO, et al. What matters in weight loss? An in-depth analysis of self-monitoring. J Med Internet Res 2017 May 12;19(5):e160. [doi: 10.2196/jmir.7457]

XSL•FO
RenderX

37. Rogers EM, Banks NF, Jenkins NDM. Acute effects of daily step count on postprandial metabolism and resting fat oxidation: a randomized controlled trial. J Appl Physiol 2023 Oct 1;135(4):812-822. [doi: 10.1152/japplphysiol.00052.2023]
38. Fukushima N, Kikuchi H, Amagasa S, et al. Exposure to prolonged sedentary behavior on weekdays rather than weekends in white-collar workers in comparison with blue-collar workers. J Occup Health 2021 Jan;63(1):e12246. [doi: 10.1002/1348-9585.12246] [Medline: 34275174]
39. Wright JD, Wang CY. Trends in intake of energy and macronutrients in adults from 1999-2000 through 2007-2008. NCHS Data Brief 2010 Nov(49):1-8. [Medline: 21211164]
40. Estruch R, Ros E, Salas-Salvadó J, et al. Primary prevention of cardiovascular disease with a Mediterranean diet supplemented with extra-virgin olive oil or nuts. N Engl J Med 2018 Jun 21;378(25):1279-1290. [doi: 10.1056/NEJMoa1800389] [Medline: 29897866]
41. Brittain EL, Han L, Annis J, et al. Physical activity and incident obesity across the spectrum of genetic risk for obesity. JAMA Netw Open 2024 Mar 4;7(3):e243821. [doi: 10.1001/jamanetworkopen.2024.3821] [Medline: 38536175]
42. Mäkinen VP, Ala-Korpela M. Influence of age and sex on longitudinal metabolic profiles and body weight trajectories in the UK Biobank. Int J Epidemiol 2024 Apr 11;53(3). [doi: 10.1093/ije/dyae055]
43. Menni C, Migaud M, Kastenmüller G, et al. Metabolomic profiling of long-term weight change: role of oxidative stress and urate levels in weight gain. Obesity (Silver Spring) 2017 Sep;25(9):1618-1624. [doi: 10.1002/oby.21922] [Medline: 28758372]
44. Srivastava S, Joseph K J V, Dristhi D, Muhammad T. Interaction of physical activity on the association of obesity-related measures with multimorbidity among older adults: a population-based cross-sectional study in India. BMJ Open 2021 May;11(5):e050245. [doi: 10.1136/bmjopen-2021-050245]
45. Sun F, Norman IJ, While AE. Physical activity in older people: a systematic review. BMC Public Health 2013 Dec;13(1). [doi: 10.1186/1471-2458-13-449]
46. Nielson R, Vehrs PR, Fellingham GW, Hager R, Prusak KA. Step counts and energy expenditure as estimated by pedometry during treadmill walking at different stride frequencies. J Phys Act Health 2011 Sep;8(7):1004-1013. [doi: 10.1123/jpah.8.7.1004] [Medline: 21885892]
47. O'Brien MW, Kivell MJ, Wojcik WR, d'Entremont G, Kimmerly DS, Fowles JR. Step rate thresholds associated with moderate and vigorous physical activity in adults. Int J Environ Res Public Health 2018 Nov 3;15(11):2454. [doi: 10.3390/ijerph15112454] [Medline: 30400331]
48. Johansson MS, Korshøj M, Schnohr P, et al. Time spent cycling, walking, running, standing and sedentary: a cross-sectional analysis of accelerometer-data from 1670 adults in the Copenhagen City Heart Study. BMC Public Health 2019 Dec;19(1). [doi: 10.1186/s12889-019-7679-z]
49. Chevance G, Golaszewski NM, Tipton E, et al. Accuracy and precision of energy expenditure, heart rate, and steps measured by combined-sensing Fitbits against reference measures: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022 ;10(4):e35626. [doi: 10.2196/35626]

Abbreviations

CVD: cardiovascular disease
LCMM: latent class mixed model
mHealth: mobile health
SHC: specific health checkup
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by L Buis; submitted 09.Jul.2025; peer-reviewed by J Dai, Y Oyende; revised version received 21.Nov.2025; accepted 26.Mar.2026; published 04.May.2026.

Please cite as:
Taguchi K, Oyama A, Kotoku J, Toki H, Yamamoto R
One-Year Trajectory of Step Counts and Weight Loss in Adults With Overweight/Obesity: Retrospective Cohort Study
JMIR Mhealth Uhealth 2026;14:e80339
URL: https://mhealth.jmir.org/2026/1/e80339
doi:10.2196/80339

© Kenshiro Taguchi, Asuka Oyama, Jun'ichi Kotoku, Hiroshi Toki, Ryohei Yamamoto. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 4.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The

XSL•FO
RenderX

complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Traditional Social Sports Games and Mental Training for Smartphone Addiction and Psychological Distress in School-Aged Adolescents: Randomized Controlled Trial

Mohamed Yaakoubi[1,2*], PhD; Ahmed Ghorbel[1,2*], PhD; Hiba Abdelkafi[3*], PhD; Liwa Masmoudi[1,2], PhD; Omar Trabelsi[1,2,3*], PhD; Safaa M Elkholi[4*], PhD; Fatma H Yagin[5*], PhD; Georgian Badicu[6*], PhD; Adnene Gharbi[2,7*], PhD

[1]High Institute of Sports and Physical Education of Kef, University of Jendouba, Kef, Tunisia

[2]Research Unit: Physical Activity, Sport and Health, UR18JS01, National Observatory of Sport, Tunis, Tunisia

[3]The High Institute of Applied Studies in Humanities of Mahdia, Department of Education Sciences, University of Monastir, Mahdia, Tunis, Tunisia

[4]Department of Rehabilitation Sciences, College of Health and Rehabilitation Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Riyadh, Saudi Arabia

[5]Department of Biostatistics, Faculty of Medicine, Malatya Turgut Ozal University, 44210, Malatya, Turkey

[6]Department of Physical Education and Special Motricity, University Transilvania of Braşov, Braşov, Romania

[7]High Institute of Sports and Physical Education of Sfax, University of Sfax, Sfax, Tunisia

[*]these authors contributed equally

Corresponding Author:
Safaa M Elkholi, PhD
Department of Rehabilitation Sciences, College of Health and Rehabilitation Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Riyadh, Saudi Arabia

Abstract

Background: Problematic smartphone use among adolescents is a growing public health concern closely associated with psychological distress and loneliness. Effective, culturally grounded, school-based interventions are needed.

Objective: The aim of this study was to assess the effects of a 12-week program combining traditional social sports games and mental exercises on smartphone addiction, nomophobia, psychological distress, and loneliness in adolescents.

Methods: In this randomized controlled trial, 69 school-recruited Tunisian adolescents (aged 14-16 years) with clinically elevated smartphone addiction scores were assigned to an experimental group (n=36, 52.2%) or a control group (n=33, 47.8%). The experimental group received a 12-week intervention comprising 4 weekly sessions integrating traditional social sports games with mental exercises, whereas the control group continued standard physical education. Outcomes (smartphone addiction, nomophobia, psychological distress, and loneliness) were assessed at baseline and after the intervention using scales validated in Arabic.

Results: Linear mixed-effects models adjusted for age, sex, and BMI revealed significant group × time interactions of moderate magnitude across all outcomes (P<.05 in all cases) favoring the experimental group. Adjusted postintervention comparisons confirmed significantly lower scores in the experimental group for smartphone addiction, nomophobia, psychological distress, and loneliness (P<.05 in all cases; partial η_p^2=0.08 - 0.12). Mediation analysis indicated that reductions in loneliness accounted for 34.4% of the intervention's effect on smartphone addiction, consistent with partial mediation.

Conclusions: A culturally adapted, school-based intervention combining traditional social sports games and mental exercises significantly reduced problematic smartphone use and improved psychological well-being. The partial mediation through reduced loneliness highlights the critical role of social connectedness in adolescent digital health interventions.

Trial Registration: Pan African Clinical Trials Registry PACTR202601838702413; https://pactr.samrc.ac.za/TrialDisplay.aspx?TrialID=39795

(JMIR Mhealth Uhealth 2026;14:e85353) doi:10.2196/85353

KEYWORDS

smartphone addiction; adolescents; loneliness; psychological distress; randomized controlled trial; Tunisia

XSL•FO
RenderX

Introduction

Adolescence constitutes a critical period of neurobiological, emotional, and social development marked by heightened sensitivity to reward processing, peer feedback, and affective stimuli. Ongoing maturation of prefrontal-limbic circuits during this stage increases impulsivity and emotional reactivity, rendering adolescents particularly vulnerable to maladaptive behavioral patterns in technology-saturated environments [1].

In Tunisia, these developmental vulnerabilities intersect with rapid digital expansion that has outpaced public health policy and digital literacy initiatives. As a result, smartphones have become central tools for managing academic demands, emotional regulation, and social relationships among adolescents [2-4]. While digital access offers clear benefits, excessive reliance on digital devices has contributed to the growing prevalence of problematic smartphone use (PSU).

PSU is characterized by compulsive, excessive smartphone engagement that results in functional impairment across academic, emotional, and social domains [5]. Among its core manifestations, nomophobia, the fear of being without one's smartphone, has emerged as a prominent indicator of psychological dependence and psychological distress [6,7]. Empirical evidence consistently links PSU and nomophobia to emotional dysregulation; reduced academic performance; and elevated symptoms of anxiety, depression, and stress [8].

Loneliness is a key psychosocial mechanism underlying PSU. Adolescents often engage in smartphone use to compensate for perceived social disconnection, yet digitally mediated interactions frequently fail to satisfy deeper relational needs. This mismatch can create a self-reinforcing cycle where increased smartphone use exacerbates loneliness, further driving compensatory digital engagement [9,10]. Therefore, interventions that disrupt this cycle by directly enhancing social connectedness are critical. Meta-analytic evidence indicates that physical activity (PA) interventions, particularly those that are group based and sustained (≥12 weeks), effectively reduce smartphone addiction [11], with loneliness established as a mediator of this effect [12,13]. This process aligns with self-determination theory (SDT) as socially engaging PA can satisfy core psychological needs such as relatedness, thereby reducing reliance on smartphones for compensatory fulfillment [14].

SDT further explicates this mechanism, positing that psychological well-being depends on the satisfaction of 3 basic psychological needs: autonomy, competence, and relatedness [15]. From this perspective, PSU may represent a maladaptive coping strategy that emerges when the need for relatedness remains unmet. Thus, loneliness functions both as an indicator of need frustration and as a mechanistic pathway sustaining compulsive smartphone use.

Building on this rationale, PA interventions, particularly structured, group-based programs, have demonstrated efficacy in reducing PSU and improving psychological well-being among adolescents [16,17]. There is further evidence suggesting that PA-based interventions are more effective when combined with psychological strategies such as mental skill training, which strengthens emotional regulation, stress tolerance, and self-regulation capacities [11].

Cultural adaptation represents a critical determinant of engagement and sustainability. In the Tunisian context, traditional social sports games (eg, seated ball game and dodgeball) constitute culturally resonant, rule-based PAs that promote face-to-face interaction, cooperation, and prosocial behavior. These activities foster peer bonding and shared enjoyment, directly targeting the SDT need for relatedness while counteracting the isolating effects of excessive screen use [18,19].

Despite this strong theoretical and empirical rationale, important gaps remain. Few controlled studies have evaluated multi-component, culturally grounded interventions that integrate structured PA with psychological skill development using a randomized controlled design in a school setting. Moreover, no study has empirically tested whether reductions in loneliness mediate the effects of such interventions on PSU despite the central role of relatedness within SDT.

To address these gaps, we developed and evaluated a 12-week school-based intervention integrating traditional social sports games with mental exercises (TSSG-ME) among Tunisian adolescents. We hypothesized that participants in the TSSG-ME group would demonstrate significantly greater reductions in smartphone addiction and psychological distress than a control group (CG). Grounded in SDT, we further hypothesized that reductions in loneliness would mediate the intervention's effect on smartphone addiction, reflecting enhanced satisfaction of the need for relatedness.

Methods

Study Design and Setting

This study was a 2-arm, parallel-group randomized controlled trial. The trial was retrospectively registered with the Pan African Clinical Trials Registry under the identification number PACTR202601838702413 on January 30, 2026. The registration was conducted retrospectively due to an administrative oversight during the initial recruitment phase as the research team prioritized data collection commencement following funding approval. This study was conducted between October 2024 and February 2025 across 3 public middle schools in the Sfax region of Tunisia during routine health screenings.

Participants

A total of 960 adolescents (aged 14-16 years) were initially recruited during routine health screenings across 3 public middle schools in the Sfax region. A total of 151 students with PSU were identified. The inclusion criteria were (1) meeting the clinical cutoff for smartphone addiction and (2) reporting structured PA not exceeding 3000 metabolic equivalent of task minutes per week as assessed using the Arabic-language International Physical Activity Questionnaire [20]. The exclusion criteria were (1) BMI below the 3rd or above the 97th percentile [21] and (2) habitual sleep duration outside the range of 6 to 10 hours per night [22]. This resulted in the exclusion

XSL•FO
RenderX

of 76 students (n=37, 48.7% due to excessive activity; n=39, 51.3% due to the BMI and sleep criteria).

Eligible participants (n=75) were randomly assigned to the experimental group (EG; n=36) or the CG (n=33) using stratified block randomization. Randomization was stratified by sex (male and female), with a combination of random block sizes of 4, 6, and 8 to ensure balanced allocation across groups throughout the recruitment period. The allocation sequence was obtained using computer-generated random numbers by an independent researcher who was not involved in participant recruitment, intervention delivery, or outcome assessment.

Allocation concealment was ensured using sequentially numbered opaque, sealed envelopes prepared by the independent researcher and opened only after completion of baseline assessments and confirmation of eligibility. Participants were enrolled by the lead author, and group assignment was implemented by a research assistant not involved in outcome assessment.

Due to the nature of the behavioral intervention, participants and intervention providers could not be blinded. However, outcome assessors and data analysts were blinded to group allocation throughout the trial.

During the trial, 8% (6/75) of the participants were excluded due to extended absences (>1 week), incomplete assessments, or newly emerging clinical conditions (EG: n=4; CG: n=2), resulting in a final analytic sample of 69 participants (EG: n=36, 52.2%; CG: n=33, 47.8%). The Little test was nonsignificant (χ^2_{18}=12.5; *P*=.82), supporting the assumption that data were missing completely at random. Given the low rate of missingness and the missing completely at random assumption, we conducted complete-case analyses without multiple imputation. Baseline comparisons confirmed equivalence between groups on all demographic and outcome variables (*P*>.05 in all cases). A detailed flowchart of participant screening, randomization, allocation, and follow-up is presented in Figure 1.

Intervention

Overview

Grounded in SDT, the intervention was designed to target basic psychological needs. The traditional social sports game component primarily addressed the need for relatedness through cooperative and socially structured play, whereas the mental exercise component targeted competence and autonomy through structured skill building and self-regulation practices. The intervention began within 1 week following randomization for both groups.

Experimental Group

The EG participated in a structured 12-week intervention consisting of four 60-minute sessions per week. Three sessions were conducted during regular school hours, and 1 extracurricular session was held on Friday afternoons [23]. Each session was divided into 2 core modules: traditional social sports games (45 minutes) and mental exercises (15 minutes). The sports module used 4 culturally embedded Tunisian games (seated ball game, dodgeball, sbeïla [a traditional chasing game], and korra teyara [a traditional target throwing game]) implemented with an emphasis on teamwork, communication, and strategic decision-making [18]. Game complexity increased over time: weeks 1 to 3 focused on rule acquisition and teamwork, whereas weeks 4 to 12 introduced dynamic variations and collaborative problem-solving strategies [24,25].

The mental exercise module followed a standardized 5-phase sequence to facilitate cognitive and physical recovery [26,27]. This sequence began with a 2-minute prerelaxation phase for session orientation followed by a 3-minute induction to calmness using diaphragmatic breathing and guided imagery. The main relaxation phase (5 minutes) incorporated Jacobson's progressive muscle relaxation, autogenic training, and mindfulness-based techniques. The session concluded with a 2-minute recovery period of light stretching and a 3-minute reflective closure involving verbal discussion or a short written reflection.

Control Group

The CG followed the standard, time-matched physical education curriculum mandated by the Tunisian Ministry of Education, consisting of 2 weekly sessions (one 2-hour and one 1-hour session) [23]. Sessions included conventional individual activities and team sports. The CG did not receive the novel social games or mental exercises, serving as a treatment-as-usual control. This design allowed for a pragmatic comparison with standard school practices but did not control for nonspecific intervention effects such as increased attention, novelty, or social interaction.

Instruction and Fidelity

Sessions were delivered by a team of 3 instructors, each with over 10 years of experience working with youth. The team comprised 2 physical education specialists with master's degrees in adapted physical activity and sport pedagogy and 1 certified mental preparation coach with a professional certificate in mental skill training for adolescents. To ensure treatment fidelity, all instructors received comprehensive training on the research protocol and adhered to a standardized intervention manual. For each session, participants were organized into 3 subgroups, enhancing peer interaction and enabling close instructor supervision. Attendance was monitored via session registers, with the EG maintaining an average attendance rate of 97.7% (SD 2.4%).

Data Collection

Procedures

The primary outcome of this study was smartphone addiction, measured using the SAS-SV. Secondary outcomes included nomophobia (Nomophobia Questionnaire; NMP-Q), psychological distress (Depression, Anxiety, and Stress Scale–21 [DASS-21] subscales), and loneliness (University of California, Los Angeles, Loneliness Scale). Trained assessors administered all instruments in Arabic during face-to-face sessions, with data collected electronically via Google Forms on digital tablets to minimize entry errors [28]. All assessments were conducted at 2 time points: baseline and immediately after the intervention.

XSL•FO
RenderX

Investigation Tools

Smartphone Addiction

Smartphone addiction was assessed using the SAS-SV, a self-report instrument designed to evaluate the severity of PSU and originally validated by Kwon et al [29]. The SAS-SV is a 10-item instrument rated on a 6-point scale (1="strongly disagree"; 6="strongly agree"), capturing dimensions such as daily life disturbance, withdrawal, cyberspace-oriented relationships, overuse, and tolerance. Sex-specific cutoff scores were used to classify participants as smartphone addicted (≥33 for girls and ≥31 for boys). Total scores range from 10 to 60, with higher scores indicating greater levels of PSU. An Arabic-adapted version of the SAS-SV**,** validated by Sfendla et al [30], was used to ensure linguistic and cultural appropriateness. In this sample, the scale demonstrated good internal consistency (Cronbach α=0.84).

Nomophobia

Nomophobia was assessed using the Arabic version of the NMP-Q, validated by Farchakh et al [31]. The 20-item instrument evaluates fear of being without a smartphone across four domains: (1) inability to communicate, (2) losing connectedness, (3) inability to access information, and (4) giving up convenience. Participants rated each item on a 7-point Likert scale (1="totally disagree"; 7="totally agree"). Total scores range from 20 to 140, with severity levels classified as absent (20), mild (21-59), moderate (60-99), and severe (100-140). The Arabic NMP-Q showed excellent internal consistency (Cronbach α=0.948) in this sample.

Depression, Anxiety, and Stress

Emotional distress was measured using the Arabic version of the DASS-21 validated by Ali et al [32]. The DASS-21 includes three 7-item subscales that independently measure depression, anxiety, and stress. Items are rated on a 4-point Likert scale (0="did not apply to me at all"; 3="applied to me very much/most of the time"). Severity was classified per subscale (eg, depression: 0‐9=normal; ≥28=extremely severe). The tool showed strong internal consistency for this age group (Cronbach α≥0.83).

Loneliness

Loneliness was assessed using the Arabic-adapted University of California, Los Angeles, Loneliness Scale–Short Form, validated by Alateeq et al [33]. This 8-item self-report tool evaluates perceived social isolation through statements such as "I don't have a friend." Responses are rated on a 4-point Likert scale (1="never"; 4="always"), with total scores ranging from 8 to 32. Higher scores indicate greater loneliness. Internal reliability was acceptable (Cronbach α=0.777), confirming suitability for adolescents in Tunisia.

Sample Size

An a priori power analysis was conducted using G*Power (version 3.1) [34] for a repeated-measure ANOVA with a within-between interaction. To detect a medium effect size (η_p^2=0.06, equivalent to *f*=0.25) for the primary group × time interaction, with α of .05 and 1 − β of 0.80, a total of 64 participants were required; the final sample comprised 69 adolescents. The choice of a medium effect size is consistent with prior intervention research on adolescent digital addiction and related school-based programs, which reports moderate intervention effects on problematic digital technology use outcomes [35].

Data Analysis

Data were analyzed using SPSS Statistics (version 28.0; IBM Corp). Baseline group differences were examined using independent 2-tailed *t* tests for continuous variables and chi-square tests for categorical variables. Primary intervention effects were evaluated using linear mixed-effects models, which included group, time, and the group × time interaction as fixed effects, with participants as a random effect. These models were adjusted for the covariates of age, sex, and BMI and tested for each outcome. Exploratory within-group pretest-posttest changes were assessed using paired *t* tests, and between-group differences at the posttest time point were examined using independent *t* tests. Analyses of covariance were conducted to compare adjusted postintervention outcomes between groups controlling for baseline scores as well as age, sex, and BMI. Effect sizes were calculated using η_p^2 and interpreted according to conventional benchmarks [36]. Mediation analyses were performed using bootstrapping procedures with 5000 resamples to estimate indirect effects and 95% CIs [37]. Statistical significance was set at an α value of .05 (2 tailed).

Ethical Considerations

This study received ethics approval from the local Committee for the Protection of Persons at the Faculty of Medicine of University of Sfax, Tunisia (approval 0554/2023). The research was conducted in full compliance with the ethical standards outlined in the Declaration of Helsinki (2013). Written informed consent was obtained from participants' legal guardians, and verbal assent was obtained from each adolescent prior to participation. All data, including smartphone use records, questionnaire responses, and behavioral observations, were anonymized. Any potentially identifiable information was stored separately on a password-protected server accessible only to the principal investigator. No financial or material compensation was provided to participants.

Results

Baseline Characteristics

A detailed flowchart of participant screening, randomization, allocation, and follow-up is presented in Figure 1. At baseline, no statistically significant differences were observed between the EG (n=36) and CG (n=33) across demographic, anthropometric, or psychological variables (Table 1). The groups were comparable in age, sex distribution, BMI, smartphone addiction, nomophobia, psychological distress (global and subscale scores), and loneliness (*P*>.05 in all cases). Effect size estimates were negligible, indicating satisfactory baseline equivalence prior to the intervention.

XSL•FO
RenderX

Figure 1. Participants' flow through the phases of the randomized controlled trial. PSU: problematic smartphone use; SAS-SV: Smartphone Addiction Scale–Short Version; TSSG-ME: traditional social sports games with mental exercises.

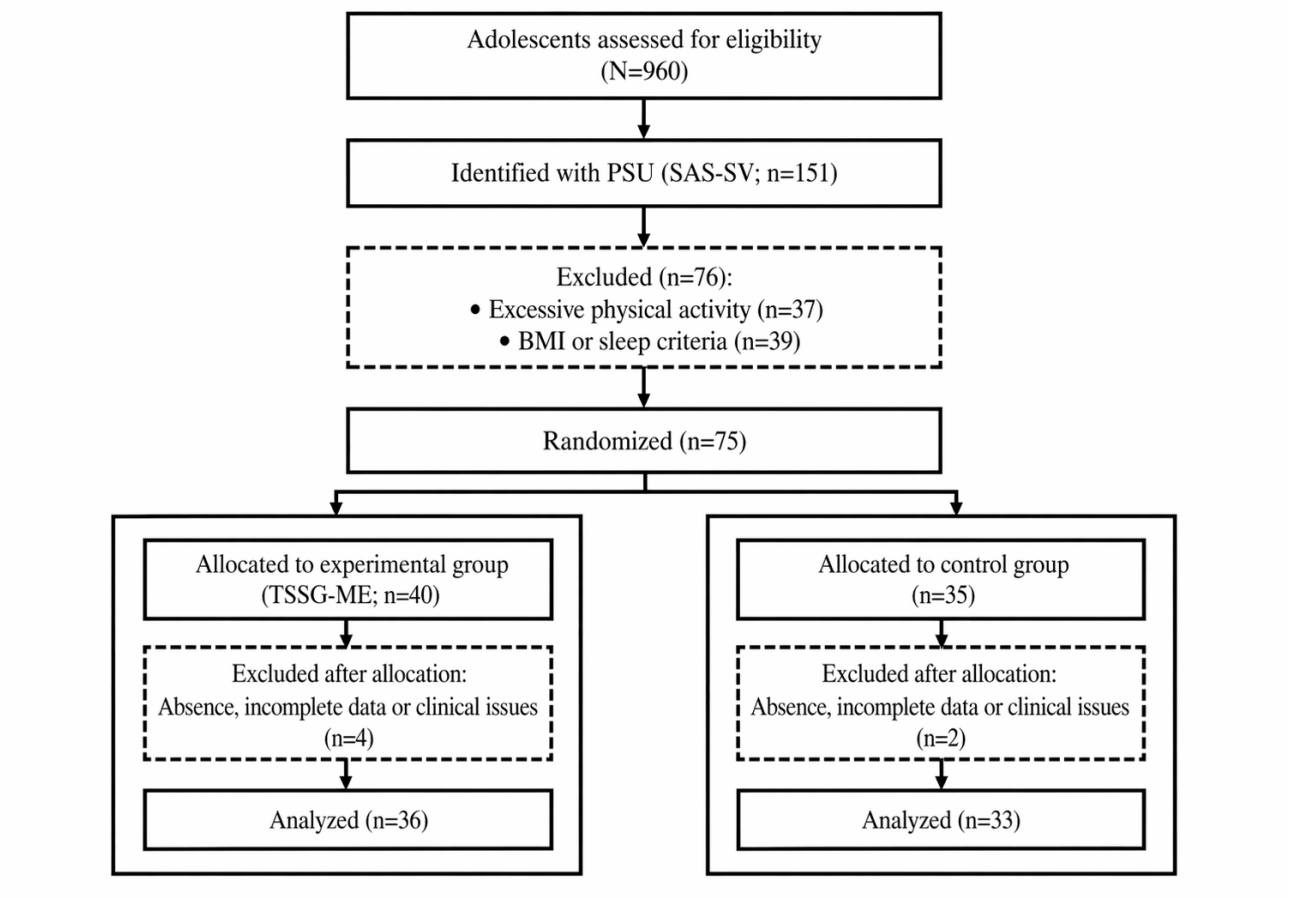


Table . Baseline characteristics of the participants (N=69)[a].

Characteristics	EG[b] (n=36)	CG[c] (n=33)
Age (y), mean (SD)	15.0 (0.6)	15.1 (0.7)
Sex (female), n (%)	16 (44.4)	14 (42.4)
BMI (kg/m^2), mean (SD)	20.7 (1.5)	20.6 (1.6)
SAS-SV[d] score (10-60), mean (SD)	38.9 (3.0)	39.0 (3.2)
NMP-Q[e] score (20-140), mean (SD)	105.2 (10.5)	106.0 (11.2)
DASS-21[f] global score (0-126), mean (SD)	70.1 (15.0)	70.8 (16.5)
DASS-21 subscale scores, mean (SD)		
Depression (0-42)	28.2 (6.5)	28.5 (7.0)
Anxiety (0-42)	22.1 (5.0)	22.3 (5.5)
Stress (0-42)	19.8 (6.0)	20.0 (6.5)
ULS-8[g] score (8-32), mean (SD)	26.8 (3.5)	27.0 (3.8)

[a]Independent *t* tests and chi-square tests confirmed no significant differences between groups at baseline. The variables of age, sex, and BMI were included as covariates in subsequent inferential models.

[b]EG: experimental group.

[c]CG: control group.

[d]SAS-SV: Smartphone Addiction Scale–Short Version.

[e]NMP-Q: Nomophobia Questionnaire.

[f]DASS-21: Depression, Anxiety, and Stress Scale–21.

[g]ULS-8: University of California, Los Angeles, Loneliness Scale.

XSL•FO
RenderX

Primary Intervention Effects (Group × Time Interaction)

Linear mixed model analyses revealed statistically significant group × time interaction effects across all primary outcomes (Table 2), consistently favoring the EG. Interaction effects were of moderate magnitude, including smartphone addiction (η_p^2=0.12; *P*=.004), nomophobia (η_p^2=0.10; *P*=.007), and global psychological distress (η_p^2=0.10; *P*=.009). Similar moderate interaction effects were observed for depression (η_p^2=0.09), anxiety (η_p^2=0.08), stress (η_p^2=0.08), and loneliness (η_p^2=0.08), indicating a consistent intervention-related improvement across behavioral and psychosocial domains.

Table . Linear mixed model results for intervention effects (group × time interaction) adjusted for age, sex, and BMI[a].

Outcome	*F* test (*df*)	*P* value	η_p^2
SAS-SV[b]	8.9 (1, 67)	.004	0.12
NMP-Q[c]	7.8 (1, 67)	.007	0.10
DASS-21[d] global	7.3 (1, 67)	.009	0.10
DASS-21 subscales			
Depression	6.9 (1, 67)	.01	0.09
Anxiety	6.2 (1, 67)	.02	0.08
Stress	5.9 (1, 67)	.02	0.08
ULS-8[e]	5.5 (1, 67)	.02	0.08

[a]*F* statistic represents fixed group × time interaction effects from linear mixed models with random intercepts for participants. Models were adjusted for age, sex, and BMI. Effect size (η_p^2) benchmarks: 0.01=small; 0.06=moderate; 0.14=large.

[b]SAS-SV: Smartphone Addiction Scale–Short Version.

[c]NMP-Q: Nomophobia Questionnaire.

[d]DASS-21: Depression, Anxiety, and Stress Scale–21.

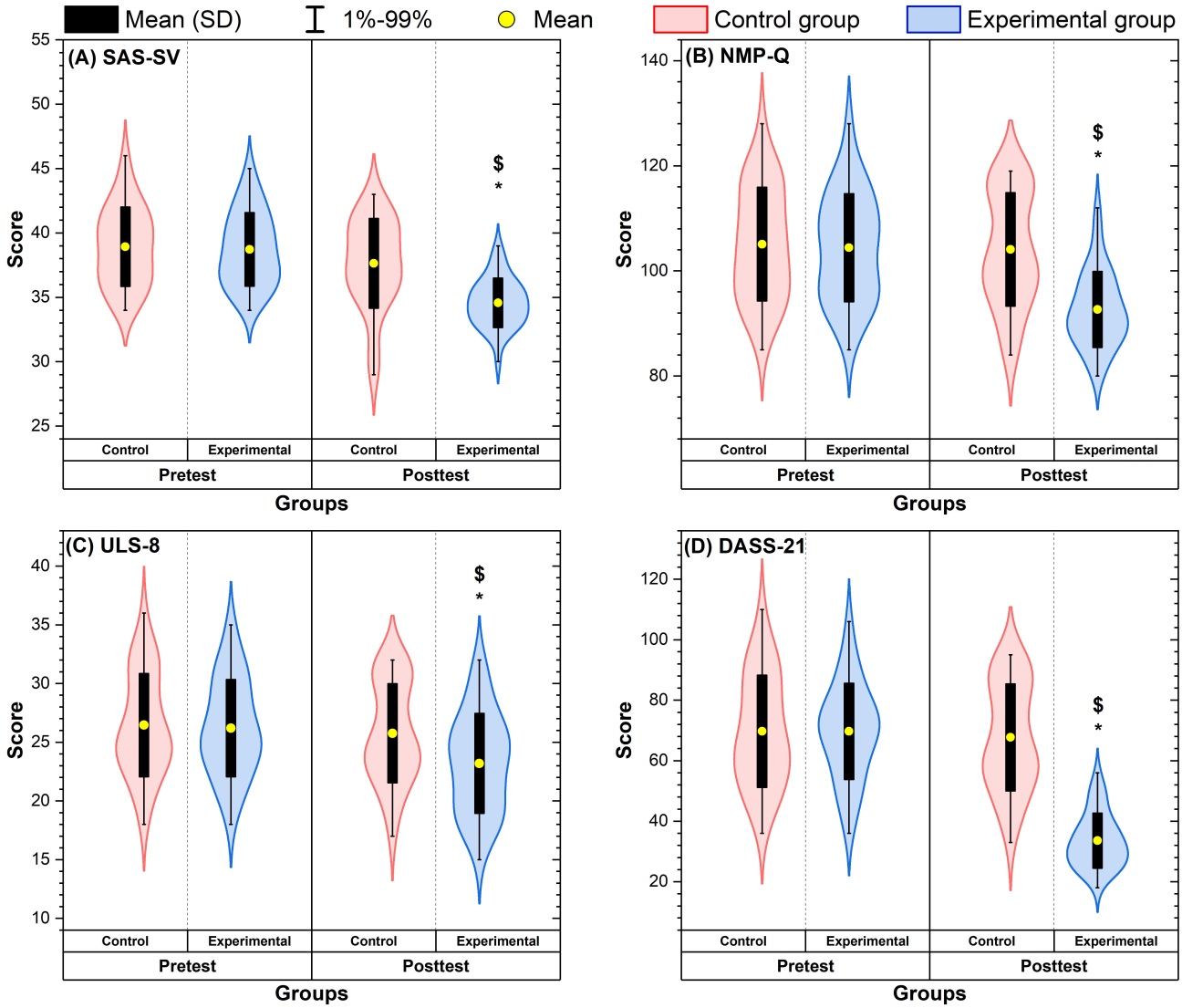
[e]ULS-8: University of California, Los Angeles, Loneliness Scale.

Pretest-Posttest Changes and Between-Group Differences

Figure 2 illustrates the distribution of pre- and postintervention scores across outcome measures for the EG and CG. Visual inspection indicates clear reductions from before to after the intervention in the EG across smartphone addiction, nomophobia, loneliness, and psychological distress (global score and subscales), whereas score distributions in the CG remained largely stable over time. Exploratory within-group comparisons showed statistically significant pretest-posttest reductions in the EG across all outcomes (*P*<.05 in all cases), whereas no significant changes were observed in the CG (*P*>.05 in all cases). Between-group comparisons at the posttest time point favored the EG across all measures (*P*≤.02).

XSL•FO
RenderX

Figure 2. Effects of the 12-week intervention on smartphone addiction and psychological outcomes. Data are presented for (A) smartphone addiction (Smartphone Addiction Scale–Short Version; SAS-SV), (B) nomophobia (Nomophobia Questionnaire; NMP-Q), (C) loneliness (University of California, Los Angeles, Loneliness Scale; ULS-8), and (D) psychological distress (Depression, Anxiety, and Stress Scale–21; DASS-21) across the experimental and control groups. Violin plots show score distributions; yellow circles indicate means, and black bars indicate means and SDs. Whiskers represent the 1st to 99th percentiles. *Significant within-group change (paired *t* test) at *P*<.05. $Significant between-group difference at the posttest time point (independent *t* test) at *P*<.05.



Adjusted Postintervention Outcomes

After controlling for baseline scores, analyses of covariance confirmed significant postintervention group differences across all outcomes (Table 3). The EG demonstrated significantly lower adjusted posttest scores than the CG for smartphone addiction (η_p^2=0.12; *P*=.004), nomophobia (η_p^2=0.11; *P*=.006), global psychological distress (η_p^2=0.12; *P*=.005), depression (η_p^2=0.10; *P*=.008), anxiety (η_p^2=0.09; *P*=.01), stress (η_p^2=0.09; *P*=.02), and loneliness (η_p^2=0.09; *P*=.01). All effects were in the moderate range, supporting the robustness of the intervention after adjustment for baseline variability.

XSL•FO
RenderX

Table . Analysis of covariance (ANCOVA) results for posttest scores adjusted for baseline scores, age, sex, and BMI[a].

Outcome	EG[b], adjusted mean (SE)	CG[c], adjusted mean (SE)	*F* test (*df*)	*P* value	η_p^2
SAS-SV[d]	34.6 (0.5)	37.8 (0.5)	9.1 (1, 66)	.004	0.12
NMP-Q[e]	95.2 (1.6)	101.7 (1.7)	8.2 (1, 66)	.006	0.11
DASS-21[f] global	58.1 (1.7)	66.4 (1.8)	8.7 (1, 66)	.005	0.12
DASS-21 subscales					
Depression	23.6 (0.8)	27.1 (0.9)	7.4 (1, 66)	.008	0.10
Anxiety	18.8 (0.7)	21.6 (0.7)	6.8 (1, 66)	.01	0.09
Stress	16.3 (0.7)	19.1 (0.8)	6.2 (1, 66)	.02	0.09
ULS-8[g]	24.9 (0.4)	26.6 (0.4)	6.5 (1, 66)	.01	0.09

[a]ANCOVA models were adjusted for baseline (pretest) scores, age, sex, and BMI. Homogeneity of regression slopes was verified for all models. Effect size (η_p^2) benchmarks: 0.01=small; 0.06=moderate; 0.14=large. The Smartphone Addiction Scale–Short Version clinical cutoffs applied were ≥31 for boys and ≥33 for girls.

[b]EG: experimental group.

[c]CG: control group.

[d]SAS-SV: Smartphone Addiction Scale–Short Version.

[e]NMP-Q: Nomophobia Questionnaire.

[f]DASS-21: Depression, Anxiety, and Stress Scale–21.

[g]ULS-8: University of California, Los Angeles, Loneliness Scale.

Mediation by Loneliness Reduction

Mediation analysis showed that reductions in loneliness partially mediated the effect of the intervention on postintervention smartphone addiction (Table 4 and Figure 3). The indirect effect was statistically significant (*ab*=−1.10, 95% CI −2.05 to −0.35), accounting for 34.4% of the total effect. The intervention significantly reduced loneliness (*a*=−1.80; *P*=.002), and higher loneliness was associated with greater smartphone addiction (*b*=0.61; *P*<.001). The direct effect of the intervention remained significant after accounting for loneliness, indicating partial mediation. No adverse events or unintended effects were reported in either group during the trial period.

Table . Simple mediation analysis (PROCESS Model 4) adjusted for baseline scores, age, sex, and BMI[a].

Path and effect	Unstandardized coefficient (SE; 95% CI)	*t* test (*df*)	*P* value
Total effect (*c*)			
[b]Effect of group on SAS-SV[b] (after)	−3.20 (0.85; −4.89 to −1.51)	−3.76 (67)	<.001
Direct effect (*c′*)			
Effect of group on SAS-SV (after)	−2.10 (0.90; −3.89 to −0.31)	−2.33 (66)	.02
Indirect effect (combined; *ab*)	−1.10 (0.45; −2.05 to −0.35)	—[c]	—
Indirect effect (path specific)			
Effect of group on loneliness (*a*)	−1.80 (0.55; −2.88 to –0.72)	−3.27 (67)	.002
Effect of loneliness on SAS-SV (*b*)	0.61 (0.11; 0.39 to 0.83)	5.55 (66)	<.001

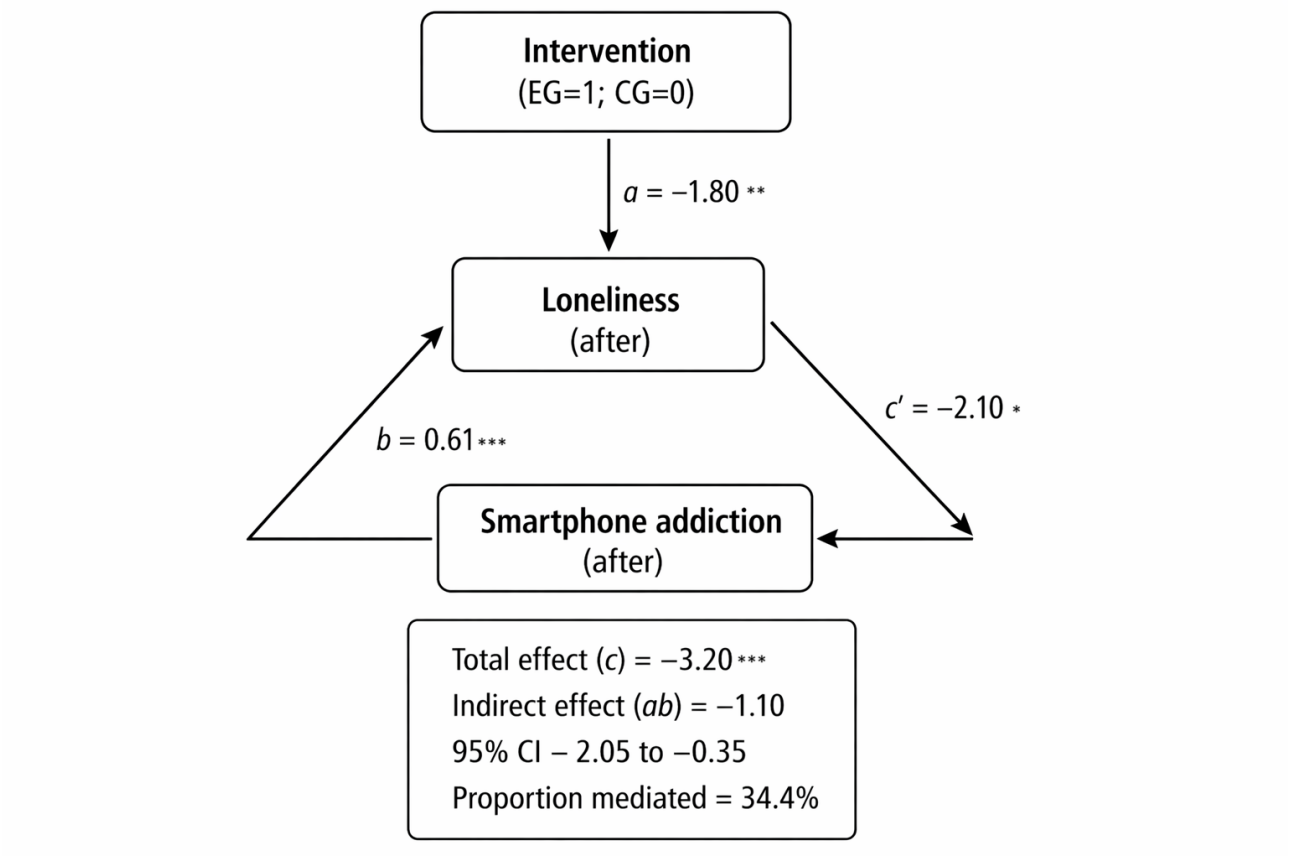
[a]Indirect effect 95% CIs were derived from 5000 bootstrap samples. The model was adjusted for age, sex, BMI, and baseline scores (Smartphone Addiction Scale–Short Version and University of California, Los Angeles, Loneliness Scale). Path *a*=intervention→loneliness; path *b*=loneliness→Smartphone Addiction Scale–Short Version. The proportion mediated was 34.4%.

[b]SAS-SV: Smartphone Addiction Scale–Short Version.

[c]Not applicable. The indirect effect (ab) is estimated using bootstrap procedures and therefore does not have a corresponding *t* value, df, or *P* value.

XSL•FO
RenderX

Figure 3. Loneliness (University of California, Los Angeles, Loneliness Scale) as a mediator of the intervention effect on smartphone addiction (Smartphone Addiction Scale–Short Version; SAS-SV)—simple mediation model (PROCESS Model 4) with 5000 bootstrap samples. The model was adjusted for age, sex, BMI, and baseline scores. The values shown are unstandardized coefficients. Path *a*=intervention→loneliness; path *b*=loneliness→SAS-SV. The indirect effect was significant (95% CI did not include 0), with 34.4% of the total effect mediated by loneliness. **P*<.05; ***P*<.01; ****P*<.001; CG: control group; EG: experimental group.



Discussion

Summary of Main Findings

This randomized controlled trial evaluated a 12-week, school-based intervention integrating traditional social sports games with mental exercises for Tunisian adolescents with elevated smartphone addiction. Linear mixed model analyses revealed significant group × time interaction effects of moderate magnitude (η_p^2=0.08-0.12), consistently favoring the EG over the CG. Participants receiving the TSSG-ME intervention demonstrated statistically significantly greater reductions in smartphone addiction, nomophobia, psychological distress (global and subscale scores), and loneliness. After adjusting for baseline scores, between-group differences at the posttest time point remained significant with moderate effect sizes. Furthermore, mediation analysis indicated that reductions in loneliness accounted for approximately 34.4% of the intervention's total effect on smartphone addiction, supporting a partial mediation model.

Interpretation and Comparison With the Literature

Reductions in smartphone addiction were among the most robust outcomes. These findings are consistent with those of prior studies showing that structured physical engagement and psychological support can reduce behavioral dependence on mobile technologies [38]. This aligns with meta-analytic findings that combined interventions integrating PA with psychological skill building produce more consistent reductions in smartphone addiction than single-component approaches [11,14]. This study contributes by demonstrating the efficacy of a culturally grounded program in a Tunisian context.

The term "traditional social sports games" denotes structured, rule-based PAs with inherent social organization and deep cultural resonance, distinguishing them from generic physical exercise [18,19]. The mental exercise component refers to a standardized protocol of evidence-based techniques adapted from sports psychology designed to build self-regulation and stress tolerance beyond simple relaxation [26,27]. The observed moderate effect sizes were clinically meaningful and consistent with those reported in other behavioral intervention trials targeting PSU [39]. They reflect a realistic and attainable impact, particularly given that our sample comprised adolescents with clinically elevated baseline symptoms who may have had greater room for improvement.

The reduction in nomophobia underscores the intervention's potential to alleviate the anxiety associated with smartphone separation. By fostering rewarding, in-person social experiences and peer validation, the traditional games likely diminished the perceived indispensability of constant digital connectivity [6]. Concurrently, the mental exercise component equipped participants with concrete skills for managing emotional arousal,

XSL•FO
RenderX

offering a healthier alternative to digital coping mechanisms [40].

Critically, our findings provide empirical support for the psychological mechanism proposed by SDT [15]. Reductions in loneliness mediated 34.4% of the intervention's total effect on smartphone addiction, highlighting a key pathway. This suggests that the TSSG-ME program enhanced the basic psychological need for relatedness through prosocial, face-to-face interaction in traditional games, reducing perceived social isolation. Consequently, adolescents' compensatory reliance on smartphones for social fulfillment likely decreased, alleviating addiction symptoms. This mechanism aligns with research identifying loneliness as a critical link between unmet social needs and PSU [41,42] and supports evidence that loneliness mediates the relationship between health-promoting behaviors and reduced maladaptive digital use [13,43]. Using culturally embedded games to foster relatedness also aligns with literature emphasizing their effectiveness in promoting sociability [44]. A key limitation is that mediation was tested with concurrent postintervention measures; temporal precedence cannot be fully established, so causal interpretation should be made cautiously.

The dual-component design posits complementary pathways: the traditional social sports games primarily targeted relatedness through social engagement and shared PA, whereas the mental exercise component targeted competence and autonomy through enhanced self-regulation [45]. Their integration likely addressed the multifaceted social-motivational and emotional-cognitive vulnerabilities underpinning smartphone addiction more comprehensively than either component alone [46]. Future research using factorial designs is necessary to disentangle their independent and synergistic contributions.

Limitations

Several limitations should be acknowledged. First, the study relied on self-reported measures, which may be subject to response bias; future research should incorporate objective metrics such as device-based screen time tracking. Second, the culturally specific sample, while emphasizing the importance of contextual adaptation, may limit the generalizability of the findings to broader populations. Third, the absence of long-term follow-up assessments precludes conclusions regarding the sustainability of the intervention effects. Fourth, the use of a passive CG (standard physical education) rather than an active, time-matched comparator (eg, generic group sports or relaxation program) restricts the ability to attribute observed outcomes specifically to the traditional social games and mental training components as opposed to nonspecific factors such as increased social engagement or novelty. Finally, although effect sizes were statistically significant and clinically meaningful, they were moderate, highlighting the complex, multifactorial nature of PSU and suggesting the need for further optimization of the intervention.

Conclusions

This study provides evidence that a culturally adapted, school-based intervention combining traditional social sports games and mental training can yield significant, moderate improvements in smartphone addiction, nomophobia, psychological distress, and loneliness among adolescents. The mediation of the treatment effect through reduced loneliness highlights the importance of enhancing real-world social connectedness, supporting a key mechanism derived from SDT. By providing alternative, healthier avenues for satisfying psychological needs for relatedness and competence, programs such as TSSG-ME may reduce reliance on maladaptive smartphone use.

Beyond individual outcomes, these findings suggest that integrating culturally relevant social activities into school-based programs may serve as a scalable strategy to support adolescent mental health in increasingly digital contexts. Policymakers and educators could leverage such approaches to mitigate the negative consequences of excessive smartphone use. Future research should use longitudinal designs, active CGs, and diverse populations to confirm causal pathways and assess long-term efficacy. However, given that mediation was examined cross-sectionally, causal interpretations should be made cautiously.

Acknowledgments

The authors express their gratitude to the Princess Nourah bint Abdulrahman University Researchers Supporting Project (PNURSP2026R145), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia, for funding this research. The authors used SciSpace and Copilot during the preparation of this work. These tools were used for literature organization and language editing. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

This study was funded by the Princess Nourah bint Abdulrahman University Researchers Supporting Project (PNURSP2026R145), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The funders had no role in the study design; data collection, analysis, and interpretation; manuscript preparation; or decision to publish.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

Conceptualization: MY, OT
Data curation: A Ghorbel
Formal analysis: A Ghorbel, FHY
Funding acquisition: SME
Investigation: MY, HA, SME, A Gharbi
Methodology: LM, OT
Project administration: HA
Resources: OT
Software: FHY
Supervision: LM, OT
Validation: LM
Visualization: A Ghorbel, FHY, GB
Writing—original draft: MY
Writing—review and editing: HA, SME, GB, A Gharbi

Conflicts of Interest

None declared.

Editorial Notice

This randomized study was retrospectively registered. The authors noted that the registration was conducted retrospectively due to an administrative oversight during the initial phase of the study, although the protocol was approved by the Committee for the Protection of Persons prior to participant recruitment. The editor granted an exception from International Committee of Medical Journal Editors rules mandating prospective registration of randomized trials because the risk of bias appeared low and the study information registered with the authors' ethics committee matched the trial reporting. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Checklist 1
CONSORT checklist.
[PDF File, 446 KB - mhealth_v14i1e85353_app1.pdf]

References

1. Marciano L, Viswanath K, Morese R, Magis-Weinberg L, Dubicka B, Weber R. Digital media, cognition, and brain development in adolescence. In: Christakis DA, Hale L, editors. Handbook of Children and Screens: Digital Media, Development, and Well-Being From Birth Through Adolescence: Springer; 2025:21-29. [doi: 10.1007/978-3-031-69362-5_4]
2. Halayem S, Nouira O, Bourgou S, Bouden A, Othman S, Halayem M. The mobile: a new addiction upon adolescents [Article in French]. Tunis Med 2010 Aug;88(8):593-596. [Medline: 20711967]
3. Maamri H, Yaich S, Ketata N, et al. Santé et mode de vie des adolescents scolarisés dans la région de Sfax, Tunisie. Rev Epidemiol Sante Publique 2023 Sep;71(Supplement 3):102021. [doi: 10.1016/j.respe.2023.102021]
4. Ben Thabet J, Ellouze AS, Ghorbel N, et al. Factors associated with internet addiction among Tunisian adolescents [Article in French]. Encephale 2019 Dec;45(6):474-481. [doi: 10.1016/j.encep.2019.05.006] [Medline: 31421811]
5. Wang A, Wang Z, Zhu Y, Shi X. The prevalence and psychosocial factors of problematic smartphone use among Chinese college students: a three-wave longitudinal study. Front Psychol 2022;13:877277. [doi: 10.3389/fpsyg.2022.877277] [Medline: 35450331]
6. Jelleli H, Hindawi O, Rebhi M, et al. Psychometric evidence of the Arabic version of Nomophobia Questionnaire among physical education students. Psychol Res Behav Manag 2023;16:2383-2394. [doi: 10.2147/PRBM.S416312] [Medline: 37408785]
7. Rodríguez-García AM, Moreno-Guerrero AJ, López Belmonte J. Nomophobia: an individual's growing fear of being without a smartphone-a systematic literature review. Int J Environ Res Public Health 2020 Jan 16;17(2):580. [doi: 10.3390/ijerph17020580] [Medline: 31963208]
8. Gupta AK, Singh PS, Krishak AK. Smartphone addiction: impact on health and well-being. Int J Community Med Public Health 2024;11(5):2100-2106. [doi: 10.18203/2394-6040.ijcmph20241213]
9. Maisarah M, Khairani M. How excessive use of smartphone are becoming loneliness in young adult. S-JPU 2023;6(2):151-169. [doi: 10.24815/s-jpu.v6i2.33282]
10. Shao S, Mao N, Zhang Z, et al. Adolescents' loneliness and problematic smartphone use: the mediating role of upward social comparison and the moderating role of self‐identity. Psychol Sch 2024;61(8):3224-3242. [doi: 10.1002/pits.23215]

XSL•FO
RenderX

11. Liu S, Xiao T, Yang L, Loprinzi PD. Exercise as an alternative approach for treating smartphone addiction: a systematic review and meta-analysis of random controlled trials. Int J Environ Res Public Health 2019 Oct 15;16(20):3912. [doi: 10.3390/ijerph16203912] [Medline: 31618879]
12. Xu J, Tang L. The relationship between physical exercise and problematic internet use in college students: the chain-mediated role of self-control and loneliness. BMC Public Health 2024 Jun 27;24(1):1719. [doi: 10.1186/s12889-024-19226-x] [Medline: 38937729]
13. Zhu X, Niu H, Zhu Q, Chen W. Physical activity and mobile phone addiction among Chinese college students: a chain mediation model of social anxiety and loneliness. Front Psychol 2025;16:1545440. [doi: 10.3389/fpsyg.2025.1545440] [Medline: 40330309]
14. Meng S, Bai G, Shen P, et al. The effectiveness of physical activity intervention on mobile phone addiction in college students: the chain mediating role of emotion regulation strategies and positive coping style. BMC Public Health 2025 Sep 25;25(1):3138. [doi: 10.1186/s12889-025-24488-0] [Medline: 40999393]
15. Deci EL, Ryan RM. Self-determination theory: a macrotheory of human motivation, development, and health. Can Psychol 2008;49(3):182-185. [doi: 10.1037/a0012801]
16. Fan H, Qi S, Huang G, Xu Z. Effect of acute aerobic exercise on inhibitory control of college students with smartphone addiction. Evid Based Complement Alternat Med 2021;2021:5530126. [doi: 10.1155/2021/5530126] [Medline: 34394381]
17. Li Z, Xia X, Sun Q, Li Y. Exercise intervention to reduce mobile phone addiction in adolescents: a systematic review and meta-analysis of randomized controlled trials. Front Psychol 2023;14:1294116. [doi: 10.3389/fpsyg.2023.1294116] [Medline: 38192396]
18. Amar IB, Jlassi A, Kacem N, Elloumi A. Towards a school physical education learning model based on traditional sport game to develop pupil well-being in Tunisia. Adv Soc Sci Manag 2023;1(1):22-36 [FREE Full text]
19. Elloumi A, Parlebas P. Sociocultural analysis of Tunisian traditional sports. Fit Per J 2009 Mar 1;8(2):136-145. [doi: 10.3900/fpj.8.2.136.e]
20. Arumugam A, Alsaafin N, Shalash RJ, et al. Concurrent validity between self-reported International Physical Activity Questionnaire Short Form and Fibion accelerometer data among young adults in the UAE. Eur J Med Res 2024 Aug 19;29(1):426. [doi: 10.1186/s40001-024-01975-5] [Medline: 39155363]
21. Growth reference data for 5-19 years. World Health Organization. URL: https://www.who.int/tools/growth-reference-data-for-5to19-years [accessed 2026-04-20]
22. Hirshkowitz M, Whiton K, Albert SM, et al. National Sleep Foundation's sleep time duration recommendations: methodology and results summary. Sleep Health 2015 Mar;1(1):40-43. [doi: 10.1016/j.sleh.2014.12.010] [Medline: 29073412]
23. Maaloul H, Chalghaf N, Hentati A, Hamani J, Azaiez F. The Tunisian school time allocated to the matter of sport and physical education. IOSR J Humanit Soc Sci 2013;15(4):61-65. [doi: 10.9790/0837-1546165]
24. Li J, Li YM, Xing YX, Zhang B, Tang Y, Drasgow F. Development and validation of a game-based assessment for complex problem solving. J Intell 2025 Jan 14;13(1):9. [doi: 10.3390/jintelligence13010009] [Medline: 39852418]
25. Richter K, Kickmeier-Rust MD, Tschirky D. Mastering the game: how level structure and game elements shape competency acquisition. Proc Eur Conf Games Based Learn 2024;18(1). [doi: 10.34190/ecgbl.18.1.2695]
26. Tsakona P, Kitsatis I, Apostolou T, Papadopoulou O, Hristara-Papadopoulou A. The effect of diaphragmatic breathing as a complementary therapeutic strategy in stress of children and teenagers 6-18 years old. Children (Basel) 2025 Jan 5;12(1):59. [doi: 10.3390/children12010059] [Medline: 39857890]
27. Retnowati R, Waluyo A, Zahra AN, Purba S. Evidence based nursing Jacobson's Progressive Muscle Relaxation untuk mengurangi kecemasan dan depresi pada pasien dengan HIV. Joting 2024;6(1):684-692. [doi: 10.31539/joting.v6i1.9001]
28. Tate A, Smallwood C. Comparing the efficiency of paper-based and electronic data capture during face-to-face interviews. PLoS One 2021;16(3):e0247570. [doi: 10.1371/journal.pone.0247570] [Medline: 33684116]
29. Kwon M, Kim DJ, Cho H, Yang S. The Smartphone Addiction Scale: development and validation of a short version for adolescents. PLoS One 2013;8(12):e83558. [doi: 10.1371/journal.pone.0083558] [Medline: 24391787]
30. Sfendla A, Laita M, Nejjar B, Souirti Z, Touhami AA, Senhaji M. Reliability of the Arabic Smartphone Addiction Scale and Smartphone Addiction Scale-Short Version in two different Moroccan samples. Cyberpsychol Behav Soc Netw 2018 May;21(5):325-332. [doi: 10.1089/cyber.2017.0411] [Medline: 29762065]
31. Farchakh Y, Hallit R, Akel M, et al. Nomophobia in Lebanon: scale validation and association with psychological aspects. PLoS One 2021;16(4):e0249890. [doi: 10.1371/journal.pone.0249890] [Medline: 33878132]
32. Ali AM, Ahmed A, Sharaf A, Kawakami N, Abdeldayem SM, Green J. The Arabic version of the Depression Anxiety Stress Scale-21: cumulative scaling and discriminant-validation testing. Asian J Psychiatr 2017 Dec;30:56-58. [doi: 10.1016/j.ajp.2017.07.018] [Medline: 28755630]
33. Alateeq D, Aljabri A, Aldogiam G, et al. Psychometric properties of the Arabic version of the UCLA Loneliness Scale among Saudi female undergraduate students. Int J Med Dev Ctries 2022;6(1):149-154. [doi: 10.24911/IJMDC.51-1638046495]
34. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]

XSL•FO
RenderX

35. Žmavc M, Horvat J, Židan M, Selak Š. The effectiveness of school-based interventions to reduce problematic digital technology use and screen time: a systematic review and meta-analysis. J Behav Addict 2025 Jun 23;14(2):571-589. [doi: 10.1556/2006.2025.00043] [Medline: 40549948]
36. Richardson JT. Eta squared and partial eta squared as measures of effect size in educational research. Educ Res Rev 2011 Jan;6(2):135-147. [doi: 10.1016/j.edurev.2010.12.001]
37. Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach: Guilford Publications; 2022.
38. Tang AC, Lee RL. Effects of a group mindfulness-based cognitive programme on smartphone addictive symptoms and resilience among adolescents: study protocol of a cluster-randomized controlled trial. BMC Nurs 2021 Jun 5;20(1):86. [doi: 10.1186/s12912-021-00611-5] [Medline: 34090443]
39. Precht LM, Mertens F, Brickau DS, et al. Engaging in physical activity instead of (over)using the smartphone: an experimental investigation of lifestyle interventions to prevent problematic smartphone use and to promote mental health. Z Gesundh Wiss 2023 Feb 9:1-19. [doi: 10.1007/s10389-023-01832-5] [Medline: 36785655]
40. Goutama E, Chandra Widjanantie S. The effect of diaphragm breathing exercises on physiological and psychological stress. Indones J Phys Med Rehabil 2023;12(2):197-205. [doi: 10.36803/indojpmr.v12i02.344]
41. Bian M, Leung L. Linking loneliness, shyness, smartphone addiction symptoms, and patterns of smartphone use to social capital. Soc Sci Comput Rev 2015 Feb;33(1):61-79. [doi: 10.1177/0894439314528779]
42. Peng Y, Mao H, Zhang B, et al. Depression and loneliness as mediators between social support and mobile phone addiction. Psychiatr Danub 2022;34(3):475-482. [doi: 10.24869/psyd.2022.475] [Medline: 36256985]
43. Zeng M, Chen S, Zhou X, Zhang J, Chen X, Sun J. The relationship between physical exercise and mobile phone addiction among Chinese college students: testing mediation and moderation effects. Front Psychol 2022;13:1000109. [doi: 10.3389/fpsyg.2022.1000109] [Medline: 36262440]
44. Karuc S, Bora MM, Bora MV. Sports participation motivation, attitudes, and digital gaming behaviors among adolescent students. Int J Technol Educ 2025;8(4):1161-1184. [doi: 10.46328/ijte.5802]
45. Ji F, Sun Q, Han W, Li Y, Xia X. How physical exercise reduces problematic mobile phone use in adolescents: the roles of expression suppression, depression, anxiety, and resilience. Psychol Res Behav Manag 2024;17:4369-4382. [doi: 10.2147/PRBM.S484089] [Medline: 39722776]
46. Liu H, Soh KG, Samsudin S, Rattanakoses W, Qi F. Effects of exercise and psychological interventions on smartphone addiction among university students: a systematic review. Front Psychol 2022;13:1021285. [doi: 10.3389/fpsyg.2022.1021285] [Medline: 36275318]

Abbreviations:

CG: control group
CONSORT: Consolidated Standards of Reporting Trials
DASS-21: Depression, Anxiety, and Stress Scale–21
EG: experimental group
NMP-Q: Nomophobia Questionnaire
PA: physical activity
PSU: problematic smartphone use
SAS-SV: Smartphone Addiction Scale–Short Version
SDT: self-determination theory
TSSG-ME: traditional social sports games with mental exercises

Edited by A Stone; submitted 06.Oct.2025; peer-reviewed by AC Lusar, V Stefanica; revised version received 02.Apr.2026; accepted 09.Apr.2026; published 04.May.2026.

Please cite as:
Yaakoubi M, Ghorbel A, Abdelkafi H, Masmoudi L, Trabelsi O, Elkholi SM, Yagin FH, Badicu G, Gharbi A
Traditional Social Sports Games and Mental Training for Smartphone Addiction and Psychological Distress in School-Aged Adolescents: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e85353
URL: https://mhealth.jmir.org/2026/1/e85353
doi:10.2196/85353

© Mohamed Yaakoubi, Ahmed Ghorbel, Hiba Abdelkafi, Liwa Masmoudi, Omar Trabelsi, Safaa M. Elkholi, Fatma H. Yagin, Georgian Badicu, Adnene Gharbi. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 4.May.2026.

XSL•FO
RenderX

This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Multimodal Personalized Mobile Health Just-in-Time Adaptive Intervention for Occupational Stress Management: Pilot Study

Beatrix Kunas[1], PhD; Oliver Jung[2], BEng; Christoph Schranz[2], MSc; Mathias Schmoigl-Tonis[2], MSc; Julian Mehlis[1], BSc; Anton-Rupert Laireiter[1], Dr

[1]Department of Psychology, University of Salzburg, Hellbrunner Str. 34, Salzburg, Austria

[2]Salzburg Research Forschungsgesellschaft mbH, Salzburg, Austria

Corresponding Author:
Beatrix Kunas, PhD
Department of Psychology, University of Salzburg, Hellbrunner Str. 34, Salzburg, Austria

Abstract

Background: As digital health solutions gain traction, there is an urgent need for effective, person-centered stress management tools for employees. Advances in wearable stress monitoring and machine learning now enable the collection of high-resolution, real-time data and the delivery of personalized interventions with respect to both timing and content. Despite this technological progress, there remains a notable paucity of mobile health (mHealth) interventions that capitalize on these capabilities to implement just-in-time adaptive interventions (JITAIs).

Objective: This pilot study aimed to examine the preliminary effects of the RELAX mHealth app, which uses a JITAI framework to reduce occupational stress, with a particular focus on initial indications of stress reduction, the influence of the app's degree of personalization on these effects, and overall user experience.

Methods: A 2-arm randomized trial was conducted with 46 employees across two phases of 3 weeks each. Participants used the RELAX mHealth app, which adapts intervention timing based on real-time stress data and tailors content according to users' preferences. Stress was assessed via multiple heart rate variability (HRV) parameters measured with wearable sensors, alongside subjective questionnaires and ecological momentary assessment (EMA) questionnaires. Usability and user satisfaction were also evaluated. In the second phase, a group randomization was implemented, with two groups of different degrees of in-app personalization.

Results: The findings were mixed. While selected subjective stress indicators decreased from baseline to the end of the study, one physiological measure increased. Retrospective chronic stress (P=.048) and arousal related to the most recent stress event prior to EMA entry (P=.02) were reduced; however, the low frequency (LF)-to-high frequency (HF) ratio of HRV increased (P=.03). Other stress variables, such as work-related stress, valence of daily experiences, and momentary arousal, showed no change over the course of the study. No between-group differences were observed with respect to the assigned degree of personalization. Additionally, usability increased (P<.001), whereas satisfaction decreased at a trend level (P=.05).

Conclusions: Although the results provide selective preliminary indications of stress reduction with app use, the variability of outcomes raises concerns about its current practical utility. The effects are also discussed with regard to the specific design of the prototype and the study setup. Technical issues and user discomfort were noted despite increasing usability over the study period, indicating areas for improvement in the developed prototype. Given the absence of a control group and the use of interventional arms only, no definitive conclusions about effects can be drawn. Nevertheless, this study highlights both the potential and the challenges of a personalized mHealth approach to occupational stress management. Further refinement of the app and more comprehensive evaluations are needed to fully understand the impact and functional usability of this JITAI-based approach in real-world settings.

Trial Registration: OSF 10.17605/OSF.IO/MYRD9; https://osf.io/myrd9/overview

International Registered Report Identifier (IRRID): RR2-10.2196/68012

(JMIR Mhealth Uhealth 2026;14:e79642) doi:10.2196/79642

KEYWORDS

occupational stress; mobile health; mHealth; just-in-time adaptive intervention concept; wearables; personalized interventions

XSL•FO
RenderX

Introduction

Mobile Health and Individual Occupational Stress

Digital health solutions, including mobile health (mHealth), hold considerable promise as a means of delivering cost-effective, low-threshold interventions to individuals via their personal smartphones [1]. This form of health care has experienced substantial growth in recent years [2], and digital solutions have the potential to deliver effective stress management, particularly for employees [3]. The habitual use of smartphones during working hours [4] presents a significant opportunity for mHealth apps to be used in close proximity to stress-inducing factors, both in terms of temporal and local proximity. However, the precise conditions of occupational stress of the individual employee can vary considerably [5], thus hindering approaches that follow a one-size-fits-all concept. If not fully addressed, individual stress can have a detrimental effect on health and well-being over time [6]. Depending on the individual situation, the consequences range from loss of performance and mental and physical issues [7-9] to economic consequences for the company [10]. It can be concluded that there is a significant need for interventions that optimally address the individual circumstances of employees to prevent the adverse consequences.

Numerous mHealth approaches for occupational stress have been developed across both the unevaluated commercial sector and those with scientifically proven effectiveness [11-13]. The one-size-fits-all approach is predominantly adopted, with a single or limited number of interventions from cognitive behavioral therapy (CBT) being offered in the form of a smartphone app [11-13]. These apps include specific interventions, such as relaxation techniques, psychoeducation, and cognitive methods, with the objective of providing users with generalized low-threshold support in managing occupational stress. The suitability of these apps for specific stress situations is determined by the user, as these apps operate in a passive manner. This highlights the limitations of such apps in providing any depth and personalized approach of traditional psychological counseling or therapy by a qualified professional. The incorporation of individual needs and situational factors, recognized as essential elements in the efficacy of psychological interventions [14], is a crucial aspect that remains beyond the capabilities of these apps.

Just-in-Time Adaptive Intervention Concept in mHealth

Technical advances in mHealth can replicate traditional face-to-face counseling if they adhere to the just-in-time adaptive intervention (JITAI) concept. JITAI refers to interventions that are delivered at opportune moments and adapt their content to the needs and conditions of the user at those moments [15]. Such an approach requires continuous transmission of real-time personal data to best adapt to prevailing circumstances [16]. This rapidly generates a substantial volume of data, necessitating the incorporation of advanced artificial intelligence (AI) techniques [17]. Algorithms and learning agents, operating discreetly in the background of the app, process the incoming data to ensure optimal personalization [18]. Therefore, the advanced capabilities of mobile devices and the promising concept of JITAIs are mutually beneficial. This approach could multiply personalization by leveraging smartphone technology to provide constant support, potentially surpassing the limitations of conventional, practitioner-based meetings.

In the context of occupational stress, JITAIs have been regarded as a particularly promising medium for providing personalized stress management assistance via smartphones, irrespective of the presence of stress during the workday [19,20]. Design methodologies incorporating personal demographic data to customize the frequency of app interventions are already in practice [21,22]. Nonetheless, these methodologies signify a rudimentary approach to the potential personalization of mobile app use. The development of advanced diagnostic methods for personal stress, such as additional self-report tools and passive sensors, has the potential to enable the app to better understand users' needs and to adapt in real time for JITAI foundations [23,24]. The use of this multimodal approach in the context of occupational stress is already underway [11,23]. The use of "wearables" is particularly advantageous in this context, as they can be worn during work and thus provide a continuous input of personal stress data for the app without needing much attention [25]. In this respect, heart rate variability (HRV) [26,27] is the stress parameter that is most commonly used and validated.

Recent Developments and Research

Developments of JITAIs for mHealth are in their infancy, primarily due to the current lack of practical implementations. In general stress management, unobtrusive wearables and app interventions can be effectively combined [28,29], with a few personalized work-stress mHealth offerings identified [23,30]. These approaches often target physical rather than mental stress [31], and most methods limit themselves to monitoring without integrating the data into interventions [24]. This results in the development of personalized apps, but not personalized interventions, thereby not aligning with the JITAI concept. Furthermore, current approaches use either the app or the wearable for measurements and interventions, exhibiting a lack of comprehensive, dynamic integration of multimodal measurement options and data for personalized interventions [23]. mHealth biofeedback apps enable interactive, tailored strategies for occupational stress by integrating real-time sensor data [32]. They provide immediate stress feedback, and when combined with breathing exercises, users can modulate heart rate (HR) through guided breathing, influencing physiological stress [19,33,34]. This interaction supports personalized data analysis and adaptive mHealth interventions. Nevertheless, the time-dependent parameter of the adaptive concept is absent in this instance, thus preventing full compliance with the JITAI concept, leaving users to manage the task of applying the interventions at the optimal time. Moreover, the selection of a singular intervention method restricts the capacity to thoroughly address the user's unique stress circumstances and individual needs.

While there is one approach that integrates several additional CBT interventions alongside biofeedback for workplace stress,

XSL•FO
RenderX

it does not leverage sensor data [19], indicating that the full benefits of ongoing stress monitoring are not fully used. At present, the application to personalize interventions using individual sensor data to tailor the full spectrum of possible interventional conditions, such as content, modules, and presentation timing of the intervention, in line with the JITAI framework, is primarily only conceptualized as such [35,36]. To the best of our knowledge, only one study has formally implemented the concept of automatically selecting CBT interventions based on prior stress indicators measured with a wearable device [37], thereby aligning with the criteria established by JITAI. However, further personalization through long-term learning of individual user preferences, a potential of learning agents, has not been addressed here.

RELAX Prototype and Research Hypotheses

A prototype based on the JITAI concept was developed and extended with a learning agent to enable longer-term adaptation of interventions within the RELAX project. The prototype is equipped with a smartphone app that gathers both direct and indirect subjective stress data. Concurrently, a wearable wristband sensor is combined to monitor cardiovascular stress indicators. The information gathered from the sensor provides objective measurements of stress levels via HRV. The AI, trained on both personal and stress-related datasets, is responsible for determining the optimal timing for app-based interventions tailored to each user. Additionally, it selects the most appropriate intervention by content for the user at that specific moment, ensuring a personalized approach to stress management. This personalized selection of interventions was preceded by an intensive training phase in which users should evaluate guided access interventions from a pool of a plethora of interventions and thus express their needs and preferences to the AI. At the evaluation time, the training phase was ongoing, but the prototype had achieved the basic JITAI criteria: interventions were triggered by real-time stress, and content could be adaptively selected. More information of the architecture of the app and the personalization process of interventions is described in the study protocol of this study [38].

This study is an analysis of the prototype's developmental status. It was conducted as a pilot study to test the potential for stress reduction through the app and its feasibility in the everyday work situations of employees. The following research hypotheses were formulated for this purpose:

- Hypothesis 1: A decrease in stress levels in comparison to the participant's baseline can be observed following the initiation of app use. It is expected that the developed prototype will join the existing effective mHealth applications for stress management. With its structured interventions and diagnostic wearable, it combines proven features [23] and extends them with a JITAI component. The JITAI approach is anticipated to further influence stress by delivering more targeted interventions than prior methods.
- Hypothesis 2: A greater degree of personalization of the app is associated with a greater observed reduction in stress. The benefits of personalization have already been described in previous research on stress management solutions [39], as well as on digital health solutions in general [40], or on behavior change apps in general [41]. Personalization, as the core of the JITAI concept, can have different degrees, thereby giving rise to the question of which nuances are decisive for effectiveness gains. There is evidence of a consistent positive correlation between the degree of personalization and the desired effect [42,43]. However, the exact influence of different forms of personalization is still largely unknown [44].

In addition to the two hypotheses on preliminary app effects, this study assesses the usability of the system. User experience is crucial for mHealth effectiveness [45]. App efficacy hinges on users' willingness to use it as intended. User concerns, particularly in the domain of health monitoring, encompass a range of issues, including data protection [46], digital stress [47], accessibility, and user control [48]. Prior mHealth work has shown mixed usability: some reported satisfaction [37] and continued use [49], while others faced technical challenges that impaired the user experience [19]. Given these concerns, the study addresses usability and includes supplementary analyses for a broader understanding.

Methods

Study Design

The design of this study is described in detail in the corresponding study protocol [38]. The experimental design entailed a 2-arm trial comprising two randomly assigned experimental groups. Participants were required to use the RELAX project app in their naturalistic working environment. The study comprised two phases of 3 weeks each (phase 1 and phase 2). In phase 2, participants were randomly assigned to two distinct levels of personalization of the app (*high* vs *low*). Time series or continuous measurements were obtained via the app and the sensor in the study phases as well as in study-related surveys before the study (T1), between the study phases (T2), and after the study (T3). During the study phases, the participants had access to personalized interventions in the app.

Recruitment was carried out as described in a study by Kunas et al [38]. The ad hoc sample was recruited through previously committed companies and by social media and public flyers. Employees who met the inclusion criteria were given access to the study infrastructure. Adherence during the study was supported by regular reminders via email, the study website, and messages on the app dashboard. Randomization to the experimental groups from phase 2 onward was blinded, as was the evaluation. The allocation of participants and the evaluation process were conducted by different parts of the study team. The participants were randomly assigned to the groups using a pseudo-random function implemented with the NumPy package in the Python programming language. Following the conclusion of the study, participants were given access to the alternative version of personalization.

Ethical Considerations

The study was ethically approved (EK-GZ 25/2023) and preregistered prior to data analysis [50]. Registration was

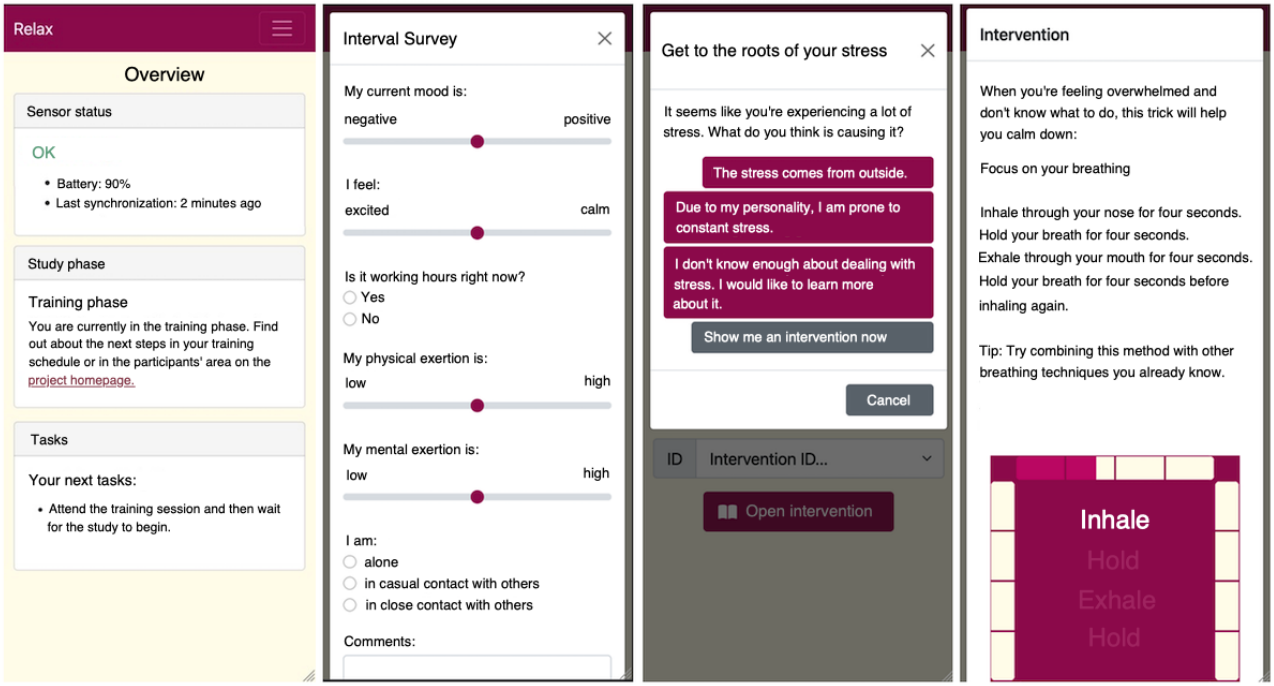
XSL•FO
RenderX

completed after data collection because modifications to the study design and the prototype were anticipated as the study progressed. The final protocol provides a detailed account of the study design and the prototype architecture and documents all changes, including, for example, the app update implemented in the second week to resolve technical issues identified during the first week. Participants gave their informed consent and were informed that they could opt out of participation at any time without facing any negative consequences. Their consent covered data analyses, especially after anonymization, including secondary analyses. The collected data were pseudonymized with a code for the duration of the study for the purpose of facilitating the identification of the individual participant in the event of technical issues and for the purpose of determining the compensation. Following the completion of the study, all data were anonymized. Upon completion of the study, participants were remunerated with a financial incentive of €80 (US $93.7). Additionally, they were permitted to retain the study sensor at their discretion.

Intervention

The intervention involved the use of the RELAX app in conjunction with the associated sensor. Further details of the app are described in the study protocol [38]. Figure 1 illustrates the user interface of the app in the English translation; the German original of the screen texts is available in the online supplementary materials in Multimedia Appendix 1.

Figure 1. User interface of the RELAX app in English translation, from left to right: dashboard, ecological momentary assessment questionnaire, first level of decision tree, and guided breathing intervention.



Participants were able to select the self-report questionnaires via the app. The questionnaires had to be completed on a daily basis in the morning and evening in both study phases, as well as several times during the day. The data from the questionnaires served as a measurement tool for the outcomes of the study and as input for the learning agent in the background of the app. In instances where the learning agent identified elevated levels of stress, automated access to a decision tree was initiated, culminating in the provision of suitable microinterventions. In the context of the physiological stress event, personalized current threshold value was calculated for the user (top 25% of values for the quotient of low frequency (LF) vs high frequency (HF), but at least 1.2; bottom 25% of values for the square root of squared mean value of the sum of the differences of successive R-R intervals, but maximum 50 ms; see [38] for detailed description of how the learning agent operated). In the event that both values exceeded or fell below the threshold values in two consecutive 5-minute windows, a stress event was annotated. The subjective triggers remained unaffected by personalization, and the threshold was on the valence scales only, with the mood being ≤3 on the Likert scale (1=negative and 7=positive).

For each branch of the decision tree, the app could recommend several microinterventions from a pool of 179 possible microinterventions for stress reduction. As described in the study protocol [38], the areas of microinterventions included several subareas of CBT, such as psychoeducation, training, or relaxation, which varied in the presentation format. Preset weightings indicated a probability of which microintervention would be triggered in an individual case (see [38] for the distribution of initial weightings). These weightings could be influenced by the user giving feedback ratings after completing the microintervention, which allowed the overall app experience to be personalized over the period of use.

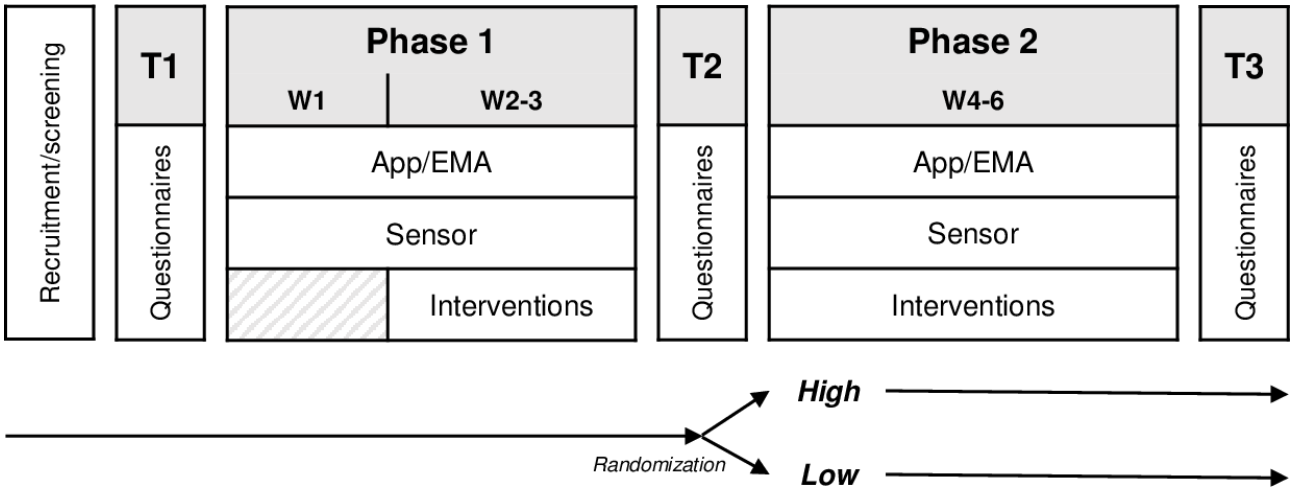
In phase 2, the versions of the app in both experimental groups differed in terms of the extent to which feedback ratings for the microinterventions received were incorporated. The version with *high* personalization modified the weighting of the probabilities for the occurrence of the individual

XSL•FO
RenderX

microintervention according to the individual ratings. In contrast, the version with *low* personalization was influenced by the aggregate ratings of the entire group of participants, indicating that no personalization per se was facilitated by the individual user. In phase 1, all participants received the same form of personalization according to the *low* group.

The sensor used in this study was the Polar Verity Sense, a wristband that uses photoplethysmography to measure various HRV parameters. Participants were required to wear the sensor continuously throughout the two study phases, with the exception of a 1-hour break each day. The passive nature of the recording process needed minimal involvement on the part of the participants, who were only required to ensure that the sensor was correctly connected to the app. The sequence of events in the study is presented in Figure 2, with the six study weeks abbreviated as W1-6.

Figure 2. Timeline of the study. EMA: ecological momentary assessment.



Outcomes

The demographic data were collected at the stage of screening for recruitment prior to the initiation of the study. The main outcomes evaluated in this study are derived from the primary outcomes of the corresponding study protocol (see Kunas et al [38] for measurement times and instruments). The collection of outcome data was facilitated by the use of standardized questionnaires and items developed by the study group via the app and online surveys for subjective data and the sensor for physiological data.

First, stress was measured in chronic form using the items of the daily ecological momentary assessment (EMA), which was created by the study team, as well as in work-related form using the standardized Valencia Eustress-Distress Appraisal Scale (VEDAS) [51] before the start of the study, between the study phases, and after the end of the study. A distinction was made between work-related distress and eustress. Stress was also measured using HRV parameters. These included the SD of R-R intervals, square root of squared mean value of the sum of the differences of successive R-R intervals, and LF/HF, which were measured continuously via the sensor. Further subjective variables encompassed the emotional states of arousal and valence, recorded at the time of entry and regarding the most recent stress event via EMA. Finally, satisfaction with the user experience and the usability of the system were measured using the standardized self-assessment Customer Satisfaction Score [52] and the System Usability Scale [53].

Statistical Analysis

The data were preprocessed according to the study protocol [38], including generating weekly data for the time series stress data. Datasets with more than 75% missing values were excluded from the analysis, and the procedure was therefore stricter than initially anticipated [38]. To measure the effect of the intervention on stress reduction, pre-post analyses were conducted in a within-subject design with the stress-related and emotional outcomes. Usability and satisfaction were compared separately between T2 and T3. For the between-subject analysis, the *high* versus *low* groups were included in all analyses. The System Usability Scale scores were converted to the standard 0 to 100 scale.

Corresponding linear mixed models (LMMs) were calculated for the analyses, with measurement times and groups as fixed factors and stress levels and usability variables, as measured by the different instruments or indicators, as dependent variables. Post hoc tests with Bonferroni correction were included for the detailed differentiation between measurement times. An α of .05 was used as the significance level, and partial eta-squared (η^2_p) indicated the effect size for the LMMs. To control for the false discovery rate, the Benjamini-Hochberg correction was applied. Only the adjusted *P* values are reported.

Results

Sample Characteristics

A total of 46 employees participated in the study after four people dropped out in the first 2 weeks of the study. The sample size exceeded the target of 31 participants, as determined by the a priori power analysis for feasibility estimation [38]. The sample consisted of 25 women (54%), 20 men (44%), and 1 diverse person (2%), with an average age of 40 (SD 9.71, range 24-62) years. The majority of the participants were Austrians (n=36, 78%), and the rest were Germans (n=10, 22%). In terms of occupation, the service sector was most strongly represented

XSL•FO
RenderX

in the sample with 17 employees (37%), followed by society, sport, or cultural sectors with 8 employees (17%) and the industrial sector with 6 employees (13%). In terms of educational level, the majority of respondents (n=20, 44%) had a postgraduate degree, followed by a university degree (20%) and a high school degree with 9 people each (20%).

On average, participants received 10.24 interventions (SD 7.41). There was no difference in the number of interventions in phase 2 between the groups ($t_{44}=-1.24$; $P=.22$). The sample sizes of the individual LMM analyses varied due to missing values, a consequence of the quality criterion on the collected data. This resulted in sample sizes ranging from 22 to 46 in the analyses. The proportion of data from the subjective questionnaires relative to the total possible data varied, on average, between 85% and 100%, depending on the study phase and parameters, and between 45% and 60% for the sensor data. Figure 3 shows the study's flow diagram, and Table 1 displays the means and SDs of the analyzed outcomes, arranged by time point or study group.

XSL•FO
RenderX

Figure 3. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the study.

XSL•FO
RenderX

Table . Study variables.

Variable	Time of measurement, mean (SD)									Groups, mean (SD)	
	T1[a]	W1[b]	W2[b]	W3[b]	T2[a]	W4[b]	W5[b]	W6[b]	T3[a]	High	Low
Work-related distress	3.61 (0.95)	—[c]	—	—	3.64 (0.78)	—	—	—	3.74 (0.65)	3.72 (0.72)	3.60 (0.87)
Work-related eustress	3.18 (0.68)	—	—	—	3.17 (0.74)	—	—	—	3.23 (0.71)	3.26 (0.74)	3.13 (0.69)
Daily chronic stress	—	3.13 (0.54)	3.07 (0.57)	3.03 (0.63)	—	3.10 (0.60)	2.90 (0.60)	2.92 (0.56)	—	3.13 (0.63)	2.91 (0.51)
Arousal (momentary)	—	3.33 (0.89)	3.29 (0.77)	3.30 (1.02)	—	3.15 (0.99)	3.12 (1.12)	3.17 (1.06)	—	3.27 (1.03)	3.18 (0.92)
Valence (momentary)	—	5.40 (0.72)	5.23 (0.68)	5.45 (0.85)	—	5.23 (0.91)	5.39 (0.89)	5.45 (0.79)	—	5.13 (0.77)	5.41 (0.87)
Arousal (most recent stress event)	—	4.38 (0.68)	4.37 (0.59)	4.28 (0.85)	—	4.13 (0.63)	4.05 (0.65)	4.03 (0.40)	—	4.23 (0.76)	4.17 (0.43)
Valence (most recent stress event)	—	4.14 (0.61)	4.06 (0.45)	4.32 (0.64)	—	4.07 (0.55)	4.05 (0.52)	4.06 (0.49)	—	4.16 (0.67)	4.07 (0.31)
SDRR[d]	—	19.90 (5.02)	19.18 (4.34)	20.62 (4.38)	—	19.03 (5.15)	19.45 (4.16)	19.41 (4.14)	—	18.66 (4.31)	20.46 (4.70)
RMSSD[e]	—	21.12 (5.60)	20.34 (4.76)	22.01 (4.81)	—	19.99 (5.74)	20.73 (4.81)	20.59 (4.44)	—	19.98 (4.66)	21.54 (5.37)
LF/HF[f]	—	1.52 (0.24)	1.60 (0.43)	1.58 (0.32)	—	1.69 (0.48)	1.65 (0.52)	1.80 (0.60)	—	1.67 (0.46)	1.61 (0.41)
Usability	—	—	—	—	37.35 (8.97)	—	—	—	46.85 (9.21)	43.26 (8.75)	41.03 (9.39)
Satisfaction	—	—	—	—	2.57 (0.64)	—	—	—	2.38 (0.76)	2.66 (0.68)	2.28 (0.68)

[a]T1-3: measurement times.

[b]W1-6: study weeks.

[c]Not applicable.

[d]SDRR: SD of R-R intervals.

[e]RMSSD: square root of squared mean value of the sum of the differences of successive R-R intervals.

[f]LF/HF: quotient of low versus high frequency.

Stress Reduction (H_1)

The overall picture reveals only isolated significant results. There was a decrease in daily chronic stress over the course of the study ($F_{5,165}$=2.89; P=.048; η^2_p=0.081). Post hoc analyses revealed no significant differences between weeks; however, there was a tendency between W1 and W5 (P=.06). Furthermore, there was a decline in the self-reported arousal of the most recent stress event, as reported from the time of app entry, that was identified over the course of the study period ($F_{5,150}$=4.52; P=.02; η^2_p=0.131). Post hoc tests revealed several effects between the weeks: Specifically, these included comparisons between W1 and W5 (P=.02), W1 and W6 (P=.02), W2 and W5 (P=.04), and W2 and W6 (P=.03). However, no significant effects were observed in other subjective parameters, such as work-related stress experience or valence.

The sensor data revealed that only one parameter exhibited significant effects during the study period, with LF/HF increasing over the weeks ($F_{5,109.01}$=3.91; P=.03; η^2_p=0.152). According to the definition of the parameter, this can be interpreted as an increase in sympathetic activity, which indicates increased physiological stress. However, none of the other HRV parameters exhibited a similar pattern. The results of the LMMs are displayed in Table 2.

XSL•FO
RenderX

Table . Linear mixed models of the study variables.

Variable	Within-subjects analysis			Between-subjects analysis		
	F (df)	*P* value	η^2_p (95% CI)	*F* (df)	*P* value	η^2_p (95% CI)
Work-related distress	0.77 (2, 90)	.696	0.017 (0.000-0.085)	0.37 (1, 44)	.73	0.008 (0.000-0.124)
Work-related eustress	0.22 (2, 90)	.80	0.005 (0.000-0.047)	0.59 (1, 44)	.67	0.013 (0.000-0.139)
Daily chronic stress	2.89 (5, 165)	.048	0.081 (0.003-0.144)	1.78 (1, 32)	.38	0.053 (0.000-0.244)
Arousal (momentary)	1.24 (5, 150)	.50	0.040 (0.000-0.085)	0.06 (1, 29)	.80	0.002 (0.000-0.118)
Valence (momentary)	2.13 (5, 150)	.17	0.066 (0.000-0.126)	0.13 (1, 29)	.81	0.005 (0.000-0.140)
Arousal (most recent stress event)	4.52 (5, 150)	.02	.131 (0.027-0.209)	0.11 (1, 29)	.81	0.004 (0.000-0.135)
Valence (most recent stress event)	1.96 (5, 150)	.19	0.061 (0.000-0.119)	0.44 (1, 29)	.73	0.15 (0.000-0.181)
SDRR[a]	1.62 (5, 110)	.26	0.069 (0.000-0.135)	1.13 (1, 21)	.50	0.051 (0.000-0.286)
RMSSD[b]	1.84 (5, 110)	.31	0.077 (0.000-0.148)	0.68 (1, 21)	.56	0.031 (0.000-0.253)
LF/HF[c]	3.91 (5, 109.01)	.03	0.152 (0.022-0.243)	0.32 (1, 20.99)	.70	0.015 (0.000-0.215)
Usability	24.80 (1, 45)	<.001	0.355 (0.138-0.522)	1.39 (1, 44)	.38	0.031 (0.000-0.177)
Satisfaction	7.12 (1, 45)	.05	0.137 (0.008-0.319)	4.03 (1, 44)	.12	0.084 (0.000-0.258)

[a]SDRR: SD of R-R intervals.

[b]RMSSD: square root of squared mean value of the sum of the differences of successive R-R intervals.

[c]LF/HF: quotient of low versus high frequency.

Group Differences (H_2)

No significant group effects were identified.

Usability and Satisfaction Explorations

With regard to the usability, an increase in positive feedback from the participants was observed between the middle and the end of the study ($F_{1,45}$=24.80; P<.001; η^2_p=0.355). A tendency of an effect also emerged in the satisfaction data, which, however, showed a significant decrease between the two measurement points ($F_{1,45}$=7.12; P=.05; η^2_p=0.137). Technical issues occurred during the study, which inevitably disrupted measurements, particularly with regard to the sensor data. These disruptions primarily affected the connection between the sensor and the app in the initial weeks of the study. Furthermore, there have also been reports of discomfort caused by the daily demands of using the app and the sensor. Regarding the latter, the battery life was criticized.

Discussion

Principal Findings

This pilot study on the mobile app for occupational stress management developed in the RELAX project, according to the JITAI concept, has taken on the first evaluation of this app. The study aimed to determine the preliminary effectiveness of the app in terms of stress reduction during the study period and whether the degree of personalization of the stress interventions modifies the influence of the app. In addition to the indications of impact, the study also sought to examine the usability of the app in order to assess its transfer to a real-world context. The study's findings, overall, were rather mixed in all respects.

Evidence of Stress Reduction

Hypothesis 1, which predicted a reduction in stress levels by the end of the study compared to the baseline, could only be proven for certain stress indicators. Specifically, the subjective chronic stress measured daily and the subjective arousal of the

XSL•FO
RenderX

most recent stress event in the daily EMA entries demonstrated a reduction in stress between given previous weeks and subsequent weeks. The substantial reduction in these stress variables occurred in the fifth and sixth week after the study began. These partial results align with prior research on stress management through mHealth [11,23] and offer preliminary indications of the app architecture's effectiveness for the pilot study.

The absence of effects on other subjective measures of stress may be explained by the nature of the variables collected. The stress markers collected at untriggered, that is, self-selected measurement points of EMA entries, could contain a bias, where participants deliberately choose the EMA time points when they are not stressed. In the recommended EMA paradigm, the temporal parameters of measurement are typically prespecified by the researcher [54]. However, in this study, the subjects were permitted to select the temporal parameters of their observations autonomously, thereby allowing for the bias. Conversely, the question regarding the intensity of the most recent stress event in these EMA surveys specifically asks about stress situations, thereby ensuring that similar negative moments are recorded in this category for direct comparison. In contrast, the actual momentary experiences at EMA entry time could offer a much wider range of possible experiences in the moment, which might not provide clear direction when aggregated.

As indicated by prior studies [55], the manner in which stress is assessed using EMA has been found to influence the frequency with which subjects disclose their stress levels. Consequently, using explicit items regarding the most recent stressful event has been shown to yield more dependable data concerning stress compared to momentary inputs that are not based on events. The same rationale can be applied to the valence measurements. At the time of EMA entry, the valence may have been without clear direction in the spectrum of potential valences. It is noteworthy that the negative valence of the most recent stress event has remained constant, despite a decline in arousal levels during the most recent stress moment. According to the findings of research conducted on mood, the arousal level is more volatile and susceptible to interventions than the valence level [56]. However, empirical studies on stress management with just-in-time interventions show that the valence of an affect can be influenced after only a few weeks [57,58], so this aspect requires further consideration.

The absence of change in work-related stress prior to and following the study necessitates an alternative explanation for the self-reports. One possibility is that the instrument used to assess these constructs, the VEDAS [51], was predominantly influenced by the content of specific work-related factors. The items in the questionnaire address specific scenarios that have been identified as potential triggers for stress, such as the pressure to take work home or to extend one's working hours. These factors might not have been adequately addressed by the interventions within the short time frame of this study. Despite the high degree of personalization of the interventions, it is possible that there was an absence of an appropriate intervention for every possible scenario. For instance, interventions involving broadband techniques, such as different types of relaxation or skills training, may not be immediately effective for a person in a highly specific scenario. A preponderance of extant literature already recommends that factors such as the individual, the context, and the content must be considered equally in a potential intervention if that intervention is to be truly effective [59]. The decision tree, with predefined structures for individual, context, and content factors, restricts this aspect. In this context, the validity of an instrument such as the VEDAS should be reconsidered, as the person-centered approach to interventions and the situation-based approach to this instrument may not be sufficiently aligned.

The physiological indicators demonstrated a contrary effect to the expected subjective stress reduction. In this case, the only observed effect in the change of stress levels was manifested as an increase in the HRV parameter of the LF/HF. In the context of the relationship between subjective and physiological stress markers, it is essential to consider the complex interplay between these two parameter classes. A recent review of the correspondence between daily stress parameters at different levels demonstrated that only approximately half of the studies examined exhibited an association between the two indicator classes [60]. It is also conceivable that the duration of the study was insufficient to elicit a discernible effect on the markers of HRV, which are generally regarded as being relatively stable. Chronic stress, as has been previously discussed in analogous studies, may not have exerted a perceptible influence due to its inherent stability [19].

The validity of the LF/HF ratio as a stress marker is also debated in the literature [61]. When derived from real-life conditions in long-term measurements, such as from everyday wearables, it appears to be an unreliable indicator [62]. In this study, LF/HF was the only HRV parameter to show a significant change, despite expectations of corresponding changes across the multiple HRV indices [63]. Taken together, these observations suggest that the physiological recordings in this study should be interpreted with caution and that their interpretive value may be limited. In addition, the shortcomings of sensor measurement, which in some cases miss up to 75% of the possible data due to limited compliance or technical challenges, must be addressed. Comparable studies in the health care sector have shown adherence rates of more than 50% for the majority of subjects, with a maximum of 95% of data achieved relative to the total possible data in the study periods [64].

Although there were promising changes in selected subjective stress markers that could indicate an accumulated intervention-related effect in the later weeks of the study, the possibility must also be considered that, given the lack of significant changes in other parameters, the overall finding is negligible, and the app was not able to reduce stress as intended. The only two stress indicators that demonstrated a reduction during the study period had to be assessed retrospectively on a daily basis. The arousal of the most recent stress event included the query of a period of several hours before the entry, and the daily subjective chronic stress required the assessment of the entire past day with regard to various stress dimensions. Consequently, despite the implementation of the EMA method, recall bias could affect the validity of the results, as the experience in question could not be measured precisely in real time [65]. As Shiffman [54] previously demonstrated, recall

XSL•FO
RenderX

bias includes the tendency to handle negative and positive experiences differently. Therefore, there may have been a tendency to recall prompted negative experiences in a biased manner, resulting in systematic results that may not reflect reality.

Analysis of Personalization and Usability

In consideration of the degree of personalization according to H_2, it was observed that both levels of personalization provided by the algorithm, high and low, contributed equally to the changes in stress levels or the absence of such changes. It is important to note again that the duration of the experiment may have been too short to fully manifest the potential of the high degree version of personalization. The necessity for the algorithm to undergo further training following the splitting into two groups may indicate that the difference in user experience was not yet substantial enough to be reflected in differences in the stress scores. The necessity of navigating the decision tree for each intervention, which serves to narrow down potentially suitable interventions on an individual basis, underscores the fact that the user had still a considerable degree of influence on the system at this particular stage of development. Consequently, the automatic selection of interventions by the app may have been marginalized. As an alternative possibility, the extensive personalization options provided by a selection of nearly 200 interventions, which may have already exceeded the threshold for a discernible effect, could have resulted in a ceiling effect, thereby preventing a potential group effect. Comparable research, which focused on an expanded range of potential interventions, has previously discussed the simplification of the app again, a discussion that occurred subsequent to their analysis of the degree of personalization and the corresponding results [19].

Comparable research so far shows that multilevel approaches to stress management, in which several levels of data are considered dynamically, show a high variability in effectiveness [23]. In this context, the persuasiveness of an intervention must always be considered [66] because this study does not show whether the participants performed the interventions at all and, if so, correctly, because they were unguided. Given the early stage of the research, it is to be expected that the desired effects cannot yet be clearly identified. At this stage of development, not only the effectiveness factors may not be fully developed, but the usability factors are also only explored and optimized in several iterations of evaluation studies. In the present case, this was an initial evaluation, which means that it will remain unclear what proportion of the pattern of results is actually attributable to effectiveness factors and what proportion to usability factors.

An increase in usability was observed over the course of the study, as shown by the self-report data between the study phases and at the end of the study. The adaptation of the app during the study period, particularly subsequent to the initial phase, in alignment with the expressed preferences of users, can be regarded as a notable improvement. Concurrently, there were indications of a decline in satisfaction at the trend level. This finding suggests a contradictory interpretation: while users initially perceived the system as more usable, their preference for it decreased over time. During the course of the study, personalization mechanisms were integrated into the background of the app to tailor interventions to individual stress levels. This was anticipated to improve the user experience and satisfaction. The potential for the daily use of the app to induce exhaustion or discomfort could offer an alternative explanation for the observed outcomes. The technical modifications implemented in the application may have promoted its use; nevertheless, the intrinsic characteristics of its daily use may have resulted in an escalation in discomfort over time. In the context of mHealth for stress management, the consideration of digital stress is imperative, as it can readily engender an effect that is contrary to the intended objective [47].

The adherence rates indicate an overall rather low level of engagement in the daily study activities. Conversely, the relatively high incidence of subjective data indicated the general willingness of the participants, suggesting that the low sensor rates are most likely attributable to technical barriers. Consequently, there is an ongoing need for technical enhancements to ensure a seamless user experience and optimize functioning. The absence of seamless functionality in a digital system can significantly compromise user experience, potentially impeding the system's intended functionality [45]. Consequently, this study sought to illustrate the potential of the JITAI concept through the use of the developed app on a system yet to be optimized.

Strengths and Limitations

The present evaluation of the RELAX app demonstrates the initial results of a mHealth approach to the practical implementation of the JITAI concept in the domain of occupational stress management. The application's objective is to deliver interventions that are not only time efficient but also content focused, with the aim of optimally addressing the individual's situation and needs. This approach builds on earlier methods that either limited themselves to the compilation of diagnostic data or offered generic, one-size-fits-all interventions [11-13]. A notable aspect of the study is the remarkably low rate of attrition, with only 4 participants (8% of the total sample) dropping out, which is particularly noteworthy given the intensive nature of the repeated daily EMAs and the wearable monitoring process that included nocturnal recordings. Dropout rates from other mHealth studies on chronic diseases exhibited comparatively higher percentages, ranging from 40% to 49% [67]. Despite the presence of incomplete data recordings, adherence remained consistent, underscoring the necessity for such interventions in employees' daily lives.

The assessment of multiple levels of stress markers made it possible to compare them, whereas previous studies of comparable approaches have focused on only a few parameters [13,23]. This approach allowed for the identification of variability in the effects across dimensions, which, although not in the expected manner, may provide valuable insights from alternative perspectives. It is acknowledged that stress markers do not always align, particularly between physiological and subjective parameters [68]. The findings from this study lay the foundation for future research to dive deeper into this divergence

XSL•FO
RenderX

and potentially link it to underlying factors within the specific mHealth context.

This study fulfills the quality requirements of a pilot study. However, it must therefore also be evaluated with the limited significance of such a study. The absence of a control group without intervention, as previously identified as a limitation in the study protocol [38], prevents clear attribution of all changes in stress parameters due to the intervention. The study design used here does not provide the internal validity required for statements on the effectiveness of the intervention. At this juncture, the outcomes of this study can only be interpreted as effects that occurred during the study period, which may provide an indication of effectiveness that must be validated with a controlled follow-up study.

Furthermore, as previously stated in the discussion, the study period was relatively brief, thereby limiting the opportunities for user intervention. The small sample size precluded the ability to draw definitive conclusions; thus, the findings should be considered exploratory in nature. This is particularly evident in the group randomization, where the sample sizes were further diminished. Following the implementation of technical stabilization updates to the application and the completion of its development phase, which will enable fully automated intervention selection, the evaluation should be expanded to include a more substantial sample size and an extended study duration.

Moreover, the employees' daily experiences in their highly individual roles and in different industries are likely to have generated such a variable set of situations that the stress events within and between individuals call into question the comparability of the EMA data. The app's capacity to adapt to each unique situation and individual user could be considered overly ambitious for the current stage of development, particularly in light of the absence of complete automation in intervention selection at the time of evaluation. Consequently, users retained a substantial degree of active involvement and control over the app's functionality. To enhance the validity and generalizability of the findings, it is recommended that future evaluations use more mature app systems that prioritize the comparability of the sample and incorporate control groups.

Conclusions

In summary, the study preliminarily examined the impact of an mHealth app developed in accordance with the JITAI concept. The findings were inconsistent, with only two subjective stress markers demonstrating the hypothesized stress reduction, while one physiological parameter exhibited a debatable increase. The validity of the attribution of the observed effects to the performance of the app is not without question. It is conceivable that these effects are, in fact, an artifact of a system that has not yet been fully developed, if the effects are at all attributable to the system. These findings give rise to questions concerning the extent of the app's limited usability in its present developmental stage, such as the ease of use of the app system, as well as the integration of generally consistent daily mHealth use in the work context. Nevertheless, the study identified several promising starting points for the future development of a functional and effective app that can address occupational mental stress in an adaptive manner in terms of time and content.

Acknowledgments

We would like to thank all participants for their contribution to the research and University Prof Dr Frank Wilhelm for his valuable input in the analysis of the data. The translation into English was supported by DeepL. Generative artificial intelligence was not used.

Funding

The study received funding from the Austrian Research Promotion Agency. This open access publication was supported by the Publication Fund of Paris Lodron University of Salzburg.

Data Availability

The datasets generated and analyzed during the study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization – BK (lead), ARL (equal), OJ (supporting)
Data curation – JM (lead), BK (equal)
Formal analysis – BK (lead), JM (supporting)
Funding acquisition – OJ (lead), ARL (equal)
Investigation – BK (lead), JM (supporting), ARL (supporting)
Methodology – BK (lead), ARL (supporting)
Project administration – ARL (supporting), BK (equal), OF (supporting)
Resources – OJ (lead), BK (equal), JM (equal), CS (equal), MS-T (equal)
Software – OJ (lead), CS (equal), MS-T (equal)
Supervision – OJ (lead), ARL (equal)
Validation – BK (lead), ARL (equal)
Visualization – BK (lead), OJ (supporting)

XSL•FO
RenderX

Writing – original draft – BK
Writing – review & editing – ARL (lead), BK (equal), OJ (supporting), CS (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
User interface of the RELAX app in German, from left to right: dashboard, ecological momentary assessment questionnaire, first level of decision tree, and guided breathing intervention.
[PNG File, 336 KB - mhealth_v14i1e79642_app1.png]

References

1. Marzano L, Bardill A, Fields B, et al. The application of mHealth to mental health: opportunities and challenges. Lancet Psychiatry 2015 Oct;2(10):942-948. [doi: 10.1016/S2215-0366(15)00268-0]
2. van de Vijver S, Tensen P, Asiki G, et al. Digital health for all: how digital health could reduce inequality and increase universal health coverage. Digit Health 2023;9:20552076231185434. [doi: 10.1177/20552076231185434] [Medline: 37434727]
3. Shah IA, Jhanjhi NZ, Humayun M, Ghosh U, et al. Health care digital revolution during COVID-19. In: Anandan R, Suseendran G, Chatterjee P, Jhanjhi NZ, Ghosh U, editors. How COVID-19 Is Accelerating the Digital Revolution: Challenges and Opportunities: Springer International Publishing; 2022:17-30. [doi: 10.1007/978-3-030-98167-9_2]
4. Tandon A, Kaur P, Ruparel N, Islam JU, Dhir A. Cyberloafing and cyberslacking in the workplace: systematic literature review of past achievements and future promises. INTR 2022 Jan 18;32(1):55-89. [doi: 10.1108/INTR-06-2020-0332]
5. Kavosi Z, Ranaei Kordshouli H, Zare Saadabadi M, Ghorbanian A. The factors related to the employees' job stress: a meta-analysis. J Health Manag Inform 2018;5(4):145-151 [FREE Full text]
6. Steed LB, Swider BW, Keem S, Liu JT. Leaving work at work: a meta-analysis on employee recovery from work. J Manage 2021 Apr;47(4):867-897. [doi: 10.1177/0149206319864153]
7. Guthier C, Dormann C, Voelkle MC. Reciprocal effects between job stressors and burnout: a continuous time meta-analysis of longitudinal studies. Psychol Bull 2020 Dec;146(12):1146-1173. [doi: 10.1037/bul0000304] [Medline: 33119345]
8. Nixon AE, Mazzola JJ, Bauer J, Krueger JR, Spector PE. Can work make you sick? A meta-analysis of the relationships between job stressors and physical symptoms. Work & Stress 2011 Jan;25(1):1-22. [doi: 10.1080/02678373.2011.569175]
9. van der Molen HF, Nieuwenhuijsen K, Frings-Dresen MHW, de Groene G. Work-related psychosocial risk factors for stress-related mental disorders: an updated systematic review and meta-analysis. BMJ Open 2020 Jul 5;10(7):e034849. [doi: 10.1136/bmjopen-2019-034849] [Medline: 32624469]
10. Nekoei A, Sigurdsson J, Wehr D. CESifo working paper: the economic burden of burnout. : CESifo GmbH; 2024. [doi: 10.2139/ssrn.4827359]
11. Egger SM, Frey S, Sauerzopf L, Meidert U. A literature review to identify effective web- and app-based mHealth interventions for stress management at work. Workplace Health Saf 2023 Oct;71(10):452-463. [doi: 10.1177/21650799231170872] [Medline: 37254448]
12. Blázquez Martín D, De La Torre I, Garcia-Zapirain B, Lopez-Coronado M, Rodrigues J. Managing and controlling stress using mHealth: systematic search in app stores. JMIR Mhealth Uhealth 2018;6(5):e111. [doi: 10.2196/mhealth.8866]
13. Stratton E, Jones N, Peters SE, Torous J, Glozier N. Digital mHealth interventions for employees: systematic review and meta-analysis of their effects on workplace outcomes. J Occup Environ Med 2021 Aug 1;63(8):e512-e525. [doi: 10.1097/JOM.0000000000002267] [Medline: 34050095]
14. Barkham M, Lambert MJ. The efficacy and effectiveness of psychological therapies. In: Barkham M, Lutz W, Castonguay LG, editors. Bergin and Garfield's Handbook of Psychotherapy and Behavior Change: Wiley & Sons; 2021:135-189.
15. Wang L, Miller LC. Just-in-the-moment adaptive interventions (JITAI): a meta-analytical review. Health Commun 2020 Nov;35(12):1531-1544. [doi: 10.1080/10410236.2019.1652388] [Medline: 31488002]
16. Müller AM, Blandford A, Yardley L. The conceptualization of a just-in-time adaptive intervention (JITAI) for the reduction of sedentary behavior in older adults. Mhealth 2017;3:37. [doi: 10.21037/mhealth.2017.08.05] [Medline: 29184889]
17. Ara A, Mifa AF. Integrating artificial intelligence and big data in mobile health: a systematic review of innovations and challenges in healthcare systems. JALHE 2024;3(1):1-16. [doi: 10.62304/jbedpm.v3i01.70]
18. Bhatt P, Liu J, Gong Y, Wang J, Guo Y. Emerging artificial intelligence-empowered mHealth: scoping review. JMIR Mhealth Uhealth 2022 Jun 9;10(6):e35053. [doi: 10.2196/35053] [Medline: 35679107]
19. Baumann H, Heuel L, Bischoff LL, Wollesen B. Efficacy of individualized sensory-based mHealth interventions to improve distress coping in healthcare professionals: a multi-arm parallel-group randomized controlled trial. Sensors (Basel) 2023 Feb 19;23(4):2322. [doi: 10.3390/s23042322] [Medline: 36850920]

XSL•FO
RenderX

20. Rodriguez JM, Aso S, Cavero C, et al. Towards digital and personalized healthcare and well-being solutions for the workplace. In: Novais P, Chatzigiannakis I, Tobe Y, Amft O, editors. Intelligent Environments 2018: IOS Press; 2018:377-383.
21. Dugas M, Gao GG, Agarwal R. Unpacking mHealth interventions: a systematic review of behavior change techniques used in randomized controlled trials assessing mHealth effectiveness. Digit Health 2020;6:2055207620905411. [doi: 10.1177/2055207620905411] [Medline: 32128233]
22. Gosetto L, Ehrler F, Falquet G. Personalization dimensions for mHealth to improve behavior change: a scoping review. Stud Health Technol Inform 2020 Nov 23;275:77-81. [doi: 10.3233/SHTI200698] [Medline: 33227744]
23. Herold M, Simbula S, Gallucci M. Can smartphone applications and wearable technologies improve workplace well-being and help manage stress? A systematic review. Curr Psychol 2024 Sep;43(36):28650-28673. [doi: 10.1007/s12144-024-06534-z]
24. Morales A, Barbosa M, Morás L, et al. Occupational stress monitoring using biomarkers and smartwatches: a systematic review. Sensors (Basel) 2022 Sep 2;22(17):6633. [doi: 10.3390/s22176633] [Medline: 36081096]
25. Amor J, James C. Unobtrusive wearable technology for health monitoring. In: Khosrow-Pour M, editor. Wearable Technologies: Concepts, Methodologies, Tools, and Applications: IGI Global; 2018:562-577. [doi: 10.4018/978-1-5225-5484-4.ch025]
26. Han HJ, Labbaf S, Borelli JL, Dutt N, Rahmani AM. Objective stress monitoring based on wearable sensors in everyday settings. J Med Eng Technol 2020 May;44(4):177-189. [doi: 10.1080/03091902.2020.1759707] [Medline: 32589065]
27. Haque Y, Zawad RS, Rony CSA, et al. State-of-the-art of stress prediction from heart rate variability using artificial intelligence. Cogn Comput 2024 Mar;16(2):455-481. [doi: 10.1007/s12559-023-10200-0]
28. González Ramírez ML, García Vázquez JP, Rodríguez MD, Padilla-López LA, Galindo-Aldana GM, Cuevas-González D. Wearables for stress management: a scoping review. Healthcare (Basel) 2023 Aug 22;11(17):2369. [doi: 10.3390/healthcare11172369] [Medline: 37685403]
29. Jerath R, Syam M, Ahmed S. The future of stress management: integration of smartwatches and HRV technology. Sensors (Basel) 2023 Aug 22;23(17):7314. [doi: 10.3390/s23177314] [Medline: 37687769]
30. Chaudhry BM, Islam A, Matthieu M. Toward designs of workplace stress management mobile apps for frontline health workers during the COVID-19 pandemic and beyond: mixed methods qualitative study. JMIR Form Res 2022 Jan 13;6(1):e30640. [doi: 10.2196/30640] [Medline: 34806985]
31. Khakurel J, Melkas H, Porras J. Tapping into the wearable device revolution in the work environment: a systematic review. ITP 2018 May 23;31(3):791-818. [doi: 10.1108/ITP-03-2017-0076]
32. Lalanza JF, Lorente S, Bullich R, García C, Losilla JM, Capdevila L. Methods for heart rate variability biofeedback (HRVB): a systematic review and guidelines. Appl Psychophysiol Biofeedback 2023 Sep;48(3):275-297. [doi: 10.1007/s10484-023-09582-6] [Medline: 36917418]
33. Chelidoni O, Plans D, Ponzo S, Morelli D, Cropley M. Exploring the effects of a brief biofeedback breathing session delivered through the BioBase app in facilitating employee stress recovery: randomized experimental study. JMIR Mhealth Uhealth 2020 Oct 15;8(10):e19412. [doi: 10.2196/19412] [Medline: 33055072]
34. Rudics E, Nagy Á, Dombi J, et al. Photoplethysmograph based biofeedback for stress reduction under real-life conditions in healthcare frontline. Appl Sci (Basel) 2023;13(2):835. [doi: 10.3390/app13020835]
35. Clarke S, Jaimes LG, Labrador MA. MStress: a mobile recommender system for just-in-time interventions for stress. In: Clarke S, Jaimes LG, Labrador MA, editors. Presented at: 2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC); Jan 8-11, 2017; Las Vegas, NV p. 1-5. [doi: 10.1109/CCNC.2017.8015367]
36. Luštrek M, Lukan J, Bolliger L, editors. Designing an intervention against occupational stress based on ubiquitous stress and context detection. Presented at: Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing; Oct 8, 2023. [doi: 10.1145/3594739.3611326]
37. Winslow BD, Kwasinski R, Hullfish J, et al. Automated stress detection using mobile application and wearable sensors improves symptoms of mental health disorders in military personnel. Front Digit Health 2022;4:919626. [doi: 10.3389/fdgth.2022.919626] [Medline: 36082233]
38. Kunas B, Jung O, Schranz C, Schmoigl-Tonis M, Ploessnig M, Laireiter AR. Personalizing mHealth interventions for occupational stress: protocol for a randomized pilot study. JMIR Res Protoc 2025 Jun 3;14(1):e68012. [doi: 10.2196/68012] [Medline: 40460419]
39. Alhasani M, Mulchandani D, Oyebode O, Baghaei N, Orji R. A systematic and comparative review of behavior change strategies in stress management apps: opportunities for improvement. Front Public Health 2022;10:777567. [doi: 10.3389/fpubh.2022.777567] [Medline: 35284368]
40. Grua EM, De Sanctis M, Malavolta I, Hoogendoorn M, Lago P. An evaluation of the effectiveness of personalization and self-adaptation for e-Health apps. Inf Softw Technol 2022 Jun;146:106841. [doi: 10.1016/j.infsof.2022.106841]
41. Tong HL, Quiroz JC, Kocaballi AB, et al. Personalized mobile technologies for lifestyle behavior change: a systematic review, meta-analysis, and meta-regression. Prev Med 2021 Jul;148:106532. [doi: 10.1016/j.ypmed.2021.106532] [Medline: 33774008]

XSL•FO
RenderX

42. Wang J, Wu Z, Choi SW, et al. The dosing of mobile-based just-in-time adaptive self-management prompts for caregivers: preliminary findings from a pilot microrandomized study. JMIR Form Res 2023 Sep 14;7(1):e43099. [doi: 10.2196/43099] [Medline: 37707948]
43. Sartori F, Savi M, Talpini J. Tailoring mHealth apps on users to support behavior change interventions: conceptual and computational considerations. Appl Sci (Basel) 2022;12(8):3782. [doi: 10.3390/app12083782]
44. Kankanhalli A, Xia Q, Ai P, Zhao X, National University of Singapore. Understanding personalization for health behavior change applications: a review and future directions. THCI 2021;13(3):316-349. [doi: 10.17705/1thci.00152]
45. Deniz-Garcia A, Fabelo H, Rodriguez-Almeida AJ, et al. Quality, usability, and effectiveness of mHealth apps and the role of artificial intelligence: current scenario and challenges. J Med Internet Res 2023 May 4;25:e44030. [doi: 10.2196/44030] [Medline: 37140973]
46. Kerr JI, Naegelin M, Benk M, et al. Investigating employees' concerns and wishes regarding digital stress management interventions with value sensitive design: mixed methods study. J Med Internet Res 2023 Apr 13;25:e44131. [doi: 10.2196/44131] [Medline: 37052996]
47. Freytag A, Hefner D. Digital stress. In: Ho EY, Bylund CL, Weert JCM, Basnyat I, Bol N, Dean M, editors. The International Encyclopedia of Health Communication: John Wiley & Sons; 2023:1-6. [doi: 10.1002/9781119678816]
48. Nasr E, Alsaggaf W, Sinnari D. Developing usability guidelines for mHealth applications (UGmHA). MTI 2023;7(3):26. [doi: 10.3390/mti7030026]
49. Kawadler JM, Hemmings NR, Ponzo S, Morelli D, Bird G, Plans D. Effectiveness of a smartphone app (BioBase) for reducing anxiety and increasing mental well-being: pilot feasibility and acceptability study. JMIR Form Res 2020 Nov 10;4(11):e18067. [doi: 10.2196/18067] [Medline: 32969341]
50. Kunas B. A personalized m-health intervention for occupational stress using multi-level diagnostics - the evaluation of the RELAX approach. : OSF; 2024. [doi: 10.17605/OSF.IO/MYRD9]
51. Rodríguez I, Kozusznik MW, Peiró JM. Development and validation of the Valencia Eustress-Distress Appraisal Scale. Int J Stress Manag 2013;20(4):279-308. [doi: 10.1037/a0034330]
52. Hill N, Brierley J. How to Measure Customer Satisfaction: Routledge; 2017. [doi: 10.4324/9781315253107]
53. Jordan PW, Thomas B, McClelland IL, Weerdmeester B. Usability Evaluation in Industry: CRC Press; 1996.
54. Shiffman S. Designing protocols for ecological momentary assessment. In: Stone A, Shiffman S, Atienza A, editors. The Science of Real-Time Data Capture: Self-Reports in Health Research: Oxford: University Press; 2007:27-53. [doi: 10.1093/oso/9780195178715.003.0003]
55. Zawadzki MJ, Scott SB, Almeida DM, et al. Understanding stress reports in daily life: a coordinated analysis of factors associated with the frequency of reporting stress. J Behav Med 2019 Jun;42(3):545-560. [doi: 10.1007/s10865-018-00008-x] [Medline: 30600403]
56. Gomez P, Zimmermann PG, Guttormsen Schär S, Danuser B. Valence lasts longer than arousal: persistence of induced moods as assessed by psychophysiological measures. J Psychophysiol 2009;23(1):7-17. [doi: 10.1027/0269-8803.23.1.7]
57. Johnson JA, Zawadzki MJ, Materia FT, White AC, Smyth JM. Efficacy and acceptability of digital stress management micro-interventions. Procedia Comput Sci 2022;206:45-55. [doi: 10.1016/j.procs.2022.09.084]
58. Smyth JM, Heron KE. Is providing mobile interventions "just-in-time" helpful? An experimental proof of concept study of just-in-time intervention for stress management. Presented at: 2016 IEEE Wireless Health (WH); Oct 25-27, 2016; Bethesda, MD, USA p. 1-7. [doi: 10.1109/WH.2016.7764561]
59. Suh J, Howe E, Lewis R, et al. Toward tailoring just-in-time adaptive intervention systems for workplace stress reduction: exploratory analysis of intervention implementation. JMIR Ment Health 2024 Sep 12;11:e48974. [doi: 10.2196/48974] [Medline: 39264703]
60. Weber J, Angerer P, Apolinário-Hagen J. Physiological reactions to acute stressors and subjective stress during daily life: a systematic review on ecological momentary assessment (EMA) studies. PLoS ONE 2022;17(7):e0271996. [doi: 10.1371/journal.pone.0271996]
61. Thomas BL, Claassen N, Becker P, Viljoen M. Validity of commonly used heart rate variability markers of autonomic nervous system function. Neuropsychobiology 2019 May 16;78(1):14-26. [doi: 10.1159/000495519]
62. Hayano J, Yuda E. Assessment of autonomic function by long-term heart rate variability: beyond the classical framework of LF and HF measurements. J Physiol Anthropol 2021;40(1):21. [doi: 10.1186/s40101-021-00272-y]
63. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health 2017;5:258. [doi: 10.3389/fpubh.2017.00258] [Medline: 29034226]
64. Barac M, Scaletty S, Hassett LC, et al. Wearable technologies for detecting burnout and well-being in health care professionals: scoping review. J Med Internet Res 2024 Jun 25;26:e50253. [doi: 10.2196/50253] [Medline: 38916948]
65. Stone AA, Shiffman S, Atienza AA, et al. Historical roots and rationale of ecological momentary assessment (EMA). In: Stone A, Shiffman S, Atienza A, editors. The Science of Real-Time Data Capture: Self-Reports in Health Research: Oxford: University Press; 2007:03-10. [doi: 10.1093/oso/9780195178715.003.0001]
66. Rivera-Romero O, Gabarron E, Ropero J, Denecke K. Designing personalised mHealth solutions: an overview. J Biomed Inform 2023 Oct;146:104500. [doi: 10.1016/j.jbi.2023.104500] [Medline: 37722446]

XSL•FO
RenderX

67. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020;22(9):e20283. [doi: 10.2196/20283]
68. Vaessen T, Rintala A, Otsabryk N, et al. The association between self-reported stress and cardiovascular measures in daily life: a systematic review. PLoS ONE 2021;16(11):e0259557. [doi: 10.1371/journal.pone.0259557] [Medline: 34797835]

Abbreviations

AI: artificial intelligence
CBT: cognitive behavioral therapy
EMA: ecological momentary assessment
HF: high frequency
HR: heart rate
HRV: heart rate variability
JITAI: just-in-time adaptive intervention
LF: low frequency
LMM: linear mixed model
mHealth: mobile health
VEDAS: Valencia Eustress-Distress Appraisal Scale

Edited by L Buis; submitted 25.Jun.2025; peer-reviewed by H Baumann, P Nohl-Deryk; revised version received 08.Oct.2025; accepted 12.Mar.2026; published 18.May.2026.

Please cite as:
Kunas B, Jung O, Schranz C, Schmoigl-Tonis M, Mehlis J, Laireiter AR
Multimodal Personalized Mobile Health Just-in-Time Adaptive Intervention for Occupational Stress Management: Pilot Study
JMIR Mhealth Uhealth 2026;14:e79642
URL: https://mhealth.jmir.org/2026/1/e79642
doi:10.2196/79642

© Beatrix Kunas, Oliver Jung, Christoph Schranz, Mathias Schmoigl-Tonis, Julian Mehlis, Anton-Rupert Laireiter. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 18.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Theory-Based Digital Intervention to Improve Maternal Oral Health Behaviors for Young Children: Quasi-Experimental Study

Mei Zhao[1], BMed, MDS; Min Liu[1], BMed, MDS, PhD; Shiyu Wang[2], BMed, MPH; Xiaoyue Zhang[3], BM, MPH; Chun Chang[3], BMed, MA, PhD; Qingping Yun[4], BMed, PhD

[1]Department of Preventive Dentistry, Beijing Stomatological Hospital, Capital Medical University, Beijing, China

[2]Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

[3]Department of Social Medicine and Health Education, School of Public Health, Peking University, Beijing, China

[4]Key Laboratory of Tropical Translational Medicine of Ministry of Education, School of Public Health, Hainan Medical University, NO.3 Xueyuan Road, Longhua district, Haikou, Hainan, China, Haikou, China

Corresponding Author:
Qingping Yun, BMed, PhD
Key Laboratory of Tropical Translational Medicine of Ministry of Education, School of Public Health, Hainan Medical University, NO.3 Xueyuan Road, Longhua district, Haikou, Hainan, China, Haikou, China

Abstract

Background: Parental oral health education is critical for preventing early childhood caries. However, few interventions are theoretically grounded or use digital approaches.

Objective: The objective of this study was to evaluate the effects of a health belief model–based digital intervention on maternal oral health behaviors.

Methods: This quasi-experimental study enrolled 648 mother-child dyads from 19 community health care centers (CHCs) in Beijing, China. CHCs were allocated to intervention or control groups depending on their voluntary adoption of the dental referral system. Ten CHCs (n=332, 52.6%) were assigned to the intervention group, where mothers received oral health education materials and had access to a dental referral system. The remaining 9 CHCs (n=316, 47.4%) served as the control group, in which mothers continued to receive standard child health care services. The primary outcome was parent-assisted toothbrushing, and the secondary outcome included other oral health behaviors, including night feeding practices, sugar intake, and dental visits. To evaluate the intervention effects on behavioral outcomes, generalized linear mixed models were used, accounting for repeated measures and potential confounding factors.

Results: Compared with the control group, the intervention group demonstrated a significant increase in parent-assisted toothbrushing, with an absolute difference of 10.3 (95% CI 3.0 to 17.6; *P*=.006) percentage points at 6 months and 1.5 (95% CI −7.2 to 10.1; *P*=.74) percentage points at 12 months. Additionally, dental visit rates were significantly higher in the intervention group at 12 months (odds ratio 4.65, 95% CI 1.30 to 16.70; *P*=.02). However, no statistically significant differences were observed between groups in nighttime feeding cessation or sugar intake control at either the 6- or 12-month follow-ups.

Conclusions: The health belief model–based digital intervention was effective in the short term for enhancing parent-assisted toothbrushing in young children, but its long-term effectiveness remains unproven. Future research should therefore prioritize exploring sustainability strategies.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2000039866; https://tinyurl.com/mr3r6aan

(JMIR Mhealth Uhealth 2026;14:e79002) doi:10.2196/79002

KEYWORDS

digital intervention; health belief model; oral health; young children; quasi-experimental study

Introduction

Background

Early childhood caries (ECC) is one of the most prevalent chronic conditions affecting children worldwide [1]. The Global Burden of Disease Study (2017) estimated approximately 532 million cases of untreated caries in primary teeth [2], and the epidemiological data from 193 countries (2007 - 2017) revealed that 57.3% of children aged 3 to 6 years were affected by dental caries [3]. In China, the situation appears even more concerning, with the national oral health survey (2015) reporting that 71.9% of children aged 5 years were affected by ECC [4]. Beyond its immediate detrimental effects on young children's quality of life, ECC significantly elevates the risk of caries in permanent dentition, resulting in long-term oral health consequences [5,6].

XSL•FO
RenderX

Parents play an important role in their young children's oral health care. Key recommended practices for infants' and toddlers' oral care include parent-assisted toothbrushing, limiting dietary sugar intake, avoiding prolonged nighttime feeding, and scheduling regular dental visits [7]. Despite the importance of these practices, most parents fail to adhere to the recommended oral health behaviors [8,9].

Although numerous oral health education programs have been designed for parents, their efficacy across different studies remains mixed [10]. A critical issue is that most interventions lacked theoretical frameworks, limiting their ability to explain intervention outcomes [10]. Moreover, existing programs predominantly emphasize home-based oral care behaviors while neglecting the persistently low rates of dental visits among young children [11,12]. This oversight is problematic because early dental visits enable oral health professionals to deliver personalized oral health education and help families adopt optimal oral health practices [13]. Nevertheless, strategies to effectively promote dental visit adherence remain poorly investigated.

Interventions grounded in theoretical frameworks tend to be more effective, underscoring the value of theory-driven approaches in health behavior change [14]. The health belief model (HBM) is a well-validated social psychological framework for health behavior modification [15]. The model's core constructs, including perceived susceptibility, perceived severity, perceived benefits, perceived barriers, and self-efficacy, offer a comprehensive theoretical framework for intervention design. A previous review has demonstrated its efficacy in improving oral health behaviors among adolescent and adult populations [16], yet its application in parental oral health education programs remains unexplored. To date, only one relevant study protocol has been published [17]. That study delivered HBM-based SMS text messaging interventions to parents of children aged 18 to 30 months, but it focused only on toothbrushing and sugar intake, without addressing early dental visits. Moreover, an existing intervention used caries prevalence and complication statistics to increase parents' awareness of ECC susceptibility and severity, but it did not address other important factors, such as perceived benefits and barriers to performing recommended oral health behaviors [18]. Therefore, the effectiveness of comprehensive HBM-based interventions in promoting early childhood oral health behaviors remains to be investigated.

Aims

To address this gap, we developed a novel digital intervention grounded in the HBM. To improve the low rate of dental visits among young children, the intervention integrated a dental referral system to facilitate access to dental care by helping parents locate nearby clinics and book appointments directly. The primary objective of this study was to evaluate the effects of this HBM-based digital intervention on improving maternal oral health behaviors for young children.

Methods

Design, Participants, and Recruitment

A parallel, quasi-experimental design was used to evaluate the effects of the HBM-based digital intervention on maternal oral health behaviors for young children aged 11 to 14 months. Participants were eligible if they met the following criteria: (1) aged 11 to 14 months, (2) at least one tooth has erupted, and (3) no cleft lip or palate. Children diagnosed by a physician with a chronic disease were excluded. The potential participants were recruited from 19 community health care centers (CHCs) in Beijing, China, between September and November 2020. Mothers of infants aged 11 to 12 months were approached during routine visits to community vaccination clinics, while those with children aged 13 to 14 months were contacted via telephone based on vaccination registry records and were invited to participate in this study. Mothers who agreed to participate were then asked to review and sign the informed consent form at the CHC.

CHCs were assigned to the intervention or control group based on their voluntary decision to implement the dental referral system with an allocation ratio of 1:1. Of the 19 CHCs, 10 (52.6%) that opted to implement the system formed the intervention group, while the remaining 9 (47.4%) served as the control group. To mitigate potential selection bias between treatment groups, we implemented controls for a set of family- and child-related factors known to be associated with young children's oral health in the literature [19-21] (refer to the Control Variables section).

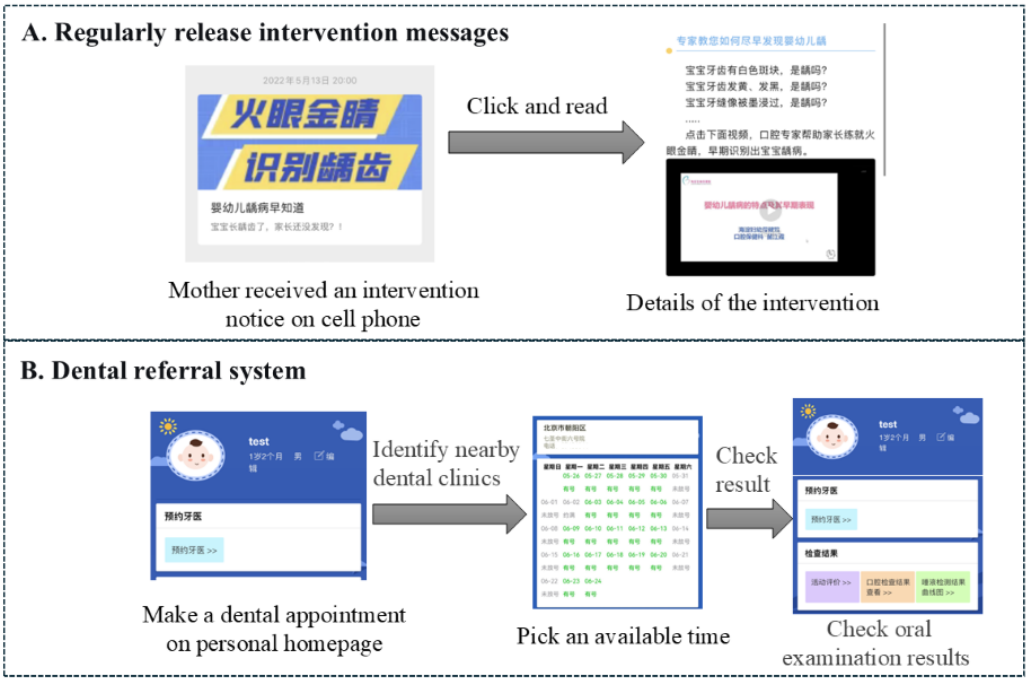
Intervention and Control Group

Mothers in the intervention group received the intervention through a WeChat (Tencent Holdings Limited) public account, via which we disseminated the intervention materials and hosted the dental referral system (Figure 1). The intervention lasted for 6 months, divided into two phases: (1) an intensive phase (first 2 months) with weekly intervention material deliveries, followed by (2) a maintenance phase (remaining 4 months) with biweekly material distribution. Throughout the intervention period, participants could use the online dental referral system to book dental appointments for their children. Guided by HBM, the intervention materials targeted the following constructs: (1) perceived susceptibility—presented Chinese epidemiological data on high prevalence of ECC [4]; (2) perceived severity—emphasized the adverse developmental consequences of untreated ECC; (3) perceived benefits—outlined advantages of oral health behaviors, including impacts on children's dental health, speech development, and long-term oral health outcomes [22,23]; (4) perceived barriers—provided practical strategies to address common obstacles (eg, managing child resistance during toothbrushing) [24]; and (5) self-efficacy—built confidence through the sharing of peer maternal experiences and success stories [25]. All content was reviewed and approved by a panel of 8 experts (including 3 oral health professionals, 2 pediatricians, and 2 health promotion specialists) to ensure scientific accuracy and appropriateness. Furthermore, the digital platform's usability and accessibility were evaluated in a 2-phase process: an initial assessment by the research team, followed

XSL•FO
RenderX

by a test with 8 mothers for user-friendliness and clarity. Feedback from these processes was incorporated prior to the final release. The detailed intervention materials, including the topic, key contents, corresponding theoretical constructs, and engagement metrics (eg, views) for each session, are provided in Multimedia Appendix 1.

Figure 1. Key elements of the intervention program. (A) Regular intervention message to parents; (B): dental referral system for appointment booking.



Mothers in the control group received standard child health care as stipulated by the national health management service for children aged 0 to 6 years [26], which for 1- to 2-year-olds includes biannual general physical examinations and feeding guidance. We did not provide any additional interventions to the control group.

Outcomes and Measurements

Due to COVID-19 pandemic–related restrictions, this study modified its primary outcome from oral health assessments to parent-assisted toothbrushing behaviors. This adaptation was made for two reasons: (1) standardized caries detection became impossible during the epidemic and (2) as a key ECC-preventive behavior [27], toothbrushing demands daily practice, reflecting ongoing parental commitment to their child's oral care. Other oral health behaviors, including nighttime feeding cessation, dental visits, and sugar intake control, served as secondary outcomes in this study. Consistent with our theoretical framework, the intervention aimed to improve maternal oral health behaviors by knowledge enhancement and psychosocial modification, with these factors concurrently assessed as secondary outcome measures.

All data were obtained using structured questionnaires at CHCs. Parent-assisted toothbrushing was operationally defined as the daily toothbrushing performed by caregivers before bedtime. Accordingly, the frequency of toothbrushing and whether it occurred before bedtime were recorded. The use of fluoridated toothpaste was also documented. Following biannual dental visit recommendations for young children, we assessed whether they had received dental visits in the preceding 6 months. Regarding nighttime feeding practices, mothers were asked whether they had discontinued nighttime feeding for their young children; responses indicating "yes" were categorized as nighttime feeding cessation. Additionally, data on the frequency of sugar intake, including candies, sugar-sweetened beverages, and bakery products, were collected.

Maternal oral health knowledge was evaluated using a 9-item questionnaire, which was developed according to Chinese infants' and toddlers' oral health guidelines (Multimedia Appendix 2) [23]. Each item offered 3 response options: "true," "false," and "don't know." Correct answers received 1 point, while incorrect or "don't know" responses scored 0.

This study also assessed HBM components, including perceived susceptibility to ECC, perceived severity of ECC, perceived benefits and barriers of oral health behaviors, and self-efficacy. All components were measured using 5-point Likert scale items (1=strongly disagree to 5=strongly agree), with cumulative scores calculated for each domain. A pilot study was conducted with 187 mothers (excluded from the main study) to assess the scales' readability and internal consistency. These scales demonstrated acceptable internal consistency, with Cronbach α coefficients ranging from 0.67 to 0.93. The items for each HBM component are presented in Multimedia Appendix 3.

Blinding

Blinding of participants was not possible due to the nature of this quasi-experimental study. Furthermore, as the primary outcome was self-reported, blinding of the outcome assessors was not applicable.

XSL•FO
RenderX

Sample Size

On the basis of our previous survey in Beijing, the rate of parent-assisted toothbrushing among 2-year-old children was 22.6%. This rate served as the reference for the control group, while the expected rate in the intervention group was set at 33% (ie, the prevention fraction of 50% based on a previous study [28]). The minimum necessary sample size was 258 participants for each arm based on a 2-sided test at the .05 level of significance and 80% power. Assuming a 20% loss of participants to follow-up, 310 participants were needed in each arm.

Control Variables

This study included 7 covariates encompassing both child and maternal characteristics. Child-related covariates included sex (male vs female), baseline age (categorized as 11 - 12 months vs 13 - 14 months), order of birth (first child vs nonfirst child) and primary caregiver status (mother vs others). Maternal characteristics included infant feeding practices, educational attainment, and household income. Infant feeding practices (0 - 6 months) were categorized into 3 groups: exclusive breastfeeding, mixed feeding, or exclusive formula feeding. Educational attainment was classified into 3 levels: college and below, bachelor's degree, or graduate degree. Based on participants' self-ranking of their family income relative to others in the local area, household income was classified into 3 tiers, including high (top 40%), middle (60% - 80%), or low (bottom 40%). The selection of birth order and infant feeding practices as covariates is supported by previous evidence linking them to children's oral health outcomes [19,20].

Follow-Up

Consistent with World Health Organization recommendations for evaluating novel digital health interventions, this study incorporated medium- and long-term impact assessments through 3-wave data collection [29]. Thus, data collection occurred at 3 time points in this study: outcome variables were assessed at baseline, 6 months (immediately after the intervention), and 12 months (6 months after the postintervention assessment). Maternal oral health knowledge and psychosocial variables were measured at baseline and 12 months, and control variables were obtained at baseline.

Statistical Analysis

Behavioral outcomes and control variables were reported as frequencies (proportions) and analyzed by chi-square test or Fisher exact test, as appropriate. Maternal oral health knowledge and psychosocial variables were described as mean (SD) and compared using 2-tailed independent *t* tests. Logistic regression generalized linear mixed models were used to assess behavioral outcome effects, and linear mixed models were applied to assess the psychosocial outcome effects. Fixed effects included time, treatment group, and their interaction (time×treatment). Random effects included community, accounting for within-cluster correlations. A statistically significant interaction indicates that the treatment effect varied over time.

Data analysis followed the intention-to-treat principles. Participants who had at least one follow-up visit were included in the primary analysis. Missing data were accounted for by a mixed-effects model using the full information maximum likelihood estimation method [30]. Sensitivity analyses were conducted to evaluate the robustness of the primary results. First, a complete-case analysis was performed. Second, given the continuous nature of the behavior, the last observation carried forward (LOCF) approach was applied under the assumption that the behavior would remain unchanged after the last measurement. Analyses were conducted using R software (version 4.4; R Foundation for Statistical Computing).

Ethical Considerations

Ethics approval (IRB00001052-19136) was obtained from the institutional review board of Peking University Health Science Center. This study was performed in accordance with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. Informed consent was obtained from all mothers. All data were anonymized prior to analysis. No personally identifiable information was included in the analysis or the reporting of findings. Participants received a pack of baby wipes (valued at RMB 50 [US $7.30]) on each visit. No additional financial compensation was offered for participation.

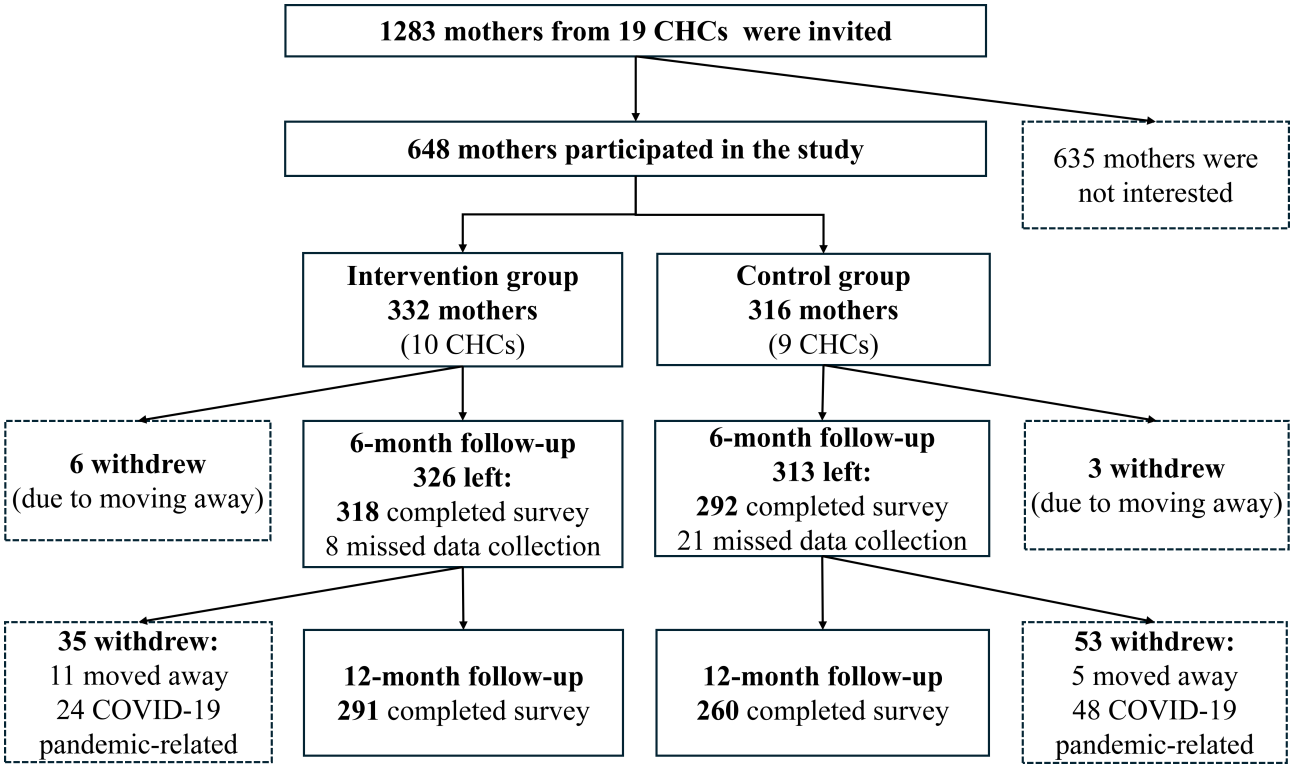
Results

Baseline Characteristics

This study enrolled 648 mother-child dyads. Of the children, 331 (51.1%) were male, and 546 (84.3%) were aged 11 to 12 months. Maternal education levels were distributed as follows (Table 1): 508 (78.4%) held at least a bachelor's degree, 103 (15.9%) had a college degree, and 37 (5.7%) completed high school or less. The baseline characteristics of participants were well matched between the intervention and control groups. Follow-up rates were 94.1% (n=610) at 6 months and 85% (n=551) at 12 months, meeting prespecified expectations (Figure 2).

XSL•FO
RenderX

Figure 2. Flow diagram of participants. CHC: community health care center.



XSL•FO
RenderX

Table . Baseline characteristics of participants.

Characteristics	Total (n=648), n (%)	Intervention group (n=332), n (%)	Control group (n=316), n (%)	*P* value
Child sex				.80
Male	331 (51.1)	168 (50.6)	163 (51.6)	
Female	317 (48.9)	164 (49.4)	153 (48.4)	
Child age group (months)				.63
11 - 12	546 (84.3)	282 (84.9)	264 (83.5)	
13 - 14	102 (15.7)	50 (15.1)	52 (16.5)	
Birth order				.36
First child	462 (71.3)	242 (72.9)	220 (69.6)	
Non–first child	186 (28.7)	90 (27.1)	96 (30.4)	
Primary caregiver				.80
Mother	294 (45.4)	149 (44.9)	145 (45.9)	
Others	354 (54.6)	183 (55.1)	171 (54.1)	
Infant feeding practice				.76
Exclusive breastfeeding	339 (52.3)	175 (52.7)	164 (51.9)	
Mixed feeding	224 (34.6)	111 (33.4)	113 (35.8)	
Exclusive formula feeding	85 (13.1)	46 (13.9)	39 (12.3)	
Maternal education				.15
High school or below	37 (5.7)	16 (4.8)	21 (6.6)	
College (2 - 3 y)	103 (15.9)	61 (18.4)	42 (13.3)	
Bachelor's degree or higher	508 (78.4)	255 (76.8)	253 (80.1)	
Household income				.30
High	157 (24.2)	75 (22.6)	82 (25.9)	
Middle	459 (70.8)	237 (71.4)	222 (70.3)	
Low	32 (5)	20 (6)	12 (3.8)	

Of the initial sample, 38 (5.9%) participants did not complete the 6-month follow-up. Compared with completers, noncompleters had significantly higher baseline self-efficacy scores (*P*=.03) and a higher prevalence of dental visits (*P*=.003) at baseline (Multimedia Appendix 4).

Behavioral Outcome

The intervention demonstrated significant short-term effects on parent-assisted toothbrushing behavior. Compared with controls, mothers in the intervention group showed a 10.3 percentage point increase (95% CI 3.0-17.6) in toothbrushing behavior, corresponding to an adjusted odds ratio (OR) of 2.43 (95% CI 1.24-4.74). The observed improvement was consistent in sensitivity analyses using complete-case analysis (adjusted OR 2.19, 95% CI 1.06-4.53; Multimedia Appendix 5) and the LOCF approach to handle missing data (adjusted OR 2.88, 95% CI 1.43-5.78; Multimedia Appendix 6). However, this improvement was not sustained at the 12-month assessment (Table 2), with an adjusted OR of 1.20 (95% CI 0.61-2.36). No statistically significant effect was found for toothbrushing with fluoridated toothpaste at this follow-up as well (Multimedia Appendix 7), with an adjusted OR of 2.12 (95% CI 0.78-5.75).

XSL•FO
RenderX

Table . Intervention effects on maternal oral health behaviors for young children.

	Intervention group, n/N (%)	Control group, n/N (%)	Rate difference, % (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
Parent-assisted brushing						
Baseline	47/332 (14.2)	47/316 (14.9)	−0.3 (−2.5 to 1.7)	.74	1.00[a]	—[b]
6-month follow-up	124/318 (39)	82/292 (28.1)	10.3 (3.0 to 17.6)	.006	2.43 (1.24 to 4.74)	.009
12-month follow-up	118/291 (40.5)	104/260 (40)	1.5 (−7.2 to 10.1)	.74	1.20 (0.61 to 2.36)	.59
Nighttime feeding cessation						
Baseline	104/332 (31.3)	79/316 (25)	1.0 (−4.5 to 6.5)	.73	1.00	—
6-month follow-up	225/318 (70.8)	214/292 (73.3)	−0.6 (−3.3 to 2.0)	.64	0.53 (0.16 to 1.73)	.29
12-month follow-up	253/291 (86.9)	222/260 (85.4)	0.0 (−0.1 to 0.2)	.67	1.21 (0.23 to 6.32)	.82
Sugar intake control						
Baseline	259/332 (78)	241/316 (76.3)	2.4 (−4.8 to 9.8)	.51	1.00	—
6-month follow-up	220/318 (69.2)	212/292 (72.6)	−3.5 (−12.2 to 9.2)	.43	0.73 (0.41 to 1.28)	.27
12-month follow-up	190/291 (65.3)	172/260 (66.2)	0.1 (−9.8 to 10.0)	.99	0.87 (0.49 to 1.53)	.62
Dental visit in past 6 months						
Baseline	9/332 (2.7)	9/316 (2.8)	−0.1 (−0.7 to 0.5)	.65	1.00	—
6-month follow-up	33/318 (10.4)	15/292 (5.1)	1.5 (−0.1 to 3.6)	.15	3.29 (0.98 to 11.05)	.05
12-month follow-up	30/291 (10.3)	13/260 (5)	2.0 (–0.5 to 4.4)	.12	4.65 (1.30 to 16.70)	.02

[a]Odds ratio value of 1.00 indicates the reference group in the model.

[b]Not applicable.

Additionally, the intervention showed a limited effect on dental visits. Mothers in the intervention group had a higher likelihood of taking their children for dental visits compared with controls, with adjusted ORs of 3.29 (95% CI 0.98 - 11.05) at 6 months and 4.65 (95% CI 1.30 - 16.70) at 12 months. However, the low incidence of dental visits led to wide CIs in both the primary and sensitivity analyses. The sensitivity analyses yielded ORs of 1.81 (95% CI 0.29-11.38) for the complete-case analysis (Multimedia Appendix 5) and 3.85 (95% CI 0.82-17.95) for the LOCF approach at 12-month follow-up (Multimedia Appendix 6). No statistically significant effects of the intervention were observed on either nighttime feeding cessation or sugar intake reduction at the 6- or 12-month follow-ups (*P*>.05).

Maternal Knowledge and Psychosocial Outcomes

We further assessed changes in maternal knowledge and psychosocial variables over the 12-month study period. The intervention significantly improved maternal oral health knowledge scores (mean difference 0.31 points; 95% CI 0.01-0.60), whereas no statistically significant changes were observed in other psychosocial components (Table 3).

XSL•FO
RenderX

Table . Intervention effects on maternal oral health knowledge and health belief model constructs.

	Intervention group, mean (SD)	Control group, mean (SD)	Mean difference (95% CI)	*P* value
Maternal oral health knowledge				
Baseline	6.24 (1.56)	6.25 (1.75)	−0.03 (−0.31 to 0.25)	.82
12-month follow-up	7.15 (1.43)	6.90 (1.51)	0.31 (0.01 to 0.60)	.04
Perceived susceptibility to ECC[a]				
Baseline	11.66 (1.72)	11.73 (1.89)	−0.08 (−0.44 to 0.27)	.65
12-month follow-up	12.11 (1.82)	11.91 (2.01)	0.21 (−0.17 to 0.59)	.28
Perceived severity of ECC				
Baseline	14.28 (1.50)	14.22 (1.70)	0.05 (−0.23 to 0.33)	.72
12-month follow-up	14.46 (1.27)	14.28 (1.64)	0.19 (−0.11 to 0.48)	.23
Perceived benefits of behaviors				
Baseline	28.89 (2.75)	28.73 (2.94)	0.15 (−0.31 to 0.60)	.53
12-month follow-up	29.40 (1.72)	29.02 (3.01)	0.38 (−0.11 to 0.86)	.14
Perceived barriers of behaviors				
Baseline	22.02 (6.13)	22.00 (6.58)	−0.03 (−1.16 to 1.10)	.96
12-month follow-up	20.82 (5.46)	20.75 (6.04)	0.19 (−1.00 to 1.37)	.76
Self-efficacy				
Baseline	13.56 (1.79)	13.72 (1.95)	−0.15 (−0.48 to 0.19)	.40
12-month follow-up	13.56 (1.78)	13.45 (2.01)	0.17 (−0.18 to 0.52)	.34

[a]ECC: early childhood caries.

Discussion

Main Findings

In this quasi-experimental study, the HBM-based digital intervention achieved a 10.3 percentage point improvement in parent-assisted toothbrushing behavior at the 6-month follow-up, but this improvement did not persist to the 12-month follow-up. Additionally, slight improvements were observed in dental visit adherence and maternal oral health knowledge at the 12-month follow-up.

Comparison With Prior Work

Previous interventions targeting mothers of young children (aged 0-3 years) primarily focused on oral health knowledge and skills without solid theoretical foundation [18,31,32]. Among the few theory-guided interventions, one used the theory of planned behavior (TPB). While this intervention improved maternal attitudes and perceived behavioral control, it did not yield significant changes in behavioral intention or toothbrushing behavior at the 3-month follow-up [28]. This limited efficacy may be attributed to 2 key conceptual omissions in TPB. First, TPB does not incorporate self-efficacy, which is crucial for sustaining parental involvement in children's oral hygiene [33]. Second, it overlooks the perceived threat, a construct whose enhancement has been linked to reduced caries incidence in children [18]. HBM integrates these essential components. Consequently, it was selected as the theoretical framework for this study.

Guided by HBM, our study achieved short-term improvements in toothbrushing behavior, likely mediated by increased maternal oral health knowledge [34], as our results indicated that maternal oral health knowledge scores increased 0.31 points from baseline. However, the intervention effects diminished by the 12-month follow-up, possibly due to persistent perceived barriers or inadequate self-efficacy reinforcement [24,35]. To facilitate lasting behavioral modifications, future research should explore the efficacy of hybrid interventions that integrate initial education with booster sessions (eg, regular reminders) to address behavioral relapses. Future practice should embed oral health promotion into routine child health services (eg, during vaccination visits), thereby creating systematic opportunities for timely problem-solving to overcome barriers and for skill-refreshing to enhance self-efficacy.

Previous interventions focused on home-based oral behaviors and largely overlooked the importance of dental visits [10,36]. To our knowledge, this study is the first to implement an online dental referral system designed to facilitate parental scheduling of dental appointments for children. However, the intervention yielded only a modest effect, with a mere 2.0 percentage point increase in dental visit rates at the 12-month follow-up. This limited effect is likely multifactorial. First, the COVID-19 pandemic created a universal barrier, during which families avoided nonessential pediatric dental care [37]. More fundamentally, our digital dental referral system primarily addressed the barriers of finding and scheduling an appointment, but it could not overcome other critical obstacles, such as financial constraints [38], and parental attitudes regarding dental

XSL•FO
RenderX

visits. If mothers primarily view dental visits as therapeutic rather than a preventive service, they are unlikely to take their children to oral health professionals in the absence of symptoms [39,40]. This suggests that future interventions need to address not only service access but also cost barriers and underlying beliefs about dental visits.

It should be noted that this study was likely underpowered to detect a significant effect on dental visits, with the low incidence of this behavior resulting in wide CIs. This limits the robustness of our findings and calls for cautious interpretation of the results. Therefore, future studies with larger sample sizes are required to evaluate the intervention's effect.

However, our intervention showed no significant effect on nighttime feeding cessation. Further analysis revealed that the mean cessation age was similar between groups (intervention: mean 13.48, SD 4.18 months vs control: mean 13.41, SD 3.86 months; *P*=.79). This overall cessation timing closely aligned with the baseline age of participants (11 - 14 months), which could partially explain the limited intervention effect.

Strengths and Limitations

This theory-based digital intervention evaluated both oral health behavior changes and shifts in maternal psychosocial factors, offering explanatory insights into the mechanisms behind behavioral improvements. However, the study had several limitations. First, randomization was not feasible due to COVID-19 pandemic–related constraints that prevented comprehensive implementation of the referral intervention across participating CHCs. The lack of randomization limited our ability to fully account for potential confounders [41]. To mitigate this, we used a parallel group comparison and adjusted for child- and maternal-related confounding variables, strengthening causal inference. Second, the COVID-19 pandemic prevented clinical assessments of oral health outcomes (eg, caries incidence), restricting our analysis to rely on self-reported behaviors, which might introduce reporting bias. Future studies can more comprehensively assess the effect of theory-based digital interventions on children's oral health.

Conclusions

The HBM-based digital intervention was effective in the short term for enhancing parent-assisted toothbrushing in young children, but its long-term effectiveness remains unproven. Future research should therefore prioritize exploring sustainability strategies.

Acknowledgments

The authors thank the children and their mothers for their involvement. No generative artificial intelligence tools were used to generate or ideate the text or visual materials included in this manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (grant 82504423).

Data Availability

The dataset used and analyzed during this study will be available from the corresponding author upon reasonable request.

Authors' Contributions

ML, CC, and QY designed the study. MZ, SW, and QY drafted the manuscript. MZ and XZ recruited participants. CC and QY were responsible for the intervention preparation. MZ, XZ, and QY collected the data, and SW and QY analyzed the data. All authors revised the manuscript and approved the final version for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Topics and key contents of intervention materials.
[DOCX File, 24 KB - mhealth_v14i1e79002_app1.docx]

Multimedia Appendix 2
Maternal oral health knowledge questionnaire.
[DOCX File, 15 KB - mhealth_v14i1e79002_app2.docx]

Multimedia Appendix 3
Items measuring health belief model constructs.
[DOCX File, 21 KB - mhealth_v14i1e79002_app3.docx]

XSL•FO
RenderX

Multimedia Appendix 4
Baseline characteristics of participants by follow‐up completion.
[DOCX File, 21 KB - mhealth_v14i1e79002_app4.docx]

Multimedia Appendix 5
Intervention effects on maternal oral health behaviors: Based on complete-case analysis.
[DOCX File, 19 KB - mhealth_v14i1e79002_app5.docx]

Multimedia Appendix 6
Intervention effects on behavioral outcomes: Based on the last observation carried forward approach.
[DOCX File, 19 KB - mhealth_v14i1e79002_app6.docx]

Multimedia Appendix 7
Intervention effects on toothbrushing with fluoridated toothpaste.
[DOCX File, 17 KB - mhealth_v14i1e79002_app7.docx]

Checklist 1
CONSORT-eHealth (V 1.6.1).
[PDF File, 1294 KB - mhealth_v14i1e79002_app8.pdf]

References

1. Chen KJ, Gao SS, Duangthip D, Lo EC, Chu CH. Prevalence of early childhood caries among 5-year-old children: a systematic review. J Investig Clin Dent 2019 Feb;10(1):e12376. [doi: 10.1111/jicd.12376] [Medline: 30499181]
2. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858. [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
3. El Tantawi M, Folayan MO, Mehaina M, et al. Prevalence and data availability of early childhood caries in 193 United Nations Countries, 2007-2017. Am J Public Health 2018 Aug;108(8):1066-1072. [doi: 10.2105/AJPH.2018.304466] [Medline: 29927650]
4. Du MQ, Li Z, Jiang H, et al. Dental caries status and its associated factors among 3- to 5-year-old children in China: a national survey. Chin J Dent Res 2018;21(3):167-179. [doi: 10.3290/j.cjdr.a41076] [Medline: 30255168]
5. Fung MH, Wong MC, Lo EC, Chu CH. Early childhood caries: a literature review. J Oral Hyg Health 2013;1(1):107. [doi: 10.4172/johh.1000107]
6. Kagihara LE, Niederhauser VP, Stark M. Assessment, management, and prevention of early childhood caries. J Am Acad Nurse Pract 2009 Jan;21(1):1-10. [doi: 10.1111/j.1745-7599.2008.00367.x] [Medline: 19125889]
7. Tinanoff N, Baez RJ, Diaz Guillory C, et al. Early childhood caries epidemiology, aetiology, risk assessment, societal burden, management, education, and policy: global perspective. Int J Paediatr Dent 2019 May;29(3):238-248. [doi: 10.1111/ipd.12484] [Medline: 31099128]
8. Trinh MV, Rhodes AL, Measey MA, Silva M. Dental visits in early life: patterns and barriers among Australian children. Aust N Z J Public Health 2022 Jun;46(3):281-285. [doi: 10.1111/1753-6405.13224] [Medline: 35298078]
9. Child oral health: habits in Australian homes. The Royal Children's Hospital Melbourne. 2018. URL: https://rchpoll.org.au/polls/child-oral-health-habits-in-australian-homes/ [accessed 2026-05-01]
10. Aliakbari E, Gray-Burrows KA, Vinall-Collier KA, et al. Home-based toothbrushing interventions for parents of young children to reduce dental caries: a systematic review. Int J Paediatr Dent 2021 Jan;31(1):37-79. [doi: 10.1111/ipd.12658] [Medline: 32333706]
11. Kasemkhun P, Jirarattanasopha V, Lertsooksawat W. Characteristics and influencing factors of the first dental visit among children in Bangkok, Thailand: a cross-sectional study. BMC Oral Health 2024 Jan 3;24(1):11. [doi: 10.1186/s12903-023-03811-4] [Medline: 38172814]
12. Mileva SP, Kondeva VK. Age at and reasons for the first dental visit. Folia Med (Plovdiv) 2010;52(4):56-61. [doi: 10.2478/v10153-010-0018-x] [Medline: 21462893]
13. American Academy of Pediatric Dentistry. Periodicity of examination, preventive dental services, anticipatory guidance/counseling, and oral treatment for infants, children, and adolescents. In: The Reference Manual of Pediatric Dentistry: American Academy of Pediatric Dentistry; 2025:312-324.
14. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4. [doi: 10.2196/jmir.1376] [Medline: 20164043]

XSL•FO
RenderX

15. Rosenstock IM. Historical origins of the health belief model. Health Educ Monogr 1974 Dec;2(4):328-335. [doi: 10.1177/109019817400200403]
16. Sanaei Nasab H, Yazdanian M, Mokhayeri Y, et al. The role of psychological theories in oral health interventions: a systematic review and meta‐analysis. Int J Dent Hyg 2019 May;17(2):142-152. [doi: 10.1111/idh.12386]
17. Wang K, Lee GH, Liu P, Gao X, Wong SY, Wong MC. Health belief model for empowering parental toothbrushing and sugar intake control in reducing early childhood caries among young children-study protocol for a cluster randomized controlled trial. Trials 2022 Apr 12;23(1):298. [doi: 10.1186/s13063-022-06208-w] [Medline: 35413872]
18. Basir L, Rasteh B, Montazeri A, Araban M. Four-level evaluation of health promotion intervention for preventing early childhood caries: a randomized controlled trial. BMC Public Health 2017 Oct 2;17(1):767. [doi: 10.1186/s12889-017-4783-9] [Medline: 28969655]
19. Liu M, Yun Q, Zhao M, et al. Association of siblings' presence and oral health-related quality of life among children: a cross-sectional study. BMC Oral Health 2021 Mar 24;21(1):153. [doi: 10.1186/s12903-021-01526-y] [Medline: 33757508]
20. Cui L, Li X, Tian Y, et al. Breastfeeding and early childhood caries: a meta-analysis of observational studies. Asia Pac J Clin Nutr 2017;26(5):867-880. [doi: 10.6133/apjcn.082016.09] [Medline: 28802297]
21. Kyan A, Takakura M, Kamiya Y, Kinjo N, Nakasone T. Socioeconomic inequalities in toothbrushing behaviours in young children: a children's lifestyle survey in a representative population of A city, Okinawa prefecture, Japan. Eur Arch Paediatr Dent 2022 Dec;23(6):969-977. [doi: 10.1007/s40368-022-00751-5] [Medline: 36063356]
22. American Academy of Pediatric Dentistry. Policy on early childhood caries (ECC): Consequences and preventive strategies. In: The Reference Manual of Pediatric Dentistry: American Academy of Pediatric Dentistry; 2025:96-100.
23. National Health Commission of the People's Republic of China. Oral health guideline for Chinese population. Chinese J Stomatology 2010;45(6):325-330. [doi: 10.3760/cma.j.issn.1002-0098.2010.06.003]
24. Aliakbari E, Gray-Burrows KA, Vinall-Collier KA, et al. Facilitators and barriers to home-based toothbrushing practices by parents of young children to reduce tooth decay: a systematic review. Clin Oral Investig 2021 Jun;25(6):3383-3393. [doi: 10.1007/s00784-021-03890-z] [Medline: 33743074]
25. Nagarkar A, Martin G, Sadaniantz K, et al. Storytelling for health promotion: a scoping review. Am J Health Promot 2026 Feb;40(2):187-209. [doi: 10.1177/08901171251365366] [Medline: 40770144]
26. He Z, Xu T. China's actions to achieve universal health coverage for children. China CDC Wkly 2022 Sep 9;4(36):802-806. [doi: 10.46234/ccdcw2022.167] [Medline: 36285278]
27. Kumar S, Tadakamadla J, Johnson NW. Effect of toothbrushing frequency on incidence and increment of dental caries: a systematic review and meta-analysis. J Dent Res 2016;95(11):1230-1236. [doi: 10.1177/0022034516655315]
28. Makvandi Z, Karimi-Shahanjarini A, Faradmal J, Bashirian S. Evaluation of an oral health intervention among mothers of young children: a clustered randomized trial. J Res Health Sci 2015;15(2):88-93. [Medline: 26175290]
29. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. World Health Organization. 2016. URL: https://www.who.int/publications/i/item/9789241511766 [accessed 2026-05-01]
30. Molenberghs G, Thijs H, Jansen I, et al. Analyzing incomplete longitudinal clinical trial data. Biostatistics 2004 Jul;5(3):445-464. [doi: 10.1093/biostatistics/5.3.445] [Medline: 15208205]
31. Mohebbi SZ, Virtanen JI, Vahid-Golpayegani M, Vehkalahti MM. A cluster randomised trial of effectiveness of educational intervention in primary health care on early childhood caries. Caries Res 2009;43(2):110-118. [doi: 10.1159/000209343] [Medline: 19321988]
32. Yu KF, Wen W, Liu P, Gao X, Lo ECM, Wong MCM. Effectiveness of family-centered oral health promotion on toddler oral health in Hong Kong. J Dent Res 2022 Mar;101(3):286-294. [doi: 10.1177/00220345211036663] [Medline: 34515552]
33. Smith SR, Kroon J, Schwarzer R, Hamilton K. Parental social-cognitive correlates of preschoolers' oral hygiene behavior: a systematic review and meta-analysis. Soc Sci Med 2020 Nov;264:113322. [doi: 10.1016/j.socscimed.2020.113322] [Medline: 32916333]
34. Naidu R, Nunn J, Irwin JD. The effect of motivational interviewing on oral healthcare knowledge, attitudes and behaviour of parents and caregivers of preschool children: an exploratory cluster randomised controlled study. BMC Oral Health 2015 Sep 2;15:101. [doi: 10.1186/s12903-015-0068-9] [Medline: 26328785]
35. Hamilton K, Cornish S, Kirkpatrick A, Kroon J, Schwarzer R. Parental supervision for their children's toothbrushing: mediating effects of planning, self-efficacy, and action control. Br J Health Psychol 2018 May;23(2):387-406. [doi: 10.1111/bjhp.12294] [Medline: 29349924]
36. Ly-Mapes O, Jang H, Al Jallad N, et al. Effectiveness of interventions on improving early-life dental care utilization: a systematic review and meta-analysis. Quintessence Int 2024 Oct 24;55(9):722-732. [doi: 10.3290/j.qi.b5640008] [Medline: 39078171]
37. Üstün N, Akgöl BB, Bayram M. Influence of COVID-19 pandemic on paediatric dental attendance. Clin Oral Investig 2021 Nov;25(11):6185-6191. [doi: 10.1007/s00784-021-03917-5] [Medline: 33786650]
38. Mousa M, Alkandari M, Alabbasi H, et al. Barriers to dental service utilization in children: a systematic review and meta-analysis. BMC Oral Health 2026 Feb 17;26(1):521. [doi: 10.1186/s12903-026-07890-x] [Medline: 41699601]
39. Alshahrani NF, Alshahrani AN, Alahmari MA, Almanie AM, Alosbi AM, Togoo RA. First dental visit: age, reason, and experiences of Saudi children. Eur J Dent 2018;12(4):579-584. [doi: 10.4103/ejd.ejd_426_17] [Medline: 30369806]

XSL•FO
RenderX

40. Al Agili DE, Farsi NJ. Need for dental care drives utilisation of dental services among children in Saudi Arabia. Int Dent J 2020 Jun;70(3):183-192. [doi: 10.1111/idj.12539] [Medline: 31912900]
41. Harris AD, McGregor JC, Perencevich EN, et al. The use and interpretation of quasi-experimental studies in medical informatics. J Am Med Inform Assoc 2006;13(1):16-23. [doi: 10.1197/jamia.M1749] [Medline: 16221933]

Abbreviations

CHC: community health care center
ECC: early childhood caries
HBM: health belief model
LOCF: last observation carried forward
OR: odds ratio
TPB: theory of planned behavior

Edited by L Buis; submitted 17.Jun.2025; peer-reviewed by AM Meldrum, O Igunma, S Jiang; revised version received 20.Apr.2026; accepted 20.Apr.2026; published 22.May.2026.

Please cite as:
Zhao M, Liu M, Wang S, Zhang X, Chang C, Yun Q
A Theory-Based Digital Intervention to Improve Maternal Oral Health Behaviors for Young Children: Quasi-Experimental Study
JMIR Mhealth Uhealth 2026;14:e79002
URL: https://mhealth.jmir.org/2026/1/e79002
doi:10.2196/79002

© Mei Zhao, Min Liu, Shiyu Wang, Xiaoyue Zhang, Chun Chang, Qingping Yun. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 22.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effectiveness of Step Goal Personalization Strategies on Physical Activity in a Mobile Health App: A Field Study

Xia Liu[1,2,3], PhD; Tammo H A Bijmolt[3], PhD; Marijke C Leliveld[3], PhD; Ernst H Noppers[4], PhD

[1]Department of Marketing, Faculty of Business Administration, Hebei University of Economics and Business, Xuefu Road 47, Shijiazhuang, China

[2]Research Centre for Science Technology and Innovation Policy, Hebei University of Economics and Business, Shijiazhuang, China

[3]Department of Marketing, Faculty of Economics and Business, University of Groningen, Groningen, The Netherlands

[4]Menzis, Groningen, The Netherlands

Corresponding Author:
Xia Liu, PhD
Department of Marketing, Faculty of Business Administration, Hebei University of Economics and Business, Xuefu Road 47, Shijiazhuang, China

Abstract

Background: Goal personalization features integrated into mobile health apps have the potential to enhance physical activity, as some evidence shows that the personalized goals generated by algorithms are more effective than default or fixed goals. However, it remains unclear whether goals set by users are more effective than fixed goals and which personalization strategy is more effective for different user segments.

Objective: This field study aimed to evaluate (1) the efficacy of 2 step goal personalization strategies—personalized-by-you and personalized-by-the-algorithm—and (2) which strategy is more effective among users with different activity levels.

Methods: All users of SamenGezond, a Dutch mobile health app, have a default goal of 2000 steps per day, 5 days a week. For this study, 2 random groups were selected, totaling 5800 users. Subsequently, an email was sent to 3800 users in group 1, asking whether they were satisfied with their current goal. Those who were not satisfied were offered 2 personalization options: to set a goal themselves or to have the algorithm integrated in the app set goals for them. In total, 1399 users responded: 230 chose to set their own goals (personalized-by-you group), 236 opted for setting the goal by the algorithm (personalized-by-the-algorithm group), and 933 chose to keep the default goal (not-changed group). The algorithm used a moving-window percentile rank method based on step data from the previous 4 weeks. Users who did not personalize retained the default goal. The remaining 2000 users in group 2 did not receive the email and also retained the default goal. To evaluate the effectiveness of step goal personalization strategies, we used propensity score matching and difference-in-difference analysis.

Results: Users in the personalized-by-you group increased weekly step count by 3793 a week, while those in the personalized-by-the-algorithm group increased by 4315 steps a week, compared with the not-changed group (users with default goals). The 2 strategies appear to have a similar effect. Interestingly, users in the not-changed group also increased their weekly steps by 1759. Furthermore, the effectiveness of each strategy varied by baseline activity level. The personalized-by-you strategy was effective for medium- (increase of 5842 steps) and high-active users (increase of 4266 steps) but not for low-active users (increase of 384 steps; P=.82). Conversely, the personalized-by-the-algorithm strategy was effective for low- (increase of 5095 steps) and medium-active users (increase of 5278 steps) but not for high-active users (increase of 1446 steps; P=.51).

Conclusions: Step goal personalization demonstrates short-term effectiveness. However, their impact varies by users' baseline activity levels, indicating the need for a tailored approach for different user segments. Future studies should examine the long-term effects of such interventions to design sustainable health behavior change strategies.

(JMIR Mhealth Uhealth 2026;14:e81779) doi:10.2196/81779

KEYWORDS

mobile health app; goal management; goal personalization; physical activity; difference-in-difference; propensity score matching

Introduction

Background

Despite the compelling evidence of its benefits, many barriers (eg, inconvenience, lack of time, and motivation) hinder individuals from regularly exercising. According to the World Health Organization, 1 in 4 people worldwide is physically inactive [1]. Overcoming these obstacles requires stronger motivation and self-regulation. Mobile health (mHealth) technologies offer significant potential as a solution to physical inactivity [2-4]. mHealth refers to the integration of mobile computing, medical sensors, and communication technologies,

XSL•FO
RenderX

designed to deliver health care services [5]. These tools allow individuals to manage goals, track physical activity, and receive personalized feedback—making exercise more accessible and convenient. However, measurement capabilities and personalized feedback alone may not be sufficient to sustain motivation. In some cases, they can even backfire, leading to demotivation if users feel pressured or coerced [6-8].

Goal Management and Goal Setting

This study focuses on goal management, a central feature in most mHealth apps, as goals are fundamental drivers of motivation and self-regulation [9]. Goal setting is one of the most frequently used and effective strategies for behavioral change [10,11]. Despite its demonstrated importance, an essential question remains: what kind of goals should be set to effectively motivate behavioral change?

This question sparked abundant research and theories on goal setting. For instance, one line of research recommends setting achievable and realistic goals as formalized in the SMART (specific, measurable, achievable, realistic, time-bound) framework [12,13]. Other research emphasizes the motivational benefits of challenging goals [14,15]. Despite that both approaches have merits, the majority of research on goal setting agrees that challenging goals may be discouraging if perceived as unattainable, while easy goals may fail to motivate people [11,14]. The effectiveness of goal setting thus depends critically on aligning goals with users' abilities, needs, and contexts [11]. This raises an essential question: how can such goals be designed and calibrated effectively?

Goal Personalization

To achieve this alignment, many mHealth apps have introduced goal personalization features. These features allow users to modify the default goal, uniform across all users, offering options to personalize the goal to better align with individual circumstances. The process of goal personalization can typically take two forms: (1) personalized-by-you (also known as "customization"), where users set their own goals, and (2) personalized-by-the-algorithm, where the app generates goals based on users' previous exercise data with an integrated algorithm [16].

Goal personalization is expected to increase physical activity compared to no personalization (with default goals) for a few reasons. First, from a goal-setting theory perspective, personalized goals—whether self-set or algorithm-generated—might be more effective than default goals because they accommodate individual abilities and situational context. In addition, from the self-regulation perspective, goal personalization may have the potential to increase physical activity compared to default goals, since the process of personalization encourages active user engagement. This may further foster a greater sense of control and autonomy [17] and reinforce users' perception of themselves as the originators of their goals [18-20]. As a result, users may experience less depletion of self-control resources and feel more energized and maintain greater commitment to achieving their goals [21]. Previous research has partly demonstrated the efficacy of step goal personalization. For example, a study found that a personalized-by-the-algorithm strategy led users to take more steps than a static and default goal (eg, 10,000 steps) [22]. Similarly, 2 more studies reported that the personalized goals generated by the algorithm led to a slower decline in step counts over time compared to nonpersonalized, static goals [23,24].

Despite these promising findings, existing research faces several limitations. First, most studies compare the effectiveness of goals personalized by the algorithm with high, fixed goals (eg, 10,000 steps), which may be challenging for most users [22-24]. It remains unclear whether personalization is still effective when compared to low default goals (eg, 2000 steps). Moreover, much of the prior research on the effect of personalized-by-the-algorithm relies on small-scale randomized controlled trials, often involving fewer than 50 participants per condition, limiting the generalizability of findings [22-24]. For example, one study stated in its limitation section a low sample size of 64 staff workers from a certain university with a dominant proportion of female participants (83%). The results may not generalize to the general public [23]. Second, as far as we know, no previous research has examined the effectiveness of the personalized-by-you approach, relative to existing default goals, in promoting physical activity, underscoring the need for further research. Third, and more importantly, no evidence exists on which personalization strategy is more effective for different types of users. The goal-setting literature offers mixed insights: some studies suggest that self-set goals may be less effective than the assigned goals due to the sustained effort required from users [23,25,26], while others argue that self-set goals promote greater autonomy and may thus be more motivating [27,28]. It is possible that the effectiveness of goal-setting approaches is moderated by other factors, such as individuals' locus of control [29] or their levels of ability [11]. Thus, it is important to investigate which goal personalization strategy is more effective across different types of users.

Collectively, these limitations underscore the need to examine the effects of both goal personalization strategies and investigate which goal personalization strategy works best for different types of users.

Objectives

To address these gaps, this study investigates whether goal personalization increases health behavior and how the different routes of setting the personalized goal affect health behavior. Specifically, this study investigates the effect of goal personalization on physical activity by examining whether 2 personalization strategies (personalized-by-you and personalized-by-the-algorithm) outperform no personalization (with uniform default goals). Additionally, it also examines which strategy is more effective among users with different baseline activity levels (ie, high-, medium-, and low-active users), providing a more comprehensive understanding of how goal personalization operates in real-world settings.

Methods

Ethical Considerations

This study utilized data from an experiment implemented by SamenGezond, which was designed to optimize the effectiveness

XSL•FO
RenderX

of the SamenGezond program. Specifically, this research involved a secondary analysis of the data from adult users of the SamenGezond, an mHealth app, collected by SamenGezond in accordance with its terms of use and privacy policy. The policy states that SamenGezond measures and stores user activities and analyzes them (1) to provide feedback and advice and (2) to test and optimize the effectiveness of SamenGezond. By agreeing to the privacy policy, users consented to the use of their data for analytical purposes. All data used in this research were fully deidentified and aggregated. The use of these data was approved by the institutional research board of the Faculty of Economics and Business of the University of Groningen (approval number FEB-20250424 - 01512).

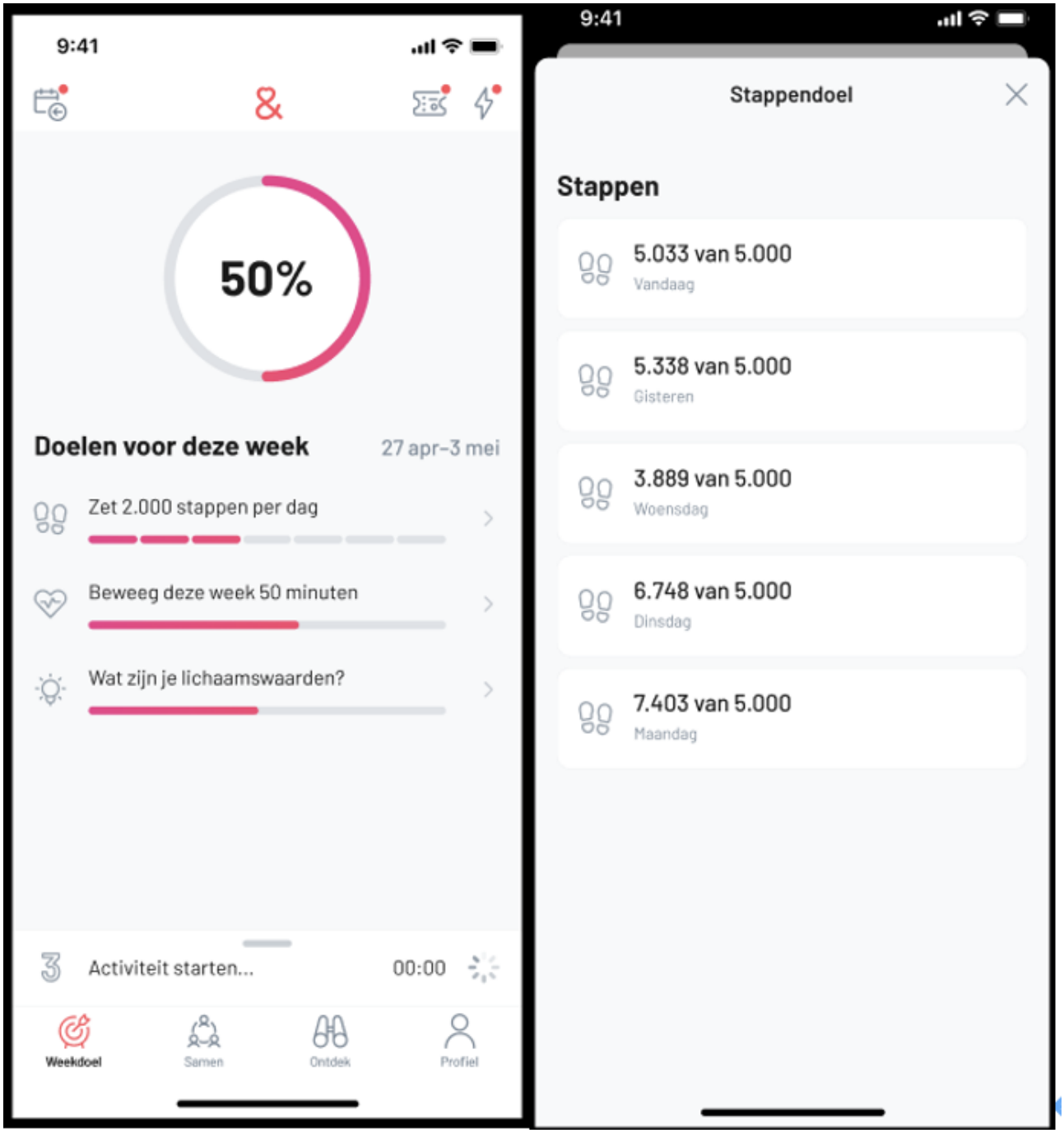
Study Design

SamenGezond (which translates to "healthy together") is a health app based in the Netherlands, launched in October 2017. Similar to other mHealth apps, it aims to enhance users' health through features such as goal management, GPS-enabled physical activity tracking, and personalized feedback. In addition, the app offers access to health-related papers, expert advice, support from a web-based coach, and opportunities to join exercise communities. Each user is initially assigned a default step goal of 2000 steps per day, 5 days a week. As shown in Figure 1, the app's main page (left panel) displays a circular progress indicator representing the percentage of the user's weekly goal achieved. The circle is color-coded according to the completion level. Below this indicator, the user's step goal is shown, translated as "Take 2000 steps per day," followed by a progress bar indicating weekly performance. Users can also view their total number of steps for each day and the step goal for the week on a separate interface (right panel of Figure 1).

To obtain a valid sample for the study, we used 3 criteria to filter the users in the database. The criteria were, first, that the user had to be a member of the app for at least 4 weeks to ensure some experience with the app; otherwise, the participants would not be able to answer questions on their experience with the step goal. Second, the user must have used this app in the last year because historical exercise data are required to generate a new step goal for users. Third, the users must have subscribed to email communication to ensure the survey can be sent. After applying these criteria, we used the R package dplyr to randomly choose users in the database, ensuring that each user had an equal probability of being included in the sample. Specifically, we used the sample_frac function to select 10% of the users for group 1 and the sample_n function to select users for group 2. As a result, 3800 users were chosen for group 1 and 2000 users were chosen for group 2.

XSL•FO
RenderX

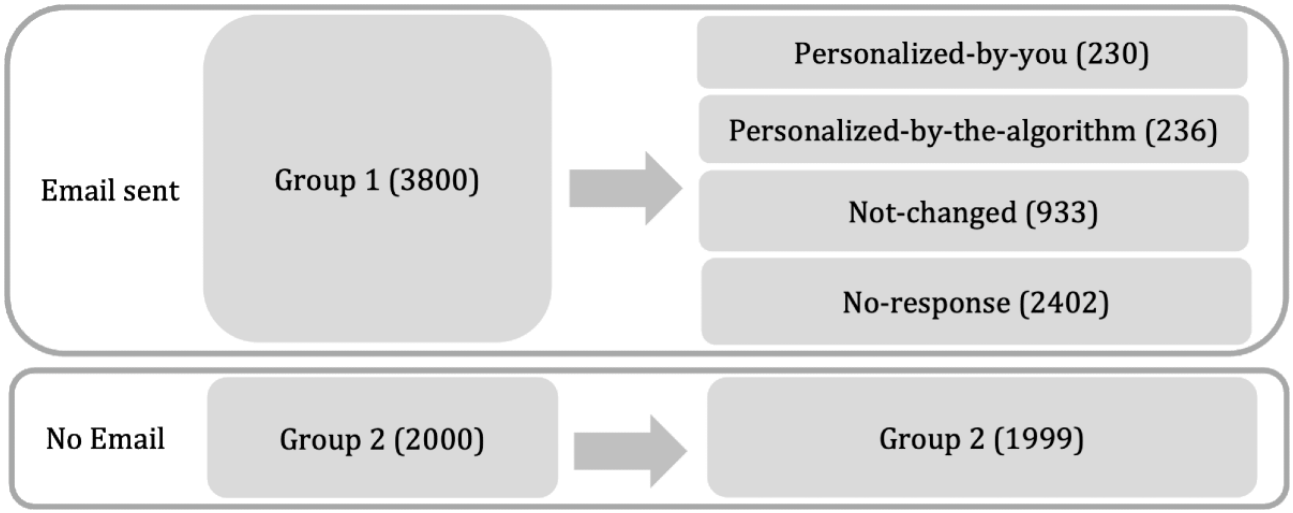
Figure 1. Screenshots of the (Dutch language) mobile health app.



The field study was conducted from week 8 to week 11 in 2023 (February 20 to March 19). A total of 5800 users who met our criteria were randomly chosen from the database by their user ID. In total, 3800 users were selected for group 1 and 2000 users for group 2. At the end of week 7, users in group 1 received an email containing a survey that asked whether they were satisfied with their current step goal. Users who were not satisfied, indicating the goal is "Too high" or "Too low," could choose to personalize their goal either by themselves or through the app (see part A of Multimedia Appendix 1 for detailed information on the survey). A total of 1399 users responded, resulting in 3 subgroups: 230 users personalized their goals themselves (personalized-by-you group), 236 users had their goals personalized by the app (personalized-by-the-algorithm group), and 933 users were satisfied with their current goal and thought their current step goal was "All right" and chose to retain the default goal (not-changed group), resulting in 3 treatment groups. The remaining 2402 users did not respond to the survey (no-response group). Users in group 2 did not receive this email. The procedure is illustrated in Figure 2. Note that when scrutinizing the data, we found that 1 participant from group 2 received and completed the survey due to a technical error. Since they chose to keep the default goal, they were classified into the not-changed group, resulting in 1999 users in the no-intervention group and 933 users in the not-changed group.

XSL•FO
RenderX

Figure 2. Experimental procedure.



In the personalized-by-you group, users set their goals within a reference range of 1000 steps to 20,000 steps. In the personalized-by-the-algorithm group, step goals were generated based on each participant's daily step count over the 4 weeks preceding the study. Not all participants have 28 observations. For example, if users only use the app for 3 days during the previous 4 weeks before the experiment, then they only have 3 observations. As the mean steps per day do not necessarily represent a "normal day" due to outliers, we adopted a moving-window percentile rank algorithm [22,24]. Specifically, step counts were sorted in ascending order for each participant, and the observation just below the median was selected as the new step goal. For example, for a participant with 10 observations, the fourth in rank was selected, and for a participant with 7 observations, the third in rank was selected. Then, we rounded the chosen step counts to the nearest hundred as users' step goals. We set limits on the new goal generated through this procedure. The users had to have at least 7 daily observations to generate a new goal based on their historical exercise. For those users, the algorithm constrains the goals from 2000 to 8000 steps. For example, for a user with 28 observations, if the 13th observation is over 8000, then the goal would be set as 8000. For participants with fewer than 7 observations, the goals were assigned to them based on the following rule: those who indicated that the default goal of 2000 steps was too high received a goal of 1500 steps, while those who felt it was too low were assigned 3500 steps, each for 5 days a week.

Statistical Analysis

For all of the users in our study (N=5800), we have their data on anonymous ID, group information, step goal, age, gender, app usage duration (measured as the total number of days they have registered as a member of this app), total steps between weeks 6 and 11, and the number of days the goal was achieved during that period. To prepare the dataset for analysis, we imputed missing age values using group means, categorized missing gender as "unknown," and replaced extreme outliers in total steps (eg, over 20 million steps per week) with the group mean for the respective week. The extreme outliers were caused by system error. An adult can only walk up to 560,000 steps a week if they walk 10 hours for 7 days based on the average walking speed. A total of 560,000 steps is far lower than 20 million steps.

A major challenge in identifying the causal effect of goal personalization on physical activity lies in the self-selection of users into different groups. The decision to personalize goals might be influenced by factors such as prior physical activity levels, app usage duration, and other individual characteristics. For instance, more active users may be more likely to engage in goal personalization. To address this issue, we utilize the propensity score matching (PSM) technique together with the difference-in-difference (DID) model, following previous research [30,31]. PSM was used to construct 3 matched control groups corresponding to the personalized-by-you, personalized-by-the-algorithm, and not-changed groups. The DID method then compares the pre- and posttreatment differences between each treatment group and its matched control group [32]. This approach helps to control the impact of exogenous and time-varying factors, such as weather, that may affect all users [32].

To ensure a sufficiently large matching pool for PSM, we combined the no-response group and group 2 (Figure 2), a total of 4401 users, as the matching group. Propensity scores were estimated using a logistic regression model based on users' age, gender, app usage duration, total steps in weeks 6 and 7, and the number of days the step goal was achieved in those weeks. We then applied 1:1 nearest neighbor matching to construct comparable control groups for each of the 3 groups. To evaluate the quality of PSM, we checked whether the variables were balanced between each treatment group and its matched control group. The mean values of the covariates, such as the mean age and gender proportion, are closely aligned across the matched pairs, confirming the success of the PSM procedure.

To estimate the effect of goal personalization on physical activity, as measured by total steps walked per week, our DID model specification is as follows:

$$(1)\text{Stepsit}=\beta 0+\beta 1\text{Personalizationit}+\theta i+\lambda t+\varepsilon it$$

In equation 1, Steps_{it} denotes the weekly total number of steps taken by user i in week t, and $\text{Personalization}_{it}$ is a binary variable indicating whether the users i personalized their goal

XSL•FO
RenderX

in week *t*. We also included individual fixed effects (θ_i) to control for unobserved heterogeneity across users and time fixed effects (λ_t) to account for exogenous influence (eg, weather). We estimated the DID model 3 times, each time comparing 1 of the 3 groups with its corresponding matched control group.

The identification of the goal personalization effect using a DID model relies on the common trend assumption. To test whether the common trend assumption is met, we split the personalization effect into different weeks by including the interactions of personalization and week. The findings indicate that there is no significant difference between week 6 and week 7 across all 3 groups, supporting the common trend assumption (shown in part B of Multimedia Appendix 1).

Results

User Statistics

The overall mean age of the 5800 users was 54.86 years (SD 10.42), with 60% (n=5800) being women. The average app usage duration was 813.77 days (SD 536.42). Tables 1-3 present detailed characteristics of users across the personalized-by-you, personalized-by-the-algorithm, and not-changed groups and their matched control groups.

Table . Summary statistics and covariate comparison before and after matching for the personalized-by-you group.

	Personalized-by-you group	Matching group		Matched group	
	Values	Values	Mean difference[a]	Values	Mean difference[a]
Age (y), mean (SD)	52.7 (10.945)	54.786 (10.015)	–0.191	52.752 (12.225)	–0.005
Gender, n (%)					
Man	57 (24.8)	1202 (27.3)	–0.059	48 (20.9)	0.091
Unknown	8 (3.5)	674 (15.3)	–0.646	7 (3.0)	0.024
Woman	165 (71.7)	2525 (57.4)	0.319	175 (76.1)	–0.097
App usage duration (day), mean (SD)	903.596 (521.482)	780.913 (542.465)	0.235	872.213 (532.278)	0.06
Total steps in week 6, mean (SD)	48,527.896 (26,162.616)	14,648.264 (21,554.015)	1.295	47,255.104 (26,591.905)	0.049
Total steps in week 7, mean (SD)	48,593.409 (27,125.127)	13,269.404 (21,143.616)	1.302	48,968.865 (26,984.531)	–0.014
Days goal achieved in week 6, mean (SD)	5.983 (1.613)	2.094 (2.716)	2.411	5.835 (1.885)	0.092
Days goal achieved in week 7, mean (SD)	5.787 (1.820)	1.896 (2.673)	2.138	5.835 (1.642)	–0.026

[a]Mean difference denotes the standard mean difference.

XSL•FO
RenderX

Table . Summary statistics and covariate comparison before and after matching for the personalized-by-the-algorithm group.

	Personalized-by-the-algorithm group	Matching group		Matched group	
	Values	Values	Mean difference[a]	Values	Mean difference[a]
Age (y), mean (SD)	51.275 (11.944)	54.786 (10.015)	–0.294	50.915 (12.910)	0.030
Gender, n (%)					
Man	56 (23.7)	1202 (27.3)	–0.084	54 (22.9)	0.020
Unknown	7 (3.0)	674 (15.3)	–0.728	6 (2.5)	0.025
Woman	173 (73.3)	2525 (57.4)	0.360	176 (74.6)	–0.029
Registration period (day), mean (SD)	779.102 (503.97)	780.913 (542.465)	–0.004	796.254 (557.702)	–0.034
Total steps in week 6, mean (SD)	48,336.182 (23,668.516)	14,648.264 (21,554.015)	1.423	45,666.670 (22,468.613)	0.113
Total steps in week 7, mean (SD)	49,459.788 (24,121.172)	13,269.404 (21,143.616)	1.500	46,349.390 (23,702.009)	0.129
Days goal achieved in week 6, mean (SD)	6.093 (1.432)	2.094 (2.716)	2.793	6.114 (1.467)	–0.015
Days goal achieved in week 7, mean (SD)	6.042 (1.467)	1.896 (2.673)	2.827	6.081 (1.449)	–0.026

[a]Mean difference denotes the standard mean difference.

Table . Summary statistics and covariate comparison before and after matching for the not-changed group.

	Not-changed group	Matching group		Matched group	
	Values	Values	Mean difference[a]	Values	Mean difference[a]
Age (y), mean (SD)	56.656 (11.399)	54.786 (10.015)	0.164	56.329 (11.512)	0.029
Gender, n (%)					
Man	297 (31.8)	1202 (27.3)	0.097	297 (31.8)	0
Unknown	26 (2.8)	674 (15.3)	–0.761	31 (3.3)	–0.033
Woman	610 (65.4)	2525 (57.4)	0.168	605 (64.8)	0.011
Registration period (day), mean (SD)	955.374 (492.763)	780.913 (542.465)	0.354	891.212 (536.883)	0.130
Total steps in week 6, mean (SD)	39,582.826 (20,642.639)	14,648.264 (21,554.015)	1.208	40,006.05 (20,862.073)	–0.021
Total steps in week 7, mean (SD)	38,890.994 (19,347.041)	13,269.404 (21,143.616)	1.324	40,306.503 (19,527.440)	–0.073
Days goal achieved in week 6, mean (SD)	5.727 (1.586)	2.094 (2.716)	2.290	5.721 (1.634)	0.003
Days goal achieved in week 7, mean (SD)	5.755 (1.487)	1.896 (2.673)	2.595	5.827 (1.43)	–0.049

[a]Mean difference denotes the standard mean difference.

Main Results

The results in Table 4 show a significant increase in physical activity among users in all 3 groups compared to their respective matched control groups. On average, users in the personalized-by-you group increased their weekly steps by 3793 (P<.001) from week 8 to week 11 after the personalization, compared to 2 weeks prior to the personalization. This corresponds to an average increase of 542 steps per day. Those in the personalized-by-the-algorithm group showed an increase of 4315 (P<.001) weekly steps (616 steps per day). These results demonstrate that both personalization strategies are effective in increasing physical activity levels. There is no significant difference between the overall effects of the 2 personalization strategies (t=0.354, P=.72). Users in the not-changed group, who did not alter their goals, also demonstrated an increase of 1759 (P<.001) per week (251 steps per day).

XSL•FO
RenderX

Table . The estimated effects of goal personalization on physical activity[a].

	Personalized-by-you, steps	Personalized-by-the-algorithm, steps	Not-changed, steps
Personalization	3793.229[b] (–1077.40)	4315.046[b] (–994.93)	1758.642[b] (–462.033)
Individual fixed effect	Yes	Yes	Yes
Week fixed effect	Yes	Yes	Yes
Observations	2760	2832	11,196
Individuals	460	472	1866

[a]Heteroskedasticity-robust standard errors are given in parentheses.

[b]*P*<.001.

Heterogeneity by Prior Activity Levels

To assess the potential heterogeneity in the goal personalization effect by prior activity, we estimate specification (equation 1) on subsamples classified according to total steps walked in week 6 and week 7. Specifically, the users in each of the 2 personalization groups were further divided into 3 groups: the first quartile (low-active users), the second and third quartiles (medium-active users), and the fourth quartile (high-active users), based on their step counts prior to the study.

The results reveal significant differences in the effects of the personalized-by-you and personalized-by-the-algorithm strategies across activity levels. Among low-active users, the personalized-by-you strategy showed no significant effect (*P*=.82, see Table 5, column 1), whereas the personalized-by-the-algorithm strategy demonstrated significant effectiveness, increasing weekly total steps by 5094 from week 8 to week 11 after the personalization, compared to 2 weeks prior to the personalization (*P*=.003; see Table 5, column 4). For medium-active users, both strategies are similarly effective: personalized-by-you increased weekly steps by 5841 (*P*<.001) and personalized-by-the-algorithm by 5278 (*P*<.001) over a 4-week period after the personalization (see Table 5, columns 2 and 5). In contrast, for high-active users, personalized-by-the-algorithm was not effective (*P*=.51, see Table 5, column 6), while personalized-by-you proved to be marginally effective (*P*=.09; see Table 5, column 3).

Table . Heterogeneous effects of personalization by prior activity[a].

	Personalized-by-you, steps			Personalized-by-the-algorithm, steps		
	Low-active users	Medium-active users	High-active users	Low-active users	Medium-active users	High-active users
Personalization	383.641 (–1694.72)	5841.661[b] (–1476.70)	4266.063[c] (–2598.36)	5093.521[d] (–1766.61)	5278.006[b] (–1476.70)	1446.37 (–2194.72)
Individual fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Week fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	828	1212	720	576	1572	684
Individuals	138	202	120	96	262	114

[a]Heteroskedasticity-robust SEs are given in parentheses.

[b]*P*<.001.

[c]*P*<.10.

[d]*P*<.01.

Robustness Check

As robustness checks, we used different matching samples for PSM. In our main study, we combined both users in the no-response group and users from group 2 as the matching sample. For robustness checks, we treated these 2 groups separately as distinct matching samples. We then reestimated the same DID models. The results remain robust, with all 3 groups (personalized-by-you, personalized-by-the-algorithm, and not-changed groups) showing significant effects (shown in parts C and D in Multimedia Appendix 1).

Discussion

Principal Results

Given the potential benefits of goal personalization on physical activity and the lack of solid evidence in previous literature, a large-scale field study was conducted to examine the potential of personalization in prompting health behavior. Our findings demonstrate that goal personalization effectively increased physical activity, resulting in an average increase of over 500 daily steps over a 4-week period. This increase is meaningful at the individual level, as prior meta-analytic evidence based on over 200,000 participants indicates that an additional 500

XSL•FO
RenderX

steps per day is associated with a 7% reduction in cardiovascular-related mortality [33]. Our findings also demonstrate that even participants who only completed the survey but did not change their goal increased their weekly steps by 1759 over a 4-week period. The increase may be attributed to a reminder effect from receiving the email and completing the survey [34,35]. However, post hoc analyses reveal that users in both goal personalization groups take significantly more steps than those in the not-changed groups. This suggests that the observed increase in steps is not solely due to the reminder effect. Specifically, compared with the unchanged group, the users in the personalized-by-you group showed a marginally significant increase (P=.09), while the users in the personalized-by-the-algorithm group showed a significant increase (P=.02).

Moreover, this study also compared the effectiveness of 2 personalization strategies. While the overall effects of these 2 strategies are similar, heterogeneity tests based on prior activity levels indicate varying effectiveness across distinct groups. The personalized-by-you strategy proves effective for medium and high-active users, whereas the personalized-by-the-algorithm significantly boosts total steps for low-active and medium-active users.

We argue the potential reason why personalized-by-the-algorithm is effective for low-active users, while personalized-by-you is not, is as follows. Low-active users may overestimate their physical abilities when personalizing the goal by themselves, leading to goals that are less realistic and motivating. In contrast, goals generated by the algorithm are based on users' previous exercise data, likely aligning better with their capabilities. This alignment enhances individuals' self-accountability and competence and drives better health outcomes [36]. A post hoc analysis shows that low-active users in the personalized-by-you group set higher goals than those in the personalized-by-the-algorithm group—3730 steps versus 2844 steps—a difference that is marginally significant (P=.07). This finding underscores the importance of personalized-by-the-algorithm in setting realistic and motivating goals for low-active users.

The reason why personalized-by-you is effective for highly active users, while personalized-by-the-algorithm is not, is as follows. High-active users, being more experienced with exercise, may be capable of setting goals that better align with their ability, situation, and standard than the algorithm. Further analysis supports this, showing that high-active users in the personalized-by-you group set lower, more realistic goals (5700 steps on average) compared to those in the personalized-by-the-algorithm group (7061 steps on average; P=.002). The higher goals set by the algorithm might be too challenging and therefore demotivating these users.

These findings advance goal-setting theory in 3 ways. First, prior research provides mixed guidance on goal type: while the SMART framework emphasizes specific and attainable goals [12,13], other work stresses specific and challenging goals [14,15]. Both approaches have limitations: too-easy goals may fail to motivate meaningful health behavior. This study indeed shows that the default 2000-step goal was less motivating than the average 4747-step goal set in the personalized groups. Conversely, difficult goals can discourage users [11,14], as this study shows that the average 7061-step goals for high-active users in the personalized-by-the-algorithm group only had a similar effect as the 2000-step goal. Our study highlights the value of personalized goals that align with individuals' abilities, characteristics, and contexts. Second, beyond goal type, the study contributes to how to best set the goals by identifying 2 distinct goal personalization strategies. Third, the study examines the moderating role of prior activity level, offering theoretical insights into the conditions under which different personalization strategies are most effective. In sum, the findings extend goal-setting theory by clarifying what types of goals should be set, how goals can be personalized (methods), and when different personalization strategies are most effective (boundary conditions).

As for practical contribution, given the findings, mHealth technology companies should prioritize algorithmic personalization or, at the very least, frame the choices of personalization in a manner that encourages low-active users to opt for algorithm-based personalization. However, for high-active users, it is important to give them the opportunity to personalize their goals by themselves, as they are more motivated by self-control, autonomy, and self-accountability and are better equipped to set goals that align with their abilities.

Limitations

This research has limitations that warrant consideration. First, the participants were not randomly assigned to different treatment groups but made the decision by themselves, which could introduce potential biases. Although we employ PSM and DID to address the issue, only a randomized field experiment can fully mitigate such biases. Second, the lack of postexperiment data prevents us from assessing the long-term sustainability of the goal personalization effect. Third, we do not have data to test whether the goal personalization effect extends to other health-related activities, such as gym attendance or weight management, which limits the generalizability of our findings. Fourth, this study compares personalized goals to a fixed 2000-step goal, without considering other types of goals (eg, 5000-step goal), which may constrain the scope of the comparison. Finally, there are other goal personalization algorithms, such as machine learning approaches that predict step goals using previous step data and goal histories [23], or simple goal adjustment approaches based on goal achievement [37,38]. Future studies can investigate the comparative effectiveness of these goal personalization methods. Furthermore, more intelligent algorithms that incorporate contextual factors—such as users' moods and weather conditions—should be developed and empirically tested in future research to examine whether they can generate more motivating personalized goals and encourage users to be more physically active.

Conclusions

This study investigated the effects of 2 goal personalization strategies on increasing users' physical activity levels through an open-access mHealth app. The results show that both strategies significantly promote physical activity—by over 500

XSL•FO
RenderX

additional steps per day—with improvements that are meaningful at the individual level, particularly in reducing health risks. Moreover, the study identified the applicability of each personalization strategy: the personalized-by-you approach is more effective for medium- and high-active users, while the personalized-by-the-algorithm strategy works better for low- and medium-active users. These insights can inform the design of the goal management features in mHealth apps to enhance the effectiveness of health interventions.

Funding

This research did not receive any funding.

Data Availability

The dataset supporting the findings of this study is available from the first author upon reasonable request due to proprietary company restrictions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey instruments, goal personalization effects, and robustness checks across sample variations.
[DOCX File, 100 KB - mhealth_v14i1e81779_app1.docx]

References

1. Global status report on physical activity 2022. : World Health Organization; 2022 URL: https://iris.who.int/server/api/core/bitstreams/8804f1b0-dbae-4e58-a251-36fd14dc7e02/content [accessed 2026-02-07]
2. Hydari MZ, Adjerid I, Striegel AD. Health wearables, gamification, and healthful activity. Manage Sci 2023 Jul;69(7):3920-3938. [doi: 10.1287/mnsc.2022.4581] [Medline: 37547027]
3. Liu Y, Jiang Z, Choi BCF. Pushing yourself harder: the effects of mobile touch modes on users' self-regulation. Inf Sys Res 2023 Sep;34(3):996-1016. [doi: 10.1287/isre.2022.1155]
4. de Vries HJ, Delahaij R, van Zwieten M, Verhoef H, Kamphuis W. The effects of self-monitoring using a smartwatch and smartphone app on stress awareness, self-efficacy, and well-being-related outcomes in police officers: longitudinal mixed design study. JMIR mHealth uHealth 2025 Jan 28;13(1):e60708. [doi: 10.2196/60708] [Medline: 39881435]
5. Estrin D, Sim I. Health care delivery. Open mHealth architecture: an engine for health care innovation. Science 2010 Nov 5;330(6005):759-760. [doi: 10.1126/science.1196187] [Medline: 21051617]
6. Constantiou I, Mukkamala A, Sjöklint M, Trier M. Engaging with self-tracking applications: how do users respond to their performance data? Eur J Inf Syst 2023 Nov 2;32(6):941-961. [doi: 10.1080/0960085X.2022.2081096]
7. Ghose A, Guo X, Li B, Dang Y. Empowering patients using smart mobile health platforms: evidence from a randomized field experiment. MIS Q 2022 Mar 1;46(1):151-192. [doi: 10.25300/MISQ/2022/16201]
8. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. JAMA 2015 Feb 3;313(5):459-460. [doi: 10.1001/jama.2014.14781] [Medline: 25569175]
9. Milyavskaya M, Werner KM. An integrative model of goal pursuit. PsyArXiv. Preprint posted online on Sep 20, 2021. [doi: 10.31234/osf.io/qydpv]
10. McEwan D, Harden SM, Zumbo BD, et al. The effectiveness of multi-component goal setting interventions for changing physical activity behaviour: a systematic review and meta-analysis. Health Psychol Rev 2016;10(1):67-88. [doi: 10.1080/17437199.2015.1104258] [Medline: 26445201]
11. Swann C, Rosenbaum S, Lawrence A, Vella SA, McEwan D, Ekkekakis P. Updating goal-setting theory in physical activity promotion: a critical conceptual review. Health Psychol Rev 2021 Mar;15(1):34-50. [doi: 10.1080/17437199.2019.1706616] [Medline: 31900043]
12. Swann C, Jackman PC, Lawrence A, et al. The (over)use of SMART goals for physical activity promotion: a narrative review and critique. Health Psychol Rev 2023 Jun;17(2):211-226. [doi: 10.1080/17437199.2021.2023608] [Medline: 35094640]
13. Latham GP. Goal setting: a five-step approach to behavior change. Organ Dyn 2003;32(3):309-318. [doi: 10.1016/S0090-2616(03)00028-7]
14. Mann T, de Ridder D, Fujita K. Self-regulation of health behavior: social psychological approaches to goal setting and goal striving. Health Psychol 2013 May;32(5):487-498. [doi: 10.1037/a0028533] [Medline: 23646832]
15. Linde JA, Jeffery RW, Finch EA, Ng DM, Rothman AJ. Are unrealistic weight loss goals associated with outcomes for overweight women? Obes Res 2004 Mar;12(3):569-576. [doi: 10.1038/oby.2004.65] [Medline: 15044676]
16. Kankanhalli A, Xia Q, Zhao X, Ai P. Understanding personalization for health behavior change applications: a review and future directions. AIS Trans Hum Comput Interact 2021;13(3):316-349. [doi: 10.17705/1thci.00152]

XSL•FO
RenderX

17. Chua RYJ, Iyengar SS. Empowerment through choice? A critical analysis of the effects of choice in organizations. Res Organ Behav 2006 Jan;27:41-79. [doi: 10.1016/S0191-3085(06)27002-3]
18. Deci EL, Ryan RM. The support of autonomy and the control of behavior. J Pers Soc Psychol 1987;53(6):1024-1037. [doi: 10.1037/0022-3514.53.6.1024]
19. Legault L, Inzlicht M. Self-determination, self-regulation, and the brain: autonomy improves performance by enhancing neuroaffective responsiveness to self-regulation failure. J Pers Soc Psychol 2013 Jul;105(1):123-138. [doi: 10.1037/a0030426] [Medline: 23106250]
20. Muraven M, Gagné M, Rosman H. Helpful self-control: autonomy support, vitality, and depletion. J Exp Soc Psychol 2008 May;44(3):573-585. [doi: 10.1016/j.jesp.2007.10.008] [Medline: 18496610]
21. Muraven M. Autonomous self-control is less depleting. J Res Pers 2008;42(3):763-770. [doi: 10.1016/j.jrp.2007.08.002] [Medline: 18704202]
22. Adams MA, Sallis JF, Norman GJ, Hovell MF, Hekler EB, Perata E. An adaptive physical activity intervention for overweight adults: a randomized controlled trial. PLoS One 2013;8(12):e82901. [doi: 10.1371/journal.pone.0082901] [Medline: 24349392]
23. Zhou M, Fukuoka Y, Mintz Y, et al. Evaluating machine learning-based automated personalized daily step goals delivered through a mobile phone app: randomized controlled trial. JMIR mHealth uHealth 2018 Jan 25;6(1):e28. [doi: 10.2196/mhealth.9117] [Medline: 29371177]
24. Adams MA, Hurley JC, Todd M, et al. Adaptive goal setting and financial incentives: a 2 × 2 factorial randomized controlled trial to increase adults' physical activity. BMC Public Health 2017 Mar 29;17(1):286. [doi: 10.1186/s12889-017-4197-8] [Medline: 28356097]
25. Shilts MK, Horowitz M, Townsend MS. Goal setting as a strategy for dietary and physical activity behavior change: a review of the literature. Am J Health Promot 2004;19(2):81-93. [doi: 10.4278/0890-1171-19.2.81] [Medline: 15559708]
26. Williamson O, Swann C, Bennett KJM, et al. The performance and psychological effects of goal setting in sport: a systematic review and meta-analysis. Int Rev Sport Exerc Psychol 2024 Dec 31;17(2):1050-1078. [doi: 10.1080/1750984X.2022.2116723]
27. Hall HK, Kerr AW. Goal setting in sport and physical activity: tracing empirical developments and establishing conceptual direction. In: Roberts GG, editor. Advances in Motivation in Sport and Exercise: Human Kinetics Publisher; 2001:183-234 URL: https://www.scribd.com/document/695639286/Glyn-Roberts-Advances-in-Motivation-in-Sport-Exercise-Human-Kinetics-2001 [accessed 2026-02-06]
28. Jeong YH, Healy LC, McEwan D. The application of Goal Setting Theory to goal setting interventions in sport: a systematic review. Int Rev Sport Exerc Psychol 2023 Dec 31;16(1):474-499. [doi: 10.1080/1750984X.2021.1901298]
29. Lambert SM, Moore DW, Dixon RS. Gymnasts in training: the differential effects of self-and coach-set goals as a function of locus of control. J Appl Sport Psychol 1999 Mar;11(1):72-82. [doi: 10.1080/10413209908402951]
30. Gu X, Kannan PK. The dark side of mobile app adoption: examining the impact on customers' multichannel purchase. J Mark Res 2021 Apr;58(2):246-264. [doi: 10.1177/0022243720988257]
31. Li X, Liao C, Xie Y. Digital piracy, creative productivity, and customer care effort: evidence from the digital publishing industry. Mark Sci 2021 Jul;40(4):685-707. [doi: 10.1287/mksc.2020.1275]
32. Angrist JD, Pischke JS. Mostly Harmless Econometrics: An Empiricist's Companion: Princeton University Press; 2009. [doi: 10.2307/j.ctvcm4j72]
33. Banach M, Lewek J, Surma S, et al. The association between daily step count and all-cause and cardiovascular mortality: a meta-analysis. Eur J Prev Cardiol 2023 Dec 21;30(18):1975-1985. [doi: 10.1093/eurjpc/zwad229] [Medline: 37555441]
34. Calzolari G, Nardotto M. Effective reminders. Manage Sci 2017 Sep;63(9):2915-2932. [doi: 10.1287/mnsc.2016.2499]
35. Agachi E, Bijmolt THA, van Ittersum K, Mierau JO. The effect of periodic email prompts on participant engagement with a behavior change mHealth app: longitudinal study. JMIR mHealth uHealth 2023 May 11;11:e43033. [doi: 10.2196/43033] [Medline: 37166974]
36. Deci EL, Ryan RM. The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 2000 Oct;11(4):227-268. [doi: 10.1207/S15327965PLI1104_01]
37. Waki K, Tsurutani Y, Waki H, et al. Efficacy of StepAdd, a personalized mHealth intervention based on social cognitive theory to increase physical activity among patients with type 2 diabetes mellitus: protocol for a randomized controlled trial. JMIR Res Protoc 2024 Feb 23;13(1):e53514. [doi: 10.2196/53514] [Medline: 38393770]
38. Sze WT, Waki K, Enomoto S, et al. StepAdd: a personalized mHealth intervention based on social cognitive theory to increase physical activity among type 2 diabetes patients. J Biomed Inform 2023 Sep;145:104481. [doi: 10.1016/j.jbi.2023.104481] [Medline: 37648101]

Abbreviations

DID: difference-in-difference
mHealth: mobile health
PSM: propensity score matching

XSL•FO
RenderX

SMART: specific, measurable, achievable, realistic, time-bound

Edited by G Badicu; submitted 03.Aug.2025; peer-reviewed by E Hargreaves, I Coppens; accepted 10.Dec.2025; published 18.Feb.2026.

Please cite as:
Liu X, Bijmolt THA, Leliveld MC, Noppers EH
Effectiveness of Step Goal Personalization Strategies on Physical Activity in a Mobile Health App: A Field Study
JMIR Mhealth Uhealth 2026;14:e81779
URL: https://mhealth.jmir.org/2026/1/e81779
doi:10.2196/81779

© Xia Liu, Tammo H A Bijmolt, Marijke C Leliveld, Ernst H Noppers. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 18.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Assessing Wearable mHealth Adherence in Underserved Adolescents and its Associations With Physical Activity, Sports, and Safety Perceptions: Prospective Cohort Study

Annabel Nunez-Gaunaurd[1*], PT, PhD; Michele Raya[2*], PT, PhD

[1]Department of Physical Therapy, Nicole Wertheim College of Nursing and Health Sciences, Florida International University, 11200 SW 8 St, AHC3-424, Miami, FL, United States

[2]Department of Physical Therapy, Miller School of Medicine, University of Miami, Coral Gables, FL, United States

[*]all authors contributed equally

Corresponding Author:
Annabel Nunez-Gaunaurd, PT, PhD
Department of Physical Therapy, Nicole Wertheim College of Nursing and Health Sciences, Florida International University, 11200 SW 8 St, AHC3-424, Miami, FL, United States

Abstract

Background: Adolescents from underserved communities, particularly Black and Hispanic youth, engage in lower levels of physical activity (PA), increasing their risk for chronic disease. Conventional interventions often face barriers such as limited access to safe environments. Wearable mobile health technologies offer scalable and context-sensitive solutions; however, predictors of sustained adherence in school-based settings among high-risk populations remain underexplored.

Objective: This study aims to examine the behavioral and contextual predictors of adherence to a consumer-grade wearable PA tracker among underserved high school students.

Methods: In this school-based observational study, 63 students (mean age 14.8, SD 1.17 years) enrolled in physical education received Fitbit devices. Adherence was defined as ≥21 valid days of step count data. Measures included self-reported PA behaviors, neighborhood perceptions, physical fitness (including anthropometrics), and device adherence. Group comparisons were conducted using *t* tests and chi-square tests. Logistic regression was used to identify predictors of adherence.

Results: Overall, 73% (46/63) of participants met the adherence threshold. Adherent students reported fewer days of moderate-to-vigorous PA (2 vs 4 days/week; *P*=.004), lower team sports participation (21/46, 46% vs 12/17, 71%; *P*=.004), and higher perceived neighborhood safety (*P*=.02). In adjusted models, lower PA frequency, greater perceived safety, and neighborhood walkability significantly predicted adherence (χ^2_6=16.23*; P*=.01, Nagelkerke R^2=0.61).

Conclusions: Wearable mobile health technologies show promise for engaging underserved adolescents in PA, particularly those with lower baseline activity and limited access to structured sports. Key predictors of adherence included perceived neighborhood walkability, team sports participation, and prior PA behavior. School-based deployment of wearable devices should emphasize personalized goals and autonomy-supportive strategies to foster sustained engagement and promote PA among high-risk youth.

(JMIR Mhealth Uhealth 2026;14:e80465) doi:10.2196/80465

KEYWORDS

mHealth; physical activity; health promotion; obesity; wearable adherence; functional measures; mobile health; wearables; adolescents; school-based

Introduction

With national and international guidelines recommending at least 60 minutes of moderate-to-vigorous physical activity (MVPA) daily for children and adolescents, physical inactivity remains a pressing public health concern [1,2]. Regular physical activity (PA) is associated with numerous health benefits, including lower resting heart rate (HR), a key indicator of cardiovascular health [3,4], improved functional fitness [5], and enhanced long-term physical and mental well-being [6]. Despite these benefits, only 22% of adolescents meet the recommended PA guidelines [7], with participation rates significantly lower among Black, Hispanic, and immigrant youth [8]. Many children and adolescents in racially diverse and low-resource neighborhoods lack access to consistent and supportive PA opportunities outside school [9]. Furthermore, the growing financial and time demands placed on parents for youth sports participation [10] have made consistent PA opportunities

XSL•FO
RenderX

increasingly inaccessible for many adolescents, particularly those from underserved communities.

Consumer-grade wearable mobile health (mHealth) technologies may offer a scalable, user-friendly means to assess and promote PA engagement [11]. Awareness and use of step-based metrics from wearables can provide an opportunity for self-monitoring, goal setting, and motivation to increase PA behavior and improve health and fitness. For example, typically step-based metrics ranging from 7000 to 11,700 steps per day are associated with meeting Centers for Disease Control and Prevention (CDC)-recommended PA levels [12-14] and with reducing cardiovascular and mortality risks [15]. Conversely, taking fewer than 7000 steps per day may indicate insufficient PA behavior [14]. Wearables can help adolescents become more aware of their PA levels and facilitate tracking of these metrics [16]. There is growing evidence supporting the role of mHealth technologies, including smartphones, apps, and wearables, in promoting positive health behaviors and some improvement in weight status among youth [17]. These devices incorporate behavioral change techniques, such as goal setting, self-monitoring, and feedback [18], and may be particularly useful in school-based programs where environmental constraints can be addressed.

Traditional PA opportunities often rely on structured sports or clinic programs outside of school, which may be inaccessible due to socioeconomic and parental involvement time constraints, limited transportation, and neighborhood safety concerns for underserved youth [19]. Additionally, limited access to safe recreational environments and health-promoting infrastructure further compounds these risks, reinforcing poor PA engagement and long-term health outcomes [20,21]. These limited access to PA opportunities disproportionately affect adolescents from underserved communities, who are at increased risk for deficiencies in motor proficiency, strength, cardiovascular endurance, and overall PA levels, factors closely linked to obesity and poor fitness [22]. Over time, these disparities contribute to the early onset of chronic health conditions such as cardiovascular disease and type 2 diabetes [23]. In contrast, wearable mHealth technologies offer a scalable, low-cost alternative that can be integrated into adolescents' daily routines, making them particularly promising for populations with limited access to conventional resources.

School-based physical education (PE) programs are a critical platform to provide accessible PA interventions [24]. Wearables integrated into school-based programs can promote autonomy and engagement even in constrained environments. Importantly, low levels of teacher burden have previously been associated with favorable perceptions of feasibility, further supporting the scalability of wearable devices (WD) applicability in school settings [25]. Prior studies suggest that wearables may positively influence PA levels by enhancing self-efficacy and behavioral regulation among adults [26]. Additionally, accuracy and behavioral impact of WD in general have mostly been examined in the adult population [26,27]. Despite the growing popularity of consumer wearables, little is known about long-term adherence to wearable technologies among high school age adolescents from underserved racial and ethnic minority groups [11,28]. There is a significant gap in understanding adherence behavioral patterns among adolescents facing environmental barriers, such as unsafe neighborhoods, low PA engagement, and limited access to recreational spaces.

Recent shifts in PA research emphasize the importance of broader contextual factors, both environmental (ie, social environments, physical settings, and access to resources) and personal (ie, enjoyment, autonomy, and perceived competence), in shaping PA behavior [29]. For example, previous studies found that wearables may be less accessible or sustainable for adults in underserved settings owing to socioeconomic barriers, limited internet access, and concerns about privacy and crime safety [30]. Additionally, research on long-term adherence remains limited, particularly among marginalized youth [31,32]. These contextual challenges may influence how adolescents engage with digital health tools, and the findings from broader populations may not be generalizable to these groups.

Bandura's Social Cognitive Theory (SCT) provides a behavioral foundation for understanding adherence to WD use, emphasizing how personal factors (self-efficacy and motivation), behavioral patterns (consistent device use), and environmental influences (social support and accessibility) interact to influence health behaviors [33]. The International Classification of Functioning, Disability, and Health (adherence) complements this perspective by situating these behaviors within a biopsychosocial framework, capturing how an individual's body functions and structures, activities and participation, and contextual factors (personal and environmental) collectively shape overall functioning and engagement in daily life [34]. Together, SCT and the International Classification of Functioning and Disability (ICF) informed the selection of study variables, including self-efficacy, readiness for change, perceived neighborhood safety, and physical fitness and function, allowing for a nuanced exploration of WD adherence and PA engagement among underserved adolescents.

To address the gap in understanding the influence of contextual factors on WD adherence, this study investigated associations among contextual factors, including sociodemographic characteristics, environmental conditions, PA behavior, and adherence to WD, among a racially and socioeconomically diverse cohort of high school students enrolled in a school-based PE program. Guided by the ICF framework, we included predictors, such as self-efficacy, readiness for change, perceived neighborhood safety, and physical fitness, each representing key domains within the ICF [34]. These factors have been validated in prior studies and are particularly relevant to understanding adolescent engagement in PA as they reflect the complex interplay between individual capabilities, motivation, and contextual influences [35]. Specifically, we examine the following: 1. whether demographic, personal, and environmental factors influenced WD adherence; and 2. Whether WD adherence was associated with differences in self-reported PA, fitness, and functional performance outcomes. Our findings aim to inform the design of future school-based wearable interventions that effectively engage adolescents from historically marginalized populations and promote PA through sustained behavioral changes.

XSL•FO
RenderX

Methods

Overview

This prospective cohort study was conducted in collaboration with local nonprofit organizations and public-school administrators. Participants were recruited using a purposive sampling strategy designed to ensure the representation of racially and socioeconomically diverse adolescents in an underserved community. School-based physical therapists and PE teachers facilitated recruitment through in-class announcements and take-home informational flyers distributed during sessions of state-mandated Health Opportunities Through Physical Education (HOPE) courses. This course emphasizes lifelong fitness, healthy behavior, and personal responsibility [36]. All students enrolled in the HOPE were invited to participate, minimizing selection bias by including students regardless of their prior PA engagement or academic standing. To reduce sampling bias and ensure equitable participation, parent consent and student assent materials emphasized the voluntary nature of the study, clarified that participation would not affect academic standing, and explained the privacy protections in place. Materials were designed to be culturally appropriate and accessible, with bilingual staff available to address questions for parents or guardians. These steps were taken to minimize the likelihood that guardians would restrict participation based on personal beliefs, misconceptions about wearable technology, or concerns about data privacy.

Although the school served as the central hub for recruitment and data collection, participants wore WD during and outside school hours, enabling continuous monitoring of PA across multiple contexts, including home and community settings. Predictors of adherence, such as neighborhood safety, team sports participation, and self-reported PA behaviors, reflect influences beyond the school setting. Nevertheless, the school played a pivotal role in supporting this study through structured integration of the PE curriculum, device distribution, and ongoing student engagement.

Study Timeline

Data collection spanned the entire academic year, with baseline assessments conducted in the first quarter and follow-up assessments conducted in the final quarter. This extended duration enabled evaluation of adherence to WD use throughout the academic year.

Procedures

Baseline assessments were conducted over 2 PE class sessions. Day 1 included demographic and PA surveys, and day 2 involved physical performance testing. In collaboration with PE instructors and school-based physical therapists, students received individualized step-count goals, starting with a 10% weekly increase from their baseline average. PE teachers provided weekly reminders to wear and synchronize their Fitbit devices. Measures such as self-efficacy and readiness for change were selected based on their established predictive value in adolescent PA engagement, as supported by Bandura's SCT [33] and integration into the WHO-ICF framework. Environmental variables were included to reflect contextual influences consistent with ICF's socio-ecological approach. Table 1 provides a summary of variables assessed and associated theoretical construct.

XSL•FO
RenderX

Table . Summary of predictors, data sources, and theoretical frameworks.

Variables	Outcome measures	Theoretical framework
Demographics	• Survey during PE[a] class	• Personal factors (ICF)[b]
Self-report-physical activity level	• PACE+[c] survey [37,38] • Fitbit Flex [39,40]	• Activities and participation (ICF) • Behavioral capability and self-regulation (SCT)[d]
Physical activity level	• Fitbit Flex [39,40]	• Activities and participation (ICF) • Self-monitoring and reinforcement (SCT)
Wearable device adherence	• Fitbit Flex via Fitabase [39,40]	• Activities and participation (ICF) • Self-regulation reinforcement (SCT)
Self-efficacy and readiness for change	• PACE+ Survey [37,38,41].	• Personal factors (ICF) • Self-efficacy and expectations (SCT)
Social support and enjoyment	• PACE+ Survey [42]	• Environmental factors (ICF) • Social support and reinforcement (SCT)
MVPA frequency-self-report	• PACE+ Survey [42]	• Activities and participation (ICF) • Behavioral capability and self-regulation (SCT)
Team sports participation	• Youth Risk Behavior Survey [43-45]	• Activities and participation (ICF) • Observational learning and social support (SCT)
Neighborhood walkability	• NEWS-Y Scale [46]	• Environmental factors (ICF)
Anthropometrics	• Height, weight, and waist circumference [47,48]	• Body functions and structures (ICF)
Vital signs	• Heart rate and blood pressure [49]	• Body functions and structures (ICF)
Cardiovascular endurance	• 6-minute walk test [50,51].	• Body Functions and Activities (ICF)
Functional mobility	• Timed up and down stairs	• Activities and participation (ICF)
Balance	• Single-leg stance test [52]	• Body functions and activities (ICF)
Upper-body strength	• 90° push-ups [53].	• Body functions and activities (ICF)

[a]PE: physical education.

[b]ICF: International Classification of Functioning and Disability.

[c]PACE: Pollution Abatement Costs and Expenditures.

[d]SCT: Social Cognitive Theory.

WD Adherence

Participants received a Fitbit Flex activity tracker and information on device use, care, and data synchronization from school-based physical therapists. These devices contain accelerometers and have demonstrated moderate to excellent validity compared to research-grade accelerometers in youth [39,40]. The Fitbit data were uploaded to Fitabase, a research platform for remote device monitoring. The weekly averages of daily steps were captured using Fitbit Flex devices [54]. Adherence days were defined as days with at least 10 hours of HR monitoring data [55], and Adherence week was defined as a week with at least 3 valid days of WD use [41]. A 21-day period of WD use stratified participants as adherent or nonadherent WD users. This threshold was chosen based on prior studies that defined high adherence to WD use as achieving at least 70% of the recommended daily use over a 30 day period [56] and systematic review benchmarks of 10 hours per day criteria for high adherence to daily use [41]. These timeframes were selected to balance early feasibility with theoretical relevance, particularly in adolescent populations, where motivation and environmental influences may fluctuate.

PA and Psychosocial Measures

PA behavior and psychosocial constructs were assessed using a patient-centered assessment and counseling for exercise and

XSL•FO
RenderX

nutrition (Pollution Abatement Costs and Expenditures) survey. This validated instrument includes items on decisional balance, self-efficacy, enjoyment, and social support [37,38]. It has demonstrated acceptable reliability (ICC=0.60 - 0.82) and moderate validity with accelerometer-derived MVPA (r=0.27 - 0.40) [41]. Additionally, the Pollution Abatement Costs and Expenditures Survey assessed all participants on: days per week of 60 minutes of MVPA, readiness for change, and perceived peer or family support [42-44]. Lastly, team sports participation was measured using the 2013 Youth Risk Behavior Survey, which has shown substantial reliability (κ=0.61 - 0.91) [45].

Neighborhood Walkability

Perceived neighborhood environment was assessed using the Neighborhood Environment Walkability Scale for Youth (NEWS-Y) [46]. The subscales measure safety from crime, traffic hazards, and access to land. Items were rated on a 4-point Likert scale, with higher scores indicating lower walkability. The NEWS-Y has demonstrated good test-retest reliability (ICC=0.56 - 0.87) and validity, as evidenced by associations with walking behaviors among youth [46].

Clinical Assessments

To evaluate physical fitness and functional performance, a series of standardized assessments was administered during PE class sessions by trained physical therapists. These measures were selected based on their reliability, validity, and relevance to the adolescent population.

Anthropometrics and Vital Signs

Height, weight, and waist circumference data were collected using standardized procedures [47]. BMI was calculated using CDC growth charts. The waist-to-height ratio (WtHR) was calculated; a ratio ≥0.5 indicated elevated central adiposity [48]. Resting HR and blood pressure were measured using standard clinical procedures and equipment [49].

Functional Fitness Assessments

Cardiovascular endurance was assessed using the 6-minute walk test, a submaximal exercise test that measures the distance an individual can walk in 6 minutes. The 6-minute walk test is widely used in pediatric populations and has demonstrated excellent reliability (ICC=0.94) and strong validity for estimating aerobic capacity [50,51].

Functional mobility was assessed using the timed up and go test, which evaluates lower-extremity strength, coordination, and mobility. Participants were timed as they ascended and descended stairs. The timed up and down stairs test has shown excellent reliability (ICC=0.978 - 0.999) and correlates well with other mobility and balance measures [57].

Balance was assessed using the timed single-leg stance test, in which participants stood on one leg for as long as possible without support. This test provides insight into postural control and neuromuscular stability and demonstrates acceptable reliability in youth populations [52].

Upper-body strength and endurance were assessed using 90° push-ups, in which participants performed push-ups until exhaustion with proper form. This test is a reliable indicator of muscular endurance (ICC=0.93) and has demonstrated moderate validity when compared to bench press performance (r=0.64; males and r=0.28; females) [53].

Data Analysis

Statistical analyses were performed using SPSS v28. Descriptive statistics were used to characterize the participants. Chi-square and independent t tests were used to compare the adherence groups in terms of demographic and outcome variables. Pearson correlation coefficient was used to assess the association between PA behaviors and psychosocial variables. Relationships were classified by effect size: 0.00 - 0.25 (none), 0.25 - 0.50 (weak), 0.50 - 0.75 (moderate), and >0.75 (strong) [58]. Binomial logistic regression was used to examine the associations between WD adherence and demographic, behavioral, and environmental predictors, including PA level, obesity status, perceived neighborhood safety, and nativity. A priori power analysis using G*Power (v3.1.9.6) determined a minimum sample size of 57 to detect medium effect sizes (Cohen d=0.5, f^2=0.30) with α=.05 and 80% power.

Ethical Considerations

This study was approved by the Western Institutional Review Board and the Research Ethics Committee of the County School Board. Informed consent was obtained from all the participants and their guardians. The eligibility criteria were high school adolescents aged 14 - 18 years who were enrolled in a HOPE course. Baseline assessments were conducted in the first academic quarter, and follow-up assessments were conducted in the final quarter. Surveys and physical fitness tests were conducted during the PE class. Race, ethnicity, and nativity (US or foreign-born) were self-reported. All collected data were stored on secure, password-protected servers; identifiable information was removed prior to analysis to ensure participant privacy and confidentiality. No financial or academic compensation was provided, and participation had no effect on course grades.

Results

Cohort Characteristics

Participant characteristics by WD adherence (N=63) are presented in Table 2. Of the total sample, 73% (46/63) met the WD adherence criteria. The average age was 14.8 (SD 1.17) years, with the majority in ninth grade (51/63, 81%). More than half of the participants were female (35/63, 56%), and 73% (46/63) received free or reduced lunch. Most were Black or African Americans (40/63, 72%), followed by Hispanics (16/63, 42%). Sixteen percent (10/63) were foreign-born, and 37% (23/63) reported maternal nativity outside the United States. The average BMI and WtHR for the sample were 23.06 (SD 4.62) kg/m² and 0.45 (SD 0.07), respectively.

XSL•FO
RenderX

Table . Demographic characteristics by wearable adherence, using wearable device (WD) and not using wearable device (N-WD).

Demographic characteristics	Total (N=63)	WD (n=46)	N-WD (n=17)	*P* value[a]
Age (years), mean (SD)	14.79 (1.17)	14.67 (1.06)	15.11 (1.41)	.13
Sex, n (%)				.41
Male	28 (44.4)	19 (41)	9 (53)	
Female	35 (55.6)	27 (59)	8 (47)	
Grade level, n (%)				.20
9th	51 (81)	40 (87)	11 (65)	
10th	2 (3.2)	1 (2.2)	1 (5.9)	
11th	3 (4.8)	1 (2.2)	2 (11.11)	
12th	7 (11.1)	4 (8.9)	3 (16.7)	
Socioeconomics, n (%)				
Free or reduced lunch	46 (73)	—[b]	—	—
Race and ethnicity, n (%)				
Black or African American	40 (73)	28 (72)	12 (75)	.55
Hispanic	16 (42)	14 (47)	2 (25)	.25
Student birth country (non-US)	10 (16)	9 (21)	1 (7)	.24
Parents birth country (1 - 2, non-US)	26 (41)	22 (48)	4 (24)	.07
Maternal birth country, n (%)				—
United States	30 (48)	19 (41)	11 (65)	
Bahamas	1 (1.6)	1 (2.2)	—	
Brazil	3 (4.8)	3 (6.5)	—	
Dominican Republic	1 (1.6)	1 (2)	—	
Ecuador	1 (1.6)	0	1 (7)	
Germany	1 (1.6)	1 (2)	—	
Guatemala	1 (1.6)	1 (2)	—	
Haiti	8 (13)	7 (15)	1 (7)	
Honduras	1 (1.6)	1 (2)	—	
Jamaica	2 (3)	2 (4)	—	
Mexico	3 (4.8)	2 (4)	1 (7)	
Peru	1 (1.6)	1 (2)	—	
Country not reported	10 (16)	7 (15)	3 (18)	

[a]Chi-square *P* value of ≤.05 considered statistically significant

[b]Not applicable.

Anthropometric Characteristics

As detailed in Table 3, 19% of participants were classified as overweight (12/63) and 15.9% were classified as obese (10/63). Also, 21% (13/63) had a WtHR >0.5, indicating elevated cardiometabolic risk. Significant differences between the WD adherence groups were found in resting HR (76.4 vs 67.4 bpm; *P*=.009) and single-leg stance duration (57 vs 59 s; *P*=.03). No statistically significant differences were found in BMI, WtHR, blood pressure, 6 MWT, timed up and down stairs, push-ups, or sit-to-stand performance.

XSL•FO
RenderX

Table . Health outcomes characteristics by wearable adherence status, using wearable device (WD) and not using wearable device (N-WD).

Characteristics	Total (N=63)	WD (n=46)	N-WD (n=17)	*P* value[a]
Anthropometrics				
BMI (kg/m^2), mean (SD)	23.06 (4.62)	23.16 (4.9)	22.79 (3.9)	.78
BMI-for-age (percentile), mean (SD)	66.25 (9.02)	66.1 (30.7)	66.64 (24.6)	.96
Underweight (<5 percentile), n (%)	3 (4.8)	3 (6.5)	0 (0)	.38
Healthy weight (5-<85 percentile), n (%)	36 (57)	25 (54)	11 (65)	.38
Overweight (≧85 percentile but ≦95 percentile), n (%)	12 (19)	8 (17)	4 (24)	.38
Obese (≧95 percentile), n (%)	10 (15.9)	9 (20)	1 (6)	.38
WHtR[b], mean (SD)	0.45 (0.07)	0.46 (0.08)	0.44 (0.04)	.20
≦0.5 normal risk, n (%)	49 (77.8)	34 (74)	15 (88)	.09
>0.5 higher risk, n (%)	13 (21)	12 (26)	1 (6)	.09
Vitals, mean (SD)				
Resting HR[c] (beats/min)	73.95 (14.3)	76.4 (15.3)	67.4 (9.1)	.01
SBP[d] (mm Hg)	114.7 (14.4)	115.2 (14.7)	113.4 (14.1)	.68
DBP[e] (mm Hg)	73.34 (12.9)	74.2 (13.6)	71.1 (10.9)	.42
Performance health outcomes, mean (SD)				
6MWT[f] (m)	503.14 (62.72)	496.5 (66.1)	520.6 (50.6)	.19
SLS[g] (right+left, 60s max)	57.41 (6.21)	56.9 (6.9)	59.4 (1.4)	.03
Timed up and down stairs (s)	9.27 (3.68)	9.6 (4.1)	8.3 (1.5)	.23
Push-ups repetitions	13.77 (10.36)	12.6 (10.5)	18 (8.8)	.11
Sit-to-stand repetitions per minute	36.7 (7.36)	36.13 (7.6)	38.47 (6.52)	.29

[a]*P* values (*P*<.05) were calculated on the basis of *t* test, *P* value of ≤.05 considered statistically significant.

[b]WHtR: waist to height ratio.

[c]HR: heart rate.

[d]SBP: systolic blood pressure.

[e]DBP: diastolic blood pressure.

[f]6MWT: six-minute walk test.

[g]SLS: single-leg stance.

PA and Environmental Perceptions

Table 4 summarizes the psychosocial and environmental characteristics based on WD adherence. On average, participants reported engaging in ≥60 minutes of MVPA at 2.86 (SD 2.17) days per week. Only 13% (8/63) participants met the daily MVPA guidelines, with males being more likely than females to do so (7/28, 25% vs 2/35, 5.7%; *P*=.06). Notably, 24% (15/63) of the participants reported no MVPA during the week, with inactivity being more prevalent among female adolescents (13/35, 36% vs 3/28, 12%). Participation in team sports was positively associated with MVPA; among the 33 adolescents who participated in sports, only 3 reported no MVPA. However, WD-adherent participants reported fewer days of MVPA (2 d vs 4 d; *P*=.004), fewer days of muscle-strengthening activities (3 vs 4.8 d; *P*=.004), and lower team sports participation (21/46, 46% vs 12/17, 71%; *P*=.004) compared to their nonadherent peers. No significant differences were found in psychosocial measures such as self-efficacy, family or friend support, or decisional balance. Interestingly, adherent participants perceived their neighborhoods as safer, with lower scores for crime (1.7

XSL•FO
RenderX

vs 2.4; *P*= .02), traffic concerns (2.1 vs 2.4; *P*=.04), and reported greater access to land-use features (3.1 vs 2.9; *P*=.04).

XSL•FO
RenderX

Table . Physical activity–related psychosocial characteristics and perceptions of neighborhood crime and walkability by wearable adherence status, using wearable device and not using wearable device.

Characteristics	Total (N=63)	WD[a] (n=46)	N-WD[b] (n=17)	*P*[c] value
PA[d] steps per day				
PA wearables, average steps/day, mean (SD)	—[e]	7875 (2696.23)	—	—
PA<10,000 steps/day, n (%)	—	38 (83)	—	—
PA <7000 steps/day, n (%)	—	19 (41)	—	—
PA <5000 steps/day, n (%)	—	7 (15)	—	—
PA self-report measures				
PA 60 min-MVPA[f] days<*7* days, n (%)[g]	54 (86)	42 (91)	12 (71)	.04
PA 60 min-MVPA days<*3* days, n (%)[g]	29 (46)	26 (57)	3 (18)	.006
PA 60 min-MVPA days 0 days, n (%)	15 (24)	13 (28)	2 (12)	.17
PA 60 min-MVPA days, mean (SD)	2.86 (2.17)	2.39 (2.07)	4.12 (1.97)	.004
PA muscle strengthening days, mean (SD)	3.49 (2.29)	3 (2.02)	4.82 (2.48)	.004
PA team sports participation, n (%)[g]	33 (52)	21 (46)	12 (71)	.08
PA team sports (number of teams), mean (SD)	1.94 (1.11)	1.76 (1.02)	2.41 (1.22)	.02
PA stage of change, mean (SD)	2.89 (1.01)	2.78 (.97)	3.23 (1.10)	.06
PA con, mean (SD)	1.81 (.75)	1.82 (.77)	1.76 (.71)	.40
PA pro, mean (SD)	3.64 (1.01)	3.77 (1)	3.29 (.97)	.06
PA self-efficacy, mean (SD)	3.12 (1.06)	3.04 (1.12)	3.37 (.82)	.15
PA-family support, mean (SD)	2.68 (1.34)	2.57 (1.12)	3 (1.42)	.12
PA-friend support, mean (SD)	2.95 (1.05)	2.90 (0.91)	3.09 (0.86)	.48
Neighborhood Walkability Scale, mean (SD)				
Land-use mix–diversity[h]	3.77 (0.83)	3.80 (0.79)	3.69 (0.96)	.65
Neighborhood recreation facilities[h]	3.61 (0.95)	3.59 (0.94)	3.66 (1)	.80
Residential density[h]	3.14 (1.10)	3.18 (1.10)	3.07 (1.14)	.74
Land-use mix–access[h]	3.04 (0.41)	3.10 (0.39)	2.86 (0.40)	.04
Street connectivity[h]	2.95 (0.61)	3 (0.58)	2.81 (0.68)	.30
Walking or cycling facilities[h]	2.99 (0.58)	2.91 (0.58)	3.20 (0.51)	.08
Neighborhood esthetics[h]	2.68 (0.58)	2.73 (0.60)	2.52 (0.51)	.22
Pedestrian and automobile traffic safety[i]	2.23 (0.41)	2.17 (0.43)	2.41 (0.31)	.04

XSL•FO
RenderX

Characteristics	Total (N=63)	WD[a] (n=46)	N-WD[b] (n=17)	*P*[c] value
Crime safety[i]	1.79 (0.68)	1.68 (0.61)	2.13 (0.78)	.02

[a]WD: wearable device.

[b]N-WD: not-using wearable device.

[c]*P* value of ≤.05 considered statistically significant.

[d]PA: physical activity.

[e]Not available.

[f]Activities require moderate physical effort and cause small increases in breathing or heart rate to vigorous physical activities that require hard physical effort and cause large increases in breathing or heart rate.

[g]Pre-not meeting Center for Disease Control and Prevention recommendations.

[h]Higher score=higher walkability.

[i]Higher score=lower walkability.

Predictors of Wearable Adherence

The results of the binomial logistic regression analysis in Table 5 indicate that the model was significant (χ^2_6=16.23; *P*=.01), explaining 61% of the variance (Nagelkerke R^2). Overall, 2 significant predictors emerged: perceived neighborhood crime safety and MVPA frequency. Adolescents engaging in <3 days per week of MVPA had significantly higher odds of WD adherence (OR 6.1, 95% CI 1.5-24) than those who met or exceeded the 3-day threshold. These findings suggest that lower levels of PA and perceptions of neighborhood safety may influence engagement with WD.

Table . Logistic regression predicting the likelihood of wearable device adherence.

Variables	β	SE	Wald (*df*)	Sig.	Exp (β)	95% CI for Exp (β)
Step 1						
Sex	−2.685	2.238	1.439 (1)	0.230	0.068	0.001-5.486
Ethnicity	1.478	1.498	0.974 (1)	0.324	4.386	0.233-82.609
Obesity status	3.036	2.178	1.943 (1)	0.163	20.813	0.291-1486.464
HR[a] (rest)	.057	0.069	0.685 (1)	0.408	1.059	0.925-1.211
PA[b] self-report	−1.234	0.620	3.957 (1)	0.047	0.291	0.086-0.982
Crime safety	−2.442	1.094	4.986 (1)	0.026	0.087	0.010-0.742
Constant	3.966	6.569	0.365	0.546	52.798	—[c]

[a]HR: heart rate.

[b]PA: physical activity.

[c]Not applicable.

Despite the early engagement, adherence to Fitbit monitoring declined sharply over time. Of the 63 students who initiated device use, 42 (67%) maintained adherence for at least 2 months. By month 3, retention dropped to 29 students (46%), and by month 4, only 21 students (33%) remained. Long-term adherence was minimal: 9 students (14%) persisted through month 5 and only 3 (5%) continued through month 7. This trajectory reflects a steep early decline followed by a gradual taper, underscoring the challenge of sustaining engagement beyond the initial months.

Discussion

Principal Findings

This study offers novel insights into WD adherence among underserved adolescents in a school-based setting, thus addressing a critical gap in the literature. While prior research has emphasized the role of personal and environmental factors in shaping PA behaviors, primarily focused on homogenous or higher-resource populations [27], few studies have explored how these influences interact with technology engagement in diverse, resource-limited populations. Notably, our findings revealed that adolescents with lower baseline PA, defined as engaging in MVPA fewer than 3 days per week and those least involved in team sports were more likely to adhere to WD use, precisely the group most in need of sustained support. Furthermore, 90% (9/10) of participants with obesity adhered to WD use, underscoring the potential of wearables to engage adolescents at elevated risk for chronic health conditions. Although counterintuitive, this pattern may reflect an opportunity to drive behavior change, wherein previously inactive youth leveraged technology as a catalyst to initiate and sustain PA. These findings align with prior research, such as the RAW-PA study [25], which demonstrated the high acceptability of wearable interventions among adolescents attending socioeconomically disadvantaged schools. To further

XSL•FO
RenderX

explore the impact of WD adherence on PA behavior, we conducted a post hoc analysis of step-count trends over the academic year. WD-adherent participants with lower baseline PA demonstrated modest but consistent increases in average daily steps, suggesting a potential behavioral change. Although further research is recommended, these results support the potential for wearables to facilitate engagement and promote sustained improvements in PA levels among adolescents, who are most in need of interventions.

Furthermore, adherence to WD use is often undefined or underreported in adolescent cohorts [59], particularly in resource-limited community-based settings. Our study addressed this gap by defining adherence and examining longitudinal patterns in a population disproportionately affected by obesity and related chronic conditions, thereby distinguishing it from short-term pilot studies. WD adherence was defined as having WD HR data for more than 10 hours per day for at least 3 days of the week for a minimum of 3 weeks, which aligns with benchmarks established in the wearable adherence literature [55,56] and indicates meaningful engagement. Furthermore, the data collection spanned the academic year, providing a flexible timeline for teachers and students to initiate WD use as needed. Although the initial WD engagement was high (46/63, 73%) following deployment, adherence declined sharply after the first month, with only 14% persisting until month 5 and 5% through month 7. Our findings in this underserved population echo more recent systematic reviews reporting mixed results for long-term adherence and call for interventions that combine technology with behavioral support and environmental modifications [60,61]. These results underscore the challenge and opportunity of maintaining engagement beyond the initial novelty period and highlight the need for tailored strategies, such as non–peer-competitive gamification and personalized goal setting.

Our study also revealed a stronger than anticipated association between perceived neighborhood safety and WD adherence. Our findings support the role of baseline PA behavior and environmental perceptions in PA adherence in previous studies, while extending this knowledge to WD use among predominantly Black or African American and Hispanic high school adolescents. Moreover, schools can leverage wearable data to personalize interventions, monitor progress, and foster peer support, thereby enhancing engagement and safety. Integrating wearables into school-based programs that address environmental constraints may offer a promising strategy to promote equitable PA engagement and reduce chronic disease risk [62,63]. Collaborations between schools, policymakers, device developers, and health care professionals are essential to ensure equitable access to supportive technologies and environments [22,64,65]. Beyond behavioral approaches, environmental modifications, such as improving walkability, access to sidewalks, and expanding green spaces, can support sustained PA.

Future school-based studies should explore adaptive strategies with WD within PE and wellness classes. For example, dynamic goal adjustment and real-time feedback help maintain motivation throughout the academic year, ensuring consistent exposure and accountability [66]. Wearables may also complement these efforts by mitigating environmental barriers, enabling self-monitoring, goal setting, and engagement in safe, controlled environments such as homes or schools. To enhance long-term engagement, future interventions should prioritize personalized goal setting over gamified peer competition. Evidence suggests that normative comparisons, such as an arbitrary selection of 10,000 steps/day or competition with peers, may yield only short-term motivation and can negatively impact autonomy [67]. Supporting adolescents in setting individualized goals and encouraging self-referenced progress may foster more sustainable PA habits and better align with their personal values and contexts [67,68]. For adolescents not engaged in team sports, additional motivational strategies, including mentorship and small-group activities, may foster social connections and reduce perceived isolation [69].

Although inconclusive, our study also found higher WD adherence rates among female participants than among their male peers, further highlighting the potential influence of safety-related factors on engagement with wearable technologies and underscoring the need for future research to explore sex-specific motivators and barriers to adherence. Furthermore, our study provided insights into the unique needs of adolescents who are first- or second-generation immigrants navigating acculturation, a group that has been shown to experience sharp declines in PA over time in the United States [70]. Despite their medically underserved status, race, ethnicity, and foreign birth did not significantly influence WD adherence in this study. These findings suggest that WD technology can be a culturally neutral and accessible tool for promoting PA across diverse populations. Further research should disaggregate data by generational status and language proficiency to better support first-generation and non–English-speaking youths [71].

Limitations

This study offers valuable insights into WD adherence among underserved adolescents; however, it had several limitations that should be addressed. The unexpectedly strong influence of perceived neighborhood safety may reflect regional or demographic differences, including potential sampling bias, and suggests that environmental safety perceptions play a more prominent role in technology engagement than previously recognized. The reliance on self-report measures introduces potential recall and social desirability biases, and the cross-sectional design limits causal inferences. The relatively small sample size and concentration of participants within a single school year (51/63, 81% in ninth grade) further constrain generalizability. This age clustering may obscure developmental differences in PA behavior and technology engagement across adolescence. Additionally, variability in school schedules, seasonal factors, and academic demands may influence adherence patterns and should be considered in future studies. Future studies should aim for broader age representations and larger multisite samples to enhance external validity and capture developmental variability.

This study highlights the potential of wearable mHealth technologies to engage underserved adolescents in PA, particularly those with lower baseline activity levels and limited access to structured sports. Our findings suggest that WD

XSL•FO
RenderX

adherence is influenced not only by individual behaviors but also by contextual factors, such as perceived neighborhood safety, an often-overlooked determinant of technology engagement among youth. The counterintuitive finding that adolescents with lower baseline PA levels were more likely to adhere to WD use warrants further investigation. Future research should explore mHealth access, usability barriers, and motivational factors for sustained WD use to improve PA. Anecdotal reports from teachers indicated that some students faced challenges with WD use, including limited Wi-Fi connectivity at home or school, forgetfulness, and declining interest over time. These barriers merit further exploration to inform the design of more accessible and engaging mHealth interventions. Despite being at a critical developmental stage for establishing lifelong health behaviors, older adolescents (aged 14 y - 19 y) remain underrepresented in mHealth research. Multischool collaborations and standardized adherence metrics are essential for scaling wearable-based interventions and strengthening cross-study comparisons.

Conclusions

This study highlights the potential of wearable mHealth technologies to engage underserved adolescents in PA, particularly those with lower baseline activity levels and limited access to structured sports. Adherence to WD use was significantly associated with contextual factors, including perceived neighborhood walkability, team sports participation, and prior PA behavior. These findings underscore the importance of designing school-based WD deployment that prioritizes autonomy-supportive strategies, such as personalized goal-setting and self-referenced progress to initiate and sustain PA opportunities. Moving beyond normative comparisons and peer-based competition may enhance sustained engagement and promote equitable health outcomes among high-risk youth.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Conflicts of Interest

None declared.

References

1. Physical activity. World Health Organization. URL: https://www.who.int/initiatives/behealthy/physical-activity [accessed 2026-01-14]
2. Physical activity guidelines for school-aged children and adolescents. US Centers for Disease Control and Prevention. URL: https://www.cdc.gov/physical-activity-education/guidelines/index.html [accessed 2026-01-14]
3. Reimers AK, Knapp G, Reimers CD. Effects of exercise on the resting heart rate: a systematic review and meta-analysis of interventional studies. J Clin Med 2018 Dec 1;7(12):503. [doi: 10.3390/jcm7120503] [Medline: 30513777]
4. Lindgren M, Robertson J, Adiels M, et al. Resting heart rate in late adolescence and long term risk of cardiovascular disease in Swedish men. Int J Cardiol 2018 May 15;259:109-115. [doi: 10.1016/j.ijcard.2018.01.110] [Medline: 29579585]
5. Gu X, Bao S, Moss S, Zhang T. The associations of physical activity and motor performance with executive functioning among young minority children. Med Sci Sports Exerc 2023;55(9S):507-508. [doi: 10.1249/01.mss.0000984536.86093.be]
6. Robinson LE, Stodden DF, Barnett LM, et al. Motor competence and its effect on positive developmental trajectories of health. Sports Med 2015 Sep;45(9):1273-1284. [doi: 10.1007/s40279-015-0351-6] [Medline: 26201678]
7. The 2024 United States report card on physical activity for children and youth. : Physical Activity Alliance URL: https://paamovewithus.org/wp-content/uploads/2024/10/2024-US-Report-Card-Executive-Summary_FINAL.pdf [accessed 2026-01-27]
8. Tandon PS, Kroshus E, Olsen K, Garrett K, Qu P, McCleery J. Socioeconomic inequities in youth participation in physical activity and sports. Int J Environ Res Public Health 2021 Jun 29;18(13):6946. [doi: 10.3390/ijerph18136946] [Medline: 34209544]
9. Klos L, Eberhardt T, Nigg C, Niessner C, Wäsche H, Woll A. Perceived physical environment and active transport in adolescents: a systematic review. J Transp Health 2023 Nov;33:101689. [doi: 10.1016/j.jth.2023.101689]
10. Project Play survey: family spending on youth sports rises 46% over five years. : Project Play; 2025 URL: https://projectplay.org/news/2025/2/24/project-play-survey-family-spending-on-youth-sports-rises-46-over-five-years [accessed 2026-01-14]
11. About one-in-five Americans use a smart watch or fitness tracker. : Pew Research Center; 2020 URL: https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five [accessed 2026-01-14]
12. McDonough DJ, Su X, Gao Z. Health wearable devices for weight and BMI reduction in individuals with overweight/obesity and chronic comorbidities: systematic review and network meta-analysis. Br J Sports Med 2021 Aug;55(16):917-925. [doi: 10.1136/bjsports-2020-103594] [Medline: 33731385]
13. Tudor-Locke C, Craig CL, Beets MW, et al. How many steps/day are enough? for children and adolescents. Int J Behav Nutr Phys Act 2011 Jul 28;8(1):78. [doi: 10.1186/1479-5868-8-78] [Medline: 21798014]

XSL•FO
RenderX

14. Mayorga-Vega D, Casado-Robles C, Viciana J, López-Fernández I. Daily step-based recommendations related to moderate-to-vigorous physical activity and sedentary behavior in adolescents. J Sports Sci Med 2019 Dec;18(4):586-595. [Medline: 31827342]
15. Paluch AE, Gabriel KP, Fulton JE, et al. Steps per day and all-cause mortality in middle-aged adults in the coronary artery risk development in young adults study. JAMA Netw Open 2021 Sep 1;4(9):e2124516. [doi: 10.1001/jamanetworkopen.2021.24516] [Medline: 34477847]
16. Creaser AV, Clemes SA, Costa S, et al. The acceptability, feasibility, and effectiveness of wearable activity trackers for increasing physical activity in children and adolescents: a systematic review. Int J Environ Res Public Health 2021 Jun 8;18(12):6211. [doi: 10.3390/ijerph18126211] [Medline: 34201248]
17. Kouvari M, Karipidou M, Tsiampalis T, et al. Digital health interventions for weight management in children and adolescents: systematic review and meta-analysis. J Med Internet Res 2022 Feb 14;24(2):e30675. [doi: 10.2196/30675] [Medline: 35156934]
18. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Mar 28;14(1):42. [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
19. Terra LF, Rezende LD, Ferreira Silva RM, et al. Interventions on barriers to the participation of adolescents in physical activity: a systematic review. Int J Environ Res Public Health 2025 May 31;22(6):881. [doi: 10.3390/ijerph22060881] [Medline: 40566312]
20. Taverno Ross SE, Sebastião E, Hall G, Bantham A. Overcoming barriers to physical activity in underserved populations. Prog Cardiovasc Dis 2021;64:64-71. [doi: 10.1016/j.pcad.2020.11.002] [Medline: 33159937]
21. Bailey ZD, Krieger N, Agénor M, Graves J, Linos N, Bassett MT. Structural racism and health inequities in the USA: evidence and interventions. Lancet 2017 Apr 8;389(10077):1453-1463. [doi: 10.1016/S0140-6736(17)30569-X] [Medline: 28402827]
22. Nunez-Gaunaurd A, Moore JG, Roach KE, Miller TL, Kirk-Sanchez NJ. Motor proficiency, strength, endurance, and physical activity among middle school children who are healthy, overweight, and obese. Pediatr Phys Ther 2013;25(2):130-138. [doi: 10.1097/PEP.0b013e318287caa3] [Medline: 23542187]
23. Pineros-Leano M, Grafft N, Aguayo L. Childhood obesity risk factors by race and ethnicity. Obesity (Silver Spring) 2022 Aug;30(8):1670-1680. [doi: 10.1002/oby.23500] [Medline: 35894074]
24. Silva ECM, Helme ZE, Silva DRP, et al. Creating active schools: what influences continuous implementation following adoption? J Phys Act Health 2026 Jan 1;23(1):51-62. [doi: 10.1123/jpah.2025-0075] [Medline: 40983312]
25. Koorts H, Salmon J, Timperio A, et al. Translatability of a wearable technology intervention to increase adolescent physical activity: mixed methods implementation evaluation. J Med Internet Res 2020 Aug 7;22(8):e13573. [doi: 10.2196/13573] [Medline: 32763872]
26. Yen HY, Liao Y, Huang HY. Smart wearable device users' behavior is essential for physical activity improvement. Int J Behav Med 2022 Jun;29(3):278-285. [doi: 10.1007/s12529-021-10013-1] [Medline: 34363130]
27. Doherty C, Baldwin M, Keogh A, Caulfield B, Argent R. Keeping pace with wearables: a living umbrella review of systematic reviews evaluating the accuracy of consumer wearable technologies in health measurement. Sports Med 2024 Nov;54(11):2907-2926. [doi: 10.1007/s40279-024-02077-2] [Medline: 39080098]
28. Kim EH, Jenness JL, Miller AB, et al. Association of demographic and socioeconomic indicators with the use of wearable devices among children. JAMA Netw Open 2023 Mar 1;6(3):e235681. [doi: 10.1001/jamanetworkopen.2023.5681] [Medline: 36995714]
29. Vella SA, Aidman E, Teychenne M, et al. Optimising the effects of physical activity on mental health and wellbeing: a joint consensus statement from Sports Medicine Australia and the Australian Psychological Society. J Sci Med Sport 2023 Feb;26(2):132-139. [doi: 10.1016/j.jsams.2023.01.001] [Medline: 36737260]
30. Chandrasekaran R, Katthula V, Moustakas E. Patterns of use and key predictors for the use of wearable health care devices by US adults: insights from a national survey. J Med Internet Res 2020 Oct 16;22(10):e22443. [doi: 10.2196/22443] [Medline: 33064083]
31. Ferguson T, Olds T, Curtis R, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022 Aug;4(8):e615-e626. [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
32. Bopp T, Vadeboncoeur J. "It makes me want to take more steps": racially and economically marginalized youth experiences with and perceptions of Fitbit Zips® in a sport-based youth development program. Sport and Health, Sport and Livelihoods 2021 Oct 22;9(2) [FREE Full text]
33. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
34. International Classification of Functioning, Disability and Health (ICF). World Health Organization. URL: https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health [accessed 2026-01-23]

XSL•FO
RenderX

35. Burton AM, Cowburn I, Thompson F, Eisenmann JC, Nicholson B, Till K. Associations between motor competence and physical activity, physical fitness and psychosocial characteristics in adolescents: a systematic review and meta-analysis. Sports Med 2023 Nov;53(11):2191-2256. [doi: 10.1007/s40279-023-01886-1] [Medline: 37542607]
36. HOPE-physical education core. CPALMS. URL: https://www.cpalms.org/PreviewCourse/Preview/4051 [accessed 2026-01-14]
37. Sallis JF, Grossman RM, Pinski RB, Patterson TL, Nader PR. The development of scales to measure social support for diet and exercise behaviors. Prev Med 1987 Nov;16(6):825-836. [doi: 10.1016/0091-7435(87)90022-3] [Medline: 3432232]
38. Prochaska JJ, Sallis JF, Long B. A physical activity screening measure for use with adolescents in primary care. Arch Pediatr Adolesc Med 2001 May;155(5):554-559. [doi: 10.1001/archpedi.155.5.554] [Medline: 11343497]
39. Sushames A, Edwards A, Thompson F, McDermott R, Gebel K. Validity and reliability of fitbit flex for step count, moderate to vigorous physical activity and activity energy expenditure. PLoS ONE 2016;11(9):e0161224. [doi: 10.1371/journal.pone.0161224] [Medline: 27589592]
40. Jones D, Crossley K, Dascombe B, Hart HF, Kemp J. Validity and reliability of the fitbit flex. Int J Sports Phys Ther 2018 Aug;13(5):860-870. [doi: 10.26603/ijspt20180860] [Medline: 30276018]
41. Chan A, Chan D, Lee H, Ng CC, Yeo AHL. Reporting adherence, validity and physical activity measures of wearable activity trackers in medical research: A systematic review. Int J Med Inform 2022 Apr;160:104696. [doi: 10.1016/j.ijmedinf.2022.104696] [Medline: 35121356]
42. Luszczynska A, Piko B, Januszewicz A. Self-efficacy and adolescents' health. In: Levesque RJR, editor. Encyclopedia of Adolescence: Springer International Publishing; 2018:3386-3395. [doi: 10.1007/978-3-319-33228-4_190]
43. Norman GJ, Sallis JF, Gaskins R. Comparability and reliability of paper- and computer-based measures of psychosocial constructs for adolescent physical activity and sedentary behaviors. Res Q Exerc Sport 2005 Sep;76(3):315-323. [doi: 10.1080/02701367.2005.10599302] [Medline: 16270708]
44. Carlson JA, Sallis JF, Wagner N, et al. Brief physical activity-related psychosocial measures: reliability and construct validity. J Phys Act Health 2012 Nov;9(8):1178-1186. [doi: 10.1123/jpah.9.8.1178] [Medline: 22207589]
45. Youth Risk Behavior Surveillance System (YRBSS). US Centers for Disease Control and Prevention (CDC). URL: http://www.cdc.gov/HealthyYouth/yrbs/index.htm [accessed 2026-01-14]
46. Rosenberg D, Ding D, Sallis JF, et al. Neighborhood Environment Walkability Scale for Youth (NEWS-Y): reliability and relationship with physical activity. Prev Med 2009;49(2-3):213-218. [doi: 10.1016/j.ypmed.2009.07.011] [Medline: 19632263]
47. NHANES Anthropometry Procedures Manual. : US Centers for Disease Control and Prevention; 2021 URL: https://wwwn.cdc.gov/nchs/data/nhanes/public/2021/manuals/2021-Anthropometry-Procedures-Manual-508.pdf [accessed 2026-01-27]
48. Gibson S, Ashwell M. A simple cut-off for waist-to-height ratio (0·5) can act as an indicator for cardiometabolic risk: recent data from adults in the health survey for England. Br J Nutr 2020 Mar 28;123(6):681-690. [doi: 10.1017/S0007114519003301] [Medline: 31840619]
49. Vos T, Lim SS, Abbafati C, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet 2020 Oct;396(10258):1204-1222. [doi: 10.1016/S0140-6736(20)30925-9]
50. Geiger R, Strasak A, Treml B, et al. Six-minute walk test in children and adolescents. J Pediatr 2007 Apr;150(4):395-399. [doi: 10.1016/j.jpeds.2006.12.052] [Medline: 17382117]
51. Li AM, Yin J, Yu CCW, et al. The six-minute walk test in healthy children: reliability and validity. Eur Respir J 2005 Jun;25(6):1057-1060. [doi: 10.1183/09031936.05.00134904] [Medline: 15929962]
52. Springer BA, Marin R, Cyhan T, Roberts H, Gill NW. Normative values for the unipedal stance test with eyes open and closed. J Geriatr Phys Ther 2007;30(1):8-15. [doi: 10.1519/00139143-200704000-00003] [Medline: 19839175]
53. The Cooper Institute. Fitnessgram and Activitygram Test Administration Manual, 4th edition: Human Kinetics; 2010.
54. Pixel and Fitbit keep you moving. Google Store. URL: https://www.fitbit.com/global/us/produ [accessed 2025-04-09]
55. Claudel SE, Tamura K, Troendle J, et al. Comparing methods to identify wear-time intervals for physical activity with the fitbit charge 2. J Aging Phys Act 2021 Jun 1;29(3):529-535. [doi: 10.1123/japa.2020-0059] [Medline: 33326935]
56. Paolillo EW, Lee SY, VandeBunte A, et al. Wearable use in an observational study among older adults: adherence, feasibility, and effects of clinicodemographic factors. Front Digit Health 2022;4:884208. [doi: 10.3389/fdgth.2022.884208] [Medline: 35754462]
57. Zaino CA, Marchese VG, Westcott SL. Timed up and down stairs test: preliminary reliability and validity of a new measure of functional mobility. Pediatr Phys Ther 2004;16(2):90-98. [doi: 10.1097/01.PEP.0000127564.08922.6A] [Medline: 17057533]
58. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice, 3rd edition: Pearson Prentice Hall; 2009.
59. Or CKL, Karsh BT. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009;16(4):550-560. [doi: 10.1197/jamia.M2888] [Medline: 19390112]
60. Zhang W, Xiong K, Zhu C, Evans R, Zhou L, Podrini C. Promoting child and adolescent health through wearable technology: a systematic review. Digit Health 2024;10. [doi: 10.1177/20552076241260507] [Medline: 38868368]

XSL•FO
RenderX

61. Chen X, Wang F, Zhang H, Lin Y, Zhu S, Yang Y. Effectiveness of wearable activity trackers on physical activity among adolescents in school-based settings: a systematic review and meta-analysis. BMC Public Health 2025 Mar 18;25(1):1050. [doi: 10.1186/s12889-025-22170-z] [Medline: 40102761]
62. Plotnikoff RC, Gebel K, Lubans DR. Self-efficacy, physical activity, and sedentary behavior in adolescent girls: testing mediating effects of the perceived school and home environment. J Phys Act Health 2014 Nov;11(8):1579-1586. [doi: 10.1123/jpah.2012-0414] [Medline: 24733181]
63. van Sluijs EMF, Ekelund U, Crochemore-Silva I, et al. Physical activity behaviours in adolescence: current evidence and opportunities for intervention. Lancet 2021 Jul 31;398(10298):429-442. [doi: 10.1016/S0140-6736(21)01259-9] [Medline: 34302767]
64. Pope M. Preventing weight gain in children who are school age and African-American. Pediatr Phys Ther 2016;28(2):207-216. [doi: 10.1097/PEP.0000000000000243] [Medline: 26914717]
65. Bezner JR. Promoting health and wellness: implications for physical therapist practice. Phys Ther 2015 Oct;95(10):1433-1444. [doi: 10.2522/ptj.20140271] [Medline: 25908523]
66. Creaser AV, Frazer MT, Costa S, Bingham DD, Clemes SA. The use of wearable activity trackers in schools to promote child and adolescent physical activity: a descriptive content analysis of school staff's perspectives. Int J Environ Res Public Health 2022 Oct 28;19(21):21. [doi: 10.3390/ijerph192114067] [Medline: 36360944]
67. Kerner C, Goodyear VA. The motivational impact of wearable healthy lifestyle technologies: a self-determination perspective on fitbits with adolescents. Am J Health Educ 2017 Sep 3;48(5):287-297. [doi: 10.1080/19325037.2017.1343161]
68. Danković G, Stantić T, Herodek R, et al. Effects of commercially available wearable devices on physical activity promotion and health in children and adolescents: systematic review. Appl Sci (Basel) 2023;13(12):7194. [doi: 10.3390/app13127194]
69. Hartman SJ, Chen R, Tam RM, Narayan HK, Natarajan L, Liu L. Fitbit use and activity levels from intervention to 2 years after: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2022 Jun 30;10(6):e37086. [doi: 10.2196/37086] [Medline: 35771607]
70. Allen ML, Elliott MN, Morales LS, Diamant AL, Hambarsoomian K, Schuster MA. Adolescent participation in preventive health behaviors, physical activity, and nutrition: differences across immigrant generations for Asians and Latinos compared with Whites. Am J Public Health 2007 Feb;97(2):337-343. [doi: 10.2105/AJPH.2005.076810] [Medline: 17138919]
71. Taverno Ross SE, Larson N, Graham DJ, Neumark-Sztainer D. Longitudinal changes in physical activity and sedentary behavior from adolescence to adulthood: comparing U.S.-born and foreign-born populations. J Phys Act Health 2014 Mar;11(3):519-527. [doi: 10.1123/jpah.2011-0359] [Medline: 23416986]

Abbreviations

HOPE: Health Opportunities through Physical Education
HR: heart rate
ICF: International Classification of Functioning and Disability
mHealth: mobile health
MVPA: moderate-to-vigorous physical activity
PA: physical activity
PE: physical education
SCT: Social Cognitive Theory
WD: wearable device
WtHR: waist-to-height ratio

Edited by G Badicu; submitted 10.Jul.2025; peer-reviewed by M Fraser, P Gorp; revised version received 30.Oct.2025; accepted 16.Nov.2025; published 29.Jan.2026.

Please cite as:
Nunez-Gaunaurd A, Raya M
Assessing Wearable mHealth Adherence in Underserved Adolescents and its Associations With Physical Activity, Sports, and Safety Perceptions: Prospective Cohort Study
JMIR Mhealth Uhealth 2026;14:e80465
URL: https://mhealth.jmir.org/2026/1/e80465
doi:10.2196/80465

© Annabel Nunez-Gaunaurd, Michele Raya. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 29.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,

XSL•FO
RenderX

provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Efficacy of a Digital Peer Support Program on Weight Management and Mental Health in University Students With Preobesity: Randomized Controlled Trial

Xingyu Liu[1], PhD; Ting Liu[2], PhD; Tao Chen[1*], PhD; Ruisi Ma[3,4*], PhD

[1]Badminton Technical and Tactical Analysis and Diagnostic Laboratory, National Academy of Badminton, Guangzhou Sport University, Guangzhou, China

[2]School of Nursing, Sun Yat-sen University, Guangzhou, China

[3]School of Physical Education, Jinan University, No. 601 Huangpu Avenue West, Tianhe District, Guangzhou, Guangdong, China

[4]Key Laboratory of Guangdong Province, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Ruisi Ma, PhD
School of Physical Education, Jinan University, No. 601 Huangpu Avenue West, Tianhe District, Guangzhou, Guangdong, China

Abstract

Background: Approximately one-third of university students are overweight or obese, and a similar proportion experience anxiety or depression. Despite the interrelated nature of weight and mental health, interventions rarely address these issues simultaneously in young adults. Digital peer support interventions have the potential to promote healthy lifestyle and mental well-being. However, evidence is limited on whether a digital peer-driven approach can concurrently improve weight management and mental health in university populations with preobesity.

Objective: This randomized controlled trial (RCT) evaluated the efficacy of a digital peer support program in concurrently improving weight management and mental health outcomes among university students with preobesity.

Methods: In a single-blind parallel group RCT, 216 students with preobesity were allocated equally among three 6-month arms, which were a peer support intervention, an active wellness control, and a waitlist control. The peer support arm began with an interactive online workshop followed by moderated WeChat (Tencent) group discussions, daily micro tasks, biweekly group challenges, and digital badges to reinforce engagement. The active control group received the same schedule and formats but focused on general wellness topics. The waitlist group completed the same assessments without any intervention during the study period. The primary outcome measured the change in BMI from baseline to 6 months. Secondary outcomes included weekly physical activity measured in metabolic equivalent of task minutes, self-esteem, loneliness, anxiety, and depression assessed at 0, 2, 4, and 6 months. Analyses used linear mixed effects models.

Results: Retention exceeded 90%. At 6 months, the peer-support group achieved a greater BMI reduction than the active control by 0.47 (95% CI −0.89 to −0.04) kg/m² and waitlist by 0.54 (95% CI −0.85 to −0.01) kg/m². Weekly metabolic equivalent of task-minutes was 129.5 higher than active control (95% CI 53.3-205.6) and 152.9 higher than waitlist (95% CI 68.4-237.4). Self-esteem increased by 1.81 points versus active control (95% CI 0.22-3.39) and 1.99 points versus waitlist (95% CI 0.21-3.76). Loneliness scores fell by 3.79 points relative to active control (95% CI −7.03 to −0.56) and by 5.02 points relative to waitlist (95% CI −8.38 to −1.66). No significant differences emerged for anxiety or depression.

Conclusions: A comprehensive digital peer-support program delivered via WeChat produced modest but clinically meaningful improvements in weight management, physical activity, self-esteem, and social connectedness among undergraduates with preobesity compared with wellness control and no intervention. These findings suggest that integrating peer support into scalable digital platforms can simultaneously address physical and psychosocial health in at-risk university populations.

Trial Registration: ClinicalTrials.gov NCT06966661; https://clinicaltrials.gov/study/NCT06966661

(JMIR Mhealth Uhealth 2026;14:e78960) doi:10.2196/78960

KEYWORDS

digital peer support; weight management; mental health; preobese; university students; randomized controlled trial

XSL•FO
RenderX

Introduction

Young adults with preobesity are defined by a BMI between 25 and 29.9 kg/m² [1]. This group represents a critical yet underaddressed population in the fight against obesity [2]. University students with preobesity have a high prevalence of excess weight, with studies reporting that roughly 20% - 30% of students fall into the overweight category, with an additional 5% - 10% already meeting obesity criteria [3,4]. Unlike individuals of normal weight, these students are already on the threshold of obesity, experiencing early weight gain–related health risks and often subtle metabolic changes [5]. Crucially, even at this preobese stage, individuals already experience adverse consequences, including heightened stress, increased depressive symptoms, and lower academic performance [6-8]. This underscores that focusing on students with preobesity is not merely a preventative measure but a response to an emerging health and psychosocial burden that distinctly affects this group.

Intervening during the preobese stage in emerging adulthood may offer a pivotal window to alter lifelong health trajectories [9]. The university years are a well-recognized turning point for weight gain and lifestyle habits, commonly known as the "freshman weight gain" phenomenon. Grounded in the socioecological model, first-year transitions expose students to interpersonal, organizational, and environmental cues that favor weight gain [10,11]. The prevention of further weight gain and the promotion of modest weight loss are widely viewed as more cost-effective and achievable than treating established obesity later on [12,13]. Overweight students generally have fewer entrenched behaviors and less severe physiological impairment than individuals with obesity, meaning early intervention can capitalize on the reversibility of risk [14]. From a public health perspective, preventing the progression from overweight to obesity avoids the need for intensive medical treatments and reduces long-term health care costs [12,15,16]. Focusing on college students with preobesity offers a critical window for high-impact preventive measures, potentially delaying cardiometabolic diseases and mental health decline before they become more difficult and expensive to manage [17,18].

Despite the clear rationale for early intervention, university students with preobesity face unique challenges that current interventions rarely address adequately [8,19]. This transitional life stage features new freedoms and psychosocial pressures. Academic stress, changing peer networks, and the shift from adolescence to adulthood can negatively affect eating, activity, and mental well-being [20]. Overweight students may struggle with body image concerns and social stigma while also encountering environments filled with unhealthy food options and sedentary habits [21]. These psychosocial stressors and environmental factors intertwine, often creating a vicious cycle where weight gain and mental health issues reinforce each other [22]. However, most existing health programs operate in isolation. Weight management initiatives in college settings tend to focus solely on diet and exercise [18], whereas campus mental health services rarely integrate lifestyle change for physical health [23]. There is a notable gap in interventions that treat physical and mental health as interconnected targets. In fact, a recent review highlighted that very few programs for university students have successfully improved both health behaviors, such as diet, physical activity, and weight-related outcomes, and mental health outcomes at the same time [8]. This lack of comprehensive approaches means students with preobesity at a crucial phase often do not receive support that addresses the full range of their needs, namely managing weight while concurrently bolstering psychological resilience and well-being. Moreover, no prior study has included both an attention-matched active control and a waitlist control arm to isolate specific peer-support benefits from general engagement effects.

Digital peer-support interventions, encompassing online group chats, moderated peer discussions, peer challenges, and digital badge systems, offer a promising and innovative solution to fill this gap for young adults with preobesity [24]. According to social cognitive theory, observing peers' successes and receiving encouragement enhances self-efficacy and outcome expectancies, thereby improving motivation, adherence, and outcomes in weight management [25]. In line with this theory, social support and positive peer influences operate through observational learning and modeling to normalize healthy behaviors and reduce the isolation or stigma an overweight student might feel, especially among university students who are highly receptive to peer influences [25,26]. By delivering such peer support via digital platforms, we can amplify its reach and convenience. Technology-based interventions are inherently adaptable, engaging students in real time, and digital platforms offer broad appeal and extensive reach across this population [27]. Through social media groups, mobile apps, or online forums, students can connect with "virtual" peers to share progress, setbacks, and encouragement [25]. Prior studies indicate that online peer communities can recreate the sense of camaraderie found in in-person support groups, providing accountability and empathy among participants with similar goals [25]. Moreover, digital peer support can be accessed on-demand and flexibly integrated into students' busy schedules, overcoming common barriers to participation such as time constraints or embarrassment in face-to-face settings [28]. Although this approach has been previously studied, existing trials seldom focus on the critical preobese life stage, compare multiple digital support modalities within the same randomized controlled trial (RCT), or include long-term follow-up and an expanded set of psychosocial outcomes.

In light of the above, this study aimed to evaluate the efficacy of a digital peer-support program on weight management and mental health in university students with preobesity. To test this, we conducted a single-blind, parallel-group, 3-arm RCT comparing a digital peer-support intervention, an attention-matched active control, and a waitlist control, with prespecified pairwise comparisons. We hypothesized that participants in the peer-support arm would demonstrate greater improvements in BMI and mental health measures than those in the active control arm, and that both intervention arms would outperform the waitlist control. By examining this novel intervention, our goal is to address an important gap in student health promotion and contribute evidence on the integration of digital peer support in weight management and mental health prevention efforts. Our intervention was grounded in social

XSL•FO
RenderX

cognitive theory and the socioecological model [29,30]. We anticipated that peer modeling within a supportive digital environment would enhance self-efficacy and motivation for healthy behaviors and provide a multilevel context for behavior change.

Methods

Study Design

We conducted a single-blind, parallel-group, superiority RCT entirely online over a 6-month intervention period to evaluate a brief digital peer-support intervention for weight management and mental health in university students with preobesity. The participants were drawn from 6 universities in Guangzhou, including Sun Yat-sen University, South China Normal University, Jinan University, South China Agricultural University, South China University of Technology, and Guangdong University of Foreign Studies. Participants were recruited between September 1 and October 31, 2024, and follow-up assessments were conducted at 2 months in November 2024, 4 months in January 2025, and at the study end on April 30, 2025. The study is reported in accordance with CONSORT (Consolidated Standards of Reporting Trials) guidelines and was registered prior to enrollment (ClinicalTrials.gov ID NCT06966661). Participants were recruited from university campuses via online advertisements, email bulletins, and social media postings. Interested individuals accessed a screening survey by clicking a link or QR code in the advertisement.

Participants and Recruitment

Eligible participants were currently enrolled university undergraduate students, aged 18 years or older, with preobese weight status (defined as BMI between 25.0 and 29.9 kg/m² based on self-reported height and weight at screening). Additional inclusion criteria were access to a stable internet connection and a device for video conferencing, to enable participation in the online program. Participants were excluded if they had a BMI in the obese range (≥30 kg/m^2) or normal range (<25 kg/m^2) at screening, had any diagnosed medical or psychiatric condition that would interfere with study participation or require alternate care (eg, uncontrolled diabetes, an eating disorder, psychosis, or acute suicidal ideation), or were already enrolled in a weight management program or receiving psychological counseling. To ensure safety, individuals with severe depression or high acute mental health risk were directed to clinical services and not enrolled. Pregnant students were also excluded due to potential weight changes unrelated to the intervention.

Randomization and Blinding

Eligible consenting participants were randomly assigned in a 1:1:1 ratio to 3 groups that were the digital peer support intervention group, the active control group receiving standard online wellness content, and the waitlist control group. Randomization was carried out using a computer-generated sequence. This sequence applied permuted blocks of varying sizes (6 and 9). The allocation list was prepared by an independent statistician who was not involved in participant recruitment or intervention delivery. To maintain allocation concealment, the randomization list was stored in a password-protected file on a secure server accessible only to the statistician. Study coordinators obtained the next assignment code via a secure messaging system only after participants completed baseline assessments and eligibility confirmation, which prevented recruiters from foreseeing upcoming assignments. Because the trial was conducted entirely online, sealed envelopes were not used. Instead, participants were assigned by the statistician through a secure web-based platform. The study used a single-blind design, and participants were informed that the trial compared 3 different online wellness programs but were not told which program was the primary intervention. Both the intervention and active control groups received online content and peer interaction for 6 months to control for expectation and exposure time; the waitlist control group did not receive any intervention activities during this period.

Facilitators and research staff who delivered the intervention could not be blinded due to the nature of the program. However, they were not involved in outcome assessments, and distinct facilitators managed each group without evaluative responsibilities. Follow-up data on BMI, self-esteem, loneliness, and other outcomes were collected via self-administered online surveys distributed by research assistants who were blind to group assignment and used anonymous study codes to track responses. These assistants did not have access to the randomization list. Data analysts remained blinded to group assignments until the primary analyses were completed.

Intervention Group

Participants in this arm began with a 15-minute live interactive workshop delivered synchronously via an online videoconferencing platform. All participants joined remotely from their own devices. A trained health educator used a shared digital whiteboard and chat functions to guide the group through evidence-based guidance on balanced eating, regular movement, stress management, and resilience building using real-time polls and interactive infographics to illustrate key points in a neutral, student-friendly language while participants added their own reflections and questions. Immediately afterward, teams of 4 to 6 students convened in a dedicated WeChat group for a 1-hour conversation combining text and voice interactions under the guidance of a trained moderator, with participants sharing brief reflections on personal challenges, exploring sustainable behavior change strategies, and agreeing on 1 or 2 small health goals for the coming week while the moderator used open-ended prompts to ensure inclusive and supportive dialogue. Over the 6-month follow-up, the WeChat group delivered daily micro tasks 7 days a week, such as minute-long nutrition quizzes, morning breathing exercises, and simple step count challenges. Each completed task earned participants a digital badge displayed in the group, and automated motivational messages reinforced progress and reminded students of their goals. Every 2 weeks, the moderator launched a friendly group challenge via the WeChat group, such as collectively walking 10,000 steps in a day to foster camaraderie and mutual accountability, and the WeChat group remained open for ongoing peer support with study staff overseeing the conversation to keep discussions respectful and on topic.

XSL•FO
RenderX

Active Control Group

Participants followed the same schedule and digital formats, including a 15-minute workshop, a 1-hour WeChat discussion, daily micro tasks, and 2-week challenges, but the content focused on general student wellness topics, such as effective study habits, time management, and healthy sleep. WeChat discussions addressed academic strategies, identification of barriers, and the setting of weekly improvement goals.

Waitlist Control Group

Participants in this arm completed all baseline and follow-up assessments on the same schedule as the other 2 groups but did not receive any workshop activity, WeChat discussions, or micro task prompts during the 6-month period. They were informed that they would gain access to the full peer-support program after the final data collection. Throughout the study, they continued their usual routines without additional contact beyond survey reminders and, at the conclusion of the follow-up, they were offered the opportunity to join the interactive workshop, the WeChat peer discussions, and to receive the series of daily micro tasks.

Procedures

After providing informed consent, each participant completed an online baseline survey capturing demographic information, self-reported height and weight, and all outcome measures. Participants were then randomized in equal proportions to 1 of the 3 arms and received secure instructions for scheduling their workshop and WeChat group times or confirmation of their waitlist status. Sessions were arranged to match participants' availability, and reminders were sent by email and text 1 to 2 days before each workshop or discussion session. Attendance and engagement were logged by facilitators, and all WeChat group activities were monitored by study staff to document participation. Completion of the daily micro-learning tasks and the awarding of digital badges were automatically recorded via backend logs in the WeChat mini-program. Each week, the data manager generated adherence reports to flag participants who completed fewer than half of the assigned tasks or who had not earned a badge in 7 consecutive days. These participants were contacted by study staff to encourage reengagement. No participants were excluded or reassigned on the basis of missed tasks. Adherence metrics were summarized descriptively and considered in exploratory analyses but did not influence eligibility or weighting in the primary available-case analyses. Follow-up assessments were delivered online at 2 months, 4 months, and 6 months after baseline using identical survey links to maintain blinding. Nonresponders received reminders at 5 and 12 days postinvitation and a final contact attempt by phone or messaging app if needed. Participants who completed each survey received a modest gift card honorarium and, after the final follow-up, the waitlist control group was provided access to the peer support program while the study was closed to further interventions.

Outcomes and Measures

The primary outcome was BMI, which was recorded from baseline to the 6-month assessment, with interim follow-up assessments at 2 months and at 4 months. Participants were instructed to weigh themselves in the morning before eating and to measure height against a flat wall, noting any deviations in the procedure.

Secondary outcomes encompassed changes in physical activity, anxiety, depression, mental well-being, perceived stress, general well-being, self-esteem, loneliness, and social connectedness. Physical activity was captured by the International Physical Activity Questionnaire, a 7-item measure recording the frequency and duration of walking, moderate-intensity, and vigorous activities over the previous 7 days [31]. Anxiety symptoms were assessed using the 7-item Generalized Anxiety Disorder (GAD) scale, which rates symptoms over the past 2 weeks from 0 (not at all) to 3 (nearly every day) for a total score of 0-21 [32]; the Chinese version demonstrates strong reliability, validity, and measurement invariance in university samples [33]. Depressive symptoms were measured by the 9-item Patient Health Questionnaire (PHQ), scored 0 - 3 per item for a total of 0-27 [34]; Chinese validation studies report high internal consistency and test-retest reliability [35]. Mental well-being was captured by the 14-item Warwick-Edinburgh Mental Well-being Scale (WEMWBS), with each item scored 1 - 5 for a total range of 14-70 [36]; the Chinese WEMWBS shows robust internal consistency and construct validity among nursing trainees and general students [37]. Perceived stress was assessed via the 10-item Perceived Stress Scale (PSS), rated 0 (never) to 4 (very often), yielding scores from 0 to 40 [38]; the Simplified Chinese version exhibits high reliability (α>.90) and a clear 2-factor structure [39]. General well-being was measured by the 5-item World Health Organization Well-Being Index (WHO-5), scored 0 (at no time) to 5 (all of the time) and converted to a 0 - 100 index [40]; the Chinese WHO-5 has demonstrated good internal consistency (α≈.85) and unidimensional fit in student and patient samples [41]. Self-esteem was measured by the 10-item Rosenberg Self-Esteem Scale on a 4-point agreement scale [42]; Chinese adaptations confirm robust factorial and criterion validity [43]. Loneliness was assessed by the 20-item University of California, Los Angeles (UCLA) Loneliness Scale Version 3, with items rated 1 (never) to 4 (often) [44]; recent validation of Chinese variants supports strong reliability and validity in adolescent and university populations [45]. Social connectedness was measured by the 8-item Social Connectedness Scale (SCS), scored 1 (strongly disagree) to 6 (strongly agree) [46]; the Chinese translation demonstrates satisfactory psychometric properties in student samples [47].

Statistical Analysis

Effect size assumptions (Cohen $d \approx 0.50$) were derived from published digital weight management trials [48]. A priori power analysis using G*Power (Heinrich-Heine-Universität Düsseldorf) for a 3-arm, 4 timepoint repeated measures design [49], with an assumed within-subject correlation of 0.5, 2-tailed α=.05, and 80% power, indicated a minimum sample of 159 participants (n=53 per arm). Allowing for up to 20% attrition, we therefore enrolled 216 participants (n=72 per arm), ensuring robust power to detect the expected effects.

Primary and secondary outcomes were measured at baseline and 2, 4, and 6 months. Analyses were performed on an

XSL•FO
RenderX

available-case basis. Participants who completed at least 1 follow-up assessment were included in the analysis. Missing data were less than 5% and under the missing at random assumption, linear mixed effects models estimated by maximum likelihood were used without additional imputation [50]. Each model specified fixed effects for study arm time and the study arm by time interaction and adjusted for the baseline value of the outcome. A random intercept for each participant accounted for within-subject correlation across repeated measures. The primary test of efficacy was the study arm by time interaction for BMI, and significance for this test was evaluated at 2-tailed P <.05. Secondary outcomes were analyzed in the same way, and false discovery rate–adjusted P values were reported to provide context for the full set of psychosocial measures. Model diagnostics included inspection of residuals for normal distribution and equal variance. All analyses were performed in R (version 4.2.2; R Foundation for Statistical Computing) with the *lme4* and *mice* packages, and results are presented alongside 95 % CIs [51].

Ethical Considerations

The study protocol was reviewed and approved by the Institutional Review Board of Guangzhou Sport University (approval number 2025LCLL-057), ensuring that the research adhered to institutional and national ethical guidelines for research involving human participants. All participants provided informed consent electronically after a one-on-one discussion with a research staff member about the study procedures and eligibility criteria. Participation was voluntary, and they were informed that they could withdraw from the study at any time without consequences. Data were collected via secure online surveys and stored on a password-protected server accessible only to the research team, and personal identifiers were removed before analysis so that only deidentified, aggregated data were used for publication to protect participant confidentiality. Participants did not receive monetary or material compensation, and digital badges and peer support interactions provided within the intervention were part of the program design and served as motivational tools rather than compensation.

Results

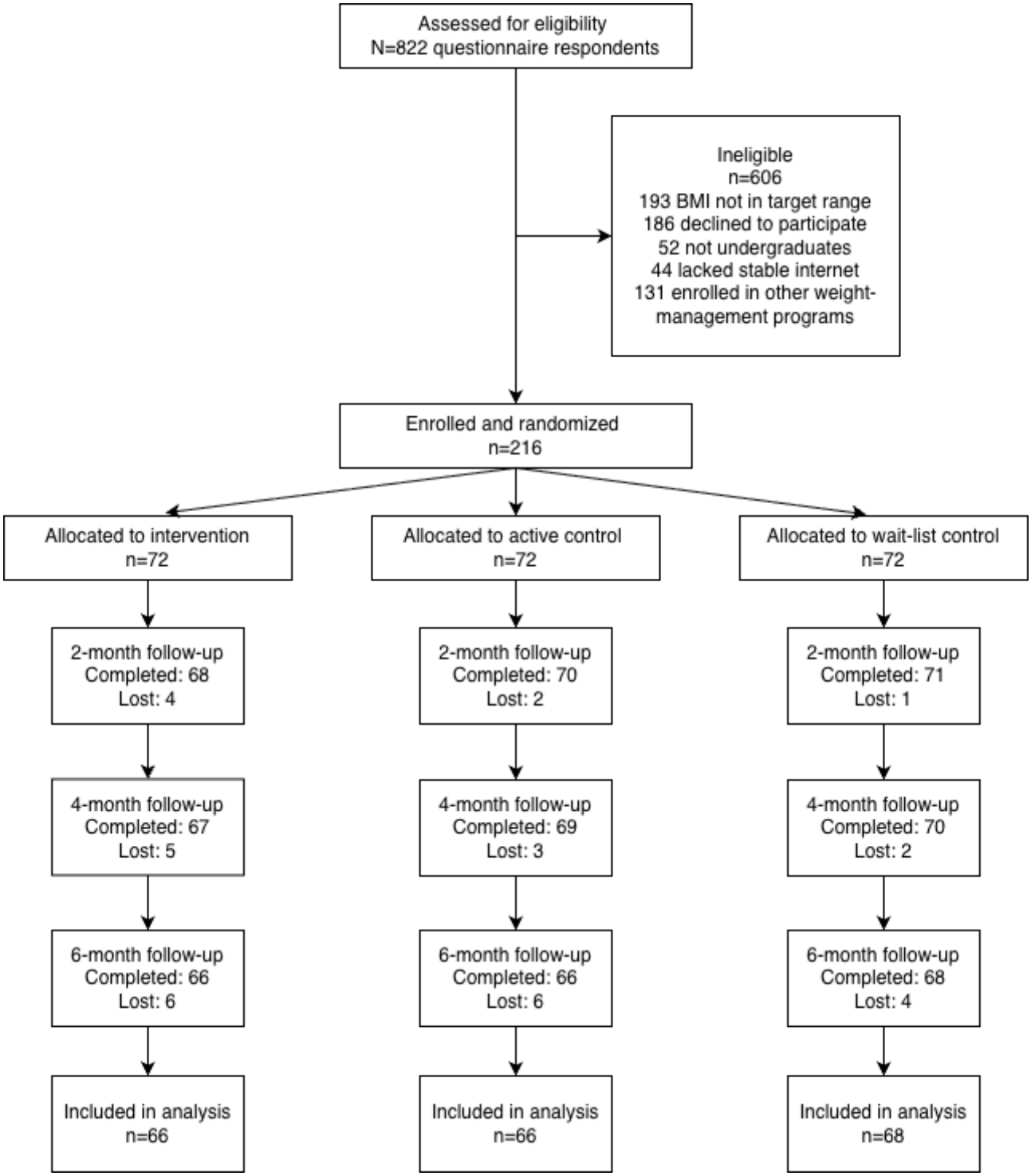
Of the 822 students assessed for eligibility, 216 met inclusion criteria and were randomized equally to the intervention (n=72), active control (n=72), and waitlist control (n=72) arms. Follow-up rates exceeded 90% at each of the 4 assessment points (Figure 1). Baseline demographic and clinical characteristics were well-balanced across groups: the overall mean age was 21.69 (SD 2.40) years, 48.15% (104/216) were female, and the mean BMI was 27.03 (SD 1.29) kg/m^2; no between-group differences reached statistical significance on any measure (Table 1).

Adherence to the program activities was high. Participants in the digital peer-support arm completed an average of 130.94 out of 181 of assigned daily micro tasks (72.34%, SD 12.38%) and attended a mean of 21.68 out of 24 scheduled WeChat group discussions (90.33%). In the active wellness control arm, participants completed a mean of 124.80 out of 181 daily tasks (68.94%, SD 13.11%) and attended 21.30 out of 24 group discussions (88.73%). These adherence estimates were obtained through daily check-ins conducted by research assistants within the WeChat groups, who recorded participant confirmations of task completion and discussion attendance in study logs rather than relying on automated device logs.

At 2 months, there were no significant group differences on any outcome. However, by 4 months (Table 2), the intervention arm had already achieved markedly higher weekly physical activity than both control groups: adjusted total metabolic equivalent of task (MET) - minutes were 124.10 MET greater than in the active control (95% CI 45.64-202.55; P<.001) and 147.32 MET greater than in the wait - list control (95% CI 62.51-232.13; P<.001). Self - esteem scores in the intervention arm exceeded those of the active control by 1.93 points (95% CI 0.35-3.52; P=.04) and those of the wait - list control by 2.08 points (95% CI 0.30-3.87; P=.04). Loneliness scores were also significantly lower among intervention participants, with scores 3.76 points below the active control (95% CI –6.93 to –.58; P=.03) and 5.09 points below the waitlist control (95% CI –8.40 to –1.78; P=.03), reflecting a greater reduction in perceived social isolation (Figure 2).

XSL•FO
RenderX

Figure 1. Flow diagram of participant recruitment, allocation, follow-up, and analysis in the randomized trial.



XSL•FO
RenderX

Table . Baseline demographic and psychosocial characteristics of participants, overall and by group.

Characteristics[a]	Total sample (N=216)	Intervention (n=72)	Active control (n=72)	Waitlist control (n=72)
Age, mean (SD)	21.69 (2.40)	21.83 (2.25)	21.21 (2.35)	22.04 (2.55)
Sex, n (%)				
Male	112 (51.90)	36 (50)	40 (55.56)	36 (50)
Female	104 (48.10)	36 (50)	32 (44.44)	36 (50)
Grade[b], n (%)				
Grade 1	60 (27.78)	20 (27.78)	15 (20.83)	25 (34.72)
Grade 2	44 (20.37)	15 (20.83)	16 (22.22)	13 (18.06)
Grade 3	59 (27.31)	23 (31.94)	19 (26.39)	17 (23.61)
Grade 4	53 (24.54)	14 (19.44)	22 (30.56)	17 (23.61)
WeChat[c] use (hrs/day), mean (SD)	1.96 (1.02)	2.04 (0.93)	2.03 (1.12)	1.82 (1)
BMI, kg/m^2, mean (SD)	27.03 (1.29)	26.93 (1.28)	27.10 (1.34)	27.06 (1.27)
MET[d], mean (SD)	628 (213.17)	646.40 (196.73)	612.44 (212.90)	625.14 (230.17)
GAD[e], mean (SD)	5.31 (3.62)	5.11 (3.69)	5.79 (3.78)	5.03 (3.38)
PHQ[f], mean (SD)	5.90 (4.53)	5.97 (4.53)	5.96 (5.04)	5.78 (4.05)
WEMWBS[g], mean (SD)	50.30 (9.45)	49.72 (8.77)	50.19 (9.75)	50.99 (9.88)
PSS[h], mean (SD)	16.83 (5.79)	16.97 (6.10)	17.25 (5.79)	16.28 (5.53)
WHO-5[i], mean (SD)	61.33 (20.15)	60.67 (19.77)	64.15 (19.63)	59.17 (20.99)
Self-esteem[j], mean (SD)	30.54 (4.84)	30.54 (5.13)	30.57 (4.06)	30.51 (5.30)
Loneliness[k], mean (SD)	40.68 (9.62)	41 (9.71)	40.07 (9.56)	40.97 (9.68)

[a]Values are presented as mean (SD) for continuous variables and n (%) for categorical variables.

[b]Grade denotes year of study.

[c]WeChat use reflects average daily hours on the platform.

[d]MET: metabolic equivalent task minutes per week, which was assessed via the International Physical Activity Questionnaire.

[e]GAD: Generalized Anxiety Disorder-7.

[f]PHQ: Patient Health Questionnaire-9.

[g]WEMWBS: Warwick-Edinburgh Mental Well-being Scale.

[h]PSS: Perceived Stress Scale.

[i]WHO‐5: 5-item World Health Organization Well‐Being Index.

[j]Self‐Esteem: Rosenberg Self‐Esteem Scale.

[k]Loneliness: University of California, Los Angeles Loneliness Scale.

XSL•FO
RenderX

Table . Adjusted linear mixed - effects model estimates with group× time interaction and pairwise contrasts.

Outcomes and timepoint[a]	Intervention	Active control	Waitlist control	Int (LRT)[b], χ^2 (df)	Intervention versus AC[c], mean difference (95% CI)	Intervention versus WLC[d], mean difference (95% CI)
BMI, kg/m²						
Baseline	26.93 (1.28)	27.10 (1.34)	27.06 (1.27)	51.57 (6)	–0.17 (–0.59 to 0.26)	–0.13 (–0.55 to 0.29)
2 months	26.84 (1.27)	27.06 (1.32)	27.02 (1.28)	—[e]	–0.21 (–0.63 to 0.21)	–0.18 (–0.60 to 0.24)
4 months	26.80 (1.25)	27.03 (1.37)	26.99 (1.32)	—	–0.24 (–0.66 to 0.18)	–0.19 (–0.61 to 0.23)
6 months	26.63 (1.28)	27.10 (1.34)	27.06 (1.27)	—	–0.47 (–0.89 to –0.04)	–0.54 (–0.85 to –0.01)
MET[f]						
Baseline	646.40 (196.73)	612.44 (212.90)	625.14 (230.17)	21 (6)	33.96 (−33.11 to 101.03)	21.26 (−48.80 to 91.32)
2 months	690.64 (243.11)	632.54 (228.78)	608.71 (268.26)	—	58.10 (−19.15 to 135.34)	81.93 (−1.84 to 165.70)
4 months	719.26 (247.92)	595.17 (231.28)	598.81 (196.49)	—	124.10 (45.64-202.55)	147.32 (62.51-232.13)
6 months	749.21 (205.77)	600.42 (231.22)	603.60 (196.60)	—	129.45 (53.28-205.61)	152.87 (68.35-237.38)
GAD[g]						
Baseline	5.11 (3.69)	5.79 (3.78)	5.03 (3.38)	13.04 (6)	–0.68 (–1.83 to 0.47)	0.08 (–1.07 to 1.24)
2 months	4.67 (3.63)	5.15 (3.84)	4.38 (3.40)	—	–0.49 (–1.64 to 0.67)	0.29 (–.86 to 1.45)
4 months	4.24 (3.55)	4.75 (3.73)	3.72 (3.09)	—	–0.51 (–1.67 to 0.64)	0.51 (–0.64 to 1.67)
6 months	3.79 (3.67)	4.38 (3.78)	3.06 (3.06)	—	–0.58 (–1.74 to 0.57)	0.74 (–0.42 to 1.89)
PHQ[h]						
Baseline	5.97 (4.53)	5.96 (5.04)	5.78 (4.05)	6.33 (6)	0.01 (–1.31 to 1.33)	–0.08 (–1.40 to 1.25)
2 months	5.49 (4.33)	5.40 (4.80)	5.25 (4.02)	—	–0.49 (–1.81 to 0.83)	–0.42 (–1.74 to 0.90)
4 months	4.78 (4.41)	4.85 (4.81)	4.71 (3.87)	—	–0.51 (–1.83 to 0.81)	–0.74 (–2.06 to 0.57)
6 months	4.12 (4.33)	4.46 (4.62)	3.83 (3.70)	—	–0.51 (–1.83 to 0.81)	–0.65 (–1.97 to 0.67)
WEMWBS[i]						
Baseline	49.72 (8.77)	50.19 (9.75)	50.99 (9.88)	7.21 (6)	–0.47 (–4 to 3.06)	–2.67 (–6.20 to 0.87)
2 months	49.54 (9.19)	50.50 (9.79)	50.76 (10.26)	—	–1.60 (–5.13 to 1.93)	–1.42 (–4.95 to 2.11)

XSL•FO
RenderX

Outcomes and timepoint[a]	Intervention	Active control	Waitlist control	Int (LRT)[b], χ^2 (df)	Intervention versus AC[c], mean difference (95% CI)	Intervention versus WLC[d], mean difference (95% CI)
4 months	49.08 (9.52)	50.25 (9.42)	50.62 (10.63)	—	–2.49 (–6.01 to 1.03)	–0.87 (–4.40 to 2.67)
6 months	48.60 (9.82)	50.04 (9.66)	50.57 (10.44)	—	–1.51 (–5.04 to 2.02)	0.11 (–3.42 to 3.65)
PSS[j]						
Baseline	16.97 (6.10)	17.25 (5.79)	16.28 (5.53)	7.46 (6)	–0.28 (–2.29 to 1.73)	–0.97 (–2.98 to 1.04)
2 months	16.21 (6.46)	16.93 (5.86)	15.53 (5.75)	—	–0.42 (–2.42 to 1.58)	–0.93 (–2.94 to 1.07)
4 months	15.50 (6.87)	16 (6.18)	14.99 (6.15)	—	–0.76 (–2.77 to 1.25)	–1.67 (–3.68 to 0.34)
6 months	14.54 (7.03)	15.56 (6.46)	14.42 (6.51)	—	–0.90 (–2.90 to 1.10)	–0.90 (–2.91 to 1.11)
WHO-5[k]						
Baseline	60.67 (19.77)	64.15 (19.63)	59.17 (20.99)	2.95 (6)	–3.49 (–10.42 to 3.45)	–1.78 (–8.71 to 5.15)
2 months	62.18 (20.64)	64.86 (19.77)	59.65 (22.38)	—	–1.11 (–8.04 to 5.82)	1.75 (–5.18 to 8.68)
4 months	62.54 (21.60)	64.96 (19.89)	60.21 (22.37)	—	–1.18 (–8.11 to 5.74)	3.75 (–3.18 to 10.68)
6 months	62.04 (22.67)	65.75 (19.89)	60.90 (22.47)	—	–0.31 (–7.24 to 6.62)	2.58 (–4.35 to 9.51)
Self-esteem[l]						
Baseline	30.54 (5.13)	30.57 (4.06)	30.51 (5.30)	5.60 (6)	–0.03 (–1.55 to 1.50)	0.03 (–1.69 to 1.75)
2 months	30.06 (5.27)	30.43 (4.15)	29.93 (5.31)	—	–0.38 (–1.94 to 1.19)	0.12 (–1.62 to 1.87)
4 months	31.82 (5.41)	29.89 (4.11)	29.74 (5.42)	—	1.93 (0.35-3.52)	2.08 (0.30-3.87)
6 months	31.44 (5.48)	29.64 (4.03)	29.46 (5.27)	—	1.81 (0.22-3.39)	1.99 (0.21-3.76)
Loneliness[m]						
Baseline	41 (9.71)	40.07 (9.56)	40.97 (9.68)	3.07 (6)	0.93 (–2.24 to 4.11)	0.03 (–3.17 to 3.22)
2 months	40.03 (9.93)	39.29 (9.78)	40.26 (10.07)	—	0.74 (–2.51 to 3.98)	–0.24 (–3.53 to 3.06)
4 months	34.49 (9.77)	38.25 (9.52)	39.58 (10.30)	—	–3.76 (–6.93 to -.58)	5.09 (–8.40 to –1.78)

XSL•FO
RenderX

Outcomes and timepoint[a]	Intervention	Active control	Waitlist control	Int (LRT)[b], χ^2 (df)	Intervention versus AC[c], mean difference (95% CI)	Intervention versus WLC[d], mean difference (95% CI)
6 months	33.55 (9.94)	37.35 (9.69)	38.57 (10.45)	—	–3.79 (–7.03 to –0.56)	–5.02 (–8.38 to –1.66)

[a]Adjusted means (SE) are estimated from linear mixed‐effects models with a random intercept for each participant, including fixed effects for group (1 = intervention, active control, 3=waitlist control), time (baseline, 2 months, 4 months, 6 months), and their interaction. Interaction effects were tested via likelihood‐ratio tests (chi-square (df)); reported chi-square and degrees of freedom refer to the comparison between the full model (with group × time) and the reduced model (without the interaction). Pairwise contrasts (intervention vs active control; intervention vs waitlist control) give mean differences and 2‐sided 95% CIs. Positive values indicate higher scores in the intervention group.

[b]LRT: likelihood ratio test.

[c] AC: active control.

[d] WLC: waitlist control.

[e]Not applicable.

[f]MET: total weekly metabolic equivalent minutes.

[g]GAD: Generalized Anxiety Disorder‐7.

[h] PHQ: Patient Health Questionnaire‐9.

[i]WEMWBS: Warwick–Edinburgh Mental Well‐being Scale.

[j]PSS: Perceived Stress Scale.

[k]WHO‐5: 5-item World Health Organization Well‐Being Index.

[l]Self‐Esteem: Rosenberg Self‐Esteem Scale.

[m]Loneliness: UCLA Loneliness Scale.

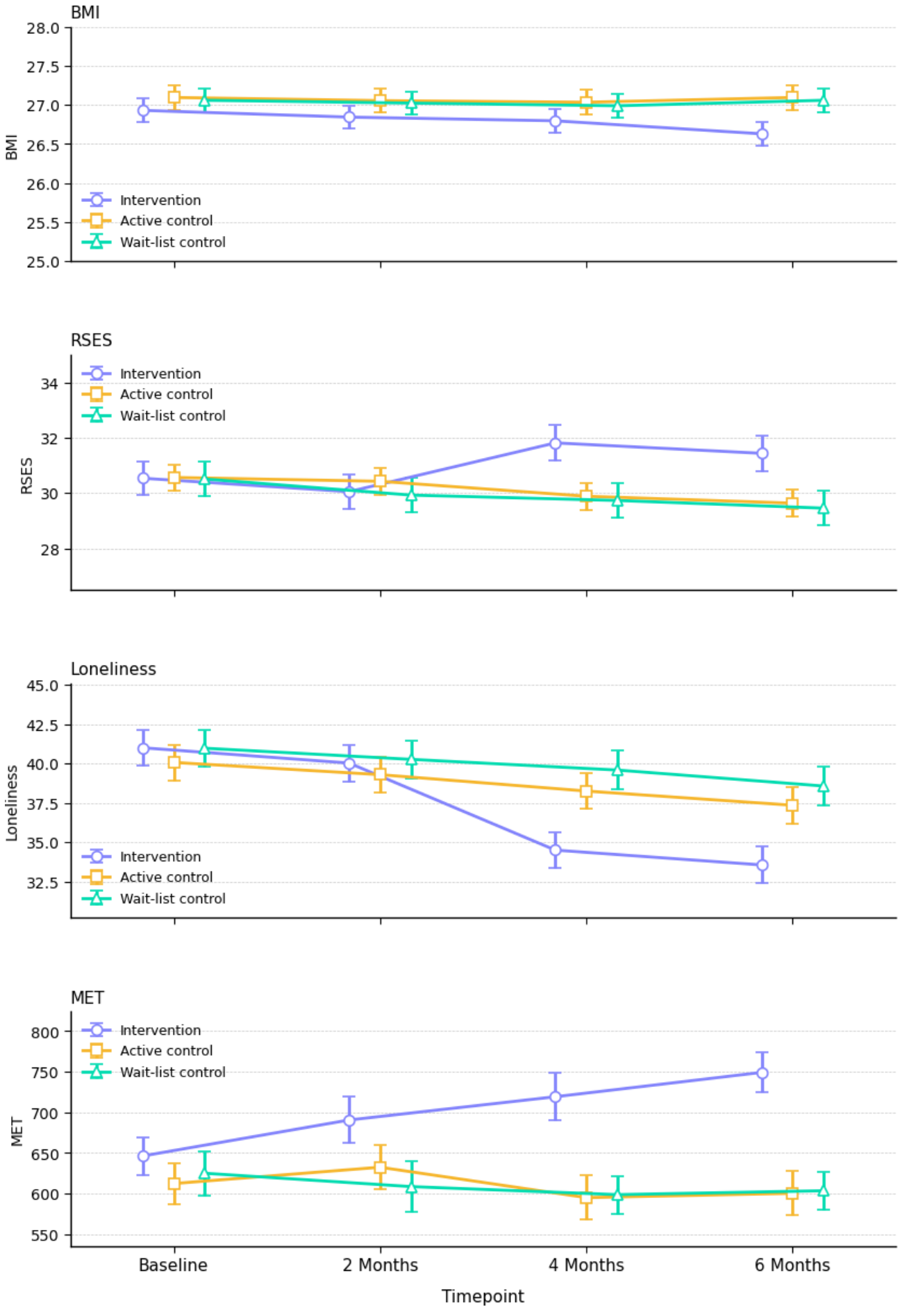
At 6 months, these benefits not only persisted but, in the case of BMI, became statistically significant. Although BMI differences were nonsignificant at 4 months, by 6 months the intervention group's mean BMI was 0.47 kg/m² lower than that of the active control (95% CI –0.89 to –0.04; *P*=.04) and 0.54 kg/m² lower than that of the waitlist control (95% CI –0.85 to –0.01; *P*=.03). Weekly MET‐minutes remained elevated among intervention participants (129.45 MET-minutes greater than active control; 95% CI 53.28-205.61, *P*<.001; 152.87 MET greater than waitlist control; 95% CI 68.35-237.38, *P*<.001). Similarly, self‐esteem differences persisted: intervention scores were 1.81 points higher than active control (95% CI 0.22-3.39; *P*=.02) and 1.99 points higher than waitlist control (95% CI 0.21-3.76; *P*=.03). Loneliness remained lower in the intervention arm, with mean scores 3.79 points below active control (95% CI –7.03 to –0.56; *P*=.03) and 5.02 points below waitlist control (95% CI –8.38 to –1.66; *P*<.001; Figure 2; Table 2).

Besides these intervention effects, no differences in mental health, specifically depression, anxiety, and stress, were observed between conditions, and no adverse or unintended events occurred in any study arm. Participants' evaluations of intervention quality did not differ between the intervention and control conditions, and follow‐up evaluations indicated high overall satisfaction rates among study participants.

XSL•FO
RenderX

Figure 2. Line charts of outcomes exhibiting significant changes over time in the intervention, active control, and waitlist control groups (mean, SE). MET: metabolic equivalent of task; RSES: Rosenberg Self-Esteem Scale.

XSL•FO
RenderX



XSL•FO
RenderX

Discussion

Principal Findings

In this RCT, a digitally delivered peer-support intervention yielded significant improvements in both physical and psychological outcomes among university students with preobesity. Compared with participants in the active control and waitlist control groups, those receiving digital peer support experienced a greater reduction in BMI over 6 months, accompanied by marked increases in weekly physical activity. Notably, the intervention also promoted significant enhancements in self-esteem and produced a meaningful decrease in loneliness scores. These convergent improvements demonstrate that a well-structured digital peer-support program can effectively address weight-related metrics while simultaneously bolstering mental well-being in a high-risk young adult population.

Comparison with Prior Work

When contextualized within the broader literature, our results both corroborate and extend existing findings. Prior meta-analyses of peer-support interventions have reported modest reductions in weight and BMI, often limited to decreases on the order of 0.5 - 1.0 kg and 0.1 - 0.2 kg/m², respectively [48]. In contrast, the BMI improvement observed here exceeded those averages, suggesting that delivering peer support through a comprehensive digital platform may enhance engagement and efficacy. Similarly, earlier work by Yeo and colleagues found that digital peer-support programs led to higher levels of physical activity [24]. Our trial confirms that young adults can achieve and maintain substantial activity gains when supported by peers in a virtual setting. More striking is the contrast between our findings on psychosocial outcomes and those of many prior weight management trials. Whereas some research has found limited or no improvement in self-esteem or loneliness following peer support, often attributing these null effects to insufficient social cohesion or lack of personalized engagement [8], our intervention clearly produced robust psychological benefits. The structure of our digital program, which combined interactive workshops, ongoing peer dialogues, gamified challenges, and personalized goal-setting, likely fostered a sense of community and accountability that translated into higher self-esteem and reduced perceived social isolation.

Potential Mechanisms

Several features of our study design and implementation may underlie these positive outcomes. First, we targeted college students with “preobesity,” a population at a critical inflection point that is often overlooked by both obesity-prevention initiatives and mental health programs [2]. By focusing on individuals with BMI between 25 and 29.9 kg/m², our intervention capitalized on a phase when behavior change is particularly feasible, as these students are frequently less entrenched in unhealthy habits and experience fewer physiological impediments compared to individuals already in the obese range [52]. Second, the digital nature of our peer-support platform enabled consistent, real-time interaction and social reinforcement, meeting students in their preferred communication channels. Daily micro-tasks, biweekly group challenges, and digital badges created an environment in which healthy behaviors became normalized and visible, thereby leveraging social comparison and accountability to strengthen motivation. Third, the inclusion of both an attention-matched active control and a waitlist control increased the rigor of our design. By demonstrating that the intervention’s benefits exceeded those of an equally engaging, wellness-focused program, we confirmed that the peer-support component itself was responsible for the superior outcomes, rather than the generic effects of participation or attention.

Behavioral and psychosocial theories provide mechanistic explanations for the observed effects. According to social cognitive theory, observing peers’ successes and receiving encouragement enhances self-efficacy, which in turn promotes adoption and maintenance of healthy behaviors [53]. In our study, participants saw fellow group members achieving step-count goals, sharing healthy meal ideas, and offering positive feedback, all of which contributed to a collective sense of possibility. As self-efficacy increased, participants likely felt more capable of meeting weekly activity targets and monitoring their weight, resulting in the higher MET scores and lower BMI values observed at 4 and 6 months. Concurrently, the online peer-support format fostered social connectedness and reduced feelings of isolation, with participants reporting a growing sense of belonging as they exchanged challenges and celebrated accomplishments in a respectful, nonjudgmental space [54]. Improvements in self-esteem likely stemmed from personal achievements, such as meeting fitness goals and from external validation like peer praise, thereby reinforcing a more positive self-concept [55]. Moreover, reductions in loneliness may have mitigated stress and negative affect, both of which are known barriers to sustained lifestyle change [54]. By alleviating emotional distress and enhancing social support simultaneously, the intervention created a synergistic environment in which mental well-being and physical health could rise together [56].

Implications

Finally, the integration of physical, behavioral, and psychosocial measures in our outcome assessment underscores the value of holistic intervention approaches. In contrast to many programs that focus narrowly on either weight loss or mental health, our trial demonstrates that digital peer support can generate parallel benefits across multiple domains. This alignment is particularly important in the university setting, where academic pressures, social transitions, and evolving peer networks intersect to shape students’ health behaviors and psychological states. By addressing these factors in concert, providing evidence-based guidance on nutrition and exercise alongside structured opportunities for emotional support, the intervention enabled participants to tackle weight management without sacrificing mental health, and vice versa. Such a comprehensive strategy not only enhances immediate outcomes, such as lower BMI, higher activity levels, improved self-esteem, and reduced loneliness, but also establishes a foundation for sustained long-term well-being. Given the extensive reach and accessibility of digital platforms, our findings underscore the importance of integrating peer-supported programming as a foundational element of university health promotion strategies.

XSL•FO
RenderX

Limitations

This study has several limitations. First, the use of self-reported weight and physical activity data may have introduced measurement error, given the well-established divergence between self-reported and objectively measured values. Second, relying on BMI as the sole physical outcome may not adequately capture changes in body composition, as BMI does not differentiate between fat and lean mass. Third, the study sample consisted exclusively of digitally literate, university students with preobesity, which may limit the generalizability of the findings to other populations, including individuals in different age groups, weight categories, or with limited digital access or engagement. Fourth, we observed no significant changes in some secondary outcomes. This null finding may be attributable to the relatively short follow-up period and limited sensitivity of the self-report instruments. Long-term follow-up is needed to determine whether improvements in BMI, physical activity, and mental health are sustained beyond the intervention and to evaluate any delayed effects on the incidence of obesity and mental health disorders. Fifth, we could not directly verify completion of some self-guided micro tasks at the individual level; future trials should incorporate objective sensing or device-based validation to confirm task completion without increasing participant burden.

Conclusions

In this RCT among university students with preobesity, a digital peer-support intervention led to improvements in both physical and mental health outcomes compared to a control condition. Participants receiving the program attained a modest reduction in BMI and increased physical activity, while also reporting higher self-esteem and lower loneliness than controls. These findings highlight the promise of interventions that leverage peer support via digital platforms, an accessible and replicable approach to addressing emerging health risks in young adults. The intervention's digital, peer-driven design enables broad dissemination across campus settings and beyond, engaging at-risk students who may not use traditional health services. Notably, this trial is among the first to target the preobesity stage with a holistic approach that concurrently addresses weight management and mental health. By intervening early with an integrated program, this approach prevents progression to obesity while mitigating psychological distress, illustrating a novel paradigm for young adult health promotion.

Acknowledgments

The authors thank all the participants for their time and involvement in this study, research assistants for their hard work throughout the project, and all the faculty members who helped coordinate and support the study. The authors further confirm that no generative artificial intelligence tools were used in the drafting or editing of this manuscript.

Funding

This work was supported by the Guangdong Provincial Sports Bureau Science & Technology Innovation and Sports Culture Development Program (GDSS2024N069) and by the Guangdong Provincial Education Science Planning Leading Group Office (Higher Education Special) (2024GXJK439).

Data Availability

The datasets generated or analyzed during this study are available in the Harvard Dataverse [57].

Authors' Contributions

Conceptualization: XL, TC, and RM
Methodology: RM and TL
Formal analysis: TM and TL
Data curation: TM
Investigation: XL and RM
Software and visualization: RM
Writing – original draft preparation: XL and RM
Writing – review and editing: XL and RM
Supervision: XL and TC

Conflicts of Interest

None declared.

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 1362 KB - mhealth_v14i1e78960_app1.pdf]

References

XSL•FO
RenderX

1. Cole TJ, Lobstein T. Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. Pediatr Obes 2012 Aug;7(4):284-294. [doi: 10.1111/j.2047-6310.2012.00064.x] [Medline: 22715120]
2. Bailey CP, Sharma S, Economos CD, Hennessy E, Simon C, Hatfield DP. College campuses' influence on student weight and related behaviours: a review of observational and intervention research. Obes Sci Pract 2020 Dec;6(6):694-707. [doi: 10.1002/osp4.445] [Medline: 33354347]
3. Odlaug BL, Lust K, Wimmelmann CL, et al. Prevalence and correlates of being overweight or obese in college. Psychiatry Res 2015 May 30;227(1):58-64. [doi: 10.1016/j.psychres.2015.01.029] [Medline: 25770354]
4. Ilić M, Pang H, Vlaški T, Grujičić M, Novaković B. Prevalence and associated factors of overweight and obesity among medical students from the Western Balkans (South-East Europe Region). BMC Public Health 2024 Jan 2;24(1):29. [doi: 10.1186/s12889-023-17389-7] [Medline: 38166959]
5. Dipietro L, Zhang Y, Mavredes M, et al. Physical activity and cardiometabolic risk factor clustering in young adults with obesity. Med Sci Sports Exerc 2020 May;52(5):1050-1056. [doi: 10.1249/MSS.0000000000002214] [Medline: 31764468]
6. Emon HH, Sarker S, Lima MSA, et al. Prevalence of overweight and obesity and their impact on academic performance and psychological well-being among university students in 2024 in Bangladesh. In: Ahmed SK, editor. PLoS ONE 2024;19(12):e0315321. [doi: 10.1371/journal.pone.0315321] [Medline: 39656712]
7. Dakanalis A, Voulgaridou G, Alexatou O, et al. Overweight and obesity is associated with higher risk of perceived stress and poor sleep quality in young adults. Medicina (Kaunas) 2024 Jun 14;60(6):983. [doi: 10.3390/medicina60060983] [Medline: 38929600]
8. Streram S, Burrows T, Duncan MJ, Hutchesson M. Health behaviour interventions to improve mental health outcomes for students in the university setting: a systematic review of randomised controlled trials. Int J Behav Nutr Phys Act 2025 Mar 11;22(1):32. [doi: 10.1186/s12966-025-01718-7] [Medline: 40069770]
9. Lanoye A, Brown KL, LaRose JG. The transition into young adulthood: a critical period for weight control. Curr Diab Rep 2017 Oct 2;17(11):114. [doi: 10.1007/s11892-017-0938-4] [Medline: 28971312]
10. Vadeboncoeur C, Townsend N, Foster C. A meta-analysis of weight gain in first year university students: is freshman 15 a myth? BMC Obes 2015;2(1):22. [doi: 10.1186/s40608-015-0051-7] [Medline: 26217537]
11. West DS, Monroe CM, Turner-McGrievy G, et al. A technology-mediated behavioral weight gain prevention intervention for college students: controlled, quasi-experimental study. J Med Internet Res 2016 Jun 13;18(6):e133. [doi: 10.2196/jmir.5474] [Medline: 27296086]
12. Harrison S, Dixon P, Jones HE, Davies AR, Howe LD, Davies NM. Long-term cost-effectiveness of interventions for obesity: a Mendelian randomisation study. In: Veerman JL, editor. PLoS Med 2021 Aug;18(8):e1003725. [doi: 10.1371/journal.pmed.1003725] [Medline: 34449774]
13. Thorpe KE, Joski PJ. Estimated reduction in health care spending associated with weight loss in adults. JAMA Netw Open 2024 Dec 2;7(12):e2449200. [doi: 10.1001/jamanetworkopen.2024.49200] [Medline: 39636635]
14. Bjerregaard LG, Jensen BW, Ängquist L, Osler M, Sørensen TIA, Baker JL. Change in overweight from childhood to early adulthood and risk of type 2 diabetes. N Engl J Med 2018 Apr 5;378(14):1302-1312. [doi: 10.1056/NEJMoa1713231] [Medline: 29617589]
15. Gortmaker SL, Wang YC, Long MW, et al. Three interventions that reduce childhood obesity are projected to save more than they cost to implement. Health Aff (Millwood) 2015 Nov;34(11):1932-1939. [doi: 10.1377/hlthaff.2015.0631] [Medline: 26526252]
16. The Heavy Burden of Obesity: The Economics of Prevention: OECD; 2019. [doi: 10.1787/67450d67-en]
17. Dhar D, Packer J, Michalopoulou S, et al. Assessing the evidence for health benefits of low-level weight loss: a systematic review. Int J Obes (Lond) 2025 Feb;49(2):254-268. [doi: 10.1038/s41366-024-01664-7] [Medline: 39487296]
18. Pfisterer J, Rausch C, Wohlfarth D, Bachert P, Jekauc D, Wunsch K. Effectiveness of physical-activity-based interventions targeting overweight and obesity among university students-a systematic review. Int J Environ Res Public Health 2022 Aug 1;19(15):9427. [doi: 10.3390/ijerph19159427] [Medline: 35954789]
19. Cao B, Xu J, Li R, Teopiz KM, McIntyre RS, Chen H. Interventions targeting comorbid depression and overweight/obesity: a systematic review. J Affect Disord 2022 Oct 1;314:222-232. [doi: 10.1016/j.jad.2022.07.027] [Medline: 35878825]
20. Oftedal S, Fenton S, Hansen V, et al. Changes in physical activity, diet, sleep, and mental well-being when starting university: a qualitative exploration of Australian student experiences. J Am Coll Health 2024 Dec;72(9):3715-3724. [doi: 10.1080/07448481.2023.2194426] [Medline: 37014766]
21. Lee KM, Hunger JM, Tomiyama AJ. Weight stigma and health behaviors: evidence from the eating in America study. Int J Obes (Lond) 2021 Jul;45(7):1499-1509. [doi: 10.1038/s41366-021-00814-5] [Medline: 33934109]
22. Flint SW, Vázquez-Velázquez V, Le Brocq S, Brown A. The real-life experiences of people living with overweight and obesity: a psychosocial perspective. Diabetes Obes Metab 2025 Apr;27 Suppl 2(Suppl 2):35-47. [doi: 10.1111/dom.16255] [Medline: 39931901]
23. Jones RA, Lawlor ER, Birch JM, et al. The impact of adult behavioural weight management interventions on mental health: a systematic review and meta-analysis. Obes Rev 2021 Apr;22(4):e13150. [doi: 10.1111/obr.13150] [Medline: 33103340]

XSL•FO
RenderX

24. Yeo G, Fortuna KL, Lansford JE, Rudolph KD. The effects of digital peer support interventions on physical and mental health: a review and meta-analysis. Epidemiol Psychiatr Sci 2025 Feb 13;34:e9. [doi: 10.1017/S2045796024000854] [Medline: 39945388]
25. Ufholz K. Peer support groups for weight loss. Curr Cardiovasc Risk Rep 2020 Oct;14(10):19. [doi: 10.1007/s12170-020-00654-4]
26. Osborn TG, Town R, Ellis R, Buckman JEJ, Saunders R, Fonagy P. Implementing peer support in higher education: a feasibility study. SSM Ment Health 2022 Dec;2:100175. [doi: 10.1016/j.ssmmh.2022.100175] [Medline: 37916032]
27. D'Adamo L, Paraboschi L, Grammer AC, et al. Reach and uptake of digital mental health interventions based on cognitive-behavioral therapy for college students: a systematic review. J Behav Cogn Ther 2023 Jun;33(2):97-117. [doi: 10.1016/j.jbct.2023.05.002] [Medline: 37724304]
28. Toon J, Geneva M, Sharpe P, Lavin J, Bennett S, Avery A. Weight loss outcomes achieved by adults accessing an online programme offered as part of Public Health England's "Better Health" campaign. BMC Public Health 2022 Jul 30;22(1):1456. [doi: 10.1186/s12889-022-13847-w] [Medline: 35907834]
29. Stokols D. Establishing and maintaining healthy environments. Toward a social ecology of health promotion. Am Psychol 1992 Jan;47(1):6-22. [doi: 10.1037//0003-066x.47.1.6] [Medline: 1539925]
30. McLeroy KR, Bibeau D, Steckler A, Glanz K. An ecological perspective on health promotion programs. Health Educ Q 1988;15(4):351-377. [doi: 10.1177/109019818801500401] [Medline: 3068205]
31. Craig CL, Marshall AL, Sjöström M, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
32. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
33. Zhang C, Wang T, Zeng P, et al. Reliability, validity, and measurement invariance of the General Anxiety Disorder Scale among Chinese Medical University students. Front Psychiatry 2021;12:648755. [doi: 10.3389/fpsyt.2021.648755] [Medline: 34093269]
34. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
35. Wang W, Bian Q, Zhao Y, et al. Reliability and validity of the Chinese version of the Patient Health Questionnaire (PHQ-9) in the general population. Gen Hosp Psychiatry 2014;36(5):539-544. [doi: 10.1016/j.genhosppsych.2014.05.021] [Medline: 25023953]
36. Tennant R, Hiller L, Fishwick R, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Nov 27;5(1):63. [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
37. Fung SF. Psychometric evaluation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) with Chinese university students. Health Qual Life Outcomes 2019 Mar 14;17(1):46. [doi: 10.1186/s12955-019-1113-1] [Medline: 30871563]
38. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [doi: 10.2307/2136404] [Medline: 6668417]
39. Lu W, Bian Q, Wang W, Wu X, Wang Z, Zhao M. Chinese version of the Perceived Stress Scale-10: a psychometric study in Chinese university students. In: Dang Y, editor. PLoS ONE 2017;12(12):e0189543. [doi: 10.1371/journal.pone.0189543] [Medline: 29252989]
40. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176. [doi: 10.1159/000376585] [Medline: 25831962]
41. Fung SF, Kong CYW, Liu YM, et al. Validity and psychometric evaluation of the Chinese version of the 5-item WHO Well-Being Index. Front Public Health 2022;10:872436. [doi: 10.3389/fpubh.2022.872436] [Medline: 35433612]
42. Rosenberg M, Schooler C, Schoenbach C, Rosenberg F. Global self-esteem and specific self-esteem: different concepts, different outcomes. Am Sociol Rev 1995 Feb;60(1):141. [doi: 10.2307/2096350]
43. Jiang C, Zhu Y, Luo Y, et al. Validation of the Chinese version of the Rosenberg Self-Esteem Scale: evidence from a three-wave longitudinal study. BMC Psychol 2023 Oct 18;11(1):345. [doi: 10.1186/s40359-023-01293-1] [Medline: 37853499]
44. Russell DW. UCLA Loneliness Scale (version 3): reliability, validity, and factor structure. J Pers Assess 1996 Feb;66(1):20-40. [doi: 10.1207/s15327752jpa6601_2] [Medline: 8576833]
45. Ip H, Suen YN, Hui LMC, Cheung C, Wong SMY, Chen EYH. Psychometric properties of the variants of the Chinese UCLA Loneliness Scales and their associations with mental health in adolescents. Sci Rep 2024 Oct 21;14(1):24663. [doi: 10.1038/s41598-024-75739-w] [Medline: 39433867]
46. Lee RM, Robbins SB. Measuring belongingness: the social connectedness and the social assurance scales. J Couns Psychol 1995;42(2):232-241. [doi: 10.1037//0022-0167.42.2.232]
47. Yoon E, Jung KR, Lee RM, Felix-Mora M. Validation of social connectedness in mainstream society and the ethnic community scales. Cultur Divers Ethnic Minor Psychol 2012 Jan;18(1):64-73. [doi: 10.1037/a0026600] [Medline: 22250899]
48. Gold N, Yau A, Rigby B, Dyke C, Remfry EA, Chadborn T. Effectiveness of digital interventions for reducing behavioral risks of cardiovascular disease in nonclinical adult populations: systematic review of reviews. J Med Internet Res 2021 May 14;23(5):e19688. [doi: 10.2196/19688] [Medline: 33988126]

XSL•FO
RenderX

49. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
50. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods 2002 Jun;7(2):147-177. [Medline: 12090408]
51. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1). [doi: 10.18637/jss.v067.i01]
52. Burgess E, Hassmén P, Pumpa KL. Determinants of adherence to lifestyle intervention in adults with obesity: a systematic review. Clin Obes 2017 Jun;7(3):123-135. [doi: 10.1111/cob.12183] [Medline: 28296261]
53. Islam KF, Awal A, Mazumder H, et al. Social cognitive theory-based health promotion in primary care practice: a scoping review. Heliyon 2023 Apr;9(4):e14889. [doi: 10.1016/j.heliyon.2023.e14889] [Medline: 37025832]
54. Bravata DM, Kim J, Russell DW, Goldman R, Pace E. Digitally enabled peer support intervention to address loneliness and mental health: prospective cohort analysis. JMIR Form Res 2023 Nov 6;7:e48864. [doi: 10.2196/48864] [Medline: 37930770]
55. Asadabadi FS, Karami K. The effect of exercise training videos on self-esteem of 7-10-year-old obese girls: a randomized controlled trial. BMC Public Health 2025 May 7;25(1):1684. [doi: 10.1186/s12889-025-22832-y] [Medline: 40336009]
56. Holt-Lunstad J. Social connection as a critical factor for mental and physical health: evidence, trends, challenges, and future implications. World Psychiatry 2024 Oct;23(3):312-332. [doi: 10.1002/wps.21224] [Medline: 39279411]
57. Efficacy of a digital peer support program on weight management and mental health in pre-obese university students: a randomized controlled trial. HARVARD Dataverse. URL: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/BZGDBB [accessed 2026-01-07]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
GAD-7: Generalized Anxiety Disorder Scale 7
MET: metabolic equivalent of task
PHQ-9: Patient Health Questionnaire-9
PSS: Perceived Stress Scale
RCT: randomized controlled trial
SCS: Social Connectedness Scale
UCLA: University of California, Los Angeles Loneliness Scale
WEMWBS: Warwick-Edinburgh Mental Well-being Scale
WHO-5: 5-item World Health Organization Well-Being Index

Edited by L Buis; submitted 12.Jun.2025; peer-reviewed by D Tang, J Liu; revised version received 03.Oct.2025; accepted 16.Dec.2025; published 21.Jan.2026.

Please cite as:
Liu X, Liu T, Chen T, Ma R
Efficacy of a Digital Peer Support Program on Weight Management and Mental Health in University Students With Preobesity: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e78960
URL: https://mhealth.jmir.org/2026/1/e78960
doi:10.2196/78960

© Xingyu Liu, Ting Liu, Tao Chen, Ruisi Ma. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 21.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Analysis of Training Behavior in Users of a Fitness App: Cross-Sectional Study

Andrea Fuente-Vidal[1,2], PhD; Roger Prat[1], MSc; Juan Manuel Arribas-Marin[3,4], PhD; Oscar Bastidas-Jossa[5], PhD; Myriam Guerra-Balic[1], MD, PhD; Begonya Garcia-Zapirain[5], PhD; Joel Montane[1,6*], PhD; Javier Jerez-Roig[2,7,8*], PhD

[1]Faculty of Psychology, Education and Sports Sciences (FPCEE) Blanquerna, Universitat Ramon Llull, Barcelona, Spain

[2]Research Group on Methodology, Methods, Models and Outcomes of Health and Social Sciences (M3O), Faculty of Health Sciences and Welfare, Centre for Health and Social Care Research (CESS), Universitat de Vic - Universitat Central de Catalunya, Vic, Spain

[3]San Juan de Dios University School of Nursing and Physical Therapy, Comillas Pontifical University, Madrid, Spain

[4]San Juan de Dios Foundation, Madrid, Spain

[5]eVIDA Research Group, Universidad de Deusto, Bilbao, Spain

[6]School of Health Sciences, Blanquerna, Universitat Ramon Llull, Barcelona, Spain

[7]Institute of Sport Science and Innovations, Lithuanian Sports University, Kaunas, Lithuania

[8]Institute for Research and Innovation in Life and Health Sciences in Central Catalonia (IRIS-CC), Universitat de Vic - Universitat Central de Catalunya, Vic, Spain

[*]these authors contributed equally

Corresponding Author:
Joel Montane, PhD
Faculty of Psychology, Education and Sports Sciences (FPCEE) Blanquerna
Universitat Ramon Llull
Carrer del Císter, 34
Barcelona, 08022
Spain
Phone: 34 932 53 30 00
Email: joelmm@blanquerna.url.edu

Abstract

Background: Mobile health (mHealth) apps are increasingly being used to promote physical activity (PA) and can support exercise uptake and maintenance. Despite their potential, these tools face high dropout rates and inconsistent adherence, posing a significant challenge. Understanding how users engage with fitness apps is essential for improving user experience and health outcomes.

Objective: This study aims to analyze user behavior patterns in the Mammoth Hunters (MH) fitness app (Mammoth Hunters SL), focusing on retention (days from registration to user's last recorded training session), average weekly training frequency, and adherence (alignment between planned and actual training). We examined how these outcomes are influenced by sociodemographic, motivational, and other variables.

Methods: This cross-sectional study involved 2771 Mammoth Hunters app users. In a subsample (n=289), training data were complemented by motivational data acquired through online surveying via an ad-hoc scale (internal consistency >0.83) based on the self-determination theory (SDT). Descriptive statistics and nonparametric tests (Kruskal-Wallis, Dunn post-hoc, and Spearman correlation) were used to assess correlation between sociodemographic, motivation, and training behavior variables.

Results: Mean retention (days) was significantly higher among males than females (135 vs 109, respectively; $P<.01$), users in the subscription vs free plan (154 vs 81; $P<.001$), active or very active individuals vs inactive, midbuilt vs thin body types (132 vs 120; $P=.001$), and those with slightly lower BMI. Users pursuing antiaging or muscle gain goals showed longer retention than those aiming to lose weight (gain: 132, antiaging: 128, lose weight: 116; $P<.001$). Average weekly frequency (sessions per week) of training was statistically significantly different by sex (male: 1.9 vs female: 1.8; $P=.04$), body type (thin: 1.96 vs mid: 1.77; $P=.04$), activity level (very active: 2.05 vs inactive: 1.83; $P=.04$), and motivation type (extrinsic introjected motivation correlated positively: $r=0.17$; $P<.05$), but did not correlate with perceived difficulty or fitness goals. Adherence, defined as actual vs targeted training frequency, was only significantly different among body types, with thin users showing higher adherence than the midbuilt

XSL•FO
RenderX

group (57% vs 52.1%; *P*=.02). Intrinsic motivation showed a positive correlation with retention (*r*=0.19; *P*=.002), as did identified motivation (*r*=0.12; *P*<.05).

Conclusions: This study shows that retention is influenced by demographic factors, with males, subscribers, previously active, midbuilds, those aiming to gain muscle, and individuals with autonomous types (ie, intrinsic and identified) of motivation displaying greater long-term participation. These findings provide valuable preliminary insight into the complexities of exercise training behavior in apps. They suggest that training frequency, retention, and adherence do not respond to the same factors. App developers, researchers, and trainers should assess these variables separately and develop strategies accordingly.

(JMIR Mhealth Uhealth 2026;14:e72201) doi:10.2196/72201

KEYWORDS

fitness app; physical activity; exercise adherence; retention; motivation; mHealth

Introduction

Adherence to Physical Activity and Exercise

Physical activity (PA) and exercise are fundamental components of a healthy lifestyle, with well-established benefits for physical and mental well-being. PA, as defined by the World Health Organization (WHO), encompasses any bodily movement that results in energy expenditure, while exercise is considered a structured subset of PA, performed with the intent of improving or maintaining physical fitness [1]. Regular engagement in PA is crucial for reducing the risk of chronic diseases, yet adherence to recommended activity levels remains a global challenge [2]. Despite the widespread awareness of PA benefits, sustaining an active lifestyle is often hindered by behavioral, environmental, and psychological barriers [3]. Understanding factors that influence adherence is therefore critical for improving PA participation and ensuring long-term engagement.

Influence of mobile health on PA Behaviors

In recent years, mobile health (mHealth) apps have emerged as a potential solution to bridge the gap between PA recommendations and actual adherence. Fitness apps, a subset of mHealth, offer structured training programs, progress tracking, and personalized feedback, aiming to enhance motivation and user engagement. The widespread availability of smartphones has contributed to a surge in fitness app usage, with millions of users accessing digital exercise programs globally [4]. These apps incorporate behavior change techniques such as goal setting, social support, and gamification to facilitate sustained exercise habits [5]. However, despite their potential, high attrition rates and inconsistent long-term adherence pose significant challenges to their effectiveness [6]. Recent evidence reinforces that these barriers still persist across different populations and intervention designs. For example, several authors reported significant dropout rates even in gamified or socially incentivized fitness apps [7,8]. Similarly, previous studies highlighted continued adherence challenges in young individuals or older adults despite tailored mHealth interventions [6,7,9].

Adherence to exercise, particularly in digital interventions, remains a complex issue, often inconsistently defined across studies. Traditional adherence models typically assess exercise frequency, duration, and intensity, yet these criteria may not fully capture engagement in app-based fitness programs [3]. Furthermore, users may abandon apps due to technical difficulties, loss of motivation, or unrealistic expectations [10]. Consequently, understanding the determinants of fitness app adherence requires a multidimensional approach, integrating psychological, technological, and behavioral perspectives [11].

The challenges surrounding fitness app adherence are compounded by factors such as user characteristics, app usability, and the broader social and environmental contexts in which users engage with digital interventions. Studies have highlighted that individual attributes such as age, sex, health consciousness, and baseline PA levels may influence the likelihood of sustained engagement with fitness apps [12]. Additionally, app design elements, including intuitive navigation, feedback mechanisms, and interactive features, play a crucial role in user retention [13]. Social and motivational factors, such as competition, social support, and reinforcement strategies, have also been shown to impact adherence levels in digital exercise interventions [5]. Recent evidence highlights the importance of incorporating behavioral theories and enhancing usability and perceived value in reducing attrition and promoting sustained engagement with mHealth tools. For instance, recent findings emphasize the relevance of behavioral theories in crafting more effective mHealth interventions, showing how tailored features can reduce dropout and improve user retention [14]. Similarly, perceived value and usability have been identified as key drivers of long-term engagement with digital health tools [15]. Other studies suggest that personalization, motivational strategies, and social features are critical to increasing user commitment [12,16], overall highlighting the multifactorial nature of adherence and reinforcing the need for user-centered app design approaches.

Given the rising reliance on digital solutions for health and fitness, it is imperative to explore how different aspects of fitness apps contribute to sustained PA engagement. Previous research presents mixed findings on the long-term efficacy of fitness apps in promoting adherence, with some studies reporting positive behavior changes and others limited long-term impact [8,17].

Understanding real users' training behavior, beyond theoretical frameworks or self-reported intentions, is essential to identify how engagement translates into real-world usage. Analyzing app usage data provides evidence of behavior patterns, allowing researchers to identify which user profiles are more likely to sustain app use. This study establishes a theoretical and practical differentiation between adherence and retention, and how they

XSL•FO
RenderX

relate to user motivation to exercise. This is a critical matter, since sustained usage is key to ensuring the long-term impact of digital health interventions [18].

Study Goal

This study aims to explore the factors influencing the training behavior of users of a fitness app, focusing specifically on exercise adherence, retention, and motivation, and to explore how these outcomes are influenced by sociodemographic, motivational, and training-related factors. We hypothesize that sociodemographic characteristics, motivation types, and training behaviors significantly influence users' retention, adherence, and frequency of training. Understanding these factors is essential for optimizing fitness app design, improving intervention strategies, and ultimately promoting long-term participation in exercise.

Methods

Data Collection and Processing

Data were collected in collaboration with Mammoth Hunters (MH; Mammoth Hunters SL), a fitness app that focused on high-intensity interval exercises to improve strength, endurance, and mobility. MH delivered structured programs rooted in functional movement, making it an ideal platform for investigating digital fitness adherence, motivation, and retention. MH was launched in 2014 by a team of fitness experts and scientists from Barcelona, Spain. A free version with limited access to certain workouts and features was available upon registration, while the Pro version (per subscription) provided full access to personalized plans, a greater variety of workouts, and advanced tracking tools. MH ceased operations in September 2021, being one of the most widely used high-intensity training apps worldwide, having accumulated a total of 719,421 users. The company's shutdown, as well as its noninvolvement in any of the stages of study, ensured no conflict of interest.

The study used a cross-sectional design. Data in the MH app database included user registries ranging from November 21, 2020, to May 27, 2022. Motivation data were collected via online surveying on March 20, 2022. Data cleaning and descriptive analyses were conducted using the R programming language (version 4.3.1; R Core Team) in the R Studio environment software (version 2023.9.1.494; Posit, PBC).

The initial MH dataset contained 5858 entries, which corresponded to users who had granted informed consent to share their deidentified data for analysis. To ensure the accuracy and relevance of the data, several cleaning steps were executed, including a convenience selection of the most relevant variables and exclusion of registries with insufficient or missing data (see Table S1 in Multimedia Appendix 1 for more detail of the MH app's original variables).

Outliers were identified through data visualization and consultation of descriptive statistics. A decision was made to remove all outliers to avoid distortion in the analysis, based on two main reasons: (1) certain tests run by MH staff members had intentionally introduced outlier scores to facilitate their identification and removal, and (2) some outliers resulted from the arbitrary temporal cutoff applied to the dataset, specifically, some participants had only just begun training with the app shortly before the data extraction date, leading to unrealistic extreme values in some outcome variables (eg, extremely low retention or extremely high adherence). Therefore, all values exceeding Q3+1.5×IQR or falling below Q1–1.5×IQR were removed.

The motivation-related outcome variables were determined using the means of the composite scores from the observable items corresponding to each factor in the scale, allowing us to define the latent variables. Outcome variables retention, frequency, and adherence were derived from existing variables on the mobile app (eg, number of sessions executed, last training date, and user sign-up date) to enhance analytical depth. User retention was calculated as the number of days between a user's initial registration date within the fitness app and their last recorded training session. Weekly training frequency was calculated by dividing the total number of training sessions completed by a user by the total number of weeks from their first to their last executed session. Adherence was quantified as the percentage to which a user's actual average weekly training frequency aligned with their initial plan (self-declared upon registration). Additionally, data types were adjusted as required to ensure compatibility and accuracy.

Following the described data cleaning steps, the final dataset comprised 2771 user entries. See Table 1 for more details on study variables.

XSL•FO
RenderX

Table 1. Description of explanatory and outcome variables.

Explanatory variables	Type	Description
Sex	Categorical; sociodemographic	User biological sex, with 2 categories: female and male.
Body type	Categorical; sociodemographic	Self-reported body type selected by the user at registration, out of 3 available categories: thin, mid, and strong.
Activity level	Categorical; sociodemographic	User activity level at the time of registration in the fitness app, with 3 categories: inactive, active, and very active.
Fitness goal	Categorical; sociodemographic	The goal the user aims to achieve through app use (selected from 3 available categories: lose weight, gain muscle, and antiaging).
Pro version	Categorical; training	Indicative of app user being subscribed to a payment ("Pro") program or not. Two categories: yes and no.
Training schedule	Categorical; training	Time of day in which the user executes most (>50%) of their training sessions. Processed into 3 categories: morning (5:30-12:30 hours), afternoon (12:31-20 hours), and night (20-5:29 hours).
Age	Numerical; sociodemographic	User's reported age at registration.
BMI	Numerical; sociodemographic	User BMI calculated from their declared height and weight.
Subjective body fat	Numerical; sociodemographic	Users' self-reported body fat.
Difficulty	Numerical; training	Average perceived exertion reported at the end of the training session. 0 (lowest)-10 (highest).
Enjoyment	Numerical; training	Average user-reported enjoyment after each training session. 0 (lowest)-5 (highest).
SDT[a]-based variables		
Intrinsic motivation	Numerical; motivation	Average score of the intrinsic motivation items on the scale: a decimal number between 1 (lowest) and 5 (highest).
Identified extrinsic motivation	Numerical; motivation	Average score of the identified extrinsic motivation items on the scale: a decimal number between 1 (lowest) and 5 (highest).
Introjected extrinsic motivation	Numerical; motivation	Average score of the introjected extrinsic motivation items on the scale: a decimal number between 1 (lowest) and 5 (highest).
Outcome variables		
Retention	Numerical; training	Measured as the total number of days from the user registration date in the app to their last recorded training session.
Frequency, weekly average	Numerical; training	Calculated by dividing the total number of user sessions by the total number of weeks between their first and last recorded sessions.
Adherence	Percentage; training	Defined as the percentage alignment between the user's actual weekly training frequency and their predefined weekly training goal.

[a]SDT: self-determination theory.

Motivational data, which had been previously collected (March 20, 2022) by means of an ad-hoc scale (Table S2 in Multimedia Appendix 1), provided insight as to the motivational regulation of a subsample (n=753) of MH users. The scale was based on the self-determination theory (SDT) [19,20]. It showed good fit indices and a 3-factor structure as confirmed per exploratory and confirmatory factor analyses, with internal consistency indices >0.830 for the 3 subscales (intrinsic, identified extrinsic, and introjected extrinsic motivations). Data obtained through surveying (n=753) and data obtained from the MH fitness app (n=2771) were then merged, and a sample consisting of n=328, for which both training and motivational data were available, was obtained. Thirty-nine registries had to be disregarded due to missing data for the calculation of adherence and weekly training frequency outcome variables. A resulting total of 289 was complete for all explanatory and outcome study variables.

XSL•FO
RenderX

Descriptive Analysis for Sociodemographic, Training, and Motivation Variables

Following data cleaning, descriptive statistics were computed to summarize and describe the dataset's characteristics. Frequencies and percentages were calculated for categorical variables to provide an overview of their distribution and proportions within the sample. For numerical variables, measures of central tendency (mean and median) and measures of dispersion (minimum, maximum, and quartiles) were obtained to characterize data distribution and its variability.

Inferential Analysis of Explanatory Variables

Normality was assessed using the D'Agostino-Pearson test, along with skewness and kurtosis coefficients to quantify distributional properties. Additionally, histograms and quantile-quantile plots were inspected to visually evaluate deviations from normality. The results indicated significant departures from normality, and nonparametric tests were used for subsequent analyses. The Kruskal-Wallis test was conducted to evaluate differences among groups, followed by post-hoc analysis using the Bonferroni correction to adjust for multiple comparisons. The effect size was assessed using Dunn, which quantifies the magnitude of observed differences. To examine relationships between numeric variables, Spearman correlation tests were performed. The Holm correction was applied to control for multiple comparisons and to adjust the significance levels accordingly. For training behavior analysis, the categorical variables evaluated were sex, Pro version, self-declared level of previous PA, body type, fitness goal, and training schedule. Additionally, explanatory numerical variables included age, BMI, subjective body fat, perceived difficulty, and enjoyment. Intrinsic, identified extrinsic, and introjected extrinsic motivations were also considered (Table 1). The relationship of all these variables was analyzed with three outcome variables: adherence, frequency, and retention.

Inferential Analysis of Outcome Variables

Adherence was calculated as the percentage match between the target weekly frequency, as selected by the user upon sign-up, and the actual, executed weekly training frequency. The latter was averaged by dividing the total number of executed sessions by the total number of weeks from the sign-up date to the last executed session for the given user. Finally, retention was measured as the total number of days from the user sign-up date to the user's last recorded training session (Table 1). Outcome variables were analyzed through the same procedures as explanatory variables.

Ethical Considerations

Ethical approval for this study was obtained from the Research Ethics Committee of Universitat Ramon Llull in March 2020 (reference code 1920003P). All included users provided informed consent to use their data for research purposes, either through the app at registration or through the motivational survey. All data were anonymized before analysis, ensuring the privacy and confidentiality of participants in compliance with data protection regulations. No compensation was provided to participants, as the data were collected retrospectively and only for research purposes.

Results

Descriptive Results for Sociodemographic, Training, and Motivation Variables

Our sample consisted of 2771 MH users. Of them, a 64.8% majority identified as male, and 35.2% as female. Their age range spanned from 21 to 64 years, with a median age of 43 years and a mean age of 42.45 years. Users' fitness goals varied, with the largest segment (46.6%) aiming to "gain muscle" mass, followed by those wanting to "lose weight" at 32%. A smaller portion, 21.4%, pursued "antiaging" benefits. Table 2 provides full details on sample description and other results.

XSL•FO
RenderX

Table 2. Descriptive characteristics of the Mammoth Hunters user sample (N=2771).

Variable	Values
Sex, n (%)	
Male	1796 (64.8)
Female	975 (35.2)
Age (years)	
Mean (SD)	42.45 (9.8)
Median (IQR)	43 (21)
Range (minimum-maximum)	21-64
Body type, n (%)	
Thin	1269 (45.8)
Midbuild	1419 (51.2)
Strong	191 (6.9)
Body fat	
Mean (SD)	22.2 (6.4)
Median (IQR)	20 (14)
Range (minimum-maximum)	6-40
BMI	
Mean (SD)	23.46 (2.9)
Median (IQR)	23.44 (3.8)
Subscription type, n (%)	
Pro (paid) users	1726 (62.3)
Standard (free) users	768 (27.7)
Physical activity level, n (%)	
Active	1640 (59.2)
Very active	454 (16.4)
Inactive	675 (24.4)
Actual training schedule, n (%)	
Morning	928 (33.5)
Afternoon	1372 (49.5)
Night	471 (17)
Training difficulty (1-10)	
Mean (SD)	5.56 (1.78)
Median (IQR)	5.40 (2.10)
Range (minimum-maximum)	1-9.5
Enjoyment (1-5)	
Mean (SD)	3.58 (0.82)
Median (IQR)	3.50 (1.20)
Motivation (1-5)	
Intrinsic	
Mean (SD)	4.01 (0.74)
Median (IQR)	4 (0.90)
Range (minimum-maximum)	1.5-5
Identified extrinsic	

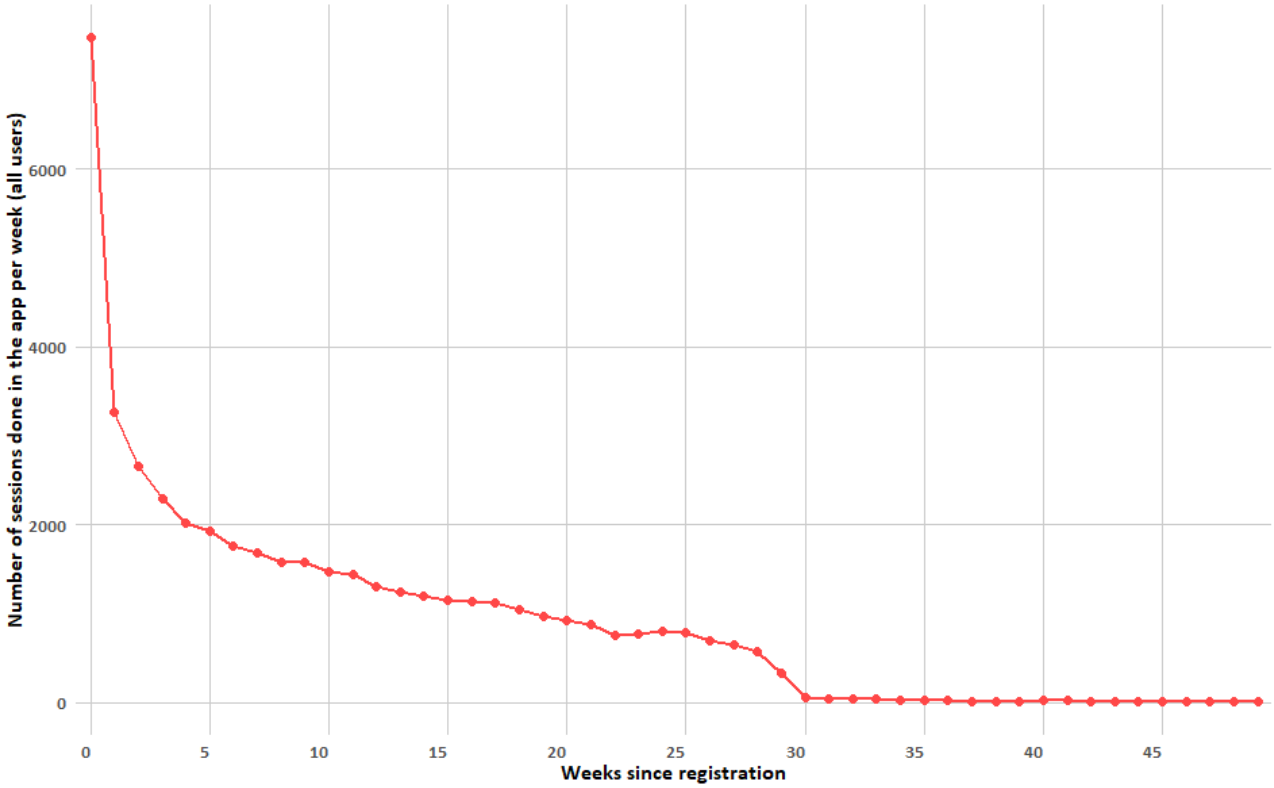
XSL•FO
RenderX

Variable	Values
Mean (SD)	4.42 (0.50)
Median (IQR)	4.67 (0.67)
Range (minimum-maximum)	3-5
Introjected extrinsic	
Mean (SD)	4.22 (0.61)
Median (IQR)	4.33 (1)
Range (minimum-maximum)	1-5
Retention (days)	
Mean (SD)	125.99 (92.60)
Median (IQR)	132.72 (144.36)
Range (minimum-maximum)	3.49-410.82
Training frequency (sessions per week)	
Mean (SD)	1.87 (1.52)
Median (IQR)	1.62 (1.84)
Range (minimum-maximum)	0.07-6.59
Adherence (%)	
Mean (SD)	54.24 (32.81)
Median (IQR)	47.31 (52.81)
Range (minimum-maximum)	1.18-166.67

Figure 1 illustrates the number of training sessions completed each week, by the total number of users, over a 49-week span. At the beginning, there was a sharp peak in the number of training sessions, with 7469 sessions recorded in the first week after user registration, for a total of 2771 users. Following this peak, the number of sessions decreased rapidly over the next several weeks. It declined to 2295 by the end of the first month (a reduction of 69.3%), to 1678 by the end of the second month, and to 1448 (a reduction of 80.6%) by the end of the third. By around week 10, the decline began to stabilize, though a gradual downward trend persisted. By the 30th week, the number of sessions plateaued at a much lower level, approximately below 100 sessions in a week.

XSL•FO
RenderX

Figure 1. Total number of sessions in time (per training week) for all users (N=2771).



The time from user registration to their first training session ranged from 0 to 319.84 days, with a median delay of 7.98 days and a mean of 30 days. Most users initiated training within the first few days after registering, and the frequency declined steeply after 10 days. Delays beyond 50 days were rare, and only a very small proportion of users waited more than 100 days. A small subgroup of 31 users showed exceptionally long delays between 200 and 350 days.

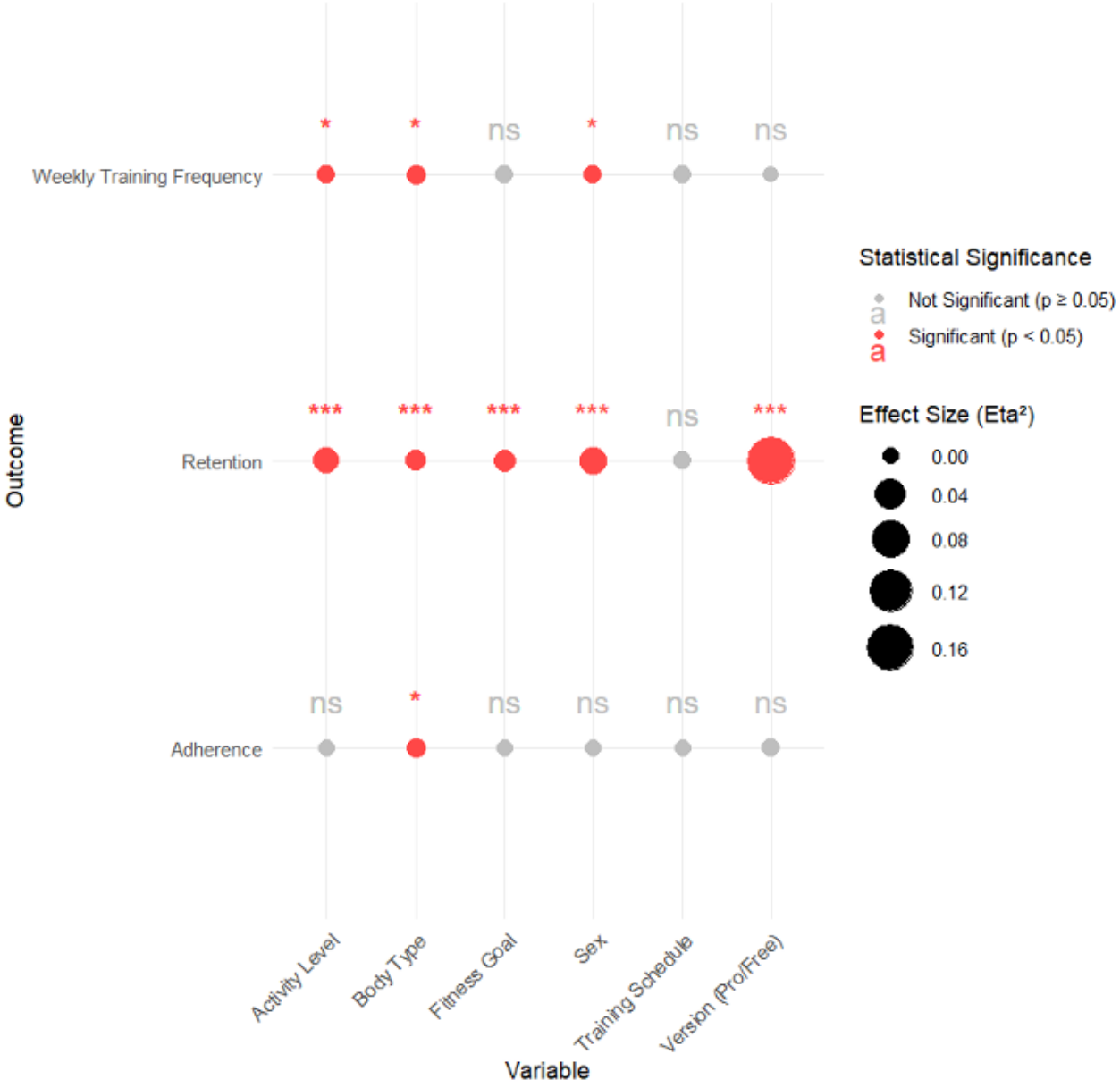
Differences Between Training Outcome Variables

Retention Results

All categorical variables except for training schedule showed statistically significant differences to retention (Table 3). For the sex variable, retention values were statistically significantly (P<.001) higher in the “male” group when compared to the “female” group, though the effect size was small (Dunn r=0.14). For the Pro version variable, indicative of whether the user was or was not subscribed for service at the time of data download, retention values were statistically significant (P<.001) with a moderate-to-large effect size (Dunn r=0.41), higher in the “yes” group than in the “no” group.” In regard to activity level, retention values were statistically significant (P<.001, Dunn r=0.11; and P<.001, Dunn r=0.07, respectively) and were higher for the “active” and “very active” groups than for the “inactive” group. No statistically significant (P>.05, Dunn r=0.03) differences were found in retention values between the “active” and “very active” groups. Refer to Figure 2 for further details on the Kruskal-Wallis results for the outcome variables.

XSL•FO
RenderX

Figure 2. Summary of Kruskal-Wallis findings for each outcome variable.



XSL•FO
RenderX

Table 3. Retention (in days), relative to categorical variables levels according to the multivariate analysis of variance (n = 2771).

Variable and group	Mean (SD)	H[a]	η^2 [b]	*P* [c]
Sex		56.151	0.019	<.001[d]
Female	109.43 (75.51)			
Male	134.99 (73.86)			
Pro version		459.965	0.163	<.001[d]
No	80.73 (70.11)			
Yes	153.53 (64.47)			
Activity level		47.414	0.016	<.001[d]
Inactive	109.74 (74.84)			
Active	130.69 (75.65)			
Very active	139.72 (69.57)			
Body type		14.303	0.004	.001[e]
Mid	132.37 (74.16)			
Thin	120.16 (76.27)			
Strong	117.44 (75.63)			
Fitness goal		21.982	0.007	<.001[d]
Antiaging	128.45 (70.23)			
Gain	132.15 (76.62)			
Lose	115.8 (75.18)			
Training schedule		3.613	0.000	.16
Afternoon	125.51 (74.23)			
Morning	128.49 (76.54)			
Night	119.07 (77.71)			

[a]H: Kruskal-Wallis H value.

[b]η2: eta squared.

[c]*P:* Kruskal-Wallis significance.

[d]*P*<.001.

[e]*P*<.01**.**

For the body type variable, the “mid” group retention values were statistically significantly higher (*P*<.002; Dunn *r* 0.07) than those in the “thin” group. No statistically significant differences were found in retention values between the “mid” and “strong” groups, nor between the “strong” and “thin” groups (*P*>.05; Dunn *r*=0.04 and *P*>.05; Dunn *r*=0.00, respectively). In the fitness goal variable, retention values were significantly lower for the “lose weight” group compared to the “antiaging” and “gain muscle” groups (*P*<.02; Dunn *r*=0.05 and *P*<.001; Dunn *r*=0.09, respectively) (Table 4).

XSL•FO
RenderX

Table 4. Post-hoc comparison of retention by categorical variable levels.

Variable	Comparison	U[a]	*P* value
Sex	• Female-male	• –7.493	• <.001b
Pro version	• No-yes	• –21.447	• <.001b
Activity level	• Active-inactive • Active-very active • Inactive-very active	• 6.067 • –1.796 • –5.641	• <.001b • .22 • <001b
Body type	• Mid-strong • Mid-thin • Strong-thin	• 2.034 • 3.579 • –0.234	• .13 • <.001c • >.99
Fitness goal	• Antiaging-gain • Antiaging-lose • Gain-lose	• –0.607 • 2.858 • 4.597	• >.99 • .01d • <.001b
Training schedule	• Afternoon-morning • Afternoon-night • Morning-night	• –1.062 • 1.266 • 1.850	• .87 • .62 • .19

[a]*U:* Standardized test statistic.

[b]***P*<.001**.

[c]*P*<.01.

[d]*P*<.05.

Regarding explanatory numerical variables, all of them showed statistically significant correlations with retention. Age had a moderate positive correlation (r=0.21; P<.001) with retention, while subjective body fat and BMI showed a low negative correlation (r=–0.13; P<.001; r=–0.06; P<.01, respectively). Training difficulty had a moderate, positive correlation (r=0.24; P<.001), and enjoyment had a low, positive correlation (r=0.11; P<.001).

Of all motivation dimensions, intrinsic motivation had the highest positive correlation (r=0.19; P<.01) with retention. Identified extrinsic motivation had a small, statistically significant, positive correlation (r=0.12; P<.05). Introjected extrinsic motivation had a much lower, positive, and nonstatistically significant correlation (r=0.07; P>.05; Table 5).

XSL•FO
RenderX

Table 5. Results of the correlation tests for motivation variables (N=289).

Variable	r [a]	*P* value
Retention		
INTRINS[b]	0.19	.01[c]
IDENT, extr.[d]	0.12	<.05[c]
INTROJ, extr.[e]	0.07	.22
Frequency, weekly		
INTRINS	0.001	>.99
IDENT, extr.	0.008	>.99
INTROJ, extr.	0.168	.01[c]
Adherence		
INTRINS	–0.002	>.99
IDENT, ext.	0.021	>.99
INTROJ, ext.	0.138	.06

[a]*r:* Spearman correlation coefficient.

[b]INTRINS: Intrinsic motivation.

[c]*P*<.05.

[d]IDENT, extr.: Identified extrinsic motivation.

[e]INTROJ, extr.: Introjected extrinsic motivation.

Average Weekly Frequency Results

Weekly frequency was found to be statistically significantly associated with sex, activity level, and body type (*P*<.05) and not significantly related to the Pro version, fitness goal, or training schedule (Table 6).

XSL•FO
RenderX

Table 6. Average weekly training frequency relative to categorical variables levels. Multivariate analysis of variance (N=2771).

Variable	Mean (SD)	H[a]	η^2 [b]	*P* [c]
Sex		3.950	0.001	.04[d]
Female	1.82 (1.35)			
Male	1.9 (1.29)			
Pro version		0.122	-0.001	.78
No	1.94 (1.45)			
Yes	1.83 (1.21)			
Activity level		6.199	0.001	.04[d]
Inactive	1.83 (1.34)			
Active	1.85 (1.29)			
Very active	2.05 (1.31)			
Body type		8.324	0.002	.02[d]
Mid	1.77 (1.23)			
Thin	1.96 (1.37)			
Strong	2.02 (1.39)			
Fitness goal		4.363	<0.001	.11
Antiaging	1.79 (1.28)			
Gain	1.91 (1.28)			
Lose	1.85 (1.36)			
Training schedule		4.519	0.001	.10
Afternoon	1.82 (1.30)			
Morning	1.93 (1.32)			
Night	1.94 (1.30)			

[a]H: Kruskal-Wallis H value.

[b]$\eta 2$: eta squared.

[c]*P*: Kruskal-Wallis significance.

[d]$P<.05$.

For the sex variable, figures in the "male" group were statistically significant ($P<.005$; Dunn $r=0.04$) and higher than those for the "female" group. In the activity level variable, the "very active" group values were significantly higher ($P<.05$; Dunn $r=0.05$) than those in the "inactive" group. Neither the "active" versus "inactive" groups nor the "active" versus "very active" groups showed any statistically significant differences in weekly training frequency values. For the body type variable, values in the "thin" group were statistically significantly ($P<.04$; Dunn $r=0.05$) higher than those of the "mid" group. No statistically significant differences were found between the "mid" and "strong" groups, nor between the "strong" and "thin" groups, for weekly training frequency values. For more detail on multivariate analyses of variance results for weekly training frequency, please refer to Table 7.

XSL•FO
RenderX

Table 7. Post-hoc comparison of average weekly training frequency by categorical variable levels.

Variable	Comparison	U [a]	*P* value
Sex	• Female-male	• –1.988	• .04b
Pro version	• No-yes	• 0.349	• .73
Activity level	• Active-inactive • Active-very active • Inactive-very active	• 0.599 • –2.218 • –2.427	• >.99 • .08 • .04b
Body type	• Mid-strong • Mid-thin • Strong-thin	• –1.989 • –2.501 • 0.727	• .14 • .04b • >.99
Fitness goal	• Antiaging-gain • Antiaging-lose • Gain-lose	• –1.716 • –0.372 • 1.679	• .26 • >.99 • .28
Training schedule	• Afternoon-morning • Afternoon-night • Morning-night	• –1.824 • –1.508 • –0.379	• .20 • .40 • >.99

[a]U = standardized test statistic.

[b]P<.05.

After corrections for multiple comparisons, none of the explanatory numerical variables reflected statistically significant correlations with weekly training frequency.

Regarding types of motivation, only introjected extrinsic motivation correlated significantly with training frequency *(*r=0.17; P<.05), after corrections for multiple comparisons, with a moderate-low positive correlation. Intrinsic and identified extrinsic motivations showed close to no correlation with frequency (r=0.00 and P>.05 for both) (Table 5).

Adherence Results

Adherence was found to be associated with one of the categorical variables (Table 8). These differences were found in body type, which presented a highly statistically significant (P<.005; Dunn r=–0.06) difference between the groups “mid” and “thin”. In particular, the values for the “mid” group were statistically significantly lower than those for the “thin” group. No statistically significant differences were found between groups “mid” and “strong,” nor between “strong” and “thin,” in adherence to the training program.

XSL•FO
RenderX

Table 8. Adherence, relative to categorical variables levels. Multivariate analysis of variance (N=2771).

Variable and group	Mean (SD)	H[a]	η^2 [b]	*P* [c]
Sex		1.479	0.000	.22
Female	54.02 (38.07)			
Male	54.35 (34.88)			
Pro version		3.260	0.000	.07
No	57.35 (39.85)			
Yes	52.35 (33.37)			
Activity level		2.394	0.000	.30
Inactive	53.76 (38.29)			
Active	53.91 (35.09)			
Very active	57.01 (35.39)			
Body type		7.322	0.002	.03[d]
Mid	52.1 (35.28)			
Thin	56.96 (37.32)			
Strong	52.01 (31.57)			
Fitness goal		2.960	0.000	.23
Antiaging	52.52 (34.7)			
Gain	55.34 (35.53)			
Lose	53.48 (37.37)			
Training schedule		2.843	0.000	.24
Afternoon	53.12 (35.66)			
Morning	56.05 (36.95)			
Night	53.61 (34.38)			

[a]*H*: Kruskal Wallis H value.

[b]$\eta 2$: eta squared.

[c]*P:* Kruskal-Wallis significance.

[d]$P<.05$.

No statistically significant differences were found between adherence and any of the groups in any of the remaining categorical variables (Table 9).

XSL•FO
RenderX

Table 9. Post-hoc comparison of adherence by categorical variables levels.

Variable	Comparison	U [a]	P value
Sex	• Female-male	• –1.216	• .22
Pro version	• No-yes	• 1.806	• .07
Activity Level	• Active-inactive • Active-very active • Inactive-very active	• 0.832 • –1.082 • –1.539	• >.99 • .84 • .37
Body Type	• Mid-strong • Mid-thin • Strong-thin	• –0.517 • –2.703 • –0.836	• >.99 • .02b • >.99
Fitness Goal	• Antiaging-gain • Antiaging-lose • Gain-lose	• –1.282 • –0.096 • 1.499	• .60 • >.99 • .40
Training schedule	• Afternoon - Morning • Afternoon - Night • Morning - Night	• –1.680 • –0.558 • 0.453	• .28 • >.99 • >.99

[a]U: Mann-Whitney standardized test statistic.

[b]P<.05.

For explanatory numerical variables, none reflected statistically significant correlations with adherence. Similarly, no types of motivation showed statistically significant correlations with adherence. Introjected extrinsic motivation rendered the highest (r= 0.14; P>.05) positive correlation, while intrinsic and identified extrinsic motivation showed little to no correlation (r=–0.00; P>.05, and r=0.02; P>.05, respectively; Table 5).

Summary of All Findings From the Inferential Analysis

A summary of all findings from the inferential analysis is presented in Table 10.

Table 10. Summary of statistically significant findings of the post-hoc inferential analyses.

Variable	Retention	Weekly training frequency	Adherence
Sex	Male>female	Male>female	—[a]
Pro version	Yes>no	—	—
Activity level	Active>inactive Very active>inactive	Very active>inactive	—
Fitness goal	Antiaging>lose Gain>lose	—	—
Body type	Mid>thin	Thin>mid	Thin>mid
Training schedule	—	—	—
Age	Positive correlation	—	—
BMI	Negative correlation	—	—
Subjective body fat	Negative correlation	—	—
Difficulty	Positive correlation	—	—
Enjoyment	Positive correlation	—	—
Intrinsic motivation	Positive correlation	—	—
Identified extrinsic motivation	Positive correlation	—	—
Introjected extrinsic motivation	—	—	—

[a]Not applicable.

XSL•FO
RenderX

Discussion

Summary of Key Findings

This study examined training behaviors among users of the MH fitness app and identified key factors associated with training behavior. Variables including adherence and retention were evaluated, with the latter having shown greater relevance in users long-term maintenance of training behavior. The main findings pointed at paid subscription and intrinsic motivation as being the most determinant factors to user retention. Other variables that correlated with retention included sex, body type, BMI, and fitness goal. In contrast, adherence was only linked to body type, while training frequency varied slightly by sex, activity level, motivation, and body type.

This piece of research involved 2771 individuals and is possibly one of the largest existing cohort studies of fitness app users to date. Previous large cohorts include the Konstanz Life Study with 1236 users of either fitness or nutrition apps [21]. Some systematic reviews have covered samples of 3555 participants from a total of 22 interventions (n=833 in the largest single study) [8] or 1622 total participants from 6 different studies [7]. Our work possibly also covers the longest time duration (18 months). Previous research has been 2-24 weeks [7], up to 5 months [22], or even 6 months in some cases [8].

Our sample figures fall within the "expected" ranges for a fitness app that offers high-intensity training, delivered electronically. Results are also in line with the systematic review by Stecher et al [8], which included participants between 10.6 and 61.5 years of age and found a mean of 39.6 (SD 6.5) years. Participants in this study presented some features worth noting, which were probably specific to our sample population. The majority (75.6%) of them were previously "active" or "very active." This was most likely due to the fact that all data registries were obtained from an app update (MH version 2.0) which, naturally, received many of its users from the previous version. This could also partially explain why 62% of our users were on the Pro version (paid subscription). MH always offered a free training program upon first registration, so the newest users would be expected to be on a free deal, while more experienced users would naturally progress to payment modes.

The studied sample primarily pursued "muscle gain" or "weight loss" fitness goals. A remarkably small (21.4%) percentage trained for "antiaging" purposes. We initially interpreted this finding as a sign that individuals were focusing mainly on "appearance," but this would have to be further investigated, as muscle gain [23,24], as well as weight loss [25], are also known markers of improved health [25] and consequently better aging.

Attrition Rates and Perceived Difficulty

It is well-established that attrition rates in mobile apps are extremely high. Meyerowitz-Katz et al [26], in their 2020 meta-analysis, stated that up to 98% of people only use apps for a short period of time. Our results fully align with this marked tendency, as we appreciated a remarkable drop in the number of training sessions within the first few weeks of enrollment. There was an observable decline of 69.3% by the end of the first month, a reduction of 77.5% by the end of month 2, and an 80.6% decline by the end of month 3. These figures strike even harder if we assume that many enrollments allegedly came from MH users who were transitioning from the old to the new version of the app. Participants in our study preferred "afternoon" (12:31-20:00 hours) training sessions and declared mean rates of session "difficulty" of 5.56, over a total of 10 points. The "difficulty" variable and its results need to be interpreted with caution. In our study, "difficulty" was an equivalent of perceived exertion, and it aimed to be indicative of how hard the session had felt to the user. However, this data were inquired once the user had not only finished the training but also finished the cool-down phase, and this could have led to respondents underrating the perceived exertion derived from the main block of training. Contrary to our expectations, difficulty in our sample showed a strong positive correlation to retention, which could be interpreted as a sign that challenge fosters engagement. Indeed, there is previous evidence that complex, vigorous, or hybrid activities correlate with intrinsic motivation [27], which commonly underlines activity retention. Regarding constructs of adherence, this finding could also reflect a self-selection bias, where more committed users are more likely to opt for challenging sessions, thus reinforcing their engagement over time.

Factors Influencing Training Frequency

Frequency of training seemed not to be related to factors such as age, BMI, declared enjoyment or perceived difficulty, subscription vs nonsubscription, declared fitness goal, or preferred training schedule. However, statistically significant differences were observed based on sex, previous activity level, motivation, and body type. Frequency of training was greater in the introjected motivation group, in males, in the very active vs inactive, and in the thin vs mid groups. One could argue that the controlled and external regulation of introjected motivation could explain the increased frequency observed in this group. This would partially align with previous research that points to the primacy of extrinsic motivation in exercise contexts [28]. The fact that introjected motivation seems to encourage higher training frequencies but no longer retention or higher adherence might be indicative of an enthusiasm that is not sustained over time. As to the user's previous activity level, while it seems logical that highly active individuals would train more often, this could be influenced by their prior engagement with the MH app or other forms of PA. If they were former MH users, their higher frequency could indicate loyalty, whereas if their activity stemmed from external sources, it is noteworthy that they also engaged frequently with the app. In contrast, inactive individuals may have felt overwhelmed by structured training. Previous research highlights differences in how beginners perceive social comparison and networking features in fitness apps, as well as how exercise proficiency affects adherence [29]. Additionally, attitudes toward PA significantly impact behavior, with Feng et al [30] showing that greater activity levels correspond to deeper integration and sustained engagement. Another possible explanation is that very active users may use more app features, enhancing their overall experience and leading to higher engagement [30]. Our results should, however, be interpreted with caution, since despite statistical significance, effect sizes

XSL•FO
RenderX

were small to very small, which is indicative of them having limited practical implications.

Reflections on Adherence to mHealth Training

In our study, adherence did not correlate with age, sex, previous level of activity, declared fitness goal, being on a free plan versus subscription mode, training schedule, perceived difficulty, or enjoyment in sessions.

Only one statistically significant difference was found for adherence, and it was for the "thin" group, which showed higher adherence than the "mid" group. Both frequency and adherence in this study were correlated with "thin" body type, but in both cases the effect size was small, so the association may not imply high practical impact. Notably, no motivation type proved to be more relevant for adherence, in spite of several authors having pointed to the more autonomous regulations of motivation leading to increased adherence and persistence [31]. Recent evidence confirms that maintaining physical activity remains challenging for healthy adults, with persistent individual-level barriers (ie, lack of motivation, attitudes, and concerns about physical changes) [32]. Adherence results in our study ranged from 1.2% to 166.7%, which was an impactfully wide range. It is important to note that intensity and duration data were not consistently available across users, which limited our ability to construct the adherence measure. Adherence in this study was based only on training frequency, a limitation that highlights the need for more standardized and comprehensive adherence metrics in future app-based exercise research. These results brought us to the following insights. Adherence is rather a measure of precise forecasting, as it basically depends on the ability to foresee future behavior. In that case, several personal characteristics may come into play, which have not been assessed in this study, such as the concept of self-efficacy, ambition, the ability to plan in advance, or the ability to pursue goals. Similarly, in healthy adults, psychological factors such as self-efficacy, enjoyment, and planning were significant predictors of long-term adherence to PA, emphasizing the relevance of individual motivational and behavioral traits in sustained engagement [33]. In addition, a study identified lack of time, motivation, and fatigue as frequent barriers to PA in healthy young adults, while enjoyment and social support emerged as consistent facilitators [34]. We found the lowest adherence rates for those who trained 6 times per week, while the highest adherence values were for those who trained twice per week. Based on our results, individuals with lower frequency expectations managed better to fulfill their target plan and were, consequently, more adherent. Again, we see a disadvantage in how different researchers seem to measure and define adherence, in addition to the fact that electronically delivered interventions often lack a detailed reporting of it [35]. We note that in our study, adherence was based on training frequency (% of targeted versus actual), which naturally correlates both variables. In contrast, retention and adherence operate on different parameters, especially when the exercise program is nonprescribed, lacks external obligation, and has no set duration.

Retention as a Key Variable, Distinct From Adherence

In this study, retention correlated significantly with most study variables. There was higher retention in the male group, in subscribers, in "active" and "very active," in "mid" body types versus "thin," and also higher retention when the fitness goal was "antiaging" or "gain muscle" versus "lose weight." "Pro version" users exhibited higher retention, aligning with previous research linking price to commitment [29,36,37], suggesting that subscription may indicate greater interest. The effect size for our finding was moderate-to-large, which points at subscription possibly being the most determinant factor in long-term training behavior. All other correlations had small to very small effect sizes. "Thin" body types, which had shown correlation with frequency and adherence, did not display higher retention, potentially reflecting an initial enthusiasm that wanes over time. The finding that "antiaging" goals led to higher retention than "lose weight" aligns with theories suggesting that health-oriented goals promote sustained engagement. However, the fact that "gain muscle" goals also outperformed "lose weight" in retention suggests that aesthetic-driven objectives may still play a role in long-term engagement, challenging this interpretation.

The study found that only intrinsic motivation had a statistically significant positive correlation with retention, while no such correlation was observed between intrinsic motivation and adherence. The distinction between retention and adherence is emphasized, as these concepts are considered distinct. In line with previous evidence [28,38-40], our study reflects that intrinsic and identified autonomously regulated motivations are the strongest correlated with retention. There is, however, previous evidence that points at extrinsic regulations of motivation as possibly the most important ones for exercise contexts [28,41]. We agree with Wilson's statement that future research with larger sample sizes is recommended, considering potential variations in extrinsic motivation types [28] and a revision of the commonly accepted theory that intrinsic motivation is the most desirable to engage in and sustain exercise activities.

To translate our findings into practical applications, we suggest that fitness app developers consider tailoring features to specific user subgroups (ie, providing targeted support or content adaptations for users with a "mid" body type). Moreover, including motivational aspects that support intrinsic regulation, such as goal-tracking tools, personalized feedback, and autonomy-enhancing design, may further increase retention and adherence.

Strengths and Limitations

The present study focused on analyzing user training behavior by means of a cross-sectional study conducted on 2771 MH app users over a period of 18 months. To the authors' knowledge, the largest study previously available was a cohort study conducted under the Konstanz Life Study, which followed a total of 1236 users of either fitness or nutrition apps [21]. Other revisions involved larger samples, such as that by Stecher et al [8], (with 3555 participants across 22 interventions) or He et al [7] (with 1622 participants from 6 studies). Previous studies followed participants for periods of 2 to 24 months [7,8,22].

XSL•FO
RenderX

Based on this evidence, our study could be the largest of its kind in sample size and follow-up period so far. Nonetheless, we acknowledge that broader meta-analyses may include larger cumulative samples and aggregated durations across multiple interventions and apps [26].

The use of real-world app data, combined with motivational surveys, provides valuable insights into user behavior, retention, and adherence patterns. Additionally, the study uses robust statistical analyses, including nonparametric tests and multiple correction methods, ensuring the reliability of the findings.

However, the research also has limitations. The cross-sectional design prevents establishing causal relationships between motivation, training behavior, and adherence. The dataset is limited to users of a single fitness app (MH), potentially restricting generalizability to other platforms with different features or user demographics. We obtained informed consent from 5858 users. Of those registries, 2771 were complete and eligible for analysis. This loss should be acknowledged as the fact that we only managed to merge motivation and training data for a total of 289 participants, which limits the statistical power of our motivation-related analyses. Finally, adherence was measured in terms of training frequency, which may not fully capture engagement in app-based fitness programs, highlighting the need for more nuanced adherence metrics in future research.

Conclusions

This study provides crucial insights into the exercise behavior and retention patterns of MH app users, highlighting key factors that influence user engagement. New insights are shared in regard to how motivation relates to training behavior with fitness apps. Clear differentiations are presented between adherence and retention, as conceptualized by the study authors. Fitness apps are a promising tool toward more active lifestyles, but we are yet lacking a sound understanding of related human behavior. Strategies such as gamification, goal-setting, or prompting are available to app developers to increase user engagement. However, longitudinal studies and mixed methods approaches are needed both to study causality and explore qualitative drivers to trainee behavior.

Acknowledgments

We would like to thank all users of MH app who consent to share their anonymized data for research purposes. All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

Funding

This study was supported with research funds provided by Aristos Campus Mundus projects (ACM2022_25 and ACM2023_19) and Fundació Naccari-Ravà 2022. The funding bodies had no role in study design, data collection, data analysis, or manuscript preparation.

Data Availability

The datasets analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: BG-Z, JMA-M, JM, JJ-R
Data curation: OB-J
Formal analysis: AF-V, RP, JMA-M
Funding acquisition: JMA-M, BG-Z, JM
Investigation: AF-V
Methodology: AF-V, RP, BG-Z, MG-B
Project administration: JM
Supervision: JM, BG-Z, JJ-R
Validation: AF-V
Visualization: JM, JJ-R
Writing – original draft: AF-V, JM
Writing – review & editing: AF-V, RP, JMA-M, OB-J, MG-B, JM, JJ-R

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

Complete description of the Mammoth Hunters dataset, including all variable categories and the full motivation survey used in the study.
[DOCX File , 20 KB - mhealth_v14i1e72201_app1.docx]

References

1. WHO guidelines on physical activity and sedentary behaviour: at a glance Internet. World Health Organization. Geneva PP - Geneva: World Health Organization URL: https://iris.who.int/handle/10665/337001 [accessed 2025-12-04]
2. Rodgers WM, Hall CR, Duncan LR, Pearson E, Milne M. Becoming a regular exerciser: examining change in behavioural regulations among exercise initiates. Psychology of Sport and Exercise 2010;11(5):378-386 [FREE Full text] [doi: 10.1016/j.psychsport.2010.04.007]
3. Bailey DL, Holden MA, Foster NE, Quicke JG, Haywood KL, Bishop A. Defining adherence to therapeutic exercise for musculoskeletal pain: a systematic review. Br J Sports Med 2020;54(6):326-331 [FREE Full text] [doi: 10.1136/bjsports-2017-098742] [Medline: 29875278]
4. Number of US Fitness Health App Users Internet. Statista. URL: https://www.statista.com/statistics/1154994/number-us-fitness-health-app-users/ [accessed 2025-12-04]
5. Chatterjee A, Prinz A, Gerdes M, Martinez S. Digital interventions on healthy lifestyle management: systematic review. J Med Internet Res 2021;23(11):e26931 [FREE Full text] [doi: 10.2196/26931] [Medline: 34787575]
6. Yang Y, Boulton E, Todd C. Measurement of adherence to mHealth physical activity interventions and exploration of the factors that affect the adherence: scoping review and proposed framework. J Med Internet Res 2022;24(6):e30817 [FREE Full text] [doi: 10.2196/30817] [Medline: 35675111]
7. He Z, Hassan MA, Saiz-González P, Ryu S, Wang R, Gao Z. Smartphone app-based interventions on physical activity behaviors and psychological correlates in healthy young adults: a systematic review. PLoS One 2024;19(4):e0301088 [FREE Full text] [doi: 10.1371/journal.pone.0301088] [Medline: 38578729]
8. Stecher C, Pfisterer B, Harden SM, Epstein D, Hirschmann JM, Wunsch K, et al. Assessing the pragmatic nature of mobile health interventions promoting physical activity: systematic review and meta-analysis. JMIR Mhealth Uhealth 2023;11:e43162 [FREE Full text] [doi: 10.2196/43162] [Medline: 37140972]
9. Ozdamli F, Milrich F. Positive and negative impacts of gamification on the fitness industry. Eur J Investig Health Psychol Educ 2023;13(8):1411-1422 [FREE Full text] [doi: 10.3390/ejihpe13080103] [Medline: 37623300]
10. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
11. Lu T, Lin Q, Yu B, Hu J. A systematic review of strategies in digital technologies for motivating adherence to chronic illness self-care. NPJ Health Syst 2025;2(1). [doi: 10.1038/s44401-025-00017-4]
12. Fang P, Shi S, Menhas R, Laar RA, Saeed MM. Demographic characteristics and digital platforms for physical activity among the Chinese residents during the COVID-19 pandemic: a mediating analysis. JMDH 2022;15:515-529. [doi: 10.2147/jmdh.s354984]
13. Al-Shamaileh O, Sutcliffe A. Why people choose apps: an evaluation of the ecology and user experience of mobile applications. Int J Hum Comput Stud 2023;170:102965. [doi: 10.1016/j.ijhcs.2022.102965]
14. Peng X, Menhas R, Dai J, Younas M. The COVID-19 Pandemic and overall wellbeing: mediating role of virtual reality fitness for physical-psychological health and physical activity. Psychol Res Behav Manag 2022;15:1741-1756. [doi: 10.2147/prbm.s369020]
15. Liu R, Menhas R, Dai J, Saqib ZA, Peng X. Fitness apps, live streaming workout classes, and virtual reality fitness for physical activity during the COVID-19 lockdown: an empirical study. Front Public Health 2022;10:852311 [FREE Full text] [doi: 10.3389/fpubh.2022.852311] [Medline: 35812515]
16. Yang J, Menhas R, Dai J, Younas T, Anwar U, Iqbal W, et al. Virtual reality fitness (VRF) for behavior management during the COVID-19 pandemic: a mediation analysis approach. Psychol Res Behav Mana 2022;Volume 15:171-182. [doi: 10.2147/prbm.s350666]
17. Luo C. An exploration of the impact of fitness apps on individual exercise behaviour. Rev Psicol del Deport 2024;33:385-393.
18. Boucher EM, Raiker JS. Engagement and retention in digital mental health interventions: a narrative review. BMC Digit Heal 2024;2(1):52. [doi: 10.1186/s44247-024-00105-9]
19. Deci EL, Ryan RM. Self-determination theory: a macrotheory of human motivation, development, and health. Can Psychol / Psychol Can 2008;49(3):182-185. [doi: 10.1037/a0012801]
20. Ryan RM, Deci EL. Intrinsic and extrinsic motivation from a self-determination theory perspective: definitions, theory, practices, and future directions. Contemp Educ Psychol 2020;61:101860. [doi: 10.1016/j.cedpsych.2020.101860]
21. König LM, Sproesser G, Schupp HT, Renner B. Describing the process of adopting nutrition and fitness apps: behavior stage model approach. JMIR Mhealth Uhealth 2018;6(3):e55 [FREE Full text] [doi: 10.2196/mhealth.8261] [Medline: 29535078]
22. Herrmann LK, Kim J. The fitness of apps: a theory-based examination of mobile fitness app usage over 5 months. Mhealth 2017;3:2 [FREE Full text] [doi: 10.21037/mhealth.2017.01.03] [Medline: 28293619]

XSL•FO
RenderX

23. Wang DXM, Yao J, Zirek Y, Reijnierse EM, Maier AB. Muscle mass, strength, and physical performance predicting activities of daily living: a meta-analysis. J Cachexia Sarcopenia Muscle 2020;11(1):3-25 [FREE Full text] [doi: 10.1002/jcsm.12502] [Medline: 31788969]
24. Westcott WL. Resistance training is medicine: effects of strength training on health. Curr Sports Med Rep 2012;11(4):209-216. [doi: 10.1249/jsr.0b013e31825dabb8]
25. Jayedi A, Khan TA, Aune D, Emadi A, Shab-Bidar S. Body fat and risk of all-cause mortality: a systematic review and dose-response meta-analysis of prospective cohort studies. Int J Obes (Lond) 2022;46(9):1573-1581. [doi: 10.1038/s41366-022-01165-5] [Medline: 35717418]
26. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
27. Hsu RMCS, Cardoso FL, Varella MAC, Pires EM, Valentova JV. Comparing different typologies of physical activities with a focus on motivation. Front Psychol 2022;13:790490 [FREE Full text] [doi: 10.3389/fpsyg.2022.790490] [Medline: 35645925]
28. Wilson PM, Rodgers WM, Fraser SN, Murray TC. Relationships between exercise regulations and motivational consequences in university students. Res Q Exerc Sport 2004;75(1):81-91. [doi: 10.1080/02701367.2004.10609136] [Medline: 15532364]
29. Hu J, He W, Zhang J, Song J. Examining the impacts of fitness app features on user well-being. Inf Manag 2023;60(5):103796 [FREE Full text] [doi: 10.1016/j.im.2023.103796]
30. Feng B. Exploration on how integration degree into sport APPs affect Users? Attitude and behavioural habit towards physical exercise. Rev Psicol del Deport 2024;33:14-23.
31. Rodrigues F, Bento T, Cid L, Pereira Neiva H, Teixeira D, Moutão J, et al. Can interpersonal behavior influence the persistence and adherence to physical exercise practice in adults? A systematic review. Front Psychol 2018;9:2141 [FREE Full text] [doi: 10.3389/fpsyg.2018.02141] [Medline: 30459690]
32. Gabay M, Oravitan M. The factors affecting adherence to physical activity in fitness facility settings: a narrative review. Timisoara Phys Educ Rehabil J 2022;15:46-61. [doi: 10.2478/tperj-2022-0013]
33. Watson-Mackie K, Arundell L, Lander N, McKay FH, Jerebine A, Venetsanou F, et al. Technology-supported physical activity and its potential as a tool to promote young women's physical activity and physical literacy: systematic review. J Med Internet Res 2024;26:e52302 [FREE Full text] [doi: 10.2196/52302] [Medline: 39423006]
34. Duffey K, Barbosa A, Whiting S, Mendes R, Yordi Aguirre I, Tcymbal A, et al. Barriers and facilitators of physical activity participation in adolescent girls: a systematic review of systematic reviews. Front Public Health 2021;9:743935 [FREE Full text] [doi: 10.3389/fpubh.2021.743935] [Medline: 34722450]
35. Valenzuela T, Okubo Y, Woodbury A, Lord SR, Delbaere K. Adherence to technology-based exercise programs in older adults: a systematic review. J Geriatr Phys Ther 2018;41:49-61. [doi: 10.1519/jpt.0000000000000095]
36. Oyibo K. Investigating the key persuasive features for fitness app design and extending the persuasive system design model: a qualitative approach. Proc Int Symp Hum Factors Ergon Healthc 2021;10(1):47-53 [FREE Full text] [doi: 10.1177/2327857921101022]
37. James TL, Bélanger F, Lowry PB. The mediating role of fitness technology enablement of psychological need satisfaction and frustration on the relationship between goals for fitness technology use and use outcomes. J Assoc Inf Syst Internet 2022;23:913-965 [FREE Full text] [doi: 10.17705/1jais.00745]
38. Kuroda Y, Sato Y, Ishizaka Y, Yamakado M, Yamaguchi N. Exercise motivation, self-efficacy, and enjoyment as indicators of adult exercise behavior among the transtheoretical model stages. Glob Health Promot 2012;19(1):14-22. [doi: 10.1177/1757975911423073] [Medline: 24801311]
39. Ryan RM, Deci EL. Overview of Self-Determination theory: An Organismic-Dialectical Perspective. Handb self-Determination Res. Rochester, NY, US: University of Rochester Press; 2002:3-33.
40. Teixeira PJ, Carraça EV, Markland D, Silva MN, Ryan RM. Exercise, physical activity, and self-determination theory: a systematic review. Int J Behav Nutr Phys Act 2012;9:78 [FREE Full text] [doi: 10.1186/1479-5868-9-78] [Medline: 22726453]
41. Ryan RM. Psychological needs and the facilitation of integrative processes. J Pers 1995;63(3):397-427. [doi: 10.1111/j.1467-6494.1995.tb00501.x] [Medline: 7562360]

Abbreviations

MH: Mammoth Hunters
mHealth: mobile health
PA: physical activity
SDT: self-determination theory
WHO: World Health Organization

XSL•FO
RenderX

Edited by L Buis; submitted 13.Feb.2025; peer-reviewed by RN Al-Jabali, R Menhas; comments to author 11.Jun.2025; revised version received 23.Jul.2025; accepted 06.Nov.2025; published 08.Jan.2026.

Please cite as:
Fuente-Vidal A, Prat R, Arribas-Marin JM, Bastidas-Jossa O, Guerra-Balic M, Garcia-Zapirain B, Montane J, Jerez-Roig J
Analysis of Training Behavior in Users of a Fitness App: Cross-Sectional Study
JMIR Mhealth Uhealth 2026;14:e72201
URL: https://mhealth.jmir.org/2026/1/e72201
doi:10.2196/72201
PMID:

©Andrea Fuente-Vidal, Roger Prat, Juan Manuel Arribas-Marin, Oscar Bastidas-Jossa, Myriam Guerra-Balic, Begonya Garcia-Zapirain, Joel Montane, Javier Jerez-Roig. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 08.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of a Hybrid Community-Based Heart-Healthy Lifestyle Intervention: Three-Arm Randomized Controlled Trial Integrating mHealth and Motivational Interviewing

Jina Choo[1,2], DrPH, PhD; Yura Shin[1,2], MSN; Songwhi Noh[1,2], MSN; Juneyoung Lee[3], PhD

[1]College of Nursing, Korea University, Seoul, Republic of Korea

[2]Transdisciplinary Major in Learning Health Systems, Graduate School, Korea University, Seoul, Republic of Korea

[3]Department of Biostatistics, College of Medicine, Korea University, Seoul, Seoul, Republic of Korea

Corresponding Author:
Jina Choo, DrPH, PhD
College of Nursing
Korea University
145, Anam-ro, Seongbuk-gu
Seoul, 02841
Republic of Korea
Phone: 82 232904925
Email: jinachoo@korea.ac.kr

Abstract

Background: Limited empirical evidence exists on the effectiveness of a hybrid approach to heart-healthy lifestyle interventions that integrates mobile health (mHealth) technology with face-to-face counseling. Moreover, its superiority over exclusive mHealth use in promoting heart-healthy behavioral outcomes within a community setting remains unclear.

Objective: This study aims to evaluate the effectiveness of a hybrid community-based approach to heart-healthy lifestyle intervention incorporating a mobile app and motivational interviewing among community-dwelling adults without a history of cardiovascular disease.

Methods: We conducted a 3-arm, parallel-group, randomized controlled trial with assessments at baseline and after 12 weeks. A total of 75 participants, each presenting at least 1 component of metabolic syndrome and no history of cardiovascular disease, were randomly assigned to 1 of 3 groups: hybrid (n=25), mobile (n=25), or control (n=25). Participants were recruited through an online platform. The hybrid group underwent a 12-week hybrid intervention combining a mobile app (ie, “My HeartHELP”) and face-to-face motivational interviewing led by a nursing researcher. The mobile group used only the mobile app, while the control group received written material on general heart health. The intervention was facilitated by 3 trained nursing researchers. The primary outcome was a composite score of “heart-healthy behaviors,” while secondary outcomes included scores for heart-healthy “information,” “self-efficacy,” “motivation,” and cardiovascular parameters. The trial was conducted in 2 rounds from October 2022 to May 2023. An intention-to-treat analysis was performed.

Results: Of the 75 participants, 72 (96%) completed this study. Compared with the control group, both the hybrid and mobile intervention groups demonstrated significantly greater improvements in behavioral outcomes, including composite heart-healthy behavior ($F_{2,69}$=7.25, P=.001), its theoretical predictors—heart-healthy motivation ($F_{2,69}$=8.54, P<.001) and self-efficacy for diet ($F_{2,69}$=4.87, P=.01) and exercise ($F_{2,69}$=5.48, P=.006)—as well as fasting glucose levels ($F_{2,69}$=3.90, P=.03) following the 12-week intervention. Particularly, the hybrid group—unlike the mobile group—showed significantly greater improvement in dietary behavior, a subscale of heart-healthy behavior, compared with the control group, and demonstrated significantly greater improvements in interest or enjoyment, a core subscale of intrinsic motivation, than the mobile and control groups.

Conclusions: The hybrid community-based heart-healthy lifestyle intervention—integrating a mobile app and motivational interviewing—demonstrated overall effectiveness comparable to the mobile app alone, while yielding greater improvements in dietary behavior and core intrinsic motivation. These findings highlight the potential of mHealth apps as practical, stand-alone tools to promote cardiovascular health, particularly in community settings with limited access to in-person professional support. However, incorporating motivational interviewing may further enhance internalized motivation and complex behavior changes over time. Health professionals can therefore adopt mHealth either independently or in combination with motivational interviewing.

XSL•FO
RenderX

Future studies should optimize integration strategies to enhance effectiveness and evaluate the long-term sustainability of such hybrid approaches.

Trial Registration: ISRCTN Registry ISRCTN83643383; https://www.isrctn.com/ISRCTN83643383

International Registered Report Identifier (IRRID): RR2-10.1161/circ.147.suppl_1.P147

(JMIR Mhealth Uhealth 2026;14:e76521) doi:10.2196/76521

KEYWORDS

healthy lifestyle; cardiovascular diseases; primary prevention; mobile applications; motivational interviewing; mobile phone

Introduction

Cardiovascular disease remains the leading cause of mortality worldwide [1]. To mitigate cardiovascular disease risk, the World Health Organization advocates for lifestyle modification interventions targeting smoking, poor dietary habits, physical inactivity, obesity, and alcohol consumption [2]. Evidence indicates that multiple unhealthy lifestyle behaviors significantly elevate cardiovascular disease risk [3], whereas adopting multiple healthy lifestyle behaviors can reduce all-cause mortality by up to 58%-66% [3,4]. Despite these findings, only a small proportion of individuals at risk of cardiovascular disease in the United States have achieved these recommended lifestyle changes [5]. Consequently, there is a pressing need for innovative, integrated strategies to effectively promote heart-healthy lifestyles in a comprehensive approach.

To date, mobile health (mHealth) interventions using websites, smartphones, or videoconferencing (online) have been used to improve heart health outcomes, such as cardiovascular risk parameters [6]. mHealth may help offset limited professional human resources and the time and effort required for expert-centered cardiovascular health care in the community [7]. However, evidence on the effectiveness of mHealth interventions in changing behaviors remains limited [8]. When studied, they have primarily focused on a single approach, such as physical activity [9] or weight loss [10], rather than a comprehensive strategy targeting multiple lifestyle behaviors.

Furthermore, mHealth interventions may face challenges in maximizing educational effects due to a lack of empathic support from the absence of active human interaction [11]. The primary limitation of mHealth interventions may be characterized as the lack of a robust therapeutic alliance, as highlighted by Müssener [12]. To overcome these limitations, mHealth interventions may incorporate tailored interactions using real-time feedback messages as an innovative strategy to strengthen therapeutic alliance and promote heart-healthy behavioral changes [9]. For this purpose, mHealth interventions require technologies that integrate theory-based cognitive behavioral strategies, such as goal-setting, self-monitoring, and reinforcement or feedback [13]. Based on this background, we developed a mobile app called “My HeartHELP,” incorporating 3 behavioral strategies to promote comprehensive heart-healthy behaviors: providing information, encouraging self-monitoring, and delivering real-time automatized feedback messages [14]. The behavioral strategies of “My HeartHELP” may be innovative in the ability to address multiple heart-healthy behaviors simultaneously while tailoring real-time feedback to individual behavioral outcomes. Hence, evidence on the effectiveness of the “My HeartHELP” app on heart-healthy behavioral outcomes is needed.

Nurse-led face-to-face behavioral lifestyle interventions are still required, as they elicit small-to-moderate effects on behavioral changes [15]. However, their effectiveness in modifying lifestyle behaviors remains unclear in community settings without a system that educates and supports community-dwelling individuals in self-management. Among the several behavioral intervention modes proposed, motivational interviewing has been recognized as a potential approach to facilitate behavioral changes through person-centered care, fostering active interaction between the client and provider within the client’s psychosocial context [13]. This method involves counseling techniques such as reflections, affirmations, open questions, and summarizations, which are well-received by clients because they address internal conflicts and guide goal-setting [16]. By focusing on connecting behaviors to desired outcomes and evoking an individual’s intrinsic motivation, motivational interviewing encourages people to take an active role in their own change processes [17]. In the context of behavioral interventions, motivational interviewing has proven effective in managing substance use, smoking cessation, and physical activity [18]. Therefore, motivational interviewing may complement mHealth by addressing its shortcomings and could be integrated into mHealth interventions in community settings.

A hybrid approach integrating mHealth with face-to-face motivational interviewing counseling presents a promising alternative for addressing the limitations of mHealth. This approach offers 2 key advantages. First, it combines the technological convenience of mHealth with the person-centered care of face-to-face interventions, thereby complementing the constraints of online interventions in terms of therapeutic alliance [19]. Second, in-person interactions can enhance engagement in and adherence to lifestyle behaviors recommended by health care expert providers [20]. In this regard, the American Heart Association has posited that a combined approach involving multiple modalities of behavioral interventions targeting cardiovascular risk factor reduction may be more effective than a single-modality approach [13]. Furthermore, the addition of clinical expert counseling to mobile-based cardiovascular health programs has not only been recommended [21] but has also been associated with greater reductions in systolic blood pressure when nurse-led case management was integrated with mHealth interventions, compared to mHealth alone [22]. However, empirical research on whether hybrid interventions that combine online mHealth

XSL•FO
RenderX

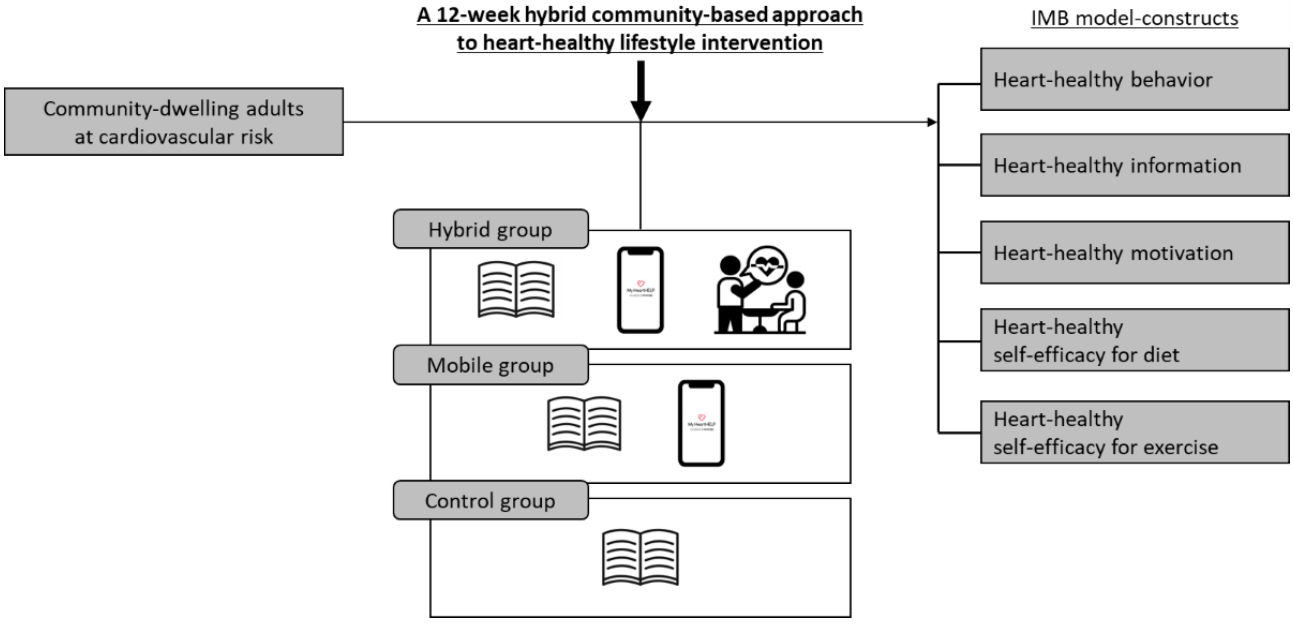
and offline face-to-face interventions are more effective than mHealth alone remains limited.

To address the current gap in the literature, we developed and implemented a hybrid community-based approach to a heart-healthy behavioral lifestyle intervention that combines the use of the "My HeartHELP" mobile app with face-to-face motivational interviewing to maximize behavioral outcomes in cardiovascular prevention and promotion. However, little information exists on whether such a hybrid approach is effective and, furthermore, whether it is superior to mobile app use alone in achieving heart-healthy behavioral outcomes within a community setting.

Meanwhile, the Information-Motivation-Behavioral Skills (IMB) model (Multimedia Appendix 1) served as the basis for developing the behavioral strategies of the hybrid community-based approach to a heart-healthy behavioral lifestyle intervention and determining the outcome variables to evaluate its effectiveness. The IMB model explains behavioral changes using constructs such as information, motivation, and behavioral skills [23]. Theoretically, motivation in the IMB model is directly associated with behavioral skills (ie, self-efficacy) and indirectly associated with behavior via self-efficacy [23]. Motivation enhancement may effectively change heart-healthy behaviors when used as part of lifestyle interventions [24]. In particular, the hybrid community-based approach was designed to enhance personal motivation by delivering information and behavioral skills for change through mobile text messaging and individually tailored counseling, while fostering self-monitoring to promote greater self-awareness and adoption of heart-healthy behaviors. Furthermore, its group sessions—grounded in motivational interviewing—were intended to facilitate the exchange of behavioral strategies and experiences through group dynamics, thereby strengthening social support and shared norms and, in turn, enhancing social motivation.

Therefore, the hybrid intervention based on the IMB model has a theoretical foundation aimed at informing individuals about cardiovascular health through frequent text messaging (heart-healthy information) and encouraging them through mHealth and motivational interviewing to promote intrinsic motivation (heart-healthy motivation). This, in turn, enhances behavioral skills (heart-healthy self-efficacy for diet and exercise) through cognitive-behavioral strategies and facilitates behavioral changes (heart-healthy behavior; Figure 1).

Figure 1. A theoretical framework of this study. The hybrid group received a 12-week hybrid intervention combining a mobile app (ie, "My HeartHELP") and face-to-face motivational interviewing; the mobile group used the mobile app (My HeartHELP), and the control group received a brochure on general heart-health. IMB: Information-Motivation-Behavior Skills.



We aimed to examine the effectiveness of a 12-week hybrid community-based heart-healthy lifestyle intervention among community-dwelling adults without a history of cardiovascular disease (Figure 1). We tested 2 hypotheses: first, a mobile group would be more likely to improve heart-healthy information, motivation, behavioral skills, and behavior (ie, IMB model-constructs) and cardiovascular parameters than a control group. Second, a hybrid group would be more likely to improve these outcomes than the mobile or control groups.

Methods

Study Design

This study (clinical trial number ISRCTN83643383) was a randomized controlled trial with 3 arms, parallel groups, and a 12-week follow-up period: a hybrid group, a mobile app group, and a control group. The pretest (T1) was conducted before the 12-week intervention, while the posttest (T2) was conducted after the intervention. This study was conducted in 2 cohort rounds to facilitate the effective implementation of the intervention by optimizing resources within a small research team. The rationale for the separate rounds was based on the

XSL•FO
RenderX

study by Burke et al [25]. The first round was conducted between November 2022 and February 2023, and the second round between February and May 2023 in a community-based setting in Seoul, South Korea.

As this study adopted a prospective, randomized, open-label, blinded end point design [26], double blinding was not feasible due to the nature of the interventions, which involved mobile app use and counseling. Specifically, neither study participants nor interviewers were blinded to this study's groups, but outcome assessors were blinded. We report our trial using the CONSORT (Consolidated Standards of Reporting Trials) guidelines (checklist provided in Multimedia Appendix 2) [27].

Participants

This study's participants comprised 75 community-dwelling adults, and the inclusion criteria were as follows: (1) aged between 20 and 64 years, and (2) having at least one metabolic syndrome according to the National Cholesterol Education Program Expert Panel [28]. The exclusion criteria were as follows: (1) medically diagnosed with diabetes mellitus and taking hypoglycemics or insulin, (2) medically diagnosed with either cardiovascular or psychiatric disease (ie, major depression or anxiety disorder), (3) physical activity limitations, (4) cognitive problems, and (5) inability to use mobile apps.

Participants were recruited using a notice on the internal bulletin board of the university with which the researchers were affiliated, as well as on the external bulletin boards of mobile communication platforms. The first round of recruitment was conducted from October 17 to November 11, 2022, and the second round from January 11 to February 3, 2023. When interested candidates contacted the researchers by phone to participate in this study, the researchers determined whether they were eligible and informed them that they could be randomly assigned to 3 groups. After agreeing to participate, participants were asked to attend a preliminary investigation session in person, where the purpose, methods, and procedures of this study were explained in detail based on this study's protocol, and written informed consent was obtained.

Sample Size

The minimum sample size per group was calculated based on the effect size of the composite heart-healthy behavior—the primary outcome variable in this study—between the mobile and control groups. According to Park [29], the mean difference (MD; Δ=change in mean composite heart-healthy behavior score from pretest to posttest) was 0.66 and –0.15 in intervention and control groups, respectively. However, we conservatively assumed that there would be no difference in the composite heart-healthy behavior between the pre- and posttests (ie, 0.0) in the control group. Thus, the effect size assumed in this study was 0.66.

The SD of the mean score changes for each group was estimated as follows, based on the results from the dissertation by Park [29]: pretest score SDs were assumed to be 0.47 and 0.56, and the posttest score SDs were 0.41 and 0.50 for the intervention and control groups, respectively. By conservatively assuming a correlation coefficient of 0.5 between pre- and posttests, and using the method proposed by Abrams et al [30], we projected the SD of the mean score changes from pre- to posttests to be 0.443 and 0.533 for the intervention and control groups, respectively. With these SDs of mean score changes and, under a 2-sided significance level (α) of 2.5%, a minimum of 20 subjects per group was needed to ensure a statistical power (1-β) of 95% to detect the effect size of 0.66. Assuming a 20% dropout rate during this study's period, we planned to enroll 25 participants in each group. Statistical software PASS 2020 (version. 20.0.8; NCSS LLC) was used to calculate the sample size.

Random Allocation

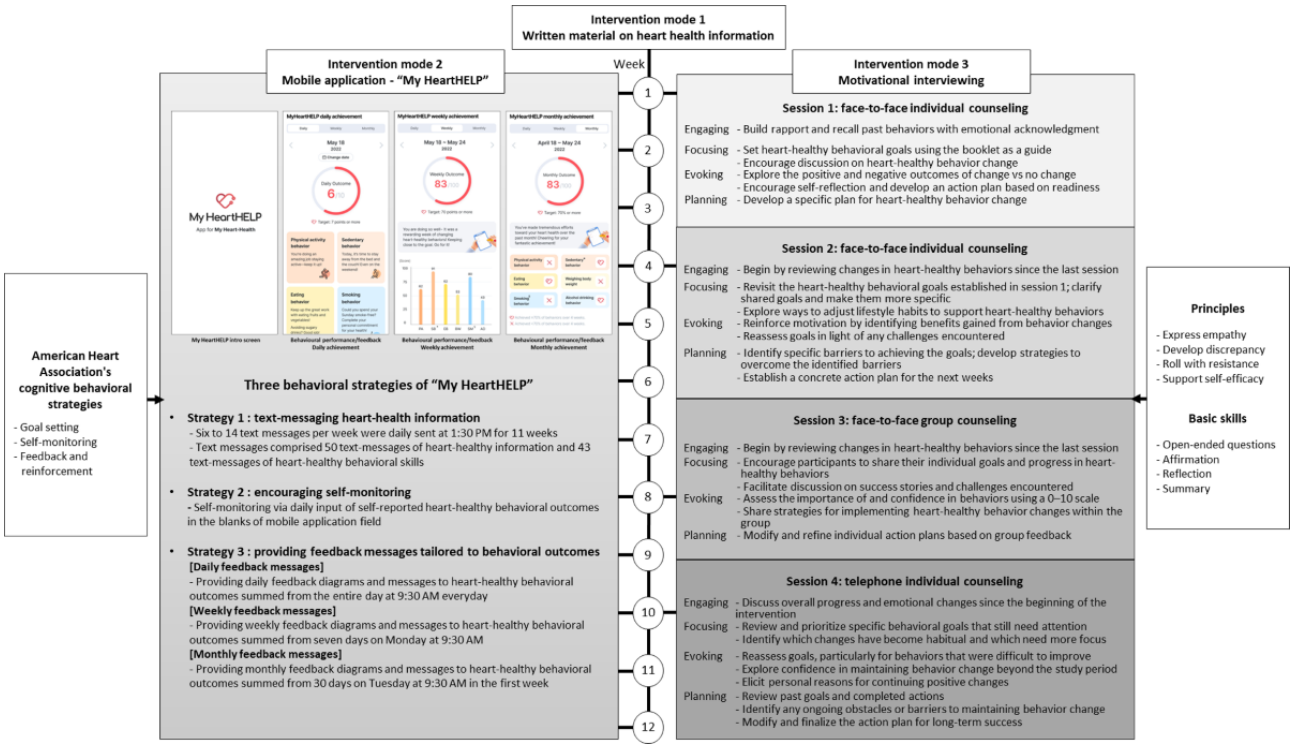
Random allocation in this study was conducted by the principal investigator using an age- and gender-stratified block randomization method with an allocation ratio of 1:1:1 across the 3 study arms. A randomization list was generated for each of the 10 age- and gender-stratified groups (ie, 20-29, 30-39, 40-49, 50-59, and 60-65 years, by gender) using R software (version 3.6.0; R Foundation for Statistical Computing). Study participants were listed in the order in which they were recruited, and each participant was assigned to an appropriately stratified group based on age and sex. Finally, a total of 75 participants were allocated to the hybrid (n=25), mobile (n=25), or control (n=25) groups.

Study Intervention

An outline and details of the hybrid intervention are presented in Figure 2, which contains three intervention modes: (1) intervention mode 1: written materials on general heart-health information, (2) intervention mode 2: the "My HeartHELP" mobile app [14], and (3) intervention mode 3: motivational interviewing counseling. General heart health information was delivered via a brochure, including information on cardiovascular diseases, cardiovascular risk factors, and general heart-healthy lifestyle changes. The duration of intervention modes 2 and 3 was 12 weeks. Specifically, the mobile app was designed for daily use throughout the intervention period, while motivational interviewing was delivered in 4 sessions, each lasting 2 hours, conducted at weeks 1, 4, 8, and 10.

XSL•FO
RenderX

Figure 2. An outline for the 12-week hybrid community-based heart-healthy lifestyle intervention. The 12-week hybrid intervention consisted of the three intervention modes: (1) intervention mode 1: written materials on general heart-health information, (2) intervention mode 2: the "My HeartHELP" mobile app, and (3) intervention mode 3: motivational interviewing counseling. General heart health information was delivered via a brochure, including information on cardiovascular diseases, cardiovascular risk factors, and general heart-healthy lifestyle changes. Please see Multimedia Appendix 3 for a larger version of this figure.



The hybrid group participated in all 3 intervention modes. The mobile group received 2 intervention modes of written materials and the "My HeartHELP" app. The control group received only the written materials.

Participants in both the hybrid and mobile groups attended separate 1-hour preintervention education sessions, which provided an overview of the 12-week intervention schedule and instructions on using the "My HeartHELP" app. A printed booklet containing a QR code for app download was distributed, and participants were informed that the app was freely available for research use. Each session also included a hands-on practice session, during which the research team assisted participants with app registration and guided them in entering sample data into the app.

The "My HeartHELP" app was implemented as an online mode, consisting of 3 strategies: strategy 1—text messaging with heart-health information, strategy 2—encouraging the self-monitoring of 6 lifestyle behaviors (physical activity, nonsedentary behavior, healthy dietary behaviors, nonsmoking, nonalcohol binge drinking, and daily weighing of body weight) [14], and strategy 3—providing automated or tailored feedback messages tailored to individual behavioral outcomes from self-monitoring (Figure 2). The text messaging intervention was delivered daily at 1:30 PM over a period of 1 to 11 weeks and comprised 50 messages with heart-health information and 43 messages focused on heart-healthy behavioral skills. Encouraging self-monitoring was also a daily practice, requiring participants to record behavioral outcomes for the 6 behaviors by entering data into the designated input fields of the app. Automated or tailored feedback messaging, including diagrams and texts, was daily, weekly, and monthly based. As a preparatory step toward developing automated and tailored feedback messaging, the research team established behavioral target goals and their ranges for daily, weekly, and monthly outcomes based on 6 heart-healthy behaviors. According to the ranges, a pool of feedback text messages was developed and embedded into the app's algorithm. For daily outcomes, scores obtained from self-monitoring of the 6 heart-healthy behaviors during the previous day were categorized as either >7 or <7, with the daily goal set at >7 out of a possible 10 points. The daily feedback messages were presented as a diagram and texts at 9:30 AM every day. For weekly outcomes, the average scores over the previous 7 consecutive days were converted to a 100-point scale and classified into 3 categories: ≥70, 50-69, and <50. The weekly goal was set at ≥70. The weekly feedback messages were delivered on Monday at 9:30 AM. For monthly outcomes, the mean scores across the preceding 4 weeks were dichotomized into ≥70 and <70. The monthly feedback messages were delivered as diagrams and text messages on Tuesday at 9:30 AM in the first week. The app automatically delivered tailored feedback messages through an embedded algorithm that operated according to predefined behavioral target goals and their corresponding ranges. Further details are available in the feasibility and acceptability study of the "My HeartHELP" mobile app [14].

Motivational interviewing counseling was an offline face-to-face intervention conducted in 4 sessions: sessions 1 (week 1) and 2 (week 4) for individualized counseling, session 3 (week 8) for group counseling, and session 4 (week 10) for telephone counseling (Figure 2). Each session followed the core processes

XSL•FO
RenderX

of engaging, focusing, evoking, and planning, as outlined in Figure 2 [17]. Engaging focused on building trust, understanding participants' unique perspectives, reviewing past behaviors, and evaluating emotional and behavioral changes [17]. Focusing aimed to clarify shared goals, making them more specific and actionable, while evoking strengthened participants' motivations for change by imagining potential outcomes and assessing their importance and confidence. Planning facilitated the development of practical strategies by identifying barriers to behavior change and determining ways to overcome them. The intervention was tailored to individual lifestyle contexts and used the principles of facilitating empathy, discrepancy, resistance, and self-efficacy along with open-ended questions, affirmations, reflections, and summaries techniques. Additional details are presented in Figure 2.

Fidelity of Intervention

To ensure the fidelity of the "My HeartHELP" app implementation, 2 measures used in previous studies [14,31] were applied: access and self-monitoring rates. Participants in both the hybrid and mobile groups were instructed to access the app daily and self-monitor their behavioral outcomes (ie, fill in the blanks of the behavioral components) daily for 12 weeks. Among those enrolled in the hybrid and mobile groups, 36/47 (76.6%) accessed the app at least once daily, while 45.7/47 (97.3%) engaged daily in self-monitoring of heart-healthy behaviors throughout the 12 weeks.

To ensure the fidelity of motivational interviewing, all authors completed training and certification through motivational interviewing courses, including 10 hours of fundamental information and 8 hours focused on health care [32]. One author implemented the motivational interviewing sessions over 12 weeks as an interventionist, with assistance from another author. To maintain counseling quality, an activity sheet template was developed and used for each session. Regarding attendance adherence, participants in the hybrid group attended 23/23 (100%) of the sessions.

Measures

Heart-Healthy Behavioral Outcomes Variables

The primary outcome was the composite score of heart-healthy behaviors. Secondary outcomes included scores for heart-healthy knowledge, heart-healthy self-efficacy, heart-healthy motivation, and cardiovascular parameters. All outcomes were measured at baseline (T1) and after 12 weeks (T2) using self-reported questionnaires, anthropometric measurements, and blood sampling.

Composite heart-healthy behavior refers to the extent to which participants practiced heart-healthy behaviors for cardiovascular health. This was measured using the Management Behaviors of Metabolic Syndrome Evaluation Tool developed by Kang [33] for individuals with metabolic syndrome. This tool consists of 36 questions: 8 on "physical activity and weight control," 16 on "dietary habits," 3 on "drinking and smoking," 3 on "stress," 2 on "sleep and rest," and 4 on "health check-up and management." The average score was calculated by summing the scores for each item (rated from "never" to "always" on a 4-point Likert scale) and dividing it by the number of items. Higher scores indicated greater heart-healthy behavior. Cronbach α was 0.92 in the study by Kang [33] and 0.91 in this current study.

Heart-healthy information refers to the level of knowledge regarding cardiovascular disease prevention. This was measured using the Heart-Healthy Information Questionnaire developed by Choo et al [34]. This tool consists of 50 questions, each answered as "true," "false," or "don't know." Participants who answered correctly received 1 point. If participants answered incorrectly or did not know, they received 0 points. The total score ranged from 0 to 50 points. The Kuder-Richardson formula 20 was 0.85 in the study by Choo et al [34] and 0.81 in this current study.

Motivation refers to intrinsic motivation levels for practicing heart-healthy behaviors. It was measured using the Intrinsic Motivation Inventory developed by McAuley et al [35], which we modified for the present study. Based on the inventory guidance indicating that subscales may be selectively used according to the study's focus [36], we used 5 of the 7 subscales, excluding 2 subscales. The pressure or tension subscale was excluded because it is more relevant to stressful activities such as competitive sports [37,38] than to heart-healthy behaviors. The relatedness subscale was also excluded because its items capture relational closeness—such as perceived distance, trust, and friendship with specific individuals—which are less relevant to the autonomous, self-directed process of adopting heart-healthy lifestyle habits. Consequently, 5 of the original 7 subscales were used—interest or enjoyment (7 items), perceived competence (6 items), effort or importance (5 items), perceived choice (7 items), and value or usefulness (7 items)—yielding a total of 32 items. The English version of the modified Intrinsic Motivation Inventory was minimally adjusted to fit heart-healthy behaviors, then translated into Korean separately by 3 Korean nursing scholars and consolidated into a single Korean version. A native English speaker back-translated this version, and the final English version was reviewed and confirmed by the original translators. Participants rated items using a 7-point Likert scale ranging from "completely disagree" to "strongly agree," with responses summed and averaged for each participant. Cronbach α was 0.85 in the study by McAuley et al [35] and 0.88 in this current study.

Heart-healthy self-efficacy refers to participants' confidence in their ability to engage in heart-healthy eating and exercise. It was measured using 2 instruments: the Self-Efficacy for Diet and Self-Efficacy for Exercise tools developed by Sallis et al [39]. The Self-Efficacy for Diet instrument comprises 20 items, while the Self-Efficacy for Exercise instrument has 12 items. Scores for eating and exercise habits were calculated by summing the responses on a 5-point Likert scale, ranging from "I cannot do this at all" to "I can definitely do this." Higher scores indicate higher self-efficacy levels. Cronbach α for eating was 0.84 in the study by Shin and Lach [40] and 0.91 in this current study. Cronbach α for exercise was 0.90 in the study by Sallis et al [39] and 0.93 in this current study.

Cardiovascular Parameters

Cardiovascular parameters comprised BMI, waist circumference, blood pressure, fasting glucose, total cholesterol,

XSL•FO
RenderX

low-density lipoprotein cholesterol, and high-density lipoprotein cholesterol. Body weight was measured after an overnight fast using the InBody270 scale (InBody Co, Ltd). Before weighing, participants wore a study-provided gown and removed their shoes. Height was measured using a wall-mounted stadiometer. BMI was calculated as weight (kg) divided by height (m^2). Waist circumference (cm) was measured twice using a Gullick II measuring tape at the midpoint between the lowest rib and the iliac crest, with the average of the 2 measurements used. Blood samples were drawn from the antecubital vein in the morning after a 10-hour overnight fast, without participants taking any current medications, including antihypertensive or lipid-lowering drugs. The samples were analyzed as described in a previous study [41].

Statistical Analysis

Participants' demographic and baseline characteristics were summarized using descriptive statistics, including frequency, percentage, mean, and SD. To test the homogeneity of the 3 groups' baseline measures, chi-square tests were used for categorical variables, while a 1-way ANOVA was performed for continuous variables (ie, age and study variables).

Both intention-to-treat and per-protocol approaches were used. The intention-to-treat analysis included all randomized participants regardless of adherence to the intervention, whereas the per-protocol population consisted of those who strictly adhered to the intervention protocol [42]. Missing values, which occurred due to nonresponse to the measurements, were not imputed but analyzed as observed. Analysis of covariance was used to compare outcome variables among the hybrid, mobile, and control groups at T2. In this analysis, T2 scores served as the dependent variable, study groups as the main exposure variables, and baseline scores for each respective outcome variable as confounders to adjust for individual differences at the start of the intervention. As a post hoc analysis, the Tukey least significant difference multiple comparison test was performed to identify significant differences between study groups in outcome variable changes. The analyses further examined significant group differences in the subscales of composite heart-healthy behavior and motivation. All statistical analyses were conducted using SPSS/WIN (version 28.0; SPSS Inc), with a 2-sided $P<.05$ considered statistically significant.

Ethical Considerations

This study was approved by the Institutional Review Board of Korea University (KUIRB-2022-0287-01). All participants provided written informed consent. All procedures were performed in accordance with the ethical standards of the Institutional Research Committee and the 2013 Declaration of Helsinki [43].

Safety and security procedures related to the use of the mobile app were carefully implemented. To ensure data privacy, access to the secure data server was restricted to a single designated member of the research team. Data were downloaded and processed offline to enhance data protection and confidentiality. Preintervention education sessions were conducted to inform participants about relevant safety and privacy measures. In addition, real-time user support was offered via KakaoTalk (Kakao Corp), a widely used mobile messaging platform in South Korea, enabling participants to directly communicate with the research team throughout the intervention period. Participants in the control group, mobile group, and hybrid group who completed only the pretest received a KR ₩30,000 (US $21) gift voucher. Participants who successfully completed all stages of the study received monetary compensation according to group assignment: KR ₩150,000 (US $104) for the control group, KR ₩250,000 (US $172) for the mobile group, and KR ₩300,000 (US $207) for the hybrid group. All details related to compensation were fully disclosed to participants during the informed consent process before their enrollment in this study.

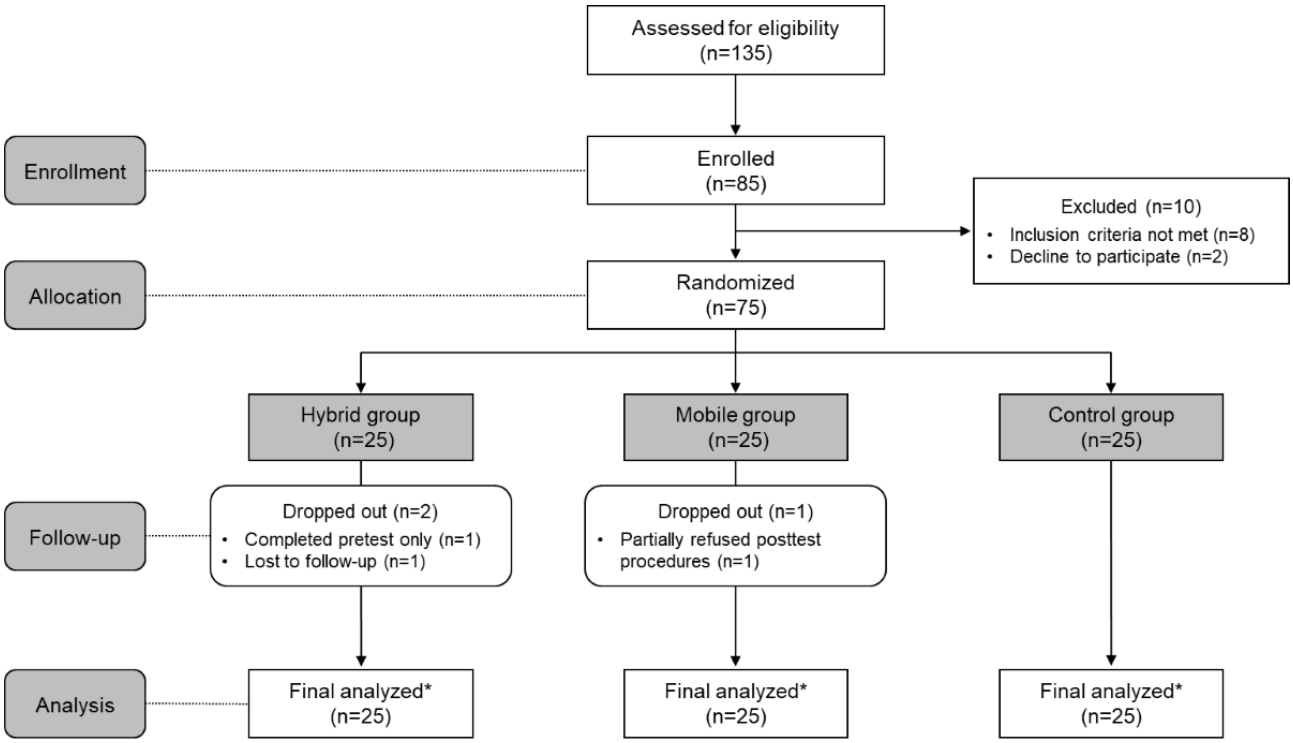
Results

Participants' General Characteristics

Three participants dropped out of this study for the following reasons: one participant from the hybrid group due to time constraints (n=1), another from the hybrid group due to loss to follow-up (n=1), and one from the mobile group who declined to participate in the posttest (n=1; Figure 3). Since the findings from both the intention-to-treat and per-protocol analyses yielded identical statistical significance, the results are presented based on the intention-to-treat analysis.

XSL•FO
RenderX

Figure 3. Participant flow in this study. The hybrid group received a 12-week hybrid intervention combining a mobile app (ie, "My HeartHELP") and face-to-face motivational interviewing; the mobile group used the mobile app, and the control group received written material on general heart-health information. *All participants were included in an intention-to-treat analysis.



The participants had a mean age of 43.6 (SD 11.3) years (Table 1). Of all participants, 46 out of 75 (61.3%) were women, and 64 out of 75 (85.3%) were college-educated. Additionally, 32 out of 75 (42.7%) had a median income level greater than or equal to that of the general South Korean population [44]. Among the participants, 52 out of 75 (69.3%) were used, and 16 out of 75 (21.3%) were taking antihypertensive or lipid-lowering medications. No significant differences in general characteristics were observed between the groups (Table 1).

XSL•FO
RenderX

Table 1. Participants' general characteristics (N=75) those of the general South Korean population [44]. Among the participants, 52/75 (69.3%) were employed, and 16/75 (21.3%) were taking antihypertensive or lipid-lowering medications. No significant differences in general characteristics were observed between the groups.

	All (N=75)	Hybrid group (n=25)	Mobile group (n=25)	Control group (n=25)	Chi-square (*df*)	*F* test (*df*)	*P* value
Age (years), mean (SD)	43.6 (11.3)	43.4 (11.0)	43.9 (12.3)	43.6 (11.0)		0.01 (2, 69)	.99
Gender, n (%)					0.11 (2, 69)		.94
Women	46 (61.3)	16 (64)	15 (60)	15 (60)			
Men	29 (38.7)	9 (36)	10 (40)	10 (40)			
Education, n (%)					0.21 (2, 69)		.90
More than college-educated	64 (85.3)	22 (88)	21 (84)	21 (84)			
Less than college-educated	11 (14.7)	3 (12)	4 (16)	4 (16)			
Monthly household income[a], 10,000 won, n (%)					2.07 (2, 69)		.36
>500	32 (42.7)	11 (44)	8 (32)	13 (52)			
<500	43 (57.3)	14 (56)	17 (68)	12 (48)			
Employed, n (%)					1.63 (2, 69)		.44
Yes	52 (69.3)	15 (60)	18 (72)	19 (76)			
No	23 (30.7)	10 (40)	7 (28)	6 (24)			
Medications[b], n (%)					1.11 (2, 69)		.57
Yes	16 (21.3)	5 (20)	4 (16)	7 (28)			
No	59 (78.7)	20 (80)	21 (84)	18 (72)			

[a]Monthly household income was classified using a cutoff of 5,000,000 won (US $3450; IQR 1767-5300), which represents the median income of the general population in South Korea [44].

[b]Taking either antihypertensives or lowering lipids medications.

Heart-Healthy Behavioral Outcome Variables

Table 2 summarizes the results of the differential effects of the 3 groups on the primary and secondary outcome variables over a 12-week intervention period. Significant differences were observed among the 3 groups in the scores for composite heart-healthy behavior ($F_{2,69}$=7.25, *P*=.001), motivation ($F_{2,69}$=8.54, *P*<.001), self-efficacy for diet ($F_{2,69}$=4.87, *P*=.01), and self-efficacy for exercise ($F_{2,69}$=5.48, *P*=.006). Compared to the control group, the hybrid and mobile groups demonstrated significantly greater increases in the scores for composite heart-healthy behavior (hybrid vs control: MD 0.37, SD 0.10, *P*<.001; mobile vs control: MD 0.26, SD 0.10; *P*=.01), motivation (hybrid vs control: MD 0.74, SD 0.18; *P*<.001; mobile vs control: MD 0.42, SD 0.18; *P*=.02), self-efficacy for diet (hybrid vs control: MD 8.44, SD 2.83; *P*=.004; mobile vs control: MD 6.29, SD 2.77; *P*=.03), and self-efficacy for exercise (hybrid vs control: MD 5.01, SD 2.27; *P*=.03>; mobile vs control: MD 7.23, SD 2.24; *P*=.002). There were no significant differences in the heart-healthy information scores among the 3 groups.

XSL•FO
RenderX

Table 2. Effectiveness of a hybrid community-based heart-healthy lifestyle intervention on heart-healthy behavioral outcome variables (N=75). T1 and T2 refer to the baseline time point and the time point after 12 weeks, respectively.

	Hybrid group (n=25)	Mobile group (n=25)	Control group (n=25)	*F* test (*df*)	*P* value
Composite heart-healthy behavior, mean (SD)					
Overall				7.25 (2, 69)	.001
T1	2.4 (0.46)	2.4 (0.47)	2.4 (0.42)		
T2	2.9 (0.46)[a]	2.8 (0.42)[a]	2.5 (0.50)[b]		
Physical activity and weight control				4.02 (2, 69)	.02
T1	1.9 (0.73)	2.2 (0.71)	2.0 (0.57)		
T2	2.5 (0.56)[a]	2.7 (0.75)[a]	2.2 (0.68)[b]		
Dietary habits				5.04 (2, 69)	.009
T1	2.3 (0.50)	2.3 (0.60)	2.3 (0.55)		
T2	2.9 (0.51)[a]	2.7 (0.58)[a,b]	2.5 (0.62)[b]		
Drinking and smoking				0.06 (2, 69)	.94
T1	3.3 (0.91)	3.2 (0.81)	3.1 (0.92)		
T2	3.3 (0.85)	3.3 (0.91)	3.2 (0.89)		
Stress				1.54 (2, 69)	.22
T1	3.2 (0.69)	3.0 (0.70)	2.9 (0.53)		
T2	3.4 (0.54)	3.2 (0.66)	3.0 (0.63)		
Sleep and rest				1.88 (2, 69)	.16
T1	2.9 (0.69)	2.7 (0.64)	2.8 (0.60)		
T2	3.1 (0.73)	3.1 (0.74)	2.8 (0.74)		
Health management				8.67 (2, 69)	<.001
T1	2.4 (0.59)	2.2 (0.55)	2.3 (0.69)		
T2	2.9 (0.59)[a]	2.7 (0.60)[a]	2.3 (0.70)[b]		
Heart-healthy information, mean (SD)				2.68 (2, 69)	.08
T1	43.0 (4.04)	41.9 (4.04)	38.6 (6.49)		
T2	45.5 (3.03)	44.0 (3.01)	41.6 (4.12)		
Heart-healthy motivation, mean (SD)					
Overall				8.54 (2, 69)	<.001
T1	4.4 (0.80)	4.6 (0.74)	4.4 (0.56)		
T2	5.5 (0.60)[a]	5.2 (0.80)[a]	4.7 (0.76)[b]		
Effort or importance				9.45 (2, 69)	<.001
T1	3.8 (1.22)	4.3 (1.24)	3.9 (1.07)		
T2	5.5 (0.73)[a]	5.4 (0.93)[a]	4.5 (1.15)[b]		
Interest or enjoyment				10.48 (2, 69)	<.001
T1	3.7 (1.24)	4.1 (1.21)	4.0 (0.95)		
T2	5.5 (1.02)[a]	5.1 (1.09)[b]	4.4 (1.09)[c]		
Perceived choice				1.83 (2, 69)	.17
T1	4.7 (0.70)	4.6 (0.61)	4.5 (0.78)		
T2	5.0 (0.65)	4.9 (0.60)	4.6 (0.87)		
Perceived competence				3.28 (2, 69)	.04

XSL•FO
RenderX

	Hybrid group (n=25)	Mobile group (n=25)	Control group (n=25)	*F* test (*df*)	*P* value
T1	3.5 (1.28)	3.9 (1.13)	3.8 (0.83)		
T2	4.8 (1.04)[a]	4.7 (1.16)[a,b]	4.2 (1.13)[b]		
Value or usefulness				1.85 (2, 69)	.16
T1	5.9 (0.95)	6.0 (0.80)	5.7 (0.81)		
T2	6.5 (0.67)	6.1 (1.21)	5.9 (0.95)		
Heart-healthy self-efficacy for diet, mean (SD)				4.87 (2, 69)	.01
T1	71.5 (13.94)	71.2 (13.79)	75.4 (11.86)		
T2	79.8 (11.48)[a]	77.3 (10.62)[a]	73.1 (12.89)[b]		
Heart-healthy self-efficacy for exercise, mean (SD)				5.48 (2, 69)	.006
T1	37.4 (11.26)	38.9 (10.25)	36.2 (8.32)		
T2	40.3 (10.07)[a]	43.5 (7.43)[a]	34.8 (10.32)[b]		

[a]Different superscripts indicate a statistically significant difference by Tukey least significant difference multiple comparison.

[b]Different superscripts indicate a statistically significant difference by Tukey least significant difference multiple comparison.

[c]Different superscripts indicate a statistically significant difference by Tukey least significant difference multiple comparison.

Regarding the subscales of heart-healthy behavior, significant group differences were observed in physical activity and weight control ($F_{2,69}$=4.02, *P*=.02), dietary habits ($F_{2,69}$=5.04, *P*=.009), and health management ($F_{2,69}$=8.67, *P*<.001; Table 2). Specifically, both the hybrid and mobile groups showed significantly greater increases in physical activity and weight control (hybrid vs control: MD 0.38, SD 0.17; *P*=.03; mobile vs control: MD 0.45, SD 0.17; *P*=.01), as well as health management (hybrid vs control: MD 0.62, SD 0.16; *P*<.001; mobile vs control: MD 0.49, SD 0.16; *P*=.002), compared with the control group. Moreover, the hybrid group (MD 0.42, SD 0.13; *P*=.002), but not the mobile group, demonstrated significantly greater improvements in dietary habit scores compared with the control group.

Regarding the subscales of heart-healthy motivation, significant group differences were observed in effort or importance ($F_{2,69}$=9.45, *P*<.001), interest or enjoyment ($F_{2,69}$=10.48, *P*<.001), and perceived competence ($F_{2,69}$=3.28, *P*=.04; Table 2). The hybrid group demonstrated significantly greater increases in interest or enjoyment than both the mobile (MD 0.54, SD 0.27; *P*=.045) and control groups (MD 1.21, SD 0.27; *P*<.001), while the mobile group also showed a significantly greater increase than the control group (MD 0.67, SD 0.26; *P*=.01). Both the hybrid (MD 1.00, SD 0.25; *P*<.001) and mobile (MD 0.79, SD 0.24; *P*=.002) groups showed significantly greater increases in effort or importance compared to the control group; however, there were no significant differences between the hybrid and mobile groups. Finally, the hybrid group (MD 0.70, SD 0.28; *P*=.01), but not the mobile group, demonstrated a significantly greater increase in perceived competence compared to the control group.

Cardiovascular Parameters

Table 3 summarizes the differential effects of the 3 groups on cardiovascular parameters over 12 weeks. Significant differences were observed between the 3 groups in fasting glucose levels ($F_{2,69}$=3.90, *P*=.03), with significant differences between the hybrid and control groups (*P*=.02) and the mobile and control groups (*P*=.02). However, there were no significant differences in the other parameters among the 3 groups.

XSL•FO
RenderX

Table 3. Effectiveness of a hybrid community-based heart-healthy lifestyle intervention on cardiovascular parameters (N=75). T1 and T2 refer to the baseline time point and the time point after 12 weeks, respectively.

	Hybrid group (n=25)	Mobile group (n=25)	Control group (n=25)	*F* test (*df*)	*P* value
BMI (kg/m^2), mean (SD)				0.82 (2, 69)	.45
T1	28.2 (4.84)	26.6 (3.73)	26.2 (4.14)		
T2	27.9 (4.79)	26.3 (3.93)	26.2 (4.01)		
WC[a] (cm), mean (SD)				0.20 (2, 69)	.82
T1	92.1 (10.47)	88.9 (10.39)	87.8 (11.26)		
T2	89.5 (9.39)	87.0 (10.47)	85.3 (10.73)		
Systolic BP[b], mean (SD)				0.07 (2, 69)	.93
T1	127.0 (18.36)	124.3 (19.10)	124.1 (18.77)		
T2	123.8 (14.77)	121.3 (17.15)	120.3 (12.30)		
Fasting glucose, mean (SD)				3.90 (2, 69)	.03
T1	93.9 (13.62)	92.9 (12.31)	92.2 (10.47)		
T2	90.6 (7.90)[c]	89.2 (8.98)[c]	93.2 (10.12)[d]		
LDL[e] cholesterol, mean (SD)				0.38 (2, 69)	.69
T1	128.0 (26.67)	123.5 (35.48)	126.0 (34.77)		
T2	129.9 (21.88)	122.5 (30.42)	123.0 (36.08)		
HDL[f] cholesterol, mean (SD)				0.78 (2, 69)	.46
T1	53.2 (14.81)	53.6 (13.20)	55.4 (14.86)		
T2	55.0 (13.83)	57.7 (13.70)	56.9 (13.18)		
Triglycerides, mean (SD)				0.57 (2, 69)	.57
T1	248.3 (372.18)	169.8 (86.41)	155.7 (104.73)		
T2	163.1 (83.39)	130.5 (65.80)	145.0 (85.40)		

[a]WC: waist circumference.

[b]BP: blood pressure.

[c]Different superscripts indicate a statistically significant difference by Tukey least significant difference multiple comparison.

[d]Different superscripts indicate a statistically significant difference by Tukey least significant difference multiple comparison.

[e]LDL: low-density lipoprotein.

[f]HDL: high-density lipoprotein.

Discussion

Principal Results

We found that, compared to the control group, both the hybrid and mobile intervention groups demonstrated significantly greater improvements in heart-healthy behavioral outcomes, including composite heart-healthy behavior, its theoretical predictors (heart-healthy motivation and self-efficacy), and fasting glucose levels following the 12-week hybrid intervention for individuals at cardiovascular risk. In particular, the hybrid group—unlike the mobile group—showed significantly greater improvement in dietary behavior, a subscale of composite heart-healthy behavior, compared with the control group, and also demonstrated significantly greater improvements in interest or enjoyment, a core subscale of intrinsic motivation, than the mobile and control groups [45]. These findings suggest that the hybrid community-based heart-healthy lifestyle intervention—integrating the mobile app and motivational interviewing—demonstrated overall effectiveness comparable to the mobile app alone, while yielding greater improvements in dietary behavior and core intrinsic motivation.

The hybrid and mobile groups were more likely to increase the composite score of heart-healthy behaviors than the control group after the 12-week intervention, with no significant difference observed between the 2 groups. However, the hybrid group—unlike the mobile group—showed significantly greater improvement in dietary behavior, a subscale of composite heart-healthy behavior, compared with the control group. The finding regarding composite heart-healthy behavior suggests that motivational interviewing may not have demonstrated superiority over the use of a mobile app alone, indicating that the “My HeartHELP” mobile app itself may effectively facilitate overall improvements in heart-healthy behaviors. The “My HeartHELP” app uniquely incorporates 6 key heart-healthy

XSL•FO
RenderX

behaviors simultaneously, which may significantly affect the composite score of heart-healthy behaviors. Empirically, this positive finding may be explained by the fact that the "My HeartHELP" app incorporates evidence-based behavioral change strategies [10,13,14,23], including text messaging for enhancing information, encouraging self-monitoring, and providing feedback or reinforcement in line with individuals' behavioral goal setting.

To date, few mHealth studies on cardiovascular health have primarily relied on theory-based smartphone apps targeting multiple heart-healthy behaviors, as in our study. Most studies, by contrast, have used a variety of mHealth modalities, such as commercial apps, telephone, web-based, email, and SMS text messaging, each corresponding to a single specific behavior, such as physical activity, weight loss, or smoking [20]. Furthermore, most previous studies have predominantly emphasized health outcomes as direct indicators of effectiveness [6], without accounting for the internal validity of mHealth interventions through behavioral change mechanisms. In this regard, our study may be unique in its integration of a mobile app with evidence-based behavioral change strategies specifically designed to promote multiple heart-healthy behaviors and in its comprehensive evaluation of both behavioral changes and health outcomes, including cardiovascular parameters [14].

Meanwhile, the hybrid group, but not the mobile group, showed a significantly greater improvement in the subscale for dietary behavior after 12 weeks, compared to the control group. This finding indicates that motivational interviewing may have had an additional effect on changing dietary habits through the use of the "My HeartHELP" app. Motivational interviewing has traditionally demonstrated significant effectiveness in promoting behavioral changes in substance use, especially in diverse psychological contexts [17]. Dietary behavior involves adherence to diverse dietary guidelines and the consumption of various food groups [46], making effective intervention challenging without considering the psychological and social contexts of individuals [47]. In this regard, motivational interviewing was incorporated in the present study to mitigate the limitations of mobile apps in facilitating close expert interaction. This approach may be effective in facilitating a person-centered understanding of the psychosocial context of complex eating behavior and allows for personalized adjustments to eating behaviors, thus enhancing the intervention's effectiveness. This highlights motivational interviewing's particular efficacy for complex behavioral domains that require nuanced, individualized support beyond what a mobile app alone can provide.

Our findings demonstrated that both the hybrid and mobile groups exhibited greater increases in overall scores of the heart-healthy motivation compared to the control group, with no significant difference between the hybrid and mobile groups. This current study did not fully support the hypothesis that motivational interviewing would be more effective than using a mobile app alone [48,49]. Nevertheless, the hybrid group demonstrated significantly greater increases in a core subscale of intrinsic motivation—interest or enjoyment [50]—than the mobile group. In this context, the above findings may not entirely rule out the possibility that motivational interviewing could serve as an effective intervention component for strengthening intrinsic motivation [24,51], potentially offering advantages beyond the use of a mobile app alone. This finding also aligns with the conceptualization of intrinsic motivation as measured by the 4 items for interest or enjoyment in the Situational Motivation Scale [45].

However, the lack of overall statistical superiority of motivational interviewing may be explained from 2 perspectives. First, the "My HeartHELP" app incorporated substantial motivational elements, including self-monitoring, automated and tailored feedback, and goal attainment encouragement [52]. Second, the limited dosage of the motivational interviewing—4 sessions over 12 weeks—may have been insufficient to produce statistically significant additive effects across all motivational outcomes [53]. In line with this dosage, the quality and intensity of motivational interviewing delivered in this study may have constrained its effectiveness, beyond that of the mobile app alone, in fostering individuals' autonomy to adjust behavioral goal settings, as both the mobile and hybrid groups adhered to the same predefined behavioral goals. Future research should optimize the dosage of motivational interviewing and enhance its quality of autonomy-supportive interventions, thereby more strategically and intensively enabling participants to independently design their activities and personalize goal-setting, as emphasized in self-determination theory [54].

Theoretically grounded in the IMB model [23], the observed increase in heart-healthy motivation in both the hybrid and mobile groups can be interpreted as evidence that enhanced motivation may have served as a foundational mechanism strengthening the mediating variable—heart-healthy self-efficacy—which, in turn, contributed to improvements in heart-healthy behavior (Multimedia Appendix 1). Consistent with this theoretical pathway, our path analytic data (not shown) indicated that increases in overall heart-healthy motivation scores were significantly and positively associated with increases in self-efficacy for diet (β=.32, *P*=.02) and exercise (β=.38, *P*=.005), as well as with improvements in heart-healthy behaviors (β=.27, *P*=.04). These findings may be attributable to the intervention strategies used in this study, which fostered self-directed, health-oriented internalization through self-monitoring and cognitive reappraisal—encompassing the daily input and reflection on 6 core heart-healthy behaviors via the app and motivational interviewing that encouraged participants to recognize progress and use self-management. Subsequently, self-efficacy for diet and exercise may have been enhanced through text messaging and tailored feedback reflecting individual success (eg, mastery experience) delivered via "My HeartHELP" [14], along with professional encouragement and persuasive support provided through motivational interviewing. Collectively, these cognitive and behavioral mechanisms may have contributed to the observed improvements in overall heart-healthy behaviors in either the hybrid or mobile group.

In our findings, the significantly increased self-efficacy for diet and exercise after 12 weeks of using the "My HeartHELP" app may be explained by the efficacy of the behavioral strategies embedded in the app. Goal-setting and self-monitoring can

XSL•FO
RenderX

contribute to mastery experiences [55] by allowing participants to observe their own progress and achieve incremental success, reinforcing their belief in their ability to sustain heart-healthy behaviors [55]. Moreover, the "My HeartHELP" app's personalized feedback messaging delivered daily, weekly, and monthly based on individuals' behavioral outcomes could function as a form of verbal persuasion, as also described by Bandura [55], reinforcing self-efficacy through experts' positive reinforcement and targeted feedback on behavioral progress. Moreover, the "My HeartHELP" app delivered 43 text messages on behavioral skills, informing users on how to apply these skills in daily life. Mobile apps are currently limited in their ability to sensitively recognize and address individuals' emotional states. However, the 4 core principles of motivational interviewing (ie, expressing empathy, developing discrepancies, rolling with resistance, and supporting self-efficacy) are thought to be effective in addressing such emotional states [17]. Previous studies have suggested that motivational interviewing, as a behavioral intervention, can be highly effective in enhancing individuals' self-efficacy [51]. Nevertheless, the lack of differences observed between the hybrid and mobile-only groups may be due to the strong self-efficacy effects of the mobile app overshadowing the effects of motivational interviewing.

We found no significant differences in heart-healthy information levels among the 3 groups following the 12-week intervention. All participants were provided with educational materials on cardiovascular health; the mobile group received cardiovascular information via text messaging, while the hybrid group received both text messages and individualized information. However, an analysis of the item-response rate revealed that participants in all 3 groups continued to provide incorrect responses to certain items on the Heart-Healthy Information Questionnaire at posttest—items that were nearly identical to those answered incorrectly at baseline [34]—despite completing the 12-week intervention. These findings suggest that interventionists should proactively address specific questionnaire items that participants struggled with at the baseline assessment. Integrating targeted educational reinforcement into the intervention design may be necessary to enhance knowledge acquisition and retention.

Strengths and Limitations of This Study

This study underscored the potential of mHealth within a community setting for behavioral lifestyle interventions by optimizing user-interventionist interactions through tailored feedback delivered via in-app text messaging. To the best of our knowledge, this study is the first to reveal the effectiveness of a hybrid approach combining mobile apps and motivational interviewing counseling on heart-healthy behavioral changes, compared to mobile apps alone, within a community setting. Moreover, this indicated that the mHealth intervention using the mobile app effectively facilitated significant changes in heart-healthy behaviors, and the validity of these changes was confirmed through significant improvements in behavioral predictors based on the IMB model. This is particularly meaningful, as it establishes the internal validity of the intervention, demonstrating that the observed behavioral changes were systematically driven by theoretical constructs.

Nevertheless, this study has several limitations. First, dropouts (2 subjects in the hybrid group and 1 in the mobile group) may have led to a bias influencing the validity of the results, even though we performed an intention-to-treat analysis. Second, the absence of participant blinding in this study's design may have introduced bias into this study's findings. Third, because all the participants in this study were Korean, the results cannot be generalized to other ethnic groups. Fourth, because the survey was conducted online, participants with lower digital literacy were likely underrepresented. This is reflected in the socioeconomic profile of our participants, with a high proportion of college-educated individuals (64/75, 85.3%) compared with the national average in South Korea [36]. These characteristics may have facilitated greater engagement with the mHealth intervention and should be considered when generalizing the findings to populations with lower education levels or limited digital access. Fifth, this study lacks a qualitative component, which limits the understanding of why the intervention was effective from the participants' perspective. Qualitative feedback, such as through user interviews, could have provided invaluable insight into which app features were most engaging or how the motivational interviewing sessions were perceived. Sixth, given that this study targeted multiple healthy lifestyle behaviors, appropriate behavioral and self-efficacy measures specifically designed for the comprehensive assessment of heart-healthy behaviors are lacking. Moreover, the measure of heart-healthy motivation used in this study excluded 2 subscales—relatedness and pressure or tension—because their items were deemed inappropriate for the context of adopting heart-healthy behaviors, although these dimensions, particularly relatedness, may play an essential role. The rationale for this exclusion has been described above in the Methods section. The instrument was developed and refined through a rigorous translation and back-translation process; however, the absence of a formal validation process, including factor analysis and psychometric testing, warrants cautious interpretation of the present findings. Therefore, a subsequent validation study should be conducted among the Korean population. In light of these limitations, future research should prioritize the development and validation of robust assessment tools that accurately capture the multidimensional nature of heart-healthy behavioral constructs, thereby ensuring a more precise evaluation of intervention outcomes.

Implications for Policy and Practice

The use of a mobile app alone may be effective in facilitating changes in heart-healthy behaviors when integrated with evidence-based behavioral strategies. Given the limitations of expert-centered cardiovascular care in community settings, a mobile heart-healthy intervention led by nurses or other health professionals may serve as an efficient and accessible alternative. Although motivational interviewing did not demonstrate universal superiority over the mobile app alone, it remains a valuable component for behavioral interventions. More importantly, the hybrid approach used in this study demonstrated the potential of motivational interviewing not only to enhance intervention effects—particularly in improving intricate heart-healthy dietary behaviors—but also to promote sustainability by reinforcing intrinsic motivation, specifically

XSL•FO
RenderX

interest or enjoyment [45]. We therefore strongly recommend integrating motivational interviewing into mHealth interventions, especially when targeting subtle behavioral changes and fostering core intrinsic motivation, as supported by our findings on the motivation subscales. Furthermore, the intervention could be further strengthened by integrating periodic counseling sessions using the motivational interviewing modality alongside the mobile intervention. Therefore, workforce training programs should be implemented to equip health professionals with the skills necessary to deliver mHealth interventions and motivational interviewing as components of cardiovascular health promotion. This approach could enhance the effectiveness and accessibility of heart health interventions in community settings.

Recommendations for Further Research

Future studies should explore the long-term effects of mobile apps and their combinations with motivational interviewing on heart-healthy behavioral outcomes. Additionally, optimizing the frequency and delivery of motivational interviewing sessions should be considered, particularly within the constraints of limited resources in community settings, to enhance the sustainability and effectiveness of behavioral interventions.

Conclusions

The hybrid community-based heart-healthy lifestyle intervention—integrating the mobile app and motivational interviewing—demonstrated comparable overall effectiveness to the mobile app alone, yet achieved greater improvements in intrinsic motivation (interest or enjoyment) and dietary behavior. These findings highlight the potential of mHealth apps as practical, stand-alone tools to promote cardiovascular health, particularly in community settings with limited access to in-person professional support. However, incorporating motivational interviewing may further enhance internalized motivation and sustain complex behavior changes over time. Health professionals can therefore adopt mHealth either independently or in combination with motivational interviewing to optimize heart-healthy behavioral outcomes. Future studies should optimize integration strategies to enhance effectiveness and evaluate the long-term sustainability of such hybrid approaches.

Funding

This research was supported by the National Research Foundation of Korea (NRF) grant from the Korean Government (No. NRF-2019R1A2C1004116) and that from the Ministry of Science and ICT (Information and Communication Technology; No. RS-2024-00336847).

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: JC
Data curation: YS, SN
Formal analysis: YS
Funding acquisition: JC
Methodology: JC
Project administration: JC, SN
Resources: YS, SN
Supervision: JC, JL
Validation: JC, JL
Visualization: YS
Writing – original draft: JC
Writing – review & editing: JC, YS, SN, JL

Conflicts of Interest

None declared.

Multimedia Appendix 1
The hypothetical model based on constructs of the IMB model. IMB: Information-Motivation-Behavior. Reprinted with permission from the American Psychological Association. 2025 American Psychological Association. License obtained.
[PNG File , 53 KB - mhealth_v14i1e76521_app1.png]

Multimedia Appendix 2
CONSORT-EHEALTH (V1.6.1) checklist.
[PDF File (Adobe PDF File), 1318 KB - mhealth_v14i1e76521_app2.pdf]

XSL•FO
RenderX

Multimedia Appendix 3
Expanded outline of the 12-week hybrid community-based heart-healthy lifestyle intervention.
[PNG File , 1008 KB - mhealth_v14i1e76521_app3.png]

References

1. World Heart Report 2023: confronting the world's number one killer. World Heart Federation. Geneva, Switzerland: World Heart Federation; 2023. URL: https://world-heart-federation.org/wp-content/uploads/World-Heart-Report-2023.pdf [accessed 2025-12-20]
2. Cardiovascular diseases (CVDs). World Health Organization. Geneva, Switzerland: World Health Organization; 2025. URL: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) [accessed 2025-12-20]
3. Wu J, Feng Y, Zhao Y, Guo Z, Liu R, Zeng X, et al. Lifestyle behaviors and risk of cardiovascular disease and prognosis among individuals with cardiovascular disease: a systematic review and meta-analysis of 71 prospective cohort studies. Int J Behav Nutr Phys Act 2024;21(1):42 [FREE Full text] [doi: 10.1186/s12966-024-01586-7] [Medline: 38650004]
4. Loef M, Walach H. The combined effects of healthy lifestyle behaviors on all cause mortality: a systematic review and meta-analysis. Prev Med 2012;55(3):163-170. [doi: 10.1016/j.ypmed.2012.06.017] [Medline: 22735042]
5. Fang J, Moore L, Loustalot F, Yang Q, Ayala C. Reporting of adherence to healthy lifestyle behaviors among hypertensive adults in the 50 states and the District of Columbia, 2013. J Am Soc Hypertens 2016;10(3):252-262.e3 [FREE Full text] [doi: 10.1016/j.jash.2016.01.008] [Medline: 26851000]
6. Sequi-Dominguez I, Alvarez-Bueno C, Martinez-Vizcaino V, Fernandez-Rodriguez R, Del Saz Lara A, Cavero-Redondo I. Effectiveness of mobile health interventions promoting physical activity and lifestyle interventions to reduce cardiovascular risk among individuals with metabolic syndrome: systematic review and meta-analysis. J Med Internet Res 2020;22(8):e17790 [FREE Full text] [doi: 10.2196/17790] [Medline: 32865503]
7. Piette JD, List J, Rana GK, Townsend W, Striplin D, Heisler M. Mobile Health devices as tools for worldwide cardiovascular risk reduction and disease management. Circulation 2015;132(21):2012-2027 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.008723] [Medline: 26596977]
8. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR mHealth uHealth 2020;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
9. Khunti K, Griffin S, Brennan A, Dallosso H, Davies MJ, Eborall HC, et al. Promoting physical activity in a multi-ethnic population at high risk of diabetes: the 48-month PROPELS randomised controlled trial. BMC Med 2021;19(1):130 [FREE Full text] [doi: 10.1186/s12916-021-01997-4] [Medline: 34078362]
10. Burke LE, Sereika SM, Parmanto B, Bizhanova Z, Kariuki JK, Cheng J, et al. Effect of tailored, daily feedback with lifestyle self-monitoring on weight loss: the SMARTER randomized clinical trial. Obesity (Silver Spring) 2022;30(1):75-84. [doi: 10.1002/oby.23321] [Medline: 34898011]
11. Giebel GD, Speckemeier C, Abels C, Plescher F, Börchers K, Wasem J, et al. Problems and barriers related to the use of digital health applications: scoping review. J Med Internet Res 2023;25:e43808 [FREE Full text] [doi: 10.2196/43808] [Medline: 37171838]
12. Müssener U. Digital encounters: human interactions in mHealth behavior change interventions. Digit Health 2021;7:20552076211029776 [FREE Full text] [doi: 10.1177/20552076211029776] [Medline: 34262783]
13. Artinian NT, Fletcher GF, Mozaffarian D, Kris-Etherton P, Van Horn L, Lichtenstein AH, American Heart Association Prevention Committee of the Council on Cardiovascular Nursing. Interventions to promote physical activity and dietary lifestyle changes for cardiovascular risk factor reduction in adults: a scientific statement from the American Heart Association. Circulation 2010;122(4):406-441 [FREE Full text] [doi: 10.1161/CIR.0b013e3181e8edf1] [Medline: 20625115]
14. Choo J, Noh S, Shin Y. Evaluating feasibility and acceptability of the "My HeartHELP" mobile app for promoting heart-healthy lifestyle behaviors: mixed methods study. JMIR Form Res 2025;9:e66108 [FREE Full text] [doi: 10.2196/66108] [Medline: 40315608]
15. Patnode CD, Redmond N, Iacocca MO, Henninger M. Behavioral counseling interventions to promote a healthy diet and physical activity for cardiovascular disease prevention in adults without known cardiovascular disease risk factors: updated evidence report and systematic review for the US preventive services task force. JAMA 2022;328(4):375-388. [doi: 10.1001/jama.2022.7408] [Medline: 35881116]
16. Polcin DL, Sterling J, Brown T, Brown M, Buscemi R, Korcha R. Client and therapist views about intensive and standard motivational interviewing. J Contemp Psychother 2015;45(3):167-176 [FREE Full text] [doi: 10.1007/s10879-014-9280-1] [Medline: 26185335]
17. Miller W, Rollnick S. Motivational Interviewing: Helping People Change. 3rd ed. New York (NY): Guilford Press; 2012.
18. Frost H, Campbell P, Maxwell M, O'Carroll RE, Dombrowski SU, Williams B, et al. Effectiveness of motivational interviewing on adult behaviour change in health and social care settings: a systematic review of reviews. PLoS One 2018;13(10):e0204890 [FREE Full text] [doi: 10.1371/journal.pone.0204890] [Medline: 30335780]
19. Rotger JM, Cabré V. Therapeutic alliance in online and face-to-face psychological treatment: comparative study. JMIR Ment Health 2022;9(5):e36775 [FREE Full text] [doi: 10.2196/36775] [Medline: 35499910]

XSL•FO
RenderX

20. Cajita MI, Zheng Y, Kariuki JK, Vuckovic KM, Burke LE. mHealth technology and CVD risk reduction. Curr Atheroscler Rep 2021;23(7):36. [doi: 10.1007/s11883-021-00927-2] [Medline: 33983491]
21. Yun YH, Kang E, Cho YM, Park SM, Kim Y, Lee H, et al. Efficacy of an electronic health management program for patients with cardiovascular risk: randomized controlled trial. J Med Internet Res 2020;22(1):e15057 [FREE Full text] [doi: 10.2196/15057] [Medline: 32012053]
22. Ogedegbe G, Teresi JA, Williams SK, Ogunlade A, Izeogu C, Eimicke JP, et al. Home blood pressure telemonitoring and nurse case management in Black and Hispanic patients with stroke: a randomized clinical trial. JAMA 2024;332(1):41-50. [doi: 10.1001/jama.2024.6609] [Medline: 38842799]
23. Fisher W, Fisher J, Harman J. The information-motivation-behavioral skills model: a general social psychological approach to understanding and promoting health behavior. In: Social Psychological Foundations of Health and Illness. New York (NY): Wiley; 2003:82-106.
24. Nuss K, Moore K, Marchant T, Courtney JB, Edwards K, Sharp JL, et al. The combined effect of motivational interviewing and wearable fitness trackers on motivation and physical activity in inactive adults: a randomized controlled trial. J Sports Sci 2023;41(1):45-55. [doi: 10.1080/02640414.2023.2195228] [Medline: 36966352]
25. Burke LE, Choo J, Music E, Warziski M, Styn MA, Kim Y, et al. PREFER study: a randomized clinical trial testing treatment preference and two dietary options in behavioral weight management — rationale, design and baseline characteristics. Contemp Clin Trials 2006;27(1):34-48. [doi: 10.1016/j.cct.2005.08.002] [Medline: 16233990]
26. Hansson L, Hedner T, Dahlöf B. Prospective randomized open blinded end-point (PROBE) study. A novel design for intervention trials. Blood Press 1992;1(2):113-119 [FREE Full text] [doi: 10.3109/08037059209077502] [Medline: 1366259]
27. Eysenbach G. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
28. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation,Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. Circulation 2002;106(25):3143-3421. [Medline: 12485966]
29. Park S. Effects of a Community-Based Nursing Program for Promoting Cardiometabolic Health Behaviors Among Postmenopausal Middle-Aged Women: on a Basis of the Extended Theory of Planned Behavior [dissertation]. Seoul (KR): Korea University; 2019.
30. Abrams KR, Gillies CL, Lambert PC. Meta-analysis of heterogeneously reported trials assessing change from baseline. Stat Med 2005;24(24):3823-3844. [doi: 10.1002/sim.2423] [Medline: 16320285]
31. Symer MM, Abelson JS, Milsom J, McClure B, Yeo HL. A mobile health application to track patients after gastrointestinal surgery: results from a pilot study. J Gastrointest Surg 2017;21(9):1500-1505. [doi: 10.1007/s11605-017-3482-2] [Medline: 28685388]
32. Miller W, Rollnick S, Moyers T. Motivational interviewing. 2021. URL: https://psychwire.com/motivational-interviewing [accessed 2025-12-20]
33. Kang S. The validity and reliability of a lifestyle evaluation tool for patients with metabolic syndrome. J Korean Acad Fundam Nurs 2010;17(4):487-497 [FREE Full text]
34. Choo J, Noh S, Moon J, Shin Y. Development and psychometric testing of the Heart-Healthy Information Questionnaire. Eur J Cardiovasc Nurs 2023;22(3):299-310. [doi: 10.1093/eurjcn/zvac055] [Medline: 35766172]
35. McAuley E, Duncan T, Tammen VV. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: a confirmatory factor analysis. Res Q Exerc Sport 1989;60(1):48-58. [doi: 10.1080/02701367.1989.10607413] [Medline: 2489825]
36. Korea Disease Control and Prevention Agency. The eighth korea national health and nutrition examination survey (KNHANES VIII) Internet. Cheongju (KR): KDCA; 2021. URL: https://knhanes.kdca.go.kr/knhanes/eng/main.do [accessed 2026-01-03]
37. Cocca A, Veulliet N, Niedermeier M, Drenowatz C, Cocca M, Greier K, et al. Psychometric parameters of the intrinsic motivation inventory adapted to physical education in a sample of active adults from Austria. Sustainability 2022;14(20):13681. [doi: 10.3390/su142013681]
38. Gibbens B. Measuring student motivation in an introductory biology class. Am Biol Teach 2019;81(1):20-26. [doi: 10.1525/abt.2019.81.1.20]
39. Sallis JF, Pinski RB, Grossman RM, Patterson TL, Nader PR. The development of self-efficacy scales for healthrelated diet and exercise behaviors. Health Educ Res 1988;3(3):283-292. [doi: 10.1093/her/3.3.283]
40. Shin C, Lach H. Nutritional issues of Korean Americans. Clin Nurs Res 2011;20(2):162-180. [doi: 10.1177/1054773810393334] [Medline: 21160079]
41. Choo J, Lee J, Cho J, Burke LE, Sekikawa A, Jae SY. Effects of weight management by exercise modes on markers of subclinical atherosclerosis and cardiometabolic profile among women with abdominal obesity: a randomized controlled trial. BMC Cardiovasc Disord 2014;14:82 [FREE Full text] [doi: 10.1186/1471-2261-14-82] [Medline: 25011384]
42. Tripepi G, Chesnaye NC, Dekker FW, Zoccali C, Jager KJ. Intention to treat and per protocol analysis in clinical trials. Nephrology (Carlton) 2020 Jul;25(7):513-517. [doi: 10.1111/nep.13709] [Medline: 32147926]

XSL•FO
RenderX

43. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
44. Ministry of Health and Welfare. The Median Income Level of Korea. Sejong (KR): Ministry of Health and Welfare; 2024. URL: https://www.mohw.go.kr/menu.es?mid=a10708010900 [accessed 2026-01-03]
45. Guay F, Vallerand RJ, Blanchard C. On the assessment of situational intrinsic and extrinsic motivation: the Situational Motivation Scale (SIMS). Motiv Emotion 2000;24(3):175-213. [doi: 10.1023/a:1005614228250]
46. Nocella G, Srinivasan C. Adherence to WHO's nutrition recommendations in the UK: dietary patterns and policy implications from a national survey. Food Policy 2019;86:101719. [doi: 10.1016/j.foodpol.2019.05.002]
47. Stevenson RJ. Psychological correlates of habitual diet in healthy adults. Psychol Bull 2017;143(1):53-90. [doi: 10.1037/bul0000065] [Medline: 27618545]
48. Young HM, Miyamoto S, Dharmar M, Tang-Feldman Y. Nurse coaching and mobile health compared with usual care to improve diabetes self-efficacy for persons with type 2 diabetes: randomized controlled trial. JMIR mHealth uHealth 2020;8(3):e16665 [FREE Full text] [doi: 10.2196/16665] [Medline: 32130184]
49. Wong AKC, Bayuo J, Wong FKY, Chow KKS, Wong SM, Lau ACK. The synergistic effect of nurse proactive phone calls with an mHealth app program on sustaining app usage: 3-arm randomized controlled trial. J Med Internet Res 2023;25:e43678 [FREE Full text] [doi: 10.2196/43678] [Medline: 37126378]
50. Intrinsic motivation inventory: constructing the IMI for your study internet. Center for Self-Determination Theory. Rochester (NY): University of Rochester; 2022. URL: https://selfdeterminationtheory.org/wp-content/uploads/2022/02/IMI_Complete.pdf [accessed 2025-12-20]
51. Khadoura K, Shakibazadeh E, Mansournia M, Aljeesh Y, Fotouhi A. Effectiveness of motivational interviewing on medication adherence among Palestinian hypertensive patients: a clustered randomized controlled trial. Eur J Cardiovasc Nurs 2021;20(5):411-420. [doi: 10.1093/eurjcn/zvaa015] [Medline: 34009313]
52. Ambeba E, Ye L, Sereika S, Styn M, Acharya S, Sevick M, et al. The use of mHealth to deliver tailored messages reduces reported energy and fat intake. J Cardiovasc Nurs 2015;30(1):35-43 [FREE Full text] [doi: 10.1097/JCN.0000000000000120] [Medline: 24434827]
53. Lee WW, Choi K, Yum RW, Yu DS, Chair S. Effectiveness of motivational interviewing on lifestyle modification and health outcomes of clients at risk or diagnosed with cardiovascular diseases: a systematic review. Int J Nurs Stud 2016;53:331-341. [doi: 10.1016/j.ijnurstu.2015.09.010] [Medline: 26493130]
54. Self-determination theory. Wikipedia. 2025. URL: https://en.wikipedia.org/wiki/Self-determination_theory [accessed 2025-12-20]
55. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
IMB: Information-Motivation-Behavioral Skills
MD: mean difference
mHealth: mobile health

Edited by L Buis; submitted 25.Apr.2025; peer-reviewed by P Pichayapinyo, H Sadeghsalehi; comments to author 24.Sep.2025; revised version received 16.Oct.2025; accepted 17.Dec.2025; published 15.Jan.2026.

Please cite as:
Choo J, Shin Y, Noh S, Lee J
Effectiveness of a Hybrid Community-Based Heart-Healthy Lifestyle Intervention: Three-Arm Randomized Controlled Trial Integrating mHealth and Motivational Interviewing
JMIR Mhealth Uhealth 2026;14:e76521
URL: https://mhealth.jmir.org/2026/1/e76521
doi:10.2196/76521
PMID:41407528

©Jina Choo, Yura Shin, Songwhi Noh, Juneyoung Lee. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 15.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete

XSL•FO
RenderX

bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Brief Web-Based and Mobile Intervention of Intermittent Fasting With Meal Support for Weight Loss Among Adults With Overweight and Obesity in Japan: Pilot Randomized Controlled Trial

Takashi Noda[1], MPH; Tomonari Shimamoto[1], DPH; Kosuke Kiyohara[2], DPH; Mie Imanaka[3], MPH, RD; Norihiro Nishioka[1], MD, PhD; Yukiko Tateyama[1], PhD; Taku Iwami[1], MD, PhD

[1]Department of Preventive Services, School of Public Health, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[2]Department of Food Science, Faculty of Home Economics, Otsuma Women's University, Tokyo, Japan

[3]Department of Health and Nutrition, Faculty of Nursing and Nutrition, The University of Shimane, Izumo, Japan

Corresponding Author:
Taku Iwami, MD, PhD
Department of Preventive Services
School of Public Health, Graduate School of Medicine
Kyoto University
Yoshida Konoe-cho
Sakyo-ku
Kyoto, 606-8315
Japan
Phone: 81 757534400
Email: iwami.taku.8w@kyoto-u.ac.jp

Abstract

Background: Intermittent fasting emerges as a promising dietary approach against obesity, offering a cost-effective strategy for implementation via web-based platforms. We developed a Brief Online Intermittent Fasting Program (OIF), featuring a self-administered, weekly 1-day fasting regimen with replacement meals delivery, online guidance, and app messaging to support adherence.

Objective: This pilot study aimed to assess the preliminary effectiveness, feasibility, and safety of the OIF on weight loss in adults with overweight and obesity in Japan. Secondary objectives were to assess its effects on body composition and metabolic markers.

Methods: This 12-week, 1:1 randomized controlled trial recruited adults with overweight and obesity (BMI from 23 to <35) in 1 university, 1 hospital, and 2 company offices. Participants were randomized into 2 groups stratified by sex and age (<40 or ≥40 years). The intervention group received very low-calorie (407 kcal) meal replacements for weekly intermittent fasting, online guidance via Zoom (Zoom Video Communications, Inc), and app messages encouraging fasting and healthy lifestyles. The control group received app messages promoting healthy lifestyles only. Interventions were administered by a nonblinded researcher. The primary outcome was the change in body weight after 12 weeks, analyzed using intention-to-treat principles and adjusted for sex, age, and baseline weight. Secondary outcomes encompassed body composition, blood pressure, biomarkers (eg, hemoglobin A_{1c}, triglycerides, and cholesterol), quality of life, physical activity, intervention adherence, and adverse events.

Results: A total of 57 individuals were enrolled (28 in the intervention group and 29 in the control group). At 12 weeks, 25 participants in the intervention group and 27 participants in the control group completed follow-up. The baseline median weight was 75.8 (IQR 68.3-80.6) kg for the intervention group and 74.8 (IQR 69.8-81.8) kg for the control group. The mean weight change was –0.9 (SD 1.9) kg in the intervention group and +0.6 (SD 1.4) kg in the control group. The adjusted between-group difference in weight change was statistically significant at –1.6 (95% CI –2.5 to –0.8) kg. Fat mass change was not statistically significant (–0.1, 95% CI –1.3 to 1.4 kg), but muscle mass reduction was implied (–1.3, 95% CI –2.5 to –0.2 kg). Intervention adherence was 79% (22/28) in the intervention group. No serious adverse events were reported, and there were no significant changes in key biomarkers, such as hemoglobin A_{1c} or quality of life.

Conclusions: The OIF demonstrated effectiveness in promoting modest weight loss among adults with overweight and obesity over 12 weeks, with high feasibility and safety indicated by low dropout rates and absence of serious adverse events. However,

XSL•FO
RenderX

the observed reduction in muscle mass indicates a need for program refinement, such as incorporating exercise guidance, to optimize health outcomes.

Trial Registration: UMIN-CTR UMIN000050437; https://tinyurl.com/4x5h2t2x

(JMIR Mhealth Uhealth 2026;14:e58930) doi:10.2196/58930

KEYWORDS

intermittent fasting; eHealth; mHealth; mobile health; weight loss; diet; overweight; obesity; randomized controlled trial

Introduction

Obesity is a well-known public health challenge, which increases the risk of cardiovascular diseases and mortality [1]. According to the World Health Organization, at least 2.8 million people die each year as a result of being overweight or obese [2], with the obesity epidemic still escalating at an alarming rate [3]. Moreover, the recent COVID-19 pandemic has further deteriorated our lifestyles and made the development of remote obesity interventions, using the internet, apps, and other web-based technologies, an even more urgent issue [4-6]. Although many web-based programs have been developed, with some degree of effectiveness acknowledged [7,8], few have been successfully implemented in society to yield significant results. Also, criticism exists toward the previous clinical trial approaches, which have focused too much on internal validity [9], indicating a demand for the development of highly feasible online weight loss programs that are easily applied and disseminated into society.

Recently, intermittent fasting (IF) has emerged as a popular strategy for combating obesity [10]. IF is a simple dietary restriction regimen that incorporates periodic fasting (energy restriction of less than 600 kcal) into normal eating habits [11,12]. Recent systematic reviews and meta-analyses have demonstrated the effectiveness of IF for weight loss compared to ad libitum eating [13]. The mechanisms behind weight loss with IF are related to energy restriction, increased fat metabolism, and improved glucose metabolism [14], which are common to any dietary restriction. IF may also provide cardiovascular health benefits. For example, studies with women with obesity show that a weekly IF intervention over 2 months significantly reduced resting heart rate, blood glucose, insulin, and homocysteine levels. These effects are likely associated with reductions in adipokines, including leptin, IL-6, tumor necrosis factor-alpha, and insulin-like growth factor-1, suggesting that IF could mitigate cardiovascular disease risk [15,16]. Additional benefits reported in recent studies include decreased hunger, reduced inflammation, improved cellular repair, regulated hormone production, enhanced immune function, adjusted circadian rhythms, and a more diverse gut microbiome [14,17-21]. Therefore, IF holds significant potential as a promising nonpharmacological approach to improving health at the population level [18]. We particularly focused on the simplicity and feasibility of IF, because a truly successful weight loss approach needs to be simple and sustainable, capable of being incorporated into daily life as lifelong dietary customs [18,21]. Since IF does not require detailed examination of food content or energy calculations, it is easy to implement with brief web-based guidance, having a high possibility of being applied as a cost-effective population approach. However, to the best of our knowledge, no studies have implemented IF as a web-based intervention, leaving a potentially significant solution to obesity unexplored.

To assess the effect of web-based IF administration, we have developed the "Brief Online Intermittent Fasting Program (OIF)," which features self-led fasting practice supported with meal delivery, online guidance, and app messages. The purposes of this pilot study were to evaluate the preliminary effectiveness, feasibility, and safety of the 12-week OIF on weight loss in adults with overweight and obesity in Japan. Secondary objectives were to assess its effects on body composition and metabolic markers.

Methods

Study Design

This pilot study is a 1:1 parallel-group randomized controlled trial (RCT) comparing the intervention group following the OIF with the control group receiving minimal care.

Participants and Setting

Eligibility Criteria

In order to benefit from weight reduction and ensure the safe and stable administration of the OIF, the following eligibility criteria were established.

The inclusion criteria are as follows: individuals with overweight or obesity and with a BMI between 23 and under 35, adults aged 20 years to less than 65 years at the time of consent, those capable of communicating in Japanese, owners of a smartphone with iOS or Android, those who were familiar with using apps on a smartphone and using email and Zoom (Zoom Video Communications, Inc) on either a smartphone or a personal computer, those willing to visit one of the research facilities twice, and consenting to provide blood samples through self-puncture at the measurements. We defined overweight as having a BMI of 23 or above, while the global standard for overweight is a BMI of 25 or above. This adjustment aligns with guidelines from the National Institute for Health and Care Excellence, which recommend lowering the overweight threshold to 23 for Asian populations to enhance metabolic disease prevention in populations with lower obesity prevalence [22].

The exclusion criteria include any of the following: individuals with a history of heart disease, kidney disease, psychiatric disorders, or any serious medical conditions, individuals using diabetes medication (to minimize the risk of hypoglycemia), smokers or heavy drinkers (consuming alcohol on average more

XSL•FO
RenderX

than 4 drinks per day or 14 drinks per week), night shift workers, individuals who may face significant environmental changes during the trial period, individuals with eating disorders, food allergies or alcohol dependence who require specific dietary supervision, women who are breastfeeding, pregnant, or planning to become pregnant during the study, lean individuals with a body fat percentage below 10% for men and 20% for women, athletes engaging in more than 12 hours of training or sports practice per week or those belonging to university sports clubs or professional teams aiming for competition, and any other individuals deemed unsuitable for participation by the researchers.

Setting and Recruitment Procedure

Participant recruitment, measurement, and registration were conducted at Kyoto University and 3 collaborating facilities in Japan: Kusaka Hospital, Cold Storage Japan Inc, and Buddy Training Co, Ltd. We recruited participants through our website (Department of Preventive Services at School of Public Health, Kyoto University), our X (formerly Twitter; X Corp), posters and flyers, and the OReC—a web platform dedicated to research recruitment [23]. Further details on the recruitment procedure and materials are presented in Multimedia Appendix 1.

Sample Size

As this is a pilot RCT, a formal power-based sample size calculation for definitive hypothesis testing was not the primary objective. The main goals were to obtain estimates of effect size to inform the design of a future, larger trial. Based on recommendations for pilot study sample sizes [24], a range of 10-30 participants per group is often considered sufficient for these purposes. To ensure our sample size was within a reasonable range, we referred to previous literature [25,26] and pre-estimated the effect size to be approximately Cohen d=0.5. A sample of 10 participants per group would provide a preliminary indication of such an effect. To account for the potentially high dropout rate common in web-based intervention trials, we determined our target sample size to be 20 participants per group, for a total of 40 participants.

Randomization

Allocation

The researcher (TN) weekly sent the information (ID, sex, and age) of the registered participants to another researcher (KK) via email. He conducted a 1:1 stratified randomization using computer-generated random sequences, with sex (male or female) and age (<40 or ≥40 years) as stratifying factors. The block sizes were randomly chosen to be either 2 or 4. Allocation concealment was maintained throughout the study.

Blinding

Participants were not blinded due to the nature of the dietary intervention. The implementer of the intervention was also not blinded since it was carried out by the researcher (TN), who knew the allocation results. Although the messages sent via app were also administered by TN and not blinded, the timing of delivery and the number of characters in the messages were standardized across both groups. However, to minimize potential bias, outcome assessors and a statistician analyst were blinded, as suggested by best practices in behavioral studies [27].

Intervention

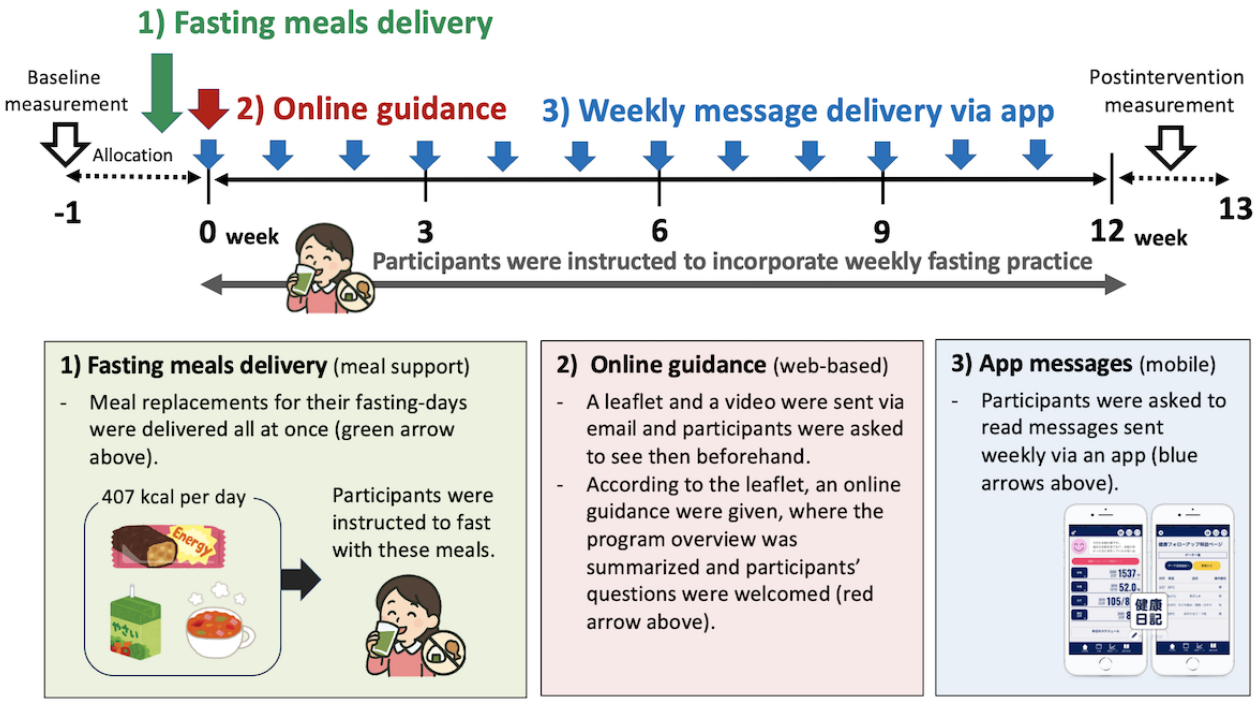
A nonblinded researcher (TN) administered all the interventions for both groups. Detailed information on interventions was found in Multimedia Appendix 2.

Intervention Group (OIF)

Participants in the intervention group were engaged in the OIF (Figure 1). In this study, fasting was defined as consuming the provided fasting meals by 8 PM without any additional food intake. The OIF comprises three components: (1) meal support (nonweb-based component), (2) a web-based component, and (3) a mobile component. Further details are described in Multimedia Appendix 2.

XSL•FO
RenderX

Figure 1. Overview of the 12-week "Brief Online Intermittent Fasting Program (OIF).".



1. Fasting meals delivery: After the allocation, a 12-day supply of fasting meals was delivered to the intervention group in one batch. Fasting meals consisted of commercially available products that were low in energy, low in saturated fat, and plant-based and were constructed based on previous studies [28,29]. Briefly, it included 1 bottle of tomato juice, 2 bottles of fruit and vegetable juice, 2 servings of instant miso soup, 1 soy bar, and 1 multivitamin-mineral supplement, totaling 407 kcal per day, with 13% protein, 26% fat, and 61% carbohydrates. We developed this 12-week regimen in order to theoretically promote weight loss of 2.4-2.7 kg, which corresponds to a 3% loss of initial weight for a potential participant with a weight of 80 kg and 90 kg, respectively.
2. Online fasting guidance: We sent participants a leaflet detailing OIF (an A4 sheet, double-sided) via email upon allocation. They were directed to watch the prerecorded video (about 9 minutes) before the online guidance and then attended the guidance session via Zoom, which was held on their first day of the trial period and averaged 3.5 (SD 4.1) minutes per participant. They were instructed to replace one day's meals with the provided fasting meals, with a recommendation to fast on a holiday to minimize their burden.
3. Weekly message delivery via app: Participants were instructed to install an app, "Kenko-Nikki (Healthtech Laboratory [HTL], Inc)" [30], detailed in Multimedia Appendix 3. Participants received weekly messages in Japanese (mean 382, SD 126 characters) containing general health advice on diet, exercise, sleep, and encouragement and reminders for their fasting practice through the app.

Control Group—Minimal Care

Similar to the intervention group, participants were directed to read weekly messages containing general health advice on diet, exercise, and sleep. The content of these messages differed partly from those sent to the intervention group, as outlined in Multimedia Appendix 2.

Outcome Definition and Measurement

Data Collection

Overview

Sex, age, medical history, current medical visits, and medication status were collected via a web form at the time of preregistration. This information was reverified at the baseline measurement, along with identity confirmation. The following assessment data were measured at 2 time points—baseline and after 12 weeks, with the exception of adverse events and mild symptoms, adherence, and participants' feedback, which were measured once at 12 weeks, and of self-recorded weight and step counts and fasting status, which were supposed to be measured every day.

Anthropometric Measures

Body weight, fat mass, muscle mass, and basal metabolic rate (BMR) were measured by using the body composition analyzer (MC-780A-N; TANITA Corporation). Height was measured 3 times at the baseline, barefoot, using a portable height measure (Seca 213 l; Seca Japan), and the median value was used. The BMI was calculated based on these data.

Blood Pressure

Measurements of systolic blood pressure (SBP) and diastolic blood pressure (DBP) were taken twice on the upper arm opposite the dominant hand using a blood pressure monitor

XSL•FO
RenderX

(HEM–7511T; OMRON Corporation) while the participant was seated. The average value of these two measurements was used.

Blood Biomarkers

Hemoglobin A_{1c} (HbA_{1c}), triglycerides, total cholesterol, high-density-lipoprotein cholesterol (HDL-C), and low-density-lipoprotein cholesterol (LDL-C) were measured using a point-of-care biochemical analysis device (cobas b 101 plus; Roche Diagnostics KK). Blood sampling involved participants self-puncturing the fingertip of the middle finger on the opposite side of their dominant hand using a spring-loaded lancet (Nipro LS lancet 30G 1.0 mm; Nipro Corporation). Subsequently, research staff aspirated the blood onto the reagent disk for analysis in the device.

Quality of Life

Quality of life (QOL) scores were also evaluated, which were measured using an electronic scoring system for the Short-Form 12-Item Survey-version 2 (SF-12v2) Japanese version [31] (Qualitest). Participants responded to questions on the web system regarding their lives over the past month via their smartphones. These responses were then calculated through the system to obtain physical and mental 2-component summary scores, which are standardized T-scores with a mean of 50 (SD 10), where higher scores indicate better health.

To assess participants' subjective health status, responses to the first question of the SF-12v2 questionnaire, "In general, would you say your health is:," were further analyzed. Response options included "(1) Excellent," (2) Very good," (3) Good," (4) Unsatisfactory," and "(5) Poor."

Physical Activity

Participants' physical activities were self-reported using the short form of the International Physical Activity Questionnaire [32]. The questions were prepared on a Google Form, and participants were asked to respond via smartphone about their physical activity over the past month at the time of final measurement. The amount of total physical activity was calculated as metabolic equivalent of task-minutes per week (METs).

Adverse Events and Mild Symptoms

For the evaluation of the safety, the number of adverse events and the number of individuals experiencing them were investigated. Participants were instructed to consult anytime throughout the study regarding any adverse events or symptoms. In addition, they were asked at the 12-week measurement about the incidence of 20 symptoms prelisted from a prior study [33] (sleep difficulty, hunger, fatigue, headache, diarrhea, sensitivity to cold, dry mouth, back pain, bad breath, muscle pain, abdominal bloating, cravings, vertigo, blurred vision, restless leg, skin rash, nausea, palpitation, dyspepsia, and muscular cramp) via a web form. We primarily evaluated the difference in the number of these 20 events and the number of participants who reportedly experienced them. Other adverse events were also self-reported separately through a free-text web questionnaire at the measurement at 12 weeks.

Adherence

Adherence in the intervention group was defined by the number of actual fasts, based on self-report via a web questionnaire at the 12-week measurement. Adherence in the control group was defined by the read status of messages on the app, with data obtained from the server. The definition of adherence is detailed in the statistical analysis section.

Participants' Feedback

Participants' feedback on each intervention was collected through a web questionnaire at the 12-week measurement.

Self-Recorded Weight and Step Counts

Participants were directed to record their self-measured weight and step counts via the app every day throughout the trial, which was predetermined to be assessed only exploratorily.

Fasting Status

Participants in the intervention group were directed to record the following information on their fasting day: the date, the consumption status of each fasting meal (the extent to which they ate and whether they ate other food or not), and the planned date of the next fasting day. This was also to be exploratorily assessed.

Outcome

The primary outcome was the change in body weight. The changes in BMI, fat mass, muscle mass, BMR, SBP, DBP, HbA_{1c}, triglyceride, HDL-C, and LDL-C were evaluated as secondary outcomes. After formal analysis execution, the number of participants achieving a minimal clinically important change (MIC) was added as an exploratory outcome for post hoc analysis.

Statistical Analysis

Overview

The statistical analysis plan (Multimedia Appendix 4) was fixed before data locking. After finalizing the data, the precoded analysis script (R [R Foundation for Statistical Computing], version 4.3.2) was executed. Statistical estimations were presented with 95% CI. Our predetermined α level was .05. Missing data in outcomes were imputed using the Baseline Observation Carried Forward (BOCF) approach. As this was a pilot study, no formal interim analyses or prespecified stopping guidelines were planned.

Baseline Analysis and Definition of Analysis Population

Baseline characteristics were presented as the number and percentage for sex (male), while continuous variables were presented as median and IQR.

Under the intention-to-treat (ITT) principle, all the allocated participants were regarded as the primary analysis population (ITT).

The per protocol set (PPS) included individuals in the intervention group who adhered to the fasting regimen (reportedly had fasted as instructed at least 10 out of 12 times) and participants in the control group who adhered to the minimal care (had read at least 10 out of the 12 app messages, as

XSL•FO
RenderX

confirmed by server data). Additionally, within the intervention group, participants who reportedly had fasted 10 or more times out of 12 in accordance with our instructions were defined as "adherent," while those who reported fasting 12 times were categorized as "fully adherent."

Safety analyses were conducted on the completers set (CS), defined as participants who were followed up at the 12-week measurement.

In the discussion, we also presented the mean (SD) values of QOL scores for considering generalizability. Additionally, we assessed the subjective health status of participants by analyzing responses to the SF-12v2 questionnaire and comparing them with those from a similar question in a national annual survey conducted in 2022 [34]. We compare the proportions of participants who self-reported positive health status in the SF-12v2 questionnaire's first question, "In general, would you say your health is:" with those from the reference survey, which asked, "How do you rate your current health status?" with similar responses ranging from 1 to 5. In the SF-12v2, the response options were "(1) Excellent, (2) Very good, (3) Good, (4) Unsatisfactory, (5) Poor," while in the reference survey, they were "(1) Good, (2) Satisfactory, (3) Average, (4) Unsatisfactory, (5) Poor." Recognizing the differences in response expression between the 2 surveys, we focused on respondents who considered themselves healthy, combining responses categorized as 1-3. This comparison was predetermined in the protocol.

Analysis of Primary Outcome

The change in body weight was presented as the mean and SD for each group. The primary analysis was to estimate the between-group mean difference under the ITT principle, adjusted for sex, age, and baseline weight using a generalized linear regression model, presented along with a 95% CI and *P* value. The regression coefficients for each covariate were also reported with their 95% CIs. Sensitivity analyses included unadjusted comparisons of crude values (2-tailed Welch *t* test) and per-protocol analysis within the PPS (using the same adjustment as the primary analysis). Additionally, in case of missing weight data, an analysis incorporating the worst-case scenario (imputing the maximum weight gain for the intervention group and the maximum weight loss for the control group from observed data) was predetermined to be added to the sensitivity analysis.

Analysis of Secondary Outcomes

For continuous secondary outcomes, crude (unadjusted) between-group mean differences and 95% CIs were calculated using the Welch *t* test. Total physical activity was also assessed to identify within-group differences, using the Wilcoxon signed-rank test for paired comparisons. In the safety analysis, the number of adverse events and the number of individuals experiencing them were investigated, and the between-group differences in incidence proportions were analyzed by the chi-square test. Additionally, within-group differences of QOL were analyzed by the Wilcoxon signed-rank test. The proportion of follow-up rate, adherence rate, and message-read rate was presented as percentages of the ITT population.

Exploratory Analysis

Self-recorded weight and step counts were predetermined to be assessed for discussion as needed, in cases of high dropout or missing data, for instance. As a post hoc analysis for the primary outcome, the number of participants achieving a MIC was compared between groups, with *P* values reported from the Fisher exact test. In addition, we included an additional sensitivity analysis for CS, which only analyzed the participants who completed the 12-week follow-up.

Methods Alteration After Study Commencement

After the recruitment began, 3 collaborative facilities were added, and the initial eligibility criterion of having a BMI between 25 and less than 35 was modified to a BMI between 23 and less than 35. This change required a revision of the protocol and a subsequent review and reapproval by the Ethics Committee. Aspects of the statistical analysis not explicitly outlined in the original protocol were detailed in the statistical analysis plan (Multimedia Appendix 4), with its final version being fixed prior to the data lock.

Ethical Considerations

This study was conducted with the approval of the Kyoto University Graduate School and Faculty of Medicine Ethics Committee and under the authorization of the dean (approval number C1625). The study was registered in the University Hospital Medical Information Network-Clinical Trials Registry (UMIN-CTR), a clinical trial registration service that meets the standards of the International Committee of Medical Journal Editors, on February 27, 2023 (UMIN000050437). Written informed consent was obtained from all participants. To protect privacy, participant data were pseudonymized. The correspondence table linking personal information to the data was stored on a separate USB drive and kept in a locked cabinet within a locked room. This table was deleted after the analysis to prevent reidentification. Participants received a 5000 Japanese yen Amazon gift card as compensation. Other ethical considerations and the consent document were described in Multimedia Appendix 1. This trial was conducted in accordance with the CONSORT (Consolidated Standards of Reporting Trials) 2025 and CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines (Multimedia Appendix 5). There was no patient or public involvement in the design, conduct, or reporting of the study.

Results

Participants Flow

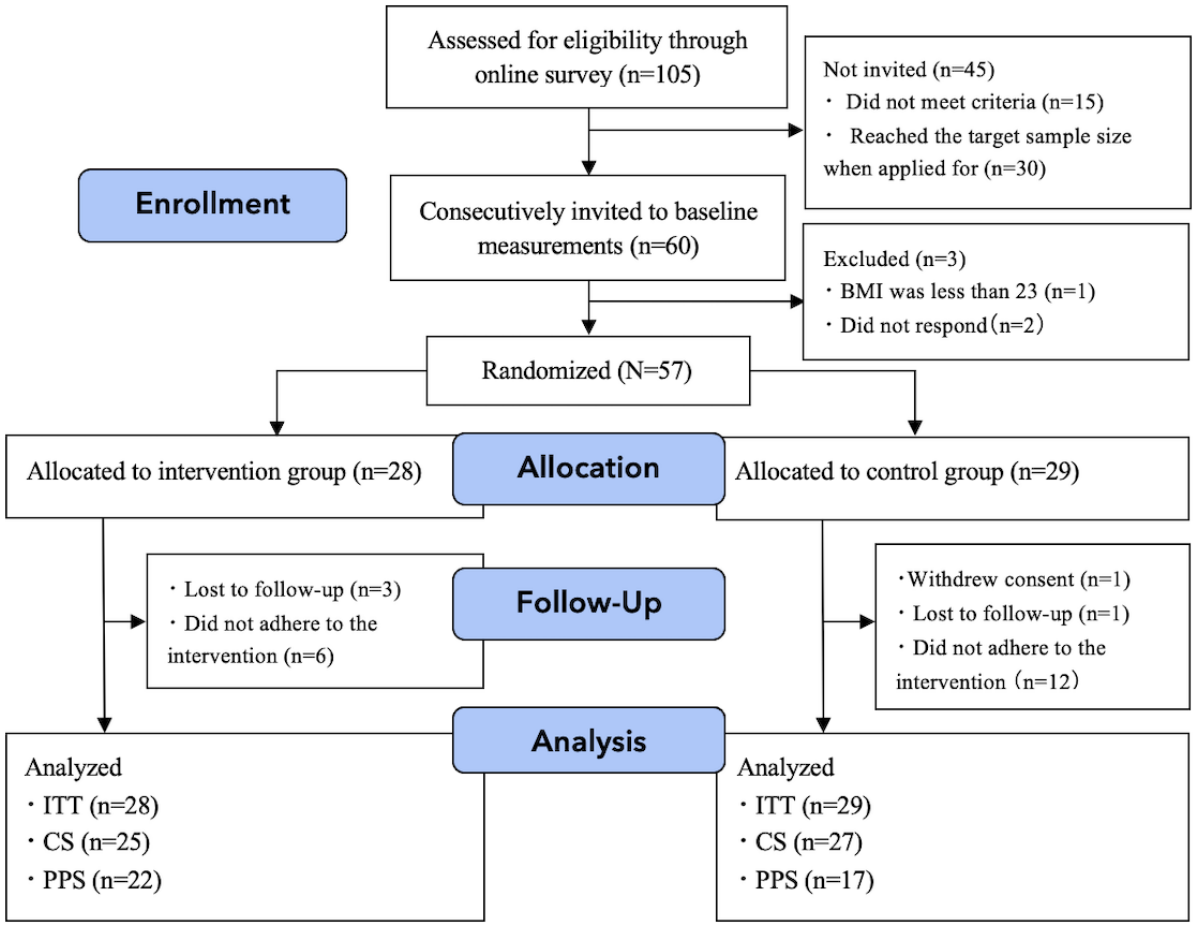
Participant recruitment took place from July 20, 2023, to October 10, 2023. The follow-up measurement of all the participants ended on January 12, 2024.

Among the 60 individuals invited to the baseline measurement, 57 were eventually registered, with 28 allocated to the intervention group and 29 to the control group (Figure 2). In the intervention group, 3 participants were lost to follow-up (dropped out) due to discontinued contact. In the control group, a participant withdrew consent due to "difficulties in continuing

XSL•FO
RenderX

participation," but agreed to our using his baseline data, allowing us to include him in the ITT population. Another participant was lost to follow-up (dropped out) in the control group. Consequently, of the 57 participants in the ITT population, 89% (25/28) in the intervention group and 93% (27/29) in the control group were followed up (defined together as the CS). The PPS was defined based on adherence, with 22 (vs 17) participants in the intervention (vs control) group. Among the 28 participants in the intervention group, 22 (79%) participants were defined as "adherent" (those who fasted 10 times or more; included in PPS), while 19 (68%) were defined as "fully adherent" (those who fasted 12 times). Among the 29 participants in the control group, 17 (59%) were adherent (who had read 10 app messages or more) and thus were included in PPS. Similarly, 15 (54%) participants in the intervention group had read app messages.

Figure 2. Participants flowchart. CS: completers set; ITT: intention-to-treat; PPS: per protocol set.



Participants Characteristics

Table 1 presents the baseline characteristics of participants in each group. The proportion of men was high in both groups (19/28, 68% in the intervention group vs 19/29, 66% in the control group). The median age was 34 (IQR 29-47) vs 38 (IQR 29-52) years, the median body weight was 75.8 (IQR 68.3-80.6) vs 74.8 (IQR 69.8-81.8) kg, and the median BMI was 27.1 (IQR 24.8-29.8) vs 26.6 (IQR 24.7-28.5) for the intervention (vs control) group. The number of registered participants by facility was as follows: 22 (12 vs 10) participants at Kyoto University, 11 (7 vs 4) participants at Kusaka Hospital, 10 (4 vs 6) participants at Cold Storage Japan Inc, and 14 (5 vs 9) participants at Buddy Training Co, Ltd. Significant imbalances were not observed between the 2 groups across all characteristics.

XSL•FO
RenderX

Table 1. Baseline characteristics of the intervention and control groups (N=57)a.

Characteristic	Intervention (n=28)	Control (n=29)
Sex (male), n (%)	19 (68)	19 (66)
Age (years), median (IQR)	34 (29-47)	38 (29-52)
Weight (kg), median (IQR)	75.8 (68.3-80.6)	74.8 (69.8-81.8)
BMI (kg/cm^2), median (IQR)	27.1 (24.8-29.8)	26.6 (24.7-28.5)
Height (cm), median (IQR)	168 (163-171)	170 (163-175)
Fat mass (kg), median (IQR)	22 (18-28)	22 (18-29)
Muscle mass (kg), median (IQR)	52 (42-56)	52 (41-57)
BMR[b] (kcal), median (IQR)	1529 (1382-1682)	1563 (1369-1703)
SBP[c] (mm Hg), median (IQR)	128 (118-141)	130 (117-142)
DBP[d] (mm Hg), median (IQR)	87 (79-96)	90 (80-99)
HbA_{1c}[e] (%), median (IQR)	5.4 (5.2-5.6)	5.4 (5.2-5.5)
TG[f] (mg/dL), median (IQR)	107 (70-209)	131 (111-208)
Total-C[g] (mg/dL), median (IQR)	187 (165-212)	196 (176-220)
HDL-C[h] (mg/dL), median (IQR)	57 (45-65)	57 (48-66)
LDL-C[i] (mg/dL), median (IQR)	103 (89-122)	112 (80-128)
Total physical activity (METs[j]), median (IQR)	999 (399-1706)	1074 (827-1497)
QOL[k] score[l], median (IQR)		
Physical component summary	50 (47-52)	52 (48-54)
Mental component summary	59 (54-63)	55 (49-58)
Subjective health status[m], n (%)		
Excellent	3 (11)	3 (10)
Very good	9 (32)	5 (17)
Good	15 (54)	15 (52)
Fair	0 (0)	5 (17)
Poor	1 (4)	1 (3)
Registered facility, n (%)		
Kyoto University	12 (43)	10 (35)
Kusaka Hospital	7 (25)	4 (14)
Cold Storage Japan Inc	4 (14)	6 (21)
Buddy Training Co, Ltd.	5 (18)	9 (31)

[a]Data were presented as numbers for sex (male) and registered facilities, and as medians (IQR) for all other variables. The values were rounded for all the variables except for weight, BMI, and hemoglobin A_{1c}.

[b]BMR: basal metabolic rate.

[c]SBP: systolic blood pressure.

[d]DBP: diastolic blood pressure.

[e]HbA_{1c}: hemoglobin $A_{1c.}$

[f]TG: triglyceride.

[g]Total-C: total cholesterol.

[h]HDL-C: high-density-lipoprotein cholesterol.

[i]LDL-C: low-density-lipoprotein cholesterol.

[j]METs: metabolic equivalent of task-minutes per week.

[k]QOL: quality of life.

XSL•FO
RenderX

[l]SF-12v2 (Short-Form 12-Item Survey-version 2) Japanese version was used to obtain each 2-component summary score, which is standardized T-scores with a mean of 50 (SD 10), where higher scores indicate better health.

[m]Calculated from the responses to the first question of the SF-12v2 questionnaire, "In general, would you say your health is:" with options ranging from (1) Excellent, (2) Very good, (3) Good, (4) Unsatisfactory, and (5) Poor.

Changes in Primary Outcome

Primary Analysis

Table 2 shows the primary results of this study. After 12 weeks, the mean weight change in the intervention group was –0.9 (SD 1.9) kg, while it was 0.6 (SD 1.4) kg in the control group. The primary analysis revealed an adjusted mean difference of –1.6 (95% CI –2.5 to –0.8; *P*<.001) kg. The regression coefficients for covariates were as follows: baseline weight=–0.2 (95% CI –0.6 to 0.2) kg per 10-kg increase; sex (male)=0.5 (95% CI –0.5 to 1.4) kg; and age=–0.4 (95% CI –0.7 to 0.0) kg per 10-year increase.

Table 2. The effect of the intervention between the intervention and control groups (N=57)a.

Variable	Intervention (n=28)	Control (n=29)	Adjusted mean difference (95% CI)[b]	*P* value
Weight at baseline (kg), median (IQR)	75.8 (68.3-80.6)	74.8 (69.8-81.8)	—[c]	—
Weight at 12 weeks (kg), median (IQR)	74.5 (66.3-79.5)	74.8 (69.1-83.1)	—	—
Weight change (kg), mean (SD)	–0.9 (1.9)	0.6 (1.4)	–1.6 (–2.5 to –0.8)[b]	<.001

[a]Intention-to-treat analysis. Data were presented as median (IQR) for each measured weight and as mean (SD) for the weight change.

[b]Adjusted for stratifying factors (age and sex) and baseline weight by using the generalized linear regression model.

[c]Not applicable.

Sensitivity Analysis

Among the sensitivity analyses, the crude comparison showed a mean difference of –1.5 (95% CI –2.4 to –0.7; *P*=.001) kg, and the per-protocol analysis revealed a mean difference of –1.8 (95% CI –2.9 to –0.6; *P*=.004) kg. Additional sensitivity analysis for CS presented similar results; a mean difference of –1.8 (95% CI –2.7 to –0.9; *P*=.004) kg. In a worst-case scenario, where missing data were imputed (+3.3 kg to the weight change for the 3 missing participants in the intervention group and –4.7 kg to the weight change for the 2 missing in the control group), the adjusted mean difference was –1.0 (95% CI –2.1 to 0.2; *P*=.10) kg, and the crude mean difference was –0.9 (95% CI –2.0 to 0.3; *P*=.13) kg.

Exploratory Analysis

The post hoc analysis revealed that more participants (8/28, 29%; *P*=.01) achieved a MIC (3% reduction in initial weight) in the intervention group compared to the control group (1/29, 3%).

Changes in Secondary Outcomes

Changes in Continuous Variables

The results for secondary outcomes were presented in Table 3, showing the within-group changes and between-group differences. In line with weight changes, the BMI decreased more in the intervention group compared to the control group (mean difference=–0.4, 95% CI –0.1 to –0.8 kg). However, no significant difference was observed in fat mass changes (mean difference=–0.1, 95% CI –1.3 to 1.4 kg), whereas a decrease in muscle mass was implied (mean difference=–1.3, 95% CI –2.5 to –0.2 kg). In alignment with the changes in body composition, a decrease was suggested in the BMR (mean difference=–37.8, 95% CI –67.5 to –8.2 kcal). No significant differences were observed in other outcomes. The mean differences (95% CI) for each outcome were as follows: SBP= –4.6 (–10.6 to 1.3) mm Hg, DBP=–1.1 (–5.6 to 3.3) mm Hg, HbA_{1c}=–0.0 (–0.2 to 0.1) %, TG=–30.8 (–71.3 to 9.7) mg/dL, total cholesterol=–8.2 (–22.7 to 6.4) mg/dL, HDL-C=2.2 (–2.7 to 7.1) mg/dL, LDL-C=–4.1 (–17.1 to 8.9) mg/dL, physical QOL=0.6 (–2.6 to 3.9), mental QOL=–0.7 (–4.0 to 2.3), subjective health status=–0.2 (–0.5 to 0.3), and total physical activity=291 (–508 to 1090) METs.

XSL•FO
RenderX

Table 3. The changes in secondary outcomes from baseline to 3 months. (N=57)a.

Variable	Intervention (n=28), mean (SD)	Control (n=29), mean (SD)	Between-group difference[b], mean difference (95% CI)
BMI (kg/cm^2)	–0.3 (0.7)	0.2 (0.5)	–0.4 (–0.8 to –0.1)
Fat mass (kg)	–0.2 (1.7)	–0.3 (3.1)	0.1 (–1.3 to 1.4)
Muscle mass (kg)	–0.6 (1.0)	0.8 (2.8)	–1.3 (–2.5 to –0.2)
BMR[c] (kcal)	–17.9 (31.9)	19.9 (72.0)	–37.8 (–67.5 to –8.2)
SBP[d] (mm Hg)	–3.3 (9.4)	1.3 (12.7)	–4.6 (–10.6 to 1.3)
DBP[e] (mm Hg)	–3.6 (6.6)	–2.4 (9.8)	–1.1 (–5.6 to 3.3)
HbA_{1c}[f] (%)	–0.1 (0.2)	–0.0 (0.3)	0.0 (–0.2 to 0.1)
TG[g] (mg/dL)	1.1 (64.8)	31.9 (86.4)	–30.8 (–71.3 to 9.7)
Total-C[h] (mg/dL)	3.4 (24.4)	11.6 (30.2)	–8.2 (–22.7 to 6.4)
HDL-C[i] (mg/dL)	2.4 (7.4)	0.2 (10.8)	2.2 (–71.3 to 9.7)
LDL-C[j] (mg/dL)	0.9 (25.3)	5.0 (23.7)	–4.1 (–17.1 to 8.9)
QOL[k] score[l]			
Physical component summary	0.5 (5.8)	–0.1 (6.3)	0.6 (–2.6 to 3.9)
Mental component summary	–0.5 (6.8)	0.2 (4.6)	–0.7 (–4.0 to 2.3)
Subjective health status[m]	–0.3 (0.6)	–0.1 (0.9)	–0.2 (–0.5 to 0.3)
Total physical activity (METs[n])	350 (1679)	59 (1295)	291 (–508 to 1090)

[a]Intention-to-treat analysis. Data were presented as mean (SD) for the change in each outcome and as mean difference (95% CI) for the difference between groups in the change scores.

[b]The difference in secondary outcomes between groups was not adjusted and was presented as crude values, with 95% CI of the Welch *t* test.

[c]BMR: basal metabolic rate.

[d]SBP: systolic blood pressure.

[e]DBP: diastolic blood pressure.

[f]HbA_{1c}: hemoglobin $A_{1c.}$

[g]TG: triglyceride.

[h]Total-C: total cholesterol.

[i]HDL-C: high-density-lipoprotein cholesterol.

[j]LDL-C: low-density-lipoprotein cholesterol.

[k]QOL: quality of life.

[l]SF-12v2 (Short-Form 12-Item Survey-version 2) Japanese version was used to obtain each 2-component summary score, which is standardized T-scores with a mean of 50 (SD 10), where higher scores indicate better health.

[m]Calculated from the responses to the first question of the SF-12v2 questionnaire, “In general, would you say your health is:” with options ranging from (1) Excellent, (2) Very good, (3) Good, (4) Unsatisfactory, and (5) Poor.

[n]METs: metabolic equivalents-minutes per week.

Safety Analysis

Table 4 displays the reported number of events and the incidence—the number of participants who experienced it—for each of the 20 mild symptoms. A larger proportion of participants reported hunger in the intervention group (22/25, 89%; *P*<.001) compared to the control group (3/27, 11%). Additionally, fatigue (16/25, 64% vs 6/27, 22%; *P*=.06) and cravings (15/25, 60% vs 3/27, 11%; *P*=.01) were also more reported in the intervention group (vs control group). On the other hand, the participants in the control group reported hunger (43 vs 27) and cravings (70 vs 42) more frequently compared to the intervention group. Adverse events reported in free text included influenza-like illness (1/25 vs 3/27), sinusitis (0/25 vs 1/27), viral gastroenteritis (1/25 vs 0/27), loss of appetite (1/25 vs 3/27), and drowsiness (2/25 vs 0/27). Regarding the within-group difference in QOL scores, no significant changes were observed before and after in either group: physical QOL (intervention: mean 0.51, SD 5.84 vs control: mean –0.13, SD 6.27), and the mental QOL (mean –0.54, SD 6.84 vs mean 0.15, SD 4.60).

XSL•FO
RenderX

Table 4. Self-reported mild symptoms in the intervention and control groups (N=52)a.

Mild symptoms	Intervention (n=25)		Control (n=27)		*P* value[b]
	Events	Incidence (%)	Events	Incidence (%)	
Hunger	27	22 (88)	43	3 (11)	<.001
Fatigue	187	16 (64)	27	6 (22)	.01
Headache	125	12 (48)	74	7 (25)	.17
Cravings	42	15 (60)	70	3 (11)	.001
Diarrhea	30	5 (20)	68	7 (25)	.86
Dry mouth	118	8 (32)	68	4 (14)	.25
Back pain	66	6 (24)	21	6 (22)	>.99
Sensitivity to cold	36	5 (20)	25	6 (22)	>.99
Muscle pain	24	4 (16)	4	5 (18)	>.99
Abdominal bloating	9	4 (16)	14	5 (18)	>.99
Vertigo	17	4 (16)	9	5 (18)	>.99
Sleep difficulty	77	4 (16)	24	4 (14)	>.99
Dyspepsia	13	2 (8)	57	5 (18)	.48
Bad breath	7	4 (16)	12	2 (7)	.59
Blurred vision	6	2 (8)	0	4 (14)	.74
Nausea	0	3 (12)	13	2 (7)	.93
Muscular cramp	9	3 (12)	8	2 (7)	.93
Palpitation	1	1 (4)	7	3 (11)	.66
Skin rash	22	0 (0)	21	3 (11)	.26
Restless leg	4	2 (8)	2	0 (0)	.44

[a]Analysis population was the completers set. Participants were asked to report the number of incidences of 20 mild symptoms listed in the web questionnaire by answering 0-90 times per symptom.

[b]*P* values were presented for the incidences.

Discussion

Principal Findings

This study aimed to evaluate the effectiveness, feasibility, and safety of a 12-week OIF among adults with overweight and obesity in Japan. Overall, the program demonstrated effectiveness in promoting modest weight loss over 12 weeks, with high feasibility and safety indicated by low dropout rates and absence of serious adverse events.

The participants in the intervention group tended to lose weight (mean –0.9, SD 1.9 kg), whereas the control group, on average, gained weight (mean 0.6, SD 1.4 kg). Generally, individuals with overweight tend to continue to gain weight [35], but the result suggests this increase can be mitigated by the OIF. The adjusted mean difference was –1.6 (95% CI –2.5 to –0.8) kg, and more participants achieved a MIC in the intervention group compared to the control group (8/28 vs 1/29 participants). Among the sensitivity analyses, the per-protocol analysis showed a subtle larger effect (mean difference –1.8, 95% CI –2.9 to –0.6 kg), implying the efficacy of the OIF. The effect size calculated by the weight change (SD) was large (Cohen *d*=0.95), but the effect size based on the after-weight (SD) was small (Cohen *d*=0.19).

The dropout rate was low in the intervention group (3/28, 11%). Additionally, adherence rate exceeded our expectations, with 79% (22/28) of “adherent” and 68% (17/28) of “fully adherent,” indicating a modest level of feasibility. However, there were also concerns regarding their self-reported mild symptoms. Based on the incidence, more participants reported hunger or cravings in the intervention group, although the participants in the control group reported them more frequently in terms of the number of events they experienced. There was a slight trend that participants in the intervention group had more headaches (both in events and incidence), but the result was not statistically significant. Except for hunger, fatigue, and cravings, there were no statistically significant differences in reported symptoms between the groups.

Concerning safety, the intervention showed encouraging indications. No serious adverse events were reported, and QOL remained stable (insignificant within-group differences in QOL score between before and after, and insignificant between-group differences in change in QOL) throughout the study. These findings, coupled with the modest adherence observed, suggest that the intervention is generally safe. Nonetheless, the intervention group reported more hunger, fatigue, and cravings,

XSL•FO
RenderX

underscoring the importance of providing thorough explanations and follow-up.

Comparison to Prior Work

Many studies have demonstrated the effectiveness of web-based weight loss interventions [7,8,25]. A previous RCT among adults with overweight in Japan showed a moderate weight loss effect from a 12-week app intervention (adjusted mean difference=–1.60, 95% CI –2.83 to –0.38; *P*=.01) [26]. A recent meta-analysis indicated that web-based interventions facilitate weight loss compared to minimal care control groups (mean difference=–1.40, 95% CI –1.98 to –0.82; *P*<.001) [8]. This study observed similar effects to these. However, we initially had expected a potential weight loss of 2.7 kg through the OIF, but it was 0.9 kg, and the observed between-group difference was –1.6 kg, with both changes falling short of the expectation. A possible reason for the slight weight loss in the intervention group and the weight gain in the control group could be attributed to the timing of the study. The study was conducted from September 2023 to January 2024, and most participants completed their follow-up measurements during the Christmas and New Year holidays, which possibly contributed to the weight gain [36]. Since total physical activity did not change, another possible reason should also lie in diet. Participants in the intervention group might feel hungrier, which is consistent with the previous studies [37,38], and might have compensated for the energy reduced by fasting through consuming more food on nonfasting days. This was implied by several participants' feedback in the questionnaire. However, this is just speculative since prior works have demonstrated the lack of compensation for energy intake in IF [39-41], and this study did not collect information on a regular diet. The next study should consider comprehensive dietary assessments to elucidate these findings further.

This study showed no significant reduction in fat mass (mean difference=–0.1, 95% CI –1.3 to 1.4 kg) but a decrease in muscle mass (mean difference=–1.3, 95% CI –2.5 to –0.2 kg). However, it is natural that glycogen stored in muscles is used first during the initial phase of short-term weight loss, accompanied by a temporary decrease in water and muscle mass [42]. Thus, the observed reduction in muscle mass may not directly raise concerns. A systematic review and meta-analysis comparing conventional dietary restrictions with IF have not found evidence that IF promotes greater muscle loss [43]. Furthermore, other studies have shown that incorporating exercise during IF can increase muscle mass and exercise performance [44,45]. A recent systematic review has also concluded that combining IF with strength training can maintain lean body mass [46]. These findings suggest that the OIF does not necessarily worsen body composition, but on its own, it might not provide sufficient health benefits. Weight loss interventions should ideally include exercise therapy and comprehensive lifestyle modifications [47,48], a principle that we conclude also applies to online fasting interventions.

Strengths and Limitations

One of the strengths of this study lies in its study design. A randomized controlled design using minimal care as a control enabled us to obtain more valid estimates than study designs with single-arm or no-intervention controls. The sample size that was large for a pilot study led to demonstrating statistically significant results for weight loss, justifying conducting future trials.

In addition, we comprehensively evaluated the effects of the intervention, measuring multiple outcomes, allowing us to obtain invaluable insights into our program. For example, no observed change in physical activity implied that the OIF might facilitate weight loss through reduced energy intake.

This study has several limitations. First, we lacked important data on physical activity and dietary habits. In this study, physical activity was measured by self-report and was considered to be of low accuracy, which might prevent us from detecting potential changes. Additionally, the lack of data on a regular diet limited the interpretation of the results. In particular, evaluating why weight loss was below expectations is quite challenging. Future studies should evaluate the physical activity and dietary habits more accurately in order to distinguish the efficacy of IF from other lifestyle factors.

Second, our analysis is limited by the way of handling the missing values. We used the BOCF approach to impute missing body weight data. However, since missing data on body weight after the intervention could occur systematically, this approach may have introduced bias and compromised our conclusions. Notably, the statistically significant between-group difference in weight change diminished when considering the worst-case scenario (imputing +3.3 kg of weight change for the 3 missing participants in the intervention group and –4.7 kg for the 2 missing participants in the control group). Nevertheless, we believe that the BOCF approach is unlikely to have significantly overestimated the intervention effect. Participants in the control group were unlikely to lose weight without intervention and might even have gained weight as previously reported [35], whereas those in the intervention group were more likely to achieve weight loss. Thus, imputing the outcome weight with the baseline value is more likely to lead to diluting the weight loss effect, not overestimating it. Furthermore, our analysis demonstrated that a greater proportion of participants in the intervention group achieved a clinically meaningful weight loss (MIC; a 3% reduction in initial weight) compared to the control group, supporting the effectiveness of the intervention.

Third, as for the participants' recruitment, we used convenience sampling and cannot assume that it was a random sampling. Notably, participants included several direct acquaintances of the researchers (TN and TI). The proportion of participants considering themselves to be healthy was 88% (50/57), which is comparable to the results from the reference national survey [34], with 86.16% (91768/106507) for the total population and 89.66% (50816/56675) for those aged 20-65 years, implying that they were not far from the generally healthy population with overweight and obesity. However, our participants were exclusively Japanese, and it is unclear whether the findings would be applicable in other countries or regions with different lifestyles. Thus, the generalizability of the conclusions remains unclear. Nevertheless, we believe that our setting with multiple facilities possibly contributed to recruiting participants with a broad range of characteristics. In the future trial, we plan to

XSL•FO
RenderX

broaden the generalizability of the findings by using systematic recruitment methods and conducting the study in other countries.

Finally, we did not systematically collect data on concomitant care that participants may have received during the trial, such as seeking external dietary advice or using other commercial weight-loss programs. Therefore, we cannot completely rule out the possibility that such unmeasured care could have influenced the results.

Conclusions

The 12-week OIF can promote weight loss in apparently healthy adults with overweight and obesity; yet, it may not result in improvements in body composition. The observed low dropout rate and modest adherence, with no serious adverse events or significant changes in QOL, indicate the feasibility and safety of the program.

Acknowledgments

Special thanks to TANITA Corporation for lending the body composition analyzer at no cost. We extend our gratitude to Fujita Hideo and his coworkers at Kusaka Hospital, Daigo Goto at Cold Storage Japan Inc, and Yuzo Hoshino at Buddy Training Co, Ltd. for their collaboration. Our thanks also go to Toshiaki Furukawa, Yusuke Ogawa, and Hiroshi Okada for their invaluable advice. We are grateful to Ryuta Onodera, Daisuke Sasaki, and Nami Noda for their dedicated support as research staff. Finally, our sincere thanks to Kazuya Goto and Rina Kitabe for their thorough review of the manuscript. In preparing this manuscript, the author (TN) used the generative artificial intelligence ChatGPT-4 (OpenAI) for the English translation and proofreading of the text. The authors have carefully reviewed all content generated with this technology, making necessary modifications, and assume full responsibility for the submitted work.

Funding

This research was funded by a donation from Ritsuan Inc, which had no role in any part of the research process.

Data Availability

The datasets analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: TN, TS, KK, MI
Data curation: TN, NN
Formal analysis: TN, NN
Funding acquisition: TN, TS
Investigation: TN, KK, MI
Methodology: TN, TS, KK, MI, NN, YT
Project administration: TN, TS, TI
Resources: TN, TS, YT, TI
Software: TN
Supervision: TI
Visualization: TN
Writing–original draft: TN
Writing–review & editing: TN, TS, KK, MI, NN, YT, TI

Conflicts of Interest

Ritsuan Inc, a funder of this study, did not have any role in the design, conduct, analysis, and reporting of the study. The closed, test version of the publicly available “Kenko-Nikki” app was developed and provided for this study at a low cost by HTL (Healthtech Laboratory, Inc). The principal investigator (TI) serves as an external director (but unpaid) at HTL, and uses intellectual property that he has transferred to HTL in this study. He also accepts other collaborative research projects that use the intellectual property, making him subject to auditing. TS is employed through joint research funds between HTL and Kyoto University, and uses the intellectual property mentioned above in this study. YT is also employed under the joint research funds. TN owns the Nihonsouken LLC, running dietary guidance business that involves fasting and selling health food products that are partly related to fasting. These conflicts of interest have been reported in advance to the Kyoto University Conflict of Interest Management Office and were audited by the Audit Unit of the Department of Clinical Research Facilitation, Institute for Advancement of Clinical and Translational Science at Kyoto University Hospital.

Multimedia Appendix 1
Supplementary information on participant recruitment.
[DOCX File , 1245 KB - mhealth_v14i1e58930_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Supplementary information on the interventions.
[DOCX File , 173 KB - mhealth_v14i1e58930_app2.docx]

Multimedia Appendix 3
Supplementary information on research app.
[PDF File (Adobe PDF File), 2948 KB - mhealth_v14i1e58930_app3.pdf]

Multimedia Appendix 4
Statistical analysis plan.
[PDF File (Adobe PDF File), 864 KB - mhealth_v14i1e58930_app4.pdf]

Multimedia Appendix 5
CONSORT 2025 checklist.
[PDF File (Adobe PDF File), 2055 KB - mhealth_v14i1e58930_app5.pdf]

References

1. Finucane MM, Stevens GA, Cowan MJ, Danaei G, Lin JK, Paciorek CJ, Global Burden of Metabolic Risk Factors of Chronic Diseases Collaborating Group (Body Mass Index). National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9·1 million participants. Lancet 2011;377(9765):557-567 [FREE Full text] [doi: 10.1016/S0140-6736(10)62037-5] [Medline: 21295846]
2. Obesity. World Health Organization. URL: https://www.who.int/news-room/facts-in-pictures/detail/6-facts-on-obesity [accessed 2024-02-29]
3. Arroyo-Johnson C, Mincey KD. Obesity epidemiology worldwide. Gastroenterol Clin North Am 2016;45(4):571-579 [FREE Full text] [doi: 10.1016/j.gtc.2016.07.012] [Medline: 27837773]
4. Bakaloudi DR, Barazzoni R, Bischoff SC, Breda J, Wickramasinghe K, Chourdakis M. Impact of the first COVID-19 lockdown on body weight: a combined systematic review and a meta-analysis. Clin Nutr 2022;41(12):3046-3054 [FREE Full text] [doi: 10.1016/j.clnu.2021.04.015] [Medline: 34049749]
5. Fraticelli F, Nicola MD, Vitacolonna E. A nutritional web-based approach in obesity and diabetes before and during the COVID-19 lockdown. J Telemed Telecare 2023;29(2):91-102 [FREE Full text] [doi: 10.1177/1357633X20966933] [Medline: 33081596]
6. Ammar A, Brach M, Trabelsi K. Effects of COVID-19 home confinement on eating behaviour and physical activity: results of the ECLB-COVID19 international online survey. Nutrients 2020;12:1583. [doi: 10.37473/dac/10.1101/2020.05.04.20072447]
7. Hutchesson MJ, Rollo ME, Krukowski R, Ells L, Harvey J, Morgan PJ, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. Obes Rev 2015;16(5):376-392. [doi: 10.1111/obr.12268] [Medline: 25753009]
8. Sorgente A, Pietrabissa G, Manzoni GM, Re F, Simpson S, Perona S, et al. Web-based interventions for weight loss or weight loss maintenance in overweight and obese people: a systematic review of systematic reviews. J Med Internet Res 2017;19(6):e229 [FREE Full text] [doi: 10.2196/jmir.6972] [Medline: 28652225]
9. Akers JD, Estabrooks PA, Davy BM. Translational research: bridging the gap between long-term weight loss maintenance research and practice. J Am Diet Assoc 2010;110(10):1511-22, 1522.e1 [FREE Full text] [doi: 10.1016/j.jada.2010.07.005] [Medline: 20869490]
10. Ye Y, Zhang M, Lin Z, Tang L. Is intermittent fasting better than continuous energy restriction for adults with overweight and obesity? Diabetes Metab Syndr Obes 2022;15:2813-2826 [FREE Full text] [doi: 10.2147/DMSO.S376409] [Medline: 36134390]
11. Anton SD, Moehl K, Donahoo WT, Marosi K, Lee SA, Mainous AG, et al. Flipping the metabolic switch: understanding and applying the health benefits of fasting. Obesity (Silver Spring) 2018;26(2):254-268 [FREE Full text] [doi: 10.1002/oby.22065] [Medline: 29086496]
12. Allaf M, Elghazaly H, Mohamed O, Fareen MFK, Zaman S, Salmasi AM, et al. Intermittent fasting for the prevention of cardiovascular disease. Cochrane Database Syst Rev 2021;1(1):CD013496 [FREE Full text] [doi: 10.1002/14651858.CD013496.pub2] [Medline: 33512717]
13. Morales-Suarez-Varela M, Collado Sánchez E, Peraita-Costa I, Llopis-Morales A, Soriano JM. Intermittent fasting and the possible benefits in obesity, diabetes, and multiple sclerosis: a systematic review of randomized clinical trials. Nutrients 2021;13(9):3179 [FREE Full text] [doi: 10.3390/nu13093179] [Medline: 34579056]
14. de Cabo R, Mattson MP. Effects of intermittent fasting on health, aging, and disease. N Engl J Med 2019;381(26):2541-2551. [doi: 10.1056/NEJMra1905136] [Medline: 31881139]

XSL•FO
RenderX

15. Klempel MC, Kroeger CM, Bhutani S, Trepanowski JF, Varady KA. Intermittent fasting combined with calorie restriction is effective for weight loss and cardio-protection in obese women. Nutr J 2012;11:98 [FREE Full text] [doi: 10.1186/1475-2891-11-98] [Medline: 23171320]
16. Kroeger CM, Klempel MC, Bhutani S, Trepanowski JF, Tangney CC, Varady KA. Improvement in coronary heart disease risk factors during an intermittent fasting/calorie restriction regimen: relationship to adipokine modulations. Nutr Metab (Lond) 2012;9(1):98 [FREE Full text] [doi: 10.1186/1743-7075-9-98] [Medline: 23113919]
17. Di Francesco A, Di Germanio C, Bernier M, de Cabo R. A time to fast. Science 2018;362(6416):770-775 [FREE Full text] [doi: 10.1126/science.aau2095] [Medline: 30442801]
18. Patterson RE, Sears DD. Metabolic effects of intermittent fasting. Annu Rev Nutr 2017;37:371-393 [FREE Full text] [doi: 10.1146/annurev-nutr-071816-064634] [Medline: 28715993]
19. Jamshed H, Beyl R, Della Manna D, Yang E, Ravussin E, Peterson C. Early time-restricted feeding improves 24-Hour glucose levels and affects markers of the circadian clock, aging, and autophagy in humans. Nutrients 2019;11(6):1234 [FREE Full text] [doi: 10.3390/nu11061234] [Medline: 31151228]
20. Manoogian EN, Panda S. Circadian rhythms, time-restricted feeding, and healthy aging. Ageing Res Rev 2017;39:59-67 [FREE Full text] [doi: 10.1016/j.arr.2016.12.006] [Medline: 28017879]
21. Wilhelmi de Toledo F, Grundler F, Sirtori CR, Ruscica M. Unravelling the health effects of fasting: a long road from obesity treatment to healthy life span increase and improved cognition. Ann Med 2020;52(5):147-161 [FREE Full text] [doi: 10.1080/07853890.2020.1770849] [Medline: 32519900]
22. Obesity in adults: prevention and lifestyle weight management programmes. National Institute for Health and Care Excellence. 2016. URL: https://www.nice.org.uk/guidance/qs111 [accessed 2023-07-31]
23. Open Research Cafe. URL: https://orec.space [accessed 2023-11-29]
24. Whitehead AL, Julious SA, Cooper CL, Campbell MJ. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. Stat Methods Med Res 2016;25(3):1057-1073 [FREE Full text] [doi: 10.1177/0962280215588241] [Medline: 26092476]
25. Muramoto A, Matsushita M, Kato A, Yamamoto N, Koike G, Nakamura M, et al. Three percent weight reduction is the minimum requirement to improve health hazards in obese and overweight people in Japan. Obes Res Clin Pract 2014;8(5):e466-e475. [doi: 10.1016/j.orcp.2013.10.003] [Medline: 25263836]
26. Nakata Y, Sasai H, Gosho M, Kobayashi H, Shi Y, Ohigashi T, et al. A smartphone healthcare application, CALO mama Plus, to promote weight loss: a randomized controlled trial. Nutrients 2022;14(21):4608 [FREE Full text] [doi: 10.3390/nu14214608] [Medline: 36364870]
27. Hróbjartsson A, Thomsen ASS, Emanuelsson F, Tendal B, Hilden J, Boutron I, et al. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. BMJ 2012;344:e1119. [doi: 10.1136/bmj.e1119] [Medline: 22371859]
28. Wei M, Brandhorst S, Shelehchi M, Mirzaei H, Cheng CW, Budniak J, et al. Fasting-mimicking diet and markers/risk factors for aging, diabetes, cancer, and cardiovascular disease. Sci Transl Med 2017;9(377):eaai8700 [FREE Full text] [doi: 10.1126/scitranslmed.aai8700] [Medline: 28202779]
29. Wilhelmi de Toledo F, Buchinger A, Burggrabe H, Hölz G, Kuhn C, Lischka E, Medical Association for FastingNutrition (Ärztegesellschaft für Heilfasten und Ernährung, ÄGHE. Fasting therapy - an expert panel update of the 2002 consensus guidelines. Forsch Komplementmed 2013;20(6):434-443. [doi: 10.1159/000357602] [Medline: 24434758]
30. "Health Diary" is a free app that records your daily health status. Healthtech.Lab Inc. URL: https://htech-lab.co.jp/products/kenkounikki.html [accessed 2024-02-21]
31. Ware J, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
32. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
33. Wilhelmi de Toledo F, Grundler F, Bergouignan A, Drinda S, Michalsen A. Safety, health improvement and well-being during a 4 to 21-day fasting period in an observational study including 1422 subjects. PLoS One 2019;14(1):e0209353 [FREE Full text] [doi: 10.1371/journal.pone.0209353] [Medline: 30601864]
34. Overview of the 2022 (Reiwa 4) comprehensive survey on living conditions of the people. Ministry of Health, Labour and Welfare. 2022. URL: https://www.mhlw.go.jp/toukei/saikin/hw/k-tyosa/k-tyosa22/index.html [accessed 2024-02-25]
35. Katsoulis M, Lai AG, Diaz-Ordaz K, Gomes M, Pasea L, Banerjee A, et al. Identifying adults at high-risk for change in weight and BMI in England: a longitudinal, large-scale, population-based cohort study using electronic health records. Lancet Diabetes Endocrinol 2021;9(10):681-694 [FREE Full text] [doi: 10.1016/S2213-8587(21)00207-2] [Medline: 34481555]
36. Helander EE, Wansink B, Chieh A. Weight gain over the holidays in three countries. N Engl J Med 2016;375(12):1200-1202. [doi: 10.1056/NEJMc1602012] [Medline: 27653588]

XSL•FO
RenderX

37. Sundfør TM, Svendsen M, Tonstad S. Effect of intermittent versus continuous energy restriction on weight loss, maintenance and cardiometabolic risk: a randomized 1-year trial. Nutr Metab Cardiovasc Dis 2018;28(7):698-706. [doi: 10.1016/j.numecd.2018.03.009] [Medline: 29778565]
38. Heilbronn LK, Smith SR, Martin CK, Anton SD, Ravussin E. Alternate-day fasting in nonobese subjects: effects on body weight, body composition, and energy metabolism. Am J Clin Nutr 2005;81(1):69-73 [FREE Full text] [doi: 10.1093/ajcn/81.1.69] [Medline: 15640462]
39. Huang A, Henderson G, Profeta A, Pfeiffer M, Feinstein LH, deLahunta M, et al. Lack of compensation of energy intake explains the success of alternate day feeding to produce weight loss. Physiol Behav 2023;263:114128. [doi: 10.1016/j.physbeh.2023.114128] [Medline: 36805441]
40. Varady KA, Bhutani S, Klempel MC, Kroeger CM, Trepanowski JF, Haus JM, et al. Alternate day fasting for weight loss in normal weight and overweight subjects: a randomized controlled trial. Nutr J 2013;12(1):146 [FREE Full text] [doi: 10.1186/1475-2891-12-146] [Medline: 24215592]
41. Ravussin E, Beyl RA, Poggiogalle E, Hsia DS, Peterson CM. Early time-restricted feeding reduces appetite and increases fat oxidation but does not affect energy expenditure in humans. Obesity (Silver Spring) 2019;27(8):1244-1254 [FREE Full text] [doi: 10.1002/oby.22518] [Medline: 31339000]
42. Denke MA. Metabolic effects of high-protein, low-carbohydrate diets. Am J Cardiol 2001;88(1):59-61. [doi: 10.1016/s0002-9149(01)01586-7] [Medline: 11423059]
43. Enríquez Guerrero A, San Mauro Martín I, Garicano Vilar E, Camina Martín MA. Effectiveness of an intermittent fasting diet versus continuous energy restriction on anthropometric measurements, body composition and lipid profile in overweight and obese adults: a meta-analysis. Eur J Clin Nutr 2021;75(7):1024-1039. [doi: 10.1038/s41430-020-00821-1] [Medline: 33293678]
44. Keenan SJ, Cooke MB, Hassan EB, Chen WS, Sullivan J, Wu SX, et al. Intermittent fasting and continuous energy restriction result in similar changes in body composition and muscle strength when combined with a 12 week resistance training program. Eur J Nutr 2022;61(4):2183-2199 [FREE Full text] [doi: 10.1007/s00394-022-02804-3] [Medline: 35084574]
45. Martínez-Rodríguez A, Rubio-Arias JA, García-De Frutos JM, Vicente-Martínez M, Gunnarsson TP. Effect of high-intensity interval training and intermittent fasting on body composition and physical performance in active women. Int J Environ Res Public Health 2021;18(12):6431 [FREE Full text] [doi: 10.3390/ijerph18126431] [Medline: 34198554]
46. Keenan S, Cooke MB, Belski R. The effects of intermittent fasting combined with resistance training on lean body mass: a systematic review of human studies. Nutrients 2020;12(8):2349 [FREE Full text] [doi: 10.3390/nu12082349] [Medline: 32781538]
47. Hall ME, Cohen JB, Ard JD, Egan BM, Hall JE, Lavie CJ, American Heart Association Council on Hypertension; Council on Arteriosclerosis, ThrombosisVascular Biology; Council on LifestyleCardiometabolic Health;Stroke Council. Weight-loss strategies for prevention and treatment of hypertension: a scientific statement from the American heart association. Hypertension 2021;78(5):e38-e50. [doi: 10.1161/HYP.0000000000000202] [Medline: 34538096]
48. Wharton S, Lau DC, Vallis M, Sharma AM, Biertho L, Campbell-Scherer D, et al. Obesity in adults: a clinical practice guideline. CMAJ 2020;192(31):E875-E891 [FREE Full text] [doi: 10.1503/cmaj.191707] [Medline: 32753461]

Abbreviations

BMR: basal metabolic rate
BOCF: Baseline Observation Carried Forward
CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Applications and Online Telehealth
CS: completers set
DBP: diastolic blood pressure
HbA1c: hemoglobin A1c
HDL-C: high-density-lipoprotein cholesterol
HTL: Healthtech Laboratory, Inc
IF: intermittent fasting
ITT: intention-to-treat
LDL-C: low-density-lipoprotein cholesterol
METs: metabolic equivalent of task-minutes per week
MIC: minimal clinically important change
OIF: Brief Online Intermittent Fasting Program
PPS: per protocol set
QOL: quality of life
RCT: randomized controlled trial
SBP: systolic blood pressure

XSL•FO
RenderX

SF-12v2: Short-Form 12-Item Survey-version 2
UMIN-CTR: University Hospital Medical Information Network-Clinical Trials Registry

Edited by L Buis; submitted 03.Apr.2024; peer-reviewed by J Zheng, J Mistry, R Bipat; comments to author 17.Sep.2024; revised version received 31.Dec.2024; accepted 30.Sep.2025; published 26.Jan.2026.

Please cite as:
Noda T, Shimamoto T, Kiyohara K, Imanaka M, Nishioka N, Tateyama Y, Iwami T
A Brief Web-Based and Mobile Intervention of Intermittent Fasting With Meal Support for Weight Loss Among Adults With Overweight and Obesity in Japan: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e58930
URL: https://mhealth.jmir.org/2026/1/e58930
doi:10.2196/58930
PMID:

©Takashi Noda, Tomonari Shimamoto, Kosuke Kiyohara, Mie Imanaka, Norihiro Nishioka, Yukiko Tateyama, Taku Iwami. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 26.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of Smartphone-Based Dyadic Interventions to Increase Physical Activity in Romantic Couples: Microrandomized Trial

Patrick Stefan Höhener[1], PhD; Robert Tobias[1], PhD, ING; James Martin Allen[1,2], PhD; Pascal Küng[1], PhD; Urte Scholz[1], Prof Dr

[1]Department of Psychology, University of Zurich, Zurich, Switzerland

[2]School of Health & Wellbeing, University of Glasgow, Glasgow, United Kingdom

Corresponding Author:
Patrick Stefan Höhener, PhD
Department of Psychology
University of Zurich
Binzmühlestrasse 14
8050
Zurich
Switzerland
Phone: 41 44 635 72 45
Email: patrick.hoehener@psychologie.uzh.ch

Abstract

Background: Social exchange processes, such as social support and social control, can promote health behavior change. However, these processes are often neglected when studying health behavior change and designing interventions. Intervening on these social exchange processes using dyadic interventions may provide a promising approach to promote health behaviors.

Objective: This study aimed to investigate the effects of dyadic interventions to increase moderate-to-vigorous physical activity (MVPA) in romantic couples. Furthermore, we explored how the target, type, and timing of the interventions affect their effectiveness.

Methods: In total, 38 romantic couples (mean age 34.01, SD 11.03 y) were recruited through online advertisements and participated in a smartphone-based microrandomized trial over 55 days consisting of control and intervention phases. The fully automated dyadic interventions included a one-time psychoeducation intervention, weekly dyadic and collaborative planning, and dyadic just-in-time adaptive interventions (JITAIs). MVPA was measured through daily diaries and wrist-worn accelerometers. We used multilevel modeling to estimate the effect of the intervention phase and weighted and centered estimation for microrandomized trials to estimate the treatment effects of dyadic and collaborative planning, as well as the dyadic JITAIs.

Results: Participants indicated higher device-based (b=5.88, SE=3.04, t_{3665}=1.93; *P*=.03) and self-reported (b=8.26, SE=3.88, t_{3904}=2.13; *P*=.01) MVPA during the intervention phase compared with the control phase. Dyadic and collaborative planning did not increase device-based (b=6.31, SE=5.18; *P*=.12) but only self-reported (b=14.25, SE=5.16; *P*=.005) MVPA. However, the effects of the 2 kinds of planning on self-reported MVPA disappeared when additional covariates were included (b=0.14, SE=3.32; *P*=.48). Furthermore, the dyadic JITAIs targeting both the actor and the partner increased device-based (actor: b=11.17, SE=3.18; *P*<.001; partner: b=7.23, SE=3.60; *P*=.03) and self-reported (actor: b=17.34, SE=3.65; *P*<.001; partner: b=11.82, SE=4.10; *P*<.001) MVPA. However, the effects of the dyadic JITAIs targeting the actor disappeared for self-reported MVPA (b=2.20, SE=3.22; *P*=0.25) when additional covariates were included. Exploratory analyses revealed that different types and timings of dyadic JITAIs were differentially effective.

Conclusions: This study demonstrated the promising effects of dyadic interventions to promote MVPA and highlighted the importance of the target, type, and timing of the interventions. Further research should investigate the mechanisms underlying the effects of dyadic interventions on health behaviors.

Trial Registration: ISRCTN registry ISRCTN15673058; https://www.isrctn.com/ISRCTN15673058

(JMIR Mhealth Uhealth 2026;14:e67136) doi:10.2196/67136

XSL•FO
RenderX

KEYWORDS

health behavior change; just-in-time adaptive interventions; dyadic interventions; romantic couples; physical activity; microrandomized trial; mobile phone

Introduction

Background

Regular engagement in physical activity is essential for maintaining overall health and reducing the risk of chronic and noncommunicable diseases, including heart disease, stroke, diabetes, and cancer [1]. The World Health Organization recommends at least 75 to 150 minutes of moderate-to-vigorous physical activity (MVPA) per week. Even though most people know of its importance and intent to engage in MVPA, many struggle to engage in the recommended amount of MVPA [2]. For example, in Switzerland, where the study took place, about 24% of adults do not meet these recommendations [3].

Theoretical and empirical research has been conducted to support people in increasing their MVPA [4]. This research has primarily focused on behavior change on the individual level and neglected social relationships and social processes [5]. However, since people are embedded in social networks, health behavior change often occurs in a social context. These social networks promote interdependence among people, and especially close partners can influence each other's health behaviors [6,7]. Thus, involving close social relationships in the behavior change process and implementing dyadic interventions provides a promising approach to promoting health behavior change in people with low physical activity [8].

Dyadic Interventions for Health Behavior Change

Implementing dyadic interventions makes it possible to capitalize on the interpersonal resources of both partners, encouraging mutual support and enhancing investment in the partners' behavior change, thereby facilitating behavior change [9-11]. The empirical evidence regarding the effectiveness of dyadic interventions is promising but mixed. On the one hand, some studies have shown that dyadic interventions are effective in increasing physical activity [12] and are more effective than control interventions, including equivalent interventions targeting individuals [9]. On the other hand, systematic reviews found mixed evidence for improved physical health [13], and they did not demonstrate superior effectiveness in increasing physical activity compared with comparison groups involving individual or control interventions [11].

Dyadic interventions may be categorized along the continuum of dyadic interventions into different prototypes of dyadic interventions depending on the degree of partner involvement (eg, cross-over and joint interventions) [10]. Cross-over interventions prompt interactions between the individuals by explicitly instructing one dyad member to engage with or clearly refer to the other dyad member, but the intervention is not administered to both. Joint interventions actively involve both dyad members and consider the dyad as a unit, making it impossible to be delivered to only 1 dyad member present. Cross-over and joint interventions may work through different mechanisms and thus may differ in their effectiveness. However, little research has investigated the differences in their effectiveness, calling for a more nuanced examination of these 2 prototypes of dyadic interventions [10].

Dyadic interventions may work by applying different dyadic behavior change techniques (DBCTs) [10] that explicitly elicit various social exchange processes influencing health behavior change [14]. Social exchange processes are interactions between dyad members that affect one or both individuals' behaviors, emotions, or cognitions [15]. There are various social exchange processes, such as providing social support, exerting social control, or collaboratively establishing goals and plans. Social support refers to providing psychological, material, or informational resources intended to benefit the ability to cope with stressors, solve problems, and pursue life opportunities [16,17]. One common distinction of social support is between emotional and instrumental support. Emotional support concerns the emotional well-being of the recipient [18]. Regarding physical activity, this form of support may include encouragement and comfort, guidance, and help to stay motivated and committed. Instrumental support is defined as providing the recipient with practical help and resources, such as advice or assistance [18]. Regarding physical activity, this form of support could include providing help while exercising or taking over daily chores to free up time for the partner to engage in MVPA. Empirical evidence showed that emotional and instrumental support can be associated with higher levels of physical activity, but there is also substantial heterogeneity in the associations (eg, studies by Kouvonen et al [19], Rackow et al [20], and Scarapicchia et al [21]).

Another social exchange process is social control, which refers to a deliberate and intentional attempt to change what another person thinks, feels, or does toward an outcome desired by the person exerting control [22]. Social control can be categorized into positive or negative social control based on the means used to influence behavior [23]. Positive social control includes strategies such as persuasion, modeling, and discussions. Negative social control includes, for example, coercion, social pressure, elicitation of negative emotions, and withdrawal. These strategies may differ in effectiveness when used to promote health behaviors in the partner. A meta-analysis has found that positive control was associated with better health behaviors, whereas negative control was related to worse health behaviors [22]. Furthermore, negative control relates to increased reactance-related responses and resistance to change, rendering it ineffective in encouraging health behavior change [7]. So far, there is almost no research on the effectiveness of interventions promoting positive social control and reducing negative social control. This gap will be addressed in this study.

While social support and positive social control can enhance health behaviors, poorly executed or nonresponsive support may yield adverse outcomes, including diminished well-being, heightened stress, and decreased performance [24]. Therefore, when deploying dyadic interventions that leverage social

XSL•FO
RenderX

exchange processes, it seems crucial for partners to have the knowledge and skills to offer effective support. However, little research has included information on skillful social support in interventions.

Social exchange processes can also play a role in goal setting and action planning. Goal setting refers to defining behaviors or states one wants to accomplish in the future [25]. Action planning refers to linking behaviors to specific cues by specifying when, where, and how to act [26]. Action planning can be augmented by coping planning, a barrier-focused self-regulation strategy in which an association between anticipated barriers and suitable solutions to overcome these barriers is made [26]. Combining action planning with coping planning can have additive and synergistic effects to promote health behaviors [27]. Furthermore, setting goals and planning physical activities with a partner may facilitate adherence to these goals and plans [5,28]. There are 2 forms of planning in the dyadic context. Dyadic planning involves creating plans with a partner on when, where, and how one partner will implement a behavior [29]. Collaborative planning entails creating joint plans with a partner on when, where, and how both partners will engage in a behavior [30]. There is mixed evidence for the effectiveness of dyadic planning in promoting health behaviors. While some studies have found positive associations between dyadic planning and goal progress [31] and plan enactment [32], other studies have not supported its effectiveness [33,34]. However, empirical evidence generally supports the effectiveness of collaborative planning to foster health behaviors [28,35,36].

Just-in-Time Adaptive Interventions

Technological advancements and the widespread use of devices, such as smartphones and smartwatches, have opened up new possibilities for enhancing the design and implementation of interventions [37]. These technologies allow for tracking health behaviors and context variables, which provide information about the individual's current state. This information can be integrated with theoretical and empirical knowledge to tailor the intervention content to individual needs and deliver interventions at the right time. Such interventions are known as just-in-time adaptive interventions (JITAIs) [38]. A meta-analytical review found moderate to large effects of JITAIs compared with waitlist-control conditions and non-JITAI treatments to improve health behaviors [39]. However, regarding physical activity and sedentary behavior, the evidence supporting the effectiveness of JITAIs is mixed [37]. However, so far, only JITAIs aimed at individuals have been implemented. Combining DBCTs with JITAIs and intervening on a dyadic level may provide a promising approach to increasing their effectiveness in promoting MVPA.

This Study

This study aimed to examine the effectiveness of dyadic interventions in promoting MVPA in romantic couples. Specifically, we investigated the effectiveness of dyadic interventions containing various DBCTs that elicit social exchange processes for increasing MVPA in romantic couples. We implemented an intensive-longitudinal study over 55 days consisting of phases without dyadic interventions (ie, control phase) and with dyadic interventions, including a one-time psychoeducation intervention on skilled support, weekly planning interventions, and dyadic JITAIs (ie, intervention phase). In the psychoeducation intervention on skilled support, participants learned how to support their partner appropriately to increase their MVPA. In the planning interventions, the participants set a goal and planned the physical activities for the upcoming week. Finally, there were dyadic JITAIs targeting various social exchange processes to increase MVPA.

To investigate the effectiveness of the dyadic interventions, we compared the MVPA during the intervention phase with the control phase. To isolate the effects of the different dyadic interventions, we investigated the effectiveness of the planning interventions and the dyadic JITAIs separately. Note that we did not examine the effect of the skilled support intervention since this was a one-time intervention taking place right after the baseline week and is, therefore, highly correlated with the intervention phase.

We proposed the following hypotheses, all of which were preregistered on Open Science Framework (OSF):

- Hypothesis 1: Couples show higher MVPA (device-based and self-reported) during the intervention phase than during the control phase.
- Hypothesis 2: On days for which participants had planned to be physically active during the planning intervention, they engage in more MVPA (device-based and self-reported) compared with days for which they had not planned any physical activity.
- Hypothesis 3a: Participants indicate higher MVPA (device-based and self-reported) on days when a dyadic JITAI targeted the actor's MVPA compared with days without a dyadic JITAI targeting the actor.
- Hypothesis 3b: Participants indicate higher MVPA (device-based and self-reported) on days when a dyadic JITAI targeted the partner compared with days without a dyadic JITAI targeting the partner.

Additionally, we conducted 3 exploratory analyses. First, we explored the effect of different types of dyadic JITAIs by differentiating between cross-over and joint interventions. Second, we explored the effectiveness of the dyadic JITAIs depending on their timing (ie, when they were sent). There were dyadic JITAIs sent before the planning intervention, before a planned activity, and in the evening. These 3 timings target different aspects of the behavior change process, which may vary in their effectiveness. Finally, we explored how long the effects of the dyadic JITAIs last. Dyadic JITAIs may make lasting changes in the interaction patterns between the partners, allowing their effects to be maintained over time. Thus, these changes in the social exchange processes may go beyond the specific day targeted by the dyadic JITAIs, facilitating engagement in MVPA at a later occasion.

Methods

Funding and Preregistration

This study is part of the "Time and Ties: Dynamic modelling of temporal patterns in dyadic health behaviour change" project

XSL•FO
RenderX

funded by the Swiss National Science Foundation (grant 10001C_197471 / 1). A comprehensive description of the project can be found on the OSF page of the Time and Ties project [40]. The preregistration, including hypotheses, inclusion and exclusion criteria, and planned primary analyses, can be found on the OSF page of the study [41]. The study differs from the preregistration in that we conducted the dyadic analyses for indistinguishable instead of distinguishable dyads. This decision was made to be inclusive of same-gender couples. The completed CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist can be found in Multimedia Appendix 1.

Ethical Considerations

The study received approval from the Ethics Committee of the Faculty of Arts and Social Sciences of the University of Zurich (approval 21.9.11). All participants provided online informed consent using a checkbox before enrolling in the study and were free to withdraw at any time (the informed consent is provided in Multimedia Appendix 2). All collected data are stored in anonymized form on the server of the University of Zurich and are only accessible to the project team. All participants received 150 Swiss Francs (approximately US $189) for their complete participation.

Participants

Data collection took place from June to October 2022. Romantic couples were recruited from the Swiss population through online advertisements on Facebook (Meta) and Instagram (Meta). Inclusion criteria were (1) both partners must be at least 18 years old, (2) they must be in a romantic relationship with each other and live together, (3) both partners must be physically active for less than 150 minutes per week, (4) both partners must have the intention to be more physically active, (5) both partners must not have severe health conditions preventing physical activity, (6) both partners must have a smartphone that they use regularly and have sufficient literacy to operate it independently, and (7) both partners must speak German fluently.

Descriptive statistics of the sample are presented in Table 1.

Table 1. Descriptive statistics of the sample characteristics (76 participants; 38 couples).

Variable	Value
Gender, n (%)	
Women	39 (51)
Men	37 (49)
Education, n (%)	
Vocational education	8 (11)
High school diploma	24 (32)
Bachelor's degree	23 (30)
Master's degree	21 (28)
Married, n (%)	26 (34)
Having children, n (%)	21 (28)
Age (y), mean (SD; range)	34.01 (11.03; 19-60)
BMI (kg/m^2), mean (SD; range)	24.94 (4.11; 16.37-33.95)
Relationship duration (y), mean (SD; range)	9.23 (9.09; 0.58-36.00)
Cohabitation duration (y), mean (SD; range)	7.53 (9.20; 0.25-33.00)
Age of children (n=21; y), mean (SD; range)	16.16 (9.24; 4-32)

Study Design

We conducted an intensive-longitudinal microrandomized trial over 55 days. This study was the first empirical study within the larger Time and Ties project and aimed to inform a computational modeling approach on dyadic interventions as well as to test and improve the smartphone-based data gathering and intervention provision. Thus, the study was not powered for advanced statistical analyses of dyadic intervention effects. From this perspective, it can be considered a pilot trial. An overview of the larger Time and Ties project, including this study, can be found on OSF. The study was smartphone-based, using a self-developed app in collaboration with a company specializing in app development (Multimedia Appendix 3). Participants read the information about the study and provided written informed consent before the start of the study. They were informed that various interventions would be implemented during the study. After registration, the participants received a link to download and sign in to the app. Throughout the study, participants completed daily diaries and wore wrist-worn accelerometers. The study consisted of phases without dyadic interventions (ie, control phase) and with dyadic interventions, including a 1-time psychoeducation intervention about skilled support, weekly planning interventions, and dyadic JITAIs (ie, intervention phase). Random allocation was implemented using a computer program that assigned couples to the intervention conditions characterized with different intervention phases, while ensuring equal sample sizes across groups (ie, blocked

XSL•FO
RenderX

randomization). Each group started with a 1-week control phase to establish a baseline. Afterwards, group A received a 7-week intervention phase. Group B received a 3-week intervention phase followed by a 4-week control phase. Group C received a 4-week control phase followed by a 3-week intervention phase at the end of the study (Figure S1 in Multimedia Appendix 4). During the intervention phase, various dyadic interventions were implemented on designated days, but there were also randomly assigned control days without any interventions. Participants were aware that there were interventions during the study, but they did not know which intervention conditions there were. At the end of the study, participants completed a questionnaire similar to the baseline questionnaire, where they also had the opportunity to provide qualitative feedback. All questionnaires and interventions were sent using a smartphone app designed for this study.

Interventions

Skilled Support Intervention

The psychoeducation intervention about the principle of skilled support took place on the first Saturday after the 1-week control phase. During this intervention, couples learned about different types of social support and social control, the importance of the right timing and equity of support, and how to support each other effectively to help engage in MVPA (DBCTs: "one partner receives education for supporting the other partner," which was applied by both partners) [24]. Furthermore, partners discussed and reported how and when they would like to be supported and how they communicate about social support and social control (DBCTs: "one partner identifies preferred support strategies for the other partner," "one partner gives feedback on support provision of the other partner," and "one partner practices communication skills for health behavior of the other partner;" all of which were applied by both partners).

Planning Interventions

The planning interventions occurred each Sunday during the intervention phase and consisted of 3 parts. First, the couple set their weekly goal by indicating the desired duration of MVPA they aimed to achieve in the upcoming week (DBCTs: "the couple sets a goal for the couple" and "the couple commits to a goal of the couple") [42]. Second, they planned physical activities for the upcoming week. Thereby, they could either plan dyadically (ie, plan together activities they want to do individually) [29] or collaboratively (ie, plan together activities they want to do together) [30]. They planned the activity, timing, location, duration, and whether they would participate individually or together in the planned activity (DBCTs: "the couple plans for one partner" and "the couple plans for the couple"). Third, they engaged in coping planning by anticipating potential barriers that may prevent them from engaging in the planned activities and discussing solutions to overcome them (DBCTs: "the couple creates a coping plan for one partner" and "the couple creates a coping plan for the couple") [26].

Dyadic JITAIs

The dyadic JITAIs were designed in accordance with the recommendations proposed by Nahum-Shani et al [38]. The JITAIs targeted various social exchange processes within the couple that were hypothesized to promote engagement in MVPA (a complete list of all the DBCTs that were implemented can be found on OSF). Note that the target of the JITAI is not necessarily the person who receives the JITAI. For example, a JITAI prompting to provide emotional support to engage in MVPA may be sent to one partner (ie, execution level) but targeting the other (ie, target level) [10]. Depending on the content of the JITAIs, they could target either one or both partners' MVPA. However, due to the interdependence between the partners, the effects may spill over to the nontargeted partner [43]. Thus, this study included JITAIs targeting the actor's MVPA and JITAIs targeting the partner's MVPA. Note that both JITAIs targeting the actor and the partner may be present on the same day (ie, through joint or 2 cross-over interventions targeting different partners).

There were 2 types of dyadic JITAIs: cross-over and joint JITAIs [10]. Cross-over JITAIs were sent to one partner and included instructions to interact with the other partner (eg, provide emotional support to the partner). Joint JITAIs were sent to both partners and included instructions to both partners to actively interact with each other (eg, jointly create a list of advantages of physical activity). The JITAIs were sent to the participants at a specified time, hypothesized to be the most appropriate for the JITAI content. There were three timings when the JITAIs could be triggered: (1) before the planning intervention, (2) before a planned activity, and (3) in the evening. The dyadic JITAIs before the planning intervention targeted the social exchange processes during the planning; those before the planned activity targeted the social exchange processes (eg, social support and social control) before and during the activity; and those in the evening targeted the social exchange processes during the reflection about the engagement in MVPA. Depending on the timing of the JITAIs, the expected effect might not manifest on the same day the couple received the dyadic JITAI but may be on a subsequent day they intended to engage in MVPA (Figures S2-S4 in Multimedia Appendix 4).

Each day, a randomization process either selected a JITAI or not, the latter serving as a control condition. Multimedia Appendix 4 provides a description of the dyadic JITAIs and the selection process. All interventions were stand-alone interventions and were triggered automatically without human involvement.

Measures

Intervention Variables

This study comprised several variables for the different intervention components. First, the study was categorized into the control phase (ie, phase without any dyadic interventions, serving as the reference category) and intervention phase (ie, phase with dyadic interventions). Second, planning was categorized into days without planned physical activities (coded as 0) and days for which the couples had planned physical activities during the planning intervention (coded as 1). Third, regarding the dyadic JITAIs, we included various dummy variables to describe whether a specified JITAI targeted the MVPA that day or did not target the day. We defined two variables for JITAIs targeting the actor and the partner: (1) the

XSL•FO
RenderX

dyadic JITAI for the actor compared days on which a dyadic JITAI targeted the actor's MVPA (coded as 1) with days without dyadic JITAI targeting the actor's MVPA (coded as 0), and (2) the dyadic JITAI for the partner compared days on which a dyadic JITAI targeted the partner's MVPA (coded as 1) with days without dyadic JITAI targeting the partner's MVPA (coded as 0). For the exploratory analyses, we created two variables representing the types of JITAI: (1) cross-over JITAI (coded as 1) versus no cross-over JITAI (coded as 0) targeting the MVPA on that day, and (2) joint JITAI (coded as 1) versus no joint JITAI (coded as 0) targeting the MVPA on that day. Joint JITAIs were coded as targeting both the actor's and partner's MVPA. Moreover, we created three variables representing the timing of the JITAI: (1) JITAI before the planning intervention (coded as 1) versus no JITAI before the planning intervention (coded as 0) targeting the MVPA that day, (2) JITAI before the activity (coded as 1) versus no JITAI before the activity (coded as 0) targeting the MVPA that day, and (3) JITAI in the evening (coded as 1) versus no JITAI in the evening (coded as 0) targeting the MVPA that day. Finally, we computed three lagged terms of the JITAI variables to describe how many occasions ago the JITAI targeted the MVPA: (1) JITAI targeting the MVPA 1 occasion before (coded as 1) compared with no JITAI targeting the MVPA one occasion before (coded as 0), (2) JITAI targeting the MVPA 2 occasions before (coded as 1) compared with no JITAI targeting the MVPA 2 occasions before (coded as 0), and (3) JITAI targeting the MVPA 3 occasions before (coded as 1) compared with no JITAI targeting the MVPA 3 occasions before (coded as 0).

Physical Activity

Physical activity was assessed using both device-based and self-reported methods. The device-based physical activity in minutes of MVPA per day was measured using wrist-worn accelerometers (ActiGraph CentrePoint Insight Watches [Ametris]). The acceleration is measured on 3 axes from which a single vector magnitude count is calculated. Values above 2690 counts per minute were classified as MVPA [44]. The device-based MVPA was filtered for wear time of at least 10 hours per day, according to Choi et al [45], and for awake time, according to the algorithm proposed by Tracy et al [46]. Overall, 11.3% (n=474) of the responses on device-based MVPA are missing after excluding days with low wear-compliance. The self-reported MVPA was measured by adding up the 2 questions capturing the MVPA they did alone ("How many minutes did you spend today alone doing moderate-to-vigorous physical activity?") and with their partner ("How many minutes did you spend today together with your partner doing moderate-to-vigorous physical activity?"). Overall, 5.6% (n=236) of the responses for self-reported MVPA are missing.

Covariates

We assessed the intervention group with 2 time-invariant dummy covariates (group A serving as the reference category) to address differences due to different sequences in the intervention and control phases. Furthermore, we assessed various time-variant covariates. We included the wear time of the wrist-worn accelerometer (minutes per day) because increased wear time increases the potential duration of MVPA that can be recorded. Moreover, we included the time (per 7 days) to address potential time trends [47]. Furthermore, we included dummy variables for the weekend (weekday serving as the reference category) to address differences in the behavior between weekdays and weekends, and skilled support intervention (before the intervention serving as the reference category) to address the potential effects of this intervention on the MVPA. Finally, we assessed the barriers and facilitating factors. Barriers and facilitating factors were assessed with the item "What has made your physical activity easier or more difficult today?" on a bipolar scale from −5 to 5, on which the participants responded to 8 influences that hindered or facilitated engagement in physical activity (eg, "hindering conditions for physical activity" to "beneficial conditions for physical activity"). The scores of the barriers and facilitating factors were summed up into 2 separate scores, indicating the total barriers and facilitating factors today. We included the barriers and facilitating factors to control the internal and external influences that may change the probability or duration of the engagement in MVPA and to test if the interventions effectively increased MVPA over and above these factors [48].

Data Analysis

To estimate the treatment effect of the intervention phase, we calculated multilevel models for intensive longitudinal data and indistinguishable dyads [49]. This approach can be used to analyze dyadic data by accounting for the nonindependence of observations within dyads. We included the intervention phase as the predictor variable and the device-based or self-reported MVPA as the outcome variables.

We used weighted and centered least-squares estimation for microrandomized trials to estimate the treatment effects of the planning interventions and the dyadic JITAIs [50]. This approach addresses the potential biases of time-varying treatment effects and provides consistent causal effects in microrandomized trials. We extended this approach to dyadic data by considering the dyads as the units of analysis. To be inclusive of same-gender couples in our study, we treated the partners as indistinguishable rather than distinguishable based on gender [49]. We included the planning or the JITAI targeting the actor and partner as predictor variables, respectively. We ran separate models with device-based and self-reported MVPA as outcome variables. In all models, complete case analyses were conducted. Thus, days with missing data were excluded from the data analyses. We included time as a covariate in all models [47]. Additionally, in analyses of device-based MVPA, we included the wear time of the wrist-worn accelerometer as an additional covariate.

Furthermore, we conducted sensitivity analyses to account for potential influences of covariates. In these sensitivity analyses, we additionally included variables indicating the intervention group, weekend, skilled support intervention, barriers, and facilitating factors as covariates. All covariates were grand mean-centered. All analyses were complete-case analyses.

We ran 3 separate models for the exploratory analyses. In these exploratory analyses, we included the dyadic JITAI type (ie, cross-over and joint), the JITAI timing (ie, before planning,

XSL•FO
RenderX

before the activity, and evening), and the lagged terms of the JITAI as dummy variables.

Software

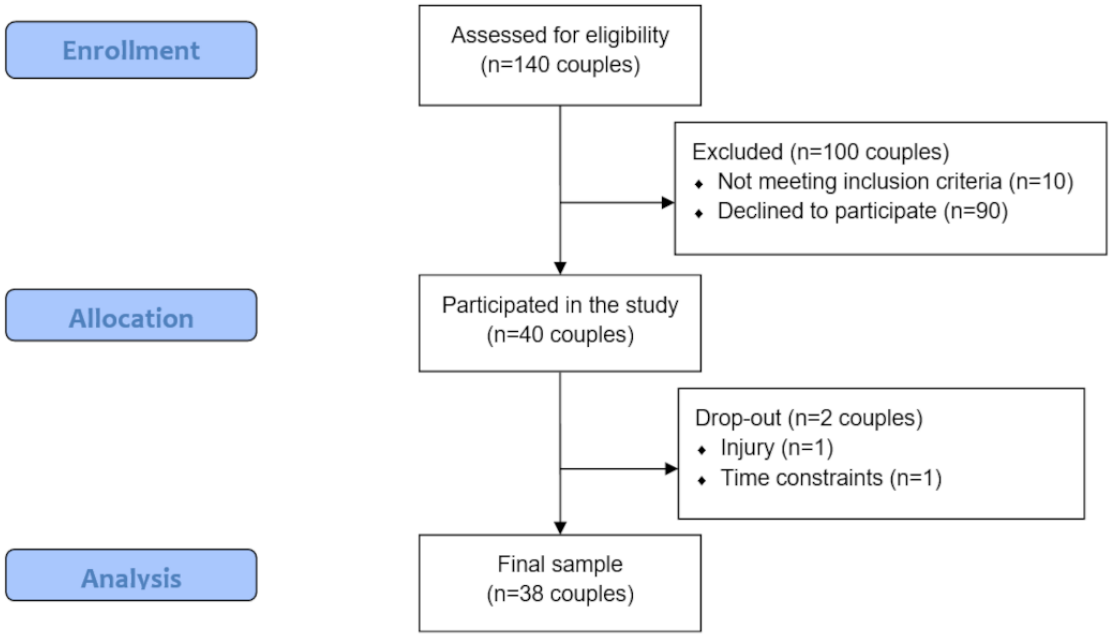
The analyses were conducted with R in RStudio [51,52]. We used the nlme package (v3.1-157) to estimate the effect of the intervention phase [53] and the xgeepack package (v1.3.9) to estimate the causal treatment effects of dyadic and collaborative planning and dyadic JITAIs [50,54,55]. All R codes are provided on OSF.

Results

Overview

Initially, 140 couples signed up for the study. However, 100 of them either chose not to participate or were excluded during screening. Thus, a total of 40 couples (80 participants) participated in the study, which accounted for an expected attrition rate of 20%. The recruitment stopped when the planned sample size was reached. Furthermore, 1 couple dropped out of the study due to an injury of 1 partner, and 1 couple discontinued the study because of time issues, leaving a total sample of 38 couples (37 mixed-gender and 1 same-gender couple). The CONSORT (Consolidated Standards of Reporting Trials) flow diagram is shown in Figure 1. The mean duration of MVPA was 116.64 (SD 35.40) minutes per day for device-based MVPA and 30.24 (SD 22.19) minutes per day for self-reported MVPA. There is a moderate to high correlation between the device-based and self-reported MVPA, with r_b=0.57 (P<.001) at the between-person and r_w=0.46 (P<.001) at the within-person level.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.



Tables 2 and 3 present the effects of the intervention phase on device-based and self-reported MVPA, expressed as the difference in minutes between days during the intervention phase compared with days during the control phase. In line with hypothesis 1, the results indicated that during the intervention phase, the couples had significantly higher device-based and self-reported MVPA than during the control phase. Furthermore, the intercepts and slopes showed significant variability, indicating heterogeneity in the mean levels of MVPA and the effects of the intervention phases on the MVPA between the couples.

XSL•FO
RenderX

Table 2. Effects of the intervention phase on device-based moderate-to-vigorous physical activity.

Effect[a]	Estimate, b (SE)	90% CI	*t* test (*df*) or *z* score	*P* value[b]
Fixed effects				
Intercept	106.50 (4.39)	99.29 to 113.72	24.28 (3665)[c]	<.001
Intervention phase[d]	5.88 (3.04)	0.88 to 10.89	1.93 (3665)[c]	.03
Time[e]	−1.05 (0.85)	−2.44 to 0.34	−1.24 (3665)[c]	.22
Wear time[f]	0.15 (0.01)	0.13 to 0.17	13.04 (3665)[c]	<.001
Random effects				
Intercept	24.30 (3.30)	19.48 to 30.33	7.37[g]	<.001
Intervention phase	8.59 (3.60)	4.51 to 16.36	2.39[g]	.02
Time	4.21 (0.83)	3.06 to 5.79	5.06[g]	<.001
Wear time	0.06 (0.01)	0.05 to 0.08	5.69[g]	<.001
Residuals	49.89 (0.65)	48.83 to 50.97	76.53[g]	<.001
Autocorrelation	0.27 (0.02)	0.24 to 0.30	15.67[g]	<.001

[a]Number of couples=38; number of days=55; number of cases in the analysis=3706.

[b]All *P* values are 2-tailed except those of the hypothesized effect of the intervention phase, where 1-tailed *P* values are used.

[c]*t* test.

[d]Intervention phase (coded as 1) included all days during the period when the couple received dyadic interventions and was compared with the control phase (coded as 0).

[e]Time was grand mean-centered per 7 days.

[f]Wear time of the accelerometers in minutes was grand mean-centered.

[g]*z* score.

XSL•FO
RenderX

Table 3. Effects of the intervention phase on self-reported moderate-to-vigorous physical activity.

Effects[a]	Estimate, b (SE)	90% CI	*t* test (*df*) or *z* score	*P* value[b]
Fixed effects				
Intercept	26.65 (3.17)	21.44 to 31.86	8.41 (3904)[c]	<.001
Intervention phase[d]	8.26 (3.88)	1.88 to 14.63	2.13 (3904)[c]	.02
Time[e]	−2.28 (0.95)	−3.84 to −0.71	−2.40 (3904)[c]	.02
Random effects				
Level-2				
Intercept	16.95 (2.50)	13.33 to 21.54	6.79[f]	<.001
Intervention phase	19.01 (3.50)	14.11 to 25.62	5.43[f]	<.001
Time	5.20 (0.79)	4.06 to 6.67	6.55[f]	<.001
Level-1				
Residuals	49.14 (0.57)	48.22 to 50.08	86.93[f]	<.001
Autocorrelation	0.03 (0.02)	0.00 to 0.06	1.74[f]	.08

[a]Number of couples=38; number of days=55; number of cases in the analysis=3994.

[b]All *P* values are 2-tailed except those of the hypothesized effect of the intervention phase, where 1-tailed *P* values are used.

[c]*t* test.

[d]Intervention phase (coded as 1) included all days during the period when the couple received dyadic interventions and was compared to the control phase (coded as 0).

[e]Time was grand mean-centered per 7 days.

[f]*z* score.

The average treatment effects of the planning interventions on device-based and self-reported MVPA in minutes are presented in Table 4. The results indicate that on days the participants planned to be physically active during the planning interventions, they indicated higher self-reported but not device-based MVPA. Thus, hypothesis 2 is only supported for self-reported but not device-based MVPA.

Table 4. Effects of the dyadic and collaborative planning interventions on device-based and self-reported moderate-to-vigorous physical activity.

Parameter[a]	Device-based MVPA[b,c]			Self-reported MVPA[d]		
	Estimate, b (SE)	90% CI	*P* value[e]	Estimate, b (SE)	90% CI	*P* value[e]
Intercept	107.59 (3.79)	101.19 to 114.00	<.001	30.27 (3.51)	24.34 to 36.21	<.001
Planning[f]	6.31 (5.18)	−2.45 to 15.06	.12	14.25 (5.16)	5.53 to 22.98	.005
Time[g]	−1.59 (0.87)	−3.06 to −0.12	.08	−2.27 (0.85)	−3.71 to −0.83	.01
Wear time[h]	0.16 (0.01)	0.13 to 0.18	<.001	—[i]	—	—

[a]Number of couples=38; number of days=55.

[b]MVPA: moderate-to-vigorous physical activity.

[c]Number of cases in the analysis for device-based moderate-to-vigorous physical activity=3706.

[d]Number of cases in the analysis for self-reported moderate-to-vigorous physical activity=3944.

[e]All *P* values are 2-tailed except those of the hypothesized effects of the planning interventions, where 1-tailed *P* values are used.

[f]Days with planned physical activities (coded as 1) were compared with days without any planned physical activities (coded as 0).

[g]Time was grand mean-centered per 7 days.

[h]Wear time of the accelerometers in minutes was grand mean-centered.

[i]Not applicable.

The average treatment effects of the dyadic JITAIs on device-based and self-reported MVPA in minutes are presented in Table 5. In line with hypothesis 3, the results indicated that both the dyadic JITAIs targeting the actor as well as the partner significantly increased the device-based and self-reported MVPA.

XSL•FO
RenderX

Table 5. Effects of the dyadic just-in-time adaptive interventions on device-based and self-reported moderate-to-vigorous physical activity.

Parameter[a]	Device-based MVPA[b,c]			Self-reported MVPA[d]		
	Estimate, b (SE)	90% CI	*P* value[e]	Estimate, b (SE)	90% CI	*P* value[e]
Intercept	107.62 (3.73)	101.31 to 113.93	<.001	30.17 (3.47)	24.30 to 36.04	<.001
Dyadic $JITAI_{Actor}$[f,g]	11.17 (3.18)	5.79 to 16.55	<.001	17.34 (3.65)	11.17 to 23.51	<.001
Dyadic $JITAI_{Partner}$[h]	7.23 (3.60)	1.14 to 13.33	.03	11.82 (4.10)	4.89 to 18.75	.003
Time[i]	−1.64 (0.79)	−2.97 to −0.30	.046	−2.17 (0.73)	−3.41 to −0.93	.006
Wear time[j]	0.16 (0.01)	0.13 to 0.18	<.001	—[k]	—	—

[a]Number of couples=38; number of days=55.

[b]MVPA: moderate-to-vigorous physical activity.

[c]Number of cases in the analysis for device-based moderate-to-vigorous physical activity=3706.

[d]Number of cases in the analysis for self-reported moderate-to-vigorous physical activity=3944.

[e]All *P* values are 2-tailed except those of the effects of the hypothesized dyadic just-in-time adaptive interventions, where 1-tailed *P* values are used.

[f]JITAI: just-in-time adaptive intervention.

[g]Days on which a dyadic just-in-time adaptive intervention targeted the actor's moderate-to-vigorous physical activity (coded as 1) were compared with days without dyadic just-in-time adaptive interventions targeting the actor's moderate-to-vigorous physical activity (coded as 0).

[h]Days on which a dyadic just-in-time adaptive intervention targeted the partner's moderate-to-vigorous physical activity (coded as 1) were compared with days without dyadic just-in-time adaptive interventions targeting the partner's moderate-to-vigorous physical activity (coded as 0).

[i]Time was grand mean-centered per 7 days.

[j]Wear time of the accelerometers in minutes was grand mean-centered.

[k]Not applicable.

Sensitivity Analyses

Results of the sensitivity analyses, including the effect of the intervention phase, the average treatment effects of the planning, and the dyadic JITAIs, are reported in Tables S1-S3 in Multimedia Appendix 5. The pattern of the effects of the intervention phase on the device-based and self-reported MVPA remained the same after controlling for various covariates, suggesting the robustness of the results. However, the average treatment effect of the planning intervention on self-reported MVPA disappeared when controlling for covariates. Similarly, the average treatment effect of the dyadic JITAIs targeting the actor on the self-reported MVPA was no longer significant when adding the covariates. We further explored which covariates were responsible for these different results compared with the main analyses. Excluding the barriers and facilitating factors from the sensitivity analyses led to the same patterns as in the main analyses, suggesting that the differences were mainly driven by the barriers and facilitating factors (Table S4 in Multimedia Appendix 5). Moreover, the skilled support intervention resulted in a negative effect on the self-reported MVPA in some sensitivity analyses.

Exploratory Analyses

We conducted 3 exploratory analyses to gain insights into the effects of the type, timing, and temporal dynamics of the dyadic JITAIs (Tables 6-8). Regarding the type of dyadic JITAIs (ie, cross-over or joint), the results showed significant effects of cross-over JITAIs for both targeting the actor as well as the partner on device-based and self-reported MVPA. That means that on days when cross-over dyadic JITAIs were sent, couples were more physically active. Joint JITAIs did not increase device-based but increased self-reported MVPA, indicating that on days when joint dyadic JITAIs were sent, couples reported being more physically active. Furthermore, the effect sizes of the cross-over and joint JITAIs were comparable for both device-based and self-reported MVPA.

XSL•FO
RenderX

Table 6. Effects of the exploratory analyses of cross-over and joint just-in-time adaptive interventions on device-based and self-reported moderate-to-vigorous physical activity.

Parameter[a]	Device-based MVPA[b,c]			Self-reported MVPA[d]		
	Estimate, b (SE)	90% CI	*P* value[e]	Estimate, b (SE)	90% CI	*P* value[e]
Intercept	107.64 (3.75)	101.30 to 113.98	<.001	30.17 (3.49)	24.27 to 36.08	<.001
Cross-over $JITAI_{Actor}$[f,g]	12.31 (4.16)	5.26 to 19.34	.006	13.68 (4.24)	6.51 to 20.86	.003
Cross-over $JITAI_{Partner}$[h]	8.68 (4.24)	1.50 to 15.86	.049	11.91 (5.22)	3.09 to 20.74	.03
Joint JITAI[i]	8.87 (6.00)	−1.30 to 19.04	.15	19.20 (9.56)	3.03 to 35.69	.05
Time[j]	−1.59 (0.78)	−2.91 to −0.28	.049	−2.06 (0.78)	−3.37 to −0.75	.01
Wear time[k]	0.16 (0.01)	0.13 to 0.18	<.001	—[l]	—	—

[a]Number of couples=38; number of days=55.

[b]MVPA: moderate-to-vigorous physical activity.

[c]Number of cases in the analysis for device-based moderate-to-vigorous physical activity=3706.

[d]Number of cases in the analysis for self-reported moderate-to-vigorous physical activity=3944.

[e]All *P* values are 2-tailed.

[f]JITAI: just-in-time adaptive intervention.

[g]Days on which a cross-over JITAI targeted the actor's MVPA (coded 1) were compared to days without cross-over JITAIs targeting the actors's MVPA (coded 0).

[h]Days on which a cross-over JITAI targeted the partner's MVPA (coded 1) were compared to days without cross-over JITAIs aiming at the partner's MVPA (coded 0).

[i]Days on which a joint JITAI targeted both's MVPA (coded 1) were compared to days without joint JITAIs (coded 0).

[j]Time was grand-mean-centered per 7 days.

[k]Wear time of the accelerometers in minutes was grand-mean-centered.

[l]Not applicable.

XSL•FO
RenderX

Table 7. Effects of the exploratory analyses of the timing of the dyadic just in-time adaptive interventions on device-based and self-reported moderate-to-vigorous physical activity.

Parameter[a,b,c]	Device-based MVPA[d,e]			Self-reported MVPA[f]		
	Estimate, b (SE)	90% CI	*P* value[g]	Estimate, b (SE)	90% CI	*P* value[g]
Intercept	107.65 (3.72)	101.34 to 113.97	<.001	30.16 (3.45)	24.30 to 36.02	<.001
Dyadic JITAI[h] planning$_{Actor}$	16.59 (4.74)	8.54 to 24.64	.002	17.92 (6.15)	7.48 to 28.36	.007
Dyadic JITAI activity$_{Actor}$	−6.82 (5.37)	−15.94 to 2.30	.21	10.87 (6.39)	0.03 to 21.71	.10
Dyadic JITAI evening$_{Actor}$	4.17 (3.29)	−1.42 to 9.77	.22	−1.69 (6.11)	−12.06 to 8.68	.74
Dyadic JITAI planning$_{Partner}$	6.57 (5.11)	−2.11 to 15.25	.21	13.97 (6.18)	3.47 to 24.46	.03
Dyadic JITAI activity$_{Partner}$	8.07 (5.65)	−1.53 to 17.67	.16	9.55 (2.99)	4.48 to 14.62	.003
Dyadic JITAI evening$_{Partner}$	3.54 (3.82)	−2.95 to 10.03	.36	5.41 (3.43)	−0.41 to 11.24	.12
Time[i]	−1.66 (0.77)	−2.98 to −0.34	.04	−2.15 (0.69)	−3.33 to −0.98	.004
Wear time[j]	0.16 (0.01)	0.13 to 0.18	<.001	—[k]	—	—

[a]Number of couples=38; number of days=55.

[b]The “Actor” and “Partner” subscripts indicate who was targeted by the dyadic just-in-time adaptive intervention. For just-in-time adaptive interventions targeting the actor, days on which a dyadic just-in-time adaptive intervention targeted the actor’s moderate-to-vigorous physical activity (coded as 1) were compared with days without dyadic just-in-time adaptive interventions targeting the actor’s moderate-to-vigorous physical activity (coded as 0). For just-in-time adaptive interventions targeting the partner, days on which a dyadic just-in-time adaptive intervention targeted the partner’s moderate-to-vigorous physical activity (coded as 1) were compared with days without dyadic just-in-time adaptive interventions targeting the partner’s moderate-to-vigorous physical activity (reference category, coded as 0).

[c]Dyadic just-in-time adaptive interventions may be triggered before the planning intervention (labeled planning), before a planned activity (labeled activity), and in the evening (labeled evening).

[d]MVPA: moderate-to-vigorous physical activity.

[e]Number of cases in the analysis for device-based moderate-to-vigorous physical activity=3706.

[f]Number of cases in the analysis for self-reported moderate-to-vigorous physical activity=3944.

[g]All *P* values are 2-tailed.

[h]JITAI: just-in-time adaptive intervention.

[i]Time was grand-mean-centered per 7 days.

[j]Wear time of the accelerometers in minutes was grand mean-centered.

[k]Not applicable.

XSL•FO
RenderX

Table 8. Effects of the exploratory analyses of the lagged effects of the dyadic just-in-time adaptive interventions on device-based and self-reported moderate-to-vigorous physical activity.

Parameter[a,b,c]	Device-based MVPA[d,e]			Self-reported MVPA[f]		
	Estimate, b (SE)	90% CI	*P* value[g]	Estimate, b (SE)	90% CI	*P* value[g]
Intercept	107.63 (3.73)	101.28 to 113.97	<.001	30.05 (3.51)	24.08 to 36.02	<.001
Dyadic $JITAI_{Actor}$[h]	7.79 (3.12)	2.47 to 13.10	.02	9.23 (3.44)	3.38 to 15.07	.01
Dyadic $JITAI_{Actor\ lag1}$	5.69 (3.28)	0.10 to 11.27	.09	10.71 (3.41)	4.90 to 16.51	.004
Dyadic $JITAI_{Actor\ lag2}$	0.22 (2.55)	−4.13 to 4.56	.93	6.38 (2.65)	1.87 to 10.89	.02
Dyadic $JITAI_{Actor\ lag3}$	1.13 (4.27)	−6.14 to 8.40	.79	5.25 (5.16)	−3.52 to 14.03	.31
Dyadic $JITAI_{Partner}$	4.10 (4.04)	−2.79 to 10.99	.32	7.66 (4.27)	0.39 to 14.92	.08
Dyadic $JITAI_{Partner\ lag1}$	3.22 (2.87)	−1.68 to 8.11	.27	4.35 (2.68)	−0.20 to 8.91	.12
Dyadic $JITAI_{Partner\ lag2}$	1.61 (3.18)	−3.80 to 7.03	.62	−0.96 (2.76)	−5.65 to 3.73	.73
Dyadic $JITAI_{Partner\ lag3}$	0.44 (4.10)	−6.54 to 7.42	.92	−2.33 (4.35)	−9.73 to 5.07	.60
Time[i]	−1.82 (0.83)	−3.23 to −0.41	.04	−2.56 (0.73)	−3.81 to −1.31	.002
Wear time[j]	0.16 (0.01)	0.13 to 0.18	<.001	—[k]	—	—

[a]Number of couples=38; number of days=55.

[b]The "Actor" and "Partner" subscripts indicate who was targeted by the dyadic just-in-time adaptive intervention. For just-in-time adaptive interventions targeting the actor, days on which a dyadic just-in-time adaptive intervention targeted the actor's moderate-to-vigorous physical activity (coded as 1) were compared with days without dyadic just-in-time adaptive interventions targeting the actor's moderate-to-vigorous physical activity (coded as 0). For just-in-time adaptive interventions targeting the partner, days on which a dyadic just-in-time adaptive intervention targeted the partner's moderate-to-vigorous physical activity (coded as 1) were compared with days without dyadic just-in-time adaptive interventions targeting the partner's moderate-to-vigorous physical activity (coded as 0).

[c]The 3 lag parameters indicate how many occasions ago the just-in-time adaptive interventions targeted the moderate-to-vigorous physical activity.

[d]MVPA: moderate-to-vigorous physical activity.

[e]Number of cases in the analysis for device-based moderate-to-vigorous physical activity=3706.

[f]Number of cases in the analysis for self-reported moderate-to-vigorous physical activity=3944.

[g]All *P* values are 2-tailed.

[h]JITAI: just-in-time adaptive intervention.

[i]Time was grand mean-centered per 7 days.

[j]Wear time of the accelerometers in minutes was grand mean-centered.

[k]Not applicable.

The dyadic JITAIs varied in effectiveness depending on the time they were sent. The dyadic JITAIs sent before the planning intervention targeting the actor increased the device-based and self-reported MVPA. The dyadic JITAIs sent before the planning intervention targeting the partner increased the self-reported but not the device-based MVPA. Furthermore, the dyadic JITAIs sent before the planned activity targeting the actor did not increase device-based nor self-reported MVPA. The dyadic JITAIs sent before the planned activity targeting the partner increased self-reported, but not the device-based MVPA. Finally, neither the dyadic JITAIs sent in the evening targeting the actor nor the partner had any significant effect on the self-reported or device-based MVPA.

Regarding the temporal dynamics of the JITAIs, there was evidence for lagged effects of the dyadic JITAIs targeting the actor on self-reported and, to a lesser extent, on device-based MVPA. This finding indicates that the dyadic JITAIs targeting the actor not only affected the MVPA at the time they were initially aimed at, but their effects lasted and influenced subsequent self-reported MVPA. There was no evidence of lagged effects of dyadic JITAIs targeting the partner on device-based or self-reported MVPA.

Discussion

Principal Findings

This study demonstrated the promising effects of smartphone-based dyadic interventions in promoting MVPA in romantic partners. Participants reported higher device-based and self-reported MVPA during the intervention phase than during the control phase. Furthermore, planning physical activities increased the participants' self-reported but not device-based MVPA. Dyadic JITAIs targeting the actor as well as the partner increased the device-based and self-reported MVPA. However, sensitivity analyses indicated that some of these effects were not robust, suggesting that contextual factors play a relevant role in the effectiveness of dyadic JITAIs. Exploratory analyses showed that cross-over JITAIs promoted device-based and self-reported MVPA, and joint JITAIs promoted self-reported MVPA. Moreover, dyadic JITAIs aiming

XSL•FO
RenderX

at different times (ie, before a planning intervention, before a planned activity, and in the evening) were differently effective in increasing MVPA. Finally, there was evidence for lagged effects of the dyadic JITAIs targeting the actor for device-based and self-reported MVPA but not for lagged effects of dyadic JITAIs targeting the partner.

Effects of the Dyadic Interventions

Consistent with past research, this study found promising effects of dyadic interventions to increase physical activity. Overall, the complex intervention, including different dyadic interventions, was effective in increasing both device-based and self-reported MVPA (hypothesis 1). However, there was heterogeneity in the effects of the intervention phase across couples. This heterogeneity may reflect differences in couple dynamics or insufficient knowledge of moderating factors interacting with the intervention or control phase [48]. A better understanding of these factors may help improve the effectiveness of dyadic interventions [39,56]. Future research may investigate intrapersonal characteristics, relationship dynamics, and the broader social context to provide valuable insights into factors relevant to intervention development and implementation [56].

This study found significant effects of the planning intervention on the self-reported but not device-based MVPA (hypothesis 2). These results are comparable with previous studies that found mixed effects of dyadic and collaborative planning [28,31,32,35]. In contrast to previous studies, this study used a within-couple design and compared days with planned activities to days without planned activities instead of comparing the effects of various planning interventions (eg, dyadic, collaborative, and individual) on subsequent physical activity. Additionally, the participants could freely choose and vary if they wanted to plan dyadically or collaboratively. This approach aimed to enable participants to plan according to their needs and preferences to compensate for potential challenges associated with each form of planning. However, these attempts to better tailor the planning interventions to couples' needs and to use more contingent outcomes [57] were not sufficient to enhance the planning intervention's effectiveness, at least with regard to the device-based measure of MVPA. As discussed previously [58], it seems necessary to identify under what circumstances and for whom these forms of planning unfold the most beneficial effects.

To our knowledge, our study was the first to implement dyadic JITAIs to promote MVPA in romantic couples. In line with previous studies on the effectiveness of JITAIs on health behaviors [39], this study showed promising effects of dyadic JITAIs in promoting MVPA (hypothesis 3a). Furthermore, not only the dyadic JITAIs targeting the actor but also those targeting the partner increased the MVPA (hypothesis 3b). This finding may be explained by both the dyadic nature of the JITAIs as well as the interdependence between close partners in a romantic relationship [6]. One partner's engagement may motivate the other partner to engage in physical activity as well [59]. Furthermore, helping the partner engage in physical activity may have positive outcomes for the provider [43]. For example, providing social support or control to the partner may reiterate the importance of physical activity for the provider [60], increase self-esteem [61], and contribute to feeling more energized [62], promoting engagement in physical activity. Thus, deciding who to target with dyadic interventions appears to be an essential design choice when developing and implementing dyadic interventions.

The effect of the planning and the effect of the dyadic JITAIs targeting the actor on self-reported MVPA disappeared when controlling for additional covariates. Exploratory analyses suggested that these discrepancies were driven by the reported barriers and facilitating factors. Potentially, the barriers and facilitating factors may act as mediators in explaining the JITAIs' effects on MVPA. The planning and JITAIs prompted processes to help engage in MVPA, which may have removed barriers or facilitated engagement in MVPA. Furthermore, it is also possible that the causality goes in the other direction, in that engagement and nonengagement in MVPA influenced the perception of barriers and facilitating factors.

We found an unexpected negative effect of the skilled support intervention on self-reported MVPA in some of the analyses. However, it is important to note that this part of the intervention was not randomized.

This study found some differences in the effects of the dyadic interventions on device-based and self-reported MVPA. Interestingly, in this study, the device-based measure of MVPA was much higher than the self-reported measure. These 2 measures represent related but distinct indicators of MVPA [63]. There are several explanations for these differences. First, self-reported and device-based MVPA may capture different aspects of physical activity [64]. For example, self-reported physical activity may refer more closely to the activities perceived as physical activity by the participants, whereas device-based physical activity may also include other forms of movement, such as naturally-occurring MVPA. Furthermore, it may be possible that engaging in planned physical activities (captured by both the device-based and self-reported MVPA) leads to compensatory effects in that the engagement in naturally occurring MVPA (captured only by device-based MVPA) decreases [65], leading to some discrepancies between the measures. Second, there may be response biases in the self-reported but not device-based MVPA measure [64]. Participants may indicate higher MVPA if they received an intervention because they expect it to be effective. Finally, the device-based MVPA may have misclassified some motions as MVPA, which would also explain why the device-based MVPA was substantially higher than the self-reported MVPA [65].

Exploratory Analyses

Results from the exploratory analyses showed that cross-over JITAIs increased device-based and self-reported MVPA, and joint JITAIs increased self-reported MVPA while having similar effect sizes. Given the exploratory nature of our analyses, results need to be replicated in confirmatory designs and analyses in the future. In addition to examining the main effects of these different prototypes of dyadic interventions [10] and despite their comparable effect sizes found in our study, these 2 types of interventions may target different mechanisms of action [10,14]. For example, joint interventions, such as discussions

XSL•FO
RenderX

and collaborative planning, may trigger interactions in which both partners are equally involved. In contrast, cross-over interventions, such as prompting to provide social support and to exert social control, may trigger one-sided interactions where one partner is more involved than the other. Future research is needed to better understand the potential different effects of cross-over and joint interventions and how they may vary under specific circumstances.

The second set of exploratory analyses showed that dyadic JITAIs with different timings varied in effectiveness. Specifically, there was stronger evidence for the effectiveness of the dyadic JITAIs sent before the planning intervention and before the planned activity than those sent in the evening. These dyadic JITAIs targeted different aspects of the behavior change process. The dyadic JITAIs before the planning interventions aimed at increasing the quality of the dyadic and collaborative plans, a factor central to the effectiveness of planning [58]. The dyadic JITAIs before the planned activity aimed at helping to engage in this activity. Sending dyadic JITAIs at this specific time may be effective because they directly enhance commitment to the planned activities. The dyadic JITAIs in the evening aimed at the partners to reflect on their goal progress, discussing things that went well or poorly regarding physical activity, or prompting support or positive control in the future. A reason for the lack of effectiveness may be that these dyadic JITAIs were not matched to a particular state of vulnerability or opportunity since they did not target immediate preparation or engagement in MVPA directly. Furthermore, the participants may not have had the opportunity to engage in the prompted task because the interventions were sent too late in the day. Thus, future studies should also examine intervention fidelity and intervention engagement, which may vary with different timings of the interventions.

The final exploratory analyses showed that dyadic JITAIs may have effects over and above their targeted time point. Implementing dyadic JITAIs to improve the social exchange processes may translate to improved interaction patterns that sustainably facilitate engagement in health behaviors [11]. Since many health behaviors, such as physical activity, require consistent engagement over extended periods of time, this would be promising for promoting longer-term behavior change [66]. Specifically, this study found evidence for long-term effects of dyadic JITAIs targeting the actor but not for those targeting the partner. These results suggest that different mechanisms may explain these effects [43]. For example, receiving social support and control may establish a subjective norm or change attitudes regarding physical activity [67], which translates into regular engagement in physical activity. In contrast, providing social support and control may promote MVPA by increasing positive effects [43] and energy levels [61], which may entail more transient effects. Future research is needed to investigate these underlying mechanisms in more detail.

Past research has illustrated a need to better understand the boundary conditions moderating the effectiveness of (dyadic) JITAIs [39]. The exploratory analyses of this study contribute to this call by illustrating preliminary insights into the importance of the type and timing of the dyadic JITAIs for their effectiveness. Additionally, it showed the potential long-term effects of dyadic JITAIs, providing new insights and opportunities for health behavior promotion. Given the exploratory nature, however, the next steps are to replicate findings with confirmatory analyses. Moreover, in future studies with more participants, it might also be worthwhile to examine interactions between the different factors examined here. For example, it might well be that joint JITAIs work better when being sent in the evening to prompt joint reflection of past interactions, compared with cross-over JITAIs that might work better before a planning intervention. It is also possible that joint JITAIs might have longer-lasting effects, as the positive dyadic dynamics in supportive interactions promoted by these kinds of dyadic JITAIs might add to the maintenance of joint behavior change attempts.

Strengths and Limitations

This study has various strengths. First, the study investigated health behavior change from a dyadic rather than individual perspective, allowing for a more complete understanding of health behavior change processes in romantic relationships. Second, the microrandomized trial design allowed us to experimentally investigate the dyadic health behavior change as a process over time and to examine daily dynamics [47]. This also allowed us to examine the causal effects of the interventions on the outcomes [50]. Third, we were able to investigate shorter-term as well as lagged effects of the dyadic JITAIs. Finally, including device-based and self-reported MVPA measurements offered a more comprehensive understanding of the effects of dyadic interventions on MVPA [63], as well as addressed potential methodological challenges inherent to the measurement methods.

Despite these strengths, this study also has its limitations. As it was a pilot study, the sample size of 38 romantic couples was relatively small. This relatively small sample size may limit the statistical power, which may be especially the case for the exploratory analyses, where there have been fewer instances of interventions. However, the relatively large number of days examined per couple (ie, 55 days) partly compensates for this small sample size [47]. The limited number of couples also restricted the possibility of conducting more detailed analyses exploring the effectiveness of interventions targeting specific social exchange processes (eg, emotional support and positive social control) and potential mechanisms underlying the intervention effects or interaction effects as outlined above. Furthermore, there were different types of dyadic interventions (ie, skilled support intervention, planning interventions, and JITAIs) implemented in the study. These interventions were designed to build upon and facilitate each other. For example, the dyadic JITAIs were built on the knowledge of skilled support and were triggered in reference to the planned activities. However, this also implies that it remains unclear how exactly and to what extent these different dyadic interventions influenced each other (eg, if the dyadic JITAIs would have been effective without an initial skilled support intervention). Finally, the study targeted inactive romantic couples who expressed a willingness to increase their physical activity. Thus, the study sample was self-selected and therefore may not fully represent the general internet population.

XSL•FO
RenderX

Conclusion

This study extended the common approach of promoting health behavior change of individuals by including partners in the behavior change process. This allowed us to implement DBCTs that address social interactions between the partners. Overall, this study demonstrated the effectiveness of dyadic interventions, including planning interventions and dyadic JITAIs. Furthermore, the exploratory analyses provided first hints to the assumptions that conditions, such as the type, timing, and target, moderate the effectiveness of dyadic JITAIs, providing new insights into conceptual and design elements of dyadic JITAIs. Future research is needed to get a more comprehensive understanding of the effectiveness of dyadic interventions. Specifically, the moderating variables influencing the effectiveness of dyadic interventions and the mechanisms explaining dyadic interventions should be examined.

Acknowledgments

No generative artificial intelligence has been used.

Funding

This study, as well as PSH and JMA, was funded by the Swiss National Science Foundation (grant 10001C_197471). Furthermore, JMA was funded for part of this work by the Medical Research Council (MR/X003264/1 and MC_UU_00022/3) and Chief Scientist Office (SPHSU18).

Data Availability

Data are available upon reasonable request.

Authors' Contributions

Conceptualization: PSH, US
Formal analysis: PSH
Funding acquisition: RT, US
Methodology: PSH
Project administration: RT, US
Resources: RT, US
Supervision: US
Writing – original draft: PSH
Writing – review and editing: RT, JMA, PK, US

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 324 KB - mhealth_v14i1e67136_app1.pdf]

Multimedia Appendix 2
Informed Consent.
[DOCX File , 17 KB - mhealth_v14i1e67136_app2.docx]

Multimedia Appendix 3
Screenshots Study App.
[PPTX File , 267 KB - mhealth_v14i1e67136_app3.pptx]

Multimedia Appendix 4
Dyadic Just-in-time Adaptive Interventions.
[DOCX File , 336 KB - mhealth_v14i1e67136_app4.docx]

Multimedia Appendix 5
Additional Results.
[DOCX File , 43 KB - mhealth_v14i1e67136_app5.docx]

XSL•FO
RenderX

References

1. Physical activity. World Health Organization. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/physical-activity [accessed 2025-12-16]
2. Feil K, Fritsch J, Rhodes RE. The intention-behaviour gap in physical activity: a systematic review and meta-analysis of the action control framework. Br J Sports Med 2023;57(19):1265-1271. [doi: 10.1136/bjsports-2022-106640] [Medline: 37460164]
3. Körperliche Aktivität. Bundesamt für Statistik. 2022. URL: https://www.bfs.admin.ch/bfs/de/home/statistiken/gesundheit/determinanten/koerperliche-aktivitaet.html [accessed 2025-12-16]
4. Gourlan M, Bernard P, Bortolon C, Romain AJ, Lareyre O, Carayol M, et al. Efficacy of theory-based interventions to promote physical activity. A meta-analysis of randomised controlled trials. Health Psychol Rev 2016;10(1):50-66. [doi: 10.1080/17437199.2014.981777] [Medline: 25402606]
5. Scholz U, Berli C, Lüscher J, Knoll N. Dyadic behavior change interventions. In: Hagger MS, Cameron LD, Hamilton K, Hankonen N, Lintunen T, editors. Handb Behav Change. England: Cambridge University Press; 2020:632-648.
6. Huelsnitz CO, Jones RE, Simpson JA, Joyal-Desmarais K, Standen EC, Auster-Gussman LA, et al. The dyadic health influence model. Pers Soc Psychol Rev 2022;26(1):3-34. [doi: 10.1177/10888683211054897] [Medline: 34873983]
7. Scholz U, Stadler G, Berli C, Lüscher J, Knoll N. How do people experience and respond to social control from their partner? Three daily diary studies. Front Psychol 2020;11:613546 [FREE Full text] [doi: 10.3389/fpsyg.2020.613546] [Medline: 33519637]
8. Cobb LK, Godino JG, Selvin E, Kucharska-Newton A, Coresh J, Koton S. Spousal influence on physical activity in middle-aged and older adults: The ARIC study. Am J Epidemiol 2016;183(5):444-451 [FREE Full text] [doi: 10.1093/aje/kwv104] [Medline: 26337074]
9. Carr RM, Prestwich A, Kwasnicka D, Thøgersen-Ntoumani C, Gucciardi DF, Quested E, et al. Dyadic interventions to promote physical activity and reduce sedentary behaviour: systematic review and meta-analysis. Health Psychol Rev 2019;13(1):91-109 [FREE Full text] [doi: 10.1080/17437199.2018.1532312] [Medline: 30284501]
10. Di Maio S, Villinger K, Knoll N, Scholz U, Stadler G, Gawrilow C, et al. Compendium of dyadic intervention techniques (DITs) to change health behaviours: a systematic review. Health Psychol Rev 2024;18(3):538-573 [FREE Full text] [doi: 10.1080/17437199.2024.2307534] [Medline: 38437798]
11. Richards EA, Franks MM, McDonough MH, Porter K. 'Let's move:' a systematic review of spouse-involved interventions to promote physical activity. International Journal of Health Promotion and Education 2017;56(1):51-67. [doi: 10.1080/14635240.2017.1415160]
12. Gellert P, Ziegelmann JP, Warner LM, Schwarzer R. Physical activity intervention in older adults: does a participating partner make a difference? Eur J Ageing 2011;8(3):211 [FREE Full text] [doi: 10.1007/s10433-011-0193-5] [Medline: 28798651]
13. Ramchand R, Ahluwalia SC, Xenakis L, Apaydin E, Raaen L, Grimm G. A systematic review of peer-supported interventions for health promotion and disease prevention. Prev Med 2017;101:156-170. [doi: 10.1016/j.ypmed.2017.06.008] [Medline: 28601621]
14. Rook KS, August KJ, Sorkin D. Social network functions and health. In: Handb Stress Sci Biol Psychol. Princeton, New Jersey: Health Springer Publishing Company; 2011:123-135.
15. Radtke T, Luszczynska A, Schenkel K, Biddle S, Scholz U. A cluster randomized controlled trial comparing the effectiveness of an individual planning intervention with collaborative planning in adolescent friendship dyads to enhance physical activity (TWOgether). BMC Public Health 2018;18(1):911 [FREE Full text] [doi: 10.1186/s12889-018-5818-6] [Medline: 30041603]
16. Cohen S. Social relationships and health. Am Psychol 2004;59(8):676-684. [doi: 10.1037/0003-066X.59.8.676] [Medline: 15554821]
17. Feeney BC, Collins NL. A new look at social support: a theoretical perspective on thriving through relationships. Pers Soc Psychol Rev 2015;19(2):113-147 [FREE Full text] [doi: 10.1177/1088868314544222] [Medline: 25125368]
18. Schwarzer R, Knoll N. Social support. In: Kaptein A, Weinman J, editors. Health Psychol 2nd ed Great Britain. Oxford: Blackwell; 2010:283-293.
19. Kouvonen A, De Vogli R, Stafford M, Shipley MJ, Marmot MG, Cox T, et al. Social support and the likelihood of maintaining and improving levels of physical activity: the whitehall II study. Eur J Public Health 2012;22(4):514-518 [FREE Full text] [doi: 10.1093/eurpub/ckr091] [Medline: 21750013]
20. Rackow P, Berli C, Lüscher J, Luszczynska A, Scholz U. Emotional or instrumental support? Distinct effects on vigorous exercise and affect. Psychology of Sport and Exercise 2017;33:66-74. [doi: 10.1016/j.psychsport.2017.07.011]
21. Scarapicchia TMF, Amireault S, Faulkner G, Sabiston CM. Social support and physical activity participation among healthy adults: a systematic review of prospective studies. International Review of Sport and Exercise Psychology 2016;10(1):50-83. [doi: 10.1080/1750984x.2016.1183222]
22. Craddock E, vanDellen MR, Novak SA, Ranby KW. Influence in relationships: A meta-analysis on health-related social control. Basic and Applied Social Psychology 2015;37(2):118-130. [doi: 10.1080/01973533.2015.1011271]

XSL•FO
RenderX

23. Lewis MA, Butterfield RM. Antecedents and reactions to health-related social control. Pers Soc Psychol Bull 2005;31(3):416-427. [doi: 10.1177/0146167204271600] [Medline: 15657456]
24. Rafaeli E, Gleason MEJ. Skilled support within intimate relationships. J of Family Theo & Revie 2009;1(1):20-37. [doi: 10.1111/j.1756-2589.2009.00003.x]
25. Shilts MK, Horowitz M, Townsend MS. Goal setting as a strategy for dietary and physical activity behavior change: a review of the literature. Am J Health Promot 2004;19(2):81-93. [doi: 10.4278/0890-1171-19.2.81] [Medline: 15559708]
26. Sniehotta F, Schwarzer R, Scholz U, Schüz B. Action planning and coping planning for long-term lifestyle change: theory and assessment. Eur. J. Soc. Psychol 2005;35(4):565-576 [FREE Full text] [doi: 10.1002/ejsp.258]
27. Wiedemann AU, Lippke S, Reuter T, Ziegelmann JP, Schwarzer R. How planning facilitates behaviour change: Additive and interactive effects of a randomized controlled trial. Eur. J. Soc. Psychol 2011;41(1):42-51. [doi: 10.1002/ejsp.724]
28. Szczuka Z, Kulis E, Boberska M, Banik A, Kruk M, Keller J, et al. Can individual, dyadic, or collaborative planning reduce sedentary behavior? A randomized controlled trial. Soc Sci Med 2021;287:114336 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114336] [Medline: 34482277]
29. Burkert S, Scholz U, Gralla O, Roigas J, Knoll N. Dyadic planning of health-behavior change after prostatectomy: a randomized-controlled planning intervention. Soc Sci Med 2011;73(5):783-792. [doi: 10.1016/j.socscimed.2011.06.016] [Medline: 21807446]
30. Prestwich A, Conner M, Lawton R, Bailey W, Litman J, Molyneaux V. Individual and collaborative implementation intentions and the promotion of breast self-examination. Psychology & Health 2005;20(6):743-760. [doi: 10.1080/14768320500183335]
31. Katzenelenbogen O, Knoll N, Stadler G, Bar-Kalifa E. The role of individual and dyadic planning in couples' daily goal pursuits. Pers Soc Psychol Bull 2022;48(2):239-253. [doi: 10.1177/0146167221997630] [Medline: 33783241]
32. Keller J, Fleig L, Hohl DH, Wiedemann AU, Burkert S, Luszczynska A, et al. Which characteristics of planning matter? Individual and dyadic physical activity plans and their effects on plan enactment. Soc Sci Med 2017;189:53-62. [doi: 10.1016/j.socscimed.2017.07.025] [Medline: 28783502]
33. Keller J, Hohl DH, Hosoya G, Heuse S, Scholz U, Luszczynska A, et al. Long-term effects of a dyadic planning intervention with couples motivated to increase physical activity. Psychology of Sport and Exercise 2020;49:101710. [doi: 10.1016/j.psychsport.2020.101710]
34. Knoll N, Hohl DH, Keller J, Schuez N, Luszczynska A, Burkert S. Effects of dyadic planning on physical activity in couples: A randomized controlled trial. Health Psychol 2017;36(1):8-20. [doi: 10.1037/hea0000423] [Medline: 27642760]
35. Kulis E, Szczuka Z, Keller J, Banik A, Boberska M, Kruk M, et al. Collaborative, dyadic, and individual planning and physical activity: A dyadic randomized controlled trial. Health Psychol 2022;41(2):134-144. [doi: 10.1037/hea0001124] [Medline: 34968130]
36. Prestwich A, Conner MT, Lawton RJ, Ward JK, Ayres K, McEachan RRC. Randomized controlled trial of collaborative implementation intentions targeting working adults' physical activity. Health Psychol 2012;31(4):486-495. [doi: 10.1037/a0027672] [Medline: 22468716]
37. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019;16(1):31 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
38. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
39. Wang L, Miller LC. Just-in-the-moment adaptive interventions (JITAI): A meta-analytical review. Health Commun 2020;35(12):1531-1544. [doi: 10.1080/10410236.2019.1652388] [Medline: 31488002]
40. OSF. URL: https://osf.io/d98yp/ [accessed 2026-01-15]
41. Time and Ties: Intervention Effects. OSF. URL: https://osf.io/q3vmu/ [accessed 2026-01-15]
42. Scholz U, Berli C. A dyadic action control trial in overweight and obese couples (DYACTIC). BMC Public Health 2014;14:1321 [FREE Full text] [doi: 10.1186/1471-2458-14-1321] [Medline: 25540972]
43. Berli C, Schwaninger P, Scholz U. "We Feel Good": daily support provision, health behavior, and well-being in romantic couples. Front Psychol 2020;11:622492 [FREE Full text] [doi: 10.3389/fpsyg.2020.622492] [Medline: 33536986]
44. Sasaki J, John D, Freedson P. Validation and comparison of ActiGraph activity monitors. J Sci Med Sport 2011;14(5):411-416. [doi: 10.1016/j.jsams.2011.04.003] [Medline: 21616714]
45. Choi L, Liu Z, Matthews C, Buchowski M. Validation of accelerometer wear and nonwear time classification algorithm. Med Sci Sports Exerc 2011;43(2):357-364 [FREE Full text] [doi: 10.1249/MSS.0b013e3181ed61a3] [Medline: 20581716]
46. Tracy JD, Acra S, Chen KY, Buchowski MS. Identifying bedrest using 24-h waist or wrist accelerometry in adults. PLoS One 2018;13(3):e0194461 [FREE Full text] [doi: 10.1371/journal.pone.0194461] [Medline: 29570740]
47. Bolger N, Laurenceau JP. Intensive Longitudinal Methods: an Introduction to Diary and Experience Sampling Research. New York: Guilford Press; 2013.
48. Rothman AJ, Sheeran P. The operating conditions framework: Integrating mechanisms and moderators in health behavior interventions. Health Psychol 2021;40(12):845-857. [doi: 10.1037/hea0001026] [Medline: 32914997]

XSL•FO
RenderX

49. Ledermann T, Kenny DA. Analyzing dyadic data with multilevel modeling versus structural equation modeling: A tale of two methods. J Fam Psychol 2017;31(4):442-452. [doi: 10.1037/fam0000290] [Medline: 28165269]
50. Boruvka A, Almirall D, Witkiewitz K, Murphy SA. Assessing time-varying causal effect moderation in mobile health. J Am Stat Assoc 2018;113(523):1112-1121 [FREE Full text] [doi: 10.1080/01621459.2017.1305274] [Medline: 30467446]
51. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020. URL: https://www.R-project.org/ [accessed 2025-12-16]
52. Posit Team. RStudio: integrated development environment for R. Posit Software, PBC. 2023. URL: http://www.posit.co/ [accessed 2025-12-16]
53. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. 2022. URL: https://CRAN.R-project.org/package=nlme [accessed 2025-12-16]
54. Qian T, Walton AE, Collins LM, Klasnja P, Lanza ST, Nahum-Shani I, et al. The microrandomized trial for developing digital interventions: experimental design and data analysis considerations. Psychol Methods 2022;27(5):874-894 [FREE Full text] [doi: 10.1037/met0000283] [Medline: 35025583]
55. Halekoh U, Højsgaard S, Yan J. The R package geepack for generalized estimating equations. J Stat Softw 2006;15(2). [doi: 10.18637/jss.v015.i02]
56. Rothman AJ, Simpson JA, Huelsnitz CO, Jones RE, Scholz U. Integrating intrapersonal and interpersonal processes: a key step in advancing the science of behavior change. Health Psychol Rev 2020;14(1):182-187. [doi: 10.1080/17437199.2020.1719183] [Medline: 31959071]
57. Sniehotta FF. Towards a theory of intentional behaviour change: plans, planning, and self-regulation. Br J Health Psychol 2009;14(Pt 2):261-273. [doi: 10.1348/135910708X389042] [Medline: 19102817]
58. Hagger MS, Luszczynska A. Implementation intention and action planning interventions in health contexts: state of the research and proposals for the way forward. Appl Psychol Health Well Being 2014;6(1):1-47. [doi: 10.1111/aphw.12017] [Medline: 24591064]
59. Jackson SE, Steptoe A, Wardle J. The influence of partner's behavior on health behavior change: the English longitudinal study of ageing. JAMA Intern Med 2015;175(3):385-392. [doi: 10.1001/jamainternmed.2014.7554] [Medline: 25599511]
60. Hoffman RK, Jobe MC, Dodge T. A brief self-persuasion intervention to strengthen health-promoting dietary intentions through autonomous motivation. Appetite 2023;180:106371. [doi: 10.1016/j.appet.2022.106371] [Medline: 36402411]
61. Nelson SK, Layous K, Cole SW, Lyubomirsky S. Do unto others or treat yourself? The effects of prosocial and self-focused behavior on psychological flourishing. Emotion 2016;16(6):850-861 [FREE Full text] [doi: 10.1037/emo0000178] [Medline: 27100366]
62. Liu Y, Kornfield R, Shaw BR, Shah DV, McTavish F, Gustafson DH. Giving and receiving social support in online substance use disorder forums: how self-efficacy moderates effects on relapse. Patient Educ Couns 2020;103(6):1125-1133 [FREE Full text] [doi: 10.1016/j.pec.2019.12.015] [Medline: 31901364]
63. Marasso D, Lupo C, Collura S, Rainoldi A, Brustio PR. Subjective versus objective measure of physical activity: A systematic review and meta-analysis of the convergent validity of the physical activity questionnaire for children (PAQ-C). Int J Environ Res Public Health 2021;18(7):3413 [FREE Full text] [doi: 10.3390/ijerph18073413] [Medline: 33806106]
64. Reilly JJ, Penpraze V, Hislop J, Davies G, Grant S, Paton JY. Objective measurement of physical activity and sedentary behaviour: review with new data. Arch Dis Child 2008;93(7):614-619. [doi: 10.1136/adc.2007.133272] [Medline: 18305072]
65. Montoye AH, Clevenger KA, Pfeiffer KA, Nelson MB, Bock JM, Imboden MT, et al. Development of cut-points for determining activity intensity from a wrist-worn ActiGraph accelerometer in free-living adults. J Sports Sci 2020;38(22):2569-2578. [doi: 10.1080/02640414.2020.1794244] [Medline: 32677510]
66. Hawlader MDH, Mozid N, Sharmin S, Monju IH, Ahmed SB, Sarker W, et al. The art of forming habits: applying habit theory in changing physical activity behaviour. J Public Health (Berl.) 2022;31(12):2045-2057. [doi: 10.1007/s10389-022-01766-4]
67. Riccio MT, Shrout PE, Balcetis E. Interpersonal pursuit of intrapersonal health goals: Social cognitive–motivational mechanisms by which social support promotes self‐regulatory success. Social & Personality Psych 2019;13(10):e12495. [doi: 10.1111/spc3.12495]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
DBCT: dyadic behavior change technique
JITAI: just-in-time adaptive intervention
MVPA: moderate-to-vigorous physical activity
OSF: Open Science Framework

XSL•FO
RenderX

Edited by L Buis; submitted 07.Oct.2024; peer-reviewed by M vanDellen, J Novak, LO Wilhelm, H Namba; comments to author 20.Jun.2025; accepted 15.Nov.2025; published 27.Jan.2026.

Please cite as:
Höhener PS, Tobias R, Allen JM, Küng P, Scholz U
Effectiveness of Smartphone-Based Dyadic Interventions to Increase Physical Activity in Romantic Couples: Microrandomized Trial
JMIR Mhealth Uhealth 2026;14:e67136
URL: https://mhealth.jmir.org/2026/1/e67136
doi:10.2196/67136
PMID:

©Patrick Stefan Höhener, Robert Tobias, James Martin Allen, Pascal Küng, Urte Scholz. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 27.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Adolescents' Engagement With an mHealth Multiple Health Behavior Change Intervention (LIFE4YOUth): Mixed Methods and Qualitative Comparative Analysis

Anna Seiterö[1], PhD; Pontus Henriksson[1], PhD; Kristin Thomas[1,2], PhD; Marie Löf[2], PhD; Marcus Bendtsen[1], PhD; Ulrika Müssener[1], PhD

[1]Department of Health, Medicine, and Caring Sciences, Linköping University, Linköping, Östergötland, Sweden

[2]Department of Medicine, Huddinge, Karolinska Institutet, Stockholm, Sweden

Corresponding Author:
Anna Seiterö, PhD
Department of Health, Medicine, and Caring Sciences
Linköping University
Campus US
Linköping, Östergötland, SE-581 83
Sweden
Phone: 46 700 89 53 76
Email: anna.seitero@liu.se

Abstract

Background: Behavior change interventions delivered through mobile phones often have low engagement among end users.

Objective: This study aimed to explore factors influencing engagement among Swedish high school students with access to LIFE4YOUth, a mobile-based multiple behavior change intervention targeting physical activity, diet, alcohol consumption, and cigarette smoking. Special emphasis was placed on understanding low engagement.

Methods: A sequential explanatory mixed methods design was used. Quantitative usage data from 377 students were analyzed to describe engagement patterns. This was followed by qualitative data collection through 3 focus groups and 2 individual interviews (n=20), analyzed using inductive content analysis. Finally, qualitative comparative analysis (QCA) was used to integrate findings and identify configurations of psychosocial and behavioral conditions associated with low engagement. The results from all phases were interpreted and discussed as a whole.

Results: A majority (253/377, 67%) of participants showed low engagement, with 62% (158/253) never interacting with the intervention beyond receiving weekly SMS text messaging. Focus group discussions revealed 3 overarching categories influencing engagement: perceived importance of behavior change, user experiences, and environment of use. In total, 48% (121/253) of the low-engaged participants were represented by 1 of 3 configurations, which described participants' characteristics as unmotivated high-needers, motivated low-needers, and dissatisfied needers. Robustness tests confirmed the stability of the unmotivated high-needers configuration.

Conclusions: LIFE4YOUth (Linköping University) did not engage high school students with multiple risk behaviors who were content with their lives and did not consider healthy behaviors as very important. However, positive experiences of being both confirmed and encouraged may explain engagement among students engaged in a combination of health-risk and health-promoting behaviors. Future research could explore how tailoring the number of behaviors targeted by mHealth interventions for adolescents might increase engagement and, in turn, behavioral outcomes.

(JMIR Mhealth Uhealth 2026;14:e88054) doi:10.2196/88054

KEYWORDS

mHealth; adolescents; health behaviors; user engagement; mixed methods; qualitative comparative analysis

XSL•FO
RenderX

Introduction

Health-risk behaviors among adolescents, such as insufficient physical activity, poor diet, heavy episodic drinking, and cigarette smoking, call for effective strategies to increase adolescents' engagement in healthy behaviors [1-3]. Research shows that interventions that use mobile health (mHealth) technology have the potential to influence adolescents' health behaviors [4-6]. This finding is promising because it implies that many adolescents can be reached concurrently at low costs regardless of geographical location [7].

Despite the potential of mHealth to reach adolescents anytime and anywhere while preserving their integrity in relation to peers and professionals [8], evidence indicates that it is often challenging to engage adolescents with mHealth tools [9,10]. The mHealth engagement typically includes behavioral, cognitive, and emotional aspects of the interaction between the user (eg, adolescent) and the technical platform, such as a mobile phone health app. Although these dimensions of engagement interact [11], the behavioral aspect has received the most attention in the scientific literature to date [12,13]. For example, studies that rely on usage data combined with user demographic data to investigate factors associated with, and potentially predictive of, intervention usage in general [14] and effective engagement in particular [10,15,16].

However, aside from the limited number of empirical studies that seek to understand and explain engagement patterns from broader perspectives, including cognitive and emotional aspects of engagement, several important knowledge gaps remain. First, since the use of multiple behavior change approaches to tackle the co-occurrence of health risk behaviors is still in its infancy, there is limited knowledge about user engagement with mHealth interventions that target multiple behaviors simultaneously [17,18]. Second, the existing evidence on contextual factors associated with mHealth engagement is primarily based on adult populations [17,19,20], underscoring the need for studies focusing specifically on adolescents' engagement with mHealth behavior change interventions [13]. To address these gaps, studies have been conducted to examine design features associated with sustained engagement [21,22] and to identify effective strategies for increasing engagement among adolescent users of mHealth behavior change tools [23]. However, knowledge about what influences and explains engagement patterns among adolescents with access to an mHealth multiple behavior change intervention is still limited.

Consequently, this mixed methods study aimed to investigate explanations for engagement among high school students with access to an mHealth multiple behavior change intervention (LIFE4YOUth; Linköping University), with a specific emphasis on understanding low engagement. The following research questions (RQs) were examined:

- RQ1: What characterizes high school students' engagement with the LIFE4YOUth intervention during a 16-week intervention period?
- RQ2: What conditions are plausible explanations for high school students' engagement with the LIFE4YOUth intervention?
- RQ3: Which configurations of psychosocial and behavioral conditions may explain low engagement?

Methods

Study Design

This study was part of a project that overall aimed to evaluate the effects of a multiple behavior change mHealth intervention (LIFE4YOUth) targeting Swedish high school students [24]. A sequential 3-phased explanatory mixed methods design [25] with qualitative comparative analysis (QCA) [26] was used to understand engagement patterns. Integration occurred at each stage of the study. In phase 1 (RQ1), baseline and usage data were analyzed using descriptive statistics to give an overall idea of how high school students engaged with LIFE4YOUth during a 16-week intervention period. Results from Phase 1 were used to develop an interview guide and diagrams depicting engagement patterns for the sequential phase. In phase 2 (RQ2), high school students participating in focus groups were encouraged to interpret the depicted engagement patterns through discussion and reflection based on their experiences with LIFE4YOUth. An inductive approach of qualitative content analysis was used for analyzing the qualitative data [27]. Explanations for low engagement discussed among focus group participants were thereafter used to conceptualize and build the conditions that were further investigated in Phase 3. In phase 3 (RQ3), conditions revealed from phase 2 were systematically compared using QCA to identify combinations of conditions (ie, configurations) that specifically occurred among low-engaged participants. In line with the sequential explanatory mixed methods design, the qualitative and quantitative results were ultimately integrated and interpreted together to provide a comprehensive understanding of explanations for engagement among high school students with access to an mHealth multiple behavior change intervention (LIFE4YOUth). All procedures are reported in accordance with the COREQ (Consolidated Criteria for Reporting Qualitative Research) 32-item checklist [28] and the Good Reporting of A Mixed Methods Study (GRAMMS) checklist [29].

LIFE4YOUth Intervention

The LIFE4YOUth intervention was developed as part of the Mobile health Multiple lifestyle Behaviors Interventions across the Lifespan (MoBILE) research program, which designs and evaluates mHealth interventions tailored for different populations [30]. LIFE4YOUth was designed to promote healthy behaviors among Swedish high school students. Three intervention components were central: (1) a weekly screening and feedback feature; (2) a dashboard with factual information and exercises for all health behaviors; and (3) an SMS text messaging service that can be activated by end users that includes automatic SMS text messaging programs for each health behavior (physical activity, diet, alcohol consumption, and cigarette smoking), as described in detail elsewhere [24,31]. All content was tailored to the target group [32] and based on recommendations provided by the National Board of Health and Welfare in Sweden [33], previous studies [34,35], and health behavior theories [36,37]. The LIFE4YOUth trial indicated a potential of LIFE4YOUth to increase participants' weekly time

XSL•FO
RenderX

spent in moderate to vigorous physical activity after 2 and 4 months and to a small extent increase their short-term intake of fruit and vegetables. The effects on alcohol consumption and cigarette smoking were less convincing [31].

Phase 1: Characterizing Engagement Patterns (RQ1)

Data Collection

Quantitative data were collected from participants (n=377) in the intervention arm of the LIFE4YOUth trial [31]. The trial participants were recruited from high schools across Sweden, including vocational training and education preparing students for higher studies. The eligibility criteria were being a high school student, not meeting health recommendations for either physical activity, diet, alcohol consumption, or cigarette smoking, and owning a mobile phone. No exclusion criteria were applied. A total of 1398 students registered interest in the study, of whom 890 (64%) consented to participate, and 760 completed the baseline questionnaire, which was used to assess eligibility. Four (0.5%) students did not meet the eligibility criteria and were therefore excluded. All participants were enrolled in the trial between September 2020 and June 2023.

Engagement Measures

Three engagement measures were used to describe participants' engagement with LIFE4YOUth: (1) activation of the SMS text messaging program to support behavior change (yes or no), (2) the number of weeks completing the weekly screening assessment (1-16), and (3) the total number of sessions interacting with the content, defined as a continuous series of screen views initiated by actively clicking on symbols to navigate between dashboard content (zero to infinity). A new screen view after 30 minutes of inactivity was considered a new session [38].

Explanatory Measures

Explanatory measures referred to baseline variables that were assumed to influence participants' engagement with the intervention. These included 4 psychosocial and 7 health behavioral variables. The choice of psychosocial variables was based on common constructs in behavior change theories [36,37] assumed to mediate the effects of LIFE4YOUth [39]. Importance, confidence, and know-how were generally measured using the following questions: (1) "How important do you think it is to improve your lifestyle or sustain your healthy behaviors?"; (2) "How confident are you that you will be able to change your lifestyle or sustain your healthy behaviors?"; (3) "How well do you know how to change your health behaviors?" Responses were given on a 10-point Likert scale. As adolescents' satisfaction with life is also associated with their health behaviors [40,41], life satisfaction was also included as a psychosocial variable; and assumed to influence engagement. Life satisfaction was measured using Cantril's ladder [42]: (4) "Thinking about your own life and personal circumstances, how satisfied are you with your life as a whole?" Responses were recorded on an 11-point Likert scale ranging from 0 (not at all satisfied) to 10 (completely satisfied).

The behavioral variables assumed to influence engagement included (1) weekly time (minutes) spent in moderate to vigorous physical activity (MVPA); (2) number of daily portions (100 g) of fruit and vegetables; (3) number of weekly portions of candy or snacks; (4) number of weekly sugary drinks (33 cL) consumed; (5) monthly frequency of heavy episodic drinking (ie, ≥ 4 standard drinks at 12 g pure alcohol on one occasion); (6) number of weekly standard drinks of alcohol consumed; and (7) smoking any cigarette in the past week (yes or no). All variables except for smoking were measured numerically through self-reported data, as described in detail elsewhere [31].

Data Analysis

The k-means function in R (version 1.3.1073; R Core Team) was used to create 3 engagement clusters (high, medium, and low) based on participants' (1) activation of text-message programs, (2) the number of weekly screening assessments, and (3) the number of sessions interacting with the dashboard content. Two-sample *t* tests and chi-square tests were used to test for statistically significant differences between clusters with respect to baseline characteristics, using a significance level of 0.05. The k-means clustering was also used to identify crossover points for all psychosocial variables included in the QCA.

Phase 2: Identifying Conditions Explaining Engagement Patterns (RQ2)

Data Collection

High school students participating in the LIFE4YOUth trial [31] were invited to participate in a focus group interview. Two high schools in the region of Östergötland invited students to focus groups using a convenient sampling strategy [27]. The schools were chosen based on their different educational profiles and the fact that contact with staff was established. Students were recruited either by written information provided by the school nurse or by oral information from teachers. All students in 2 classes were invited to participate during school hours. Three students within these classes declined participation. A total of 3 focus groups and 2 individual interviews with a total of 20 high school students were conducted by AS (a female PhD candidate and experienced school nurse) and UM (a female senior associate professor with expertise in qualitative methodology). No one other than the students and moderators, who were not familiar with the participants, attended the focus groups.

Both interviews and focus groups, which involved 4-8 participants, were audio-recorded and conducted at the participants' school during school hours between February 2023 and April 2023. A semistructured interview guide with predefined themes was used to facilitate the focus groups. The discussions were initiated by showing participants diagrams illustrating engagement patterns (derived from Phase 1) and asking them to reflect on the factors that might explain each pattern based on their experiences with LIFE4YOUth. The duration of the sessions varied between 33 and 46 (mean 40, SD 6) minutes. Field notes were written directly after each session.

Data Analysis

The audio-recorded data were transcribed verbatim by professionals and analyzed inductively using qualitative content

XSL•FO
RenderX

analysis [27]. All transcripts were carefully read by AS and UM and checked for accuracy against the audio recordings (AS). The coding procedure was performed using the software (NVivo 12 Plus; QSR International Pty Ltd) to organize similar content in the data using preliminary labels (ie, codes; AS). After establishing a coding scheme, all transcripts were reread to systematically code all relevant text units (AS). Text extracts for each code were read and reread to get an idea of similarities and differences among data within codes and between codes. Codes with similar content were combined to create categories, which were checked against the transcripts to verify meaningfulness and to ensure that all relevant data were captured by the categories. UM was involved in all analytical procedures through repeated discussions until consensus was reached.

Phase 3: Identifying Combinations of Conditions Explaining Low Engagement (RQ3)

Building Integration Using QCA

A crisp set QCA [26,43] was conducted to explore whether the conditions derived from the qualitative findings helped explain the low-engagement cluster identified in Phase 1. Crisp sets were used because of our interest in exploring whether the presence rather than the degree (ie, fuzzy set) of a condition was related to engagement. While the qualitative findings from Phase 2 informed the selection and definition of conditions, k-means clustering was used to determine the crossover points used to calibrate each condition. All participants at or above the crossover point were coded as 1 (fully in the set), while cases below the crossover point were coded as 0 (fully out). For negatively framed conditions, the calibration was reversed: cases at or below the crossover point were coded as 1, and cases above the crossover point were coded as 0. A truth table [26,44] was constructed to obtain an overview of all possible combinations of conditions (ie, configurations) and their coverage in empirical data. The analysis also included configurations represented by only one participant, which is coherent with good practice of QCA [45,46]. An enhanced standard analysis [47] was carried out by logically minimizing the truth table. A consistency threshold at 0.75 was deemed enough to reveal descriptive patterns in the data. The minimization procedure, as with all other QCA procedures, was performed using the QCA packages (version 3.19) and SetMethods (version 4.0) in R (version 1.3.1073). The algorithm used to minimize the truth table starts at the upper row and systematically works through each truth table row with an outcome value of 1 to determine whether a condition can be eliminated from the configuration. The goal of this procedure is to find the briefest way to summarize all available information from the truth table, that is, the most parsimonious solution [26]. All configurations included in the most parsimonious solution indicate alternative pathways to low engagement. All configurations were interpreted holistically based on knowledge gained from the qualitative phase and were labeled to describe the overall characteristics conveyed by each pathway.

Robustness Tests

The analytical procedures were repeated using the negation of the outcome (not low engagement), in accordance with recommendations [45,46]. The model analytics included interpretation of consistency and coverage, evaluating the simplifying assumptions used in the logical minimization (ie, whether possible configurations that were not represented by data were used in the minimization), and assessing model ambiguity (ie, whether the minimization revealed more than one valid way of summarizing the truth table).

Following Oana and Schneider's approach [48], robustness was further assessed using a fit-oriented and case-oriented approach to measure how sensitive the solution was to analytical decisions. The obtained solution was tested against alternative solutions derived from the following analytical decisions using a consistency threshold of 0.8 instead of 0.75, not excluding participants with activated SMS text messaging programs from the low engagement outcome, and using 5 as a frequency threshold (ie, including configurations represented by at least 5 participants in the minimization of the truth table). The obtained solution was compared to the minimal test solution, which is the area where all possible solutions agree, and the maximal test solution, which is the entire area of possible solutions. This comparison was made to determine how much the 2 sets of solutions overlapped. Robustness was also assessed in terms of the proportion of participants that would remain in the alternative solution.

Ethical Considerations

Students were informed about the LIFE4YOUth intervention through printed advertisements distributed in high school settings. Before providing consent to participate in the study, all participants received written and oral information that briefly described the characteristics of the 16-week digital intervention, the purpose and design of the study, the types of data collected, how data will be handled, and that participation is voluntary and without any financial compensation. All participants were further informed that they were free to decide if, how, and when to use the digital intervention, and that they had the right to withdraw from the study at any point without consequences. When applicable, participants were informed about the role of the focus group moderators as part of the research team. Informed consent was collected from all participants (aged 15 years or older) but not from their guardians. All quantitative data were pseudoanonymized using a user identity, which only one co-author (MB) had access to. To further ensure privacy and confidentiality, only those involved in the collection and analysis of focus group data (AS and UM) had access to personal information linked to the data. The procedures to collect and use quantitative engagement data (Swedish Ethical Review Authority, Dnr 2019-03813, Dnr 2020-03538) and qualitative data (Dnr 2022-02493-02) were consistent with the ethical approval from the Regional Ethical Committee in Linköping, Sweden. No financial compensation was offered for participation.

Results

Overview

The characteristics of the study sample are presented in Table 1. The median age of participants was 17 years. Females were overrepresented (265/377, 70%).

XSL•FO
RenderX

Table 1. Overview of participants specified per dataset and engagement pattern.

Participant characteristics	Qualitative (n=20)	Quantitative engagement (n=377)			
		Total (n=377)	Low (n=253)	Medium or high (n=124)	*P* value[a]
Age (years), mean (SD)	17-19[b]	17.18 (1.20)	17.19 (1.19)	17.15 (1.23)	.81
Sex, n (%)					.11
Female	15 (75)	265 (70)	179 (71)	86 (69)	
Male	5 (25)	112 (30)	74 (29)	38 (31)	
Parents' education[c], n (%)					.58
Primary	—[d]	25 (7)	17 (7)	8 (6)	
Secondary	—	131 (35)	90 (36)	41 (33)	
Tertiary	—	221 (59)	146 (56)	75 (60)	
Economy[e], n (%)					.29
Very good	—	112 (30)	76 (30)	36 (29)	
Average	—	232 (62)	158 (62)	74 (60)	
Not so good	—	30 (8)	16 (6)	14 (11)	
Not good at all	—	3 (1)	3 (1)	0 (0)	
Region of birth[f], n (%)					.81
Sweden	—	324 (86)	219 (87)	105 (85)	
Other	—	53 (14)	34 (13)	19 (15)	
Parents' region of birth[f], n (%)					.31
Sweden	—	268 (71)	181 (72)	87 (70)	
Other	—	109 (29)	72 (28)	37 (30)	
Satisfaction with life[g], mean (SD)		6.83 (1.99)	6.91 (1.98)	6.67 (2.03)	.28
Psychosocial variables[h], mean (SD)					
Importance	—	6.66 (2.53)	6.45 (2.58)	7.07 (2.39)	.03
Knowledge	—	6.65 (2.31)	6.99 (2.32)	6.59 (2.27)	.05
Confidence	—	6.24 (2.52)	6.17 (2.54)	6.38 (2.48)	.45
Behavioral variables, mean (SD)					
Weekly time spent in MVPA[i] (minutes)	—	323 (316)	321(310)	326 (329)	.89
Number of daily portions (100 g) of fruit and vegetables consumed	—	1.43 (1.14)	1.37 (1.08)	1.56 (1.26)	.14
Number of weekly sugary drinks (33 cL) consumed	—	3.27 (4.45)	3.53 (4.75)	2.75 (3.73)	.11
Number of weekly portions of candy or snacks consumed	—	4.60 (5.55)	4.29 (5.19)	5.24 (6.19)	.12
Monthly frequency of heavy episodic drinking (ie, ≥4 standard drinks)	—	1.21 (2.86)	1.47 (3.35)	0.69 (1.22)	.01
Number of weekly standard drinks (12 g pure alcohol) consumed	—	1.75 (4.10)	2.03 (4.49)	1.18 (3.11)	.06
Smoking any cigarette in the past week, n (%)	—	68 (18)	53 (21)	15 (12)	.05

[a]Based on 2 sample *t* tests of means and chi-square tests for proportions.

XSL•FO
RenderX

[b]Range for the qualitative dataset.

[c]Assessed as "Please select the highest education for your mother and father."

[d]Not available.

[e]Assessed as "How would you describe the economic situation in your family?"

[f]Assessed as "Where were you/your parents born?"

[g]Cantril's ladder: "Thinking about your own life and personal circumstances, how satisfied are you with your life as a whole?" Responses were given on an 11-point Likert scale where 0 represents not at all satisfied.

[h]Defined as follows: Importance: "How important do you think it is to improve your lifestyle or sustain your healthy behaviors?" Confidence: "How confident are you that you will be able to change your lifestyle or sustain your healthy behaviors?" Knowledge: "How well do you know how to change your health behaviors?" Responses were provided on a 10-point Likert scale.

[i]MVPA: moderate to vigorous physical activity.

The majority were born in Sweden (324/377, 86%) with highly educated Swedish-born parents. An overview of participants' baseline characteristics across all engagement levels is provided in Table S1 in Multimedia Appendix 1.

Phase 1: High School Students' Engagement With LIFE4YOUth (RQ1)

Defining Engagement Patterns

The results from this phase describe the characteristics of participants' engagement with the LIFE4YOUth intervention. The results from the k-means clustering (Table S2 in Multimedia Appendix 1) informed the thresholds used to define 3 engagement patterns. Low engagement was defined as 2 or fewer weeks (out of 16) of completing the weekly screenings, a maximum of 3 sessions of engagement with the dashboard content, and no activated SMS text messaging program. High engagement was defined as completing the weekly screening in at least 9 weeks, or by 12 or more engagement sessions during the intervention period, or by having activated at least one SMS text messaging program. Engagement between low and high was defined as a medium level of engagement.

Engagement With the LIFE4YOUth Intervention

Among participants who engaged with LIFE4YOUth at least once (219/377, 58%) during the 16-week intervention period, the weekly screening was completed on average in 4 (SD 4) weeks, and the dashboard content was accessed on average 5 (SD 5.5) sessions. The proportion of participants who completed the weekly screening assessment was almost identical to the proportion per week who engaged with the dashboard content. Twenty-one percent of participants (45/219) who engaged at least once activated one or more SMS text messaging programs focusing on physical activity, diet, alcohol consumption, or cigarette smoking. Table 2 presents an overview of engagement activities by engagement cluster.

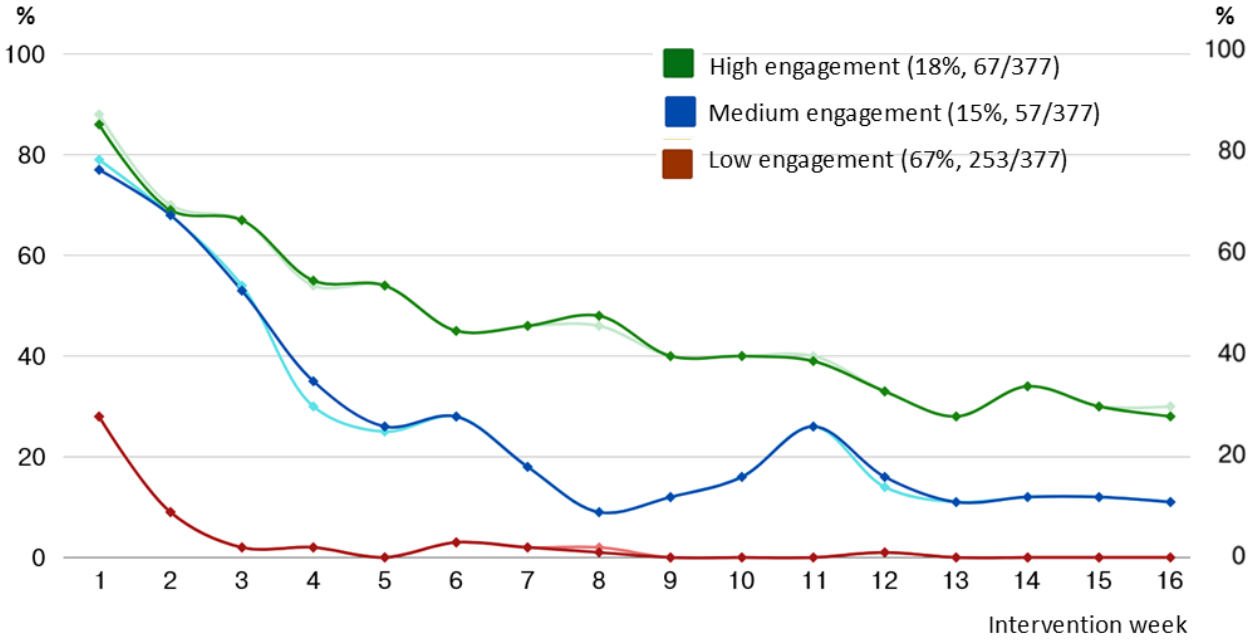
In total, 18% (67/377) were defined as high-engaged, 15% (57/377) were defined as medium-engaged, and 67% (253/377) were defined as low-engaged (Figure 1). Sixty-two percent (158/253) of the low-engaged participants never engaged actively. At baseline, low-engaged participants reported significantly higher frequency of heavy episodic drinking, lower levels of perceived importance, and higher levels of knowledge of how to change their behaviors. These differences were investigated further in the subsequent phases to gain a better understanding of what explains engagement in general (Phase 2) and low engagement in particular (Phase 3). Engagement patterns were similar throughout the study period, with no statistically significant differences during or after the COVID-19 pandemic (Table S3 in Multimedia Appendix 1).

Table 2. Cluster-specific engagement measures.

Engagement measures	Low (n=253)	Medium (n=57)	High (n=67)
SMS text messaging programs, n (%)	0 (0)	0 (0)	45 (67)
Weekly screenings, mean (SD)	0.49 (0.69)	4.25 (1.66)	7.46 (5.28)
Session episodes, mean (SD)	0.54 (0.79)	5.35 (2.37)	9.75 (7.40)

XSL•FO
RenderX

Figure 1. Proportion of participants who completed the weekly screening (darker color) and engaged with the dashboard content (lighter color), by engagement cluster and intervention week.



Phase 2: Conditions Explaining High School Students' Engagement With LIFE4YOUth (RQ2)

Categories of Conditions That Overall Influence Engagement

The results from this phase describe conditions identified among a subset of participants (n=20) from phase 1 who took part in focus groups to provide insights and help interpret the previously identified engagement patterns. Three categories of plausible explanations for engagement with LIFE4YOUth were discussed among focus group participants: (1) perceived importance of health behavior change, (2) user experiences, and (3) environment of use. Descriptions of these categories are followed by suggestions regarding how these conditions were perceived to explain medium and high engagement as well as low engagement.

Perceived Importance of Health Behavior Change

Focus group participants considered the perceived importance of healthy behaviors as a product based on their interest, willingness, and intention to act to accomplish change. Experiences of problems, such as being dissatisfied with life or oneself, were considered reasons for being interested in behavior change. However, the components of willingness and intention were perceived to be influenced by students' beliefs about their ability to cope with the required efforts given the current circumstances in their lives. Thus, the perceived importance of health behavior change as a condition for engagement incorporates confidence in behavior change. Indeed, participants stressed that one can be both interested and have a desire to change something but lack the confidence to make changes, a constraint that impacts perceptions about the importance of a behavior change, as illustrated in the quotations below:

> *When one isn't doing well, one may easily think of oneself as incapable and avoid trying to establish new habits because being too busy thinking about other things and doesn't want any changes when one is at least somewhat happy.* [Female]

> *It's all about your attitude to change behaviors, [. . .] to notice the tips and acknowledge what they might do for you, you need to have a desire to be helped.* [Male]

> *Well, the first two times I used it [LIFE4YOUth], I browsed and read. After a while, I realized that there's nothing I want to actively change right now.* [Female]

User Experiences

The discussions among participants revealed that the experience of gaining something expected from the intervention—for example, gaining knowledge and adequate support—is critical for an overall positive user experience. The conditions the user experiences further include reactions elicited by the intervention, such as positive emotions of inspiration or negative emotions of defense. Emotions evoked can vary depending on current behaviors and social values. Indeed, some participants perceived that even study content that was not part of the intervention (eg, baseline weight question) elicited negative emotions:

> *It [study content] can be a reminder to always have to think about it, and when life struggles on and you know that 'yeah, I am aware about what I am doing may not be the best, but I don't have time to think about it'.* [Female]

> *Participant 1: I think that people always seek validation [. . .], it doesn't have to be something material, but as you said about feedback . . . Receiving feedback, such as grades or whatever, [. . .] is crucial, because if you do all the things and get nothing in return, then you will probably not . . . Participant 2: I also think that you want to gain something from it*

XSL•FO
RenderX

[the intervention]. Such as knowledge or, as you said, validation, and insights on how to improve oneself. [Male]

Environment of Use

The condition environment of use refers to social circumstances that influence when and how high school students engage with LIFE4YOUth. The participants emphasized that many things require high school students' attention that compete with how much time they spend with LIFE4YOUth. In fact, participants argued that it may be merely a lucky coincidence if a LIFE4YOUth SMS text messaging is received at a suitable time, which may slightly increase the likelihood of engaging with the intervention by clicking on the link in the SMS text messaging. On the other hand, receiving the SMS text messaging while busy means that the message will likely be forgotten:

Depending on when it is received [the LIFE4YOUth text message], one has a varied amount of time to actually respond to the questions. If I was studying at home and received a [LIFE4YOUth] text message, [. . .] I would likely take a break. But if you're with friends and get a text-message, it's not like, 'Yeah, chill, I should just respond to some questions.' [Male]

There are other [mobile-phone health] apps that only aims at smoking cessation [. . .], but this app [LIFE4YOUth] should aid . . . it serves several purposes. However, the other app may only have one function, and that is to make you stop smoking. [Female]

Suggested Explanations for Medium or High Engagement With LIFE4YOUth

Participants reasoned that when one is satisfied with life, one tends to be happier, more confident, curious, and interested in improving oneself. Satisfied people who are engaged in some behaviors considered healthy and in some behaviors considered risky can experience that LIFE4YOUth confirms their healthy behaviors and encourages them to avoid their risky behaviors. The experience of being validated and encouraged can spark ambitions and increase the likelihood of end users continuing to be engaged with LIFE4YOUth. Thus, positive reactions were considered a mechanism that explains engagement. Furthermore, the content provides tips for the adaptation of health-promoting behaviors. According to the participants, when experiencing the tips as desirable and helpful based on a learning need, end users may engage with LIFE4YOUth to a greater extent, indicating that meeting expectations also constitutes a mechanism that drives engagement.

Furthermore, participants thought that end users who are not very happy with their lives and want a change to feel better might consider LIFE4YOUth as a tool worth trying, even when not being convinced that a mobile phone app will help them. Under such circumstances, expectations regarding one's ability to improve health behaviors and the perceived helpfulness of LIFE4YOUth were lower than when the goal was to successfully change those behaviors. That is, the influence of confidence was perceived to be less meaningful in such situations.

Suggested Explanations for Low Engagement With LIFE4YOUth

Participants emphasized that end users with several health-risk behaviors are unlikely to want to be reminded of this on a regular basis. This was deemed to be stressful, that is, a negative reaction to LIFE4YOUth. In addition, participants believed that one may not agree with indicators about health behaviors being considered a health risk. Some participants underlined that behaviors vary from week to week and that their behaviors overall might be adequate. Furthermore, LIFE4YOUth addresses multiple health behaviors, and end users with limited knowledge, or who seek in-depth knowledge about a specific behavior, may find the content too broad. According to participants' reasoning, the risk of being disappointed increases as end users' expectations or needs increase. Higher needs can, for example, depend on current health behaviors or on overall circumstances that impact end users' capabilities to manage their situation.

Two additional explanations for low engagement were raised during the discussions. It was not assumed that end users experiencing life challenges possessed the necessary energy or interest to engage in behavior change. In fact, participants generally found it far too easy to disengage because no one pursued them or even encouraged them, as parents might with their schoolwork. When being satisfied with life, on the other hand, one may have no desire to change anything, and engagement with LIFE4YOUth might not be considered as worthwhile. Furthermore, the weekly SMS text messaging itself served as a reminder about health behaviors. Some participants did not believe that active engagement was therefore needed to benefit from the intervention.

Phase 3: Combinations of Conditions Explaining Low Engagement (RQ3)

Defining Conditions for the QCA

The integrated results from this phase describe how the conditions identified in phase 2 may help explain low engagement within the phase 1 sample (n=377). Based on the discussions among focus group participants, 5 conditions were selected, defined, and included in the QCA. These were (1) engagement in health-risk behaviors; (2) engagement in health-promoting behaviors; (3) being dissatisfied with life; (4) being high-skilled in how to change behaviors; and (5) considering healthy behaviors as very important.

Engagement in health-risk behaviors was defined as smoking ≥1 cigarette in the past week, or (if ≤18 years old) drinking any alcohol in the past week, or (if ≥18 years old) having consumed ≥10 standard drinks of alcohol in the past week, or engaging in heavy episodic drinking ≥1 time in the past month. Health-promoting behaviors were defined as ≥420 minutes of MVPA in the past week or ≥3 daily portions of fruit and vegetables in the past week, combined with either ≤4 portions of candy and cakes in the past week or ≤2 cans (33 cL) of sugary drinks in the past week. These cut-off points align with the thresholds used in the screening and feedback component of LIFE4YOUth. Dissatisfaction with life was defined based on the results of the k-means clustering (≤5 on a 11-point Likert scale). Being highly skilled in how to change health behaviors,

XSL•FO
RenderX

as well as considering healthy behaviors as very important, was defined as ≥8 on a 10-point Likert scale based on the k-means clustering.

Findings From the QCA

Table 3 shows an overview of all 32 possible combinations of conditions (ie, configurations) and their representation among participants in this dataset. The minimization of these configurations and their empirical representation assisted in identifying 3 configurations that contribute to explaining low engagement. No simplifying assumptions were applied to obtain the most parsimonious solution, and no model ambiguity occurred, meaning that none of the configurations associated with low engagement were also associated with non–low-engagement. The most parsimonious solution model was identical to the conservative and intermediate solution models, indicating a high degree of robustness in the identified configurations. The robustness test identified one configuration that remained stable across all tested changes, including participants who were not engaged in health-promoting behaviors but were engaged in health-risk behaviors, did not consider it important to change their health behaviors, and lacked knowledge of how to change them were likely to show low engagement with LIFE4YOUth. Details of the robustness tests are provided in Tables S4 and S5 in Multimedia Appendix 1.

The 26% raw coverage presented in Table 4 indicates that 26% (66/253) of all low-engaged participants demonstrated the configuration labeled unmotivated high needers. Within this configuration, participants engaged in health-risk behaviors but not in health-promoting behaviors, were not dissatisfied with their lives, and did not consider healthy behaviors to be very important. The consistency value of 0.83 means that, of all participants who were described by this configuration, 83% (69/80) were low-engaged. The configurations motivated low needers and dissatisfied needers to describe 2 additional pathways to low engagement. The first pathway reflects being engaged in health-promoting behaviors, considering healthy behaviors to be very important, and not being dissatisfied with life. The second pathway reflects being engaged in health-risk behaviors, not being very skilled in how to change behaviors, and being dissatisfied with life. In total, 48% (121/253) of all low-engaged participants in this dataset were described by one of these configurations, while 19% (29/151) of those described by the configurations were not low-engaged.

XSL•FO
RenderX

Table 3. Truth table, logically possible combinations of conditions in relation to the outcome.

LS[a]	HPB[b]	HRB[c]	IMP[d]	KNOW[e]	OUT[f]	n (%)	Incl[g]
1	1	1	1	0	1	1 (0)	1.000
1	0	1	0	0	1	11 (4)	0.909
0	0	1	0	0	1	57 (23)	0.842
0	1	1	1	1	1	5 (2)	0.800
0	0	1	0	1	1	23 (9)	0.783
0	1	0	1	1	1	13 (5)	0.769
1	0	1	1	0	1	13 (5)	0.769
0	1	0	1	0	1	16 (6)	0.750
0	1	1	1	0	1	8 (3)	0.750
1	1	1	0	0	1	4 (2)	0.750
0	0	1	1	0	0	15 (6)	0.733
0	0	0	0	0	0	39 (15)	0.692
0	1	1	0	0	0	22 (9)	0.667
1	1	1	1	1	0	6 (2)	0.667
0	1	0	0	1	0	8 (3)	0.625
1	0	0	0	0	0	15 (6)	0.600
1	0	0	1	1	0	5 (2)	0.600
1	1	0	0	0	0	5 (2)	0.600
1	0	0	1	0	0	12 (5)	0.583
0	0	1	1	1	0	16 (6)	0.562
0	1	0	0	0	0	9 (4)	0.556
0	0	0	1	1	0	20 (8)	0.550
0	0	0	0	1	0	12 (5)	0.500
0	0	0	1	0	0	24 (9)	0.500
1	0	1	1	1	0	2 (1)	0.500
0	1	1	0	1	0	5 (2)	0.400
1	0	1	0	1	0	3 (1)	0.333
1	0	0	0	1	0	2 (1)	0.000
1	1	0	1	0	0	6 (2)	0.000
1	1	0	0	1	__[h]	0 (0)	—
1	1	0	1	1	—	0 (0)	—
1	1	1	0	1	—	0 (0)	—

[a]LS: being dissatisfied with life (≤5 on a 11-point Likert scale).

[b]HPB: health-promoting behavior (≥420 min MVPA or ≥3 daily portions [100 g] of fruit and vegetables and <4 portions of candy or cakes or <2 cans [33 cL] of sugary drinks).

[c]HRB: health-risk behavior (smoking ≥1 cigarette or being <18 years old and drinking any alcohol or being ≥18 years old and having consumed ≥10 standard drinks [12 g pure alcohol] or being engaged in heavy episodic drinking [≥4 standard drinks on a single occasion] at least once in the past month).

[d]IMP: considering healthy behaviors as very important (≥8 on a 10-point scale).

[e]KNOW: being high-skilled in how to change behaviors (≥8 on a 10-point scale).

[f]OUT: 1=≥75% of all participants within the configuration had low engagement, 0=<75% had low engagement.

[g]Incl: quantifies the degree of a set relationship where 1 indicates a strong relationship between the configuration and the outcome.

[h]Not available.

XSL•FO
RenderX

Table 4. The most parsimonious solution contains three configurations that were associated with low engagement with LIFE4YOUth. Overall consistency: 0.81; overall coverage: 0.48.

Configuration[a]	HRB[b]	HPB[c]	LS[d]	KNOW[e]	IMP[f]	Raw[g]	Unique[h]	Consistency
Unmotivated high-needers	Yes	No	No	Yes or No	No	0.26	0.26	0.83
Motivated low-needers	Yes or No	Yes	No	Yes or No	Yes	0.13	0.13	0.76
Dissatisfied needers	Yes	Yes or No	Yes	No	Yes or No	0.10	0.10	0.83

[a]Labeled according to the overall characteristics of the configurations, as interpreted from the findings reported in the qualitative phase.

[b]HRB: health-risk behavior.

[c]HPB: health-promoting behavior.

[d]LS: Being dissatisfied with life.

[e]KNOW: High-skilled in how to change behaviors.

[f]IMP: Considering healthy behaviors as very important.

[g]Raw: the proportion of participants in the outcome set that were covered by the solution term.

[h]Unique: the proportion of participants in the outcome set that are also in the solution term and not covered by any other solution term.

Discussion

Principal Findings

This mixed methods study aimed to deepen the understanding of the conditions that influence and explain high school students' engagement with an mHealth tool (LIFE4YOUth) targeting multiple health risk behaviors (physical activity, diet, alcohol consumption, and cigarette smoking). Overall, the majority of participants who were given access to the LIFE4YOUth intervention engaged to a low extent (Textbox 1). Participants' perceived importance of behavior change, their experiences with the intervention, and external circumstances that affected their ability to respond to prompts immediately may have influenced their level of engagement. By using participants' reasoning about explanations for low engagement as a basis for defining conditions in the QCA, the integrated results revealed 3 configurations associated with low engagement. These configurations constitute possible pathways for explaining why adolescents with access to a behavior change tool targeting multiple health-risk behaviors hardly ever interact with the tool. In the discussion that follows, these pathways are interpreted in relation to the findings obtained from the focus groups (phase 2) as well as theory and previous research.

Focus group participants suggested that having multiple risk behaviors—such as insufficient engagement in physical activity and unhealthy dietary behaviors alongside alcohol consumption and cigarette smoking—could explain low engagement among individuals who are content with their lives and do not consider behavior change to be important. This expectation was supported by the quantitative data through the configuration of unmotivated high needers. Indeed, 62% (157/253) of the low-engaged participants disengaged without being exposed to any intervention content other than the weekly SMS text messaging, which raises the question of why so many participants did not interact with the dashboard content even once. It is unlikely that this can be explained by participants' experiences of the LIFE4YOUth content. However, it is possible that low-engaged participants had no interest in changing their behaviors but enrolled in the trial for other reasons [49]. Therefore, it is possible that the characteristics shared by the low-engaged participants are similar to those of individuals who declined participation in the first place. Low interest in mHealth behavior change interventions can be understood from a novelty-complexity perspective, as suggested by Crutzen and Ruiter [50], who developed a conceptual model on mechanisms that are assumed to influence people's interest in such tools. People may be interested in behavior change interventions if they perceive the intervention to be novel and the level of complexity of end users' "problems" is neither too high nor too low. High-level complexity problems are thought to reduce people's self-efficacy and interest, whereas low-level complexity problems are thought to be simple enough to manage without the assistance of digital tools. Both scenarios can account for low interest in LIFE4YOUth among motivated low-needers, but also unmotivated high-needers and dissatisfied needers who may lack resources relative to their needs. Therefore, it is possible that focusing on multiple behaviors simultaneously within a single intervention became a barrier for some participants to even consider the idea of behavior change.

Although engagement in multiple health risk behaviors and dissatisfaction with life in combination with other explanatory conditions may explain low engagement in some cases, the reverse relationship may also apply. For example, Thornton et al [51] found no evidence of an association between the number of risk behaviors or psychological distress and adolescents' level of engagement with an app targeting multiple health risk behaviors. This finding is noteworthy, as it resonates with explanations discussed by participants in our focus groups, who suggested that experiencing poor mental well-being might spark interest in using mHealth tools to initiate positive change. This interpretation further aligns with the QCA results for the negated outcome (ie, medium or high engagement; Table S6 in Multimedia Appendix 1), which suggest that dissatisfaction with life, together with nonengagement in health-risk behaviors and other conditions, may be associated with continued engagement in LIFE4YOUth. Thus, adolescents who are struggling in some aspect of their lives may find mHealth lifestyle tools appealing. However, careful tailoring might be needed to empower individuals to believe in their ability to change.

XSL•FO
RenderX

The concept of response efficacy refers to the belief that the "problem" is manageable through the intervention, which is an essential mechanism that can drive continued engagement [50]. Our findings, as well as previous research [9,52,53], show that adolescents who overcome barriers and invest effort in mHealth lifestyle tools may experience these tools as encouraging and remain engaged over time. However, sustained engagement with mHealth tools is not a desired outcome itself [16]. Despite a large proportion of participants in the intervention group of the LIFE4YOUth trial engaging with the content only a few times over the 16-week intervention period, effect estimates [31] indicated positive effects on MVPA after 2 and 4 months and a small 2-month effect on fruit and vegetable intake. Given these effects, the overall level of engagement observed among participants in this study may have been adequate for gaining meaningful insights and reminders that influenced some health behaviors.

Our finding that a large proportion of adolescents engaged with the intervention only a limited number of times further aligns with previous research reporting high levels of nonusage attrition among adolescents using mHealth tools for health promotion [9,52,54]. Peuters et al [9,52] and Maenhout et al [54] explored reasons for engagement and nonengagement with an app targeting physical activity, diet, and sleep among adolescents aged 12-15 years and found that limited interest in behavior change contributed to low engagement. Consequently, it has been suggested that strategies are needed to first increase adolescents' awareness and motivation to change and thereafter integrate mHealth tools into their everyday lives [54,55].

Textbox 1. Joint display summarizing the results of each phase of this mixed methods study.

Summary of quantitative, qualitative, and integrated findings.

Quantitative findings (research question 1 [RQ1])

A total of 67% (253/377) of participants actively engaged with the intervention content ≤3 times over 16 weeks. Low-engaged participants were significantly more likely to report heavy episodic drinking (mean 1.47, SD 3.35; *P*=.01), lower perceived importance of behavior change (mean 6.45, SD 2.58; *P*=.03), and higher knowledge of how to change their behaviors (mean 6.99, SD 2.32; *P*=.05).

Qualitative findings (RQ2)

Three main categories described conditions influencing high school students' engagement with LIFE4YOUth: (1) perceived importance of health behavior change, (2) user experiences, and (3) environment of use. Low engagement was explained by pressure experienced by participants with poor health behaviors and dissatisfaction with life, or alternatively by high life satisfaction and no desire to change behaviors.

Integrated findings (RQ3)

Three configurations associated with low engagement were identified: unmotivated high-needers, motivated low-needers, and dissatisfied needers. These configurations accounted for 48% of all low-engaged participants. Unmeasured circumstances related to user experiences and the environment of usage may explain the remaining cases.

Implications

Our findings on possible explanations for low engagement as well as medium and higher levels of engagement may help other researchers interpret mHealth engagement data gathered from adolescents. Also, the results presented in this study may be useful in the design of future mHealth interventions aimed at promoting healthy behaviors in adolescents, as they indicate that the multiple behavior approach seemed less feasible for adolescents with the highest need for behavior change. On the other hand, many adolescents are engaged in a combination of health-risk and health-promoting behaviors [56], and our findings suggest that adolescents who experience being both confirmed and encouraged by the intervention may benefit from the multiple-behavior approach. To further examine participants' engagement with LIFE4YOUth, future research should investigate the quality of adolescents' engagement with intervention content in more depth, based on self-authored data as well as user-experience data collected from a larger sample of participants. Future research could explore how engagement levels relate to the effectiveness of multiple behavior change interventions in adolescents. It would also be valuable to understand the characteristics of adolescents who are not drawn to digital behavior change support and what types of health-promoting support, if any, they might prefer.

Strengths and Limitations

This study used a novel design to gain a more nuanced understanding of the conditions and mechanisms that influence and explain high school students' engagement with a stand-alone mHealth multiple behavior change intervention. The use of a mixed methods design with QCA for integration increased rigor and provided insights beyond those obtained from either quantitative or qualitative data alone, which is considered a strength. There are, however, several important limitations to consider. First, the use of self-reported data on health behaviors increases the risk of recall bias. Importance, confidence, and know-how were additionally measured using a general scale, which limited the ability to capture variability in these constructs across different behaviors [57]. Second, the convenience sampling [58] of the focus groups limited diversity in terms of, for example, socioeconomic status and educational background. The time since focus group participants accessed the intervention also varied, which may have resulted in more speculative statements rather than those based on participants' experiences. Furthermore, the choice of conditions for the QCA was limited to the data collected in the LIFE4YOUth trial. Several perspectives highlighted by focus group participants, including stress and LIFE4YOUth experiences, were therefore not included in the QCA. Nevertheless, this study was conducted iteratively, with methodological decisions based on the data rather than the reverse.

XSL•FO
RenderX

Finally, all data were collected within the context of a randomized controlled trial, which does not replicate naturalistic circumstances. Biases may have been introduced by the research setting due to research participation effects [59], and it is possible that high school students would have engaged with LIFE4YOUth differently if they had not been part of a research study. We believe, however, that the transparent reporting allows readers to assess whether the findings are valid and generalizable to other contexts.

Conclusions

Several conditions were identified that may influence and explain high school students' engagement with the multiple behavior change mHealth intervention (LIFE4YOUth). Low-engaged high school students were more likely to be described as unmotivated high-needers, motivated low-needers, or dissatisfied needers. Emotions and reactions evoked by the intervention, as well as perceptions of whether the intervention aligned with high school students' support needs, were suggested to explain engagement. In addition, the timing of SMS text messaging prompts was also considered an important condition influencing all levels of engagement. The findings from this study indicate that the multiple-behavior approach was a decisive factor for students with multiple risk behaviors not to remain engaged with LIFE4YOUth, whereas it may have been a better fit for students who felt both confirmed and encouraged by this approach. Future studies can build on these findings to investigate how tailoring the number of behaviors addressed by mHealth tools targeting adolescents may optimize engagement and behavioral outcomes.

Acknowledgments

We would like to express our gratitude to the high school students who participated in this study as well as staff at the schools for their assistance. We are also very grateful to all participants and experts at the International QCA Development Workshop 2023 in Antwerpen, who markedly contributed through insightful discussions and suggestions that improved the study.

The authors declare the use of generative AI in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to generative artificial intelligence (GAI) tools under full human supervision:

1. The GAI tool used was Microsoft Copilot version number 2.20260101.44.0.

2. Responsibility for the final manuscript lies entirely with the authors.

3. GAI tools are not listed as authors and do not bear responsibility for the final outcomes.

4. Declaration submitted by the first author.

Funding

This research was funded by the Swedish Research Council for Health, Working Life and Welfare 2018-01410; ML. The funder had no involvement in the study design, data collection, analysis, interpretation, or the writing of the manuscript.

Data Availability

The data that support the findings of this study are available from the first author, but restrictions apply to the availability of these data to protect study participant privacy. However, pseudo-anonymized data are available upon reasonable request.

Authors' Contributions

AS, PH, KT, ML, MB, and UM performed conceptualization. AS performed data curation. AS and UM performed investigation. AS, PH, MB, and UM conducted formal analysis. AS, PH, KT, ML, MB, and UM contributed to interpretation of data. AS wrote the original draft. AS, PH, KT, ML, MB, and UM performed review and editing. All authors approved the final version of the manuscript.

Conflicts of Interest

MB owns a private company (Alexit AB) that maintains and distributes evidence-based lifestyle interventions to be used by the public and in health care settings. Alexit AB played no role in developing the intervention, study design, data analysis, data interpretation, or writing of this report.

Multimedia Appendix 1
Complementary analyses of engagement patterns, including the validity of the identified clusters, the influence of the COVID-19 pandemic on engagement, and the robustness of the results.
[DOCX File , 130 KB - mhealth_v14i1e88054_app1.docx]

References

XSL•FO
RenderX

1. Gore FM, Bloem PJN, Patton GC, Ferguson J, Joseph V, Coffey C, et al. Global burden of disease in young people aged 10-24 years: a systematic analysis. Lancet 2011 Jun 18;377(9783):2093-2102. [doi: 10.1016/S0140-6736(11)60512-6] [Medline: 21652063]
2. Patton GC, Sawyer SM, Santelli JS, Ross DA, Afifi R, Allen NB, et al. Our future: a Lancet commission on adolescent health and wellbeing. Lancet 2016;387(10036):2423-2478 [FREE Full text] [doi: 10.1016/S0140-6736(16)00579-1] [Medline: 27174304]
3. Akseer N, Mehta S, Wigle J, Chera R, Brickman ZJ, Al-Gashm S, et al. Non-communicable diseases among adolescents: current status, determinants, interventions and policies. BMC Public Health 2020;20(1):1908 [FREE Full text] [doi: 10.1186/s12889-020-09988-5] [Medline: 33317507]
4. Rodríguez-González P, Iglesias D, Fernandez-Rio J, Gao Z. Effectiveness of interventions using apps to improve physical activity, sedentary behavior and diet: an umbrella review. Complement Ther Clin Pract 2023;50:101711. [doi: 10.1016/j.ctcp.2022.101711] [Medline: 36509032]
5. Staiger PK, O'Donnell R, Liknaitzky P, Bush R, Milward J. Mobile apps to reduce tobacco, alcohol, and illicit drug use: systematic review of the first decade. J Med Internet Res 2020;22(11):e17156 [FREE Full text] [doi: 10.2196/17156] [Medline: 33231555]
6. Talens C, da Quinta N, Adebayo FA, Erkkola M, Heikkilä M, Bargiel-Matusiewicz K, et al. Mobile- and web-based interventions for promoting healthy diets, preventing obesity, and improving health behaviors in children and adolescents: systematic review of randomized controlled trials. J Med Internet Res 2025;27:e60602 [FREE Full text] [doi: 10.2196/60602] [Medline: 40392587]
7. Odendaal WA, Anstey Watkins J, Leon N, Goudge J, Griffiths F, Tomlinson M, et al. Health workers' perceptions and experiences of using mHealth technologies to deliver primary healthcare services: a qualitative evidence synthesis. Cochrane Database Syst Rev 2020;3(3):CD011942 [FREE Full text] [doi: 10.1002/14651858.CD011942.pub2] [Medline: 32216074]
8. Ferretti A, Hubbs S, Vayena E. Global youth perspectives on digital health promotion: a scoping review. BMC Digit Health 2023;1(1). [doi: 10.1186/s44247-023-00025-0]
9. Peuters C, DeSmet A, Maenhout L, Cardon G, Debeer D, Crombez G. Adolescent engagement with a multicomponent mhealth tool: identifying usage patterns, determinants, and health behavior change in an intervention trial. JMIR Mhealth Uhealth 2025;13:e59041 [FREE Full text] [doi: 10.2196/59041] [Medline: 40825232]
10. Eaton C, Vallejo N, McDonald X, Wu J, Rodríguez R, Muthusamy N, et al. User engagement with mhealth interventions to promote treatment adherence and self-management in people with chronic health conditions: systematic review. J Med Internet Res 2024;26:e50508 [FREE Full text] [doi: 10.2196/50508] [Medline: 39316431]
11. O'Brien HL, Roll I, Kampen A, Davoudi N. Rethinking (Dis)engagement in human-computer interaction. Comput Hum Behav 2022;128:107109. [doi: 10.1016/j.chb.2021.107109]
12. Nahum-Shani I, Yoon C. Towards the science of engagement with digital interventions. Curr Dir Psychol Sci 2024;33(4):239-246. [doi: 10.1177/09637214241254328] [Medline: 39552747]
13. Whitehead L, Robinson S, Arabiat D, Jenkins M, Morelius E. The report of access and engagement with digital health interventions among children and young people: systematic review. JMIR Pediatr Parent 2024 Jan 17;7:e44199. [doi: 10.2196/44199] [Medline: 38231560]
14. Psihogios AM, King-Dowling S, O'Hagan B, Darabos K, Maurer L, Young J, et al. Contextual predictors of engagement in a tailored mHealth intervention for adolescent and young adult cancer survivors. Ann Behav Med 2021;55(12):1220-1230. [doi: 10.1093/abm/kaab008] [Medline: 33674863]
15. Li Y, Guo Y, Hong YA, Zeng Y, Monroe-Wise A, Zeng C, et al. Dose-response effects of patient engagement on health outcomes in an mHealth intervention: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2022;10(1):e25586. [doi: 10.2196/25586] [Medline: 34982724]
16. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
17. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, et al. Factors Influencing Adherence to mHealth Apps for Prevention or Management of Noncommunicable Diseases: Systematic Review. J Med Internet Res 2022 May 25;24(5):e35371 [FREE Full text] [doi: 10.2196/35371] [Medline: 35612886]
18. Duan Y, Shang B, Liang W, Du G, Yang M, Rhodes RE. Effects of eHealth-based multiple health behavior change interventions on physical activity, healthy diet, and weight in people with noncommunicable diseases: systematic review and meta-analysis. J Med Internet Res 2021;23(2):e23786 [FREE Full text] [doi: 10.2196/23786] [Medline: 33616534]
19. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
20. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]

XSL•FO
RenderX

21. Schwarz A, Winkens LHH, de Vet E, Ossendrijver D, Bouwsema K, Simons M. Design features associated with engagement in mobile health physical activity interventions among youth: systematic review of qualitative and quantitative studies. JMIR Mhealth Uhealth 2023;11:e40898 [FREE Full text] [doi: 10.2196/40898] [Medline: 36877551]
22. Ghosh P, Proffitt R, Bosworth KT, Koopman RJ, Flowers L, Wilson G, et al. mHealth app features that facilitate adolescent use for lifestyle management, and are endorsed by caregivers and health care providers. Mhealth 2024;10:21. [doi: 10.21037/mhealth-24-3] [Medline: 39114461]
23. Thornton L, Corliss C, Deen H, Teesson M, Champion KE, Partridge SR, et al. The triple E project: a factorial randomised controlled trial to enhance engagement with eHealth approaches to improve health risk behaviours among adolescents. BMC Public Health 2024;24(1):2697 [FREE Full text] [doi: 10.1186/s12889-024-20124-5] [Medline: 39363157]
24. Bendtsen M, Seiterö A, Bendtsen P, Henriksson H, Henriksson P, Thomas K, et al. mHealth intervention for multiple lifestyle behaviour change among high school students in Sweden (LIFE4YOUth): protocol for a randomised controlled trial. BMC Public Health 2021;21(1):1406 [FREE Full text] [doi: 10.1186/s12889-021-11446-9] [Medline: 34271882]
25. Creswell JW, Plano Clark VL. Designing and Conducting Mixed Methods Research. Thousand Oaks, CA: SAGE Publications; 2017.
26. Kahwati L, Kane H. Qualitative Comparative Analysis in Mixed Methods Research and Evaluation. Thousand Oaks, CA: SAGE Publications; 2020.
27. Patton MQ. Qualitative Research and Evaluation Methods - Integrating Theory and Practice. Thousand Oaks, California: SAGE Publications; 2015.
28. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
29. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy 2008;13(2):92-98. [doi: 10.1258/jhsrp.2007.007074] [Medline: 18416914]
30. Bendtsen M, Bendtsen P, Henriksson H, Henriksson P, Müssener U, Thomas K, et al. The mobile health multiple lifestyle behavior interventions across the lifespan (MoBILE) research program: protocol for development, evaluation, and implementation. JMIR Res Protoc 2020;9(4):e14894 [FREE Full text] [doi: 10.2196/14894] [Medline: 32310147]
31. Seiterö A, Henriksson P, Thomas K, Henriksson H, Löf M, Bendtsen M, et al. Effectiveness of a mobile phone-delivered multiple health behavior change intervention (LIFE4YOUth) in adolescents: randomized controlled trial. J Med Internet Res 2025;27:e69425 [FREE Full text] [doi: 10.2196/69425] [Medline: 40262133]
32. Müssener U, Thomas K, Linderoth C, Löf M, Åsberg K, Henriksson P, et al. Development of an intervention targeting multiple health behaviors among high school students: participatory design study using heuristic evaluation and usability testing. JMIR Mhealth Uhealth 2020;8(10):e17999 [FREE Full text] [doi: 10.2196/17999] [Medline: 33118942]
33. National guidelines: unhealthy lifestyle habits. Socialstyrelsen (the National Board of Health and Welfare in Sweden). URL: https://www.socialstyrelsen.se/kunskapsstod-och-regler/regler-och-riktlinjer/nationella-riktlinjer/ [accessed 2021-07-12]
34. Bendtsen M, Bendtsen P, Müssener U. Six-month outcomes from the nexit junior trial of a text messaging smoking cessation intervention for high school students: randomized controlled trial with Bayesian analysis. JMIR Mhealth Uhealth 2021;9(10):e29913 [FREE Full text] [doi: 10.2196/29913] [Medline: 34673532]
35. Thomas K, Müssener U, Linderoth C, Karlsson N, Bendtsen P, Bendtsen M. Effectiveness of a text messaging-based intervention targeting alcohol consumption among university students: randomized controlled trial. JMIR Mhealth Uhealth 2018;6(6):e146 [FREE Full text] [doi: 10.2196/mhealth.9642] [Medline: 29941417]
36. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
37. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
38. How a web session is defined in universal analytics. Analytics Help. 2017. URL: https://support.google.com/analytics/answer/2731565#zippy=%2Cin-this-article [accessed 2023-11-19]
39. Hedin L, Seiterö A, Crawford J, Bendtsen M, Löf M. Mediated effects of LIFE4YOUth-a mobile health intervention for multiple lifestyle behavior change among high school students in Sweden: findings from a randomized controlled trial. BMC Public Health 2025;25(1):922 [FREE Full text] [doi: 10.1186/s12889-025-22097-5] [Medline: 40055714]
40. Mahon C, Howard E, O'Reilly A, Dooley B, Fitzgerald A. A cluster analysis of health behaviours and their relationship to mental health difficulties, life satisfaction and functioning in adolescents. Prev Med 2022;164:107332. [doi: 10.1016/j.ypmed.2022.107332] [Medline: 36336163]
41. McGovern CM, Militello LK, Arcoleo KJ, Melnyk BM. Factors associated with healthy lifestyle behaviors among adolescents. J Pediatr Health Care 2018;32(5):473-480 [FREE Full text] [doi: 10.1016/j.pedhc.2018.04.002] [Medline: 29941235]
42. Mazur J, Szkultecka-Dębek M, Dzielska A, Drozd M, Małkowska-Szkutnik A. What does the cantril ladder measure in adolescence? Arch Med Sci 2018;14(1):182-189 [FREE Full text] [doi: 10.5114/aoms.2016.60718] [Medline: 29379549]
43. Rohlfing I. The choice between crisp and fuzzy sets in qualitative comparative analysis and the ambiguous consequences for finding consistent set relations. Field Methods 2019;32(1):75-88. [doi: 10.1177/1525822x19896258]
44. Oana IE, Schneider CQ, Thomann E. Qualitative Comparative Analysis with R: A Beginner's Guide. Cambridge: Cambridge University Press; 2021.

XSL•FO
RenderX

45. Thomann E, Ege J, Paustyan E. Approaches to qualitative comparative analysis and good practices: a systematic review. Swiss Polit Sci Rev 2022;28(3):557-580. [doi: 10.1111/spsr.12503]
46. Elgin DJ, Erickson E, Crews M, Kahwati LC, Kane HL. Applying qualitative comparative analysis in large-N studies: a scoping review of good practices before, during, and after the analytic moment. Qual Quant 2024;58(5):4241-4256. [doi: 10.1007/s11135-024-01849-2]
47. Schneider CQ, Wagemann C. Doing justice to logical remainders in QCA: moving beyond the standard analysis. Disabil Rehabil 2012;66(1):211-220. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]
48. Oana IE, Schneider CQ. A robustness test protocol for applied QCA: theory and R software application. Sociol Methods Res 2021;53(1):57-88. [doi: 10.1177/00491241211036158]
49. Saarijärvi M, Wallin L, Moons P, Gyllensten H, Bratt EL. Factors affecting adolescents' participation in randomized controlled trials evaluating the effectiveness of healthcare interventions: the case of the STEPSTONES project. BMC Med Res Methodol 2020;20(1):205 [FREE Full text] [doi: 10.1186/s12874-020-01088-7] [Medline: 32746862]
50. Crutzen R, Ruiter R. Interest in behaviour change interventions: a conceptual model. Eur Heal Psychol 2015;17(1):6-11 [FREE Full text]
51. Thornton L, Brown HM, Osman B, Stewart C, Whife J, Champion KE, et al. Factors associated with adolescents' engagement with a healthy lifestyles app. Procedia Comput Sci 2022;206:56-67. [doi: 10.1016/j.procs.2022.09.085]
52. Peuters C, Maenhout L, Cardon G, De Paepe A, DeSmet A, Lauwerier E, et al. A mobile healthy lifestyle intervention to promote mental health in adolescence: a mixed-methods evaluation. BMC Public Health 2024;24(1):44 [FREE Full text] [doi: 10.1186/s12889-023-17260-9] [Medline: 38166797]
53. Paz Castro R, Haug S, Filler A, Kowatsch T, Schaub MP. Engagement within a mobile phone-based smoking cessation intervention for adolescents and its association with participant characteristics and outcomes. J Med Internet Res 2017;19(11):e356 [FREE Full text] [doi: 10.2196/jmir.7928] [Medline: 29092811]
54. Maenhout L, Peuters C, Cardon G, Crombez G, DeSmet A, Compernolle S. Nonusage attrition of adolescents in an mHealth promotion intervention and the role of socioeconomic status: secondary analysis of a 2-arm cluster-controlled trial. JMIR Mhealth Uhealth 2022;10(5):e36404 [FREE Full text] [doi: 10.2196/36404] [Medline: 35536640]
55. Melo GLR, Santo RE, Mas Clavel E, Bosque Prous M, Koehler K, Vidal-Alaball J, et al. Digital dietary interventions for healthy adolescents: a systematic review of behavior change techniques, engagement strategies, and adherence. Clin Nutr 2025;45:176-192. [doi: 10.1016/j.clnu.2025.01.012] [Medline: 39827503]
56. Whitaker V, Oldham M, Boyd J, Fairbrother H, Curtis P, Meier P, et al. Clustering of health-related behaviours within children aged 11-16: a systematic review. BMC Public Health 2021;21(1):137 [FREE Full text] [doi: 10.1186/s12889-020-10140-6] [Medline: 33446174]
57. Noar SM, Chabot M, Zimmerman RS. Applying health behavior theory to multiple behavior change: considerations and approaches. Prev Med 2008;46(3):275-280. [doi: 10.1016/j.ypmed.2007.08.001] [Medline: 17825898]
58. Moser A, Korstjens I. Series: practical guidance to qualitative research. Part 3: sampling, data collection and analysis. Eur J Gen Pract 2018;24(1):9-18 [FREE Full text] [doi: 10.1080/13814788.2017.1375091] [Medline: 29199486]
59. Barned C, Dobson J, Stintzi A, Mack D, O'Doherty KC. Children's perspectives on the benefits and burdens of research participation. AJOB Empir Bioeth 2018;9(1):19-28. [doi: 10.1080/23294515.2018.1430709] [Medline: 29369013]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
GRAMMS: Good Reporting of a Mixed Method Study
mHealth: mobile health
MoBILE: Mobile health Multiple lifestyle Behaviors Interventions across the Lifespan
MVPA: moderate to vigorous physical activity
QCA: qualitative comparative analysis
RQ: research question

Edited by A Schwartz, A Stone; submitted 20.Nov.2025; peer-reviewed by T Baranowski; comments to author 02.Jan.2026; accepted 30.Jan.2026; published 12.Mar.2026.

Please cite as:
Seiterö A, Henriksson P, Thomas K, Löf M, Bendtsen M, Müssener U
Adolescents' Engagement With an mHealth Multiple Health Behavior Change Intervention (LIFE4YOUth): Mixed Methods and Qualitative Comparative Analysis
JMIR Mhealth Uhealth 2026;14:e88054
URL: https://mhealth.jmir.org/2026/1/e88054
doi:10.2196/88054
PMID:

XSL•FO
RenderX

©Anna Seiterö, Pontus Henriksson, Kristin Thomas, Marie Löf, Marcus Bendtsen, Ulrika Müssener. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 12.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Assessing Mental Health and Emotional States by Using Smartphone Photoplethysmography–Based Digital Pulse Waveform Analysis: Cross-Sectional Observational Study

Ivan Liu[1,2,3], DPhil; Luming Hu[1,2,3], DPhil; Jing Luo[3], MSc; Chang Liu[2], BS; Qi Zhong[3], MSc; Shiguang Ni[4], DPhil

[1]Bay Area School of Applied Psychological Sciences, Beijing Normal University, Zhuhai, Guangdong, China

[2]Department of Psychology, Faculty of Arts and Sciences, Beijing Normal University, Zhuhai, Guangdong, China

[3]Faculty of Psychology, Beijing Normal University, Beijing, China

[4]Shenzhen International Graduate School, Tsinghua University, Shenzhen, Guangdong, China

Corresponding Author:
Shiguang Ni, DPhil
Shenzhen International Graduate School
Tsinghua University
University Town of Shenzhen, Nanshan District
Shenzhen, Guangdong, 518055
China
Phone: 86 755 2603 6097
Email: ni.shiguang@sz.tsinghua.edu.cn

Abstract

Background: Pulse characteristics are well-established biomarkers of physical health; however, their relevance to psychological well-being remains insufficiently explored. A key barrier is the difficulty of acquiring pulse recordings and blood pressure measurements of adequate quality outside clinical or laboratory settings by using accessible measurement approaches.

Objective: This study aimed to examine the feasibility of using smartphone photoplethysmography to extract fingertip pulse-waveform features and to evaluate their associations with psychological measures. It further aimed to systematically compare time-, curvature-, and frequency-domain pulse-waveform features in relation to psychological variables.

Methods: A total of 127 students and university employees in Shenzhen, China, were recruited. Participants recorded repeated 4-minute fingertip videos by using a custom smartphone app while a fingertip oximeter simultaneously acquired reference pulse signals. Smartphone videos were converted into photoplethysmography signals, segmented into beat-to-beat intervals, and summarized into time-, curvature-, and frequency-domain features, with normalization for heart rate and stature. Psychological well-being and mental health were assessed using the Satisfaction With Life Scale, Subjective Vitality Scale, Positive and Negative Affect Schedule, Patient Health Questionnaire-9, Generalized Anxiety Disorder-7, and the Self-Assessment Manikin. Associations between pulse-waveform features and psychological measures were examined using univariate regression with participant-level aggregation and cluster-robust standard errors. Random forest models evaluated multivariate predictive performance by using participant-level cross-validation. Agreement between smartphone-derived and oximeter-derived waveform features was assessed using Bland–Altman analysis.

Results: Correlation analyses revealed strong within-domain associations among time-, curvature-, and frequency-domain pulse-waveform features, with comparatively weaker cross-domain correlations. A correlation-based feature-selection procedure reduced multicollinearity and yielded a final set of 7 features: estimated reflection index, crest time (CT), the third curvature minimum (F/A), the fourth curvature minimum (H/A), the first power spectrum density component, the baseline of Fourier decomposition (V0), and systolic blood pressure. Univariate regression analyses indicated that negative psychological states were primarily associated with time- and curvature-domain features. Depressive symptoms were significantly related to F/A, V0, and the estimated reflection index. Anxiety showed an association with F/A, and negative affect was associated with CT and F/A. In contrast, positive affect measures showed fewer and weaker associations. Valence was related to F/A and H/A, whereas arousal was associated with CT and H/A. Random forest models demonstrated statistically significant but modest predictive performance for negative mental health outcomes, with weaker performance for positive affect. Bland–Altman analyses indicated minimal systematic bias for outcomes with significant predictive correlations. Comparisons with an oximeter showed significant correlations and acceptable agreement, with time-domain features demonstrating greater robustness than reflection-based metrics.

XSL•FO
RenderX

Conclusions: Smartphone-based photoplethysmography can capture pulse-waveform features associated with psychological measures, particularly negative psychological states. However, predictive performance remains limited, and variability in signal quality from user-operated recordings poses a practical challenge.

(JMIR Mhealth Uhealth 2026;14:e81301) doi:10.2196/81301

KEYWORDS

mental health; mental well-being; happiness; smartphone photoplethysmography; digital pulse waveform; multimodal emotion recognition; affective computing; atherosclerosis; blood pressure

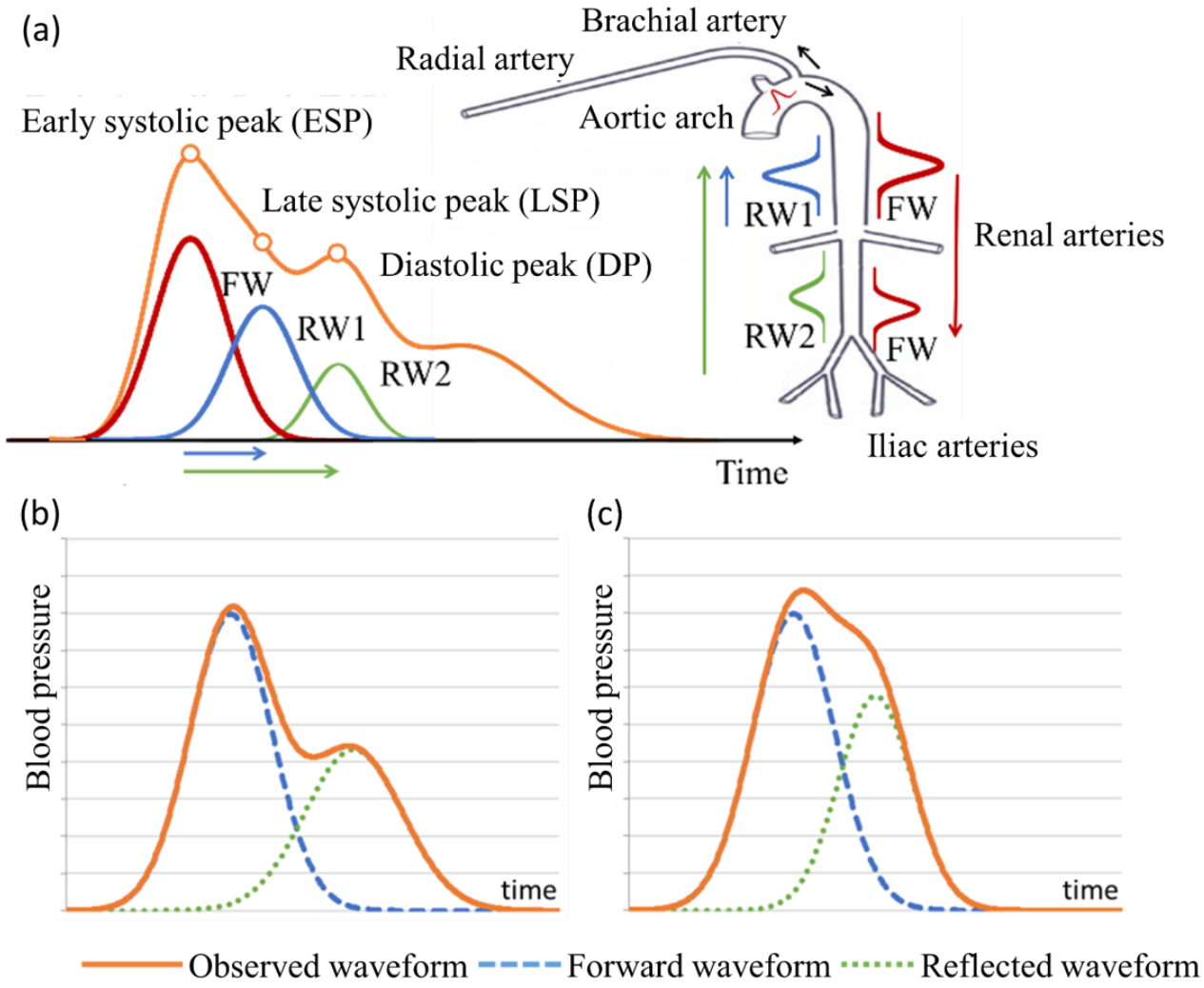
Introduction

Assessing physiological health by monitoring pulse rhythm and waveforms has been practiced for millennia [1,2]. Through the analysis of pulse-waveform contours, clinicians gain insight into circulatory status and related physiological influences [3]. Numerous studies have associated waveform characteristics with factors including height, ethnicity [4], sex [5], respiration [6], and autonomic nervous system activity [7], and have used these features to predict medical conditions such as hypertension [8], heart failure [9], lipid metabolism disorders [3], renal disease [10], and diabetes mellitus [11].

The relationship between pulse waveforms and various health conditions can be explained by arterial stiffening. During ventricular systole, the left ventricle ejects oxygenated blood into the aorta, generating a forward pressure wave that travels through the arterial tree. The recorded pulse waveform reflects the superposition of this forward wave and the wave reflected from peripheral arterial sites [12] (Figure 1A) [8,13]. In healthy arteries, the compliant arterial walls transmit pulse energy more slowly, causing the reflected waves to arrive later, producing a waveform characterized by a pronounced systolic peak and a well-defined dicrotic notch (Figure 1B). With aging or pathological stiffening, often initiated by endothelial injury and atherosclerotic remodeling [14], arterial compliance decreases, causing the reflected wave to return earlier, which augments the forward wave, raises systolic pressure, and blunts the dicrotic notch [15] (Figure 1C). Thus, specific modifications in waveform morphology serve as indicators of cardiovascular aging and related pathologies [16].

XSL•FO
RenderX

Figure 1. (A) The reflected waves and the shape of the radial arterial pulse waveform (reproduced from Liu et al [13], which is published under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License [17]). (B) An illustration of the radial arterial pulse waveform of healthy young adults (adapted from Baruch et al [8], which is published under Creative Commons Attribution 2.0 Generic License [18]). (C) The pulse waveform of people with signs of arterial aging. The reflected wave arrives earlier, boosts the systolic blood pressure, and changes the pulse contour [12,19]. FW: forward wave; RW: reflected wave.



Because enduring psychological states and personality traits modulate physiological regulatory systems, particularly inflammatory pathways [20] and sympathetic nervous system activation [21], it is reasonable to assume that psychological factors also influence vascular waveforms. Empirical evidence has supported the assumption that individuals with psychological conditions, such as chronic stress [22], depression [23], bipolar disorder [24], anxiety [25], and panic disorder [26], frequently exhibit arterial stiffening. Consequently, maladaptive psychological states may be detectable via pulse waveform analysis. In support of this premise, early investigations have reported significant associations between pulse waveform characteristics and mental health issues, including depression [27], anxiety [18,28], affective states [29], chronic fatigue [30], and cognitive stress [31,32].

Although early studies have linked pulse waveform morphology to mental health, real-world applications remain limited. A key challenge lies in the cost and complexity of acquiring waveform features outside clinical environments. Most prior research has focused on aortic waveforms, obtained using invasive equipment and specialized personnel in hospital settings [33], thereby limiting scalability. More recently, a small number of pioneering studies have used photoplethysmography to capture peripheral arterial signals (eg, from radial or digital arteries) [34,35]. However, despite its lower technical requirements, photoplethysmography-based pulse analysis still typically relies on dedicated medical devices [36], which constrains its accessibility and widespread deployment.

To bridge the gap in the literature, this study has two primary objectives:

First, it proposes and validates an innovative technique for extracting fingertip (digital artery) pulse-waveform features pertinent to psychological evaluation using only a standard smartphone. Over the past decade, smartphone camera–based acquisition of cardiac signals has garnered substantial attention, and numerous investigations have demonstrated that heart rate (HR) and heart rate variability (HRV) derived from smartphone cameras can reflect alterations in psychological states [37]. This approach uses the phone's built-in camera to record subtle, cardiac-induced skin color fluctuations at the fingertip and

XSL•FO
RenderX

converts them into radial artery waveforms via photoplethysmography. However, the capacity of this technology to capture detailed pulse waveform morphology and its application to mental health assessment remains underexplored.

Second, existing studies have typically examined a limited number of pulse-waveform features in isolation, and comprehensive comparisons, particularly for digital pulse-waveform features, are largely lacking. Although prior research has explored associations between pulse-waveform characteristics and psychological states, it has focused almost exclusively on aortic indices, such as the augmentation index [38]. Therefore, evidence linking peripheral waveforms (eg, radial or digital artery signals) to mental health outcomes remains sparse. Given that peripheral waveform morphology differs substantially from aortic contours due to local anatomical and physiological factors [16], these signals warrant independent and systematic investigation. On the other hand, studies of radial or digital artery morphology have generally considered only a narrow subset of available features. To address these gaps, this study conducts an extensive examination of time-domain, frequency-domain, and curvature-domain pulse-waveform features, evaluates their interrelationships, and investigates their associations with psychological health.

Methods

Participant Recruitment

This study recruited a total of 127 participants, comprising university students and staff members, from the last author's affiliated institution in Shenzhen, China (72/127, 56.69% male; mean age 22.78, SD 1.97 years). Participants were recruited through on-campus posters and electronic posters distributed via student social media platforms. Eligibility was restricted to university students and staff, ensuring that all participants were at least 18 years of age and had access to university facilities. Participants were required to have no clinically diagnosed mental disorders or cardiovascular conditions. No additional exclusion criteria were applied.

Study Procedures and Data Collection

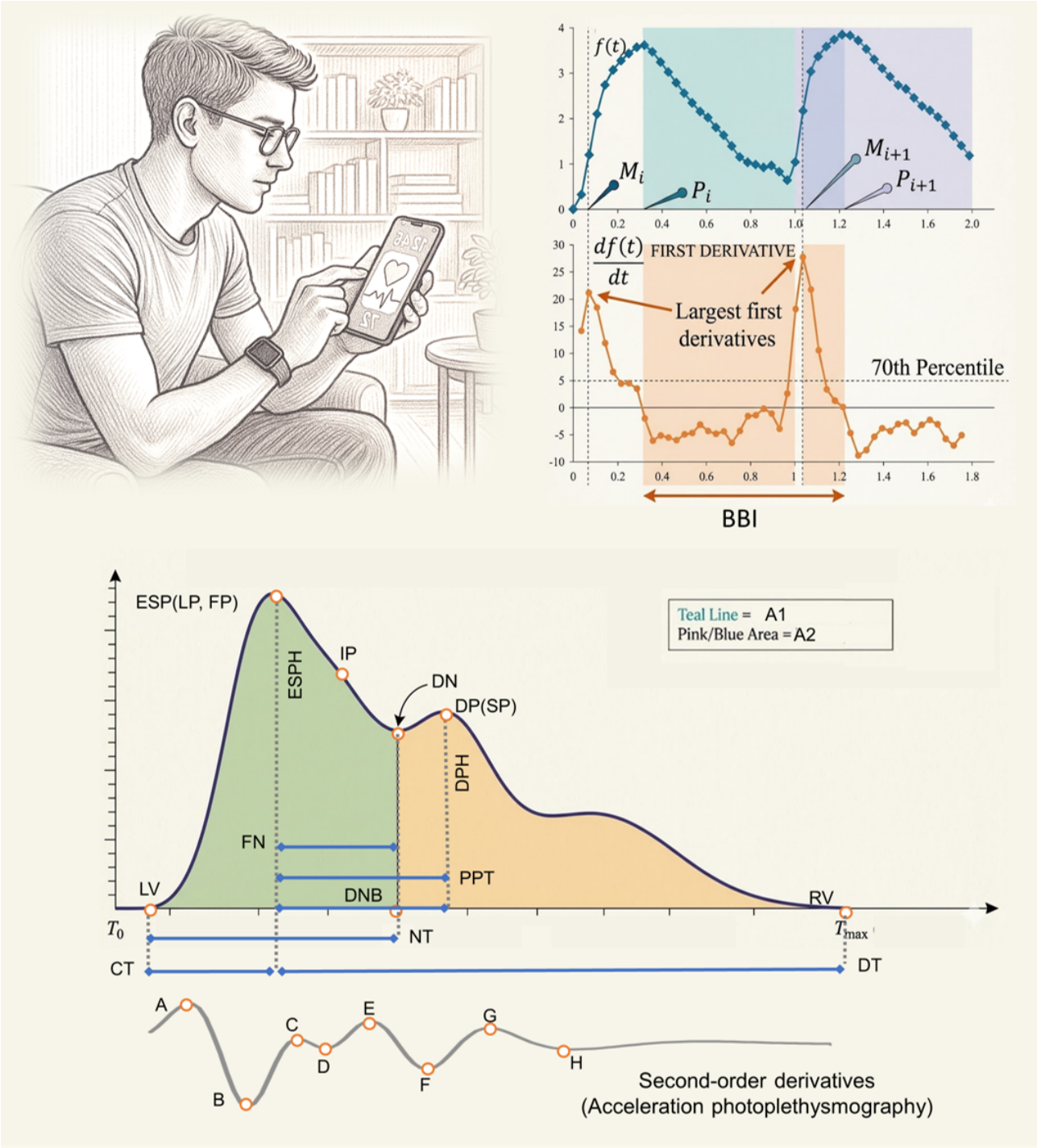
During data collection, participants were instructed to rest their arm on a table, hold a Mi 8 SE smartphone (Xiaomi) in their left hand, and place their index finger over the rear camera. Using the Heartily Happy app, developed by our team and publicly available on Google Play at the time of the study, participants recorded 7 fingertip videos, each lasting 4 minutes, at a resolution of 120×160 pixels. To further evaluate the accuracy and practical applicability of the proposed approach, participants were additionally instructed to place the index finger of their right hand on a US Food and Drug Administration–cleared fingertip oximeter (FDA 510(k): K082641; CMS50D, Contec Medical Systems Co Ltd) to simultaneously record pulse waveforms. This study is part of a broader research program examining the relationship between pulse waveforms and mental health. The same dataset has also been used in studies investigating blood pressure [13] and HRV [37].

Signal Processing

The Heartily Happy app used the smartphone's built-in flashlight to illuminate the fingertip and captured video at approximately 30 frames per second (Figure 2) [13,39]. Raw video frames, initially recorded in YUV format, were converted to red green blue color model for subsequent analysis. Waveform signals were extracted by computing a SD–weighted average across the 3 color channels.

XSL•FO
RenderX

Figure 2. Smartphone photoplethysmography data collection and feature extraction. BBI: beat-to-beat interval; CT: crest time; DN: dicrotic notch; DP: diastolic peak; DPH: diastolic peak height; DT: diastolic time; ESP: the early systolic left peak; ESPH: early systolic peak height; FN: the interval between first peak to notch; FP: first peak; IP: inflection point; LP: left peak; LV: left valley; NT: notch time; PPG: photoplethysmography; PPT: peak-to-peak time; RV: reflected wave; SP: second peak.



Following peak detection, both photoplethysmography and oximetry signals were segmented into beat-to-beat intervals (BBIs). Waveform features were calculated for each BBI, and the median value of each feature across all BBIs was used to represent the sample. Detailed signal-processing procedures are described in [13,39].

Because pulse waveform length is inversely related to HR, following a conventional, although controversial, practice in aortic blood pressure research, waveform features were normalized to a reference HR of 75 beats per minute to control for HR-related effects [40-42]. In addition, since the timing of reflected waves depends on vascular path length, which is associated with body height, waveform features are often further normalized by stature; for example, the stiffness index is defined as body height divided by peak-to-peak time (PPT) [43]. Therefore, in this study, waveform features were normalized to a reference height of 170 cm to reduce stature-related interindividual variability. Sex was not included as a separate normalization factor, as its effects on waveform morphology are largely mediated by differences in body height.

XSL•FO
RenderX

Waveform Metrics

This study considers 3 categories of waveform features commonly cited in pulse waveform analysis but underexplored in psychological research (Table 1 [13]). Time-domain indicators encompass measures of intervals between characteristic waveform points, amplitude differences among these points, and the total area under the waveform curve. Curvature-domain features are derived from the second derivative of the photoplethysmography waveform, often termed acceleration photoplethysmography, and quantify waveform contour and curvature [44]. Frequency-domain features are obtained via Fourier transform decomposition, with power spectral density functions characterizing the strength of each frequency component (Figure 3) [13].

XSL•FO
RenderX

Table 1. Waveform features (adapted from Liu et al [13]).

Name	Definition and calculation details
Time-domain features	
SI[a]	The ratio of body height to PPT[b] [45,46].
PPT	The interval between the ESP[c], (also known as the LP[d] or the FP[e]) and the DP[f]. The RV[g] is the lowest point in the BBI[h], and the preceding BBI's RV is termed the LV[i]. The DP or SP is the first point between ESP and RV with a zero first derivative and a negative second derivative. If no second peak exists between ESP and RV, indicating a consistently negative slope from LP to RV, the DP is identified as the point with the minimum second derivative between ESP and RV.
EPPT	Estimated PPT
RI[j]	The ratio of the DPH[k] to the estimated ESPH[l] is a critical measure in this analysis.
ERI[m]	Estimated RI. Since the initial findings indicated that the auto-exposure adjustment feature of smartphone cameras could result in waveforms that appear smoother than those predicted by theoretical models, this study incorporates the ERI, as proposed by [13]. The ERI, defined as the ratio between the second peak and the estimated early systolic peak, serves as a substitute for the widely used Reflection Index to account for these discrepancies.
ESI[n]	Estimated SI
CT[o]	The interval between the LV and ESP [47].
NT[p]	The IP[q] is defined as the last point before the second peak where the first derivative shifts from positive to negative. The lowest point between the end-systolic point (ESP) and the dicrotic peak DP[r] is called the DN[s]. NT is the interval from the LV to the DN. In cases where there is no obvious second peak, the IP is designated as the DN.
IPP[t]	The percentage of BBIs in a sample where there is no obvious second peak and the IP is designated as the DN.
DT[u]	The interval between ESP and the RV [48]
IS[v]	The slope of the curve at the inflection point.
RCA	The ratio of CT to NT.
RDA	The ratio of NT to the combined duration of CT and DT.
FN	The interval between the first peak to notch.
SR	The interval between the second peak to RV
NV	The interval between the notch to RV
NH	Notch height.
SPH	Second peak height.
A1	The area under the curve from LV to the DN [32,49,50]. When there is no obvious second peak, IP is used as a surrogate for the DN. Since calculating the actual area involves more complex methods, this study approximates A1 as the polygon formed by LV, ESP, DN, and the DNB.
A2	The area under the curve from DN to RV [49]. This study approximates A2 as the triangle comprising IP, DN, and RV.
Inflection point area (IPA)	A2/A1 [49].
Curvature-domain features (acceleration photoplethysmography)	
A	The first local maximum of the second derivative is determined by the point with the largest second derivative before ESP.
B	The first local minimum of the second derivative, determined by the point with the smallest second derivative before ESP.
C	The second local maximum of the second derivative is determined by the point with the largest second derivative, point B, and point E.
D	The second local minimum of the second derivative is determined by the point with the smallest second derivative between points B and point E.
E	The third local maximum of the second derivative is determined by the point with the largest second derivative between ESP and point F.
F	The third local minimum of the second derivative is determined by the point with the smallest second derivative between the ESP and RV.
G	The fourth local maximum of the second derivative is determined by the point with the largest second derivative between point F and point H.

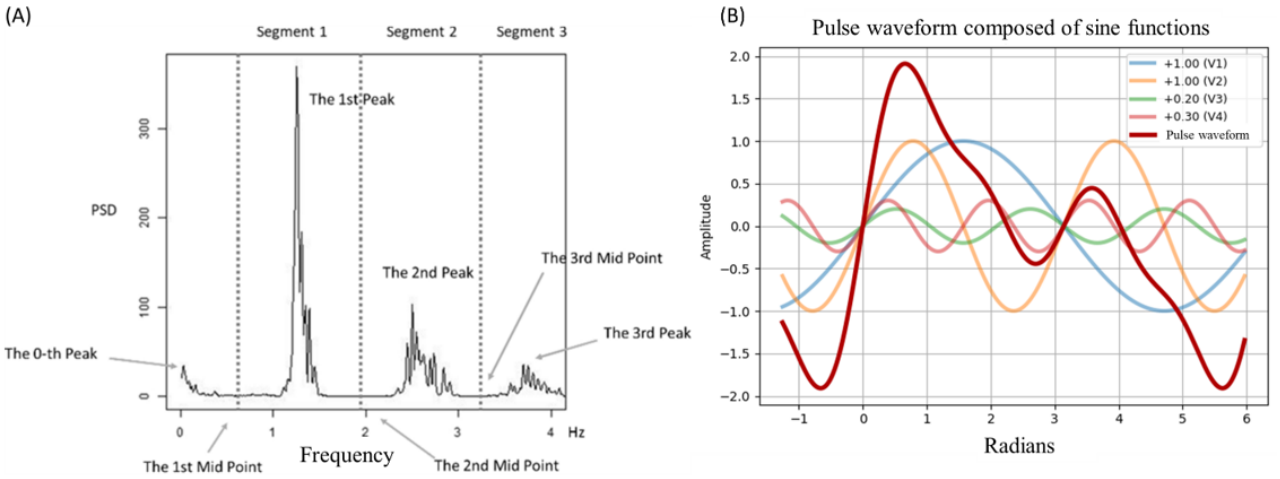
XSL•FO
RenderX

Name	Definition and calculation details
H	The fourth local minimum of the second derivative is determined by the point with the smallest second derivative between point F and RV.
B/A, E/A, F/A, G/A, and H/A	|B/A|, |E/A|, |F/A|, |G/A| and |H/A|. Since B/A, F/A, and H/A were negative, we converted them to their absolute values for more intuitive interpretation. C/A and D/A were not used in this study due to difficulties in identifying points C and D under poor signal quality.
AI[w]	AI is typically defined as B/A – (C/A + D/A + E/A) [51]. However, in this study, AI is defined as the difference between B/A and E/A only, due to difficulties in identifying points C and D under poor signal quality, and because the values of C and D are close to zero.
Frequency-domain features	
PSDi[x]	The PSD function of the pulse waveform exhibits distinct peaks corresponding to fundamental sine-wave components and their harmonics. Traditional approaches often quantify harmonic strength by peak heights; however, this metric is vulnerable to random noise. To address this, the strength of the i-th harmonic (S) is defined as the area under the PSD curve between the midpoint separating the ith and (i + 1)th peaks and the subsequent midpoint. Each midpoint is located halfway between 2 consecutive spectral peaks. This interval-based integration reduces the influence of noise on individual peak amplitudes and yields a more robust estimate of harmonic strength by effectively averaging the PSD over each frequency band (Figure 3). Since the is influenced by signal quality, the value of S_0typically represents the average energy of the entire waveform, which is less relevant to the waveform contour. Additionally, the values of $S_i$for i>6 are usually very small. Therefore, the energy strength of the i-th harmonic (PSD_i) is normalized as: .
Vi	The weight of the i-th sinusoidal function that composes the waveform. Since any function can be decomposed into a Fourier series, Liu et al [39] proposed using the sum of sinusoidal functions to represent the waveform for further analysis. So, for each sample, f(t) can be decomposed as .
NHA[y]	[52]
IHAR[z]	(1-NHA)/IPA [52]

[a]SI: stiffness index.
[b]PPT: peak-to-peak time.
[c]ESP: early systolic peak.
[d]LP: left peak.
[e]FP: first peak.
[f]DP: diastolic peak.
[g]RV: right valley.
[h]BBI: beat-to-beat interval.
[i]LV: left valley.
[j]RI: reflection index.
[k]DPH: diastolic peak height.
[l]ESPH: estimated early systolic peak height.
[m]ERI: estimated reflection index.
[n]ESI: estimated SI.
[o]CT: crest time.
[p]NT: notch time.
[q]IP: inflection point.
[r]DP: dicrotic peak.
[s]DN: dicrotic notch.
[t]IPP: inflection point percentage.
[u]DT: diastolic time.
[v]IS: Inflection slope.
[w]AI: aging index.
[x]PSDi: the i-th relative power spectrum density.
[y]NHA: normalized harmonic area.
[z]IHAR: inflection and harmonic area ratio.

XSL•FO
RenderX

Figure 3. (A) An illustration of the power spectrum density generated by fast Fourier transform (reproduced from Liu et al [13], which is published under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License [17]). (B) An illustration of pulse waveform segmentation using sinusoidal functions.



Psychological Measures

To assess well-being, 3 self-report measures were administered: the Satisfaction With Life Scale (SWLS) [53], the Subjective Vitality Scale (SVS) [54], and the positive and negative affect schedule (PANAS) [55]. The SWLS comprises 5 items rated on a 7-point Likert scale ranging from "strongly disagree" to "strongly agree." The SVS comprises 7 statements scored on a 7-point scale from "not at all true" to "very true." The PANAS consists of 20 adjectives assessing positive and negative affect over the past week, with responses provided on a 5-point scale (1="very slightly or not at all," 5="extremely"). Separate scores for positive affect (PANAS-P) and negative affect (PANAS-N) were computed to enhance interpretability.

Common mental health measures included the Patient Health Questionnaire-9 (PHQ-9) [56] and the Generalized Anxiety Disorder-7 (GAD-7) scale [57]. The PHQ-9 includes 9 items rated from 0 ("not at all") to 3 ("nearly every day"), with total scores categorizing depressive symptom severity from minimal to severe. The GAD-7 comprises 7 items using the same response scale and classifies anxiety levels as normal, mild, moderate, or severe based on the summed score.

Momentary emotional states were captured using the Self-Assessment Manikin (SAM) [58], a nonverbal pictorial instrument representing affective valence and arousal. Participants selected the figure that best reflected their current level of pleasure (valence) and activation (arousal). The dominance dimension was omitted, as it was not directly relevant to the study's objectives. All questionnaires were completed before signal collection.

Data Analysis

Associations between individual pulse-waveform features and psychological measures were first examined using simple linear regression, which provides a transparent and interpretable baseline for evaluating feature–outcome relationships and facilitates comparison with prior psychophysiological studies relying on univariate analyses.

The data exhibited a nested structure: psychological scales were administered once per participant, whereas pulse-waveform features consisted of multiple repeated measurements per participant. Treating all observations as independent under standard ordinary least squares assumptions would therefore be inappropriate and could result in underestimated standard errors and biased statistical inference. To explicitly address this nonindependence, 2 complementary analytic strategies were implemented.

The first strategy involved participant-level aggregation, whereby waveform features were summarized for each individual using median values and were entered into ordinary least squares models. This approach mitigates within-subject dependence by aligning the unit of analysis for physiological features with that of the psychological measures. The second strategy applied cluster-robust standard errors at the original observation level, allowing statistical inference to account for repeated measurements within participants while preserving within-person variability.

From a theoretical perspective, participant-level aggregation may be more appropriate in this context, as cluster-robust standard errors can be less suitable when the dependent variable is constant across observations within the same cluster [59]. However, this aggregation approach necessarily entails a loss of within-participant information. Taken together, these 2 analytical strategies offer complementary protections against bias arising from nonindependent observations.

In addition, because multiple waveform features were tested following feature selection, statistical significance was evaluated using a Bonferroni-corrected threshold [60,61]. Although conservative, this correction was adopted to control the error rate, with the acknowledgment that it may exclude features of potential substantive relevance [62].

To evaluate predictive utility beyond univariate linear models, machine learning methods were further used to model psychological outcomes from multivariate pulse-waveform features. These methods are well-suited to this context because they can capture nonlinear relationships, complex feature

XSL•FO
RenderX

interactions, and high-dimensional structures that may not be adequately modeled using traditional regression. Model performance was assessed using 5-fold cross-validation to obtain stable estimates of generalization performance while maintaining sufficient training data in each fold. Among the evaluated algorithms, random forest (RF) models were selected due to their robustness to multicollinearity, resistance to overfitting in moderate sample sizes, and ability to model nonlinear associations without strong parametric assumptions.

In each fold, the RF model was trained on data from a subset of participants and evaluated on held-out participants. Performance was quantified using out-of-sample metrics, including mean absolute error and the correlation between predicted and observed values, averaged across folds. The RF models were implemented using the default hyperparameters provided by scikit-learn. No manual hyperparameter tuning was performed. Specifically, the number of trees was set to 100, node splitting used the mean squared error criterion, trees were grown to full depth (no maximum depth specified), and bootstrap sampling was enabled. Other parameters remained at their library default values.

To prevent data leakage and preserve independence between training and evaluation, all observations from a given participant were assigned exclusively to either the training or test set. For all data analyses, an IQR multiplier of 1.5 was applied for outlier removal.

Agreement between predicted and observed values was further assessed using Bland–Altman analysis. Unlike correlation-based metrics, this approach evaluates absolute agreement and identifies potential systematic bias between predictions and ground-truth values, providing a complementary assessment of model validity that is particularly relevant for evaluating practical predictive performance.

For oximeter-photoplethysmography agreement analysis, 3 time-domain pulse waveform features commonly reported in prior studies, RI, FN, and diastolic time, were selected as representative indicators. Pearson correlation analyses were conducted to assess the linear association between the 2 measurement systems for each feature. Because correlation does not imply measurement agreement or interchangeability, Bland–Altman analysis was additionally performed to evaluate method agreement by plotting the differences between measurements against their means, thereby visualizing systematic bias and the 95% limits of agreement. The mean difference indicates directional bias, whereas the limits of agreement define the expected range of measurement differences.

Agreement was considered acceptable when most observations fell within the limits of agreement without systematic trends across the measurement range. Together, correlation and Bland–Altman analyses enabled a comprehensive evaluation of both association strength and measurement agreement between the smartphone-based photoplethysmography system and the clinical-grade fingertip oximeter.

Ethical Considerations

The study protocol was approved by the Ethics Board of the Department of Psychology at Tsinghua University (number 201910). All study procedures were conducted in accordance with institutional and national ethical guidelines for human subjects research. Written informed consent was obtained from all participants prior to participation.Each participant received 50 renminbi (approximately US $7) as compensation for participation. Participants were informed about the study procedures and their right to withdraw at any time without penalty. All data were deidentified prior to analysis to protect participant privacy. Access to the original identifiable data were restricted to the first author, and no personally identifiable information was included in the analyses or reported results.

Results

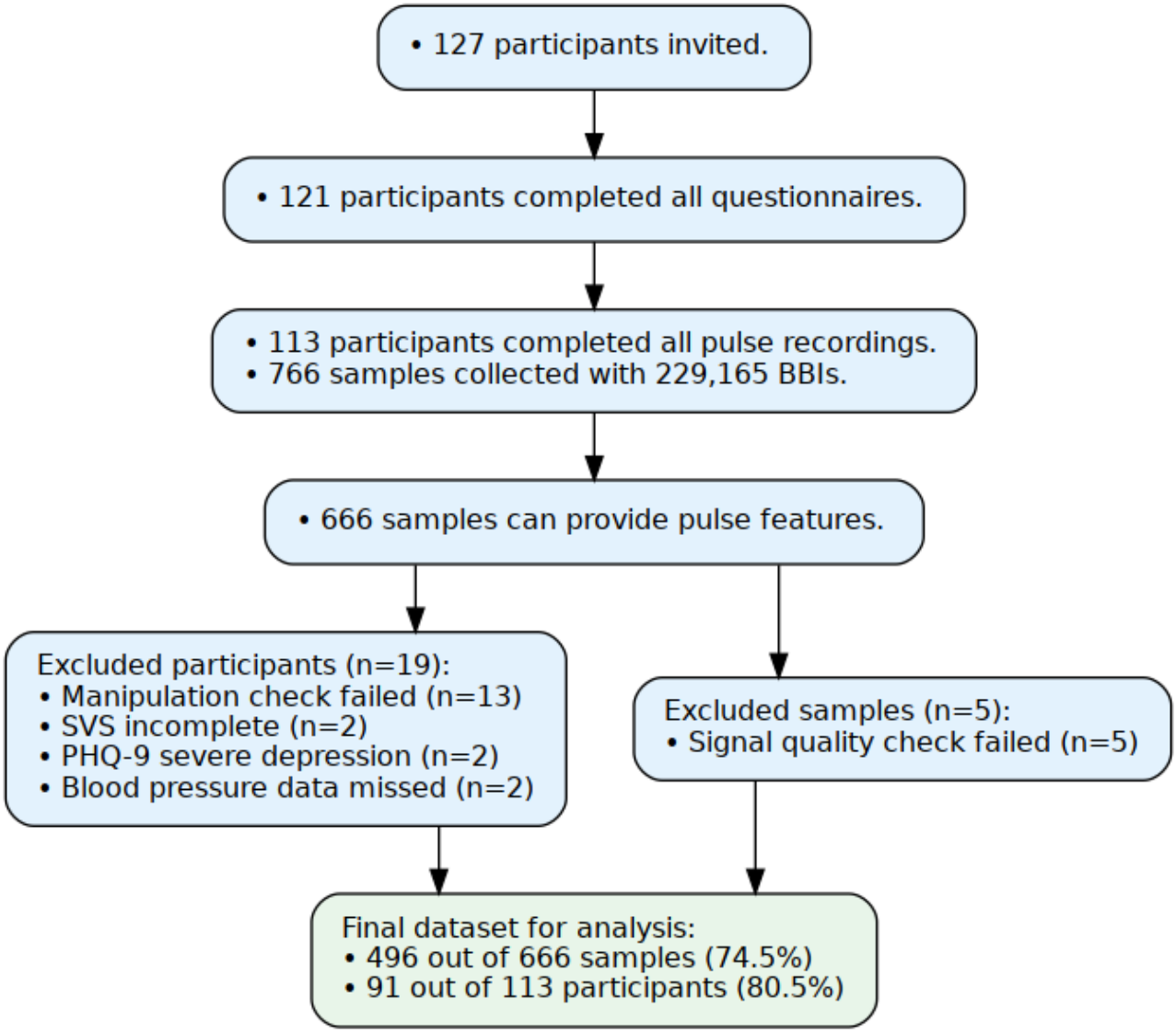
Data Collection

Of the 127 participants recruited, 121 completed all questionnaires (71/121, 58.68% male; mean age 22.79, SD 1.87 years), and 113 of these participants completed the pulse-waveform recording (67/113, 59.29% male; mean age 22.74, SD 1.83 years). The 2 participants who exhibited severe depressive symptoms, as indicated by questionnaire assessments, were excluded at this stage.

From the remaining 113 participants, a total of 766 waveform samples were collected (mean 6.78 samples per participant), comprising 229,165 BBIs (BBIs; mean 299 BBIs per sample). Of these, 100 samples were excluded due to partial feature loss (666 remaining; 86.9%), primarily resulting from poor signal quality that prevented reliable peak and valley detection (eg, HR could not be computed), although all recordings were successfully completed. Additionally, 19 participants were excluded because of problematic questionnaire responses identified through manipulation checks or severe depressive symptoms, and 5 samples with a signal quality index below 0.5, as proposed by Liu et al [39], were removed. After all exclusions, 496 valid samples from 91 participants were retained for analysis (Figure 4).

XSL•FO
RenderX

Figure 4. Participant and data filtering process. BBI: beat-to-beat interval; PHQ-9: Patient Health Questionnaire-9; SVS: Subjective Vitality Scale.

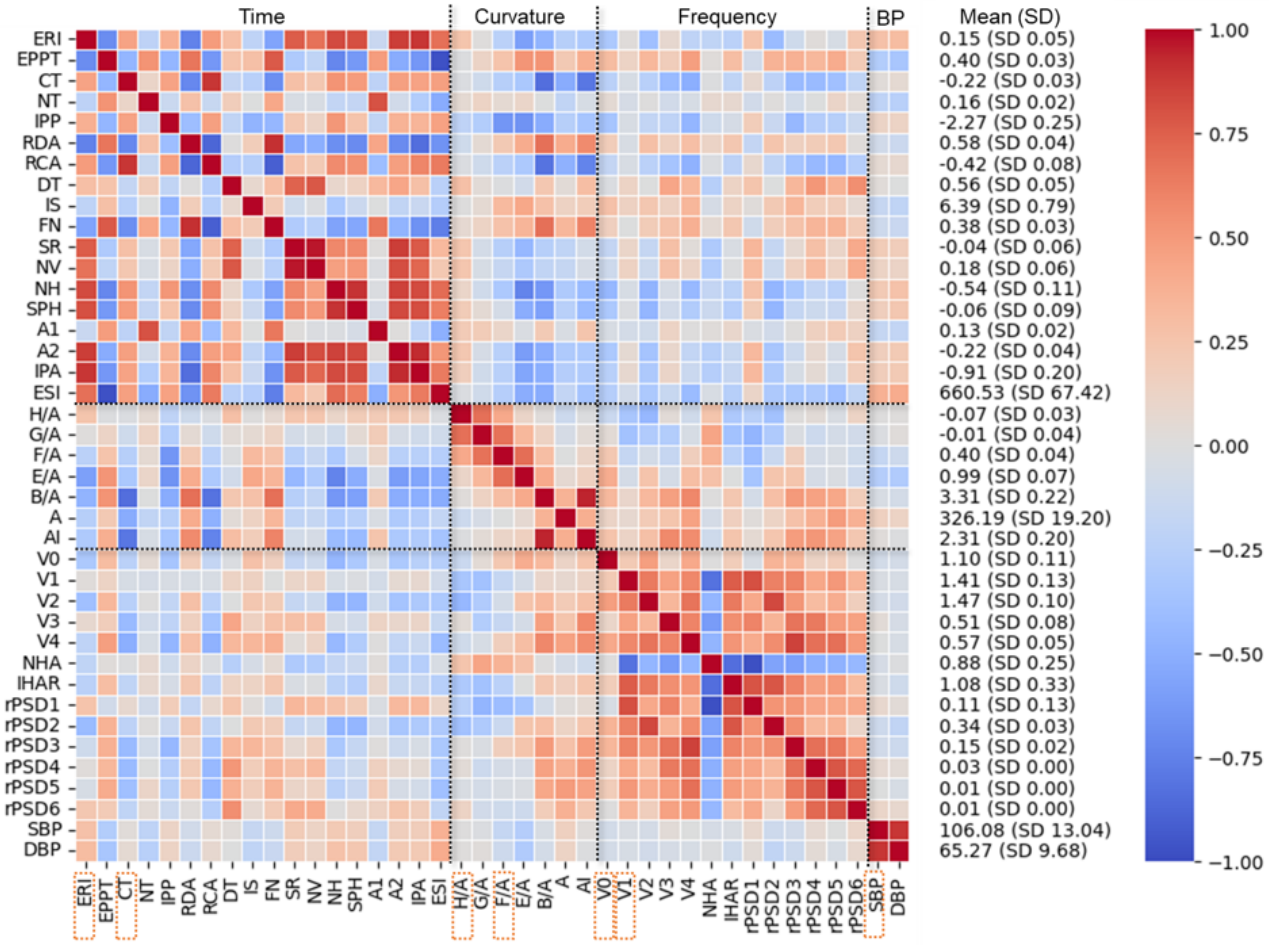


Correlation Among Pulse Waveform Features

Features exhibited strong within-domain correlations across the time, curvature, and frequency domains, whereas cross-domain correlations were comparatively weak (Figure 5). To address high collinearity in the data, a systematic feature-selection procedure was applied prior to statistical modeling to reduce multicollinearity and improve model stability, parameter estimation, and inferential validity. Within each domain, feature pairs with absolute correlations exceeding a prespecified threshold of 0.5 were considered highly collinear and addressed through sequential elimination, after which the top 2 variables were retained for further analysis. Following this procedure, the final feature set used in subsequent analyses comprised estimated reflection index, crest time (CT), the third curvature minimum (F/A), the fourth curvature minimum (H/A), the first power spectrum density component (rPSD1), the baseline of Fourier decomposition (V0), and systolic blood pressure.

XSL•FO
RenderX

Figure 5. The correlation matrix of pulse waveform features. The orange boxes indicate the selected variables after correlation coefficient threshold filtering. DBP: diastolic blood pressure; DT: diastolic time; EPPT: estimated PPT; ERI: estimated reflection index; ESI: estimated SI; ESP: the early systolic peak; ESPH: early systolic peak height; IP: inflection point; IPA: inflection point area. (A2/A1); IPP: inflection point percentage; IS: inflection slope; NH: notch height; NT: notch time; NV: the interval between the second peak to right valley; PPT: peak-to-peak time; PSD: power spectrum density; RCA: the ratio of CT to NT; RDA: the ratio of NT to the combined duration of CT and DT; RI: reflection index; RV: right valley; SBP: systolic blood pressure; SI: stiffness index; SP: second peak; SPH: second peak height; SR: the interval between the second peak to right valley.

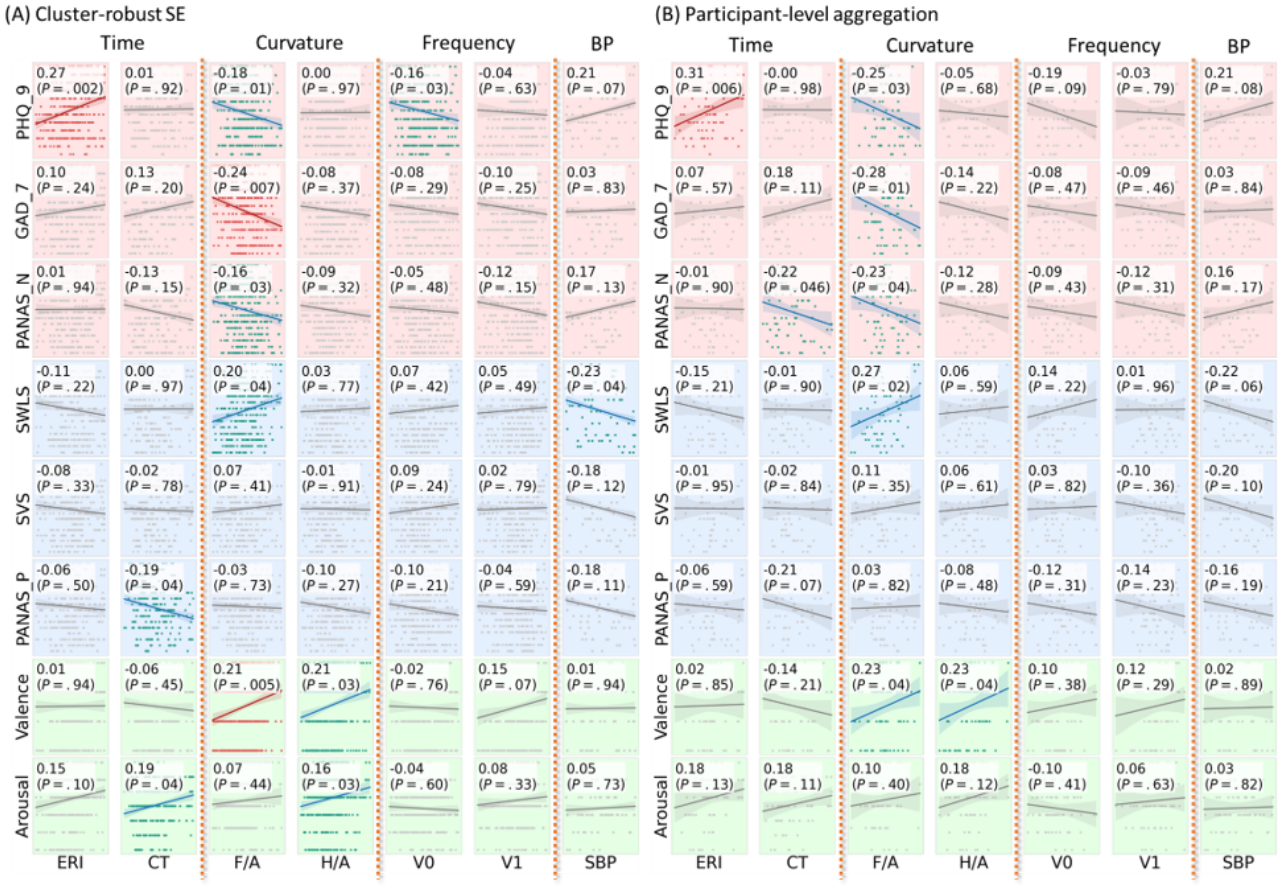


Linear Regression and Correlation Analysis

In univariate regression and correlation analyses, we examined associations among 8 psychological measures, 9 pulse waveform features, and blood pressure. The results indicated that negative psychological states were primarily associated with time- and curvature-domain waveform features (Figure 6). Depressive symptoms (PHQ-9) were significantly associated with F/A and V0, with the association between PHQ-9 and estimated reflection index remaining significant after Bonferroni correction. Generalized anxiety (GAD-7) also showed a Bonferroni-corrected significant association with F/A. Negative affect (PANAS-N) was associated with both CT and F/A.

XSL•FO
RenderX

Figure 6. Univariate relationships between each independent and dependent variable. Green markers and regression lines denote *P* value less than .05 and red markers denote *P* value less than the Bonferroni corrected criteria .05/9=.0056. CT: crest time; ERI: estimated reflection index; GAD-7: Generalized Anxiety Disorder-7; PANAS: positive and negative affect schedule; PANAS-N: negative affect; PANAS-P: positive affect; PHQ-9: Patient Health Questionnaire-9; SBP: systolic blood pressure; SVS: Subjective Vitality Scale; SWLS: Satisfaction with Life Scale.



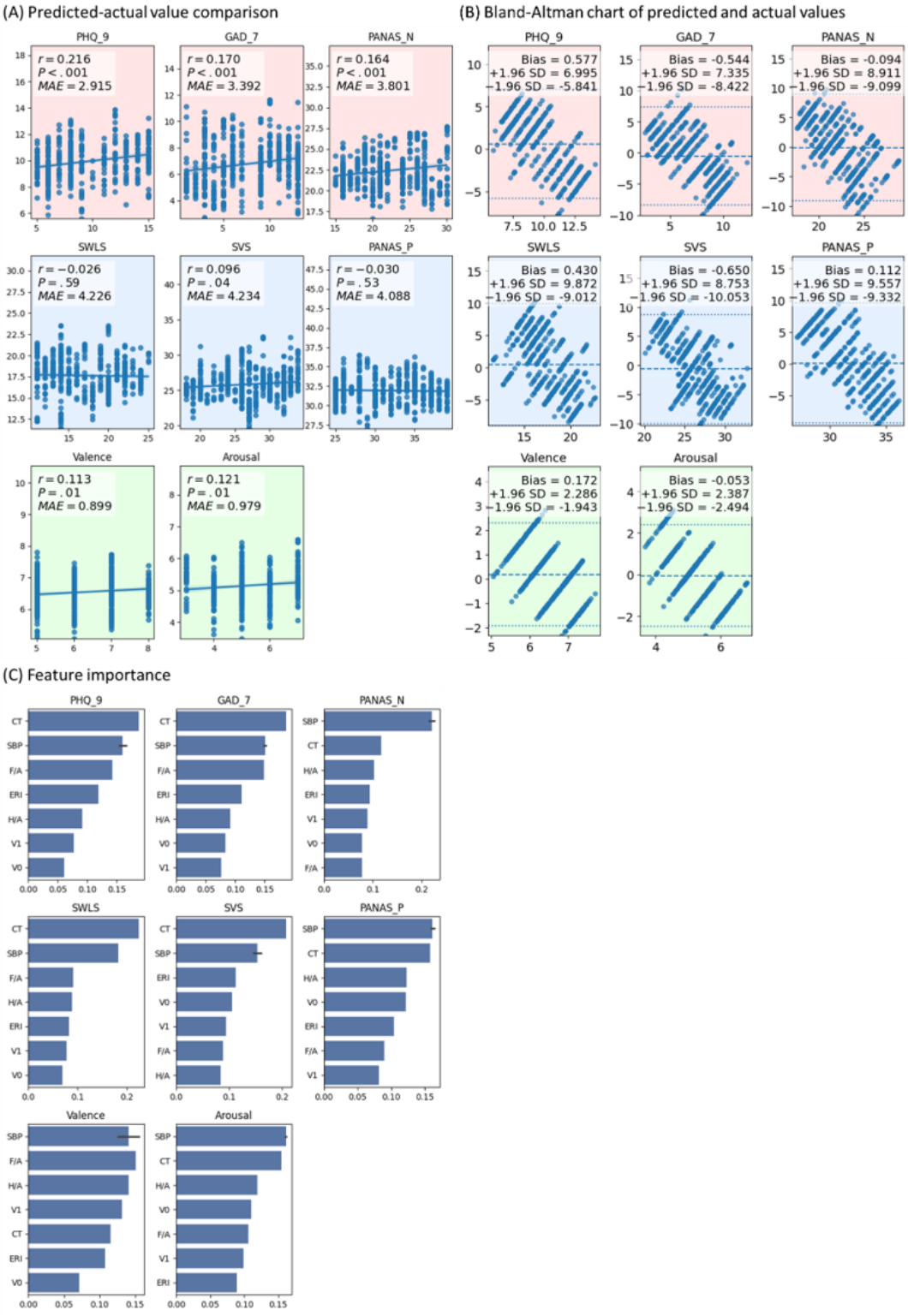
In contrast, positive affect exhibited fewer significant associations with waveform features: SWLS was correlated with F/A, and PANAS-P was correlated with CT. Regarding momentary emotional states, valence was associated with F/A and H/A, whereas arousal was associated with CT and H/A.

Machine Learning

In this study, 9 features selected through feature engineering were used as inputs to predict mental health outcomes. The machine learning results (Figure 7) indicated that waveform-related features were effective in predicting mental disorders (*P*<.001 for PHQ-9, GAD-7, and PANAS-N), whereas predictive performance for positive emotional states was substantially weaker. With respect to short-term emotional fluctuations, heartbeat waveform features showed no clear association with valence but were evidently influenced by arousal. Bland–Altman analysis further demonstrated that, for outcomes with significant predictive correlations, the machine learning predictions exhibited slight but not substantial systematic bias. The 3 most important features identified by the RF models across all dependent variables were CT, F/A, and systolic blood pressure.

XSL•FO
RenderX

Figure 7. Machine learning results. GAD-7: Generalized Anxiety Disorder-7; MAE: mean absolute error; PANAS: positive and negative affect schedule; PANAS-N: negative affect; PANAS-P: positive affect; PHQ-9: Patient Health Questionnaire-9; SVS: Subjective Vitality Scale; SWLS: Satisfaction with Life Scale.

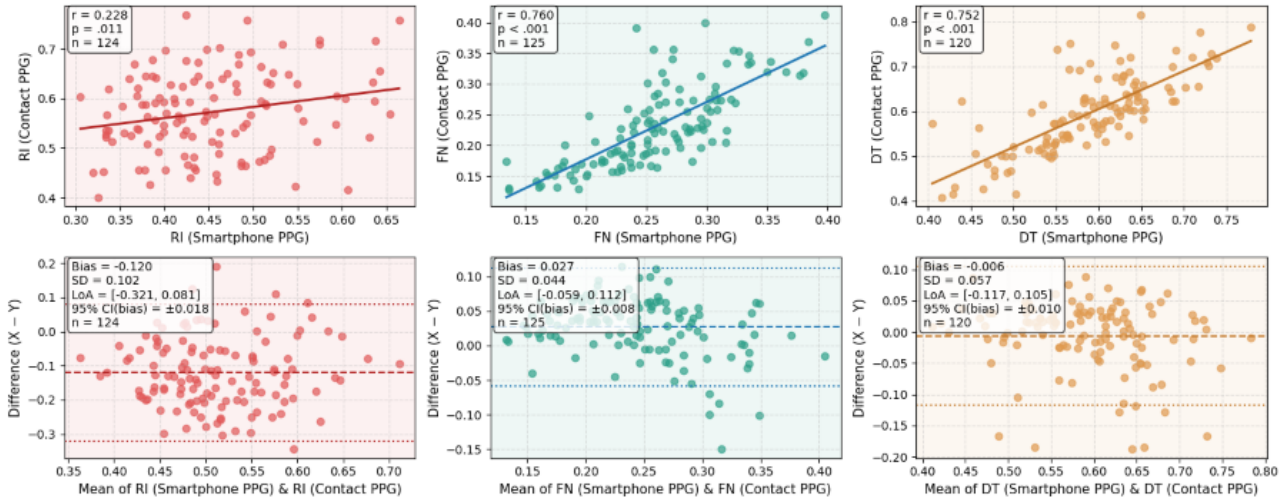


Oximeter Comparison

The results (Figure 8) indicate reasonable agreement between waveform features obtained from the oximeter and those derived from the smartphone-based photoplethysmography method. For all 3 features analyzed, correlation coefficients were statistically significant (P<.05), and Bland–Altman analyses revealed minimal systematic bias. However, consistent with the optical boundary effects inherent to smartphone photoplethysmography acquisition discussed in the waveform metrics section, the reflection index exhibited some degree of distortion. In contrast, time-related waveform features were largely unaffected by these optical limitations and showed more consistent agreement across the 2 measurement modalities.

XSL•FO
RenderX

Figure 8. The agreement between the proposed smartphone photoplethysmography-based pulse waveform features and those obtained by oximeter. DT: diastolic time; FN: the interval between first peak to notch; PPG: photoplethysmography; RI: reflection index.



Discussion

Principal Findings

Overall, the findings support the study's hypotheses and objectives by demonstrating that fingertip pulse-waveform signals recorded using a standard smartphone can be used to predict individual psychological health status. This work makes 3 principal contributions.

First, we introduce and validate a cost-effective, noninvasive approach that relies solely on a smartphone camera to capture subtle skin color variations, reconstruct digital arterial pulse waveforms, and infer mental health status. Across both univariate linear regression and multivariate machine learning analyses, waveform features showed consistent and statistically significant associations with psychological indicators. Beyond extending smartphone-based waveform analysis to psychological research, these findings provide additional empirical support for pulse-waveform–based psychological assessment more broadly [34,35], as well as for the use of smartphones in related physiological applications such as HRV and blood pressure analysis [63,64].

Second, our results reveal associations between pulse-waveform features and psychological variables across multiple feature domains. Consistent with findings reported by Kontaxis et al [34] and Celka et al [65], negative psychological conditions were strongly linked to waveform characteristics, suggesting that long-term psychological states may induce structural or functional changes in the arterial system. However, contrary to our initial expectations, these associations were primarily observed in curvature- and time-domain features rather than in the frequency domain. As one of the first studies to incorporate frequency-domain pulse-waveform features into mental health research, we suggest that such features may encode relevant information, although this was not supported by the present dataset.

The relatively weaker associations observed between positive affect and pulse-waveform features were also unexpected but appear reasonable in retrospect. Given the well-documented negative correlation between positive and negative affect and evidence that positive affect plays a protective role against adverse psychological states [66], we anticipated that positive affect would exhibit associations with waveform features comparable to those observed for negative affect. Although the observed relationships were directionally consistent with this hypothesis, they were generally weaker and less statistically robust. This pattern remains theoretically consistent with prior psychological research, which emphasizes that positive and negative affect are distinct, partially independent constructs that can co-occur within individuals despite their negative correlation [67]. Together, these findings suggest that both affective dimensions relate to cardiovascular functioning, but that negative affect exerts a more pronounced influence.

Short-term emotional fluctuations appear to exert limited influence on vascular wall properties. However, because arousal is closely linked to HRV [68], which directly affects the timing of cardiac contractions, it is reasonable that arousal was associated primarily with time-domain waveform features. This interpretation aligns with earlier work in affective computing by Healey and Picard [69], who demonstrated that physiological signals reliably capture arousal but are less effective at distinguishing affective valence. Although the present findings are supported by emerging evidence, further studies using higher-quality data and more controlled experimental designs are needed to confirm and extend these results.

Limitations and Future Directions

Despite the theoretical and practical innovations of this study, 3 limitations warrant cautious interpretation of the results.

First, although the machine learning models achieved significant predictive performance, the level of agreement between the proposed method and the target psychological measures remained limited. Similarly, comparisons between waveform features derived from the smartphone-based photoplethysmography method and those obtained from a reference oximeter demonstrated statistically significant associations, but the magnitude of agreement was insufficient for precise interchangeability. Taken together, these findings indicate that, while pulse waveform features are systematically

XSL•FO
RenderX

influenced by mental health status, the current models primarily capture correlational relationships rather than yielding predictions with practical accuracy.

Second, of the 766 waveform samples collected, 100 (13.1%) were classified as corrupted and excluded, a substantial loss rate that may undermine the study's validity. In uncontrolled, real-world settings, slight finger movement or fluctuations in ambient illumination can distort the photoplethysmography signals [70]. Future research can therefore use enhanced image capture protocols [71], integrate advanced signal processing and noise reduction algorithms [72], and implement quality filters to discard low-fidelity recordings [73]. Moreover, the limited and variable frame rates of smartphone cameras (typically 10-60 Hz) constrain temporal resolution compared with medical-grade devices sampling up to 1000 Hz [74]. Because smartphone camera frame rates are not fixed, this instability further exacerbates data-collection challenges [75]. The effects of reduced and unstable sampling rates on photoplethysmography signal integrity remain poorly understood and warrant further investigation [76].

Third, the sample consisted primarily of young adults with similar educational backgrounds, which may limit the generalizability of the findings. Because vascular changes associated with psychological factors often develop over time, associations between mental health and pulse-waveform morphology may be more pronounced in older adults or individuals with chronic conditions. Future studies should therefore include participants across broader age ranges and with more diverse health profiles. Additionally, the slight gender imbalance in the sample may have introduced bias in sex-specific waveform effects. Other potential confounders, such as sleep quality, smoking, diet, and medication use, were not controlled for and should be addressed in future research to reduce unexplained variance.

Conclusions

Pulse waveform analysis has traditionally been applied to the assessment of physical health, whereas its extension to mental health, via links between psychological states and vascular stiffness, remains relatively novel. Conventional waveform acquisition methods are often costly and inaccessible, which has limited their adoption in psychological research. This study uses smartphone-based photoplethysmography to analyze digital arterial waveforms, thereby providing a noninvasive and accessible approach for psychological assessment. The findings demonstrate significant associations between psychological variables and pulse-waveform features, particularly within the time- and curvature-domains, and thus contribute to the advancement of psychophysiological research grounded in hemodynamic principles. Nevertheless, technical limitations (eg, signal variability and low, unstable sampling rates) and sample-related constraints (eg, demographic homogeneity and uncontrolled confounders) should be addressed in future studies.

Acknowledgments

Generative artificial intelligence (ChatGPT) was used for proofreading and in preparing Figure 2.

Funding

This work was supported by the Shenzhen R&D Sustainable Development Funding (KCXFZ20230731093600002), Guangdong Digital Mental Health and Intelligent Generation Laboratory (2023WSYS010), Guangdong Philosophy and Social Science Foundation (GD23SQXY01), Shenzhen Education Science Policy Research Project (ZDZC23013), National Natural Science Foundation of China (32371121), Guangdong Science and Technology Programme (2024A0505050012), Guangdong Medical Scientific Research Foundation (B2024262), and Zhuhai City Medical Scientific Research Foundation (2420009000094).

Data Availability

The datasets generated or analyzed during this study are available in the GitHub repository [77].

Authors' Contributions

Conceptualization: IL
Data curation: IL
Funding acquisition: SN
Investigation: IL
Supervision: SN
Writing – original draft: IL (lead) and QZ (supporting)
Writing – review & editing: IL (lead), LH (supporting), JL (supporting), and CL (supporting)

Conflicts of Interest

None declared.

References

XSL•FO
RenderX

1. Ghasemzadeh N, Zafari AM. A brief journey into the history of the arterial pulse. Cardiol Res Pract 2011;2011:164832 [FREE Full text] [doi: 10.4061/2011/164832] [Medline: 21811677]
2. de Sá Ferreira A, Lopes AJ. Pulse waveform analysis as a bridge between pulse examination in Chinese medicine and cardiology. Chin J Integr Med 2013;19(4):307-314. [doi: 10.1007/s11655-013-1412-z] [Medline: 23546634]
3. Weber T, Auer J, O'Rourke MF, Kvas E, Lassnig E, Berent R, et al. Arterial stiffness, wave reflections, and the risk of coronary artery disease. Circulation 2004;109(2):184-189. [doi: 10.1161/01.CIR.0000105767.94169.E3] [Medline: 14662706]
4. Kohara K, Tabara Y, Oshiumi A, Miyawaki Y, Kobayashi T, Miki T. Radial augmentation index: a useful and easily obtainable parameter for vascular aging. Am J Hypertens 2005;18(1 Pt 2):11S-14S. [doi: 10.1016/j.amjhyper.2004.10.010] [Medline: 15683726]
5. Yim YK, Lee C, Lee HJ, Park KS. Gender and measuring-position differences in the radial pulse of healthy individuals. J Acupunct Meridian Stud 2014;7(6):324-330 [FREE Full text] [doi: 10.1016/j.jams.2014.06.014] [Medline: 25499566]
6. Addison PS. Respiratory effort from the photoplethysmogram. Med Eng Phys 2017;41:9-18 [FREE Full text] [doi: 10.1016/j.medengphy.2016.12.010] [Medline: 28126420]
7. Khanokh B, Slovik Y, Landau D, Nitzan M. Sympathetically induced spontaneous fluctuations of the photoplethysmographic signal. Med Biol Eng Comput 2004;42(1):80-85. [doi: 10.1007/BF02351014] [Medline: 14977226]
8. Baruch MC, Warburton DE, Bredin SS, Cote A, Gerdt DW, Adkins CM. Pulse decomposition analysis of the digital arterial pulse during hemorrhage simulation. Nonlinear Biomed Phys 2011;5(1):1 [FREE Full text] [doi: 10.1186/1753-4631-5-1] [Medline: 21226911]
9. Liu C, Zheng D, Zhao L, Liu C. Gaussian fitting for carotid and radial artery pressure waveforms: comparison between normal subjects and heart failure patients. Biomed Mater Eng 2014;24(1):271-277. [doi: 10.3233/BME-130808] [Medline: 24211907]
10. Savage MT, Ferro CJ, Pinder SJ, Tomson CRV. Reproducibility of derived central arterial waveforms in patients with chronic renal failure. Clin Sci (Lond) 2002;103(1):59-65. [doi: 10.1042/cs1030059] [Medline: 12095404]
11. Gajdova J, Karasek D, Goldmannova D, Krystynik O, Schovanek J, Vaverkova H, et al. Pulse wave analysis and diabetes mellitus. A systematic review. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub 2017;161(3):223-233 [FREE Full text] [doi: 10.5507/bp.2017.028] [Medline: 28627523]
12. Zanoli L, Rastelli S, Inserra G, Castellino P. Arterial structure and function in inflammatory bowel disease. World J Gastroenterol 2015;21(40):11304-11311 [FREE Full text] [doi: 10.3748/wjg.v21.i40.11304] [Medline: 26523102]
13. Liu ISC, Liu F, Zhong Q, Ni S. A finger on the pulse of cardiovascular health: estimating blood pressure with smartphone photoplethysmography-based pulse waveform analysis. Biomed Eng Online 2025;24(1):36 [FREE Full text] [doi: 10.1186/s12938-025-01365-w] [Medline: 40108587]
14. Millasseau SC, Ritter JM, Takazawa K, Chowienczyk PJ. Contour analysis of the photoplethysmographic pulse measured at the finger. J Hypertens 2006;24(8):1449-1456. [doi: 10.1097/01.hjh.0000239277.05068.87] [Medline: 16877944]
15. O'Rourke MF, Pauca A, Jiang XJ. Pulse wave analysis. Br J Clin Pharmacol 2001;51(6):507-522 [FREE Full text] [doi: 10.1046/j.0306-5251.2001.01400.x] [Medline: 11422010]
16. Seldenrijk A, van Hout HP, van Marwijk HWJ, de Groot E, Gort J, Rustemeijer C, et al. Depression, anxiety, and arterial stiffness. Biol Psychiatry 2011;69(8):795-803. [doi: 10.1016/j.biopsych.2010.12.034] [Medline: 21334599]
17. Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by-nc-nd/4.0/ [accessed 2026-03-13]
18. Attribution 2.0 Generic (CC BY 2.0). Creative Commons. URL: https://creativecommons.org/licenses/by/2.0/ [accessed 2026-03-13]
19. Lacolley P, Regnault V, Segers P, Laurent S. Vascular smooth muscle cells and arterial stiffening: relevance in development, aging, and disease. Physiol Rev 2017;97(4):1555-1617 [FREE Full text] [doi: 10.1152/physrev.00003.2017] [Medline: 28954852]
20. Howren MB, Lamkin DM, Suls J. Associations of depression with C-reactive protein, IL-1, and IL-6: a meta-analysis. Psychosom Med 2009;71(2):171-186. [doi: 10.1097/PSY.0b013e3181907c1b] [Medline: 19188531]
21. Lambert EA, Thompson J, Schlaich M, Laude D, Elghozi JL, Esler MD, et al. Sympathetic and cardiac baroreflex function in panic disorder. J Hypertens 2002;20(12):2445-2451. [doi: 10.1097/00004872-200212000-00025] [Medline: 12473870]
22. Vlachopoulos C, Kosmopoulou F, Alexopoulos N, Ioakeimidis N, Siasos G, Stefanadis C. Acute mental stress has a prolonged unfavorable effect on arterial stiffness and wave reflections. Psychosom Med 2006;68(2):231-237. [doi: 10.1097/01.psy.0000203171.33348.72] [Medline: 16554388]
23. Wu Y, Sun D, Wang B, Li Y, Ma Y. The relationship of depressive symptoms and functional and structural markers of subclinical atherosclerosis: A systematic review and meta-analysis. Eur J Prev Cardiol 2018;25(7):706-716. [doi: 10.1177/2047487318764158] [Medline: 29537293]
24. Rybakowski JK, Wykretowicz A, Heymann-Szlachcinska A, Wysocki H. Impairment of endothelial function in unipolar and bipolar depression. Biol Psychiatry 2006;60(8):889-891. [doi: 10.1016/j.biopsych.2006.03.025] [Medline: 16730331]
25. Yeragani VK, Tancer M, Seema K, Josyula K, Desai N. Increased pulse-wave velocity in patients with anxiety: implications for autonomic dysfunction. J Psychosom Res 2006;61(1):25-31. [doi: 10.1016/j.jpsychores.2005.10.011] [Medline: 16813842]

XSL•FO
RenderX

26. Cicek Y, Durakoglugil ME, Kocaman SA, Guveli H, Cetin M, Erdogan T, et al. Increased pulse wave velocity in patients with panic disorder: independent vascular influence of panic disorder on arterial stiffness. J Psychosom Res 2012;73(2):145-148. [doi: 10.1016/j.jpsychores.2012.05.012] [Medline: 22789419]
27. Sanchez-Gonzalez MA, May RW, Brown PC, Koutnik AP, Fincham FD. Depressive symptoms contribute to increased wave reflection during cold pressor test in young adult men. Am J Hypertens 2013;26(6):778-783. [doi: 10.1093/ajh/hpt012] [Medline: 23449606]
28. Yanartas O, Sunbul M, Durmus E, Kivrak T, Senkal Z, Subasi N, et al. Severity of depression and anxiety symptoms is associated with increased arterial stiffness in depressive disorder patients undergoing psychiatric treatment. Psychiatry Clin Psychopharmacol 2016;26(3):287-293. [doi: 10.5455/bcp.20160325085828]
29. Li F, Yang L, Shi H, Liu C. Differences in photoplethysmography morphological features and feature time series between two opposite emotions: Happiness and sadness. Artery Research 2017;18:7-13. [doi: 10.1016/j.artres.2017.02.003]
30. Park YB, Park YJ, Ko YI. Relationships of pulse waveform parameters to mood states and chronic fatigue. J Altern Complement Med 2012;18(11):1050-1060. [doi: 10.1089/acm.2011.0430] [Medline: 23072268]
31. Celka P, Charlton PH, Farukh B, Chowienczyk P, Alastruey J. Influence of mental stress on the pulse wave features of photoplethysmograms. Healthc Technol Lett 2020;7(1):7-12 [FREE Full text] [doi: 10.1049/htl.2019.0001] [Medline: 32190335]
32. Charlton PH, Celka P, Farukh B, Chowienczyk P, Alastruey J. Assessing mental stress from the photoplethysmogram: a numerical study. Physiol Meas 2018;39(5):054001 [FREE Full text] [doi: 10.1088/1361-6579/aabe6a] [Medline: 29658894]
33. Cruickshank K, Riste L, Anderson SG, Wright JS, Dunn G, Gosling RG. Aortic pulse-wave velocity and its relationship to mortality in diabetes and glucose intolerance: an integrated index of vascular function? Circulation 2002;106(16):2085-2090. [doi: 10.1161/01.cir.0000033824.02722.f7] [Medline: 12379578]
34. Kontaxis S, Gil E, Marozas V, Lazaro J, Garcia E, Posadas-de Miguel M, et al. Photoplethysmographic waveform analysis for autonomic reactivity assessment in depression. IEEE Trans Biomed Eng 2021;68(4):1273-1281. [doi: 10.1109/TBME.2020.3025908] [Medline: 32960759]
35. Udhayakumar R, Rahman S, Buxi D, Macefield VG, Dawood T, Mellor N, et al. Measurement of stress-induced sympathetic nervous activity using multi-wavelength PPG. R Soc Open Sci 2023;10(8):221382 [FREE Full text] [doi: 10.1098/rsos.221382] [Medline: 37650068]
36. Budidha K, Kyriacou PA. Photoplethysmography for quantitative assessment of sympathetic nerve activity (SNA) during cold stress. Front Physiol 2018;9:1863 [FREE Full text] [doi: 10.3389/fphys.2018.01863] [Medline: 30687108]
37. Liu I, Ni S, Peng K. Happiness at your fingertips: assessing mental health with smartphone photoplethysmogram-based heart rate variability analysis. Telemed J E Health 2020;26(12):1483-1491. [doi: 10.1089/tmj.2019.0283] [Medline: 32101084]
38. Yasmin, Brown MJ. Similarities and differences between augmentation index and pulse wave velocity in the assessment of arterial stiffness. QJM 1999;92(10):595-600. [doi: 10.1093/qjmed/92.10.595] [Medline: 10627881]
39. Liu I, Ni S, Peng K. Enhancing the robustness of smartphone photoplethysmography: a signal quality index approach. Sensors (Basel) 2020;20(7):1923 [FREE Full text] [doi: 10.3390/s20071923] [Medline: 32235543]
40. Mortensen K, Aydin MA, Rybczynski M, Baulmann J, Schahidi NA, Kean G, et al. Augmentation index relates to progression of aortic disease in adults with marfan syndrome. Am J Hypertens 2009;22(9):971-979. [doi: 10.1038/ajh.2009.115] [Medline: 19574960]
41. Sahebkar A, Pećin I, Tedeschi-Reiner E, Derosa G, Maffioli P, Reiner Ž. Effects of statin therapy on augmentation index as a measure of arterial stiffness: a systematic review and meta-analysis. Int J Cardiol 2016;212:160-168. [doi: 10.1016/j.ijcard.2016.03.010] [Medline: 27038725]
42. Stoner L, Faulkner J, Lowe A, M Lambrick D, M Young J, Love R, et al. Should the augmentation index be normalized to heart rate? J Atheroscler Thromb 2014;21(1):11-16 [FREE Full text] [doi: 10.5551/jat.20008] [Medline: 24257465]
43. Said MA, Eppinga RN, Lipsic E, Verweij N, van der Harst P. Relationship of arterial stiffness index and pulse pressure with cardiovascular disease and mortality. J Am Heart Assoc 2018;7(2):e007621. [doi: 10.1161/JAHA.117.007621] [Medline: 29358193]
44. Takazawa K, Tanaka N, Fujita M, Matsuoka O, Saiki T, Aikawa M, et al. Assessment of vasoactive agents and vascular aging by the second derivative of photoplethysmogram waveform. Hypertension 1998;32(2):365-370. [doi: 10.1161/01.hyp.32.2.365] [Medline: 9719069]
45. Saini S, Saxena Y, Gupta R. Arterial compliance and autonomic functions in adult male smokers. J Clin Diagn Res 2016;10(5):CC12-CC16 [FREE Full text] [doi: 10.7860/JCDR/2016/19547.7831] [Medline: 27437209]
46. Du Y, Wang L, Li S, Zhi G, Li D, Zhang C. Influential factors for pressure pulse waveform in healthy young adults. Biomed Mater Eng 2015;26(Suppl 1):S497-S505. [doi: 10.3233/BME-151339] [Medline: 26406041]
47. Přibil J, Přibilová A, Frollo I. First-step PPG signal analysis for evaluation of stress induced during scanning in the open-air MRI device. Sensors (Basel) 2020;20(12):3532 [FREE Full text] [doi: 10.3390/s20123532] [Medline: 32580364]
48. El-Hajj C, Kyriacou P. A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure. Biomedical Signal Processing and Control 2020;58:101870. [doi: 10.1016/j.bspc.2020.101870]

XSL•FO
RenderX

49. Park J, Seok HS, Kim SS, Shin H. Photoplethysmogram analysis and applications: an integrative review. Front Physiol 2021;12:808451 [FREE Full text] [doi: 10.3389/fphys.2021.808451] [Medline: 35300400]
50. Pierce DR, Doma K, Leicht AS. Acute effects of exercise mode on arterial stiffness and wave reflection in healthy young adults: a systematic review and meta-analysis. Front Physiol 2018;9:73 [FREE Full text] [doi: 10.3389/fphys.2018.00073] [Medline: 29487535]
51. Shin H, Noh G, Choi BM. Photoplethysmogram based vascular aging assessment using the deep convolutional neural network. Sci Rep 2022;12(1):11377. [doi: 10.1038/s41598-022-15240-4] [Medline: 35790836]
52. Wang L, Pickwell-Macpherson E, Liang YP, Zhang YT. Noninvasive cardiac output estimation using a novel photoplethysmogram index. Annu Int Conf IEEE Eng Med Biol Soc 2009;2009:1746-1749. [doi: 10.1109/IEMBS.2009.5333091] [Medline: 19963762]
53. Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. J Pers Assess 1985;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
54. Ryan RM, Frederick C. On energy, personality, and health: subjective vitality as a dynamic reflection of well-being. J Pers 1997;65(3):529-565. [doi: 10.1111/j.1467-6494.1997.tb00326.x] [Medline: 9327588]
55. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
56. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
57. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
58. Bradley MM, Lang PJ. Measuring emotion: the Self-Assessment Manikin and the semantic differential. J Behav Ther Exp Psychiatry 1994;25(1):49-59. [doi: 10.1016/0005-7916(94)90063-9] [Medline: 7962581]
59. Abadie A, Athey S, Imbens GW, Wooldridge JM. When should you adjust standard errors for clustering? The Quarterly Journal of Economics 2023;138(1):1-35. [doi: 10.1093/qje/qjac038]
60. Bender R, Lange S. Adjusting for multiple testing—when and how? J Clin Epidemiol 2001 Apr;54(4):343-349. [doi: 10.1016/S0895-4356(00)00314-0] [Medline: 11297884]
61. Thiese MS, Ronna B, Ott U. P value interpretations and considerations. J Thorac Dis 2016;8(9):E928-E931 [FREE Full text] [doi: 10.21037/jtd.2016.08.16] [Medline: 27747028]
62. Morawski K, Ghazinouri R, Krumme A, Lauffenburger JC, Lu Z, Durfee E, et al. Association of a smartphone application with medication adherence and blood pressure control: the MedISAFE-BP randomized clinical trial. JAMA Intern Med 2018;178(6):802-809 [FREE Full text] [doi: 10.1001/jamainternmed.2018.0447] [Medline: 29710289]
63. Fujita D, Suzuki A. Evaluation of the possible use of PPG waveform features measured at low sampling rate. IEEE Access 2019;7:58361-58367. [doi: 10.1109/access.2019.2914498]
64. Junior AD, Murali S, Rincon F, Atienza D. Estimation of blood pressure and pulse transit time using your smartphone. : ACM; 2015 Presented at: DSD '15: Proceedings of the 2015 Euromicro Conference on Digital System Design; August 26, 2015; United States p. 173-180 URL: https://ieeexplore.ieee.org/abstract/document/7302267/ [doi: 10.1109/dsd.2015.90]
65. Celka P, Charlton PH, Farukh B, Chowienczyk P, Alastruey J. Influence of mental stress on the pulse wave features of photoplethysmograms. Healthc Technol Lett 2020;7(1):7-12 [FREE Full text] [doi: 10.1049/htl.2019.0001] [Medline: 32190335]
66. Tugade MM, Fredrickson BL. Resilient individuals use positive emotions to bounce back from negative emotional experiences. J Pers Soc Psychol 2004;86(2):320-333. [doi: 10.1037/0022-3514.86.2.320] [Medline: 14769087]
67. Fredrickson BL. Updated thinking on positivity ratios. American Psychologist 2013;68(9):814-822 [FREE Full text] [doi: 10.1037/a0033584]
68. Wascher CAF. Heart rate as a measure of emotional arousal in evolutionary biology. Philos Trans R Soc Lond B Biol Sci 2021;376(1831):20200479 [FREE Full text] [doi: 10.1098/rstb.2020.0479] [Medline: 34176323]
69. Healey J, Picard R. Digital processing of affective signals. 1998 Presented at: Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP-98; May 12-15, 1998; Seattle, Washington, USA p. 3749-3752 URL: https://ieeexplore.ieee.org/abstract/document/679699/ [doi: 10.1109/icassp.1998.679699]
70. Garcia-Agundez A, Dutz T, Goebel S. Adapting smartphone-based photoplethysmograpy to suboptimal scenarios. Physiol Meas 2017;38(2):219-232. [doi: 10.1088/1361-6579/aa51db] [Medline: 28099163]
71. Nam Y, Nam YC. Photoplethysmography signal analysis for optimal region-of-interest determination in video imaging on a built-in smartphone under different conditions. Sensors (Basel) 2017;17(10):2385 [FREE Full text] [doi: 10.3390/s17102385] [Medline: 29048394]
72. Morelli D, Bartoloni L, Colombo M, Plans D, Clifton DA. Profiling the propagation of error from PPG to HRV features in a wearable physiological-monitoring device. Healthc Technol Lett 2018;5(2):59-64. [doi: 10.1049/htl.2017.0039] [Medline: 29750114]
73. Desquins T, Bousefsaf F, Pruski A, Maaoui C. A survey of photoplethysmography and imaging photoplethysmography quality assessment methods. Applied Sciences 2022;12(19):9582. [doi: 10.3390/app12199582]

XSL•FO
RenderX

74. Kok BE, Coffey KA, Cohn MA, Catalino LI, Vacharkulksemsuk T, Algoe SB, et al. How positive emotions build physical health: perceived positive social connections account for the upward spiral between positive emotions and vagal tone. Psychol Sci 2013;24(7):1123-1132. [doi: 10.1177/0956797612470827] [Medline: 23649562]
75. Garcia-Agundez A, Dutz T, Goebel S. Adapting smartphone-based photoplethysmograpy to suboptimal scenarios. Physiol Meas 2017;38(2):219-232. [doi: 10.1088/1361-6579/aa51db] [Medline: 28099163]
76. Béres S, Hejjel L. The minimal sampling frequency of the photoplethysmogram for accurate pulse rate variability parameters in healthy volunteers. Biomedical Signal Processing and Control 2021;68:102589. [doi: 10.1016/j.bspc.2021.102589]
77. GitHub. URL: https://github.com/8n98324n/pwa_mental [accessed 2026-02-17]

Abbreviations

BBI: beat-to-beat interval
CT: crest time
GAD-7: Generalized Anxiety Disorder-7
HR: heart rate
HRV: heart rate variability
PANAS-N: negative affect
PANAS-P: positive affect
PANAS: positive and negative affect schedule
PHQ-9: Patient Health Questionnaire-9
RF: random forest
SAM: Self-Assessment Manikin
SVS: Subjective Vitality Scale
SWLS: Satisfaction With Life Scale

Edited by L Buis; submitted 26.Jul.2025; peer-reviewed by V Vojdanovski, C Wang, JR Cox; comments to author 25.Nov.2025; revised version received 01.Feb.2026; accepted 02.Feb.2026; published 16.Mar.2026.

Please cite as:
Liu I, Hu L, Luo J, Liu C, Zhong Q, Ni S
Assessing Mental Health and Emotional States by Using Smartphone Photoplethysmography–Based Digital Pulse Waveform Analysis: Cross-Sectional Observational Study
JMIR Mhealth Uhealth 2026;14:e81301
URL: https://mhealth.jmir.org/2026/1/e81301
doi:10.2196/81301
PMID:41629260

©Ivan Liu, Luming Hu, Jing Luo, Chang Liu, Qi Zhong, Shiguang Ni. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 16.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Reducing Sedentary Time After Knee Replacement Using a Multicomponent mHealth Intervention: Randomized Controlled Trial

Christine A Pellegrini[1], PhD; Clare L Kennerley[1], PhD; Sara Wilcox[1,2], PhD; Jungwha Lee[3], PhD; Katherine DeVivo[4], PhD; Kailyn Horn[1], PT, DPT; Scott Jamieson[1], MS; Jeffrey Hopkins[5,6], MD; Harley T Davis[5], PhD; J Benjamin Jackson III[5,6], MD, MBA

[1]Department of Exercise Science, Arnold School of Public Health, University of South Carolina, Columbia, SC, United States

[2]Prevention Research Center, University of South Carolina, Columbia, SC, United States

[3]Department of Preventive Medicine, Northwestern University, Chicago, IL, United States

[4]School of Health Sciences, Columbia College, Columbia, SC, United States

[5]Prisma Health, Columbia, SC, United States

[6]School of Medicine Columbia, University of South Carolina, Columbia, SC, United States

Corresponding Author:
Christine A Pellegrini, PhD
Department of Exercise Science
Arnold School of Public Health
University of South Carolina
915 Greene Street
Columbia, SC, 29208
United States
Phone: 1 803 777 0911
Email: pellegca@mailbox.sc.edu

Abstract

Background: Total knee replacement (TKR) is a common surgery for end-stage knee osteoarthritis. Although reductions in pain and improvements in mobility occur after surgery, physical activity levels often do not change. Given the challenges of increasing physical activity in this population, targeting reductions in sedentary behavior may be a first step; however, no prior studies have examined the feasibility and effects of a sedentary reduction intervention after TKR.

Objective: This study examined the effects of a 2-month multicomponent mobile health sedentary reduction intervention (*NEAT!2*) on sedentary time in adults with TKR.

Methods: Adults (N=83; mean age 65.3, SD 9.4 years; mean BMI 32.7, SD 6.9 kg/m^2; 62/83, 74.7% female; 64/83, 77.1% White) with a TKR ≤1 year ago were randomized to the *NEAT!2* group (n=42, 50.6%) or the attention-matched control group (n=41, 49.4%). The *NEAT!2* intervention focused on reducing sedentary time via a smartphone app designed to interrupt prolonged bouts (≥30 minutes) of sedentary behavior and through 5 coaching calls emphasizing goal setting and problem solving. The control group focused on surgery recovery via an app or website and 5 educational calls. Sedentary time, total physical activity, physical function, and pain were measured at 2 and 5 months. Linear mixed-effects models examined intervention effects and differences between groups at each time point.

Results: Retention was 96% and 95% at 2 and 5 months, respectively, with no differences between groups. Participants in the *NEAT!2* group completed an average of 4.95 (SD 0.2) calls, used the app on an average of 40.3 (SD 13.8) days (out of 56 days), and received an average of 9.6 (SD 6.0) notifications per day. At 5 months, there was a significant increase in sit-to-stand transitions in the *NEAT!2* group and a substantial decrease in the control group, resulting in a significant difference between groups (mean difference 4.06, 95% CI 0.13-7.99; *P*=.04); however, the *NEAT!2* intervention did not result in significant effects on any of the other study outcomes at 2 or 5 months. Additionally, more days of app use were associated with greater increases in moderate-to-vigorous intensity physical activity (*r*=0.335; 95% CI 0.017-0.585; *P*=.04).

XSL•FO
RenderX

Conclusions: This study highlights the challenges of reducing sitting time in adults with TKR. Future studies should explore alternative behavior change techniques across different levels of influence (eg, environmental and social) to support interventions implemented within the first year after TKR.

Trial Registration: ClinicalTrials.gov NCT04482400; https://clinicaltrials.gov/ct2/show/NCT04482400

(JMIR Mhealth Uhealth 2026;14:e83148) doi:10.2196/83148

KEYWORDS

arthroplasty; sedentary behavior; mobile health; mHealth; physical activity

Introduction

Total knee replacement (TKR) is one of the most common surgical procedures performed in the United States, and its prevalence is expected to continue rising rapidly [1,2]. TKR is considered an effective treatment for severe knee osteoarthritis, with most patients experiencing reduced pain and improved physical function and quality of life following the procedure [3-5]. However, postoperatively, most adults struggle to increase physical activity [6], further increasing their risk of all-cause mortality, cardiovascular disease, and disability, with less than 5% of adults with TKR meeting physical activity guidelines [7]. Several interventions have been developed to target physical activity after TKR; however, most have had limited or modest success in increasing steps or moderate-intensity physical activity [8-10].

In addition to struggling to increase physical activity, adults with TKR continue to spend excessive time in sedentary behaviors [11,12]. Sedentary behavior is defined as any waking behavior in a seated or reclining position at ≤1.5 metabolic equivalents of task [13]. Among inactive adults, higher sedentary time is associated with an increased risk of all-cause mortality [14]. Given the low levels of physical activity after surgery, targeting reductions in sedentary behavior may be a more feasible first-step approach to increasing overall activity in this population. Sedentary reduction interventions tailored to other populations, such as office workers, have been shown to be effective [15,16], and a recent study demonstrated that multicomponent sedentary behavior interventions were well accepted and effective in older adults [17]; however, no prior studies have examined the effects of a sedentary reduction intervention after TKR.

The purpose of this study was to examine the effects of a multicomponent mobile health (mHealth) sedentary reduction intervention on sedentary time in adults with TKR at the end of the intervention (2 months) and following a 3-month maintenance period (5 months). Changes in physical activity, physical function, and pain were also examined at each time point. Additionally, this study examined the relationship between adherence to the mHealth program and changes in sedentary time, physical activity, physical function, and pain.

Methods

Study Design

This study was a randomized clinical trial, and full study details have been previously published [18]. Participants with TKR were randomized to either the mHealth sedentary reduction intervention (the *NEAT!2* group) or an attention control group. Outcomes were assessed at baseline, at the end of the intervention (2 months), and following a 3-month maintenance period.

Participants

Participants were recruited and enrolled between August 2020 and April 2024. Initially, participants were recruited at least 7 days prior to knee replacement and then started the program 4 weeks after surgery; however, due to recruitment challenges because of the COVID-19 pandemic, recruitment and eligibility criteria were modified in October 2021. Following the change, participants were eligible if they had a unilateral TKR within <1 year of the baseline assessment. Additional inclusion criteria included being aged 40 to 79 years, having a smartphone that was accessible and near them most of the day, being willing to download the study smartphone apps, self-reporting spending at least 7 hours a day sitting, and being English-speaking. Participants were excluded if they had any contraindications to activity, had a mobility-limiting comorbidity (eg, spinal stenosis), were scheduled for surgery (eg, TKR on the contralateral knee) within the next 6 months, or did not have ≥4 days of valid ActiGraph (ARCHIMED) accelerometer wear (≥10 hours/day) at baseline. There was no racial or gender bias in the selection of participants.

Participants were primarily recruited from orthopedic centers within a local health care system in Columbia, South Carolina. Study candidates were contacted via postcards, mailings, emails, or telephone calls. Additional recruitment strategies included flyers posted on the university campus or in local businesses or clinics, emails to university faculty and staff, and social media. All recruitment materials directed interested candidates to complete an online or telephone screening with study staff. Eligible candidates were invited to an in-person session to review complete study details and complete the written informed consent process if interested and eligible.

Randomization

After completing the baseline assessment and verifying accelerometer wear time, participants were contacted via telephone to remind them of the study details. Additionally, to promote retention and prevent differential attrition, the benefits and barriers to both general research study participation and specific study participation were reviewed, and participants were asked to discuss their personal perspectives on the pros and cons of being randomized to either study condition [19]. Participants who expressed continued interest in participating were randomized to one of two conditions: (1) *NEAT!2* or (2)

XSL•FO
RenderX

control. Randomization was created by the biostatistician, and participants were stratified by age (<65 years and ≥65 years). Participants were randomly assigned in a 1:1 ratio to either the *NEAT!2* group or the control group using permuted block sizes of 2 or 4.

Randomized Conditions

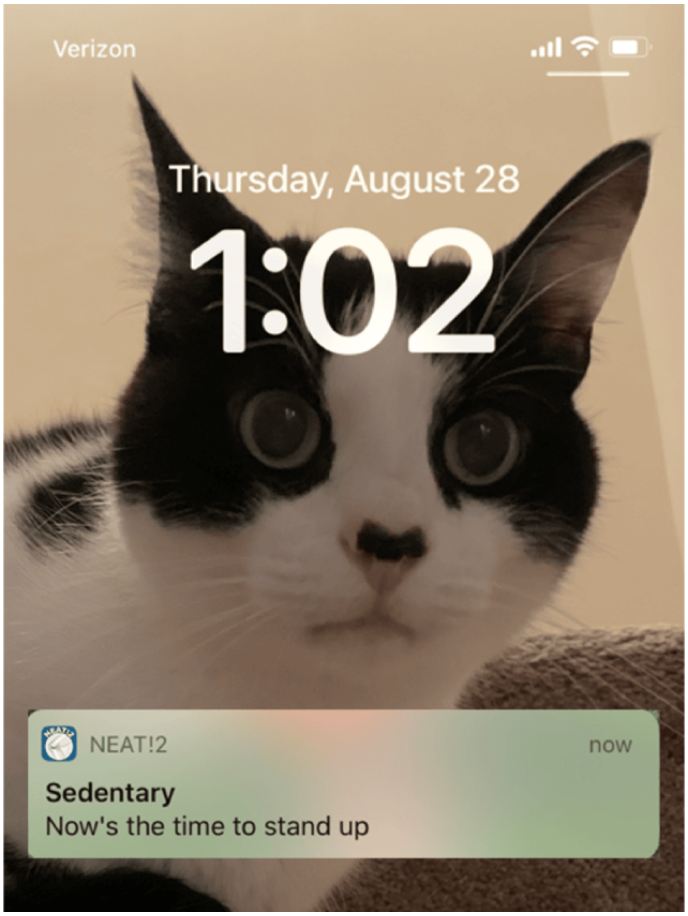
NEAT!2

Participants randomized to *NEAT!2* received the mHealth and coaching intervention guided by the Dual-Process Theory [20,21]. Participants were given goals to reduce sedentary time, starting with a reduction of 30 minutes per day and progressing every 2 weeks to a final goal of reducing sedentary time by 90 minutes per day. To help participants reach this goal, the *NEAT!2* app was turned on after randomization and displayed the overall daily goal. The app was designed to detect prolonged sedentary time using the accelerometer and activity recognition libraries on the participant's Android or iOS device. Specifically, when 30 minutes of continuous inactivity was detected, the app generated a sound or vibration and placed a notification on the participant's phone (Figure 1). The notifications varied in content, but each instructed participants to take a break from sitting, either by standing or walking for at least 2 minutes. The *NEAT!*2 prompts were designed to target automatic (implicit) processes.

In conjunction with the *NEAT!2* app, controlled (conscious) processes were targeted through goal setting, education, problem solving, and behavioral feedback. Participants received a coaching call every 2 weeks during the 2-month intervention (5 calls in total). Calls were intended to last 10 to 15 minutes and focused on app use, goal setting, and problem solving around reducing sedentary time, with coaches guided by semistructured call scripts. All coaches had a background in exercise science, psychology, public health, or a related field.

Between the 2- and 5-month assessments, participants received monthly phone calls from research staff members who did not conduct the coaching calls. These calls focused solely on surgery recovery and were designed to help with retention during the follow-up period. Participants were given the option to continue using the *NEAT!2* app during this period; however, use of the app or sedentary time was not discussed on calls.

Figure 1. Screenshot of an example NEAT!2 app notification.



Attention-Matched Control

Participants randomized to the control group received an attention-matched education program focused on surgery recovery, which included a smartphone app or website (My Knee Guide) and regular calls. Control participants were encouraged to use the app or website to plan recovery activities and read about others' experiences after TKR. Participants also received calls from the study coaches every 2 weeks (5 calls in total) to discuss surgery recovery and education after TKR (eg, dealing with fatigue, emotional self-care, and reducing injuries at home). If participants brought up unintended content (eg, sedentary behavior or physical activity), coaches were trained to redirect the conversation.

XSL•FO
RenderX

Similar to *NEAT!2* participants, control participants received monthly phone calls on surgery recovery between the 2- and 5-month assessments. Following the 5-month assessment, the control group was offered access to the *NEAT!2 app*.

Treatment Fidelity

For both groups, all coaching sessions were audiotaped, and a 15% sample of calls was rated every quarter. Fidelity assessments evaluated unintended and intended content based on randomized conditions. If fidelity fell below 80% or coaches discussed unintended content, coaches were retrained.

Measures

Assessments were completed either in an orthopedic clinic or a campus research center at baseline and at 2-month and 5-month follow-ups. Study assessors were blinded to group allocation and had undergone training on all assessment procedures. All questionnaires were completed online via REDCap (Research Electronic Data Capture) [22].

Outcomes

The primary outcome of the study was the percentage of time spent in sedentary behavior as assessed by ActiGraph at 2 months. Secondary outcomes included total physical activity time (light-, moderate-, and vigorous-intensity physical activity), physical function, and pain. Exploratory outcomes included sedentary time assessed by activPAL, knee symptoms, general health, sleep disturbance, and habit strength.

Sedentary Time and Physical Activity

At each assessment, participants were asked to wear an ActiGraph GT9X Link waist-worn accelerometer for 7 days during waking hours (except during water-based activities). A minimum of 4 days of valid wear time consisting of at least 10 hours/day was required for inclusion in analyses [23]. Nonwear time was defined as ≥90 minutes with 0 activity counts, allowing for up to 2 minutes of <100 counts/minute [24]. Sedentary time was defined as <100 counts/minute, and total activity as ≥100 counts/minute [23]. Average daily sedentary time (minutes/day), percentage of the waking day spent in sedentary time, and weekly total physical activity (light-, moderate-, and vigorous-intensity activity) were calculated within the ActiLife (version 6; Ametris) software.

In addition to ActiGraph monitors, the activPAL software (version 8; PAL Technologies) and Sedentary Time and Activity Reporting Questionnaire [25] were used as exploratory measures of sedentary time at each time point. The activPAL monitor was waterproofed and taped onto participants' thighs using waterproof tape. Participants were asked to wear the monitor for 24 hours a day during each 7-day period. A minimum of 4 days of valid wear time, defined as 24 hours, was required for inclusion in analyses. The time spent sitting or lying, standing, and walking; transitions; and step counts were determined using the activPAL. The Sedentary Time and Activity Reporting Questionnaire [25] was also administered and is a self-reported measure of habitual sedentary behavior. The average number of hours per day spent sitting on weekdays and weekends was calculated.

Physical Function

Physical function was assessed using the Chair Stand Test, Timed Up and Go test, and 6-minute walk test. All physical function tests were completed following the Osteoarthritis Research Society International–recommended procedures [26]. During the Chair Stand Test, participants were asked to complete as many chair stand repetitions as possible during a 30-second period. The Timed Up and Go test assessed the time, in seconds, taken to rise from a chair, walk 3 meters, turn, walk back to the chair, and sit down. The 6-minute walk test evaluated the maximal distance, in feet, that a participant could walk during a 6-minute period.

Patient-Reported Outcomes

Pain and related self-reported outcomes were assessed using the Western Ontario and McMaster Universities Osteoarthritis Index [27] and Knee Injury and Osteoarthritis Outcome Score [28]. The Western Ontario and McMaster Universities Osteoarthritis Index is a 15-item survey assessing pain, stiffness, and functional limitations over the last 48 hours on a 5-point Likert scale. Scores range from 0 to 20 for pain, 0 to 8 for stiffness, and 0 to 68 for physical function, with higher scores indicating more symptoms. The Knee Injury and Osteoarthritis Outcome Score assesses overall knee health and includes 5 subscales: pain, symptoms, activities of daily living, sport and recreation, and knee-related quality of life. Scores on each subscale range from 0 (no knee issues) to 100 (extreme knee issues). General health, sleep disturbance, and mobility were also assessed using Patient-Reported Outcomes Measurement Information System computer-adaptive tests via REDCap. Each survey generates a *t* score, where scores <50 indicate poorer outcomes, while scores >50 reflect better outcomes.

Habit strength was assessed with 4 items for sitting, stretching, and exercising from the Self-Report Habit Index [29]. Participants were asked to rate each behavior from 1 (strongly disagree) to 7 (strongly agree) on the following items: "I do automatically," "I do without having to consciously remember," "I do without thinking," and "I start doing before I realize I'm doing it." Higher scores indicate greater habit strength.

Demographics and Intervention Adherence

Demographics and medical history were obtained at baseline. Height (m) and weight (kg) were measured in light clothing and without shoes at each time point. BMI was subsequently calculated as kg/m^2.

Adherence to the *NEAT!2* intervention was assessed by the percentage of coaching calls completed (completed calls/5 total calls), the total days the *NEAT!2* app was used (56 possible days), and the response to *NEAT!2* notifications. Total days the *NEAT!2* app was used were determined as the number of unique days participants received notifications through the app. The response to *NEAT!2* notifications was determined by the percentage of notifications for which movement or activity was detected by the app within 5 minutes of notification delivery.

Adverse Events

Injuries and illnesses were assessed at each assessment, as well as during follow-up calls. All adverse events were monitored

XSL•FO
RenderX

and rated based on their severity, expectedness, and relatedness to the study.

Statistical Analysis

Descriptive statistics were used to describe participants' baseline characteristics. Mean (SD) for continuous variables and frequency (percentage) for categorical variables were reported. Linear mixed-effects models were used to evaluate whether the *NEAT!2* sedentary reduction intervention led to greater improvements in sedentary time (ActiGraph minutes/day and percentage of the day spent sedentary), physical activity, physical function, and pain than the control group at 2 and 5 months. Models included group (*NEAT!2* vs control), time (baseline, 2-month follow-up, and 5-month follow-up), and group×time interaction terms. Models accounted for repeated measures with a random intercept to allow for baseline differences. Estimated regression coefficients and corresponding 95% CIs were computed for each outcome. Potential confounders included age, sex, BMI, and comorbidities. ActiGraph total and moderate-to-vigorous intensity physical activity were adjusted for monitor wear time. To examine the association between adherence to the *NEAT!2* intervention and changes in outcomes (sedentary time, physical activity, physical function, and pain), bootstrap Spearman correlation coefficients and corresponding 95% Cis, accounting for the small sample size, were computed between adherence to the *NEAT!2* intervention and changes in outcomes. All hypothesis tests were 2-sided with statistical significance set at $P<.05$. No adjustments were made for multiple comparisons. All statistical analyses were performed in SAS (version 9.4; SAS Institute Inc).

The sample size calculations have been reported previously [18]. In brief, having 40 participants per group complete the 2-month and 5-month assessments would provide 90% power to detect a difference of 8.1% reduction in sedentary time (ie, effect size of 0.7) between participants in the *NEAT!2* and control groups. A reduction of 8% was chosen as it is approximately equivalent to a 90-minute reduction in sedentary time over the course of a waking day.

Ethical Considerations

All study procedures were approved by the University of South Carolina's Institutional Review Board (approval number Pro00092132) on September 19, 2019, and all procedures were conducted in accordance with the Helsinki Declaration of 1975. All participants provided written informed consent before participation. The trial was registered on Clinicaltrials.gov (NCT04482400). A CONSORT checklist is provided in Multimedia Appendix 1. Participants received US $15 for completing each follow-up assessment, and all research data were deidentified.

Results

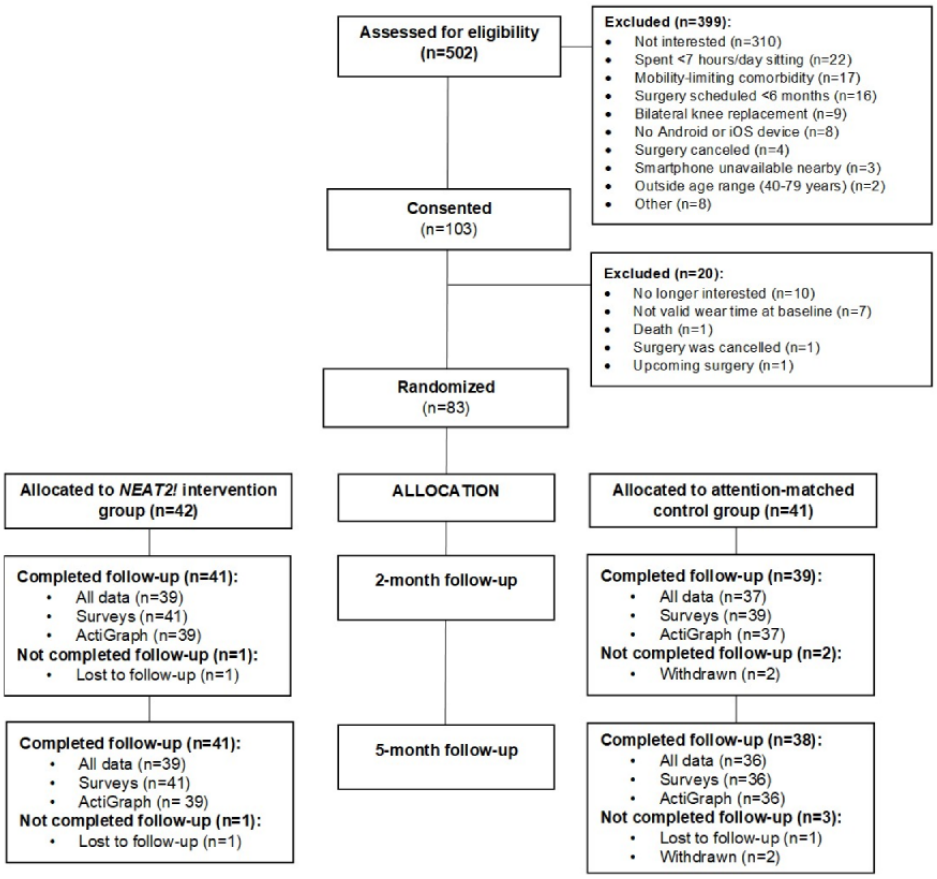
Participant Characteristics, Retention, and Safety

Of the 502 individuals screened for eligibility, 103 (20.5%) were eligible and provided consent (Figure 2). The primary reasons participants were excluded were lack of interest (n=310, 77.7%), reporting sitting <7 hours/day (n=22, 5.5%), or having a mobility-limiting comorbidity (n=17, 4.3%). Following consent, 83 (80.6%) participants were eligible and underwent randomization (n=42, 50.6% to the *NEAT!2* group and n=41, 49.4% to the control group). Retention was 96% and 95% at 2-month follow-up and 5-month follow-up, respectively, with no differences between groups at either time point.

Participants in the *NEAT!2* and control groups had similar characteristics at baseline (Table 1). Participants had a mean age of 65.3 (SD 9.4) years, were primarily female (n=62, 74.7%) and White (n=64, 77.1%), and had an average BMI of 32.7 (SD 6.9) kg/m^2. Most participants (n=44, 53%) were employed full-time or part-time. The most common medical conditions included hypertension (n=48, 57.8%), depression or anxiety (n=27, 32.5%), sleep apnea (n=22, 26.5%), and diabetes (n=18, 21.7%). In total, 9 (10.8%) participants were recruited prior to surgery (mean 12.0, SD 3.5 days before surgery), and 74 (83.2%) participants were recruited a median of 213 (IQR 87-300.5, range 38- 351) days after surgery. During the study period, 32 adverse events (2 prior to randomization, 8 in the control group, and 22 in the *NEAT!2* group) and 6 serious adverse events (3 in the control group and 1 in the *NEAT!2* group) occurred. The serious adverse events included 1 death and 5 surgeries or hospitalizations. Only 1 adverse event (nerve entanglement in the back) was deemed possibly related to the intervention; all others were deemed unrelated to the study.

XSL•FO
RenderX

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow diagram for the study.



XSL•FO
RenderX

Table 1. Baseline characteristics of the participants by full sample and by group allocation.

Characteristics	Total sample (n=83)	*NEAT!2* group (n=42)	Control group (n=41)
Age (years), mean (SD)	65.3 (9.4)	65.5 (9.8)	65.1 (9.0)
Sex, n (%)			
Female	62 (74.7)	32 (76.2)	30 (73.2)
Male	21 (25.3)	10 (23.8)	11 (26.8)
BMI (kg/m^2), mean (SD)	32.7 (6.9)	33.7 (7.1)	31.7 (6.6)
Days since knee surgery, mean (SD)	175.3 (121.8)	192.8 (122.6)	157.4 (119.8)
Race, n (%)			
Black or African American	19 (22.9)	11 (26.2)	8 (19.5)
White	64 (77.1)	31 (73.8)	33 (80.5)
Ethnicity: not Hispanic or Latinx, n (%)	83 (100)	42 (100)	41 (100)
Marital status, n (%)			
Married	51 (61.4)	30 (71.4)	21 (51.2)
Not married	32 (38.6)	12 (28.6)	20 (48.8)
Education, n (%)			
Below a college degree	26 (31.3)	11 (26.2)	15 (36.6)
College degree or above	57 (68.7)	31 (73.8)	26 (63.4)
Employment status, n (%)			
Full-time employment	33 (39.8)	17 (40.5)	16 (39.0)
Part-time employment	11 (13.2)	5 (11.9)	6 (14.6)
Retired or not working	39 (47.0)	20 (47.6)	19 (46.3)
Household income (US $), n (%)			
<100,000	41 (49.4)	22 (52.4)	19 (46.3)
≥100,000	33 (39.8)	16 (38.1)	17 (41.5)
Prefer not to answer	9 (10.8)	4 (9.5)	5 (12.2)
Comorbidities, n (%)			
Diabetes	18 (21.7)	9 (21.4)	9 (22.0)
Cancer	12 (14.4)	8 (19.0)	4 (9.8)
Depression and/or anxiety	27 (32.5)	15 (35.7)	12 (29.3)
High blood pressure	48 (57.8)	24 (57.1)	24 (58.5)
Thyroid problems	17 (20.5)	10 (23.8)	7 (17.1)
Sleep apnea	22 (26.5)	12 (28.6)	10 (24.4)

Intervention Effects on Primary and Secondary Outcomes

Valid ActiGraph data were provided by 82 (98.8%) participants at baseline (mean 6.6, SD 0.8 valid days), 76 (91.6%) participants at 2 months (mean 6.4, SD 0.9 valid days), and 75 (90.4%) participants at 5 months (mean 6.4, SD 0.9 valid days). Valid activPAL data were provided by 81 (97.6%) participants at baseline (mean 6.5, SD 0.7 valid days), 74 (89.2%) participants at 2 months (mean 6.2, SD 0.6 valid days), and 74 (89.2%) participants at 5 months (mean 6.2, SD 0.8 valid days). The intervention effects on sedentary behavior and physical activity outcomes are shown in Table 2. At baseline, the average percentage of the waking day spent in sedentary time was 73.26% (SD 7.47%) in the *NEAT!2* group and 74.13% (SD 6.17%) in the control group. At 5 months, ActiGraph-assessed moderate-to-vigorous physical activity (MVPA; mean difference −24.23, 95% CI −46.91 to −1.55; adjusted for weekly wear time; P=.04) and activPAL sit-to-stand transitions (mean difference 4.06, 95% CI 0.13-7.99; P=.04) were different between groups. The *NEAT!2* intervention did not result in statistically significant effects on any other sedentary behavior or physical activity outcomes at 2 or 5 months, and no other differences were observed between groups. The intervention effects on physical function and patient-reported outcomes are shown in Table 3. Exploratory outcomes are shown in

XSL•FO
RenderX

Multimedia Appendix 2. There were also no significant effects across time or by group.

XSL•FO
RenderX

Table 2. Sedentary behavior and physical activity outcomes at baseline, at 2-month follow-up, and at 5-month follow-up by group allocation.

Outcomes and time points	*NEAT!2* group (n=42), mean (SE)	Control group (n=41), mean (SE)	Group×time interaction effects in mixed-effects models	
			β coefficient (95% CI)	*P* value
ActiGraph				
Percentage of time sedentary (% per day)				
Baseline	73.26 (1.06)	74.05 (1.07)	—[a]	—
2-month follow-up	73.55 (1.08)	74.36 (1.09)	−0.02 (−2.36 to 2.33)	.99
5-month follow-up	74.73 (1.08)	73.74 (1.10)	1.78 (−0.57 to 4.14)	.14
Sedentary time (minutes per day[b])				
Baseline	658.77 (12.30)	659.41 (12.06)	—	—
2-month follow-up	644.18 (12.61)	658.90 (12.37)	−0.76 (−21.28 to 19.75)	.94
5-month follow-up	642.90 (12.71)	673.56 (12.37)	15.39 (−5.23 to 36.01)	.14
Total physical activity (light-, moderate-, or vigorous-intensity; minutes per week[b])				
Baseline	1642.76 (72.66)	1471.93 (73.93)	—	—
2-month follow-up	1530.88 (74.07)	1390.24 (75.37)	−14.86 (−141.59 to 111.86)	.82
5-month follow-up	1510.26 (74.07)	1424.61 (75.83)	−96.25 (−223.63 to 31.12)	.14
Moderate-to-vigorous physical activity (minutes per week[b])				
Baseline	78.67 (86.06)	51.08 (71.06)	—	—
2-month follow-up	58.51 (74.62)	49.76 (60.68)	−18.51 (−41.07 to 4.06)	.11
5-month follow-up	60.67 (89.10)	57.81 (65.74)	−24.23 (−46.91 to −1.55)	.04
Wear time (minutes per day)				
Baseline	900.42 (64.65)	888.42 (81.99)	—	—
2-month follow-up	972.74 (193.88)	923.29 (209.03)	18.52 (−12.32 to 49.36)	.88
5-month follow-up	944.02 (120.31)	924.37 (228.56)	18.62 (−12.37 to 49.61)	.73
activPAL				
Sitting time (minutes per day)				
Baseline	626.20 (17.95)	633.12 (18.17)	—	—
2-month follow-up	630.46 (18.22)	611.82 (18.79)	25.55 (−16.71 to 67.81)	.23
5-month follow-up	623.79 (18.45)	622.74 (18.54)	7.96 (−34.25 to 50.18)	.71
Standing time (minutes per day)				
Baseline	213.13 (11.82)	191.05 (11.96)	—	—
2-month follow-up	219.87 (11.96)	196.72 (12.29)	1.08 (−23.41 to 25.57)	.93
5-month follow-up	221.51 (12.08)	207.17 (12.16)	−7.73 (−32.19 to 16.74)	.53
Walking time (minutes per day)				
Baseline	79.52 (4.52)	73.34 (4.58)	—	—
2-month follow-up	78.81 (4.57)	75.01 (4.68)	−2.38 (−10.99 to 6.23)	.58
5-month follow-up	78.41 (4.61)	76.18 (4.64)	−3.95 (−12.55 to 4.66)	.37
Sit-to-stand transitions (transitions per day)				
Baseline	45.82 (1.86)	47.21 (1.89)	—	—
2-month follow-up	48.38 (1.89)	46.46 (1.94)	3.31 (−0.62 to 7.24)	.10
5-month follow-up	47.66 (1.91)	45.00 (1.92)	4.06 (0.13 to 7.99)	.04
Steps (steps per day)				

XSL•FO
RenderX

Outcomes and time points	*NEAT!2* group (n=42), mean (SE)	Control group (n=41), mean (SE)	Group×time interaction effects in mixed-effects models	
			β coefficient (95% CI)	*P* value
Baseline	5959.90 (368.33)	5325.10 (372.83)	—	—
2-month follow-up	5920.49 (372.12)	5413.79 (381.33)	−128.10 (−823.13 to 566.93)	.72
5-month follow-up	5956.20 (375.19)	5615.94 (377.97)	−294.54 (−988.81 to 399.72)	.40

[a]Not applicable.

[b]Adjusted for either daily or weekly wear time.

Table 3. Physical function and patient-reported outcomes at baseline, at 2-month follow-up, and at 5-month follow-up by group allocation.

Outcomes and time points	*NEAT!2* group (n=42), mean (SE)	Control group (n=41), mean (SE)	Group×time interaction effects in mixed-effects models	
			β coefficient (95% CI)	*P* value
6-minute walk test (feet)				
Baseline	1231.36 (46.69)	1207.42 (47.25)	__[a]	—
2-month follow-up	1328.44 (46.84)	1330.05 (47.85)	−7.79 (−29.70 to 14.12)	.48
5-month follow-up	1342.40 (46.84)	1342.89 (48.15)	−7.45 (−29.61 to 14.70)	.51
30-second Chair Stand Test (repetitions)				
Baseline	9.48 (0.43)	9.34 (0.43)	—	—
2-month follow-up	10.85 (0.43)	10.39 (0.44)	0.33 (−0.46 to 1.12)	.41
5-month follow-up	10.54 (0.43)	10.40 (0.44)	0.00 (−0.79 to 0.80)	.996
Timed Up and Go test (seconds)				
Baseline	10.28 (0.37)	10.20 (0.38)	—	—
2-month follow-up	9.54 (0.38)	9.45 (0.38)	0.01 (−0.74 to 0.77)	.97
5-month follow-up	9.25 (0.38)	9.54 (0.39)	−0.37 (−1.13 to 0.39)	.34
WOMAC[b] pain score				
Baseline	5.21 (0.53)	5.37 (0.54)	—	—
2-month follow-up	2.98 (0.53)	3.67 (0.54)	−0.54 (−1.63 to 0.55)	.33
5-month follow-up	3.10 (0.53)	3.03 (0.55)	0.23 (−0.87 to 1.32)	.69
WOMAC stiffness score				
Baseline	3.07 (0.27)	3.15 (0.27)	—	—
2-month follow-up	2.55 (0.27)	2.17 (0.28)	0.45 (−0.23 to 1.14)	.19
5-month follow-up	2.28 (0.27)	1.95 (0.28)	0.41 (−0.28 to 1.10)	.24
WOMAC function score				
Baseline	17.95 (1.64)	16.68 (1.66)	—	—
2-month follow-up	11.96 (1.66)	12.45 (1.68)	−1.76 (−5.59 to 2.08)	.37
5-month follow-up	11.33 (1.66)	11.23 (1.70)	−1.18 (−5.05 to 2.70)	.55

[a]Not applicable.

[b]WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index (higher scores on WOMAC indicate more knee symptoms).

A sensitivity analysis was performed including covariates such as BMI, marital status, cancer, depression, and the interaction between baseline physical activity and BMI. Effect estimates were similar across all covariate specifications. Compared with the primary outcome of ActiGraph percentage of time sedentary (group × time interaction effect −0.02, 95% CI −2.36 to 2.33 at 2M; 1.78, 95% CI −0.57 to 4.14 at 5M), estimates from the model adjusted for BMI, marital status, cancer, and depression (0.01, 95% CI −2.34 to 2.35 at 2M; 1.82, 95% CI −0.54 to 4.17 at 5M) and the model accounting for interaction of baseline low physical activity and BMI (0.28, 95% CI −2.01 to 2.57 at 2M; 2.18, 95% CI −0.12 to 4.49 at 5M) did not change conclusions. After accounting for covariates, the *NEAT!2* intervention did not result in statistically significant improvement on any of the

XSL•FO
RenderX

other sedentary behavior or physical activity outcomes at 2 or 5 months.

Intervention Engagement

On average, participants completed an average of 4.8 (SD 0.6) calls of the possible 5 calls, with no difference between randomized groups. Average call times for *NEAT!2* participants (mean 13.0, SD 7.2 minutes) were greater than those for participants in the control group (mean 10.65, SD 4.8 minutes; $P<.001$). Participants in the *NEAT!2* condition used the app for an average of 40.3 (SD 3.82) days over the 56-day intervention, with participants receiving an average of 9.7 (SD 6.0) notifications per day. Participants responded to *NEAT!2* notifications (movement detected by the smartphone) on an average of 9.7% (SD 7.8%; range 2%-42%) of occasions, with an average time between notification and movement of 2.4 (0.4) minutes.

Association Between Intervention Adherence and Changes in Outcomes

A higher percentage of days of *NEAT!2* app use was associated with greater increases in MVPA ($r=0.335$; 95% CI 0.017-0.585; $P=.04$). There were no other significant relationships between adherence to call completion or response to *NEAT!2* notifications and changes in sedentary time, physical activity, physical function, and pain.

Discussion

This study examined the effects of a multicomponent mHealth sedentary reduction intervention (*NEAT!2*) on sedentary time in adults with recent TKR. Additionally, this study examined the dose-response relationship between adherence to the mHealth program and changes in each outcome. Despite high adherence to the program, the *NEAT!2* intervention only led to improvements in sit-to-stand transitions assessed by activPAL at 5 months. The intervention did not result in any other statistically significant changes in the study outcomes of sedentary time, physical activity, physical function, or pain. Results did not change even after adjusting for covariates. Although there were no between-group differences in outcomes, there was a positive association between adherence to the *NEAT!2* app and MVPA, suggesting that those who used the app more frequently had greater increases in MVPA.

Although some previous interventions have been effective in reducing sedentary time in non-TKR populations [15-17], the literature is mixed, particularly among older adults [30]. A recent systematic review by Chastin et al [30] highlighted the uncertainties regarding whether sedentary reduction interventions can effectively reduce sedentary time in older adults; however, the review only included 7 studies. The lack of changes in sedentary time in this *NEAT!2* mHealth intervention among adults <1 year after TKR is consistent with this review [30], highlighting the challenges of reducing sitting time in an older clinical population, as well as the need for more studies examining reductions in sedentary behavior in older adults.

This intervention was guided by the Dual-Process Theory [20,21] and targeted both automatic and conscious individual-level behaviors. The behavior change techniques implemented in this study included strategies, such as reminders to interrupt sedentary behavior (targeting automatic processes) and goal setting (targeting conscious processes), to target total sedentary time. Previously, mHealth interventions, including disruptive smartphone apps designed to interrupt sedentary behavior, have shown effectiveness in reducing total sedentary time among various populations, including individuals with overweight or obesity and diabetes [15,31]; however, these strategies may not have been sufficient to change daily sitting time in adults after TKR within this brief 2-month intervention. The intervention increased sit-to-stand transitions at 5 months, but to change total sedentary time, providing real-time feedback may be necessary, as this approach has been an effective behavior change technique for physical activity [32]. Unfortunately, there were no devices at the time, to our knowledge, that could accurately provide feedback on sitting time. Furthermore, many effective sedentary reduction programs have included environmental changes and the provision of standing desks, primarily for office workers [33]. Although this approach may work for working adults, approximately half of the participants in this study were retired or not working; thus, standing desks may not be an effective strategy for this population. Environmental strategies for retired adults could include modifying the home by adding standing-friendly spaces (eg, high tables or counters) for activities such as puzzles, rearranging furniture and living space to encourage standing and light movement, or incorporating visual cues or reminders (eg, notes near the television) to break up sitting time. Future interventions may need to increase the focus on environmental, as well as cultural and social factors that could be influencing sedentary behavior more than individual-level factors [30,34].

Potential postoperative and COVID-19 pandemic–related barriers faced by the current participants may have also reinforced sedentary behaviors [35] and prevented the intervention from being effective. Additionally, previous research has suggested that adults with TKR and osteoarthritis are unaware of the health benefits of light-intensity activity and often think that sedentary time should be replaced with moderate-intensity activity [34]. This intervention primarily promoted light-intensity activity such as standing or light walking; however, the results suggested that the more days the app was used, the greater the increase in MVPA. Despite coaches encouraging the replacement of sedentary time with light-intensity activity, participants may still have focused on replacing sedentary time with MVPA such as through brisk walk. Although increasing MVPA would provide substantial health benefits, it is unreasonable to think that all excess sedentary time could be replaced with MVPA. More effort is needed to raise awareness of the health consequences of excess sedentary time and the benefits of replacing sedentary time with activity of any intensity in older adults [36].

This study had several strengths. In addition to being one of the first studies to examine a sedentary reduction program after TKR, the use of multiple objective and subjective measurement tools provides further evidence of the challenges and

XSL•FO
RenderX

inconsistencies in measuring sedentary time in this population. However, this study has also had several limitations. The COVID-19 pandemic resulted in necessary changes to recruitment whereby some participants completed baseline testing before surgery, and others completed testing postoperatively. Although this was accounted for in the analyses, the COVID-19 pandemic may have had other effects on participants and outcomes in both the intervention and control groups. Additionally, the prevalence of depression and/or anxiety (32.5%) was higher than anticipated and may have affected adherence to the intervention. The sample size was smaller than anticipated, resulting in lower power to detect changes and there was potential for inter- and intraindividual variation in capturing habitual sedentary and physical activity behaviors, which may have influenced the results. Furthermore, there is likely an error in the determination of responsiveness to the *NEAT!2* notifications, as the determination of response was based on movement detected on the smartphone. It is possible that participants may have stood up or walked around but did not move their phone; thus, the app would not have been able to detect that response. Additionally, the ActiGraph activity monitor cannot distinguish sitting from standing. Future studies may need to consider using more innovative wearables that can detect postural changes.

A multicomponent mHealth sedentary reduction intervention led to changes at 5M in sit-to-stand transitions but was not effective in reducing sedentary time in adults who had undergone TKR; however, participants who used the app more frequently throughout the intervention had greater increases in MVPA. This study highlights the challenges of reducing sitting time in this population and presents direction for future studies, which should explore alternative behavior change techniques across different levels of influence (eg, environmental and social) within interventions during the first year after TKR.

Funding

Research reported in this publication was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health (grant R21AR074780). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

JBJ is a consultant for Vilex Inc. All other authors have no conflicts of interest to disclose.

Multimedia Appendix 1
CONSORT (Consolidated Standards of Reporting Trials) checklist.
[PDF File (Adobe PDF File), 98 KB - mhealth_v14i1e83148_app1.pdf]

Multimedia Appendix 2
Exploratory outcomes by randomized condition at 2 and 5 months.
[DOCX File , 19 KB - mhealth_v14i1e83148_app2.docx]

References

1. Singh JA, Yu S, Chen L, Cleveland JD. Rates of total joint replacement in the United States: future projections to 2020-2040 using the national inpatient sample. J Rheumatol 2019 Sep;46(9):1134-1140. [doi: 10.3899/jrheum.170990] [Medline: 30988126]
2. Shichman I, Roof M, Askew N, Nherera L, Rozell JC, Seyler TM, et al. Projections and epidemiology of primary hip and knee arthroplasty in Medicare patients to 2040-2060. JB JS Open Access 2023 Feb 28;8(1):e22 [FREE Full text] [doi: 10.2106/JBJS.OA.22.00112] [Medline: 36864906]
3. Woodland N, Takla A, Estee MM, Franks A, Bhurani M, Liew S, et al. Patient-reported outcomes following total knee replacement in patients aged 65 years and over-a systematic review. J Clin Med 2023 Feb 17;12(4):1613 [FREE Full text] [doi: 10.3390/jcm12041613] [Medline: 36836148]
4. Neuprez A, Neuprez AH, Kaux JF, Kurth W, Daniel C, Thirion T, et al. Total joint replacement improves pain, functional quality of life, and health utilities in patients with late-stage knee and hip osteoarthritis for up to 5 years. Clin Rheumatol 2020 Mar;39(3):861-871 [FREE Full text] [doi: 10.1007/s10067-019-04811-y] [Medline: 31720892]
5. Vina ER, Hannon MJ, Kwoh CK. Improvement following total knee replacement surgery: exploring preoperative symptoms and change in preoperative symptoms. Semin Arthritis Rheum 2016 Apr;45(5):547-555 [FREE Full text] [doi: 10.1016/j.semarthrit.2015.10.002] [Medline: 26573377]

XSL•FO
RenderX

6. Hammett T, Simonian A, Austin M, Butler R, Allen KD, Ledbetter L, et al. Changes in physical activity after total hip or knee arthroplasty: a systematic review and meta-analysis of six- and twelve-month outcomes. Arthritis Care Res (Hoboken) 2018 Jun;70(6):892-901 [FREE Full text] [doi: 10.1002/acr.23415] [Medline: 28898559]
7. Kahn TL, Schwarzkopf R. Does total knee arthroplasty affect physical activity levels? Data from the osteoarthritis initiative. J Arthroplasty 2015 Sep;30(9):1521-1525. [doi: 10.1016/j.arth.2015.03.016] [Medline: 25882608]
8. Christiansen CL, Kline PW, Anderson CB, Melanson EL, Sullivan WJ, Richardson VL, et al. Optimizing total knee arthroplasty rehabilitation with telehealth physical activity behavior change intervention: a randomized clinical trial. Phys Ther 2024 Oct 02;104(10):pzae088. [doi: 10.1093/ptj/pzae088] [Medline: 39001713]
9. Losina E, Collins JE, Deshpande BR, Smith SR, Michl GL, Usiskin IM, et al. Financial incentives and health coaching to improve physical activity following total knee replacement: a randomized controlled trial. Arthritis Care Res (Hoboken) 2018 May;70(5):732-740 [FREE Full text] [doi: 10.1002/acr.23324] [Medline: 28732147]
10. White DK, Jakiela J, Liles S, Master H, Voinier D, Brunette M, et al. Effectiveness of a physical therapist-administered physical activity intervention after total knee replacement: a randomized trial. Phys Ther 2026 Jan 02;106(1):pzaf140. [doi: 10.1093/ptj/pzaf140] [Medline: 41347910]
11. Frimpong E, McVeigh JA, Meiring RM. Sedentary behavior in patients with knee osteoarthritis before and after total knee arthroplasty: a systematic review. J Aging Phys Act 2018 Oct 01;26(4):671-681. [doi: 10.1123/japa.2017-0214] [Medline: 29252080]
12. Webber SC, Strachan SM, Pachu NS. Sedentary behavior, cadence, and physical activity outcomes after knee arthroplasty. Med Sci Sports Exerc 2017 Jun;49(6):1057-1065. [doi: 10.1249/MSS.0000000000001207] [Medline: 28099297]
13. Tremblay MS, Aubert S, Barnes JD, Saunders TJ, Carson V, Latimer-Cheung AE, et al. Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project process and outcome. Int J Behav Nutr Phys Act 2017 Jun 10;14(1):75 [FREE Full text] [doi: 10.1186/s12966-017-0525-8] [Medline: 28599680]
14. Ekelund U, Tarp J, Fagerland MW, Johannessen JS, Hansen BH, Jefferis BJ, et al. Joint associations of accelero-meter measured physical activity and sedentary time with all-cause mortality: a harmonised meta-analysis in more than 44 000 middle-aged and older individuals. Br J Sports Med 2020 Dec;54(24):1499-1506 [FREE Full text] [doi: 10.1136/bjsports-2020-103270] [Medline: 33239356]
15. Compernolle S, DeSmet A, Poppe L, Crombez G, De Bourdeaudhuij I, Cardon G, et al. Effectiveness of interventions using self-monitoring to reduce sedentary behavior in adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2019 Aug 13;16(1):63 [FREE Full text] [doi: 10.1186/s12966-019-0824-3] [Medline: 31409357]
16. Shrestha N, Kukkonen-Harjula KT, Verbeek JH, Ijaz S, Hermans V, Pedisic Z. Workplace interventions for reducing sitting at work. Cochrane Database Syst Rev 2018 Dec 17;12(12):CD010912 [FREE Full text] [doi: 10.1002/14651858.CD010912.pub5] [Medline: 30556590]
17. Rosenberg DE, Zhu W, Greenwood-Hickman MA, Cook AJ, Florez Acevedo S, McClure JB, et al. Sitting time reduction and blood pressure in older adults: a randomized clinical trial. JAMA Netw Open 2024 Mar 04;7(3):e243234 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.3234] [Medline: 38536177]
18. Pellegrini CA, Lee J, DeVivo KE, Harpine CE, Del Gaizo DJ, Wilcox S. Reducing sedentary time using an innovative mHealth intervention among patients with total knee replacement: rationale and study protocol. Contemp Clin Trials Commun 2021 Jun 18;22:100810 [FREE Full text] [doi: 10.1016/j.conctc.2021.100810] [Medline: 34195473]
19. Jake-Schoffman DE, Brown SD, Baiocchi M, Bibeau JL, Daubenmier J, Ferrara A, et al. Methods-motivational interviewing approach for enhanced retention and attendance. Am J Prev Med 2021 Oct;61(4):606-617 [FREE Full text] [doi: 10.1016/j.amepre.2021.04.005] [Medline: 34544560]
20. Groves PM, Thompson RF. Habituation: a dual-process theory. Psychol Rev 1970 Sep;77(5):419-450. [doi: 10.1037/h0029810] [Medline: 4319167]
21. Conroy DE, Maher JP, Elavsky S, Hyde AL, Doerksen SE. Sedentary behavior as a daily process regulated by habits and intentions. Health Psychol 2013 Nov;32(11):1149-1157 [FREE Full text] [doi: 10.1037/a0031629] [Medline: 23477579]
22. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
23. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
24. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. Med Sci Sports Exerc 2011 Feb;43(2):357-364 [FREE Full text] [doi: 10.1249/MSS.0b013e3181ed61a3] [Medline: 20581716]
25. Lynch BM, Friedenreich CM, Khandwala F, Liu A, Nicholas J, Csizmadi I. Development and testing of a past year measure of sedentary behavior: the SIT-Q. BMC Public Health 2014 Sep 01;14:899 [FREE Full text] [doi: 10.1186/1471-2458-14-899] [Medline: 25179551]
26. Dobson F, Hinman RS, Roos EM, Abbott JH, Stratford P, Davis AM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. Osteoarthritis Cartilage 2013 Aug;21(8):1042-1052 [FREE Full text] [doi: 10.1016/j.joca.2013.05.002] [Medline: 23680877]

XSL•FO
RenderX

27. McConnell S, Kolopack P, Davis AM. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. Arthritis Rheum 2001 Oct;45(5):453-461. [doi: 10.1002/1529-0131(200110)45:5<453::aid-art365>3.0.co;2-w] [Medline: 11642645]
28. Roos EM, Lohmander LS. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. Health Qual Life Outcomes 2003 Nov 03;1:64 [FREE Full text] [doi: 10.1186/1477-7525-1-64] [Medline: 14613558]
29. Verplanken B, Orbell S. Reflections on past behavior: a self-report index of habit strength. J Appl Soc Psychol 2006 Jul 31;33(6):1313-1330. [doi: 10.1111/j.1559-1816.2003.tb01951.x]
30. Chastin S, Gardiner PA, Harvey JA, Leask CF, Jerez-Roig J, Rosenberg D, et al. Interventions for reducing sedentary behaviour in community-dwelling older adults. Cochrane Database Syst Rev 2021 Jun 25;6(6):CD012784 [FREE Full text] [doi: 10.1002/14651858.CD012784.pub2] [Medline: 34169503]
31. Ueno DT, Guerra PH, Christofoletti AE, Bonolo A, Nakamura PM, Kokubun E. Mobile health apps to reduce sedentary behavior: a scoping review. Health Promot Int 2022 Apr 29;37(2):daab124. [doi: 10.1093/heapro/daab124] [Medline: 34392354]
32. Ferguson T, Olds T, Curtis R, Blake H, Crozier AJ, Dankiw K, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022 Aug;4(8):e615-e626 [FREE Full text] [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
33. Silva H, Ramos PG, Teno SC, Júdice PB. The impact of sit-stand desks on full-day and work-based sedentary behavior of office workers: a systematic review. Hum Factors 2025 Jul;67(7):695-713. [doi: 10.1177/00187208241305591] [Medline: 39626101]
34. Webber SC, Ripat JD, Pachu NS, Strachan SM. Exploring physical activity and sedentary behaviour: perspectives of individuals with osteoarthritis and knee arthroplasty. Disabil Rehabil 2020 Jul;42(14):1971-1978. [doi: 10.1080/09638288.2018.1543463] [Medline: 30656981]
35. Pellegrini CA, Ledford G, Chang RW, Cameron KA. Understanding barriers and facilitators to healthy eating and physical activity from patients either before and after knee arthroplasty. Disabil Rehabil 2018 Aug;40(17):2004-2010 [FREE Full text] [doi: 10.1080/09638288.2017.1323026] [Medline: 28475433]
36. Mankowski RT, Anton SD, Axtell R, Chen SH, Fielding RA, Glynn NW, et al. Device-measured physical activity as a predictor of disability in mobility-limited older adults. J Am Geriatr Soc 2017 Oct;65(10):2251-2256 [FREE Full text] [doi: 10.1111/jgs.15037] [Medline: 28799216]

Abbreviations

mHealth: mobile health
MVPA: moderate-to-vigorous physical activity
REDCap: Research Electronic Data Capture
TKR: total knee replacement

Edited by J Zhou; submitted 29.Aug.2025; peer-reviewed by T Sjoros, D Rosenberg; comments to author 06.Mar.2026; revised version received 17.Apr.2026; accepted 30.Apr.2026; published 26.May.2026.

Please cite as:
Pellegrini CA, Kennerley CL, Wilcox S, Lee J, DeVivo K, Horn K, Jamieson S, Hopkins J, Davis HT, Jackson III JB
Reducing Sedentary Time After Knee Replacement Using a Multicomponent mHealth Intervention: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e83148
URL: https://mhealth.jmir.org/2026/1/e83148
doi:10.2196/83148
PMID:

©Christine A Pellegrini, Clare L Kennerley, Sara Wilcox, Jungwha Lee, Katherine DeVivo, Kailyn Horn, Scott Jamieson, Jeffrey Hopkins, Harley T Davis, J Benjamin Jackson III. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 26.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Automated Physical Activity Support for Adults and Youth From Low-Income Communities: Single-Arm Pilot Study

Jordan A Carlson[1], PhD; Frank Materia[2,3], PhD; Mallory Moon[1], MPH; Suryeon Ryu[1], PhD; Cory Yeager[1], MS; Chelsea Steel[1], MS, MPH; Kacee Shields[1], MBA; Harpreet Singh Gill[4], MBA; Jannette Berkley-Patton[5], PhD; Delwyn Catley[6], PhD

[1]Center for Children's Healthy Lifestyles & Nutrition, Children's Mercy Hospital, Kansas City, MO, United States

[2]Department of Otolaryngology-Head and Neck Surgery, University of Kansas Medical Center, Kansas City, KS, United States

[3]Department of Population Health, University of Kansas Medical Center, Kansas City, KS, United States

[4]Research Informatics, Children's Mercy Hospital, Kansas City, MO, United States

[5]Department of Biomedical and Health Informatics, School of Medicine, University of Missouri–Kansas City, Kansas City, MO, United States

[6]Department of Psychology, San Diego State University, San Diego, CA, United States

Corresponding Author:
Jordan A Carlson, PhD
Center for Children's Healthy Lifestyles & Nutrition
Children's Mercy Hospital
2401 Gillham Rd
Kansas City, MO, 64108
United States
Phone: 1 816 234 9240
Email: jacarlson@cmh.edu

Abstract

Background: Mobile health (mHealth) interventions are growing in popularity, but less research has focused on low-income families, particularly interventions integrating wearable devices with automated personalized messages.

Objective: We tested a preliminary wearable-integrated mHealth intervention with initial personalization elements among adults and youth from low-income urban communities, focusing on feasibility, acceptability, and preliminary evidence of physical activity behavior.

Methods: Participants were 83 adults and 31 youth recruited through community health events held in low-income urban communities. Using a single-arm pre-post design, participants were enrolled into a 7-week beta-version mHealth intervention that integrated a Garmin activity monitor with automated text messages. Messages were sent 4 days/week and focused on increasing step counts using theory-based behavior change techniques related to goal setting, self-monitoring, reinforcement, contextual factors, and self-efficacy. Most messages were personalized by including calculations based on the step-count and step-goal data, using branching logic, and using 2-way question-and-response messages. Feasibility measures included enrollment, retention, fidelity of message delivery, and adherence to wearing the Garmin device. Acceptability measures included survey items and engagement with responding to 2-way messages. Changes in daily steps were explored using mixed-effects linear regression.

Results: Enrollment and eligibility rates were 64% (84/132, adults) and 63% (31/49, youth), retention for physical activity measures was 84% (70/83) and 77% (24/31), and 99% (3910/3955) of the intended messages were delivered. Adults and youth adhered to wearing the Garmin on 82% (45/56) and 79% (44/56) of the study days, respectively. Overall acceptability ratings were 83% to 100%, with 97% (75/77) of adults and 100% (27/27) of youth indicating they would recommend the program. Adults and youth replied to a mean of 2.6 (SD 2.2) and 3.2 (SD 2.7) of the 7 text messages that asked for a reply, with higher engagement among adults who participated with their child. Pre-post changes in daily steps were β=240 (95% CI –387 to 866) for adults and β=413 (95% CI –877 to 1703) among youth, with larger changes observed among those in the highest tertile of engagement (adults: β=584, 95% CI –784 to 1952; n=19; youth: β=941, 95% CI –827 to 2709; n=11) and those who were meeting less than two-thirds of the physical activity guideline at baseline (adults: β=609, 95% CI –30 to 1247; n=47; youth: β=1406, 95% CI –94 to 2907; n=22).

Conclusions: Personalized mHealth physical activity interventions integrating wearable step trackers with automated text messaging appear to be feasible and acceptable among adults and youth from low-income communities. Step-count findings show

XSL•FO
RenderX

promise for the intervention's ability to support individuals who are further from meeting physical activity guidelines and warrant more research among parent–child dyads. Overall, findings support additional research to optimize and evaluate similar interventions within this population group using fully powered randomized controlled trials.

Trial Registration: ClinicalTrials.gov NCT05110508; https://clinicaltrials.gov/ct2/show/NCT05110508

(JMIR Mhealth Uhealth 2026;14:e76991) doi:10.2196/76991

KEYWORDS

African American; community engagement; medically underserved; text messaging; wearables

Introduction

Increasing physical activity is central to the prevention and control of many diseases and promotion of general wellness for adults and youth [1]. Physical activity interventions are especially important for low-income population groups, which can be underresourced and have been shown to engage in less leisure-time physical activity and experience higher rates of related chronic diseases than other groups [2-5]. While behavior change interventions have been effective for increasing physical activity in multiple population groups [6,7], participation can be difficult because of burdensome protocols, times constraints, and scheduling conflicts, particularly for those from marginalized backgrounds [8-11]. Mobile health (mHealth) approaches have helped mitigate some of these barriers by providing greater flexibility for participants, which can improve the reach of behavior change interventions [12]. mHealth can also improve the reach and scalability of interventions through automation, such as delivering content through automated messages and/or website- or app-based platforms. Mobile technologies are now highly prevalent among low-income groups, with more than 84% of adults in low-income households having a smartphone [13].

mHealth interventions have been effective for increasing physical activity in multiple population groups, but there has been less research in low-income groups [14-19]. The few prior mHealth studies addressing physical activity among low-income groups have shown promise for this approach [20-23]. However, there is a need for more mHealth research among these groups, especially studies that include both adults and youth. This family-inclusive approach could improve reach by allowing parents/caregivers and children to enroll together and may extend physical activity support within family networks. Another gap in prior work among low-income groups is that while interventions have used wearables and websites or apps and content is often tailored to the focus population, the extent to which the content is personalized or individualized for each participant has been limited in many studies [12,20-23]. Personalization better addresses the real-life circumstances and contexts people experience and can be accomplished in multiple ways, including customizing delivery timing, using participant information within content/messages, and using participant or environmental information to select and/or customize goals, feedback, prompts, or other content based on decision trees, algorithms, or even artificial intelligence [14,24-28]. Wearable physical activity trackers are particularly useful for personalizing content on physical activity goal monitoring, although there is a need for more research in low-income groups to investigate the integration of physical activity wearables with automated, personalized messaging using text messages [20-23].

This pilot study aimed to help address gaps in prior research by exploring a personalized, wearable-integrated, family-inclusive physical activity mHealth intervention among both adults and youth from low-income communities. The study used a beta-version intervention to collect initial information on the intervention approach to inform a more robust version of the intervention and evaluation in a future randomized clinical trial. The beta-version incorporated basic elements of personalization through integration of step-count and adaptive step-goal information into messages, use of branching logic to select messages based on step counts and goal achievement, 2-way question-and-response messaging, and use of participant names in messages. The evaluation focused on feasibility and acceptability (primary outcomes) while exploring preliminary evidence of physical activity behavior change.

Methods

Ethical Considerations

The study protocol was approved by Children's Mercy Hospital's Internal Review Board on August 24, 2021 (IRB study number STUDY00001939), and participants were compensated up to US $95 for completing study assessments. Prior to receiving the Garmin monitor, all participants provided written informed consent or child assent, and parents provided written parental permission. All data were deidentified prior to data storage and analyses. Study reporting was aligned with the Transparent Reporting of Evaluations with Nonrandomized Designs statement [29]. The only deviation from the original clinical trials registration was that step counts, rather than both step counts and physical activity minutes, were used as the physical activity outcome measure, as detailed below.

Participants and Procedures

The study used a single-arm pre-post design, with all participants receiving the same active intervention, termed ActiveKC. Adults and youth were included, with those aged 8 years or older being eligible. Participants were required to be able to communicate in English and could be an adult, child, or parent–child dyad enrolled together. Either the participant or parent was required to have access to a smartphone. A priori eligibility criteria included that participants be insufficiently active, defined as not meeting step-count–based physical activity guidelines [30-32], because of the intervention's focus on helping move individuals toward meeting these guidelines. Interested adults and youth wore a step tracker (Garmin Vívofit 4; Garmin

XSL•FO
RenderX

International, Inc) for a 7-day run-in period and were excluded from the study if they averaged ≥7500 steps/day (adults) or ≥12,000 steps/day (youth). Individuals were also ineligible if they did not wear the Garmin device for ≥3 days during the run-in period because of perceived difficulty in their ability to receive the intervention content as intended, thus impacting internal validity. The selection of wear-time thresholds is detailed below.

Participants were recruited from 6 zip codes within eastern Kansas City, Missouri, that were considered low income by the study team because of having the highest scores on the Centers for Disease Control and Prevention Social Vulnerability Index [33] among all zip codes within Kansas City, Missouri. Social Vulnerability Index scores for the 6 selected zip codes ranged from 0.67 to 0.95 on a range of 0 (lowest vulnerability) to 1 (highest vulnerability) based on measures of socioeconomic status, household composition/disability, minority status and language, and housing and transportation. The 6 zip codes also had a large population of Black residents, with 44% to 88% of the residents being Black individuals. Recruitment occurred at COVID-19 vaccination and health screening community events hosted by neighborhood, youth, business, faith-based, and health organizations in collaboration with university health professional schools, and during well-child visits within a large hospital-based pediatric primary care health system serving primarily uninsured and underinsured families from the same communities where the events were held. At the community events, the study team operated a study-information booth and distributed study fliers to event attendees. In the health care system, study team members used a recruitment flier and script to offer enrollment to families while they were waiting to see their provider or just after seeing their provider. For both recruitment approaches, individuals were enrolled on site during the event or clinic visit, except in a few circumstances when a separate enrollment appointment was warranted so that a family member who was not present could be included. These separate appointments occurred in community locations, the study team's office, or remotely over the telephone. Study enrollment occurred between September 2021 and February 2022.

Intervention

The beta-version ActiveKC intervention was informed by prior qualitative research conducted with parents and children from the same low-income Kansas City communities of focus in this study [34]. A primary objective of the ActiveKC pilot project was to obtain participants' input on design elements and intervention content while receiving a beta-version mHealth intervention. This participant input has been reported elsewhere.

ActiveKC was a 7-week mHealth physical activity intervention that integrated the Garmin Vívofit 4 with automated intervention text messages developed by the study team. The messages were sent to the participant's phone number using Microsoft Azure Communication Services and a real-time management application custom built for this project (Figure 1). For youth without a personal mobile phone, text messages were sent to the parent. Study staff helped set up the Garmin monitor, Garmin Connect smartphone app, and Garmin account; set the monitor to automatically synchronize with the Connect app; and demonstrated how to use the monitor and Connect app. Study staff monitored the data reports through the Garmin application programming interface (API) on a weekly basis and reached out to participants who did not have data within the prior 4 days. These participants were contacted via text message and phone call to provide technical assistance around synchronizing their Garmin watch with the Garmin Connect app on their smartphone (eg, because of problems with automatic synchronization). Participants could also request technical support from study staff.

The intervention focused on step counts because they are readily interpretable and preferred by many participants and because step counts and other measures of total physical activity have been consistently linked with health outcomes [35-40]. The intervention content used behavior change techniques [41] that were grounded in Social Cognitive Theory [42] and emphasized active living for building habitual physical activity into daily routines, including walking and movement throughout the day [43]. The automated text messages were sent 4 days/week and were personalized using the participant's Garmin data obtained through the Garmin API, using 2-way messaging starting with a question that was used to tailor the content of the second message, and using the participant's first name.

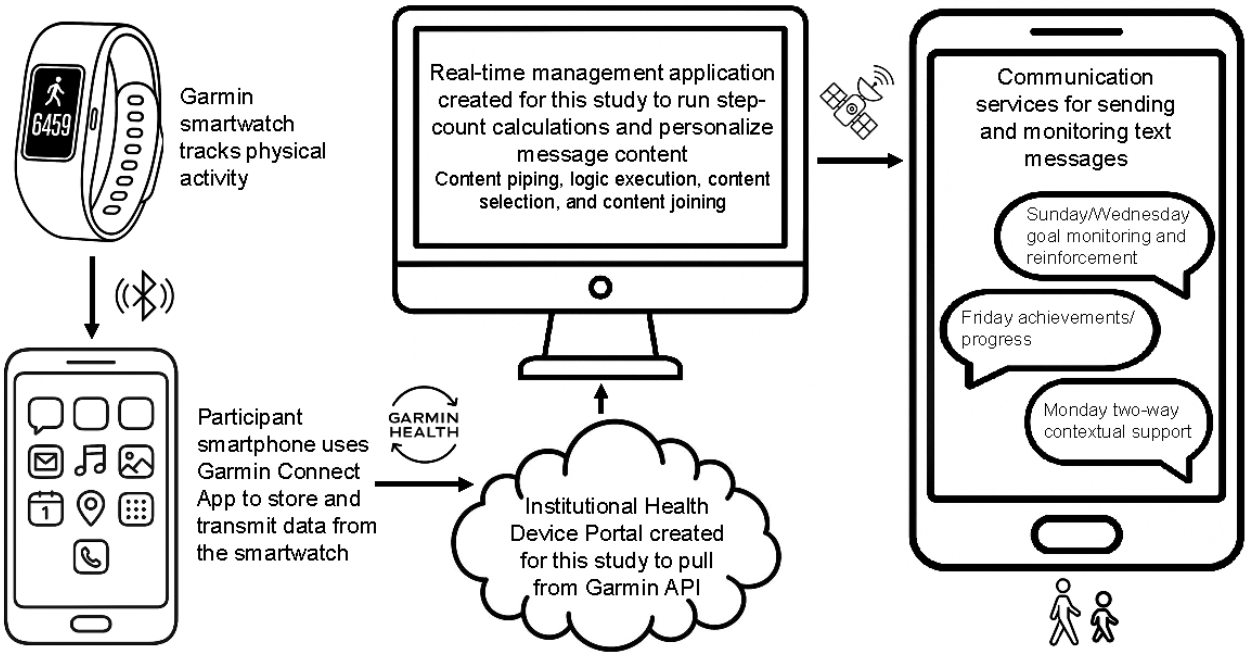
Step goals, monitoring, and reinforcement were core components of the intervention because of their grounding in Social Cognitive Theory [42] and because of consistent evidence showing their important role in behavior change [44]. The activity monitor tracked the participant's daily steps and step goal, and the text messages focused on weekly goals, with each week spanning Sunday to Saturday. The daily step goal was set by the activity monitor based on the participant's daily steps during their baseline week and adjusted each day using Garmin's adaptive algorithm that increased or decreased the goal based on the participant's step counts over the previous days. The weekly goal was the same for all participants: meeting their personalized daily goal on ≥4 days/week. Goal-monitoring messages were sent on Sundays (past-week review) and Wednesdays (midweek check-in). These messages were personalized based on step-count and goal-achievement data from the past week (past-week review; eg, "Congratulations Mallory, you met your step goal on 6 days this past week!") or prior days in the current week (midweek check-in; eg, "Mallory, there's still enough time to meet your step goal on 4 days this week. You just need 2 more days including today. You can do it!"). Conditional branching logic was used in these messages to include language that was reinforcing (eg, "You're a natural!") or encouraging (eg, "You can do it!") based on whether the goal was met or on track to be met [41]. The Sunday messages also included a success story from a fictional participant in the intervention, targeting self-efficacy through vicarious learning with a similar model [45]. The Wednesday messages included a new tip on incorporating habitual physical activity into daily routines (eg, "Weekly activity tip: Sitting too much is bad for us even when we meet our step goal. Stand up every 20-30 minutes and move around for a minute or two!"). Friday messages were also personalized using data from the wearable tracker, revealing the number of miles the participant had walked since the start of the intervention and referencing

XSL•FO
RenderX

an area or origin and destination with a similar distance (eg, the deepest part of the ocean). Throughout the study, participants were sent automated text messages to remind them to wear and/or synchronize the wearable tracker with the Garmin Connect app if data were not recorded on ≥3 consecutive days.

Each Monday, the participant received a 2-way message on a new topic related to multilevel contextual factors: knowledge, values, problem solving, planning, built environment, social support, and maintenance. The use of 2-way messaging allowed the system to send personalized information on the topic based on the participant's response to the first message. The first message comprised a multiple-choice assessment of the context the participant was currently experiencing in relation to the topic. For example, the message for the problem-solving topic asked the participant to choose the top physical activity barrier they were facing that day or week, and the second message provided behavioral support for problem solving the selected barrier (eg, "When your energy is low and being active sounds like a chore, start small. Stand up, wiggle your limbs, and walk at a low pace until you are re-energized!"). Participants were then sent a link to a study web page with more information on the topic and on the unselected response options. Participants had 48 hours to respond to each Monday message to receive a response from the system.

Figure 1. Overview of the ActiveKC Mobile Health System Architecture.



Measures

Sociodemographic Characteristics

Participant sociodemographic characteristics were collected via survey and included age, sex at birth, race/ethnicity, and adult or parent education (categorized as college degree or higher vs no college degree).

Feasibility

Enrollment and retention measures included the proportion of individuals who were enrolled out of all individuals who expressed interest and were assessed for eligibility, as well as the proportion of those enrolled who completed follow-up measurements (Garmin measures at the intervention midpoint and end point and survey measures at the intervention end point). Fidelity was based on the number of total messages per participant that were not received (eg, due to the participant's phone service being disconnected), which was recorded through the messaging services provider. Adherence to wearing the Garmin device was based on the number of days the Garmin was worn across the intervention. The number of participants whom staff needed to contact to provide technical support was also tracked.

Acceptability

The primary items used to capture overall acceptability were: "How difficult or easy was it to use the program?" ("very difficult," "somewhat difficult," "somewhat easy," or "very easy"); "How would you rate the frequency of the messages?" ("too few," "just right," or "too many"); "How much, if at all, did the program help you to be more active?" ("not at all," "somewhat," or "a lot"); and "Would you recommend that your family and friends use the program?" ("yes" or "no"). A total of 4 additional items had participants rate how helpful they perceived each of the primary 4 components of the intervention (step-tracker device, text-message content, ability to respond to messages, and web resources), with response options of "not at all," "somewhat," "very," and "extremely." Engagement with the text messages was captured as an additional indicator of acceptability based on text-message logs that tracked the number of times each participant responded to a 2-way message that asked for a reply, with up to 7 responses possible.

XSL•FO
RenderX

Physical Activity

Participants were asked to wear the Garmin Vívofit 4 step tracker throughout the intervention to facilitate intervention content and evaluation. Garmin step trackers have good validity for measuring step counts in adults and youth [46-48]. At the start of the screening period, study staff helped each participant set up a Garmin Connect account on their smartphone to facilitate automatic transfer of data into the Garmin API. Youth without a smartphone were connected to another device in the household (eg, tablet or parent's smartphone). In the present analyses, daily step counts were evaluated over a 1-week baseline prior to the start of the intervention (week 0, also referred to as the run-in period), the intervention midpoint (weeks 3 and 4), and the end of the intervention (weeks 6 and 7). An exception was that the baseline period could have been 5 or 6 days if the participant was enrolled on a Monday or Tuesday. To account for nonwear time, 15-minute periods with a maximum motion intensity of 0, indicating no movement of the watch, were considered nonwear. Participant-level estimates for each assessment period were based on only valid days, defined as having ≥8 hours of wear time and ≥100 total steps. Participant-level estimates were set to missing if the participant did not have ≥3 valid days during the assessment period. These wear time criteria aimed to capture habitual physical activity while minimizing missing data [49,50]. The outcome variable of interest was an average daily step count for the participant at each assessment period. Minutes of physical activity was explored. However, those results are not reported because the Garmin model that was used calculated physical activity based on step counts and the results were redundant with the step-count outcome. Participants' awareness of their physical activity levels was measured at baseline and at the end of the intervention with 1 item asking whether they generally knew how much physical activity or steps they obtained each day, with a 4-point response scale ranging from "strongly disagree" to "strongly agree."

Analysis

Descriptive statistics were summarized using frequencies, means, and SDs. Feasibility and acceptability benchmarks that were perceived to be reasonable targets in the current context were selected to help guide the interpretation of findings and were generally based on prior pilot behavioral and mHealth intervention studies [51,52]. These benchmarks included ≥60% for enrollment/eligibility, ≥80% for retention and fidelity, ≥70% for adherence, ≥80% for acceptability ratings, and ≥70% for engagement [51,52]. ANOVA and chi-square tests were used to investigate differences by eligibility/enrollment status, baseline physical activity, and level of engagement. Engagement as indicated by responses to messages was investigated as a categorical variable based on tertiles, which resulted in the following groups: replied to 0-1 messages, replied to 2-4 messages, and replied to 5-7 messages. We used 2-tailed *t* tests to investigate differences from preintervention to postintervention in participants' awareness of the amount of physical activity they engaged in. Changes in daily step counts over time were explored using mixed-effects linear regression models, with participant entered as the nesting variable and time entered as a categorical fixed effect (coded as 0=baseline, 1=midpoint, and 2=end point) to test differences at both the midpoint and end point. Models were adjusted for daily wear time (minutes/day), number of valid wear days, and proportion of valid wear days that were weekdays during each measurement period as time-varying fixed effects. These models were explored in the full adult and youth sample as well as post hoc by subgroup based on meeting less than two-thirds of the recommended daily steps at baseline based on the physical activity guidelines (<5000 steps/day for adults and <8000 for youth) [30-32], which is aligned with clinical trials that have focused on individuals with lower activity. These participants were of interest given the low levels of physical activity in the US population, particularly among low-income groups [5]. There is also greater potential for improvement among these individuals and a need for interventions that can help individuals move from low to moderate levels of physical activity [1,4,5]. The same models were also explored within subgroups based on level of engagement with the messages asking for a response, to capture dose-response and help inform whether observed increases in step counts were potentially attributable to greater exposure to intervention content. An intent-to-treat approach was used by including those lost to follow-up in the mixed-effects models and using restricted maximum likelihood estimation to account for the missing Garmin data. Effect sizes (Cohen *d*) were calculated by dividing the mean change in daily steps from baseline to the end of the intervention by the SD in this change. Given this pilot study was not powered to detect effectiveness, interpretations of the data were focused more on effect sizes than *P* values. All analyses were conducted separately for adults and youth using SPSS (version 27.0, SPSS Inc).

Results

Feasibility

Figure 2 presents study enrollment, exclusions, and retention information based on the CONSORT (Consolidated Standards of Reporting Trials) guidelines. Among adults, 64% (84/132) of those assessed for eligibility were enrolled into the intervention, although 1 participant did not receive most of the intervention text messages because of their phone being disconnected or because they changed phone numbers. Similar enrollment rates were observed for youth, with 63% (31/49) of those assessed for eligibility being enrolled. A total of 19% (25/132) of adults were ineligible because of the step-count criteria during the run-in period, whereas no youth were ineligible based on these criteria. No individuals who were screened for eligibility were ineligible because of lack of access to a smartphone.

The analytic sample comprised 83 adults and 31 youth. Among adult participants, 88% (73/83) were female, 84% (70/83) were Black or African American individuals, and the mean age was 50 years (range 19-88 years) (Table 1). A total of 77% (24/31) of the youth participants were female, 77% (24/31) were Black or African American individuals, and the mean age was 13 years (range 8-17 years). A total of 84% (26/31) of the youth participants had a parent who also participated. The eligible/enrolled (n=83) and ineligible/not enrolled (n=48) adults were similar with regards to age and race/ethnicity, although

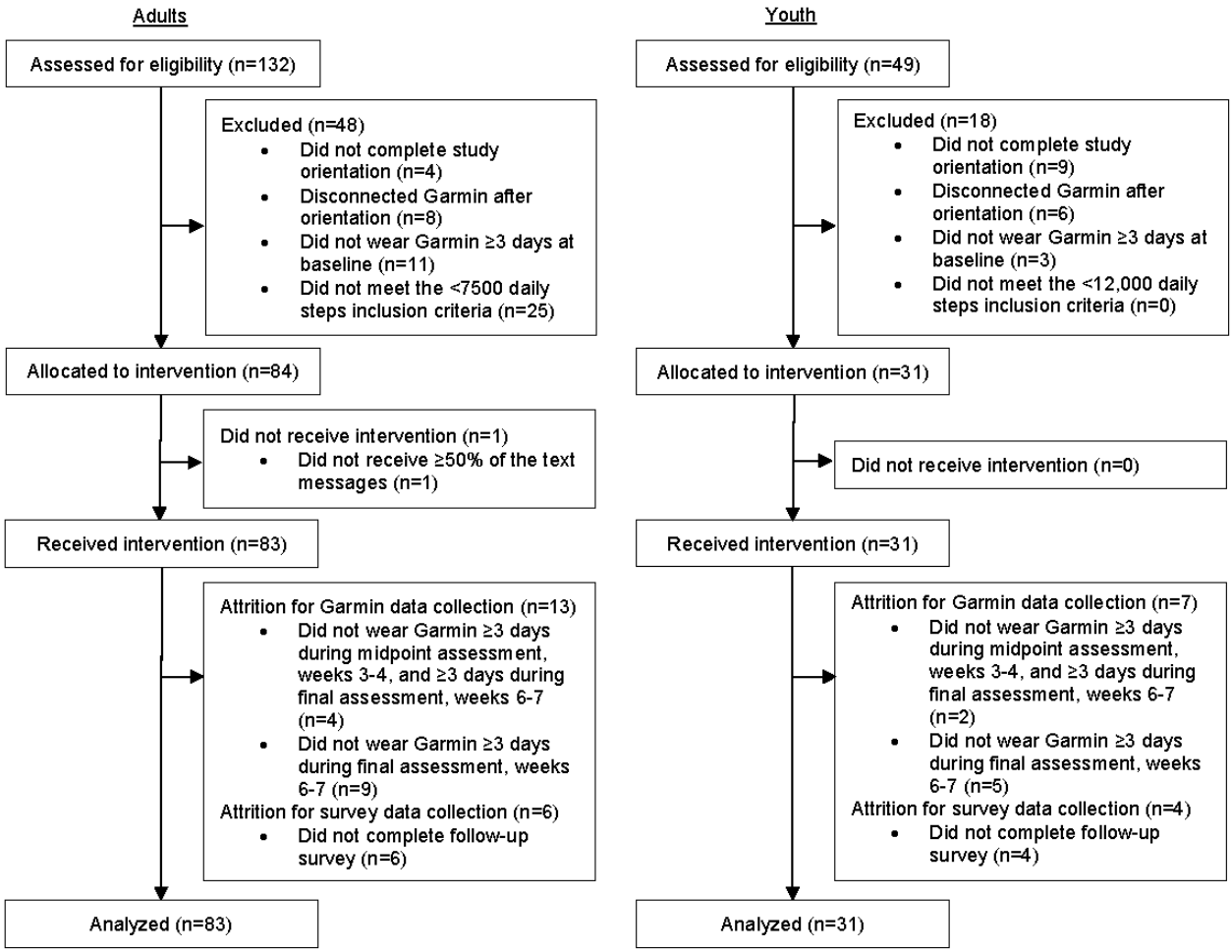
XSL•FO
RenderX

enrolled adults were significantly (P<.05) more likely to have a college degree (30/80, 37% vs 7/46, 15%) and nonsignificantly more likely to be female (73/83, 88% vs 39/48, 75%; Table S1 in Multimedia Appendix 1). The eligible/enrolled (n=31) and ineligible/not enrolled (n=18) youth were similar with regards to age, although enrolled youth were significantly more likely to be female (24/31, 77% vs 4/12, 33%), significantly more likely to be Black individuals (24/31, 77% vs 9/17, 53%), and nonsignificantly more likely to have a parent with a college degree (6/29, 21% vs 0/12, 0%; Table S2 in Multimedia Appendix 1). For 48% (15/31) of the youth participants, messages were sent to the parent's phone. These youth were significantly younger than those whose messages were sent to their own phone (mean age 10.5, SD 1.9 years vs mean age 15.1, SD 1.9 years) and nonsignificantly more likely to have a parent with a college degree (4/14, 29% vs 1/14, 7%), although the 2 groups were similar with regards to sex, race/ethnicity, or parental education (Table S2 in Multimedia Appendix 1).

Retention rates were 95% (79/83) among adults and 94% (29/31) among youth for the step-tracker midpoint assessment, and 84% (70/83) among adults and 77% (24/31) among youth for the step-tracker end point assessment. Retention was 93% (77/83) among adults and 87% (27/31) among youth for the survey end point assessment. Across the 3955 messages that were sent, a total of 1.1% (45/3955) failed to be delivered, largely because of mobile phone numbers being inactive for a period of time. No participants experienced adverse events related to the intervention.

For adherence to wearing the step tracker, adults and youth had a mean of 46 (SD 13) valid wear days and 44 (SD 13) valid wear days, respectively, over the 54- to 56-day period spanning the 5- to 7-day baseline and 7-week intervention (ie, 45/56, 82% and 44/56, 79%). Wear time at each study assessment period is summarized in Table S3 in Multimedia Appendix 1. Study staff contacted 10 total participants to troubleshoot technical problems, all of which were related to a lack of recent Garmin data. Of these, 9 participants were documented as experiencing technical problems related to their Garmin monitor synchronizing with the Garmin app on their smartphone, while the other participant was believed to have disengaged from wearing the study monitor.

Figure 2. Study CONSORT (Consolidated Standards of Reporting Trials) diagram.



XSL•FO
RenderX

Table 1. Participant sociodemographic characteristics.

Characteristic	Adults (n=83)	Youth (n=31)
Age (years), range	19-88	8-17
Age (years), mean (SD)	49.6 (17.0)	12.9 (3.0)
Female, n (%)	73 (88)	24 (77)
Race/ethnicity, n (%)		
Black/African American	70 (84)	24 (77)
White non-Hispanic	11 (13)	6 (19)
Hispanic/Latino	0 (0)	0 (0)
Other	2 (2)	1 (3)
No college degree, n (%)[a]	50 (63)	23 (79)[b]

[a]Percentages were calculated based on nonmissing responses (3 and 2 participants had missing responses in the adult and youth columns, respectively).
[b]Refers to parent education.

Acceptability

Among adults, 88% (69/78) indicated the intervention was somewhat or very easy to use, 83% (65/78) rated the frequency of messages as just right, 99% (77/78) believed the program helped them be more active, and 97% (75/77) indicated they would recommend the program to family and friends (Table S3 in Multimedia Appendix 1). Among youth, 96% (27/28) indicated the program was somewhat or very easy to use, 93% (26/28) rated the frequency of messages as just right, 100% (28/28) believed the program helped them be somewhat or a lot more active, and 100% (28/28) indicated they would recommend the program to family and friends. The step tracker device was the highest rated component of the intervention, followed by the text message content and ability to respond to messages, and these results were similar for adults and youth (Table S3 in Multimedia Appendix 1).

Adults replied to a mean of 2.6 (SD 2.2) of the 7 text messages that asked for a reply within the allotted 48-hour response period. Youth replied to a mean of 3.2 (SD 2.7) messages. Among adults, those who had a child who also participated replied to more messages than those without a participating child (mean 3.3, SD 2.2 messages vs mean 2.4, SD 2.2 messages), although this difference was nonsignificant (2-tailed t_{81}=1.5; P=.12). Adults who replied to more messages were significantly (P<.05) more likely to be White non-Hispanic and have a college degree, and nonsignificantly more likely to be female and younger in age (Table S1 in Multimedia Appendix 1). Youth who replied to more messages, including those with messages sent to the parent's phone, were similar to those who replied to fewer messages with regard to age, sex, race/ethnicity, and parent education (Table S2 in Multimedia Appendix 1).

Physical Activity

Adults and youth had a mean of 4761 (SD 1719) daily step counts and 6618 (SD 2229) daily step counts at baseline, respectively. Adults and youth experienced small and nonsignificant increases in their daily step counts from baseline to the end of the intervention (weeks 6-7) of a mean of 240 (SD 1799; d=0.13) steps/day and 413 (SD 2698; d=0.15) steps/day, respectively, based on the regression analysis (Table 2).

Exploratory analyses suggested that participants with lower daily step counts at baseline experienced larger increases over the course of the intervention, with adults and youth who entered the study with a daily step count that was less than two-thirds of the recommended daily steps experiencing nonsignificant increases of a mean of 609 (SD 1534; d=0.40) steps/day and 1406 (SD 2445; d=0.58) steps/day by the end of the intervention, respectively (Table 2). This reflected increases of 18% and 26% over baseline. As shown in Tables S1 and S2 in Multimedia Appendix 1, both adults and youth with less than two-thirds of the recommended daily steps at baseline were similar to those with at least two-thirds of the recommended daily steps at baseline with regard to age, sex, race/ethnicity, and adult education.

Exploratory analyses also suggested that participants who had higher engagement with the intervention messages experienced larger increases over the course of the intervention, with adults and youth in the highest tertile of engagement experiencing nonsignificant increases of a mean of 584 (SD 1744; d=0.33) steps/day and 941 (SD 2465; d=0.38) steps/day by the end of the intervention, respectively (Table 2). This reflected increases of 12% (584/4838) and 15% (941/6321) over baseline. Among youth, data suggested that larger increases in daily steps were experienced by those whose parent was receiving messages about the child (mean 530, SD 2690 steps/day; d=0.20), rather than the child receiving the messages on their own mobile phone.

Increases were observed in participants' awareness of their physical activity levels, from a mean of 2.7 (SD 1.0) to 3.1 (SD 0.8) in adults (2-tailed t_{53}=–2.6; P=.01) and a mean of 2.6 (SD 0.7) to 2.9 (SD 0.7) in youth (2-tailed t_{16}=–1.8; P=.10).

XSL•FO
RenderX

Table 2. Daily step counts over the course of the intervention. Estimated means, SEs, unstandardized regression coefficients (β), 95% CIs, and *P* values were estimated from mixed-effects linear regression models with adjustment for daily wear time, number of wear days, and proportion of wear days that were weekdays during each measurement period. Each subsample was investigated in a separate model.

Sample and subsample	Frequency, n	Daily step count, estimated mean (SE)			Change in daily steps at intervention midpoint		Change in daily steps at intervention end point	
		Baseline[a]	Intervention midpoint[b]	Intervention end point[c]	β (95% CI)	*P* value	β (95% CI)	*P* value
Adults								
Full sample	83	4761 (198)	4908 (216)	5000 (244)	148 (–448 to 743)	.62	240 (–387 to 866)	.45
By baseline activity								
≥5000 daily steps at baseline	36	6379 (128)	6256 (299)	6124 (329)	–123 (–793 to 547)	.71	–256 (–971 to 460)	.48
<5000 daily steps at baseline	47	3404 (162)	4049 (210)	4013 (272)	645 (105 to 1185)	.02	609 (–30 to 1247)	.06
By engagement								
Replied to 0-1 messages	32	4830 (337)	4925 (316)	4822 (357)	94 (–843 to 1031)	.84	–8 (997 to 981)	.99
Replied to 2-4 messages	32	4821 (329)	4909 (344)	4797 (376)	89 (–913 to 1091)	.86	–24 (–1057 to 1010)	.96
Replied to 5-7 messages	19	4838 (353)	4740 (491)	5423 (564)	–99 (–1377 to 1180)	.88	584 (–784 to 1952)	.39
Youth								
Full sample	31	6618 (411)	6517 (366)	7031 (482)	–101 (–1221 to 1018)	.86	413 (–877 to 1703)	.52
By baseline activity								
≥8000 daily steps at baseline	9	9590 (317)	7150 (357)	7765 (323)	–2440 (–3482 to –1398)	<.001	–1825 (–2805 to –846)	.002
<8000 daily steps at baseline	22	5386 (289)	6194 (502)	6793 (654)	807 (–389 to 2004)	.18	1406 (–94 to 2907)	.06
By engagement								
Replied to 0-1 messages	11	6913 (547)	7005 (911)	6125 (1212)	91 (–2236 to 2419)	.93	–788 (–3959 to 2382)	.57
Replied to 2-4 messages	9	6980 (681)	5974 (820)	7057 (1042)	–1006 (–3490 to 1478)	.40	77 (–2732 to 2885)	.95
Replied to 5-7 messages	11	6321 (596)	6531 (409)	7262 (577)	209 (–1333 to 1752)	.78	941 (–827 to 2709)	.28
By phone used								
Child's phone	16	7485 (759)	6254 (688)	6809 (978)	–1230 (–3701 to 1240)	.31	–676 (–3518 to 2165)	.63
Parent's phone	15	6397 (594)	6425 (462)	6927 (549)	28 (–1512 to 1568)	.97	530 (–1133 to 2192)	.52

[a]Baseline=week 0

[b]Intervention midpoint=weeks 3-4.

[c]Intervention end point=weeks 6-7.

Discussion

Overview

Findings generally supported high feasibility and acceptability of the beta-version mHealth intervention among adult and youth participants from marginalized, low-income communities. Most benchmarks were achieved for enrollment/eligibility, retention, fidelity, adherence, and acceptability, although level of engagement based on replies to messages was moderate overall and highly variable. The step-count data provided preliminary evidence in support of increased physical activity, particularly among those who had a low amount of physical activity at baseline (less than two-thirds of the daily recommendation) or who were more engaged in the intervention. The engagement and physical activity findings were most promising among

XSL•FO
RenderX

youth, particularly those who were younger and participated as part of a dyad whereby messages were sent to the parent. These findings warrant expansion of the beta-version intervention to further incorporate end-user preferences around content and personalized components along with rigorous evaluation using a randomized controlled trial design.

Feasibility

Feasibility was supported across all indicators for both adults and youth, including rates of enrollment/eligibility (31/49, 63% to 84/132, 64%), retention (24/31, 77% to 77/83, 93%), fidelity of message delivery (3910/3955, 99%), and adherence to wearing the step tracker (44/56, 79% to 46/56, 82% of days). All individuals screened had a smartphone, which reflects the growing ubiquity in smartphone use among low-income populations [13] and supports the potential for mHealth interventions to have high reach among this population group. The exclusion of 19% (25/132) of adults because of already meeting the physical activity guideline aligns with the low levels of physical activity in the US population and generally lower levels among low-income groups [4], although it was alarming that none of the youth who were screened were already meeting the physical activity guideline, highlighting the particular importance of efforts to increase physical activity among youth from these communities. The enrolled participants generally reflected the target population based on sociodemographic factors, although findings suggest a need to improve reach among boys, men, and adults with lower education. Engaging men in health behavior interventions is a known challenge and may warrant interventions tailored for adult men from specific population groups [53]. The low rate of enrollment among boys warrants further investigation and differs from other family-based interventions and interventions among children, which have generally had similar reach among boys and girls [54,55]. The retention rate in youth for the final physical activity assessment period warrants improvement, with potential strategies including incentivization tied to wearing the monitor for specific time periods and even using a separate monitor (eg, research-grade accelerometer) for baseline and end point evaluation measurement given the need to retain participants in assessments even if they disengage from the intervention.

The high level of adherence to wearing the step tracker among adults and youth can also be viewed as a marker of engagement in the intervention, supporting both feasibility and acceptability. Prior studies have observed similar levels of adherence during relatively brief (eg, 12-week) interventions, with poorer adherence over longer periods of time and at follow-up periods [56], indicative of the challenges with ongoing adherence to wearing step trackers. Adherence is especially important in wearable-integrated interventions that rely on activity data. Given the vast body of evidence supporting self-monitoring as one of the most important components in physical activity and behavior change interventions [44,57], there is a need for more research on longer-term adherence to consumer wearable trackers [58]. Future mHealth research should consider testing whether intermittent periods of wearing the tracker can provide benefits that translate to periods of nonwear and testing “back-up” strategies for supporting goal setting and monitoring during periods of nonwear.

Many participants experienced technical problems with the step tracker, primarily with the tracker failing to regularly synchronize with the Garmin app on their smartphone. This challenge has been documented across many prior studies and monitor brands and models and thus is a critical issue facing wearable-integrated mHealth research [59]. Study staff were able to work with many of these participants to achieve resolution, although anecdotal evidence that we were unable to quantify suggested that some problems were recurring and that some participants needed to regularly engage with their Garmin app to ensure proper synchronizing. The exact source of the problem could vary and was often unknown, and the process for implementing and testing solutions could vary somewhat across smartphone software type and version, requiring staff to develop detailed guidance protocols to capture the different scenarios encountered. Overall, there is a need for more research to guide effective and efficient technical support for wearable-based physical activity interventions to reduce missing data and prevent disruptions in delivery of tailored intervention content, particularly among population groups with less experience with technology.

Acceptability

The high overall acceptability ratings of the intervention among adults and youth (65/78, 83% to 28/28, 100%) show that participants generally enjoyed the automated intervention. The step tracker device was a favorite aspect of the intervention, especially among youth. This was likely attributable to the behavioral monitoring features it provided, given participants became more aware of their physical activity during the intervention and given the entry-level device had limited additional features (eg, no internet connectivity or text message alerts). The acceptability ratings suggested that participants liked the message frequency of approximately once per day on 4 days of the week, which is similar to the messaging frequency used in some prior physical activity interventions [15,20,60]. This somewhat frequent delivery of messages helps facilitate timely interaction and check-ins, allowing content to be personalized to experiences related to that day or even time of day. However, the acceptability ratings of the text messages were lower than the overall acceptability ratings for the intervention, and engagement based on replies to messages was moderate and highly variable across participants. These findings indicate a need for refined message content and potentially message types and targets. Qualitative research from this study sample, which is being reported elsewhere in more detail, suggested that participants generally preferred personalized messages over other messages, particularly the 2-way messages and the messages that were customized using their Garmin data, although present findings suggest that even these message types warrant further optimization. Future studies should use additional user-centered design strategies to collect participant input on specific messages and personalization components to optimize content, such as iterative prototyping and co-design activities [61-64]. There is also a need to explore additional approaches for personalization within this study population, such as integrating periodic questionnaires, ecological momentary assessments, location tracking, and/or meteorological monitoring [26-28]. Study designs such as

XSL•FO
RenderX

microrandomized and multifactorial trials allow for testing different types of messages and personalization elements and thus are ideal next steps for optimizing this and similar interventions [65].

Participant engagement with intervention content is a common concern in behavioral interventions, including those using mHealth [66]. The participant characteristics among those who were more likely to engage in responding to messages can help to inform the population groups that future versions of the intervention should focus on, although research is also needed to identify strategies for improving engagement among groups that may be more difficult to engage. The present engagement findings generally suggest that wearable-integrated, text message–based mHealth interventions may be most promising among youth and among younger adults, and that additional strategies are needed to support engagement among older adults, Black individuals, individuals with lower education, and men. Features that have been shown to increase engagement include personalization, which was a focus of this study but through basic initial elements, as well as gamification, social interaction, and live human integration (eg, coaching) [67,68]. Present findings also suggest that participating together as a parent–child dyad may result in higher engagement, which warrants more family-based research in this area. Theory-based frameworks such as Family Systems Theory [69] can guide family-based physical activity interventions and be used to inform personalized content, so their use should be investigated within this context to better address engagement from both the parent and child. It is also important to consider that, given the flexibility and scalability of mHealth tools, lower levels of engagement may be acceptable, especially if behavioral outcomes are still achieved.

The web-based content was rated as the least helpful aspect of the intervention, warranting more exploration of strategies for providing helpful behavior change content beyond what is provided in the text messages themselves. Through qualitative research in this study sample, we discovered that some participants were concerned about clicking on website links that were provided in the text messages, which may have impacted acceptability. Future studies should track website metrics such as visits by each participant, which we have had success doing in other studies through use of participant-specific URLs for each study web page. More research is needed to test strategies for increasing engagement in asynchronous intervention content, which could include personalization of content within web pages, enhancements to credibility, enhancement to aesthetics, multimedia content (eg, pictures, videos, and audio), gamification, peer-to-peer engagement, and/or reinforcing engagement [70,71].

Physical Activity

The magnitude of the changes in daily step counts was small and not likely to be meaningful in the full sample, although the changes appeared meaningful for participants who were more engaged in the intervention or who were further from meeting the recommended daily steps at baseline. The association between engagement and physical activity has also been observed in prior research [72] and suggests that the observed changes were more likely to be attributable to intervention participation than to an external or confounding factor. These exploratory findings were also useful in informing who may be more likely to benefit from future iterations of the intervention, pointing to those with low levels of activity (meeting less than two-thirds of the guideline) and parent–child dyads (ie, children with parent involvement). The approximately 500 steps/day to approximately 1400 steps/day changes in daily steps among these subgroups of youth are comparable to prior digital health interventions in youth from different population groups and to amounts recommended for health improvements, which reflect changes of approximately 1000 steps/day (reflecting approximately 10 minutes/day of moderate-to-vigorous physical activity [MVPA]) [24,30,73,74]. Similarly, the approximately 500 steps/day to approximately 600 steps/day changes in adults are comparable to prior studies in adults and are within the range of 500 steps/day to 600 steps/day (reflecting approximately 30-35 minutes/week of MVPA) that has been shown to produce clinically meaningful health improvements [1,15,16,31,32,75]. However, caution should be used when considering these changes alongside findings from prior studies given the pilot nature of this study and lack of a control group. Overall, the physical activity findings warrant additional research on mHealth physical activity interventions among low-income populations to focus on optimization and effectiveness testing in randomized controlled trials.

Strengths and Limitations

Study strengths and innovations included the focus on a population group that has been historically underrepresented in research and the integration of device-based physical activity data with automated messages that were personalized based on goal achievements. The inclusion of both adults and youth was a study strength given this approach can increase reach and allow families to participate together. However, the intervention content did not differ between adults and youth in this beta-version intervention, and qualitative data collected as part of this study (reported separately) suggest future iterations of the intervention should heavily tailor the content for different age groups (eg, children, adolescents, parents, and other adults). While the intervention content was informed through community engagement, more work is needed to refine the content through additional cocreation strategies to optimize its relevance for the population group. The study design did not support investigation of which intervention components or text messages had the greatest impact on engagement and physical activity. Since message frequency did not vary across weeks or across participants, we were not able to explore the impact of message frequency. We were unable to track whether participants viewed text messages or web pages, which are important aspects of engagement that should be assessed in future research. Survey measures generally relied on single-item ratings that did not have established psychometric properties and thus should be interpreted cautiously. The lack of a control group prevented the ability to distinguish the effects of the intervention from natural changes in physical activity that may have occurred over time and thus limited the ability to make causal inferences. Few of the daily step increases observed were statistically significant, which was likely due to the small sample size and wide CIs, so

XSL•FO
RenderX

results should be confirmed in larger samples. The step count data collected from the Garmin device did not provide insights into activity intensity, so we were unable to distinguish light activity from MVPA. Since the study focused on a low-income group, the results may not generalize to other population groups.

Conclusions

This study showed high feasibility and acceptability for using basic personalized mHealth physical activity intervention components that integrate wearable step trackers with automated text messaging among adults and youth from marginalized, low-income communities. More robust development and testing of this type of intervention appears particularly promising for children with low physical activity through involvement from a parent (ie, dyadic approach). The daily step increases associated with the brief beta-version intervention warrant larger-scale effectiveness research in randomized controlled trials. Given the importance of content personalization, future research is warranted to evaluate additional strategies for delivering highly personalized and engaging content for marginalized population groups.

Acknowledgments

The authors thank Shauna Kim, Maurice Miller, Amy Moran, Maddy Pilla, Shea Roy, and the study participants for their contributions to this work.

Figure 1 was created with assistance from the generative artificial intelligence tool Microsoft Copilot. Generative artificial intelligence was not used in any other portion of the manuscript writing.

Funding

Funding was provided by Jackson County, Missouri (contract number O.5506). Garmin Inc provided some of the monitors for this study. The funders had no role in the design or conduct of this study or in the interpretation or publication of findings.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: JAC, FM, JB-P, DC
Data curation: MM, KS, HSG
Formal analysis: CY, CS, SR
Funding acquisition: JB-P
Investigation: JAC, FM, MM, DC
Methodology: JAC, FM, MM, CS, HSG, DC
Project administration: MM, KS
Resources: HSG, JB-P
Software: HSG
Supervision: JAC, MM
Validation: JAC, CS
Visualization: HSG
Writing – original draft: JAC
Writing – review & editing: JAC, FM, MM, SR, CY, CS, KS, HSG, JB-P, DC

Conflicts of Interest

None declared.

Multimedia Appendix 1
Comparison of participant sociodemographic characteristics, adults.
[DOCX File , 25 KB - mhealth_v14i1e76991_app1.docx]

References

1. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA. The physical activity guidelines for Americans. JAMA 2018;320(19):2020-2028 [FREE Full text] [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
2. Ogden CL, Fakhouri TH, Carroll MD, Hales CM, Fryar CD, Li X. Prevalence of obesity among adults, by household income and education - United States, 2011-2014. MMWR Morb Mortal Wkly Rep 2017;66(50):1369-1373 [FREE Full text] [doi: 10.15585/mmwr.mm6650a1] [Medline: 29267260]

XSL•FO
RenderX

3. Ogden CL, Carroll MD, Fakhouri TH, Hales CM, Fryar CD, Li X. Prevalence of obesity among youths by household income and education level of head of household - United States 2011-2014. MMWR Morb Mortal Wkly Rep 2018;67(6):186-189 [FREE Full text] [doi: 10.15585/mmwr.mm6706a3] [Medline: 29447142]
4. Watson KB, Whitfield G, Chen TJ, Hyde ET, Omura JD. Trends in aerobic and muscle-strengthening physical activity by race/ethnicity across income levels among US adults, 1998-2018. J Phys Act Health 2021;18(S1):S45-S52 [FREE Full text] [doi: 10.1123/jpah.2021-0260] [Medline: 34465650]
5. Carlson JA, Moon M, Steel C, Bai Y, Dodson EA, Dooley EE. Results from the 2024 United States report card on physical activity for children and youth. J Phys Act Health 2026;23(3):293-300. [doi: 10.1123/jpah.2024-0872] [Medline: 41352335]
6. Love R, Adams J, van Sluijs EMF, Foster C, Humphreys D. A cumulative meta-analysis of the effects of individual physical activity interventions targeting healthy adults. Obes Rev 2018;19(8):1164-1172 [FREE Full text] [doi: 10.1111/obr.12690] [Medline: 29701299]
7. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011;11:119 [FREE Full text] [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
8. Jimanez E, Czaja J. Recruitment and retention: two of the most important, yet challenging, tasks in behavioral intervention research. In: Behavioral Intervention Research: Designing, Evaluating, and Implementing. New York: Springer Publishing Company; 2016.
9. Clayton P, Connelly J, Ellington M, Rojas V, Lorenzo Y, Trak-Fellermeier MA. Facilitators and barriers of children's participation in nutrition, physical activity, and obesity interventions: a systematic review. Obes Rev 2021;22(12):e13335 [FREE Full text] [doi: 10.1111/obr.13335] [Medline: 34472191]
10. McVay MA, Yancy WS, Bennett GG, Jung S, Voils CI. Perceived barriers and facilitators of initiation of behavioral weight loss interventions among adults with obesity: a qualitative study. BMC Public Health 2018;18(1):854 [FREE Full text] [doi: 10.1186/s12889-018-5795-9] [Medline: 29996812]
11. Withall J, Jago R, Fox KR. Why some do but most don't. Barriers and enablers to engaging low-income groups in physical activity programmes: a mixed methods study. BMC Public Health 2011;11:507 [FREE Full text] [doi: 10.1186/1471-2458-11-507] [Medline: 21711514]
12. Mendoza-Vasconez AS, Linke S, Muñoz M, Pekmezi D, Ainsworth C, Cano M. Promoting physical activity among underserved populations. Curr Sports Med Rep 2016;15(4):290-297 [FREE Full text] [doi: 10.1249/JSR.0000000000000276] [Medline: 27399827]
13. Mobile fact sheet. Pew Research Center. 2024. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2026-05-13]
14. Davis A, Sweigart R, Ellis R. A systematic review of tailored mHealth interventions for physical activity promotion among adults. Transl Behav Med 2020;10(5):1221-1232. [doi: 10.1093/tbm/ibz190] [Medline: 33044542]
15. Smith DM, Duque L, Huffman JC, Healy BC, Celano CM. Text message interventions for physical activity: a systematic review and meta-analysis. Am J Prev Med 2020;58(1):142-151 [FREE Full text] [doi: 10.1016/j.amepre.2019.08.014] [Medline: 31759805]
16. Mönninghoff A, Kramer JN, Hess AJ, Ismailova K, Teepe GW, Tudor Car L. Long-term effectiveness of mhealth physical activity interventions: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2021;23(4):e26699 [FREE Full text] [doi: 10.2196/26699] [Medline: 33811021]
17. He Z, Wu H, Yu F, Fu J, Sun S, Huang T. Effects of smartphone-based interventions on physical activity in children and adolescents: systematic review and meta-analysis. JMIR Mhealth Uhealth 2021;9(2):e22601 [FREE Full text] [doi: 10.2196/22601] [Medline: 33522980]
18. Wang J, Zhu Z, Shuling Z, Fan J, Jin Y, Gao Z. Effectiveness of mHealth app-based interventions for increasing physical activity and improving physical fitness in children and adolescents: systematic review and meta-analysis. JMIR Mhealth Uhealth 2024;12:e51478 [FREE Full text] [doi: 10.2196/51478] [Medline: 38687568]
19. Eckerstorfer LV, Tanzer NK, Vogrincic-Haselbacher C, Kedia G, Brohmer H, Dinslaken I. Key elements of mhealth interventions to successfully increase physical activity: meta-regression. JMIR Mhealth Uhealth 2018;6(11):e10076 [FREE Full text] [doi: 10.2196/10076] [Medline: 30425028]
20. Joseph RP, Royse KE, Benitez TJ. A systematic review of electronic and mobile health (e- and mhealth) physical activity interventions for African American and Hispanic women. J Phys Act Health 2019;16(3):230-239. [doi: 10.1123/jpah.2018-0103] [Medline: 30782040]
21. McCoy P, Leggett S, Bhuiyan A, Brown D, Frye P, Williams B. Text messaging: an intervention to increase physical activity among African American participants in a faith-based, competitive weight loss program. Int J Environ Res Public Health 2017;14(4):326 [FREE Full text] [doi: 10.3390/ijerph14040326] [Medline: 28353650]
22. Olofintuyi A, Roberson K, Ilesanmi A, Dearman C, Carr L, Onsomu E. The use of text messaging to promote physical activity in African-American college students: a feasibility study. J Natl Soc Allied Health 2018;15(1):30-44 [FREE Full text] [Medline: 35222808]

XSL•FO
RenderX

23. Pathak LE, Aguilera A, Williams JJ, Lyles CR, Hernandez-Ramos R, Miramontes J. Developing messaging content for a physical activity smartphone app tailored to low-income patients: user-centered design and crowdsourcing approach. JMIR Mhealth Uhealth 2021;9(5):e21177 [FREE Full text] [doi: 10.2196/21177] [Medline: 34009130]
24. Baumann H, Fiedler J, Wunsch K, Woll A, Wollesen B. mHealth interventions to reduce physical inactivity and sedentary behavior in children and adolescents: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2022;10(5):e35920 [FREE Full text] [doi: 10.2196/35920] [Medline: 35544294]
25. Murimi MW, Nguyen B, Moyeda-Carabaza AF, Lee HJ, Park OH. Factors that contribute to effective online nutrition education interventions: a systematic review. Nutr Rev 2019;77(10):663-690. [doi: 10.1093/nutrit/nuz032] [Medline: 31290970]
26. Nahum-Shani I, Smith S, Spring B, Collins L, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
27. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019;16(1):31 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
28. De Santis KK, Jahnel T, Matthias K, Mergenthal L, Al Khayyal H, Zeeb H. Evaluation of digital interventions for physical activity promotion: scoping review. JMIR Public Health Surveill 2022;8(5):e37820 [FREE Full text] [doi: 10.2196/37820] [Medline: 35604757]
29. Des Jarlais DC, Lyles C, Crepaz N, TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health 2004;94(3):361-366. [doi: 10.2105/ajph.94.3.361] [Medline: 14998794]
30. Tudor-Locke C, Craig CL, Beets MW, Belton S, Cardon GM, Duncan S. How many steps/day are enough? For children and adolescents. Int J Behav Nutr Phys Act 2011;8:78 [FREE Full text] [doi: 10.1186/1479-5868-8-78] [Medline: 21798014]
31. Tudor-Locke C, Craig CL, Brown WJ, Clemes SA, De Cocker K, Giles-Corti B. How many steps/day are enough? For adults. Int J Behav Nutr Phys Act 2011;8:79 [FREE Full text] [doi: 10.1186/1479-5868-8-79] [Medline: 21798015]
32. Tudor-Locke C, Craig CL, Aoyagi Y, Bell RC, Croteau KA, De Bourdeaudhuij I. How many steps/day are enough? For older adults and special populations. Int J Behav Nutr Phys Act 2011;8:80 [FREE Full text] [doi: 10.1186/1479-5868-8-80] [Medline: 21798044]
33. Flanagan B, Gregory E, Hallisey E, Heitgerd J, Lewis B. A social vulnerability index for disaster management. J Homel Secur Emerg Manag 2011;8(1):3. [doi: 10.2202/1547-7355.1792]
34. Forseth B, Ortega A, Hibbing P, Moon M, Steel C, Singh M. Adding family digital supports to classroom-based physical activity interventions to target in- and out-of-school activity: an evaluation of the Stay Active intervention during the COVID-19 pandemic. J Healthy Eat Act Living 2021;1(4):226-240 [FREE Full text] [Medline: 37771566]
35. Bassett DR, Toth LP, LaMunion SR, Crouter SE. Step counting: a review of measurement considerations and health-related applications. Sports Med 2017;47(7):1303-1315 [FREE Full text] [doi: 10.1007/s40279-016-0663-1] [Medline: 28005190]
36. Poitras VJ, Gray CE, Borghese MM, Carson V, Chaput J, Janssen I. Systematic review of the relationships between objectively measured physical activity and health indicators in school-aged children and youth. Appl Physiol Nutr Metab 2016;41(6 Suppl 3):S197-S239 [FREE Full text] [doi: 10.1139/apnm-2015-0663] [Medline: 27306431]
37. Saint-Maurice PF, Troiano RP, Bassett DR, Graubard BI, Carlson SA, Shiroma EJ. Association of daily step count and step intensity with mortality among US adults. JAMA 2020;323(12):1151-1160 [FREE Full text] [doi: 10.1001/jama.2020.1382] [Medline: 32207799]
38. Hamaya R, Shiroma EJ, Moore CC, Buring JE, Evenson KR, Lee I. Time- vs step-based physical activity metrics for health. JAMA Intern Med 2024;184(7):718-725. [doi: 10.1001/jamainternmed.2024.0892] [Medline: 38767892]
39. Kraus WJ, Janz KF, Powell KE, Campbell WW, Jakicic JM, Troiano RP. Daily step counts for measuring physical activity exposure and its relation to health. Med Sci Sports Exerc 2019;51(6):1206-1212 [FREE Full text] [doi: 10.1249/MSS.0000000000001932] [Medline: 31095077]
40. Ekelund U, Tarp J, Steene-Johannessen J, Hansen BH, Jefferis B, Fagerland MW. Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis. BMJ 2019;366:l4570 [FREE Full text] [doi: 10.1136/bmj.l4570] [Medline: 31434697]
41. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
42. Bandura A. Social Foundations of Thought and Action. Englewood Cliffs: Prentice-Hall; 1986:23-28.
43. Physical Activity Guidelines for Americans. 2nd Edition.: U.S. Department of Health and Human Services; 2018. URL: https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf [accessed 2026-05-13]
44. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
45. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]

XSL•FO
RenderX

46. Evenson KR, Spade CL. Review of validity and reliability of garmin activity trackers. J Meas Phys Behav 2020;3(2):170-185 [FREE Full text] [doi: 10.1123/jmpb.2019-0035] [Medline: 32601613]
47. Šimůnek A, Dygrýn J, Jakubec L, Neuls F, Frömel K, Welk GJ. Validity of Garmin Vívofit 1 and Garmin Vívofit 3 for school-based physical activity monitoring. Pediatr Exerc Sci 2019;31(1):130-136. [doi: 10.1123/pes.2018-0019] [Medline: 30304974]
48. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
49. Cain KL, Sallis JF, Conway TL, Van Dyck D, Calhoon L. Using accelerometers in youth physical activity studies: a review of methods. J Phys Act Health 2013;10(3):437-450 [FREE Full text] [doi: 10.1123/jpah.10.3.437] [Medline: 23620392]
50. Troiano R, Berrigan D, Dodd K, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
51. Teresi JA, Yu X, Stewart AL, Hays RD. Guidelines for designing and evaluating feasibility pilot studies. Med Care 2022;60(1):95-103 [FREE Full text] [doi: 10.1097/MLR.0000000000001664] [Medline: 34812790]
52. Ying X, Ehrhardt S. Pilot trial characteristics, postpilot design modifications, and feasibility of full-scale trials. JAMA Netw Open 2023;6(9):e2333642 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.33642] [Medline: 37707813]
53. Waters L, Galichet B, Owen N, Eakin E. Who participates in physical activity intervention trials? J Phys Act Health 2011;8(1):85-103. [doi: 10.1123/jpah.8.1.85] [Medline: 21297189]
54. Brown HE, Atkin AJ, Panter J, Wong G, Chinapaw MJM, van Sluijs EMF. Family-based interventions to increase physical activity in children: a systematic review, meta-analysis and realist synthesis. Obes Rev 2016;17(4):345-360 [FREE Full text] [doi: 10.1111/obr.12362] [Medline: 26756281]
55. Rhodes RE, Hollman H, Sui W. Family-based physical activity interventions and family functioning: a systematic review. Fam Process 2024;63(1):392-413. [doi: 10.1111/famp.12864] [Medline: 36748347]
56. Hartman SJ, Chen R, Tam RM, Narayan HK, Natarajan L, Liu L. Fitbit use and activity levels from intervention to 2 years after: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2022;10(6):e37086 [FREE Full text] [doi: 10.2196/37086] [Medline: 35771607]
57. Rose T, Barker M, Maria Jacob C, Morrison L, Lawrence W, Strömmer S, et al. A systematic review of digital interventions for improving the diet and physical activity behaviors of adolescents. J Adolesc Health 2017;61(6):669-677 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.05.024] [Medline: 28822682]
58. Grady A, Pearson N, Lamont H, Leigh L, Wolfenden L, Barnes C. The effectiveness of strategies to improve user engagement with digital health interventions targeting nutrition, physical activity, and overweight and obesity: systematic review and meta-analysis. J Med Internet Res 2023;25:e47987 [FREE Full text] [doi: 10.2196/47987] [Medline: 38113062]
59. Phillips SM, Cadmus-Bertram L, Rosenberg D, Buman MP, Lynch BM. Wearable technology and physical activity in chronic disease: opportunities and challenges. Am J Prev Med 2018;54(1):144-150 [FREE Full text] [doi: 10.1016/j.amepre.2017.08.015] [Medline: 29122356]
60. Ludwig K, Arthur R, Sculthorpe N, Fountain H, Buchan DS. Text messaging interventions for improvement in physical activity and sedentary behavior in youth: systematic review. JMIR Mhealth Uhealth 2018;6(9):e10799 [FREE Full text] [doi: 10.2196/10799] [Medline: 30224335]
61. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
62. Wang T, Zhu H, Qian S, Giunti G, Goossens R, Melles M. Designing digital patient experiences: the digital health design framework. Appl Ergon 2024;119:104289 [FREE Full text] [doi: 10.1016/j.apergo.2024.104289] [Medline: 38688183]
63. Tzimourta KD. Human-centered design and development in digital health: approaches, challenges, and emerging trends. Cureus 2025;17(6):e85897. [doi: 10.7759/cureus.85897] [Medline: 40548156]
64. Griffin AC, Javier SJ, Golding M, Runnels TW, Matthias MS, Shimada SL. Designing patient-friendly messages: tutorial on applying human-centered, self-determination theory with AI considerations. J Med Internet Res 2025;27:e78173 [FREE Full text] [doi: 10.2196/78173] [Medline: 41105949]
65. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. Health Psychol 2015;34S:1220-1228 [FREE Full text] [doi: 10.1037/hea0000305] [Medline: 26651463]
66. Domin A, Spruijt-Metz D, Theisen D, Ouzzahra Y, Vögele C. Smartphone-based interventions for physical activity promotion: scoping review of the evidence over the last 10 years. JMIR Mhealth Uhealth 2021;9(7):e24308 [FREE Full text] [doi: 10.2196/24308] [Medline: 34287209]
67. Schwarz A, Winkens LHH, de Vet E, Ossendrijver D, Bouwsema K, Simons M. Design features associated with engagement in mobile health physical activity interventions among youth: systematic review of qualitative and quantitative studies. JMIR Mhealth Uhealth 2023;11:e40898 [FREE Full text] [doi: 10.2196/40898] [Medline: 36877551]

XSL•FO
RenderX

68. Alencar M, Johnson K, Gray V, Mullur R, Gutierrez E, Dionico P. Telehealth-based health coaching increases m-health device adherence and rate of weight loss in obese participants. Telemed J E Health 2020;26(3):365-368 [FREE Full text] [doi: 10.1089/tmj.2019.0017] [Medline: 30994410]
69. Rhodes RE, Guerrero MD, Vanderloo LM, Barbeau K, Birken CS, Chaput J. Development of a consensus statement on the role of the family in the physical activity, sedentary, and sleep behaviours of children and youth. Int J Behav Nutr Phys Act 2020;17(1):74 [FREE Full text] [doi: 10.1186/s12966-020-00973-0] [Medline: 32539730]
70. Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H. Design features for improving mobile health intervention user engagement: systematic review and thematic analysis. J Med Internet Res 2020;22(12):e21687 [FREE Full text] [doi: 10.2196/21687] [Medline: 33295292]
71. Oakley-Girvan I, Yunis R, Longmire M, Ouillon JS. What works best to engage participants in mobile app interventions and e-health: a scoping review. Telemed J E Health 2022;28(6):768-780 [FREE Full text] [doi: 10.1089/tmj.2021.0176] [Medline: 34637651]
72. Mclaughlin M, Delaney T, Hall A, Byaruhanga J, Mackie P, Grady A. Associations between digital health intervention engagement, physical activity, and sedentary behavior: systematic review and meta-analysis. J Med Internet Res 2021;23(2):e23180 [FREE Full text] [doi: 10.2196/23180] [Medline: 33605897]
73. Cai S, Wang H, Zhang Y, Zhao T, Yuan X, Deng H. Could physical activity promote indicators of physical and psychological health among children and adolescents? An umbrella review of meta-analyses of randomized controlled trials. World J Pediatr 2025;21(2):159-173. [doi: 10.1007/s12519-024-00874-3] [Medline: 39847308]
74. Stojanović S, Andrieieva O, Trajković N. Associations between number of steps and health outcomes in children and adolescents: a systematic review and meta-analysis. BMC Public Health 2024;24(1):3310 [FREE Full text] [doi: 10.1186/s12889-024-20835-9] [Medline: 39604923]
75. Wahlich C, Chaudhry UAR, Fortescue R, Cook DG, Hirani S, Knightly R. Effectiveness of adult community-based physical activity interventions with objective physical activity measurements and long-term follow-up: a systematic review and meta-analysis. BMJ Open 2020;10(5):e034541 [FREE Full text] [doi: 10.1136/bmjopen-2019-034541] [Medline: 32371512]

Abbreviations

API: application programming interface
CONSORT: Consolidated Standards of Reporting Trials
mHealth: mobile health
MVPA: moderate-to-vigorous physical activity

Edited by A Stone, L Buis; submitted 20.May.2025; peer-reviewed by A Graham, E Soltero, N Lukkahatai; comments to author 23.Sep.2025; revised version received 07.May.2026; accepted 08.May.2026; published 05.Jun.2026.

Please cite as:
Carlson JA, Materia F, Moon M, Ryu S, Yeager C, Steel C, Shields K, Gill HS, Berkley-Patton J, Catley D
Automated Physical Activity Support for Adults and Youth From Low-Income Communities: Single-Arm Pilot Study
JMIR Mhealth Uhealth 2026;14:e76991
URL: https://mhealth.jmir.org/2026/1/e76991
doi:10.2196/76991
PMID:

©Jordan A Carlson, Frank Materia, Mallory Moon, Suryeon Ryu, Cory Yeager, Chelsea Steel, Kacee Shields, Harpreet Singh Gill, Jannette Berkley-Patton, Delwyn Catley. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 05.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Tobacco Use Trajectories and Associated Changes in Biometrics and Sleep During the First 72 Weeks of Wearable Membership: Observational Cohort Study

Dylan J Curran[1,2]; Josh Leota[3,4,5], PhD; William von Hippel[1], PhD; Finnbarr Fielding[1], BS; Christopher J Chapman[1], MS; Jenna G Cohen[1], PhD; David M Presby[3,6,7], PhD; Reha Jhunjhunwala[8], DMD; Kristen E Holmes[1], PhD; Gregory J Grosicki[1], PhD

[1]Department of Performance Science and Health Outcomes, WHOOP, Inc, Boston, MA, United States

[2]School of Mathematics, Georgia Institute of Technology, Atlanta, GA, United States

[3]Department of Research, Algorithms, and Data Science, WHOOP, Inc, Boston, MA, United States

[4]Department of Medicine, Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA, United States

[5]School of Psychological Sciences, Monash University, Victoria, Australia

[6]Swiss Institute of Bioinformatics, Lausanne, Switzerland

[7]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland

[8]Department of Digital Health, WHOOP, Inc, Boston, MA, United States

Corresponding Author:
Gregory J Grosicki, PhD
Department of Performance Science and Health Outcomes
WHOOP, Inc
1 Kenmore Square
Boston, MA, 02215
United States
Phone: 1 703 303 7812
Email: greg.grosicki@whoop.com

Abstract

Background: Tobacco use remains a leading preventable cause of morbidity and mortality. Digital health tools and wearable technologies offer scalable opportunities for behavioral self-monitoring. However, real-world evidence characterizing long-term tobacco use trajectories and associated physiological changes during wearable adoption is limited.

Objective: This study aims to characterize longitudinal trajectories of self-reported tobacco use during the first 72 weeks of wearable adoption and to examine associations between tobacco use and wearable-derived cardiopulmonary and sleep measures.

Methods: We analyzed data from 12,678 new wearable members (18-79 years) who contributed up to 72 weeks of daily self-reported tobacco use and wearable-derived biometric data. Longitudinal trajectories of tobacco use were examined across prespecified 12-week quarters (Q1-Q6) using generalized linear mixed-effects models. Associations between tobacco use and wearable-derived nocturnal resting heart rate (RHR), heart rate variability (HRV), respiratory rate (RR), and sleep duration were evaluated using linear mixed-effects models that accounted for within- and between-person variation and adjusted for demographic and temporal covariates.

Results: Across 3,765,573 person-days, the estimated daily probability of tobacco use declined from 55.1% (95% CI 53.8-56.4) during Q1 to 27.2% (95% CI 26.1-28.3) during Q6, representing an absolute reduction of 27.9 percentage points (95% CI –28.4 to –27.4). Among tobacco users with end-of-follow-up data, over one-quarter (1404/4975, 28.22%) reported no tobacco use during Q6. Greater logging engagement was associated with larger reductions in tobacco use; each 10-percentage-point increase in engagement corresponded to a 0.92-percentage-point greater decline from Q1 to Q6 (95% CI –1.58 to –0.26). Following tobacco use days, RHR was 1.71 beats/minute higher (95% CI 1.70-1.73), HRV was 3.54 ms lower (95% CI –3.59 to –3.49), RR was 0.19 breaths/minute higher (95% CI 0.19-0.20), and sleep duration was 9.78 minutes shorter (95% CI –10.08 to –9.49) relative to nonuse days. Reductions in tobacco use over time were associated with directionally favorable changes in RHR (P=.001), RR (P=.002), and sleep duration (P=.02).

XSL•FO
RenderX

Conclusions: In this large-scale, observational, real-world study of wearable users, the probability of tobacco use declined over the first 72 weeks of adoption. Among participants with end-of-follow-up data, more than one-quarter reported no tobacco use during Q6. Tobacco use was consistently associated with less favorable cardiopulmonary and sleep measures, while reductions in tobacco use over time co-occurred with directionally favorable changes in these measures, although such changes may also reflect broader lifestyle or health changes. Overall, this study provides large-scale, longitudinal evidence that sustained reductions in tobacco use co-occur with favorable changes in physiological markers within a digital self-monitoring environment. As the findings derive from a single commercial ecosystem, independent replication in noncommercial, multiplatform settings will be needed to establish generalizability.

(JMIR Mhealth Uhealth 2026;14:e98116) doi:10.2196/98116

KEYWORDS

wearable technology; digital health; cigarettes; behavior change; cardiopulmonary

Introduction

Tobacco use remains a leading global cause of morbidity and mortality [1], largely due to its contributions to cardiovascular disease, cancer, and chronic respiratory disease [2]. Although age-standardized rates of tobacco-related mortality have declined over the past century, the absolute burden remains high. For example, in the 2019 Global Burden of Disease study [3], tobacco use was associated with an estimated 8.7 million deaths and 229.8 million disability-adjusted life years globally. Despite widespread awareness of health risks and strong motivation to quit among tobacco users, the habitual and addictive nature of tobacco use makes sustained cessation challenging [4]. A variety of pharmacological and behavioral cessation interventions have been developed to address this challenge, but these approaches vary widely in cost, effectiveness, and scalability [5].

Digital health technologies are increasingly recognized as promising tools for supporting behavior change across a range of health behaviors [6,7]. Prior work suggests that digital health platforms and self-monitoring tools may facilitate reductions in substance use through increased awareness, feedback, and habit formation [8]. Work in psychological control theory provides a potential explanation for this change, which posits that awareness of discrepancies between desired states and observed states may prompt behavior change [9]. Still, evidence for digital tobacco cessation interventions remains mixed. A recent systematic review reported potential benefits of digital health approaches but noted substantial heterogeneity in intervention design, participant characteristics, and outcomes [10]. Importantly, the evidence base is characterized by relatively modest sample sizes, raising questions about generalizability and persistence of observed effects. Indeed, a recent systematic review and meta-analysis reported that digital health interventions may be most effective in the short (<3 months) and medium term (3-9 months), with less certainty regarding longer-term maintenance of cessation [11]. Collectively, these findings highlight both the promise and limitations of existing digital tobacco cessation interventions and emphasize the need for scalable, real-world evidence examining how continuous self-monitoring may relate to changes in tobacco use over extended periods.

Leveraging a large real-world dataset, we examined longitudinal trajectories of self-reported tobacco use (ie, use of tobacco products as reported in a smartphone app) during the first 72 weeks of wearable membership and how trajectories differed by age, biological sex, baseline tobacco use frequency, and engagement with optional in-app goal setting. We also evaluated associations between tobacco use and wearable-derived measures, including resting heart rate (RHR), heart rate variability (HRV), respiratory rate (RR), and sleep duration, and assessed whether changes in tobacco use were associated with corresponding changes in these outcomes. We conducted these analyses using the WHOOP platform (WHOOP, Inc), which served 2 primary purposes in the study. First, it provided the combination of features required for this investigation: daily user-initiated self-reports of tobacco use, continuous nightly cardiorespiratory and sleep measures, and a recently onboarded member base of sufficient scale to support longitudinal analyses across 72 weeks. Second, it functioned as a digital self-monitoring environment in which users receive insights on their sleep, recovery, and strain; accordingly, this study characterizes tobacco use trajectories and associated physiological changes within this larger environment rather than any isolated platform feature. The platform has previously been leveraged for analogous real-world longitudinal characterization of health behavior trajectories [6,7,12], including a recent companion analysis of alcohol use [7].

Methods

Overview

We conducted a 72-week observational cohort study of WHOOP members who contributed daily self-reported tobacco use, and continuous wearable-derived cardiopulmonary and sleep measures. Participants self-reported tobacco use through a daily morning prompt within the WHOOP smartphone app, with the option to set tobacco-related goals at any point during the observation period. Detailed descriptions of the study population, wearable device and measures, self-reported tobacco use, and statistical analyses follow.

Study Population

We studied international adults aged 18-79 years who joined the WHOOP platform between January 2022 and June 2024 and met prespecified reporting criteria. Participants enrolled on a rolling basis and contributed up to 504 days (72 weeks) of data. Time was structured into consecutive, nonoverlapping 7-day weeks anchored to each participant's join date. Quarters

XSL•FO
RenderX

(Q1-Q6) were defined relative to each individual's join date as consecutive 12-week intervals.

Eligibility required at least 8 weeks with tobacco use reporting (≥10% of the 72-week observation window). Eligible weeks were defined as those with either complete reporting (ie, explicit "yes" or "no" entries for all 7 days) or at least 3 tobacco use entries to capture meaningful reporting even if all 7 days were not logged. These weeks did not need to be consecutive and could occur at any time during the observation period. These thresholds were selected a priori to balance adequate longitudinal coverage with retention of a broadly representative cohort rather than restricting analyses to only highly consistent reporters.

Ethics Considerations

All members consented to the use of their deidentified data for research purposes at account creation under WHOOP's Terms of Service. The study protocol was reviewed and approved under expedited procedures by an independent institutional review board (Salus Institutional Review Board; approval number 251121). The study is reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for observational cohort studies (Multimedia Appendix 1).

Wearable Device and Measures

WHOOP (WHOOP, Inc) is a subscription-based wearable platform consisting of a wrist-worn device and companion smartphone app used to derive sleep, recovery, and strain metrics that provide members with personalized insights and recommendations (Figure S1 in Multimedia Appendix 2). The WHOOP strap combines accelerometry and photoplethysmography to detect major sleep periods of at least 1 hour in duration. Within each detected sleep period, WHOOP classifies each 30-second interval into 1 of 4 stages: wake, light sleep, slow-wave sleep (SWS), and rapid eye movement sleep. Throughout each sleep period, the WHOOP device derives peak-to-peak heartbeat intervals from photoplethysmography to estimate HRV, calculated as the root mean square of successive differences. Beat-to-beat intervals, along with features derived from frequency transformations, are used to estimate RHR throughout the night. To generate nightly RHR and HRV values, WHOOP first filters out epochs identified as wake or as having low signal quality. The remaining epochs are then aggregated using a weighted average, in which epochs with a higher likelihood of SWS and epochs closer to the end of sleep are assigned greater weights. RR is calculated as the median respirations per minute derived from interbeat intervals throughout sleep. Sleep duration is defined as total sleep time (light, SWS, and rapid eye movement). WHOOP has undergone independent validation against gold-standard reference methods in peer-reviewed studies. Compared with electrocardiography, WHOOP-derived heart rate measures demonstrate 99% agreement, and compared with polysomnography, WHOOP demonstrates 86%-89% agreement for 2-stage sleep categorization [13,14].

Self-Reported Tobacco Use

Daily tobacco use was self-reported through the WHOOP smartphone app via the WHOOP Journal feature, a customizable morning prompt through which users confirmed behaviors from the prior day (Figure S1B in Multimedia Appendix 2). Users were prompted to indicate whether they used tobacco ("yes" or "no") and could optionally report the number of times they used tobacco that day (ie, tobacco use instances). These self-reported entries are user-initiated and configurable within the smartphone app; tobacco logging was not required for device use. Primary analyses included only days with explicit "yes" or "no" responses, whereas days with missing responses were used solely to calculate logging engagement, defined as the proportion of active days on the platform with a reported tobacco status.

Members could also set personalized tobacco-related goals within the app, which could range from abstaining entirely to limiting tobacco use on specific days. Goals could be created, modified, or abandoned at any time. Due to heterogeneity in goal types and because goals could be initiated at any point during follow-up, goal setting was analyzed as a fixed binary indicator of whether a member set any tobacco-related goal at any time during the 72-week observation window (any goal vs none). As this approach collapses a dynamic, heterogeneous behavior into a single time-invariant indicator, analyses involving goal setting should be interpreted as exploratory.

Statistical Analysis

Descriptive statistics are reported as means (SDs). Tobacco use frequency was modeled as the daily probability of tobacco use within each user-week using generalized linear mixed-effects models (GLMMs) with a binomial distribution and logit link. Weekly outcomes were specified as the number of tobacco use and nontobacco use days, allowing the binomial denominator to vary according to the number of completed tobacco use entries that week (range 3-7 days). Fixed effects included prespecified 12-week quarters (Q1-Q6), age, biological sex, season, total weeks contributed, and the proportion of days in each user-week that fell on a weekend (Saturday or Sunday entries, which reflect Friday or Saturday tobacco use), with participant-level random intercepts. Models were first fit in the overall sample, followed by evaluation of effect modification by biological sex, age group (young=18-39 years, n=10,248; middle=40-59 years, n=2294; and old=60-79 years, n=136), baseline tobacco use tertiles, and tobacco-related goal setting (goal setters vs nongoal setters). Estimated marginal means were computed on the response scale to obtain model-predicted daily tobacco use probabilities by quarter. Sensitivity analyses used inverse probability of censoring weighting (IPCW) to assess whether differential dropout biased the primary estimates. A logistic regression predicted quarter-to-quarter retention from quarter, age, sex, BMI, and the participant's mean tobacco use proportion in the current quarter. Each participant's IPCW weight was the inverse of their cumulative predicted retention probability across all prior quarter transitions, winsorized at the 1st and 99th percentiles. Weights were applied at the participant-quarter level in the GLMM. To assess the robustness of findings to incomplete tobacco use reporting, we conducted multiple imputation by chained equations under a

XSL•FO
RenderX

missing-at-random assumption. Of all active platform-days contributed by participants (ie, days the wearable was worn; N=4,436,523), 619,295 (13.96%) lacked an explicit yes/no tobacco use response. As the analytic sample additionally required ≥3 reported days per person-week, this day-level denominator exceeds the analytic person-day total. The imputation model included age, biological sex, BMI, weekday versus weekend, season, habitual tobacco use (ie, proportion of days with tobacco use), and the corresponding night's wearable-derived measures (RHR, HRV, RR, and sleep duration) decomposed into within- and between-person components. Tobacco use trajectory analyses were re-fit across imputed datasets and pooled using Rubin's rules [15]. Further sensitivity analyses were conducted among members contributing at least 1 week with complete tobacco use instances reporting (12,271/12,678, 96.79%). Weekly use instances were modeled using negative binomial GLMMs with a log link.

As a complementary descriptive analysis, we characterized the proportion of participants reporting tobacco use during Q1 who achieved sustained nonuse, defined as reporting zero tobacco use days during Q6, conditional on contributing at least 3 weeks in each quarter. We also examined whether logging engagement was associated with changes in tobacco use frequency. For each participant, logging engagement was defined as the proportion of their total active days on the platform during the observation period with a reported tobacco use status (yes/no). Linear models evaluated associations between logging engagement and change in tobacco use frequency from Q1 to Q6, adjusting for baseline (Q1) tobacco use, total platform days, age, sex, and BMI.

Associations between tobacco use and biometric and sleep outcomes over the observation period were evaluated using linear mixed-effects models that decomposed tobacco use into within-person (binary daily deviation from an individual's mean) and between-person (proportion of days with tobacco use) components. The between-person component was defined as each individual's mean across all days with a tobacco use entry over the full 72-week observation window. Outcomes included nightly RHR, HRV, RR, and total sleep duration. Additional analyses examined sleep onset time (bedtime) and the percentage of time in bed spent awake as complementary indicators of sleep continuity and quality. Models included random intercepts for participants and were adjusted for age, biological sex, BMI, weekday versus weekend, and season. For interpretability and visualization, models were fit with between-person tobacco use specified either as a continuous variable or as tertiles (low, medium, and high).

To evaluate whether changes in tobacco use were associated with changes in physiological outcomes, we used a change-score framework restricted to participants with data in Q1 and Q6 (n=5776). Quarterly averages were computed for each participant, and change scores were calculated as the difference between Q6 and Q1. Change in tobacco use was defined as the absolute percentage-point change between quarters. Models adjusted for baseline (Q1) tobacco use percentage, the baseline outcome value, age, biological sex, BMI, and Q1 season. Change in tobacco use was modeled both continuously and categorically in 5 prespecified groups: large increase (≥25 percentage-point increase; n=597), small increase (≥5 to <25 percentage-point increase; n=749), no meaningful change (>−5 to <5 percentage-point change; n=1643), small decrease (≤−5 to >−25 percentage-point decrease; n=1264), and large decrease (≤−25 percentage-point decrease; n=1523). These categorical thresholds were prespecified to facilitate interpretation of the change-score analyses and to distinguish meaningful behavior change from minor fluctuation. Absolute, rather than relative, change was used because it more directly reflects the magnitude of behavioral exposure relevant to physiological outcomes.

Across analyses, post hoc comparisons were adjusted for multiple testing as appropriate. Comparisons across quarters were evaluated relative to Q1 using Dunnett-adjusted contrasts, whereas Tukey-adjusted pairwise comparisons were used for categorical group comparisons. All analyses were conducted in R (version 4.4.2; R Foundation), with statistical significance set at α=.05.

Results

Sample Characteristics

The final dataset included 12,678 participants (n=2678, 21.12%, females) and consisted of 3,765,573 person-days, aggregated into 596,589 person-weeks. The mean age was 32.4 (SD 8.9) years and the mean BMI was 25.7 (SD 4.3) kg/m^2. Participants contributed a mean of 47.1 (SD 18.2) weeks (median 48 weeks; range 8-72 weeks), with an average of 6.3 (SD 1.1) completed tobacco status ("yes" or "no") entries per week. The number of contributing participants declined from 12,626 in Q1 to 5804 in Q6 (Tables S1-S3 in Multimedia Appendix 2). A total of 52 participants met eligibility criteria through reporting that began after Q1, and therefore, are not represented in Q1 counts. Reporting engagement remained relatively high across follow-up, with a mean of 6.5 (SD 0.9) completed tobacco entries per week in Q1 and 6.1 (SD 1.2) in Q6.

Daily Tobacco Use Probability Declines Over 72 Weeks

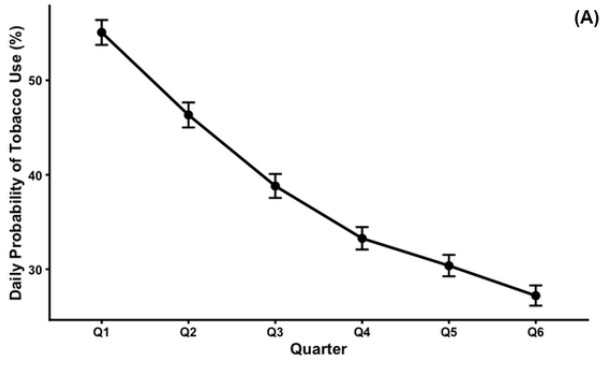
The predicted probability of daily tobacco use declined over the 72-week observation window (Figure 1A), from 55.1% (95% CI 53.8-56.4) in Q1 to 27.2% (95% CI 26.1-28.3) by Q6, representing an absolute decrease of 27.9 percentage points (95% CI –28.4 to –27.4). Estimates were similar in a sensitivity analysis restricted to person-weeks with complete 7-day reporting (ie, 7 explicit "yes" or "no" entries), which included 374,497 of 596,589 (62.77%) person-weeks. In this restricted sample, the predicted daily probability of tobacco use declined from 54.9% (95% CI 53.4-56.4) in Q1 to 26.0% (95% CI 24.8-27.2) in Q6, corresponding to an absolute decrease of 28.9 percentage points (95% CI –29.5 to –28.3; data not shown).

To further test the robustness of these observations, additional sensitivity analyses were conducted. When analyses were restricted to users with all 72 weeks of follow-up (n=1049), results remained consistent (Figure S2 in Multimedia Appendix 2), with the daily probability of tobacco use declining from 66.9% (95% CI 62.1-71.4) in Q1 to 37.4% (95% CI 32.6-42.5) in Q6, a decrease of 29.5 percentage points (95% CI –30.4 to –28.6). Models incorporating IPCW produced estimates consistent with the primary analysis (Table S4 in Multimedia Appendix 2), with the IPCW-adjusted daily probability of

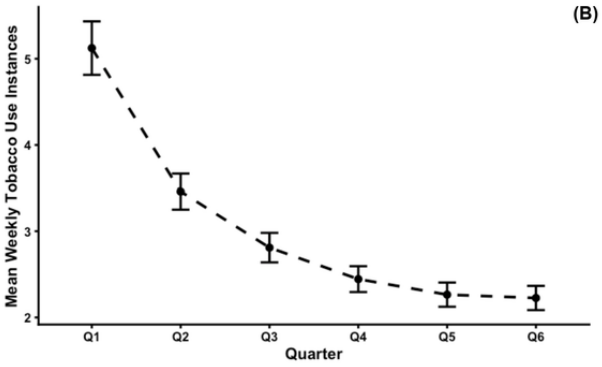
XSL•FO
RenderX

tobacco use declining from 55.3% (95% CI 53.9-56.6) in Q1 to 27.3% (95% CI 26.2-28.4) in Q6, a decrease of 28.0 percentage points (95% CI –28.4 to –27.5). Trajectory estimates derived from multiple-imputation analyses were similarly consistent (Figure S3 in Multimedia Appendix 2), with a Q1-to-Q6 decrease of 25.6 percentage points (95% CI –26.1 to –25.1). Results were also similar when analyses were restricted to weeks with complete 7-day tobacco instance data (ie, all "yes" entries specified the number of use instances, 321,925/596,589, 53.96%, person-weeks; Figure 1B); the model-estimated weekly number of tobacco use instances declined from 5.12 (95% CI 4.81-5.43) in Q1 to 2.23 (95% CI 2.09-2.37) in Q6, a decrease of 2.89 use instances per week (95% CI –3.14 to –2.65).

Figure 1. Adjusted tobacco use over 72 weeks. (A) Adjusted daily probability of tobacco use. (B) Adjusted mean weekly tobacco use instances. Marginal estimates from generalized linear mixed-effects models are shown for 6 sequential 12-week quarters (Q1-Q6). Points represent model-based estimates, and error bars indicate 95% CIs. Absolute changes from Q1 and corresponding Dunnett-adjusted *P* values are based on model contrasts. Panel A is based on a binomial model estimating the probability of a tobacco-use day, whereas Panel B is based on a negative binomial model estimating weekly tobacco use instances. Both models included quarter, age at activation, sex, season, proportion of weekend days, and total weeks contributed as fixed effects, with a random intercept for each participant. N/A: not applicable.



Quarter	Daily probability (%) (95% CI)	Absolute change vs Q1 (pp, 95% CI)	*P* value
Q1	55.1 (53.8 to 56.4)	N/A	N/A
Q2	46.3 (45.0 to 47.7)	–8.8 (–9.0 to –8.5)	<.001
Q3	38.8 (37.6 to 40.1)	–16.3 (–16.6 to –16.0)	<.001
Q4	33.3 (32.1 to 34.5)	–21.8 (–22.2 to –21.5)	<.001
Q5	30.4 (29.3 to 31.5)	–24.7 (–25.1 to –24.3)	<.001
Q6	27.2 (26.1 to 28.3)	–27.9 (–28.4 to –27.4)	<.001



Quarter	Weekly use instances (95% CI)	Absolute change vs Q1 (95% CI)	*P* value
Q1	5.12 (4.81 to 5.43)	N/A	N/A
Q2	3.46 (3.25 to 3.67)	–1.66 (–1.82 to –1.51)	<.001
Q3	2.81 (2.64 to 2.98)	–2.31 (–2.51 to –2.12)	<.001
Q4	2.44 (2.29 to 2.59)	–2.68 (–2.90 to –2.45)	<.001
Q5	2.26 (2.12 to 2.40)	–2.86 (–3.10 to –2.62)	<.001
Q6	2.23 (2.09 to 2.37)	–2.89 (–3.14 to –2.65)	<.001

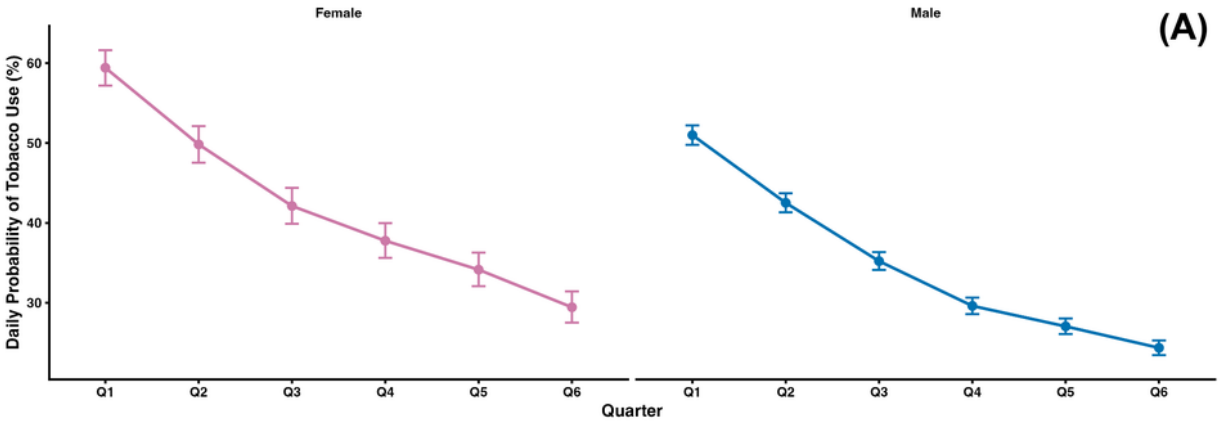
Females and Middle-Aged Adults Show Greater Relative Declines in Tobacco Use

The decline in use probability from Q1 to Q6 differed by biological sex (β=–.08 on the log-odds scale, 95% CI –0.12 to –0.05; *P*<.001), indicating a greater relative decline among females. From Q1 to Q6, decreases were 30.0 percentage points in females (95% CI –30.9 to –29.1) and 26.6 percentage points in males (95% CI –27.2 to –26.1; Figure 2A). Trends were similar in sensitivity analyses of tobacco use instances (Figure 2B), with decreases of 3.07 instances per week in females (95% CI –3.52 to –2.60) and 2.86 instances per week in males (95% CI –3.09 to –2.63).

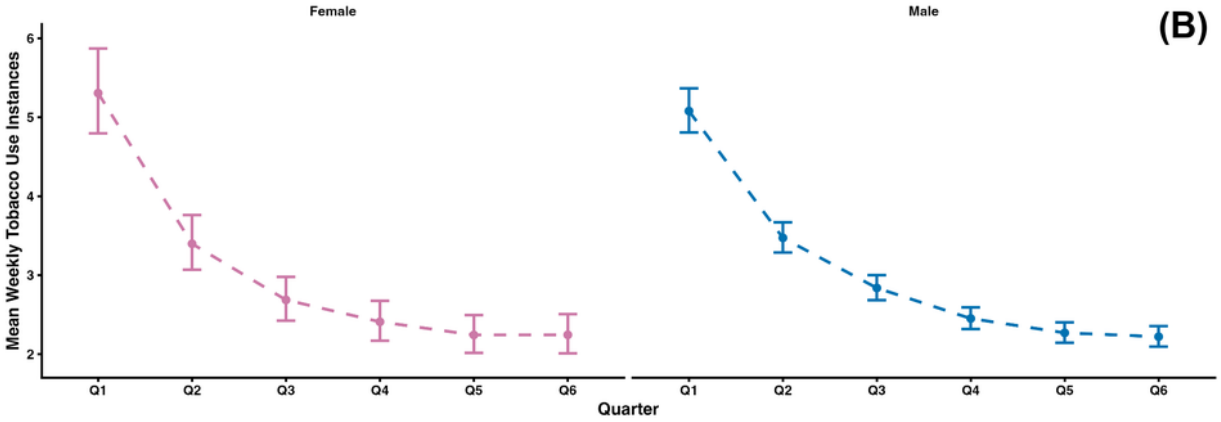
Trajectories of tobacco use probabilities varied by age group (Figure 3A). Across follow-up, all age groups demonstrated declines in tobacco use probability, although the decline from Q1 to Q6 differed by age group (*P*s≤.001). Middle-aged adults showed greater relative declines from Q1 to Q6 than both younger adults (β=–.13 on the log-odds scale, 95% CI –0.17 to –0.09; *P*<.001) and older adults (β=–.19, 95% CI –0.32 to –0.07; *P*=.001), with no significant difference between younger and older adults (*P*=.49). From Q1 to Q6, the model-estimated daily probability of tobacco use decreased by 31.0 percentage points (95% CI –31.8 to –30.1) in the middle-aged group, compared with 27.1 percentage points (95% CI –27.6 to –26.5) and 26.8 percentage points (95% CI –29.9 to –23.7) in the youngest and oldest age groups, respectively. Decreases in weekly tobacco use instances were also observed across age groups (Figure 3B).

XSL•FO
RenderX

Figure 2. Adjusted tobacco use over 72 weeks, stratified by biological sex. (A) Adjusted daily probability of tobacco use. (B) Adjusted mean weekly tobacco use instances. Marginal estimates from generalized linear mixed-effects models with a quarter × sex interaction are shown for 6 sequential 12-week quarters (Q1-Q6), separately for males (blue) and females (pink). Points represent model-based estimates, and error bars indicate 95% CIs. Absolute changes from Q1 and corresponding Dunnett-adjusted *P* values are based on within-sex model contrasts. The quarter-by-sex interaction was assessed by comparing the Q6 versus Q1 contrast between sexes on the model scale (log odds for panel A and log instances for panel B). Panel A is based on a binomial model estimating the probability of a tobacco-use day, whereas Panel B is based on a negative binomial model estimating weekly tobacco use instances. All models included quarter, sex, their interaction, age at activation, season, proportion of weekend days, and total weeks contributed as fixed effects, with a random intercept for each participant. N/A: not applicable.



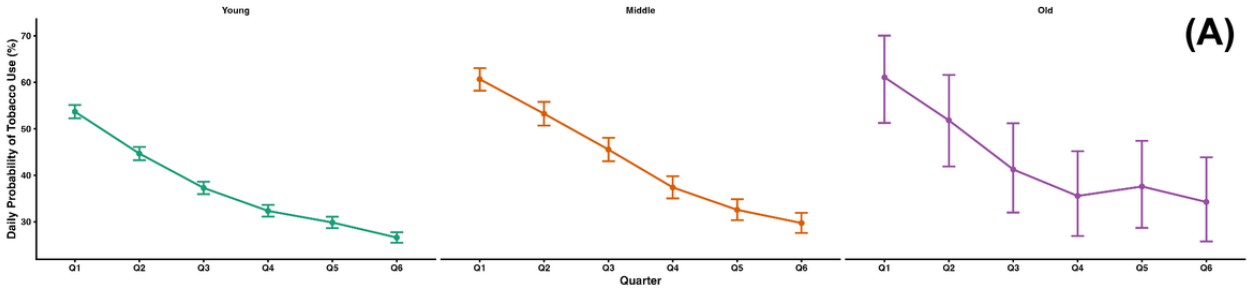
	Females			Males		
Quarter	Daily probability (%) (95% CI)	Absolute change vs Q1 (pp, 95% CI)	*P* value	Daily probability (%) (95% CI)	Absolute change vs Q1 (pp, 95% CI)	*P* value
Q1	59.4 (57.2 to 61.6)	N/A	N/A	51.0 (49.8 to 52.2)	N/A	N/A
Q2	49.8 (47.5 to 52.1)	–9.6 (–10.2 to –9.0)	<.001	42.5 (41.3 to 43.7)	–8.5 (–8.8 to –8.2)	<.001
Q3	42.1 (39.9 to 44.4)	–17.3 (–17.9 to –16.7)	<.001	35.2 (34.1 to 36.4)	–15.8 (–16.1 to –15.4)	<.001
Q4	37.8 (35.6 to 40.0)	–21.6 (–22.3 to –21.0)	<.001	29.6 (28.6 to 30.7)	–21.4 (–21.8 to –20.9)	<.001
Q5	34.1 (32.1 to 36.3)	–25.3 (–26.0 to –24.5)	<.001	27.0 (26.1 to 28.0)	–24.0 (–24.4 to –23.5)	<.001
Q6	29.4 (27.5 to 31.4)	–30.0 (–30.9 to –29.1)	<.001	24.4 (23.5 to 25.3)	–26.6 (–27.2 to –26.1)	<.001



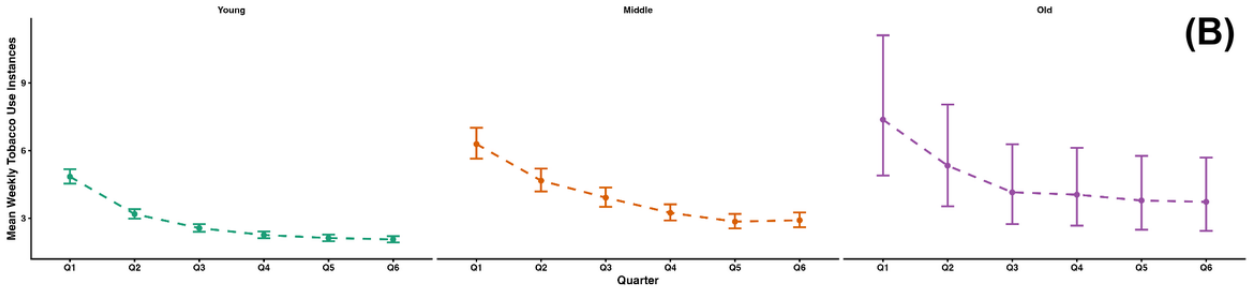
	Females			Males		
Quarter	Weekly use instances (95% CI)	Absolute change vs Q1 (95% CI)	*P* value	Weekly use instances (95% CI)	Absolute change vs Q1 (95% CI)	*P* value
Q1	5.31 (4.79 to 5.87)	N/A	N/A	5.08 (4.81 to 5.36)	N/A	N/A
Q2	3.40 (3.07 to 3.76)	–1.91 (–2.23 to –1.59)	<.001	3.47 (3.29 to 3.67)	–1.61 (–1.76 to –1.46)	<.001
Q3	2.69 (2.42 to 2.98)	–2.62 (–3.01 to –2.23)	<.001	2.84 (2.69 to 3.00)	–2.24 (–2.43 to –2.05)	<.001
Q4	2.41 (2.17 to 2.67)	–2.90 (–3.32 to –2.47)	<.001	2.45 (2.32 to 2.59)	–2.63 (–2.84 to –2.42)	<.001
Q5	2.24 (2.01 to 2.49)	–3.07 (–3.51 to –2.61)	<.001	2.27 (2.14 to 2.40)	–2.81 (–3.03 to –2.59)	<.001
Q6	2.24 (2.01 to 2.51)	–3.07 (–3.52 to –2.60)	<.001	2.22 (2.10 to 2.35)	–2.86 (–3.09 to –2.63)	<.001

XSL•FO
RenderX

Figure 3. Adjusted tobacco use over 72 weeks, stratified by age group. (A) Adjusted daily probability of tobacco use. (B) Adjusted mean weekly tobacco use instances. Marginal estimates from generalized linear mixed-effects models with a quarter × age-group interaction are shown for 6 sequential 12-week quarters (Q1-Q6), separately for younger (18-39 years; green), middle-aged (40-59 years; orange), and older (60-79 years; purple) adults. Points represent model-based estimates, and error bars indicate 95% CIs. Absolute changes from Q1 and corresponding Dunnett-adjusted *P* values are based on within-age-group model contrasts. The quarter × age-group interaction was assessed by comparing the Q6 versus Q1 contrast across age groups on the model scale (log odds for panel A and log instances for panel B). Panel A is based on a binomial model estimating the probability of a tobacco-use day, whereas panel B is based on a negative binomial model estimating weekly tobacco use instances. All models included quarter, age group, their interaction, sex, season, proportion of weekend days, and total weeks contributed as fixed effects, with a random intercept for each participant.



Quarter	Young: Daily probability (%) (95% CI)	Young: Absolute change vs Q1 (pp, 95% CI)	Young: *P* value	Middle: Daily probability (%) (95% CI)	Middle: Absolute change vs Q1 (pp, 95% CI)	Middle: *P* value	Old: Daily probability (%) (95% CI)	Old: Absolute change vs Q1 (pp, 95% CI)	Old: *P* value
Q1	53.7 (52.3 to 55.1)	N/A	N/A	60.7 (58.2 to 63.1)	N/A	N/A	61.1 (51.3 to 70.1)	N/A	N/A
Q2	44.7 (43.3 to 46.1)	–9.0 (–9.3 to –8.7)	<.001	53.3 (50.7 to 55.8)	–7.4 (–8.1 to –6.8)	<.001	51.8 (41.9 to 61.6)	–9.3 (–11.7 to –6.8)	<.001
Q3	37.3 (36.0 to 38.6)	–16.4 (–16.7 to –16.1)	<.001	45.5 (43.0 to 48.1)	–15.2 (–15.8 to –14.4)	<.001	41.3 (32.0 to 51.2)	–19.8 (–22.4 to –17.3)	<.001
Q4	32.4 (31.1 to 33.6)	–21.3 (–21.7 to –20.9)	<.001	37.4 (35.0 to 39.8)	–23.3 (–24.0 to –22.6)	<.001	35.6 (27.0 to 45.2)	–25.5 (–28.2 to –22.9)	<.001
Q5	29.9 (28.7 to 31.1)	–23.8 (–24.3 to –23.4)	<.001	32.6 (30.4 to 34.9)	–28.1 (–28.9 to –27.3)	<.001	37.6 (28.7 to 47.4)	–23.5 (–26.3 to –20.7)	<.001
Q6	26.6 (25.5 to 27.8)	–27.1 (–27.6 to –26.5)	<.001	29.7 (27.6 to 31.9)	–31.0 (–31.8 to –30.1)	<.001	34.3 (25.8 to 43.9)	–26.8 (–29.9 to –23.7)	<.001



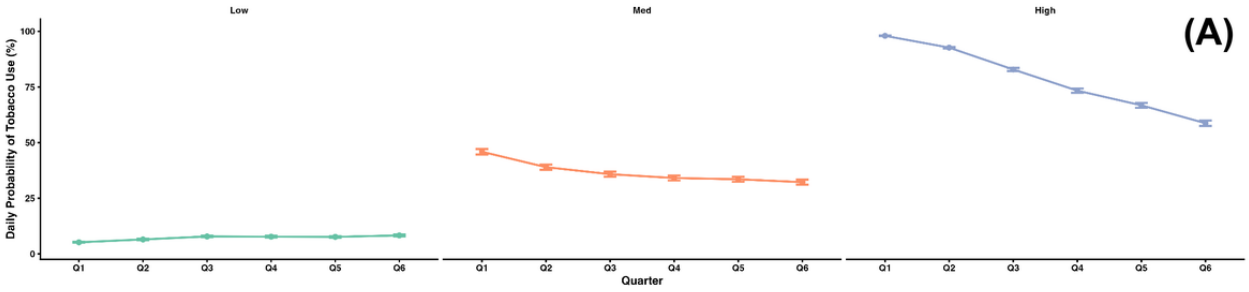
Quarter	Young: Weekly use instances (95% CI)	Young: Absolute change vs Q1 (95% CI)	Young: *P* value	Middle: Weekly use instances (95% CI)	Middle: Absolute change vs Q1 (95% CI)	Middle: *P* value	Old: Weekly use instances (95% CI)	Old: Absolute change vs Q1 (95% CI)	Old: *P* value
Q1	4.8 (4.5 to 5.2)	N/A	N/A	6.3 (5.6 to 7.0)	N/A	N/A	7.4 (4.9 to 11.1)	N/A	N/A
Q2	3.2 (3.0 to 3.4)	–1.6 (–1.8 to –1.5)	<.001	4.7 (4.2 to 5.2)	–1.6 (–2.0 to –1.3)	<.001	5.3 (3.5 to 8.0)	–2.1 (–3.6 to –0.5)	.004
Q3	2.6 (2.4 to 2.7)	–2.2 (–2.5 to –2.1)	<.001	3.9 (3.5 to 4.4)	–2.4 (–2.8 to –2.0)	<.001	4.2 (2.7 to 6.3)	–3.2 (–5.2 to –1.2)	<.001
Q4	2.3 (2.1 to 2.4)	–2.5 (–2.8 to –2.4)	<.001	3.2 (2.9 to 3.6)	–3.1 (–3.5 to –2.6)	<.001	4.0 (2.7 to 6.1)	–3.4 (–5.4 to –1.3)	<.001
Q5	2.1 (2.0 to 2.3)	–2.7 (–3.0 to –2.5)	<.001	2.9 (2.6 to 3.2)	–3.4 (–4.0 to –2.9)	<.001	3.8 (2.5 to 5.8)	–3.6 (–5.8 to –1.4)	<.001
Q6	2.1 (1.9 to 2.2)	–2.7 (–3.0 to –2.5)	<.001	2.9 (2.6 to 3.3)	–3.4 (–3.9 to –2.9)	<.001	3.7 (2.4 to 5.7)	–3.7 (–5.9 to –1.4)	<.001

High Baseline Tobacco Users Show the Largest Declines

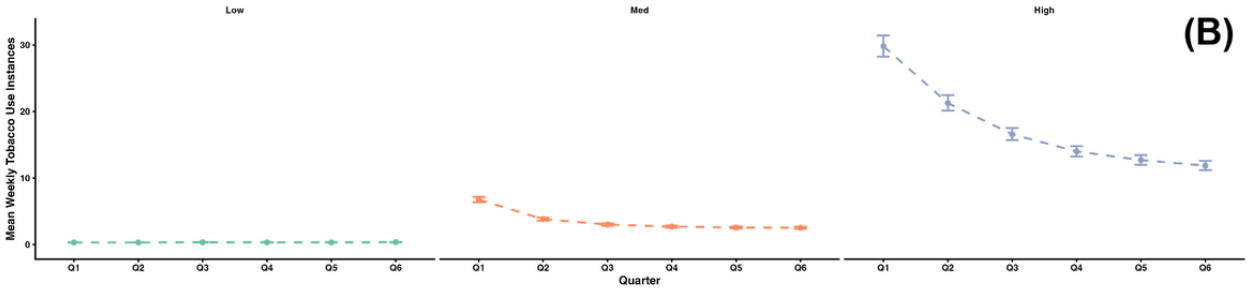
Longitudinal tobacco use trajectories differed by baseline tobacco use frequency (*P*<.001; Figure 4A), with progressively larger relative reductions in use across the low-, medium-, and high-use tertiles. Relative declines were greater in the medium-use tertile than in the low-use tertile (β=–1.08 on the log-odds scale, 95% CI –1.12 to –1.04) and greater in the high-use tertile than in the medium-use tertile (β=–2.96, 95% CI –2.99 to –2.92). From Q1 to Q6, the model-estimated daily probability of tobacco use increased by 3.1 percentage points (95% CI 2.8-3.4) in the low-use tertile but decreased by 13.7 percentage points (95% CI –14.2 to –13.1) and 39.3 percentage points (95% CI –40.8 to –37.8) in the medium- and high-use tertiles, respectively. Comparable trends were observed for the model-estimated mean weekly total number of use instances (Figure 4B).

XSL•FO
RenderX

Figure 4. Adjusted tobacco use over 72 weeks, stratified by baseline use tertile. (A) Adjusted daily probability of tobacco use. (B) Adjusted mean weekly tobacco use instances. Marginal estimates from generalized linear mixed-effects models with a quarter × tertile interaction are shown for 6 sequential 12-week quarters (Q1-Q6), separately for participants in the low (teal), medium (orange), and high (purple) tertiles. Points represent model-based estimates, and error bars indicate 95% CIs. Absolute changes from Q1 and corresponding Dunnett-adjusted *P* values are based on within-tertile model contrasts. The quarter × tertile interaction was assessed by comparing the Q6 versus Q1 contrast across tertiles on the model scale (log odds for panel A and log instances for panel B). Panel A is based on a binomial model estimating the probability of a tobacco-use day, with tertiles defined by the observed Q1 tobacco use proportion. Panel B is based on a negative binomial model estimating weekly tobacco use instances, with tertiles defined by the observed Q1 tobacco use instances. All models included quarter, tertile, their interaction, sex, age at activation, season, proportion of weekend days, and total weeks contributed as fixed effects, with a random intercept for each participant. N/A: not applicable.



	Low			Medium			High		
Quarter	Daily probability (%) (95% CI)	Absolute change vs Q1 (pp, 95% CI)	*P* value	Daily probability (%) (95% CI)	Absolute change vs Q1 (pp, 95% CI)	*P* value	Daily probability (%) (95% CI)	Absolute change vs Q1 (pp, 95% CI)	*P* value
Q1	5.2 (5.0 to 5.5)	N/A	N/A	45.9 (44.7 to 47.1)	N/A	N/A	98.0 (97.9 to 98.1)	N/A	N/A
Q2	6.5 (6.2 to 6.8)	1.3 (1.1 to 1.5)	<.001	38.9 (37.8 to 40.1)	−7.0 (−7.3 to −6.6)	<.001	92.7 (92.4 to 93.0)	−5.3 (−5.6 to −5.0)	<.001
Q3	7.9 (7.5 to 8.2)	2.7 (2.4 to 2.9)	<.001	35.8 (34.7 to 37.0)	−10.1 (−10.5 to −9.7)	<.001	82.9 (82.2 to 83.6)	−15.1 (−15.9 to −14.4)	<.001
Q4	7.7 (7.4 to 8.1)	2.5 (2.3 to 2.8)	<.001	34.1 (33.0 to 35.2)	−11.8 (−12.2 to −11.3)	<.001	73.4 (72.4 to 74.3)	−24.6 (−25.8 to −23.5)	<.001
Q5	7.6 (7.3 to 8.0)	2.4 (2.2 to 2.7)	<.001	33.5 (32.4 to 34.7)	−12.4 (−12.9 to −11.9)	<.001	66.8 (65.7 to 67.9)	−31.2 (−32.5 to −29.9)	<.001
Q6	8.3 (7.9 to 8.7)	3.1 (2.8 to 3.4)	<.001	32.2 (31.1 to 33.4)	−13.7 (−14.2 to −13.1)	<.001	58.7 (57.5 to 59.9)	−39.3 (−40.8 to −37.8)	<.001



	Low			Medium			High		
Quarter	Weekly use instances (95% CI)	Absolute change vs Q1 (95% CI)	*P* value	Weekly use instances (95% CI)	Absolute change vs Q1 (95% CI)	*P* value	Weekly use instances (95% CI)	Absolute change vs Q1 (95% CI)	*P* value
Q1	0.32 (0.30 to 0.34)	N/A	N/A	6.76 (6.38 to 7.16)	N/A	N/A	29.81 (28.25 to 31.45)	N/A	N/A
Q2	0.31 (0.29 to 0.33)	−0.01 (−0.02 to 0.01)	.880	3.84 (3.62 to 4.07)	−2.92 (−3.20 to −2.64)	<.001	21.28 (20.15 to 22.46)	−8.53 (−9.43 to −7.63)	<.001
Q3	0.35 (0.33 to 0.38)	0.03 (0.01 to 0.05)	<.001	3.01 (2.83 to 3.19)	−3.75 (−4.07 to −3.43)	<.001	16.58 (15.69 to 17.52)	−13.23 (−14.35 to −12.10)	<.001
Q4	0.33 (0.31 to 0.35)	0.01 (−0.01 to 0.04)	.255	2.71 (2.55 to 2.88)	−4.05 (−4.39 to −3.71)	<.001	14.01 (13.24 to 14.81)	−15.80 (−17.06 to −14.54)	<.001
Q5	0.33 (0.31 to 0.35)	0.01 (−0.01 to 0.03)	.505	2.55 (2.40 to 2.72)	−4.21 (−4.56 to −3.85)	<.001	12.71 (11.99 to 13.46)	−17.10 (−18.44 to −15.76)	<.001
Q6	0.37 (0.34 to 0.40)	0.05 (0.03 to 0.08)	<.001	2.53 (2.37 to 2.71)	−4.23 (−4.59 to −3.86)	<.001	11.87 (11.18 to 12.60)	−17.94 (−19.33 to −16.53)	<.001

Nongoal Setters Show Greater Relative Declines in Tobacco Use

A total of 1587 members set at least 1 tobacco-related goal during follow-up. Compared with members who did not set goals (n=11,091), goal-setting members were similar in age (mean 32.1, SD 8.8 years vs mean 32.4, SD 8.9 years; *P*=.13) but had lower BMI (mean 25.4, SD 4.1 kg/m^2 vs mean 25.8, SD 4.3 kg/m^2; *P*=.003). A greater proportion of goal-setting members were female (371/1587, 23.38% vs 2307/11,091, 20.80%; *P*=.02). The decline from Q1 to Q6 differed by goal-setting status (β=.20 on the log-odds scale; *P*<.001), indicating different longitudinal trajectories between groups. From Q1 to Q6, declines were 25.0 percentage points in goal-setting members (95% CI –25.9 to –24.0) and 28.2 percentage points in members who did not set goals (95% CI –28.8 to –27.7; Figure S4A in Multimedia Appendix 2). Declines in tobacco use instances were also observed among goal-setting and non-goal-setting members (Figure S4B in Multimedia Appendix 2).

Over One-Quarter of Participants Report Nonuse by Q6

Among participants contributing at least 3 weeks of data in both Q1 and Q6, slightly over a quarter (1404/4975, 28.22%) reported sustained nonuse by the end of follow-up, defined as zero tobacco use days during Q6. To assess potential attrition bias, baseline Q1 characteristics (tobacco use, logging frequency, sex, age, BMI, RHR, HRV, RR, and sleep duration) were compared between participants retained through Q6 and those lost to follow-up. Effect sizes were consistent with

XSL•FO
RenderX

trivial-to-small differences (Cohen *d* range=0.02-0.22; Table S5 in Multimedia Appendix 2) [16].

The prevalence of sustained nonuse was similar by biological sex, occurring in 270 of 981 (27.5%) females and 1134 of 3994 (28.4%) males. However, participants in the lowest tertile of tobacco use probability exhibited the highest prevalence of self-reported nonuse (658/1552, 42.4%), followed by the medium- (386/1712, 22.5%) and high-use tertiles (360/1711, 21%).

In a sensitivity analysis restricted to participants with complete daily tobacco use reporting across all 12 weeks of both Q1 and Q6, the prevalence of sustained nonuse was similar (408/1623, 25.1%).

Logging Engagement Is Associated With Larger Reductions in Tobacco Use

In a complementary engagement analysis, we examined whether the proportion of days with completed tobacco logging was associated with a percentage-point change in tobacco use from Q1 to Q6. Heavier baseline (Q1) tobacco users had higher logging engagement (β=15.0; 95% CI 12.0-18.0). After adjustment for baseline tobacco use, total observed study days, age, biological sex, and BMI, logging engagement was independently associated with larger reductions in tobacco use (Figure S5 in Multimedia Appendix 2); each 10-percentage-point increase in engagement corresponded to a 0.92-percentage-point greater decline from Q1 to Q6 (95% CI –1.58 to –0.26).

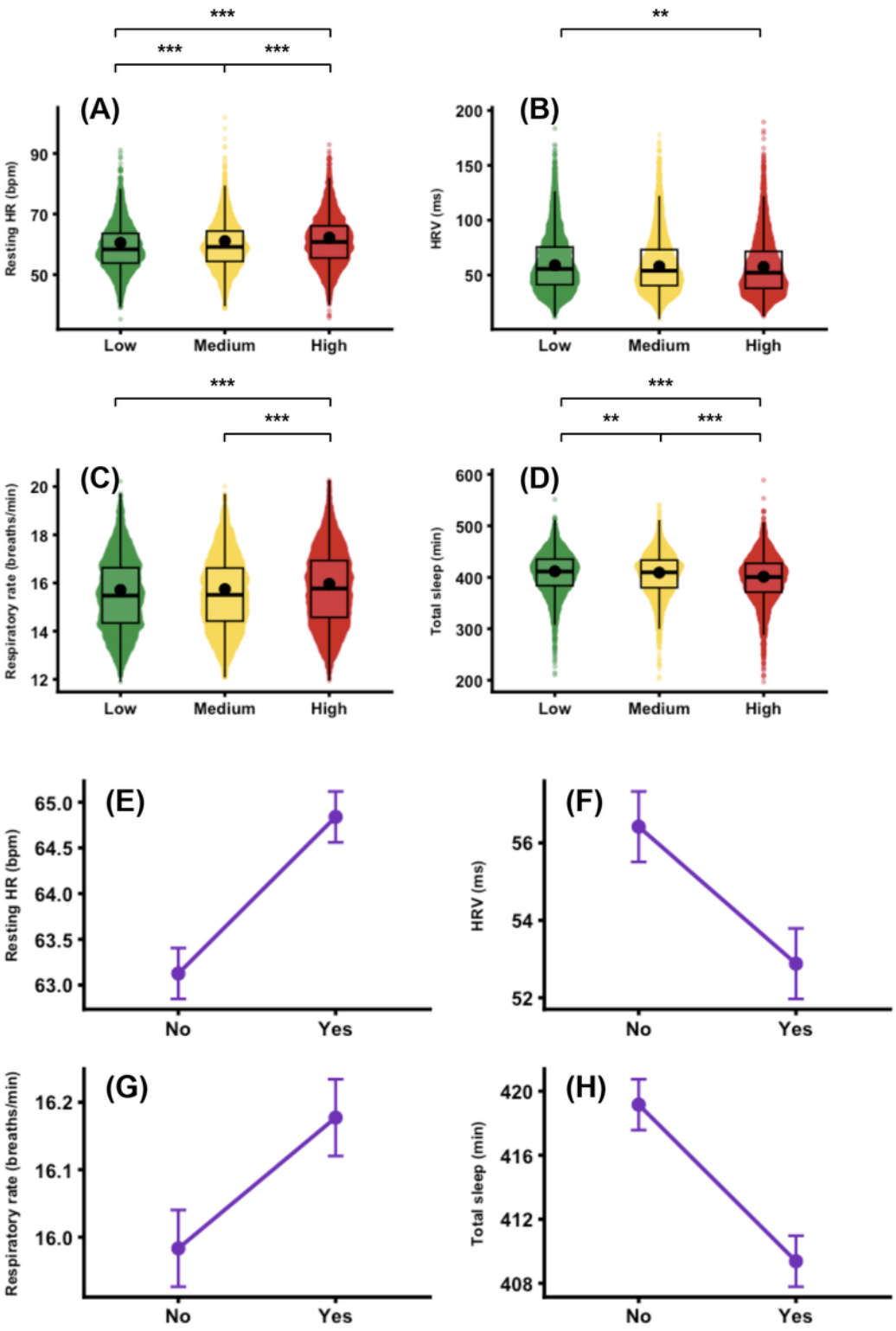
Higher Tobacco Use Is Associated With Higher RHR, Lower HRV, Higher RR, and Shorter Sleep

In models that decomposed tobacco use into between- and within-person components, tobacco use was consistently associated with worse biometric and sleep measures. At the between-person level (Figure S6 in Multimedia Appendix 2), continuous models indicated that higher tobacco use was associated with increased RHR (0.22 beats/minute per 10 percentage-point increase in tobacco use days; 95% CI 0.18-0.26), reduced HRV (–0.19 ms per 10 percentage-point increase in tobacco use days; 95% CI –0.31 to –0.07), increased RR (0.03 breaths/minute per 10 percentage-point increase in tobacco use days; 95% CI 0.03-0.04), and reduced total sleep duration (–1.39 minutes per 10 percentage-point increase in tobacco use days; 95% CI –1.60 to –1.18). Consistent with these continuous associations, categorical models comparing tobacco use tertiles demonstrated significant differences in biometric and sleep measures (Figure 5A-D). Relative to the low-use tertile, individuals in the high-use tertile exhibited a 1.75 beats/minute higher RHR (95% CI 1.43-2.06, *P*<.001), a 1.58 ms lower HRV (95% CI –2.59 to –0.58, *P*=.006), a 0.25 breaths/minute higher RR (95% CI 0.19-0.31, *P*<.001), and a 10.55-minute shorter total sleep duration (95% CI –12.28 to –8.81, *P*<.001).

Within individuals, days with reported tobacco use were associated with acute adverse changes in overnight physiology and sleep measures compared with tobacco-free days (Figure 5E-H). On nights following tobacco use days, RHR was 1.71 beats/minute higher (95% CI 1.70-1.73), HRV was 3.54 ms lower (95% CI –3.59 to –3.49), RR was 0.19 breaths/minute higher (95% CI 0.19-0.20), and total sleep duration was 9.78 minutes shorter (95% CI –10.08 to –9.49) relative to nonuse days (*P*s<.001). Exploratory analyses suggested that these reductions in sleep duration were largely accompanied by later sleep onset, with tobacco use days associated with a 13.5-minute later bedtime (95% CI 13.0-14.0) but only a modest increase in the percentage of time in bed spent awake (0.20%; 95% CI 0.19-0.22).

XSL•FO
RenderX

Figure 5. Associations of tobacco use with physiological outcomes. (A-D) Between-person associations of tobacco use frequency with physiological outcomes, stratified by overall tobacco use tertile. (E-H) Within-person associations of tobacco use on a given day with concurrent physiological outcomes. Outcomes included resting heart rate (HR; beats/minute), heart rate variability (HRV; ms), respiratory rate (breaths/minute), and total sleep duration (minutes), each modeled separately using linear mixed-effects models with a random intercept for participant and fixed effects for tobacco variables, biological sex, age at activation, BMI, season, and weekday versus weekend. In panels A-D, individual data points (quasi-random jitter) represent each participant's mean outcome value and are colored by tobacco use tertile: low (green), medium (yellow), and high (red). Boxplots summarize the distribution within each tertile. Black points with error bars represent estimated marginal means and 95% CIs; pairwise comparisons were adjusted using the Tukey method. In panels E-H, points represent model-based estimates of population-average outcome values for days without tobacco use (no) and days with tobacco use (yes), with continuous covariates held at their sample means and categorical covariates at their modal values. Error bars indicate 95% CIs. pp: percentage point. ***P*<.01, ****P*<.001.



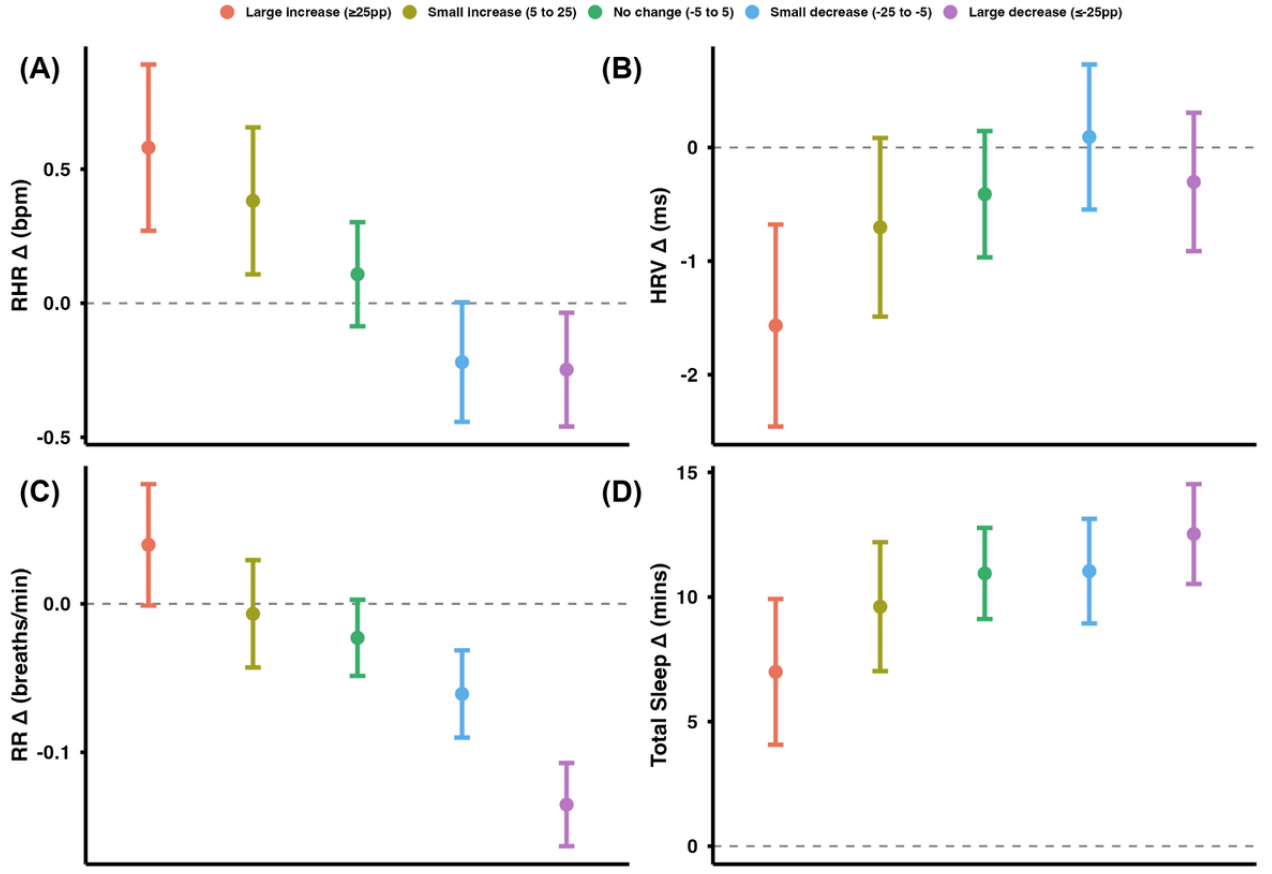
XSL•FO
RenderX

Reductions in Tobacco Use Are Associated With Favorable Changes in Biometrics and Sleep

Reductions in tobacco use between Q1 and Q6 were associated with directionally favorable changes in biometric and sleep measures. In continuous models (Figure S7 in Multimedia Appendix 2), reductions in tobacco use were associated with lower RHR, higher HRV, lower RR, and longer sleep duration. Categorical models reinforced these findings. For RHR (Figure 6A), participants with large decreases in tobacco use demonstrated a 0.25 beats/minute decrease (95% CI –0.46 to –0.04), which differed significantly from the changes observed among those with small (0.38 beats/minute; 95% CI 0.11-0.66, *P*=.001) or large increases (0.58 beats/minute; 95% CI 0.27-0.89, *P*<.001) in tobacco use. For HRV (Figure 6B), participants with large increases in tobacco use experienced HRV reductions of 1.57 ms (95% CI –2.46 to –0.68) that differed from the HRV changes observed among participants with small decreases in tobacco use (0.09 ms; 95% CI –0.54 to 0.73, between-group *P*=.01).

Regarding RR (Figure 6C), participants with large decreases in tobacco use demonstrated a 0.14 breaths/minute reduction (95% CI –0.16 to –0.11), which was significantly greater in magnitude than the reductions observed in all other groups (*P*<.001 vs the large-increase, small-increase, and no-change groups; *P*=.002 vs the small-decrease group). Total sleep duration increased across all groups (Figure 6D), but larger increases were observed among participants with large reductions in tobacco use (12.5 minutes; 95% CI 10.5-14.5) than among those with large increases in tobacco use (7.00 minutes; 95% CI 4.07-9.92; between-group *P*=.02). To investigate trajectories over time, visualizations of quarter-by-quarter tobacco-change-group interactions for each outcome were also constructed (Figure S8 in Multimedia Appendix 2).

Figure 6. Association between change in tobacco use frequency and change in physiological outcomes from Q1 to Q6. Model-based estimated marginal means (points) and 95% CIs (error bars) are shown for changes in each outcome across 5 categories of tobacco use change: large decrease (≤−25 percentage points [pp]), small decrease (−25 to −5 pp), no change (−5 to +5 pp), small increase (+5 to +25 pp), and large increase (≥+25 pp). Dashed horizontal lines indicate no change. Outcomes include resting heart rate (beats/minute; A), heart rate variability (ms; B), respiratory rate (breaths/minute; C), and total sleep duration (minute; D). Each outcome was modeled separately. Models included baseline tobacco use frequency, baseline outcome value, age, biological sex, BMI, and season as covariates. Pairwise comparisons between tobacco-use change categories were adjusted using the Tukey method.



Discussion

Principal Findings

In this 72-week study of more than 12,000 wearable device adopters, we examined longitudinal trajectories of tobacco use and their associations with cardiopulmonary and sleep-related biomarkers. Two principal findings emerged. First, daily tobacco use probability and total use instances declined over time at the population level, and more than one-quarter of participants who reported tobacco use at baseline achieved sustained nonuse by the end of follow-up. Second, tobacco use was consistently

XSL•FO
RenderX

associated with less favorable cardiopulmonary and sleep-related biomarkers at both the acute, within-person level (ie, days of use vs nonuse) and the habitual, between-person level (differences in overall use patterns). Importantly, reductions in tobacco use over time were accompanied by directionally favorable shifts in RHR, RR, and sleep duration.

Across the 72-week observation period, tobacco use probability declined substantially at the population level, with reductions robust to multiple sensitivity analyses, including IPCW and restriction to complete reporting. Prior wearable-based approaches to tobacco reduction have focused on real-time detection of smoking behavior using event-triggered feedback [17], with a recent 8-week randomized controlled trial reporting a reduction of 10 cigarettes per day and 11% of participants achieving 7-day biochemically confirmed abstinence [18]. By contrast, our study did not involve real-time tobacco use detection. Instead, participants engaged in active self-monitoring through daily tobacco use entries, a behavior independently shown to promote reductions in tobacco use [19], alongside continuous passive feedback on physiological responses to use. Such feedback on sleep and recovery may surface discrepancies between an individual's physiological state and desired state, consistent with a self-regulation framework; however, this study cannot establish whether, or to what extent, these elements contributed to behavior change. Furthermore, over a substantially longer follow-up period, we observed a reduction of approximately 3 tobacco use instances per week and a substantial proportion of participants achieving sustained nonuse (1404/4975, 28.22%), although fundamental differences in study design, outcome definitions, and follow-up duration preclude direct comparison. Nonetheless, these findings are consistent with the possibility that physiological self-monitoring may complement existing pathways to tobacco reduction. Future work evaluating the combination of real-time use detection and continuous physiological feedback is warranted.

Although declines in tobacco use were detected in all demographic subgroups, they were greater in females than in males, and middle-aged adults showed greater relative reductions than younger and older adults. Women and younger adults are more likely to enroll in digital tobacco cessation programs [20], yet women have historically demonstrated lower quit rates than men despite greater tobacco-related disease risk [21,22]. The reasons for sex differences remain unclear but do not appear to be fully explained by demographic, socioeconomic, or addiction-related factors [23], making the greater relative decline among females in this study notable. Although our observational design and modest number of female participants limit mechanistic interpretation, one plausible contributor is that women's smoking behavior is more strongly reinforced by nonnicotine, behavioral, and sensory cues than men's [24], such that behavioral feedback environments like the one studied here may be disproportionately associated with reductions in this group. Regarding age, it has been posited that differences in digital health engagement may contribute to differences in intervention efficacy. However, while younger adults are generally more likely to use technology, older adults are increasingly adopting digital health–related tools [25], with some evidence suggesting comparable outcomes across age groups [26]. Rather, we observed the greatest relative declines in tobacco use among middle-aged adults, in line with prior evidence from a systematic review and meta-analysis that used a cruder framework comparing individuals aged <40 and ≥40 years [11]. The factors driving greater reductions specifically among middle-aged adults represent an important area for future inquiry.

Declines in tobacco use were most pronounced among participants with the highest baseline use, whereas those in the low-use tertile showed modest increases over time. This pattern may reflect greater motivation for reduction among heavier users and is consistent with pharmacotherapy trials showing larger relative treatment effects among individuals with higher baseline nicotine dependence [27]. However, these observations should be interpreted cautiously, given the potential for regression to the mean [28].

A counterintuitive finding was that members who did not set tobacco-related goals exhibited greater relative reductions in tobacco use than those who did, which contrasts with the broader behavior change literature on goal setting. Notably, a similar pattern was observed in our recent companion analysis of alcohol use trajectories among WHOOP members [7]. Although the mechanisms underlying this finding cannot be resolved by our observational design, several nonmutually exclusive explanations warrant consideration. First, goal-setting members had a substantially higher baseline tobacco use probability, which may reflect greater nicotine dependence or longer use histories that could render reductions more difficult to achieve (ie, the hardening hypothesis) [29]. Second, members may self-select into goal setting precisely because they anticipate difficulty reducing use, such that group differences may reflect underlying addiction severity rather than an effect of goal setting, per se. Third, our binary indicator of goal setting captured any goal set at any point during the 72-week observation window. As goals could be set, modified, or abandoned over time, treating goal setting as a time-invariant exposure may introduce time-related bias. We accordingly interpret this exploratory finding as hypothesis-generating and suggest that future work examine goal type, timing, and persistence as time-varying exposures.

Participants with larger reductions in tobacco use exhibited directionally favorable changes in cardiopulmonary and sleep-related biomarkers, consistent with known acute and chronic physiological effects of tobacco exposure. Acute tobacco exposure reduces vagal-cardiac activity [30,31], reflected by increased RHR and reduced HRV, and nicotine exposure disrupts central nervous system regulation of sleep [32]. Meanwhile, real-world monitoring studies have shown lower RHR during periods of smoking abstinence [33]. These findings align with our within-person results, in which day-to-day tobacco use was associated with unfavorable changes in cardiopulmonary indices and shorter sleep duration. Notably, while associations between tobacco use reductions and population-level favorable shifts in cardiopulmonary and sleep-related measures were modest, changes were graded according to the degree of tobacco reduction. This dose-response pattern supports the interpretation that the observed associations are biologically meaningful rather than statistical artifacts of a

XSL•FO
RenderX

large sample. Moreover, because tobacco use is a chronic, repeated exposure, modest differences may compound over time and yield clinically relevant cumulative effects, particularly given that even small increases in RHR have been linked to elevated cardiovascular and all-cause mortality in large epidemiological cohorts [34].

At the between-person level, habitual tobacco use has been associated with higher RHR, lower HRV, and poorer sleep outcomes, with evidence of dose-response associations based on use frequency and intensity [35-37]. Consistent with these cross-sectional associations, we observed that reductions in tobacco use were associated with modest but statistically significant improvements in RHR, RR, and sleep duration. The relationship between tobacco use and HRV was more nuanced. Participants with large increases in tobacco use exhibited significant HRV reductions relative to those with small decreases in use, but HRV did not significantly improve among participants who reduced their tobacco use. Prior cessation studies suggest that HRV improvements may occur transiently following smoking cessation, peaking within the first week and attenuating thereafter [38]. Longer-term autonomic adaptations may also be influenced by concurrent metabolic changes following cessation, such as weight gain and reductions in insulin sensitivity that have been reported after tobacco cessation [39], which could counteract sustained HRV improvements. These findings suggest that cardiac autonomic function may follow a different temporal trajectory than other cardiopulmonary and sleep-related measures following reductions in tobacco use.

Implications

Our study has several potential implications for the broader mobile health (mHealth) community. First, this study highlights how wearables can combine data from free-living populations with long observation windows, complementing work conducted in more controlled environments. Second, these findings suggest that further investigation of the combined effects of continuous health measurement and behavioral logging in tobacco cessation programs may be particularly valuable. Third, the association between logging engagement and reductions in tobacco use suggests that future evaluations may benefit from including logging engagement as a covariate. Fourth, to support broader generalizability, it will be important to replicate these findings outside of a single commercial environment.

Limitations

Several limitations should be considered when interpreting these findings. First, tobacco use was self-reported and not biochemically verified, raising the possibility of recall error, underreporting, and social desirability bias, the degree of which cannot be quantified from the available data. However, the graded dose-response pattern between self-reported tobacco reduction and physiological change, with the largest reported reductions corresponding to the greatest favorable shifts in RHR, RR, and sleep duration, provides objective corroboration that these self-reports reflect meaningful behavior change. Second, participants were subscription-based wearable users and may have been more health conscious or had greater financial resources than the general population. The additional requirement for sustained reporting engagement (≥8 weeks) may have further enriched the sample with more motivated individuals, limiting generalizability. Relatedly, we lacked information on participants' tobacco use histories and concurrent use of cessation supports such as nicotine replacement therapy, pharmacotherapy, or behavioral counseling, any of which may have contributed to the observed reductions independent of the wearable platform. Third, tobacco exposure was assessed as daily use probability and the number of weekly use instances, without the ability to distinguish among tobacco products (eg, cigarettes, cigars, chewing tobacco), which may have heterogeneous physiological effects. Fourth, we were unable to account for a range of co-occurring behavioral and health factors that may independently influence cardiopulmonary physiology and sleep, including alcohol intake, caffeine consumption, physical activity, acute illness, medication use, and psychological stress. Fifth, without a control group, observed changes cannot be attributed causally to wearable adoption or self-monitoring and may reflect broader secular trends [40]. Although the 72-week follow-up represents a relatively long observation period compared with prior digital health studies, longer durations will be needed to evaluate the persistence of observed changes. Additionally, outcomes were limited to wearable-derived measures of RHR, HRV, RR, and sleep; future studies incorporating clinical end points such as blood pressure, body weight, or blood biomarkers would provide greater insight into the health implications of tobacco reduction. Finally, the study was conducted within a single wearable ecosystem (WHOOP), and several authors hold financial interests in the company. While analytic and editorial decisions were firewalled from commercial functions (see the "Conflicts of Interest" section), readers should consider this context when interpreting the findings. Nonetheless, these limitations are offset by several strengths, including the large sample size, real-world setting, extended follow-up, and high-frequency longitudinal assessment of tobacco use alongside continuous physiological monitoring.

Conclusions

In this 72-week observational study of more than 12,000 wearable device users, self-reported tobacco use declined substantially over time, with greater reductions among females, middle-aged adults, and individuals with medium-to-high baseline use. More than one-quarter of baseline tobacco users with data available in Q6 reported sustained nonuse by the end of follow-up. Tobacco use was consistently associated with less favorable cardiopulmonary and sleep-related measures, and reductions in use were accompanied by modest but directionally favorable changes in RHR, RR, and sleep duration. Although these findings cannot establish causality, they suggest that sustained reductions in tobacco use co-occur with favorable shifts in objective physiological measures over time. However, the observed physiological changes may also reflect broader lifestyle or health changes. Together, these results highlight the potential relevance of wearable technologies and digital health platforms as settings in which tobacco use reductions and associated physiological differences are observed.

XSL•FO
RenderX

Acknowledgments

The authors acknowledge the technical assistance of the WHOOP Data Science and Data Engineering teams for their contributions to feature development and data infrastructure. We also thank the broader WHOOP employee community for its continued support of the company's research initiatives. The authors used the generative artificial intelligence tool ChatGPT (OpenAI) for language editing.

Funding

This work was supported by WHOOP, Inc through salary support provided to authors DJC, JL, WvH, FF, CJC, DMP, RJ, KEH, and GJG. No specific grant numbers are associated with this work. WHOOP, Inc provided general institutional support but had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

The data and code supporting the findings of this study are not publicly available because of the intellectual property interests of WHOOP, Inc. Deidentified participant data and a data dictionary may be made available upon reasonable request to WHOOP (research@whoop.com). Access will require submission of a methodologically sound proposal approved by WHOOP's research team and execution of a signed data use agreement. WHOOP will respond to requests within 4 weeks.

Authors' Contributions

Conceptualization: DJC, JL, WvH, FF, CJC, JGC, DMP, RJ, KEH, GJG
Data curation: DJC, GJG
Formal analysis: DJC (lead), GJG (lead), JL (supporting)
Project administration: KEH, GJG
Visualization: DJC, GJG
Writing – original draft: DJC, GJG
Writing – reviewing & editing: DJC, JL, WvH, FF, CJC, JGC, DMP, RJ, KEH, GJG

Conflicts of Interest

The study was funded by WHOOP, Inc. DJC, JL, WvH, FF, CJC, JGC, DMP, RJ, KEH, and GJG are employees of WHOOP, Inc, and JL, FF, CJC, DMP, RJ, KEH, and GJG hold stock options as part of their employee compensation, representing a potential financial interest. All analyses were performed solely by the research and data science teams. Marketing, enterprise, and other nonresearch groups at WHOOP, Inc had no role in study approval, data interpretation, manuscript preparation, or decisions regarding submission or publication.

Multimedia Appendix 1
Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist.
[DOC File , 90 KB - mhealth_v14i1e98116_app1.doc]

Multimedia Appendix 2
Engagement characteristics and sensitivity analyses with supporting tables (n=5) and figures (n=8).
[DOCX File , 4850 KB - mhealth_v14i1e98116_app2.docx]

References

1. Roy A, Rawal I, Jabbour S, Prabhakaran D. Chapter 4: Tobacco and cardiovascular disease: a summary of evidence. In: Prabhakaran D, Anand S, Gaziano TA, Mbanya JC, Wu Y, Nugent R, editors. Cardiovascular, Respiratory, and Related Disorders, 3rd Edition. Washington, DC: The International Bank for Reconstruction and Development/The World Bank; Nov 17, 2017:e1.
2. Vallée A. Associations between tobacco smoking and mortality: a sex-stratified cohort analysis. Eur J Public Health 2025 Dec 01;35(6):1212-1218 [FREE Full text] [doi: 10.1093/eurpub/ckaf194] [Medline: 41134686]
3. He H, Pan Z, Wu J, Hu C, Bai L, Lyu J. Health effects of tobacco at the global, regional, and national levels: results from the 2019 Global Burden of Disease Study. Nicotine Tob Res 2022 Apr 28;24(6):864-870. [doi: 10.1093/ntr/ntab265] [Medline: 34928373]
4. Roh S. Scientific evidence for the addictiveness of tobacco and smoking cessation in tobacco litigation. J Prev Med Public Health 2018 Jan;51(1):1-5 [FREE Full text] [doi: 10.3961/jpmph.16.088] [Medline: 29397646]

XSL•FO
RenderX

5. Onwuzo CN, Olukorode J, Sange W, Orimoloye DA, Udojike C, Omoragbon L, et al. A review of smoking cessation interventions: efficacy, strategies for implementation, and future directions. Cureus 2024 Jan;16(1):e52102 [FREE Full text] [doi: 10.7759/cureus.52102] [Medline: 38344627]
6. Grosicki GJ, Fielding F, Kim J, Chapman CJ, Olaru M, Hippel WV, et al. Wearing WHOOP more frequently is associated with better biometrics and healthier sleep and activity patterns. Sensors (Basel) 2025 Apr 12;25(8):s25082437 [FREE Full text] [doi: 10.3390/s25082437] [Medline: 40285124]
7. Grosicki GJ, Hippel WV, Fielding F, Chapman CJ, Presby DM, Leota J, et al. Alcohol use trajectories during the first 72 weeks of WHOOP wearable platform membership: observational cohort study. JMIR Mhealth Uhealth 2026 Apr 27;14:e91288-e91288 [FREE Full text] [doi: 10.2196/91288] [Medline: 42044360]
8. Scott CK, Dennis ML, Gustafson DH. Using smartphones to decrease substance use via self-monitoring and recovery support: study protocol for a randomized control trial. Trials 2017 Aug 10;18(1):374 [FREE Full text] [doi: 10.1186/s13063-017-2096-z] [Medline: 28797307]
9. Carver CS, Scheier MF. Control theory: a useful conceptual framework for personality–social, clinical, and health psychology. Psychological Bulletin 1982;92(1):111-135. [doi: 10.1037/0033-2909.92.1.111]
10. Wasnik M, Dave B, Vadher V. Effectiveness of digital intervention for tobacco cessation among adults: a systematic review. Addict Health 2025 Jan;17:1566. [doi: 10.34172/ahj.1566] [Medline: 40458272]
11. Li S, Li Y, Xu C, Tao S, Sun H, Yang J, et al. Efficacy of digital interventions for smoking cessation by type and method: a systematic review and network meta-analysis. Nat Hum Behav 2025 Oct 02;9(10):2054-2065. [doi: 10.1038/s41562-025-02295-2] [Medline: 40897803]
12. Grosicki GJ, Kim J, Fielding F, Jasinski SR, Chapman C, Hippel WV, et al. Heart and health behavior responses to GLP-1 receptor agonists: a 12-wk study using wearable technology and causal inference. Am J Physiol Heart Circ Physiol 2025 Feb 01;328(2):H235-H241. [doi: 10.1152/ajpheart.00809.2024] [Medline: 39705534]
13. Miller DJ, Lastella M, Scanlan AT, Bellenger C, Halson SL, Roach GD, et al. A validation study of the WHOOP strap against polysomnography to assess sleep. J Sports Sci 2020 Nov;38(22):2631-2636. [doi: 10.1080/02640414.2020.1797448] [Medline: 32713257]
14. Miller DJ, Sargent C, Roach GD. A validation of six wearable devices for estimating sleep, heart rate and heart rate variability in healthy adults. Sensors (Basel) 2022 Aug 22;22(16):s22166317 [FREE Full text] [doi: 10.3390/s22166317] [Medline: 36016077]
15. Rubin D. Multiple Imputation: Flexible Imputation of Missing Data, Second Edition. Boca Raton, FL: Chapman and Hall/CRC; 2018:29.
16. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013 Nov 26;4:863. [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
17. Imtiaz MH, Ramos-Garcia RI, Wattal S, Tiffany S, Sazonov E. Wearable sensors for monitoring of cigarette smoking in free-living: a systematic review. Sensors (Basel) 2019 Oct 28;19(21):s19214678 [FREE Full text] [doi: 10.3390/s19214678] [Medline: 31661856]
18. Bold KW, Mestre LM, Garrison KA, Gueorguieva R, O'Malley SS, Fucito LM. Feasibility and efficacy of a real-time smoking intervention using wearable technology. PLOS Digit Health 2025 Nov;4(11):e0001086 [FREE Full text] [doi: 10.1371/journal.pdig.0001086] [Medline: 41212887]
19. McFall RM. Effects of self-monitoring on normal smoking behavior. J Consult Clin Psychol 1970 Oct;35(2):135-142. [doi: 10.1037/h0030087] [Medline: 5474283]
20. Fradkin N, Zbikowski SM, Christensen T. Analysis of demographic characteristics of users of a free tobacco cessation smartphone app: observational study. JMIR Public Health Surveill 2022 Mar 09;8(3):e32499 [FREE Full text] [doi: 10.2196/32499] [Medline: 35262491]
21. Bohadana A, Nilsson F, Rasmussen T, Martinet Y. Gender differences in quit rates following smoking cessation with combination nicotine therapy: influence of baseline smoking behavior. Nicotine Tob Res 2003 Feb;5(1):111-116. [doi: 10.1080/1462220021000060482] [Medline: 12745512]
22. Ward KD, Klesges RC, Zbikowski SM, Bliss RE, Garvey AJ. Gender differences in the outcome of an unaided smoking cessation attempt. Addict Behav 1997;22(4):521-533. [doi: 10.1016/s0306-4603(96)00063-9] [Medline: 9290861]
23. Sheikhattari P, Barsha RAA, Assari S. Gender differences in quit rates in a tobacco cessation program: in search of demographic, socioeconomic, health, or behavioral explanatory mechanisms. J Biomed Life Sci 2025;5(1):6017. [doi: 10.31586/jbls.2025.6017] [Medline: 40264471]
24. Perkins KA, Donny E, Caggiula AR. Sex differences in nicotine effects and self-administration: review of human and animal evidence. Nicotine Tob Res 1999 Dec;1(4):301-315. [doi: 10.1080/14622299050011431] [Medline: 11072427]
25. Fahey MC, Huang S, Dahne J. Technology use for health-related purposes among US older adults (65+ years). J Am Geriatr Soc 2024 Dec;72(12):3913-3916. [doi: 10.1111/jgs.19122] [Medline: 39158124]
26. Kwon DM, Santiago-Torres M, Mull KE, Sullivan BM, Bricker JB. Older adults who smoke: do they engage with and benefit from web-based smoking cessation interventions? Prev Med 2022 Aug;161:107118. [doi: 10.1016/j.ypmed.2022.107118] [Medline: 35718119]

XSL•FO
RenderX

27. Thomas KH, Dalili MN, López-López JA, Keeney E, Phillippo DM, Munafò MR, et al. Comparative clinical effectiveness and safety of tobacco cessation pharmacotherapies and electronic cigarettes: a systematic review and network meta-analysis of randomized controlled trials. Addiction 2022 Apr;117(4):861-876 [FREE Full text] [doi: 10.1111/add.15675] [Medline: 34636108]
28. Linden A. Assessing regression to the mean effects in health care initiatives. BMC Med Res Methodol 2013 Sep 28;13:119 [FREE Full text] [doi: 10.1186/1471-2288-13-119] [Medline: 24073634]
29. Hughes JR. The hardening hypothesis: is the ability to quit decreasing due to increasing nicotine dependence? A review and commentary. Drug Alcohol Depend 2011 Sep 01;117(2-3):111-117 [FREE Full text] [doi: 10.1016/j.drugalcdep.2011.02.009] [Medline: 21411244]
30. Niedermaier ON, Smith ML, Beightol LA, Zukowska-Grojec Z, Goldstein DS, Eckberg DL, et al. Influence of cigarette smoking on human autonomic function. Circulation 1993 Aug;88(2):562-571. [doi: 10.1161/01.cir.88.2.562] [Medline: 8339419]
31. Wolk R, Shamsuzzaman ASM, Svatikova A, Huyber CM, Huck C, Narkiewicz K, et al. Hemodynamic and autonomic effects of smokeless tobacco in healthy young men. J Am Coll Cardiol 2005 Mar 15;45(6):910-914 [FREE Full text] [doi: 10.1016/j.jacc.2004.11.056] [Medline: 15766828]
32. Singh N, Wanjari A, Sinha AH. Effects of nicotine on the central nervous system and sleep quality in relation to other stimulants: a narrative review. Cureus 2023 Nov;15(11):e49162 [FREE Full text] [doi: 10.7759/cureus.49162] [Medline: 38130519]
33. Herbec A, Parker E, Ubhi HK, Raupach T, West R. Decrease in resting heart rate measured using smartphone apps to verify abstinence from smoking: an exploratory study. Nicotine Tob Res 2020 Jul 16;22(8):1424-1427 [FREE Full text] [doi: 10.1093/ntr/ntaa021] [Medline: 31971595]
34. Aune D, Sen A, ó'Hartaigh B, Janszky I, Romundstad P, Tonstad S, et al. Resting heart rate and the risk of cardiovascular disease, total cancer, and all-cause mortality - a systematic review and dose-response meta-analysis of prospective studies. Nutr Metab Cardiovasc Dis 2017 Jun;27(6):504-517 [FREE Full text] [doi: 10.1016/j.numecd.2017.04.004] [Medline: 28552551]
35. Papathanasiou G, Georgakopoulos D, Papageorgiou E, Zerva E, Michalis L, Kalfakakou V, et al. Effects of smoking on heart rate at rest and during exercise, and on heart rate recovery, in young adults. Hellenic J Cardiol 2013;54(3):168-177. [Medline: 23685653]
36. Murgia F, Melotti R, Foco L, Gögele M, Meraviglia V, Motta B, et al. Effects of smoking status, history and intensity on heart rate variability in the general population: the CHRIS study. PLoS One 2019;14(4):e0215053 [FREE Full text] [doi: 10.1371/journal.pone.0215053] [Medline: 30964923]
37. Li W, Kalan ME, Kondracki AJ, Gautam P, Jebai R, Erinoso O, et al. Poor sleep duration and e-cigarette/cigarette use among US adults with cardiovascular diseases: findings from the 2022 BRFSS cross-sectional study. Sleep Breath 2024 Dec;28(6):2701-2710. [doi: 10.1007/s11325-024-03140-y] [Medline: 39192028]
38. Yotsukura M, Koide Y, Fujii K, Tomono Y, Katayama A, Ando H, et al. Heart rate variability during the first month of smoking cessation. Am Heart J 1998 Jun;135(6 Pt 1):1004-1009. [doi: 10.1016/s0002-8703(98)70065-1] [Medline: 9630104]
39. Bush T, Lovejoy JC, Deprey M, Carpenter KM. The effect of tobacco cessation on weight gain, obesity, and diabetes risk. Obesity (Silver Spring) 2016 Sep 29;24(9):1834-1841 [FREE Full text] [doi: 10.1002/oby.21582] [Medline: 27569117]
40. Meza R, Cao P, Jeon J, Warner KE, Levy DT. Trends in US adult smoking prevalence, 2011 to 2022. JAMA Health Forum 2023 Dec 01;4(12):e234213 [FREE Full text] [doi: 10.1001/jamahealthforum.2023.4213] [Medline: 38038988]

Abbreviations

GLMM: generalized linear mixed-effects model
HRV: heart rate variability
IPCW: inverse probability of censoring weighting
mHealth: mobile health
Q: quarter
RHR: resting heart rate
RR: respiratory rate
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
SWS: slow-wave sleep

XSL•FO
RenderX

Edited by A Stone; submitted 15.Apr.2026; peer-reviewed by C Pinkerton, T Ishinuki, E Hurwitz, WJ He; comments to author 11.May.2026; revised version received 04.Jun.2026; accepted 23.Jun.2026; published 03.Jul.2026.

Please cite as:
Curran DJ, Leota J, Hippel WV, Fielding F, Chapman CJ, Cohen JG, Presby DM, Jhunjhunwala R, Holmes KE, Grosicki GJ
Tobacco Use Trajectories and Associated Changes in Biometrics and Sleep During the First 72 Weeks of Wearable Membership: Observational Cohort Study
JMIR Mhealth Uhealth 2026;14:e98116
URL: https://mhealth.jmir.org/2026/1/e98116
doi:10.2196/98116
PMID:42336399

©Dylan J Curran, Josh Leota, William von Hippel, Finnbarr Fielding, Christopher J Chapman, Jenna G Cohen, David M Presby, Reha Jhunjhunwala, Kristen E Holmes, Gregory J Grosicki. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 03.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Week-Ahead Prediction of High-Risk Drinking Episodes Among Young Adults Using Wearable Biosignals and Psychological Vulnerabilities: Prospective Observational Machine Learning Study

Jae Seok Kwak[1*], PhD; Hae Kook Lee[2*], MD, PhD; Sun-Jin Jo[3*], PhD; Jun Hyuk Kwon[4*], BE; Sun Jung Kwon[5*], PhD; Yena Kim[5*], PhD; Haejung Lee[6*], MSc

[1]Department of Psychiatry, College of Medicine, The Catholic University of Korea, Seocho-gu, Seoul, Republic of Korea

[2]Department of Psychiatry, The Catholic University of Korea Uijeongbu St. Mary's Hospital, Uijeongbu, Gyeonggi-do, Republic of Korea

[3]Department of Addiction Studies, Graduate School, The Catholic University of Korea, Bucheon-si, Gyeonggi-do, Republic of Korea

[4]Department of Industrial Engineering, Kongju National University, Cheonan, Chungcheongnam-do, Republic of Korea

[5]Department of Counseling Psychology, Korea Baptist Theological University and Seminary, Yuseong-gu, Daejeon, Republic of Korea

[6]AI Healthcare Division, PCN, Gangnam-gu, Seoul, Republic of Korea

[*]all authors contributed equally

Corresponding Author:
Hae Kook Lee, MD, PhD
Department of Psychiatry
The Catholic University of Korea Uijeongbu St. Mary's Hospital
65-1, Geumo-Dong
Uijeongbu, Gyeonggi-do, 11765
Republic of Korea
Phone: 82 318203050
Email: nplhk@catholic.ac.kr

Abstract

Background: Although machine learning has increasingly been used to predict mental health symptoms and maladaptive behaviors, real-world prediction of addiction-related risk remains limited. Emotional and temperamental vulnerabilities are established correlates of alcohol-related problems, yet few studies have integrated these factors with wearable-derived biosignals in alcohol-risk prediction models.

Objective: This study evaluated whether machine learning models could predict weekly high-risk drinking episodes among young adults with elevated alcohol-use risk by integrating wearable-derived health data with baseline emotional and personality vulnerability indicators.

Methods: In this prospective observational study, adults in their 20s completed weekly self-report surveys and wore Fitbit devices for 4 weeks. Features from week *t* were used to predict the Alcohol Use Disorders Identification Test-Korean version (AUDIT-K)–based high-risk drinking label at week *t*+1. Positive labels were defined using AUDIT-K high-risk drinking cutoffs, with scores of ≥20 for men and ≥10 for women. Extreme gradient boosting (XGBoost) and random forest models were evaluated across self-report-only, wearable-only, and integrated feature sets using 5-fold participant-level grouped cross-validation.

Results: A total of 206 participants contributed 620 week-level observations, of which 85 (13.7%) were labeled as positive high-risk drinking episodes. In participant-level grouped cross-validation, the integrated random forest model showed the most favorable sensitivity-oriented performance, with a mean accuracy of 0.617 (SD 0.078), recall/sensitivity of 0.653 (SD 0.144), area under the receiver operating characteristic curve (ROC AUC) of 0.681 (SD 0.079), and area under the precision-recall curve (PR AUC) of 0.255 (SD 0.090). The integrated XGBoost model achieved an accuracy of 0.670 (SD 0.089), recall/sensitivity of 0.399 (SD 0.174), ROC AUC of 0.651 (SD 0.089), and PR AUC of 0.228 (SD 0.074). Shapley additive explanations analyses indicated that both baseline vulnerability indicators and wearable-derived weekly summaries contributed to model predictions.

Conclusions: Integrating baseline emotional and personality vulnerability indicators with wearable-derived weekly health signals may provide useful information for week-ahead prediction of high-risk drinking episodes. These findings provide preliminary support for wearable-assisted alcohol-risk stratification, although the modest positive predictive performance indicates

XSL•FO
RenderX

that external validation and more proximal within-person measures are needed before real-world early-warning or just-in-time adaptive intervention applications.

(JMIR Mhealth Uhealth 2026;14:e88223) doi:10.2196/88223

KEYWORDS

digital phenotyping; high-risk drinking; machine learning; risk prediction; wearable devices

Introduction

Background

According to the 2021 National Mental Health Survey conducted by the Ministry of Health and Welfare, the 12-month prevalence of alcohol use disorder among Korean adults was 2.6%, yet only 2.6% of those individuals reported receiving any treatment during the past year [1]. In addition, data from the Korea Community Health Survey [2] showed that monthly alcohol consumption increased from 53.7% in 2021 to 58.0% in 2023, a rise of 4.3 percentage points. The prevalence of high-risk drinking also increased from 11.0% in 2021 to 13.2% in 2023. Rising rates of alcohol use and high-risk drinking heighten the likelihood of early alcohol misuse and chronic alcohol dependence, underscoring an urgent need for targeted public-health strategies.

The COVID-19 pandemic further exacerbated alcohol-related problems. According to the 2020 Alcohol Industry Information Survey [3], the incidence of home drinking (72.2%) and drinking alone (74.9%) increased substantially. A twin study of adults also indicated that pandemic-related social restrictions increased stress and anxiety, which in turn led to greater alcohol consumption as a coping mechanism [4]. Collectively, these findings, combined with the limited accessibility of in-person treatment during the pandemic, have intensified the emphasis on digital interventions as an alternative means of addressing alcohol-related problems [5].

Research efforts have increasingly focused on addressing alcohol-related problems by leveraging widely accessible technologies such as smartphones and wearable devices, particularly in response to low treatment use, limited-service accessibility, and poor treatment awareness among individuals with alcohol problems [6-8]. Wearable technologies offer notable advantages by continuously monitoring individuals' behaviors and physiological states in real time, enabling timely and context-sensitive interventions [9]. Moreover, wearing such devices can enhance users' sense of empowerment, motivation, and accountability [6], suggesting that wearable-supported approaches may contribute positively to reducing alcohol use and improving outcomes among individuals at risk.

A growing body of research has attempted to predict alcohol-related problems at an early stage. One representative line of work uses smartphone sensors and machine learning to detect or forecast drinking and binge-drinking episodes among young adults. Prior studies have demonstrated the feasibility of classifying and predicting high-risk drinking using passively collected indicators, such as accelerometer data, call logs, and digital device use patterns [10]. More recent work has combined extreme gradient boosting (XGBoost)–based prediction models with artificial intelligence (AI) to anticipate same-day binge drinking 1 to 6 hours in advance, with the aim of enabling just-in-time adaptive interventions (JITAIs) [11]. Early studies further established the technical validity of detecting weekend nighttime drinking or heavy drinking using smartphone sensors with high accuracy [12].

In parallel, digital marker–based screening approaches have emerged, including models that estimate alcohol risk from Instagram images and text using neural network analysis [13]. In the neuroimaging domain, multimodal classifiers integrating structural, task-based, and resting-state connectivity data have demonstrated significant accuracy in distinguishing individuals with alcohol dependence [14]. Similarly, a model combining default mode network functional connectivity, neuropsychological indicators, and impulsivity features using random forest algorithms has been reported [15]. Deep-learning classification using electroencephalography signals has also shown high accuracy even with a limited number of channels and small sample sizes, highlighting the promise of noninvasive physiological markers [16].

Taken together, research across digital sensing, neuroimaging, physiological signals, and behavioral data demonstrates a growing potential for machine learning–based early detection and prediction of alcohol-related behaviors. These approaches collectively aim to balance measurement burden, real-time monitoring capability, and generalizability, thereby advancing the development of lightweight and scalable predictive models for use in real-world settings.

Despite their well-established relevance, emotional problems and personality vulnerabilities have been insufficiently explored in AI and machine learning research on alcohol use. Systematic reviews and empirical studies consistently show that psychological factors such as depression [17,18], anxiety [17,19], and stress [20,21], as well as personality-related vulnerabilities including emotional instability [22], harm avoidance [22-25], and impulsivity [22,26,27], are among the most robust predictors of alcohol dependence and problematic drinking. Exploring whether these emotional and personality vulnerabilities interact with real-world physiological rhythms to predict high-risk drinking at an early stage represents a meaningful and necessary research avenue.

Despite the growing body of international research, machine learning studies that use smartphones or wearable devices to predict alcohol-related problems are virtually absent in South Korea. Although national statistics indicate rising prevalence and increasing rates of hazardous drinking, treatment use, and service accessibility remain critically low. Developing machine learning models capable of detecting early indicators of problematic drinking in individuals' daily lives, therefore,

XSL•FO
RenderX

represents a highly significant public-health priority. Accordingly, this study aims to develop a predictive model for high-risk drinking episodes by passively and unobtrusively collecting real-world biodata through wearable devices and integrating these signals with individuals' emotional and personality vulnerability profiles. This approach seeks to estimate the onset of high-risk alcohol use in advance, enabling timely and personalized preventive intervention.

Objectives

This study aims to develop machine learning algorithms capable of predicting the risk of emerging alcohol-related problems among high-risk drinkers using easily accessible tools such as smartphones and wearable devices. To achieve this goal, we collected real-time health data (eg, heart rate, activity, and sleep) via a wearable device and gathered self-reported emotional and personality vulnerability measures through a smartphone interface. We then compared the predictive performance of machine learning models trained on (1) self-report psychological data alone, (2) wearable-derived health data alone, and (3) an integrated feature set combining both.

This approach aligns with prior machine learning research demonstrating that predictive accuracy for mental health outcomes, such as emotional distress, stress responses, and depressive symptoms, improves when physiological sensing data and psychological-behavioral data are jointly modeled [28-30]. Accordingly, this study aimed to develop and evaluate machine learning models for week-ahead prediction of Alcohol Use Disorders Identification Test-Korean version (AUDIT-K)–defined high-risk drinking episodes by comparing self-report-only, wearable-only, and integrated feature sets.

Methods

Participants

Participants were recruited through Kiwisurvey, an online panel platform operated by Korea Research International, where an announcement for a mobile-based health behavior monitoring study was posted. Individuals who expressed interest and met the eligibility criteria were screened for inclusion. Eligible participants were Korean adults aged 19-29 years residing in the Seoul or Gyeonggi metropolitan areas, with no visual or hearing impairments. To maintain gender balance, a 1:1 sex allocation strategy was applied during recruitment. Additionally, to ensure sufficient representation of high-risk drinkers, at least 30% of the sample was required to meet the AUDIT-K threshold for hazardous drinking (≥10 for men and ≥6 for women). This recruitment criterion was distinct from the prediction outcome label, which was defined using the AUDIT-K high-risk drinking threshold.

Two offline orientation sessions were held on June 24 and July 1, 2023, and participants were asked to attend one of the sessions. During these sessions, researchers provided detailed instructions regarding study procedures, wearable device setup, and informed consent requirements. A total of 209 individuals provided written informed consent and were enrolled in the study. Of these, 206 participants completed the full 4-week monitoring period and were included in the final analyses. Two participants withdrew consent, and one was excluded due to missing data. Participants received an incentive equivalent to approximately US $100 upon completing the 4-week follow-up period.

Measures and Procedure

After providing written informed consent, participants received a baseline survey link via their personal smartphones and completed a series of self-report questionnaires. Baseline assessments included sociodemographic characteristics such as gender, age, education level, monthly personal allowance, and residential status. Participants also completed the AUDIT-K at baseline and once weekly for 4 consecutive weeks to assess alcohol-use risk.

The AUDIT-K is a validated screening instrument based on the *International Classification of Diseases, Tenth Revision* diagnostic criteria that assesses alcohol use disorder risk and associated drinking problems. The Korean version consists of 10 items measuring drinking frequency and quantity, symptoms of dependence, and alcohol-related problems [31]. Validation studies have established AUDIT-K cutoff scores of ≥10 for men and ≥6 for women to classify hazardous drinking and ≥20 for men and ≥10 for women to classify high-risk drinking. In this study, the hazardous drinking cutoff was used only for recruitment enrichment, whereas the high-risk drinking cutoff was used to define the binary prediction outcome. The measure demonstrated excellent test-retest reliability in prior validation research (r=0.929). In this study, internal consistency at baseline was acceptable (Cronbach α=0.794).

Baseline and week-4 surveys included the Patient Health Questionnaire-9 (PHQ-9), the Generalized Anxiety Disorder scale (GAD-7), and the Perceived Stress Scale-10 (PSS-10). The PHQ-9 demonstrated excellent internal consistency at the time of development (Cronbach α=0.89) [32]. A Korean validation study [33] recommended a cutoff score of 9, reporting strong reliability (Cronbach α=0.95); in this study, internal consistency was acceptable (Cronbach α=0.845). The GAD-7 similarly showed high reliability in its original development (Cronbach α=0.92) [34]. The Korean validated version [35] also demonstrated strong internal consistency (Cronbach α=0.92), and reliability in this sample was high (Cronbach α=0.906). Both the PHQ-9 and GAD-7 assess subjective symptoms experienced over the past 2 weeks using a Likert scale ranging from 0 ("not at all") to 3 ("nearly every day"). Higher total scores indicate greater severity of depressive or anxiety symptoms, respectively.

We used the Korean validated version of the scale, which measures the degree to which individuals perceive their lives as stressful and comprises 2 factors, negative perception and positive perception, rated on a 5-point Likert scale [36]. Higher total scores indicate greater perceived stress, with possible scores ranging from 0 to 40. Established interpretive ranges categorize scores of 14-16 as mild stress, 17-18 as moderate stress, and ≥19 as severe stress. The original PSS-10 demonstrated strong internal consistency (Cronbach α=0.84-0.86), and reliability in this study was also high (Cronbach α=0.815).

XSL•FO
RenderX

Personality vulnerabilities were assessed using the Personality Vulnerability Scale, a measure developed by the Korean Psychological Association [37] based on transdiagnostic approaches. The scale was designed to identify and screen for dispositional psychological risk factors, including cognitive, emotional, and behavioral processes, that interact with stressful environments to trigger or maintain various mental health problems. The original instrument consists of 58 items rated on a 4-point Likert scale (0="not at all" to 3="very much"), organized into 3 overarching domains (emotional and impulse regulation, information processing, and interpersonal/social functioning) and 8 subcomponents.

For the purposes of this study, we extracted 3 subscales known to be closely associated with risky alcohol use: emotional instability, impulsivity, and harm avoidance. Internal consistency at baseline was high for all subscales (Cronbach α=0.891 for emotional instability, 0.808 for impulsivity, and 0.870 for harm avoidance). Reliability in the current dataset also demonstrated excellent internal consistency (Cronbach α=0.870, 0.759, and 0.851, respectively).

All questionnaire data, including baseline and follow-up assessments, were collected via a smartphone-accessible web-based survey platform. Physiological and behavioral data—such as sleep, heart rate, and step count—were passively and continuously recorded in real time using a wearable device (Fitbit Charge 5; Fitbit Inc).

Statistical Analysis

Supervised classification models were applied to predict a binary outcome variable aggregated at the weekly level. The dependent variable, weekly high-risk drinking episode, was defined using the AUDIT-K high-risk drinking cutoffs, with positive cases coded as ≥20 for men and ≥10 for women. Predictor variables consisted of self-report measures (depression, anxiety, stress, and personality vulnerabilities) and wearable-derived features (summarized metrics of heart rate, activity, and sleep). Missing data were handled according to predefined rules. The full list of features included in the analysis is presented in Table 1.

For psychological predictors in the primary models, only baseline scores of emotional and personality vulnerability indicators were included. These included depression, anxiety, perceived stress, emotional instability, impulsivity, and harm avoidance, and were treated as baseline vulnerability indicators rather than time-varying weekly predictors. Accordingly, these variables were interpreted as between-person vulnerability indicators, whereas week-to-week variation in the prediction task was primarily captured by wearable-derived weekly summaries.

Preliminary exploratory analyses examined whether adding repeated weekly self-report symptom scores improved week-ahead prediction. These weekly self-report scores referred to repeated PHQ-9, GAD-7, and PSS-10 assessments collected during follow-up and paired with the subsequent week's high-risk drinking label when available. These repeated weekly self-report models were treated as exploratory analyses and were not retained as the final primary modeling approach because they differed from the primary models in feature definition, available feature-label pairs, missing-data structure, and conceptual interpretation. In the primary models, self-report psychological and personality measures were conceptualized as baseline vulnerability indicators rather than rapidly changing weekly predictors. This decision was based on the theoretical nature and assessment time frames of the PHQ-9, GAD-7, PSS-10, and personality vulnerability measures, which are intended to capture symptom severity or relatively stable vulnerability profiles over specified recent periods rather than short-term week-to-week fluctuations. Therefore, the primary modeling framework fixed these self-report vulnerability indicators at baseline and focused on whether week-varying wearable-derived summaries could improve the prediction of the following week's high-risk drinking label. Because these exploratory models differed from the primary models in feature definition, available feature-label pairs, and missing-data structure, their performance estimates should not be directly compared with the primary model results reported in the Results section. Details are provided in Multimedia Appendix 1.

Wearable-derived features were constructed through a multistep aggregation process. Raw Fitbit-derived heart rate, activity, calorie expenditure, distance, and sleep data were first processed at their available sampling resolution and transformed into minute-level, hour-binned, daily, or sleep episode–level summaries, depending on the type of signal. These intermediate summaries were then aggregated to the weekly level to create the final model features. Thus, variables such as steps_mean and calories_mean represent weekly summaries of minute-level activity or energy-expenditure metrics, whereas variables such as avg_heart_rate_1hour and total_calories represent weekly summaries of hour-binned metrics. Sleep-related variables, such as total_sleep_duration_weekly_avg and sleep_fragmentation_weekly_avg, represent weekly averages of daily or sleep-episode-level sleep summaries**.** The resulting weekly wearable-derived features were paired with baseline psychological variables to predict the high-risk drinking label at week *t*+1 from characteristics observed during week *t*.

XSL•FO
RenderX

Table 1. Variables used in machine learning models for predicting weekly high-risk drinking episodes.

Category and variable	Description
Dependent variable	
y	Weekly alcohol episode (1=yes, 0=no; AUDIT-K[a]: male ≥20, female ≥10)
Psychometric data	
Emotional instability	Total score for emotional instability at baseline
Harm avoidance	Total score for harm avoidance at baseline
Impulsivity	Total score for impulsivity at baseline
PHQ[b]_mean	Mean score for depressive items at baseline
PHQ_std	Standard deviation for depressive items at baseline
GAD[c]_mean	Mean score for anxiety items at baseline
GAD_std	Standard deviation for anxiety items at baseline
PSS[d]_mean	Mean score for stress items at baseline
PSS_std	Standard deviation for stress items at baseline
Physiological data (wearable)	
heart_rate_mean	Weekly mean heart rate
heart_rate_std	Weekly standard deviation for heart rate
heart_rate_max	Weekly maximum heart rate
heart_rate_min	Weekly minimum heart rate
steps_mean	Weekly mean for minute-level step count
steps_std	Weekly standard deviation for minute-level step count
steps_max	Weekly maximum for minute-level step count
steps_min	Weekly minimum for minute-level step count
distance_mean	Weekly mean for minute-level distance (m)
distance_std	Weekly standard deviation for minute-level distance
distance_max	Weekly maximum for minute-level distance
distance_min	Weekly minimum for minute-level distance
calories_mean	Weekly mean for minute-level calories burned
calories_std	Weekly standard deviation for minute-level calories burned
calories_max	Weekly maximum for minute-level calories burned
calories_min	Weekly minimum for minute-level calories burned
avg_heart_rate_1hour	Weekly mean for hour-binned average heart rate
std_heart_rate_1hour	Weekly mean for hour-binned heart rate standard deviation
total_steps	Weekly mean for hour-binned total steps
total_distance	Weekly mean for hour-binned total distance
total_calories	Weekly mean for hour-binned total calories burned
active_minutes	Weekly mean for hour-binned minutes with active_rate ≥2
ratio_light_sleep	Weekly mean for hour-binned light-sleep ratio
ratio_sleep	Weekly mean for hour-binned total-sleep ratio
ratio_deep_sleep	Weekly mean for hour-binned deep-sleep ratio
movement_burst_count	Weekly count for ≥3-minute activity bursts within 10-minute windows
resting_duration	Weekly summary of the longest continuous resting duration with active_rate=1 (min)
step_zero_streak	Weekly summary for continuous inactivity duration with steps=0

XSL•FO
RenderX

Category and variable	Description
total_sleep_duration_weekly_avg	Weekly average for daily total sleep duration
sleep_fragmentation_weekly_avg	Weekly average number of daily sleep-stage transitions
max_hr_spike_weekly_avg	Weekly average number of ≥10 bpm heart-rate spikes per minute
hr_variability_window_weekly_avg	Weekly average for window-based heart rate variability (SD)
nighttime_hr_avg_weekly_avg	Weekly average for mean heart rate between 00:00 and 06:00
unexpected_activity_at_night_occurred	Weekly indicator of whether active_rate ≥3 occurred between 00:00 and 06:00
nocturnal_walking_occurred	Weekly indicator of whether steps or distance >0 occurred during sleep
resting_high_calorie_occurred	Weekly indicator of whether high calorie expenditure occurred during inactivity (steps=0 and active_rate=1)

[a]AUDIT-K: Alcohol Use Disorders Identification Test-Korean version.

[b]PHQ: Patient Health Questionnaire-9.

[c]GAD: Generalized Anxiety Disorder-7.

[d]PSS: Perceived Stress Scale-10.

To clarify temporal alignment, each participant's data were organized into Monday to Sunday calendar weeks. Wearable-derived signals collected during week *t* were aggregated only after the week *t* observation window was complete. Weekly survey responses were collected following each calendar week and were used to define the alcohol-risk label for that corresponding week. Prediction was conceptualized as occurring after week *t* feature aggregation but before week *t*+1 outcome information was available; therefore, no information from the target week was used to construct predictors for that target label. Figure 1 illustrates the temporal alignment used to construct week-ahead feature-label pairs. Model evaluation was conducted separately using participant-level grouped cross-validation. The analytical task was defined as a week-ahead binary classification problem in which features aggregated at week *t* were used to predict the positive label at week *t*+1. For each participant, week indices were chronologically ordered, and (t, t+1) feature-label pairs were generated only when the label for week *t*+1 was observed. Observations from the final week, for which no subsequent label existed, were excluded from model training and evaluation (Figure 1). Wearable signals and self-report indicators were summarized from daily raw data into weekly statistics, including means, SDs, quantiles, and proportional measures (eg, proportion of time spent in specific sleep stages). Temporal context variables, such as weekday/weekend indicators and the number of valid observation days per week, were also included in the feature set. Missing data were handled according to the nature and distribution of each variable. Weeks with insufficient within-week observations were excluded entirely, whereas simple imputation or model-tolerant missingness strategies were applied when appropriate. All scaling and imputation procedures were derived exclusively from statistics calculated within each training fold (eg, fold-specific mean and SD) to strictly prevent information leakage into validation folds.

Model comparisons were conducted across three feature configurations: (1) self-report features only, (2) wearable-derived health features only, and (3) an integrated feature set combining self-report and wearable-derived health data. Two supervised classifiers, XGBoost and random forest, were implemented. The primary model evaluation framework was a 5-fold participant-level stratified grouped cross-validation. Participant ID was used as the grouping variable, ensuring that all week-level observations from the same participant were assigned to the same fold and did not appear simultaneously in both the training and held-out folds. Label stratification was applied as far as possible to maintain a comparable distribution of positive high-risk drinking labels across folds. The cross-validation split was generated using a fixed random seed of 42, and fixed random seeds were also applied to stochastic model-training procedures to support reproducibility. This procedure was used to prevent subject-level information leakage arising from repeated weekly observations within individuals.

Within each cross-validation iteration, the held-out fold was used only for final performance estimation. All preprocessing steps, including missing-data handling, scaling where applicable, and any feature transformation, were fitted exclusively on the training portion of that fold and then applied to the held-out fold. Class imbalance was addressed within the training portion of each fold using algorithm-specific class weighting rather than oversampling or undersampling. For XGBoost, the positive-class weight was adjusted based on the ratio of negative to positive labels within the training portion of each fold. For random forest, balanced class weights were applied. No oversampling, undersampling, synthetic minority oversampling technique, or synthetic data generation was applied to the held-out fold. Hyperparameter tuning was conducted within the training portion of each fold using internal validation, and the held-out fold was not used for model selection or threshold determination. Details of the Optuna hyperparameter search spaces, number of trials, and key hyperparameter ranges are provided in Multimedia Appendix 2.

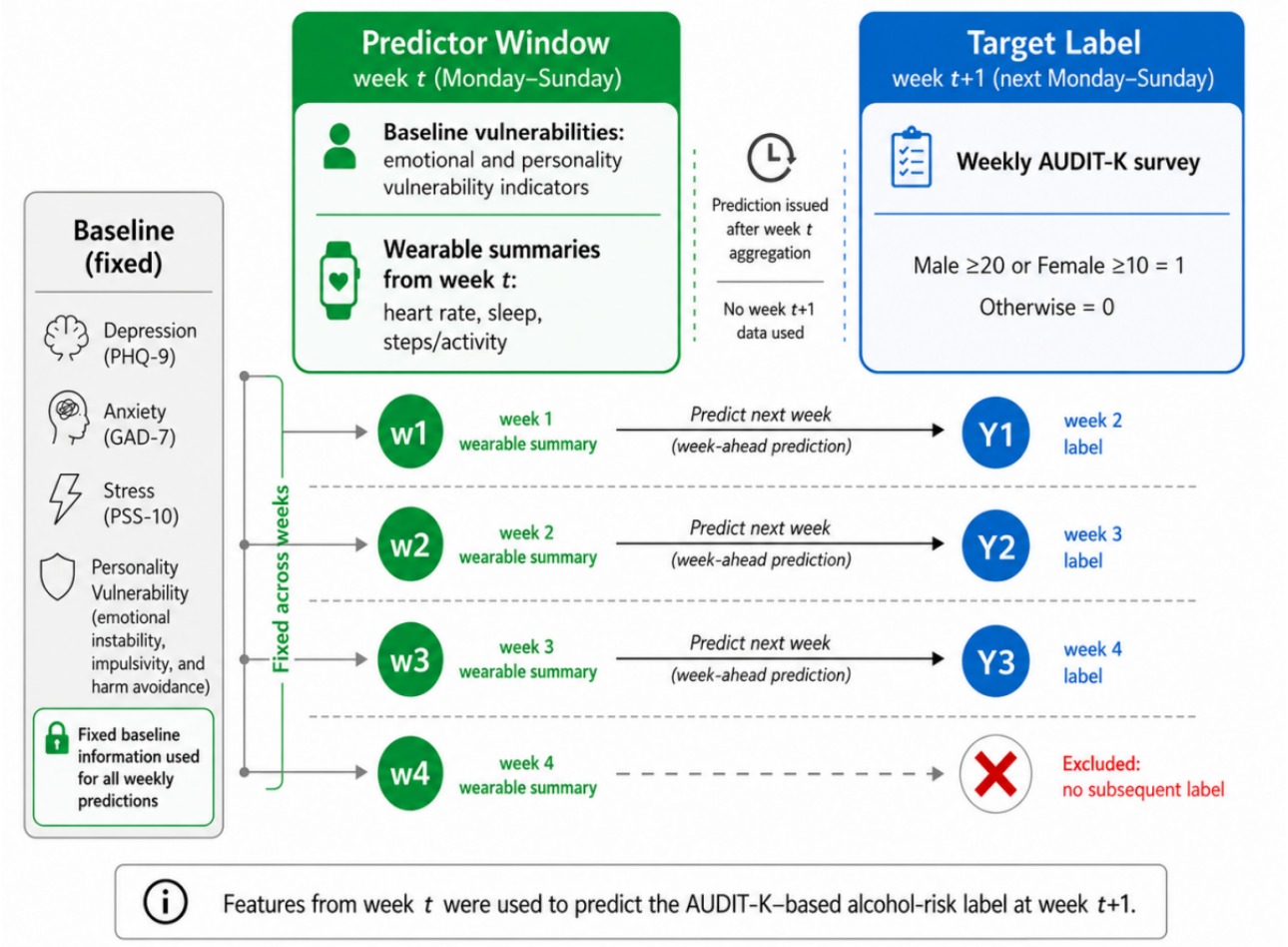
Because the intended use of the model was early detection of future high-risk drinking episodes, recall/sensitivity was prespecified as the primary performance metric. Threshold-dependent metrics were calculated using a sensitivity-oriented decision threshold of 0.20 to prioritize early detection while monitoring precision/positive predictive value (PPV). Precision/PPV, recall/sensitivity, negative predictive

XSL•FO
RenderX

value (NPV), and F_1-score were calculated at this operating point. Model discrimination was evaluated using the area under the receiver operating characteristic curve (ROC AUC) and the area under the precision-recall curve (PR AUC), which are threshold-independent metrics. Final performance estimates were reported as the mean and SD across the 5 participant-level grouped cross-validation folds. All analyses were conducted in Python using *scikit-learn*, *xgboost*, and *Optuna*.

To improve transparency and completeness of reporting, the Methods and Results sections were reviewed and revised with reference to the TRIPOD+AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis Plus Artificial Intelligence) reporting guidance for prediction model studies using machine learning methods.

Figure 1. Week-ahead prediction framework and feature-label alignment. AUDIT-K: Alcohol Use Disorders Identification Test-Korean version; GAD-7: Generalized Anxiety Disorder-7; PHQ-9: Patient Health Questionnaire-9; PSS-10: Perceived Stress Scale-10.



Ethical Considerations

All study procedures were conducted in accordance with the Declaration of Helsinki. The institutional review board of the College of Medicine, The Catholic University of Korea, approved this study (MC22OIDI0117). All participants attended an in-person orientation session, during which the study objectives and data collection procedures were fully explained, and written informed consent was obtained from all participants prior to enrollment.

To protect participant privacy and confidentiality, identifying information was stored separately from the research data. Online self-report data and wearable-derived biometric data were managed using study registration numbers after direct identifiers were removed. The linkage file connecting personal identifiers with study registration numbers was stored in encrypted form, and access was restricted to designated data-management personnel. Electronic files and computers were password-protected, and physical documents, including consent forms, were stored in double-locked storage at Korea Research International. Unauthorized access, viewing, or modification of the study data was not permitted under the study data management procedures.

Participants received a coffee voucher worth KRW 5000 (approximately US $3.24) after completing the orientation, consent forms, and baseline questionnaires. Participants who completed the study received KRW 100,000 (approximately US $64.71) as a single bank transfer. Consent for collecting bank account information for compensation payment was included in the informed consent form.

XSL•FO
RenderX

Results

Participant Characteristics

Table 2 summarizes the general characteristics of the 206 participants included in the analysis. The sample showed a balanced gender distribution, with 102 (49.5%) men and 104 (50.5%) women. Participants were primarily in their mid to late 20s: 129 (62.6%) individuals were aged 25-29 years, and 77 (37.4%) individuals were aged 19-24 years, with a mean age of 25.21 (SD 2.54) years. Regarding educational attainment, 118 (57.3%) participants had completed a university degree, 60 (29.1%) were currently enrolled as undergraduates, 7 (3.4%) were on a leave of absence, 12 (5.8%) held a graduate-level degree, and 9 (4.4%) had a high school education or below. Monthly personal allowance was ≤US $500 for 95 (46.1%) participants and US $501-US $1000 for 74 (35.9%) participants, with an average of US $783 (SD 558), assuming an approximate conversion rate of US $1=KRW 1000. In terms of living arrangements, 115 (55.8%) participants lived with their parents, 82 (39.8%) lived independently, 6 (2.9%) resided in dormitories, and 3 (1.5%) lived with relatives.

Table 2. General characteristics of participants (N=206).

Variable and category	Value
Gender, n (%)	
Male	102 (49.5)
Female	104 (50.5)
Age (years), n (%)	
19-24	77 (37.4)
25-29	129 (62.6)
Age (years), mean (SD)	25.21 (2.54)
Education level, n (%)	
High school or below	9 (4.4)
Undergraduate (enrolled)	60 (29.1)
Undergraduate (on leave)	7 (3.4)
Bachelor's degree	118 (57.3)
Graduate school or higher	12 (5.8)
Monthly personal allowance (US $)[a], n (%)	
≤500	95 (46.1)
501-1000	74 (35.9)
Monthly personal allowance (US $), mean (SD)	783 (558)
Housing type, n (%)	
Living independently	82 (39.8)
Parents' home	115 (55.8)
Dormitory	6 (2.9)
Relative's home	3 (1.5)

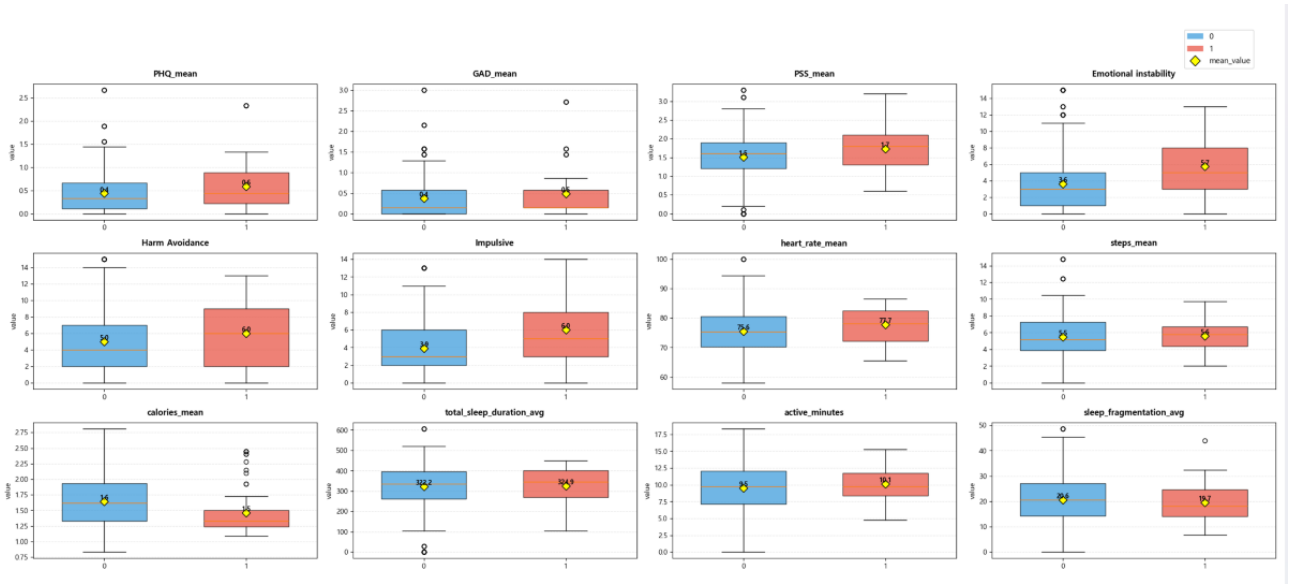
[a]Refers to the monthly allowance an individual uses for personal expenses (eg, transportation, eating out, and cultural activities). For convenience, US $1=KRW 1000 was used for conversion.

Figure 2 presents group comparisons between the nonepisode group (0) and the episode group (1) across psychological, personality, and wearable-derived physiological and behavioral indicators. The episode group showed higher levels of depressive symptoms (PHQ_mean), anxiety (GAD_mean), and perceived stress (PSS_mean), as well as higher scores on personality vulnerability traits, including emotional instability, harm avoidance, and impulsivity. These patterns suggest that individuals in the episode group were more emotionally vulnerable and exposed to greater negative affective states.

Clear differences were also observed in wearable-derived health metrics. The episode group exhibited a higher average heart rate (heart_rate_mean) but lower average daily steps (steps_mean) and calorie expenditure (calories_mean). Additionally, they demonstrated shorter total sleep duration (total_sleep_duration_avg) and greater sleep fragmentation (sleep_fragmentation_avg). Collectively, these findings indicate that the episode group experienced reduced behavioral stability, heightened physiological arousal, and disruptions in sleep-activity rhythms compared to the nonepisode group.

XSL•FO
RenderX

Figure 2. Comparison of psychological, personality, and wearable-derived behavioral indicators between groups. GAD: Generalized Anxiety Disorder-7; PHQ: Patient Health Questionnaire-9; PSS: Perceived Stress Scale-10.



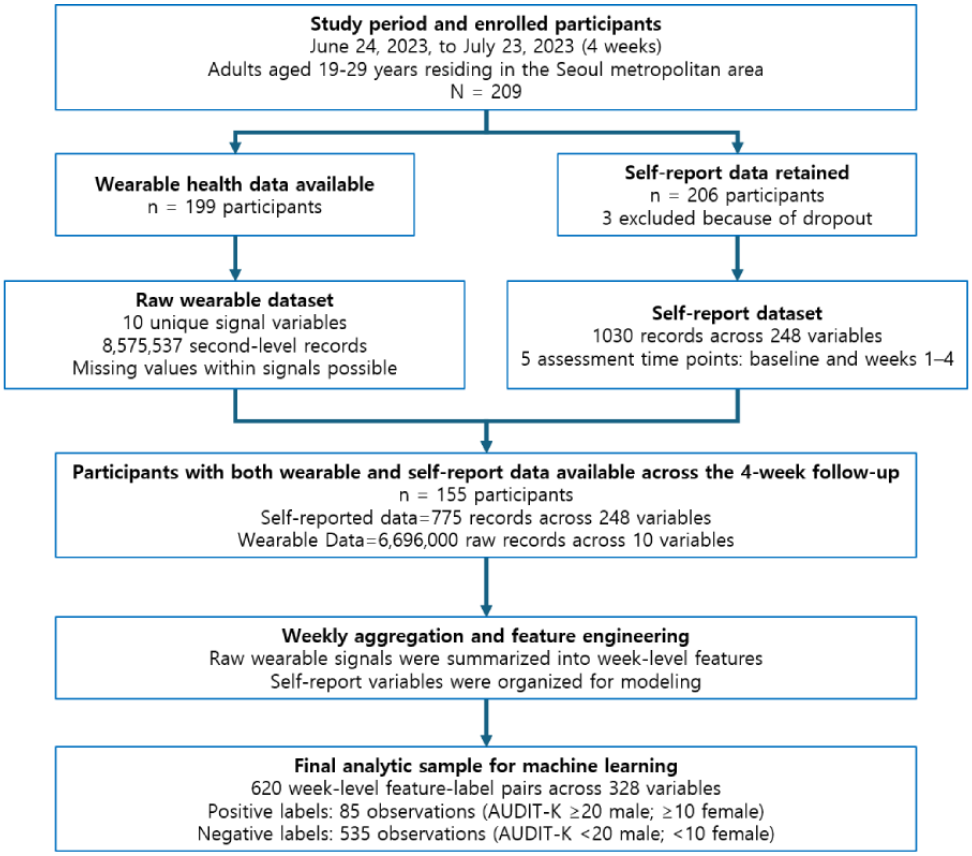
Data Overview

A total of 620 week-level observations were obtained from 206 participants across the 4-week monitoring period. Based on the AUDIT-K high-risk drinking thresholds (≥20 for men and ≥10 for women), 85 (13.7%) observations were labeled as positive high-risk drinking episodes, and 535 (86.3%) observations were labeled as negative episodes.

Model development and evaluation were conducted using 5-fold participant-level grouped cross-validation. All week-level observations from the same participant were kept within the same fold to avoid subject-level leakage. Therefore, model performance was estimated from held-out participants rather than from randomly split week-level observations. Within each fold, preprocessing, class-imbalance handling through algorithm-specific class weighting, hyperparameter tuning, and model fitting were performed using only the training portion, and the held-out fold was used exclusively for performance estimation. The overall modeling workflow is summarized in Figure 3. Accordingly, all primary performance estimates were derived from participant-level grouped cross-validation rather than from randomly split week-level observations.

XSL•FO
RenderX

Figure 3. Study sample and derivation of the analytic dataset. AUDIT-K: Alcohol Use Disorders Identification Test-Korean version.



Performance of High-Risk Drinking Episode Prediction Models

Table 3 summarizes model performance estimated using 5-fold participant-level grouped cross-validation. To prevent information leakage, all preprocessing procedures, class-imbalance correction, hyperparameter tuning, and model fitting were performed exclusively within the training portion of each fold; the held-out fold was used only for final performance estimation. Because the intended application was early detection of future high-risk drinking episodes, threshold-dependent metrics were calculated using a sensitivity-oriented decision threshold of 0.20, while ROC AUC and PR AUC were evaluated as threshold-independent discrimination metrics.

Across the evaluated models, the integrated random forest model showed the most favorable early-detection profile. This model achieved a mean accuracy of 0.617 (SD 0.078), precision/PPV of 0.209 (SD 0.069), recall/sensitivity of 0.653 (SD 0.144), NPV of 0.919 (SD 0.036), F_1-score of 0.315 (SD 0.097), ROC AUC of 0.681 (SD 0.079), and PR AUC of 0.255 (SD 0.090). The integrated XGBoost model showed higher accuracy but lower sensitivity, with a mean accuracy of 0.670 (SD 0.089), precision/PPV of 0.172 (SD 0.039), recall/sensitivity of 0.399 (SD 0.174), NPV of 0.880 (SD 0.059), F_1-score of 0.236 (SD 0.066), ROC AUC of 0.651 (SD 0.089), and PR AUC of 0.228 (SD 0.074).

For models trained on self-report features only, the XGBoost classifier achieved a mean accuracy of 0.396 (SD 0.225), precision/PPV of 0.128 (SD 0.044), recall/sensitivity of 0.630 (SD 0.356), NPV of 0.884 (SD 0.085), F_1-score of 0.203 (SD 0.076), ROC AUC of 0.603 (0.084), and PR AUC of 0.213 (0.023). The random forest classifier showed a mean accuracy of 0.491 (SD 0.057), precision/PPV of 0.150 (SD 0.044), recall/sensitivity of 0.607 (SD 0.221), NPV of 0.882 (SD 0.086), F_1-score of 0.234 (SD 0.061), ROC AUC of 0.569 (SD 0.137), and PR AUC of 0.226 (SD 0.080).

For wearable-derived health features only, the XGBoost model yielded a mean accuracy of 0.727 (SD 0.086), precision/PPV of 0.208 (SD 0.095), recall/sensitivity of 0.393 (SD 0.246), NPV of 0.888 (SD 0.069), F_1-score of 0.260 (SD 0.133), ROC AUC of 0.649 (SD 0.087), and PR AUC of 0.248 (SD 0.070). The random forest model achieved a mean accuracy of 0.601 (SD 0.056), precision/PPV of 0.185 (SD 0.061), recall/sensitivity of 0.581 (SD 0.147), NPV of 0.902 (SD 0.051), F_1-score of 0.278 (SD 0.086), ROC AUC of 0.610 (SD 0.076), and PR AUC of 0.264 (SD 0.107).

Overall, the integrated random forest model showed the highest recall/sensitivity and ROC AUC among the evaluated models, suggesting that combining baseline psychological vulnerability indicators with wearable-derived weekly summaries may improve sensitivity for detecting future high-risk drinking episodes. However, precision/PPV remained modest across models, indicating that the models should be interpreted as preliminary risk-stratification tools rather than stand-alone decision systems.

The confusion matrices and ROC curves for the integrated XGBoost and random forest models were generated from pooled out-of-fold predictions across participant-level grouped

XSL•FO
RenderX

cross-validation folds and are presented in Figures 4 and 5, respectively.

Table 3. Machine learning model performance for week-ahead prediction of high-risk drinking episodes[a].

Feature set	Model	Accuracy, mean (SD)	Precision/PPV[b], mean (SD)	Recall/sensitivity, mean (SD)	NPV[c], mean (SD)	F_1-score, mean (SD)	ROC AUC[d], mean (SD)	PR AUC[e], mean (SD)
Self-report only	XGBoost[f]	0.396 (0.225)	0.128 (0.044)	0.630 (0.356)	0.884 (0.085)	0.203 (0.076)	0.603 (0.084)	0.213 (0.023)
Self-report only	Random forest	0.491 (0.057)	0.150 (0.044)	0.607 (0.221)	0.882 (0.086)	0.234 (0.061)	0.569 (0.137)	0.226 (0.080)
Health data only	XGBoost	0.727 (0.086)	0.208 (0.095)	0.393 (0.246)	0.888 (0.069)	0.260 (0.133)	0.649 (0.087)	0.248 (0.070)
Health data only	Random forest	0.601 (0.056)	0.185 (0.061)	0.581 (0.147)	0.902 (0.051)	0.278 (0.086)	0.610 (0.076)	0.264 (0.107)
Integrated	XGBoost	0.670 (0.089)	0.172 (0.039)	0.399 (0.174)	0.880 (0.059)	0.236 (0.066)	0.651 (0.089)	0.228 (0.074)
Integrated	Random forest	0.617 (0.078)	0.209 (0.069)	0.653 (0.144)	0.919 (0.036)	0.315 (0.097)	0.681 (0.079)	0.255 (0.090)

[a]Values are presented as mean (SD) across 5 participant-level grouped cross-validation folds. Threshold-dependent metrics were calculated using a decision threshold of 0.20. Class imbalance was addressed using algorithm-specific class weighting within the training portion of each fold. Area under the receiver operating characteristic curve and area under the precision-recall curve are threshold-independent discrimination metrics. Precision is equivalent to positive predictive value, and recall is equivalent to sensitivity.

[b]PPV: positive predictive value.

[c]NPV: negative predictive value.

[d]ROC AUC: area under the receiver operating characteristic curve.

[e]PR AUC: area under the precision-recall curve.

[f]XGBoost: extreme gradient boosting.

Figure 4. Prediction performance of the integrated extreme gradient boosting model. PR AUC: area under the precision-recall curve; ROC: receiver operating characteristic curve.

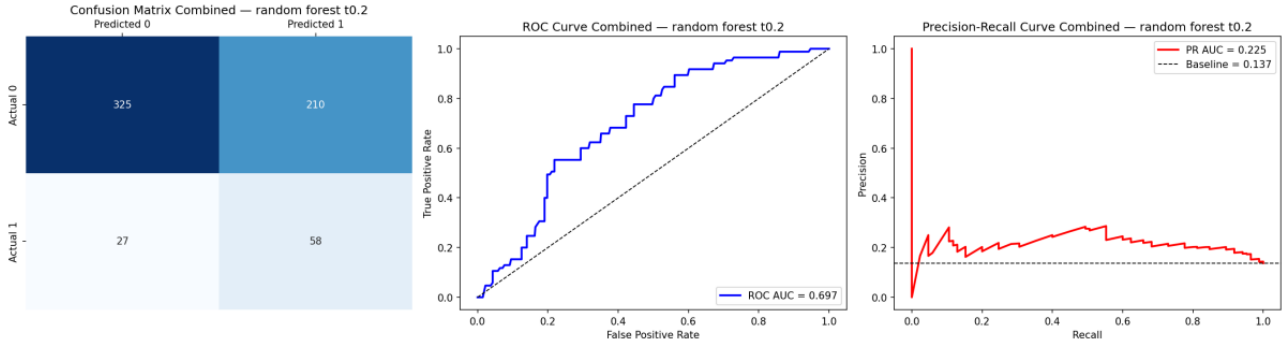
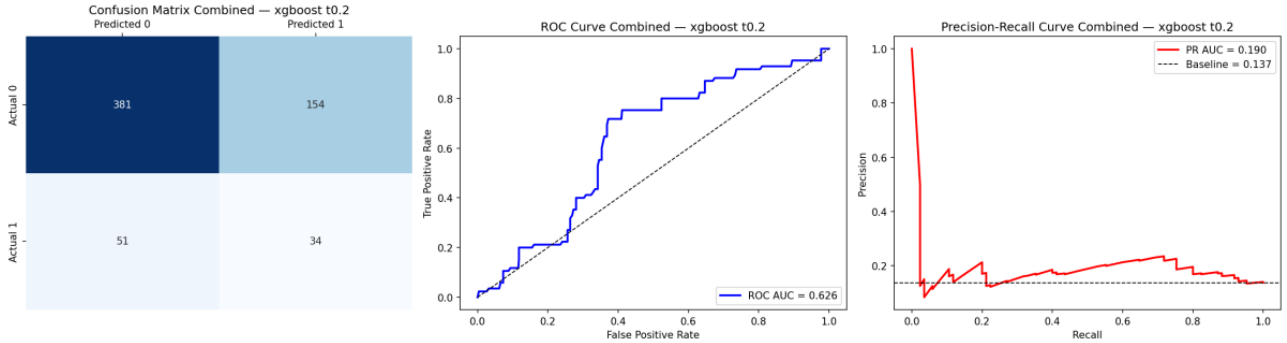


Figure 5. Prediction performance of the integrated random forest model. PR AUC: area under the precision-recall curve; ROC: receiver operating characteristic curve.



XSL•FO
RenderX

Model Explainability Using Shapley Additive Explanations

We examined which features contributed most strongly to the performance of each machine learning model. Shapley additive explanations (SHAP) summary plots and bar plots visually illustrate the variables that contributed most to model predictions (Figures 6 and 7 and Table 4). Because SHAP values quantify feature contributions to model output rather than causal effects, the identified variables were interpreted as predictive correlates rather than physiological precursors or causal mechanisms.

Figure 6. Shapley additive explanations (SHAP) summary plots for the integrated models: extreme gradient boosting (XGBoost; left) and random forest (right).

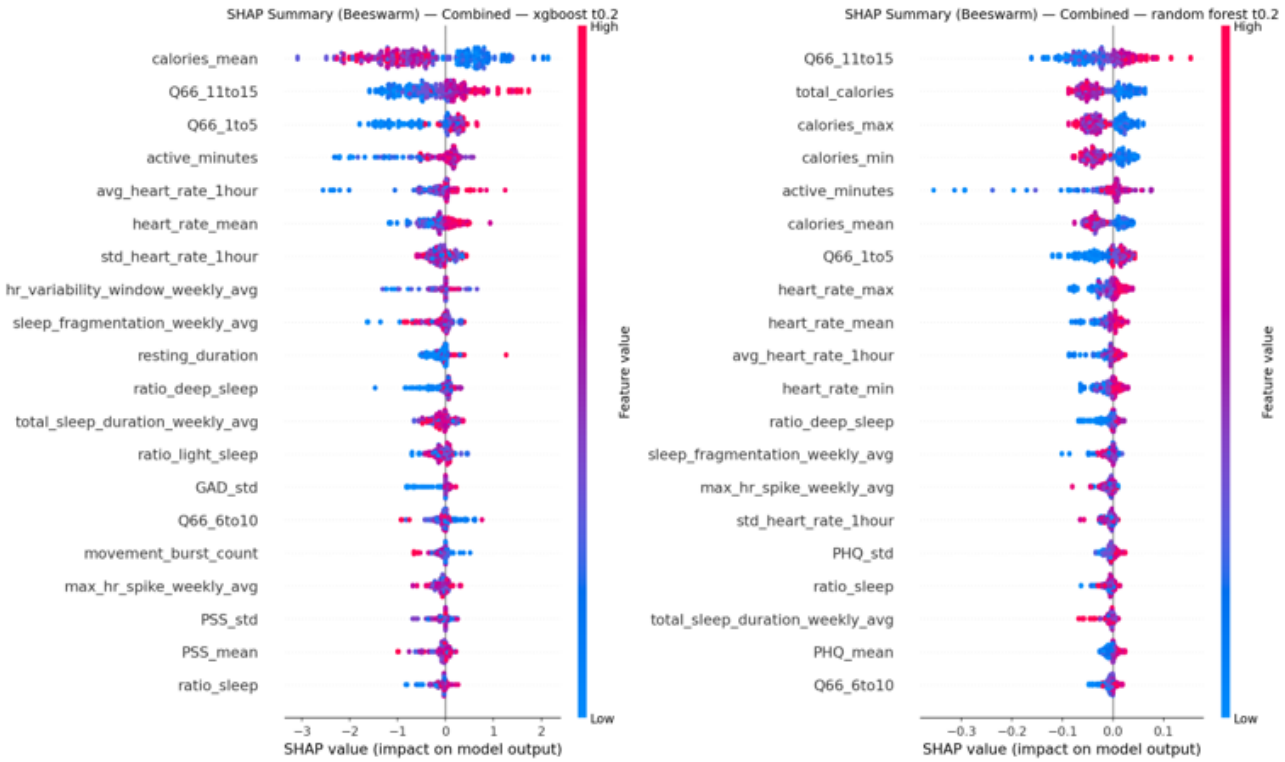
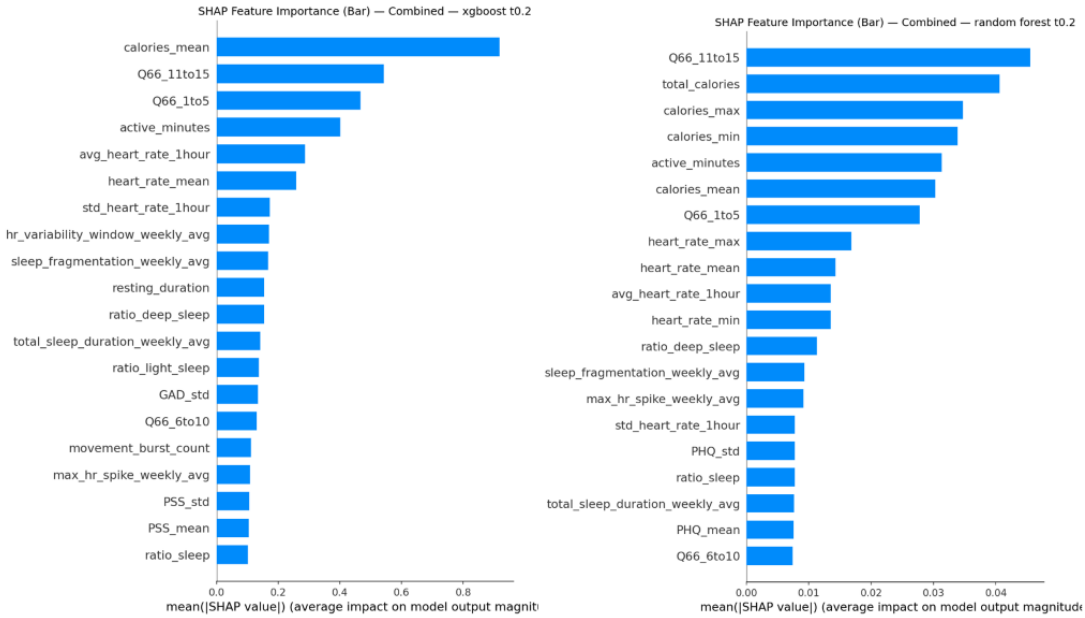


Figure 7. Top 20 features ranked by mean absolute Shapley additive explanations (SHAP) values: integrated extreme gradient boosting (XGBoost; left) and integrated random forest (right).



XSL•FO
RenderX

Table 4. Top-ranked Shapley additive explanations (SHAP) feature importance scores for the integrated extreme gradient boosting (XGBoost) and random forest models[a].

Rank and feature	SHAP feature importance score
XGBoost	
1. Mean calorie expenditure (calories_mean)	0.918
2. Impulsivity (Q66_11to15)	0.543
3. Emotional instability (Q66_1to5)	0.467
4. Active minutes (active_minutes)	0.401
5. Average hourly heart rate (avg_heart_rate_1hour)	0.287
6. Mean heart rate (heart_rate_mean)	0.259
7. Hourly heart rate variability (std_heart_rate_1hour)	0.174
8. Weekly heart rate variability (hr_variability_window_weekly_avg)	0.170
9. Sleep fragmentation (sleep_fragmentation_weekly_avg)	0.168
10. Resting duration (resting_duration)	0.155
11. Ratio of deep sleep (ratio_deep_sleep)	0.155
12. Total sleep duration (total_sleep_duration_weekly_avg)	0.142
13. Ratio of light sleep (ratio_light_sleep)	0.138
14. Anxiety variability (GAD_std)	0.135
15. Harm avoidance (Q66_6to10)	0.131
Random forest feature	
1. Impulsivity (Q66_11to15)	0.045
2. Total calories (total_calories)	0.041
3. Maximum calories (calories_max)	0.035
4. Minimum calories (calories_min)	0.034
5. Active minutes (active_minutes)	0.031
6. Mean calorie expenditure (calories_mean)	0.030
7. Emotional instability (Q66_1to5)	0.028
8. Maximum heart rate (heart_rate_max)	0.017
9. Mean heart rate (heart_rate_mean)	0.014
10. Average hourly heart rate (avg_heart_rate_1hour)	0.014
11. Minimum heart rate (heart_rate_min)	0.014
12. Ratio of deep sleep (ratio_deep_sleep)	0.011
13. Sleep fragmentation (sleep_fragmentation_weekly_avg)	0.009
14. Maximum heart rate spikes (max_hr_spike_weekly_avg)	0.009
15. Hourly heart rate variability (std_heart_rate_1hour)	0.008

[a]The table presents the top 15 features ranked by SHAP feature importance scores in the integrated models combining self-report and wearable-derived health features. SHAP feature importance scores were calculated as the average absolute SHAP values across observations for each feature and were used to rank the relative contribution of predictors within each model. SHAP values indicate relative feature contributions to model output and should be interpreted as predictive correlates rather than causal effects.

To interpret the integrated XGBoost model, we conducted SHAP analyses to assess feature importance and contribution patterns. Based on the SHAP summary plot and mean absolute SHAP values, the most influential features included mean calorie expenditure (calories_mean), impulsivity (Q66_11to15), emotional instability (Q66_1to5), active minutes, average heart rate per hour (avg_heart_rate_1hour), and mean heart rate (heart_rate_mean). These findings indicate that the integrated model relied on both baseline vulnerability indicators and wearable-derived weekly summaries when classifying next-week high-risk drinking labels.

Among baseline psychological indicators, impulsivity and emotional instability demonstrated prominent contributions to model predictions. Among wearable-derived features, calorie expenditure, active minutes, and heart-rate summaries

XSL•FO
RenderX

contributed prominently to the model output. This pattern suggests that baseline emotional and personality vulnerability indicators provided between-person risk information, whereas wearable-derived summaries contributed week-varying behavioral and physiological information. In the integrated XGBoost model, lower values of calorie expenditure tended to contribute positively to model output for high-risk classification, whereas higher values of impulsivity and emotional instability generally contributed toward higher predicted risk. These patterns should be interpreted as model-identified predictive correlates associated with the classification of subsequent high-risk drinking labels.

To interpret the integrated random forest model, we conducted SHAP analyses to identify the features that contributed most strongly to next-week predictions of high-risk drinking episodes. Based on the SHAP summary plot and rankings of mean absolute SHAP values, the most influential variables included impulsivity (Q66_11to15), total calories, maximum calories, minimum calories, active minutes, mean calories, emotional instability (Q66_1to5), maximum heart rate, mean heart rate, and average heart rate per hour.

The random forest model similarly showed that both baseline vulnerability indicators and wearable-derived activity, heart rate, sleep, or calorie-related summaries contributed to prediction outcomes. Among baseline psychological variables, impulsivity and emotional instability demonstrated consistent contributions. Among wearable-derived variables, calorie-related indices, active minutes, and heart rate indicators were prominent contributors. The integrated random forest model showed a particularly strong contribution of impulsivity, followed by calorie-related wearable features. These findings indicate that SHAP-identified features helped explain model classification patterns, but they do not establish causal pathways or temporally specific physiological mechanisms.

Because this was an observational sensing study, wearable-derived predictors may also reflect unmeasured contextual or behavioral factors, including physical activity routines, sleep debt, acute illness, academic or work schedules, weekend social activity, shift work, device wear time, and consumer-device estimation error. Therefore, SHAP-identified features should be interpreted as predictive correlates that require validation in future studies using more granular contextual, behavioral, and physiological data.

Discussion

Principal Findings

This study developed machine learning models that integrate psychological and personality vulnerability factors with wearable-derived physiological and behavioral signals to predict weekly high-risk drinking episodes among young adults. Several key findings emerged.

First, the integrated models combining biosignals and psychological-personality vulnerabilities showed a modest but meaningful early-detection profile under 5-fold participant-level grouped cross-validation compared with models using either feature set alone. The integrated XGBoost model achieved a mean accuracy of 0.670 (SD 0.089) and a mean ROC AUC of 0.651 (SD 0.089), while the integrated random forest model achieved a mean accuracy of 0.617 (SD 0.078) and a ROC AUC of 0.681 (SD 0.079). The corresponding PR AUC values were mean 0.228 (SD 0.074) for XGBoost and mean 0.255 (SD 0.090) for random forest. These cross-validated estimates suggest that the integrated models provided modest but useful week-ahead predictive information for high-risk drinking episodes under participant-level grouped validation. In particular, the integrated random forest model showed the most favorable sensitivity-oriented performance, with a mean recall/sensitivity of 0.653 (SD 0.144) and a mean NPV of 0.919 (SD 0.036). However, precision/PPV remained modest at a mean of 0.209 (SD 0.069), indicating that the model should be interpreted as a preliminary risk-stratification tool rather than a stand-alone decision system.

These findings are broadly consistent with previous machine learning and AI-based studies addressing alcohol dependence and substance-use risk. For example, a study detecting substance-use risk from Instagram images reported an ROC AUC of 0.65 [13], and a multimodal neuroimaging classifier for chronic alcohol dependence achieved an accuracy of 79.3% [14]. Research predicting high-risk drinking patterns among medical students reported an area under the curve of 0.738 [38]. Smartphone sensor–based studies forecasting heavy drinking showed a wider performance range, with accuracies from 94.3% to 96.6% [11,12]. However, direct comparison across studies should be made cautiously because prediction horizons, outcome definitions, data modalities, validation schemes, operating thresholds, and participant-level leakage controls differ substantially across studies.

In comparison, a distinctive strength of this study lies in its integration of behavioral biosignals with psychological and personality vulnerabilities that are theoretically and empirically central to alcohol-use problems. Rather than relying on a single modality or purely behavioral data, the present models incorporated emotional vulnerability, including depression, anxiety, and perceived stress, and personality traits such as emotional instability and impulsivity. This multimodal approach may improve week-ahead risk estimation by combining stable vulnerability profiles with week-varying physiological and behavioral information. The integrated model should therefore be interpreted as combining 2 distinct sources of predictive information: stable between-person vulnerability captured by baseline psychological and personality indicators, and week-varying behavioral and physiological information captured by wearable-derived summaries. Accordingly, the model's week-ahead performance does not necessarily imply that baseline psychological variables captured short-term within-person psychological change. Rather, these variables likely contributed to identifying individuals with higher underlying vulnerability, while wearable-derived weekly features provided the time-varying component of the prediction.

Examining the most influential predictors showed that both baseline vulnerability indicators and wearable-derived weekly summaries contributed to model prediction. In the integrated XGBoost model, the features with the highest mean absolute SHAP values included mean calorie expenditure, impulsivity,

XSL•FO
RenderX

emotional instability, active minutes, average heart rate per hour, and mean heart rate. In the integrated random forest model, the highest-ranking SHAP features included impulsivity, total calories, maximum calories, minimum calories, active minutes, mean calories, emotional instability, maximum heart rate, mean heart rate, and average heart rate per hour.

Across both integrated models, impulsivity and emotional instability emerged as prominent baseline vulnerability indicators, while calorie-related metrics, active minutes, and heart-rate summaries were consistently identified among the most influential wearable-derived features. This pattern suggests that the models combined relatively stable between-person vulnerability information with week-level behavioral and physiological summaries when estimating the likelihood of a subsequent high-risk drinking label. The convergence of feature rankings across the 2 algorithms also supports the relevance of integrating psychological vulnerability profiles with passive wearable signals for week-ahead alcohol-risk prediction. Although these SHAP-identified features helped explain the models' classification patterns, they should be interpreted as predictive correlates rather than causal mechanisms. Because this was an observational sensing study, activity-, calorie-, and heart rate–related features may also reflect broader contextual factors such as habitual physical activity, sleep debt, illness, academic or work schedules, weekend social activity, shift work, or device wear time. Future studies incorporating more granular contextual information and person-centered deviations from individual baselines will be needed to clarify which wearable-derived changes are most specific to alcohol-related risk.

These findings align with prior work using mobile sensor data, which demonstrated that increased physical movement—such as greater acceleration and travel distance—preceded binge-drinking episodes by 1-6 hours [11]. Other studies have similarly reported that fluctuations in activity levels serve as meaningful predictors of drinking behavior [10,12], supporting the relevance of activity-related digital signals as predictive features for drinking-related outcomes. Although this study focused on weekly episode prediction rather than immediate predrinking states, the finding that activity- and calorie-related wearable summaries contributed to model prediction is directionally consistent with previous work on sensor-based alcohol-risk prediction. However, these similarities should be interpreted cautiously because this study did not directly measure immediate predrinking states and cannot determine whether the observed wearable patterns reflected drinking-related processes or other contextual factors. Prior research has also shown that individuals who drink tend to be more physically active than nondrinkers [39], and population-based health surveillance data in Korea likewise indicate that among adults aged 20-39 years, higher levels of alcohol consumption are associated with greater physical activity [40].

A notable finding was that calorie-related variables contributed strongly to model predictions in both integrated models and differed between the high-risk episode and nonepisode groups. Unlike heart rate or step count, both of which largely reflect gross physical movement, calorie expenditure is a composite wearable-derived metric that may be influenced by sleep-wake rhythm, arousal level, metabolic factors, device estimation algorithms, and the intensity and quality of physical activity. However, this pattern should not be interpreted as evidence of altered metabolic efficiency or disrupted biological rhythms. Rather, it indicates that calorie-related wearable summaries were useful predictive features in this model. Although prior research has suggested that alcohol use may be associated with disruptions in biological rhythms [41], this observational sensing data cannot determine whether the observed calorie-related patterns reflected alcohol-related physiological processes or other contextual factors. Because calorie estimates from consumer wearables can be affected by activity type, device wear time, individual physiology, and algorithmic estimation error, this finding should be interpreted cautiously and validated in future studies using more granular contextual and physiological data.

Taken together, the findings suggest that integrating wearable-derived physiological signals with psychological and personality-based vulnerability factors may provide useful information for week-ahead prediction of high-risk drinking episodes. In particular, personality vulnerabilities such as impulsivity and emotional instability, together with activity-, calorie-, and heart rate–related wearable summaries, emerged as predictive correlates of subsequent high-risk drinking labels. These results provide preliminary support for risk-stratified digital monitoring approaches, while also indicating that this model is not yet sufficient as a stand-alone early-warning or intervention-triggering system. Future JITAI-oriented models may require more proximal within-person signals, such as ecological momentary assessment (EMA)–based affective states, individualized deviations from wearable baselines, and contextual data that can distinguish alcohol-related risk signals from broader daily-life variability.

From a translational perspective, these findings are best interpreted as supporting a preliminary risk-stratification approach rather than immediate deployment as a stand-alone early-warning system. A wearable-assisted prediction system of this kind could be most realistically implemented as part of a stepped-care workflow in which model-generated risk signals prompt low-burden self-monitoring feedback, brief motivational messages, or clinician review when elevated risk persists. In primary care, community mental health, university counseling, or addiction service settings, such a system could support clinicians, counselors, or care coordinators in identifying young adults who may benefit from closer monitoring or preventive intervention. This implementation pathway is consistent with evidence suggesting that practitioner-delivered alcohol interventions may have greater short-term effects than digitally delivered interventions alone, while digital interventions may still provide scalable benefits over time and at the population level [42]. Therefore, the model may be most useful when integrated into ongoing professional support or postintervention follow-up, where weekly risk estimates can help guide timely preventive feedback, brief intervention, or referral.

Even when predictive performance remains preliminary, low-burden digital risk feedback may have public health value if it helps individuals recognize periods of elevated vulnerability

XSL•FO
RenderX

and prepare preventive coping strategies before high-risk drinking occurs. This aligns with a precautionary public health perspective, in which early, proportionate, and low-risk preventive actions may be justified when potential harms are substantial, and the intervention burden is minimal [43]. However, responsible deployment would require higher and externally validated predictive performance, clear escalation pathways, user consent, transparent alert logic, and safeguards to prevent unnecessary alarm, stigma, or overintervention. In this regard, the model may be better suited to an institution- or service-mediated business to business to consumer–type model than to a stand-alone direct-to-consumer application [44]. In such a model, individuals would use the mobile app and wearable device in daily life, while health care providers, counseling services, public health agencies, university counseling centers, workplace mental health programs, or addiction care coordinators would provide oversight, interpret risk signals, and connect users to appropriate support when needed. In the South Korean context, this approach would also require careful attention to privacy, data minimization, secure storage, governance of sensitive behavioral and health data, regulatory classification, and clear rules regarding who is authorized to access and act on model-generated risk information.

Limitations

Despite the meaningful findings of this study, several limitations should be acknowledged. First, the sample consisted exclusively of young adults in their 20s, which constrains the generalizability of the results. High-risk drinking patterns are known to vary substantially across age groups, occupational contexts, and sociocultural environments; therefore, the predictive performance demonstrated in this cohort may not translate directly to older adults or clinical populations with more severe alcohol-related problems. Nevertheless, given the rising prevalence of problematic drinking among young adults in South Korea [1,2], early detection and intervention efforts targeting this age group carry important preventive value. Future studies should expand the model to a broader age range to evaluate its robustness across more diverse populations.

Second, this study developed prediction models using relatively short-term longitudinal data from 5 assessment time points over a 4-week follow-up period. As a result, the models were unable to capture long-term fluctuations in drinking behavior, seasonal or contextual influences, and broader temporal dynamics that may affect alcohol use risk. Future work should incorporate longer monitoring periods to more precisely characterize how changes in biosignal patterns interact with shifts in drinking risk over time. Although certain mental health conditions, such as mood or anxiety disorders, often show seasonal or cyclical patterns and therefore require extended longitudinal observation [28,30], such temporal regularity is not a defining feature of addictive behaviors. Indeed, prior studies have reported that using only 1-3 days of past smartphone sensor data was sufficient to enhance model performance for predicting drinking episodes [12]. From an implementation standpoint, accumulating excessively long digital histories may not be advantageous. In real-world mobile apps, storing and processing large volumes of behavioral and physiological data can impose substantial data usage burdens, reduce algorithmic efficiency, and negatively affect user adherence [45]. Thus, while long-term data collection can provide analytic benefits, it may conflict with the practical requirement for lightweight, low-burden, and scalable digital health solutions.

Third, psychological and personality vulnerability variables were measured only at baseline and were therefore modeled as time-invariant predictors. As a result, these variables likely captured between-person differences in underlying vulnerability rather than within-person psychological changes preceding a high-risk drinking episode. Although wearable-derived weekly summaries provided time-varying behavioral and physiological information, this model did not directly estimate individualized deviations from each participant's own psychological baseline. This distinction is important for interpreting the model as a week-ahead risk prediction tool. Emotional states and stress levels can fluctuate considerably on a week-to-week basis, and treating these variables as static may have influenced both predictive performance and interpretability. The findings suggest that stable vulnerability profiles can improve risk stratification when combined with weekly wearable signals, but they should not be interpreted as demonstrating that baseline psychological variables function as proximal early-warning signals.

However, the psychological measures used in this study, including depression (PHQ-9), anxiety (GAD-7), perceived stress (PSS-10), and personality vulnerability indicators, were originally developed to capture relatively stable symptom severity or vulnerability profiles rather than rapid state-level changes [32,34,46]. Consequently, although repeated weekly self-report symptom scores may provide additional predictive information, they differ conceptually from the baseline vulnerability indicators used in the primary models. The primary modeling framework, therefore, treated psychological and personality measures as baseline vulnerability indicators and focused on whether week-varying wearable-derived summaries could contribute to the prediction of the following week's high-risk drinking label.

Nevertheless, research using EMA for affective disorders has demonstrated that negative affect often shows substantial moment-to-moment fluctuation, whereas positive affect tends to exhibit dampened diurnal variation [47,48]. Emotional states also react strongly to daily stressors [49]. Therefore, future models should incorporate EMA-based affective states, repeated brief affect measures, or person-centered deviation features to better capture within-person changes that may be more actionable for JITAI delivery.

Fourth, because this was an observational sensing study, wearable-derived predictors may have been influenced by unmeasured contextual and behavioral factors. For example, physical activity routines, sleep debt, acute illness, academic or work schedules, weekend social activity, shift work, device wear time, and other daily-life contexts may affect heart rate, sleep, step count, and calorie estimates independently of alcohol-risk processes. Therefore, SHAP-identified features should be interpreted as predictive correlates rather than causal mechanisms or physiological precursors. Future studies should incorporate contextual measures, EMA-based state assessments,

XSL•FO
RenderX

and person-centered deviation features to better distinguish alcohol-related risk signals from broader daily-life variability.

Fifth, the prediction model in this study defined weekly high-risk drinking episodes using a binary threshold based on the AUDIT-K. Because AUDIT-K scores rely on self-reported information, they are subject to recall bias and social desirability bias, which may limit the precision of the outcome labels. It is possible that the model's predictive performance would differ if more granular event-level drinking assessments or biosensing-based alcohol detection technologies were used.

At the same time, many machine learning and AI studies aiming to predict substance use or alcohol-related risk rely on self-reported labels for the outcome variable [10-12,28,30,50]. Although the reliability of such outcome labels may be debated, they remain the most practical and widely adopted approach in real-world digital phenotyping research. Nonetheless, future studies will benefit from incorporating passive and low-burden outcome measurement tools—such as continuous biosensing devices or ecological momentary detection of drinking events—which may enhance labeling accuracy and further improve predictive performance.

Sixth, although participant-level grouped cross-validation was used to reduce subject-level leakage, the number of positive high-risk drinking episodes within each held-out fold was relatively small. This may have contributed to variability in sensitivity and precision estimates across folds. In addition, the sensitivity-oriented operating threshold improved detection of positive episodes but resulted in modest precision/PPV, indicating that the model may generate false-positive alerts if applied as a stand-alone early-warning tool. Therefore, this model should be interpreted as a preliminary risk-stratification approach rather than a clinically deployable decision system.

Finally, the predictive modeling in this study was limited to 2 tree-based algorithms, XGBoost and random forest. These models were chosen to balance interpretability and predictive performance, particularly given our focus on identifying the likelihood of entering the high-risk drinking category in the following week. However, tree-based models have only limited capacity to capture temporal dependencies or long-range sequential patterns inherent in week-to-week behavioral fluctuations. Future research should incorporate more advanced time-series deep learning approaches, such as long short-term memory networks, temporal convolutional networks, or transformer-based architectures, which may better model temporal dynamics and improve prediction accuracy as longer-term longitudinal data become available.

Conclusions

This study developed machine learning models integrating emotional and personality vulnerability factors with wearable-derived physiological and behavioral signals to predict weekly high-risk drinking episodes among young adults. The findings suggest that combining stable baseline vulnerability indicators with week-level wearable summaries may provide useful information for week-ahead risk estimation. Activity-, sleep-, heart rate-, and vulnerability-related features contributed to model prediction, although these features should be interpreted as predictive correlates rather than causal mechanisms or proximal physiological precursors.

Importantly, this work provides preliminary evidence for the feasibility of wearable-assisted risk stratification for alcohol-related risk in South Korea. Rather than establishing a fully deployable early-warning system, these findings offer an initial basis for future digital monitoring approaches. Future studies involving broader age groups, longer-term longitudinal data, more proximal within-person measures, and external validation will be essential to enhance the generalizability, clinical use, and JITAI relevance of prediction models for high-risk drinking.

Acknowledgments

We extend our deepest gratitude to all members of the research team who have worked tirelessly over a long period to achieve optimal model performance.

Data Availability

The datasets generated and analyzed during this study contain sensitive personal health and behavioral information and are therefore not publicly available. Deidentified data may be made available from the corresponding author upon reasonable request and with appropriate institutional approval.

Funding

This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant RS-2022-KH125882).

Authors' Contributions

JSK, SJJ, HKL, JHK, SJK, YK, and HL designed and conceptualized the study. JSK and SJJ developed the methodology and contributed to writing the manuscript. JSK was responsible for project administration and visualization. JHK curated and processed the data and supported data analysis. HKL provided funding acquisition, supervision, and resources. SJK and YK contributed to conceptualization and provided resources. HL performed data curation, investigation, and software implementation. All authors contributed to writing, reviewing, and editing the manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Exploratory model performance when weekly self-report scores were included.
[DOCX File , 1502 KB - mhealth_v14i1e88223_app1.docx]

Multimedia Appendix 2
Optuna hyperparameter search spaces.
[DOCX File , 23 KB - mhealth_v14i1e88223_app2.docx]

References

1. Rim SJ, Hahm BJ, Seong SJ, Park JE, Chang SM, Kim BS, et al. Prevalence of mental disorders and associated factors in Korean adults: National Mental Health Survey of Korea 2021. Psychiatry Investig 2023;20(3):262-272. [doi: 10.30773/pi.2022.0307] [Medline: 36990670]
2. Korea Disease Control and Prevention Agency. Community Health Survey Report. Seoul, Korea: Korea Disease Control and Prevention Agency; 2023.
3. Ministry of Agriculture Food and Rural Affairs. Liquor Industry Information Survey. Sejong, Korea: Ministry of Agriculture, Food and Rural Affairs; 2020.
4. Avery AR, Tsang S, Seto EYW, Duncan GE. Stress, anxiety, and change in alcohol use during the COVID-19 pandemic: findings among adult twin pairs. Front Psychiatry 2020;11:571084 [FREE Full text] [doi: 10.3389/fpsyt.2020.571084] [Medline: 33088277]
5. Giansanti D. The role of the mHealth in the fight against the COVID-19: successes and failures. Healthcare (Basel) 2021;9(1):58 [FREE Full text] [doi: 10.3390/healthcare9010058] [Medline: 33429866]
6. Ryan J, Edney S, Maher C. Anxious or empowered? A cross-sectional study exploring how wearable activity trackers make their owners feel. BMC Psychol 2019;7(1):42 [FREE Full text] [doi: 10.1186/s40359-019-0315-y] [Medline: 31269972]
7. Cormack F, McCue M, Taptiklis N, Skirrow C, Glazer E, Panagopoulos E, et al. Wearable technology for high-frequency cognitive and mood assessment in major depressive disorder: longitudinal observational study. JMIR Ment Health 2019;6(11):e12814 [FREE Full text] [doi: 10.2196/12814] [Medline: 31738172]
8. Nadal C, Earley C, Enrique A, Vigano N, Sas C, Richards D, et al. Integration of a smartwatch within an internet-delivered intervention for depression: protocol for a feasibility randomized controlled trial on acceptance. Contemp Clin Trials 2021;103:106323 [FREE Full text] [doi: 10.1016/j.cct.2021.106323] [Medline: 33621632]
9. Gomes N, Pato M, Lourenço AR, Datia N. A survey on wearable sensors for mental health monitoring. Sensors (Basel) 2023;23(3):1330 [FREE Full text] [doi: 10.3390/s23031330] [Medline: 36772370]
10. Bae S, Chung T, Ferreira D, Dey AK, Suffoletto B. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: implications for just-in-time adaptive interventions. Addict Behav 2018;83:42-47 [FREE Full text] [doi: 10.1016/j.addbeh.2017.11.039] [Medline: 29217132]
11. Bae SW, Suffoletto B, Zhang T, Chung T, Ozolcer M, Islam MR, et al. Leveraging mobile phone sensors, machine learning, and explainable artificial intelligence to predict imminent same-day binge-drinking events to support just-in-time adaptive interventions: algorithm development and validation study. JMIR Form Res 2023;7:e39862 [FREE Full text] [doi: 10.2196/39862] [Medline: 36809294]
12. Bae S, Ferreira D, Suffoletto B, Puyana JC, Kurtz R, Chung T, et al. Detecting drinking episodes in young adults using smartphone-based sensors. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017;1(2):1-36 [FREE Full text] [doi: 10.1145/3090051] [Medline: 35146236]
13. Hassanpour S, Tomita N, DeLise T, Crosier B, Marsch LA. Identifying substance use risk based on deep neural networks and Instagram social media data. Neuropsychopharmacology 2019;44(3):487-494 [FREE Full text] [doi: 10.1038/s41386-018-0247-x] [Medline: 30356094]
14. Guggenmos M, Schmack K, Veer IM, Lett T, Sekutowicz M, Sebold M, et al. A multimodal neuroimaging classifier for alcohol dependence. Sci Rep 2020;10(1):298 [FREE Full text] [doi: 10.1038/s41598-019-56923-9] [Medline: 31941972]
15. Kamarajan C, Ardekani BA, Pandey AK, Kinreich S, Pandey G, Chorlian DB, et al. Random forest classification of alcohol use disorder using fMRI functional connectivity, neuropsychological functioning, and impulsivity measures. Brain Sci 2020;10(2):115 [FREE Full text] [doi: 10.3390/brainsci10020115] [Medline: 32093319]
16. Menon S, Swathi J, Anit S, Nair A, Sarath S. Driver face recognition and sober drunk classification using thermal images. 2019 Presented at: International Conference on Communication and Signal Processing (ICCSP); 2026 June 20; Melmaruvathur, Tamil Nadu, India. [doi: 10.1109/iccsp.2019.8697908]
17. Lannoy S, Duka T, Carbia C, Billieux J, Fontesse S, Dormal V, et al. Emotional processes in binge drinking: a systematic review and perspective. Clin Psychol Rev 2021;84:101971 [FREE Full text] [doi: 10.1016/j.cpr.2021.101971] [Medline: 33497920]

XSL•FO
RenderX

18. Ning K, Gondek D, Patalay P, Ploubidis GB. The association between early life mental health and alcohol use behaviours in adulthood: a systematic review. PLoS One 2020;15(2):e0228667 [FREE Full text] [doi: 10.1371/journal.pone.0228667] [Medline: 32069303]
19. D'Aquino S, Kumar A, Riordan B, Callinan S. Long-term effects of alcohol consumption on anxiety in adults: a systematic review. Addict Behav 2024;155:108047 [FREE Full text] [doi: 10.1016/j.addbeh.2024.108047] [Medline: 38692070]
20. Weckesser LJ, Pilhatsch M, Muehlhan M, Endraß T, Miller R. A systematic review with meta-analysis on the relation between acute stress, alcohol consumption and cortisol levels in individuals with a personal, familial or no alcohol use disorder. Transl Psychiatry 2025;15(1):423 [FREE Full text] [doi: 10.1038/s41398-025-03641-8] [Medline: 41115889]
21. Pedersen DE. Gender differences in college binge drinking: examining the role of depression and school stress. Soc Sci J 2019;50(4):521-529. [doi: 10.1016/j.soscij.2013.03.003]
22. Adan A, Forero DA, Navarro JF. Personality traits related to binge drinking: a systematic review. Front Psychiatry 2017;8:134 [FREE Full text] [doi: 10.3389/fpsyt.2017.00134] [Medline: 28804465]
23. Galen LW, Henderson MJ, Whitman R. The utility of novelty seeking, harm avoidance, and expectancy in the prediction of drinking. Addict Behav 1997;22(1):93-106. [doi: 10.1016/s0306-4603(96)00018-4] [Medline: 9022875]
24. Koller G, Zill P, Skoruppa T, Bondy B, Preuss U, Soyka M. Low level of harm avoidance is associated with serotonin transporter functional haplotype in alcohol-dependent individuals. Psychiatr Genet 2008;18(2):59-63. [doi: 10.1097/ypg.0b013e3282f60333]
25. Johansson A, Hansen S. Novelty seeking and harm avoidance in relation to alcohol drinking in intact rats and following axon-sparing lesions to the amygdala and ventral striatum. Alcohol Alcohol 2002;37(2):147-156. [doi: 10.1093/alcalc/37.2.147] [Medline: 11912070]
26. McNamara IA, King SE, Corbin WR, Fromme K. A longitudinal examination of relations between competitive athletic participation, drinking norms, impulsivity, and sensation seeking and binge drinking throughout college. Psychol Addict Behav 2022;36(7):837-848. [doi: 10.1037/adb0000849]
27. Sliedrecht W, Roozen HG, Witkiewitz K, de Waart R, Dom G. The association between impulsivity and relapse in patients with alcohol use disorder: a literature review. Alcohol Alcohol 2021;56(6):637-650. [doi: 10.1093/alcalc/agaa132] [Medline: 33382416]
28. Lee HJ, Cho CH, Lee T, Jeong J, Yeom JW, Kim S, et al. Prediction of impending mood episode recurrence using real-time digital phenotypes in major depression and bipolar disorders in South Korea: a prospective nationwide cohort study. Psychol Med 2022;53(12):5636-5644. [doi: 10.1017/s0033291722002847]
29. Aqajari SAH, Labbaf S, Tran PH, Nguyen B, Mehrabadi MA, Levorato M. Context-aware stress monitoring using wearable and mobile technologies in everyday settings. arXiv. Preprint posted online on December 14, 2023 2023. [doi: 10.48550/arXiv.2401.05367]
30. Cho CH, Lee T, Kim MG, In HP, Kim L, Lee HJ. Mood prediction of patients with mood disorders by machine learning using passive digital phenotypes based on the circadian rhythm: prospective observational cohort study. J Med Internet Res 2019;21(4):e11029 [FREE Full text] [doi: 10.2196/11029] [Medline: 30994461]
31. Joe KH, Chai SH, Park AR, Lee HK, Shin IH, Min SH. Optimum cut-off score for screening of hazardous drinking using the Korean version of Alcohol Use Disorder Identification Test (AUDIT-K). J Korean Acad Addict Psychiatry 2009;13(1):34-40 [FREE Full text]
32. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary care evaluation of mental disorders. Patient Health Questionnaire. JAMA 1999;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
33. An JY, Seo ER, Lim KH, Shin JH, Kim JB. Standardization of the Korean version of screening tool for depression (Patient Health Questionnaire-9, PHQ-9). J Korean Soc Biol Ther Psychiatry 2013;19(1):47-56 [FREE Full text]
34. Spitzer RL, Kroenke K, Williams JBW, Lowe B. A brief measure for assessing Generalized Anxiety Disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
35. Seo J, Park S. Validation of the Generalized Anxiety Disorder-7 (GAD-7) and GAD-2 in patients with migraine. J Headache Pain 2015;16(1):97 [FREE Full text] [doi: 10.1186/s10194-015-0583-8] [Medline: 26596588]
36. Park JH, Seo YS. Validation of the perceived stress scale (PSS) on samples of korean university students. Korean J Psychol Gen 2010;29(3):611 [FREE Full text]
37. Korean Psychological Association. Enhancing the precision psychological testing system and improving the overall psychological testing regime [Report in Korean]. Military Manpower Administration. 2021. URL: https://tinyurl.com/3k2a2kw3 [accessed 2026-06-28]
38. Marcon G, de Ávila Pereira F, Zimerman A, da Silva BC, von Diemen L, Passos IC, et al. Patterns of high-risk drinking among medical students: a web-based survey with machine learning. Comput Biol Med 2021;136:104747. [doi: 10.1016/j.compbiomed.2021.104747] [Medline: 34449306]
39. Piazza-Gardner AK, Barry AE. Examining physical activity levels and alcohol consumption: are people who drink more active? Am J Health Promot 2012;26(3):e95-e104. [doi: 10.4278/ajhp.100929-lit-328]
40. Oh S. Association between alcohol drinking status and physical activity in Korean adults using the Community Health Survey. J Korea Contents Assoc 2023;23(6):556-567. [doi: 10.5392/jkca.2023.23.06.556]

XSL•FO
RenderX

41. el-Guebaly N. Alcohol, alcoholism, and biological rhythms. Alcohol Clin Exp Res 1987;11(2):139-143. [doi: 10.1111/j.1530-0277.1987.tb01277.x] [Medline: 3296830]
42. Beyer FR, Kenny RPW, Johnson E, Caldwell DM, Garnett C, Rice S, et al. Practitioner and digitally delivered interventions for reducing hazardous and harmful alcohol consumption in people not seeking alcohol treatment: a systematic review and network meta-analysis. Addiction 2023;118(1):17-29 [FREE Full text] [doi: 10.1111/add.15999] [Medline: 35815387]
43. Goldstein BD. The precautionary principle also applies to public health actions. Am J Public Health 2001;91(9):1358-1361. [doi: 10.2105/ajph.91.9.1358] [Medline: 11527755]
44. Grustam AS, Vrijhoef HJM, Cordella A, Koymans R, Severens JL. Care coordination in a business-to-business and a business-to-consumer model for telemonitoring patients with chronic diseases. Int J Care Coord 2017;20(4):135-147 [FREE Full text] [doi: 10.1177/2053434517747908] [Medline: 29276610]
45. Onnela J. Opportunities and challenges in the collection and analysis of digital phenotyping data. Neuropsychopharmacology 2021;46(1):45-54 [FREE Full text] [doi: 10.1038/s41386-020-0771-3] [Medline: 32679583]
46. Cohen S. Perceived stress in a probability sample of the United States. In: Spacapan S, Oskamp S, editors. The Social Psychology of Health. Thousand Oaks, California: Sage Publications, Inc; 1988:31-67.
47. Peeters F, Berkhof J, Delespaul P, Rottenberg J, Nicolson NA. Diurnal mood variation in major depressive disorder. Emotion 2006;6(3):383-391. [doi: 10.1037/1528-3542.6.3.383] [Medline: 16938080]
48. Murray G. Diurnal mood variation in depression: a signal of disturbed circadian function? J Affect Disord 2007;102(1-3):47-53. [doi: 10.1016/j.jad.2006.12.001] [Medline: 17239958]
49. Suls J, Green P, Hillis S. Emotional reactivity to everyday problems, affective inertia, and neuroticism. Pers Soc Psychol Bull 1998;24(2):127-136. [doi: 10.1177/0146167298242002]
50. Epstein DH, Tyburski M, Kowalczyk WJ, Burgess-Hull AJ, Phillips KA, Curtis BL, et al. Prediction of stress and drug craving ninety minutes in the future with passively collected GPS data. NPJ Digit Med 2020;3(1):26 [FREE Full text] [doi: 10.1038/s41746-020-0234-6] [Medline: 32195362]

Abbreviations

AI: artificial intelligence
AUDIT-K: Alcohol Use Disorders Identification Test-Korean version
EMA: ecological momentary assessment
GAD-7: Generalized Anxiety Disorder-7
JITAI: just-in-time adaptive intervention
NPV: negative predictive value
PHQ-9: Patient Health Questionnaire-9
PPV: positive predictive value
PR AUC: area under the precision-recall curve
PSS-10: Perceived Stress Scale-10
ROC AUC: area under the receiver operating characteristic curve
SHAP: Shapley additive explanations
TRIPOD+AI: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis Plus Artificial Intelligence
XGBoost: extreme gradient boosting

Edited by Z Liang; submitted 21.Nov.2025; peer-reviewed by M Al-Agil, CY Su; comments to author 11.Jan.2026; revised version received 16.Jun.2026; accepted 18.Jun.2026; published 10.Jul.2026.

Please cite as:
Kwak JS, Lee HK, Jo SJ, Kwon JH, Kwon SJ, Kim Y, Lee H
Week-Ahead Prediction of High-Risk Drinking Episodes Among Young Adults Using Wearable Biosignals and Psychological Vulnerabilities: Prospective Observational Machine Learning Study
JMIR Mhealth Uhealth 2026;14:e88223
URL: https://mhealth.jmir.org/2026/1/e88223
doi:10.2196/88223
PMID:

©Jae Seok Kwak, Hae Kook Lee, Sun-Jin Jo, Jun Hyuk Kwon, Sun Jung Kwon, Yena Kim, Haejung Lee. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use,

XSL•FO
RenderX

distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Brief Web-Based Psychological Intervention for Depressive Symptoms Among Chinese Adolescents: Randomized Controlled Trial

Pei Liu[1*], MD; Xinwei Wang[2*], BMSc; Xiaoxia Duan[1], MD; Mingjing Situ[1], MD; Yujie Tao[1], MD; Zhaozhi Yang[1], MD; Shengnan Zhao[1], MD; Lei Li[1], MD; Sixun Li[1], MM; Yuchu Jiang[1], MM; Manxue Zhang[1], MD; Di Jing[1], MSc; Xinyi Yu[1], BS; Yi Huang[1,3], MD

[1]Mental Health Center, West China Hospital, Sichuan University, Chengdu, China

[2]Schulich School of Medicine and Dentistry, Western University, London, ON, Canada

[3]Laboratory of Child and Adolescent Psychiatry, Mental Health Center, West China Hospital, Sichuan University, Chengdu, China

[*]these authors contributed equally

Corresponding Author:
Yi Huang, MD
Mental Health Center
West China Hospital
Sichuan University
28 Dianxin South St., Wuhou Dist.
Chengdu
China
Phone: 86 19982012721
Email: huangyu@scu.edu.cn

Abstract

Background: Adolescents with depressive symptoms are at increased risk of social functional impairment and are more likely to develop major depressive disorder. Digital interventions offer advantages, such as high accessibility, especially for adolescents. However, evidence on the effectiveness of web-based psychological interventions in alleviating depressive symptoms among adolescents remains limited.

Objective: This study aims to develop a brief web-based psychological intervention tailored to the developmental and psychological characteristics of Chinese adolescents and to evaluate its effectiveness and influencing factors.

Methods: In a 2-arm randomized controlled trial, adolescents aged 12 to 18 years with depressive symptoms were recruited from high schools in China. Eligible participants were randomly assigned in a 1:1 ratio to the intervention group (n=212) or the control group (n=224). Participants in the intervention group received a 4-week brief web-based emotional cognitive training program, whereas those in the control group received a web-based psychoeducation program. Nonprofessional helpers provided minimal support through text messages or phone calls. The primary outcome was depressive symptom severity. Secondary outcomes included the depressive symptom remission rate, changes in anxiety symptoms, suicidal ideation, and resilience. Data were collected at baseline (T0), postintervention (T1), and at 1- and 3-month follow-ups (T2 and T3). Statistical analyses included analysis of covariance (ANCOVA), logistic regression analysis, and mediation analysis.

Results: Two-way repeated-measures ANCOVA revealed a significant time effect ($F_{3,1461}$=82.515; P<.001; η^2_p=0.140) on depressive symptom severity, with significantly lower scores at T1, T2, and T3 than at T0. No significant group effect ($F_{1,1461}$=1.039; P=.31; η^2_p=0.001) or time-by-group interaction ($F_{3,1461}$=2.424; P=.06; η^2_p=0.005) for depressive symptom severity was observed. Post hoc exploratory analysis showed that among adolescents with anxiety symptoms, the intervention group exhibited a higher depressive symptom remission rate at T1 than the control group (adjusted odds ratio 1.39, 95% CI 1.05-1.85; P=.02; number needed to treat=7). Additionally, the intervention group showed significantly greater improvements in the interpersonal support dimension of resilience at T1 (least squares mean difference=1.79, 95% CI 0.53-3.06; Cohen d=0.37, 95% CI 0.11-0.63; P=.006; false discovery rate–adjusted P=.04). Exploratory mediation analysis identified a significant indirect effect between the intervention and depressive symptom remission via the interpersonal support dimension of resilience at T1 (indirect effect=0.08, 95% CI 0.02-0.14; P=.009).

XSL•FO
RenderX

Conclusions: The brief web-based psychological intervention showed no significant difference from web-based psychoeducation in reducing depressive symptoms across the 3-month follow-up period. Post hoc exploratory analyses detected a potential between-group difference in depressive symptom remission among adolescents with comorbid anxiety symptoms, with interpersonal support showing a significant mediating association. These exploratory findings may provide new insights into identifying the target population suitable for brief web-based psychological interventions and inform the development of tailored interventions in the future.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2400090235; https://www.chictr.org.cn/showproj.html?proj=243991

(JMIR Mhealth Uhealth 2026;14:e86134) doi:10.2196/86134

KEYWORDS

adolescents; brief web-based psychological intervention; depressive symptoms; randomized controlled trial; RCT; web-based psychoeducation

Introduction

Major depressive disorder (MDD) is a prevalent mental disorder among the general population and has ranked fourth in the global burden of disease in recent years [1]. Evidence suggests that 19% of MDD cases first emerge during adolescence [2]. Given that adolescents are in a critical developmental period characterized by emotional and cognitive growth [3], they are particularly susceptible to various life events that may trigger depressive symptoms and potentially lead to the onset of MDD [4]. The global prevalence of depressive symptoms among adolescents during the COVID-19 pandemic was 25.2% [5]. Adolescents with depressive symptoms may experience impairments in academic and occupational functioning [6,7], and they remain at elevated risk of developing MDD compared with healthy individuals [8-10]. Therefore, psychological interventions targeting adolescents with depressive symptoms represent a critical approach for the early prevention and early intervention of depression.

Cognitive behavioral therapy (CBT) has been widely recommended as a standard intervention for adolescents with depressive symptoms or mild depressive disorders across many countries [11-14]. Although research suggests that CBT is the most evidence-based psychological intervention for depressive symptoms [15], several barriers hinder its widespread implementation, including high intervention costs, substantial time investment from therapists, and limited accessibility [16-18]. School-based CBT interventions can cost up to GBP £43 (US $32) per student [16]. During a full course of CBT treatment, therapists typically spend between 5 and 24 hours providing direct guidance to each participant [17]. Approximately 75% of adolescents with mental health problems have not accessed mental health services [18]. This reluctance to seek help is primarily attributed to a lack of accurate knowledge about mental disorders, high levels of stigma, significant economic burdens, and the inherent inaccessibility of psychotherapy itself [19].

Recent advances in psychotherapy have revealed that specific functional modules within effective interventions demonstrate significant therapeutic value in reducing symptoms [20]. Previous studies have demonstrated that behavioral activation, whether delivered alone or in combination with cognitive restructuring [21,22], as well as psychoeducation [23], mindfulness, relaxation training [24], and peer support modules, can effectively reduce depressive symptoms [25]. Digital therapeutics integrated with internet technology, as well as web-based psychological interventions derived from traditional psychological approaches, have increasingly attracted attention in the field of psychotherapy [26]. Indeed, recent studies have demonstrated that brief web-based psychotherapy provides significant therapeutic effects on depressive symptoms [27-29]. These interventions offer several advantages, including shorter treatment durations, reduced constraints of time and space, high accessibility, and strong scalability. For example, the World Health Organization has released a brief web-based psychological intervention called “Step-by-Step” for adult depressive symptoms [30]. Evidence indicates that “Step-by-Step” can effectively reduce depressive symptoms in both adults and university students, with effects maintained for up to a 3-month follow-up period [31,32].

For adolescent depressive symptoms, studies have demonstrated the efficacy of brief web-based psychological interventions in reducing depressive symptoms. A randomized controlled trial (RCT) involving 128 female adolescents with mild to moderate depressive symptoms evaluated a web-based social cognitive training program. Results showed that the intervention group exhibited significantly lower symptom severity than the blank control group at the 3-month follow-up [33]. A systematic review indicates that most current brief web-based interventions for adolescent depression incorporate psychoeducational components, whereas control groups are typically assigned to waitlist or blank controls [34]. Such designs may introduce placebo effects and overstate the ability to assess the specific therapeutic benefits of the interventions. Therefore, it is essential to implement psychoeducation as an active control condition within an RCT framework to more rigorously evaluate the efficacy of brief CBT-oriented web-based interventions.

This study aims to examine whether a brief web-based psychological intervention can reduce depressive symptoms in Chinese adolescents aged 12 to 18 years and assess the efficacy and sustainability of the intervention effects over a 1- to 3-month follow-up period. Additionally, this study seeks to explore the potential factors influencing the effectiveness of the intervention, such as resilience.

XSL•FO
RenderX

Methods

Ethical Considerations

The study protocol was approved by the Ethics Committee on Biomedical Research at West China Hospital, Sichuan University (approval number 2024-1225), and registered in the Chinese Clinical Trial Registry (ChiCTR2400090235). This study was reported in accordance with the CONSORT-EHEALTH checklist (Multimedia Appendix 1). All participants and their parents provided informed consent. Participants received US $15 for completing each of the 1-month and 3-month follow-up assessments. Participants were not compensated for engaging in the intervention. All data were deidentified before analysis to ensure participant confidentiality.

Study Design

This study used a 2-arm RCT design. From October 2024 to January 2025, eligible adolescents were recruited from 2 junior high schools and 1 senior high school in Chengdu. A total of 457 participants were initially enrolled, and 436 completed randomization and received the assigned interventions. The inclusion criteria were as follows: (1) age between 12 and 18 years; (2) a score of ≥10 on the Patient Health Questionnaire-8 (PHQ-8). Given the sensitivity of items related to "suicide" and "self-harm" in school-based universal screening, the final item of the PHQ-9 was omitted, and the PHQ-8 was used for screening; and (3) all adolescent participants volunteered to take part in the study, and both the participants and their parents provided written informed consent. Outcome assessments were conducted at 4 time points: baseline (T0), postintervention (T1), 1-month follow-up (T2), and 3-month follow-up (T3).

Procedure

From October 2024 to January 2025, adolescents aged 12 to 18 years with depressive symptoms were recruited from schools to participate in this study. A total of 457 participants were recruited, and 436 underwent randomization. A dedicated researcher stratified the participants by age, gender, school, and depressive symptom severity using a random number generator and subsequently allocated them to either the intervention group or the control group in a 1:1 ratio. Participants in both groups received their assigned interventions and completed outcome assessments at 4 time points. Participants and outcome assessors were not informed of their assigned group or the intervention content delivered to the comparator arm; however, formal verification of blinding status was not conducted during the trial.

Interventions

Participants in the intervention group received a 4-week, once-weekly web-based psychological intervention through a WeChat mini-program accessible via mobile phone or computer (a screenshot is provided in Figure S1 in Multimedia Appendix 2), whereas those in the control group received web-based psychoeducation with the same frequency and delivered via the identical platform. More details of the intervention modules are provided in Table S1 in Multimedia Appendix 2. Each participant was provided a unique login account and password. Participants were not informed about the content of the other group's intervention. Four postgraduate students majoring in psychology or psychiatry without medical licenses served as nonprofessional helpers, providing minimal support for platform access and login issues via text messages or phone calls. They did not perform any clinical assessment or therapeutic intervention. The average time commitment per participant was 15 minutes per week for both arms.

Safety Monitoring

All participants completed the PHQ-8 and the Beck Scale for Suicide Ideation–Chinese Version (BSI-CV) at screening and throughout the intervention period via the online platform. A licensed psychiatrist supervised the risk assessment. An automated alert system was activated when participants scored ≥6 on the BSI-CV to flag suicide risk for psychiatrist review. Based on this review, the psychiatrist determined final study eligibility and decided whether clinical referral was warranted. If a participant explicitly reported worsening depressive symptoms, expressed suicidal ideation, or asked for professional help, the nonprofessional helpers documented the participant's exact words and escalated the case within 2 hours to the supervising psychiatrist via a secure channel. The psychiatrist then contacted the participant directly, assessed the need for referral, and, if necessary, withdrew the participant from the study to arrange appropriate care.

Outcome Measurements

Overview

Data were collected via the same online platform used for the interventions, with participants completing self-report questionnaires at 4 time points (T0, T1, T2, and T3).

PHQ-8

The primary outcome was depressive symptom severity measured by the PHQ-8 [35]. A total score of ≥10 is considered clinically significant and suggests the presence of depressive symptoms [35]. In this study, remission of depressive symptoms was defined as a score of <10 [36]. Score changes were defined as follows: T1–T0 (postintervention minus baseline), T2–T0 (1-month follow-up minus baseline), and T3–T0 (3-month follow-up minus baseline).

Generalized Anxiety Disorder-7

Anxiety symptom severity was assessed using the Generalized Anxiety Disorder-7 [37]. A total score of ≥10 is considered indicative of clinically significant anxiety symptoms and may suggest the presence of an anxiety disorder [37].

BSI-CV

Suicidal ideation severity was assessed using the BSI-CV [38]. A total score of ≥6 demonstrates the highest classification accuracy for predicting future suicidal behavior [39].

Adolescent Resilience Scale

Resilience was evaluated using the Adolescent Resilience Scale [40]. It is a self-report instrument comprising 27 items organized into 5 dimensions: emotion regulation, interpersonal support, family support, positive cognition, and focused targeting. Higher

XSL•FO
RenderX

total scores reflect greater resilience and a stronger capacity to cope with stress and adversity.

Difficulties in Emotion Regulation Scale

Emotional regulation difficulties were assessed using the Difficulties in Emotion Regulation Scale [41]. It is a self-report measure consisting of 36 items organized into 6 dimensions: lack of emotional awareness, lack of emotional clarity, nonacceptance of emotional responses, impulse control difficulties, difficulties engaging in goal-directed behavior, and limited access to emotion regulation strategies. Higher total scores indicate greater difficulties in emotion regulation.

Adolescent Social Support Rating Scale

Social support levels were assessed using the Adolescent Social Support Rating Scale [42]. It is a self-report measure consisting of 17 items organized into 3 dimensions: subjective support, objective support, and support utilization. The average total score reflects the overall level of perceived social support, with higher scores indicating greater availability and more effective use of social support resources.

Socioeconomic Status Index

The socioeconomic status (SES) index was calculated using Ren's method [43], which used principal component analysis to generate standardized scores for parents' educational level, parents' occupation type, and family monthly income. Factor loadings were used as weights and multiplied by the corresponding *Z* scores; these weighted values were then summed and divided by the eigenvalue of the first principal component: $SES = (\beta_1 \times Z_{\text{parents' education level}} + \beta_2 \times Z_{\text{parents' occupation type}} + \beta_3 \times Z_{\text{family monthly income}})/f$. In this equation, β_1, β_2, and β_3 represent factor loadings, while f denotes the eigenvalue. The final derived formula in this study is: $SES = (0.84 \times Z_{\text{parents' education level}} + 0.84 \times Z_{\text{parents' occupation type}} + 0.68 \times Z_{\text{family monthly income}})/0.62$.

Acceptability and Adverse Events

Regarding the acceptability of the brief web-based psychological intervention program, we analyzed completion rates for both the intervention and control groups at T1, T2, and T3. Additionally, qualitative feedback regarding user experience and suggestions was collected from a subset of students. Technical issues encountered during implementation were reported by the backend developers.

Adverse events, including suicide attempts, self-harm, physical discomfort, and symptom deterioration, were systematically monitored and documented through reports from participants, guardians, or school personnel. Students who reported severe adverse events, such as suicide attempts or self-harm, were followed up and received appropriate psychological or medical support as necessary.

Statistical Analysis

The sample size was determined using G*Power software (version 3.1). Based on effect size data from a previous RCT, the between-group difference in postintervention depressive symptom scores assessed by the PHQ-9 scale yielded a Hedges *g* of 0.35 [31]. In that trial, the intervention group received a brief web-based psychological intervention, whereas the control group was offered brief web-based psychoeducation and referrals to local mental health services for professional assistance [31]. The statistical power analysis assumed a 2-tailed significance level of α=.05, a desired power of 80% (β=0.20), and a 1:1 allocation ratio between groups, yielding a minimum required sample size of 130 participants per group and a total of 260 participants for the study. To account for anticipated participant attrition, we applied a conservative dropout rate of 40%, informed by a pooled attrition rate of 35.5% estimated from a meta-analysis of smartphone-delivered psychological interventions [44]. The adjusted sample size was set at 217 participants per group, resulting in a target total sample size of 434. Eventually, 457 adolescents with depressive symptoms were initially recruited, and 21 participants withdrew before randomization, leaving a final sample of 436 randomized participants.

To assess whether the achieved sample size provided adequate power to detect the effect size observed in this trial, we performed a post hoc sensitivity analysis. Given the final sample of 436 participants (intervention group=212 and control group=224), a 2-tailed independent-samples *t* test with α=.05 and power (1–β)=0.80 could detect a minimum standardized mean difference of Cohen *d*=0.27, computed using the R (version 4.5.1; R Foundation for Statistical Computing) *pwr* package (version 1.3-0). This sensitivity analysis indicates that the trial could reliably identify small effects above Cohen *d*=0.27, whereas true treatment effects smaller than this threshold could not be statistically distinguished from 0.

Data were organized and analyzed using R, including the *lme4* (version 1.1-37) and *lmerTest* (version 3.1-3) packages, SPSS (version 25.0; IBM Corp), and Mplus (version 8.3; 2012-2018 Muthen & Muthen) software. Demographic characteristics and baseline outcome measures were compared between the intervention and control groups via 2-tailed independent-samples *t* tests and chi-square tests. For data with multiple time points (scale scores of depressive symptoms or other outcome indicators), 2-way repeated-measures analyses of covariance (ANCOVAs) were used, with time (T0, T1, T2, and T3), group (intervention vs control) effects, and the time-by-group interaction as independent variables. Baseline values of the outcome indicators, age, gender, educational stage, only-child status, family structure, and SES index were included as covariates. For change scores in depressive symptoms or other outcome indicators, 1-way ANCOVAs were used, with group as the independent variable and baseline values of the outcome indicators, age, gender, educational stage, only-child status, family structure, and SES index as covariates. Effect sizes for main effects and interactions were evaluated using η^2_p, with 0.01, 0.06, and 0.14 representing small, medium, and large effects, respectively [45]. Between-group effect sizes were calculated via least squares mean differences (LSMDs) and Cohen *d*, along with corresponding 95% CIs. Cohen *d* effect magnitudes were interpreted using conventional thresholds: small (*d*=0.20), medium (*d*=0.50), and large (*d*=0.80) [45]. Logistic regression analyses were conducted to examine the association between group assignment and depressive symptom

XSL•FO
RenderX

remission rates at each assessment time point (T1, T2, and T3), adjusting for baseline PHQ-8 score, age, gender, educational stage, only-child status, family structure, and SES index. Number needed to treat (NNT) and odds ratios (ORs) with 95% CIs were derived from the adjusted models [45]. A *P* value of <.05 was considered statistically significant throughout the study. To control for inflated type I error from multiple comparisons, we adopted stratified Benjamini-Hochberg false discovery rate (FDR) correction for indicators in between-group comparisons [46,47]. FDR correction was applied to scores and score changes of anxiety symptoms and suicidal ideation, as well as total and subscale scores and corresponding score changes for resilience, difficulties in emotion regulation, and social support.

To explore potential influencing factors of the intervention and the underlying psychological mechanisms, outcome variables with significant between-group differences were selected as mediators. With group assignment as the independent variable and outcome scores or rates as the dependent variable, we adjusted for age, gender, and baseline levels of each outcome. Mediation analyses were performed using maximum likelihood estimation and 1000 bootstrap resamples to evaluate the significance of indirect effects [48].

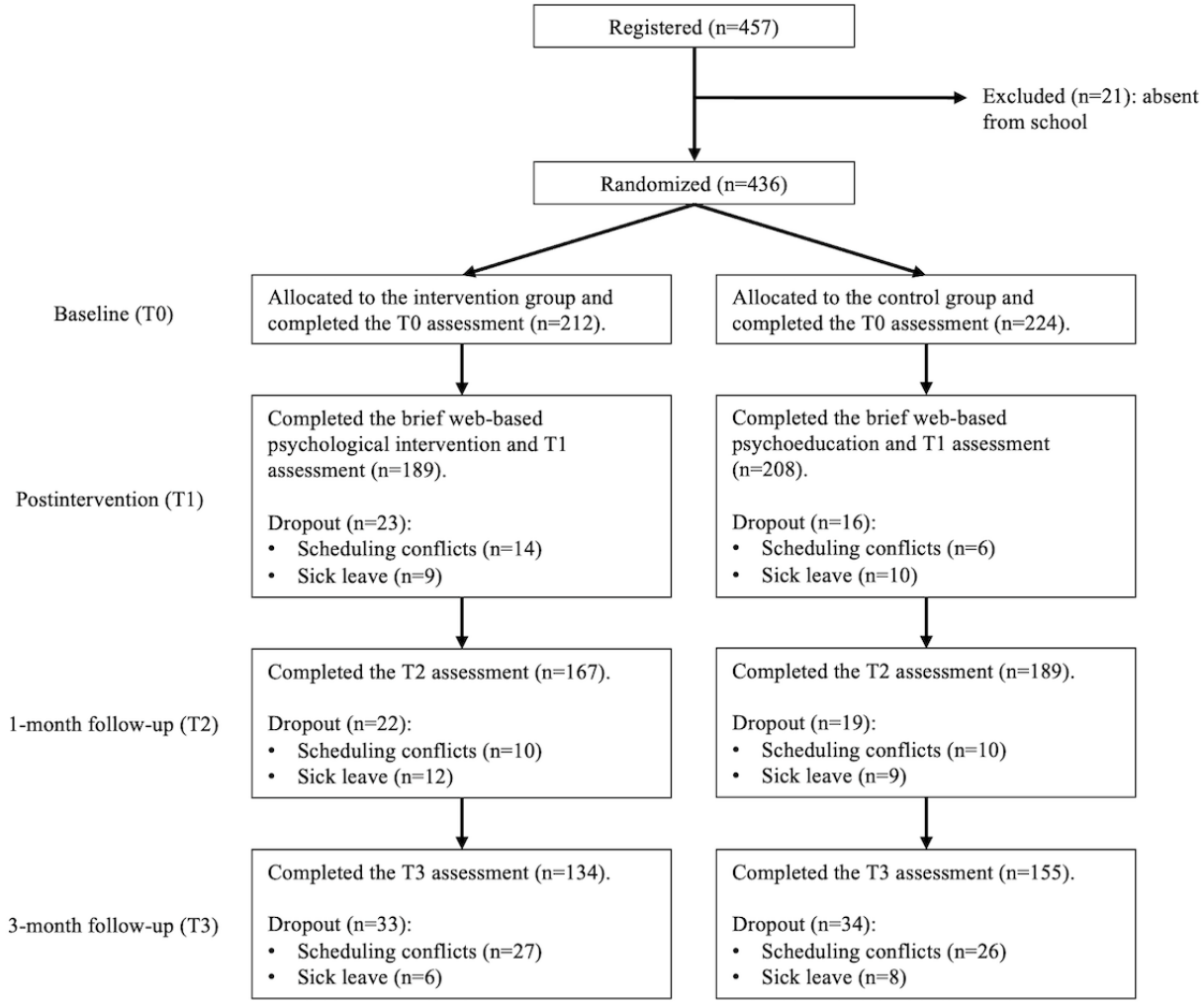
Results

Participants

From October 2024 to January 2025, adolescents aged 12 to 18 years with depressive symptoms were recruited from schools to participate in this study. A total of 457 participants were recruited, and 436 underwent randomization. Dropout rates differed across time points (Figure 1). For the intervention group (n=212), the rates were 10.9% (23/212) at T1, 10.4% (22/212) at T2, and 15.6% (33/212) at T3. For the control group (n=224), the corresponding rates were 7.1% (16/224) at T1, 8.5% (19/224) at T2, and 15.2% (34/224) at T3. Among the full sample of 436 participants, the mean age was 15.02 (SD 1.49) years. In terms of demographic characteristics, 75.2% (328/436) of participants were female, and 24.8% (108/436) were male; 30.7% (134/436) were junior high school students, while 69.3% (302/436) were high school students. Additionally, 27.3% (119/436) of participants were only children. Regarding family structure, 67% (292/436) of participants came from 2-parent families (residing with both biological or adoptive parents), 9.4% (41/436) from single-parent families (residing with only 1 parent), 6.4% (28/436) from stepfamilies (residing with 1 biological parent and a stepparent), and 17.2% (75/436) did not live with either parent but resided with grandparents or other relatives instead. The SES index ranged from –5.18 to 9.41, with a mean value of 0 (SD 2.99). The mean baseline PHQ-8 score in the full sample was 14.43 (SD 3.02), ranging from 10 to 24. No significant differences were found between the intervention and control groups regarding age, gender, educational stage, only-child status, family structure, SES index, and any outcome indicators (all $P \geq .05$; Table 1).

XSL•FO
RenderX

Figure 1. Flow diagram.



XSL•FO
RenderX

Table 1. Demographic characteristics and baseline levels of outcome indicators for the intervention and control groups.

Characters and baseline data	Intervention group (n=212)	Control group (n=224)	*t* test (*df*)/chi-square (*df*)	*P* value
Age (years), mean (SD)	15.07 (1.43)	14.98 (1.56)	0.617 (434)[a]	.54
Gender, n (%)			0.599 (1)[b]	.44
Male	56 (26.4)	52 (23.2)		
Female	156 (73.6)	172 (76.8)		
Educational stage, n (%)			0.200 (1)[b]	.65
Junior high school	63 (29.7)	71 (31.7)		
Senior high school	149 (70.3)	153 (68.3)		
Only child status, n (%)			2.860 (1)[b]	.09
Only child	50 (23.6)	69 (30.8)		
Non-only child	162 (76.4)	155 (69.2)		
Family structure, n (%)[c]			3.939 (3)[b]	.27
2-parent family	139 (65.6)	153 (68.3)		
Single-parent family	16 (7.5)	25 (11.2)		
Stepfamily	17 (8)	11 (4.9)		
No parent present	40 (18.9)	35 (15.6)		
Socioeconomic status index, mean (SD)	–0.09 (3.02)	0.08 (2.97)	–0.600 (434)[a]	.55
Depressive symptom severity, mean (SD)	14.62 (3.01)	14.26 (3.04)	1.240 (434)[a]	.22
Anxiety symptom severity, mean (SD)	11.43 (4.89)	11.35 (5.21)	0.168 (434)[a]	.87
Presence of anxiety symptoms[d], n (%)			0.003 (1)[b]	.96
With anxiety symptoms	134 (63.2)	141 (62.9)		
Without anxiety symptoms	78 (36.8)	83 (37.1)		
Suicidal ideation severity, mean (SD)	13.48 (7.63)	12.98 (8.66)	0.633 (431.81)[a]	.53
Resilience, mean (SD)	72.30 (15.13)	73.77 (17.56)	–0.940 (434)[a]	.35
Emotion regulation	13.95 (4.56)	14.23 (5.41)	–0.595 (426.99)[a]	.55
Interpersonal support	15.11 (4.97)	15.96 (5.36)	–1.709 (434)[a]	.09
Family support	16.59 (5.03)	16.52 (5.32)	0.140 (434)[a]	.09
Positive cognition	12.69 (3.41)	12.76 (3.65)	–0.204 (434)[a]	.84
Focused targeting	13.96 (4.37)	14.30 (4.41)	–0.814 (434)[a]	.42
Difficulties in emotion regulation, mean (SD)	113.80 (18.40)	114.35 (20.89)	–0.292 (434)[a]	.77
Lack of emotional awareness	16.33 (4.77)	16.00 (4.69)	0.739 (434)[a]	.46
Lack of emotional clarity	15.37 (3.29)	15.43 (3.81)	–0.164 (430.53)[a]	.87
Nonacceptance of emotional responses	18.98 (5.32)	19.26 (5.51)	–0.544 (434)[a]	.59
Impulse control difficulties	19.09 (4.96)	19.62 (5.38)	–1.052 (434)[a]	.29
Difficulties engaging in goal-directed behavior	17.54 (4.17)	17.77 (4.49)	–0.542 (434)[a]	.59
Limited access to emotion regulation strategies	26.47 (6.37)	26.27 (7.06)	0.309 (434)[a]	.76
Social support, mean (SD)	3.04 (0.93)	3.17 (1.02)	–1.352 (433.27)[a]	.18
Subjective support	3.01 (1.12)	3.12 (1.13)	–1.030 (434)[a]	.30

XSL•FO
RenderX

Characters and baseline data	Intervention group (n=212)	Control group (n=224)	*t* test (*df*)/chi-square (*df*)	*P* value
Objective support	3.25 (1.03)	3.34 (1.07)	–0.842 (434)[a]	.40
Support utilization	2.86 (1.06)	3.04 (1.20)	–1.660 (432.06)[a]	.10

[a]Indicates 2-tailed independent-samples *t* test.

[b]Indicates chi-square test.

[c]Family structure was classified into 4 types. 2-parent family: residing with both biological or adoptive parents; single-parent family: residing with only 1 parent; stepfamily: residing with 1 biological parent and a stepparent; no parent present: not living with either parent but staying with grandparents or other relatives.

[d]The criteria for determining the presence of anxiety symptoms were as follows: a total score of ≥10 on the Generalized Anxiety Disorder-7 indicates the presence of anxiety symptoms, whereas a score of <10 indicates the absence of anxiety symptoms.

Main Results

For all participants, Table 2 presents the ANCOVA-adjusted estimated means and standard errors of each continuous outcome indicator at each time point for the intervention and control groups. The primary outcome (depressive symptom severity, measured by the PHQ-8 score) demonstrated a significant time effect through 2-way repeated-measures ANCOVA ($F_{3,1461}$=82.515; P<.001; η^2_p=0.140), but the group effect ($F_{1,1461}$=1.039; P=.31; η^2_p=0.001) and time-by-group interaction ($F_{3,1461}$=2.424; P=.06; η^2_p=0.005) were not significant. Post hoc comparisons revealed that depressive symptom severity at T1, T2, and T3 was significantly lower than that at T0 in both the intervention and control groups (T1–T0: Cohen d=–0.91, 95% CI –1.05 to –0.77; T2–T0: Cohen d=–0.90, 95% CI –1.05 to –0.76; T3–T0: Cohen d=–0.89, 95% CI –1.05 to –0.74). Significant time effects were also observed for almost all secondary outcomes (P_s<.05; Table S2 in Multimedia Appendix 2). Following FDR correction for multiple testing, none of the secondary outcomes exhibited a statistically significant group effect or time-by-group interaction (all FDR-adjusted P≥.05; Table S2 in Multimedia Appendix 2).

XSL•FO
RenderX

Table 2. Estimated means and standard errors for continuous outcome indicators at each time point for the intervention and control groups (adjusted for baseline covariates using analysis of covariance).

Outcome indicator and group[a]	Baseline, mean (SE)	Posttest, mean (SE)	1-month follow-up, mean (SE)	3-month follow-up, mean (SE)
Depressive symptom severity				
Intervention	14.77 (0.43)	9.54 (0.45)	10.04 (0.48)	10.73 (0.53)
Control	14.65 (0.42)	10.25 (0.44)	9.79 (0.45)	9.23 (0.50)
Anxiety symptom severity				
Intervention	11.98 (0.38)	9.11 (0.40)	9.16 (0.42)	8.96 (0.47)
Control	11.91 (0.38)	9.36 (0.39)	8.57 (0.40)	8.26 (0.44)
Suicidal ideation severity				
Intervention	13.58 (0.52)	10.88 (0.55)	10.09 (0.58)	9.04 (0.64)
Control	13.41 (0.51)	10.53 (0.53)	9.34 (0.55)	9.22 (0.60)
Resilience dimension				
Resilience				
Intervention	72.00 (0.89)	78.60 (0.94)	79.30 (0.99)	78.70 (1.09)
Control	72.50 (0.88)	77.00 (0.91)	77.40 (0.94)	77.30 (1.03)
Emotion regulation				
Intervention	13.60 (0.34)	16.90 (0.37)	16.80 (0.38)	17.00 (0.42)
Control	13.70 (0.34)	15.90 (0.35)	16.00 (0.37)	16.70 (0.40)
Interpersonal support				
Intervention	15.10 (0.30)	17.40 (0.32)	17.50 (0.34)	12.50 (0.37)
Control	15.50 (0.30)	16.50 (0.31)	16.90 (0.32)	11.70 (0.35)
Family support				
Intervention	16.30 (0.32)	17.00 (0.34)	17.20 (0.35)	16.80 (0.39)
Control	16.30 (0.31)	17.20 (0.32)	17.20 (0.34)	16.70 (0.37)
Positive cognition				
Intervention	12.80 (0.28)	12.50 (0.30)	12.60 (0.31)	17.70 (0.35)
Control	12.80 (0.28)	12.50 (0.29)	12.60 (0.30)	17.90 (0.33)
Focused targeting				
Intervention	14.20 (0.34)	14.70 (0.36)	15.00 (0.38)	14.50 (0.42)
Control	14.30 (0.34)	15.00 (0.35)	14.70 (0.36)	14.30 (0.39)
Emotion regulation dimension				
Difficulties in emotion regulation				
Intervention	115.00 (1.34)	113.00 (1.42)	108.00 (1.49)	110.00 (1.65)
Control	115.00 (1.32)	113.00 (1.37)	111.00 (1.42)	110.00 (1.56)
Lack of emotional awareness				
Intervention	16.00 (0.40)	16.50 (0.43)	17.30 (0.45)	17.00 (0.50)
Control	15.80 (0.40)	16.60 (0.41)	16.80 (0.43)	16.80 (0.47)
Lack of emotional clarity				
Intervention	15.50 (0.22)	15.60 (0.23)	14.90 (0.24)	15.30 (0.27)
Control	15.50 (0.22)	15.70 (0.22)	14.90 (0.23)	15.30 (0.26)
Nonacceptance of emotional responses				
Intervention	19.40 (0.45)	19.30 (0.48)	17.80 (0.50)	18.00 (0.55)
Control	19.50 (0.44)	19.30 (0.46)	18.60 (0.48)	18.10 (0.52)

XSL•FO
RenderX

Outcome indicator and group[a]	Baseline, mean (SE)	Posttest, mean (SE)	1-month follow-up, mean (SE)	3-month follow-up, mean (SE)
Impulse control difficulties				
Intervention	19.50 (0.35)	19.20 (0.38)	18.20 (0.39)	18.40 (0.44)
Control	19.70 (0.35)	19.40 (0.36)	18.50 (0.38)	18.40 (0.41)
Difficulties engaging in goal-directed behavior				
Intervention	17.80 (0.29)	16.20 (0.31)	15.60 (0.32)	16.00 (0.35)
Control	17.90 (0.28)	16.40 (0.29)	16.40 (0.31)	16.20 (0.33)
Limited access to emotion regulation strategies				
Intervention	26.90 (0.48)	26.00 (0.51)	24.40 (0.53)	24.80 (0.59)
Control	26.90 (0.47)	26.10 (0.49)	25.50 (0.51)	25.50 (0.56)
Social support dimension				
Social support				
Intervention	3.07 (0.06)	3.12 (0.07)	3.25 (0.07)	3.32 (0.08)
Control	3.11 (0.06)	3.08 (0.07)	3.05 (0.07)	3.22 (0.07)
Subjective support				
Intervention	3.02 (0.07)	3.13 (0.08)	3.24 (0.08)	3.28 (0.09)
Control	3.07 (0.07)	3.09 (0.07)	3.05 (0.08)	3.25 (0.08)
Objective support				
Intervention	3.27 (0.07)	3.15 (0.07)	3.33 (0.08)	3.35 (0.08)
Control	3.31 (0.07)	3.20 (0.07)	3.17 (0.07)	3.30 (0.08)
Support utilization				
Intervention	2.89 (0.07)	3.06 (0.08)	3.17 (0.08)	3.29 (0.09)
Control	2.97 (0.07)	2.97 (0.07)	2.93 (0.08)	3.12 (0.08)

[a]At baseline, postintervention, 1-month follow-up, and 3-month follow-up, the intervention group included 212, 189, 167, and 134 participants, respectively, while the control group included 224, 208, 189, and 155 participants, respectively.

Furthermore, a significant between-group difference was observed in the change score for the interpersonal support dimension of resilience from T0 to T1 (intervention group: mean change 2.03, SD 0.53; control group: mean change 0.64, SD 0.52; LSMD=1.39, 95% CI 0.42-2.35; Cohen *d*=0.29, 95% CI 0.09-0.49; *P*=.005; FDR-adjusted *P*=.03) through 1-way ANCOVA (Table S3 in Multimedia Appendix 2). Moreover, there was no significant difference in the depressive symptom remission rate between the groups at any time point through logistic regression analysis (Table S4 in Multimedia Appendix 2).

In terms of acceptability, completion rates for both the intervention and control groups across all time points are presented in Table 3. Chi-square tests revealed no significant between-group differences at any assessment point (all *P*≥.05). Furthermore, some participants reported that the visual and textual content was engaging. However, participants reported difficulty finding a quiet environment for audio playback, and the password reset process was troublesome when they forgot their credentials. The backend developers resolved the network latency issue during intervention delivery and shortened the password reset procedure to improve user experience.

No severe adverse events, such as suicide attempts or self-harm behaviors, were reported during the intervention or follow-up period. In this trial, 3 participants withdrew because of self-reported worsening depressive symptoms that required clinical referral. No physical discomfort related to program use was reported.

Table 3. Comparison of completion rates between the intervention and control groups across different time points.

Time points	Intervention group (n=212), n (%)	Control group (n=224), n (%)	Chi-square (*df*)	*P* value
Postintervention	189 (89.2)	208 (92.9)	1.837 (1)	.18
1-month follow-up	167 (78.8)	189 (84.4)	2.281 (1)	.13
3-month follow-up	134 (63.2)	155 (69.2)	1.748 (1)	.19

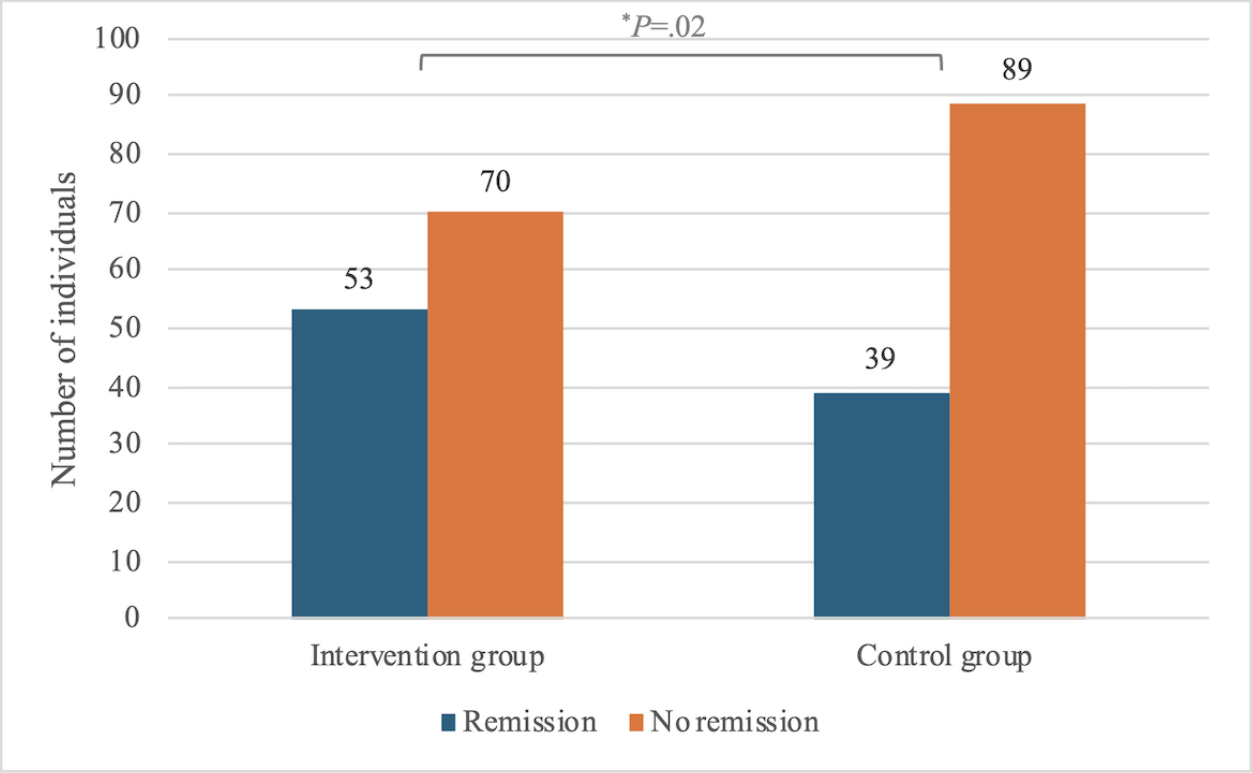
XSL•FO
RenderX

Post Hoc Exploratory Analysis Results

Adolescents with depressive symptoms may exhibit substantial individual heterogeneity, and intervention effects may differ between those with and without anxiety symptoms. Therefore, participants were divided into 2 subgroups (with anxiety symptoms vs without anxiety symptoms), and the statistical analyses were performed separately within each subgroup. Results are reported in the subsequent section. Demographic characteristics and baseline scale score differences between the intervention and control groups within each subgroup are presented in Table S5 in Multimedia Appendix 2.

Among adolescents with anxiety symptoms, logistic regression analyses indicated a significantly higher depressive symptom remission rate at T1 in the intervention group relative to the control group (intervention: 53/123, 43.1%; control: 39/128, 30.5%; adjusted OR=1.39, 95% CI 1.05-1.85; *P*=.02; NNT=7; Figure 2 and Table S4 in Multimedia Appendix 2). No significant between-group differences were detected in raw scores or changes in depressive symptom severity at any time point in this subgroup (all *P*≥.05; Tables S6-S8 in Multimedia Appendix 2). One-way ANCOVA further showed that the intervention group exhibited substantially greater improvements in the interpersonal support dimension of resilience from T0 to T1 (intervention group: mean change 2.75, SD 0.69; control group: mean change 0.96, SD 0.64; LSMD=1.79, 95% CI 0.53-3.06; Cohen *d*=0.37, 95% CI 0.11-0.63; *P*=.006; FDR-adjusted *P*=.04; Table S8 in Multimedia Appendix 2).

Figure 2. The number of individuals with remission and no remission of depressive symptoms in the intervention and control groups among adolescents with anxiety symptoms at postintervention. Remission of depressive symptoms is defined as a Patient Health Questionnaire-8 (PHQ-8) score of <10 (below the clinical cutoff point). **P*<.05; ***P*<.01; ****P*<.001.

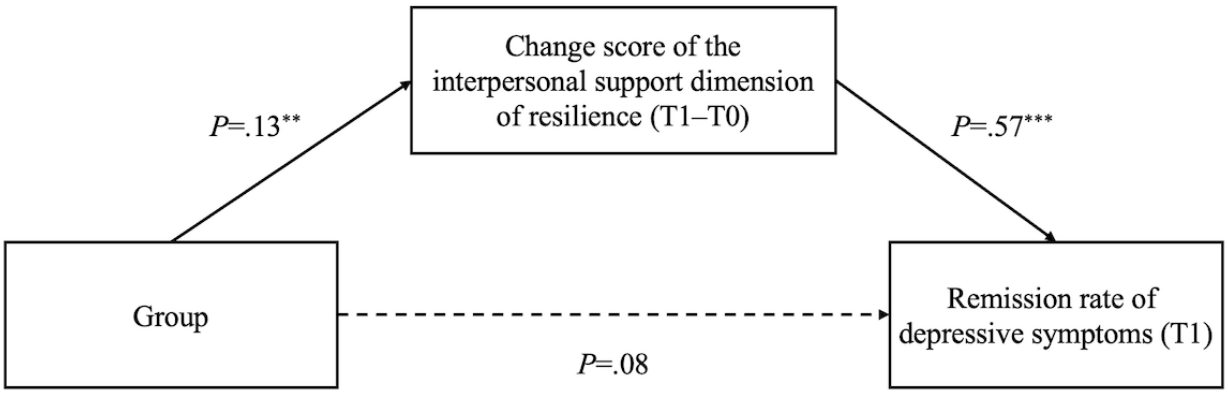


Building on these findings observed among adolescents with anxiety symptoms, we constructed an exploratory mediation model. The analysis revealed a statistical association between group assignment and changes in the interpersonal support dimension of resilience from T0 to T1 (β=0.13, 95% CI 0.03-0.23; *P*=.008). In turn, changes in interpersonal support were associated with depressive symptom remission at T1 (β=0.57, 95% CI 0.41-0.69; *P*<.001). A significant indirect effect was observed across the sequence: group → changes in interpersonal support from T0 to T1 → depressive symptom remission at T1 (β=0.08, 95% CI 0.02-0.14; *P*=.009; Figure 3). The proportion mediated was 49.8% (95% CI 15.5%-169.5%). The upper bound exceeding 100% reflects sampling variability and the instability of ratio-based mediation metrics when the total effect is not statistically significant. We therefore focus on the indirect effect as the primary evidence for mediation.

Among adolescents without anxiety symptoms, no significant between-group differences were found in depressive symptom severity, any other outcome measures, or depressive symptom remission rates (all *P*≥.05; all FDR-adjusted *P*≥.05; Tables S4, S6, S9, and S10 in Multimedia Appendix 2).

XSL•FO
RenderX

Figure 3. The mediating mechanism on remission of depressive symptoms. The group is a binary categorical variable, with the "intervention group" coded as 1 and the "control group" coded as 0. T0 and T1 refer to the baseline and postintervention time points, respectively. *P<.05; **P<.01; ***P<.001.



Discussion

Principal Findings

This study reveals that both the brief web-based psychological intervention and web-based psychoeducation are associated with short-term reductions in adolescent depressive symptoms, with improvements sustained from postintervention to the 1- to 3-month follow-up period. No significant between-group differences were observed in any outcome measures. However, unplanned, hypothesis-generating post hoc exploratory analyses among adolescents with anxiety symptoms indicated that the intervention group exhibited a higher depressive symptom remission rate and greater improvement in the interpersonal support dimension of resilience at T1 relative to the control group. Additionally, the association between the intervention and depressive symptom remission was partially explained by improvements in interpersonal support. Given that these findings came from unplanned post hoc analyses, they should be interpreted cautiously and require further validation in rigorous trials.

Both groups exhibited significant reductions in depressive symptom severity up to the 3-month follow-up, which aligns with prior studies of web-based psychological interventions for youth. RCTs have shown that when the control group receives web-based psychoeducation or supportive therapy, both the intervention and control groups exhibit significant decreases in depressive symptom scores, with effects lasting up to the 1- to 3-month follow-up [49,50]. These findings align with the present results, suggesting that web-based psychoeducation alone may relieve depressive symptoms via emotion recognition, regulation training, and social skills modules, which may help adolescents build self-confidence and improve emotional well-being, thereby reducing depressive symptoms. Consistent with several previous trials, we found no group differences in depressive symptom severity when the control group received web-based psychoeducation [51,52]. By contrast, when the control group received treatment as usual or a blank control, the intervention group showed significantly lower depressive symptom scores postintervention or at the 3- to 6-month follow-up [33,53,54]. Notably, significant between-group differences were observed in several subdimensions of resilience and social support. However, these differences did not survive after multiple-comparison correction (Tables S2 and S3 in Multimedia Appendix 2), indicating that they should be interpreted as exploratory rather than confirmatory. Nonetheless, the observed trend toward improved adolescent resilience and social support following web-based psychological intervention aligns with prior evidence [55,56]. This may be attributed to the combined effects of identifying personal strengths within the psychoeducation module of the brief web-based psychological intervention, as well as the skill modules of behavioral activation, cognitive restructuring, peer support, mindfulness, and relaxation training [57-59]. In summary, the brief web-based intervention showed no advantages over web-based psychoeducation in reducing depressive symptoms across all assessed time points. Future work should extend follow-up to 6 to 12 months to evaluate long-term effects and add a blank control group to identify the specific components of psychoeducation that contribute to its effectiveness.

Among adolescents with anxiety symptoms, post hoc analyses revealed a higher postintervention depressive symptom remission rate in the intervention group, whereas no group difference was found in overall depressive symptom changes from baseline to postintervention and the 1- to 3-month follow-ups. Previous studies have reported contradictory findings regarding effects on depressive symptom severity and remission rate. Two RCTs with larger sample sizes (n=160 and n=187) yielded results similar to those of this study, showing significantly higher depressive symptom remission rates in the intervention group but no significant between-group differences in depressive symptom scores [60,61]. In contrast, an RCT with a smaller sample size (n=34) reported the opposite results: significant differences in depressive symptom severity but not in depressive symptom remission rate [50]. In these studies, the control groups received either web-based psychoeducation or treatment as usual, and criteria for depressive symptom remission were both defined as a score less than the clinical cutoff on the depression scales. These findings suggest that larger sample sizes may be more likely to detect significant differences in depressive symptom remission rates. Notably, using a PHQ-8 cutoff of 10 for both enrollment and remission may lead to biased classification due to mean regression. To address this concern, we excluded participants with baseline PHQ-8 scores of 10 to 11 (full sample: 39/436, 8.9%; subgroup

XSL•FO
RenderX

with anxiety symptoms: 17/275, 6.2%) and reconducted sensitivity analyses for all significant outcomes. The results remained statistically significant, demonstrating that our main findings are relatively robust (Multimedia Appendix 2).

Both the full sample and the subgroup with anxiety symptoms showed small but statistically significant between-group differences in changes in the interpersonal support dimension of resilience from baseline to postintervention (full sample: Cohen *d*=0.29, 95% CI 0.09-0.49; subgroup with anxiety symptoms: Cohen *d*=0.37, 95% CI 0.11-0.63). Despite statistical significance, the small effect size suggests limited clinical practicality, as such minor changes are unlikely to improve daily functioning or clinical outcomes. This exploratory result can only serve as a preliminary reference for intervention design, rather than evidence of robust clinical efficacy. Mediation analyses further revealed that improved interpersonal support partially mediated the association between the intervention and depressive symptom remission in the anxiety subgroup. As a core component of adolescent resilience, interpersonal support reflects the ability to seek comfort and express emotions through social connections [40]. The peer support module in our intervention may have strengthened this capacity, helping adolescents access more psychological resources, cope better with stress, and recover from distress [62]. A prior RCT involving 91 adolescents with MDD found that higher baseline levels of resilience predicted lower depressive symptom levels after receiving CBT [63]. These findings indicate that the interpersonal support dimension of resilience not only serves as a protective factor against depressive symptoms before intervention but may also predict better postintervention outcomes. Therefore, future web-based psychological interventions should prioritize resilience-building strategies. Unfortunately, these benefits were not maintained at the 1- to 3-month follow-ups, highlighting the need for longer follow-up assessments to evaluate long-term efficacy.

Compared with adolescents with anxiety symptoms, adolescents without anxiety symptoms showed no significant between-group differences in any outcome measures, which is consistent with previous research. In an RCT involving 151 adults with MDD, participants in the intervention group received traditional face-to-face cognitive therapy, whereas those in the control group received interpersonal psychotherapy. The study found that anxiety symptoms positively predicted greater reductions in depressive symptoms in the cognitive therapy group at postintervention [64]. Another RCT including 268 adults with persistent depressive disorder found that participants receiving face-to-face CBT exhibited significantly lower depressive symptom scores and a higher remission rate among those with comorbid anxiety disorders than among those without comorbid anxiety disorders [65]. The co-occurrence of depressive and anxiety symptoms during adolescence is relatively common [66]. Another study showed that among youth with anxiety disorders receiving self-guided CBT, the subgroup exhibiting severe depressive and anxiety symptoms, poor sleep quality, and high levels of negative emotion experienced greater alleviation of both depressive and anxiety symptom severity following intervention [67]. Multiple large-scale epidemiological studies in China have reported a comorbidity rate of depression and anxiety of 11.3% among adolescents [68]. Compared with individuals presenting with either depressive or anxiety symptoms alone, those with both depressive and anxiety symptoms may experience more severe impairments in social functioning, underscoring the need for targeted interventions in this population [69]. The observed intervention effects exhibited considerable heterogeneity within the adolescent population. Future research should aim to refine intervention protocols and develop more tailored psychological intervention strategies to enhance treatment outcomes for specific subgroups.

In terms of intervention acceptability, the overall completion rate reached 89.2% at postintervention, comparable to the average rate of 88.3% reported in similar web-based psychological interventions [70]. We summarized participants' reasons for withdrawal: some dropped out because of academic conflicts, for which we coordinated with teachers to adjust schedules and facilitate participation; others withdrew because of technical failures, which were resolved promptly by the technical team. Future improvements should optimize intervention content, technical delivery, and implementation based on withdrawal patterns. No severe adverse events, such as suicide attempts or self-harm behaviors, occurred during the trial.

Limitations

First, this study lacked a no-intervention blank control group. Both the web-based psychological intervention and psychoeducation reduced adolescents' depressive symptoms, yet no significant differences were found between the 2 interventions. Adding a blank control group in future research will help identify the effective components of psychoeducation. Second, no formal clinical diagnostic assessments were conducted, so the intervention efficacy among adolescents with diagnosed depression remains unknown. Standard diagnostic procedures are needed in future studies to stratify participants accurately and design targeted interventions. Third, data on family-level factors, including siblings' mental health and parental physical or psychiatric illness, were not collected. These factors influence the development of adolescent depression and intervention responses [71-74]. Systematic assessment of family characteristics is warranted to explore how family dynamics interact with online interventions. Fourth, rigorous allocation concealment was implemented during randomization to prevent predictable group assignment prior to enrollment. Nevertheless, we did not conduct formal checks to confirm whether participants and outcome assessors remained unaware of group allocation after randomization, which may introduce performance bias and detection bias. Future trials should strictly enforce and verify blinding protocols. Fifth, a post hoc sensitivity analysis revealed that, given the final sample size (n=436), the study could detect only between-group differences of Cohen $d \geq 0.27$. All observed effect sizes for the primary outcome were below this threshold. Therefore, the study was underpowered to detect the true effect size achievable when comparing 2 active interventions, which may partly explain the nonsignificant group effects and interactions. Future trials should consider larger sample sizes or alternative designs to detect small but potentially meaningful effects. Finally, follow-up was limited to 1 and 3 months after the intervention. Extended

XSL•FO
RenderX

follow-up of 6 to 12 months is necessary to assess the long-term outcomes of this intervention.

Conclusions

In this RCT comparing a brief web-based psychological intervention with an active web-based psychoeducation control among adolescents with depressive symptoms, the intervention showed no superior effects relative to the active control in reducing depressive symptom severity at any follow-up time point. However, in exploratory subgroup analyses among adolescents with comorbid anxiety symptoms, the intervention suggested a possible improvement in depressive symptom remission at postintervention, which was partially mediated by improvements in the interpersonal support dimension of resilience at that time point. These findings suggest that future studies should strengthen the peer support components, optimize intervention content to deliver targeted psychological support tailored to adolescents with and without anxiety symptoms, and extend the follow-up period to assess the long-term effects of the intervention on depressive symptom reduction.

Acknowledgments

The authors gratefully acknowledge funding support for this research from the Brain Science and Brain-like Intelligence Technology–National Science and Technology Major Project (grant number 2022ZD0209100/2022ZD0209104) and the Achievement Transformation Demonstration Project of Chengdu City (grant number 2022-YF09-00010-SN). Generative AI tools were used for English editing and checking for typographical errors.

Funding

This research was funded by the Brain Science and Brain-like Intelligence Technology–National Science and Technology Major Project (grant number 2022ZD0209100/2022ZD0209104) and the Achievement Transformation Demonstration Project of Chengdu City (grant number 2022-YF09-00010-SN).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: PL, YH, XD, XY
Data curation: PL, XD
Formal analysis: PL, XW, YH
Funding acquisition: YH
Investigation: MS, YT, PL, ZY
Methodology: PL, YH, XD, XY
Project administration: PL, XW, YH
Resources: XD, MZ, DJ
Software: XY
Supervision: PL, XW, YH
Validation: YH, XD, SZ, LL
Visualization: SL, YJ
Writing—original draft: PL
Writing—review and editing: YH

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2926 KB - mhealth_v14i1e86134_app1.pdf]

Multimedia Appendix 2
Supplementary materials: detailed intervention information and complete results encompassing secondary outcome measures and post hoc exploratory analyses.
[DOCX File , 489 KB - mhealth_v14i1e86134_app2.docx]

References

XSL•FO
RenderX

1. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021;398(10312):1700-1712 [FREE Full text] [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
2. Rohde P, Lewinsohn PM, Klein DN, Seeley JR, Gau JM. Key characteristics of major depressive disorder occurring in childhood, adolescence, emerging adulthood, adulthood. Clin Psychol Sci 2013;1(1):10.1177/2167702612457599. [doi: 10.1177/2167702612457599] [Medline: 24273703]
3. Hagan CC, Graham JME, Wilkinson PO, Midgley N, Suckling J, Sahakian BJ, et al. Neurodevelopment and ages of onset in depressive disorders. Lancet Psychiatry 2015;2(12):1112-1116. [doi: 10.1016/S2215-0366(15)00362-4] [Medline: 26613851]
4. Graber JA. Pubertal timing and the development of psychopathology in adolescence and beyond. Horm Behav 2013;64(2):262-269. [doi: 10.1016/j.yhbeh.2013.04.003] [Medline: 23998670]
5. Racine N, McArthur BA, Cooke JE, Eirich R, Zhu J, Madigan S. Global prevalence of depressive and anxiety symptoms in children and adolescents during COVID-19: a meta-analysis. JAMA Pediatr 2021;175(11):1142-1150. [doi: 10.1001/jamapediatrics.2021.2482] [Medline: 34369987]
6. Lee RSC, Hermens DF, Redoblado-Hodge MA, Naismith SL, Porter MA, Kaur M, et al. Neuropsychological and socio-occupational functioning in young psychiatric outpatients: a longitudinal investigation. PLoS One 2013;8(3):e58176 [FREE Full text] [doi: 10.1371/journal.pone.0058176] [Medline: 23469268]
7. Thomas A, Jing M, Chen H, Crawford EL. Taking the good with the bad? Social media and online racial discrimination influences on psychological and academic functioning in Black and Hispanic youth. J Youth Adolesc 2023;52(2):245-257. [doi: 10.1007/s10964-022-01689-z] [Medline: 36229754]
8. Cuijpers P, Smit F. Subthreshold depression as a risk indicator for major depressive disorder: a systematic review of prospective studies. Acta Psychiatr Scand 2004;109(5):325-331 [FREE Full text] [doi: 10.1111/j.1600-0447.2004.00301.x] [Medline: 15049768]
9. Hill RM, Pettit JW, Lewinsohn PM, Seeley JR, Klein DN. Escalation to major depressive disorder among adolescents with subthreshold depressive symptoms: evidence of distinct subgroups at risk. J Affect Disord 2014;158:133-138. [doi: 10.1016/j.jad.2014.02.011] [Medline: 24655777]
10. Jinnin R, Okamoto Y, Takagaki K, Nishiyama Y, Yamamura T, Okamoto Y, et al. Detailed course of depressive symptoms and risk for developing depression in late adolescents with subthreshold depression: a cohort study. Neuropsychiatr Dis Treat 2017;13:25-33 [FREE Full text] [doi: 10.2147/NDT.S117846] [Medline: 28053534]
11. Clinical practice guideline for the treatment of depression across three age cohorts. American Psychological Association. 2019. URL: https://www.apa.org/depression-guideline [accessed 2025-04-08]
12. Guidelines for adolescent depression in primary care (GLAD-PC) toolkit. The Reach Institute. 2018. URL: https://thereachinstitute.org/glad-pc-toolkit-is-here-to-help-you-treat-depression/ [accessed 2025-04-08]
13. Depression in children and young people: identification and management. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/ng134 [accessed 2025-04-08]
14. Li LJ, Ma X. Chinese Guideline for the Prevention and Treatment in Depressive Disorder. 2nd Edition. Beijing, China: Chinese Medical Multimedia Press; 2015.
15. Cuijpers P, Quero S, Noma H, Ciharova M, Miguel C, Karyotaki E, et al. Psychotherapies for depression: a network meta-analysis covering efficacy, acceptability and long-term outcomes of all main treatment types. World Psychiatry 2021 Jun;20(2):283-293 [FREE Full text] [doi: 10.1002/wps.20860] [Medline: 34002502]
16. Caldwell DM, Davies SR, Thorn JC, Palmer JC, Caro P, Hetrick SE. School-Based Interventions to Prevent Anxiety, Depression and Conduct Disorder in Children and Young People: A Systematic Review and Network Meta-Analysis. Southampton, England: NIHR Journals Library; 2021.
17. James AC, Reardon T, Soler A, James G, Creswell C. Cognitive behavioural therapy for anxiety disorders in children and adolescents. Cochrane Database Syst Rev 2020 Nov 16;11(11):CD013162 [FREE Full text] [doi: 10.1002/14651858.CD013162.pub2] [Medline: 33196111]
18. Lightning review: access to child and adolescent mental health services. Children's Commissioner for England. 2016 May 28. URL: https://tinyurl.com/2stkubj6 [accessed 2025-04-08]
19. Aguirre Velasco A, Cruz ISS, Billings J, Jimenez M, Rowe S. What are the barriers, facilitators and interventions targeting help-seeking behaviours for common mental health problems in adolescents? A systematic review. BMC Psychiatry 2020;20(1):293 [FREE Full text] [doi: 10.1186/s12888-020-02659-0] [Medline: 32527236]
20. Wang C, Lei MC, Zhang N. Symptom-based psychotherapy decision-making under the framework of psychiatric medicine: taking cognitive behavioral therapy skills as an example. Chin J Psychiatry 2022;55(1):3-7. [doi: 10.1891/9780826110084.0008]
21. Cuijpers P, van Straten A, Warmerdam L. Behavioral activation treatments of depression: a meta-analysis. Clin Psychol Rev 2007;27(3):318-326. [doi: 10.1016/j.cpr.2006.11.001] [Medline: 17184887]
22. Stiles-Shields C, Montague E, Kwasny MJ, Mohr DC. Behavioral and cognitive intervention strategies delivered via coached apps for depression: pilot trial. Psychol Serv 2019;16(2):233-238. [doi: 10.1037/ser0000261] [Medline: 30407055]

XSL•FO
RenderX

23. Yang Y, Wang H, Sha W, Guo X, Deng W, Wang J, et al. Short video-based mental health intervention for depressive symptoms in junior high school students: a cluster randomized controlled trial. Psychol Res Behav Manag 2023;16:4169-4181 [FREE Full text] [doi: 10.2147/PRBM.S433467] [Medline: 37868654]
24. Moritz S, Cludius B, Hottenrott B, Schneider BC, Saathoff K, Kuelz AK, et al. Mindfulness and relaxation treatment reduce depressive symptoms in individuals with psychosis. Eur Psychiatry 2015;30(6):709-714. [doi: 10.1016/j.eurpsy.2015.05.002] [Medline: 26163302]
25. Schleider JL, Dobias ML, Sung JY, Mullarkey MC. Future directions in single-session youth mental health interventions. J Clin Child Adolesc Psychol 2020;49(2):264-278 [FREE Full text] [doi: 10.1080/15374416.2019.1683852] [Medline: 31799863]
26. DTx definition. Digital Therapeutics Alliance. 2023. URL: https://dtxalliance.org/understanding-dtx/ [accessed 2025-04-08]
27. Grudin R, Ahlen J, Mataix-Cols D, Lenhard F, Henje E, Månsson C, et al. Therapist-guided and self-guided internet-delivered behavioural activation for adolescents with depression: a randomised feasibility trial. BMJ Open 2022;12(12):e066357. [doi: 10.1136/bmjopen-2022-066357] [Medline: 36572500]
28. Nicol G, Wang R, Graham S, Dodd S, Garbutt J. Chatbot-delivered cognitive behavioral therapy in adolescents with depression and anxiety during the COVID-19 pandemic: feasibility and acceptability study. JMIR Form Res 2022;6(11):e40242 [FREE Full text] [doi: 10.2196/40242] [Medline: 36413390]
29. Schleider JL, Mullarkey MC, Fox KR, Dobias ML, Shroff A, Hart EA, et al. A randomized trial of online single-session interventions for adolescent depression during COVID-19. Nat Hum Behav 2022;6(2):258-268. [doi: 10.1038/s41562-021-01235-0] [Medline: 34887544]
30. Carswell K, Harper-Shehadeh M, Watts S, Van't Hof E, Abi Ramia J, Heim E, et al. Step-by-Step: a new WHO digital mental health intervention for depression. Mhealth 2018;4:34 [FREE Full text] [doi: 10.21037/mhealth.2018.08.01] [Medline: 30225240]
31. Li G, Sit HF, Chen W, Wu K, Sou EKL, Wong M, et al. A WHO digital intervention to address depression among young Chinese adults: a type 1 effectiveness-implementation randomized controlled trial. Transl Psychiatry 2024;14(1):102. [doi: 10.1038/s41398-024-02812-3] [Medline: 38378687]
32. Cuijpers P, Heim E, Abi Ramia J, Burchert S, Carswell K, Cornelisz I, et al. Effects of a WHO-guided digital health intervention for depression in Syrian refugees in Lebanon: a randomized controlled trial. PLoS Med 2022;19(6):e1004025. [doi: 10.1371/journal.pmed.1004025] [Medline: 35737665]
33. Moeini B, Bashirian S, Soltanian AR, Ghaleiha A, Taheri M. Examining the effectiveness of a web-based intervention for depressive symptoms in female adolescents: applying social cognitive theory. J Res Health Sci 2019 Aug 19;19(3):e00454 [FREE Full text] [Medline: 31586376]
34. Liu P, Situ M, Duan X, Sun H, Wang M, Tao Y, et al. The acceptability and effects of internet-based cognitive behavioral therapy on depressive symptoms and remission in 13- to 17-year-old adolescents: a systematic review and meta-analysis. J Evid Based Med 2024;17(4):740-757. [doi: 10.1111/jebm.12657] [Medline: 39582126]
35. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
36. Arevian AC, Jones F, Tang L, Sherbourne CD, Jones L, Miranda J, Community Partners in Care Writing Group. Depression remission from community coalitions versus individual program support for services: findings from community partners in care, Los Angeles, California, 2010-2016. Am J Public Health 2019;109(S3):S205-S213. [doi: 10.2105/AJPH.2019.305082] [Medline: 31242001]
37. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
38. Beck AT, Steer RA, Ranieri WF. Scale for suicide ideation: psychometric properties of a self-report version. J Clin Psychol 1988;44(4):499-505. [doi: 10.1002/1097-4679(198807)44:4<499::aid-jclp2270440404>3.0.co;2-6]
39. de Beurs DP, Fokkema M, O'Connor RC. Optimizing the assessment of suicidal behavior: the application of curtailment techniques. J Affect Disord 2016;196:218-224 [FREE Full text] [doi: 10.1016/j.jad.2016.02.033] [Medline: 26938964]
40. HU YQ, Gan YQ. Development and psychometric validity of the resilience scale for Chinese adolescents. Acta Psychol Sin 2008;40(8):902-912. [doi: 10.3724/sp.j.1041.2008.00902]
41. Gratz KL, Roemer L. Multidimensional assessment of emotion regulation and dysregulation: development, factor structure, and initial validation of the difficulties in emotion regulation scale. J Psychopathol Behav Assess 2004;26(1):41-54 [FREE Full text] [doi: 10.1023/B:JOBA.0000007455.08539.94]
42. Ye YM, Dai XY. Development of social support scale for university students. Chin J Clin Psychol 2008;16(5):456-458. [doi: 10.16128/j.cnki.1005-3611.2008.05.019]
43. Ren CR. Measurement methodology on social economic status index of students. J Educ Stud 2010;6(5):77-82. [doi: 10.14082/j.cnki.1673-1298.2010.05.010]
44. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
45. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

XSL•FO
RenderX

46. Glickman ME, Rao SR, Schultz MR. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. J Clin Epidemiol 2014;67(8):850-857. [doi: 10.1016/j.jclinepi.2014.03.012] [Medline: 24831050]
47. Sun L, Craiu RV, Paterson AD, Bull SB. Stratified false discovery control for large-scale hypothesis testing with application to genome-wide association studies. Genet Epidemiol 2006;30(6):519-530. [doi: 10.1002/gepi.20164] [Medline: 16800000]
48. Wang MC. Latent Variable Modeling Using Mplus. 1st Edition. Chongqing, China: Chongqing University Press; 2014.
49. Lehtimaki S, Martic J, Wahl B, Foster KT, Schwalbe N. Evidence on digital mental health interventions for adolescents and young people: systematic overview. JMIR Ment Health 2021;8(4):e25847 [FREE Full text] [doi: 10.2196/25847] [Medline: 33913817]
50. Stasiak K, Hatcher S, Frampton C, Merry SN. A pilot double blind randomized placebo controlled trial of a prototype computer-based cognitive behavioural therapy program for adolescents with symptoms of depression. Behav Cogn Psychother 2014;42(4):385-401. [doi: 10.1017/S1352465812001087] [Medline: 23253641]
51. Poppelaars M, Tak YR, Lichtwarck-Aschoff A, Engels RCME, Lobel A, Merry SN, et al. A randomized controlled trial comparing two cognitive-behavioral programs for adolescent girls with subclinical depression: a school-based program (Op Volle Kracht) and a computerized program (SPARX). Behav Res Ther 2016;80:33-42. [doi: 10.1016/j.brat.2016.03.005] [Medline: 27019280]
52. Wright B, Tindall L, Hargate R, Allgar V, Trépel D, Ali S. Computerised cognitive-behavioural therapy for depression in adolescents: 12-month outcomes of a UK randomised controlled trial pilot study. BJPsych Open 2019;6(1):e5. [doi: 10.1192/bjo.2019.91] [Medline: 31829300]
53. Do R, Lee S, Kim JS, Cho M, Shin H, Jang M, et al. Effectiveness and dissemination of computer-based cognitive behavioral therapy for depressed adolescents: effective and accessible to whom? J Affect Disord 2021;282:885-893. [doi: 10.1016/j.jad.2020.12.177] [Medline: 33601732]
54. Smith P, Scott R, Eshkevari E, Jatta F, Leigh E, Harris V, et al. Computerised CBT for depressed adolescents: randomised controlled trial. Behav Res Ther 2015;73:104-110. [doi: 10.1016/j.brat.2015.07.009] [Medline: 26301756]
55. Fazia T, Bubbico F, Nova A, Buizza C, Cela H, Iozzi D, et al. Improving stress management, anxiety, and mental well-being in medical students through an online mindfulness-based intervention: a randomized study. Sci Rep 2023;13(1):8214 [FREE Full text] [doi: 10.1038/s41598-023-35483-z] [Medline: 37217666]
56. Bostock S, Crosswell AD, Prather AA, Steptoe A. Mindfulness on-the-go: effects of a mindfulness meditation app on work stress and well-being. J Occup Health Psychol 2019;24(1):127-138 [FREE Full text] [doi: 10.1037/ocp0000118] [Medline: 29723001]
57. Ma L, Zhang Y, Huang C, Cui Z. Resilience-oriented cognitive behavioral interventions for depressive symptoms in children and adolescents: a meta-analytic review. J Affect Disord 2020;270:150-164. [doi: 10.1016/j.jad.2020.03.051] [Medline: 32339107]
58. Bisson JI, Ariti C, Cullen K, Kitchiner N, Lewis C, Roberts NP, et al. Guided, internet-based, cognitive behavioural therapy for post-traumatic stress disorder: pragmatic, multicentre, randomised controlled non-inferiority trial (RAPID). BMJ 2022;377:e069405. [doi: 10.1136/bmj-2021-069405] [Medline: 35710124]
59. Schleider JL, Weisz JR. Little treatments, promising effects? Meta-analysis of single-session interventions for youth psychiatric problems. J Am Acad Child Adolesc Psychiatry 2017;56(2):107-115. [doi: 10.1016/j.jaac.2016.11.007] [Medline: 28117056]
60. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MFG. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. BMJ 2012;344:e2598. [doi: 10.1136/bmj.e2598] [Medline: 22517917]
61. Peake E, Miller I, Flannery J, Chen L, Lake J, Padmanabhan A. Preliminary efficacy of a digital intervention for adolescent depression: randomized controlled trial. J Med Internet Res 2024 Mar 07;26:e48467 [FREE Full text] [doi: 10.2196/48467] [Medline: 38324367]
62. Zhou R, Zhang YJ. Design group intervention training for middle school students? psychological resilience based on rational emotive behavior. J Dalian Educ Univ 2015;31(4):48-50 [FREE Full text]
63. Konradt CE, Cardoso TDA, Mondin TC, Souza LDDM, Kapczinski F, da Silva RA, et al. Impact of resilience on the improvement of depressive symptoms after cognitive therapies for depression in a sample of young adults. Trends Psychiatry Psychother 2018;40(3):226-231. [doi: 10.1590/2237-6089-2017-0047] [Medline: 30304118]
64. van Bronswijk SC, Lemmens LH, Huibers MJ, Arntz A, Peeters FP. The influence of comorbid anxiety on the effectiveness of cognitive therapy and interpersonal psychotherapy for major depressive disorder. J Affect Disord 2018;232:52-60. [doi: 10.1016/j.jad.2018.02.003] [Medline: 29477584]
65. Assmann N, Schramm E, Kriston L, Hautzinger M, Härter M, Schweiger U, et al. Moderating effect of comorbid anxiety disorders on treatment outcome in a randomized controlled psychotherapy trial in early-onset persistently depressed outpatients. Depress Anxiety 2018;35(10):1001-1008. [doi: 10.1002/da.22839] [Medline: 30199128]
66. Cole DA, Peeke LG, Martin JM, Truglio R, Seroczynski AD. A longitudinal look at the relation between depression and anxiety in children and adolescents. J Consult Clin Psychol 1998;66(3):451-460. [doi: 10.1037//0022-006x.66.3.451] [Medline: 9642883]

XSL•FO
RenderX

67. Giber B, Schier MM, Bress JN, Mir Z, Diaz JL, Falk A, et al. Uncovering the heterogeneity in young adult anxiety: subtypes predict outcomes of a CBT mobile app. Mol Psychiatry 2026;31(2):908-914. [doi: 10.1038/s41380-025-03188-6] [Medline: 40866542]
68. Wang Y, Zhang X, Li Y, Qin H, Li X. Gender differences in the prevalence, correlated factors and comorbidity of depression in adolescents: a cross-sectional study in Shanghai, China. Front Public Health 2024;12:1436413 [FREE Full text] [doi: 10.3389/fpubh.2024.1436413] [Medline: 39712306]
69. Choi KW, Kim YK, Jeon HJ. Comorbid anxiety and depression: clinical and conceptual consideration and transdiagnostic treatment. Adv Exp Med Biol 2020;1191:219-235. [doi: 10.1007/978-981-32-9705-0_14] [Medline: 32002932]
70. Dülsen P, Baumeister H. Internet- and mobile-based anxiety and depression interventions for children and adolescents: efficacy and negative effects - a systematic review and meta-analysis. Eur Child Adolesc Psychiatry 2025;34(1):101-121. [doi: 10.1007/s00787-024-02404-y] [Medline: 38430237]
71. He J, Zhong Y, Song Y, Luo J, Lin C, Wu Y, et al. The impact of only-child status on adolescent mental health: a multi-center cross-sectional study using propensity score matching in Western China. BMC Public Health 2025 Jun 07;25(1):2140 [FREE Full text] [doi: 10.1186/s12889-025-23383-y] [Medline: 40483412]
72. Downey DB, Cao R. Number of siblings and mental health among adolescents: evidence from the U.S. and China. J Fam Issues 2023;45(11):2822-2850. [doi: 10.1177/0192513x231220045]
73. Bratu EA, Moroianu L, Isailă O, Ple ea-Condratovici C, Avram O, Drima E. Parental mental health and suicidal behavior as predictors of adolescent suicidal ideation and attempts: a systematic review and meta-analysis. J Clin Med 2025;14(19):6860. [doi: 10.3390/jcm14196860] [Medline: 41095939]
74. Le Bas G, Aarsman SR, Rogers A, Macdonald JA, Misuraca G, Khor S, et al. Paternal perinatal depression, anxiety, and stress and child development: a systematic review and meta-analysis. JAMA Pediatr 2025;179(8):903-917. [doi: 10.1001/jamapediatrics.2025.0880] [Medline: 40522669]

Abbreviations

ANCOVA: analysis of covariance
BSI-CV: Beck Scale for Suicide Ideation–Chinese Version
CBT: cognitive behavioral therapy
FDR: false discovery rate
LSMD: least squares mean difference
MDD: major depressive disorder
NNT: number needed to treat
OR: odds ratio
PHQ-8: Patient Health Questionnaire-8
RCT: randomized controlled trial
SES: socioeconomic status

Edited by L Buis; submitted 19.Oct.2025; peer-reviewed by M Pritchard, Y Cao; comments to author 30.Apr.2026; revised version received 23.Jun.2026; accepted 24.Jun.2026; published 07.Aug.2026.

Please cite as:
Liu P, Wang X, Duan X, Situ M, Tao Y, Yang Z, Zhao S, Li L, Li S, Jiang Y, Zhang M, Jing D, Yu X, Huang Y
Brief Web-Based Psychological Intervention for Depressive Symptoms Among Chinese Adolescents: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e86134
URL: https://mhealth.jmir.org/2026/1/e86134
doi:10.2196/86134
PMID:

©Pei Liu, Xinwei Wang, Xiaoxia Duan, Mingjing Situ, Yujie Tao, Zhaozhi Yang, Shengnan Zhao, Lei Li, Sixun Li, Yuchu Jiang, Manxue Zhang, Di Jing, Xinyi Yu, Yi Huang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 07.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Mobile Virtual Reality–Based Intervention for Stress Reduction Among Young Adults: Pilot Randomized Controlled Trial

Houlin Hong[1,2], MPH; Matthew R Lootens[1,2], PhD; Rebecca Bachman[1,2], MLA; Luis David Olivera Leon[1,2], MD, MPH; Ann F Dunlap[1,2], MPH; Emilio Goldenhersch[3,4], BA; Nicolas Rosencovich[3,5], BSc; Cristian Waitman[3,6], BA; Katarzyna E Wyka[1,2], PhD; Terry T-K Huang[1,2], MPH, MBA, PhD

[1]Center for Systems and Community Design, City University of New York, Graduate School of Public Health and Health Policy, New York City, NY, United States

[2]NYU-CUNY Prevention Research Center, New York City, NY, United States

[3]MindCo Health, San Francisco, CA, United States

[4]Universidad de Flores, Laboratorio de Investigación en Neurociencia y Ciencias Sociales, Ciudad Autónoma de Buenos Aires, Argentina

[5]Universidad Nacional de Córdoba, Escuela de Ingeniería Biomédica, Córdoba, Argentina

[6]Universidad Siglo XXI, Córdoba, Argentina

Corresponding Author:
Terry T-K Huang, MPH, MBA, PhD
Center for Systems and Community Design
City University of New York, Graduate School of Public Health and Health Policy
55 W 125th St
New York City, NY, 10027
United States
Phone: 1 646 364 9600
Email: terry.huang@sph.cuny.edu

Abstract

Background: Young adults experience disproportionately high levels of stress, creating significant health challenges. This demographic often underuses traditional stress management methods due to perceived ineffectiveness, inconvenience, or stigma. Given these challenges, mobile virtual reality (VR) offers a novel, engaging, and accessible alternative stress management tool that may be more appealing to young adults.

Objective: This pilot randomized controlled trial (RCT) evaluated the preliminary effects of MindCo Relief (MR), a mobile VR-based stress management app, accompanied by online coaching, on perceived stress, resilience, anxiety, depressive symptoms, and quality of life (QoL).

Methods: Young adults aged 18 to 29 years were recruited to participate in the study. A total of 70 participants with moderate-to-high stress levels at baseline were randomized to an intervention (n=34) group and a control (n=36) group. Moderate-to-high stress was defined as a sum score of 14 or more on the Perceived Stress Scale–10 (PSS-10). The intervention group completed an 8-week VR mindfulness program with access to in-person coaching, while the control group received no intervention. The primary outcome was self-reported perceived stress (PSS-10). Secondary measures included resilience (measured by the Connor-Davidson Resilience Scale–2), anxiety (measured by the Generalized Anxiety Disorder–7 scale), depressive symptoms (measured by the Patient Health Questionnaire–9), and QoL (measured by the General Well-Being Schedule). Primary analyses used linear mixed-effects models to assess the intervention effect over the 8-week study period. Secondary analyses examined whether the percentage of program modules completed and the number of coaching sessions attended were associated with changes in the primary and secondary outcomes.

Results: The mean age of the participants was 23 (SD 3.41) years, and they were predominantly female (n=52, 74%), students (n=46, 66%), and insured (n=63, 90%). No significant differences were observed between the intervention and control groups in terms of sex (*P*=.89), student status (*P*=.56), insurance status (*P*=.68), race and ethnicity (*P*=.18), or age (*P*=.32). Linear mixed-effects models showed a significant interaction between time and intervention, with the intervention group showing greater reductions in stress (β=−3.84, SE 1.37; *P*=.007), anxiety (β=−3.21, SE 0.97; *P*=.002), and depressive symptoms (β=−3.13, SE 1.13; *P*=.007) and a greater increase in QoL (β=9.99, SE 2.92; *P*=.001). No significant interaction was found for resilience

XSL•FO
RenderX

(β=0.83, SE 0.43; *P*=.06). In the secondary analyses, neither the percentage of modules completed nor the number of coaching sessions attended was significantly associated with changes in the primary or secondary outcomes.

Conclusions: This study provided preliminary evidence that mobile VR can help reduce stress among young adults by providing immersive environments that promote relaxation and adaptive stress responses. The findings support further exploration of VR's role in mental health care. Future research should examine the efficacy and long-term effects of this mobile VR software using larger, more diverse samples. Strategies to sustain VR engagement are also warranted.

Trial Registration: ClinicalTrials.gov NCT06970652; https://clinicaltrials.gov/study/NCT06970652

(JMIR Mhealth Uhealth 2026;14:e83438) doi:10.2196/83438

KEYWORDS

virtual reality; stress management; young adults; mindfulness; randomized control trial

Introduction

Young adults aged 18 to 29 years in the United States experience disproportionately high levels of stress, making them one of the most vulnerable groups for mental health disorders [1-3]. Several often co-occurring challenges of young adulthood, such as pressures to succeed in school, establish a career, maintain relationships, and manage finances, contribute to higher stress levels in this age group than among older adults [2,4]. Repeated exposure to social comparisons and expectations from frequent use of social media and other digital technologies may further exacerbate feelings of stress [5,6].

Stress can have a profound impact on both mental and physical health. Studies have shown that prolonged stress is a precursor to severe psychiatric conditions, such as major depressive disorder, bipolar disorder, and posttraumatic stress disorder (PTSD) [7]. Furthermore, adverse psychological health conditions, including stress, anxiety, and depression, are considered risk factors for neurological, metabolic, and cardiovascular diseases (CVDs) [8,9]. The psychological effects of frequent stress and the perception of daily stress may disrupt short-term stress responses and hinder healthy adaptation. Consequently, these disruptions can increase the risk of chronic diseases such as CVD [10].

For young adults, the consequences of untreated stress extend to diminished academic and occupational performance and reduced overall quality of life (QoL), highlighting the urgent need for accessible and effective interventions [11]. Despite studies showing that physical activity is positively associated with mental well-being, only a quarter of US adults meet the National Physical Activity Guidelines [9,12,13]. Although traditional methods such as cognitive behavioral therapy (CBT) and mindfulness-based stress reduction therapy have also been shown to be effective for managing stress and anxiety, the uptake of these treatments among young adults is often hindered by perceptions that such care is ineffective or inconvenient [14-16]. Studies have also shown that stigma is a major barrier to mental health help seeking among college students, as they feel embarrassed and fear being labeled "crazy" by others [17,18]. Young adults also lack knowledge about mental health services and have difficulty identifying symptoms of mental illness. Consequently, they often continue to normalize their elevated distress and prefer self-reliance over professional help [3,18-20]. Moreover, traditional therapeutic approaches often lack personalization and do not resonate or align with young adults' individual needs and lifestyles [20].

Given these challenges, innovative stress management interventions that are flexible, accessible, and appealing to young adults are needed. Virtual reality (VR) is a novel and promising stress management tool, providing an immersive and engaging experience that can be more appealing than traditional approaches to many individuals [21]. VR technology uses computer modeling to generate interactive 3D visual environments, with additional auditory and haptic feedback also possible [22-24]. VR allows users to experience and interact with the generated environment as if they were physically present, setting it apart from other non-VR apps, websites, and video-based relaxation programs [25]. Compared to imagery and 2D technologies, VR triggers stronger emotional engagement and physiological responses, making it a more efficient medium for immersive experiences [26]. By fostering a strong sense of presence in a calming, controllable environment, VR can support short-term emotion regulation, which in turn may improve subjective well-being and QoL [27]. Mobile VR operating on smartphones can enhance VR's accessibility and portability, making it a potential solution for addressing stress on a larger scale [20].

Despite the growing interest in VR apps, there is limited research on the effectiveness of VR in stress reduction and management, particularly among young adults. Some studies have shown promising results, indicating that VR is associated with stress reduction [28-30]. However, most of the literature to date on VR therapy focuses on addictions or severe mental health conditions, such as social anxiety disorder, obsessive-compulsive disorder, and PTSD [20,22,31]. Although mobile technology has evolved rapidly over the past 5 years, studies examining VR use for stress remain scarce [27].

This pilot study used MindCo Relief (MR), a smartphone VR app designed for stress reduction, as the intervention in a randomized controlled trial (RCT) to test whether this mobile VR behavior change program would reduce perceived stress among young adults aged 18 to 29 years. We hypothesized that after the 8-week intervention period, use of the MR app would be positively associated with decreases in perceived stress, depressive symptoms, and anxiety and increases in resilience and QoL.

XSL•FO
RenderX

Methods

Study Design

This study was an 8-week pilot RCT conducted to examine the feasibility and preliminary impact of MR on stress reduction among young adults. Block randomization by sex was used to ensure a balanced distribution between male and female participants. The random allocation sequence was generated using a computer-based random number generator, with a block size of 2. Participants were assigned sequentially to the intervention or control group according to the generated allocation sequence. Enrollment occurred from October 2023 to January 2024. The first participant began the trial in October 2023, and the last completed the study in May 2024. This study was reported in accordance with the CONSORT (Consolidated Standards of Reporting Trials) checklist, which can be found in Multimedia Appendix 1.

Participants

Study participants were screened for eligibility using an online REDCap-based survey (Vanderbilt University) [32,33]. Eligibility criteria included being aged 18 to 29 years, having a moderate or high level of perceived stress at baseline, having access to a smartphone running the Android or iOS operating system, and being able to install and run the MR app. Perceived stress was measured using the Perceived Stress Scale–10 (PSS-10) [34,35]. Individuals participating in other stress reduction programs or with photosensitive epilepsy were excluded. Moderate stress was defined as a PSS-10 score of at least 14 but no more than 26, and high stress was defined as a score of 27 or greater. Participants were recruited primarily through online postings (on university websites and social media platforms) directed at individuals residing in New York City. Recruitment materials emphasized stress management and targeted college students and young adults meeting the inclusion criteria. After enrollment, participants were required to visit our study site to complete randomization.

Intervention Details

Participants in the intervention group received access to the MR smartphone app and a smartphone-compatible VR headset, both developed by MindCo Health [36]. All participants in the intervention group received VR headsets directly from the research team at our study site. The app contained a stress reduction program that comprised a structured sequence of 76 modules designed to be completed over 8 weeks. Each module was designed to be completed in approximately 7 minutes and had to be completed before progressing to the next module. Forty modules used VR. In these modules, participants placed their personal smartphones running the MR app in the VR headset, and the smartphone provided all the audiovisual content. All other modules used smartphones without the VR headset.

The program was grounded in a CBT framework and integrated mindfulness-based stress reduction principles with structured psychoeducation and immersive natural environments [37-39]. VR modules served 2 complementary functions. First, educational VR sessions placed participants in a simulated therapy office environment where a prerecorded clinician delivered CBT-based psychoeducational content focused on stress physiology, cognitive appraisal, maladaptive thought patterns, and coping strategies. This immersive format was designed to enhance attentional engagement and contextual encoding of therapeutic material. Second, mindfulness-based VR sessions transported participants to simulated natural environments (eg, forests and beaches), where guided practices such as breath awareness, body scan, muscle relaxation, and compassion exercises were delivered within biophilic settings to promote attentional stabilization and relaxation.

VR module completion was automatically tracked through the app backend. To initiate immersive playback, participants were required to grant gyroscope and rotation permissions, activating head-tracking mode. VR videos could not be fast-forwarded, and module completion status was recorded only after full playback duration. Progression to subsequent modules required completion of prior modules. Backend logs captured the playback duration and activation of immersive mode. As is typical in remote digital health interventions, direct verification of sustained headset wear or continuous visual attention was not feasible. Self-reflective journaling modules reinforced the integration of psychoeducational and mindfulness content by prompting participants to identify stress triggers, bodily sensations, and cognitive responses.

In addition to the structured program, the app provided access to 30-minute online coaching sessions. Each participant in the intervention group was allowed to schedule up to 3 sessions during the 8-week participation period. Coaches were non–US-licensed psychologists based in Argentina, trained in CBT, and followed a protocol developed to support the MR intervention. The protocol promoted engagement with the MR modules aligned with the participants' needs at the time of the coaching session. The control group received no active intervention throughout the study period.

Primary and Secondary Outcomes

The primary study outcome was perceived stress, measured by PSS-10, a widely used, validated, and easy-to-use questionnaire [34,40,41]. The PSS-10 includes 10 questions, each scored from 0 to 4. A sum score is calculated as the total PSS score, with a higher score indicating a higher stress level. Secondary outcomes were also collected via surveys, including resilience, anxiety, depression, and QoL. Resilience was measured using the Connor-Davidson Resilience Scale–2 (CD-RISC2), with a total score ranging from 0 to 8, where higher scores indicate greater resilience [42]. Anxiety symptoms were assessed using the Generalized Anxiety Disorder-7 scale (GAD-7), and depressive symptoms were measured using the Patient Health Questionnaire–9 (PHQ-9). Both instruments evaluated symptom frequency over the past 2 weeks using a 4-point scale with answer options as follows: 0="Not at all," 1="Several days," 2="More than half the days," and 3="Nearly every day." Total scores range from 0 to 21 for the GAD-7 and 0 to 27 for the PHQ-9, with higher scores indicating greater symptom severity [43-46]. QoL was measured using the General Well-Being Schedule (GWS), which comprises 18 items assessing psychosocial well-being; higher scores indicate better overall

XSL•FO
RenderX

QoL [47-49]. All measures have demonstrated good reliability and validity in adult populations and have been widely used in young adults of the study cohort [41,50-52]. Assessments for all outcomes were conducted at baseline and again at 8 weeks after the conclusion of the intervention.

Covariates

Demographic variables were collected from the participants at study enrollment, including age, sex (male vs female), race and ethnicity (Asian, Black, other, Hispanic, or White), student status, and health insurance status (public, private, or none). Age and student status were examined to account for potential differences in digital program use, as recent systematic reviews have indicated that these factors significantly influence adherence to digital health interventions [53,54]. Additionally, race and ethnicity were explored to assess potential differences in intervention outcomes. Reports and studies have shown that mental health outcomes and stress exposure differ across racial and ethnic groups in New York City, with structural and environmental inequities contributing to higher psychological distress among Black and Latino adults [55-57]. A previous study has also indicated that insurance status may affect mental health conditions [57].

Statistical Analysis

Demographics were summarized by treatment group, and chi-square tests were used to examine differences in categorical variables by intervention status. A 2-tailed *t* test was used to compare age between the intervention and control groups. Additional comparisons were conducted between individuals who enrolled in the study and those who were eligible but did not participate, as well as between participants who completed the follow-up assessment and those who were lost to follow-up.

Linear mixed-effects models with repeated measures were used to regress primary and secondary outcomes on time, intervention, and the interaction between time and intervention. The models also included fixed effects for sex, as sex was used for block randomization. With 2 waves of data collection, a compound symmetry covariance structure was used. Other covariates that were balanced between the 2 groups were not included in the model.

In the secondary analysis, we also examined whether program engagement was associated with improvements in outcomes. Changes in scores for the primary and secondary outcomes were first calculated as the difference between follow-up and baseline measurements and then modeled as a function of program engagement. Program engagement was assessed by the percentage of module completion and the number of in-person coaching sessions attended. Both variables were analyzed as continuous and categorical variables, with each specification evaluated in a separate regression model. When analyzed as categorical variables, the percentage of module completion was divided into quartiles, and the number of coaching sessions was dichotomized as <2 sessions vs ≥2 sessions.

The primary analysis was conducted on an intention-to-treat (ITT) basis. No participants with relevant observed data were excluded, and the analysis followed the participants' assigned group, regardless of the actual treatment or dose received. The secondary engagement analysis, which modeled change scores as a function of the dose received, was restricted to participants with both baseline and follow-up measurements. Statistical analyses were performed using R (version 4.4.1; R Foundation for Statistical Computing) [58,59]. The α level was set at .05.

Ethical Considerations

The study was approved by the City University of New York Institutional Review Board (2023-0338-PHHP), which determined the study to be of minimal risk. This study was registered at ClinicalTrials.gov (NCT06970652). All participants were provided with detailed information regarding the study's purpose, procedures, potential risks and benefits, data confidentiality, and their rights as research participants. Informed consent was obtained electronically through REDCap prior to study participation.

Participation was entirely voluntary, and participants were informed that they could withdraw at any time without penalty. To protect privacy and confidentiality, data were deidentified before analysis, and only the research team had access to the final analytic dataset. Personally identifiable information was stored separately from study data on secure, password-protected servers. Participants who completed all 8 weeks of the study and provided postintervention data were compensated with a US $50 electronic gift card.

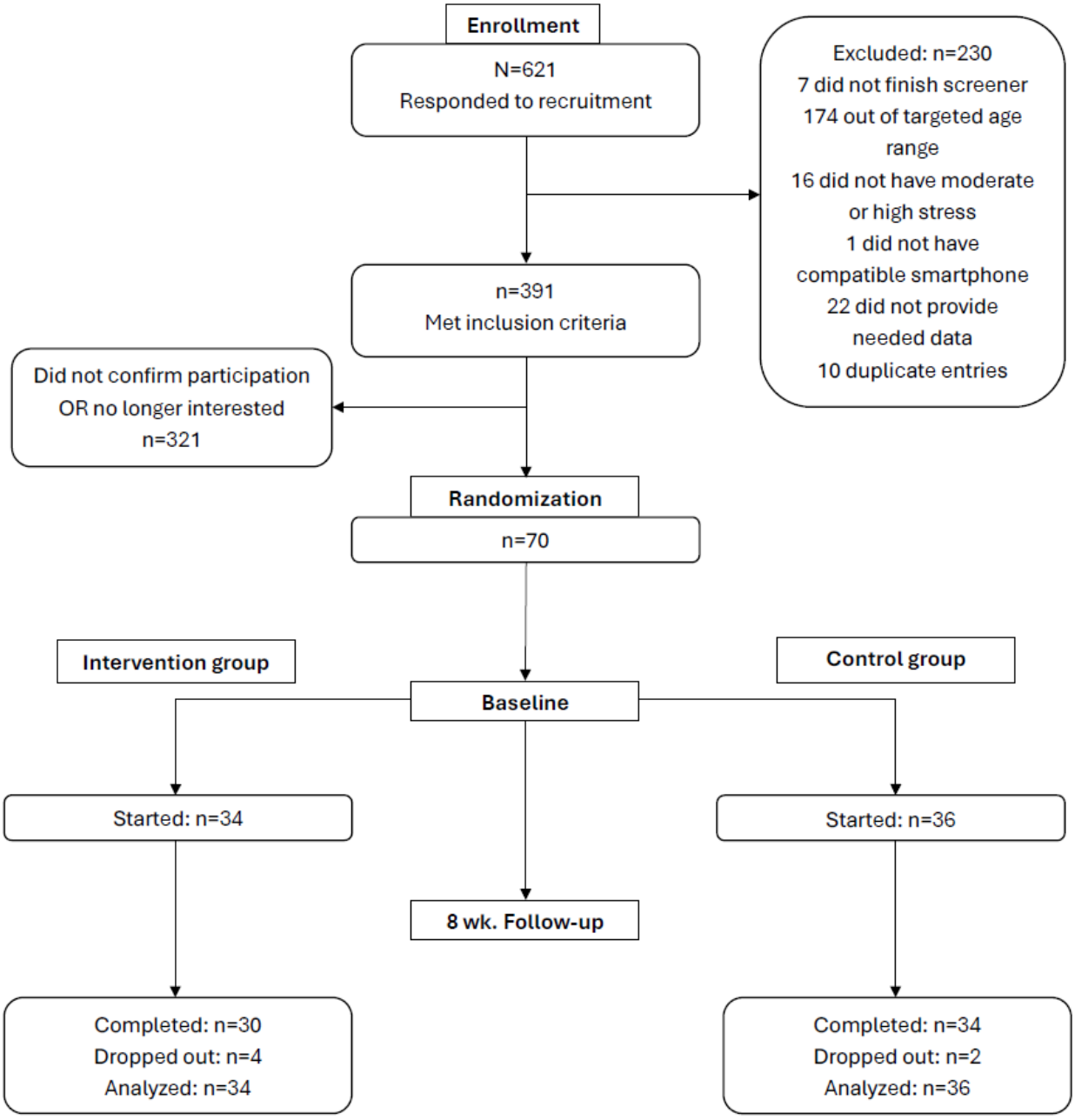
Results

Sample Characteristics

A total of 621 individuals responded to the online recruitment, of whom 230 were excluded based on the study's inclusion and exclusion criteria. Among the remaining respondents, 283 individuals completed the consent form, and 70 enrolled in the study and were randomized to either the intervention or control group. Figure 1 shows the sampling flowchart. Demographic comparisons between participants who enrolled and those who were eligible but did not enroll are shown in Table S1 in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 1. CONSORT flow diagram.



Of the 70 enrolled participants, 34 were randomized to the intervention group and 36 to the control group. The mean age was 23.4 (SD 3.41) years, and the majority were female (n=52, 74%). Forty-six participants were college students, and 63 reported health insurance coverage, either private (n=24) or public (n=39). We found no significant differences in demographic variables between the intervention and control groups at baseline (all $P>.05$; Table 1). Six participants dropped out of the study and failed to provide postintervention data (n=4 in the intervention group; n=2 in the control group). Additional comparisons showed no demographic differences between those who were lost to follow-up and those who provided follow-up data (Table S2 in Multimedia Appendix 1).

XSL•FO
RenderX

Table 1. Demographic characteristics of study participants (n=70).

Characteristics	Overall	Control group (n=36)	Intervention group (n=34)	*P* value
Sex, n (%)				.89
Male	18 (26)	9 (25)	9 (26)	
Female	52 (74)	27 (75)	25 (74)	
Student status, n (%)				.56
Not a student	24 (34)	14 (39)	10 (29)	
Student	46 (66)	22 (61)	24 (71)	
Insurance status, n (%)				.68
Public	39 (56)	22 (61)	17 (50)	
Private	24 (34)	11 (31)	13 (38)	
None	7 (10)	3 (8)	4 (12)	
Race and ethnicity, n (%)				.18
Asian	24 (34)	12 (33)	12 (35)	
Black	9 (13)	4 (11)	5 (15)	
Hispanic	22 (31)	12 (33)	10 (29)	
White	11 (16)	8 (22)	3 (9)	
Other	4 (6)	0 (0)	4 (12)	
Age (years), mean (SD)	23.36 (3.41)	23.75 (3.51)	22.94 (3.30)	.32

Overall, the sample showed moderate levels of stress (PSS-10: mean 25.87, SD 4.84; 23/70, 33% with high stress), moderate resilience (CD-RISC2: mean 4.52, SD 1.50), moderate anxiety (GAD-7: mean 11.41, SD 4.61), moderate depression (PHQ-9: mean 11.99, SD 6.26), and severe distress (GWS: mean 55.66, SD 13.23). There were no significant differences between the intervention and control groups in perceived stress, anxiety, depressive symptoms, resilience, or QoL at baseline (all *P*>.05).

Primary Outcome

The intervention group showed a decrease in average stress scores from 25.92 (SE 1.03) at baseline to 18.21 (SE 1.03) after the intervention. Linear mixed-effects regression showed a significant interaction between the intervention group and time for PSS-10 (β=−3.84, SE 1.37; *P*=.007; Table 2).

Table 2. Mean scores for primary and secondary outcomes at baseline and follow-up (n=70)[a].

Outcomes	Control group (n=36), mean (SE)		Intervention group (n=34), mean (SE)		Interaction, β (SE)	*P* value
	Baseline	Follow-up	Baseline	Follow-up		
Perceived Stress Scale–10	25.25 (0.96)	21.38 (0.98)	25.92 (0.98)	18.21 (1.03)	−3.84 (1.37)	*.007* [b]
Connor-Davidson Resilience Scale–2	4.68 (0.28)	4.81 (0.27)	4.33 (0.27)	5.29 (0.29)	0.83 (0.43)	.06
Generalized Anxiety Disorder–7 scale	11.43 (0.84)	9.68 (0.85)	10.98 (0.85)	6.02 (0.89)	−3.21 (0.97)	*.002*
Patient Health Questionnaire–9	12.15 (1.14)	10.42 (1.15)	11.78 (1.16)	6.93 (1.20)	−3.13 (1.13)	*.007*
General Well-Being Schedule	55.31 (2.53)	59.72 (2.56)	56.65 (2.57)	71.05 (2.67)	9.99 (2.92)	*.001*

[a]Reported means and interaction estimates were derived from linear mixed-effects models that included intervention, time, sex, and the time×intervention interaction.

[b]Italicized *P* values indicate statistical significance.

Secondary Outcomes

The intervention group showed significant reductions in GAD-7 and PHQ-9 scores and significant increases in CD-RISC2 and GWS scores from baseline to follow-up. Significant interactions were also observed for GAD-7 (β=−3.21, SE 0.97; *P*=.002), PHQ-9 (β=−3.13, SE 1.13; *P*=.007), and GWS (β=9.99, SE 2.92; *P*=.001), but not for CD-RISC2 (β=0.83, SE 0.43; *P*=.06). Detailed results from the sex-adjusted models are shown in Table 2, and results from the unadjusted models are provided in Table S3 in Multimedia Appendix 1.

XSL•FO
RenderX

Intervention Use

At the end of the intervention period, participants in the intervention group completed an average of 35% of the MR program. When examining the relationship between changes in outcomes and the percentage of modules completed, we did not find a significant interaction between time and the percentage of modules completed. Module completion was also not a significant predictor of changes in outcomes when examined by quartile. The average number of assisted coaching sessions used by participants in the intervention group was 1.3 out of a maximum of 3. Among individuals who attended any coaching sessions, the average number of sessions used was 2.2. No associations were found between the number of coaching sessions and changes in the primary and secondary outcomes. Study participants reported no harmful or unintended consequences.

Discussion

Principal Findings

This pilot study aimed to explore the feasibility of mobile VR technology as a potential tool for reducing stress and managing anxiety among young adults. The results showed that after 8 weeks of using a VR-based mindfulness program, the intervention group had significant reductions in stress, anxiety, and depressive symptoms and an increase in QoL.

Our findings are consistent with those of the limited previous studies demonstrating that VR technology, particularly mobile VR, is effective in reducing stress and managing mental health challenges. Gao et al [28] found that the use of mobile VR led to a reduction in stress. Similar findings were reported by McGarry et al [60], who found that mobile VR reduced psychological stress and was perceived as a useful and effective strategy for stress management by young adults aged 18 to 25 years. These 2 studies, however, did not have an active control group. In contrast, a mixed methods study with a control group that examined the short-term pre-post intervention effect on psychological well-being also reported a large decrease in stress, anxiety, and sadness in the intervention group [30]. These findings should be interpreted cautiously, as all 3 studies were limited to sample sizes of <20 participants and brief VR exposures ranging from a single session to a maximum of 14 days. Furthermore, these studies focused on generic VR apps that used scenery for relaxation, whereas our study used an integrated app designed for stress reduction through a CBT-based structured protocol.

Despite the positive findings from our study, participant engagement remains an important area for improvement. The average completion rate of VR sessions was 35% in this 8-week, 76-module intervention. Although modest, this level of completion is consistent with adherence patterns observed in multiweek, self-directed digital mental health programs, where attrition over time is common and full completion rates often fall within similar ranges (reported ranges of 0.5% to 28.6%), particularly in the absence of performance-based incentives [61]. Digital mindfulness and CBT interventions frequently demonstrate a progressive decline in engagement across sequential modules, especially in young adult populations [30,62]. In the present study, the requirement to complete modules sequentially and, for 40 modules, to set up immersive VR hardware may have introduced additional user burden relative to purely smartphone-based apps. Although monthly coaching sessions were incorporated to support adherence, the frequency of support may not have been sufficient to fully counteract typical digital attrition. Future iterations may benefit from increased personalization, adaptive pacing, or gamification strategies to enhance sustained engagement.

Our study did not find a dose-response relationship. In the regression models examining the preintervention and postintervention changes as a function of module completion, there was a trend in which participants in the highest quartile had the greatest reduction in PSS-10, GAD-7, and PHQ-9 scores compared to those in the lowest quartile. However, the differences did not reach statistical significance. Although dose-response relationships have been observed in VR interventions for other conditions such as PTSD and chronic pain, there is a lack of consensus regarding such relationships in VR-based stress management [63,64]. Future studies should explore effective engagement strategies and dose-response effects to strengthen VR-based stress management programs.

The ways in which VR reduces stress and anxiety are likely to be multifaceted. VR creates an immersive environment that can transport users to calm, relaxing settings, and this immersion can help redirect users away from real-world stressors and promote relaxation [65,66]. The VR app used by our participants incorporated guided imagery and relaxation techniques, such as deep breathing exercises and mindfulness practices. In addition, the program also facilitated psychoeducational content grounded in CBT. These techniques help users build coping skills, reduce stress responses, and achieve relaxation [65].

Unlike the significant findings for stress, we did not find a significant improvement in resilience associated with MR. Resilience, defined as the ability to adapt to adversity, is a critical component of mental health and well-being [67]. However, resilience may be influenced by a broader range of factors beyond the scope of our intervention, such as long-term life experiences, social support, and personality traits, which may require more extended or targeted interventions to observe measurable changes. There is a limited body of research examining resilience as an outcome of VR-based stress interventions, but given the potential of resilience to buffer against a poor ability to cope with stress, further research in this area is warranted.

An important observation from our study is the improvement in self-reported QoL following the intervention. QoL is a multifaceted construct of well-being, encompassing physical, psychological, and social dimensions [68]. Stress and anxiety are known to be negatively associated with QoL, contributing to emotional instability, cognitive difficulties, and possibly impaired social interactions [69]. Studies have shown that interventions aimed at reducing stress can improve QoL [27]. This study provides further evidence supporting the relationship between stress and QoL. Unlike generic relaxation VR, which typically consists solely of passive, scenery-based relaxation audiovisual components without targeted educational or skill

XSL•FO
RenderX

training content, the intervention used in this study was designed for stress management with structured techniques and guided practices. This targeted approach may be more effective than generic relaxation VR in improving psychological well-being and stress management. Future research should explore how integrating adaptive content and personalized engagement strategies can further optimize VR interventions for sustained improvements in QoL.

As VR technology continues to evolve, its integration into mental health care systems offers more possibilities. Mobile VR solutions, in particular, are advantageous due to their high accessibility and portability, making them widely available and suitable for individuals regardless of location [70]. Furthermore, the increasing prevalence of telehealth provides an opportunity to complement remote therapy with VR-based interventions, offering individuals accessible tools for managing psychological distress at home. The feasibility and cost-effectiveness of integrating VR into existing mental health care services will be an important next step toward maximizing VR's potential impact.

Strength and Limitations

Although this study used an RCT design, several limitations remain. First, our study had a small sample size and a relatively short study duration, limiting the ability to assess the intervention's long-term effects. Participants were not blinded due to the nature of the VR intervention, which may introduce expectancy effects in both the intervention and control groups. Although we observed significant reductions in stress, anxiety, and depressive symptoms over the trial period, the duration of these effects remains uncertain. Additionally, all outcomes were measured using self-reported surveys, which may be subject to reporting bias. The generalizability of our findings is also limited by the self-selected nature of our sample, which consisted primarily of young adults in New York City. The inclusion criteria were necessary to account for confounders but may have inadvertently excluded individuals who would benefit from the intervention in real-world settings. Although many eligible individuals did not proceed to randomization, this attrition occurred between screening and in-person equipment collection, suggesting that logistical requirements may have posed a barrier to study participation. Further research with larger, more diverse samples and extended follow-up periods is necessary to determine the sustained benefits of VR-based stress management interventions over time.

Conclusions

This study provides preliminary evidence supporting the use of mobile VR technology to reduce stress, anxiety, and depressive symptoms and to improve QoL among young adults. Given that this study had a moderate sample size, a low program completion rate, and a sample that was predominantly female, self-selected, and mostly urban, the generalizability of the study's results should be approached with caution. Future research should examine the efficacy of this VR app with larger, more diverse samples to assess its long-term mental health impact.

Acknowledgments

The authors thank all study participants who devoted their time to this study. The authors declare that no generative AI tools were used in the writing of this manuscript, and only the writing tool, Grammarly (Grammarly Inc) was used to suggest language improvements.

Data Availability

The dataset used and/or analyzed in this study is available from the corresponding author upon reasonable request.

Funding

No external funding was received. This pilot study was supported by internal funding from the CUNY Graduate School of Public Health and Health Policy.

Authors' Contributions

Conceptualization: EG, TT-KH
Data curation: HH, MRL, NR
Formal analysis: HH, MRL
Funding acquisition: NR, TT-KH
Investigation: RB, LDOL, AFD, KEW
Methodology: HH, EG, KEW, TT-KH
Project administration: MRL, AFD
Resources: EG, NR, CW, TT-KH
Software: HH, MRL, EG, NR, CW
Supervision: TT-KH
Validation: MRL, KEW
Writing—original draft: HH, MRL
Writing—review and editing: HH, MRL, RB, EG, KEW, TT-KH

XSL•FO
RenderX

Conflicts of Interest

EG, NR, and CW are cofounders of MindCo Health, a start-up company that developed the MindCo Relief app used in this study. TH is affiliated with an investment fund that holds an equity interest in MindCo Health.

Editorial Notice

This randomized study was retrospectively registered. The authors noted that the registration was conducted retrospectively due to administrative oversight. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials, because the risk of bias appears low and the study information registered with authors' ethics committee matches the trial reporting. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
Supplementary file with additional characteristics and outcome data.
[DOCX File , 31 KB - mhealth_v14i1e83438_app1.docx]

Multimedia Appendix 2
CONSORT (Consolidated Standards of Reporting Trials) checklist.
[PDF File (Adobe PDF File), 290 KB - mhealth_v14i1e83438_app2.pdf]

References

1. Acharya L, Jin L, Collins W. College life is stressful today - emerging stressors and depressive symptoms in college students. J Am Coll Health 2018 Oct;66(7):655-664. [doi: 10.1080/07448481.2018.1451869] [Medline: 29565759]
2. Stefaniak AR, Blaxton JM, Bergeman CS. Age differences in types and perceptions of daily stress. Int J Aging Hum Dev 2022 Mar;94(2):215-233. [doi: 10.1177/00914150211001588] [Medline: 33739147]
3. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
4. Matud MP, Díaz A, Bethencourt JM, Ibáñez I. Stress and psychological distress in emerging adulthood: a gender analysis. J Clin Med 2020 Sep 04;9(9):2859 [FREE Full text] [doi: 10.3390/jcm9092859] [Medline: 32899622]
5. Steele RG, Hall JA, Christofferson JL. Conceptualizing digital stress in adolescents and young adults: toward the development of an empirically based model. Clin Child Fam Psychol Rev 2020 Mar;23(1):15-26. [doi: 10.1007/s10567-019-00300-5] [Medline: 31392451]
6. Dissing AS, Jørgensen TB, Gerds TA, Rod NH, Lund R. High perceived stress and social interaction behaviour among young adults. A study based on objective measures of face-to-face and smartphone interactions. PLoS One 2019 Jul 26;14(7):e0218429 [FREE Full text] [doi: 10.1371/journal.pone.0218429] [Medline: 31348787]
7. Davis MT, Holmes SE, Pietrzak RH, Esterlis I. Neurobiology of chronic stress-related psychiatric disorders: evidence from molecular imaging studies. Chronic Stress (Thousand Oaks) 2017;1:2470547017710916 [FREE Full text] [doi: 10.1177/2470547017710916] [Medline: 29862379]
8. Levine GN, Cohen BE, Commodore-Mensah Y, Fleury J, Huffman JC, Khalid U, et al. Psychological health, well-being, and the mind-heart-body connection: a scientific statement from the American Heart Association. Circulation 2021 Mar 09;143(10):e763-e783 [FREE Full text] [doi: 10.1161/CIR.0000000000000947] [Medline: 33486973]
9. O'Connor DB, Thayer JF, Vedhara K. Stress and health: a review of psychobiological processes. Annu Rev Psychol 2021 Jan 04;72:663-688. [doi: 10.1146/annurev-psych-062520-122331] [Medline: 32886587]
10. Tyra AT, Soto SM, Young DA, Ginty AT. Frequency and perceptions of life stress are associated with reduced cardiovascular stress-response adaptation. Int J Psychophysiol 2020 Nov;157:51-60. [doi: 10.1016/j.ijpsycho.2020.08.006] [Medline: 32976887]
11. Ribeiro Í, Pereira R, Freire IV, de Oliveira BG, Casotti CA, Boery EN. Stress and quality of life among university students: a systematic literature review. Health Prof Educ 2018 May 15;4(2):70-77. [doi: 10.1016/j.hpe.2017.03.002]
12. Vankim NA, Nelson TF. Vigorous physical activity, mental health, perceived stress, and socializing among college students. Am J Health Promot 2013;28(1):7-15 [FREE Full text] [doi: 10.4278/ajhp.111101-QUAN-395] [Medline: 23470187]
13. Jimenez MP, DeVille NV, Elliott EG, Schiff JE, Wilt GE, Hart JE, et al. Associations between nature exposure and health: a review of the evidence. Int J Environ Res Public Health 2021 Apr 30;18(9):4790 [FREE Full text] [doi: 10.3390/ijerph18094790] [Medline: 33946197]
14. Kaczkurkin AN, Foa EB. Cognitive-behavioral therapy for anxiety disorders: an update on the empirical evidence. Dialogues Clin Neurosci 2015 Sep;17(3):337-346 [FREE Full text] [doi: 10.31887/DCNS.2015.17.3/akaczkurkin] [Medline: 26487814]

XSL•FO
RenderX

15. González-Valero G, Zurita-Ortega F, Ubago-Jiménez JL, Puertas-Molero P. Use of meditation and cognitive behavioral therapies for the treatment of stress, depression and anxiety in students. A systematic review and meta-analysis. Int J Environ Res Public Health 2019 Nov 10;16(22):4394 [FREE Full text] [doi: 10.3390/ijerph16224394] [Medline: 31717682]
16. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
17. Vidourek RA, King KA, Nabors LA, Merianos AL. Students' benefits and barriers to mental health help-seeking. Health Psychol Behav Med 2014 Jan 01;2(1):1009-1022 [FREE Full text] [doi: 10.1080/21642850.2014.963586] [Medline: 25750831]
18. Mitchell C, McMillan B, Hagan T. Mental health help-seeking behaviours in young adults. Br J Gen Pract 2017 Jan;67(654):8-9 [FREE Full text] [doi: 10.3399/bjgp17X688453] [Medline: 28034926]
19. Park MJ, Kim DJ, Lee U, Na EJ, Jeon HJ. A literature overview of virtual reality (VR) in treatment of psychiatric disorders: recent advances and limitations. Front Psychiatry 2019 Jul 19;10:505 [FREE Full text] [doi: 10.3389/fpsyt.2019.00505] [Medline: 31379623]
20. Goldenhersch E, Thrul J, Ungaretti J, Rosencovich N, Waitman C, Ceberio MR. Virtual reality smartphone-based intervention for smoking cessation: pilot randomized controlled trial on initial clinical efficacy and adherence. J Med Internet Res 2020 Jul 29;22(7):e17571 [FREE Full text] [doi: 10.2196/17571] [Medline: 32723722]
21. Maples-Keller JL, Yasinski C, Manjin N, Rothbaum BO. Virtual reality-enhanced extinction of phobias and post-traumatic stress. Neurotherapeutics 2017 Jul;14(3):554-563 [FREE Full text] [doi: 10.1007/s13311-017-0534-y] [Medline: 28512692]
22. Emmelkamp PM, Meyerbröker K. Virtual reality therapy in mental health. Annu Rev Clin Psychol 2021 May 07;17:495-519. [doi: 10.1146/annurev-clinpsy-081219-115923] [Medline: 33606946]
23. Milgram P, Kishino F. A taxonomy of mixed reality visual displays. IEICE Trans Inf Syst 1994 Dec;12(12):1321-1329 [FREE Full text]
24. Skarbez R, Smith M, Whitton MC. Revisiting Milgram and Kishino's reality-virtuality continuum. Front Virtual Real 2021 Mar 24;2. [doi: 10.3389/frvir.2021.647997]
25. Valmaggia LR, Latif L, Kempton MJ, Rus-Calafell M. Virtual reality in the psychological treatment for mental health problems: an systematic review of recent evidence. Psychiatry Res 2016 Feb 28;236:189-195. [doi: 10.1016/j.psychres.2016.01.015] [Medline: 26795129]
26. Kuhne C, Kecelioglu ED, Maltby S, Hood RJ, Knott B, Ditton E, et al. Direct comparison of virtual reality and 2D delivery on sense of presence, emotional and physiological outcome measures. Front Virtual Real 2023 Aug 17;4. [doi: 10.3389/frvir.2023.1211001]
27. Xu J, Khanotia A, Juni S, Ku J, Sami H, Lin V, et al. Effectiveness of virtual reality-based well-being interventions for stress reduction in young adults: systematic review. JMIR Ment Health 2024 Mar 29;11:e52186 [FREE Full text] [doi: 10.2196/52186] [Medline: 38551625]
28. Gao K, Boyd C, Wiederhold MD, Wiederhold BK. VR mobile solutions for chronic stress reduction in young adults. Stud Health Technol Inform 2014;199:88-93. [Medline: 24875697]
29. Vaquero-Blasco MA, Perez-Valero E, Morillas C, Lopez-Gordo MA. Virtual reality customized 360-degree experiences for stress relief. Sensors (Basel) 2021 Mar 22;21(6):2219 [FREE Full text] [doi: 10.3390/s21062219] [Medline: 33810135]
30. Riches S, Kaleva I, Nicholson SL, Payne-Gill J, Steer N, Azevedo L, et al. Virtual reality relaxation for stress in young adults: a remotely delivered pilot study in participants' homes. J Technol Behav Sci 2024 Mar 28;9:771-783. [doi: 10.1007/s41347-024-00394-x]
31. Jackson RW, Cao-Nasalga A, Chieng A, Pirkl A, Jagielo AD, Xu C, et al. Adding virtual reality mindful exposure therapy to a cancer center's tobacco treatment offerings: feasibility and acceptability single-group pilot study. JMIR Form Res 2024 Jul 23;8:e54817 [FREE Full text] [doi: 10.2196/54817] [Medline: 39042439]
32. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
33. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
34. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [doi: 10.2307/2136404]
35. Taylor JM. Psychometric analysis of the ten-item Perceived Stress Scale. Psychol Assess 2015 Mar;27(1):90-101. [doi: 10.1037/a0038100] [Medline: 25346996]
36. Reduce stress. MindCo Health. URL: https://mindcohealth.com/reduce-stress/ [accessed 2025-01-22]
37. Niazi AK, Niazi SK. Mindfulness-based stress reduction: a non-pharmacological approach for chronic illnesses. N Am J Med Sci 2011 Jan;3(1):20-23 [FREE Full text] [doi: 10.4297/najms.2011.320] [Medline: 22540058]
38. Kabat-Zinn J. Some reflections on the origins of MBSR, skillful means, and the trouble with maps. Contemp Buddhism 2011 Jun 14;12(1):281-306. [doi: 10.1080/14639947.2011.564844]

XSL•FO
RenderX

39. Shapiro S, Goldenhersch E, Toye E. The transformative power of mindfulness: enhancing behavioral change for a healthier lifestyle. In: Liamputtong P, editor. Handbook of Concepts in Health, Health Behavior and Environmental Health. Singapore: Springer; Feb 07, 2025.
40. Alharbi H, Alshehry A. Perceived stress and coping strategies among ICU nurses in government tertiary hospitals in Saudi Arabia: a cross-sectional study. Ann Saudi Med 2019;39(1):48-55 [FREE Full text] [doi: 10.5144/0256-4947.2019.48] [Medline: 30712051]
41. Lee EH. Review of the psychometric evidence of the Perceived Stress Scale. Asian Nurs Res (Korean Soc Nurs Sci) 2012 Dec;6(4):121-127 [FREE Full text] [doi: 10.1016/j.anr.2012.08.004] [Medline: 25031113]
42. Connor KM, Davidson JR. Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). Depress Anxiety 2003;18(2):76-82. [doi: 10.1002/da.10113] [Medline: 12964174]
43. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
44. Generalized Anxiety Disorder 7-item (GAD-7). National HIV Curriculum. URL: https://www.hiv.uw.edu/page/mental-health-screening/gad-7 [accessed 2025-07-08]
45. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
46. Patient Health Questionnaire-9 (PHQ-9). National HIV Curriculum. URL: https://www.hiv.uw.edu/page/mental-health-screening/phq-9 [accessed 2025-07-08]
47. Dupuy HJ. Self-representations of general psychological well-being of American adults. In: Proceedings of the American Public Health Association Meeting. 1978 Presented at: APHA 1978; Oct 15-19, 1978; Los Angeles, CA. [doi: 10.1037/t04083-000]
48. Taylor JE, Poston WS2, Haddock CK, Blackburn GL, Heber D, Heymsfield SB, et al. Psychometric characteristics of the General Well-Being Schedule (GWB) with African-American women. Qual Life Res 2003 Feb;12(1):31-39. [doi: 10.1023/a:1022052804109] [Medline: 12625516]
49. Vaishnavi S, Connor K, Davidson JR. An abbreviated version of the Connor-Davidson Resilience Scale (CD-RISC), the CD-RISC2: psychometric properties and applications in psychopharmacological trials. Psychiatry Res 2007 Aug 30;152(2-3):293-297 [FREE Full text] [doi: 10.1016/j.psychres.2007.01.006] [Medline: 17459488]
50. Johnson SU, Ulvenes PG, Øktedalen T, Hoffart A. Psychometric properties of the General Anxiety Disorder 7-item (GAD-7) scale in a heterogeneous psychiatric sample. Front Psychol 2019 Aug 06;10:1713 [FREE Full text] [doi: 10.3389/fpsyg.2019.01713] [Medline: 31447721]
51. Shaff J, Kahn G, Wilcox HC. An examination of the psychometric properties of the Patient Health Questionnaire-9 (PHQ-9) in a Multiracial/ethnic population in the United States. Front Psychiatry 2024 Jan 16;14:1290736 [FREE Full text] [doi: 10.3389/fpsyt.2023.1290736] [Medline: 38293592]
52. Gaston JE, Vogl L. Psychometric properties of the General Well-Being Index. Qual Life Res 2005 Feb;14(1):71-75. [doi: 10.1007/s11136-004-0793-z] [Medline: 15789942]
53. Tuitert I, Marinus JD, Dalenberg JR, van 't Veer JT. Digital health technology use across socioeconomic groups prior to and during the COVID-19 pandemic: panel study. JMIR Public Health Surveill 2024 Sep 13;10:e55384 [FREE Full text] [doi: 10.2196/55384] [Medline: 39269755]
54. Arsandaux J, Montagni I, Macalli M, Texier N, Pouriel M, Germain R, et al. Mental health condition of college students compared to non-students during COVID-19 lockdown: the CONFINS study. BMJ Open 2021 Aug 19;11(8):e053231 [FREE Full text] [doi: 10.1136/bmjopen-2021-053231] [Medline: 34413111]
55. Bravo AJ, Wedell E, Villarosa-Hurlocker MC, Looby A, Dickter CL, Schepis TS. Perceived racial/ethnic discrimination among young adult college students: prevalence rates and associations with mental health. J Am Coll Health 2023 Oct;71(7):2062-2073. [doi: 10.1080/07448481.2021.1954012] [Medline: 34398695]
56. Hamwey M, Norman C, Suss R, Tuazon E, Carey E, McAteer J, et al. The state of mental health of New Yorkers. New York City Department of Health and Mental Hygiene. 2024 May. URL: https://www.nyc.gov/assets/doh/downloads/pdf/mh/state-of-mental-health-new-yorkers.pdf [accessed 2026-07-28]
57. Ward BW, Martinez ME. Health insurance status and psychological distress among US adults aged 18-64 years. Stress Health 2015 Oct;31(4):324-335 [FREE Full text] [doi: 10.1002/smi.2559] [Medline: 24403273]
58. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2022. URL: https://www.R-project.org/ [accessed 2026-07-29]
59. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
60. McGarry S, Brown A, Gardner M, Plowright C, Skou R, Thompson C. Immersive virtual reality: an effective strategy for reducing stress in young adults. Br J Occup Ther 2023 Aug;86(8):560-567. [doi: 10.1177/03080226231165644] [Medline: 40337613]
61. Linardon J. Rates of attrition and engagement in randomized controlled trials of mindfulness apps: systematic review and meta-analysis. Behav Res Ther 2023 Nov;170:104421 [FREE Full text] [doi: 10.1016/j.brat.2023.104421] [Medline: 37862854]

XSL•FO
RenderX

62. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
63. Deng W, Hu D, Xu S, Liu X, Zhao J, Chen Q, et al. The efficacy of virtual reality exposure therapy for PTSD symptoms: a systematic review and meta-analysis. J Affect Disord 2019 Oct 01;257:698-709. [doi: 10.1016/j.jad.2019.07.086] [Medline: 31382122]
64. Maddox T, Oldstone L, Linde-Zwirble W, Bonakdar R, Maddox R, Sackman J, et al. Differential treatment response to virtual reality in high-impact chronic pain: secondary analysis of a randomized trial. Sci Rep 2025 Apr 25;15(1):14430 [FREE Full text] [doi: 10.1038/s41598-025-98716-3] [Medline: 40281004]
65. Chan SH, Qiu L, Esposito G, Mai KP, Tam KP, Cui J. Nature in virtual reality improves mood and reduces stress: evidence from young adults and senior citizens. Virtual Real 2021 Nov 26:1-16 [FREE Full text] [doi: 10.1007/s10055-021-00604-4] [Medline: 34849087]
66. Ladakis I, Filos D, Chouvarda I. Virtual reality environments for stress reduction and management: a scoping review. Virtual Real 2024 Feb 27;28:50. [doi: 10.1007/s10055-024-00943-y]
67. Davydov DM, Stewart R, Ritchie K, Chaudieu I. Resilience and mental health. Clin Psychol Rev 2010 Jul;30(5):479-495 [FREE Full text] [doi: 10.1016/j.cpr.2010.03.003] [Medline: 20395025]
68. Teoli D, Bhardwaj A. Quality of life (archived). StatPearls 2023:1-2 [FREE Full text] [Medline: 30725647]
69. Wilmer MT, Anderson K, Reynolds M. Correlates of quality of life in anxiety disorders: review of recent research. Curr Psychiatry Rep 2021 Oct 06;23(11):77 [FREE Full text] [doi: 10.1007/s11920-021-01290-4] [Medline: 34613508]
70. Jingili N, Oyelere SS, Nyström MB, Anyshchenko L. A systematic review on the efficacy of virtual reality and gamification interventions for managing anxiety and depression. Front Digit Health 2023 Nov 7;5:1239435 [FREE Full text] [doi: 10.3389/fdgth.2023.1239435] [Medline: 38026832]

Abbreviations

CBT: cognitive behavioral therapy
CD-RISC2: Connor-Davidson Resilience Scale–2
CONSORT: Consolidated Standards of Reporting Trials
CVD: cardiovascular disease
GAD-7: Generalized Anxiety Disorder–7
GWS: General Well-Being Schedule
ITT: intention-to-treat
MR: MindCo Relief
PHQ-9: Patient Health Questionnaire–9
PSS-10: Perceived Stress Scale–10
PTSD: posttraumatic stress disorder
QoL: quality of life
RCT: randomized controlled trial
VR: virtual reality

Edited by A Stone, L Buis; submitted 05.Sep.2025; peer-reviewed by F Balieva, R Endrighi, R Nooripour; comments to author 10.Feb.2026; revised version received 29.Jun.2026; accepted 16.Jul.2026; published 21.Aug.2026.

Please cite as:
Hong H, Lootens MR, Bachman R, Olivera Leon LD, Dunlap AF, Goldenhersch E, Rosencovich N, Waitman C, Wyka KE, Huang TTK
A Mobile Virtual Reality–Based Intervention for Stress Reduction Among Young Adults: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e83438
URL: https://mhealth.jmir.org/2026/1/e83438
doi:10.2196/83438
PMID:

©Houlin Hong, Matthew R Lootens, Rebecca Bachman, Luis David Olivera Leon, Ann F Dunlap, Emilio Goldenhersch, Nicolas Rosencovich, Cristian Waitman, Katarzyna E Wyka, Terry T-K Huang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 21.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete

XSL•FO
RenderX

bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

User Engagement and Feature Preferences in an AI-Powered mHealth Intervention for Diabetes Prevention: Secondary Analysis of a Randomized Controlled Trial

Benjamin Lalani[1], BS; Gabriela Siew[2]; Yllka Valdez[2], BA; Aliyah Shehadeh[2], BS; Daniel Zade[2], BS; Kristin Riekert[3], PhD; Nestoras Mathioudakis[2], MHS, MD; AI-DPP Study Group[4]

[1]Harvard Medical School, Harvard University, Boston, MA, United States

[2]Division of Endocrinology, Diabetes & Metabolism, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States

[3]Division of Pulmonary and Critical Care Medicine, Department of Medicine, Johns Hopkins Medicine, Baltimore, MD, United States

[4]See Acknowledgments

Corresponding Author:
Nestoras Mathioudakis, MHS, MD
Division of Endocrinology, Diabetes & Metabolism
Department of Medicine
Johns Hopkins University School of Medicine
1830 E. Monument Street, Suite 333`
Baltimore, MD, 21287
United States
Phone: 1 667 306 8085
Email: nmathio1@jhmi.edu

Abstract

Background: Prediabetes is highly prevalent and increasing globally, yet lifestyle interventions remain underused. AI-driven mobile health (mHealth) tools can help scale diabetes prevention efforts, but the key factors driving their success are not well understood.

Objective: This post hoc secondary analysis of a randomized controlled trial (RCT) aimed to characterize the most valued features and the role of user engagement in outcomes of a fully automated mHealth intervention for diabetes prevention.

Methods: Data from 151 participants with prediabetes and overweight or obesity who were assigned to an AI-based diabetes prevention program (Sweetch) in a parent RCT (NCT05056376) were analyzed. Engagement (defined as the total number of days the app was used) was categorized into tertiles (low, medium, and high). Baseline characteristics were compared across engagement groups using ANOVA, Kruskal-Wallis, and chi-square tests, and regression models assessed the association between engagement and achievement of diabetes risk reduction outcomes (≥5% weight loss, ≥4% weight loss with ≥150 min/week of physical activity, or ≥0.2 percentage point reduction in hemoglobin A_{1c} [HbA_{1c}] at 12 months). Perceived usefulness of intervention features was surveyed at 12 months.

Results: Median engagement was 98 (IQR 34-232) days. Older age (P<.001) and lower baseline BMI (P=.04) were significantly associated with higher engagement. Compared with low engagement, high engagement was associated with greater odds of achieving the composite diabetes risk reduction outcome (odds ratio [OR] 2.59, 95% CI 1.11-6.01; P=.03), ≥5% weight loss (OR 3.31, 95% CI 1.16-9.42; P=.03), and ≥0.2 percentage point reduction in HbA_{1c} (OR 3.57, 95% CI 1.19-10.75; P=.02). Participants most frequently rated weight tracking, physical activity tracking, and the digital body weight scale as the features that were most helpful for achieving their health goals.

Conclusions: Higher engagement with an AI-driven intervention requiring no human intervention was associated with improved diabetes risk reduction. Contrary to concerns about lower digital literacy, older adults engaged with the intervention more than younger adults. Features related to weight and physical activity tracking were most valued by patients in the program.

Trial Registration: ClinicalTrials.gov NCT05056376; https://clinicaltrials.gov/study/NCT05056376

(JMIR Mhealth Uhealth 2026;14:e92981) doi:10.2196/92981

XSL•FO
RenderX

KEYWORDS

artificial intelligence; digital health; mobile app; prediabetes; obesity; overweight; diabetes prevention; lifestyle intervention; user engagement; reinforcement learning

Introduction

Background

Mobile health (mHealth) technologies offer a promising modality for delivering diabetes prevention interventions at scale, with the potential to reach large populations at significantly lower costs than traditional, human-based interventions [1,2]. Various digital health tools, including mobile apps, web-based platforms, SMS text messaging, telemedicine, activity monitors, and voice calls, have been used to support diabetes prevention and weight management. These tools have yielded promising results in previous studies [3-5], and consequently, the Centers for Disease Control and Prevention (CDC) has recognized asynchronous, digital diabetes prevention programs (DPPs) within the National DPP framework [6].

Despite demonstrated efficacy in controlled settings, mHealth interventions have not translated into sustained real-world impact in diabetes prevention. Even with the availability of digital DPPs, which lower barriers to access compared with in-person programs, fewer than 1% of eligible adults participate in a DPP [7]. Engagement in intensive behavioral interventions is influenced by the perceived relevance, timing, and personalization of intervention content. With the emergence of AI, mHealth platforms have the potential to address these limitations by dynamically tailoring content delivery, feedback, and behavioral prompts to individual users at scale [8]. Unlike rule-based or static digital programs, AI-driven systems can adapt to user behavior over time, a capability that may enhance long-term engagement and increase participation. At the same time, skepticism persists regarding user willingness to engage with fully automated AI-based health interventions [9]. Concerns regarding trust and the perceived lack of human support raise important questions about whether AI-driven programs can sustain meaningful engagement over extended periods.

One such platform is Sweetch, an AI-driven, fully automated mHealth intervention that delivers personalized recommendations to promote physical activity and weight loss. A prior pilot study demonstrated the feasibility, acceptability, and short-term effectiveness of the Sweetch platform in adults with prediabetes [10]. A follow-up randomized, noninferiority clinical trial demonstrated that Sweetch was noninferior to the human coach–based DPP in reducing diabetes risk among patients with prediabetes [11]. However, beyond its clinical performance, the patterns, determinants, and perceived drivers of program engagement have not yet been examined. In addition, given the multicomponent nature of the Sweetch intervention, it remains unclear which features were perceived as most important in achieving these outcomes. Identifying these features may help inform the design and scaling of digital DPPs in the early stages of development [6].

Study Objectives

We conducted a post hoc secondary analysis of a randomized controlled trial (RCT) to evaluate engagement with and perceived feature-specific usefulness of the AI-driven DPP over a 12-month period. The objectives of this study were to (1) characterize longitudinal engagement with an AI-driven DPP and identify participant characteristics associated with higher engagement, (2) evaluate the association between engagement and clinical outcomes, and (3) assess participants' perceived usefulness of individual intervention features at the end of the intervention period. By elucidating how users interact with a fully automated AI-based DPP in real-world conditions, this study aims to inform the design, optimization, and scaling of future AI-driven diabetes prevention interventions.

Methods

Study Design and Population

The parent study was a multisite, pragmatic, parallel-group, 1:1 randomized noninferiority trial comparing a fully automated, AI-based DPP with a standard human-coached DPP over 12 months. Participants were randomized to either referral to an AI-powered DPP delivered through a mobile app and Bluetooth-enabled digital scale or referral to a CDC-recognized, human coach–led DPP delivered remotely over 12 months. The AI-led intervention provided personalized behavioral recommendations, educational content, self-monitoring tools, and adaptive push notifications without human coaching, whereas the comparator consisted of trained lifestyle coaches delivering the CDC PreventT2 curriculum through an initial core phase of 16 weekly sessions, followed by a maintenance phase of biweekly to monthly sessions. Both interventions were delivered independently of the study team after referral. The present study is a post hoc secondary analysis restricted to participants randomized to the AI-led DPP arm. The protocol and primary outcomes of the parent trial have been described elsewhere [11,12].

The parent trial was conducted from October 11, 2021, to December 16, 2024, with a planned 12-month follow-up for all participants. The 368 participants enrolled in the parent trial had laboratory-confirmed prediabetes and a race-specific BMI in the overweight or obese range (≥25 kg/m^2 or ≥23 kg/m^2 for Asian individuals). Exclusion criteria included a prior diagnosis of diabetes, severe cardiovascular conditions, factors affecting accuracy of hemoglobin A_{1c} (HbA_{1c}), cognitive or psychiatric barriers to participation, and conditions or medications expected to substantially affect body weight or blood glucose.

In the parent trial, eligible participants were randomized in a 1:1 ratio to the AI-based or human coach–led DPP using centrally administered REDCap (Vanderbilt University) software. Randomization was stratified by recruitment site and baseline HbA_{1c} (≤6% vs 6.1%-6.4%) using randomly permuted block sizes of 2, 4, and 8. The allocation sequence was

XSL•FO
RenderX

concealed from research coordinators responsible for participant enrollment and assignment until randomization occurred. Participants continued to receive routine medical care throughout the study. To minimize potential confounding, participation in other structured lifestyle interventions targeting nutrition, weight management, or diabetes prevention was not permitted, and medications expected to substantially influence body weight or glycemic control were prohibited during follow-up.

For this single-arm analysis, participants were included if they (1) were assigned to the AI arm of the trial and (2) had available outcomes data at 12 months. Participants were excluded from this analysis if they initiated any medications that could impact study outcomes during the intervention period, including antihyperglycemic agents, weight loss medications, or systemic steroids. No separate sample-size calculation was performed for this post hoc secondary analysis.

The clinical trial was conducted at 2 sites in the United States: Johns Hopkins Hospital, an urban tertiary academic medical center in Baltimore, Maryland; and Reading Hospital (Tower Health) in Reading, Pennsylvania, which serves a mixed urban, suburban, and rural population. The parent trial was prospectively registered at ClinicalTrials.gov (NCT05056376).

Outcomes

The parent trial prespecified a primary composite outcome assessed at 12 months, defined as maintenance of HbA_{1c} <6.5% throughout follow-up and achievement of at least 1 of the following: (1) ≥5% weight loss, (2) ≥4% weight loss plus ≥150 minutes/week of moderate-to-vigorous physical activity (MVPA), or (3) an absolute HbA_{1c} reduction of ≥0.2 percentage points. Secondary end points were the individual components of the composite primary end point; attainment of CDC-recommended weekly physical activity levels at 12 months; and continuous changes in body weight, HbA_{1c}, and average weekly physical activity over 12 months.

This secondary analysis focused on identifying factors associated with participant engagement and evaluating whether engagement level was associated with achieving end points related to weight loss, physical activity, and HbA_{1c} reduction. Thus, engagement served as the primary exposure variable, and the outcomes corresponded to the original end points from the parent trial.

Although the intervention incorporated Bluetooth-enabled devices and smartphone-derived data to support personalized recommendations, study outcomes were assessed independently. Body weight and HbA_{1c} were measured during study visits by trained research coordinators, and physical activity outcomes were derived from ActiGraph wrist monitors (Signant Health) worn for 7 consecutive days each month.

Engagement Definition and Data Collection

Engagement definitions were not prespecified because engagement data were collected passively by the manufacturer's platform, and the research team did not have access to these data until after trial completion. Participant engagement was automatically tracked by the Sweetch app (Sweetch Health Ltd) during months 1 to 11 of the trial and transferred to the research team after trial completion. Engagement was defined as the number of days over the course of the trial on which the Sweetch app was opened, and participants were categorized into tertiles based on cumulative use. Passive background processes (eg, automatic sensor data collection or data synchronization without opening the app) were not counted as engagement days. Engagement during month 12 was not included because participants completed their final study visits on different dates, resulting in unequal observation periods during the final month. Restricting engagement analyses to months 1 through 11 ensured a comparable exposure period for all participants before assessment of the 12-month outcomes. In addition, the number of days on which the app was opened in each month of the intervention period was assessed. Baseline characteristics were compared between engagement groups.

REDCap electronic data capture tools were used to collect participant baseline data and follow-up measures [12]. Baseline data were collected by trained research coordinators and included point-of-care HbA_{1c} measurements, height, weight, and demographics. Follow-up assessments at 6 and 12 months included repeat HbA_{1c} measurements, body weight, and surveys of perceived usefulness of intervention features. Physical activity was measured using ActiGraph wrist monitors worn for 7 consecutive days each month from randomization through month 11. MVPA was derived from raw accelerometer data using the Montoye wrist-worn cut point (≥3941 counts/min) [13]. Weekly MVPA was estimated as the mean across 11 monitoring periods.

Intervention Description

The AI-based DPP used in this study was Sweetch. Participants received a health kit from Sweetch Health Ltd within 8 to 12 days after randomization. The kit included a Bluetooth-enabled digital body weight scale and instructions for app registration and synchronization. To simulate real-world conditions, research staff did not intervene in participants' initiation or ongoing engagement with the app.

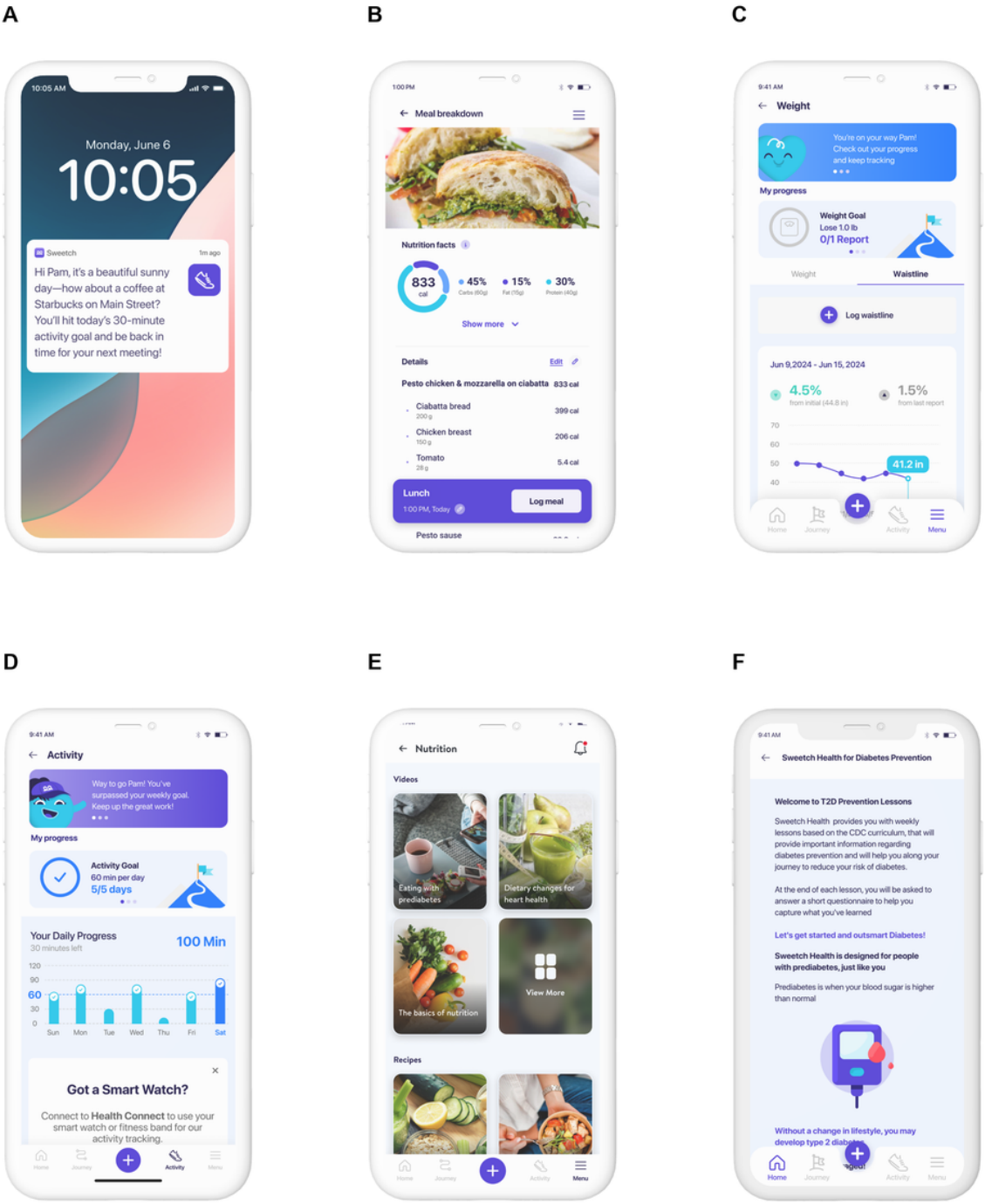
The intervention was fully automated, meaning that it was delivered without human coaching or scheduled human support after enrollment. Personalized recommendations and behavioral prompts were generated automatically by the AI-based platform, and use of the platform's features (eg, tracking tools and the health education library) was driven solely by the participants themselves.

The manufacturer was permitted to implement updates to the intervention during the study period (refer to Multimedia Appendix 1 for details on the timing and content of these updates).

The key features of the Sweetch intervention are displayed in Figure 1 and are summarized here.

XSL•FO
RenderX

Figure 1. Features of the Sweetch intervention. (A) Just-in-time adaptive push notification, (B) meal-logging screen with a nutrition breakdown, (C) weight-tracking dashboard showing trend and goal progress, (D) physical activity module summarizing daily and weekly activity, (E) nutrition content library featuring videos and recipes, and (F) educational lesson on type 2 diabetes prevention.



Push notifications enable personalized, real-time nudges promoting adherence to physical activity, weight management, and dietary goals (Figure 1A). These notifications are generated based on both active data inputs (eg, body weight and meal logging) and passive data inputs (eg, geolocation, smartphone accelerometer data, and calendar) and use a reinforcement learning algorithm that iteratively adapts its recommendations by assigning rewards or penalties based on user behavior [14,15].

Nutrition tracking allows structured meal logging, including options for manual entry using a food library and automated photo-based meal detection. This feature replaced self-rated meal quality assessments in previous versions of the app (Figure 1B).

Weight tracking tracks weight trends over time through a Bluetooth-enabled digital scale, manual entries, or synchronization from third-party platforms. Visual graphs and logs allow users to monitor progress, while seamless integration

XSL•FO
RenderX

with the Bluetooth-enabled scale ensures accurate, up-to-date data (Figure 1C).

Physical activity tracking automatically monitors users' daily activities using smartphone accelerometers or integrated wearable devices (eg, Apple Health and Google Fit; Figure 1B). Users can also manually log activities such as Pilates, yoga, and swimming. The app provides insights into activity levels and sets adaptive, personalized goals to encourage gradual improvement (Figure 1D).

An educational content library provides evidence-based resources and lessons on type 2 diabetes (T2D) prevention, accessible within the app interface (Figures 1E and 1F)

"My Achievements" tracks user progress by displaying milestones and rewarding positive behaviors, offering visual motivation to continue healthy habits

"My Challenges" presents tailored health goals for users to complete, fostering resilience and a sense of accomplishment with each task achieved

"My Competitions" introduces friendly competition by allowing users to compare progress with peers, enhancing motivation and accountability through engaging contests

"My Leaderboard" displays user rankings based on activity and achievements, promoting motivation and community engagement through visible progress tracking

Assessment of Perceived Feature Usefulness

After the intervention period (12 months), participants completed a survey assessing the perceived usefulness of specific app features in helping them achieve their health goals. The survey administered to participants is provided in Multimedia Appendix 1. Participants rated the following features on a 5-point Likert scale (1="strongly disagree" to 5="strongly agree"): push notifications, My Achievements, My Challenges, My Competitions, My Leaderboard, T2D prevention lessons, physical activity tracking, weight tracking, and the digital body weight scale. Four features were either discontinued before the trial (My Supporter and My Habits) or added during the intervention period (nutrition tracking and the educational content library) after survey finalization and were therefore not included in the feature usefulness assessment.

Statistical Analysis

Baseline characteristics were summarized using medians and IQRs for continuous variables and frequencies with percentages for categorical variables. Differences in baseline characteristics across engagement categories (low, moderate, and high tertiles of total app use days) were assessed using 1-way ANOVA for normally distributed continuous variables, Kruskal-Wallis tests for nonnormally distributed continuous variables, and the Pearson chi-square tests for categorical variables. Monthly engagement patterns were examined descriptively across engagement groups.

To evaluate associations between engagement and study outcomes, logistic regression models were used for binary end points, including the primary composite end point and its individual components (eg, ≥5% weight loss, ≥0.2 percentage point reduction in HbA_{1c}, and ≥4% weight loss combined with ≥150 minutes per week of physical activity), and linear regression models were used for continuous outcomes, including changes in body weight, HbA_{1c}, and mean physical activity from baseline to 12 months. As an additional analysis, engagement was modeled as a continuous exposure, defined as total app-open days over the 330-day study period and scaled per 10-day increase for interpretability. For all analyses evaluating HbA_{1c} change at 12 months, models were restricted to participants with a baseline HbA_{1c} in the prediabetes range (5.7%-6.4%).

We performed a targeted sensitivity analysis adjusting for baseline BMI, baseline HbA_{1c}, and baseline physical activity because these baseline measures correspond directly to the outcomes being evaluated (weight loss, HbA_{1c} reduction, and achievement of physical activity targets) and therefore could influence participants' likelihood of achieving the study outcomes. Additional adjustment for age, sex, race, education, and study site was not performed because the limited number of outcome events would have resulted in low events-per-variable ratios and an increased risk of model overfitting and unstable coefficient estimates.

Missing data were handled using a complete-case approach. Participants without 12-month outcome data were excluded from the analytic cohort. Analyses of perceived feature usefulness included only participants who completed the 12-month follow-up survey. There was no missing engagement data for any participant, as these data were collected continuously and automatically. No imputation of missing data was performed for this secondary analysis.

All statistical tests were 2-tailed, and a significance threshold of $P<.05$ was used.

Due to the post hoc nature of the study, these analyses were exploratory, and the findings are presented as hypothesis generating.

Ethical Considerations

The parent RCT (ClinicalTrials.gov; NCT05056376) was conducted in accordance with the Declaration of Helsinki and approved by the institutional review boards at Johns Hopkins Hospital and Reading Hospital (Tower Health). All participants provided written informed consent before randomization, and the original informed consent form and institutional review board approval permitted secondary analyses of the study data without additional participant contact or consent. This secondary analysis used deidentified data collected as part of the approved parent trial and did not require additional participant contact or consent. Participants received compensation for study-related assessments, as specified in the parent trial protocol [11,12]. No additional compensation was provided for this secondary analysis.

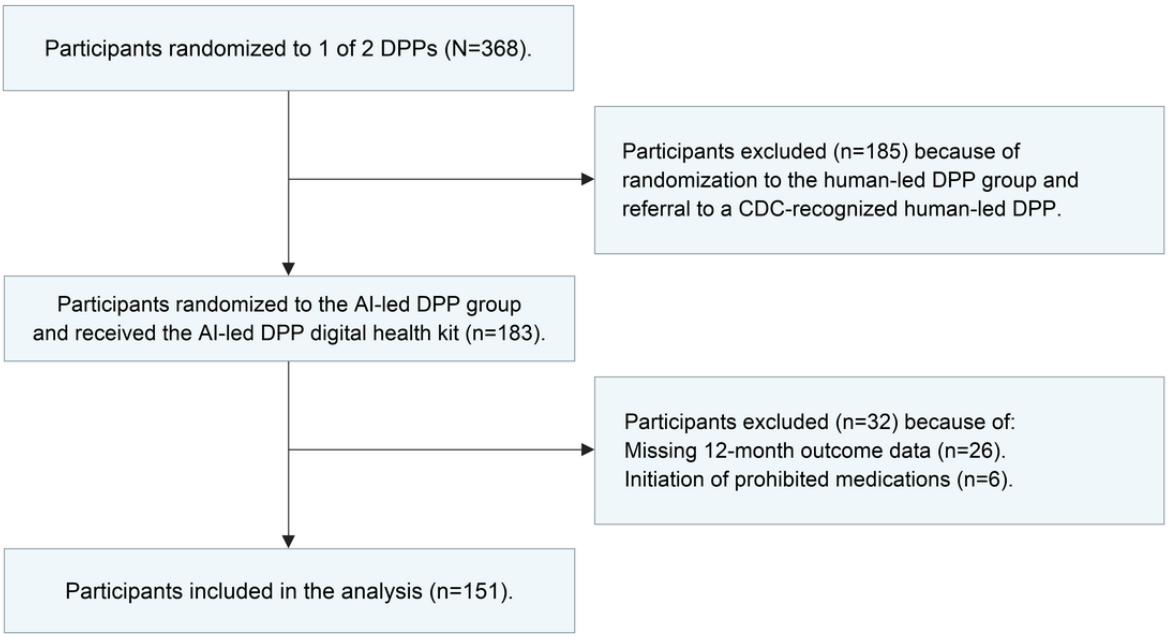
XSL•FO
RenderX

Results

Participants

Figure 2 illustrates participant flow from randomization through exclusions, resulting in 151 participants included in the final analytic sample; baseline characteristics are shown in Table 1. Participants were predominantly female and White, with high educational attainment and most being married or partnered. Recruitment occurred across 2 health systems. The median BMI was 32.3 (IQR 28.2-35.9) kg/m^2, and the mean HbA_{1c} at baseline was 5.8% (SD 0.3%).

Figure 2. Study flow diagram. CDC: Centers for Disease Control and Prevention; DPP: diabetes prevention program.



XSL•FO
RenderX

Table 1. Baseline characteristics of the study population (N=151).

Characteristics	Values
Site, n (%)	
Johns Hopkins Hospital	93 (61.6)
Reading Hospital (Tower Health)	58 (38.4)
Age (years), median (IQR)	61 (52-69)
Sex, n (%)	
Female	96 (63.6)
Male	55 (36.4)
Race, n (%)	
Asian	10 (6.6)
Black or African American	35 (23.2)
White or Caucasian	99 (65.6)
>1 race	3 (2)
Other	4 (2.6)
Ethnicity, n (%)	
Not Hispanic or Latino	146 (96.7)
Hispanic or Latino	3 (2)
Declined to answer	1 (0.7)
Unknown	1 (0.7)
Marital status, n (%)	
Married or partnered	102 (67.5)
Previously married (divorced, separated, or widowed)	23 (15.2)
Single or other	26 (17.2)
Educational attainment, n (%)	
High school or less	22 (14.6)
Some college or associate degree	31 (20.5)
Bachelor degree	39 (25.8)
Graduate or professional degree	59 (39.1)
BMI, median (IQR)	32.3 (28.2-35.9)
BMI classification, n (%)	
Overweight	44 (29.1)
Class I obesity	59 (39.1)
Class II obesity	28 (18.5)
Class III obesity	20 (13.2)
Hemoglobin A_{1c} (%), mean (SD)	5.8 (0.3)
Moderate-to-vigorous physical activity (min/week), median (IQR)	250 (112-404)
Diet quality, n (%)	
Healthiest diet (1-5)	58 (38.4)
Moderately healthy diet (6-10)	42 (27.8)
Least healthy diet (11-16)	51 (33.8)

XSL•FO
RenderX

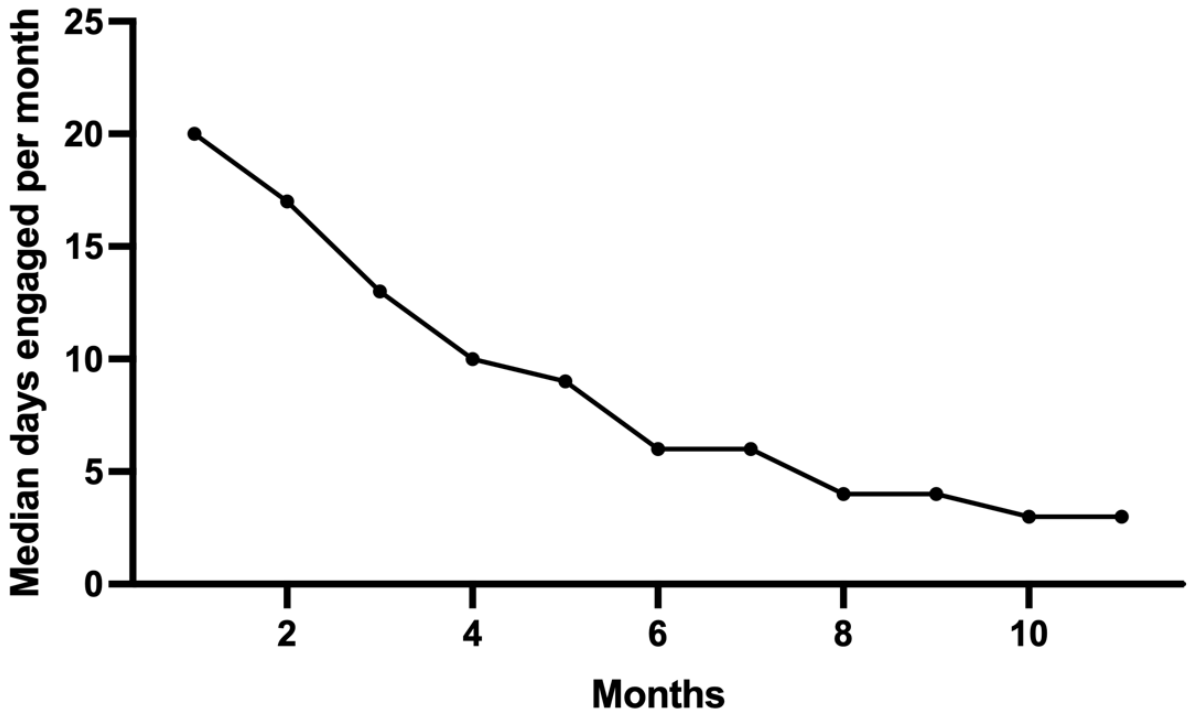
Engagement Patterns During the Intervention Period

Table 2 shows the distribution of total days engaged throughout the trial. Median engagement was 98 (IQR 34-232) days. The distribution was roughly bimodal, and engagement tertiles were defined as low (≤49 open days), moderate (50-175 open days), and high (>175 open days). The median number of days engaged per month among the cohort was 9 (IQR 3-21) days. Monthly engagement patterns (showing the median number of days engaged each month among the cohort) revealed a gradual decrease in app engagement over time (Figure 3).

Table 2. Distribution of total days engaged during the trial period (N=151).

Days engaged during trial period	Participants, n (%)
0	22 (14.6)
30	23 (15.2)
60	20 (13.2)
90	13 (8.6)
120	11 (7.3)
150	10 (6.6)
180	4 (2.6)
210	8 (5.3)
240	9 (6)
270	8 (5.3)
300	9 (6)
330	14 (9.3)

Figure 3. Engagement with the intervention during the study period. The median monthly engagement trajectory for all participants is shown over the intervention period.



Factors Associated With Increased Engagement

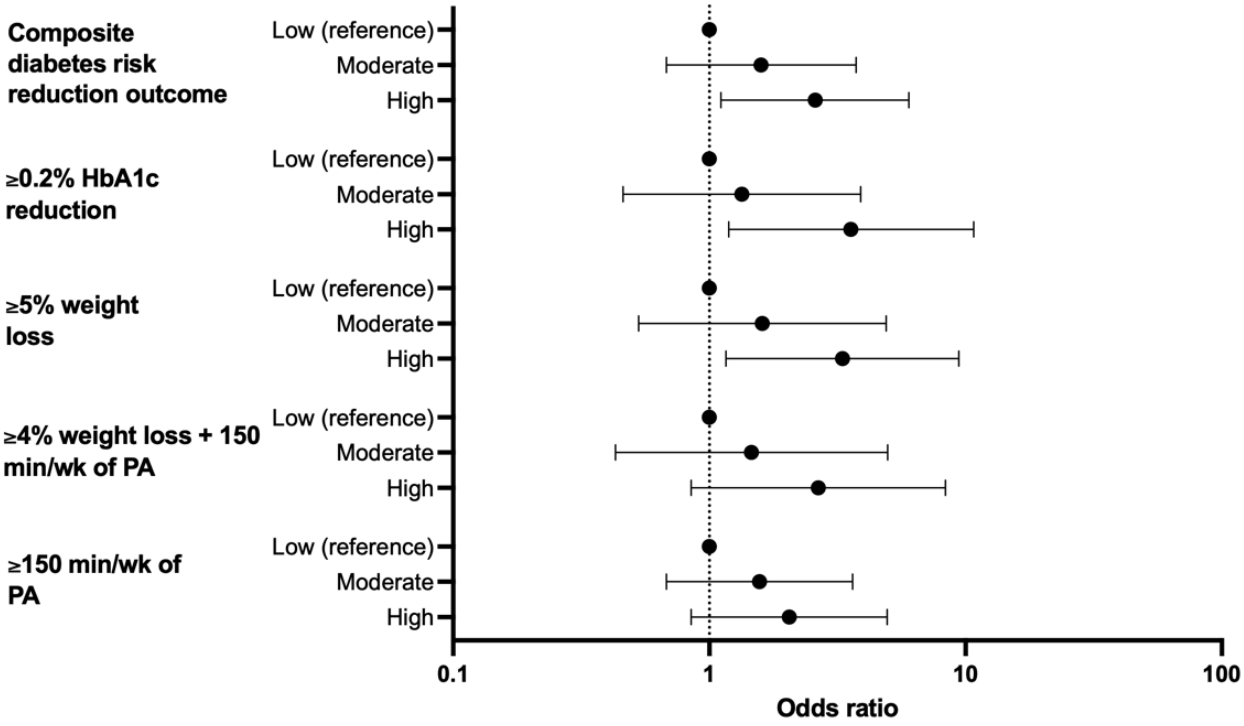
Several baseline factors were associated with engagement level. Participants in the low engagement group were significantly younger (median 52, IQR 42-63 years) than those in the moderate (median 63, IQR 55-68 years) and high (median 65, IQR 58-70 years) engagement groups (P<.001). They also had a higher median BMI (33.5, IQR 30.4-37.6 kg/m^2) than those in the high engagement group (30.8 kg/m^2, IQR 27.5-34.0; P=.04). There were no significant differences across engagement levels by sex, race, educational attainment, physical activity, or diet quality (Table S1 in Multimedia Appendix 1).

XSL•FO
RenderX

Association Between Engagement and Outcomes

Higher engagement was associated with greater odds of achieving the composite diabetes risk reduction benchmark (odds ratio [OR] 2.59, 95% CI 1.11-6.01; *P*=.03), ≥5% weight loss (OR 3.31, 95% CI 1.16-9.42; *P*=.03), and a ≥0.2 percentage point reduction in HbA_{1c} (OR 3.57, 95% CI 1.19-10.75; *P*=.02) compared with low engagement (Figure 4). Moderate engagement was not significantly associated with any outcome. Engagement level was not significantly associated with achieving ≥4% weight loss combined with 150 minutes per week of physical activity or with meeting the physical activity threshold alone. The sensitivity analysis of the primary composite outcome, adjusting for baseline BMI, baseline HbA_{1c}, and baseline physical activity, yielded results consistent with the primary analysis (Table S2 in Multimedia Appendix 1).

Figure 4. Odds ratios (ORs) for clinical outcomes at 12 months by engagement level. OR and 95% CI are shown for each engagement category, with low engagement as the reference group. Engagement categories represent tertiles based on the cumulative number of days the app was opened during the trial. HbA1c: hemoglobin A1c; PA: physical activity.



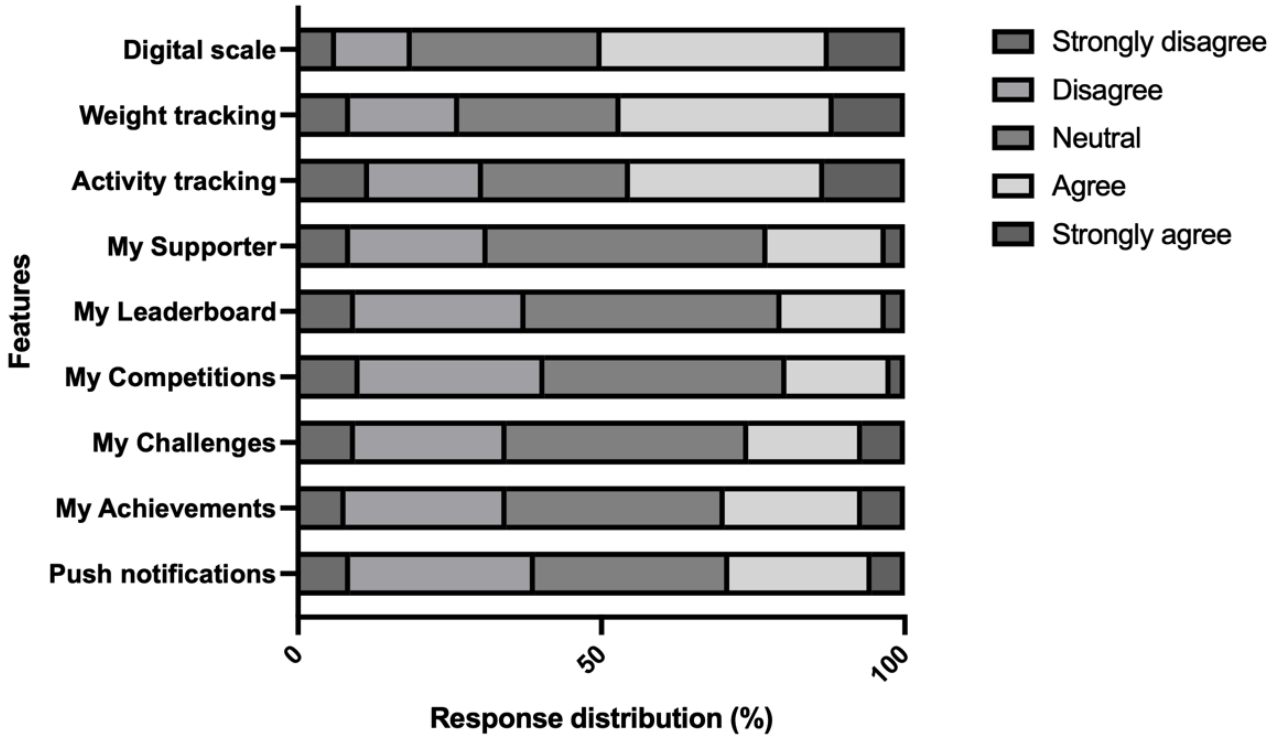
In linear regression analyses of continuous outcomes, high engagement was associated with greater weight loss at 12 months, whereas no significant associations were observed for changes in HbA_{1c} or physical activity (Table S3 in Multimedia Appendix 1). When engagement was modeled as a continuous variable, higher engagement was associated with modest increases in the odds of achieving the composite diabetes risk reduction outcome, ≥0.2 percentage point reduction in HbA_{1c}, and ≥5% weight loss, with no significant associations observed for physical activity outcomes (Table S4 in Multimedia Appendix 1).

User Preferences After 12 Months of Intervention

Survey data were available for 128 of 151 (84.8%) participants included in the analytic sample. Participants' perceptions of whether individual app features helped them achieve their goals are summarized in Figure 5, with corresponding raw counts and percentages reported in Table S5 in Multimedia Appendix 1. For several features, a neutral response was the most frequently selected among the 128 participants, including achievements (n=46, 35.9%), challenges (n=51, 39.8%), competitions (n=51, 39.8%), leaderboard (n=54, 42.2%), and T2D prevention lessons (n=59, 46.1%).

XSL•FO
RenderX

Figure 5. Perceived usefulness of features in the intervention. The horizontal stacked bar chart shows the percentage distribution of responses to survey items assessing whether each app feature helped participants achieve their goals. Responses were recorded on a 5-point Likert scale ranging from "strongly disagree" to "strongly agree." Each bar represents 1 feature and sums to 100% of responses (n=128 per item).



In contrast, tracking-oriented features elicited more favorable perceptions. For activity tracking, 58 (45.3%) participants reported agreeing or strongly agreeing, compared with 39 (30.5%) reporting disagreeing or strongly disagreeing. Similar patterns were observed for weight tracking, with 60 (46.9%) participants reporting agreeing or strongly agreeing. The digital scale demonstrated the most positive response distribution overall, with 64 (50%) participants agreeing or strongly agreeing.

For push notifications, among the 128 participants, 39 (30.5%) indicated disagreement, 41 (32%) were neutral, 30 (23.4%) indicated agreement, and 7 (5.5%) indicated strong agreement.

Discussion

Principal Findings

In this secondary analysis of an RCT, engagement (ie, greater amounts of app use) with an AI-led DPP was associated with achieving the CDC's diabetes risk reduction benchmark and recommended thresholds for weight loss and HbA_{1c} reduction. These findings reinforce app engagement as an important mechanism through which mHealth interventions may exert clinical benefit and extend prior digital health literature by demonstrating this relationship in the context of a fully automated AI-based program delivered over a year-long intervention period [3,16-18].

Factors Associated With Engagement

Engagement varied by participant characteristics. Older participants demonstrated higher levels of engagement, challenging assumptions that older adults engage less with digital health interventions [19,20]. This may reflect greater health motivation, more available time, or a stronger perceived need among older users. Participants with higher baseline BMI engaged less with the intervention, which may be a result of greater physical, behavioral, or psychological barriers to sustained self-monitoring and physical activity. Because participants with higher baseline BMI are at greater metabolic risk, lower engagement in this subgroup raises the possibility that additional tailoring, support, or re-engagement strategies may be needed to ensure that fully automated DPPs do not inadvertently widen disparities by preferentially benefiting individuals at lower metabolic risk.

Engagement with the intervention declined over time, which may reflect a natural waning of novelty, behavior change fatigue, and/or a perceived decline in the app's ongoing value. Alternatively, it may indicate that users acquired the knowledge or behavioral reinforcement they needed early on, reducing their reliance on the app over time. The Hawthorne effect, whereby initial behavior changes are driven by awareness of being observed, may also explain early spikes followed by tapering engagement [21]. The observed decline in engagement highlights the need for future work to identify strategies to maintain or re-engage users over time, particularly in longer-term prevention programs.

Interpretation of the Engagement-Outcome Association

This exploratory analysis identified an association between higher engagement with the intervention and a greater likelihood of achieving diabetes risk reduction outcomes. However, several alternative explanations should be considered. Participants who engaged with the app more frequently may have differed

XSL•FO
RenderX

systematically from those with lower engagement in ways not captured by measured baseline characteristics, including motivation, self-regulation, or readiness for behavior change. Reverse causality is also possible, as participants who experienced early weight loss or other positive reinforcement may have become more engaged over time rather than engagement itself driving these improvements. Consequently, the observed associations should not be interpreted as causal, and residual confounding cannot be excluded.

Notably, the engagement metric used in this analysis may not fully capture all aspects of intervention engagement. Some components of the Sweetch intervention, including meal logging, weight tracking, educational content, and progress monitoring, require active interaction with the app. In contrast, the AI-generated push notifications were designed to deliver personalized behavioral support passively and could influence participants without requiring the app to be opened. Furthermore, the lifestyle behaviors targeted by the intervention occur outside the app, meaning that behavior change initiated through either active or passive intervention components may be sustained without ongoing app interaction.

Perceived Usefulness of Features

Weight tracking, physical activity tracking, and use of the digital scale were more frequently endorsed as helpful in achieving health goals than social or gamified features, such as leaderboards and competitions, which were often rated as neutral. This aligns with findings from a large-scale evaluation of the NHS digital DPP, which found that participants most frequently engaged with tracking features, while peer support was rarely used [22].

Push notifications, which serve as the primary application for AI-driven personalization and are intended to approximate aspects of human coaching, were not perceived as among the most useful features, whereas participants more frequently endorsed self-monitoring features that do not rely directly on AI (eg, weight tracking, physical activity tracking, and the connected digital scale). These findings suggest that the contribution of AI-driven personalization to user engagement and clinical benefit remains uncertain and highlight the need for future studies to better define the most effective role of AI within digital DPPs.

In the context of the parent trial's demonstration of noninferiority to a human-coached DPP, these findings raise the possibility that comparable clinical benefit may have been achieved through mechanisms distinct from human support, with advanced, low-burden self-monitoring capabilities playing a central role. Other programs, such as Lark [23], have similarly adopted AI-based personalization through chatbot interfaces, reflecting a broader trend toward individualized digital experiences. However, the relative value of these features remains uncertain from the user perspective. As digital DPPs continue to scale, the findings from this study underscore the need to better understand how users interact with individual components and suggest that refining core tracking functionality may be as important as, or potentially more impactful than, expanding complex personalization systems.

Strengths

This study has strengths. It draws on prospectively collected data from a RCT with a 12-month follow-up period and high retention. Unlike prior studies that report engagement descriptively without treating it as a formal research end point, this study prospectively measured engagement, analyzed its association with clinical outcomes, and examined use patterns over time [24]. The pragmatic nature of the trial, specifically the absence of prompts from research staff to engage with the intervention, allowed participant behavior to more closely reflect real-world dynamics. The multisite design and diverse participant demographics enhance generalizability, and the use of validated, policy-relevant outcomes for diabetes risk reduction and objectively measured physical activity data strengthen the reliability of the findings.

Limitations

Our study has limitations. Because the research team did not have access to the engagement data structure until after trial completion, engagement definitions and analytic thresholds could not be prespecified. Consequently, this post hoc secondary analysis should be considered exploratory and hypothesis generating, and the observed associations warrant confirmation in prospectively designed studies. Residual confounding by measured baseline factors (eg, sex, age, race, and education) and unmeasured baseline characteristics (eg, health status, motivation, and readiness for behavior change) cannot be excluded. Given the limited number of outcome events, adjusted multivariable models were not constructed because of potential overfitting and unstable coefficient estimates. Engagement was defined solely by the number of days the app was opened, which may not capture the quality or type of user interaction, although this is a commonly used metric for measuring engagement in mHealth interventions [24]. Engagement was categorized using sample-dependent tertiles due to the lack of accepted criteria for thresholds of meaningful or adequate use, limiting interpretability and comparability across studies. Consequently, the definitions of low, moderate, and high engagement used in this study may not generalize to other digital interventions or real-world populations. In addition, feature usefulness was assessed only at the 12-month follow-up and may therefore not reflect the perspectives of participants who disengaged previously and did not complete the survey. Iterative updates to the Sweetch platform during the study period could have influenced user experience, engagement patterns, and participants' perceptions of individual intervention features, as not all participants were necessarily exposed to identical versions of the app throughout the study. Finally, the trial was conducted during the COVID-19 pandemic, which may have affected participants' health behaviors or ability to engage with the app.

Conclusions

In summary, this study found that higher engagement with a fully automated, AI-driven DPP was associated with clinically meaningful diabetes risk reduction outcomes. Engagement was notably higher among older adults and individuals with lower BMI, challenging assumptions about digital literacy barriers and suggesting possible differences in motivation or perceived

XSL•FO
RenderX

need. Among the app's features, integrated tracking tools, such as weight and activity monitoring, were rated most positively. Future research should further investigate the relative value of specific features within digital DPPs, particularly in fully automated platforms, and explore how such programs perform in broader real-world settings beyond the trial environment. As digital health interventions continue to expand, identifying the most effective and user-preferred components will be essential for designing scalable, cost-effective, and sustainable tools for diabetes prevention.

Acknowledgments

The authors thank the participants in this trial, particularly for their participation during the challenging period of the COVID-19 pandemic. The members of the AI-DPP Study Group are Nestoras N Mathioudakis, MD, MHS; Benjamin Lalani, BS; Mohammed S Abusamaan, MD, MPH; Adrian Dobs, MD, MHS; Jeromie Ballreich, PhD, MHS; Nisa Maruthur, MD, MHS; John McGready, PhD; Kristin Riekert, PhD; Amal Wanigatunga, PhD; Brian Kane, MD; Aliyah Shehadeh, BS; Defne Alver, BS, MHS; Benjamin Ringham, BS, MS; Fatima Vandi, BS, MS; Daniel Zade, BS; Pamela Morrero, CCRP; Amy Schiller, BS; Mary (Ellie) Alderfer, MSN, RN; Nae-Yuh Wang, PhD; Seth Martin, MD, MHS; and Sara Benjamin-Neelon, PhD, JD, MPH. The authors also thank the program directors and lifestyle coaches from the participating human coach-based diabetes prevention programs who implemented the intervention and provided attendance data to the coordinating center: Amanda Kulp, Chris Yocom, and Debbie Zlomek (Pottstown Medical Specialists Inc DPP); Dawn Batman, Tray Flynn, and Doreen Hespell (Montgomery County DPP); Megan Brown, Nicole Cronin, and Sahirmayi Chalimila (Brancati Center DPP); Mary Jane E Vaeth, Alexandra Morán, Dylan Blough, and Angela Ginn-Meadow (University of Maryland Medical Center DPP). The authors further acknowledge the team at Sweetch Health Ltd for their vision and successful delivery of the AI-led DPP: Yossi Bahagon, Miri Mishor-Goldenberg, Jennifer Meller, Noam Diamant, Bar Harpazi, Roy Sonnenberg, and Dana Chanan. None of these acknowledged individuals received direct compensation for their contributions. Generative AI (ChatGPT version 5.5; OpenAI) was used solely for ideation and brainstorming during manuscript preparation. It was not used to generate scientific content, interpret study findings, or draft the final manuscript. All final writing, revisions, and editorial decisions were completed and verified by the authors.

Data Availability

As of the time of this submission, the authors do not have institutional review board (IRB) approval to share this dataset. However, the consent language used in the study may permit the data to be submitted to a controlled-access repository (eg, Vivli) following deidentification and IRB approval of a change in research.

Funding

This study was funded by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK; grant R01DK125780). This study was also supported by the Johns Hopkins Institute for Clinical and Translational Research (ICTR), which is funded in part by a grant from the National Center for Advancing Translational Sciences (NCATS; grant UL1TR001079), a component of the National Institutes of Health (NIH), and the NIH Roadmap for Medical Research. The funders had no role in the study design, data collection, data analysis, interpretation of the results, writing of the manuscript, or the decision to submit the manuscript for publication.

Authors' Contributions

Conceptualization: BL, NM
Data curation: BL, DZ, NM
Formal analysis: BL, NM
Funding acquisition: NM
Investigation: BL, GS, YV, AS, DZ, KR, NM
Methodology: BL, KR, NM
Project administration: BL, NM
Resources: NM
Software: BL, NM
Supervision: NM
Validation: BL, DZ, NM
Visualization: BL
Writing—original draft: BL, NM
Writing—review and editing: BL, GS, YV, AS, DZ, KR, NM

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Survey materials, software update details, baseline characteristics stratified by engagement level, baseline characteristics stratified by engagement levels, regression analyses examining associations between engagement and primary, continuous, and composite trial outcomes, and participant feature preference survey results.
[DOCX File , 186 KB - mhealth_v14i1e92981_app1.docx]

Multimedia Appendix 2
CONSORT checklist.
[PDF File (Adobe PDF File), 140 KB - mhealth_v14i1e92981_app2.pdf]

References

1. Rinaldi G, Hijazi A, Haghparast-Bidgoli H. Cost and cost-effectiveness of mHealth interventions for the prevention and control of type 2 diabetes mellitus: a systematic review. Diabetes Res Clin Pract 2020 Apr;162:108084. [doi: 10.1016/j.diabres.2020.108084] [Medline: 32061819]
2. Park S, Ballreich J, Ward T, Shi L. Cost-effectiveness analysis of a digital diabetes-prevention programme versus an in-person diabetes-prevention programme in people with prediabetes in the United States. Diabetes Obes Metab 2024 Oct;26(10):4522-4534. [doi: 10.1111/dom.15807] [Medline: 39056211]
3. Katula JA, Dressler EV, Kittel CA, Harvin LN, Almeida FA, Wilson KE, et al. Effects of a digital diabetes prevention program: an RCT. Am J Prev Med 2022 Apr;62(4):567-577 [FREE Full text] [doi: 10.1016/j.amepre.2021.10.023] [Medline: 35151522]
4. Auster-Gussman LA, Lockwood KG, Graham SA, Pitter V, Branch OH. Engagement in digital health app-based prevention programs is associated with weight loss among adults age 65. Front Digit Health 2022 May 19;4:886783 [FREE Full text] [doi: 10.3389/fdgth.2022.886783] [Medline: 35663278]
5. Joiner KL, Nam S, Whittemore R. Lifestyle interventions based on the diabetes prevention program delivered via eHealth: a systematic review and meta-analysis. Prev Med 2017 Jul;100:194-207 [FREE Full text] [doi: 10.1016/j.ypmed.2017.04.033] [Medline: 28456513]
6. Lalani B, Shim J, Vadini V, Valdez Y, Zade D, Mathioudakis N. Consumer-oriented review of digital diabetes prevention programs: insights from the CDC's Diabetes Prevention Recognition Program. Front Clin Diabetes Healthc 2025 Apr 16;6:1562108 [FREE Full text] [doi: 10.3389/fcdhc.2025.1562108] [Medline: 40308291]
7. Ackermann RT. The U.S. National Diabetes Prevention Program (NDPP) shows promise as a cost-effective implementation strategy. Diabetes Care 2025 Jul 01;48(7):1150-1151. [doi: 10.2337/dci24-0100] [Medline: 40540680]
8. Lalani B, Herur R, Zade D, Collins G, Dishong DM, Mehta S, et al. Applications of artificial intelligence and machine learning in prediabetes: a scoping review. J Diabetes Sci Technol 2025 Jul 08:19322968251351995 (forthcoming). [doi: 10.1177/19322968251351995] [Medline: 40626375]
9. Osnat B. Patient perspectives on artificial intelligence in healthcare: a global scoping review of benefits, ethical concerns, and implementation strategies. Int J Med Inform 2025 Nov;203:106007. [doi: 10.1016/j.ijmedinf.2025.106007] [Medline: 40494217]
10. Everett E, Kane B, Yoo A, Dobs A, Mathioudakis N. A novel approach for fully automated, personalized health coaching for adults with prediabetes: pilot clinical trial. J Med Internet Res 2018 Feb 27;20(2):e72 [FREE Full text] [doi: 10.2196/jmir.9723] [Medline: 29487046]
11. Mathioudakis N, Lalani B, Abusamaan MS, Alderfer M, Alver D, Dobs A, et al. An AI-powered lifestyle intervention vs human coaching in the diabetes prevention program: a randomized clinical trial. JAMA 2025 Dec 16;334(23):2079-2089. [doi: 10.1001/jama.2025.19563] [Medline: 41144242]
12. Abusamaan MS, Ballreich J, Dobs A, Kane B, Maruthur N, McGready J, et al. Effectiveness of artificial intelligence vs. human coaching in diabetes prevention: a study protocol for a randomized controlled trial. Trials 2024 May 16;25(1):325 [FREE Full text] [doi: 10.1186/s13063-024-08177-8] [Medline: 38755706]
13. Montoye AH, Clevenger KA, Pfeiffer KA, Nelson MB, Bock JM, Imboden MT, et al. Development of cut-points for determining activity intensity from a wrist-worn ActiGraph accelerometer in free-living adults. J Sports Sci 2020 Nov;38(22):2569-2578. [doi: 10.1080/02640414.2020.1794244] [Medline: 32677510]
14. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
15. Systems and methods for prediction, prevention and management of chronic and autoimmune diseases. Google Patents. 2024. URL: https://patents.google.com/patent/US20240296958A1/en [accessed 2026-08-17]
16. Lavikainen P, Mattila E, Absetz P, Harjumaa M, Lindström J, Järvelä-Reijonen E, et al. Digitally supported lifestyle intervention to prevent type 2 diabetes through healthy habits: secondary analysis of long-term user engagement trajectories in a randomized controlled trial. J Med Internet Res 2022 Feb 24;24(2):e31530 [FREE Full text] [doi: 10.2196/31530] [Medline: 35200147]

XSL•FO
RenderX

17. Toro-Ramos T, Michaelides A, Anton M, Karim Z, Kang-Oh L, Argyrou C, et al. Mobile delivery of the diabetes prevention program in people with prediabetes: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 08;8(7):e17842 [FREE Full text] [doi: 10.2196/17842] [Medline: 32459631]
18. Sepah SC, Jiang L, Ellis RJ, McDermott K, Peters AL. Engagement and outcomes in a digital Diabetes Prevention Program: 3-year update. BMJ Open Diabetes Res Care 2017 Sep 07;5(1):e000422 [FREE Full text] [doi: 10.1136/bmjdrc-2017-000422] [Medline: 28948027]
19. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health 2021 Aug 17;21(1):1556 [FREE Full text] [doi: 10.1186/s12889-021-11623-w] [Medline: 34399716]
20. Bertolazzi A, Quaglia V, Bongelli R. Barriers and facilitators to health technology adoption by older adults with chronic diseases: an integrative systematic review. BMC Public Health 2024 Feb 16;24(1):506 [FREE Full text] [doi: 10.1186/s12889-024-18036-5] [Medline: 38365698]
21. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J Clin Epidemiol 2014 Mar;67(3):267-277 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.08.015] [Medline: 24275499]
22. Ross J, Hawkes RE, Miles LM, Cotterill S, Bower P, Murray E. Design and early use of the nationally implemented Healthier You National Health Service Digital Diabetes Prevention Programme: mixed methods study. J Med Internet Res 2023 Aug 17;25:e47436 [FREE Full text] [doi: 10.2196/47436] [Medline: 37590056]
23. Graham SA, Pitter V, Hori JH, Stein N, Branch OH. Weight loss in a digital app-based diabetes prevention program powered by artificial intelligence. Digit Health 2022 Oct 9;8:20552076221130619 [FREE Full text] [doi: 10.1177/20552076221130619] [Medline: 36238752]
24. Eaton C, Vallejo N, McDonald X, Wu J, Rodríguez R, Muthusamy N, et al. User engagement with mHealth interventions to promote treatment adherence and self-management in people with chronic health conditions: systematic review. J Med Internet Res 2024 Sep 24;26:e50508 [FREE Full text] [doi: 10.2196/50508] [Medline: 39316431]

Abbreviations

CDC: Centers for Disease Control and Prevention
DPP: diabetes prevention program
HbA1c: hemoglobin A1c
mHealth: mobile health
MVPA: moderate-to-vigorous physical activity
OR: odds ratio
RCT: randomized controlled trial
T2D: type 2 diabetes

Edited by L Buis; submitted 20.Feb.2026; peer-reviewed by O Pitsun, M Chakit; comments to author 11.Jun.2026; revised version received 14.Jul.2026; accepted 15.Jul.2026; published 02.Sep.2026.

Please cite as:
Lalani B, Siew G, Valdez Y, Shehadeh A, Zade D, Riekert K, Mathioudakis N, AI-DPP Study Group
User Engagement and Feature Preferences in an AI-Powered mHealth Intervention for Diabetes Prevention: Secondary Analysis of a Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e92981
URL: https://mhealth.jmir.org/2026/1/e92981
doi:10.2196/92981
PMID:

©Benjamin Lalani, Gabriela Siew, Yllka Valdez, Aliyah Shehadeh, Daniel Zade, Kristin Riekert, Nestoras Mathioudakis, AI-DPP Study Group. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 02.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Engagement With a Relaxation-Based Mobile Health Intervention for Perioperative Anxiety: Prospective Longitudinal Cohort Mixed Methods Study

Xinghui Yan[1], PhD; Afton L Hassett[2], PsyD; Jennifer F Waljee[3,4], MD, MPH, MSc; Mark W Newman[1,5], PhD; Sun Young Park[1,6], PhD; Rongqi Bei[1], MS; Noelle E Carlozzi[7], PhD

[1]School of Information, University of Michigan, Ann Arbor, MI, United States

[2]Chronic Pain & Fatigue Research Center, Department of Anesthesiology, University of Michigan, Ann Arbor, MI, United States

[3]Department of Surgery, Michigan Medicine, Ann Arbor, MI, United States

[4]Department of Surgery, Indiana University, Indianapolis, IN, United States

[5]Electrical Engineering and Computer Science, College of Engineering, University of Michigan, Ann Arbor, MI, United States

[6]Penny W Stamps School of Art and Design, University of Michigan, Ann Arbor, MI, United States

[7]Center for Clinical Outcomes Development and Application (CODA), Departments of Physical Medicine and Rehabilitation, and Surgery, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Noelle E Carlozzi, PhD
Center for Clinical Outcomes Development and Application (CODA)
Departments of Physical Medicine and Rehabilitation, and Surgery
University of Michigan
2800 Plymouth Rd
Ann Arbor, MI, 48109
United States
Phone: 1 734 763 8917
Email: carlozzi@med.umich.edu

Abstract

Background: Relaxation-based mobile health (mHealth) interventions hold strong potential to address perioperative anxiety and pain management in a scalable manner. However, there is limited research on how surgical patients engage with these interventions and what aspects may be most helpful.

Objective: This study had 3 goals: (1) explore how users interact with a relaxation-based mHealth intervention during the perioperative period; (2) understand how user performance and perceptions relate to perceived benefits; and (3) identify design considerations to enhance user engagement and behavior change outcomes in mHealth interventions.

Methods: We conducted a prospective longitudinal cohort, mixed methods study to evaluate user engagement with MiCarePath, a relaxation-based mHealth intervention. A total of 19 perioperative patients (mean age 41.3 years; 14/19, 74%, female) undergoing elective surgery and reporting sometimes-to-always anxiety were enrolled using 3 recruitment methods. Participants used the app from 10 days before surgery to 4 weeks after surgery. We collected quantitative data on video engagement metrics (eg, duration of watching, response rate to video prompts) and ecological momentary assessments of anxiety and pain. Concurrently, we conducted semistructured, data-prompted interviews at 3 time points to assess user perceptions. These data were integrated using descriptive statistics, repeated-measures ANOVA, and qualitative thematic analysis to examine the relationship between user performance and perceived benefits.

Results: Eleven participants reported developing or improving relaxation-based self-management strategies (high benefit), whereas 8 reported little benefit from the intervention. The high-benefit group showed significantly greater total watching time (difference 223.8 minutes, 95% CI 95.54-366.46 minutes, P<.001), higher adherence to video prompts (73.45% vs 34.60%, P<.001), and greater perceived helpfulness of the content (79.80% vs 25.50%, P<.001) compared with the low-benefit group. A significant group-by-time interaction effect was observed for engagement ($F_{1,17}$=17.02, P<.001). No difference was observed in delay in response to video prompts. However, within each group, participant perceptions were not fully reflected in performance. Some participants in the high-benefit group showed decreased video watching after surgery but still reported feeling engaged with the intervention. Interview data revealed that “intentional use” behavior—where users actively seek content to manage

XSL•FO
RenderX

symptoms independent of prompts—distinguished these participants. Those who developed intentional use were able to practice relaxation techniques without the app and resume app engagement following contextual disruptions.

Conclusions: This study explored user engagement with a relaxation-based mHealth intervention for perioperative care, advancing the literature by integrating user perceptions and performance for digital anxiety and pain management. Results indicate that perceived benefit may not always be reflected in user performance. Intentional use emerges as a promising indicator of effective mHealth engagement in perioperative care. Future designs should aim to foster intentional use through features such as reflection prompts and adaptive notifications. Future work should also develop computational models to detect intentional use and evaluate adaptive interventions in larger cohorts.

(JMIR Mhealth Uhealth 2026;14:e64278) doi:10.2196/64278

KEYWORDS

mHealth intervention; perioperative anxiety; pain management; user engagement; relaxation; mindfulness

Introduction

Background

Preoperative anxiety is common and closely intertwined with pain experiences both before and after surgery. It affects a significant proportion of individuals across different ages and health conditions, with estimates suggesting that up to 50% of patients experience heightened anxiety in anticipation of surgery [1]. Some studies report even higher rates, with up to 80% of patients across different surgical fields experiencing anxiety due to elevated stress and fear of surgery [2]. The experience of preoperative anxiety, if not addressed in a timely and appropriate manner, often extends throughout the entire perioperative period. Greater preoperative anxiety is associated with adverse outcomes, including increased patient-reported pain, prolonged opioid use, and postoperative misuse [3-6]. These interrelated challenges pose a substantial clinical burden, increasing both clinical workload and patient morbidity [7]. Despite the severity of the problem, managing both anxiety and pain throughout the perioperative period remains a critical challenge. Unlike general chronic disease populations, perioperative patients often face unique, acute stressors that complicate self-management. These include physical mobility restrictions, the cognitive load of managing new medication regimens (eg, opioids), and the emotional burden of anticipated surgical outcomes, all of which significantly complicate recovery for patients undergoing surgery [3,8].

In recent years, mobile health (mHealth) interventions have been shown to promote well-being through personalized care and behavioral support [9-11] and have the potential to improve perioperative outcomes by delivering interventions that support healthy behaviors [10,12]. With integrated educational and psychological strategies delivered via an mHealth platform, patients can receive support without the need for additional clinical appointments or programs [13,14], allowing more time to build resilience and receive behavioral support before and after surgery [15]. Some interventions focus on behavioral pain management and self-regulation, helping patients actively participate in their recovery and pain control [16,17]. A growing number of intervention apps have adopted a relaxation-based approach, leveraging evidence-based mind-body techniques to alleviate perioperative anxiety and pain [18]. These include guided mindfulness, breathing exercises, meditation [19], progressive muscle relaxation [20], and music-based interventions delivered via digital devices [21]. Emerging evidence suggests that these technologies show promise in supporting patients' emotional and physical well-being in a scalable manner during the perioperative period.

Research Gap

While mHealth interventions show promise for managing preoperative anxiety and pain, sustained user engagement remains a major challenge in this and most mHealth settings [22-26]. User engagement refers to interaction with and use of mHealth tools and can be measured by user perceptions and performance [22,23]. The effectiveness and desired outcomes of mHealth interventions are often hindered by poor uptake [27], early attrition [28], significant declines in adherence over 4-8 weeks of an intervention [29], intentional nonadherence to the intervention [30], and even abandonment [23] (eg, [29,31,32]). While broader mHealth studies have examined engagement, there is a scarcity of studies examining both user performance and perceptions in relaxation-based interventions for surgical patients [15,20,33]. Prior work often focused on general populations or chronic conditions, lacking the specific context of the acute perioperative phase and longitudinal data tracking usage from presurgery to recovery. Consequently, the intersection of objective usage metrics and subjective user experience remains underexplored in the perioperative patient population.

Given the acute stress of surgical anticipation, significant time constraints, and the physical burden of postoperative pain, these patients often navigate competing medical priorities that may limit their capacity to engage with digital tools [34,35]. Consequently, prior research on digital interventions for pain and anxiety management in surgical settings has reported low uptake and high attrition due to substantial time demands and emotional burden [16,34]. However, existing studies have relied heavily on quantitative metrics (eg, system usage data) to assess these challenges, often neglecting the affective and cognitive dimensions of the user experience [35]. Taken together, only a few studies have examined both user performance and perceptions through a longitudinal lens specifically within the acute perioperative phase, as most research focuses on general wellness or chronic conditions. This gap is particularly significant for perioperative patients, whose unique stressors and needs may shape their interaction with mHealth interventions for anxiety and pain management. Thus, a comprehensive understanding of the relationship between

XSL•FO
RenderX

objective usage and subjective experience remains underexplored in this population.

To address this gap, we investigated perioperative patient engagement with a relaxation-based mHealth intervention, MiCarePath. MiCarePath was designed to provide perioperative patients with low-cost, engaging education and relaxation-based exercises to help reduce both anxiety and pain. It includes 3 short educational videos about surgery in general (eg, what to expect, pain as a normal part of healing), the role of anxiety in the experience of pain, the use of opioids and other strategies to mitigate surgical pain, and the potential benefits of learning behavioral skills to reduce anxiety and pain. MiCarePath also includes self-directed, evidence-based relaxation exercises, such as guided imagery meditation videos and deep breathing training, which have been demonstrated to reduce perioperative anxiety and pain [3,4]. In addition, MiCarePath incorporates ecological momentary assessments of anxiety and pain, as well as questionnaires to measure health and behavioral outcomes (eg, pain, anxiety, self-management strategies) at different time points during the study.

Regarding user performance, we primarily focused on the length of interaction with different videos, the delay in responding to video prompts, and the response rate to video prompts. To examine user perceptions, we conducted semistructured interviews informed by the data-prompted interview approach [36] both before and following surgery. We used video engagement data, momentary anxiety and pain levels, and feedback related to the relaxation videos to probe participant decision-making and actions in response to video prompts. Findings from this study will inform the design of relaxation-based interventions for perioperative anxiety and pain management and enhance understanding of user perceptions and performance with such interventions.

Objective

We aimed to better understand perioperative patients' engagement with a relaxation-based mHealth intervention. We examined both user perceptions and performance during the pre- and postsurgical periods. This understanding is critical for designing more effective, patient-centered interventions that promote adherence and improve clinical outcomes. We conducted a mixed methods analysis to address the following research questions: (1) How do patients act on video prompts and engage with relaxation exercise videos during the perioperative period? (2) How can user perceptions and performance be aligned with an intervention to better understand user engagement? and (3) How might mHealth interventions be better designed to enhance user engagement and ultimately improve behavioral outcomes?

Methods

Intervention Design: The MiCarePath App

The MiCarePath app was designed to help surgical patients manage perioperative anxiety and pain through education and the teaching of relaxation skills. The app was built using MyDataHelps [37], a research platform that enables survey data collection and the delivery of mHealth interventions. The MiCarePath intervention included 5 components: (1) 3 surgery-related educational videos assigned to participants at the start of the intervention, covering expectations for surgery, safe opioid use, and the benefits of learning anxiety and pain self-management strategies; (2) relaxation exercise videos for day-to-day pain and anxiety management, including guided imagery meditation videos and a deep breathing pacer; (3) push notifications that provide reminders to watch relaxation exercise videos; (4) mindfulness-related behavioral suggestions; and (5) ecological momentary assessments of anxiety and pain (1 item each).

Prior literature shows that the timing of intervention delivery affects user receptivity and engagement with intervention content [12,25,26]. While we did not aim to evaluate the effectiveness of the intervention in a fully randomized controlled trial setting, we recognized the importance of randomizing the timing of video prompts to examine user experience. This approach allowed us to explore context-specific variations in user decision-making and actions in response to different video prompts. Accordingly, each day, the app delivered a video prompt at a random time within 1 of the following 4 time frames: (1) morning: 9:00 AM-noon; (2) early afternoon: noon-3:00 PM; (3) late afternoon: 3:00-6:00 PM; and (4) evening: 6:00-9:00 PM. Similar to the video prompts, participants completed momentary assessments of pain and anxiety 4 times per day, prompted within the same time frames. Considering the capabilities of MyDataHelps, we randomly generated 3 unique prompting schedules (ie, when to deliver a relaxation exercise video and ecological momentary assessments) and randomly assigned participants to one of these schedules, forming 3 groups. Participants within the same group received intervention prompts and self-report questions at the same time.

The primary intervention content of the MiCarePath app was the relaxation exercise videos, which were presented to participants once per day and were also the focus of this study. The videos, ranging from approximately 10 to 15 minutes in length, immerse users in various relaxation and meditation scenarios, such as a forest, seaside, or garden. These videos were developed and produced specifically for the MiCarePath app. Participants could switch to a different video or watch multiple videos, and they also had the option to proactively watch a video without being prompted. After completing relaxation exercises, the app prompted them to answer 3 or 4 questions designed to collect user feedback, including whether the video was watched, the perceived helpfulness of the video (helpful, neutral, or not helpful), activities before responding to the prompt, and reasons for nonadherence (if relevant). To monitor users' anxiety and pain levels, we implemented ecological momentary assessments using two 10-point single-item rating scales (ie, "From 1-10, what is your level of anxiety/pain right now?") [37].

To preliminarily assess the major outcomes of the MiCarePath intervention, the app collected data on participants' pain, anxiety, and any self-management strategies used over the course of the intervention. These measures were completed at study entry and for 2 weeks following surgery. The small sample size precluded analysis of the intervention's effectiveness;

XSL•FO
RenderX

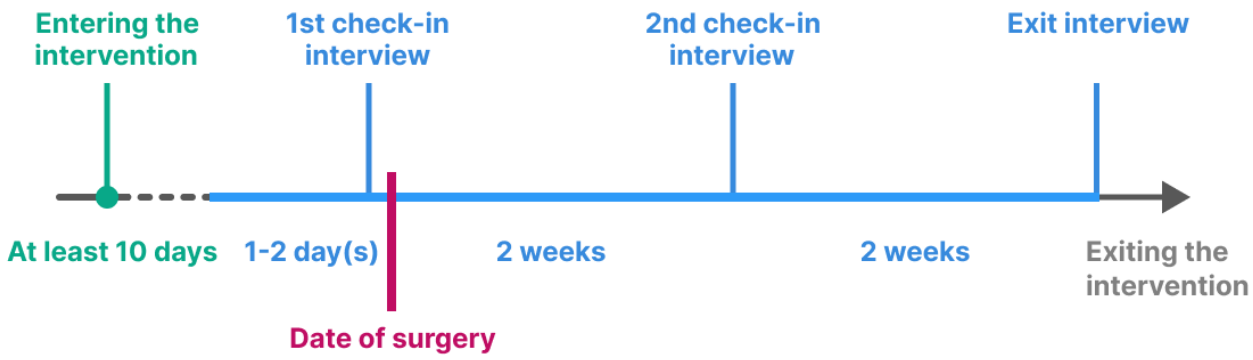
instead, we examined participants' user experience over time to explore intervention engagement.

Study Procedures

The study team first invited individuals to participate. The study used a fully remote design, with participant coordination conducted via email. Individuals who indicated interest were screened for eligibility criteria, consented, and then provided with instructions to download and install the MiCarePath project within the MyDataHelps app. The intervention period for each participant lasted approximately 40 days; participants began using the app at least 10 days before surgery and were encouraged to continue for 1 month following surgery. This timeline allowed participants to become familiar with the intervention, as well as to learn and practice relaxation techniques in preparation for surgery and during recovery. We conducted 3 qualitative interviews (each approximately 15-20 minutes): (1) before surgery (1-2 days prior); (2) 2 weeks after surgery; and (3) upon study completion (4 weeks following surgery, approximately 30 minutes). These interviews, conducted at multiple time points over the course of the intervention, allowed us to examine changes in participant perceptions and experiences. Figure 1 presents the study timeline. Additionally, participants were asked to complete questionnaires at these time points, including measures of anxiety and pain levels, self-management strategies, and medication use.

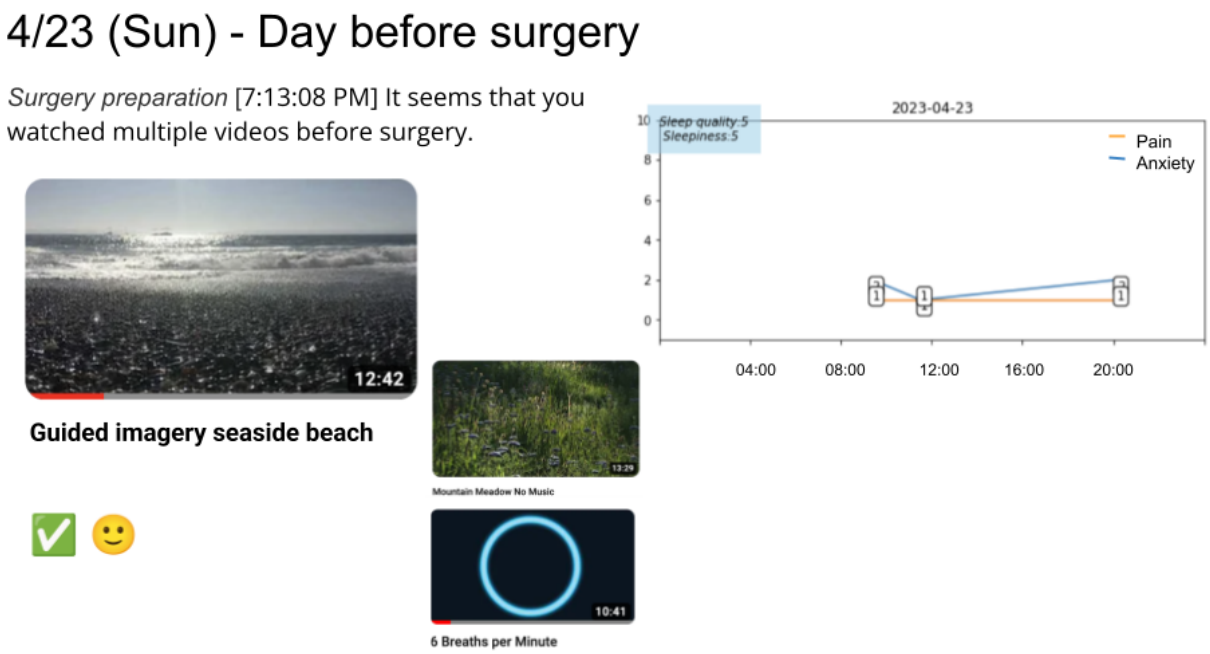
Interviews were informed by a data-prompted interview approach, which helps researchers gain rich insights into participants' in-the-field experiences by using collected data to guide conversations [38-40]. Two authors (XY and RB), both graduate researchers trained in human-computer interaction, developed the interview guide and conducted the interviews remotely via Zoom (Zoom Communications). During each interview, the research investigator first introduced the purpose of the interview and their background, and began with general questions about health and surgical status. We then presented data visualizations to participants displaying information about their actions in response to video prompts (ie, whether they chose to watch the video or not, or switched to a different video), video engagement (eg, completion of a video), momentary anxiety and pain, and activities before watching a video. Interview materials were presented using Google Slides (Google LLC; see Figure 2 for an example). Specifically, we probed user actions when (1) the participant chose not to watch a video; (2) the participant appeared to watch a video at an atypical time or in a unique context; and (3) the participant expressed an attitude that differed from their usual perceptions of the videos. When presenting the visualizations, the researcher described the data and asked participants to recall their experiences around the time they responded to a video prompt. Follow-up questions were used to explore possible reasons underlying participant decision-making and actions. We concluded the interviews by asking participants about other techniques or behavioral support they received during that period.

Figure 1. Study timeline. Participants attended 2 check-in interviews and 1 exit interview.



XSL•FO
RenderX

Figure 2. Example of the visualization shown during interviews. Text indicated when participants received a video prompt and their subsequent actions (eg, watching the video). A video screenshot was used to prompt recall of past experiences, and same-day pain and anxiety scores were displayed. Icons indicated video completion and perceived helpfulness.



Participant Recruitment

Participants were recruited using 3 sampling methods: study posts on a university-based research study site, snowball sampling of clinical providers' patient contacts, and review of surgical schedules for patients undergoing elective surgery. We used multiple recruitment methods to maximize enrollment of a diverse cohort of end users. Eligible participants were (1) at least 18 years of age; (2) fluent in English; (3) undergoing elective surgery at least 10 days after the consent date (to provide sufficient time to engage with the app); (4) willing to participate in surveys through 4 weeks after surgery; (5) able to access a device (smartphone or tablet, such as an iPad; Apple Inc) to use the MyDataHelps platform for completing anxiety and pain assessments and engaging with the MiCarePath intervention content; and (6) endorsing surgery-related anxiety, as indicated by a response of 3-5 to the following statement: "I feel anxious when thinking of my upcoming surgery (1=never, 2=rarely, 3=sometimes, 4=often, and 5=always)." We focused on patients who reported surgery-related anxiety, as these patients may be more interested in learning relaxation techniques and motivated to do so before surgery. The enrollment process commenced in October 2022 and concluded in July 2023. We aimed to enroll 15-25 participants to achieve thematic saturation and allow for subgroup comparisons [41,42]. The final sample size (N=19) is consistent with prior mixed methods and feasibility studies examining mHealth engagement in clinical contexts [29,43]. Given the exploratory aim of identifying engagement patterns and intentional use behaviors, this sample size was deemed sufficient to support thematic analysis and descriptive quantitative comparisons rather than hypothesis testing.

Ethical Considerations

Ethics approval was obtained in 2022 from the University of Michigan, Ann Arbor Institutional Review Board (approval number IRBMED HUM00217162). Individuals who contacted the study team were screened for eligibility, and those who were eligible provided e-consent before enrollment in study activities. Participants were instructed to install the MyDataHelps app during a brief onboarding session and were provided with a 1-page user manual for reference.

To ensure participant data confidentiality, all collected data were deidentified and replaced with unique participant codes. Interviews were audio-recorded, and no images of participants were obtained or reported in the manuscript. The document linking these codes to personally identifiable information was stored separately from the study data on a secure, firewalled university server, with access strictly restricted to the research team. Regarding data security, the MyDataHelps platform uses industry-standard encryption protocols. It is built on the HIPAA (Health Insurance Portability and Accountability Act)–compliant HIEBus platform, which securely handles protected health information in accordance with clinical standards.

Participants were compensated up to US $60 for completing surveys and interviews. They were not compensated for their engagement with video prompts to mitigate potential intervention effects related to incentives.

Data Analysis

Characterization of Perceived Benefit and Mixed Methods Integration

As the key aim of the MiCarePath intervention is to introduce relaxation techniques to perioperative patients and encourage

XSL•FO
RenderX

them to adopt or maintain relaxation as a self-management strategy, we used participant-reported use of strategies at week 2 after surgery—a multiple-choice question framed as follows: "Do you currently use strategies to decrease anxiety? (check all that apply)."—and frequency of use as criteria to assess perceived benefit from the intervention (ie, whether patients learned and developed a new self-management strategy as intended). We categorized participants into 2 groups based on their reported benefit from the intervention: a high-benefit group, comprising those who reported developing a new self-management strategy or maintaining relaxation techniques, and a low-benefit group, comprising those who did not report these outcomes. Our focus was on identifying commonalities within each perceived benefit group (high vs low) and differences between the 2 groups to identify phenomena indicative of users' perceived benefit. We first followed a sequential explanatory strategy [43,44], using the collected data to inform the check-in and exit interviews. During the data analysis phase, researchers applied a concurrent triangulation approach to conduct qualitative and quantitative analyses simultaneously and then integrated both sets of results for interpretation. When discrepancies were identified, we used participants' interview data to interpret changes in perceptions and performance.

Quantitative Data Analysis

We analyzed 4 key video engagement metrics: (1) average length of video engagement per week (in minutes), (2) response rate to video prompts (defined as whether a participant opened the notification), (3) response delay (measured in seconds from notification to video engagement), and (4) reported helpfulness of videos (categorized as helpful, neutral, or not helpful). These measures served as proxy indicators of user interest and depth of engagement with different relaxation videos [45]. User interaction data were collected using Google Analytics 4 and Google Tag Manager [46], which recorded time-stamped events such as video start and stop times, video content type, and engagement duration. Weekly engagement data (eg, postoperative week 1, week 2) were treated as a categorical within-participant variable to enable temporal comparisons. Missing data resulting from participant dropout were handled using listwise deletion, as analyses focused on within-participant trends rather than population-level inference. Given the exploratory nature of the study and small sample size, analyses were interpreted descriptively rather than as confirmatory tests. We began with descriptive statistical analyses to summarize trends in video engagement and perceived helpfulness. We then applied mixed-design analysis of variance (mixed ANOVA) to examine weekly video viewing time and response delay between the 2 groups. Before conducting ANOVA, we assessed the assumptions of normality and sphericity. When the Mauchly test indicated a violation of sphericity, we applied the Greenhouse-Geisser correction. Given the exploratory nature of the study, we did not apply corrections for multiple comparisons. All statistical analyses were conducted in Python (version 3.8.7; Python Foundation) using Jupyter Notebook.

Qualitative Data Analysis

We conducted a 2-stage thematic analysis of interview data, drawing on data-prompted interview analysis practices [38-40]. Interviews were audio-recorded and transcribed using auto-generated transcripts, which were subsequently proofread by the first author (XY) for accuracy. Researchers first reviewed participants' transcripts and annotated critical incidents, a widely used technique in qualitative research in which researchers focus on participants' recalled events (eg, how they responded to an intervention prompt) to reconstruct past behaviors and identify notable behavioral patterns [47]. Data visualizations presented during the interviews (eg, engagement trends) were used to contextualize participant responses and anchor incidents during analysis. In the first stage, 2 researchers (XY and RB) engaged in memo writing [48] to summarize each participant's key characteristics (including prior meditation experience and coping strategies for anxiety and pain management) and general experience with the intervention (including development of new self-management strategies and overall attitudes toward the intervention). Using a spreadsheet, researchers conducted a deductive analysis guided by the study's research questions, focusing on key experiential domains such as perceptions, reported behaviors, and changes before and after surgery. The spreadsheet format allowed for flexible categorization, efficient organization, and filtering of quotes by contextual variables (eg, timing relative to surgery, user group). Drawing on participants' objective engagement data, the researchers examined alignment between perceptions and performance, particularly across the 2 user groups. Check-in interview data were used to help explain changes in engagement over time. This triangulation process allowed us to contextualize subjective experience with objective intervention engagement patterns, specifically examining alignment between user perceptions and performance across the 2 user groups.

Researchers then applied in vivo coding [49,50] to uncover emergent themes grounded in participants' language through an iterative codebook development process. They collaboratively coded around 20% of the transcripts (4/19, 21%), establishing intercoder reliability through discussion and alignment on code definitions. Once a consistent agreement was reached, the remaining transcripts were independently coded using the finalized codebook. The broader research team, including senior members (NEC and MWN), held regular debriefing meetings to review and refine emerging themes. Data saturation was assessed through ongoing review of new transcripts and was considered achieved when no substantial new codes or themes emerged.

To ensure the quality and transparency of our mixed methods approach, this study followed the GRAMMS (Good Reporting of a Mixed Methods Study; Multimedia Appendix 1) framework [51] and the COREQ (Consolidated Criteria for Reporting Qualitative Research; Multimedia Appendix 2) guidelines [52].

Results

Overview

A total of 22 participants were enrolled in this study. Of these, 19 participants successfully completed the study; 1 participant

XSL•FO
RenderX

dropped out before surgery, and 2 dropped out after surgery. Sixteen participants completed all 3 interviews, while 3 were either unavailable due to physical therapy or remained unresponsive to the research team's phone calls or emails. Table 1 presents demographic characteristics, with a focus on prior experience and whether participants developed or maintained relaxation as a self-management strategy upon completing the intervention, which was used as the criterion to differentiate users' perceived benefit.

Based on participant responses to the self-management strategy survey items and the reported frequency of use in the baseline and exit surveys, participants were categorized into 2 groups: a high-benefit group (n=11) and a low-benefit group (n=8). In the high-benefit group, 7 participants incorporated relaxation as a new self-management strategy, and 4 reported increasing the frequency of its use. By contrast, participants in the low-benefit group reported that they did not develop a new strategy.

Table 1. Demographic characteristics of the participants (N=19) who completed this study.

Characteristics	Perioperative patients, n
Age group (years)	
20-34	6
35-49	4
50-64	5
65-80	4
Gender	
Male	4
Female	14
Chose not to disclose	1
Prior experience with relaxation exercises	
Yes	5
No	14
Self-management strategy	
Developed a new self-management strategy	7
Increasingly applied relaxation due to the intervention	4
Did not develop a new strategy	8

Comparing User Perceptions and Performance Between High- and Low-Benefit Groups

User Performance

All participants viewed the 3 educational videos upon entering the intervention. Table 2 presents the average total watching time, response rate to video prompts, delay in responding to a video prompt, and perceived helpfulness of the videos. Overall, results showed that, except for the average delay in responding to a video prompt, participants in the 2 groups exhibited significant differences in average total watching time (P <.001) and average response rate to video prompts (P<.001; also see Tables 2 and 3; Figures 3 and 4). The average weekly watching time and CIs of both high- and low-perceived benefit groups are presented in Table 3.

Before surgery, participants in the high-benefit group generally adhered to video prompts, with a mean nonadherence rate of only 11% (14/127). Four participants reported intentionally choosing their preferred relaxation exercises at least once, despite being recommended a different one. Three participants from the high-benefit group reported that they sometimes watched more than 1 video when prompted. In comparison, participants who perceived little or low benefit initially tried relaxation exercises before surgery, with an adherence rate of 55% (53/97) to video prompts, but most did not finish viewing the entire video. After surgery, 4 of 8 participants in the low-benefit group rarely attended to any video prompts.

A mixed ANOVA was conducted to examine changes in watching time across 5 perioperative time points (T1: before surgery; T2-T5: week 1 to week 4 after surgery) between the 2 groups. The data showed that, in the high-benefit group, the average weekly watching time started at 99.6 (95% CI 70.9-128.2) minutes at baseline (T1) and was maintained at 55.4 (95% CI 31.7-79.1) minutes by the end of the study (week 4, T5). Although there was a nominal reduction of 44.2 minutes, high interparticipant variability across the 5 time points resulted in a nonsignificant main effect of time (P=.10).

By contrast, participants in the low-benefit group exhibited a significant decline in watching time after surgery ($F_{4,28}$=19.76, P<.001). While they began with moderate engagement at baseline (mean 47.0 minutes; 95% CI 25.6-68.4 minutes)—approximately half the watching time of the high-benefit group, the average weekly watching time in the low-benefit group had dropped to only 4.3 (95% CI 0.0-12.6) minutes. At baseline (T1), there was already a significant difference in watching time between the high- and low-benefit

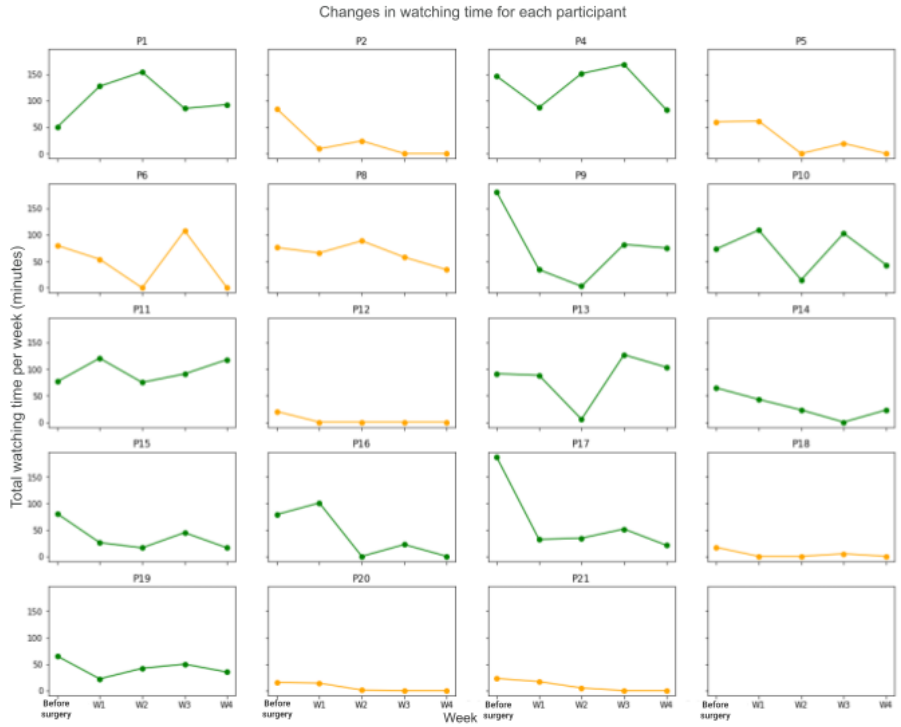
XSL•FO
RenderX

groups ($F_{1,17}$=21.62, P<.001), suggesting that participants in the low-benefit group were not initially engaged with the intervention. Moreover, preliminary analyses indicated an interaction effect between perioperative time and participants' perceived benefit level. The high-benefit group sustained 55.6% (55.4/99.6) of their baseline engagement by week 4 (maintaining a mean of 55.4 minutes), whereas the low-benefit group exhibited a significantly greater decline in watching time over time than the high-benefit group ($F_{1,17}$=17.02, P<.001). Conversely, no significant difference in video responsiveness was observed between the 2 groups (P=.81). The average delay in attending to video prompts was comparable between high-benefit (5.54 hours) and low-benefit (5.15 hours) participants.

Table 2. The average total watching time, average response rate to video prompts, average delay in attending to a video prompt, and the rate of perceiving videos as helpful from the ecological momentary assessment. Except for the delay in attending to a video prompt, participants in the high-benefit group acted significantly more engaged with the intervention as opposed to the low-benefit group.

Group variable	High-perceived benefit (H) (n=11), mean (SD); 95% CI	Low-perceived benefit (n=8), mean (SD); 95% CI	*P* value
Total watching time on average (minutes)	340.8 (152.3); 238.5-443.1	117 (115.5); 20.4-213.6	<.001
Average response rate to video prompts (%)	73.45 (15.36); 63.13-83.78	34.60 (18.26); 19.35-49.90	<.001
Average delay in attending a video prompt (hours)	5.54 (4.02); 2.84-8.24	5.15 (3.42); 2.29-8.01	.81
Rate of perceiving videos as helpful (%)	79.80 (9.25); 73.18-86.42	25.50 (14.76); 13.20-37.80	<.001

Figure 3. Weekly video engagement time for each participant relative to surgery (before surgery and weeks 1-4 after surgery). Participants in the high-benefit group are shown in green and those in the low-benefit group in yellow.



XSL•FO
RenderX

Figure 4. Participants' average daily video engagement (minutes watched) over the intervention period, from enrollment to up to 4 weeks after surgery. For each group, the figure shows individual trajectories, the group mean (bold lines), and within-group variability (shaded areas).

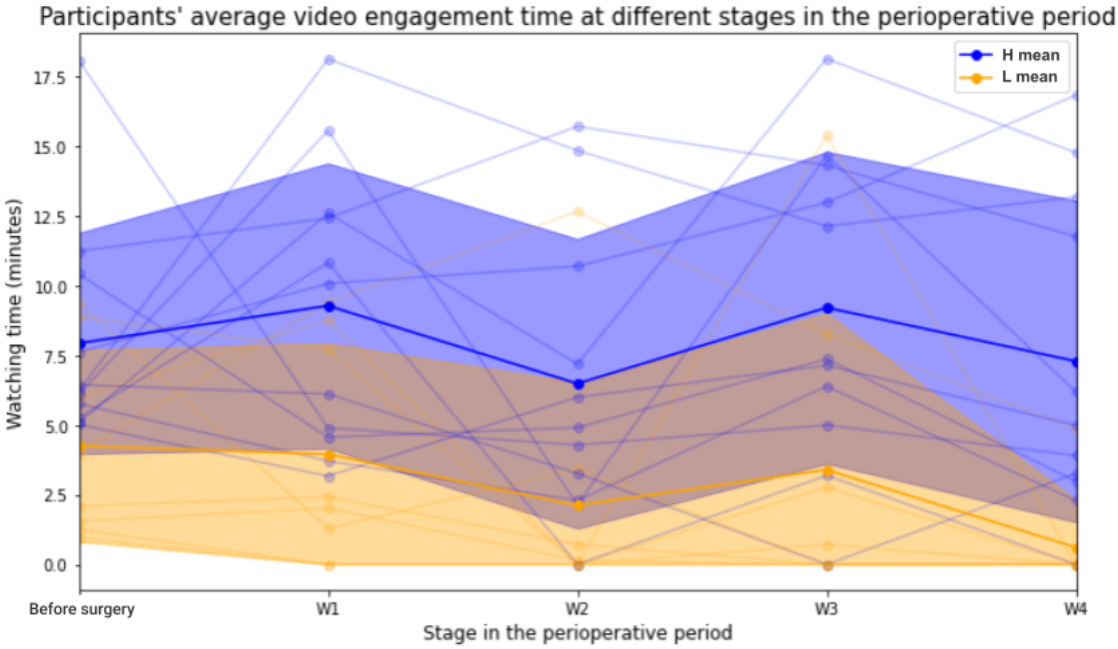


Table 3. The average weekly watching time and CIs of both high- and low-perceived benefit groups.

Time point	High-perceived benefit (n=11), mean; 95% CI	Low-perceived benefit (n=8), mean; 95% CI
T1 (before surgery)	99.6; 70.9-128.2	47.0; 25.6-68.4
T2 (week 1)	71.8; 47.8-95.9	25.8; 5.6-46.1
T3 (week 2)	48.7; 15.9-81.5	14.3; 0.0-35.9
T4 (week 3)	65.3; 39.5-91.2	23.2; 0.0-50.7
T5 (week 4)	55.4; 31.7-79.1	4.3; 0.0-12.6

User Perceptions

High-Benefit Group (n=11)

All participants in the high-benefit group reported incorporating relaxation practices as a new self-management strategy or reinforcing their previous experience with relaxation. Participants with prior experience continued to engage with the videos, viewing at least one video per week during the intervention. For example, P4 had over 20 years of meditation experience and told us, "The meditation videos in this app were good because they were new to me. So it gave me a fresher feeling." Among participants without prior experience (7/11), individuals gradually became familiar with relaxation techniques and appreciated having a tool to cope with anxiety and pain. Their positive experience with the intervention appeared to stem from 1 or more helpful relaxation exercises. For example, P9 tried garden imagery meditation for the first time and appreciated that the video offered less instruction and greater flexibility for imagination: "There's a little bit less instruction. The video does not talk as much, but it will give good ideas of going through the (garden) path and everything, and then just sitting down and relaxing. There's a little bit of freedom just to do whatever. And sometimes I like to imagine lots of flowers, or sometimes I just relax, and it's nice." Similarly, P11 gradually became more engaged with the relaxation videos, and increased interaction reinforced continued use of the intervention: "I did some meditation sometimes, but on and off. But with this [intervention], it's helping me to remind me to keep meditating and relaxing."

Low-Benefit Group (n=8)

As noted in user performance, participants who perceived low benefit were not engaged with the intervention from the outset. Most participants in this group reported that they felt it was less necessary to incorporate the intervention into their self-care, owing to satisfaction with existing self-management strategies. For example, P6 reported being satisfied with his current strategy: "As long as I am just sitting in this massage chair, it's not too painful, and I just watch television or read....This is the first time I really ever even thought about it [meditation] or seen anything about it." For several participants, although they completed the educational videos (including the rationale for learning relaxation techniques) upon entering the intervention, they remained unreceptive to relaxation exercises and intentionally ignored video prompts, as reported by P5: "I'm also very busy with work, and I don't have time [to use the

XSL•FO
RenderX

intervention app]. If I have time, I'll be relaxing and thinking about the surgery, and then I'll get the notification. And I have to think about it again. It's not that stressful, but sometimes I feel like I have one more thing to do now."

Discrepancy Between Participants' Performance and Perceptions

As previously noted, for most participants, those who derived the most benefit from the intervention exhibited significantly higher quantity and quality of video engagement with the MiCarePath app (ie, average total watching time, $P<.001$; average response rate to video prompts, $P<.001$; and rate of perceiving videos as helpful, $P<.001$). However, when aligning each participant's performance data with their perceptions of the app obtained from interviews, we observed variability within both groups. Briefly, a few participants in the high-benefit group showed lower video engagement at certain points during the intervention than some participants who reported little benefit. Additionally, although these participants appeared to perceive benefit from the MiCarePath app, their video engagement fluctuated throughout the intervention and occasionally showed signs of disengagement (eg, P14, P20, P21; see Figure 3). Conversely, 2 participants in the low-benefit group exhibited higher adherence rates before surgery, despite reporting that they did not perceive any benefit from using the intervention.

High-Perceived Benefit With Low Intervention Engagement After Surgery

Five participants in the high-benefit group reported reduced use of the intervention after surgery, while others reported maintaining a similar level of use as before; this was also confirmed by their performance data. However, these 5 participants still reported that the videos were helpful and that they derived benefit from the relaxation exercises. Notably, all 5 participants refrained from using the intervention for at least one week during the postoperative period (1 participant did not use the intervention in week 1 after surgery, and 4 in week 2 after surgery). From the interviews, one possible explanation was that these participants were prescribed other treatments or therapy after surgery and therefore, felt less able to attend to the video prompts. Nevertheless, they expressed satisfaction whenever they engaged with a relaxation video. In P17's case, he was proactively engaged with the videos before surgery to cope with abrupt, severe back pain, resulting in an average of 18 minutes of watching time per day: "That [the back pain] was a big thing. I was dealing with the pain and the anxiety that it might affect my heart surgery." He reported a positive experience with the app: "For most of the videos, I ended up closing my eyes. I pictured some places as depicted by the video." After surgery, P17 reported being busy and less attentive to video prompts (adherence rate dropped from 100% before surgery to 34% in weeks 1-2 after surgery). Although the frequency of relaxation practice decreased, he reported becoming more adapted to "imagining my own vision while watching videos." Consistent with P17's experience, other participants reported continued benefit from the app by gradually grasping relaxation techniques (eg, "learned a lot from the app"), although resumption of regular responsibilities after surgery (eg, work) limited their availability to engage in relaxation practice as before.

Low-Perceived Benefit Despite High Engagement With The Intervention

Adherence Without Perceived Benefit in the Low-Benefit Group

Regarding the low-benefit group's experience, 2 participants (P6 and P8) showed high adherence to video prompts and short response delays, often demonstrating better adherence than some participants in the high-benefit group. During the first interview before surgery, both P6 and P8 expressed doubt about the value of relaxation exercises. However, because both were retired, they had greater availability and could respond quickly to video prompts. For example, P6 told us: "If I saw there was a notification, I would just open it and play the video, sometimes doing other tasks." As noted, they typically played the recommended video in the background but were not fully engaged with the content. After surgery, because P6 and P8 did not perceive benefit, they became less compliant with video prompts and mostly chose not to engage with the videos. For the remaining participants who perceived low benefit, although they sometimes complied with video prompts, they did not report a positive experience or develop a routine of practicing relaxation. Interview data indicated that these participants either had difficulty finding time to watch the videos or watched them without full engagement. For example, P18 watched a video before sleep but did not perceive sufficient benefit: "On that day, I just happened to do the breathing exercises when I went to bed. It set the stage for sleep, but certainly not worth that much focus of time."

Overall, by aligning individual participants' performance data with their perceptions of the intervention, we found that user engagement with the intervention videos (eg, length of video engagement, delay in attending to a video prompt) may not fully reflect their perceived engagement or ability to derive benefit.

Intentional Use as a Key Phenomenon to Differentiate the 2 Groups

As presented above, participants' performance was largely aligned with perceived benefit from the intervention; however, in some cases, we observed discrepancies between performance and perceptions of the intervention. By triangulating user performance data, self-reported video helpfulness, and interview insights—particularly from participants who showed such discrepancies—we identified intentional use of the intervention content as a key factor in the interview data. We defined "intentional use" as using the intervention content purposefully for desired health-related effects, namely, managing anxiety and pain in our research context. This phenomenon can be considered a qualitative dimension of user engagement that differed between participants who perceived high benefit and those who perceived little benefit.

Participants Developed Intentional Use of the Intervention (n=7)

Participants in the high-benefit group reported that they would actively engage in relaxation exercises using the MiCarePath

XSL•FO
RenderX

app when feeling anxious or in pain, even without prompts. P16, for example, reported experiencing anxiety symptoms for years and being aware of her anxiety patterns. With the intervention, she became more conscious of using it to address her specific patterns of anxiety: “I'm usually already relaxed in the morning, but more at a higher anxiety state in the afternoon. So as long as I had time to watch the videos, I would open the app and practice mindfulness.”

Some participants with little prior experience with relaxation exercises appeared to shift in their engagement with the intervention, from more compliance-based to more intention-driven. For example, P10 reported that she initially responded to prompts to watch relaxation videos before surgery. After surgery, she became more intentional about practicing relaxation to cope with postoperative pain and avoid opioid use: “I was doing it as soon as I got the alerts, but then I learned that I could do them [do meditation exercises] later on. Sometimes it alerted me when I was in the middle of working. This was before my surgery. But now it's like I try to do it at times when I know that I'm stressed, to really get the full effect of it.” P15 gradually developed a habit of intentionally choosing a short relaxation video to address unexpected anxiety episodes, as expressed in the exit interview: “I think, it was going to be the difference between watching or not watching the video, [less about which video to watch]. So sometimes when I felt anxious, I would open the app and pick the shorter one because my goal was just to address my anxiety at that moment.” P19 also reported actively using the intervention; interestingly, she noted that the videos tended to be more helpful when she was more relaxed: “If I do it when I’m more relaxed, I actually pay more attention and the videos will actually help me better more. So I tried to watch them when I felt calm.”

Although some participants perceived little benefit from the intervention, they occasionally reported actively engaging with relaxation exercises without prompts, albeit without developing a clear intention or habit. For example, P8 recalled waiting at a doctor’s appointment and thinking of the intervention. He tried a relaxation exercise during that stressful moment and, for the first time, perceived the video as helpful. He told us, “The (doctor) appointment reminded me of the app so I pulled it up while waiting at the doctor’s place. Other than that, I haven’t experienced [anxiety] a lot, so I could not tell the difference [in my anxiety change before and after watching a video].” Compared with P8, although P5 exhibited low engagement and perceived little benefit from the intervention, she liked the 2-minute diaphragmatic breathing video and actively tried it on 1 occasion when she had time: “A week or 2 before surgery like I said, I was busy preparing like with work and with my house. But I thought about trying the 2-min video (6 breaths per minute), it’s kind of telling you how to do it....But it’s hard to start a habit [practicing relaxation].”

Participants Applied Relaxation Techniques Without Using the App (n=5)

A subset of participants (n=5) in the high-benefit group learned and remembered scripts of the relaxation exercises and were able to practice meditation without using the app. This may partly explain why these participants, despite subjectively feeling engaged with the intervention, showed lower engagement with relaxation videos (eg, shorter or less frequent use). For example, P10 watched her preferred meditation video multiple times and was able to remember parts of the instructions, which were “replayed” in her mind when she felt anxious. She told us, “I feel like I’ve been able to learn and remember different parts of the video [meditation instructions] even if I did not listen to it. I can still think myself through some of the stuff to relax and to get myself out of a stressful moment. So that’s pretty cool. It’s a skill that I’ve built up over time that I didn’t have at the beginning of all of this [this intervention].” Similarly, P17 preferred the short breathing pacer exercise and actively applied the learned skill over time: “[While practicing deep breathing], you just focus on your breathing, and you don’t think about ‘oh, my heart is racing’. So I could apply this when I felt overwhelmed.”

Moreover, there were situations in which participants intended to watch a video but were unable to interact with the app due to contextual constraints. As a result, some participants applied relaxation techniques without using the app. For example, P11 became capable of proactively using deep breathing techniques when feeling overwhelmed, even when her anxiety was unrelated to surgery, as expressed in the second check-in interview: “I am a contract-based teacher and every time I get a different teaching task. I can be anxious. During the day with my students in the classroom, I couldn’t play the video, so I just reminded myself to take deep breaths when it was overwhelming.”

Participants With Intentional Use of the Intervention Could Resume Engagement After Experiencing Disruptions

Surgery is considered a major life disruption that requires time, can prompt significant anxiety, and may require additional therapy for recovery. Along with surgery, other contextual disruptions (eg, going on vacation) can lead to temporary or sustained disengagement from mHealth interventions, as reported in prior work [42]. Based on responses from the check-in interviews, several participants in the high-benefit group appeared able to resume engagement after experiencing contextual disruptions to their daily routines. P13 learned to practice deep breathing through the app and was able to intentionally apply the technique when feeling stressed, noting, “I think it benefits me even right now, as I am talking about it. I am slowing down my breathing a little bit.” In week 3 after surgery, she did not use the app at all due to a family trip. Upon returning, she resumed use and achieved a level of video engagement similar to before (127 minutes over the 2 weeks after the trip, compared with 91 minutes before surgery; see Figure 3). By contrast, P2 (low perceived benefit) initially complied with video prompts before surgery but did not find it easy to integrate the meditation exercises into her routines after surgery (eg, “I was not quite engaged,” “Maybe it might help to watch videos right before sleep.”). Following recovery, she reported physical and mental fatigue that disrupted her routine and led to continued disengagement (eg, total watching time was 35 minutes over 4 weeks after surgery).

XSL•FO
RenderX

Discussion

Principal Findings

We report data from an exploratory field study of 19 perioperative patients on their experience with and use of a relaxation-based mHealth intervention, MiCarePath, designed to target perioperative anxiety and pain management. Findings indicate 2 distinct user groups: those who perceived high benefit from the intervention (n=11) and those who perceived little to no benefit (n=8). Participants in the high-benefit group remained more engaged with relaxation exercises before and after surgery, showing significant differences in interaction duration with videos (P<.001), average response rates (P<.001), and rate of perceiving videos as helpful (P<.001) compared with participants who reported limited or no benefit. However, for some participants, video engagement did not fully reflect perceived benefit, suggesting a discrepancy between user performance and perceptions of the intervention. Interview data indicated that this discrepancy may be attributed to 2 factors: (1) although participants' use of the intervention decreased after surgery, they were still able to apply relaxation techniques learned from the videos when needed, even without rewatching them; and (2) participants may have adhered to video prompts without feeling genuinely engaged in the activities, leading to gradual disengagement over time. To characterize this discrepancy, we propose intentional use of the intervention content as an important phenomenon that may distinguish participants who derive high benefit from those who do not. Intentional use can be considered a qualitative dimension of engagement, shifting the focus from the quantity of interaction (eg, amount, frequency, duration) to its quality. This concept complements quantitative measures by capturing whether user interaction with intervention content stems from motivation and intention, thereby distinguishing between passive adherence to intervention delivery and the strategic application of intervention strategies in daily life [53,54]. In this study, participants demonstrating intentional use were able to resume engagement with the intervention after experiencing contextual disruptions to their daily lives.

Theoretical Implications: Intentional Use and Behavioral Change Frameworks

Our interview data suggest that intentional use of an intervention involves 2 key factors: (1) belief in its ability to alleviate anxiety and pain, and (2) the capacity to schedule and integrate practice into daily routines. These findings align with the Health Action Process Approach (HAPA) [55,56] and the Theory of Planned Behavior (TPB) [57], both of which emphasize intention as central to behavior formation. More specifically, intentional use maps onto key constructs within these frameworks. Participants' belief in the effectiveness of relaxation techniques parallels outcome expectancies in HAPA and attitudes in TPB, while their ability to apply these techniques within daily routines reflects task self-efficacy in HAPA and perceived behavioral control in TPB. HAPA divides behavior change into 2 stages: the preintentional phase, in which intentions are formed, and the postintentional phase, in which intentions are enacted [55,56]. Within this framework, intention is shaped by task self-efficacy, outcome expectancies, and risk perception [55]. Our study suggests that participants who developed intentional use behaviors may be transitioning into the postintentional phase, actively engaging with the intervention rather than passively receiving content. The identified factor of outcome expectancies reflects participants' belief in achieving desired outcomes (eg, reduced anxiety or pain), which may explain why some repeatedly watched the same video, driven by anticipated benefits. TPB similarly links intention to attitudes, subjective norms, and perceived behavioral control [57]. Here, attitudes are shaped by the perceived effectiveness of the behavior [57], consistent with participants' belief in the value of relaxation techniques. Perceived behavioral control extends beyond self-efficacy to include external factors such as time, resources, and environment [58], aligning with participants' ability to apply relaxation techniques in real-life contexts. Thus, we suggest that intentional use arises when users believe in the effectiveness of the intervention and feel capable of incorporating it into their routines, consistent with facilitators identified in prior work as central to engagement with digital health interventions [54,59].

Furthermore, our findings show that even participants who perceived high benefit may temporarily "disengage" from the intervention, as they may be prescribed additional treatments after surgery. The perioperative context introduces acute stressors—such as surgical anticipation, postoperative pain, and competing clinical demands [60-62]—that can temporarily disrupt observable engagement without necessarily undermining user intention. Our findings suggest that intention may remain stable even when actions on intervention content through the app fluctuate, indicating a decoupling between system-recorded engagement and the underlying behavior change process. This aligns with Yardley's model, which outlines 4 phases of user engagement with digital behavior change interventions: initial engagement with the intervention only (phase 1), intervention-mediated behavior change (phase 2), sustained behavior change independent of the platform (phase 3), and potential reengagement if needed (phase 4). This model suggests that users may disengage from an intervention platform (eg, an app) while remaining engaged in the behavior change process [63]. Similarly, Perski et al [64] defined engagement as both the extent of usage and a subjective experience, specifically a user's attention, interest, and affect related to the intervention.

Taken together, user engagement with an intervention—particularly effective engagement—reflects the intervention-mediated behavior change process [63,65]. To better connect the design of mHealth interventions with behavior change theories, designers need to consider the desired type and level of behavior change mediated by the intervention. This perspective aligns with prior agenda-setting work in behavior change and digital health intervention design [66,67]. Such consideration is critical, as the user's behavior change process may not be entirely driven by the intervention or easily detectable [68]. Furthermore, our research offers a perspective that extends behavioral frameworks by emphasizing contextual disruption as a moderating factor in intention enactment, rather than as a failure of engagement. In this respect, our findings contribute to a more nuanced understanding of user engagement

XSL•FO
RenderX

that is sensitive to an individual's short-term self-care trajectory. Given the variability in users' motivation, needs, prior skills, and changing contexts, the target level of engagement can be tailored to individual status (eg, existing skills and strategies) and evolving contexts (eg, before and after surgery), rather than defined by uniform expectations of intervention use.

Design Implications: Supporting Intentional Use in Perioperative mHealth Interventions

Overview

In this section, we discuss design implications centered on the concept of intentional use. These implications include (1) accounting for both user performance and perceptions in understanding perioperative patients' engagement; (2) considering intentional use in the evaluation of mHealth interventions; and (3) designing mHealth interventions to support and shape intentional use.

Accounting for Both User Performance and Perceptions in Understanding Perioperative Patients' Engagement With an Intervention

Assessing user engagement with mHealth interventions requires both objective performance measures and subjective user perceptions, as relying solely on objective metrics may misrepresent engagement by prioritizing frequency over quality [69]. Our study with perioperative patients revealed a potential discrepancy between performance with relaxation videos and perceived benefits. While several low-benefit users consistently followed video prompts with minimal delay, they did not report meaningful benefits. By contrast, intentional use—where users actively incorporate relaxation as a self-management strategy—emerged as a potentially valuable indicator of "effective engagement" and behavior change [63]. This suggests that engagement is not solely driven by system reminders (eg, push notifications), but also by users' growing recognition of when and how to use the intervention to achieve desired outcomes. As prior research highlights varying definitions and measures of engagement, the concept of "effective engagement"—interactions that lead to meaningful behavioral outcomes [70]—offers a more appropriate lens for evaluation. Given the importance of intentional use in shaping user engagement and behavior change, future mHealth interventions should integrate this factor into system design to enable better personalization and optimization in alignment with behavioral theories.

Considering Intentional Use to Evaluate mHealth Interventions

Findings from our study show that participants who benefited most from the app were able to intentionally engage with the intervention content at preferred and needed moments. However, in practice, it remains unclear how to leverage this insight to better understand user experience with mHealth interventions and to optimize intervention algorithms for adaptive interventions. In our study, check-in interviews—through which we learned about participants' gradual development of intentional use behaviors over time—proved useful. This suggests that periodic assessment of intentional use behavior, as well as progress in developing such behavior, may be valuable for inferring user engagement and determining whether desirable behavioral outcomes (eg, learning a new self-management strategy) are likely to occur.

Indeed, measuring user intention is challenging due to its subjective nature, highlighting the need to develop appropriate measures in future research. Intervention designers may consider using self-report methods to periodically assess intention in using preferred intervention content (eg, new skills, health behaviors) and associated contextual conditions (eg, before sleep, before a doctor's appointment). One caveat is the potential user burden associated with manual data collection, which may negatively affect overall adherence to the intervention [71]. Furthermore, it remains unclear whether users can accurately express their perceptions and reflect on their intent, particularly when they are in an intermediate stage of intentional use (ie, not fully aware of their intent but gradually developing the inclination to use the intervention) [72]. Future work should continue to explore factors that may influence or indicate intentional use. For example, prior research suggests that engaging with intervention content at a consistent time may indicate higher engagement [73]. In this direction, it would be valuable to explore how to computationally characterize the trajectory of developing intentional use, enabling digital tools to adapt intervention strategies accordingly.

Designing to Shape Intentional Use With mHealth Interventions

Our study highlights the potential to shape intentional use of an intervention to achieve behavioral goals. At present, it remains unclear how different system features, such as user onboarding or a feedback dashboard, may support the development of intentional use. Our interview findings suggest that intentional use may involve a belief in the intervention's ability to produce desired benefits. Accordingly, intervention systems could encourage users to reflect on the benefits they experience [74]. For example, after a relaxation exercise, the app could prompt a simple reflection question: "How do you feel after watching this?" This may help users recognize positive effects over time and reinforce belief in the intervention's benefits. As intentional use develops—for instance, as users gain a better understanding of when to proactively engage with intervention content—systems could introduce features such as implementation intentions, prompting users to specify actions (eg, engaging with the intervention) and associated conditions (eg, needed moments) [75]. Additionally, interventions could prompt users to reflect on and actively plan future activities based on their detected intentions (eg, [76,77]). For example, a lightweight planning feature might ask, "Would you like to schedule a time for your next relaxation session?" and synchronize with users' calendars to provide timely nudges [76]. Overall, these design strategies align with behavioral theory by reinforcing self-efficacy and autonomy, rather than promoting compliance with fixed intervention schedules [78].

XSL•FO
RenderX

Closing Remarks: Practical and Theoretical Implications

Theoretical Implications

Intentional, self-initiated use of the intervention—without reliance on system notifications—emerges as a key behavior distinguishing high-benefit users. This finding underscores the theoretical importance of autonomous engagement in mHealth interventions.

Practical Implications for the Design and Assessment of mHealth Interventions

Given the potential discrepancies between user performance and perceptions, it is important to consider both aspects when assessing user engagement. When examining user perceptions, priority should be given to self-reported intention, such as perceived behavioral control [79]. Future research should explore the development of assessment methods that effectively capture users' intentions. When evaluating intervention effectiveness, systems should account for user behaviors that occur outside the system but are reinforced by in-system engagement, to support a more holistic view of users' progress in the behavior change process. With this perspective, intervention designers should consider incorporating reflective features that promote intentional use or implementation intentions—for example, by prompting users to specify actions and contextual conditions. Looking ahead, future research could investigate gamified meditation exercises to strengthen user intention, alongside visual progress trackers (eg, integrating in-system and off-system activities) to reinforce habit formation.

Limitations

Although this exploratory study revealed important insights into perioperative patient engagement with a relaxation-based intervention, several limitations should be noted.

First, the study was conducted with a small sample of perioperative patients. As participants were expected to use the intervention around the surgical period, recruitment was subject to strict timing constraints. Despite employing 3 recruitment approaches, we encountered difficulties enrolling participants, primarily due to limited availability and high levels of preoperative stress. These recruitment constraints may have introduced selection bias, as individuals who chose to participate may have been more motivated, more comfortable with technology, or more open to relaxation-based approaches than the broader perioperative population. Nevertheless, we were able to obtain insights from participants with varying levels of intervention engagement, which helped inform refinement of the intervention design and evaluation of its effectiveness.

Second, the study sample was relatively homogeneous, with participants predominantly female and all fluent in English. The predominance of female participants may reflect prior findings that meditation apps tend to be more favored and well-received by females than males [80]. This homogeneity may limit the generalizability of the findings to more diverse populations, including individuals from different cultural backgrounds, non–English-speaking patients, and those with varying levels of health literacy and access to technology. Future work should intentionally recruit more diverse samples to examine how cultural, linguistic, and sociodemographic factors shape engagement with relaxation-based mHealth interventions.

Third, the current intervention does not offer adaptive or personalized tailoring. Findings from this study highlight the importance of facilitating intentional use of interventions. Building on this direction, there is an opportunity to implement a just-in-time adaptive intervention system for perioperative patients and to explore approaches that enhance user proactivity and intentional use over time. Fourth, we conducted multiple check-in interviews during the intervention, which may have been perceived as intrusive and could have influenced user perceptions and experiences. To mitigate this, unless participants asked about a feature in the app, we did not prompt them to engage with the intervention or evaluate their performance. Nevertheless, some participants may have inferred the study's desired outcomes and, as a result, expressed more favorable views of the intervention during interviews. Accordingly, interview questions were designed to focus on eliciting detailed accounts of participants' experiences in context. Participants may also have had difficulty recalling details of their engagement with the intervention, leading to potential recall bias. To minimize this, we used a data-prompted interview approach, presenting participants with their anxiety and pain scores, along with screenshots of relaxation videos, to support recall. Finally, several participants dropped out during the postoperative period due to unavailability or recovery, and they may have lacked interest in the intervention; however, we were unable to collect their feedback.

Conclusions

This exploratory study examined perioperative patients' engagement with a relaxation-based mHealth intervention (ie, the MiCarePath app) by jointly considering user performance, subjective perceptions, and retrospective perceptions over 3 weeks. Our findings highlight the importance of understanding how patients interpret and apply intervention content within the constraints of the perioperative context. By integrating user engagement data with interview data, this work demonstrates that observable interaction with an intervention may not fully capture perceived value or its role in supporting patients' self-management during the perioperative period.

A central contribution of this study is the identification of intentional use as a qualitative dimension of engagement that helps explain discrepancies between performance and perceived benefit. Participants who reported greater benefit were not necessarily those with consistently high levels of interaction with intervention content, but rather those who developed the ability to intentionally use relaxation strategies to manage anxiety and pain—including outside the app and after disruptions such as surgery or competing care demands. This finding extends the existing literature on user engagement by emphasizing the quality and purpose of interaction, particularly in short-term, high-burden clinical settings where sustained adherence may be difficult to achieve.

These insights have important implications for the design and evaluation of relaxation-based mHealth interventions in perioperative care. Interventions that support the development

XSL•FO
RenderX

of intentional use may better promote nonpharmacological anxiety and pain management, complement clinical care, and reduce reliance on intensive in-person support or medication-based approaches. Future research should explore how intentional use can be fostered and detected over time, as well as how adaptive intervention strategies can be integrated into perioperative care pathways to support scalable, patient-centered outcomes.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The GRAMMS (Good Reporting of a Mixed Methods Study) checklist.
[PDF File (Adobe PDF File), 49 KB - mhealth_v14i1e64278_app1.pdf]

Multimedia Appendix 2
The COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[PDF File (Adobe PDF File), 424 KB - mhealth_v14i1e64278_app2.pdf]

References

1. Friedrich S, Reis S, Meybohm P, Kranke P. Preoperative anxiety. Curr Opin Anaesthesiol 2022 Dec 01;35(6):674-678. [doi: 10.1097/ACO.0000000000001186] [Medline: 36131642]
2. Rajjoub R, Sammak SE, Rajjo T, Rajjoub NS, Hasan B, Saadi S, et al. Meditation for perioperative pain and anxiety: a systematic review. Brain Behav 2024 Jul 29;14(7):e3640 [FREE Full text] [doi: 10.1002/brb3.3640] [Medline: 39073307]
3. Gurram P, Narayanan V, Chandran S, Ramakrishnan K, Subramanian A, Kalakumari AP. Effect of heartfulness meditation on anxiety and perceived pain in patients undergoing impacted third molar surgery. J Oral Maxillofac Surg 2021 Oct;79(10):2060.e1-2060.e7. [doi: 10.1016/j.joms.2021.04.027] [Medline: 34097862]
4. Gorsky K, Black ND, Niazi A, Saripella A, Englesakis M, Leroux T, et al. Psychological interventions to reduce postoperative pain and opioid consumption: a narrative review of literature. Reg Anesth Pain Med 2021 Oct;46(10):893-903. [doi: 10.1136/rapm-2020-102434] [Medline: 34035150]
5. Hanley AW, Gililland J, Garland EL. To be mindful of the breath or pain: comparing two brief preoperative mindfulness techniques for total joint arthroplasty patients. J Consult Clin Psychol 2021 Jul;89(7):590-600 [FREE Full text] [doi: 10.1037/ccp0000657] [Medline: 34165999]
6. Davin SA, Savage J, Thompson NR, Schuster A, Darnall BD. Transforming standard of care for spine surgery: integration of an online single-session behavioral pain management class for perioperative optimization. Front Pain Res (Lausanne) 2022;3:856252 [FREE Full text] [doi: 10.3389/fpain.2022.856252] [Medline: 35599968]
7. Doan LV, Blitz J. Preoperative assessment and management of patients with pain and anxiety disorders. Curr Anesthesiol Rep 2020 Jan;10(1):28-34 [FREE Full text] [doi: 10.1007/s40140-020-00367-9] [Medline: 32435161]
8. Grutz K, Poch N. Meditation for preoperative anxiety and postoperative pain in bariatric surgery. J Perianesth Nurs 2021 Oct;36(5):586-590. [doi: 10.1016/j.jopan.2021.04.003] [Medline: 34688457]
9. Nahum-Shani I, Smith S, Spring B, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
10. Lee JA, Choi M, Lee SA, Jiang N. Effective behavioral intervention strategies using mobile health applications for chronic disease management: a systematic review. BMC Med Inform Decis Mak 2018 Mar 20;18(1):12 [FREE Full text] [doi: 10.1186/s12911-018-0591-0] [Medline: 29458358]
11. Pan S, Rong LQ. Mobile applications in clinical and perioperative care for anesthesia: narrative review. J Med Internet Res 2021 Sep 17;23(9):e25115 [FREE Full text] [doi: 10.2196/25115] [Medline: 34533468]
12. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019 Apr 03;16(1):31 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
13. Kondylakis H, Chicchi Giglioli IA, Katehakis DG, Aldemir H, Zikas P, Papagiannakis G, et al. A digital health intervention for stress and anxiety relief in perioperative care: protocol for a feasibility randomized controlled trial. JMIR Res Protoc 2022 Nov 29;11(11):e38536 [FREE Full text] [doi: 10.2196/38536] [Medline: 36445734]
14. Matthews P, Topham P, Caleb-Solly P. Interaction and engagement with an anxiety management app: analysis using large-scale behavioral data. JMIR Ment Health 2018 Sep 14;5(4):e58 [FREE Full text] [doi: 10.2196/mental.9235] [Medline: 30287415]
15. Kerimaa H, Hakala M, Haapea M, Vähänikkilä H, Serlo W, He HG, et al. Effectiveness of a mobile app intervention for preparing preschool children and parents for day surgery: randomized controlled trial. J Med Internet Res 2023 Sep 29;25:e46989 [FREE Full text] [doi: 10.2196/46989] [Medline: 37773624]

XSL•FO
RenderX

16. Darnall B, Ziadni M, Krishnamurthy P, Flood P, Heathcote LC, Mackey IG, et al. "My surgical success": effect of a digital behavioral pain medicine intervention on time to opioid cessation after breast cancer surgery-a pilot randomized controlled clinical trial. Pain Med 2019 Nov 01;20(11):2228-2237 [FREE Full text] [doi: 10.1093/pm/pnz094] [Medline: 31087093]
17. Pronk Y, Peters MCWM, Sheombar A, Brinkman JM. Effectiveness of a mobile eHealth app in guiding patients in pain control and opiate use after total knee replacement: randomized controlled trial. JMIR Mhealth Uhealth 2020 Mar 13;8(3):e16415 [FREE Full text] [doi: 10.2196/16415] [Medline: 32167483]
18. Rand A. The role of mind-body interventions in reducing perioperative anxiety and pain. Journal of Anesthesiology and Pain Research 2023 Apr 05;6(2):1-2 [FREE Full text] [doi: 10.37421/2684-5997.2023.6.177]
19. Kondylakis H, Giglioli IAC, Katehakis D, Aldemir H, Zikas P, Papagiannakis G, et al. Stress reduction in perioperative care: feasibility randomized controlled trial. J Med Internet Res 2025 Jan 07;27(1):e54049 [FREE Full text] [doi: 10.2196/54049] [Medline: 39773866]
20. Loh EW, Shih HF, Lin CK, Huang TW. Effect of progressive muscle relaxation on postoperative pain, fatigue, and vital signs in patients with head and neck cancers: a randomized controlled trial. Patient Educ Couns 2022 Jul;105(7):2151-2157. [doi: 10.1016/j.pec.2021.10.034] [Medline: 34785078]
21. Park S, Lee S, Howard S, Yi J. Technology-based music interventions to reduce anxiety and pain among patients undergoing surgery or procedures: systematic review of the literature. JMIR Mhealth Uhealth 2024 Jul 08;12:e48802 [FREE Full text] [doi: 10.2196/48802] [Medline: 38976863]
22. Cole-Lewis H, Ezeanochie N, Turgiss J. Understanding health behavior technology engagement: pathway to measuring digital behavior change interventions. JMIR Form Res 2019 Oct 10;3(4):e14052 [FREE Full text] [doi: 10.2196/14052] [Medline: 31603427]
23. Fowers R, Berardi V, Huberty J, Stecher C. Using mobile meditation app data to predict future app engagement: an observational study. J Am Med Inform Assoc 2022 Nov 14;29(12):2057-2065 [FREE Full text] [doi: 10.1093/jamia/ocac169] [Medline: 36164826]
24. O'Brien H, Toms EG. What is user engagement? A conceptual framework for defining user engagement with technology. J Am Soc Inf Sci 2008 Feb 28;59(6):938-955. [doi: 10.1002/asi.20801]
25. Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H. Design features for improving mobile health intervention user engagement: systematic review and thematic analysis. J Med Internet Res 2020 Dec 09;22(12):e21687 [FREE Full text] [doi: 10.2196/21687] [Medline: 33295292]
26. Vaghefi I, Tulu B. The continued use of mobile health apps: insights from a longitudinal study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e12983 [FREE Full text] [doi: 10.2196/12983] [Medline: 31469081]
27. Jardine J, Nadal C, Robinson S, Enrique A, Hanratty M, Doherty G. Between rhetoric and reality: real-world barriers to uptake and early engagement in digital mental health interventions. ACM Trans Comput-Hum Interact 2024;31(2):1-59. [doi: 10.1145/3635472]
28. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
29. Mikolasek M, Witt CM, Barth J. Effects and implementation of a mindfulness and relaxation app for patients with cancer: mixed methods feasibility study. JMIR Cancer 2021 Jan 13;7(1):e16785 [FREE Full text] [doi: 10.2196/16785] [Medline: 33439132]
30. Badawy SM, Shah R, Beg U, Heneghan MB. Habit strength, medication adherence, and habit-based mobile health interventions across chronic medical conditions: systematic review. J Med Internet Res 2020 Apr 28;22(4):e17883 [FREE Full text] [doi: 10.2196/17883] [Medline: 32343250]
31. Beatty L, Kemp E, Binnion C, Turner J, Milne D, Butow P, et al. Uptake and adherence to an online intervention for cancer-related distress: older age is not a barrier to adherence but may be a barrier to uptake. Support Care Cancer 2017 Jun;25(6):1905-1914. [doi: 10.1007/s00520-017-3591-1] [Medline: 28155018]
32. Kubo A, Altschuler A, Kurtovich E, Hendlish S, Laurent CA, Kolevska T, et al. A pilot mobile-based mindfulness intervention for cancer patients and their informal caregivers. Mindfulness (N Y) 2018 Dec;9(6):1885-1894 [FREE Full text] [doi: 10.1007/s12671-018-0931-2] [Medline: 30740187]
33. Wantanakorn P, Harintajinda S, Chuthapisith J, Anurathapan U, Rattanatamrong P. A new mobile application to reduce anxiety in pediatric patients before bone marrow aspiration procedures. Hosp Pediatr 2018 Oct;8(10):643-650. [doi: 10.1542/hpeds.2018-0073] [Medline: 30213798]
34. Nilsson O, Stenman M, Letterstål A, Hultgren R. A randomized clinical trial of an eHealth intervention on anxiety in patients undergoing abdominal aortic aneurysm surgery. Br J Surg 2021 Aug 19;108(8):917-924 [FREE Full text] [doi: 10.1093/bjs/znab151] [Medline: 34021309]
35. Madujibeya I, Lennie T, Aroh A, Chung ML, Moser D. Measures of engagement with mHealth interventions in patients with heart failure: scoping review. JMIR Mhealth Uhealth 2022 Aug 22;10(8):e35657 [FREE Full text] [doi: 10.2196/35657] [Medline: 35994345]
36. Taki S, Lymer S, Russell CG, Campbell K, Laws R, Ong KL, et al. Assessing user engagement of an mHealth intervention: development and implementation of the growing healthy app engagement index. JMIR Mhealth Uhealth 2017 Jun 29;5(6):e89 [FREE Full text] [doi: 10.2196/mhealth.7236] [Medline: 28663164]

XSL•FO
RenderX

37. CareEvolution LLC. URL: https://careevolution.com/mydatahelps/ [accessed 2026-04-26]
38. Kwasnicka D, Dombrowski SU, White M, Sniehotta FF. 'It's not a diet, it's a lifestyle': a longitudinal, data-prompted interview study of weight loss maintenance. Psychol Health 2019 Aug;34(8):963-982. [doi: 10.1080/08870446.2019.1579913] [Medline: 30905184]
39. Kwasnicka D, Potthoff S, Hagger MS, Vandelanotte C, Rebar A, Short CE, et al. Challenging assumptions underlying physical activity promotion for health care professionals in Australia: a data-prompted interview study. Health Promot J Austr 2024 Apr;35(2):542-550. [doi: 10.1002/hpja.784] [Medline: 37537885]
40. Cole A, O'Brien H. Using data-prompted interviews in interactive information retrieval research: a reflection on the study of self-efficacy when learning using search. 2023 Mar Presented at: CHIIR '23: ACM SIGIR Conference on Human Information Interaction and Retrieval; March 19-23, 2023; Austin, TX. [doi: 10.1145/3576840.3578285]
41. Vasileiou K, Barnett J, Thorpe S, Young T. Characterising and justifying sample size sufficiency in interview-based studies: systematic analysis of qualitative health research over a 15-year period. BMC Med Res Methodol 2018 Nov 21;18(1):148 [FREE Full text] [doi: 10.1186/s12874-018-0594-7] [Medline: 30463515]
42. Yan X, Newman MW, Park SY, Sander A, Choi SW, Miner J, et al. Identifying design opportunities for adaptive mHealth interventions that target general well-being: interview study with informal care partners. JMIR Form Res 2023 Oct 24;7:e47813 [FREE Full text] [doi: 10.2196/47813] [Medline: 37874621]
43. Naughton F, Hopewell S, Lathia N, Schalbroeck R, Brown C, Mascolo C, et al. A context-sensing mobile phone app (Q Sense) for smoking cessation: a mixed-methods study. JMIR Mhealth Uhealth 2016 Sep 16;4(3):e106 [FREE Full text] [doi: 10.2196/mhealth.5787] [Medline: 27637405]
44. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. Health Serv Res 2013 Dec;48(6 Pt 2):2134-2156 [FREE Full text] [doi: 10.1111/1475-6773.12117] [Medline: 24279835]
45. Couper MP, Alexander GL, Zhang N, Little RJA, Maddy N, Nowak MA, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010 Nov 18;12(4):e52 [FREE Full text] [doi: 10.2196/jmir.1430] [Medline: 21087922]
46. Tagging for Google Analytics. Google LLC/Alphabet Inc. URL: https://developers.google.com/analytics/devguides/collection/ga4/tag-options [accessed 2026-04-26]
47. The critical incident technique. Psychol Bull 1954 Jul;51(4):327-358. [doi: 10.1037/h0061470] [Medline: 13177800]
48. Birks M, Chapman Y, Francis K. Memoing in qualitative research: probing data and processes. J Res Nurs 2008;13(1):68-75. [doi: 10.1177/1744987107081254]
49. Manning J. In vivo coding. In: The International Encyclopedia of Communication Research Methods. Hoboken, NJ: Wiley; Nov 07, 2017:1-2.
50. Raj S, Toporski K, Garrity A, Lee J, Newman M. "My blood sugar is higher on the weekends": finding a role for context and context-awareness in the design of health self-management technology. 2019 May 02 Presented at: CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland, UK. [doi: 10.1145/3290605.3300349]
51. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy 2008 Apr;13(2):92-98. [doi: 10.1258/jhsrp.2007.007074] [Medline: 18416914]
52. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
53. Kelders SM, van Zyl LE, Ludden GDS. The concept and components of engagement in different domains applied to eHealth: a systematic scoping review. Front Psychol 2020;11:926 [FREE Full text] [doi: 10.3389/fpsyg.2020.00926] [Medline: 32536888]
54. Kelders SM, Kip H, Beerlage-de Jong N, Köhle N. What does it mean to be engaged with digital health interventions? A qualitative study into the experiences of engaged users and the views of professionals. Digit Health 2024 Oct;10:20552076241283530 [FREE Full text] [doi: 10.1177/20552076241283530] [Medline: 39376944]
55. Schwarzer R. Health action process approach (HAPA) as a theoretical framework to understand behavior change. Actual Psicol 2016;30(121):119. [doi: 10.15517/ap.v30i121.23458]
56. Sutton S. How does the health action process approach (HAPA) bridge the intention–behavior gap? An examination of the model's causal structure. Applied Psychology 2008 Jan 30;57(1):66-74. [doi: 10.1111/j.1464-0597.2007.00326.x]
57. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
58. Ajzen I. Perceived behavioral control, self-efficacy, locus of control, and the theory of planned Behavior. J Appl Soc Psychol 2002;32(4):665-683. [doi: 10.1111/j.1559-1816.2002.tb00236.x]
59. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]

XSL•FO
RenderX

60. Shebl MA, Toraih E, Shebl M, Tolba AM, Ahmed P, Banga HS, et al. Preoperative anxiety and its impact on surgical outcomes: a systematic review and meta-analysis. J Clin Transl Sci 2025 Jan;9(1):e33. [doi: 10.1017/cts.2025.6] [Medline: 40052059]
61. Gan TJ. Poorly controlled postoperative pain: prevalence, consequences, and prevention. J Pain Res 2017 Sep;10:2287-2298 [FREE Full text] [doi: 10.2147/JPR.S144066] [Medline: 29026331]
62. Sierra MC, To H, Song WJ. Use and application of mHealth technologies in perioperative surgical care: narrative review. JMIR Mhealth Uhealth 2025 Oct 10;13:e52206 [FREE Full text] [doi: 10.2196/52206] [Medline: 41072920]
63. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
64. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
65. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
66. Hekler E, Klasnja P, Froehlich J, Buman M. Mind the theoretical gap: interpreting, using, and developing behavioral theory in HCI research. 2013 Presented at: CHI '13: The SIGCHI Conference on Human Factors in Computing Systems; April 27, 2013 to May 2, 2013; Paris, France. [doi: 10.1145/2470654.2466452]
67. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
68. Klasnja P, Consolvo S, Pratt W. How to evaluate technologies for health behavior change in HCI research. In: CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: CHI '11: CHI Conference on Human Factors in Computing Systems; May 7-12, 2011; Vancouver, BC, Canada p. 3063-3072. [doi: 10.1145/1978942.1979396]
69. Eaton C, Vallejo N, McDonald X, Wu J, Rodríguez R, Muthusamy N, et al. User engagement with mHealth interventions to promote treatment adherence and self-management in people with chronic health conditions: systematic review. J Med Internet Res 2024 Sep 24;26:e50508 [FREE Full text] [doi: 10.2196/50508] [Medline: 39316431]
70. Alshurafa N, Jain J, Alharbi R, Iakovlev G, Spring B, Pfammatter A. Is more always better?: discovering incentivized mHealth intervention engagement related to health behavior trends. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2018 Dec 17;2(4):1-26. [doi: 10.1145/3287031]
71. Yan X, Raj S, Huang B, Park SY, Newman MW. Toward lightweight in-situ self-reporting: an exploratory study of alternative smartwatch interface designs in context. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2020 Dec 18;4(4):1-22 [FREE Full text] [doi: 10.1145/3432212]
72. Nisbett R, Wilson TD. Telling more than we can know: verbal reports on mental processes. Psychological Review 1977 May;84(3):231-259. [doi: 10.1037/0033-295X.84.3.231]
73. Stecher C, Sullivan M, Huberty J. Using personalized anchors to establish routine meditation practice with a mobile app: randomized controlled trial. JMIR Mhealth Uhealth 2021 Dec 22;9(12):e32794 [FREE Full text] [doi: 10.2196/32794] [Medline: 34941558]
74. Li I, Dey AK, Forlizzi J. Using context to reveal factors that affect physical activity. ACM Trans Comput-Hum Interact 2012;19(1):1-21. [doi: 10.1145/2147783.2147790]
75. Gollwitzer P. Implementation intentions: strong effects of simple plans. American Psychologist 1999 Jul;54(7):493-503 [FREE Full text] [doi: 10.1037/0003-066X.54.7.493]
76. Xu K, Yan X, Newman M. Understanding people's experience for physical activity planning and exploring the impact of historical records on plan creation and execution. 2022 Presented at: CHI '22: The 2022 CHI Conference on Human Factors in Computing Systems; April 29-May 5, 2022; New Orleans, LA. [doi: 10.1145/3491102.3501997]
77. Sniehotta FF, Scholz U, Schwarzer R. Bridging the intention–behaviour gap: planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. Psychology & Health 2005 Apr;20(2):143-160. [doi: 10.1080/08870440512331317670]
78. Alberts L, Lyngs U, Lukoff K. Designing for sustained motivation: a review of self-determination theory in behaviour change technologies. Interacting with Computers 2024 Sep 18;38:447-468. [doi: 10.1093/iwc/iwae040]
79. Conner M, Armitage CJ. Extending the theory of planned behavior: a review and avenues for further research. J Applied Social Pyschol 2006 Jul 31;28(15):1429-1464. [doi: 10.1111/j.1559-1816.1998.tb01685.x]
80. Upchurch DM, Johnson PJ. Gender differences in prevalence, patterns, purposes, and perceived benefits of meditation practices in the United States. J Womens Health (Larchmt) 2019 Mar;28(2):135-142 [FREE Full text] [doi: 10.1089/jwh.2018.7178] [Medline: 30543475]

XSL•FO
RenderX

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
GRAMMS: Good Reporting of a Mixed Methods Study
HAPA: Health Action Process Approach
HIPAA: Health Insurance Portability and Accountability Act
mHealth: mobile health
TPB: Theory of Planned Behavior

Edited by S Brini; submitted 13.Jul.2024; peer-reviewed by R Nooripour, U Tandon; comments to author 06.Feb.2025; revised version received 04.Apr.2026; accepted 06.Apr.2026; published 18.Sep.2026.

Please cite as:
Yan X, Hassett AL, Waljee JF, Newman MW, Park SY, Bei R, Carlozzi NE
Engagement With a Relaxation-Based Mobile Health Intervention for Perioperative Anxiety: Prospective Longitudinal Cohort Mixed Methods Study
JMIR Mhealth Uhealth 2026;14:e64278
URL: https://mhealth.jmir.org/2026/1/e64278
doi:10.2196/64278
PMID:

©Xinghui Yan, Afton L Hassett, Jennifer F Waljee, Mark W Newman, Sun Young Park, Rongqi Bei, Noelle E Carlozzi. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 18.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Association Between Intradialytic Foot Temperature Gradients and Severe Peripheral Artery Disease in Patients Undergoing Hemodialysis: Prospective Cohort Study

Donna Shu-Han Lin[1,2], MD; Chih-Cheng Wu[2,3,4,5], MD, PhD; Chung-Wei Yang[3,6], MD; Jen-Kuang Lee[3,7,8,9,10], MD, PhD; Chien-Yi Hsu[11,12], MD, PhD

[1]Department of Internal Medicine, Division of Cardiology, Shin Kong WHS Memorial Hospital, Taipei, Taiwan
[10]Telehealth Center, National Taiwan University Hospital, Taipei, Taiwan
[11]Taipei Heart Institute, Taipei Medical University, No. 252, Wuxing St, Taipei, Taiwan, Taiwan
[12]Division of Cardiology, Department of Internal Medicine, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan
[2]Department of Internal Medicine, Division of Cardiology, National Taiwan University, Hsin-Chu Branch, Hsinchu, Taiwan
[3]Department of Internal Medicine, College of Medicine, National Taiwan University Hospital, Taipei, Taiwan
[4]Institute of Biomedical Engineering, National Tsing-Hua University, Hsinchu, Taiwan
[5]Institute of Cellular and System Medicine, National Health Research Institutes, Zhunan, Taiwan
[6]Department of Internal Medicine, Division of Nephrology, National Taiwan University Hospital, Hsin-Chu Branch, Hsinchu, Taiwan
[7]Department of Internal Medicine, Division of Cardiology, National Taiwan University Hospital, Taipei, Taiwan
[8]Department of Laboratory Medicine, National Taiwan University College of Medicine, Taipei, Taiwan
[9]Cardiovascular Center, National Taiwan University Hospital, Taipei, Taiwan

Corresponding Author:
Chien-Yi Hsu, MD, PhD
Taipei Heart Institute, Taipei Medical University, No. 252, Wuxing St, Taipei, Taiwan, Taiwan

Abstract

Background: Peripheral artery disease (PAD) is prevalent in hemodialysis (HD) patients but remains difficult to diagnose due to unreliable ankle-brachial index (ABI) measurements caused by vascular calcification. We hypothesized that monitoring foot temperatures and their changes during HD, provoked by hemodynamic shifts, could serve as a novel diagnostic tool for detecting severe PAD.

Objective: This study aims to evaluate the diagnostic utility of intradialytic foot temperature monitoring using wearable mobile health (mHealth) sensors for detecting severe PAD in patients undergoing HD.

Methods: In this prospective observational study, we recruited 109 patients undergoing chronic maintenance HD. Skin temperatures were monitored continuously at the instep and sole during HD sessions using wearable sensors (Temp Pal, iWEECARE Co, Ltd), which are wireless, nanochip-based mHealth thermometers with ±0.05 °C accuracy. The diagnostic performance of temperature parameters, specifically the temperature difference between the wrist and foot at the end of HD, was evaluated against vascular duplex ultrasound (DUS). Severe PAD was defined as abnormal (monophasic or absent) flow in the distal posterior tibial artery (PTA) or anterior tibial artery (ATA).

Results: Patients with abnormal tibial flow on DUS exhibited significantly different temperature gradients between the wrist and foot at the end of dialysis compared to those with normal flow. Receiver operating characteristic (ROC) analysis demonstrated that the wrist-foot temperature difference at the sole effectively discriminated abnormal ATA flow (area under the curve [AUC] 64.5%, 95% CI 57.8% - 70.9%) and abnormal PTA flow (AUC 62.0%, 95% CI 55.2% - 68.5%). Notably, an instep wrist-foot temperature difference cutoff of –2.2 °C or less yielded a specificity of 97.6% (95% CI 94.0%-99.3%) for detecting abnormal PTA flow.

Conclusions: Intradialytic foot temperature monitoring, particularly the temperature gradient between the wrist and foot, significantly correlates with severe tibial artery flow abnormalities. This method offers an accessible, noninvasive screening adjunct to the ABI for detecting severe PAD in the HD population. These findings indicate an association rather than a standalone diagnostic test; given the modest discriminative ability, external validation is warranted.

(JMIR Mhealth Uhealth 2026;14:e96782) doi:10.2196/96782

KEYWORDS

mobile health; mHealth; sensors; peripheral artery disease; hemodialysis; remote patient monitoring

XSL•FO
RenderX

Introduction

Patients with end-stage renal disease (ESRD) are at elevated risk of multivascular cardiovascular disease, including coronary artery disease (CAD), cerebrovascular disease, and peripheral artery disease (PAD) [1,2], compared to the general population. In addition, the presentation of PAD in patients undergoing dialysis manifests a worse prognosis than that in patients without ESRD, with greater risk of major adverse limb events, including amputation or hospitalization for chronic limb-threatening ischemia (CLTI), and mortality [3-6]. Indeed, patients with ESRD receiving dialysis are at even higher risk of PAD and adverse limb outcomes than those in earlier stages of chronic kidney disease. As with most diseases, earlier detection is pivotal to the optimal management of these patients. However, diagnosis of PAD remains challenging in the hemodialysis (HD) population for several reasons. First, symptoms are often mild, atypical, or even absent in this group due to limited ambulation or concomitant neuropathy [7]. Second, the severe vascular calcification that is commonly seen in patients with ESRD often leads to falsely elevated ankle-brachial index (ABI) measurements due to noncompressible vessels [8]. Third, the presence of below-the-ankle disease in renal-diabetic foot ulcers is associated with higher risk of treatment failure [9] but is less likely to be detected using conventional ABI or vascular duplex measurements in the tibial arteries. Several noninvasive modalities are used to assess PAD, including the ABI, toe-brachial index, Doppler waveform analysis, duplex ultrasonography, and computed tomographic or magnetic resonance angiography [10]. Infrared thermography, which maps skin-temperature variation as a surrogate for perfusion, has also been applied to identify ischemic changes, although it requires dedicated cameras and is sensitive to ambient conditions [11,12].

Foot temperature monitoring has previously been advocated as a means of surveillance for the development of diabetic foot ulcers [13-15]. This was based on the rationale that inflammation usually precedes overt ulceration in neuropathic ulcers, which can be detected by an increase in local foot temperature. However, for ischemic ulcers seen in PAD, temperatures are usually low. We previously demonstrated that an increase in foot temperature of 2 °C or greater after endovascular therapy (EVT) for chronic foot ulcers in the absence of active infection was associated with better wound healing compared to those with a foot temperature increase of less than 2 °C after EVT [16]. As yet, how temperatures relate to outcomes in asymptomatic patients is uncertain. We hypothesized that foot temperatures in asymptomatic patients also relate to the degree of ischemia and may provide significant diagnostic value. In addition, in patients with ESRD whose ABIs are often falsely negative due to noncompressible vessels, foot temperatures may have higher sensitivity in diagnosing PAD than conventional ABIs and offer greater accessibility than vascular duplex ultrasound (DUS).

Because symptoms related to PAD often manifest in patients with ESRD during their HD sessions due to the significant fluid shifts and hemodynamic changes, we initiated a prospective study that investigated the associations among ABI, DUS measurements, and intradialytic foot temperatures. Our goal was to explore the diagnostic value of intradialytic foot temperatures in identifying severe PAD, potentially offering a readily accessible means of early detection in this vulnerable patient group.

Methods

Study Cohort

This prospective, observational cohort study was conducted at 2 tertiary centers in Taiwan: National Taiwan University Hospital, Hsinchu Branch, and Shin Kong Wu Ho-Su Memorial Hospital. Patients with ESRD undergoing maintenance HD were screened for inclusion. Exclusion criteria included age under 20 years, prior lower-limb amputations (minor or major), and refusal to provide informed consent. Each lower limb was evaluated as an independent unit. Following enrollment, data on participant characteristics, biochemical profiles, and ABI measurements were collected. The presence of severe PAD was confirmed using DUS as the gold standard, defined by monophasic or absent flow in the distal posterior tibial artery (PTA) and distal anterior tibial artery (ATA) at 5 cm above the ankle.

Ethical Considerations

This study was approved by the Institutional Review Board (IRB) of both the National Taiwan University Hospital, Hsinchu Branch (IRB number 111 - 105-E), and Shin Kong Wu Ho-Su Memorial Hospital (IRB number 20230702D); written informed consent was obtained from all participants.

Data Collection

Following study enrollment, a systematic review of electronic medical records and baseline clinical assessments of the participants was conducted. Demographic information, including age and sex, and a detailed history of comorbid conditions such as hypertension, diabetes mellitus (DM), hyperlipidemia, CAD, congestive heart failure, and prior cerebrovascular or myocardial events were recorded. Current medication profiles were also documented, specifically the use of erythropoietin, vitamin D analogs, antihypertensive agents, and iron supplements. Comprehensive laboratory data were gathered, including complete blood counts, serum albumin, and markers of dialysis adequacy such as predialysis and postdialysis blood urea nitrogen and Kt/V. Additionally, we recorded metabolic indicators including fasting blood glucose, lipid panels, and bone mineral markers such as calcium, phosphorus, and intact parathyroid hormone levels. As part of the initial vascular screening, the ABI was measured for each limb.

Foot Temperature Measurements

To record skin temperatures, contact thermometers were placed on the instep and sole of both feet, as well as the wrist of the nonaccess arm (Figures 1 and 2). The instep (dorsal midfoot) overlies the anterior tibial and dorsalis pedis angiosomes, whereas the sole reflects the posterior tibial angiosome; monitoring both sites therefore samples the 2 principal tibial-artery territories assessed by DUS, and these acral sites are among the earliest to display perfusion-related thermal

XSL•FO
RenderX

changes. We used Temp Pal (iWEECARE Co Ltd), a wireless, mobile health (mHealth)–based thermometer system that allows for simultaneous, multisite continuous monitoring. Five devices were paired via Bluetooth to a single iOS mobile device running the Temp Pal app, which captured and uploaded continuous temperature data to a cloud-based server for export. This system uses an autocalibration algorithm to ensure a high degree of measurement reliability, maintaining accuracy and precision within ±0.05 °C. Baseline body temperature was measured at the start and end of each HD session using an infrared ear thermometer, with simultaneous readings obtained from the skin sensors. Throughout the entire procedure, skin temperatures at the wrist and both feet were monitored continuously to capture fluctuations related to intradialytic hemodynamic shifts. Furthermore, clinical parameters, including dialysate temperature, blood pressure, and heart rate, were systematically recorded during the session.

Figure 1. Schematic of foot temperature monitoring using Temp Pal nanochip-based thermometer.

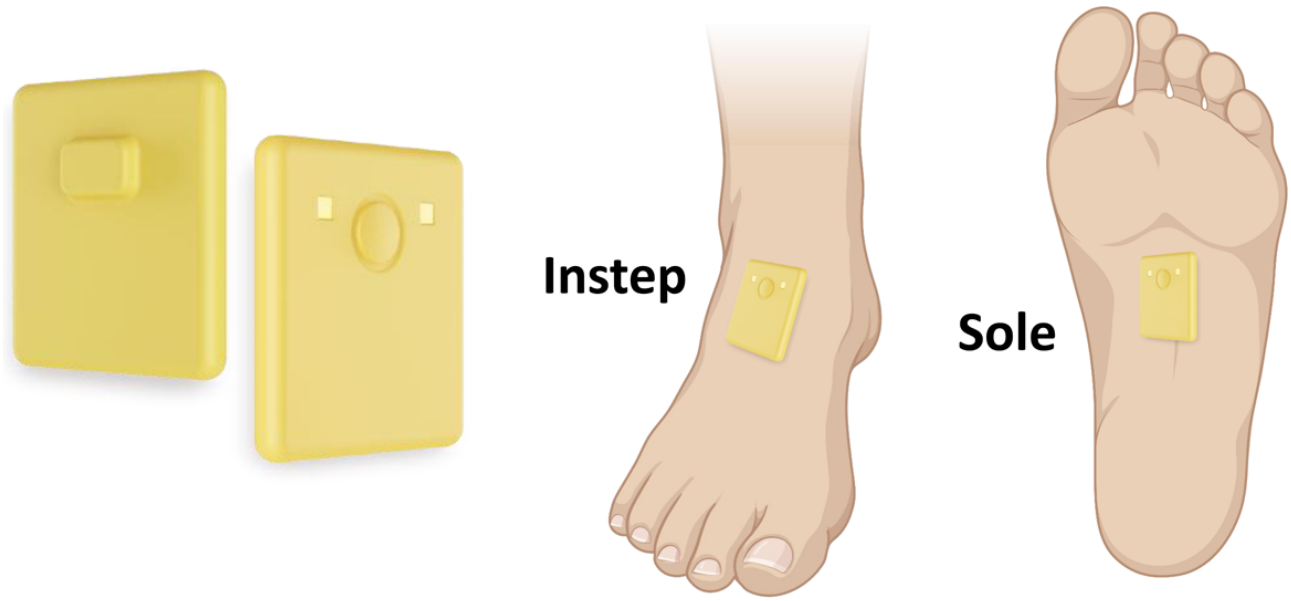


Figure 2. Example photos showing intradialytic foot and wrist temperature monitoring with Temp Pal nanochip-based thermometer.



Vascular DUS Measurements

To evaluate the presence and severity of PAD, all participants underwent DUS performed by experienced technicians. DUS served as the gold standard for this study, specifically focusing on flow patterns in the tibial arteries. Measurements were obtained from the distal PTA and ATA at a standardized location 5 cm above the ankle. Severe PAD was defined by the identification of monophasic or absent flow in either the distal PTA or ATA. These flow patterns were contrasted with normal (triphasic or biphasic) flow to categorize the limbs for comparative analysis of intradialytic temperature shifts.

Statistical Analysis

Inferential analyses were conducted at the limb level; 109 patients contributed 218 limbs eligible for analysis. Temperature parameters were compared between limbs with normal vs abnormal distal PTA/ATA flow using generalized estimating equation (GEE) with a normal distribution and an identity link, accounting for the intrapatient correlation between the 2 limbs of the same individual by using robust (empirical) standard errors. We further performed multivariable GEE analyses to investigate the independent associations of each temperature parameter and ABI value with abnormal distal PTA/ATA flow. The models were adjusted for age, DM, CAD, congestive heart failure, Kt/V (Daugirdas formula), and calcium × phosphorus product. We intentionally limited the adjustment to 6 confounding variables to avoid model overfitting, given that the number of events was 51 for abnormal distal PTA flow and 48 for abnormal distal ATA flow.

Discriminative performance of abnormal distal PTA/ATA flow was evaluated using receiver operating characteristic (ROC) curve analysis, with area under the curve (AUC) calculated as

XSL•FO
RenderX

the primary metric. The 95% CI for the AUC was derived using DeLong nonparametric method. When the AUC was statistically significant (ie, the 95% CI did not include 0.5), the optimal cutoff was determined using the Youden index to balance sensitivity and specificity. To determine if temperature monitoring offers incremental diagnostic value beyond the ABI, the statistical significance of differences between the respective AUC values was assessed via DeLong test. Statistical significance was defined as a 2-sided *P* value of less than .05. All analyses were conducted with MedCalc Statistical Software (version 20.019; MedCalc Software).

Results

Baseline Characteristics of the Study Cohort

The study included 109 patients with ESRD undergoing maintenance HD. The mean age of the participants was 65.2 (SD 11.9) years, and 56.9% (n=62) were male participants. Comorbid conditions were highly prevalent in this cohort, with hypertension (n=89, 81.7%) and DM (n=61, 55.9%) being the most frequent, followed by gout (n=29, 26.6%), CAD (n=26, 23.9%), and congestive heart failure (n=25, 22.9%). A smaller proportion of patients had a history of malignancy (n=6, 5.5%), cerebrovascular accidents (n=4, 3.7%), and myocardial infarctions (n=2, 1.8%). Regarding pharmacological management, the majority of patients were receiving erythropoietin (n=97, 88.9%) and iron supplements (n=84, 77.1%), while 58.7% (n=64) were on antihypertensive medications. Laboratory assessments revealed a mean hemoglobin of 11.6 (SD 1.2) g/dL and mean albumin of 3.99 (SD 0.37) g/dL. Dialysis adequacy was reflected by a mean Daugirdas Kt/V of 1.62 (SD 0.21). Biochemical markers related to bone and mineral metabolism showed a mean calcium of 8.92 (SD 0.98) mg/dL, mean phosphorus of 5.03 (SD 1.35) mg/dL, and median intact parathyroid hormone level of 290.7 (IQR 106.1-570.8) pg/mL. Iron status was characterized by a mean transferrin saturation of 27.9% (SD 11.8%) and median ferritin level of 567 (IQR 333-738) ng/mL. Detailed demographic, clinical, and biochemical data are summarized in Table 1.

XSL•FO
RenderX

Table . Demographics and characteristics of the 109 patients.

Variable	Descriptive statistics
Demographics	
Age (y), mean (SD)	65.2 (11.9)
Sex (male), n (%)	62 (56.9)
Comorbid condition, n (%)	
Hypertension	89 (81.7)
Diabetes mellitus	61 (55.9)
Hyperlipidemia	47 (43.1)
Coronary artery disease	26 (23.9)
Congestive heart failure	25 (22.9)
Chronic obstructive pulmonary disease	1 (0.9)
Gout	29 (26.6)
Chronic hepatitis	11 (10.1)
Neuropathy	2 (1.8)
Malignancy	6 (5.5)
History of prior event, n (%)	
Myocardial infarction	2 (1.8)
Cerebrovascular accident	4 (3.7)
Gastrointestinal bleeding	0 (0)
Parathyroidectomy	2 (1.8)
Medication, n (%)	
Erythropoietin	97 (88.9)
Vitamin D	11 (10.1)
Antihypertensive drugs	64 (58.7)
Iron supplements	84 (77.1)
Complete blood count, mean (SD)	
White blood cell count (1000/μL)	6.5 (2.1)
Red blood cell (1000/μL)	3.6 (0.5)
Hemoglobin (g/dL)	11.6 (1.2)
Hematocrit (%)	35.1 (3.5)
Mean corpuscular volume (fL)	93.5 (7.4)
Platelet (1000/μL)	168.1 (58.0)
Biochemistry	
Albumin (g/dL), mean (SD)	3.99 (0.37)
Alanine aminotransferase (U/L), mean (SD)	18.5 (10.8)
Aspartate aminotransferase (U/L), mean (SD)	16.6 (13.1)
Alkaline phosphatase (U/L), median (IQR)	70 (57-84)
Total cholesterol (mg/dL), mean (SD)	147.9 (36.0)
Triglyceride (mg/dL), median (IQR)	138.5 (90.5-207.0)
Fasting blood glucose (mg/dL), mean (SD)	136.6 (48.4)
Predialysis blood urea nitrogen (mg/dL), mean (SD)	83.2 (24.6)
Postdialysis blood urea nitrogen (mg/dL), mean (SD)	21.6 (7.8)
Creatinine (mg/dL), mean (SD)	11.7 (2.7)

XSL•FO
RenderX

Variable	Descriptive statistics
Uric acid (mg/dL), mean (SD)	7.6 (1.9)
Sodium (mEq/L), mean (SD)	137.1 (3.3)
Potassium (mEq/L), mean (SD)	4.7 (0.70)
Calcium (mg/dL), mean (SD)	8.92 (0.98)
Phosphorus (mg/dL), mean (SD)	5.03 (1.35)
Urea reduction ratio, Kt/V, mean (SD)	58.0 (30.8)
Gotch, Kt/V, mean (SD)	1.37 (0.18)
Daugirdas, Kt/V, mean (SD)	1.62 (0.21)
Iron (ug/dL), mean (SD)	64.3 (27.3)
Unsaturated iron-binding capacity (ug/dL), mean (SD)	173.3 (47.2)
Total iron binding capacity (ug/dL), mean (SD)	231.7 (42.2)
Ferritin (ng/mL), median (IQR)	567 (333-738)
Transferrin saturation (%), mean (SD)	27.9 (11.8)
Intact parathyroid hormone (pg/mL), median (IQR)	290.7 (106.1-570.8)
Calcium × phosphorus, mean (SD)	45.6 (13.1)

Foot Temperature Parameters and ABI in Limbs With and Without Severe PAD

A total of 218 limbs were evaluated to analyze the relationship between distal arterial flow and various physiological parameters (Table 2). The parameters analyzed comprised the ABI and the temperature parameters, namely, wrist-to-foot and ear-to-foot temperature differences at the end of HD and absolute foot temperature before and at the end of HD. Limbs with abnormal flow in either the distal PTA or ATA exhibited significantly lower ABI values compared to those with normal flow (mean 0.90, SD 0.26 vs mean 1.13, SD 0.20 for PTA; mean 0.89, SD 0.27 vs mean 1.13, SD 0.200 for ATA; both $P<.001$). However, it is noteworthy that despite these significant differences, limbs with confirmed severe PAD still exhibited mean ABI values as high as 0.89 (SD 0.27) to 0.90 (SD 0.26), which border the clinically normal range and highlight the limitations of ABI in representing disease severity in this population.

XSL•FO
RenderX

Table . Comparison of temperature parameters in limbs with normal and abnormal distal posterior tibial artery (PTA)/anterior tibial artery (ATA) flow.

Variable	Limbs (N=218), mean (SD)	Distal PTA flow			Distal ATA flow		
		Abnormal (n=51), mean (SD)	Normal (n=167), mean (SD)	*P* value	Abnormal (n=48), mean (SD)	Normal (n=170), mean (SD)	*P* value
Ankle-brachial index (n=203)	1.08 (0.23)	0.90 (0.26)	1.13 (0.20)	<.001	0.89 (0.27)	1.13 (0.20)	<.001
Instep (°C)							
Temperature difference between the wrist and foot at the end of the HD[a] session	1.50 (2.68)	0.75 (3.15)	1.73 (2.48)	.051	0.61 (2.91)	1.75 (2.57)	.046
Temperature difference between the ear and foot at the end of the HD session	5.38 (2.54)	4.92 (2.66)	5.52 (2.50)	.10	4.69 (2.59)	5.57 (2.50)	.09
Foot temperature before the HD session	30.51 (2.80)	30.73 (2.77)	30.45 (2.82)	.45	30.62 (2.59)	30.48 (2.87)	.99
Foot temperature at the end of the HD session	30.69 (2.50)	31.14 (2.66)	30.55 (2.44)	.11	31.41 (2.55)	30.48 (2.46)	.08
Sole (°C)							
Temperature difference between the wrist and foot at the end of the HD session	1.95 (2.73)	1.05 (3.05)	2.23 (2.57)	.02	0.81 (2.51)	2.27 (2.71)	.006
Temperature difference between the ear and foot at the end of the HD session	5.83 (2.68)	5.22 (2.64)	6.02 (2.68)	.04	4.90 (2.42)	6.10 (2.70)	.03
Foot temperature before the HD session	30.17 (2.67)	30.29 (2.69)	30.13 (2.67)	.48	30.44 (2.47)	30.09 (2.72)	.33
Foot temperature at the end of the HD session	30.23 (2.60)	30.84 (2.63)	30.05 (2.58)	.04	31.20 (2.36)	29.96 (2.61)	.02

[a]HD: hemodialysis.

Regarding intradialytic temperature dynamics, the temperature difference between the wrist and the foot at the end of the HD session was significantly smaller in limbs with abnormal flow (Table 2). Specifically, at the sole, the wrist-foot temperature difference was mean 1.05 (SD 3.05) °C in the abnormal PTA group compared to mean 2.23 (SD 2.57) °C in the normal group (*P*=.02). A similar pattern was observed at the instep, reaching statistical significance for ATA flow (*P*=.046) but showing only a borderline association for PTA flow (*P*=.051). While foot temperatures before the HD session did not differ significantly between groups, the foot temperature at the end of the session tended to be higher in limbs with abnormal ATA flow compared to those with normal flow (mean 31.41, SD 2.55 °C vs mean 30.48, SD 2.46 °C at the instep; *P*=.08; mean 31.20, SD 2.36 °C vs mean 29.96, SD 2.61 °C at the sole; *P*=.02).

Additionally, a multivariable GEE model, adjusting for age, DM, CAD, congestive heart failure, Kt/V (Daugirdas formula), and the calcium × phosphorus product, revealed that an increased ABI was significantly associated with a reduced risk of abnormal distal PTA flow (odds ratio [OR] 0.73, 95% CI 0.63 - 0.83) and ATA flow (OR 0.71, 95% CI 0.59 - 0.85; Table 3). Moreover, wrist-to-foot temperature differences at the end of the HD session, evaluated at the instep and sole, respectively, were significantly associated with abnormal distal PTA flow at the sole (OR 0.81, 95% CI 0.68 - 0.97), with a borderline, nonsignificant association at the instep (OR 0.84, 95% CI 0.70 - 1.003), and with abnormal distal ATA flow at both the

XSL•FO
RenderX

instep (OR 0.84, 95% CI 0.71 - 0.99) and sole (OR 0.81, 95% CI 0.70 - 0.95). A similar pattern of results was observed for foot temperature at the end of the HD session.

Table . Associations of each temperature parameter and the ankle-brachial index (ABI) with abnormal distal posterior tibial artery (PTA)/anterior tibial artery (ATA) flow using a generalized estimating equation model[a].

Variable	Distal PTA flow		Distal ATA flow	
	OR[b] (95% CI)	*P* value	OR (95% CI)	*P* value
ABI (n=203)	0.73 (0.63 - 0.83)	<.001	0.71 (0.59 - 0.85)	<.001
Instep				
Temperature difference between the wrist and foot at the end of the HD[c] session	0.84 (0.70 - 1.003)	.053	0.84 (0.71 - 0.99)	.04
Temperature difference between the ear and foot at the end of the HD session	0.86 (0.75 - 0.995)	.04	0.83 (0.71 - 0.98)	.03
Foot temperature before the HD session	1.10 (0.93 - 1.31)	.27	1.06 (0.89 - 1.26)	.53
Foot temperature at the end of the HD session	1.16 (0.99 - 1.35)	.07	1.21 (1.02 - 1.44)	.03
Sole				
Temperature difference between the wrist and foot at the end of the HD session	0.81 (0.68 - 0.97)	.02	0.81 (0.70 - 0.95)	.008
Temperature difference between the ear and foot at the end of the HD session	0.84 (0.73 - 0.97)	.02	0.82 (0.70 - 0.95)	.008
Foot temperature before the HD session	1.11 (0.93 - 1.32)	.27	1.14 (0.93 - 1.40)	.20
Foot temperature at the end of the HD session	1.20 (1.02 - 1.40)	.03	1.24 (1.06 - 1.46)	.008

[a]Adjusted for age, diabetes mellitus, coronary artery disease, congestive heart failure, Daugirdas Kt/V, and calcium × phosphorus.

[b]OR: odds ratio.

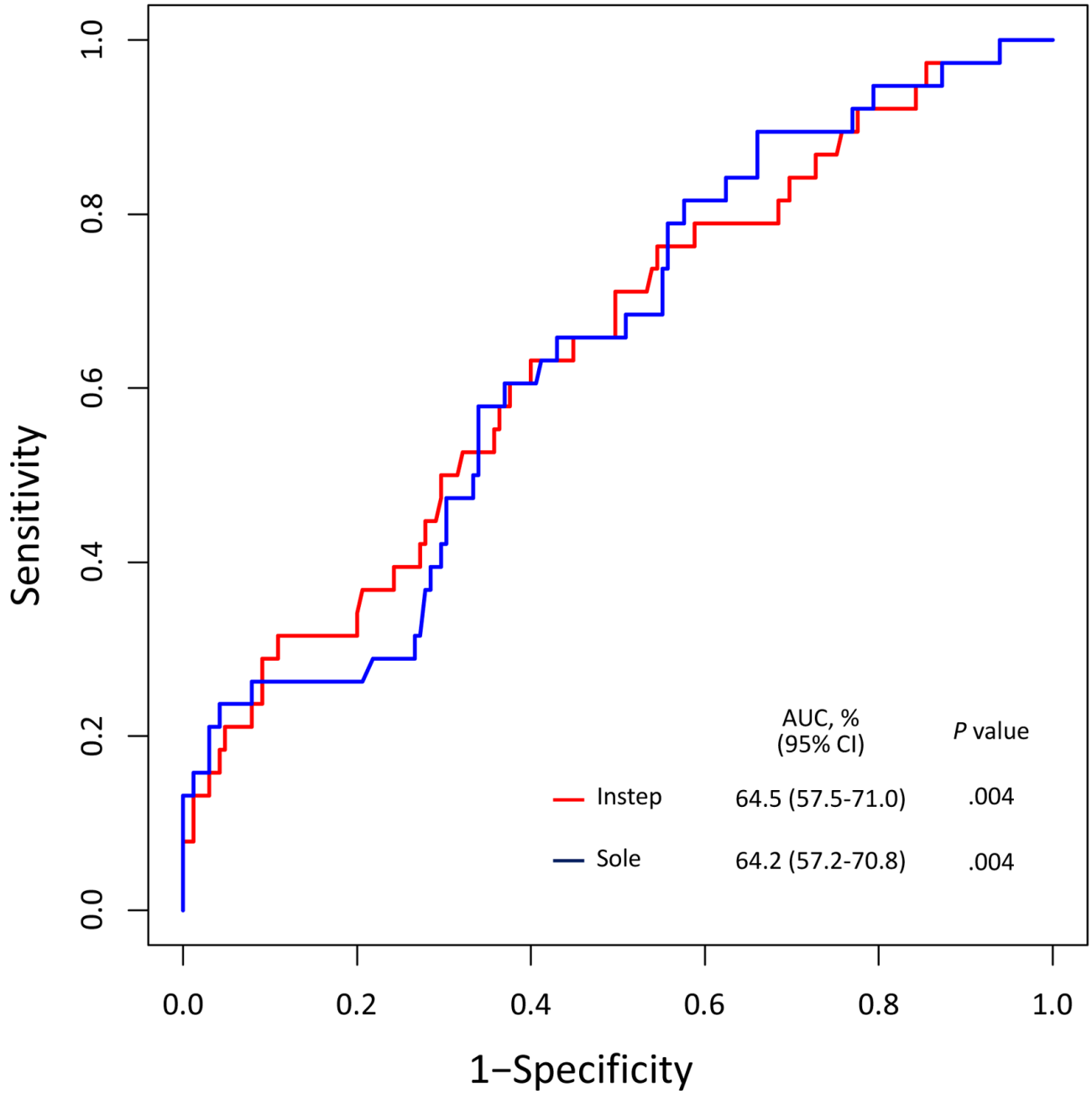
[c]HD: hemodialysis.

Diagnostic Performance of Foot Temperature Parameters in Discriminating an ABI Value of Less Than 0.9

The diagnostic utility of intradialytic temperature monitoring was evaluated for its ability to predict an ABI value of less than 0.9 (Table 4). ROC analysis revealed that the temperature difference between the wrist and foot at the end of the HD session was the most effective parameter for this purpose. At the instep, this gradient achieved an AUC of 64.5% (95% CI 57.5% - 71.0%). Similarly, the gradient measured at the sole demonstrated a significant AUC of 64.2% (95% CI 57.2% - 70.8%; Figure 3). For the instep measurement, an optimal cutoff of 1.25 °C or less yielded a sensitivity of 63.2% (95% CI 46.0% - 78.2%) and a specificity of 60.0% (95% CI 52.1% - 67.5%). At the sole, a cutoff of 2.83 °C or less provided a higher sensitivity of 81.6% (95% CI 65.7% - 92.3%) but lower specificity of 42.4% (95% CI 34.8% - 50.3%). Other parameters, including absolute foot temperatures before or after HD and the ear-foot temperature gradient, did not reach statistical significance in discriminating limbs with an ABI value of less than 0.9 (Table 4).

XSL•FO
RenderX

Figure 3. Receiver operating characteristic curves for wrist-to-foot temperature differences measured at the end of the hemodialysis session, at the instep and sole, for the discrimination of ankle-brachial index value of less than 0.9. AUC: area under the curve.



XSL•FO
RenderX

Table . Summary of temperature parameters of interest for discriminating limbs with an ankle-brachial index value of less than 0.9.

Position/parameter	AUC[a], % (95% CI)	Cutoff, °C[b]	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Instep				
Temperature difference between the wrist and foot at the end of the HD[c] session	64.5 (57.5 - 71.0)[d]	≤1.25	63.2 (46.0 - 78.2)	60.0 (52.1 - 67.5)
Temperature difference between the ear and foot at the end of the HD session	53.2 (46.1 - 60.3)	≤3.71	39.5 (24.0 - 56.6)	72.7 (65.3 - 79.4)
Foot temperature before the HD session	50.9 (43.8 - 58.0)	≤29.4	18.4 (7.7 - 34.3)	67.3 (59.5 - 74.4)
Foot temperature at the end of the HD session	54.2 (47.1 - 61.2)	>31.12	55.3 (38.3 - 71.4)	60.0 (52.1 - 67.5)
Sole				
Temperature difference between the wrist and foot at the end of the HD session	64.2 (57.2 - 70.8)[d]	≤2.83	81.6 (65.7 - 92.3)	42.4 (34.8 - 50.3)
Temperature difference between the ear and foot at the end of the HD session	53.9 (46.8 - 60.9)	≤4.25	44.7 (28.6 - 61.7)	71.5 (64.0 - 78.3)
Foot temperature before the HD session	50.5 (43.4 - 57.6)	>32.15	13.2 (4.4 - 8.1)	70.9 (63.3 - 77.7)
Foot temperature at the end of the HD session	55.2 (48.1 - 62.2)	>31.58	47.4 (31.0 - 64.2)	70.9 (63.3 - 77.7)

[a]AUC: area under the curve.

[b]According to the Youden index.

[c]HD: hemodialysis.

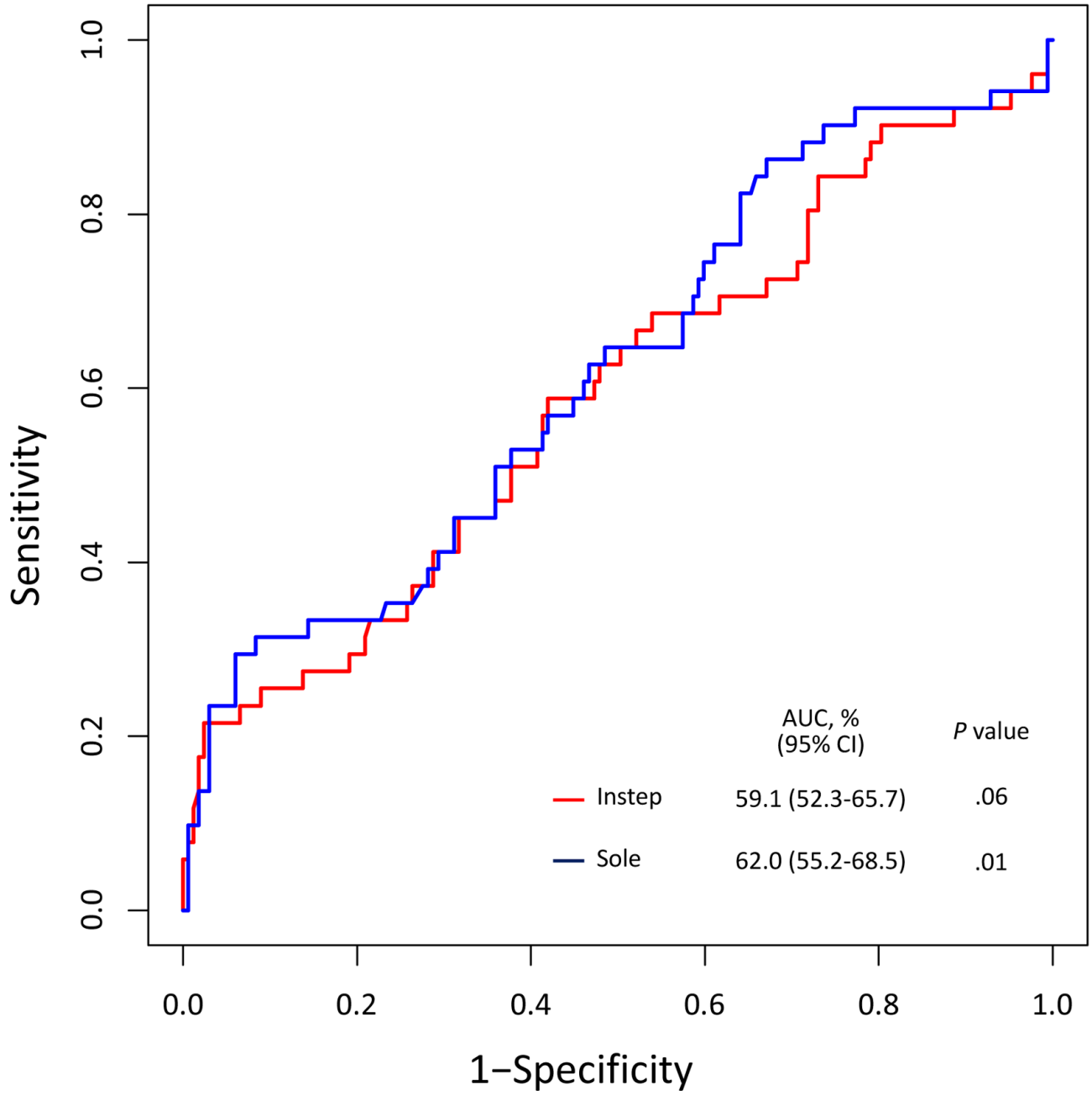
[d]$P<.05$.

Diagnostic Performance of Foot Temperature Parameters in Predicting the Presence of Severe PAD

ROC analysis was performed to further evaluate the diagnostic utility of various temperature parameters for identifying severe PAD (Table 5). The temperature difference between the wrist and the foot at the end of the HD session emerged as a significant indicator of arterial flow abnormalities. Specifically, for the detection of abnormal distal PTA flow, the wrist-foot temperature difference at the sole yielded an AUC of 62.0% (95% CI 55.2% - 68.5%; Figure 4). At the instep, this parameter showed an AUC of 59.1% (95% CI 52.3% - 65.7%). Notably, using a cutoff value of –2.2 °C or less for the wrist-foot difference at the instep provided a high specificity of 97.6% (95% CI 94.0% - 99.3%) for identifying abnormal PTA flow. The temperature parameters demonstrated even greater discriminative power for abnormal distal ATA flow. The wrist-foot temperature difference at the end of dialysis showed an AUC of 64.5% (95% CI 57.8% - 70.9%) at the sole and 62.0% (95% CI 55.2% - 68.5%) at the instep (Figure 5). Furthermore, absolute foot temperature at the end of the HD session was also a significant indicator of abnormal ATA flow, with an AUC of 64.5% (95% CI 57.7% - 70.8%) at the sole and 61.5% (95% CI 54.7% - 68.0%) at the instep. According to Youden index, a sole temperature cutoff of greater than 30.44 °C at the end of HD yielded a sensitivity of 70.8% (95% CI 55.9% - 83.0%) and a specificity of 58.2% (95% CI 50.4% - 65.7%) for detecting abnormal ATA flow (Table 5).

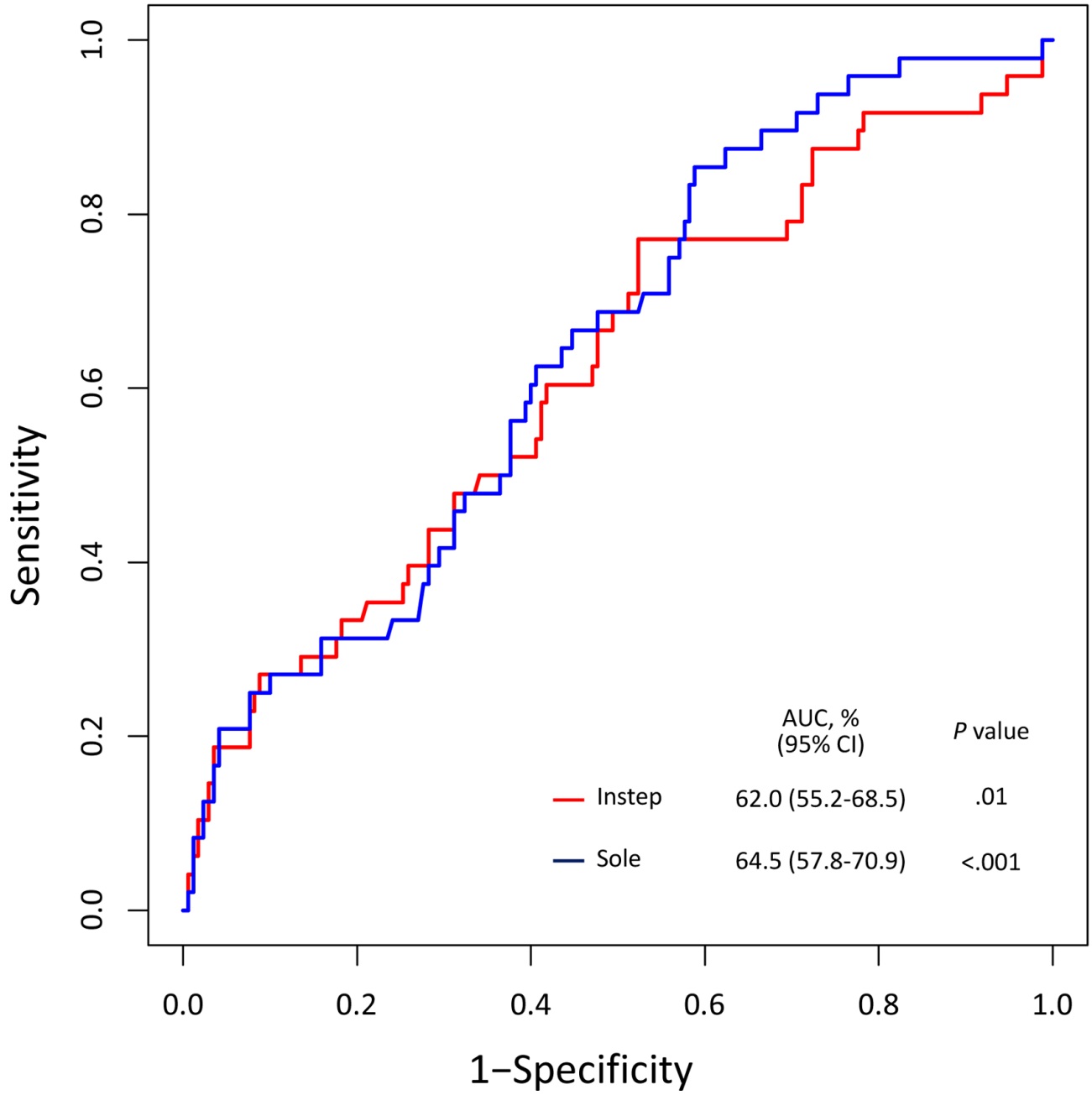
XSL•FO
RenderX

Figure 4. Receiver operating characteristic curves of wrist-to-foot temperature differences measured at the end of the hemodialysis session, at the instep and sole, for the discrimination of abnormal distal posterior tibial artery (PTA) flow. AUC: area under the curve.



XSL•FO
RenderX

Figure 5. Receiver operating characteristic curves of the wrist-to-foot temperature differences measured at the end of the hemodialysis session, at the instep and sole, for the discrimination of abnormal distal anterior tibial artery (ATA) flow. AUC: area under the curve.



XSL•FO
RenderX

Table . Diagnostic value of temperature parameters for distinguishing between normal and abnormal distal posterior tibial artery (PTA)/anterior tibial artery (ATA) flow.

Outcome/position/parameter	AUC[a], % (95% CI)	Cutoff (°C)[b]	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Abnormal distal PTA flow				
Instep				
Temperature difference between the wrist and foot at the end of the HD[c] session	59.1 (52.3 - 65.7)[d]	≤–2.2	21.6 (11.3 - 35.3)	97.6 (94.0 - 99.3)
Temperature difference between the ear and foot at the end of the HD session	57.3 (50.5 - 64.0)[d]	≤4.5	54.9 (40.3 - 68.9)	67.1 (59.4 - 74.1)
Foot temperature before the HD session	52.7 (45.9 - 59.5)	—[e]	—	—
Foot temperature at the end of the HD session	57.2 (50.4 - 63.9)[d]	>31.92	51.0 (36.6 - 65.2)	71.9 (64.4 - 78.5)
Sole				
Temperature difference between the wrist and foot at the end of the HD session	62.0 (55.2 - 68.5)[d]	≤–1.2	29.4 (17.5 - 43.8)	94.0 (89.3 - 97.1)
Temperature difference between the ear and foot at the end of the HD session	59.4 (52.6 - 66.0)[d]	≤4.87	52.9 (38.5 - 67.1)	67.7 (60.0 - 74.7)
Foot temperature before the HD session	52.8 (45.9 - 59.6)	—	—	—
Foot temperature at the end of the HD session	59.4 (52.5 - 65.9)[d]	>30.44	66.7 (52.1 - 79.2)	57.5 (49.6 - 65.1)
Abnormal distal ATA flow				
Instep				
Temperature difference between the wrist and foot at the end of the HD session	62.0 (55.2 - 68.5)[d]	≤1.94	77.1 (62.7 - 88.0)	47.7 (39.9 - 55.4)
Temperature difference between the ear and foot at the end of the HD session	61.5 (54.7 - 68.0)[d]	≤4.5	60.4 (45.3 - 74.2)	68.2 (60.7 - 75.2)
Foot temperature before the HD session	50.2 (43.4 - 57.1)[d]	>29.35	81.3 (67.4 - 91.1)	32.4 (25.4 - 39.9)
Foot temperature at the end of the HD session	61.5 (54.7 - 68.0)[d]	>31.16	66.7 (51.6 - 79.6)	61.8 (54.0 - 69.1)
Sole				
Temperature difference between the wrist and foot at the end of the HD session	64.5 (57.8 - 70.9)[d]	≤2.88	85.4 (72.2 - 93.9)	41.2 (33.7 - 49.0)
Temperature difference between the ear and foot at the end of the HD session	63.8 (57.1 - 70.2)[d]	≤4.87	58.3 (43.2 - 72.4)	68.8 (61.3 - 75.7)
Foot temperature before the HD session	53.8 (46.9 - 60.6)	—	—	—
Foot temperature at the end of the HD session	64.5 (57.7 - 70.8)[d]	>30.44	70.8 (55.9 - 83.0)	58.2 (50.4 - 65.7)

[a]AUC: area under the curve.

[b]According to the Youden index.

[c]HD: hemodialysis.

[d]$P<.05$.

XSL•FO
RenderX

[e]Not applicable

Lastly, we performed DeLong test to examine whether the addition of temperature monitoring yields incremental diagnostic value over the baseline ABI (Table 6). Regarding abnormal distal PTA flow, temperature parameters at both the instep and sole failed to offer incremental diagnostic value over an ABI value of less than 0.9. Conversely, relatively stronger diagnostic utility was observed for abnormal distal ATA flow. For the instep measurements, although the improvements did not achieve statistical significance, the ΔAUC values increased by roughly 1% (95% CI –5.5% to 7.5%) to 4% (95% CI –2.1% to 10.1%). Notably, for the sole measurements, the wrist-to-foot temperature difference at the end of dialysis yielded a significant increase in AUC (ΔAUC 5.6%, 95% CI 0.2%-11.0%). Additionally, a potential trend toward improved AUC was noted for foot temperature at the end of the HD session (ΔAUC 5.1%, 95% CI –1.0% to 11.2%).

XSL•FO
RenderX

Table . Added predictive value of foot temperature parameters over ankle-brachial index (ABI) value of less than 0.9 for abnormal distal posterior tibial artery (PTA) and anterior tibial artery (ATA) flow.

Outcome/position/parameter	ABI <0.9 alone (95% CI)	Temperature+ABI <0.9 (95% CI)	ΔAUC[a], % (95% CI)	*P* value
Abnormal distal PTA flow				
Instep				
Temperature difference between the wrist and foot at the end of the HD[b] session	69.4 (61.6-77.1)	70.4 (60.3-80.5)	1.0 (–5.5 to 7.5)	.76
Temperature difference between the ear and foot at the end of the HD session	69.4 (61.6-77.1)	70.1 (59.9-80.2)	0.7 (–5.7 to 7.0)	.84
Foot temperature before the HD session	69.4 (61.6-77.1)	71.3 (61.3-81.2)	1.9 (–4.9 to 8.6)	.58
Foot temperature at the end of the HD session	69.4 (61.6-77.1)	69.9 (59.6-80.2)	0.5 (–6.0 to 7.1)	.88
Sole				
Temperature difference between the wrist and foot at the end of the HD session	69.4 (61.6-77.1)	71.4 (61.6-81.3)	2.0 (–4.4 to 8.4)	.53
Temperature difference between the ear and foot at the end of the HD session	69.4 (61.6-77.1)	70.6 (60.8-80.4)	1.2 (–4.7 to 7.1)	.69
Foot temperature before the HD session	69.4 (61.6-77.1)	68.0 (57.6-78.5)	–1.3 (–8.0 to 5.3)	.69
Foot temperature at the end of the HD session	69.4 (61.6-77.1)	70.3 (60.3-80.4)	0.9 (–5.3 to 7.2)	.77
Abnormal distal ATA flow				
Instep				
Temperature difference between the wrist and foot at the end of the HD session	71.0 (62.9-79.2)	75.0 (65.7-84.4)	4.0 (–2.1 to 10.1)	.19
Temperature difference between the ear and foot at the end of the HD session	71.0 (62.9-79.2)	74.8 (64.8-84.8)	3.7 (–2.7 to 10.2)	.26
Foot temperature before the HD session	71.0 (62.9-79.2)	72.1 (62.1-82.2)	1.1 (–5.2 to 7.4)	.73
Foot temperature at the end of the HD session	71.0 (62.9-79.2)	74.5 (64.3-84.7)	3.5 (–3.2 to 10.1)	.31
Sole				
Temperature difference between the wrist and foot at the end of the HD session	71.0 (62.9-79.2)	76.6 (68.0-85.2)	5.6 (0.2 to 11.0)	.04
Temperature difference between the ear and foot at the end of the HD session	71.0 (62.9-79.2)	76.3 (67.0-85.5)	5.2 (–0.6 to 11.1)	.08
Foot temperature before the HD session	71.0 (62.9-79.2)	73.7 (63.9-83.5)	2.6 (–3.8 to 9.0)	.42
Foot temperature at the end of the HD session	71.0 (62.9-79.2)	76.1 (66.7-85.6)	5.1 (–1.0 to 11.2)	.10

[a]AUC: area under the curve.

[b]HD: hemodialysis.

XSL•FO
RenderX

Discussion

Principal Findings

In this prospective cohort study of patients with ESRD undergoing maintenance HD, we explored the diagnostic potential of intradialytic foot temperatures for detecting severe PAD. First, we found that in this population, where arterial calcification is prominent, the ABI may remain near-normal despite the presence of severe PAD as confirmed by DUS. Second, absolute foot temperatures at the end of the dialysis session were significantly higher in patients with severe ATA disease compared with those without. Finally, the temperature gradient between the wrist and foot served as a significant predictor of severe ATA or PTA disease, achieving an AUC of up to 64.5% (95% CI 57.8%-70.9%) for abnormal ATA flow. These results suggest that monitoring foot temperatures during HD provides an accessible and noninvasive means of screening for PAD in the vulnerable ESRD population. Nonetheless, the discriminative ability of these temperature parameters was modest (AUCs of approximately 0.59 - 0.65), and because all estimates were derived and tested within the same cohort without internal or external validation, these findings should be interpreted as demonstrating an association rather than establishing a standalone diagnostic tool.

The diagnosis of PAD in patients with ESRD remains a significant clinical challenge due to the inherent limitations of conventional screening tools. The ABI, while widely used as a first-line diagnostic test, frequently yields falsely elevated or "normal" results in the HD population [17,18]. This is due to extensive arterial calcification, which renders the crural vessels noncompressible and prevents the accurate measurement of distal blood pressure [8]. This leads to a high rate of underdiagnosis in a population that is otherwise at extreme risk for CLTI and amputation. Furthermore, renal insufficiency is frequently associated with a more distal, crural distribution of atherosclerosis, which further complicates pressure-based assessments that are more sensitive to proximal disease. In our cohort, limbs confirmed to have severe PAD via DUS still exhibited mean ABI values as high as 0.89 (SD 0.27) to 0.90 (SD 0.26), suggesting that relying on ABI alone would have missed a substantial proportion of patients with severe tibial artery disease. We also demonstrated the feasibility of detecting severe PAD through the temperature gradient between the wrist and foot at the end of the HD session. For example, an instep wrist-foot temperature difference cutoff of –2.2 °C or less achieved a remarkable specificity of 97.6% (95% CI 94.0%-99.3%) for detecting abnormal flow in the PTA. By leveraging the hemodynamic stress during the dialysis process, this method provides a functional assessment of perfusion that is independent of vessel wall compressibility. Intradialytic thermography offers a potential screening adjunct, identifying high-risk patients who would otherwise be cleared by traditional, yet possibly less reliable, pressure-derived measurements.

Based on the established relationship between thermodynamics and hemodynamics, the use of skin temperature during HD as a proxy for arterial perfusion provides a novel physiological entry point for detecting PAD in patients with ESRD. Skin temperature has been previously recognized as an indicator of cutaneous blood flow, with infrared thermography and contact sensors being used to map ischemic zones in high-risk populations [19-21]. Prior studies have demonstrated that abnormal skin temperatures in the feet correlate significantly with reduced ABI values and abnormal limb arterial ultrasonography [20,22]. Our team previously reported that a significant postprocedural increase in foot temperature serves as a robust predictor of wound healing after EVT [16], reinforcing that thermal changes can directly reflect perfusion. However, interpretation of foot temperatures varies in different settings and can be challenging. In patients with DM, focal increases in foot temperatures are predictive of ulceration [15,23], while globally warmer feet suggest autosympathectomy related to diabetic neuropathy, with subsequent arteriovenous shunting and increased blood flow to the skin [19,22]. Thus, absolute foot temperature alone may be confounded and insufficient to reflect PAD severity. Similar to the use of myocardial stress testing to unmask underlying CAD, this present study advances this diagnostic concept by leveraging HD as a "thermal stress test" for the peripheral vasculature. To ensure diagnostic precision, the temperature at the wrist of the nonaccess arm served as a proximal, well-perfused reference. This approach allowed for the isolation of local thermal changes elicited specifically by vascular stress, thus controlling for systemic temperature fluctuations related to the dialysate itself. We found that while baseline temperatures before dialysis were similar between groups, the temperature at the end of the session significantly differentiated those with severe PAD. Specifically, the wrist-foot difference at the sole effectively discriminated abnormal flow in both the PTA and ATA. Interestingly, it is a lack of temperature decrease during HD that identifies severe PAD in our cohort (ie, those with severe PAD have higher foot temperatures). This paradoxical finding conforms with literature [18,19], suggesting disrupted vasoconstriction mechanisms in these patients. Our study demonstrates that wearable mHealth sensors provide a physiological foundation for future research into the underlying mechanisms of PAD in patients with ESRD.

Socioeconomic disparities significantly delay the diagnosis of PAD, leading to higher amputation rates in underserved populations [24]. Patients with ESRD, in particular, are frequently overlooked, with limited access to specialized care due to ambulatory difficulties and socioeconomic disadvantages [7]. Traditional screening for PAD via ABI or DUS is often limited by accessibility, resulting in a reactive approach where referral is initiated only after ulcers or gangrene appear. Intradialytic foot temperature monitoring offers a solution to overcome these barriers by using the Temp Pal, a wearable mHealth thermometer system that provides high-fidelity, continuous monitoring with minimal overhead. While studies have demonstrated the feasibility of smart-socks [21,25], such specialized apparel is unavailable in many regions, including Taiwan. In contrast, cutaneous thermometers are significantly cheaper, highly accessible, and reusable, making them a more practical choice for global implementation. More importantly, integrating this monitoring method into routine dialysis sessions allows patients to undergo screening without additional travel or time commitments, ensuring accessibility to those with limited mobility or resources. Our ultimate goal is to mitigate

XSL•FO
RenderX

health disparities within the high-risk population with ESRD, thereby improving both the trajectory of vascular disease and long-term survival. These potential benefits, improved screening access, reduced disparities, and earlier identification, were not evaluated in the present study and remain hypotheses to be tested in future implementation and outcome studies.

Limitations

While this study offers a novel approach to PAD screening, several limitations must be acknowledged. First, the sample size of 109 patients is relatively small and was recruited from 2 tertiary centers within a single geographic region. As our cohort consisted exclusively of an East Asian population in Taiwan, the generalizability of our findings to other ethnic groups remains uncertain. Given that racial and socioeconomic disparities significantly influence the presentation and progression of CLTI, further research in more diverse populations is necessary to validate our findings. Second, we used DUS flow patterns at 5 cm above the ankle as our gold standard for defining severe PAD. Although DUS is a clinically standard and reliable tool for assessing arterial patency, it primarily evaluates macrovascular flow. We did not incorporate other perfusion markers, such as transcutaneous oxygen tension or skin perfusion pressure, which might have provided a better understanding of how intradialytic temperature shifts correlate with actual tissue oxygenation. Finally, the sensitivity of intradialytic temperature monitoring for detecting severe tibial artery flow abnormalities varied considerably across parameters and measurement sites, with a reciprocal trade-off in specificity. In addition, this study was not designed to assess long-term outcomes. Although early detection is critical in PAD, there is a lack of evidence that this screening method leads to improved limb salvage rates or survival benefits. Future longitudinal studies are required to determine if initiating early interventions based on these thermal warning signs can successfully reduce major adverse limb events, major adverse cardiovascular events, or mortality associated with PAD in the population with ESRD. In addition, as most patients were asymptomatic and were screened irrespective of symptoms, formal clinical PAD staging (eg, Fontaine or Rutherford categories) was not performed; disease severity was instead defined objectively by DUS flow patterns. Moreover, several factors may have affected the accuracy of our temperature measurements: the ambient temperature of the dialysis unit was not systematically recorded, the access (fistula) arm was not measured, and a single nonaccess arm was used as the wrist reference for both legs, which could introduce bias if that reference limb itself had impaired perfusion.

Conclusions

This study demonstrated that intradialytic foot temperature monitoring is a potential diagnostic adjunct for detecting severe PAD in patients with ESRD. We found that in patients with ESRD, the ABI can remain near-normal despite severe disease, whereas the wrist-foot temperature difference at the end of the HD session was significantly associated with abnormal tibial flow, although sensitivity and specificity varied considerably across temperature parameters. Temperature monitoring with mHealth devices during HD may provide an accessible, low-cost screening solution that mitigates health disparities in patients with ESRD. Further longitudinal studies are required to confirm the survival benefits of this novel method. Because the observed discriminative ability was modest and has not yet been externally validated, intradialytic temperature monitoring should be regarded as a complementary adjunct to, rather than a replacement for, established vascular assessments such as the ABI.

Acknowledgments

The authors thank two biostatisticians, Mr Alfred Hsing-Fen Lin and Mr Jia-Kia Wu, who served at Raising Statistics Consultant Inc, for their assistance with statistical analysis during the completion of the manuscript.

Generative AI (Claude) tools were used solely for language editing (grammar and wording) of the manuscript. No content, data, analyses, or references were generated by AI. All authors reviewed and edited the text and take full responsibility for the content of this publication.

Funding

The present work was supported by the following research grants: Shin Kong Wu Ho-Su Memorial Hospital and Taipei Medical University (grant number SKH-TMU-113-01); NSTC114-2314-B-038-126 and NSTC114-2314-B-038-128 from the National Science and Technology Council (NSTC) of Taiwan; and research funds 114TMUH-C01-1 and 114TMUH-SP-03 from Taipei Medical University Hospital, Taipei, Taiwan. The funders had no role in the study design, collection, analysis, or interpretation of data, writing of the report, or decision to submit the article for publication.

Data Availability

Data sharing may be available to investigators whose proposed use of the data has been approved by an independent review committee, upon reasonable request to the corresponding author.

Authors' Contributions

DL contributed to conceptualization, methodology, investigation, data curation, formal analysis, and writing of the original draft. CCW contributed to investigation, resources, and review and editing of the manuscript. CWY contributed to investigation (patient

XSL•FO
RenderX

recruitment and data acquisition), resources, and review and editing of the manuscript. JKL contributed to conceptualization, methodology, supervision, and review and editing of the manuscript. CYH contributed to conceptualization, funding acquisition, project administration, supervision, and review and editing of the manuscript, and is the corresponding author. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. O'Hare AM, Hsu CY, Bacchetti P, Johansen KL. Peripheral vascular disease risk factors among patients undergoing hemodialysis. J Am Soc Nephrol 2002 Feb;13(2):497-503. [doi: 10.1681/ASN.V132497] [Medline: 11805180]
2. Kundhal K, Lok CE. Clinical epidemiology of cardiovascular disease in chronic kidney disease. Nephron Clin Pract 2005;101(2):c47-c52. [doi: 10.1159/000086221] [Medline: 15942250]
3. Willey J, Mentias A, Vaughan-Sarrazin M, McCoy K, Rosenthal G, Girotra S. Epidemiology of lower extremity peripheral artery disease in veterans. J Vasc Surg 2018 Aug;68(2):527-535.e5. [doi: 10.1016/j.jvs.2017.11.083] [Medline: 29588132]
4. Eggers PW, Gohdes D, Pugh J. Nontraumatic lower extremity amputations in the Medicare end-stage renal disease population. Kidney Int 1999 Oct;56(4):1524-1533. [doi: 10.1046/j.1523-1755.1999.00668.x] [Medline: 10504504]
5. Lin CS, Chen SJ, Sung CC, et al. Hemodialysis is associated with increased peripheral artery occlusive disease risk among patients with end-stage renal disease: a nationwide population-based cohort study. Medicine (Baltimore) 2015 Jul;94(28):e1164. [doi: 10.1097/MD.0000000000001164] [Medline: 26181560]
6. Lee CC, Wu CJ, Chou LH, et al. Peripheral artery disease in peritoneal dialysis and hemodialysis patients: single-center retrospective study in Taiwan. BMC Nephrol 2012 Sep 3;13:100. [doi: 10.1186/1471-2369-13-100] [Medline: 22943313]
7. Matsuzawa R, Aoyama N, Yoshida A. Clinical characteristics of patients on hemodialysis with peripheral arterial disease. Angiology 2015 Nov;66(10):911-917. [doi: 10.1177/0003319715572678] [Medline: 25694516]
8. Adragao T, Pires A, Branco P, et al. Ankle--brachial index, vascular calcifications and mortality in dialysis patients. Nephrol Dial Transplant 2012 Jan;27(1):318-325. [doi: 10.1093/ndt/gfr233] [Medline: 21551082]
9. Meloni M, Izzo V, Giurato L, et al. Peripheral arterial disease in patients with renal-diabetic foot ulcers. J Wound Care 2021 Aug 2;30(8):660-664. [doi: 10.12968/jowc.2021.30.8.660] [Medline: 34382847]
10. Antonopoulos CN, Lazaris A, Venermo M, Geroulakos G. Predictors of wound healing following revascularization for chronic limb-threatening ischemia. Vasc Endovascular Surg 2019 Nov;53(8):649-657. [doi: 10.1177/1538574419868863] [Medline: 31405350]
11. Ilo A, Romsi P, Mäkelä J. Infrared thermography as a diagnostic tool for peripheral artery disease. Adv Skin Wound Care 2020 Sep;33(9):482-488. [doi: 10.1097/01.ASW.0000694156.62834.8b] [Medline: 32810061]
12. de Deus Passos M, da Rocha AF. Evaluation of infrared thermography with a portable camera as a diagnostic tool for peripheral arterial disease of the lower limbs compared with color Doppler ultrasonography. Arch Med Sci Atheroscler Dis 2022;7:e66-e72. [doi: 10.5114/amsad/150716] [Medline: 36158072]
13. Bus SA, Aan de Stegge WB, van Baal JG, Busch-Westbroek TE, Nollet F, van Netten JJ. Effectiveness of at-home skin temperature monitoring in reducing the incidence of foot ulcer recurrence in people with diabetes: a multicenter randomized controlled trial (DIATEMP). BMJ Open Diabetes Res Care 2021 Sep;9(1):e002392. [doi: 10.1136/bmjdrc-2021-002392] [Medline: 34493496]
14. Ena J, Carretero-Gomez J, Arevalo-Lorido JC, Sanchez-Ardila C, Zapatero-Gaviria A, Gómez-Huelgas R. The association between elevated foot skin temperature and the incidence of diabetic foot ulcers: a meta-analysis. Int J Low Extrem Wounds 2021 Jun;20(2):111-118. [doi: 10.1177/1534734619897501] [Medline: 32106729]
15. Armstrong DG, Holtz-Neiderer K, Wendel C, Mohler MJ, Kimbriel HR, Lavery LA. Skin temperature monitoring reduces the risk for diabetic foot ulceration in high-risk patients. Am J Med 2007 Dec;120(12):1042-1046. [doi: 10.1016/j.amjmed.2007.06.028] [Medline: 18060924]
16. Lin DSH, Lee JK. Mobile health-based thermometer for monitoring wound healing after endovascular therapy in patients with chronic foot ulcer: prospective cohort study. JMIR mHealth uHealth 2021 May 7;9(5):e26468. [doi: 10.2196/26468] [Medline: 33960955]
17. Antônio de Carvalho Abreu J, de Oliveira RA, Martin AA. Correlation between ankle-brachial index and thermography measurements in patients with peripheral arterial disease. Vascular 2022 Feb;30(1):88-96. [doi: 10.1177/1708538121996573] [Medline: 33752532]
18. Ríos-González BE, López-Barragán L, Saldaña-Cruz AM, et al. Foot sole temperature correlates with ankle-brachial index, pulse wave velocity, and arterial age in diabetic patients without diagnosis of peripheral arterial disease. J Clin Med 2024 Oct 25;13(21):6383. [doi: 10.3390/jcm13216383] [Medline: 39518522]
19. Gatt A, Cassar K, Falzon O, et al. The identification of higher forefoot temperatures associated with peripheral arterial disease in type 2 diabetes mellitus as detected by thermography. Prim Care Diabetes 2018 Aug;12(4):312-318. [doi: 10.1016/j.pcd.2018.01.001] [Medline: 29396205]

XSL•FO
RenderX

20. Hayashi A, Shichiri M. Use of noncontact infrared skin thermometer for peripheral arterial disease screening in patients with and without diabetes. Angiology 2020 Aug;71(7):650-657. [doi: 10.1177/0003319720920162] [Medline: 32319312]
21. Golledge J, Fernando M, Lazzarini P, Najafi B, Armstrong DG. The potential role of sensors, wearables and telehealth in the remote management of diabetes-related foot disease. Sensors (Basel) 2020 Aug 13;20(16):4527. [doi: 10.3390/s20164527] [Medline: 32823514]
22. Ramalho Rosas R, Moraes Silveira FT, Cavalcante Quintela RV, et al. Isolated and grouped relationships between type 2 diabetes and peripheral arterial disease on skin temperature in the foot plantar region. J Therm Biol 2024 Oct;125:103956. [doi: 10.1016/j.jtherbio.2024.103956] [Medline: 39305796]
23. Lavery LA, Higgins KR, Lanctot DR, et al. Preventing diabetic foot ulcer recurrence in high-risk patients: use of temperature monitoring as a self-assessment tool. Diabetes Care 2007 Jan;30(1):14-20. [doi: 10.2337/dc06-1600] [Medline: 17192326]
24. Klein AJ, Hawkins BM. Addressing disparities in chronic limb-threatening ischemia care: what are we waiting for? Vasc Med 2021 Apr;26(2):123-125. [doi: 10.1177/1358863X21992432] [Medline: 33606966]
25. El-Nahas M, El-Shazly S, El-Gamel F, Motawea M, Kyrillos F, Idrees H. Relationship between skin temperature monitoring with smart socks and plantar pressure distribution: a pilot study. J Wound Care 2018 Aug 2;27(8):536-541. [doi: 10.12968/jowc.2018.27.8.536] [Medline: 30086248]

Abbreviations

ABI: ankle-brachial index
ATA: anterior tibial artery
AUC: area under the curve
CAD: coronary artery disease
CLTI: chronic limb-threatening ischemia
DM: diabetes mellitus
DUS: duplex ultrasound
ESRD: end-stage renal disease
EVT: endovascular therapy
GEE: generalized estimating equation
HD: hemodialysis
IRB: Institutional Review Board
mHealth: mobile health
OR: odds ratio
PAD: peripheral artery disease
PTA: posterior tibial artery
ROC: receiver operating characteristic

Edited by J Zhou; submitted 31.Mar.2026; peer-reviewed by C Kapatais, G Piva; revised version received 18.Sep.2026; accepted 18.Sep.2026; published 06.Nov.2026.

Please cite as:
Lin DSH, Wu CC, Yang CW, Lee JK, Hsu CY
Association Between Intradialytic Foot Temperature Gradients and Severe Peripheral Artery Disease in Patients Undergoing Hemodialysis: Prospective Cohort Study
JMIR Mhealth Uhealth 2026;14:e96782
URL: https://mhealth.jmir.org/2026/1/e96782
doi:10.2196/96782

© Donna Shu-Han Lin, Chih-Cheng Wu, Chung-Wei Yang, Jen-Kuang Lee, Chien-Yi Hsu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 6.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobile AI-Assisted Oral Lesion Triage for Community Oral Cancer Screening: Prospective Field Evaluation Study

Hao-Yun Liu[1,2], MD; Yu-Cheng Huang[3], MD; Si-Wei Chen[4], BS; Ling-Ying Wei[5], DDS; Cheng-Ying Chou[4], PhD; Jang-Jaer Lee[5], DDS, PhD; Shyh-Jye Chen[2,3], MD, PhD; Chien-Chang Lee[6], MD, PhD

[1]Department of Medical Imaging, National Taiwan University Hospital Hsin-Chu Branch, Hsin-Chu, Taiwan

[2]College of Medicine, National Taiwan University, Taipei, Taiwan

[3]Department of Medical Imaging, National Taiwan University Hospital, No. 7, Chung-Shan South Road, Taipei, Taiwan

[4]Department of Biomechatronics Engineering, National Taiwan University, Taipei, Taiwan

[5]Department of Dentistry, College of Medicine, National Taiwan University, Taipei, Taiwan

[6]Department of Emergency Medicine, National Taiwan University Hospital, Taipei, Taiwan

Corresponding Author:
Shyh-Jye Chen, MD, PhD
College of Medicine, National Taiwan University, Taipei, Taiwan

Abstract

Background: AI-based tools for oral cancer screening have shown promising performance in curated image datasets, but their performance in real-world mobile community screening workflows remains less clear.

Objective: This study evaluated the prospective field performance of a parameter-locked, mobile AI-assisted oral lesion triage system embedded in a human-supervised community oral cancer screening workflow.

Methods: We conducted a prospective community-based field evaluation in eastern Taiwan from June 17 to December 31, 2025. Trained nonspecialist personnel acquired standard white-light intraoral images using handheld mobile devices during routine screening. The AI system incorporated on-site image quality assessment and lesion-level risk stratification into green, yellow, and red triage categories. Technically inadequate images after repeated acquisition attempts were classified as ungradable and excluded from performance analysis. Three board-certified oral and maxillofacial specialists established an operational clinical reference standard through structured image review and consensus adjudication. The primary outcome was lesion-level identification of high-risk lesions requiring specialist referral, defined as red versus nonred triage. Diagnostic performance was estimated with Wilson 95% CIs; precision-recall performance and preconsensus specialist agreement were also assessed.

Results: Among 602 screened participants, 4283 interpretable intraoral images were included in expert adjudication and lesion-level analysis. The AI system flagged 68 red lesions; 12 (17.6%) were confirmed as high risk, and 4 additional high-risk lesions were identified during expert review, yielding 16 high-risk events for the primary red-triage analysis. For high-risk lesion identification, sensitivity was 75% (12/16; 95% CI 50.5%-89.8%) and positive predictive value was 17.6% (12/68; 95% CI 10.4%-28.4%); specificity, negative predictive value, and accuracy were 98.6% (3928/3984), 99.9% (3928/3932), and 98.5% (3940/4000), respectively. Area under the receiver operating characteristic curve values were 0.957 for red, 0.931 for yellow, and 0.964 for green triage; precision-recall analysis showed area under the precision-recall curve values of 0.8235, 0.9052, and 0.9476, respectively. Preconsensus specialist agreement was high, with a Fleiss κ of 0.921 and a Gwet AC1 of 0.981.

Conclusions: This mobile AI-assisted oral lesion triage system was feasibly integrated into a real-world community screening workflow operated by trained nonspecialist personnel. In this low-prevalence field setting, red triage showed moderate sensitivity and modest positive predictive value, supporting its use as human-supervised referral-prioritization support rather than autonomous diagnosis. Further multicenter studies with blinded assessment, participant-level outcomes, workflow-efficiency measures, and longitudinal follow-up are needed before clinical impact and scalability can be determined.

(JMIR Mhealth Uhealth 2026;14:e101721) doi:10.2196/101721

KEYWORDS

artificial intelligence; AI; oral cancer; mobile health; deep learning; diagnostic imaging; telemedicine

Introduction

Oral cancer (OC) remains a substantial global public health burden and is associated with considerable yet often underrecognized socioeconomic consequences. Major risk factors include tobacco use, excessive alcohol consumption, and betel quid chewing, with a subset of cases associated with human papillomavirus infection [1]. Malignant transformation

XSL•FO
RenderX

commonly develops along a continuum of precursor conditions collectively referred to as oral potentially malignant disorders (OPMDs), which encompass clinical and histopathologic alterations of the oral mucosa and carry variable risks of subsequent progression to cancer [2]. Although surgical resection with or without adjuvant therapy remains the cornerstone of treatment, the overall 5-year survival rate for OC has remained approximately 40% to 50% [3]. These observations underscore the importance of early detection and risk-stratified referral, particularly in community screening settings where subtle oral mucosal lesions may be encountered before specialist evaluation.

In Taiwan, a nationwide OC screening program targeting high-risk populations, including individuals with tobacco, alcohol, or betel quid use, has been implemented for more than two decades, reflecting the recognized clinical and public health value of early detection. Nevertheless, the effectiveness of population-level screening remains challenged by the subtle and heterogeneous appearance of early oral mucosal lesions, as well as variability in lesion recognition among frontline screening personnel. These challenges may be amplified in remote community-based screening settings, where access to specialist oral oncology expertise is limited and screening is often performed under field conditions rather than in specialist clinics. Consequently, misclassification or underrecognition of OPMDs remains a long-standing issue [4]. These workflow-level constraints highlight the need for scalable, mobile-compatible tools that can support image acquisition, lesion risk stratification, and referral prioritization while preserving specialist oversight.

With advances in AI and deep learning, image-based tools for detecting oral mucosal lesions have progressed from proof-of-concept classification models toward clinically oriented screening support. Prior studies have shown that AI systems can identify OPMDs and OC from intraoral images with favorable classification or detection performance across diverse datasets [5-13]. Smartphone-based and mobile digital platforms have also been explored as accessible approaches for OC screening in community or resource-limited settings [14,15]. In addition, recent work has incorporated multimodal imaging, lesion segmentation, and risk-based triage frameworks to improve lesion detection, boundary delineation, and referral decision support [16-19]. However, most existing evidence remains focused on model development, internal validation, retrospective datasets, or clinic-based populations, which may not fully capture deployment-related factors such as image acquisition variability, low-prevalence predictive values, and reference standard uncertainty. Prospective evidence on how mobile AI-assisted triage performs when embedded in real-world community screening workflows, particularly with trained nonspecialist image acquisition and on-site quality assurance and expert oversight, remains limited.

To address these clinical and workflow needs, our group previously developed a deep learning system capable of simultaneously identifying multiple coexisting oral mucosal lesions from standard white-light intraoral images collected through Taiwan's population-based screening program [20]. Unlike approaches centered on single-disease classification, this system was designed to reflect the clinical complexity of routine oral examinations, in which benign lesions, OPMDs, and highly suspicious lesions may coexist within the same individual. In internal validation, the model demonstrated favorable diagnostic performance, supporting further prospective evaluation of its potential role as triage support for community-based screening. However, model performance in curated or development datasets does not necessarily translate into reliable operation in mobile field settings. In community screening workflows, image quality, operator experience, oral cavity exposure, lighting conditions, and lesion prevalence may differ substantially from those observed in controlled datasets. Therefore, prospective evaluation within routine screening workflows is needed to determine whether AI-assisted oral lesion triage can operate reliably under real-world conditions and support referral prioritization without replacing specialist oversight.

Although recent multicenter and clinically oriented studies have reported robust performance of AI systems for OPMD-related detection and classification tasks, much of the existing evidence has been generated from controlled image datasets or clinical populations rather than prospective community screening workflows [21,22]. Direct evidence remains limited regarding how mobile AI-assisted oral lesion triage performs when integrated into real-world field screening, particularly when images are acquired by trained nonspecialist personnel using handheld mobile devices. In this study, we evaluated the prospective field performance of a previously developed, parameter-locked AI-based oral lesion triage system embedded within routine community OC screening. Rather than focusing on model development or diagnostic classification alone, we assessed a human-supervised mobile health workflow that incorporated on-site image quality assurance; green, yellow, and red lesion-level triage; and structured expert adjudication.

Methods

Study Design and Setting

This prospective, community-based field evaluation study assessed a previously developed and parameter-locked AI-assisted oral lesion triage system within routine OC screening workflows. The system was deployed from June 17, 2025, to December 31, 2025, during community screening activities in eastern Taiwan, a region with a relatively high OC burden and limited access to specialist oral oncology services. This study was designed to evaluate workflow-integrated performance rather than model development; no retraining, optimization, or post hoc threshold adjustment was performed during the study period. No formal sample size calculation was performed; the sample size reflected all eligible participants screened during the prospective field deployment period. Screening procedures followed the standard operating protocols of Taiwan's national OC prevention program and were conducted by trained nonspecialist personnel to reflect frontline community practice. Intraoral image acquisition using handheld mobile devices, automated on-site image quality assessment, and AI-generated triage outputs were performed concurrently at screening sites under routine field conditions. Representative images of the

XSL•FO
RenderX

community screening environment are provided in Figure S1 in Multimedia Appendix 1.

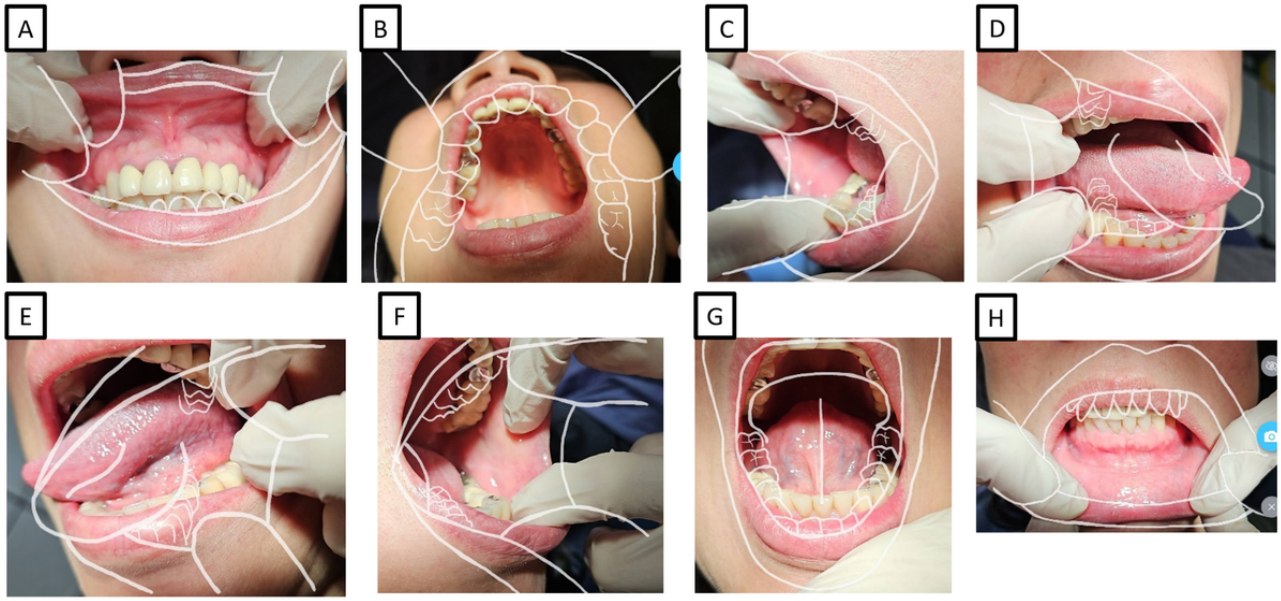
Participants and Recruitment

Community residents aged 18 years or older who voluntarily participated in routine OC screening activities were eligible for inclusion. Participants were recruited from community-based screening events rather than specialist referral clinics, thereby reflecting a frontline screening population with a low expected prevalence of high-risk oral lesions. Exclusion criteria were refusal to undergo intraoral imaging, persistently inadequate image quality despite repeated acquisition attempts, or a history of OC treatment resulting in substantial postsurgical anatomic alteration that precluded standardized mucosal assessment. In accordance with Taiwan's national OC screening protocols, each participant underwent standardized intraoral photographic documentation as part of the screening workflow. Screening-related clinical and imaging data were collected prospectively at the screening sites to reflect real-world community-level heterogeneity. Individual-level demographic variables and behavioral risk factors, including age, sex, tobacco use, alcohol use, and betel quid use, were not collected in the analytic dataset.

Mobile Image Acquisition and On-Site Quality Assurance

Intraoral images were acquired during routine community screening by trained nonspecialist personnel using handheld mobile devices. Each participant underwent standardized photographic documentation of the oral cavity according to Taiwan's national screening protocols, with images obtained from predefined mucosal regions to support comprehensive visual assessment (Figure 1). Image acquisition was performed under real-world field conditions without specialized photographic equipment, thereby preserving variability in illumination, patient positioning, oral opening, and operator technique encountered in frontline community screening practice.

Figure 1. A standard oral photograph sample using augmented reality lines to guide the correct angles and positions for examining different parts of the oral mucosa.



Immediately after acquisition, each image was assessed on-site by the AI-based image quality assurance module. The module evaluated whether the image provided adequate focus sharpness, illumination, mucosal visibility, and absence of motion artifacts for downstream lesion assessment. Images that failed to meet predefined quality criteria triggered immediate repeat acquisition, with up to 3 attempts permitted for each target view. Images that remained inadequate after repeated acquisition attempts were classified as ungradable and were excluded from AI triage output and subsequent performance analysis. This quality assurance step was incorporated into the mobile screening workflow to reduce the impact of technically inadequate images while preserving the practical constraints of community-based image acquisition.

AI Triage System and Field Deployment Workflow

The AI system evaluated in this study was a previously developed oral lesion triage model trained on 8336 clinical intraoral images collected between 2021 and 2024 at a tertiary medical center. Before field deployment, the model architecture, parameters, and triage thresholds were fixed, and no retraining, fine-tuning, or post hoc threshold adjustment was performed during the study period. The system was designed to identify multiple coexisting oral mucosal lesions within standard white-light intraoral images and generate lesion-level bounding box annotations with risk-based triage outputs to support screening workflow decisions.

For field deployment, the system incorporated 2 sequential modules within the mobile screening workflow. The first was an automated image quality assurance module, which assessed whether acquired images were technically adequate for lesion evaluation and triggered immediate repeat acquisition when predefined quality criteria were not met. The second was a multi-lesion detection and risk stratification module, which classified screening findings into 3 clinically actionable triage

XSL•FO
RenderX

categories: green, indicating no actionable lesion or no evident malignant potential; yellow, indicating lesions requiring structured follow-up; and red, indicating lesions requiring specialist referral because of high-risk clinical features. AI-generated outputs were produced on-site as triage support and were not used as autonomous diagnostic decisions. Representative AI-generated outputs for the green, yellow, and red triage categories are provided in Figure S2 in Multimedia Appendix 1. Additional technical details of model development, internal validation, and comparative training outcomes are provided in Table S1 in Multimedia Appendix 1.

Operational Clinical Reference Standard

The reference standard in this study was defined as an operational clinical reference standard established through structured expert adjudication, reflecting decision-making in community-based OC screening rather than definitive histopathologic diagnosis. Three board-certified oral and maxillofacial specialists independently reviewed the complete intraoral image set for each participant. Reviewers assessed the appropriateness of the AI-generated triage classification and evaluated the presence, location, and risk category of clinically relevant lesions across the full image set.

Reviewers were aware of the AI-generated triage outputs, including green, yellow, and red classifications, during the evaluation process. This design reflected the workflow-integrated nature of the field evaluation and, therefore, constituted an unblinded operational reference standard rather than a blinded diagnostic validation reference standard. In addition to assessing AI-flagged findings, reviewers were instructed to comprehensively examine all original images to identify clinically significant lesions that may not have been detected by the AI system, thereby incorporating potential false-negative findings into the reference standard. Initial independent assessments were recorded for all cases and were used to calculate preconsensus interrater agreement among the 3 specialists. When all reviewers agreed, the concordant classification was adopted. For cases with disagreement, a formal adjudication meeting was conducted, and the final consensus decision was established as the operational clinical reference standard.

Histopathologic examination remains the definitive standard for confirming oral malignancy and epithelial dysplasia. However, systematic biopsy of all lesions or all screened participants is not feasible or ethically appropriate in population-level community screening. In routine screening practice, referral and follow-up decisions are commonly based on clinical visual assessment by experienced specialists, with biopsy reserved for lesions considered clinically suspicious or requiring diagnostic confirmation. Furthermore, in the context of multifocal and heterogeneous OPMDs, localized biopsy sampling may not fully represent overall lesion risk because of inherent sampling limitations [23]. Because this study evaluated AI-assisted triage within a screening workflow rather than final disease diagnosis, structured expert consensus was used to determine clinically relevant risk categories for performance evaluation. This study was designed and reported in accordance with the STARD (Standards for Reporting of Diagnostic Accuracy Studies) 2015 reporting guidelines, and the completed checklist is provided in Checklist 1.

Outcome Measures

The primary outcome was the lesion-level performance of the AI system for identifying high-risk oral mucosal lesions requiring specialist referral, corresponding to the red triage category. For the primary analysis, each AI-detected or expert-adjudicated lesion was classified using a red vs nonred framework, and lesion-level 2 × 2 contingency tables were constructed using expert adjudication as the comparator. Given the low expected prevalence of high-risk lesions in community screening, interpretation of the primary red triage outcome emphasized sensitivity and positive predictive value (PPV).

Secondary outcomes included category-specific lesion-level performance for the yellow and green triage categories, image-level agreement between AI outputs and expert adjudication where applicable, and discriminative performance based on AI detection confidence using receiver operating characteristic (ROC) and precision-recall analyses. The 3 triage categories were defined according to clinical actionability and were informed by previously reported gradients of malignant transformation risk in OPMDs [24-28]. In this framework, red indicated lesions requiring specialist referral because of high-risk clinical features; yellow indicated lesions requiring structured follow-up because of low to intermediate risk features; and green indicated no actionable lesion or findings without evident malignant potential.

For category-specific analyses, each triage level was evaluated using a one vs the rest framework, with red compared with nonred categories, yellow compared with nonyellow categories, and green compared with nongreen categories. False-negative cases in the red category were further reviewed descriptively to characterize potential failure patterns relevant to real-world mobile screening deployment.

Statistical Analysis

All analyses were conducted according to the predefined outcome framework. The primary analysis evaluated lesion-level performance of the AI system for identifying high-risk oral mucosal lesions requiring specialist referral using the expert-adjudicated lesion-level classification as the comparator. Secondary analyses included category-specific lesion-level performance for the yellow and green triage categories, as well as image-level agreement between AI outputs and expert adjudication where applicable.

For each analysis, diagnostic performance metrics, including sensitivity, PPV, specificity, negative predictive value (NPV), and overall accuracy, were calculated with corresponding 95% CIs using the Wilson method. Discriminative performance was further assessed using ROC curves and the area under these curves (ROC-AUC) based on AI detection confidence scores for each triage category. Precision-recall curves and the area under these curves (PR-AUC) were also generated to evaluate positive triage performance in the low-prevalence screening setting. Preconsensus interrater agreement among the 3 specialists was summarized using the complete agreement rate, average pairwise observed agreement, Fleiss κ, and Gwet AC1.

XSL•FO
RenderX

All analyses were descriptive in nature, and no formal hypothesis testing was performed. Given the prospective field evaluation setting and the exploratory objective of workflow-integrated assessment, statistical comparisons between groups were not prespecified. Analyses were conducted using the Python software (version 3.11; Python Software Foundation) with standard scientific computing libraries.

Ethical Considerations

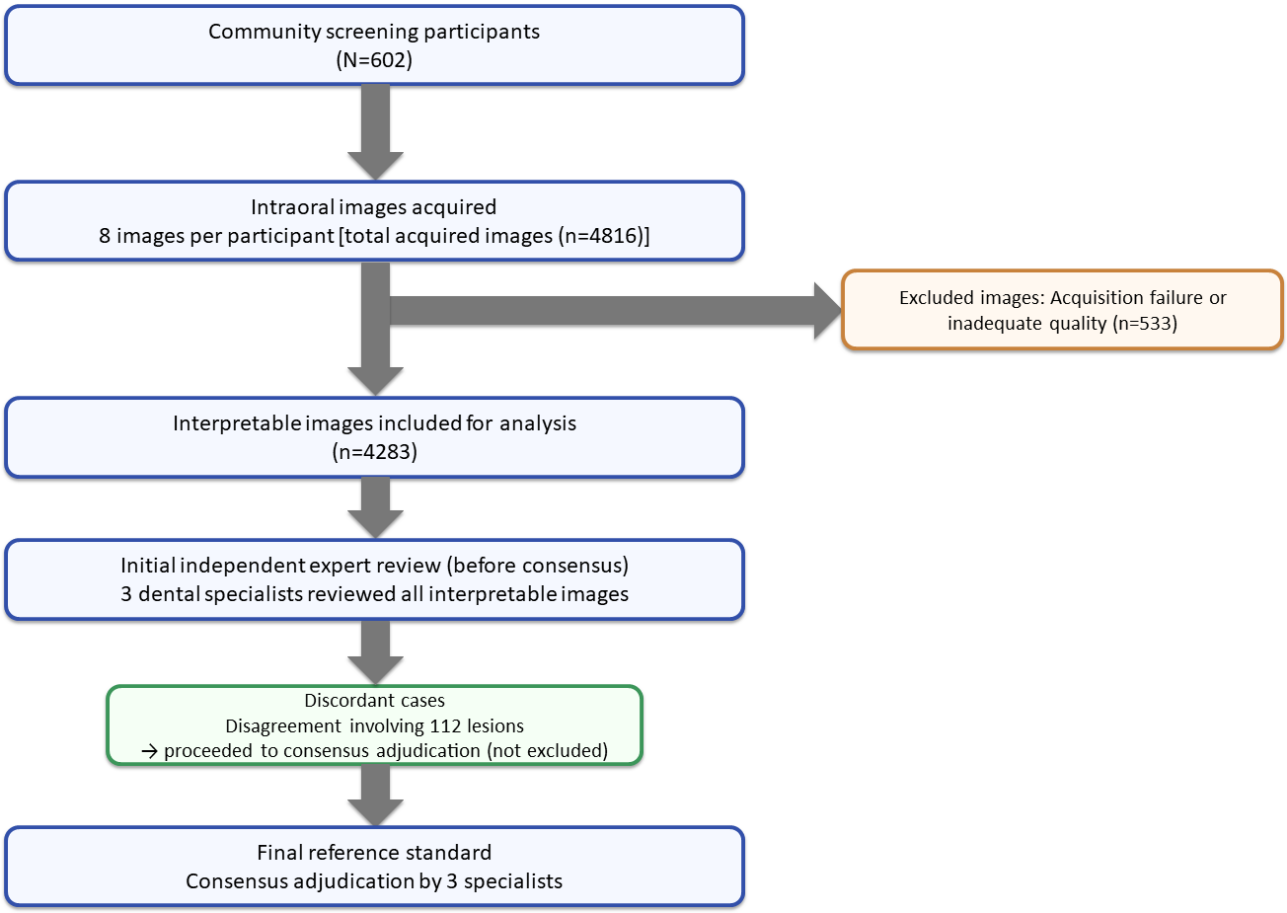
This study was approved by the Research Ethics Committee of National Taiwan University Hospital (approval number NTUH 202504075 RIND). All procedures were conducted in accordance with local regulations and the principles of the Declaration of Helsinki. Screening-related clinical and imaging data were deidentified before analysis and stored in secure institutional systems. Participation was voluntary, and informed consent was obtained according to the approved study protocol. No financial compensation was provided to participants.

Results

Participant and Image Flow

Between June 2025 and December 2025, a total of 602 participants underwent community-based OC screening. In accordance with the standardized screening protocol, multiple intraoral images were acquired for each participant, with a mean of approximately 8 images per participant. Of 4816 acquired intraoral images, 533 (11.1%) were excluded because of acquisition failure or inadequate quality, leaving 4283 (88.9%) interpretable images for expert adjudication and lesion-level performance analysis. The participant and image flow is summarized in Figure 2.

Figure 2. STARD (Standards for Reporting of Diagnostic Accuracy Studies)–consistent participant and image flow diagram.



Expert Adjudication and Final Analysis Set

All interpretable images were initially reviewed independently by 3 board-certified oral and maxillofacial specialists without discussion. During this initial review, cases for which all 3 specialists reached the same initial classification were summarized as a preconsensus process measure, as shown in Table 1. Preconsensus lesion-level agreement among the 3 specialists was high, with a complete agreement rate of 97.5%, a Fleiss κ of 0.921, and a Gwet AC1 of 0.981. This preconsensus summary was used to describe the initial concordant expert review process and was not used as the primary outcome analysis. Among these concordant initial assessments, 91% (3898/4283) of the images were classified as showing no detectable lesion.

XSL•FO
RenderX

Table . Lesion-level performance based on concordant initial independent expert assessments before consensus adjudication.

	TPs[a], n	FPs[b], n	FNs[c], n	TNs[d], n	Sensitivity (%; 95% CI)	Specificity (%; 95% CI)	PPV[e] (%; 95% CI)	NPV[f] (%; 95% CI)	Accuracy (%; 95% CI)
Red	10	12	4	3898	71.4 (45.4-88.3)	99.7 (99.5-99.8)	45.5 (26.9-65.3)	99.9 (99.7-100)	99.6 (99.3-99.7)
Yellow	240	16	9	3898	96.4 (94.0-98.1)	99.6 (99.3-99.7)	93.8 (90.1-96.2)	99.8 (99.6-99.9)	99.4 (99.2-99.6)
Green	206	4	16	3898	92.8 (88.6-95.5)	99.9 (99.7-100)	98.1 (95.2-99.3)	99.6 (99.3-99.7)	99.5 (99.3-99.7)

[a]TP: true positive.

[b]FP: false positive.

[c]FN: false negative.

[d]TN: true negative.

[e]PPV: positive predictive value.

[f]NPV: negative predictive value.

Disagreement among the 3 specialists was identified for 112 lesions, indicating that a subset of oral mucosal findings remained clinically equivocal during independent expert assessment. These discordant lesions subsequently underwent formal consensus adjudication to establish the final operational clinical reference standard. After consensus adjudication, all interpretable images, including those with initial discordance, were incorporated into the final lesion-level analysis. In the final adjudicated dataset, 91.7% (3928/4283) of the images were confirmed as not containing detectable lesions, and the final adjudicated performance is shown in Table 2.

Table . Final adjudicated lesion-level performance after expert consensus.

	TPs[a], n	FPs[b], n	FNs[c], n	TNs[d], n	Sensitivity (%; 95% CI)	Specificity (%; 95% CI)	PPV[e] (%; 95% CI)	NPV[f] (%; 95% CI)	Accuracy (%; 95% CI)
Red	12	56	4	3928	75.0 (50.5-89.8)	98.6 (98.2-98.9)	17.6 (10.4-28.4)	99.9 (99.7-100)	98.5 (98.1-98.8)
Yellow	281	64	22	3928	92.7 (89.3-95.2)	98.4 (98.0-98.7)	81.4 (77.0-85.2)	99.4 (99.2-99.6)	98.0 (97.5-98.4)
Green	221	13	48	3928	82.2 (77.1-86.3)	99.7 (99.4-99.8)	94.4 (90.7-96.7)	98.8 (98.4-99.1)	98.6 (98.1-98.9)

[a]TP: true positive.

[b]FP: false positive.

[c]FN: false negative.

[d]TN: true negative.

[e]PPV: positive predictive value.

[f]NPV: negative predictive value.

Primary Outcome: High-Risk Red Triage Categorization Performance

For the primary lesion-level analysis of high-risk lesions corresponding to the red triage category, the AI system flagged 68 lesions during screening. Following expert consensus adjudication, 17.6% (n=12) of these lesions were confirmed as true high-risk lesions. In addition, 4 lesions not initially classified as red by the AI system, including 2 lesions not detected by the model and 2 lesions classified as yellow or green, were adjudicated as high risk, yielding 16 high-risk lesions for the primary red triage category analysis.

In the final adjudicated analysis, sensitivity for high-risk lesion identification was 75.0% (95% CI 50.5%-89.8%), and PPV was 17.6% (95% CI 10.4%-28.4%). Specificity, NPV, and overall accuracy were 98.6% (95% CI 98.2%-98.9%), 99.9% (95% CI 99.7%-100%), and 98.5% (95% CI 98.1%-98.8%), respectively (Table 2). In operational terms, the red triage category generated 68 flags among the 4283 interpretable images, corresponding to 15.9 red flags, 2.8 true-positive red flags, and 13.1 false-positive red flags per 1000 interpretable images.

Secondary Outcomes: Yellow and Green Triage Categorization Performance

In secondary lesion-level analyses based on the final adjudicated reference standard, category-specific performance was evaluated for the yellow and green triage categories. For the yellow category, sensitivity was 92.7% (95% CI 89.3%-95.2%), and PPV was 81.4% (95% CI 77.0%-85.2%); specificity, NPV, and overall accuracy were 98.4% (95% CI 98.0%-98.7%), 99.4% (95% CI 99.2%-99.6%), and 98.0% (95% CI 97.5%-98.4%), respectively. For the green category, sensitivity was 82.2% (95% CI 77.1%-86.3%), and PPV was 94.4% (95% CI 90.7%-96.7%); specificity, NPV, and overall accuracy were

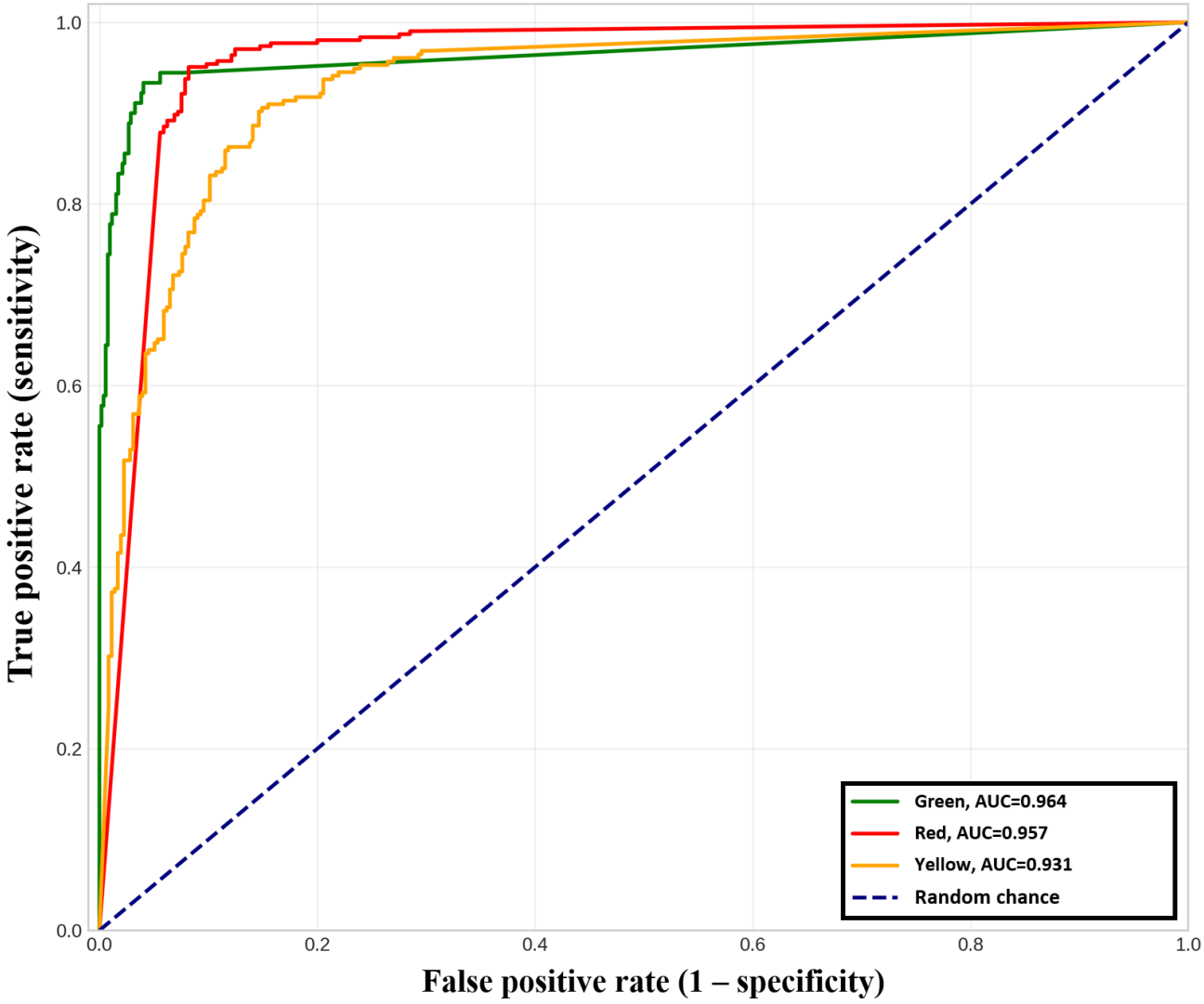
XSL•FO
RenderX

99.7% (95% CI 99.4%-99.8%), 98.8% (95% CI 98.4%-99.1%), and 98.6% (95% CI 98.1%-98.9%), respectively (Table 2).

Discriminative Performance Based on Detection Confidence

To evaluate discriminative performance across lesion risk categories, ROC curves were constructed using AI detection confidence scores. The expert-adjudicated classification served as the operational clinical reference standard for ROC analysis. For each triage category, a one vs the rest analysis was performed, with the target category treated as positive and all other categories combined as negative. ROC-AUC values were 0.957 for red, 0.931 for yellow, and 0.964 for green (Figure 3). Precision-recall analysis yielded PR-AUC values of 0.8235 for red, 0.9052 for yellow, and 0.9476 for green. The corresponding precision-recall curves are shown in Figure S3 in Multimedia Appendix 1.

Figure 3. Receiver operating characteristic curves for AI-based oral lesion risk stratification showing the discriminative performance of the AI system for green, yellow, and red risk categories using the expert-adjudicated operational clinical reference standard. The diagonal dashed line represents random classification. AUC: area under the curve.



Discussion

Principal Findings

In this prospective community-based field evaluation study, we assessed a previously developed, parameter-locked, mobile AI-assisted oral lesion triage system within a routine community OC screening workflow. Under field conditions, including image acquisition by trained nonspecialist personnel using handheld mobile devices, on-site image quality assurance, AI-assisted lesion-level risk stratification, and expert adjudication, the system was feasibly integrated into the screening workflow and provided preliminary lesion-level performance estimates for the high-risk red category. The primary red triage category analysis was based on 16 high-risk lesions and showed moderate sensitivity with modest PPV, reflecting the low-prevalence structure of community screening. Unlike model development or retrospective validation studies, this study evaluated how an AI-assisted triage system functioned when embedded in a real-world mobile screening workflow, where image quality, operator experience, lesion distribution, and expert interpretation reflect frontline community practice. These findings suggest that mobile AI-assisted oral lesion triage may help bridge the gap between algorithmic validation and practical field evaluation. The system should therefore be understood not as an autonomous diagnostic tool but as human-supervised triage

XSL•FO
RenderX

support within a digital health workflow designed to assist risk stratification and referral prioritization while preserving expert oversight.

Interpretation in the Context of Community Screening

Community-based OC screening differs fundamentally from the controlled image datasets commonly used in AI research. Prospective field evaluation can therefore reveal deployment-related constraints, including acquisition variability, low-prevalence predictive values, and reference standard uncertainty, that may not be fully captured in model development or retrospective validation studies [29,30]. More broadly, recent clinical AI evaluation commentaries have similarly emphasized that aggregate metrics or direct output-only evaluation may be insufficient to assess workflow-level reliability, interpretability, and deployment safety [31,32]. In routine mobile screening workflows, intraoral images are acquired on-site by trained nonspecialist personnel under variable lighting, positioning, oral opening, and operator conditions. In addition, the prevalence of high-risk lesions within community screening populations is low, resulting in a predominance of normal or nonspecific findings and a markedly imbalanced risk distribution. In this study, despite high overall preconsensus interrater agreement, disagreement among the 3 independent specialists was observed for 112 lesions before consensus adjudication, underscoring that a subset of oral mucosal findings remained clinically equivocal even under expert review. The contrast between the preconsensus summary and the final adjudicated analysis illustrates the clinical uncertainty inherent in field screening and supports the need for structured expert adjudication when evaluating AI-assisted triage systems. Within such workflows, image quality assurance is a central deployment requirement because reliable risk stratification depends not only on model performance but also on the quality and consistency of field-acquired images.

Interpretation of Predictive Values in Low-Prevalence Screening

The high-risk red triage category showed moderate sensitivity and modest PPV despite high specificity and NPV. This finding should be interpreted within the epidemiologic structure of community-based OC screening. Population-based studies from Taiwan have reported that the prevalence of highly suspicious or referral-requiring oral lesions in community screening settings is approximately 0.9% to 2.2% [26-28]. In such low-prevalence settings, even screening tools with high sensitivity and specificity may yield constrained PPVs because predictive values are strongly influenced by baseline disease prevalence [33,34]. Therefore, the PPV observed for the red triage category should not be interpreted as a stand-alone diagnostic measure. It should also not be interpreted as an operational measure of positive triage yield and expert review burden. Rather, it should be considered in relation to the intended role of AI-assisted triage: prioritizing a smaller subset of lesions for further specialist evaluation while preserving human oversight in a predominantly low-risk screening population [35-37].

Contextual Risk Enrichment and Quality Management Relevance

As an exploratory contextual analysis, PPV lift was used to describe the degree of risk enrichment achieved by the red triage category relative to the expected baseline event rate in the screened population. Using reported baseline prevalence estimates for high-risk oral lesions in Taiwanese community screening populations and the observed red category sensitivity and specificity, the estimated PPV lift suggested that AI-assisted triage may concentrate high-risk findings within the red category [38-40]. PPV has also been used as a quality management indicator in Taiwan's oral mucosal screening program, although such benchmarks were developed for program-level screening evaluation rather than lesion-level AI triage assessment [41,42]. Although direct comparison is limited by differences in outcome definitions, analysis level, and reference standards, the lesion-level PPV of 17.6% observed for AI red triage may be viewed as a contextual signal of risk enrichment under low-prevalence community screening conditions. Because the present analysis did not directly measure participant-level referral yield, specialist workload, cost-effectiveness, or downstream clinical outcomes, this finding should be interpreted as a potential prioritization signal within a supervised triage workflow rather than direct evidence of clinical benefit.

Human-Supervised Triage and Expert Oversight

The structured expert adjudication process also provides insights into how AI outputs were incorporated into the screening workflow. Prior studies have cautioned that AI-driven decision support may introduce automation bias or anchoring effects if clinicians overrely on model outputs, particularly when explainability alone is insufficient to ensure appropriate clinical interpretation [43]. In the present study, the presence of 112 discordant lesions before consensus adjudication and the identification of 4 AI-missed high-risk lesions during expert review suggest that expert review included active assessment beyond AI-flagged outputs; however, because reviewers were aware of AI outputs, the process remains susceptible to AI-related review or incorporation bias. Instead, discordant or equivocal findings were subjected to formal consensus adjudication, and additional high-risk lesions not classified as red by the AI system were incorporated into the final operational clinical reference standard. This process helped maintain a practical distinction between AI-generated triage suggestions and expert-derived clinical reference decisions within the workflow. These findings are consistent with broader recommendations that medical AI systems should support rather than replace clinical judgment, with appropriate oversight and accountability maintained throughout deployment [44-46].

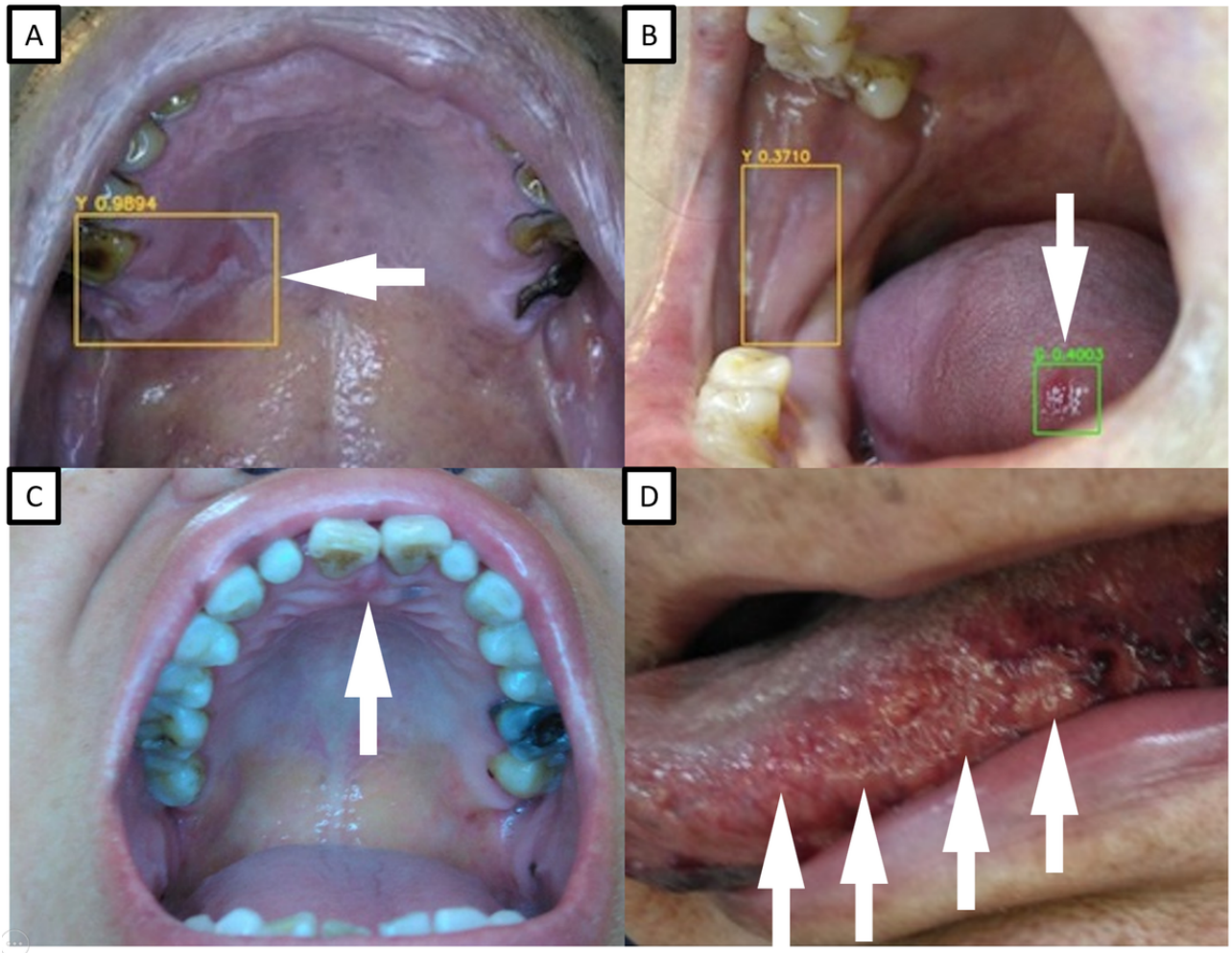
Failure Patterns and Field Deployment Implications

The 4 lesions adjudicated as high risk but not initially classified as red by the AI system provide clinically informative examples of deployment-relevant limitations under mobile field screening conditions. These cases included lesions with borderline visual characteristics, such as mixed erythematous and leukoplakic appearances, lesions located near image margins, and lesions with limited local contrast or incomplete visibility. These findings suggest that subtle morphology, partial lesion capture,

XSL•FO
RenderX

peripheral lesion location, and low contrast may reduce model sensitivity for high-risk triage in field-acquired images. From a workflow perspective, these failure patterns underscore the importance of maintaining human oversight, comprehensive image review, and structured expert adjudication in AI-assisted community screening, particularly for lesions that fall along a continuum of clinical risk. Representative examples are shown in Figure 4.

Figure 4. Representative false-negative cases in the high-risk (red) triage category during AI-based oral cancer screening (n=4): (A) a borderline lesion with mixed erythematous and leukoplakic appearance that was initially classified by the AI system as low risk (yellow) and subsequently reclassified as high risk (red) following expert consensus; (B) a lesion (green) located near the image margin that was not flagged as high risk by the AI system, where proximity to the image boundary may have influenced model detection; and (C and D) lesions not identified by the AI system that were subsequently recognized by experts upon case-by-case image review at the locations indicated by the white arrows and classified as high risk.



Practical Implications for Community Screening

The findings of this study support the potential value of AI-assisted oral lesion triage in remote community screening settings, where access to specialist oral oncology expertise may be limited. By combining on-site image quality assurance, risk-stratified AI outputs, and expert adjudication, this mobile screening workflow may support referral prioritization under specialist oversight rather than directly establishing improved diagnostic performance by frontline personnel. Because the fixed red triage operating point yielded modest PPV and measurable false positive review burden, future implementation studies should evaluate whether alternative operating points could better balance sensitivity, PPV, and specialist workload. Within Taiwan's existing national OC screening program, such systems could serve as adjunctive tools for referral prioritization rather than replacements for clinical judgment. This approach may be particularly relevant for high-risk or geographically underserved populations, where scalable triage support could help bridge gaps between community screening and specialist evaluation. Further studies are needed to determine whether mobile AI-assisted triage can improve referral efficiency, specialist workload, cost-effectiveness, and downstream clinical outcomes.

Limitations

Several limitations warrant consideration. First, the primary performance analysis was conducted at the lesion level rather than as a participant-level clinical outcome analysis. Although this approach was aligned with the lesion-level AI outputs and expert adjudication used in the present workflow, future studies should evaluate participant-level referral decisions, referral completion, and downstream clinical outcomes. Second, histopathologic confirmation was not systematically obtained for all lesions or screened participants. Expert consensus therefore served as an operational clinical reference standard

XSL•FO
RenderX

for triage evaluation rather than a definitive diagnostic gold standard. Third, expert adjudication was not blinded to AI outputs, reflecting the workflow-integrated design of this study but introducing potential AI-related review or incorporation bias. This unblinded operational reference standard may have influenced expert classification and may have overestimated agreement between AI triage and expert adjudication. Fourth, the number of high-risk lesions was limited, with only 16 high-risk lesions in the primary red triage category analysis. This was consistent with the low-prevalence structure of community screening but constrained the precision of sensitivity and PPV estimates. Fifth, performance analysis was restricted to interpretable images. Images that remained technically inadequate after repeated acquisition attempts were excluded, which may have introduced selection bias and may underestimate the practical burden of image acquisition in field screening. Sixth, this was a single-region field evaluation, and mobile device characteristics, operator training, acquisition conditions, local screening infrastructure, and population risk profiles may affect generalizability to other settings. Individual-level demographic variables and behavioral risk factors were not collected in the analytic dataset, precluding subgroup analyses by age, sex, tobacco use, alcohol use, or betel quid use. Finally, this study did not include a usual care comparator arm; direct evaluation of nonspecialist diagnostic performance; or formal measurement of acquisition time or repeat acquisition rate, specialist workload, cost-effectiveness, or longitudinal outcome linkage. Future multicenter studies incorporating blinded assessment, workflow efficiency metrics, registry linkage, and longitudinal follow-up will be needed to evaluate generalizability, clinical impact, and scalability.

Conclusions

In this prospective community-based field evaluation, a parameter-locked mobile AI-assisted oral lesion triage system was integrated into a routine OC screening workflow operated by trained nonspecialist personnel. Under field conditions characterized by mobile image acquisition, low prevalence of high-risk lesions, and variability in image quality and lesion presentation, the system provided preliminary lesion-level evidence of red triage categorization performance, with moderate sensitivity and modest PPV for high-risk lesion identification. These findings suggest that mobile AI-assisted triage may support referral prioritization within a human-supervised community screening workflow while preserving the need for expert oversight. Further multicenter studies with blinded assessment, participant-level outcome evaluation, workflow efficiency measures, and longitudinal follow-up are needed to determine the clinical impact and scalability of AI-assisted OC screening in real-world practice.

Acknowledgments

The authors thank the community screening personnel and participating residents for their contributions to this field evaluation. They also thank the oral and maxillofacial specialists involved in expert image review and consensus adjudication. Generative AI tools were used to assist with language editing and drafting during manuscript revision. All scientific content, analyses, interpretations, and final manuscript text were reviewed, verified, and approved by the authors. No generative AI tool was listed as an author.

Funding

This study was supported by an investigator-initiated clinical research program of National Taiwan University Hospital: “Application and validation of a mobile device–based deep learning system for assisting the screening of oral potentially malignant lesions and oral cancer” (NTUH 202504075 RIND). The funder had no role in the study design, data collection, data analysis, data interpretation, manuscript preparation, or the decision to submit the manuscript for publication.

Data Availability

The datasets generated and analyzed during this study are not publicly available because they contain clinical images and screening-related information that may compromise participant privacy and are subject to institutional data governance restrictions. Deidentified data may be available from the corresponding author on reasonable request, subject to institutional approval and applicable data-use agreements.

Authors' Contributions

Conceptualization: HYL, YCH, SJC, CCL
Data curation: HYL, YCH, SWC, CYC
Formal analysis: HYL, YCH, SWC
Funding acquisition: SJC
Investigation: HYL, YCH, LYW, JJL
Methodology: HYL, YCH, SWC, CYC, SJC
Software: SWC, CYC
Supervision: SJC, CCL
Validation: HYL, YCH, LYW, JJL
Writing—original draft: HYL

XSL•FO
RenderX

Writing—review and editing: all authors
CYC and SJC served as co-corresponding authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary methods, model training outcomes, representative community screening environment, representative AI-generated triage outputs, and precision-recall curves.
[DOCX File, 3592 KB - mhealth_v14i1e101721_app1.docx]

Checklist 1
STARD 2015 checklist.
[DOCX File, 35 KB - mhealth_v14i1e101721_app2.docx]

References

1. Mehrtash H, Duncan K, Parascandola M, et al. Defining a global research and policy agenda for betel quid and areca nut. Lancet Oncol 2017 Dec;18(12):e767-e775. [doi: 10.1016/S1470-2045(17)30460-6] [Medline: 29208442]
2. Warnakulasuriya S, Kujan O, Aguirre-Urizar JM, et al. Oral potentially malignant disorders: a consensus report from an international seminar on nomenclature and classification, convened by the WHO Collaborating Centre for Oral Cancer. Oral Dis 2021 Nov;27(8):1862-1880. [doi: 10.1111/odi.13704] [Medline: 33128420]
3. Zeng H, Chen W, Zheng R, et al. Changing cancer survival in China during 2003-15: a pooled analysis of 17 population-based cancer registries. Lancet Glob Health 2018 May;6(5):e555-e567. [doi: 10.1016/S2214-109X(18)30127-X] [Medline: 29653628]
4. Liao DZ, Schlecht NF, Rosenblatt G, et al. Association of delayed time to treatment initiation with overall survival and recurrence among patients with head and neck squamous cell carcinoma in an underserved urban population. JAMA Otolaryngol Head Neck Surg 2019 Nov 1;145(11):1001-1009. [doi: 10.1001/jamaoto.2019.2414] [Medline: 31513264]
5. Thapa P, Singh V, Bhatt S, et al. Multimodal fluorescence imaging and spectroscopic techniques for oral cancer screening: a real-time approach. Methods Appl Fluoresc 2023 Sep 13;11(4). [doi: 10.1088/2050-6120/acf6ac] [Medline: 37666247]
6. Kouketsu A, Doi C, Tanaka H, et al. Detection of oral cancer and oral potentially malignant disorders using artificial intelligence-based image analysis. Head Neck 2024 Sep;46(9):2253-2260. [doi: 10.1002/hed.27843] [Medline: 38860703]
7. Lin H, Chen H, Weng L, Shao J, Lin J. Automatic detection of oral cancer in smartphone-based images using deep learning for early diagnosis. J Biomed Opt 2021 Aug;26(8):086007. [doi: 10.1117/1.JBO.26.8.086007] [Medline: 34453419]
8. Tanriver G, Soluk Tekkesin M, Ergen O. Automated detection and classification of oral lesions using deep learning to detect oral potentially malignant disorders. Cancers (Basel) 2021 Jun 2;13(11):2766. [doi: 10.3390/cancers13112766] [Medline: 34199471]
9. Warin K, Limprasert W, Suebnukarn S, Jinaporntham S, Jantana P, Vicharueang S. AI-based analysis of oral lesions using novel deep convolutional neural networks for early detection of oral cancer. PLoS One 2022;17(8):e0273508. [doi: 10.1371/journal.pone.0273508] [Medline: 36001628]
10. Li L, Pu C, Tao J, et al. Development of an oral cancer detection system through deep learning. BMC Oral Health 2024 Dec 4;24(1):1468. [doi: 10.1186/s12903-024-05195-5] [Medline: 39633342]
11. Keser G, Pekiner FN, Bayrakdar İ, Çelik Ö, Orhan K. A deep learning approach to detection of oral cancer lesions from intra oral patient images: a preliminary retrospective study. J Stomatol Oral Maxillofac Surg 2024 Oct;125(12 Suppl 2):101975. [doi: 10.1016/j.jormas.2024.101975] [Medline: 39043293]
12. Chen R, Wang Q, Huang X. Intelligent deep learning supports biomedical image detection and classification of oral cancer. Technol Health Care 2024;32(S1):465-475. [doi: 10.3233/THC-248041] [Medline: 38759069]
13. Schmidl B, Hütten T, Pigorsch S, et al. Artificial intelligence for image recognition in diagnosing oral and oropharyngeal cancer and leukoplakia. Sci Rep 2025 Jan 29;15(1):3625. [doi: 10.1038/s41598-025-85920-4] [Medline: 39880876]
14. Wongsapai M, Wudtijureepun K, Tamdee D, Suthachai T, Wuttisarnwattana P, Suebnukarn S. Oral cancer surveillance in remote areas using a smartphone digital platform. Stud Health Technol Inform 2025 Aug 7;329:1898-1899. [doi: 10.3233/SHTI251270] [Medline: 40776287]
15. Prabhune AG, Srihari VR, Shree S, Katiyar M, Thampi V. Development and validation of an AI-based application for early detection and risk stratification of oral potentially malignant disorders. J Oral Biol Craniofac Res 2025;15(6):1806-1812. [doi: 10.1016/j.jobcr.2025.10.017] [Medline: 41234509]
16. Horvath D, Fekete A, Martinekova P, et al. Artificial intelligence helps diagnose oral potentially malignant disorders: a systematic review and meta-analysis. JDR Clin Trans Res 2026 Jan 21:23800844251403962. [doi: 10.1177/23800844251403962] [Medline: 41562124]

XSL•FO
RenderX

17. Nemichand, Shukla S, Singha Deo B, et al. Smartphone-based bimodal device (SBBD) for oral precancer diagnosis and biopsy guidance in clinical settings. Opt Lett 2025 Mar 15;50(6):1993-1996. [doi: 10.1364/OL.555254] [Medline: 40085611]
18. Thakuria T, Mahanta LB, Khataniar SK, Goswami RD, Baruah N, Bharali T. Smartphone-based oral lesion image segmentation using deep learning. J Imaging Inform Med 2025 Dec;38(6):4358-4376. [doi: 10.1007/s10278-025-01455-0] [Medline: 40032764]
19. Mitbander R, Brenes D, Coole JB, et al. Development and evaluation of an automated multimodal mobile detection of oral cancer imaging system to aid in risk-based management of oral mucosal lesions. Cancer Prev Res (Phila) 2025 Apr 1;18(4):197-207. [doi: 10.1158/1940-6207.CAPR-24-0253] [Medline: 39817650]
20. Hsu Y, Chou CY, Huang YC, et al. Oral mucosal lesions triage via YOLOv7 models. J Formos Med Assoc 2025 Jul;124(7):621-627. [doi: 10.1016/j.jfma.2024.07.010] [Medline: 39003230]
21. Araújo AL, da Silva AV, Gonçalves AR, et al. Two-step pipeline for oral diseases detection and classification: a deep learning approach. Front Oral Health 2025;6:1659323. [doi: 10.3389/froh.2025.1659323] [Medline: 41221302]
22. Cao YQ, Zhang JY, Lu M, et al. A deep learning method for diagnosis of oral potentially malignant disorders. J Dent 2026 Apr;167:106138. [doi: 10.1016/j.jdent.2025.106138] [Medline: 41016582]
23. Essat M, Cooper K, Bessey A, Clowes M, Chilcott JB, Hunter KD. Diagnostic accuracy of conventional oral examination for detecting oral cavity cancer and potentially malignant disorders in patients with clinically evident oral lesions: systematic review and meta-analysis. Head Neck 2022 Apr;44(4):998-1013. [doi: 10.1002/hed.26992] [Medline: 35092324]
24. Speight PM, Khurram SA, Kujan O. Oral potentially malignant disorders: risk of progression to malignancy. Oral Surg Oral Med Oral Pathol Oral Radiol 2018 Jun;125(6):612-627. [doi: 10.1016/j.oooo.2017.12.011] [Medline: 29396319]
25. Iocca O, Sollecito TP, Alawi F, et al. Potentially malignant disorders of the oral cavity and oral dysplasia: a systematic review and meta-analysis of malignant transformation rate by subtype. Head Neck 2020 Mar;42(3):539-555. [doi: 10.1002/hed.26006] [Medline: 31803979]
26. Chiang WF, Liu SY, Lin JF, et al. Malignant development in patients with oral potentially malignant disorders detected through nationwide screening: outcomes of 5-year follow-up at a single hospital. Head Neck 2020 Jan;42(1):67-76. [doi: 10.1002/hed.25973] [Medline: 31589002]
27. Chiu SF, Ho CH, Chen YC, et al. Malignant transformation of oral potentially malignant disorders in Taiwan: an observational nationwide population database study. Medicine (Baltimore) 2021 Mar 5;100(9):e24934. [doi: 10.1097/MD.0000000000024934] [Medline: 33655959]
28. Chen SH, Hsiao SY, Chang KY, Chang JY. New insights into oral squamous cell carcinoma: from clinical aspects to molecular tumorigenesis. Int J Mol Sci 2021 Feb 24;22(5):2252. [doi: 10.3390/ijms22052252] [Medline: 33668218]
29. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. Clin Chem 2015 Dec;61(12):1446-1452. [doi: 10.1373/clinchem.2015.246280]
30. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open 2016 Nov 14;6(11):e012799. [doi: 10.1136/bmjopen-2016-012799] [Medline: 28137831]
31. Hu Y. Toward retrieval-grounded evaluation for conversational large language model-based risk assessment. JMIR AI 2026 Mar 12;5:e90759. [doi: 10.2196/90759] [Medline: 41818631]
32. Hu Y. Comment on “Classifying the clinical significance of common breast pain symptoms using a large language model, ChatGPT (GPT-4)”. Clin Imaging 2026 Apr;132:110741. [doi: 10.1016/j.clinimag.2026.110741] [Medline: 41671902]
33. Maxim LD, Niebo R, Utell MJ. Screening tests: a review with examples. Inhal Toxicol 2014 Nov;26(13):811-828. [doi: 10.3109/08958378.2014.955932] [Medline: 25264934]
34. Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. Front Public Health 2017;5:307. [doi: 10.3389/fpubh.2017.00307] [Medline: 29209603]
35. Webb MP, Sidebotham D. Bayes’ formula: a powerful but counterintuitive tool for medical decision-making. BJA Educ 2020 Jun;20(6):208-213. [doi: 10.1016/j.bjae.2020.03.002] [Medline: 33456952]
36. Williams C, Sahagún BE, Su MK. Biostatistics and epidemiology principles for the toxicologist: the “testy” test characteristics part II: positive predictive value and negative predictive value. J Med Toxicol 2023 Jul;19(3):298-302. [doi: 10.1007/s13181-023-00953-3] [Medline: 37314615]
37. Ben-Haim Y, Dacso CC. Interpreting PPV and NPV of diagnostic tests with uncertain prevalence. Rambam Maimonides Med J 2024 Jul 30;15(3):e0013. [doi: 10.5041/RMMJ.10527] [Medline: 39088705]
38. Isakov O, Riesel D, Leshchinsky M, et al. Development and validation of a colorectal cancer prediction model: a nationwide cohort-based study. Dig Dis Sci 2024 Jul;69(7):2611-2620. [doi: 10.1007/s10620-024-08427-4] [Medline: 38662163]
39. Lau K, Hart GR, Deng J. Predicting time-to-first cancer diagnosis across multiple cancer types. Sci Rep 2025 Jul 9;15(1):24790. [doi: 10.1038/s41598-025-08790-w] [Medline: 40634390]
40. Katki HA. Quantifying risk stratification provided by diagnostic tests and risk predictions: comparison to AUC and decision curve analysis. Stat Med 2019 Jul 20;38(16):2943-2955. [doi: 10.1002/sim.8163] [Medline: 31037749]
41. Principles for quality management of oral mucosal examination in preventive oral health services [Report in Chinese]. Ministry of Health and Welfare, Taiwan. URL: https://www.mohw.gov.tw/dl-95288-76196c53-62cc-4831-835d-bd3dddf4375e.html [accessed 2026-05-05]

XSL•FO
RenderX

42. 2017 hospital cancer screening quality improvement programme: tender briefing [106年醫院癌症篩檢品質精進計畫招標說明]. Health Promotion Administration, Ministry of Health and Welfare, Taiwan. 2017. URL: https://pfp.hpa.gov.tw/News/4d4859a6-b440-41c6-9ad3-8268a026e482/ [illegible] 13%E5%AD%90%E6%8A%B9%E9%A6%96%E7%AF%A9%E5%B9%B4%E9%BD%A1.pdf [accessed 2026-05-05]
43. Khera R, Simon MA, Ross JS. Automation bias and assistive AI: risk of harm from AI-driven clinical decision support. JAMA 2023 Dec 19;330(23):2255-2257. [doi: 10.1001/jama.2023.22557] [Medline: 38112824]
44. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019 Oct 29;17(1):195. [doi: 10.1186/s12916-019-1426-2] [Medline: 31665002]
45. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
46. Saldivia-Siracusa C, Carlos de Souza ES, Barros da Silva AV, et al. Automated classification of oral potentially malignant disorders and oral squamous cell carcinoma using a convolutional neural network framework: a cross-sectional study. Lancet Reg Health Am 2025;47:101138. [doi: 10.1016/j.lana.2025.101138] [Medline: 40519355]

Abbreviations

NPV: negative predictive value
OC: oral cancer
OPMD: oral potentially malignant disorder
PPV: positive predictive value
PR-AUC: area under the precision-recall curve
ROC: receiver operating characteristic
ROC-AUC: area under the receiver operating characteristic curve
STARD: Standards for Reporting of Diagnostic Accuracy Studies

Edited by L Buis; submitted 19.May.2026; peer-reviewed by S Zhuang, Y Hu; revised version received 18.Aug.2026; accepted 26.Aug.2026; published 20.Oct.2026.

Please cite as:
Liu HY, Huang YC, Chen SW, Wei LY, Chou CY, Lee JJ, Chen SJ, Lee CC
Mobile AI-Assisted Oral Lesion Triage for Community Oral Cancer Screening: Prospective Field Evaluation Study
JMIR Mhealth Uhealth 2026;14:e101721
URL: https://mhealth.jmir.org/2026/1/e101721
doi:10.2196/101721

© Hao-Yun Liu, Yu-Cheng Huang, Si-Wei Chen, Ling-Ying Wei, Cheng-Ying Chou, Jang-Jaer Lee, Shyh-Jye Chen, Chien-Chang Lee. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 20.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

CogniPaz, an Ecologically Oriented Gamified mHealth Tool for Detecting Mild Cognitive Impairment: Prospective Cohort Diagnostic Accuracy Study

Carlos Estebas-Armas[1], MD; Alejandro Medina-López[1], MD; Daniel Blanco-de-Miguel[1], MSc; Itsaso Losantos-García[2], MSc; María Hernández-Barral[1], MD; Elisa Alonso-López[1,3], MD, PhD; Sergio Henche-Rodríguez[1], MSc; Oscar Ocampo-Pérez[1], MD; Gerardo Ricardo Zmork-Martínez[1], MD; María Alonso-de-Leciñana[1], MD, PhD; Ana Frank-García[1,3], MD, PhD; Ángel Martín-Montes[1,3], MD, PhD

[1]Department of Neurology and Stroke Center, La Paz University Hospital, Universidad Autónoma de Madrid, Hospital La Paz Institute for Health Research, Paseo de la Castellana, 241, Madrid, Spain

[2]Biostatistics Department, La Paz University Hospital, Universidad Autónoma de Madrid, Hospital La Paz Institute for Health Research, Madrid, Spain

[3]Networking Research Center on Neurodegenerative Diseases (CIBERNED), Instituto de Salud Carlos III, Madrid, Spain

Corresponding Author:
Ángel Martín-Montes, MD, PhD
Department of Neurology and Stroke Center, La Paz University Hospital, Universidad Autónoma de Madrid, Hospital La Paz Institute for Health Research, Paseo de la Castellana, 241, Madrid, Spain

Abstract

Background: Discriminating mild cognitive impairment (MCI) from subjective cognitive decline (SCD), and healthy controls (HC) remains challenging. Digital tools may overcome paper-and-pencil screening's limitations.

Objective: This study evaluated the diagnostic accuracy and usability of CogniPaz (Daniel Blanco), an ecologically oriented, mobile health (mHealth), gamified cognitive assessment tool exploring memory and executive functions, in discriminating MCI (particularly amnestic phenotypes [aMCI]) from SCD, and HC.

Methods: This prospective observational study recruited a convenience sample of consecutive patients with cognitive complaints and HC at La Paz University Hospital Cognitive Disorders Unit from April 2025 to January 2026. Demographic and clinical characteristics, and Montreal Cognitive Assessment (MoCA) scores were collected. An evaluator blinded to the MoCA scores oversaw the automated, self-administered CogniPaz assessment. Participants were classified into HC (no cognitive complaints, MoCA score≥26), SCD (cognitive complaints, MoCA score≥26), and MCI (cognitive complaints without functional impairment, MoCA score<26, subdivided into amnestic [aMCI] and nonamnestic phenotypes). Receiver operating characteristic (ROC) analyses using age-adjusted models, the DeLong test, and bootstrap validation evaluated diagnostic accuracy across groups. The correlation between MoCA and CogniPaz scores was assessed using Spearman rank correlation. Participants with available biomarker data underwent exploratory analysis of CogniPaz's ability to detect biologically confirmed any-cause neurodegeneration (bND). Usability was tested using an ad hoc Likert questionnaire (1 - 9 points).

Results: The participants (n=84, 57.1% female; median age 72.0, IQR 62.5-77.0 years) included 27 HC, 17 with SCD, 22 with aMCI, and 18 with nonamnestic MCI. The aMCI group was older than the SCD group (age: median 76.5, IQR 70.8-78.8 vs median 59.0, IQR 50.0-69.0 years; Kruskal-Wallis/Dunn, *P*=.003). CogniPaz scores correlated with MoCA scores (ρ=0.81, 95% CI 0.75 - 0.89; *P*<.001). CogniPaz showed good concurrent accuracy, independently of age: MCI vs HC (area under the curve [AUC]=0.880, 95% CI 0.798 - 0.961), MCI vs SCD (AUC=0.915, 95% CI 0.841 - 0.990), aMCI vs HC (AUC=0.910, 95% CI 0.830 - 0.990), and aMCI vs SCD (AUC=0.943, 95% CI 0.878 - 1.000). At the provisional <28.5 cutoff point, sensitivity ranged from 85.0% (34/40, 95% CI 70.2 - 94.3) to 90.9% (20/22, 95% CI 70.8 - 98.9) and specificity from 74.1% (20/27, 95% CI 53.7 - 88.9) to 82.4% (14/17, 95% CI 56.6 - 96.2) across comparisons. CogniPaz suggested capacity to detect bND (AUC=0.810, 95% CI 0.661 - 0.958), with potential incremental value over MoCA and age (odds ratio [OR]=0.78, 95% CI 0.60 - 0.94; *P*=.02). Median administration time was 9.39 (IQR 7.53 - 11.38) minutes with 100% (84/84) completion. User satisfaction was high (median score 8.4/9, IQR 7.9 - 8.6; Cronbach α=0.84).

Conclusions: CogniPaz is an innovative, user-friendly mHealth tool that addresses classical cognitive screening tools' limitations through a brief, multidomain, minimally supervised gamified format, with high patient acceptance. It correlates with MoCA scores, shows good concurrent capacity to differentiate MCI, particularly aMCI, from SCD and HC, and appears promising for identifying bND. Pending external longitudinal validation, these findings support CogniPaz as a candidate for early cognitive screening in primary care.

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2026;14:e99688) doi:10.2196/99688

KEYWORDS

mild cognitive impairment; amnestic mild cognitive impairment; subjective cognitive decline; digital biomarkers; cognitive assessment; mobile applications; early diagnosis; Alzheimer disease; neurodegenerative diseases; ecological validity; usability

Introduction

Mild cognitive impairment (MCI) is a transitional clinical state between normal aging and dementia, characterized by objective cognitive decline with largely preserved functional independence [1]. Within MCI, amnestic forms (aMCI) are associated with an increased risk of progression to dementia in Alzheimer disease (AD) [2]. Subjective cognitive decline (SCD) refers to a self-perceived, persistent decline in cognitive capacity, typically involving memory, despite normal performance on standard cognitive tests, and is now considered a risk state for future MCI and dementia [3]. Differentiating aMCI from SCD is clinically relevant but challenging because complaints and even subtle performance differences frequently overlap, and early deficits may be masked by cognitive reserve. Early and accurate identification of individuals at increased risk for neurodegenerative disease is essential to guide monitoring, risk factor management, and potential disease-modifying interventions.

Traditional paper-and-pencil screening tools, such as the Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MoCA), have important roles but can exhibit ceiling effects, limited sensitivity to early changes, and constraints in repeated administration [4,5]. These limitations have prompted a rapid shift toward digital tools for cognitive assessments. Recent narrative and systematic reviews indicate substantial growth in digital cognitive tools for detecting MCI and dementia, including web-based batteries, tablet-based neuropsychological platforms, and smartphone apps [6-9]. Digital assessments can improve standardization, capture reaction times, automate scoring, and facilitate remote or high-frequency testing. Moreover, the scalability of these tools makes them ideal candidates to meet the growing demand for screening in aging populations [5]. Indeed, the integration of emerging digital technologies is increasingly recognized as a crucial strategy for developing scalable algorithms that facilitate early AD diagnosis and provide globally accessible screening guidance [10-12]. Meta-analytic evidence suggests that digital cognitive tools generally achieve good sensitivity and specificity for MCI and dementia detection, but performance varies by domain, platform, and comparison group [13-16].

However, important limitations remain. Recent reviews show that although digital cognitive assessment for older adults has expanded quickly, validation remains uneven: across web-, tablet-, and smartphone-based tools, construct validity and robust normative data are still incompletely documented [5,17]. Usability and implementation barriers remain a recurring problem, including difficulties navigating software, unstable connectivity, lengthy assessment times, and limited adaptability for people with less education and lower digital literacy [18,19].

These gaps support the development of brief, user-friendly tools that can be deployed in real-world pathways. Many digital tools have been validated primarily to distinguish dementia or global cognitive impairment from normal cognition, rather than to separate SCD from MCI in real-world memory clinic populations [5,20]. Ecological validity and user engagement, particularly among older patients and patients with anxiety, are often underreported, and the integration of digital tests into existing pathways for specialized care is still evolving [5,14,19,20]. Moreover, there is a paucity of tools capable of exploring executive functions alongside episodic memory in a brief, gamified format that can be administered easily by nonspecialist personnel [21]. In this regard, incorporating principles from adaptive mobile health (mHealth) systems, which use human-centered design to dynamically adjust to the user's needs and have proven effective in improving engagement and usability [22,23], offers a valuable framework for developing cognitive tools tailored to older adults.

To address these needs, we developed CogniPaz (Daniel Blanco), an ecologically oriented, tablet-based, gamified digital cognitive screening tool designed for rapid administration by nonspecialist clinicians (eg, general physicians or nurses) as a screening tool to identify MCI in patients consulting for SCD. Accordingly, the primary objectives of this prospective study were to evaluate the diagnostic accuracy of CogniPaz in discriminating MCI, particularly aMCI, from SCD and healthy controls (HC), to examine its correlation with MoCA scores, and to assess its feasibility and usability. In a secondary exploratory analysis, we also evaluated whether CogniPaz could identify biologically confirmed any-cause neurodegeneration (bND) in a subset of participants with available biomarker data.

Methods

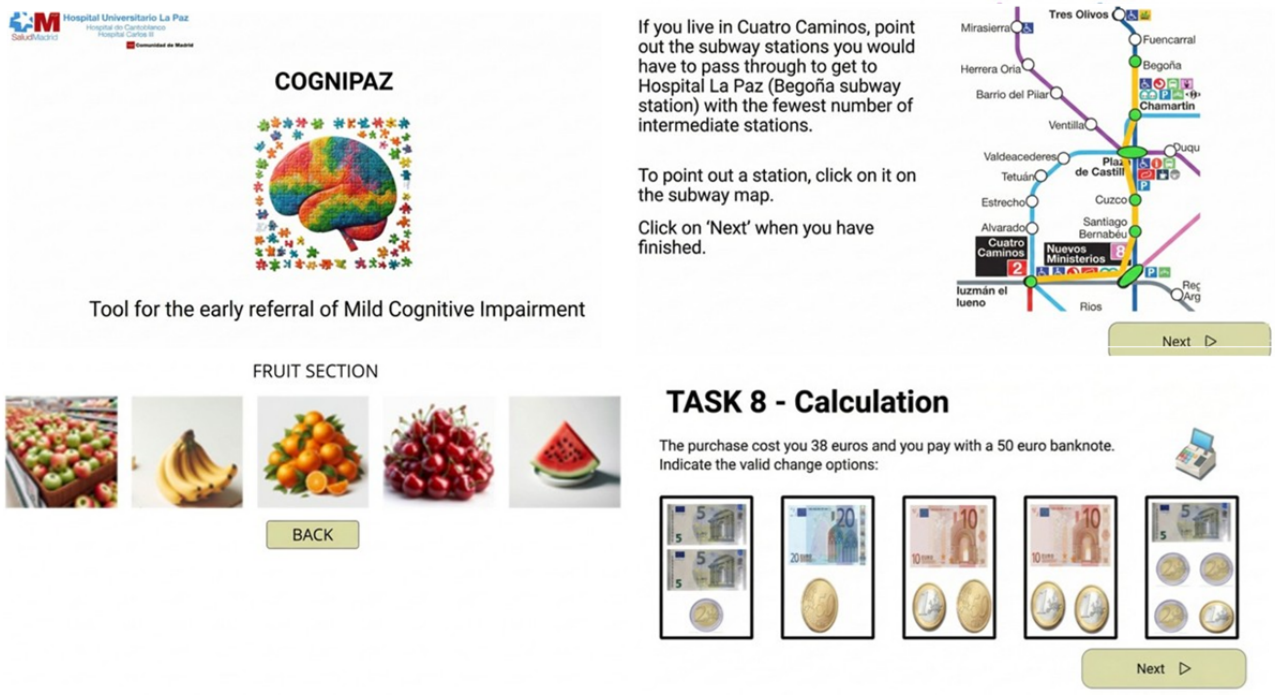
Instrumentation

CogniPaz was designed as a digital cognitive battery test. The study version of CogniPaz was implemented as an Android tablet app using Ionic/Angular and Capacitor. The protocol comprised 9 tasks, yielding a maximum total score of 38 points (Figure 1).

The device automatically provided the initial instructions to the participants before they began the tasks. To ensure accessibility, the platform featured an intuitive visual interface with minimal text to reduce linguistic bias, large font sizes for mild visual impairment, and an integrated audio guide for clear instructions. The assessment is structured as follows:

XSL•FO
RenderX

Figure 1. Examples of tasks included in the CogniPaz cognitive screening tool (English translation). The complete battery, in its original Spanish version, is provided in Multimedia Appendix 1.



Attention and Orientation (16 Points)

Task 1: Verbal Attention With Visual Input-Encoding

The participants encoded a list of 6 words (salmón [salmon], hamburguesas [hamburgers], azúcar [sugar], leche [milk], plátanos [bananas], and lejía [bleach]) presented sequentially in full-screen mode at 2-second intervals. This task was performed across 2 trials; after each presentation, participants verbally recalled the items using the tablet microphone; speech was processed transiently, and the resulting text was scored automatically by the app (6+6 points).

Task 3.1: Verbal Attention With Auditory Input-Encoding

The participants listened to a simulated TV news broadcast regarding a fire. Immediately after, they were required to manually type 3 specific details regarding the news story, including the number of victims, the distance in meters from the fire station, and the time of the event. They were also explicitly instructed to memorize these details, as they would be asked about them again later in the session (3 points).

Task 3.2: Orientation

The participants were asked to input the current date (day, month, and year) using the digital interface (1 point).

Attention and Working Memory (1 Point for Each Subtask, 3 Points)

Tasks 4 and 5: Forward and Backward Digit Span

The participants listened to and repeated increasingly long sequences of numbers, first in the same order, and then in reverse order.

Task 6: Visuospatial Sequencing

The participants were required to observe and replicate a sequence of changing colored circles by tapping them in the correct order.

Executive Function (4 Points)

Task 2: Planning

A metro route map was displayed; participants had to analyze the map and select the specific path that contains the fewest stations between 2 designated points (1 point).

Task 8: Calculation

In a simulated functional task, participants were presented with a total purchase costing €38 (US $42.02; €1=US $1.1057 as of April 4, 2025) and must calculate and select all the necessary currency denominations to provide the correct change from a €50 bill (US $55.29; 3 points).

Memory Domains (15 Points)

Task 7: Verbal Memory With Visual Input-Delayed Recall

This task evaluated the retrieval of the shopping list. It consists of a delayed free recall phase (Task 7.1), in which participants had to recall the original items (6 points). If free recall was incomplete, a facilitated cued recall phase (Task 7.2) was triggered to differentiate retrieval from storage deficits. A digital map of a supermarket was displayed, organized into various product sections (eg, bakery and butchery). When the participant tapped on a section, the interface displayed images of products in that category. The participant then had to identify and select the correct items from the original list (6 points). In both phases, the scoring algorithm applied a one-point penalty for each incorrect response. If a participant achieves a perfect score of

XSL•FO
RenderX

6 points in the free recall phase (Task 7.1), the facilitated cued recall phase (Task 7.2) was not administered, and the maximum score of 6 points for that phase was automatically awarded. The combined maximum for Task 7 is 12 points.

Task 9: Verbal Memory With Auditory Input—Delayed Recall

To assess long-term retrieval of the news broadcast, participants had to provide the 3 specific details previously encoded (3 points).

Tasks were scored based on the number of correctly recalled elements (1 point per item for Tasks 1, 3.1, 7, 9), on a dichotomous pass/fail basis (1/0 points for Tasks 2, 3.2, 4, 5, 6), or per correct denomination (1 point per item for Task 8). A total CogniPaz score was computed by summing the task scores; higher scores indicate better cognitive performance. Domain-specific scores for episodic memory and executive function were also generated. A full video demonstration in Spanish is available in Multimedia Appendix 1.

To ensure accurate and standardized quantification, CogniPaz used a deterministic rule-based scoring algorithm. For speech-based verbal recall tasks, the system transiently processed textual hypotheses returned by the native speech-recognition service, applying uppercase conversion and Spanish accent normalization. Target matching was then executed using a Levenshtein distance threshold of ≤1 to tolerate minor articulation, spelling, or speech-recognition variations. Further details on software architecture, data flow, and task-level scoring rules are provided in Multimedia Appendix 2.

A sanitized source-code artifact of the CogniPaz app is provided in Multimedia Appendix 3. This artifact is not intended to serve as a production deployment package or a fully executable clinical system. End-to-end execution requires a separately configured institutional REDCap (Vanderbilt University) backend project with matching fields, authorized private credentials, and institutionally approved deployment settings.

Upon completion, participants completed a custom-designed satisfaction questionnaire rated on a 9-point Likert scale (1=minimum; 9=maximum). This ad hoc, nonvalidated instrument covered 7 dimensions in total (general perception, ease of use, information clarity, task clarity, adequate time, task ease, and perceived utility), and was designed to capture aspects specific to a gamified, minimally supervised tablet-based assessment that are not fully covered by generic instruments. It was also evaluated for internal consistency using Cronbach alpha, which was computed among responders, alongside item-rest correlations and alpha-if-item-deleted statistics to assess the coherence of the construct.

Conditions and Design

This was a single-center, prospective observational study of patients with cognitive complaints and HC presenting to the Cognitive Disorders Unit of La Paz University Hospital (Madrid, Spain), a tertiary referral unit specializing in cognitive neurodegenerative diseases. In accordance with nonexperimental, observational designs, the participants' cognitive conditions were naturally observed, and no experimental manipulation was performed. This study was reported in adherence to the 2015 STARD (Standards for Reporting of Diagnostic Accuracy Studies) guidelines, with the completed checklist provided in .

Inclusion and Exclusion

The general inclusion criteria were patients with no diagnosis of dementia presenting with cognitive complaints. The comparison group of HC consisted of companions of the patients and other healthy volunteers recruited at the hospital who were age-matched to the patient group. The exclusion criteria were (1) established dementia, defined as significant functional impairment that compromised independence in instrumental activities of daily living (Clinical Dementia Rating [CDR] global score ≥1); (2) sensory or motor deficits (eg, severe visual impairment and upper limb paresis) precluding the use of a tablet; or (3) a significant language barrier or inability to communicate proficiently.

Participant Characteristics

Participants were classified into 3 diagnostic groups (HC, SCD, and MCI) based on a comprehensive clinical assessment integrating medical history, neurological examination, and standard neuropsychological testing. For subgroup analyses, the MCI group was further stratified into 2 phenotypes: aMCI and nonamnestic MCI. The MoCA, a widely validated instrument with established reliability and validity in the Spanish population [24], was used as the reference standard due to its sensitivity for MCI screening [25,26], and its potential value in predicting underlying AD pathology [27].

The MCI group was defined according to established clinical criteria [4,28]. Participants reported a persistent decline in cognitive capacity relative to their previous level of functioning, unrelated to an acute event, and this decline was corroborated by an informant. This was accompanied by objective impairment in at least one cognitive domain (MoCA <26, adjusted for education), while maintaining largely preserved independence in functional abilities (CDR global score of 0.5). For subgroup analyses, the aMCI subgroup was defined by predominant episodic memory impairment, whereas nonamnestic patients with MCI exhibited impairments in other cognitive domains [29].

The SCD group was defined according to the SCD-initiative framework [30]. Participants reported analogous subjective concerns regarding persistent cognitive decline but, in contrast to the MCI group, exhibited normal performance on standardized cognitive tests (MoCA ≥26, adjusted for education) and independence in functional abilities (CDR global score of 0).

HC were participants with an absence of active neurological or psychiatric disorders, no significant subjective cognitive concern, and performance within normal limits on screening (MoCA ≥26, adjusted for education).

Sampling Procedures

Participants were recruited using a consecutive convenience sampling procedure during scheduled clinical visits between April 2025 and January 2026. During this period, all consecutive patients and their accompanying individuals attending new

XSL•FO
RenderX

clinical consultations were systematically assessed for eligibility and recruited accordingly by the attending clinical and research staff to minimize selection bias. Potential candidates were provided with a detailed explanation of the study procedures prior to obtaining written informed consent. No financial compensation or other incentives were provided for participation.

Sample Size, Power, and Precision

No formal sample size calculation was performed because no previous data were available to estimate the expected effect sizes and because of the exploratory nature of the study. Therefore, a convenience sample was used based on feasibility criteria, including the expected recruitment capacity and available resources.

Measures and Covariates

Major demographic and clinical characteristics (including age, sex, education level, family history of AD, and MoCA scores) were collected and compared across the diagnostic groups (HC, SCD, MCI, aMCI, and nonamnestic MCI) to assess baseline differences. Structural neuroimaging (magnetic resonance imaging [MRI]), fluorodeoxyglucose positron emission tomography (FDG-PET), amyloid-PET, and cerebrospinal fluid (CSF) biomarkers (amyloid beta 42 [Aβ42], amyloid beta 42/40 ratio [Aβ42/40], phosphorylated tau [p-tau], and total tau [t-tau]) [31] were also reviewed in participants who had undergone these procedures, based on the routine clinical judgment of the attending physician. The primary investigational measures derived from the CogniPaz platform included the total cognitive score (ranging from 0 to 38) and the total administration time. Feasibility covariates included the completion rate, technical malfunctions, and the frequency of required external assistance. User satisfaction was measured using the 9-point ad hoc Likert scale questionnaire.

Data Collection

Participants completed the assessments during a single, in-person session in a quiet room at the cognitive disorders unit. To reflect routine clinical workflow, the order of administration was fixed (MoCA followed by CogniPaz). CogniPaz was administered on a 9-inch touchscreen device (Lenovo Tab M9, Android 12), and performance metrics (including correct responses, error counts, and completion times) were automatically captured by the software. To ensure secure data gathering, the tablet front end transmitted structured assessment data securely via HTTPS POST requests to a REDCap-compatible API. The app used an automated offline contingency mechanism (Ionic Storage) to temporarily queue data locally and prevent loss during Wi-Fi interruptions in the clinical environment. The research team recorded any technical malfunctions, data loss incidents, and the frequency and nature of required external assistance. Upon completion of the cognitive tasks, participants filled out the satisfaction questionnaire.

Quality of Measurements

To ensure standardization and enhance the quality of measurements, the research team attended an introductory training session prior to the study. To minimize interrater variability during data collection, the majority of the CogniPaz administrations were overseen by a single trained investigator. Supervision was minimal, limited strictly to technical support or task clarification to prevent administration bias.

Masking

Routine clinical assessment and MoCA administration were performed by neurologists blinded to the participants' CogniPaz scores. Subsequently, the self-administered, automated CogniPaz assessment was overseen by researchers who were entirely blinded to the participants' clinical diagnosis and MoCA results.

Data Diagnostics

A predefined protocol was established to manage data integrity prior to inferential testing. Post-data-collection exclusion criteria were set to disqualify sessions with incomplete data tracking resulting from technical disruptions or participant withdrawal. No missing data imputation methods were planned; instead, a complete-case analysis approach was prespecified for any incomplete records. In case of missing data, the Little test of missing completely at random (MCAR test) was performed to determine if missing data were missing completely at random. We also compared participants who completed the satisfaction questionnaire ("responders") with those who did not ("nonresponders") across demographic and clinical variables, using Welch *t* test for continuous variables and Fisher exact test for categorical variables, to assess whether noncompletion was associated with participant characteristics. Statistical outliers in continuous variables were defined using the IQR method. The normality of continuous-variable distributions was systematically evaluated using the Shapiro-Wilk test to guide the selection between parametric and nonparametric inferential strategies. Data transformations were not anticipated, as nonparametric methods were preferred to robustly handle potential skewness in neurocognitive metrics and session administration times.

Data Analysis

Continuous variables were presented as medians and IQR if nonnormal distribution and heterogeneity of variances were observed. Between-group comparisons of continuous variables were conducted using the Kruskal-Wallis test, run separately across the 3 primary diagnostic groups (HC, SCD, and MCI) and, within MCI, across its 2 subtypes (aMCI and nonamnestic MCI). Significant global results were followed by Dunn post hoc pairwise comparisons with Bonferroni correction.

Categorical variables were analyzed based on the distribution of expected frequencies. The chi-square test was used for variables meeting the assumption of expected cell counts ≥5 (sex and education level). Fisher exact test was used for variables with sparse data violating these assumptions (family history of dementia) and for all post hoc pairwise comparisons to ensure robustness against small subgroup sizes.

Correlations between CogniPaz global scores and MoCA indices were evaluated using Spearman rank coefficients. The 95% CIs for these correlation coefficients were derived using bootstrap resampling with 5000 replications. Diagnostic accuracy of CogniPaz was examined using receiver operating characteristic (ROC) curve analyses for the following four comparisons: (1)

XSL•FO
RenderX

MCI vs SCD, (2) MCI vs HC, (3) aMCI vs SCD, and (4) aMCI vs HC. For the secondary exploratory subanalysis, ROC analysis evaluated the tool's capacity to discriminate between biologically confirmed versus nonconfirmed ND. For each analysis, the area under the curve (AUC) with 95% CIs was calculated.

The Youden Index was initially used to identify the mathematically optimal threshold for each diagnostic comparison. To optimize the tool for clinical screening, a unified provisional threshold strategy was then applied, prioritizing high sensitivity to minimize false negatives. At the provisional, hypothesis-generating, unified cutoff point, we reported sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and overall diagnostic accuracy, each with 95% Clopper-Pearson CI. Additionally, positive and negative likelihood ratios (LR and LR-) were computed, with 95% CI derived using the log-normal method, to provide prevalence-independent measures of diagnostic performance. To characterize performance around each cutoff and support its interpretation as provisional, we calculated a threshold-performance table showing all metrics with 95% CI across a grid of candidate cutoffs (24.5, 26.5, and 28.5) for every comparison. It should be noted that MoCA scores contributed to the diagnostic group; accordingly, the primary diagnostic accuracy analyses constitute concurrent criterion validity assessment and are subject to potential incorporation bias.

To evaluate whether the tool's performance was influenced by demographic factors, we compared unadjusted models (CogniPaz score alone) against adjusted models incorporating those variables that showed significant baseline differences between groups. Using the DeLong test for correlated ROC curves, we assessed whether the inclusion of these variables significantly improved diagnostic accuracy or if the tool's performance remained independent of demographic characteristics. Furthermore, to quantify the specific explanatory power of the tool, we estimated the proportion of variance explained by each model using McFadden pseudo-R^2. The unique variance attributable to CogniPaz was calculated as the absolute increase in explained variance (incremental R^2) observed when adding the total score to the base model containing the confounding variables.

Internal validation was performed using bootstrap resampling with 1000 iterations to assess the robustness of diagnostic estimates. Optimism-corrected AUCs, sensitivities, and specificities were derived, with corresponding 95% CIs. Leave-one-out cross-validation (LOOCV) was conducted to confirm the stability of the results.

To assess the added value of CogniPaz relative to conventional cognitive screening, we conducted a secondary analysis restricted to the biomarker-confirmed subsample with a concurrently administered MoCA. CogniPaz and MoCA discrimination for confirmed bND were compared using the DeLong test for 2 correlated ROC curves, given that both measures were obtained in the same patients. To evaluate the potential incremental value of CogniPaz beyond MoCA and age, we fitted 2 nested logistic regression models predicting confirmed status: a base model including MoCA total score and age, and a full model additionally including the CogniPaz total score. Model improvement was assessed with a likelihood-ratio test. Effect estimates for CogniPaz in the full model were reported as adjusted odds ratios (ORs) with 95% CIs.

The 4 primary ROC comparisons, the secondary bND subanalysis, and the grid of candidate cutoffs were prespecified in the analysis plan. Dunn post hoc pairwise comparisons were Bonferroni-corrected; no further correction for multiplicity was applied across the prespecified primary comparisons, and the bND subanalysis and threshold-performance grid are reported as exploratory and hypothesis-generating. Reported *P* values should be interpreted accordingly.

Statistical analyses were performed using R software (version 4.4.2; R Foundation for Statistical Computing). All tests were 2-tailed, and a P value <.05 was considered statistically significant.

Ethical Considerations

The study protocol adhered to the principles of the Declaration of Helsinki and was approved by the Clinical Research Ethics Committee of La Paz University Hospital (HULP Code: PI-6915) on January 22, 2025 (Acta No. 02/2025). All participants provided written informed consent prior to enrollment. All data were deidentified at the source and stored in compliance with the General Data Protection Regulation (GDPR; EU 2016/679) and Spanish Organic Law 3/2018 on secure servers at the IdiPAZ Research Institute. No participant audio recordings or raw speech-recognition strings were stored or transmitted. No individual-level identifiable data or images of individual participants are reported. No financial compensation was provided to participants.

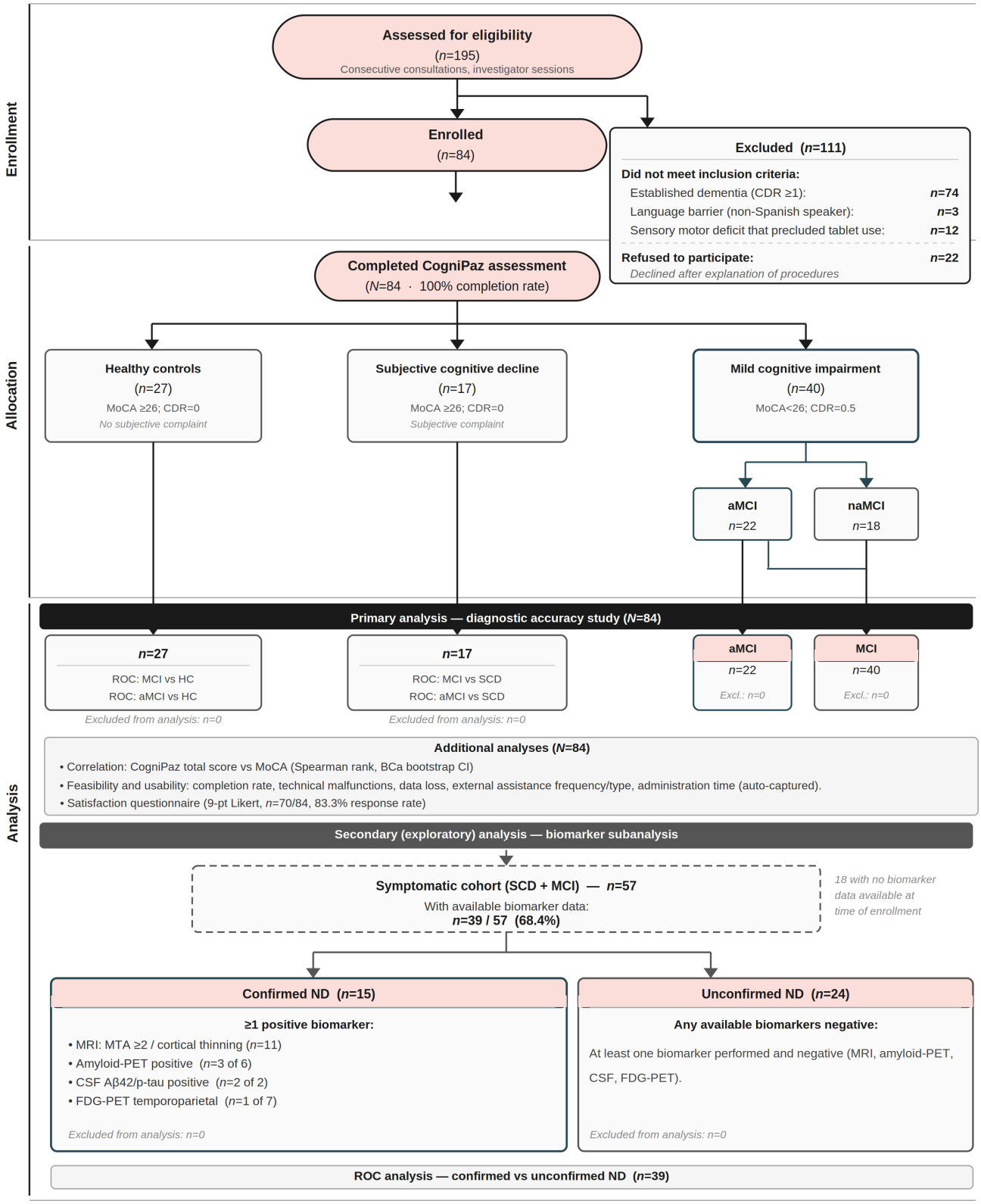
Results

Participants

A total of 195 consecutive individuals were assessed for eligibility. Of these, 111 (56.9%) were excluded, primarily due to established dementia (74/111, 66.7%) or refusal to participate (22/111, 19.8%). A total of 84 (43.1%) participants provided informed consent, met all criteria, and were formally enrolled. The test completion rate among enrolled participants was 100% (84/84), with no missing data or dropouts during the assessment. The complete participant flow, including detailed reasons for exclusion, allocation to primary diagnostic groups, and patients included in the exploratory biomarker subanalysis cohort, is depicted in Figure 2.

XSL•FO
RenderX

Figure 2. Participants flowchart, adapted from JARS-Quant guidelines. Aβ42: amyloid beta 42; aMCI: amnestic mild cognitive impairment; CDR: clinical dementia rating; CSF: cerebrospinal fluid; FDG-PET: fluorodeoxyglucose positron emission tomography; HC: healthy controls; MCI: mild cognitive impairment; MoCA: Montreal cognitive assessment; MRI: magnetic resonance imaging; MTA: medial temporal atrophy; naMCI: nonamnestic mild cognitive impairment; ND: neurodegeneration; PET: positron emission tomography; p-tau: phosphorylated tau; ROC: receiver operating characteristic; SCD: subjective cognitive decline.



Sample Characteristics

The study cohort comprised 84 participants. Final diagnoses were 27 (32.1%) HC, 17 (20.2%) SCD, and 40 (47.6%) MCI. Within the MCI group, 22 (55.0%) participants were classified as having aMCI and 18 (45.0%) as nonamnestic MCI. Detailed demographic and clinical characteristics are summarized in Table 1.

XSL•FO
RenderX

Table . Demographic and clinical characteristics and median scores of the study population.

Characteristic	Total (n=84)	HC[a] (n=27)	SCD[b] (n=17)	MCI[c] (n=40)	aMCI[d] (n=22)	naMCI[e] (n=18)	*P* value[f]
Age, median (IQR)	72.0 (62.5-77.0)	74.0 (59.5-77.5)	59.0 (50.0-69.0)	74.0 (69.0-78.0)	76.5 (70.8-78.8)	71.5 (68.0-74.0)	.003[g]
Women, n (%)	48 (57.1)	11 (40.7)	9 (52.9)	28 (70.0)	16 (72.7)	12 (66.7)	.12[h]
Education (college/professional training), n (%)	57 (67.9)	19 (70.4)	14 (82.4)	24 (60.0)	14 (63.6)	10 (55.6)	.37[h]
Family history of Alzheimer disease, n (%)	25 (29.8)	3 (11.1)	5 (29.4)	17 (42.5)	8 (36.4)	9 (50.0)	.02[j]
MoCA[i] scores, median (IQR)	26.0 (23.3-27.0)	27.0 (26.0-29.0)	27.0 (27.0-28.0)	23.0 (21.0-25.0)	23.0 (21.0-24.0)	24.0 (22.0-25.0)	<.001[g]
CogniPaz scores, median (IQR)	28.0 (22.0-31.0)	30.0 (29.0-34.0)	31.0 (30.0-33.0)	22.5 (16.5-26.3)	19.0 (14.0-24.0)	24.0 (21.3-27.0)	<.001[g]

[a]HC: healthy control.

[b]SCD: subjective cognitive decline.

[c]MCI: mild cognitive impairment.

[d]aMCI: amnestic mild cognitive impairment.

[e]naMCI: nonamnestic mild cognitive impairment.

[f]*P* values correspond to the global comparison across 4 diagnostic groups: HC, SCD, aMCI, and naMCI.

[g]Kruskal-Wallis test.

[h]Chi-square test.

[i]MoCA: Montreal Cognitive Assessment.

[j]Fisher exact test.

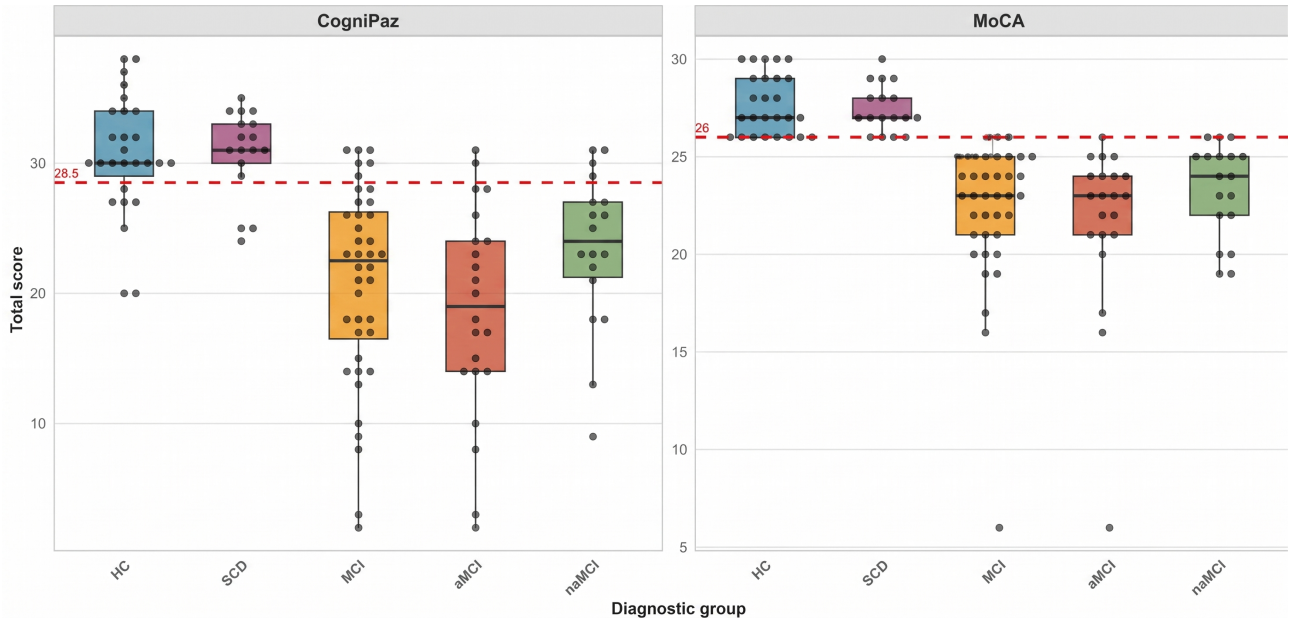
For significant continuous variables, post hoc pairwise comparisons were performed using the Dunn test (Bonferroni-adjusted), run separately for the 3 primary groups (HC, SCD, and MCI) and for the 2 MCI subtypes (aMCI and nonamnestic MCI). Significant differences ($P<.05$) were found for age (MCI vs SCD, $P=.02$; aMCI vs SCD, $P=.003$); family history (HC vs MCI; $P=.006$); MoCA score (MCI and aMCI vs HC and SCD, all $P<.001$); and CogniPaz score (MCI and aMCI vs HC and SCD, all $P<.001$).

The median age of the cohort was 72 years. Significant age differences were observed between groups, primarily driven by SCD group being younger than those with MCI ($P=.02$) and aMCI ($P=.003$). Sex distribution and educational levels were well-balanced across all diagnostic categories. A family history of AD was significantly more prevalent in the clinical groups compared to the control group; specifically, in patients with MCI ($P=.006$).

Figure 3 illustrates the significant differences in MoCA and CogniPaz scores observed across diagnostic groups.

XSL•FO
RenderX

Figure 3. Distribution of baseline CogniPaz and Montreal Cognitive Assessment (MoCA) total scores across diagnostic groups. Boxplots show the median and IQR, and dots represent individual participants. aMCI: amnestic mild cognitive impairment; HC: healthy controls; MCI: mild cognitive impairment; naMCI: nonamnestic mild cognitive impairment; SCD: subjective cognitive decline.



MoCA scores were comparable between HC and SCD, confirming the objective normality of the SCD group on standard testing. In contrast, MCI and patients with aMCI showed significantly lower scores, with pairwise comparisons confirming significant (P<.001) differences compared to both HC and SCD groups. Similarly, CogniPaz total scores differed significantly between groups. While HC and SCD participants maintained preserved performance, with no significant differences between them, MCI—and particularly the aMCI subgroup—exhibited lower scores. Pairwise comparisons confirmed that patients with aMCI scored significantly lower than both HC (P<.001) and SCD (P<.001) participants, supporting the tool's ability to discriminate objective deficits in the target population.

MoCA scores displayed a ceiling effect in the HC and SCD groups and revealed no significant differences between aMCI and nonamnestic MCI phenotypes (P=.23). In contrast, CogniPaz scores exhibited a wider distribution. While aMCI participants obtained numerically lower scores than nonamnestic MCI cases, this difference did not reach statistical significance (P=.07).

Feasibility and Administration Time

All recruited participants completed the 9 CogniPaz tasks, achieving a 100% (84/84) task completion rate across all clinical sessions, with no indeterminate results, technical failures, software crashes, or data synchronization losses due to application malfunction reported during the study; thus, all recruited participants who completed the protocol were included in the final analysis. Regarding administration time, the total sample followed a nonnormal distribution. The median total administration time was 9.39 minutes (IQR 7.53 - 11.38). Notably, when stratified by diagnosis, the individual subgroups displayed normal distributions. This difference suggests that the global nonnormality is driven by a distinct performance clustering between groups. As illustrated in Figure 4 and detailed in Table 2, both HC and SCD groups completed the assessment significantly faster than participants with MCI.

XSL•FO
RenderX

Figure 4. Distribution of CogniPaz administration times across diagnostic groups. Violin plots represent the distribution density of administration times (minutes), embedded boxplots show the median and interquartile range, and dots represent individual participants. aMCI: amnestic mild cognitive impairment; HC: healthy controls; MCI: mild cognitive impairment; naMCI: nonamnestic mild cognitive impairment; SCD: subjective cognitive decline.

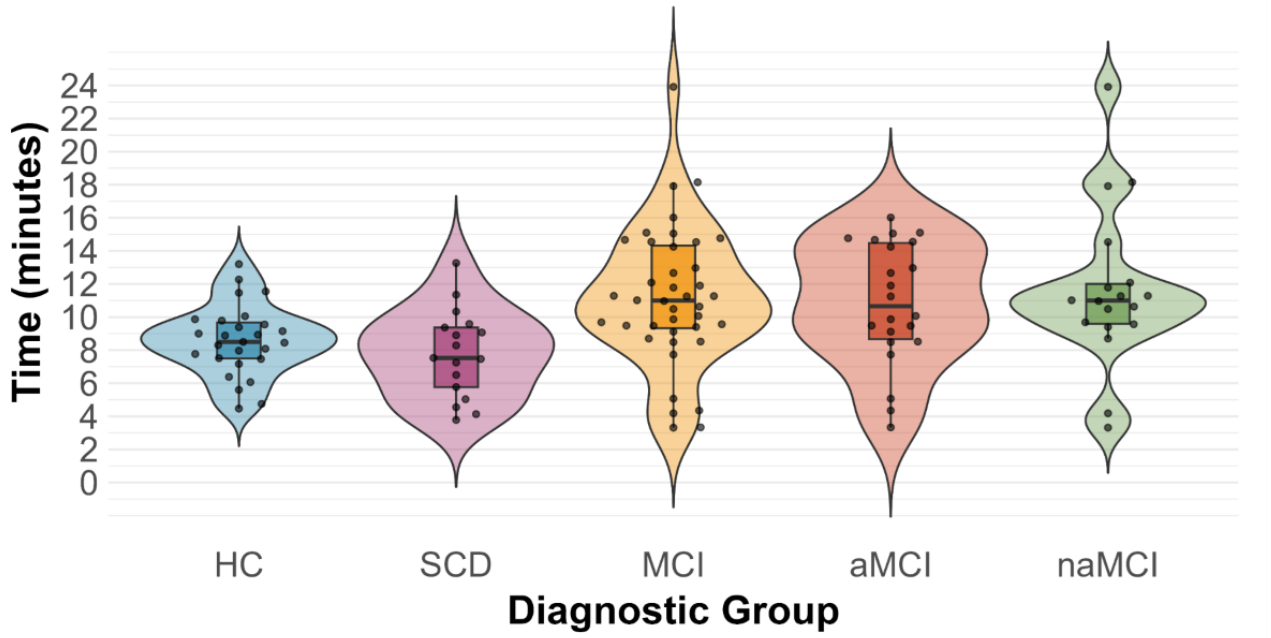


Table . Comparison of completion times (minutes) across diagnostic groups.

Group	N	Median IQR	*P* value (vs HC)[a]	*P* value (vs SCD)[b]	*P* value (vs MCI)[c]	*P* value (vs aMCI)[d]	*P* value (vs naMCI[e])[f]
Total	84	9.39 (7.53-11.38)	—[g]	—	—	—	—
HC	27	8.50 (7.49-9.67)	—	.99	*.02*[h]	.10	.06
SCD	17	7.53 (5.77-9.37)	.99	—	*.007*	*.04*	*.02*
MCI	40	10.99 (9.33-14.32)	*.02*	*.007*	—	—	—
aMCI	22	10.66 (8.67-14.47)	.10	*.04*	—	—	.99
naMCI	18	10.99 (9.60-12.01)	.06	*.02*	—	.99	—

[a]HC: healthy controls.

[b]SCD: subjective cognitive decline.

[c]MCI: mild cognitive impairment.

[d]aMCI: amnestic mild cognitive impairment.

[e]naMCI: nonamnestic mild cognitive impairment.

[f]Kruskal-Wallis test for intergroup comparisons followed by Dunn test for pairwise comparisons, run separately for the 3 primary groups (HC, SCD, and MCI) and for the 2 MCI subtypes (aMCI and nonamnestic MCI).

[g]Not applicable.

[h]Italicized values indicate statistically significant differences, $P<.05$.

Correlation Between CogniPaz Scores and MoCA Scores

The CogniPaz total score exhibited a strong positive correlation with the MoCA total score (Spearman ρ=0.81, 95% CI 0.75 - 0.89; $P<.001$). Furthermore, domain-specific validity was supported by high correlations between the memory-related CogniPaz tasks and the corresponding MoCA memory indices (ρ=0.61-0.68; $P<.001$), suggesting that the digital metrics accurately capture the targeted cognitive constructs.

Diagnostic Accuracy

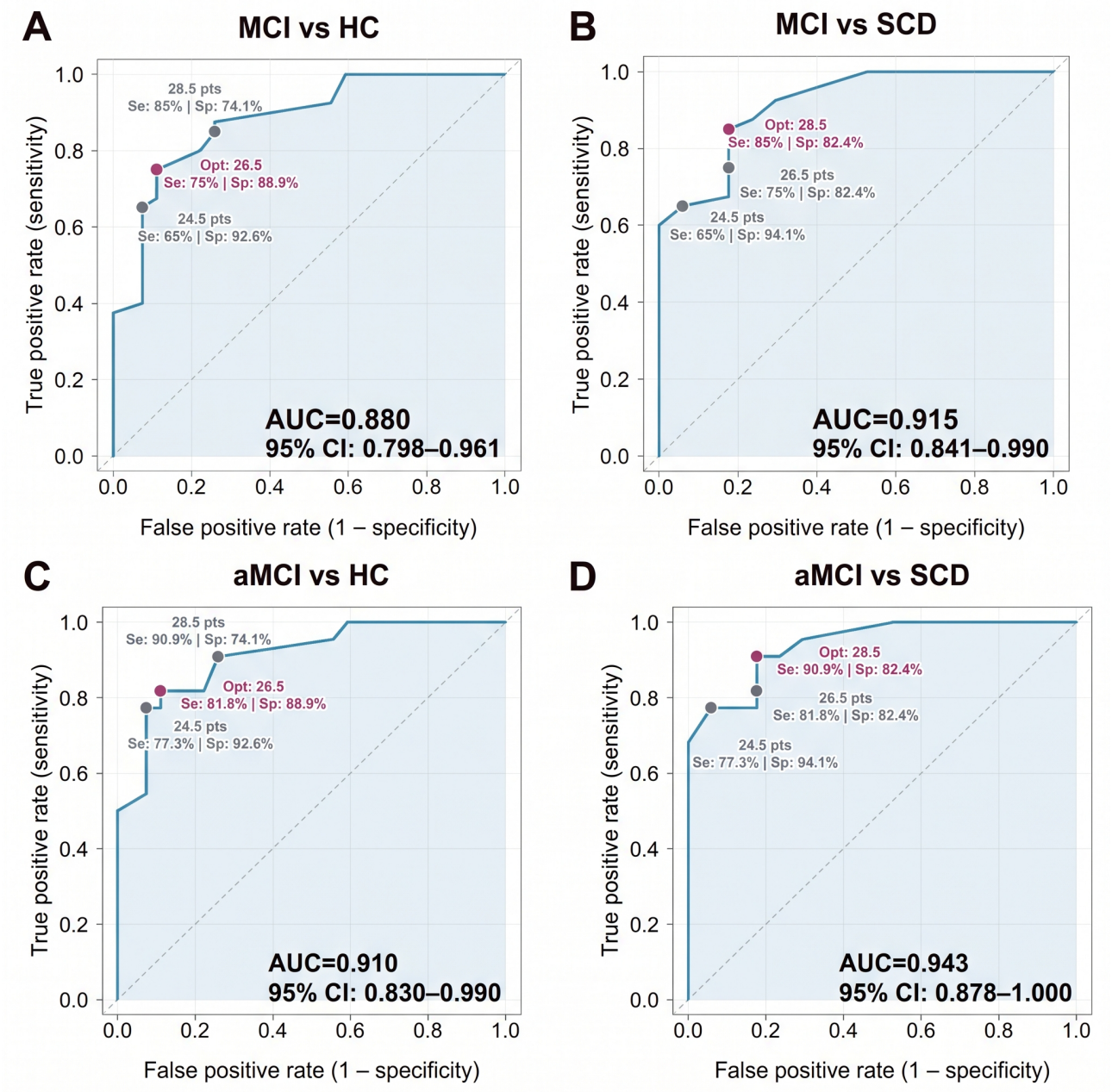
CogniPaz discriminated between MCI and HC and between MCI and SCD, with numerically higher accuracy for the amnestic MCI subgroup than for the overall MCI group. Figure 5 presents the ROC curve analyses for all diagnostic

XSL•FO
RenderX

comparisons. Because aMCI is a subset of MCI, the corresponding AUCs are not statistically independent, and relative performance across these comparisons should be interpreted descriptively rather than as formally tested differences.

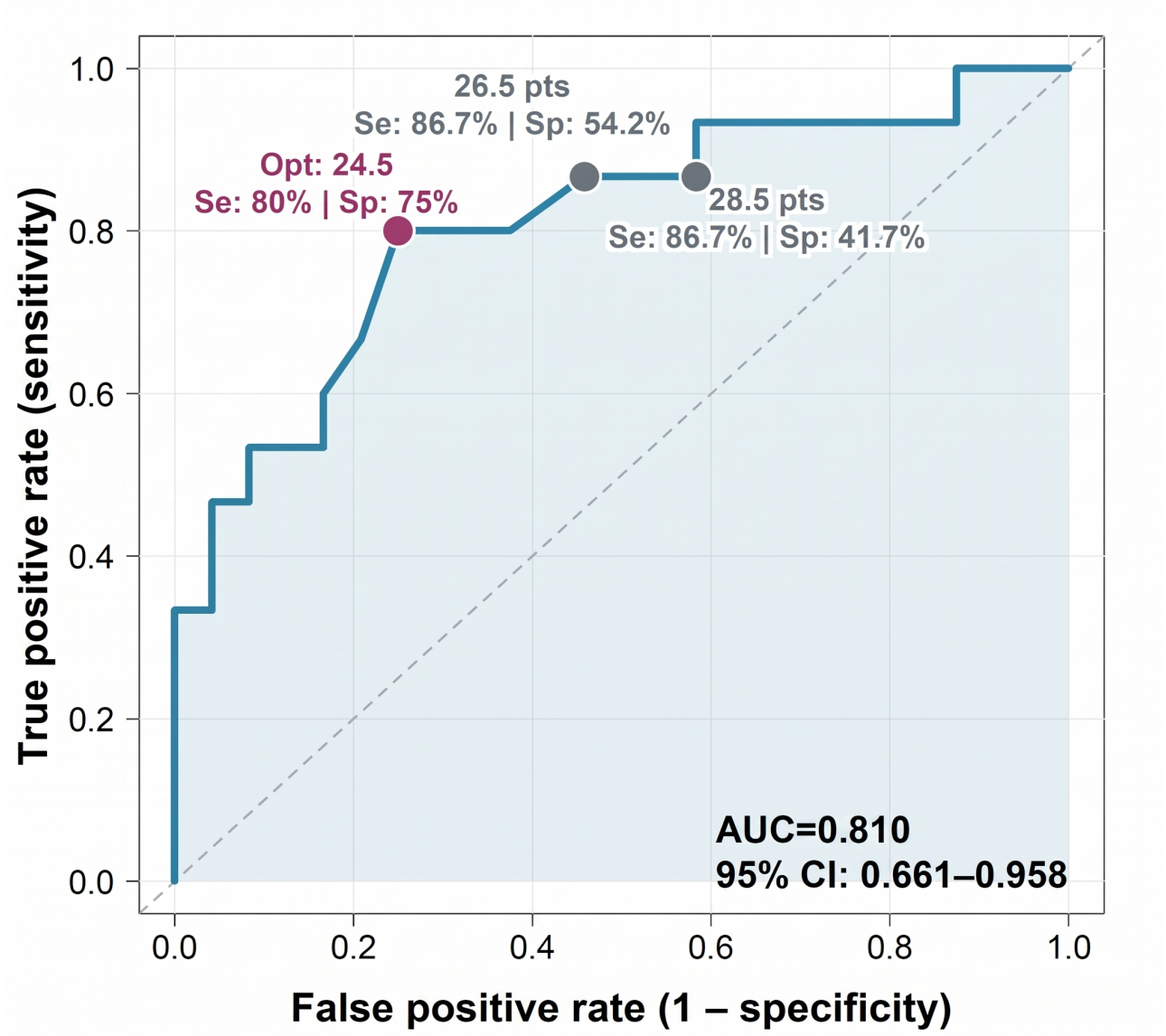
We further conducted a prespecified secondary exploratory subanalysis to evaluate the tool's capacity to identify underlying bND within patients in the symptomatic cohort (SCD + MCI). Among patients with available biomarker data (n=39; 68.4% of the 57 symptomatic participants), 15 (38.5%) had confirmatory data, and 24 (61.5%) did not. The confirmed group included patients with hippocampal atrophy (medial temporal atrophy [MTA] score ≥2) and/or a reduced cortical thickness in the precuneus, entorhinal cortex, and posterior cingulate regions (11/37 MRIs performed, 29.7%), positive amyloid-PET (3/6, 50.0%), positive CSF biomarkers (2/2, 100.0%), or suggestive FDG-PET hypometabolism, defined as a temporoparietal pattern (1/7, 14.3%); the remaining cases showed nonspecific, patchy frontoparietal (2/7, 28.6%) or parietal (1/7, 14.3%) patterns, or were entirely normal (3/7, 42.9%) [32]. In this specific subsample, the tool achieved an AUC of 0.810 (95% CI 0.661 - 0.958) for discriminating both categories. ROC curve analyses are presented in Figure 6.

Figure 5. Receiver operating characteristic (ROC) curves showing the diagnostic performance of CogniPaz for identifying mild cognitive impairment (MCI) and amnestic mild cognitive impairment (aMCI) in the study population. (A) MCI versus healthy controls (HC). (B) MCI versus subjective cognitive decline (SCD). (C) aMCI versus HC. (D) aMCI versus SCD. The highlighted point in red indicates the optimal cutoff determined using the Youden index. Sensitivity and specificity at the 24.5-, 26.5-, and 28.5-point cutoffs are displayed for each comparison. The diagonal dashed line represents chance-level discrimination. aMCI: amnestic mild cognitive impairment; AUC: area under the curve; HC: healthy controls; MCI: mild cognitive impairment; Opt: optimal cutoff; Pts: points; Se: sensitivity; Sp: specificity.



XSL•FO
RenderX

Figure 6. Receiver operating characteristic (ROC) curve showing the diagnostic performance of CogniPaz for identifying biologically confirmed any-cause neurodegeneration among symptomatic participants (subjective cognitive decline and mild cognitive impairment) with available biomarker data recruited from a tertiary memory clinic. The highlighted point in red indicates the optimal cutoff determined using the Youden index. The diagonal dashed line represents chance-level discrimination. Sensitivity and specificity at 24.5, 26.5, and 28.5 cutoffs are shown, highlighting the optimal threshold. AUC: area under the curve; Opt: optimal cutoff; Pts: points; Se*:* sensitivity; Sp: specificity.



To formally address the added value of CogniPaz over conventional screening of neurodegenerative etiology, we analyzed the biomarker-available subsample (n=39; 15 confirmed bND, 24 not confirmed bND). CogniPaz showed higher discrimination than MoCA for confirmed bND (AUC=0.810, 95% CI 0.661 - 0.958, vs AUC=0.661, 95% CI 0.471 - 0.851; DeLong Z=1.97; P=.05). In multivariable analysis, adding the CogniPaz total score to a model including MoCA and age significantly improved model fit (likelihood-ratio test: χ^2_1=7.46; P=.006), with each additional CogniPaz point associated with lower odds of confirmed bND independently of MoCA and age (adjusted OR=0.78, 95% CI 0.60 - 0.94; P=.02).

To account for the significant differences in age identified in the baseline characteristics, we analyzed unadjusted and age-adjusted regression models. The inclusion of age did not yield a statistically significant improvement in diagnostic accuracy for any comparison, as detailed in Table 3.

XSL•FO
RenderX

Table . Comparison of diagnostic accuracy between unadjusted and age-adjusted CogniPaz models across the diagnostic subgroup comparisons.

Diagnostic comparison	Unadjusted AUC[a] (95% CI)	Age-adjusted AUC (95% CI)
MCI[b] vs HC[c]	0.880 (0.798 - 0.961)	0.901 (0.828 - 0.974)
MCI vs SCD[d]	0.915 (0.841 - 0.990)	0.921 (0.847 - 0.995)
aMCI[e] vs HC	0.910 (0.830 - 0.990)	0.924 (0.855 - 0.994)
aMCI vs SCD	0.943 (0.878 - 1.000)	0.943 (0.878 - 1.000)
Confirmed ND vs Unconfirmed ND[f]	0.810 (0.661 - 0.958)	0.811 (0.660 - 0.961)

[a]AUC: area under the curve.

[b]MCI: mild cognitive impairment.

[c]HC: healthy controls.

[d]SCD: subjective cognitive decline.

[e]aMCI: amnestic mild cognitive impairment.

[f]ND: neurodegeneration.

The analysis of explained variance demonstrated that CogniPaz accounted for a substantial proportion of the unique variance in diagnostic classification (20.9% - 38.8%), whereas age contributed only marginal additional variance (1.4% - 2.9%). These findings suggest that the discriminative ability of the tool may reflect disease-specific cognitive impairment rather than age differences between the diagnostic groups.

To optimize clinical utility, we evaluated the most discriminative cutoff point of CogniPaz scores based on the Youden index, which identified the optimal threshold according to the balance between sensitivity and specificity depending on the comparison group. To distinguish MCI from HC, the optimal threshold was identified at 26.5 points. At this cutoff, CogniPaz demonstrated a sensitivity of 81.8% and specificity of 88.9% for aMCI vs HC (PPV: 85.7%, NPV: 85.7%) and a sensitivity of 75.0% and specificity of 88.9% for MCI vs HC (PPV: 90.9%, NPV: 70.6%). Conversely, to distinguish MCI from SCD, the optimal threshold shifted to 28.5 points, as shown in Figure 5.

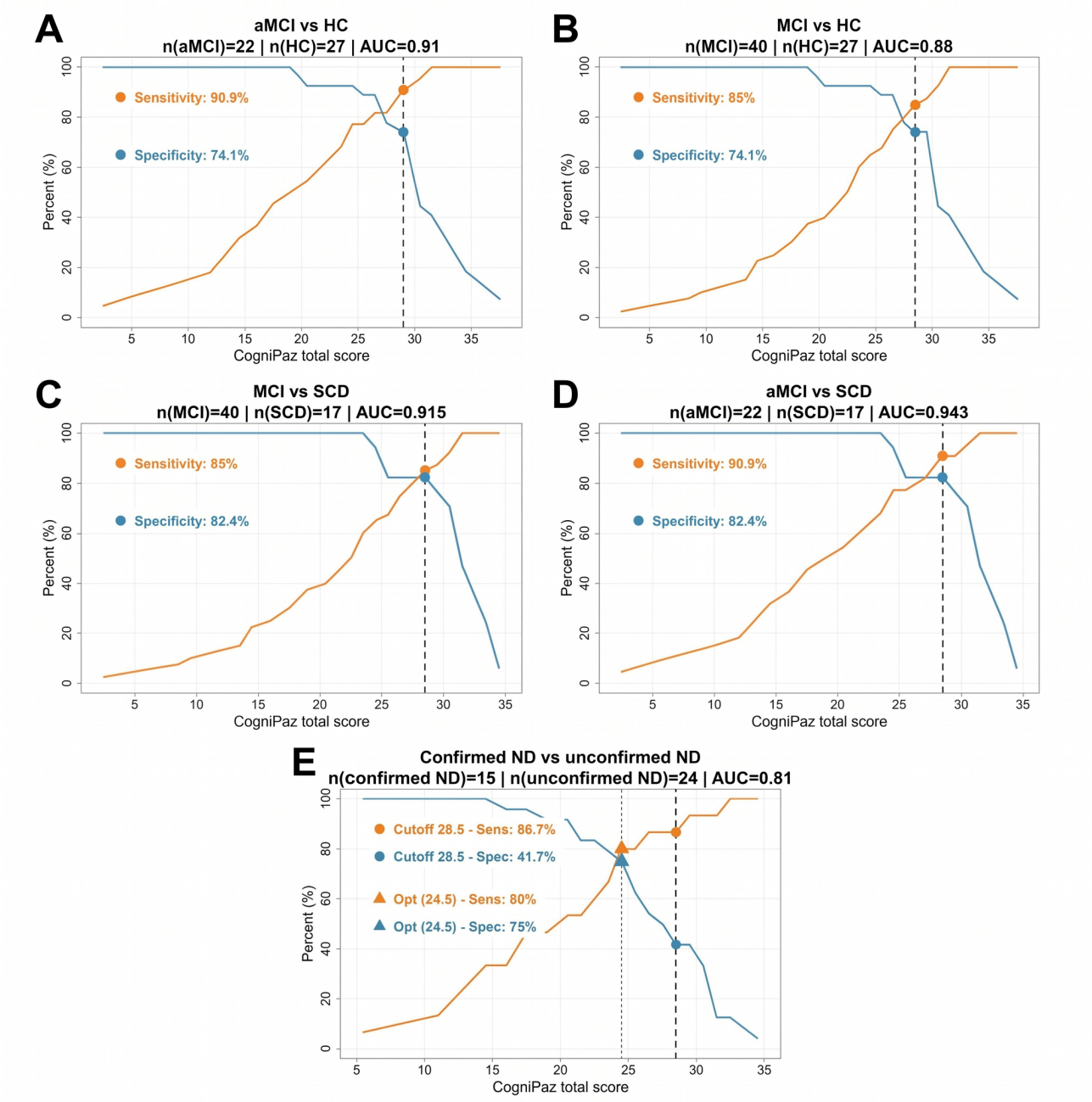
Since the primary objective of CogniPaz is to serve as an effective screening instrument, we prioritized sensitivity to minimize the risk of missing cases by establishing a provisional unified cutoff of 28.5 points (scores below this threshold indicate cognitive impairment). The shift from 26.5 to 28.5 resulted in an improvement in sensitivity for both aMCI vs HC (from 81.8% to 90.9%) and MCI vs HC (from 75.0% to 85.0%). This gain in sensitivity came with a trade-off in specificity, which decreased from 88.9% to 74.1% for both MCI vs HC and aMCI vs HC, as depicted in Figure 7.

In the subanalysis of bND, although the 28.5-point cutoff maintained good sensitivity (86.7%), this came at the cost of lower specificity (41.7%). The optimal 24.5 threshold provided a more balanced performance for identifying bND, increasing specificity to 75.0% and overall accuracy to 76.9% vs 59.0% at the 28.5 cutoff. The 95% CIs for all metrics at every candidate cutoff are provided in Table S1 in Multimedia Appendix 4.

At this level, 81.0% of the total cohort (68/84 participants) was correctly classified. Table 4 presents the final performance metrics at the provisional 28.5-point cutoff for the clinical comparisons and, separately, at the 24.5-point cutoff for the exploratory bND subanalysis; the complete threshold-performance table across all candidate cutoffs (24.5, 26.5, and 28.5), with 95% CIs for every metric, is provided in Table S1 in Multimedia Appendix 4.

XSL•FO
RenderX

Figure 7. Sensitivity and specificity of the CogniPaz digital cognitive screening tool across the evaluated diagnostic comparisons. (A) amnestic mild cognitive impairment (aMCI) versus healthy controls (HC). (B) mild cognitive impairment (MCI) versus HC. (C) MCI versus subjective cognitive decline (SCD). (D) aMCI versus SCD. (E) Biologically confirmed versus unconfirmed neurodegeneration. Orange and blue curves represent sensitivity and specificity, respectively, across the range of CogniPaz total scores. Dashed vertical lines indicate the provisional 28.5-point cutoff; in panel E, the dotted vertical line indicates the alternative 24.5-point cutoff. aMCI: amnestic mild cognitive impairment; HC: healthy controls; MCI: mild cognitive impairment; ND*:* neurodegeneration; SCD: subjective cognitive decline.



XSL•FO
RenderX

Table . Diagnostic performance and classification accuracy of the CogniPaz tool across clinical comparisons at the 28.5 cutoff and at the 24.5 cutoff for biologically confirmed any-cause neurodegeneration.

Diagnostic comparison	TP[a]	TN[b]	FP[c]	FN[d]	Sens[e] (95% CI)	Spec[f] (95% CI)	PPV[g] (95% CI)	NPV[h] (95% CI)	LR[i]+ (95% CI)	LR-[j] (95% CI)	Accuracy (95% CI)
MCI[k] vs HC[l]	34	20	7	6	85.0 (70.2 - 94.3)	74.1 (53.7 - 88.9)	82.9 (67.9 - 92.8)	76.9 (56.4 - 91.0)	3.28 (1.71 - 6.28)	0.20 (0.09 - 0.44)	80.6 (69.1 - 89.2)
aMCI[m] vs HC	20	20	7	2	90.9 (70.8 - 98.9)	74.1 (53.7 - 88.9)	74.1 (53.7 - 88.9)	90.9 (70.8 - 98.9)	3.51 (1.83 - 6.72)	0.12 (0.03 - 0.47)	81.6 (68.0 - 91.2)
MCI vs SCD[n]	34	14	3	6	85.0 (70.2 - 94.3)	82.4 (56.6 - 96.2)	91.9 (78.1 - 98.3)	70.0 (45.7 - 88.1)	4.82 (1.71 - 13.56)	0.18 (0.08 - 0.39)	84.2 (72.1 - 92.5)
aMCI vs SCD	20	14	3	2	90.9 (70.8 - 98.9)	82.4 (56.6 - 96.2)	87.0 (66.4 - 97.2)	87.5 (61.7 - 98.4)	5.15 (1.83 - 14.51)	0.11 (0.03 - 0.42)	87.2 (72.6 - 95.7)
Confirmed ND[o] vs unconfirmed ND (28.5 cutoff)	13	10	14	2	86.7 (59.5 - 98.3)	41.7 (22.1 - 63.4)	48.1 (28.7 - 68.1)	83.3 (51.6 - 97.9)	1.49 (1.00 - 2.20)	0.32 (0.08 - 1.26)	59.0 (42.1 - 74.4)
Confirmed ND vs unconfirmed ND (24.5 cutoff)	12	18	6	3	80.0 (51.9 - 95.7)	75.0 (53.3 - 90.2)	66.7 (41.0 - 86.7)	85.7 (63.7 - 97.0)	3.20 (1.53 - 6.69)	0.27 (0.09 - 0.75)	76.9 (60.7 - 88.9)

[a]TP: true positive.

[b]TN: true negative.

[c]FP: false positive.

[d]FN: false negative.

[e]Sens: sensitivity.

[f]Spec: specificity.

[g]PPV: positive predictive value.

[h]NPV: negative predictive value.

[i]LR : positive likelihood ratio.

[j]LR : negative likelihood ratio.

[k]MCI: mild cognitive impairment.

[l]HC: healthy controls.

[m]aMCI: amnestic mild cognitive impairment.

[n]SCD: subjective cognitive decline.

[o]ND: neurodegeneration.

Internal validation using 1000 bootstrap resamples confirmed the robustness of these diagnostic estimates. The optimism-corrected AUCs were virtually identical to the apparent values for all comparisons (eg, aMCI vs SCD: apparent AUC=0.943 vs optimism-corrected AUC=0.942; 95% CI 0.866 - 0.989). Sensitivity and specificity at the 28.5 cutoff remained stable after correction (aMCI vs SCD: sensitivity 90.9% vs 90.7%; specificity 82.4% vs 82.5%), indicating minimal overfitting and supporting the generalizability of the findings within the target population (see Table S2 in Multimedia Appendix 5 for complete bootstrap results). LOOCV yielded consistent results, further reinforcing the stability of the model.

Usability and Satisfaction

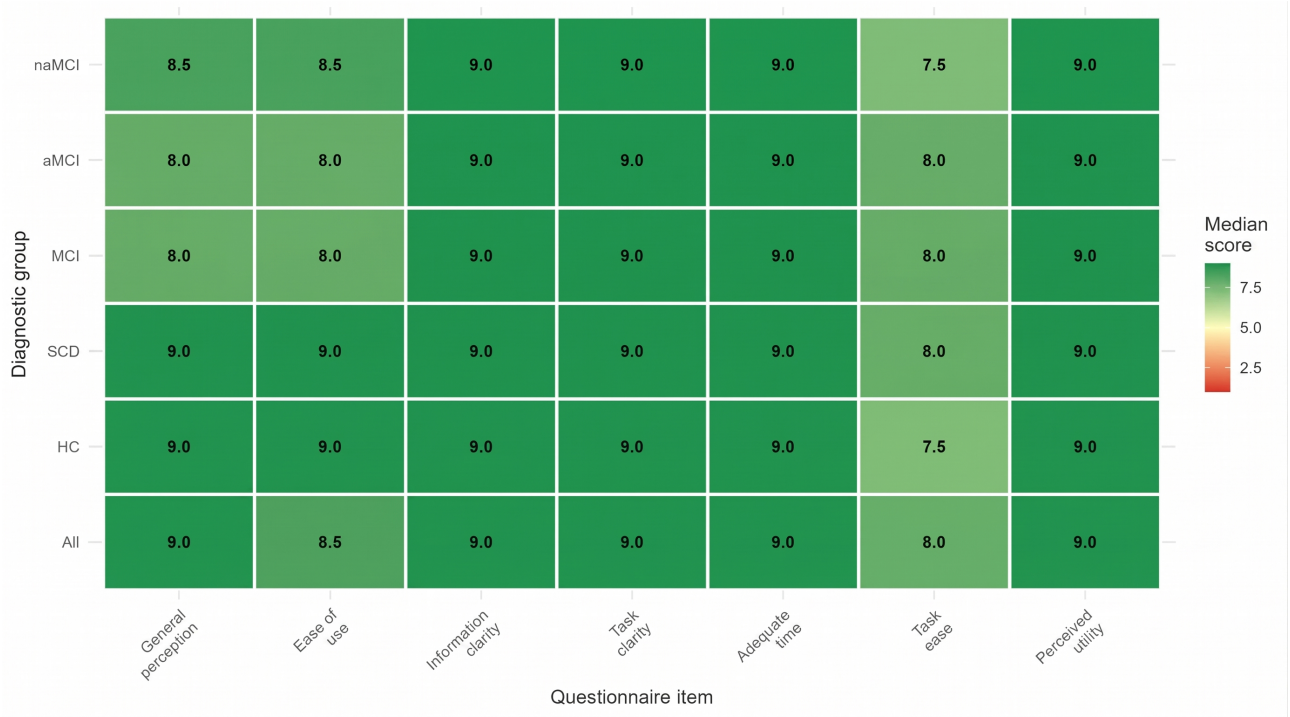
Of the 84 participants, 70 (83.3%) completed the usability questionnaire and 14 (16.7%) did not. Responders and nonresponders did not differ significantly in age, sex, diagnostic group, or MoCA score. Nonresponders more frequently required assistance with the tablet during testing (69.2% vs 41.2%; *P*=.08). Nonresponders also had a significantly higher prevalence of first-degree family history of AD (57.1% vs 25.4%; *P*=.03). The Little MCAR test applied to the 7

XSL•FO
RenderX

questionnaire items alone did not reject the null hypothesis of data missing completely at random (χ^2_6=8.86; *P*=.18); however, when age, sex, family history, need for tablet assistance, and MoCA score were included, the test rejected MCAR (χ^2_{49}=69.77; *P*=.03), in line with the association between nonresponse and family history noted above. Missingness in this questionnaire was therefore better characterized as missing at random (MAR), conditional on observed covariates, rather than missing completely at random. The 7-item questionnaire showed good internal consistency (Cronbach α=0.84, 95% CI 0.78 - 0.89), with all item-rest correlations ≥0.46 and no item improving reliability if removed.

Overall usability ratings were high, with a median score of 8.43 (IQR 7.86 - 8.57) on a 9-point scale. Item-level analysis indicated high ratings across domains, with median scores of 9.0 for “Perceived Utility,” “Information Clarity,” “Task Clarity,” and “Adequate Time“ (all IQR 8.0 - 9.0) across all diagnostic groups. “Tasks Ease“ yielded the lowest median score (median 8.0, IQR 6.25 - 9.0), as illustrated in Figure 8.

Figure 8. Median satisfaction scores by diagnostic group and item. Heatmap colors and cell values represent the median score for each questionnaire item on a 1 - 9 scale. “ALL” indicates the overall study population. aMCI: amnestic mild cognitive impairment; HC: healthy controls; MCI: mild cognitive impairment; naMCI: nonamnestic mild cognitive impairment; SCD: subjective cognitive decline.



While HC and SCD participants reported high and homogeneous satisfaction scores (HC: median 8.50, IQR 8.00-8.61; SCD: median 8.43, IQR 8.29-8.57), the MCI subgroups exhibited wider IQR, reflecting a much broader dispersion in user experience. Specifically, the aMCI subgroup showed a median satisfaction score of 8.17, but with a wider dispersion (IQR 7.29-8.57) compared to the groups with no cognitive impairment; similarly, the nonamnestic MCI phenotype presented a median satisfaction score of 8.00 (IQR 7.61-8.61). Statistical analysis confirmed no significant correlation between age and satisfaction scores (Spearman ρ=–0.16; *P*=.20).

Moreover, the use of the tool required minimal supervision: 54.3% of the cohort completed the protocol with zero assistance, while the remaining participants required only minor guidance (limited to 1 - 2 interventions), which primarily involved assistance with microphone usage or clarification of the task instructions. However, the rate of independent use varied significantly by group: 87.5% for SCD, 69.2% for HC, and only 30.8% for the MCI group (ie, most patients required some support). A significant negative correlation was found between the frequency of assistance and total CogniPaz scores (*r*=–0.505; *P*<.001). Participants requiring guidance achieved significantly lower global scores (mean 21.51, SD 8.44) compared to those who were fully autonomous (mean 29.48, SD 5.17; *P*<.001). No systematic adverse reactions (eg, frustration and fatigue) were reported.

Discussion

Principal Findings

In this study, CogniPaz showed good, concurrent capacity to differentiate MCI, particularly the amnestic phenotype, from HC and SCD. It also exhibited a strong correlation with MoCA and proved feasible for routine clinical use, achieving good user satisfaction and requiring minimal supervision. Finally, the tool suggested a preliminary capacity to identify underlying bND among symptomatic patients.

Related Work

Different digital tools aimed at assessing cognitive disorders are available, with varying degrees of validation [5,6,15]. Established computerized batteries, such as the Cogstate Brief Battery (CBB), which focuses on processing speed and learning, have demonstrated moderate accuracy for discriminating MCI

XSL•FO
RenderX

from cognitively unimpaired participants (AUC≈0.75), achieving improved performance (AUC up to 0.93) only when biomarker-defined prodromal AD is considered [33]. Similarly, BrainCheck (BrainCheck, Inc), a multidomain battery, has reported sensitivities and specificities between 77% and 88% for MCI and dementia classification [34], while Philips IntelliSpace Cognition (ISC) offers tablet-based versions of conventional tests, with evidence of measurement equivalence [35].

Beyond the mere digitization of traditional assessments, recent diagnostic approaches have incorporated automated analysis of process-based metrics and machine learning algorithms to refine early detection [5,6,15,16]. A linear support vector model using digital metrics from the "What was where?" Oxford Memory Task (OMT) has shown high diagnostic accuracy across the AD continuum. This tablet-based tool achieved an AUC of 0.82 in discriminating between HC and those with SCD and an even higher AUC of 0.92 for distinguishing SCD from MCI. While its performance in classifying MCI versus AD dementia was lower (AUC=0.75), the tool's identification accuracy was a top predictor of hippocampal atrophy, performing as effectively as standard neuropsychological screenings [36]. Mobile-based solutions like the Digital Assessment of Cognitive Impairment (DACI) have achieved AUCs around 0.87 - 0.90 in identifying cognitively impaired older adults, with administration times under 5 minutes [37], while ALTOIDA-iADL (Altoida Inc), using augmented reality to simulate instrumental activities of daily living, demonstrated robust diagnostic accuracy (AUC≈0.78) in distinguishing nondegenerative MCI from prodromal AD in patients characterized by cerebrospinal fluid biomarkers [38]. Other emerging digital solutions include remote episodic memory composites, unsupervised paired-associate learning tasks, and app-based screeners such as digiDEM-SCREEN (digiDEM Bayern project, Friedrich-Alexander-Universität Erlangen-Nürnberg) [9,39].

Expanding into self-administered and remote solutions presents both opportunities and trade-offs regarding diagnostic certainty. The Geras Solutions Cognitive Test (GSCT) proved viable for unsupervised screening (AUC=0.740 for MCI vs SCD, AUC=0.960 for detecting dementia) [40,41], while the remote, unsupervised Cambridge Neuropsychological Test Automated Battery (CANTAB) Paired Associates Learning (PAL) demonstrated feasibility in the Brain Health Registry, despite modest accuracy (AUC=0.66 - 0.68 for self-reported MCI) compared to supervised versions [42]. In parallel, the digitalization of the "gold standard" MoCA has evolved from the eMoCA (which established measurement equivalence with the paper version; concordance correlation coefficient=0.84; mean difference <2 points [43]) to self-administered solutions like MoCA Solo [44] and MoCA XpressO. However, recent validation of XpressO highlights the complexity of unassisted screening: while effective at extremes, it yielded an "indeterminate range" (scores 43 - 71) where diagnostic certainty is reduced, achieving good sensitivity and specificity only outside this broad window [45].

Finally, this digital transformation is exemplified by large-scale initiatives aiming to democratize early detection. Prominent examples include SCAP-AD (Cognitive Screening and Personalized Approach for Alzheimer's and other Dementias, Fundación CIEN) [46], a research-driven project focused on developing screening algorithms for AD in the general population, and industry collaborations, such as "The Mind Guardian" (Samsung) [47], which promotes accessibility and cognitive monitoring.

From a digital health and implementation science perspective, CogniPaz should be considered a minimally supervised mHealth cognitive-assessment device rather than a standalone diagnostic algorithm. As highlighted in recent mHealth literature, the successful adoption of digital systems relies heavily on human-centered design, proof-of-concept validation, and thorough usability assessment [23]. The novelty of CogniPaz lies in applying these core mHealth design principles to a screening environment. Consequently, its primary value proposition is not based on diagnostic superiority alone but on how it overcomes practical implementation barriers. By integrating a multi-domain evaluation into an ecological and gamified format, the brief assessment can be entirely self-administered. As demonstrated in our feasibility analysis, these characteristics allow for minimal oversight from nonspecialist personnel, such as general nurses or accompanying family members, offering a highly scalable triage solution that may help mitigate administrative bottlenecks in real-world clinical settings.

Beyond these tools, CogniPaz offers several advantages based on its unique clinical and design features. While most published digital tools primarily contrast MCI or dementia with normal cognition or examine biomarker-defined AD stages [5,6,13-15,17,20,48], CogniPaz demonstrates good accuracy in distinguishing aMCI—the primary target for early interventions—from SCD, thus allowing identification of clinically relevant early cognitive deficits. Early cognitive deficits represent a diagnostic challenge frequently encountered in specialized settings but are less often studied in digital validation. In this regard, CogniPaz displays good accuracy, identifying a provisional threshold that maximizes sensitivity. While digital adaptations of traditional scales like eMoCA or MoCA XpressO can be hampered by broad 'indeterminate ranges' where diagnostic confidence drops [43,45], CogniPaz provides a provisional threshold for classification. With its good sensitivity and specificity for detecting aMCI versus SCD, it makes it a more practical instrument for rapid clinical triage. Specifically, the tool's likelihood ratios indicate that, while a positive result provides moderate evidence for confirming aMCI, a negative result offers evidence for ruling it out—a particularly valuable combination for clinical decision-making in memory clinic settings. Our findings suggest that CogniPaz may assist in establishing a clinical suspicion of bND versus nonneurodegenerative etiologies, a capability rarely available in brief, digital screening tools. In particular, its preliminary performance in identifying bND etiology suggests a high rule-out capacity that could be useful in clinical triage. Within this biomarker-confirmed subsample, CogniPaz showed higher point-estimate discrimination than the MoCA, and its addition to a model already containing MoCA score and age significantly improved model fit, a pattern that may suggest potential incremental value beyond conventional screening.

XSL•FO
RenderX

Our data indicate that CogniPaz scores are not biased by age, a limitation of many computerized batteries—such as CBB, CANTAB, BrainCheck, and Philips ISC—that require the use of complex, age-stratified normative databases to distinguish pathology from normal aging [6,17]. Since the tool's discriminative power appears to be driven by disease-specific deficits rather than age-related slowing, it allows for direct clinical interpretation without the need for lookup tables. This makes CogniPaz inherently more practical for rapid screening in diverse older adult populations.

Compared to other digital tools that emphasize episodic memory or processing speed [5,6,13-15,17,20,48], CogniPaz integrates memory and executive tasks for simultaneous assessment of memory and executive functions. Its strong correlation with the MoCA suggests that it effectively captures overall cognitive functioning, while its multidomain approach may enhance sensitivity to the early dysexecutive changes often observed in prodromal neurodegenerative syndromes. Notably, CogniPaz provides a more granular evaluation of episodic memory than traditional screens; for instance, while the MoCA uses a 5-word recall task, CogniPaz uses a 6-word list, presented through dual modalities—verbal memory with visual and auditory input. Despite this increased depth and the inclusion of multimodal assessments, the tool remains efficient. With a brief administration time, CogniPaz is competitive with existing digital batteries and shorter than many paper-based screens, successfully balancing diagnostic precision with the brevity required for routine clinical use.

The tool's gamified and ecologically oriented design ensures high user acceptance, a critical requirement for digital health interventions highlighted by previous research [5,6,13-15,17,20,48], especially for older adults and individuals with lower educational levels. In our sample, user satisfaction was high across all domains, and no major usability issues were reported. While some participants acknowledged the inherent difficulty of certain tasks—likely reflecting the subjective challenge associated with objective performance deficits in cognitively impaired individuals—high ratings in "perceived utility" across all diagnostic groups indicate that patients recognized and valued the tool's clinical relevance. Such intuitive and engaging interfaces may mitigate test anxiety and promote long-term engagement, thereby facilitating repeated or longitudinal assessments.

Limitations

Several limitations must be acknowledged. The sample size, particularly for the SCD group, was modest, and confidence intervals around AUC estimates (as well as around LRs) remain relatively wide. This low level of precision reflects the exploratory nature of the study of a novel tool with a modest sample size, broadly comparable to other early-stage digital cognitive assessment-tool validation studies [19,34,39,44]. Nevertheless, the robustness of these preliminary findings is supported by the study's prospective design and its execution within a real-world clinical setting. That said, larger, multicenter, prospective validation studies are necessary to confirm these findings and ensure their generalizability across more diverse populations and clinical environments.

Consistent with real-world clinical demographics, MCI participants were significantly older than SCD participants. While age differences can theoretically confound cognitive performance, statistical comparison of ROC curves demonstrated that adding age to the predictive models did not affect accuracy for any comparison. Furthermore, analysis of explained variance revealed that CogniPaz scores accounted for most of the predictive power, whereas age contributed marginally. This supports the tool's validity as a disease-detection instrument, independent of demographic confounders.

Regarding clinical characterization, we acknowledge a lack of systematic quantification of potential confounders. Unlike explanatory trials with strict exclusionary criteria, our study did not exclude patients with common comorbidities, such as depressive symptoms, vascular pathology, or sensory deficits, nor did we grade their severity using standardized scales. While this decision preserves the ecological validity of a "real-world" clinical cohort, it limits our ability to determine the extent to which these nonneurodegenerative factors—rather than incipient neurodegeneration—may have contributed to performance variability, particularly in the SCD group.

Biomarker data (eg, amyloid/tau PET and CSF markers) were not systematically available for the entire cohort, as this was not the primary aim of the study. This limitation reflects the real-world setting of a tertiary memory clinic, where many participants were patients with mild symptoms, attending their first specialist consultation, at which point invasive or expensive confirmatory tests had not yet been performed. Consequently, our preliminary biological validation represents an exploratory subanalysis. While we identified a subset of bND cases with good sensitivity, the small sample size limits the statistical power and generalizability of these findings. This may have resulted in an underestimation of the tool's diagnostic accuracy in this group, as the "unconfirmed" group may include individuals with latent neurodegenerative pathology who had not yet undergone definitive testing. These preliminary findings warrant future studies with systematic biomarker profiling before establishing CogniPaz as a digital proxy for bND and should be regarded as hypothesis-generating rather than conclusive evidence of added value.

The composite bND outcome combined heterogeneous biomarker modalities (MRI, amyloid-PET, CSF, and FDG-PET) that were obtained through routine clinical judgment rather than a uniform protocol. Participants were not verified with the same set of tests, and the modalities differ in their sensitivity and specificity for detecting neurodegeneration, so the "confirmed" and "unconfirmed" labels may not represent a fully consistent construct across participants.

The usability questionnaire used was an ad hoc, nonvalidated instrument rather than a standardized tool integrating both researcher and patient perspectives, allowing a more comparable, reproducible appraisal [49]. The 7-item scale nonetheless showed good internal consistency, supporting its interpretation as a coherent construct within this study. Furthermore, a minority of participants did not complete the questionnaire. While responders and nonresponders did not differ significantly on most measured demographic or cognitive variables, the Little

XSL•FO
RenderX

MCAR test formally rejected the assumption of data missing completely at random, indicating that nonresponse was systematically related to observed participant characteristics rather than occurring purely by chance. This reinforces the need for caution when interpreting aggregate usability scores. Future iterations of CogniPaz should incorporate a validated usability instrument administered with strategies to minimize missing data, particularly among participants at higher risk of non-response.

We also did not include longitudinal follow-up in this initial, preliminary validation. Consequently, we were unable to estimate the test-retest reliability of the CogniPaz battery, and we cannot yet determine the prognostic value of CogniPaz scores for predicting conversion from SCD to MCI or from MCI to dementia. Longitudinal digital assessments, including remote or home-based administration to formally assess test-retest reliability and clinical trajectories, could be considered for future work.

We acknowledge a potential incorporation bias, as the MoCA test was utilized as one of the operational criteria for defining the aMCI group. Theoretically, validating a new tool against a standard that forms part of the diagnosis could inflate diagnostic accuracy estimates. However, the clinical classification of aMCI was not determined by MoCA cutoffs alone but through a comprehensive clinical adjudication process involving the CDR, functional history, and expert neurological consensus. Therefore, the strong relationship observed between CogniPaz and MoCA should be interpreted as evidence of solid concurrent validity against the current standard of care, demonstrating that our digital tool successfully captures the same cognitive constructs as traditional, paper-based screening, without being solely a statistical artifact of the inclusion criteria. Beyond incorporation bias in the statistical sense, the index test and the reference standard tap partially overlapping cognitive constructs by design: both CogniPaz and the MoCA assess verbal word-list recall, and aMCI is itself defined by predominant episodic memory impairment. Part of the observed correlation and of the discriminative accuracy for aMCI may therefore reflect shared method rather than fully independent validity.

The optimal cutoff score of 28.5 points was derived and tested within the same cohort, introducing a risk of overfitting. Although internal validation using bootstrap resampling suggested stability of the estimates, this does not replace external validation. Consequently, this threshold should be considered provisional and hypothesis-generating, and its generalizability to other populations and clinical settings remains unproven. The enrollment design presents an inherent risk of spectrum bias. Because this study was conducted within a tertiary memory clinic, the clinical spectrum of our cohort may not fully mirror that of general community or primary care populations where digital screening tools are frequently deployed. Using a convenience sample of accompanying individuals as HC could artificially amplify the contrast between typical aging and cognitive impairment. The relationship of each control to a study patient (eg, spouse, other relative, or unrelated volunteer) was not systematically recorded, so shared lifestyle, socioeconomic, or genetic factors within this group could not be assessed in more detail. This may be partly mitigated by the study's focus on differentiating subtle cognitive states, such as aMCI from SCD, rather than extreme phenotypes. Independent confirmation of both the cutoff and the tool's generalizability in larger, multicenter cohorts spanning primary care and community settings is therefore an essential prerequisite before clinical implementation can be recommended.

We were unable to calculate the word error rate (WER) or perform a post hoc inter-rater agreement analysis between the automated system and manual transcriptions. To prioritize participant privacy and satisfy regulatory constraints, the application was architected to fully discard all audio recordings and raw speech-recognition text strings immediately after task execution. While this design prevents retrospective acoustic or linguistic auditing, the platform demonstrated reliable technical feasibility during clinical evaluation, with zero technical failures or data losses. Potential automatic speech recognition (ASR) errors were mitigated in real time through live user-interface safeguards (including real-time word visualization, the ability to manually delete misrecognized items, and the option to retry the task), requiring only minimal examiner supervision. Nevertheless, future iterations of the platform should incorporate prospective, item-level logging alongside parallel manual validation protocols to formally quantify ASR diagnostic accuracy.

Finally, this study focused on supervised administration in a specialized hospital setting; we did not examine unsupervised or remote use, which is increasingly relevant for population-level screening and monitoring [17]. In our study, the need for external support was not attributable to interface complexity but was strongly driven by the participants' cognitive status. Collectively, these findings suggest that requests for assistance reflect the executive and memory deficits inherent to the MCI condition, rather than limitations in the clarity or usability of the CogniPaz platform. Additional work is needed to determine whether the gamified interface of CogniPaz translates into consistent performance under remote conditions and how best to adapt or simplify tasks, if needed.

Future Directions

Future research on CogniPaz should pursue several lines. First, larger normative datasets stratified by age, education, and culture are necessary to refine cutoffs and ensure fair interpretation across populations. Furthermore, longitudinal validation through repeated assessments in SCD and MCI cohorts will help determine whether CogniPaz scores and trajectories predict clinical progression or stability. Integrating these digital metrics with multimodal biomarkers—including neuroimaging, fluid and serum markers, and genetic risk profiles—may further enhance early detection and allow for the derivation of composite risk scores. Finally, evaluating CogniPaz in nonspecialized settings is a critical next step to establish its utility as a first-line triage tool; to this end, future validation studies will be extended to the primary care settings within the La Paz University Hospital health area to assess the tool's effectiveness and feasibility as a scalable screening instrument in real-world community practice.

XSL•FO
RenderX

Conclusions

CogniPaz is an innovative mHealth tool for early screening of cognitive impairment, designed as a multidomain, gamified, and ecologically oriented framework that includes multimodal memory and executive tasks. It differs from traditional screening tools by offering a brief, self-administered format that requires minimal supervision, makes use of the inherent advantages of digital platforms, and targets the clinically demanding aMCI-versus-SCD boundary. In doing so, it brings to the field a practical, user-friendly instrument capable of addressing shortcomings of classical screening tools. CogniPaz shows good concurrent capacity to distinguish MCI, particularly the amnestic phenotype, from SCD and HC, alongside high user acceptance. Additionally, it showed preliminary, hypothesis-generating capacity to identify bND. These results require further external, multicenter, longitudinal validation to confirm the diagnostic capacity and prognostic value of the tool in daily practice. If validated, this digital approach may provide more efficient pathways for cognitive assessment in primary care and earlier triage of patients with cognitive decline to specialized care.

Acknowledgments

The authors thank the La Paz Hospital Institute for Health Research and the Spanish Ministerio de Sanidad for their institutional support.

Generative AI Use

The authors declare the use of generative AI in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: literature search and systematization, writing the literature review, code generation, code optimization, creation of algorithms for data analysis, data curation and organization, data analysis, text generation, proofreading and editing, summarizing text, formulation of conclusions, translation, reformatting, bias analysis and potential discrimination assessment, ethical risk analysis, monitoring compliance with ethical standards, quality assessment, trend identification, identification of limitations, recommendations, and publication support.

The GenAI tools used were Claude (Anthropic, Sonnet 4.6), DeepSeek (R1), ChatGPT (OpenAI, 4.1), and Gemini (Google, flash, pro).

Responsibility for the final manuscript lies entirely with the authors.

GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

Declaration submitted by: CE-A and AM-M

Funding

This study has been funded by Instituto de Salud Carlos III (ISCIII) through the project "PI25/01343" to AM-M and co-funded by the European Union. This grant also supported the research contract of CE-A. This competitive public research grant covered the article processing charge (APC). Besides, this work was supported by a “Juan Rodés” contract "JR24/00004" to EA-L, funded by Instituto de Salud Carlos III (ISCIII) and co-funded by the European Union. Role of the funder: the sponsor had no role in the design of the study; collection, analysis, or interpretation of data; writing of the manuscript; or decision to submit the manuscript for publication.

Data Availability

The datasets and analysis code generated during the current study are available from the corresponding author upon reasonable request. Data are shared in anonymized form in accordance with institutional ethics requirements.

Authors' Contributions

Conceptualization: CE-A, AM-L, DB-d-M, MA-d-L, AM-M
Data curation: CE-A, AM-L, DB-d-M, IL-G
Formal analysis: CE-A, DB-d-M, IL-G
Funding acquisition: CE-A, AF-G, AM-M
Investigation: CE-A, AM-L, DB-d-M, IL-G, MH-B, EA-L, SH-R, OO-P, AM-M
Methodology: CE-A, AM-L, MA-d-L, AM-M
Resources: CE-A, AM-M
Software: DB-d-M, AM-M
Supervision: CE-A, AM-M
Validation: CE-A, IL-G, MA-d-L, AM-M
Writing–original draft: CE-A
Writing–review & editing: CE-A, IL-G, MH-B, EA-L, SH-R, GRZ-M, MA-d-L, AF-G, AM-M

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Video demonstration of the CogniPaz tool in use (Spanish).
[MP4 File, 168330 KB - mhealth_v14i1e99688_app1.mp4]

Multimedia Appendix 2
Technical software documentation, architecture diagrams, data-flow descriptions, and scoring workflows.
[DOC File, 1144 KB - mhealth_v14i1e99688_app2.DOC]

Multimedia Appendix 3
Sanitized source-code artifact of the CogniPaz application.
[ZIP File, 27473 KB - mhealth_v14i1e99688_app3.zip]

Multimedia Appendix 4
Table S1, complete threshold-performance data across all candidate cutoffs (24.5, 26.5, 28.5) for every comparison, including all metrics with 95% CI.
[DOCX File, 19 KB - mhealth_v14i1e99688_app4.docx]

Multimedia Appendix 5
Table S2, complete bootstrap validation results.
[XLSX File, 8 KB - mhealth_v14i1e99688_app5.xlsx]

Checklist 1
Completed STARD 2015 checklist.
[PDF File, 112 KB - mhealth_v14i1e99688_app6.pdf]

References

1. Petersen RC, Caracciolo B, Brayne C, Gauthier S, Jelic V, Fratiglioni L. Mild cognitive impairment: a concept in evolution. J Intern Med 2014 Mar;275(3):214-228. [doi: 10.1111/joim.12190] [Medline: 24605806]
2. Oltra-Cucarella J, Ferrer-Cascales R, Alegret M, et al. Risk of progression to Alzheimer's disease for different neuropsychological mild cognitive impairment subtypes: a hierarchical meta-analysis of longitudinal studies. Psychol Aging 2018 Nov;33(7):1007-1021. [doi: 10.1037/pag0000294] [Medline: 30284855]
3. Jessen F, Amariglio RE, Buckley RF, et al. The characterisation of subjective cognitive decline. Lancet Neurol 2020 Mar;19(3):271-278. [doi: 10.1016/S1474-4422(19)30368-0] [Medline: 31958406]
4. Petersen RC. Mild cognitive impairment. Continuum (Minneap Minn) 2016 Apr;22(2 Dementia):404-418. [doi: 10.1212/CON.0000000000000313] [Medline: 27042901]
5. Cornacchia E, Bonvino A, Scaramuzzi GF, et al. Digital screening for early identification of cognitive impairment: a narrative review. Wiley Interdiscip Rev Cogn Sci 2025;16(4):e70009. [doi: 10.1002/wcs.70009] [Medline: 40616364]
6. Henkel C, Seibert S, Nichols Widmann C. Current advances in computerized cognitive assessment for mild cognitive impairment and dementia in older adults: a systematic review. Dement Geriatr Cogn Disord 2025;54(2):109-119. [doi: 10.1159/000541627] [Medline: 39342930]
7. Thompson LI, De Vito AN, Kunicki ZJ, et al. Psychometric and adherence considerations for high-frequency, smartphone-based cognitive screening protocols in older adults. J Int Neuropsychol Soc 2024 Oct;30(8):785-793. [doi: 10.1017/S1355617724000328] [Medline: 39301587]
8. Butler PM, Yang J, Hobbs M, et al. Digital detection of MCI: single‐handed smartphone screening at‐scale. Alzheimers Dement 2025 Dec;21(S2):e106559. [doi: 10.1002/alz70856_106559]
9. Berron D, Glanz W, Clark L, et al. A remote digital memory composite to detect cognitive impairment in memory clinic samples in unsupervised settings using mobile devices. NPJ Digit Med 2024 Mar 26;7(1):79. [doi: 10.1038/s41746-024-00999-9] [Medline: 38532080]
10. Sabbagh MN, Boada M, Borson S, et al. Rationale for early diagnosis of mild cognitive impairment (MCI) supported by emerging digital technologies. J Prev Alzheimers Dis 2020;7(3):158-164. [doi: 10.14283/jpad.2020.19] [Medline: 32463068]
11. Thompson LI, Strenger JR, Emrani S, et al. 91 remote smartphone-based assessment predicts standard neuropsychological test performance and cerebral amyloid status in cognitively normal older adults. J Int Neuropsychol Soc 2023 Nov;29(s1):493-494. [doi: 10.1017/S1355617723006392]

XSL•FO
RenderX

12. Sánchez-Cabaco A, Palacios-Vicario B, De La Torre L, García-García R, Cacho Gutiérrez J, Prieto Fernández P. Screening cognitive impairment in older adults: an ICT-based study. J Ageing Longev 2025;5(2):15. [doi: 10.3390/jal5020015]
13. Ding Z, Lee TL, Chan AS. Digital cognitive biomarker for mild cognitive impairments and dementia: a systematic review. J Clin Med 2022 Jul 19;11(14):4191. [doi: 10.3390/jcm11144191] [Medline: 35887956]
14. Bonvino A, Cornacchia E, Scaramuzzi GF, et al. Digital tools for mild cognitive impairment: a systematic review and meta-analysis of diagnostic accuracy and methodological challenges. Neuropsychol Rev 2025 Oct 22:1-13. [doi: 10.1007/s11065-025-09680-8] [Medline: 41126003]
15. Magno M, Martins AI, Pais J, Silva AG, Rocha NP. Diagnostic accuracy of digital solutions for screening for cognitive impairment: a systematic review and meta-analysis. Applied Sciences 2024;14(6):2640. [doi: 10.3390/app14062640]
16. Molinari-Ulate M, Mahmoudi A, Parra-Vidales E, Muñoz-Sánchez JL, Franco-Martín MA, van der Roest HG. Digital health technologies supporting the application of comprehensive geriatric assessments in long-term care settings or community care: a systematic review. Digit Health 2023;9:20552076231191008. [doi: 10.1177/20552076231191008] [Medline: 37529535]
17. Tsoy E, Zygouris S, Possin KL. Current state of self-administered brief computerized cognitive assessments for detection of cognitive disorders in older adults: a systematic review. J Prev Alzheimers Dis 2021;8(3):267-276. [doi: 10.14283/jpad.2021.11] [Medline: 34101783]
18. Fowe IE, Sanders EC, Boot WR. Understanding barriers to the collection of mobile and wearable device data to monitor health and cognition in older adults: A scoping review. AMIA Jt Summits Transl Sci Proc 2023;2023:186-195. [Medline: 37350920]
19. Wu M, Feng J, Sun R, et al. Validity and usability for digital cognitive assessment tools to screen for mild cognitive impairment: a randomized crossover trial. J Neuroeng Rehabil 2025 Jun 11;22(1):132. [doi: 10.1186/s12984-025-01665-1] [Medline: 40500744]
20. Cubillos C, Rienzo A. Digital cognitive assessment tests for older adults: systematic literature review. JMIR Ment Health 2023 Dec 8;10:e47487. [doi: 10.2196/47487] [Medline: 38064247]
21. Libon DJ, Swenson R, Jannati A, et al. A 10‐minute digitally administered and scored neuropsychological protocol assessing executive control and episodic memory. Alzheimer's & Dementia 2023 Dec;19(S11):e082141. [doi: 10.1002/alz.082141]
22. Young SR, Shono Y, Hauner KK, et al. Psychometric properties of MyCog 2.0: a human‐centered cognitive screening tool for older adults. Int J Methods Psych Res 2025 Jun;34(2). [doi: 10.1002/mpr.70020]
23. Balderas-Díaz S, Rodríguez-Fórtiz MJ, Garrido JL, Bellido-González M, Guerrero-Contreras G. A psycho-educational intervention programme for parents with SGA foetuses supported by an adaptive mHealth system: design, proof of concept and usability assessment. BMC Med Inform Decis Mak 2022 Nov 11;22(Suppl 4):291. [doi: 10.1186/s12911-022-02036-9] [Medline: 36357878]
24. Arjol D, Unzueta-Arce J, Mora-Simón S, et al. Use and psychometric properties of the MoCA in Spain: a systematic review with meta-analysis and meta-regression. Neuropsychol Rev 2026 Jul 1:1-27. [doi: 10.1007/s11065-026-09703-y] [Medline: 42380691]
25. Martinelli JE, Cecato JF, Bartholomeu D, Montiel JM. Comparison of the diagnostic accuracy of neuropsychological tests in differentiating Alzheimer's disease from mild cognitive impairment: can the montreal cognitive assessment be better than the cambridge cognitive examination? Dement Geriatr Cogn Dis Extra 2014 May;4(2):113-121. [doi: 10.1159/000360279] [Medline: 24987399]
26. Roalf DR, Moberg PJ, Xie SX, Wolk DA, Moelter ST, Arnold SE. Comparative accuracies of two common screening instruments for classification of Alzheimer's disease, mild cognitive impairment, and healthy aging. Alzheimers Dement 2013 Sep;9(5):529-537. [doi: 10.1016/j.jalz.2012.10.001] [Medline: 23260866]
27. Clarens MF, Calandri I, Helou MB, Martín ME, Chrem Méndez P, Crivelli L. Utility of a screening test (MoCa) to predict amyloid physiopathology in mild cognitive impairment. J Appl Cogn Neurosci 2020;1(1):87-91. [doi: 10.17981/JACN.1.1.2020.13]
28. Winblad B, Palmer K, Kivipelto M, et al. Mild cognitive impairment--beyond controversies, towards a consensus: report of the International Working Group on mild cognitive impairment. J Intern Med 2004 Sep;256(3):240-246. [doi: 10.1111/j.1365-2796.2004.01380.x] [Medline: 15324367]
29. Petersen RC. Mild cognitive impairment as a diagnostic entity. J Intern Med 2004 Sep;256(3):183-194. [doi: 10.1111/j.1365-2796.2004.01388.x] [Medline: 15324362]
30. Jessen F, Amariglio RE, van Boxtel M, et al. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. Alzheimers Dement 2014 Nov;10(6):844-852. [doi: 10.1016/j.jalz.2014.01.001] [Medline: 24798886]
31. Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging‐Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011 May;7(3):270-279. [doi: 10.1016/j.jalz.2011.03.008] [Medline: 21514249]
32. Jack CR Jr, Bennett DA, Blennow K, et al. A/T/N: an unbiased descriptive classification scheme for Alzheimer disease biomarkers. Neurology 2016 Aug 2;87(5):539-547. [doi: 10.1212/WNL.0000000000002923] [Medline: 27371494]

XSL•FO
RenderX

33. Alden EC, Pudumjee SB, Lundt ES, et al. Diagnostic accuracy of the cogstate brief battery for prevalent MCI and prodromal AD (MCI A+ T+) in a population-based sample. Alzheimers Dement 2021;17:584-594. [doi: 10.1002/alz.12219]
34. Ye S, Sun K, Huynh D, et al. A computerized cognitive test battery for detection of dementia and mild cognitive impairment: instrument validation study. JMIR Aging 2022 Apr 15;5(2):e36825. [doi: 10.2196/36825] [Medline: 35436212]
35. Vermeent S, Spaltman M, van Elswijk G, Miller JB, Schmand B. Philips IntelliSpace Cognition digital test battery: equivalence and measurement invariance compared to traditional analog test versions. Clin Neuropsychol 2022 Nov;36(8):2278-2299. [doi: 10.1080/13854046.2021.1974565] [Medline: 34528868]
36. Toniolo S, Attaallah B, Maio MR, et al. Performance and validation of a digital memory test across the Alzheimer's disease continuum. Brain Commun 2025;7(1):fcaf024. [doi: 10.1093/braincomms/fcaf024] [Medline: 39886066]
37. Ko HJ, Kim JS, Kim W, et al. Optimizing mobile cognitive assessment reduces administration time while maintaining screening accuracy in older adults. Sci Rep 2025 Aug 19;15(1):30356. [doi: 10.1038/s41598-025-14130-9] [Medline: 40830390]
38. Tort‐Merino A, Tarnanas I, Bügler M, et al. ALTOIDA‐iADL for the diagnosis of mild cognitive impairment and early Alzheimer's disease. Alzheimers Dement 2021 Dec;17(S12):e057982. [doi: 10.1002/alz.057982]
39. Zeiler M, Dietzel N, Kammerer K, et al. Digital screening tool for the assessment of cognitive impairment in unsupervised setting-digiDEM-SCREEN: study protocol for a validation study. BMJ Open 2025 Jan 11;15(1):e087256. [doi: 10.1136/bmjopen-2024-087256] [Medline: 39800401]
40. Bloniecki V, Hagman G, Ryden M, Kivipelto M. Digital screening for cognitive impairment - a proof of concept study. J Prev Alzheimers Dis 2021;8(2):127-134. [doi: 10.14283/jpad.2021.2] [Medline: 33569558]
41. Bloniecki V, Ulfvarson J, Javanshiri K, Hagman G, Freund-Levi Y, Nordström A. The geras solutions cognitive test for assessing cognitive impairment: normative data from a population-based cohort. J Prev Alzheimers Dis 2023;10(2):207-211. [doi: 10.14283/jpad.2023.9] [Medline: 36946447]
42. Ashford MT, Aaronson A, Kwang W, et al. Unsupervised online paired associates learning task from the Cambridge Neuropsychological Test Automated Battery (CANTAB®) in the Brain Health Registry. J Prev Alzheimers Dis 2024;11(2):514-524. [doi: 10.14283/jpad.2023.117] [Medline: 38374758]
43. Berg JL, Durant J, Léger GC, Cummings JL, Nasreddine Z, Miller JB. Comparing the electronic and standard versions of the Montreal Cognitive Assessment in an outpatient memory disorders clinic: a validation study. J Alzheimers Dis 2018 Feb 6;62(1):93-97. [doi: 10.3233/JAD-170896]
44. Gruber J, Krieger J, Jansen R, et al. A Usability Study of MoCA solo, a digital AI‐assisted version of the MoCA test. Alzheimer Dement 2025 Dec;21(S3). [doi: 10.1002/alz70857_097196]
45. Klil-Drori S, Bodenstein KC, Sun S, et al. Montreal Cognitive Assessment (MoCA) XpressO: validation of a digital self-administered cognitive prescreening tool. J Am Geriatr Soc 2024 Aug;72(8):2516-2522. [doi: 10.1111/jgs.18902] [Medline: 38558263]
46. Research project for the early detection of alzheimer's disease. SCAP-AD | CIEN. URL: https://www.fundacioncien.es/proyectos-de-investigacion/scap-ad [accessed 2026-01-28]
47. The mind guardian: detección temprana del deterioro coginitivo. Samsung España. URL: https://www.samsung.com/es/tecnologiaconproposito/accesibilidad-bienestar/themindguardian/ [accessed 2026-01-28]
48. Bastardo R, Pavão J, Martins AI, et al. A Scoping Review of Digital Solutions That Might Be Used as Cognitive Screening Instruments of Community-Dwelling Older Adults: Procedia Computer Science. Elsevier B.V; 2021:956-963. [doi: 10.1016/j.procs.2021.12.097]
49. Liu P, Wang L, Wang F. Evaluation of Chinese HIV mobile apps by researchers and patients with HIV: quality evaluation study. JMIR Mhealth Uhealth 2024 Jan 26;12:e52573. [doi: 10.2196/52573] [Medline: 38277215]

Abbreviations

AD: Alzheimer disease
aMCI: amnestic mild cognitive impairment
ASR: automatic speech recognition
AUC: area under the curve
Aβ42: amyloid beta 42
Aβ42/40: amyloid beta 42/40 ratio
bND: biologically confirmed any-cause neurodegeneration
CANTAB: Cambridge Neuropsychological Test Automated Battery
CBB: Cogstate Brief Battery
CDR: Clinical Dementia Rating
CSF: cerebrospinal fluid
DACI: Digital Assessment of Cognitive Impairment
FDG-PET: fluorodeoxyglucose positron emission tomography
GDPR: General Data Protection Regulation

XSL•FO
RenderX

GSCT: Geras Solutions Cognitive Test
HC: healthy controls
ISC: IntelliSpace Cognition
LOOCV: leave-one-out cross-validation
LR : positive likelihood ratio
LR : negative likelihood ratio
MAR: missing at random
MCAR: missing completely at random
MCI: mild cognitive impairment
mHealth: mobile health
MMSE: Mini-Mental State Examination
MoCA: Montreal Cognitive Assessment
MRI: magnetic resonance imaging
MTA: medial temporal atrophy
NPV: negative predictive value
OMT: Oxford Memory Task
OR: odds ratio
p-tau: phosphorylated tau
PAL: Paired Associates Learning
PET: positron emission tomography
PPV: positive predictive value
ROC: receiver operating characteristic
SCAP-AD: Cognitive Screening and Personalized Approach for Alzheimer and other dementias
SCD: subjective cognitive decline
STARD: Standards for Reporting of Diagnostic Accuracy Studies
t-tau: total tau
WER: word error rate

Edited by S Brini; submitted 28.Apr.2026; peer-reviewed by G Guerrero-Contreras, P Liu; revised version received 28.Aug.2026; accepted 08.Sep.2026; published 04.Nov.2026.

Please cite as:
Estebas-Armas C, Medina-López A, Blanco-de-Miguel D, Losantos-García I, Hernández-Barral M, Alonso-López E, Henche-Rodríguez S, Ocampo-Pérez O, Zmork-Martínez GR, Alonso-de-Leciñana M, Frank-García A, Martín-Montes Á
CogniPaz, an Ecologically Oriented Gamified mHealth Tool for Detecting Mild Cognitive Impairment: Prospective Cohort Diagnostic Accuracy Study
JMIR Mhealth Uhealth 2026;14:e99688
URL: https://mhealth.jmir.org/2026/1/e99688
doi:10.2196/99688

© Carlos Estebas-Armas, Alejandro Medina-López, Daniel Blanco-de-Miguel, Itsaso Losantos-García, María Hernández-Barral, Elisa Alonso-López, Sergio Henche-Rodríguez, Oscar Ocampo-Pérez, Gerardo Ricardo Zmork-Martínez, María Alonso-de-Leciñana, Ana Frank-García, Ángel Martín-Montes. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 4.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Passive Screening for Depressive Symptoms Using Daily Wrist Actigraphy and Deep Learning: Model Development and Validation Study

Doljinsuren Enkhbayar[1], MS; Somin Oh[1], MS; Jinhee Lee[2], MD, PhD; Min-Hyuk Kim[2], MD, PhD; Erdenebayar Urtnasan[3*], PhD; Jaehong Key[1*], PhD

[1]Department of Biomedical Engineering, Mirae campus, Yonsei University, 1, Yeonsedae-gil, Heungeop-myeon, Wonju-si, Gangwan-do, Republic of Korea

[2]Department of Psychiatry, Wonju College of Medicine, Yonsei University Mirae Campus, Wonju, Gangwon-do, Republic of Korea

[3]Division of AI Semiconductor, Yonsei University Mirae Campus, Wonju, Gangwon-do, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jaehong Key, PhD
Department of Biomedical Engineering, Mirae campus, Yonsei University, 1, Yeonsedae-gil, Heungeop-myeon, Wonju-si, Gangwan-do, Republic of Korea

Abstract

Background: Depressive symptoms are common yet often underrecognized in routine care, underscoring the need for scalable screening approaches beyond episodic self-report assessments. Wearable actigraphy can passively and continuously capture daily activity and 24-hour rest–activity rhythms associated with depressive symptom burden. However, the performance of artificial intelligence (AI) models that leverage actigraphy data for depressive symptom screening remains insufficiently established.

Objective: This study aimed to develop and evaluate AI-based models for passive screening of depressive symptoms from daily wrist actigraphy data.

Methods: We analyzed actigraphy recordings from 1160 Hispanic/Latino adults in the Hispanic Community Health Study/Study of Latinos who completed the 10-item Center for Epidemiologic Studies Depression scale (CESD-10), a self-reported depressive symptom screening scale. Multichannel actigraphy data, including activity counts, light exposure, and wake status, were used as inputs to 5 deep learning architectures to classify CESD-10–defined depressive symptom groups, comparing mild and higher symptoms with the normal group.

Results: Actigraphy-derived behavioral markers differed across depressive symptom groups, showing lower daytime activity and altered circadian rest–activity organization with increasing symptom burden. Among the 5 deep learning architectures evaluated, the long short-term memory model consistently demonstrated the strongest overall discrimination. In held-out testing, the long short-term memory model achieved a macro-averaged area under the receiver operating characteristic curve of 0.80, with the strongest discrimination observed for the higher depressive symptom group (area under the receiver operating characteristic curve 0.889). These findings indicate improved model discrimination with increasing symptom severity, although false-positive rates remained notable across both classification tasks.

Conclusions: Our study suggests that actigraphy-derived data can support AI-based classification of depressive symptoms. An actigraphy-based AI model may serve as a scalable, passive, and noninvasive complementary signal to aid early screening alongside traditional depressive symptom assessments before clinical diagnosis.

(JMIR Mhealth Uhealth 2026;14:e91479) doi:10.2196/91479

KEYWORDS

actigraphy; wearable devices; depressive symptoms; depression screening; artificial intelligence; deep learning; passive monitoring

Introduction

Depression is a highly prevalent and debilitating psychiatric condition globally, with its burden increasing steadily in recent decades [1]. Despite its substantial impact on individuals and society, the mechanisms underlying depression remain complex and incompletely understood. Clinically, depression encompasses a broad spectrum of psychological, behavioral, and somatic symptoms, such as fatigue, sleep disturbances, persistent low mood, changes in appetite, anhedonia, and weight changes, impaired concentration, low self-esteem, and suicidal ideation [2,3]. These manifestations can markedly impair daily functioning, interpersonal relationships, and self-care [4].

XSL•FO
RenderX

Depression nevertheless remains underdiagnosed and undertreated in routine clinical practice, representing a critical gap in mental health care [2]. Many individuals are not identified during standard assessments, and screening is reported to occur in only 4.2% of adults without a prior diagnosis [5]. The nonspecific nature of depressive symptoms, which frequently overlap with somatic complaints such as fatigue or pain, can lead to misdiagnosis or attribution to other medical conditions [6,7]. Consequently, a substantial proportion of affected individuals remain untreated, increasing the likelihood of persistent symptoms and long-term functional impairment.

Given these challenges, there is increasing interest in objective behavioral markers that can complement traditional self-report tools [8]. Sedentary behavior, reduced physical activity, and sleep disruption have been consistently linked to depressive symptoms [9]. However, many studies rely primarily on self-reported measures, which are susceptible to subjectivity and limitations in memory accuracy. In contrast, wearable devices can passively and continuously monitor activity and sleep, providing an objective approach to capturing behavioral changes relevant to depressive symptoms [10].

Passive monitoring modalities differ in their technical characteristics, the type of behavioral information captured, and their degree of standardization. Consumer wearables, such as Fitbit devices [11] or the Apple Watch [12], rely on proprietary, closed-source algorithms that limit methodological transparency and reproducibility. Smartphones capture behavior indirectly through usage-based proxies, such as screen time, GPS mobility [13], and call logs [14], which are susceptible to device heterogeneity and user-specific confounds. Research-grade actigraphy, by contrast, uses validated, open-standard triaxial accelerometry with well-characterized signal processing pipelines, making it particularly suitable for controlled cohort studies and systematic architectural model comparison.

Actigraphy, based on wrist-worn triaxial accelerometers, provides an unobtrusive and continuous method for monitoring physical activity, rest-activity cycles, and light exposure over extended periods [15]. It can reliably estimate sleep-wake patterns and behavioral rhythms, enabling the characterization of reduced activity levels, increased sedentary time, and irregular sleep-wake schedules commonly observed among individuals with elevated depressive symptoms [16]. Evidence from meta-analyses indicates that depressive symptoms are associated with lower overall activity, poor sleep efficiency, longer sleep-onset latency, greater wake after sleep onset, and attenuated circadian amplitude compared with the normal group [17]. Collectively, these actigraphy-derived measures complement questionnaires and clinical interviews by providing a quantitative representation of circadian dysregulation and reduced activity. They may also support more effective screening of depressive symptoms than self-report alone [18].

Recent advances in artificial intelligence (AI) have expanded the utility of actigraphy for screening depressive symptoms [19]. Prior work has ranged from conventional machine learning models trained in manually engineered features, such as step-derived activity summaries and circadian indices, to end-to-end deep learning approaches [20]. While feature-based methods, including gradient-boosted ensembles, can yield moderate performance, they may be less effective at capturing time-resolved structure in actigraphy signals and can be susceptible to overfitting when many correlated predictors are included [21]. Accordingly, deep learning models have been applied directly to raw or minimally processed actigraphy sequences. Existing work has examined convolutional and recurrent architectures [22], including long short-term memory (LSTM) and gated recurrent unit (GRU) models, alongside transformers, Gramian angular field–based convolutional neural networks (CNNs), and mixed-input frameworks that incorporate sleep-stage features or questionnaire data, reporting area under the curve (AUC) values ranging approximately between 0.68 and 0.94 [20]. However, many studies have relied on small or selected samples, focused on extreme symptom groupings, or required additional clinical inputs, limiting generalizability and scalability [23]. Moreover, most studies evaluate a single architecture, which restricts comparative insight into sequence-modeling strategies [24]. Therefore, systematic comparisons of end-to-end deep learning models trained on multichannel actigraphy in large, ethnically diverse cohorts using validated self-report screening tools remain needed.

Despite these advances, actigraphy-based approaches for depressive symptoms have shown modest and variable performance, with reported accuracy and AUC values spanning a wide range across studies [25]. Much of the prior work has been developed using relatively small or demographically narrow samples, which increases the risk of overfitting and may limit the generalizability of findings to broader populations. Moreover, both machine learning and deep learning models can be difficult to interpret, particularly when complex architectures or extensive feature engineering are used, making it challenging to relate predictions to meaningful behavioral or physiological patterns [26]. Limited interpretability can hinder clinical uptake, as clinicians may be reluctant to act on algorithmic outputs without a clear connection to recognizable patterns in patients' daily activities and sleep. Accordingly, there is a need for actigraphy-based screening models that achieve robust discrimination, with transparency and clinical interpretability identified as important directions for future development [27].

We developed and evaluated an AI-based screening approach that leverages wrist-worn actigraphy to identify depressive symptoms, using the 10-item Center for Epidemiologic Studies Depression scale (CESD-10) as the reference screening tool. We trained and compared 5 deep learning architectures on 7 days of multichannel actigraphy time series to classify CESD-10 depressive symptoms. We hypothesized that a higher symptom burden would be reflected in characteristic behavioral signatures, including reduced overall activity and disrupted rest-activity rhythms, which sequence models could learn from raw actigraphy patterns. This study evaluates whether daily actigraphy-based AI models can support objective and unobtrusive screening for depressive symptoms.

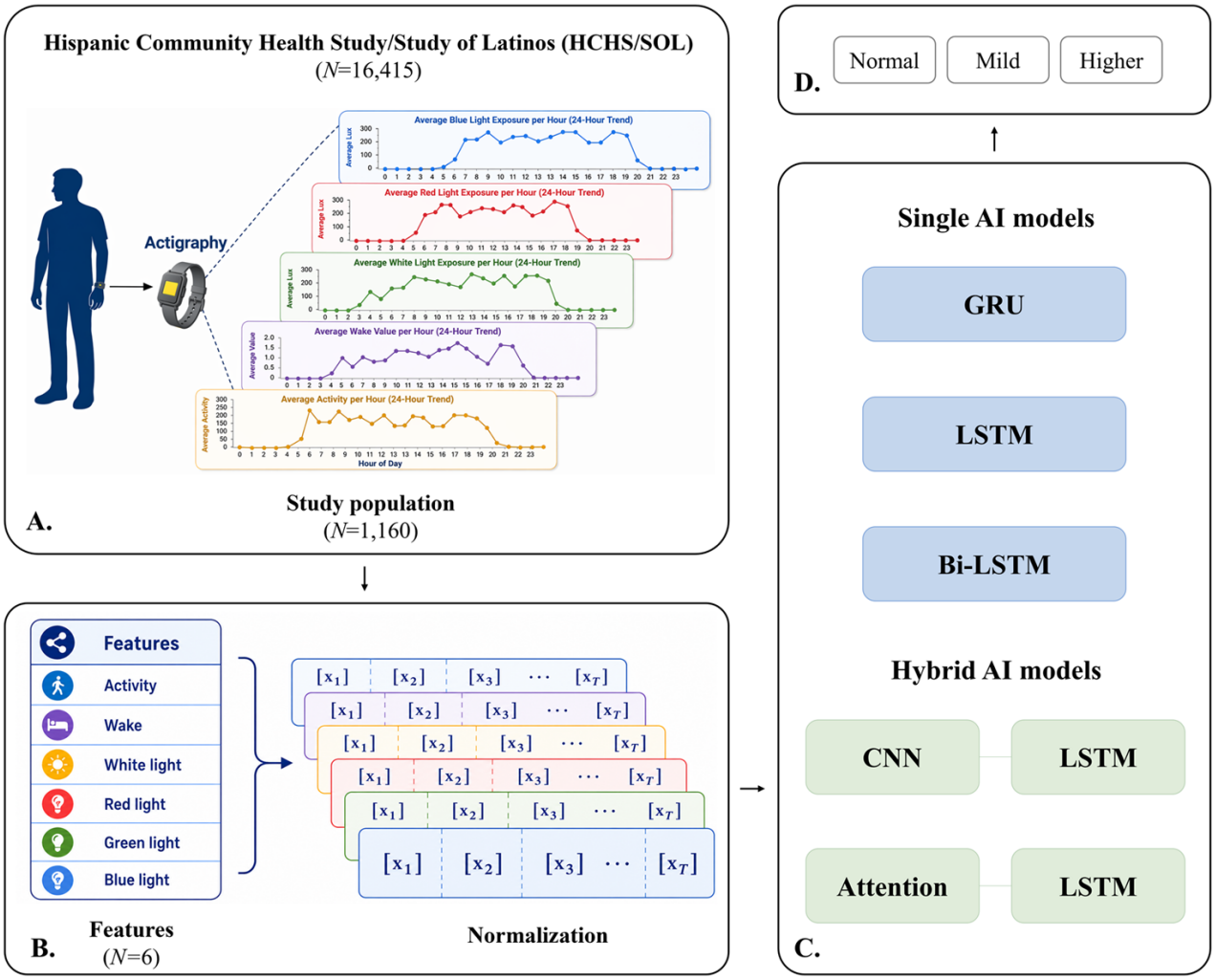
XSL•FO
RenderX

Methods

Study Design and Participants

The Hispanic Community Health Study/Study of Latinos (HCHS/SOL) is a multicenter, population-based cohort study that enrolled 16,415 noninstitutionalized Hispanic/Latino adults aged 18 to 75 years across 4 US field centers in Miami, San Diego, the Bronx, and Chicago. Baseline examinations from 2008 to 2011 included standardized assessments of sociodemographic characteristics, lifestyle factors, medical comorbidities, and cardiometabolic biomarkers [28]. The Sueño ancillary study recruited 2087 adults aged 18 - 65 years from the parent cohort for actigraphy-based assessment of sleep and circadian rhythms (Figure 1A).

Figure 1. Overview of the methodological framework for screening depressive symptoms using actigraphy-derived data. (A) Study design and participants, (B) data processing and feature extraction, (C) artificial intelligence–based models, and (D) output. AI: artificial intelligence, Bi-LSTM: bidirectional long short-term memory, CNN: convolutional neural network, GRU: gated recurrent unit, LSTM: long short-term memory.



In the parent HCHS/SOL Sueño study, exclusion criteria included pregnancy, cognitive dysfunction affecting reliable self-report or informed consent, and severe mobility limitations such as hemiparesis or quadriparesis. Actigraphy was collected over 7 days using an Actiwatch Spectrum device (Philips Respironics) worn on the nondominant wrist to continuously record activity counts, multiwavelength light exposure, and sleep-wake status. After standardized quality-control procedures that excluded recordings with prolonged off-wrist intervals or implausible signal patterns, valid actigraphy data were available for 1887 participants. We then excluded 217 participants with missing demographic or psychometric data, yielding 1670 participants eligible for propensity score matching. Using a 1:1 propensity score matching approach matched on age, sex, and BMI, we derived a final analytic cohort of 1160 participants (71.4% female, aged 18 - 65 years, BMI 21.9-41.2 kg/m²), comprising 580 participants with normal depressive symptom scores and 580 with depressive symptoms. The depressive symptom group was further classified into mild and higher subgroups based on CESD-10 scores (Figure S1 in Multimedia Appendix 1).

All dataset partitioning and model evaluation were performed at the participant level. The matched cohort was then partitioned into training (70%, n=812), validation (10%, n=116), and test (20%, n=232) sets, with no participant appearing in more than one split. A fixed hold-out design was selected over cross-validation because the sequential, time-structured nature of actigraphy data makes random k-fold partitioning susceptible to temporal leakage across participants. The large number of

XSL•FO
RenderX

epoch-level training samples (>17 million) per channel further mitigates overfitting risk despite the moderate participant count. Epoch-level counts for each channel are provided in Table 1 for descriptive purposes only; all model development and performance evaluations were conducted at the participant level.

Data access, documentation, and governance were provided through the National Sleep Research Resource (NSRR) under a controlled-access data use agreement. Institutional review board approvals were in place at all participating sites, and written informed consent was obtained from all participants.

Table . Total epoch counts for each actigraphy channel across the training, validation, and test splits.

Features	Training set	Validation set	Test set	Total
Activity	17,336,985	2,487,580	4,898,740	24,723,305
Wake	17,325,681	2,486,083	4,895,117	24,706,881
White light	17,335,426	2,487,011	4,898,349	24,720,786
RGB[a] light	17,336,985	2,487,580	4,898,740	24,723,305
Total	69,335,077	9,948,254	19,590,946	98,874,277

[a]R: red, G: green, and B: blue light channels have an equal number of epochs.

Depressive Symptoms Assessment

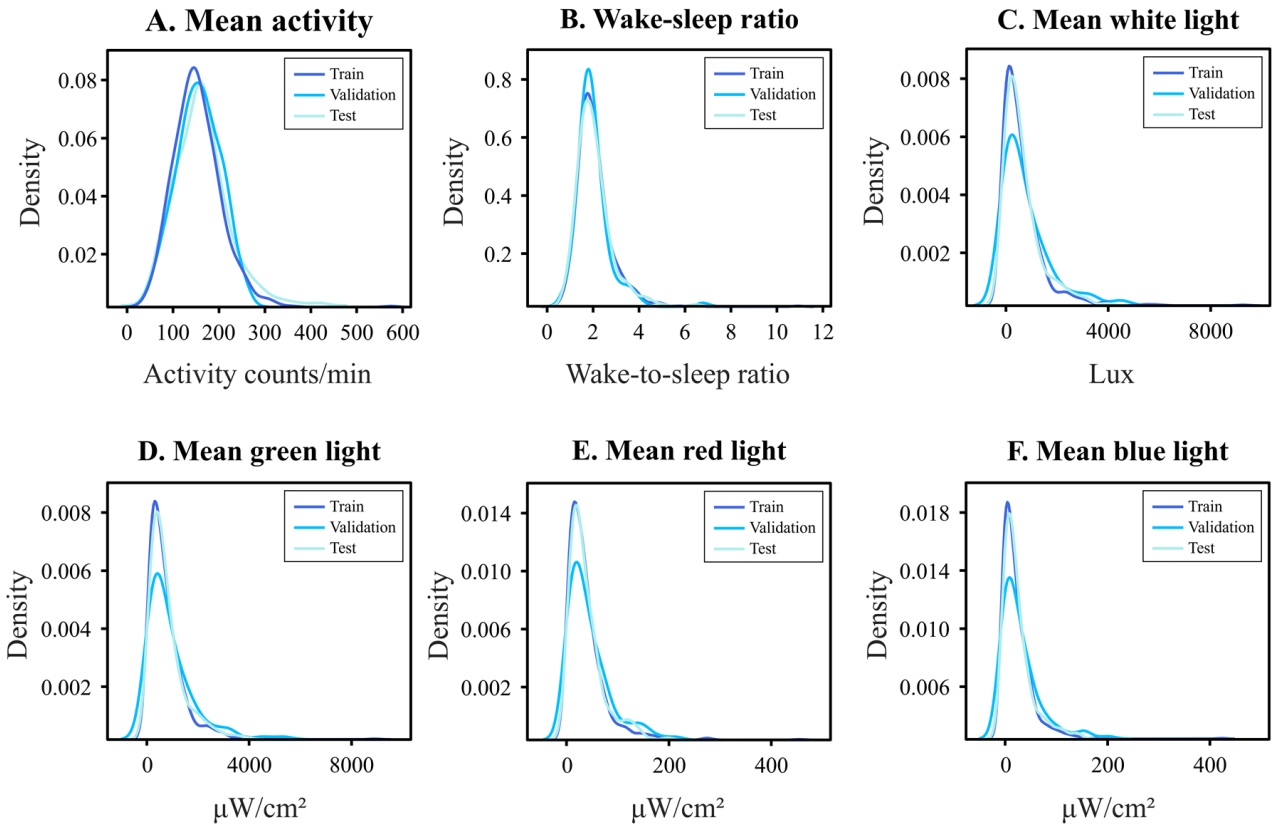
Depressive symptoms were measured using the CESD-10, a validated self-report screening tool widely used in epidemiologic and community-based studies [29]. Items are rated on a 4-point response scale from 0 to 3, yielding a total score from 0 to 30, with higher scores indicating greater depressive symptom burden over the past week [30]. Following prior HCHS/SOL analyses, participants with CESD-10 scores of 10 or higher were classified as having elevated depressive symptoms [31]. For the present analyses, CESD-10 scores were further categorized into three groups to capture graded symptom burden: 0 to 9 (normal), 10 to 14 (mild), and ≥15 (higher depressive symptoms). These categories were used to describe baseline characteristics and to evaluate trends in sociodemographic, behavioral, and clinical factors across increasing symptom severity. CESD-10 has demonstrated good psychometric properties across diverse populations, supporting its use as a community-based screening measure for depressive symptoms [32].

Data Processing and Feature Extraction

Raw wrist actigraphy recordings were preprocessed to support downstream model development and evaluation. For each participant, signals were segmented into 30-second epochs containing activity counts, binary wake status, and light exposures in the white, red, green, and blue channels. Epochs flagged as invalid or off-wrist under the HCHS/SOL quality control criteria were removed before analysis. After data cleaning, the remaining time series were standardized using *z* score normalization. To prevent information leakage, the mean and SD were estimated using the training set only and applied unchanged to the validation and test sets for consistent scaling across splits (Figure 1B). The normalized sequences were segmented into nonoverlapping windows of 2000 consecutive epochs, corresponding to approximately 16.7 hours, providing fixed-length inputs for the deep learning architectures. The resulting inputs comprised the following six synchronized channels: activity (Figure 2A), wake status (Figure 2B), white light exposure (Figure 2C), green light exposure (Figure 2D), red light exposure (Figure 2E), and blue light exposure (Figure 2F). Across the training, validation, and test splits, this yielded approximately 24 to 25 million 30-second epochs per channel in total (Table 1). Minor differences in epoch counts across channels reflect channel-specific quality-control exclusions applied independently during preprocessing. Overall epoch proportions across splits remained consistent with the participant-level 70/10/20 allocation. This segmentation approach provided a large and representative set of training samples while maintaining consistent preprocessing across datasets and minimizing the risk of distributional shifts introduced by data handling.

XSL•FO
RenderX

Figure 2. Density distributions of *z* score–normalized actigraphy features across dataset splits. Kernel density curves for the training, validation, and test sets show substantial overlap, indicating comparable feature distributions across splits and supporting consistent data representation during model development and evaluation. (A) Mean activity, (B) wake-to-sleep ratio, (C) mean white light, (D) mean green light, (E) mean red light, and (F) mean blue light.



To assess consistency across dataset splits, we examined the distributions of key actigraphy-derived features within the training, validation, and test sets (Figure 2). The *z* score–normalized distributions showed substantial overlap, indicating comparable feature profiles across splits and supporting the appropriateness of the partitioning. This consistency helps preserve statistical comparability between datasets and strengthens the reliability of model development and evaluation.

In addition to the raw actigraphy time series used for deep learning, we computed summary measures to describe 24-hour rest–activity rhythms for descriptive analyses and between-group comparisons. For each valid day, mean activity counts were calculated within eight 3-hour clock-time intervals spanning midnight to 2:59 AM through 9 PM to 11:59 PM and then averaged across monitoring days. Using the 30-second activity signal, we quantified circadian rhythm metrics, including intradaily variability, interdaily stability, and relative amplitude derived from the most active 10-hour window (M10) and the least active 5-hour window (L5), along with the onset times of M10 and L5. We also fit a 24-hour cosinor model to each participant's activity profile to estimate the midline statistic of rhythm, log-transformed amplitude, and acrophase. For each metric, values were aggregated across valid days to obtain a participant-level estimate, which was compared across CESD-10 depressive symptom categories.

AI-Based Models

We implemented 5 deep learning architectures to classify depressive symptoms using multichannel wrist actigraphy data, including a GRU [33], LSTM [34], bidirectional LSTM (Bi-LSTM) [35], 1-dimensional convolutional neural network–long short-term memory (CNN-LSTM) [36], and attention-LSTM [37] (Figure 1C). These architectures were selected to represent diverse sequential modeling strategies, ranging from conventional recurrent networks to bidirectional and hybrid designs, thereby enabling a systematic comparison of their ability to capture temporal dependencies in multichannel actigraphy data. Each model leveraged different modeling strengths, from local feature extraction and long-range temporal dependency learning to bidirectional sequence modeling, enabling identification of the most effective approach for classifying depressive symptom severity from wrist-worn sensor data.

All models received input sequences of 2000 time steps with 6 synchronized channels, including activity counts, wake status, and multiwavelength light exposure (white, red, green, and blue), and produced a 3-class probability distribution through a softmax activation function. A dropout rate of 0.3 was selected based on prior LSTM-based studies [34] and preliminary validation experiments comparing rates of 0.2, 0.3, and 0.5, with 0.3 yielding the best validation performance. The final class prediction was determined by selecting the class with the highest softmax probability (Table 2).

XSL•FO
RenderX

Table . Summary of deep learning model architecture and training configurations. All models were trained using the AdamW optimizer with a mini-batch size of 32, focal loss with class-specific weighting, and a Cosine Annealing Warm Restarts learning rate scheduler.

Model	Key components	Units or filters	Learning rate	Max epochs (early stop)	Total parameters
Single AI[b] models					
GRU[f]	4 GRU layers + BN[a] + Dense	60 - 120 units	3×10^{-4}	150 (patience=30)	170,345
LSTM[c]	4 LSTM layers + BN + Dense	60 - 120 units	3×10^{-4}	150 (patience=30)	258,585
Bi-LSTM[d]	4 Bi-LSTM layers + BN + Dense	60 - 120 units (bidirectional)	3×10^{-4}	150 (patience=30)	518,065
Hybrid AI models					
CNN-LSTM[e]	3 Conv1D + MaxPooling + 2 LSTM + Dense	32 - 128 filters, 64 LSTM units	3×10^{-4}	150 (patience=30)	86,817
Attention LSTM	3 Conv1D + MaxPooling + Attention + LSTM + Dense	32 - 128 filters, 64 LSTM units, 128d attention	3×10^{-4}	150 (patience=30)	136,097

[a]BN: batch normalization.

[b]AI: artificial intelligence.

[c]LSTM: long short-term memory.

[d]Bi-LSTM: bidirectional long short-term memory.

[e]CNN-LSTM: convolutional neural network–long short-term memory.

[f]GRU: gated recurrent unit.

Single AI Models

The GRU, LSTM, and Bi-LSTM models were implemented using 4 stacked recurrent layers with 60 - 120 hidden units, followed by batch normalization and a fully connected output layer (Table S2 and Figures S2-S3 in Multimedia Appendix 1). GRU and LSTM models use different gating formulations but are both widely applied to sequential physiological signals. In contrast, the Bi-LSTM processes each sequence in both forward and backward directions, allowing the model to integrate information from both earlier and later time points within the same window (Table S3 and Figure S6 in Multimedia Appendix 1).

Hybrid AI Models

The CNN-LSTM model consisted of three 1D convolutional blocks with filter sizes of 32 - 128, each followed by rectified linear unit activation and max-pooling, and 2 LSTM layers with 64 hidden units (Table S4 and Figure S7 in Multimedia Appendix 1). The attention-LSTM extended this architecture with a single-head attention module before the LSTM layer (Table S5 and Figure S8 in Multimedia Appendix 1).

Model Training

All models were trained for up to 150 epochs using a mini-batch size of 32, with early stopping applied for 30 consecutive epochs. A Cosine Annealing Warm Restarts learning rate scheduler was applied to facilitate convergence. Hyperparameters, including learning rate (3×10^{-4}), hidden units, and dropout rate, were selected through a structured manual search over predefined ranges, with final configurations chosen to maximize validation AUC. To further optimize classification performance, threshold calibration was performed on the validation set using the Nelder-Mead optimization method to maximize macro F_1, with calibrated thresholds applied during test set evaluation.

Class Imbalance

The training cohort reflected moderate class imbalance, with 580 participants classified as normal, 318 as mild, and 262 as higher depressive symptoms (Figure 1D). This was addressed using a combination of 3 strategies. First, sequence-level oversampling was applied to the training set, augmenting minority class samples with Gaussian jitter and random scaling until each class reached 65% of the majority class size (target ratio=0.65). Second, weighted random sampling was applied during mini-batch construction to further balance class representation. Third, focal loss with class-specific weights (mild: 1.8×, higher: 2.0× relative to normal) was used to penalize misclassification of minority classes during training.

Experiments and Evaluations

All experiments were implemented in Python (version 3.11.4; Python Software Foundation), with model development and training carried out using the Keras API within TensorFlow 2.x [38]. Statistical analyses were carried out using SciPy (version 1.8.1; SciPy developers) [39], and machine learning metrics were computed using scikit-learn (version 1.1.2; scikit-learn developers) [40].

Training and evaluation were performed on a Windows 11 workstation equipped with an NVIDIA GTX 1080 Ti GPU. For efficient computation, the dataset was divided into minibatches of 256 samples. Model parameters were updated using gradient accumulation after each minibatch.

XSL•FO
RenderX

Model performance on the test set was summarized using standard classification metrics, including precision, specificity, sensitivity, accuracy, F_1-score, and the area under the receiver operating characteristic curve (AUROC). To assess the robustness of model performance, 95% CIs for AUROC were estimated using bootstrap resampling (1000 iterations) on the test set. All metrics were computed from the test-set confusion matrix. Metric definitions are provided below [41]:

Precision = TP / (TP + FP)

Specificity = TN / (TN + FP)

Sensitivity (Recall) = TP / (TP + FN)

Accuracy = (TP + TN) / (TP + TN + FP + FN)

F_1-score = 2 × (Precision × Recall) / (Precision + Recall)

For multi-class classification, accuracy was computed within a one-versus-rest framework, in which each depressive symptom group was designated as the positive class, and the remaining two groups were combined as the negative class. Per-class accuracy thus reflects the proportion of correctly classified instances, both true positives and true negatives, within each binary classification setting.

Here, TP represents depressed participants correctly classified as depressed, and TN represents nondepressed participants correctly classified as nondepressed. FP refers to nondepressed participants misclassified as depressed, whereas FN indicates depressed participants misclassified as nondepressed.

Ethical Considerations

This study was a secondary analysis of deidentified wrist-worn actigraphy and associated covariate data from the HCHS/SOL, obtained through the NSRR. The original HCHS/SOL study received institutional review board approvals at all participating field centers (University of North Carolina at Chapel Hill, Albert Einstein College of Medicine, University of Illinois at Chicago, University of Miami, and San Diego State University), and written informed consent was obtained from all participants. The dataset analyzed in this study was deidentified before release in accordance with the Health Insurance Portability and Accountability Act and contained no direct or indirect personal identifiers. This study involved secondary analysis of existing deidentified data and did not involve recruitment of participants, interaction with participants, intervention, or access to identifiable private information. Because this study used only deidentified secondary data and did not involve human participants as defined under applicable institutional and national research ethics regulations, additional institutional review board approval and informed consent were not required for this secondary analysis. Data were accessed under an approved Data Access and Use Agreement with Brigham and Women's Hospital through the NSRR (NHLBI R24 HL114473). No attempt was made to identify or reidentify any participant. All data were fully deidentified; no identifiable participant information was accessed, and no identifying details appear in this study.

Results

Demographic Characteristics

A total of 1160 participants with valid baseline actigraphy and CESD-10 data were included. Based on CESD-10 scores, 580 (49.1%) participants were classified into the normal group (CESD-10 <10), 318 (27.9%) into the mild symptom group (CESD-10 =10 - 14), and 262 (23%) into the higher symptom group (CESD-10 ≥15). Higher depressive symptom burden was associated with a greater prevalence of being unmarried, currently unemployed, and current smoking (all P<.001; Table 3).

XSL•FO
RenderX

Table . Baseline demographic and clinical characteristics stratified by depressive symptom group. Group differences were assessed using chi-square tests for categorical variables and one-way ANOVA for continuous variables, as appropriate.

Measure	Normal (n=580)	Mild (n=318)	Higher (n=262)	*P* value[a]
Demographic				
Sex, n (%)				.28
Female	405 (69.83)	226 (71.07)	197 (75.19)	
Male	175 (30.17)	92 (28.93)	65 (24.80)	
Age (years), n (%)				.39
18 - 29	54 (9.31)	26 (8.18)	28 (10.69)	
30 - 39	85 (14.66)	52 (16.35)	31 (11.83)	
40 - 49	208 (35.86)	122 (38.36)	86 (32.82)	
50 - 64	233 (40.17)	118 (37.11)	117 (44.66)	
BMI (kg/m²), mean (SD)	30.53 (6.20)	30.65 (6.63)	30.48 (6.20)	.96
Married, n (%)	138 (23.79)	109 (34.28)	93 (35.50)	<.001
Education, n (%)				.08
No HS[b] or GED[c]	127 (39.94)	47 (30.72)	85 (32.44)	
≤HS or GED	106 (33.33)	61 (39.87)	103 (39.31)	
>HS or GED	85 (26.73)	45 (29.41)	74 (28.24)	
Unemployed	263 (45.34)	173 (54.40)	181 (69.08)	<.001
Lifestyle and behavioral, n (%)				
Smoker	97 (16.72)	83 (26.10)	91 (35.94)	<.001
Alcohol use	253 (43.62)	147 (46.23)	108 (40.44)	.48
Comorbidities, n (%)				
Hypertension	176 (30.34)	102 (32.08)	93 (35.27)	.33
Diabetes mellitus	106 (18.28)	71 (22.33)	54 (20.68)	.33
Cerebrovascular disease	12 (2.07)	11 (3.46)	6 (2.22)	.43
Coronary heart disease	37 (6.38)	18 (5.66)	19 (7.13)	.74
Chronic kidney disease	267 (46.11)	138 (43.81)	135 (52.03)	.14
Dyslipidemia	220 (38.00)	104 (33.02)	99 (33.44)	.24

[a]*P* values reflect between-group comparisons based on the tests described above.

[b]HS: high school.

[c]GED: general educational development.

In contrast, demographic characteristics and cardiometabolic comorbidities were comparable across depressive symptom groups, with no statistically significant differences observed in age, sex, education, BMI, alcohol use, diabetes, hypertension, cerebrovascular disease, chronic kidney disease, coronary heart disease, or dyslipidemia (all *P*>.05). In this community-based cohort, higher depressive symptom burden was more strongly associated with sociodemographic and behavioral factors, particularly marital status, unemployment, and smoking, than with cardiometabolic comorbidity. Specifically, unemployment exceeded two-thirds in the higher symptom group, and current smoking increased across depressive categories (Table 3).

Daily Light Exposure and Rest Activity

Actigraphy-derived light exposure and rest-activity characteristics by depressive symptom groups are provided in Table 4. Time spent above 500 and 1000 lux did not differ significantly across groups, although values were numerically lower among participants with depressive symptoms. In contrast, mean daytime white-light exposure was reduced in the higher symptom group compared with the normal group.

XSL•FO
RenderX

Table . Wrist actigraphy–based light exposure, 24-h rest-activity, and clock-time movement indices across 10-item Center for Epidemiologic Studies Depression scale depressive symptom groups.

Feature	Normal (n=580)	Mild (n=318)	Higher (n=262)	*P* value
Daily light exposure, mean (SD)				
Avg bright ≥500 lux (min/day)	220.94 (176.09)	201.47 (162.08)	200.98 (164.43)	.05[a]
Avg bright ≥1000 lux (min/day)	163.90 (147.78)	146.44 (135.65)	145.60 (136.20)	.05
Mean daytime white light (lux)	1464.38 (1557.69)	1399.92 (1534.62)	1124.74 (1279.01)	.01
Nonparametric 24 h rest-activity, mean (SD)				
Intradaily variability	0.97 (0.1)	0.83 (0.2)	0.91 (0.3)	<.001
Interdaily stability	0.25 (0.14)	0.14 (0.15)	0.10 (0.14)	.05
M10[b] activity	168.57 (51.71)	164.56 (63.10)	156.69 (52.79)	<.001
M10 onset	547.47 (122.05)	579.67 (127.29)	590.60 (126.52)	<.001
L5[c] activity	12.62 (12.41)	14.78 (18.75)	15.92 (20.83)	<.001
L5 onset	341.43 (113.39)	309.39 (98.90)	289.69 (115.26)	<.001
Cosinor-based circadian rhythm, mean (SD)				
MESOR[d]	583.84 (195.76)	400.56 (186.32)	304.50 (172.49)	<.001
Log amplitude	4.84 (1.53)	4.76 (1.47)	4.56 (1.31)	.01
Acrophase	9.78 (3.73)	9.68 (4.17)	9.40 (3.20)	.01
Clock-time activity, mean (SD)				
Midnight-2:59 AM (midnight)	29.63 (34.60)	37.30 (41.85)	37.65 (37.70)	.01
3 AM-5:59 AM (predawn)	20.04 (26.53)	20.34 (26.97)	20.82 (28.04)	.05
6 AM-8:59 AM (early morning)	98.19 (66.03)	84.90 (57.21)	73.74 (53.92)	<.001
9 AM-11:59 AM (morning)	157.35 (68.08)	144.57 (70.02)	140.09 (60.05)	.01
Noon-2:59 PM (early afternoon)	170.09 (58.00)	170.68 (72.76)	156.42 (61.30)	.01
3 PM-5:59 PM (mid-afternoon)	161.72 (55.01)	162.35 (70.33)	152.04 (58.18)	.05
6 PM-8:59 PM (evening)	147.88 (51.90)	144.45 (58.38)	134.61 (58.43)	<.001
9 PM-11:59 PM (late-evening)	96.46 (47.25)	100.77 (54.39)	93.16 (56.49)	.05

[a]*P*=.05: statistically significant.

[b]M10: average activity within the most active 10-h period of the day.

[c]Average activity within the least active 5-h period of the day.

[d]MESOR: midline estimating statistics of rhythm.

Across nonparametric rest-activity measures, higher depressive symptom burden was associated with lower M10 activity and higher L5 activity, consistent with attenuation of the 24-hour rest-activity amplitude. M10 and L5 onset times were also shifted in symptomatic groups, indicating a modest change in the timing of the most and least active periods. Cosinor-derived metrics showed concordant patterns, with midline estimating statistics of rhythm and log-transformed amplitude decreasing from the normal to the higher depressive group, while acrophase occurred slightly earlier in participants with greater symptom burden. In contrast, interdaily stability and intradaily variability differed minimally between groups, suggesting that day-to-day regularity and fragmentation were largely preserved despite lower overall activity and reduced rhythmic amplitude.

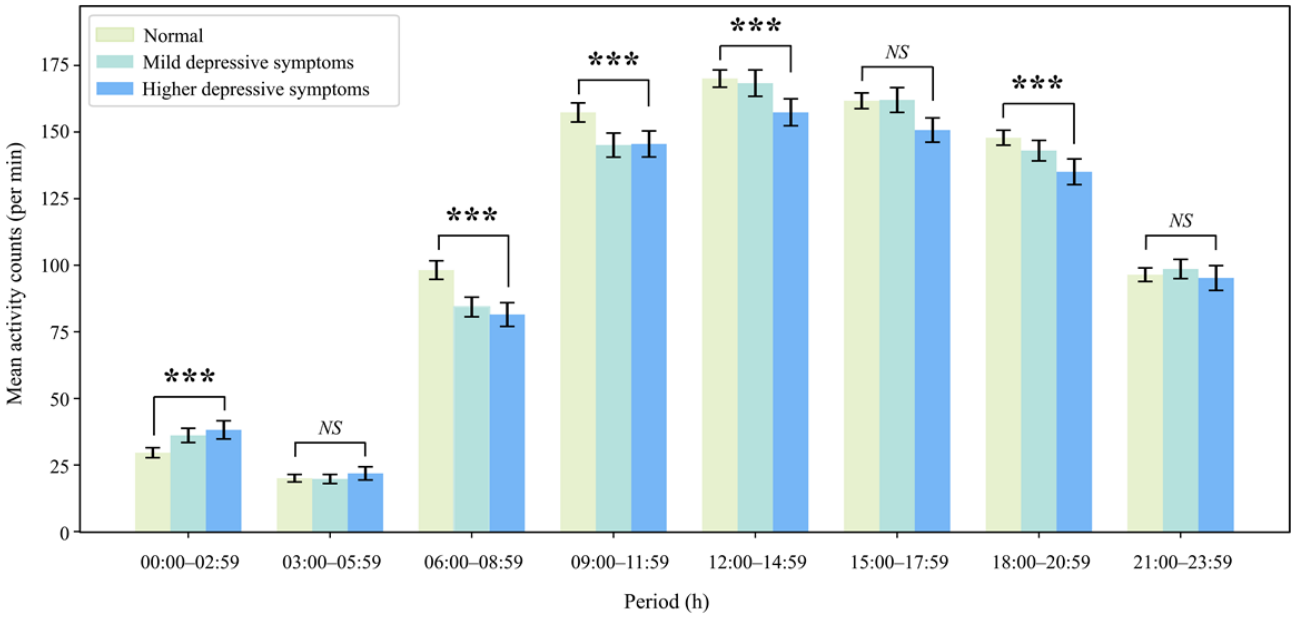
Time-of-Day Activity

Across the 24-hour day, mean activity declined as depressive symptom burden increased, with the largest separations observed

XSL•FO
RenderX

during daytime periods (Figure 3). The higher-symptom group showed lower activity in the morning (6 AM-8:59 AM and 9 AM-11:59 AM), midday (noon to 2:59 PM), and early evening (6 PM-8:59 PM) relative to the normal group. In contrast, late-evening activity (9 PM-11:59 PM) and several late-night intervals differed only modestly between groups. Notably, activity between midnight and 2:59 AM was slightly higher in the higher-symptom group, suggesting a small redistribution of activity toward the nighttime period.

Figure 3. Mean activity counts across 3-h intervals by depressive symptom group. *P* values were calculated using unpaired 2-tailed student *t* tests (NS: not significant; ***$P<.001$).



AI Model Performance for Depressive Symptom Screening

Across the 5 deep learning architectures, we evaluated classification performance on the test set, with 95% CIs (Table 5). Among all models, discrimination was consistently higher for the higher depressive symptom group than for the mild group, suggesting that actigraphy-derived behavioral patterns may better differentiate higher symptom burden from normal. LSTM demonstrated the strongest overall performance, achieving the highest AUROC of 0.640 (95% CI 0.526 - 0.736) for the mild depressive symptom group and 0.889 (95% CI 0.829 - 0.942) for the higher depressive symptom group, followed by Bi-LSTM with a comparable AUROC of 0.881 (95% CI 0.817 - 0.939). Model performance across the training, validation, and test sets is provided in Table S1 in Multimedia Appendix 1.

XSL•FO
RenderX

Table . Multiclass artificial intelligence model performance for depressive symptom screening.

Model	Normal (CESD[a] <10)		Mild (CESD 10 - 14)		Higher (CESD ≥15)	
	Accuracy[h] (95% CI)	AUC[b] (95% CI)	Accuracy[h] (95% CI)	AUC[b] (95% CI)	Accuracy[h] (95% CI)	AUC[b] (95% CI)
GRU[c]	0.634 (0.573-0.698)	0.811 (0.743-0.871)	0.639 (0.577-0.698)	0.557 (0.450-0.660)	0.786 (0.730-0.835)	0.777 (0.690-0.867)
LSTM[d]	0.726 (0.669 - 0.778)	0.857 (0.798 - 0.907)	0.682 (0.625 - 0.738)	0.640 (0.526 - 0.736)	0.850 (0.802 - 0.895)	0.889 (0.829 - 0.942)
Bi-LSTM[e]	0.625 (0.560 - 0.681)	0.855 (0.791 - 0.905)	0.577 (0.516 - 0.637)	0.551 (0.443 - 0.651)	0.854 (0.806 - 0.899)	0.881 (0.817 - 0.939)
CNN-LSTM[f]	0.597 (0.540 - 0.661)	0.832 (0.772 - 0.887)	0.630 (0.573 - 0.690)	0.579 (0.460 - 0.684)	0.758 (0.702 - 0.810)	0.807 (0.731 - 0.878)
Attention-LSTM[g]	0.706 (0.645 - 0.758)	0.845 (0.785 - 0.896)	0.646 (0.589 - 0.702)	0.536 (0.432 - 0.634)	0.859 (0.810 - 0.903)	0.861 (0.785 - 0.930)

[a]CESD: Center for Epidemiologic Studies Depression scale.

[b]AUC: area under the curve.

[c]GRU: gated recurrent unit.

[d]LSTM: long short-term memory.

[e]Bi-LSTM: bidirectional long short-term memory.

[f]CNN-LSTM: convolutional neural network–long short-term memory.

[g]Attention-LSTM: attention-based long short-term memory.

[h]Values are presented as point estimates with 95% confidence intervals (CI) in parentheses.

The confusion matrices summarize test-set classification outcomes across 3 depressive symptom groups—normal, mild, and higher—for 5 deep learning models: GRU (Figure 4A), LSTM (Figure 4B), Bi-LSTM (Figure 4C), CNN-LSTM (Figure 4D), and Attention-LSTM (Figure 4E). Recall for the mild symptom group was comparatively limited across all models, with recall ranging from 41% to 51.3%, reflecting frequent misclassification into adjacent categories. For the higher depressive symptom group, CNN-LSTM achieved the highest recall at 68.7% (22/32), followed by Bi-LSTM at 65.7% (21/32) (Figures 4C and 4D). Across all models, misclassification was most prevalent in the mild depressive symptom group, while higher depressive symptoms were least likely to be misclassified as normal across all models.

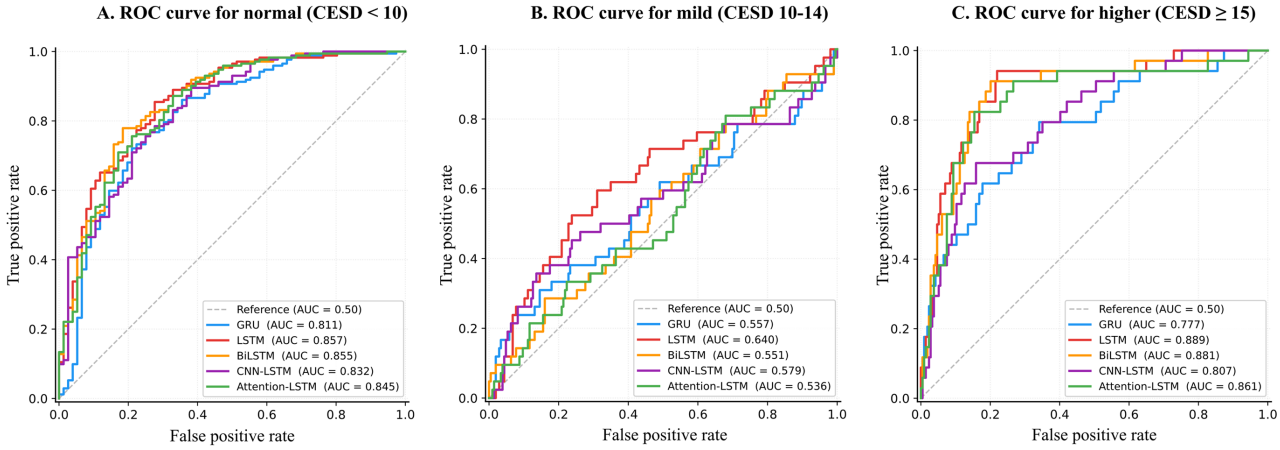
XSL•FO
RenderX

Figure 4. Confusion matrices for depressive symptom classification on the test set. Models evaluated include (A) gated recurrent unit (GRU), (B) long short-term memory (LSTM), (C) bidirectional long short-term memory (Bi-LSTM), (D) convolutional neural network–long short-term memory (CNN-LSTM), and (E) attention-LSTM.

XSL•FO
RenderX

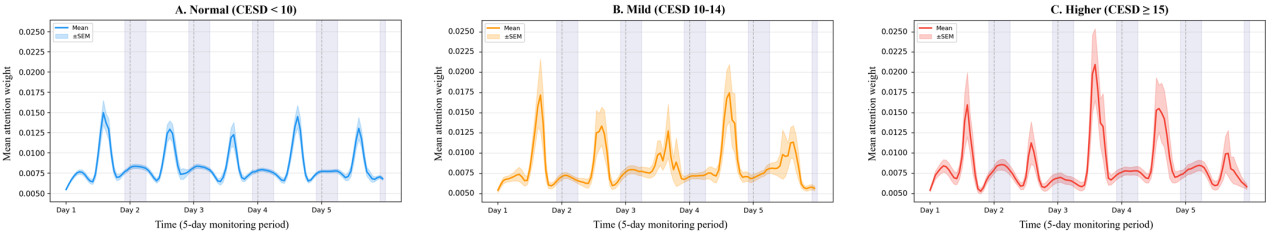
We further evaluated model discrimination using receiver operating characteristic curves across 5 deep learning models (Figure 5). For the normal group, all models achieved strong discrimination, with AUC values ranging from 0.811 to 0.857 (Figure 5A). In contrast, for the mild depressive symptom group (Figure 5B), discriminative performance was substantially lower across all models, with AUC values ranging from 0.536 to 0.640, with LSTM achieving the highest AUC of 0.640, reflecting the inherent challenge of classifying mild depressive symptoms. For the higher depressive symptom group (Figure 5C), models showed improved performance with AUC values ranging from 0.777 to 0.889, with LSTM achieving the highest AUC of 0.889, followed by Bi-LSTM at 0.881 and attention-LSTM at 0.861.

Figure 5. Receiver operating characteristic (ROC) curves for multiclass classification of depressive symptoms across 5 artificial intelligence models (gated recurrent unit [GRU], long short-term memory [LSTM], bidirectional long short-term memory [Bi-LSTM], convolutional neural network–long short-term memory [CNN-LSTM], and attention-LSTM) using a one-versus-rest framework. ROC curves are presented for each depressive symptom group: (A) normal (Center for Epidemiologic Studies Depression scale [CESD] <10), (B) mild (CESD 10-14), and (C) higher (CESD ≥15). Each curve represents the trade-off between the true positive rate (sensitivity) and false positive rate (1 − specificity) for a given model, with the dashed diagonal line indicating reference-level discrimination (area under the curve [AUC]=0.50).



The attention-LSTM model consistently assigned higher attention weights to daytime periods across all depressive symptom groups throughout the 5-day monitoring period. The normal group demonstrated stable (Figure 6A), rhythmic daily peaks, whereas the mild group showed more variable patterns with pronounced spikes at days 1, 3, and 4 (Figure 6B). The higher group exhibited the largest attention weight peaks, particularly at days 3 and 4, with wider SE of the mean (SEM) bands indicating greater interindividual variability (Figure 6C). Heatmap visualization further confirmed the concentration of attention weights between 12 hours and 16 hours (Figure S9 in Multimedia Appendix 1). To further quantify these patterns, peak attention timepoints were identified for each group (Table S6 and Figure S10 in Multimedia Appendix 1). The normal group peaked at day 1, 14 hours (mean 0.0150, SD 0.0197), the mild group at day 4, 15 hours (mean 0.0175, SD 0.0230), and the higher group showed concentrated peaks on day 3 between 13 hours and 15 hours (mean 0.0181, SD 0.0210), with all 3 peaks occurring within the same day, indicating sustained afternoon behavioral disruption as a key discriminative feature for severe depressive symptoms.

Figure 6. Mean attention weights of the attention-LSTM model across a 5-day monitoring period by depressive symptom group. Mean attention weights are plotted over the 5-day monitoring period for each depressive symptom group. (A) Normal (Center for Epidemiologic Studies Depression scale [CESD] <10), (B) mild (CESD 10-14), and (C) higher (CESD ≥15). Solid lines represent the group-level mean attention weight, and shaded regions indicate ±1 SE of the mean (SEM). Navy-shaded vertical bands denote approximate nighttime periods. The attention-LSTM model consistently assigned higher weights to daytime periods across all groups, reflecting the importance of active-hour behavioral signals in depressive symptom classification.



Discussion

Principal Findings

This study systematically compared 5 deep learning architectures using multichannel wrist-worn actigraphy to screen for CESD-10–defined depressive symptoms in a propensity score–matched HCHS/SOL cohort. Across both symptom-burden comparisons, discrimination was consistently higher for the higher depressive symptom group than for the mild group. In the mild depressive symptom comparison, LSTM achieved the strongest overall discrimination, with a test-set AUROC of 0.640. In the higher depressive symptom comparison, LSTM also showed the best performance, achieving

XSL•FO
RenderX

a test-set AUROC of 0.889, followed closely by Bi-LSTM with an AUROC of 0.881 (Table 5). The confusion matrix findings further indicated fewer missed cases in the higher symptom burden comparison, whereas false-positive classifications remained substantial in both settings (Figure 4). Collectively, these findings suggest that actigraphy-based sequence models may better distinguish higher symptom burden from typical behavioral patterns than mild symptoms. This difference may reflect the more heterogeneous and subtle behavioral signatures associated with mild depressive burden.

From a clinical and public health perspective, these findings support actigraphy-based deep learning as a scalable, passive approach for depressive symptom screening rather than a diagnostic test. In settings where depressive symptoms are underrecognized during routine care, model outputs could complement self-report screening by providing objective, longitudinal behavioral information that may prompt follow-up assessment, particularly among individuals with higher symptom burden.

Clinical Utility and Implementation Considerations

Classification threshold selection is a key consideration for real-world deployment. In this study, threshold calibration using the Nelder-Mead method on the validation set yielded model-specific optimal thresholds that maximized macro-F_1 performance. Lower thresholds may be preferred in population-level screening to prioritize sensitivity and reduce missed cases, whereas higher thresholds may be more appropriate in resource-limited settings to improve specificity and reduce unnecessary follow-up. As illustrated by the receiver operating characteristic curves (Figure 5), operating points can be adjusted to reflect these trade-offs; however, the optimal threshold should ultimately be determined empirically through prospective validation in the target population.

The confusion matrices (Figure 4) further illuminate the practical implications of these classification outcomes. For the mild depressive symptom group, the attention-LSTM produced 61 false positives and 23 false negatives (Figure 4E), reflecting a greater burden of overidentification than missed cases. For the higher depressive symptom group, despite CNN-LSTM achieving the highest recall, the Bi-LSTM yielded 11 false negatives with a false-positive rate of 12.5% (n=25), suggesting a more favorable trade-off between sensitivity and specificity (Figure 4C). These findings highlight a key characteristic of screening-oriented models: while sensitivity can be improved, false-positive rates may remain substantial, especially for the mild depressive symptom group.

In real-world clinical settings, false-positive predictions may increase patient burden, clinician workload, and health care costs. To address this, we propose a stepped implementation pathway in which a positive actigraphy-based screen is followed by a low-burden secondary assessment, such as a validated Patient Health Questionnaire-9 (PHQ-9) [42], rather than immediate referral for specialist evaluation. Repeating actigraphy-based screening over time may further help confirm persistent behavioral patterns before clinical referral. Individuals who screen positive across both stages could then be prioritized for formal clinical assessment by a mental health professional. The proposed model is not intended to replace clinical diagnosis but to function as a complementary screening signal within existing care pathways, enabling earlier identification in settings where depressive symptoms remain underrecognized during routine care.

Comparison With Prior Work

Prior studies using smartphones and consumer wearables have reported predictive value for depressive outcomes; however, as discussed in the Introduction section, these modalities differ fundamentally from research-grade actigraphy in sensor standardization, algorithmic transparency, and signal continuity, limiting direct comparisons. Among smartphone-based approaches, Saeb et al [13] demonstrated that GPS-derived features, including location variance and circadian movement, were significantly correlated with PHQ-9 depressive symptom severity ($|r| \geq 0.43$). Ikäheimonen et al [14] showed that smartphone behavioral features, including screen-off events, communication patterns, and location data, classified depressive states with an accuracy of 82% (95% CI 80% - 84%) in a clinically diagnosed sample. Consumer wearable studies have leveraged heart rate variability and circadian rhythm features for depressive symptom screening; for instance, Rykov et al [11] used Fitbit-derived features and reported an accuracy of 0.80, though performance varied across subgroups, reflecting the influence of population characteristics and feature selection on model generalizability.

Beyond modality, this study also differs from prior actigraphy work in its modeling approach. Incorporating multi-wavelength light exposure alongside activity and wake status offers a richer behavioral representation than studies relying on activity alone. Performance differences across model architectures were not uniform, suggesting that model selection should align with the intended screening target and the acceptable trade-off between missed cases and false alarms.

Table 6 summarizes representative studies that differ in outcome definitions, populations, and input features. Notably, all prior studies employed binary classification frameworks, whereas the present study simultaneously classified three depressive symptom groups, which inherently increases task complexity and limits direct performance comparison. Nonetheless, these studies provide useful context, as the present results extend prior evidence that AI models can support passive screening of depressive symptoms from wearable data. Specifically, Vahedifard et al [23] presented engineered actigraphy-based activity features using XGBoost for 4-class classification of inpatient adolescents with bipolar disorder. Price et al [24] applied a Gramian angular field CNN approach and highlighted the informativeness of nighttime motor patterns, although sensitivity was comparatively lower, potentially reflecting differences in feature scope and labeling strategy.

XSL•FO
RenderX

Table . Performance comparison of actigraphy-based models for 10-item Center for Epidemiologic Studies Depression scale defined depressive symptom screening.

Research work	Count, n[a]	Models	Class	Accuracy	Sensitivity	Specificity	AUC[b]
Jacobsen et al 2020 [43]	178	CNN[c]	Binary	0.73	0.65	0.78	0.83
Rykov et al 2021 [11]	267	XGBoost[d]	Binary	0.80	0.82	0.78	0.75
Kim et al 2021 [44]	47	LR[e]	Binary	0.91	0.88	0.94	0.96
Ghate et al 2023 [45]	59	LOOCV[f]	Binary	0.79	0.59	0.86	—[u]
Chen et al 2023 [46]	80	RF[h] and KNN[i]	Binary	0.66	0.67	0.68	0.68
Lee et al 2024 [26]	352	ResNet	Binary	0.88	0.83	0.91	0.94
Price et al 2024 [27]	1232	Transf[s]	Binary	0.83	0.80	0.86	0.89
Price et al 2024 [24]	8378	CNN-GAF[t]	Binary	—	0.45	0.82	0.68
Our study, 2026	1160	LSTM[r]	3-class[m]	0.75	0.60	0.83	0.80

[a]n is the number of participants included in the analysis.

[b]AUC: area under the curve.

[c]CNN: convolutional neural network.

[d]XGBoost: Extreme Gradient Boosting.

[e]LR: linear regression.

[f]LOOCV: leave-one-out cross-validation.

[g]Not available.

[h]RF: random forest.

[i]KNN: k-nearest neighbors.

[j]Transf: transfer learning.

[k]CNN-GAF: convolutional neural network–Gramian angular field.

[l]LSTM: long short-term memory.

[m]Values represent macro-averaged one-versus-rest performance metrics across three classes.

Prior work has evaluated deep learning approaches for depression screening using either actigraphy as a stand-alone modality or in combination with complementary modalities (Table 6). Lee et al [26] implemented a mixed-input framework integrating actigraphy-derived step counts with sleep stage information and survey measures, reporting an AUC of 0.94. Price et al [27] presented transformer-based architectures to model long-range temporal dependencies in actigraphy sequences, yielding an AUC of 0.89. Jacobsen et al [43] used CNN-based models to distinguish individuals with and without depression, reporting an accuracy of 0.73.

Additional studies have incorporated transfer learning and real-world wearable platforms. Ghate et al [45] proposed a real-time screening pipeline using transfer learning on Fitbit data, achieving 79% accuracy with leave-one-out cross-validation in a small sample. Rykov et al [11] evaluated an XGBoost (Extreme Gradient Boosting)–based machine learning model using wearable-derived digital biomarkers, such as circadian rhythm regularity and nighttime heart rate variation, for binary classification of depressive symptom status in a contrasted subsample, achieving an AUC of 0.75.

Strengths and Limitations

Unlike studies that rely primarily on engineered features or small-sample transfer-learning settings, we trained 5 deep learning models directly on raw multichannel actigraphy signals in a large, ethnically diverse Hispanic/Latino community cohort (n=1160), a population that remains underrepresented in actigraphy-based mental health research. All models used relatively lightweight architectures ranging from 86,817 to 518,065 parameters, supporting computational feasibility for real-world deployment. We further characterized symptom burden across three depressive symptom groups and reported complementary discrimination and error-profile metrics to quantify trade-offs between missed cases and false alarms. The use of 6 synchronized actigraphy channels, including activity counts, wake status, and multiwavelength light exposure, provided a richer behavioral representation than studies relying on single-channel inputs.

Several limitations should be considered. First, depressive symptom status was defined using the CESD-10, a self-reported screening tool rather than a structured clinical diagnosis;

XSL•FO
RenderX

therefore, our models estimate symptom burden, not clinician-diagnosed major depressive disorder. Second, actigraphy nonwear and missing segments may introduce residual measurement noise despite quality-control procedures. Third, external validation in independent cohorts is needed to establish transportability across diverse populations. Fourth, attention weight patterns in the attention-LSTM reflect selective temporal weighting rather than formal clinical explanations; individual-level attribution methods, such as Shapley Additive Explanations–based approaches, represent an important direction for future work.

Future Research

Future studies should validate these findings using clinically adjudicated outcomes across diverse subgroups and comorbid conditions. Longitudinal monitoring designs should be explored to assess whether actigraphy-derived behavioral signatures can track symptom trajectories and detect early relapses or recovery patterns. Additionally, the generalizability of models trained on research-grade actigraphy to consumer wearables and smartphone-based platforms warrants investigation. Establishing interpretable and externally validated models will be essential for translating actigraphy-based screening into real-world mental health workflows.

Conclusions

In a matched Hispanic/Latino community-based cohort, deep learning models trained on multichannel wrist-worn actigraphy demonstrated stronger discrimination for higher depressive symptoms than for mild depressive symptoms, with LSTM achieving the strongest overall performance and a macro-averaged AUROC of 0.80. These findings suggest that actigraphy-derived data can support AI-based passive screening of depressive symptoms; however, false-positive rates remain notable, and external validation and improved interpretability are needed before clinical deployment. An actigraphy-based AI model may serve as a scalable, passive, and noninvasive complementary signal to aid early screening alongside traditional depressive symptom assessments.

Acknowledgments

During the preparation of this paper, the authors used ChatGPT (OpenAI) and Claude (Anthropic) for language editing and grammar refinement. The authors reviewed and edited all content and take full responsibility for the accuracy, originality, and integrity of the paper.

Funding

This research was funded by the Basic Science Research Program through the National Research Foundation of Korea (NRF), supported by the Ministry of Education (RS-2023-00249148). This study was funded by the Regional Innovation System and Education (RISE) program through the Gangwon RISE Center, supported by the Ministry of Education and the Gangwon State, Republic of Korea (grant 2025-RISE-10-006). This research was supported by the Ministry of Science and ICT (MSIT), Korea, under the National Program in Medical AI Semiconductor (2024-0-0097), supervised by the Institute of Information and Communications Technology Planning and Evaluation (IITP) in 2026.

Authors' Contributions

DE conceived the study design and concept, conducted data preprocessing, coding, data analysis, feature engineering, and wrote of the paper. SO conducted data preprocessing and model development and revised the paper. The EU contributed to data collection, study concept and design, data interpretation, and paper editing and revision. JL and MHK contributed to data interpretation and data analysis and critically revised the paper. JK supported the study design and concept and edited and revised the paper. EU is a co-corresponding author and can be reached via email at edenbyra@yonsei.ac.kr.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Detailed model descriptions and architectures (gated recurrent unit, long short-term memory, bidirectional long short-term memory, convolutional neural network–long short-term memory, and attention-long short-term memory), layer-wise specifications, multiclass performance metrics, and attention-based long short-term memory weight analyses including heatmaps and peak timepoint distributions.

[DOCX File, 3629 KB - mhealth_v14i1e91479_app1.docx]

References

1. MacKinnon DF, Chen LN. Depressive disorders. In: Tasman’s Psychiatry, 5th edition: Springer International Publishing; 2024:1823-1880. [doi: 10.1007/978-3-030-51366-5_20]

XSL•FO
RenderX

2. Proudman D, Greenberg P, Nellesen D. The growing burden of major depressive disorders (MDD): implications for researchers and policy makers. Pharmacoeconomics 2021 Jun;39(6):619-625. [doi: 10.1007/s40273-021-01040-7] [Medline: 34013439]
3. Rose AL, Hopko DR, Lejuez CW, et al. Major depressive disorder. In: Functional Analysis in Clinical Treatment, 5th edition: Elsevier; 2020. [doi: 10.1016/B978-0-12-805469-7.00015-2]
4. Iliou K, Balaris D, Dokali AM, Fotopoulos V, Kouletsos A, Katsiana A. Exploring the effects of major depressive disorder on daily occupations and the impact of psychotherapy: a literature review. Cureus 2024 Mar;16(3):e55831. [doi: 10.7759/cureus.55831] [Medline: 38590479]
5. Garcia ME, Hinton L, Neuhaus J, Feldman M, Livaudais-Toman J, Karliner LS. Equitability of depression screening after implementation of general adult screening in primary care. JAMA Netw Open 2022 Aug 1;5(8):e2227658. [doi: 10.1001/jamanetworkopen.2022.27658] [Medline: 35980633]
6. Bransfield RC, Friedman KJ. Differentiating psychosomatic, somatopsychic, multisystem illnesses and medical uncertainty. Health Care (Don Mills) 2019;7(4):114. [doi: 10.3390/healthcare7040114]
7. Iancu SC, Wong YM, Rhebergen D, van Balkom A, Batelaan NM. Long-term disability in major depressive disorder: a 6-year follow-up study. Psychol Med 2020 Jul;50(10):1644-1652. [doi: 10.1017/S0033291719001612]
8. Richter T, Fishbain B, Richter-Levin G, Okon-Singer H. Machine learning-based behavioral diagnostic tools for depression: advances, challenges, and future directions. J Pers Med 2021 Sep 26;11(10):957. [doi: 10.3390/jpm11100957] [Medline: 34683098]
9. W Adams Z, McClure EA, Gray KM, Danielson CK, Treiber FA, Ruggiero KJ. Mobile devices for the remote acquisition of physiological and behavioral biomarkers in psychiatric clinical research. J Psychiatr Res 2017 Feb;85:1-14. [doi: 10.1016/j.jpsychires.2016.10.019] [Medline: 27814455]
10. Lekkas D, Gyorda JA, Price GD, Jacobson NC. Depression deconstructed: wearables and passive digital phenotyping for analyzing individual symptoms. Behav Res Ther 2023 Sep;168:104382. [doi: 10.1016/j.brat.2023.104382] [Medline: 37544229]
11. Rykov Y, Thach TQ, Bojic I, Christopoulos G, Car J. Digital biomarkers for depression screening with wearable devices: cross-sectional study with machine learning modeling. JMIR Mhealth Uhealth 2021 Oct 25;9(10):e24872. [doi: 10.2196/24872] [Medline: 34694233]
12. Lui GY, Loughnane D, Polley C, Jayarathna T, Breen PP. The Apple Watch for monitoring mental health-related physiological symptoms: literature review. JMIR Ment Health 2022 Sep 7;9(9):e37354. [doi: 10.2196/37354] [Medline: 36069848]
13. Saeb S, Lattie EG, Schueller SM, Kording KP, Mohr DC. The relationship between mobile phone location sensor data and depressive symptom severity. PeerJ 2016;4:e2537. [doi: 10.7717/peerj.2537] [Medline: 28344895]
14. Ikäheimonen A, Luong N, Baryshnikov I, et al. Predicting and monitoring symptoms in patients diagnosed with depression using smartphone data: observational study. J Med Internet Res 2024 Dec 3;26:e56874. [doi: 10.2196/56874] [Medline: 39626241]
15. Liu F, Wanigatunga AA, Schrack JA. Assessment of physical activity in adults using wrist accelerometers. Epidemiol Rev 2021 Dec 30;43(1):65-93. [doi: 10.1093/epirev/mxab004]
16. Schuch F, Vancampfort D, Firth J, et al. Physical activity and sedentary behavior in people with major depressive disorder: a systematic review and meta-analysis. J Affect Disord 2017 Mar 1;210:139-150. [doi: 10.1016/j.jad.2016.10.050] [Medline: 28033521]
17. Riemann D, Krone LB, Wulff K, Nissen C. Sleep, insomnia, and depression. Neuropsychopharmacol 2020 Jan;45(1):74-89. [doi: 10.1038/s41386-019-0411-y]
18. Abdullah S, Murnane EL, Matthews M. Circadian computing: sensing, modeling, and maintaining biological rhythms. In: Kumar S, Nilsen WJ, Pavel M, editors. Mobile Health: Sensors, Analytic Methods, and Applications: Springer; 2017:35-58. [doi: 10.1007/978-3-319-51394-2_3]
19. Chen ZS, Kulkarni P(, Galatzer-Levy IR, Bigio B, Nasca C, Zhang Y. Modern views of machine learning for precision psychiatry. Patterns 2022 Nov;3(11):100602. [doi: 10.1016/j.patter.2022.100602]
20. Yu Y, Si X, Hu C, Zhang J. A review of recurrent neural networks: LSTM cells and network architectures. Neural Comput 2019 Jul;31(7):1235-1270. [doi: 10.1162/neco_a_01199] [Medline: 31113301]
21. Ismail Fawaz H, Forestier G, Weber J, Idoumghar L, Muller PA. Deep learning for time series classification: a review. Data Min Knowl Disc 2019 Jul 1;33(4):917-963. [doi: 10.1007/s10618-019-00619-1]
22. Ma Z, Mei G. A hybrid attention-based deep learning approach for wind power prediction. Appl Energy 2022 Oct;323:119608. [doi: 10.1016/j.apenergy.2022.119608]
23. Vahedifard F, Birmaher B, Iyengar S, et al. Maximum and minimum activity in inpatient adolescents with bipolar disorders: daily-variability classification of actigraphy pattern with artificial intelligence. Psychiatry Res Commun 2025 Jun;5(2):100212. [doi: 10.1016/j.psycom.2025.100212] [Medline: 40575150]
24. Price GD, Heinz MV, Collins AC, Jacobson NC. Detecting major depressive disorder presence using passively-collected wearable movement data in a nationally-representative sample. Psychiatry Res 2024 Feb;332:115693. [doi: 10.1016/j.psychres.2023.115693] [Medline: 38194801]

XSL•FO
RenderX

25. Mienye ID, Swart TG, Obaido G. Recurrent neural networks: a comprehensive review of architectures, variants, and applications. Information 2024;15(9):517. [doi: 10.3390/info15090517]
26. Lee TR, Kim GH, Choi MT. Geriatric depression and anxiety screening via deep learning using activity tracking and sleep data. Int J Geriat Psychiatry 2024 Feb;39(2). [doi: 10.1002/gps.6071]
27. Price GD, Collins AC, Mackin DM, Heinz MV, Jacobson NC. Use of passively collected actigraphy data to detect individual depressive symptoms in a clinical subpopulation and a general population. J Psychopathol Clin Sci 2025 Jan;134(1):31-40. [doi: 10.1037/abn0000933] [Medline: 39432354]
28. Sorlie PD, Avilés-Santa LM, Wassertheil-Smoller S, et al. Design and implementation of the Hispanic Community Health Study/Study of Latinos. Ann Epidemiol 2010;20(8):629-641. [doi: 10.1016/j.annepidem.2010.03.015]
29. Khatun MA, Yousuf MA, Ahmed S, et al. Deep CNN-LSTM with self-attention model for human activity recognition using wearable sensor. IEEE J Transl Eng Health Med 2022;10:2700316. [doi: 10.1109/JTEHM.2022.3177710] [Medline: 35795873]
30. Ramasubramanian K, Singh A. Deep learning using Keras and TensorFlow. In: Machine Learning Using R: With Time Series and Industry-Based Use Cases in R: Apress; 2019:667-688. [doi: 10.1007/978-1-4842-4215-5_11]
31. Nguyen HH, Adams PN, Miller SJ. Building Statistical Models in Python: Develop Useful Models for Regression, Classification, Time Series, and Survival Analysis: Packt Publishing; 2023.
32. Sharma A, Shrimali VR, Beyeler M. Machine Learning for Opencv 4: Intelligent Algorithms for Building Image Processing Apps Using Opencv 4, Python, and Scikit-Learn: Packt Publishing Ltd; 2019.
33. Yacouby R, Axman D. Probabilistic extension of precision, recall, and F1 score for more thorough evaluation of classification models. Presented at: Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems; 2020; Online p. 79-91. [doi: 10.18653/v1/2020.eval4nlp-1.9]
34. Minaeva O, Riese H, Lamers F, Antypa N, Wichers M, Booij SH. Screening for depression in daily life: development and external validation of a prediction model based on actigraphy and experience sampling method. J Med Internet Res 2020 Dec 1;22(12):e22634. [doi: 10.2196/22634] [Medline: 33258783]
35. Gallo LC, Penedo FJ, Carnethon M, et al. The Hispanic Community Health Study/Study of Latinos Sociocultural Ancillary Study: sample, design, and procedures. Ethn Dis 2014;24(1):77-83. [Medline: 24620452]
36. Cheng S, Chan ACM. The Center for Epidemiologic Studies Depression Scale in older Chinese: thresholds for long and short forms. Int J Geriat Psychiatry 2005 May;20(5):465-470. [doi: 10.1002/gps.1314]
37. González P, Nuñez A, Merz E, et al. Measurement properties of the Center for Epidemiologic Studies Depression Scale (CES-D 10): findings from HCHS/SOL. Psychol Assess 2017 Apr;29(4):372-381. [doi: 10.1037/pas0000330] [Medline: 27295022]
38. Muhuri PS, Chatterjee P, Yuan X, Roy K, Esterline A. Using a Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) to classify network attacks. Information 2020;11(5):243. [doi: 10.3390/info11050243]
39. Alizadegan H, Rashidi Malki B, Radmehr A, Karimi H, Ilani MA. Comparative study of long short-term memory (LSTM), bidirectional LSTM, and traditional machine learning approaches for energy consumption prediction. Energy Exploration & Exploitation 2025 Jan;43(1):281-301. [doi: 10.1177/01445987241269496]
40. Amtmann D, Kim J, Chung H, et al. Comparing CESD-10, PHQ-9, and PROMIS depression instruments in individuals with multiple sclerosis. Rehabil Psychol 2014 May;59(2):220-229. [doi: 10.1037/a0035919] [Medline: 24661030]
41. Mohebbi M, Nguyen V, McNeil JJ, et al. Psychometric properties of a short form of the Center for Epidemiologic Studies Depression (CES-D-10) scale for screening depressive symptoms in healthy community dwelling older adults. Gen Hosp Psychiatry 2018;51:118-125. [doi: 10.1016/j.genhosppsych.2017.08.002] [Medline: 28890280]
42. Williams N. PHQ-9. Occup Med 2014 Mar 1;64(2):139-140. [doi: 10.1093/occmed/kqt154]
43. Jakobsen P, Garcia-Ceja E, Riegler M, et al. Applying machine learning in motor activity time series of depressed bipolar and unipolar patients compared to healthy controls. PLoS ONE 2020;15(8):e0231995. [doi: 10.1371/journal.pone.0231995] [Medline: 32833958]
44. Kim H, Lee SH, Lee SE, Hong S, Kang HJ, Kim N. Depression prediction by using ecological momentary assessment, Actiwatch data, and machine learning: observational study on older adults living alone. JMIR Mhealth Uhealth 2019 Oct 16;7(10):e14149. [doi: 10.2196/14149] [Medline: 31621642]
45. Ghate R, Kalnad N, Walambe R, Kotecha K. Real-Time inferencing using transfer learning for a screening of depression detection using actigraphy. In: Artificial Intelligence: Theory and Applications: Springer Nature; 2024:327-336. [doi: 10.1007/978-981-99-8476-3_27]
46. Chen J, Chan NY, Li CT, et al. Multimodal digital assessment of depression with actigraphy and app in Hong Kong Chinese. Transl Psychiatry 2024 Mar 18;14(1):150. [doi: 10.1038/s41398-024-02873-4] [Medline: 38499546]

Abbreviations

AI: artificial intelligence
AUC: area under the curve
AUROC: area under the receiver operating characteristic curve

XSL•FO
RenderX

Bi-LSTM: bidirectional long short-term memory
CESD-10: 10-item Center for Epidemiologic Studies Depression scale
CNN: convolutional neural network
CNN-LSTM: convolutional neural network–long short-term memory
GRU: gated recurrent unit
HCHS/SOL: Hispanic Community Health Study/Study of Latinos
HS/GED: High school / General Educational Development
L5: average activity within the least active 5-hour period of the day
LSTM: long short-term memory
M10: average activity within the most active 10-hour period of the day
NSRR: National Sleep Research Resource
PHQ-9: Patient Health Questionnaire-9
XGBoost: Extreme Gradient Boosting

Edited by Z Liang; submitted 19.Jan.2026; peer-reviewed by A Teles, D Shin; revised version received 03.Jun.2026; accepted 04.Jun.2026; published 15.Jul.2026.

Please cite as:
Enkhbayar D, Oh S, Lee J, Kim MH, Urtnasan E, Key J
Passive Screening for Depressive Symptoms Using Daily Wrist Actigraphy and Deep Learning: Model Development and Validation Study
JMIR Mhealth Uhealth 2026;14:e91479
URL: https://mhealth.jmir.org/2026/1/e91479
doi:10.2196/91479

© Doljinsuren Enkhbayar, Somin Oh, Jinhee Lee, Min-Hyuk Kim, Erdenebayar Urtnasan, Jaehong Key. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 15.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Smartphone-Based Physical Performance and Multidimensional Determinants of Self-Reported Knee Pain in Community-Dwelling Older Adults: Cross-Sectional Machine Learning and Network Analysis Study

Ui-jae Hwang[1], PhD; Peng Xia[1], PhD; Tianxiang Fan[1], MS; Arnold YL Wong[1], PhD; Oh-yun Kwon[2], PhD; Siu-ngor Fu[1], PhD

[1]Department of Rehabilitation Sciences, Hong Kong Polytechnic University, 11 Yuk Choi Road, Hung Hom, Kowloon, Hong Kong, China (Hong Kong)

[2]Department of Physical Therapy, Yonsei University, Wonju-si, Republic of Korea

Corresponding Author:
Ui-jae Hwang, PhD
Department of Rehabilitation Sciences, Hong Kong Polytechnic University, 11 Yuk Choi Road, Hung Hom, Kowloon, Hong Kong, China (Hong Kong)

Abstract

Background: Knee pain affects 22.9% of individuals aged 40 years and over globally and is associated with body function, activity, environmental, and personal factors described in the *International Classification of Functioning, Disability and Health* (ICF) model. Most prior studies examined isolated risk factors using conventional regression.

Objective: This study aimed to examine self-reported knee pain in community-dwelling older adults by combining machine learning and partial-correlation network analysis with smartphone-based physical performance measurement.

Methods: This cross-sectional study included 852 adults aged 60 years and older. Baseline assessment of 38 variables across 7 ICF-aligned domains was reduced to 21 predictors using prespecified, rule-based criteria (events per variable=14.95). Walking speed, sit-to-stand, gait knee flexion, and gait asymmetry were measured using a validated vision-based smartphone app. Missing data were addressed by multiple imputation (m=5) with Rubin rules. Six algorithms (logistic regression with elastic-net penalty as baseline, k-nearest neighbors, random forest, Extreme Gradient Boosting, Light Gradient Boosting Machine, and support vector machine) were trained using nested 5×5 cross-validation, and the best model was interpreted using Shapley Additive Explanations and partial dependence plots. Partial-correlation network analysis examined the 10 top-ranked features and the knee-pain node, with bootstrap stability assessment and a sensitivity analysis excluding EuroQol 5-Dimension (EQ-5D) Pain/Discomfort.

Results: Knee pain prevalence was 36.9% (314/852). Pooled area under the receiver operating characteristic curve values ranged from 0.692 to 0.723, with random forest highest (area under the receiver operating characteristic curve=0.723, 95% CI 0.689 - 0.757). The top Shapley Additive Explanations features were EQ-5D Pain/Discomfort, EQ-5D Utility, and house estate. In the 11-node network, only EQ-5D Pain/Discomfort (partial correlation=0.260) and house estate (0.158) had direct edges with knee pain. Sit-to-stand showed the highest strength centrality (0.679) without a direct edge, acting as a hub linking body function, body structure, mental health, and activity variables. Shapley Additive Explanations importance and network strength were weakly correlated (ρ=0.018). The sensitivity analysis preserved the pattern (best area under the receiver operating characteristic curve=0.724).

Conclusions: Knee pain in older adults was associated with variables spanning 7 ICF domains, with EQ-5D Pain/Discomfort and housing environment (lower-income rental estate residence) as direct correlates and sit-to-stand as a network hub. Combining machine learning and partial-correlation network analysis may inform multidisciplinary biopsychosocial assessment, pending confirmation in prospective and externally validated studies.

(JMIR Mhealth Uhealth 2026;14:e92466) doi:10.2196/92466

KEYWORDS

machine learning; network analysis; knee pain; aging; digital health; geriatric assessment

XSL•FO
RenderX

Introduction

Self-reported knee pain is one of the most common musculoskeletal complaints among community-dwelling older adults and a leading contributor to disability, reduced physical activity, and impaired quality of life [1,2]. The Global Burden of Disease 2021 study estimated that 595 million people worldwide were living with osteoarthritis in 2020, corresponding to 7.6% of the global population, with knee being the most common site; compared with 2020, the number of cases of knee osteoarthritis is projected to increase by 74.9% by 2050, alongside continued increases in population aging and obesity prevalence [3,4]. Identifying which factors are associated with self-reported knee pain in community settings, before structural disease and clinical referral, is therefore a continuing research priority.

Knee pain in older adults is conceptualized as multidimensional. The *International Classification of Functioning, Disability and Health* (*ICF*) framework, built on the biopsychosocial model, organizes health into interacting domains spanning body functions and structures, activities and participation, and environmental and personal factors [5-7]. In knee osteoarthritis populations, physical performance, psychological status, sleep, and environmental context each carry independent associations with pain experience and functional limitation [8-10]. The *ICF* specifies that health outcomes such as knee pain arise from a structure in which some variables show direct associations with the outcome and others act through intermediate variables. Although this direct-versus-indirect distinction is part of the framework, prior studies of knee pain have not quantitatively separated the two types of association in a single analysis. As a result, it remains unclear which domains carry direct associations with knee pain in older adults and which contribute through other variables in the multidimensional structure.

Two analytical paradigms have been applied to multidimensional pain data, but they answer different questions and have typically been used in isolation. Machine learning (ML) classification with Shapley Additive Explanations (SHAP)–based explanation quantifies the marginal predictive importance of each feature, capturing potentially nonlinear contributions that traditional regression may miss [11,12]. Recent reviews of ML applications in knee osteoarthritis highlight the rapid growth of this methodology across clinical, structural, and surgical endpoints, while also noting that most studies focus on imaging-based outcomes and that few have addressed self-reported pain or nonimaging multidimensional determinants in older adults [2,3]. Partial-correlation network analysis characterizes the conditional dependency structure among variables, with centrality indices summarizing how each variable connects to the broader system after controlling for all others [13,14]. When the two paradigms are reported jointly on the same data, it becomes possible to determine whether a variable's contribution to the model output is mirrored by a direct conditional dependency with the outcome, or whether that contribution is accounted for by other variables in the network. This joint use operationalizes the direct and indirect pathways described by the *ICF*, but applications of this combined analysis to knee pain remain limited [8,15]. Mobile health measures of physical performance, such as smartphone-based gait and sit-to-stand assessments [16], need to be integrated with questionnaire-based body function and environmental variables within a single *ICF*-aligned analysis.

The aim of this study was to characterize how variables across seven *ICF*-aligned domains, including smartphone-derived physical performance, are associated with self-reported knee pain in community-dwelling older adults, and to distinguish those that show direct conditional dependencies with knee pain from those whose contribution is accounted for by other variables in the multidimensional structure. We hypothesized that the seven *ICF* domains would show heterogeneous patterns of association with knee pain, with some domains contributing through direct conditional dependencies and others through indirect pathways involving other variables, in line with the multidimensional structure described by the framework.

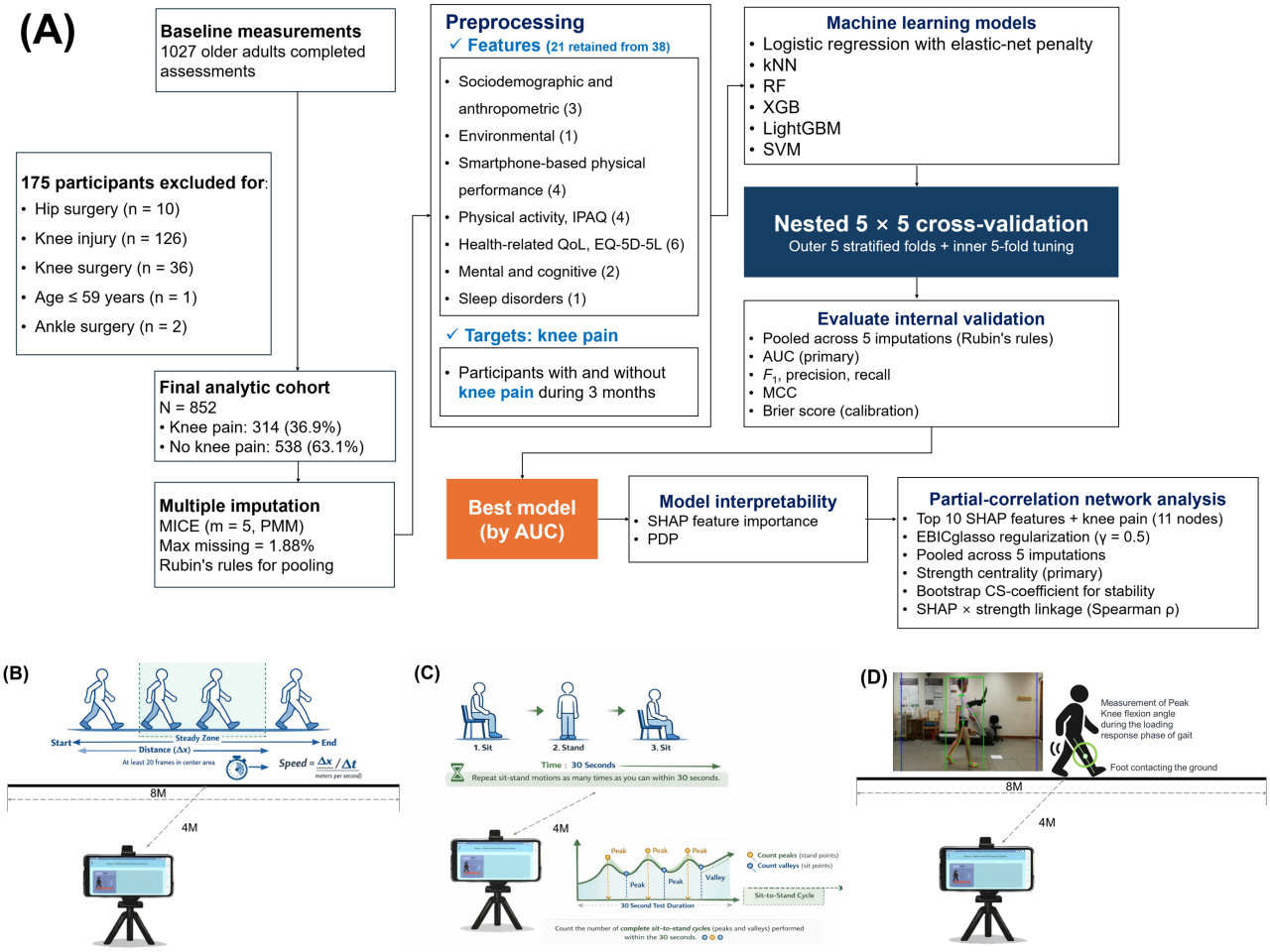
Methods

Study Design and Participation

A total of 1027 community-dwelling Hong Kong residents aged 60 years or older were recruited for the study. Eligible participants were those living independently in housing estates, able to walk independently for at least 8 meters, and capable of providing informed consent. Exclusion criteria included (1) severe cognitive impairment preventing informed consent; (2) inability to communicate in Cantonese or read Traditional Chinese; (3) history of hip surgery, knee surgery, or ankle surgery within the past 12 months; (4) recent knee injury (within the past 6 mo); and (5) diagnosed neurological diseases including Parkinson disease, stroke, multiple sclerosis, or other conditions. Of the 1027 participants initially screened, 852 participants (83%) met all inclusion criteria and completed the comprehensive baseline assessment protocol. Study flowchart and participant characteristics are presented in Figure 1 and Table 1, respectively. Reporting follows the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis + AI statement for prediction model studies using AI and ML [17], consistent with the *Journal of Medical Internet Research*–recommended guidelines for ML predictive models in biomedical research [18].

XSL•FO
RenderX

Figure 1. Study flowchart and analytical framework and smartphone-based physical performance assessment. (A) Study flowchart. (B) Walking speed measurement across an 8-meter walkway with smartphone positioned 4 meters perpendicular to walking path. (C) Sit-to-stand performance assessment with automated cycle counting over 30 seconds using smartphone-based motion analysis. (D) Knee flexion angle measurement during early stance phase of gait using computer vision pose estimation to detect loading response. AUC: area under the receiver operating characteristic curve; EQ-5D-5L: EuroQol 5-Dimensions 5-Levels questionnaire; IPAQ: International Physical Activity Questionnaire; kNN: k-nearest neighbors; LightGBM: Light Gradient Boosting Machine; MCC: Matthews correlation coefficient; MICE: multiple imputation by chained equations; PDP: partial dependence plots; PMM: predictive mean matching; QoL: quality of life; RF: random forest; SHAP: Shapley Additive Explanations; SVM: support vector machine; XGB: Extreme Gradient Boosting.



XSL•FO
RenderX

Table . Baseline characteristics of the analytic cohort by self-reported knee pain status.

Characteristics	Knee pain, no (n=538)	Knee pain, yes (n=314)	*P* value	Effect size
Sociodemographic and anthropometric				
Age (years), mean (SD)	76.20 (7.77)	75.53 (7.24)	.21	0.09
Gender (female), n (%)	403 (74.9)	257 (81.8)	.02	0.08
Height (m), mean (SD)	1.57 (0.08)	1.56 (0.08)	.06	0.14
Weight (kg), mean (SD)	57.00 (10.28)	58.33 (10.87)	.08	−0.13
BMI (kg/m²), mean (SD)	22.97 (3.54)	23.82 (3.84)	.001	−0.23
Environmental				
House estate, n rental (%)	336 (62.5)	258 (82.2)	<.001	0.20
Smartphone-based physical performance, mean (SD)				
Walking speed (m/s)	1.20 (0.29)	1.15 (0.26)	.02	0.17
Sit-to-stand (reps/30 s)	11.41 (3.52)	10.50 (3.16)	<.001	0.27
Gait knee flexion (degree)	20.93 (4.88)	20.41 (4.58)	.12	0.11
Gait asymmetry (%)	8.77 (7.69)	9.48 (9.10)	.25	−0.09
Physical activity (IPAQ[a]), mean (SD)				
Sitting time (min/day)	292.55 (134.27)	331.17 (138.94)	<.001	−0.28
Walking (MET[b]-min/week)	1800.28 (1519.42)	1377.88 (1090.70)	<.001	0.31
Moderate PA[c] (MET-min/week)	1231.56 (1778.13)	1076.45 (1291.01)	.15	0.10
Vigorous PA (MET-min/week)	449.11 (1920.25)	81.92 (471.79)	<.001	0.24
Health-related quality of life (EQ-5D-5L)[d], mean (SD)				
Mobility (1-5)	1.16 (0.47)	1.38 (0.68)	<.001	−0.39
Self-care (1-5)	1.03 (0.19)	1.04 (0.23)	.38	−0.07
Usual activities (1-5)	1.08 (0.33)	1.17 (0.47)	.003	−0.24
Pain/Discomfort (1-5)	1.41 (0.63)	1.92 (0.79)	<.001	−0.73
Anxiety/Depression (1-5)	1.19 (0.51)	1.29 (0.56)	.02	−0.17
Utility index (HK-specific)[e]	0.93 (0.11)	0.86 (0.16)	<.001	0.57
Mental and cognitive, mean (SD)				
Mental health (ReQoL[f] total)	32.38 (5.94)	31.10 (6.13)	.006	0.21
ERA-12[g] total	35.86 (14.68)	35.81 (15.22)	.97	0.00
Sleep, mean (SD)				
Sleep disorders (count, 0 - 6)	1.08 (0.93)	1.44 (1.14)	<.001	−0.36

[a]IPAQ: International Physical Activity Questionnaire.

[b]MET: metabolic equivalent of task.

[c]PA: physical activity.

[d]EQ-5D-5L: EuroQol 5-Dimension 5-Level.

[e]HK: Hong Kong.

[f]ReQoL: Recovering Quality of Life.

[g]ERA-12: 12-item Expectations Regarding Aging.

XSL•FO
RenderX

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board of The Hong Kong Polytechnic University (HSEARS20220628001). All participants provided written informed consent before data collection, after the nature and possible consequences of the study were explained. Video recordings captured during the smartphone-based assessments were transmitted to institutional servers for automated pose-estimation analysis; all data were deidentified before analysis and stored in a password-protected database accessible only to the research team. No identifiable participant information is presented in this paper or its supplementary materials. As part of the wider project, participants who completed the exercise program and achieved a predefined fitness criterion at the 12-month follow-up received a health hamper valued at HKD 200 (approximately US $26).

Knee Pain Outcome

The primary outcome was the presence of self-reported knee pain experienced within the preceding 3 months, ascertained through structured interviewer-administered questionnaires conducted by trained research assistants. Knee pain was defined as any ache, discomfort, or pain sensation experienced in one or both knees during this 3-month period [1]. Participants were classified as having knee pain (Yes) or not (No) based on this question.

Data Collection and Measurements

Overview

Comprehensive baseline assessment included variables organized under seven domains aligned with the *ICF* framework: sociodemographic and anthropometric characteristics, environmental factors, smartphone-based physical performance, physical activity, health-related quality of life, mental and cognitive functioning, and sleep. All instruments used validated Traditional Chinese versions with established psychometric properties in older adult populations.

Sociodemographic and anthropometric measures comprised age, sex, and BMI (calculated from measured height and weight). Housing type (elderly housing vs rental estate) was used as an environmental indicator. The Hong Kong Housing Society operates a spectrum of estates serving different income groups, with rental housing estates serving lower-income households and senior (older) housing serving middle- to higher-income groups; in this cohort, rental estate residence therefore reflects a relatively lower socioeconomic position. Health-related quality of life was assessed using the EuroQol 5-Dimension 5-Level (EQ-5D-5L) instrument [19], confirming scores for five dimensions (Mobility, Self-care, Usual Activities, Pain/Discomfort, and Anxiety/Depression), the visual analog scale, and the Hong Kong–specific utility index. Aging expectations were assessed using the 12-item Expectations Regarding Aging questionnaire total score [20]. Mental health–related quality of life was assessed using the Recovering Quality of Life total score [21]. Sleep was assessed using the Sleep Disorders Questionnaire covering six categories (insomnia, periodic limb movement, circadian rhythm disorder, movement disorder, parasomnia, and sleep apnea); a total disorder count from 0 to 6 was used for analysis [22]. Physical activity was assessed using the International Physical Activity Questionnaire (IPAQ), from which we derived sitting time (minutes per day) and metabolic equivalent of task (MET) minutes per week for walking, moderate, and vigorous activities [23].

Smartphone-Based Physical Performance Assessment

Physical performance assessment was performed at community-dwelling centers using a research prototype smartphone-based digital health system that used computer vision pose estimation algorithms for mobility evaluation (Figure 1) [16]. Correlation analyses of the smartphone measurement system showed agreement for walking speed measurements (r=0.975) and knee flexion assessments (r=0.907 to 0.923), with test-retest reliability values (intraclass correlation coefficient=0.86 to 0.94) and no significant systematic bias detected following algorithm calibration [16]. The experimental setup involved positioning an Android (Google LLC) smartphone equipped with a single camera on a 0.9-meter height stand, placed 4 meters perpendicular to the midpoint of an 8-meter walkway within the dwelling centers. The smartphone was installed with a vision-based gait analysis app that uses pose estimation technology to detect and track anatomical landmarks throughout movement sequences. The app connects to a backend infrastructure featuring cloud-based processing capabilities through an Nginx load balancer (F5, Inc.) capable of handling multiple simultaneous connections. Video data is transmitted frame-by-frame in real-time to a Flask framework for pose detection [16]. Processed data, including raw video files and numerical results, are stored in a MySQL (Oracle Corporation) database with web-based researcher portals providing access to results.

For walking speed (m/s) [24,25], participants walked naturally across a standardized 8-meter walkway while being recorded using smartphone cameras positioned perpendicular to the walking path. The analysis system used human detection followed by pose estimation to track 17 key anatomical landmarks, including head, thorax, pelvis, and bilateral extremity joints with subpixel accuracy. Spatial calibration was applied by matching the median-detected human height across all frames with participant-reported height, enabling conversion from pixel coordinates to real-world measurements without external reference objects. The system automatically identified steady-state walking zones by tracking pelvis position along the trajectory, requiring consistent walking velocity within a predefined spatial range (1/5 to 4/5 of walkway length) for a minimum of 20 consecutive frames to exclude acceleration and deceleration phases. Walking speed was computed as the change in pelvis horizontal position divided by elapsed time within identified zones, with median values calculated across multiple segments and temporal smoothing applied using a 5-frame median filter.

Sit-to-stand performance was assessed through automated movement cycle counting using pose-based motion analysis over 30 seconds [26,27]. The system tracked pelvis vertical position and bilateral knee flexion angles, with hip height

XSL•FO
RenderX

changes serving as the primary indicator for transitions supplemented by knee angle verification. Complete cycles were identified using peak and valley detection in hip trajectories with dynamic thresholds established based on movement range (25th to 75th percentile of hip height range) to accommodate individual differences. Each detected peak required a corresponding valley for cycle validation, counting only complete cycles within the time window from the first detected movement. Quality control included 5-frame smoothing filters and automatic switching to knee flexion angle-based analysis using predefined thresholds (110 to 160 degrees range) when hip height detection showed inconsistencies.

Bilateral knee flexion angles were measured during early stance phase using three-point angular calculations incorporating hip, knee, and ankle joint centers [28,29]. The algorithm computed angles between thigh and shank segments using arctangent-based trigonometric calculations, with early stance phase automatically identified through ankle crossing patterns and calf angle peaks. The system detected when the contralateral ankle passed the ipsilateral ankle, followed by identification of maximum calf angle relative to vertical, indicating early stance initiation. Measurements were direction-adjusted based on walking trajectory, analyzing right leg early stance for rightward walking and left leg for leftward movement. Peak knee flexion angles were identified across multiple gait cycles with median values calculated to represent typical performance while minimizing outlier influence.

Statistical Analysis

All ML analyses were conducted using R statistical software (version 4.5.3; R Core Team). Baseline characteristics of the analytic cohort were compared between participants with and without self-reported knee pain using independent 2-tailed *t* tests (continuous variables) or χ^2 tests (categorical variables).

Premodeling Variable Reduction

To address events-per-variable (EPV) constraints raised in the literature for prediction model development [30,31], we applied a prespecified, rule-based variable reduction prior to modeling. The 38 candidate predictor variables were reduced to 21 based on conceptual and statistical overlap (Table S1 in Multimedia Appendix 1): six Sleep Disorders Questionnaire items were collapsed into a single sleep-disorder count (0 to 6); bilateral knee flexion was averaged into one variable, with absolute bilateral asymmetry retained as a separate gait-quality measure; redundant IPAQ frequency or duration counts were dropped in favor of four MET-summary variables (sitting, walking, moderate, and vigorous); the EQ-5D visual analog scale was excluded because it overlaps with the five dimensions and the utility index; and Expectations Regarding Aging subscale scores were collapsed into the Expectations Regarding Aging-12 total. With 314 self-reported knee-pain events in the 852 analytic cohort, this supported an EPV of 14.95, exceeding the conventional threshold of 10 or greater for stable prediction model estimation [31]. Data-driven variable selection was not applied, and the rule-based reduction was conducted before any outcome-aware modeling step.

Multiple Imputation

Across the 21 retained variables, the maximum item-level missingness was 1.88%, with a median of 1.17% (IQR 0.47%-1.29%; Table S2 in Multimedia Appendix 1). Missing values were addressed using multiple imputation by chained equations with predictive mean matching, m=5 imputed datasets, and 30 iterations per chain. Convergence was inspected using trace plots (Figure S1 in Multimedia Appendix 1). All subsequent analyses were conducted on each of the 5 imputed datasets, and results were pooled using Rubin's rules. The class distribution (knee pain "Yes" [314/852] 36.9%, "No" [538/852] 63.1%) constituted mild imbalance, addressed through imbalance-aware metrics (area under the receiver operating characteristic curve [AUC], F_1-score, and Matthews correlation coefficient [MCC]) rather than algorithmic resampling.

Nested Cross-Validation and Comparison With Logistic Regression Baseline

Six classification algorithms were compared: logistic regression (LR) with elastic-net penalty (LR; clinically interpretable baseline), k-nearest neighbors, random forest (RF), Extreme Gradient Boosting (XGB), Light Gradient Boosting Machine (LightGBM), and support vector machine (SVM) with radial basis function kernel. Performance was estimated using nested 5×5 cross-validation: 5 stratified outer folds for unbiased performance estimation, with an inner 5-fold cross-validation within each outer training fold for hyperparameter grid-search tuning (Table S3 in Multimedia Appendix 1). Continuous predictors were centered and scaled within each fold. This procedure was repeated independently on each of the 5 multiply-imputed datasets, and pooled estimates with 95% CIs were obtained via Rubin rules. For each algorithm, we report pooled AUC, F_1-score, precision, recall, Matthews correlation coefficient [MCC], accuracy, and Brier score with 95% CIs.

SHAP-Based Marginal Importance and Partial Dependence

Each variable's marginal contribution to the model output was quantified using SHAP values computed with the fastshap package and 50 Monte Carlo permutation samples on the best-performing caret-compatible algorithm [11]. SHAP values were computed within each imputation and averaged. We report mean absolute SHAP for global ranking and beeswarm plots for per-participant distributions. To visualize the marginal model–implied relationship between each top-ranked feature and the predicted probability of knee pain, partial dependence plots (PDPs) were generated for the top 9 SHAP features using the iml package with a 30-point grid. PDPs are model-dependent visualizations that do not establish causal effects or constitute clinical thresholds; we therefore frame all PDP-derived patterns as exploratory.

Partial-Correlation Network Analysis With Paradigm Concordance and Pathway Decomposition

To characterize the conditional dependency structure among predictor variables and the outcome, we estimated partial-correlation networks using graphical LASSO with extended Bayesian information criterion (EBICglasso) regularization, Spearman correlations to accommodate

XSL•FO
RenderX

nonnormal distributions, EBIC tuning parameter gamma of 0.5, and edge thresholding [32]. Network nodes comprised the top 10 SHAP-ranked features plus the binary knee pain outcome (n=11). This prespecified threshold was chosen to keep the node count compatible with stable EBICglasso estimation at our sample size, to focus the network on the predictors carrying the most model-relevant information, and to preserve interpretability; the sensitivity analysis confirmed a consistent network structure. One prespecified rule was applied: when EQ-5D Utility appeared in the top 10 alongside one or more individual EQ-5D dimensions, Utility was excluded (because it is a deterministic weighted combination of the dimensions) and replaced with the next-ranked SHAP feature. Networks were estimated separately on each imputation, and partial-correlation matrices were averaged. Three centrality measures were computed (strength, closeness, and betweenness); strength was prespecified as the primary measure, as closeness and betweenness are known to be unstable in cross-sectional networks of modest size. Centrality stability was assessed using case-dropping bootstrap (1000 resamples, first imputation), with the correlation-stability (CS) coefficient reported [33].

To examine paradigm concordance, we prespecified the Spearman rank correlation between SHAP importance and each centrality measure across the 10 network feature nodes (excluding the knee-pain node) as a quantitative test. We additionally decomposed each variable's contribution into a direct pathway, operationalized as the absolute partial correlation with the knee-pain node, and an indirect pathway, operationalized as the residual SHAP magnitude after subtracting the direct edge contribution. This decomposition is descriptive, not causal: it characterizes whether a variable's contribution to the model's output corresponds to a direct conditional dependency with the outcome or is instead accounted for by conditional dependencies with other variables in the network.

Sensitivity Analysis for Outcome-Feature Conceptual Overlap

EQ-5D Pain/Discomfort is a self-rated pain item that shares conceptual content with the outcome. To examine whether model results depended on this overlap, we conducted a prespecified sensitivity analysis excluding EQ-5D Pain/Discomfort from the feature set (n=20 variables; EPV=15.7). The full pipeline was repeated unchanged, and results are reported in Table S4 in Multimedia Appendix 1.

Results

Baseline Characteristics

Of the 852 participants, 314 (36.9%) reported knee pain in the preceding 3 months and 538 (63.1%) did not. Baseline characteristics by knee pain status are presented in Table 1. The two groups were comparable in age (2-tailed t_{693}=1.27; P=.21), with a small difference in the proportion of female participants (χ^2_1=5.08; P=.02). Participants with knee pain had higher BMI (t_{613}=−3.22; P=.001), were more frequently rental estate residents (χ^2_1=35.57; P<.001), and showed slower walking speed (t_{702}=2.41; P=.02) and fewer sit-to-stand repetitions (t_{713}=3.85; P<.001). They reported more sitting time (t_{635}=−3.81; P<.001), less walking-related (t_{816}=4.61; P<.001), and vigorous physical activity (all P<.001), poorer health-related quality of life across multiple EQ-5D dimensions, lower mental health scores (t_{640}=2.74; P=.006), and a higher count of sleep disorder categories (t_{553}=−4.88; P<.001). The largest between-group difference was observed for EQ-5D Pain/Discomfort (Cohen d=−0.73), followed by EQ-5D Utility (Cohen d=0.57) and EQ-5D Mobility (Cohen d=−0.39).

Model Performance

Pooled performance metrics for the six classification algorithms are presented in Table 2 and Figure 2. AUC values clustered within a narrow range from 0.692 (95% CI 0.661-0.724) for LightGBM to 0.723 (95% CI 0.686-0.760) for RF. The LR baseline showed a pooled AUC of 0.721 (95% CI 0.691-0.751), placing it within 0.002 of RF and 0.001 of SVM. F_1-score, recall, and MCC were modest across all algorithms (MCC range 0.252-0.328), and LR produced the lowest Brier score (0.200), indicating the best calibration. Performance was stable across the 5 imputations, with within-imputation variation in mean AUC less than 0.02 for every algorithm (Figure 2D). The discrimination performance of all six algorithms was effectively comparable; RF was retained as the best-performing algorithm by pooled AUC for SHAP analysis.

XSL•FO
RenderX

Table . Pooled performance metrics for the six classification algorithms.

Algorithm	AUC[a], pooled estimate (95% CI)	F_1-score, pooled estimate (95% CI)	Precision, pooled estimate (95% CI)	Recall, pooled estimate (95% CI)	MCC[b], pooled estimate (95% CI)	Brier, pooled estimate (95% CI)
RF[c]	0.723 (0.686 - 0.760)	0.510 (0.473 - 0.547)	0.649 (0.604 - 0.694)	0.423 (0.376 - 0.469)	0.325 (0.279 - 0.371)	0.201 (0.191 - 0.211)
SVM[d]	0.722 (0.687 - 0.757)	0.523 (0.472 - 0.575)	0.636 (0.584 - 0.688)	0.451 (0.375 - 0.527)	0.328 (0.276 - 0.380)	0.202 (0.190 - 0.215)
LR[e]	0.721 (0.691 - 0.751)	0.466 (0.337 - 0.595)	0.666 (0.597 - 0.736)	0.368 (0.232 - 0.505)	0.312 (0.212 - 0.412)	0.200 (0.192 - 0.208)
XGB[f]	0.707 (0.680 - 0.734)	0.530 (0.498 - 0.562)	0.598 (0.552 - 0.643)	0.480 (0.429 - 0.531)	0.306 (0.260 - 0.351)	0.211 (0.197 - 0.225)
kNN[g]	0.695 (0.650 - 0.740)	0.484 (0.426 - 0.542)	0.565 (0.507 - 0.623)	0.427 (0.357 - 0.498)	0.252 (0.179 - 0.325)	0.211 (0.195 - 0.227)
LightGBM[h]	0.692 (0.661 - 0.724)	0.530 (0.487 - 0.573)	0.574 (0.523 - 0.626)	0.497 (0.434 - 0.559)	0.289 (0.229 - 0.350)	0.249 (0.215 - 0.283)

[a]AUC: area under the receiver operating characteristic curve.

[b]MCC: Matthews correlation coefficient.

[c]RF: random forest.

[d]SVM: support vector machine.

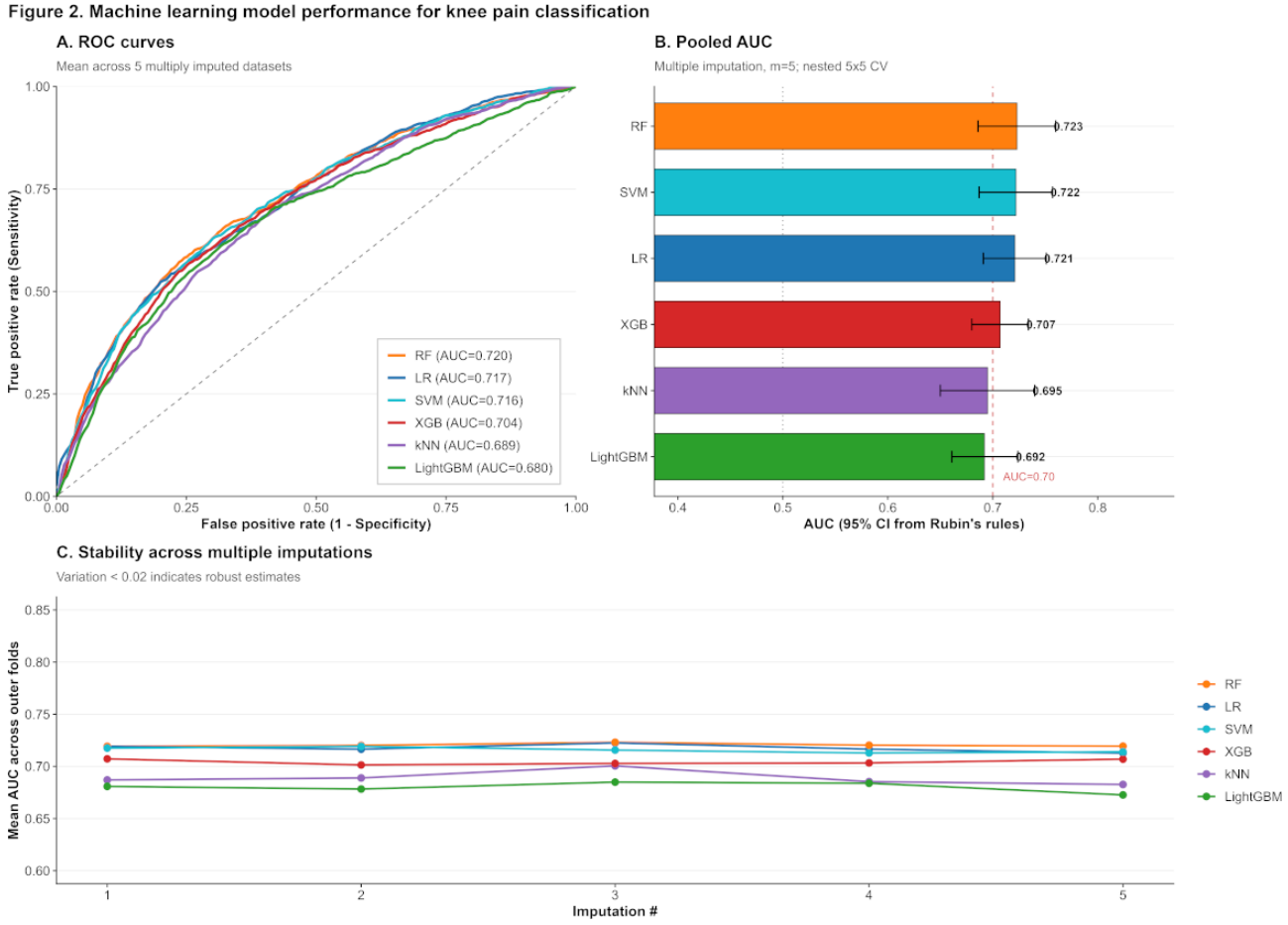
[e]LR: logistic regression.

[f]XGB: Extreme Gradient Boosting.

[g]kNN: k-nearest neighbor.

[h]LightGBM: Light Gradient Boosting Machine

Figure 2. Machine learning model performance for knee pain classification. (A) Receiver operating characteristic (ROC) curves; (B) Pooled area under the ROC curve (AUC); (C) Stability across the 5 multiply-imputed datasets. AUC: area under the receiver operating characteristic curve; CV: cross-validation; ROC: receiver operating characteristic; kNN: k-nearest neighbors; LightGBM: Light Gradient Boosting Machine; LR: logistic regression with elastic-net penalty; MCC: Matthew correlation coefficient; RF: random forest; SVM: support vector machine; XGB: Extreme Gradient Boosting.



XSL•FO
RenderX

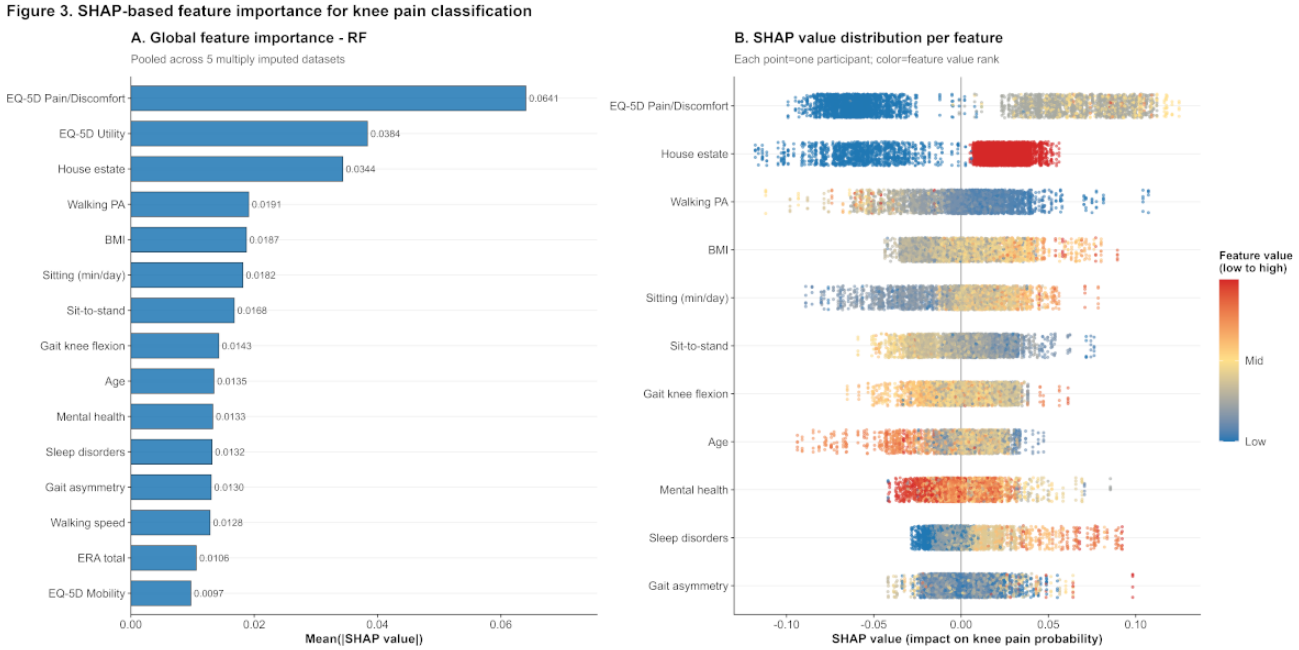
Values are pooled estimates (95% CI) across five multiply-imputed datasets, obtained via Rubin's rules from nested 5×5 cross-validation. Algorithms are ordered by pooled AUC.

Pooled performance of six classification algorithms (LR with elastic-net penalty, k-nearest neighbors, RF, XGB, LightGBM, and SVM with radial basis function kernel) estimated using nested 5×5 cross-validation on five multiply-imputed datasets and pooled via Rubin rules.

SHAP-Based Marginal Contribution

Pooled SHAP-based feature importance for the RF model is presented in Figure 3, with values for all 21 variables in Table S5 in Multimedia Appendix 1. The three highest-ranked features were EQ-5D Pain/Discomfort (mean |SHAP|=0.064), EQ-5D Utility (0.038), and house estate (0.034). Subsequent features clustered at substantially lower magnitudes (|SHAP| range 0.013-0.019) and included physical activity, anthropometric, smartphone-based physical performance, and mental and sleep variables. The SHAP value distributions in Figure 3 characterized the direction of each association: higher EQ-5D Pain/Discomfort scores, lower EQ-5D Utility, rental estate residence, lower walking physical activity and sit-to-stand repetitions, higher BMI and sitting time, and higher sleep disorder counts were each associated with increased predicted probability of knee pain, consistent with the group differences observed in Table 1.

Figure 3. SHAP-based feature importance for knee pain classification. (A) Feature importance for mean absolute SHAP value for the top 15 features; (B) SHAP value distribution per feature (beeswarm plot). ERA: Expectations Regarding Aging; EQ-5D: EuroQol 5-Dimension; SHAP: Shapley Additive Explanations; BMI: body mass index; PA: physical activity.

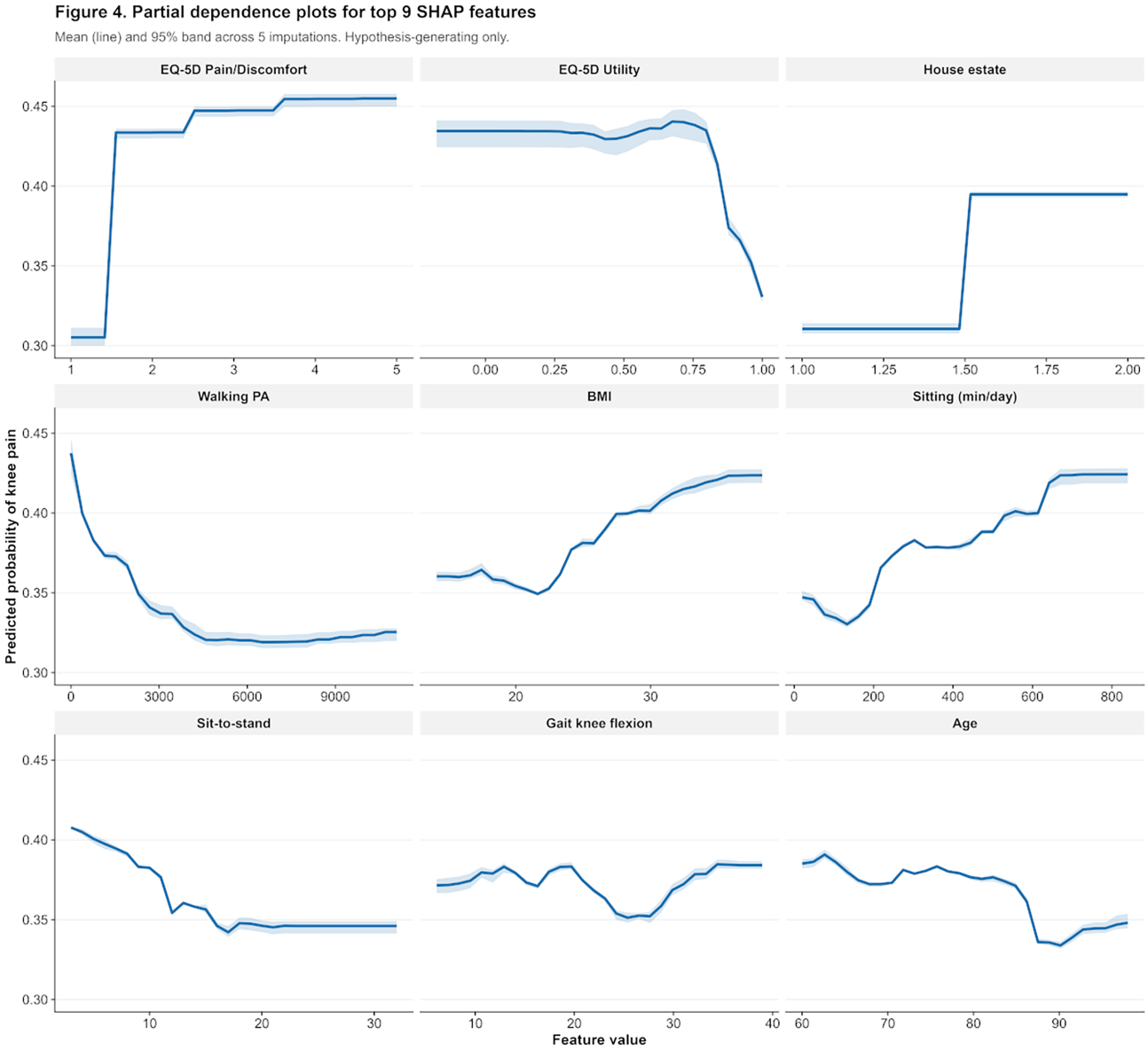


Partial Dependence Plots

PDPs for the top 9 SHAP-ranked features are shown in Figure 4. The model-implied probability of knee pain rose sharply between EQ-5D Pain/Discomfort scores of 1 and 2 (from approximately 0.31 to 0.43), reaching a plateau near 0.45. The relationship for EQ-5D Utility was relatively flat across most of the range and decreased steeply at utility values above approximately 0.85. House estate showed a step function consistent with its binary coding. For continuous features, the model-implied probability of knee pain decreased with higher walking physical activity and sit-to-stand repetitions and increased with higher BMI and sitting time. Gait knee flexion and age showed comparatively flat partial dependence functions. These patterns describe model-implied marginal relationships and are presented as exploratory; they do not establish causal effects or clinical thresholds.

XSL•FO
RenderX

Figure 4. Partial dependence plots for the top 9 SHAP-ranked features. Features are arranged in descending order of SHAP importance: EQ-5D Pain/Discomfort, EQ-5D Utility, house estate, walking physical activity, body mass index, sitting time, sit-to-stand repetitions, gait knee flexion, and age. EQ-5D: EuroQol 5-Dimension 5-Level; PA: physical activity; SHAP: Shapley Additive Explanations.



Network Structure, Paradigm Concordance, and Pathway Decomposition

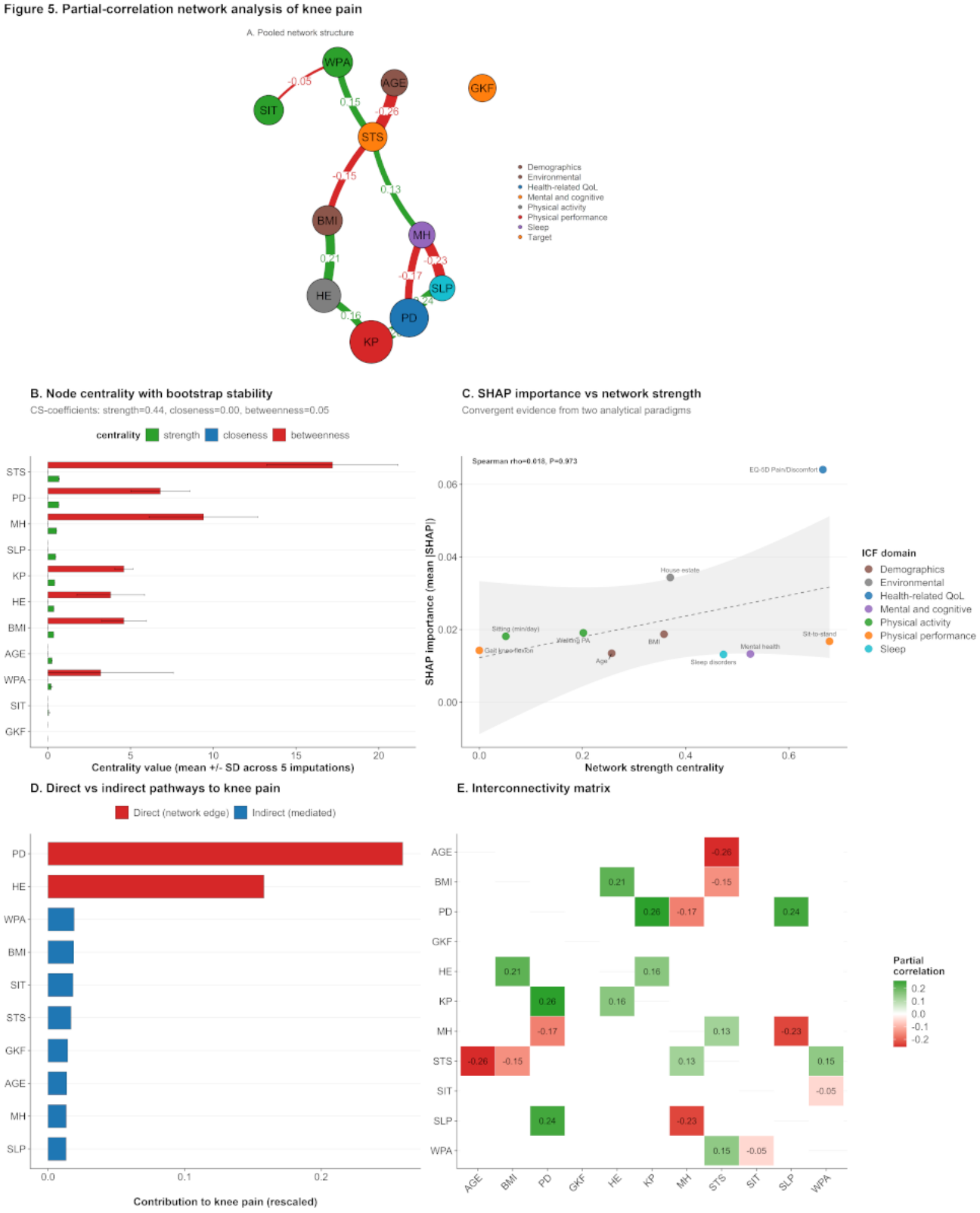
The pooled partial-correlation network for the 11 nodes is shown in Figure 5A, with the full pooled partial-correlation matrix in Table S6 in Multimedia Appendix 1. After EBICglasso regularization and edge thresholding, 11 conditional dependencies survived. The highest strength centrality values were observed for sit-to-stand (0.679), EQ-5D Pain/Discomfort (0.666), and mental health (0.526), with the knee pain node ranking 5th of 11 (0.418; Figure 5). The case-dropping bootstrap CS-coefficient was 0.36 for strength (acceptable stability), and 0.00 and 0.05 for closeness and betweenness, respectively.

Two direct edges connected the knee pain node to the predictor set: a positive edge with EQ-5D Pain/Discomfort (partial correlation=0.260) and a positive edge with house estate (0.158). The remaining nine network edges represented conditional dependencies among the predictors after controlling for all others. The prespecified Spearman rank correlation between SHAP importance and network strength centrality across the 10 feature nodes was ρ=0.018 (*P*=.97), indicating no monotonic association between the two paradigms (Figure 5).

The decomposition of each variable's contribution into direct and indirect pathways (Figure 5D) showed a clear binary pattern. EQ-5D Pain/Discomfort and house estate carried their contributions entirely through direct conditional dependencies, with edge magnitudes (0.260 and 0.158) larger than their corresponding mean absolute SHAP values (0.064 and 0.034). The remaining eight predictors had no direct edge to knee pain after regularization (Figure 5), indicating that their contributions to the model output were accounted for by their conditional dependencies with other variables in the network. This pattern of two variables with direct dependencies and eight contributing only through indirect pathways is consistent with the multidimensional structure described by the *ICF* framework.

XSL•FO
RenderX

Figure 5. Partial-correlation network, centrality, paradigm concordance, and pathway decomposition. (A) Pooled network structure, (B) node centrality with bootstrap stability, (C) linkage between SHAP importance and network strength centrality, (D) direct versus indirect pathways to knee pain, and (E) interconnectivity matrix. CS: correlation-stability; EBICglasso: graphical LASSO with extended Bayesian information criterion regularization; GKF: gait knee flexion; HE: house estate; *ICF*: *International Classification of Functioning, Disability and Health*; KP: knee pain; LASSO: Least Absolute Shrinkage and Selection Operator; MH: mental health; ML: machine learning; PD: EQ-5D-5L Pain/Discomfort dimension; SHAP: SHapley Additive exPlanations; SIT: sitting timea; SLP: sleep disorders; STS: sit-to-stand; WPA: walking physical activity; QoL: quality of life.



Sensitivity Analysis

Excluding EQ-5D Pain/Discomfort (n=20 variables; EPV=15.7) preserved the qualitative pattern of findings (Table S4 in Multimedia Appendix 1, Figures S1 and S2 in Multimedia Appendix 1). The pooled best AUC was 0.724 for SVM (ΔAUC =+0.001 vs main analysis; Figure S1 in Multimedia Appendix

XSL•FO
RenderX

1). The top SHAP feature shifted from EQ-5D Pain/Discomfort to EQ-5D Utility, but the next-ranked variables (house estate, walking physical activity, and BMI) remained the same (Figure S2 in Multimedia Appendix 1). House estate retained a direct conditional dependency with knee pain (partial correlation=0.20). The Spearman correlation between SHAP importance and network strength remained close to zero (ρ=−0.10; *P*=.78), and the CS-coefficient for strength was 0.44. The principal findings, including the modest discrimination performance, the predominance of indirect pathways from physical performance and mental health variables to knee pain, and the divergence between SHAP and network rankings, did not depend on inclusion of EQ-5D Pain/Discomfort.

Discussion

Principal Findings

This study examined self-reported knee pain in 852 community-dwelling older adults in Hong Kong using ML and partial-correlation network analysis. Based on the *ICF* model, knee pain in older age is associated with body function and structure, activity, participation, environmental, and personal factors [6,9,34]. These findings support this multidimensional view. Across the 21 predictors covering seven *ICF* domains, no single variable was dominant, and the model output reflected a distributed contribution from the health-related quality of life, environmental, physical performance, physical activity, mental health, and sleep domains. These results agree with the biopsychosocial view of knee pain in older adults [35], and with the projected rise of knee osteoarthritis to 595 million prevalent cases globally by 2050 [4,36]. Combining SHAP and network analysis allowed the assessment of which variables relate to knee pain and how those variables relate to one another.

The six classification algorithms showed similar discrimination, with pooled AUC ranging from 0.692 to 0.723. LR (AUC 0.721, 95% CI, 0.691–.751) was within 0.002 of RF and 0.001 of SVM, and showed the lowest Brier score (0.200) among the six algorithms. These results agree with prior reports that ML did not outperform LR in clinical prediction studies [37], and with recent recommendations that adequate sample size and prespecified variable selection, rather than algorithmic complexity, are the main determinants of stable AI-based prediction [38]. The discrimination level (best AUC 0.723) was comparable with prior ML models for knee pain or knee osteoarthritis in older adults, in which reported AUCs ranged from 0.70 to 0.80 [39-41]. With mild class imbalance (36.9% vs 63.1%), algorithmic resampling can degrade calibration without improving discrimination [42]; imbalance-aware metrics (AUC, F_1-score, MCC, and Brier score) were therefore used [43]. The convergence of LR and ML performance suggests that, when EPV is 14.95 and the predictor set is conceptually prespecified, the benefit of more flexible algorithms over a regularized LR is small in cross-sectional knee pain classification.

SHAP-based importance and PDPs showed the model-implied association of each variable with the predicted probability of knee pain. The three highest-ranked features were EQ-5D Pain/Discomfort (mean |SHAP|=0.064), EQ-5D Utility (0.038), and house estate (0.034). The remaining features clustered at lower magnitudes (mean |SHAP| range, 0.013 - 0.019) and included variables from the physical activity, sociodemographic, smartphone-based physical performance, mental health, and sleep domains. The SHAP-derived directions of association agreed with the baseline comparisons: lower walking-related physical activity and sit-to-stand repetitions, higher BMI and sitting time, lower mental health scores, and a higher count of sleep disorder categories were each associated with an increased predicted probability of knee pain. These patterns agree with prior reports linking reduced gait speed and lower-extremity capacity to musculoskeletal pain in older adults [44], and with the bidirectional association between sleep disturbance and chronic musculoskeletal pain [45]. Because PDPs show model-implied marginal associations rather than causal effects, the inflection points described below are reported as hypothesis-generating and require prospective validation. The predicted probability of knee pain rose sharply between EQ-5D Pain/Discomfort scores of 1 and 2 (0.31 to 0.43) and declined above an EQ-5D Utility of 0.85, while for the modifiable variables it declined steeply with walking physical activity up to 3000 MET·minutes/week, with sit-to-stand performance up to 15 repetitions, and rose above a BMI of 22 kg/m² and above a sitting time of 150 minutes/day. These patterns are consistent with prior reports linking sedentary behavior, low physical activity, higher BMI, and reduced lower-extremity capacity to musculoskeletal pain in older adults, but the exact values should not be treated as clinical cutoffs.

Network analysis showed how variables relate to one another after adjusting for the rest of the network. In the pooled 11-node network, only two predictors had direct edges with the knee pain node after EBICglasso regularization: EQ-5D Pain/Discomfort (partial correlation=0.260) and house estate (0.158). The remaining nine edges were among predictors. The Spearman correlation between SHAP importance and network strength centrality across the 10 feature nodes was ρ=0.018, indicating that the two analyses ranked variables differently and captured distinct facets of the multidimensional structure. Sit-to-stand showed this difference clearly. Although its marginal SHAP magnitude was modest (mean |SHAP|=0.0168; rank 7 of 21) and no direct edge to knee pain survived regularization, sit-to-stand showed the highest strength centrality (0.679). Sit-to-stand therefore acted as a central physical performance hub whose conditional dependencies extended across multiple *ICF* domains, including body structure and function (BMI, partial correlation = −0.146; age, −0.260), mental and cognitive function (mental health, +0.128), and activity (walking physical activity, +0.150). This pattern is consistent with the role of the sit-to-stand test as an integrative measure of lower-extremity power, balance, and global physical capacity in older adults [46,47]. Sit-to-stand therefore appears not as a direct correlate of knee pain, but as a multidimensional hub that links body function, body structure, activity, and mental health variables, which agrees with the biopsychosocial pathways described in the *ICF* model [15,48,49]. The direct edge between house estate and knee pain reflects rental estate residency as an indicator of lower socioeconomic position in Hong Kong, and is consistent with the well-documented inverse association between socioeconomic status and chronic pain [50], and with

XSL•FO
RenderX

reports that public rental housing in Hong Kong is associated with poorer health-related behaviors and chronic disease outcomes [51]. Beyond reflecting socioeconomic position, the housing categories in this cohort were defined primarily by income level rather than by physical or care-related features, so the mechanisms underlying the association with knee pain remain uncertain and warrant further study. The direct edge between EQ-5D Pain/Discomfort and knee pain partly reflects conceptual overlap, as both capture self-reported pain; however, the two are distinct constructs (general pain or discomfort versus knee-specific pain), and a sensitivity analysis excluding EQ-5D Pain/Discomfort preserved the qualitative pattern of findings, indicating that our conclusions do not depend on this predictor.

Several implications for clinical practice and future research are noted. First, the multidimensional pattern across *ICF* domains supports multidisciplinary, biopsychosocial assessment for older adults with knee pain, rather than the use of isolated impairment-based measures [34,35]. Second, the central network position of sit-to-stand, despite its weak direct association with knee pain, indicates that this measure was conditionally connected to body structure, mental health, and activity domains within the observed cross-sectional network, in which sit-to-stand was simultaneously connected to BMI and age, mental health, and walking physical activity. Sit-to-stand performance was therefore associated with multiple *ICF* domains in this network, a pattern that would be missed by considering its marginal SHAP rank alone and that prospective studies should test directly. This relationship is consistent with reports linking sit-to-stand power to functional outcomes in advanced knee osteoarthritis [26], and with the general role of sit-to-stand as a functional vital sign in geriatric assessment [46]. Third, the smartphone-based digital health system used in this study, validated for walking speed (r=0.975), knee flexion (r=0.907 - 0.923), and test-retest reliability (intraclass correlation coefficient=0.86 - 0.94) [16], provides a feasible method for measuring *ICF* body function and activity domain variables in community settings. Because these inferences are hypothesis-generating, independent external validation in separate cohorts is a necessary next step before any clinical implementation, alongside prospective longitudinal studies.

This study had several limitations. First, the cross-sectional design did not allow causal inference; the direct and indirect pathways describe conditional dependencies and model-implied associations, not mechanistic effects. Second, knee pain was measured by a single self-report item covering the preceding 3 months and did not distinguish etiology (osteoarthritis, rheumatoid arthritis, meniscal injury, or referred pain), severity, or laterality; medical diagnoses and medication use for knee pain (such as oral or topical nonsteroidal anti-inflammatory drugs, duloxetine, or intra-articular injections) were not ascertained and could not be included in the modeling. These omissions limit the clinical interpretability of the outcome and may contribute to the modest discrimination observed. Third, participants were recruited from eleven Hong Kong Housing Society estates by convenience sampling and were predominantly female (77.5%); generalizability to other regions, ethnicities, sex distributions, and socioeconomic strata is therefore limited. Fourth, no external validation was performed; nested 5×5 cross-validation provided unbiased internal estimates but not external transportability. Fifth, strength centrality showed acceptable bootstrap stability (CS-coefficient=0.36), but closeness and betweenness were unstable; these two measures were therefore not interpreted further. Sixth, the interpretive tools have inherent limits: PDPs show model-implied marginal associations, not clinical thresholds; SHAP values are model-dependent and may split importance among correlated predictors; and centrality metrics, especially closeness and betweenness, can be unstable in modest cross-sectional networks, which is why strength was pre-specified as the primary measure. Finally, the smartphone-based assessment used a research-prototype application that is not commercially available, and broader translation will require additional regulatory and engineering work. Future prospective and longitudinal studies with larger and more diverse cohorts, severity-graded outcomes, and independent external validation are required to test the hypotheses generated by this work.

Conclusions

This study combined ML and partial-correlation network analysis to examine self-reported knee pain in community-dwelling older adults within the *ICF* framework. The six classification algorithms showed pooled AUC values ranging from 0.692 to 0.723, indicating modest discrimination across the predictor set. Network analysis showed that only EQ-5D Pain/Discomfort and house estate had direct edges with knee pain, while sit-to-stand acted as a central hub linking body function, body structure, mental health, and activity variables. These cross-sectional, hypothesis-generating findings may inform multidisciplinary biopsychosocial assessment, but not causal or intervention conclusions, pending prospective external validation.

Acknowledgments

We would like to thank all participants in our study for their active participation and cooperation.

Funding

This research was supported by Research Centre for Innovative Technologies for Chronic Musculoskeletal Pain (grant number: 1-CE0K) in Hong Kong Polytechnic University.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request. Code for all statistical analyses may be available upon reasonable request from the corresponding author.

XSL•FO
RenderX

Authors' Contributions

Conceptualization, writing—original draft, and visualization: UH and SF
Supervision and project administration: SF
Data curation, validation, and software: UH, PX, and TF
Data curation and formal analysis: AYLW
Methodology and writing—review & editing: UH, OK, and AYLW

Conflicts of Interest

None declared.

Multimedia Appendix 1
The full list of baseline assessment variables by ICF domain, missing data summary, hyperparameter grids, SHAP-based feature importance for all 21 predictors, the pooled partial correlation matrix, and results of the sensitivity analysis excluding the EQ-5D Pain/Discomfort dimension
[DOCX File, 148 KB - mhealth_v14i1e92466_app1.docx]

References

1. Hunter CW, Deer TR, Jones MR, et al. Consensus guidelines on interventional therapies for knee pain (STEP guidelines) from the American Society of Pain and Neuroscience. J Pain Res 2022;15:2683-2745. [doi: 10.2147/JPR.S370469] [Medline: 36132996]
2. Joseph GB, McCulloch CE, Nevitt MC, Lane NE, Majumdar S, Link TM. Machine learning models for clinical and structural knee osteoarthritis prediction: recent advancements and future directions. Osteoarthr Cartil Open 2025 Sep;7(3):100654. [doi: 10.1016/j.ocarto.2025.100654] [Medline: 40799630]
3. Castagno S, Gompels B, Strangmark E, et al. Understanding the role of machine learning in predicting progression of osteoarthritis. Bone Joint J 2024 Nov 1;106-B(11):1216-1222. [doi: 10.1302/0301-620X.106B11.BJJ-2024-0453.R1] [Medline: 39481441]
4. Steinmetz JD, Culbreth GT, Haile LM, et al. Global, regional, and national burden of osteoarthritis, 1990–2020 and projections to 2050: a systematic analysis for the Global Burden of Disease Study 2021. The Lancet Rheumatology 2023 Sep;5(9):e508-e522. [doi: 10.1016/S2665-9913(23)00163-7] [Medline: 37675071]
5. Bagraith KS, Strong J. The International Classification of Functioning, Disability and Health (ICF) can be used to describe multidisciplinary clinical assessments of people with chronic musculoskeletal conditions. Clin Rheumatol 2013 Mar;32(3):383-389. [doi: 10.1007/s10067-012-2130-1] [Medline: 23318704]
6. Fréz AR, Alouche SR, Binda AC, Vieira GW, Bueno BAM, Cabral CMN. Development of a core set for knee dysfunction based on the international classification of functioning, disability and health: a cross-sectional study. Arch Phys Med Rehabil 2021 Apr;102(4):571-581. [doi: 10.1016/j.apmr.2020.10.139] [Medline: 33278366]
7. Weigl M, Wild H. European validation of The Comprehensive International Classification of Functioning, Disability and Health Core Set for Osteoarthritis from the perspective of patients with osteoarthritis of the knee or hip. Disabil Rehabil 2018 Dec;40(26):3104-3112. [doi: 10.1080/09638288.2017.1377295] [Medline: 28911242]
8. Solmi M, Koyanagi A, Thompson T, Fornaro M, Correll CU, Veronese N. Network analysis of the relationship between depressive symptoms, demographics, nutrition, quality of life and medical condition factors in the Osteoarthritis Initiative database cohort of elderly North-American adults with or at risk for osteoarthritis – CORRIGENDUM. Epidemiol Psychiatr Sci 2020;29:e14. [doi: 10.1017/S2045796019000064]
9. Pollard B, Johnston M, Dieppe P. Exploring the relationships between International Classification of Functioning, Disability and Health (ICF) constructs of impairment, activity limitation and participation Restriction in people with osteoarthritis prior to joint replacement. BMC Musculoskelet Disord 2011 May 16;12(1):97. [doi: 10.1186/1471-2474-12-97] [Medline: 21575238]
10. Oberhauser C, Escorpizo R, Boonen A, Stucki G, Cieza A. Statistical validation of the brief International Classification of Functioning, Disability and Health Core Set for osteoarthritis based on a large international sample of patients with osteoarthritis. Arthritis Care Res 2013 Feb;65(2):177-186 [FREE Full text] [doi: 10.1002/acr.21775]
11. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. arXiv. Preprint posted online on Nov 25, 2017. [doi: 10.48550/arXiv.1705.07874]
12. Kim JH. Comparison of predictive models for knee pain and analysis of individual and physical activity variables using interpretable machine learning. Knee 2025 Jun;54:146-153. [doi: 10.1016/j.knee.2025.02.006] [Medline: 40043328]
13. Borsboom D, Deserno MK, Rhemtulla M, et al. Network analysis of multivariate data in psychological science. Nat Rev Methods Primers 2021;1:58. [doi: 10.1038/s43586-021-00055-w]
14. Gevers-Montoro C, Liew BXW, Deldar Z, et al. A network analysis on biopsychosocial factors and pain-related outcomes assessed during a COVID-19 lockdown. Sci Rep 2023 Mar 16;13(1):4399. [doi: 10.1038/s41598-023-31054-4] [Medline: 36928233]

XSL•FO
RenderX

15. Courbalay A, Jobard R, Descarreaux M, Bouvard B. Direct and indirect relationships between physical activity, fitness level, kinesiophobia, and health-related quality of life in patients with rheumatic and musculoskeletal diseases: a network analysis. J Pain Res 2021;14:3387-3399. [doi: 10.2147/JPR.S323424] [Medline: 34737633]
16. Leung KL, Li Z, Huang C, Huang X, Fu SN. Validity and reliability of gait speed and knee flexion estimated by a novel vision-based smartphone application. Sensors (Basel) 2024 Nov 28;24(23):7625. [doi: 10.3390/s24237625] [Medline: 39686162]
17. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024 Apr 16;385:e078378. [doi: 10.1136/bmj-2023-078378] [Medline: 38626948]
18. Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016 Dec 16;18(12):e323. [doi: 10.2196/jmir.5870] [Medline: 27986644]
19. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011 Dec;20(10):1727-1736. [doi: 10.1007/s11136-011-9903-x] [Medline: 21479777]
20. Sarkisian CA, Steers WN, Hays RD, Mangione CM. Development of the 12-item expectations regarding aging survey. Gerontologist 2005 Apr;45(2):240-248. [doi: 10.1093/geront/45.2.240] [Medline: 15799989]
21. Keetharuth AD, Brazier J, Connell J, et al. Recovering Quality of Life (ReQoL): a new generic self-reported outcome measure for use with people experiencing mental health difficulties. Br J Psychiatry 2018 Jan;212(1):42-49. [doi: 10.1192/bjp.2017.10] [Medline: 29433611]
22. Hombali A, Seow E, Yuan Q, et al. Prevalence and correlates of sleep disorder symptoms in psychiatric disorders. Psychiatry Res 2019 Sep;279:116-122. [doi: 10.1016/j.psychres.2018.07.009] [Medline: 30072039]
23. Tomioka K, Iwamoto J, Saeki K, Okamoto N. Reliability and validity of the International Physical Activity Questionnaire (IPAQ) in elderly adults: the Fujiwara-kyo study. J Epidemiol 2011;21(6):459-465. [doi: 10.2188/jea.je20110003] [Medline: 21946625]
24. Bindawas SM. Relationship between frequent knee pain, obesity, and gait speed in older adults: data from the osteoarthritis Initiative. Clin Interv Aging 2016;11:237-244. [doi: 10.2147/CIA.S100546] [Medline: 26955266]
25. Seydi M, Delbaere K, Han DU, Chan L, Ambrens M, van Schooten KS. The effect of pain on gait in older people: a systematic review and meta-analysis. J Pain 2025 Apr;29:104758. [doi: 10.1016/j.jpain.2024.104758] [Medline: 39672448]
26. Langgård Jørgensen S, Mechlenburg I, Bagger Bohn M, Aagaard P. Sit-to-stand power predicts functional performance and patient-reported outcomes in patients with advanced knee osteoarthritis. A cross-sectional study. Musculoskelet Sci Pract 2024 Feb;69:102899. [doi: 10.1016/j.msksp.2023.102899] [Medline: 38141496]
27. Khuna L, Soison T, Plukwongchuen T, Tangadulrat N. Reliability and concurrent validity of 30-s and 5-time sit-to-stand tests in older adults with knee osteoarthritis. Clin Rheumatol 2024 Jun;43(6):2035-2045. [doi: 10.1007/s10067-024-06969-6] [Medline: 38634965]
28. McCarthy I, Hodgins D, Mor A, Elbaz A, Segal G. Analysis of knee flexion characteristics and how they alter with the onset of knee osteoarthritis: a case control study. BMC Musculoskelet Disord 2013 May 21;14(1):169. [doi: 10.1186/1471-2474-14-169] [Medline: 23692671]
29. Li Z, Leung KL, Huang C, et al. Higher knee flexion moment during walking is associated with a lower risk of knee pain developing among the elderly after 24 months. Eur J Phys Rehabil Med 2023 Jun;59(3):386-395. [doi: 10.23736/S1973-9087.23.07798-5] [Medline: 37195650]
30. Austin PC, Steyerberg EW. Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. Stat Methods Med Res 2017 Apr;26(2):796-808. [doi: 10.1177/0962280214558972] [Medline: 25411322]
31. Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: part II - binary and time - to - event outcomes. Stat Med 2019 Mar 30;38(7):1276-1296. [doi: 10.1002/sim.7992] [Medline: 30357870]
32. Epskamp S, Fried EI. A tutorial on regularized partial correlation networks. Psychol Methods 2018 Dec;23(4):617-634. [doi: 10.1037/met0000167] [Medline: 29595293]
33. Epskamp S, Borsboom D, Fried EI. Estimating psychological networks and their accuracy: a tutorial paper. Behav Res Methods 2018 Feb;50(1):195-212. [doi: 10.3758/s13428-017-0862-1] [Medline: 28342071]
34. Logerstedt DS, Snyder-Mackler L, Ritter RC, Axe MJ, Orthopedic Section of the American Physical Therapy Association. Knee pain and mobility impairments: meniscal and articular cartilage lesions. J Orthop Sports Phys Ther 2010 Jun;40(6):A1-A35. [doi: 10.2519/jospt.2010.0304] [Medline: 20511698]
35. Bayrak G, Alkan H. Factors influencing pain intensity in knee osteoarthritis: a cross-sectional biopsychosocial perspective. BMC Musculoskelet Disord 2025 Feb 25;26(1):196. [doi: 10.1186/s12891-025-08450-0] [Medline: 40001137]
36. Cui A, Li H, Wang D, Zhong J, Chen Y, Lu H. Global, regional prevalence, incidence and risk factors of knee osteoarthritis in population-based studies. EClinicalMedicine 2020 Dec;29-30:100587. [doi: 10.1016/j.eclinm.2020.100587] [Medline: 34505846]

XSL•FO
RenderX

37. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019 Jun;110:12-22. [doi: 10.1016/j.jclinepi.2019.02.004] [Medline: 30763612]
38. Riley RD, Ensor J, Snell KIE, et al. Importance of sample size on the quality and utility of AI-based prediction models for healthcare. Lancet Digit Health 2025 Jun;7(6):100857. [doi: 10.1016/j.landig.2025.01.013] [Medline: 40461350]
39. Chen T, Or CK. Automated machine learning-based prediction of the progression of knee pain, functional decline, and incidence of knee osteoarthritis in individuals at high risk of knee osteoarthritis: data from the osteoarthritis initiative study. Digit Health 2023;9:20552076231216419. [doi: 10.1177/20552076231216419] [Medline: 38033512]
40. Jamshidi A, Leclercq M, Labbe A, et al. Identification of the most important features of knee osteoarthritis structural progressors using machine learning methods. Ther Adv Musculoskelet Dis 2020;12:1759720X20933468. [doi: 10.1177/1759720X20933468] [Medline: 32849918]
41. Nielsen RL, Monfeuga T, Kitchen RR, et al. Data-driven identification of predictive risk biomarkers for subgroups of osteoarthritis using interpretable machine learning. Nat Commun 2024 Apr 1;15(1):2817. [doi: 10.1038/s41467-024-46663-4] [Medline: 38561399]
42. Sadeghi S, Khalili D, Ramezankhani A, Mansournia MA, Parsaeian M. Diabetes mellitus risk prediction in the presence of class imbalance using flexible machine learning methods. BMC Med Inform Decis Mak 2022 Feb 10;22(1):36. [doi: 10.1186/s12911-022-01775-z] [Medline: 35139846]
43. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019 Apr 4;380(14):1347-1358. [doi: 10.1056/NEJMra1814259] [Medline: 30943338]
44. Middleton A, Fritz SL, Lusardi M. Walking speed: the functional vital sign. J Aging Phys Act 2015 Apr;23(2):314-322. [doi: 10.1123/japa.2013-0236] [Medline: 24812254]
45. Runge N, Ahmed I, Saueressig T, et al. The bidirectional relationship between sleep problems and chronic musculoskeletal pain: a systematic review with meta-analysis. Pain 2024 Nov 1;165(11):2455-2467. [doi: 10.1097/j.pain.0000000000003279] [Medline: 38809241]
46. van Lummel RC, Walgaard S, Pijnappels M, et al. Physical performance and physical activity in older adults: associated but separate domains of physical function in old age. PLoS ONE 2015;10(12):e0144048. [doi: 10.1371/journal.pone.0144048] [Medline: 26630268]
47. Yee XS, Ng YS, Allen JC, et al. Performance on sit-to-stand tests in relation to measures of functional fitness and sarcopenia diagnosis in community-dwelling older adults. Eur Rev Aging Phys Act 2021 Jan 8;18(1):1. [doi: 10.1186/s11556-020-00255-5] [Medline: 33419399]
48. Åkerblom S, Cervin M, Perrin S, Rivano Fischer M, Gerdle B, McCracken LM. A network analysis of clinical variables in chronic pain: a study from the Swedish Quality Registry for Pain Rehabilitation (SQRP). Pain Med 2021 Jul 25;22(7):1591-1602. [doi: 10.1093/pm/pnaa473] [Medline: 33706371]
49. Ruocco Verengue B, Dos Santos PG, Miura BG, et al. Network analysis of cognitive, physical/functional, sensory, and pain processing-related factors in individuals with knee osteoarthritis. Musculoskelet Sci Pract 2025 Nov;80:103434. [doi: 10.1016/j.msksp.2025.103434] [Medline: 41086590]
50. Prego-Domínguez J, Khazaeipour Z, Mallah N, Takkouche B. Socioeconomic status and occurrence of chronic pain: a meta-analysis. Rheumatology (Oxford) 2021 Mar 2;60(3):1091-1105. [doi: 10.1093/rheumatology/keaa758] [Medline: 33276382]
51. Zhang T, Huang B, Wong H, Wong SYS, Chung RYN. Public rental housing and obesogenic behaviors among adults in Hong Kong: mediator role of food and physical activity environment. Int J Environ Res Public Health 2022 Mar 3;19(5):2960. [doi: 10.3390/ijerph19052960] [Medline: 35270652]

Abbreviations

AUC: area under the receiver operating characteristic curve
EBICglasso: extended Bayesian Information Criterion graphical LASSO
EPV: events per variable
EQ-5D: EuroQol 5-Dimension
***ICF*:** *International Classification of Functioning, Disability and Health*
LASSO: Least Absolute Shrinkage and Selection Operator
LightGBM: Light Gradient Boosting Machine
LR: logistic regression
MCC: Matthews Correlation Coefficient
ML: machine learning
PDP: partial dependence plot
RF: random forest
SHAP: SHapley Additive exPlanations
SVM: support vector machine

XSL•FO
RenderX

XGBoost: Extreme Gradient Boosting

Edited by R Chandrasekaran; submitted 29.Jan.2026; peer-reviewed by C Baxter, GGD Alencar, N Meng, P Kellmeyer; revised version received 23.Jun.2026; accepted 29.Jun.2026; published 03.Aug.2026.

Please cite as:
Hwang UJ, Xia P, Fan T, Wong AYL, Kwon OY, Fu SN
Smartphone-Based Physical Performance and Multidimensional Determinants of Self-Reported Knee Pain in Community-Dwelling Older Adults: Cross-Sectional Machine Learning and Network Analysis Study
JMIR Mhealth Uhealth 2026;14:e92466
URL: https://mhealth.jmir.org/2026/1/e92466
doi:10.2196/92466

© Ui-jae Hwang, Peng Xia, Tianxiang Fan, Arnold YL Wong, Oh-yun Kwon, Siu-ngor Fu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 3.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Utility of Tablet-Based Eye Tracking for Early Screening of Poststroke Cognitive Impairment: Diagnostic Cohort Study

Boxu Xie[1,2,3], BM; Ronghui Ren[1,2,3], BM; Yuting Zhang[1,2,3], BM; Xuan Wang[1,2,3], MM; Junli Zhen[1,2,3], Dr med

[1]Key Laboratory of Clinical Neurology, Ministry of Education, Hebei Medical University, 215 Heping West Road, Xinhua District, Shijiazhuang, Hebei, China

[2]Department of Neurology, the Second Hospital of Hebei Medical University, Shijiazhuang, Hebei, China

[3]Key Neurological Laboratory of Hebei Province, Shijiazhuang, Hebei, China

Corresponding Author:
Junli Zhen, Dr med
Key Laboratory of Clinical Neurology, Ministry of Education, Hebei Medical University, 215 Heping West Road, Xinhua District, Shijiazhuang, Hebei, China

Abstract

Background: Poststroke cognitive impairment (PSCI) is a common and disabling complication after stroke; however, early screening remains challenging due to limited access to neuropsychological testing and the high cost of neuroimaging. Portable, tablet-based eye-tracking technology may offer a scalable, low-cost solution for early PSCI detection.

Objective: This study aimed to evaluate the clinical utility of a tablet-based, AI-driven eye-tracking system for early screening of PSCI at 3 months after acute ischemic stroke. We sought to quantify oculomotor-cognitive associations and develop a practical nomogram for individualized risk prediction.

Methods: We prospectively enrolled 142 hospitalized patients with acute cerebral infarction between May 2023 and October 2024, of whom 122 completed the 3-month follow-up and were included in the final analysis, along with 20 healthy community-dwelling controls. All patients underwent tablet-based eye tracking (visual paired comparison and antisaccade tasks) during the acute phase, as well as baseline and 3-month neuropsychological assessments. PSCI was defined using validated cutoffs. Multivariable logistic regression was used to identify independent predictors, and a nomogram was constructed. Internal validation was performed using bootstrap resampling (1000 samples).

Results: At 3 months, out of 122 patients, 47 (38.5%) met PSCI criteria. Compared with patients with non-PSCI (n=75), patients with PSCI showed significantly prolonged correct saccade latency (median 322.96, IQR 209.45-445.59 ms vs 194.55, IQR 141.50-299.75 ms; Z=−4.03, P<.001), increased uncorrected error rate (median 30.00%, IQR 15.00%-42.00% vs 5.00%, IQR 0.00%-28.00%; Z=−4.24, P<.001), and reduced novelty preference ratio (median 1.44, IQR 0.97-1.70 vs 2.12, IQR 1.27-4.56; Z=−3.44, P=.001). Multivariable analysis identified 4 independent predictors of 3-month PSCI: older age (odds ratio [OR] 1.067 per year, 95% CI 1.009 - 1.129; P=.02), lower education level (OR 0.841 per year, 95% CI 0.708 - 0.999; P=.049), higher NIHSS (National Institutes of Health Stroke Scale) scores (OR 1.557 per point, 95% CI 1.075 - 2.256; P=.02), and prolonged correct saccade latency (OR 1.004 per ms, 95% CI 1.000 - 1.007; P=.04). A nomogram incorporating these 4 factors achieved good discriminative performance (area under the receiver operating characteristic curve 0.86, 95% CI 0.793 - 0.927) with satisfactory calibration.

Conclusions: Age, education, admission NIHSS, and correct saccade latency were identified as possible independent predictors of 3-month PSCI in this cohort. The tablet-based eye-tracking system, when combined with clinical variables, may represent a feasible approach for early PSCI screening. A nomogram based on these variables demonstrated high accuracy and potential clinical utility for early PSCI identification. This approach may facilitate early identification of high-risk patients and enable timely, personalized interventions in resource-limited settings.

(JMIR Mhealth Uhealth 2026;14:e83194) doi:10.2196/83194

KEYWORDS

acute ischemic stroke; poststroke cognitive impairment; eye tracking; tablet assessment; cognitive screening; mobile health technology

XSL•FO
RenderX

Introduction

Globally, stroke ranks as the second leading cause of mortality and frequently induces lifelong disabilities, including motor, cognitive, linguistic, and psychological impairments [1]. Among these sequelae, poststroke cognitive impairment (PSCI) constitutes a prevalent and debilitating neurological complication. PSCI is clinically defined as a syndrome fulfilling diagnostic criteria for cognitive impairment manifesting within 6 months post stroke. Formal neuropsychological assessment at 3 to 6 months post stroke is typically required for diagnosis [2]. This condition manifests clinically as deficits in memory, language comprehension, perceptual processing, visuospatial reasoning, and executive functioning [3]. Despite advances in acute stroke management, PSCI persists as a highly prevalent and debilitating condition [4], associated with elevated 5-year mortality and depression risks [5], thereby substantially contributing to long-term disability and reduced quality of life [6].

Early identification and management of PSCI represent critical clinical priorities [7]. Conventional tools such as the Mini-Mental State Examination (MMSE) and Montreal Cognitive Assessment (MoCA) exhibit limitations, including susceptibility to age, education, and emotional confounders, alongside significant time and resource requirements [8]. In recent years, mobile health (mHealth) technologies have shown great potential in enhancing the accessibility and scalability of medical assessments [9-11]. The integration of eye tracking into portable tablet devices allows for standardized, low-cost cognitive screening outside traditional clinical settings [12,13]. This approach aligns with growing emphasis on digital biomarkers and decentralized health care models, particularly in neurology and rehabilitation [14].

AI technology has gained substantial research interest owing to its objectivity and efficiency. AI-enhanced eye tracking quantitatively records ocular movements and gaze positions temporally and across task conditions. For target image projection and maintenance on the retinal fovea, coordinated eye movements are essential, assessed through saccadic, smooth pursuit, fixation, and visual search tasks [15]. As a noninvasive technique, eye tracking can identify cognitive impairment, monitor disease progression, and elucidate underlying cognitive processes [16]. Evidence suggests visual processing deficits may precede memory impairment in mild cognitive impairment (MCI), manifesting in altered eye movement patterns [17,18]. Patients with MCI frequently demonstrate elevated saccadic error rates, prolonged latency, and inverse correlations between erroneous saccade frequency and neuropsychological test performance [19]. In smooth pursuit tasks, patients with Alzheimer disease (AD) exhibit increased pursuit latency, reduced acceleration and gain, and frequent compensatory saccades correlating with cognitive decline [20].

These oculomotor abnormalities are not random; they directly reflect the functional state of the underlying neural circuits [21]. Precision eye tracking enables real-time parameter quantification, facilitating objective cognitive assessment [22]. Thus, AI-enhanced eye tracking captures millisecond-level ocular dynamics, offering an objective, quantifiable methodology for cognitive impairment evaluation.

Building on this evidence, researchers have recently extended eye tracking to stroke populations. Recent studies have applied eye tracking in stroke populations for various purposes, including augmenting stroke diagnosis [23], recognizing stroke through free viewing of faces [24], assessing the impact of poststroke fatigue on saccadic control [25], and evaluating memory-guided saccades in patients with subacute and chronic stroke [26]. These works confirm the feasibility of eye tracking in stroke-related conditions; however, none address early cognitive impairment. Regarding PSCI prediction, Zhou et al [27] built a nomogram using clinical and inflammatory markers. However, that model did not include oculomotor biomarkers. More recently, Chen et al [28] applied AI-assisted eye tracking with gait measurements to detect cognitive impairment in cerebral small vessel disease, directly supporting the utility of AI-driven oculomotor assessment in cerebrovascular disorders. A systematic review has further confirmed the diagnostic potential of eye tracking for AD and MCI, strengthening the rationale for applying oculomotor paradigms to cognitive screening [29]. Similarly, Maldonado-Díaz et al [30] used eye tracking to assess visual attention during virtual reality balance training in older adults with mild to moderate cognitive impairment, further illustrating the applicability of oculomotor measures across different cognitive impairment severities.

Despite these advances, no study has yet integrated a tablet-based, AI-enhanced eye-tracking system into a practical nomogram specifically for early PSCI screening. This gap motivated this study.

This study uses a noninvasive eye-tracking technology developed by Neuro Weave Ltd, which utilizes standard smartphone or tablet cameras to track subtle dynamic eye movements. By performing visual paired comparisons (VPCs) and antisaccade tasks on a tablet device, the assessment can be completed within minutes, significantly reducing time and labor compared to traditional paper-based scales. A study using this technology in patients with AD demonstrated a sensitivity of 82% and specificity of 91%, indicating high discriminative ability and low misdiagnosis risk in screening for AD [12]. Similarly, using the same tablet-based AI-driven eye-tracking technology, Wei et al [13] validated its performance in a large community-based cohort in China, achieving an area under the receiver operating characteristic curve (AUC) of 0.80 for detecting cognitive impairment and a high negative predictive value (87.2%), supporting its feasibility for population-level screening in resource-limited settings.

In summary, early warning and risk identification for PSCI are crucial. This study aims to explore the clinical utility of AI-based eye-tracking technology implemented on a tablet platform for screening early PSCI. By analyzing the relationship between eye movement characteristics and cognitive function in patients with stroke, we seek to develop and validate an AI eye-tracking model for early PSCI screening. The study hopes to provide an economical, convenient, and objective auxiliary diagnostic tool for early PSCI detection, thereby supporting clinical intervention

XSL•FO
RenderX

and rehabilitation strategies to improve long-term prognosis and quality of life for patients with stroke.

Methods

Participants

We prospectively recruited 142 patients with acute cerebral infarction who were admitted to the Department of Neurology at the Second Hospital of Hebei Medical University between May 2023 and October 2024. Of these, 122 completed the 3-month follow-up assessments and were included in the final analysis, along with 20 healthy community-dwelling controls. Inclusion criteria for patients with stroke were as follows: (1) age 18 to 80 years; (2) within 14 days after symptom onset; (3) fulfilling the American Heart Association/American Stroke Association diagnostic criteria for stroke [31]; (4) capacity to undergo neuropsychological testing, physical examination, and brain magnetic resonance imaging (MRI); (5) corrected visual acuity sufficient for testing; and (6) provision of informed consent. Exclusion criteria included the following: (1) significant aphasia; (2) pregnancy, active malignancy, major organ failure, or life expectancy less than 1 year; (3) prestroke dementia or neurodegenerative disorders (eg, Lewy body dementia, AD, or secondary dementias); (4) Informant Questionnaire on Cognitive Decline in the Elderly score ≥4; and (5) congenital intellectual disability or severe neuropsychiatric conditions. Control participants were recruited using the following criteria: (1) age 18 to 80 years, (2) capacity to undergo neuropsychological testing, (3) corrected visual acuity sufficient for testing, and (4) provision of informed consent. Exclusion criteria included the following: (1) history of cognitive disorders, congenital intellectual disability, or other severe neuropsychiatric comorbidities; (2) history of stroke; or (3) current cognitive impairment defined as below-threshold scores on either MMSE or MoCA. Potentially eligible participants were identified based on admission records and clinical diagnosis. A purposive sampling strategy was used.

Ethical Considerations

This study was approved by the Research Ethics Committee of the Second Hospital of Hebei Medical University (reference number 2023-C023) on March 31, 2023. All procedures followed the Declaration of Helsinki. Before enrollment, the researchers explained the study objectives in detail to the participants and their legal guardians, and written informed consent was obtained from the participants themselves or their guardians. All participants were informed that they could withdraw from the study at any time without penalty. Participant data were anonymized and deidentified before analysis; no personal identifiers are reported. No compensation was provided to participants for their involvement in this study.

Neuropsychological Assessment

All participants completed cognitive and mood assessments, including the MMSE, MoCA, Self-Rating Anxiety Scale, and Self-Rating Depression Scale. Patients with stroke underwent repeated assessments at 3 months post stroke. PSCI diagnosis required meeting either criterion: MMSE score <17 (illiterate), <20 (1‐6 y education), or <24 (≥7 y education) or MoCA score <22 (validated cutoff demonstrating optimal sensitivity/specificity) [32,33]. Patients were stratified into PSCI and non-PSCI (NPSCI) subgroups based on 3-month cognitive performance. All assessments were administered by neurologists certified in standardized neuropsychological testing protocols. The neurologists who administered the 3-month neuropsychological assessments were blinded to the patients' baseline eye-tracking results and the initial clinical assessments.

Eye-Tracking Technology Task

Eye-tracking data were acquired during the acute phase of cerebral infarction using a Xiaomi Pad 5 Pro 12.4 Android tablet (12.4-inch display, 2560×1600 resolution, 60 Hz refresh rate) with a front-facing camera capable of recording 1080p video at 30 frames per second, running a proprietary research software developed by Shanghai Neuroweave Technology Co, Ltd, for Android 8.0+ devices. This software is a research tool rather than a commercial retail product. Its validity for cognitive screening has been demonstrated in previous studies [12,13]. The system initiated recording upon task start, transmitting videos to Alibaba Cloud for processing via an AI eye-tracking model that integrated computer vision and deep learning architectures for facial landmark detection, head pose estimation, pupil tracking, and gaze vector calculation. Demographic data were input preassessment via user interface, followed by head pose calibration requiring participants to position within an on-screen virtual frame to standardize distance and minimize motion artifacts. The 6-minute protocol comprised (1) pretask calibration optimizing device-specific tracking, (2) VPC phase 1: five 5-second image sets for memorization, (3) antisaccade task (~3 min): central fixation with rapid gaze shifts to contralateral space upon peripheral red dot appearance, and (4) VPC phase 2: ten 6-second sets (1 familiar/1 novel image per set) where participants viewed novel stimuli while camera-recorded videos underwent cloud-based processing to extract quantitative eye movement-related parameters. After processing the video, a series of eye movement–related parameters were established, as shown in Table 1. An example diagram of the eye-tracking task is shown in Figure 1.

XSL•FO
RenderX

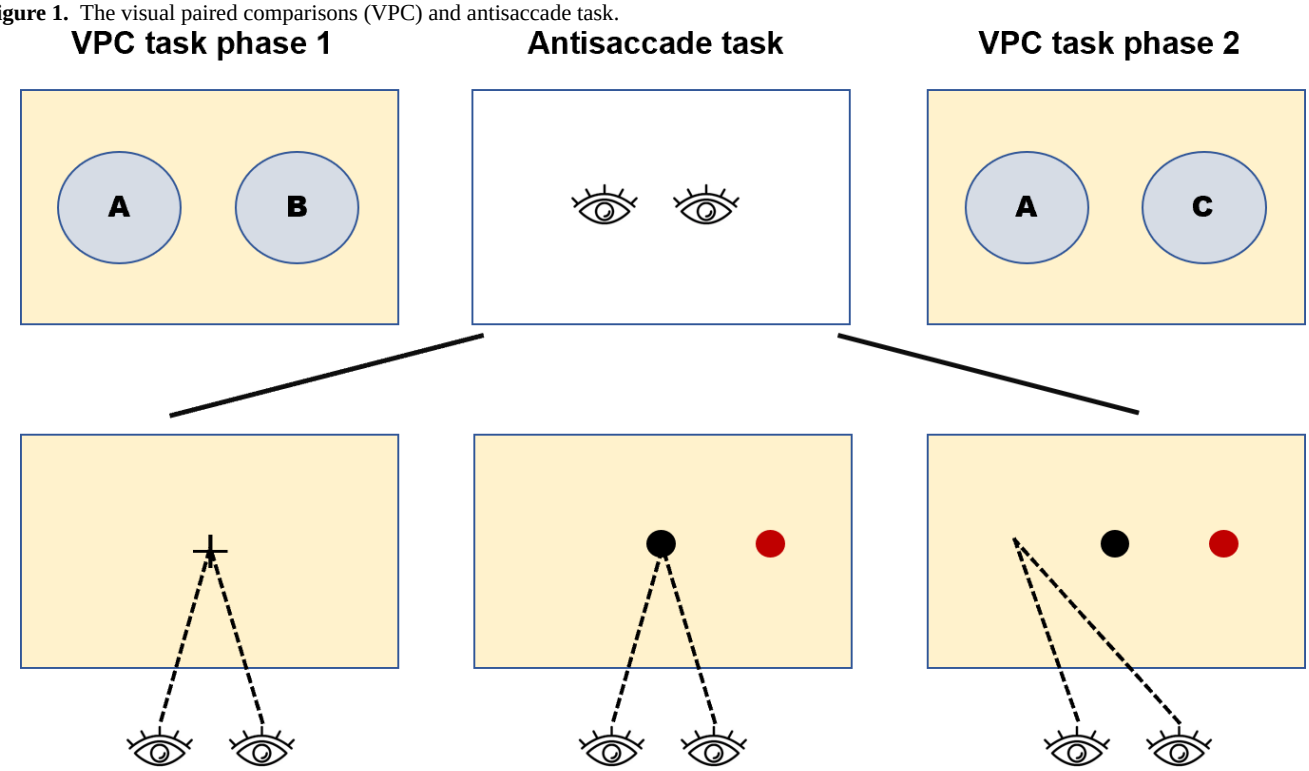
Table . Oculomotor parameters quantified during eye-tracking assessment.

Parameter	Definition	Unit/formula
Antisaccade task		
Latency	Time delay for saccade initiation after target onset	ms
Correct saccade latency	Latency of the first directionally accurate saccade	ms
Error rate	Percentage of initial saccades with incorrect direction	%
Peak velocity (horizontal)	Maximum horizontal velocity during saccades	Pixels/ms
Correction latency	Delay in generating corrective saccades after erroneous initial saccades	ms
Uncorrected error rate	Percentage of trials where corrective saccades remained directionally inaccurate	%
Corrected saccade latency	Latency of the first accurate saccade post correction	ms
Presaccadic fixation count	Number of fixation events before target onset	Count
Postsaccadic fixation count	Number of fixation events after target onset	Count
Fixation stability	Number of fixation clusters prior to target appearance	Count
Blinks during fixation	Fixation events contaminated by blinks	Count
Initial saccadic gain	Horizontal displacement amplitude of the first saccade	Pixels
Terminal saccadic gain ratio	Ratio of horizontal distance: Δ (first saccade end point → initial fixation)/Ref (final presaccade fixation → target)	Δ/Ref (unitless)
VPC[a] task		
Novelty preference ratio	Dwell time ratio (novel image/familiar image)	Ratio (unitless)
Total fixation duration	Summed duration of all fixation events	ms
Fixation events with blinks	Fixation epochs interrupted by blinks	Count

[a]VPC: visual paired comparisons.

XSL•FO
RenderX

Figure 1. The visual paired comparisons (VPC) and antisaccade task.



All tests were conducted in an inpatient setting, in a dedicated examination room to ensure a quiet and controlled environment. No other patients were present during testing. Patients remained seated during the test. The tablet was mounted on a stand and placed on a desk at a comfortable viewing distance. A single investigator who had undergone standardized protocol training administered the test. Before the test started, the investigator used a pre-prepared script to explain the procedure to the patient. Once the test began, the software provided voice guidance (on-screen text and narration) to instruct the patient through each step.

Head pose calibration required the patient to align their face within an on-screen virtual frame. The adjustment process typically took about 3 seconds; once the patient's head was correctly positioned, calibration was completed immediately. If the initial alignment failed, the investigator gently reminded the patient to stay still, and the patient readjusted within another few seconds. This process did not cause noticeable fatigue, even in patients with hand tremors. The complete testing procedure (including calibration, VPC, and antisaccade tasks) took approximately 6 minutes per participant, which was well tolerated.

Statistical Analysis

As the index test comprised multiple continuous oculomotor parameters, no single test positivity cutoff was predefined. Instead, a multivariable logistic regression model was constructed to combine these metrics with clinical variables for PSCI risk prediction. The model's performance was evaluated using AUC, sensitivity, and specificity at the optimal probability threshold derived from the receiver operating characteristic curve. All statistical analyses were performed using IBM SPSS 27.0 and R software (v4.2.2; R Foundation for Statistical Computing). The sample size was determined based on feasibility (all eligible patients enrolled during the study period); no formal sample size calculation was performed. Continuous variables with normal distribution were expressed as mean (SD) and compared using independent *t* tests or one-way ANOVA. Nonnormally distributed continuous variables were reported as median (IQR) and analyzed using Kruskal-Wallis *H* tests or Mann-Whitney *U* tests. Categorical variables were described as frequencies (%) and compared via chi-square tests or Fisher exact tests. Variables with $P<.05$ in univariate logistic regression analyses for PSCI were incorporated into multivariate binary logistic regression using stepwise selection to construct a PSCI risk prediction model. Model performance was validated internally. Statistical significance was defined as $P<.05$. Missing data on the 3-month follow-up were handled by complete case analysis; only patients with complete data were included in the final analysis.

Results

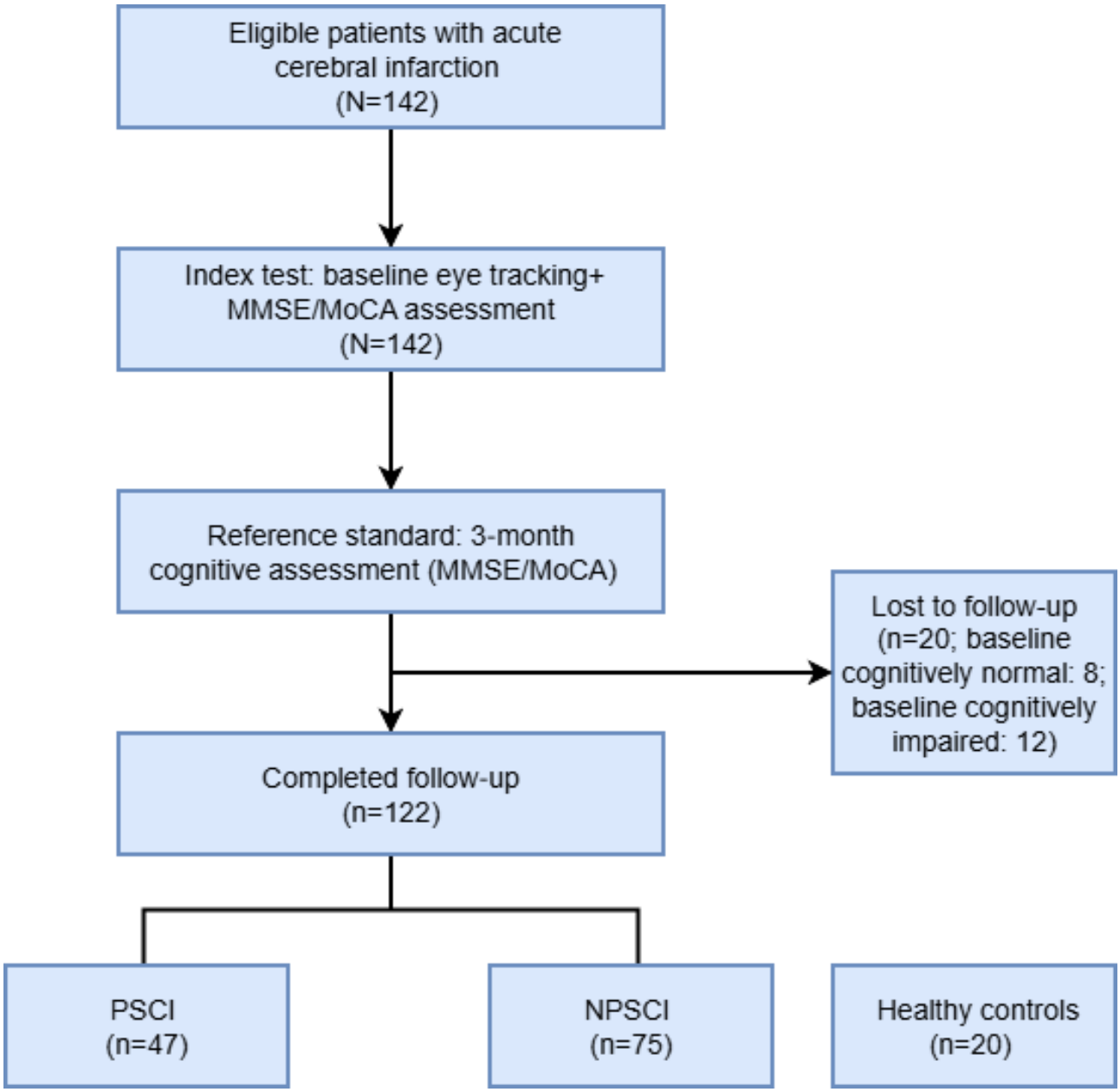
General Clinical Data Analysis of the Patients

This study enrolled 142 patients with acute cerebral infarction who met the inclusion/exclusion criteria, all of whom underwent baseline neuropsychological assessments during hospitalization. Of these, 122 completed 3-month follow-up evaluations, comprising the study cohort. PSCI developed in 47 patients (38.5% incidence), with 75 classified as NPSCI (Figure 2; see Checklist 1). Compared to patients with NPSCI, the PSCI group had a significantly higher proportion of females (18/47, 38% vs 14/75, 19%; $\chi^2_1=5.76$, $P=.02$), older age (mean 64.23, SD 8.95 vs mean 56.68, SD 10.42; $t_{120}=-4.11$, $P<.001$), lower education level (6.00, IQR 6.00 - 9.00 vs 9.00, IQR 6.00 - 12.00; $Z=-3.68$, $P<.001$), and elevated admission neurological severity scores (NIHSS [National Institutes of Health Stroke Scale]=3.00, IQR 2.00 - 6.00 vs 2.00, IQR

XSL•FO
RenderX

1.00 - 3.00; Z=−3.68, P<.001; mRS [Modified Rankin Scale]=2.00, IQR 1.00 - 4.00 vs 1.00, IQR 1.00 - 3.00; Z=−3.06, P=.002), whereas no significant differences existed in BMI, history of prior stroke, hypertension, diabetes, coronary artery disease, smoking, alcohol use, or anxiety/depression prevalence (Table 2).

Figure 2. STARD (Standards for Reporting Diagnostic Accuracy Studies) flow diagram. MMSE: Mini-Mental State Examination; MoCA: Montreal Cognitive Assessment; NPSCI: non–poststroke cognitive impairment; PSCI: poststroke cognitive impairment.



XSL•FO
RenderX

Table . Comparative analysis of demographic and clinical characteristics across study groups.

Variable	PSCI[a] (n=47)	NPSCI[b] (n=75)	HC[c] (n=20)	PSCI vs NPSCI		NPSCI vs HC	
				$t/Z/\chi^2$ (*df*)	*P* value	$t/Z/\chi^2$ (*df*)	*P* value
Age (y)	64.23 (8.95)[d]	56.68 (10.42)[d]	61.00 (44.75-62.75)[e]	−4.11 (120)[f]	<.001	−0.17	.87
Education level (y)	6.00 (6.00-9.00)[e]	9.00 (6.00-12.00)[e]	9.00 (6.75-12.00)[e]	−3.68	<.001	−0.44	.66
BMI (kg/m^2)	25.06 (23.31-27.12)[e]	25.96 (23.94-27.92)[e]	25.00 (2.76)[d]	−1.33	.18	1.52	.13
NIHSS[g]	3.00 (2.00-6.00)[e]	2.00 (1.00-3.00)[e]	—[h]	−3.68	<.001	—	—
mRS[i]	2.00 (1.00-4.00)[e]	1.00 (1.00-3.00)[e]	—	−3.06	.002	—	—
Baseline MMSE[j]	21.00 (17.00-25.00)[e]	27.00 (25.00-29.00)[e]	28.50 (28.00-30.00)[e]	−6.49	<.001	−1.63	.10
Baseline MoCA[k]	15.00 (12.00-19.00)[e]	24.00 (20.00-26.00)[e]	26.00 (26.00-27.00)[e]	−7.47	<.001	−2.45	.01
Follow-up MMSE	22.00 (17.00-24.00)[e]	28.00 (27.00-29.00)[e]	—	−8.40	<.001	—	—
Follow-up MoCA	16.00 (12.00-20.00)[e]	25.00 (23.00-27.00)[e]	—	−9.18	<.001	—	—
Sex, n (%)				5.76 (1)[l]	.02	10.75 (1)[l]	.001
Male	29 (62)	61 (81)	9 (45)	—	—	—	—
Female	18 (38)	14 (19)	11 (55)	—	—	—	—
History of stroke, n (%)	15 (32)	13 (17)	—	3.47 (1)[l]	.06	—	—
Hypertension, n (%)	32 (68)	49 (65)	9 (45)	0.10 (1)[l]	.75	2.75 (1)[l]	.10
Diabetes, n (%)	10 (21)	17 (23)	3 (15)	0.03 (1)[l]	.86	0.02 (1)[l]	.66
Coronary artery disease, n (%)	1 (2)	3 (4)	1 (5)	—	>.99	—	>.99
Smoking, n (%)	15 (32)	32 (43)	6 (30)	1.41 (1)[l]	.24	0.89 (1)[l]	.36
Alcohol use, n (%)	12 (26)	23 (31)	7 (35)	0.37 (1)[l]	.54	0.14 (1)[l]	.71
Anxiety, n (%)	7 (15)	6 (8)	0 (0)	1.44 (1)[l]	.23	—	.34
Depression, n (%)	7 (15)	5 (7)	0 (0)	1.38 (1)[l]	.24	—	.58

[a]PSCI: poststroke cognitive impairment.

[b]NPSCI: non–poststroke cognitive impairment.

[c]HC: healthy control.

[d]Continuous variables with normal distribution were expressed as mean (SD).

[e]Nonnormally distributed continuous variables were reported as median (IQR).

[f]For continuous variables, independent-samples *t* tests were used for normally distributed data.

[g]NIHSS: National Institutes of Health Stroke Scale.

[h]Not applicable.

[i]mRS: Modified Rankin Scale.

[j]MMSE: Mini-Mental State Examination.

[k]MoCA: Montreal Cognitive Assessment.

[l]For categorical variables, χ^2 tests were used to compare proportions between groups.

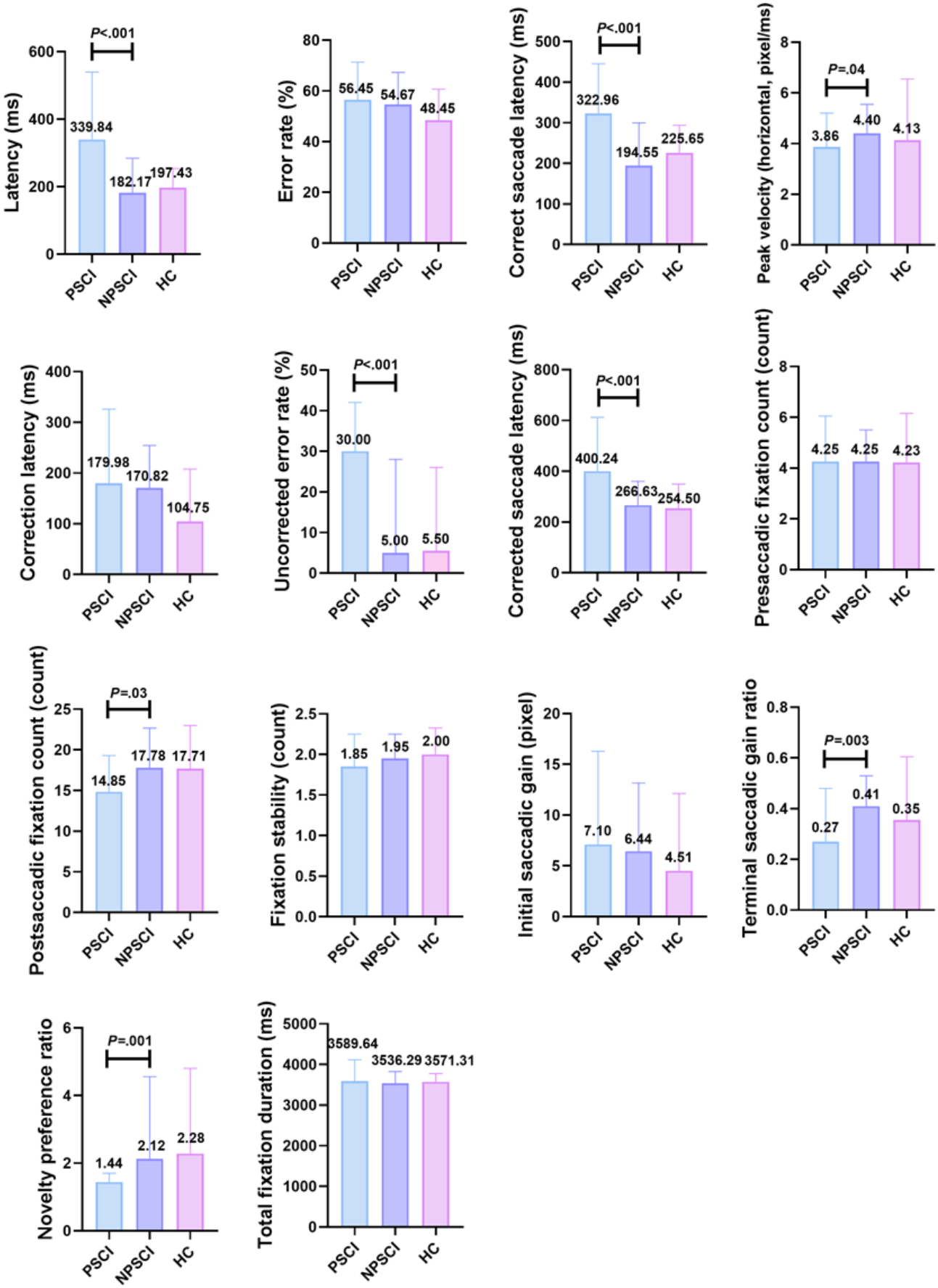
XSL•FO
RenderX

Analysis of Eye-Tracking Metrics

Analysis of eye-tracking metrics revealed that patients with PSCI exhibited markedly prolonged saccadic processing, demonstrating significantly higher saccade latency (Z=−4.77, P<.001), correct saccade latency (Z=−4.03, P<.001), uncorrected error rate (Z=−4.24, P<.001), and corrected saccade latency (Z=−4.62, P<.001) compared to NPSCI controls. Conversely, the PSCI group showed significantly reduced oculomotor efficiency, characterized by lower peak horizontal velocity (Z=−2.01, P=.04), diminished postsaccadic fixation count (Z=−2.20, P=.03), impaired terminal saccadic gain ratio (Z=−2.97, P=.003), and decreased novelty preference ratio (Z=−3.44, P=.001). However, fundamental visuomotor integration remained intact between groups, with no significant differences in error rate, correction latency, presaccadic fixation count, fixation stability, blinks during fixation, initial saccadic gain, total fixation duration, or VPC fixation events with blinks (all P>.05; Figure 3; see Multimedia Appendix 1 for exact P values). These findings suggest that executive-attentional deficits in PSCI selectively disrupt higher-order saccadic control and novelty processing, while preserving basic oculomotor functions.

XSL•FO
RenderX

Figure 3. Comparative analysis of quantitative oculomotor metrics across study cohorts. HC: healthy control; NPSCI: non–poststroke cognitive impairment; PSCI: poststroke cognitive impairment.



XSL•FO
RenderX

Comparative Analysis of Demographic and Eye-Tracking Metrics Between Healthy Controls and NPSCI Groups

To investigate whether oculomotor performance in patients with NPSCI aligns with neurologically intact individuals, we compared quantitative eye-tracking metrics between 20 age-matched healthy control (HC) and the NPSCI cohort. Comprehensive analysis demonstrated no statistically significant differences in any eye-tracking parameters (all $P>.05$; Figure 3; see Multimedia Appendix 2 for exact P values), including saccade latency, novelty preference ratio, and fixation stability metrics, confirming that oculomotor function in patients with NPSCI remains within normative ranges despite cerebrovascular pathology. These findings suggest preserved visuomotor network integrity in patients living with stroke without cognitive impairment.

Logistic Regression Analysis of Eye-Tracking Metrics in Patients With PSCI

Binary logistic regression analysis identified significant predictors of PSCI, with demographic factors including sex (male vs female, with female as the reference; OR 0.370, 95% CI 0.162 - 0.845; $P=.02$), advanced age (OR 1.084, 95% CI 1.038 - 1.132 per year; $P<.001$), and lower education level (OR 0.801, 95% CI 0.710 - 0.904 per year; $P<.001$), alongside clinical indicators of higher admission NIHSS (OR 1.463, 95% CI 1.202 - 1.781 per point; $P<.001$) and elevated mRS scores (OR 1.701, 95% CI 1.220 - 2.373 per point; $P=.002$). Critically, oculomotor metrics independently predicted PSCI risk: prolonged saccade latency (OR 1.005, 95% CI 1.003 - 1.007 per ms; $P<.001$), increased correct saccade latency (OR 1.004, 95% CI 1.002 - 1.006 per ms; $P=.001$), elevated uncorrected error rate (OR 1.043, 95% CI 1.021 - 1.066 per %; $P<.001$), extended corrected saccade latency (OR 1.005, 95% CI 1.003 - 1.007 per ms; $P<.001$), reduced postsaccadic fixation count (OR 0.949, 95% CI 0.903 - 0.998 per count; $P=.04$), and diminished novelty preference ratio (OR 0.686, 95% CI 0.532 - 0.884 per unit; $P=.004$). These findings demonstrate that oculomotor dysfunction significantly contributes to PSCI pathogenesis beyond traditional clinical factors (complete models in Table 3).

Table . Univariate logistic regression analysis.

Variable	*B* (SE)	Wald	*P* value	OR[a] (95% CI)
Sex (male vs female)[b]	−0.995 (0.422)	5.565	.02	0.370 (0.162-0.845)
Age (y)	0.081 (0.022)	13.432	<.001	1.084 (1.038-1.132)
Education level (y)	−0.222 (0.062)	12.953	<.001	0.801 (0.710-0.904)
NIHSS[c]	0.381 (0.100)	14.366	<.001	1.463 (1.202-1.781)
mRS[d]	0.531 (0.170)	9.799	.002	1.701 (1.220-2.373)
Latency (ms)	0.005 (0.001)	16.824	<.001	1.005 (1.003-1.007)
Correct saccade latency (ms)	0.004 (0.001)	11.192	.001	1.004 (1.002-1.006)
Uncorrected error rate (%)	0.042 (0.011)	15.076	<.001	1.043 (1.021-1.066)
Corrected saccade latency (ms)	0.005 (0.001)	16.014	<.001	1.005 (1.003-1.007)
Postsaccadic fixation count (count)	−0.052 (0.025)	4.222	.04	0.949 (0.903-0.998)
Novelty preference ratio	−0.377 (0.130)	8.464	.004	0.686 (0.532-0.884)

[a]OR: odds ratio.

[b]Sex was coded as male=1 and female=0; therefore, an OR<1 indicates a lower risk for males relative to females.

[c]NIHSS: National Institutes of Health Stroke Scale.

[d]mRS: Modified Rankin Scale.

Multivariate logistic regression analysis was performed after excluding variables with variance inflation factors greater than 5 to mitigate multicollinearity, retaining significant predictors from univariate screening. The final validated model demonstrated that advanced age (OR 1.067, 95% CI 1.009 - 1.129 per year; $P=.02$), lower education level (OR 0.841, 95% CI 0.708 - 0.999 per year; $P=.049$), higher NIHSS scores (OR 1.557, 95% CI 1.075 - 2.256 per point; $P=.02$), and prolonged correct saccade latency (OR 1.004, 95% CI 1.000 - 1.007 per ms; $P=.04$) independently predicted PSCI development. Clinically, each 1-year age increase elevates PSCI risk by 6.7% (OR 1.067), while each additional education year reduces risk by 15.9% ($1/0.841\approx1.189$). A 1-point NIHSS increase confers 55.7% higher PSCI probability, and critically, every 100-ms increase in correct saccade latency nearly doubles PSCI likelihood (OR 1.49, 95% CI 1.00 - 2.22, calculated as 1.004^{100}). These results establish eye-tracking metrics as novel predictors complementing traditional stroke severity indices (complete models in Table 4).

XSL•FO
RenderX

Table . Multivariate logistic regression analysis.

Variable	*B* (SE)	Wald	*P* value	OR[a] (95% CI)
Sex (male vs female)[b]	−0.383 (0.633)	0.366	.55	0.682 (0.197-2.357)
Age (y)	0.065 (0.029)	5.200	.02	1.067 (1.009-1.129)
Education level (y)	−0.173 (0.088)	3.867	.049	0.841 (0.708-0.999)
NIHSS[c]	0.443 (0.189)	5.482	.02	1.557 (1.075-2.256)
mRS[d]	−0.221 (0.349)	0.402	.53	0.801 (0.404-1.590)
Correct saccade latency (ms)	0.004 (0.002)	4.342	.04	1.004 (1.000-1.007)
Uncorrected error rate (%)	0.014 (0.015)	0.885	.35	1.014 (0.985-1.043)
Postsaccadic fixation count (count)	−0.023 (0.039)	0.357	.56	0.977 (0.905-1.054)
Novelty preference ratio	−0.023 (0.141)	0.027	.87	0.977 (0.741-1.289)
Constant	−4.451 (2.309)	3.717	.05	0.012

[a]OR: odds ratio.

[b]Sex was coded as male=1 and female=0; therefore, an OR<1 indicates a lower risk for males relative to females.

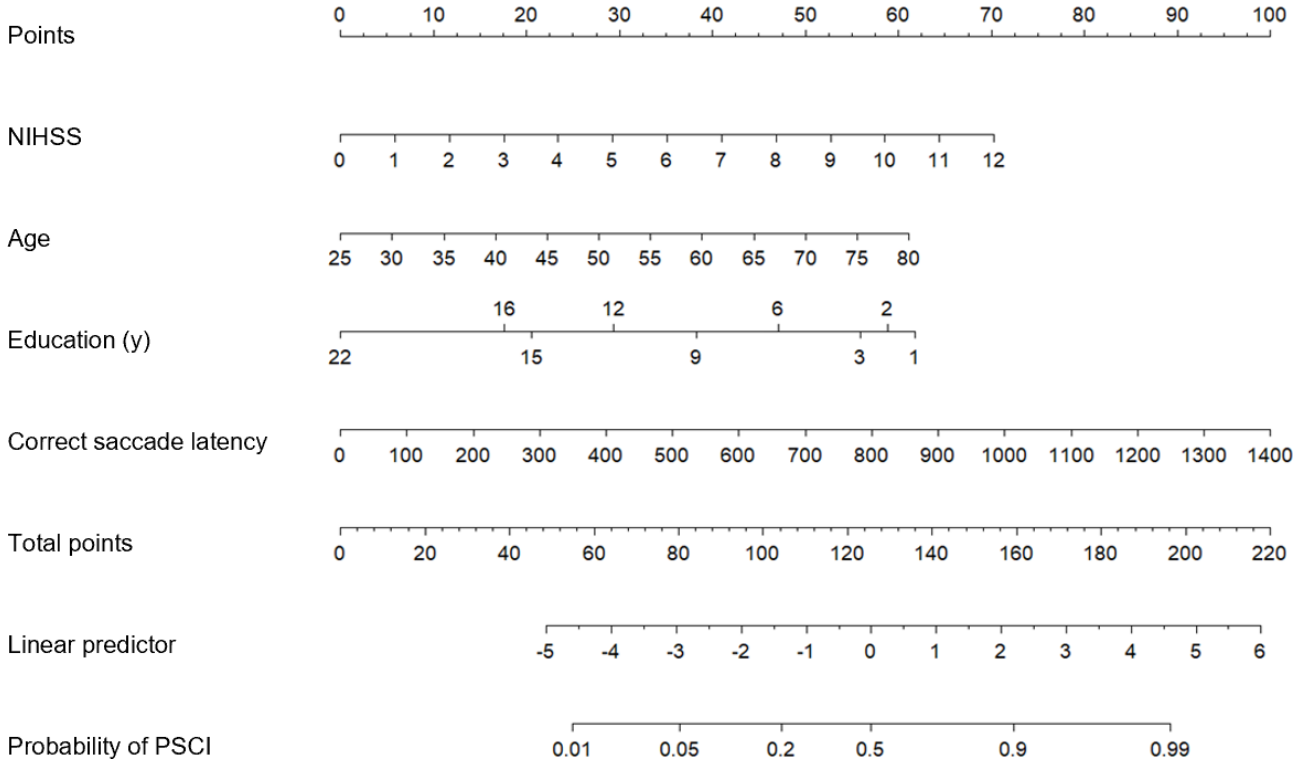
[c]NIHSS: National Institutes of Health Stroke Scale.

[d]mRS: Modified Rankin Scale.

Development of a Multivariable Predictive Model for PSCI Incorporating a Risk Nomogram

Based on the multivariate logistic regression analysis, advanced age, lower education level, higher NIHSS scores, and prolonged correct saccade latency were identified as independent predictors of PSCI in patients with acute ischemic stroke. These factors were integrated into a clinically actionable point-based nomogram (Figure 4) for individualized PSCI risk estimation at 3 months post stroke. The nomogram operates through three sequential steps: (1) variable scoring: assign points for each predictor (age, education, NIHSS, and latency) using the top axis scale. (2) Total points calculation: sum all individual scores to determine aggregate risk points. (3) Probability conversion: map total points to the bottom probability axis, directly indicating PSCI likelihood. For example, a 70-year-old (50 points) with 6 years of education (35 points), NIHSS=4 (23 points), and correct saccade latency=500 ms (35 points) would have 143 total points, corresponding to 70% PSCI probability. This visual tool enables rapid risk stratification at bedside without computational tools, demonstrating how oculomotor metrics enhance traditional clinical prediction.

Figure 4. Nomogram model for predicting poststroke cognitive impairment (PSCI) risk. NIHSS: National Institutes of Health Stroke Scale.

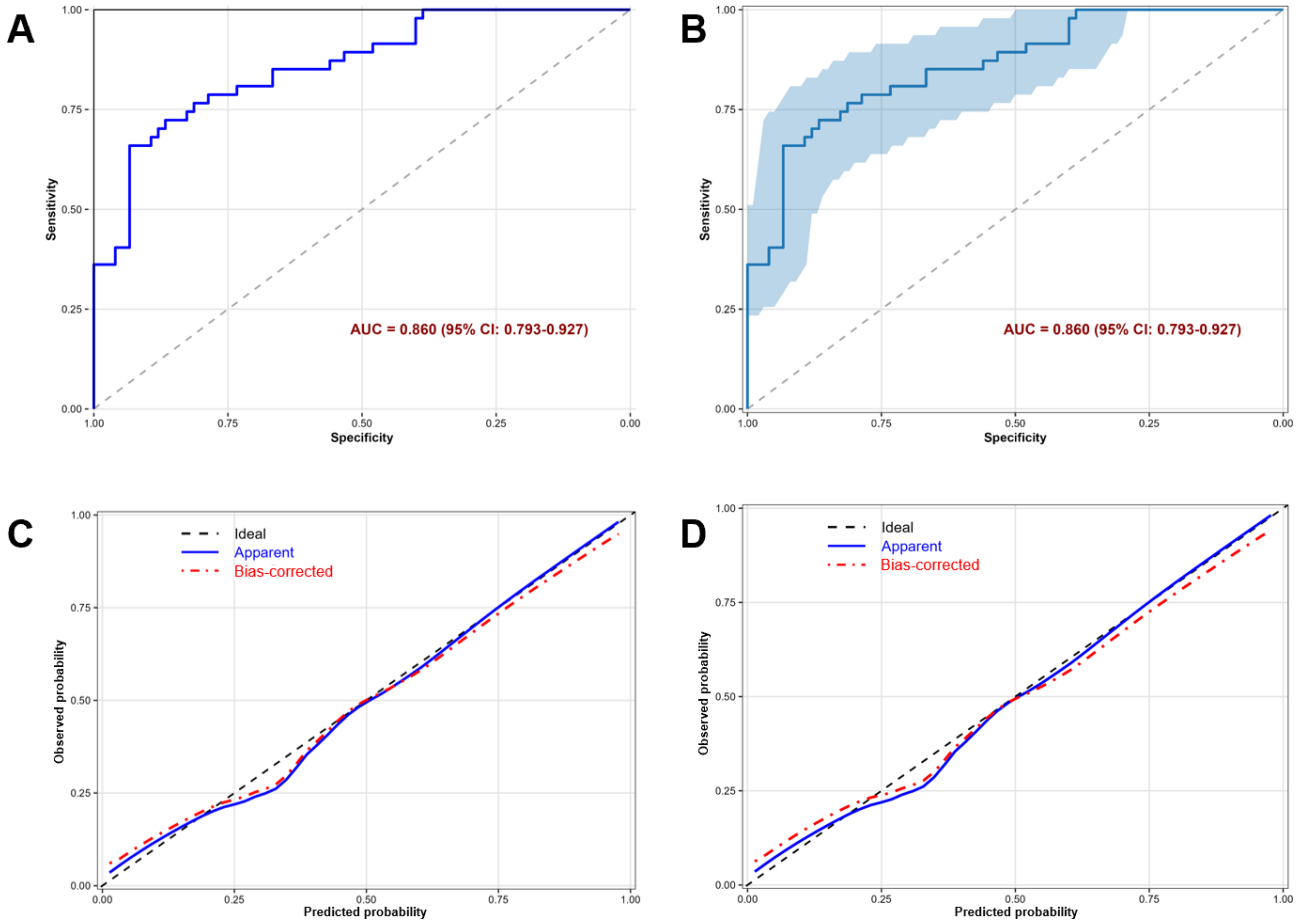


XSL•FO
RenderX

Validation of the PSCI Risk Nomogram: Discriminatory Accuracy and Calibration Performance

The discriminative ability of the PSCI risk prediction model was evaluated using the concordance index (C-index), which is mathematically equivalent to the AUC. The model achieved a C-index of 0.860 (95% CI 0.793 - 0.927, sensitivity=93.3%, and specificity=66.0%; Figure 5A). Internal validation with 1000 bootstrap resamples yielded a calibrated C-index of 0.860. Both values significantly exceeded the reference threshold of 0.75, indicating robust discriminative performance (Figure 5B).

Figure 5. Performance and validation of the nomogram model. (A) The model achieved a concordance index (C-index) of 0.860 (95% CI 0.793 - 0.927). (B) Internal validation with 1000 bootstrap resamples yielded a calibrated C-index of 0.860. (C) Calibration was assessed via a calibration plot comprising 3 characteristic curves: the apparent curve representing raw model predictions, the bias-corrected curve reflecting bias-adjusted predictions, and the ideal curve denoting perfect prediction; both the raw and bias-corrected curves demonstrated close alignment with the ideal trajectory, confirming satisfactory model calibration. (D) Calibration curve with internal validation. AUC: area under the receiver operating characteristic curve.



Calibration was assessed via a calibration plot comprising 3 characteristic curves: the apparent curve representing raw model predictions, the bias-corrected curve reflecting bias-adjusted predictions, and the ideal curve denoting perfect prediction. Both the raw and bias-corrected curves demonstrated close alignment with the ideal trajectory, confirming satisfactory model calibration (Figure 5C). The apparent performance line, bias-corrected performance line, and ideal line also show close alignment during internal model validation using bootstrap resampling (Figure 5D).

Comparative Analysis of Neuroimaging Metrics in PSCI vs NPSCI Cohorts

The analysis of admission brain MRI scans revealed a significantly higher prevalence of frontal lobe infarcts (16/47, 34% vs 11/75, 15%; χ^2_1=6.29, *P*=.01), temporal lobe involvement (13/47, 28% vs 6/75, 8%; χ^2_1=8.49, *P*=.004), larger infarct volumes (χ^2_2=11.27, *P*=.004), and more severe white matter hyperintensities (WMHs; Fazekas grade ≥2: 22/47, 47% vs 15/75, 20%; χ^2_2=10.00, *P*=.007) in patients with PSCI vs NPSCI (Table 5). No significant intergroup differences (*P*>.05) were observed in anterior/posterior circulation distribution or moderate-to-severe large vessel stenosis/occlusion (see Table 5 for exact *P* values). These findings implicate frontal-temporal lesions and white matter disruption as neuroanatomical substrates for PSCI.

XSL•FO
RenderX

Table . Comparative analysis of neuroimaging parameters between poststroke cognitive impairment (PSCI) and non–poststroke cognitive impairment (NPSCI) cohorts.

Variable	PSCI (n=47), n (%)	NPSCI (n=75), n (%)	χ^2 (*df*)	*P* value
Stroke infarct location				
Frontal lobe	16 (34)	11 (15)	6.29 (1)	.01
Parietal lobe	9 (19)	6 (8)	3.33 (1)	.07
Temporal lobe	13 (28)	6 (8)	8.49 (1)	.004
Occipital lobe	8 (17)	9 (12)	0.61 (1)	.44
Internal capsule	2 (4)	3 (4)	—[a]	>.99
Basal ganglia	24 (51)	37 (49)	0.04 (1)	.85
Diencephalon	4 (9)	12 (16)	1.42 (1)	.23
Brainstem	11 (23)	16 (21)	0.07 (1)	.79
Cerebellum	1 (2)	3 (4)	—	>.99
Anterior/posterior circulation				
Anterior circulation	34 (72)	44 (59)	2.34 (1)	.13
Posterior circulation	16 (34)	31 (41)	0.65 (1)	.42
Stroke infarct size	—	—	11.27 (2)	.004
Lacunar infarction (<1.5 cm)	25 (53)	58 (78)	—	—
Small subcortical infarction (1.5-3 cm)	9 (19)	12 (16)	—	—
Large territorial infarction (>3 cm)	13 (28)	5 (7)	—	—
Vascular stenosis (≥50%)	17 (36)	18 (24)	2.09 (1)	.15
WMH[b]	—	—	10.00 (2)	.007
Grade 0 (absent)	13 (28)	28 (37)	—	—
Grade I (punctate foci)	12 (26)	32 (43)	—	—
Grade≥II (early confluent or confluent foci)	22 (47)	15 (20)	—	—

[a]Not applicable.

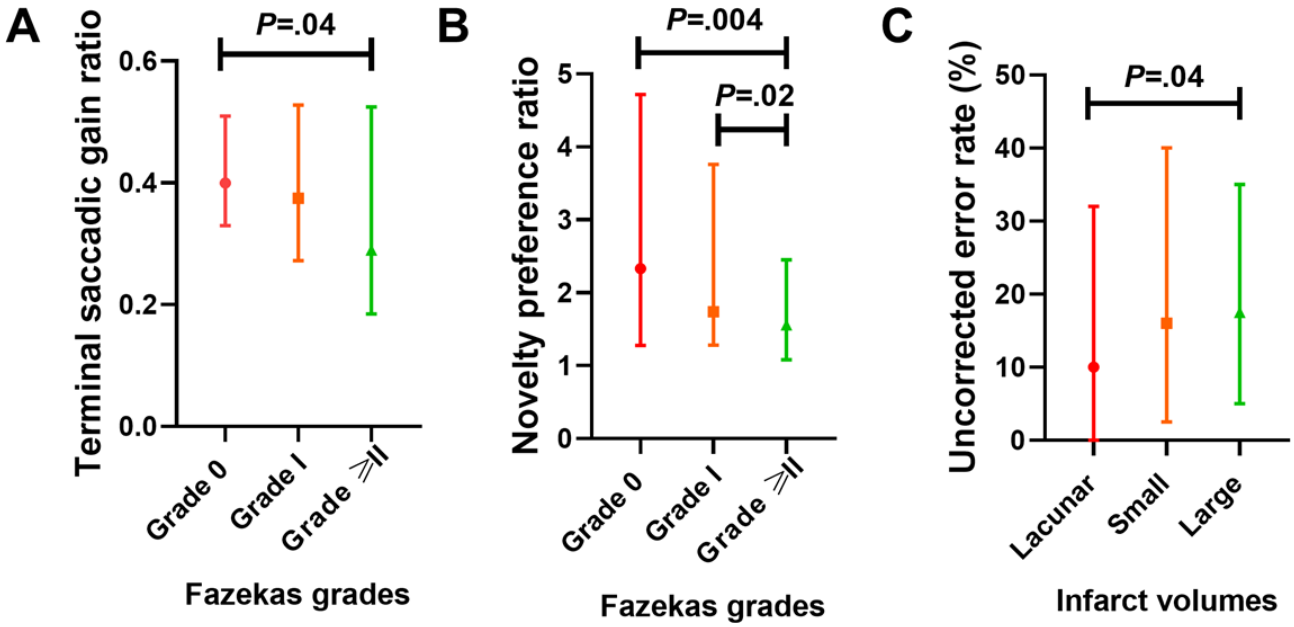
[b]WMH: white matter hyperintensity.

Quantitative Correlation Analysis Between WMHs and Eye-Tracking Parameters

These stratified analyses were exploratory and not prespecified. Stratified by Fazekas grades on admission brain MRI, the 122 patients with cerebral infarction were categorized into grade 0 (n=41), grade I (n=44), and grade≥II (n=37) cohorts. Significant intergroup differences emerged in terminal saccadic gain ratio (H=6.35, P=.04) and novelty preference ratio (H=11.78, P=.003), while other oculomotor metrics showed no statistical significance (all P>.05; see Multimedia Appendix 3 for exact P values). Post hoc analyses revealed that for terminal saccadic gain ratio, grade≥II demonstrated a 33% reduction vs grade 0 (0.27, IQR 0.17-0.49 vs 0.40 IQR 0.33-0.51; Z=19.94, P=.04; Figure 6A), but no differences between grade≥II vs I (Z=13.20, P=.28) or grade I vs 0 (Z=6.74, P>.99). For novelty preference ratio, grade≥II showed 39% and 18% reductions vs grade 0 (1.43, IQR 1.00-1.63 vs 2.33, IQR 1.27-4.71; Z=25.57, P=.004; Figure 6B) and grade I (1.43, IQR 1.00-1.63 vs 1.74, IQR 1.28-3.76; Z=21.88, P=.02), respectively, with no grade 0-I difference (Z=3.69, P>.99). This dose-dependent impairment pattern indicates that severe WMHs (Fazekas≥II) selectively disrupt visuospatial calibration and novelty detection, implicating compromised comprehension and executive function and significant memory decline.

XSL•FO
RenderX

Figure 6. Stratified comparative analysis of eye-tracking metrics by white matter hyperintensity (WMH) severity and infarct volume strata. (A) Comparison of terminal saccadic gain ratio across different WMH severity grades. (B) Comparison of novelty preference ratio across WMH severity grades. (C) Comparison of uncorrected error rate across stroke infarct volume strata.



Quantitative Correlation Analysis Between Infarct Volumes and Eye-Tracking Parameters

These stratified analyses were exploratory and not prespecified. Stratified by cerebral infarct volume on admission MRI, the 122 patients were categorized into lacunar infarction (<1.5 cm, n=83), small infarction (1.5 - 3 cm, n=21), and large infarction (>3 cm, n=18) cohorts. Statistical analysis revealed significant intergroup differences exclusively in uncorrected saccadic error rate (H=6.22, P=.045; Figure 6C), while other oculomotor metrics showed no significance (all P>.05; see Multimedia Appendix 4 for exact P values). Post hoc comparisons demonstrated patients with large infarction exhibited 18.0% higher uncorrected error rates vs lacunar infarction (28.00, IQR 15.00-44.00 vs 10.00, IQR 0.00-32.00; Z=−22.21, P=.04), but there were no significant differences between large vs small infarction (Z=−13.90, P=.65) or small vs lacunar infarction (Z=−8.30, P>.99). The dose-dependent trend (large>small>lacunar) further suggests that infarct volume thresholds may dictate cognitive-motor integration failure.

Discussion

Principal Findings

PSCI represents a prevalent disabling sequela of stroke, characterized by an insidious onset that frequently delays diagnosis until 3 to 6 months post event. This diagnostic window highlights the urgent need for early detection tools, as untreated PSCI leads to irreversible cognitive decline, severely impacting quality of life and imposing substantial health care system burdens [34]. Current diagnostic limitations such as neuroimaging costs, psychometric scale inaccuracies in patients with aphasic/hemiparetic, and delayed assessment prompted our development of an accessible eye-tracking predictive model.

Among 122 patients with acute ischemic stroke, we observed a 3-month PSCI incidence of 38.5% (47/122), consistent with established epidemiological data [35]. Multivariable regression analysis identified key independent predictors of PSCI: demographic factors, including female sex, advanced age, and lower education level; clinical indicators of higher admission NIHSS score and elevated mRS score; and the oculomotor biomarker of prolonged correct saccade latency. Significantly, eye-tracking biomarkers differentiated PSCI from patients with NPSCI, demonstrating increased saccade latencies and error rates alongside decreased peak velocity, terminal saccadic gain, and novelty preference ratio, while patients with NPSCI exhibited normative oculomotor profiles indistinguishable from HCs. This specificity supports the validity of our integrated nomogram as a rapid, cost-effective tool for acute-phase PSCI risk stratification, facilitating timely interventions during the period of peak neuroplasticity.

Prediction Model for PSCI Combining AI Eye Tracking and Clinical Demographics

Advanced age and lower educational attainment represent nonmodifiable risk factors for PSCI, consistent with established epidemiological evidence [36,37]. A multicenter cross-sectional study demonstrated escalating PSCI incidence with age, demonstrating 1.887-fold elevated risk in patients ≥75 years compared to younger cohorts [1]. Furthermore, PSCI prevalence exhibits a socioeconomic gradient, disproportionately affecting individuals with lower education and income. The neurobiological mechanisms of education-cognition relationships are explained by 2 conceptual frameworks: the cognitive reserve hypothesis posits that advanced education enhances neural resilience against age-related decline, while the compensatory plasticity model suggests that higher education enables functional reorganization through preserved cognitive domains compensating for impaired functions [38].

NIHSS and mRS scores, which are validated measures of stroke severity, constitute robust predictors of long-term functional outcomes. Higher scores correlate with increased disability and

XSL•FO
RenderX

mortality and, specifically, predict the development of PSCI (adjusted OR 1.56 per NIHSS point increase in our cohort) [39]. Pathophysiologically, elevated NIHSS scores reflect more extensive neuroanatomical damage, heightening vulnerability to cognitive domain impairment.

Although female patients showed higher PSCI prevalence in our cohort (18/47, 38% vs 14/75, 19% in NPSCI), it was not an independent predictor in multivariate modeling. This observation aligns with reports of poorer poststroke functional outcomes in women, including greater activity limitations, reduced quality of life, and higher depression rates [40,41]. Notably, while PSCI incidence is compared between sexes, cognitive phenotypes differ: female patients demonstrate predominant deficits in attention, executive function, and language domains, while male patients display greater impairment in verbal memory [40]. Screening tool performance also varies by sex, with MMSE showing higher sensitivity but lower specificity for PSCI detection in women and MoCA exhibiting sex-dependent psychometric properties. These findings necessitate sex-stratified interpretation of cognitive screening instruments to reduce diagnostic inaccuracies.

Strong neuroanatomical connections exist between higher-order cognition and core oculomotor systems. Abstract cognitive tasks engage oculomotor circuits in primates, as manifested through cognitively modulated microsaccades [42]. Observed disparities in eye-tracking metrics between PSCI and NPSCI cohorts reflect multidimensional cognitive dysfunction encompassing attentional control, memory consolidation, and executive processing. Within the antisaccade paradigm, saccadic latency, the interval between stimulus onset and saccade initiation, reflects premotor decision-making processes. Primate neurophysiological studies confirm that saccade programming activates decision-related neurons in frontal eye fields and superior colliculi, wherein saccadic accuracy correlates with neuronal discharge patterns in oculomotor-integration zones [43-45]. Parameters, including latency, amplitude, peak velocity, and positional error, provide quantifiable proxies for decision accuracy and temporal processing [45].

Compared to prosaccades, antisaccades require heightened attentional and executive resources. This task recruits the dorsolateral prefrontal cortex, anterior cingulate cortex, supplementary eye field, and parietal eye field, with increased frontal activation correlating with enhanced antisaccade performance [28]. In our cohort, prolonged saccadic latency in patients with PSCI indicates impaired executive-attentional network function, consistent with frontoparietal dysfunction delaying stimulus-response integration [12,46,47]. Crucially, corrected saccade latency, requiring directional error correction, was identified as an independent PSCI predictor, indicating sensitivity to disrupted feedback monitoring in prefrontal-basal ganglia circuits [48,49]. During the VPC task, reduced novelty preference ratio in patients with PSCI indicated hippocampal-dependent memory impairment. Patients with NPSCI exhibited preserved familiarity discrimination, whereas patients with PSCI allocated equivalent dwell times to novel and familiar images, suggesting declarative memory consolidation failure. Notably, patients with NPSCI demonstrated normative eye-tracking profiles indistinguishable from HCs, confirming stroke-specific cognitive pathology rather than global oculomotor disruption. These oculomotor signatures provide direct windows into disrupted cortical-subcortical circuitry in PSCI.

Current evidence supports the diagnostic utility of eye tracking across neurodegenerative disorders: AD (AUC=0.91) [12], Parkinson dementia, and schizophrenia. Our integrated PSCI prediction model demonstrated comparable discriminative capacity (AUC=0.86, sensitivity=93.3%, and specificity=66.0%). Independent predictors included advanced age, lower education, higher NIHSS scores, and prolonged correct saccade latency, establishing this framework as a rapid biomarker-based stratification tool.

Correlations Between Neuroimaging Markers and Quantitative Eye-Tracking Metrics

These behavioral patterns implicate underlying neural substrates, which were validated through structural neuroimaging analyses. Neuroimaging analysis confirmed significant associations between PSCI and frontal/temporal lobe infarcts, larger infarct volumes, and severe WMH (Fazekas grade ≥ II). Frontal lesions disrupt attention-executive networks [50], while temporal infarcts impair hippocampal memory consolidation and language processing [35,51]. Moderate-to-severe WMH are established risk factors for PSCI [52,53], likely through the disruption of neuronal networks subserving cognitive reserve. Notably, left anterior thalamic radiation WMH specifically correlates with attention/executive dysfunction, whereas occipital callosal radiation lesions predominantly affect processing speed [54]. Larger infarct volumes predict greater cognitive vulnerability through extensive neocortical involvement. Thus, infarction location/size and WMH severity collectively contribute to cognitive decline.

These oculomotor biomarkers reflect underlying neuroimaging pathology. Significant oculomotor correlates were identified: WMH severity is inversely associated with terminal saccadic gain and novelty preference ratios, while infarct volume is positively correlated with uncorrected saccadic error rates (significantly higher in large vs lacunar infarcts). Mechanistically, reduced terminal gain indicates visuospatial-executive impairment; diminished novelty preference reflects declarative memory deficits; and elevated uncorrected errors signify attentional monitoring dysfunction [45]. These findings establish oculomotor metrics as structural pathology surrogates. Reduced terminal gain/novelty preference indicates severe WMH, while elevated uncorrected errors associate with large infarct volume. For patients contraindicated for neuroimaging, eye tracking provides a clinically feasible alternative for domain-specific cognitive assessment.

Implementing accessible methodologies for early cognitive impairment prediction is paramount in stroke care. Eye-tracking technology offers distinct clinical advantages by integrating audiovisual task instructions, including on-screen text, subtitles, and voice-over narration, to ensure accessibility across varying levels of literacy and physical ability. By relying on gaze-based responses, the method eliminates the need for verbal or manual interaction, making it particularly suitable for patients with aphasia or dominant-hand paresis. As a mobile screening tool,

XSL•FO
RenderX

it enables efficient cognitive phenotyping through quantitative oculomotor biomarkers, delivering actionable clinical insights promptly.

Furthermore, integrated within mHealth platforms, tablet-based eye tracking supports scalable, low-cost screening beyond traditional clinical settings. This approach facilitates real-time data acquisition and seamless telehealth integration, promoting equitable access to follow-up care—especially for individuals in remote or underserved regions. Beyond accelerating clinical decision-making, this mHealth-enabled paradigm allows decentralized longitudinal monitoring of disease progression and treatment response, including remotely administered home assessments.

This study pioneers the multidisciplinary integration of eye-tracking analytics with clinical demographics to develop a predictive model for PSCI. We established neurobiological correlations between oculomotor signatures and structural neuroimaging markers, thereby expanding the repertoire of neurobehavioral biomarkers while offering a cost-effective framework for stratifying post-stroke cognitive risk.

Limitations

Several limitations of this study should be acknowledged. First, the sample size was relatively modest (122 patients with complete follow-up, along with 20 HCs) and derived from a single center, which may limit statistical power and generalizability of the findings. Future multicenter studies with larger cohorts are needed to validate and refine our predictive model. Second, the follow-up period was limited to 3 months, which may not capture later-onset cognitive decline or the long-term trajectory of PSCI. Extended follow-up (eg, 6 or 12 mo) would provide a more comprehensive understanding of the evolution of cognitive impairment and the stability of eye-tracking biomarkers. Third, as an observational diagnostic study with a relatively small sample size, our findings should be considered exploratory rather than confirmatory. Prospective randomized controlled trials with adequate sample size and appropriate blinding are required to further validate the diagnostic sensitivity and specificity of this tablet-based eye-tracking tool, particularly if it is to be considered as a software-as-a-medical-device for regulatory or clinical registration purposes. Fourth, the oculomotor battery was restricted to VPC and antisaccade tasks. Although these tasks are well-established for assessing executive function and memory, incorporating additional paradigms such as smooth pursuit, free viewing, or memory-guided saccades could offer a more complete assessment of the spectrum of cognitive deficits in PSCI. Fifth, our study focused exclusively on acute cerebral infarction; therefore, the findings may not directly generalize to other stroke subtypes (eg, hemorrhagic stroke) or to diverse ethnic and demographic populations. Validation across different stroke types and settings is warranted. Sixth, while we correlated eye-tracking metrics with WMH severity and infarct volume, we did not integrate more advanced neuroimaging biomarkers (eg, diffusion tensor imaging or resting-state functional MRI). Future studies linking oculomotor impairments to structural and functional connectivity measures would further strengthen the neurobiological basis of our findings. Seventh, the study employed a purposive (nonprobability) sampling strategy, which may introduce selection bias and limit the representativeness of the sample. Eighth, there was heterogeneity in acute stroke treatments among participants (eg, intravenous thrombolysis and endovascular thrombectomy), which could influence cognitive outcomes; although we adjusted for baseline NIHSS and mRS scores, residual confounding cannot be excluded.

Conclusions

Age, educational level, admission NIHSS score, and correct saccade latency were identified as independent predictors of 3-month PSCI and were incorporated into a nomogram model. The model demonstrated robust performance, achieving high predictive accuracy (AUC=0.86), strong discriminative capacity, and clinically useful interpretability. This tool demonstrates potential for early identification of patients with high-risk PSCI, thereby enabling timely and personalized interventions. Its integration into standard clinical pathways—particularly through mHealth platforms—may significantly enhance strategic decision-making and improve long-term outcomes for vulnerable populations.

Acknowledgments

The authors sincerely thank Shanghai Neuroweave Technology Co, Ltd, for the provision of eye-tracking instrumentation and Huashan Hospital, Fudan University, for institutional support of this investigation. The authors declare the use of generative AI (GAI) in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GAI tools under full human supervision: proofreading and editing. The GAI tool used was DeepSeek-V3.2. Responsibility for the final manuscript lies entirely with the authors. GAI tools are not listed as authors and do not bear responsibility for the final outcomes. Declaration submitted by: collective responsibility.

Funding

This work was partially supported by the Health Commission of Hebei Province's Tracking Project (number C20260099), the Natural Science Foundation of Hebei Province (number C2020206056), and the National Natural Science Foundation of China (number 82271504). The Government of Hebei Province has funded the project for the cultivation of outstanding clinical medical talents (number ZF2024061) and Innovation Research Team of Hebei Medical University (number 2023LCTD-A2). The funders had no role in study design, data collection, analysis, interpretation, or manuscript preparation.

XSL•FO
RenderX

Data Availability

The datasets generated and analyzed during this study are not publicly available because they are still being used for ongoing follow-up studies. However, these datasets, along with the full study protocol, are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: BX, JZ
Data curation: BX, RR, YZ
Formal analysis: RR
Supervision: JZ
Writing – original draft: BX
Writing – review and editing: XW, JZ
All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Comparative analysis of eye-tracking metrics between poststroke cognitive impairment and non–poststroke cognitive impairment cohorts.
[DOCX File, 20 KB - mhealth_v14i1e83194_app1.docx]

Multimedia Appendix 2
Comparative analysis of eye-tracking metrics between healthy control and non–poststroke cognitive impairment cohorts.
[DOCX File, 20 KB - mhealth_v14i1e83194_app2.docx]

Multimedia Appendix 3
Analysis of eye-tracking metrics by white matter hyperintensity severity grade.
[DOCX File, 23 KB - mhealth_v14i1e83194_app3.docx]

Multimedia Appendix 4
Analysis of eye-tracking metrics by infarct volume strata.
[DOCX File, 22 KB - mhealth_v14i1e83194_app4.docx]

Checklist 1
STARD-2015 checklist.
[DOCX File, 42 KB - mhealth_v14i1e83194_app5.docx]

References

1. He A, Wang Z, Wu X, et al. Incidence of post-stroke cognitive impairment in patients with first-ever ischemic stroke: a multicenter cross-sectional study in China. Lancet Reg Health West Pac 2023;33:100687. [doi: 10.1016/j.lanwpc.2023.100687] [Medline: 37181529]
2. Dowling NM, Johnson S, Nadareishvili Z. Poststroke cognitive impairment and the risk of recurrent stroke and mortality: systematic review and meta-analysis. J Am Heart Assoc 2024 Sep 17;13(18):e033807. [doi: 10.1161/JAHA.123.033807] [Medline: 39239841]
3. Rost NS, Brodtmann A, Pase MP, et al. Post-stroke cognitive impairment and dementia. Circ Res 2022 Apr 15;130(8):1252-1271. [doi: 10.1161/CIRCRESAHA.122.319951] [Medline: 35420911]
4. Gallucci L, Sperber C, Guggisberg AG, et al. Post-stroke cognitive impairment remains highly prevalent and disabling despite state-of-the-art stroke treatment. Int J Stroke 2024 Oct;19(8):888-897. [doi: 10.1177/17474930241238637] [Medline: 38425239]
5. Chi X, Fan X, Fu G, Liu Y, Zhang Y, Shen W. Research trends and hotspots of post-stroke cognitive impairment: a bibliometric analysis. Front Pharmacol 2023;14:1184830. [doi: 10.3389/fphar.2023.1184830] [Medline: 37324494]
6. Weaver NA, Kuijf HJ, Aben HP, et al. Strategic infarct locations for post-stroke cognitive impairment: a pooled analysis of individual patient data from 12 acute ischaemic stroke cohorts. Lancet Neurol 2021 Jun;20(6):448-459. [doi: 10.1016/S1474-4422(21)00060-0] [Medline: 33901427]

XSL•FO
RenderX

7. Ashburner JM, Chang Y, Porneala B, et al. Predicting post-stroke cognitive impairment using electronic health record data. Int J Stroke 2024 Oct;19(8):898-906. [doi: 10.1177/17474930241246156] [Medline: 38546170]
8. Ramos-Henderson M, Calderón C, Domic-Siede M. Education bias in typical brief cognitive tests used for the detection of dementia in elderly population with low educational level: a critical review. Appl Neuropsychol Adult 2025;32(1):253-261. [doi: 10.1080/23279095.2022.2155521] [Medline: 36519252]
9. Grossman JT, Frumkin MR, Rodebaugh TL, Lenze EJ. mHealth assessment and intervention of depression and anxiety in older adults. Harv Rev Psychiatry 2020;28(3):203-214. [doi: 10.1097/HRP.0000000000000255] [Medline: 32310834]
10. Gu B, Kim HS, Kim H, Yoo JI. Advancements in wearable sensor technologies for health monitoring in terms of clinical applications, rehabilitation, and disease risk assessment: systematic review. JMIR mHealth uHealth 2026 Jan 9;14:e76084. [doi: 10.2196/76084] [Medline: 41511829]
11. Noda K, Kageyama I, Kobayashi Y, Lim Y, Sengoku S, Kodama K. Leveraging mHealth wearables for managing patients with Alzheimer's disease: a scoping review. Drug Discov Today 2025 May;30(5):104363. [doi: 10.1016/j.drudis.2025.104363] [Medline: 40250750]
12. Li Q, Yan J, Ye J, et al. Construction of a prediction model for Alzheimer's disease using an AI-driven eye-tracking task on mobile devices. Aging Clin Exp Res 2024 Dec 27;37(1):9. [doi: 10.1007/s40520-024-02882-9] [Medline: 39725797]
13. Wei M, Li J, You T, et al. An accessible and efficient mobile eye-tracking application for community-based cognitive impairment screening in China. Alzheimers Res Ther 2025 Nov 17;17(1):250. [doi: 10.1186/s13195-025-01884-7] [Medline: 41250208]
14. Trinh S, Skoll D, Saxon LA. Health care 2025: how consumer-facing devices change health management and delivery. J Med Internet Res 2025 Apr 23;27:e60766. [doi: 10.2196/60766] [Medline: 40267475]
15. Skaramagkas V, Giannakakis G, Ktistakis E, et al. Review of eye tracking metrics involved in emotional and cognitive processes. IEEE Rev Biomed Eng 2023;16:260-277. [doi: 10.1109/RBME.2021.3066072] [Medline: 33729950]
16. Riek HC, Brien DC, Coe BC, et al. Cognitive correlates of antisaccade behaviour across multiple neurodegenerative diseases. Brain Commun 2023;5(2):fcad049. [doi: 10.1093/braincomms/fcad049] [Medline: 36970045]
17. Shaikh AG, Zee DS. Eye movement research in the twenty-first century-a window to the brain, mind, and more. Cerebellum 2018 Jun;17(3):252-258. [doi: 10.1007/s12311-017-0910-5] [Medline: 29260439]
18. Eraslan Boz H, Koçoğlu K, Akkoyun M, Tüfekci IY, Ekin M, Akdal G. Visual search in Alzheimer's disease and amnestic mild cognitive impairment: an eye-tracking study. Alzheimers Dement 2024 Feb;20(2):759-768. [doi: 10.1002/alz.13478] [Medline: 37774122]
19. Opwonya J, Wang C, Jang KM, Lee K, Kim JI, Kim JU. Inhibitory control of saccadic eye movements and cognitive impairment in mild cognitive impairment. Front Aging Neurosci 2022;14:871432. [doi: 10.3389/fnagi.2022.871432] [Medline: 35478701]
20. Lin J, Xu T, Yang X, et al. A detection model of cognitive impairment via the integrated gait and eye movement analysis from a large Chinese community cohort. Alzheimers Dement 2024 Feb;20(2):1089-1101. [doi: 10.1002/alz.13517] [Medline: 37876113]
21. Fang Z, Dang Y, Ping A, et al. Human high-order thalamic nuclei gate conscious perception through the thalamofrontal loop. Science 2025 Apr 4;388(6742):eadr3675. [doi: 10.1126/science.adr3675] [Medline: 40179184]
22. Li H, Peng A, Lai W, et al. Impacts of education level on Montreal Cognitive Assessment and saccades in community residents from Western China. Clin Neurophysiol 2024 May;161:27-39. [doi: 10.1016/j.clinph.2024.02.017] [Medline: 38432186]
23. Abul Hassan M, Aldridge CM, Zhuang Y, et al. Approach to quantify eye movements to augment stroke diagnosis with a non-calibrated eye-tracker. IEEE Trans Biomed Eng 2023 Jun;70(6):1750-1757. [doi: 10.1109/TBME.2022.3227015] [Medline: 37015585]
24. Lu Q, Zeng Y, Wang X, Chen Y, Deng J, Yan C. Eye-tracking during free visual exploration of familiar dramatic character faces facilitates rapid and accurate stroke recognition. Front Neurosci 2025;19:1692719. [doi: 10.3389/fnins.2025.1692719] [Medline: 41567510]
25. Jamali A, Baluchnejadmojarad T, Mehdizadeh H, Jazaeri SZ, Fallah S, Taghizadeh G. Impact of post stroke fatigue on saccadic eye movement control and learning through inflammatory mechanisms. Sci Rep 2025 Nov 20;15(1):40911. [doi: 10.1038/s41598-025-20200-9] [Medline: 41266558]
26. tefănescu E, Balea M, Chelaru VF, et al. Memory-guided saccades in subacute and chronic stroke: secondary data analysis of the N-PEP-12 clinical study. Biomedicines 2024 Jul 26;12(8):1678. [doi: 10.3390/biomedicines12081678] [Medline: 39200143]
27. Zhou H, Han Y, Xie D, et al. Predictors and clinical implications of post-stroke cognitive impairment: a retrospective study. Sci Rep 2025 Jul 7;15(1):24198. [doi: 10.1038/s41598-025-08048-5] [Medline: 40624177]
28. Chen H, Du H, Yi F, et al. Artificial intelligence-assisted oculo-gait measurements for cognitive impairment in cerebral small vessel disease. Alzheimers Dement 2024 Dec;20(12):8516-8526. [doi: 10.1002/alz.14288] [Medline: 39410879]
29. Verón GL, Juantorena GE, Keller G, Crivelli L, Kamienkowski JE. Eye tracking as a diagnostic tool in Alzheimer's disease, mild cognitive impairment, and related dementias: a systematic review. Alzheimers Dement (Amst) 2026;18(1):e70238. [doi: 10.1002/dad2.70238] [Medline: 41476954]

XSL•FO
RenderX

30. Maldonado-Díaz M, Jara-Vargas G, González-Seguel F. Visual attention during non-immersive virtual reality balance training in older adults with mild to moderate cognitive impairment: an eye-tracking study. Front Aging Neurosci 2025;17:1671477. [doi: 10.3389/fnagi.2025.1671477] [Medline: 41195081]
31. Powers WJ, Rabinstein AA, Ackerson T, et al. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2019 Dec;50(12):e344-e418. [doi: 10.1161/STR.0000000000000211] [Medline: 31662037]
32. Salvadori E, Cova I, Mele F, Pomati S, Pantoni L. Prediction of post-stroke cognitive impairment by Montreal Cognitive Assessment (MoCA) performances in acute stroke: comparison of three normative datasets. Aging Clin Exp Res 2022 Aug;34(8):1855-1863. [doi: 10.1007/s40520-022-02133-9] [Medline: 35441928]
33. Wei X, Liu Y, Li J, et al. MoCA and MMSE for the detection of post-stroke cognitive impairment: a comparative diagnostic test accuracy systematic review and meta-analysis. J Neurol 2025 May 18;272(6):407. [doi: 10.1007/s00415-025-13146-5] [Medline: 40383729]
34. Rost NS, Meschia JF, Gottesman R, et al. Cognitive impairment and dementia after stroke: design and rationale for the DISCOVERY study. Stroke 2021 Aug;52(8):e499-e516. [doi: 10.1161/STROKEAHA.120.031611] [Medline: 34039035]
35. Lee M, Yeo NY, Ahn HJ, et al. Prediction of post-stroke cognitive impairment after acute ischemic stroke using machine learning. Alzheimers Res Ther 2023 Aug 31;15(1):147. [doi: 10.1186/s13195-023-01289-4] [Medline: 37653560]
36. Ma Y, Yang Y, Wang X, et al. Prevalence and risk factors of poststroke cognitive impairment: a systematic review and meta-analysis. Public Health Nurs 2025;42(2):1047-1059. [doi: 10.1111/phn.13503] [Medline: 39702976]
37. Chau JPC, Lo SHS, Zhao J, et al. Prevalence of post-stroke cognitive impairment and associated risk factors in Chinese stroke survivors. J Neurol Sci 2023 Dec 15;455:122805. [doi: 10.1016/j.jns.2023.122805] [Medline: 37995462]
38. Díaz-Venegas C, Samper-Ternent R, Michaels-Obregón A, Wong R. The effect of educational attainment on cognition of older adults: results from the Mexican Health and Aging Study 2001 and 2012. Aging Ment Health 2019 Nov;23(11):1586-1594. [doi: 10.1080/13607863.2018.1501663] [Medline: 30449138]
39. Ji W, Wang C, Chen H, Liang Y, Wang S. Predicting post-stroke cognitive impairment using machine learning: a prospective cohort study. J Stroke Cerebrovasc Dis 2023 Nov;32(11):107354. [doi: 10.1016/j.jstrokecerebrovasdis.2023.107354] [Medline: 37716104]
40. Exalto LG, Weaver NA, Kuijf HJ, et al. Sex differences in poststroke cognitive impairment: a multicenter study in 2343 patients with acute ischemic stroke. Stroke 2023 Sep;54(9):2296-2303. [doi: 10.1161/STROKEAHA.123.042507] [Medline: 37551589]
41. Gall S, Phan H, Madsen TE, et al. Focused update of sex differences in patient reported outcome measures after stroke. Stroke 2018 Mar;49(3):531-535. [doi: 10.1161/STROKEAHA.117.018417] [Medline: 29438087]
42. Rosen MC, Freedman DJ. Multiplexing of cognitive encoding by oculomotor networks leads to incidental gaze shifts. Proc Natl Acad Sci U S A 2025 Apr 15;122(15):e2422331122. [doi: 10.1073/pnas.2422331122] [Medline: 40198709]
43. Levi AJ, Huk AC. Interpreting temporal dynamics during sensory decision-making. Curr Opin Physiol 2020 Aug;16:27-32. [doi: 10.1016/j.cophys.2020.04.006] [Medline: 32864535]
44. Polden M, Crawford TJ. Eye movement latency coefficient of variation as a predictor of cognitive impairment: an eye tracking study of cognitive impairment. Vision (Basel) 2023 May 1;7(2):38. [doi: 10.3390/vision7020038] [Medline: 37218956]
45. Spering M. Eye movements as a window into decision-making. Annu Rev Vis Sci 2022 Sep 15;8(1):427-448. [doi: 10.1146/annurev-vision-100720-125029] [Medline: 35676097]
46. Yang Q, Wang T, Su N, Xiao S, Kapoula Z. Specific saccade deficits in patients with Alzheimer's disease at mild to moderate stage and in patients with amnestic mild cognitive impairment. Age (Dordr) 2013 Aug;35(4):1287-1298. [doi: 10.1007/s11357-012-9420-z] [Medline: 22576337]
47. Wolf A, Tripanpitak K, Umeda S, Otake-Matsuura M. Eye-tracking paradigms for the assessment of mild cognitive impairment: a systematic review. Front Psychol 2023;14:1197567. [doi: 10.3389/fpsyg.2023.1197567] [Medline: 37546488]
48. Donnarumma F, Prevete R, Chersi F, Pezzulo G. A programmer-interpreter neural network architecture for prefrontal cognitive control. Int J Neural Syst 2015 Sep;25(6):1550017. [doi: 10.1142/S0129065715500173] [Medline: 25986752]
49. Nakajima M, Schmitt LI, Halassa MM. Prefrontal cortex regulates sensory filtering through a basal ganglia-to-thalamus pathway. Neuron 2019 Aug 7;103(3):445-458.e10. [doi: 10.1016/j.neuron.2019.05.026] [Medline: 31202541]
50. Levy R. The prefrontal cortex: from monkey to man. Brain 2024 Mar 1;147(3):794-815. [doi: 10.1093/brain/awad389] [Medline: 37972282]
51. Wang F, Hua S, Zhang Y, et al. Association between small vessel disease markers, medial temporal lobe atrophy and cognitive impairment after stroke: a systematic review and meta-analysis. J Stroke Cerebrovasc Dis 2021 Jan;30(1):105460. [doi: 10.1016/j.jstrokecerebrovasdis.2020.105460] [Medline: 33227579]
52. Filler J, Georgakis MK, Dichgans M. Risk factors for cognitive impairment and dementia after stroke: a systematic review and meta-analysis. Lancet Healthy Longev 2024 Jan;5(1):e31-e44. [doi: 10.1016/S2666-7568(23)00217-9] [Medline: 38101426]

XSL•FO
RenderX

53. Georgakis MK, Fang R, Düring M, et al. Cerebral small vessel disease burden and cognitive and functional outcomes after stroke: a multicenter prospective cohort study. Alzheimers Dement 2023 Apr;19(4):1152-1163. [doi: 10.1002/alz.12744] [Medline: 35876563]
54. Coenen M, de Kort FA, Weaver NA, et al. Strategic white matter hyperintensity locations associated with post-stroke cognitive impairment: a multicenter study in 1568 stroke patients. Int J Stroke 2024 Oct;19(8):916-924. [doi: 10.1177/17474930241252530] [Medline: 38651756]

Abbreviations

AD: Alzheimer disease
AUC: area under the receiver operating characteristic curve
C-index: concordance index
HC: healthy control
MCI: mild cognitive impairment
mHealth: mobile health
MMSE: Mini-Mental State Examination
MoCA: Montreal Cognitive Assessment
MRI: magnetic resonance imaging
mRS: Modified Rankin Scale
NIHSS: National Institutes of Health Stroke Scale
NPSCI: non–poststroke cognitive impairment
OR: odds ratio
PSCI: poststroke cognitive impairment
VPC: visual paired comparisons
WMH: white matter hyperintensity

Edited by A Stone, L Buis; submitted 02.Sep.2025; peer-reviewed by C Wang, C Baxter; revised version received 26.Jun.2026; accepted 26.Jun.2026; published 06.Aug.2026.

Please cite as:
Xie B, Ren R, Zhang Y, Wang X, Zhen J
Utility of Tablet-Based Eye Tracking for Early Screening of Poststroke Cognitive Impairment: Diagnostic Cohort Study
JMIR Mhealth Uhealth 2026;14:e83194
URL: https://mhealth.jmir.org/2026/1/e83194
doi:10.2196/83194

© Boxu Xie, Ronghui Ren, Yuting Zhang, Xuan Wang, Junli Zhen. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 6.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Facilitators of and Barriers to Implementation of a Tablet-Based Digital Health Program for Colorectal Cancer Screening in Primary Care: Qualitative Pragmatic Implementation Study

Mark Cromo[1], BS; Aliza Randazzo[2], BS; David P Miller Jr[3], MD, MS; Kristie L Foley[2], MS, PhD; Ajay Dharod[3], MD; Anna C Snavely[4], PhD; Elena Wright[2], MS; Holly Brower[5], PhD; Mark Dignan[1], MPH, PhD

[1]College of Medicine, Department of Internal Medicine, University of Kentucky, Lexington, KY, United States

[2]Department of Implementation Science, Wake Forest University School of Medicine, Winston-Salem, NC, United States

[3]Department of Internal Medicine, Department of Implementation Science, Wake Forest University School of Medicine, Winston-Salem, NC, United States

[4]Department of Biostatistics and Data Science, Wake Forest University School of Medicine, Winston-Salem, NC, United States

[5]School of Business, Wake Forest University, Winston-Salem, NC, United States

Corresponding Author:
Mark Dignan, MPH, PhD
College of Medicine
Department of Internal Medicine
University of Kentucky
760 Press Ave Room 343
Lexington, KY, 40536-0679
United States
Phone: 1 8599130207
Email: mark.dignan@uky.edu

Abstract

Background: Colorectal cancer (CRC) incidence and mortality rates continue to be elevated even though effective screening methods are widely available. To increase CRC screening in primary care practices, our team developed a tablet-based digital health program (mPATH) designed to identify patients needing CRC screening, provide education, and empower patients to request a screening test via the program.

Objective: This study aimed to qualitatively assess facilitators of and barriers to implementing and maintaining mPATH in primary care clinics.

Methods: In a pragmatic implementation trial, clinics were randomized to receive only in-person training and technological support via phone or email (low touch) or added levels of support, such as at-elbow support during launch, regular check-ins, memos, and reports (high touch). After implementation and data collection were concluded, we conducted telephone interviews with health care providers, clinic managers, and front desk and nursing staff recruited from 8 primary care clinics of varying sizes and with varying degrees of implementation of mPATH. The interviews were designed to collect perceived facilitators of and barriers to using mPATH. All interviews were administered via telephone by a single project staff member with no prior contact with participants. Interviews were audio-recorded, and 2 study team members independently coded each interview transcript and developed a codebook to identify meaningful categories in the dataset. The coders met periodically to resolve discrepancies. Data within each category were abstracted and synthesized into themes. Themes were determined inductively by prevalence and salience in the data per the principles of thematic analysis.

Results: A total of 33 interviews were completed between September 2021 and April 2023 with health care providers (n=8, 24.2%), clinic managers (n=9, 27.3%), nursing staff (n=8, 24.2%), and front desk staff (n=8, 24.2%). Interviews averaged 26.7 (SD 4.9) minutes. Barriers and facilitators identified varied among clinic sites and by clinic role. Overall, the primary factors supporting the implementation of mPATH were health care provider and staff buy-in, perceived potential time savings, and workflow improvement. The primary barriers identified were perceived lack of need for the program and technical issues. There was no significant indication that clinic size or randomization to low- or high-touch training and support played a role in the decision to continue or stop using the program.

XSL•FO
RenderX

Conclusions: Implementation of a tablet-based CRC screening tool in primary care practices is feasible with health care provider and staff buy-in and validation of potential time savings and workflow improvements but may be limited by perceived lack of need for the program and technical issues.

(JMIR Mhealth Uhealth 2026;14:e74206) doi:10.2196/74206

KEYWORDS

clinical decision support; mobile health technology; electronic health records; medical informatics applications; primary health care

Introduction

Colorectal cancer (CRC) is the third most common cancer diagnosed among men (15.1/100,000) and women (10.7/100,000) in the United States, and it is the second leading cause of cancer-related death [1,2]. In 2025, an estimated 154,000 Americans will be diagnosed with CRC, and more than 52,000 will die from CRC [1]. The US Preventive Services Task Force recommends that average-risk adults be screened for CRC starting at the age of 45 years using any of several efficacious tests [3]. Unfortunately, although CRC screening rates have increased in recent decades, they remain relatively low [4] at approximately 70% and fall below the goal of 80% in every community promoted by the American Cancer Society National Colorectal Cancer Roundtable [5].

Multilevel barriers contribute to low CRC screening rates. Some patients are unaware of screening methods, and those aware of stool-based screening may be reluctant to provide stool samples [6,7]. For colonoscopy, patients report difficulty completing the preparation and may have other barriers such as cost, the need to take time off work, and arranging for transportation [8]. Decision aids have been developed and tested for prostate cancer, lung cancer, and CRC screening to assist health care providers and patients in identifying suitable screening methods [9-12], but primary care clinicians report that they lack the time needed for shared decision-making [13-16].

With the goal of creating a decision aid that addresses the patient, health care provider, and system barriers to CRC screening and requires minimal time from health care providers to implement, our group developed and tested a tablet-based app (mPATH) that identifies patients overdue for CRC screening, provides education, and allows them to request a screening test via the program. In a prior randomized controlled trial conducted in community-based primary care practices, mPATH doubled the proportion of patients who completed CRC screening [17].

To determine the optimal strategy for implementing the app in routine care, we then conducted a pragmatic cluster randomized type III hybrid implementation-effectiveness trial in 18 community-based primary care practices to evaluate a “low-touch” vs “high-touch” bundle of implementation strategies [18]. Low-touch clinics received an on-site training session on use of the app and access to ongoing technical support. High-touch clinics received the same support plus at-elbow support on the day the program was launched in the clinic, as well as periodic coaching calls throughout the 6-month implementation period and regular audit and feedback reports on program use. We observed initial moderate use of the program, which decreased over time with no difference by implementation strategy [19-21]. We also observed variation in use of the program across practices and staff. To identify and better understand the factors that led to higher or lower use of the program, we conducted in-depth interviews with clinicians and staff who participated in the trial.

Methods

Overview of the mPATH Program

On the basis of formative research with clinical staff and health care providers [18], the mPATH app was divided into 2 separate tablet programs: mPATH-CheckIn and mPATH-CRC. mPATH-CheckIn was provided to patients by front desk staff at check-in to be completed in the waiting room. mPATH-CheckIn included the health system’s mandated screening items for depression, fall risk, and intimate partner violence. An application programming interface connection to the electronic health record (EHR) imported CRC screening history. For age-eligible patients with no record of current CRC screening, mPATH-CheckIn asked additional questions to determine whether screening had been completed outside the health system and whether the patient had risk factors that would change screening recommendations. Those patients indicated as eligible or due for screening on mPATH-CheckIn were then provided with a separate tablet (mPATH-CRC) by nursing staff when they transitioned from the waiting room to an examination room, which displayed a screening decision aid video and closed by allowing patients to request a screening test, either a fecal immunochemical test or a colonoscopy. mPATH-CRC then immediately sent an EHR in-basket message to the health care provider, alerting them that the patient had used the program and summarizing the outcome. When the health care provider arrived in the examination room, a pop-up alert let the provider order the requested CRC screening test with minimal clicks.

Recruitment

We intentionally selected 8 primary care clinics from the 18 enrolled in the randomized trial described previously [20], varying by practice size, participant group, and degree of mPATH adoption, with the aim to conduct interviews with a representative sample of the primary care clinics in our randomized trial. Our sample included 3 small (1-3 health care providers) and 5 large (>3 health care providers) clinics. The sample included clinics that continued use of mPATH throughout the 6-month implementation period and those that discontinued use. Characteristics of the 8 clinics selected for interviews are shown in Table 1.

XSL•FO
RenderX

Table 1. Clinic characteristics.

Clinic	Size	High or low touch	Number of clinical personnel	Used mPATH for 12 months	Discontinued mPATH use before 12 months
A	Large	High	26	✓	
B	Small	High	8	✓	
C	Large	High	23	✓	
D	Large	High	14		✓
E	Small	Low	12	✓	
F	Large	Low	16		✓
G	Large	Low	16		✓
H	Small	Low	12		✓

Interview Development

Semistructured interview guides were developed by the investigators and pretested to ensure clarity. Separate interview guides were developed for (1) front desk or administrative staff and (2) clinical personnel or practice managers. Because the front desk and administrative staff did not interact with the mPATH-CRC program, their interviews only asked about mPATH-CheckIn, whereas the clinical personnel and practice manager interviews included questions about both mPATH-CheckIn and mPATH-CRC. The interview guides were designed to collect descriptive data on the interviewee and explore the following constructs: (1) overall assessment of the effectiveness of mPATH-CheckIn for screening for depression, fall risk, and home safety and effectiveness of mPATH-CRC for CRC screening; (2) perceived barriers and facilitators as related to mPATH and clinic flow; (3) impact of mPATH on screening and clinic operations; (4) resources or training that were provided to the practice to support implementation; (5) program champion and leadership functioning; and (6) scale-up and dissemination. Because the trial took place during the COVID-19 pandemic, all interviewees were also asked how COVID-19 affected the implementation of mPATH.

Data Collection, Management, and Analyses

Study personnel identified potential interviewees in collaboration with each practice's clinic champion, a clinical staff member who served as the primary contact and was responsible for promoting mPATH internally and encouraging its adoption. Because we had previously conducted extensive usability testing of the app with patients, this study focused on health care providers and clinic staff [18]. To avoid influencing the randomized trial results, interviews were conducted after 1 year of implementation data collection was completed or after they stopped using the program, if earlier. All interviews were conducted via telephone by a single independent project staff member with no prior contact with participants. Interviews were audio-recorded, transcribed, deidentified, and reviewed. Transcripts were imported into ATLAS.ti (Scientific Software Development GmbH) to store and manage the data. Two study team members (AR and MC) inductively developed a codebook to identify meaningful categories in the dataset. Each transcript was independently coded by the 2 study team members, who met periodically to resolve discrepancies and revise the codebook as needed. Response saturation was identified when interview responses were repetitive rather than providing new information. The interview data were analyzed, and differences among participant groups (health care providers, nursing staff, clinic managers, and front desk staff) were labeled. Data within each category were abstracted and synthesized into themes. Themes were determined inductively by their prevalence and salience in the data per the principles of thematic analysis.

Ethical Considerations

The Wake Forest University School of Medicine Institutional Review Board approved this study with a waiver of signed informed consent (IRB00048919). This study was also approved by the University of Kentucky Office of Research Integrity under a reliance agreement. Participants were provided with a study information sheet at the time of interview scheduling describing the purpose of the study and their rights to participate; risks and benefits and financial disclosures were declared. Participants provided verbal consent at the beginning of the interview to proceed and have their responses recorded. Interviews were conducted in private offices for confidentiality, and any participant or staff identifiers that were used in the interview discussion were removed from the transcripts. All interviewees were offered a US $50 gift card as compensation for their time.

Results

Overview

A total of 33 interviews were completed between September 2021 and April 2023 with health care providers (n=8, 24.2%), clinic managers (n=9, 27.3%), nursing staff (n=8, 24.2%), and front desk staff (n=8, 24.2%). Interviews averaged 25 minutes in length. Study staff identified the most prevalent facilitators of and barriers to mPATH implementation according to the methods described above. While some implementation facilitators and barriers were common to both mPATH-CheckIn and mPATH-CRC, there was some divergence of opinions on the 2, as described below.

Implementation Facilitators

Staff and Health Care Provider Buy-In

Staff buy-in was identified as a significant factor in the success of mPATH implementation. Clinic staff were initially hesitant about implementing a new program and new workflows to

XSL•FO
RenderX

accommodate mPATH, but this concern was alleviated with experience using the program:

> *I think that the biggest challenge would be, people worried about their workflow, and doctors are like everybody else. They do a lot of things just because that's the way they do it. Implementing anything new maybe makes everyone a little nervous, but I think in general, everybody felt the way I did, that it didn't disrupt workflows any significant way, so it was probably scary at the beginning, and didn't prove to be much trouble, and once you do a new thing for a little while, you are okay.* [Clinic A; health care provider]

> *I think the number one most important factor was the—for lack of a better term—buy-in by the docs and the staff. Initially, folks were very enthusiastic about it.* [Clinic D; health care provider]

Buy-in from front desk staff was noted as important in mPATH-CheckIn because they were responsible for cleaning, storing, charging, and distributing mPATH-CheckIn tablets; instructing patients on use; answering patient questions; and preparing the tablets for each patient. Health care provider buy-in was also very important for the implementation of mPATH-CRC as they felt responsible for CRC screening in their eligible patients. Nursing staff bridged both mPATH-CheckIn and mPATH-CRC implementation workflows, and buy-in from them was important for smooth implementation and transition between both apps. mPATH-CheckIn had the most impact on nursing staff because it improved their workflow by conducting the required check-in screening, which, in turn, increased their buy-in to the program.

Training and Support

Interviewees found the in-clinic training sessions to be effective, especially the "hands-on" aspect of being able to test the tablets:

> *Yes, when we got trained on it, it was, basically, real-time training, and so she [mPATH research team staff member] would come in and help us on a day when we had patients. She walked through one patient with us, and just be there, step by step.* [Clinic E; management]

Leadership Support and Champion

Before implementation, each clinic identified a champion to lead implementation efforts. The practice manager or clinical coordinator was often identified as the champion. Staff recognized the importance of the clinic champion role and felt that mPATH implementation was well supported at the clinic leadership level.

Efficiency, Workflow, and Time Savings

Interviewees agreed that mPATH-CheckIn was a significant time saver and improved workflow during patient check-in and rooming, as patients answered questions directly into the tablet and answers were transferred from the tablet directly into the EHR:

> *Oh, 'cause it saves so much time between them already answering them, and they just pull it over into our system. Versus me going through each question myself and having to put those answers in. I just pull it over in there, and I didn't have to go into any of those questions and open it up and ask.* [Clinic G; nurse]

Interviewees agreed that mPATH-CheckIn standardized and automated the assessment process and allowed more patients to be screened for depression, fall risk, and household violence than with previous workflows:

> *I do feel like it's made the screening part much easier and faster, and we've been able to screen more patients...Typically, we wouldn't get the chance to ask the questions just because of time and things like that. Typically, now they just do it in the waiting room when they're checking in. I feel like a large majority of our patients are getting screened now.* [Clinic E; health care provider]

Patient Openness

Patient openness was noted by a significant number of interviewees as an important benefit of mPATH-CheckIn, specifically regarding depression questions. Staff perceived that asking depression questions on the tablet instead of verbally helped patients be more open and forthcoming with their answers. On the basis of interview responses, more patients with depression symptoms were identified through the tablet-based depression questionnaire than through previous procedures:

> *I think it actually helps the patient sometimes. It's easier for them just to answer the question on the [tablet] rather than verbalize it 'cause sometimes they're embarrassed to verbalize it to someone in-person. I think it's easier for them just to answer it on the [tablet].* [Clinic C; nurse]

Patient Knowledge

Interviewees overall felt that the mPATH-CRC module with its educational video improved patient knowledge of CRC screening and the screening options available. It helped clarify the importance of screening, generate questions, and open the discussion when the health care provider entered the examination room:

> *I think that the content behind...[mPATH-CRC]...of giving our patients a visual, kind of audio/video on education with colon cancer screening, why they should be screened, and how they can do that is really helpful because I do that education with my patients, but some people learn better in a different manner, so that gives them a more visual or auditory way to learn. Then, they have, of course, the opportunity to ask more questions.* [Clinic F; health care provider]

Potential Impact on CRC Screening

Interviewees expressed that they believed that mPATH-CRC helped increase CRC screening or at least had the potential to increase screening:

XSL•FO
RenderX

> *I think the people that were on the fence of doing any type of screening, I think it did help do some screening. I think before our people said, nope, I'm not doing that—and I think after the mPATH, more people were open toward it.* [Clinic C; health care provider]

Implementation Barriers

Analysis of the interviews also identified barriers to implementing mPATH in the clinics.

COVID-19 Pandemic–Related Barriers

Interviewees acknowledged that there were intervals of time when they did not use mPATH due to the pandemic or that there were protocols for disinfecting the mPATH tablets that slowed implementation:

> *...some of the patients, especially with COVID, would complain that they didn't want to touch it (the tablet), so for a while, we could not use it at all, but now that we've added it back, some patients like it, and some patients absolutely refuse to do it.* [Clinic A; management]

However, not all clinics reported being affected by the pandemic:

> *I don't think it affected it a whole lot. We sterilize our live stuff anyway. The only extra thing was more sterilization, but COVID really didn't affect it a lot. We don't really give the iPad to sick patients. None of those sick visits get the iPads, so it really didn't affect it as much. Now, of course, patient volume was down during COVID, so we didn't use it as often but other than that.* [Clinic C; management]

Technical Issues and EHR Limitations

Although, overall, interviews indicated that mPATH-CheckIn improved workflow, several interviewees (clinics B and G) voiced concerns about occasional data transmission errors between mPATH-CheckIn and the EHR:

> *There seemed to be an intermittent problem with the Medicare wellness visit when we used the mPATH for check-in. It would pull in some erroneous values. We weren't able to figure out if that was part of mPATH or something else...* [Clinic G; health care provider]

Typically, when an issue arose where data were missing, staff would revert to their previous workflow of asking patients the mPATH-CheckIn questions verbally and then manually entering the data into the EHR.

There was also a perception that mPATH failed to recognize patients who received CRC screening outside the health system. Two clinics (A and F) with high numbers of patients who received screening outside the health system reported issues with mPATH-CRC properly identifying patients' CRC screening status:

> *...almost all of our people are screened with outside of [University] system, and mPATH can't see it. mPATH almost always tells me my patient hasn't had a screen, and I just have to look anyway and see that they have, and then I have to make sure that it's documented in a way that's acceptable to our system, but being on the periphery like this, I don't think mPATH stops me from basically having to do that same process and look. It always tells me they don't have it, and most of them do. If our patients were getting it through a [University] gastroenterologist and it was in our [EHR], I think that would be different, but it doesn't seem to identify the ones of ours that are done in community clinics, outside of our network.* [Clinic A; health care provider]

This added to staff burden as the health care provider or nurse had to double-check or verify the screening status of each patient before giving the patient the mPATH-CRC tablet.

Lack of Need or Demand for the Program

Specific to mPATH-CRC, while interviewees were generally positive about the potential of the program to improve patient knowledge and potentially help increase screening rates, most reported that they infrequently used mPATH-CRC, primarily due to patient eligibility factors. Clinics reported that many of their patients were outside of screening age or were up-to-date with CRC screening due to a colonoscopy within the previous 10 years.

Efficiency, Workflow, and Time Barriers

There were some barriers noted that impacted workflow, specifically the transition between the mPATH-CheckIn and mPATH-CRC apps. As patients were given mPATH-CheckIn in the waiting room and then mPATH-CRC after they were taken back to an examination room, any delays at the front end of the workflow had the potential to carry over to the back end. This particularly occurred with patients with limited technical skills who required more time with the tablets:

> *I feel like, for our patients that were more technologically inclined, it, in some cases, was able to help facilitate that data gathering more quickly. For other patients that were less technologically inclined, we found that often they hadn't made it through or had only partially made it through the questionnaire. Then the MAs [medical assistants] would go back in and redo the process.* [Clinic G; health care provider]

For mPATH-CRC specifically, several clinics said that the 5-minute screening decision aid video was too long and had a negative impact on workflow, especially health care provider workflow:

> *Because I personally liked that they can watch the video, but some of our providers do not like it, because it obviously delays them going into the room a little bit. That's the main hiccup that we've had is just the providers have to wait however many minutes. I don't remember how many minutes the video is, but they haven't wanted to do that.* [Clinic A; nurse]

XSL•FO
RenderX

Overall Recommendation for mPATH

Health care providers and staff from 4 clinics reported that they were still using mPATH at the time of their interviews, whereas 1 clinic reported continuing use of mPATH-CheckIn only. Interviewees from 3 clinics reported that they had discontinued use of the program at the time of their interviews. However, those who had discontinued the program reported that they liked it and would have continued using it if it were up to them individually. The decision to discontinue the program was often made by other clinic leadership who were not interviewed or was a group decision among health care providers and staff as a whole. All interviewees were positive about recommending that other clinics at least try mPATH to determine whether it was a good fit for them.

Discussion

Principal Findings

We found that the opinions of the clinical staff and health care providers about implementing the tablet-based screening program were impacted by staff buy-in, how well the program was incorporated into existing workflows, and the additional value it delivered. Concern related to technical issues was the main barrier described by interviewees. Those interviewed considered mPATH-CheckIn to be beneficial as a time saver and as a means to streamline and improve efficiency in the patient data collection process at check-in by providing an easy-to-use technology for increasing their quality metrics for depression, fall risk, and household violence and abuse. These factors seemed to serve as motivators for clinics to use mPATH, especially the clinical staff.

While the clinic staff interviewed expressed appreciation for the concept of mPATH-CRC "in theory," it was challenging to implement due to a variety of factors. Health care providers were most opinionated about mPATH-CRC compared to nursing staff, perhaps due to health care providers expressing their feelings of "ownership" and "responsibility" for ensuring that their patients were up-to-date with CRC screening. The nursing staff interviewed had a very limited or peripheral role in the CRC screening process—primarily reviewing patient screening history and notifying health care providers if the patient was due for screening. The health care providers interviewed seemed more resistant to changing their current workflows to accommodate mPATH-CRC and often indicated that "they do what mPATH-CRC does anyway," seeing it as a duplication or add-on to their existing preferred process. Another major factor in implementing mPATH-CRC was the belief that most patients were already up-to-date with CRC screening and did not need mPATH-CRC. Finally, mPATH-CheckIn was intended to be used with every patient aged >18 years, and mPATH-CRC was intended to be used just with those patients aged >50 years who were due for screening, reflecting screening guidelines in place at the start of the study. The need to differentiate patients by age may have impacted clinics' comfort level with the mPATH-CRC program as it was not routinely used for all adult patients.

Our findings are consistent with those of other trials that have reported difficulty implementing digital health interventions in primary care environments. In a hybrid implementation-effectiveness study of a digital health program to support patients with substance use disorders in federally qualified health centers, the program was offered to fewer than 10% of eligible patients [22]. Mei and Vaez [23] implemented electronic screening for substance use and mental health in a community-based university health system–affiliated primary care practice; although screening completion increased, the absolute improvement was modest at 9%. Similarly, Yakovchenko et al [24] implemented an automatic SMS text messaging program to help with hepatitis C virus medication adherence in Veterans Health Administration primary care clinics, with fewer than 20% of approached patients progressing to active use of the system.

Despite being conducted in markedly different health system contexts (federally qualified health centers, the US Department of Veterans Affairs health system, and a university-affiliated primary care practice), all 3 studies reported similar barriers to implementation, including workflow disruptions, technical challenges, and increased clinician burden. Each study also identified perceived time savings, leadership support, and staff engagement as important facilitators of adoption. Consistent with these findings, published literature reviews have repeatedly identified workflow integration as one of the strongest determinants of digital health implementation, with tools perceived to save time more likely to be adopted and those perceived to add burden less likely to be used [12,25-27].

Our study extends this literature by highlighting the importance of routinization. Participants specifically mentioned that the mPATH-CRC program was less likely to be used because it was only administered to age-eligible patients who needed screening. In contrast, the mPATH-CheckIn program was seen as easier to administer because it was designed to be used with every patient at every visit. In addition, because our study was conducted during the COVID-19 pandemic, our findings yielded unique insights into how public health concerns can affect implementation. Approximately half of the interviewees reported that patients were concerned about germs and touching the tablets. Although clinics used antibacterial tablet cases and implemented cleaning protocols that were communicated to patients, these reassurances were not always adequate to allay concerns. In addition, pandemic-related workflows, such as restricted tablet distribution to patients who were well, limited implementation in some clinics.

Limitations

The program was implemented in clinics all within the same health system and the same EHR system, which may have impacted opinions, perceptions, and technical issues regarding connectivity and integration. While we sought candid interview responses by ensuring confidentiality and using an external interviewer, social desirability bias may have affected responses. The fact that the study was conducted in the same health system to which the research team is affiliated may have also impacted responses. As stated previously, we conducted interviews after all quantitative implementation outcome data were collected, which was 1 year after program launch. This time delay may have impacted respondents' memories and perceptions of the

XSL•FO
RenderX

initial training and implementation activities. Finally, because we previously found that the program had excellent patient usability [13,15], and as staff were responsible for administering the program, we did not include patients in our interviews. Patient-level factors not identified by clinical staff may have affected implementation.

Conclusions

When implementing tablet-based digital health programs in primary care clinics, demonstrating the programs' ability to improve care, fit into usual workflows, and ideally save time is critical. Providing hands-on training before or at the start of implementation and identifying a suitable clinic champion can help increase buy-in and comfort with the program. For decision aids related to cancer or other health screenings, the demand for the program should be weighed against potential barriers to ensure that the program can be implemented and sustained effectively. Review of potential implementation facilitators and barriers can help maximize the uptake of digital health programs and enhance continued use.

Acknowledgments

This study was funded by the US Department of Health and Human Services, National Cancer Institute (R01CA218416), and Wake Forest University School of Medicine Comprehensive Cancer Center (P30CA012197). No generative artificial intelligence was used in any portion of the manuscript.

Data Availability

Data sharing is not applicable to this paper as no datasets other than transcripts were generated or analyzed during this study.

Authors' Contributions

MC and AR developed the interview guide and conducted the interviews, conducted complete analysis of the data, and were major contributors in writing the manuscript. KLF and EW contributed to the interview and were major contributors in writing the manuscript. MBD and AD contributed to the interview guide and to writing the manuscript. ACS contributed to writing the manuscript. DPM contributed to the interview guides and was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

DPM and AD are coinventors of mPATH. DPM, AD, and Wake Forest University Health Sciences have an ownership interest in the mPATH app. Neither DPM nor AD participated in data collection or analysis. KLF and MD report grants from the National Cancer Institute (NCI) during the conduct of this study. ACS reports grants from the National Institutes of Health and the NCI during the conduct of this study and grants and personal fees from Shattuck Labs outside the submitted work. AD reports grants from the NCI during the conduct of this study. DPM reports grants from the NCI during the conduct of this study. All other authors declare no other conflicts of interest.

References

1. Cancer facts and figures. American Cancer Society. 2025. URL: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/2025-cancer-facts-figures.html [accessed 2026-02-03]
2. Screening for colorectal cancer. Centers for Disease Control and Prevention. 2025 Feb 26. URL: https://www.cdc.gov/colorectal-cancer/screening/index.html [accessed 2025-12-11]
3. US Preventive Services Task Force. URL: https://www.uspreventiveservicestaskforce.org/uspstf/ [accessed 2025-12-11]
4. 80% in every community. American Cancer Society. URL: https://nccrt.org/our-impact/80-in-every-community [accessed 2025-12-11]
5. Coughlin SS. Social determinants of colorectal cancer risk, stage, and survival: a systematic review. Int J Colorectal Dis 2020 Jun 21;35:985-995. [doi: 10.1007/s00384-020-03585-z] [Medline: 32314192]
6. Travis E, Kerrison RS, O'Connor DB, Ashley L. Barriers and facilitators to colonoscopy for cancer detection: patient and practitioner perspectives. Psychol Health 2024 Sep;39(9):1263-1283 [FREE Full text] [doi: 10.1080/08870446.2022.2141241] [Medline: 36373225]
7. Luque JS, Wallace K, Blankenship BF, Roos LG, Berger FG, LaPelle NR, et al. Formative research on knowledge and preferences for stool-based tests compared to colonoscopy: what patients and providers think. J Community Health 2018 Dec;43(6):1085-1092 [FREE Full text] [doi: 10.1007/s10900-018-0525-x] [Medline: 29767821]
8. Estevan-Vilar M, Parker LA, Caballero-Romeu JP, Ronda E, Hernández-Aguado I, Lumbreras B. Barriers and facilitators of shared decision-making in prostate cancer screening in primary care: a systematic review. Prev Med Rep 2024 Jan;37:102539 [FREE Full text] [doi: 10.1016/j.pmedr.2023.102539] [Medline: 38179441]
9. Jimbo M, Shultz CG, Nease DE, Fetters MD, Power D, Ruffin MT4. Perceived barriers and facilitators of using a web-based interactive decision aid for colorectal cancer screening in community practice settings: findings from focus groups with

XSL•FO
RenderX

primary care clinicians and medical office staff. J Med Internet Res 2013 Dec 18;15(12):e286 [FREE Full text] [doi: 10.2196/jmir.2914] [Medline: 24351420]
10. Wolf AM, Oeffinger KC, Shih TY, Walter LC, Church TR, Fontham ET, et al. Screening for lung cancer: 2023 guideline update from the American Cancer Society. CA Cancer J Clin 2024;74(1):50-81 [FREE Full text] [doi: 10.3322/caac.21811] [Medline: 37909877]
11. Abubaker-Sharif M, Shusted C, Myers P, Myers R. Primary care physician perceptions of shared decision making in lung cancer screening. J Cancer Educ 2022 Aug;37(4):1099-1107. [doi: 10.1007/s13187-020-01925-9] [Medline: 33230673]
12. Scholl I, LaRussa A, Hahlweg P, Kobrin S, Elwyn G. Organizational- and system-level characteristics that influence implementation of shared decision-making and strategies to address them - a scoping review. Implement Sci 2018 Mar 09;13(1):40 [FREE Full text] [doi: 10.1186/s13012-018-0731-z] [Medline: 29523167]
13. Almadi MA, Mosli MH, Bohlega MS, Al Essa MA, AlDohan MS, Alabdallatif TA, et al. Effect of public knowledge, attitudes, and behavior on willingness to undergo colorectal cancer screening using the health belief model. Saudi J Gastroenterol 2015;21(2):71-77 [FREE Full text] [doi: 10.4103/1319-3767.153814] [Medline: 25843192]
14. Dalton AF, Golin CE, Morris C, Kistler CE, Dolor RJ, Bertin KB, et al. Effect of a patient decision aid on preferences for colorectal cancer screening among older adults: a secondary analysis of a randomized clinical trial. JAMA Netw Open 2022 Dec 01;5(12):e2244982 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.44982] [Medline: 36469317]
15. Ankolekar A, Dahl Steffensen K, Olling K, Dekker A, Wee L, Roumen C, et al. Practitioners' views on shared decision-making implementation: a qualitative study. PLoS One 2021;16(11):e0259844 [FREE Full text] [doi: 10.1371/journal.pone.0259844] [Medline: 34762683]
16. Schroy PC3, Duhovic E, Chen CA, Heeren TC, Lopez W, Apodaca DL, et al. Risk stratification and shared decision making for colorectal cancer screening: a randomized controlled trial. Med Decis Making 2016 May;36(4):526-535 [FREE Full text] [doi: 10.1177/0272989X15625622] [Medline: 26785715]
17. Miller DPJ, Spangler JG, Case LD, Goff DCJ, Singh S, Pignone MP. Effectiveness of a web-based colorectal cancer screening patient decision aid: a randomized controlled trial in a mixed-literacy population. Am J Prev Med 2011 Jun;40(6):608-615 [FREE Full text] [doi: 10.1016/j.amepre.2011.02.019] [Medline: 21565651]
18. Puccinelli-Ortega N, Cromo M, Foley KL, Dignan MB, Dharod A, Snavely AC, et al. Facilitators and barriers to implementing a digital informed decision making tool in primary care: a qualitative study. Appl Clin Inform 2022 Jan;13(1):1-9 [FREE Full text] [doi: 10.1055/s-0041-1740481] [Medline: 34986491]
19. Miller DPJ, Denizard-Thompson N, Weaver KE, Case LD, Troyer JL, Spangler JG, et al. Effect of a digital health intervention on receipt of colorectal cancer screening in vulnerable patients: a randomized controlled trial. Ann Intern Med 2018 Apr 17;168(8):550-557 [FREE Full text] [doi: 10.7326/M17-2315] [Medline: 29532054]
20. Miller DPJ, Snavely AC, Dharod A, Wright E, Randazzo A, Bundy R, et al. High-touch vs low-touch strategy for implementing a CRC screening digital health intervention: a randomized clinical trial. JAMA Intern Med 2025 Jul 01;185(7):788-795. [doi: 10.1001/jamainternmed.2025.0779] [Medline: 40323625]
21. Snavely AC, Foley K, Dharod A, Dignan M, Brower H, Wright E, et al. Effectiveness and implementation of mPATH™-CRC: a mobile health system for colorectal cancer screening. Trials 2023 Apr 14;24(1):274 [FREE Full text] [doi: 10.1186/s13063-023-07273-5] [Medline: 37060023]
22. Quanbeck AR, Gustafson DH, Marsch LA, McTavish F, Brown RT, Mares ML, et al. Integrating addiction treatment into primary care using mobile health technology: protocol for an implementation research study. Implement Sci 2014 May 29;9:65 [FREE Full text] [doi: 10.1186/1748-5908-9-65] [Medline: 24884976]
23. Mei C, Vaez K. Success factors for and barriers to integration of electronic mental health screening in primary care. Comput Inform Nurs 2020 Mar 16;38(9):441-450. [doi: 10.1097/CIN.0000000000000617] [Medline: 33955369]
24. Yakovchenko V, Hogan TP, Houston TK, Richardson L, Lipschitz J, Petrakis BA, et al. Automated text messaging with patients in department of veterans affairs specialty clinics: cluster randomized trial. J Med Internet Res 2019 Aug 04;21(8):e14750-e14758 [FREE Full text] [doi: 10.2196/14750] [Medline: 31444872]
25. Palacholla RS, Fischer N, Coleman A, Agboola S, Kirley K, Felsted J, et al. Provider- and patient-related barriers to and facilitators of digital health technology adoption for hypertension management: scoping review. JMIR Cardio 2019 Mar 26;3(1):e11951 [FREE Full text] [doi: 10.2196/11951] [Medline: 31758771]
26. Whitelaw S, Pellegrini DM, Mamas MA, Cowie M, Van Spall HG. Barriers and facilitators of the uptake of digital health technology in cardiovascular care: a systematic scoping review. Eur Heart J Digit Health 2021 Mar;2(1):62-74 [FREE Full text] [doi: 10.1093/ehjdh/ztab005] [Medline: 34048508]
27. Aanes SG, Wiig S, Nieder C, Haukland EC. Implementing digital patient-reported outcomes in routine cancer care: barriers and facilitators. ESMO Real World Data Digit Oncol 2024 Dec;6:100088. [doi: 10.1016/j.esmorw.2024.100088]

Abbreviations

CRC: colorectal cancer
EHR: electronic health record

XSL•FO
RenderX

Edited by L Buis; submitted 20.Mar.2025; peer-reviewed by O Enaworu, T Kayode, G Ely; comments to author 21.Oct.2025; revised version received 20.Jan.2026; accepted 28.Jan.2026; published 13.Mar.2026.

Please cite as:
Cromo M, Randazzo A, Miller Jr DP, Foley KL, Dharod A, Snavely AC, Wright E, Brower H, Dignan M
Facilitators of and Barriers to Implementation of a Tablet-Based Digital Health Program for Colorectal Cancer Screening in Primary Care: Qualitative Pragmatic Implementation Study
JMIR Mhealth Uhealth 2026;14:e74206
URL: https://mhealth.jmir.org/2026/1/e74206
doi:10.2196/74206
PMID:

©Mark Cromo, Aliza Randazzo, David P Miller Jr, Kristie L Foley, Ajay Dharod, Anna C Snavely, Elena Wright, Holly Brower, Mark Dignan. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 13.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Gait Analysis for Identifying Normal Cognition, Subjective Cognitive Decline, and Mild Cognitive Impairment in Parkinson Disease: Diagnostic Study

Juan Huang[2*], MD; Lingyu Wu[3,4,5*], MS; Hui Wang[1*], MD; Lin Chen[1], MD; Binbin Hu[1], MD; Fei Zhang[3,4,5], MS; Kang Ren[3,4,5], PhD; Yun Ling[3,4,5], BS; Zhonglue Chen[3,4,5], MS; Wei Huang[1], MD

[1]Department of Neurology, Jiangxi Medical College, Nanchang University, The Second Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China

[2]Department of Neurology, The First affiliated Hospital of Nanhua University, Hengyang, China

[3]GYENNO Science Co., Ltd., Shenzhen, China

[4]HUST-GYENNO CNS Intelligent Digital Medicine Technology Center, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, Hubei, China

[5]Guangdong Provincial Engineering Technology Research Center for Medical Artificial Intelligence in Neurological Disease, Shenzhen, Guangdong, China

[*]these authors contributed equally

Corresponding Author:
Wei Huang, MD
Department of Neurology
Jiangxi Medical College, Nanchang University
The Second Affiliated Hospital of Nanchang University
No. 1, Minde Road
Nanchang, Jiangxi, 330006
China
Phone: 86 13677080198
Email: 13677080198@163.com

Abstract

Background: Patients with Parkinson disease (PD) along with subjective cognitive decline (PD-SCD) are considered an intermediate status between those with normal cognition (PD-NC) and those with mild cognitive impairment (PD-MCI). Wearable digital monitoring technologies and machine learning models offer significant potential for assessing cognitive impairment in patients with PD.

Objective: We aimed to evaluate the utility of wearable technology and machine learning for identifying ordinal cognitive stages (OCS) in PD based on timed up and go (TUG) tests (including single-task TUG [TUGst] and dual-task TUG [TUGdt]).

Methods: Patients with PD along with SCD, MCI, and NC were recruited in a movement disorder clinic. Patients performed TUGst and TUGdt gait trials wearing a motor function and motor symptom quantitative assessment system. In total, 209 kinematic parameters were synthesized for individual TUG to illustrate patients' motion profiles. We constructed dual-task cost parameters (DTC), describing the magnitude of the effect of the cognitive challenge on motion performance. Covariate-adjusted ordered logistic regression was used to compare parameter differences among 3 groups. Multiple machine learning models were used to classify the participants into 3 cognitive impairment levels, with features being selected based on *P* values from intergroup statistical tests. The total population was randomly divided into a training set and an independent validation set in a 7:3 ratio, and 10-fold cross-validation was used in the training set. Furthermore, this study used permutation importance and Shapley Additive Explanations (SHAP) analysis (including summary plots, bar plots, and waterfall plots) to explain the feature importance of the final model.

Results: The study included 65 age-matched patients (PD-NC: PD-SCD: PD-MCI= 14:21:30). Forty-five kinematic parameters were significantly different ($P<.05$) among the 3 groups, distributed across TUGst (n=25), TUGdt (n=12), and DTC (n=8) paradigms. Gait phase analysis revealed 35 parameters from walking phases, 9 from stand-to-sit transitions, and 1 from sit-to-stand transitions. Feature type distribution demonstrated predominance of variability features (n=20), followed by pace (n=12) and axial (n=8) characteristics. TUGdt paradigm analysis revealed pronounced movement differences between PD-MCI and both

XSL•FO
RenderX

PD-NC and PD-SCD groups, particularly in variability, amplitude, pace, and axial domains. Cross-paradigm analysis identified consistent significant differences in specific features. These findings provide objective kinematic biomarkers for early cognitive state identification in Parkinson disease, with TUGdt parameters demonstrating superior discriminative capacity.

Conclusions: This suggests patients with PD-SCD could have early kinetic signs of cognitive impairment, positioning them between PD-NC and PD-MCI, and our multiclass support vector machine classification model with kinematic parameters achieved a recall rate above 0.70 in both training and validation datasets. The feature importance analysis revealed that DTC_Trunk-Right Rotation Max, DTC_Trunk-Max Transverse Angular Velocity, and dTUG_Lumbar-Right Sway Max Std were the most critical features for distinguishing cognitive states, providing scientific evidence for cognitive function screening based on kinematic parameters.

(JMIR Mhealth Uhealth 2026;14:e69273) doi:10.2196/69273

KEYWORDS

Parkinson disease; subjective cognitive decline; mild cognitive impairment; wearable sensor; kinematic parameters; machine learning

Introduction

Parkinson disease (PD) with subjective cognitive decline (SCD) [1] is significantly associated with cognitive impairment, including attention and executive function abilities [2]. At the time of this writing, it is suggested that the cognition performance of patients with PD can be classified into three stages: (1) SCD as the prodromal phase (PD-SCD), (2) followed by MCI (PD-MCI), and (3) finally leading to dementia [3-6]. Longitudinal studies revealed that PD-SCD is more likely to develop into PD-MCI, and PD-SCD may have the potential to predict cognitive decline later [7]. Thus, it is essential to identify individuals with PD-SCD as early as possible.

Brain imaging showed functional and structural correlates between motor control and cognition [8]. Evidence implicates dopamine-dependent basal ganglia dysfunction in PD-related dual-task gait impairments. As the basal ganglia regulate automatic motor control, their degeneration may hinder concurrent motor-cognitive processing during walking [9]. In PD, basal ganglia dysfunction reduces movement automaticity, forcing increased cognitive control of actions. The functional magnetic resonance imaging [10] shows patients with PD overrecruit the premotor and prefrontal cortices during dual-task arm movements versus controls, a pattern also observed in single-task walking, suggesting pervasive compensatory cognitive engagement [11]. Notably, increased quantitative susceptibility mapping values in the cuneus significantly correlated with poorer Montreal Cognitive Assessment (MoCA) performance. The patients with PD-MCI showed elevated quantitative susceptibility mapping values in the cuneus, precuneus, caudate head, fusiform gyrus, and orbitofrontal cortex relative to cognitively normal PD controls [12-14]. Collectively, PD-related dopaminergic degeneration concurrently impairs motor and cognitive circuits, with particular dysfunction in the basal ganglia [15].

Timed Up and Go (TUG) tests (eg, single-task TUG [TUGst] tests and dual-task cognition-gait TUG [TUGdt] tests) have been commonly used to assess motor performance and illustrate the association between motor performance and cognition. The TUGdt test reflects motor-cognitive interactions by challenging the cognitive component of locomotion and thus can provide insight into the mechanisms underlying brain motor control and cognitive performance [16,17]. Gait modifications (ie, dual-task costs [DTCs]) are interpreted to reflect the magnitude of changes in motor performance when cortical attention is involved in gait processes [17].

Wearable digital monitoring technologies combined with machine learning models potentially have considerable value for patients with PD and cognitive impairment [18,9]. Previous studies have demonstrated that PD-MCI exhibits greater postural instability and gait disturbances compared to those without MCI (PD-noMCI), particularly under dual-task conditions [19,20]. Russo et al [21,22], further corroborated the strong association between gait dysfunction and MCI in PD, reporting that patients with PD-MCI present with cognitive decline alongside more severe axial motor symptoms and increased postural flexion. Additionally, meta-analyses by other researchers have underscored the diagnostic value of integrating Kinect-based gait analysis—a critical behavioral marker—with machine learning for identifying cognitive impairment in PD [23,24].

The diagnostic criteria for PD-SCD are mainly based on subjective scales [25-27], and the variety of very diverse instruments used to identify SCD could potentially introduce bias. Thus, research on SCD has predominantly focused on areas such as assessment questionnaires, neuroimaging, and associations with emotion and/or objective cognitive functions [28,29]. There have been few studies combining wearable devices and machine learning methodologies for early screening. This gap is important because larger longitudinal studies have shown that patients with Parkinson disease—subjective cognitive decline have a higher likelihood of progressing to Parkinson disease—mild cognitive impairment within a 3-year follow-up period, highlighting the predictive value of PD-SCD for subsequent cognitive decline [7].

In this study, we hypothesized that patients with PD-SCD were intermediate in cognitive status between patients with PD with normal cognition (PD-NC) and PD-MCI. This study aimed to evaluate the utility of wearable technology and machine learning for identifying ordinal cognitive stages (OCS) in patients with PD based on TUG tests (eg, single-task TUG [TUGst] tests and dual-task cognition-gait TUG [TUGdt] tests), which have been commonly used to assess motor performance and demonstrate the association between motor performance and cognition [17].

XSL•FO
RenderX

To our knowledge, this study is pioneering in its inclusion of individuals with PD-SCD within a framework that uses wearable sensors and machine learning for screening.

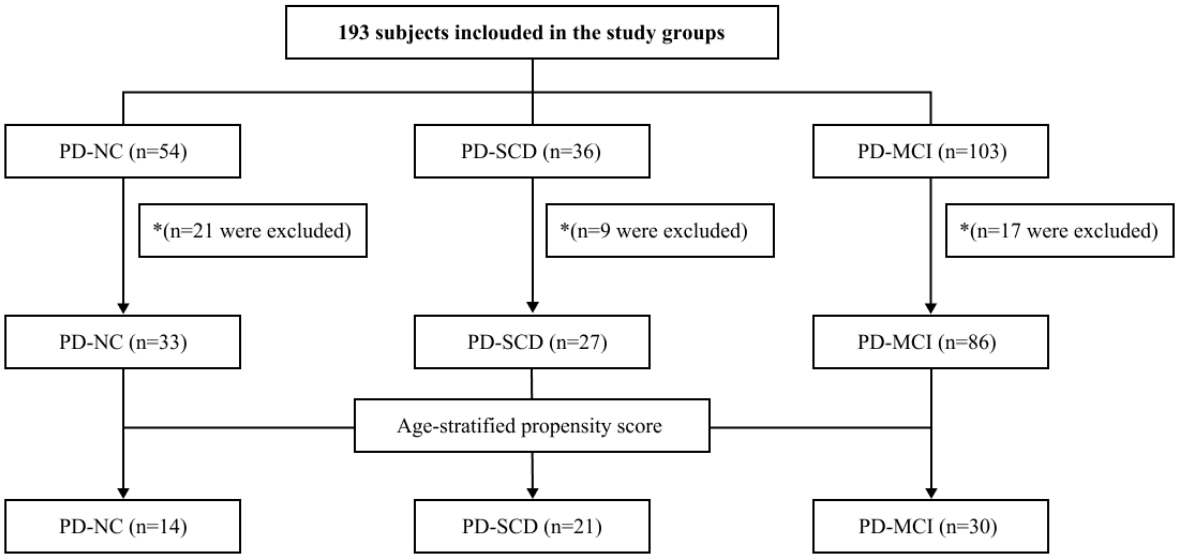
Methods

Recruitment

The data collection process in this study used a rigorous screening and matching strategy (Figure 1). The initial research cohort comprised 193 participants who were preliminarily categorized into 3 groups based on cognitive functional status: PD with normal cognition (PD-NC; n=54), PD with subjective cognitive decline (PD-SCD; n=36), and PD with mild cognitive impairment (PD-MCI; n=103). During the first-stage screening process, some participants were excluded due to incomplete demographic records, partial scale completion, or missing gait measurement data. Specifically, 21 cases were excluded from the PD-NC group, 9 cases from the PD-SCD group, and 17 cases from the PD-MCI group. After preliminary screening, the sample sizes for each group were 33 cases for PD-NC, 27 cases for PD-SCD, and 86 cases for PD-MCI. To control for the influence of age as a significant confounding factor on research outcomes, we used age-stratified propensity score matching for all participants. This matching process ensured balanced age distribution across the 3 groups, effectively controlling for age differences' impact on cognitive function and gait performance. After age matching, the final sample size included in the analysis was 65 participants, consisting of 14 cases in the PD-NC group, 21 cases in the PD-SCD group, and 30 cases in the PD-MCI group. This matching strategy not only enhanced the internal validity of the study but also provided a more reliable data foundation for subsequent machine learning model analysis.

A consecutive inclusion of 65 age-matched patients from a movement disorder clinic was undertaken (PD-NC [n=14]; PD-SCD [n=21]; and PD-MCI [n=30]). Both physical and thorough interviews were done with them. Age, gender, education level, and duration of disease were recorded. All of the patients were diagnosed with idiopathic PD according to the International Parkinson and Movement Disorder Society (MDS) clinical diagnostic criteria. The inclusion criteria were as follows: (1) diagnosis of PD according to the criteria of the MDS, (2) stages 1-3 on the Hoehn-Yahr (H-Y) scale, (3) stable use of antiparkinsonian medication based on medical history, and (4) ability to walk independently for at least 5 minutes. Patients were excluded based on the following criteria: (1) those who diagnosed dementia, (2) a diagnose of secondary PD, (3) medical history indicates having history of brain disorders (eg, cerebral infarction, cerebral hemorrhage, or brain tumor), and (4) those with serious psychiatric diseases (eg, Hamilton Anxiety Scale [HAMA] ≥21 [30], Hamilton Depression Scale [HAMD] ≥17 [31], or obsessive-compulsive disorder). This study was approved by the Research Ethics Committee of the Second Affiliated Hospital of Nanchang University.

Figure 1. The data screening flowchart. PD-NC: Parkinson disease-normal cognition; PD-SCD: Parkinson disease-subjective cognitive decline; *The data are excluded because of the incomplete demographic records, partial scale completion, or the incomplete gait measurement. PD-MCI: Parkinson disease-mild cognitive impairment;.



Diagnostic Criteria for PD-NC, PD-SCD, and PD-MCI

The participants underwent a comprehensive neurocognitive assessment. Motor symptom severity was evaluated using the MDS Unified Parkinson Disease Rating Scale (MDS-UPDRS) and H-Y stage [32]. Cognitive function was assessed using the MoCA. The MoCA covers multiple cognitive domains, including memory, language, complex visuospatial processing, and executive function, and has a maximum score of 30. Emotional features were assessed with the HAMA and HAMD, which has been validated for use with patients with PD.

According to the guidelines for the diagnosis of dementia in PD published by the MDS, a cut-off score of <26 is recommended for PD-MCI; additionally, 1 point is added to the total score (if less than 30) for participants with ≤12 years of

XSL•FO
RenderX

education [33]. In this study, we define normal global cognition as MoCA ≥25/30 for participants with ≤12 years of education and 26/30 for participants with >12 years of education. Participants answered item 1.1 of MDS-UPDRS Part I, and based on the score, the participants were divided into 2 groups: a score of ≥1 indicated those with subjective cognitive complaints and a score of <1 indicated those without subjective cognitive complaints. Thus, we defined PD-SCD as individuals with item 1.1 of MDS-UPDRS Part I ≥1 as well as normal objective global cognition; PD-NC was identified as patients with PD with item 1.1 of MDS-UPDRS Part I <1 and normal objective global cognition [34,35].

Single-Task TUG (TUGst) and Dual-Task TUG (TUGdt) Gait Tests

Participants performed single-task TUG (TUGst) and dual-task TUG (TUGdt) gait trials wearing a Motor Function and Motor Symptom Quantitative Assessment System (MATRIX; GYENNO SCIENCE). The MATRIX consists of 10 inertial sensors (ie, 10 data recording channels) sampling at 100 Hz. Each inertial sensor consists of a (1) triaxial accelerometer with range=±16 g and sensitivity=16,384 LSB/g and a (2) triaxial gyroscope with range=±2000 dps and sensitivity=131 LSB/dps. The chest sensor was placed on the sternum of the chest, and the lumbar sensor was attached to the fifth lumbar vertebra. Two wrist sensors were bilaterally placed on the dorsal side of the wrist. Two thigh sensors were bilaterally placed 7 cm above the knee, while 2 shank sensors were bilaterally placed 7 cm below the knee. Two-foot sensors were bilaterally placed at the instep (dorsal side of the metatarsus) of each foot. All sensors were tightened to designated locations using straps. As a 3-axis accelerometer and a 3-axis gyro are combined in one sensor, each sensor of the MATRIX records 6 separate signals synchronously. Therefore, 60 signals (10 sensors×6 signals) were recorded in computers for feature extraction of a single TUG trial per participant. Please refer to previous research literature for the specific attached position and detailed metadata of sensors [36,37].

During the TUGst test, without the use of any mobility aids, participants stood up from a chair, walked 5 meters in a straight line, turned 180°, walked straight back to the chair, turned another 180°, and sat on the chair at their usual pace. During the TUGdt test, participants performed the usual-pace TUG test while subtracting 7 serially from 100 (ie, 93, 86, 79, and so on). To avoid learning effects and fatigue, participants performed both tests quietly and completed only one trial of each test.

Kinematic Parameters

To illustrate motion profiles of the arms, lumbar spine, trunk, feet, and shanks, 209 kinematic parameters were synthesized for individual TUG trials by gait event recognition based on signal data stored in computers [36,37]. Parameters were listed in Multimedia Appendix 1. This process involved several key aspects: data acquisition and preprocessing using MATRIX with 10 inertial sensors, each recording 6 independent signals (triaxial accelerometer and gyroscope), resulting in 60 raw signals per TUG trial; gait event recognition through algorithmic identification of key gait events, dividing the trial into walking phase, stand-to-sit transition, sit-to-stand transition, and turning phases; parameter extraction and classification based on gait event recognition, extracting 209 kinematic parameters from raw signals covering spatial and temporal characteristics such as velocity, range of motion, stride length, cadence, swing, and stance, as well as asymmetry features including velocity asymmetry and range of motion (RoM) asymmetry; and anatomical plane analysis considering sagittal, coronal, and transverse planes across multiple body segments including arms, lumbar spine, trunk, feet, and shanks. These parameters represent motion asymmetry for bilateral limbs, kinematic variability (SD of parameters), and task-related spatial and temporal characteristics, including velocity (degree per sec), range of motion (RoM, degree), peak velocity (degree per sec), and RoM (degree) of 3 anatomical planes, respectively (parameters the same as lumbar), stride velocity (m/s), cadence (step per min), gait cycle (sec), stride length (cm), step length (cm), swing (%), stance (%), peak angular velocity (degree per sec), RoM (degree), velocity asymmetry (%), RoM asymmetry (%), absolute deviation between limbs, walk duration (sec), and turning duration (sec). According to parameter definitions, 209 parameters were categorized into variability features, amplitude features, pace features, speed features, axial features, and asymmetry features of motion, ensuring comprehensive capture of motion characteristics across all body segments during TUG tasks and providing rich kinematic information for cognitive state classification. Please refer to previous research literature for specific algorithms [37].

Furthermore, DTC parameters, describing the magnitude of the effect of the cognitive challenge on motion performance, were calculated with equation

, in which i refers to each of the 209 parameters.

Data Analysis

Statistics

Quantitative variables following a normal distribution were represented with mean and SD (μ+σ), and statistical differences among multiple groups were tested with one-way ANOVA. Pairwise comparisons between group levels used a student *t* test. Nonnormal quantitative and qualitative or ordinal variables were summarized with median and IQR (lower quartile part or 25th percentile-upper quartile part or 75th percentile) and number of samples and population percentage n (%), respectively. The Kruskal-Wallis rank sum test was used to compare univariate differences on demographic, clinic characteristics, and kinematic parameters among 3 groups. A pairwise Wilcoxon rank sum test was used for 2×2 multiple comparisons between group levels. The ordered logistic regression with covariates (ie, education, age, length of the calf, weight, and BMI) adjusted was used to compare differences on kinematic parameters among multiple groups. Hypothesis testing for categorical variables was performed with the Pearson chi-square test or Fisher exact test. Statistical analyses were conducted using R (version 4.1.0; R Foundation for Statistical Computing) with R Studio (version 1.4.1717; RStudio, PBC).

XSL•FO
RenderX

Feature Selection, Modeling, and Interpretability

Kinematic parameters that statistically significantly different among 3 groups (with covariates adjusted in ordered logistic regression) were selected as candidate predictors, resulting in 45 kinematic features. Then, for candidate predictors, 2×2 multiple comparisons were applied to multiply compare the differences between PD-MCI and PD-SCD, PD-MCI and PD-NC, or PD-NC and PD-SCD, respectively. Parameters that significantly different among one of three comparisons were selected as predictors. To optimize model performance and prevent overfitting given the limited sample size (n=65), we selected the top 10 most discriminative features based on their statistical significance and clinical relevance, resulting in a final set of 10 kinematic features as predictors. This feature selection strategy was chosen to balance model complexity with available data, as including too many features relative to sample size can lead to overfitting and reduced generalization ability, while too few features may not capture sufficient discriminative information for accurate classification.

The total population was randomly divided into a training set and an independent validation set in a 7:3 ratio, and random forest, support vector machine (SVM), and K-nearest neighbors algorithms were used to classify the participants into 3 cognitive impairment levels (ie, PD-MCI, PD-SCD, and PD-NC). The training set was used to train and evaluate model performance, and the independent test set was used to test the final model. Ten-fold cross-validation was used to evaluate the classification performance of these models in the training set. Recall, precision, and accuracy were used to measure the accuracy and calculated as follows:

True positive = the number of cases correctly classified as PD-MCI, PD-SCD, or PD-NC;

False positive = the number of cases incorrectly classified as PD-MCI, PD-SCD, or PD-NC;

True negative = the number of cases correctly classified as PD-MCI, PD-SCD, or PD-NC;

False negative = number of cases incorrectly classified as PD-MCI, PD-SCD, or PD-NC;

(1.1)

(1.2)

(1.3)

Due to the small sample size (n=65) after age-matching, the model performance might be unstable with repeated data splits. Therefore, the additional validation using the bootstrap method to calculate CIs for recall, precision, and accuracy to evaluate the reliability of model performance.

Furthermore, we used the permutation importance by randomly shuffling the values of a feature and observing the resulting degradation in model performance and mean absolute Shapley Additive Explanations (SHAP) value [38] (mean [|SHAP|]) to measure feature importance. SHAP deconstructs a prediction into a sum of contributions from each of the model's input variables, thus explaining how individual predictions are made by a machine learning model. And the waterfall plot based on SHAP was employed to visually display how the SHAP values (evidence) of each feature move the model output from our prior expectation under the background data distribution to the final model prediction given the evidence of all the features. The bottom of a waterfall plot starts as the expected value of the model output, and then each row shows how the positive (red) or negative (blue) contribution of each feature moves the value from the expected model output over the background dataset to the model output for this prediction. Mean (|SHAP|) is, on average, how much each variable impacts the predicted probability of participants being PD-NC, PD-MCI, and PD-SCD. The mean absolute SHAP value is a metric that quantifies the average absolute magnitude of SHAP values across all features in a model. Machine learning was conducted using Python (version 3.8.1627; Python Software Foundation) [39].

Ethical Considerations

This study was approved by the Research Ethics Committee of the Second Affiliated Hospital of Nanchang University, with all participants providing informed consent. Data were anonymized using ID numbers, and follow-up participants received travel compensation.

Results

Demographic and Clinic Characteristics in Groups of PD-NC, PD-SCD, and PD-MCI

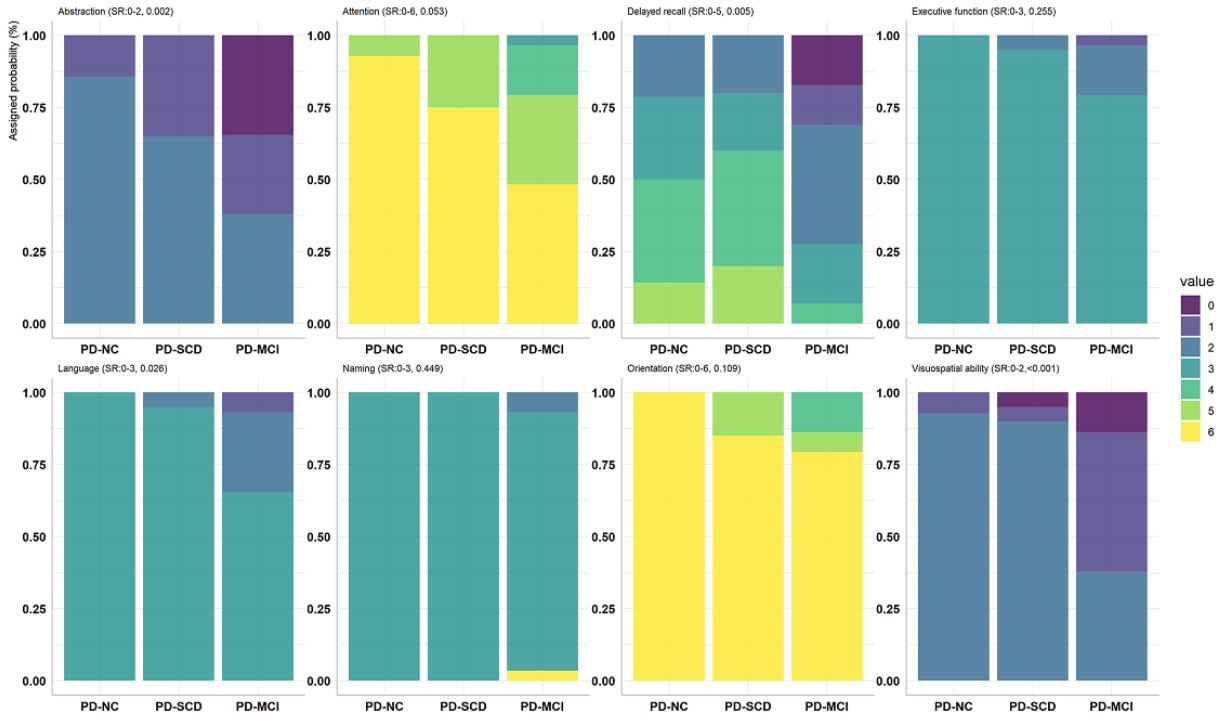
There were no statistically significant differences among the PD-NC, PD-SCD, and PD-MCI groups in terms of demographic characteristics (including age, sex, height, thigh length, weight, education level, disease duration, and H-Y stage) or scores on MDS-UPDRS I, MDS-UPDRS II, MDS-UPDRS III, HAMA, and HAMD (P>.05).

The diagnostic criteria for PD-NC, PD-SCD, and PD-MCI were based on UPDRS1.1 scores and MoCA total scores, which showed statistically significant intergroup differences (P<.001; Multimedia Appendix 2). The MoCA scale and its subdomain scores exhibited different patterns of differences among the 3 groups (Figure 2). In terms of MoCA total scores, the PD-MCI group (23.00, 95% CI 22.00-25.00) was significantly lower than both the PD-NC group (28.00, 95% CI 27.00-29.00) and the PD-SCD group (27.00, 95% CI 27.00-28.00), while no significant difference was observed between PD-NC and PD-SCD groups. Among the MoCA subdomain scores, statistically significant differences were found among the 3 groups in abstraction (P=.002), delayed recall (P=.005), and visuospatial ability (P<.001), with the PD-MCI group showing significantly lower performance in these cognitive domains compared to PD-NC and PD-SCD groups. In attention, the differences among the 3 groups approached statistical significance (P=.05), with the PD-MCI group showing a trend toward decreased attention. However, no statistically significant differences were observed among the 3 groups in executive function (P=.25), language (P=.83), naming (P=.45), and orientation (P=.11; Figure 2). These findings align with the disease progression pattern of PD. As a chronic

XSL•FO
RenderX

neurodegenerative condition, cognitive function may gradually deteriorate with disease progression, particularly in higher-order cognitive functions such as memory, visuospatial ability, and abstract thinking. The significant decline in delayed recall and visuospatial ability observed in patients with PD-MCI reflects the typical characteristics of PD-related cognitive impairment. In contrast, language, naming, and executive functions remain relatively preserved in the early stages, which is consistent with the domain-specific pattern of cognitive impairment in PD.

Figure 2. Distribution of Montreal Cognitive Assessment (MoCA) subdomain scores among Parkinson disease-normal cognition (PD-NC), Parkinson disease-subjective cognitive decline (PD-SCD), and Parkinson disease-mild cognitive impairment (PD-MCI). PD-MCI: Parkinson disease-mild cognitive impairment; PD-NC: Parkinson disease-normal cognition; PD-SCD: Parkinson disease-subjective cognitive decline.



Characteristics of Kinematic Parameters in Groups of PD-NC, PD-SCD, and PD-MCI

Screening procedures identified 45 kinematic parameters meeting the established criteria, distributed across TUGst (n=25), TUGdt (n=12), and DTC (n=8) paradigms. Gait phase analysis revealed 35 parameters from walking phases, 9 from stand-to-sit transitions, and 1 from sit-to-stand transitions. Feature type distribution demonstrated a predominance of variability features (n=20), followed by pace (n=12) and axial (n=8) characteristics. This distribution pattern reflects differential impacts of cognitive decline on gait control mechanisms, with movement precision (variability) and temporal coordination (pace) emerging as critical cognitive function indicators. TUGdt paradigm analysis revealed pronounced movement differences between PD-MCI and both PD-NC and PD-SCD groups, particularly in variability, amplitude, pace, and axial domains. Notable examples include “Shank-Max Sagittal Angular Velocity R Std” (21.12 vs 28.38 vs 21.38) and “Trunk-Left Rotation Max Std” (1.94 vs 1.99 vs 1.33) in variability measures. Amplitude analysis of “Shank-Symbolic Symmetry Index” demonstrated compensatory patterns (14.27 vs 9.75 vs 12.62), suggesting potential compensatory mechanisms during SCD stages. TUGst paradigm differences primarily manifested in pace characteristics, exemplified by “Double Support” (23.16 [4.76] vs 18.55 [3.04] vs 20.87 [3.75]) and “Stance _MAXLR” (62.17 [2.75] vs 60.06 [1.95] vs 61.17 [2.09]). DTC parameter analysis revealed a concentration of differences in variability features, including “Gait Speed R.Std” (–0.97 vs –0.30 vs –0.14) and “Shank-Max Sagittal Angular Velocity R.Std” (–1.67 vs –0.33 vs –0.30), demonstrating cognitive load effects on motor performance. Cross-paradigm analysis identified consistent significant differences in specific features, such as “Step Length_ABSLR” in DTC (–0.36 vs 0.16 vs 0.14), establishing step length asymmetry as a robust cognitive function biomarker. These findings provide objective kinematic biomarkers for early cognitive state identification in PD, with TUGdt parameters demonstrating superior discriminative capacity.

Multimedia Appendix 3 results demonstrate systematic kinematic feature differences across cognitive states, primarily in movement precision (variability) and temporal coordination (pace) domains. Enhanced differences in TUGdt paradigms align with theoretical frameworks predicting motor control degradation under increased cognitive load conditions. Compensatory pattern identification (eg, preserved performance in SCD stages) suggests early-stage compensatory mechanisms maintaining motor function. However, progressive cognitive decline appears to compromise these mechanisms, resulting in significant motor deterioration in MCI stages. These findings

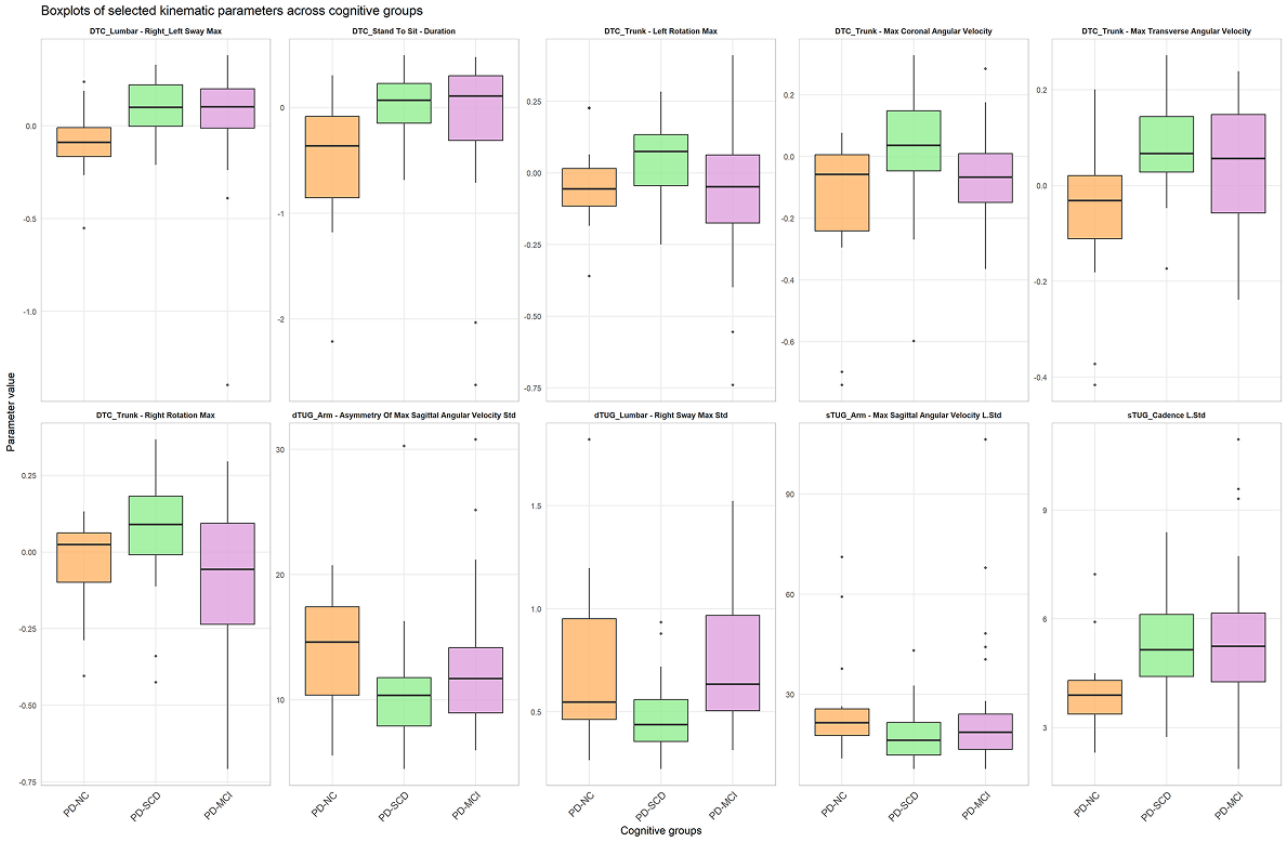
XSL•FO
RenderX

establish objective kinematic biomarkers for early cognitive state identification and monitoring in PD, with TUGdt variability features demonstrating particular clinical translational potential.

Figure 3 presents boxplots illustrating the distribution of 10 selected kinematic parameters across the PD-NC, PD-SCD, and PD-MCI cognitive groups, visually demonstrating the differences among these groups for each parameter. The box in each boxplot represents the 25th percentile (Q1, bottom line of the box), the median (central line of the box), and the 75th percentile (Q3, top line of the box). The bottom and top of the error bars (whiskers) indicate the "Minimum" (Q1-1.5×IQR) and "Maximum" (Q3+1.5×IQR), where IQR is the interquartile range. Blue dots outside this range represent outliers.

As observed from the figure, for several DTC related parameters, such as "DTC_Lumbar-Right_Left Sway Max," "DTC_Stand To Sit-Duration," "DTC_Trunk-Max Transverse Angular Velocity," and "DTC_Trunk-Right Rotation Max," the median values for the PD-NC group are generally lower than those for the PD-SCD and PD-MCI groups, indicating lower motor costs under dual-task conditions in the normal cognition group. This aligns with the theory that impaired cognitive function leads to decreased motor control ability. For instance, "DTC_Stand To Sit-Duration" shows a median of approximately –0.05 for the PD-NC group, while the medians for the PD-SCD and PD-MCI groups are significantly higher than 0.20, suggesting that patients with cognitive impairment are more affected by cognitive challenges during the stand-to-sit transition. In dTUG (dual-task TUG) -related parameters, such as "dTUG_Arm-Asymmetry Of Max Sagittal Angular Velocity SD" and "dTUG_Lumbar-Right Sway Max SD," the median values for the PD-SCD group appear slightly lower than those for the PD-NC and PD-MCI groups. This might reflect a compensatory mechanism or different motor strategies used during the SCD stage. For sTUG (single-task TUG) related parameters, such as "sTUG_Cadence L SD," the median for the PD-NC group (approximately 3.5) is notably lower than that for both PD-SCD and PD-MCI groups (both approximately 5.5), indicating lower gait variability in cognitively normal patients under single-task conditions, whereas patients with cognitive impairment exhibit higher gait variability. These visual differences are consistent with the statistically significant results reported in Multimedia Appendix 3, further supporting the potential of these kinematic parameters as biomarkers for distinguishing different cognitive states.

Figure 3. Multiple-comparison results of represented kinematic parameters between pairs of groups. Multiple-comparison results of represented kinematic parameters between pairs of groups. DTC: dual-task cost; PD-MCI: Parkinson disease-mild cognitive impairment; PD-NC: Parkinson disease-normal cognition; PD-SCD: Parkinson disease-subjective cognitive decline.



Modeling and Interpretability

The best-performing model in the 3-class classification was the SVM model. Its classification accuracies on the training and test sets were 0.8 (0.69-0.86) and 0.6 (0.45-0.73), respectively, indicating a good performance on the training set but reduced performance on the test set (Multimedia Appendix 4). Regarding recognition capability per category on the training set, PD-MCI had a relatively higher recall (0.71), while the PD-NC and PD-SCD categories exhibited relatively higher precision, at 0.90 and 0.86, respectively. On the test set, PD-MCI maintained a relatively higher recall (0.78), while PD-NC and PD-SCD

XSL•FO
RenderX

demonstrated relatively lower precision, at 0.75 and 0.29, respectively. Based on the receiver operating characteristic (ROC) curves (Figure 4), the macro-average ROC AUC on the training set was 0.87, compared to 0.60 on the test set.

Integrating participants' cognitive levels allows for a deeper explanation of the mechanistic differences in classifying MCI, SCD, and NC (Multimedia Appendix 5, Figure 5, and Multimedia Appendix 6). Regarding the distinction between MCI and NC, cognitive level analysis revealed significant differences between the 2 groups across multiple cognitive dimensions: the MoCA total score of the MCI group was significantly lower than that of the NC group (effect size, Cohen d=2.672, indicating a large effect); delayed recall ability was significantly lower in the MCI group than in the NC group (Cohen d=1.915); visuospatial ability was significantly lower in the MCI group than in the NC group (Cohen d=1.354). These significant differences in objective cognitive function provide a cognitive load basis for TUG task features to effectively distinguish MCI from NC. Due to limited cognitive resources, patients with MCI exhibit significant degradation in motor performance during dual-task TUG. This cognitive-motor interaction pattern can be effectively captured by TUG features.

Regarding the distinction between SCD and NC, cognitive level analysis revealed the fundamental cause of confusion. The SCD and NC groups showed no significant difference in MoCA total score (P=.14). Except for abstract ability, no significant differences were found in other cognitive subdomains, with most comparisons showing small-to-medium effect sizes.

This similarity in objective cognitive function explains why TUG features struggle to distinguish SCD from NC: although patients with SCD subjectively perceive cognitive changes, their objective cognitive resources are sufficient, leading to similar performance in dual-task TUG as NC, resulting in TUG features failing to effectively distinguish these 2 groups.

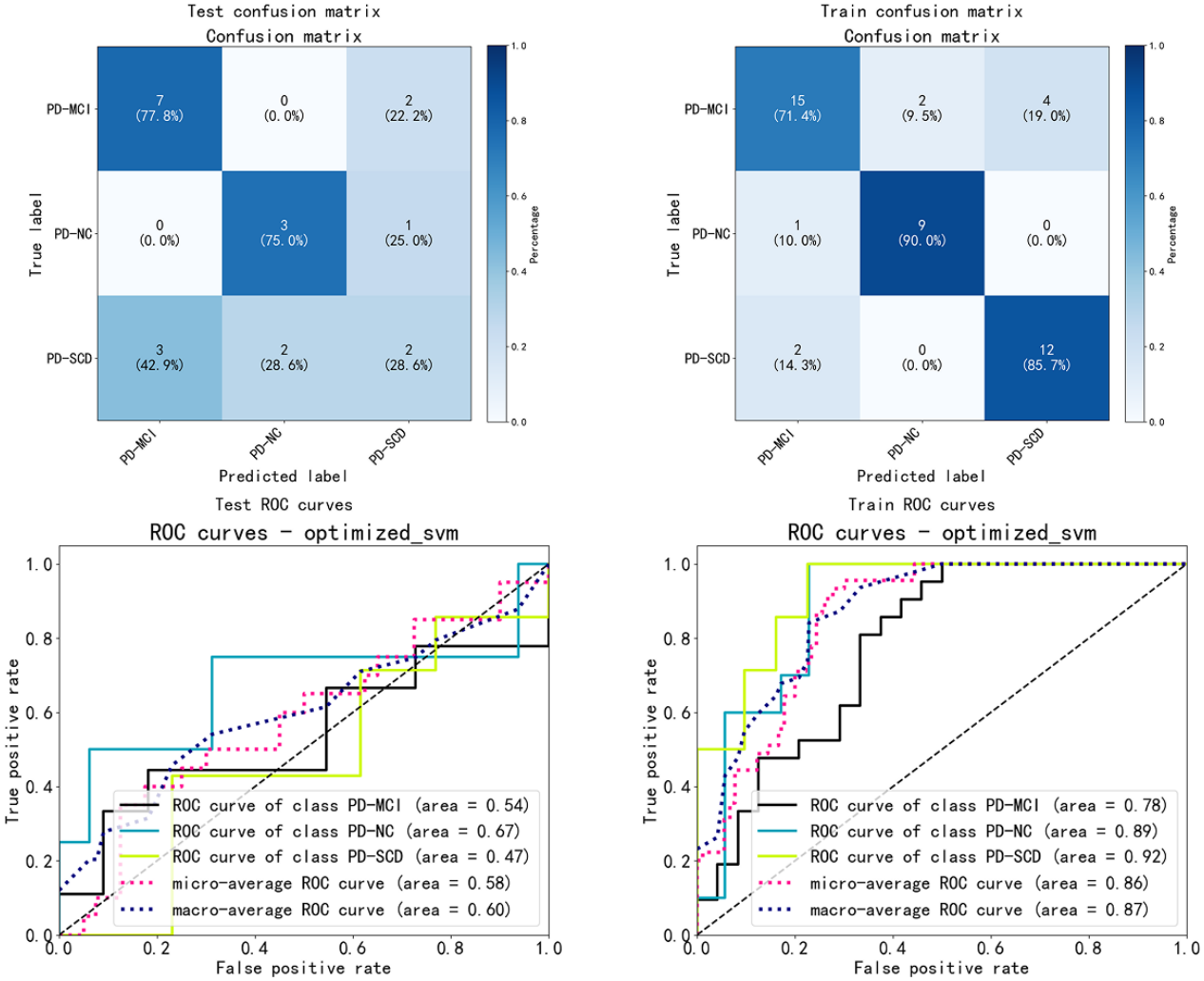
Concerning the distinction between MCI and SCD, cognitive level analysis showed significant differences in objective cognitive function between the 2 groups: the MoCA total score of the MCI group was significantly lower than that of the SCD group (Cohen d=2.301, large effect); delayed recall ability was significantly lower in the MCI group than in the SCD group (Cohen d=1.423); visuospatial ability was significantly lower in the MCI group than in the SCD group (Cohen d=1.528). However, despite significant cognitive differences between MCI and SCD, TUG features still faced difficulties distinguishing them, with 42.9% of SCD samples misclassified as MCI.

Feature importance analysis of cognitive levels further validated the cognitive level analysis results. The MoCA total score had the highest discriminative power ($F_{2,16}$=18.94; P<.001), followed by delayed recall ($F_{2,16}$=11.70; P<.001). This aligns highly with the significant decline in delayed recall observed in the MCI group in the cognitive level analysis (effect size, Cohen d=1.915; Figure 5). The decline in delayed recall ability may directly impact patient performance in dual-task TUG, as this task requires simultaneously processing motor and cognitive demands; the efficiency of cognitive resource allocation directly affects motor performance.

Furthermore, to gain a deeper understanding of the model's decision-making mechanism, we used 2 feature importance analysis methods (Figure 6). Permutation importance analysis evaluated the impact of features on model performance by randomly shuffling feature values, revealing that DTC_Trunk-Right Rotation Max (importance value=0.145) was the most critical feature, followed by DTC_Trunk-Max Transverse Angular Velocity (0.090) and dTUG_Lumbar-Right Sway Max Std (0.080). These findings emphasize the importance of trunk motor control under dual-task conditions for distinguishing cognitive states. SHAP analysis provided 3 different perspectives on feature importance assessment. The SHAP feature importance bar plot showed that sTUG_Cadence L Std (mean [SHAP]=0.23) had the highest average impact, followed by DTC_Trunk-related features. This discrepancy with permutation importance results reflects the different mechanisms of the 2 analysis methods; permutation importance focuses on the impact of feature absence on overall model performance, while SHAP analysis focuses on the direct impact of feature value changes on model output. The SHAP feature importance summary plot further revealed the complex relationship between feature values and their impact on model output. For most DTC features, higher feature values tended to produce larger positive SHAP values, pushing the model output toward higher values, indicating that increased trunk kinematic parameters under dual-task conditions are associated with cognitive decline. This pattern supports the cognitive-motor interaction theory, suggesting that when cognitive resources are limited, individuals exhibit greater motor variability under dual-task conditions. SHAP waterfall plots provided detailed explanations at the individual sample level, demonstrating how each feature drives predictions from baseline values to final outputs. For example, in correctly classified PD-MCI samples, DTC_Trunk-Max Transverse Angular Velocity and DTC_Trunk-Right Rotation Max contributed most strongly to predictions, while correctly classified PD-NC samples exhibited relatively stable dual-task motor features. These individual-level analyses revealed the model's decision-making mechanism in distinguishing different cognitive states, particularly the challenges faced when processing groups with similar cognitive states (such as PD-SCD and PD-NC; Multimedia Appendix 7).

XSL•FO
RenderX

Figure 4. Confusion matrices and receiver operating characteristic (ROC) curves. PD-MCI: Parkinson disease-mild cognitive impairment; PD-NC: Parkinson disease-normal cognition; PD-SCD: Parkinson disease-subjective cognitive decline; ROC: receiver operating characteristic curve.



XSL•FO
RenderX

Figure 5. Comprehensive analysis of cognitive and motor features across Parkinson disease-normal cognition (PD-NC), Parkinson disease-subjective cognitive decline (PD-SCD), and Parkinson disease-mild cognitive impairment (PD-MCI) groups. MoCA: Montreal Cognitive Assessment; PD-MCI: Parkinson disease-mild cognitive impairment; PD-NC: Parkinson disease-normal cognition; PD-SCD: Parkinson disease-subjective cognitive decline.

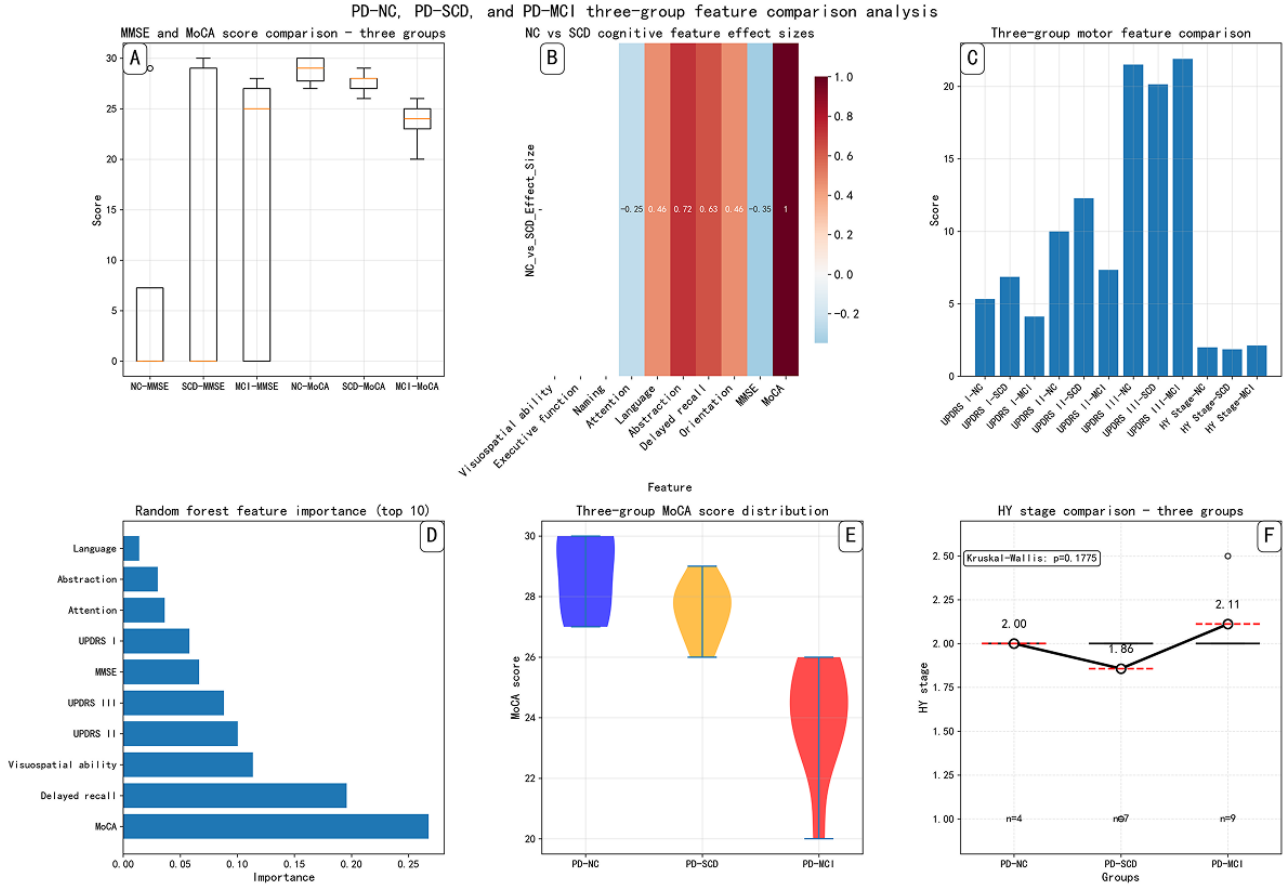
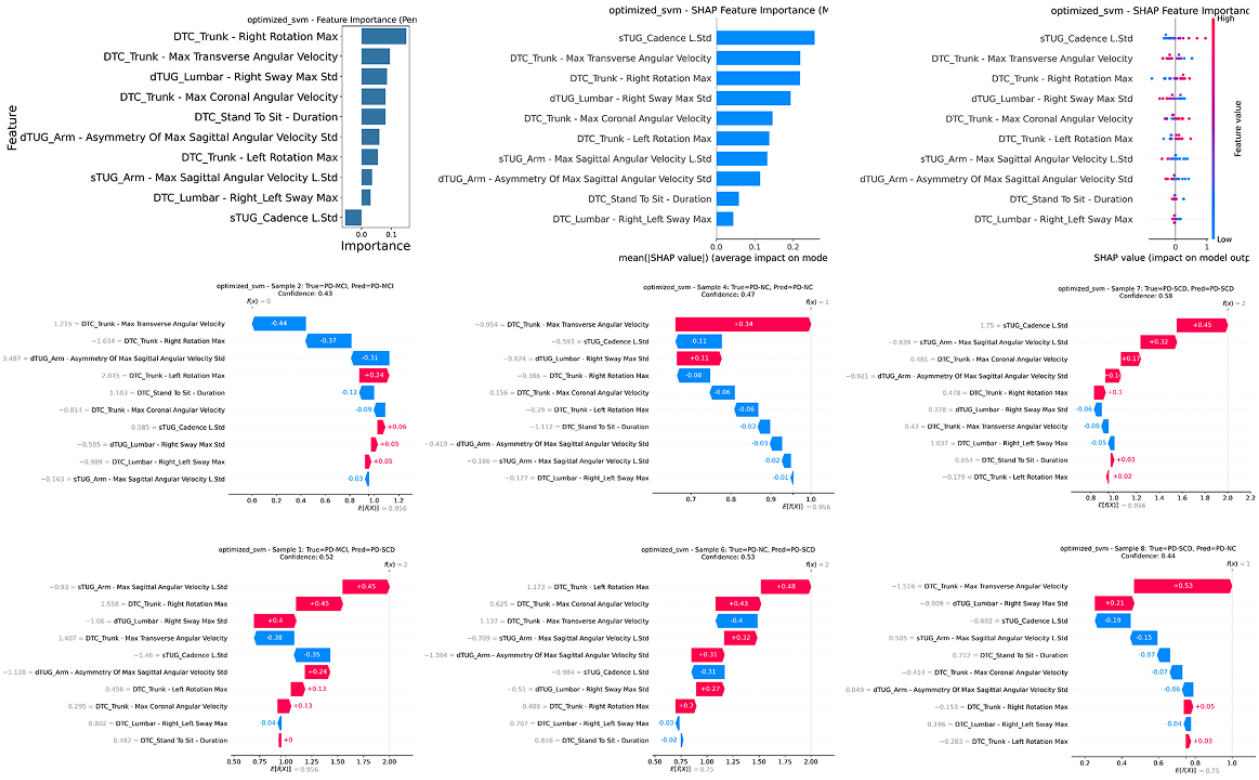


Figure 6. Model feature importance analysis-permutation importance and Shapley Additive Explanations (SHAP) analysis. DTC: dual-task cost; SHAP: Shapley Additive Explanations.



XSL•FO
RenderX

Discussion

Principal Results

In this study, we included 65 patients with PD exhibiting various cognitive states (PD-NC: 14 cases, PD-SCD: 21 cases, and PD-MCI: 30 cases) and explored the method of using kinematic characteristics for the stratified cognitive stage classification of patients with PD by combining wearable sensors with machine learning. We first described the gait characteristics of PD-NC, PD-SCD, and PD-MCI, respectively; second, we explored the factors associated with gait kinematic parameters and cognitive status; next, we analyzed significantly different gait parameters among groups; and finally, we established a model that relies on various TUG task outcomes, which can distinguish individuals with PD-NC, PD-SCD, and PD-MCI. Our results demonstrated that kinematic features are associated with cognitive stages, patients with PD-SCD have underlying kinetic impairments similar to those with PD-MCI, and the diagnostic SVM model showed potential in differentiating cognitive impairments. The SVM model achieved classification accuracies of 0.8 (0.69-0.86) and 0.6 (0.45-0.73) on the training and test sets, respectively, with PD-MCI demonstrating the highest recall rates (0.71 on the training set, 0.78 on the test set) and macro-average ROC AUC values of 0.87 and 0.60 for the training and test sets.

Regarding the classification performance, the model demonstrated effective differentiation between MCI and NC groups due to significant differences in objective cognitive function (MoCA total score effect size, Cohen *d*=2.672), which provided a cognitive load basis for TUG task features to effectively distinguish these groups. However, the model faced significant challenges in distinguishing PD-SCD from both PD-NC and PD-MCI groups. Concerning the distinction between SCD and NC, cognitive level analysis revealed the fundamental cause of confusion. By definition, SCD refers to a subjective perception of cognitive decline, which is not necessarily aligned with the actual cognitive level. For instance, substantial cognitive impairment may occur without being perceived, whereas minor declines may be perceived as significant. Thus, it is difficult to determine which condition reflects a higher level of cognition. Consequently, NC and SCD may not exhibit clear differences in cognitive impairment; the 2 groups showed no significant difference in MoCA total score (*P*=.14). It is correspondingly challenging to distinguish between them based on motor performance. This suggests that SCD may represent a subtype with a higher risk of progressing to MCI, rather than a group that is necessarily closer to MCI in terms of cognitive function.

Regarding the distinction between SCD and MCI, cognitive level analysis showed significant differences in objective cognitive function between the 2 groups (MoCA total score effect size, Cohen *d*=2.301). However, despite significant cognitive differences between MCI and SCD, TUG features still faced difficulties distinguishing them, with 42.9% of SCD samples misclassified as MCI. This seemingly paradoxical result highlights an important mechanism of TUG task features: although MoCA assessment revealed significant cognitive differences between SCD and MCI, this difference primarily manifests in static cognitive tests. In the dynamic context of dual-task TUG, cognitive differences may be insufficient to produce noticeable variations in motor performance. Specifically, while patients with SCD had significantly higher MoCA scores than patients with MCI, their relative cognitive advantage during dual-task TUG might be counterbalanced by other factors, such as attentional allocation strategies, task priority adjustments, or motor control patterns. Furthermore, UPDRS score analysis revealed that the SCD group had the lowest UPDRS III (motor examination) score (20.143), indicating relatively better motor function in patients with SCD. This motor advantage might further compensate for the impact of cognitive differences on TUG performance. This underscores that performance in the TUG task, as a cognitive-motor dual-task paradigm, is not solely influenced by static cognitive function. Instead, it results from the combined effects of dynamic cognitive-motor interactions, attentional allocation, motor control, and other factors. When cognitive differences are counterbalanced by other factors in a dynamic task context, TUG feature-based classification models struggle to accurately distinguish different cognitive states.

Additionally, the limited sample size (n=65) after age-matching may have contributed to model performance instability, with the training set containing only 45 samples and the test set containing only 20 samples, including only 7 SCD samples in the test set, resulting in insufficient sample representativeness that hindered the model's ability to adequately validate SCD classification performance.

Comparison With Prior Work

Demographic Comparation With Other Studies

Consistent with other studies [8,20,40] patients with PD-MCI showed older age, a higher proportion of lower education, worse motor function assessed by MDS-UPDRS-III, and worse performance in global cognition and cognitive domains than the other 2 groups. Additionally, we observed a decrease across the 3 groups in various cognitive domains, including visuospatial ability, executive function, attention, language, abstract thinking, memory, and orientation. Specifically, the PD-NC group had the highest proportion of perfect scores, followed by patients with PD-SCD, with the PD-MCI group having the lowest proportion. These findings here were partially consistent with other studies, which reported that patients with PD-SCD exhibited poorer performance in attention, executive abilities, memory, and visuospatial functions compared to individuals with PD-NC [2,41-43]. Therefore, we believe that the findings of this study have strong generalizability.

Gait Performance Comparison Among Paradigms

In this study, we identified 45 distinct gait parameters that demonstrated significant variance across the 3 participant groups in the TUGst, TUGdt test, and DTC test. These parameters encompassed a range of dynamic attributes, including motion amplitude, pace, speed, and axial alignment, observed during the 5-meter walking, 180° turning, and sit-to-stand phases. Our findings are in line with prior cross-sectional research that has

XSL•FO
RenderX

established robust correlations between gait characteristics—such as pace, rhythm, and asymmetry—and cognitive function, as measured by the TUGst or TUGdt [22,44]. A study by Kim et al [45] revealed that individuals with PD who experience cognitive impairment exhibit a marked reduction in gait speed and stride length when contrasted with those without cognitive impairment. This underscored the adverse effect that cognitive impairment could have on the mobility of patients with PD. Our research observed a decline in step length, stride length, gait speed, and swing among all 3 participant groups (PD-NC, PD-SCD, and PD-MCI) during TUGdt as opposed to the TUGst. Also, kinematic characteristics observed during the TUGdt test exhibited a more pronounced association with the ordinal cognitive stages when compared to other assessments. Specifically, Shank-Max Sagittal Angular Velocity, Trunk-Sway Max SD, Shank-Swing Speed, Shank Forward-Backward Swing Max, gait speed, and straight-walking duration (first segment of the TUGdt test or DTC) were the top 6 features associated with cognitive performance. The result here was adjusted with covariates. This suggests that the introduction of cognitive tasks during the TUGst can disrupt the execution of motor tasks, consequently diminishing the participants' walking capabilities.

Gait Performance Comparison Across Participant Groups

In our analysis, we observed distinct patterns in gait performance across the different participant groups, highlighting the intricate relationship between cognitive function and motor control in PD. Here is a refined presentation of our findings:

1. Gait speed and swing speed decline: a pronounced decreasing trend in gait speed and swing speed was noted as cognitive function worsened across the 3 groups. This aligns with clinical observations where individuals with PD-MCI tend to exhibit poorer gait patterns [20]. Mechanistic insights suggest that the altered neural networks and neurotransmitter deficiencies in PD are linked to both gait disturbances and cognitive impairments [19,46]. The initiation of swinging motion is a complex interplay between passive trunk movement and the precise activation of muscles, regulated by the basal ganglia [47], indicating that cognitive decline may be associated with a slower swing speed.

2. Step and stride length variations: a significant difference in step length and stride length was found between PD-MCI and PD-NC groups, with PD-MCI showing shorter lengths, which is consistent with previous research [48,49]. Interestingly, the PD-SCD group showed longer step and stride lengths, though without statistical significance, possibly indicating early compensatory mechanisms. Neuroimaging studies have consistently demonstrated elevated activation levels in the prefrontal cortex during cognitive tasks in individuals with SCD when compared to those with normal cognition [50,51]. The prefrontal cortex is known for its involvement in attention and executive functions, both of which are closely linked to performance in the TUGdt test [52,53], the enhanced functional connectivity in this region may be an underlying neural mechanism for the decline in TUGdt performance [54]. Additional studies also demonstrated the presence of compensatory neural activity in the early stages of cognitive impairment, and the cognitive strategies can compensate for the decline in gait function [55,56]. Nevertheless, it is crucial to recognize this scarcity of research on the brain function of individuals with PD-SCD. To substantiate our hypothesis and findings, there is a clear need for more extensive, large-scale investigations into brain function within this cohort.

3. Poorest gait performance in patients with PD-MCI: patients with PD-MCI showed significantly poorer gait performance than the other 2 groups. Specifically, in the TUGdt test, both the PD-NC and PD-SCD exhibited shorter walking duration, longer step length and stride length, faster walking speed, swing speed, and angular velocity than the group of PD-MCI, suggesting that individuals with PD-MCI have poorer walking ability, and this was supported by previous studies [8,57,58].

Furthermore, when comparing the PD-MCI and PD-NC groups, we noticed significantly reduced angular velocity during turning and maximum sagittal plane angular velocity during walking in PD-MCI than in PD-NC in TUGdt (Multimedia Appendix 2). PD is characterized by a reduction in the range of motion and muscle strength at the hip, knee, and ankle joints [59,60]. Consequently, those with PD-MCI might struggle with activities like turning, which demand heightened coordination and cognitive capacity [61,62]. When comparing the PD-MCI and PD-SCD groups, we observed that the PD-MCI group had longer walking duration, support phase, stance, shorter step length, stride length, walking speed, arm swing speed, and arm swing amplitude in both TUGst and TUGdt. Our results support earlier research showing that severe cognitive impairments can lead to greater motor skill deficits in gait activities [63,64].

4. Subtle gait changes in PD-SCD: our study suggests that PD-SCD individuals show subtle changes in gait, which is supported by studies demonstrating individuals with poor walking ability during the TUG test have a significantly higher risk of developing MCI compared to those with normal walking ability [48,65].

The control of trunk posture is controlled by the frontal lobe [66,67]. Individuals with PD who have abnormal posture also experience noticeable clinical symptoms of impaired attention and executive ability [68], suggesting that there may be a shared brain activity in the frontal lobe between trunk posture abnormalities and attention and executive ability.

Modeling and Explanation

To automatically obtain patients' cognitive impairment level, 10 kinematic parameters were selected as predictors to construct a multiclass SVM classification model. The recall rate for PD-MCI is above 0.7, while the accuracy for the PD-SCD and PD-NC categories is relatively low. To our knowledge, this pioneering machine learning model, despite its moderate accuracy, represents the first attempt to delineate the cognitive impairment stages in patients with PD, particularly for the PD-SCD stage. Overall, motion speed–related features appeared to contribute most to distinguishing among PD-NC, PD-SCD, and PD-MCI groups. Differences across feature-importance analysis methods reflect the complexity of model decision-making. Permutation importance emphasizes the overall contribution of features, SHAP analysis focuses on the

XSL•FO
RenderX

direct impact of features on model output, and waterfall plots provide detailed explanations at the individual-sample level. These complementary analysis methods collectively revealed the important role of dual-task kinematic features in cognitive state assessment, particularly DTC features related to trunk motor control, providing scientific evidence for cognitive function screening–based on kinematic parameters.

Limitations

There are several limitations in our study. First, the MDS-UPDRS Part I 1.1 score ≥1 evaluating SCD was too simple compared with a comprehensive screening scale. However, SCD identified by MDS-UPDRS Part I 1.1 has been revealed to be associated with the development of PD-MCI, which validates the clinical significance of our measurement tool. Second, we used a short MoCA scale to assess global cognition rather than a comprehensive neuropsychological test focused on each cognitive domain. It is necessary to confirm our results with comprehensive neuropsychological tests. Next, the classification performance of the model needs to be further improved. Furthermore, a larger sample and follow-up multiple-center studies are essential to validate our cross-sectional results.

Conclusions

Our study examines kinematic features to distinguish between 65 patients with PD, including PD-NC, PD-SCD, and PD-MCI. Forty-five gait parameters differed significantly by group in the TUGst, TUGdt tests and DTC, with TUGdt parameters showing stronger cognitive association. Patients with PD-MCI had worse gait performance than patients with PD-NC and PD-SCD. The PD-SCD group showed reduced gait speed and swing speed but longer steps and strides, and our multiclass SVM classification model with kinematic parameters achieved a recall rate above 0.7 in both training and validation datasets. This suggests patients with PD-SCD could have early kinetic signs of cognitive impairment, positioning them between PD-NC and PD-MCI, and kinematic parameters could serve as predictors to distinguish the cognitive stage of patients with PD.

Acknowledgments

We thank all the patients who participated in this study.

Funding

Science and Technology Planning Project of Shenzhen Municipality (KCXFZ20201221173605013).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

All authors declare they have no financial interests. LW, FZ, KR, YL, and ZC are employees of GYENNO Science Co, Ltd. There is no nonfinancial interest in this study.

Multimedia Appendix 1
Supplementary table kinematic features description.
[DOCX File , 30 KB - mhealth_v14i1e69273_app1.docx]

Multimedia Appendix 2
Demographic and clinic characteristics among PD-NC, PD-SCD, and PD-MCI.
[DOCX File , 21 KB - mhealth_v14i1e69273_app2.docx]

Multimedia Appendix 3
Table 2 Intergroup differences of kinematic parameters in PD-NC, PD-SCD, PD-MCI.
[DOCX File , 29 KB - mhealth_v14i1e69273_app3.docx]

Multimedia Appendix 4
Table 3 The accuracy in train set and test set.
[DOCX File , 16 KB - mhealth_v14i1e69273_app4.docx]

Multimedia Appendix 5
Cognitive features analysis - three groups comparison.
[DOCX File , 19 KB - mhealth_v14i1e69273_app5.docx]

XSL•FO
RenderX

Multimedia Appendix 6
The permutation importance of selected features.
[DOCX File , 19 KB - mhealth_v14i1e69273_app6.docx]

Multimedia Appendix 7
Analysis of Cognitive Level Overlap and Classification Confusion.
[DOCX File , 136 KB - mhealth_v14i1e69273_app7.docx]

References

1. Jessen F, Amariglio RE, van Boxtel M, Breteler M, Ceccaldi M, Chételat G, Subjective Cognitive Decline Initiative (SCD-I) Working Group. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. Alzheimers Dement 2014;10(6):844-852 [FREE Full text] [doi: 10.1016/j.jalz.2014.01.001] [Medline: 24798886]
2. Koster DP, Higginson CI, MacDougall EE, Wheelock VL, Sigvardt KA. Subjective cognitive complaints in Parkinson disease without dementia: a preliminary study. Appl Neuropsychol Adult 2015;22(4):287-292. [doi: 10.1080/23279095.2014.925902] [Medline: 25495957]
3. Erro R, Santangelo G, Picillo M, Vitale C, Amboni M, Longo K, et al. Link between non-motor symptoms and cognitive dysfunctions in de novo, drug-naive PD patients. J Neurol 2012;259(9):1808-1813. [doi: 10.1007/s00415-011-6407-0] [Medline: 22310940]
4. Jones JD, Uribe C, Bunch J, Thomas KR. Beyond PD-MCI: objectively defined subtle cognitive decline predicts future cognitive and functional changes. J Neurol 2021;268(1):337-345 [FREE Full text] [doi: 10.1007/s00415-020-10163-4] [Medline: 32804281]
5. Ophey A, Krohm F, Kalbe E, Greuel A, Drzezga A, Tittgemeyer M, et al. Neural correlates and predictors of subjective cognitive decline in patients with Parkinson's disease. Neurol Sci 2022;43(5):3153-3163 [FREE Full text] [doi: 10.1007/s10072-021-05734-w] [Medline: 34820745]
6. Yoo HS, Kwon H, Chung SJ, Sohn YH, Lee J, Lee PH. Neural correlates of self-awareness of cognitive deficits in non-demented patients with Parkinson's disease. Eur J Neurol 2021;28(12):4022-4030. [doi: 10.1111/ene.15095] [Medline: 34478599]
7. Mills KA, Schneider RB, Saint-Hilaire M, Ross GW, Hauser RA, Lang AE, et al. Cognitive impairment in Parkinson's disease: associations between subjective and objective cognitive decline in a large longitudinal study. Parkinsonism Relat Disord 2020;80:127-132. [doi: 10.1016/j.parkreldis.2020.09.028] [Medline: 32987359]
8. Ramírez F, Gutiérrez M. Dual-Task Gait as a Predictive Tool for Cognitive Impairment in Older Adults: A Systematic Review. Front Aging Neurosci 2021;13:769462 [FREE Full text] [doi: 10.3389/fnagi.2021.769462] [Medline: 35002676]
9. Burtscher J, Bourdillon N, Janssen Daalen JM, Patoz A, Bally JF, Kopp M, et al. Movement analysis in the diagnosis and management of Parkinson's disease. Neural Regen Res 2025 Feb 01;20(2):485-486 [FREE Full text] [doi: 10.4103/NRR.NRR-D-24-00207] [Medline: 38819059]
10. Wu T, Hallett M. Neural correlates of dual task performance in patients with Parkinson's disease. J Neurol Neurosurg Psychiatry 2008;79(7):760-766. [doi: 10.1136/jnnp.2007.126599] [Medline: 18006652]
11. Jones D, Rochester L, Birleson A, Hetherington V, Nieuwboer A, Willems A, et al. Everyday walking with Parkinson's disease: understanding personal challenges and strategies. Disabil Rehabil 2008;30(16):1213-1221. [doi: 10.1080/09638280701828955] [Medline: 18608374]
12. Uchida Y, Kan H, Sakurai K, Arai N, Kato D, Kawashima S, et al. Voxel-based quantitative susceptibility mapping in Parkinson's disease with mild cognitive impairment. Mov Disord 2019;34(8):1164-1173. [doi: 10.1002/mds.27717] [Medline: 31091347]
13. Uchida Y, Kan H, Sakurai K, Inui S, Kobayashi S, Akagawa Y, et al. Magnetic susceptibility associates with dopaminergic deficits and cognition in Parkinson's disease. Mov Disord 2020;35(8):1396-1405. [doi: 10.1002/mds.28077] [Medline: 32369660]
14. Shibata H, Uchida Y, Inui S, Kan H, Sakurai K, Oishi N, et al. Machine learning trained with quantitative susceptibility mapping to detect mild cognitive impairment in Parkinson's disease. Parkinsonism Relat Disord 2022;94:104-110. [doi: 10.1016/j.parkreldis.2021.12.004] [Medline: 34906915]
15. Zgaljardic DJ, Borod JC, Foldi NS, Mattis PJ, Gordon MF, Feigin A, et al. An examination of executive dysfunction associated with frontostriatal circuitry in Parkinson's disease. J Clin Exp Neuropsychol 2006;28(7):1127-1144 [FREE Full text] [doi: 10.1080/13803390500246910] [Medline: 16840240]
16. Collyer TA, Murray AM, Woods RL, Storey E, Chong TT, Ryan J, et al. Association of dual decline in cognition and gait speed with risk of dementia in older adults. JAMA Netw Open 2022;5(5):E2214647 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.14647] [Medline: 35639376]
17. Montero-Odasso M, Verghese J, Beauchet O, Hausdorff JM. Gait and cognition: a complementary approach to understanding brain function and the risk of falling. J Am Geriatr Soc 2012;60(11):2127-2136 [FREE Full text] [doi: 10.1111/j.1532-5415.2012.04209.x] [Medline: 23110433]

XSL•FO
RenderX

18. Park J, Lee HJ, Park JS, Kim CH, Jung WJ, Won S, et al. Development of a gait feature-based model for classifying cognitive disorders using a single wearable inertial sensor. Neurology 2023 Jul 04;101(1):e12-e19. [doi: 10.1212/WNL.0000000000207372] [Medline: 37188539]
19. Amboni M, Barone P, Iuppariello L, Lista I, Tranfaglia R, Fasano A, et al. Gait patterns in Parkinsonian patients with or without mild cognitive impairment. Mov Disord 2012;27(12):1536-1543. [doi: 10.1002/mds.25165] [Medline: 23032876]
20. Amboni M, Ricciardi C, Cuoco S, Donisi L, Volzone A, Ricciardelli G, et al. Mild cognitive impairment subtypes are associated with peculiar gait patterns in Parkinson's disease. Front Aging Neurosci 2022;14:781480 [FREE Full text] [doi: 10.3389/fnagi.2022.781480] [Medline: 35299943]
21. Russo M, Amboni MT, Barone P, Pellecchia MT, Romano M, Ricciardi C, et al. Identification of a gait pattern for detecting mild cognitive impairment in Parkinson's disease. Sensors (Basel) 2023;23(4):1985 [FREE Full text] [doi: 10.3390/s23041985] [Medline: 36850582]
22. Morris R, Lord S, Bunce J, Burn D, Rochester L. Gait and cognition: Mapping the global and discrete relationships in ageing and neurodegenerative disease. Neurosci Biobehav Rev 2016 May;64:326-345. [doi: 10.1016/j.neubiorev.2016.02.012] [Medline: 26915926]
23. Altham C, Zhang H, Pereira E. Machine learning for the detection and diagnosis of cognitive impairment in Parkinson's disease: a systematic review. PLoS One 2024;19(5):e0303644 [FREE Full text] [doi: 10.1371/journal.pone.0303644] [Medline: 38753740]
24. Seifallahi M, Galvin JE, Ghoraani B. Detection of mild cognitive impairment using various types of gait tests and machine learning. Front Neurol 2024;15:1354092 [FREE Full text] [doi: 10.3389/fneur.2024.1354092] [Medline: 39055321]
25. Baschi R, Nicoletti A, Restivo V, Recca D, Zappia M, Monastero R. Frequency and correlates of subjective memory complaints in Parkinson's disease with and without mild cognitive impairment: data from the Parkinson's disease cognitive impairment study. J Alzheimers Dis 2018;63(3):1015-1024. [doi: 10.3233/JAD-171172] [Medline: 29710711]
26. Siciliano M, Trojano L, De Micco R, Russo A, Tedeschi G, Tessitore A. Subjective memory decline in Parkinson's disease patients with and without fatigue. Parkinsonism Relat Disord 2020;70:15-19. [doi: 10.1016/j.parkreldis.2019.11.017] [Medline: 31786424]
27. Lehrner J, Moser D, Klug S, Gleiß A, Auff E, Pirker W, et al. Subjective memory complaints, depressive symptoms and cognition in Parkinson's disease patients. Eur J Neurol 2014;21(10):1276-1284. [doi: 10.1111/ene.12470] [Medline: 24909761]
28. Huang J, Yuan X, Chen L, Hu B, Jiang L, Shi T, et al. Subjective cognitive decline in patients with Parkinson's disease: an updated review. Front Aging Neurosci 2023;15:1117068 [FREE Full text] [doi: 10.3389/fnagi.2023.1117068] [Medline: 37304074]
29. Oedekoven C, Egeri L, Jessen F, Wagner M, Dodel R. Subjective cognitive decline in idiopathic Parkinson´s disease: a systematic review. Ageing Res Rev 2022;74:101508. [doi: 10.1016/j.arr.2021.101508] [Medline: 34740867]
30. Hamilton M. The assessment of anxiety states by rating. Br J Med Psychol 1959;32(1):50-55. [doi: 10.1111/j.2044-8341.1959.tb00467.x] [Medline: 13638508]
31. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23(1):56-62 [FREE Full text] [doi: 10.1136/jnnp.23.1.56] [Medline: 14399272]
32. Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, Movement Disorder Society UPDRS Revision Task Force. Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): scale presentation and clinimetric testing results. Mov Disord 2008;23(15):2129-2170. [doi: 10.1002/mds.22340] [Medline: 19025984]
33. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
34. Han LL, Wang L, Xu ZH, Liang XN, Zhang MW, Fan Y, et al. Disease progression in Parkinson's disease patients with subjective cognitive complaint. Ann Clin Transl Neurol 2021;8(10):2096-2104 [FREE Full text] [doi: 10.1002/acn3.51461] [Medline: 34595848]
35. Xiao Y, Ou R, Yang T, Liu K, Wei Q, Hou Y, et al. Different associated factors of subjective cognitive complaints in patients with early- and late-onset Parkinson's disease. Front Neurol 2021;12:749471 [FREE Full text] [doi: 10.3389/fneur.2021.749471] [Medline: 34887827]
36. Lin S, Gao C, Li H, Huang P, Ling Y, Chen Z, et al. Wearable sensor-based gait analysis to discriminate early Parkinson's disease from essential tremor. J Neurol 2023;270(4):2283-2301 [FREE Full text] [doi: 10.1007/s00415-023-11577-6] [Medline: 36725698]
37. Cai G, Shi W, Wang Y, Weng H, Chen L, Yu J, et al. Specific distribution of digital gait biomarkers in Parkinson's disease using body-worn sensors and machine learning. J Gerontol A Biol Sci Med Sci 2023;78(8):1348-1354. [doi: 10.1093/gerona/glad101] [Medline: 37067827]
38. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with eplainable AI for trees. Nat Mach Intell 2020;2(1):56-67 [FREE Full text] [doi: 10.1038/s42256-019-0138-9] [Medline: 32607472]

XSL•FO
RenderX

39. Van RG, Drake FL. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace; 2009.
40. Pieruccini-Faria F, Black SE, Masellis M, Smith EE, Almeida QJ, Li KZH, et al. Gait variability across neurodegenerative and cognitive disorders: results from the Canadian consortium of neurodegeneration in aging (CCNA) and the gait and brain study. Alzheimers Dement 2021;17(8):1317-1328 [FREE Full text] [doi: 10.1002/alz.12298] [Medline: 33590967]
41. Mills KA, Mari Z, Pontone GM, Pantelyat A, Zhang A, Yoritomo N, et al. Cognitive impairment in Parkinson's disease: association between patient-reported and clinically measured outcomes. Parkinsonism Relat Disord 2016;33:107-114 [FREE Full text] [doi: 10.1016/j.parkreldis.2016.09.025] [Medline: 27733275]
42. Nakhla MZ, Holiday KA, Filoteo JV, Zlatar ZZ, Malcarne VL, Lessig S, et al. Informant-reported cognitive decline is associated with objective cognitive performance in Parkinson's disease. J Int Neuropsychol Soc 2021;27(5):439-449. [doi: 10.1017/S1355617720001137] [Medline: 33292885]
43. Hong JY, Lee JE, Sohn YH, Lee PH. Neurocognitive and atrophic patterns in Parkinson's disease based on subjective memory complaints. J Neurol 2012;259(8):1706-1712. [doi: 10.1007/s00415-011-6404-3] [Medline: 22289965]
44. Rajtar-Zembaty A, Sałakowski A, Rajtar-Zembaty J, Starowicz-Filip A, Skalska A. Slow gait as a motor marker of mild cognitive impairment? The relationships between functional mobility and mild cognitive impairment. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2019;26(4):521-530. [doi: 10.1080/13825585.2018.1495690] [Medline: 29985737]
45. Kim SM, Kim DH, Yang Y, Ha SW, Han JH. Gait patterns in Parkinson's disease with or without cognitive impairment. Dement Neurocogn Disord 2018;17(2):57-65 [FREE Full text] [doi: 10.12779/dnd.2018.17.2.57] [Medline: 30906393]
46. Aarsland D, Batzu L, Halliday GM, Geurtsen GJ, Ballard C, Ray Chaudhuri K, et al. Parkinson disease-associated cognitive impairment. Nat Rev Dis Primers 2021;7(1):47. [doi: 10.1038/s41572-021-00280-3] [Medline: 34210995]
47. Meyns P, Bruijn SM, Duysens J. The how and why of arm swing during human walking. Gait Posture 2013;38(4):555-562. [doi: 10.1016/j.gaitpost.2013.02.006] [Medline: 23489950]
48. Montero-Odasso MM, Sarquis-Adamson Y, Speechley M, Borrie MJ, Hachinski VC, Wells J, et al. Association of dual-task gait with incident dementia in mild cognitive impairment: results from the gait and brain study. JAMA Neurol 2017;74(7):857-865 [FREE Full text] [doi: 10.1001/jamaneurol.2017.0643] [Medline: 28505243]
49. Rosano C, Aizenstein H, Brach J, Longenberger A, Studenski S, Newman AB. Special article: gait measures indicate underlying focal gray matter atrophy in the brain of older adults. J Gerontol A Biol Sci Med Sci 2008;63(12):1380-1388 [FREE Full text] [doi: 10.1093/gerona/63.12.1380] [Medline: 19126852]
50. Erk S, Spottke A, Meisen A, Wagner M, Walter H, Jessen F. Evidence of neuronal compensation during episodic memory in subjective memory impairment. Arch Gen Psychiatry 2011;68(8):845-852. [doi: 10.1001/archgenpsychiatry.2011.80] [Medline: 21810648]
51. Rodda JE, Dannhauser TM, Cutinha DJ, Shergill SS, Walker Z. Subjective cognitive impairment: increased prefrontal cortex activation compared to controls during an encoding task. Int J Geriatr Psychiatry 2009;24(8):865-874. [doi: 10.1002/gps.2207] [Medline: 19259973]
52. Filmer HL, Mattingley JB, Dux PE. Improved multitasking following prefrontal tDCS. Cortex 2013;49(10):2845-2852. [doi: 10.1016/j.cortex.2013.08.015] [Medline: 24079917]
53. Johansen-Berg H, Gutman DA, Behrens TEJ, Matthews PM, Rushworth MFS, Katz E, et al. Anatomical connectivity of the subgenual cingulate region targeted with deep brain stimulation for treatment-resistant depression. Cereb Cortex 2008;18(6):1374-1383 [FREE Full text] [doi: 10.1093/cercor/bhm167] [Medline: 17928332]
54. Crockett RA, Hsu CL, Best JR, Liu-Ambrose T. Resting state default mode network connectivity, dual task performance, gait speed, and postural sway in older adults with mild cognitive impairment. Front Aging Neurosci 2017;9:423 [FREE Full text] [doi: 10.3389/fnagi.2017.00423] [Medline: 29311906]
55. Cohen JA, Verghese J, Zwerling JL. Cognition and gait in older people. Maturitas 2016;93:73-77. [doi: 10.1016/j.maturitas.2016.05.005] [Medline: 27240713]
56. Bookheimer SY, Strojwas MH, Cohen MS, Saunders AM, Pericak-Vance MA, Mazziotta JC, et al. Patterns of brain activation in people at risk for Alzheimer's disease. N Engl J Med 2000;343(7):450-456 [FREE Full text] [doi: 10.1056/NEJM200008173430701] [Medline: 10944562]
57. Holtzer R, Verghese J, Xue X, Lipton RB. Cognitive processes related to gait velocity: results from the einstein aging study. Neuropsychology 2006;20(2):215-223. [doi: 10.1037/0894-4105.20.2.215] [Medline: 16594782]
58. Yuan J, Blumen HM, Verghese J, Holtzer R. Functional connectivity associated with gait velocity during walking and walking-while-talking in aging: a resting-state fMRI study. Hum Brain Mapp 2015;36(4):1484-1493 [FREE Full text] [doi: 10.1002/hbm.22717] [Medline: 25504964]
59. Albani G, Cimolin V, Fasano A, Trotti C, Galli M, Mauro A. "Masters and servants" in parkinsonian gait: a three-dimensional analysis of biomechanical changes sensitive to disease progression. Funct Neurol 2014;29(2):99-105 [FREE Full text] [Medline: 25306119]
60. Pistacchi M, Gioulis M, Sanson F, De Giovannini E, Filippi G, Rossetto F, et al. Gait analysis and clinical correlations in early Parkinson's disease. Funct Neurol 2017;32(1):28-34 [FREE Full text] [doi: 10.11138/fneur/2017.32.1.028] [Medline: 28380321]

XSL•FO
RenderX

61. Dibilio V, Stummer C, Drenthen L, Bloem BR, Nonnekes J, Weerdesteyn V. Secondary task performance during challenging walking tasks and freezing episodes in Parkinson's disease. J Neural Transm (Vienna) 2016;123(5):495-501 [FREE Full text] [doi: 10.1007/s00702-016-1516-7] [Medline: 27032775]
62. Rochester L, Galna B, Lord S, Burn D. The nature of dual-task interference during gait in incident Parkinson's disease. Neuroscience 2014;265:83-94. [doi: 10.1016/j.neuroscience.2014.01.041] [Medline: 24508154]
63. Morenilla L, Márquez G, Sánchez JA, Bello O, López-Alonso V, Fernández-Lago H, et al. Postural stability and cognitive performance of subjects with Parkinson's disease during a dual-task in an upright stance. Front Psychol 2020;11:1256 [FREE Full text] [doi: 10.3389/fpsyg.2020.01256] [Medline: 32903649]
64. Raffegeau TE, Krehbiel LM, Kang N, Thijs FJ, Altmann LJP, Cauraugh JH, et al. A meta-analysis: Parkinson's disease and dual-task walking. Parkinsonism Relat Disord 2019;62:28-35 [FREE Full text] [doi: 10.1016/j.parkreldis.2018.12.012] [Medline: 30594454]
65. Rosso AL, Metti AL, Faulkner K, Redfern M, Yaffe K, Launer L, et al. Complex walking tasks and risk for cognitive decline in high functioning older adults. J Alzheimers Dis 2019;71(s1):S65-S73 [FREE Full text] [doi: 10.3233/JAD-181140] [Medline: 30814353]
66. Fredericks D, Norton JC, Atchison C, Schoenhaus R, Pill MW. Parkinson's disease and Parkinson's disease psychosis: a perspective on the challenges, treatments, and economic burden. Am J Manag Care 2017;23(5 Suppl):S83-S92 [FREE Full text] [Medline: 28715903]
67. Varghese JP, Merino DM, Beyer KB, McIlroy WE. Cortical control of anticipatory postural adjustments prior to stepping. Neuroscience 2016;313:99-109. [doi: 10.1016/j.neuroscience.2015.11.032] [Medline: 26608123]
68. Abe K, Uchida Y, Notani M. Camptocormia in Parkinson's disease. Parkinsons Dis 2010;2010:267640 [FREE Full text] [doi: 10.4061/2010/267640] [Medline: 20948888]

Abbreviations

DTC: dual-task cost
dTUG: dual-task Timed Up and Go
HAMA: Hamilton Anxiety Scale
HAMD: Hamilton Depression Scale
H-Y: Hoehn-Yahr
MATRIX: Motor Function and Motor Symptom Quantitative Assessment System
MCI: mild cognitive impairment
MDS: Movement Disorder Society
MDS-UPDRS: Movement Disorder Society-Unified Parkinson Disease Rating Scale
MoCA: Montreal Cognitive Assessment
NC: normal cognition
OCS: ordinal cognitive stages
PD: Parkinson disease
ROC: receiver operating characteristic
RoM: range of motion
SCD: subjective cognitive decline
SHAP: Shapley Additive Explanations
sTUG: single-task Timed Up and Go
SVM: support vector machine
TUG: Timed Up and Go
TUGdt: dual-task cognition-gait timed up and go
TUGst: single-task Timed Up and Go

Edited by L Buis; submitted 26.Nov.2024; peer-reviewed by Y Uchida, VN Bobic, M Russo; comments to author 31.Mar.2025; revised version received 18.Aug.2025; accepted 05.Feb.2026; published 24.Jun.2026.

Please cite as:
Huang J, Wu L, Wang H, Chen L, Hu B, Zhang F, Ren K, Ling Y, Chen Z, Huang W
Gait Analysis for Identifying Normal Cognition, Subjective Cognitive Decline, and Mild Cognitive Impairment in Parkinson Disease: Diagnostic Study
JMIR Mhealth Uhealth 2026;14:e69273
URL: https://mhealth.jmir.org/2026/1/e69273
doi:10.2196/69273
PMID:

XSL•FO
RenderX

©Juan Huang, Lingyu Wu, Hui Wang, Lin Chen, Binbin Hu, Fei Zhang, Kang Ren, Yun Ling, Zhonglue Chen, Wei Huang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 24.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Co-Designed Text Messaging Intervention for Parents of Children With Early Childhood Caries: Randomized Controlled Trial

Rebecca Chen[1,2], BOH, MPH, PhD; Michelle Irving[2], MHS, PhD; Harleen Kumar[3,4], BDSc, DCD; Neeta Prabhu[4,5], BDS, MDSc, MPhil; Bradley Christian[1,6], BDS, MDSc, PhD; Woosung Sohn[1], DDS, DrPH, PhD; Heiko Spallek[4], MSBA, DMD, Dr Med Dent; Simone Marschner[2], PhD; Clara K Chow[2], MBBS, PhD

[1]Population Oral Health, School of Dentistry, Faculty of Medicine and Health, The University of Sydney, Surry Hills, New South Wales, Australia

[2]Westmead Applied Research Centre, Faculty of Medicine and Health, The University of Sydney, Hawkesbury Road, Westmead, New South Wales, Australia

[3]Paediatric Dentistry Department, Sydney Dental Hospital, Oral Health Services, Sydney Local Health District, Surry Hills, New South Wales, Australia

[4]School of Dentistry, Faculty of Medicine and Health, The University of Sydney, Surry Hills, New South Wales, Australia

[5]Paediatric Dentistry Department, Oral Health Services, Western Sydney Local Health District, Westmead, New South Wales, Australia

[6]Australian Centre for Integration of Oral Health, Australian Centre for Integration of Oral Health, School of Nursing & Midwifery, Western Sydney University, Liverpool, NSW, Australia

Corresponding Author:
Rebecca Chen, BOH, MPH, PhD
Population Oral Health, School of Dentistry, Faculty of Medicine and Health, The University of Sydney, Surry Hills, New South Wales, Australia

Abstract

Background: Early childhood caries (ECC) is a prevalent yet preventable disease that significantly impacts young children's health and quality of life. The World Health Organization's *Global Strategy and Action Plan on Oral Health 2023–2030* highlights the potential of mobile health programs to enhance engagement and support sustained oral health behavior change.

Objective: The aim of this study was to assess the effectiveness of the SmilesUp 12-week SMS text messaging–based education and support program to improve parents' management of their children's oral health behaviors compared to standard clinical care.

Methods: We conducted a multisite, assessor-blinded, parallel-design randomized controlled trial among parents of children aged 0 to 6 years who had an ECC diagnosis and attended 1 of 6 public dental clinics in metropolitan, regional, and rural New South Wales, Australia. Participants were randomly assigned to either an intervention group receiving the fully automated 12-week SmilesUp program (3 SMS text messages weekly) or a control group. Both groups continued to receive usual standard dental clinical treatment. The primary behavioral outcome was parent-reported twice-daily toothbrushing with fluoride toothpaste. Secondary outcomes included adherence to oral health–promoting bedtime routines and child consumption of sugar-sweetened foods and beverages. Oral health self-efficacy was measured using 6 behavioral theory–mapped items across oral hygiene, diet, and bedtime routines.

Results: A total of 154 participants were enrolled and randomized (n=78, 50.6% in the control group; n=76, 49.4% in the intervention group) across 6 sites, with 80.5% (124/154) completing the follow-up assessment. High program adherence was observed, with 96.1% (73/76) of intervention participants completing the program. At 12 weeks, a greater proportion of SmilesUp participants reported brushing their children's teeth twice daily (45/61, 73.8% vs 42/63, 66.7%; adjusted odds ratio [aOR] 2.45, 95% CI 0.95-6.74; *P*=.07), though this difference was not statistically significant. Significant improvements were observed in adherence to bedtime routines (54/61, 88.5% vs 45/63, 71.4%; aOR 3.25, 95% CI 1.15-10.18; *P*=.03) as well as oral health knowledge on toothbrushing (46/50, 92.0% vs 41/52, 78.8%; aOR 5.47, 95% CI 1.48-25.19; *P*=.02). No statistically significant differences were found in sugar-sweetened food (10/61, 16.4% vs 4/63, 6.3%; aOR 3.72, 95% CI 1.02-17.14; *P*=.06) or beverage (25/61, 41.0% vs. 27/63, 42.9%; aOR 1.08, 95% CI 0.49-2.30; *P*=.85) consumption.

Conclusions: These findings suggest that co-designed SMS text messaging–based programs show promise and may represent a feasible strategy for oral health behavior change. However, larger-scale and longer-term studies are needed to confirm sustained effectiveness using clinical outcomes. Future research should also explore ways mobile health can be scaled and cost-effectively embedded into comprehensive, interdisciplinary primary models of care.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12623000325606; https://anzctr.org.au/Trial/Registration/TrialReview.aspx?ACTRN=12623000325606

International Registered Report Identifier (IRRID): RR2-10.1371/journal.pone.0310561

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2026;14:e91720) doi:10.2196/91720

KEYWORDS

mobile health; mHealth; dental caries; prevention; early childhood caries; co-design

Introduction

Early childhood caries (ECC) affects more than 600 million children worldwide and remains one of the most common yet preventable childhood diseases [1]. Untreated ECC can lead to significant pain, infection, and impaired quality of life, often resulting in costly dental procedures, including treatments requiring general anesthesia (GA) [2]. In New South Wales (NSW), Australia, 1 in 250 children requires a GA admission for planned dental treatment, with 8.4% (1979/23,559) readmitted within 2 years and a further 9% (1641/18,194) of their siblings requiring similar care [3]. Similar rates in the United Kingdom also indicate that dental caries is the leading cause of hospital admissions for children aged 5 to 9 years (2015-2018) [4]. Preventive behaviors to reduce the risk of ECC include twice-daily toothbrushing with fluoride toothpaste, limiting sugar intake, and avoiding milk or sugar-sweetened beverages at bedtime [5]. While primary care dental clinics play a key role in diagnosis and clinical treatment, parents and caregivers are central to forming and maintaining the preventive behaviors that underpin children's oral health [6]. The World Health Organization's (WHO) 2019 *Ending Childhood Dental Caries* report emphasizes that parental knowledge, motivation, and capability—conceptualized as oral health self-efficacy—are essential in establishing oral health–promoting behaviors [6]. Systematic reviews [5] informing this WHO report and the 2023 update to Australia's national oral health promotion messages [7] reinforce the preventive role that parents play. However, increasing parental knowledge alone is insufficient for sustained behavior change [8]. Traditional chairside advice and demonstration have limited effectiveness because behavior change requires repeated, practical, and accessible prompts that build parental oral health literacy and capability within the home environment [8,9].

Recognizing this gap, the WHO *Global Strategy and Action Plan on Oral Health 2023–2030* [10] advocates for the use of digital innovations, including mobile health (mHealth) [11]. mHealth interventions have shown to improve health behaviors for a broad range of cardiometabolic and chronic health conditions by increasing knowledge, capability, and motivation [12-14]. They are scalable, accessible, and effective across a range of noncommunicable diseases [12-14], yet their use in early childhood oral health remains emerging [15]. Existing programs for oral health largely target older adults and adolescents [16]. They narrowly focus on toothbrushing [17], are not of high quality [18,19], and are rarely grounded in behavioral theory [20,21].

To address these limitations, our research group undertook a comprehensive co-design process with parents to develop SmilesUp, a theory-informed mHealth intervention tailored to parental needs, preferences, and identified barriers and enablers [22]. The SmilesUp SMS text messaging program aims to enhance parental oral health self-efficacy by providing tailored information, including practical video demonstrations, and timely prompts delivered directly to parents within the home environment, where oral health habits are formed [22]. This randomized controlled trial (RCT) aimed to evaluate the impact of the SmilesUp SMS text messaging program on parental oral health behaviors, knowledge, oral health self-efficacy (capability, opportunity, and motivation), and overall engagement, addressing a critical gap in mHealth strategies to prevent ECC.

Methods

Study Design

This single-blinded parallel RCT (1:1 allocation ratio) was assessor blinded and carried out within NSW public dental clinics. The protocol has been previously published [23], and a concise summary of our methods is provided below. The trial adhered to the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth; Checklist 1) reporting guidelines.

Participant Eligibility and Recruitment

Parents of children diagnosed with ECC who were either awaiting or undergoing dental treatment in the NSW public dental service in Australia were approached face-to-face for the study at 6 community dental health services, including 2 tertiary pediatric sites across metropolitan, rural, and regional areas. These areas collectively serve a population of approximately 4 million people and cover an area of 155,000 square kilometers. All clinics primarily offer a fee-free dental service for families unable to afford private care, with clinical and preventive care provided based on existing NSW public dental policies [24]. Prior to enrollment, all participants were given access to a participant information and consent form via eConsent through the REDCap platform (Multimedia Appendix 1). Parents of children who met all the following criteria were eligible to participate: (1) a clinical diagnosis of ECC, (2) age of 0 to 6 years inclusive, and (3) awaiting (or having attended) a public NSW dental service for dental caries treatment. Only 1 parent and 1 child per family were recruited. Parents were excluded if they (1) were already participating in an SMS text messaging–based study on lifestyle risk factors for health, (2) did not have proficient English-language skills to provide written informed consent, or (3) were unable to comply with study requirements. Parents of these children must also have basic computer and smartphone literacy and own an operational mobile phone capable of sending and receiving SMS text messages and accessing the internet.

Procedures

Following completion of eConsent (Multimedia Appendix 1) through the REDCap web link provided via their smartphones, general demographic data, including the parent's gender,

XSL•FO
RenderX

ethnicity, country of birth, age category (18-20, 20-25, 26-35, and 36-45 years), and level of education, were collected via self-report. Information on the child's age, number and order of siblings, Socio-Economic Indexes for Areas (SEIFA) rank [25,25] and whether the child had a planned dental procedure requiring GA was also collected from the parent. At the 12-week follow-up, data were collected through an online questionnaire available via the REDCap link that was provided as an SMS text message to the parent at the conclusion of the program. If the questionnaire was not completed, follow-up phone calls were made at the end of the 12 weeks.

Randomization and Blinding

Participants were randomly assigned to either the intervention or control group at a 1:1 allocation ratio stratified by recruitment site (community dental clinics vs tertiary pediatric settings) and modality of care (dental procedures requiring GA vs dental procedures not requiring GA). Randomization was generated and accessed only by the study statistician using the *randomizeR* library in R (version 4.3.2; R Foundation for Statistical Computing) and was based on permuted block sizes (2,4,6) [26]. Clinicians recruited parent participants who self-enrolled in the study, and participants were randomly assigned to the study groups via computer generation and notified via SMS text message. All researchers were blinded as the TextCare system automatically implemented the SMS text messaging program in the intervention arm. Due to the nature of the intervention, participants were not blinded, but dental care providers and those analyzing the data were.

Intervention

The SmilesUp intervention is a 12-week SMS text messaging program co-designed with parents and dental practitioners to target key ECC-related oral health behaviors, including twice-daily toothbrushing with fluoride toothpaste, bedtime routines, and consumption of sugar-sweetened foods and drinks [22]. The intervention aligns with WHO guidance to improve oral health through behavior change and is informed by the behavior change wheel (BCW) theory [27]. The BCW synthesizes 19 major behavior change theories to identify the capabilities, opportunities, and motivations required and the most appropriate intervention functions to support effective behavior change [27]. Messages aimed to improve oral health literacy and self-efficacy and provide brief reminders, reassurance, and links to smartphone-accessible video resources (Multimedia Appendix 2).

Participants directly received 36 semipersonalized messages (3 per week) to their mobile phones, delivered through the TextCare platform via an automated and preprogrammed schedule shaped by consumer engagement. Personalization included the children's names and parents' choice of morning or evening SMS text message delivery. The program was primarily one way but allowed for two-way communication, monitored asynchronously by a clinician researcher. Participants requiring clinical assistance received automated messages that contained the contact details for their nearest NSW public dental clinic or emergency services if the dental concern was urgent. Feedback on the program, including notifications on whether messages were received and read, was monitored asynchronously by a clinician researcher, with additional feedback collected at the 12-week follow-up.

The SmilesUp program was implemented alongside standard clinical dental care. The control group received standard care only, consisting of clinical assessment, treatment, and preventative advice provided by a dental practitioner. After completing the study period, parents in the control group were offered access to the SmilesUp intervention.

Outcomes

The primary outcome was achieving the target behavior of brushing the children's teeth twice a day with fluoride toothpaste. This item was analyzed dichotomously in response to an existing question used by NSW Health to evaluate other oral health programs: "How often do you brush your child's teeth with toothpaste?" [28]. The primary outcome was achieved if participants responded "Twice a day" or "More than twice a day"; all other responses were classified as not meeting the outcome. Secondary outcomes were oral health–promoting behaviors, including use of fluoride toothpaste, reduced sugar-sweetened beverage consumption, reduced sugar-sweetened food consumption, bedtime routines related to the frequency and timing of milk consumption at night, and water consumption. Outcomes were ascertained through parental self-report using a standardized NSW Health questionnaire [28]. All definitions for primary and secondary behavioral outcomes are described in an a priori statistical analysis plan, which can be found in Multimedia Appendix 1.

Additional oral health self-efficacy outcomes were assessed using an adapted questionnaire based on the capability, opportunity, and motivation–behavior (COM-B) framework [29]. These questions were optional to reduce survey fatigue at enrollment. Existing oral health self-efficacy tools primarily focus on oral hygiene and are not typically grounded in behavioral theory [21]. In contrast, the questionnaire used in this study was aligned with the COM-B framework, which also underpinned the co-design of the SmilesUp intervention [22]. This approach was selected because it could be applied to a broad range of oral health–promoting behaviors addressed by the intervention (oral hygiene, diet, and bedtime routines). Accordingly, the questionnaire applied the COM-B framework to each behavior and addressed (1) physical capability, (2) psychological capability (ie, knowledge capability), (3) physical opportunity, (4) social opportunity, (5) reflective motivation, and (6) automatic motivation (Multimedia Appendix 1). Acceptance of the program was also assessed via the number of participants who completed the SmilesUp SMS text messaging intervention [23].

Sample Size Calculation

A sample size of 150 participants provided 80% statistical power to detect a 25% increase in the target toothbrushing behavior in the intervention group from 53% to 78% compared to the control group at 12 weeks considering a 5% significance level and a 25% dropout rate. A cross-sectional study of a similar Australian population accessing public dental services found that 53% of children aged 3 to 8 years brushed their teeth twice a day with toothpaste [30].

XSL•FO
RenderX

Statistical Analysis

Our statistical analysis followed an intention-to-treat principle and a prespecified statistical analysis plan (Multimedia Appendix 1). To assess the effectiveness of the intervention, a logistic regression model adjusting for whether parents performed toothbrushing for their children twice a day with toothpaste at baseline and the stratification variables was conducted, and adjusted odds ratios (aORs) with 95% CIs were reported. Similarly, all secondary behavioral outcomes were analyzed using the same logistic regression model. A statistical significance level of .05 was used. An interaction analysis of treatment and prespecified covariates (planned dental procedure requiring GA vs non-GA and tertiary vs nontertiary sites) was conducted to explore the heterogeneity of our results. No multiplicity adjustments were performed for secondary outcomes as they were exploratory by nature. If missingness was greater than 5%, missingness at random was assessed using 2-tailed *t* tests or chi-square tests for the relevant covariate. Additionally, a sensitivity analysis was conducted using multivariate imputation by chained equations with 50 imputations. No interim analyses were needed in alignment with stopping guidelines. All analyses were undertaken using R.

Ethical Considerations

This study was conducted in accordance with the principles of the Declaration of Helsinki. The protocol was reviewed and approved by the Western Sydney Local Health District Human Research Ethics Committee (2022/ETH01920). All participants took part voluntarily without compensation and provided written informed consent prior to enrollment, with the ability to opt out at any time. To ensure privacy and confidentiality, all collected data were deidentified. Additionally, the trial was prospectively registered with the Australian New Zealand Clinical Trials Registry (ACTRN12623000325606).

Results

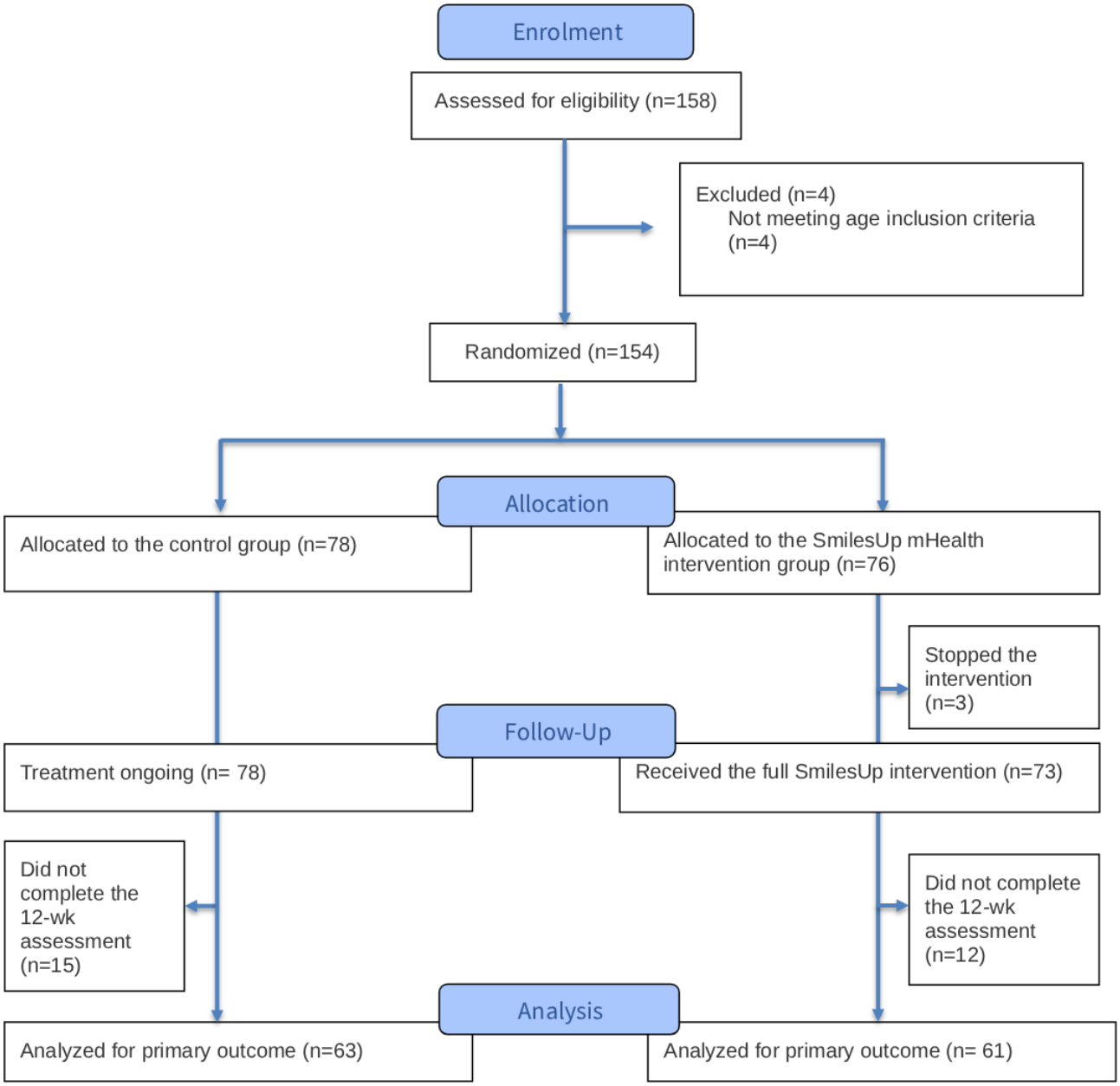
Study Participants

Of the 154 participants who were recruited and randomized in this study between August 24, 2023, and May 10, 2024, a total of 76 (49.4%) in the intervention group and 78 (50.6%) in the control group (n=124, 80.5%) completed follow-up data collection of primary and secondary behavioral outcomes at the end of the 12-week intervention (Figure 1). Additional outcomes focused on oral health literacy and self-efficacy were collected from 66.2% (n=102) of the participants at 12 weeks. Recruitment was completed once the anticipated sample size was reached.

Most parent participants (144/154, 93.5%) were above 25 years of age, 74.0% (114/154) were female, 47.4% (73/154) had completed a university degree, 57.1% (88/154) were born overseas, 18.8% (29/154) used general health apps, and 5.2% (8/154) used oral health apps (Table 1). The mean age of the children was 4.2 (SD 1.7) years, 31.8% (49/154) had a planned dental procedure requiring GA, 83.1% (128/154) had a sibling, 21.4% (33/154) had 4 or more siblings, 33.8% (52/154) had a younger sibling in the family, and 37.7% (58/154) were from lower–socioeconomic status areas as determined via postcodes (Socio-Economic Indexes for Areas 1-3; Table 2).

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram.



XSL•FO
RenderX

Table . Baseline demographic characteristics of parents by treatment group (N=154).

Characteristics	Control (n=78), n (%)	Intervention (n=76), n (%)	Overall, n (%)
Parent's gender			
Men	20 (25.6)	20 (26.3)	40 (26.0)
Women	58 (74.4)	56 (73.7)	114 (74.0)
Parent's age (y)			
18-25	3 (3.8)	3 (3.9)	6 (3.9)
26-35	31 (39.7)	25 (32.9)	56 (36.4)
36-45	35 (44.9)	43 (56.6)	78 (50.6)
≥46	6 (7.7)	4 (5.3)	10 (6.5)
Prefer not to answer	3 (3.8)	1 (1.3)	4 (2.6)
Parent's level of education			
University degree	34 (43.6)	39 (51.3)	73 (47.4)
Nonuniversity degree	39 (50)	35 (46.1)	74 (48.1)
Prefer not to answer	5 (6.4)	2 (2.6)	7 (4.5)
Country of birth			
Australia	32 (41.0)	34 (44.7)	66 (42.9)
Overseas	46 (59.0)	41 (53.9)	88 (57.1)
Do you use oral health apps?	5 (6.4)	3 (3.9)	8 (5.2)
Do you use other health apps?	16 (20.5)	13 (17.1)	29 (18.8)

XSL•FO
RenderX

Table . Baseline demographic characteristics of the children by treatment group (N=154).

Characteristics	Control (n=78)	Intervention (n=76)	Overall
Child's age (y), mean (SD)	4.3 (1.5)	4.1 (1.8)	4.2 (1.7)
Is your child receiving or have they received care with a general anesthetic?, n (%)	25 (32.1)	24 (31.6)	49 (31.8)
Total number of children in your family, n (%)			
1	13 (16.7)	13 (17.1)	26 (16.9)
2	36 (46.2)	29 (38.2)	65 (42.2)
3	11 (14.1)	19 (25)	30 (19.5)
4 or more	18 (23.1)	15 (19.7)	33 (21.4)
Do you have any other children who are younger than the child receiving dental treatment? ("yes" responses), n (%)[a]	25 (32.1)	27 (35.5)	52 (33.8)
Do you have any other children who are older than the child receiving dental treatment? ("yes" responses), n (%)	36 (46.2)	28 (36.8)	64 (41.6)
I have children that are both older and younger ("yes" responses), n (%)	9 (11.5)	9 (11.8)	18 (11.7)
Not applicable to me (I have one child) ("yes" responses), n (%)	13 (16.7)	16 (21.1)	29 (18.8)
SEIFA[b] (IRSAD[c] quintile), n (%)			
1 (most disadvantaged)	16 (20.5)	9 (11.8)	25 (16.2)
2	8 (10.3)	12 (15.8)	20 (13.0)
3	7 (9.0)	6 (7.9)	13 (8.4)
4	18 (23.1)	15 (19.7)	33 (21.4)
5 (most advantaged)	29 (37.2)	34 (44.7)	63 (40.9)

[a]Responses for the four questions on siblings do not sum to 100% because multiple responses were allowed.

[b]SEIFA: Socio-Economic Indexes for Areas.

[c]IRSAD: Index of Relative Socio-economic Advantage and Disadvantage.

Primary Outcomes

Of the 76 parents who were randomized to the intervention group, 73 (96.1%) completed the 12-week SmilesUp program. At 12 weeks, 73.8% (45/61) of participants who received the SmilesUp SMS text messaging program reported brushing their children's teeth twice a day compared to 66.7% (42/63) in the control group (aOR 2.45, 95% CI 0.95-6.74; *P*=.07; Table 3). The primary outcome was missing for 19.5% (30/154) and similar in both groups. Examination for bias indicated missingness to be at random (Multimedia Appendix 3). The sensitivity analysis using multiple imputation by chained equations analysis demonstrated that the study groups yielded similar results (aOR 1.93, 95% CI 0.75-4.96; *P*=.17) and was not statistically significant.

Our interaction analysis showed similar intervention effects among families who had or had not undergone a planned dental procedure requiring GA (*P*=.95) and whether families were recruited from tertiary or nontertiary settings (*P*=.12).

XSL•FO
RenderX

Table . Primary and secondary behavioral outcomes.

	Control, n (%)		Intervention, n (%)		aOR[a] (95% CI)	*P* value[b]
	Baseline (n=78)	12 wk (n=63)	Baseline (n=76)	12 wk (n=61)		
Primary outcome						
Met recommended toothbrushing requirements	49 (62.8)	42 (66.7)	43 (56.6)	45 (73.8)	2.45 (0.95-6.74)	.07
Secondary outcomes						
Met recommended fluoride toothpaste use	68 (87.2)	62 (98.4)	62 (81.6)	57 (93.4)	0.27 (0.01-2.18)	.27
Met oral health–promoting bedtime routines	58 (74.4)	45 (71.4)	60 (78.9)	54 (88.5)	3.25 (1.15-10.18)	.03
Met recommended sugar-sweetened beverage consumption	27 (34.6)	27 (42.9)	20 (26.3)	25 (41.0)	1.08 (0.49-2.39)	.85
Met recommended sugar-sweetened food consumption	8 (10.3)	4 (6.3)	7 (9.2)	10 (16.4)	3.72 (1.02-17.14)	.06
Met recommended water consumption	74 (94.9)	57 (90.5)	73 (96.1)	60 (98.4)	6.53 (0.90-139.76)	.11

[a]aOR: adjusted odds ratio.

[b]*P* value using logistic regression adjusting for the corresponding outcome at baseline and stratification variables (general anesthesia [GA] vs non-GA and tertiary vs nontertiary sites).

Secondary Outcomes

For secondary outcomes, intervention participants reported better oral health bedtime routines (54/61, 88.5% vs, 45/63, 71.4%; aOR 3.25, 95% CI 1.15-10.18; *P*=.03). Sugar-sweetened food consumption showed a consistent direction of effect, although it was not statistically significant (aOR 3.72, 95% CI 1.02-17.14; *P*=.06). There was no statistically significant difference between the groups for sugar-sweetened beverage consumption (aOR 1.08, 95% CI 0.49-2.39; *P*=.85), use of fluoride toothpaste (aOR 0.27, 95% CI 0.01-2.18; *P*=.27), or adequate water consumption (aOR 6.53, 95% CI 0.90-139.76; *P*=.11; Table 3).

Oral Health Self-Efficacy Outcomes

Additional outcomes of oral health self-efficacy were reported in a subset of participants in the intervention group (50/102, 49.0%) and the control group (52/102, 51.0%). They were overall suggestive of a pattern of improved self-efficacy, although few measures met statistical significance. A greater percentage of intervention participants reported improved knowledge on how to brush their children's teeth (aOR 5.47, 95% CI 1.48-25.19; *P*=.02) and how to maintain a healthy bedtime routine (aOR 5.76, 95% CI 1.38-32.16; *P*=.03). There was no difference in knowledge on providing healthy drink and snack options (aOR 2.7, 95% CI 0.60-14.97; *P*=.21) or automatic motivation regarding oral health–promoting bedtime routines (aOR 3.01, 95% CI 0.96-10.54; *P*=.07; Table 4).

XSL•FO
RenderX

Table . Oral health self-efficacy outcomes.

Statement (capability, opportunity, and motivation–behavior domain related to oral health self-efficacy)	...brush my child's teeth (oral hygiene)		...provide healthy drink and snack options (diet)		...support oral health–promoting bedtime routines	
	aOR[a] (95% CI)	*P* value[b]	aOR (95% CI)	*P* value	aOR (95% CI)	*P* value
I can physically... (physical capability)	1.13 (0.38-3.33)	.82	0.830 (0.17-3.91)	.81	5.19 (1.16-31.33)	.045
I have all the knowledge to... (psychological capability)	5.47 (1.48-25.19)	.02	2.7 (0.60-14.97)	.21	5.76 (1.38-32.16)	.03
I have all the things I need to... (physical opportunity)	0.98 (0.08-12.04)	.99	1.66 (0.36-8.32)	.51	4.42 (0.92-28.38)	.08
People around me think it is normal for me to... (social opportunity)	2.34 (0.55-11.78)	.26	1.56 (0.40-6.54)	.53	2.22 (0.64-8.53)	.22
I am motivated to... (reflective motivation)	0.87 (0.24-3.01)	.83	1.77 (0.15-22.59)	.64	2.03 (0.40-12.20)	.40
This is something I do automatically without thinking about whether to... (automatic motivation)	1.50 (0.60-3.84)	.39	2.43 (0.70-9.54)	.18	3.01 (0.96-10.54)	.07

[a]aOR: adjusted odds ratio.

[b]*P* value using logistic regression adjusting for the corresponding outcome at baseline and stratification variables (general anesthesia [GA] vs non-GA and tertiary vs nontertiary sites).

Discussion

Principal Findings

The SmilesUp trial is the first SMS text messaging randomized controlled clinical trial specific for ECC home management, using the BCW framework [27] tested in an Australian public dental setting. Most participants completed the SmilesUp program, and while the primary outcome (twice-daily brushing) was not statistically significantly different, some secondary outcomes suggested improvements in oral health behaviors and oral health self-efficacy. Together, these findings suggest that the program may have potential benfit if implemented at scale; however, larger and longer confirmatory trials are needed to determine effectiveness over time.

In our sample, 18.8% (29/154) of parents reported using health apps, but only 5.2% (8/154) used oral health apps. This highlights both a gap and an opportunity to design high-quality mHealth oral health programs that respond to growing demand for digital preventive health care [31-33]. No interaction effects were observed among families whose children had a planned dental procedure requiring GA and those children who received care in community dental clinics without a GA. These findings indicate an opportunity to test the effect for families using community dental clinics with and without a GA in a future full-scale RCT and ensure that the study is adequately powered to enable subgroup analyses. Additionally, the intervention was highly acceptable, with a 96.1% (73/76) completion rate, indicating strong feasibility for a future confirmatory trial. Given the growing consumer preference for and acceptance of digital health support, particularly in the post–COVID-19 context [3,34], value-based health care systems should prioritize the integration of effective prevention-focused digital tools into routine health care delivery [31,34].

Comparison With Prior Studies

This trial extends existing evidence in the field by demonstrating that mHealth interventions showed a consistent direction of effect to influence behaviors alongside improved parental knowledge on ECC risk factors. Our findings on parental knowledge are consistent with those of the systematic review by Ajay et al [35] on mHealth interventions targeting ECC risk factors, which showed improvements in oral health knowledge only. Importantly, this study moves beyond knowledge outcomes by providing evidence that mHealth interventions have the potential to affect behavioral outcomes, including oral

XSL•FO
RenderX

health–promoting bedtime routines. Existing peer-reviewed literature demonstrates that changes in these behavioral outcomes are associated with favorable ECC clinical indicators, including the decayed, missing, filled teeth index; caries incidence; and the need for repeat invasive dental treatment [1,5,6].

Strengths

The strengths of this study include its randomized clinical trial design; its evaluation of a broad set of oral health–promoting behavioral outcomes; and its implementation across a diverse participant population in metropolitan, regional, and rural NSW. The findings demonstrate the feasibility of an SMS text messaging–based program irrespective of socioeconomic status, mobile phone type, and geographical location, as well as the low cost of implementation for health services [36]. Importantly, the intervention was co-designed with parents and pre-evaluated for acceptability and usefulness [22], which likely contributed to the high engagement and completion rates. Evaluating the intervention alongside standard clinical care demonstrated the added value of the intervention and highlights the importance of multilevel approaches to support parents in maintaining sustained preventative oral health–promoting behaviors [37].

Limitations

This study has limitations, including a follow-up rate of 80.5% (124/154), although this was balanced between the intervention and control groups [38]. The use of self-reported outcomes may have introduced social desirability bias compared to the use of clinical measures such as plaque scores; however, the randomized design ensured that this was balanced between the groups. The proportion of families reporting the primary outcome (twice-daily toothbrushing) at baseline was higher than anticipated (92/154, 59.7% compared to the anticipated 53%). Despite achieving close to the anticipated target of 78%, the study had limited ability to detect a statistically significant difference. This smaller-than-expected difference reduced the contrast between groups and, consequently, diminished the detectable effect size. Given the suggestions of benefit and low cost of the intervention, larger RCTs powered to detect smaller effect sizes would be justifiable. Recruitment within public dental clinics introduced selection bias; however, the inclusion of participants from diverse settings in metropolitan, regional, and rural areas may also support representativeness for the health system. The high level of missingness could have influenced the results; however, this was accounted for as participant characteristics were similar between groups (Multimedia Appendix 3). In fact, the high completion rate of participants previously not engaged in the target behavior of toothbrushing suggests that the intervention is particularly relevant for those most in need, highlighting potential benefits in health equity (Multimedia Appendix 3). The short follow-up period (12 weeks) limits conclusions regarding long-term behavioral maintenance and related clinical outcomes, highlighting the need for a full-scale RCT [38]. These future studies should explore whether mHealth initiatives such as SmilesUp can be effective in preventing longer-term clinical and service use related to ECC. Recruiting from a larger sample may also enhance the generalizability of these findings.

Future Directions

Clinically, the results of the SmilesUp trial indicate that digital interventions have the potential to improve oral health behaviors and parental self-efficacy, which may translate into reduced ECC outcomes. Compared to traditional chairside advice, mHealth offers a less resource intensive approach that directly supports families in their home settings. The limited effect on some behaviors, including sugar consumption, also highlights the need for multipronged and interdisciplinary collaborative mHealth approaches to address common risk factors [39,40]. From a policy perspective, future efforts should focus on integrating mHealth into value-based, interdisciplinary models of care that recognize preventative behavior change in home settings as a critical component of health care delivery, complementing clinical dental treatment to achieve long-term oral and general health [38].

A process evaluation is also required to inform implementation strategies for scale-up, including culturally appropriate adaptation of the programs and alignment with other interdisciplinary family-focused preventive mHealth programs delivered in primary care settings [41]. Interoperability with electronic oral health records could enhance uptake and data collection to allow health systems to prospectively and longitudinally measure the program's long-term effects, including clinical dental caries measures (decayed, missing, and filled teeth index), service use, and hospital admissions related to ECC [33,38]. Leveraging technologies including AI may further support tailored content delivery [42] and continuous improvement within learning health systems [43]. Future research should examine the impact of scaling mHealth programs that leverage AI on sustained oral health behavior change, long-term effectiveness, and cost-effectiveness.

Conclusions

The SmilesUp mHealth program signals benefits and the potential of mHealth interventions to enhance oral health behaviors and underscores the need for larger studies with longer follow-up to determine the sustainability of behavior change and its translation into clinical outcomes. Future research should also explore optimal strategies for embedding mHealth innovations within broader primary prevention and interdisciplinary primary care frameworks and assess long-term preventative effectiveness and cost-effectiveness when delivered at scale.

Acknowledgments

The authors wish to acknowledge clinical leads from local health district (LHD) sites Lara Mayze, Trolisa Knudsen (Southern and Murrumbidgee LHDs), Grace Wong (Northern Sydney LHD), and Tanya Kumar (Nepean Blue Mountains LHD).

XSL•FO
RenderX

Generative AI Use

Generative AI was not used to develop the manuscript.

Funding

This study was funded by the Western Sydney Local Health District. The funders did not have any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Trial protocol and statistical analysis plan including the participant information and consent form.
[PDF File, 1292 KB - mhealth_v14i1e91720_app1.pdf]

Multimedia Appendix 2
Sample messages from the SmilesUp program.
[PDF File, 50 KB - mhealth_v14i1e91720_app2.pdf]

Multimedia Appendix 3
Missingness data.
[DOCX File, 30 KB - mhealth_v14i1e91720_app3.docx]

Checklist 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File, 1154 KB - mhealth_v14i1e91720_app4.pdf]

References

1. Early childhood caries: IAPD Bangkok Declaration. Int J Paediatr Dent 2019 May;29(3):384-386. [doi: 10.1111/ipd.12490] [Medline: 31099129]
2. Tsai C, Li A, Brown S, et al. Early childhood caries sequelae and relapse rates in an Australian public dental hospital. Int J Paediatr Dent 2023 Jan;33(1):1-11. [doi: 10.1111/ipd.12969] [Medline: 35484872]
3. Chen R, Schneuer FJ, Irving MJ, et al. Socio-demographic and familial factors associated with hospital admissions and repeat admission for dental caries in early childhood: a population-based study. Community Dent Oral Epidemiol 2022 Dec;50(6):539-547. [doi: 10.1111/cdoe.12708] [Medline: 34837420]
4. The Lancet Child Adolescent Health. Oral health: oft overlooked. Lancet Child Adolesc Health 2019 Oct;3(10):663. [doi: 10.1016/S2352-4642(19)30275-5] [Medline: 31526741]
5. Moynihan P, Tanner LM, Holmes RD, et al. Systematic review of evidence pertaining to factors that modify risk of early childhood caries. JDR Clin Trans Res 2019 Jul;4(3):202-216. [doi: 10.1177/2380084418824262] [Medline: 30931717]
6. Ending childhood dental caries: WHO implementation manual. World Health Organization. 2020. URL: https://www.who.int/publications/i/item/ending-childhood-dental-caries-who-implementation-manual [accessed 2026-09-21]
7. Welti R, Chinotti M, Walsh O, et al. Oral health messages for Australia: a national consensus statement. Aust Dent J 2023 Dec;68(4):247-254. [doi: 10.1111/adj.12973] [Medline: 37665214]
8. Faghihian R, Faghihian E, Kazemi A, Tarrahi MJ, Zakizade M. Impact of motivational interviewing on early childhood caries: a systematic review and meta-analysis. J Am Dent Assoc 2020 Sep;151(9):650-659. [doi: 10.1016/j.adaj.2020.06.003] [Medline: 32854867]
9. Faisal MR, Mishu MP, Jahangir F, et al. The effectiveness of behaviour change interventions delivered by non-dental health workers in promoting children's oral health: a systematic review and meta-analysis. PLoS One 2022;17(1):e0262118. [doi: 10.1371/journal.pone.0262118] [Medline: 35015771]
10. Global strategy and action plan on oral health 2023–2030. World Health Organization. 2024. URL: https://www.who.int/publications/i/item/9789240090538 [accessed 2026-09-21]
11. Mobile technologies for oral health: an implementation guide. World Health Organization. 2021. URL: https://www.who.int/publications/i/item/9789240035225 [accessed 2026-09-22]
12. Al Mahmud A, Joachim S, Jayaraman PP, et al. Digital health interventions to support chronic disease management: systematic scoping review. JMIR Mhealth Uhealth 2026 Jan 14;14:e63742. [doi: 10.2196/63742] [Medline: 41533959]

XSL•FO
RenderX

13. Zhang Y, Ngai FW, Yang Q, Xie YJ. Effectiveness of digital health interventions on sedentary behavior among patients with chronic diseases: systematic review and meta-analysis. JMIR Mhealth Uhealth 2025 Jun 24;13:e59943. [doi: 10.2196/59943] [Medline: 40554804]
14. Chow CK, Klimis H, Thiagalingam A, et al. Text messages to improve medication adherence and secondary prevention after acute coronary syndrome: the TEXTMEDS randomized clinical trial. Circulation 2022 May 10;145(19):1443-1455. [doi: 10.1161/CIRCULATIONAHA.121.056161] [Medline: 35533220]
15. Recommendations on digital interventions for health system strengthening: WHO guideline. World Health Organization. 2019. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2026-09-22]
16. Carrouel F, Lan R, Saliasi I, et al. Effectiveness of different methods of interdental hygiene in daily practice among young adults: protocol for a randomized, single-blind controlled trial. JMIR Res Protoc 2025 Dec 4;14:e85154. [doi: 10.2196/85154] [Medline: 41343847]
17. Toniazzo MP, Nodari D, Muniz FW, Weidlich P. Effect of mHealth in improving oral hygiene: a systematic review with meta-analysis. J Clin Periodontol 2019 Mar;46(3):297-309. [doi: 10.1111/jcpe.13083] [Medline: 30761580]
18. Hyzy M, Bond R, Mulvenna M, et al. Quality of digital health interventions across different health care domains: secondary data analysis study. JMIR Mhealth Uhealth 2023 Nov 23;11:e47043. [doi: 10.2196/47043] [Medline: 37995121]
19. Chen R, Santo K, Wong G, et al. Mobile apps for dental caries prevention: systematic search and quality evaluation. JMIR Mhealth Uhealth 2021 Jan 13;9(1):e19958. [doi: 10.2196/19958] [Medline: 33439141]
20. Murariu A, Bobu L, Gele u GL, et al. The impact of mobile applications on improving oral hygiene knowledge and skills of adolescents: a scoping review. J Clin Med 2025 Apr 23;14(9):2907. [doi: 10.3390/jcm14092907] [Medline: 40363939]
21. Buchanan H, Newton JT, Baker SR, Asimakopoulou K. Adopting the COM-B model and TDF framework in oral and dental research: a narrative review. Community Dent Oral Epidemiol 2021 Oct;49(5):385-393. [doi: 10.1111/cdoe.12677] [Medline: 34270110]
22. Chen R, Irving MJ, Tsai C, et al. Co-design and development of the SmilesUp text messaging intervention using behavioral theory to support parents of children with early childhood caries: mixed methods study. JMIR Pediatr Parent 2025 Nov 18;8:e72107. [doi: 10.2196/72107] [Medline: 41252553]
23. Chen R, Irving M, Christian B, et al. SmilesUp text message intervention for early childhood dental caries prevention: a protocol for a randomised controlled trial. PLoS One 2024;19(9):e0310561. [doi: 10.1371/journal.pone.0310561] [Medline: 39348422]
24. Priority oral health program (POHP) and waiting list management. NSW Health. 2017. URL: https://www1.health.nsw.gov.au/pds/Pages/doc.aspx?dn=PD2017_023 [accessed 2023-08-01]
25. 2039.0 - Information paper: an introduction to Socio-Economic Indexes for Areas (SEIFA), 2006. : Australian Bureau of Statistics; 2008 URL: https://www.abs.gov.au/ausstats/abs@.nsf/mf/2039.0 [accessed 2026-09-22]
26. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2023. URL: https://www.r-project.org/ [accessed 2026-09-22]
27. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42. [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
28. Balasubramanian M, Ghanbarzadegan A, Sohn W, et al. Primary school mobile dental program in New South Wales, Australia: protocol for the evaluation of a state government oral health initiative. BMC Public Health 2023 Feb 20;23(1):363. [doi: 10.1186/s12889-023-15241-6] [Medline: 36803579]
29. Keyworth C, Epton T, Goldthorpe J, Calam R, Armitage CJ. Acceptability, reliability, and validity of a brief measure of capabilities, opportunities, and motivations ("COM-B"). Br J Health Psychol 2020 Sep;25(3):474-501. [doi: 10.1111/bjhp.12417] [Medline: 32314500]
30. Tadakamadla SK, Rathore V, Mitchell AE, Kaul A, Morawska A. Child- and family-level factors associated with toothbrushing frequency in a sample of Australian children. Int J Paediatr Dent 2022 Sep;32(5):639-648. [doi: 10.1111/ipd.12942] [Medline: 34811821]
31. Väyrynen E, Hakola S, Keski-Salmi A, Jämsä H, Vainionpää R, Karki S. The use of patient-oriented mobile phone apps in oral health: scoping review. JMIR Mhealth Uhealth 2023 Sep 6;11:e46143. [doi: 10.2196/46143] [Medline: 37672331]
32. Poirier B, Jensen E, Sethi S. The evolution of the teledentistry landscape in Australia: a scoping review. Aust J Rural Health 2022 Aug;30(4):434-441. [doi: 10.1111/ajr.12874] [Medline: 35567780]
33. Bamgboje-Ayodele A, Paonne A, Evans E, et al. Facilitators and barriers to implementing a patient portal at a dental hospital from the implementers' perspectives: qualitative study. J Med Internet Res 2025 Nov 18;27:e78979. [doi: 10.2196/78979] [Medline: 41251416]
34. Fernández CE, Maturana CA, Coloma SI, Carrasco-Labra A, Giacaman RA. Teledentistry and mHealth for promotion and prevention of oral health: a systematic review and meta-analysis. J Dent Res 2021 Aug;100(9):914-927. [doi: 10.1177/00220345211003828] [Medline: 33769123]
35. Ajay K, Azevedo LB, Haste A, et al. App-based oral health promotion interventions on modifiable risk factors associated with early childhood caries: a systematic review. Front Oral Health 2023;4:1125070. [doi: 10.3389/froh.2023.1125070] [Medline: 36968137]

XSL•FO
RenderX

36. Dobson R, Whittaker R, Abroms LC, et al. Don't forget the humble text message: 25 years of text messaging in health. J Med Internet Res 2024 Dec 17;26:e59888. [doi: 10.2196/59888] [Medline: 39689299]
37. Nelson S, Albert JM, Selvaraj D, et al. Multilevel interventions and dental attendance in pediatric primary care: a cluster randomized clinical trial. JAMA Netw Open 2024 Jul 1;7(7):e2418217. [doi: 10.1001/jamanetworkopen.2024.18217] [Medline: 38980678]
38. Casamassimo PS, Nowak AJ. Benefits of early dental care now hard to refute. JAMA Netw Open 2019 Mar 1;2(3):e190673. [doi: 10.1001/jamanetworkopen.2019.0673] [Medline: 30874774]
39. Della Vecchia C, Leroy T, Bauquier C, et al. Willingness of French general practitioners to prescribe mHealth apps and devices: quantitative study. JMIR Mhealth Uhealth 2022 Feb 11;10(2):e28372. [doi: 10.2196/28372] [Medline: 35147508]
40. Sutherland R, Brown A, Nathan N, et al. A multicomponent mHealth-based intervention (SWAP IT) to decrease the consumption of discretionary foods packed in school lunchboxes: type I effectiveness-implementation hybrid cluster randomized controlled trial. J Med Internet Res 2021 Jun 24;23(6):e25256. [doi: 10.2196/25256] [Medline: 34185013]
41. Christian B, George A, Veginadu P, et al. Strategies to integrate oral health into primary care: a systematic review. BMJ Open 2023 Jul 5;13(7):e070622. [doi: 10.1136/bmjopen-2022-070622] [Medline: 37407034]
42. Trivedi R, Shaw T, Sheahen B, Chow CK, Laranjo L. Patient perspectives on conversational artificial intelligence for atrial fibrillation self-management: qualitative analysis. J Med Internet Res 2025 Mar 12;27:e64325. [doi: 10.2196/64325] [Medline: 40073398]
43. Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. J Dent Res 2020 Jul;99(7):769-774. [doi: 10.1177/0022034520915714] [Medline: 32315260]

Abbreviations

aOR: adjusted odds ratio
BCW: behavior change wheel
COM-B: capability, opportunity, and motivation–behavior
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
ECC: early childhood caries
GA: general anesthesia
mHealth: mobile health
NSW: New South Wales
RCT: randomized controlled trial
WHO: World Health Organization

Edited by L Buis; submitted 19.Jan.2026; peer-reviewed by P Dhawan, R Musa; revised version received 09.Jul.2026; accepted 10.Aug.2026; published 03.Nov.2026.

Please cite as:
Chen R, Irving M, Kumar H, Prabhu N, Christian B, Sohn W, Spallek H, Marschner S, Chow CK
Co-Designed Text Messaging Intervention for Parents of Children With Early Childhood Caries: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e91720
URL: https://mhealth.jmir.org/2026/1/e91720
doi:10.2196/91720

© Rebecca Chen, Michelle Irving, Harleen Kumar, Neeta Prabhu, Bradley Christian, Woosung Sohn, Heiko Spallek, Simone Marschner, Clara K Chow. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 3.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Time to Cancer Screening Completion Following an SMS Reminder Intervention in a Large Federally Qualified Health Center Network: Secondary Data Analysis

Tonghui Xu[1], PhD; Summer Chavez[1,2], DO; Ben King[1,2], PhD; Gonzalo Ramirez-Pulido[1], BS; Chinedum O Ojinnaka[3], PhD; Daniel Osayi[2], BS; Valery Kounga[2], PharmD; Omolola E Adepoju[1,2], PhD

[1]Humana Integrated Health Systems Sciences Institute, University of Houston, 5055 Medical Cir, Houston, TX, United States

[2]Tilman J. Fertitta Family College of MedicineX, University of Houston, Houston, TX, United States

[3]Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, IN, United States

Corresponding Author:
Omolola E Adepoju, PhD
Humana Integrated Health Systems Sciences Institute, University of Houston, 5055 Medical Cir, Houston, TX, United States

Abstract

Background: Delays in completing cancer screening diminish the preventive benefits of early detection, particularly among women receiving care in Federally Qualified Health Centers (FQHCs). Although many patients receive SMS reminders and complete screening, less is known about how quickly they complete testing or which patient-level and structural factors are associated with delays.

Objective: This study examined factors associated with time to cancer screening completion among women aged 50 years or older who received SMS reminders and completed screening across a large FQHC network in Texas. The study also compared time to completion across three cancer screening tests: human papillomavirus (HPV) or the Papanicolaou test (hereinafter “Pap test”), mammography, and the fecal immunochemical test (FIT) or Cologuard screening.

Methods: We conducted a secondary data analysis using electronic health record (EHR) data from a 56-clinic FQHC network in Texas. The initial cohort included 1803 women aged 50 years or older who (1) were overdue for HPV or Pap testing, mammography, or FIT or Cologuard screening, and (2) received at least three SMS reminders. Of those, 551 completed the screening and constituted the analytic cohort for this study. The outcome was the number of days from the initial SMS reminder to documented completion of the overdue screening test in the EHR. Kaplan-Meier methods were used to estimate time to completion by screening modality. A multivariable Cox proportional hazards model assessed associations of screening modality, sociodemographic, and clinical characteristics, and self-reported health-related social needs with the rate of screening completion.

Results: Overall, 40.8% (n=212) of patients overdue for HPV or Pap testing completed their screening, while 21.1% (n=138) of those overdue for mammography and 32.1% (n=201) of those overdue for FIT or Cologuard screening completed their respective screening. Median time to completion was 72.5 (95% CI 64 - 86) days for HPV or Pap screening and 52.0 days for both mammography (95% CI 43 - 64) and FIT or Cologuard screening (95% CI 52 - 53). In the adjusted model, screening completion was faster for FIT or Cologuard screening (hazard ratio [HR] 1.65, 95% CI 1.34 - 2.05) and mammography (HR 1.41, 95% CI 1.11 - 1.78) than for HPV or Pap screening. Patient-reported transportation limitation was associated with slower screening completion (HR 0.74, 95% CI 0.55 - 0.99).

Conclusions: These findings demonstrate meaningful variation in both the completion and timeliness of overdue cancer screening across screening modalities. Although HPV or Pap testing had the highest overall completion rate, time to completion was significantly shorter for mammography and FIT or Cologuard screening. The association between transportation limitations and delayed screening further underscores the influence of access-related barriers on timely preventive care. This suggests that efforts to improve cancer screening should extend beyond patient outreach to incorporate modality-specific strategies and interventions that address structural barriers to screening completion.

(JMIR Mhealth Uhealth 2026;14:e93912) doi:10.2196/93912

KEYWORDS

cancer screening; SMS text messaging; transportation; social determinants of health; mammography; FIT; HPV testing; PAP testing

XSL•FO
RenderX

Introduction

Routine cancer screening is a significant contributor to the persistent decline in cancer mortality [1]; Federally Qualified Health Centers (FQHCs) provide preventive and primary care to almost 30 million individuals in the United States; yet, preventable cancer screening rates remain significantly lower among patients who receive care in FQHCs than among the general population [2-4]. Past studies comparing rates of cancer screenings have found persistent disparities in FQHC populations, regardless of cancer type including breast (FQHC: 45.4% vs general population: 78.2%), colorectal (FQHC: 40.2% vs general population: 72.3%), or cervical cancer (FQHC: 51.0% vs general population: 82.9%) [3,4]. FQHCs are required to report standardized quality measures, including breast, cervical, and colorectal cancer screening aligned with US Preventive Services Task Force (USPSTF) recommendations, and higher performance on these measures is associated with better preventive care delivery and progress toward national screening goals [5-7]. A growing body of evidence, including past research by our team, suggests that SMS text reminders for FQHC patients overdue for their cancer screenings may be an effective strategy toward reducing this disparity in preventive care [8-10].

Meta-analytic and clinical trial evidence support SMS reminders as a useful nudge that increases screening uptake for cervical, breast, and CRC screening, though absolute effects vary (often ~2 - 5 percentage points in mammography trials and larger in targeted fecal immunochemical test [FIT] programs) [6,7,10-18]. Studies also vary in the type of cancer, addition of phone calls or mailed notices, and message content [6,7,10-18]. Most previous studies have evaluated whether SMS-based interventions increase cancer screening completion. However, less attention has been paid to the timeliness of completion among patients who ultimately complete screening. One prior study in FQHC settings reported median intervals of 72 and 83 days from an abnormal FIT result to follow-up colonoscopy [19], suggesting that substantial delays may persist even among patients who ultimately complete the recommended follow-up. Patients receiving care at FQHCs often face socioeconomic and structural barriers [20] and longer times to screening completion, which may reduce opportunities for timely detection and follow-up and diminish the potential benefits of cancer screening.

Populations served by FQHCs are especially impacted by social determinants of health (SDOH), with common unmet social needs including transportation, food insecurity, and social support contributing to lower cancer screening rates. Individuals who lack reliable transportation are less likely to obtain guideline recommended cancer screenings, with one study reporting 41% lower odds of mammography adherence, independent of other SDOH [21,22]. However, it is unclear to what extent and which specific SDOH may facilitate or act as a barrier to SMS text reminder implementation in safety net populations. Previous studies have shown that time to screening completion varies across patient groups. For example, Vives et al [23] found that used FIT kits were returned sooner by women, older adults, individuals with higher deprivation scores, and those with prior screening experience. Similarly, a recent study found that younger rural adults took longer to return FIT kits, whereas women returned them sooner than men [24]. These findings demonstrate the value of examining time to completion in addition to overall screening completion.

Building on this prior body of evidence, this study examined factors associated with time to cancer screening completion among women aged 50 years or older who received SMS reminders and completed screening within a large FQHC network in Texas. The screening tests included human papillomavirus (HPV) or Papanicolaou testing (hereinafter “Pap testing”), mammography, the FIT, and Cologuard screening. Studies of this nature highlight how screening modality, socioeconomic factors, and health-related social needs may influence delays in completing recommended preventive care.

Methods

Data Source

Secondary electronic health record (EHR) data were obtained from a quality improvement (QI) intervention that examined the impact of SMS text message reminders on cancer screening test completion. The QI intervention was conducted at a large FQHC network in Texas. Because age ranges differ across the three cancer screening modalities, we restricted the analytic sample to women aged 50 years or older who had received at least three reminder SMS messages (three consecutive weeks) to complete any of the following overdue cancer screening tests: (1) HPV or Pap test, (2) mammography, or (3) FIT or Cologuard screening. For FIT and Cologuard screening, the FQHC provided home delivery of the test kits.

Intervention Design and Setting

The QI initiative used SMS messages to contact patients identified through the EHR as overdue for breast, cervical, or colorectal cancer screening. The present study examined the time from the first SMS reminder to EHR-documented completion of the overdue screening test and factors associated with the rate of completion. The rolling intervention was implemented from May 2023 through June 2024. The unit of analysis was the screening episode. Approximately 5% of patients were eligible for two screening modalities and could therefore contribute one observation to each relevant modality-specific analysis.

Participants and Eligibility

Patients were eligible for our analysis if they were (1) female and aged 50 years or older at the time of the first SMS reminder, (2) overdue for HPV or Pap testing, mammography, or FIT or Cologuard screening as of May 2023, (3) received at least three SMS reminders related to the overdue test, and (4) subsequently completed the corresponding screening test. Patients whose screening had been completed before the initial SMS reminder were excluded. Because this study focused on the timeliness of screening completion among patients who received the intervention messages and subsequently completed screening, we excluded patients in the control group, those who did not complete screening, and those who opted out.

XSL•FO
RenderX

The QI Intervention

During the intervention period, the FQHCs implemented an SMS text messaging reminder initiative for eligible female patients who were overdue for cancer screening. SMS messages were sent in English and Spanish and tailored to the cancer screening test for which the patient was overdue. The intervention messages were tailored to each patient's cancer test type: colorectal, breast, or cervical cancer screening. Patients received between 3 and 16 messages during the intervention period. The timing and frequency of messages varied across weeks and intervention pathways. Messages were generally sent between 9 AM and 11:30 AM Central Time, most often on Tuesdays through Fridays. Most messages were interactive, allowing patients to respond directly to the SMS messages or click embedded links to watch a short video, or to access additional information and scheduling resources. Some messages provided information about cancer screening tests, some included links to schedule cancer screening appointments, and some provided educational links about the benefits of screening and the risks of not completing screening.

Examples of the breast messages included, "Dense breasts have more fibrous tissue than fat. It is harder to spot cancer in dense breasts on a mammogram. If your mammogram shows dense breasts, you might be asked to get an ultrasound or MRI for a better view. Do it! Learn more: https://msg.care/8SSoY."

Another example was, "Hi %{patient_first_name}! %{provider_short_name} reminds you that you are due for a mammogram This is an important screening test for breast cancer. Call us at ###-###-#### to schedule."

Outcome and Follow-Up

The event was completion of the screening test for which the patient was overdue, as documented in the EHR. Time to completion was calculated as the number of days from the date of the first SMS reminder to the date of documented test completion. Patients without documented screening completion or who opted out were excluded from this analysis. The latest EHR observation date available in our dataset was July 2024, and the maximum observed time to completion was 249 days. Information on death, transfer of care, and screening completed outside the FQHC network was not consistently available and therefore could not be incorporated.

Covariates

Covariates were selected a priori based on their plausible relationships with access to, and completion of, preventive screening. EHR-derived variables included age at the first SMS reminder, insurance status, primary language, race and ethnicity, completion of an annual wellness visit during the preceding (12 mo per calendar year), BMI category, and documented histories of depression, hypertension, diabetes, and family history of cancer.

Age was categorized as 50 - 54, 55 - 59, 60 - 64, or 65 years or older. Insurance was categorized as uninsured, private insurance, Medicaid, or Medicare. Primary language was categorized as English, or Spanish or another language. Race and ethnicity were categorized as non-Hispanic White, non-Hispanic Black, Hispanic, or another race or multiple races and ethnicities. Overweight or obesity was defined as a BMI of ≥25 kg/m². Health-related social needs (HRSN) were obtained from patients' responses to the program's routine screening questionnaire and included social isolation, food insecurity, housing insecurity, and transportation limitations. Each measure was coded as positive or negative. The most recent response recorded before the first SMS reminder was used.

Missing Data

EHR-derived demographic and clinical covariates were complete in the source data. Missingness was present in the self-reported HRSN measures: 50% for food insecurity, 50% for housing insecurity, 44% for transportation limitations, and 43% for social isolation. No values were imputed. We compared demographic characteristics between observations included in and excluded from the complete-case analysis and found no substantial differences between the groups.

Statistical Analysis

We first summarized the characteristics of the study population using descriptive statistics. The time-to-event outcome, defined as the number of days from the start of the intervention to screening completion, was visualized using Kaplan-Meier curves. We plotted stratified curves for each cancer screening test type to visually compare completion trends. The proportional hazards assumption was assessed using Schoenfeld residuals, and no violations were detected (all P values were >.05). Subsequently, a Cox proportional hazards model was applied to quantify the associations between time to screening completion and the independent variables among patients who completed screening [25]. Hazard ratios (HRs) and their 95% CIs were derived by exponentiating the model coefficients. We also computed multicollinearity using variance inflation factors (VIFs) to ensure model stability. All analyses were performed using R (version 4.4.2) in RStudio (Posit, PBC), using the survival package. All tests were 2-sided with a significance level of P<.05.

Ethical Considerations

The University of Houston's Institutional Review Board reviewed and approved this secondary analysis of EHR data (STUDY00004081). The board approved waivers of informed consent because the study used existing clinical and QI data and involved no additional participant contact. The analytic dataset was deidentified before sharing with the research team for analysis and was stored on access-controlled institutional systems. Patients received no compensation for their inclusion in this secondary data analysis. This manuscript was prepared in accordance with STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for cohort studies.

Results

Table S1 in Multimedia Appendix 1 shows screening completion rates overall and by screening type. The total sample of patients who received SMS reminders included 1803 individuals. In the HPV or Pap test group, 212 (40.38%) patients completed screening, while 313 (59.62%) did not complete screening. In

XSL•FO
RenderX

the mammography group, 138 (21.10%) patients completed screening, while 516 (78.90%) did not complete screening. In the FIT or Cologuard screening group, 201 (32.21%) patients completed screening, while 423 (67.79%) did not complete screening. Overall, 551 (30.56%) patients completed screening, and 1252 (69.44%) did not complete screening. Table 1 presents the descriptive statistics of the study sample. Among participants who completed a cancer screening test, 38.48% (n=212) were in the HPV or Pap test group, 25.05% (n=138) in the mammogram group, and 36.47% (n=201) in the FIT or Cologuard screening group. Regarding age group, 26.32% (n=145) of participants who completed a test were aged 50 - 54 years, 26.68% (n=147) were aged 55 - 59 years, 24.32% (n=134) were aged 60 - 64 years, and 22.68% (n=125) were aged 65 years or older. The proportion of participants whose primary language was English was 53.36% (n=294). In terms of insurance coverage, 52.63% (n=290) of patients were uninsured, 28.86% (n=159) had private insurance, 4.72% (n=26) had Medicaid, and 13.79% (n=76) had Medicare. Over half of the participants were Hispanic (n=302, 54.81%), 24.32% (n=134) were non-Hispanic White, 9.07% (n=50) were Black/African American, and 11.80% (n=65) were of other races (Asian, bi- or multiracial, and other racial categories). Regarding health-related variables, 25.41% (n=140) of participants had depressive disorder, 59.17% (n=326) had hypertension, 88.75% (n=489) had overweight or obesity, 34.48% (n=190) had diabetes, 7.80% (n=43) had a family history of cancer, and 21.78% (n=120) had a wellness visit in 2022. Additionally, 12.52% (n=69) of patients reported social isolation, 4.90% (n=27) reported having food insecurity, 13.79% (n=76) reported having housing insecurity, and 9.98% (n=55) had transportation limitations.

XSL•FO
RenderX

Table . Sociodemographic characteristics of the study sample (n=551).

Variable	Patients, n (%)
Cancer screening test	
HPV[a] or Pap[b] test	212 (38.48)
Mammogram	138 (25.05)
Fecal immunochemical test or Cologuard screening	201 (36.47)
Age (years)	
50 - 54	145 (26.32)
55 - 59	147 (26.68)
60 - 64	134 (24.32)
65+	125 (22.68)
Primary language	
English	294 (53.36)
Spanish and others	257 (46.64)
Insurance type	
Uninsured	290 (52.63)
Private	159 (28.86)
Medicaid	26 (4.72)
Medicare	76 (13.79)
Race/ethnicity	
Non-Hispanic White	134 (24.32)
Black or African American	50 (9.07)
Hispanic	302 (54.81)
Others	65 (11.80)
Depressive disorder	
No	411 (74.59)
Yes	140 (25.41)
Overweight or obesity	
No	62 (11.25)
Yes	489 (88.75)
Hypertension	
No	225 (40.83)
Yes	326 (59.17)
Diabetes	
No	361 (65.52)
Yes	190 (34.48)
Family cancer history	
No	508 (92.20)
Yes	43 (7.80)
Wellness visits in the prior year	
No	431 (78.22)
Yes	120 (21.78)
Social isolation	
No	482 (87.48)

XSL•FO
RenderX

Variable	Patients, n (%)
Yes	69 (12.52)
Food insecurity	
No	524 (95.11)
Yes	27 (4.90)
Housing insecurity	
No	475 (86.21)
Yes	76 (13.79)
Transportation limitation	
No	496 (90.02)
Yes	55 (9.98)

[a]HPV: human papillomavirus.

[b]Pap: Papanicolaou.

Table 2 shows the median survival times by type of cancer screening. The median survival times for the overall cohort, HPV or Pap testing, mammography, and FIT or Cologuard screening groups, were 59.0 (95% CI 53 - 65), 72.5 (95% CI 64 - 86), 52.0 (95% CI 43 - 64), and 52.0 (95% CI 52 - 53) days, respectively. Figure 1 shows Kaplan-Meier curves for time to cancer screening completion by type of cancer screening. Higher survival probabilities indicate a greater proportion of patients who had not yet completed screening at a given time point (ie, slower completion). The HPV or Pap testing cohort showed the highest survival probabilities, suggesting the slowest completion of follow-up. For the overall cohort, the completion period ranged from 0 to 248 days; for the FIT or Cologuard screening cohort, from 0 to 168 days; for the mammography cohort, from 0 to 248 days; and for the HPV or Pap testing cohort, from 0 to 248 days.

Table . Median survival times by cancer screening type.

Test	Events, n	Survival time (days), median (95% CI)
Overall	551	59.0 (53-65)
HPV[a] or Pap[b]	212	72.5 (64-86)
Mammography	138	52.0 (43-64)
FIT[c] or Cologuard screening	201	52.0 (52-53)

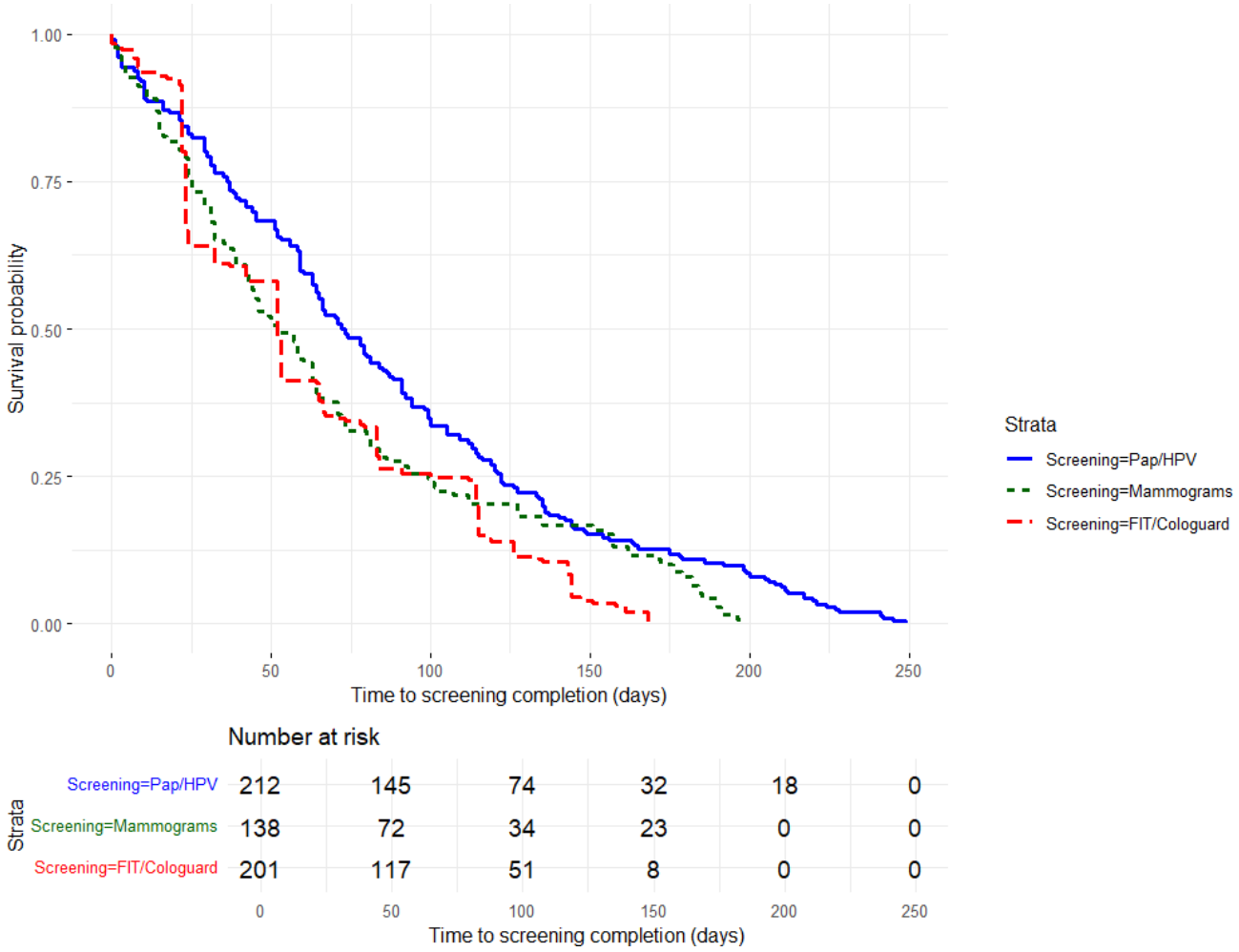
[a]HPV: human papillomavirus.

[b]Pap: Papanicolaou.

[c]FIT: fecal immunochemical test.

XSL•FO
RenderX

Figure 1. Survival probability by cancer screening type. FIT: fecal immunochemical test; HPV: human papillomavirus; Pap: Papanicolaou.



Finally, Table 3 presents the results of the Cox proportional hazards model. The model demonstrated no evidence of problematic multicollinearity (all VIFs<5). Compared to patients who were overdue for HPV or Pap screening, those in the FIT or Cologuard screening group (HR 1.65, 95% CI 1.34 - 2.05; *P*<.001) and the mammography group (HR 1.41, 95% CI 1.11 - 1.78; *P*<.001) had significantly shorter times to screening completion. Screening positive for transportation as a social need was associated with delayed screening completion (HR 0.74, 95% CI 0.55 - 0.99; *P*=.045).

XSL•FO
RenderX

Table . Cox proportional hazards model.

Variable	HR[a]	SE	*P* value	95% CI
Cancer screening test				
HPV[b] or Pap[c] test	Ref[d]	—[e]	—	—
Mammogram	1.41	0.11	<.001	1.34-2.05
FIT[f] or Cologuard screening	1.65	0.12	<.001	1.11-1.78
Age (years)				
50 - 54	Ref	—	—	—
55 - 59	0.85	0.12	.19	[0.67, 1.08]
60 - 64	1.08	0.13	.51	0.85-1.39
65+	0.95	0.14	.74	0.75-1.32
Primary language				
English	Ref	—	—	—
Spanish and others	0.86	0.11	.19	0.69-1.08
Insurance type				
Uninsured	Ref	—	—	—
Private	0.99	0.10	.99	0.81-1.23
Medicaid	1.29	0.21	.23	0.85-1.96
Medicare	0.91	0.15	.53	0.85-1.23
Race or ethnicity				
White	Ref	—	—	—
Black or African American	0.84	0.18	.34	0.59-1.20
Hispanic	1.14	0.14	.34	0.87-1.50
Others	1.04	0.17	.82	0.75-1.45
Depressive disorder				
No	Ref	—	—	—
Yes	1.11	0.11	.34	0.90-1.37
Overweight or obesity				
No	Ref	—	—	—
Yes	0.83	0.12	.97	0.61-1.11
Hypertension				
No	Ref	—	—	—
Yes	0.85	0.10	.11	0.69-1.04
Diabetes				
No	Ref	—	—	—
Yes	1.04	0.10	.73	0.84-1.27
Wellness visits in 2022				
No	Ref	—	—	—
Yes	1.16	0.11	.19	0.93-1.43
Family member has a history of cancer				
No	Ref	—	—	—
Yes	1.09	0.17	.62	0.78-1.53
Social isolation				

XSL•FO
RenderX

Variable	HR[a]	SE	*P* value	95% CI
No	Ref	—	—	—
Yes	0.97	0.14	.83	0.74-1.27
Food insecurity				
No	Ref	—	—	—
Yes	1.04	0.22	.99	0.66-1.53
Housing insecurity				
No	Ref	—	—	—
Yes	0.96	0.13	.79	0.73-1.26
Transportation limitation				
No	Ref	—	—	—
Yes	0.74	0.15	.045	0.55-0.99

[a]HR: hazard ratio.

[b]HPV: human papillomavirus.

[c]Pap: Papanicolaou.

[d]Ref: reference group.

[e]Not applicable.

[f]FIT: fecal immunochemical test.

Discussion

Principal Findings

In this study of women aged 50 years or older who completed cancer screening after receiving SMS reminders within a large FQHC network, we observed notable differences in the timeliness of screening completion across screening modalities. Compared to the HPV or Pap testing cohort, patients exhibited shorter times to screening completion for breast and colorectal cancer screening tests. Importantly, screening positive for transportation as a social need was associated with delayed screening completion, suggesting that transportation barriers may be an important factor related to timely preventive care. These findings suggest that social needs and screening modality may jointly contribute to suboptimal preventive cancer screening behaviors.

Although patients received similar intervention messages for breast, cervical, and colorectal cancer screening, the time required to complete screening differed across the three screening modalities. The finding of faster time to screening for FIT or Cologuard screening and mammography completion, compared to HPV or Pap testing, is noteworthy, and aligns with prior research elucidating variations in barriers to cancer screening across multiple sites [26,27]. Thus, our study findings may be reflective of these variations. The ability to complete FIT and Cologuard screening tests at home and return them by mail may reduce barriers such as scheduling conflicts, transportation challenges, and time constraints. The longer time to HPV or Pap testing completion may be attributable to delays in obtaining appointments with primary care providers or gynecologists, which may exceed wait times for mammography services. These differences and perceived burden may explain the shorter completion time we observed among patients assigned to the HPV or Pap testing intervention. Augmenting text message reminders with patient navigation to assist with identifying and addressing barriers to screening could reduce time to screening [28].

The observed association between transportation challenges and delayed screening completion has practical implications. Individuals may face transportation challenges due to transportation costs or difficulty navigating public transit [29,30]. This insight offers an actionable strategy for improving screening completion rates, especially among FQHC patients, and potentially reducing cancer outcomes disparities. For example, others have explored the use of mobile mammograms [31], in partnership with community stakeholders, to provide mammograms within the community and reduce transportation-related challenges. Additionally, strategies such as ride-share programs and insurance-covered transportation, which have been used to address transportation barriers to cancer treatment, may also offer an effective approach to addressing similar screening barriers [32-34]. For patients who ultimately complete screening, transportation assistance may reduce delays and facilitate more timely completion.

Limitations and Strengths

This study is not without limitations. First, due to the focus on one FQHC network in Texas and the inclusion of only women aged 50 years and older, the generalizability of the findings may be limited. Second, because patient-level information on death, transfer of care, loss to follow-up, and the end of observation was unavailable, censoring could not be reliably defined for patients without documented screening completion. This analysis was restricted to patients with valid documented completion dates. Accordingly, the findings represent associations among screening completers and should not be interpreted as factors associated with whether screening was completed or generalized to patients who did not complete screening or opted out. The study, however, focuses on patients in an FQHC population

XSL•FO
RenderX

who received SMS reminders and completed screening, allowing us to identify factors associated with faster or delayed completion among eventual completers. Third, there are other contextual factors, such as clinic-level workflows, provider recommendation patterns, and individual health beliefs, which may influence time to screening completion that are not captured in this study.

Additionally, our analysis treated each data point as independent. In reality, some individuals were likely eligible for more than one type of screening test, introducing potential correlation across observations. Still, because the study examined completion of each screening test separately, this limitation is unlikely to meaningfully affect the findings. Furthermore, HRSN variables relied on self-reported information, which may be subject to missingness or reporting bias. Patients with missing HRSN values were excluded from the analysis, which may have reduced the analytic sample size and introduced potential selection bias if those patients differed from those with complete HRSN data. Finally, due to data limitations, we were unable to control for other potential barriers to screening, such as scheduling conflicts, fear, anxiety, or preferred language other than English, which may have resulted in residual confounding.

Despite these limitations, this study has several strengths. First, our focus on patients served by FQHCs adds to the literature, given their lower screening uptake and longer time to screening completion than those of the general population. Second, our comparison of multiple screening modalities provides stakeholders with insights into differences in completion time and identifies which screening pathways may require more time or additional support for these three common cancers. Third, our inclusion of demographic, social needs, and health-related factors increases the robustness of the findings. Fourth, identifying transportation as a barrier highlights a clear and actionable target for interventions aimed at promoting timely completion of recommended cancer screening.

Conclusions

In closing, our findings show that the same SMS reminder intervention have varied effects across cancer screening modalities, suggesting that additional demands, such as structural barriers faced by underserved patients, influence timely cancer completion. Although all patients in this study ultimately completed screening, transportation challenges were associated with delayed completion, reinforcing the need for targeted transportation support in FQHC settings. These findings suggest that SMS reminders alone may be insufficient to address structural barriers. Efforts to reduce these barriers may promote faster screening completion, enhance the effectiveness of SMS-based interventions, and help reduce inequities in cancer prevention outcomes.

Acknowledgments

Generative AI was not used at any point in the writing, drafting, or editing of this manuscript.

Funding

The authors declared no financial support was received for this work.

Data Availability

The data generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: OEA
Data curation: TX, OEA
Formal analysis: TX
Methodology: TX, OEA
Writing – original draft: TX, SC, BTK, GRP, COO
Writing – review & editing: TX, SC, BTK, GRP, COO, DO, VK, OEA

Conflicts of Interest

None declared.

Multimedia Appendix 1
Event rates by screening type (n=1803).
[DOCX File, 14 KB - mhealth_v14i1e93912_app1.docx]

References

1. Bhardwaj N, Herndon AT, Kuo YF, Porterfield LR. Text messaging intervention for Pap smear uptake: A single-institution study. Mhealth 2023;9(34):34. [doi: 10.21037/mhealth-23-11] [Medline: 38023781]
2. Firmino-Machado J, Varela S, Mendes R, Moreira A, Lunet N, SCAN-Cervical Cancer collaborators. Stepwise strategy to improve cervical cancer screening adherence (SCAN-Cervical Cancer) - Automated text messages, phone calls and reminders:

XSL•FO
RenderX

Population based randomized controlled trial. Prev Med 2018 Sep;114:123-133. [doi: 10.1016/j.ypmed.2018.06.004] [Medline: 29894717]
3. Amboree TL, Montealegre JR, Parker SL, et al. National breast, cervical, and colorectal cancer screening use in federally qualified health centers. JAMA Intern Med 2024 Jun 1;184(6):671-679. [doi: 10.1001/jamainternmed.2024.0693] [Medline: 38683574]
4. Kao R, Patel AK, Sadigh G. The role of federally qualified health centers in achieving equitable cancer screening. J Am Coll Radiol 2025 Jul;22(7):718-721. [doi: 10.1016/j.jacr.2025.03.005] [Medline: 40157408]
5. Ruco A, Dossa F, Tinmouth J, et al. Social media and mHealth technology for cancer screening: Systematic review and meta-analysis. J Med Internet Res 2021 Jul 30;23(7):e26759. [doi: 10.2196/26759] [Medline: 34328423]
6. Kerrison RS, Shukla H, Cunningham D, Oyebode O, Friedman E. Text-message reminders increase uptake of routine breast screening appointments: A randomised controlled trial in a hard-to-reach population. Br J Cancer 2015 Mar 17;112(6):1005-1010. [doi: 10.1038/bjc.2015.36] [Medline: 25668008]
7. Mehta SJ, Rhodes C, Linn KA, et al. Behavioral Interventions to Improve Breast Cancer Screening Outreach: Two Randomized Clinical Trials. JAMA Intern Med 2024 Jul 1;184(7):761-768. [doi: 10.1001/jamainternmed.2024.0495] [Medline: 38709509]
8. Adepoju OE, Xu T, Angelocci T. From SMS Text Messaging to Screening Completion: Understanding How SMS Responses Influence Cancer Screening Uptake in Federally Qualified Health Center Clinics. Health Serv Res 2026 Jun;61(3):e70134. [doi: 10.1111/1475-6773.70134] [Medline: 42206315]
9. Angelocci T, Xu T, Talaparthy T, Corral C, Adepoju OE. Leveraging text messaging and behavior theory to improve colorectal cancer screening in federally qualified health centers: Cohort study. J Med Internet Res 2026 May 13;28:e86408. [doi: 10.2196/86408] [Medline: 42127425]
10. Levitz CE, Kuo E, Guo M, et al. Using text messages and fotonovelas to increase return of home-mailed colorectal cancer screening tests: Mixed methods evaluation. JMIR Cancer 2023 Mar 7;9:e39645. [doi: 10.2196/39645] [Medline: 36881466]
11. Nanda AD, Reifel KR, Mann MP, et al. Text-based intervention increases mammography uptake at an Urban Safety-Net hospital. Ann Surg Oncol 2022 Oct;29(10):6199-6205. [doi: 10.1245/s10434-022-12130-x] [Medline: 35976462]
12. Hamad W, Purushotham A, Hughes S, Round T. Text-message reminders to increase participation in colorectal cancer screening. Br J Gen Pract 2024 Apr;74(741):152. [doi: 10.3399/bjgp24X736761] [Medline: 38538132]
13. Vives N, Travier N, Farre A, et al. Effectiveness and acceptability of targeted text message reminders in colorectal cancer screening: Randomized controlled trial (M-TICS Study). JMIR Public Health Surveill 2024 Jul 31;10(57959):e57959. [doi: 10.2196/57959] [Medline: 39083331]
14. Oyalowo A, Forde KA, Lamanna A, Kochman ML. Effect of patient-directed messaging on colorectal cancer screening: A randomized clinical trial. JAMA Netw Open 2022 Mar 1;5(3):e224529. [doi: 10.1001/jamanetworkopen.2022.4529] [Medline: 35357457]
15. Huguet N, Hodes T, Holderness H, Bailey SR, DeVoe JE, Marino M. Community health centers' performance in cancer screening and prevention. Am J Prev Med 2022 Feb;62(2):e97-e106. [doi: 10.1016/j.amepre.2021.07.007] [Medline: 34663549]
16. Martinez-Gutierrez J, Jhingan E, Angulo A, Jimenez R, Thompson B, Coronado GD. Cancer screening at a federally qualified health center: A qualitative study on organizational challenges in the era of the patient-centered medical home. J Immigr Minor Health 2013 Oct;15(5):993-1000. [doi: 10.1007/s10903-012-9701-8] [Medline: 22878911]
17. Resources USH, Administration S. Uniform Data System (UDS) quality measure benchmarks: Breast, cervical and colorectal cancer screening. 2025. URL: https://bphc.hrsa.gov/sites/default/files/bphc/data-reporting/uds-cqm-benchmarks.pdf [accessed 2026-08-28]
18. Joseph NT, Namuli A, Kakuhikire B, et al. Implementing community-based human papillomavirus self-sampling with SMS text follow-up for cervical cancer screening in rural, southwestern Uganda. J Glob Health 2021;11(4036):04036. [doi: 10.7189/jogh.11.04036] [Medline: 35003710]
19. Coronado GD, Kihn-Stang A, Slaughter MT, et al. Follow-up colonoscopy after an abnormal stool-based colorectal cancer screening result: Analysis of steps in the colonoscopy completion process. BMC Gastroenterol 2021 Sep 28;21(1):356. [doi: 10.1186/s12876-021-01923-1] [Medline: 34583638]
20. Shin E, Jung D, Rajbhandari-Thapa J. Cancer screening in federally qualified health centers by neighborhood social vulnerability: A national study. J Gen Intern Med 2026 Mar 12. [doi: 10.1007/s11606-026-10355-0] [Medline: 41820739]
21. Davis C. Overcoming transportation barriers to cancer screening. Prevent Cancer Foundation. 2023. URL: https://preventcancer.org/article/overcoming-transportation-barriers-to-cancer-screening/ [accessed 2026-08-28]
22. Wang C. The impact of car ownership and public transport usage on cancer screening coverage: Empirical evidence using a spatial analysis in England. J Transp Geogr 2016 Oct;56:15-22. [doi: 10.1016/j.jtrangeo.2016.08.012] [Medline: 27829709]
23. Vives N, Milà N, Binefa G, et al. Role of community pharmacies in a population-based colorectal cancer screening program. Prev Med 2021 Apr;145:106420. [doi: 10.1016/j.ypmed.2021.106420] [Medline: 33422578]
24. Beseler CL, Kim J, Leypoldt M, et al. Investigating FIT kit completion for CRC screening in younger adults in rural areas. Discov Soc Sci Health 2024;4(1). [doi: 10.1007/s44155-024-00102-3]

XSL•FO
RenderX

25. Cox DR. Regression models and life-tables. J R Stat Soc B 1972 Jan 1;34(2):187-202. [doi: 10.1111/j.2517-6161.1972.tb00899.x]
26. Aguiar-Ibáñez R, Mbous Y, Sharma S, Chakali R, Chawla E. Barriers to cancer screening uptake and approaches to overcome them: A systematic literature review. Front Oncol 2025;15:1575820. [doi: 10.3389/fonc.2025.1575820] [Medline: 40842583]
27. Gurayah AA, An A, Kuchakulla M, et al. Barriers to health care and cancer screening. JAMA Netw Open 2026 Apr 1;9(4):e267024. [doi: 10.1001/jamanetworkopen.2026.7024] [Medline: 41979877]
28. Gonzalez Mosquera DA, Mateo Chavez MB, Bandi SSS, et al. Improving diagnostic colonoscopy completion after positive stool-based screening in health-disparity populations: A systematic review and meta-analysis. J Gen Intern Med 2026 Aug 19. [doi: 10.1007/s11606-026-10706-x] [Medline: 42618855]
29. Syed ST, Gerber BS, Sharp LK. Traveling towards disease: Transportation barriers to health care access. J Community Health 2013 Oct;38(5):976-993. [doi: 10.1007/s10900-013-9681-1] [Medline: 23543372]
30. Wolfe MK, McDonald NC, Holmes GM. Transportation barriers to health care in the United States: Findings from the National Health Interview Survey, 1997-2017. Am J Public Health 2020 Jun;110(6):815-822. [doi: 10.2105/AJPH.2020.305579] [Medline: 32298170]
31. Bakhai S, Ansari M, Sadeghi C, Reynolds JL. Advancing health equity in improving breast cancer screening with the use of a mobile mammography bus in marginalised population: Quality improvement project. BMJ Open Qual 2024 Jan 4;13(1):e002482. [doi: 10.1136/bmjoq-2023-002482] [Medline: 38176952]
32. Pringle S, Ko EM, Doherty M, Smith AJB. Addressing transportation barriers in oncology: Existing programs and new solutions. Support Care Cancer 2024 Apr 29;32(5):317. [doi: 10.1007/s00520-024-08514-2] [Medline: 38684580]
33. Yabroff KR, Breen N, Vernon SW, Meissner HI, Freedman AN, Ballard-Barbash R. What factors are associated with diagnostic follow-up after abnormal mammograms? Findings from a U.S. National Survey. Cancer Epidemiol Biomarkers Prev 2004 May;13(5):723-732. [Medline: 15159302]
34. Sabatino SA, Thompson TD, White MC, et al. Cancer Screening Test Receipt - United States, 2018. MMWR Morb Mortal Wkly Rep 2021 Jan 15;70(2):29-35. [doi: 10.15585/mmwr.mm7002a1] [Medline: 33444294]

Abbreviations

EHR: electronic health record
FIT: fecal immunochemical test
FQHC: Federally Qualified Health Center
HPV: human papillomavirus
HR: hazard ratio
HRSN: health-related social needs
QI: quality improvement
SDOH: social determinants of health
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
USPSTF: US Preventive Services Task Force
VIF: variance inflation factor

Edited by L Buis; submitted 21.Feb.2026; peer-reviewed by B Singh, R Ferrari; revised version received 20.Jul.2026; accepted 18.Aug.2026; published 03.Sep.2026.

Please cite as:
Xu T, Chavez S, King B, Ramirez-Pulido G, Ojinnaka CO, Osayi D, Kounga V, Adepoju OE
Time to Cancer Screening Completion Following an SMS Reminder Intervention in a Large Federally Qualified Health Center Network: Secondary Data Analysis
JMIR Mhealth Uhealth 2026;14:e93912
URL: https://mhealth.jmir.org/2026/1/e93912
doi:10.2196/93912

© Tonghui Xu, Summer Chavez, Ben King, Gonzalo Ramirez-Pulido, Chinedum O Ojinnaka, Daniel Osayi, Valery Kounga, Omolola E Adepoju. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 3.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Individualized Treatment Effects of a Digital Smoking Cessation Intervention Among Individuals Looking Online for Help: Secondary Analysis of a Randomized Controlled Trial

Joel Crawford[1], PhD; Jenny Blomqvist[1], BSc; Katarina Ulfsdotter Gunnarsson[1], PhD; Preben Bendtsen[1,2], MD; Marcus Bendtsen[1], PhD

[1]Department of Health, Medicine and Caring Services, Linköping University, Linköping, Sweden

[2]Department of Medical Specialist, Motala Hospital, Motala, Sweden

Corresponding Author:
Joel Crawford, PhD
Department of Health, Medicine and Caring Services, Linköping University, Linköping, Sweden

Abstract

Background: Smoking is a leading cause of mortality and morbidity worldwide. Efforts to reduce smoking prevalence have used SMS text message–based interventions, which typically send participants a series of short, informational, motivational, and practical messages over a set period. Evidence highlights the efficacy of using this approach to support smoking cessation, with such trials typically reporting the average treatment effects, in which causal inference is made regarding the average effect of a treatment on a heterogeneous sample. Nonetheless, using this approach to assessing treatment effects means we are unable to account for individual factors that impact the effectiveness of a treatment on outcomes, such as age, gender, and genetics.

Objective: This study aimed to estimate the individualized effects of an SMS text message–based smoking cessation intervention to ascertain which individuals benefited the most and least during an effectiveness trial.

Methods: Data from a randomized controlled trial including 1012 adults from the Swedish general population were used. The trial assessed the effects of an SMS text message–based intervention, NEXit (Nicotine Exit), that aimed to change behavior by increasing the importance of change, boosting knowledge on how to change, and instilling confidence for change. Outcomes assessed in the trial were prolonged abstinence and point prevalence of smoking cessation. Individualized treatment effects were modeled using baseline factors (demographics, psychosocial variables, and past behavior) to study who benefited the most and least from the intervention.

Results: For prolonged abstinence, there was evidence of heterogeneous effects, with those benefiting the most from NEXit being older adults, female participants, individuals with high confidence in their ability to quit, and those who believed that quitting was important. For point prevalence abstinence, older individuals and those reporting high confidence in the ability to quit, the importance of quitting, and knowledge for change benefited the most. For both outcomes, individuals who reported smoking for a longer duration and smoking more at baseline benefited less.

Conclusions: The results demonstrate how individuals respond differently to an SMS text message–based smoking cessation intervention. This provides an insight into who benefits the most and least from the intervention in terms of demographics, baseline characteristics, and behaviors. The study highlights which individuals need to be specifically targeted and/or have content developed to suit their individual needs to further reduce the prevalence of smoking.

Trial Registration: ISRCTN Registry ISRCTN13455271; http://www.isrctn.com/ISRCTN13455271

(JMIR Mhealth Uhealth 2026;14:e63578) doi:10.2196/63578

KEYWORDS

individualized treatment effects; randomized controlled trial; text message–based smoking cessation intervention; smoking cessation; smoking; Sweden; behavior change; text messages; smoker

Introduction

Smoking remains a leading cause of morbidity and mortality across the globe, with recent figures highlighting approximately 8 million deaths worldwide and 200 million disability-adjusted life years attributed to smoking [1,2]. In Sweden, the prevalence of smoking within the general population has consistently decreased over the last 2 decades, with the Public Health Agency of Sweden reporting that prevalence rates dropped to 6% for women and 5% for men in 2022 [3]. Despite this, smoking remains a leading cause of mortality and morbidity within Sweden [4] and is more prevalent in groups with lower

XSL•FO
RenderX

socioeconomic status [5]. To stem the associated morbidity and mortality and reduce health inequalities, effective interventions aimed at reducing the prevalence of smoking within the general population are needed.

Current efforts to reduce smoking prevalence have used SMS text message–based interventions, which typically comprised a series of short messages sent to participants over a designated time period (usually 2-3 mo) [6,7]. The aim of these interventions is to provide motivation to quit, support for quitting, and reinforcement of the decision to quit [7]. A series of meta-analyses have highlighted the effectiveness of SMS text message–based smoking cessation interventions, and all concluded positive effects were observed for both self-reported and biochemically verified abstinence [6-10].

In Sweden, a recent randomized controlled trial (RCT) estimated the effectiveness of receiving a SMS text message–based smoking cessation intervention against existing public support (ie, referral to the national Quitline and website, slutarokalinjen.se) [11]. The trial targeted the general population using the NEXit (Nicotine Exit) protocol [12]. Effect estimates were made at 3 and 6 months post randomization for both prolonged abstinence (ie, the Russell standard definition of not having smoked more than 5 cigarettes in the past 8 wk) and point prevalence of abstinence. At 3 months post randomization, the odds ratio (OR) of prolonged abstinence was 2.15 (95% compatibility interval [CoI] 1.51-3.06), and at 6 months, the OR of prolonged abstinence was 2.38 (95% CoI 1.62-3.57). For point prevalence of abstinence, the OR at 3 months was 1.70 (95% CoI 1.18-2.44), and at 6 months, it was 1.49 (95% CoI 1.03-2.14).

While the findings suggest the intervention is effective for smoking cessation, the results of the RCT were assessed at the group level, meaning we are unable to account for how individual differences impact outcomes. This is pertinent, as not everyone in the intervention group successfully quit smoking, meaning that some individuals benefited more from the intervention than others. Individual differences in terms of fixed (eg, age, ethnicity, gender) and modifiable (eg, socioeconomic status, psychological) health factors have been shown to impact abstinence rates [13-16], which can potentially account for why some individuals benefit more from treatment than others.

Currently, Sweden has a national target of becoming a “smoke-free” state (ie, <5% prevalence in the general population) [17]. To achieve this target, smoking cessation support such as the NEXit protocol needs to become specifically targeted to those who continue to smoke tobacco despite policy measures and interventions. The aim of this study is to ascertain who benefited the most and least in the recent NEXit trial [11] through a secondary analysis of the group-level data, which estimated the average treatment effect of NEXit in the Swedish general population. We aim to achieve this by assessing how baseline factors influence outcomes of the trial at the individual level.

Methods

Study Design

The current study investigated the individualized effects of the NEXit protocol using data from a 2-arm, parallel-group (1:1) RCT. The trial was registered in the ISRCTN registry on December 3, 2020 (ISRCTN13455271), and a protocol and an analysis plan were published prospectively [12].

Participants and Settings

A total of 1012 participants from the Swedish general public were recruited from 2 settings: (1) online advertisements, in which individuals clicked on a link that took them to an information page with instructions for signing up, and (2) primary health care units in southern Sweden, which used printed media (leaflets and posters) containing instructions for signing up. Individuals were required to be aged 18 years or older, have access to a mobile phone, and smoke at least one cigarette per week to participate in the trial. Regardless of medium, those interested in participating sent an SMS text message to a dedicated number and received a reply within 5 minutes with a link to the study information and the informed consent form. After providing consent, participants were asked to complete a baseline survey, which was used to assess eligibility. Immediately afterward, eligible participants were randomly assigned to 1 of 2 study groups.

Ethical Considerations

The trial received ethical approval from the Swedish Ethical Review Authority on June 16, 2020 (Dnr 2020 - 01427). Participants provided informed consent prior to the commencement of the main trial, which included consent for their deidentified data to be analyzed and published. Trial data are deidentified, and participants in the main trial did not receive compensation.

Interventions

Both the control and intervention groups were offered treatment as usual and were thus not prohibited from using publicly available smoking cessation support aids. Treatment as usual was referral to a national smoking cessation helpline and website; those recruited from primary care units also had the option of meeting with a nurse or smoking cessation counselor. Additionally, participants in the intervention group were given access to the NEXit SMS text messaging intervention. The intervention had 2 fairly similar versions: a general version and a version for those undergoing elective surgery (within 3 mo of signing up). The intervention consists of a 1- to 4-week motivational phase in which participants were encouraged to set a date on which to quit, followed by a 12-week program consisting of 157 cessation support messages. Messages were typically informational and encouraging, and they provided tips and techniques for creating a smoke-free environment and lifestyle. These included distraction techniques, tips to avoid weight gain, tips to cope with cravings, tips on how to avoid triggers, tips to self-regulate, and practical tips to help restructure the physical environment. The design of the messages was informed by intervention mapping [18] and the behavior change wheel [19], with various behavior change techniques being

XSL•FO
RenderX

utilized, including 1.2 problem solving, 1.4 action planning, 5.1 information about health consequences, 5.2 salience of consequences, 8.2 behavioral substitution, 8.4 habit reversal, 9.1 credible source, 10.9 self-reward, 12 restricting the physical environment, 12.3 avoidance to cues, and 12.4 distraction. Messages were sent daily to participants, and the elective surgery version contained additional messages regarding the impact of smoking on surgery and recovery. Participants could also text and request additional support if they had relapsed or were experiencing cravings.

Outcomes and Measures

The primary outcomes of the trial were assessed at 3 and 6 months post randomization:

1. Prolonged abstinence, using the Russell standard definition of not having smoked more than 5 cigarettes in the past 8 weeks at the 3-month follow-up and in the past 5 months at the 6-month follow-up
2. Point prevalence, defined as not smoking any cigarettes during the past 4 weeks at both follow-up intervals

Baseline factors assessed in this secondary analysis of individualized effects include factors that are associated with cessation, including demographics (age and gender) and smoking behaviors (number of years smoking, amount of cigarettes smoked per week, use of snus [akin to snuff tobacco], number of past quit attempts, and nicotine dependence score using the Fagerström scale, which uses 6 items to assess intensity of physical addiction, including quantity and frequency of smoking) [13,20-24]. The recruitment setting (online or primary care unit) and whether participants were undergoing elective surgery were also assessed. Psychosocial variables (importance of quitting, confidence in the ability to quit, and knowledge on how to quit) were assessed. These factors are reflective of the behavioral and control beliefs that drive behavior, as outlined in expectancy-value models of health [25,26]. Extant literature highlights that the demographic factors (age and gender) are associated with other baseline factors used in our modeling [13,20-22].

Randomization and Blinding

Randomization was completed using a computer-generated random sequence, stratified according to which version of the intervention was appropriate (general or surgery). Block randomization, using block sizes of 2 and 4, ensured equal numbers of participants were allocated to the groups. Participants received notification of group allocation via SMS text message. Research personnel were blinded to allocation, and all procedures of the study were automated, except follow-up calls to nonresponders. For a full description of all study procedures, see Bendtsen et al [12].

Statistical Model

The observed outcome (Y) was modeled as a function of group allocation (G) and baseline variables (X) to estimate the individualized treatment effect $\delta(xi)=P(Yi|G=1,X=xi)-P(Yi=|G=0,X=xi)$, following the approach by Hoogland et al [27]. Here, δxi represents the difference between the predicted probability of smoking cessation (Yi) for an individual (i) with baseline factors xi, if they had access to the intervention (G=1) versus if they did not (G=0). Individualized treatment effects were modeled for each participant as a function of baseline factors to study which individuals had the most and least benefit from the intervention.

Smoking cessation outcomes at 3 and 6 months post randomization were modeled using multilevel logistic regression with a time-by-group interaction term. Available data were used to estimate the model and predict outcomes for all randomized participants. The models included participant-level adaptive intercepts, covariates for baseline factors, and interactions between baseline factors and group allocation. Bayesian inference was used to estimate the parameters of the models [28]. To promote parsimonious models, priors that shrunk estimates toward the null for baseline factor interaction coefficients were used, specifically Cauchy priors centered at the null with a standard normal hyperprior for the scale parameter [28]. We used Student t test priors centered at 0, with 3 df and a scale of 2.5 for noninteraction coefficients and standard normal priors for adaptive and fixed intercepts. Individualized treatment effects were calculated from posterior predictive draws, resulting in posterior distributions over treatment effects for each individual. Medians of these posterior distributions were used as point estimates of δxi.

Linear regression was used to study associations between standardized estimated individualized effects and baseline factors. Student t test priors centered at 0 with 3 df and a scale of 2.5 for all coefficients in these models. Medians of posterior distributions of covariates were used as point estimates and created 95% CoI using the 2.5 and 97.5 percentiles of the posterior distributions. Estimation was done using CmdStan version 2.33.0 (Stan Project), which is a shell interface for Stan [29]. Plots were generated using the ggplot2 library in R.

Results

Participant Retention and Missing Data

Table 1 presents baseline factors for the 1012 participants recruited to the trial. At the 3-month follow-up interval, data were available for 66% of participants in the intervention group and 69% of participants in the control group. The corresponding proportions for the 6-month interval were 64% for both groups. These data were used to estimate the multilevel logistic regression models, which were subsequently used to estimate individualized treatment effects for all randomized participants. As reported in the primary findings of the trial [11], there was evidence that older participants were more likely to respond to follow-up at the 6-month interval, but no other associations were found between response and baseline factors. The CONSORT (Consolidated Standards of Reporting Trials) flowchart and checklist are presented in Figure 1 and Checklist 1.

XSL•FO
RenderX

Table . Baseline factors of participants.

	Total (N=1012)	Intervention (n=505)	Control (n=507)
Woman[a], n (%)	820 (81)	406 (80.4)	414 (81.7)
Age (years), mean (SD)	45.4 (14.0)	45 (13.9)	45.7 (14.1)
Years of smoking, mean (SD)	25.3 (14.6)	24.7 (14.3)	25.9 (14.9)
Daily smokers (vs weekly smokers), n (%)	981 (96.9)	489 (96.8)	492 (97)
Cigarettes smoked per week, mean (SD)	101 (46.2)	101.4 (47.3)	100.6 (45.1)
Use of snus[b], n (%)			
Daily	63 (6.2)	27 (5.3)	36 (7.1)
Weekly or monthly	90 (8.9)	45 (8.9)	45 (8.9)
No	859 (84.9)	433 (85.7)	426 (84)
Fagerström test for nicotine dependence, mean (SD)	5 (2.2)	5 (2.2)	5 (2.2)
Quit attempts[c], mean (SD)	7.2 (13.7)	7.0 (12.7)	7.5 (14.6)
Cessation counseling experience, n (%)			
Yes, now	32 (3.2)	13 (2.6)	19 (3.7)
Yes	192 (19)	95 (18.8)	97 (19.1)
No	788 (77.9)	397 (78.6)	391 (77.1)
Used quit smoking helpline, n (%)	13.6 (138)	14.5 (73)	12.8 (65)
Importance of quitting[d], mean (SD)	9.4 (1.3)	9.4 (1.3)	9.5 (1.2)
Confidence in ability to quit[d], mean (SD)	6.2 (2.5)	6.3 (2.5)	6.2 (2.6)
Knowledge of how to quit[d], mean (SD)	5.5 (2.6)	5.5 (2.7)	5.5 (2.5)
Elective surgery[e], n (%)	63 (6.2)	31 (6.1)	32 (6.3)

[a]The baseline questionnaire included a category “Other”; however, it was not chosen by any participant.

[b]Snus is a moist oral tobacco product, which is common in Sweden, sometimes translated as snuff.

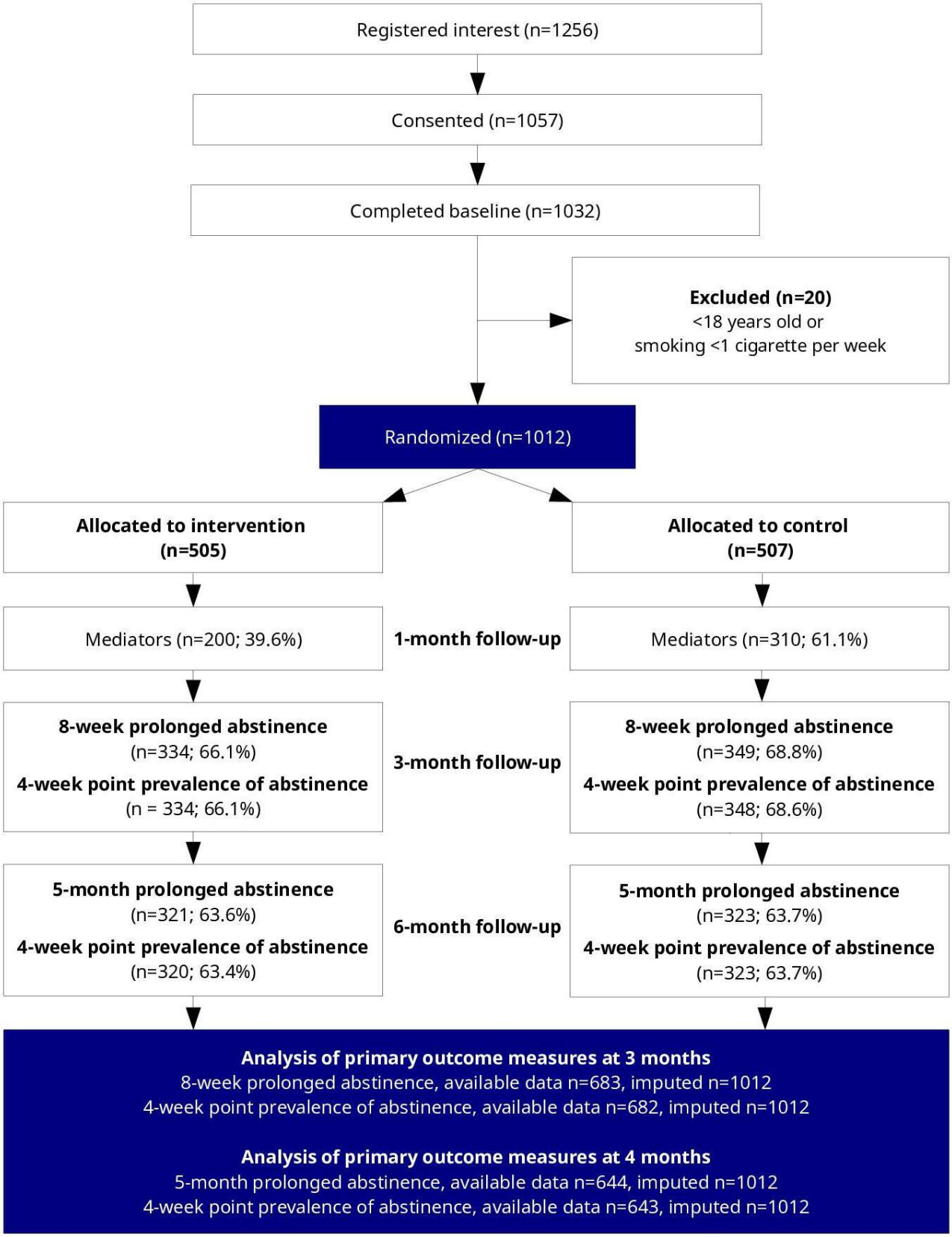
[c]Participants were asked about the lifetime number of quit attempts.

[d]Three single-item measures were used to assess the importance, confidence, and know-how regarding smoking cessation. Responses ranged from 0 to 10, with 10 representing the highest agreement (ie, very important, very confident, very knowledgeable). The same items were used at follow-up as hypothesized mediators of effects.

[e]Participants self-reporting that they had planned elective surgery received additional intervention materials regarding smoking and surgery.

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flowchart.



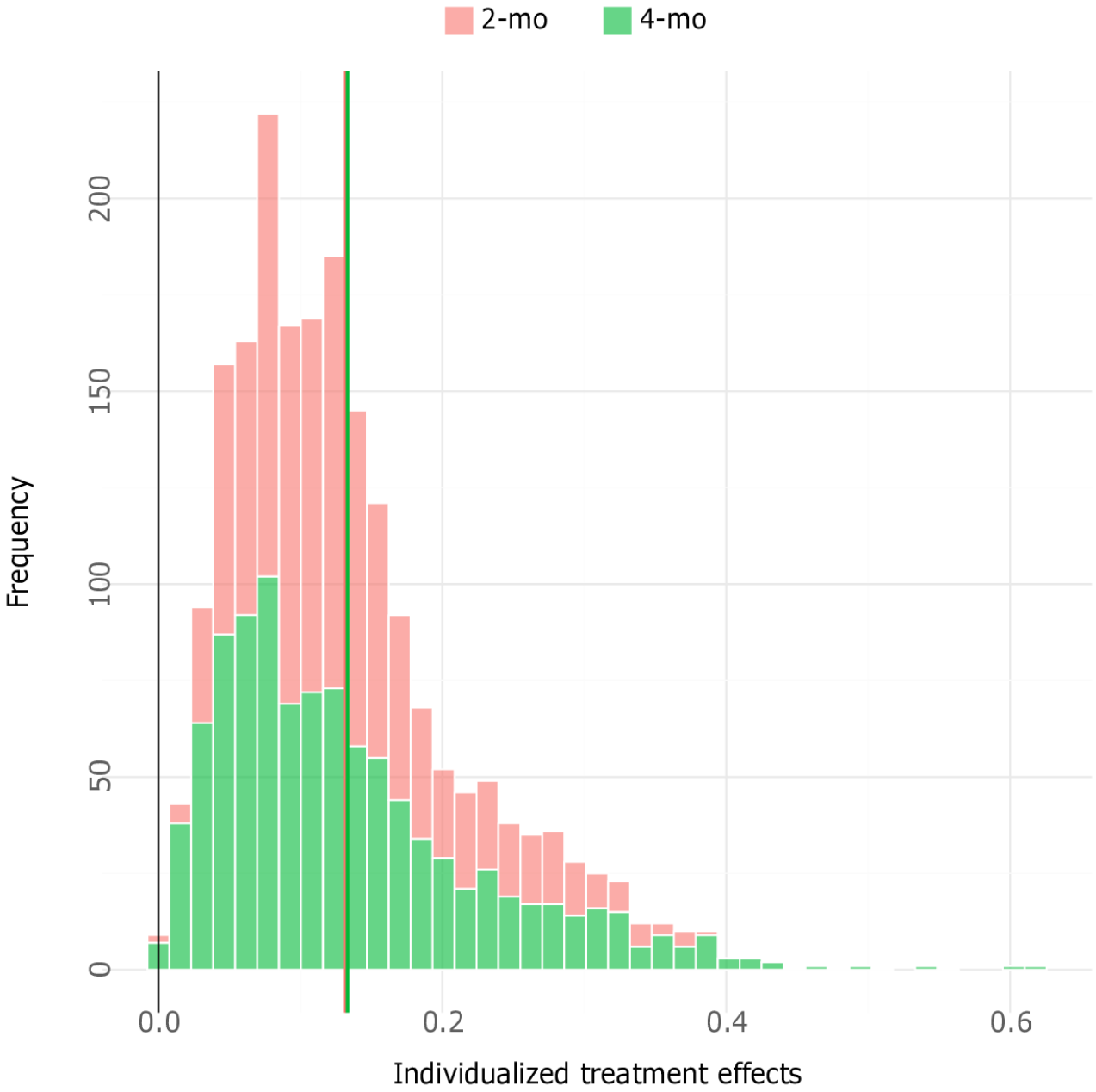
Baseline Factors and Individualized Treatment Effects

In Figure 2, the distribution of effects for 8-week or 5-month abstinence is presented for both the 3- and 6-month follow-up intervals, and Figure 3 provides the distribution of effects for point prevalence abstinence for both follow-ups. The figures show the effects of NEXit at the individual level, expressed as differences in probability (percentage points) of smoking cessation between having received the intervention and the control. From the figures, it is evident that most participants were predicted to benefit from the intervention; however, more

XSL•FO
RenderX

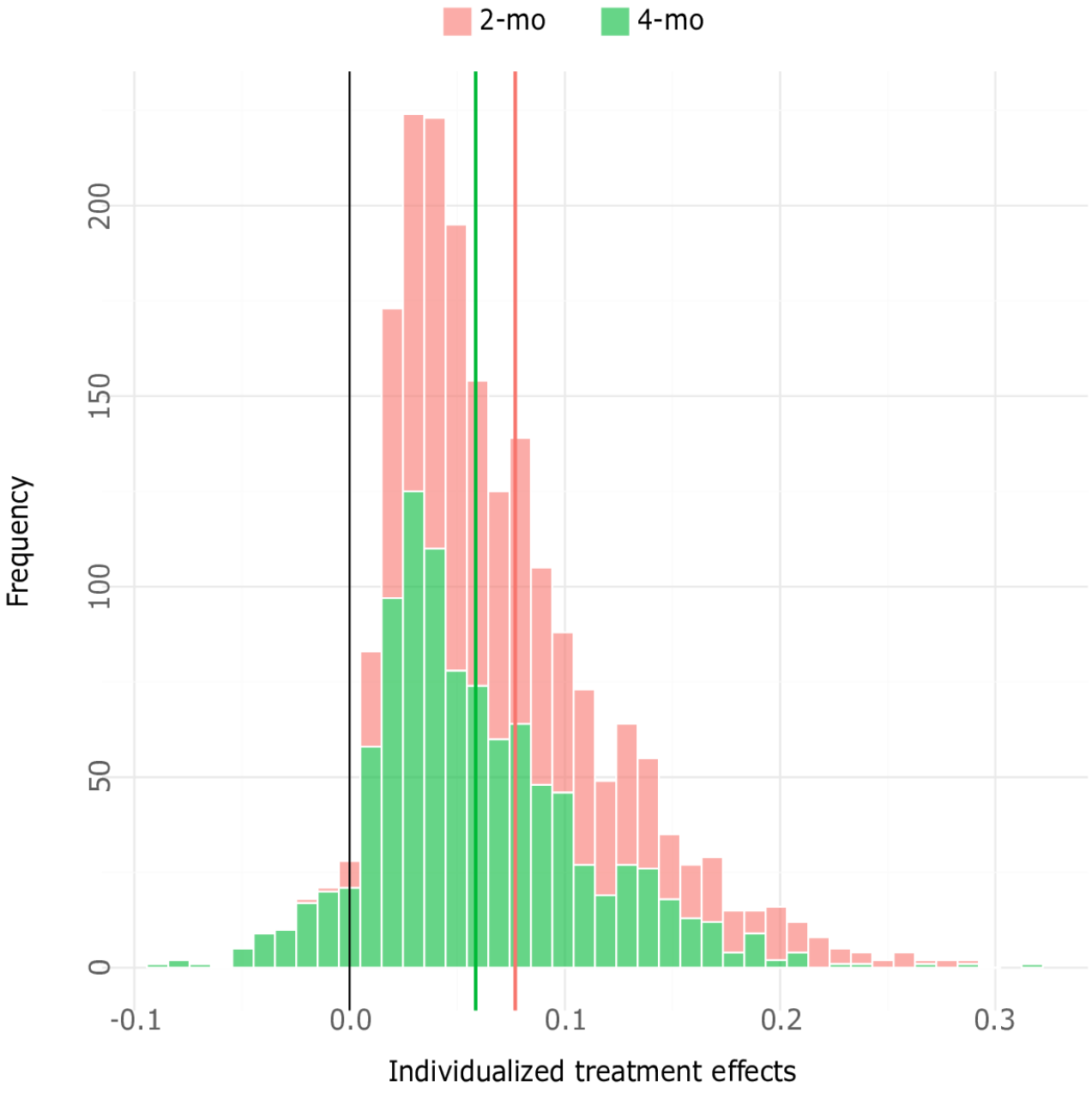
so in terms of prolonged abstinence at the 6-month follow-up interval. The mean individualized effect for point prevalence at the 3-month interval was 3.9 (SD 5.3) percentage points and 2.7 (SD 3.9) percentage points at the 6-month interval. For prolonged abstinence, the mean individualized effect was 7.5 (SD 8.5) percentage points at 3 months and 7.8 (SD 10.7) percentage points at 6 months. This indicates that the intervention was most effective in producing prolonged abstinence, but more importantly, that individualized effects were overdispersed and showed heterogeneity of effects in the population.

Figure 2. Distribution of effects for prolonged abstinence.



XSL•FO
RenderX

Figure 3. Distribution of effects for point prevalence abstinence.



Prolonged Abstinence

In Table 2, associations between baseline factors and standardized individualized treatment effects on prolonged abstinence are presented for both follow-up intervals. For both time points (3 and 6 mo), there was marked evidence for positive associations between individualized effects and gender, age, importance, and confidence. Meanwhile, there was marked evidence for negative associations between individualized effect and number of cigarettes smoked at baseline and years of smoking. Associations between individualized effects and recruitment setting, and past professional cessation support, were found for the 6-month follow-up period.

XSL•FO
RenderX

Table . Baseline factors and standardized individualized effects on prolonged abstinence.

	8 wk		5 mo	
	Est[a], OR[b] (95% CoI[c])	Prob[d] (%)	Est[a], OR (95% CoI)	Prob[d] (%)
Man vs woman	0.34 (0.19 to 0.49)	>99.9	0.18 (0.03 to 0.32)	99.1
Age	0.02 (0.01 to 0.03)	>99.9	0.02 (0.02 to 0.03)	>99.9
Importance	0.09 (0.05 to 0.14)	>99.9	0.1 (0.05 to 0.14)	>99.9
Confidence	0.05 (0.03 to 0.08)	>99.9	0.06 (0.03 to 0.08)	>99.9
Know-how	0.01 (-0.01 to 0.04)	83.7	0.02 (0.0 to 0.05)	97.6
Cigarettes smoked	–0.01 (–0.01 to –0.01)	>99.9	–0.01 (–0.01 to –0.004)	>99.9
Years smoking	–0.01 (–0.02 to –0.004)	99.8	–0.02 (–0.03 to –0.01)	>99.9
Quit attempts	–0.01 (–0.01 to –0.004)	>99.9	0.01 (0.006 to 0.014)	>99.9
Snus, weekly or monthly vs daily	–0.08 (–0.37 to 0.21)	70.5	–0.11 (–0.40 to 0.18)	76.0
Snus, none vs daily	–0.11 (–0.35 to 0.12)	83.1	0.02 (–0.21 to 0.25)	56.1
Fagerström score	0.01 (–0.02 to 0.04)	72.9	0.02 (–0.01 to 0.06)	89.4
Elective surgery	0.63 (0.40 to 0.86)	>99.9	0.50 (0.27 to 0.73)	>99.9
Online vs primary care	–0.09 (–0.37 to 0.20)	71.7	–0.46 (–0.74 to –0.17)	99.9
Quitline	–0.03 (–0.20 to 0.13)	64.2	0.14 (–0.03 to 0.30)	95.1
Past cessation support	–0.12 (–0.27 to 0.03)	93.5	–0.21 (–0.36 to –0.06)	99.7
Current cessation support	0.24 (–0.08 to 0.56)	92.6	0.17 (–0.15 to 0.48)	84.6

[a]Est: median of the posterior distribution of standardized individual effects with 95% CoIs.

[b]OR: odds ratio.

[c]CoI: compatibility interval.

[d]Prob: proportion of the posterior distribution above or below the null in the direction of the median.

Point Prevalence Abstinence

In Table 3, associations between baseline factors and standardized individualized treatment effects on point prevalence are presented for both follow-up intervals. For both time points, there was marked evidence of positive associations between individualized effect and age, importance, confidence, know-how, and elective surgery at the 6-month interval. Negative associations between individualized effects and the number of cigarettes smoked at baseline and years of smoking were found.

XSL•FO
RenderX

Table . Baseline factors and standardized individualized effects on point prevalence.

	3 mo		6 mo	
	Est[a], OR[b] (95% CoI[c])	Prob[d] (%)	Est[a], OR (95% CoI)	Prob[d] (%)
Man vs Woman	0.05 (–0.10 to 0.20)	73.7	0.011 (–0.13 to 0.16)	55.7
Age	0.03 (0.02 to 0.03)	>99.9	0.02 (0.011 to 0.03)	>99.9
Importance	0.05 (0.01 to 0.10)	98.5	0.07 (0.03 to 0.12)	99.9
Confidence	0.05 (0.02 to 0.07)	>99.9	0.04 (0.02 to 0.07)	>99.9
Know-How	0.02 (–0.004 to 0.04)	94.4	0.02 (–0.001 to 0.05)	97.0
Cigarettes smoked	–0.01 (–0.01 to –0.01)	>99.9	–0.01 (–0.01 to –0.01)	>99.9
Years smoking	–0.02 (–0.03 to –0.02)	>99.9	–0.03 (–0.04 to –0.02)	>99.9
Quit attempts	–0.002 (–0.01 to 0.002)	84.6	0.002 (–0.002 to 0.01)	85.7
Snus, weekly or monthly vs daily	–0.03 (–0.32 to 0.26)	57.2	–0.43 (–0.71 to –0.14)	99.9
Snus, none vs daily	–0.12 (–0.35 to 0.11)	84.4	–0.13 (–0.91 to 0.65)	63.7
Fagerström score	–0.02 (–0.05 to 0.01)	86.6	0.01 (–0.02 to 0.04)	75.1
Elective surgery	0.20 (–0.03 to 0.43)	95.7	0.37 (0.15 to 0.60)	99.9
Online vs primary care	0.19 (–0.09 to 0.47)	90.4	0.01 (–0.27 to 0.29)	51.5
Quitline	0.09 (–0.08 to 0.25)	84	0.01 (–0.16 to 0.18)	53.9
Past cessation support	–0.04 (–0.19 to 0.11)	71.7	–0.04 (–0.19 to 0.11)	70.7
Current cessation support	–0.12 (–0.45 to 0.21)	76.1	–0.13 (–0.45 to 0.19)	78.6

[a]Est: Median of the posterior distribution of standardized individual effects with 95% compatibility intervals (CoI).

[b]OR: odds ratio.

[c]CoI: compatibility interval.

[d]Prob: Proportion of the posterior distribution above or below the null in the direction of the median.

Interactions

To assess whether the heterogeneous main associations for age and gender were qualified by interactions, models with added interaction terms were estimated. The only interaction for which there was marked evidence was between gender and importance to quit smoking, which was associated with individualized effects of the intervention on prolonged abstinence. This suggests that the main association of gender for this outcome was accounted for by its interaction with importance. Please see Table 4 for estimates of the interaction and Figure 4 for plots illustrating these interactions.

Table . Interaction associations of gender and baseline importance to quit smoking on prolonged abstinence.

	8 wk		5 mo	
	Est[a], OR[b] (95% CoI[c])	Prob[d] (%)	Est[a], OR (95% CoI)	Prob[d] (%)
Man vs woman	0.004 (–0.12 to 0.28)	56.6	0.0 (–0.22 to 0.18)	50
Importance	0.06 (0.011 to 0.11)	99.4	0.08 (0.03 to 0.13)	>99.9
Man vs woman × importance	0.03 (0.01 to 0.05)	98.4	0.02 (–0.004 to 0.05)	95.4

[a]Est: Median of the posterior distribution of standardized individual effects with 95% compatibility intervals (CoI).

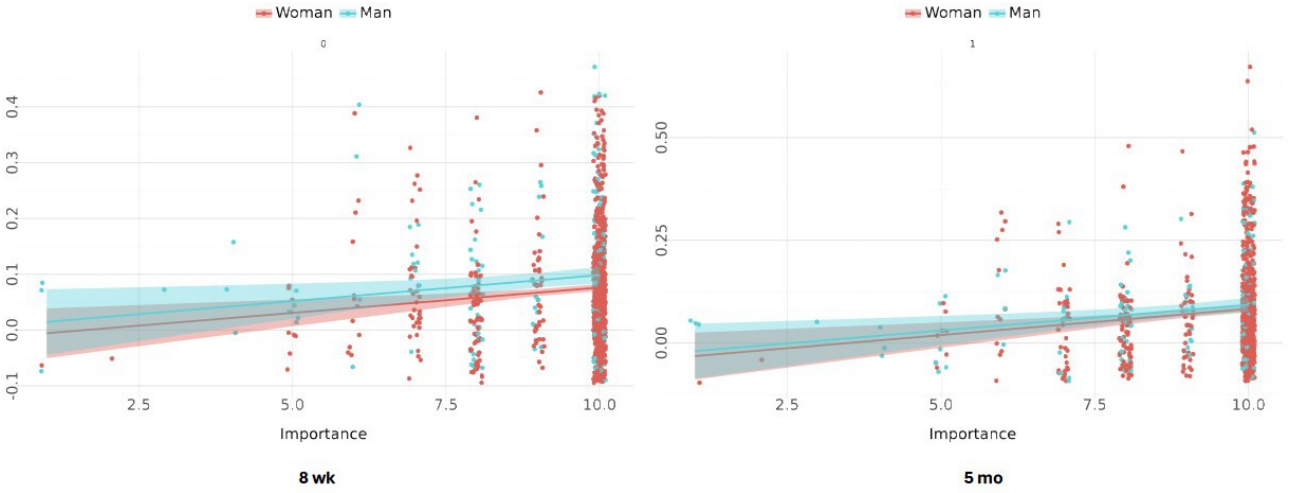
[b]OR: odds ratio.

[c]CoI: compatibility interval.

[d]Prob: Proportion of the posterior distribution above or below the null in the direction of the median.

XSL•FO
RenderX

Figure 4. Visualization of interaction associations of gender and baseline importance to quit smoking on prolonged abstinence.



Discussion

Age

The current study assessed the individualized treatment effects of the recent NEXit SMS text message–based intervention [11]. We found an association between age and individualized effects (ie, as age increases beyond the mean value), with older participants benefiting more than younger participants for both prolonged and point prevalence abstinence. This finding contrasts with evidence comparing effects across trials, which highlights that effective interventions for the adult population in general have similar effects for younger adults [30]. This disparity may be a result of the content of the intervention, which specifically focused on support to restructure the physical environment rather than support for the social environment. This is important, as younger adults have a high prevalence of social smoking [20], which may account for this effect. Future interventions for this subsample could include aspects targeting social factors. Nonetheless, evidence also highlights that younger adults are more likely to initiate cessation compared to older adults and smoke less heavily [13,20]; therefore, the results of the current analysis are encouraging, as they suggest that the subsample who are less likely to quit and who smoke more benefit the most from NEXit. The results also suggest that older adults in Sweden are willing to engage with digital health interventions. This finding contrasts with evidence that suggests a "digital divide" exists between older and younger adults in terms of acceptance and uptake of digital health interventions [31]. The current findings suggest that SMS text message–based interventions may be particularly apt for older adults.

Confidence and Know-How

Baseline confidence was associated with individualized effects for both prolonged and point prevalence abstinence, suggesting that those who had greater confidence in their ability to quit benefited more from NEXit. This finding supports extant literature that highlights higher rates of self-efficacy (ie, confidence) are associated with a greater likelihood of successful smoking cessation [32-34]. The results for know-how varied, with no association found for the first time point for both prolonged and point prevalence abstinence; however, an association was found at the second time point for both outcome measures, with those displaying a higher rating of know-how benefiting more. It may be the case that participants had been lacking knowledge on how to quit and maintain abstinence over a short period of time, while they may have had knowledge on how to sustain abstinence over a longer period. This suggests that NEXit may be beneficial to those who are unsure of how to initially stop smoking and maintain the initial phase of abstinence.

Gender

Associations between both importance and gender with individualized effects were found, with those who deemed it of greater importance to quit benefiting more, while effects were more pronounced for men compared to women. The finding for importance supports extant literature that has assessed motivation to quit (ie, the reasons smokers consider it important to quit), which suggests that higher motivation increases the likelihood of cessation and predicts abstinence [34-37]. Results compared across effectiveness trials highlight that women are less likely to remain abstinent than men [22], and the findings of the current study support this suggestion. Nonetheless, the association between effect and gender dissipated when adding the interaction term of gender and importance to the model, suggesting that, between men and women who deemed it of greater importance to quit, the effects of NEXit were greater for men. Future studies should consider exploring why importance may not be a driving force for change among women but is so for men. It should be noted that the sample included more women than men, reflecting greater treatment seeking among women. The parent trial was pragmatic in the sense that it was an effectiveness trial recruiting participants in a way that mimicked how a real-world implementation could work (ie, online advertisements). Therefore, our findings should most conservatively be generalized to those seeking help online, and not the general population, where the prevalence of smoking is more balanced.

On the other hand, for point prevalence, there were no marked associations between effects or interactions with respect to gender. This suggests that rates of relapse (or lack of) may have been similar between women and men, which is in contrast to evidence suggesting that women are more likely to relapse than

XSL•FO
RenderX

men [21]. The content for craving provided by NEXit may have been equally effective for the needs of both genders; evidence suggests that affective factors such as mood, anger, and perceived stress predict relapse in women, while the motivation to reduce craving is a unique relapse predictor in men [38]. The support for craving had specific content relating to stress management, reframing perspectives, and motivation.

The results highlight that confidence and importance (especially for men) are potential focal considerations for smoking cessation. Confidence and importance are integral parts of the decision to quit smoking, especially for individuals who are contemplating or preparing for a change and who may feel ambivalent about quitting [39]. It may be that those who reported the importance of quitting and their confidence as high were better psychologically prepared to change behavior. This suggests that future interventions should target improvements in both factors.

Smoking and Number of Quit Attempts

Associations between the number of cigarettes smoked and years of smoking with individualized effects were found for both prolonged and point prevalence abstinence, with those who smoked less and for a shorter duration benefiting the most. This suggests that NEXit is less beneficial for those who smoke more and have done so for a longer duration. Theories of addiction posit that when an addiction progresses, such as nicotine dependence, control over consumption shifts from goal-directed behavior to automatic or habitual motivation [40], hence those who smoke more and for longer *may* have been driven more by habit than goal-directed behavior. While the content of NEXit did include advice on how to avoid smoking cues, this may not have been enough for these heavier, longer-term smokers. Providing greater support for identifying what triggers habitual smoking and advice on how to plan for encountering triggers or cues may be beyond the scope of an SMS text message–based intervention. Potentially, this could be better achieved by smoking cessation counseling or through the usage of an interactive digital smoking cessation intervention, whereby end users are encouraged to record how they are triggered, along with creating their own plans to overcome their triggers, which can then be sent as reminders.

For prolonged abstinence, an association between quit attempts and individualized effect was found at 8 weeks, with those having fewer attempts at baseline benefiting more. The underlying action of NEXit may account for this finding, with the intervention theorized to be driven by instilling confidence in ability and raising the importance of quitting. Evidence highlights that self-efficacy (ie, confidence) and motivation (ie, importance) are key factors in smoking cessation for individuals who are less likely to make a quit attempt [37]. It may be that the mediating effects of confidence and importance on prolonged abstinence had a greater impact on this group than on those with greater recorded quit attempts at baseline. Nonetheless, at the 5-month follow-up, an individualized effect on prolonged abstinence was found for those reporting more quit attempts at baseline. This effect is potentially driven by these individuals building more confidence and gaining knowledge for quitting over the course of the trial.

Degree of Dependence

Finally, for the dependence score, only an association with individualized effect for point prevalence at 3 months was found, with those reporting less dependence benefiting more. No other associations with dependence were observed. Dependence as a factor in cessation and abstinence is well established; the likelihood of quitting is much greater at low levels of dependence, in comparison to high levels [41]. The disparity in our findings with the literature may be due to the methods used to assess dependency. We used a self-report measure, whereas many studies use biochemical measures of dependence [42]. While the Fagerström scale has been shown to predict abstinence across trials [24], self-report measures are more prone to biases in comparison to objective measures such as a saliva cotinine test. While future studies should consider both self-reports and biochemical measures, it may not be feasible for the digital modality. Furthermore, evidence highlights selection biases present in saliva testing [43], while recommendations suggest self-report measures are more appropriate for general population studies [43]. Nonetheless, our results for dependence could be interpreted to mean that, by and large, NEXit is equally effective regardless of dependence severity.

Limitations

The results provide insight into the individualized effects of the recent NEXit trial; however, they should be interpreted with some considerations. Typically in RCTs, the randomized component is tested, rather than a comparison across treatments, meaning individuals may adhere to or engage in an intervention at their own preference. Therefore, the results should be considered in terms of having access to the intervention rather than using it. Indeed, the aim of the trial was to estimate the effectiveness of providing a digital smoking cessation intervention; therefore, all intercurrent events were acceptable. We do note that since we did not collect data on the use of other interventions throughout the trial period, we cannot investigate if the NEXit intervention increased the utilization of other treatments.

Another limitation of this study is that a statistical model was used to predict and contrast outcomes. This means that the estimated effects are not based entirely on observed data but on predicted data. These predictions may have considerable uncertainty that is not readily quantifiable, as no individuals received both conditions; thus, no individual contrasts exist against which predictions can be compared. The result of this is that when comparing individualized effects against baseline factors, there is uncertainty that is not accounted for. More research is needed to develop methods to estimate the uncertainty of the statistical approach taken [27].

It should be noted that the relationships between baseline characteristics and individualized effects are associational and should not be interpreted as causal. Confounding and collider bias can substantially alter estimated associations. Further, the current study investigated several associations between baseline characteristics and individualized effects. The Bayesian inference paradigm informs which associations are stronger and weaker relative to one another and which associations have the

XSL•FO
RenderX

relatively strongest evidence, conditional on the data and prior knowledge. By using Student *t* test priors centered at 0, there was weak shrinkage induced in these estimations to protect against overestimation. Thus, the result of this study is all the associations presented and not only those that pass a prespecified threshold of significance (eg, null-hypothesis testing). In all, the findings presented in this study should be seen as exploratory, and future studies should look to confirm the associations as causal.

Conclusions

In conclusion, the results of this study highlight that individuals participating in the RCT of the NEXit intervention had heterogeneous outcomes, which were not discernible from the primary analysis of the RCT that compared the intervention and control groups [11]. The intervention was most effective among men, older individuals, those who smoked less at baseline and for a shorter duration, and those who had confidence in their ability and deemed it important to quit. The results have the potential to help refine smoking cessation interventions so that we can target those who persist in smoking despite current intervention efforts and policy measures.

Acknowledgments

Artificial intelligence chatbots were not used in any part of the manuscript writing.

Funding

This study was conducted under the auspices of a grant from The Swedish Cancer Society (Cancerfonden, 23 2888; principal investigator: MB). The parent trial was funded by Region Östergötland (Forsknings-ALF, Dnr: LIO-896081). The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the General Data Protection Regulation but are available upon reasonable request to marcus.bendtsen@liu.se.

Authors' Contributions

Data analysis: MB
Data collection: JB, MB
Editing manuscript: KUG, PB, MB
Writing – lead and editing: JC

Conflicts of Interest

MB and PB own a private company (Alexit AB) that maintains and distributes evidence-based lifestyle interventions to be used by the public and in health care settings. Alexit AB played no role in developing the intervention, study design, data analysis, data interpretation, or writing of this report. Services developed and maintained by Alexit AB were used for sending SMS text messages and data collection. All other authors declare no conflicts of interest.

Checklist 1
CONSORT checklist.
[PDF File, 239 KB - mhealth_v14i1e63578_app1.pdf]

References

1. Reitsma MB, Kendrick PJ, Ababneh E, GBD 2019 Tobacco Collaborators. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990-2019: a systematic analysis from the Global Burden of Disease Study 2019. Lancet 2021 Jun 19;397(10292):2337-2360. [doi: 10.1016/S0140-6736(21)01169-7] [Medline: 34051883]
2. Tobacco. World Health Organization. URL: https://www.who.int/health-topics/tobacco#tab=tab_1 [accessed 2026-01-06]
3. Vuxnas bruk av tobaks- och nikotinprodukter [Article in Swedish]. Folkhälsomyndigheten. URL: https://www.folkhalsomyndigheten.se/livsvillkor-levnadsvanor/andts/andts-anvandning-och-ohalsa/anvandning-och-omfattning-av-andts-i-befolkningen/anvandning-av-tobaks-och-nikotinprodukter/vuxnas-bruk-av-tobaks--och-nikotinprodukter/ [accessed 2025-01-06]
4. Tobacco and nicotine products. Public Health Agency of Sweden. URL: https://www.folkhalsomyndigheten.se/the-public-health-agency-of-sweden/living-conditions-and-lifestyle/andtg/tobacco/ [accessed 2026-01-06]

XSL•FO
RenderX

5. Molarius A, Hellstrand M, Engström S. Social differences in who receives questions and advice about smoking habits when visiting primary care - results from a population based study in Sweden in 2012. Prev Med Rep 2017 Mar;5:236-240. [doi: 10.1016/j.pmedr.2016.12.016] [Medline: 28127526]
6. Spohr SA, Nandy R, Gandhiraj D, Vemulapalli A, Anne S, Walters ST. Efficacy of SMS text message interventions for smoking cessation: a meta-analysis. J Subst Abuse Treat 2015 Sep;56:1-10. [doi: 10.1016/j.jsat.2015.01.011] [Medline: 25720333]
7. Scott-Sheldon LAJ, Lantini R, Jennings EG, et al. Text messaging-based interventions for smoking cessation: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016 May 20;4(2):e49. [doi: 10.2196/mhealth.5436] [Medline: 27207211]
8. Mason M, Ola B, Zaharakis N, Zhang J. Text messaging interventions for adolescent and young adult substance use: a meta-analysis. Prev Sci 2015 Feb;16(2):181-188. [doi: 10.1007/s11121-014-0498-7] [Medline: 24930386]
9. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev 2019 Oct 22;10(10):CD006611. [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
10. Zhou X, Wei X, Cheng A, et al. Mobile phone-based interventions for smoking cessation among young people: systematic review and meta-analysis. JMIR Mhealth Uhealth 2023 Sep 12;11:e48253. [doi: 10.2196/48253] [Medline: 37706482]
11. Blomqvist J, Gunnarsson KU, Bendtsen P, Bendtsen M. Effects of a text messaging smoking cessation intervention amongst online help-seekers and primary health care visitors: findings from a randomised controlled trial. BMC Med 2023 Oct 4;21(1):382. [doi: 10.1186/s12916-023-03073-5] [Medline: 37794399]
12. Bendtsen M, Thomas K, Linderoth C, Bendtsen P. Effects of a text messaging smoking cessation intervention among online help seekers and primary health care visitors in Sweden: protocol for a randomized controlled trial using a Bayesian group sequential design. JMIR Res Protoc 2020 Dec 3;9(12):e23677. [doi: 10.2196/23677] [Medline: 33269703]
13. Arancini L, Borland R, Le Grande M, et al. Age as a predictor of quit attempts and quit success in smoking cessation: findings from the International Tobacco Control Four-Country survey (2002-14). Addiction 2021 Sep;116(9):2509-2520. [doi: 10.1111/add.15454] [Medline: 33651412]
14. Nollen NL, Mayo MS, Sanderson Cox L, et al. Factors that explain differences in abstinence between Black and White smokers: a prospective intervention study. J Natl Cancer Inst 2019 Oct 1;111(10):1078-1087. [doi: 10.1093/jnci/djz001] [Medline: 30657926]
15. Kock L, Brown J, Hiscock R, Tattan-Birch H, Smith C, Shahab L. Individual-level behavioural smoking cessation interventions tailored for disadvantaged socioeconomic position: a systematic review and meta-regression. Lancet Public Health 2019 Dec;4(12):e628-e644. [doi: 10.1016/S2468-2667(19)30220-8] [Medline: 31812239]
16. Boudrez H. Psychological factors and long-term abstinence after smoking cessation treatment. J Smok Cessat 2009 May 1;4(1):10-17. [doi: 10.1375/jsc.4.1.10]
17. Hallengren L. Rökfritt sverige 2025 [Report in Swedish]. : Sveriges Riksdag; 2017 URL: https://data.riksdagen.se/fil/4E855498-EF93-426C-9420-E13DA6E88E18 [accessed 2026-01-06]
18. Eldredge LKB, Markham CM, Ruiter RA, Fernandez ME. Planning Health Promotion Programs: An Intervention Mapping Approach, 4th edition: Jossey-Bass Inc; 2016. URL: https://cris.maastrichtuniversity.nl/en/publications/planning-health-promotion-programs-an-intervention-mapping-approa/ [accessed 2026-01-06]
19. Miche S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions, 1st edition: Silverback Publishing; 2014. URL: https://www.behaviourchangewheel.com/about-book [accessed 2026-01-06]
20. Song AV, Ling PM. Social smoking among young adults: investigation of intentions and attempts to quit. Am J Public Health 2011 Jul;101(7):1291-1296. [doi: 10.2105/AJPH.2010.300012] [Medline: 21566040]
21. Swan GE, Ward MM, Carmelli D, Jack LM. Differential rates of relapse in subgroups of male and female smokers. J Clin Epidemiol 1993 Sep;46(9):1041-1053. [doi: 10.1016/0895-4356(93)90172-w] [Medline: 8263577]
22. Smith PH, Bessette AJ, Weinberger AH, Sheffer CE, McKee SA. Sex/gender differences in smoking cessation: a review. Prev Med 2016 Nov;92:135-140. [doi: 10.1016/j.ypmed.2016.07.013] [Medline: 27471021]
23. Gram IT, Antypas K, Wangberg SC, Løchen ML, Larbi D. Factors associated with predictors of smoking cessation from a Norwegian internet-based smoking cessation intervention study. Tob Prev Cessat 2022;8:38. [doi: 10.18332/tpc/155287] [Medline: 36382026]
24. Fagerström K, Russ C, Yu CR, Yunis C, Foulds J. The Fagerström test for nicotine dependence as a predictor of smoking abstinence: a pooled analysis of varenicline clinical trial data. Nicotine Tob Res 2012 Dec;14(12):1467-1473. [doi: 10.1093/ntr/nts018] [Medline: 22467778]
25. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
26. Maddux JE, Rogers RW. Protection motivation and self-efficacy: a revised theory of fear appeals and attitude change. J Exp Soc Psychol 1983 Sep;19(5):469-479. [doi: 10.1016/0022-1031(83)90023-9]
27. Hoogland J, IntHout J, Belias M, et al. A tutorial on individualized treatment effect prediction from randomized trials with a binary endpoint. Stat Med 2021 Nov 20;40(26):5961-5981. [doi: 10.1002/sim.9154] [Medline: 34402094]
28. Bendtsen M. A gentle introduction to the comparison between null hypothesis testing and Bayesian analysis: reanalysis of two randomized controlled trials. J Med Internet Res 2018 Oct 24;20(10):e10873. [doi: 10.2196/10873] [Medline: 30148453]

XSL•FO
RenderX

29. Stan Development Team. Stan documentation. Stan. 2023. URL: https://mc-stan.org/docs/ [accessed 2026-01-06]
30. Suls JM, Luger TM, Curry SJ, Mermelstein RJ, Sporer AK, An LC. Efficacy of smoking-cessation interventions for young adults: a meta-analysis. Am J Prev Med 2012 Jun;42(6):655-662. [doi: 10.1016/j.amepre.2012.02.013] [Medline: 22608385]
31. McDonough CC. The effect of ageism on the digital divide among older adults. Gerontol Geriatr Med 2016 Jun 16;2(1):1-7. [doi: 10.24966/GGM-8662/100008]
32. Gallus S, Cresci C, Rigamonti V, et al. Self-efficacy in predicting smoking cessation: a prospective study in Italy. Tob Prev Cessat 2023;9(April):15. [doi: 10.18332/tpc/162942] [Medline: 37125003]
33. Klemperer EM, Mermelstein R, Baker TB, et al. Predictors of smoking cessation attempts and success following motivation-phase interventions among people initially unwilling to quit smoking. Nicotine Tob Res 2020 Aug 24;22(9):1446-1452. [doi: 10.1093/ntr/ntaa051] [Medline: 32236417]
34. Boardman T, Catley D, Mayo MS, Ahluwalia JS. Self-efficacy and motivation to quit during participation in a smoking cessation program. Int J Behav Med 2005;12(4):266-272. [doi: 10.1207/s15327558ijbm1204_7] [Medline: 16262545]
35. Wee LH, West R, Bulgiba A, Shahab L. Predictors of 3-month abstinence in smokers attending stop-smoking clinics in Malaysia. Nicotine Tob Res 2011 Feb;13(2):151-156. [doi: 10.1093/ntr/ntq221] [Medline: 21186253]
36. Williams GC, Gagné M, Ryan RM, Deci EL. Facilitating autonomous motivation for smoking cessation. Health Psychol 2002 Jan;21(1):40-50. [doi: 10.1037/0278-6133.21.1.40] [Medline: 11846344]
37. Jardin BF, Carpenter MJ. Predictors of quit attempts and abstinence among smokers not currently interested in quitting. Nicotine Tob Res 2012 Oct;14(10):1197-1204. [doi: 10.1093/ntr/nts015] [Medline: 22387995]
38. Nakajima M, al'Absi M. Predictors of risk for smoking relapse in men and women: a prospective examination. Psychol Addict Behav 2012 Sep;26(3):633-637. [doi: 10.1037/a0027280] [Medline: 22352701]
39. DiClemente CC, Prochaska JO. Self-change and therapy change of smoking behavior: a comparison of processes of change in cessation and maintenance. Addict Behav 1982;7(2):133-142. [doi: 10.1016/0306-4603(82)90038-7] [Medline: 7102444]
40. Doñamayor N, Ebrahimi C, Arndt VA, Weiss F, Schlagenhauf F, Endrass T. Goal-directed and habitual control in human substance use: state of the art and future directions. Neuropsychobiology 2022;81(5):403-417. [doi: 10.1159/000527663] [Medline: 36349761]
41. Caponnetto P, Polosa R. Common predictors of smoking cessation in clinical practice. Respir Med 2008 Aug;102(8):1182-1192. [doi: 10.1016/j.rmed.2008.02.017] [Medline: 18586479]
42. Jarvis MJ, Sutherland G. Tobacco smoking. In: Bellack AS, Hersen M, editors. Comprehensive Clinical Psychology: Pergamon; 1998:645-674. [doi: 10.1016/B0080-4270(73)00101-2]
43. SRNT Subcommittee on Biochemical Verification. Biochemical verification of tobacco use and cessation. Nicotine Tob Res 2002 May;4(2):149-159. [doi: 10.1080/14622200210123581] [Medline: 12028847]

Abbreviations

CoI: compatibility interval
CONSORT: Consolidated Standards of Reporting Trials
NEXit: Nicotine Exit
OR: odds ratio
RCT: randomized controlled trial

Edited by L Buis; submitted 24.Jun.2024; peer-reviewed by H Tateno, K Bold, MC Fahey, M Mansour, T Lid; revised version received 21.Nov.2025; accepted 28.Dec.2025; published 11.Feb.2026.

Please cite as:
Crawford J, Blomqvist J, Ulfsdotter Gunnarsson K, Bendtsen P, Bendtsen M
Individualized Treatment Effects of a Digital Smoking Cessation Intervention Among Individuals Looking Online for Help: Secondary Analysis of a Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e63578
URL: https://mhealth.jmir.org/2026/1/e63578
doi:10.2196/63578

© Joel Crawford, Jenny Blomqvist, Katarina Ulfsdotter Gunnarsson, Preben Bendtsen, Marcus Bendtsen. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

WeChat-Based Intervention for Glycemic Control in Patients With Type 2 Diabetes Mellitus: Multicenter Randomized Controlled Trial

Chuanfen Zheng[1], MM; Xiong Dou[1], MM; Xiaotao Xiong[1], MM; En-Yu Lei[1], BSc; Ming Jiang[2], MM; Yulin Wu[2], MM; Jing Yu[3], BSc; Xianjun Wang[4], MM; Ling Zhang[1], MM; Honghui Rong[1], MM; Lu Lu[1], MM; Fengju Li[1], MM; Ting Luo[1], MM; Xiangyu Ma[5], PhD; Ji-An Chen[1], PhD

[1]Department of Health Education, College of Preventive Medicine, Army Medical University, 30 Gaotanyan Street, Shapingba District, Chongqing, China

[2]Xinqiao Community Health Service Center, Shapingba District, Chongqing, China

[3]Chenjiaqiao Community Health Service Center, Shapingba District, Chongqing, China

[4]Yubeilu Community Health Service Center, Shapingba District, Chongqing, China

[5]Department of Epidemiology, College of Preventive Medicine, Army Medical University, Chongqing, China

Corresponding Author:
Ji-An Chen, PhD
Department of Health Education, College of Preventive Medicine, Army Medical University, 30 Gaotanyan Street, Shapingba District, Chongqing, China

Abstract

Background: China's diabetes epidemic faces critical gaps in glycemic control, with only 50.1% of treated patients achieving hemoglobin A_{1C} (HbA_{1c}) targets in 2021. Conventional interventions struggle with scalability in primary care, particularly for vulnerable populations.

Objective: This study aimed to evaluate the use of a WeChat-based health education tool (the WeWalk mini program, the Bayu Health public account, and a WeChat group) for improving glycemic control in community-dwelling patients with type 2 diabetes mellitus.

Methods: This multicenter randomized controlled trial enrolled 600 adults with type 2 diabetes from 3 communities in Chongqing, randomly allocating participants 1:1 to either a 12-week WeChat-based intervention (n=300, 50%) or a control group (n=300, 50%) in September 2020. The control group received 4 face-to-face traditional health education sessions, whereas the intervention group participated in a digital program: a 4-week course followed by an 8-week practical implementation. At baseline and 12 weeks after the intervention began, both groups were examined in terms of HbA_{1c} and fasting blood glucose (FBG) as the primary outcomes, as well as variables such as blood lipid profile, blood pressure, and physical fitness–related indexes as secondary outcomes. Longitudinal glycemic control was assessed through triplicate FBG measurements extracted from standardized electronic health records at the 2-year follow-up. Independent *t* tests or Mann-Whitney *U* tests were used to assess changes from baseline to follow-up between groups.

Results: A total of 92.7% (556/600) of the participants completed the 12-week follow-up visit. The WeChat-based intervention demonstrated superior glycemic control outcomes, with intervention participants achieving a 0.59% greater HbA_{1c} reduction than controls (−0.03% vs 0.56%; *P*<.001) and significant improvements in FBG levels (−0.69 vs 0.00 mmol/L; Δ=0.69; *P*=.001). Subgroup analysis revealed that WeChat-based health education was significantly effective in patients with diabetes with a disease duration of <10 years, educational level of junior high school or lower, and annual family income of <CN ¥50,000 (US $7172.10) . These benefits persisted throughout the 2-year follow-up, where the intervention group maintained lower FBG levels (6.87 vs 7.35 mmol/L; *P*=.001).

Conclusions: WeChat-based health education was beneficial for glycemic control in primary health care settings. However, the sustained efficacy and feasibility of this approach require further investigation.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2300071926; https://www.chictr.org.cn/showprojEN.html?proj=172083

(JMIR Mhealth Uhealth 2026;14:e80738) doi:10.2196/80738

XSL•FO
RenderX

KEYWORDS

WeChat; glycemic control; type 2 diabetes; randomized controlled trial; WeChat mini program; health education tool

Introduction

In recent years, the prevention and control of diabetes mellitus have become a global public health challenge, especially in China [1]. Effective management of type 2 diabetes mellitus (T2DM) relies critically on successful patient self-management [2], with structured diabetes self-management education (DSME) serving as a cornerstone of this process [3,4]. Despite this recognition, only 32.9% of patients with T2DM in China receive treatment, and among those treated, only 50.1% achieve adequate glycemic control [5]. This situation imposes substantial financial burdens on both patients and community health care systems [6,7].

Robust evidence from high-income countries, including Sweden [8], Spain [9], the United Kingdom [10], and New Zealand [11], has consistently demonstrated the efficacy of structured DSME programs in improving patient outcomes within primary care settings. These programs often emphasize lifestyle modifications, including balanced nutrition, regular physical activity, and weight management, as advocated by the World Health Organization [12]. Nevertheless, the traditional DSME model, which relies heavily on face-to-face and resource-intensive sessions, faces significant scalability challenges within the vast and heterogeneous primary health care environment [13]. Barriers to adopting these behaviors persist, including limited access to personalized health education and fragmented communication between patients and health care providers [14]. Therefore, building effective intervention systems for diabetes health education has become an important issue in primary health care in China.

The rapid proliferation of mobile health technologies has ushered in new opportunities for scalable diabetes management. Mobile apps have emerged as a cost-effective and efficient means of managing diabetes, primarily focusing on self-management and hospital appointments [15-17]. However, these apps often lack broad applicability for patients and family physicians in primary health care settings, neglecting the holistic needs of the community [18]. In contrast, ubiquitous social media platforms such as WeChat present a unique opportunity [19,20]. Their high penetration rate and familiar interface facilitate the delivery of structured educational content and enable continuous, interactive communication between patients and health care providers directly within the primary care context. Therefore, we designed an online free health education platform incorporating a WeChat mini program named WeWalk and a WeChat public account named Bayu Health.

This randomized controlled trial aimed to evaluate the effectiveness of WeChat-based interventions among patients with T2DM in Chinese community health centers. Our hypothesis was that mobile health education, which is based on the health belief model and social cognitive theory, would be more effective than conventional health education. We expected it to improve communication with community family physicians; foster peer-supportive social networks; promote healthy lifestyle changes; and ultimately, improve the effectiveness of diabetes self-management. By addressing these gaps and leveraging the potential of mobile health technology, this study contributes to the global effort in diabetes management.

Methods

Design

This study was designed as a 12-week, nonblinded randomized controlled trial to examine the efficacy of a WeChat-based intervention in patients with T2DM between September 2020 and December 2020. These participants were recruited from 3 community health service centers and followed up on at 12 weeks and 2 years after the intervention. Family physicians followed up on patients with diabetes via telephone to introduce the trial and determine potential participants' willingness to take part. These participants had established health records and had signed up for contracted family physician services at 3 community health service centers in Chongqing, China. Before the trial started, all the researchers involved were trained in uniform protocols, including trial procedures, questionnaires, and the use of the WeChat mini program and public account. A CONSORT (Consolidated Standards of Reporting Trials) checklist was completed to guide the trial's design, analysis, and findings (Checklist 1). Before conducting the formal research, we carried out a feasibility study, as shown in Multimedia Appendix 1.

Participants

On the basis of the inclusion and exclusion criteria, the family physician teams selected eligible participants for registration. The inclusion criteria for enrollment were as follows: (1) participants who were diagnosed with T2DM and met the 1999 World Health Organization diagnostic criteria; (2) receipt of the contracted family physician service within the validity period; (3) ability of the participants to undergo face-to-face interviews and voluntary participation in the trial; and (4) ownership of a smartphone and familiarity with the WeChat app, including chatting, subscribing to WeChat public accounts, and using WeChat mini programs. The exclusion criteria for this study were as follows: (1) participants with severe complications, such as ketoacidosis or hyperosmolar nonketotic diabetic coma; (2) participants with severe heart disease who would not tolerate moderate-intensity exercise; (3) participants who could not answer the questions clearly; and (4) participants who were taking part in other intervention studies.

A total of 900 eligible patients with T2DM were recruited by family physician teams from 3 communities, of whom 300 (33.3%) were excluded. Of these 300 excluded patients, 196 (65.3%) did not meet the inclusion criteria, 79 (26.3%) refused to participate, and 25 (8.3%) were excluded for other reasons. A total of 600 participants were randomized into the intervention group (n=300, 50%) or the control group (n=300, 50%). Moreover, 556 participants completed the 12-week follow-up

XSL•FO
RenderX

visit (intervention: n=278, 50%; control: n=278, 50%), and the retention rate was 92.7% (556/600).

Sample Size

The sample size was calculated based on a completely random design using the tests for 2 proportions. The trial was designed for analysis via 2-tailed tests, with type I and II error rates set at 0.05 and 0.2, respectively. A priori sample size calculations were conducted for the primary outcome of hemoglobin A_{1C} (HbA_{1c}) control rate (proportion of <7%). On the basis of national epidemiological data [1], the retention rate in the control group was estimated to be 35.2% in this study. We hypothesized that an 11% absolute risk reduction in the intervention group (24.2% event rate) derived from conservative adaptation of the Finnish Diabetes Prevention Study [21] would account for our digital intervention modality. With these parameters in the PASS software (version 15.0; NCSS, LLC), the total sample size required was 267 in each group. Considering a dropout rate of up to 10%, the final sample size was determined to be 300 patients in each group.

Allocation

After the willingness of patients to participate in the study was confirmed by the family physician teams, the baseline data of the participants were collected at the community health service centers. Randomization was performed by an independent statistician who generated the allocation sequence through computerized random number generation. The participants were assigned in a 1:1 ratio to the study groups, with allocation concealment ensured via opaque, sealed envelopes that were sequentially numbered.

Diabetes Health Education Team

Our multidisciplinary diabetes health education teams [22] were composed of the following: (1) a total of 18 family physicians from 6 family physician teams at 3 community health centers who were responsible for directly managing patients with diabetes and providing sustainable health education services; (2) an online health assistant who was responsible for functional maintenance and training on the health education platform; and (3) health education experts from the Military Health Education Center, including endocrinologists, rehabilitation therapists, ophthalmologists, dietitians, and coaches.

WeChat-Based Health Education Tool

The WeChat-based health education tool used in our study consisted of the WeWalk mini program, the Bayu Health public account, and the WeChat group. The WeWalk mini program, an interactive mobile health platform, integrates 5 core functional modules designed to promote diabetes self-management through physical activity monitoring and community engagement. This gamified intervention system uses behavioral economics principles and social cognitive theory to enhance users' health literacy and sustainable behavior change. Participants in the WeChat-based intervention group carried out daily health education activities to promote healthy behaviors. The WeChat public account Bayu Health was used to push diabetes self-management courses for the intervention group. The WeChat group was used mainly to organize and guide patients to participate in intervention activities. Multimedia Appendix 2 presents the functional modules of the WeWalk mini program and shows the QR codes of the WeWalk mini program and the Bayu Health public account.

Intervention Protocol

Control Group

The participants in the control group received standard health education, consisting of 4 sessions over a 12-week period provided by their family physician team. The contents of the education intervention focused on 5 key areas: diabetes knowledge, dietary control, exercise therapy, drug therapy, and self-monitoring of blood glucose. Education was provided during regular monthly follow-up visits or scheduled clinic appointments through in-person consultations.

Intervention Group

Participants in the intervention group received a 12-week theory-based diabetes self-management program delivered through an integrated WeChat platform. The intervention was structured in 2 sequential, theory-informed phases to facilitate knowledge acquisition and sustainable behavior change.

Stage 1: Structured Education (Weeks 1 - 4)

During the initial 4-week phase, the participants completed a structured self-management curriculum via the Bayu Health public account. Developed by the Military Health Education Center and aligned with China's national guidelines, the curriculum incorporated the health belief model to increase perceived susceptibility, severity, benefits, and self-efficacy. It consists of 30 evidence-based microlearning modules (videos and articles) covering lifestyle modification, glycemic control, medication adherence, and complication prevention. Health assistants reinforced engagement through daily performance summaries and the recognition of active participants in dedicated WeChat groups.

Stage 2: Guided Application (Weeks 5 - 12)

The subsequent 8-week phase focused on translating knowledge into practice via the WeWalk mini program. Grounded in principles of gamification and social cognitive theory, this stage included physician-supervised goal setting (Multimedia Appendix 3), photo-based tracking of physical activity and dietary patterns, and a gamified ranking system with virtual health points to incentivize consistent engagement. All user-generated health data were shared exclusively with the family physician teams via encrypted channels, enabling real-time clinical oversight while maintaining privacy.

Data Collection and Measurements

The participants underwent 2 follow-up assessments: after the intervention (12 weeks) and in the long term (2 years after the intervention). To assess the condition of the patients, the primary outcome measure of the study was the change in fasting blood glucose (FBG) and HbA_{1c} levels in the 2 groups at baseline and 12 weeks after the intervention began. The secondary outcomes included changes in weight, BMI, waist circumference, hip circumference, systolic blood pressure, diastolic blood pressure, and serum lipid profiles. Baseline assessments included

XSL•FO
RenderX

standardized questionnaires capturing demographics, medical and treatment history, comorbidities, and lifestyle factors, alongside fasting-state physical examinations conducted by certified personnel at community health service centers. Longitudinal glycemic control was assessed through triplicate FBG measurements extracted from standardized electronic health records at the 2-year follow-up.

Statistical Analysis

Data were double entered using EpiData (EpiData Association) and analyzed using SPSS (version 26; IBM Corp). Normally distributed variables expressed as means and SDs were compared using independent *t* tests; nonparametric variables reported as medians and IQRs were analyzed via Mann-Whitney *U* tests. Demographic (sex, educational level, and annual family income), clinical (hypoglycemic medication use and diabetic complications), and behavioral (smoking and alcohol use) variables were analyzed using the chi-square test at baseline. The chi-square test was used to test the differences in HbA_{1c} data before and after the intervention. Patients with diabetes were divided into 2 categories according to their HbA_{1c} levels: HbA_{1c}<7% (normal) and HbA_{1c}≥7% (abnormal) according to the guidelines for the prevention and treatment of T2DM in China (2017) [23]. P<.05 was considered statistically significant.

Ethical Considerations

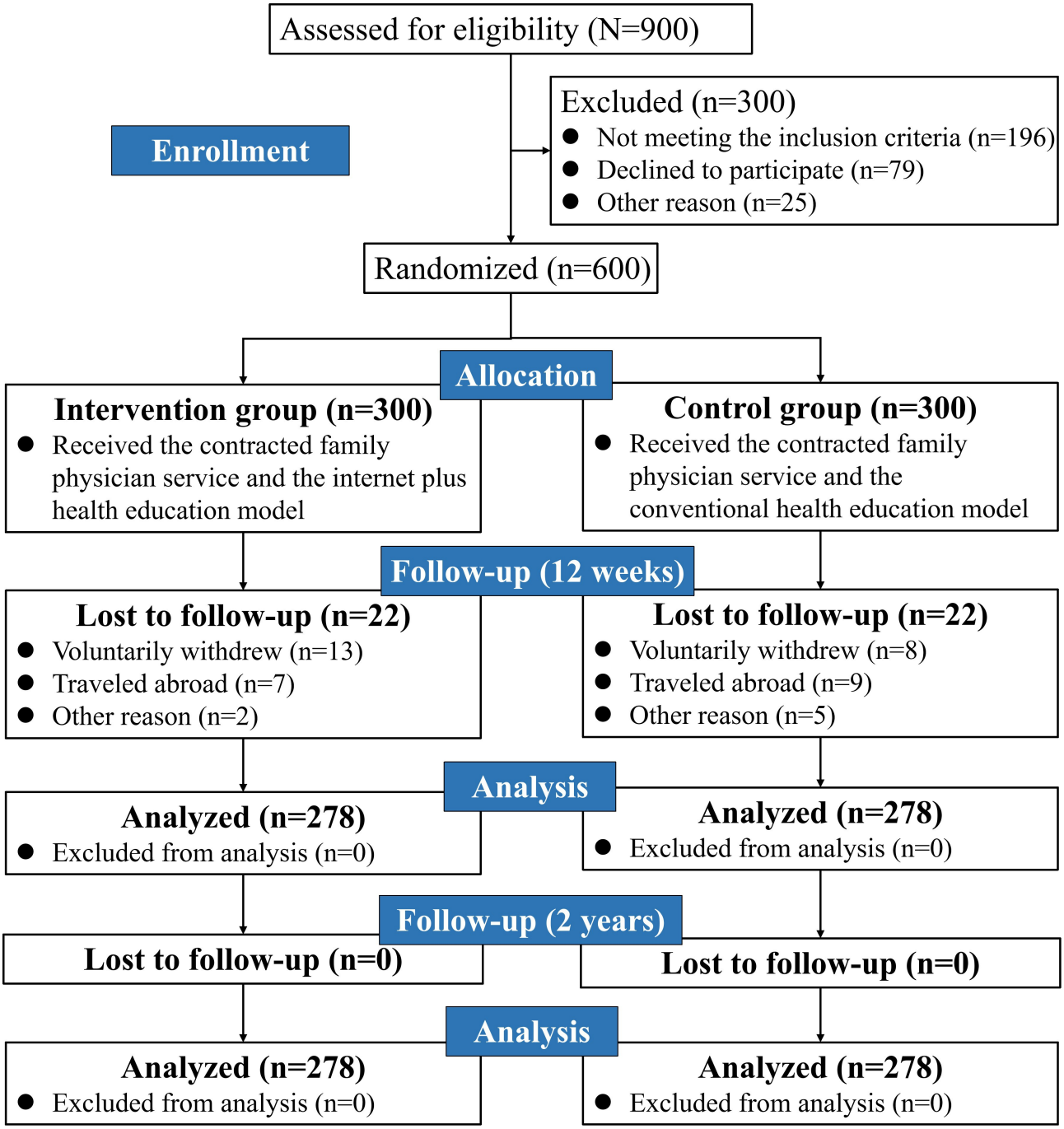
This study was approved by the Army Medical University medical ethics committee (approval 2020-007-02) and registered at the Chinese Clinical Trial Registry (ChiCTR2300071926). Written informed consent was obtained from all participants before enrollment. All data were deidentified for analysis and stored securely to ensure confidentiality. Participants received no financial compensation but were provided with free medical checkups, medication guidance, and healthy lifestyle guidance.

Results

A total of 600 participants were randomized into the intervention or control group. Follow-up at 12 weeks was completed by 46.3% (n=278) of the participants per group (overall retention: n=556, 92.7%), with no further attrition at 2 years (Figure 1). The intervention and control groups were well balanced at baseline in terms of key demographic and clinical characteristics (Table 1).

XSL•FO
RenderX

Figure 1. Participant flow diagram.



XSL•FO
RenderX

Table . Baseline characteristics of the 2 groups. The chi-square test was used to compare 2 rates or multiple component ratios.

Characteristic	Intervention (n=278)	Control (n=278)	*P* value[a]
Age (years), mean (SD)	59.38 (6.93)	58.81 (6.21)	.312
Sex (female), n (%)	159 (57.2)	158 (56.8)	.932
Disease duration (years), mean (SD)	6.29 (4.50)	6.87 (5.29)	.166
Educational level, n (%)			.472
Primary school or lower	72 (25.9)	59 (21.2)	
Junior high school	118 (42.4)	131 (47.1)	
Senior high school	70 (25.2)	69 (24.8)	
College or higher	18 (6.5)	19 (6.8)	
Annual family income, n (%)			.144
<¥30,000 (US $4303.26)	64 (23.0)	61 (21.9)	
¥30,000-¥50,000 (US $4303.26-$7172.10)	96 (34.5)	91 (32.7)	
¥50,000-¥100,000 (US $7172.10-$14,344.20)	92 (33.1)	82 (29.5)	
>¥100,000 (US $14,344.20)	26 (9.4)	44 (15.8)	
Use of hypoglycemic drugs, n (%)	243 (87.4)	237 (85.3)	.459
Diabetic complications, n (%)	92 (33.1)	99 (35.6)	.532
Smoking, n (%)			.849
No	206 (74.1)	203 (73.0)	
Yes	51 (18.3)	57 (20.5)	
Quit	21 (7.6)	18 (6.5)	
Alcohol consumption, n (%)			.639
No	212 (76.3)	207 (74.5)	
Yes	55 (19.8)	60 (21.6)	
Quit	11 (4.0)	11 (4.0)	
BMI (kg/m^2), mean (SD)	24.79 (3.05)	24.77 (3.44)	.941
Waist circumference (cm), mean (SD)	85.82 (9.02)	86.53 (8.83)	.352
Hip circumference (cm), mean (SD)	93.80 (7.02)	93.87 (7.00)	.902
WHR[b], mean (SD)	0.92 (0.07)	0.92 (0.07)	.245
FBG[c] (mmol/L), mean (SD)	8.12 (2.57)	8.39 (2.73)	.232
HbA_{1c}[d] (%), mean (SD)	7.39 (1.53)	7.44 (1.46)	.697
TC[e] (mmol/L), mean (SD)	4.77 (1.14)	4.78 (1.09)	.954
TG[f] (mmol/L), median (IQR)	1.70 (1.20-2.55)	1.91 (1.21-2.80)	.090
LDL-C[g] (mmol/L), mean (SD)	2.55 (0.87)	2.50 (0.86)	.493
HDL-C[h] (mmol/L), mean (SD)	1.31 (0.35)	1.30 (0.34)	.650
SBP[i] (mm Hg), mean (SD)	129.63 (17.80)	128.67 (16.38)	.512
DBP[j] (mm Hg), mean (SD)	77.82 (9.55)	77.09 (9.41)	.362

[a]Between-group differences for normally distributed variables were assessed using an independent-sample *t* test, whereas those for nonnormally distributed variables were assessed using a Mann-Whitney *U* test.

[b]WHR: waist-to-hip ratio.

XSL•FO
RenderX

[c]FBG: fasting blood glucose.

[d]HbA_{1c}: hemoglobin A_{1C}.

[e]TC: total cholesterol.

[f]TG: triglycerides.

[g]LDL-C: low-density lipoprotein cholesterol.

[h]HDL-C: high-density lipoprotein cholesterol.

[i]SBP: systolic blood pressure.

[j]DBP: diastolic blood pressure.

Compared with conventional education, after the 12-week intervention, the WeChat-based program resulted in statistically and clinically significant improvements in several metabolic outcomes, as detailed in Table 2. Participants in the intervention group achieved significantly greater reductions in FBG and HbA_{1c} at follow-up (P<.001 in both cases). A favorable effect was also observed on triglyceride levels, which were significantly lower in the intervention group (P=.003). The waist circumference and waist-to-hip ratio (WHR) levels in the control group were significantly higher than those in the intervention group. However, no significant between-group differences were found for total cholesterol, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, or blood pressure (P>.05 in all cases).

XSL•FO
RenderX

Table . The follow-up and change data of the 2 groups after 12 weeks of intervention.

Characteristic	Follow-up			Change		
	Intervention (n=278)	Control (n=278)	*P* value[a]	Intervention (n=278)	Control (n=278)	*P* value[a]
BMI (kg/m^2), mean (SD)	25.37 (3.08)	25.53 (3.69)	.584	0.58 (1.81)	0.76 (1.81)	.245
Waist circumference (cm), mean (SD)	85.74 (8.35)	87.73 (8.78)	.006	–0.09 (5.85)	1.21 (6.70)	.016
Hip circumference (cm), mean (SD)	94.48 (7.23)	95.42 (7.82)	.143	0.68 (6.67)	1.55 (7.20)	.142
WHR[b], mean (SD)	0.90 (0.06)	0.92 (0.06)	.019	–0.01 (0.07)	–0.00 (0.07)	.418
FBG[c] (mmol/L), mean (SD)	7.43 (1.71)	8.39 (2.52)	<.001	–0.69 (2.42)	–0.00 (2.51)	.001
HbA_{1c}[d] (%), mean (SD)	7.35 (1.49)	8.00 (1.76)	<.001	–0.03 (1.83)	0.56 (1.60)	<.001
TC[e] (mmol/L), mean (SD)	4.71 (1.02)	4.88 (1.07)	.059	–0.07 (1.05)	0.10 (0.96)	.058
TG[f] (mmol/L), median (IQR) /mean (SD)	1.40 (1.03-2.00)	1.70 (1.10-2.69)	.003	–0.58 (1.88)	–0.25 (1.59)	.025
LDL-C[g] (mmol/L), mean (SD)	2.60 (0.84)	2.59 (0.85)	.876	0.05 (0.72)	0.09 (0.75)	.529
HDL-C[h] (mmol/L), mean (SD)	1.34 (0.36)	1.33 (0.33)	.558	0.03 (0.23)	0.03 (0.29)	.861
SBP[i] (mm Hg), mean (SD)	133.62 (14.48)	134.23 (16.02)	.633	3.99 (16.46)	5.56 (17.25)	.273
DBP[j] (mm Hg), mean (SD)	79.33 (8.67)	79.39 (9.92)	.940	1.51 (10.19)	2.31 (9.91)	.353

[a]Between-group differences for normally distributed variables were assessed using an independent-sample *t* test, whereas those for nonnormally distributed variables were assessed using a Mann-Whitney *U* test.

[b]WHR: waist-to-hip ratio.

[c]FBG: fasting blood glucose.

[d]HbA_{1c}: hemoglobin A_{1C}.

[e]TC: total cholesterol.

[f]TG: triglycerides.

[g]LDL-C: low-density lipoprotein cholesterol.

[h]HDL-C: high-density lipoprotein cholesterol.

[i]SBP: systolic blood pressure.

[j]DBP: diastolic blood pressure.

The rate of normal HbA_{1c} levels was not significantly different between the 2 groups at baseline (*P*=.932), whereas it decreased significantly in the control group after the intervention (*P*=.001), as shown in Table 3. Subgroup analyses were performed according to different characteristics (sex, disease duration, educational level, and annual family income), and detailed data are shown in Table 3. Significant differences in the rate of normal HbA_{1c} levels were observed between the 2 groups after the intervention in the participants with disease duration (<10 years), educational level (junior high school or lower), and annual family income [<CN ¥50,000 (US $7172.10)].

XSL•FO
RenderX

Table . Distribution of hemoglobin A_{1C} (HbA_{1c}) levels before and after the intervention.

Characteristic	Baseline			12 week		
	HbA_{1c}<7%, n/N (%)	HbA_{1c}≥7%, n/N (%)	*P* value[a]	HbA_{1c}<7%, n/N (%)	HbA_{1c}≥7%, n/N (%)	*P* value[a]
Overall			.932			.001
Intervention	132/278 (47.5)	146/278 (52.5)		129/278 (46.4)	149/278 (53.6)	
Control	131/278 (47.1)	147/278 (52.9)		92/278 (33.1)	186/278 (66.9)	
Sex						
Female			.466			.033
Intervention	75/159 (47.2)	84/159 (52.8)		73/159 (45.9)	86/159 (54.1)	
Control	81/158 (51.3)	77/158 (48.7)		54/158 (34.2)	104/158 (65.8)	
Male			.333			.015
Intervention	57/119 (47.9)	62/119 (52.1)		56/119 (47.1)	63/119 (52.9)	
Control	50/120 (41.7)	70/120 (58.3)		38/120 (31.7)	82/120 (68.3)	
Disease duration (years)						
<10			.756			.001
Intervention	113/214 (52.8)	101/214 (47.2)		110/214 (51.4)	104/214 (48.6)	
Control	100/196 (51.0)	96/196 (49.0)		70/197 (35.5)	127/197 (64.5)	
≥10			.353			.737
Intervention	19/64 (29.7)	45/64 (70.3)		19/64 (29.7)	45/64 (70.3)	
Control	30/81 (37.0)	51/81 (63.0)		22/81 (27.2)	59/81 (72.8)	
Educational level						
Junior high school or lower			.537			<.001
Intervention	90/190 (47.4)	100/190 (52.6)		90/190 (47.4)	100/190 (52.6)	
Control	84/190 (44.2)	106/190 (55.8)		53/190 (27.9)	137/190 (72.1)	
Senior high school or higher			.451			>.999
Intervention	42/88 (47.7)	46/88 (52.3)		39/88 (44.3)	49/88 (55.7)	
Control	47/88 (53.4)	41/88 (46.6)		39/88 (44.3)	49/88 (55.7)	
Annual family income						
<CN ¥ 50,000 (US $7172.10)			.908			<.001
Intervention	80/160 (50.0)	80/160 (50.0)		76/160 (47.5)	84/160 (52.5)	
Control	75/152 (49.3)	77/152 (50.7)		43/152 (28.3)	109/152 (71.7)	
≥CN ¥ 50,000 (US $7172.10)			.953			.340
Intervention	52/118 (44.1)	66/118 (55.9)		53/118 (44.9)	65/118 (55.1)	
Control	56/126 (44.4)	70/126 (55.6)		49/126 (38.9)	77/126 (61.1)	

[a]The chi-square test was used to compare 2 rates.

Two years after the intervention, the FBG level was significantly lower in the intervention group than in the control group (mean 6.87, SD 1.48 mmol/L vs mean 7.35, SD 1.87 mmol/L; *P*=.001), but there was no significant difference in the change in FBG from baseline to the 2-year follow-up between the 2 groups. Among the participants with a disease duration of <10 years, the FBG level was significantly lower in the intervention group than in the control group (mean 6.82, SD 1.43 mmol/L vs mean 7.42, SD 2.00 mmol/L; *P*=.001). Among the participants with annual family income of <CN ¥50,000 (US $7172.10), the FBG level was significantly lower in the intervention group than in the control group (mean 6.94, SD 1.56 mmol/L vs mean 7.53, SD 2.24 mmol/L; *P*=.007; Table 4).

XSL•FO
RenderX

Table . Follow-up fasting blood glucose data 2 years after the intervention.

Characteristic	Follow-up (2 years; mmol/L), mean (SD)			Change (2 years; mmol/L), mean (SD)		
	Intervention	Control	*P* value[a]	Intervention	Control	*P* value[a]
Overall	6.87 (1.48)	7.35 (1.87)	.001	−1.25 (2.87)	−1.04 (3.14)	.403
Sex						
Female	6.91 (1.67)	7.29 (2.00)	.062	−1.11 (2.76)	−0.93 (3.11)	.569
Male	6.82 (1.17)	7.42 (1.68)	.001	−1.44 (3.00)	−1.20 (3.19)	.534
Disease duration (years)						
<10	6.82 (1.43)	7.42 (2.00)	.001	−1.13 (2.73)	−0.71 (3.01)	.150
≥10	7.03(1.61)	7.17 (1.52)	.588	−1.67 (3.28)	−1.84 (3.31)	.757
Educational level						
Junior high school or lower	6.95 (1.64)	7.37 (1.91)	.024	−1.32 (3.03)	−1.10 (3.31)	.504
Senior high school or higher	6.68 (1.03)	7.31 (1.18)	.005	−1.12 (2.48)	−0.91 (2.75)	.606
Annual family income						
<CN ¥50,000 (US $7172.10)	6.94 (1.56)	7.53 (2.24)	.007	−0.87 (2.64)	−0.71 (3.28)	.624
≥CN ¥50,000 (US $7172.10)	6.77 (1.36)	7.13 (1.26)	.034	−1.77 (3.08)	−1.45 (2.92)	.394

[a]Between-group differences were assessed using the Mann-Whitney *U* test.

Discussion

Principal Findings

This study evaluated the short-term efficacy of a WeChat-based health education tool in primary care through a 12-week randomized trial with a 2-year follow-up on glycemic control. After 12 weeks, the intervention group had improved HbA_{1c} and FBG levels compared with those of the control group, especially in patients with diabetes with a disease duration of <10 years, an educational level of junior high school or lower, and a annual family income of <CN ¥50,000 (US $7172.10) . Significant between-group differences were found in waist circumference and triglycerides, but no significant changes were observed in BMI, hip circumference, WHR, total cholesterol, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, systolic blood pressure, or diastolic blood pressure. Two years after the intervention, FBG levels were significantly lower in the intervention group. However, the change in FBG from baseline was not significantly different between the groups.

Comparison With Prior Work

Our study adds to the growing body of international evidence confirming the efficacy of digital health interventions, including smartphone apps and telemedicine, in improving diabetes outcomes [24-26]. This global consensus is strongly supported by a series of studies in China that have demonstrated the effectiveness of telemedicine and app-based interventions for glycemic control [27-29]. Specifically, with respect to the WeChat platform, our results are consistent with and reinforce findings from regional studies in the Jiangsu [19], Jilin [30], and Henan [31] provinces. The convergence of evidence from these diverse settings solidifies the view that WeChat-based health education is a viable and effective strategy for improving blood glucose levels in patients with T2DM. The intervention likely works by first improving disease knowledge through structured education [20]; then increasing self-efficacy via interactive support [32]; and, finally, enhancing adherence to medication and healthy behaviors through regular reminders and self-monitoring [33].

We found that the reduction in the HbA_{1c} level in the intervention group was lower than that reported in other studies according to a systematic review of self-management among patients with type 2 diabetes via the WeChat app [34]. A similar eHealth intervention in Shanghai demonstrated a greater HbA_{1c} reduction (7.9% to 7.3%) than our study (7.39% to 7.35%) [28]. This difference may be attributed to regional disparities in health care infrastructure [35-37] and a critical design variation: their single-platform (WeChat public account) intervention likely offered a simpler user experience, whereas our multimodule design incorporating a mini program may have increased operational complexity for participants.

We found that WeChat-based health education improved the HbA_{1c} level in patients with diabetes with low income, low educational level, and short disease duration who contracted family physicians. Currently, health apps designed to enhance self-management of chronic illnesses require continuous patient involvement as a pivotal factor influencing clinical effectiveness [38-40]. A meta-analysis revealed that the dropout rate in clinical trials for smartphone apps was significantly higher among high-income countries than among lower-middle–income countries [41]. A national web-based survey in China revealed that educational level and family income were associated with

XSL•FO
RenderX

diabetes management app use in adult patients [42]. We speculated that patients with diabetes with shorter disease durations would have better acceptance of the WeChat-based health education tool, those with low educational levels would have higher expectations for our intervention, and our free program would be more attractive to low-income patients with diabetes.

Moreover, we found that the waist circumference and WHR in the intervention group were significantly lower than those in the control group. Indeed, sex was a confounding factor, with the waist circumference and WHR of female patients being out of balance in the 2 groups before the intervention. We observed that the BMI values increased in both groups after the intervention, presumably related to the end of the intervention in the winter. People are more likely to eat more food during the Spring Festival in China [43,44]. However, WeChat-based health education was effective in helping patients with diabetes control their body shape in the winter, which was similar to the findings of other studies [45].

The intervention was constructed based on the WeChat mini program and public account, which enhances communication between family physicians and patients [46]. The family physicians' online virtual clinics were set up on the WeWalk mini program to help patients easily contact their contracted family physicians. Compared with other diabetes management mobile apps, the WeWalk mini program is free to download and use and easy to operate, reducing the likelihood of users discontinuing use [42]. Furthermore, subgroup analysis revealed that WeChat-based health education was effective for glucose control based on evidence from 3 communities. The WeChat mini program provides remote supervision and ongoing support for patients and physicians, which makes it easy to achieve continuous efficacy in the long term [47]. We suggest integrating effective, low-cost digital solutions such as this WeChat mini program into national chronic disease management strategies. Further validation is needed to assess its efficacy in rural settings with different health care infrastructures.

Limitations

Our study has several limitations. First, due to the impact of COVID-19, formal follow-up was only conducted at 12 weeks, and only the short-term effects of WeChat-based health education were observed in our study. Regarding the long-term effects of WeChat-based health education, only FBG was evaluated in this study. However, this result still needs to be further evaluated through other indicators. Second, this study was based on 3 community health service centers but was not a cluster randomized controlled trial, which may lead to insufficient accuracy of the study results. Although the authorized supervisor prevented information contamination, they may have unintentionally disclosed intervention details while managing both groups. Third, no implementation fidelity data such as user engagement and use frequency of the intervention group were collected and analyzed. Finally, a future cost-effectiveness analysis and continuous real-world optimization are crucial next steps to definitively inform policy. We could explore incorporating artificial intelligence to provide more personalized, patient-specific educational content and support, thereby enhancing the scalability and integration of this WeChat-based intervention.

Conclusions

We conducted a randomized controlled trial in community-dwelling patients with T2DM to examine the efficacy of a WeChat-based health education intervention in Chongqing, China. By strategically integrating social media platforms with primary care services, this WeChat-based health education intervention achieved significant results in improving glycemic control in patients with T2DM. Future investigations should use longitudinal designs to evaluate the sustained efficacy and feasibility of implementing WeChat-based health education.

Acknowledgments

The authors thank all the participants in the study for their involvement and help. They gratefully acknowledge family physician teams from 3 community health service centers (Xinqiao, Chenjiaqiao, and Yubeilu) in Chongqing, China. They also sincerely thank Ms Shun-Rong Huang, Ms Wen-Mei Yang, and Ms Zhi Xiao from these family physician teams for their help in sample collection.

Funding

This research was financially supported by the Technical Innovation and Application Development Special Project in Chongqing, China (grant cstc2019jscx-msxmX0198).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

CZ contributed to conceptualization, methodology, formal analysis, and writing—original draft. XD and XX contributed to software development and data curation. EYL, MJ, YW, JY, XW, LZ, HR, LL, FL, and TL were responsible for project implementation, participant recruitment, and data collection. XM contributed to the methodology and validation. JAC supervised the study, acquired funding, and reviewed the manuscript. All the authors approved the final version.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Feasibility study.
[PDF File, 61 KB - mhealth_v14i1e80738_app1.pdf]

Multimedia Appendix 2
The introduction of the WeChat mini program and public account.
[PDF File, 1903 KB - mhealth_v14i1e80738_app2.pdf]

Multimedia Appendix 3
Descriptions and screenshots of diabetes self-management goals organized into 16 items.
[PDF File, 220 KB - mhealth_v14i1e80738_app3.pdf]

Checklist 1
CONSORT-eHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (V 1.6.1).
[PDF File, 550 KB - mhealth_v14i1e80738_app4.pdf]

References

1. Sun H, Saeedi P, Karuranga S, et al. IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022 Jan;183:109119. [doi: 10.1016/j.diabres.2021.109119] [Medline: 34879977]
2. Świątoniowska N, Sarzyńska K, Szymańska-Chabowska A, Jankowska-Polańska B. The role of education in type 2 diabetes treatment. Diabetes Res Clin Pract 2019 May;151:237-246. [doi: 10.1016/j.diabres.2019.04.004] [Medline: 31063855]
3. The Lancet Diabetes Endocrinology. Diabetes education: the key to a brighter tomorrow. Lancet Diabetes Endocrinol 2022 Dec;10(12):827. [doi: 10.1016/S2213-8587(22)00323-0] [Medline: 36384161]
4. Falk EM, Staab EM, Deckard AN, et al. Effectiveness of multilevel and multidomain interventions to improve glycemic control in U.S. racial and ethnic minority populations: a systematic review and meta-analysis. Diabetes Care 2024 Sep 1;47(9):1704-1712. [doi: 10.2337/dc24-0375] [Medline: 39190927]
5. Wang L, Peng W, Zhao Z, et al. Prevalence and treatment of diabetes in China, 2013-2018. JAMA 2021 Dec 28;326(24):2498-2506. [doi: 10.1001/jama.2021.22208] [Medline: 34962526]
6. National diabetes statistics report 2020: estimates of diabetes and its burden in the United States. Centers for Disease Control and Prevention. URL: https://diabetesresearch.org/wp-content/uploads/2022/05/national-diabetes-statistics-report-2020.pdf [accessed 2021-10-24]
7. IDF Diabetes Atlas, 9th edition: International Diabetes Federation; 2019.
8. Rossen J, Yngve A, Hagströmer M, et al. Physical activity promotion in the primary care setting in pre- and type 2 diabetes - the Sophia step study, an RCT. BMC Public Health 2015 Jul 12;15:647. [doi: 10.1186/s12889-015-1941-9] [Medline: 26164092]
9. De la Fuente Coria MC, Cruz-Cobo C, Santi-Cano MJ. Effectiveness of a primary care nurse delivered educational intervention for patients with type 2 diabetes mellitus in promoting metabolic control and compliance with long-term therapeutic targets: randomised controlled trial. Int J Nurs Stud 2020 Jan;101:103417. [doi: 10.1016/j.ijnurstu.2019.103417] [Medline: 31683226]
10. Lean ME, Leslie WS, Barnes AC, et al. Primary care-led weight management for remission of type 2 diabetes (DiRECT): an open-label, cluster-randomised trial. Lancet 2018 Feb 10;391(10120):541-551. [doi: 10.1016/S0140-6736(17)33102-1] [Medline: 29221645]
11. Daly BM, Arroll B, Scragg RK. Trends in diabetes care and education by primary health care nurses in Auckland, New Zealand. Diabetes Res Clin Pract 2021 Jul;177:108903. [doi: 10.1016/j.diabres.2021.108903] [Medline: 34102248]
12. Guidance on global monitoring for diabetes prevention and control: framework, indicators and application. World Health Organization. 2024. URL: https://www.who.int/publications/i/item/9789240102248 [accessed 2026-01-17]
13. Sun J, Fan Z, Kou M, Wang X, Yue Z, Zhang M. Impact of nurse-led self-management education on type 2 diabetes: a meta-analysis. Front Public Health 2025 Aug 11;13:1622988. [doi: 10.3389/fpubh.2025.1622988] [Medline: 40860544]
14. Fitzpatrick R, Pant S, Li J, et al. Implementation of non-insulin-dependent diabetes self-management education (DSME) in LMICs: a systematic review of cost, adoption, acceptability, and fidelity in resource-constrained settings. Front Health Serv 2023;3:1155911. [doi: 10.3389/frhs.2023.1155911] [Medline: 37383485]

XSL•FO
RenderX

15. Huang Z, Lum E, Jimenez G, Semwal M, Sloot P, Car J. Medication management support in diabetes: a systematic assessment of diabetes self-management apps. BMC Med 2019 Jul 17;17(1):127. [doi: 10.1186/s12916-019-1362-1] [Medline: 31311573]
16. Veazie S, Winchell K, Gilbert J, et al. Rapid evidence review of mobile applications for self-management of diabetes. J Gen Intern Med 2018 Jul;33(7):1167-1176. [doi: 10.1007/s11606-018-4410-1] [Medline: 29740786]
17. Gao C, Zhou L, Liu Z, Wang H, Bowers B. Mobile application for diabetes self-management in China: do they fit for older adults? Int J Med Inform 2017 May;101:68-74. [doi: 10.1016/j.ijmedinf.2017.02.005] [Medline: 28347449]
18. Fleming GA, Petrie JR, Bergenstal RM, Holl RW, Peters AL, Heinemann L. Diabetes digital app technology: benefits, challenges, and recommendations. A consensus report by the European Association for the Study of Diabetes (EASD) and the American Diabetes Association (ADA) diabetes technology working group. Diabetes Care 2020 Jan;43(1):250-260. [doi: 10.2337/dci19-0062] [Medline: 31806649]
19. Xia SF, Maitiniyazi G, Chen Y, et al. Web-based TangPlan and WeChat combination to support self-management for patients with type 2 diabetes: randomized controlled trial. JMIR Mhealth Uhealth 2022 Mar 30;10(3):e30571. [doi: 10.2196/30571] [Medline: 35353055]
20. Wu Y, Min H, Li M, et al. Effect of Artificial Intelligence-based Health Education Accurately Linking System (AI-HEALS) for type 2 diabetes self-management: protocol for a mixed-methods study. BMC Public Health 2023 Jul 11;23(1):1325. [doi: 10.1186/s12889-023-16066-z] [Medline: 37434126]
21. Lindström J, Ilanne-Parikka P, Peltonen M, et al. Sustained reduction in the incidence of type 2 diabetes by lifestyle intervention: follow-up of the Finnish Diabetes Prevention Study. Lancet 2006 Nov 11;368(9548):1673-1679. [doi: 10.1016/S0140-6736(06)69701-8] [Medline: 17098085]
22. Sneha S, Thalla S, Rischie I, Shahriar H. Health internet technology for chronic conditions: review of diabetes management apps. JMIR Diabetes 2021 Aug 31;6(3):e17431. [doi: 10.2196/17431] [Medline: 34463627]
23. Chinese Diabetes Society, National Office for Primary Diabetes Care. National guidelines for the prevention and control of diabetes in primary care (2018). Zhonghua Nei Ke Za Zhi 2018 Dec 1;57(12):885-893. [doi: 10.3760/cma.j.issn.0578-1426.2018.12.003] [Medline: 30486556]
24. Lim SL, Ong KW, Johal J, et al. Effect of a smartphone app on weight change and metabolic outcomes in Asian adults with type 2 diabetes: a randomized clinical trial. JAMA Netw Open 2021 Jun 1;4(6):e2112417. [doi: 10.1001/jamanetworkopen.2021.12417] [Medline: 34081137]
25. Kruse CS, Mileski M, Heinemann K, Huynh H, Leafblad A, Moreno E. Analyzing the effectiveness of mHealth to manage diabetes mellitus among adults over 50: a systematic literature review. J Multidiscip Healthc 2023 Jan 12;16:101-117. [doi: 10.2147/JMDH.S392693] [Medline: 36660039]
26. Mueller S, Dinges SM, Gass F, et al. Telemedicine-supported lifestyle intervention for glycemic control in patients with CHD and T2DM: multicenter, randomized controlled trial. Nat Med 2025 Apr;31(4):1203-1213. [doi: 10.1038/s41591-025-03498-w] [Medline: 39920395]
27. Jia W, Zhang P, Zhu D, et al. Evaluation of an mHealth-enabled hierarchical diabetes management intervention in primary care in China (ROADMAP): a cluster randomized trial. PLoS Med 2021 Sep 21;18(9):e1003754. [doi: 10.1371/journal.pmed.1003754] [Medline: 34547030]
28. Feng Y, Zhao Y, Mao L, et al. The effectiveness of an eHealth family-based intervention program in patients with uncontrolled type 2 diabetes mellitus (T2DM) in the community via WeChat: randomized controlled trial. JMIR Mhealth Uhealth 2023 Mar 20;11:e40420. [doi: 10.2196/40420] [Medline: 36939825]
29. Dong Y, Wang P, Dai Z, et al. Increased self-care activities and glycemic control rate in relation to health education via WeChat among diabetes patients: a randomized clinical trial. Medicine (Baltimore) 2018 Dec;97(50):e13632. [doi: 10.1097/MD.0000000000013632] [Medline: 30558051]
30. Sun C, Sun L, Xi S, et al. Mobile phone-based telemedicine practice in older Chinese patients with type 2 diabetes mellitus: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 4;7(1):e10664. [doi: 10.2196/10664] [Medline: 30609983]
31. Zhang Y, Chu L. Effectiveness of systematic health education model for type 2 diabetes patients. Int J Endocrinol 2018 Jul 25;2018:6530607. [doi: 10.1155/2018/6530607] [Medline: 30147724]
32. Zhu L, Wang J, Pan Z, et al. Effectiveness of a family-based self-management intervention for type 2 diabetes patients receiving family doctor contract services: a community-based randomized controlled trial. J Prim Care Community Health 2025;16:21501319251330384. [doi: 10.1177/21501319251330384] [Medline: 40162887]
33. Zhang B, Kalampakorn S, Powwattana A, Sillabutra J, Liu G. A transtheoretical model-based online intervention to improve medication adherence for Chinese adults newly diagnosed with type 2 diabetes: a mixed-method study. J Prim Care Community Health 2024;15:21501319241263657. [doi: 10.1177/21501319241263657] [Medline: 39077970]
34. Yang J, Yang H, Wang Z, et al. Self-management among type 2 diabetes patients via the WeChat application: a systematic review and meta-analysis. J Clin Pharm Ther 2021 Feb;46(1):4-16. [doi: 10.1111/jcpt.13264] [Medline: 33085789]
35. Dai H, Tang L, Wang Z, et al. Facilitate signing with the family doctor: a study of the practice in Shanghai, China. Int J Gen Med 2021 Oct 18;14:6907-6917. [doi: 10.2147/IJGM.S332890] [Medline: 34703295]
36. Issom DZ, Henriksen A, Woldaregay AZ, Rochat J, Lovis C, Hartvigsen G. Factors influencing motivation and engagement in mobile health among patients with sickle cell disease in low-prevalence, high-income countries: qualitative exploration of patient requirements. JMIR Hum Factors 2020 Mar 24;7(1):e14599. [doi: 10.2196/14599] [Medline: 32207692]

XSL•FO
RenderX

37. Scott AR, Alore EA, Naik AD, Berger DH, Suliburk JW. Mixed-methods analysis of factors impacting use of a postoperative mHealth app. JMIR Mhealth Uhealth 2017 Feb 8;5(2):e11. [doi: 10.2196/mhealth.6728] [Medline: 28179215]
38. Agarwal P, Mukerji G, Desveaux L, et al. Mobile app for improved self-management of type 2 diabetes: multicenter pragmatic randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 10;7(1):e10321. [doi: 10.2196/10321] [Medline: 30632972]
39. Sze WT, Kow SG. Perspectives and needs of Malaysian patients with diabetes for a mobile health app support on self-management of diabetes: qualitative study. JMIR Diabetes 2023 Oct 23;8:e40968. [doi: 10.2196/40968] [Medline: 37870903]
40. Krall JS, Childs B, Mehrotra N. Mobile applications to support diabetes self-management education: patient experiences and provider perspectives. J Diabetes Sci Technol 2023 Sep;17(5):1206-1211. [doi: 10.1177/19322968231174037] [Medline: 37162000]
41. Iqhrammullah M, Yudhistira Refin R, Fitria Andika F, et al. Dropout rate in clinical trials of smartphone apps for diabetes management: a meta-analysis. Diabetes Res Clin Pract 2024 Jun;212:111723. [doi: 10.1016/j.diabres.2024.111723] [Medline: 38830484]
42. Zhang Y, Li X, Luo S, et al. Use, perspectives, and attitudes regarding diabetes management mobile apps among diabetes patients and diabetologists in China: national web-based survey. JMIR Mhealth Uhealth 2019 Feb 8;7(2):e12658. [doi: 10.2196/12658] [Medline: 30735147]
43. Xu H, Cao L, Li J, et al. Is Chinese spring festival a key point for glycemic control of patients with type 2 diabetes mellitus in China? Front Public Health 2022 Dec 22;10:975544. [doi: 10.3389/fpubh.2022.975544] [Medline: 36620247]
44. Guo K, Ye J, Zhang L, et al. Effect of the Chinese New Year public holiday on the glycemic control of T1DM with intensive insulin therapy. Front Endocrinol (Lausanne) 2022 Jun 28;13:915482. [doi: 10.3389/fendo.2022.915482] [Medline: 35837316]
45. Han Y, Ye X, Li X, et al. Comparison of an online versus conventional multidisciplinary collaborative weight loss programme in type 2 diabetes mellitus: a randomized controlled trial. Int J Nurs Pract 2023 Feb;29(1):e13126. [doi: 10.1111/ijn.13126] [Medline: 36567135]
46. Liu K, Xie Z, Or CK. Effectiveness of mobile app-assisted self-care interventions for improving patient outcomes in type 2 diabetes and/or hypertension: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2020 Aug 4;8(8):e15779. [doi: 10.2196/15779] [Medline: 32459654]
47. Duan Y, Li X, Guo L, Liang W, Shang B, Lippke S. A WeChat mini program-based intervention for physical activity, fruit and vegetable consumption among Chinese cardiovascular patients in home-based rehabilitation: a study protocol. Front Public Health 2022 Mar 22;10:739100. [doi: 10.3389/fpubh.2022.739100] [Medline: 35392478]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
DSME: diabetes self-management education
FBG: fasting blood glucose
HbA_{1c}: hemoglobin A_{1C}
T2DM: type 2 diabetes mellitus
WHR: waist-to-hip ratio

Edited by L Buis; submitted 16.Jul.2025; peer-reviewed by Q Liu, SS Palve, Y Wan; revised version received 26.Dec.2025; accepted 05.Jan.2026; published 20.Feb.2026.

Please cite as:
Zheng C, Dou X, Xiong X, Lei EY, Jiang M, Wu Y, Yu J, Wang X, Zhang L, Rong H, Lu L, Li F, Luo T, Ma X, Chen JA
WeChat-Based Intervention for Glycemic Control in Patients With Type 2 Diabetes Mellitus: Multicenter Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e80738
URL: https://mhealth.jmir.org/2026/1/e80738
doi:10.2196/80738

© Chuanfen Zheng, Xiong Dou, Xiaotao Xiong, En-Yu Lei, Ming Jiang, Yulin Wu, Jing Yu, Xianjun Wang, Ling Zhang, Honghui Rong, Lu Lu, Fengju Li, Ting Luo, Xiangyu Ma, Ji-An Chen. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 20.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons

XSL•FO
RenderX

Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

WeChat-Delivered Mobile Medical Nutrition Therapy Intervention in Gestational Diabetes Mellitus: Randomized Controlled Trial

XiaoRong Wang[1], PhD; MiaoMiao Bian[2], MSN; Fang Li[3], BMed; Chang Zhou[4], BSN; LiNing Mu[5], MSN

[1]Community Healthcare Center, Jinan Central Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, China

[2]Cardiovascular Department, Jinan Central Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, China

[3]Graduate School, Shandong First Medical University, Jinan, Shandong, China

[4]School of Nursing and Rehablitation, Shandong University, Jinan, Shandong, China

[5]Diabetes Nursing Specialist Clinic, Jinan Central Hospital Affiliated to Shandong First Medical University, 105 Jiefang Road, Lixia District, Jinan, Shandong, China

Corresponding Author:
LiNing Mu, MSN
Diabetes Nursing Specialist Clinic, Jinan Central Hospital Affiliated to Shandong First Medical University, 105 Jiefang Road, Lixia District, Jinan, Shandong, China

Abstract

Background: Medical nutrition therapy (MNT) serves as the foundational intervention in the clinical management of gestational diabetes mellitus (GDM) management. However, inadequate supportive care often hinders patients' ability to sustain dietary modifications and self-management behaviors, particularly for complex regimens. Flexible online interventions are thus gaining interest as adjuncts to clinical care, with the potential to improve the outcomes of GDM self-management.

Objective: This study sought to evaluate the benefits of WeChat-delivered medical nutrition therapy (WeMNT), a WeChat-delivered MNT intervention, for patients with GDM.

Methods: A parallel-group, single-blind randomized controlled trial (with outcome assessor blinding) was implemented at a university hospital's obstetric clinic. Eligible participants were those with a 24‐ to 28-week singleton pregnancy, a GDM diagnosis confirmed by the 75 g oral glucose tolerance test, no requirement for insulin therapy, and the ability to use a smartphone and WeChat, as well as to communicate in Chinese. Participants were randomized 1:1 to the intervention (n=47, 50%) or control group (n=47, 50%) using a random number table. The WeMNT intervention was designed per the behavior change wheel framework: (1) analyze capability-, opportunity-, and motivation-based barriers related to MNT adherence; (2) formulate targeted intervention functions; and (3) select evidence-based behavior change techniques. Three WeChat mini-programs served as the core delivery platform to implement the multifaceted intervention functions. The primary outcomes measured in this study included fasting blood glucose, 2-hour postprandial glucose (2hPG), hemoglobin A_{1c} (HbA_{1c}), and gestational weight gain (GWG). The secondary outcomes assessed included obstetric complications and neonatal parameters. Intention-to-treat analysis using adjusted generalized linear mixed models was conducted to evaluate the effects of the intervention.

Results: From March 2023 to October 2023, a total of 94 participants were enrolled in this study. Among the participants, 88 (93.62%) individuals (mean age 32.18, SD 5.04 years) successfully completed the study, 44 (46.81%) in the intervention group and 44 (46.81%) in the control group. Approximately 68.18% (30/44) of the participants in the intervention group demonstrated sustained adherence, as expected, adhered to the diet plan ≥60% of the days. Compared with the control group, the intervention group showed a significantly greater decrease in GWG over time (group×time interaction: β=−1.96, 95% CI −3.58 to −0.34; *P*=.02), with no significant effects on fasting blood glucose. Conversely, significant decreases in 2hPG (β=−0.12, 95% CI −0.19 to −0.04; *P*<.001) and HbA_{1c} (β=−0.49, 95% CI −0.74 to −0.23; *P*<.001) were noted in the intervention group. No unexpected maternal adverse events occurred, and no significant effects on other health outcomes were detected.

Conclusions: As a WeChat-delivered, behavior change wheel–informed intervention, WeMNT successfully reduced 2hPG, HbA_{1c}, and GWG in GDM management. Its integrated self-management model yields robust evidence, validating this patient-centered tool as a viable option for clinical adoption.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2300078227; https://tinyurl.com/4st5cvb2

(JMIR Mhealth Uhealth 2026;14:e67410) doi:10.2196/67410

KEYWORDS

gestational diabetes mellitus; medical nutrition therapy; behavior change techniques; gestational weight gain; mobile health; mHealth

XSL•FO
RenderX

Introduction

Background

Gestational diabetes mellitus (GDM) refers to a prevalent pregnancy-associated metabolic disorder, defined by the onset or initial identification of glucose intolerance during gestation [1]. According to the International Diabetes Federation, the age-standardized prevalence of GDM in low- and middle-income countries increased from 6.8% in 2011 to 8.8% in 2021, representing a weighted relative increase of 29.4% (95% CI 22.1% - 37.2%) after adjusting for diagnostic criteria and study heterogeneity, in parallel with the growing epidemic of obesity [2,3]. Suboptimal glycemic regulation during GDM is associated with increased maternal risks, including preeclampsia, cesarean delivery, and postpartum hemorrhage. Beyond the perinatal period, women with a history of GDM face an increased likelihood of developing chronic conditions such as type 2 diabetes, metabolic syndrome, and cardiovascular disease in the long run [4]. Offspring exposed to maternal hyperglycemia face immediate risks of macrosomia, birth trauma, neonatal hypoglycemia, hyperbilirubinemia, and polycythemia. These infants also demonstrate increased susceptibility to obesity and diabetes later in life.

Medical nutrition therapy (MNT) is the primary approach for managing GDM and is frequently effective in achieving desired blood glucose levels [5]. The significant benefits of dietary interventions in GDM management also include improved maternal outcomes and reduced neonatal macrosomia risk compared with conventional care [6]. Specifically, low–glycemic index (GI) diets have been shown to attenuate glycemic excursions while mitigating the need for pharmacological interventions [6,7]. These outcomes are hypothesized to result from delayed carbohydrate absorption and enhanced satiety signaling associated with low-GI macronutrient profiles [7]. However, controversy persists regarding optimal carbohydrate modulation strategies. While certain trials advocate restrictive carbohydrate regimens (≤40% of total energy intake) for acute glycemic reduction, emerging data suggest that liberal intake of high-fiber complex carbohydrates (eg, whole grains and legumes) may offer superior long-term metabolic benefits [8]. This paradigm prioritizes glycemic stability and preserves maternal nutritional adequacy, potentially reducing the risk of ketogenesis compared with severe caloric restriction [9]. Regardless of whether a restrictive or more liberal carbohydrate regimen is adopted, individualized nutritional plans that consider mothers' specific characteristics (eg, health status, weight, ethnicity, and cultural preferences) are crucial for effective GDM management [10]. A joint care team comprising endocrinologists, obstetricians, certified dietitians, and diabetes education specialists is recommended to develop individualized treatment plans [5].

Enhanced dietary support significantly improved carbohydrate adherence in high-risk pregnancies [11]. Among obese women (BMI ≥30 kg/m^2), those receiving weekly 90-minute dietitian-led group counseling on carbohydrate-controlled diets (40% - 45% of total calories) demonstrated a 47% reduction in days with excessive carbohydrate intake (≥10% above target) compared with controls over 24 weeks (adjusted odds ratio 0.53, 95% CI 0.39 - 0.72; P<.001) [11]. A dietary plan is key to managing GDM, yet women often struggle to adopt and sustain new dietary habits and self-management behaviors [12]. Long-term adherence to complex dietary plans is challenging without continuous support [13]. Providing adequate self-management support empowers patients to control their diet and blood glucose levels. As they gain self-management skills, positive behavioral changes become evident [13]. Prescribed treatment plans instill a strong sense of responsibility in women with GDM, motivating them to strictly adhere to dietary and self-management guidelines [12]. Nevertheless, their clinical appointments with health care professionals are frequently too brief to provide the in-depth, comprehensive guidance necessary for effective GDM management [14,15]. Mobile health (mHealth) and digital interventions have become essential complementary tools for supporting women with GDM and health care professionals. Recent studies report an increasing reliance on digital platforms because of their flexibility and convenience, which help overcome limited clinical contact time and provide continuous, personalized support for self-management [16,17]. Both women and health care professionals expressed openness to using technology as an adjunct to standard care but emphasized the need for improvements in usability, such as enhanced content layout, user interface design, and data visualization [17]. Participants also highlighted the importance of features supporting diverse data recording, personalization, comprehensive information delivery, and effective communication with health care providers and peer networks [17]. Currently, various mHealth tools often do not fully address the needs of women with GDM [18]. Both patients and health care professionals have stressed the importance of improving usability, personalization, information adequacy, and communication functionalities [18]. These findings highlight the ongoing need to develop and optimize mHealth solutions to better support self-management among women with GDM.

Theoretical Framework

According to the literature, current behavior change interventions often have 2 major limitations: insufficient theoretical support and underuse of evidence-based frameworks [19]. To address these limitations, we adopted the behavior change wheel (BCW) as our core theoretical foundation [19]. With the embedded capability, opportunity, motivation behavior (COM-B) model, the BCW is distinguished in supporting researchers to analyze the origins of behavioral barriers and design theory-informed interventions [19]. Specifically, capability refers to an individual's psychological (eg, knowledge and cognitive skills) or physical (eg, motor ability) capacity to perform the target behavior; opportunity encompasses external enablers or constraints, including the physical environment (eg, access to resources) and social support (eg, family encouragement); and motivation integrates conscious decision-making (eg, goal setting) and subconscious drives (eg, emotional attitudes) that influence behavioral intent [19]. In the system, COM-B elements interact dynamically to shape behavior, enabling designers to first identify context-specific barriers to the target behavior and then map them to targeted

XSL•FO
RenderX

intervention functions most likely to drive meaningful change. This systematic mapping, in turn, facilitates the precise translation of intervention functions into fine-grained, measurable, and replicable behavior change techniques (BCTs)—a critical step in ensuring the scientific rigor and scalability of the intervention [19]. By grounding the intervention in the BCW, we establish a transparent, theory-driven link between identified behavioral barriers, intervention design choices, and measurable outcomes—addressing the field's need for more structured, evidence-based approaches to behavior change intervention development.

Goal of This Study

Building on the BCW framework, we developed WeChat-delivered medical nutrition therapy (WeMNT), a WeChat-delivered MNT program tailored for women with GDM. On the basis of clinical expertise and patient feedback, we hypothesized that WeMNT will improve the effectiveness of GDM management. This study aimed to evaluate the benefits of WeMNT, a WeChat-delivered MNT intervention program developed from clinical knowledge and the BCW framework, for women with GDM.

Methods

Study Design and Participants

This single-blind, parallel-group randomized controlled trial (RCT)—with blinding limited specifically to outcome assessors—was carried out at the obstetric clinic affiliated with a university hospital. The reporting of this study adhered to the CONSORT (Consolidated Standards of Reporting Trials) guidelines (Checklist 1). The inclusion criteria included singleton pregnancy at 24 to 28 weeks, GDM diagnosed by a 75 g oral glucose tolerance test (OGTT) [20], no insulin requirement, and the ability to use a smartphone with WeChat and to communicate in Chinese.

Sample Size Calculation

The study was powered based on the primary outcome of hemoglobin A_{1c} (HbA_{1c}). To determine the appropriate sample size, we first identified the effect size based on the relevant literature. A meta-analysis revealed that telemedicine interventions for women with GDM yielded an effect size of 0.63 (Cohen *d*, ie, a standardized mean difference between groups) for HbA_{1c} reduction; specifically, studies using WeChat as the intervention tool (6 RCTs in total) demonstrated a greater effect size of 0.84 for HbA_{1c} improvement [21]. Considering both clinical relevance and methodological conservatism, we determined a minimum effect size (Cohen *d*) of 0.65 for the sample size calculation.

G*Power 3.1 software [22] was used to perform the a priori sample size calculation using the following key parameters: effect size (*d*)=0.65, significance level (α)=.05 (2-tailed), and statistical power (1−β)=.80. The calculation results indicated a basic sample size of 78 participants. To address a potential 15% attrition rate, a common consideration in longitudinal clinical studies, the final sample size was adjusted to 94 participants, which ensured adequate statistical power for detecting the prespecified effect on the primary outcome (HbA_{1c}).

Recruitment and Randomization

Participants were enrolled during their standard antenatal care appointments. Clinic physicians and nursing staff assisted in recruitment by disseminating study information to potential participants who met the eligibility criteria and had recently completed the OGTT. Interested candidates underwent an initial eligibility assessment by the research team. A follow-up appointment was then scheduled at the clinic, typically within 1 week of the OGTT, to discuss the study in detail, obtain written informed consent, and initiate enrollment.

After providing informed consent, eligible participants were randomly assigned to either the intervention group (n=47, 50%) or the control group (n=47, 50%) in a 1:1 allocation ratio, with randomization sequences generated by a computer. To ensure allocation concealment, sequentially numbered, sealed, opaque envelopes were used. This process preserved the integrity of the randomization and minimized selection bias.

Intervention

To achieve the study's objective, we followed the 3-stage protocol specified in the BCW intervention design guide [19]. First, preliminary assessments and literature review were conducted, and the COM-B barriers impeding GDM patients' adherence to a personalized MNT plan were analyzed. The identified distinct COM-related barriers to MNT adherence include the following:

- Capability: deficits in nutritional knowledge and inability to formulate personalized MNT plans, accurately weigh food portions, or maintain dietary records
- Opportunity: lack of household weighing tools, work-related time constraints for learning and implementing MNT, and insufficient social support
- Motivation: inadequate awareness of MNT's clinical importance, concerns about disrupting daily routines, and anxieties regarding the potential impacts of MNT

Second, targeted intervention functions—including education, enablement, modeling, and environmental restructuring—were selected to address these COM-specific barriers. Finally, building on this foundation, evidence-based BCTs were further identified to operationalize the selected functions [23]. Specifically, a group of BCTs validated for type 2 diabetes self-management in the literature was adopted [24].

For delivery, WeChat mini-programs were prioritized as the core platform because of their high accessibility and user familiarity. Three WeChat mini-programs were designated as the core delivery platform to implement diverse intervention functions and facilitate participants' acquisition of relevant knowledge and practical skills. A schematic overview outlines the BCW-informed intervention development, COM categories, barriers, BCTs used, and WeMNT intervention operationalization, including BCT embedded in mini-program features (Table 1).

XSL•FO
RenderX

Table . Intervention development: capability, opportunity, and motivation (COM) categories, barriers, behavior change techniques (BCTs), and mini-programs.

COM categories and key barriers	Intervention functions	Policy categories	BCTs	WeChat platform	Intervention contents
Capability					
Deficits in nutritional knowledge	Education	Communication	• BCT4.1 Instruction on how to perform a behavior • BCT 4.2 Information about antecedents	• Mini-program "Diet Diary-Food Charts"	• Via the program, women learn GI[a] knowledge and practical nutritional skills (BCT4.1) and get nutritional choice antecedent info (BCT4.2).
Inability to accurately understand food portions and formulate personalized MNT[b] plans	Training	Service provision	• BCT4.1 Instruction on how to perform a behavior • BCT6.1 Demonstration of the behavior • BCT1.4 Action planning	• Mini-program "Diet Diary"	• Via the program, dietitians teach women to input clinical data, select food types and weights/portions, and make personal MNT plans (BCT4.1), demonstrate plan design and adjustment (BCT6.1), and cocreate daily meal schedules with reminders (BCT1.4).
Inability to maintain dietary records	Training	Service provision	• BCT2.3 Self-monitoring of behavior	• Mini-program "Diet Diary"	• Women use the program to log daily food portion weights and receive alerts on whether intake meets, falls below, or exceeds targets.
Opportunity					
Lack of household weighing tools	Environmental restructuring	Service provision	• BCT12.5 Adding objects to the environment		• The research team provided digital food scales to support accurate measurement.
Work-related time constraints for implementing MNT	Environmental restructuring	Service provision	• BCT7.1 Prompts/cues	• Mini-program "Diet Diary"	• Aligned with the MNT plan, the mini-program sends timed pop-up reminders (eg, "Check meal compliance 10 minutes premeal," "Log daily diet 5 minutes before bedtime").

XSL•FO
RenderX

COM categories and key barriers	Intervention functions	Policy categories	BCTs	WeChat platform	Intervention contents
Insufficient social support	Modeling	Service provision	• BCT12.2 Restructuring the social environment	• Group discussion	• A nurse moderated a WeChat group for peer support; women shared experiences/strategies weekly.
Motivation					
Inadequate awareness of MNT's clinical importance	Education	Service provision	• BCT4.3 Information about health consequences • BCT2.2 Feedback on behavior	• Mini-program "Diet Diary"; Video call	• Via the program, women self-learn MNT's medical value, disease-prognosis links, and noncompliance risks to fill cognitive gaps (BCT4.3). Persistent glucose issues triggered video consultations and potential multidisciplinary referrals (BCT2.2).
Worries about MNT disrupting daily routines	Modeling	Service provision	• BCT3.3 Social support	• Group discussion	• A nurse moderated a WeChat group for peer support; women shared experiences or strategies weekly.
Fears of MNT causing potential health impacts	Enablement	Service provision	• BCT2.4 Self-monitoring of the outcomes • BCT2.2 Feedback on behavior	• Mini-program "Pregnancy Sugar Weight Record" and "Fetal Weight Assessor" • Instant message or voice call	• Women input blood glucose, body weight, and fetal weight into the programs, which generate linked trend curves for all 3 and alert if values meet, fall below, or exceed targets. • Diabetes nurses weekly reviewed backend data, provided WeChat feedback, and adjusted meal plans with dietitians for off-target patients.

[a]GI: glycemic index.

[b]MNT: medical nutrition therapy.

Intervention Features

In this mHealth study, we used well-established, publicly available, nonprofit WeChat mini-programs as the intervention platform. First, the "Diet Diary" mini-program was developed by a certified dietitian, Zhongyi Gu, as a nonprofit public health tool aligned with the American Diabetes Association guidelines for GDM [5]. This mini-program integrates personalized meal planning, real-time dietary tracking, and weekly professional feedback to facilitate effective dietary management for users. Dietitians leveraged the mini-program to deliver step-by-step instructions on MNT plan formulation, demonstrating both the

XSL•FO
RenderX

plan design logic and mini-program operation through real-time interface walkthroughs to teach women how to use the platform independently. They then collaborated with women using the mini-program's action planning module to develop structured daily meal plans that specified food types and weights for each meal, with embedded reminders and real-time adjustment features tailored to individual clinical needs and daily routines. On the basis of the mini-program's "Food Charts" self-learning feature, women acquired knowledge about the GIs of 476 common foods, food selection, weighing, recording, and meal arrangement; access to antecedent information about nutritional choices; and learned about the health consequences of dietary behaviors linked to GI.

The second mini-program is the "Pregnancy Sugar Weight Record" mini-program. This mini-program offers 3 evidence-based modules: weight monitoring with personalized reference ranges based on prepregnancy BMI (pBMI), blood glucose monitoring with goals set with diabetes nurses, and an automated alert system for weight gain deviations. Weight measurements are standardized using calibrated digital scales and automatically synchronized with user profiles, with graphical trend visualization for weight gain deviations exceeding ±10% of target ranges.

The third mini-program is the "Fetal Weight Assessor" mini-program. This mini-program is designed for fetal growth monitoring, allowing participants to input key ultrasound parameters after routine prenatal examinations to estimate fetal weight. It also generates percentile curves to visualize growth trajectories relative to gestational age, aiding in comprehensive fetal growth assessment (Multimedia Appendix 1).

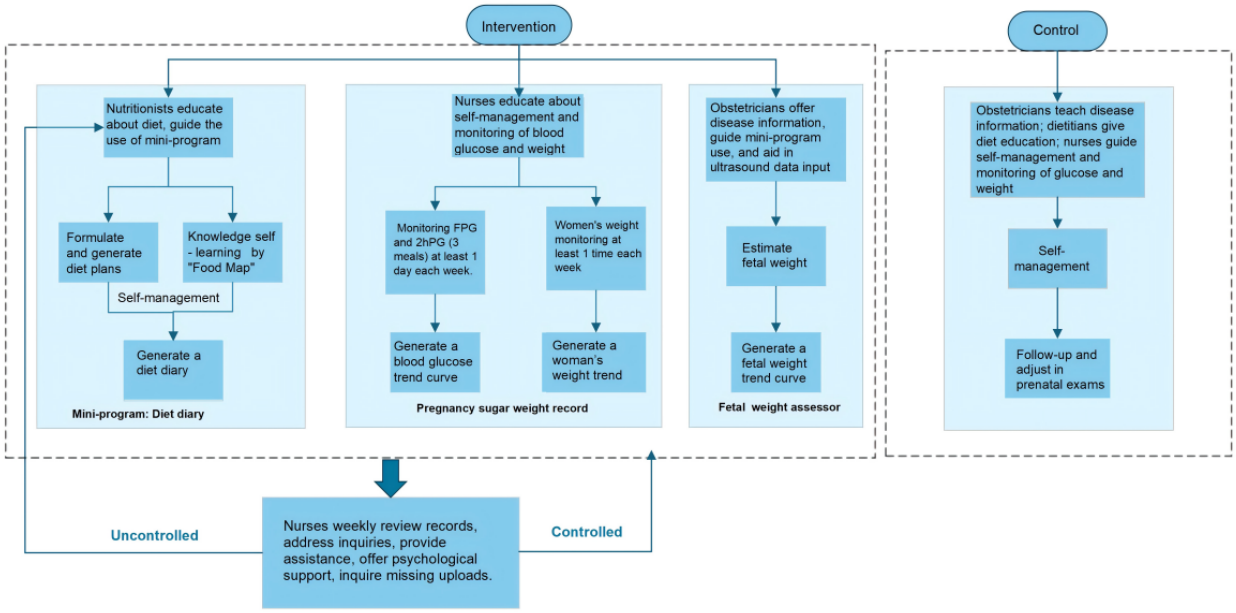
Intervention Process

Upon enrollment, each participant received an individualized MNT prescription developed by a study dietitian using the "Diet Diary" mini-program, which incorporated personal parameters (age, gestational age, and pBMI) and specified precise gram weights of food types for daily meals and snacks. Participants recorded all food and beverage intake (type and weight) daily via the same mini-program, with digital food scales provided to support accurate measurement, and used the "Food Charts" self-learning module to learn as needed about food selection, GI values, and portion control techniques. Concurrently, women logged blood glucose and body weight data weekly into the "Pregnancy Sugar Weight Record" mini-program for continuous health monitoring. Those unable to input data independently could use paper records, audio or video recordings, or photos, with nurses assisting with data entry. Obstetricians delivered disease information to the participants, instructed them on the operation of the "Fetal Weight Assessor" mini-program, and supported the accurate input of ultrasound examination data during scheduled antenatal visits. Following this step, the system calculated and estimated fetal weight using the collected ultrasound parameters. A dynamic fetal weight trend curve was generated to visualize the changes in fetal weight over time.

A multidisciplinary team (certified dietitians, nurses, and obstetricians) provided comprehensive support throughout the study. Diabetes nurses reviewed participants' dietary logs, blood glucose levels, and weight records weekly via the WeChat backend, delivering personalized feedback and evaluations through WeChat messages or voice calls—acknowledging adherence achievements, addressing inquiries, and offering constructive suggestions. For participants with persistently unmet glycemic targets, nurses collaborated with dietitians to analyze issues and adjust meal plans (eg, modifying carbohydrate distribution or introducing alternative foods). A real-time consultation channel ensured that dietary inquiries received responses within 12 hours, and the team routinely audited 20% of the records to maintain intervention adherence standards. A nurse-modified WeChat peer support group facilitated weekly discussions in which women exchanged experiences, practical strategies, and insights into glycemic management and MNT implementation, fostering mutual encouragement. For participants facing persistent glucose management challenges for more than 2 weeks, the care team arranged 15 to 30 minutes of individual video consultations to provide focused support; these sessions could also lead to referrals for further multidisciplinary evaluation if needed. Individualized WeChat communications (messages, voice recordings, and video calls) were additionally provided for participants with specific needs, offering one-on-one dietary consultation and guidance. Throughout the intervention, all communications maintained an empathetic, nonjudgmental tone, and participants retained the flexibility to adjust their involvement or pause monitoring as needed. Figure 1 shows the implementation process of the study.

XSL•FO
RenderX

Figure 1. Implementation process of the study. FPG: fasting blood glucose; 2hPG: 2-hour postprandial glucose.



Control Group: Standard Antenatal Care

Participants in the control group received routine GDM management in accordance with national guidelines. This included dietary guidance through printed educational materials, such as the Pagoda of Balanced Diet for Pregnant Women, which emphasized whole grains, lean proteins, and low-GI fruits. Biweekly 30-minute counseling sessions were conducted to address general dietary principles and portion sizes. Clinical monitoring involved regular weight, blood pressure, and capillary blood glucose measurements (fasting and postprandial) during scheduled antenatal visits, with participants recording self-monitored glucose values in paper logs. Additionally, exercise recommendations were provided to all study participants, including both the intervention and control groups, who were advised to engage in moderate-intensity walking for 30 minutes, 3 to 4 times weekly. Table 2 provides a detailed comparison of management approaches across modules between the WeMNT intervention and control groups.

Table . Comparison of management approaches across modules: WeChat-delivered medical nutrition therapy (WeMNT) versus standard care.

Component	Intervention group (WeMNT)	Control group
Dietary education, planning, and monitoring	Knowledge self-learning by "Food Charts" using the GI[a] database (476 foods), AI[b]-driven personalized meal plans, and real-time nutrient monitoring and analysis	Static PDF guidelines and general portion advice
Weight monitoring	Automated IOM[c] guideline alerts and trend visualization dashboards	Manual clinic measurements and verbal feedback
Glucose tracking	App-integrated data synchronization and weekly dietitian review	Paper-based logs and monthly physician review
Fetal surveillance	Ultrasound parameter input for fetal weight estimation and growth percentile curves	Standard ultrasound reports
Review, feedback, and support frequency	Daily self-monitoring+weekly professional feedback	Biweekly counseling+routine antenatal check-ups

[a]GI: glycemic index.

[b]AI: artificial intelligence.

[c]IOM: Institute of Medicine.

Data Collection

Data collection was performed by research nurses who were blinded to group allocation. Baseline data were collected during the OGTT, which was conducted between the 24th and 28th gestational weeks, and the intervention was initiated at the 28th gestational week. Two intermediate follow-up assessments were planned from intervention initiation to delivery, scheduled at the 32nd (1-mo follow-up) and 36th weeks (2-mo follow-up). The study end point was defined as delivery, including preterm delivery (consistent with clinical criteria, ie, delivery between the 28th and 37th gestational weeks) and term delivery (delivery

XSL•FO
RenderX

at or after the 38th gestational week). Participants with preterm delivery may miss the follow-up assessments scheduled thereafter and reach the study end point ahead of schedule (early closure) compared to term delivery (routine closure). Research nurses contacted participants by phone or WeChat to schedule data collection sessions, and all assessments were completed by participants in the obstetric clinic, with onsite researcher assistance available if needed. HbA_{1c} was measured twice, specifically at baseline and at the end point. To prevent contamination between groups, data collection procedures for the intervention and control groups were scheduled on separate dates at the study site.

Ethical Considerations

This study was approved by the hospital institutional ethics committee (approval 2023-109-02) and abided by the principles of the Declaration of Helsinki. All patients who participated in the study participated voluntarily and obtained the informed consent from each patient in writing. Patients had the right to withdraw from the study at any time. All exported data must undergo anonymization by the data manager before statistical analysis can be conducted. Individuals cannot be identified in any results presentation. No financial compensation was provided to avoid coercion.

Outcome Measures

The primary outcome measures of this study included glycemic control status and weight changes during pregnancy. Glycemic control was assessed using fasting blood glucose (FPG) and 2hPG levels, both of which were measured on a Hitachi 7600 analyzer (Hitachi High-Technologies Corporation) using venous plasma samples. Additionally, HbA_{1c} was determined through high-performance liquid chromatography. Gestational weight gain (GWG) was defined as the weight accrued during pregnancy and was computed as the numerical difference between the measured weight values at each follow-up time point and the baseline. The secondary outcome measures included obstetric complications and neonatal outcomes, such as gestational hypertension, preterm birth (<37 wk), and neonatal birth weight. The amniotic fluid index (AFI) was included as a secondary outcome because of its reported association with GDM-related fetal oligohydramnios and polyhydramnios, which may increase the risks of preterm labor and cesarean delivery. AFI was measured by ultrasound using the 4-quadrant technique after the 36th week and prior to delivery, with oligohydramnios defined as an AFI<5 cm and polyhydramnios as an AFI>25 cm, respectively. All measurements were conducted in strict adherence to standardized clinical practices.

Statistical Analysis

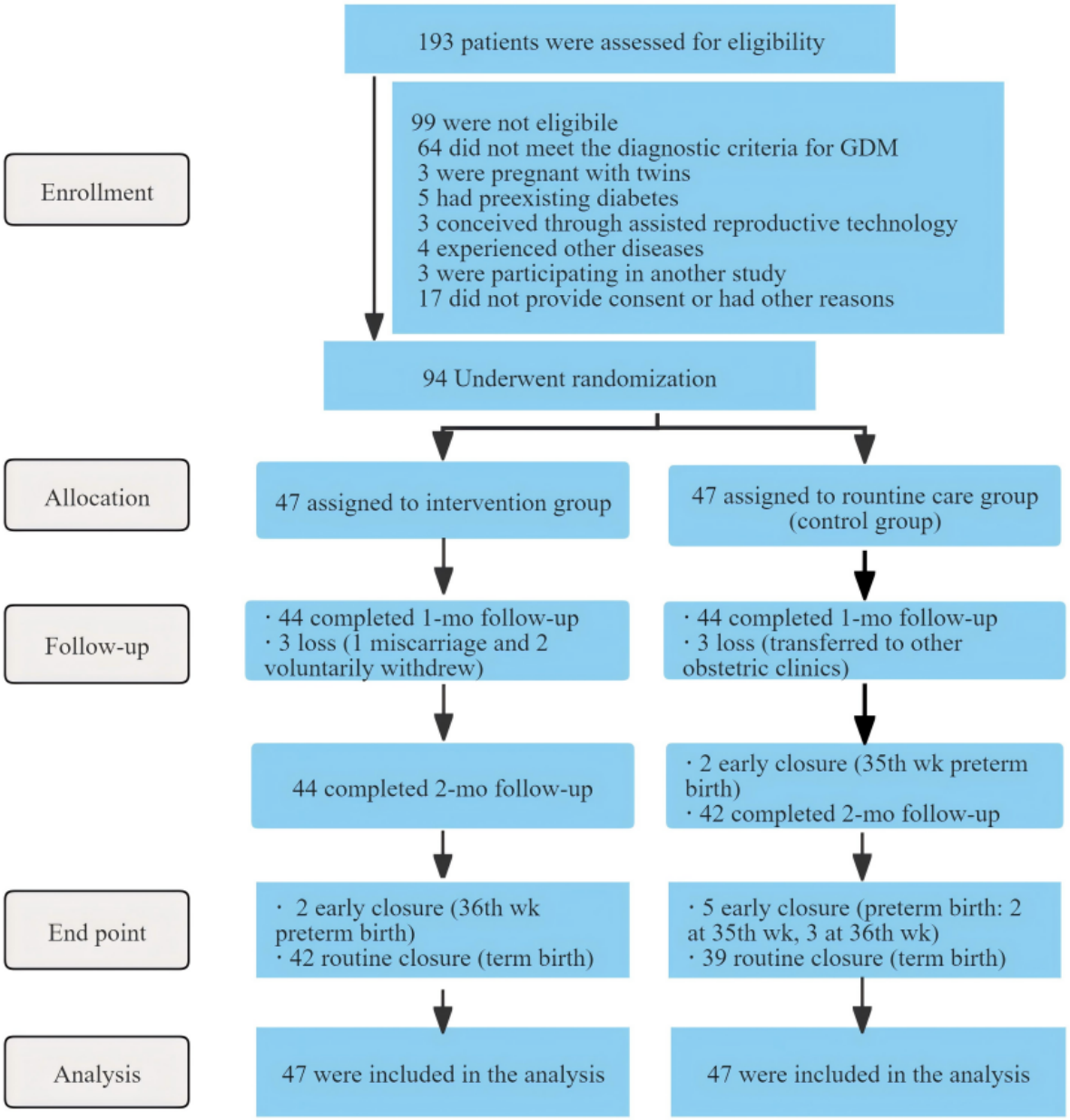
Descriptive statistical analyses were performed, whereby continuous variables were summarized as means (SDs) and categorical variables were presented as frequencies and corresponding percentages. The primary efficacy analysis adhered to the intention-to-treat (ITT) principle, wherein all randomized participants (n=94) were analyzed according to their original group assignments—this included individuals who were excluded during the study or lost to follow-up. Among the 94 participants, 6 participants (6.38%) had transferred to other clinics, experienced miscarriage, or withdrew (missing data, Figure 2). Little's missing completely at random tests were performed on the postintervention outcome variables. The results were not statistically significant (P=.17 and P>.05), indicating that the missing data mechanism was completely at random.

Generalized linear mixed models (GLMMs) were used to assess the intervention's effects on primary outcomes (FPG, 2hPG, and GWG) across different time points: 4 (time: baseline, 1 mo, 2 mo, and end point)×2 (group: intervention vs control). General sociodemographic characteristics and blood glucose management–related data were incorporated as covariates in this model to control for potential confounding factors. Additionally, a distinct 2 (time: baseline and end point)×2 (group: intervention vs control) GLMM was established to examine longitudinal changes in HbA_{1c} levels and to compare between-group differences in HbA_{1c}. This analytical approach was selected because it can account for within-subject correlations and manage missing data under the missing-at-random assumption, enabling the use of all available data points without the need for multiple imputation. All GLMMs incorporated fixed effects for 3 key terms: the intervention group, time point, and the group×time interaction between group and time; a random intercept for each participant (1|ID) was also included to account for interindividual variability. To verify the robustness of the study findings, a per-protocol analysis was additionally performed, including only participants who completed the full intervention course and all scheduled follow-up assessments.

The secondary outcomes (maternal and fetal clinical measures) were analyzed using per-protocol methods. For the assessment of between-group disparities, we used independent-samples 2-tailed t tests to compare continuous variables (eg, AFI) and chi-square tests to analyze categorical variables such as the incidence of preeclampsia. All analyses were conducted using R software (R Foundation for Statistical Computing). The GLMMs were fitted using the lme4 package. A 2-sided P value <.05 was considered to indicate statistical significance.

XSL•FO
RenderX

Figure 2. The participant flow chart of the study. GDM: gestational diabetes mellitus.



Results

Overview

Between March 2023 and October 2023, a total of 94 participants were recruited for this study. Among them, 88 (93.62%) participants (mean age 32.18, SD 5.04 y) successfully completed the study: 44 (46.81%) participants in the intervention group and 44 (46.81%) in the control group. No adverse events or unintended effects were observed in either group throughout the study period. The participant flow and reasons for dropout are shown in Figure 2.

Participant Engagement

The backend records of the WeChat app showed that in the intervention group, 87.2% (41/47) of the participants used the "Diet Diary" mini-program for daily dietary logging, whereas 78.7% (37/47) used the "Pregnancy Sugar Weight Record" and "Fetal Weight Assessor" programs for weight tracking. For the remaining participants who encountered difficulties inputting the data themselves, nurses assisted in entering the information into the mini-programs. All the participants explored the "food charts" module, learning the GI values of 476 foods, food selection, and weighing. Initially, they engaged with these tools 2 to 3 times daily for 20 to 45 minutes, but usage decreased as they became more familiar with food attributes. They only revisited the information when needed. Among the 44 participants who completed the study, 54.5% (24/44) adhered to the diet plan ≥80% of the days, 13.6% (6/44) followed it for 60% to 79% of the days, 11.4% (5/44) followed it for 40% to 59% of the days, another 11.4% (5/44) for 20% to 39% of the

XSL•FO
RenderX

days, and 9.1% (4/44) for <20% of the days. In the control group, all 44 participants who completed the study reported independently studying the provided materials and following the dietary guidance for 60% to 80% of the days. No harms or unintended effects were found in each group.

Baseline Demographic Characteristics and Clinical Data

The sample had a mean age of 32.18 (SD 5.04; range 22 - 45) years (Table 3). The mean pBMI was 24.97 (SD 4) kg/m². For glycemic indicators, the mean FPG was 5.39 (SD 0.67) mmol/L, the mean 2hPG was 8.66 (SD 1.24) mmol/L, and the mean HbA_{1c} in the sample was 7.2% (SD 0.69).

Table . Baseline data across the intervention and control groups.

Characteristics	Total (N=94)	Intervention (n=47)	Control (n=47)
Age (y), mean (SD)	32.18 (5.04)	32.74 (4.91)	31.62 (5.15)
pBMI[a] (kg/m^2), mean (SD)	24.97 (4)	24.58 (4.47)	25.37 (3.48)
FPG[b] (mmol/L), mean (SD)	5.39 (0.67)	5.37 (0.83)	5.40 (0.47)
2hPG[c] (mmol/L), mean (SD)	8.66 (1.24)	8.70 (1.33)	8.41 (1.69)
HbA_{1c}[d] (%), mean (SD)	7.2 (0.69)	7.23 (0.68)	7.16 (0.71)
Weight status, n (%)			
Normal weight	48(51.0)	24 (51.06)	24 (51.06)
Overweight	37(39.4)	18 (38.30)	19 (40.43)
Obese	9 (9.6)	5 (10.64)	4 (8.51)
Household income (RMB per month), n (%)			
<3000	7 (7.4)	3 (6.38)	4 (8.51)
3000 - 6000	24 (25.5)	20 (42.55)	17 (36.17)
6000 - 9000	40 (42.6)	19 (40.43)	20 (42.55)
>9000	23 (24.5)	5 (10.64)	6 (12.77)
Education years, n (%)			
1 - 9	8 (8.5)	6 (12.77)	2 (4.26)
10 - 12	30 (31.9)	14 (29.79)	16 (34.04)
13 - 16	45 (47.9)	21 (44.68)	24 (51.06)
≥17	11(11.7)	6 (12.77)	5 (10.64)

[a]pBMI: prepregnancy BMI.

[b]FPG: fasting blood glucose.

[c]2hPG: 2-hour postprandial glucose.

[d]HbA_{1c}: hemoglobin A_{1c}.

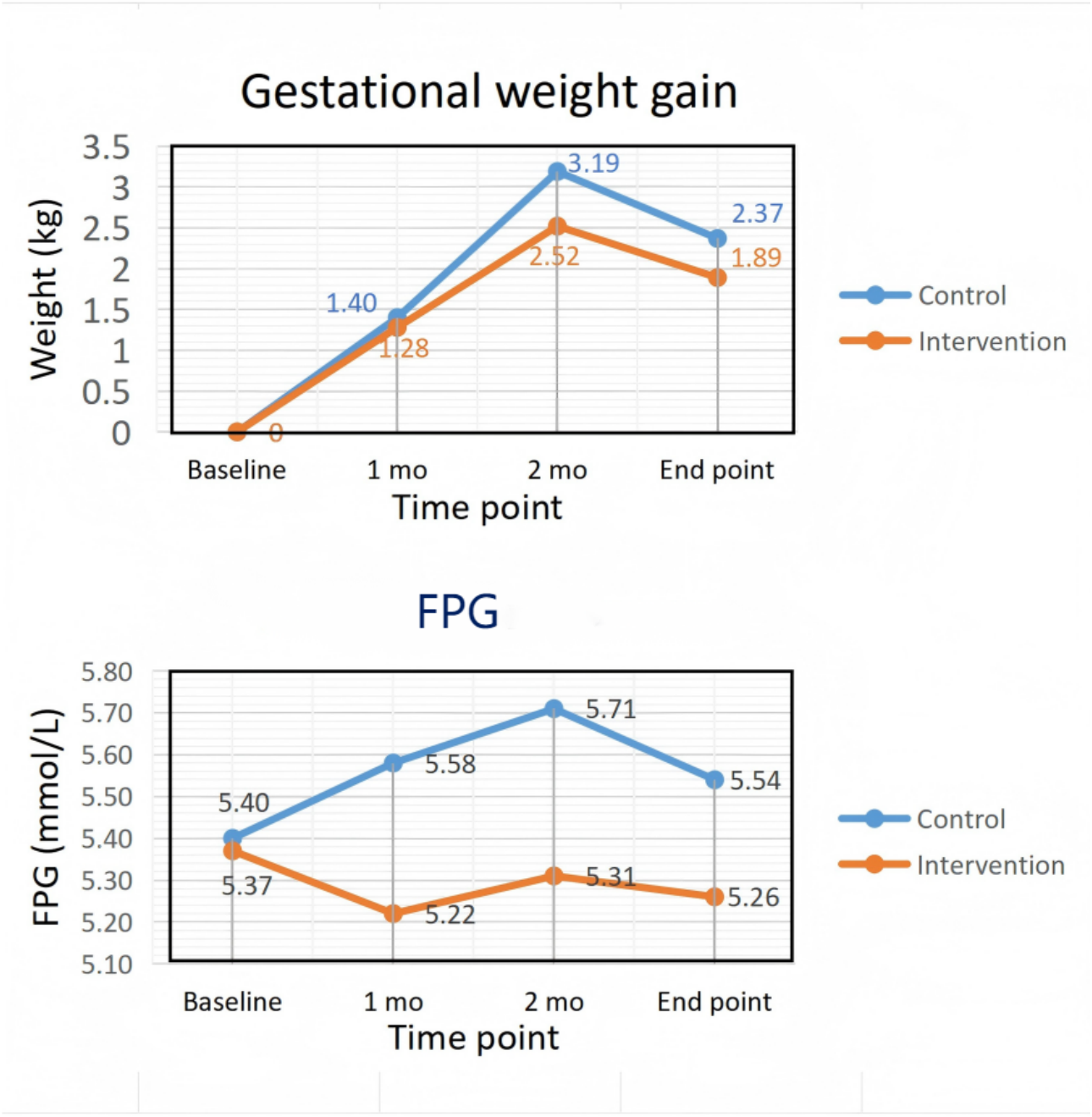
Changes in Primary Outcomes Over Time Between the mHealth and Control Groups

The GWG progressively increased over time in both groups, with the control group exhibiting greater increases (Figure 3). By the end point, the control group's mean weight had increased by 15.11 (SD 12.56) kg, whereas the intervention group's mean weight had increased by 13.31 (SD 12.21) kg (Table 4).

FPG levels fluctuated between the groups (Figure 3). In the control group, the FPG concentration increased by 0.17 (SD 0.61) mmol/L at 1 month and 0.32 (SD 0.70) mmol/L at 2 months but then slightly decreased by 0.15 (SD 0.62) mmol/L at the end point (Table 4). In contrast, the levels in the intervention group gradually decreased, with FPG levels decreasing by 0.15 (SD 0.70) mmol/L at 1 month, 0.06 (SD 0.78) mmol/L at 2 months, and 0.11 (SD 0.56) mmol/L at the end point.

XSL•FO
RenderX

Figure 3. Changes in gestational weight gain and fasting blood glucose (FPG) between the intervention and control groups over time.



XSL•FO
RenderX

Table . Comparisons of the primary outcomes between the intervention and control groups.

Variables	Control group				Intervention group			
	Baseline (n=47)	1 mo (n=44)	2 mo (n=44)	End point (n=44)	Baseline (n=47)	1 mo (n=44)	2 mo (n=42)	End point (n=44)
GWG[a] (kg), mean (SD)	0	1.40 (0.37)	3.19 (0.77)	2.37 (0.67)	0	1.18 (0.32)	2.52 (0.63)	1.89 (0.51)
FPG[b] (mmol/L), mean (SD)	5.39 (0.49)	5.56 (0.61)	5.71 (0.70)	5.54 (0.62)	5.37 (0.86)	5.22 (0.70)	5.31 (0.78)	5.26 (0.56)
2hPG[c] (mmol/L), mean (SD)	8.49 (1.08)	8.62 (1.38)	8.79 (1.21)	8.91 (1.23)	8.60 (1.30)	8.34 (1.69)	8.25 (1.56)	8.08 (1.51)
HbA_{1c}[d,e] (%), mean (SD)	7.14 (0.72)	N/A[f]	N/A	6.71 (0.62)	7.21 (0.70)	N/A	N/A	6.29 (0.75)

[a]GWG: gestational weight gain

[b]FPG: fasting blood glucose.

[c]2hPG: 2-hour postprandial glucose.

[d]HbA_{1c}: hemoglobin A_{1c}.

[e]HbA_{1c} was measured only at baseline and at the final end point of the study.

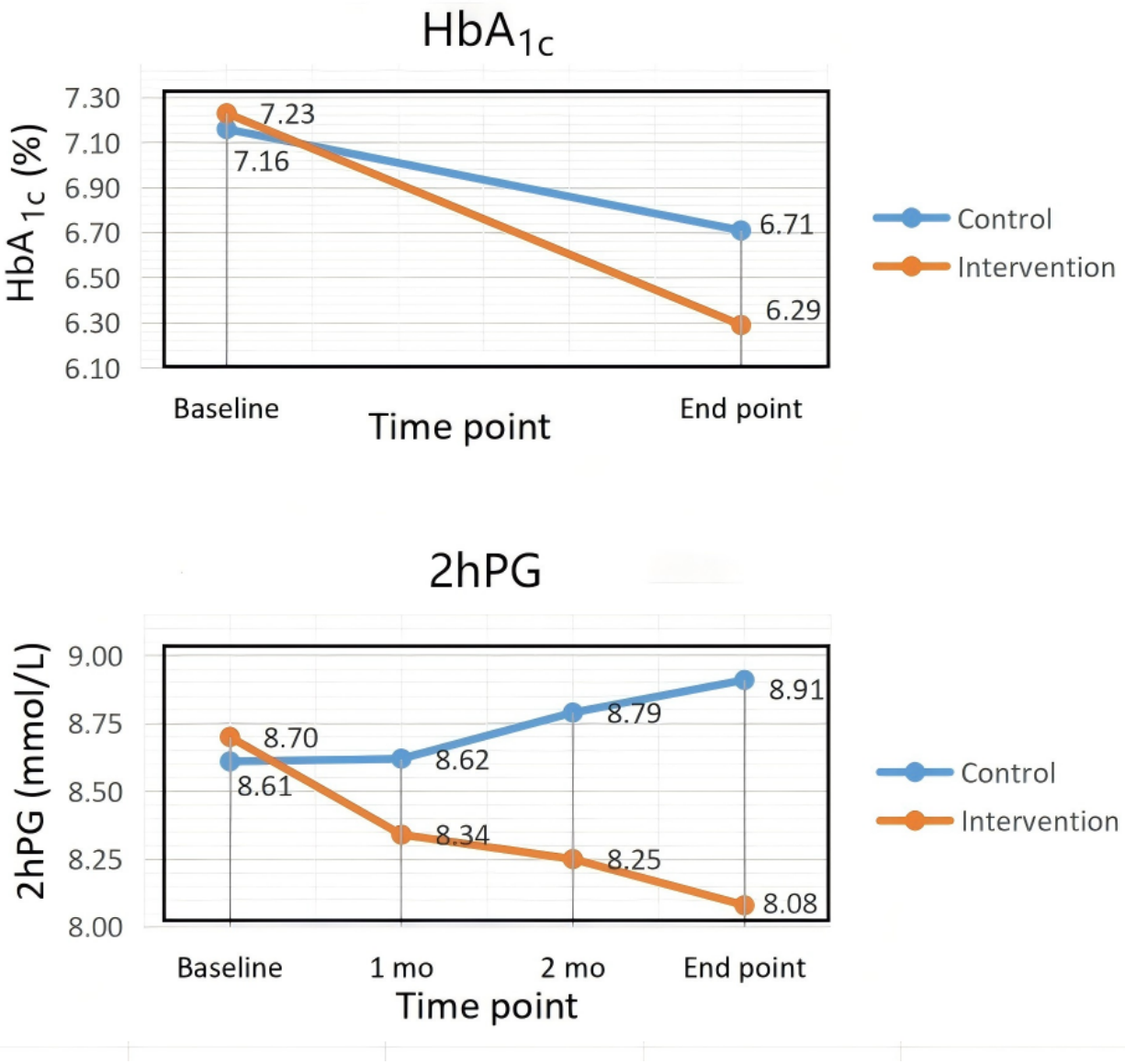
[f]N/A: not applicable.

Both groups showed a decreasing trend in HbA_{1c} levels (Figure 4), with the intervention group demonstrating substantially greater decreases. By the end point, the control group's mean HbA_{1c} had decreased to 6.71% (SD 0.62), representing a reduction of 0.43% (Table 4). The mean HbA_{1c} in the intervention group decreased to 6.29% (SD 0.75), indicating a more pronounced reduction of 0.92% from baseline.

Both groups showed an increase in 2hPG over time, but the intervention group demonstrated greater decreases (Figure 4). By the end point, the control group's mean 2hPG had risen to 8.91 (SD 1.23) mmol/L, representing an increase of 0.42 mmol/L (Table 4). In contrast, the mean 2hPG of the intervention group decreased to 8.08 (SD 1.51) mmol/L, indicating a reduction of 0.52 mmol/L from baseline.

XSL•FO
RenderX

Figure 4. Changes in hemoglobin A1c (HbA_{1c}) and 2-hour postprandial glucose (2hPG) between the intervention and control groups over time.



Effectiveness of the Intervention Through the ITT Analysis

Adjusted GLMMs with ITT analysis revealed a significant group×time interaction for GWG (Table 5); notably, the intervention group achieved a significantly greater reduction in GWG over time compared with the control group, as indicated by the negative coefficient of the group×time interaction (β=−1.96, 95% CI −3.58 to −0.34; *P*=.02). For FPG, no significant group×time interaction effect was observed. In contrast, both 2hPG and HbA_{1c} showed significant group×time interaction effects. Compared with the control group, the intervention group demonstrated substantially greater decreases in 2hPG (β=−0.12, 95% CI −0.19 to −0.04; *P*<.001) and HbA_{1c} (β=−0.49, 95% CI −0.74 to −0.23; *P*<.001) over time. The results of the per-protocol analyses (Multimedia Appendix 2) were consistent in terms of the direction and significance of the ITT analyses.

XSL•FO
RenderX

Table . Generalized linear mixed model analyses of the efficacy of the intervention on the primary outcomes.

Outcome	β estimate (95% CI)	SE	*P* value
GWG[a] (kg)			
Time	15.10 (13.96 to 16.25)	0.59	<.001
Group	−2.29 (−7.24 to 2.67)	2.53	.37
Group×time	−1.96 (−3.58 to −0.34)	0.83	.02
FPG[b] (mmol/L)			
Time	0.02 (−0.02 to 0.07)	0.02	.26
Group	−0.01 (−0.06 to 0.04)	0.02	.65
Group×time	−0.04 (−0.10 to 0.02)	0.03	.18
2hPG[c] (mmol/L)			
Time	0.04 (−0.02 to 0.09)	0.03	.22
Group	0.02 (−0.05 to 0.08)	0.03	.65
Group×time	−0.12 (−0.19 to −0.04)	0.04	<.001
HbA_{1c}[d] (%)			
Time	−0.44 (−0.62 to −0.26)	0.09	<.001
Group	0.07 (−0.21 to 0.35)	0.14	.64
Group×time	−0.49 (−0.74 to −0.23)	0.13	<.001

[a]GWG: gestational weight gain.

[b]FPG: fasting blood glucose.

[c]2hPG: 2-hour postprandial glucose.

[d]HbA_{1c}: hemoglobin A_{1c}.

Maternal and Fetal Outcomes

No unexpected maternal events were observed among the participants. Compared with that in the control group, the AFI in the intervention group was significantly different (Table 6). However, both values were within the normal range and were not expected to have adverse effects on the fetuses. The number of cases of preeclampsia, preterm infants, low-birth-weight infants (birth weight <2500 g), and macrosomia (birth weight >4000 g) were lower in the intervention group than those in the control group, yet the differences were not statistically significant (Table 6).

Table . Comparisons of the secondary outcomes between the intervention and control groups.

Group	AFI[a,b,c] (cm), mean (SD)	Preeclampsia[d,e], n (%)	Preterm birth[f,g], n (%)	Low birth weight[h,i], n (%)	Macrosomia[j,k], n (%)
Intervention (n=44)	12.01 (3.31)	1 (2.27)	2 (4.54)	1 (2.27)	4 (9.09)
Control (n=44)	13.49 (3.19)	3 (6.82)	5 (11.36)	3 (6.82)	7 (15.91)

[a]AFI: amniotic fluid index.

[b]t_{86}=−2.15 (2 sided).

[c]*P*=.04.

[d]X^2_1=0.26.

[e]*P*=.60.

[f]X^2_1=0.62.

[g]*P*=.43.

[h]X^2_1=0.26.

[i]*P*=.60.

[j]X^2_1=0.94.

[k]*P*=.33.

XSL•FO
RenderX

Discussion

Principal Findings

Collectively, the WeMNT intervention demonstrated targeted efficacy in improving glycemic control among women with GDM, particularly in modulating GWG, 2hPG, and HbA_{1c} dynamics. While it failed to have a meaningful effect on FPG, the favorable changes in 2hPG and HbA_{1c} trajectories highlight the intervention's potential as a practical tool for GDM management. These findings underscore the importance of tailored dietary interventions to address specific glycemic outcomes during pregnancy, with further research needed to optimize the interventions for broader efficacy.

Prior studies on HbA_{1c} management have demonstrated that WeChat-delivered intervention tools—validated across 6 RCTs—yield a notable effect size of 0.84 (95% CI –0.22 to −1.46) [21]. Remarkably, in our MNT intervention cohort, HbA_{1c} levels substantially decreased. This promising outcome is most likely due to the mHealth technology we crafted. An mHealth intervention based on a behavioral model, whether used independently or in tandem with other approaches, confers a distinct advantage for effective blood glucose control [25]. Continuous self-tracking of blood glucose, a common intervention measure in GDM management, clarifies the links between glucose levels and nutrition, thereby supporting autonomous self-management without the need for direct health care assistance [26]. Using key BCTs in diabetes management, our project provides timely feedback and evaluations based on records and indicator data, acknowledging strengths and making improvement suggestions through WeChat instant messages and voice calls. It not only facilitates self-learning and monitoring of GDM and communication with health care professionals but also establishes WeChat peer groups to encourage experience sharing and mutual encouragement among patients. This feature effectively sustains participants' engagement and active involvement in managing their own care, a critical aspect of the intervention [27,28].

Evidence indicates that when HbA_{1c} levels are less than 7.3%, 2hPG may contribute up to 70% of total HbA_{1c} elevation [29]. In our GDM cohort, FPG had a mean of 5.38 (SD 0.69) mmol/L, with an average HbA_{1c} of 7.17% (SD 0.71%), suggesting that the increase in HbA_{1c} in this population is largely attributable to increased 2hPG levels. Given this relationship, dietary variations have a pronounced effect on 2hPG, rendering this parameter highly responsive to dietary interventions; HbA_{1c} consequently shares this sensitivity, as its elevation is closely linked to 2hPG fluctuations. In contrast, the intervention yielded no significant effects on FPG levels. These findings are consistent with the results of a meta-analysis focusing on dietary interventions for GDM, which reported a significant 2hPG reduction of 27.78 mg/dL (P<.001) and a transient FPG reduction of 24.07 mg/dL (P=.02) in intervention groups [6]; however, only the abovementioned results align with common observations in the field: FPG levels often do not significantly respond to dietary interventions in individuals with GDM [30].

Maternal weight gain throughout pregnancy constitutes a fundamental physiological process of gestation. Our regression model revealed a highly significant time main effect (β=15.10, 95% CI 13.96 - 16.25; P<.001), indicating that a substantial increase in maternal weight occurs as pregnancy progresses. Notably, mobile phones have emerged as a promising tool for GWG management in recent years. Prior studies have consistently demonstrated that smartphone apps exhibit remarkable efficacy in promoting healthy weight gain during pregnancy [31,32]. These technologies overcome the limitations of traditional face-to-face guidance by providing personalized dietary recommendations, supporting pregnant women in maintaining optimal weight trajectories during gestation. The findings of this study validated the effectiveness of mHealth interventions in controlling GWG. Our regression analysis revealed a significant group×time interaction effect (β=−1.96, 95% CI −3.58 to −0.34; P=.02), indicating a significantly gentler pattern of weight gain in the intervention group compared with the control group. While the effect size was small and its impact on final weight gain was limited, its value in controlling dynamic weight growth remained undiminished. These results align with the conclusions of previous studies, further confirming the potential of mHealth tools as an adjunct to routine antenatal care.

The unique advantages of our mHealth intervention lie in its targeted timing and strategy design, which address the unmet needs of pregnant women in the third trimester. Our intervention started at week 28 (the start of the third trimester), a time when physical constraints hinder exercise-based weight management [25]. Unlike exercise-reliant programs, it integrates core nutritional guidance (BCT 4.1/4.2), MNT plan design support (BCT 6.1), scheduled meal reminders (BCT 1.4), food portion logging, and real-time intake target alerts—helping participants use science-backed diets to curb excessive weight gain amid inevitable late-pregnancy physiological weight growth. A second key advantage is its focus on collaborative clinical and self-management efforts, which are critical for addressing inappropriate GWG and protecting maternal and child health. This dual-focused approach bridges the gap between clinical recommendations and real-world adherence.

Regarding other maternal and fetal outcomes, a comprehensive meta-analysis of 32 RCTs confirmed that mHealth strategies effectively improve glycemic control and reduce the risks of adverse maternal and fetal outcomes [19]. Specifically, the risk ratios were 0.49 for macrosomia (95% CI 0.30 - 0.80; P<.01), 0.48 for preeclampsia (95% CI 0.40 - 0.58; P<.01), and 0.27 for preterm birth (95% CI 0.20 - 0.35; P<.01). However, our study revealed no significant effects of the intervention on these outcomes or birth weight. Notably, our limited sample size may have restricted the detection of subtle effects, highlighting the need for further investigation of these endpoints in future research.

Limitations

This study has several limitations. First, it was impossible to implement blinding for the personnel involved in the WeMNT intervention. To address this issue, we crafted a protocol to guide the intervention, refined the guidance to align with it, and

XSL•FO
RenderX

equipped the team with the skills to execute it effectively. Second, this study, which was conducted in a single region with a homogeneous population, has limitations in terms of generalizability. To bridge the implementation gap, we documented each mHealth module's design and implementation, aligned with those of the BCW. This supports cross-setting validation and real-world generalization. Third, our study focused on short-term outcomes, and long-term effects on mothers and children remain undetermined. Future research should recruit larger, heterogeneous cohorts to evaluate the efficacy across diverse populations and extended timeframes. Finally, although we adjusted for key covariates in the sensitivity analyses, there may still be unmeasured confounders (eg, genetic predisposition and family medical history) that could influence the study outcomes. Future research should use multivariate regression or mixed effects models to achieve more effective control over these factors.

Conclusions

This study demonstrated that the WeMNT program effectively improved nutritional self-management and blood glucose control in patients, with significant reductions in 2hPG, HbA_{1c}, and GWGs, supporting its clinical adoption for GDM. However, long-term maternal-child outcomes remain unexplored and warrant larger, diverse cohort studies for extended evaluation.

Funding

This work was supported by Science and Technology Development Program of Jinan Municipal Health Commission (2024205001).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

LM directed the project's design, execution, and overall coordination. XW developed the intervention strategy, led project outreach and administration, performed data analysis, and drafted the manuscript. MB implemented project activities, collected clinical data, and assisted with analysis. FL supported clinical data collection. CZ analyzed the data, refined all manuscript sections, and finalized the draft. All authors reviewed and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the WeChat mini-programs in the WeMNT intervention.
[PDF File, 1906 KB - mhealth_v14i1e67410_app1.pdf]

Multimedia Appendix 2
The results of the per-protocol analyses of the efficacy of the intervention on the primary outcome.
[DOCX File, 15 KB - mhealth_v14i1e67410_app2.docx]

Checklist 1
CONSORT-EHEALTH (V 1.6.1) checklist.
[PDF File, 4689 KB - mhealth_v14i1e67410_app3.pdf]

References

1. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care 2014 Jan;37 Suppl 1:S81-S90. [doi: 10.2337/dc14-S081] [Medline: 24357215]
2. Magliano DJ, Boyko EJ. International Diabetes Federation Diabetes Atlas 10th Edition Scientific Committee, 10th edition: International Diabetes Federation; 2021.
3. Choudhury AA, Devi Rajeswari V. Gestational diabetes mellitus - A metabolic and reproductive disorder. Biomed Pharmacother 2021 Nov;143:112183. [doi: 10.1016/j.biopha.2021.112183] [Medline: 34560536]
4. Ye W, Luo C, Huang J, Li C, Liu Z, Liu F. Gestational diabetes mellitus and adverse pregnancy outcomes: systematic review and meta-analysis. BMJ 2022 May 25;377:e067946. [doi: 10.1136/bmj-2021-067946] [Medline: 35613728]
5. American Diabetes Association Professional Practice Committee. 15. Management of diabetes in pregnancy: standards of care in diabetes-2025. Diabetes Care 2025 Jan 1;48(1 Suppl 1):S306-S320. [doi: 10.2337/dc25-S015] [Medline: 39651985]
6. Yamamoto JM, Kellett JE, Balsells M, et al. Gestational diabetes mellitus and diet: a systematic review and meta-analysis of randomized controlled trials examining the impact of modified dietary interventions on maternal glucose control and neonatal birth weight. Diabetes Care 2018 Jul;41(7):1346-1361. [doi: 10.2337/dc18-0102] [Medline: 29934478]

XSL•FO
RenderX

7. Wei J, Heng W, Gao J. Effects of low glycemic index diets on gestational diabetes mellitus: a meta-analysis of randomized controlled clinical trials. Medicine (Baltimore) 2016 May;95(22):e3792. [doi: 10.1097/MD.0000000000003792] [Medline: 27258511]
8. Hernandez TL, Mande A, Barbour LA. Nutrition therapy within and beyond gestational diabetes. Diabetes Res Clin Pract 2018 Nov;145:39-50. [doi: 10.1016/j.diabres.2018.04.004] [Medline: 29679625]
9. Sweeting A, Mijatovic J, Brinkworth GD, et al. The carbohydrate threshold in pregnancy and gestational diabetes: how low can we go? Nutrients 2021 Jul 28;13(8):2599. [doi: 10.3390/nu13082599] [Medline: 34444759]
10. Vasile FC, Preda A, tefan AG, et al. An update of medical nutrition therapy in gestational diabetes mellitus. J Diabetes Res 2021;2021:5266919. [doi: 10.1155/2021/5266919] [Medline: 34840988]
11. Vesco KK, Karanja N, King JC, et al. Efficacy of a group-based dietary intervention for limiting gestational weight gain among obese women: a randomized trial. Obesity (Silver Spring) 2014 Sep;22(9):1989-1996. [doi: 10.1002/oby.20831] [Medline: 25164259]
12. Wei X, Zou H, Zhang T, et al. Gestational diabetes mellitus: what can medical nutrition therapy do? Nutrients 2024 Apr 19;16(8):1217. [doi: 10.3390/nu16081217] [Medline: 38674907]
13. Karavasileiadou S, Almegwely W, Alanazi A, Alyami H, Chatzimichailidou S. Self-management and self-efficacy of women with gestational diabetes mellitus: a systematic review. Glob Health Action 2022 Dec 31;15(1):2087298. [doi: 10.1080/16549716.2022.2087298] [Medline: 35867537]
14. Parsons J, Sparrow K, Ismail K, Hunt K, Rogers H, Forbes A. Experiences of gestational diabetes and gestational diabetes care: a focus group and interview study. BMC Pregnancy Childbirth 2018 Jan 11;18(1):25. [doi: 10.1186/s12884-018-1657-9] [Medline: 29325518]
15. Wah YYE, McGill M, Wong J, Ross GP, Harding AJ, Krass I. Self-management of gestational diabetes among Chinese migrants: a qualitative study. Women Birth 2019 Feb;32(1):e17-e23. [doi: 10.1016/j.wombi.2018.03.001] [Medline: 29693546]
16. Hewage S, Audimulam J, Sullivan E, Chi C, Yew TW, Yoong J. Barriers to gestational diabetes management and preferred interventions for women with gestational diabetes in Singapore: mixed methods study. JMIR Form Res 2020 Jun 30;4(6):e14486. [doi: 10.2196/14486] [Medline: 32602845]
17. Sabag A, Houston L, Neale EP, et al. Supports and barriers to lifestyle interventions in women with gestational diabetes mellitus in Australia: a national online survey. Nutrients 2023 Jan 17;15(3):487. [doi: 10.3390/nu15030487] [Medline: 36771195]
18. Safiee L, Rough D, George P, Mudenha R. Baseline perceptions of women with gestational diabetes mellitus and health care professionals about digital gestational diabetes mellitus self-management health care technologies: interview study among patients and health care professionals. JMIR Hum Factors 2023 Dec 19;10:e51691. [doi: 10.2196/51691] [Medline: 38113070]
19. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42. [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
20. International Association of Diabetes and Pregnancy Study Groups Consensus Panel, Metzger BE, Gabbe SG, et al. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care 2010 Mar;33(3):676-682. [doi: 10.2337/dc09-1848] [Medline: 20190296]
21. Xie W, Dai P, Qin Y, Wu M, Yang B, Yu X. Effectiveness of telemedicine for pregnant women with gestational diabetes mellitus: an updated meta-analysis of 32 randomized controlled trials with trial sequential analysis. BMC Pregnancy Childbirth 2020 Apr 6;20(1):198. [doi: 10.1186/s12884-020-02892-1] [Medline: 32252676]
22. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160 [FREE Full text] [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]
23. Kebede MM, Liedtke TP, Möllers T, Pischke CR. Characterizing active ingredients of eHealth interventions targeting persons with poorly controlled type 2 diabetes mellitus using the behavior change techniques taxonomy: scoping review. J Med Internet Res 2017 Oct 12;19(10):e348. [doi: 10.2196/jmir.7135] [Medline: 29025693]
24. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
25. Zhou M, Wang L, Deng Y, Ge J, Zhao S, You H. Effects of a mobile health intervention based on behavioral integrated model on cognitive and behavioral changes in gestational weight management: randomized controlled trial. J Med Internet Res 2025 Mar 10;27:e55844. [doi: 10.2196/55844] [Medline: 40063942]
26. Kytö M, Koivusalo S, Tuomonen H, et al. Supporting the management of gestational diabetes mellitus with comprehensive self-tracking: mixed methods study of wearable sensors. JMIR Diabetes 2023 Oct 31;8:e43979. [doi: 10.2196/43979] [Medline: 37906216]
27. Leblalta B, Kebaili H, Sim R, Lee SWH. Digital health interventions for gestational diabetes mellitus: a systematic review and meta-analysis of randomised controlled trials. PLOS Digit Health 2022 Feb;1(2):e0000015. [doi: 10.1371/journal.pdig.0000015] [Medline: 36812531]

XSL•FO
RenderX

28. Xu N, Han X, Chen S, Zhang J, Gu P. Self-reported barriers in self-management of women with gestational diabetes: a systematic review of qualitative studies. Nurs Open 2023 Nov;10(11):7130-7143. [doi: 10.1002/nop2.1988] [Medline: 37700604]
29. Riddle M, Umpierrez G, DiGenio A, Zhou R, Rosenstock J. Contributions of basal and postprandial hyperglycemia over a wide range of A1C levels before and after treatment intensification in type 2 diabetes. Diabetes Care 2011 Dec;34(12):2508-2514. [doi: 10.2337/dc11-0632] [Medline: 22028279]
30. Xu J, Ye S. Influence of low-glycemic index diet for gestational diabetes: a meta-analysis of randomized controlled trials. J Matern Fetal Neonatal Med 2020 Feb;33(4):687-692. [doi: 10.1080/14767058.2018.1497595] [Medline: 30231782]
31. Li LJ, Aris IM, Han WM, Tan KH. A promising food-coaching intervention program to achieve optimal gestational weight gain in overweight and obese pregnant women: pilot randomized controlled trial of a smartphone app. JMIR Form Res 2019 Oct 24;3(4):e13013. [doi: 10.2196/13013] [Medline: 31651407]
32. Sandborg J, Söderström E, Henriksson P, et al. Effectiveness of a smartphone app to promote healthy weight gain, diet, and physical activity during pregnancy (HealthyMoms): randomized controlled trial. JMIR Mhealth Uhealth 2021 Mar 11;9(3):e26091. [doi: 10.2196/26091] [Medline: 33704075]

Abbreviations

2hPG: 2-hour postprandial glucose
AFI: amniotic fluid index
BCT: behavior change technique
BCW: behavior change wheel
COM-B: capability, opportunity, motivation, and behavior
CONSORT: Consolidated Standards of Reporting Trials
FPG: fasting blood glucose
GDM: gestational diabetes mellitus
GI: glycemic index
GLMM: generalized linear mixed model
GWG: gestational weight gain
HbA_{1c}: hemoglobin A_{1c}
ITT: intention-to-treat
mHealth: mobile health
MNT: medical nutrition therapy
OGTT: oral glucose tolerance test
pBMI: prepregnancy BMI
RCT: randomized controlled trial
WeMNT: WeChat-delivered medical nutrition therapy

Edited by L Buis; submitted 11.Oct.2024; peer-reviewed by LT Markussen, WQ Chen; revised version received 15.Jan.2026; accepted 22.Apr.2026; published 11.Jun.2026.

Please cite as:
Wang X, Bian M, Li F, Zhou C, Mu L
WeChat-Delivered Mobile Medical Nutrition Therapy Intervention in Gestational Diabetes Mellitus: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e67410
URL: https://mhealth.jmir.org/2026/1/e67410
doi:10.2196/67410

© XiaoRong Wang, MiaoMiao Bian, Fang Li, Chang Zhou, LiNing Mu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Pilot mHealth Text Messaging Program Targeting Parents During the First 2000 Days: Nonrandomized Repeat Cross-Sectional Analysis to Evaluate Feasibility, Engagement, Acceptability, and Potential Effectiveness

Jacklyn Jackson[1,2,3,4], PhD; Nayerra Hudson[1,2,3], BNAD; Tessa Delaney[1,2,3], PhD; Alison L Brown[1,2,3,4], PhD; Sienna Kavalec[1,4], BNAD; Jessica Pinfold[3], BNAD; Rebecca Sewter[3], BNAD; Hannah McCormick[3], BNAD; Rebecca Liackman[1,2,3], MD; Christophe Lecathelinais[1,2,3], DESS; Lynda Davies[2,3], MNAD, MPH; Luke Wolfenden[1,2,3,4], PhD; Daniel Groombridge[3], MPH; Belinda Tully[3,5], MPH; Sinead Redman[3], BNurs; Tauri Smart[3], BNurs; Nafiseh Ghafournia[2,6,7], PhD; Paul David Craven[3], MD; Rachel Sutherland[1,2,3,4], PhD

[1]School of Medicine and Public Health, The University of Newcastle, University Drive, Callaghan, Australia

[2]Hunter Medical Research Institute, Newcastle, Australia

[3]Hunter New England Local Health District, New Lambton, Australia

[4]National Centre of Implementation Science, The University of Newcastle, Callaghan, Australia

[5]Armajun Health Service Aboriginal Corporation, Inverell, Australia

[6]School of Humanities, Creative Industries and Social Sciences, The University of Newcastle, Callaghan, Australia

[7]Western Sydney Local Health District, Paramatta, Australia

Corresponding Author:
Rachel Sutherland, PhD
School of Medicine and Public Health, The University of Newcastle, University Drive, Callaghan, Australia

Abstract

Background: The first 2000 days can profoundly influence long-term health. Healthy Beginnings for Hunter New England Kids (HB4HNEKids) is an SMS text messaging program delivered alongside routine Child and Family Health Nursing (CFHN) care, which provides families with evidence-based, age- and stage-related preventive health information across the first 2000 days.

Objective: This pilot study aimed to explore the feasibility, engagement, and acceptability of the HB4HNEKids program. It also aimed to explore the potential effectiveness of the program at 6 and/or 12 months post partum on outcomes including breastfeeding, child diet, child movement, and parental mental well-being.

Methods: During the pilot phase (October 2021 to July 2024), project records were used to assess the number of families enrolled, number of SMS text messages sent (feasibility), and the number of opt outs (engagement). Repeat cross-sectional surveys were conducted at 5 - 7 months post partum and again at 12 - 14 months post partum using validated survey instruments. Using convenience sampling methods, survey participants consisted of birthing parents who had received HB4HNEKids and a concurrent nonrandomized comparison group that did not receive the program. Surveys assessed parental self-reported engagement with the messages, program acceptability, breastfeeding status, child diet, child movement, and parental mental well-being. Mixed linear regression analyses were conducted to calculate mean differences and odds ratios.

Results: During the pilot phase, HB4HNEKids was delivered to 6243 families (73.4% of families contacted by CFHN). A total of 383 birthing parents completed the survey at 6 months (99/383, 26% receiving HB4HNEKids), and 283 completed the survey at 12 months (104/283, 37% receiving HB4HNEKids). Of the survey participants who received HB4HNEKids (n=200), between 76% and 83% reported that they always or very often read the SMS text messages, spending on average 5 - 7 minutes engaged with the content. At both survey time points, more than 90% of participants receiving HB4HNEKids agreed that the program was acceptable. Child daily intake of vegetables was significantly higher in the HB4HNEKids group (adjusted mean difference 0.23, 95% CI 0.07-0.40; *P*=.006) than in the comparison group at 12 months. Parents receiving HB4HNEKids also reported significantly better mental well-being scores (*P*=.005). While HB4HNEKids participants reported breastfeeding rates 5 percentage points greater than comparison participants at 6 and 12 months, this result was not statistically significant. There were no statistically significant differences between HB4HNEKids, and comparison participant responses related to child movement behaviors.

Conclusions: The HB4HNEKids SMS text messaging program is feasible to deliver at scale alongside routine CFHN care and is highly acceptable and engaging to parents. This pragmatic evaluation of the pilot, embedded into usual care, indicates potential

XSL•FO
RenderX

effectiveness of the program for improving child vegetable intakes and parental mental well-being. Further evaluation of this program using robust methodology is needed to determine the effectiveness of this innovative mHealth program across the first 2000 days.

(JMIR Mhealth Uhealth 2026;14:e83162) doi:10.2196/83162

KEYWORDS

digital health; mHealth; text message; first 2000 days; child health; feasibility; acceptability; cross-sectional

Introduction

The first 2000 days of life represents a critical period between conception through to 5 years that significantly impacts on child's physical, emotional, social, and cognitive development, laying the foundations for a child's lifelong trajectories [1]. Breastfeeding practices, introduction to solid foods, development of movement behaviors, and optimal parental mental well-being can significantly influence child's health and developmental outcomes [2-4]. The mental well-being of parents can influence the quality of care given to infants and is associated with a variety of health and developmental outcomes [5-7]. While longer breastfeeding duration, child's consumption of fruits and vegetables, and increased physical activity are associated with enhanced cognitive development and reduced lifelong risk of chronic disease [2-4,8], current data suggest that these outcomes are below national and international targets [9-11].

In alignment with key international, national, and state-level health policies, Child and Family Health Nursing (CFHN) services in New South Wales (NSW), Australia, provide care and support to families following the birth of infants, with the goal of enhancing child and family health, development, and well-being during the first 2000 days [12,13]. The services offered by CFHN are primarily delivered face-to-face and align to scheduled health and developmental checks, providing lactation education, emotional support, vaccinations, preventive health advice, and screening and secondary referral pathways. However, due to various health service barriers (eg, staff capacity and geographic location), care focuses on supporting families during the first few months of the child's life, with ongoing targeted services prioritized for those identified with additional needs. Furthermore, accessing face-to-face preventative health services can be challenging for parents, given caregiving responsibilities and work commitments [14,15]. Therefore, population-wide engagement with CFHN care significantly declines over the first 2000-day period, resulting in fragmented service utilization (NSW Ministry of Health, Health System Performance Report, internal data, 2025). This highlights the need for inclusive and accessible preventive health services to ensure that all families receive timely, trusted, and evidence-based support throughout the course of the first 2000 days.

Mobile health (mHealth) interventions represent a promising strategy to overcome some of the barriers to the traditional face-to-face support offered during the first 2000 days. Given that access to mobile phones and smartphones is relatively ubiquitous [16,17], mHealth interventions, including targeted SMS text messages delivered directly to parent or carer mobile phones, could improve the delivery of preventative care and health promotion during the first 2000 days [18]. This concept is supported by formative research conducted within the Hunter New England (HNE) region of NSW, Australia, which found that parents are highly accepting of receiving CFHN support via a variety of technology modalities (including websites, email, and SMS text messaging) on a variety of child– and family health–promoting topics (including healthy eating, introduction to solids, breastfeeding, sleep, and healthy growth) [19]. Systematic reviews also demonstrate that mHealth is effective in increasing multiple health behaviors including frequency of maternal contact with antenatal care, child immunization rates, and improving maternal and child health outcomes such as anxiety and depression, obesity prevention, and exclusive breastfeeding rates [20-25]. Additionally, a randomized controlled trial (RCT) conducted under robust research conditions demonstrates the efficacy of 1-way parent-targeted SMS text messaging interventions for improving infant feeding practices (ie, bottle use at bedtime) and reducing screen time [26]. This highlights the potential for 1-way SMS text message–based interventions during the postpartum period of the first 2000 days.

Despite the potential of mHealth interventions, few effective interventions have been embedded into health service delivery and implemented in real-world settings during the first 2000 days [27-29]. For mHealth interventions to effectively improve population health, they must be integrated into the settings, services, or health systems that reach the intended target population [29]. Given the limited translation of effective mHealth programs into usual health service delivery and building on the robust research evidence [18], there is a need to evaluate the feasibility, engagement, acceptability, and potential effectiveness of research innovations when embedded into routine care. The aim of this study was to pragmatically evaluate the feasibility, engagement, and acceptability of the pilot Healthy Beginnings for Hunter New England Kids (HB4HNEKids) program when embedded into routine CFHN service delivery. This pilot also explores the potential effectiveness of HB4HNEKids for improving child and parental health and well-being outcomes at 6 and/or 12 months post partum, including (1) breastfeeding, (2) child nutrition, (3) child movement, and (4) parental mental well-being.

Methods

Study Design

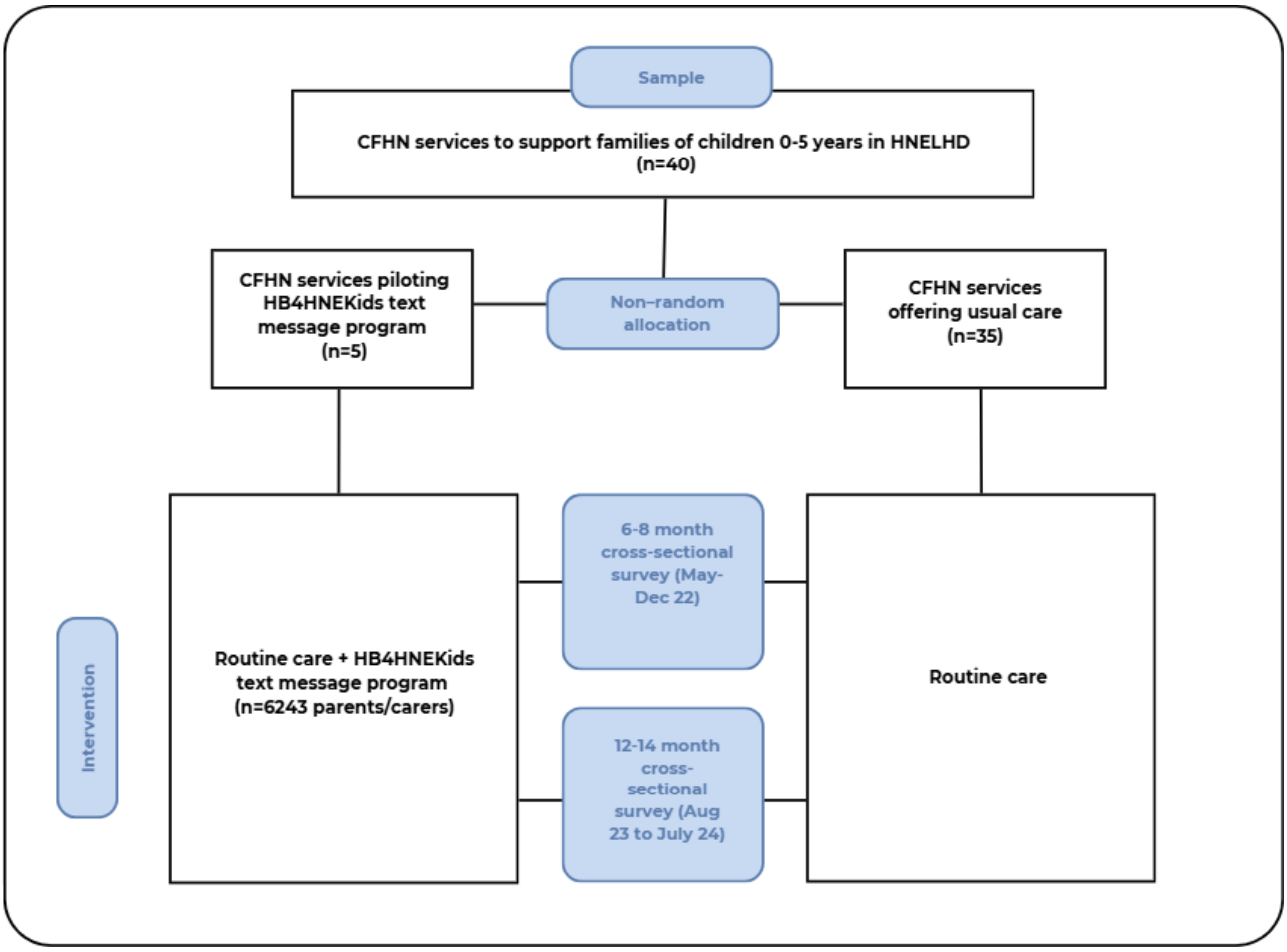
The feasibility, engagement, acceptability, and potential effectiveness of an SMS text messaging model of care delivered to families during the first 2000 days were evaluated as part of an observational study with a nonrandomized control arm. This study involved 2 cross-sectional surveys conducted with a

XSL•FO
RenderX

convenience sample of birthing parents located in the Hunter New England Local Health District (HNELHD) of NSW, Australia. As the 1-way SMS text messaging program (HB4HNEKids) was initially launched in October 2021 to parents attending 5 pilot CFHN services across 3 diverse local government areas (Armidale, Cessnock, and Newcastle/Lake Macquarie), the cross-sectional surveys include a subsample of participants who had received HB4HNEKids and a concurrent non–randomized comparison group of parents who did not receive HB4HNEKids. The first cross-sectional survey was undertaken between May and December 2022 and included birthing parents 5 - 7 months post partum. The second cross-sectional survey was conducted between August 2023 and July 2024 and included birthing parents 12 - 14 months post partum (Figure 1).

This repeat cross-sectional study is reported in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for cross-sectional studies (Checklist 1).

Figure 1. Flow of participants through the study. CFHN: Child and Family Health Nursing; HB4HNE: Healthy Beginnings for Hunter New England.



Ethical Considerations

Ethics approval for all maternal data was obtained from the Hunter New England Human Research Ethics Committee (16/11/16/4.07), Aboriginal Health and Medical Research Council (1236/16), and The University of Newcastle Human Research Ethics Committee (H-2017 - 0032). This research was conducted in compliance with informed consent guidelines, and study participant informed consent was sought prior to completion of surveys and implied upon the completion of a survey. Participant data were collected and securely stored using the REDCap (Research Electronic Data Capture), compliant with health service privacy and confidentiality policies. Data were anonymized, deidentified, and stored so that only authorized individuals could access the data. This research did not offer participants compensation for participating in the SMS text messaging program (HB4HNEKids) or the surveys.

Data Collection Procedures

Overview

Both the 6-month and 12-month surveys were developed using the REDCap (version 14.1.2; © 2025 Vanderbilt University) tool, hosted by HNELHD, which is a secure research electronic database commonly used in health service research [30,31]. Participants completed the survey either online or via computer-assisted telephone interview (CATI). Survey consent and responses were also stored in REDCap, providing a central survey database.

Eligibility for Participation in the Surveys

To evaluate the effectiveness of the pilot HB4HNEKids SMS text messaging program, a cross section of birthing parents who had accessed HNELHD maternity services was invited to participate in a survey at 2 time points. For the first survey, eligibility criteria for birthing parents were 26 - 37 weeks (ie,

XSL•FO
RenderX

approximately 6 months) post partum; those aged 18 years or older; had received antenatal care from any public antenatal service in HNE (providing antenatal care to approximately 70% of women across the district); had not had an unfortunate pregnancy-related outcome (ie, stillbirth); able to communicate in English; had previously provided consent to participate in a survey of antenatal care; and agreed to be contacted for future surveys. Eligibility for the second survey was the same as the first survey, except that birthing parents needed to be approximately 12 months post partum and have reached >34 weeks of gestation to be eligible to complete the survey.

Recruitment of Participants to Participate in the Survey

As per ethics approval, electronic medical records and previous antenatal survey data were used to identify the sample. Using this method, birthing parents who met eligibility criteria for survey participation were mailed printed information statements, providing an outline of the survey and a toll-free number to withdraw or decline participation.

Recruitment of Non-Aboriginal Participants to the Survey

Participants were sent information statements if their child met the eligibility criteria for participation. One week after information statements were mailed, birthing parents were contacted by telephone and invited to participate in the survey via a CATI. Birthing parents received up to 10 phone call attempts over a 2-week period to invite survey participation. As per ethics approval, verbal consent to participate in the surveys was sought from the birthing parent at the time of the interview. Birthing parents who declined to participate during the CATI were offered the opportunity to complete the survey online and were sent an individual survey link to their mobile number or email address. Before accessing the online survey, participants were reminded in the survey's display screen that participation was voluntary and that it was possible to decline the survey at any point. Participant consent into the study and survey completion status (both via CATI and online) were saved into a central survey database, in REDCap, held by the research team.

Recruitment of Aboriginal Participants to the Survey

Following advice provided from local cultural consultation, birthing parents identifying as Aboriginal and/or Torres Strait Islander were sent an SMS text message 3 days after the mail out of the information statement. The SMS text message offered one of three options (1) to complete the survey via CATI, (2) to complete the survey online, or (3) decline participation. Participants could respond to the SMS text message to indicate their preferred method of contact. Those who opted to complete the survey online were sent an individual survey link to their mobile number which was active for 2 weeks. Those who opted to complete the survey via telephone or did not reply to the SMS text message within 5 days received a phone call from a female interviewer and were invited to participate in the study.

The Text Messaging Program—Healthy Beginnings for Hunter New England Kids

Overview

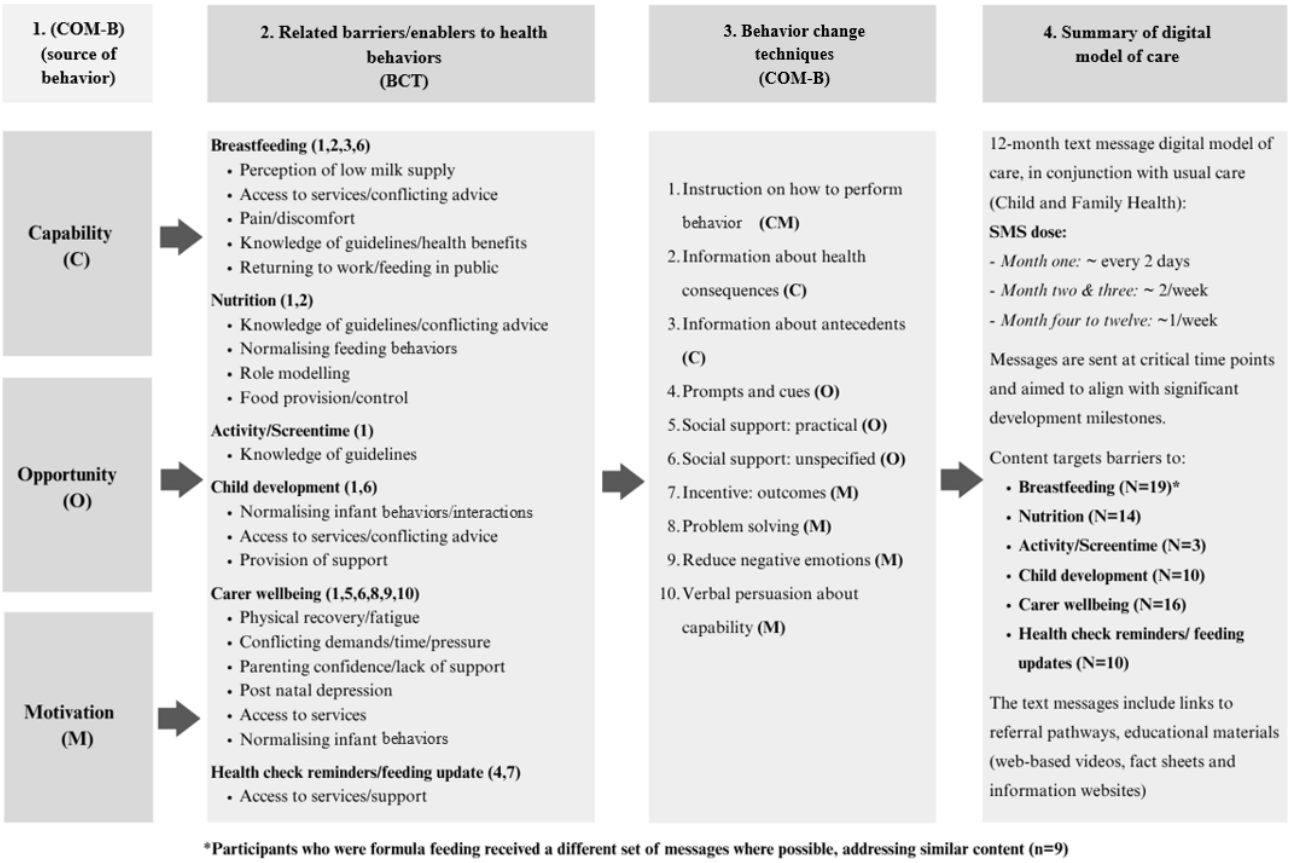
HB4HNEKids is an mHealth SMS text messaging program providing parents or carers with age- and stage-relevant preventative health information. The program is an adapted version of an RCT that demonstrated impact on reducing screen time and bottle use at bedtime at 12 months [26]. The HB4HNEKids program was adapted to strengthen a range of behavioral outcomes, supporting parents and children to develop preventive health behaviors during the first 2000 days. Implementation focused on overcoming health system and parent barriers to accessing health services and was therefore designed to be delivered alongside routine CFHN care. HB4HNEKids aims to offer timely, age- and stage-appropriate health promotion information directly to parents' and carers' mobile phones including preventive health information such as breastfeeding, infant feeding, movement behaviors, and parental well-being. The program also offers reminders for scheduled child health checks and immunizations, early developmental screening, and information on available support services.

Program Adaptation Process

HB4HNEKids was adapted from a prior RCT [26] to strengthen the effectiveness of SMS text message support offered across the first 2000 days and enhance alignment to usual CFHN care delivery. The HB4HNEKids program focuses on 6 key health behaviors, aligned with specific age and stage developmental milestones and key policy frameworks targeting the first 2000 days [13,32,33]. The process of adaptation involved (1) identifying priority health behaviors from policy frameworks, strategic plans, and expert consultation; (2) review of the literature to identify determinants of priority health behaviors; (3) use of behavioral theory (eg, Behavior Change Wheel [34] and Theoretical Domains Framework [35]) to develop SMS text messages embedding behavior change techniques to address the identified determinants; and (4) co-design of SMS text messages collaboratively with HNELHD health practitioners including CFHN, allied health professionals, public health practitioners, Aboriginal health workers, and representatives from multicultural health to ensure that wording and intent aligned with usual care (Figure 2).

XSL•FO
RenderX

Figure 2. Healthy Beginnings for Hunter New England Kids logic model.



All eligible families were offered HB4HNEKids either prior to or during the face-to-face routine 1 - to 4-week health check provided by CFHN services, as a digital program offered alongside usual care. The HB4HNEKids SMS text messages were short (often less than 160 characters) and included clickable links to evidenced-based "online" videos, fact sheets, websites, and relevant support services. Parents receiving the HB4HNEKids program were sent up to 44 SMS text messages in the first 6 months and a further 28 between 6 and 12 months to their nominated mobile number. Messages were sent 7 days a week, generally between the hours of 9 AM and 12 PM. While most messages were designed to be a 1-way SMS text message, the exception was the 2-way SMS text message sent to participants to confirm infant feeding status (on days 21, 97, and 150). Responses to feeding status were received in the messaging platform. Feeding status responses were stored in REDCap and monitored by a member of the research team each weekday. By collecting feeding status, the program content was tailored to better meet the individual needs of the participant (eg, families who were exclusively formula feeding did not receive content related to breastfeeding). Participants were sent a link on days 1, 51, and 296, which allowed participants to opt out and automatically unsubscribe from the program upon clicking. The opt out links remained valid or open allowing the participant to opt out of the program at any time.

Eligibility for Receipt of HB4HNEKids Program

All parents in HNELHD who had recently given birth are referred to CFHN services to receive ongoing care for their child until age 5 years. Parents due to receive a 1 - to 4-week health check from 1 of the 5 participating CFHN services located in the HNELHD between October 2021 and July 2024 were eligible for the HB4HNEKids SMS text message program alongside their usual CFHN care. Parents needed to have access to a mobile phone to be eligible to receive the program.

Recruitment of Women to the HB4HNEKids Program

As per the existing health service procedures, CFHN services received a referral from public and private maternity services in the HNELHD following the birth of a baby and discharge of the parent from hospital [36]. CFHN service staff then contacted birthing parents in their local catchment area to offer a 1 - to 4-week health check, which typically occurs within 3 business days of receipt of the CFHN referral. Staff from pilot CFHN services were encouraged to offer the HB4HNEKids program to all birthing parents within their catchment area. At the initial booking in contact, CFHN staff confirmed feeding status (breastfeeding, mixed feeding, and formula feeding), child's name, mobile number, and "clinician offer of service" (yes, client refused, and unable to contact after multiple attempts), and recorded this in a REDCap [31]. If CFHN service staff were unable to determine infant feeding status during the initial call (ie, due to language barriers), they would locate infant feeding status via the discharge summary provided by maternity services and enter this into REDCap.

Outcome Measures

Feasibility of Delivering HB4HNEKids via Routine CFH Care

During the relevant pilot time frame (October 2021 to July 2024), project record data (securely stored in REDCap) were

XSL•FO
RenderX

used to assess feasibility of the program as follows: (1) number of families enrolled in the HB4HNEKids program, (2) proportion of families enrolled in the HB4HNEKids program relative to the number of families contacted by CFHN services for the 1 - to 4-week appointment, (3) number of families who opted out of the HB4HNEKids program, and (4) number of messages sent as part of the HB4HNEKids program.

Engagement With HB4HNEKids Content

Within the 6-month and 12-month surveys, parents who had received HB4HNEKids were asked to report their engagement with the program SMS text message content. These items assessed how often participants read the SMS text messages and how often they clicked the website links within the SMS text messages, with response options including Always, Very often, Sometimes, Rarely, and Never. Participants were also asked to report average time spent (in minutes) engaged in the SMS text message content and links.

Acceptability of HB4HNEKids

To assess parent acceptability of the HB4HNEKids SMS text messaging program, measures of program acceptability were adapted from the psychometrically validated Acceptability of Intervention Measure by Weiner et al [37]. All survey items were assessed on a 5-point Likert scale (strongly disagree=1, disagree=2, neither agree nor disagree=3, agree=4, and strongly agree=5). Participants were coded as "agreeing" to statements, if they responded "strongly agree" or "agree" to any of the statements. Responses to Acceptability of Intervention Measure items were grouped according to (1) acceptability of the overall program, (2) acceptability of SMS text message timing, (3) acceptability of SMS text message content, and (4) acceptability of linked resources.

Potential Effectiveness

Breastfeeding Status, Introduction to Solids, and Breastfeeding Self-Efficacy at 6 and 12 Months

Infant feeding survey items were adapted from local, state, and national infant feeding surveys [38]. As per the Australian National Infant Feeding Survey, "current" breastfeeding status was reported if a parent reported "yes" to their "baby had received breastmilk since this time yesterday." Breastfeeding duration was assessed in accordance with the Centers for Disease Control and Prevention definitions of breastfeeding duration [39], as parents who had ceased breastfeeding were asked to report the age (in months) at which breastfeeding ceased. Participants were also asked to report age (in months) the infant was first introduced to soft, semisolid or solid food, using survey items from the Australian National Infant Feeding Survey [38].

The Breastfeeding Self-Efficacy Scale—Short Form tool was used to assess maternal breastfeeding self-efficacy [40], which has consistently been demonstrated as a valid and reliable measure of breastfeeding self-efficacy globally [41]. The Breastfeeding Self-Efficacy Scale—Short Form is a 14-item self-report instrument derived from the 33-item Breastfeeding Self-Efficacy Scale tool measuring breastfeeding confidence. All items are preceded by the phrase "I think I can always" and rated on a 5-point Likert scale, ranging from 1 (not at all confident) to 5 (always confident). As such, total score ranges from 14 to 70, with a higher score indicating greater levels of breastfeeding self-efficacy. The items were assessed within both the 6-month and 12-month surveys.

Child Nutrition at 12 Months

The Short Food Survey (SFS) developed by Hendrie et al [42] is a 38-item tool that captures parent-reported information on children's food group intakes and food choices and has been shown to produce a reliable estimate of compliance with dietary guidelines [42]. Four survey items from the SFS were used to assess child intake of fruit and vegetables at 12 months. Fruit intake was assessed by asking parents "How many serves of fruit does your child usually eat," including a prompt describing what a serve of fruit is (eg, 1 serve is equal to 2 small pieces of fruit, eg, apricots and plums; 1 medium apple, banana, or mandarin; or 1 cup of diced pieces, eg, diced peaches or diced fruit salad). To assess child intake of vegetables, responses to 3 survey items from the SFS were combined [42], including "How many servings of starchy vegetables (not including hot chips) does your child usually eat?" (prompt: starchy vegetables include potatoes and sweet potatoes; 1 serve is equal to ½ a cooked potato); "How many servings of salad vegetables does your child usually eat?" (prompt: salad vegetables include vegetables such as tomato, capsicum, cucumber, lettuce, and celery. A serve is equal to 1 cup of salad vegetables); and "How many servings of cooked vegetables does your child usually eat?" (prompt: a serve is equal to ½ cup of cooked vegetables).

The Children's Dietary Questionnaire [43] is a survey tool used to reliability assesses child intake patterns in either the previous week or 24 hours. We used survey items from the Children's Dietary Questionnaire to measure the mean weekly frequency of discretionary food consumption, captured by combining responses related to the frequency of child's intake of soft drinks (including cordial or sports drinks), takeaway foods (meals or snacks), chips (eg, baked potato gems/chips, hash browns, hot chops or French fries, wedges, or fried potatoes), savory snacks (eg, potato crisps, pretzels, or plain or flavored crackers), sweet snacks (eg, sweet biscuits, cakes, buns, muffins, and doughnuts), savory pastries, snack bars (eg, muesli bars or cereal bars), chocolate or lollies, and ice cream or ice-blocks. Items assessing child nutrition were included only within the 12-month survey.

Child Movement Behaviors at 12 Months

The Movement Behavior Questionnaire—Baby (MBQ-B) is a validated 6-item tool designed to be administered to parents of children younger than 18 months [44]. The MBQ-B assess frequency and duration of tummy time, time spent in active play or outdoor play, watching television, and using mobile digital devices. Items from the MBQ-B were used to assess child movement, restrained time, and screen time.

For active play, parents were asked "Thinking about the past week, on a typical day, how much time in total did you do some active play with your child? Active play could be crawling on the floor with your child, rolling around the floor with your child, playing at the park, dancing with your child, chasing your child." The response options ranged from 0 minutes per day to more than 2 hours per day. Given Australian 24-hour movement, guidelines recommend that toddlers (aged 1 - 2 years) should

XSL•FO
RenderX

achieve at least 3 hours of various physical activities each day, including energetic play, and parents reporting "more than 2 hours per day" of active play were coded as having children meeting physical activity guidelines [45]. To explore restrained time, parents were asked to report "On a typical day, how many times did you place your child in a baby carrier or sling, car seat or capsule, stroller or pram, highchair, bouncer, jolly jumper or play pen?," and "When thinking about each occasion that your child was in one of those devices, how long were they usually in it?." Given that the infant movement guidelines recommend not restraining an infant for more than 1 hour at a time [45], response options of "less than 15 minutes per day," "between 15 and 30 minutes per day," "between 30-35 minutes per day," and "between 45 to 60 minutes per day" were combined to indicate the proportion of parents compliant with this recommendation. Screen time behaviors were explored by asking participants 2 surveys items: "Thinking about the past week, on a typical day, how much time did your child spend watching television programs, videos/internet clips or movies on a television, computer or portable/mobile device such as iPad, tablet or smartphone?" (ie, passive screen time) and "Thinking about the past week, on a typical day, how much time did your child spend playing games, looking at photos, or video chatting (eg, FaceTime, Zoom, Skype) on a screen-based device such as a computer or laptop, video games console, iPad, tablet, or smartphone?" (ie, interactive screen time). Response options ranged between "0 minutes per day" and "more than 2 hours per day." Given Australian 24-hour movement guidelines recommend no screen time for children younger than 2 years [45], if a parent responded "0 minutes per day" to screen time items, they were coded as "not offering any screen time per day" indicating the proportion of children meeting screen time guidelines. These items were assessed only within the 12-month survey.

Parental Mental Well-Being at 12 Months

The Short Warwick Edinburgh Mental Wellbeing Scale (SWEMWBS) was used to capture parental mental well-being [46]. The SWEMWBS was funded by the Scottish Executive National Program for improving mental health and well-being, commissioned by NHS Health Scotland, developed by the University of Warwick and the University of Edinburgh, and is jointly owned by NHS Health Scotland, the University of Warwick, and the University of Edinburgh [47]. The SWEMWBS is a 7-item tool that has been positively worded to measure participant thoughts and feelings related to functioning over the past 2 weeks. The SWEMWBS is scored by summing the scores for each of the 7 items (scored 1 "none of the time" to 5 "all of the time") and then transformed into metric scores using the SWEMWBS conversion table [46]. Scores can range from 7 to 35, with higher scores indicating higher positive mental well-being, and has been shown to be a valid and reliable measure of mental well-being among caregivers [48]. The SWEMWBS was assessed within the 12-month survey only.

Participant Demographics and Pregnancy Characteristics

Within the survey, birthing parents were asked questions related to Aboriginal and/or Torres Strait Islander origin, country of birth, residential postcode, current employment status (full-time, part-time, casual, paid or unpaid maternity leave, unemployed, home duties, studies, retired, full-time carers, and unable to work due to health problems), timing of their return to work after birth (in months), and the child's date of birth. The survey items are based on those used in previous surveys with postpartum women [49]. Participant education status, first or subsequent pregnancy status, and maternal age data had been collected via previous surveys or medical record data. Participant residential postcode was used to determine geographical remoteness ("major cities" or "regional or remote") using the Accessibility/Remoteness Index of Australia [50]. Postcode was also used to determine socioeconomic areas ("most disadvantaged" or "least disadvantaged") using 2016 socioeconomic indexes for areas [51].

Data Analysis

Data were analyzed by an independent statistician using SAS statistical software (version 9.3, SAS Institute). Descriptive statistics were used to describe (1) characteristics of the samples, (2) feasibility outcomes, (3) program engagement outcomes, and (4) acceptability outcomes. Chi-square test or *t* test was used to compare the characteristics of participants between those who had received HB4HNEKids and those who had not (usual care control).

Sample size calculations were not conducted for this pilot feasibility study [52]. However, to explore the potential impact of the program on child breastfeeding, nutrition, and movement behaviors, as well as parental well-being outcomes (effectiveness), separate logistic regression analyses were conducted to examine differences between survey respondents who had received the HB4HNEKids program and those who had not (usual care control). Unadjusted and adjusted analyses were conducted at both time points. Analyses were adjusted for a range of key participant characteristics including parent age, Aboriginal and Torres Strait Islander status, highest level of education, socioeconomic status, and level of remoteness. Participants with missing data were excluded from analyses. Dichotomous outcomes were presented as odds ratios and 95% CIs, and continuous outcomes were presented as mean differences (MDs) and 95% CIs. A *P* value of <.05 was considered statistically significant.

Results

Feasibility of Delivering HB4HNEKids via Routine CFHN Care

During the 2.5-year pilot period (October 2021 to July 2024) of the HB4HNEKids program a total of 8501 families across the 5 CFHN services were eligible to receive the program. A total of 6243 families commenced the program (6243/8501, 73.4%), from which 430 families later withdrew, giving a 6.9% (430/6243) dropout rate. A total of 377,969 SMS text messages

XSL•FO
RenderX

were sent during the pilot period (average of 61 messages per family).

Characteristics of Cross-Sectional Survey Participants

Of the 780 birthing parents approached, a total of 383 (49% response rate) eligible participants completed the 6-month survey, whereby 26% (99) of survey participants had received the HB4HNEKids program. Of the 626 birthing parents approached, a total of 283 (45% response rate) participants completed the 12-month survey, from which 37% (104) survey participants had received the HB4HNEKids program. The 2 cross-sectional samples were similar in terms of birthing parent's age, education status, socioeconomic status, and remoteness of geographical location (Table 1). The 12-month sample included a higher proportion of participants identifying as Aboriginal and/or Torres Strait Islander (66/283, 23%), compared with the 6-month sample (43/383, 11%). By 12 months the proportion of participants on parental leave had significantly dropped, with 8% (22/283) of participants on parental leave, compared with 45% (173/383) at 6 months. Correspondingly, more participants engaged in employment at 12 months (188/283, 66% vs 114/383, 30%, respectively).

XSL•FO
RenderX

Table . Characteristics of birthing parents who participated in the 6- and 12-month survey by HB4HNEKids participation[a].

Characteristics	6-Month survey participants			12-Month survey participants		
	HB4HNEKids[b]	Usual care	Overall sample	HB4HNEKids	Usual care	Overall sample
Parents, n (%)	99 (26)	284 (74)	383 (100)	104 (37)	179 (63)	283 (100)
Age of parent (years), mean (SD)	30.6 (4.9)	30.6 (4.5)	30.6 (4.6)	31.6 (4.0)	30.9 (5.3)	31.2 (4.9)
Age of infant (months), n (%)						
5	2 (2)	5 (2)	7 (2)	—[c]	—[c]	—[c]
6	97 (98)	279 (98)	376 (98)	—[c]	—[c]	—[c]
12	—[c]	—[c]	—[c]	69 (66)	128 (71)	197 (70)
13	—[c]	—[c]	—[c]	30 (29)	41 (23)	71 (25)
14	—[c]	—[c]	—[c]	5 (5)	10 (6)	15 (5)
Aboriginal and/or Torres Strait Islander, n (%)						
Yes	8 (8)	35 (12)	43 (11)	10 (10)	56 (31)	66 (23)[d]
No	91 (92)	249 (88)	340 (89)	94 (90)	123 (69)	217 (77)[d]
Highest level of education, n (%)						
High school or less	21 (21)	79 (28)	100 (26)	18 (18)	57 (32)	75 (27)[d]
Tertiary	78 (79)	205 (72)	283 (74)	84 (82)	121 (68)	205 (73)[d]
Employment status, n (%)						
Employed	25 (25)	89 (31)	114 (30)	75 (72)	113 (63)	188 (66)
Maternity leave (paid or unpaid)	49 (50)	124 (44)	173 (45)	8 (8)	14 (8)	22 (8)
Unemployed	25 (25)	71 (25)	96 (25)	21 (20)	52 (29)	73 (26)
Socioeconomic status, n (%)						
Most disadvantaged	49 (49)	191 (68)	240 (63)[d]	42 (41)	119 (67)	161 (58)[d]
Least disadvantaged	50 (51)	90 (32)	140 (37)[d]	61 (59)	58 (33)	119 (42)[d]
Level of remoteness, n (%)						
Major city	78 (79)	156 (56)	234 (62)[d]	77 (75)	108 (61)	185 (66)[d]
Regional or remote	21 (21)	125 (44)	146 (38)[d]	26 (25)	69 (39)	95 (34)[d]

[a]Total 6-month sample (n=383): incomplete responses or missing data for socioeconomic status (n=380) and level of remoteness (n=380). Total 12-month sample (n=283): incomplete responses or missing data for highest level of education (n=280), socioeconomic status (n=280), and level of remoteness (n=280).

[b]HB4HNEKids: Healthy Beginnings for Hunter New England Kids.

[c]Not available.

[d]Statistically significant (P<.05) difference for the demographic characteristic between survey participants who had received HB4HNEKids and those who had not (usual care control group).

Participants who had received HB4HNEKids at the 6-month survey had a higher proportion from “least disadvantaged” areas and higher proportion residing in major cities than usual care control participants. Similar differences between groups were found at 12 months. Additionally, HB4HNEKids participants had a lower proportion of participants identifying as Aboriginal and/or Torres Strait Islander and a higher proportion of participants who had completed tertiary education.

XSL•FO
RenderX

Engagement With HB4HNEKids Text Message Content at 6 and 12 Months

Of the survey participants receiving the HB4HNEKids program (n=200), 76% (73/96) and 83% (86/104) of participants from the 6-month and 12-month survey, respectively, reported that they always or very often read the HB4HNEKids SMS text messages. Participants from the 6-month survey reported spending an average of 7.43 minutes engaged in the SMS text message content of HB4HNEKids, while participants from the 12-month survey reported an average of 5.83 minutes engaged in the content. In addition, 43% (41/96) of participants from the 6-month survey and 31% (32/104) of participants from the 12-month survey reported that they always or very often clicked on the website links provided in the messages (Table 2).

Table . Self-reported engagement with HB4HNEKids[a] SMS text message content and links.

	6-Month survey participants (n=96[b])	12-Month survey participants (n=104)
Parent always or very often		
Read the SMS text messages, n (%)	73 (76)	86 (83)
Clicked the website links within the SMS text messages, n (%)	41 (43)	32 (31)
Minutes on average spent engaged in the SMS text message content and links, mean (SD)	7.43 (7.91)	5.83 (5.52)

[a]HB4HNEKids: Healthy Beginnings for Hunter New England Kids.

[b]Response data missing for 3 survey participants at 6 months, as those who reported opting out of the HB4HNEKids program were not required to provide data on SMS text message engagement.

Acceptability of HB4HNEKids

At 6 months and 12 months, most parents reported that the program was acceptable (Table 3), with more than 90% of parents agreeing or strongly agreeing with statements related to the overall acceptability of the program, timing of the messages, and message content. More than 80% of parents found the embedded website links acceptable.

XSL•FO
RenderX

Table . Acceptability of the HB4HNEKids[a] SMS text messaging program.

Aim and item	6-Month survey (agree or strongly agree; n=96[b]), %	12-Month survey (agree or strongly agree; n=104), %
The overall program		
Overall, I liked the HB4HNEKids program.	93	95
Overall, the HB4HNEKids program meets my approval.	96	94
Overall, the HB4HNEKids program was appealing to me.	—[c]	90
Overall, I was satisfied with the HB4HNEKids program.	97	93
I would recommend the HB4HNEKids program to other parents/carers.	94	93
SMS text message timing		
I was happy with the timing of the delivery of the content of text messages based on [child] age and development.	97	99
SMS text message content		
I found the immunization and growth check reminder text messages acceptable.	95	91
I found the information and advice on breastfeeding/formula feeding acceptable.	95	90
I found the information and advice on introduction to solids acceptable.	95	90
Linked resources		
I like the website links.	84	86

[a]HB4HNEKids: Healthy Beginnings for Hunter New England Kids.

[b]Response data missing for 3 survey participants at 6 months, as those who reported opting out of the HB4HNEKids program were not required to provide data on program acceptability.

[c]Not available.

Potential Effectiveness

Breastfeeding Status, Breastfeeding Duration, Introduction to Solids, and Breastfeeding Self-Efficacy

The proportion of parents currently breastfeeding was not significantly different based on whether the parent had participated in the HB4HNEKids program or not at both time points (6 months and 12 months). However, at both time points, HB4HNEKids participants reported a 5% higher point prevalence of breastfeeding than the comparison group participants (6 months: 66% vs 61%; 12 months: 46% vs 41%). There was no difference in breastfeeding duration for HB4HNEKids participants compared with comparison participants. Similarly, there were no statistically significant differences between HB4HNEKids and comparison participants at both time points for mean age of solids introduction and breastfeeding self-efficacy scores (Tables 4 and 5).

XSL•FO
RenderX

Table . Breastfeeding and infant feeding outcomes measured at 6 months by HB4HNEKids[a] participation.

	HB4HNEKids	Usual care	Unadjusted estimate (95% CI), *P* value	Adjusted estimate[b] (95% CI), *P* value
Proportion currently breastfeeding, n (%)	61 (66)	166 (61)	OR[c] 1.23 (0.75 to 2.03), .41	OR 1.07 (0.62 to 1.83), .81
Breastfeeding duration in months, mean (SD)	3.80 (2.19)	3.86 (2.15)	MD[d] –0.06 (–0.57 to 0.46), .83	MD –0.17 (–0.69 to 0.35), .53
Age in months child was introduced to first solid foods, mean (SD)	5.36 (0.80)	5.32 (0.87)	MD 0.04 (–0.16 to 0.24), .70	MD 0.06 (–0.14 to 0.27), .55
Breastfeeding self-efficacy score, mean (SD)	57.38 (12.72)	55.96 (14.36)	MD 1.41 (–1.90 to 4.73), .40	MD 1.30 (–2.12 to 4.73), .45

[a]HB4HNEKids: Healthy Beginnings for Hunter New England Kids.

[b]Adjusted analysis was adjusted for participant socioeconomic status, level of remoteness, level of education, Aboriginal and Torres Strait Islander status, and age.

[c]OR: odds ratio.

[d]MD: mean difference.

Table . Breastfeeding and infant feeding outcomes measured at 12 months by HB4HNEKids[a] participation.

	HB4HNEKids	Usual care	Unadjusted estimate (95% CI), *P* value	Adjusted estimate[b] (95% CI), *P* value
Proportion currently breastfeeding, n (%)	46 (46)	68 (41)	OR[c] 1.23 (0.74 to 2.03), .43	OR 1.07 (0.62 to 1.87), .80
Breastfeeding duration in months, mean (SD)	3.07 (2.59)	3.24 (2.45)	MD[d] –0.18 (–0.80 to 0.45), .58	MD –0.37 (–1.03 to 0.29), .27
Age in months child was introduced to first solid foods, mean (SD)	5.59 (1.02)	5.52 (1.02)	MD 0.06 (–0.18 to 0.31), .61	MD 0.02 (–0.24 to 0.29), .87
Breastfeeding self-efficacy score, mean (SD)	55.48 (15.77)	55.89 (15.34)	MD –0.40 (–4.30 to 3.50), .84	MD –0.73 (–4.82 to 3.37), .73

[a]HB4HNEKids: Healthy Beginnings for Hunter New England Kids.

[b]Adjusted analysis was adjusted for participant socioeconomic status, level of remoteness, level of education, Aboriginal and Torres Strait Islander status and age.

[c]OR: odds ratio.

[d]MD: mean difference.

Child Nutrition at 12 Months

Child daily intake of vegetable serves was significantly higher in the HB4HNEKids group in both the unadjusted (MD 0.21, 95% CI 0.05-0.36; *P*=.009) and adjusted analyses (MD 0.23, 95% CI 0.07-0.40; *P*=.006) than in the comparison group. Child daily intake of fruit serves was significantly higher in the HB4HNEKids group than in the comparison group participants in the unadjusted analysis (MD 0.28, 95% CI 0.05-0.50; *P*=.02); however, this result did not remain statistically significant in the adjusted analyses (MD 0.20, 95% CI –0.04 to 0.45; *P*=.10). There was no difference in the frequency of child discretionary food intake per week between the groups (Table 6).

XSL•FO
RenderX

Table . Child nutrition, child movement, and parental mental well-being outcomes measured at 12 months by HB4HNEKids[a] participation.

	HB4HNEKids	Usual care	Unadjusted estimate (95% CI), *P* value	Adjusted estimate[b] (95% CI), *P* value
Daily intake of fruit serves, mean (SD)	2.07 (1.01)	1.79 (0.88)	MD[c] 0.28 (0.05 to 0.50), .02	MD 0.20 (–0.04 to 0.45), .10
Daily intake of vegetable serves, mean (SD)	0.74 (0.70)	0.53 (0.59)	MD 0.21 (0.05 to 0.36), .01	MD 0.23 (0.07 to 0.40), .006
Frequency of discretionary food intake per week, mean (SD)	0.99 (0.93)	1.17 (1.03)	MD –0.18 (–0.42 to 0.06), .15	MD –0.02 (–0.27 to 0.23), .88
Proportion meeting physical activity guidelines, n (%)	88 (85)	139 (78)	OR[d] 1.69 (0.88 to 3.25), .12	OR 1.71 (0.84 to 3.48), .14
Proportion meeting re-strained time guidelines, n (%)	70 (72)	131 (82)	OR 0.57 (0.31 to 1.05), .07	OR 0.62 (0.33 to 1.18), .15
Proportion meeting screen time guidelines, n (%)	24 (24)	48 (27)	OR 0.86 (0.49 to 1.51), .61	OR 0.63 (0.34 to 1.16), .14
Parental mental well-being scores, mean (SD)	26.15 (3.37)	24.22 (4.43)	MD 1.93 (0.94 to 2.92), <.001	MD 1.52 (0.47 to 2.58), .005

[a]HB4HNEKids: Healthy Beginnings for Hunter New England Kids.

[b]Adjusted analysis was adjusted for participant socioeconomic status, level of remoteness, level of education, Aboriginal and Torres Strait Islander status, and age.

[c]MD: mean difference.

[d]OR: odds ratio.

Child Movement at 12 Months

Table 6 includes unadjusted and adjusted analyses exploring the odds of participants meeting physical activity guidelines, restrained time guidelines, and screen time guidelines. There were no statistically significant differences between the proportion of HB4HNEKids participant and comparison participant responses related to meeting child movement guidelines (Table 6).

Parental Mental Well-Being at 12 Months

At 12 months, parents who had received the HB4HNEKids program reported a mean SWEMWBS score of 26.15, and comparison participants reported a mean SWEMWBS score of 24.2. Unadjusted and adjusted analyses indicate that parents who had received the HB4HNEKids program reported statistically significantly better mental well-being scores (unadjusted MD 1.93, 95% CI 0.94- 2.92, *P*<.001; adjusted MD 1.52, 95% CI 0.47-2.58, *P*=.005) than the comparison participants (Table 6).

Discussion

Principal Results

This study sought to evaluate the feasibility, engagement, and acceptability of the pilot HB4HNEKids program and explore its potential effectiveness on a range of child and parental health and well-being outcomes including breastfeeding, child nutrition, child movement, and parental mental well-being. Findings indicated that the program was feasible to deliver to families in HNELHD via CFHN services and is met with high acceptability and strong engagement from parents. Furthermore, the evaluation found that the program was likely effective in improving child vegetable intakes and parental mental well-being at 12-months. These findings provide support for further development and improvement of the program with formal testing as part of a larger RCT.

Comparison With Prior Work

This pragmatic evaluation demonstrates that the 1 - to 4-week CFHN service visit (attended by >80% of families in HNELHD) is a feasible way of enrolling new families into a mHealth program at scale. There remains scope to further optimize the reach of HB4HNEKids, given just 73.4% (6243/8501) of eligible families coming through the pilot CFHN services were enrolled into the program. However, the initial reach of the HB4HNEKids program appears relatively high in contrast with other digital health interventions (DHIs), with systematic reviews of DHIs reporting a median reach of just 21.8% - 33.6% [53]. Engagement is another critical factor in determining the success of DHIs, with systematic review evidence suggesting that engagement is associated with a variety of improved health behaviors including fruit and vegetable intake, weight status, smoking, and physical activity [54-56]. Our results indicate good participant engagement with the HB4HNEKids program, with most HB4HNEKids participants reporting that they "always" or "very often" read the SMS text messages, spending on average of 6 minutes interacting with each message. While a systematic review by Grady et al [57] found that SMS text messages are a key DHI design feature associated with engagement, the extensive co-design process undertaken in the design of the HB4HNEKids program may also be driving strong engagement rates [58]. This combined with low opt-out rates and high parent-reported acceptability, the HB4HNEKids

XSL•FO
RenderX

program holds significant promise for reaching and engaging parents in health promotion content during the first 2000 days.

Vegetable intake is a key indicator of a healthy diet providing an important source of essential vitamins and minerals [59]; however, only 18.5% of Australian children aged 2 - 3 years meet the recommended intake for vegetables, dropping to just 3.8% in children aged 4 - 8 years [60]. Although data collected from parents within the HNE region of NSW, Australia, indicate that children aged 12 months are not meeting the recommended serves of vegetables each day (2 - 3 serves), the evaluation indicates that children from families who were receiving HB4HNEKids were having 0.23 of a serve more vegetables per day than those from families in the usual care group [61] and represents an effect size consistent with other interventions that have demonstrated efficacy at increasing children's vegetable intake in home and community settings [62]. Although this effect size is modest in absolute terms, this increase aligns with the magnitude of change commonly observed in early life dietary interventions and reflects a behavioral shift that is difficult to achieve during the introduction-to-solids period [63,64]. Small increments in vegetable intake at this developmental stage may nonetheless carry clinical relevance, as early exposure contributes to flavor acceptance, supports the establishment of healthier dietary trajectories, and enhances intake of key micronutrients such as folate, vitamin A precursors, and potassium [59]. Evidence indicates that even limited increases in vegetable consumption can reduce the risk of micronutrient inadequacy [65] and are prospectively associated with healthier dietary patterns linked to lower noncommunicable disease risk in adulthood [66]. Therefore, future research to assess the sustainment of the program's effects over time is warranted to better quantify the child- and population-level benefit of the HB4HNEKids program.

Like other trials such as the original Health Beginnings RCT [26], our evaluation did not indicate a statistically significant difference in current breastfeeding rates between the HB4HNEKids and usual care participants. These findings are inconsistent with those of a recent systematic review of SMS text message–based interventions, with meta-analyses showing that they were effective at improving exclusive breastfeeding rates particularly when they engage families in the prenatal period [25]. However, given our pilot study was not powered for this outcome and still observed a 5-point prevalence difference in current breastfeeding rates (favoring HB4HNEKids participants) at 6 months and 12 months, this result is nonetheless highly encouraging. Even so, further development and optimization of the breastfeeding content offered within the program, or expanding the program to engage families earlier (ie, antenatally), could be of value.

In contrast with the original Healthy Beginnings RCT that found that families receiving SMS text message support (combined with accompanying printed resources) were significantly more likely to "never have screen time" than the control (adjusted odds ratio 1.28, 95% CI 1.08 - 1.52; $P<.05$) [26], our evaluation did not indicate any differences between HB4HNEKids and usual care participants screen time behaviors or any other movement behaviors at 12 months. While the SMS text message content of the HB4HNEKids program related to child screen time and movement behavior during the first 12 months mirrored that of the original Healthy Beginnings SMS text message content and focused on increasing parental knowledge of child movement and screen time guidelines (eg, 1 message on movement guidelines, 1 message on restrained time guidelines, and 1 message on screen time recommendations), the omission of the booklet content from HB4HNEKids could be driving this difference. However, given that formative research suggests that parents would prefer to be engaged in content focused on the benefits of certain movement and screen time behaviors, rather than the risks [67], there remains opportunity to strengthen parent-directed SMS text message content to better address known barriers and facilitators to child screen time and movement behaviors during the first 12 months.

The perinatal period demands significant emotional, psychological, social, and physical adaptation from parents and represents a period of change that can have implications on parental well-being and mental health outcomes. Parental mental health and well-being are of key importance during this early life stage, as it can influence parent-child relationships, the quality of care provided, and in turn the child's life trajectory [68-70]. Outside of the standardized mental health screening conducted by primary care services, there is limited support embedded within health services to generally support parental well-being [71,72]. Thus, the improved mental well-being scores at 12 months among those receiving the program compared with those who did not (MD 1.51, 95% CI 0.45-2.57; $P=.005$) are important.

Limitations

The results of this evaluation must be considered within the context of our study limitations. These include the use of our convenience sampling methods, which included only a small subsample of participants who had received the HB4HNEKids program, thus not providing data that are representative of the whole HNE region. Furthermore, given our response rates of approximately 45% - 49%, our findings may be subject to nonresponse bias. Given the cross-sectional nature of our evaluation, and that HB4HNEKids and comparison participants were not randomly allocated, we cannot be certain that the differences in child vegetable intakes and parental mental well-being outcomes are explicitly due to the SMS text messaging program [73]. Furthermore, most of the outcomes assessed within this evaluation were self-reported by participants and may be influenced by self-report, social desirability, or recall biases. However, where possible, validated survey tools were used. This pilot study was not powered to detect statistically significant differences between groups for many of the outcomes assessed, including proportion of currently breastfeeding and proportion of children meeting physical activity guidelines. The HB4HNEKids program is currently operating as a complimentary mHealth program alongside an existing routinely offered health service. The insight offered through this evaluation of a real-world health service intervention (while acknowledging the limitations stated) would be strengthened with the opportunity to evaluate the program using more robust methodologies and a larger sample size.

XSL•FO
RenderX

Conclusions

This evaluation of the pilot HB4HNEKids SMS text messaging program indicates that the program is feasible to deliver at scale, alongside routine CFHN care and is highly acceptable and engaging to parents. Our findings indicate possible efficacy of the program for improving child vegetable intake and parental mental well-being at 12 months. However, further evaluation of the program within larger representative samples using robust methodology is needed to determine the true impact of this innovative program.

Acknowledgments

The authors would like to acknowledge the Child and Family Health Services who contributed to enrolling participants into the program. This project was supported by HNELHD IT Services. The contents of this publication are the authors' sole responsibility and do not necessarily represent official HNELHD IT. Generative artificial intelligence tools were not used in the generation of this research study or manuscript.

Funding

In-kind support was provided by Hunter New England Local Health District. This research was funded by a Glencore Community Investment grant (TPDD13). RS is supported by a Medical Research Future Fund Fellowship (APP1150661).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: JJ, NH, TD, R Sutherland
Data duration: JJ, JP, R Sewter, HMc, RL, LD
Formal analysis: JJ, CL
Funding acquisition: NH, TD, LW, R Sutherland
Investigation: JJ, NH, TD, ALB, SK, JP, R Sewter, HMc, RL, LD, R Sutherland
Methodology: JJ, NH, TD, ALB, LW, DG, BT, SR, TS, NG, PDC, R Sutherland
Project administration: JJ, NH, ALB, SK, JP, R Sewter, HMc, RL, LD, R Sutherland
Visualization: JJ, SK, CL
Writing – original draft: JJ, NH, TD, SK, R Sutherland
Writing – review & editing: all authors.

Conflicts of Interest

None declared.

Checklist 1
STROBE checklist for cross-sectional studies.
[PDF File, 211 KB - mhealth_v14i1e83162_app1.pdf]

References

1. Moore TG, Arefadib N, Deery A, West S. The First Thousand Days: An Evidence Paper: Centre for Community Child Health, Murdoch Children's Research Institute; 2017. URL: https://www.rch.org.au/uploadedFiles/Main/Content/ccchdev/CCCH-The-First-Thousand-Days-An-Evidence-Paper-September-2017.pdf [accessed 2022-03-04]
2. Zheng M, D'Souza NJ, Atkins L, et al. Breastfeeding and the longitudinal changes of body mass index in childhood and adulthood: a systematic review. Adv Nutr 2024 Jan;15(1):100152. [doi: 10.1016/j.advnut.2023.100152] [Medline: 37977327]
3. Zeng N, Ayyub M, Sun H, Wen X, Xiang P, Gao Z. Effects of physical activity on motor skills and cognitive development in early childhood: a systematic review. Biomed Res Int 2017;2017:2760716. [doi: 10.1155/2017/2760716] [Medline: 29387718]
4. Kaikkonen JE, Mikkilä V, Magnussen CG, Juonala M, Viikari JSA, Raitakari OT. Does childhood nutrition influence adult cardiovascular disease risk?--insights from the Young Finns Study. Ann Med 2013 Mar;45(2):120-128. [doi: 10.3109/07853890.2012.671537] [Medline: 22494087]
5. Smith M. Parental mental health: disruptions to parenting and outcomes for children. Child Fam Soc Work 2004 Feb;9(1):3-11. [doi: 10.1111/j.1365-2206.2004.00312.x]
6. Kahn RS, Brandt D, Whitaker RC. Combined effect of mothers' and fathers' mental health symptoms on children's behavioral and emotional well-being. Arch Pediatr Adolesc Med 2004 Aug;158(8):721-729. [doi: 10.1001/archpedi.158.8.721] [Medline: 15289242]

XSL•FO
RenderX

7. Kamis C. The long-term impact of parental mental health on children's distress trajectories in adulthood. Soc Ment Health 2021 Mar 1;11(1):54-68. [doi: 10.1177/2156869320912520] [Medline: 34094696]
8. Horta BL, Loret de Mola C, Victora CG. Breastfeeding and intelligence: a systematic review and meta-analysis. Acta Paediatr 2015 Dec;104(467):14-19. [doi: 10.1111/apa.13139] [Medline: 26211556]
9. Vaz JS, Maia MFS, Neves PAR, Santos TM, Vidaletti LP, Victora C. Monitoring breastfeeding indicators in high‐income countries: levels, trends and challenges. Matern Child Nutr 2021 Jul;17(3):e13137. [doi: 10.1111/mcn.13137]
10. Allen CK, Assaf S, Namaste S, Benedict RK. Estimates and trends of zero vegetable or fruit consumption among children aged 6-23 months in 64 countries. PLOS Glob Public Health 2023;3(6):e0001662. [doi: 10.1371/journal.pgph.0001662] [Medline: 37368873]
11. Whiting S, Buoncristiano M, Gelius P, et al. Physical Activity, Screen Time, and Sleep Duration of Children Aged 6-9 Years in 25 Countries: An Analysis within the WHO European Childhood Obesity Surveillance Initiative (COSI) 2015-2017. Obes Facts 2021;14(1):32-44. [doi: 10.1159/000511263] [Medline: 33352575]
12. World Health Organization. Improving Early Childhood Development: WHO Guideline: World Health Organization; 2020.
13. NSW Ministry of Health. The First 2000 Days Framework: NSW Ministry of Health; 2019.
14. Herlitz L, Ashford E, Powell C, Herbert K, Morris S, Woodman J. Access to primary care for children and young people (CYP) in the UK: a scoping review of CYP's, caregivers' and healthcare professionals' views and experiences of facilitators and barriers. BMJ Open 2024 May 30;14(5):e081620. [doi: 10.1136/bmjopen-2023-081620] [Medline: 38816045]
15. Salimi N, Javan-Noughabi J, Ghavami V, Matinfar R, Kokabisaghi F. Parents' perspective on children's access to healthcare services and associated factors: a cross-sectional study. BMC Public Health 2025 Jul 2;25(1):2199. [doi: 10.1186/s12889-025-23230-0] [Medline: 40604680]
16. Number of smartphone users worldwide from 2014 to 2020. Statista. URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide [accessed 2026-05-04]
17. Mobile consumer survey: the Australian cut. Deloitte. 2017. URL: https://www.scribd.com/document/466867337/tmt-mobile-consumer-survey-2017-pdf [accessed 2026-05-04]
18. Wen LM, Rissel C, Baur LA, et al. A 3-arm randomised controlled trial of Communicating Healthy Beginnings Advice by Telephone (CHAT) to mothers with infants to prevent childhood obesity. BMC Public Health 2017 Jan 14;17(1):79. [doi: 10.1186/s12889-016-4005-x] [Medline: 28088203]
19. Delaney T, Jackson JK, Brown AL, et al. Perceived acceptability of technology modalities for the provision of universal child and family health nursing support in the first 6-8 months after birth: cross-sectional study. JMIR Pediatr Parent 2024 Sep 24;7:e59191. [doi: 10.2196/59191] [Medline: 39316424]
20. Dol J, Richardson B, Tomblin Murphy G, Aston M, McMillan D, Campbell-Yeo M. Impact of mobile health (mHealth) interventions during the perinatal period for mothers in low- and middle-income countries: a systematic review. JBI Database System Rev Implement Rep 2019;17(8):1634-1667. [doi: 10.11124/JBISRIR-2017-004022]
21. Ameyaw EK, Amoah PA, Ezezika O. Effectiveness of mHealth apps for maternal health care delivery: systematic review of systematic reviews. J Med Internet Res 2024 May 29;26:e49510. [doi: 10.2196/49510] [Medline: 38810250]
22. Oliver-Williams C, Brown E, Devereux S, Fairhead C, Holeman I. Using mobile phones to improve vaccination uptake in 21 low- and middle-income countries: systematic review. JMIR Mhealth Uhealth 2017 Oct 4;5(10):e148. [doi: 10.2196/mhealth.7792] [Medline: 28978495]
23. Mekonnen ZA, Gelaye KA, Were MC, Gashu KD, Tilahun BC. Effect of mobile text message reminders on routine childhood vaccination: a systematic review and meta-analysis. Syst Rev 2019 Dec;8(1):154. [doi: 10.1186/s13643-019-1054-0]
24. Ekambareshwar M, Ekambareshwar S, Mihrshahi S, et al. Process evaluations of early childhood obesity prevention interventions delivered via telephone or text messages: a systematic review. Int J Behav Nutr Phys Act 2021 Dec;18(1):10. [doi: 10.1186/s12966-020-01074-8]
25. Fan Y, Li J, Wong JYH, Fong DYT, Wang KMP, Lok KYW. Text messaging interventions for breastfeeding outcomes: a systematic review and meta-analysis. Int J Nurs Stud 2024 Feb;150:104647. [doi: 10.1016/j.ijnurstu.2023.104647] [Medline: 38056353]
26. Wen LM, Rissel C, Xu H, et al. Effects of telephone and short message service support on infant feeding practices, "tummy time," and screen time at 6 and 12 months of child age: a 3-group randomized clinical trial. JAMA Pediatr 2020 Jul 1;174(7):657-664. [doi: 10.1001/jamapediatrics.2020.0215] [Medline: 32282034]
27. Knop MR, Nagashima-Hayashi M, Lin R, et al. Impact of mHealth interventions on maternal, newborn, and child health from conception to 24 months postpartum in low- and middle-income countries: a systematic review. BMC Med 2024 May 15;22(1):196. [doi: 10.1186/s12916-024-03417-9] [Medline: 38750486]
28. Godinho MA, Jonnagaddala J, Gudi N, Islam R, Narasimhan P, Liaw ST. mHealth for integrated people-centred health services in the Western Pacific: a systematic review. Int J Med Inform 2020 Oct;142:104259. [doi: 10.1016/j.ijmedinf.2020.104259] [Medline: 32858339]
29. Palmer MJ, Henschke N, Bergman H, et al. Targeted client communication via mobile devices for improving maternal, neonatal, and child health. Cochrane Database Syst Rev 2020 Jul 14;8(8):CD013679. [doi: 10.1002/14651858.CD013679] [Medline: 32813276]

XSL•FO
RenderX

30. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
31. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
32. NSW Ministry of Health. NSW Healthy Eating and Active Living Strategy 2022-2032: NSW Ministry of Health; 2022.
33. Australian Digital Health Agency. National Digital Health Strategy 2023-2028: Australian Digital Health Agency; 2023.
34. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42. [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
35. Atkins L, Francis J, Islam R, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implement Sci 2017 Jun 21;12(1):77. [doi: 10.1186/s13012-017-0605-9] [Medline: 28637486]
36. Hunter New England local health district child and family health nursing services model of care. : NSW Health; 2021 URL: https://www.nsw.gov.au/departments-and-agencies/hnelhd/services/child-and-family-health/child-family-nursing [accessed 2022-03-4]
37. Weiner BJ, Lewis CC, Stanick C, et al. Psychometric assessment of three newly developed implementation outcome measures. Implementation Sci 2017 Dec;12(1):108. [doi: 10.1186/s13012-017-0635-3]
38. Australian Institute of Health and Welfare. 2010 Australian National Infant Feeding Survey: Indicator Results: Australian Institute of Health and Welfare; 2010.
39. NIS-child data analysis. US Centers for Disease Control and Prevention. 2025 Jul 8. URL: https://www.cdc.gov/breastfeeding-data/survey/analysis.html [accessed 2026-05-04]
40. Dennis CL. The Breastfeeding Self‐Efficacy Scale: psychometric assessment of the short form. J Obstet Gynecol Neonatal Nurs 2003;32(6):734-744. [doi: 10.1177/0884217503258459] [Medline: 14649593]
41. Dennis CL, McQueen K, Dol J, Brown H, Beck C, Shorey S. Psychometrics of the Breastfeeding Self-Efficacy Scale and Short Form: a systematic review. BMC Public Health 2024 Feb 29;24(1):637. [doi: 10.1186/s12889-024-17805-6] [Medline: 38419045]
42. Hendrie GA, Viner Smith E, Golley RK. The reliability and relative validity of a diet index score for 4-11-year-old children derived from a parent-reported short food survey. Public Health Nutr 2014 Jul;17(7):1486-1497. [doi: 10.1017/S1368980013001778] [Medline: 23830074]
43. Magarey A, Golley R, Spurrier N, Goodwin E, Ong F. Reliability and validity of the Children's Dietary Questionnaire; a new tool to measure children's dietary patterns. Int J Pediatr Obes 2009;4(4):257-265. [doi: 10.3109/17477160902846161] [Medline: 19922040]
44. Trost SG, Terranova CO, Brookes DSK, Chai LK, Byrne RA. Reliability and validity of rapid assessment tools for measuring 24-hour movement behaviours in children aged 0-5 years: the Movement Behaviour Questionnaire Baby (MBQ-B) and child (MBQ-C). Int J Behav Nutr Phys Act 2024 Apr 23;21(1):43. [doi: 10.1186/s12966-024-01596-5] [Medline: 38654342]
45. 24-hour movement guidelines—birth to 5 years—brochure. : Australian Government Department of Health, Disability and Ageing; 2017 URL: https://www.health.gov.au/sites/default/files/documents/2021/05/24-hour-movement-guidelines-birth-to-5-years-brochure.pdf [accessed 2025-05-18]
46. Stewart-Brown S, Tennant A, Tennant R, Platt S, Parkinson J, Weich S. Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using data from the Scottish Health Education Population Survey. Health Qual Life Outcomes 2009 Feb 19;7(1):15. [doi: 10.1186/1477-7525-7-15] [Medline: 19228398]
47. Putz R, Taggart F, Stewart-Brown S. Using WEMWBS to measure the impact of your work on mental wellbeing: A practice-based user guide. 2012. URL: https://corc.uk.net/media/1244/wemwbs_practitioneruserguide.pdf?_sm_au_=isV7BM6n6vMj3BNFc17QjB7s6k3C1 [accessed 2026-05-22]
48. Lam LT. The utility of the Short Warwick Edinburgh Mental Wellbeing Scale (SWEMWBS) in caregivers: an application of item response theory. Curr Psychol 2025 Jun;44(12):11496-11507. [doi: 10.1007/s12144-025-07971-0]
49. Reynolds R, Kingsland M, Daly J, et al. Breastfeeding practices and associations with pregnancy, maternal and infant characteristics in Australia: a cross-sectional study. Int Breastfeed J 2023 Jan 19;18(1):8. [doi: 10.1186/s13006-023-00545-5] [Medline: 36658629]
50. Remoteness areas. Australian Bureau of Statistics. 2021. URL: https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026/remoteness-structure/remoteness-areas [accessed 2026-05-04]
51. Census of population and housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2016. Australian Bureau of Statistics. 2018. URL: https://www.abs.gov.au/ausstats/abs@.nsf/mf/2033.0.55.001 [accessed 2026-05-04]
52. Abbott JH. The distinction between randomized clinical trials (RCTs) and preliminary feasibility and pilot studies: what they are and are not. J Orthop Sports Phys Ther 2014 Aug;44(8):555-558. [doi: 10.2519/jospt.2014.0110] [Medline: 25082389]
53. Moschonis G, Siopis G, Jung J, et al. Effectiveness, reach, uptake, and feasibility of digital health interventions for adults with type 2 diabetes: a systematic review and meta-analysis of randomised controlled trials. Lancet Digit Health 2023 Mar;5(3):e125-e143. [doi: 10.1016/S2589-7500(22)00233-3] [Medline: 36828606]

XSL•FO
RenderX

54. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 5;13(3):e52. [doi: 10.2196/jmir.1772] [Medline: 21821503]
55. Delaney T, Mclaughlin M, Hall A, et al. Associations between digital health intervention engagement and dietary intake: a systematic review. Nutrients 2021 Sep 20;13(9):3281. [doi: 10.3390/nu13093281] [Medline: 34579158]
56. Mclaughlin M, Delaney T, Hall A, et al. Associations between digital health intervention engagement, physical activity, and sedentary behavior: systematic review and meta-analysis. J Med Internet Res 2021 Feb 19;23(2):e23180. [doi: 10.2196/23180] [Medline: 33605897]
57. Grady A, Pearson N, Lamont H, et al. The effectiveness of strategies to improve user engagement with digital health interventions targeting nutrition, physical activity, and overweight and obesity: systematic review and meta-analysis. J Med Internet Res 2023 Dec 19;25:e47987. [doi: 10.2196/47987] [Medline: 38113062]
58. Steen M, Manschot M, Koning ND. Benefits of co-design in service design projects. Int J Design 2011;5(2):53-60 [FREE Full text]
59. Yahia EM, García-Solís P, Celis MEM. Contribution of fruits and vegetables to human nutrition and health. In: Yahia EM, editor. Postharvest Physiology and Biochemistry of Fruits and Vegetables: Woodhead Publishing; 2019:19-45. [doi: 10.1016/B978-0-12-813278-4.00002-6]
60. Children's risk factors. Australian Bureau of Statistics. 2018. URL: https://www.abs.gov.au/statistics/health/health-conditions-and-risks/childrens-risk-factors/2017-18 [accessed 2025-01-20]
61. Woodside JV, Nugent AP, Moore RE, McKinley MC. Fruit and vegetable consumption as a preventative strategy for non-communicable diseases. Proc Nutr Soc 2023 May;82(2):186-199. [doi: 10.1017/S0029665123002161] [Medline: 36785878]
62. Hendrie GA, Lease HJ, Bowen J, Baird DL, Cox DN. Strategies to increase children's vegetable intake in home and community settings: a systematic review of literature. Matern Child Nutr 2017 Jan;13(1):e12276. [doi: 10.1111/mcn.12276]
63. Birch L, Savage JS, Ventura A. Influences on the development of children's eating behaviours: from infancy to adolescence. Can J Diet Pract Res 2007;68(1):s1-s56. [Medline: 19430591]
64. Daniels LA, Mallan KM, Battistutta D, et al. Child eating behavior outcomes of an early feeding intervention to reduce risk indicators for child obesity: the NOURISH RCT. Obesity (Silver Spring) 2014 May;22(5):E104-E111. [doi: 10.1002/oby.20693] [Medline: 24415390]
65. Augusto RA, Cobayashi F, Cardoso MA, ACTION Study Team. Associations between low consumption of fruits and vegetables and nutritional deficiencies in Brazilian schoolchildren. Public Health Nutr 2015 Apr;18(5):927-935. [doi: 10.1017/S1368980014001244] [Medline: 24963861]
66. Maynard M, Gunnell D, Emmett P, Frankel S, Davey Smith G. Fruit, vegetables, and antioxidants in childhood and risk of adult cancer: the Boyd Orr cohort. J Epidemiol Community Health 2003 Mar;57(3):218-225. [doi: 10.1136/jech.57.3.218] [Medline: 12594199]
67. Hinkley T, McCann JR. Mothers' and father's perceptions of the risks and benefits of screen time and physical activity during early childhood: a qualitative study. BMC Public Health 2018 Nov 20;18(1):1271. [doi: 10.1186/s12889-018-6199-6] [Medline: 30453927]
68. McMahon CA, Barnett B, Kowalenko NM, Tennant CC. Maternal attachment state of mind moderates the impact of postnatal depression on infant attachment. J Child Psychol Psychiatry 2006 Jul;47(7):660-669. [doi: 10.1111/j.1469-7610.2005.01547.x] [Medline: 16790001]
69. Cummings EM, Davies PT. Maternal depression and child development. J Child Psychol Psychiatry 1994 Jan;35(1):73-112. [doi: 10.1111/j.1469-7610.1994.tb01133.x] [Medline: 8163630]
70. Fan X, Wu N, Tu Y, et al. Perinatal depression and infant and toddler neurodevelopment: a systematic review and meta-analysis. Neurosci Biobehav Rev 2024 Apr;159:105579. [doi: 10.1016/j.neubiorev.2024.105579] [Medline: 38342472]
71. Perinatal mental health screening in Australia. Australian Institute of Health and Welfare. 2024. URL: https://www.aihw.gov.au/reports/mothers-babies/perinatal-mental-health-screening-australia [accessed 2026-05-04]
72. Moss KM, Reilly N, Dobson AJ, Loxton D, Tooth L, Mishra GD. How rates of perinatal mental health screening in Australia have changed over time and which women are missing out. Aust N Z J Public Health 2020 Aug;44(4):301-306. [doi: 10.1111/1753-6405.12999] [Medline: 32510784]
73. Kunselman AR. A brief overview of pilot studies and their sample size justification. Fertil Steril 2024 Jun;121(6):899-901. [doi: 10.1016/j.fertnstert.2024.01.040] [Medline: 38331310]

Abbreviations

CATI: computer-assisted telephone interview
CFHN: Child and Family Health Nursing
DHI: digital health intervention
HB4HNEKids: Healthy Beginnings for Hunter New England Kids
HNE: Hunter New England
HNELHD: Hunter New England Local Health District

XSL•FO
RenderX

MBQ-B: Movement Behavior Questionnaire—Baby
MD: mean difference
mHealth: mobile health
NSW: New South Wales
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
SFS: Short Food Survey
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
SWEMWBS: Short Warwick Edinburgh Mental Wellbeing Scale

Edited by L Buis; submitted 28.Aug.2025; peer-reviewed by A Bhat, S Sahai; revised version received 05.Feb.2026; accepted 23.Mar.2026; published 15.Jun.2026.

Please cite as:
Jackson J, Hudson N, Delaney T, Brown AL, Kavalec S, Pinfold J, Sewter R, McCormick H, Liackman R, Lecathelinais C, Davies L, Wolfenden L, Groombridge D, Tully B, Redman S, Smart T, Ghafournia N, Craven PD, Sutherland R
A Pilot mHealth Text Messaging Program Targeting Parents During the First 2000 Days: Nonrandomized Repeat Cross-Sectional Analysis to Evaluate Feasibility, Engagement, Acceptability, and Potential Effectiveness
JMIR Mhealth Uhealth 2026;14:e83162
URL: https://mhealth.jmir.org/2026/1/e83162
doi:10.2196/83162

© Jacklyn Jackson, Nayerra Hudson, Tessa Delaney, Alison L Brown, Sienna Kavalec, Jessica Pinfold, Rebecca Sewter, Hannah McCormick, Rebecca Liackman, Christophe Lecathelinais, Lynda Davies, Luke Wolfenden, Daniel Groombridge, Belinda Tully, Sinead Redman, Tauri Smart, Nafiseh Ghafournia, Paul David Craven, Rachel Sutherland. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 15.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Efficacy of a Mobile Phone–Based Life-Skills Training Program for Addiction Prevention Optimized Among Adolescents With Low Program Engagement: Cluster-Randomized Controlled Trial

Severin Haug, PhD; Nikolaos Boumparis, PhD; Olivia Studhalter, MSci; Andreas Wenger, MSci; Michael P Schaub, PhD; Nikolai Kiselev, PhD

Swiss Research Institute for Public Health and Addiction ISGF at University of Zurich, Konradstrasse 32, Zurich, Switzerland

Corresponding Author:
Severin Haug, PhD
Swiss Research Institute for Public Health and Addiction ISGF at University of Zurich, Konradstrasse 32, Zurich, Switzerland

Abstract

Background: Previous studies demonstrated the effectiveness of *SmartCoach*, a mobile phone–based life-skills training program for addiction prevention among adolescents. However, socially stratifying factors, such as educational level or migration background, were associated with lower program engagement and participation. To address these disparities, we optimized and tailored program elements, particularly for subgroups with low engagement, using qualitative interview data.

Objective: This study aimed to test whether the optimized program version was superior to the original one in terms of efficacy and program use. We hypothesized that optimization and advanced tailoring would result in increased program efficacy and engagement.

Methods: A 2-arm, parallel-group, double-blind, cluster-randomized controlled trial was conducted with assessments at baseline and follow-up after 6 months. Secondary and upper secondary school classes were randomized into two groups: (1) an intervention group receiving the optimized program version and (2) a control group receiving the original version. Participants in both groups received up to 4 weekly SMS text messages over 4 months to stimulate (1) self-management skills to cope with stress, (2) self-efficacy to resist social pressure, (3) observational learning, and (4) self-regulation. The optimized program version included additional reminders, extended time for challenges, simplified language, different days for weekly messages, and additional tailoring for participants with personal or parental backgrounds from non–German-speaking countries. This subgroup received shorter video clips and information on additional stressors, including acculturation and family conflicts. The primary outcomes of this trial were the quantity of alcohol use, the number of days per month that nicotine-containing products were smoked, and cannabis use days.

Results: A total of 1175 students from 67 Swiss secondary and upper secondary school classes were invited to participate in the study between September 2023 and September 2024. Of these 1175 students, 890 (75.7%; mean age 14.9, SD 1.3 y; women: n=511, 57.4%; men: n=379, 42.6%; nonbinary: n=0, 0.0%) participated in the study. Six-month follow-up assessments were completed by 552 (62.0%) of the 890 study participants. On average, program use did not differ between the study groups, with 12.6 (SD 12.8) of 38 possible activities completed in the intervention group and 12.5 (SD 12.7) activities in the control group (P=.18). The results concerning efficacy showed no significant differences between the study groups regarding the main outcomes of alcohol use quantity (P=.91), nicotine use days (P=.45), and cannabis use days (P=.67). Similarly, no significant group differences were observed for changes in perceived stress (P=.77) and social skills (P=.45).

Conclusions: The results did not show improved program use or efficacy in the optimized version of the program. Possible explanations include the relatively minor differences between the 2 program versions and the generally low prevalence rates of substance use in this age group, which limited the statistical power of the study.

Trial Registration: ISRCTN Registry ISRCTN10423173; https://www.isrctn.com/ISRCTN10423173

(JMIR Mhealth Uhealth 2026;14:e78081) doi:10.2196/78081

KEYWORDS

life skills; substance use; prevention; tailoring; adolescents; mobile phone

XSL•FO
RenderX

Introduction

Background

Substance use remains a significant public health issue in Western Europe and Switzerland [1]. This is particularly concerning due to the developmental vulnerability of adolescents, who are at a critical stage of brain maturation, particularly in areas related to impulse control, decision-making, and emotional regulation, which makes them more susceptible to risk-taking behaviors, including experimentation with substances [2,3]. Social and environmental influences also play a significant role. Adolescents are highly influenced by peer norms and social acceptance, with peer pressure and group dynamics often encouraging substance use [4,5]. Numerous meta-analyses and systematic reviews have assessed the effectiveness of school-based programs in preventing alcohol, tobacco, and other substance use [6-8]. Despite varying methodological rigor and theoretical foundations, these reviews highlight that school-based prevention can be effective, particularly when programs are grounded in psychosocial theory, build drug resistance and general competence skills, and are delivered interactively by teachers or program providers. Such life-skills intervention programs to prevent substance use [9-11] primarily combine training in self-management skills (eg, adapting to stress, emotional self-regulation, and goal setting), social skills (eg, assertiveness and communication skills), and skills facilitating resistance to substance use (eg, opposing peer pressure to drink alcohol and identifying and resisting media influences). However, their implementation and dissemination in schools present serious challenges [12,13]. First, teachers and other professionals need the time, motivation, knowledge, and skills to deliver the program. Second, extensive resources, in terms of personnel, money, and time allocated to deliver substance use prevention, are required to prepare and administer such programs [14].

Digital interventions have great potential to overcome the aforementioned obstacles that hinder successful program implementation and larger-scale dissemination of life-skills training in schools. A hybrid life-skills training program, combining e-learning modules and in-person sessions, among US middle school students, led to significant reductions in substance use and increases in health knowledge, skills knowledge, and life skills compared to a control condition [12].

SmartCoach (Swiss Research Institute for Public Health and Addiction) presented the first automated mobile phone–based life-skills training program to prevent substance use among adolescents, whose long-term effectiveness has been tested in a controlled study [15]. It includes online feedback and individually tailored SMS text messages provided over several months. The content is based on social cognitive theory and addresses self-management skills, social skills, and substance use resistance skills. The results on its efficacy at the 18-month follow-up showed that compared to controls, those in the intervention group experienced significantly lower tobacco smoking and cannabis use prevalence, while no effects were observed for alcohol use, well-being, or social skills [16].

A major challenge in delivering life-skills training and other preventive interventions electronically is the lack of control over adolescents' engagement with the intervention. Several reviews on digital interventions to promote mental health [17,18] or to prevent substance use [19] in young people point to the relatively low levels of user engagement.

Poor program engagement typically jeopardizes program efficacy [20,21]. The *SmartCoach* efficacy study also showed several associations between engagement and intended outcomes, with adolescents who used the contests more frequently being more likely to be nonsmokers, and those who read the messages more attentively being less likely to drink in a problematic manner at the 6-month follow-up [22].

Several studies among adolescents and adults show associations between engagement in mobile health interventions and sociodemographic measures [22-24]. For example, in a SMS text messaging–based smoking cessation program in Switzerland, adolescents without an immigrant background showed higher program engagement [23]. In the *SmartCoach* program, students who attended a secondary school, as opposed to an upper secondary school, and those who reported a 2-sided immigration background (ie, both parents were not born in Switzerland), as opposed to those with no immigration background or a 1-sided immigration background, showed lower program engagement [22].

The challenge of low program use and engagement in mobile health (mHealth) interventions is increasingly discussed from scientific and practical perspectives; however, there is a lack of generalizable approaches that demonstrate how program engagement in existing mHealth programs can be systematically improved [25]. Promising approaches include advanced tailoring of intervention content [24,26] and involving adolescents in the design process to ensure the content is relevant and appealing to them [27,28].

These methods also provided the foundation for developing an optimized version of the *SmartCoach* program, which was tested in this study: first, subgroups with low program engagement were identified, and second, factors influencing user engagement and suggestions for program optimization were derived by using a participatory approach.

To identify subgroups with low program engagement, we evaluated several demographic and socioeconomic predictors of program participation and program engagement among 476 secondary and upper secondary school students in Switzerland in a previous study [29]. The results of that study concerning program engagement highlighted that a migration background, or more specifically, an origin from a non–German-speaking country, was the strongest single predictor of low program engagement [29].

Additionally, aiming to identify factors influencing user engagement and suggestions for the optimization of the *SmartCoach*, semistructured phone interviews were conducted among 171 program participants from the previous study [29]. The Capability, Opportunity, and Motivation model of behavior change [30] and the theoretical domains framework (TDF) [31] were used to assess behavioral influences [32].

XSL•FO
RenderX

Next to the general recommendations for advanced tailoring of the *SmartCoach* program content, specific recommendations for the subgroup of students originating from non–German-speaking countries were also derived from the previous qualitative study and implemented in this study (see the Methods section) [32].

Study objectives

The main objective of this study was to test program engagement and the efficacy of the *SmartCoach* version optimized in this way, compared to the original program version without advanced tailoring. We hypothesized that receiving the optimized intervention program, compared to the original version, would result in increased program efficacy, indicated by lower substance use at 6-month follow-up (primary outcomes), increased program engagement and social skills, as well as decreased levels of stress (secondary outcomes).

Methods

Study Design

This study aimed to test program engagement and the efficacy of the optimized program version with advanced tailoring compared to the original program version in a 2-arm parallel-group, double-blind, randomized controlled trial with assessments at baseline and at a 6-month follow-up.

Participants, Setting, and Procedure

In this study, secondary and upper secondary schools in the Swiss cantons of Aargau and Zurich were invited to participate by prevention experts from regional addiction prevention centers. Interested teachers were informed about the program's objectives and its implementation in their classes, and they allocated 45 to 90 minutes during regular school lessons for the program.

During this time, students were introduced to the topic through a brief workshop, informed about the program and the accompanying study, and invited to participate and sign up for the study. The workshop and information session were led by prevention experts or psychology master's students experienced in working with young people and delivering preventive interventions.

The 30‐ to 60-minute workshop aimed to spark students' interest in stress and provide basic information on its origins. Initially, students shared their understanding of stress in a group discussion. The concept of stress as an imbalance between demands and resources was then explained. Interactive exercises, such as repeatedly subtracting 7 from 996, were used to deliberately induce stress, allowing students to observe its effects on their bodies and thoughts. Finally, students described their stress levels in areas like family, friends, school, and the future using emojis.

Following this, students were introduced to the *SmartCoach* program and the related study through an introductory video [33]. Students were invited to participate in the study if they met the following criteria: (1) were at least 14 years old and (2) owned a mobile phone.

They were asked to complete an online baseline assessment and study registration using their smartphones. Students who consented to participate provided a self-defined username and mobile phone number and completed additional assessments on stress and social skills for tailoring the intervention content. They received individually tailored web-based feedback on their mobile phones. Students who did not give consent were thanked for their interest and asked to explain their reasons for not participating in a free-text field.

Those who decided to participate received personalized mobile phone–based life-skills training over the next 4 months. Follow-up assessments at month 6 in both study groups were conducted using a similar procedure: participants were invited to the online follow-up assessments via SMS text messaging, which included a link to the follow-up survey. After 3 reminders, nonresponders were additionally contacted via computer-assisted telephone interviews conducted by research assistants. Study participants were recruited between September 2023 and September 2024. The 6-month follow-up assessments were conducted between March 2024 and April 2025.

Ethical Considerations

The study was approved by the Ethics Committee of the Faculty of Arts and Sciences at the University of Zurich (approval 22.2.15; date of approval April 17, 2022). Digital informed consent was obtained from all study participants. Parental consent was required only for children under 15 years of age because the study was classified by the ethics committee as a research project with minimal risk [34]. Study participants completed the screening procedure and baseline assessment anonymously. No names were collected; instead, participants chose nicknames. Password protection and secure sockets layer encoding were used to ensure the privacy and security of data transfer. Study participants received an incentive of 10 Swiss Francs (1 Swiss Franc=US $1.23 as of June 23, 2026) for taking part in the follow-up survey.

Randomization and Allocation Concealment

To prevent spillover effects within school classes, we conducted a cluster-randomized controlled trial, using school classes as the randomization unit. Multiple classes per school were randomized. Given the diverse student populations across different secondary schools, we used stratified randomization with separate computer-generated randomization lists for each school. In addition, to ensure balanced sample sizes across study groups, we used block randomization with computer-generated randomly permuted blocks of 4 cases [35].

Prevention experts and master's students conducting the baseline assessment were blinded to the group allocation of school classes. Furthermore, the research assistants who performed the computer-assisted follow-up assessments for primary and secondary outcomes were also blinded to the group allocation. Additionally, students did not know whether they were allocated to the original or optimized programs.

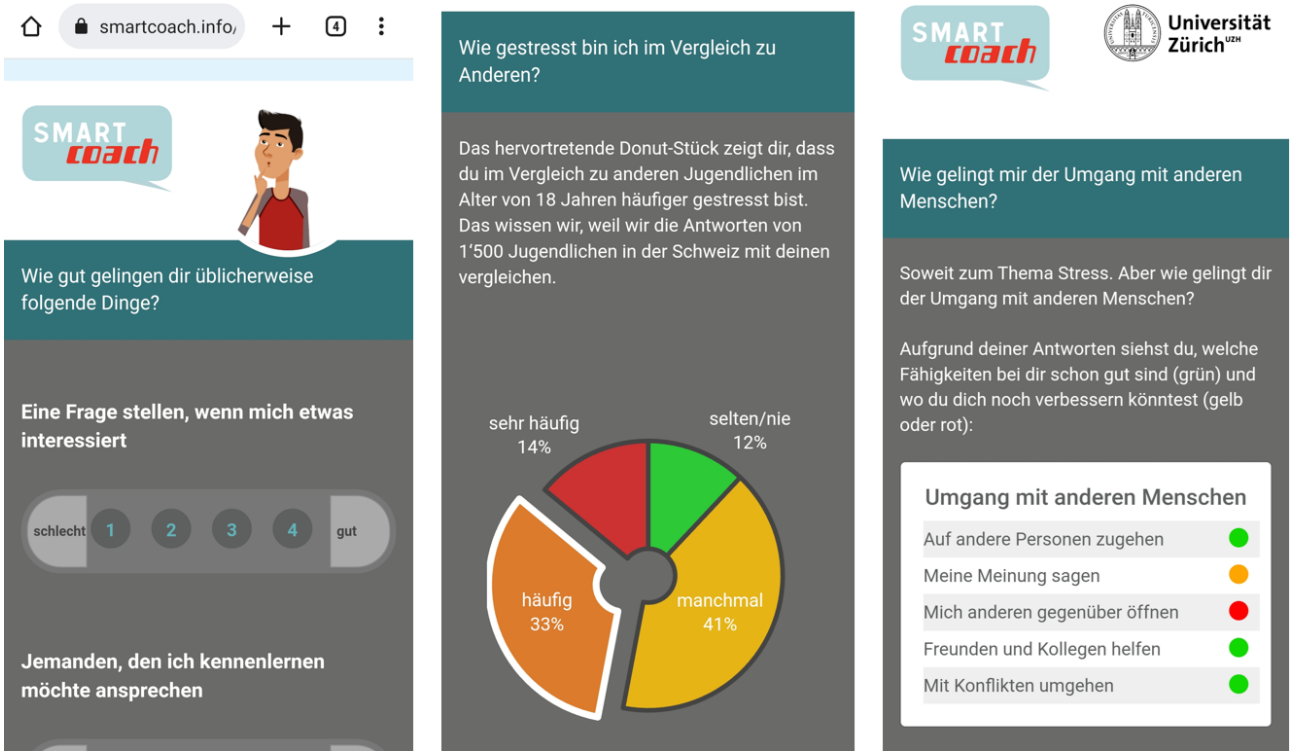
XSL•FO
RenderX

Intervention Program

Original Program Version

The intervention elements of the program were based on the social cognitive theory [36,37]. The key concepts of this theory, which were addressed within *SmartCoach*, were (1) outcome expectations, (2) self-efficacy, (3) observational learning, (4) facilitation, and (5) self-regulation. The *SmartCoach* program [33] provided tailored web-based feedback and SMS text messages to promote self-management skills, social skills, and substance use resistance skills over a period of 4 months. The intervention elements of the program were based on social cognitive theory [36,37] and the life-skills training [9]. Individually tailored web-based feedback was provided to program participants immediately after they completed the baseline assessment within their school classroom. This included textual and graphical feedback on general stress, levels of stress in various domains, individual applied and suggested coping strategies, and individual levels of social skills. Screenshots of the baseline assessment and the web-based feedback are shown in Figure 1.

Figure 1. Screenshots from the *SmartCoach* program (left: assessment of social skills; middle: feedback on level of stress; right: feedback on individual social skills).



Following the baseline assessment and feedback, program participants received between 2 and 4 individualized SMS text messages per week on their mobile phones for a period of 17 weeks. These messages were generated and sent by a fully automated (not artificial intelligence–based) system. For the first 7 weeks, the messages focused on self-management skills, for example, coping with stress, emotional self-regulation, and managing anger and frustration. In weeks 8 to 12, the messages focused on social skills, for example, making requests, refusing unreasonable requests, and meeting new people. In weeks 13 to 17, the SMS text messages focused on substance use resistance skills, for example, recognizing and resisting media influences, social norms of alcohol and tobacco use, and the associations between both self-management and social skills and substance use. The messages were tailored based on data from the baseline assessment and on SMS text messaging assessments that occurred over the course of the program.

During the 4-month coaching phase, participants received a total of 38 SMS text message prompts that invited interaction, such as replying to quizzes, retrieving media objects, and participating in challenges or contests. A friendly competition among participants was integrated into the program to stimulate engagement. Within the friendly competition, program participants could collect credits for each interaction (eg, answering monitoring SMS text messages, participating in quizzes, creating messages or pictures within contests, and accessing video links integrated into SMS text messages). The more credits participants collected, the higher their chances were of winning one of several attractive prizes, which were part of a prize draw (10 prizes with a total value of 500 Swiss Francs, eg, cinema tickets) after program completion. Participants were able to retrieve their number of credits compared to the number of credits of other program participants in their group (with a similar starting date) at any time from an individual profile page. Sample screenshots of the program from the coaching phase are shown in Figure 1.

Optimized Program Version

The optimized program version was developed as described in the "Background" section based on a quantitative study that identified students from non–German-speaking countries as a subgroup with particularly low program engagement [29], as well as on a qualitative examination of suggestions for program

XSL•FO
RenderX

optimization in this subgroup and also among other program participants [32]. The optimized program version included additional tailoring for participants with personal or parental backgrounds from non–German-speaking countries, who received shorter video clips. Furthermore, other stressors, including acculturation and family conflicts, were addressed, and the weekly SMS text messages were not always sent on Tuesdays but on different days of the week. A comprehensive overview of the program optimization version is displayed in Table 1.

Table . Program optimizations for all participants and additional optimizations for those with personal or parental backgrounds from non–German-speaking countries.

Suggestions from qualitative study [32]	Optimizations for all participants	Additional optimizations for participants with personal or parental backgrounds from non–German-speaking countries
Shorter videos	—[a]	Videos shortened to a maximum duration of 2 min.
Simplifying language	Complex or long sentences were revised and simplified to ensure clarity and accessibility.	—
General reminders	Participants who did not respond to certain prompts received additional reminder messages to encourage engagement.	—
Additional reminder message for upload within group challenges	Participants received a reminder message 1 h before the end of the group challenge submission deadline if no action was taken.	—
Reminder for voting in self-challenges	Participants received a reminder 1 h before the voting phase ended if they had not yet voted.	—
SMS on different days	Weekly interactive messages (eg, quizzes) were sent on 2 different days (Tuesdays and Fridays) rather than always on Tuesdays to better fit diverse routines.	—
Adapted timing for challenge participation	The time window for submitting entries to group challenges was adjusted to earlier in the afternoon (and prolonged from 2-4 h this way) to accommodate daily schedules.	—
Adapted timing for voting on challenges	The voting period was extended until 3 PM (an additional 3 h), allowing more flexibility for participants to engage.	—
Time for publication of group challenge results	Challenge results were sent out shortly after the extended voting period ended.	—
Regular updates on personal progress	Additional updates about collected points and status in the friendly competition were sent periodically.	—
Ensuring anonymity in group interactions	The program provided clearer and simplified explanations about anonymity to increase comfort and trust in participation.	—
More differentiated assessment of stress factors	The stress assessment was refined to distinguish between stress from family conflicts and stress from peer conflicts.	The explanation of family conflict stress included examples relating to cultural differences and intergenerational misunderstandings.
Addressing feelings of exclusion	—	New content was added to reflect on feelings of being an “outsider,” including migration-related themes like acculturation and identity development.
Addressing future-related stress	—	Specific questions and feedback regarding worries about the future were tailored to migration-related challenges (eg, education and residence status).

[a]Not applicable.

XSL•FO
RenderX

Assessments and Outcomes

Sociodemographic Data

Assessments of demographic and socioeconomic variables included the birth country of both parents and the students to determine a migration background. Since Swiss-German people are similar to Germans and Austrians in terms of language, culture, and socioeconomic background, a further variable was created based on the students' and their parents' country of birth, which describes their origin from a non–German-speaking country. Furthermore, we assessed gender, age, and perceived familial socioeconomic status from the perspective of the adolescent using a brief instrument [38,39] that was translated, adapted, and validated for German adolescents [40].

Program Use

Program use was operationalized in terms of the total number of interactions with the program that were logged by the system. During the 4-month coaching phase, participants received a total of 38 SMS text messages that prompted a total of 18 SMS text messaging–based reply activities (eg, replies to quiz questions or selection queries) and 20 media view activities (eg, views of video clips, images, or weblinks). These activities were logged through the *SmartCoach* system and were available for each program participant.

Outcomes

Baseline and follow-up assessments included the following:

- Alcohol use in the preceding 30 days was assessed by the consumption items of the Alcohol Use Disorders Identification Test (AUDIT-C) [41]. This test comprises three items assessing (1) frequency of alcohol consumption, (2) quantity of alcohol consumption, and (3) binge drinking, and it was also validated for use among adolescents [42]. Pictures were used to illustrate the quantity of a standard drink, which corresponds to 12 to 14 g of pure alcohol.
- Nicotine or tobacco use was assessed using the item: "Have you smoked a puff of nicotine-containing cigarettes, e-cigarettes, or vapes within the last 30 days?" If the participant responded "yes," we additionally assessed the number of nicotine or tobacco use days in the past 30 days.
- Cannabis use in the preceding 30 days was assessed using the item: "Have you smoked or used products containing cannabis within the last 30 days?" If the participant responded "yes," we additionally assessed the number of days they used cannabis.
- Perceived stress was assessed by a single item from the Swiss Juvenir study [43]: "How often have you had the feeling of being overstressed or overwhelmed in the last month?" with the answer options ranging from 1 "never" to 5 "all the time."
- Social skills were assessed using the brief version of the Interpersonal Competence Questionnaire-10 [44], which addresses the following domains: initiation of relationships, negative assertion, disclosure of personal information, emotional support, and conflict management.

The primary outcomes, according to the study registration, were alcohol use in the preceding 30 days, as measured by the AUDIT-C total score; the number of nicotine or tobacco use days in the past 30 days; and the number of cannabis use days in the preceding 30 days. Secondary outcomes included perceived stress, social skills, and program use captured via the total number of interactions with the program that were logged by the system.

Sample Size

The effect size calculations for this trial were based on a meta-analysis of the efficacy of SMS text messaging–based interventions for health promotion [26]. It tested a variety of participant, intervention, and methodological moderators and identified message tailoring, as well as personalization, as being significantly associated with greater intervention efficacy. The largest effect sizes were exhibited in those studies that used both message tailoring and targeting (Cohen d=0.44), followed by studies using message tailoring only (Cohen d=0.27), whereas the smallest effects were observed among studies that used only message targeting (Cohen d=0.07). Furthermore, the analysis also examined the impact of tailoring SMS text messages specifically based on demographic factors in message design. Interventions were significantly more efficacious when they used demographic factors (Cohen d=0.53) in tailoring than when they did not (Cohen d=0.22). Based on these findings, we expected an increased effectiveness of the optimized program version, which provides advanced tailoring based on several socially stratifying factors, by a small to medium magnitude. Assuming a Cohen d of 0.25, a sample size of 253 in each study group would be required to achieve 80% power for a t test (α=5%, 2-sided) to detect this difference based on a calculation using G*Power (Heinrich Heine University Düsseldorf). As secondary school students are nested within school classes, we additionally needed to consider a potential design effect for the calculation of the sample size for our study. Based on an average cluster size of 15 study participants [16] and an expected intracluster correlation coefficient of 0.05, a sample size of 430 per study group and a total of 860 study participants was required.

Data Analysis

To examine baseline differences between participants in the intervention and control groups, we performed chi-square tests for categorical variables, t-tests, and Mann-Whitney U tests for continuous variables. The same tests were applied to assess whether participants lost to follow-up differed from those who responded as a function of the study group.

We analyzed data according to the intention-to-treat principle (ITT). For the ITT analyses, we used multiple imputation procedures as described elsewhere [45]. Specifically, the imputation model included study group, demographic variables (gender, age, immigration background, and socioeconomic status), school-related factors (school type and class), baseline measures of all outcomes (stress, social skills, and substance use), and program engagement metrics. We generated 20 imputed datasets with 5 iterations each, and no systematic bias in convergence was observed. Binary variables were imputed using logistic regression, and both count and continuous variables were imputed using predictive mean matching.

XSL•FO
RenderX

For all outcomes, we used generalized linear mixed models (GLMMs) with a random intercept for the school class. For count outcomes, we used Poisson GLMMs with a log link function and included the log of the baseline measure plus one as an offset term to model relative changes. For dichotomous outcomes, we used binomial GLMMs with a logit link. For continuous outcomes, we used linear mixed models with change scores (follow-up minus baseline) as the dependent variable. All models included the study group as the primary predictor, with age, gender, school type, juvenile stress, and baseline cannabis use as covariates. We included similar covariates in the program engagement models.

We did not adjust for multiple comparisons, as this was not prespecified in the sample size calculation or study protocol. To avoid post hoc bias, analyses followed the original protocol, using predefined significance thresholds without multiplicity correction.

To identify the moderators of the primary and secondary intervention outcomes, GLMMs with a random intercept for school class were used. Interaction terms between sociodemographic characteristics (gender, migration status, origin from a non–German-speaking country, socioeconomic status) measured at baseline and study group were included, one at a time, in the previously described models.

Results with a type I error rate of $P<.05$ on 2-sided tests were considered statistically significant. Analyses were performed using R version 3.6.1 with the lme4 package [45] for fitting mixed models.

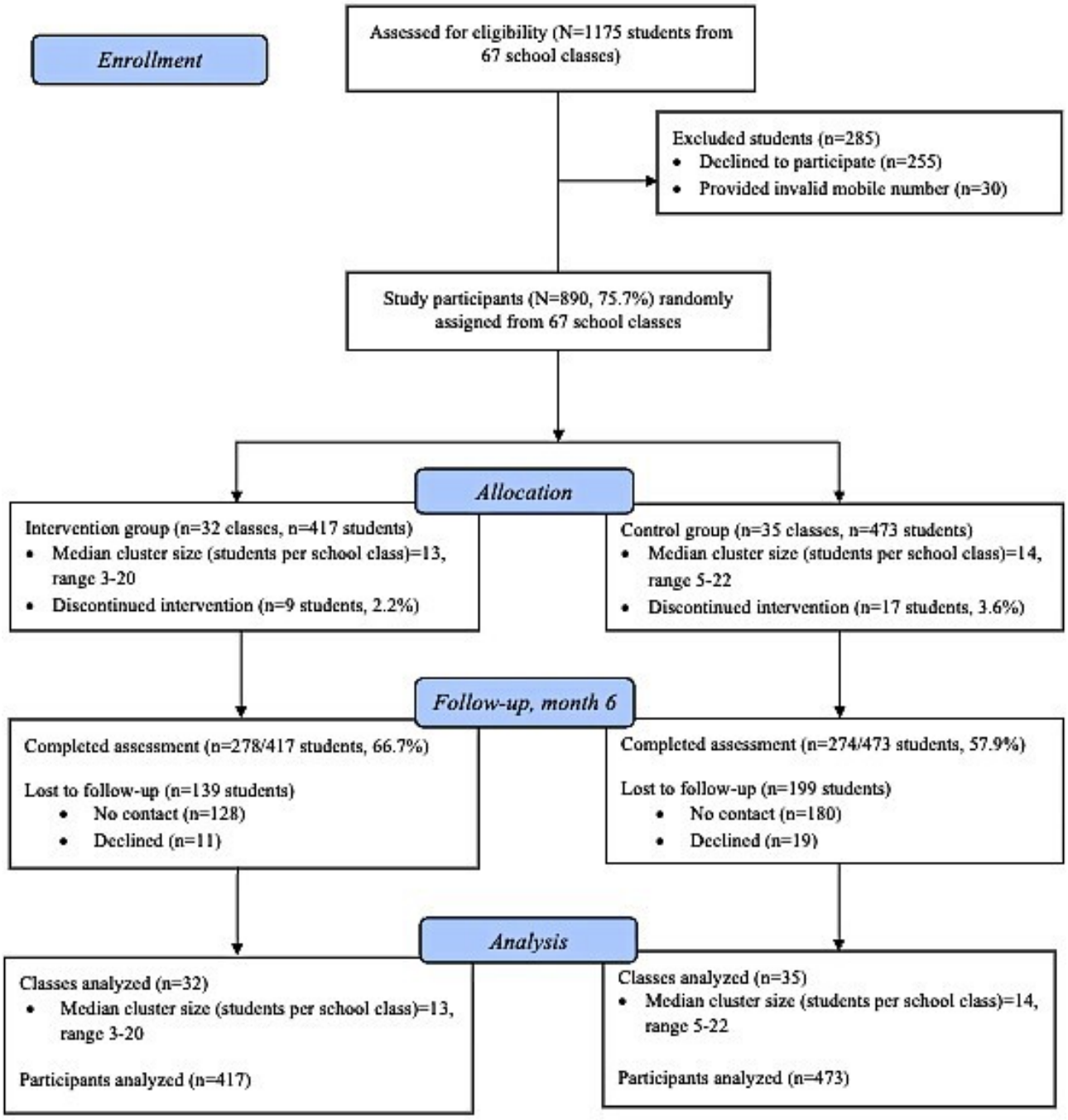
Results

Study Participants

Figure 2 depicts participants' progression through the trial. At the online screening assessment, 1175 students were present in 67 classes. Of these 1175 students, 890 (75.7%) consented to participate in the study and provided a valid mobile phone number. Of the 255 students who declined to participate, 83 (32.5%) provided usable free-text information about their reasons for not taking part in the program and associated study. The most common reason given was a lack of desire or interest in participating in the program (n=47, 57%). A total of 22 (27%) students stated that they did not have their smartphone with them; another 5 (6%) indicated that they did not have a smartphone at all. Three (4%) students said they were unwell. Two (2%) students mentioned privacy concerns, 2 (2%) did not have parental consent, and 2 (2%) experienced technical difficulties with the internet or their smartphones.

XSL•FO
RenderX

Figure 2. Flowchart of study participants.



Thirty-two classes containing 417 students in total were randomly assigned to the intervention group, and 35 classes with 473 students in total were assigned to the control group. Six-month follow-up assessments were completed by 278 of 417 (66.7%) students in the intervention groups and 274 of 473 (57.9%) students in the control group.

Baseline characteristics for the study sample are shown in Table 2. The mean age was 14.9 (SD 1.3) years, and 511 (57.4%) of the 890 study participants were women.

Baseline differences between the intervention and control groups were identified for age, type of school, and perceived stress.

Concerning attrition bias, the analysis revealed that intervention group participants who were lost to follow-up were more likely to be men (χ^2_1=8.5; *P*=.003) and had a higher cannabis use prevalence at baseline (χ^2_1=6.0; *P*=.014). Furthermore, intervention group participants lost to follow-up were more likely to have lower levels of perceived stress ($F_{1,336}$=4.8; *P*=.03) than controls.

The intracluster correlation coefficients (ICCs) for the 3 primary outcomes were as follows: for AUDIT-C, ICC of 0.12 at baseline and 0.03 at follow-up and for nicotine or tobacco use days, ICC of 0.08 at baseline and 0.03 at follow-up. For cannabis use days, the random-intercept models did not converge at baseline and follow-up, indicating negligible between-cluster variance.

XSL•FO
RenderX

Table . Baseline characteristics of the study sample.

Variable	Intervention (n=417)	Control (n=473)	Total (N=890)	*P* value[a]
Gender, n (%)				.42[b]
Man	184 (44.1)	195 (41.2)	379 (42.6)	
Woman	233 (55.9)	278 (58.8)	511 (57.4)	
Nonbinary or other	0 (0)	0 (0)	0 (0)	
Age (y), mean (SD)	15.2 (1.3)	14.7 (1.3)	14.9 (1.3)	<.001[c]
Migration background, n (%)				.29[b]
No	133 (31.9)	167 (35.3)	300 (33.7)	
Yes	284 (68.1)	306 (64.7)	590 (66.3)	
Origin from a non–German-speaking country, n (%)				.31[b]
No	166 (39.8)	205 (43.3)	371 (41.7)	
Yes	251 (60.2)	268 (56.7)	519 (58.3)	
Type of school, n (%)				.004[b]
Secondary school	264 (63.3)	343 (72.5)	607 (68.2)	
Upper secondary school	153 (36.7)	130 (27.5)	283 (31.8)	
Socioeconomic status, mean (SD)	6.6 (1.6)	6.5 (1.5)	6.6 (1.5)	.57[c]
Perceived stress, mean (SD)	3.2 (1.0)	3.4 (1.0)	3.3 (1.0)	.004[c]
Social skills (ICQ-10[d]), mean (SD)	28.6 (4.8)	28.3 (4.7)	28.4 (4.8)	.35[c]
Nicotine or tobacco use in preceding 30 d, n (%)				.31[b]
No	329 (78.9)	387 (81.8)	716 (80.4)	
Yes	88 (21.1)	86 (18.2)	174 (19.6)	
Nicotine or tobacco use days in preceding 30 d, mean (SD)	3.0 (8.0)	2.4 (7.1)	2.6 (7.5)	.23[e]
Alcohol use, preceding 30 d (AUDIT-C score[f], range 0 - 12), mean (SD)	1.2 (2.4)	1.1 (2.2)	1.1 (2.3)	.32[e]
Cannabis use, preceding 30 d, n (%)				.09[b]
No	386 (92.6)	451 (95.3)	837 (94.0)	
Yes	31 (7.4)	22 (4.7)	53 (6.0)	
Cannabis use days, preceding 30 d, mean (SD)	0.5 (3.2)	0.3 (2.0)	0.41 (2.6)	.08[e]

[a]*P* values for the comparison of the intervention and control groups.

[b]Chi-square test.

[c]*t* test.

[d]ICQ-10: Interpersonal Competence Questionnaire-10.

[e]Mann-Whitney *U* test.

[f]AUDIT-C: consumption items of the Alcohol Use Disorders Identification Test.

Program Use and Engagement

During the intervention program, which lasted for 4 months (17 wk), 9 (2.2%) of the 417 program participants in the intervention group and 17 (3.6%) of the 473 participants of the control group withdrew their participation (odds ratio 0.50, 95% CI 0.08 - 3.15; *P*=.46). Over the 4-month program, a total of 18 SMS text messaging–based reply activities (eg, replies to quiz questions or selection queries) and 20 media view activities (eg, views of video clips or images) were prompted. Considering

XSL•FO
RenderX

the total number of 38 possible activities, the mean number of activities carried out by participants was 12.6 (SD 12.8) in the intervention group and 12.5 (SD 12.7) in the control group (unstandardized regression coefficient from the linear mixed model b=−1.72, 95% CI −4.22 to 0.79; P=.18). Considering the number of SMS text messaging replies and media view activities separately, program engagement also did not differ significantly between the study groups: a mean number of 7.2 (SD 6.8) SMS text message reply activities were completed in the intervention group and 7.3 (SD 6.8) in the control group (b=−1.12, 95% CI −2.43 to 0.18; P=.10), while a mean of 5.4 (SD 6.6) media view activities were completed in the intervention group and 5.2 (SD 6.6) in the control group (b=0.18, 95% CI −0.92 to 1.28; P=.74).

Efficacy Based on 6-Month Follow-Up

The results of the complete-case (CC) and ITT analyses examining outcomes at the 6-month follow-up are presented in Tables 3 and 4. For count variables (Table 3), no significant group differences were observed for tobacco use days (ITT: rate ratio=0.75; P=.45) or cannabis use days (ITT: rate ratio=0.86; P=.67) in either the CC or ITT analyses.

Continuous outcomes (Table 4) including AUDIT-C score (ITT: P=0.91), perceived stress (ITT: P=0.77), and social skills (ITT: P=0.45) also did not differ significantly between groups in either CC or ITT analyses.

Table . Intervention effects for count outcomes.

Outcome	Intervention group (n=417)			Control group (n=473)				*P* value
	Baseline	Follow-up	Difference	Baseline	Follow-up	Difference	Rate ratio (95% CI)	
Complete-case analysis								
Nicotine or tobacco use days in preceding 30 d, mean (SD)	3.0 (8.0)	2.1 (6.8)	−0.9	2.4 (7.1)	2.4 (7.2)	0.0	0.66 (0.31-1.43)	.29
Cannabis use days, preceding 30 d, mean (SD)	0.5 (3.2)	0.2 (1.4)	−0.3	0.3 (2.0)	0.5 (3.1)	0.2	0.41 (0.11-1.47)	.17
Intention-to-treat analysis								
Nicotine or tobacco use days in preceding 30 d, mean (SD)	3.0 (8.0)	4.0 (8.7)	1.0	2.4 (7.1)	4.6 (9)	2.2	0.75 (0.36-1.57)	.45
Cannabis use days, preceding 30 d, mean (SD)	0.5 (3.2)	1.7 (4.5)	1.2	0.3 (2.0)	2.2 (5.2)	1.9	0.86 (0.43-1.73)	.67

XSL•FO
RenderX

Table . Intervention effects for continuous outcomes.

Outcome	Intervention group (n=417)			Control group (n=473)			Unstandardized regression coefficient *b* (95% CI)	*P* value
	Baseline	Follow-up	Difference	Baseline	Follow-up	Difference		
Complete-case analysis								
AUDIT-C[a] score, mean (SD)	1.2 (2.4)	1.0 (1.9)	−0.2	1.1 (2.2)	0.8 (2.1)	−0.3	0.032 (−0.338 to 0.402)	.87
Perceived stress past 30 d, mean (SD)	3.2 (1.0)	2.8 (1.0)	−0.4	3.4 (1.0)	2.8 (1.0)	−0.6	−0.013 (−0.186 to 0.160)	.89
Social skills (ICQ-10[b]), mean (SD)	28.6 (4.8)	30.9 (4.4)	2.3	28.3 (4.7)	30.8 (4.6)	2.5	−0.101 (−1.285 to 1.083)	.87
Intention-to-treat analysis								
AUDIT-C score, mean (SD)	1.2 (2.4)	1.5 (2.4)	0.3	1.1 (2.2)	1.6 (2.6)	0.5	0.030 (−0.473 to 0.533)	.91
Perceived stress past 30 d, mean (SD)	3.2 (1.0)	2.7 (1.0)	−0.5	3.4 (1.0)	2.8 (1.0)	−0.6	−0.034 (−0.262 to 0.195)	.77
Social skills (ICQ-10), mean (SD)	28.6 (4.8)	30.9 (4.4)	2.3	28.3 (4.7)	30.8 (4.6)	2.5	−0.525 (−1.881 to −0.832)	.45

[a]AUDIT-C: consumption items of the Alcohol Use Disorders Identification Test.

[b]ICQ-10: Interpersonal Competence Questionnaire-10.

Moderator Analyses

No significant demographic or socioeconomic variables moderated the effect of the intervention on the primary or secondary outcomes.

Discussion

Principal Results

This study evaluated program engagement and the efficacy of an optimized mobile phone–based life-skills training program with advanced tailoring of intervention content, compared to the original program version among adolescents. The results revealed no significant differences between participants of the 2 program versions with respect to the considered indicators of program engagement and efficacy.

The proactive invitation to participate in the program and study in Swiss secondary school classes, combined with the offer of a low-threshold mobile phone–based intervention, enabled us to reach 3 out of 4 adolescents for participation in the *SmartCoach* program and the associated study. Given that the program lasted 17 weeks and that participants had to provide their mobile phone numbers, the high participation rate of 75.7% (890/1175) is remarkable though consistent with previous studies of the program in this target group [16,29].

Concerning program attrition, the overall results were positive, with only 2.2% (9/417) of the participants in the intervention group and 3.6% (17/473) of the control group withdrawing their participation. Relating to program use, the average number of completed activities was 13 out of 38 possible activities in both study groups, with no significant between-group differences either in this total number or in the more specific activities of SMS text message replies or media views. The moderator analyses also showed no differences in the program participation variables. Therefore, the initial hypothesis that the program optimizations or the advanced tailoring of program content would result in increased program use has to be rejected. Similarly, no statistically significant group differences were observed for the primary and secondary outcomes reflecting program efficacy.

Despite previous evidence supporting the effectiveness of *SmartCoach*, the optimized version did not demonstrate significant improvements over the original program. One critical factor that may have contributed to these nonsignificant results is the lack of statistical power, which might be related to the generally low substance use prevalence rates in this age group that may not have been sufficient to detect the relatively small differences between the optimized and original program versions. In relation to the overall program content and processes, only a small proportion of the content (maximum 8%, depending on the topic choice during the assessment) was changed. This is also due to the fact that, in the qualitative program evaluations relating to the Capability, Opportunity, and Motivation model of behavior change and theoretical domains framework, most young people were very satisfied with the original program and its content, and there were

XSL•FO
RenderX

relatively few requests for changes [32]. At the same time, it must be mentioned that this previous study did not reach the entire sample for the follow-up interviews, and some decisions to change program content were based on relatively small numbers. Furthermore, these specific adaptations might not have addressed the participants' real challenges regarding program engagement or may even have had a negative impact (eg, reminders might be perceived as annoying and evoke reactance; messages on a different day of the week might be helpful for some but could also be counterproductive for another subgroup). Future studies in which optimization suggestions are assessed should therefore collect them in two steps: first, optimization suggestions should be identified qualitatively, and in a second step, an independent sample should be used to check quantitatively whether these suggestions are desired by the majority or whether they are only the preferences of a minority. Another reason for the zero effects may be that the assumed effects of advanced tailoring were overly optimistic. One reason for this is that the 3 project packages (identification of subgroups with low program engagement, qualitative exploration and implementation of program optimizations, and this randomized controlled trial to test the effectiveness of the optimized program) were planned together in advance, including power calculations. Initially, we assumed that, in addition to the variable "parental background from non–German-speaking countries," other variables such as health literacy, socioeconomic status, and gender might influence program engagement. However, it turned out that they did not have a substantial effect in this initial study, and both the number of variables that could be used for improved tailoring and the necessary scope of program optimizations were rather small [29,32].

Limitations

Main limitations of this study are that (1) all primary outcome data rely on self-report, with the associated possibility that results may have been influenced by social desirability and potential recall bias. Measures used to avoid under- or overreporting of substance use included assurances of confidentiality and anonymous assessments conducted via online surveys without personal contact, which may have increased the reliability of self-reported data. (2) Cluster randomization according to school class did not result in a balancing of all baseline characteristics, and there was selective attrition according to gender, cannabis use prevalence, and perceived stress. Although we attempted to counteract this attrition bias as effectively as possible through multiple imputation, ITT analyses, and additional complete case analyses for sensitivity testing, a certain degree of bias and its effect on the results cannot be completely ruled out. (3) The results could not be generalized to secondary and upper secondary schools in Switzerland, as we recruited a convenience sample of school classes willing to participate in the study. However, the comparison of the substance use prevalence rates among a representative sample of 15-year-old students in Switzerland [1] and the baseline characteristics of the study sample does not reveal major deviations. The 30-day point prevalence rates for nicotine use were 20% in this study and 16% for cigarette consumption, with an additional 3% for e-cigarette use in the representative survey. For cannabis use, the 30-day point prevalence was 6% in this study and 9% in the national survey.

Conclusions and Outlook

This study is among the first to evaluate the impact of a mobile phone–based life-skills training program for adolescent substance use prevention, which integrates participatory tailoring based on user feedback within a randomized controlled trial setting. The results demonstrate that the program was feasible and acceptable, as indicated by a high participation rate and relatively low program attrition. However, the optimized version did not show improved engagement or efficacy compared to the original version.

Despite the lack of superiority of the optimized program version, the *SmartCoach* program offers a low-threshold and effective intervention program to prevent substance use among adolescents [16]. The SMS text messaging–based version is currently being revised as part of a European Union–funded project to create an instant messaging version using WhatsApp (Meta Platforms). This offers significant advantages over SMS, including the ability to send media such as pictures and video clips directly, as well as the option to send messages free of charge using an existing internet connection [46]. Furthermore, this instant messaging–based coaching program will integrate storytelling elements to provide meaning to the facts presented and allow participants to connect with the information both intellectually and emotionally [47]. For example, in the new program version, participants will accompany a Stone Age man or woman (Stoney) on a journey through a picture story to discover the origins of stress and learn how to cope with stressful situations. These situations will be based on those that participants provided information about during the baseline assessment. Compared to the revisions implemented in the context of this study, these more substantial program revisions could lead to better retention and engagement, ultimately improving the efficacy of the *SmartCoach* life-skills training program.

Acknowledgments

The authors thank Andreas Filler and his team from Pathmate Technologies for setting up the server and the intervention program. The authors declare the use of generative artificial intelligence (GAI) in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GAI tools under full human supervision: proofreading, editing, and summarizing the text. The GAI tool used was ChatGPT (OpenAI). Responsibility for the final manuscript lies entirely with the authors. GAI tools are not listed as authors and do not bear responsibility for the final outcomes.

XSL•FO
RenderX

Funding

Funding for this project was provided by the Swiss National Science Foundation (grant 10001C_204412/1). The funding institution had no role in the design or conduct of the study; the management, analysis, or interpretation of data; or the preparation, review, or approval of the manuscript.

Data Availability

The datasets generated and analyzed during this study will be made available on SWISSUbase [48].

Authors' Contributions

SH and MPS were responsible for funding acquisition and conceptualization of the study. SH, NK, and OS were involved in the development of the intervention program. SH, AW, OS, and NK were responsible for data collection, data curation, and validation. AW and NB were responsible for data analysis. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

SH, NK, and OS were also involved in the development of the intervention program *SmartCoach* and its optimized program version.

Checklist 1
CONSORT-eHEALTH checklist.
[PDF File, 10236 KB - mhealth_v14i1e78081_app1.pdf]

References

1. Badura P, Eriksson C, García-Moya I, et al. A focus on adolescent social contexts in Europe, central Asia and Canada: Health Behaviour in School-Aged Children international report from the 2021/2022 survey. Volume 7. : World Health Organization Regional Office for Europe; 2024 URL: https://iris.who.int/server/api/core/bitstreams/76336dac-ba67-4a8f-90ba-621d044cc1c6/content [accessed 2026-06-20]
2. Costello EJ, Copeland W, Angold A. Trends in psychopathology across the adolescent years: what changes when children become adolescents, and when adolescents become adults? J Child Psychol Psychiatry 2011 Oct;52(10):1015-1025. [doi: 10.1111/j.1469-7610.2011.02446.x] [Medline: 21815892]
3. Hall WD, Patton G, Stockings E, et al. Why young people's substance use matters for global health. Lancet Psychiatry 2016 Mar;3(3):265-279. [doi: 10.1016/S2215-0366(16)00013-4] [Medline: 26905482]
4. Haug S, Ulbricht S, Hanke M, Meyer C, John U. Overestimation of drinking norms and its association with alcohol consumption in apprentices. Alcohol Alcohol 2011;46(2):204-209. [doi: 10.1093/alcalc/agq103] [Medline: 21262761]
5. Pischke CR, Helmer SM, McAlaney J, et al. Normative misperceptions of tobacco use among university students in seven European countries: baseline findings of the 'Social Norms Intervention for the prevention of Polydrug usE' study. Addict Behav 2015 Dec;51:158-164. [doi: 10.1016/j.addbeh.2015.07.012] [Medline: 26275842]
6. Foxcroft DR, Tsertsvadze A. Universal school–based prevention programs for alcohol misuse in young people. Cochrane Database Syst Rev 2011 May 11(5):CD009113. [doi: 10.1002/14651858.CD009113] [Medline: 21563171]
7. Stockings E, Hall WD, Lynskey M, et al. Prevention, early intervention, harm reduction, and treatment of substance use in young people. Lancet Psychiatry 2016 Mar;3(3):280-296. [doi: 10.1016/S2215-0366(16)00002-X] [Medline: 26905481]
8. Thomas RE, McLellan J, Perera R. School-based programmes for preventing smoking. Cochrane Database Syst Rev 2013 Apr 30;2013(4):CD001293. [doi: 10.1002/14651858.CD001293.pub3] [Medline: 23633306]
9. Botvin GJ, Griffin KW. Life skills training: preventing substance misuse by enhancing individual and social competence. New Dir Youth Dev 2014;2014(141):57-65. [doi: 10.1002/yd.20086] [Medline: 24753278]
10. Ellickson PL, McCaffrey DF, Ghosh-Dastidar B, Longshore DL. New inroads in preventing adolescent drug use: results from a large-scale trial of project ALERT in middle schools. Am J Public Health 2003 Nov;93(11):1830-1836. [doi: 10.2105/ajph.93.11.1830] [Medline: 14600049]
11. Weichold K, Blumenthal A. Long-term effects of the life skills program IPSY on substance use: results of a 4.5-year longitudinal study. Prev Sci 2016 Jan;17(1):13-23. [doi: 10.1007/s11121-015-0576-5] [Medline: 26202801]
12. Griffin KW, Williams C, Botvin CM, Sousa S, Botvin GJ. Effectiveness of a hybrid digital substance abuse prevention approach combining e-learning and in-person class sessions. Front Digit Health 2022;4:931276. [doi: 10.3389/fdgth.2022.931276] [Medline: 35990017]
13. Ellickson PL. You've shown the program model is effective. Now what? In: Weichold K, Giannotta F, editors. Theory-Based Approaches to Substance Misuse and Abuse Prevention in School: New Directions for Youth Development, Number 141: Jossey-Bass; 2014:95-106. [doi: 10.1002/yd.20089]

XSL•FO
RenderX

14. Domitrovich CE, Durlak JA, Staley KC, Weissberg RP. Social-emotional competence: an essential factor for promoting positive adjustment and reducing risk in school children. Child Dev 2017 Mar;88(2):408-416. [doi: 10.1111/cdev.12739] [Medline: 28213889]
15. Haug S, Paz Castro R, Wenger A, Schaub MP. Efficacy of a mobile phone–based life-skills training program for substance use prevention among adolescents: study protocol of a cluster-randomised controlled trial. BMC Public Health 2018 Sep 10;18(1):1102. [doi: 10.1186/s12889-018-5969-5] [Medline: 30200928]
16. Paz Castro R, Haug S, Wenger A, Schaub MP. Longer-term efficacy of a digital life-skills training for substance use prevention. Am J Prev Med 2022 Dec;63(6):944-953. [doi: 10.1016/j.amepre.2022.06.017] [Medline: 35985899]
17. Garrido S, Millington C, Cheers D, et al. What works and what doesn't work? A systematic review of digital mental health interventions for depression and anxiety in young people. Front Psychiatry 2019;10:759. [doi: 10.3389/fpsyt.2019.00759] [Medline: 31798468]
18. MacDougall S, Jerrott S, Clark S, Campbell LA, Murphy A, Wozney L. Text message interventions in adolescent mental health and addiction services: scoping review. JMIR Ment Health 2021 Jan 8;8(1):e16508. [doi: 10.2196/16508] [Medline: 33416504]
19. Kazemi DM, Borsari B, Levine MJ, Li S, Lamberson KA, Matta LA. A systematic review of the mHealth interventions to prevent alcohol and substance abuse. J Health Commun 2017 May;22(5):413-432. [doi: 10.1080/10810730.2017.1303556] [Medline: 28394729]
20. Achilles MR, Anderson M, Li SH, Subotic-Kerry M, Parker B, O'Dea B. Adherence to e-mental health among youth: considerations for intervention development and research design. Digit Health 2020;6:2055207620926064. [doi: 10.1177/2055207620926064] [Medline: 32547775]
21. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
22. Paz Castro R, Haug S, Debelak R, Jakob R, Kowatsch T, Schaub MP. Engagement with a mobile phone–based life skills intervention for adolescents and its association with participant characteristics and outcomes: tree-based analysis. J Med Internet Res 2022 Jan 19;24(1):e28638. [doi: 10.2196/28638] [Medline: 35044309]
23. Paz Castro R, Haug S, Filler A, Kowatsch T, Schaub MP. Engagement within a mobile phone–based smoking cessation intervention for adolescents and its association with participant characteristics and outcomes. J Med Internet Res 2017 Nov 1;19(11):e356. [doi: 10.2196/jmir.7928] [Medline: 29092811]
24. Jakob R, Harperink S, Rudolf AM, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371. [doi: 10.2196/35371] [Medline: 35612886]
25. Perski O, Hekler EB, Naughton F, Brown J. Can just-in-time adaptive interventions offer a promising solution to the "engagement crisis" in digital behaviour change interventions? Ann Behav Med 2020 May;54 Supplement 1:S191-S191.
26. Head KJ, Noar SM, Iannarino NT, Grant Harrington N. Efficacy of text messaging–based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
27. Bernaerts S, Van Daele T, Carlsen CK, Nielsen SL, Schaap J, Roke Y. User involvement in digital mental health: approaches, potential and the need for guidelines. Front Digit Health 2024;6:1440660. [doi: 10.3389/fdgth.2024.1440660] [Medline: 39238496]
28. Doerwald F, Stalling I, Recke C, et al. A rapid review of digital approaches for the participatory development of health-related interventions. Front Public Health 2024;12:1461422. [doi: 10.3389/fpubh.2024.1461422] [Medline: 39678234]
29. Haug S, Boumparis N, Wenger A, Schaub MP, Kiselev N. Predictors of youth accessibility for a mobile phone–based life skills training program for addiction prevention. Int J Environ Res Public Health 2023 Jul 17;20(14):6379. [doi: 10.3390/ijerph20146379] [Medline: 37510611]
30. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42. [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
31. Atkins L, Francis J, Islam R, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implement Sci 2017 Jun 21;12(1):77. [doi: 10.1186/s13012-017-0605-9] [Medline: 28637486]
32. Kiselev N, Schaffner R, Csepregi Z, Wenger A, Paz Castro R, Haug S. Factors influencing engagement in a digital substance use prevention program: qualitative study using the capability, opportunity, motivation model and theoretical domains framework. JMIR mHealth uHealth 2025 Aug 14;13:e64861. [doi: 10.2196/64861] [Medline: 40811785]
33. SmartCoach. URL: https://www.smartcoach.info/ [accessed 2025-05-22]
34. Guidelines for ethics applications to the Ethics Committee of the Faculty of Arts and Social Sciences. : University of Zurich; 2025 URL: https://www.phil.uzh.ch/dam/jcr:88de353c-a966-4c69-9b3e-a897f91cb503/20250924_Guidelines%20for%20Ethics%20Applications%20to%20the%20Ethics%20Committee%20of%20the%20Faculty%20of%20Arts%20and%20Social%20Sciences.pdf [accessed 2026-06-20]
35. Pocock SJ. Clinical Trials: A Practical Approach: Wiley & Sons; 1996.
36. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]

XSL•FO
RenderX

37. McAlister AL, Perry CL, Parcel GS. How individuals, environments, and health behaviour interact: social cognitive theory. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior and Health Education: Theory, Research, and Practice, 4th edition: Jossey-Bass; 2008:169-188.
38. Goodman E, Adler NE, Kawachi I, Frazier AL, Huang B, Colditz GA. Adolescents' perceptions of social status: development and evaluation of a new indicator. Pediatrics 2001 Aug;108(2):E31. [doi: 10.1542/peds.108.2.e31] [Medline: 11483841]
39. Goodman E, Huang B, Schafer-Kalkhoff T, Adler NE. Perceived socioeconomic status: a new type of identity that influences adolescents' self-rated health. J Adolesc Health 2007 Nov;41(5):479-487. [doi: 10.1016/j.jadohealth.2007.05.020] [Medline: 17950168]
40. Lampert T, Hoebel J, Kuntz B, Müters S, Kroll LE. Messung des sozioökonomischen Status und des subjektiven sozialen Status in KiGGS Welle 2 [Article in German]. J Health Monit 2018;3(1):114-133. [doi: 10.17886/RKI-GBE-2018-016]
41. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. Arch Intern Med 1998 Sep 14;158(16):1789-1795. [doi: 10.1001/archinte.158.16.1789] [Medline: 9738608]
42. Rumpf HJ, Wohlert T, Freyer-Adam J, Grothues J, Bischof G. Screening questionnaires for problem drinking in adolescents: performance of AUDIT, AUDIT-C, CRAFFT and POSIT. Eur Addict Res 2013;19(3):121-127. [doi: 10.1159/000342331] [Medline: 23183686]
43. Juvenir Studie 4.0. Zuviel Stress – zuviel Druck! Wie Schweizer Jugendliche mit Stress und Leistungsdruck umgehen [Report in German]. : Jacobs Foundation; 2015 URL: https://jacobsfoundation.org/wp-content/uploads/2017/07/Juvenir-4.0_Langfassung_DE_final.pdf [accessed 2026-06-20]
44. Coroiu A, Meyer A, Gomez-Garibello CA, Brähler E, Hessel A, Körner A. Brief form of the Interpersonal Competence Questionnaire (ICQ-15): development and preliminary validation with a German population sample. Eur J Psychol Assess 2015;31(4):272-279. [doi: 10.1027/1015-5759/a000234]
45. van Buuren S. Flexible Imputation of Missing Data: Chapman & Hall/CRC; 2012.
46. Haug S, Wenger A, Arana-Arri E, et al. Efficacy of an instant messaging–based life-skills training program to prevent behavioral risk factors among adolescents in Europe: study protocol of a cluster-randomized controlled trial. BMC Psychol 2025 Nov 24;13(1):1292. [doi: 10.1186/s40359-025-03592-1] [Medline: 41286988]
47. Brooks SP, Zimmermann GL, Lang M, et al. A framework to guide storytelling as a knowledge translation intervention for health-promoting behaviour change. Implement Sci Commun 2022 Mar 28;3(1):35. [doi: 10.1186/s43058-022-00282-6] [Medline: 35346397]
48. Efficacy of an optimized mobile phone–based life skills training program for addiction prevention among adolescents: results of a cluster-randomized controlled trial. SWISSUbase. URL: https://www.swissubase.ch/en/catalogue/studies/21038/20793/overview [accessed 2026-06-20]

Abbreviations

AUDIT-C: consumption items of the Alcohol Use Disorders Identification Test
CC: complete case
CONSORT: Consolidated Standards of Reporting Trials
GLMM: generalized linear mixed model
ICC: intracluster correlation coefficient
ITT: intention-to-treat
mHealth: mobile health

Edited by A Stone, L Buis; submitted 26.May.2025; peer-reviewed by A Minutillo, S Comulada; revised version received 22.May.2026; accepted 26.May.2026; published 15.Jul.2026.

Please cite as:
Haug S, Boumparis N, Studhalter O, Wenger A, Schaub MP, Kiselev N
Efficacy of a Mobile Phone–Based Life-Skills Training Program for Addiction Prevention Optimized Among Adolescents With Low Program Engagement: Cluster-Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e78081
URL: https://mhealth.jmir.org/2026/1/e78081
doi:10.2196/78081

© Severin Haug, Nikolaos Boumparis, Olivia Studhalter, Andreas Wenger, Michael P Schaub, Nikolai Kiselev. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 15.Jul.2026. This is an open-access article distributed under

XSL•FO
RenderX

the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Characteristics of Tailored Text Messages Associated With Increased Physical Activity Among Cardiac Rehabilitation Enrollees: Secondary Analysis of a Microrandomized Trial

Namratha Atluri[1*], MD; Kashvi Gupta[2*], MPH, MD; Tanima Basu[3], MA, MS; Evan Luff[3], MS; Jieru Shi[4], PhD; Thomas Boyden[5], MD; Bhramar Mukherjee[6], PhD; Sachin Kheterpal[7], MBA, MD; Predrag Klasnja[8], PhD; Walter Dempsey[9], PhD; Brahmajee K Nallamothu[3], MPH, MD; Jessica R Golbus[3], MS, MD

[1]Department of Internal Medicine, University of Michigan, Ann Arbor, MI, United States

[2]Division of Nephrology, Department of Internal Medicine, University of California, San Francisco, San Francisco, CA, United States

[3]Division of Cardiovascular Diseases, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, United States

[4]Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge, United Kingdom

[5]Division of Cardiovascular Diseases, Department of Internal Medicine, Corewell Health, Grand Rapids, MI, United States

[6]Department of Biostatistics, Yale School of Public Health, New Haven, CT, United States

[7]Department of Anesthesiology, University of Michigan, Ann Arbor, MI, United States

[8]School of Information, University of Michigan, Ann Arbor, MI, United States

[9]Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, United States

[*]these authors contributed equally

Corresponding Author:
Namratha Atluri, MD
Department of Internal Medicine
University of Michigan
1500 E Medical Center Dr
Ann Arbor, MI, 48109
United States
Phone: 1 734 764 1817
Email: atlurina@med.umich.edu

Abstract

Background: Emerging data suggest that text message–based mobile health interventions may enhance physical activity levels in patients with cardiovascular disease enrolled in cardiac rehabilitation. The optimal characteristics of texts that lead to maximal patient engagement and drive meaningful behavioral change are not well understood.

Objective: This study aimed to understand how text- and participant-level characteristics impact physical activity levels after text delivery.

Methods: The VALENTINE (Virtual Application-Supported Environment to Increase Exercise) study was a randomized controlled trial designed to evaluate a mobile health intervention delivered to low- and moderate-risk adults enrolled in cardiac rehabilitation. Embedded within this study was a microrandomized trial focused on the effect of texts on physical activity levels among intervention participants. Participants in the intervention group received texts through a smartwatch (Apple Watch or Fitbit Versa) that were tailored to the time of day, day of the week (weekday vs weekend), weather, and time since enrollment in cardiac rehabilitation. Texts also differed in content type (walking vs antisedentary) and in the level of personalization (inclusion of the participant's name or not). Delivery was randomized at 4 user-selected time points daily, with participants having a 25% probability of receiving a text at any time point. The primary outcome was step count 60 minutes after a decision point. This analysis focuses on the text- and participant-level factors that moderated the intervention's effect on the primary outcome. Given potential measurement differences determined a priori, analyses were stratified by device type and phase of cardiac rehabilitation and adjusted for age, sex, and baseline activity status using a generalization of regression analysis.

Results: More than 70,552 randomizations occurred in 108 participants (mean age 59.5, SD 10.7 years; n=36, 33.3% female; n=19, 17.6% non-White; n=68, 63% Apple Watch users) over 6 months. Overall, no text characteristics (including personalization with the participant's name) or participant characteristics (including baseline physical activity) consistently impacted text

XSL•FO
RenderX

responsiveness for either device type. Although the findings were not consistently significant between device types and across phases of the trial, there was a trend toward increased responsiveness to texts that promoted walking (compared to antisedentary texts) and that were delivered to younger (aged <65 years) and male participants.

Conclusions: In this randomized clinical trial, we found that tailored texts improved physical activity levels among cardiac rehabilitation enrollees in the initiation phase, but this effect was not explained by text- or participant-level moderators. Additional work is needed to explore the impact of tailoring based on an extended set of personal and environmental factors to optimize the delivery and efficacy of text message–based interventions.

Trial Registration: ClinicalTrials.gov NCT04587882; https://clinicaltrials.gov/study/NCT04587882

International Registered Report Identifier (IRRID): RR2-10.1016/j.ahj.2022.02.012

(JMIR Mhealth Uhealth 2026;14:e79792) doi:10.2196/79792

KEYWORDS

cardiac rehabilitation; mobile health; mHealth; physical activity; text messages; tailoring; personalization

Introduction

Center-based cardiac rehabilitation (CR) is an evidence-based lifestyle program that improves functional capacity and quality of life and reduces morbidity and mortality in patients with cardiovascular disease (CVD) [1-5]. Despite these well-established benefits, participation in CR remains low, and physical activity levels often decline over time after program completion [3-5]. Emerging data suggest that mobile health (mHealth) interventions may be effective in extending the benefits of center-based CR beyond the supervised setting [6-10]. mHealth interventions encompass a range of technologies, including text messages, mobile phone apps, and transmission of physiological data (eg, heart rate, step count, and blood pressure) from sensors to mobile devices and web-based platforms [9]. Among these, texts are the most widely accessible. To drive meaningful and sustainable long-term behavioral change, optimizing key text features is crucial. Through strategic timing and frequency of delivery, personalized content, and engaging user interfaces, tailored text-based interventions can promote patient engagement and usability while preventing habituation to improve clinical outcomes [9,10].

The VALENTINE (Virtual Application-Supported Environment to Increase Exercise) study evaluated the impact of an mHealth intervention combining texts, a mobile app, and a smartwatch (Apple Watch or Fitbit Versa) on physical activity levels in enrollees undergoing CR over a 6-month period [11-13]. Nested within the VALENTINE study was a microrandomized trial designed to study the effect of texts on physical activity levels throughout the day. Microrandomized trials are a novel experimental design in which interventions such as texts are repeatedly randomized for delivery at different time points throughout the day. This powerful design enables assessment of the causal effects of individual intervention components on near-term outcomes, such as step count following text delivery, which are thought to influence longer-term behavioral or health outcomes, including sustained physical activity. The primary analysis of the VALENTINE study's microrandomized trial showed that the 60-minute step count increased early after text delivery for both Apple Watch and Fitbit users, with differential results by device type thereafter [12]. Understanding which text characteristics and contextual factors promoted this increase in physical activity, at least initially, is necessary to develop more effective text interventions.

Herein, we present a prespecified secondary analysis from the VALENTINE study's microrandomized trial, evaluating the impact of text tailoring on physical activity levels. The trial leveraged real-time data from participants' smartwatches about their current environment to tailor texts with the goal of promoting low-level physical activity. In this analysis, the primary objective was to evaluate the impact of both participant-level factors and text characteristics hypothesized to moderate intervention efficacy. We hypothesized that the step count in the 60 minutes following text delivery would differ significantly based on text characteristics, such as time of delivery, personalization with the participant's name, and level of activity prior to text delivery, as well as participant-level factors such as age, sex, and baseline activity status.

Methods

Overview

The VALENTINE study was a prospective, multicenter randomized clinical trial that remotely delivered an mHealth intervention to low- and moderate-risk patients enrolled in center-based CR (ClinicalTrials.gov NCT04587882). The primary objectives of the overall trial were (1) to evaluate the efficacy of a multicomponent mHealth intervention for augmenting and extending the benefits of center-based CR and (2) to evaluate the impact of text message delivery specifically on physical activity levels. The methodological details of the study have been previously described [11-13]. Briefly, adults aged ≤75 years who had completed 2 CR visits based on a qualifying diagnosis, owned a compatible smartphone, and spoke English were randomized to the intervention or control arm of the study. Patients who were unsafe to exercise without supervision due to a high-risk cardiovascular condition were excluded.

Ethical Considerations

This study was approved by the institutional review board of the University of Michigan (HUM00162365). Informed consent was obtained from all participants before randomization. Efforts were made to preserve patient privacy whenever possible, including instructions on removing the mobile app from their

XSL•FO
RenderX

phones at the end of the study to prevent passive data collection. Other than wearable devices, no other compensation was provided for study participation.

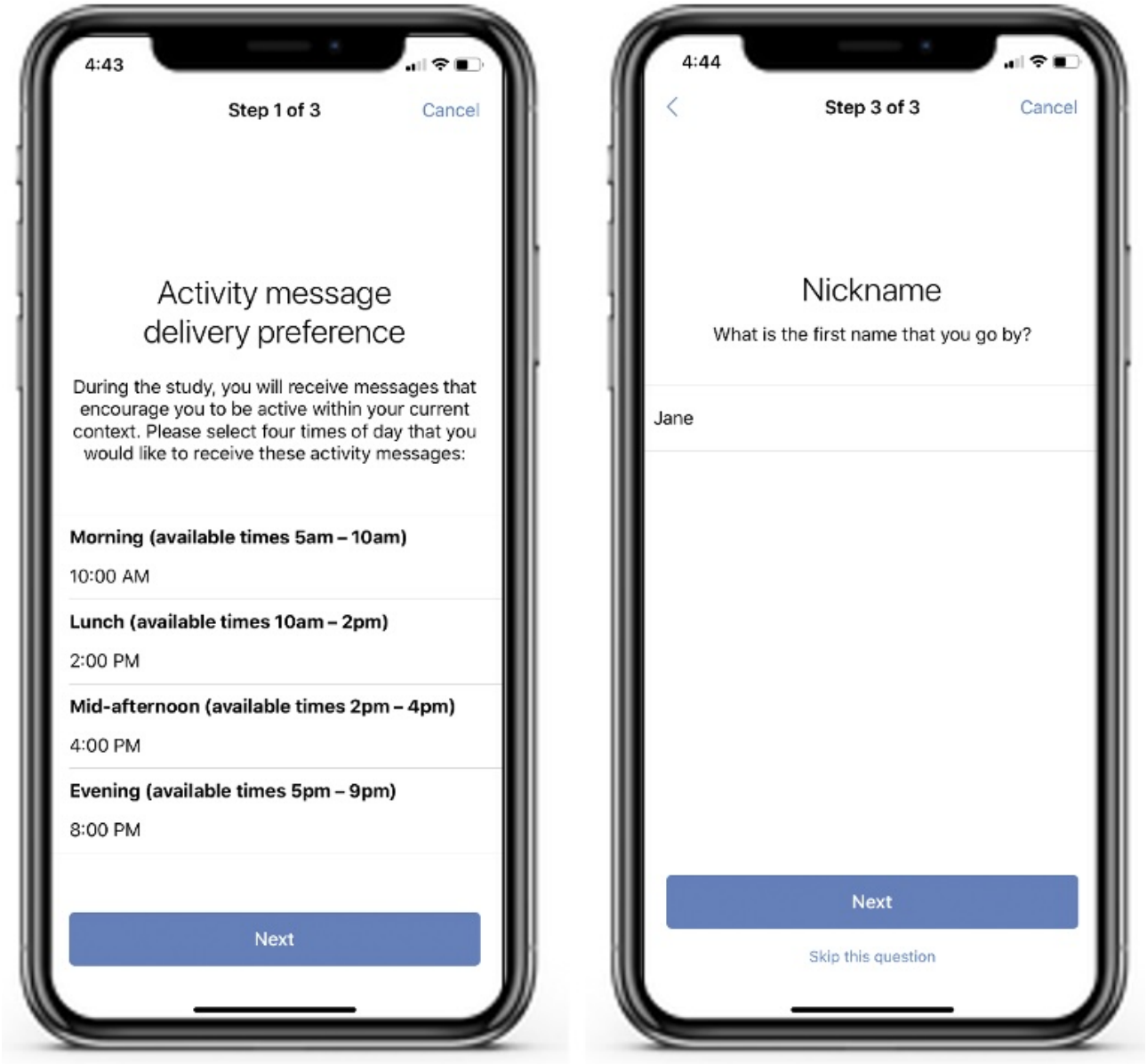
Study Procedures

The VALENTINE study launched in October 2020 at an academic medical center and in July 2021 at a large community health care center in Michigan. Both the intervention and control groups underwent remote consent and enrollment visits and received a wearable device compatible with their smartphone (Apple Watch Series 4; Apple Inc or Fitbit Versa 2; Fitbit Inc) to track daily physical activity and receive texts. Additionally, participants in both groups participated in center-based CR concurrently.

Description of the Intervention

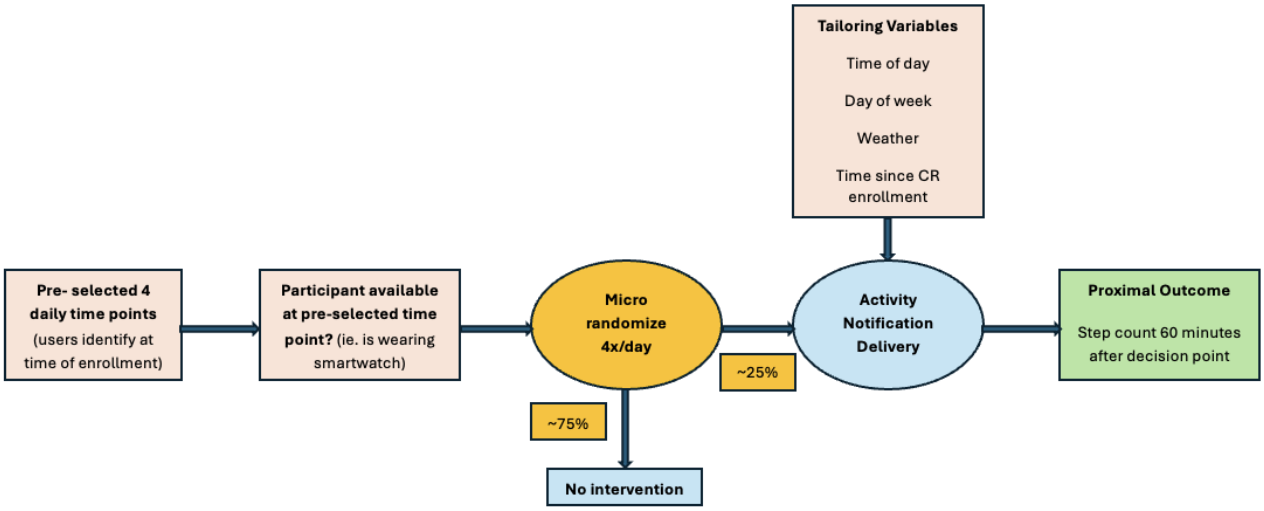
The VALENTINE study was designed to augment and extend the benefits of CR through an mHealth intervention. The intervention consisted of a mobile app that allowed activity tracking, goal setting, and microrandomized texts (Figure 1). The texts were delivered by SMS to participants' smartwatches and phones but could also be reviewed in the mobile study app. Texts were tailored to the time of day (morning, lunchtime, afternoon, or evening), day of the week (weekday vs weekend), weather (ie, indoor-based texts delivered if snow or rain was predicted), and time since enrollment in CR (corresponding to the initiation, 0-30 days; maintenance, 31-120 days; and completion, 121-182 days phases of CR). Participants were randomized to receive a text or no text at 4 user-selected time points (ie, decision points), with texts selected at random based on participants' environments (Figure 2). Participants had a 25% probability of receiving a text at each time point.

Figure 1. Screenshots from the VALENTINE (Virtual Application-Supported Environment to Increase Exercise) mobile app showing participant selection of preferred text messages delivery times and entry of a nickname for text personalization during enrollment.



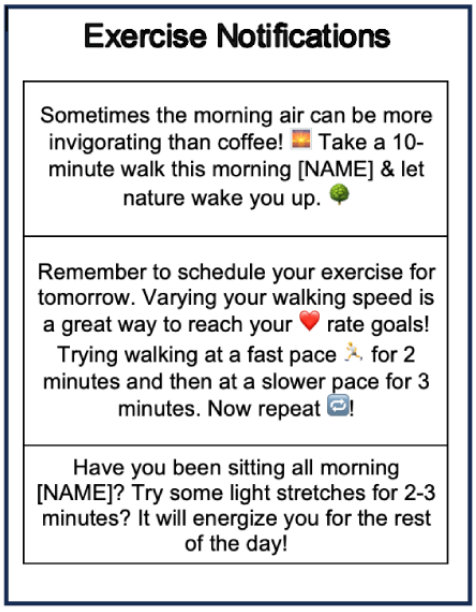
XSL•FO
RenderX

Figure 2. Microrandomized trial design diagram. CR: cardiac rehabilitation.



This secondary analysis focuses on the results of activity texts. Activity texts were designed to encourage low levels of physical activity in the hour after delivery and were written to be low intensity, short duration, and actionable in real time, as opposed to exercise, which is a planned behavior and often longer in duration and higher in intensity (Figure 3). There were a total of 284 possible activity texts that varied in recommended action (encouraging a short walk vs standing up and stretching) and framing (emphasizing the benefits of being active vs highlighting missed opportunities to meet goals). Approximately 50% of the activity texts were personalized with participants' preferred names, and approximately 30% included emojis or hyperlinks to the mobile app. The effectiveness of texts was assessed through proximal outcomes, which refer to the near-term effects of the intervention. The primary proximal outcome for activity texts was the step count 60 minutes following a decision point.

Figure 3. Example of activity text messages used in the VALENTINE (Virtual Application-Supported Environment to Increase Exercise) study.



Statistical Analysis

Baseline characteristics are described as mean (SD) for continuous variables and counts (percentage) for categorical variables. Student *t* tests (2-tailed) or chi-square tests were used for bivariate comparisons of continuous and categorical variables, respectively.

We decided a priori to analyze the data by device type (Apple Watch vs Fitbit) due to known differences in how each device measures physical activity [14]. The study was divided into initiation (days 1-30), maintenance (days 31-120), and completion (days 121-182) phases to account for participants' progression through CR. On the basis of the primary analysis results, which demonstrated an early effect of texts for both device types, this analysis focused on the overall and initiation phases of the study [12].

XSL•FO
RenderX

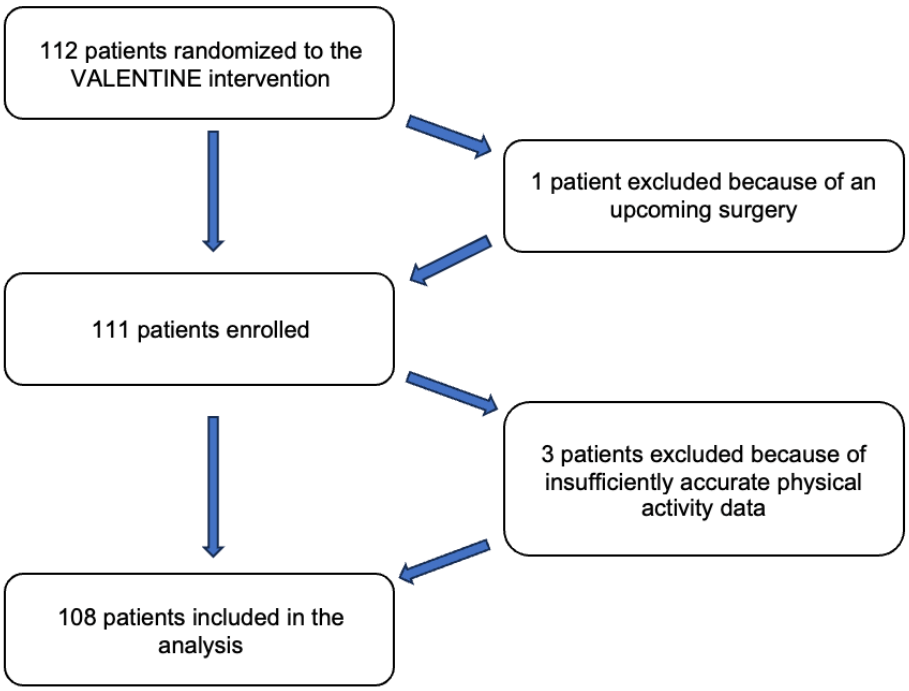
Our primary outcome for this analysis was the step count in the 60 minutes following a decision point. As the step count was positively skewed, 0.5 was added to all step counts before log transformation. All analyses were performed using a centered and weighted least squares method, a standard method to assess time-varying causal effects of a digital intervention in a microrandomized trial [15]. This method estimates parameters in models for treatment effects and allows the robust inclusion of covariates to reduce noise. Similar to the generalized estimating equations and multilevel models, this method accommodates the nested nature of the data and within-participant correlation across time and accounts for sequential randomization to estimate causal treatment effects. Additionally, regardless of sample size, this method yields consistent and asymptotically normal estimates of the marginal or moderated treatment effect of interest. We assessed causal effect moderation by the following covariates: age (<65 vs >65 years), baseline physical activity (>7000 vs <7000 steps in the first week of the study), sex, and step count 30 minutes before each decision point. Device-specific models evaluated the effect of delivering a text compared with no text at each decision point. At each decision point, participants were considered available to receive the intervention if they were wearing their smartwatches, which was operationalized by requiring at least 1 heart rate measurement on participants' smartwatches at least 30 minutes prior to a decision point. The interaction between treatment and a series of contextual factors (time of text delivery, weekend vs weekday, text content [antisedentary vs activity promoting], and personalization of the text using the participant's name), as well as participant characteristics (age category, sex, baseline activity status, presence of heart failure, and indication for CR referral), was explored. A *P* value <.05 was used to establish significance. All analyses were performed using SAS (version 14.2; SAS Institute Inc).

Results

Overview

From October 23, 2020, to March 25, 2022, 112 participants were randomized to the intervention arm of the study. Only 1 participant was excluded after randomization but before enrollment due to failure to meet the inclusion criteria (ie, upcoming surgery), and 3 had incomplete physical activity data for analysis (Figure 4). An additional 2 participants withdrew from the study after enrollment but allowed data collected up until their withdrawal date to be used. Of the 108 remaining participants, 36 (33.3%) were female, 89 (82.4%) were White, and 101 (93.5%) were non-Hispanic. Participants' mean age was 59.5 (SD 10.7) years, and their mean BMI was 30.4 (SD 6.5) kg/m^2; 68 (63%) participants owned an iPhone and were provided with an Apple Watch (Table 1). Throughout the study, these participants underwent 70,552 randomizations to receive or not receive a text over a median of 180.5 (IQR 172.5-182.0) days. Participants were available for 75.9% of decision points (77.3% Apple Watch vs 73.4% Fitbit), with availability defined as having at least 1 smartwatch heart rate measurement in the 30 minutes prior to randomization. On average, 0.91 (SD 0.86) texts per day were sent, although participants were only available to receive 0.80 texts per day.

Figure 4. CONSORT (Consolidated Standards of Reporting Trials) flow diagram. VALENTINE: Virtual Application-Supported Environment to Increase Exercise.



XSL•FO
RenderX

Table 1. Baseline characteristics of participants enrolled in the intervention arm of the VALENTINE (Virtual Application-Supported Environment to Increase Exercise) study.

Characteristics	Overall (N=108)	Fitbit (n=40)	Apple Watch (n=68)
Age (years), mean (SD)	59.5 (10.7)	58.3 (9.6)	59.6 (11.4)
Sex, n (%)			
Male	72 (66.7)	26 (65)	46 (67.6)
Female	36 (33.3)	14 (35)	22 (32.4)
Race, n (%)			
Asian	6 (5.6)	1 (2.5)	5 (7.4)
Black	4 (3.7)	4 (10)	0 (0)
Other or unknown	9 (8.3)	5 (12.5)	4 (5.9)
White	89 (82.4)	30 (75)	59 (86.8)
Ethnicity, n (%)			
Hispanic	2 (1.9)	1 (2.5)	1 (1.5)
Non-Hispanic	101 (93.5)	39 (97.5)	62 (91.2)
Unknown	5 (4.6)	0 (0)	5 (7.4)
Study site, n (%)			
Michigan Medicine	89 (82.4)	34 (85)	55 (80.9)
Spectrum Health	19 (17.6)	6 (15)	13 (19.1)
Indication for cardiac rehabilitation, n (%)			
PCI[a]	58 (20.4)	22 (55)	36 (52.9)
CABG[b]	19 (17.6)	9 (22.5)	10 (14.7)
Valve repair or replacement	22 (20.4)	7 (17.5)	15 (22.1)
PCI or CABG and valve repair or replacement	2 (1.9)	0 (0)	2 (2.9)
CAD[c] or acute coronary syndrome, not revascularized	7 (6.5)	2 (5)	5 (7.4)
BMI[d] (kg/m^2), mean (SD)	30.4 (6.5)	31.4 (7.2)	29.8 (6.0)
Comorbidities, n (%)			
Atrial fibrillation or flutter[d]	29 (26.9)	8 (20)	21 (30.9)
CAD or myocardial infarction	89 (82.4)	35 (87.5)	54 (79.4)
Diabetes mellitus[d]	30 (27.8)	17 (42.5)	13 (19.1)
Hypertension[d]	73 (67.6)	29 (72.5)	44 (64.7)
Heart failure	20 (18.5)	7 (17.5)	13 (19.1)
History of valve repair or replacement	28 (25.9)	9 (22.5)	19 (28.9)

[a]PCI: percutaneous coronary intervention.

[b]CABG: coronary artery bypass grafting.

[c]CAD: coronary artery disease.

[d]Data available for Michigan Medicine patients only.

Impact of Texts on Apple Watch Users

Baseline mean daily step count was 7536.0 (SD 3666.3) steps for Apple Watch users. During the initiation period of the study, delivering a text increased step count by a nonsignificant 10% (95% CI −1% to 20%; *P*=.05) in the 60 minutes after text delivery. Over the entire study period, walking texts increased step count more than antisedentary texts, although the difference was nonsignificant, with a relative risk ratio (RRR) of 1.08 (95% CI 1.00-1.17; *P*=.05; Table 2). In the study overall and in the initiation phase, text tailoring based on time of day, day of the week (weekend vs weekday), or personalization with participants' names did not significantly change the impact of texts on physical activity levels.

XSL•FO
RenderX

Table 2. Moderator analysis of contextually tailored text messages to promote physical activity among Apple Watch users in the VALENTINE (Virtual Application-Supported Environment to Increase Exercise) study.

Moderators	Overall			Initiation phase		
	RRR[a]	95% CI	*P* value	RRR	95% CI	*P* value
Time of day						
Morning vs lunch	0.93	0.84-1.02	.14	0.87	0.63-1.21	.42
Morning vs afternoon	0.95	0.85-1.06	.36	0.97	0.72-1.29	.81
Morning vs evening	0.93	0.84-1.03	.20	0.86	0.63-1.19	.37
Lunch vs afternoon	1.02	0.93-1.12	.65	1.10	0.86-1.42	.44
Lunch vs evening	1.00	0.91-1.10	.99	0.99	0.79-1.24	.92
Afternoon vs evening	0.98	0.90-1.06	.60	0.89	0.70-1.14	.36
Time of week: weekend vs weekday	0.97	0.90-1.05	.45	0.96	0.78-1.18	.70
Text type: walking vs antisedentary	1.08	1.00-1.17	.05	1.05	0.84-1.31	.67
Personalization: name vs no name	0.99	0.93-1.04	.58	0.88	0.75-1.03	.11
Age category: young (<65 years) vs old (>65 years)	1.04	0.97-1.11	.23	*1.27*[b]	*1.07-1.51*	*.01*
Sex: male vs female	0.99	0.93-1.05	.72	*1.27*	*1.05-1.52*	*.01*
Activity status: less active vs active	0.99	0.93-1.05	.64	0.91	0.76-1.10	.35
Heart failure: present vs absent	1.03	0.97-1.10	.31	1.02	0.81-1.30	.85
Indication for cardiac rehabilitation						
PCI[c] vs CABG[d]	1.04	0.95-1.14	.39	0.84	0.56-1.26	.40
Valve repair or replacement vs CABG	1.03	0.93-1.15	.56	0.72	0.47-1.09	.12
Valve repair or replacement vs PCI	0.99	0.91-1.08	.85	0.85	0.70-1.05	.13

[a]RRR: relative risk ratio.

[b]Italicized values indicate statistical significance.

[c]PCI: percutaneous coronary intervention.

[d]CABG: coronary artery bypass graft.

When analyzing the impact of texts based on participant characteristics, younger participants (RRR 1.27; 95% CI 1.07-1.51; *P*=.006; Table 2) and male participants (RRR 1.27; 95% CI 1.05-1.52; *P*=.01; Table 2) were more responsive to texts than older participants (age >65 years) and female participants during the initiation phase of the study. The indication for CR, the presence or absence of heart failure, and baseline physical activity levels did not significantly impact responsiveness to texts.

Impact of Texts on Fitbit Users

Baseline mean daily step count was 7011.2 (SD 3711.5) steps for Fitbit users. In the initiation phase of the study, texts increased step counts by 17% (95% CI 7%-28%; *P*<.001) in the 60 minutes after text delivery. In the initiation phase of the study, there was a trend, although not statistically significant, toward increased step count when texts were delivered at lunch compared to the afternoon (RRR 1.35; 95% CI 1.00-1.82; *P*=.05; Table 3). This trend, although not statistically significant, persisted for the overall study period (RRR 1.11; 95% CI 0.99-1.24; *P*=.07; Table 3). In the study overall and in the initiation phase, text tailoring based on day of the week or personalization with participants' names did not significantly impact step count.

XSL•FO
RenderX

Table 3. Moderator analysis of contextually tailored text messages to promote physical activity among Fitbit Versa 2 users in VALENTINE (Virtual Application-Supported Environment to Increase Exercise) study.

Moderators	Overall			Initiation phase		
	RRR[a]	95% CI	*P* value	RRR	95% CI	*P* value
Time of day						
Morning vs lunch	0.94	0.82-1.07	.34	1.02	0.72-1.44	.91
Morning vs afternoon	1.04	0.91-1.19	.60	1.38	0.96-1.97	.08
Morning vs evening	1.04	0.89-1.22	.61	1.18	0.82-1.69	.38
Lunch vs afternoon	1.11	0.99-1.24	.07	1.35	1.00-1.82	.05
Lunch vs evening	1.11	0.97-1.28	.13	1.15	0.82-1.63	.42
Afternoon vs evening	1.00	0.89-1.14	.95	0.85	0.59-1.24	.41
Time of week: weekend vs weekday	0.95	0.87-1.03	.21	0.94	0.70-1.26	.67
Text type: walking vs antisedentary	1.00	0.92-1.08	.95	0.93	0.73-1.20	.60
Personalization: name vs no name	0.99	0.93-1.06	.87	0.89	0.71-1.12	.31
Age category: young (<65 years) vs old (>65 years)	0.98	0.91-1.07	.70	0.94	0.78-1.13	.52
Sex: male vs female	*1.10*[b]	*1.02-1.18*	*.01*	1.04	0.88-1.23	.68
Activity status: less active vs active	0.95	0.88-1.02	.14	0.88	0.75-1.04	.15
Heart failure: present vs absent	0.99	0.89-1.11	.92	0.91	0.92-1.30	.31
Indication for cardiac rehabilitation						
PCI[c] vs CABG[d]	0.99	0.93-1.06	.79	0.88	0.74-1.06	.18
Valve repair or replacement vs CABG	1.01	0.92-1.11	.89	*1.31*	*1.03-1.67*	*.03*
Valve repair or replacement vs PCI	1.02	0.92-1.12	.76	*1.48*	*1.17-1.87*	*.001*

[a]RRR: relative risk ratio.

[b]Italicized values indicate statistical significance.

[c]PCI: percutaneous coronary intervention.

[d]CABG: coronary artery bypass graft.

When analyzing the impact of texts based on participant characteristics, male participants were more responsive to texts than female participants (RRR 1.10; 95% CI 1.02-1.18; *P*=.01; Table 3) in the overall study. The indication for CR significantly impacted participants' responsiveness to texts, with participants who underwent valve repair or replacement being more responsive to texts than those who underwent percutaneous coronary intervention (RRR 1.31; 95% CI 1.03-1.67; *P*=.03; Table 3) or coronary artery bypass grafting (RRR 1.48; 95% CI 1.17-1.87; *P*=.001; Table 3). Sex category, the presence or absence of heart failure, and baseline physical activity status did not significantly impact responsiveness to texts.

Discussion

Principal Findings

In the VALENTINE mHealth intervention designed to augment the benefits of CR through texts, we found that texts increased step counts in the 60 minutes following delivery by a statistically significant 17% for Fitbit users and a nonsignificant 10% for Apple Watch users in the initiation phase of the study. In this secondary analysis, the primary objective was to evaluate the intervention-level and participant-level moderators of text efficacy, examining the study period overall as well as the initiation phase, during which texts were most effective. We found that contextual factors related to texts, such as timing of delivery and content of texts, did not consistently affect step count in either Apple Watch or Fitbit users. Notably, although personalization of texts with participants' names was hypothesized to increase responsiveness based on prior studies [16-18], it did not result in increased step count compared to nonpersonalized texts for either device type in our study. Although participant-level factors, such as sex and age, occasionally had significant effects, these were not consistent across device types and phases of the trial. Notably, baseline physical activity was predicted to be a moderator of text responsiveness (participants with high baseline activity may not find texts as effective), but there was no statistically significant difference in text responsiveness between those with low and high baseline activity for either device type.

This study adds to a growing body of literature exploring the efficacy of text-based mHealth interventions and specifically ways to optimize delivery to elicit behavioral change. Prior studies have shown that texts can motivate patients with CVD to augment physical activity in the form of increased step count and time spent engaging in moderate-intensity exercise [16-23]. Meta-analyses, mostly in patient populations without CVD, have also shown that personalized and tailored texts that account

XSL•FO
RenderX

for user preferences, baseline characteristics, or real-time contextual data are more effective than generic texts in promoting engagement and responsiveness [18-21]. Although previous studies have analyzed different forms of text personalization and framing separately, no study has directly compared and evaluated multiple types of text tailoring within the same participant population, especially in patients with CVD [17-23]. Our study sought to address this gap by using a novel microrandomized trial design to randomize patients with CVD to different types of tailored texts and determine which factors were most effective for promoting physical activity.

Although the VALENTINE study did not identify text attributes that consistently increased activity levels, these results should not necessarily be interpreted as evidence that tailoring is ineffective but rather as a reflection of the complexity involved in tailoring texts. Despite focused efforts to tailor texts based on diverse factors hypothesized to be salient based on prior work [17,20,21,23], it is possible that the texts were not tailored to the most impactful factors, at least for this study population. For example, prior studies suggested that users desire personalized texts, which was enacted in the study by including participants' names in a subset of texts [16,17]. However, it is possible that participants seek personalization in other ways not included in this study, such as in the context of their specific health goals, recent activity trends, or work schedules [21-23]. Additionally, given participant heterogeneity, it is likely that there are inherent differences between participants that may make the same tailored text effective for some but not for others. More sophisticated strategies, such as the use of reinforcement learning algorithms, may be needed to account for these individual-level differences so that participants receive the types of texts to which they have responded historically [24].

Despite the results lacking statistical significance, the VALENTINE study did reveal some interesting associations that may warrant further exploration. For example, the content of texts appeared to be influential, with texts that promoted walking leading to a higher step count than texts that promoted antisedentary activities among Apple Watch users. The timing of texts may also matter, with Fitbit users responding more often to texts delivered at lunch, when participants may have had a work break to engage in physical activity. These are less explored but vital text attributes that are hypothesis-generating and should serve as early evidence for future work.

Our study also found that results varied between Apple Watch and Fitbit users, although the smaller participant sample size and lack of statistically significant results limit interpretation. Prior research has shown that key differences between the Apple Watch and the Fitbit, such as user interface, functionality, integration into daily life, and activity data collection methodology, influence user engagement and data accuracy [11-14]. Studies have also suggested differences in the characteristics of individuals who gravitate toward 1 device type over the other, with the Fitbit being marketed primarily as a health and fitness tool while the Apple Watch is a multipurpose device. However, very few studies have directly compared how the 2 devices could mediate behavioral change in mHealth interventions and what customizable factors matter for each device type. Future studies should explore how intrinsic features of wearable devices may cater to different individuals and contexts and thereby influence the efficacy of mHealth interventions.

Our study has several strengths. First, our study used a novel microrandomized design, a powerful experimental design for making causal inferences [25]. Through sequential randomized assignment, we were able to assess time-varying and contextual factors with the potential to moderate intervention efficacy, an inherent limitation of randomized controlled trials, which can only assess the impact of an intervention package in aggregate. Second, we explored multiple layers of tailoring, including text content and participant-level factors. Few studies have similarly analyzed multiple moderators of texts or tailored texts to factors such as content type (eg, walking vs antisedentary). Third, with the primary outcome being step count 60 minutes after a decision point, our study assessed the immediate and likely direct impact of texts rather than a more distant outcome that could be explained by confounding factors or other intervention components, as many mHealth interventions have multiple intervention components. Finally, our study included both Apple Watch and Fitbit users, capturing comparative data on wearable devices in the context of mHealth interventions and increasing participant inclusivity.

Our study also has limitations. First, the sample size for individual moderator subgroups was small, potentially limiting our ability to detect differences by text type or participant-level factors. Thus, this analysis should be viewed as hypothesis-generating but not confirmatory. However, we did collect a rich dataset consisting of 70,552 serial randomizations, which allowed powerful assessments of causal effects. Second, although the texts were tailored to many factors, other potentially impactful factors, such as real-time activity data or user preferences regarding text characteristics, were not incorporated. Future iterations will focus on adding this layer of personalization. Third, although baseline characteristics, such as demographics, baseline activity status, and indication for CR, were evaluated as moderators, other participant-level factors relevant to an mHealth trial were not assessed, such as digital literacy. However, given the highly accessible delivery format of texts, digital literacy is less likely to affect their impact. Fourth, although texts were delivered at a time when participants were thought to be available, text engagement data were not collected (ie, whether and when participants read the text on their smartwatch or opened it in the mobile app). Therefore, step count assessment following text delivery may not have always captured activity change after viewing a text or accounted for the fact that a participant was already active at the time of viewing the text. Nonetheless, the availability data still suggest meaningful exposure to the intervention in a timely manner. Fifth, the sample contained only low and moderate-risk patients with CVD who already owned smartphones, were enrolled in center-based CR, and had relatively high activity at baseline, suggesting a population with higher levels of digital literacy and intrinsic motivation. Given these already optimized characteristics at baseline, there may have been a ceiling effect in terms of how much more benefit participants gained from tailored texts. However, our sample is diverse in other clinically meaningful ways (more than one-third of participants were aged

XSL•FO
RenderX

≥65 years, and a subset of participants were from a community setting); therefore, our findings are still applicable to real-world CR populations. Finally, although we compared the relative impact of different tailoring variables, the study evaluated a package of tailored texts compared to no text and did not compare the impact of tailored texts to that of generic texts. Thus, we are unable to draw conclusions regarding the impact of tailoring more generally.

Conclusions

Overall, text message-based mHealth interventions are a promising approach for increasing physical activity levels among CR enrollees. In this secondary analysis of a microrandomized trial embedded within the VALENTINE study, we found that although tailored texts were associated with short-term increases in physical activity, no individual text-level or participant-level characteristic consistently moderated responsiveness across time periods and device types. These findings indicate that commonly used tailoring strategies, such as timing of text delivery, content type, or personalization with a participant's name, are insufficient on their own to optimize engagement and behavioral response in a heterogeneous patient population. Effective text tailoring likely requires a broader and more dynamic set of variables, including real-time behavioral data, evolving user preferences, and device-specific contexts. Future research should focus on developing learning algorithms that can learn and adapt to user preferences and adjust text delivery over time to more effectively augment and extend the benefits of center-based CR.

Acknowledgments

Generative artificial intelligence was not used in the preparation of this manuscript.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Funding

This work was supported by institutional grants at the University of Michigan, including a Precision Health, MCubed, and Aikens Innovation Awards.

Authors' Contributions

BM, SK, PK, BKN, T Boyden, and JRG conceived the study. NA, KG, JRG, and BKN conceived the analysis. The analysis was carried out by JRS, T Basu, and EL under the supervision of WD and JRG. NA and KG drafted the original manuscript, which was reviewed and edited by all coauthors.

Conflicts of Interest

JRG receives funding from the National Institutes of Health (NIH; L30HL143700, 1K23HL168220) and Patient-Centered Outcomes Research Institute. BKN is a principal investigator or coinvestigator on research grants from the NIH, Veterans Affairs Health Services Research and Development, and the American Heart Association. He also receives compensation as editor-in-chief of Circulation: Cardiovascular Quality & Outcomes, a journal of the American Heart Association. Finally, he is a coinventor on US utility patent number US15/356,012 (US20170148158A1) titled "Automated Analysis of Vasculature in Coronary Angiograms," which uses software technology with signal processing and machine learning to automate the reading of coronary angiograms and is held by the University of Michigan. The patent is licensed to AngioInsight Inc, in which BKN holds ownership shares and receives consultancy fees. SK is a principal investigator or coinvestigator on research grants from the NIH; Blue Cross Blue Shield of Michigan; the American Heart Association; Apple Inc; Merck & Co; and Becton, Dickinson & Company; he is also a coinventor on US patent number 62/791,257 titled "Automated System To Medical Procedures," which is held by the University of Michigan. PK is a principal investigator or a coinvestigator on research grants from the NIH. All other authors declare no other conflicts of interest.

References

1. Sandesara PB, Lambert CT, Gordon NF, Fletcher GF, Franklin BA, Wenger NK, et al. Cardiac rehabilitation and risk reduction: time to "rebrand and reinvigorate". J Am Coll Cardiol 2015 Feb 03;65(4):389-395 [FREE Full text] [doi: 10.1016/j.jacc.2014.10.059] [Medline: 25634839]
2. Dunlay SM, Pack QR, Thomas RJ, Killian JM, Roger VL. Participation in cardiac rehabilitation, readmissions, and death after acute myocardial infarction. Am J Med 2014 Jun;127(6):538-546 [FREE Full text] [doi: 10.1016/j.amjmed.2014.02.008] [Medline: 24556195]
3. Beatty AL, Brown TM, Corbett M, Diersing D, Keteyian SJ, Mola A, et al. Million hearts cardiac rehabilitation think tank: accelerating new care models. Circ Cardiovasc Qual Outcomes 2021 Oct;14(10):e008215 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.121.008215] [Medline: 34587751]

XSL•FO
RenderX

4. Hansen D, Dendale P, Raskin A, Schoonis A, Berger J, Vlassak I, et al. Long-term effect of rehabilitation in coronary artery disease patients: randomized clinical trial of the impact of exercise volume. Clin Rehabil 2010 Apr;24(4):319-327. [doi: 10.1177/0269215509353262] [Medline: 20176771]
5. Gupta R, Sanderson BK, Bittner V. Outcomes at one-year follow-up of women and men with coronary artery disease discharged from cardiac rehabilitation: what benefits are maintained? J Cardiopulm Rehabil Prev 2007;27(1):11-8; quiz 19-20. [doi: 10.1097/01.hcr.0000265015.44210.bf] [Medline: 17474639]
6. Piotrowicz E, Baranowski R, Bilinska M, Stepnowska M, Piotrowska M, Wójcik A, et al. A new model of home-based telemonitored cardiac rehabilitation in patients with heart failure: effectiveness, quality of life, and adherence. Eur J Heart Fail 2010 Feb;12(2):164-171 [FREE Full text] [doi: 10.1093/eurjhf/hfp181] [Medline: 20042423]
7. Varnfield M, Karunanithi M, Lee CK, Honeyman E, Arnold D, Ding H, et al. Smartphone-based home care model improved use of cardiac rehabilitation in postmyocardial infarction patients: results from a randomised controlled trial. Heart 2014 Nov;100(22):1770-1779 [FREE Full text] [doi: 10.1136/heartjnl-2014-305783] [Medline: 24973083]
8. Krishnamurthi N, Schopfer DW, Shen H, Rohrbach G, Elnaggar A, Whooley MA. Association of home-based cardiac rehabilitation with lower mortality in patients with cardiovascular disease: results from the Veterans Health Administration Healthy Heart Program. J Am Heart Assoc 2023 Mar 07;12(5):e025856 [FREE Full text] [doi: 10.1161/JAHA.122.025856] [Medline: 36856057]
9. Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H. Design features for improving mobile health intervention user engagement: systematic review and thematic analysis. J Med Internet Res 2020 Dec 09;22(12):e21687 [FREE Full text] [doi: 10.2196/21687] [Medline: 33295292]
10. Thomas RJ, Beatty AL, Beckie TM, Brewer LC, Brown TM, Forman DE, et al. Home-based cardiac rehabilitation: a scientific statement from the American Association of Cardiovascular and Pulmonary Rehabilitation, the American Heart Association, and the American College of Cardiology. J Am Coll Cardiol 2019 Jul 09;74(1):133-153 [FREE Full text] [doi: 10.1016/j.jacc.2019.03.008] [Medline: 31097258]
11. Jeganathan VS, Golbus JR, Gupta K, Luff E, Dempsey W, Boyden T, et al. Virtual AppLication-supported Environment To INcrease Exercise (VALENTINE) during cardiac rehabilitation study: rationale and design. Am Heart J 2022 Jun;248:53-62. [doi: 10.1016/j.ahj.2022.02.012] [Medline: 35235834]
12. Golbus JR, Gupta K, Stevens R, Jeganathan VS, Luff E, Shi J, et al. A randomized trial of a mobile health intervention to augment cardiac rehabilitation. NPJ Digit Med 2023 Sep 14;6(1):173 [FREE Full text] [doi: 10.1038/s41746-023-00921-9] [Medline: 37709933]
13. Golbus JR, Gupta K, Stevens R, Jeganathan VS, Luff E, Boyden T, et al. Understanding baseline physical activity in cardiac rehabilitation enrollees using mobile health technologies. Circ Cardiovasc Qual Outcomes 2022 Jul;15(7):e009182 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.122.009182] [Medline: 35559648]
14. Bai Y, Tompkins C, Gell N, Dione D, Zhang T, Byun W. Comprehensive comparison of Apple Watch and Fitbit monitors in a free-living setting. PLoS One 2021 May 26;16(5):e0251975 [FREE Full text] [doi: 10.1371/journal.pone.0251975] [Medline: 34038458]
15. Boruvka A, Almirall D, Witkiewitz K, Murphy SA. Assessing time-varying causal effect moderation in mobile health. J Am Stat Assoc 2018;113(523):1112-1121 [FREE Full text] [doi: 10.1080/01621459.2017.1305274] [Medline: 30467446]
16. Atluri N, Mishra SR, Anderson T, Stevens R, Edwards A, Luff E, et al. Acceptability of a text message-based mobile health intervention to promote physical activity in cardiac rehabilitation enrollees: a qualitative substudy of participant perspectives. J Am Heart Assoc 2024 Jan 16;13(2):e030807 [FREE Full text] [doi: 10.1161/JAHA.123.030807] [Medline: 38226512]
17. Rabbi M, Pfammatter A, Zhang M, Spring B, Choudhury T. Automated personalized feedback for physical activity and dietary behavior change with mobile phones: a randomized controlled trial on adults. JMIR Mhealth Uhealth 2015 May 14;3(2):e42 [FREE Full text] [doi: 10.2196/mhealth.4160] [Medline: 25977197]
18. Shariful Islam SM, Farmer AJ, Bobrow K, Maddison R, Whittaker R, Pfaeffli Dale LA, et al. Mobile phone text-messaging interventions aimed to prevent cardiovascular diseases (Text2PreventCVD): systematic review and individual patient data meta-analysis. Open Heart 2019 Oct 9;6(2):e001017 [FREE Full text] [doi: 10.1136/openhrt-2019-001017] [Medline: 31673381]
19. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]
20. Head KJ, Noar SM, Iannarino NT, Grant Harrington N. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
21. Smith DM, Duque L, Huffman JC, Healy BC, Celano CM. Text message interventions for physical activity: a systematic review and meta-analysis. Am J Prev Med 2020 Jan;58(1):142-151 [FREE Full text] [doi: 10.1016/j.amepre.2019.08.014] [Medline: 31759805]
22. Martin SS, Feldman DI, Blumenthal RS, Jones SR, Post WS, McKibben RA, et al. mActive: a randomized clinical trial of an automated mHealth intervention for physical activity promotion. J Am Heart Assoc 2015 Nov 09;4(11):e002239 [FREE Full text] [doi: 10.1161/JAHA.115.002239] [Medline: 26553211]

XSL•FO
RenderX

23. Klasnja P, Smith S, Seewald NJ, Lee A, Hall K, Luers B, et al. Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of HeartSteps. Ann Behav Med 2019 May 03;53(6):573-582 [FREE Full text] [doi: 10.1093/abm/kay067] [Medline: 30192907]
24. Liao P, Greenewald K, Klasnja P, Murphy S. Personalized HeartSteps: a reinforcement learning algorithm for optimizing physical activity. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020 Mar;4(1):18 [FREE Full text] [doi: 10.1145/3381007] [Medline: 34527853]
25. Golbus JR, Dempsey W, Jackson EA, Nallamothu BK, Klasnja P. Microrandomized trial design for evaluating just-in-time adaptive interventions through mobile health technologies for cardiovascular disease. Circ Cardiovasc Qual Outcomes 2021 Feb;14(2):e006760 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.120.006760] [Medline: 33430608]

Abbreviations

CR: cardiac rehabilitation
CVD: cardiovascular disease
mHealth: mobile health
RRR: relative risk ratio
VALENTINE: Virtual Application-Supported Environment to Increase Exercise

Edited by Z Ni; submitted 29.Jun.2025; peer-reviewed by M Fraser, R Thomas, I Ensari; comments to author 27.Oct.2025; accepted 25.May.2026; published 11.Aug.2026.

Please cite as:
Atluri N, Gupta K, Basu T, Luff E, Shi J, Boyden T, Mukherjee B, Kheterpal S, Klasnja P, Dempsey W, Nallamothu BK, Golbus JR Characteristics of Tailored Text Messages Associated With Increased Physical Activity Among Cardiac Rehabilitation Enrollees: Secondary Analysis of a Microrandomized Trial
JMIR Mhealth Uhealth 2026;14:e79792
URL: https://mhealth.jmir.org/2026/1/e79792
doi:10.2196/79792
PMID:

©Namratha Atluri, Kashvi Gupta, Tanima Basu, Evan Luff, Jieru Shi, Thomas Boyden, Bhramar Mukherjee, Sachin Kheterpal, Predrag Klasnja, Walter Dempsey, Brahmajee K Nallamothu, Jessica R Golbus. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Personalized Mobile App–Based Messaging and Step Tracking for Improving Walking Behavior in Adults: Randomized Controlled Trial

Sarang Jang[1], PhD; Aeree Sohn[1], DrPH

Department of Public Health, Sahmyook University, Seoul, Republic of Korea

Corresponding Author:
Aeree Sohn, DrPH
Department of Public Health
Sahmyook University
815, Hwarang-ro, Nowon-gu
Seoul, 01795
Republic of Korea
Phone: 82 2 3399 1669
Email: aeree@syu.ac.kr

Abstract

Background: Walking is one of the most effective forms of physical activity; however, maintaining long-term engagement in digital step-counting interventions remains challenging.

Objective: This study aimed to evaluate the effects of personalized message interventions integrating the Information-Motivation-Behavioral Skills (IMB) model and regulatory focus theory (RFT) on walking behavior.

Methods: Participants were recruited offline through community advertisements containing a QR code that linked to an online sign-up form and were screened using an online baseline questionnaire by research staff, while all intervention delivery and outcome assessments were conducted via the WalkOn mobile app. A 12-week randomized controlled trial targeting 174 adults in Mapo-gu, Seoul, Korea, was conducted from April 1 to June 22, 2024. Of the 174 randomized participants, 141 (67 in the intervention group and 74 in the control group) engaged in the intervention. Both groups received incentives based on goal achievement and general health information, whereas the intervention group received weekly personalized messages provided by research staff according to midweek step-count performance. Primary outcomes included weekly goal achievement rates for the 25,000-step goal, step-count patterns, and consecutive 2-week goal achievement failure events. Analyses were conducted using generalized estimating equations (GEEs), linear mixed models (LMMs), and Cox proportional hazards models.

Results: Of the 174 participants randomized, 141 (81%) engaged in the intervention and were included in the analyses (67 intervention; 74 control), while 33 participants did not participate after randomization. The overall 12-week mean goal achievement rate was 87% in the intervention group and 87% in the control group, with no significant difference between the groups (odds ratio [OR] 1.12, 95% CI 0.29-4.37; *P*=.87). Weekly step-count patterns also showed no significant differences (group × time interaction: estimate=–417; SE 09; *P*=.31). However, the risk of consecutive 2-week goal achievement failure events was significantly lower in the intervention group (hazard ratio [HR] 0.67, 95% CI 0.52-0.86; *P*=.002). The intervention group showed an approximately 33% lower likelihood of consecutive 2-week goal achievement failure events than the control group.

Conclusions: Personalized message interventions did not show significant effects on overall step-count increase but were associated with a reduced risk of consecutive 2-week goal achievement failure events. These findings suggest that personalized messages may play an important role in improving behavior change sustainability and failure resilience. Incorporating personalized messaging strategies may enhance the effectiveness of digital health intervention design.

Trial Registration: Clinical Research Information Service (CRIS), Republic of Korea KCT0010966; https://tinyurl.com/3jeepxve

(JMIR Mhealth Uhealth 2026;14:e84430) doi:10.2196/84430

KEYWORDS

digital health; mobile applications; mobile apps; personalized messaging; theory-based intervention; walking

XSL•FO
RenderX

Introduction

Walking is one of the most accessible and effective forms of physical activity and is consistently associated with reduced risks of cardiovascular disease, cancer, type 2 diabetes, and dementia [1-3]. In particular, according to meta-analyses and large-scale cohort studies, walking briskly for 30 minutes daily, 5 days a week, has been shown to significantly reduce cardiovascular disease risk factors, improve blood pressure control, and contribute to healthy aging [1,2,4]. Additionally, walking for 30 minutes daily is associated with substantial health benefits, including improved mental health, enhanced mobility, and increased life expectancy [5]. Walking interventions have been emphasized in rehabilitation contexts and programs for older adults, and structured and sustained walking programs have shown particularly important benefits for mobility, balance, quality of life, and stroke recovery [6]. Recent studies have shown that even short walks of 5 to 11 minutes per day are associated with reduced mortality from cancer, cardiovascular disease, and all causes [7]. Walking is an ideal tool for population-level health promotion interventions due to its accessibility and scalability, as it requires no special equipment, can be performed in various environments, and can be adapted to an individual's fitness level [8].

Despite these clear health benefits, physical inactivity remains a serious public health problem worldwide, contributing to the burden of noncommunicable diseases and reduced quality of life [9]. As an innovative approach to address this issue, digital step-counting interventions using pedometers, smartphone apps, and wearable devices have emerged as promising strategies to encourage walking behavior and increase overall physical activity levels. These interventions have received significant attention owing to their accessibility, cost-effectiveness, and scalability across diverse populations.

The proliferation of digital health technologies has created unprecedented opportunities for individuals to monitor their physical activity on their own [10], and step counters provide real-time feedback on daily activity levels and enable users to track their progress toward the recommended physical activity guidelines. The appeal of such interventions lies in the simplicity and quantifiability of step counting, which can serve as a visible and motivating goal for individuals seeking to increase their physical activity.

There is growing evidence supporting the short-term effectiveness of digital step counting in promoting actual walking behavior. According to several systematic reviews and meta-analyses, such interventions significantly increased daily step counts, with an average increase of 1126 steps per day after 4 months of pedometer use, and sustained improvements were observed even up to 3 to 4 years later [11]. Similarly, interventions using smartphone apps have demonstrated an average daily increase of 1182 steps over 3 months [12]. These increases are often accompanied by meaningful improvements in health outcomes such as weight loss, reduced BMI, lower blood pressure, decreased triglyceride levels, and improved psychological well-being [12,13].

However, despite these initial positive results, there is growing evidence that digital step-counting interventions have significant limitations in maintaining long-term walking behavior changes. The novelty and enthusiasm surrounding these technologies tend to wane over time, leading to lower participation and adherence rates. Several longitudinal studies have reported that more than half of the participants discontinue device use within 6 months of a digital step-counting intervention [14]. The novelty effect begins to fade after approximately 3 months, with the effect size in long-term interventions decreasing by approximately 40% compared with short-term interventions [15]. These sustainability issues highlight the significant gap between the short- and long-term public health impacts, suggesting that maintaining high activity levels requires addressing complex psychological, social, and environmental factors beyond simple self-monitoring.

To overcome these limitations, digital walking interventions have evolved around step-counting technology using wearable devices and mobile apps, with various efforts to increase physical activity engagement [16]. Current digital walking interventions commonly integrate behavior change techniques (BCTs) such as self-monitoring, goal setting, social comparison, and gamification, which are increasingly enhanced through personalized feedback [17,18]. Among these approaches, personalized messaging has proven to be a particularly effective form of personalized feedback that supports behavior maintenance by providing tailored information and motivation based on individual behavioral data [19]. Research shows that personalized messages elicit higher levels of engagement and adherence than generic information approaches [20,21].

However, current digital walking interventions have several limitations that restrict their effectiveness. First, the accuracy of commercially available step counters can vary significantly in real-world settings, with limited representativeness and generalizability across diverse populations [22,23]. Second, user engagement and compliance tend to decrease over time, and the long-term maintenance of behavior change remains an ongoing challenge [11,24]. Third, existing personalized messaging research primarily focuses on general encouragement or simple performance feedback and lacks systematic research on differentiated theory-based messaging strategies according to individual real-time performance levels. Fourth, most studies rely on single behavior change theories, failing to comprehensively address individuals' complex motivations and change stages. These limitations emphasize the need for more sophisticated intervention approaches beyond simple feedback mechanisms.

Digital health interventions are increasingly integrating behavior change theories to improve their effectiveness [25,26]. While the effectiveness of an intervention depends on the appropriateness of the theoretical framework rather than the mere number of theories used, this study specifically integrates the Information-Motivation-Behavioral Skills (IMB) model and regulatory focus theory (RFT) to support the design of personalized walking interventions. Existing research is limited to adaptive approaches that comprehensively address information provision, motivation, and behavioral skill acquisition while dynamically adjusting prevention- or

XSL•FO
RenderX

promotion-focused messages according to individual weekly performance changes.

The IMB model proposes that three core components are necessary for behavior change [27]: (1) accurate information related to the behavior, (2) personal and social motivation to perform the behavior, and (3) behavioral skills to effectively perform the behavior [28,29]. This model provides a structured theoretical framework supporting users in understanding relevant information, developing intrinsic and extrinsic motivation, and acquiring the skills necessary for behavior change. Similarly, the RFT explains that individual motivational orientations are distinguished as prevention and promotion foci [30-32]. RFT proposes that individuals regulate goal pursuit through 2 distinct motivational orientations: promotion focus and prevention focus [33,34]. A promotion focus is concerned with aspirations, accomplishments, and advancement, motivating individuals to pursue gains and ideal outcomes. In contrast, a prevention focus centers on safety, responsibility, and the fulfillment of obligations, where the primary motivation is loss avoidance and the maintenance of a non-loss state. This theory can support habit formation by effectively encouraging the repetition and habituation of walking by connecting goal achievement with positive reinforcement and rewards.

However, the IMB model has limitations in that it does not consider differences in individual motivational orientations and regulatory focus types and fails to explain dynamic changes in motivational needs according to performance levels [35,36]. Conversely, RFT has limitations in that it does not provide specific behavior change mechanisms, focuses primarily on motivational orientation, and does not specify what types of information or skills should be provided according to the regulatory focus [33,37]. These limitations of each theory can be addressed through complementary integration.

The integration of the IMB model and RFT operates through integrative processes that mutually complement the limitations of each theory. This study proposes an integrative approach that differentiates the IMB's information component into loss or gain frames according to RFT's prevention and promotion foci, applies motivational strategies tailored to individual regulatory focus, and provides behavioral skills differentiated by regulatory focus (prevention focus: barrier removal; promotion focus: performance optimization). Conversely, the IMB model provides a foundation for integrating systematic information provision mechanisms, specific behavioral skill acquisition processes, and multilayered motivational systems to complement RFT's limitations.

Specifically, this study proposes a performance-based dynamic integration model that dynamically determines the regulatory focus based on participants' real-time step count performance and selectively applies IMB components accordingly. Low-performance groups (<12,500 steps) activate a prevention focus emphasizing information and behavioral skills, medium-performance groups (12,500-24,999 steps) apply a prevention focus and an IMB integrated approach, and high-performance groups (≥25,000 steps) activate a promotion focus emphasizing motivation.

Based on these theoretical and practical needs, this study aims to overcome existing research limitations and present new approaches. The originality of this study lies in the following aspects. First, we developed a personalized messaging system that is responsively adjusted weekly based on the participants' real-time step count performance. Second, we systematically implemented differentiated messaging strategies according to individual performance levels by applying a theoretical framework that integrates the IMB model and RFT. Third, we comprehensively evaluated the intervention effects from the perspective of behavior change maintenance, focusing on goal achievement sustainability and consecutive failure prevention beyond a simple step count increase. This approach overcomes the limitations of existing static and uniform personalized messaging and presents the possibility of dynamic, customized interventions tailored to individual change stages and performance levels.

Therefore, this study seeks to answer the following specific research questions: (1) Do personalized theory-based messages have significant effects on weekly step count goal (25,000 steps) achievement rates over 12 weeks? (2) Does the message intervention have significant effects on participants' step count patterns over 12 weeks? and (3) Do personalized messages reduce the risk of consecutive 2-week goal achievement failures? To answer these research questions, this study adopted a randomized controlled trial (RCT) design and conducted a 12-week digital walking intervention targeting 174 adults.

Methods

Study Design

This study was a 12-week RCT conducted from April 1 to June 22, 2024, to evaluate the effects of personalized health message interventions on walking behavior in Korean adults. A parallel-group design with a 1:1 allocation ratio was used. An active control group design was adopted to isolate the specific effects of personalized messaging by providing substantial nonpersonalized intervention components to the control group.

The trial was registered retrospectively with the Clinical Research Information Service (CRIS; KCT0010966). The registered study procedures corresponded to the protocol implemented in the trial. The study protocol, including primary outcomes and statistical analysis plans, was finalized prior to data analysis. No changes were made to the primary outcomes or analytical strategies after examination of the data. The registered protocol corresponds to the procedures implemented in the trial.

Sample Size

The sample size calculation was performed using G*Power. Assuming a small effect size (Cohen d=0.20), a significance level of α=0.05, and power of 80%, the required sample size for the primary comparison was 200 participants per group. However, the final number of participants randomized was 174, with 141 engaging in the study (67 in the intervention group and 74 in the control group). The impact of this reduced sample size on statistical power and precision is discussed in the Discussion section.

XSL•FO
RenderX

Study Participants

Overview

Residents of Mapo-gu, Seoul, were recruited through voluntary participation via community offline advertisements (eg, local posters and bulletin boards containing a QR code that linked to an online sign-up form) and were screened using an online baseline questionnaire by research staff. Eligibility criteria included being an adult aged 19 years or older, owning a smartphone capable of installing a mobile app, being able to walk independently, and having signed a consent form for study participation. Those with difficulty walking or an inability to use a smartphone app were excluded. A total of 174 participants were randomized (87 per group).

Participants received incentive points for achieving weekly walking goals during the intervention period. The incentive value progressively increased every 4 weeks (weeks 1-4: 1000 points [KRW 1000; KRW 1=US $0.00074 as of April 5, 2024]; weeks 5-8: 1500 points [KRW 1500]; weeks 9-12: 2000 points [KRW 2000]) to encourage sustained participation.

Randomization and Intervention

After baseline data collection, the participants were assigned to either the intervention or control group using a computer-generated randomization sequence created with the RAND function in Microsoft Excel. Following randomization, the intervention began, and participants were monitored throughout the 12-week intervention period. Both groups were presented with a weekly walking goal of 25,000 steps and used the WalkOn mobile app (Swallaby Co) for real-time step tracking.

The intervention was implemented using the WalkOn mobile app, a commercially available smartphone app that tracks users' daily step counts using built-in motion sensors and provides real-time feedback on walking activity. Step count data were automatically recorded and stored on the system server, allowing researchers to collect participants' walking data throughout the 12-week intervention period. Program participation was defined as having at least one recorded step count through the WalkOn mobile app after the intervention began. Participants who had no recorded step count data during the first week of the intervention were classified as nonparticipants and were considered not to have initiated the program. These participants had agreed to participate but had no recorded step-count data during the first week after randomization, indicating that the WalkOn step-tracking app had not been activated on their smartphones.

The weekly goal of 25,000 steps established in this study was set by converting the World Health Organization's (WHO) physical activity recommendations for adults—"at least 150 minutes of moderate-intensity physical activity per week"—into step counts [34]. According to previous research, 30 minutes of moderate walking corresponds to approximately 3000 steps when performed at a walking cadence of approximately 100 steps per minute [38,39]. Multiple controlled laboratory studies have confirmed that a cadence of 100 steps per minute corresponds to moderate physical activity intensity (3 metabolic equivalents [METs]) [40,41]. In addition, step-count–based physical activity classifications suggest that fewer than 5000 steps per day represents a sedentary lifestyle threshold [42]. Accordingly, accumulating approximately 25,000 steps per week represents a realistic and attainable behavioral target that encourages participants to move beyond sedentary activity levels while remaining feasible for community-based populations.

Based on these considerations, this study set the weekly step goal at 25,000 steps, equivalent to approximately 5 days of moderate walking per week. This threshold also allowed meaningful midweek performance classification (<50%, 50%-99%, and ≥100%) for the personalized messaging algorithm used in the intervention. Therefore, the selected goal was considered both practically achievable and consistent with the WHO physical activity recommendation standards.

The goal was intentionally based on approximately 5 days of activity rather than 7 days to reflect the WHO recommendation structure (30 minutes of moderate physical activity on at least 5 days per week). In addition, allowing 2 potential rest days per week reflects realistic physical activity patterns in community populations and improves the feasibility and sustainability of behavioral interventions.

Intervention Group (Personalized Message Group)

The intervention group received personalized health messages via KakaoTalk (Kakao Corp) every Friday based on walking performance from Monday to Wednesday of each week, in addition to basic intervention elements.

Participants' performance groups were determined based on the cumulative step counts recorded from Monday to Wednesday of each week, rather than projected weekly averages. These midweek step totals were compared with the predefined weekly goal of 25,000 steps to classify participants into three performance categories (<50%, 50%-99%, and ≥100%). Personalized messages were developed using RFT and the IMB model.

Participants were classified into three subgroups based on the cumulative step counts recorded from Monday to Wednesday of each week: (1) fewer than 12,500 steps received prevention-focused messages emphasizing information and behavioral skills; (2) 12,500 to 24,999 steps received prevention-focused messages, including all 3 IMB model components; and (3) 25,000 steps or more (≥100% of the weekly goal) received promotion-focused messages emphasizing motivation. Participants who accumulated ≥25,000 steps between Monday and Wednesday were considered to have already achieved the weekly goal ahead of schedule and therefore received promotion-focused reinforcement messages designed to sustain their walking behavior for the remainder of the week. Motivational elements were present across the intervention through goal setting and incentive systems; however, message content for the low-performance group primarily emphasized information and behavioral skills to address practical barriers to walking. The performance categories were predefined prior to study implementation and were directly linked to the weekly goal of 25,000 steps and the associated incentive system. Because incentives were provided upon achieving the 25,000-step weekly target, proportional

XSL•FO
RenderX

progress toward this predefined goal (<50%, 50%-99%, and ≥100%) was used to tailor intervention messages. Thus, the categorization reflected meaningful stages of goal attainment rather than arbitrary thresholds. The specific message strategies and theoretical components for each performance range are summarized in Table 1. Each message was personalized with the participants' names and specific performance feedback, including concrete walking suggestions reflecting the seasonal context and local characteristics.

Table 1. Theory-based personalized message components and examples by performance group.

Performance group	Step count range	Theoretical framework	Message focus and example[a]
High-performance group	≥25,000 steps per week	• RFT[b]: promotion focus • IMB[c]: motivation	• Focus: celebrating success and maintaining engagement • Example: "Great job! You've already achieved 25,000+ steps. You've earned 1,500 points. Congratulations!"
Medium-performance group	12,500 to 24,999 steps per week	• RFT: prevention focus • IMB: information, motivation, and behavioral skills	• Focus: emphasizing proximity to goal with specific behavioral suggestions • Example: "Mid-week check: You've reached more than half of your weekly goal (25,000 steps). How about evening walks to enjoy cricket sounds?"
Low-performance group	<12,500 steps per week	• RFT: prevention focus • IMB: information and behavioral skills	• Focus: providing specific behavioral strategies and preventing goal abandonment • Example: "Did you know that just a 5-minute walk after meals can help lower your blood sugar and boost your mood? Try a short stroll today!"

[a]Messages were delivered every Friday via KakaoTalk to intervention group participants (n=67) based on Monday to Wednesday step count assessment. Each message included personalized greetings using participants' names or nicknames.

[b]RFT: regulatory focus theory.

[c]IMB: Information-Motivation-Behavioral Skills.

Control Group (Active Control Group)

The control group was designed as an active control group that received all intervention elements except personalized messages. Specifically, they received the following components:

- Same incentive system: both groups received the same progressive incentive system for achieving weekly walking goals during the intervention period.
- General health information provision:
  participants received standardized, nonpersonalized health information weekly in card news format, including information on walking, smoking cessation, alcohol moderation, and community health resources.
- Social comparison feature:
  participants could experience social comparison and competitive motivation through real-time step count comparison features within the WalkOn app.
- Same goal setting:
  participants received the same weekly walking goal of 25,000 steps as the intervention group.

Through this active control group design, we aimed to isolate the effect of personalized messages and evaluate the added value of personalization beyond general health information provision or social comparison. All intervention components except personalized messaging were identical between the 2 groups.

Operational Definition of the Theoretical Framework

The IMB and RFT integrated framework developed in this study is a system that dynamically determines RFT's prevention and promotion focus according to participants' weekly step count performance and selectively combines and applies IMB model components (information, motivation, and behavioral skills) accordingly.

Performance-Based Regulatory Focus Determination

The participants were classified into 3 groups based on their midweek step count, with the personalization algorithm dynamically determining the regulatory frame to maximize motivational relevance. For the low-performance group (<12,500 steps per week; <50% achievement), messages used a prevention focus to emphasize "loss avoidance," highlighting potential negative health outcomes and the risk of failing the weekly commitment to prevent goal abandonment. The medium-performance group (12,500-24,999 steps per week; ≥50% achievement) received messages that maintained a prevention-oriented frame while emphasizing the possibility of achieving the weekly goal. Conversely, for the high-performance group (≥25,000 steps per week; 100% achievement), a promotion focus was applied to celebrate achievement and progress, focusing on positive gains and encouraging the pursuit of higher health aspirations. This dynamic application of RFT ensures that the motivational feedback remains tailored to the individual's real-time behavioral state.

IMB Component Application Strategy

The IMB model components were applied differently for each performance range. Low-performance ranges were centered on information and behavioral skills components, providing specific implementation methods and self-monitoring strategies. Medium performance ranges provided comprehensive support by integrating all information, motivation, and behavioral skills components. High-performance ranges were centered on

XSL•FO
RenderX

motivation components to induce achievement recognition and continued participation.

The information components emphasized the importance of goal achievement and current status feedback for prevention focus, achievement results, and rewards for promotion focus. Motivation components used supportive keywords like “cheer up” and “let’s do this together” for prevention focus, and recognition keywords like “excellent” and “congratulations” for promotion focus. Behavioral skills components provided specific practice strategies for prevention focus, and performance maintenance and expansion strategies for promotion focus.

Message Generation Algorithm

Personalized messages were automatically generated using the step-by-step algorithm.

Message Generation Procedure

In the first step, participants’ weekly step count data (Monday to Wednesday step count sum, with Thursday 12 AM as the cutoff) were collected to classify performance ranges. Midweek (Monday to Wednesday) step count assessment was intentionally implemented as an early trajectory monitoring strategy to facilitate timely feedback within the same weekly goal cycle. Because the intervention was fundamentally structured around the predefined weekly target of 25,000 steps, midweek performance evaluation allowed differentiated messaging to support goal attainment before the week concluded, rather than serving as a replacement for the weekly outcome measure. The second step determined the regulatory focus (prevention or promotion) according to the performance ranges and selected the IMB components to apply. The third step generated message content according to the selected components and added personalization elements.

Message Structure Template

All messages were structured identically: (1) greeting including the personal name, (2) specific step count feedback, (3) core message by performance range, (4) specific behavioral suggestions, and (5) encouragement and closing. Each message was limited to 150 to 200 characters to maintain conciseness while preserving a friendly and personal tone to strengthen relationships with the participants. Examples of the performance-level messages are provided in Multimedia Appendix 1.

Application of Personalization Elements

All messages included the participants’ real names, actual step counts, and goal achievement rates. They also suggested seasonally appropriate activities (spring: cherry blossom walks; summer: early morning walks) and specific locations reflecting local characteristics (Hangang Park and World Cup Park). Differential messages were provided according to change patterns compared with the previous weeks: encouragement for improvement, new strategy suggestions for stagnation, and supportive messages for decline.

Outcome Measurement

Baseline characteristics, including gender and age, were measured through preintervention surveys. Step count-related outcome variables were calculated using daily step count data automatically collected via the WalkOn app, which uses built-in smartphone accelerometers with weekly verification procedures to ensure data accuracy and integrity.

Although participants could withdraw from the WalkOn community at any time, the research team had securely exported and stored all step-count data for the 12-week intervention period in Microsoft Excel files before any withdrawals, ensuring that participant withdrawal did not affect the completeness or integrity of the final dataset.

Weekly Goal Achievement Rate

Weekly goal achievement rate was defined as the proportion of participants achieving ≥25,000 steps per week over 12 weeks, with individual achievement coded as binary (1=achievement; 0=nonachievement).

Twelve-Week Step Count Patterns

Step count patterns were analyzed using total weekly steps (Monday to Sunday) as a continuous variable to examine changes in step counts over the 12-week intervention period.

Consecutive 2-Week Goal Achievement Failure Events

Consecutive 2-week goal achievement failure was defined as the first occurrence of failing to achieve the weekly 25,000-step goal for 2 consecutive weeks. This definition was selected to distinguish temporary weekly fluctuations from sustained disruptions in walking adherence. Because weekly step performance may vary due to short-term contextual factors such as weather conditions, temporary illness, or schedule changes, a single week of goal failure may reflect natural variability rather than meaningful behavioral disengagement. In contrast, consecutive failure across 2 weeks can indicate an emerging lapse in behavioral persistence and therefore provides a more stable indicator of behavioral disruption within the weekly intervention cycle used in this study.

Statistical Analysis

Overview

Baseline characteristics were summarized using descriptive statistics. In accordance with CONSORT (Consolidated Standards of Reporting Trials) recommendations for RCTs, statistical significance testing between randomized groups was not performed. Baseline characteristics were compared between participants and nonparticipants after randomization to assess potential attrition bias.

Analysis of Weekly Goal Achievement

Generalized estimating equations (GEEs) were applied to evaluate changes in weekly goal achievement rates using logit link functions and exchangeable correlation structures to analyze the 12-week 25,000-step goal achievement status.

XSL•FO
RenderX

Analysis of Step Count Patterns

Linear mixed models (LMMs) were used to evaluate the 12-week step count patterns, including subject-specific random intercepts and random slopes, to account for within-individual correlations.

Analysis of Consecutive Goal Failure Events

Kaplan-Meier survival analysis and Cox proportional hazards models were applied to evaluate the intervention effects on consecutive 2-week goal achievement failure events.

All analyses included time, group, and time × group interactions as the main variables, with gender and age as adjustment variables. Primary analyses included a modified intention-to-treat analysis targeting 141 participants who actually participated in the intervention. Because 33 participants did not initiate the intervention and had no postbaseline outcome data, inclusion in repeated-measures analyses required at least 1 postbaseline observation. Therefore, the primary analyses followed a modified intention-to-treat approach including all participants who initiated the intervention (n=141). Sensitivity analyses using a full intention-to-treat framework including all randomized participants (n=174) with multiple imputation for missing data were additionally performed to assess robustness. Missing values were handled using multiple imputation methods. All analyses were performed using SPSS (version 28.0; IBM Corp), with statistical significance set at $P<.05$.

Ethical Considerations

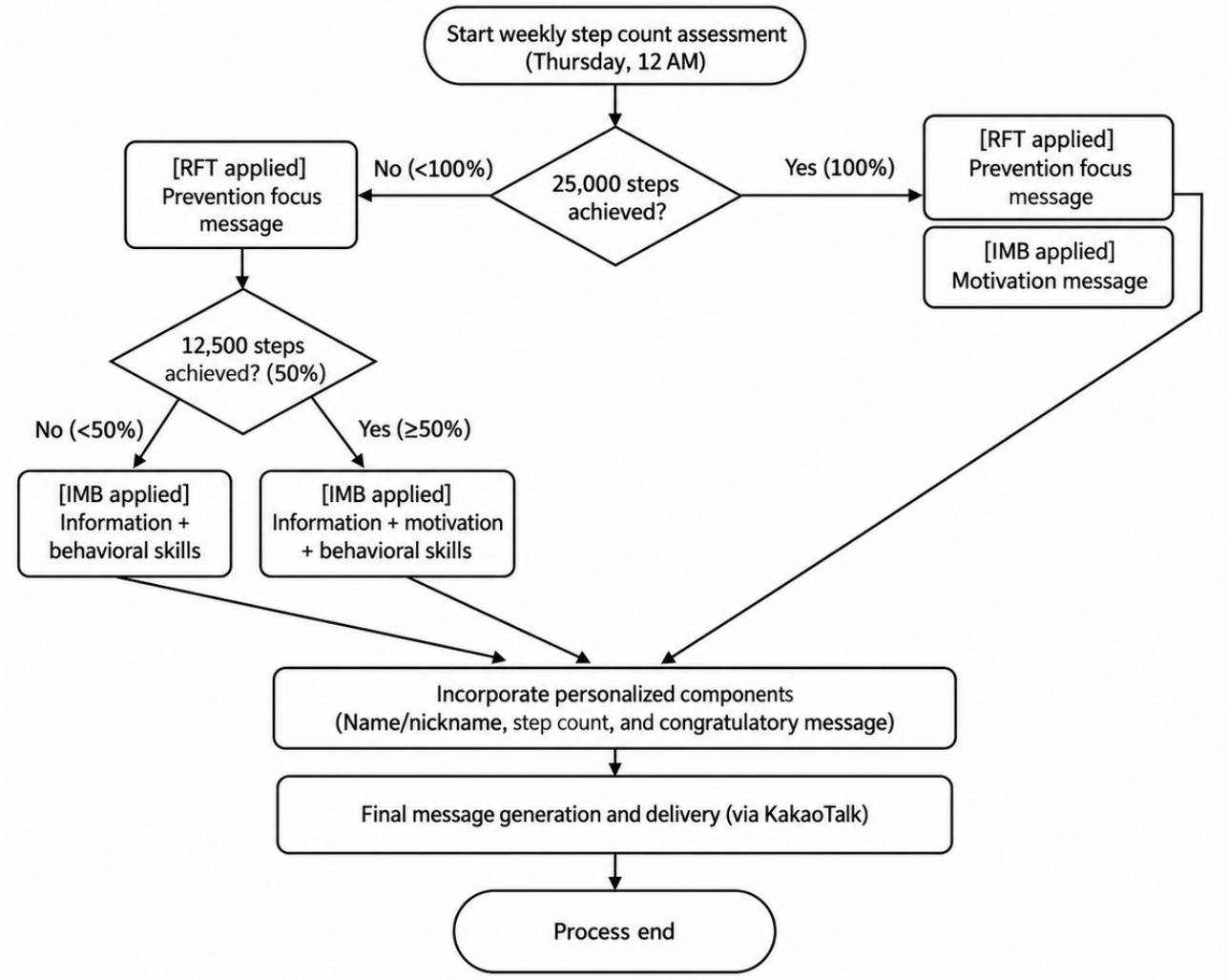
This study was approved by the Institutional Review Board of Sahmyook University (SYU 2024-02-017). All participants provided written informed consent prior to enrollment and were informed of their right to withdraw at any time without penalty. Step count data were collected through the WalkOn mobile app and deidentified prior to analysis to ensure confidentiality. Participants in both groups received identical performance-based incentives (1000-2000 points [equivalent to KRW 1000-2000] per week depending on study phase).

Results

Personalized Messaging Algorithm

The personalized messaging algorithm used to generate and deliver messages according to participants' midweek step count performance is illustrated in Figure 1.

Figure 1. Flowchart of personalized message generation process based on weekly step count assessment. IMB: Information-Motivation-Behavioral Skills; RFT: Regulatory Focus Theory.

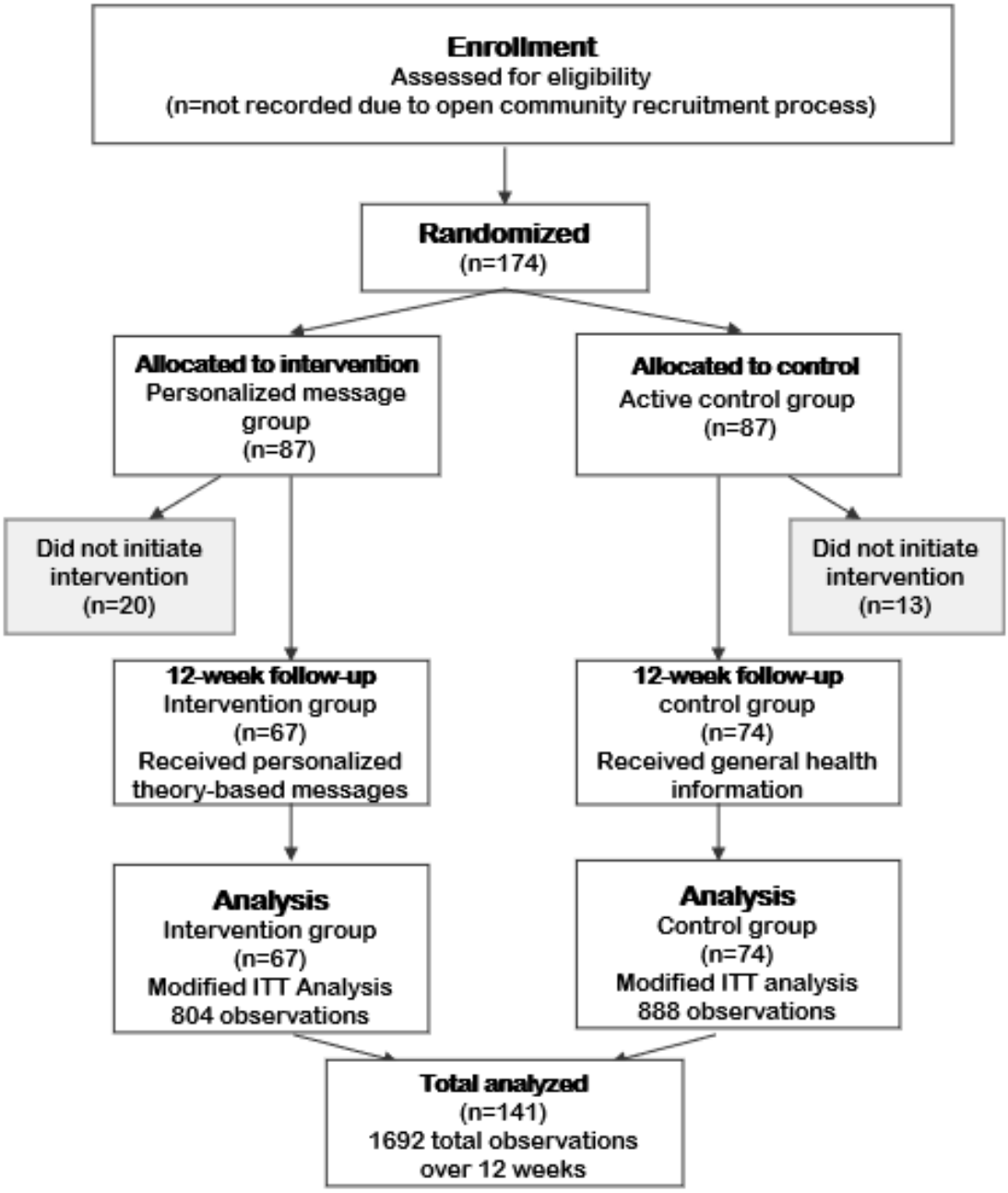


XSL•FO
RenderX

Participant Characteristics and Baseline Analysis

Of the 174 randomized participants, 141 (81%) engaged in the 12-week study, including 77% (67/87) in the intervention group and 85% (74/87) in the control group. A total of 33 participants did not participate after randomization (20/87, 23%) in the intervention group and 15% (13/87) in the control group. The participant flow is presented in Figure 2.

Figure 2. Flow of participants through the study. ITT: intention-to-treat.



Baseline characteristics of the randomized participants are presented in Table 2. In accordance with CONSORT recommendations, formal significance testing of baseline characteristics between randomized groups was not performed (CONSORT checklist is provided in Multimedia Appendix 2).

XSL•FO
RenderX

Table 2. Baseline characteristics of randomized participants and randomization verification. A total of 141 participants (67 in intervention; 74 in control) completed the 12-week intervention.

Variable	Total (n=174), n (%)	Intervention group (n=87), n (%)	Control group (n=87), n (%)
Gender			
Men	60 (34.5)	30 (35)	30 (35)
Women	114 (65.5)	57 (66)	57 (66)
Age group (years)			
20-29	17 (9.8)	10 (12)	7 (8)
30-39	26 (14.9)	11 (13)	15 (17)
40-49	58 (33.3)	24 (28)	34 (39)
50-59	54 (31)	31 (36)	23 (26)
≥60	19 (10.9)	11 (13)	8 (9)

Population Characteristics and Selection Bias Assessment

Among the 141 participants who engaged in the intervention, borderline differences in age group distribution were observed between groups (*P*=.06), whereas no significant differences were found in gender or baseline step counts.

To assess potential attrition-related selection bias, baseline characteristics were summarized using descriptive statistics for each study group (Tables S1 and S2 in Multimedia Appendix 3).

Intervention Effects on Weekly Goal Achievement Rates

Weekly goal achievement rates over the 12 weeks showed generally high levels in both groups; detailed week-by-week values are provided in Table S3 in Multimedia Appendix 4. The intervention group maintained relatively stable achievement rates during weeks 5 to 8, with the highest rate of 91% (61/67) observed in participants in weeks 4 to 6. The control group showed more variable patterns, with recovery in weeks 9 to 10 after midstudy declines.

GEE was applied to evaluate the effects of the personalized message interventions on weekly step count goal (25,000 steps) achievement rates over 12 weeks. Goal achievement status was set as a binary dependent variable (1=achievement; 0=nonachievement) using logit link functions and exchangeable correlation structures to account for internal correlations in repeated measurement data by subject. The model included time (week), group (intervention vs control), and time × group interaction as the main variables, analyzing 1692 observations (141 participants × 12 weeks).

GEE analysis results (Table 3) showed statistically significant changes in goal achievement rates over time (Wald χ^2_{11}=30.7; *P*=.001). Using week 1 as reference, most weeks did not differ significantly from week 1; however, the odds of goal achievement were significantly lower at week 12 (odds ratio [OR] 0.34, 95% CI 0.12-0.93; *P*=.04). Overall differences in goal achievement rates between intervention and control groups were not statistically significant (OR 1.12, 95% CI 0.29-4.37; *P*=.87). Additionally, time and group interaction effects were not significant (Wald χ^2_{11}=11.8; *P*=.38), confirming that personalized message interventions did not differentially influence goal achievement rate change patterns over time.

XSL•FO
RenderX

Table 3. GEE[a] model results for weekly goal achievement rates (N=1692 observations from 141 participants).

Variable	OR[b] (95% CI)	Wald chi-square (*df*)	*P value*
Time effect			
Week 1	Reference	—[c]	—
Week 2	0.81 (0.32-1.92)	0.23 (1)	.63
Week 3	0.58 (0.21-1.60)	1.11 (1)	.29
Week 4	0.80 (0.25-2.55)	0.15 (1)	.70
Week 5	0.79 (0.25-2.51)	0.17 (1)	.68
Week 6	0.57 (0.21-1.58)	1.17 (1)	.28
Week 7	0.56 (0.20-1.55)	1.24 (1)	.27
Week 8	0.65 (0.21-2.04)	0.54 (1)	.46
Week 9	1.62 (0.44-5.99)	0.52 (1)	.47
Week 10	1.59 (0.55-4.64)	0.73 (1)	.39
Week 11	0.55 (0.18-1.71)	1.07 (1)	.30
Week 12	0.34 (0.12-0.93)	4.40 (1)	.04
Intervention group (vs control)	1.12 (0.29-4.37)	0.03 (1)	.87
Time (week)	—	30.68 (11)	.001
Group	—	1.64 (1)	.20
Time × group interaction	—	11.82 (11)	.38

[a]GEE: generalized estimating equation.

[b]OR: odds ratio.

[c]Not applicable.

Step Count Patterns

LMM was applied to evaluate the effects of personalized message interventions on 12-week step count patterns. Step counts were set as continuous dependent variables, including subject-specific random intercepts and random slopes to account for within-individual correlations. The model included time (week), group (intervention vs control), and the time × group interaction as the main variables.

Weekly step counts over 12 weeks varied in both groups. The intervention group showed a lower overall mean weekly step count than the control group; detailed week-by-week values are presented in Table S4 in Multimedia Appendix 4. The intervention group showed a decreasing pattern from 57,217 steps in week 1 to 49,067 steps in week 12, whereas the control group showed a decreasing pattern from 60,940 steps in week 1 to 54,808 steps in week 12.

The intercept representing the estimated average weekly step count of the intervention group was 48,369 steps (SE 3524; *P*<.001, Table 4). Group effects representing between-group differences in average step counts were not significant (estimate 6184, SE 4856; *P*=.20), and step count changes over time (week) were also not significant (estimate 324, SE 642; *P*=.61). Moreover, the group × week interactions were not significant (estimate –417, SE 409; *P*=.31), confirming that personalized message interventions did not significantly influence step count patterns over 12 weeks.

Table 4. Linear mixed model results for weekly step count changes (N=1692 observations from 141 participants; model fit: −2 log likelihood=35,185.6; AIC[a]=35,345.6; BIC[b]=35,774.7).

Parameter	Estimate (SE)	95% CI	*P* value
Intercept	48,369 (3524)	41,437 to 55,299	<.001
Intervention group (vs control; reference: intervention)	6184 (4856)	–3369 to 15,736	.20
Week	324 (642)	–934 to 1582	.61
Group × week	–417 (409)	–1218 to 393	.31

[a]AIC: Akaike information criterion.

[b]BIC: Bayesian information criterion.

XSL•FO
RenderX

Intervention Effects on Preventing Consecutive 2-Week Goal Achievement Failure Events

Kaplan-Meier survival analysis and Cox proportional hazards models were applied to evaluate the effects of personalized message interventions on reducing the risk of consecutive 2-week goal achievement failure events. Consecutive 2-week goal achievement failure was defined as the event of interest, and time to first occurrence was used as the survival time.

During the 12-week observation period, 37% (27/74) of participants in the control group and 34% (23/67) of participants in the intervention group experienced consecutive 2-week goal achievement failure events. The average time to first consecutive failure was 4.7 weeks in the control group and 4.9 weeks in the intervention group. Detailed event frequencies and survival statistics are presented in Tables S5 and S6 in Multimedia Appendix 5.

Kaplan-Meier survival analysis demonstrated a statistically significant difference in time to consecutive 2-week goal achievement failure between groups (log-rank χ^2_1=9.9; *P*=.002; Figure 3). By the end of the 12-week follow-up, more than half of the participants in both groups had not experienced consecutive failure events; therefore, the median survival time was not reached. The intervention group showed longer survival without consecutive failure compared with the control group, indicating improved resilience against repeated goal failure.

Figure 3. Kaplan-Meier survival curves for consecutive 2-week goal achievement failure.



The Cox proportional hazards model analysis results (Table 5) showed that the intervention group had a significantly reduced risk of consecutive 2-week goal achievement failure events compared with the control group (hazard ratio [HR] 0.68, 95% CI 0.53-0.87; *P*=.002). In models adjusted for gender and age group, the intervention group still had a significantly lower risk of consecutive 2-week goal achievement failure events than the control group (HR 0.67, 95% CI 0.52-0.86; *P*=.002), representing an approximately 32.9% risk reduction.

Table 5. Cox proportional hazards model results for consecutive 2-week goal achievement failure events.

Variable	Unadjusted model, HR[a] (95% CI)	Adjusted Model[b], HR (95% CI)	*P* value
Intervention group (vs control)	0.68 (0.53-0.87)	0.67 (0.52-0.86)	.002
Women (vs men)	—[c]	0.43 (0.33–0.55)	<.001
Age group (years; reference: 20-29)			
30-39	—	1.46 (0.76–2.83)	.26
40-49	—	1.48 (0.80–2.73)	.21
50-59	—	2.00 (1.09–3.67)	.03
≥60	—	1.62 (0.83–3.17)	.16

[a]HR: hazard ratio.

[b]Adjusted for gender and age group.

[c]Not applicable.

Among the adjustment variables, women had a significantly lower risk of consecutive 2-week goal achievement failure events than men (HR 0.43, 95% CI 0.33-0.55; *P*<.001), and by age group, participants aged 50-59 years had approximately twice the risk of consecutive 2-week goal achievement failure events compared with those aged 20-29 years (HR 2.00, 95% CI 1.09-3.67; *P*=.03).

Kaplan-Meier survival curves also showed that the intervention group maintained a higher survival probability for consecutive

XSL•FO
RenderX

2-week goal achievement failure events than the control group throughout the study period (Figure 2), consistent with the Cox proportional hazards model results.

Sensitivity Analyses

Sensitivity analyses including all randomized participants (n=174) yielded results consistent in direction with the primary analysis. The intervention group showed a lower risk of consecutive 2-week goal achievement failure (HR 0.67, 95% CI 0.29-1.58), although this was not statistically significant (*P*=.37).

Discussion

Interpretation of Main Results

This RCT found that personalized theory-based messages did not significantly improve overall weekly step goal achievement rates or step count patterns over 12 weeks but significantly reduced the risk of consecutive 2-week goal achievement failures (HR 0.67, 95% CI 0.52-0.86; *P*=.002). These findings suggest that personalized messaging may enhance behavioral resilience by preventing repeated goal failures, even if it does not substantially increase overall performance levels.

These results are particularly noteworthy in the context of the study design. The control group in this study was not a simple comparison group but received substantial active intervention. The control group participants received the same progressive incentive system as the intervention group (weeks 1-4: 1000 points [KRW 1000]; weeks 5-8: 1500 points [KRW 1500]; weeks 9-12: 2000 points [KRW 2000]) and continuously received general health information through weekly walking and health promotion card news. They also experienced social comparison and competitive motivation through real-time participant step count comparison features via the WalkOn app and received the same weekly 25,000-step goal as the intervention group.

The significant effect of personalized messages on consecutive failure prevention, even in this strong control group intervention environment, suggests the unique value of personalization. Particularly, despite both groups maintaining high goal achievement rates (intervention group: 87%, control group: 87%), the approximately 33% reduction in consecutive 2-week goal achievement failure events in the intervention group demonstrates that personalized messages play a special role in improving “failure resilience” [43,44].

Compared with existing research, the results of this study have a deeper meaning. While most digital walking intervention studies have reported only short-term step count increases using minimal intervention control groups [11,12], this study verified the pure effects of personalized messages through an active control group design. This represents an important contribution by clearly identifying the actual effects of personalized interventions while addressing the methodological limitations of existing research. Consecutive failure prevention effects suggest that personalized feedback positively influences behavioral sustainability [45-47], and confirming these results under rigorous control conditions demonstrates the unique contribution of this study.

Recent evidence has also demonstrated the effectiveness of personalized digital messaging for increasing physical activity [48]. For example, an RCT using a reinforcement learning–based adaptive messaging system reported that personalized messages tailored to participants’ behavioral data significantly increased daily step counts compared with nonpersonalized messaging approaches. These findings support the broader evidence that personalization strategies can enhance behavioral engagement and physical activity in digital health interventions.

However, unlike previous studies that primarily focused on increasing overall step counts, the present study evaluated whether personalized theory-based messages could prevent repeated weekly goal achievement failure. While our intervention did not significantly increase mean step counts compared with the control condition, it significantly reduced the risk of consecutive goal achievement failure. This suggests that personalized feedback may play a particularly important role in sustaining behavioral adherence and preventing disengagement rather than simply increasing activity volume.

The goal achievement rate increase pattern observed during the middle period (weeks 4-8) is particularly interesting. This suggests that personalized messages may have acted as protective factors when the novelty effects were diminishing [49]. The maximum 6 percentage point achievement rate difference at week 6 demonstrates that personalized messages are particularly effective during the “dropout risk period” of intervention participation [49,50]. This emphasizes the need for individualized motivational support, which cannot be sufficiently provided through general health information or social comparison alone [51].

Theoretical Contribution

This study contributes theoretically by systematically implementing differentiated messaging strategies according to individual performance levels by applying a theoretical framework that integrates the IMB model and RFT. While existing research primarily relied on single theories [52-54], this study attempted an adaptive approach that comprehensively addresses information, motivation, and behavioral skills while dynamically adjusting prevention- or promotion-focused messages according to individual performance changes.

The preventive effects on consecutive 2-week goal achievement failure events align with the theoretical predictions that RFT’s prevention focus supports behavioral sustainability through loss avoidance motivation [32]. Particularly, prevention-focused messages provided to participants with fewer than 12,500 steps may have effectively prevented goal abandonment. This demonstrates the importance of theory-based approaches that consider individual motivational orientations beyond simple performance feedback in personalized messages [53,55,56].

Practical Implications

The practical implications of this study are as follows. First, the finding that personalized messages are effective in preventing consecutive failures provides important insights for developing user retention strategies in the mobile health care industry. Considering the high dropout rates (over 40%) observed in existing wearable devices and smartphone applications,

XSL•FO
RenderX

according to a meta-research study [57], the responsive messaging system developed in this study can be used as a cost-effective solution for improving user engagement.

Second, the differential effects of gender and age group (women: HR 0.43, 95% CI 0.33-0.55; age group 50-59 years: HR 2.00, 95% CI 1.09-3.67) emphasize the importance of personalized approaches that consider demographic characteristics in personalized intervention design [58,59]. This study provides evidence for establishing differentiated strategies for target populations in national-level physical activity promotion policies [60].

Third, the weekly performance-based message reflection system developed in this study can be extended to other areas of health behavior change, such as chronic disease management, smoking cessation, and dietary improvement. Particularly, it can provide important evidence for the systematic application of behavior change theories and personalized algorithm designs in digital therapeutics development [61].

Study Limitations

This study has several limitations. First, although an a priori sample size calculation was conducted, the achieved sample size was smaller than the original recruitment target. As a result, statistical power to detect small between-group differences may have been limited, and nonsignificant findings for overall goal achievement and step count patterns should be interpreted cautiously. Therefore, this trial should not be interpreted as a fully powered confirmatory trial for detecting small between-group differences. Second, the relatively short 12-week follow-up period limited the ability to assess long-term maintenance of behavior change. Given the sharp decline in engagement after 6 months reported in previous studies [62], longer-term follow-up trials are warranted. Third, higher nonparticipation rates were observed in the intervention group (24.1% vs 14.9%). This difference may reflect the additional engagement burden associated with personalized messaging or the psychological pressure of continuous performance monitoring [63]. Although no significant baseline differences were observed between participants and nonparticipants, residual attrition-related bias due to postrandomization nonparticipation cannot be excluded. Fourth, participants were recruited from a single urban community in Mapo-gu, Seoul, which may limit generalizability to other regional or cultural contexts. Moreover, the trial included progressive monetary incentives and weekly researcher monitoring, conditions that may not be replicated in routine real-world implementation. Fifth, step counts were measured using built-in smartphone accelerometers, which may not capture all forms of physical activity and may vary in accuracy across devices. Participants were not blinded to group allocation, which may have introduced performance or expectation bias, a common limitation in eHealth trials. In addition, although primary outcomes were prespecified, the analysis of multiple outcomes raises the possibility of an inflated type I error.

Future Research Directions

Based on the results of this study, we propose the following follow-up research. First, long-term follow-up studies of 6 months or more should evaluate the sustained effects of personalized messages and their impact on habit formation. Second, research is needed to develop machine learning-based adaptive messaging systems that learn individual real-time response patterns and implement more sophisticated personalization. Third, generalizability should be verified through multisite RCTs targeting diverse populations (rural areas, older adults, and patients with chronic diseases). Fourth, research on optimizing message delivery methods (push notifications, chatbots, and voice messages) and message transmission frequency is required. Finally, research is necessary to comprehensively evaluate message acceptability, fatigue, privacy concerns, and so on through user experience research to increase implementation feasibility in actual clinical settings.

Conclusion

This study empirically evaluated the effects of personalized message interventions that integrate the IMB model and RFT on walking behavior maintenance. Although these interventions did not significantly affect the overall increase in step counts, they were confirmed to be effective in preventing consecutive goal achievement failures. Personalized tailored messages did not significantly affect the overall increase in step counts but had significant effects in preventing consecutive 2-week goal achievement failures. This demonstrates that personalized messages are important variables that enhance behavioral sustainability rather than simple performance improvement. This study provides important evidence for establishing user retention strategies in digital health care intervention design, showing that personalized feedback can contribute to improving user resilience and engagement sustainability. Performance-based personalized messaging strategies can be applied to various health behavior change areas for developing digital therapeutics or healthy lifestyle habits in the future. Future research should assess the long-term sustainability of personalized messaging and determine how these findings can support the refinement of adaptive digital health personalization systems.

Acknowledgments

We thank all participants and community collaborators who contributed to this study. The authors used ChatGPT (OpenAI) for translation and language editing during manuscript preparation. No AI tool was used for data analysis, data interpretation, or generation of scientific conclusions. All AI-assisted content was critically reviewed, revised, and approved by the authors, who take full responsibility for the integrity and accuracy of the manuscript.

XSL•FO
RenderX

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT; MSIT; grant number RS-2023-00213294). This study was also supported by the Korea Disease Control and Prevention Agency and the Seoul Metropolitan Government (grant number 2022-11-001). The funders had no involvement in the study design, data collection, analysis, interpretation of the data, or the writing of the manuscript.

Authors' Contributions

SJ contributed to conceptualization, data curation, formal analysis, and software. SJ and AS contributed to funding acquisition, methodology, and writing of the original draft. All authors contributed to data interpretation, critical revision of the manuscript for important intellectual content, and final approval of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Examples of personalized messages sent by performance level.
[DOCX File , 16 KB - mhealth_v14i1e84430_app1.docx]

Multimedia Appendix 2
CONSORT checklist.
[PDF File (Adobe PDF File), 3187 KB - mhealth_v14i1e84430_app2.pdf]

Multimedia Appendix 3
Participant/nonparticipant comparison and attrition bias assessment.
[DOCX File , 17 KB - mhealth_v14i1e84430_app3.docx]

Multimedia Appendix 4
Detailed weekly step count data and weekly mean step counts.
[DOCX File , 18 KB - mhealth_v14i1e84430_app4.docx]

Multimedia Appendix 5
Additional survival analysis details.
[DOCX File , 18 KB - mhealth_v14i1e84430_app5.docx]

References

1. Ungvari Z, Fazekas-Pongor V, Csiszar A, Kunutsor SK. The multifaceted benefits of walking for healthy aging: from blue zones to molecular mechanisms. Geroscience 2023;45(6):3211-3239 [FREE Full text] [doi: 10.1007/s11357-023-00873-8] [Medline: 37495893]
2. Schaller A, Stassen G, Baulig L, Lange M. Physical activity interventions in workplace health promotion: objectives, related outcomes, and consideration of the setting-a scoping review of reviews. Front Public Health 2024;12:1353119 [FREE Full text] [doi: 10.3389/fpubh.2024.1353119] [Medline: 38406496]
3. Ryu S, Rodriguez-Gonzalez P. A review of health wearable-based physical activity interventions among children and adolescents. Int J Phys Act Health 2023;2(2):6. [doi: 10.18122/ijpah.020206.boisestate]
4. Tully MA, Cupples ME, Hart ND, McEneny J, McGlade KJ, Chan W, et al. Randomised controlled trial of home-based walking programmes at and below current recommended levels of exercise in sedentary adults. J Epidemiol Community Health 2007;61(9):778-783 [FREE Full text] [doi: 10.1136/jech.2006.053058] [Medline: 17699531]
5. Powell KE, King AC, Buchner DM, Campbell WW, DiPietro L, Erickson KI, et al. The scientific foundation for the physical activity guidelines for Americans, 2nd edition. J Phys Act Health 2018;16(1):1-11 [FREE Full text] [doi: 10.1123/jpah.2018-0618] [Medline: 30558473]
6. Winstein CJ, Stein J, Arena R, Bates B, Cherney LR, Cramer SC, et al. Guidelines for adult stroke rehabilitation and recovery: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2016;47(6):e98-e169. [doi: 10.1161/STR.0000000000000098] [Medline: 27145936]

XSL•FO
RenderX

7. Lee I, Shiroma EJ, Kamada M, Bassett DR, Matthews CE, Buring JE. Association of step volume and intensity with all-cause mortality in older women. JAMA Intern Med 2019;179(8):1105-1112 [FREE Full text] [doi: 10.1001/jamainternmed.2019.0899] [Medline: 31141585]
8. Gupta A, Bivina G, Parida M. Does neighborhood design matter for walk access to metro stations? An integrated SEM-Hybrid discrete mode choice approach. Transp Policy (Oxf) 2022;121:61-77. [doi: 10.1016/j.tranpol.2022.03.010]
9. Organization WH. Global Status Report on Physical Activity 2022. Geneva, Switzerland: World Health Organization; 2022.
10. Page EJ, Massey AS, Prado-Romero PN, Albadawi S. The use of self-monitoring and technology to increase physical activity: a review of the literature. Perspect Behav Sci 2020;43(3):501-514 [FREE Full text] [doi: 10.1007/s40614-020-00260-0] [Medline: 33029577]
11. Chaudhry UAR, Wahlich C, Fortescue R, Cook DG, Knightly R, Harris T. The effects of step-count monitoring interventions on physical activity: systematic review and meta-analysis of community-based randomised controlled trials in adults. Int J Behav Nutr Phys Act 2020;17(1):129 [FREE Full text] [doi: 10.1186/s12966-020-01020-8] [Medline: 33036635]
12. Okamoto M, Saito Y, Nakamura S, Nagasawa M, Shibuya M, Nagasaka G, et al. Smartphone-based digital peer support for a walking intervention among public officers in Kanagawa Prefecture: single-arm pre- and postintervention evaluation. JMIR Form Res 2024;8:e53759 [FREE Full text] [doi: 10.2196/53759] [Medline: 39316793]
13. Hargreaves EA, Mutrie N, Fleming JD. A web-based intervention to encourage walking (StepWise): pilot randomized controlled trial. JMIR Res Protoc 2016;5(1):e14 [FREE Full text] [doi: 10.2196/resprot.4288] [Medline: 26810251]
14. Shin G, Feng Y, Jarrahi MH, Gafinowitz N. Beyond novelty effect: a mixed-methods exploration into the motivation for long-term activity tracker use. JAMIA Open 2019;2(1):62-72 [FREE Full text] [doi: 10.1093/jamiaopen/ooy048] [Medline: 31984346]
15. Heizmann A, Chapelle C, Laporte S, Roche F, Hupin D, Le Hello C. Impact of wearable device-based interventions with feedback for increasing daily walking activity and physical capacities in cardiovascular patients: a systematic review and meta-analysis of randomised controlled trials. BMJ Open 2023;13(7):e069966 [FREE Full text] [doi: 10.1136/bmjopen-2022-069966] [Medline: 37433730]
16. Xu Y, Yue L, Wang W, Wu X, Liang Z. Gender-specific impact of self-monitoring and social norm information on walking behavior among Chinese college students assessed using WeChat: longitudinal tracking study. J Med Internet Res 2021;23(12):e29167 [FREE Full text] [doi: 10.2196/29167] [Medline: 34878992]
17. Mazéas A, Blond M, Chalabaev A, Duclos M. Validity and reliability of an app-based medical device to empower individuals in evaluating their physical capacities. PLoS One 2023;18(8):e0289874 [FREE Full text] [doi: 10.1371/journal.pone.0289874] [Medline: 37561737]
18. Susaiyah APS. Feedback-driven insight generation and recommendation for health self-management. TechRxiv. Preprint posted online on August 15, 2023 . [doi: 10.36227/techrxiv.23931570.v1]
19. Javed A, Kim DS, Hershman SG, Shcherbina A, Johnson A, Tolas A, et al. Personalized digital behaviour interventions increase short-term physical activity: a randomized control crossover trial substudy of the MyHeart Counts Cardiovascular Health Study. Eur Heart J Digit Health 2023;4(5):411-419 [FREE Full text] [doi: 10.1093/ehjdh/ztad047] [Medline: 37794870]
20. Rei D, Clavel C, Martin J, Ravenet B. Adapting goals and motivational messages on smartphones for motivation to walk. Smart Health (Amst) 2024;32:100482. [doi: 10.1016/j.smhl.2024.100482]
21. Novak D. A serious game (MyDiabetic) to support children's education in type 1 diabetes mellitus: iterative participatory co-design and feasibility study. JMIR Serious Games 2024;12:e49478 [FREE Full text] [doi: 10.2196/49478] [Medline: 38713496]
22. Adamakis M. Validity of wearable monitors and smartphone applications for measuring steps in semi-structured and free-living settings. Technologies (Basel) 2023;11(1):29. [doi: 10.3390/technologies11010029]
23. Bertapelli F, Curtis JS, Carlson B, Johnson M, Abadie B, Agiovlasitis S. Step-counting accuracy of activity monitors in persons with Down syndrome. J Intellect Disabil Res 2019;63(1):21-30. [doi: 10.1111/jir.12550] [Medline: 30239068]
24. Moy ML, Martinez CH, Kadri R, Roman P, Holleman RG, Kim HM, et al. Long-term effects of an internet-mediated pedometer-based walking program for chronic obstructive pulmonary disease: randomized controlled trial. J Med Internet Res 2016;18(8):e215 [FREE Full text] [doi: 10.2196/jmir.5622] [Medline: 27502583]
25. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
26. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
27. Fisher JD, Fisher WA. Changing AIDS-risk behavior. Psychol Bull 1992;111(3):455-474. [doi: 10.1037/0033-2909.111.3.455] [Medline: 1594721]
28. Klonoff DC. Behavioral theory: the missing ingredient for digital health tools to change behavior and increase adherence. J Diabetes Sci Technol 2019;13(2):276-281 [FREE Full text] [doi: 10.1177/1932296818820303] [Medline: 30678472]

XSL•FO
RenderX

29. Fisher JD, Amico KR, Fisher WA, Harman JJ. The information-motivation-behavioral skills model of antiretroviral adherence and its applications. Curr HIV/AIDS Rep 2008;5(4):193-203. [doi: 10.1007/s11904-008-0028-y] [Medline: 18838059]
30. Sadat Rezai L, Chin J, Casares-Li R, He F, Bassett-Gunter R, Burns C. Can message-tailoring based on regulatory fit theory improve the efficacy of persuasive physical activity systems? Information 2019;10(11):347. [doi: 10.3390/info10110347]
31. Latimer AE, Rivers SE, Rench TA, Katulak NA, Hicks A, Hodorowski JK, et al. A field experiment testing the utility of regulatory fit messages for promoting physical activity. J Exp Soc Psychol 2008;44(3):826-832 [FREE Full text] [doi: 10.1016/j.jesp.2007.07.013] [Medline: 19543443]
32. Higgins ET. Beyond pleasure and pain. Am Psychol 1997;52(12):1280-1300. [doi: 10.1037//0003-066x.52.12.1280] [Medline: 9414606]
33. Watling C, Driessen E, van der Vleuten CPM, Vanstone M, Lingard L. Understanding responses to feedback: the potential and limitations of regulatory focus theory. Med Educ 2012;46(6):593-603. [doi: 10.1111/j.1365-2923.2012.04209.x] [Medline: 22626051]
34. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020;54(24):1451-1462 [FREE Full text] [doi: 10.1136/bjsports-2020-102955] [Medline: 33239350]
35. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
36. Chang SJ, Choi S, Kim S, Song M. Intervention strategies based on information-motivation-behavioral skills model for health behavior change: A systematic review. Asian Nurs Res 2014;8(3):172-181. [doi: 10.1016/j.anr.2014.08.002]
37. Ludolph R, Schulz PJ. Does regulatory fit lead to more effective health communication? A systematic review. Soc Sci Med 2015;128:142-150. [doi: 10.1016/j.socscimed.2015.01.021] [Medline: 25617673]
38. Marshall SJ, Levy SS, Tudor-Locke CE, Kolkhorst FW, Wooten KM, Ji M, et al. Translating physical activity recommendations into a pedometer-based step goal: 3000 steps in 30 minutes. Am J Prev Med 2009;36(5):410-415. [doi: 10.1016/j.amepre.2009.01.021] [Medline: 19362695]
39. Tudor-Locke C, Sisson SB, Collova T, Lee SM, Swan PD. Pedometer-determined step count guidelines for classifying walking intensity in a young ostensibly healthy population. Can J Appl Physiol 2005;30(6):666-676. [doi: 10.1139/h05-147] [Medline: 16485518]
40. Tudor-Locke C, Aguiar EJ, Han H, Ducharme SW, Schuna JM, Barreira TV, et al. Walking cadence (steps/min) and intensity in 21-40 year olds: CADENCE-adults. Int J Behav Nutr Phys Act 2019;16(1):8 [FREE Full text] [doi: 10.1186/s12966-019-0769-6] [Medline: 30654810]
41. Tudor-Locke C, Ducharme SW, Aguiar EJ, Schuna JM, Barreira TV, Moore CC, et al. Walking cadence (steps/min) and intensity in 41 to 60-year-old adults: the CADENCE-adults study. Int J Behav Nutr Phys Act 2020;17(1):137 [FREE Full text] [doi: 10.1186/s12966-020-01045-z] [Medline: 33168018]
42. Tudor-Locke C, Craig CL, Brown WJ, Clemes SA, De Cocker K, Giles-Corti B, et al. How many steps/day are enough? For adults. Int J Behav Nutr Phys Act 2011;8:79 [FREE Full text] [doi: 10.1186/1479-5868-8-79] [Medline: 21798015]
43. Tielman ML, Neerincx MA, Brinkman W. Design and evaluation of personalized motivational messages by a virtual agent that assists in post-traumatic stress disorder therapy. J Med Internet Res 2019;21(3):e9240 [FREE Full text] [doi: 10.2196/jmir.9240] [Medline: 30916660]
44. Hsu TC, Whelan P, Gandrup J, Armitage CJ, Cordingley L, McBeth J. Personalized interventions for behaviour change: a scoping review of just-in-time adaptive interventions. Br J Health Psychol 2025;30(1):e12766. [doi: 10.1111/bjhp.12766] [Medline: 39542743]
45. Cadmus-Bertram LA, Marcus BH, Patterson RE, Parker BA, Morey BL. Randomized trial of a Fitbit-based physical activity intervention for women. Am J Prev Med 2015;49(3):414-418 [FREE Full text] [doi: 10.1016/j.amepre.2015.01.020] [Medline: 26071863]
46. Cadmus-Bertram L, Marcus BH, Patterson RE, Parker BA, Morey BL. Use of the Fitbit to measure adherence to a physical activity intervention among overweight or obese, postmenopausal women: self-monitoring trajectory during 16 weeks. JMIR Mhealth Uhealth 2015;3(4):e96 [FREE Full text] [doi: 10.2196/mhealth.4229] [Medline: 26586418]
47. Ellingson LD, Lansing JE, DeShaw KJ, Peyer KL, Bai Y, Perez M, et al. Evaluating motivational interviewing and habit formation to enhance the effect of activity trackers on healthy adults' activity levels: randomized intervention. JMIR Mhealth Uhealth 2019;7(2):e10988 [FREE Full text] [doi: 10.2196/10988] [Medline: 30762582]
48. Aguilera A, Arévalo Avalos M, Xu J, Chakraborty B, Figueroa C, Garcia F, et al. Effectiveness of a digital health intervention leveraging reinforcement learning: results from the Diabetes and Mental Health Adaptive Notification Tracking and Evaluation (DIAMANTE) randomized clinical trial. J Med Internet Res 2024;26:e60834 [FREE Full text] [doi: 10.2196/60834] [Medline: 39378080]
49. Schroé H, Crombez G, De Bourdeaudhuij I, Van Dyck D. Investigating when, which, and why users stop using a digital health intervention to promote an active lifestyle: secondary analysis with a focus on health action process approach–based psychological determinants. JMIR Mhealth Uhealth 2022;10(1):e30583 [FREE Full text] [doi: 10.2196/30583] [Medline: 35099400]

XSL•FO
RenderX

50. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
51. Sharpe EE, Karasouli E, Meyer C. Examining factors of engagement with digital interventions for weight management: rapid review. JMIR Res Protoc 2017;6(10):e205 [FREE Full text] [doi: 10.2196/resprot.6059] [Medline: 29061557]
52. Choi J, Park YC, Choi S. Development of a mobile-based self-management health alarm program for obese children in South Korea and a test of its feasibility for metabolic outcomes: a study based on the information-motivation-behavioral skills model. Child Health Nurs Res 2021;27(1):13-23 [FREE Full text] [doi: 10.4094/chnr.2021.27.1.13] [Medline: 35004493]
53. Ferrari M, Speight J, Beath A, Browne JL, Mosely K. The information-motivation-behavioral skills model explains physical activity levels for adults with type 2 diabetes across all weight classes. Psychol Health Med 2021;26(3):381-394. [doi: 10.1080/13548506.2020.1749292] [Medline: 32266821]
54. Fuglestad PT, Rothman AJ, Linde JA. Applying regulatory focus theory to encourage weight loss in a self-directed intervention. Int J Behav Med 2024;31(2):331-337. [doi: 10.1007/s12529-023-10180-3] [Medline: 37118346]
55. Pfeffer I. Regulatory fit messages and physical activity motivation. J Sport Exerc Psychol 2013;35(2):119-131. [doi: 10.1123/jsep.35.2.119] [Medline: 23535971]
56. Suarez JR, Blount A, Lafontant K, Park JH, Xie R, Lighthall N, et al. Association between prevention focus and sedentary behavior in older adults: cross-sectional study. Interact J Med Res 2025;14:e63280 [FREE Full text] [doi: 10.2196/63280] [Medline: 40313184]
57. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
58. Wang H, Lehman JD. Using achievement goal-based personalized motivational feedback to enhance online learning. Educ Technol Res Dev 2021;69(2):553-581. [doi: 10.1007/s11423-021-09940-3]
59. Mateo-Orcajada A, Vaquero-Cristóbal R, Abenza-Cano L. Gender and academic year as moderators of the efficacy of mobile app interventions to promote physical activity in adolescents: a randomized controlled trial. Humanit Soc Sci Commun 2023;10(1):1-15 [FREE Full text] [doi: 10.1057/s41599-023-02502-3]
60. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
61. Lauffenburger JC, Yom-Tov E, Keller PA, McDonnell ME, Crum KL, Bhatkhande G, et al. The impact of using reinforcement learning to personalize communication on medication adherence: findings from the REINFORCE trial. NPJ Digit Med 2024;7(1):39 [FREE Full text] [doi: 10.1038/s41746-024-01028-5] [Medline: 38374424]
62. Mönninghoff A, Kramer JN, Hess AJ, Ismailova K, Teepe GW, Tudor Car L, et al. Long-term effectiveness of mHealth physical activity interventions: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2021;23(4):e26699 [FREE Full text] [doi: 10.2196/26699] [Medline: 33811021]
63. Jesuthasan J, Low M, Ong T. The impact of personalized human support on engagement with behavioral intervention technologies for employee mental health: an exploratory retrospective study. Front Digit Health 2022;4:846375 [FREE Full text] [doi: 10.3389/fdgth.2022.846375] [Medline: 35574254]

Abbreviations

BCT: behavior change technique
CONSORT: Consolidated Standards of Reporting Trials
CRIS: Clinical Research Information Service
GEE: generalized estimating equation
HR: hazard ratio
IMB: Information-Motivation-Behavioral Skills
LMM: linear mixed models
MET: metabolic equivalent
OR: odds ratio
RCT: randomized controlled trial
RFT: regulatory focus theory
WHO: World Health Organization

XSL•FO
RenderX

Edited by L Buis; submitted 19.Sep.2025; peer-reviewed by M Arévalo Avalos, YTD Cheung; comments to author 09.Jan.2026; revised version received 15.Jun.2026; accepted 15.Jun.2026; published 17.Sep.2026.

Please cite as:
Jang S, Sohn A
Personalized Mobile App–Based Messaging and Step Tracking for Improving Walking Behavior in Adults: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e84430
URL: https://mhealth.jmir.org/2026/1/e84430
doi:10.2196/84430
PMID:

©Sarang Jang, Aeree Sohn. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 17.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Behavioral Interactions Shaping the Use of a Mobile Health Intervention for Diabetic Foot Self-Care Among People With Diabetes and Caregivers: COM-B and TDF Qualitative Analysis

Sofia Malena Rebull[1], BA; Guillermo Almeida-Huanca[1], BA; Josselyn Jauregui Mantilla[1], BA; Lourdes Cruzado-Castro[1], MPH, MPM; Sebastian García Cruz[1], MA; Pablo Perel[2], MD, MSc, PhD; Juan Jaime Miranda[1,2,3], MD, MSc, PhD; Juned Siddique[4], MS, DrPH; Jacqueline Seiglie[5], MSc, MD; Hueiming Liu[3,6], MD, MPH, PhD; Stephen Jan[6,7], BEc, MEc, PhD; Laura Downey[6,7], PhD; David Beran[8], MPH, PhD; David G Armstrong[9], MD, PhD, DPM, MSc; German Comina[10], MSc, PhD; María Lazo-Porras[1], MD, MSc, PhD; Jessica Hanae Zafra-Tanaka[1], MD, PhD

[1]CRONICAS Center of Excellence in Chronic Diseases, Universidad Peruana Cayetano Heredia, Av. Armendariz 445, Miraflores, Lima, Lima Province, Peru

[10]Research Group in Engineering for Health and the Environment (SAMA-UNI), National University of Engineering Peru, Lima, Lima Province, Peru

[2]Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, England, United Kingdom

[3]Sydney School of Public Health, Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia

[4]Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[5]Department of Medicine, Massachusetts General Hospital, Boston, MA, United States

[6]The George Institute for Global Health, UNSW Sydney, Sydney, New South Wales, Australia

[7]The George Institute for Global Health, Imperial College London, London, England, United Kingdom

[8]Division of Tropical and Humanitarian Medicine, Faculty of Medicine, University of Geneva and Geneva University Hospitals, Geneva, Switzerland

[9]Department of Surgery, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Jessica Hanae Zafra-Tanaka, MD, PhD
CRONICAS Center of Excellence in Chronic Diseases, Universidad Peruana Cayetano Heredia, Av. Armendariz 445, Miraflores, Lima, Lima Province, Peru

Abstract

Background: Mobile health (mHealth) interventions show promise for supporting diabetic foot self-care; however, sustained engagement remains a challenge, particularly in low- and middle-income countries. While educational strategies are commonly used to address diabetic foot ulcers, they often fail to translate knowledge into long-term self-care. Understanding behavioral mechanisms underlying mHealth use, including the role of caregivers, is critical for improving implementation and effectiveness.

Objective: This study explored how behavioral interactions between people with diabetes and caregivers shape the use of an mHealth app for diabetic foot self-care.

Methods: This qualitative explanatory study was embedded within a parent study of an mHealth app for diabetic foot care that was conducted in primary care settings in Peru. It used the COM-B (capability, opportunity, motivation–behavior) model and the Theoretical Domains Framework (TDF) to identify mechanisms relevant for intervention design and scale-up. Semistructured interviews were conducted with people with diabetes and caregivers about their experiences related to accessing the app. Data were analyzed using deductive and inductive thematic analysis informed by the COM-B and TDF.

Results: In total, 14 people with diabetes (median age 54 y, 8 women and 6 men) and 10 caregivers (median age 28.5 y, all women) were interviewed. The app use depended on varying interactions between COM-B domains, while other behavioral determinants aligned with a single domain. Across domains, interactions between people with diabetes and caregivers also affected the app use. For example, caregivers often compensated for physical capability limitations through foot positioning or environmental adjustments. Environmental context and resources, including internet access, lighting, and device compatibility, affected the app use for all participants. Motivation for all participants was largely driven by beliefs about the app's preventive value, although reassurance from health professionals and altruism also emerged.

Conclusions: The app use often depended on collaboration between people with diabetes and caregivers as a dyad, rather than on independent users. This highlights the need to design digital interventions around caregiver demographics, competing responsibilities, and the additional responsibilities an app may introduce. Participants' preference for individual follow-up suggests

XSL•FO
RenderX

that nonphysician health personnel and other human support may help maintain continuity and reassurance alongside digital monitoring. Although the app was an early prototype intended for algorithmic training, participant experiences provided insight into real-world usability and contextual barriers for future app development.

(JMIR Mhealth Uhealth 2026;14:e101195) doi:10.2196/101195

KEYWORDS

diabetic foot; TDF; mobile health intervention; AI; low- and middle-income countries; behavioral change; facilitators; barriers; COM-B model

Introduction

It is estimated that up to one-third of people with diabetes will develop a diabetic foot ulcer (DFU) in their lifetime [1]. DFU remains one of the most severe and preventable complications of diabetes, requiring continuous self-care behaviors such as routine foot inspection and early detection of preulcerative lesions [1,2]. Mobile health (mHealth) technologies have emerged as promising tools to support these behaviors by enabling remote monitoring, patient education, and self-management support [3]. Systematic reviews suggest that mobile app–assisted interventions can improve diabetes management outcomes and facilitate self-care behaviors; however, sustained engagement with these technologies remains inconsistent in real-world settings [4,5]. Many interventions report declining use over time, indicating that the availability of digital tools alone is insufficient to ensure long-term behavioral adoption [6-8].

Caregivers frequently play a critical role in supporting diabetes self-management, particularly in low- and middle-income countries, where family members often assist with daily health tasks, monitoring, and decision-making [9-11]. For individuals who are older or have mobility limitations, visual impairments, or lower digital literacy, caregiver involvement may become essential for technological support and the navigation of app interfaces. Despite this reality, most digital health interventions conceptualize technology use as an individual behavior, rarely examining caregivers as explicit behavioral actors influencing adoption and sustained use.

Previous studies on mHealth interventions for diabetic foot prevention have primarily focused on clinical outcomes, technological performance, or usability assessments [12,13]. While these studies provide important evidence on feasibility and acceptability, they often overlook the behavioral mechanisms that determine whether individuals incorporate digital tools into everyday routines. Moreover, the factors that influence the behavior of caregivers of people with diabetes have not been explored in this context. As a result, there remains a limited understanding of how sustained engagement with mHealth technologies for diabetic foot self-monitoring is influenced by personal, physical, and social environments [7].

To better understand and improve sustained engagement with these types of interventions, behavioral theory can provide a structured framework for examining determinants of technology use. The COM-B (capability, opportunity, motivation–behavior) model conceptualizes behavior as the result of interactions between individuals' capabilities, opportunities, and motivations [14]. The Theoretical Domains Framework (TDF) complements COM-B by providing additional granularity to explore underlying mechanisms, such as knowledge, skills, social influences, reinforcement, and emotional responses, which influence behavioral performance [15]. Together, these frameworks enable a systematic analysis of how behavioral determinants interact within real-world contexts to shape the engagement of people with diabetes and caregivers with digital health interventions.

Methods

Study Context

The Foot Selfie study was established as an innovative approach to address the burden of DFU in Peru, emphasizing prevention and early identification of foot ulcers. In phase 1 of the project, an initial version of a mobile app (Android package kit [APK]) was designed with a single functionality of capturing and uploading foot images for training an AI algorithm to detect healthy skin and 6 preulcerative lesions. These lesions included fissures, erythemas, calluses, hemorrhagic calluses, discoloration, and ulcers. The first cohort of participants for image collection included 1000 adults with type 1 or type 2 diabetes mellitus in hospital settings. After initial image collection in hospital settings, a subsample of up to 200 individuals was invited, according to specific selection criteria, to use the app at home for follow-up image capture. In total, 12,164 images were collected from phase 1. During phase 2 of the parent study, individuals who participated in at-home follow-up were selected to participate in this study according to the selection criteria detailed below. At this stage, the app was designed to be fully self-directed and did not provide diagnostic feedback. Participants were informed that submitted images were not a substitute for clinical care. However, if the research team identified a concerning lesion or potential complication during study procedures, participants were contacted and advised to seek appropriate clinical evaluation. Weekly follow-up from field personnel (nurses) emerged somewhat organically and typically occurred when participants stopped using the app.

Study Design

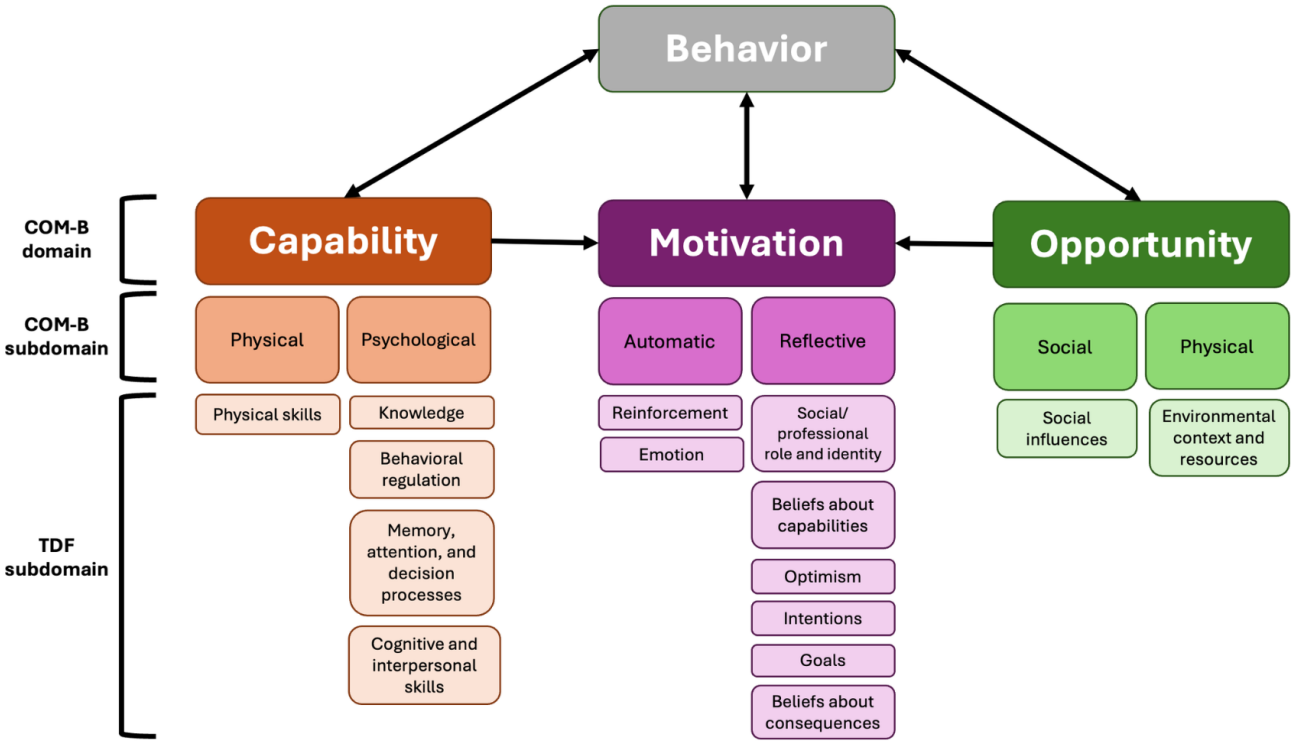
This study used a qualitative descriptive design and was explicitly informed by the COM-B model to examine determinants of app use behavior [14]. To deepen interpretation, domains from the TDF were used to provide additional context within the COM-B domains [16] (Figure 1). The target behavior for all participants was defined as the use of the app to capture

XSL•FO
RenderX

photographs of the foot for early detection of preulcerative lesions.

People with type 2 diabetes and caregivers who had participated in at-home use of the early prototype for a 1-month period and had consented to be recontacted were invited to participate in semistructured interviews. Data were collected between August 2024 and February 2025. The COREQ (Consolidated Criteria for Reporting Qualitative Research) was used to ensure detailed reporting [17].

Figure 1. Overview of the COM-B (capability, opportunity, motivation–behavior) model and Theoretical Domains Framework (TDF). This figure demonstrates the 3 COM-B domains, their subdomains, and the associated TDF domains that inform behavior. Adapted from Atkins et al [16]. This framework was applied independently to each participant—people with diabetes and caregivers—and compared afterward.



Study Setting

Qualitative data collection was conducted in Lima, Peru, a region selected for its epidemiological relevance to the burden of diabetes. Lima, the country's capital, concentrates approximately one-third of the national population and reported the highest number of diabetes cases in 2025 [18].

The people with diabetes in the study were receiving care for this condition at endocrinology and diabetic foot service units at 5 high-complexity (category III-1, according to the Peruvian Ministry of Health classification) hospitals: Hospital María Auxiliadora, Hospital Dos de Mayo, Hospital Cayetano Heredia, Hospital Arzobispo Loayza, and Hospital de Emergencias Villa El Salvador.

The multidisciplinary research team was composed of researchers with backgrounds in public health, social sciences, and medicine, with significant collective experience in diabetes research. Interview supervision and data collection were led by an anthropologist with extensive expertise in qualitative research and behavioral design.

All interviewers had prior training in qualitative interviewing, and they did not have an established personal relationship with participants at the time the interviews were conducted.

The analysis was supported by team members trained in public health, biology, and project management and was led by a researcher with a background in epidemiology and global health. Discussions were held throughout the study to consider how researchers' backgrounds could influence data collection and interpretation.

Study Participants and Sampling

For interview selection, a purposive sampling strategy with predefined user profiles was used to ensure variation in experiences with app use. Four profiles were defined: (1) people with diabetes who were frequent users, (2) people with diabetes who were infrequent users, (3) caregivers who were frequent users, and (4) caregivers who were infrequent users. Frequent users were defined as individuals who used the app 6 or more times during the month-long study period, whereas infrequent users were those who used the app fewer than 6 times that month.

Eligible participants were approached via 2 strategies. First, potential participants who had already finished using the app at home were contacted by telephone and invited to an interview. Second, potential participants who were invited to start using the app at home were also invited to participate in this study. For this option, researchers visited health care facilities and invited individuals in waiting rooms to participate. When approached, potential participants received information on the study objectives, conditions of participation, and voluntary nature of participation.

XSL•FO
RenderX

Study participants included people with diabetes and caregivers who had participated in the at-home use of the app to capture foot images. The demonstration of app interface and standard image capture is provided in Multimedia Appendix 1. The caregivers interviewed were not the caregivers of the people with diabetes interviewed in this qualitative sample, but instead they were the caregivers of other patients enrolled in the same parent study.

Data Collection

In-depth individual interviews were conducted using a semistructured guide based on the COM-B subdomains. The interview guide in both the original Spanish and its English translation is provided in Multimedia Appendix 2. The guide explored initial experiences with the app, foot-monitoring habits, frequency and conditions of use, technological barriers, digital accessibility (device type, connectivity, and storage), social support, perceptions of diabetic foot risk, and suggestions for improving the tool's design.

Questions were adapted according to participants' profiles (people with diabetes or caregivers) while maintaining conceptual consistency across interviews. When feasible, participants were asked to describe or demonstrate how they accessed the app and captured foot photographs to identify practical usability challenges.

Most interviews were conducted in participants' homes, and one interview was conducted remotely via videoconference based on participants' preferences and availability. All interviews were conducted in private settings without the presence of third parties.

Interviews lasted approximately 45 to 60 minutes. All interviews were audio-recorded with prior authorization and transcribed verbatim. Repeat interviews were not conducted, and transcripts were not returned to participants. Data collection continued until meaning saturation was reached, which is defined as the point at which additional interviews no longer generated new insights or perspectives relevant to the analytic categories, indicating sufficient depth and coverage of the study objectives [19].

Data Analysis

All interviews were anonymized using alphanumeric codes and analyzed in 2 sequential stages.

In the first stage, a deductive thematic analysis guided by the COM-B model and the TDF was conducted. The COM-B model provided the overarching behavioral structure, while the TDF offered an additional contextual layer for classifying determinants within each domain. Separate analytical matrices were constructed for people with diabetes and caregivers. Two researchers independently coded excerpts and classified them into COM-B domains and corresponding TDF subdomains, followed by consensus meetings to resolve discrepancies.

In the second stage, an inductive analysis was conducted within each domain to identify emergent cross-cutting themes among and between participant types. These themes were then compared between people with diabetes and caregivers to examine patterns shared between and distinct to the 2 groups. To do so, a second synthesis matrix was used to reorganize coded excerpts into themes while maintaining explicit links to the COM-B and TDF domains. Regular analytic meetings with the senior qualitative researcher and the broader research team were held to refine interpretations and strengthen analytic rigor. A sample matrix with domain and coding labels can be found in Multimedia Appendix 3.

Data management and analysis were manually conducted using structured Microsoft Excel matrices due to the team's familiarity with this software. No specialized qualitative software was used. All quotes presented in this paper were originally collected in Spanish and have been translated into English for reporting purposes.

Ethical Considerations

This study has received ethical approval from the following committees: the Institutional Ethics Committee of the Universidad Peruana Cayetano Heredia (reference number 210819), Lima, Peru, and the London School of Hygiene & Tropical Medicine Research Ethics Committee (reference number 30662), United Kingdom, as required for studies involving London School of Hygiene & Tropical Medicine–affiliated investigators and those funded by the UK Medical Research Council.

All participants provided written informed consent before enrollment. For qualitative interviews, additional consent was obtained for audio-recording. Participation was voluntary, and individuals could withdraw at any time without affecting their access to care.

To protect participants' privacy and confidentiality, identifying information was removed from the study dataset, and each participant was assigned a unique study code. Study data were maintained in a secure spreadsheet accessible only to the research team. Participants who attended interviews in person received 20 Peruvian soles (US $5.89) to help cover transportation costs.

Results

General Findings

A total of 24 interviews were conducted: 14 with people with type 2 diabetes and 10 with caregivers. All caregivers interviewed were identified as female, were predominantly younger than 30 years, and were almost all informal family caregivers, except for one nonfamily caregiver (Table 1). In contrast, most people with diabetes were older than 50 years and had attained only up to a secondary education.

XSL•FO
RenderX

Table . Sociodemographic characteristics of interview participants, by role (caregiver vs patient).

Participant characteristic	Caregiver (N=10)	People with diabetes (N=14)
Gender		
Male	0	6
Female	10	8
Age group (y)		
≤30	6	0
31 - 39	1	0
40 - 49	2	2
50 - 59	1	6
60 - 69	0	5
70 - 79	0	1
Highest education level attained[a]		
Primary	1	1
Secondary	5	11
Higher education[b]	4	2
History of ulceration	—[c]	6
History of amputation	—	4
History of deformity	—	7

[a]Primary and secondary education categories include participants who both completed and did not complete that level.

[b]Higher education in Peru includes postsecondary education at universities or nonuniversity higher education institutions (eg, technical or professional institutes), regardless of degree completion status.

[c]Not applicable.

Across transcripts, 239 excerpts were coded for people with diabetes and 332 for caregivers, using COM-B and TDF domains. These coded excerpts were inductively collapsed into general facilitators and barriers mapped to capability, opportunity, and motivation (Tables 2 and 3). Overall, although some themes predominantly aligned with a single COM-B domain, our analysis revealed substantial overlap across domains, with capability, opportunity, and motivation interacting dynamically to shape the app use. The following sections detail these cross-domain interactions, which are also visually summarized in Figure 2.

XSL•FO
RenderX

Table . Summary of facilitators and barriers across COM-B (capability, opportunity, motivation–behavior) domains identified among people with diabetes.

COM-B domain	General themes
Physical capability	
Facilitators	• Independent ability to position and photograph the foot
Barriers	• Limited mobility or difficulty reaching/visualizing the foot • Physical fatigue or late-night exhaustion • Need for assistance to complete positioning
Psychological capability	
Facilitators	• Development of routines and reminder strategies • Use of adaptive strategies (pillows, selfie sticks, and lamp positioning)
Barriers	• Limited digital literacy • Forgetfulness or inconsistent routine • Initial confusion regarding installation and workflow
Physical opportunity	
Facilitators	• Access to smartphones, internet connectivity, and adequate lighting • Organized home spaces and established routines • Ability to complete monitoring without traveling to hospital
Barriers	• Unstable internet connectivity and device incompatibility • App crashes, slow loading, and unclear submission confirmation • Inadequate lighting or physical setup
Social opportunity	
Facilitators	• Assistance from family members or household contacts • External reminders from relatives • Follow-up communication from health professionals
Barriers	• Dependence on others for technical or physical assistance • Inconsistent follow-up from field personnel during and after study completion
Automatic motivation	
Facilitators	• Emotional reassurance from perceived monitoring • Satisfaction from observing wound improvement • Routine formation through repeated use
Barriers	• No barrier identified
Reflective motivation	
Facilitators	• Belief in early detection and prevention benefits • Perceived convenience compared to hospital visits • Commitment to research participation and helping others
Barrier	• Doubt about whether photos were reviewed or useful • Participation primarily driven by obligation rather than perceived personal value

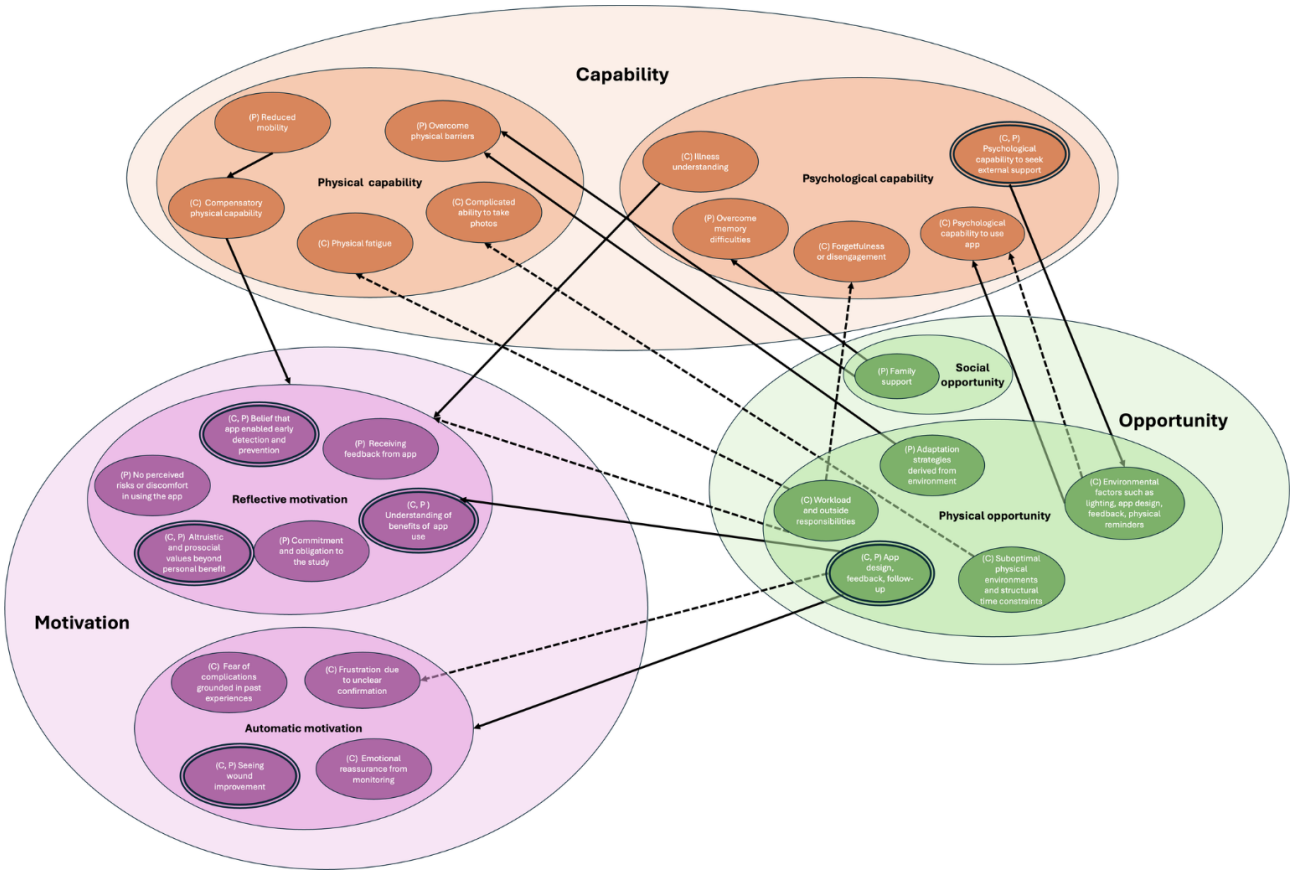
XSL•FO
RenderX

Table . Summary of facilitators and barriers across COM-B (capability, opportunity, motivation–behavior) domains identified among caregivers.

COM-B domain	General themes
Physical capability	
Facilitators	• Physical ability to position, lift, and assist people with diabetes enabled photo-taking despite mobility limitations of the person with diabetes • Physical capability is strengthened through repetition and routine performance
Barriers	• Physical fatigue from caregiving load, work, or illness reduced consistency of app use • Patient mobility limitations required assistance for task completion
Psychological capability	
Facilitators	• Knowledge of diabetic foot risk increased preparedness and engagement • Caregivers demonstrated higher digital literacy than patients • Caregivers exercised behavioral regulation and sought out answers themselves
Barriers	• Memory, lapses, attention, and cognitive overload • Initial confusion regarding installation of the app and workflow
Physical opportunity	
Facilitators	• Access to smartphones, internet connectivity, and adequate lighting • Organized home spaces and established routines • Remote monitoring reduced need for hospital travel
Barriers	• Unstable internet connectivity and device incompatibility • App crashes, slow loading, and unclear submission of pictures (lack of confirmation) • Time constraints due to work and caregiving responsibilities • Inadequate lighting or physical setup
Social opportunity	
Facilitators	• Family collaboration and task-sharing among relatives • Follow-up and accountability from field personnel • Patient reminders and mutual reinforcement within the household
Barriers	• Limited or inconsistent family support • Discontinuation of follow-up from field personnel
Automatic motivation	
Facilitators	• Emotional reassurance from perceived monitoring • Satisfaction from observing wound improvement • Fear of complications (eg, amputation) prompting vigilance • Routine formation through repeated use
Barriers	• Frustration and stress from technical failures
Reflective motivation	
Facilitators	• Belief in early detection and prevention benefits • Perceived convenience compared to hospital visits • Commitment to research participation and helping others
Barrier	• Doubt about whether photos were reviewed or useful

XSL•FO
RenderX

Figure 2. Cross-domain interactions that influence the sustained use of the app among people with diabetes and caregivers. This conceptual model illustrates how capability (purple), opportunity (green), and motivation (orange) domains interact to shape engagement with the intervention. Nodes labeled (P) reflect patient-specific influences, (C) reflect caregiver-specific influences, and double-lined (C and P) reflect shared determinants. Solid arrows indicate positive influences, whereas dashed arrows represent negative relationships.



Capability

The biggest difference observed between caregivers and people with diabetes was their capabilities. When people with diabetes had reduced mobility, caregivers' physical capabilities acted as a compensatory facilitator, enabling task completion through lifting, positioning, or adapting the environment (TDF domain: physical skills). A person's lack of physical capability served as a motivating factor for caregivers to use the app.

> *Only I could handle the app, position my mom […] I adjust her, and then I just take the photo.* [Caregiver, female, 23 years old]

Still, even when caregivers had additional skills required to use the app, physical fatigue from the caregiving load, work, or illness served as a barrier, reducing consistency and delaying task completion. In this way, people with diabetes and caregivers alike experienced physical capability barriers, but for different reasons.

> *I would get home at 10 at night… I was already sleepy, tired, and then it would be tomorrow already, and I wouldn't take it anymore.* [Caregiver, female, 50 years old]

> *Well, what happens is that since it's the feet, it's like I'm a bit overweight, but that depends on me, not on the application.* [Patient, female, 50 years old]

Interestingly, some people with diabetes used adaptation strategies to overcome these limitations, such as using a pillow or a selfie stick or taking a photo during the day to take advantage of the lighting (TDF domain: behavioral regulation).

> *On the second day it was difficult for me because I had to take the photos by myself and for [the app] to take the photo […] I had to put pillows, then I used the little stick to take photos.* [Patient, female, 41 years old]

Many people with diabetes were psychologically capable of using the app and acquired this ability through practice and other strategies, such as setting reminders and developing routines for taking pictures (TDF domain: behavioral regulation). Caregivers also experienced initial confusion with the app, which was resolved through practice, demonstration, repetition, or troubleshooting (TDF domain: memory, attention, and decision processes).

> *The first few days, yes, two or three days I would get confused [with the app] […] But afterward it became easier. Afterward you get the…the trick. The trick, as they say, right? And then normally, you send [the photo].* [Patient, male, 51 years old]

> *At first it was a little difficult, but after that…I got the hang of it.* [Caregiver, female, 25 years old]

Additionally, caregivers seemed to better understand the seriousness of diabetic foot complications (eg, infection and

XSL•FO
RenderX

amputation) and the purpose of the app as preventive monitoring, which made them more motivated and cognitively prepared to engage (TDF domain: knowledge). This understanding reframed the task from "taking photos" to preventing future harm, which supported sustained attention and follow-through.

> *I am aware that a person with diabetes… can become infected and progress to the point that they can amputate it.* [Caregiver, female, 25 years old]

Opportunity

Both caregivers and people with diabetes who had the opportunity to perform foot monitoring depended on having access to the internet, smartphones, lighting, usable home spaces, and physical reminders. For caregivers, these environmental conditions enabled the behavior and interacted with capability: poor lighting or device limitations increased physical strain and cognitive effort, while adequate setups reduced both (TDF domain: environmental context and resources). People with diabetes who had these adequate environmental conditions were able to perform the behavior consistently, while those who did not were unable to complete the task without adaptation (TDF domain: environmental context and resources).

> *We always looked for the room with good lighting.* [Caregiver, female, 21 years old]

> *Well, in my case, I have internet at home […] I never had any problems.* [Patient, female, 50 years old]

> *The first day, two days, yes, it took us a while because sometimes with the light my foot would come out crooked, then little by little I learned the position […] and at night I would turn on my lamp so that it would focus on me […] sometimes the light would reflect and I would say, no, the light is positioned wrong, I remember.* [Patient, male, 65 years old]

In terms of app performance, people with diabetes reported that although the app was often easy to use due to its intuitive interface, structural barriers—including device incompatibility, inability to verify whether photos had been successfully sent, and the need to repeatedly reenter the app workflow—created disruptions in their opportunity to complete the task consistently. For caregivers, clear visual design and feedback enabled opportunity and motivation by making photo-taking more feasible. Conversely, slow loading, crashes, unclear confirmation, and device incompatibility disrupted opportunity, decreased motivation, and simultaneously increased cognitive uncertainty and frustration.

> *I had to buy another cellphone, because the one I had wouldn't go in anymore […] No, it wouldn't allow it.* [Patient, male, 51 years old]

> *It would freeze […] I had to take the photo again.* [Caregiver, female, 27 years old]

Caregivers frequently faced structural time constraints due to work, caregiving for others, unstable schedules, and an overall lack of routine. These factors limited physical opportunities to perform the task, even when motivation and capability were present. These constraints also produced fatigue, forgetfulness, or disengagement.

> *[…] I just don't have enough time to stay and check. Even though it doesn't take long, honestly, I don't have a set schedule — sometimes I work in the morning, sometimes in the afternoon, sometimes at night.* [Caregiver, female, 23 years old]

Social support was a dominant structural enabler of participation for people with diabetes and caregivers, but the underlying mechanisms differed. Many people with diabetes depended on children, siblings, or even building staff to take photos, enter login credentials, or troubleshoot (TDF domain: social influences). For several people with diabetes, independent use was not feasible without assistance due to physical positioning challenges, digital literacy limitations, or difficulty visualizing their own feet.

> *Sometimes my husband [took the photographs]. I told him, move here or he would take about 10 photos and then I would choose.* [Patient, female, 41 years old]

> *It became difficult for me so I looked for the security guard to help me take the photo, nothing more. I did the programming and he would just take the photo.* [Patient, male, 65 years old]

In contrast, caregivers' reliance on social support did not stem from a lack of capability. Caregivers were generally able to perform the task independently; however, sustaining participation required coordination within the household. Family collaboration functioned as task-sharing rather than substitution, with relatives rotating through roles such as taking photos, providing reminders, or assisting when the primary caregiver was unavailable. Follow-up by health staff similarly reinforced continuity by serving as an external structure of accountability.

> *And if I couldn't, like I'm telling you, if I couldn't be there one or two days, my son was here with him all day. And he stayed like that. He would watch how I [took the photos], and he had already memorized it too.* [Caregiver, female, 42 years old]

> *The same doctor, the nurse, kept communicating that there was progress, and afterward she told me: 'Okay, that's fine, up to here,' and we didn't send any more photos.* [Caregiver, female, 23 years old]

Motivation

By far, the most common motivator mentioned among caregivers and people with diabetes was a belief about the consequences of app use, whether that was a belief in the potential for prevention of DFU or a belief in the convenience of taking care of your feet at home rather than in a hospital (TDF domain: beliefs about consequences). Caregivers were motivated when they believed that the app enabled early detection and prevention, reframing the behavior as proactive rather than reactive. Motivation to use the app for people with diabetes arose from personal benefits and the detection of health problems that were not initially perceived.

XSL•FO
RenderX

> *If a lesion were to suddenly appear and you let me know, we can act sooner. That way we prevent it.* [Caregiver, female, 32 years old]

> *…Because in the long run the benefit will be for us, the diagnoses will be for us, not for the person who received the photos, but for ourselves, so that's why I did it.* [Patient, female, 53 years old]

> *No, no, because I dedicate 10 minutes and that's it, I can dedicate more hours to TikTok, to Facebook, and this is something useful, and you learn. If they send me, put this on, you can eat this, I don't know, more information about my disease, great, I read it, because I read.* [Patient, female, 50 years old]

Among people with diabetes, some participated in photo-taking out of commitment to the research study, which ties into the "reflective motivation" TDF domain of social/professional role or identity. In these cases, the use of the app exclusively arises out of an obligation and a belief in the value of their commitment (TDF domain: intentions). However, some still recognized the app's usefulness, which helped reinforce more consistent use through curiosity, self-care, and control over their health.

> *More than anything because I had made a commitment to the young lady (research team member), I had committed to helping. So with the study, more than anything it was for that reason, because I am a person that if I commit to something it is because I am going to try to do it, to fulfill it.* [Patient, female, 53 years old]

An unexpected motivator that emerged for both caregivers and people with diabetes was a desire to help others or altruism. Some caregivers framed their use of the app not as a means to benefit their own family members, but as a way to contribute to research, support other people with diabetes, and improve care for the broader community (TDF domain: social/professional role and identity). This altruistic framing sustained engagement even when direct personal benefits were limited or uncertain. People with diabetes expressed a desire to help other people with diabetes and share experiences even when they felt exhausted or faced difficulties in daily life.

> *Why? Because you're going to help someone else. Well, that's what moved me, that it is to help other people to see. I think that's why it motivates you to do it, that it's help for others, right? To tell you about your experience […] I liked it because it is to help other people.* [Patient, female, 41 years old]

> *I'm not doing it only for my mom… but because I see that it's an app that is going to help other people.* [Caregiver, female, 25 years old]

For people with diabetes, automatic motivation interpreted as reinforcement functioned as a key facilitator. Receiving feedback on submitted photos or observing visible wound improvement generated satisfaction, reassurance, and increased confidence in their ability to manage their condition (TDF domain: reinforcement). These positive emotional responses and sense of relief reinforced routine use and strengthened habit formation over time. Similarly, caregivers described reinforcement through visible clinical progress or confirmation from health care staff. Observing improvement in the foot or receiving acknowledgment that images had been reviewed produced pride and relief, reinforcing continued monitoring behavior.

> *Now [the application] was useful because they have seen that I am well and they have monitored me and I like being well.* [Patient, female, 50 years old]

> *Sometimes it felt good […] even with the wound. Whenever it improves, you feel happy: 'Look at how the photo looks!' Comparing the first photo with the last is always nice because it's the change—you can see the improvement.* [Caregiver, female, 23 years old]

Emotion-driven vigilance was also prominent, particularly among caregivers (TDF domain: emotion). Fear of complications, especially amputation, served as a powerful emotional driver that heightened attention and sustained adherence to monitoring. This fear was often grounded in prior experiences with severe diabetic complications within the family or community.

> *That's what scares me—that it could get to that point… that they might have to amputate the foot.* [Caregiver, female, 21 years old]

In one particular instance involving a caregiver, frustration and uncertainty about the app undermined automatic motivation. Technical failures, unclear confirmation of image receipt, and ambiguity about whether photos were being reviewed generated stress, disengagement, and overall demotivation. Ultimately, these negative emotional responses led to the permanent abandonment of the prototype app by this caregiver.

> *That stressed me out… and that made me quit [the app].* [Caregiver, female, 21 years old]

Discussion

Principal Findings

In this study, we explored behavioral influences on the use of an mHealth app for diabetic foot care in Peru. After our analysis with the COM-B model and TDF, we found that the sustained use of the app for the detection of DFUs and preulcerative lesions was not always determined by a single behavioral domain but rather by interactions among capability, opportunity, and motivation among people with diabetes and caregivers. These interactions were strongly influenced by conditions across the social, technological, and health care systems in Peru, highlighting the importance of incorporating stakeholder experiences and behaviors from the earliest stages of implementation, including the algorithmic development of an mHealth intervention.

When discussing individual domains, physical capability barriers were particularly evident among people with diabetes who had mobility limitations, while caregivers frequently compensated by positioning the foot to take pictures or by making environmental adjustments, such as modifying light and room settings. Opportunity was strongly shaped by structural and

XSL•FO
RenderX

technological conditions, including device compatibility, internet access, and home environments, which in turn conditioned how these behaviors could be carried out. Meanwhile, motivation was largely driven by beliefs about the preventive value of the intervention, emotional responses to first-hand experiences with DFU, and reassurance that health professionals were monitoring submitted images. Other motivational influences included altruistic reasons for sustaining engagement with the app, whereas uncertainty about whether images were being received could undermine motivation. Together, these results highlight that engagement with our digital self-care intervention stems from various behavioral interactions rather than only isolated determinants.

A key contribution of this study is the identification of people with diabetes and their caregivers as a behavioral unit or dyad in the use of digital health technologies. While many mHealth interventions conceptualize self-monitoring as an empowering, individual activity [20-22], our findings show that the routine use often depended on collaboration within the household. In several multigenerational households, family members, mostly children or grandchildren, rotated responsibilities when the primary caregiver was unavailable. Previous research on remote monitoring technologies for diabetes similarly suggests that caregiver involvement can facilitate adherence and improve clinical outcomes, particularly among older adults and individuals with lower digital literacy [23].

In our sample, this dynamic was also reflected in the demographic profile of participants. People with diabetes were substantially older than caregivers (median age of 54 y vs 28.5 y) and had generally attained lower levels of formal education. These differences may have contributed to variations in both familiarity with smartphone interfaces and illness-related knowledge, as reported in other studies [24,25]. Although smartphone ownership is widespread in Peru [26-28], several people with diabetes described challenges related to positioning the foot for photographs, reading small text, or navigating app steps. Despite these physical struggles, people with diabetes whom we interviewed still demonstrated adequate psychological capability to understand the purpose of the intervention and even used adaptation strategies to continue using the app. In these cases, caregivers did not replace the decision-making capacity of people with diabetes but instead bridged physical and informational gaps associated with performing the task. By supporting the practical and cognitive demands associated with technology use, caregivers enabled participation among people with diabetes who might otherwise be excluded from technological interventions [29,30].

Building on the concept of a dyadic relationship between people with diabetes and caregivers, the involvement of caregivers also introduces additional complexity. While digital technologies are often promoted as tools that reduce caregiver burden, evidence suggests that they may simultaneously introduce new responsibilities and stressors [31]. Studies examining the impact of assistive technologies on informal caregivers have found that while these tools can improve communication, monitoring, and perceived safety, they may also generate frustration when technical difficulties arise or when the technology requires additional time and attention to operate [31].

Importantly, the caregiving role itself is highly gendered, which may intersect with the age-related differences in digital literacy that we observed. In our sample, all caregivers interviewed were women, a pattern that reflects broader trends documented in the literature, where informal caregiving responsibilities disproportionately fall on female family members, particularly in Latin American contexts [31-33]. Although we used a purposive sampling strategy to capture a diverse range of app-use profiles (eg, frequent and infrequent users) as well as variation in gender, our caregiver sample did not include male caregivers. As a result, our findings may not fully capture the behaviors and interactions between people with diabetes and male caregivers. This gendered pattern is important because research examining the use of technology among informal female caregivers has shown that, while digital tools may support caregiving tasks, they can also add to existing responsibilities and contribute to declines in caregivers' physical and mental well-being over time [32]. This is especially relevant in settings where caregiving responsibilities are often balanced alongside employment, household duties, or educational commitments.

Similar patterns emerged in our investigation. Caregivers frequently described time constraints, competing responsibilities, and fatigue that limited their ability to perform monitoring tasks consistently. However, most participants did not describe the app itself as a direct burden. Instead, the challenges they described were primarily related to the broader demands of caregiving and daily life, such as work or school obligations. Only one caregiver explicitly reported disengaging from the app due to frustration with technical difficulties, which were also reported by Holthe et al [34] as a source of stress for caregivers working with assistive technology. These instances suggest that even relatively simple digital monitoring interventions can introduce new routines and obligations that must be considered during intervention design.

Another important finding relates to participants' strong preference for follow-up and reassurance from health care professionals. Both caregivers and people with diabetes described feeling more motivated to use the app when they believed that a clinician was reviewing their images or monitoring their progress. Previous studies of digital interventions have similarly shown that remote monitoring had a reassuring effect on people with diabetes and allowed them to share the responsibility of their health care decisions with others [21,29]. Interestingly, although these studies involved clinicians directly reviewing patients' results, our mHealth app did not include clinicians' review of submitted images for diagnosis. Instead, follow-up only occurred if any lesions were found or if participants had discontinued the app use. During interviews, participants often interpreted this follow-up communication and the submission of photos as physicians reviewing their images, even though this was not outlined in the participant information and consent forms. This finding in particular has been considered for future iterations of the app within the Foot Selfie study. Future versions of the app are set to provide clearer feedback confirming whether images were successfully submitted and clarifying that diagnostic feedback is AI-generated rather than determined by a physician.

XSL•FO
RenderX

This demonstrated preference for follow-up may be related to the broader preference for clinician-guided care and paternalism described in Peru and other contexts outside of Western medicine [35-37]. Although paternalism is often defined as physicians making decisions for people with diabetes without regard for patient autonomy, in this cultural context, the desire for authority may instead reflect how people with diabetes interpret care and attention from health care professionals, as seen in other qualitative studies [38]. In other words, follow-up appeared to signal concern and commitment from clinicians, which in turn increased the perceived value of the app among participants. This interpretation is consistent with a preliminary study from our research group, in which both the intervention and control arms of a DFU prevention trial remained more engaged than expected, suggesting that regular contact and follow-up may itself influence self-care behaviors [39]. Brandt et al [40] reinforced this finding and found that participants were not as motivated or adherent with a mobile self-help intervention in Peru when they were left without the support of a nurse. Taken together, these findings suggest that sustained engagement in our study may have been driven in part by participants' perception that a health professional was monitoring their submissions, rather than by the digital intervention alone.

Participants' desire for continuity of care in this study raises questions about the sustainability of implementing such an intervention if its primary goal is to address gaps in access to care and promote preventive self-management. The feasibility of having health care professionals regularly reviewing submissions is limited, particularly in a health care system such as Peru's that is already resource-constrained and where people with diabetes often experience long wait times to access specialist care [37]. In this context, AI-assisted identification of preulcerative lesions may provide a form of timely preliminary review, helping prioritize clinicians' attention for participants who need further evaluation. As a whole, future implementation of the app may be more effective if it incorporates other actors, including caregivers, nonphysician health personnel, and AI-assisted triage, to support continuity without relying solely on regular clinician review.

Limitations

Although the use of the COM-B model combined with the TDF provided a structured approach to understanding behavioral determinants of app use, the interpretation of qualitative interview data remains inherently subjective. The analytic framework may also have shaped how themes were identified and interpreted. To enhance analytic rigor, regular meetings were held within the research team to review coding decisions and discuss interpretations. Because the analysis relied on predefined theoretical domains, it is possible that certain factors influencing behavior were interpreted primarily through the lens of COM-B and TDF, despite regular team-based reviews of coding decisions.

During interviews, participants may have exhibited social desirability bias by portraying the app more favorably, even though the interviewers were not previously known to them. Despite sampling participants who were not frequent users of the app, it is still possible that individuals with negative experiences might not have fully disclosed them due to this bias. To mitigate this, we conducted semistructured interviews, and interviewers were trained to build trust, maintain confidentiality, and remain neutral. Additionally, the caregivers we interviewed were not the caregivers of the people with diabetes included in our sample, limiting our ability to directly assess concordance or discordance in beliefs within patient-caregiver dyads.

Implications

Importantly, the app used during the study was an early prototype developed for algorithmic training. While participants reported several technical challenges that at times affected engagement, these experiences provided critical insight into real-world usability and contextual barriers. Incorporating user feedback from this early stage allowed for refinement of the app through subsequent cocreation workshops with stakeholders and engineers, resulting in an improved version better aligned with the needs of people with diabetes in Peru.

Our findings also imply that app design should prioritize usability across diverse technological contexts, including compatibility with older devices and clear feedback confirming successful data submission. Alongside this, interventions should explicitly consider caregivers as key actors in digital self-monitoring programs, particularly for older adults or individuals with limited mobility. Training materials, reminders, and communication features may therefore need to address both people with diabetes and caregivers. Prior research examining dyadic relationships in chronic heart failure care similarly highlights how interactions between people with heart failure and caregivers can influence disease management behaviors, suggesting that approaches engaging both members of the caregiving dyad may strengthen the effectiveness of self-management interventions [25]. Furthermore, since the use of mHealth tools can, at times, be a burden for caregivers, informal task-sharing among other household members should be considered for implementation. Finally, the strong desire for professional follow-up suggests that future versions of the intervention may benefit from hybrid models combining self-monitoring with targeted human support. The AI-assisted identification of concerning findings could help guide review by nurses or other nonphysician health professionals, with escalation when needed, preserving the preventive focus of the intervention while supporting realistic continuity of care.

Conclusions

As mHealth interventions for diabetic foot self-care continue to expand, our study demonstrates how examining behavioral interactions between people with diabetes and caregivers can inform the development of person-centered mobile interventions. We found that sustained engagement with the app for diabetic foot self-care emerges from dynamic interactions between capability, opportunity, and motivation rather than any single behavioral determinant for participants. In this study, the app frequently depended on collaboration between people with diabetes and various household caregivers or family members, who together functioned as a behavioral unit that jointly enabled participation in digital monitoring. This central finding

XSL•FO
RenderX

underscores the importance of designing digital health interventions that account for household caregiving dynamics, technological constraints, and user perceptions of clinical support. As these insights are incorporated into future versions of the app, considering behavioral and contextual factors may improve the implementation and long-term adoption of mHealth tools aimed at preventing diabetic foot complications in Peru and other low- and middle-income settings, as well as within health care organizations.

Acknowledgments

The authors thank all study participants and their caregivers for their time, commitment, and collaboration throughout the study. They also acknowledge the field team for their essential role in participant training, image collection, and follow-up activities. The authors also thank Pedro Távara and Ana Bautista for their valuable contribution as interviewers in Piura and Lima, respectively. They are especially grateful to the health care professionals from the endocrinology services and the administrative staff of the 5 participating public hospitals in Lima, including Hospital Nacional Cayetano Heredia, Hospital María Auxiliadora, Hospital Nacional Dos de Mayo, Hospital Nacional Arzobispo Loayza, and Hospital de Emergencias Villa El Salvador, for their support, collaboration, and facilitation of study activities. Finally, the authors would like to thank Lucas Taipe for his assistance in developing an app prototype with German Comina for this study. The authors acknowledge that ChatGPT (Open AI) was used to help translate anonymized interview quotes and the interview guide from Spanish to English. Native, Spanish-speaking authors reviewed and edited the final version of each quote included in this manuscript for correctness and accuracy.

Funding

This study was supported by the Medical Research Council of the United Kingdom (grant number MR/X004163/1). The funder had no role in the conceptualization and design of this study and did not have any role in the collection of data or the decision to publish. SMR was supported by the US Fulbright Open Research Award Program in Peru.

Data Availability

The research team is committed to broad and open dissemination of findings. All peer-reviewed journal articles resulting from this study will be published in an open-access format and made freely available at the time of publication. Deidentified anonymized datasets and study materials will be available upon reasonable request to the corresponding author, in line with institutional policies and data-sharing agreements.

Authors' Contributions

ML-P and JJ Miranda conceived the Foot Selfie study. JJ Miranda, ML-P, JHZ-T, PP, J Siddique, J Seiglie, HL, SJ, LD, DB, DGA, and GC secured study funding. LC-C led day-to-day study planning and implementation activities. JHZ-T and JJ Mantilla conceived this substudy and led the overall study design. SMR, GA-H, and SGC worked on transcript analysis and thematic coding of the data. SMR, JHZ-T, GA-H, JJ Mantilla, and LC-C drafted the manuscript. SMR, JHZ-T, GA-H, JJ Miranda, LC-C, SGC, JJ Mantilla, PP, J Siddique, J Seiglie, HL, SJ, LD, DB, DGA, GC, and ML-P contributed to the refinement of the study, reviewed, and approved the final manuscript.

Conflicts of Interest

The authors declare no financial conflicts of interest. JJ Miranda, ML-P, JHZ-T, PP, J Siddique, J Seiglie, HL, SJ, LD, DB, DGA, and GC are co-investigators on the parent grant that supported the broader project from which this study emerged. GC, JJ Mantilla, SGC, and LC-C also held roles related to the development of the app and/or the qualitative analysis reported in this manuscript. These roles are disclosed for transparency.

Multimedia Appendix 1
Foot Selfie mobile app used for foot image capture.
[DOCX File, 186 KB - mhealth_v14i1e101195_app1.docx]

Multimedia Appendix 2
Semistructured interview guide for people with diabetes and caregivers regarding the use and accessibility of the Foot Selfie mobile app in both original Spanish and its English translation.
[DOCX File, 15 KB - mhealth_v14i1e101195_app2.docx]

Multimedia Appendix 3
A sample matrix with domain and coding labels for COM-B (capability, opportunity, motivation–behavior) and Theoretical Domains Framework (TDF) transcript analysis.

XSL•FO
RenderX

[XLSX File, 150 KB - mhealth_v14i1e101195_app3.xlsx]

References

1. Armstrong DG, Boulton AJM, Bus SA. Diabetic foot ulcers and their recurrence. N Engl J Med 2017 Jun 15;376(24):2367-2375. [doi: 10.1056/NEJMra1615439] [Medline: 28614678]
2. Bus SA, Sacco ICN, Monteiro-Soares M, et al. Guidelines on the prevention of foot ulcers in persons with diabetes (IWGDF 2023 update). Diabetes Metab Res Rev 2024 Mar;40(3):e3651. [doi: 10.1002/dmrr.3651] [Medline: 37302121]
3. Istepanian RSH, Al-Anzi TM. m-Health interventions for diabetes remote monitoring and self management: clinical and compliance issues. mHealth 2018;4:4. [doi: 10.21037/mhealth.2018.01.02] [Medline: 29552566]
4. Hou C, Xu Q, Diao S, Hewitt J, Li J, Carter B. Mobile phone applications and self-management of diabetes: a systematic review with meta-analysis, meta-regression of 21 randomized trials and GRADE. Diabetes Obes Metab 2018 Aug;20(8):2009-2013. [doi: 10.1111/dom.13307] [Medline: 29582538]
5. Wu X, Guo X, Zhang Z. The efficacy of mobile phone apps for lifestyle modification in diabetes: systematic review and meta-analysis. JMIR mHealth uHealth 2019 Jan 15;7(1):e12297. [doi: 10.2196/12297] [Medline: 30664494]
6. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11. [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
7. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
8. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567. [doi: 10.2196/14567] [Medline: 31573916]
9. Rosland AM, Heisler M, Piette JD. The impact of family behaviors and communication patterns on chronic illness outcomes: a systematic review. J Behav Med 2012 Apr;35(2):221-239. [doi: 10.1007/s10865-011-9354-4] [Medline: 21691845]
10. Niño-de-Guzman Quispe E, Bracchiglione J, Ballester M, et al. Patients' and informal caregivers' perspectives on self-management interventions for type 2 diabetes mellitus outcomes: a mixed-methods overview of 14 years of reviews. Arch Public Health 2023 Aug 4;81(1):140. [doi: 10.1186/s13690-023-01153-9] [Medline: 37537669]
11. Wolff JL, Roter DL. Family presence in routine medical visits: a meta-analytical review. Soc Sci Med 2011 Mar;72(6):823-831. [doi: 10.1016/j.socscimed.2011.01.015] [Medline: 21353358]
12. Najafi B, Reeves ND, Armstrong DG. Leveraging smart technologies to improve the management of diabetic foot ulcers and extend ulcer-free days in remission. Diabetes Metab Res Rev 2020 Mar;36 Suppl 1(S1):e3239. [doi: 10.1002/dmrr.3239] [Medline: 31909547]
13. Méndez NG, Aguilera MFG, Muñoz E, et al. New technologies applied in self-care to patients with diabetic foot ulcers: a scoping review. Diabetol Metab Syndr 2025 Jul 11;17(1):262. [doi: 10.1186/s13098-025-01822-5] [Medline: 40646659]
14. Michie S, van Stralen MM, West R. The Behaviour Change Wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42. [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
15. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012 Apr 24;7:37. [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
16. Atkins L, Francis J, Islam R, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implement Sci 2017 Jun 21;12(1):77. [doi: 10.1186/s13012-017-0605-9] [Medline: 28637486]
17. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
18. Sala situacional de diabetes [Article in Spanish]. Centro Nacional de Epidemiología, Prevención y Control de Enfermedades (CDC Perú). URL: https://app7.dge.gob.pe/maps/sala_diabetes/ [accessed 2026-03-23]
19. Saunders B, Sim J, Kingstone T, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907. [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
20. Sousa P, Martinho R, Parreira P, Luo G. Editorial: mHealth tools for patient empowerment and chronic disease management. Front Psychol 2023;14:1206567. [doi: 10.3389/fpsyg.2023.1206567] [Medline: 37388655]
21. Walker RC, Tong A, Howard K, Darby N, Palmer SC. Patients' and caregivers' expectations and experiences of remote monitoring for peritoneal dialysis: a qualitative interview study. Perit Dial Int 2020 Nov;40(6):540-547. [doi: 10.1177/0896860820927528] [Medline: 33084514]
22. Paton J, Abey S, Hendy P, Williams J, Collings R, Callaghan L. Behaviour change approaches for individuals with diabetes to improve foot self-management: a scoping review. J Foot Ankle Res 2021 Jan 6;14(1):1. [doi: 10.1186/s13047-020-00440-w] [Medline: 33407755]
23. Seng JJB, Gwee MFR, Yong MHA, Kwan YH, Thumboo J, Low LL. Role of caregivers in remote management of patients with type 2 diabetes mellitus: systematic review of literature. J Med Internet Res 2023 Sep 11;25(1):e46988. [doi: 10.2196/46988] [Medline: 37695663]
24. Jung SO, Son YH, Choi E. E-health literacy in older adults: an evolutionary concept analysis. BMC Med Inform Decis Mak 2022 Jan 31;22(1):28. [doi: 10.1186/s12911-022-01761-5] [Medline: 35101005]

XSL•FO
RenderX

25. Yoong SQ, Tan R, Jiang Y. Dyadic relationships between informal caregivers and older adults with chronic heart failure: a systematic review and meta-synthesis. Eur J Cardiovasc Nurs 2024 Dec 16;23(8):833-854. [doi: 10.1093/eurjcn/zvae093] [Medline: 38875163]
26. Digital Government in Peru: Working Closely with Citizens. : Organization for Economic Co-operation and Development (OECD); 2019 URL: https://www.oecd.org/content/dam/oecd/en/publications/reports/2019/06/digital-government-in-peru_2cf04c8f/0c1eb85b-en.pdf [accessed 2026-09-19]
27. Kemp S. Digital 2026: peru. DataReportal. 2025. URL: https://datareportal.com/reports/digital-2026-peru [accessed 2026-03-24]
28. Bécares B. En 2020, el 90% de teléfonos en Perú serán smartphones [Article in Spanish]. Silicon. 2015. URL: https://www.siliconweek.com/mobility/smartphones/en-2020-3-de-cada-telefonos-en-peru-seran-smartphones-64380 [accessed 2026-03-24]
29. Van Grootven B, Irusta LA, Christiaens W, et al. Experiences with the implementation of remote monitoring in patients with COVID-19: a qualitative study with patients and healthcare professionals. J Nurs Scholarsh 2023 Jan;55(1):67-78. [doi: 10.1111/jnu.12814] [Medline: 36165577]
30. Houlding E, Mate KKV, Engler K, et al. Barriers to use of remote monitoring technologies used to support patients with COVID-19: rapid review. JMIR mHealth uHealth 2021 Apr 20;9(4):e24743. [doi: 10.2196/24743] [Medline: 33769943]
31. Aparicio A, Inostroza-Correa MA, Miranda P, et al. The impact of technology use for care by informal female caregivers on their well-being: a scoping review. Syst Rev 2025 Apr 17;14(1):89. [doi: 10.1186/s13643-025-02817-z] [Medline: 40241162]
32. Fernández MB, Herrera MS. Health effects of dependent older people caregiving by family members [Article in Spanish]. Rev Med Chil 2020 Jan;148(1):30-36. [doi: 10.4067/S0034-98872020000100030] [Medline: 32730433]
33. Bernabe-Ortiz A, Diez-Canseco F, Vásquez A, Miranda JJ. Disability, caregiver's dependency and patterns of access to rehabilitation care: results from a national representative study in Peru. Disabil Rehabil 2016;38(6):582-588. [doi: 10.3109/09638288.2015.1051246] [Medline: 26017542]
34. Holthe T, Jentoft R, Arntzen C, Thorsen K. Benefits and burdens: family caregivers' experiences of assistive technology (AT) in everyday life with persons with young-onset dementia (YOD). Disabil Rehabil Assist Technol 2018 Nov;13(8):754-762. [doi: 10.1080/17483107.2017.1373151] [Medline: 28891356]
35. Thompson GA, Segura J, Cruz D, Arnita C, Whiffen LH. Cultural differences in patients' preferences for paternalism: comparing Mexican and American patients' preferences for and experiences with physician paternalism and patient autonomy. Int J Environ Res Public Health 2022 Aug 26;19(17):10663. [doi: 10.3390/ijerph191710663] [Medline: 36078378]
36. Norman I. Blind trust in the care-giver: is paternalism essential to the health-seeking behavior of patients in sub-Saharan Africa? Adv Appl Sociol 2015;05(2):94-104. [doi: 10.4236/aasoci.2015.52008]
37. Málaga G, Romero ZO, Málaga AS, Cuba-Fuentes S. Shared decision making and the promise of a respectful and equitable healthcare system in Peru. Z Evid Fortbild Qual Gesundhwes 2017 Jun;123-124:81-84. [doi: 10.1016/j.zefq.2017.05.021] [Medline: 28539192]
38. Thompson GA, Whiffen LH. Can physicians demonstrate high quality care using paternalistic practices? A case study of paternalism in Latino physician-patient interactions. Qual Health Res 2018 Oct;28(12):1910-1922. [doi: 10.1177/1049732318783696] [Medline: 29962283]
39. Lazo-Porras M, Bernabe-Ortiz A, Taype-Rondan A, et al. Foot thermometry with mHeath-based supplementation to prevent diabetic foot ulcers: a randomized controlled trial. Wellcome Open Res 2020;5:23. [doi: 10.12688/wellcomeopenres.15531.2] [Medline: 32923686]
40. Brandt LR, Hidalgo L, Diez-Canseco F, et al. Addressing depression comorbid with diabetes or hypertension in resource-poor settings: a qualitative study about user perception of a nurse-supported smartphone app in Peru. JMIR Ment Health 2019 Jun 18;6(6):e11701. [doi: 10.2196/11701] [Medline: 31215511]

Abbreviations

APK: Android package kit
COM-B: capability, opportunity, motivation–behavior
COREQ: Consolidated Criteria for Reporting Qualitative Research
DFU: diabetic foot ulcer
mHealth: mobile health
TDF: Theoretical Domains Framework

XSL•FO
RenderX

Edited by L Buis; submitted 13.May.2026; peer-reviewed by A Searle, D Levy; revised version received 01.Sep.2026; accepted 02.Sep.2026; published 10.Nov.2026.

Please cite as:
Rebull SM, Almeida-Huanca G, Jauregui Mantilla J, Cruzado-Castro L, García Cruz S, Perel P, Miranda JJ, Siddique J, Seiglie J, Liu H, Jan S, Downey L, Beran D, Armstrong DG, Comina G, Lazo-Porras M, Zafra-Tanaka JH
Behavioral Interactions Shaping the Use of a Mobile Health Intervention for Diabetic Foot Self-Care Among People With Diabetes and Caregivers: COM-B and TDF Qualitative Analysis
JMIR Mhealth Uhealth 2026;14:e101195
URL: https://mhealth.jmir.org/2026/1/e101195
doi:10.2196/101195

© Sofia Malena Rebull, Guillermo Almeida-Huanca, Josselyn Jauregui Mantilla, Lourdes Cruzado-Castro, Sebastian García Cruz, Pablo Perel, Juan Jaime Miranda, Juned Siddique, Jacqueline Seiglie, Hueiming Liu, Stephen Jan, Laura Downey, David Beran, David G Armstrong, German Comina, María Lazo-Porras, Jessica Hanae Zafra-Tanaka. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Development and Launch of a Dutch Mobile App (MediMama) on Over-the-Counter Medication Safety During Pregnancy and Breastfeeding: Development and Usability Study

Veronique YF Maas[1], PhD; Maud de Feijter[1], PhD; Anneke LM Passier[1], PhD; Miranda HM van Tuyl[1], MSc; Agnes C Kant[1,2], Prof Dr; Maartje Conijn[1], PhD

[1]Department Teratalogy Information Service, Netherlands Pharmacovigilance Centre Lareb, Goudsbloemvallei 7, 's-Hertogenbosch, The Netherlands
[2]Department of Clinical Pharmacology and Toxicology, Leiden University Medical Centre, Leiden, The Netherlands

Corresponding Author:
Veronique YF Maas, PhD
Department Teratalogy Information Service, Netherlands Pharmacovigilance Centre Lareb, Goudsbloemvallei 7, 's-Hertogenbosch, The Netherlands

Abstract

Background: Over-the-counter (OTC) medicines are frequently used during pregnancy. As these medicines are often used without medical supervision, accessible and reliable safety information is essential. However, finding reliable and understandable information on the safety of these medicines during pregnancy is often experienced as difficult. Hence, there is a need for a new easily accessible electronic health (eHealth) tool that empowers women to actively seek information to support safer self-medication practices during pregnancy and breastfeeding.

Objective: This study aimed to describe the development and dissemination process of a Dutch mobile app providing reliable safety information on OTC medicines during pregnancy and breastfeeding using a development and formative evaluation approach.

Methods: The app was developed over a 2-year project comprising 5 phases, including preparation, development, preimplementation, implementation, and evaluation. Mixed-method strategies, including questionnaires, focus groups, and user feedback rounds, were applied to involve the target population in the development process. Medicine safety information in the app was based on the latest scientific evidence. First-year app-usage outcomes included app downloads, usage patterns, and information-seeking behavior.

Results: Input from 253 potential users formed the foundation for the development of the MediMama app (Netherlands Pharmacovigilance Centre Lareb), with users expressing a need for clear, reliable, and easily accessible information on medication safety during pregnancy and breastfeeding. The app was launched on Mother's Day 2024 and provides safety information on over 250 OTC medicines, including supplements and herbal remedies, across 27 medicine categories. Promotion occurred through multiple online and offline channels. During its first year, the MediMama app was downloaded 22,415 times, with an average of 370 unique daily users, indicating substantial user engagement. Information on paracetamol (acetaminophen) and nasal sprays was most frequently accessed, reflecting the need for information on commonly used OTC medicines among the target population.

Conclusions: One year after its launch, the MediMama app is considered a promising tool in maternity care, meeting the target population's need for accessible OTC medicine safety information. The app aims to support informed decision-making, contributing to safer medication use during pregnancy and breastfeeding. Further research is required to evaluate the effectiveness of the implementation strategy, as well as the app's impact on maternal medication use behaviors and health outcomes.

(JMIR Mhealth Uhealth 2026;14:e85948) doi:10.2196/85948

KEYWORDS

medicines; pregnancy; breastfeeding; lactation; mobile app; development; safety; information; over-the-counter medicines.

Introduction

Medication use during pregnancy and breastfeeding is common, with an estimated 80% - 90% of women reporting the use of at least one (prescribed) medicine during pregnancy [1,2]. Nonprescription or over-the-counter (OTC) medicines are frequently used to manage pregnancy-related or seasonal symptoms such as heartburn, pain, or cold symptoms [3,4]. Although readily available in supermarkets and drugstores, OTC medicines are not without risk. Inappropriate use of some OTC medicines may increase the risk of congenital anomalies, adverse perinatal outcomes, and long-term health consequences for the offspring [3,5]. For instance, use of nonsteroidal anti-inflammatory drugs such as ibuprofen or naproxen in the second half of pregnancy has been associated with higher risks

XSL•FO
RenderX

of constriction of the ductus arteriosus, oligohydramnios, and fetal renal dysfunction [6-8]. Similarly, the use of certain antihistamines during breastfeeding may reduce milk supply or cause sedation and irritability in infants, especially at high doses or with prolonged use [9,10]. While some women may be unaware of these potential risks, others tend to overestimate teratogenic risks and may be reluctant to use necessary medications to manage health conditions (eg, fever or pain during pregnancy), which in some cases could result in greater harm than the medications themselves [11,12].

Empowering women to make informed medication choices is therefore essential. Compared with prescription medicines, OTC medicines are widely available, less stringently regulated, and frequently used without professional guidance [3]. This shifts the responsibility for safety assessment to the individual user. Uninformed decisions about taking OTC medicines (including herbs, homeopathic remedies, and supplements) may lead to the usage of unsafe medicines without the user even being aware. While teratogenic risk information is available online, it has been suggested to be unclear, hard to find, or challenging to interpret for the lay audience. Hence, new interventions tailored to the target population are needed. Electronic health (eHealth) apps are described as the next generation in prenatal care and have been demonstrated to improve antenatal care quality, medication adherence, and pregnancy outcomes [13-16]. Research suggests that up to 95% of pregnant women search online for health information [17,18]. Mobile apps have become one of the primary sources of information and tend to rapidly replace traditional methods of obtaining information from health professionals and written resources [13,19-21]. Consequently, there are more apps available with pregnancy-related information compared to any other medical topic at the time of this writing [22].

While online information is easily accessible for women of reproductive age, it also entails risks, particularly regarding the reliability and accuracy of online sources [23,24]. Conflicting information from multiple online sources is common and can result in women seeking additional unreliable sources, experiencing increased anxiety or even avoiding necessary medication [18,25,26]. For example, previous research suggests that the main reasons women are hesitant to take medication during lactation include fear of harming the infant, lack of clear and consistent information, and conflicting advice (eg, between health care providers and the often outdated and overly restrictive information provided in the product label) [27-31]. This can lead women to discontinue breastfeeding or avoid necessary medication, even though most medications are compatible with breastfeeding [32,33]. Therefore, it is important to develop a single tool that is integrated across all stages of maternity care, providing reliable, clear, and comprehensible safety information.

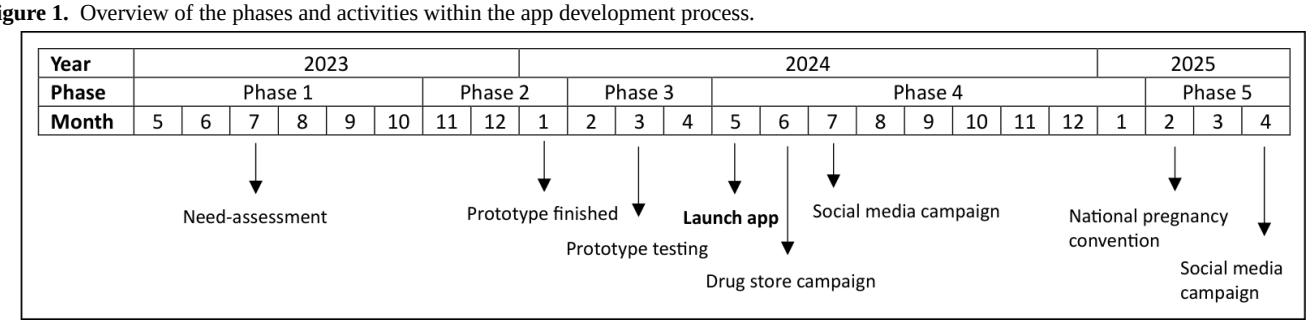
Although several apps providing medication safety information for pregnancy or breastfeeding have been developed, awareness and use of these tools remain suboptimal, as they are often not easily accessible or widely known to users [34-36]. In addition, since brands of OTC medicines, use patterns, and regulatory recommendations on safety during pregnancy and lactation differ between countries, each country bears responsibility for providing, maintaining, and effectively disseminating locally relevant recommendations on medication safety [37,38]. In response to these challenges, this study aims to describe the development and dissemination process of a free and easily accessible Dutch mobile app on the safety of OTC medicines during pregnancy and breastfeeding.

Methods

Overview

The app is developed by the Dutch Teratology Information Service (TIS), part of Pharmacovigilance Centre Lareb. The Dutch TIS is the national knowledge center for medication safety during pregnancy and breastfeeding [39]. TIS provides evidence-based information on over 1100 (primarily prescription) drugs through a knowledge bank on its public website, which receives approximately 1.5 million visits annually [40]. Next to the knowledge bank, there was a desire for easier and faster access to information on OTC medications, for example, when standing in the drugstore. Inspired by a widely used Dutch app on the safety of food during pregnancy (the ZwangerHap app; The Netherlands Nutrition Centre) [41], a grant proposal for the development of a new app on the safety of OTC medication was submitted to and approved by the Netherlands Organization for Health Research and Development (ZonMw; grant number: 10140162210007). The 2-year development and implementation project started on May 1, 2023. This study followed a development and formative evaluation approach in which data were collected through questionnaires and focus groups to iteratively refine the new mobile app. This study followed 5 consecutive phases, including preparation, development, preimplementation, implementation, and evaluation (Figure 1). The month numbers displayed in Figure 1 represent the month of the year.

Figure 1. Overview of the phases and activities within the app development process.



XSL•FO
RenderX

Phase 1. Preparation Phase

First, a project group consisting of representatives from the professional associations of midwives and gynecologists, as well as a delegate from the umbrella organizations representing drugstores, was assembled. Next, a needs assessment was conducted among the target population using 2 tailored online questionnaires, one for women who were pregnant or gave birth within the past year and one for health care professionals working with this group. These were the only inclusion criteria for participating in the needs assessment. Questions focused on information-seeking behavior, content preferences, and distribution strategies. Given their brief and exploratory nature, the questionnaires primarily assessed opinions and prior behaviors and were therefore not pilot-tested nor formally validated. A translated version of one of the distributed questionnaires is provided as supplemental material. Participants were recruited using convenience and snowball sampling approaches, primarily through the Dutch Pregnancy Drug Registry (as part of the Dutch TIS), social media, an umbrella organization of drugstores, and the TIS's professional network. Together, these recruitment strategies aimed to provide a comprehensive representation of the target population, thereby reducing the risk of sampling bias. The target sample size of 100 completed questionnaires was determined based on feasibility considerations and the exploratory aims of the study. The completed questionnaires were analyzed descriptively and extensively discussed within the project group, and the outcomes were used as a basis for the development phase.

Phase 2. Development Phase

During the development phase, an initial draft of the app was devised together with a marketing agency and information technology partner. When developing the app prototype, insights from the needs assessment were used, and discussions with the project group addressed which barriers health care professionals needed to overcome before recommending the app. Key decisions included the navigation structure, type of information, required warnings, interface design, app name, and integration with the existing TIS knowledge bank on the website for extended content. A comprehensive list of OTC medicines, such as analgesics, antacids, vitamins, and herbal supplements, was based on the most commonly used OTC medicines during pregnancy in the Dutch Pregnancy Drug Registry and further extended with OTC medicines identified during field visits to pharmacies, drugstores, and supermarkets. Medicine safety information included in the app was drafted by teratology experts from the Dutch TIS, based on the latest scientific available evidence. Content drafting was supervised by a medical doctor and a pharmacist. All displayed content was fully managed in-house, allowing for continuous updates and modifications as needed. Finally, in consultation with legal experts, privacy terms and disclaimers were carefully drafted. By the end of this phase, a fully functional app prototype was completed.

Phase 3. Preimplementation Phase

Before the launch, a prototype of the app was extensively tested among the target population using a mixed method approach by collecting both qualitative data (focus groups) and quantitative data (questionnaires). First, focus groups were conducted consisting of target users and teratology experts. Participants were recruited through convenience and snowball-based approaches, primarily via midwifery practices, pregnancy fitness classes, and professional networks. Each focus group was scheduled for a maximum of one hour. Participants first completed exercises designed to facilitate navigation through the app, after which the facilitator led a group discussion following a predesigned script, eliciting feedback aimed at improving the app. Sessions were audio-recorded with consent, and all feedback points were documented and descriptively analyzed without the use of an analytical framework. We aimed to conduct 5 focus groups with a minimum of 5 participants each, a target determined pragmatically in light of feasibility and the short data collection timeframe. Subsequently, an online questionnaire, mirroring the focus group tasks, was distributed among health care providers and women who were pregnant or gave birth within the past year. These were the only inclusion criteria for administering the questionnaire. The aim was to collect 10 completed questionnaire responses. All feedback from the focus groups and questionnaires was evaluated, and final adjustments to the app were made accordingly.

Phase 4. Implementation Phase

The launch of the app was supported by both broad and targeted communication campaigns. Dissemination relied on an informing strategy, engaging stakeholders to promote the app through online and offline channels. Health care professionals were positioned as key figures by recommending the app during consultations, as part of a motivational approach. Booster campaigns were periodically initiated to reach the continuously changing target population. All feedback received from app users was carefully reviewed, and whenever possible, immediate additions or changes were made to the content.

Phase 5. Evaluation Phase

Evaluation metrics included the number of app downloads, App Store ratings, user feedback, number of daily active users, most searched medications, and total number of page views. Tracking of anonymous in-app user behavior commenced 5 months postlaunch, following an app release. A cautious initial target of 15,000 downloads was set for the first year, based on the approximate annual birth rate of 170,000 in the Netherlands, with the understanding that awareness and adoption would take time to build [42].

Ethical Considerations

This study did not require approval from a medical ethics review committee, as it did not qualify as medical research involving human subjects under the Dutch Medical Research Involving Human Subjects Act. All collected data were anonymized before analysis to protect participant privacy and confidentiality. Participation was voluntary, and informed consent was obtained from all participants prior to participation. Only focus group participants received a gift voucher of their choice as a token of appreciation for their time and contribution.

XSL•FO
RenderX

Results

Phase 1. Preparation Phase

A total of 217 individuals completed the needs assessment questionnaire; 184 (84.8%) were women who were pregnant or had given birth within the past year, and 33 (15.2%) were health care professionals (Table 1). Among the women, 160 out of 184 (87.0%) were pregnant at the moment of administering the questionnaire (mainly in the second trimester), and 10 out of the 24 women who recently gave birth (41.7%) were still breastfeeding. The majority of women (134/184, 72.8%) reported actively seeking safety information before using any medication. The need for information on OTC medicines (142/184, 77.2%) and health products (eg, supplements, herbs, and vitamins; 122/184, 66.3%) was higher than the need for information on prescribed medicines (70/184, 38.0%). Preferred sources of information were official (government) websites (151/184, 82.1%), product information (125/184, 67.9%), and health care providers (115/184, 62.5%). The primary motivation for seeking safety information was (1) to determine whether a medicine can be used during pregnancy and breastfeeding or not (2), which medicine is the preferred treatment option for specific (pregnancy) complaints, and (3) what the possible adverse drug reactions are. Open-ended responses underscored challenges in finding trustworthy and accessible information, and frustration with inconsistent recommendations across sources.

XSL•FO
RenderX

Table . Results of the needs-assessment questionnaires conducted in the preparation phase data are presented as n (%).

Characteristic	Value
Women (n=184)	
Health status, n (%)	
Pregnant, 1st trimester	34 (18.5)
Pregnant, 2nd trimester	89 (48.4)
Pregnant, 3rd trimester	37 (20.1)
Gave birth <1 year, still breastfeeding	10 (5.4)
Gave birth <1 year, not breastfeeding	14 (7.6)
First-time pregnancy, n (%)	
Yes	97 (52.7)
No	87 (47.3)
Frequency of information-seeking behavior before medicine use, n (%)	
Always looks up information	134 (72.8)
Often looks up information	19 (10.3)
Sometimes, only when in doubt	26 (14.1)
(Almost) never	5 (2.7)
Searched information on[a], n (%)	
OTC medicines[b]	142 (77.2)
Other health products[c]	122 (66.3)
Prescribed medicines	70 (38.0)
No need for more information	30 (16.3)
Preferred channel to receive information[a], n (%)	
Online (neutral evidence-based medicine websites)	151 (82.1)
Product information (medicine leaflet)	125 (67.9)
Health care provider	115 (62.5)
Online (pregnancy websites or forums)	92 (50.0)
Shop assistant	28 (15.2)
Family and friends	15 (8.2)
Online pregnancy forum	13 (7.1)
Likelihood of using a new mobile app for OTC medicines[d]	
Mean (SD)	7.8 (2.3)
Score≥7	146 (79.3)
Health care professionals (n=33)	
Profession, n (%)	
Drug store employee	20 (60.6)
Pharmacist	5 (15.2)
Midwife	3 (9.1)
Assistant pharmacist	2 (6.1)
Other	3 (9.1)
Will your job be easier if women use a new mobile app for OTC medicines[e]	
Mean (SD)	6.0 (2.1)

XSL•FO
RenderX

Characteristic	Value
Score≥7	16 (48.4)
Likelihood of using a new mobile app for OTC medicines[d]	
Mean (SD)	6.0 (3.5)
Score≥7	20 (60.6)
Likelihood of recommending a new mobile app for OTC medicines[d]	
Mean (SD)	7.3 (2.4)
Score≥7	25 (75.8)

[a]Multiple answers possible.

[b]OTC medicines: over-the-counter medicines.

[c]The exact answer option was "Information about health products other than OTC medicines (eg, supplements, vitamins, and herbal remedies)".

[d]On a scale of 1 (very unlikely) to 10 (very likely).

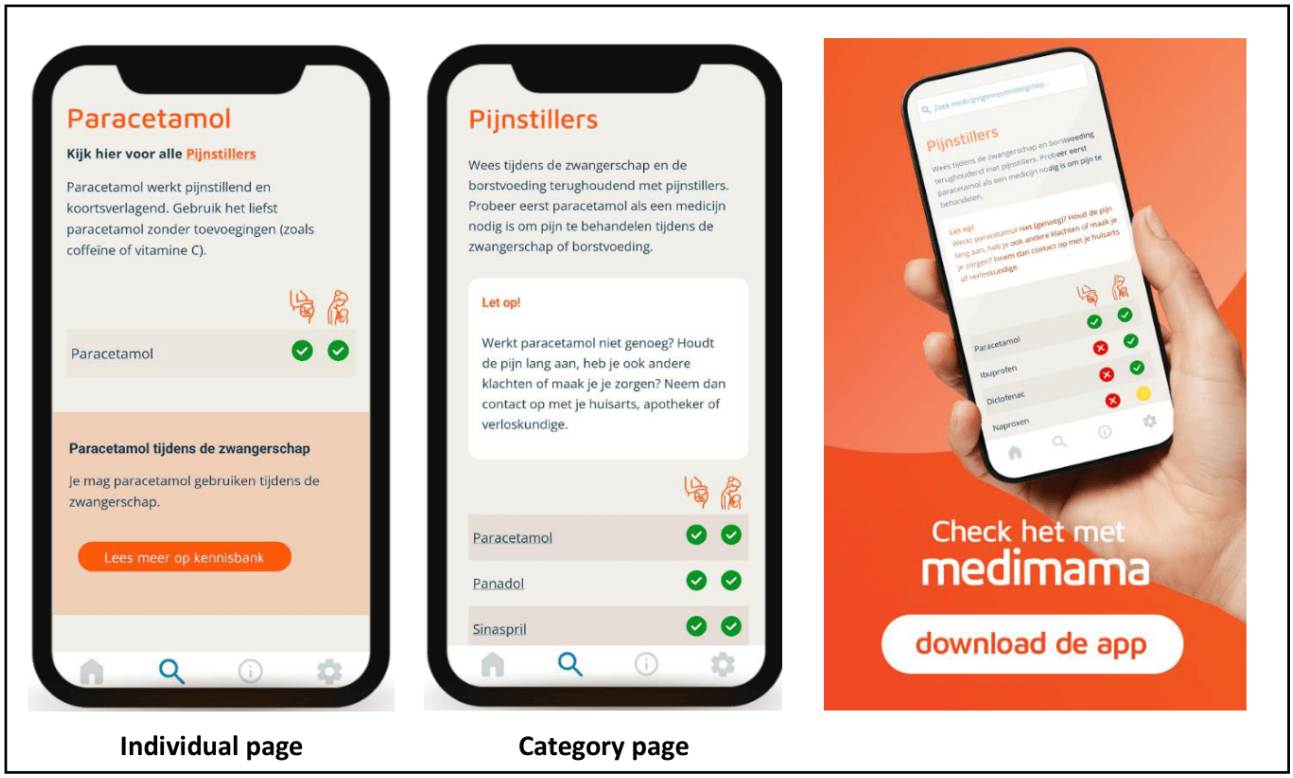
[e]On a scale of 1 (Totally disagree) to 10 (totally agree).

Health care professionals (n=33) reported using the knowledge bank of TIS on the website frequently and appreciated the clear distinction between pregnancy and breastfeeding guidance. The main reasons for seeking safety information were (1) to determine whether or not a medicine can be used during pregnancy and breastfeeding (2), what the possible adverse drug reactions are, and (3) which medicine is the preferred treatment option for specific (pregnancy) complaints. While some believed the app could reduce their workload, others expected women would still seek personal advice. Finally, health care professionals were moderately likely to use the app themselves (mean 6.0, SD 3.5) but more likely to recommend it to patients (mean 7.3, SD 2.4 out of a scale from 1 to 10).

Phase 2. Development Phase

The primary focus in the user experience testing was to make sure that the required information was found as quickly as possible and was comprehensive and easy to understand. For this purpose, all medicines in the app are classified within 3 risk categories using the traffic light colors, including green (can be used), yellow (be careful and read more information), and red (not to be used). Safety information in the app is distributed through 2 kinds of information pages (Figure 2):

Figure 2. Overview of the different pages and the look and feel of the MediMama app.



1. An individual page contains short and easily understandable product-specific safety and usage information. The app contains 260 individual pages, including medicines, products, vitamins, and herbals. On all individual pages, a referral button to the specific information on the TIS website is placed for more detailed information.

XSL•FO
RenderX

2. The category page provides an overview of a group of medicines or products where all risk categories from all relevant medicines are displayed in a table. The app contains 28 category pages, for instance, analgesics, antacids, or hay fever tablets. Additionally, this page provides lifestyle recommendations to relieve the (pregnancy) complaint before the initiation of medication.

Users can search by active ingredient, brand name, or health complaints. The app name, MediMama (Netherlands Pharmacovigilance Centre Lareb), was selected in collaboration with followers of a large Dutch pregnancy platform. The look and feel of the MediMama app are displayed in Figure 2. Based on user preferences, the app was developed as a native app as well as a web-based version.

Phase 3. Preimplementation Phase

The prototype was tested in 5 focus groups (n=23 participants) and via an online questionnaire (n=14; Table 2). During the focus groups, thematic saturation was observed. Participants in the focus groups reported that the app was easy to use, had an appealing look and feel, and the displayed safety information with the 3 colors was a big plus. Minor technical issues were noted on certain devices. Key suggestions included shortening texts, improving ease of reading, removing account registration, improving the disclaimer text, and considering adding more information about medication usage, such as duration and dosage. The online questionnaire was completed by 6 out of 14 (42.9%) women who were pregnant during the administration of the questionnaire or gave birth within one year of completing the questionnaire and 8 out of 14 (57.1%) health care providers with different backgrounds (Table 2). All questionnaire respondents rated the app's ease of use ≥8/10. Most women reported that they would probably (3/14, 21.4%) or definitely (11/14, 78.6%) use or recommend the app. Suggestions were made to improve specific texts (eg, the disclaimer) and to add references to other sources of information for more detailed information on, for example, medication usage or the health condition. All health care providers agreed with the statement that the information in the app is appropriate and suitable for women who are pregnant or breastfeeding.

XSL•FO
RenderX

Table . Results of the mixed methods prototype testing of the MediMama app conducted in the preimplementation phase.

Category	Findings
Focus groups (n=23 participants)	
Focus group sessions	
Online (team session)	Teratology experts (n=9)
Local gym	Pregnant women or women who gave birth <1 year (n=4)
Online	Pregnant women or women who gave birth <1 year (n=3)
Online	Women who gave birth <1 year (n=2)
Midwifery practice	Pregnant women (n=4) and midwife (n=1)
Examples of technical issues	Disclaimer page crashed on some devices; formatting inconsistencies across devices; design differences between dark and light phone settings
Examples of positive aspects	Appealing layout and app name; easy to search for information; positive feedback on referrals to other platforms for more detailed information
Examples of suggestions to improve	Texts could be shorter and include less jargon; remove the option of creating a personal profile; include the category page in search results when searching for an individual medicine; onboarding information can be overwhelming
Online questionnaires (n=14 participants)	
Target population	
Pregnant	Second trimester (n=1); third trimester (n=2)
Gave birth <1 year	Breastfeeding (n=2); not breastfeeding anymore (n=1)
General practitioner	3
Midwife	2
Other HCP[a]	3
MediMama app easy to use[b]	
Mean	8.6
Score ≥8	14 (100)
Examples of positive aspects	Clear, short, and understandable information; simple design with clear symbols and colors; clear distinction between pregnancy and breastfeeding; useful referrals to other platforms for more detailed information
Examples of suggestions to improve	Improve information in the disclaimer and onboarding; explicitly refer users to the prescribing doctor for information on prescribed medication; consider including preconception recommendations; be cautious when categorizing all vitamins as “green” (safe to use)
Would you use or recommend the MediMama app?	
Definitely not or maybe	0 (0)
Probably	3 (21.4)
Definitely	11 (78.6)

[a]HCP: health care professional.

[b]On a scale from 1 (difficult to use) to 10 (very easy to use).

Phase 4. Implementation Phase

The MediMama app was launched on Mother’s Day (May 12th), 2024, and has been freely available in the app stores since then [43,44]. During the first year, the MediMama app was promoted through various channels, with the 4 main promotional strategies being the following:

1. Drug store campaign: in over 800 drug stores, MediMama “shelf cards” were displayed. This generated significant visibility and raised awareness among store employees and visitors.
2. Online social media campaign: targeted campaigns via Instagram (Meta Platforms), Facebook (Meta Platforms), and Google Play (Google Inc) were conducted. Campaigns were updated every few weeks to optimize reach. These platforms use advanced algorithms and machine learning to optimize advertorial performance.

XSL•FO
RenderX

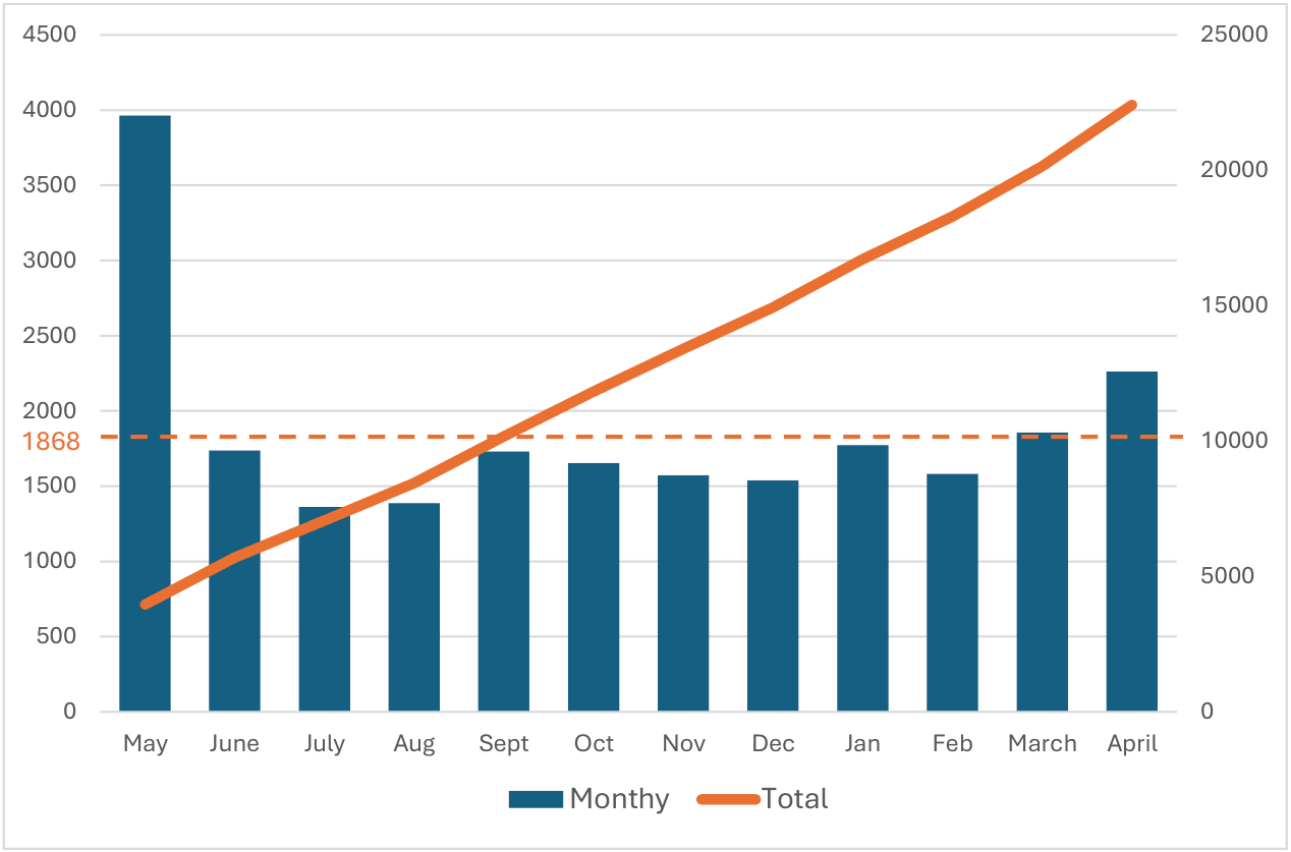
3. Dutch pregnancy convention: over 30,000 (pregnant) visitors attended a 4-day Dutch pregnancy convention. At which the MediMama app was promoted through a booth, daily master classes, and giveaways.
4. To reach health care providers from various disciplines, the development and launch of the app were promoted through channels as presentations at national medical conferences, posters and flyers in waiting rooms, local health care provider meetings, publications in professional journals, and news items from several professional associations.

User feedback led to several app improvements. In response to frequent requests, the most searched prescription medicines were also added. The prescribed medicines in the app all receive the "yellow risk category" (be careful and read more information) and directly refer to the medicine-specific information on the TIS website for more detailed information and the advice to discuss this medication use with their physician.

Phase 5. Evaluation Phase

In its first year, the app was downloaded 22,415 times, averaging 1868 downloads per month (Figure 3). Peak activity occurred in the launch month (n=3963). After the initial peak of reaching the early adopters, the average daily number of downloads gradually increased from 44 per day up to 75 per day. The social media campaign proved to be one of the most effective strategies to reach the target population. On average, app users viewed 3 pages per session; in total, on average, 24,352 pages were viewed each month (Figure 4). The number of daily app users increased from 207 per day to 370 daily app users over the course of 6 months. However, the growth in the number of downloads and app users did not follow a consistent upward trend each month over the course of the first year, resulting in some fluctuation in usage (eg, the short month of February or seasonal changes). To date, the MediMama app has received 11 reviews across both app stores, all of which have awarded it the maximum rating of 5 stars. Most of the feedback received via email and social media concerned the need for the app to be available in multiple languages and to include information on prescribed medications. Paracetamol (acetaminophen) and nasal sprays turned out to be the most frequently searched and viewed medicines in the app (Table 3). While category pages received more overall views, individual pages were more frequently accessed via search. This indicates a distinction in user behavior; searches were primarily aimed at specific medicines, whereas the higher number of views on category pages suggests that users actively navigated through the app to gain a more comprehensive understanding of an entire group of medicines. Finally, seasonal trends in search terms are visible corresponding with the seasonal health complaints (Figure 5). For example, in the flu season (December-February), information-seeking behaviors leaned toward medicines like nasal sprays or cough syrups, while a peak in hay fever medicine (eg, cetirizine tablets) was noticeable from March onward.

Figure 3. Overview of the MediMama downloads within the first year**.** The left Y-axis scale represents the number of monthly downloads (blue); the right Y-axis scale represents the total number of downloads (orange); the orange dotted line represents the average number of monthly downloads; the numbers within the blue bars represent the average number of daily downloads.



XSL•FO
RenderX

Figure 4. Overview of the MediMama users. The left Y-axis scale represents the number of monthly views of all MediMama pages within the app (blue); the right Y-axis scale represents the monthly number of app users (orange); the dotted orange line represents the average number of monthly page views; the numbers within the blue bars represent the average number of daily app users.

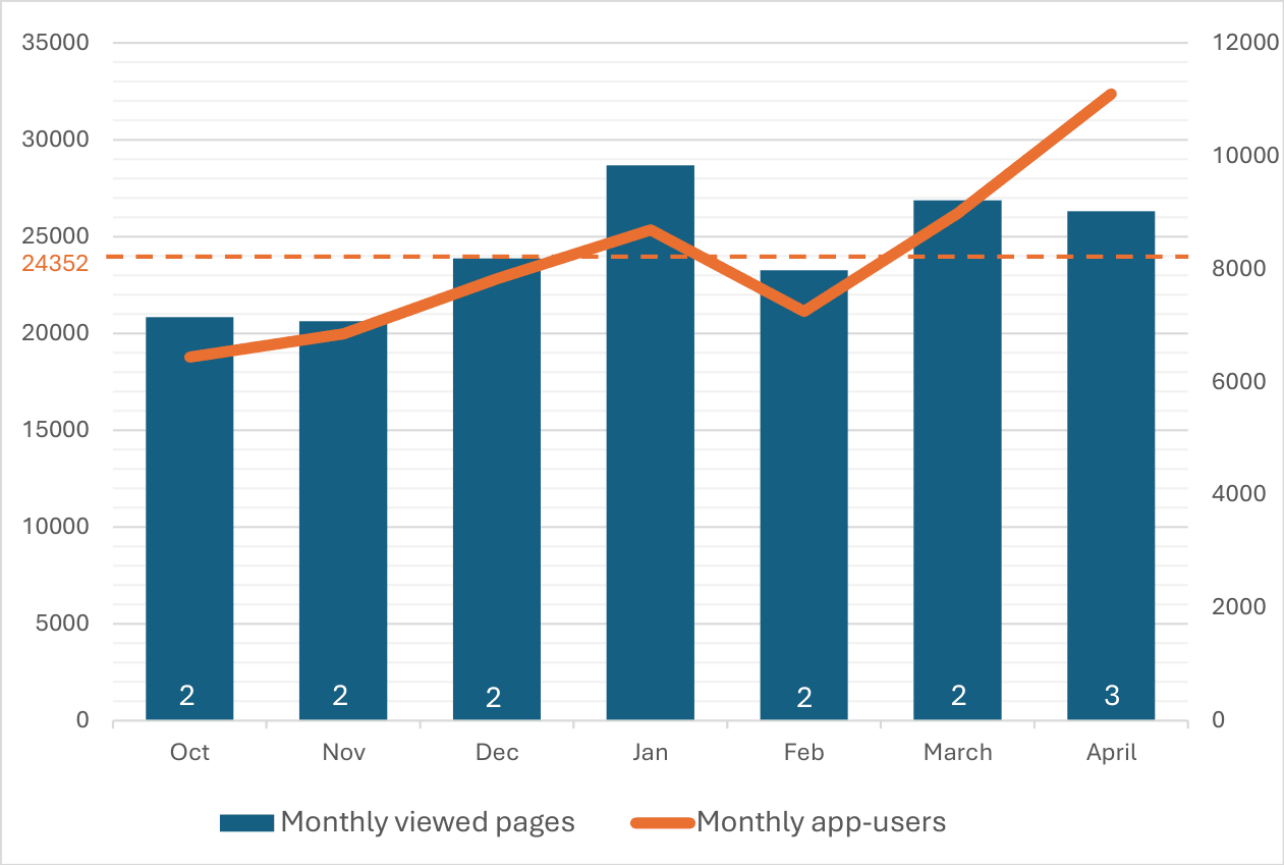


Table . Top 10 most searched terms and viewed medicine pages in the MediMama app.

Most viewed		Top	Most searched	
Type[a]	Title		Title	Type
I	Paracetamol[b]	1	Paracetamol[b]	I
C	Nasal sprays for colds	2	Nasal sprays for colds	C
C	Sore throat relievers	3	Ibuprofen	I
I	Xylometazoline nasal spray	4	Cough remedies	C
C	Cough remedies	5	Strepsils cough drop	I
C	Painkillers	6	Antacids	C
C	Antacids	7	Rennie[c]	I
I	Ibuprofen	8	Sore throat relievers	C
I	Strepsils cough drop	9	Xylometazoline nasal spray	I
C	Hemorrhoid remedies	10	Tablets for allergy and hay fever	C

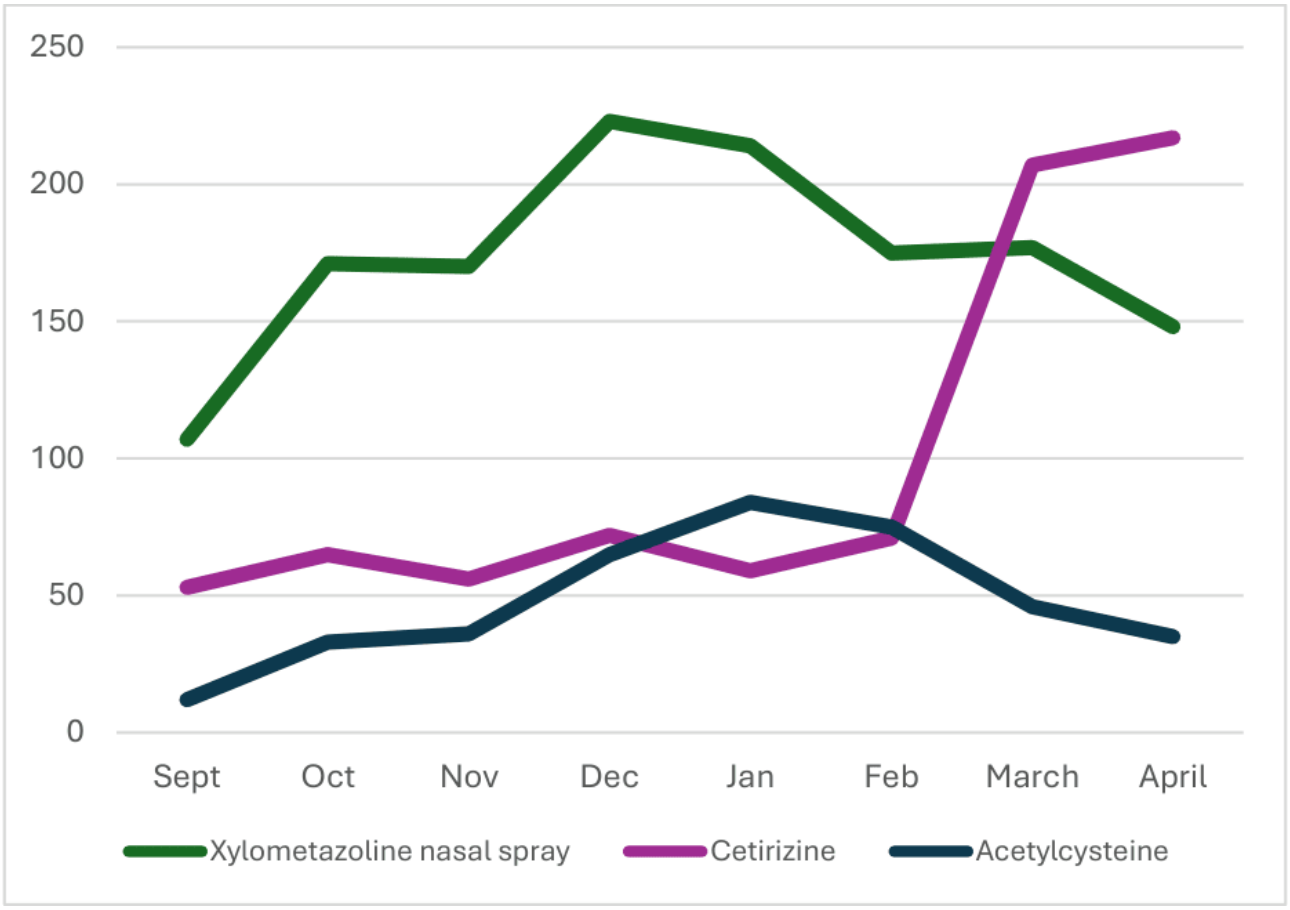
[a]Either an individual page (I) or a category page (C)

[b]Also known as acetaminophen.

[c]Brandname of an antacids containing calcium carbonate and magnesium carbonate.

XSL•FO
RenderX

Figure 5. Examples of seasonal trends in search terms for 3 medicines in the MediMama app. Numbers correspond with the monthly number of times xylometazoline nasal spray, cetirizine (hay fever medicine), and acetylcysteine (cough syrup) were searched.



Discussion

Key Findings

This study details the development and dissemination process of MediMama, a Dutch mobile app providing evidence-based information on the safety of OTC medicines and products during pregnancy and breastfeeding. Preimplementation research suggested that the preferred route for receiving safety information on medicine was online and emphasized the need for a new mobile app. Involving the target population throughout development ensured alignment with user needs. One year postlaunch, the MediMama app is considered a promising eHealth tool, addressing the demand for accessible, reliable OTC medicine safety information in maternity care.

Comparison to the Literature

Previous eHealth apps in maternity care have demonstrated a successful tool in improving safe medication use and overall maternal health outcomes during pregnancy and lactation [15]. Educational apps and clinical decision support tools have been shown to improve medication adherence, increase antenatal care attendance, and support behavioral changes for topics such as healthy diet, smoking cessation, and vaccination uptake [16,45-47]. Usability and acceptability of mobile health apps during pregnancy are generally high, with pregnant women and health care providers reporting satisfaction and willingness to recommend these tools [16,45,48,49]. Some previously developed apps on the safety of medication use during pregnancy or breastfeeding have been developed, for instance, LactSafe (NIH LactMed - MotherToBaby (OTIS)) [34], MommyMeds (InfantRisk Center) [35], and Meds4Mums2B (Innovative Health Initiative (IHI) ConcePTION, in collaboration with the Medicines and Healthcare products Regulatory Agency (MHRA)) [36]. The MediMama app differs from existing tools by being freely accessible worldwide, providing comprehensive safety information for both pregnancy and breastfeeding based on scientific literature. In addition, the MediMama app has been scientifically evaluated, and data on usability and user satisfaction are available. Because of limited available evidence, the effectiveness of mobile health apps in improving maternal knowledge, behavior, and perinatal outcomes remains inconclusive [45,50].

Despite their potential, eHealth tools during pregnancy also carry certain risks with regard to self-treatment [19]. Self-medication during pregnancy is prevalent and can lead to inappropriate drug selection, misdiagnosis, dosing errors, delayed medical help–seeking, and inefficient treatment which may result in adverse maternal and fetal outcomes [51,52]. The MediMama app aims to mitigate these risks by using intuitive, color-coded guidance that is immediately interpretable at a glance and by advising on how to treat a complaint through lifestyle changes instead of medicines (ie, for heartburn). With regard to the quality of eHealth apps, previous reviews elaborated on the growing concern over the quality of several (commercial) pregnancy apps with regard to poor or unclear evidence base of content, lack of expert involvement, and poor

XSL•FO
RenderX

validation [19,53-56]. Emerging artificial intelligence–based health tools further complicate this landscape, with risks of misinformation due to limited contextualization and transparency [57,58]. To navigate these waters of conflicting health information, health care providers can play a central role in advocating for evidence-based trustworthy health information provided by noncommercial public institutions.

Pregnancy and breastfeeding represent periods of heightened health awareness, yet implementation of interventions is challenging due to the dynamic and transient target population [59,60]. The relatively short duration of pregnancy and the subsequent breastfeeding period means that the target group is constantly changing, with individuals entering and exiting the eligible target population in a matter of months. During the implementation phase of the MediMama app, we also experienced the considerable effort it required to reach this constantly changing target population. We have experienced that focusing on online marketing and involving health care providers in disseminating the app could be a promising and sustainable approach to reaching this constantly changing target population.

Strengths and Limitations

This is the first study to systematically describe the development and dissemination process of a mobile app focused on medicine safety in pregnancy and breastfeeding, using mixed methods to capture user and provider perspectives. The diverse set of offline and online strategies used to distribute the MediMama app enabled us to reach this constantly changing target population. Due to this multifaceted promotional approach, the target aim of 15,000 downloads within the first year was reached within the first 7 months. The MediMama app has been well received; the number of daily users keeps increasing, and reviewers praise its simplicity, trustworthiness of the information, and their privacy protection since the app is used anonymously.

However, anonymity limits evaluation of user demographics, hindering assessment of whether the app reaches intended subgroups (eg, pregnant women vs health care providers) and its impact on clinical workload. Consequently, it is difficult to assess whether the app conveys the appropriate tone of voice in presenting safety information. Further user research is therefore needed. Another limitation of our study is that the preimplementation recruitment via the pregnancy registry may have introduced selection bias and the use of unvalidated and not pilot-tested questionnaires to assess user needs. This may have affected the quality and reliability of the collected data. However, the questionnaires were developed in close alignment with the study objectives, supporting their face validity despite the lack of formal validation. Finally, most suggested feedback after the launch of MediMama was making the information in the app available in multiple languages. By limiting the language options to Dutch only, we were likely unable to reach some of the most vulnerable populations, including individuals with limited digital literacy or access, who might benefit most from health guidance. However, vulnerable pregnant women may require additional and fundamentally different forms of health education because a “one-size-fits-all” approach is unlikely to reach a heterogeneous pregnant population with varying levels of health literacy [60].

Future Research

Prior to the large-scale implementation of mobile health apps during pregnancy, more research is needed on their effectiveness. As previous studies suggested, there is still much to learn on the impact of mobile health apps during pregnancy with respect to cost-effectiveness, potential to reduce health care costs and system burden, and their potential positive effect on maternal and neonatal outcomes [19,45]. Since the MediMama app is freely accessible and operates without a profit-generating model, the lack of secured funding poses a significant risk to its ongoing maintenance and sustainability, potentially jeopardizing its continued availability. Further research is needed to identify sustainable approaches to integrate these interventions within maternity care to ensure their continued maintenance, updating, and optimization. In addition, to reach the entire Dutch population of pregnant and breastfeeding women, establishing and maintaining long-term collaboration with health care professionals, such as midwives and gynecologists, is critical. While the target user group changes over time, health care professionals represent a stable and trusted gateway to this population. Further evaluation of effective strategies to continuously reach the different target populations is warranted. Finally, since the multiplicity of mobile apps in pregnancy can be overwhelming and choosing the best one is proven to be difficult for pregnant women, future efforts should explore unified, multilingual platforms that balance usability with comprehensive, evidence-based content [61].

Conclusion

This study describes the development and dissemination process of the MediMama app, designed to provide easily accessible and evidence-based safety information on OTC medicines during pregnancy and breastfeeding. One year postlaunch, MediMama has been a well-received and frequently used eHealth resource in the Netherlands, effectively meeting the information needs of its target population. While initial feedback and usage patterns of MediMama are promising, the behavioral impact of the app remains untested, and its usability has yet to be formally assessed, warranting further evaluation. Further research is required to evaluate the effectiveness of the implementation strategy, as well as the app’s impact on maternal medication use behaviors and health outcomes. Mobile health is emerging as a pivotal tool in maternity care; increased emphasis is required on the continued development of these tools to improve the health of future generations.

XSL•FO
RenderX

Acknowledgments

We would like to thank all the volunteers who provided feedback in the different steps of the process. We gratefully acknowledge the contributions of our project group, including representatives from KNOV, NVOG, and CBD, as well as our IT and design partners, for their dedicated efforts in developing the MediMama app into the product it is today.

Funding

This research and the development of the MediMama app was funded by The Netherlands Organisation for Health Research and Development (ZonMw). Grant number: 10140162210007. The funder had no involvement in the study design, data collection, analysis, interpretation, or the writing of the manuscript.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Funding acquisition: VYFM, MdF, ACK
Conceptualization: VYFM, MdF, ALMP, MHMvT, ACK, MC
Investigation: VYFM, MdF, ALMP, MC
Supervision: MHMvT, ACK
Writing – original draft: VYFM
Writing – review & editing: VYFM, MdF, ALMP, MHMvT, ACK, MC.
All authors reviewed the results and approved the final version of this manuscript.

Conflicts of Interest

None declared.

References

1. Lupattelli A, Spigset O, Twigg MJ, et al. Medication use in pregnancy: a cross-sectional, multinational web-based study. BMJ Open 2014 Feb 17;4(2):e004365. [doi: 10.1136/bmjopen-2013-004365] [Medline: 24534260]
2. Sanders J, Blaylock R, Dean C, Petersen I, Trickey H, Murphy C. Women's experiences of over-the-counter and prescription medication during pregnancy in the UK: findings from survey free-text responses and narrative interviews. BMJ Open 2023 Mar 1;13(3):e067987. [doi: 10.1136/bmjopen-2022-067987] [Medline: 36858469]
3. Powers EA, Tewell R, Bayard M. Over-the-counter medications in pregnancy. Am Fam Physician 2023 Oct;108(4):360-369. [Medline: 37843943]
4. Mansour O, Russo RG, Straub L, et al. Prescription medication use during pregnancy in the United States from 2011 to 2020: trends and safety evidence. Am J Obstet Gynecol 2024 Aug;231(2):250. [doi: 10.1016/j.ajog.2023.12.020] [Medline: 38128861]
5. Zafeiri A, Raja EA, Mitchell RT, Hay DC, Bhattacharya S, Fowler PA. Maternal over-the-counter analgesics use during pregnancy and adverse perinatal outcomes: cohort study of 151 141 singleton pregnancies. BMJ Open 2022 May 3;12(5):e048092. [doi: 10.1136/bmjopen-2020-048092] [Medline: 35504638]
6. Dathe K, Frank J, Padberg S, Hultzsch S, Beck E, Schaefer C. Fetal adverse effects following NSAID or metamizole exposure in the 2nd and 3rd trimester: an evaluation of the German Embryotox cohort. BMC Pregnancy Childbirth 2022 Aug 26;22(1):666. [doi: 10.1186/s12884-022-04986-4] [Medline: 36028798]
7. D'Ambrosio V, Vena F, Scopelliti A, et al. Use of non-steroidal anti-inflammatory drugs in pregnancy and oligohydramnios: a review. J Matern Fetal Neonatal Med 2023 Dec;36(2):2253956. [doi: 10.1080/14767058.2023.2253956] [Medline: 38092425]
8. Koren G, Florescu A, Costei AM, Boskovic R, Moretti ME. Nonsteroidal antiinflammatory drugs during third trimester and the risk of premature closure of the ductus arteriosus: a meta-analysis. Ann Pharmacother 2006 May;40(5):824-829. [doi: 10.1345/aph.1G428] [Medline: 16638921]
9. Ngo E, Spigset O, Lupattelli A, et al. Antihistamine use during breastfeeding with focus on breast milk transfer and safety in humans: a systematic literature review. Basic Clin Pharmacol Toxicol 2022 Jan;130(1):171-181. [doi: 10.1111/bcpt.13663] [Medline: 34587362]
10. Ito S, Blajchman A, Stephenson M, Eliopoulos C, Koren G. Prospective follow-up of adverse reactions in breast-fed infants exposed to maternal medication. Am J Obstet Gynecol 1993 May;168(5):1393-1399. [doi: 10.1016/s0002-9378(11)90771-6] [Medline: 8498418]
11. Ceulemans M, Lupattelli A, Nordeng H, Odalovic M, Twigg M, Foulon V. Women's beliefs about medicines and adherence to pharmacotherapy in pregnancy: opportunities for community pharmacists. Curr Pharm Des 2019;25(5):469-482. [doi: 10.2174/1381612825666190321110420] [Medline: 30907309]

XSL•FO
RenderX

12. Amundsen S, Øvrebø TG, Amble NMS, Poole AC, Nordeng H. Risk perception, beliefs about medicines and medical adherence among pregnant and breastfeeding women with migraine: findings from a cross-sectional study in Norway. BMJ Open 2019 Feb 27;9(2):e026690. [doi: 10.1136/bmjopen-2018-026690] [Medline: 30819714]
13. van den Heuvel JF, Groenhof TK, Veerbeek JH, et al. eHealth as the next-generation perinatal care: an overview of the literature. J Med Internet Res 2018 Jun 5;20(6):e202. [doi: 10.2196/jmir.9262] [Medline: 29871855]
14. Frid G, Bogaert K, Chen KT. Mobile health apps for pregnant women: systematic search, evaluation, and analysis of features. J Med Internet Res 2021 Oct 18;23(10):e25667. [doi: 10.2196/25667] [Medline: 34524100]
15. Hsieh MF, Slavin V, Huang MZ, Wang HH, Hsieh HF. mHealth use during pregnancy: a systematic review. Int J Gynaecol Obstet 2025 Oct;171(1):124-130. [doi: 10.1002/ijgo.70205] [Medline: 40347100]
16. Ameyaw EK, Amoah PA, Ezezika O. Effectiveness of mHealth apps for maternal health care delivery: systematic review of systematic reviews. J Med Internet Res 2024 May 29;26:e49510. [doi: 10.2196/49510] [Medline: 38810250]
17. Bert F, Gualano MR, Brusaferro S, et al. Pregnancy e-health: a multicenter Italian cross-sectional study on Internet use and decision-making among pregnant women. J Epidemiol Community Health 2013 Dec 1;67(12):1013-1018. [doi: 10.1136/jech-2013-202584] [Medline: 24072743]
18. Bjelke M, Martinsson AK, Lendahls L, Oscarsson M. Using the internet as a source of information during pregnancy - a descriptive cross-sectional study in Sweden. Midwifery 2016 Sep;40:187-191. [doi: 10.1016/j.midw.2016.06.020] [Medline: 27450590]
19. Carrandi A, Hayman M, Harrison CL. Safety considerations for assessing the quality of apps used during pregnancy: a scoping review. Digit Health 2023;9:20552076231198683. [doi: 10.1177/20552076231198683] [Medline: 37675058]
20. Lee Y, Moon M. Utilization and content evaluation of mobile applications for pregnancy, birth, and child care. Healthc Inform Res 2016 Apr;22(2):73-80. [doi: 10.4258/hir.2016.22.2.73] [Medline: 27200216]
21. Hsiao SH, Hsu HY, Erickson SR, Lu TS, Wu CH. Seeking health information online among U.S. pregnant women: findings from the 2009-2018 National Health Interview Surveys. Women Health 2023 Jan;63(1):59-69. [doi: 10.1080/03630242.2022.2158413] [Medline: 36546357]
22. Tripp N, Hainey K, Liu A, et al. An emerging model of maternity care: smartphone, midwife, doctor? Women Birth 2014 Mar;27(1):64-67. [doi: 10.1016/j.wombi.2013.11.001] [Medline: 24295598]
23. Wallwiener S, Müller M, Doster A, et al. Pregnancy eHealth and mHealth: user proportions and characteristics of pregnant women using web-based information sources-a cross-sectional study. Arch Gynecol Obstet 2016 Nov;294(5):937-944. [doi: 10.1007/s00404-016-4093-y] [Medline: 27084763]
24. Narasimhulu DM, Karakash S, Weedon J, Minkoff H. Patterns of internet use by pregnant women, and reliability of pregnancy-related searches. Matern Child Health J 2016 Dec;20(12):2502-2509. [doi: 10.1007/s10995-016-2075-0] [Medline: 27456311]
25. Hämeen-Anttila K, Nordeng H, Kokki E, et al. Multiple information sources and consequences of conflicting information about medicine use during pregnancy: a multinational internet-based survey. J Med Internet Res 2014 Feb 20;16(2):e60. [doi: 10.2196/jmir.2939] [Medline: 24565696]
26. De Santis M, De Luca C, Quattrocchi T, et al. Use of the internet by women seeking information about potentially teratogenic agents. Eur J Obstet Gynecol Reprod Biol 2010 Aug;151(2):154-157. [doi: 10.1016/j.ejogrb.2010.04.018] [Medline: 20478650]
27. Spencer JP, Thomas S, Trondsen Pawlowski RH. Medication safety in breastfeeding. Am Fam Physician 2022 Dec;106(6):638-644. [Medline: 36521462]
28. Sachs HC, Committee On Drugs. The transfer of drugs and therapeutics into human breast milk: an update on selected topics. Pediatrics 2013 Sep;132(3):e796-e809. [doi: 10.1542/peds.2013-1985] [Medline: 23979084]
29. Breastfeeding challenges: ACOG committee opinion, number 820. Obstet Gynecol 2021 Feb 1;137(2):e42-e53. [doi: 10.1097/AOG.0000000000004253] [Medline: 33481531]
30. Nörby U, Noël-Cuppers B, Hristoskova S, et al. Online information discrepancies regarding safety of medicine use during pregnancy and lactation: an IMI ConcePTION study. Expert Opin Drug Saf 2021 Sep;20(9):1117-1124. [doi: 10.1080/14740338.2021.1935865] [Medline: 34080506]
31. Winterfeld U, Hodson K, Berlin M, et al. Communicating medication risks in pregnancy: towards shared decision making. Front Drug Saf Regul 2025;5:1712216. [doi: 10.3389/fdsfr.2025.1712216] [Medline: 41199813]
32. Hanley L. Medication use during lactation: either a potential contributor to premature weaning or evidence-based support of the mother/child couplet. Clin Ther 2020 Mar;42(3):393-400. [doi: 10.1016/j.clinthera.2020.01.019] [Medline: 32113701]
33. Bertino E, Varalda A, Di Nicola P, et al. Drugs and breastfeeding: instructions for use. J Matern Fetal Neonatal Med 2012 Oct;25 Suppl 4(sup4):78-80. [doi: 10.3109/14767058.2012.715034] [Medline: 22958026]
34. Quick answers for expecting and nursing parents. LactSafe. URL: https://app.lactsafe.com [accessed 2025-09-30]
35. MommyMeds. Google Play. URL: https://play.google.com/store/apps/details?id=com.infantrisk.mommymeds.paid&hl=en [accessed 2025-09-30]
36. Alexe A, Lewis D, Harrison K. An innovative mobile app for the provision of medical information and collection of safety data on exposure to medicines during pregnancy or breastfeeding: an IHI ConcePTION contribution. Repro Female Child Health 2024 Mar;3(1) [FREE Full text] [doi: 10.1002/rfc2.74]

XSL•FO
RenderX

37. Warrer P, Aagaard L, Hansen EH. Comparison of pregnancy and lactation labeling for attention-deficit hyperactivity disorder drugs marketed in Australia, the USA, Denmark, and the UK. Drug Saf 2014 Oct;37(10):805-813. [doi: 10.1007/s40264-014-0215-2] [Medline: 25112669]
38. Noh Y, Yoon D, Song I, Jeong HE, Bae JH, Shin JY. Discrepancies in the evidence and recommendation levels of pregnancy information in prescription drug labeling in the United States, United Kingdom, Japan, and Korea. J Womens Health (Larchmt) 2018 Sep;27(9):1086-1092. [doi: 10.1089/jwh.2017.6792] [Medline: 29878862]
39. Teratology information services (TIS). Netherlands Pharmacovigilance Centre Lareb. URL: https://www.lareb.nl/en/pages/teratology-information-services-tis [accessed 2025-09-30]
40. Medication use around pregnancy. Bijwerkingen Centrum Lareb. URL: https://www.lareb.nl/mvm-kennis [accessed 2025-09-30]
41. Pregnancy app: zwangerhap. Voedingscentrum. URL: https://www.voedingscentrum.nl/nl/thema/apps-en-tools-voedingscentrum/zwangerschap-app-zwangerhap [accessed 2025-09-30]
42. Geboorte (birth). Centraal Bureau van de statistiek (Statistics Netherlands of Central Bureau for Statistics). URL: https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/bevolkingsgroei/geboren-kinderen [accessed 2025-10-01]
43. MediMama: zwanger-medicijnen (medimama: pregnancy medicines). App Store. URL: https://apps.apple.com/nl/app/medimama-zwanger-medicijnen/id6499128340 [accessed 2025-09-30]
44. MediMama: zwanger-medicijnen (medimama: pregnancy medicines). Google Play. URL: https://play.google.com/store/apps/details?id=com.lareb.medimama&hl=en [accessed 2025-09-30]
45. Overdijkink SB, Velu AV, Rosman AN, van Beukering MD, Kok M, Steegers-Theunissen RP. The usability and effectiveness of mobile health technology-cased lifestyle and medical intervention apps supporting health care during pregnancy: systematic review. JMIR Mhealth Uhealth 2018 Apr 24;6(4):e109. [doi: 10.2196/mhealth.8834] [Medline: 29691216]
46. Mohamed H, Ismail A, Sutan R, Rahman RA, Juval K. A scoping review of digital technologies in antenatal care: recent progress and applications of digital technologies. BMC Pregnancy Childbirth 2025 Feb 13;25(1):153. [doi: 10.1186/s12884-025-07209-8] [Medline: 39948493]
47. Hussain T, Smith P, Yee LM. Mobile phone-based behavioral interventions in pregnancy to promote maternal and fetal health in high-income countries: systematic review. JMIR Mhealth Uhealth 2020 May 28;8(5):e15111. [doi: 10.2196/15111] [Medline: 32463373]
48. Lazarevic N, Lecoq M, Bœhm C, Caillaud C. Pregnancy apps for self-monitoring: scoping review of the most popular global apps available in Australia. Int J Environ Res Public Health 2023 Jan 5;20(2):1012. [doi: 10.3390/ijerph20021012] [Medline: 36673768]
49. Carter J, Sandall J, Shennan AH, Tribe RM. Mobile phone apps for clinical decision support in pregnancy: a scoping review. BMC Med Inform Decis Mak 2019 Nov 12;19(1):219. [doi: 10.1186/s12911-019-0954-1] [Medline: 31718627]
50. Daly LM, Horey D, Middleton PF, Boyle FM, Flenady V. The effect of mobile app interventions on influencing healthy maternal behavior and improving perinatal health outcomes: systematic review. JMIR Mhealth Uhealth 2018 Aug 9;6(8):e10012. [doi: 10.2196/10012] [Medline: 30093368]
51. Rahmani A, Hamanajm SA, Fallahi A, Gheshlagh RG, Dalvand S. Prevalence of self-medication among pregnant women: a systematic review and meta-analysis. Nurs Midwifery Stud 2019;8(4):169-175. [doi: 10.4103/nms.nms_77_18]
52. Niriayo YL, Mohammed K, Asgedom SW, Demoz GT, Wahdey S, Gidey K. Self-medication practice and contributing factors among pregnant women. PLoS ONE 2021;16(5):e0251725. [doi: 10.1371/journal.pone.0251725] [Medline: 34014975]
53. Stevens WJM, van der Sande R, Beijer LJ, Gerritsen MG, Assendelft WJ. eHealth apps replacing or complementing health care contacts: scoping review on adverse effects. J Med Internet Res 2019 Mar 1;21(3):e10736. [doi: 10.2196/10736] [Medline: 30821690]
54. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020 Feb 1;27(2):330-340. [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
55. Hayman MJ, Alfrey KL, Waters K, et al. Evaluating evidence-based content, features of exercise instruction, and expert involvement in physical activity apps for pregnant women: systematic search and content analysis. JMIR Mhealth Uhealth 2022 Jan 19;10(1):e31607. [doi: 10.2196/31607] [Medline: 35044318]
56. Brown HM, Bucher T, Collins CE, Rollo ME. A review of pregnancy iPhone apps assessing their quality, inclusion of behaviour change techniques, and nutrition information. Matern Child Nutr 2019 Jul;15(3):e12768. [doi: 10.1111/mcn.12768] [Medline: 30569549]
57. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
58. Mohsin Khan M, Shah N, Shaikh N, Thabet A, Alrabayah T, Belkhair S. Towards secure and trusted AI in healthcare: a systematic review of emerging innovations and ethical challenges. Int J Med Inform 2025 Mar;195:105780. [doi: 10.1016/j.ijmedinf.2024.105780] [Medline: 39753062]
59. Johnston CA, Moreno JP. Pregnancy: a time of increased focus on health. Am J Lifestyle Med 2014;8(2):93-96. [doi: 10.1177/1559827613514695]

XSL•FO
RenderX

60. Smit AJ, Al-Dhahir I, Schiphof-Godart L, Breeman LD, Evers AW, Joosten KF. Investigating eHealth lifestyle interventions for vulnerable pregnant women: scoping review of facilitators and barriers. J Med Internet Res 2024 Dec 20;26:e54366. [doi: 10.2196/54366] [Medline: 39705692]
61. Mazaheri Habibi MR, Moghbeli F, Langarizadeh M, Fatemi Aghda SA. Mobile health apps for pregnant women usability and quality rating scales: a systematic review. BMC Pregnancy Childbirth 2024 Jan 5;24(1):34. [doi: 10.1186/s12884-023-06206-z] [Medline: 38183035]

Abbreviations:

eHealth: electronic health
OTC: over-the-counter
TIS: Teratology Information Service

Edited by M Sobolev; submitted 16.Oct.2025; peer-reviewed by A Allahham, JP Spencer, M Giner-Soriano, R Nakijoba; revised version received 22.Jan.2026; accepted 24.Feb.2026; published 02.Apr.2026.

Please cite as:
Maas VYF, de Feijter M, Passier ALM, van Tuyl MHM, Kant AC, Conijn M
Development and Launch of a Dutch Mobile App (MediMama) on Over-the-Counter Medication Safety During Pregnancy and Breastfeeding: Development and Usability Study
JMIR Mhealth Uhealth 2026;14:e85948
URL: https://mhealth.jmir.org/2026/1/e85948
doi:10.2196/85948

© Veronique YF Maas, Maud de Feijter, Anneke LM Passier, Miranda HM van Tuyl, Agnes C Kant, Maartje Conijn. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 2.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Acceptability, Feasibility, and Outcome Responsiveness of the Joint Effort Mobile App for Promoting Lower-Risk Cannabis Use Among Young Adults: Pilot Randomized Controlled Trial

José Côté[1,2,3], PhD; Gabrielle Chicoine[3,4], PhD; Patricia Auger[2,3], MSc; Billy Vinette[1,3], PhD; Geneviève Rouleau[3,5,6,7,8], PhD; Marc-André Maheu-Cadotte[1], PhD; M Gabrielle Pagé[2,9], PhD; Judith Lapierre[10,11], PhD; Shalini Lal[2,12,13], PhD; Christine Genest[1,14], PhD; Guillaume Fontaine[15,16,17,18,19], PhD; Sylvie Cossette[1,20], PhD; Jinghui Cheng[21], PhD; Didier Jutras-Aswad[2,22], MD

[1]Faculty of Nursing, Université de Montréal, 2375 Chem. de la Côte-Sainte-Catherine, Montreal, QC, Canada
[10]Faculty of Nursing, Université Laval, Québec, QC, Canada
[11]Center for Research on Social Innovations, Université Laval, Québec, QC, Canada
[12]School of Rehabilitation, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada
[13]Research Centre, Douglas Mental Health University Institute, Montreal, QC, Canada
[14]Centre d'étude sur le trauma, Centre de recherche de l'Institut universitaire de santé mentale de Montréal, Montreal, QC, Canada
[15]Ingram School of Nursing, Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada
[16]Centre for Clinical Epidemiology, Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, QC, Canada
[17]Centre for Nursing Research, Jewish General Hospital, Montreal, QC, Canada
[18]Viral Hepatitis Clinical Research Program, Kirby Institute, University of New South Wales, Sydney, Sydney, Australia
[19]Centre for Implementation Research, Methodological and Implementation Research Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada
[2]Research Centre, Centre Hospitalier de l'Université de Montréal, Montreal, QC, Canada
[20]Montreal Heart Institute, Montreal, QC, Canada
[21]Department of Computer Engineering and Software Engineering, Polytechnique Montréal, Montreal, QC, Canada
[22]Department of Psychiatry and Addictology, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada
[3]Research Chair in Innovative Nursing Practices, Montreal, QC, Canada
[4]Knowledge Translation Program, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, ON, Canada
[5]Department of Nursing, Université du Québec en Outaouais, Gatineau, QC, Canada
[6]Institute for Health System Solutions and Virtual Care, Women's College Hospital, Toronto, ON, Canada
[7]Institut du Savoir Montfort, Ottawa, ON, Canada
[8]Centre de recherche du Centre intégré de santé et de services sociaux de l'Outaouais, Gatineau, QC, Canada
[9]Department of Anesthesiology and Pain Medicine, Faculty of Medecine, Université de Montréal, Montreal, QC, Canada

Corresponding Author:
José Côté, PhD
Faculty of Nursing, Université de Montréal, 2375 Chem. de la Côte-Sainte-Catherine, Montreal, QC, Canada

Abstract

Background: Cannabis use (CU) among young adults continues to be an important public health issue. Interventions to support lower-risk CU during young adulthood can improve health outcomes. Mobile applications constitute a promising mode of service delivery. However, there is a lack of evidence-based apps specifically developed for young adult cannabis users.

Objective: This study aimed to evaluate the acceptability of a novel mobile app intervention (Joint Effort) and to assess the feasibility and outcome responsiveness of the study procedures used.

Methods: A pilot study with a parallel-group randomized trial design was conducted with Canadian-based university students aged 18 - 30 years reporting using cannabis ≥1 day in the past month. Participants were randomly assigned on a 1:1 ratio to either an experimental group (EG) involving the use of the Joint Effort mobile app or to a control group (CG) involving a web-based brief normative feedback message. The Joint Effort mobile app was designed to support CU self-management. This theory-informed behavior change intervention aims to reinforce the use of protective behavioral strategies by targeting intention, attitude, social norms, and self-efficacy. The app's acceptability was assessed via uptake, engagement, and appreciation. The feasibility of study procedures was assessed via recruitment time, recruitment rate, and attrition rate. Outcome responsiveness was informed by participant-reported outcomes: CU frequency, intention to take action on CU, protective behavioral strategies use, severity of

XSL•FO
RenderX

dependence, and psychological distress. All data were collected using a web-based survey at baseline, one-month (T1), and 2-month (T2) postbaseline. Descriptive analyses were carried out on all outcomes.

Results: The recruitment period lasted 124 days, and the recruitment rate was 56% (99/178). The final dataset analyzed included 80 participants (39 in EG and 41 in CG). Mean age was 23.4 (SD 2.6) years, and 66% (53/80) self-identified as women. Study attrition was 18% (14/80). User uptake of the Joint Effort app (ie, proportion of participants in the EG who downloaded the app) was estimated at 59% (23/39), and the average time spent on it per participant was 8.2 minutes (SD 7.3; median 7.5, IQR 5.7). The app obtained a mean total score on the User Engagement Scale-Short Form of 3.8/5 (SD 0.5) and a mean app quality total score of 4.2/5 (SD 0.5) on the end user version of the Mobile App Rating Scale. The proportion of participants who reported daily CU in the past month decreased from 13% (5/39) at baseline to 4% (1/24) at T2 in the EG and from 7% (3/41) to 6% (2/36) in the CG.

Conclusions: Joint Effort appears to be a promising, acceptable, and scalable mobile app to help young adult cannabis users who wish to better manage their CU. Findings should inform future randomized controlled trials to assess the efficacy of this mobile-based intervention for cannabis users.

Trial Registration: ClinicalTrials.gov NCT05099016; https://clinicaltrials.gov/study/NCT05099016

(JMIR Mhealth Uhealth 2026;14:e71957) doi:10.2196/71957

KEYWORDS

mobile health; mHealth; app; digital health applications; cannabis; behavior change; young adults; students; harm reduction

Introduction

Background

Cannabis is one of the most commonly used substances for nonmedical purposes throughout the world. The prevalence of past-year cannabis use (CU) in the global community 15‐64 years of age in 2018 was estimated at 3.8%, which represented nearly 200 million people [1]. Attitudes toward CU have evolved substantially over the past 25 years [2]. The legalization of medical and recreational CU has expanded in Canada and the United States, and the legal status of CU has been shifting in other countries as well [3]. In the last decade, the consumption of cannabis products has increased considerably, particularly among young adults, typically 18‐25 years old [4-6]. In this population group, the prevalence of past-year CU was found to be 25% or higher in high-use regions such as North America, Oceania, and West Africa [1]. Moreover, and important for how it might impact potential life-course outcomes, the median age of onset of regular CU is 18‐19 years [7], and CU level and frequency begin to increase in very early adulthood and peak in middle age [8,9].

CU is associated with a wide range of health harms, including risk for CU disorder [10,11], short-term acute risks (eg, cannabis-impaired driving and related injuries/death, acute intoxication, impaired cognitive functioning), and other negative social consequences (eg, poor schooling or work performance, social isolation, family violence, stigmatization, poor living conditions) [12-14]. Importantly, CU has been linked to long-term adverse health outcomes such as cardiovascular problems, brain development issues, lung tissue injury, and mental health problems [15-18]. In the context of increased availability of high-THC products, country- or state-level legalization reforms allowing recreational CU, and public opinion to the effect that cannabis is a low-risk substance [6,19,20], there is a need for interventions to promote lower-risk CU and to prevent escalation of CU-related consequences among young adults.

Across the intervention continuum, CU interventions range from public awareness campaigns and prevention practices to treatment programs for CU disorder. Among these, lower-risk CU interventions [10] constitute a wide range of public health prevention or intervention tools inspired by harm reduction, a framework that places substance use on a spectrum from total abstinence to continued heavy use [21]. Generally, lower-risk CU interventions are based on health behavior change theories [22-24] and existing harm-reduction prevention approaches (eg, low-risk drinking guidelines, guidance for safe drug use practices) [25,26].

In recent years, particularly, there has been growing interest in the development of digital interventions for the promotion of lower-risk CU [27-30]. Digital modalities not only render interventions more accessible and cost-effective, but they also offer innovative opportunities for delivering personalized and real-time adaptive interventions that meet the needs and preferences of young adults [31-33]. An important body of literature documents the development, adaptation, and evaluation of digital interventions for lower-risk CU among young adults designed to be accessed online (ie, web- or internet-based) or via a mobile device (eg, cell phone, tablet, smartwatch) [28,29,34-40]. Altogether, the scientific evidence on the effects of digital interventions for lower-risk CU on health outcomes has been mixed. Specifically, while several studies have shown that digital interventions can reduce CU among young adults [36-42], other studies have found no effects on use [35] and cannabis-related consequences [34]. Furthermore, a scoping review by Sedrati et al [39] and a recent systematic review by Côté et al [34] both concluded that, despite considerable advances in digital interventions for lower-risk CU among young adults, most of the interventions developed and tested to date are web-based programs. These asynchronous digital tools typically use websites or online platforms, accessed mainly via computers, and consist primarily of didactic, static information (eg, facts, psychoeducational material) [40]. This means that content is neither individualized to the needs and characteristics of intended end users nor tailored in real time. Therefore, more

XSL•FO
RenderX

research is needed to develop and test the acceptability and effectiveness of digital interventions using next-generation mobile health (mHealth) tools (eg, mobile apps, social media) that, undoubtedly, would be more consistent with the technological habits of today's young adults.

Related to these concerns, digital CU interventions for young adults also pose unique challenges in terms of implementation and utilization. For one, adequate engagement is crucial for digital interventions to be successful [43,44]. However, research has shown that engagement tends to be low and that attrition is common [33,35]. This concern is particularly salient with respect to the subpopulation of young adults who do not perceive their CU as problematic and, consequently, have lower readiness to change. In addition, existing digital CU interventions were developed primarily for individuals with a CU disorder (5.1% of people aged 12 years or older in the United States) [45] or mental health needs [31,38,46]. This excludes a larger proportion of young adults who do not experience severe problems from their CU [37]. For example, recent studies have demonstrated the benefits of digital interventions aimed at supporting young adults after substance-use treatment programs [47-49]. However, these interventions are not tailored to individuals with lower severity substance-use who are not necessarily seeking to change their behavior or to prevent relapse [33]. Finally, many digital tools developed specifically for young adult cannabis users include therapist support [50-56]. From a public health perspective, this reduces scalability and may be unnecessary for people with lower CU severity.

In 2018, recognizing the need for a publicly available evidence-based app to promote lower-risk CU tailored to the needs of young adults [57], we developed a mobile app named Joint Effort. The detailed rationale for developing this new intervention is described elsewhere [58]. This mobile app aims to support CU self-management and reinforce protective behavioral strategies (PBS) use among young adult cannabis users. The intervention applies behavior change principles [59,60] that have previously demonstrated their efficacy in trials of interventions targeting risky behaviors [21,61,62]. Consistent with the current best practice in developing mHealth apps [63,64], our team's prior work used an iterative intervention mapping and co-design approach [65-67] in a series of formative studies [58,68,69] to develop, inform, and refine Joint Effort. Additional formative work was required to determine the app's acceptability and feasibility before its efficacy and effectiveness in improving health outcomes among young adult cannabis users could be tested. Against this backdrop, we conducted a pilot parallel-arm randomized controlled trial (RCT) to assess the acceptability of the Joint Effort app in terms of user uptake, engagement, and appreciation in a sample of young adult university students. We also assessed the feasibility of our study procedures and the responsiveness of outcome measures.

Aim and Objectives

The objectives of this study were threefold:

1. To assess the acceptability of the Joint Effort app in terms of user uptake, engagement, and appreciation.
2. To assess the feasibility of study procedures and document the research process throughout the study period.
3. To assess outcome responsiveness in participant-reported outcomes, including CU, intention to take action on CU, PBS use, severity of dependence, and psychological distress.

Methods

Study Design

The study design was a pilot RCT [70] where a sample of young adult university students who reported CU in the past month was split into two parallel arms: (1) an experimental group (EG) that was given access to the Joint Effort mobile app; and (2) a control group (CG) that received one brief normative web-based feedback message regarding their CU and had access to publicly available information on lower-risk CU guidelines.

The study was registered in the ClinicalTrials.gov Protocol Registration and Results System (NCT05099016). The study reporting followed the CONSORT extension for Pilot and Feasibility Trials [71] checklist (Checklist 1) in reporting the trial.

Setting and Sample

The target population for this study consisted of young adults studying at the Université de Montréal, a world-class, predominantly French-language university with approximately 67,000 students, located in a metropolitan region of Eastern Canada. We used a recruitment strategy centered on social media ads and posters on university campuses to obtain a broad and diverse sample of university student cannabis users. Individuals were eligible for this study if they met four criteria: (1) 18 to 30 years of age; (2) enrolled in an undergraduate or graduate program at the Université de Montréal; (3) self-reported at least one episode of past-month CU; and (4) had an iPhone to download the app to (Joint Effort was available only as an iOS app at the time of study).

Teare et al [72] recommend that at least 70 participants (35 per group) are needed in a pilot RCT to ensure a statistically precise measurement of each variable. To compensate for potential attrition at follow-up (ie, we expected up to 35% attrition at the 2-month postbaseline assessment) [34,35], we sought to recruit a total of 110 participants in order to end up with a final sample size of at least 70 completers.

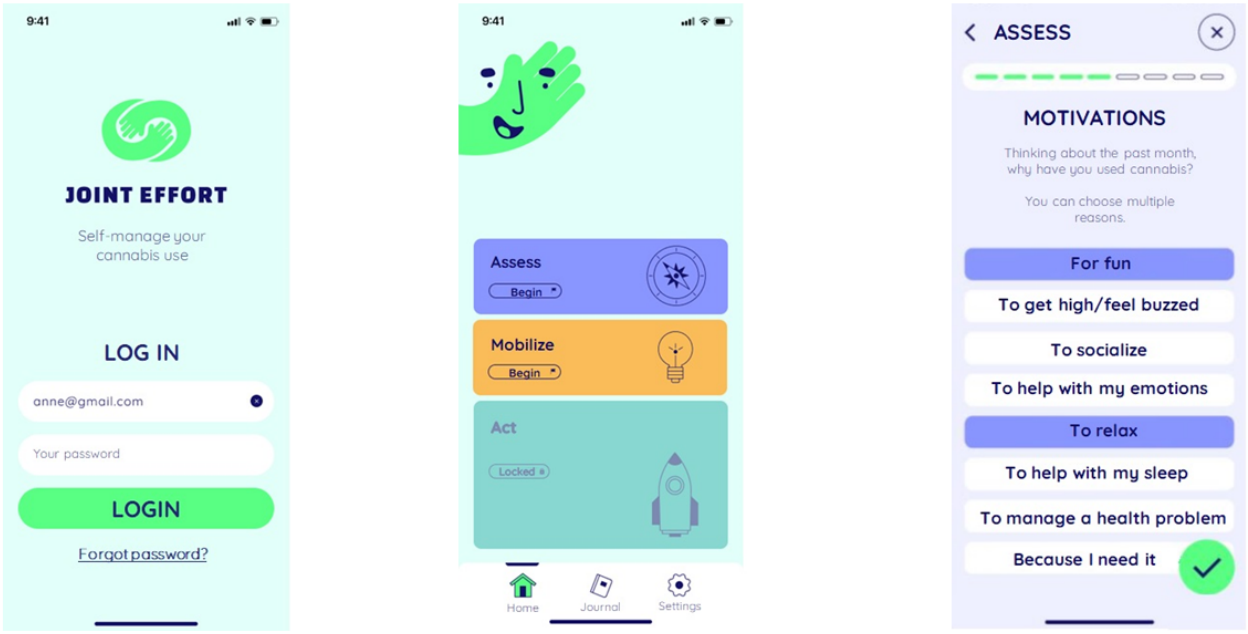
Intervention Characteristics: A Mobile App for Lower-Risk CU

The Joint Effort mobile app was designed to promote CU self-management and reinforce PBS use among young adult cannabis users. Developed based on Ajzen's Theory of Planned Behavior [59], Joint Effort focuses on 4 core determinants, namely, intention, attitude, social norms, and self-efficacy. To address these, various evidence-based, theory-driven intervention strategies and methods are used, including personalized feedback, persuasive communication, positive reinforcement, self-observation, and activation of intention. A detailed account of how the principles of the Theory of Planned Behavior were integrated in the app is presented elsewhere [58]. The intervention content is organized into five sections: (1) "Assess" to allow users to gain a higher awareness of their CU, (2) "Mobilize" to support their decision-making process, (3)

XSL•FO
RenderX

"Act" to guide them in establishing an action plan, (4) "Strengthen" to consolidate their change in behavior, and (5) "Observe" to monitor their CU. It addresses various topics such as CU prevalence, potential consequences, PBS, benefits of change, and possible difficult situations. The information is provided through general information messages, questionnaires (checkbox answers, reflection questions), and personalized feedback. The app also includes an optional consumption diary where users can enter information regarding their CU and customize certain features of their logbook for personalized monitoring (eg, frequency of daily entries, format, and concentration of cannabis used, personal notes). The intervention was designed as a self-managed exercise in prevention, given that it proposes simple strategies easy to apply and adopt without the help of outside resources. Joint Effort was available only in French and as an iOS mobile app (ie, for iPhones running on iOS version 13 or 14) at the time of the study. A detailed description of the mobile app intervention and its content, in accordance with the Template for Intervention Description and Replication (TIDieR) [73] Checklist, is available in Checklist 2 and in a publication elsewhere [58]. Figure 1 presents screenshots from the Joint Effort mobile app showing its general look and feel, as well as how features are organized within the app.

Figure 1. Screenshots from the Joint Effort mobile app.



Study Groups

Participants randomly assigned to the EG were invited to download and use the Joint Effort mobile app on their iPhones. Following randomization, members of the research team created a user account for each of them based on their chosen username/email collected during study registration. Participants received an automatically personalized email containing an App Store link to download the mobile app. Access to Joint Effort was unlimited in terms of frequency and duration of use over 8 weeks once their profile had been created by the research team. To optimize engagement with the intervention, optional pop-up notifications were embedded within the mobile app. Two types of notifications were sent: (1) to encourage users to continue with their action plan (scheduled 3, 5, and 7 days after completion of the "Act" section) and (2) to remind users about the app (triggered 9 days after their last login).

Participants randomly assigned to the CG received one brief normative web-based feedback message regarding their CU based on their past-month frequency of use reported during the baseline assessment. They were also sent links to access standardized information about lower-risk CU guidelines on reliable French public websites tailored for the Quebec context (eg, public health agencies, Quebec ministries of health). A single reminder inviting them to visit (or revisit) the recommended websites was sent via email 7 days after baseline.

Randomization and Blinding

As recommended in the CONSORT guidelines [71,74], participants were randomized only after having provided consent and completed a web-based survey at baseline. Participants had an equal chance of being allocated to the EG or the CG (1:1 ratio). The allocation sequence was generated through a permuted block randomization list to ensure that the 2 groups were as even as possible during data collection. The allocation process (concealment and implementation) was entirely computerized. The sequence was programmed within the study website, and participants were automatically informed of the intervention they were assigned to by email. The study was partially single-blinded in that participants in the 2 study groups knew the intervention they were working with, but were unaware whether it was the experimental condition or not. The research

XSL•FO
RenderX

coordinators were not blinded to assignments since they had to create unique profiles for each participant and to monitor follow-up.

Measures

Acceptability was defined as the "suitability of the intervention from the perspective of the population of interest" [75], that is, young adult cannabis users. Acceptability of the Joint Effort mobile app was assessed via 3 distinct yet interrelated constructs, namely, uptake, engagement, and appreciation, only among participants from the EG.

Uptake was defined as the act of downloading the Joint Effort mobile app via the Apple iOS App Store [76]. Uptake rate was measured as a percentage calculated by comparing the number of EG participants who downloaded Joint Effort against the total number of participants randomized to the EG.

Engagement was defined as a quality of the user experience [77] characterized by a substantial exposure to the intervention content among uptakers [78]. Engagement was assessed among participants randomized to the EG using a hybrid approach [79] that included both objective and subjective measures: (1) quantity of exposure, that is, frequency and duration of use of Joint Effort, was collected automatically when users logged in to the mobile app (data extracted by the research team at the end of the 2-month intervention period); and (2) self-report measure of user engagement using the User Engagement Scale-Short Form (UES-SF) [80] after one month of using the app to minimize memory bias. Used in various digital domains, the UES-SF is a self-completed questionnaire that has demonstrated adequate construct validity and good internal consistency [80-82]. It consists of 12 items meant to measure 4 dimensions of engagement across 4 subscales labeled Aesthetic appeal, Focused attention, Perceived usability, and Reward factor (possible total score range: 12 - 60). In this study, a French adaptation of the original English version of the UES-SF was used [83]. This French translation has shown satisfying preliminary psychometric properties in a sample of 57 Canadian-based university students (McDonalds' omega [ω] coefficients of the French UES-SF subscales range from 0.8 to 0.9) [83].

Participant appreciation of Joint Effort was assessed with the end user version of the Mobile App Rating Scale (uMARS) [84]. Considering the detailed nature of the uMARS items, this questionnaire was completed 2 weeks after randomization to minimize memory bias. The uMARS is a reliable method to assess the quality of mobile-app tools that has demonstrated excellent internal consistency with a Cronbach α of 0.90 for the full scale, high internal consistency for the subscales, and good levels of interclass correlation coefficients at follow-up [84]. The uMARS comprises 20 items across 4 objective quality subscales—Engagement, Functionality, Aesthetic, and Information—and a subjective quality subscale, all rated on a 5-point Likert scale (possible total score range: 20 - 100). Stoyanov et al [84] suggested determining the app quality mean score by computing the mean score for each objective quality subscale and averaging the four. The uMARS was translated into French by our research team following a rigorous translation process, similar to the one described in Côté et al [68,69].

Feasibility was defined as success in executing the study design and research procedures as planned [75]. In this study, it was assessed based on 3 parameters: recruitment feasibility, study attrition, and research procedures.

Recruitment feasibility was assessed by measuring recruitment time and recruitment rate. The recruitment time was calculated as the number of days from the time the study was advertised to the time our target sample size of 110 randomized participants was reached. The recruitment rate was measured as a percentage calculated by comparing the number of participants randomized to the study groups against the total number of individuals deemed eligible to participate [85].

The study attrition rate was measured as a percentage calculated by comparing the number of participants enrolled in the study who completed only the baseline assessment (ie, noncompleters) against the number of those enrolled in the study who completed the baseline assessment and at least one follow-up survey (ie, completers).

We also sought to determine the feasibility of the research procedures of a study conducted entirely via the web. To this end, we documented any issue that arose during recruitment or data collection (eg, data security and privacy, filtering fraudulent applications and information), as well as app-level problems (eg, technical issues with the operating system, challenges with software updates, and screen functionality).

The outcome responsiveness was assessed based on the following patient-reported outcomes: CU, intention to take action on CU, PBS use, severity of cannabis dependence, and psychological distress.

CU was measured by frequency of use via the following question adapted from the French version of the Canadian Community Health Survey [86]: "How often did you use cannabis in the past month?." The extent to which participants experienced days of CU in the past month was assessed on the following 6-category rating scale: never; one day in the past month; 2-3 days in the past month; 1-2 days per week in the past month; 3-4 days per week in the past month; 5-6 days per week in the past month; and every day.

Intention to take action on CU referred to "readiness to engage in a behavior" [87] and was assessed via a 3-item French-language questionnaire developed for the purpose of this study, based on Ajzen's guidelines for measuring Theory of Planned Behavior variables [88]. The questionnaire consisted of the following three items: (1) "I intend to take action on my CU in the coming month"; (2) "In the coming month, the chances that I will take action on my CU are…"; and (3) "In the coming month, I will take action on my cannabis use." These were rated on a 7-point Likert scale ranging, respectively, from "strongly disagree" to "strongly agree," "very low" to "very high," and "very unlikely" to "very likely." The final measure of intention to change was arrived at by summing the numbers from 1 to 7 corresponding to answers on the 3 items (possible score range: 3 to 21). The lower the total score, the weaker the intention to take action on CU in the coming month. In a previous study among young cannabis users, this scale showed good internal consistency with a Cronbach α of 0.87 [89].

XSL•FO
RenderX

PBS use was assessed via the short form of the 17-item Protective Behavioral Strategies for Marijuana (PBSM-17) Scale [60], a reliable instrument to identify and measure lower-risk CU strategies used by individuals before, during, and after CU. Methodological work was previously undertaken by our research team to culturally adapt and translate the PBSM-17 into French and evaluate its psychometric properties, as well as those of the original English version, in a sample of 211 bilingual, Canadian-based young adult university students who reported past-month CU [68,69]. Both the French and English versions demonstrated satisfactory psychometric properties, with good internal consistency (Cronbach α=0.88 and 0.91, for the French and English versions, respectively). Each item of the PBSM-17 concerns a particular PBS. Respondents are invited to rate the frequency of use for each strategy on a 6-point Likert scale ranging from never (value of 1) to always (value of 6). The PBSM-17 is scored by summing the values of the responses (possible score range: 17 to 102). The lower the total score, the lesser the use of PBSs.

Severity of cannabis dependence was assessed via the Severity of Dependence Scale (SDS) [90], a self-report instrument designed to measure the degree of dependence experienced by users of different types of drugs, including cannabis. A French language version of the SDS was used in this study [91]. The instrument consists of 5 items covering different psychological aspects of dependence: feeling of impaired control, anxiety, preoccupations, wish to stop, and compulsive use. The response choices for items 1 to 4 are scored as follows: 0=never/almost never, 1=sometimes, 2=often, and 3=always/nearly always. Those for item 5 are scored as follows: 0=not difficult, 1=quite difficult, 2=very difficult, and 3=impossible. The possible score ranges from 0 to 15. The cannabis version of the SDS has demonstrated satisfactory psychometric properties in samples of adults and adolescents from the general population as well as in clinical samples (eg, patients with a schizophrenia spectrum disorder) [92-96] and, more recently, it proved to have good internal consistency (Cronbach α=0.70) in a sample of 577 young adult (18 - 30 y old) frequent and heavy cannabis users [97]. Previous studies have set empirically derived optimal cut-off scores for cannabis dependence at 2 to 4 or greater [92,96,98]. In this study, the cut-off score was set at 4 or greater.

Psychological distress was assessed via the Kessler Psychological Distress Scale–6 items (K6) [99], a psychometrically robust measure of psychological distress in adult populations used in epidemiological surveys globally, including the World Health Organization World Mental Health Survey [100]. The internal consistency of the K6 has been found to be adequate in general population samples (Cronbach α=0.8 to 0.9) [101-104], and one study [105] found the K6 to be valid and reliable in an epidemiological sample of Canadian youth (2010 respondents 15 to 19 years old), demonstrating a high internal consistency (Cronbach α=0.9). The K6 is a nonspecific measure of psychological distress consisting of 6 questions regarding whether the respondent had felt nervous, hopeless, restless, or fidgety in the past month, had felt so depressed that nothing could cheer them up, had felt that everything was an effort, and had felt worthless. Each item is rated on a 5-point Likert scale ranging from “none of the time” (value of 0) to “all of the time” (value of 4) (possible score range: 0 - 24). A total score of 0 to 12 indicates that severe mental illness is unlikely; a total score of 13 or greater indicates that severe mental illness is likely [106]. The French translation of the scale was used [107].

Study Procedures

The study was conducted entirely via web over a 6-month period (November 19, 2021, to May 20, 2022). It was hosted on a secure website developed for the purposes of this study. Recruitment was carried out via social media ads (ie, public and private Université de Montréal student Facebook groups) and posters on university campuses. Interested participants were directed to the study website to find out what the study was about. Once there, they were then asked to complete a web-based screening assessment (ie, 4 questions to confirm eligibility) and a consent form. Eligible participants were invited to register by choosing a username and providing contact information (ie, email address). To complete the registration process, participants had to click on a verification link sent by email, which allowed them to activate their profile and, subsequently, receive a unique link generated automatically to be directed to the web-based baseline survey to be completed.

The web-based baseline survey included a brief sociodemographic questionnaire specifically developed for the purposes of this study. It covered various individual-level parameters, including gender, age, ethnicity, educational level, and use of cannabis, alcohol, or other drugs (eg, cocaine, amphetamines, hallucinogens). The survey also included a question about CU frequency in the past month and instruments to assess outcome responsiveness (ie, intention to take action on CU, PBSM-17, SDS, and K6).

The same web-based survey without the sociodemographic questionnaire was administered to all participants at 2 other time points: 1 month (T1) and 2 months (T2) postbaseline. EG participants, who were invited to download the Joint Effort app and use it for 2 months, were also asked to complete questionnaires regarding acceptability variables, including the uMARS at 2 weeks postbaseline and the UES-SF scales at T1. A question asking participants to confirm their date of birth was included in the survey at each time point for verification purposes. Analytics data on participant use of Joint Effort, including quantity of exposure and usage patterns, were collected automatically by a web-based platform developed specifically for this study, hosted on a secure server. A research assistant could be contacted during the study period by email to answer questions and help participants with issues regarding the app and web-based surveys.

All data collected during the study period were stored digitally at the study website until downloaded onto a secure virtual server at the research center for analysis. Given that fraudulent (or inconsistent) data are a well-known problem in web-based survey studies [108,109], an in-depth examination of the data collected was carried out to assess their reliability and accuracy, and to double-check the identity of the participants before performing any statistical analysis. This involved verifying that the dates of birth provided by participants were consistent across time points and confirming that the email addresses provided

XSL•FO
RenderX

by participants were valid. Participants with inconsistent data or contact information were considered "fraudulent participants" [110] and, therefore, not included in the final dataset analyzed.

Data Analysis

All statistical analyses were conducted in R (version 4.2.3; R Foundation for Statistical Computing). Analyses were carried out by an independent, experienced biostatistician not blinded to participant group assignment. As per protocol, the biostatistician cleaned the dataset to enhance data consistency and quality. No interim analysis was performed, and, as recommended for pilot study designs [111], no hypothesis testing (ie, inferential analysis) was carried out. Descriptive analyses of sample characteristics included computing means and SDs for continuous variables and the frequency of distribution for categorical variables. Descriptive analyses were carried out on the acceptability and feasibility variables as well. Finally, descriptive statistics were calculated to explore outcome responsiveness, and the same analytical approach was used for all outcomes.

Descriptive analyses, including means and SDs for continuous variables and frequency of distribution for categorical variables, were conducted for sample characteristics, acceptability/feasibility variables, and outcome responsiveness.

Ethical Considerations

Ethical approval was obtained from the Research Ethics Committee of the Centre Hospitalier de l'Université de Montréal (CHUM; #21.196) and from the Comité d'éthique de la recherche en sciences et en santé of the Université de Montréal (CERSES; #2021 - 1223). Informed consent was obtained from all participants prior to study activities. Participants received financial compensation for their time and effort in the form of C $15 and C $25 (US $11 and US $18) Amazon.ca gift certificates at T1 and T2, respectively. All collected data have been deidentified, and results are reported at the sample level.

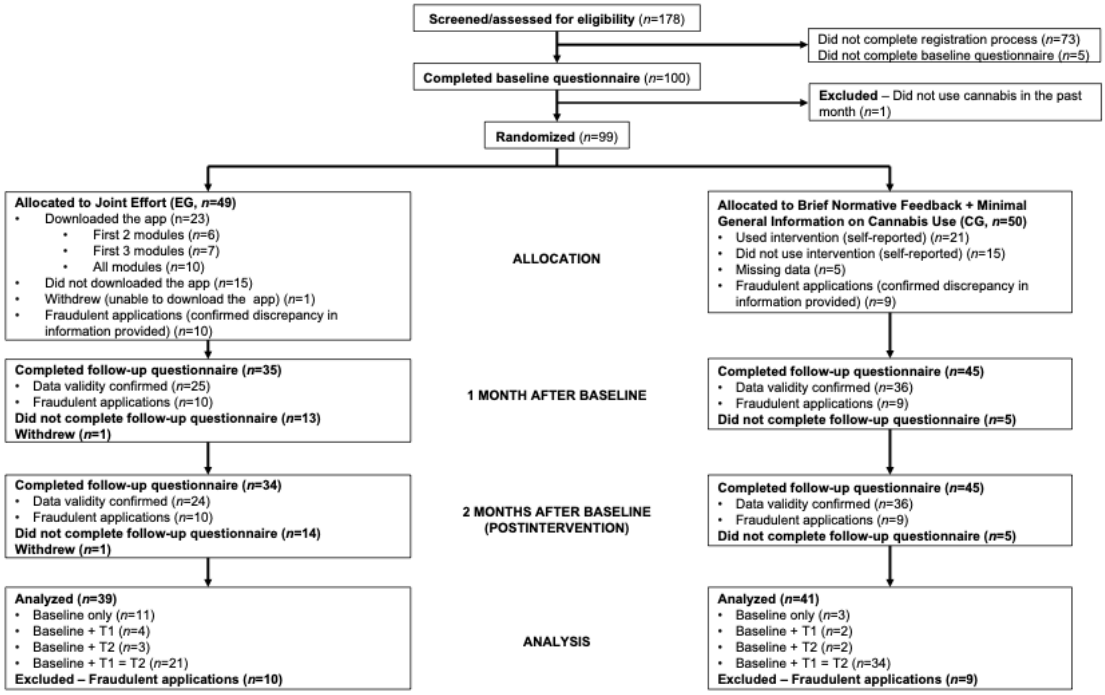
Results

Feasibility of Study Procedures

Recruitment Feasibility

As a result of our recruitment strategies, 178 people were deemed eligible after completing the screening survey (Figure 2 shows the study flow diagram). Of these, 73 were excluded for not completing the registration process, and 5 others for not completing the baseline survey. A total of 100 participants completed the baseline survey, and one participant was excluded afterward (ie, this participant was eligible according to the screening survey, but reported no CU in the baseline survey). The remaining 99 were randomized to the EG and CG, yielding a 56% (99/178) recruitment rate. Of these, 19 were considered fraudulent participants and eliminated for providing discrepant dates of birth or contact information on the registration form and the baseline or follow-up survey (10/49 participants in the EG and 9/50 participants in the CG). Consequently, there were 80 participants (39 EG and 41 CG) in the final dataset analyzed.

Figure 2. Study flow diagram with a pilot randomized controlled trial design based on Eldridge et al [70]. CG: control group; EG: experimental group.



As for recruitment time, it took 124 days total from the day the study began to be advertised on social media and on campus to the day our target sample size of 110 randomized participants was reached (from November 19, 2021, to March 22, 2022).

Study Attrition

Of the 80 participants, 61 (76%) completed the assessment at one month postbaseline (T1) and 60 (75%) completed the assessment at 2 months postbaseline (T2). In all, 66 participants were considered completers for having completed the

XSL•FO
RenderX

assessments at T0 and T1, T0 and T2, or all 3 assessments (ie, T0, T1, and T2). In the CG, 38 of 41 participants (93%) were considered completers, compared with 28 of 39 (72%) in the EG. The overall attrition rate was 18%, as 14 of the 80 participants (CG: 3/41; EG: 11/39) were considered noncompleters.

Sample Characteristics

The characteristics of participants at baseline are presented in Table 1. Age ranged from 19 to 29 years (mean 23.4, SD 2.6). Two-thirds identified as women (53/80, 66%), nearly 3-quarters were born in Canada (58/80, 73%), and 3-quarters were undergraduate students (60/80, 75%). Just under half were classified as occasional cannabis users (34/80, 43% CU 1-3 d/mo) and one out of ten as daily cannabis users (8/80, 10%). More than half of the participants reported using alcohol regularly (44/80, 55% alcohol use 1-4 d/wk), and the vast majority did not use other drugs (67/80, 84%).

XSL•FO
RenderX

Table . Participant characteristics at baseline (T0) for the overall sample (n=80), by study group and by study completion pattern.

Characteristic	Total sample (n=80)	Study groups		Study completion pattern	
		EG[a] (n=39)	CG[b] (n=41)	Completers (n=66)	Noncompleters (n=14)
Age (years), mean (SD)	23.4 (2.6)	23.6 (2.5)	23.3 (2.7)	23.5 (2.6)	22.7 (2.4)
Gender, n (%)					
Male	26 (33)	9 (23)	17 (41)	22 (33)	4 (29)
Female	53 (66)	29 (74)	24 (59)	43 (65)	10 (71)
Did not want to answer	1 (1)	1 (3)	0 (0)	1 (2)	0 (0)
Born in Canada, n (%)					
Yes	58 (73)	28 (72)	30 (73)	48 (73)	10 (71)
No	20 (25)	10 (26)	10 (24)	16 (24)	4 (29)
Did not want to answer	2 (3)	1 (3)	1 (2)	2 (3)	0 (0)
University program, n (%)					
Undergraduate	60 (75)	28 (72)	32 (78)	47 (71)	13 (93)
Master's	12 (15)	6 (15)	6 (15)	12 (18)	0 (0)
Doctorate	7 (9)	5 (13)	2 (5)	6 (9)	1 (7)
Did not want to answer	1 (1)	0 (0)	1 (2)	1 (2)	0 (0)
Recruitment strategy[c], n (%)					
Facebook ad	60 (75)	30 (77)	30 (73)	51 (72)	9 (64)
Posters on campus	8 (10)	4 (10)	4 (10)	6 (9)	2 (14)
Other[d]	15 (19)	6 (16)	9 (19)	12 (17)	3 (21)
Did not want to answer	2 (3)	1 (3)	1 (2)	2 (3)	0 (0)
Past-month cannabis use[e], n (%)					
Occasional[f]	34 (43)	17 (44)	17 (41)	28 (42)	6 (43)
Regular[g]	38 (48)	17 (44)	21 (51)	32 (49)	6 (43)
Every day	8 (10)	5 (13)	3 (7)	6 (9)	2 (14)
Past-month alcohol use[e], n (%)					
Never	12 (15)	7 (18)	5 (12)	10 (15)	2 (14)
Occasional[f]	23 (29)	10 (26)	13 (32)	19 (29)	4 (29)
Regular[g]	44 (55)	21 (54)	23 (56)	36 (55)	8 (57)
Every day	1 (1)	1 (3)	0 (0)	1 (2)	0 (0)
Past-month drug use[e,h], n (%)					
Never	67 (84)	33 (85)	34 (83)	57 (86)	10 (71)
Occasional[f]	13 (16)	6 (15)	7 (17)	9 (14)	4 (29)
Regular[g]	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

XSL•FO
RenderX

Characteristic	Total sample (n=80)	Study groups		Study completion pattern	
		EG[a] (n=39)	CG[b] (n=41)	Completers (n=66)	Noncompleters (n=14)
Every day	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

[a]EG: experimental group.

[b]CG: control group.

[c]More than one recruitment strategy possible.

[d]Other included: receiving a personalized email message or a friend referral.

[e]Frequency of use was determined by the number of days on which cannabis, alcohol, or other drugs were used in the past month.

[f]Occasional: 1 - 3 days per month.

[g]Regular: 1 - 6 days per week.

[h]Included but was not limited to cocaine, amphetamines, and hallucinogens.

Joint Effort Acceptability

Uptake

Of the 39 participants randomized to the EG, 23 downloaded the Joint Effort mobile app to their iPhone for an uptake rate of 59%.

Engagement

Of the confirmed uptakers, 43% (10/23) viewed all the sections of the mobile app, while 26% (6/23) viewed only the first two (ie, “Assess” and “Mobilize”), and 30% (7/23) viewed only the first three (ie, “Assess,” “Mobilize,” and “Act”). In terms of quantity of exposure, the mean number of screens viewed per participant was 32.4 (Min-Max=15-89; SD 18.0). Regarding duration of exposure, participants spent an average of 8.2 minutes (Min-Max=53 s-33.5 min; SD 7.3 min) and a median of 7.5 minutes (IQR 5.7) using the app. The total period of use ranged from 1 to 59 days (mean 19.5; SD 21.9) with a 5-day median (IQR 36).

Regarding subjective engagement, the mean total score on the UES-SF was 3.8 (SD 0.46) at T1. Descriptive statistics for the UES-SF subscales and items are given in Table 2. Among the 4 UES-SF subscales, Aesthetic Appeal obtained the highest score (mean 4.3; SD 0.7), followed by Perceived usability (mean 4.2; SD 0.9), while Focused attention obtained the lowest score (mean 2.8; SD 0.9).

Table . Participant subjective engagement scores on the User Engagement Scale-Short Form (UES-SF), one-month postbaseline (T1) (n=18).

Subscale and item	Mean (SD)[a]
Focused attention	2.8 (0.9)
I lost myself in this experience.	2.2 (1.1)
The time I spent using Joint Effort just slipped away.	3.2 (0.9)
I was absorbed in this experience.	2.9 (0.9)
Perceived usability	4.2 (0.9)[b]
I felt frustrated while using this Joint Effort mobile app.	4.0 (1.1)[b]
I found this Joint Effort mobile app confusing to use.	4.1 (0.9)[b]
Using this Joint Effort mobile app was taxing.	4.4 (0.8)[b]
Aesthetic Appeal	4.3 (0.7)
This Joint Effort mobile app was attractive.	4.1 (0.9)
This Joint Effort mobile app was esthetically appealing.	4.4 (0.6)
This Joint Effort mobile app appealed to my senses.	4.3 (0.8)
Reward factor	4.0 (0.7)
Using Joint Effort mobile app was worthwhile.	4.1 (0.8)
My experience was rewarding.	3.6 (0.8)
I felt interested in this experience.	4.3 (0.6)
Overall score	3.8 (0.5)

[a]Possible score range of 1 to 5 (1=strongly disagree; 5=strongly agree); the higher the score, the higher the subjective engagement.

[b]Reverse scored. (1=strongly agree; 5=strongly disagree); the higher the score, the higher the level of perceived usability.

XSL•FO
RenderX

Appreciation of Joint Effort

Sixteen participants in the EG provided data regarding their appreciation of Joint Effort 2 weeks after randomization. The app obtained a mean app quality total score of 4.2 out of 5 (SD 0.5). Among the 5 uMARS subscales, Aesthetics (mean 4.4, SD 0.6), Information (mean 4.3, SD 0.5), and Functionality (mean 4.3, SD 0.5) scored high, while Engagement (mean 3.8, SD 0.6) and App subjective quality (mean 3.3, SD 0.6) scored lower (Multimedia Appendix 1). The Quantity of information item (”Is the information within the app comprehensive but concise?”) included in the Information subscale was the item that scored highest (mean 4.6, SD 0.6) on the 5-point Likert scale, while the Willingness to pay item (“Would you pay for this app?”) part of the App subjective quality scale scored lowest (mean 1.9, SD 0.8).

Outcome Responsiveness

Results for CU frequency, intention to take action on CU, PBS use, severity of cannabis dependence, and psychological distress among participants in the EG and the CG at the 3 measurement time points are presented in Table 3.

Table . Participant outcomes at baseline (T0) and follow-up assessments (T1, T2), by study group (n=80).

Variable	Baseline (T0)		1 month (T1)		2 months (T2)	
	CG[a] (n=41)	EG[b] (n=39)	CG (n=36)	EG (n=25)	CG (n=36)	EG (n=24)
Cannabis use (past month), n (%)						
Never	0 (0)	0 (0)	4 (11)	3 (12)	6 (17)	4 (17)
Occasional (1 - 3 days/month)	17 (41)	17 (44)	11 (31)	7 (28)	16 (44)	8 (33)
Regular (1 - 6 days/week)	21 (51)	17 (44)	19 (53)	13 (52)	12 (33)	11 (46)
Everyday	3 (7)	5 (13)	2 (6)	2 (8)	2 (6)	1 (4)
Intention to take action[c], mean (SD)	12.5 (4.3)	12.4 (5.2)	13.6 (5.4)	14.3 (4.6)	13.4 (6.0)	14.3 (5.3)
PBS-17[d], mean (SD)	72.9 (12.5)	73.7 (15.1)	76.4 (12)	78.8 (15.8)	80.2 (12.3)	80.0 (13.0)
SDS[e,f], n (%)						
No dependence	26 (65)	21 (54)	N/A[g]	N/A[g]	24 (71)	15 (68)
Dependence[h]	14 (35)	18 (46)	N/A[g]	N/A[g]	10 (29)	7 (32)
Psychological distress[i,j], n (%)						
Low/moderate[k]	28 (68)	28 (76)	N/A[g]	N/A[g]	24 (69)	20 (83)
Severe[l]	13 (32)	9 (24)	N/A[g]	N/A[g]	11 (31)	4 (17)

[a]CG: control group.

[b]EG: experimental group.

[c]Intention to take action on cannabis use; possible score range 3 - 21.

[d]PBSM-17: 17-item Protective Behavioral Strategies for Marijuana Scale; possible score range 17 - 102 (summed score).

[e]SDS: Severity of Dependence Scale (for cannabis); possible score range 0 - 15.

[f]1 missing data at T0, 4 missing data at T2.

[g]Outcome not assessed at T1.

[h]Cannabis dependence=score≥4.

[i]Kessler Screening Scale for Psychological Distress (K6); possible scores range 0 - 24.

[j]2 missing data at T0, 1 missing data at T2.

[k]Low/moderate psychological distress=scores between 0 and 13.

[l]Severe psychological distress=scores between 14 and 24.

Most participants in both the EG and the CG self-assessed their past-month CU frequency as between occasional and regular at baseline, and these proportions remained mostly similar at T1 and T2. The proportion of participants who reported daily CU in the past month decreased from 13% (5/39) at baseline to 4% (1/24) at T2 in the EG and from 7% (3/41) to 6% (2/36) in the CG.

Participants in both the EG and the CG expressed a moderate level of intention to take action on their CU at baseline, and this score tended to increase at T1 and T2 for participants in the EG. Similarly, participants in the EG and CG showed a moderate level of PBS use at baseline, and this result was found to increase in both groups at follow-up time points. Regarding psychological distress, the proportion of participants in the EG

XSL•FO
RenderX

who scored a severe level of psychological distress decreased from 24% (9/39) at baseline to 17% at T2 (4/24), while in the CG, the proportion remained quite similar from baseline (13/41, 32%) to T2 (11/36, 31%). Finally, 46% (18/39) and 35% (14/41) of the participants in the EG and the CG, respectively, scored from 4 to 15 on the SDS at baseline, which indicated cannabis dependence. At T2, the proportions decreased to 32% (7/24) and 29% (10/36), respectively.

Discussion

Principal Findings

This pilot RCT sought to evaluate the acceptability of the Joint Effort mobile app in a sample of Canadian-based university students who used cannabis in the past month, examine the feasibility of the research procedures, and provide data regarding outcome responsiveness.

Acceptability of the Joint Effort Mobile App

Engaging adequately with a mobile app presupposes first downloading it and then using it to an extent corresponding to the optimal exposure threshold. This usage will be largely—but not solely—influenced by the user's appreciation of the app. Therefore, we set out to assess user acceptability of the Joint Effort app via 3 distinct but interrelated parameters, namely, uptake, engagement, and appreciation.

Almost 60% of the participants in the EG downloaded the app (23/39). Thus, many participants who decided to participate in this study failed to use the proposed mobile app. One explanation could be related to potential technical issues with accessing the app and user account setup. Some users contacted the research team for assistance reinitializing their password to access the app. We can hypothesize that others faced the same problem but did not bother to contact us. Various factors can influence uptake, such as basic technological skills, app awareness, app availability and accessibility, social influences, curiosity, and user guidance [76]. In the context of this study, lack of user guidance might have had a negative/deteriorative impact on uptake. Further exploration of barriers and facilitators to uptake and strategies to optimize app download is needed.

Results suggest that Joint Effort was well received, deemed easy to use, and favorably rated by young adult cannabis users who used the proposed mobile app. For instance, the app obtained a good overall score on the uMARS (4.20/5, SD 0.5). Aesthetics was the dimension that obtained the highest score, followed by information, functionality, engagement, and subjective quality. These results are consistent with those obtained by Santesteban-Echarri et al [49], who reported a good uMARS overall score (mean 3.8, SD 0.5) in their pilot test of a mobile-based app to monitor CU among youth at high clinical risk for psychosis.

Engaging with a mobile app requires an autonomous process that depends on one's commitment to the behavior to be changed and one's self-management capabilities. In a systematic review of influences on uptake of and engagement with health and well-being smartphone apps across various fields (41 studies), Szinay et al [76] found that rapid disengagement was common and that uptake and engagement appeared to be influenced by a wide range of individual-level factors related to capability, opportunity, and motivation. In this study, in terms of quantity of exposure, 43% (10/23) of the downloaders viewed all sections, while 57% (13/23) completed the first 2 or 3 only. Regarding duration of exposure, our results showed a wide variation across participants in terms of frequency and time of use. This last finding might reflect the different purposes for which Joint Effort was being used.

Relatedly, in a systematic review of the concept of adherence to electronic health technology (34 studies), Sieverink et al [112] found that most of the time participants did not use technologies as suggested, whether in terms of features used, frequency of use, time of use, or place of use. These findings raise an important question, namely, whether all users need to experience all elements of a technology to achieve effects, especially considering that the intended engagement with an app may vary based on the utility objectives of the app and desired outcomes, as well as the user's personal goals and substance use behaviors. Successful engagement may not be a function solely of the amount of exposure (ie, the more use, the better). This emphasizes the importance of measuring engagement in different but complementary ways.

In this study, in addition to objective engagement (ie, frequency and duration of app usage data metrics collected automatically), we also measured subjective engagement using a self-reported questionnaire. The results showed that Aesthetic Appeal was the UES-SF subscale that obtained the highest score, followed by the Reward factor. Currently, there is no standardized approach in the existing body of literature to assessing engagement in mHealth interventions, as engagement depends on the targeted behaviors and how they are conceptualized, operationalized, and measured [113-115]. For instance, a scoping review of mHealth intervention apps [113] found that 20% of the studies included did not measure user engagement and, of those that did quantifiably (37/54), only 8% (3/37) used an established scale to evaluate engagement outcomes. Consequently, it is difficult to compare our results regarding user engagement with the Joint Effort app against those of other mobile app interventions for CU. Enhancing participant engagement with mHealth apps is undoubtedly a priority for wellness and health care. It remains a significant challenge for many app developers, researchers, and clinical teams [78,113].

Feasibility of Study Procedures

Recruitment primarily via social media ads resulted in the recruitment of our target sample over the span of 18 weeks. This underscores the feasibility of such methods but also highlights that recruitment costs and duration are key factors that must be considered from the outset to ensure successful recruitment. Although representativeness in a small pilot study cannot be determined, larger studies support the utility of web-based recruitment approaches [116,117]. In this pilot study, the recruitment rate was comparable to results from prior formative research on digital interventions targeting young adult cannabis users [33,118,119].

The overall study attrition rate was 18% (14/80), though there were more noncompleters in the EG than in the CG at the 2-month postbaseline assessment. In a similar pilot study

XSL•FO
RenderX

intended to examine the feasibility and acceptability of a smartphone app intervention for adult cannabis users seeking to reduce or cease use [119], attrition rates of 16% and 32% were reported at postintervention (after 4 weeks of using the app) and at one-month follow-up, respectively. Similarly, in a scoping review (54 studies) of the components of mHealth intervention apps that encourage or hinder engagement, participant retention ranged from 14% to 100%, the average was 68% (SD 25%), and the median 79% [113]. Notably, only one-third of the studies included in this review (18/54) reported retention rates of 80% or higher. Against these numbers, the attrition rate (14/80, 18%) or retention rate (66/80, 83%) observed in this study can be considered acceptable.

Participant recruitment and retention are recognized as two of the most important challenges when conducting clinical trials [120]. These challenges are all the greater when research is conducted online. For example, a systematic review and meta-analysis (17 studies) by Meyerowitz-Katz et al [121] found that the "pooled estimate for dropout rates in trials of app-based interventions for chronic diseases was 43% over a variety of timelines". In another review (62 studies), Amagai et al [122] reported that mHealth apps were hampered by substantial participant dropout or attrition and suggested that factors related to particular app elements (eg, feedback, reminders, in-app support from coaches and peers) and research strategies (eg, compensation and niche samples) should be considered to promote retention. Of note, participant engagement in mHealth apps and study retention (or attrition) are distinct constructs that should not be used interchangeably. Engagement in a given intervention can affect a study's retention/attrition rate. Given the interconnected nature of these constructs, an integrated approach to analyzing recruitment, retention/attrition, and engagement outcomes should be used in evaluations of mHealth apps to broaden the evidence base but, more importantly, to determine how different strategies impact these three (ie, participant recruitment, retention, and engagement).

A second feasibility objective of this pilot trial was to document the research process throughout the study period in order to pinpoint and learn from the challenges and other aspects of the procedures applied before moving on to conduct a large-scale web-based RCT of the Joint Effort app. A major challenge faced during this study was the detection and management of fraudulent applications and the identification of information provided by potentially eligible or eligible participants. After reviewing completed surveys for indicators of fraudulent or suspicious data, 19% (19/99) of enrolled participants were found to have provided inconsistent information and, subsequently, were excluded from analysis. Similarly, other web-based trials or survey studies that monitored misrepresentation and fraud have consistently reported fraud rates of 20% to 35% [123-126]. Unfortunately, fraudulent behavior is a fairly widespread phenomenon in online research [127-129], which can have a significant impact on data integrity and quality, sample size and composition, and statistical precision [108,130,131].

Some researchers have outlined processes or methods, indicators, and logistic frameworks to address potential fraud [129,132,133]. In this pilot study, several strategies were rigorously incorporated to prevent, detect, and respond to fraudulent behavior by applicants/participants. These included automated procedures to prevent the same individual from registering more than once with the same email or username, as well as human monitoring (eg, checking for multiple registrations occurring almost simultaneously, identifying suspicious/fake email addresses, and unrealistic survey completion times). We also included combinations of survey questions repeated at each measurement time point (baseline, T1, and T2) designed to identify inconsistent responses in a participant's identity information (eg, date of birth). This method has been identified as an effective strategy for accurately detecting inconsistencies and preventing data quality threats, including fraudulent entries from automated bots and duplicate submissions [108,134,135]. Finally, financial compensation was given only upon completion of the follow-up assessments, as it has been shown that financial incentives can motivate applicants to register more than once [133] and relatively immediate payments have been found to yield greater rates of fraudulent behavior by participants [127].

Although all the strategies implemented in this pilot study certainly contributed to preventing, detecting, and handling fraud, the fact remains that it may be impossible to completely eliminate it. In an online cross-sectional study (414 survey entries), Ballard and colleagues [123] found that, consistent with our observations, even when protocols were put in place to improve fraud detection, 28.7% of submitted web-based surveys were "fraudulent" and an additional 10.1% were cases of "potential fraud." These authors also suggested that researchers should have a fraud detection algorithm in place prior to data collection and use a combination of indicators and methods to systematically review all online applications and verify participant identity and eligibility [123]. Based on the results of this study and our experience designing and implementing pilot RCTs using social media recruitment and web-based data collection, we encourage researchers conducting studies of mHealth apps to be vigilant about the potential for fraud, to keep detailed records of the verification process, and to report their study findings accordingly. Ultimately, there is a need for further reflection regarding the advantages and disadvantages associated with fraud prevention and detection methods, given that ethical tensions can emerge between preserving the integrity of research versus protecting the privacy and confidentiality of study participants.

End User Profile and Outcome Responsiveness

In terms of participant characteristics, a higher proportion of women was reached compared with men. This adds to the growing body of literature indicating that CU treatment-related differences exist between men and women [10,136-138]. Although help-seeking behaviors to address CU problems are rare among both young adult men and young adult women [139,140], women are more likely to seek out anonymous substance use resources, such as web-based or mobile app interventions [33,141]. Women may also more readily recognize the need to change their CU in an attempt to better manage problems related to their CU. Lending support to this hypothesis, research has shown that women often endorse external reasons for change, such as experiencing cannabis-related negative consequences and social acceptability, whereas men seem to

XSL•FO
RenderX

be motivated more by internal reasons, such as self-control and self-efficacy [142-145]. This means that women may be more willing to make behavioral changes and access strategies and motivational support to prevent or reduce the consequences/risks associated with CU. Given that cannabis-related negative consequences are disproportionally more common among young adult males [10,146], it would be worthwhile for future studies to ask young men directly about what they consider motivational or attractive to initiate change. In sum, our findings highlight the challenges of reaching target populations and developing mobile app CU interventions that take into account gender differences across a range of other clinical correlates (eg, readiness to change, sources of motivation to behavior change) that can affect young adults in their decision to take steps towards better CU management.

Regarding change in outcomes over time, our results showed that participants in both the EG and the CG indicated a moderate level of intention to take action on CU at baseline, and that the scores in this regard tended to increase at the follow-up assessments for participants in the EG. Similarly, participants in the EG and the CG indicated a moderate level of PBS use at baseline, and this level was found to increase in both groups over time. Regarding CU frequency, most participants in both groups self-assessed it as occasional to regular at baseline, with similar proportions at T1 and T2. Moreover, from baseline to postintervention, the proportion of participants in the EG who reported daily CU in the past month decreased by 9%, and those who reached the cut-off for cannabis dependence decreased by 12%. Overall, our results indicate that participants responded positively to Joint Effort and demonstrated some potential for behavior change as a result.

Limitations

This study, focusing on user acceptability, study feasibility, and outcome responsiveness, presents some limitations.

One of the major limitations related to our Internet-based recruitment strategy—which is an inherent challenge in online intervention trials—is the discovery of fraudulent identities after participants were enrolled in the study. This is a critical issue as it directly altered the randomization process and, to some extent, may have impacted the integrity of the research. Although several precautions were taken throughout the study to prevent and detect fraudulent applications using both automated detection procedures and human monitoring, our fraud detection methods were not infallible. This limitation is partially due to the lack of in-person interactions between research team members and participants, which increases the risk that applicants are not who they claim to be. The participation incentives offered in the study are another factor that may have contributed to increased fraud. A multi-layered approach to fraud prevention and detection may be warranted in future studies of mHealth apps to help reduce the inclusion of fraudulent applications [127,129,134]. Despite the numerous risks associated with web-based recruitment, this method nevertheless has significant advantages over other traditional recruitment strategies. This pilot study thus affords valuable lessons for subsequent testing of Joint Effort and highlights the need for careful implementation of processes to help address potential fraud in future online trials and in web-based survey research in general.

We focused our means of recruitment mainly on social media and posters on campus in an attempt to reach a diverse population of university students. This strategy allowed us to recruit young adult cannabis users from a large university student population and achieve our sample size goal over a relatively short period of time. However, this recruitment method may have introduced a self-selection bias. We certainly recruited people interested in engaging with the mobile app and motivated to better manage their CU, which may not be representative of the target population. Also, the fact that the app was available on iOS only might have led to a different sociodemographic sample. The decision to develop the mobile application on iOS was based on financial limitations (ie, the cost associated with the development on both smartphone operating systems exceeded the budget). Also, based on a previous study conducted among a similar sample [68,69], more respondents used an iPhone to complete the survey compared with an Android (respectively 41% vs 19%) (data not published). This target population likely had a higher level of digital literacy than the broader population, a factor that potentially limits the transferability of our findings.

Another limitation is related to samples of young adult cannabis users that were relatively small at each time point, with the EG experiencing greater attrition than the CG. Consequently, and despite the generally positive findings supporting the potential of the Joint Effort app, further study is required with a larger sample to determine not only the efficacy of Joint Effort on CU, but also the mediators to use (eg, intention to use, PBS) and the consequences of CU (eg, cannabis dependence). These findings justify further testing of Joint Effort in a fully powered trial.

Conclusions

Most existing interventions for young adult cannabis users wishing to better manage their use are limited by low uptake due to accessibility and stigma concerns. Advancements in technology have brought new ways in which a wide range of digital interventions can be developed and delivered. Joint Effort offers the advantage of convenient delivery via a smartphone app, which considerably minimizes these accessibility and stigma concerns. This pilot study provides support for Joint Effort's acceptability within the target population as an mHealth intervention for lower-risk CU and suggests that the study design used is feasible. This warrants future testing in a fully powered trial (ClinicalTrials.gov NCT05620433).

Acknowledgments

We thank Emma Février for her support with data collection and Michael Yu for his contribution to data analysis. We also thank our corporate partners, iTECH GO, Akufen, and Osedea, and, last but not least, the students who participated in the research.

XSL•FO
RenderX

Funding

This study was supported by a grant from the Ministère de la Santé et des Services Sociaux du Québec. The views and opinions expressed in this manuscript do not necessarily reflect those of the funding entities. J Côté is Chairholder of the Research Chair in Innovative Nursing Practices. J Cheng is the Chairholder of the Canada Research Chair in User Experience Design for Data-Driven Systems. GC is the recipient of a Fellowship Award from the Canadian Institute of Health Research (CIHR). GF is the recipient of a Junior 1 Research Scholar Award from the Fonds de recherche du Québec – Santé (FRQS). SL is supported through the Canada Research Chairs program. MGP is a Junior 2 Research Scholar supported by the Fonds de la recherche du Québec – Santé (FRQS). DJA is a Senior Clinical Research Scholar supported by FRQS.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

J Côté and PA contributed to conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, visualization, and writing the original draft. GC contributed to conceptualization, methodology, formal analysis, writing the original draft, visualization, and project administration. BV contributed to formal analysis, investigation, writing the original draft, visualization, and project administration. SC, CG, SL, JL, MGP, GR, and MAMC contributed to conceptualization, funding acquisition, methodology, and writing (review and editing). J Cheng contributed to conceptualization, funding acquisition, methodology, software, and writing (review and editing). GF contributed to conceptualization, formal analysis, funding acquisition, methodology, and writing (review and editing). DJA contributed to conceptualization, formal analysis, funding acquisition, methodology, supervision, and writing (review and editing).

Conflicts of Interest

DJA received study material from Cardiol Therapeutics for a clinical trial funded by the Quebec Ministry of Health and Social Services. All of the other authors declare they have no conflict of interest.

Multimedia Appendix 1
Joint Effort mean scores on the user version of the Mobile App Rating Scale (uMARS).
[PDF File, 211 KB - mhealth_v14i1e71957_app1.pdf]

Checklist 1
CONSORT 2010 extension for Pilot and Feasibility Trials Checklist.
[PDF File, 207 KB - mhealth_v14i1e71957_app2.pdf]

Checklist 2
TIDier Checklist.
[PDF File, 371 KB - mhealth_v14i1e71957_app3.pdf]

References

1. World drug report 2020: drug use and health consequences. United Nations Office on Drugs and Crime.: United Nations publication; 2020. URL: https://wdr.unodc.org/wdr2020/field/WDR20_Booklet_2.pdf [accessed 2023-12-22]
2. Rotermann M. What has changed since cannabis was legalized? Health Rep 2020 Feb 19;31(2):11-20. [doi: 10.25318/82-003-x202000200002-eng] [Medline: 32073644]
3. Page RL 2nd, Allen LA, Kloner RA, et al. Medical marijuana, recreational cannabis, and cardiovascular health: a scientific statement from the American Heart Association. Circulation 2020 Sep 8;142(10):e131-e152. [doi: 10.1161/CIR.0000000000000883] [Medline: 32752884]
4. Degenhardt L, Stockings E, Patton G, Hall WD, Lynskey M. The increasing global health priority of substance use in young people. Lancet Psychiatry 2016 Mar;3(3):251-264. [doi: 10.1016/S2215-0366(15)00508-8] [Medline: 26905480]
5. Hall W, Stjepanović D, Caulkins J, et al. Public health implications of legalising the production and sale of cannabis for medicinal and recreational use. The Lancet 2019 Oct;394(10208):1580-1590. [doi: 10.1016/S0140-6736(19)31789-1]
6. ElSohly MA, Chandra S, Radwan M, Majumdar CG, Church JC. A comprehensive review of cannabis potency in the United States in the last decade. Biol Psychiatry Cogn Neurosci Neuroimaging 2021 Jun;6(6):603-606. [doi: 10.1016/j.bpsc.2020.12.016] [Medline: 33508497]
7. Chan GCK, Becker D, Butterworth P, et al. Young-adult compared to adolescent onset of regular cannabis use: a 20-year prospective cohort study of later consequences. Drug Alcohol Rev 2021 May;40(4):627-636. [doi: 10.1111/dar.13239] [Medline: 33497516]

XSL•FO
RenderX

8. Sun Y, Liu B, Wallace RB, Bao W. Association of cannabis use with all-cause and cause-specific mortality among younger- and middle-aged U.S. adults. Am J Prev Med 2020 Dec;59(6):873-879. [doi: 10.1016/j.amepre.2020.07.010]
9. Degenhardt L, Chiu WT, Sampson N, et al. Toward a global view of alcohol, tobacco, cannabis, and cocaine use: findings from the WHO world mental health surveys. PLoS Med 2008 Jul 1;5(7):e141. [doi: 10.1371/journal.pmed.0050141] [Medline: 18597549]
10. Fischer B, Robinson T, Bullen C, et al. Lower-risk cannabis use guidelines (LRCUG) for reducing health harms from non-medical cannabis use: a comprehensive evidence and recommendations update. Int J Drug Policy 2022 Jan;99:103381. [doi: 10.1016/j.drugpo.2021.103381] [Medline: 34465496]
11. Zhu H, Wu LT. Trends and correlates of cannabis-involved emergency department visits: 2004 to 2011. J Addict Med 2016;10(6):429-436. [doi: 10.1097/ADM.0000000000000256] [Medline: 27574753]
12. Volkow ND, Baler RD, Compton WM, Weiss SRB. Adverse health effects of marijuana use. N Engl J Med 2014 Jun 5;370(23):2219-2227. [doi: 10.1056/NEJMra1402309]
13. Yang LH, Wong LY, Grivel MM, Hasin DS. Stigma and substance use disorders: an international phenomenon. Curr Opin Psychiatry 2017 Sep;30(5):378-388. [doi: 10.1097/YCO.0000000000000351] [Medline: 28700360]
14. Pearson MR. A meta-analytic investigation of the associations between cannabis use and cannabis-related negative consequences. Psychol Addict Behav 2019 May;33(3):190-196. [doi: 10.1037/adb0000452]
15. The health and social effects of nonmedical cannabis use 2016. World Health Organization. URL: https://apps.who.int/iris/handle/10665/251056 [accessed 2023-12-22]
16. National Academies of Sciences E, and Medicine. The Health Effects of Cannabis and Cannabinoids: The Current State of Evidence and Recommendations for Research: The National Academies Press; 2017. URL: https://doi.org/10.17226/24625 [doi: 10.17226/24625]
17. Hasin D, Walsh C. Cannabis use, cannabis use disorder, and comorbid psychiatric illness: a narrative review. JCM 2021;10(1):15. [doi: 10.3390/jcm10010015] [Medline: 33920719]
18. Memedovich KA, Dowsett LE, Spackman E, Noseworthy T, Clement F. The adverse health effects and harms related to marijuana use: an overview review. CMAJ Open 2018;6(3):E339-E346. [doi: 10.9778/cmajo.20180023] [Medline: 30115639]
19. Carliner H, Brown QL, Sarvet AL, Hasin DS. Cannabis use, attitudes, and legal status in the U.S.: a review. Prev Med 2017 Nov;104:13-23. [doi: 10.1016/j.ypmed.2017.07.008] [Medline: 28705601]
20. Salas-Wright CP, Vaughn MG, Perron BE, Gonzalez JMR, Goings TC. Changing perspectives on marijuana use during early adolescence and young adulthood: evidence from a panel of cross-sectional surveys. Drug Alcohol Depend 2016 Dec;169(5–10):5-10. [doi: 10.1016/j.drugalcdep.2016.09.026]
21. Peterson R, Kramer MP, Pinto D, et al. A comprehensive review of measures of protective behavioral strategies across various risk factors and associated PBS-related interventions. Exp Clin Psychopharmacol 2021 Jun;29(3):236-250. [doi: 10.1037/pha0000498] [Medline: 34264735]
22. Cilliers F, Schuwirth L, Vleuten C. Health behaviour theories: a conceptual lens to explore behaviour change. In: Cleland J, Durning SJ, editors. Researching Medical Education: John Wiley & Sons, Ltd; 2015:141-154. [doi: 10.1002/9781118838983]
23. Davis ML, Powers MB, Handelsman P, Medina JL, Zvolensky M, Smits JAJ. Behavioral therapies for treatment-seeking cannabis users: a meta-analysis of randomized controlled trials. Eval Health Prof 2015 Mar;38(1):94-114. [doi: 10.1177/0163278714529970] [Medline: 24695072]
24. Webb TL, Sniehotta FF, Michie S. Using theories of behaviour change to inform interventions for addictive behaviours. Addiction 2010 Nov;105(11):1879-1892. [doi: 10.1111/j.1360-0443.2010.03028.x] [Medline: 20670346]
25. Holmes J, Angus C, Meier PS, Buykx P, Brennan A. How should we set consumption thresholds for low risk drinking guidelines? achieving objectivity and transparency using evidence, expert judgement and pragmatism. Addiction 2019 Apr;114(4):590-600. [doi: 10.1111/add.14381] [Medline: 30133036]
26. Norberg MM, Kezelman S, Lim-Howe N. Primary prevention of cannabis use: a systematic review of randomized controlled trials. PLoS ONE 2013;8(1):e53187. [doi: 10.1371/journal.pone.0053187] [Medline: 23326396]
27. Perski O, Hébert ET, Naughton F, Hekler EB, Brown J, Businelle MS. Technology-mediated just-in-time adaptive interventions (JITAIs) to reduce harmful substance use: a systematic review. Addiction 2022 May;117(5):1220-1241. [doi: 10.1111/add.15687] [Medline: 34514668]
28. Kazemi DM, Borsari B, Levine MJ, Li S, Lamberson KA, Matta LA. A systematic review of the mHealth interventions to prevent alcohol and substance abuse. J Health Commun 2017 May;22(5):413-432. [doi: 10.1080/10810730.2017.1303556] [Medline: 28394729]
29. Nesvåg S, McKay JR. Feasibility and effects of digital interventions to support people in recovery from substance use disorders: systematic review. J Med Internet Res 2018 Aug 23;20(8):e255. [doi: 10.2196/jmir.9873]
30. Stockings E, Hall WD, Lynskey M, et al. Prevention, early intervention, harm reduction, and treatment of substance use in young people. Lancet Psychiatry 2016 Mar;3(3):280-296. [doi: 10.1016/S2215-0366(16)00002-X]
31. Bedrouni W. On the use of digital technologies to reduce the public health impacts of cannabis legalization in Canada. Can J Public Health 2018 Dec;109(5-6):748-751. [doi: 10.17269/s41997-018-0117-7] [Medline: 30097890]

XSL•FO
RenderX

32. Guo H, Yang H, Yuan G, et al. Effectiveness of information and communication technology(ICT) for addictive behaviors: an umbrella review of systematic reviews and meta-analysis of randomized controlled trials. Comput Human Behav 2023 Oct;147:107843. [doi: 10.1016/j.chb.2023.107843] [Medline: 36268221]
33. Coughlin LN, Nahum-Shani I, Philyaw-Kotov ML, et al. Developing an adaptive mobile intervention to address risky substance use among adolescents and emerging adults: usability study. JMIR Mhealth Uhealth 2021 Jan 15;9(1):e24424. [doi: 10.2196/24424] [Medline: 33448931]
34. Côté J, Chicoine G, Vinette B, et al. Digital interventions for recreational cannabis use among young adults: systematic review, meta-analysis, and behavior change technique analysis of randomized controlled studies. J Med Internet Res 2024 Apr 17;26:e55031. [doi: 10.2196/55031] [Medline: 38630515]
35. Beneria A, Santesteban-Echarri O, Daigre C, et al. Online interventions for cannabis use among adolescents and young adults: systematic review and meta-analysis. Early Interv Psychiatry 2022 Aug;16(8):821-844. [doi: 10.1111/eip.13226] [Medline: 34464502]
36. Halladay J, Scherer J, MacKillop J, et al. Brief interventions for cannabis use in emerging adults: a systematic review, meta-analysis, and evidence map. Drug Alcohol Depend 2019 Nov 1;204(107565):107565. [doi: 10.1016/j.drugalcdep.2019.107565] [Medline: 31751868]
37. Hoch E, Preuss UW, Ferri M, Simon R. Digital Interventions for problematic cannabis users in non-clinical settings: findings from a systematic review and meta-analysis. Eur Addict Res 2016;22(5):233-242. [doi: 10.1159/000445716] [Medline: 27160333]
38. Imtiaz S, Roerecke M, Kurdyak P, Samokhvalov AV, Hasan OSM, Rehm J. Brief interventions for cannabis use in healthcare settings: systematic review and meta-analyses of randomized trials. J Addict Med 2020;14(1):78-88. [doi: 10.1097/ADM.0000000000000527] [Medline: 32012140]
39. Sedrati H, Belrhiti Z, Nejjari C, Ghazal H. Evaluation of mobile health apps for non-medical cannabis use: a scoping review. Procedia Comput Sci 2022;196:581-589. [doi: 10.1016/j.procs.2021.12.052]
40. Tait RJ, Spijkerman R, Riper H. Internet and computer based interventions for cannabis use: a meta-analysis. Drug Alcohol Depend 2013 Dec 1;133(2):295-304. [doi: 10.1016/j.drugalcdep.2013.05.012] [Medline: 23747236]
41. Boumparis N, Loheide-Niesmann L, Blankers M, et al. Short- and long-term effects of digital prevention and treatment interventions for cannabis use reduction: a systematic review and meta-analysis. Drug Alcohol Depend 2019 Jul 1;200:82-94. [doi: 10.1016/j.drugalcdep.2019.03.016] [Medline: 31112834]
42. Olmos A, Tirado-Muñoz J, Farré M, Torrens M. The efficacy of computerized interventions to reduce cannabis use: a systematic review and meta-analysis. Addict Behav 2018 Apr;79:52-60. [doi: 10.1016/j.addbeh.2017.11.045] [Medline: 29248863]
43. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
44. Thornton L, Brown HM, Osman B, et al. Factors associated with adolescents' engagement with a healthy lifestyles app. Procedia Comput Sci 2022;206:56-67. [doi: 10.1016/j.procs.2022.09.085]
45. Key substance use and mental health indicators in the United States: results from the 2020 National Survey on Drug Use and Health (HHS publication nopep21-07-01-003, NSDUH series h-56). Substance Abuse and Mental Health Services Administration (SAMHSA).: Center for Behavioral Health Statistics and Quality, SAMHSA; 2021. URL: https://www.samhsa.gov/data/sites/default/files/reports/rpt35325/NSDUHFFRPDFWHTMLFiles2020/2020NSDUHFFR1PDFW102121.pdf [accessed 2023-12-22]
46. Adams Z, Grant M, Hupp S, et al. Acceptability of an mHealth app for youth with substance use and mental health needs: iterative, mixed methods design. JMIR Form Res 2021 Dec 24;5(12):e30268. [doi: 10.2196/30268] [Medline: 34951593]
47. Howlett N, García-Iglesias J, Bontoft C, et al. A systematic review and behaviour change technique analysis of remotely delivered alcohol and/or substance misuse interventions for adults. Drug Alcohol Depend 2022 Oct 1;239:109597. [doi: 10.1016/j.drugalcdep.2022.109597] [Medline: 35963209]
48. Scott CK, Dennis ML, Johnson KA, Grella CE. A randomized clinical trial of smartphone self-managed recovery support services. J Subst Abuse Treat 2020 Oct;117:108089. [doi: 10.1016/j.jsat.2020.108089] [Medline: 32811628]
49. Santesteban-Echarri O, Haffey P, Kim G, Tang J, Addington J. A mobile-based app to monitor cannabis use among youth at clinical high risk (CHR) for psychosis: feasibility and acceptability of LooseLeaf. Schizophr Res 2020 Aug;222:505-506. [doi: 10.1016/j.schres.2020.05.031] [Medline: 32522468]
50. Baumgartner C, Schaub MP, Wenger A, et al. CANreduce 2.0 adherence-focused guidance for internet self-help among cannabis users: three-arm randomized controlled trial. J Med Internet Res 2021;23(4):e27463. [doi: 10.2196/27463]
51. Bonar EE, Chapman L, Pagoto S, et al. Social media interventions addressing physical activity among emerging adults who use cannabis: a pilot trial of feasibility and acceptability. Drug Alcohol Depend 2023 Jan 1;242:109693. [doi: 10.1016/j.drugalcdep.2022.109693] [Medline: 36442441]
52. Bonar EE, Goldstick JE, Chapman L, et al. A social media intervention for cannabis use among emerging adults: randomized controlled trial. Drug Alcohol Depend 2022 Mar 1;232:109345. [doi: 10.1016/j.drugalcdep.2022.109345] [Medline: 35144238]

XSL•FO
RenderX

53. Jonas B, Tensil MD, Tossmann P, Strüber E. Effects of treatment length and chat-based counseling in a web-based intervention for cannabis users: randomized factorial trial. J Med Internet Res 2018 May 8;20(5):e166. [doi: 10.2196/jmir.9579] [Medline: 29739738]
54. Schaub MP, Wenger A, Berg O, et al. A web-based self-help intervention with and without chat counseling to reduce cannabis use in problematic cannabis users: three-arm randomized controlled trial. J Med Internet Res 2015 Oct 13;17(10):e232. [doi: 10.2196/jmir.4860] [Medline: 26462848]
55. Sinadinovic K, Johansson M, Johansson AS, Lundqvist T, Lindner P, Hermansson U. Guided web-based treatment program for reducing cannabis use: a randomized controlled trial. Addict Sci Clin Pract 2020 Feb 18;15(1):9. [doi: 10.1186/s13722-020-00185-8] [Medline: 32070417]
56. Tossmann HP, Jonas B, Tensil MD, Lang P, Strüber E. A controlled trial of an internet-based intervention program for cannabis users. Cyberpsychol Behav Soc Netw 2011 Nov;14(11):673-679. [doi: 10.1089/cyber.2010.0506] [Medline: 21651419]
57. Ramo DE, Popova L, Grana R, Zhao S, Chavez K. Cannabis mobile apps: a content analysis. JMIR Mhealth Uhealth 2015 Aug 12;3(3):e81. [doi: 10.2196/mhealth.4405] [Medline: 26268634]
58. Côté J, Auger P, Chicoine G, et al. A mobile app (joint effort) to support cannabis use self-management and reinforce the use of protective behavioral strategies: development process and usability testing. JMIR Form Res 2025 Jun 23;9:e71924. [doi: 10.2196/71924] [Medline: 40550117]
59. Ajzen I. From intentions to actions: a theory of planned behavior. In: Kuhl J, Beckmann J, editors. Action Control: From Cognition to Behavior: Springer; 1985:11-39.
60. Pedersen ER, Huang W, Dvorak RD, Prince MA, Hummer JF, (The Marijuana Outcomes Study Team). The protective behavioral strategies for marijuana scale: further examination using item response theory. Psychol Addict Behav 2017;31(5):548-559. [doi: 10.1037/adb0000271]
61. Pearson MR. Use of alcohol protective behavioral strategies among college students: a critical review. Clin Psychol Rev 2013 Dec;33(8):1025-1040. [doi: 10.1016/j.cpr.2013.08.006] [Medline: 24036089]
62. Riggs NR, Conner BT, Parnes JE, Prince MA, Shillington AM, George MW. Marijuana eCHECKUPTO GO: effects of a personalized feedback plus protective behavioral strategies intervention for heavy marijuana-using college students. Drug Alcohol Depend 2018 Sep 1;190:13-19. [doi: 10.1016/j.drugalcdep.2018.05.020] [Medline: 29960918]
63. Fedele DA, Cushing CC, Fritz A, Amaro CM, Ortega A. Mobile health interventions for improving health outcomes in youth: a meta-analysis. JAMA Pediatr 2017 May 1;171(5):461-469. [doi: 10.1001/jamapediatrics.2017.0042] [Medline: 28319239]
64. Schnall R, Rojas M, Bakken S, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251. [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
65. Garrett JJ. The Elements of User Experience: User-Centered Design for the Web and Beyond, 2nd edition: New Riders; 2011.
66. The field guide to human-centered design: design kit. IDEO.org. 2015. URL: https://www.designkit.org/resources/1.html [accessed 2023-12-01]
67. Maguire M. Methods to support human-centred design. Int J Hum Comput Stud 2001 Oct;55(4):587-634. [doi: 10.1006/ijhc.2001.0503]
68. Côté J, Auger P, Pagé GM, et al. French translation, cultural adaptation and assessment of preliminary psychometric properties of the protective behavioral strategies for marijuana scale. Can J Psychiatry 2022 Aug;67(8):608-615. [doi: 10.1177/07067437211025216] [Medline: 34160302]
69. Côté J, Cossette S, Auger P, et al. Psychometric properties of the French and English short form of the protective behavioural strategies for marijuana scale in Canadian university students. BMJ Open 2022 Apr 6;12(4):e053715. [doi: 10.1136/bmjopen-2021-053715] [Medline: 35387810]
70. Eldridge SM, Lancaster GA, Campbell MJ, et al. Defining feasibility and pilot studies in preparation for randomised controlled trials: development of a conceptual framework. PLoS One 2016;11(3):e0150205. [doi: 10.1371/journal.pone.0150205] [Medline: 26978655]
71. Eldridge SM, Chan CL, Campbell MJ, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Pilot Feasibility Stud 2016;2(64):64. [doi: 10.1186/s40814-016-0105-8] [Medline: 27965879]
72. Teare MD, Dimairo M, Shephard N, Hayman A, Whitehead A, Walters SJ. Sample size requirements to estimate key design parameters from external pilot randomised controlled trials: a simulation study. Trials 2014 Jul 3;15(264):264. [doi: 10.1186/1745-6215-15-264] [Medline: 24993581]
73. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014 Mar 7;348(mar07 3):g1687-g1687. [doi: 10.1136/bmj.g1687]
74. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
75. Feeley N, Cossette S, Côté J, et al. The importance of piloting an RCT intervention. Can J Nurs Res 2009 Jun;41(2):85-99. [Medline: 19650515]

XSL•FO
RenderX

76. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572. [doi: 10.2196/17572] [Medline: 32348255]
77. O'Brien H. Theoretical perspectives on user engagement. In: O'Brien H, Cairns P, editors. Why Engagement Matters: Cross-Disciplinary Perspectives of User Engagement in Digital Media: Springer International Publishing; 2016:1-26. [doi: 10.1007/978-3-319-27446-1_1]
78. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
79. O'Brien HL, Lebow M. Mixed‐methods approach to measuring user experience in online news interactions. J Am Soc Inf Sci Tec 2013 Aug;64(8):1543-1556. [doi: 10.1002/asi.22871]
80. O'Brien HL, Cairns P, Hall M. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. Int J Hum Comput Stud 2018 Apr;112:28-39. [doi: 10.1016/j.ijhcs.2018.01.004]
81. O'Brien H. Translating theory into methodological practice. In: O'Brien H, Cairns P, editors. Why Engagement Matters: Cross-Disciplinary Perspectives of User Engagement in Digital Media: Springer International Publishing; 2016:27-52.
82. Wiebe EN, Lamb A, Hardy M, Sharek D. Measuring engagement in video game-based environments: investigation of the user engagement scale. Comput Human Behav 2014 Mar;32:123-132. [doi: 10.1016/j.chb.2013.12.001]
83. Fontaine G, Cossette S, Maheu-Cadotte MA, et al. Translation, adaptation and preliminary psychometric evaluation of a measure of engagement and a measure of cognitive load in a digital learning context [Article in French]. Pédagogie médicale 2019;20(2):79-90. [doi: 10.1051/pmed/2020009]
84. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the mobile application rating scale (uMARS). JMIR Mhealth Uhealth 2016;4(2):e72. [doi: 10.2196/mhealth.5849]
85. Bricca A, Swithenbank Z, Scott N, et al. Predictors of recruitment and retention in randomized controlled trials of behavioural smoking cessation interventions: a systematic review and meta-regression analysis. Addiction 2022 Feb;117(2):299-311. [doi: 10.1111/add.15614] [Medline: 34159677]
86. Canadian community health survey - annual component (CCHS). Statistics Canada. 2023. URL: https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3226 [accessed 2024-03-11]
87. Fishbein M, Ajzen I. Predicting and Changing Behavior the Reasoned Action Approach: Psychology Press; 2010.
88. Ajzen I, Sheikh S. Action versus inaction: anticipated affect in the theory of planned behavior. J Applied Social Pyschol 2013 Jan;43(1):155-162. [doi: 10.1111/j.1559-1816.2012.00989.x]
89. Côté J, Tessier S, Gagnon H, April N, Rouleau G, Chagnon M. Efficacy of a web-based tailored intervention to reduce cannabis use among young people attending adult education centers in Quebec. Telemed J E Health 2018 Nov;24(11):853-860. [doi: 10.1089/tmj.2017.0144] [Medline: 29466093]
90. Gossop M, Darke S, Griffiths P, et al. The severity of dependence scale (SDS): psychometric properties of the SDS in English and Australian samples of heroin, cocaine and amphetamine users. Addiction 1995 May;90(5):607-614. [doi: 10.1046/j.1360-0443.1995.9056072.x] [Medline: 7795497]
91. Tremblay J, Blanchette-Martin N. Manuel d'utilisation du déBA-alcool/drogues/jeu-8 - recherche et intervention sur les substances psychoactives. Québec, QC: Centre Intégré Universitaire de Santé et de Services Sociaux de La Capitale-Nationale. 2016. URL: https://www.ciusss-capitalenationale.gouv.qc.ca/sites/d8/files/docs/MissionUniversitaire/Recherche/grouperecherche/DSMD_manuel_utilisation_deba.pdf [accessed 2023-11-14]
92. Martin G, Copeland J, Gates P, Gilmour S. The severity of dependence scale (SDS) in an adolescent population of cannabis users: reliability, validity and diagnostic cut-off. Drug Alcohol Depend 2006 Jun;83(1):90-93. [doi: 10.1016/j.drugalcdep.2005.10.014]
93. Swift W, Copeland J, Hall W. Choosing a diagnostic cut‐off for cannabis dependence. Addiction 1998 Nov;93(11):1681-1692. [doi: 10.1046/j.1360-0443.1998.931116816.x]
94. Hides L, Dawe S, Young RM, Kavanagh DJ. The reliability and validity of the severity of dependence scale for detecting cannabis dependence in psychosis. Addiction 2007 Jan;102(1):35-40. [doi: 10.1111/j.1360-0443.2006.01669.x] [Medline: 17207121]
95. Kedzior KK, Badcock JC, Martin-Iverson MT. Validity and consistency of self-reports regarding substance use in general research volunteers, including regular cannabis users and schizophrenia patients. Subst Use Misuse 2006;41(5):743-750. [doi: 10.1080/10826080500411502] [Medline: 16603458]
96. Piontek D, Kraus L, Klempova D. Short scales to assess cannabis-related problems: a review of psychometric properties. Subst Abuse Treat Prev Policy 2008 Dec 2;3(25):19055741. [doi: 10.1186/1747-597X-3-25] [Medline: 19055741]
97. van der Pol P, Liebregts N, de Graaf R, Korf DJ, van den Brink W, van Laar M. Reliability and validity of the severity of dependence scale for detecting cannabis dependence in frequent cannabis users. Int J Methods Psychiatr Res 2013 Jun;22(2):138-143. [doi: 10.1002/mpr.1385] [Medline: 23670783]
98. Steiner S, Baumeister SE, Kraus L. Severity of dependence scale: establishing a cut-off point for cannabis dependence in the German adult population. SUCHT 2008 Sep;54(S1):57-63. [doi: 10.1024/2008.07.07]

XSL•FO
RenderX

99. Kessler RC, Barker PR, Colpe LJ, et al. Screening for serious mental illness in the general population. Arch Gen Psychiatry 2003 Feb;60(2):184-189. [doi: 10.1001/archpsyc.60.2.184] [Medline: 12578436]
100. Kessler RC, Green JG, Gruber MJ, et al. Screening for serious mental illness in the general population with the K6 screening scale: results from the WHO World Mental Health (WMH) survey initiative. Int J Methods Psych Res 2010 Jun;19(S1):4-22. [doi: 10.1002/mpr.310]
101. Staples LG, Dear BF, Gandy M, et al. Psychometric properties and clinical utility of brief measures of depression, anxiety, and general distress: the PHQ-2, GAD-2, and K-6. Gen Hosp Psychiatry 2019;56:13-18. [doi: 10.1016/j.genhosppsych.2018.11.003] [Medline: 30508772]
102. Mewton L, Kessler RC, Slade T, et al. The psychometric properties of the Kessler Psychological Distress Scale (K6) in a general population sample of adolescents. Psychol Assess 2016 Oct;28(10):1232-1242. [doi: 10.1037/pas0000239] [Medline: 26619095]
103. Kang YK, Guo WJ, Xu H, et al. The 6-item Kessler psychological distress scale to survey serious mental illness among Chinese undergraduates: psychometric properties and prevalence estimate. Compr Psychiatry 2015 Nov;63:105-112. [doi: 10.1016/j.comppsych.2015.08.011]
104. Peiper N, Clayton R, Wilson R, Illback R. The performance of the K6 Scale in a large school sample. Psychol Assess 2015 Mar;27(1):228-238. [doi: 10.1037/pas0000025]
105. Ferro MA. The psychometric properties of the Kessler Psychological Distress Scale (K6) in an epidemiological sample of Canadian youth. Can J Psychiatry 2019 Sep;64(9):647-657. [doi: 10.1177/0706743718818414] [Medline: 30602296]
106. Canuel M, Duhoux A, Brunet A, Lesage AD. Information sheet 11: Kessler Psychological Distress Scale – 6 items (K6) (section 6.2.4). In: Post-Disaster Mental Health Impacts Surveillance Toolkit: Professional Practice Guide: Institut national de santé publique du Québec; 2019. URL: https://www.inspq.qc.ca/en/post-disaster-mental-health-impacts-surveillance-toolkit
107. K10 and K6 scales. National Comorbidity Survey (NCS). 2005. URL: https://www.hcp.med.harvard.edu/ncs/k6_scales.php [accessed 2023-11-14]
108. Panesar P, Mayo SJ. “Taking out the trash”: strategies for preventing and managing fraudulent data in web-surveys. Can Oncol Nurs J 2023;33(2):283-284. [Medline: 37152829]
109. Giebel GD, Speckemeier C, Abels C, et al. Problems and barriers related to the use of digital health applications: scoping review. J Med Internet Res 2023;25:e43808. [doi: 10.2196/43808]
110. Côté J, Rouleau G, Ramirez-Garcia MP, Auger P, Thomas R, Leblanc J. Effectiveness of a web-based intervention to support medication adherence among people living with HIV: web-based randomized controlled trial. JMIR Public Health Surveill 2020 Apr 20;6(2):e17733. [doi: 10.2196/17733] [Medline: 32310145]
111. Thabane L, Ma J, Chu R, et al. A tutorial on pilot studies: the what, why and how. BMC Med Res Methodol 2010;10(1):20053272. [doi: 10.1186/1471-2288-10-1]
112. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 6;19(12):e402. [doi: 10.2196/jmir.8578] [Medline: 29212630]
113. Oakley-Girvan I, Yunis R, Longmire M, Ouillon JS. What works best to engage participants in mobile app interventions and e-Health: a scoping review. Telemed J E Health 2022 Jun;28(6):768-780. [doi: 10.1089/tmj.2021.0176] [Medline: 34637651]
114. Kelders SM, van Zyl LE, Ludden GDS. The concept and components of engagement in different domains applied to eHealth: a systematic scoping review. Front Psychol 2020;11:926. [doi: 10.3389/fpsyg.2020.00926] [Medline: 32536888]
115. Short CE, DeSmet A, Woods C, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018;20(11):e292. [doi: 10.2196/jmir.9397]
116. Bauermeister JA, Zimmerman MA, Johns MM, Glowacki P, Stoddard S, Volz E. Innovative recruitment using online networks: lessons learned from an online study of alcohol and other drug use utilizing a web-based, respondent-driven sampling (webRDS) strategy. J Stud Alcohol Drugs 2012 Sep;73(5):834-838. [doi: 10.15288/jsad.2012.73.834]
117. Helms YB, Hamdiui N, Kretzschmar MEE, et al. Applications and recruitment performance of web-based respondent-driven sampling: scoping review. J Med Internet Res 2021 Jan 15;23(1):e17564. [doi: 10.2196/17564] [Medline: 33448935]
118. Tatar O, Abdel-Baki A, Dyachenko A, et al. Evaluating preferences for online psychological interventions to decrease cannabis use in young adults with psychosis: an observational study. Psychiatry Res 2023 Aug;326:115276. [doi: 10.1016/j.psychres.2023.115276] [Medline: 37301021]
119. Albertella L, Gibson L, Rooke S, Norberg MM, Copeland J. A smartphone app intervention for adult cannabis users wanting to quit or reduce their use: a pilot evaluation. J Cannabis Res 2019 Aug 16;1(1):9. [doi: 10.1186/s42238-019-0009-6] [Medline: 33526112]
120. Kearney A, Butlin L, Coffey T, et al. The overlap between randomised evaluations of recruitment and retention interventions: an updated review of recruitment (Online Resource for Recruitment in Clinical triAls) and retention (Online Resource for Retention in Clinical triAls) literature. Clin Trials 2024 Oct;21(5):640-649. [doi: 10.1177/17407745241238444] [Medline: 38576071]

XSL•FO
RenderX

121. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283. [doi: 10.2196/20283]
122. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022 Apr 26;24(4):e35120. [doi: 10.2196/35120] [Medline: 35471414]
123. Ballard AM, Cardwell T, Young AM. Fraud detection protocol for web-based research among men who have sex with men: development and descriptive evaluation. JMIR Public Health Surveill 2019 Feb;5(1):e12344. [doi: 10.2196/12344]
124. Young AM, Ballard AM, Cooper HLF. Novel recruitment methods for research among young adults in rural areas who use opioids: cookouts, coupons, and community-based staff. Public Health Rep 2020 Nov;135(6):746-755. [doi: 10.1177/0033354920954796]
125. Bowen AM, Daniel CM, Williams ML, Baird GL. Identifying multiple submissions in internet research: preserving data integrity. AIDS Behav 2008 Nov;12(6):964-973. [doi: 10.1007/s10461-007-9352-2]
126. Frandsen M, Walters J, Ferguson SG. Exploring the viability of using online social media advertising as a recruitment method for smoking cessation clinical trials. Nicotine Tob Res 2014 Feb;16(2):247-251. [doi: 10.1093/ntr/ntt157] [Medline: 24127266]
127. Glazer JV, MacDonnell K, Frederick C, Ingersoll K, Ritterband LM. Liar! Liar! identifying eligibility fraud by applicants in digital health research. Internet Interv 2021 Sep;25:100401. [doi: 10.1016/j.invent.2021.100401] [Medline: 34094883]
128. Pozzar R, Hammer MJ, Underhill-Blazey M, et al. Threats of bots and other bad actors to data quality following research participant recruitment through social media: cross-sectional questionnaire. J Med Internet Res 2020 Oct 7;22(10):e23021. [doi: 10.2196/23021]
129. Teitcher JEF, Bockting WO, Bauermeister JA, Hoefer CJ, Miner MH, Klitzman RL. Detecting, preventing, and responding to “fraudsters” in internet research: ethics and tradeoffs. J Law Med Ethics 2015;43(1):116-133. [doi: 10.1111/jlme.12200] [Medline: 25846043]
130. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of eHealth interventions: systematic review of the literature. J Med Internet Res 2018 May 1;20(5):e10235. [doi: 10.2196/10235] [Medline: 29716883]
131. Chandler J, Sisso I, Shapiro D. Participant carelessness and fraud: consequences for clinical research and potential solutions. J Abnorm Psychol 2020 Jan;129(1):49-55. [doi: 10.1037/abn0000479] [Medline: 31868387]
132. Pullen Sansfaçon A, Gravel E, Gelly MA. Dealing with scam in online qualitative research: strategies and ethical considerations. Int J Qual Methods 2024 Jan;23:16094069231224610. [doi: 10.1177/16094069231224610]
133. Lawlor J, Thomas C, Guhin AT, et al. Suspicious and fraudulent online survey participation: Introducing the REAL framework. Methodol Innov 2021 Sep;14(3):20597991211050467. [doi: 10.1177/20597991211050467]
134. Bonett S, Lin W, Sexton Topper P, et al. Assessing and improving data integrity in web-based surveys: comparison of fraud detection systems in a COVID-19 study. JMIR Form Res 2024 Jan 12;8:e47091. [doi: 10.2196/47091] [Medline: 38214962]
135. Choi SH, Mitchell J, Lipkus I. Lessons learned from an online study with dual-smoker couples. Am J Health Behav 2017 Jan;41(1):61-66. [doi: 10.5993/ajhb.41.1.6] [Medline: 27935791]
136. Walukevich-Dienst K, Neighbors C, Buckner JD. Online personalized feedback intervention for cannabis-using college students reduces cannabis-related problems among women. Addict Behav 2019 Nov;98:106040. [doi: 10.1016/j.addbeh.2019.106040]
137. Greaves L, Hemsing N. Sex and gender interactions on the use and impact of recreational cannabis. Int J Environ Res Public Health 2020 Jan 14;17(2):509. [doi: 10.3390/ijerph17020509] [Medline: 31947505]
138. Choi S, Adams SM, Morse SA, MacMaster S. Gender differences in treatment retention among individuals with co-occurring substance abuse and mental health disorders. Subst Use Misuse 2015 Apr;50(5):653-663. [doi: 10.3109/10826084.2014.997828] [Medline: 25587672]
139. Sherman BJ, McRae-Clark AL, Baker NL, et al. Gender differences among treatment-seeking adults with cannabis use disorder: clinical profiles of a women and men enrolled in the achieving cannabis cessation-evaluating N-acetylcysteine treatment (ACCENT) study. Am J Addict 2017 Mar;26(2):136-144. [doi: 10.1111/ajad.12503] [Medline: 28152236]
140. Caldeira KM, Kasperski SJ, Sharma E, et al. College students rarely seek help despite serious substance use problems. J Subst Abuse Treat 2009 Dec;37(4):368-378. [doi: 10.1016/j.jsat.2009.04.005] [Medline: 19553064]
141. Buscemi J, Murphy JG, Martens MP, McDevitt-Murphy ME, Dennhardt AA, Skidmore JR. Help-seeking for alcohol-related problems in college students: correlates and preferred resources. Psychol Addict Behav 2010 Dec;24(4):571-580. [doi: 10.1037/a0021122]
142. Halladay JE, Boyle MH, Munn C, Jack SM, Georgiades K. Sex differences in the association between cannabis use and suicidal ideation and attempts, depression, and psychological distress among Canadians. Can J Psychiatry 2019 May;64(5):345-350. [doi: 10.1177/0706743718804542]
143. Rochat L, Rothen S, Edel Y, et al. Measurement invariance of the marijuana motives measure among men and women using stop cannabis app. Addict Behav 2024 Jan;148:107866. [doi: 10.1016/j.addbeh.2023.107866] [Medline: 37748226]
144. Chauchard E, Levin KH, Copersino ML, Heishman SJ, Gorelick DA. Motivations to quit cannabis use in an adult non-treatment sample: are they related to relapse? Addict Behav 2013 Sep;38(9):2422-2427. [doi: 10.1016/j.addbeh.2013.04.002] [Medline: 23685328]

XSL•FO
RenderX

145. Sherman BJ, Baker NL, McRae-Clark AL. Gender differences in cannabis use disorder treatment: change readiness and taking steps predict worse cannabis outcomes for women. Addict Behav 2016 Sep;60:197-202. [doi: 10.1016/j.addbeh.2016.04.014] [Medline: 27156221]
146. Grigsby TJ, Lopez A, Albers L, Rogers CJ, Forster M. A scoping review of risk and protective factors for negative cannabis use consequences. Subst Abuse 2023;17:11782218231166622. [doi: 10.1177/11782218231166622] [Medline: 37056398]

Abbreviations

CG: control group
CU: cannabis use
EG: experimental group
K6: Kessler Psychological Distress Scale–6 items
mHealth: mobile health
PBS: protective behavioral strategies
PBSM-17: 17-item Protective Behavioral Strategies for Marijuana
RCT: randomized controlled trial
SDS: Severity of Dependence Scale
TIDieR: Template for Intervention Description and Replication
UES-SF: User Engagement Scale-Short Form
uMARS: user version of the Mobile App Rating Scale

Edited by L Buis; submitted 31.Jan.2025; peer-reviewed by A Eisingerich, N Ribeiro, Y Cui; revised version received 05.Dec.2025; accepted 22.Apr.2026; published 28.May.2026.

Please cite as:
Côté J, Chicoine G, Auger P, Vinette B, Rouleau G, Maheu-Cadotte MA, Pagé MG, Lapierre J, Lal S, Genest C, Fontaine G, Cossette S, Cheng J, Jutras-Aswad D
Acceptability, Feasibility, and Outcome Responsiveness of the Joint Effort Mobile App for Promoting Lower-Risk Cannabis Use Among Young Adults: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e71957
URL: https://mhealth.jmir.org/2026/1/e71957
doi:10.2196/71957

© José Côté, Gabrielle Chicoine, Patricia Auger, Billy Vinette, Geneviève Rouleau, Marc-André Maheu-Cadotte, M Gabrielle Pagé, Judith Lapierre, Shalini Lal, Christine Genest, Guillaume Fontaine, Sylvie Cossette, Jinghui Cheng, Didier Jutras-Aswad. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 28.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Personal Health App and Wearable Co-Design Framework for Rare and Complex Diseases: User-Centered, Collaborative Co-Design Study

Sarah Margaret Goodday[1,2], MSc, PhD; Emma Karlin[2], MPH; Madeleine Sorensen[2], MPH; Robin Yang[2], MSc; Paul Gordon[2], MA; Caresse Opoku[2], MBS; Daniel Vuong[2], BS; Jenee Wilson[2], MBS; Diane McKenzie[2], MHA; Jules Piccotti[3]; Massimiliano Tavanti[3], BSc; Megan Fitzgerald[4], PhD; Yochai Re'em[4], MD; Hannah Wei[4], BS; Megan Golden[5], JD; Daniel Morgan[5], BBE, MEng; Michele Manion[6], BS; Ricardo Mosquera[7], MD; Tricha Shivas[8], MBE; Elise Hoover[8], MPH; Allison Peck[9], BS; Nathan Peck[9], MBA, MS; Zollie Yavarow[9], PhD; Heidi Bjornson-Pennell[10], JD; Andra Stratton[10], MS; Samali Anova Sahoo[11], BS; Douglas Arbittier[11], MD, MBA; Elizabeth Arbittier[11], VMD; Renee Goff[2]; Donna Harvin-Graham[2], MBA; Nancy Howard[8], MS; Camille Knudsen[9], BA, BS; Mary Oldham[2], BSc; Stephen Friend[1,2], MD, PhD

[1]Department of Psychiatry, University of Oxford, Oxford, United Kingdom

[2]4YouandMe, New York, NY, United States

[3]AgileOps, Siena, Italy

[4]Patient-Led Research Collaborative, Calabasis, DC, United States

[5]Mission: Cure, New York, NY, United States

[6]Primary Ciliary Dyskinesia (PCD) Foundation, Rochester, NY, United States

[7]Department of Pediatrics, The University of Texas Health Science Center at Houston, Houston, TX, United States

[8]Foundation for Sarcoidosis Research, Chicago, IL, United States

[9]Cure VCP Disease Inc., Warner Robins, GA, United States

[10]Biohub, San Francisco, CA, United States

[11]University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Sarah Margaret Goodday, MSc, PhD
4YouandMe
185 Great Neck Road
New York, NY, 11021
United States
Phone: 1 9147135614
Email: sarah@4youandme.org

Abstract

Background: End user co-design in the personal digital health technology space is underdeveloped. Clinical uptake of personal digital health technologies has been poor, highlighting a need to cocreate solutions with end users.

Objective: The study aimed to describe an “end user” co-design framework in the development of 5 prototype personal health apps for patients with different rare or complex diseases.

Methods: A patient-led, user-centered, collaborative personal health app plus wearable plug-in co-design methodology was developed. Five prototype apps were developed for end users with long COVID-19, pancreatitis, primary ciliary dyskinesia, sarcoidosis, and valosin-containing protein disease by a multidisciplinary partnership including patients, app design and development experts, user experience experts, clinicians, and patient-driven organizations. Phase 1 involved a 6-month co-design process with 5 modules involving patient-driven organizations that included the codevelopment of specifications through group workshops and independent exercises that defined the goals, content, features, and user experience of each app. Phase 2 involved app build-out, internal alpha testing, and beta study preparations. Phase 3 involved a usability beta testing study in which end users used the app and associated wearable/smart devices (Oura ring, Lumia ear device, Empatica EmbracePlus, and MIR Spirobank Spirometer)

XSL•FO
RenderX

for up to 5 months. Participant feedback was documented continuously and systematically, centering on the following themes: functionality, usability, harms, benefits, self-explorations, and beta testing study details related to retention and adherence.

Results: While unique app goals were codeveloped by each disease group, a central goal across groups was to develop a personal health app enabling users to track subjective, self-reported symptoms, objective measures of health, and unique modifiers of symptoms. A total of 239 end user participants participated in the beta testing pilot study. Enrollment and retention rates were high, ranging from 94% to 100% and 92.2% to 100%, respectively. All active participants gave some form of feedback: there were 257 unique participant suggestions of how to specifically modify or improve the study app experience. Participant feedback themes commonly centered around customization to reduce daily burden and improve personal tailoring of the app. Participants' desires surrounding symptom displays were heterogeneous.

Conclusions: Personal health app co-design is rooted in a complex digital landscape that requires a significant amount of up-front effort and time. However, the up-front investment of time can result in rich and diverse end user feedback that could save time in the app development trajectory to implementation. This paper provides a co-design framework and the building blocks of 5 prototype personal health apps with publicly available open-source code on GitHub. These prototypes could be leveraged for improving understanding of, communicating symptoms of, and providing n-of-1 suggestions for rare or complex diseases, providing benefit to patient communities and individual patients.

(JMIR Mhealth Uhealth 2026;14:e85425) doi:10.2196/85425

KEYWORDS

co-design; digital health; l ong COVID-19; mobile health; pancreatitis; primary ciliary dyskinesia; PCD; sarcoidosis; user-centered; valosin-containing protein; VCP disease; wearables

Introduction

Rare and novel conditions can present unique challenges for researchers, clinicians, and patients due to their low prevalence, phenotypic heterogeneity, and lack of natural history data. For rare and novel conditions, there is limited understanding of factors that trigger or modify symptom and disease progression. There is a need for patients to be better equipped with tools to follow their symptoms, along with insights about what improves and worsens their symptoms; for clinicians to be equipped with more accurate knowledge to help guide their patients; and for patient communities to better understand natural histories of their disease to guide research and clinical activities.

Personal digital health technologies (P-DHTs [1]) such as smartphone apps and wearable devices harness the potential for real-world, high-resolution, and longitudinal health monitoring at the individual context [2]. Consumer-grade wearable and smart devices are capable of tracking semicontinuous biometric measures (eg, heart rate, heart rate variability, body temperature, blood pressure, oxygen saturation, and respiratory function), behavior (eg, activity, relative location, and sleep duration and quality), and social activity (eg, social media and phone use) [2]. Smartphone active data can capture high-frequency, in-the-moment subjective assessments of symptom experiences, and the individual-level context of a unique patient's environment. Smartphone passive information, such as phone calls, text, and app usage, daily activities, and movement data, could reflect proxies of an individual's daily habits, levels of fatigue, and cognitive disturbance in addition to facets of disease burden and quality of life [3]. Collectively, P-DHTs could detect objective measures of key symptoms that are unique and common across different rare or complex diseases. This rich information can be returned to the user in near real time, and if returned in the correct way, could empower patients to glean new insights into their condition and its modifiers—a "compass" view.

The challenge in the collection of such multimodal objective and subjective information is how to effectively return this back to the user, or draw associations that are insightful [4]. Several consumer brand wearable devices enable users to follow specific measures of health via an associated app [5]. However, there is a lack of user-driven personalization and limited options to combine information from multiple P-DHTs among existing apps, limiting a true multimodal approach to better understanding health. Furthermore, existing apps are largely proprietary-based, come with a paywall, hide algorithms behind provided data and insights, and are not open source. These realities hinder progress to modify and tailor to the individual or population and limit the patient community's ability to benefit from collective insights [2]. An additional challenge is to inspire a user to engage with the app in the long term, so that health-related insights can be made. Most consumer apps target the achievement of fitness goals that are not appropriate for those with chronic illnesses [6-9]. Further, existing health-related apps show strikingly poor engagement [10-13]. Evidence suggests that this low engagement is in part the result of a lack of personalization and, in turn, low perceived value of the tool [1].

Co-design has become a nebulous concept with numerous definitions and applications. While widely used, its applications are underdeveloped, particularly in the P-DHT space [14]. Co-design, sometimes referred to as participatory design, is often described as a method of including end users in a design process to ensure their needs and voices are present to improve the impact of the intended outcomes. However, the level of participation and timing in the design process varies across definitions and applications. Collaborative design considers all participants as experts in their respective domains. User-centered design is an approach that centers on the needs and preferences of the end user through an iterative process in real-world situations during the entire design timeline to better understand when and where a particular health care tool or intervention

XSL•FO
RenderX

will work or fail [15]. Patient-centered design aims to empower patients as agents of their own health and emphasizes patient acceptability and usability [16,17], while patient-led design assigns patients the role of partners, as opposed to end users, or collaborators in the design process—another approach aimed at increasing agency in health. The latter two approaches have unique relevance in the context of P-DHTs that could empower patients to take greater control over their health, potentially without the guiding role of the clinician, shifting the power balance between doctor and patients [4]. There are several other approaches used that describe variants of the approaches listed above (eg, human-centered [18,19], collaborative, and person-based [20]).

In the past 5 years, a number of digital health co-design frameworks have surfaced that show promise in the cocreation of P-DHTs for a number of applications, including co-designing a smartphone app to improve access to mental health care [21], interventions for obesity [22], and support for people with dementia [23]. While these modern applications include powerful inclusion of end users from the inception of digital design, challenges remain. The complexity in both the development and use of a P-DHT requires a certain level of digital literacy that can result in roadblocks in empowering end users to co-design [23]. Further, there is still an under-involvement of end users in the final usability stages. The needs and realities of how an end user will want to engage with a personal health app can only be realized through the long-term use of a tool through different health states. Usability testing in current co-design applications is limited, and the definition of what is usable is currently vague.

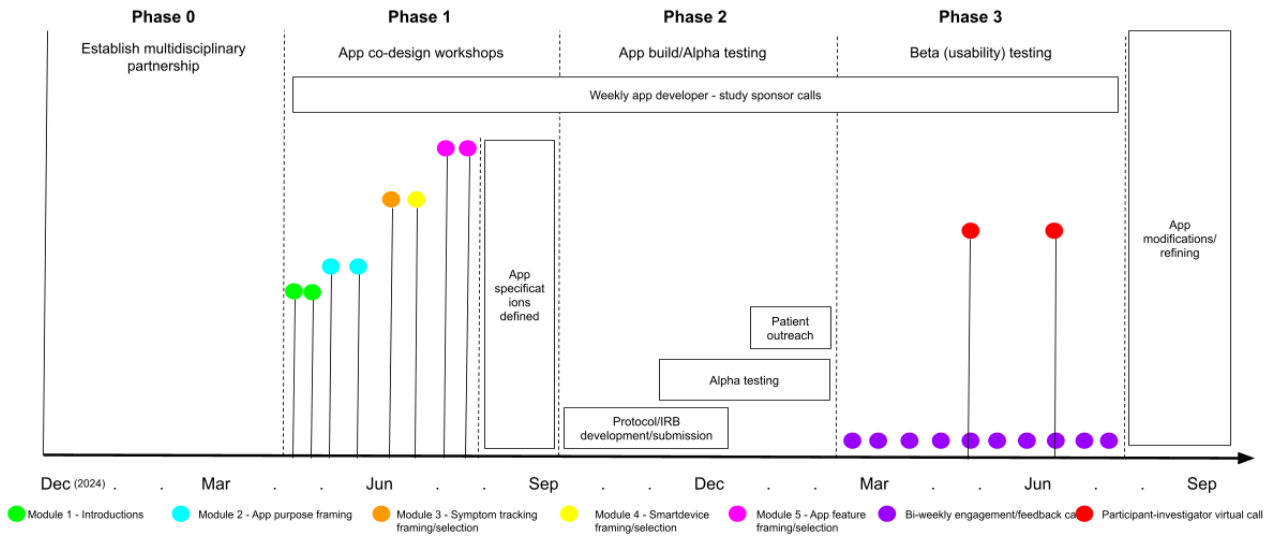
The objective of this paper is to describe an approach to a user-centered, collaborative co-design process in the development of 5 prototype apps for different end user populations, including individuals with long COVID-19, pancreatitis, sarcoidosis, valosin-containing protein (VCP) disease, and primary ciliary dyskinesia (PCD). These apps were codeveloped as part of the Biohub's Rare As One Project, which aimed to support co-design prototype symptom tracking and transmitting apps for rare or complex conditions. In this paper, we discuss lessons learned and provide a health app co-design framework for others hoping to develop a personal health app, including app development novices.

Methods

App Co-Design Synopsis

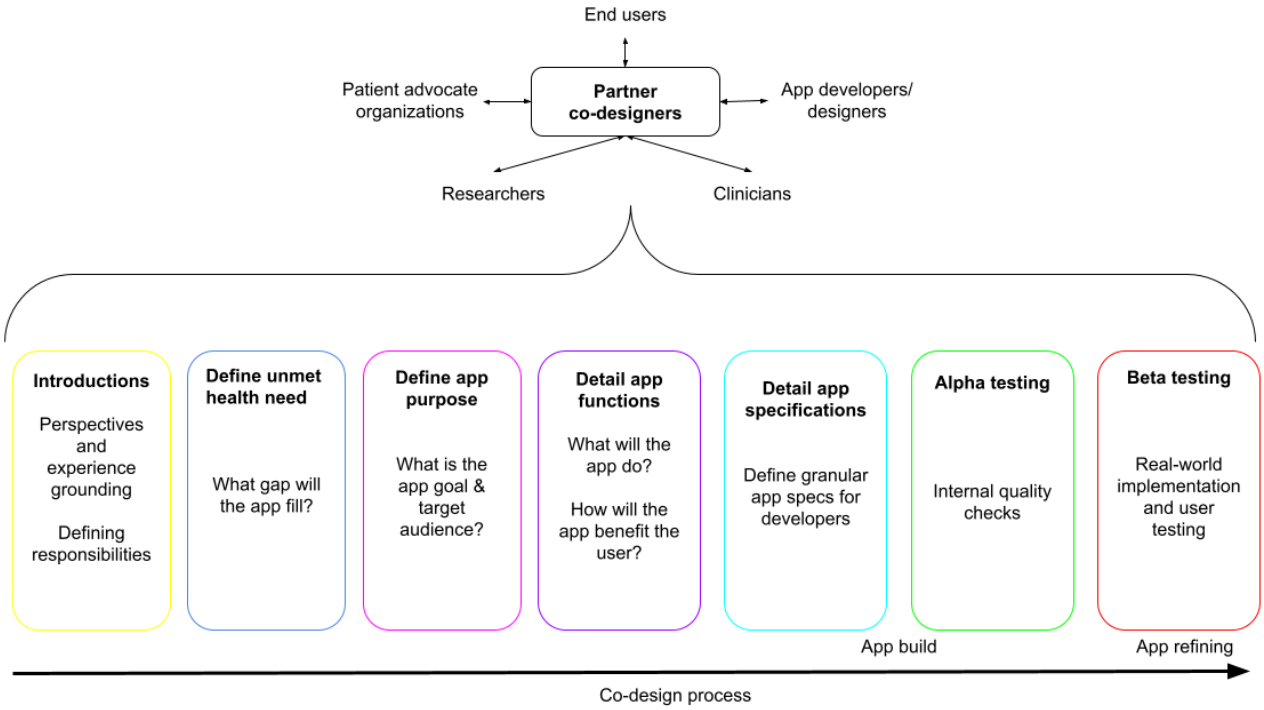
This project involved a multidisciplinary partnership between the sponsor, 4YouandMe [24] (a nonprofit digital health organization that includes expertise in app design, user experience [UX] and remote digital health app–based research), AgileOps (a software development company), and 5 different disease areas including a patient-driven organization (PDG) lead (patient advocacy groups), an individual with lived experience of the condition (herein referred to as end user), a clinician with condition-specific experience (selected by the PDG), device partners (Empatica, Lumia Health, MIR, and Oura), and Biohub (the study funder). All partners were involved in an extensive 6-month co-design phase (phase 1) that was conducted entirely remotely through frequent virtual meetings and workshops. This was followed by a study preparatory phase to build out the first versions of the prototype apps, prepare institutional review board (IRB) materials, and start patient outreach to identify end user testers (phase 2). Phase 3 involved a 5-month beta testing pilot longitudinal study of approximately 50 participants for each prototype app (Figure 1). The co-design methodology adopted and extended principles from many co-design concepts (user-centered, collaborative, and patient-led), where end users were placed at the center of the design process alongside other equal multidisciplinary partners(Figure 2). We adopted a "by each other for each other" approach where disease experts (end users, clinicians, and PDGs) were not expected to design themselves, but were empowered and provided with the tools to co-design alongside app development experts (app developers, UX experts, and digital health researchers). This approach resulted in significant investment of time from all partners involved. The PDGs received small grants to compensate for their invested time.

Figure 1. Co-design project methodology timeline.



XSL•FO
RenderX

Figure 2. Patient-led, user-centered, and collaborative personal digital health technology co-design.



Development of Multidisciplinary Team (Phase 0)

The first project phase involved the curation of the project team. 4YouandMe worked with Biohub to select 5 different PDGs. These groups were selected based on the following criteria: PDGs having motivation and capacity to participate in the patient-led design process, the ability to identify a clinician representative to participate in the co-design, and the ability to recruit 50 participants for the pilot study. The groups were additionally curated to include diverse disease areas, but with some overlap in symptoms, so that insights could be made from each group. The final selected groups included the following: Patient-led Research Collaborative (long COVID-19), Mission Cure (pancreatitis), Foundation for Sarcoidosis Research, Cure VCP Disease, Inc, and the Primary Ciliary Dyskinesia Foundation.

App Co-Design (Phase 1)

A 6-month app co-design phase was conducted involving the codevelopment of specifications through group workshops and independent exercises developed by the study sponsor that defined the goal, content, features, and UX of each app. This initial phase involved 5 modules (Table 1). The workshops provided an avenue to ground the app development experts in the experience and needs of patients with the disease, and the disease experts in aspects of the app development process. Each workshop included grounding virtual presentations by specific partners depending on the intended outcome, and all included an informal discussion period (Table 1). The co-design process was iterative and bidirectional, where partners were grounded on key concepts in the app development process and were asked to complete independent exercises (Multimedia Appendix 1) that were then reviewed in subsequent workshops to find common ground and build key stepping stones of the app development roadmap. The app development experts synchronously drafted app specification documents in collaboration with the disease experts, which were then used to communicate the final app specifications to the app builders. An example of an app specification document can be found in Multimedia Appendix 2. At each co-design step, PDGs added substantive considerations for end user needs that shaped the app specifications.

This app co-design was guided by core principles, including the following: (1) personal health apps must give back to the end user—what is the app's gift or benefit to the end user; (2) personal health apps must minimize the risk of instigating negative health outcomes; (3) personal health apps aimed at providing insights on specific diseases must be reflective of the unique needs of the patients in those communities; and (4) a personal health app's success cannot be fully measured without a significant amount of usability testing.

XSL•FO
RenderX

Table 1. App co-design module descriptions.

Module, purpose, and workshops (presenter)	Outcome
Module 1. Introductions	
Introduction between partners, description of partner goals, roles, and responsibilities	
Introduce approach, and app development process (sponsor) • What is the current state of the health app market? • What are health apps possibilities and limitations? • Co-design approach (“by each other for each other”)	• Educate disease experts about the app development process, realities
Introduce disease (PDGs[a]) • What is the disease epidemiology, burden? • What are the patient and clinician unmet needs, gaps in care • What is the PDGs past work/experience in digital health research app development?	• Educate app development experts about each disease, patient, and clinician perspectives. • Gauge all partners’ experience in app co-design
Module 2. App purpose/function framing	
Co-design primary app goals and key functionalities	
Defining the app purpose • Sponsor synthesis and discussion of PDG’s suggested patient and clinician needs • App guiding principles • Description of app possible attributes • Independent exercise 1-Preference of potential app purpose/attributes	• App goal and purpose framed and understood by all partners
Module 3. Symptom selection	
Co-design a list of core symptoms to be tracked and transmitted	
Describing subjective and objective symptom tracking through P-DHT[b] possibilities and realities (sponsor) • What symptoms are apps/wearables able to track well, not well? • Challenges in symptom tracking (confounding, uniqueness of individuals, linking objective measures to subjective symptoms) • Important of individual context • Introducing app-based visual tools for following symptoms, gaining health insights • Independent exercise 2-define specific app attributes	• Disease experts grounded on app/wearable realities in following specific symptoms and states of health
Review and discuss independent exercises 1 and 2 (all partners)	• Define app attributes and key symptoms to be tracked and transmitted
Module 4. Digital tool selection/dashboard	
Consider which on the market wearable or smart device could enable the objective tracking of the identified key symptoms, and introduce digital dashboards.	
Introduction to potential digital tool options (sponsor) • Grounding on current on the market wearable/smart devices realities and limitations • Sponsor digital tool suggestions based on PDGs disease/symptoms • Introduction to digital dashboard (ways to follow/share symptoms, experiences) • Introduction to triggers/modifiers (context labels) and how they can be captured through dashboards • Independent exercise 3-dashboard preferences	• Wearable/smart device selected • Disease experts grounded on digital dashboard purpose, realities, limits
Module 5. App feature/UX[c] introduction/app attribute refining	
Determine specific surveys/active tasks, or develop new tasks to enable the subjective tracking of the identified key symptoms, and introduce and refine app features and functionality, and UX.	
Introduction to potential app features and UX/app specification review (sponsor) • Introduce possible app features (beyond surveys/tasks) to capture symptoms and context • Brief introduction to app UX • Provide and review app specification documentation • Independent exercise 4-full app specification review	• App content (surveys/tasks) fully defined in the form of an app specification document
1 on 1 UX session to finalize app feel, and feature function	

XSL•FO
RenderX

Module, purpose, and workshops (presenter)	Outcome
1 on 1 with UX expert (UX expert) • Introduction to app UX • Sponsor synthesis and discussion of PDG's suggested features • Independent exercise 5-UX preferences	• App UX defined

[a]PDG: patient-driven organization.

[b]P-DHT: personal digital health technology.

[c]UX: user experience.

Study Preparatory Phase (Phase 2)

A 5-month study preparatory phase was conducted to prepare for phase 3 of the co-design (a pilot beta testing study with up to 50 end user participants for each prototype app). During this phase, the technology partner built the first version of each app, while in parallel, IRB materials (protocols, consent forms, and participant-facing materials) were codeveloped with all partners and submitted for approval, wearable/smart device companies were engaged, and agreements were completed. During the second half of this preparatory phase, internal alpha testing was conducted by the study partners, and patient outreach began to identify end user participants for each prototype app. Monthly one-on-one virtual meetings were conducted to give disease experts updates on app development and to co-design IRB documents. This period also included an alpha testing phase, where all partners performed an initial usability testing of the app to identify initial bugs before opening up the app to participants.

App Optimization—Beta Testing Pilot Study (Phase 3)

Overview

A 5-month pilot study was conducted with US-based end user participants to perform initial usability testing of the app. Study advertisements were co-designed with each PDG. Participants were recruited through each PDG organization using social media and newsletter advertisements (see Multimedia Appendix 3 for an example) with a cover letter from each PDG introducing the pilot study. Participants were instructed to contact a 4YouandMe study representative and complete an initial screening phone call where they were explained the study, their eligibility was confirmed, and they were given instructions on how to download the app. Briefly, participants were eligible if they were 18 years or older (except for the PCD beta testers that included a minor version of the app, inviting 14-17 year olds), had a suspected or clinician-confirmed diagnosis of each disease, and owned a personal iPhone/Android smartphone (see Multimedia Appendix 4 for full inclusion criteria). The app included an e-consent where participants officially agreed to participate, including a parental authorization mechanism for participating minors (PCD app). Post enrollment, participants were mailed their cohort-specific wearable/smart devices. Participants completed participation in an entirely remote setting.

Beta Pilot Study Procedures

Enrolled participants were instructed to download the prototype app, provisioned with specific wearable and/or smart devices (Multimedia Appendix 5), and instructed to use these devices continuously during their participation. Participants were asked to complete biweekly phone check-in calls with a dedicated engagement specialist. These calls aimed to support participants, work through technological issues, and collect their feedback as beta testers. Engagement specialists entered participant feedback into structured and unstructured fields in Research Electronic Data Capture—a secure data management database. Participant feedback was documented continuously and systematically by engagement specialists, centering on the following themes: functionality, usability, harms, benefits, self-explorations, and others (Textbox 1). Participants were also invited to participate in 2 optional virtual investigator-participant meetings where they had a chance to meet the study team, get study updates, and ask the study team questions in real time, anonymously.

XSL•FO
RenderX

Textbox 1. Beta testing app themes collated from participants.

Functionality

- Are the study tools (study app, wearables, and smart devices) functioning the way they should based on their intended use—to enable the user to track diverse symptoms and modifiers of their condition and potentially provide insight into factors that impact symptoms and outcomes, or general health insight relating to their condition. Are there factors about the participants' conditions that are not being tracked that should be? Are there missing functionalities that could help achieve the study tools goals?

Usability

- Are the study tools usable? Are there internal or external factors that prevent the participant from using the tools for their intended use (eg, physical barriers such as mobility challenges, mental barriers such as fatigue or cognitive disturbances, and technological barriers such as WiFi connectivity)?

Benefits

- Are there reported benefits from using the study tools? Do participants perceive value in using the study tools? Benefits might include factors such as improved health insight and awareness, or positive behavior modification.

Harms

- Are there reported harms from using the study tools? Harms might include factors such as increased anxiety or other mental health concerns from tracking certain symptoms or being asked certain questions at high frequencies, irritability from task burden, or injuries from using the wearable/smart devices.

Self-explorations

- How are participants using the app to learn about themselves and conduct self-explorations? How specifically are they using the compass feature? For example, are they using this feature to track specific symptoms/features, to track specific modifiers of symptoms (eg, diet), or to try to learn associations between factors that impact symptoms?

Ethical Considerations

This study was approved by Advarra IRB (Pr00082545) and all participants digitally provided informed consent and parental authorization (where appropriate). All personal data were kept confidential by the internal research team. The study data collected were stored in deidentified form (stripped of personal identifiers) and attached to a unique identifier. Participants received US $50 for every month of participation in the form of a gift card and were also offered to keep the Oura ring at the end of the study.

Prioritizing Participant Feedback

Participants communicated their feedback to their engagement specialists through text, email, video calls, and, where appropriate, sent screenshots. Participant feedback was documented systematically by engagement specialists, which was then collated by the app development team and reviewed with PDGs for prioritization. This project had time and funding constraints, meaning that not all participant suggestions could be implemented. Suggestions were condensed from duplicates across participants, and by setting aside those that were out of scope (ie, could not be achieved within the current time frame, or within the current budget). PDGs were asked to rate each feedback from least desired, desired, to most desired. The app development team worked to determine which of the desired and most desirable suggestions could feasibly be implemented within the allotted time frame and budget. This synthesis was conducted in 2 waves during the beta study, with the goal of having time to implement most suggestions so that the participants could view and test themselves. Suggestions that were not possible to implement were documented and are made available to others wishing to view (see Data Availability section).

App Development Process

App development was undertaken by the technology partner (AgileOps) using an iterative, Agile workflow aligned to the multiphase co-design program. Requirements were derived from phase 1 specification documents and UX exercises, translated into shared low- and high-fidelity Figma prototypes for rapid validation with the project team and ultimately all the stakeholders involved in the co-design exercise, and then implemented in short, time-boxed iterative development sprints (recurring unit of time by which the development process is organized) of an approximate length of 2 weeks each with traceability to the originating workshop artifacts. Every app change was mapped to a requirement or user story traceable to co-design artifacts (workshop outcomes, spec docs, and Figma frames). Weekly product reviews with PDGs and the sponsor confirmed scope, prioritized participant-reported issues, and aligned releases with IRB amendment needs. All 5 study apps were developed natively for iOS and Android and distributed via the Apple App Store and Google Play Store. The supporting backend services were deployed on cloud infrastructure to enable secure survey delivery and wearable data acquisition. Continuous integration and delivery followed trunk-based development with mandatory code review, automated linting and tests, and staged promotion (development → staging → production). Mobile releases used internal/beta channels for rapid usability checks prior to public availability during phase 3. A server-driven "study engine" delivered survey schemas, schedules, notifications, and visualization configurations by remote configuration, allowing same-day adjustments while

XSL•FO
RenderX

preserving data integrity through schema versioning. This engine enabled real-time incorporation of participant feedback during active use.

In terms of accessibility and performance, frontends follow web content accessibility guidelines–informed patterns (contrast, type scaling, and motion reduction). Mobile apps are optimized for low-bandwidth scenarios with appropriate error and conflict handling. The Compass Chart experience specifically used progressive data loading and elastic queries to keep visualizations responsive even as longitudinal data grows.

Quality and safety practices included a testing pyramid (unit, integration, and end-to-end), synthetic monitors of critical user journeys (enrollment, consent, survey completion, and device synchronization), centralized observability (logs/metrics/traces), and auditing of administrative and export actions. All network communication used transport layer security; storage and secrets followed cloud key-management and least-privilege access controls. These measures supported the usability study while meeting governance needs associated with IRB-approved procedures.

Architectural Stack and Methodology

A key element that facilitated the continuous development process was the methodological and technological stack adopted for the project, which was intentionally fine-tuned for quick prototyping, iterative development, and continuous release without compromising the quality and security of the output. This was achieved by combining modern development frameworks (specifically Ruby on Rails application layer exposing versioned application programming interfaces), web consoles for research/clinical use, and native iOS/Android apps for participants to support robust device integrations and accessibility features, all backed by a cloud-first infrastructure containerized and orchestrated on Amazon Elastic Kubernetes Service (Kubernetes) to support isolated workloads, horizontal scaling, and controlled rollouts. Data from surveys, active tasks, and connected devices were normalized through event-driven extract, transform, load pipelines and indexed in Elasticsearch to enable longitudinal visualization and query of symptoms and objective signals.

Analysis

Phase 1 of this co-design involved building out app specification documents. All of the resulting app prototype themes/attributes and outcomes are described in tabular format. Participant insight and feedback from phase 3 were collected through regular phone discussions every 2 weeks with study engagement specialists, in addition to an exit interview after completing the study. A thematic analysis [25,26] of qualitative data captured during these interactions was conducted to identify themes for app improvement. Participant engagement was measured through calculating retention rates (number of dropouts/total enrolled) and median adherence to study activities over study follow-up. Adherence to app-related tasks was defined as the proportion of tasks completed divided by the total number of tasks that could have been completed over each participant's unique study follow-up. For example, all apps included a daily quick activity. For a participant with a follow-up of 4 months, the total expected number of daily quick activities is 120. If a participant completed 95 daily quick activities, their adherence to this specific task would be 79%. Adherence to the check-in calls every 2 weeks was defined as the number of calls completed divided by the total number of expected calls over a unique participant's follow-up period. Medians and ranges were used as adherence distributions were nonnormally distributed. Adherence to the wearable/smart devices was estimated using different daily wear criteria. For Oura ring data, daily wear was defined as the presence of an "Oura Sleep Score." For the Lumia device, daily wear was defined as the presence of at least one "heart rate" value. For the Empatica device, daily wear was defined as the presence of at least one "electrodermal activity (EDA)" value. For the MIR Spirometer, daily wear was defined as the recording of any spirometry data.

Results

Co-Design Outcomes

Multimedia Appendix 6 outlines the codeveloped app goals and specifications for each prototype app. While there was overlap in unique goals, themes, and key symptoms to track across the prototype apps, each app included unique content. Each group had an interest in building longitudinal symptom maps and having the ability to share symptoms with clinicians. Further, all groups had an interest in enabling the user to follow aspects of burden of disease, routines, and potential triggers that modify symptoms or disease outcomes. This stage in co-design was critical for deciding on the prototype app's key functionalities and features to ensure that the tool could achieve the common and specific objectives. While most groups included a target age range of >18 years, given the mostly adult onset of symptoms, the PCD app included a minor app version (aged 14-17 years) in light of the presence of symptoms earlier in life.

Select wearable/smart devices were selected for each group to capture objective measures of key symptoms (section 4B in Multimedia Appendix 6). All groups used the Oura ring as a measure of daytime activity and sleep quality. Groups that had a specific interest in following the symptom of pain additionally included the Empatica EmbracePlus smart wristband device, owing to its ability to track EDA (skin conductance), which could reflect an objective proxy of pain. Additionally, the long COVID-19 group had an interest in following dysautonomia (dysregulation in the autonomic nervous system), for which EDA could be a useful measure. The Lumia ear device was additionally used in the long COVID-19 group owing to its unique ability to capture blood flow to the head as another useful potential measure of dysautonomia. Groups that had an interest in following symptoms relating to breathing, or respiratory function, additionally included a spirometer (MIR Spirobank Spirometer). Please see Multimedia Appendix 5 for a detailed description of selected wearables and smart devices. Taken together, other than design and targeted symptoms (which defined the app survey/task content and wearable/smart device selection), the apps were similar in design with the same core features.

XSL•FO
RenderX

Prototype Apps and Their Features

Specific app features were developed to accommodate the intended app themes and goals. All prototype apps included daily and intermittent prompted surveys and active tasks to collect subjective self-reported information from participants that were intended to capture each symptom in order to build longitudinal symptom maps across conditions (Figure 3).

In keeping with many group themes of enabling better abilities to track and understand certain symptoms and have the ability to better understand triggers and modifiers of symptoms and disease, additional optional app features were co-designed. These included the "I've Noticed..." feature, the "Reflections" feature, the Compass Chart, and Compass Log.

The "I've Noticed..." feature (section 5A in Multimedia Appendix 7) enables the end user to tag momentary experiences at any time in the app in either a written text format, or audio or video recording. The logged experiences from the I've Noticed feature are then time-stamped and logged in an additional diary format feature called "Compass Log" (section 5B in Multimedia Appendix 7), where the end user can return to these logged experiences and even edit them. Participants were instructed to try to use this feature to log in-the-moment potential triggers/modifiers that could help explain fluctuations in their symptoms, or for logging dynamic changes in their symptoms. For example, for the VCP group, this feature was useful for logging falls and near falls when they occurred. This feature was developed to enable a user to tag momentary context in their day or night that could lend insight into potential triggers or modifiers of their condition.

The "Reflections" feature (section 5C in Multimedia Appendix 7), while similar in format to the "I've Noticed..." feature, served a different purpose. The "Reflections" feature was designed to empower the user to have a space within the app to take the time to consider certain aspects of their condition and reflect on certain aspects that improve or deteriorate their condition. This feature prompted participants daily to complete a reflection where they could write a written note, or complete an audio or video recording. Prior to completing a reflection, this feature instructed participants to complete a grounding exercise, such as meditation, going for a walk, etc, to help prepare themselves to make a reflection about their day or their week in relation to their condition. The logged reflections were made available in the end users' Compass Log, where they could revisit their logged entries. This feature was developed to empower the end user to reflect on a deeper level of possible triggers and modifiers of their condition.

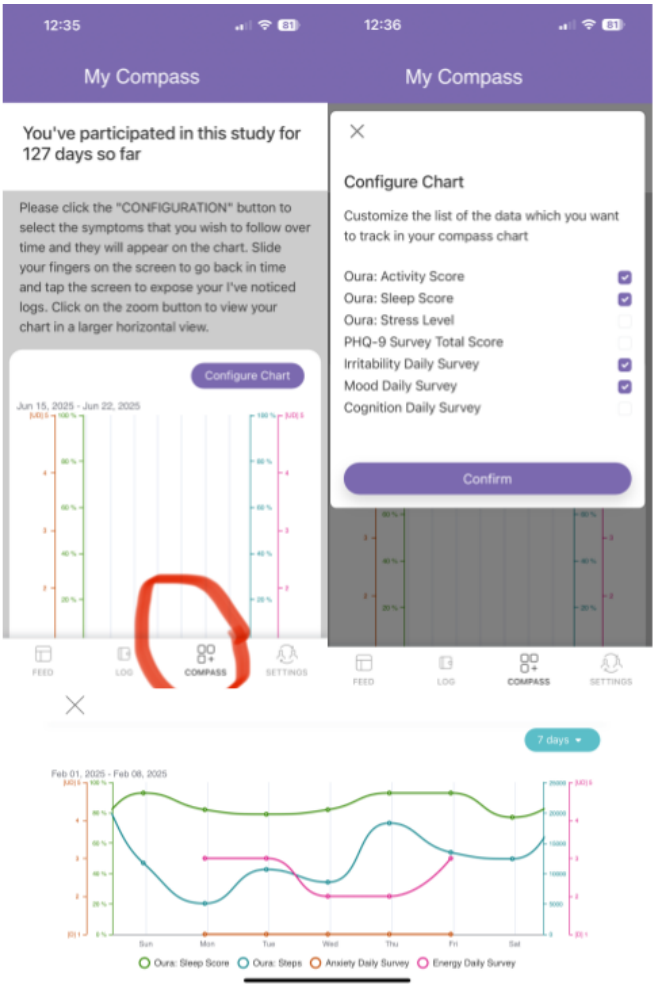
The Compass Chart (Figure 4) feature enabled participants to view select symptoms or wearable/smart data streams retrospectively over time. Participants could select from a dropdown menu of up to 10 items to track and were able to view 4 of these items in their chart at a time. This feature was designed to provide the user with a tool to follow, and potentially learn new insights about their condition in terms of symptom patterns and factors that improved or worsened their condition.

Figure 3. The prototype apps.



XSL•FO
RenderX

Figure 4. My Compass Chart feature.



Beta Testing Pilot Study Results

Overview

A total of 239 end user participants participated in the beta testing pilot study that was conducted between February 2025 and June 2025. See Multimedia Appendix 8 for specific enrollment flow details of each group. Enrollment and retention rates were high, ranging from 94% to 100% and 92.2% to 100%, respectively (Table 2). In most disease groups, exit interview completion was high (long COVID-19: 47/50; pancreatitis: 35/44, 80%; sarcoidosis: 44/50, 80%; VCP: 42/44, 95.4%) except for the PCD group (23/51, 45.1%), who were mostly minors. Demographic characteristics can be found in Multimedia Appendix 9. In all but the sarcoidosis cohort, beta testing participants were predominantly White and female. Mean age was in the middle age category (40-55 years) in all groups except for the PCD cohort, which included minors, with a mean age of 31 (SD 15) years.

XSL•FO
RenderX

Table 2. Median adherence rates for study activities across disease groups.

	Long COVID-19 (n=49)	Pancreatitis (n=44)	PCD[a] (n=51)	Sarcoidosis (n=50)	VCP[b] disease (n=44)
Study weeks, mean (SD)	14.2 (3.6)	14.5 (3.2)	11.9 (3.5)	18.0 (0.6)	13.7 (3.1)
Retention, n (%)	49 (100)	42 (95.5)	47 (92.2)	47 (94.0)	44 (100)
Biweekly check-in calls (%), median (IQR)	100 (100-100)	85.7 (61.7-100)	71.4 (46.4-92.9)	85.7 (71.4-100)	100 (100-100)
Daily quick activities (%), median (IQR)	36.8 (22.1-60.3)	39.9 (23.5-68.2)	39.8 (24.2-55.4)	51.2 (25.1-67.3)	43.1 (29.7-64.2)
Weekly in-app survey (%), median (IQR)	90.9[c] (73.3-100)	97.2[c] (89.2-100)	100[c] (80.2-100)	—[d]	72.4[e] (54.2-80.0)
Biweekly in-app survey (%), median (IQR)	100[f] (75-100)	100[f] (88.9-100)	100[f] (79.2-100)	100[f] (89.5-100)	100[f] (100-100)
Oura ring use (%), median (IQR)	89.3 (71.4-95.0)	67.1 (46.2-92.5)	26.3 (15.5-50.4)	66.4 (52.6-82.1)	83.9 (65.1-94.1)
Empatica Embrace Plus use (%), median (IQR)	69.9 (39.4-98.2)	69.7 (29.9-93.1)	—	—	—
MIR Spirometer use (%), median (IQR)	—	—	50.0 (33.3-62.5)	—	62.5 (47.2-81.7)
Lumia use (%), median (IQR)	62.5 (33.7-80.0)	—	—	—	—

[a]PCD: primary ciliary dyskinesia.

[b]VCP: valosin-containing protein.

[c]Weekly Fatigue Assessment Scale.

[d]Not available.

[e]Weekly Pain Severity Scale.

[f]9-item Patient Health Questionnaire.

Participant Adherence

Table 2 includes median adherence to study activities by disease group. With the exception of the PCD group, all groups demonstrated excellent adherence to participating in biweekly check-in calls with research staff and to weekly and biweekly in-app surveys. Adherence to using the study smart devices demonstrated acceptability, with participants using the Lumia ear device >60% of the time, Oura ring use >65% of the time (except for the PCD group), Empatica use >55% of the time, and MIR Spirometer use >50% of the time. All devices except the Oura ring had a delayed integration into the study, meaning that the time that participants had to interact with these devices was shorter than the entire study duration. The PCD group was the only group that included minor participants (aged 14-17 years). This disease group demonstrated consistently lower adherence to most study activities.

As this was a pilot beta testing study, the app was undergoing modifications, and a number of active tasks and features had periods of malfunction. Accordingly, the adherence estimates relating to the app tasks should be interpreted with caution, as some of the low adherence may be attributable to app error as opposed to noncompliance.

Participant Beta Tester Feedback

During the participant beta testing phase, 215 unique participant reports of prototype app issues were documented, while 112 unique participant reports of wearable/smart device issues were reported. There were 257 unique participant suggestions of how to modify or improve the study app experience. User feedback was coded into the following core themes: 81 functionality, 51 usability, 28 benefits, 20 harms, 71 self-explorations, and 6 others. Several features were implemented during the beta testing period, and some shortly after the study closed, based on participant feedback. For example, new surveys or new items to existing surveys were included, screening questions, and features to instruct participants not to complete tasks when they cannot safely be completed (eg, walk tests) to ease user burden, color contrast improvements, including a dark mode option for users with visual impairments, labeling features with emojis to improve ease and personalized “I’ve noticed”/”reflection” entries. See Multimedia Appendix 10 for all features implemented.

Over 90% (175/191) of participants across all disease groups (who completed an exit interview) reported that they received some benefit from using the prototype apps and associated wearable/smart devices (long COVID-19: 42/47, 89.4%; pancreatitis: 34/35, 97.1%; sarcoidosis: 41/44, 83.2%; VCP: 38/42, 90.5%; PCD: 20/23, 87.0%). Overall, 71.2% (136/191) of all participants rated their likelihood to use their respective app in the future between 7-10, on a 1-10 rating scale, with 10 being the most likely to use the app (Multimedia Appendix 11). Participants reported benefits of the prototype apps and associated wearable/smart devices, including increased self-awareness of symptom modifiers (eg, sleep and stress patterns that trigger exacerbations) and behavior modification through the encouragement of health-promoting behaviors. Participants indicated an appreciation for the comprehensive symptom tracking that empowered them to feel accountable for making healthier choices and the desire to share their health data with others. Participants also reported that they felt supported in participating in the beta testing study. Frequent check-in calls with engagement specialists and intermittent virtual progress meetings with the investigator team were viewed positively by participants.

XSL•FO
RenderX

While benefit was reported by the majority of participants, a number of subthemes emerged during the beta testing usability period (Multimedia Appendix 12). There was a desire to improve the functionality of the apps through better personalized notifications when tasks are due, better task instructions, and better app navigation through the app's core features. There was some overlap in feedback between the core themes of functionality and usability, where, for example, more personal customization was a desire, which alters initial functionality, but would likely lead to better usability. Numerous participants had different opinions on how frequently they wanted to complete certain surveys. For example, all prototype apps included a daily quick activity survey that prompted the user to report on their mood, energy, cognition, pain, and other symptoms. A number of participants suggested they track these symptoms multiple times a day to capture daily variations in fluctuating symptoms, while others preferred a longer cadence (eg, weekly), and some participants suggested tracking additional symptoms, not present in the quick activity. A solution would be to enable participants to set their own cadence and even the content of some surveys. However, in making these decisions, one must weigh the consequences of enabling too much customization where data cannot be harmonized across users for health research, or returned data are no longer useful for the individual. There were unique usability preferences from the long COVID-19 group, many of whom reported vision problems. These reports included enabling a dark mode option or better color contrast.

There were 20/239 (8.4%) unique participant reports of nonserious harms associated with using the prototype apps and associated wearable/smart devices (10 long COVID-19, 5 pancreatitis, and 5 VCP). Most of these were related to frustration relating to unclear instructions on certain tasks and having to complete tasks that did not feel relevant to the user. Many of the prototype apps used select ResearchKit's active tasks (eg, 2-minute walk task). Many of these "plug-in" tasks are not customizable, which introduces challenges to include adequate instructions to tailor these tasks for each end user population. There were also some reports of burden and fatigue from too many tasks, particularly when symptoms were static. A small number of participants reported a negative emotional impact from using the prototype apps and associated wearable/devices. For example, there was a report that the use of the Oura ring and returned summary scores made a participant feel worse about their health (VCP participant). For some, persistent, daily tasks felt like reminders of disease progression and were emotionally draining.

In addition to these reported harms, there were several reports of skin irritation issues associated with wearing the Empatica EmbracePlus wristband at the site of wear that appeared to be mostly the result of contact dermatitis. This was investigated with the device company, and Empatica has reported 0.1% skin irritation rate for EmbracePlus, according to tests performed following the ISO 10993 standard series. Both of the populations using this device may experience heightened autonomic activity from pain (pancreatitis) and from other unknown causes (long COVID-19) that could have contributed to a higher number of reported irritations than typical for this device. Participants were instructed to stop wearing the device if skin irritation occurred.

Participants were encouraged to use the prototype app Compass feature to log experiences and symptom modifiers, and use the Compass Chart to track certain symptoms over time. This feature also overlaid select feature streams from some of the wearable/smart devices. The goal of this feature was to enable the end user to start to self-explore their condition, and potentially start to make some insights into what makes them better or worse. There was again overlap in feedback themes, with this core theme relating to the desire for customization in what participants were enabled to track via the app, and how this was transmitted back to them via the app. Participants' desires for what and how to transmit via the app were heterogeneous. However, there was a general consensus that personalized labeling (eg, using icons and color coding) within the Compass Chart would be useful, and ways to export and share data with others for a deeper review of personal health data.

PDG Co-Designer Experience

The participating PDGs reported positive experiences relating to being able to shape an app, learning from other PDGs and from an app development team, exploring the use of P-DHT data that is important to patients, and providing patient communities with the opportunity to pilot test an app for them.

Negative experiences were reported relating to mismatches in project objectives. Some of the PDGs anticipated that the data could be used to inform insights about each condition or the use of the tools for following certain symptoms—there was an expectation that direct benefit would be translated to the participating communities. This highlights the critical importance of repeating and discussing project objectives throughout the entire co-design process. To effectively co-design and beta test a prototype app, end users must use and interact with the app, which means data is produced. However, with dynamic changes occurring within the app, and the resolving of early app bugs that can result in significant gaps in data, the data produced from such early prototype stages is not advisable to be used for making health-related insights that one would make from an observational study.

PDG groups were heavily involved in the first 2 project phases, but were less involved in phase 3—the beta testing study, which was more focused on end user involvement. There was a general desire from PDGs to be more involved in the beta testing study to relay updates back to their community, and to better understand insights from the data in near real time.

Some PDGs reported wanting more one-on-one time with the development team. A number of group workshops were held, but all co-designers felt that more time was needed to work through questions and complex ideas.

App Development Team Co-Designer Experience

The app development team found benefit in the co-design approach in producing a prototype app that could have taken multiple usability studies to achieve. The initial PDG input on app goal and specifications was instrumental in yielding an

XSL•FO
RenderX

early prototype that already had elements of disease-specific personalization. This allowed the beta testing period to focus on more nuanced features of the app as opposed to common user feedback on relevance. In the engagement of ~50 end user beta testers for months, the sheer volume of detailed app improvement input was overwhelming, much more than the app development team has experienced from past P-DHT studies. Further, while the app development team has reported high retention and adherence rates from past P-DHT studies [1], the observed enrollment and retention rates of end users in this pilot study were exceptionally high.

The app development team underestimated the time and resource burden of the co-design approach, especially in the development of 5 different apps. There was an expectation that more overlap in needs across the disease groups would be present; however, the initial co-design phase highlighted each group's unique needs, which surpassed the app development team's initial expectations.

Discussion

Overview

In this paper, we describe a patient-led, user-centered, and collaborative personal health app plus wearable plug-in co-design methodology. Five prototype apps were developed for end users with long COVID-19, pancreatitis, sarcoidosis, PCD, and VCP disease. End users were placed at the center of the design process alongside other multidisciplinary partners driven by nonprofits and community-based organizations, including app design and development experts, UX experts, clinicians, and PDGs. Using a "by each other for each other" approach, this co-design process included facilitators to guide the co-design process for disease experts; however, all partners were considered experts in the process in their own respective roles. After the development of the prototype apps, this approach included a 3- to 5-month usability testing period with 239 end users across different rare or complex conditions. A flexible study engine was built that facilitated real-time and frequent app modifications while the app was in active use. This flexible study engine accelerated and empowered the co-design process.

While this co-design methodology required significant effort and time from all partners involved, it yielded deep insights into end user driven prototypes. Beta tester participants demonstrated exceptionally high enrollment rates, retention, and acceptable adherence to using the apps and devices for several months during the usability testing period. A total of 16 of 191 participants actively engaged in providing feedback on biweekly check-in calls, exit interviews, and virtual calls, while 327 unique P-DHT bugs and 257 unique suggestions to improve the app were reported. Collectively, participant suggestions resulted in numerous app fixes and improvements, many of which were implemented while participants were actively using the app, enabling them to see the result of their feedback. Further, pairing participants with an engagement specialist offered a mechanism to communicate complicated app/wearable-related concepts, which would not be possible without these frequent touch points.

While the impact of the co-design approach cannot be formally evaluated from this study, as there was no comparison, the app development team reported a more streamlined process in producing an early personal health app prototype that already included disease-specific personalization. This enables a deeper beta testing focus on more nuanced features of the app. Through both the engagement of the end user participants from the PDGs and the results of the initial co-design stage, the volume of app improvement suggestions was much more than the team had experienced from past P-DHTs studies.

A common participant feedback theme centered around customization of what the app tracked, how often it tracked, and how it transmitted this information back. Further, participants expressed a desire for more instructions and tailoring of certain tasks so that they were relevant to them and their situation. For example, in the VCP and long COVID-19 end user groups, some users experienced phases where they could not walk safely and expressed a desire to temporarily turn off certain app activities (eg, the walk test) so that they were not repeatedly prompted to complete them. A challenge for the co-design team in implementing participant feedback is balancing this feedback with the personal health apps' intended goals. If the app is purely for the participant, then the co-designer may increase the level of customization. However, if information from the app is intended to be shared with others (eg, clinicians or researchers for analysis), the co-designer has to balance the level of customization with the level of data harmonization that is required to meet the app's goals. This highlights the continued importance of repetitive grounding of the P-DHT intended goal and use for all co-designers involved.

The information collected from P-DHTs holds promise in providing clinicians with richer, higher-resolution data of their patients' health in between health care visits and new knowledge for PDG research. However, a key promise is in empowering patients with the tools to follow their health, and gain insight about what makes them better or worse, in their own unique context, without the sole reliance on health care visits [2,4]. This could be enabled by augmenting P-DHTs with artificial intelligence (AI) tools that give n-of-1 suggestions for improved health [27-29]. AI-driven insights from P-DHT n-of-1 data could yield new insights into rare or complex disease symptom trajectories, and modifiers of disease exacerbations that account for individual-level variability. However, challenges exist in enabling AI suggestion engines to work effectively. This suggestion engine requires adequate "label" context data from the user, which is notoriously difficult to capture, as it often requires a high burden from the user, and much user validation work is needed to determine the accuracy of these tools' suggestions.

There are additional challenges in the general use of P-DHTs. P-DHTs involve a certain level of digital literacy and usability to be feasible for use in the long term and integrated into patients' daily lives. Further, there are disconnects in values and approaches to bring a P-DHT from design to implementation, resulting from multiple stakeholders between health care and industry that often leave the end user at the outskirts of the process [14]. Finally, the existing health app market is saturated by companies with pay walls, and that hold

XSL•FO
RenderX

data in silos that can result in a health data market that takes advantage of patients and their sensitive data. Taken together, P-DHTs hold promise to increase patient agency, putting them in the driver's seat of their health. However, for this promise to become a reality, patients and community organizations must first be empowered and given the time to cocreate these tools for themselves, backed by the appropriate companies.

A number of robust P-DHT patient-centered and patient-led co-design approaches have emerged in the past 5-10 years [21-23,30,31]. A common challenge across these instances is how to adequately ground an end user (especially in a remote co-design environment) so that they are empowered to co-design along with other partners with expertise in app design/development, and digital health research [23]. Further, most existing frameworks do not extend the co-design period into the usability testing period and tend to use small user testing groups for short periods of time [12]. The value and shortcomings of a personal health app cannot be realized until significant usage, which is a key shortcoming of existing P-DHTs on the market and is likely in part the result of poor engagement after the first few weeks of use [11,13]. Accordingly, as opposed to using a mobile app rating scale as a fixed metric of user acceptability and usability, which is commonly applied in other app co-design approaches, we implemented a dynamic feedback system that incorporated user feedback over several months.

Limitations of Approach

The personal health app co-design approach implemented here yielded several lessons learned. While 6 months was allotted for the initial co-design to build out the first app prototypes that included 8 workshops and several break-out meetings, more time was needed to ground and work through complex processes. Some workshops could have been split into 2 or 3 additional workshops. Further, there were occasional mismatches in co-design objectives throughout the project period. For example, the objective of this co-design was to produce prototype apps and conduct usability testing to determine their initial feasibility, and while that objective was clearly communicated at the start of the process, there was disappointment from some PDGs that the pilot data could not be used to yield new information about the disease or to inform clinical decision-making. Revisiting co-design objectives more frequently and understanding stakeholder goals (and limitations) would be useful for grounding all partners back to the same objective.

This co-design process took a significant amount of time and effort from all partners (>1 year with considerable full-time equivalents from some partners). While this can be seen as a limitation, it may be the reality of co-designing a meaningful personal health app. This needed investment should be clearly defined at the outset to manage expectations appropriately. However, this up-front investment could yield a shorter time to implementation. Further, there was a benefit in the ability to have conversations across five different PDGs, where different disease areas were able to learn from one another.

Each app was iteratively modified, even during the beta testing usability period. This approach is uncommon in app development but has value by enabling the end user to see the result of their feedback and further usability testing of a new fix or improvement in real time. A consequence of this approach is that the release of a new or modified feature by a change in code can inadvertently impact functionality in another app feature. While this scenario occurred on a few occasions, the app development team anticipated this reality and had the appropriate resources ready to troubleshoot these inadvertent app bugs when they arose. In the end, this did not significantly impact the co-design timeline and demonstrates that an app does not necessarily have to be locked for modifications during active use. A further challenge with the aforementioned real-time modification approach is the need for frequent research ethics board modifications to amend study protocols, which, depending on the review board, can be associated with time delays and additional costs.

While the initial app prototypes offered benefit to many of the end user testers, there was a proportion of end users (8%) who did not report any benefit. There will likely always be a proportion of individuals who will not receive benefit from a P-DHT–delivered tool for health. A challenge for app co-designers is in defining the success of the personal health app and understanding which patients are most likely to engage and find benefit. In the context of the P-DHT space, broad benefit across a user population may not be a reality, and success may be defined under a different lens relating to narrower n-of-1 outcome definitions.

As evidenced by the low adherence to most tasks in the PCD group, there were challenges in engaging a minor population compared to an adult population. This cohort had a significantly lower completion rate of biweekly check-in calls (<40% of calls), compared to up to 100% for most other groups. This check-in call has been found to be essential in maintaining engagement in the participation in digital health research activities [1]. Minors may require alternate methods of checking in beyond a phone call. Further work is needed to better understand how to engage a youth population in the use of P-DHTs for health. As the sample size in each group was ~50 participants, the numbers were not high enough to examine differences in usability, adherence, or other feedback by factors such as sociodemographic characteristics. However, with the exception of the Sarcoidosis cohort, all beta tester participants were predominantly White. Therefore, these co-design outcomes cannot generalize to diverse populations.

Importantly, the initial purpose of the Rare As One project was to develop at a high level symptom tracking and transmitting apps. In light of this, the broad goal of each app was loosely already defined, and the grounding exercises were designed to aid in tailoring the goals for each unique disease group. For others wishing to use a similar framework, the goal framing workshops may include different grounding exercises (eg, tools that provide access to care or health knowledge) than those configured here.

Future Directions and Conclusions

This user-centered, collaborative, co-design work was intended to yield open-source code and usability data for 5 prototype apps for long COVID-19, pancreatitis, sarcoidosis, VCP disease,

XSL•FO
RenderX

and PCD. Our aim is to enable others to leverage this open-source work (all available via GitHub) to use the built study engine, prototype apps, extract specific features, or modify it for the same or different clinical populations. It is our hope that other PDGs, funders, or commercial entities might leverage these open-access tools, and if not, that this framework might be used and modified for additional use cases. More beta testing and co-design will likely be needed to further evaluate the potential benefit these apps could yield for end users.

Personal health app co-design is rooted in a complex digital landscape that requires a significant amount of up-front effort and time. Despite the up-front investment in time, taking this co-design approach may yield shortcuts in the longer-term implementation of a personal health app. For groups willing to provide this up-front investment, here we provide a co-design framework and the building blocks of five prototype personal health apps that could be leveraged for improving understanding of, and communicating symptoms of, rare or complex diseases. This reflects a potentially novel resource for PDGs and communities actively engaged in research. The ultimate goal of this personal health app co-design is to develop n-of-1 suggestion tools that could empower end users to glean insight into what day-to-day factors improve or exacerbate their condition, which could be transformative for personalized health care, particularly in the rare or complex disease landscape.

Acknowledgments

We would like to acknowledge all co-design partners, including Biohub, the patient-driven organizations (Patient-led Research Collaboration [long COVID-19], Mission Cure [pancreatitis], Foundation for Sarcoidosis Research [FSR], Cure VCP Disease, Inc, and the Primary Ciliary Dyskinesia Foundation), and device companies (Empatica, Lumia, MIR, and Oura). We would like to acknowledge and thank all of the end user co-designers who participated in the beta testing study. FSR is grateful to Biohub's Rare as One Initiative for providing a grant to support FSR's participation in this project.

Funding

This work has been made possible by Biohub through grant 2024-342789 from the Chan Zuckerberg Initiative donor advised fund, an advised fund of the Silicon Valley Community Foundation.

Data Availability

Among participants who opted in, coded study data (deidentified) are made available on the Synapse platform [32] at Sage Bionetworks [33] and other platforms to be announced with 4YouandMe's website [24] and can be freely accessed by any researcher who becomes "qualified" by becoming a registered and certified Synapse user [34], and by meeting the specific conditions of use that require submitting an intended data use statement alongside an institutional review board–approved protocol. The specific Synapse Project page can be found on [35] among registered Synapse users.

Source code for the prototype apps is available at the following reproducibility Github link and website page: iOS shared codebase [36], Android shared codebase [37], and web codebase [38].

The investigators encourage others to access, modify, and build upon these open-source prototype apps.

Authors' Contributions

SMG wrote the manuscript and coled the co-design development and beta testing study. EK managed the beta testing study. RY conducted the data analysis. MS, PG, CO, DV, and JW were engagement specialists for the beta testing study. JP and MT developed the mobile apps. MF, YR, HW, MG, DM, MM, RM, TS, EH, AP, NP, and ZY were patient-driven organization partners and coled the co-design development and drove recruitment. HB-P and AS, representing the funders, coled the co-design development. SAS, DA, EA, RG, DH-G, NH, CK, and MO were beta testing participants and provided comments on the written manuscript. SF led the co-design development. All authors contributed to the editing of the manuscript.

Conflicts of Interest

TS and EH from the Foundation for Sarcoidosis Research (FSR) received funding support from numerous pharmaceutical companies associated with program implementation, and patient-centered clinical trial design and recruitment. FSR receives funding from numerous nonprofit organizations and community-based and family foundations. HB-P is Director at Biohub that funded this research. AS is Program Manager at Biohub that funded this research.

Multimedia Appendix 1
Phase 1 co-design activities.
[DOCX File , 14 KB - mhealth_v14i1e85425_app1.docx]

Multimedia Appendix 2

XSL•FO
RenderX

App specification document example (pancreatitis app).
[XLSX File (Microsoft Excel File), 154 KB - mhealth_v14i1e85425_app2.xlsx]

Multimedia Appendix 3
Sample beta testing study advertisement.
[PDF File (Adobe PDF File), 902 KB - mhealth_v14i1e85425_app3.pdf]

Multimedia Appendix 4
Eligibility criteria of beta testing participants.
[DOCX File , 8 KB - mhealth_v14i1e85425_app4.docx]

Multimedia Appendix 5
Wearable/smart device descriptions.
[DOCX File , 7 KB - mhealth_v14i1e85425_app5.docx]

Multimedia Appendix 6
App goals and specifications.
[DOCX File , 10 KB - mhealth_v14i1e85425_app6.docx]

Multimedia Appendix 7
Prototype app features.
[DOCX File , 1180 KB - mhealth_v14i1e85425_app7.docx]

Multimedia Appendix 8
Enrollment flows.
[DOCX File , 845 KB - mhealth_v14i1e85425_app8.docx]

Multimedia Appendix 9
Demographic characteristics of beta testing participants across cohorts.
[XLSX File (Microsoft Excel File), 15 KB - mhealth_v14i1e85425_app9.xlsx]

Multimedia Appendix 10
App features implementd from direct participant feedback.
[DOCX File , 285 KB - mhealth_v14i1e85425_app10.docx]

Multimedia Appendix 11
Proportion of participants rating their likelihood to use the app in the future on a 1 (least likely) to 10 (most likely) rating scale.
[DOCX File , 51 KB - mhealth_v14i1e85425_app11.docx]

Multimedia Appendix 12
Participant beta testing feedback themes—thematic analysis.
[DOCX File , 13 KB - mhealth_v14i1e85425_app12.docx]

References

1. Goodday SM, Karlin E, Brooks A, Chapman C, Harry C, Lugo N, et al. Value of engagement in digital health technology research: evidence across 6 unique cohort studies. J Med Internet Res 2024;26:e57827 [FREE Full text] [doi: 10.2196/57827] [Medline: 39226552]
2. Ginsburg GS, Picard RW, Friend SH. Key issues as wearable digital health technologies enter clinical care. N Engl J Med 2024;390(12):1118-1127. [doi: 10.1056/NEJMra2307160] [Medline: 38507754]
3. Langholm C, Kowatsch T, Bucci S, Cipriani A, Torous J. Exploring the potential of Apple SensorKit and digital phenotyping data as new digital biomarkers for mental health research. Digit Biomark 2023;7(1):104-114. [doi: 10.1159/000530698] [Medline: 37901364]
4. Goodday SM, Geddes JR, Friend SH. Disrupting the power balance between doctors and patients in the digital era. Lancet Digit Health 2021;3(3):e142-e143 [FREE Full text] [doi: 10.1016/S2589-7500(21)00004-2] [Medline: 33516653]
5. Friend SH, Ginsburg GS, Picard RW. Wearable digital health technology. N Engl J Med 2023;389(22):2100-2101. [doi: 10.1056/NEJMe2303219] [Medline: 38048193]

XSL•FO
RenderX

6. Giebel GD, Speckemeier C, Abels C, Plescher F, Börchers K, Wasem J, et al. Problems and barriers related to the use of digital health applications: scoping review. J Med Internet Res 2023;25:e43808 [FREE Full text] [doi: 10.2196/43808] [Medline: 37171838]
7. Mattison G, Canfell O, Forrester D, Dobbins C, Smith D, Töyräs J, et al. The influence of wearables on health care outcomes in chronic disease: systematic review. J Med Internet Res 2022;24(7):e36690 [FREE Full text] [doi: 10.2196/36690] [Medline: 35776492]
8. Smits M, Kim CM, van Goor H, Ludden GDS. From digital health to digital well-being: systematic scoping review. J Med Internet Res 2022;24(4):e33787 [FREE Full text] [doi: 10.2196/33787] [Medline: 35377328]
9. Lodewyk K, Wiebe M, Dennett L, Larsson J, Greenshaw A, Hayward J. Wearables research for continuous monitoring of patient outcomes: a scoping review. PLoS Digit Health 2025;4(5):e0000860 [FREE Full text] [doi: 10.1371/journal.pdig.0000860] [Medline: 40343891]
10. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020;3:21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
11. Mobile app retention rate: what's a good retention rate? Upland. 2021. URL: https://uplandsoftware.com/localytics/resources/blog/mobile-apps-whats-a-good-retention-rate/ [accessed 2026-06-18]
12. Inal Y, Wake JD, Guribye F, Nordgreen T. Usability evaluations of mobile mental health technologies: systematic review. J Med Internet Res 2020;22(1):e15337 [FREE Full text] [doi: 10.2196/15337] [Medline: 31904579]
13. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022;24(4):e35120 [FREE Full text] [doi: 10.2196/35120] [Medline: 35471414]
14. Duffy A, Christie GJ, Moreno S. The challenges toward real-world implementation of digital health design approaches: narrative review. JMIR Hum Factors 2022;9(3):e35693 [FREE Full text] [doi: 10.2196/35693] [Medline: 36083628]
15. Sanders EBN, Stappers PJ. Co-creation and the new landscapes of design. CoDesign 2008;4(1):5-18. [doi: 10.1080/15710880701875068]
16. Tang C, Lorenzi N, Harle CA, Zhou X, Chen Y. Interactive systems for patient-centered care to enhance patient engagement. J Am Med Inform Assoc 2016;23(1):2-4 [FREE Full text] [doi: 10.1093/jamia/ocv198] [Medline: 26912537]
17. Van den Bulck SA, Hermens R, Slegers K, Vandenberghe B, Goderis G, Vankrunkelsven P. Designing a patient portal for patient-centered care: cross-sectional survey. J Med Internet Res 2018;20(10):e269 [FREE Full text] [doi: 10.2196/jmir.9497] [Medline: 30287416]
18. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (integrate, design, assess, and share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
19. Wheelock A, Bechtel C, Leff B. Human-centered design and trust in medicine. JAMA 2020;324(23):2369-2370. [doi: 10.1001/jama.2020.21080] [Medline: 33320236]
20. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
21. Vial S, Boudhraâ S, Dumont M, Tremblay M, Riendeau S. Developing a mobile app with a human-centered design lens to improve access to mental health care (Mentallys project): protocol for an initial co-design process. JMIR Res Protoc 2023;12:e47220 [FREE Full text] [doi: 10.2196/47220] [Medline: 37606978]
22. Martin A, Caon M, Adorni F, Andreoni G, Ascolese A, Atkinson S, et al. A mobile phone intervention to improve obesity-related health behaviors of adolescents across Europe: iterative co-design and feasibility study. JMIR Mhealth Uhealth 2020;8(3):e14118 [FREE Full text] [doi: 10.2196/14118] [Medline: 32130179]
23. Sakamoto M, Guo YPE, Wong KLY, Mann J, Berndt A, Boger J, et al. Co-design of a digital app "WhatMatters" to support person-centred care: a critical reflection. Int J Geriatr Psychiatry 2023;38(10):e6014. [doi: 10.1002/gps.6014] [Medline: 37828681]
24. 4YouandMe. URL: https://4youandme.org/ [accessed 2026-06-19]
25. Naeem M, Ozuem W, Howell K, Ranfagni S. A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. Int J Qual Methods 2023;22. [doi: 10.1177/16094069231205789]
26. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2008;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
27. Raza S, Rahman M, Kamawal S, Toroghi A, Raval A, Navah F, et al. A comprehensive review of recommender systems: transitioning from theory to practice. Comput Sci Rev 2026 Feb;59:100849. [doi: 10.1016/j.cosrev.2025.100849]
28. Cosentino J, Belyaeva A, Liu X, Furlotte N, Yang Z, Lee C. Towards a personal health large language model. arXiv Preprint posted online on June 10, 2024. [doi: 10.48550/arXiv.2406.06474]
29. Merrill MA, Paruchuri A, Rezaei N, Kovacs G, Perez J, Liu Y, et al. Transforming wearable data into personal health insights using large language model agents. arXiv Preprint posted online on June 10, 2024. [doi: 10.48550/arXiv.2406.06464]

XSL•FO
RenderX

30. Fox S, Brown LJE, Antrobus S, Brough D, Drake RJ, Jury F, et al. Co-design of a smartphone app for people living with dementia by applying agile, iterative co-design principles: development and usability study. JMIR Mhealth Uhealth 2022;10(1):e24483 [FREE Full text] [doi: 10.2196/24483] [Medline: 35029539]
31. Marwaa MN, Guidetti S, Ytterberg C, Kristensen HK. Using experience-based co-design to develop mobile/tablet applications to support a person-centred and empowering stroke rehabilitation. Res Involv Engagem 2023;9(1):69 [FREE Full text] [doi: 10.1186/s40900-023-00472-z] [Medline: 37620982]
32. Synapse. URL: https://www.synapse.org/ [accessed 2026-06-19]
33. Sage Bionetworks. URL: https://sagebionetworks.org/ [accessed 2026-06-19]
34. Docs. Synapse. URL: https://help.synapse.org/docs/User-Account-Tiers.2007072795.html [accessed 2026-06-19]
35. A collaborative platform for open biomedical science. Synapse. URL: https://www.synapse.org/Synapse:syn70090241/wiki/635388 [accessed 2026-06-19]
36. 4YouandMeiOS. GitHub. URL: https://github.com/4YouandMeData/4YouandMeiOS [accessed 2026-06-19]
37. 4YouandMeAndroid. GitHub. URL: https://github.com/4YouandMeData/4YouandMeAndroid/tree/develop [accessed 2026-06-19]
38. 4YouandMeBackend. GitHub. URL: https://github.com/4YouandMeData/4YouandMeBackend [accessed 2026-06-19]

Abbreviations

AI: artificial intelligence
EDA: electrodermal activity
IRB: institutional review board
PCD: primary ciliary dyskinesia
PDG: patient-driven organization
P-DHT: personal digital health technology
UX: user experience
VCP: valosin-containing protein

Edited by L Buis; submitted 22.Oct.2025; peer-reviewed by F Elkourdi, R Pryss; comments to author 03.Feb.2026; revised version received 08.Jun.2026; accepted 09.Jun.2026; published 20.Jul.2026.

Please cite as:
Goodday SM, Karlin E, Sorensen M, Yang R, Gordon P, Opoku C, Vuong D, Wilson J, McKenzie D, Piccotti J, Tavanti M, Fitzgerald M, Re'em Y, Wei H, Golden M, Morgan D, Manion M, Mosquera R, Shivas T, Hoover E, Peck A, Peck N, Yavarow Z, Bjornson-Pennell H, Stratton A, Anova Sahoo S, Arbittier D, Arbittier E, Goff R, Harvin-Graham D, Howard N, Knudsen C, Oldham M, Friend S
A Personal Health App and Wearable Co-Design Framework for Rare and Complex Diseases: User-Centered, Collaborative Co-Design Study
JMIR Mhealth Uhealth 2026;14:e85425
URL: https://mhealth.jmir.org/2026/1/e85425
doi:10.2196/85425
PMID:

©Sarah Margaret Goodday, Emma Karlin, Madeleine Sorensen, Robin Yang, Paul Gordon, Caresse Opoku, Daniel Vuong, Jenee Wilson, Diane McKenzie, Jules Piccotti, Massimiliano Tavanti, Megan Fitzgerald, Yochai Re'em, Hannah Wei, Megan Golden, Daniel Morgan, Michele Manion, Ricardo Mosquera, Tricha Shivas, Elise Hoover, Allison Peck, Nathan Peck, Zollie Yavarow, Heidi Bjornson-Pennell, Andra Stratton, Samali Anova Sahoo, Douglas Arbittier, Elizabeth Arbittier, Renee Goff, Donna Harvin-Graham, Nancy Howard, Camille Knudsen, Mary Oldham, Stephen Friend. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 20.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring the Design Space of Glanceable Smartwatch Feedback Displays: Experimental Study

Yuxuan Li[1], PhD; Mark Newman[1], PhD; Predrag Klasnja[1], PhD

School of Information, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Yuxuan Li, PhD
School of Information
University of Michigan
Leinweber Computer Science & Information Building
2200 Hayward St
Ann Arbor, MI, 48109
United States
Phone: 1 2179049289
Email: liyuxuan@umich.edu

Abstract

Background: Self-monitoring technologies are commonly used to promote health behavior change, with glanceable displays offering continuous feedback throughout the day. Yet, it is still unclear how various aspects of these glanceable representations affect their interpretability and usability.

Objective: This study aimed to investigate the effects of 3 design factors—stylization, granularity, and salience—on users' ability to understand glanceable smartwatch-based feedback on daily step goals.

Methods: We conducted an online simulation study to examine how 3 design dimensions—stylization, salience, and granularity—influence the effectiveness of glanceable feedback displays. Stylization and salience were crossed in a 2×2 factorial design, while granularity varied from 1% to 20% progress increments. A total of 202 Amazon Mechanical Turk participants were randomly assigned to 1 of 16 smartwatch display conditions. In each condition, participants viewed feedback on daily step progress and estimated the level of progress shown. We measured estimation error and questionnaire-assessed perceived usability and acceptability. The collected data were analyzed using generalized estimating equations and linear regression.

Results: High stylization reduced accuracy (+4.52 error points; $P<.001$) and negatively affected perceptions across 6 dimensions, including comprehension ($P=.003$), complexity ($P<.001$), and usability ($P=.001$). Granularity had a nonlinear effect: error was lowest around 5%-10%, with sharp increases at 20%. The 10% level also received the most favorable ratings, for example, comprehension (+0.656; $P=.003$). Salience had no effect. Previous smartwatch users were less accurate than never-users (+7.46 points) but rated displays as more useful ($P=.002$) and easier to focus on ($P<.001$). Current users gave similarly positive ratings on attention and usefulness.

Conclusions: These findings could help researchers design effective glanceable smartwatch feedback displays and expand the design space for glanceable feedback.

(JMIR Mhealth Uhealth 2026;14:e81972) doi:10.2196/81972

KEYWORDS

glanceable displays; health informatics; mobile health; physical activity; smartwatches

Introduction

Self-monitoring is widely used to support behavior change in health-related contexts, and meta-analyses have found self-monitoring to be highly effective for physical activity, diet, and weight loss [1-4]. Increasingly, mobile devices such as smartphones and smartwatches have become a common way of providing self-monitoring interventions [5-8]. One of their key strengths for supporting self-monitoring is that they enable users to see self-monitoring feedback frequently—increasing its salience and, thus, its potential to shape users' behavior. Mobile devices can increase feedback frequency by leveraging multiple modalities, including inside apps, push notifications, text messages, and, crucially, glanceable displays such as lock screens (phones) and clock faces (smartwatches).

XSL•FO
RenderX

Glanceable displays refer to the designs that present important, noncritical information in an always-visible but nonintrusive form, allowing data to move seamlessly from the periphery of a user's awareness to the focus of attention and back again [9]. Glanceable displays are a promising approach to supporting self-monitoring, as they enable researchers and intervention designers to create feedback representations that a person can see throughout the day without having to remember to go into an app. By ensuring feedback is always visible, these displays allow users to maintain awareness of their health metrics throughout the day without the burden of needing to navigate to an app or manage incessant push notifications [10-12].

As the smartwatch market has matured, health and fitness have become key domains supported by these devices, and glanceable displays are a central medium through which smartwatches support self-monitoring of health-related metrics. For instance, the Apple Watch includes a complication, which is an element of the clock face that displays information other than the time, for the Activity app, shown by default on most built-in clock faces, which displays feedback on minutes of activity, calories burned, and the number of hours of nonsedentary time. Similarly, most default Fitbit clock faces display step count and heart rate. The information provided in such feedback displays is visible every time the user looks at the watch, providing ongoing awareness about one's level of activity. Moreover, Apple Watch, Fitbit, Garmin, and Wear OS platforms all support some level of clock face customization, enabling a range of feedback—as well as forms of that feedback—to be incorporated into the clock face. Given how frequently watch wearers glance at their watches each day, for a growing number of smartwatch users, feedback shown on their watches represents a powerful way to benefit from self-monitoring without the burden of receiving incessant notifications.

However, little is known about what makes a glanceable display effective, and, specifically, what design dimensions influence the understandability, usability, and behavioral impact of such displays. Prior research has demonstrated the potential of glanceable displays but has often yielded ambiguous results regarding design specifics. For instance, in the *UbiFit Garden* study, researchers found that a stylized virtual garden wallpaper effectively supported physical activity over 3 months [10], and the *BeWell* system successfully used animated sea creatures to represent sleep and social contact [12]. However, in other studies, such as *Gardy* [11], researchers found that highly stylized displays (eg, growing trees and mushrooms) were less effective than lower-stylized versions. Qualitative data from *Gardy* suggested that the issue might not have been stylization itself, but rather a lack of granularity—the inability of users to extract accurate progress data from the image. Because previous designs have varied across multiple dimensions simultaneously, it remains unclear how different dimensions of glanceable displays individually impact the understandability and usability of health feedback.

This question is particularly important in the context of smartwatch displays, which are constrained by the small screen size, the need to display time clearly first and foremost, and the very brief interactions with the feedback. To optimize the design of smartwatch-based feedback displays so that these displays can effectively support health-behavior change and be sufficiently well-liked so users are willing to keep them on their watches, we need better evidence for how different aspects of the feedback design affect how they are perceived and used.

In this paper, we report on a web-based factorial experiment with 202 participants that aimed to begin filling this knowledge gap by systematically examining the influence of 3 design dimensions—granularity of feedback, its salience, and whether the feedback is stylized—on the effectiveness of a smartwatch-based glanceable display to communicate progress on a daily step goal. Representing progress on health goals—for physical activity, sleep, diet, etc—is a common use of feedback displays on smartwatches, so the findings from our study may be applicable to other domains of health behavior change beyond physical activity.

On a methodological note, while our simulation-based study could not examine the impact of the tested design dimensions on behavior change—a research question that would require a field experiment—the size of the study, which was powered to detect small effect sizes, enabled us to examine perceptual properties and subjective judgments of the tested design dimensions with a level of rigor that could not be achieved in small-scale field deployments that are feasible for early-stage design research. As such, this study presents an example of resource-efficient optimization research that is increasingly advocated for in health and behavioral sciences [13-15].

Methods

Overview

To begin to understand how design affects usability and effectiveness of glanceable displays, we conducted an online simulation-based experiment that manipulated 3 design dimensions of glanceable feedback displays: stylization, salience, and granularity. In the experiment, stylization and salience were crossed in a 2×2 factorial structure, while granularity was varied across multiple levels, ranging from showing progress in 1% to 20% increments. A total of 202 Amazon Mechanical Turk (MTurk) participants were randomly assigned to 1 of 16 interface conditions. In each condition, participants were presented with a smartwatch display showing feedback on progress toward a daily step goal and were asked to estimate the level of progress that was shown.

The primary aim of the study was to investigate how the design dimensions influenced users' accuracy in assessing goal progress displayed on a smartwatch at a glance. The primary outcome was the error in participants' estimates of the displayed progress. A secondary aim was to examine how each design dimension affected perceived usability and acceptability.

Rationale for the Use of an Online Simulation-Based Experiment

Given our goal of systematically evaluating the impact of design variations on users' ability to rapidly comprehend step goal progress at a glance, we opted to conduct an online simulation-based experiment. Like any research methodology, this approach presents a set of trade-offs. Its most obvious

XSL•FO
RenderX

limitation is the lack of ecological validity—the fact that participants did not interact with a real smartwatch in their natural environments. However, given our research question, this limitation is less critical than it might first appear, and it is made up for by the method's advantages, including feasibility and scalability. Our rationale for using an online experiment was threefold.

First, the primary question addressed by this study is how variations in design influence participants' ability to accurately perceive the level of progress toward a goal. This question concerns visual perception (ie, the ability to perceive small differences between representations) and interpretability of different types of visualization (ie, the ability to tell what the feedback display is trying to show). Such questions have routinely been studied using surveys and online experiments, including in the context of graphical representations for communicating health risks and results from medical tests [16-21]. Importantly, for our research question, an online experiment provides a *conservative* measure of the potential impact of different feedback designs, since participants are able to focus their attention on the experimental task without the distraction of other tasks that may interfere with how accurately they perceive feedback in the midst of everyday life.

Second, an online experiment enabled us to feasibly study the impacts of design with a high level of statistical rigor. Because we were concerned that small details of each design could impact performance in unforeseen ways, we created a total of 16 different interfaces. Since we expected the differences in objective performance across interfaces to be fairly small, we determined that over 200 participants would be needed to estimate these effects (power calculations are presented below), and that such a number could only be reasonably obtained by using an online platform such as Amazon MTurk. Conducting an equivalent study in the field would be prohibitively expensive, with the time and resource costs far outstripping the importance of the research question the study was trying to answer.

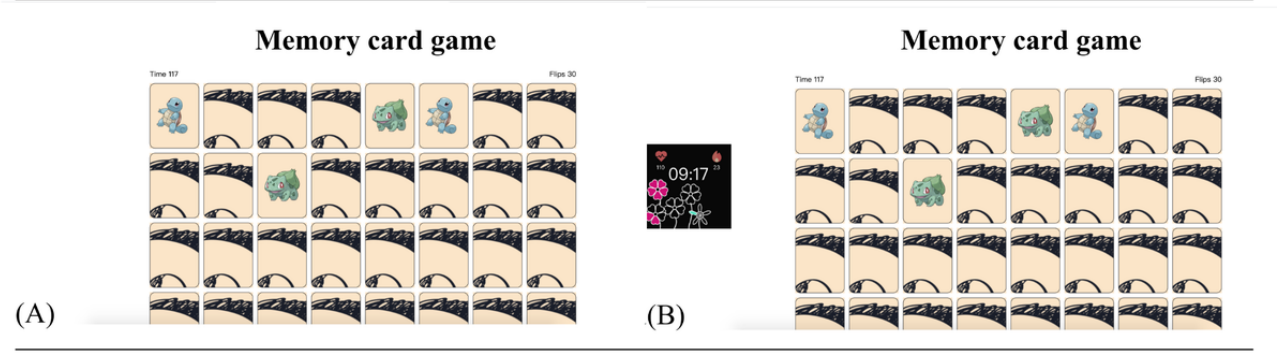
Third, using a simulation greatly facilitated our measurement approach. To measure user performance, we needed to compare participants' estimates of their step goal progress to a "true" value; in addition, we needed to be able to ensure that participants were making a reasonable effort to adhere to the study procedures. Given this need for the objective "ground truth" for comparison and to know the precise state of the watch display when the participant was engaging with the experimental task, we determined that a simulation was necessary. Controlling the "glancing" behavior in the wild would be impossible, and while moderated sessions in a user testing laboratory could have been used in place of a simulation, they would not allow us to scale effectively without huge expenses that would not be merited by the expected significance of the results.

Experimental Task

To frame the overall experiment, participants were told that they had recently bought a smartwatch, and they were using it to track their steps as they tried to increase their activity and walk 10,000 steps per day. Each experimental task was further set up with a brief scenario: participants were told that they had an important video call in a few minutes and that while they were waiting for the meeting to start, they decided to play an online game of Memory. To make sure they were not late for the meeting, they were asked to tap the spacebar a few minutes after they started playing the game to check the time, as they were told that the simulation time did not flow at the same rate as real-world time.

Tapping the spacebar brought up an image of a smartwatch that displayed time and step-goal progress using the design for the condition to which the participant was randomized. The image of the display was sized to approximate the size of a smartwatch viewed from a foot and a half (50 cm) away, the approximate distance of the wrist from the eyes when a person looks at a watch. The smartwatch image remained on the screen for 5 seconds, after which it was hidden to mimic the act of putting the arm down (Figure 1). As soon as they saw the time and activity progress, participants could resume playing the game (ie, the game remained active even while the smartwatch display was showing) in order to mimic the return to the primary task after glancing at a watch. Overall, 15 seconds after the smartwatch display was hidden, the game was interrupted, and participants were shown a 2-question questionnaire asking them to report (1) the time they saw on the display and (2) the percentage of completion of the daily step goal shown on the watch.

Figure 1. (A) Web screen without the display. (B) Web screen with the display.



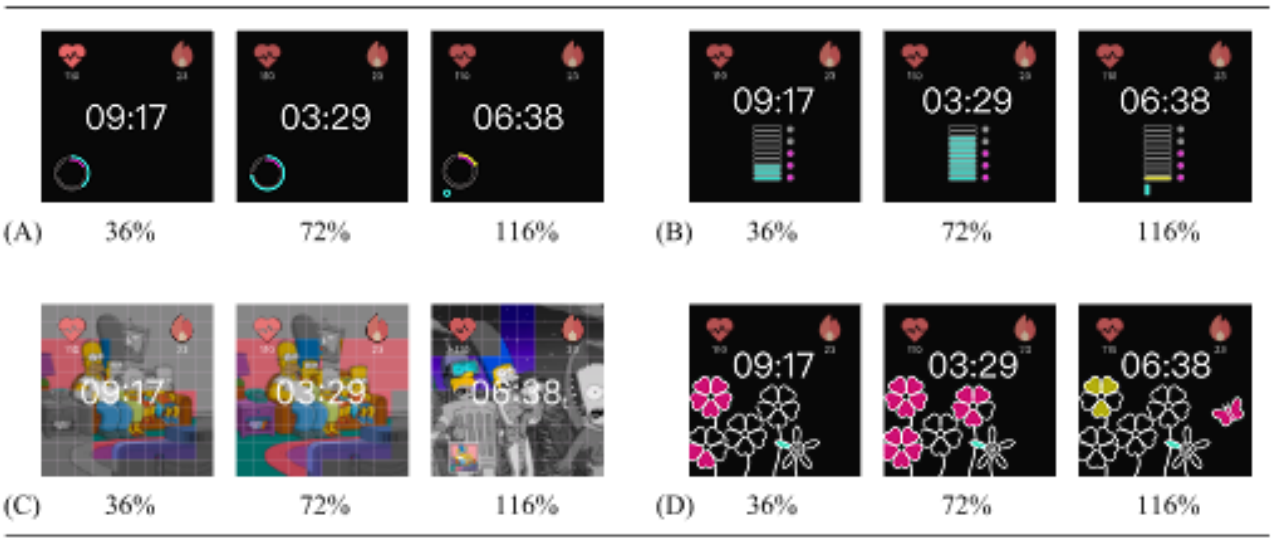
XSL•FO
RenderX

Study Procedures

All study procedures were hosted on a study website, which was linked from the task we created on the Amazon MTurk platform. When they entered the study website, participants were presented with consent information that described the study procedures, after which they were presented with a brief training that explained the display design for their assigned condition. After reading the training material, participants were given a training task where they were asked to estimate the progress on the daily step goal shown on the display that used their assigned design. After they entered their estimate, the system provided the correct answer to further assist with learning how to interpret the display accurately.

Following the training, participants were presented with 3 experimental tasks. Each experimental task followed the structure outlined in the *Experimental Task* section. However, each scenario was contextualized to represent a different time of day—morning, midday, or afternoon—and featured a distinct level of step-goal completion when the smartwatch was displayed (Figure 2). After they completed all 3 experimental tasks, participants were asked to complete a questionnaire that assessed their perceptions of the usability and interpretability of the smartwatch display they used in the experiment.

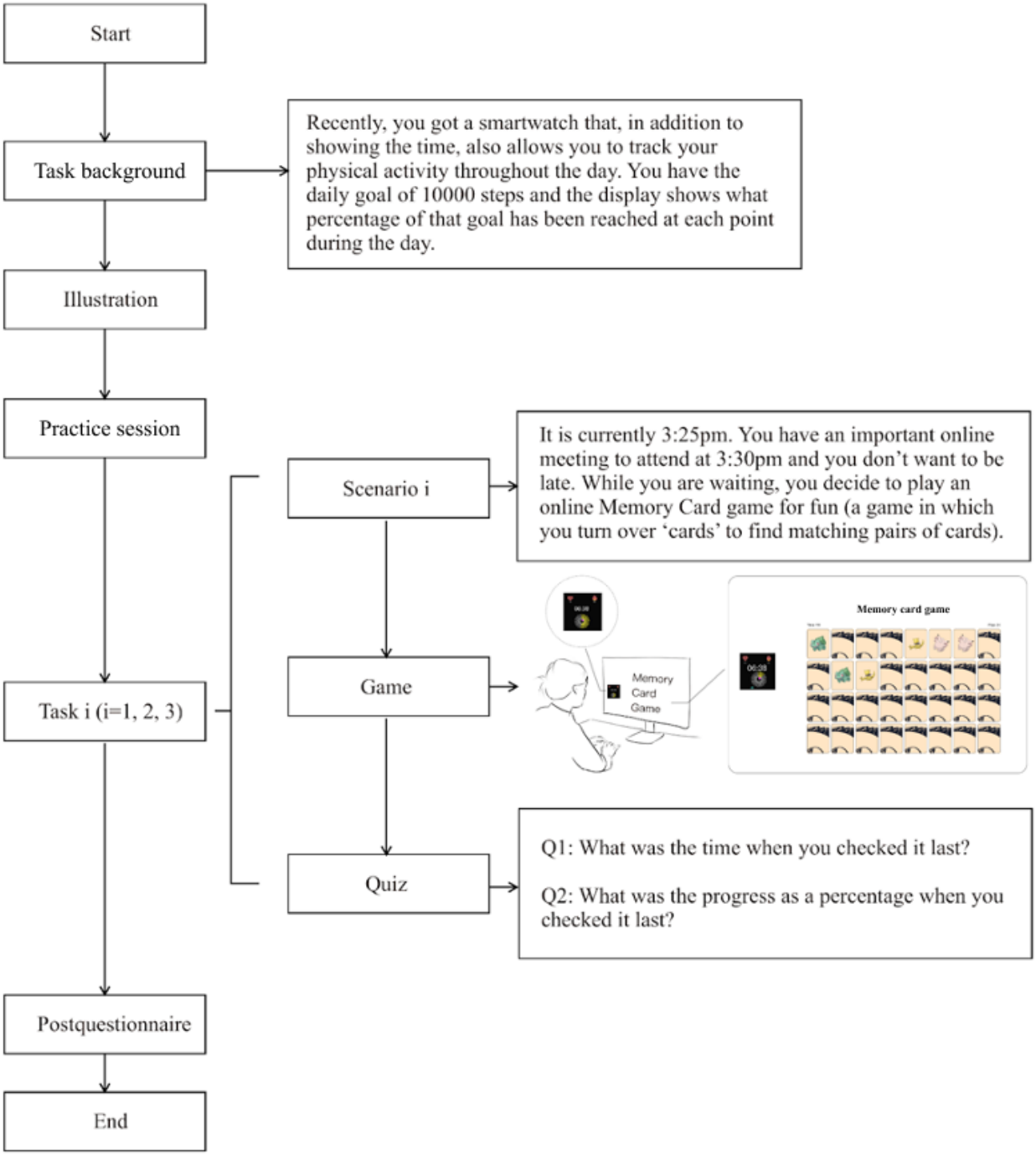
Figure 2. Four example displays shown at different times of day (morning, midday, and afternoon), each paired with varying levels of step-goal completion. Practice sessions were provided to help participants learn how to interpret the progress indicators. (A) Panel E in Figure 4: 1% increments. Progress turns from blue to yellow at 100%, with a small blue circle added when the goal is exceeded (eg, 116%). (B) Panel C in Figure 4: 10% increments. At 36%, 3 bars are filled. Upon overreaching the goal, the color changes to yellow and a small blue bar appears. (C) Panel N in Figure 4: 1% increments. A grayscale Simpsons image gradually regains color square by square. Exceeding the goal triggers a new image, and a small completed image appears below. (D) Panel K in Figure 4: 5% increments. Each pink flower has 5 petals; only full 5% units are shown. Beyond the goal, the flower turns yellow, and a pink butterfly appears.



After they submitted the questionnaire, participants were issued a code that they could enter in the MTurk task form to request payment for completing the task. Participants were compensated US $2.50 for completing the experiment. The compensation was chosen to match a US $15/hour wage, adjusted for the duration of the task (~10 minutes). The full flow of the experiment is shown in Figure 3.

XSL•FO
RenderX

Figure 3. Study flow depicting participants' progress through the training session, 3 experimental tasks, and the poststudy questionnaire.



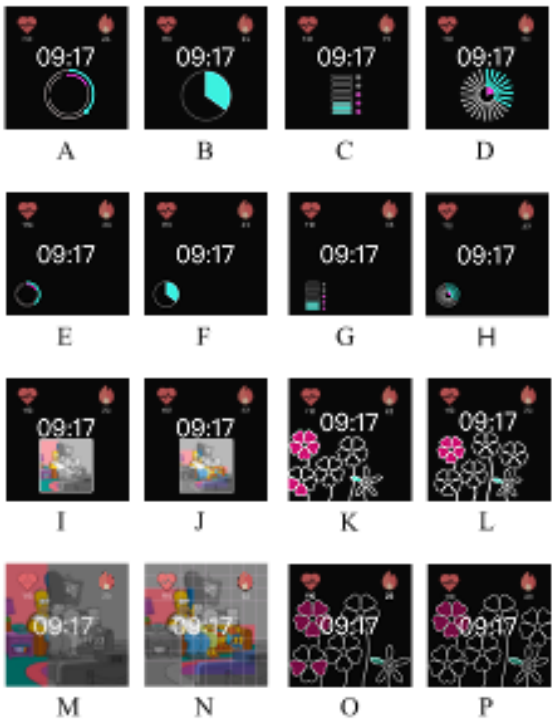
Intervention

Overview

The main manipulation in the experiment was the design of the display shown to each participant. We designed 16 unique interfaces by varying 3 design dimensions: stylization (low, high), salience (low, high), and granularity (1%, 4%, 5%, 10%, 20%). Stylization and salience were manipulated at 2 levels each, while granularity varied across 5 uneven levels (Figure 4). The design dimensions tested in the experiment were as follows:

XSL•FO
RenderX

Figure 4. A total of 16 unique interfaces used in the study: (A) Apple Watch–inspired design using a progress circle that directly reflects the current value. (B) Apple Watch-inspired design using the pie chart that directly reflects the current value. (C) Ten horizontal bars, each representing 10% progress (eg, 36% colors 3 bars and 72% colors 7 bars). (D) Bars arranged in a circle, each representing 4% (eg, 36% colors 9 bars and 72% colors 18 bars). (E-F) Low-salience versions of panels A and B, with reduced size. (G-H) Low-salience versions of C and D, with reduced size. (I) A Simpsons figure progressively colored from left to right in 1% increments. (J) Simpsons figure divided into 100 squares; each 1% of progress colors 1 square. (K) Flower petals representing 5% each (eg, 36% of progress colors 7 petals and 72% colors 14 petals). (L) Whole flowers representing 20% each (eg, 36% of progress colors 1 flower and 72% colors 3 flowers). (M-P) Low-salience versions of panels I-L, featuring reduced saturation and brightness, along with increased transparency.



interface	stylization	salience	granularity
A,B	low	high	1%
C	low	high	10%
D	low	high	4%
E,F	low	low	1%
G	low	low	10%
H	low	low	4%
I,J	high	high	1%
K	high	high	5%
L	high	high	20%
M,N	high	low	1%
O	high	low	5%
P	high	low	20%

Stylization

Stylization refers to the use of "signs that have an intermediate degree of transparency to the signified object" [9]. In mobile health, stylized feedback is sometimes used to obscure the nature of the metrics that are being tracked, in order to display health information on displays that may be visible to other people while maintaining user privacy. We used metaphors, such as the use of flowers as iconic signs, to signify the information in the feedback displays. We chose to test stylization due to its potential to increase engagement, as high-stylized displays can be designed to be attractive and can be highly customizable [22]. In addition, high-stylized displays can be privacy-preserving, making them a good match for displaying feedback on sensitive types of health information. On the other hand, interpreting high-stylized displays requires learning, and they may not be as precise for displaying progress as graphs commonly used for data visualization. In our experiment, low-stylized displays used standard graphical formats such as bar charts, pie charts, or circles (eg, panel A in Figure 4). The selection of these charts aligns with prior research in glanceable displays, as these formats are efficient for rapid information retrieval and quick understanding on smartwatches [23-25]. High-stylized displays (eg, panels J and O in Figure 4) incorporated more expressive visuals, including cartoon characters and flowers. Inspired by the customizable cartoon interfaces on the Apple Watch, such as Snoopy, we chose pictures from "*The Simpsons*"—a widely watched cartoon in the United States—to represent progress in the high-stylized condition. Additionally, drawing inspiration from the *Ubifit Garden* [26], we used flowers as an alternative high-stylized representation to design a visually engaging form of feedback.

Granularity

Granularity refers to the size of the increment used to display progress. High-granularity feedback can represent small changes (eg, each 1% of goal attainment), while low-granularity feedback represents larger increments (eg, every 10% or 20% of progress). Due to the small size of the smartwatch screen and the need to display multiple pieces of information, including time, designers may be tempted to design smartwatch feedback that uses lower data granularity in order to satisfy aesthetic considerations and feedback readability. On the other hand, high-granularity feedback has been shown to aid progress estimation and increase physical activity motivation [27,28].

To evaluate these effects, we selected 5 granularity levels: 1%, 4%, 5%, 10%, and 20%. These levels are not evenly spaced, as they are exploratory design choices based on our design inspirations and established user interface paradigms in the wearable market. The 1% granularity level represents the high-precision displays found in Apple Watch and Garmin designs, where progress is mapped to exact pixel or degree

XSL•FO
RenderX

movements. The 4%, 5%, and 10% levels reflect common progress bar increments found in Fitbit interfaces (eg, progress lights in the Fitbit Flex), where progress is often discretized into "steps" or blocks to enhance glanceability. Finally, the 20% level was inspired by HCI designs such as the *UbiFit Garden* [26], which uses discrete icons (eg, 1 flower) to represent the completion of a significant sub-goal. We chose the 20% step for this purpose as it represents a moderately large increment (ie, 1/5th) commonly used to represent goal progress.

Salience

Salience refers to the degree to which a stimulus captures attention by standing out from its environment. Rather than being a single attribute, salience is a composite of visual factors—such as contrast, size, and positioning—that determine how much "attentional grab" a display exerts [29-31]. Salience is known to influence the perception of peripheral displays [32], yet operationalizing this concept for smartwatch interfaces proved nontrivial, as different design styles required different levers to manipulate user attention.

For low-stylized displays, which used standard graphs, we operationalized salience primarily through size and chart position [29]. High-salience interfaces featured progress feedback centered on the screen at a height exceeding 100 points, demanding immediate focus. Conversely, low-salience versions used a complication style—placed in the screen's corner at a height of 60 points—following Apple Watch Human Interface Guidelines. Here, the reduced size and peripheral placement minimized the display's ability to easily compete for the user's attention vs the time display.

The operationalization became more nuanced for high-stylized interfaces, where simply shrinking the designs would have rendered them unreadable due to their visual complexity. Instead, salience was manipulated by varying the relationship between the representation and the background. The high-salience condition used a centrally positioned 128-point display that remained a distinct, foreground element (eg, panel I in Figure 4). In contrast, the low-salience condition used a 300-point display that covered the entire screen but was designed to recede into the background (eg, panel M in Figure 4). By reducing the saturation and brightness while increasing transparency, this larger display was made less visually prominent to avoid distracting from the watch's primary timekeeping function.

Randomization

When they entered the experiment, each participant was automatically randomized, with the probability of 1 out of 16 (6.25%), to 1 of the 16 experimental designs of progress feedback. All instructions for interpreting the display, the test session at the start of the experiment, as well as each of the 3 experimental tasks used the feedback design to which the participant was randomized.

Each participant completed 3 experimental tasks that used different levels of step-goal progress: 36%, 72%, and 116%. The numbers were chosen to represent low, moderate, and goal-met-and-exceeded progress on the daily step goal, and to not be round, easily guessable numbers. To control for order effects, the order of the 3 tasks was randomized for each participant. This meant that, for instance, a participant may see the interface that represents 116% goal attainment first, followed by interfaces that show 36% and 72% goal attainment. Each level of progress was shown to each participant only once.

Measures

The objective of our study was to find out how the tested design dimensions affect the ability of participants to accurately assess the smartwatch progress feedback display at a glance. Consequently, our primary outcome was the absolute error of the participants' progress estimates, calculated as the absolute difference between the progress estimate and the true progress value, which allowed us to assess the overall accuracy of participants' ratings without taking into consideration whether their ratings were above or below the true value of the goal progress.

Our secondary outcomes included the self-reported ratings of the displays' usability and usefulness. These were assessed using a questionnaire that participants completed after they completed all 3 experimental tasks. The questionnaire was based on the System Usability Scale [33] and a questionnaire used by Shami et al [34] to assess the perceived strengths and weaknesses of glanceable displays. The constructs assessed and the questions used are shown in Textbox 1. Questions Q6-Q7 and Q12-Q14 were adapted from the System Usability Scale; questions Q1-Q5 and Q8-Q11 (except Q10) were adapted from Shami et al [34]. All questions were assessed on a 5-point Likert scale that ranged from *strongly disagree* to *strongly agree*. Q10 was an attention check question that we used to filter out Amazon MTurk workers who were answering questions without paying even minimal attention.

XSL•FO
RenderX

Textbox 1. Poststudy questionnaire. Constructs and corresponding question items are listed below.

Noticeability

- Q1: I could easily notice the progress display.
- Q2: I could notice the change in my progress across 3 scenarios.

Comprehension

- Q3: I felt confident in being able to understand the information regarding my progress.
- Q4: I felt confident in being able to understand the information in the progress display just by glancing at it.

Relevance

- Q5: The display provided me with the information I needed to track my progress.

Complexity

- Q6: The smartwatch display showed too much distracting information.
- Q7: I found the smartwatch display unnecessarily complex.

Division of attention

- Q8: I was able to maintain focus on playing the game while checking the time and my progress.
- Q9: I was able to shift my attention between playing the game and checking the smartwatch display easily.

Attention check

- Q10: Please skip this question and move to the next one. This is just to screen out random clicking.

Attractiveness

- Q11: I found the design of the smartwatch display attractive.

Usability

- Q12: I think the way the progress was shown is useful.
- Q13: I think that I would like to use this smartwatch display on an ongoing basis.
- Q14: I would imagine that most people would learn to use this system very quickly.

Statistical Analysis

Sample Size and Power

The study was powered to detect the main effects of the 3 design dimensions on the primary outcome, the absolute value of the error in progress ratings. While we could not find any existing literature that would readily provide data on the sizes and distributions of the likely effects for the measure we were interested in, we started with the data from the work of Gouveia et al, who designed 21 glanceable displays for showing activity progress [1]. To derive a functional baseline, we examined their *Goal Completion* condition (mean 5340, SD 4528), which emulates established commercial standards (eg, Fitbit Flex), and compared it against their most divergent design, the *Gardy* condition (mean 3760, SD 3511). Using the pooled standard deviation method, this comparison yielded an estimated effect size of approximately d=0.39. However, to remain conservative and ensure the study could detect even subtler differences between our own design iterations, we assumed a small effect size of d=0.2. Further, we assumed a moderate, 65% correlation between repeated measures. With 90% power and the alpha level set to .05, the required sample size was 204.

Analysis

For the analyses of the progress estimation error rates, we used the generalized estimating equation (GEE), which is an extension of linear regression that takes into account the nested nature of repeated-measure data. Given that stylization and salience had 2 levels, but granularity had 5 levels, we estimated the main effects for stylization and salience in a single model, as is typically done in factorial experiments, but we fit a separate model to estimate the effect of granularity. To improve model fit, in all analyses, the estimated error was square-root transformed—that is, the dependent variable was defined as:



Diagnostic plots for the GEE model were assessed to ensure valid statistical inference. The observed striations in the residuals vs fitted plot were confirmed to be artifacts of the discrete, integer-based nature of the raw data rather than model misspecification. The residual distribution confirms that the model appropriately accounts for the data structure.

To examine responses from the poststudy questionnaire, we first created scores for each construct measured by the questionnaire (ie, noticeability, comprehension, relevance,

XSL•FO
RenderX

complexity, division of attention, attractiveness, and usability) by averaging each participant's answers to the questions for that construct. We then fitted linear regression models that used these construct scores as dependent variables and indicators for the design dimensions as predictors.

In line with best practices for factorial experiments [35], models analyzing the stylization and salience design dimensions used effect coding to code design-dimension indicators. With effect coding, 1 level of a factor (eg, high-salience) is coded as +1, and the other level of the factor (eg, low-salience) is coded as –1 in the models. The use of effect coding enables a marginal interpretation of the main effects for each factor: if the β coefficient for a particular design dimension (eg, salience) is found to be significant, that finding can be interpreted as indicating that that design dimension has an impact on the dependent variable (accuracy of progress estimation, usability, etc) averaging across other tested factors and other measured and unmeasured influences.

Eligibility, Recruitment, and Data Integrity Verification

Recruitment was done through the Amazon MTurk platform. We created an MTurk task that linked to the study website that hosted the experiment. Eligible participants were required to be 18 years of age or older and could take part in the experiment only once. Attempts at participating more than once were recognized based on the MTurk worker ID and were filtered out prior to the data analysis.

To ensure high data quality—MTurk workers sometimes click through a task to get paid without actually engaging with the task [36,37]—prior to conducting the experiment, we developed 2 attention checks that we used as criteria for *a priori* exclusion of participants from the dataset. The first check used a metric called string edit distance [38], which measures the similarity of 2 strings in terms of the number of operations (insertions, deletions, etc) that are required to transform one string into the other. In our pilot work, we found that the edit distance of 2 effectively identified responses to the goal-progress estimation that were clearly fallacious (eg, the same number was provided as an answer to all 3 experimental tasks). We, therefore, removed from the analytical data set all participants for whom the edit distance between their estimates of goal progress and the correct answer was greater than or equal to 2. The second exclusion criterion used the attention check question in the poststudy questionnaire (Q10 in Textbox 1). The question asked participants to skip the question and leave it unanswered. Participants who answered the question (ie, who clicked on any radio button on the Likert scale for that question) were excluded from the dataset.

The study protocol was approved by our university's Behavioral and Social Sciences Institutional Review Board.

Results

Participants Sample

We recruited 400 participants on Amazon MTurk, targeting adults who were interested in smartwatches and physical activity. Participants were recruited in 4 waves of 100 participants each. Each wave of recruitment was followed by the exclusion of participants who violated one of the criteria for data integrity validation. We would then start the next wave if the required number of participants had not been reached.

After applying data integrity criteria, 202 participants remained in the final dataset. Of these, 35.64% (n=72) identified as female. In terms of smartwatch usage, 29.2% (n=59) were current users, 38.61% (n=78) were former users, and 32.18% (n=65) had never used a smartwatch. The age distribution was skewed toward younger and middle-aged adults. The largest age group was 25-34 years (n=87, 43.07%), followed by 35-44 years (n=57, 28.21%) and 45-54 years (n=33, 16.34%). Participants aged 18-24 years accounted for 6.93% (n=14) of the sample, while those aged 55-64 years and older comprised 3.47% (n=7) and 1.98% (n=4), respectively.

Stylization Negatively Affects the Accuracy of Progress Assessment

Table 1 presents the results of the GEE analysis examining the effects of salience and stylization on the accuracy of participants' goal-progress estimates. Since the outcome variable was square-root transformed, the reported β coefficients should be squared to interpret effects on the original scale. To make the results interpretable, we note that the median absolute error across all conditions was 14 (IQR 4-47) points. Our findings indicate that stylization significantly reduces the accuracy of progress assessments. Compared to low-stylized representations, high-stylized ones led to an increase in error by 4.52 points, rising from 8.98 for low-stylized conditions to 13.50 points for high-stylized conditions. We found no significant impact for salience, however.

Table 1. Generalized estimating equation model results examining the effects of stylization, salience, age, gender, and user status on the accuracy of progress estimates (square-root transformed outcome).

	Coefficient	SE	*z*	*P* value	95% CI
Intercept	3.335	0.529	6.307	<.001	2.299 to 4.372
Stylization	0.339	0.122	2.764	.006	0.098 to 0.579
Salience	–0.011	0.121	–0.090	.93	–0.248 to 0.226
Age	0.021	0.012	1.783	.07	–0.002 to 0.044
Gender	0.049	0.124	0.391	.70	–0.195 to 0.292
Current user	0.256	0.336	0.763	.45	–0.402 to 0.914
Previous user	0.976	0.287	3.395	.001	0.412 to 1.539

XSL•FO
RenderX

Nonlinear Effects of Granularity on Estimated Absolute Error

Table 2 presents the results of a GEE analysis examining the effects of salience, stylization, and granularity on the accuracy of participants' goal-progress estimates. By including all 3 factors in a single model, we isolated the marginal impact of each variable. In this combined linear model, granularity demonstrated a marginally significant independent effect on estimation error (P=.08). However, because granularity was unevenly distributed across feedback designs—with half of the panels using the finest granularity level (1%)—and preliminary diagnostics suggested a nonlinear relationship, we conducted a follow-up analysis. As reported in Table 3, we used a model specifically parameterized with a quadratic term to account for this complexity. These results revealed a U-shaped relationship: error decreased as granularity *decreased*, up to a point, after which further decreases in granularity led to higher error.

Table 2. Generalized estimating equation model results examining the effects of stylization, salience, granularity, age, gender, and user status on the accuracy of progress estimates (square-root transformed outcome).

	Coefficient	SE	z	*P* value	95% CI
Intercept	3.230	0.541	5.973	<.001	2.170 to 4.290
Stylization	0.311	0.124	2.507	.01	0.068 to 0.554
Salience	–0.007	0.121	–0.056	.96	–0.248 to 0.226
Granularity	0.033	0.019	1.732	.08	–0.004 to 0.070
Age	0.020	0.012	1.652	.10	–0.004 to 0.043
Gender	0.034	0.124	0.272	.79	–0.210 to 0.278
Current user	0.254	0.335	0.759	.45	–0.402 to 0.911
Previous user	0.961	0.283	3.396	.001	0.406 to 1.515

Table 3. Generalized estimating equation model results examining the nonlinear effects of granularity on the accuracy of progress estimates (square-root transformed outcome).

	Coefficient	SE	*z*	*P* value	95% CI
Intercept	4.599	0.214	21.465	<.001	4.179 to 5.019
Granularity	–0.114	0.071	–1.599	.11	–0.254 to 0.026
Granularity squared	0.008	0.003	2.437	.015	0.002 to 0.015

We examined 5 levels of feedback display granularity: 1%, 4%, 5%, 10%, and 20%. Given that the outcome variable was square-root transformed, predicted values were squared to interpret error on the original scale. Decreasing granularity from 1% to 4% reduced the predicted error from 20.19 to 18.24, a decrease of 1.95 points. A further decrease in granularity from 4% to 5% resulted in a smaller reduction of 0.36 points, lowering the predicted error to 17.88 at the 5% granularity level. Transitioning from 5% to 10%, the error began to increase slightly, rising by 0.26 points to 18.14 at 10% granularity. When granularity decreased from 10% to 20%, the predicted error sharply increased by 12.32 points, reaching 30.46. These findings reveal a U-shaped relationship between granularity and error, suggesting that when feedback is presented on smartwatch-sized screens, moderate levels of granularity enhance accuracy, whereas high granularity and overly coarse representations hinder users' ability to estimate goal progress correctly.

According to the model, the optimal granularity level is around 7.1%. Since this value was not directly tested in our feedback designs, it suggests that the lowest predicted error would occur at a granularity level of 5% or 10%.

Stylization Negatively Affects Perception of Comprehension, Relevance, Complexity, Attention, Attractiveness, and Usability Ratings

As shown in Textbox 1, we grouped the subjective perception items from the postquestionnaire into 7 categories: noticeability, comprehension, relevance, complexity, division of attention, attractiveness, and usability. To examine the impact of stylization and salience on participants' subjective perceptions of the feedback displays, we conducted linear regression analyses.

Tables 4-9 present the results of the linear regression analyses examining the effects of stylization and salience on participants' perceptions of the interfaces, including comprehension, relevance, complexity, division of attention, attractiveness, and usability (see Table S1 in Multimedia Appendix 1 for the impact of stylization on noticeability). Perceptions were measured using a 5-point Likert scale across 7 categories. Stylization was found to negatively influence 6 of the 7 dimensions of user perceptions: comprehension, relevance, complexity, attention, attractiveness, and usability. When using high-stylized feedback designs, participants reported greater difficulty comprehending the information (–0.486 points; P=.003), perceived the interface as more complex (–0.690 points; P<.001), and felt that less relevant information was provided (–0.476 points; P=.002).

XSL•FO
RenderX

Additionally, they found it harder to maintain focus or shift attention between the interface and the game (–0.378 points; *P*=.022), and rated the interfaces as less attractive (–0.338 points; *P*=.047) and less usable (–0.532 points; *P*=.001).

Table 4. Linear regression analysis of the effects of stylization and salience on comprehension scores (reference group: people who have never used a smartwatch before).

	Coefficient	SE	*t* test (*df*)	*P* value	95% CI
Intercept	3.415	0.339	10.069 (195)	<.001	2.746 to 4.084
Stylization	–0.243	0.079	–3.060 (195)	.003	–0.400 to –0.086
Salience	0.146	0.078	1.886 (195)	.06	–0.007 to 0.300
Age	–0.005	0.008	–0.701 (195)	.48	–0.021 to 0.010
Gender	0.116	0.082	1.419 (195)	.16	–0.045 to 0.278
Current user	0.317	0.198	1.602 (195)	.11	–0.073 to 0.708
Previous user	0.351	0.190	1.846 (195)	.07	–0.024 to 0.727

Table 5. Linear regression analysis of the effects of stylization and salience on relevance scores (reference group: people who have never used a smartwatch before).

	Coefficient	SE	*t* test (*df*)	*P* value	95% CI
Intercept	3.460	0.321	10.775 (195)	<.001	2.827 to 4.093
Stylization	–0.238	0.075	–3.170 (195)	.002	–0.387 to –0.090
Salience	0.103	0.074	1.402 (195)	.16	–0.042 to 0.248
Age	0.001	0.007	0.198 (195)	.84	–0.013 to 0.016
Gender	0.137	0.078	1.762 (195)	.08	–0.016 to 0.290
Current user	0.222	0.187	1.182 (195)	.24	–0.148 to 0.591
Previous user	0.283	0.180	1.572 (195)	.12	–0.072 to 0.639

Table 6. Linear regression analysis of the effects of stylization and salience on complexity scores (reference group: people who have never used a smartwatch before).

	Coefficient	SE	*t* test (*df*)	*P* value	95% CI
Intercept	3.004	0.393	7.644 (195)	<.001	2.229 to 3.779
Stylization	–0.345	0.092	–3.745 (195)	<.001	–0.526 to –0.163
Salience	0.031	0.090	0.339 (195)	.73	–0.147 to 0.208
Age	0.004	0.009	0.426 (195)	.67	–0.014 to 0.021
Gender	0.084	0.095	0.885 (195)	.38	–0.103 to 0.272
Current user	0.019	0.229	0.084 (195)	.93	–0.433 to 0.472
Previous user	–0.754	0.221	–3.417 (195)	.001	–1.189 to –0.319

Table 7. Linear regression analysis of the effects of stylization and salience on division of attention scores (reference group: people who have never used a smartwatch before).

	Coefficient	SE	*t* test (*df*)	*P* value	95% CI
Intercept	2.750	0.350	7.849 (195)	<.001	2.059 to 3.441
Stylization	–0.189	0.082	–2.309 (195)	.022	–0.351 to –0.028
Salience	0.136	0.080	1.701 (195)	.09	–0.022 to 0.295
Age	0.003	0.008	0.320 (195)	.75	–0.013 to 0.018
Gender	0.104	0.085	1.227 (195)	.22	–0.063 to 0.271
Current user	0.559	0.204	2.733 (195)	.007	0.156 to 0.962
Previous user	0.874	0.197	4.444 (195)	<.001	0.486 to 1.262

XSL•FO
RenderX

Table 8. Linear regression analysis of the effects of stylization and salience on attractiveness scores (reference group: people who have never used a smartwatch before).

	Coefficient	SE	*t* test (*df*)	*P* value	95% CI
Intercept	3.286	0.361	9.093 (195)	<.001	2.573 to 3.998
Stylization	–0.169	0.085	–2.001 (195)	.047	–0.336 to –0.002
Salience	0.007	0.083	0.082 (195)	.93	–0.156 to 0.170
Age	–0.001	0.008	–0.141 (195)	.89	–0.017 to 0.015
Gender	0.074	0.087	0.849 (195)	.40	–0.098 to 0.247
Current user	0.163	0.211	0.773 (195)	.44	–0.253 to 0.579
Previous user	0.367	0.203	1.812 (195)	.07	–0.033 to 0.767

Table 9. Linear regression analysis of the effects of stylization and salience on usability scores (reference group: people who have never used a smartwatch before).

	Coefficient	SE	*t* test (*df*)	*P* value	95% CI
Intercept	3.369	0.324	10.409 (195)	<.001	2.730 to 4.007
Stylization	–0.266	0.076	–3.505 (195)	.001	–0.415 to –0.116
Salience	0.110	0.074	1.487 (195)	.14	–0.036 to 0.256
Age	–0.010	0.007	–1.331 (195)	.18	–0.024 to 0.005
Gender	0.102	0.078	1.301 (195)	.19	–0.053 to 0.256
Current user	0.512	0.189	2.709 (195)	.007	0.139 to 0.884
Previous user	0.566	0.182	3.118 (195)	.002	0.208 to 0.924

10% Granularity Yields the Most Favorable Perceptions

Compared to modeling granularity as a continuous predictor, treating it as a categorical variable provided a clearer understanding of how different granularity levels impacted subjective responses. Tables 10-12 present the results of linear regression analyses examining the effects of granularity on participants' perceptions of the interfaces, including comprehension, complexity, and division of attention (see Tables S2-S5 in Multimedia Appendix 1 for the impact of granularity on noticeability, relevance, attractiveness, and usability).

Table 10. Linear regression analysis of the effects of granularity on comprehension scores (reference group: 1% granularity).

	Coefficient	SE	*t* test (*df*)	*P* value	95% CI
Intercept	3.435	0.110	31.132 (197)	<.001	3.217 to 3.653
Granularity=4%	0.065	0.230	0.283 (197)	.78	–0.388 to 0.518
Granularity=5%	–0.435	0.265	–1.642 (197)	.10	–0.957 to 0.087
Granularity=10%	0.656	0.222	2.961 (197)	.003	0.219 to 1.093
Granularity=20%	–0.046	0.283	–0.163 (197)	.87	–0.603 to 0.511

Table 11. Linear regression analysis of the effects of granularity on complexity scores (reference group: 1% granularity).

	Coefficient	SE	*t* test (*df*)	*P* value	95% CI
Intercept	2.810	0.133	21.067 (197)	<.001	2.547 to 3.073
Granularity=4%	0.223	0.278	0.804 (197)	.42	–0.324 to 0.771
Granularity=5%	–0.429	0.320	–1.340 (197)	.18	–1.060 to 0.202
Granularity=10%	0.766	0.268	2.860 (197)	.005	0.238 to 1.294
Granularity=20%	0.107	0.342	0.312 (197)	.76	–0.567 to 0.780

XSL•FO
RenderX

Table 12. Linear regression analysis of the effects of granularity on division of attention scores (reference group: 1% granularity).

	Coefficient	SE	*t* test (*df*)	*P* value	95% CI
Intercept	3.295	0.117	28.121 (197)	<.001	3.064 to 3.526
Granularity=4%	0.305	0.244	1.250 (197)	.21	–0.176 to 0.786
Granularity=5%	–0.390	0.281	–1.387 (197)	.17	–0.945 to 0.164
Granularity=10%	0.523	0.235	2.224 (197)	.027	0.059 to 0.987
Granularity=20%	0.038	0.300	0.128 (197)	.90	–0.553 to 0.630

Among the 5 granularity levels (1%, 4%, 5%, 10%, and 20%), the 10% granularity level yielded the most favorable perceptions. Relative to the 1% granularity group, participants in the 10% group reported the interface as easier to comprehend (+0.656 points; *P*=.003), less complex (+0.766 points; *P*=.005), and easier to maintain focus on (+0.523 points; *P*=.03).

To further evaluate the performance of the 10% group, we reran the analysis using the 10% group as the reference category (Tables S6-S12 in Multimedia Appendix 1). Consistent with the previous comparison, we found that participants in all other granularity groups rated the feedback designs as harder to comprehend (1%: –0.656; *P*=.003; 4%: –0.591; *P*=.04; 5%: –1.091; *P*<.001; 20%: –0.702; *P*=.03). Participants in the 1% (–0.766 points; *P*=.005) and 5% (–1.195 points; *P*=.002) groups also perceived the feedback designs as more complex compared to the 10% group. Similarly, participants in the 1% (–0.523 points; *P*=.03) and 5% (–0.913 points; *P*=.006) groups found it more difficult to maintain focus, compared to the 10% group. Additionally, compared to the 10% group, participants in the 5% group perceived the feedback design as harder to notice changes in (–0.634 points; *P*=.024), less informative (–0.779 points; *P*=.009), and less useful overall (–0.848 points; *P*=.006).

Effects of Smartwatch Experience on Accuracy and Subjective Perceptions

Prior smartwatch experience was associated with decreased accuracy in progress assessment. As shown in Table 1, compared to people who have never used a smartwatch before, being a previous smartwatch user—but not a current user—increased the error by approximately 7.46 points, with the error rising from 11.13 to 18.59 points (median absolute error 14, IQR 4-47 points).

Despite exhibiting lower accuracy in progress estimation compared to participants who had never used a smartwatch, previous smartwatch users reported more favorable subjective perceptions across several measures. As seen in Tables 6, 8, and 9, compared to participants who had never used a smartwatch, previous smartwatch users found the interface easier to maintain focus on (+0.874 points, *P*<.001) and more useful (+0.566 points, *P*=.002). However, they also perceived the feedback displays as more complex, rating them 0.754 points lower in simplicity *P*=.001). In Table 13, compared to current smartwatch users, previous smartwatch users also perceived the feedback design as harder to notice changes in (–0.773 points, *P*=.001).

Table 13. Linear regression analysis of the effects of stylization and salience on noticeability scores (reference group: current smartwatch user).

	Coefficient	SE	*t* test (*df*)	*P* value	95% CI
Intercept	3.253	0.407	8.003 (195)	<.001	2.452 to 4.055
Stylization	–0.689	0.184	–3.745 (195)	<.001	–1.052 to –0.326
Salience	0.061	0.180	0.339 (195)	.73	–0.294 to 0.416
Age	0.004	0.009	0.426 (195)	.67	–0.014 to 0.021
Gender	0.168	0.190	0.885 (195)	.38	–0.207 to 0.544
Never used user	–0.019	0.229	–0.084 (195)	.93	–0.472 to 0.433
Previous user	–0.773	0.224	–3.449 (195)	.001	–1.215 to –0.331

Current smartwatch users did not show significant differences in accuracy compared to participants who had never used a smartwatch, but reported more positive subjective perceptions. Specifically, as seen in Table S1 in Multimedia Appendix 1, Tables 7 and 9, they rated the feedback designs as easier to notice changes in (+0.354 points; *P*=.053), easier to maintain or shift attention (+0.559 points; *P*=.007), and more useful (+0.512 points; *P*=.007), all on a 5-point scale.

Discussion

Principal Results

Our study examined how variations across 3 design dimensions—stylization, salience, and granularity—influenced participants' ability to interpret glanceable smartwatch feedback displays.

Results showed that high stylization negatively affected the accuracy of progress assessments. Regarding granularity, levels of 5% and 10% could be seen as a sweet spot, resulting in lower interpretation errors. Participants in the 10% granularity group

XSL•FO
RenderX

also reported more favorable subjective perceptions compared to those in other granularity conditions. Salience did not influence either the accuracy of interpreting progress or participants' subjective perceptions.

Interestingly, participants who had previously used a smartwatch but were no longer actively using one performed worse in assessing progress compared to those who had never used a smartwatch. However, both previous and current smartwatch users reported more positive subjective evaluations of feedback displays than participants who had never used a smartwatch.

Increased Interpretation Errors Caused by Stylization Present Design Trade-Offs

Our results indicate that high stylization negatively affected interpretation accuracy, increasing progress errors by 4.52 points. This finding should be considered alongside several factors that may influence the real-world viability of stylized displays. First, the observed accuracy gap might narrow with increased familiarity; while our study used a brief practice period, long-term use of a specific stylized interface could result in users learning to better interpret its visual patterns, potentially improving performance over time. Second, stylization may serve as a privacy-preserving layer, offering users a degree of discretion by preventing onlookers from easily discerning personal data. Third, the aesthetic qualities of such interfaces might boost user engagement and check frequency, which could enhance overall effectiveness [22,39,40].

In contrast, low-stylized displays are preferred when feedback precision is critical for decision-making. This is particularly relevant in safety-critical contexts or data-driven health management, such as monitoring blood glucose levels [41] or tracking caloric balance. In these scenarios, the requirement for high levels of precision makes stylized feedback a suboptimal choice. Ultimately, the use of stylization involves a trade-off between the necessity for interpretative accuracy and the user's potential preference for discretion and aesthetic engagement.

Finding the Sweet Spot in Granularity for Glanceable Displays

We found that the relationship between granularity and accuracy in progress assessment is nonlinear. Accuracy initially increases as granularity decreases (eg, from 1% to 5%), but the rate of increase slows down between 5% and 10% and then reverses direction. This pattern indicates the presence of an inflection point—an optimal granularity level, estimated at 5%-10% in our data—at which accuracy peaks. Beyond this point, accuracy declines as granularity increases further.

These results are not completely unexpected, as there are situations where high levels of feedback granularity may not be helpful. For instance, in very small feedback displays, such as smartwatch complications, feedback granularity can quickly exceed users' ability to visually discriminate between states of the display. In such situations, a lower-granularity display may more effectively convey changes in the state through more easily perceived jumps than a high-granularity display. Similarly, in many contexts, there are limits to the level of granularity that is needed for decisions. For instance, Hawley et al [16] and Zikmund-Fisher et al [20] found that variations in the granularity of the graphs about the risk of taking different medications did not have a significant impact on people's choices of which medicine to take. The researchers used n-in-100 (ie, among 100 people, the number of people who have side effects) as low granularity and n-in-1000 as high granularity. Their results indicated that when the level of granularity met the upper limit of the demand, participants' assessments became unaffected by further increases in granularity.

Our result shows that the sweet spot for goal-progress feedback displays may be at 5% or 10% granularity, which aligns with previous research. For instance, Kong et al [42] observed that participants estimating the proportion between 2 rectangles (displayed in a constrained space) tended to round their answers to multiples of 5. Our results showed the same tendency—participants frequently reported progress estimates in multiples of 5. This suggests that interfaces designed with 5% or 10% granularity may support faster and more accurate assessments, as users can more easily map perceived progress to familiar reference points.

Taken together, these results suggest that pursuing very high granularity does not necessarily improve users' ability to notice changes or make informed decisions. Conversely, very low granularity may increase errors in progress assessment and fall short in contexts where greater precision is required [41]. A granularity level of 5% or 10% appears to strike an effective balance—supporting precision, enabling users to notice changes, and reducing estimation errors.

Prior Use Followed by Abandonment May Reflect Interpretation Difficulties

Lastly, our results indicate that previous smartwatch users made more errors in perceiving progress compared to those who had never used a smartwatch. Specifically, their error was 7.46 points higher—representing the largest effect on interpretation accuracy observed in our study. While we are not sure why this discrepancy exists, one intriguing possibility is that people who are less effective at interpreting information displayed on smartwatches are more likely to abandon them. This possibility might be reflected in participants' subjective perceptions. Compared to individuals who have never used a smartwatch, previous smartwatch users rated the feedback display as more complex (–0.754 points; P=.001; Table 6). Even when compared to current smartwatch users, previous smartwatch users still perceived the feedback display as more complex (–0.773 points; P=.001; Table 13). However, as shown in Table 7 and Table 9, previous smartwatch users also found the interface more useful (+0.566 points; P=.002; Table 9) and reported it was easier to maintain focus while using it (+0.874 points; P<.001; Table 7). Current smartwatch users similarly rated the feedback display as more useful (+0.512 points; P=.007; Table 9) and easier to focus on (+0.559 points; P=.007; Table 7). Notably, current users did not report any statistically significant negative perceptions of the feedback display interface.

If this is the case, researchers who develop smartwatch-based interventions may want to use prior smartwatch use as a tailoring variable for providing additional training or using alternative forms of intervention feedback. Investigating how predictive

XSL•FO
RenderX

this variable is for other intervention effects could be an interesting avenue for future research.

Challenges in Operating Design Dimensions From Theories to Practices

Our experience in constructing these interfaces suggests a broader challenge regarding how to cleanly operationalize design dimensions when transitioning from theories to practical applications. While design dimensions such as granularity, stylization, and salience are conceptually distinct, their implementation in a real-world context like smartwatch displays may be inherently constrained by practical requirements.

For instance, the tension between exact estimations and low-granularity designs reflects the trade-off between accuracy and interpretability. For glanceable displays, while providing accurate data is important, designs need to remain interpretable to be efficiently accessed under cognitive load. While asking participants to estimate the range of progress might have aligned more naturally with low-granularity interfaces, using an exact estimation task across all conditions allowed us to observe how different designs influence this cognitive balance. This indicates that practical constraints can make it difficult to cleanly separate a single design dimension from its functional impact on performance. Another operationalization challenge for stylization and granularity is further discussed in the Limitations section.

These observations suggest that design dimensions that are theoretically distinct are often not orthogonal in practice. Operationalizing a single dimension in isolation, without accounting for these functional interdependencies, risks producing interfaces that are either visually incoherent or practically irrelevant.

Limitations

Methodological Limitations

As we note in the Method section, the main trade-off of our decision to use an online experiment involves sacrificing some ecological validity for the increased statistical rigor and feasibility of studying many design variations for feedback displays. One other notable limitation of the study is that it did not allow us to study additional questions about the influence of time and attentional context on display interpretation. In real-world use, over time, users are likely to learn how to interpret any (reasonable) type of feedback display. So, it is possible that the differences in interpretation accuracy we found in our study for unfamiliar displays would gradually decrease with repeated use. Similarly, how accurately a person interprets a feedback display at any given time is likely to have to do in part with why they are looking at the watch. If the person is in a hurry and just wants to see what time it is, they may not notice the feedback display at all. If they are trying to check goal progress, they would pay closer attention to the display. While our study was closer to the latter setting, carefully studying how exactly time and context affect feedback interpretation would require a different study design. Whether a simulation could still be used to study those processes is an interesting question that would be worthwhile for design researchers to consider. The answer will likely depend on the exact processes that are being studied and whether a simulation could incorporate the key factors hypothesized to influence those processes. Determining this involves thinking carefully about the causal processes that underlie the phenomenon of interest.

Design Space Constraints

Another limitation of this work concerns the interdependence of the design dimensions in practice. While we conceptually defined granularity, stylization, and salience as distinct factors, they were not always perfectly decoupled in our experimental stimuli. For instance, changes in granularity were occasionally tied to shifts in representation type, such as transitioning from a circular progress ring to a segmented bar chart to accommodate higher data resolution.

This confounding reflects a considered trade-off between experimental control and the practical applicability of the stimuli. Given the exploratory nature of mapping this design space, we aimed to present participants with functional and coherent interfaces that mirror the types of glanceable displays that can be used in everyday contexts. Strictly isolating a single dimension by increasing granularity without adjusting the visual representation could have resulted in artifacts that were visually awkward or practically illegible. While this approach prioritized the holistic integrity of the designs, we acknowledge that it limits our ability to isolate the independent impact of each individual dimension. Future work could build on these findings by employing more constrained experimental setups to further disentangle these interconnected factors.

Conclusions

In this paper, we presented a study that tested 16 glanceable feedback displays to examine the effects of 3 design dimensions—stylization, salience, and granularity—on the accuracy of progress assessment. Our findings show that stylization negatively affected both the accuracy of assessing goal progress and perceptions of comprehension, relevance, complexity, attention, attractiveness, and usability. In contrast, granularity levels of 5% and 10% may be optimal for displaying smartwatch feedback. We found no significant differences between high-salience and low-salience interfaces in terms of interpretive accuracy and subjective perceptions. Interestingly, participants who had previously used a smartwatch but were no longer using one exhibited larger errors in interpreting progress feedback. However, both previous and current smartwatch users reported higher subjective perceptions compared to those who had never used a smartwatch. Our results extend the research on glanceable feedback and can help researchers design effective glanceable smartwatch feedback displays.

XSL•FO
RenderX

Acknowledgments

We used the generative AI tool ChatGPT to improve grammar and refine the writing style. We ensured that the integrity and originality of the content were maintained throughout.

Data Availability

The datasets analyzed during this study are available from the corresponding author upon reasonable request.

Funding

This study was funded by internal university funds.

Authors' Contributions

Conceptualization: YL (equal), MWN, (equal), PK (equal)
Investigation: YL
Data analysis: YL (lead), PK (supporting)
Visualization: YL (lead), MWN (supporting), PK (supporting)
Writing—original draft: YL (lead), MWN (supporting), PK (supporting)
Writing—review & editing: PK (lead), YL (supporting), MWN (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Model diagnostics and secondary tables for the findings.
[DOCX File , 276 KB - mhealth_v14i1e81972_app1.docx]

References

1. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
2. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
3. Murray JM, Brennan SF, French DP, Patterson CC, Kee F, Hunter RF. Effectiveness of physical activity interventions in achieving behaviour change maintenance in young and middle aged adults: a systematic review and meta-analysis. Soc Sci Med 2017;192:125-133. [doi: 10.1016/j.socscimed.2017.09.021] [Medline: 28965003]
4. Richardson CR, Newton TL, Abraham JJ, Sen AA, Jimbo M, Swartz AM. A meta-analysis of pedometer-based walking interventions and weight loss. Ann Fam Med 2008;6(1):69-77 [FREE Full text] [doi: 10.1370/afm.761] [Medline: 18195317]
5. Choe EK, Consolvo S, Watson NF, Kientz JA. Opportunities for computing technologies to support healthy sleep behaviors. 2011 Presented at: CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; May 7, 2011; Vancouver, BC, Canada p. 3053-3062. [doi: 10.1145/1978942.1979395]
6. Froehlich J, Dillahunt T, Klasnja P, Mankoff J, Consolvo S, Harrison B, et al. UbiGreen: investigating a mobile tool for tracking and supporting green transportation habits. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA: Association for Computing Machinery; 2009:1043-1052.
7. Preuveneers D, Berbers Y. Mobile phones assisting with health self-care: a diabetes case study. 2008 Presented at: Proceedings of the 10th International Conference on Human Computer Interaction With Mobile Devices and Services; September 2-5, 2008; Amsterdam, The Netherland p. 177-186. [doi: 10.1145/1409240.1409260]
8. Siek KA, Connelly KH, Rogers Y, Lambert D, Welch JL. When do we eat? An evaluation of food items input into an electronic food monitoring application. 2006 Presented at: 2006 Pervasive Health Conference and Workshops; November 29-December 1, 2006; Innsbruck, Austria p. 1-10. [doi: 10.1109/PCTHEALTH.2006.361684]
9. Pousman Z, Stasko J. A taxonomy of ambient information systems: four patterns of design. 2006 Presented at: Proceedings of the Working Conference on Advanced Visual Interfaces; May 23-26, 2006; Venice (Venezia), Italy p. 67-74. [doi: 10.1145/1133265.1133277]
10. Consolvo S, Klasnja P, McDonald DW, Avrahami D, Froehlich J, LeGrand L, et al. 2008 Presented at: Proceedings of the 10th International Conference on Ubiquitous Computing; September 21, 2008; New York, NY, USA p. 54-63. [doi: 10.1145/1409635.1409644]
11. Gouveia R, Pereira F, Karapanos E, Munson SA, Hassenzahl M. Exploring the design space of glanceable feedback for physical activity trackers. : Association for Computing Machinery; 2016 Presented at: Proceedings of the 2016 ACM

XSL•FO
RenderX

International Joint Conference on Pervasive and Ubiquitous Computing; September 12, 2016; New York, NY, USA p. 144-155. [doi: 10.1145/2971648.2971754]
12. Lane N, Rabbi M, Lin M, Yang X, Lu H, Ali S, et al. BeWell: a smartphone application to monitor, model and promote wellbeing. 2011 Presented at: Proceedings of the 5th International ICST Conference on Pervasive Computing Technologies for Healthcare; May 23-26, 2011; Dublin, Ireland p. 23-36. [doi: 10.4108/icst.pervasivehealth.2011.246161]
13. Collins LM. Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: The Multiphase Optimization Strategy (MOST). In: Statistics for Social and Behavioral Sciences. Cham, Switzerland: Springer International Publishing; 2018.
14. Czajkowski SM, Powell LH, Adler N, Naar-King S, Reynolds KD, Hunter CM, et al. From ideas to efficacy: the ORBIT model for developing behavioral treatments for chronic diseases. Health Psychol 2015;34(10):971-982 [FREE Full text] [doi: 10.1037/hea0000161] [Medline: 25642841]
15. Hekler EB, Klasnja P, Riley WT, Buman MP, Huberty J, Rivera DE, et al. Agile science: creating useful products for behavior change in the real world. Transl Behav Med 2016;6(2):317-328 [FREE Full text] [doi: 10.1007/s13142-016-0395-7] [Medline: 27357001]
16. Hawley ST, Zikmund-Fisher B, Ubel P, Jancovic A, Lucas T, Fagerlin A. The impact of the format of graphical presentation on health-related knowledge and treatment choices. Patient Educ Couns 2008;73(3):448-455. [doi: 10.1016/j.pec.2008.07.023] [Medline: 18755566]
17. McCaffery KJ, Dixon A, Hayen A, Jansen J, Smith S, Simpson JM. The influence of graphic display format on the interpretations of quantitative risk information among adults with lower education and literacy: a randomized experimental study. Med Decis Making 2012;32(4):532-544 [FREE Full text] [doi: 10.1177/0272989X11424926] [Medline: 22074912]
18. Witteman HO, Fuhrel-Forbis A, Wijeysundera HC, Exe N, Dickson M, Holtzman L, et al. Animated randomness, avatars, movement, and personalization in risk graphics. J Med Internet Res 2014;16(3):e80-e77 [FREE Full text] [doi: 10.2196/jmir.2895] [Medline: 24642037]
19. Zikmund-Fisher BJ, Dickson M, Witteman HO. Cool but counterproductive: interactive, web-based risk communications can backfire. J Med Internet Res 2011;13(3):e60-e31 [FREE Full text] [doi: 10.2196/jmir.1665] [Medline: 21868349]
20. Zikmund-Fisher BJ, Ubel PA, Smith DY, Derry HA, McClure JB, Stark A, et al. Communicating side effect risks in a tamoxifen prophylaxis decision aid: the debiasing influence of pictographs. Patient Educ Couns 2008;73(2):209-214 [FREE Full text] [doi: 10.1016/j.pec.2008.05.010] [Medline: 18602242]
21. Zikmund-Fisher BJ, Witteman HO, Fuhrel-Forbis A, Exe NL, Kahn VC, Dickson M. Animated graphics for comparing two risks: a cautionary tale. J Med Internet Res 2012;14(4):e106-e551 [FREE Full text] [doi: 10.2196/jmir.2030] [Medline: 22832208]
22. Consolvo S, Klasnja P, McDonald DW, Landay JA. Designing for healthy lifestyles: design considerations for mobile technologies to encourage consumer health and wellness. Found Trends Hum Comput Interact 2014;6(3-4):167-315. [doi: 10.1561/1100000040]
23. Blascheck T, Besancon L, Bezerianos A, Lee B, Islam A, He T, et al. Studies of part-to-whole glanceable visualizations on smartwatch faces. 2023 Presented at: 2023 IEEE 16th Pacific Visualization Symposium (PacificVis); 4/2023; Seoul, Korea, Republic of p. 187-196. [doi: 10.1109/PacificVis56936.2023.00028]
24. Islam A, Aravind R, Blascheck T, Bezerianos A, Isenberg P. Preferences and effectiveness of sleep data visualizations for smartwatches and fitness bands. 2022 Presented at: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems; April 28, 2022; New York, NY, USA p. 1-17 URL: https://dl.acm.org/doi/10.1145/3491102.3501921 [doi: 10.1145/3491102.3501921]
25. While Z, Blascheck T, Gong Y, Isenberg P, Sarvghad A. Glanceable data visualizations for older adults: establishing thresholds and examining disparities between age groups. 2024 Presented at: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems; May 11-16, 2024; Honolulu, Hawaii, USA p. 1-17. [doi: 10.1145/3613904.3642776]
26. Consolvo S, McDonald DW, Toscos T, Chen MY, Froehlich J, Harrison B, et al. Activity sensing in the wild: a field trial of UbiFit Garden. 2008 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 6, 2008; New York, NY, USA p. 1797-1806. [doi: 10.1145/1357054.1357335]
27. Neshati A, Sakamoto Y, Leboe-McGowan LC, Leboe-McGowan J, Serrano M, Irani P. G-Sparks: glanceable sparklines on smartwatches. 2019 Presented at: Proceedings of the 45th Graphics Interface Conference on Proceedings of Graphics Interface 2019; June 3, 2019; Waterloo, CAN p. 1-9. [doi: 10.20380/GI2019.23]
28. Schiewe A, Krekhov A, Kerber F, Daiber F, Krüger J. A study on real-time visualizations during sports activities on smartwatches. 2020 Presented at: 19th International Conference on Mobile and Ubiquitous Multimedia; November 22-25, 2020; Essen, Germany p. 18-31. [doi: 10.1145/3428361.3428409]
29. Islam A, He T, Bezerianos A, Blascheck T, Lee B, Isenberg P. Visualizing information on smartwatch faces: a review and design space. Inf Vis 2026;25(1):21-40. [doi: 10.1177/14738716251360339]
30. Lee J, Wickens C, Boyle L. Designing for People: An Introduction to Human Factors Engineering. Charleston, SC, USA: CreateSpace Independent Publishing Platform; 2017.
31. Still J, Still M. Influence of visual salience on webpage product searches. ACM Trans Appl Percept 2019;16(1):3:1-3:11. [doi: 10.1145/3301413]

XSL•FO
RenderX

32. Matthews T. Designing and evaluating glanceable peripheral displays. 2006 Presented at: Proceedings of the 6th Conference on Designing Interactive Systems; June 26-28, 2006; University Park, Pennsylvania, USA p. 343-345. [doi: 10.1145/1142405.1142457]
33. Brooke J. SUS: a quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester BA, editors. Usability Evaluation in Industry. London, UK: Taylor & Francis; 1995:189-194.
34. Shami NS, Leshed G, Klein D. Context of use evaluation of peripheral displays (CUEPD). 2005 Presented at: INTERACT 2005 (10th IFIP TC13 International Conference on Human-Computer Interaction); September 12-16, 2005; Rome, Italy p. 579-587. [doi: 10.1007/11555261_47]
35. Collins LM, Dziak JJ, Kugler KC, Trail JB. Factorial experiments: efficient tools for evaluation of intervention components. Am J Prev Med 2014;47(4):498-504 [FREE Full text] [doi: 10.1016/j.amepre.2014.06.021] [Medline: 25092122]
36. Goodman J, Cryder CE, Cheema A. Data collection in a flat world: the strengths and weaknesses of Mechanical Turk samples. Behav Decis Mak 2012;26(3):213-224. [doi: 10.1002/bdm.1753]
37. Ibarra JL, Agas JM, Lee M, Pan JL, Buttenheim AM. Comparison of online survey recruitment platforms for hard-to-reach pregnant smoking populations: feasibility study. JMIR Res Protoc 2018;7(4):e101-e455 [FREE Full text] [doi: 10.2196/resprot.8071] [Medline: 29661751]
38. Levenshtein V. Binary codes capable of correcting deletions, insertions, and reversals. Sov Phys Dokl 1965:707-710 [FREE Full text]
39. Feter N, Dos Santos TS, Caputo EL, da Silva MC. What is the role of smartphones on physical activity promotion? A systematic review and meta-analysis. Int J Public Health 2019;64(5):679-690 [FREE Full text] [doi: 10.1007/s00038-019-01210-7] [Medline: 30758514]
40. Murnane EL, Jiang X, Kong A, Park M, Shi W, Soohoo C, et al. Designing ambient narrative-based interfaces to reflect and motivate physical activity. 2020 Presented at: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu, Hawaii, USA p. 1-14 URL: https://europepmc.org/abstract/MED/33880463 [doi: 10.1145/3313831.3376478]
41. Bailey K, Little J, Jung ME. Self-monitoring using continuous glucose monitors with real-time feedback improves exercise adherence in individuals with impaired blood glucose: a pilot study. Diabetes Technol Ther 2016;18(3):185-193. [doi: 10.1089/dia.2015.0285] [Medline: 26885934]
42. Kong N, Heer J, Agrawala M. Perceptual guidelines for creating rectangular Treemaps. IEEE Trans Vis Comput Graph 2010;16(6):990-908. [doi: 10.1109/TVCG.2010.186] [Medline: 20975136]

Abbreviations

GEE: generalized estimating equation
MTurk: Mechanical Turk

Edited by F Clemente; submitted 06.Aug.2025; peer-reviewed by X Su, V Gupta, R Gouveia; comments to author 06.Mar.2026; accepted 16.Jun.2026; published 23.Jul.2026.

Please cite as:
Li Y, Newman M, Klasnja P
Exploring the Design Space of Glanceable Smartwatch Feedback Displays: Experimental Study
JMIR Mhealth Uhealth 2026;14:e81972
URL: https://mhealth.jmir.org/2026/1/e81972
doi:10.2196/81972
PMID:

©Yuxuan Li, Mark Newman, Predrag Klasnja. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 23.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Configurations of Quality Dimensions Associated With Researcher and Patient Ratings of HIV/AIDS Mobile Health Apps in China: Fuzzy-Set Qualitative Comparative Analysis

Peng Liu[1], MSM; Fuzhi Wang[1], PhD; Changzheng Yang[2], DBA

[1]School of Medical Informatics and Engineering, Bengbu Medical University, Bengbu, China

[2]School of Journalism and Communication, Central China Normal University, Wuhan, Hubei, China

Corresponding Author:
Peng Liu, MSM
School of Medical Informatics and Engineering
Bengbu Medical University
2600 Donghai Avenue, Anhui
Bengbu, 233030
China
Phone: 86 18609623699
Email: 0200210@bbmu.edu.cn

Abstract

Background: Mobile health (mHealth) apps have become important tools for HIV/AIDS management. Although the Mobile App Rating Scale and its user version are widely used, how different stakeholder groups combine quality dimensions into overall judgments remains unclear. Discrepancies between researcher and patient evaluations may affect app adoption and clinical effectiveness.

Objective: This exploratory study aimed to compare configurational pathways associated with high subjective quality ratings of HIV/AIDS apps between researchers and patients using fuzzy-set qualitative comparative analysis.

Methods: Twenty-one Chinese HIV/AIDS mobile apps from early 2023 were evaluated by 3 researchers using the original Mobile App Rating Scale and 2 HIV-positive patients using the user version of the Mobile App Rating Scale. Six conditions were examined: engagement, functionality, aesthetics, information quality, privacy protection, and developer type. Necessary and sufficient configurations were identified for both groups. Wilcoxon signed-rank tests compared raw ratings between groups. Sensitivity analyses covered 4 calibration anchor sets, 4 raw-consistency cutoffs, 4 proportional reduction in inconsistency cutoffs, and multivalue operationalization of privacy protection.

Results: Researcher and patient ratings differed significantly for functionality (median difference 0.50, IQR 0.00 to 0.63; P=.003) and engagement (−0.20, IQR −0.40 to 0.10; P=.04), but not for aesthetics (P=.16), information (P=.07), or subjective quality (P=.06). Five sufficient configurations were identified for researchers (solution consistency=0.981, coverage=0.649) and 4 for patients (solution consistency=0.998, coverage=0.796). Configurations including privacy protection appeared only among researchers; its absence appeared in 3 patient configurations, reflecting the rarity of protection in the sample (5/21 apps) rather than lower patient ratings for protected apps. For patients, functionality (0.907) and information quality (0.915) exceeded the necessity threshold but were strongly correlated (r=0.816), so functionality was retained on substantive grounds rather than as an independent necessary condition. Illustrative cases were Wu Ai Fang Hua and Ai Yi Jian for the researcher-oriented pathway and Chabei for the patient-oriented pathway. Sensitivity analyses left the configurations unchanged across all anchor sets and raw-consistency cutoffs; 1 researcher configuration did not survive the stricter proportional reduction in inconsistency cutoffs, and the patient necessity result was anchor-dependent.

Conclusions: The 2 evaluator panels' assessments of HIV/AIDS mHealth apps appeared to follow different configurational logic in this sample, with several equifinal pathways to high subjective quality in each group. Because the patient panel comprised 2 evaluators, and 21 apps were analyzed with 6 conditions, the findings should be read as hypothesis-generating rather than representative of people living with HIV. The findings nonetheless suggest that both expert and end-user perspectives are needed during mHealth development, and that, because privacy protection was rare and not required for a favorable patient judgment, its provision in a population highly vulnerable to disclosure-related harm needs to be ensured by developers and regulators rather than left to user demand.

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2026;14:e88896) doi:10.2196/88896

KEYWORDS

HIV/AIDS; mobile health; mHealth; app quality; fsQCA; patient evaluation; MARS; uMARS; qualitative comparative analysis; fuzzy-set qualitative comparative analysis; Mobile App Rating Scale; user version of Mobile App Rating Scale

Introduction

Background

The global HIV/AIDS epidemic continues to pose significant public health challenges, with approximately 39 million people living with HIV worldwide as of 2022 [1]. Mobile health (mHealth) apps have emerged as innovative tools to support HIV prevention, treatment adherence, and self-management [2,3]. These digital interventions offer potential benefits including improved medication adherence, enhanced patient-provider communication, and increased access to health information [4,5]. Health app use is now widespread, but the behavioral functionality actually implemented in these apps varies widely [6,7].

The quality of mHealth apps is crucial for their effectiveness and user adoption [8,9]. Various stakeholders, including researchers, health care providers, and patients, evaluate app quality from different perspectives, potentially leading to divergent assessments [10]. Researchers typically use standardized evaluation frameworks emphasizing technical functionality, evidence-based content, and privacy protection [11,12]. In contrast, patients may prioritize usability, engagement features, and personal relevance [13,14].

The Mobile App Rating Scale (MARS) has become a widely adopted tool for assessing mHealth app quality across 4 dimensions: engagement, functionality, aesthetics, and information quality [15]. The user version of the Mobile App Rating Scale (uMARS) was subsequently developed to enable end users to evaluate apps from their own perspective [16,17]. While both versions provide comprehensive evaluation frameworks, they do not explicitly address how different evaluator groups weight these dimensions or how they combine to produce overall quality judgments. In HIV/AIDS care, this omission is consequential, because apps in this field must simultaneously deliver clinical utility and protect a health status that carries a high risk of stigma and discrimination [18,19].

Theoretical Framework

Overview

Three complementary theoretical perspectives motivate the comparison undertaken here and were used to select conditions and to interpret results.

Task-Technology Fit and Technology Acceptance

Task-technology fit theory holds that a technology improves performance to the extent that its capabilities match the requirements of the user's task [20]. The unified theory of acceptance and use of technology (UTAUT) and its consumer extension (UTAUT2) further specify performance expectancy, effort expectancy, and hedonic motivation as proximal determinants of use intention [21,22]. For a person living with HIV, the salient tasks are concrete: booking an anonymous test, obtaining reliable information, tracking medication. Performance expectancy and effort expectancy therefore map closely onto the MARS functionality and engagement dimensions and predict that these should dominate patient judgments. Researchers, who rate apps without a personal task to accomplish, are instead trained to apply normative criteria, so their judgments should weight information quality and safeguards more heavily.

Privacy Calculus and the Privacy Paradox

Privacy calculus theory models disclosure as a trade-off in which anticipated benefits are weighed against perceived risk [23]. The trade-off is systematically asymmetric when risks are abstract and delayed while benefits are immediate and concrete, producing the well-documented gap between stated privacy concern and observed behavior [24]. This framework predicts that privacy protection will function as a normative criterion for expert raters but will be discounted by end users, and it makes the direction of any observed divergence theoretically interpretable rather than merely descriptive.

eHealth Literacy and the Evaluator Effect

eHealth literacy encompasses the skills required to appraise and apply digital health information [25]. Features such as data-minimization practices and consent mechanisms are largely invisible in the interface and require appraisal skills that end users typically lack; visible features such as layout and interactivity do not. Independently of literacy, the usability literature documents an "evaluator effect": trained evaluators applying the same instrument to the same artifact reach substantially different conclusions [26]. Together these perspectives imply that expert and user ratings should not be treated as interchangeable estimates of a single latent quality, which is the assumption implicitly embedded in current app-recommendation practice.

Configurational Logic

Each of the 3 perspectives outlined in the Theoretical Framework section describes a combination rather than an isolated effect: fit depends on functionality and content matching a task; the privacy calculus depends on risk relative to benefit. Conventional regression estimates the average, independent, symmetric contribution of each predictor [27,28] and cannot represent such conjunctions. Set-theoretic methods represent them directly through equifinality, conjunctural causation, and asymmetry [29-32].

Gap in Knowledge

Despite growing recognition of the importance of patient perspectives in mHealth evaluation, few studies have systematically compared how researchers and patients evaluate app quality using configuration-based approaches [17]. Previous research has examined individual predictors of app quality ratings [33,34] but has not addressed: (1) whether multiple

XSL•FO
RenderX

combinations of conditions are associated with high quality ratings (equifinality), (2) whether the same condition contributes differently depending on the conditions it is combined with (conjunctural causation), and (3) how evaluation patterns differ between professional researchers and end users. Studies in the HIV/AIDS app domain specifically have quantified mean quality scores and interrater agreement [35-39] but have not examined how dimensions combine. Furthermore, HIV/AIDS populations face disease-related stigma, acute privacy concerns about sensitive health information, and specific self-management needs that may shape their quality perceptions differently from general user populations [18,19,40].

Study Objectives

This exploratory study aimed to: (1) identify the configurational pathways associated with high subjective quality ratings in HIV/AIDS apps as evaluated by researchers and by patients; (2) compare necessary and sufficient conditions between the two evaluator groups; (3) examine the degree of alignment or divergence in evaluation patterns; and (4) derive design-relevant propositions for HIV/AIDS app development that address both normative rigor and the user-side priorities observed in this panel. Given the panel and case numbers described in the Evaluator Panels and Sample Size Considerations section, the objective is to generate structured hypotheses about how quality judgments are formed, not to estimate population-level parameters.

Methods

Study Design

Overview

This cross-sectional study used a comparative evaluation design to assess HIV/AIDS mobile apps. Two distinct evaluator groups were recruited: (1) professional researchers with expertise in medical informatics, computer science, and social medicine; and (2) individuals living with HIV/AIDS. The study was conducted between February and May 2023, following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for systematic app selection [41]. The ratings analyzed here were collected for, and descriptively reported in, a previous quality evaluation of the same 21 apps [35]; the present study is a secondary, configurational analysis of that dataset, but the fuzzy-set qualitative comparative analysis (fsQCA), Wilcoxon, and sensitivity analyses reported here are new.

Unit of Analysis and Human Participation

Two distinct matters are addressed separately here. First, the unit of analysis is the mobile app: the 21 apps are the cases on which every datum is indexed, and all calibration, necessity, sufficiency, and coverage parameters refer to sets of apps rather than to sets of people. The evaluators are not cases; within this design each evaluator functions as a measurement instrument applied to every app, which is why evaluator reliability is assessed as interrater agreement rather than treated as sampling error. Second, and separately, 5 individuals produced those measurements, 2 of them people living with HIV. The first point concerns how the ratings are analyzed; the second concerns what was owed to the individuals who produced them. The first is set out in the Evaluator Panels and Sample Size Considerations section, and the second in the Recruitment and Eligibility, Data Collected, Basis on Which Prospective Ethics Review Was Not Sought, and Protections Applied sections.

Recruitment and Eligibility

The 2 patient evaluators were recruited entirely through the HIV prevention and control team of the local Center for Disease Control and Prevention (CDC), which is the institution that maintains the follow-up records for people living with HIV in this region. The research team specified the eligibility criteria; the CDC team applied them within its own records, identified individuals who met them, and invited participation. The research team did not receive the follow-up register, did not screen or select individuals, and did not at any stage collect, request, verify or record the serostatus of either evaluator, nor their names, contact details, age, sex, place of residence or any other personal information. Serostatus was therefore never an item of information held by the research team: it was a condition applied by the CDC within data the CDC already held for its own statutory purposes. Written informed consent was obtained from both evaluators before participation; in order to protect their privacy, the consent process was conducted by the CDC team, which collected and retained the signed consent forms on behalf of the study, so that no document bearing a participant's name passed to the research team. What the research team received were completed rating forms, identified by a study code.

Data Collected

The data generated in this study were MARS and uMARS scores for apps that are publicly listed in Chinese app stores and in WeChat. No biological sample, medical record, clinical measurement, health-record extract, or laboratory result was obtained from any evaluator, and no demographic, behavioral, or health variable describing an evaluator entered the dataset or any analysis. The research team therefore collected no personal information of any kind about the evaluators; the only data collected were ratings of publicly available software, held under study codes on an encrypted institutional server.

Basis on Which Prospective Ethics Review Was Not Sought

Prospective ethics review was not applied for, and the basis for that is stated here so that readers can assess it. The applicable instrument is the Measures for Ethical Review of Life Science and Medical Research Involving Humans, issued jointly by the National Health Commission, the Ministry of Education, the Ministry of Science and Technology, and the National Administration of Traditional Chinese Medicine (Guo Wei Ke Jiao Fa [2023] No. 4) [42] and in force from 18 February 2023, that is, for the whole of the evaluation period. Article 3 defines the scope of the Measures as research conducted with human beings as subjects, or conducted using human biological samples or personal information data, including health records and behavioral data; Article 50 defines a research participant correspondingly as a person who provides biological samples, information data, health records or behavioral data for the

XSL•FO
RenderX

research. On both definitions, this study lies outside the scope of the Measures: no human being was a subject of investigation, no biological sample was used, and no personal information data, health record, or behavioral record of any evaluator was provided to, held by, or analyzed by the research team. What the evaluators supplied were ratings of external artifacts, which are properties of the apps rather than information about the raters. Alternatively, were the ratings to be regarded as information data supplied by identifiable individuals, Article 32(2) provides that research using anonymized information data may be exempt from ethics review where it causes no bodily harm and involves neither sensitive personal information nor commercial interest; each condition is satisfied here, since ratings were coded at source, no procedure was applied to any person, no sensitive personal information reached the research team, and the study had neither commercial sponsorship nor commercial interest. The risk to participants was no greater than the minimal-risk threshold at which Article 31 permits review by expedited procedure.

This reporting position is the prevailing one in the MARS/uMARS app-quality literature. App-quality evaluations of this design are routinely published without prospective ethics review on the ground that the object of evaluation is publicly available software rather than persons [43], and the same holds where lay end users rather than experts complete the instrument: in the Greek validation of the uMARS, in which 91 informed participants each rated an app twice, the authors reported that ethics approval was not required because the study consisted of completion of the uMARS and involved no clinical procedure [44].

Protections Applied

Although the study lies outside the scope of mandatory review, the protections owed to participants were applied in full and in accordance with the principles of the Declaration of Helsinki. Both patient evaluators were informed of the purpose, procedures, and time commitment of the study and of their right to withdraw at any point without giving a reason and without consequence for their care, and both gave written informed consent before participating; the signed forms were collected and retained by the CDC team on their behalf. Recruitment through the CDC follow-up channel was chosen because it is the only route by which people living with HIV in this region can be approached without their status being disclosed to a third party. No information capable of identifying an evaluator appears in this article or in any associated data file.

App Selection and Data Collection

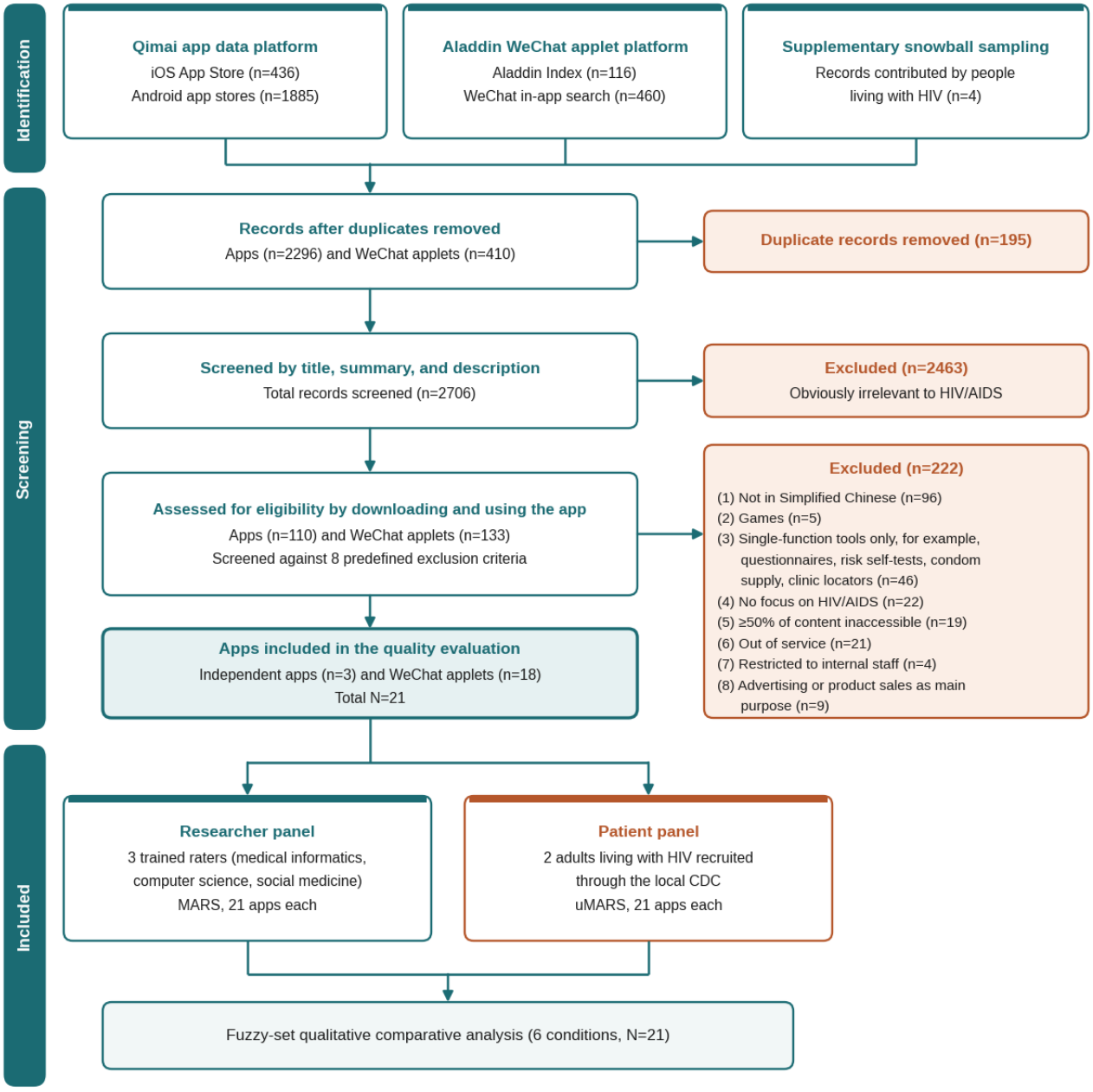
A systematic search was conducted on the Qimai app data platform, Aladdin WeChat applet data platform, and WeChat from February 18-19, 2023. Search terms included “HIV”, “AIDS”, “Human Immunodeficiency Virus”, “Acquired Immunodeficiency Syndrome”, and their Chinese equivalents. The Qimai Keyword Expansion Assistant was used to identify relevant apps with a relevance score exceeding 50% and a search index surpassing 4605.

WeChat applets, lightweight apps embedded within the WeChat platform, were included due to their widespread use in China. The Aladdin Index platform was used to identify relevant WeChat applets. Additionally, a snowball sampling method was implemented through collaboration with local CDC HIV prevention teams to identify apps commonly used within the HIV/AIDS community but potentially missed by platform searches.

The exclusion criteria for this study were the following: (1) non-Simplified Chinese language, (2) games, (3) apps with limited functionality, (4) apps not primarily focused on HIV/AIDS content, (5) more than 50% inaccessible content, (6) terminated services, (7) apps restricted to internal use, and (8) apps primarily intended for advertising or product sales. Following systematic screening, 21 HIV/AIDS apps were included for evaluation. The selection process and the subsequent 2-panel evaluation design are shown in Figure 1 [35].

XSL•FO
RenderX

Figure 1. Flow diagram of the identification, screening, and inclusion of Chinese HIV/AIDS mobile apps, and of the 2-panel evaluation design [35]. CDC: Center for Disease Control and Prevention; MARS: Mobile App Rating Scale; uMARS: user version of Mobile App Rating Scale.



Evaluation Tools: MARS and uMARS

Two versions of the MARS were used in this study to accommodate the different evaluator groups. The original MARS was used for researcher evaluations, while uMARS was used for patient evaluations [15,17].

The original MARS is a validated 23-item scale designed for expert assessment of app quality across four objective dimensions: (1) engagement (entertainment, interest, customization, interactivity, and target group); (2) functionality (performance, ease of use, navigation, and gestural design); (3) aesthetics (layout, graphics, and visual appeal); and (4) information (quality, quantity, credibility, and evidence-base). Each item is rated on a 5-point scale (1=inadequate to 5=excellent). MARS also includes a subjective quality subscale, which served as the outcome variable in this study [15].

The uMARS is a 20-item adaptation designed for end user evaluation, retaining the same 4 dimensions but with simplified language and reduced technical terminology to enhance accessibility for nonexpert raters [17]. Both MARS and uMARS have demonstrated good reliability and validity across diverse mHealth apps and languages [45-47]. The use of parallel instruments allowed for meaningful comparison between expert and user perspectives while ensuring each group could evaluate apps using an appropriate tool.

Evaluator Panels and Sample Size Considerations

Researcher Panel

Three researchers conducted the app evaluations using the original MARS and were purposively selected to cover the 3 competence domains that MARS scoring requires: health information content, software engineering, and app-quality

XSL•FO
RenderX

methodology. Rater 1 holds a master's degree in Medical Informatics with extensive experience in medical information analysis. Rater 2 holds a master's degree in Computer Science with software development expertise. Rater 3 holds a PhD in Social Medicine with substantial experience in mobile app evaluation.

Before formal evaluation, all 3 raters completed MARS training by reviewing instructional videos and conducting pilot evaluations on 4 randomly selected apps (2 WeChat applets and 2 independent apps). For the formal evaluation, raters 1 and 2 independently assessed all 21 apps. To ensure consistency, evaluators completed assessment of all apps for dimension N before proceeding to dimension N+1. When discrepancies occurred between raters 1 and 2, rater 3 reassessed the app, and a consensus score was reached through structured discussion.

Patient Panel

A total of 2 patient evaluators were recruited through collaboration with local CDC HIV prevention and control teams. Eligibility criteria, applied by the CDC team within its own follow-up records rather than by the research team, were the following: (1) confirmed HIV-positive diagnosis on the CDC follow-up register; (2) age ≥18 years; (3) ability to read and understand Simplified Chinese; and (4) experience using mobile apps. Both participants provided written informed consent, which the CDC team collected and retained on behalf of the study to protect their privacy, and both were assured of anonymity and data confidentiality; no serostatus information and no personal identifiers were transferred to the research team at any stage (see the Study Design and Ethical Considerations sections). Recruitment was purposive rather than probabilistic, with the aim of obtaining evaluators whose lived experience made them able to judge the apps as intended users.

Recruiting patient evaluators proved substantially more difficult and time-consuming than recruiting expert raters, and the size of the patient panel reflects that difficulty rather than a design preference. People living with HIV are a small and highly stigmatized minority in China, and they attach exceptional importance to the confidentiality of their serostatus. The dominant concern voiced by those who were approached and declined was not the time required by the rating protocol but the possibility that participation would create a record linking them by name to HIV-related research, however strong the assurances of anonymity. Recruitment could proceed only through the local CDC's routine follow-up visits, which is the sole channel through which people living with HIV in this region can be approached without their status being disclosed to a third party, and a considerable period elapsed before 2 individuals willing to take part were identified. No demographic detail about the 2 participants is reported, because none was collected: in a 2-person panel recruited through a single follow-up program, age, sex and city in combination would be potentially identifying, and the research team therefore asked the CDC team not to transfer such details.

Patient evaluators completed the uMARS rather than the original MARS. They were provided with brief uMARS training materials adapted for nontechnical users, emphasizing the evaluation of user experience rather than technical specifications. Patients independently evaluated all 21 apps over a 4-week period and submitted their ratings through a secure online platform. Adequate time was provided for patients to thoroughly explore each app before rating. Patient scores were averaged across the 2 evaluators for each app and dimension.

Sample Size Considerations

Two aspects of sample size warrant explicit justification: the number of evaluators and the number of apps.

Number of Evaluators

The use of 2 to 3 trained raters is not a constraint specific to this study but the established design convention of the MARS/uMARS app-quality literature, including in this journal. The MARS was developed and validated on independent ratings by a small expert panel and reported excellent interrater reliability under those conditions (intraclass correlation coefficient [ICC]=0.79) [15]. The convention has been maintained consistently since. Table 1 summarizes evaluator panel size across published MARS-based app-quality evaluations in *JMIR mHealth and uHealth* and closely related outlets, spanning instrument-validation studies, single-disease app reviews, and cross-national comparisons.

XSL•FO
RenderX

Table 1. Evaluator panel size in published MARS[a]-based app-quality evaluations. Studies are ordered by the number of apps evaluated.

Study	Health domain	Instrument	Apps evaluated (n)	Evaluators (n)
Miró and Llorens-Vernet [48], 2021	Chronic conditions (interrater reliability study)	MARS, MAG[b]	4	8
Fernández-Sánchez et al [36], 2025	HIV and STI[c] prevention	MARS	6	2 (plus a third for disputes)
Shen et al [49], 2021	Drug-drug interaction checking	MARS	7	2
Schmeelk et al [50], 2022	COVID-19 symptom monitoring	MARS	20	2 (5 for calibration)
This study	HIV/AIDS	MARS, uMARS[d]	21	3 researchers; 2 patients
Mani et al [51], 2015	Mindfulness	MARS	23	2 (second rater on 30% subset)
Amor-García et al [52], 2020	Genitourinary cancer	MARS	46	2
Yamamoto et al [53], 2022	Mental health (instrument validation)	MARS-J[e]	50	2
Messner et al [47], 2020	Mental health (instrument validation)	MARS-G[f]	104	2 per app

[a]MARS: Mobile App Rating Scale.

[b]MAG: Mobile App Development and Assessment Guide.

[c]STI: sexually transmitted infection.

[d]uMARS: user version of the Mobile App Rating Scale.

[e]MARS-J: Japanese version of the Mobile App Rating Scale.

[f]MARS-G: German version of the Mobile App Rating Scale.

Two features of this evidence are relevant here. First, panel size does not scale with the number of apps: 2 raters were used whether 6, 7, 20, 23, 46, 50, or 104 apps were evaluated [36,47,49-53]. This is because the rater in this paradigm functions as a measurement instrument applied to each app rather than as a sampled unit of analysis. The unit of analysis is the app; the psychometric requirement is demonstrable agreement between raters, not a large rater sample. Reliability is accordingly evaluated as interrater agreement, as reported in the Interrater Agreement section, rather than treated as sampling error. Second, 3 raters—the number used for our researcher panel—is at the upper rather than the lower end of the distribution [35,54]; most comparable studies use two. Beyond the studies tabulated, the same convention is evident across app domains in this journal, including drug-drug interaction checkers [55], medication adherence apps [56], apps for deaf and hard-of-hearing users [57], hematological condition apps [58,59], traditional vs modern medicine apps in China [60], noncommunicable disease management apps [61], and apps for men who have sex with men in China [39], as well as in the Korean [46] and Arabic [62] MARS validations, in reviews of depression apps in German app stores [34] and of commercial weight-management apps [63], and in the 2 published MARS-based evaluations of HIV/AIDS apps [37,38].

The patient panel requires a separate justification, because the logic that supports a small expert panel does not transfer directly to end users: experts are trained to converge on a normative standard, whereas end users are informative partly because they differ from one another. Three considerations support the present design. First, HIV-positive status is a stigmatized and legally protected attribute in China, and recruitment is correspondingly difficult: people living with HIV could be approached only through CDC follow-up visits, and the prevailing reason for declining was concern that participation would create documentation linking the individual to HIV-related research. Second, both patient evaluators rated all 21 apps on all dimensions, so the design is complete rather than sparse: there are no missing cells, and each app is evaluated by the same evaluators, which is the condition under which paired comparison with the researcher panel is valid. Third, agreement between the 2 patient raters was examined empirically rather than assumed, and the limits of agreement were narrow relative to the rating range (see Statistical Analysis). We therefore treat the patient results as characterizing the configurational structure observed in this panel and frame them as exploratory and hypothesis-generating throughout.

Terminology

Because a 2-person panel cannot stand for a population, a fixed vocabulary is used throughout. The patient panel and the 2 patient evaluators refer to the 2 individuals who completed the uMARS; patient configurations and patient-side pathways refer to configurations recovered from those 2 evaluators' averaged ratings of the 21 apps; and statements about the researcher panel are likewise restricted to the 3 raters described in the Researcher Panel section. No expression in this paper is intended to characterize the preferences of people living with HIV in China as a population, or of Chinese app users generally. Where the text refers to patient evaluations, it refers to this panel.

Number of Apps

Twenty-one apps are the complete set of Chinese HIV/AIDS apps meeting the inclusion criteria at the time of the search, identified from 2901 screened records. This is not a sample drawn from a larger population, but a census of an intentionally narrow population, and the small size is itself a substantive finding about the Chinese HIV/AIDS app market [35].

XSL•FO
RenderX

Comparable app-quality evaluations in this journal have analyzed as few as 4, 6, and 7 apps (Table 1), and evaluations of 20 apps have supported published quality conclusions [50].

For the configurational analysis specifically, qualitative comparative analysis (QCA) was developed for small- and intermediate-N designs and is conventionally applied to 10-50 cases, where intimate case knowledge is available, and each case can be interpreted individually [64,65]. With N=21, the present study lies within this range and, unlike most regression-based designs, permits every case to be inspected individually against its configuration. Six conditions with 21 cases nevertheless produce substantially limited diversity: of the 64 logically possible configurations, only 14 (researcher data) and 15 (patient data) contained at least one case with membership above 0.5. We therefore report the intermediate solution, which uses only easy counterfactuals [66,67]; interpret configurations alongside the specific apps that instantiate them; and report the sensitivity analyses described in the Robustness and Sensitivity Analyses section.

Privacy Protection Assessment

Privacy protection was coded by 2 researchers independently from the apps themselves and their store listings. An app was coded as protected (code 1) only if it met all three of the following criteria: (1) a privacy policy accessible from within the app; (2) explicit disclosure of what data are collected and how they are used; and (3) user control over personal data, such as account or data deletion. All other apps were coded as unprotected (code 0). Coding disagreement occurred for 2 of 21 apps and was resolved by discussion with reference to the app interface; agreement before discussion was 90.5% (19/21).

We acknowledge that this binary treatment simplifies a multidimensional construct. Comprehensive instruments for privacy assessment in mHealth evaluate policy fairness against data-protection regulation, permission and data-sharing behavior, and technical safeguards, and they yield graded rather than dichotomous scores [68-71]. Two considerations motivated the binary coding here. First, the empirical distribution was strongly bimodal: apps either provided all 3 elements or none of them, and no app in this sample provided partial protection in the sense of a policy without disclosure or without user control. Second, a graded index estimated from 21 cases would add analytic flexibility without adding usable variance. To test whether this choice drove the results, we repeated the entire analysis with privacy specified as a 3-level fuzzy set (0=no protection; 0.5=policy, disclosure, and user control; 1=the preceding elements plus an in-app privacy password), reported in the Alternative Privacy Operationalization section.

fsQCA

Conditions and Outcome

Six conditions were examined: engagement, functionality, aesthetics, and information quality (continuous, from MARS/uMARS); privacy protection (binary); and developer type (binary: 1=government organization, 0=nongovernment). The outcome was the subjective quality rating. Conditions were selected from the MARS/uMARS framework and from prior literature on determinants of mHealth app quality [11,15,33].

Calibration

Calibration transforms raw scores into fuzzy-set membership between 0 and 1 [66,72]. Continuous conditions were calibrated using the direct method with 3 anchors set at the 95th percentile (full membership), the 50th percentile (crossover), and the 5th percentile (full nonmembership). Percentile-based anchoring is standard where no external substantive benchmark exists for the measure, and is recommended when scores are instrument-specific and their absolute values are not interpretable against a criterion [73,74]. This applies here: a MARS engagement score of 3.0 has no external meaning as a threshold of “high engagement,” whereas its position within the observed distribution of HIV/AIDS apps does. Anchors were derived separately for the researcher and patient datasets, since the 2 instruments produce different rating distributions and a common anchor would confound instrument scaling with evaluator judgment. Membership scores falling exactly at 0.50 after calibration were adjusted upward by 0.001, so that no case was left at the point of maximum ambiguity; this mattered most for patient functionality, since 14 of the 21 apps received the same functionality rating and that rating is the crossover value. Where a percentile anchor coincided with the crossover, which occurred only for patient functionality under the wider anchor sets used in the sensitivity analysis, the affected anchor was set to the nearest observed value beyond the crossover. Binary conditions were retained as crisp sets. Because the choice of percentile anchors is a researcher decision rather than a property of the data, its consequences were examined explicitly in the sensitivity analysis.

Necessity

A condition is necessary if the outcome cannot occur without it [66,75]. Necessity was assessed by consistency, with the conventional threshold of 0.90 [67], and evaluated jointly with coverage to guard against trivial necessity. Both the presence and the absence of each condition were tested.

Sufficiency

Truth tables were constructed for both datasets. A frequency cutoff of 1 was applied, as is standard for small-N designs [64], together with a raw consistency cutoff of 0.80 and a proportional reduction in inconsistency (PRI) consistency cutoff of 0.70. Boolean minimization produced complex, parsimonious, and intermediate solutions; the intermediate solution is reported as the primary result [67]. Following Fiss [76], conditions appearing in both the parsimonious and the intermediate solution are designated core, and those appearing only in the intermediate solution are designated peripheral.

Robustness and Sensitivity Analyses

Four sets of checks were performed, following current QCA reporting recommendations [64,74,77]. First, all analyses were repeated under 4 calibration anchor sets (95/50/5, 90/50/10, 85/50/15, and 80/50/20 percentiles). Second, the truth table analysis was repeated across raw consistency cutoffs of 0.75, 0.80, 0.85, and 0.90 crossed with PRI cutoffs of 0.65, 0.70, 0.75, and 0.80. Third, the analysis was repeated with the 3-level privacy set described in the Privacy Protection Assessment section. Fourth, necessity consistencies were recomputed under

XSL•FO
RenderX

each anchor set. The results are considered robust where the set of configurations and their substantive interpretation are unchanged, and consistency remains above the 0.80 threshold; any specification under which this does not hold is reported explicitly.

Statistical Analysis

Overview

Descriptive and comparison statistics were computed in Python 3.12; calibration, necessity analysis, and truth-table analysis were conducted in fsQCA (version 4.1; University of California). Descriptive statistics were computed for all variables. Because the 2 evaluator groups rated the same 21 apps, differences in raw dimension scores were tested with Wilcoxon signed-rank tests, with the matched rank-biserial correlation as the effect size; the nonparametric test was chosen because dimension scores were bounded, skewed, and drawn from a small sample. Agreement between groups was quantified with Pearson and Spearman correlations. Given 5 prespecified comparisons, *P* values should be interpreted descriptively rather than as confirmatory tests.

Interrater Agreement

Agreement was examined within each panel separately, and because the 2 panels completed different instruments, no estimate reported in this section combines researcher and patient ratings. Within the researcher panel the 2 independent raters were raters 1 and 2, each of whom scored all 21 apps on the original MARS before any discrepancy was referred to rater 3 and before consensus scores were formed; Bland-Altman analysis of that MARS dataset [35,78] gave a mean difference of 0.01 (SD 0.10; 95% limits of agreement −0.19 to 0.21) for app quality and 0.07 (SD 0.49; 95% limits of agreement −0.89 to 1.03) for subjective quality. Within the patient panel, the 2 independent raters were the 2 patient evaluators whose ratings are analyzed in the present paper, each of whom scored all 21 apps on the uMARS before the 2 sets of scores were averaged; Bland-Altman analysis of that uMARS dataset gave a mean difference of 0.13 (SD 0.26; 95% limits of agreement −0.38 to 0.64) for app quality and 0.23 (SD 0.45; 95% limits of agreement −0.66 to 1.12) for subjective quality. Both uMARS estimates therefore derive specifically from the 2 patient evaluators included here, and no researcher rating enters them; conversely, the 2 MARS estimates derive only from raters 1 and 2, and no patient rating enters them.

Limits of agreement for the 2 app-quality scores are narrow relative to the 1-5 rating range, indicating close agreement within both panels; the wider limits for the single-item subjective quality scores are consistent with the greater variability expected of subjective judgments. Bland-Altman analysis was the prespecified agreement method in the parent evaluation, and it is the appropriate choice for 2 raters because it quantifies both systematic bias and the range of disagreement on the original measurement scale rather than expressing agreement relative to between-app variance [78]. An ICC could not be computed retrospectively for either panel, because the archived dataset retains only the postconsensus researcher scores and the averaged patient scores; the rater-level score files are no longer available, and the Bland-Altman estimates reported in this section are those computed in the parent evaluation while those files were still held [35]. Future apps of this design should archive rater-level scores so that both agreement statistics can be reported.

Ethical Considerations

This study evaluated publicly available mobile apps. The data analyzed were MARS and uMARS ratings of those apps; no biological sample, clinical record, health record, or personal information of any evaluator was collected, held, or analyzed by the research team. The serostatus of the 2 patient evaluators was neither collected nor verified by the research team: eligibility was applied by the collaborating CDC HIV prevention and control team within its own follow-up records, and no identifier or personal detail was transferred to the research team. Prospective ethics review was not sought, on the basis set out in the Methods: the study lies outside the scope defined in Article 3 of the Measures for Ethical Review of Life Science and Medical Research Involving Humans (Guo Wei Ke Jiao Fa [2023] No. 4) [42] and would in the alternative fall within the exemption in Article 32(2) of the same instrument. This position is consistent with comparable app-quality evaluations, including those in which lay end users completed the instrument [43,44]. The study was nonetheless conducted in accordance with the principles of the Declaration of Helsinki: both patient evaluators were informed of the purpose and procedures of the study and of their right to withdraw at any time without consequence, and both provided written informed consent, the signed forms having been collected and retained by the CDC team on behalf of the study to protect the participants' privacy.

Results

Descriptive Statistics and Group Comparison

Twenty-one HIV/AIDS apps were evaluated by both researcher and patient groups. Table 2 presents the descriptive statistics for all variables. Researcher ratings demonstrated moderate variability across conditions, with functionality receiving the highest mean score (4.28, SD 0.39) and engagement the lowest (mean 2.40, SD 0.53). Subjective quality ratings from researchers averaged 2.47 (SD 0.65), ranging from 1.13 to 4.13.

XSL•FO
RenderX

Table 2. Descriptive statistics for researcher (Mobile App Rating Scale) and patient (user version of the Mobile App Rating Scale) ratings of 21 HIV/AIDS apps [35].

Variable	R[a] mean (SD)	R min[b]	R max[c]	P[d] mean (SD)	P min	P max
Engagement	2.40 (0.53)	1.60	3.60	2.56 (0.63)	1.80	4.70
Functionality	4.28 (0.39)	3.25	4.75	3.84 (0.59)	1.50	4.50
Aesthetics	3.40 (0.49)	2.33	4.17	3.22 (0.48)	1.67	4.17
Information	3.83 (0.33)	3.33	4.33	3.90 (0.74)	1.75	4.75
Subjective quality	2.47 (0.65)	1.13	4.13	2.75 (0.61)	1.00	3.75

[a]R: researcher.

[b]min: minimum.

[c]max: maximum.

[d]P: patient.

Patient evaluations showed similar patterns but with notable differences. Patients rated engagement slightly higher than researchers (mean 2.56, SD 0.63), though still the lowest among all dimensions. Information quality received the highest patient rating (mean 3.90, SD 0.74), followed by functionality (mean 3.84, SD 0.59). Patient subjective quality ratings averaged 2.75 (SD 0.61), ranging from 1.00 to 3.75, indicating somewhat higher overall satisfaction compared to researchers.

Wilcoxon signed-rank tests on the paired ratings (Table 3) showed that the 2 groups differed significantly on functionality, which researchers rated substantially higher than patients (median difference 0.50, IQR 0.00 to 0.63; W=14.0, P=.003, rank-biserial r=0.82), and on engagement, which patients rated slightly higher (median difference −0.20, IQR −0.40 to 0.10; W=57.5, P=.04, r=−0.50). Differences in information quality (P=.07), subjective quality (P=.06), and aesthetics (P=.16) did not reach significance. Rank correlations between the groups were moderate for engagement (ρ=0.64, P=.002) and information (ρ=0.63, P=.002) but near 0 for functionality (ρ=0.20, P=.38). Functionality is thus the dimension on which the groups both differ most in level and agree least in ordering, whereas engagement is rated similarly and ranked consistently. This pattern provides the empirical warrant for analyzing the 2 groups separately rather than pooling them. Dimension-level differences between the 2 groups are displayed in Figure 2.

Table 3. Comparison of researcher and patient ratings across the same 21 apps (Wilcoxon signed-rank tests and rank correlations).

Dimension	R[a] mean (SD)	P[b] mean (SD)	Median difference (IQR)	W[c]	P value	Rank-biserial r	Pearson r	Spearman ρ
Engagement	2.40 (0.53)	2.56 (0.63)	−0.20 (−0.40 to 0.10)	57.5	.04	−0.50	0.71	0.64
Functionality	4.28 (0.39)	3.84 (0.59)	0.50 (0.00 to 0.63)	14.0	.003	0.82	0.06	0.20
Aesthetics	3.40 (0.49)	3.22 (0.48)	0.17 (−0.17 to 0.51)	67.5	.16	0.36	0.23	0.32
Information	3.83 (0.33)	3.90 (0.74)	−0.21 (−0.42 to 0.01)	56.0	.07	−0.47	0.48	0.63
Subjective quality	2.47 (0.65)	2.75 (0.61)	−0.38 (−0.75 to 0.25)	54.0	.06	−0.49	0.39	0.45

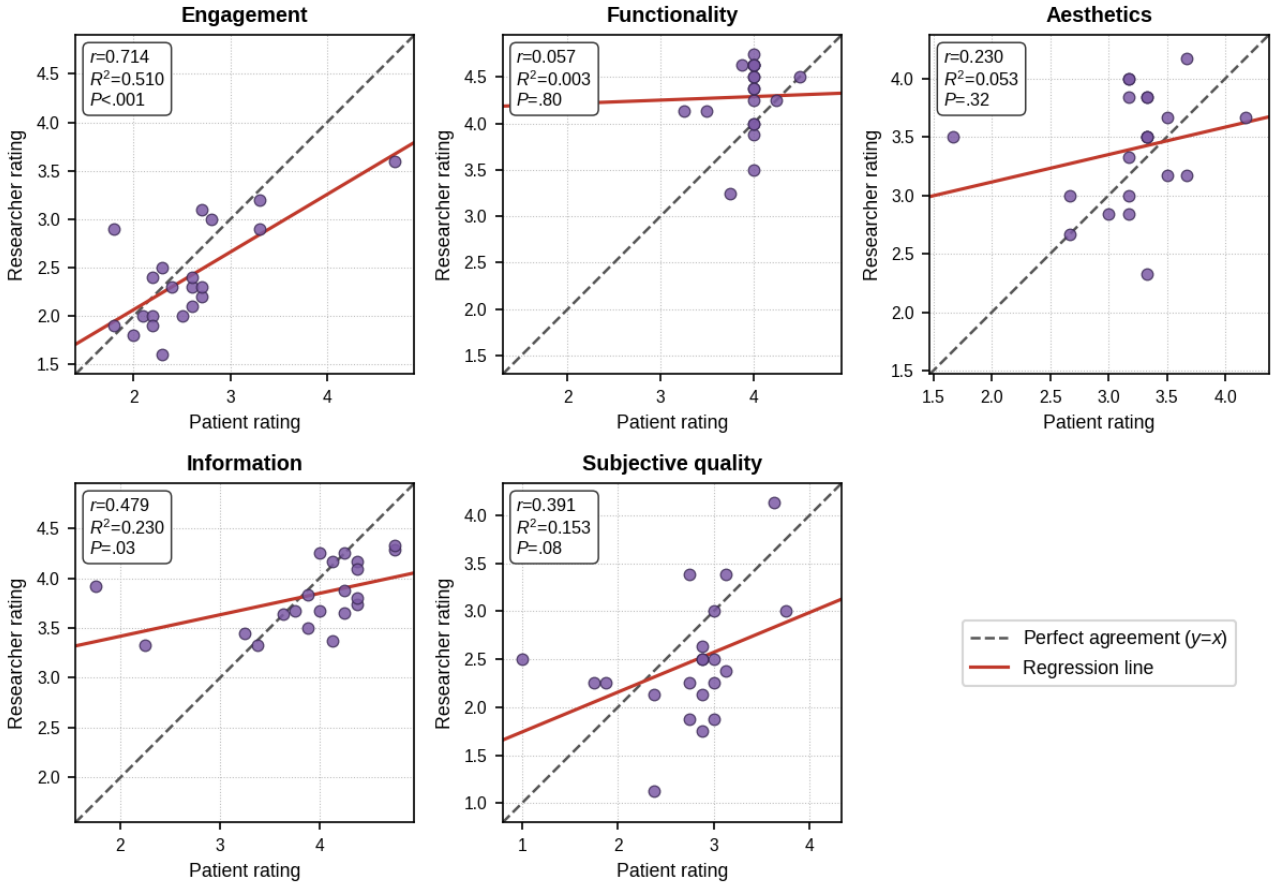
[a]R: researcher.

[b]P: patient.

[c]W: Wilcoxon signed-rank test statistic.

XSL•FO
RenderX

Figure 2. Dimension-level comparison of researcher and patient ratings.

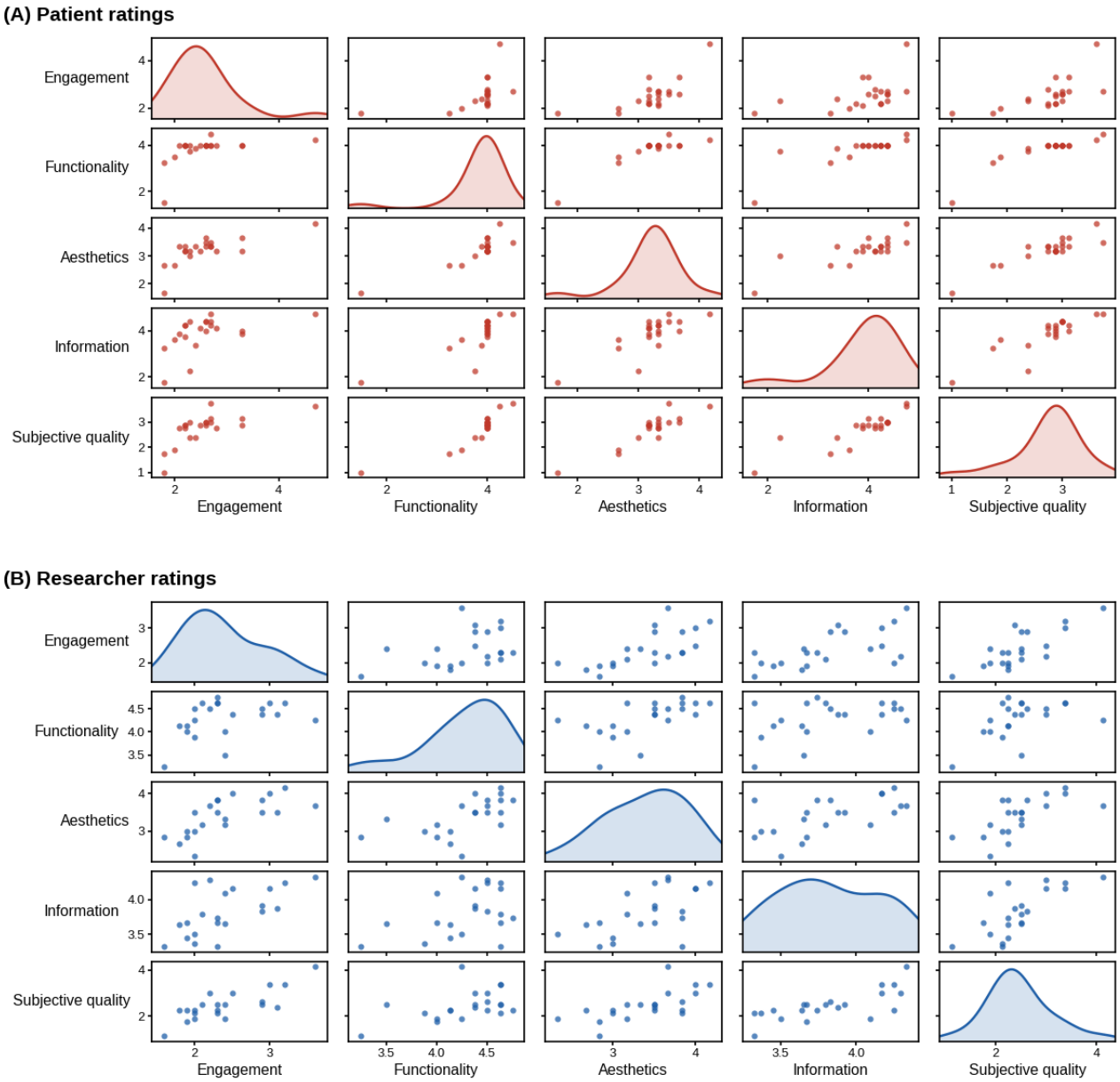


Association Structure Within Each Group

Scatter matrix analysis (Figure 3) revealed divergent evaluation frameworks. Patient ratings showed tight linear patterns, most notably for functionality-subjective quality (r=0.890; P<.001). In contrast, researcher ratings exhibited greater scatter for this relationship (r=0.489; P=.02, n=21; Δr=0.401), indicating more compartmentalized assessment. Patient judgments were therefore more strongly organized around a single evaluative axis, while researcher judgments were more compartmentalized across dimensions. The bivariate relationships that differ most between the 2 groups are shown side by side in Figure 4.

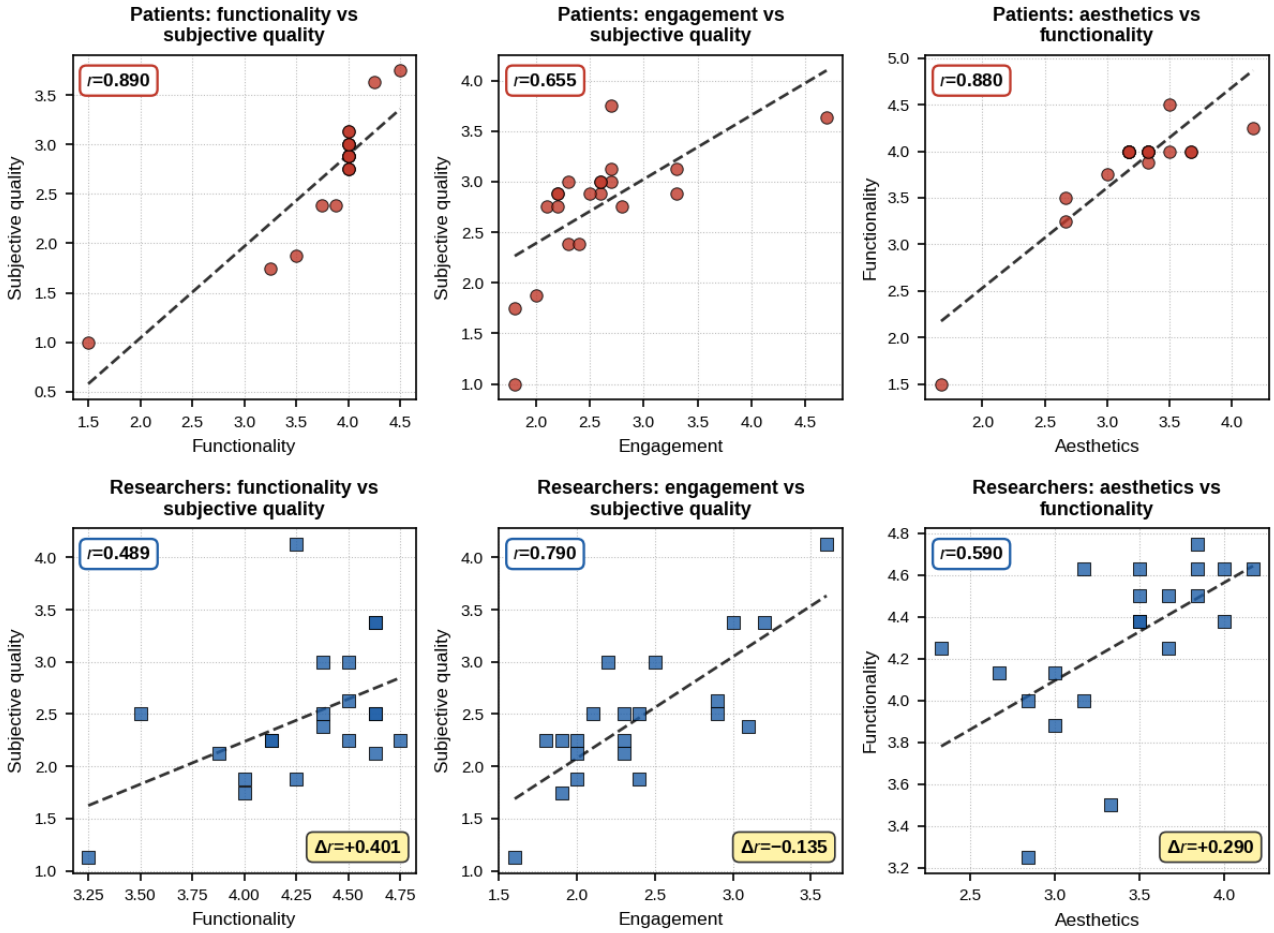
XSL•FO
RenderX

Figure 3. Scatter matrices for patient (A) and researcher (B) ratings.



XSL•FO
RenderX

Figure 4. Key bivariate relationships compared between groups.



Calibration

Calibration anchors differed between researcher and patient datasets, reflecting distinct rating distributions (Table 4). For researchers, the full membership threshold for engagement was 3.20, compared to 3.30 for patients. Functionality calibration showed more pronounced differences: researchers' full membership threshold was 4.63 vs 4.25 for patients. Information quality thresholds demonstrated the greatest divergence: researchers set full membership at 4.29, whereas patients used 4.75. For the outcome variable (subjective quality), full membership thresholds were 3.38 for researchers and 3.63 for patients.

Table 4. Calibration anchors.

Variable	Type	R[a] full[b]	R crossover	R nonmem.[c]	P[d] full	P crossover	P nonmem.
Engagement	Continuous	3.20	2.30	1.80	3.30	2.50	1.80
Functionality	Continuous	4.63	4.38	3.50	4.25	4.00	3.25
Aesthetics	Continuous	4.00	3.50	2.67	3.67	3.33	2.67
Information	Continuous	4.29	3.80	3.33	4.75	4.13	2.25
Privacy	Binary	1	N/A[e]	0	1	N/A	0
Developer type	Binary	1	N/A	0	1	N/A	0
Subjective quality	Continuous	3.38	2.38	1.75	3.63	2.88	1.75

[a]R: researcher.

[b]full: full membership.

[c]nonmem.: full nonmembership.

[d]P: patient.

[e]N/A: not applicable.

These calibration differences underscore the importance of group-specific analysis in fsQCA, as identical raw scores may represent different levels of membership in the fuzzy set of "high quality" depending on the evaluator group's reference

XSL•FO
RenderX

frame. They should not be read as evidence that one group applied a stricter standard than the other in any absolute sense, since MARS and uMARS items are not scaled to a common metric.

Necessary Conditions

Necessity analysis results are presented in Table 5 and Figure 5. For researcher evaluations, no single condition met the conventional threshold (consistency ≥0.90) for being considered necessary. Information quality demonstrated the highest necessity consistency (0.845), followed by functionality (0.821) and aesthetics (0.820). Engagement showed lower necessity (0.795), while privacy protection (0.322) and government developer type (0.242) were far below the threshold. Their negations showed moderate consistency, indicating that nongovernmental apps without formal privacy features could still receive high researcher ratings.

Table 5. Necessity analysis for high subjective quality.

Condition	R[a] cons.[b]	R cov.[c]	R ~[d]cons.	R ~cov.	P[e] cons.	P cov.	P ~cons.	P ~cov.
Engagement	0.795	0.807	0.490	0.460	0.829	0.844	0.594	0.547
Functionality	0.821	0.719	0.456	0.501	0.907	0.940	0.721	0.653
Aesthetics	0.820	0.815	0.506	0.484	0.841	0.870	0.627	0.568
Information	0.845	0.804	0.482	0.482	0.915	0.862	0.594	0.588
Privacy	0.322	0.660	0.678	0.434	0.269	0.546	0.731	0.463
Developer type	0.242	0.354	0.758	0.554	0.283	0.410	0.717	0.520

[a]R: researcher.

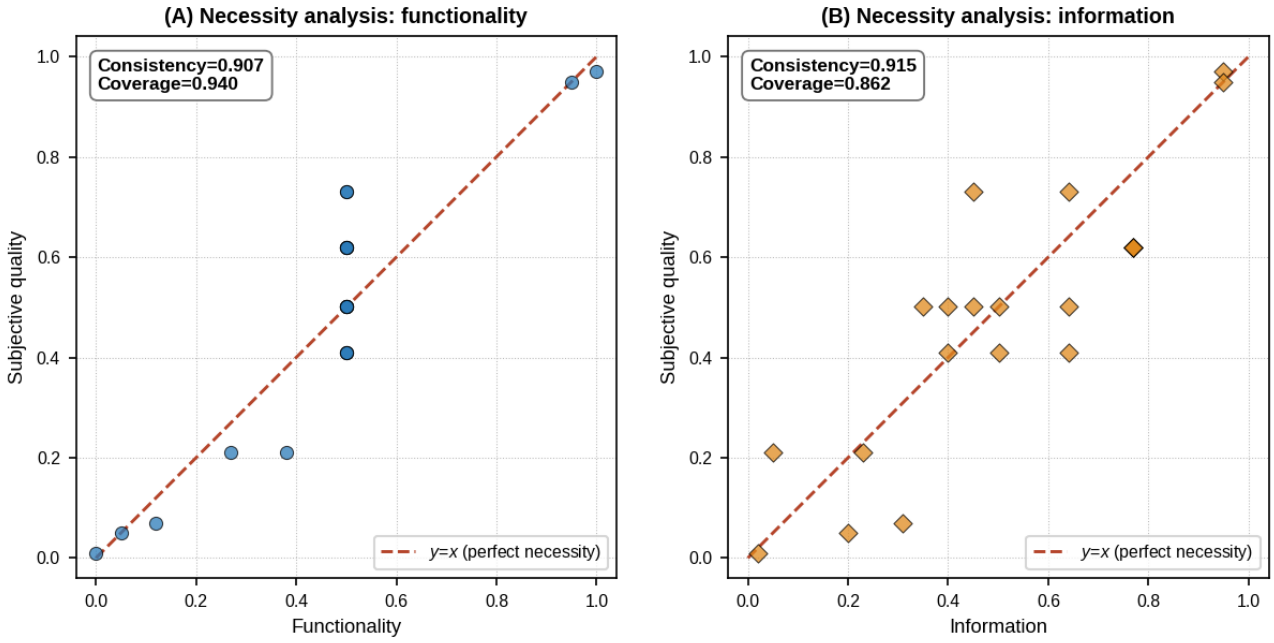
[b]cons.: consistency.

[c]cov.: coverage.

[d]~: negation of the condition.

[e]P: patient.

Figure 5. Necessity XY plots.



For patient evaluations, functionality (0.907) and information quality (0.915) both exceeded the threshold. However, these 2 conditions were highly correlated in the patient data (r=0.816), which means they cannot be treated as independent necessary conditions; a high correlation between candidate necessary conditions is a known source of overidentification in necessity analysis [67]. Following Schneider and Wagemann [67], we resolved this on substantive grounds: functionality is the precondition for reaching content at all, since a user must be able to navigate an app before its information can be judged, and information quality is best read as an extension of that access rather than a separate requirement. Functionality is therefore reported as the single candidate necessary condition for patients. We emphasize that with 21 cases and 1 panel, this should be treated as a proposition for testing rather than as an established necessity claim. Aesthetics (0.841) and engagement (0.829) approached but did not meet the threshold; privacy (0.269) and government developer type (0.283) did not.

XSL•FO
RenderX

Sufficient Configurations: Researchers

Sufficiency analysis was conducted to identify configurational pathways leading to high subjective quality ratings. Five sufficient configurations were identified for high subjective quality ratings among researchers (Table 6). The overall solution consistency was 0.981, and solution coverage was 0.649. These values exceed the recommended threshold of 0.80 for consistency and indicate acceptable empirical coverage.

Table 6. Sufficient configurations for high subjective quality—researchers.

Configuration	Causal path	Consistency	Raw coverage	Unique coverage
R1	~[a]Developer × privacy × engagement × aesthetics × information	1.000	0.159	0.033
R2	~Developer × privacy × functionality × aesthetics × information	1.000	0.163	0.037
R3	~Developer × engagement × functionality × aesthetics × information	0.985	0.463	0.293
R4	Privacy × engagement × functionality × aesthetics × information	1.000	0.175	0.049
R5	~Developer × ~privacy × engagement × ~functionality × ~aesthetics × ~information	0.942	0.110	0.066
Solution	N/A[b]	0.981	0.649	N/A

[a]~: negation of the condition.

[b]N/A: not applicable.

Configuration R3 (nongovernment developer × high engagement × high functionality × high aesthetics × high information quality) demonstrated the highest raw coverage (0.463) and unique coverage (0.293). This pathway accounted for the largest proportion of high-quality outcomes. The consistency of 0.985 indicates that this configuration almost always leads to high subjective ratings.

Configurations R1 and R2 both combined privacy protection with aesthetics and information quality, each at consistency 1.000, differing only in whether engagement or functionality completed the combination. Within this dataset, the two therefore behave as substitutable elements once privacy is present, though with a single case supporting each pathway, this substitutability is a description of these cases rather than a generalizable finding.

Configuration R4 combined all positive conditions (privacy, engagement, functionality, aesthetics, and information) without specifying developer type. This comprehensive pathway achieved perfect consistency (1.000) with moderate coverage (0.175).

Configuration R5 is a contrarian pathway: nongovernmental apps without privacy features, with engagement but low functionality, aesthetics, and information quality (consistency 0.942, raw coverage 0.110). Only one app instantiates it, and it is reported for completeness rather than as a recommended design route.

Sufficient Configurations: Patients

Four sufficient configurations emerged from patient evaluations (Table 7). The overall solution consistency was 0.998, and solution coverage was 0.796. Both metrics exceeded researcher evaluation results, indicating stronger explanatory power.

Table 7. Sufficient configurations for high subjective quality—patients.

Configuration	Causal path	Consistency	Raw coverage	Unique coverage
P1	~[a]Developer × ~privacy × engagement × functionality × aesthetics	1.000	0.285	0.018
P2	~Privacy × engagement × functionality × aesthetics × information	0.996	0.474	0.207
P3	~Developer × engagement × functionality × aesthetics × information	1.000	0.482	0.215
P4	~Developer × ~privacy × ~engagement × functionality × ~aesthetics × information	1.000	0.263	0.089
Solution	N/A[b]	0.998	0.796	N/A

[a]~: negation of the condition.

[b]N/A: not applicable.

Configuration P3 (nongovernment developer × high engagement × high functionality × high aesthetics × high information quality) demonstrated the highest coverage (raw=0.482, unique=0.215). This pathway closely mirrors researcher configuration R3,

XSL•FO
RenderX

suggesting one shared route to a favorable judgment across groups.

Configuration P2 (absence of privacy × high engagement × high functionality × high aesthetics × high information quality) achieved near-perfect consistency (0.996) with substantial coverage (0.474). The appearance of privacy absence as a component distinguishes patient from researcher solutions, but it should be read against the distribution of the condition: 16 of 21 apps lacked privacy protection, so most apps that were strong on the 4 MARS dimensions were also, necessarily, apps without it. It does not indicate that the patient panel rated protected apps lower. In set-theoretic terms, protected apps were slightly more consistently members of the high-rating set than unprotected ones (0.546 vs 0.463; Table 5), and the 2 apps the patient panel rated highest, Chabei and Hong Feng Wan, both offered protection.

Configuration P1 emphasized engagement, functionality, and aesthetics without requiring high information quality. This pathway indicates that visually appealing and functionally robust apps can achieve high patient ratings even with moderate information content.

Configuration P4 is compensatory: apps lacking engagement and aesthetics still achieved high ratings through functionality and information quality (consistency 1.000, raw coverage 0.263).

Comparison Across Groups

Table 8 presents a consolidated comparison of configurations across evaluator groups. Several patterns warrant attention.

Table 8. Configurational comparison—researchers and patients.

Antecedent condition	R1	R2	R3	R4	R5	P1	P2	P3	P4
Engagement	✓[a]	—[b]	✓	✓	✓	✓	✓	✓	X[c]
Functionality	—	(✓)[d]	(✓)	(✓)	X	(✓)	(✓)	(✓)	✓
Aesthetics	✓	✓	✓	✓	(X)[e]	✓	✓	✓	X
Information	✓	✓	✓	✓	(X)	—	✓	✓	✓
Privacy	✓	✓	—	✓	(X)	(X)	(X)	—	(X)
Developer type	X	X	X	—	X	X	—	X	X
Consistency	1.000	1.000	0.985	1.000	0.942	1.000	0.996	1.000	1.000
Raw coverage	0.159	0.163	0.463	0.175	0.110	0.285	0.474	0.482	0.263
Unique coverage	0.033	0.037	0.293	0.049	0.066	0.018	0.207	0.215	0.089
Solution consistency	0.981	N/A[f]	N/A	N/A	N/A	0.998	N/A	N/A	N/A
Solution coverage	0.649	N/A	N/A	N/A	N/A	0.796	N/A	N/A	N/A

[a]✓: core condition present.

[b]Em dashes indicate that the condition may be either present or absent.

[c]X: core condition absent.

[d](✓): peripheral condition present.

[e](X): peripheral condition absent.

[f]N/A: not applicable.

First, functionality appeared as a core or peripheral condition in most configurations for both groups. For patients, functionality was present in all 4 configurations. For researchers, functionality appeared in 3 of 5 configurations. This finding aligns with necessity analysis results showing functionality as highly important for both groups.

Second, privacy protection plays different roles. It is a core condition in 3 researcher configurations, whereas its negation appears in 3 of 4 patient configurations. Since privacy protection was present in only 5 of 21 apps, both patterns rest on few cases. The contrast is a difference in configurational role, not a difference in valuation: the negated condition enters the patient solution because unprotected apps dominate the sample, and protected apps were, if anything, rated slightly more consistently highly by the patient panel (Table 5). The skewed distribution of this condition is a limitation of the sample rather than a finding about the evaluators.

Third, engagement carried different weight in the 2 panels. Three of 4 patient configurations included engagement as a core condition, and in 2 of those it is one of only 3 core conditions. For researchers, engagement was a core condition in 4 of 5 configurations, but in 3 of those it appeared alongside 3 or 4 further core conditions.

Fourth, nongovernmental development appears across both groups' high-performing configurations. Because government-developed apps in this sample were predominantly county-level CDC applets with limited interactive design, this condition plausibly indexes development resources and design intent rather than ownership as such.

Fifth, solution coverage differed substantially between groups (researchers: 0.649; patients: 0.796). The higher patient coverage indicates that the identified configurations account for a larger proportion of the outcomes rated highly by the 2 patient evaluators.

XSL•FO
RenderX

Robustness and Sensitivity Analyses

Calibration Anchors

Recomputing every configuration under 4 anchor sets left the configurational structure unchanged (Table 9). Across all anchor sets, every configuration retained consistency at or above 0.942, and solution-level values remained high (researchers: consistency 0.966-0.983, coverage 0.639-0.679; patients: consistency 0.997-1.000, coverage 0.582-0.796). The baseline column of Table 9 is the set of values reported in Tables 6 and 7. Coverage was more anchor-sensitive than consistency, and most so for the patient configurations, which is expected because relaxing the full-membership anchor rescales membership scores in the outcome.

Table 9. Sensitivity of configuration consistency and raw coverage to calibration anchors[a].

Config.[b]	Cons.[c] 95/50/5	Cons. 90/50/10	Cons. 85/50/15	Cons. 80/50/20	Cov.[d] 95/50/5	Cov. 90/50/10	Cov. 85/50/15	Cov. 80/50/20
R1	1.000	1.000	1.000	1.000	0.159	0.159	0.159	0.167
R2	1.000	1.000	1.000	1.000	0.163	0.159	0.152	0.166
R3	0.985	0.982	0.987	0.959	0.463	0.454	0.422	0.442
R4	1.000	1.000	1.000	1.000	0.175	0.167	0.154	0.162
R5	0.942	0.948	0.944	0.944	0.110	0.103	0.084	0.082
P1	1.000	1.000	1.000	1.000	0.285	0.233	0.182	0.111
P2	0.996	1.000	1.000	1.000	0.474	0.359	0.316	0.219
P3	1.000	1.000	1.000	1.000	0.482	0.397	0.384	0.306
P4	1.000	1.000	1.000	1.000	0.263	0.219	0.181	0.153

[a]Values were recomputed from the calibrated data using identical procedures; the 95/50/5 column reproduces Tables 6 and 7.

[b]Config.: configuration.

[c]Cons.: consistency.

[d]Cov.: coverage.

Necessity Across Anchors

Necessity consistencies were recomputed under each anchor set. For researchers, no condition reached 0.90 under any specification, so the conclusion that no single condition is necessary for researcher judgments is not an artifact of the anchors. For patients, functionality (0.907) and information quality (0.915) exceeded the threshold at the baseline anchors only; under the three wider sets neither reached it (functionality 0.744-0.804; information quality 0.824-0.860). Two features of the patient data account for this. First, patient functionality ratings are concentrated at 4.00, so that the 80th, 85th, and 90th percentiles coincide with the crossover, and the full-membership anchor had to be set to the nearest observed value beyond it. Second, widening the anchors compresses membership scores toward the crossover and thereby lowers necessity consistency for conditions whose upper values are closely spaced. The patient necessity result is therefore anchor-dependent, which reinforces rather than qualifies the caution stated in the Necessary Conditions section: functionality is offered as a proposition for testing and not as an established necessary condition. Privacy protection remained far below threshold in every specification for both groups (0.264-0.343). The full necessity sensitivity results are given in Table 10.

XSL•FO
RenderX

Table 10. Necessity consistency for high subjective quality under 4 calibration anchor sets[a].

Group	Condition	95/50/5	90/50/10	85/50/15	80/50/20
Researchers	Engagement	0.795	0.787	0.767	0.783
Researchers	Functionality	0.821	0.780	0.738	0.756
Researchers	Aesthetics	0.820	0.809	0.806	0.811
Researchers	Information	0.845	0.850	0.804	0.839
Researchers	Privacy	0.322	0.329	0.326	0.343
Researchers	Developer type	0.242	0.236	0.238	0.226
Patients	Engagement	0.829	0.730	0.788	0.765
Patients	Functionality	0.907	0.804	0.795	0.744
Patients	Aesthetics	0.841	0.805	0.738	0.645
Patients	Information	0.915	0.849	0.860	0.824
Patients	Privacy	0.269	0.266	0.277	0.264
Patients	Developer type	0.283	0.284	0.279	0.288

[a]Values are necessity consistencies. The conventional threshold is 0.90. Column headings give the percentile anchors used for full membership, crossover, and full nonmembership.

Consistency and PRI Thresholds

Fourteen of the 64 truth table rows were empirically instantiated in the researcher data and 15 in the patient data. Under the baseline cutoffs, 6 researcher rows and 5 patient rows met both criteria. Across the 16 combinations of raw consistency cutoffs (0.75, 0.80, 0.85, 0.90) and PRI cutoffs (0.65, 0.70, 0.75, 0.80), the retained set was invariant in the patient data. In the researcher data, it was invariant with respect to the raw consistency cutoff, but the row instantiated by the Beijing AIDS Association applet, whose PRI consistency is 0.712, dropped out at PRI cutoffs of 0.75 and above, which removes configuration R5 from the solution. R5 is the contrarian pathway, is instantiated by that single case, and has the lowest consistency and the lowest raw coverage of the configurations reported; the other 4 researcher configurations were unaffected at every combination tested. The solutions are therefore insensitive to the raw consistency cutoff, and sensitive to the PRI cutoff only in respect of R5.

Alternative Privacy Operationalization

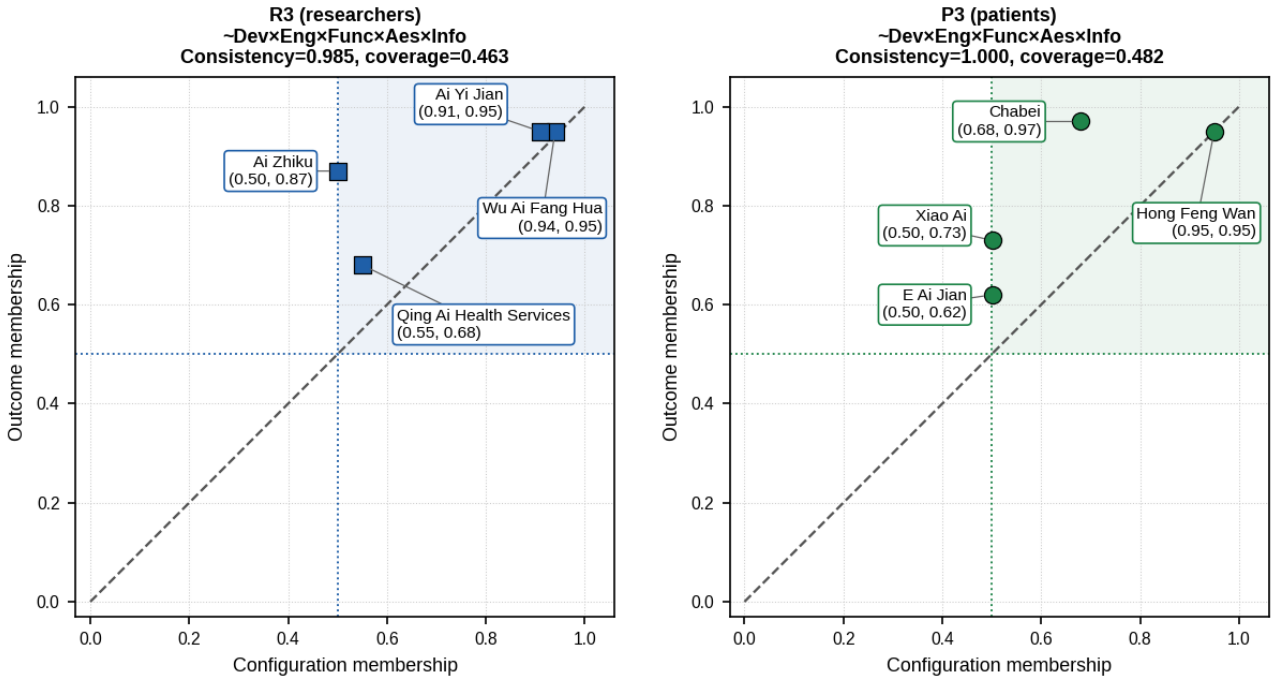
Re-running the analysis with the 3-level privacy set left every configuration unchanged in substance: the truth tables retained the same rows, Boolean minimization returned the same 5 researcher and 4 patient configurations, and the parsimonious solutions were identical. Consistency shifted by at most 0.018 (patient P4: 1.000 to 0.982), and no configuration fell below 0.947. Coverage changes were likewise small, the largest being increases of 0.052 in the raw coverage of patient P1 (0.285 to 0.337) and of 0.051 in that of P2 (0.474 to 0.525). Necessity consistency for privacy protection remained low under the graded specification (researchers 0.278; patients 0.223). The limited role of privacy in the patient configurations is therefore not an artifact of dichotomization: only 5 apps offered protection of any kind, only 2 of those added an in-app privacy password, and grading the 5 does not alter how privacy combines with the other conditions.

Case-Level Analysis

XY plots were constructed to visualize the cases instantiating the highest-coverage configurations, R3 and P3 (Figure 6). All illustrative cases demonstrated configuration membership scores above 0.5 with corresponding outcome membership scores above 0.5, confirming sufficiency at the case level. No deviant cases were identified.

XSL•FO
RenderX

Figure 6. XY plots for case-level analysis of configurations R3 (researchers) and P3 (patients). Both configurations share the same structure (~Dev×Eng×Func×Aes×Info) and are the highest-coverage pathway for each evaluator group. Each point is an app; the values in parentheses are configuration membership and outcome membership, and values shown as 0.50 are 0.501 after the adjustment described in the Calibration section. Aes: aesthetics; Dev: government developer; Eng: engagement; Func: functionality; Info: information quality; ~: negation of the condition.



For researchers, configuration R3 contained the most cases (n=4), including Wu Ai Fang Hua (0.94, 0.95) and Ai Yi Jian (0.91, 0.95). For patients, configuration P3 was best exemplified by Chabei (0.68, 0.97).

R3 and P3 share an identical structure (~developer [~ denotes negation of the condition] × engagement × functionality × aesthetics × information), yet different apps sit at the top of each. This indicates that the same combination of dimensions is satisfied by different apps depending on who is rating. These apps are presented as illustrations of each pathway, not as validated exemplars: with 1 to 4 cases per configuration, they show what a configuration looks like in practice and support interpretation, but they do not constitute evidence that these are the best available apps.

Discussion

Principal Findings

Configurational pathways to high subjective quality in HIV/AIDS apps were examined using fsQCA. Three key findings emerged. First, equifinality was demonstrated: 5 configurations were identified for researchers, and 4 for patients, and no single pathway was required for high-quality ratings. Second, the 2 panels' evaluation priorities differed in this sample. Privacy protection was a core condition in 3 researcher configurations and appeared as a present condition in none of the patient ones, whose configurations instead combined engagement, functionality, aesthetics, and information quality with the absence of privacy protection. Third, functionality emerged as foundational: it reached the necessity threshold in the patient data and appeared in most configurations for researchers. Wilcoxon tests independently confirmed that functionality is where the 2 groups diverge most in level and agree least in rank order.

These findings are broadly consistent with the theoretical framework outlined earlier, with one important qualification. The prominence of functionality and engagement in patient configurations matches the task-technology fit and performance-expectancy predictions [20-22]: for a person managing HIV, an app earns a favorable judgment by letting them accomplish something. The privacy calculus prediction that end users discount abstract, delayed risks [23,24] is not, however, supported by these data. Privacy protection was absent from the patient pathways because it was absent from most apps, not because protected apps were rated lower; the 2 apps the patient panel rated highest both offered protection. What the data do show is that protection was not required for a favorable patient judgment, which is consistent with the eHealth literacy account that such features are hard for end users to appraise [25], but equally consistent with users valuing protection when it is present. The compartmentalized structure of researcher ratings, in contrast, is what one would expect from trained raters applying a normative checklist rather than pursuing a task.

Interpreting the Divergence

Three nonexclusive mechanisms may explain the divergence, and they carry different implications.

If the mechanism is appraisal capacity—privacy features are simply not visible or assessable to end users [25,69]—the appropriate response is disclosure design: making protective

XSL•FO
RenderX

features legible at the point of use, as nutrition labeling does for food.

If the mechanism is privacy calculus—users perceive the features but discount them against immediate benefit [23,24]—disclosure alone will not change judgments, and the appropriate response is to reduce the cost of the protective option rather than to inform users about it.

If the mechanism is displaced trust—users infer safety from health-service branding or from CDC association—then the risk is concentrated precisely in the apps that appear most institutionally credible, and the appropriate response is certification rather than user education.

The present data cannot distinguish these mechanisms, because no qualitative data on evaluator reasoning were collected. This is a limitation of the design and the most important priority for follow-up work: think-aloud protocols or postrating interviews with a larger patient panel would discriminate among the three, and the practical recommendation differs substantially depending on which holds.

Illustrative Cases and Design Implications

Case-level analysis identifies apps that instantiate each pathway and, more usefully for developers, the concrete features they share; these features are the level at which user-centered design guidance for HIV apps has previously been formulated [79].

Wu Ai Fang Hua and Ai Yi Jian illustrate the researcher-oriented pathway R3. Common features include a structured knowledge base with attributed sources and update dates, medication and follow-up reminders, and a web-based community component. The design lesson is not "score highly on four dimensions," but that content credibility signals—visible sourcing, dating, and institutional attribution—are what distinguish these apps from otherwise similar applets.

Chabei illustrates the patient-oriented pathway P3. Its distinguishing features are a shallow navigation hierarchy in which testing appointment booking is reachable in 2 taps from launch, plain-language content without clinical jargon, and a peer community area. Its high outcome membership (0.97) relative to its configuration membership (0.68) suggests that features outside the MARS/uMARS subscales contribute to patient satisfaction—most plausibly the anonymity and low-friction access that a WeChat applet affords, since no download, registration, or app-store trace is required.

Configuration P4 is the practically relevant pathway for resource-constrained developers: high ratings from this patient panel were achieved through functionality and information quality alone, without investment in engagement or visual design. For county-level CDC teams building service applets on limited budgets, this suggests that reliable booking and accurate, current content are the necessary investments, and that gamified or visually elaborate features are optional rather than foundational.

Conversely, configuration R5 contained a single marginal case (Beijing AIDS Association: 0.58, 0.59) whose recognition rested on engagement despite low scores elsewhere, a reminder that expert endorsement does not transfer automatically to user acceptance.

Recommendations for Developers

Overview

The configurations and the illustrative cases translate into a small number of concrete design decisions. They are stated below at feature level, because that is the level at which a development team acts, and recommendations that follow directly from the observed cases are distinguished from those that are proposals consistent with the findings but not themselves tested here. Table 11 sets each against the configuration from which it derives.

XSL•FO
RenderX

Table 11. Feature-level design recommendations derived from the observed configurations and illustrative cases[a].

Configuration (panel)	Illustrative cases	Observed feature pattern	Recommendation	Evidential status
P3, P4 (patients)	Chabei	Core service task 2 taps from launch; content readable without registration	Keep booking, reminders, and result checking within 2 taps; allow anonymous reading	Follows from observed cases
R3 (researchers)	Wu Ai Fang Hua; Ai Yi Jian	Attributed and dated knowledge base; reminders; community area	Show source, date, and reviewing institution on every content item; keep a visible update log	Follows from observed cases
R1, R2, R4 (researchers) vs P1-P4 (patients)	—[b]	Privacy protection core in 3 researcher configurations; present in none of the patient configurations	Enable protection by default; keep notification text free of HIV-related content; summarize storage on one screen at first launch	Proposal consistent with the finding
P4 (patients)	—	High ratings from functionality and information quality alone	Fund booking reliability and content currency first; treat engagement and visual design as later increments	Follows from observed cases
R5 (researchers)	Beijing AIDS Association	Engagement present, other dimensions low; single marginal case	Do not treat engagement as a substitute for the service task: this route appeared only in the researcher solution	Single marginal case; researcher side only
P3 (patients)	Chabei	Outcome membership (0.97) above configuration membership (0.68); applet format	Consider applet delivery where feasible, treating the absence of download, icon, and store record as a privacy feature	Interpretation; not directly measured
Across panels	—	Divergence between panels in the appraisal of privacy protection	Prefer platform- or CDC-level privacy certification to user-facing privacy education	Policy proposal consistent with the finding

[a]Recommendations marked as proposals are design implications consistent with the observed configurations but not themselves tested in this study. Configuration labels follow Table 8.

[b]Not available.

Keep the Primary Service Task Within 2 Taps of Launch

Functionality entered all 4 patient configurations and 3 of the 5 researcher ones, and it is the condition retained as the candidate necessary condition for the patient panel; Chabei, the illustrative case for the patient pathway P3, placed testing appointment booking 2 taps from launch. Appointment booking, reminder setting, and result checking should not sit behind a menu tree, a splash sequence, or a sign-in screen.

Do Not Require Registration to Read Information

Information quality entered every patient configuration except P1, and registration is the point at which an identifiable record is created. The recruitment experience reported in Methods indicates how heavily that possibility weighs on this population. Content should be readable anonymously, with account creation required only for functions that cannot operate without it.

Make Content Provenance Visible on Every Item

The 2 apps illustrating R3, Wu Ai Fang Hua and Ai Yi Jian, were distinguished from otherwise similar applets by a knowledge base whose entries carried attributed sources and update dates. Each content item should therefore name its source, carry a publication or revision date and, where applicable, identify the institution that reviewed it, with an in-app record of when the content set was last updated as a whole.

Treat Protective Settings as Defaults Rather Than as Options

Privacy protection was a core condition in 3 researcher configurations and appeared as a present condition in none of the patient ones. Whatever the mechanism, a design that requires the user to seek protection out will under-protect this population. Proposals consistent with this finding, though not tested in this study, are to enable any available protection by default rather than on request, to write notification and reminder text that never names HIV status or medication, and to present what is stored and where on a single plain-language screen at first launch rather than within a terms-of-use document.

Fund Function and Content Before Engagement and Visual Design

Configuration P4 reached high ratings in the patient panel on functionality and information quality alone. For a county-level team, this makes booking reliability and content currency the first 2 line items and gamified or visually elaborate features later increments. Configuration R5 makes a related point from the other side. One app reached a favorable researcher judgment on engagement alone, with functionality, aesthetics, and information quality all low; but that route appeared only in the researcher solution, rests on a single case, and has the lowest consistency and the lowest raw coverage of any configuration reported. Engagement can carry an expert's overall impression of an app, and on the evidence here it does not stand in for the service task on the user side.

XSL•FO
RenderX

Consider an Applet Build Where the Service Permits It

The outcome membership for Chabei (0.97) exceeded its configuration membership (0.68), which suggests a contribution from something outside the rated dimensions, most plausibly that a WeChat applet requires no download, leaves no icon on the home screen, and creates no app-store record. Where the service can be delivered in that format, the distribution format is itself a privacy feature. This reading is an interpretation of a single case rather than a measured effect.

Prefer Certification to User-Facing Privacy Education

Because privacy protection did not appear as a present condition in any patient pathway, publishing more privacy information is unlikely on its own to alter user-side judgments. A visible privacy label issued at platform or CDC level would place the appraisal burden where the capacity to carry it lies. This is a policy proposal consistent with the observed divergence rather than a finding of the study.

Comparison With Prior Research

Prior MARS-based work has documented correlational differences between expert and user ratings [35,80] and has evaluated HIV/AIDS apps against mean-score benchmarks [36-38]. The present study extends this literature by describing how each group's judgments are organized rather than only whether they differ: the same dimensions enter different combinations for different evaluators.

The centrality of functionality is consistent with usability findings across health app domains, including behavioral intervention technologies [81] and mental health apps [82]. This supports the interpretation that functional performance serves as a baseline requirement rather than a differentiator.

The privacy finding aligns with the broader digital health literature documenting weak privacy practice and limited user appraisal of it [68-71,83], and with the privacy paradox [23,24]. For HIV/AIDS populations, the stakes are higher than in most app categories, since disclosure carries risks of stigma, relationship loss, and discrimination [18,19].

Limitations

Overview

Several limitations bound the claims that can be made from these data.

Patient Panel Size

The patient perspective rests on 2 evaluators whose ratings were averaged. Although a 2-person panel is consistent with the evaluator conventions of this literature (Table 1), and although both evaluators completed the full protocol for all 21 apps with narrow limits of agreement, 2 end users cannot capture the heterogeneity of people living with HIV in China. It is therefore not possible to determine from these data whether the patient configurations reflect preferences shared across this population or the perspectives of these 2 individuals. The patient findings are accordingly presented as exploratory and hypothesis-generating: the claim supported by these data is that a configurational logic distinct from the expert one is observable, not that it is the logic of the patient population.

Case Number and Conditions

Twenty-one apps with 6 conditions produce pronounced limited diversity: only 14 of 64 logically possible configurations (researcher data) and 15 of 64 (patient data) contained empirical cases. Several configurations rest on one or two apps, and unique coverage values below 0.05 (R1, R2, R4, and P1) indicate pathways with little independent empirical weight. The sensitivity analyses show the solutions are largely stable under alternative analytic choices, with the exceptions reported in the Robustness and Sensitivity Analyses section of the Results, but stability under respecification is not the same as generalizability to other app populations.

Rater Panel Composition

Researcher ratings came from 3 individuals from 2 institutions; although interrater agreement was close, evaluator effects cannot be excluded [26], and a differently composed expert panel might weight dimensions differently.

Absence of Qualitative Data

No interview or think-aloud data were collected, so the mechanisms proposed for the observed divergence—appraisal capacity, privacy calculus, displaced trust—are interpretations consistent with theory rather than findings.

Privacy Measurement

Even under the graded specification, privacy was assessed from user-facing artifacts. Actual data handling, third-party sharing, and permission behavior were not examined, and apps that display a policy may not comply with it [70,71].

External Validation

Quality was operationalized as rated subjective quality. No downstream outcome—sustained use, adherence, testing uptake, or clinical benefit—was measured, so the configurations describe pathways to a favorable evaluation, not to health impact.

Context and Time

Only Simplified Chinese apps available in early 2023 were examined. The market changes quickly, WeChat applets are a China-specific format, and quality drivers may differ in other linguistic and regulatory settings.

Future Research Directions

Priorities follow directly from the limitations. Recruiting a substantially larger and more diverse patient panel, ideally stratified by age, sex, time since diagnosis, and eHealth literacy, would allow the patient configurations reported here to be tested rather than proposed. Adding qualitative data would discriminate among the mechanisms proposed for the privacy divergence. Extending the analysis to a larger app corpus, or pooling across countries, would reduce limited diversity and permit conditions to be added without exceeding what the case number supports. Linking configurational membership to downstream use and clinical outcomes would establish whether pathways to high perceived quality are also pathways to benefit. Finally, experimental manipulation of specific features would move the design implications from association toward causal evidence.

XSL•FO
RenderX

Conclusions

Researcher and patient evaluations of Chinese HIV/AIDS mobile apps followed different configurational logics in this exploratory study. Privacy protection entered only the researcher pathways to high subjective quality, whereas the patient panel's pathways combined engagement, functionality, aesthetics, and information quality with its absence; functionality entered all 4 patient configurations and 3 of the 5 researcher ones, and met the necessity threshold in the patient data at the baseline calibration. Multiple equifinal pathways were identified in each group, so no single feature combination is required for a favorable evaluation.

These patterns were largely stable across calibration anchors, consistency thresholds, and an alternative graded operationalization of privacy; the exceptions are that 1 researcher configuration did not survive the stricter PRI cutoffs and that the patient necessity result held only at the baseline anchors. They rest, however, on 21 apps rated by 3 researchers and 2 patients, and the patient findings in particular should be read as hypotheses about how end users form quality judgments rather than as estimates of what people living with HIV prefer.

Two implications are nonetheless warranted. First, expert and end user assessments are not interchangeable, and app recommendation and development processes that rely on either alone risk producing apps that are respectively unused or underprotected. Second, privacy protection was rare in this market and was not required for a favorable patient judgment, although the apps the patient panel rated highest did offer it. In a population for which the cost of unrecognized disclosure risk is not measured in inconvenience, protection therefore needs to be ensured by developers and regulators rather than left to user demand.

Acknowledgments

The authors sincerely acknowledge the editors and anonymous reviewers for their insights and comments to further improve the quality of the paper. The authors also thank the HIV prevention and control team of the Bengbu Center for Disease Control and Prevention, which carried out the recruitment of the patient evaluators, applied the eligibility criteria within its own follow-up records, and collected and retained the consent documentation on behalf of the study. The authors declare that no generative AI tools were used in the preparation of this paper.

Data Availability

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

Funding

This work was supported by the Youth Project of the Anhui Provincial Social Science Planning Fund (AHSKYQ2025D110). The funder had no involvement in this study's design, data collection, analysis, interpretation, or the writing of this paper.

Authors' Contributions

PL, CY, and FW conceived the research idea and jointly developed the study design. PL prepared the first draft of the paper. FW supervised the study. CY was responsible for data curation and reanalysis. All authors contributed to revising the paper and approved the final version for submission.

Conflicts of Interest

None declared.

References

1. Joint United Nations Programme on HIV/AIDS. Global HIV & AIDS statistics — fact sheet. UNAIDS. URL: https://www.unaids.org/en/resources/fact-sheet
2. Muessig KE, Nekkanti M, Bauermeister J, Bull S, Hightow-Weidman LB. A systematic review of recent smartphone, internet and Web 2.0 interventions to address the HIV continuum of care. Curr HIV/AIDS Rep 2015;12(1):173-190 [FREE Full text] [doi: 10.1007/s11904-014-0239-3] [Medline: 25626718]
3. Schnall R, Travers J, Rojas M, Carballo-Diéguez A. eHealth interventions for HIV prevention in high-risk men who have sex with men: a systematic review. J Med Internet Res 2014;16(5):e134 [FREE Full text] [doi: 10.2196/jmir.3393] [Medline: 24862459]
4. Finitsis DJ, Pellowski JA, Johnson BT. Text message intervention designs to promote adherence to antiretroviral therapy (ART): a meta-analysis of randomized controlled trials. PLoS One 2014;9(2):e88166 [FREE Full text] [doi: 10.1371/journal.pone.0088166] [Medline: 24505411]
5. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]

XSL•FO
RenderX

6. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR mHealth uHealth 2015;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
7. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR mHealth uHealth 2015;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
8. Grundy QH, Wang Z, Bero LA. Challenges in assessing mobile health app quality: a systematic review of prevalent and innovative methods. Am J Prev Med 2016;51(6):1051-1059. [doi: 10.1016/j.amepre.2016.07.009] [Medline: 27659122]
9. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
10. Vaghefi I, Tulu B. The continued use of mobile health apps: insights from a longitudinal study. JMIR mHealth uHealth 2019;7(8):e12983 [FREE Full text] [doi: 10.2196/12983] [Medline: 31469081]
11. Nouri R, Kalhori SRN, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inf Assoc 2018;25(8):1089-1098. [doi: 10.1093/jamia/ocy050] [Medline: 29788283]
12. Hutton L, Price BA, Kelly R, McCormick C, Bandara AK, Hatzakis T, et al. Assessing the privacy of mHealth apps for self-tracking: heuristic evaluation approach. JMIR mHealth uHealth 2018;6(10):e185 [FREE Full text] [doi: 10.2196/mhealth.9217] [Medline: 30348623]
13. Bol N, Helberger N, Weert JCM. Differences in mobile health app use: a source of new digital inequalities? Inf Soc 2018;34(3):183-193. [doi: 10.1080/01972243.2018.1438550]
14. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
15. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
16. LeBeau K, Huey LG, Hart M. Assessing the quality of mobile apps used by occupational therapists: evaluation using the user version of the Mobile Application Rating Scale. JMIR mHealth uHealth 2019;7(5):e13019. [doi: 10.2196/13019] [Medline: 31140436]
17. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR mHealth uHealth 2016;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
18. Gillette E, Naanyu V, Nyandiko W, Chory A, Scanlon M, Aluoch J, et al. HIV-related stigma shapes research participation for youth living with HIV in Kenya. J Int Assoc Providers AIDS Care (JIAPAC) 2023;22:23259582231170732 [FREE Full text] [doi: 10.1177/23259582231170732] [Medline: 37101381]
19. Earnshaw VA, Chaudoir SR. From conceptualizing to measuring HIV stigma: a review of HIV stigma mechanism measures. AIDS Behav 2009;13(6):1160-1177. [doi: 10.1007/s10461-009-9593-3] [Medline: 19636699]
20. Goodhue DL, Thompson RL. Task-technology fit and individual performance. MIS Q 1995;19(2):213-236. [doi: 10.2307/249689]
21. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
22. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157-178. [doi: 10.2307/41410412]
23. Dinev T, Hart P. An extended privacy calculus model for e-commerce transactions. Inf Syst Res 2006;17(1):61-80. [doi: 10.1287/isre.1060.0080]
24. Barth S, de Jong MDT. The privacy paradox–investigating discrepancies between expressed privacy concerns and actual online behavior–a systematic literature review. Telematics Inf 2017;34(7):1038-1058. [doi: 10.1016/j.tele.2017.04.013]
25. Norman CD, Skinner HA. eHEALS: the eHealth Literacy Scale. J Med Internet Res 2006;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
26. Hertzum M, Jacobsen NE. The evaluator effect: a chilling fact about usability evaluation methods. Int J Hum Comput Interact 2001;13(4):421-443. [doi: 10.1207/s15327590ijhc1304_05]
27. Woodside AG. Moving beyond multiple regression analysis to algorithms: calling for adoption of a paradigm shift from symmetric to asymmetric thinking in data analysis and crafting theory. J Bus Res 2013;66(4):463-472. [doi: 10.1016/j.jbusres.2012.12.021]
28. Misangyi VF, Greckhamer T, Furnari S, Fiss PC, Crilly D, Aguilera R. Embracing causal complexity: the emergence of a neo-configurational perspective. J Manage 2017;43(1):255-282. [doi: 10.1177/0149206316679252]
29. Greckhamer T, Misangyi VF, Elms H, Lacey R. Using qualitative comparative analysis in strategic management research. Organ Res Methods 2008;11(4):695-726. [doi: 10.1177/1094428107302907]
30. Ragin CC. Fuzzy-Set Social Science. Chicago, IL: University of Chicago Press; 2000.

XSL•FO
RenderX

31. Ragin CC. Qualitative comparative analysis using fuzzy sets (fsQCA). In: Rihoux B, Ragin CC, editors. Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques. Thousand Oaks, CA: SAGE Publications, Inc; 2009:87-122.
32. Thomas J, O'Mara-Eves A, Brunton G. Using qualitative comparative analysis (QCA) in systematic reviews of complex interventions: a worked example. Syst Rev 2014;3:67 [FREE Full text] [doi: 10.1186/2046-4053-3-67] [Medline: 24950727]
33. Singh K, Drouin K, Newmark LP, Rozenblum R, Lee J, Landman A, et al. Developing a framework for evaluating the patient engagement, quality, and safety of mobile health applications. Issue Brief (Commonw Fund) 2016;5:1-11. [Medline: 26934758]
34. Terhorst Y, Rathner EM, Baumeister H, Sander LB. ["Help from the app store?": a systematic review of depression apps in German app stores] [Article in German]. Verhaltenstherapie 2018 Jun 7;28(2):101-112 [FREE Full text] [doi: 10.1159/000481692]
35. Liu P, Wang L, Wang F. Evaluation of Chinese HIV mobile apps by researchers and patients with HIV: quality evaluation study. JMIR mHealth uHealth 2024;12:e52573 [FREE Full text] [doi: 10.2196/52573] [Medline: 38277215]
36. Fernández-Sánchez H, Salazar-Alberto J, Fernández-Delgado JCM, Galvin A, Mugavero MJ, Rodriguez-Diaz CE, et al. Mobile apps for HIV and sexually transmitted infection prevention in Canada, Mexico, and the United States: environmental scan. JMIR mHealth uHealth 2025;13:e72009 [FREE Full text] [doi: 10.2196/72009] [Medline: 41237280]
37. Raeesi A, Khajouei R, Ahmadian L. Evaluating and rating HIV/AIDS mobile apps using the feature-based application rating method and mobile app rating scale. BMC Med Inf Decis Making 2022;22(1):281 [FREE Full text] [doi: 10.1186/s12911-022-02029-8] [Medline: 36310157]
38. Raeesi A, Khajouei R, Ahmadian L. Evaluation of HIV/AIDS-related mobile health applications content using an evidence-based content rating tool. BMC Med Inf Decis Making 2021;21(1):135. [doi: 10.1186/s12911-021-01498-7] [Medline: 33892705]
39. Yang G, Long J, Luo D, Xiao S, Kaminga AC. The characteristics and quality of mobile phone apps targeted at men who have sex with men in china: a window of opportunity for health information dissemination? JMIR mHealth uHealth 2019;7(3):e12573 [FREE Full text] [doi: 10.2196/12573] [Medline: 30916658]
40. Pellowski JA, Kalichman SC, Matthews KA, Adler N. A pandemic of the poor: social disadvantage and the U.S. HIV epidemic. Am Psychol 2013;68(4):197-209 [FREE Full text] [doi: 10.1037/a0032694] [Medline: 23688088]
41. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
42. National Health Commission, Ministry of Education, Ministry of Science and Technology, National Administration of Traditional Chinese Medicine. Measures for Ethical Review of Life Science and Medical Research Involving Humans (Guo Wei Ke Jiao Fa 2023 No. 4) [in Chinese]. Central People's Government of the People's Republic of China. URL: https://www.gov.cn/zhengce/zhengceku/2023-02/28/content_5743658.htm [accessed 2026-10-06]
43. Yang S, Bui CN, Park K. Mobile health apps for breast cancer: content analysis and quality assessment. JMIR mHealth uHealth 2023;11:e43522 [FREE Full text] [doi: 10.2196/43522] [Medline: 36821352]
44. Chasiotis G, Stoyanov SR, Karatzas A, Gravas S. Greek validation of the user version of the Mobile Application Rating Scale (uMARS). J Int Med Res 2023;51(3):03000605231161213 [FREE Full text] [doi: 10.1177/03000605231161213] [Medline: 36924222]
45. Domnich A, Arata L, Amicizia D, Signori A, Patrick B, Stoyanov S, et al. Development and validation of the Italian version of the Mobile Application Rating Scale and its generalisability to apps targeting primary prevention. BMC Med Inf Decis Making 2016;16:83 [FREE Full text] [doi: 10.1186/s12911-016-0323-2] [Medline: 27387434]
46. Hee Ko KK, Kim SK, Lee Y, Lee JY, Stoyanov SR. Validation of a Korean version of Mobile App Rating Scale (MARS) for apps targeting disease management. Health Informatics J 2022;28:14604582221091975 [FREE Full text] [doi: 10.1177/14604582221091975] [Medline: 35404685]
47. Messner EM, Terhorst Y, Barke A, Baumeister H, Stoyanov S, Hides L, et al. The German version of the Mobile App Rating Scale (MARS-G): development and validation study. JMIR mHealth uHealth 2020;8(3):e14479 [FREE Full text] [doi: 10.2196/14479] [Medline: 32217504]
48. Miró J, Llorens-Vernet P. Assessing the quality of mobile health-related apps: interrater reliability study of two guides. JMIR mHealth uHealth 2021;9(4):e26471. [doi: 10.2196/26471] [Medline: 33851931]
49. Shen C, Jiang B, Yang Q, Wang C, Lu KZ, Gu M, et al. Mobile apps for drug-drug interaction checks in Chinese app stores: systematic review and content analysis. JMIR mHealth uHealth 2021;9(6):e26262 [FREE Full text] [doi: 10.2196/26262] [Medline: 33962910]
50. Schmeelk S, Davis A, Li Q, Shippey C, Utah M, Myers A, et al. Monitoring symptoms of COVID-19: review of mobile apps. JMIR mHealth uHealth 2022;10(6):e36065 [FREE Full text] [doi: 10.2196/36065] [Medline: 35609313]
51. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR mHealth uHealth 2015;3(3):e82 [FREE Full text] [doi: 10.2196/mhealth.4328] [Medline: 26290327]
52. Amor-García M, Collado-Borrell R, Escudero-Vilaplana V, Melgarejo-Ortuño A, Herranz-Alonso A, Arranz Arija J, et al. Assessing apps for patients with genitourinary tumors using the Mobile Application Rating Scale (MARS): systematic

XSL•FO
RenderX

search in app stores and content analysis. JMIR mHealth uHealth 2020;8(7):e17609 [FREE Full text] [doi: 10.2196/17609] [Medline: 32706737]
53. Yamamoto K, Ito M, Sakata M, Koizumi S, Hashisako M, Sato M, et al. Japanese version of the Mobile App Rating Scale (MARS): development and validation. JMIR mHealth uHealth 2022;10(4):e33725 [FREE Full text] [doi: 10.2196/33725] [Medline: 35197241]
54. Bricca A, Pellegrini A, Zangger G, Ahler J, Jäger M, Skou ST. The quality of health apps and their potential to promote behavior change in patients with a chronic condition or multimorbidity: systematic search in app store and Google Play. JMIR mHealth uHealth 2022;10(2):e33168 [FREE Full text] [doi: 10.2196/33168] [Medline: 35119367]
55. Kim BY, Sharafoddini A, Tran N, Wen EY, Lee J. Consumer mobile apps for potential drug-drug interaction check: systematic review and content analysis using the Mobile App Rating Scale (MARS). JMIR mHealth uHealth 2018;6(3):e74 [FREE Full text] [doi: 10.2196/mhealth.8613] [Medline: 29592848]
56. Santo K, Richtering SS, Chalmers J, Thiagalingam A, Chow CK, Redfern J. Mobile phone apps to improve medication adherence: a systematic stepwise process to identify high-quality apps. JMIR mHealth uHealth 2016;4(4):e132 [FREE Full text] [doi: 10.2196/mhealth.6742] [Medline: 27913373]
57. Romero RL, Kates F, Hart M, Ojeda A, Meirom I, Hardy S. Quality of deaf and hard-of-hearing mobile apps: evaluation using the Mobile App Rating Scale (MARS) with additional criteria from a content expert. JMIR mHealth uHealth 2019;7(10):e14198 [FREE Full text] [doi: 10.2196/14198] [Medline: 31670695]
58. Narrillos-Moraza Á, Gómez-Martínez-Sagrera P, Amor-García M, Escudero-Vilaplana V, Collado-Borrell R, Villanueva-Bueno C, et al. Mobile apps for hematological conditions: review and content analysis using the Mobile App Rating Scale. JMIR mHealth uHealth 2022;10(2):e32826. [doi: 10.2196/32826] [Medline: 35171109]
59. Báez GN, Rodríguez RH, Fernández GM, Abdel-Kader ML. Smartphone apps for patients with hematologic malignancies: systematic review and evaluation of content. JMIR mHealth uHealth 2022;10(9):e35851. [doi: 10.2196/35851] [Medline: 36121688]
60. Liu X, Jin F, Hsu J, Li D, Chen W. Comparing smartphone apps for traditional Chinese medicine and modern medicine in China: systematic search and content analysis. JMIR mHealth uHealth 2021;9(3):e27406. [doi: 10.2196/27406] [Medline: 33759798]
61. Cheah KJ, Manaf ZA, Ludin AFM, Razalli NH, Mohd Mokhtar N, Md Ali SH. Mobile apps for common noncommunicable disease management: systematic search in app stores and evaluation using the Mobile App Rating Scale. JMIR mHealth uHealth 2024;12:e49055 [FREE Full text] [doi: 10.2196/49055] [Medline: 38532298]
62. Bardus M, Awada N, Ghandour LA, Fares EJ, Gherbal T, Al-Zanati T, et al. The Arabic version of the Mobile App Rating Scale: development and validation study. JMIR mHealth uHealth 2020;8(3):e16956. [doi: 10.2196/16956] [Medline: 32130168]
63. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016;13:35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
64. Greckhamer T, Furnari S, Fiss PC, Aguilera RV. Studying configurations with qualitative comparative analysis: best practices in strategy and organization research. Strategic Organ 2018;16(4):482-495. [doi: 10.1177/1476127018786487]
65. Greckhamer T, Misangyi VF, Fiss PC. The two QCAs: from a small-N to a large-N set theoretic approach. Res Sociol Organ 2013;38:49-75. [doi: 10.1108/S0733-558X(2013)0000038007]
66. Ragin CC. Redesigning Social Inquiry: Fuzzy Sets and Beyond. Chicago, IL: University of Chicago Press; 2008.
67. Schneider CQ, Wagemann C. Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis. Cambridge: Cambridge University Press; 2012.
68. Benjumea J, Ropero J, Rivera-Romero O, Dorronzoro-Zubiete E, Carrasco A. Assessment of the fairness of privacy policies of mobile health apps: scale development and evaluation in cancer apps. JMIR mHealth uHealth 2020;8(7):e17134 [FREE Full text] [doi: 10.2196/17134] [Medline: 32720913]
69. Benjumea J, Ropero J, Rivera-Romero O, Dorronzoro-Zubiete E, Carrasco A. Privacy assessment in mobile health apps: scoping review. JMIR mHealth uHealth 2020;8(7):e18868 [FREE Full text] [doi: 10.2196/18868] [Medline: 32459640]
70. Papageorgiou A, Strigkos M, Politou E, Alepis E, Solanas A, Patsakis C. Security and privacy analysis of mobile health applications: the alarming state of practice. IEEE Access 2018;6:9390-9403. [doi: 10.1109/ACCESS.2018.2799522]
71. Huckvale K, Torous J, Larsen ME. Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation. JAMA Netw Open 2019;2(4):e192542 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.2542] [Medline: 31002321]
72. Fiss PC. A set-theoretic approach to organizational configurations. Acad Manage Rev 2007;32(4):1180-1198. [doi: 10.5465/amr.2007.26586092]
73. Ragin CC. Set relations in social research: evaluating their consistency and coverage. Polit Anal 2006;14(3):291-310. [doi: 10.1093/pan/mpj019]
74. Pappas IO, Woodside AG. Fuzzy-set qualitative comparative analysis (fsQCA): guidelines for research practice in information systems and marketing. Int J Inf Manage 2021;58:102310. [doi: 10.1016/j.ijinfomgt.2021.102310]

XSL•FO
RenderX

75. Dul J. Necessary condition analysis (NCA): logic and methodology of "necessary but not sufficient" causality. Organ Res Methods 2016;19(1):10-52. [doi: 10.1177/1094428115584005]
76. Fiss PC. Building better causal theories: a fuzzy set approach to typologies in organization research. AMJ 2011;54(2):393-420. [doi: 10.5465/amj.2011.60263120]
77. Schneider CQ, Wagemann C. Standards of good practice in qualitative comparative analysis (QCA) and fuzzy-sets. Comp Sociol 2010;9(3):397-418. [doi: 10.1163/156913210x12493538729793]
78. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476):307-310. [Medline: 2868172]
79. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inf 2016;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
80. Powell AC, Torous J, Chan S, Raynor GS, Shwarts E, Shanahan M, et al. Interrater reliability of mHealth app rating measures: analysis of top depression and smoking cessation apps. JMIR mHealth uHealth 2016;4(1):e15 [FREE Full text] [doi: 10.2196/mhealth.5176] [Medline: 26863986]
81. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
82. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018;21(3):116-119 [FREE Full text] [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
83. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Privacy and security in mobile health apps: a review and recommendations. J Med Syst 2015;39(1):181. [doi: 10.1007/s10916-014-0181-3] [Medline: 25486895]

Abbreviations

CDC: Center for Disease Control and Prevention
fsQCA: fuzzy-set qualitative comparative analysis
ICC: intraclass correlation coefficient
MARS: Mobile App Rating Scale
mHealth: mobile health
PRI: proportional reduction in inconsistency
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QCA: qualitative comparative analysis
uMARS: user version of Mobile App Rating Scale
UTAUT: unified theory of acceptance and use of technology
UTAUT2: extended unified theory of acceptance and use of technology

Edited by Z Ni; submitted 03.Dec.2025; peer-reviewed by F Carrouel, R Xu, S Liu; comments to author 10.Aug.2026; revised version received 29.Sep.2026; accepted 30.Sep.2026; published 22.Oct.2026.

Please cite as:
Liu P, Wang F, Yang C
Configurations of Quality Dimensions Associated With Researcher and Patient Ratings of HIV/AIDS Mobile Health Apps in China: Fuzzy-Set Qualitative Comparative Analysis
JMIR Mhealth Uhealth 2026;14:e88896
URL: https://mhealth.jmir.org/2026/1/e88896
doi:10.2196/88896
PMID:

©Peng Liu, Fuzhi Wang, Changzheng Yang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 22.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Privacy Policy Compliance of Mobile Sports and Health Apps in China: Scale Development, Data Analysis, and Prospects for Regulatory Reform

Rengui Guo*, PhD; Fanhong Chen*, PhD

School of Law, Central South University, No.932, Lushan South Road, Yuelu District, Changsha, China

*all authors contributed equally

Corresponding Author:
Fanhong Chen, PhD
School of Law, Central South University, No.932, Lushan South Road, Yuelu District, Changsha, China

Abstract

Background: Driven by technological advancements, the proliferation of mobile sports and health apps has revolutionized health management by improving efficiency, cost-effectiveness, and accessibility. While the widespread adoption of these platforms has transformed public health practices and social well-being in China, emerging evidence suggests that inadequacies in their privacy policies may compromise personal information (PI) protection.

Objective: This study aimed to conduct a systematic evaluation of privacy policy compliance among 286 mobile sports and health apps in the Chinese Mainland, benchmarking them against the Personal Information Protection Law and associated PI regulatory guidelines.

Methods: This study develops a privacy policy compliance indicator scale based on the information life cycle and the legal framework for PI protection in the Chinese Mainland. This scale consists of 5 level 1 indicators and 37 level 2 indicators that assess the privacy policy compliance.

Results: The privacy policy compliance of 286 sports and health apps generally performed worse, with only a minimal number (n=11, 3.8%) of apps scoring above 90 points (rated as excellent), nearly half (n=121, 42.3%) of apps scored below 60 points (rated as unqualified). Among the 5 level 1 evaluation indicators for privacy compliance in sports and health apps, the compliance rate for PI collection (mean 74%, SD 25.8%) is the highest, while the compliance rate for PI storage (mean 53.5%, SD 28.4%) is the lowest. The compliance rates for privacy policies across the remaining 3 level 1 evaluation indicators, such as PI usage (mean 54.2%, SD 24.4%), PI entrusted processing, sharing, transferring, and disclosing (mean 62.2%, SD 19.8%), and PI security and feedback (mean 61.7%, SD 21.3%), fall around 60%. Out of 37, 17 level 2 evaluation indicators show a compliance rate below 60%. The compliance rate with privacy policies for 5 level 2 evaluation indicators is exceptionally high, including collection subject (mean 97.2%, SD 16.5%), collection type (mean 99%, SD 10.2%), collection purpose (mean 96.2%, SD 19.3%), reasons for sharing, transferring, and disclosing PI (mean 91.6%, SD 27.8%), and feedback channel (mean 93.4%, SD 24.9%). Notably, 3 indicators exhibit compliance rates below 20%, including sensitive information storage (mean 14%, SD 34.7%), constraints of automatic decision-making (mean 9.4%, SD 29.3%), and deceased user rule (mean 5.2%, SD 22.3%). Authorization for sensitive information (mean 29.4%, SD 45.6%) lagged behind general information (mean 83.6%, SD 37.1%).

Conclusions: Although some apps have established commendable policies, there are gaps that compromise the efficacy of PI protection. Considering this, this paper proposes targeted actions for 3 stakeholders: users, regulators, and legislators. Only through coordinated action can the app ecosystem close the compliance gaps, reduce PI protection risks, and restore user trust in digital services.

(JMIR Mhealth Uhealth 2026;14:e73651) doi:10.2196/73651

KEYWORDS

sports and health apps; personal data; sensitive information; privacy policy compliance; personal information protection; regulatory reform

Introduction

With the popularity of mobile phones and people's increasing interests in sports and health management [1-3], mobile sports and health apps, which are closely related to physical exercise and provide exercise knowledge, assist with physical activities, track workout data, and support health management, have emerged and are constantly growing [3-7]. Currently, the app stores on the 2 major mobile operating systems, Android (Google) and Apple iOS, offer a vast array of apps related to

XSL•FO
RenderX

health monitoring [8,9], calorie management [10-12], fitness guidance [13,14], health consultation [15,16], and other sports and health services [17]. These apps have enhanced the efficiency, cost-effectiveness, and accessibility of sports and health management, enabling people who are concerned about health through sports to learn exercise programs in real-time, track their diet, and improve their overall health condition using the apps [18-21]. It is reported that the extensive usage of these apps has significantly enhanced the social and public health landscape in China [22].

However, the rapid growth of mobile sports and health apps poses a threat to users' personal privacy and data security [23-26]. Streaking privacy, which occurs when sports and health apps collect, process, and use personal exercise and physical health information without authorization, has caused serious legal issues and attracted widespread social attention around the world [27-29]. For instance, the US Federal Trade Commission ordered the Easy Healthcare Corporation to pay a $100,000 civil penalty for its Premom app collecting and sharing a significant amount of users' personal information (PI) with third-party advertisers in violation of privacy protection, eavesdropping, and other applicable laws and regulations [30]. A Norwegian nonprofit organization has discovered that 10 of the most popular apps on Google Play Store, including sports and health apps, transmit users' sensitive PI to third parties without their permission [31]. In China, these challenges are more acute. In December 2018, China's Cyberspace Administration carried out a special regulatory action targeting numerous apps, including those related to sports and health, and discovered that 3496 apps were illegally collecting and using PI [32].

As a response, many countries, including the United States [33,34] and the European Union [35-37], have enacted bills to strengthen PI protection [38]. The United States has enacted laws such as the Health Insurance Portability and Accountability Act [39], the Health Breach Notification Rule [40], and the Children's Online Privacy Protection Rule [41], which stipulate the obligation of health care organizations to safeguard health data, require health care organizations to establish a consent mechanism for sharing patients' health data with third parties, outline the data breach notification mechanism for health care organizations and the legal responsibilities for disclosing patients' sensitive data [42]. The European Union protects PI and privacy in a unified manner through the General Data Protection Regulation, grants 8 rights to users, and requires data processors to process personal data under the individual consent mechanism [43].

In response to the risks of PI leakage and privacy abuse posed by sports and health apps, China's PI protection legal system adopts an approach [44], aiming to protect users' privacy data and promote market development [27]. In 2019, the Cyberspace Administration of China made policies, such as methods for determining illegal collection and use of personal information by apps and provisions on the scope of necessary personal information for common types of mobile internet applications, to regulate the collection and use of PI by apps [45]. In light of the noncoercive nature of these policies and in view of the urgent need for PI protection [46], China enacted the Personal Information Protection Law (PIPL) in 2021 [47]. While public reports held that the PIPL has established a basic legal framework for PI collection, processing, and usage [48-50], academic scholars still argue that the PIPL should be significantly improved [51-54]. Meanwhile, since December 2021, the National Technical Committee 260 on Cybersecurity of Standardization Administration of China has successively made voluntary national standards, such as the Practical Guide to Cybersecurity Standards—Guidelines for the Classification and Grading of Network Data [55], the Information Security Technology - Personal Information Security Engineering Guidelines (PI Guidelines) [56], Information Security Technology—Implementation Guidelines for Notices and Consent in Personal Information Processing [57], and Information Security Technology—Guide for De-identifying Personal Information [58] to classify and grade PI for protection. Additionally, the Ministry of Industry and Information Technology of China regularly reports and requires rectification of apps that violate users' rights and interests [59]. According to public reports, a total of 297 apps were required to rectify their illegal collection of PI in 2024 [60].

Under the above legal requirements, mobile sports and health apps, which collect and use users' PI as the premise of their normal operation, need to establish an internal compliance management system for personal privacy and data security [44]. In this system, developing privacy policies stands in the foreground, because the formulation and implementation of privacy policies ensures that the apps' market operation behavior meets the requirements of laws and regulations, thereby achieving a balance between internal self-discipline and external supervision. As required by PIPL and PI Guidelines, a legitimate and valid privacy policy is comprised of the collection, storage, use, entrusted processing, sharing, transfer, disclosure, consultation, and feedback of PI [61,62]. Accordingly, this paper attempts to measure these aspects of mobile sports and health apps' privacy policy because the degree of PI protection is correlated with how well the app's privacy policy conforms with PIPL and PI Guidelines.

There is a well-established body of work on privacy policies and PI protection in existing studies. For example, Tangari et al [4] explored free fitness apps on the Google Play Store in Australia and found that 88% of the apps will potentially collect user data. Parker et al [63] investigated and rated medicine-related apps for the Android mobile platform available in the medical store category of Google Play in the United Kingdom, United States, Canada, and Australia. Grundy et al [64] explored the risks posed by rated medicines-related apps when sharing user data. Alfawzan et al [65] emphasized the privacy policies of 23 most popular apps on women's health in the Apple operating system and the Android system. Shipp and Blasco [66] assessed the privacy practices of a set of 30 Android menstruation-related apps that track users' reproductive cycle, sex life, and health. Although research on the privacy policy compliance of mobile health (mHealth) apps is expanding, studies specifically examining the privacy policy compliance of mHealth apps closely related to sports still require further in-depth investigation. Scholars have conducted comparative in-depth research on the privacy policy compliance of China's

XSL•FO
RenderX

mHealth apps [67-69], focusing primarily on mobile mental health [29,70,71], mobile hospitals [28,72,73], contact tracing [74,75], traditional Chinese medicine treatments [76], and chronic disease management [22,77]. Research on the privacy policy compliance of Chinese mobile sports and health apps remains limited. Current research on privacy policy compliance for mobile sports and health apps is limited to a few jurisdictions, with a significant lack of attention still directed toward the privacy policy compliance issues of Chinese mobile sports and health apps. Currently, China boasts an exceptionally large user base for mobile sports and health apps. These apps collect users' PI, particularly sensitive data, such as gender, location, exercise routes, heart rate, workout duration, weight, and exercise frequency. If the PI collected by these mobile sports and health apps is not gathered, used, and shared in a lawful, transparent, and trustworthy manner, it not only jeopardizes users' personal privacy and safety but also poses a threat to national data security. Therefore, privacy policy compliance for mobile sports and health apps requires thorough analysis in the Chinese Mainland.

In order to comprehensively evaluate the compliance rate of privacy policies of sports and health apps in the Chinese Mainland, this paper deploys scenario-based contextual analysis [78] and information life cycle theory [79]. Nissenbaum [78] argued that the contextual integrity theory should serve as the cornerstone of privacy protection, because different contextual frameworks shape the norms of information flow across various scenarios, and these norms determine the legitimacy of information processing. There are typically 2 types of PI protection scenarios. The first type vertically categorizes different types of information based on the varying levels of processing risks, and the other type horizontally focuses on the different stages of information processing. Regarding the former, both domestic and international legislation generally adopt a dichotomy to differentiate between sensitive PI and general PI. This paper also follows this classification to evaluate their respective levels of protection compliance. For the latter, combining with the information life cycle stipulated in the PIPL, this study proposes 5 stages of PI collection, PI storage, PI usage, PI entrusted processing, sharing, transferring, and disclosing, and PI security and feedback, and further uses them as 5 primary indicators in the scale.

This study uses the legal framework for PI protection to develop a privacy policy compliance indicator scale and evaluate the compliance degree of privacy policies of 286 sports and health apps. In the Methods section, we focus on selecting, collecting, and analyzing sample apps, developing a compliance indicator scale, and designing the scoring process. In the Results section, we examine the level of PI protection of each privacy policy and report the results of the compliance assessment. In the Discussion section, we summarize the overall landscape and existing shortcomings in privacy policy compliance among sports and health apps and propose regulatory reforms to strengthen PI protection.

This study aims to (1) select, collect, and analyze privacy policies of sports and health apps developed for users in the Chinese Mainland, (2) establish a privacy policy compliance indicator scale based on the PIPL and PI Guidelines [22,27] (3) evaluate the compliance level of privacy policies of sports and health apps by scoring each indicator, and (4) provide suggestions on how to improve PI protection in the Chinese Mainland. This study enhances the discourse on balancing personal information protection with sustainable innovation in sports and health apps. By underscoring the necessity of enhancing privacy policy compliance and offering improvement suggestions, it offers insights for policymakers, developers, operators, and users across various nations regarding PI protection.

Methods

Study Design

We first comprehensively collected 714 apps from the app stores of Google Android and Apple iOS, the 2 mobile phone operating systems in the Chinese Mainland, between August 2 and August 6, 2025. Then we proceeded to examine the program description of apps to review their features and locate their privacy policies on August 6, 2025. Subsequently, we downloaded the privacy policies of the remaining apps, eliminating those that either did not allow a privacy policy download or lacked a privacy policy altogether between August 7 and August 8, 2025. Finally, we analyzed the compliance level of privacy policies of 286 valid sports and health apps with the PIPL and PI Guidelines and further proposed improvement suggestions regarding the gaps identified in the process of compliance evaluation from August 10 to August 20, 2025.

Sample Selection and Inclusion Criteria

This study focused on the privacy policy compliance of sports and health apps. We selected the apps from the app stores of Android and Apple, the 2 popular mobile operating systems in the Chinese mainland. We opened the app stores (Yingyongshangdian in Chinese) section, selected the function of app category (Fenlei in Chinese), and put in keywords "sports and health" (Yundong Jiankang in Chinese) to filter apps. Through the comprehensive collection and search methods described above, we have gathered the initial sports and health sample apps. Then, we carefully reviewed the feature descriptions of each of the initial sports and health sample apps we collected, while simultaneously downloading their privacy policy documents. The apps included in the sample must meet the following 2 criteria. First, they must be explicitly designed for a diverse user base to offer sport guidance, real-time exercise data tracking, fitness education, and other sports and health services, and second, they must be intended for ordinary individuals rather than for sports and health management organizations or governments. The apps with the following criteria were excluded from the sample. First, the app provides only low-fat food options, recipes, and dietary control recommendations. Second, the app provides only psychological issue diagnosis, stress relief, and emotional support functions. Third, the app provides only sleep aid and sleep quality monitoring functions. Fourth, the app solely collects physical data, analyzes health issues, and provides fitness recommendations. Fifth, the app exclusively offers exercise instruction content and videos, along with coaching services. Sixth, the app solely provides virtual fitness game services.

XSL•FO
RenderX

Seventh, the app exclusively offers sports event booking, live streaming, and participation services. Eighth, the app solely provides massage services. Ninth, the app either lacks a privacy policy or its privacy policy web page fails to display effectively. By searching, analyzing, and filtering sports and health apps, we obtained the final number of sample apps used for privacy policy compliance analysis.

Development of the Privacy Policy Compliance Indicator Scale

We constructed a privacy policy compliance indicator scale from the perspective of the scenario-based requirement [78] and the information life cycle [79] to assess the compliance between the privacy policies of sports and health apps and the PIPL and the PI Guidelines. The process of developing the privacy policy compliance indicator scale is as follows.

We first comprehensively reviewed the provisions of the PIPL and PI Guidelines and then extracted an appropriate privacy compliance evaluation legal framework applicable to sports and health apps, from which we produce level 1 evaluation indicators. According to the information life cycle, sports and health apps operate through the processes of collection, storage, usage, processing, sharing, transfer, disclosure, information consultation, suggestion, and feedback of PI [22,27]. In light of this, we argue that the legal framework is comprised of five stages: (1) the collection of PI (Articles 6, 13, 14, 17, 28, 29, and 30 of the PIPL; and Articles 2, 4, and 5 of the PI Guidelines); (2) the storage of PI (Articles 17, 19, 39, and 40 of the PIPL; and Article 6 of the PI Guidelines); (3) the use of PI (Articles 6, 14, 15, 16, 24, 45, 46, 47, 48, 49, and 50 of the PIPL; and Articles 7 and 8 of the PI Guidelines); (4) the entrusted processing, the sharing, the transfer, and the disclosure of PI (Articles 21, 22, 23, 26, 27, 38, and 39 of the PIPL; and Article 9 of the PI Guidelines); (5) the consultation and feedback on PI (Articles 50, 57, and 65 of the PIPL; and Article 10 of the PI Guidelines). In accordance with this legal framework, we set 5 level 1 evaluation indicators: the collection of PI, the storage of PI, the use of PI, the entrusted processing, the sharing, the transfer, and the disclosure of PI, and the consultation and feedback on PI.

To conduct a more precise evaluation of the privacy policy, we further divide the 5 level 1 evaluation indicators into 37 level 2 evaluation indicators, thus establishing our privacy policy compliance indicator scale (Table 1). The level 1 evaluation indicators include PI collection, PI storage, PI usage, PI entrusted processing, sharing, transferring, and disclosing, and PI security and feedback. Specifically, the level 1 evaluation indicator of PI collection comprises 9 level 2 evaluation indicators. They are the collection subject, policy update, app scope, collection type, collection purpose, processing rule, authorization for general information, authorization for sensitive information, and exceptions for explicit authorization. Another level 1 evaluation indicator of PI storage consists of 4 level 2 evaluation indicators: storage time, storage place, PI deidentification, and sensitive information storage. We further break down the level 1 evaluation indicator of PI usage into 14 level 2 evaluation indicators, including PI access control, deidentified display and use, purpose restrictions for PI usage, authorization for the usage purpose change, constraints of automatic decision-making, query right, correction right, deletion right, copy right, account cancelation right, withdraw or change authorization, asking response right, deceased user rule, and complaint mechanism. Concerning the level 1 evaluation indicator of PI entrusted processing, sharing, transferring, and disclosing, we further categorize it into 5 level 2 evaluation indicators: requirements for entrusted controller processing PI, reasons for sharing, transferring, and disclosing PI, security measures of sharing, transferring, and disclosing PI, special circumstances without consent, and cross-border transmission requirements. The last level 1 evaluation indicator is the PI security and feedback, which encompasses 5 level 2 evaluation indicators of security incident response mechanism, security event notification mechanism, feedback channel, feedback period, and external dispute resolution mechanisms.

XSL•FO
RenderX

Table . Level 1 and level 2 evaluation indicators of privacy policy compliance of sports and health apps.

Level 1 evaluation indicator	Level 2 evaluation indicator
PI[a] collection	• Collection subject • Policy update • Application scope • Collection type • Collection purpose • Processing rule • Authorization for general information • Authorization for sensitive information • Exceptions for explicit authorization
PI storage	• Storage time • Storage place • PI deidentification • Sensitive information storage
PI usage	• PI access control • Deidentified display and use • Purpose restrictions for PI usage • Authorization for the usage purpose change • Constraints of automatic decision-making • Query rightCorrection right • Deletion rightCopy right • Account cancelation right • Withdraw or change authorization • Asking response right • Deceased user rule • Complaint mechanism
PI entrusted processing, sharing, transferring, and disclosing	• Requirements for entrusted controller processing PI • Reasons for sharing, transferring, and disclosing PI • Security measures of sharing, transferring, and disclosing PI • Special circumstances without consent • Cross-border transmission requirements
PI security and feedback	• Security incident response mechanism • Security event notification mechanism • Feedback channel • Feedback period • External dispute resolution mechanisms

[a]PI: personal information.

The PIPL and PI Guidelines are of significant importance for the protection of personal data and privacy. These provisions are legally mandatory rules that must be strictly adhered to. Therefore, we treat all 5 level 1 evaluation indicators equally and assign them a value of 1. It means that if these 37 level 2 evaluation indicators are accurately and effectively stated in the privacy policy of the sports and health app, they are assigned a value of 1, and if not, they are assigned a value of 0.

After classifying the level 1 evaluation indicators and applying the scoring rule, we got a total score (37 points) for each app's privacy policies. Then, the final score is calculated on a 100-point scale. This means each app's privacy policy score (on a 37-point scale) is converted to a 100-point scale to determine the final score. The evaluation results were further divided into 4 levels: excellent for a compliance score above 90 points, good for a compliance score in the range of 80 - 90 points, qualified for a compliance score in the range of 60 - 79 points, and unqualified for a compliance score below 60 points. The privacy policy compliance tiering of sports and health apps is illustrated in Table 2.

Finally, we calculated and summarized the score of each level 2 evaluation indicator for all sample apps. From August 21 to December 25, 2025, 2 independent raters (RG and FC) participated in the scoring process. This process was divided into 2 stages. To ensure reliability and consistency of the scoring results, in the initial scoring stage, all raters independently evaluated 30 randomly selected apps (10.5% of the total), achieving a correlation coefficient of 0.983 (P<.001) among different raters within the same group, indicating that the evaluations were closest to consistency. After evaluation, the raters gathered together to discuss the differences and reasons for the initial scoring and ultimately formed an almost consistent scoring standard. In this way, subjectivity in scoring is limited, although not eliminated. In the second scoring stage, we divided the sample apps into 2 groups of 143 each. Each rater randomly selects a group of apps to score independently.

XSL•FO
RenderX

Table . The privacy policy compliance tiers of sports and health apps.

Final score for apps	Privacy policy compliance tiers
≥90	Excellent
80-90	Good
60-80	Qualified
<60	Unqualified

Results

Sample Collection

From August 2, 2025, we conducted the collection, selection, evaluation, and analysis of the privacy policies of the sports and health apps. We gathered an initial set of 714 sports and health apps from the Android App Store and Apple App Store.

We collected 328 sports and health apps from the Android App Store. After excluding 118 apps due to reasons such as irrelevance to sports and health management services, the final valid count was 210 apps. The reasons and numbers for the 118 excluded apps are as follows: (1) 11 apps were excluded for providing only diet control and fat loss services, (2) 5 apps were excluded for providing only body data collection and testing services, (3) 27 apps were excluded for offering solely exercise instruction and learning services, (4) 17 apps were excluded for providing only virtual gaming services, (5) 35 apps were excluded for offering only sports event booking and live streaming services, (6) 5 apps were excluded for providing only sports venue booking services, (7) 16 apps were excluded for other reasons, and (8) 1 app was excluded because its privacy policy web page could not be accessed.

We collected 386 sports and health apps from the Apple App Store. After excluding 273 apps unrelated to sports and health management services, 113 valid apps remained. The reasons and quantities for excluding the 113 apps are as follows: (1) 28 apps were excluded for providing only dietary control and fat loss services; (2) 45 apps were excluded for offering solely psychological counseling and mental health guidance; (3) 36 apps were excluded for providing only sleep issue monitoring and sleep aid services; (4) 44 apps were excluded for offering exclusively physical data collection and health monitoring services; (5) 3 apps were excluded for solely providing exercise instruction and learning services; (6) 1 app was excluded for solely providing virtual gaming services; (7) 11 apps were excluded for solely providing massage and relaxation services; (8) 10 apps were excluded for solely providing traditional Chinese medicine consultation and treatment services; (9) 59 apps were excluded for other reasons including menstrual cycle prediction and care, vaccination, parenting, etc; (10) 5 apps were excluded for failing to open their privacy policy webpage; and (11) 31 apps were excluded for lacking a privacy policy. Therefore, we excluded 391 apps whose functions were not centered around sports. The reasons and number of apps excluded are listed in Table 3; the names, reasons, and number of apps excluded are presented in Multimedia Appendix 1. Ultimately, we collected 323 valid sports and health apps (the characteristics of the 323 valid sports and health apps are illustrated in Multimedia Appendix 2). The characteristics of the 323 valid sports and health apps are presented in Table 4.

Table . The reasons and number of apps excluded.

Reasons for apps excluded	Apps excluded (n=391), n
Dietary control	40
Mental health	45
Sleep monitoring	36
Body monitoring	49
Physical education	30
Video games	18
Massage services	11
Traditional Chinese medicine diagnosis and treatment	10
Sports events	35
Venue booking	5
Other reasons (menstrual care, childcare, sports knowledge dissemination, etc)	75
Unable to open the webpage	6
No privacy policy	31

XSL•FO
RenderX

Table . The characteristics of the 323 valid sports and health apps.

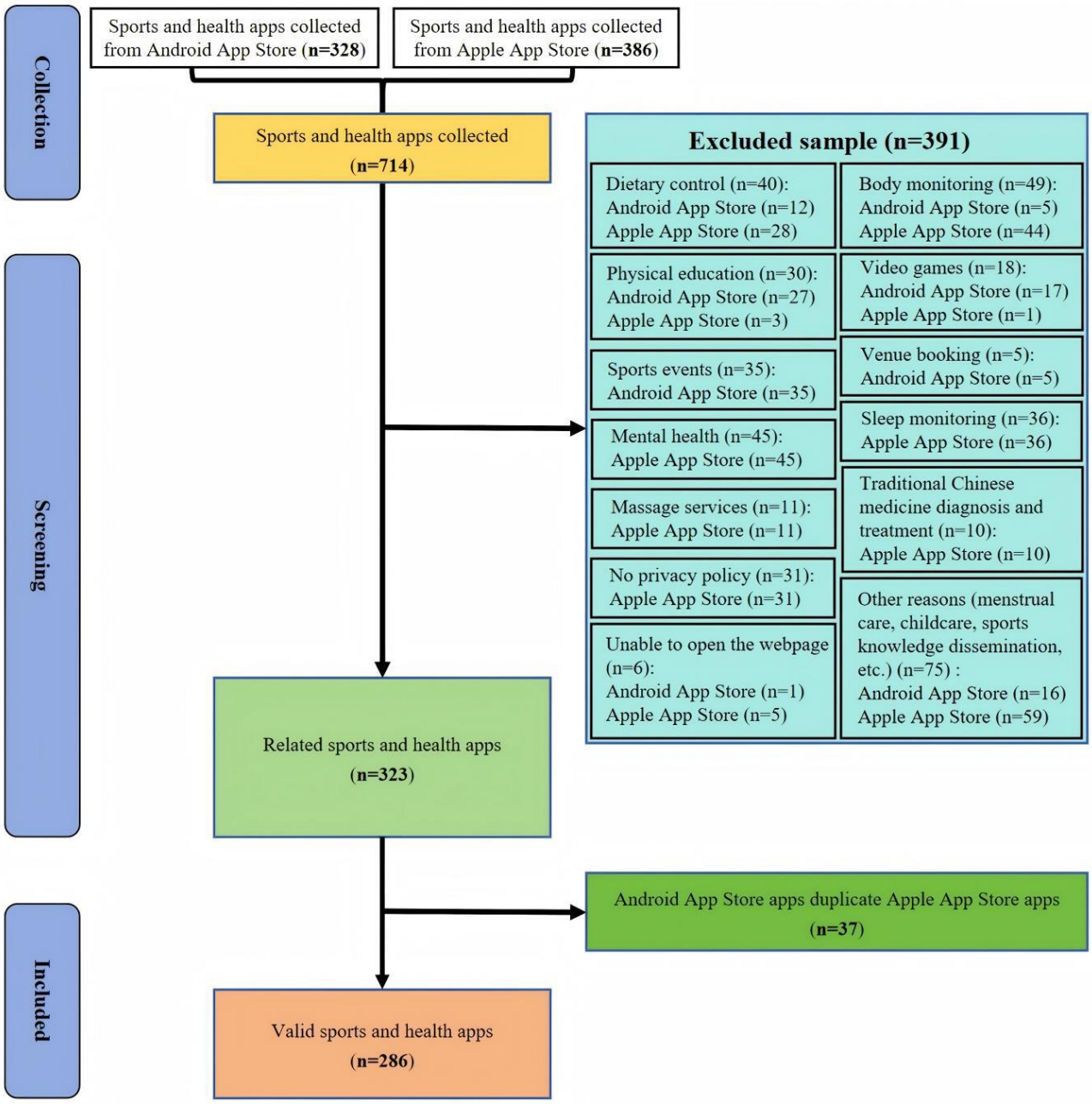
Characteristics	Apps (n=323), n (%)
Apps source	
Android App Store	210 (65)
Apple App Store	113 (35)
Number of Android apps downloads	
≤10,000	46 (21.9)
10,000-100,000	42 (20)
100,000-10,00,000	57 (27.1)
1,000,000-1,00,00,000	43 (20.5)
10,000,000-1,00,000,000	18 (8.6)
>10,00,00,000	4 (1.9)
Age-based app classification in Android apps	
≥3	90 (42.8)
≥8	18 (8.6)
≥12	30 (14.3)
≤16	25 (11.9)
≥18	47 (22.4)
Age-based app classification of Apple apps	
≥4	76 (67.3)
≥9	3 (2.6)
≥12	13 (11.5)
≤17	21 (18.6)
App rating of Android apps	
<1	60 (28.6)
1-2	38 (18.1)
2-3	32 (15.2)
3≤n-4	37 (17.6)
4-5	25 (11.9)
5	18 (8.6)
App rating of Android apps	
<1	2 (1.8)
12	2 (1.8)
23	10 (8.8)
34	14 (12.4)
45	81 (71.7)
5	4 (3.5)
App rating of total apps	
1	62 (19.2)
1-2	40 (12.4)
2-3	42 (13)
3-4	51 (15.8)
4-5	106 (32.8)
5	22 (6.8)

XSL•FO
RenderX

After carefully verifying the app names and company names of 323 valid sports and health apps, we found that among the 210 valid Android apps and 113 valid Apple iOS apps, 37 apps belonged to the same company and shared identical privacy policies. Therefore, we only need to count the privacy policies of duplicate apps from 1 operating system's app store. Since accessing privacy policies on Apple devices is more user-friendly than on Android devices, the privacy policies for these 37 duplicate apps were collected from the Apple App Store and included in the total count of apps collected from the Apple App Store. We subsequently removed these duplicate apps from the Android App Store, resulting in a final count of 286 apps. The methods and procedures for collecting, selecting, and excluding sample apps are presented in Figure 1. By entering the introduction pages of 286 sports and health apps one by one, we downloaded the privacy policy texts. After that, we thoroughly and carefully read the privacy policy documents of these apps from August 10 to August 20, 2025.

This study examined the privacy policies of 286 sports and health apps and assessed whether these privacy policies comply with the PIPL and the PI Guidelines in the Chinese Mainland. The names, compliance indicators, and scores of sports and health apps are presented in Multimedia Appendix 3. Our findings reveal a complex landscape of privacy policy compliance among sports and health apps. Among 286 sports and health apps examined, we observed a spectrum ranging from commendably high privacy compliance to moderate compliance and to a surprisingly poor compliance rate.

Figure 1. Flow diagram of the privacy policy collection and identification process.

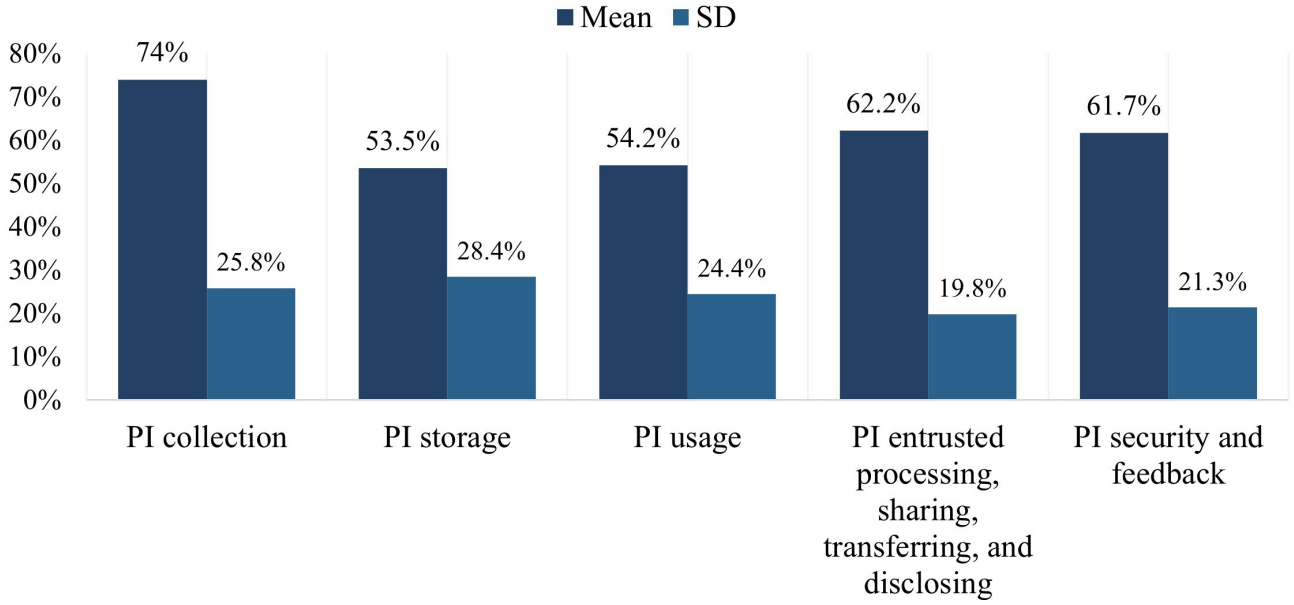


XSL•FO
RenderX

Compliance Evaluation

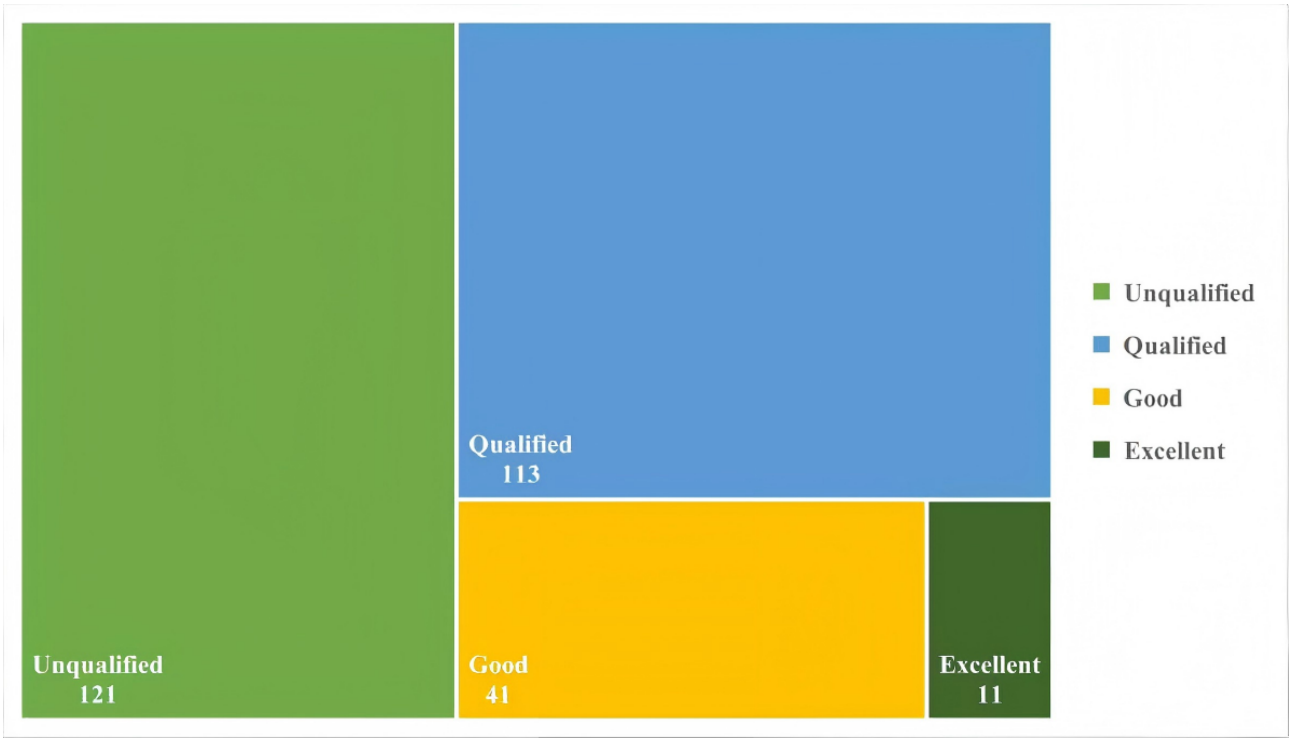
The privacy policy compliance of 286 sports and health apps exhibits significant variation overall (Multimedia Appendix 3). The privacy policy compliance rate for level 1 evaluation indicators across these apps exhibits a tiered pattern. Among level 1 evaluation indicators, PI collection had the highest privacy policy compliance rate (mean 74%, SD 25.8%), while PI storage had the lowest compliance rate (mean 53.5%, SD 28.4%). PI usage also scored low in compliance (mean 54.2%, SD 24.4%). PI security and feedback (mean 61.7%, SD 21.3%) and PI entrusted processing, sharing, transferring, and disclosing (mean 62.2%, SD 19.8%) scored relatively higher. The overall privacy policy compliance status of these level 1 evaluation indicators is illustrated in Figure 2.

Figure 2. Compliance evaluation results of level 1 evaluation indicators of the apps' privacy policies. PI: personal information.



In general, after categorizing the privacy policy compliance scores of 286 sports and health apps into 4 levels (excellent, good, qualified, and unqualified), it was found that 3.8% (n=11) of apps achieved excellent compliance, 41 apps scored good, 113 apps scored qualified, while nearly half (n=121, 42.3%) scored unqualified. Compliance evaluation results on the overall compliance level of the apps' privacy policies are illustrated in Figure 3.

Figure 3. Compliance evaluation results on the overall compliance level of the apps' privacy policies.



XSL•FO
RenderX

From the perspective of the information life cycle [58], the privacy policy compliance of sports and health apps exhibits significant diversity in level 2 evaluation indicators for the information life cycle. This reflects significant variations in the implementation of privacy policy compliance among sports and health apps, with some rules being poorly enforced. The level 2 evaluation indicators' scores of sports and health apps' privacy policy compliance are reflected in Table 5.

XSL•FO
RenderX

Table . Level 2 evaluation indicators of privacy policy compliance score.

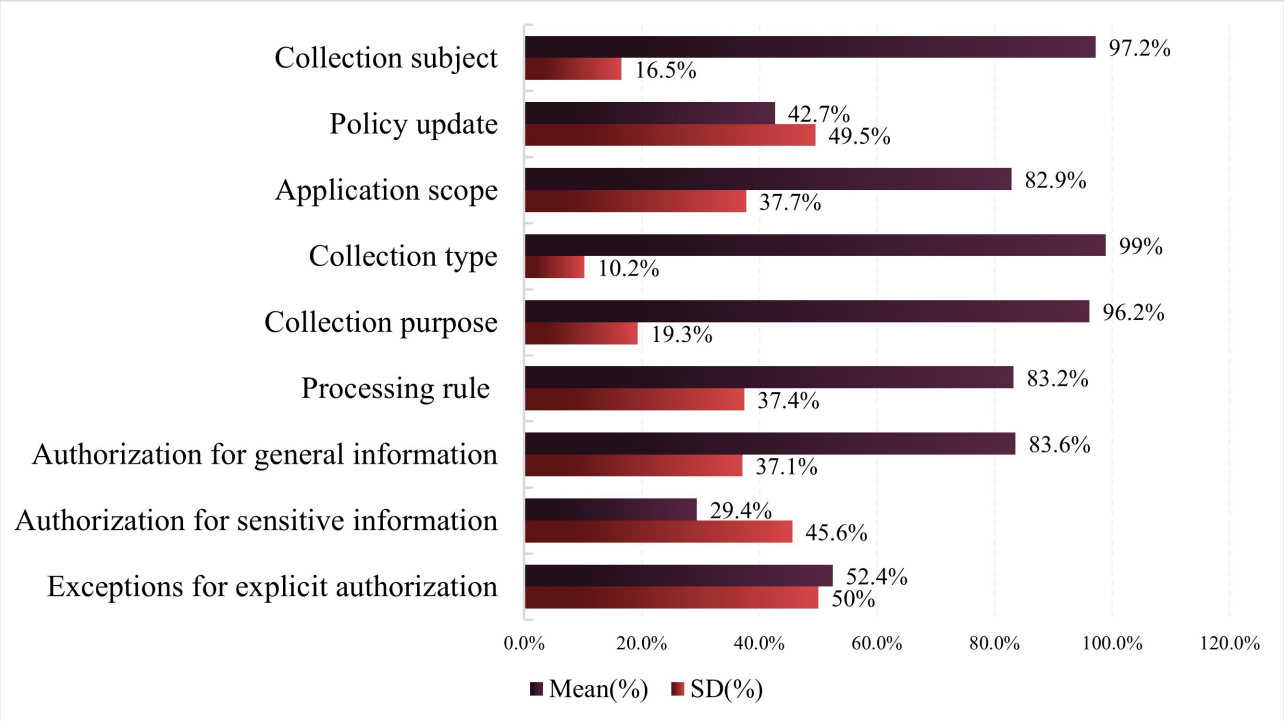
Level 1 and level 2 indicator	Mean (SD)
PI[a] collection	74% (25.8%)
Collection subject	97.2% (16.5%)
Policy update	42.7% (49.5%)
Application scope	82.9% (37.7%)
Collection type	99% (10.2%)
Collection purpose	96.2% (19.3%)
Processing rule	83.2% (37.4%)
Authorization for general information	83.6% (37.1%)
Authorization for sensitive information	29.4% (45.6%)
Exceptions for explicit authorization	52.4% (50%)
PI storage	53.5% (28.4%)
Storage time	69.6% (46.1%)
Storage place	78% (41.5%)
PI deidentification	52.4% (50%)
Sensitive information storage	14% (34.7%)
PI usage	54.2% (24.4%)
PI access control	67.5% (46.9%)
De-identified display and use	34.6% (47.7%)
Purpose restrictions for PI usage	77.6% (41.8%)
Authorization for the usage purpose change	53.5% (50%)
constraints of automatic decision-making	9.4% (29.3%)
Query right	80.1% (40%)
Correction right	70.6% (45.6%)
Deletion right	79.7% (40.3%)
Copy right	46.5% (50%)
Account cancelation right	61.4% (48.8%)
Withdraw or change authorization	75.2% (43.3%)
Asking response right	45.1% (49.8%)
Deceased user rule	5.2% (22.3%)
Complaint mechanism	51.7% (50.1%)
PI entrusted processing, sharing, transferring, and disclosing	62.2% (19.8%)
Requirements for entrusted controller processing PI	52.4% (50%)
Reasons for sharing, transferring, and disclosing PI	91.6% (27.8%)
Security measures of sharing, transferring, and disclosing PI	63.6% (48.2%)
Special circumstances without consent	65.4% (47.7%)
Cross-border transmission requirements	37.8% (48.6%)
PI security and feedback	61.7% (21.3%)
Security incident response mechanism	68.5% (46.5%)
Security event notification mechanism	55.6% (49.8%)
Feedback channel	93.4% (24.9%)
Feedback period	55.6% (49.8%)
External dispute resolution mechanisms	35.3% (47.9%)

XSL•FO
RenderX

[a]PI: personal information.

For level 1 evaluation metric PI collection, PI collection demonstrated the highest level of privacy policy compliance (mean 74%, SD 25.8%) (Figure 4). This indicates that sports and health apps generally prioritize compliance in PI collection, informing users about the methods and purposes of personal information gathering, thereby fundamentally ensuring users' right to know regarding the collection of their PI. This promotes users' understanding of how their PI is used and enhances the transparency, integrity, and security of personal data usage within sports and health apps. The level 1 evaluation indicators PI collection showed significant differences in privacy policy compliance scores across its 9 level 2 evaluation indicators. Collection type demonstrated the highest privacy policy compliance (mean 99%, SD 10.2%), indicating that sports and health apps place strong emphasis on disclosing the types of PI collected to users, ensuring users are aware that their data, particularly sensitive data, have been gathered. Sports and health apps also exhibit high privacy policy compliance rates for collection subject (mean 97.2%, SD 16.5%) and collection purpose (mean 96.2%, SD 19.3%). This indicates that the vast majority of such apps clearly state the name of the data controller and the purpose of data collection within their privacy policies. Informing users about the collection subject and collection purpose enables them to understand who collects their PI and for what processing activities it is used, thereby assisting users in future legal actions regarding PI breaches. The compliance rate for application scope (mean 82.9%, SD 37.7%), processing rule (mean 83.2%, SD 37.4%), and authorization for general information (mean 83.6%, SD 37.1%) all exceeded 80%. Policy update (mean 42.7%, SD 49.5%) and exceptions for explicit authorization (mean 52.4%, SD 50%) showed lower compliance rates, approaching half. The privacy policy compliance rate for authorization for sensitive information was the lowest (mean 29.4%, SD 45.6%), indicating that the vast majority of sports and health apps lack effective provisions in their privacy policies regarding consent mechanisms for collecting sensitive PI. Since sports and health apps inevitably collect users' sensitive PI, the absence of consent mechanisms for such data collection leaves personal sensitive information highly vulnerable and susceptible to leakage.

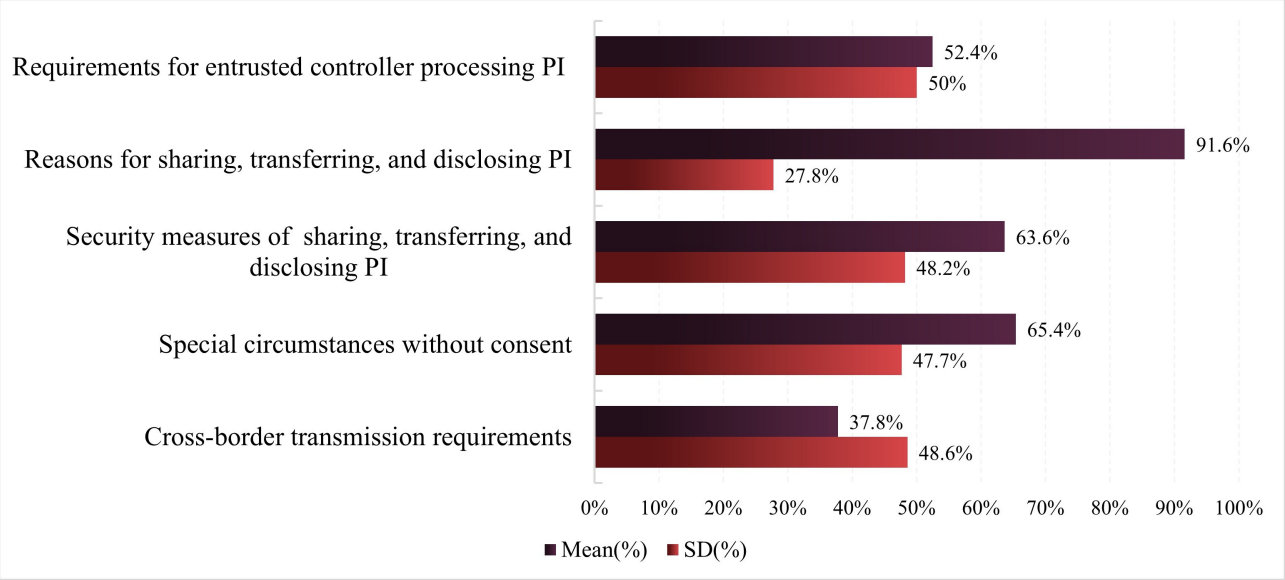
Figure 4. Privacy policy compliance evaluation results of personal information collection.



The level 1 evaluation indicator PI entrusted processing, sharing, transferring, and disclosing also demonstrated high privacy policy compliance (mean 62.2%, SD 19.8%) (Figure 5). Except for cross-border transmission requirements (mean 37.8%, SD 48.6%), the privacy policy compliance rate for level 2 evaluation indicators under level 1 PI entrusted processing, sharing, transferring, and disclosing exceeded 50%. The privacy policy compliance rate for reasons for sharing, transferring, and disclosing PI was the highest (mean 91.6%, SD 27.8%), indicating that sports and health apps highly prioritize explaining these reasons to users. This assists in safeguarding users' right to informed consent during the sharing, transferring, and disclosing of PI. The compliance rates of privacy policies for sports and health apps in the requirements for entrusted controller processing PI (mean 52.4%, SD 50%), security measures for sharing, transferring, and disclosing PI (mean 63.6%, SD 48.2%), and special circumstances without consent (mean 65.4%, SD 47.7%) all exceed one-half.

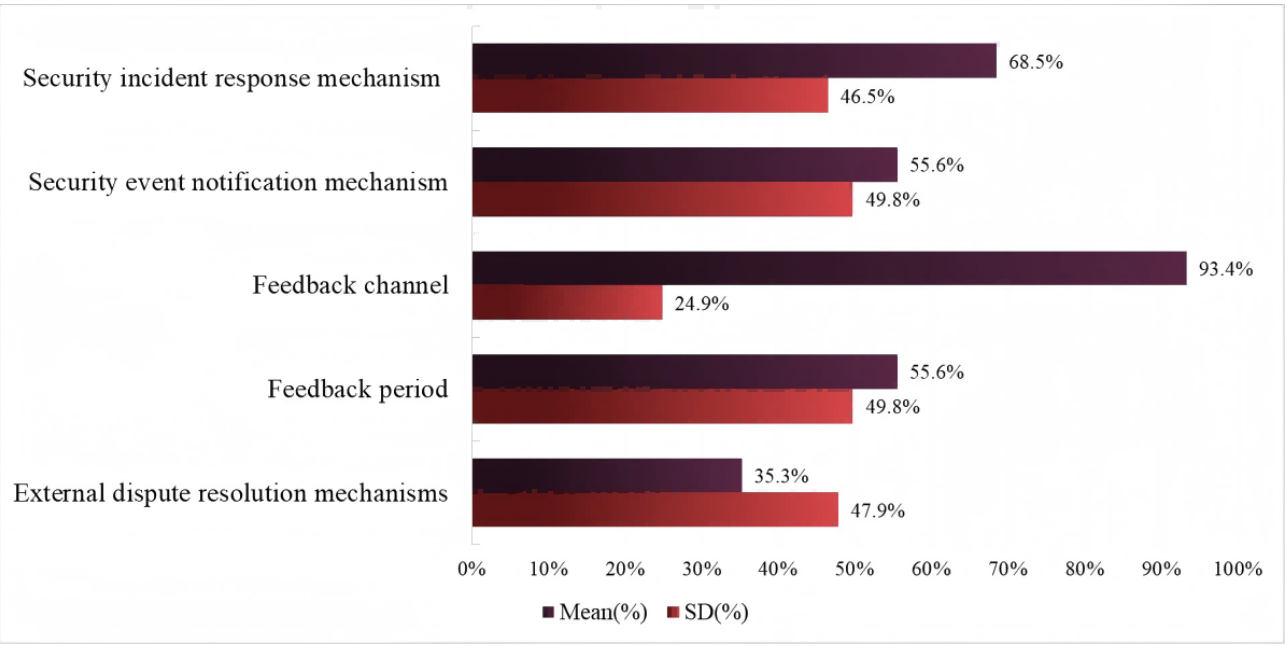
XSL•FO
RenderX

Figure 5. Privacy policy compliance evaluation results of personal information entrusted processing, sharing, transferring, and disclosing. PI: personal information.



Regarding level 1 evaluation indicators, PI security and feedback, the privacy policy compliance rate of sports and health apps in PI security and feedback exceeds half (mean 61.7%, SD 21.3%) (Figure 6). The privacy policy compliance score for the level 2 evaluation indicator feedback channel is the highest (mean 93.4%, SD 24.9%), reflecting that sports and health apps pay great attention to providing users with effective feedback channels in case of disputes regarding the collection, use, and processing of PI. In terms of privacy policy compliance, more than half of the sports and health apps meet the level 2 evaluation indicators: security incident response mechanism (mean 68.5%, SD 46.5%), security event notification mechanism (mean 55.6%, SD 49.8%), and feedback period (mean 55.6%, SD 49.85%). This indicates that over 50% of sports and health apps have established PI security protection mechanisms and security incident reporting procedures, ensuring the safety of users' PI and their right to be informed in the event of information leakage. The compliance of external dispute resolution mechanisms in privacy policies is very poor, with one-third of sports and health apps not having established external resolution mechanisms for PI disputes, which means they do not clearly inform users that they can sue in court if a dispute occurs.

Figure 6. Privacy policy compliance evaluation results of personal information security and feedback.

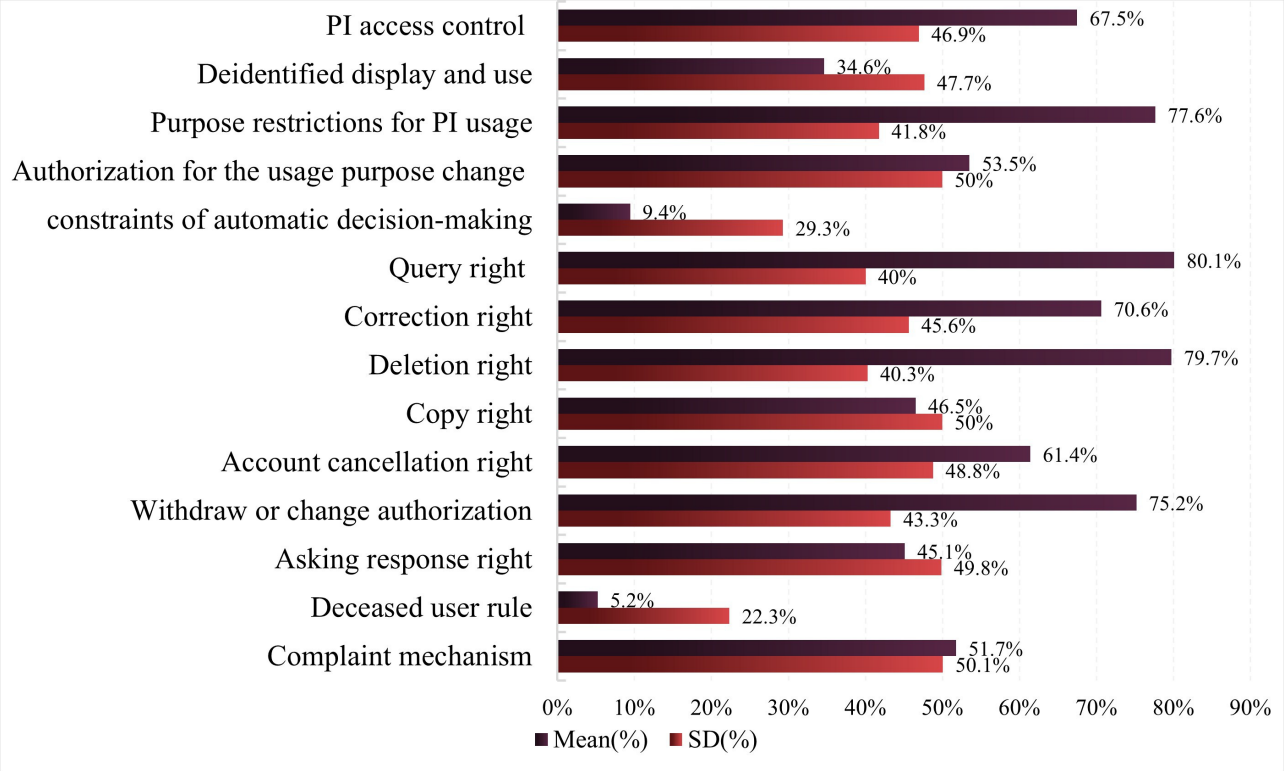


The level 1 evaluation indicator PI usage is relatively low (mean 54.2%, SD 24.4%) for privacy policy compliance. This indicates that nearly half of the sports health apps do not strictly enforce the protective regulations for PI usage (Figure 7). Sports and health apps generally provide good protection for users' PI rights, with the compliance rate for the query right being the highest (mean 80.1%, SD 40.0%). Compliance rates for the correction right (mean 70.6%, SD 45.6%), deletion right (mean

XSL•FO
RenderX

79.7%, SD 40.3%), account cancelation right (mean 61.4%, SD 48.8%), and withdraw or change authorization (mean 75.2%, SD 43.3%) also exceed half, indicating that the vast majority of sports and health apps can effectively protect users' basic PI rights. However, in terms of user personal information rights, the privacy compliance rates for copyright (mean 46.5%, SD 50%) and asking response right (mean 45.1%, SD 49.8%) are relatively poor, especially for the deceased user rule, which has the worst privacy compliance (mean 5.2%, SD 22.3%). On the other hand, the privacy compliance rates for level 2 evaluation indicators such as PI access control (mean 67.5%, SD 46.9%) and purpose restrictions for PI usage (mean 77.6%, SD 41.8%) are relatively good. The privacy policy compliance rate for authorization for the usage purpose change (mean 53.5%, SD 50%) and the complaint mechanism (mean 51.7%, SD 50.1%) have both exceeded half. Only one-third of sports and health apps meet privacy policy compliance requirements in deidentified display and use (mean 34.6%, SD 47.7%). It is particularly noteworthy that the compliance rate of privacy policies regarding constraints of automatic decision-making is very low (mean 9.4%, SD 29.3%), indicating that the vast majority of sports and health apps do not have restrictions on the automatic decision-making mechanism. This is detrimental to users' ability to decide how their PI is used and processed, leading to negative impacts on the security of PI and users' peace of mind [80].

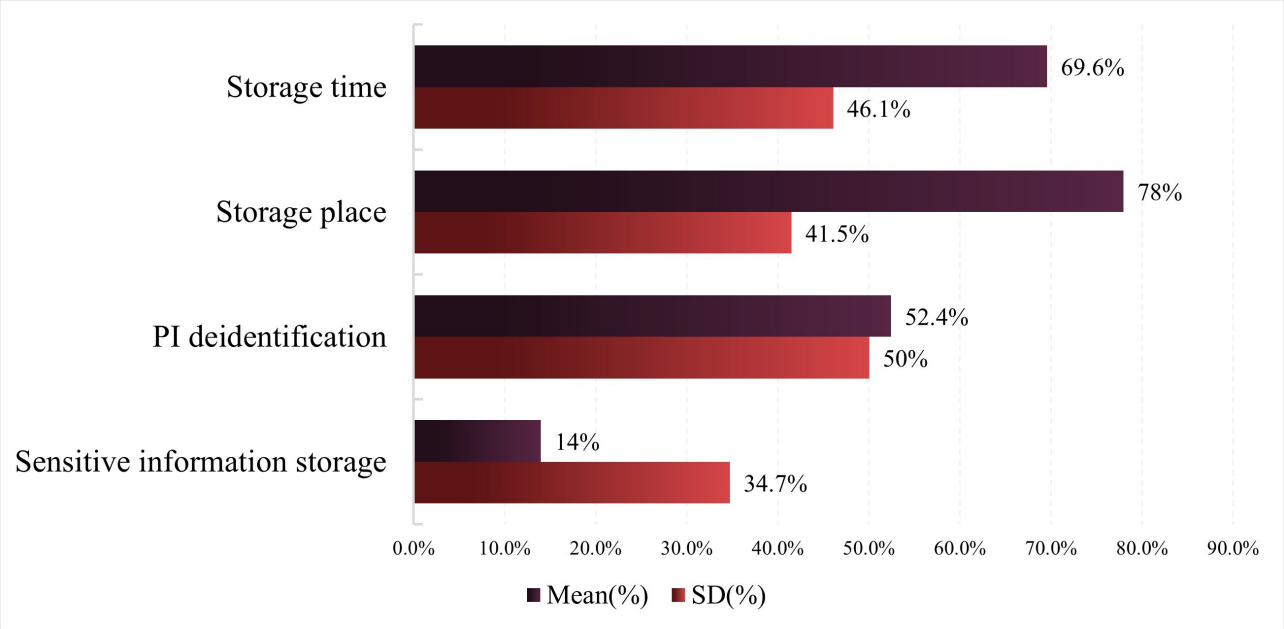
Figure 7. Privacy policy compliance evaluation results of personal information usage. PI: personal information.



The compliance rate of privacy policies for sports and health apps regarding the level 1 evaluation indicator PI storage is the lowest (mean 53.5%, SD 28.4%), with about half of the apps failing to meet the compliance requirements for PI storage (Figure 8). The compliance rate of privacy policies for sports and health apps regarding the level 2 evaluation indicator storage place is the highest (mean 78%, SD 41.5%), while the compliance extent for storage time (mean 69.6%, SD 46.1%) and PI deidentification (mean 52.4%, SD 50%) also performs relatively well. However, the compliance rate of privacy policies for sensitive information storage is the worst among sports and health apps, indicating that the vast majority of apps do not have specific provisions and settings for the storage of users' sensitive PI, which is detrimental to the protection of users' sensitive PI and data security.

XSL•FO
RenderX

Figure 8. Privacy policy compliance evaluation results of personal information storage. PI: personal information.



Discussion

Principal Findings

We developed a privacy policy evaluation scale by integrating scenario-based contextual analysis theory and information life cycle theory, grounded in the provisions of the PIPL and PI Guidelines, to assess privacy policies of mobile sports and health apps. Our defined context encompasses scenarios where users use these apps to record physical activity data, exercise routes, and fitness instructions, all directly related to physical exercise.

This study's findings demonstrate that 286 evaluated apps in the Chinese Mainland are failing to meet core PI protection requirements, with critical gaps in the overall compliance of the 286 apps, policy update, sensitive PI protection, key user rights, PI transferring, and PI security management.

First, our analysis indicates that the overall compliance level of privacy policies among 286 sports and health apps is not high. The compliance levels varied (mean 61%, SD 24.3%), with a minority of apps demonstrating strong compliance while others showed lower compliance. Sports and health apps with low compliance primarily violate the PIPL and PI Guidelines, which share common principles with many other jurisdictions. Furthermore, low privacy policy compliance of apps can have several direct and indirect negative consequences for users, including security risks, loss of control, and intrusive marketing. This indicates the need to strengthen PI regulatory oversight for sports and health apps and establish standardized, actionable practices.

Second, data analysis suggests that PI collection has the highest level of compliance. According to Article 14 of the PIPL, if the collection purpose, method, or types are changed, a separate consent shall be obtained from users. However, data analysis indicates that compliance rates regarding privacy policy updates are unsatisfactory, with only 122 out of 286 apps adhering to this requirement. Indeed, many app developers operate under the assumption that continued use equates to user consent, rather than proactively seeking explicit approval. This lack of compliance undermines both user dignity and security, increasing the risk of data misuse and breaches. To address this, we argue that apps should instead secure explicit consent through clear pop-up notifications, much like during the initial sign-up process. Users should be required to actively confirm their agreement; otherwise, access to the app's services should be limited or denied.

Third, we found that the privacy policies of sports and health apps exhibit significant shortcomings in protecting sensitive PI. It should be noted that the PIPL (especially Chapter II, Section 2) and the Guidelines (especially Article 5.4) provide special protection for the collection of sensitive PI. Substantively, Article 28 of the PIPL specifies the substantive conditions for processing sensitive PI, namely, PI processors can process sensitive PI only when there is a specific purpose and when it is of necessity, under the circumstances where strict protective measures are taken. Procedurally, Articles 29 and 30 of the PIPL and Article 5.4b of the PI Guidelines require that a processor processing sensitive PI shall notify the individual of the necessity of processing their sensitive PI and the impact it has on their rights and interests," and that sensitive PI processors must obtain informed, voluntary, specific, clear, unequivocal, and separate consent from PI subjects. The substantive and procedural requirements have also been highlighted by the Notice of the Ministry of Industry and Information Technology on Further Enhancing the Service Capacity of Mobile Internet Applications [81]. However, our data analysis shows that the compliance rate of sensitive PI protection (mean 29.4%, SD 45.6%) is significantly lower than that of general PI protection (mean 83.6%, SD 37.1%). This disparity not only indicates how inadequate the present security measures are for protecting sensitive PI, but it also continuously lowers users' awareness of security when providing sensitive PI. Additionally, it is also important to note that although a total of 40 apps referenced sensitive PI storage, none comply with the storage requirements

XSL•FO
RenderX

of the PI Guidelines, which mandate encryption (Article 6.3.a) and, in principle, the elimination of original personal biometric data (Article 6.3.c), including samples and photographs. In the 40 apps, one just lists the types of sensitive PI, such as personal identity information and biometric information. The other apps fail to explicitly state the protection measures for sensitive PI; instead, their statements imply that adequate protection measures have been taken and that they would do their best to protect sensitive PI. This reveals a serious security concern, because if sensitive PI is leaked, tampered with, or used illegally, it may endanger the user's personal or property safety.

Fourth, our analysis indicates that comprehensive compliance with the protection of users' PI rights remains weak. In terms of user rights, although most apps effectively list a series of user rights, including query rights, correction rights, deletion rights, withdraw or change authorization rights, and account cancellation rights, less than half of the apps give users the right to obtain a copy of PI and ask for a response. Notably, only 27 of the 286 apps provide users with the right to refuse using automated decision-making mechanisms. Currently, personalized recommendation technologies and artificial intelligence technologies are advancing rapidly. Automated decision-making mechanisms based on algorithms have been widely adopted in the operation of sports and health apps. The vast majority of these apps do not grant users the right to opt out of automated decision-making processes. This results in the infringement of users' autonomy over the collection, use, and processing of their personal information, violating the principles of legitimacy and transparency enshrined in PIPL. Consequently, users may be misled in their app usage, experience improper collection of their information, and face frequent disruption from app notifications. Fewer apps have provided deceased user rules, that is, unless otherwise arranged by the deceased before their death, their close relatives have the rights to access, copy, correct, delete, and other rights related to the deceased's relevant PI. This will trigger disputes over the usage and inheritance of deceased individuals' digital assets, hindering the normal operation and development of sports and health apps.

Fifth, our findings reveal that privacy policies for sports and health apps require further refinement regarding regulations on sharing, transferring, and disclosing PI. While most apps state the reasons for sharing, transferring, and disclosing PI, significantly fewer of them will take technical measures to ensure information security and obtain the user's separate consent. Notably, even fewer apps' privacy policies specify that when PI is shared or transferred to overseas processors, the user should be separately informed of the purpose of data export and the receiver, and the user's authorization should be separately obtained. This is obviously lower than the standards stipulated in Article 23 of the PIPL and Article 9 of the PI Guidelines, which may damage the users' ability to monitor the whole PI processing process, their security awareness of transferring PI, and their trust in apps, and the principle of informed consent.

Finally, our analysis reveals that sports and health apps do not provide users with clear response times for feedback. User data collection occurs in real time, and this frequent gathering inevitably leads to issues or disputes during usage. At such times, a timely and effective feedback and dispute resolution mechanism becomes crucial for resolving problems, maintaining user retention, and helping apps build a positive reputation. However, many sports and health apps merely provide feedback channels without establishing timelines for addressing concerns. This lack of accountability discourages apps from proactively and effectively resolving feedback issues and disputes. Additionally, many sports and health apps lack robust PI security safeguards and reporting mechanisms, making personal data vulnerable to infringement and posing serious risks of data breaches. This will lead to a decline in users' trust in the app's privacy protection, making users more cautious about providing personal information, especially sensitive personal information, and may lead to users uninstalling the app.

Recommendations

The study's findings collectively demonstrate that 286 evaluated apps are failing to meet core PI protection requirements. Critical gaps in PI protection are primarily attributed to 3 factors. First, there is a deficiency of awareness among users regarding PI protection, reducing incentives for app developers to prioritize privacy. Second, existing PI guidelines lack legally binding force, resulting in inconsistent adherence and voluntary compliance. Third, insufficient regulatory oversight and weak enforcement mechanisms fail to deter noncompliant practices. These factors collectively undermine the effectiveness of current PI protection frameworks, highlighting the need for targeted action from 3 stakeholders: users, regulators, and legislators. We argue that only through coordinated action can the app ecosystem close the compliance gaps identified in this study, reduce PI protection risks, and restore user trust in digital services.

First, to strengthen app privacy compliance, it is essential to enhance users' PI rights awareness [82]. We argue that interventions can be scenario-embedded and accessible. For example, app developers are encouraged to integrate snapshots of PI rights into app onboarding when users consent to privacy policies, paired with visual resources that translate legal jargon into simple guidance [83]. This ensures users learn rights when they interact with PI-related features, not just through disconnected policy texts. Additionally, official complaint channels should be easily accessible to hold apps accountable for noncompliance, which can drive industry-wide improvement [84].

Second, the regulatory mechanism for sports and health apps should be standardized. Under the current legal framework, the regulatory mechanism for apps can be characterized as self-regulation by app operators, with government oversight as a supplementary measure. Moreover, in terms of the institutional framework in the Chinese Mainland, regulators are dispersed across many different government departments, resulting in a lack of enthusiasm for law enforcement and of technical capacity for supervision. As a result, the regulation of apps in practice is inadequate. In this context, we recommend creating a unique data regulator to assess and oversee the privacy protection of apps and raising the industry's entry barrier through the issuance of compliance identification or certifications [64]. This regulator can strengthen enforcement of low-compliance areas, such as

XSL•FO
RenderX

sensitive PI storage and policy updates, through regular audits and penalties for noncompliance [63].

Finally, laws and regulations for the protection of personal health information should be improved. The Civil Code [85], the Cybersecurity Law of China [86], the Law of the People's Republic of China on Basic Medical and Health Care and the Promotion of Health [87], and the PIPL formulate the legal framework for protecting personal health information. While these laws are comprehensive, the majority of them contain oath and principle clauses that are not applicable in specific cases. Although many of these principles and oath clauses are further crafted in the PI Guidelines, these guidelines do not have the force of law; instead, they are merely recommended standards and are not mandatory by nature. In this regard, we suggest that the legal effect of adopting the PI Guidelines should be clarified in laws and regulations. In this way, the app operator should specify in the first part of the privacy policy whether the PI Guidelines are adopted, and once adopted, they will be legally binding [66].

Contributions

This study holds significant implications for both policymakers and scholars. First, based on the principal findings—particularly that the privacy protection requirements established by the PIPL demonstrate a higher level of compliance compared with the PI Guidelines—we propose that legislation should strengthen the binding force of the PI Guidelines, thereby providing guidance for legislative reform. Second, the current low compliance with privacy policies may be linked to inadequate regulatory enforcement. In response to the fragmented regulatory oversight in practice, we recommend the establishment of an independent regulator to enhance the effectiveness of privacy protection enforcement. Finally, this study addresses a gap in the existing literature. While existing studies have explored the relationship between health apps and privacy protection in Australia [4,88], between mHealth apps and privacy protection [89-93], and between health code apps and privacy protection in China [94,95], the legal compliance of sports and health apps with the PIPL and PI Guidelines, as well as China's legislative framework for PI protection, was not thoroughly examined in these studies. This study fills this gap by examining the privacy policies of 286 sports and health apps across 37 evaluation dimensions and assessing whether these privacy policies comply with the PIPL and PI Guidelines.

Limitations

However, this study also has limitations. Although we developed an indicator scale to assess the compliance level of privacy policies of the 286 sports and health apps, this evaluation process does not concern the actual implementation of different technical measures, resulting in the omission of evaluation criteria and the lack of targeted recommendations for improvement from a technical point of view [96]. For instance, an app may explicitly state in its policy that it encrypts sensitive PI during storage but may not actually implement consistent encryption protocols in its backend systems. We hope that future research can treat this study as a starting point to further analyze the processing of PI and the design and implementation of PI protection measures in the information system, by using technical methods, such as network traffic analysis or third-party data flow auditing, and adopting the research method of interviewing relevant personnel including product managers, research and development engineers, PI protection officers, legal personnel, system architects, security administrators, operation and maintenance personnel, human resources personnel, system users, and so on.

Conclusions

The extensive usage of sports and health apps is significantly enhancing the social and public health landscape in the Chinese Mainland. Nonetheless, the security threats to PI protection arising from privacy policy compliance have not received adequate attention. This paper's assessment of the overall compliance level of 286 sports and health apps and the individual evaluation of 37 privacy policies reveals a complex picture of PI protection. Although some apps establish commendable compliance policies, there are some shortcomings. These shortcomings not only pose security threats to app users but may also hinder operators in optimizing and developing app functionalities. In light of this, this paper puts forward suggestions for improvement from the perspectives of user awareness, regulatory enforcement, and legislative reforms.

Acknowledgments

We thank Jing Guo, Xiaolan Ji, and Jiamei Xie for assistance with data collection.

Funding

This research is funded by the National Social Science Fund Project of China (22CFX073 and 24CFX023) and China National Intellectual Property Administration Soft Science Research Project (SS25-A-05) .

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

RG performed the conceptualization, methodology, investigation, data curation, formal analysis, writing–original draft, and writing–review & editing. FC made the conceptualization, methodology, data curation, formal analysis, writing–original draft, and writing–review & editing.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Excluded samples.
[XLSX File, 33 KB - mhealth_v14i1e73651_app1.xlsx]

Multimedia Appendix 2
The characteristics of the valid sports and health apps.
[XLSX File, 24 KB - mhealth_v14i1e73651_app2.xlsx]

Multimedia Appendix 3
Names, compliance indicators, and scores of sports and health apps.
[XLSX File, 75 KB - mhealth_v14i1e73651_app3.xlsx]

Checklist 1
PRISMA-LSR checklist.
[PDF File, 105 KB - mhealth_v14i1e73651_app4.pdf]

References

1. Tuitert I, Marinus JD, Dalenberg JR, van 't Veer JT. Digital health technology use across socioeconomic groups prior to and during the COVID-19 pandemic: panel study. JMIR Public Health Surveill 2024 Sep 13;10:e55384. [doi: 10.2196/55384] [Medline: 39269755]
2. Mosa ASM, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012 Jul 10;12:67. [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
3. Weinstein RS, Lopez AM, Joseph BA, et al. Telemedicine, telehealth, and mobile health applications that work: opportunities and barriers. Am J Med 2014 Mar;127(3):183-187. [doi: 10.1016/j.amjmed.2013.09.032] [Medline: 24384059]
4. Tangari G, Ikram M, Ijaz K, Kaafar MA, Berkovsky S. Mobile health and privacy: cross sectional study. BMJ 2021 Jun 16;373:n1248. [doi: 10.1136/bmj.n1248] [Medline: 34135009]
5. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Privacy and security in mobile health apps: a review and recommendations. J Med Syst 2015 Jan;39(1):181. [doi: 10.1007/s10916-014-0181-3] [Medline: 25486895]
6. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295. [doi: 10.2196/jmir.7740] [Medline: 28838887]
7. Maaß L, Freye M, Pan CC, Dassow HH, Niess J, Jahnel T. The definitions of health apps and medical apps from the perspective of public health and law: qualitative analysis of an interdisciplinary literature overview. JMIR Mhealth Uhealth 2022 Oct 31;10(10):e37980. [doi: 10.2196/37980] [Medline: 36315221]
8. Gamache R, Kharrazi H, Weiner JP. Public and population health informatics: the bridging of big data to benefit communities. Yearb Med Inform 2018 Aug;27(1):199-206. [doi: 10.1055/s-0038-1667081] [Medline: 30157524]
9. Betthauser LM, Stearns-Yoder KA, McGarity S, Smith V, Place S, Brenner LA. Mobile app for mental health monitoring and clinical outreach in veterans: mixed methods feasibility and acceptability study. J Med Internet Res 2020 Aug 11;22(8):e15506. [doi: 10.2196/15506] [Medline: 32779572]
10. Hahn SL, Kaciroti N, Eisenberg D, Weeks HM, Bauer KW, Sonneville KR. Introducing dietary self-monitoring to undergraduate women via a calorie counting app has no effect on mental health or health behaviors: results from a randomized controlled trial. J Acad Nutr Diet 2021 Dec;121(12):2377-2388. [doi: 10.1016/j.jand.2021.06.311] [Medline: 34427188]
11. Chew HSJ, Koh WL, Ng J, Tan KK. Sustainability of weight loss through smartphone apps: systematic review and meta-analysis on anthropometric, metabolic, and dietary outcomes. J Med Internet Res 2022 Sep 21;24(9):e40141. [doi: 10.2196/40141] [Medline: 36129739]
12. Antoun J, Itani H, Alarab N, Elsehmawy A. The effectiveness of combining nonmobile interventions with the use of smartphone apps with various features for weight loss: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022 Apr 8;10(4):e35479. [doi: 10.2196/35479] [Medline: 35394443]
13. Sakane N, Suganuma A, Domichi M, et al. The effect of a mHealth app (KENPO-app) for specific health guidance on weight changes in adults with obesity and hypertension: pilot randomized controlled trial. JMIR Mhealth Uhealth 2023 Apr 12;11:e43236. [doi: 10.2196/43236] [Medline: 37043287]
14. Lugones-Sanchez C, Sanchez-Calavera MA, Repiso-Gento I, et al. Effectiveness of an mHealth intervention combining a smartphone app and smart band on body composition in an overweight and obese population: randomized controlled trial (EVIDENT 3 Study). JMIR Mhealth Uhealth 2020 Nov 26;8(11):e21771. [doi: 10.2196/21771] [Medline: 33242020]

XSL•FO
RenderX

15. Mano R. Mobile health apps and health management behaviors: cost-benefit modeling analysis. JMIR Hum Factors 2021 Apr 22;8(2):e21251. [doi: 10.2196/21251] [Medline: 33885372]
16. Cho J. The impact of post-adoption beliefs on the continued use of health apps. Int J Med Inform 2016 Mar;87:75-83. [doi: 10.1016/j.ijmedinf.2015.12.016] [Medline: 26806714]
17. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3:14. [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]
18. Yerrakalva D, Yerrakalva D, Hajna S, Griffin S. Effects of mobile health app interventions on sedentary time, physical activity, and fitness in older adults: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e14343. [doi: 10.2196/14343] [Medline: 31778121]
19. Modave F, Bian J, Leavitt T, Bromwell J, Harris Iii C, Vincent H. Low quality of free coaching apps with respect to the American College of Sports Medicine Guidelines: a review of current mobile apps. JMIR Mhealth Uhealth 2015 Jul 24;3(3):e77. [doi: 10.2196/mhealth.4669] [Medline: 26209109]
20. Tighe J, Shand F, Ridani R, Mackinnon A, De La Mata N, Christensen H. Ibobbly mobile health intervention for suicide prevention in Australian Indigenous youth: a pilot randomised controlled trial. BMJ Open 2017 Jan 27;7(1):e013518. [doi: 10.1136/bmjopen-2016-013518] [Medline: 28132007]
21. Grundy Q. A review of the quality and impact of mobile health apps. Annu Rev Public Health 2022 Apr 5;43:117-134. [doi: 10.1146/annurev-publhealth-052020-103738] [Medline: 34910582]
22. Ni Z, Wang Y, Qian Y. Privacy policy compliance of chronic disease management apps in China: scale development and content evaluation. JMIR Mhealth Uhealth 2021 Jan 28;9(1):e23409. [doi: 10.2196/23409] [Medline: 33507159]
23. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3:4. [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
24. Lupton D, Jutel A. “It’s like having a physician in your pocket!” A critical analysis of self-diagnosis smartphone apps. Soc Sci Med 2015 May;133:128-135. [doi: 10.1016/j.socscimed.2015.04.004] [Medline: 25864149]
25. Flaherty JL. Digital diagnosis: privacy and the regulation of mobile phone health applications. Am J Law Med 2014;40(4):416-441. [Medline: 27530051]
26. Herriger C, Merlo O, Eisingerich AB, Arigayota AR. Context-contingent privacy concerns and exploration of the privacy paradox in the age of AI, augmented reality, big data, and the internet of things: systematic review. J Med Internet Res 2025 May 14;27:e71951. [doi: 10.2196/71951] [Medline: 40367513]
27. Tovino SA. Privacy and security issues with mobile health research applications. J Law Med Ethics 2020 Mar;48(1_suppl):154-158. [doi: 10.1177/1073110520917041] [Medline: 32342741]
28. Jiang J, Zheng Z. Medical information protection in internet hospital apps in China: scale development and content analysis. JMIR Mhealth Uhealth 2024 Jun 21;12:e55061. [doi: 10.2196/55061] [Medline: 38904994]
29. Lin X, Wu X, Zhu Z, Chen D, Li H, Lin R. Quality and privacy policy compliance of mental health care apps in China: cross-sectional evaluation study. J Med Internet Res 2025 Jul 3;27:e66762. [doi: 10.2196/66762] [Medline: 40607739]
30. McGee MK. FTC fines fertility app vendor, bars it from data sharing. Gov Info Security. URL: https://www.govinfosecurity.com/ftc-fines-fertility-app-vendor-bars-from-data-sharing-a-22100 [accessed 2024-09-13]
31. Out of control-how consumers are exploited by the online advertising industry. : Forbrukerrådet; 2020 URL: https://www.forbrukerradet.no/rapporter/out-of-control/ [accessed 2024-09-13]
32. The National Cyberspace Administration of China will launch a special campaign to rectify app chaos, and the full process governance will become the norm. Cyberspace Administration of China. URL: https://www.cac.gov.cn/2018-12/28/c_1123919199.htm [accessed 2024-12-11]
33. Paul P, Abdul H. Data protection laws and regulations USA 2024-2025. ICLG. URL: https://iclg.com/practice-areas/data-protection-laws-and-regulations/usa [accessed 2024-11-04]
34. Andrew K, Willy M. Data privacy and protection in the US: a sign of bipartisan progress. iapp. URL: https://iapp.org/news/a/data-privacy-and-protection-in-the-us-a-sign-of-bipartisan-progress/ [accessed 2024-12-23]
35. Mulder T. Health apps, their privacy policies and the GDPR. Eur J Law Technol 2019;10(1) [FREE Full text]
36. Lauristin M. Draft report on the proposal for a regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing directive 2002/58/EC (regulation on privacy and electronic communications). : European Parliament; 2017 URL: https://www.europarl.europa.eu/doceo/document/LIBE-PR-606011_EN.pdf?redirect [accessed 2024-12-23]
37. Consolidated text: regulation (EU) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (general data protection regulation). European Union. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02016R0679-20160504 [accessed 2024-11-04]
38. Bardus M, Al Daccache M, Maalouf N, Al Sarih R, Elhajj IH. Data management and privacy policy of COVID-19 contact-tracing apps: systematic review and content analysis. JMIR Mhealth Uhealth 2022 Jul 12;10(7):e35195. [doi: 10.2196/35195] [Medline: 35709334]
39. Health information privacy. US Department of Health & Human Services. URL: https://www.hhs.gov/hipaa/ [accessed 2024-12-23]

XSL•FO
RenderX

40. Health breach notification rule. Federal Trade Commission. URL: https://www.ftc.gov/legal-library/browse/rules/health-breach-notification-rule [accessed 2025-01-20]
41. Children's online privacy protection rule ("COPPA"). Federal Trade Commission. URL: https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa [accessed 2025-01-20]
42. Yang J. A comprehensive guide to US data protection legislation. Safety internal reference. URL: https://www.secrss.com/articles/11839 [accessed 2025-01-20]
43. General data protection regulation (GDPR). Intersoft Consulting. URL: https://gdpr-info.eu/ [accessed 2025-01-20]
44. Yin K, Zhang G, Deng S, Li H. China's final personal information protection compliance audit rules and faqs. Fangda Partners. URL: https://www.fangdalaw.com/en/content/details32_9931.html [accessed 2026-01-20]
45. Notification regarding the illegal and irregular collection and use of personal information by 129 apps including Keep. Cyberspace Administration of China. URL: https://www.cac.gov.cn/2021-06/11/c_1624994586637626.htm [accessed 2025-01-20]
46. Jiang J, Zheng Z. Personal information protection and privacy policy compliance of health code apps in China: scale development and content analysis. JMIR Mhealth Uhealth 2023 Nov 14;11:e48714. [doi: 10.2196/48714] [Medline: 37990813]
47. Calzada I. Citizens' data privacy in China: the state of the art of the Personal Information Protection Law (PIPL). Smart Cities 2022;5(3):1129-1150. [doi: 10.3390/smartcities5030057]
48. The 55th statistical report on the internet development in China. CNNIC. URL: https://www.100ec.cn/detail--6646318.html [accessed 2025-09-05]
49. Expert interpretation | personal information protection law: establishing a fundamental law for digital society governance and digital economy development. Cyberspace Administration of China. URL: https://www.cac.gov.cn/2021-08/25/c_1631491548474109.htm [accessed 2025-09-05]
50. 2023 National Internet Users' Satisfaction Survey on Cybersecurity: special report on personal information protection and data security. : Wuzhen Digital Civilization Research Institute; 2024 URL: https://www.iscn.org.cn/uploadfile/2024/0829/zhuanti4.pdf [accessed 2026-01-20]
51. Liu Q. On the impact assessment of personal information protection: focusing on Articles 55 and 56 of Personal Information Protection Act. 2022. (5) p. 40-50 URL: https://link.oversea.cnki.net/doi/10.13806/j.cnki.issn1008-7095.2022.05.004 [accessed 2026-01-28]
52. Jet D, Ken D. The comparison between China's PIPL and EU's GDPR: practitioners' perspective. Mondaq. 2021. URL: https://www.mondaq.com/china/data-protection/1122748/the-comparison-between-chinas-pipl-and-eus-gdpr-practitioners-perspective [accessed 2025-03-01]
53. Daniel S. China's PIPL vs the GDPR: a comparison. TeachPrivacy. URL: https://teachprivacy.com/chinas-pipl-vs-gdpr-a-comparison [accessed 2025-03-01]
54. Li W, Chen J. From brussels effect to gravity assists: understanding the evolution of the GDPR-inspired personal information protection law in China. Computer Law & Security Review 2024 Sep;54(9):105994. [doi: 10.1016/j.clsr.2024.105994]
55. Practical Guide to Cybersecurity Standards-Guidelines for the classification and grading of network data [Article in Chinese]. : National Information Security Standardization Technical Committee; 2021 URL: https://www.tc260.org.cn/upload/2021-12-31/1640948142376022576.pdf [accessed 2025-03-01]
56. Information security technology - personal information security engineering guidelines. National Public Service Platform for Standards Information. URL: https://std.samr.gov.cn/gb/search/gbDetailed?id=EB58F4DA9267B2A2E05397BE0A0A7D33 [accessed 2025-03-01]
57. Implementation guidelines for notification and consent in information security technology and personal information processing. National Public Service Platform for Standards Information. URL: https://std.samr.gov.cn/gb/search/gbDetailed?id=FC816D04FFD262EBE05397BE0A0AD5FA [accessed 2025-03-01]
58. Information security technology: guidelines for personal information de-identification. National Public Service Platform for Standards Information. URL: https://std.samr.gov.cn/gb/search/gbDetailed?id=91890A0DA4AB80C6E05397BE0A0A065D [accessed 2025-03-01]
59. Notification regarding apps that infringe on user rights (Batch 1 of 2023, Batch 27 in total). Ministry of Industry and Information Technology of the People's Republic of China. URL: https://wap.miit.gov.cn/jgsj/xgj/gzdt/art/2023/art_f3c61d42b107479abbc0fa115efffc10.html [accessed 2025-03-01]
60. Notification Regarding Apps (SDKS) that Infringe on User Rights (Batch 1 of 2024, Batch 36 in Total). Ministry of Industry and Information Technology of the People's Republic of China. URL: https://wap.miit.gov.cn/jgsj/xgj/gzdt/art/2024/art_a6f9db30756c4464804355afb64723ff.html [accessed 2025-01-22]
61. Liu YL, Huang L, Yan W, Wang X, Zhang R. Privacy in AI and the IoT: the privacy concerns of smart speaker users and the Personal Information Protection Law in China. Telecomm Policy 2022 Aug;46(7):102334. [doi: 10.1016/j.telpol.2022.102334]
62. Tang Z, hu Y(, smith MD. Gaining trust through online privacy protection: self-regulation, mandatory standards, or Caveat Emptor. Journal of Management Information Systems 2008 Apr;24(4):153-173. [doi: 10.2753/MIS0742-1222240406]

XSL•FO
RenderX

63. Parker L, Halter V, Karliychuk T, Grundy Q. How private is your mental health app data? An empirical study of mental health app privacy policies and practices. Int J Law Psychiatry 2019 May;64:198-204. [doi: 10.1016/j.ijlp.2019.04.002]
64. Grundy Q, Chiu K, Held F, Continella A, Bero L, Holz R. Data sharing practices of medicines related apps and the mobile ecosystem: traffic, content, and network analysis. BMJ 2019 Mar 20;364:l920. [doi: 10.1136/bmj.l920] [Medline: 30894349]
65. Alfawzan N, Christen M, Spitale G, Biller-Andorno N. Privacy, data sharing, and data security policies of women's mHealth apps: scoping review and content analysis. JMIR Mhealth Uhealth 2022 May 6;10(5):e33735. [doi: 10.2196/33735] [Medline: 35522465]
66. Shipp L, Blasco J. How private is your period?: a systematic analysis of menstrual app privacy policies. Proceedings on Privacy Enhancing Technologies 2020 Oct 1;2020(4):491-510. [doi: 10.2478/popets-2020-0083]
67. Shi X. Reducing privacy risks of China's healthcare big data through the policy framework. Front Public Health 2024;12:1414076. [doi: 10.3389/fpubh.2024.1414076] [Medline: 39022418]
68. Sun L, Buijsen M. Mobile health in China: does it meet availability, accessibility, acceptability and quality standards? Health Policy Technol 2022 Sep;11(3):100660. [doi: 10.1016/j.hlpt.2022.100660]
69. Liu K, Xu G, Zhang X, Xu G, Zhao Z. Evaluating the privacy policy of Android apps: a privacy policy compliance study for popular apps in China and Europe. Sci Program 2022 Aug 23;2022:1-15. [doi: 10.1155/2022/2508690]
70. Shang J, Wei S, Jin J, Zhang P. Mental health apps in China: analysis and quality assessment. JMIR Mhealth Uhealth 2019 Nov 7;7(11):e13236. [doi: 10.2196/13236] [Medline: 31697245]
71. Zhang X, Lewis S, Chen X, Berry N, Bucci S. Technology use and attitudes towards digital mental health in people with severe mental health problems: a survey study in China. Front Psychiatry 2023;14:1261795. [doi: 10.3389/fpsyt.2023.1261795] [Medline: 38076680]
72. Wang Y, Liu Y, Shi Y, Yu Y, Yang J. User perceptions of virtual hospital apps in China: systematic search. JMIR Mhealth Uhealth 2020 Aug 12;8(8):e19487. [doi: 10.2196/19487] [Medline: 32687480]
73. Lian L, Feng Y, Zhao Y, Lei L, Daying F, Xiaohua W. Ethical dilemmas and solutions for patient privacy protection in internet hospital. Medicine & Philosophy 2025;46(14):22-27. [doi: 10.12014/j.issn.1002-0772.2025.14.05]
74. Zhang M, Chow A, Smith H. COVID-19 contact-tracing apps: analysis of the readability of privacy policies. J Med Internet Res 2020 Dec 3;22(12):e21572. [doi: 10.2196/21572] [Medline: 33170798]
75. Shen C, Jiang B, Yang Q, et al. Mobile apps for drug-drug interaction checks in Chinese app stores: systematic review and content analysis. JMIR Mhealth Uhealth 2021 Jun 15;9(6):e26262. [doi: 10.2196/26262] [Medline: 33962910]
76. Liu XH, Jin F, Hsu J, Li DN, Chen W. Comparing smartphone apps for traditional Chinese medicine and modern medicine in China: systematic search and content analysis. JMIR Mhealth Uhealth 2021 Mar 24;9(3):e27406. [doi: 10.2196/27406] [Medline: 33759786]
77. Meishan Z, Fengjuan W. The privacy protection dilemmas and strategies for chronic disease big data management. Medicine & Philosophy 2024;45(5):21-26. [doi: 10.12014/j.issn.1002-0772.2024.05.05]
78. Nissenbaum H. Privacy as contextual integrity [Symposium]. Wash L Rev 2004;79(1):119 [FREE Full text]
79. Lyu X, Liu X, Yang T, Fu C. Intervention strategies for health rumors: an overview based on the information lifecycle theory. Advances in Psychological Science 2024;32(4):664. [doi: 10.3724/SP.J.1042.2024.00664]
80. Katherine H. Websites' global privacy control compliance at scale and over time. Presented at: Proceedings of the 34th USENIX Security Symposium; Aug 13-15, 2025 URL: https://www.usenix.org/system/files/usenixsecurity25-hausladen.pdf [accessed 2025-08-22]
81. Notice from the Ministry of Industry and Information Technology on further enhancing mobile internet application service capabilities. Ministry of Industry and Information Technology. URL: https://www.gov.cn/zhengce/zhengceku/2023-03/02/content_5744106.htm [accessed 2025-08-01]
82. Chen M, Ding F. Research on the construction of a compliance evaluation index system for government website privacy policies. J Intell 2025;44(3):198-206. [doi: 10.3969/j.issn.1002-1965.2025.03.024]
83. Jing Y, Jiejie Z, Kailiang Z. Evaluation index system construction and empirical study of user privacy literacy on open government data platform. Library Work and Study 2024;1(12):17-26. [doi: 10.16384/j.cnki.lwas.2024.12.012]
84. Hendricks-Sturrup R, Lu CY. An assessment of perspectives and concerns among research participants of childbearing age regarding the health-relatedness of data, online data privacy, and donating data to researchers: survey study. J Med Internet Res 2023 Mar 10;25:e41937. [doi: 10.2196/41937] [Medline: 36897637]
85. Civil Code of the People's Republic of China. Trans-lex. URL: https://www.trans-lex.org/601705/_/civil-code-of-the-peoples-republic-of-china-/ [accessed 2025-03-01]
86. Cybersecurity law of the People's Republic of China. China Securities Regulatory Commission. URL: https://www.dataguidance.com/sites/default/files/en_cybersecurity_law_of_the_peoples_republic_of_china_1.pdf [accessed 2025-03-01]
87. Law of the people's republic of China on basic medical and health care and the promotion of health. The National People's Congress of the People's Republic of China. URL: http://www.npc.gov.cn/englishnpc/c2759/c23934/202012/t20201225_384343.html [accessed 2025-03-01]
88. Australian mHealth apps assessment framework. : Australasian Institute of Digital Health; 2024 URL: https://digitalhealth.org.au/wp-content/uploads/2024/06/AIDH_mHealth-Framework-Options.pdf [accessed 2025-03-01]

XSL•FO
RenderX

89. Ge F, Qian H, Lei J, et al. Experiences and challenges of emerging online health services combating COVID-19 in China: retrospective, cross-sectional study of internet hospitals. JMIR Med Inform 2022 Jun 1;10(6):e37042. [doi: 10.2196/37042] [Medline: 35500013]
90. Jiang X, Xie H, Tang R, et al. Characteristics of online health care services from China's largest online medical platform: cross-sectional survey study. J Med Internet Res 2021 Apr 15;23(4):e25817. [doi: 10.2196/25817] [Medline: 33729985]
91. Theodos K, Sittig S. Health information privacy laws in the digital age: HIPAA doesn't apply. Perspect Health Inf Manag 2021;18(Winter):1l. [Medline: 33633522]
92. Shahid J, Ahmad R, Kiani AK, Ahmad T, Saeed S, Almuhaideb AM. Data protection and privacy of the internet of healthcare things (IoHTs). Appl Sci (Basel) 2022;12(4):1927. [doi: 10.3390/app12041927]
93. Zhou SL, Jia X, Skinner SP, Yang W, Claude I. Lessons on mobile apps for COVID-19 from China. J Saf Sci Resil 2021 Jun;2(2):40-49. [doi: 10.1016/j.jnlssr.2021.04.002] [Medline: 40477492]
94. Liang F. COVID-19 and health code: how digital platforms tackle the pandemic in China. Soc Media Soc 2020 Jul;6(3):2056305120947657. [doi: 10.1177/2056305120947657] [Medline: 34192023]
95. Cong W. From pandemic control to data-driven governance: the case of China's health code. Front Polit Sci 2021 Apr;3. [doi: 10.3389/fpos.2021.627959]
96. Alamri H, Maple C, Mohamad S, Epiphaniou G. Do the right thing: a privacy policy adherence analysis of over two million apps in Apple iOS App Store. Sensors (Basel) 2022 Nov 19;22(22):8964. [doi: 10.3390/s22228964] [Medline: 36433560]

Abbreviations

mHealth: mobile health
PI: personal information
PIPL: Personal Information Protection Law

Edited by L Buis; submitted 09.Mar.2025; peer-reviewed by J Erickson, P Galetsi, Z Yu; revised version received 13.Nov.2025; accepted 22.Dec.2025; published 11.Feb.2026.

Please cite as:
Guo R, Chen F
Privacy Policy Compliance of Mobile Sports and Health Apps in China: Scale Development, Data Analysis, and Prospects for Regulatory Reform
JMIR Mhealth Uhealth 2026;14:e73651
URL: https://mhealth.jmir.org/2026/1/e73651
doi:10.2196/73651

© Rengui Guo, Fanhong Chen. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Perceived Sensitivity of Sensor-Based Digital Health Data: Qualitative Interview Study

Christine Deeney[1], BA; Anika Sonig[1], BA; Meghan E Hurley[1], MA; Birkan Tunç[2,3], PhD; Eric A Storch[4], PhD; John D Herrington[2,3], PhD; Jennifer Blumenthal-Barby[1], PhD; Kristin Kostick-Quenet[1], PhD

[1]Center for Medical Ethics and Health Policy, Baylor College of Medicine, One Baylor Plaza, Suite 310D, Houston, TX, United States

[2]Department of Child and Adolescent Psychiatry and Behavioral Services, Children's Hospital of Philadelphia, Philadelphia, PA, United States

[3]Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[4]Menninger Department of Psychiatry and Behavioral Services, Baylor College of Medicine, Houston, TX, United States

Corresponding Author:
Meghan E Hurley, MA
Center for Medical Ethics and Health Policy, Baylor College of Medicine, One Baylor Plaza, Suite 310D, Houston, TX, United States

Abstract

Background: Digital health tools are increasingly used in mental health care to passively collect patient data and analyze health status outside of clinical settings. While technologies such as digital phenotyping, affective computing, and computational behavioral analysis offer new insights into symptom manifestation in daily life, they generate large volumes of potentially sensitive data that raise significant data privacy concerns, requiring high levels of patient awareness and consent. Empirical research is lacking on stakeholder understandings toward the sensitivity of these data and expectations for data stewardship, perspectives that are critical for developing robust informed consent and data protection policies for digital health data use.

Objective: This study aimed to explore key stakeholder perspectives on the sensitivity of computer perception (CP) data, trust in existing data protections, willingness to share CP data externally, and desire for transparency of CP data transactions outside of the clinical space.

Methods: As part of a larger, multisite study, we conducted qualitative interviews (n=40) via Zoom (Zoom Communications, Inc) with 20 adolescents (aged 12 - 17 years) familiar with CP tools and their caregivers (n=20). Interviews consisted of a series of open-ended questions regarding stakeholders' perspectives on privacy, data security, and the use and exchange of CP data. We developed a qualitative codebook to identify and label thematic patterns in responses to questions addressing the topics above, using thematic content analysis to identify themes inductively. Each interview was coded by merging work from at least two separate coders, and several team members contributed to qualitative analysis.

Results: Most adolescents and caregivers viewed CP data as highly sensitive and expressed a reluctance to share these data beyond their clinical teams. While many participants expressed trust in existing data protections to protect CP data, they often misunderstood or overestimated the extent of protections to safeguard CP data.

Conclusions: Our findings underscore the critical need for clear and effective patient communication and education about the risks, benefits, and protections associated with CP data through informed consent protocols. To promote greater transparency, understanding, and trust, we recommend 5 strategies: educating patients about data protection; studying secondary data exchange and reidentification risks; strengthening transparency regulations; improving data traceability mechanisms, such as distributed ledger technologies, to enhance data traceability and auditability; and adopting dynamic consent models.

(JMIR Mhealth Uhealth 2026;14:e78788) doi:10.2196/78788

KEYWORDS

mobile health; mHealth; trust; data sharing; data protection; data privacy; transparency; digital phenotyping; affective computing

Introduction

Mental health care is undergoing rapid change as researchers and clinicians adopt digital tools and mHealth (mobile health) apps to support diagnosis and monitoring. Approaches such as digital phenotyping [1-5], affective computing [6,7], and computational behavioral analysis [8,9] involve algorithmic analysis of continuously and passively collected data from patients outside of clinical settings in an attempt to offer more ecologically valid insights into how symptoms manifest in daily life. We refer to these approaches as computer perception (CP) technologies, a term that builds on computer vision but reflects a broader set of sensing methods, including voice, movement, location, and physiological signals (eg, heart rate and skin

XSL•FO
RenderX

conductance). Despite their nuanced distinctions [5], these CP approaches generate large volumes of potentially sensitive data to enable various forms of clinical inference, particularly when triangulated and used alongside remotely administered self-report measures ("ecological momentary assessments"), often via mHealth platforms. Given their promise for data-centric and value-based care, CP technologies are increasingly applied across psychiatry and other areas of medicine (eg, cardiology [10], pain management [11], maternal health [12], and asthma [13]) that benefit from insight into behavioral patterns outside the clinic.

At the same time, CP tools also introduce significant ethical and privacy concerns, necessitating robust approaches to informed consent and data protection that directly respond to how individuals understand the sensitivity of these data and how they are—or should be—protected. A major challenge in digital health is the misalignment between stakeholder expectations and actual data governance practices [14-16]. Individuals may assume these data are safeguarded under existing health privacy laws, such as the HIPAA (Health Insurance Portability and Accountability Act), even when these data fall outside formal regulatory coverage [17-19]. Prior work shows that when perceived safeguards prove illusory, this misalignment can erode trust, reduce willingness to share data, and compromise sustained engagement with digital health tools [20].

This paper builds on this literature by presenting results from in-depth interviews exploring stakeholder understandings of the sensitivity of CP data, trust in existing data protections, willingness to share CP data for clinical or research purposes, and preferences for transparency around CP data transactions outside the clinical space. Although our empirical data were collected in the United States and reference US health privacy frameworks, the ethical and practical challenges we examine are not unique to North America. Across jurisdictions, rapid growth in digital health and sensing technologies has outpaced public understanding of how personal data is governed, shared, and reused. Even under the European Union's General Data Protection Regulation (GDPR), which provides broader formal protections for certain categories of personal data, enforcement gaps, limited transparency, and user misunderstandings persist, particularly for passively collected behavioral and biometric data [21].

Many CP data are not directly clinical in nature. Unlike traditional medical data, CP data may be collected for clinical purposes while remaining usable for nonclinical applications, such as marketing or profiling [19,22]. Importantly, CP data often fall between established categories of personal data, combining elements of identity data (eg, location, voice, facial features, and device identifiers) and clinical data (eg, inferred symptoms or functional states), even when they are not explicitly diagnostic on their own. Public perspectives on digital health data indicate that users may not perceive their self-tracked health information with the same sensitivity as traditional personal health information (PHI), despite the fact that these data can reveal sensitive information about a person's health or reproductive status [19].

As these data are not always classified as PHI, these "nontraditional health" (NTH) data often receive weaker legal protections, particularly in jurisdictions where health privacy laws apply narrowly to clinical records and covered entities. In the United States, for example, many CP data fall outside the scope of HIPAA, whereas in other jurisdictions, such as the European Union, broader data protection frameworks such as the GDPR may apply, at least in principle, to biometric and behavioral data. However, this regulatory ambiguity creates gaps in transparency, accountability, and user awareness [23,24]. Additionally, public misunderstandings regarding GDPR's scope in relation to biometric or behavioral tracking data contribute to ongoing challenges in ensuring informed consent and ethical data management [25].

We argue that this misalignment reflects a more fundamental lack of computational mechanisms, alongside supportive regulatory frameworks, that enable transparent tracking of patient and consumer data. This lack of visibility into how data moves across clinical, research, and commercial contexts perpetuates misunderstandings, power imbalances, and trust gaps between individuals, institutions, and technology systems, leading to collective vulnerabilities in privacy, autonomy, and equitable access to digital health benefits.

These ambiguities leave NTH data vulnerable to unwanted or undisclosed forms of data access and exchange by third-party data controllers (eg, large technology companies), particularly given that many mHealth apps lack adequate data security and confidentiality assurances [26]. Further, because many CP data are generated in the context of research-industry partnerships, uncertainties persist around what may be done with these data, to which entities control or claim ownership rights, where data are stored (eg, hospital vs corporate servers), and who holds decision-making authority over data access and exchange (eg, researchers, companies, or data brokers). A growing body of research suggests that public trust in data protection measures varies by institutional affiliation, with greater skepticism toward commercial data handlers [27].

Such concerns extend across data types, ranging from PHI to NTH data, collected within health care [19,22,28,29] as well as outside of health care, including browser history and consumer behavior data used for targeted marketing and personalization [29,30]. Extensive empirical literature documents public concerns about data privacy, transparency, and control and a desire for greater authority over what personal information is collected and who has access to it [27,31-35]. While numerous studies have examined perspectives on privacy, security, and confidentiality in mHealth apps [36-38], and others have explored attitudes toward CP technologies in relation to acceptability, adoption of, and adherence [37-45], few specifically investigate how individuals understand the handling of CP data after collection or how these understandings align with the realities of health information exchange (HIE) in the United States and elsewhere.

Patients and research participants may assume these data constitute PHI when collected in clinical or research settings, leading to misinterpretations of the degree to which these data are protected under data privacy laws—much as consumers

XSL•FO
RenderX

often assume their personal data are protected by default on websites or mobile apps without reading the terms of service or privacy policies [46-48]. Such misinterpretations are compounded by limited transparency about the benefits and risks of collecting CP data at the time of consent, as computational analysis capabilities continue to expand.

Examining alignment or misalignment in understandings about the nature and governance of CP data is therefore critical for developing patient education and informed consent protocols that clearly articulate potential benefits and risks. In this paper, we present findings from a qualitative investigation of adolescents' and caregivers' perspectives on sharing passively collected CP data. Our study aimed to explore key stakeholder perspectives on: (1) the perceived sensitivity of CP data; (2) trust in existing data protections governing CP data; (3) willingness to share CP data for clinical or research purposes; and (4) preferences for transparency around CP data practices and transactions—with the goal of informing more robust informed consent protocols and data governance recommendations for CP and related digital health technologies. We focus on CP data collected in settings where research and clinical care increasingly overlap, such as clinical trials embedded in health care systems, validation studies for tools intended for clinical deployment, and digital health platforms developed through academic-industry partnerships. While research and clinical contexts are governed by distinct ethical and regulatory standards, participants' expectations about data protection and use often span, or fail to distinguish between, these domains. We conclude by offering targeted recommendations to improve transparency and informed consent practices surrounding the collection and exchange of CP and other digital health data.

Methods

Study Design

This research aimed to explore ethical considerations for translating CP technologies into clinical care. Through in-depth, semistructured interviews, we explored perspectives on benefits, risks, and concerns around the integration of CP technologies, which also involved a supplementary exploration of specific concerns around privacy, data security, and perspectives toward the use and exchange of CP data.

Participants

As part of a multisite study funded by the National Center for Advancing Translational Sciences (R01TR004243), we recruited 40 participants between January 2023 and August 2023. Participants were recruited from a sister research study (R01MH125958) conducted within a clinical research setting, focused on validating CP tools designed to quantify objective digital biobehavioral markers of socioemotional functioning. Participants in our sister study were recruited from a clinical care setting in a large academic medical center using word of mouth. As part of that study, CP data were collected in a laboratory setting for research purposes with the stated goal of informing future clinical applications. As such, participant perspectives reflect experiences situated at the boundary of research and clinical care. Our choice to recruit from this sister study was also to ensure familiarity with and exposure to the collection of CP data, to enhance the ecological validity of our findings. Participants included adolescents (n=20) aged 12 - 17 years and their caregivers (typically biological parents, see Table 1). The adolescents had previously received neurodevelopmental (ie, autism) and/or psychiatric diagnoses of autism, anxiety, or depression and potentially diagnoses that often accompany these (eg, attention deficit/hyperactivity disorder or obsessive compulsive disorder). Diagnostic presentations for all adolescents were confirmed through the administration of standardized clinician interviews. Participants were told about this study by a research coordinator from our sister study. Individuals who expressed interest in participating were referred to this study by the sister study's coordinator and then contacted by a research assistant (RA) via phone or email to schedule a virtual interview. Our response rate was approximately 63% for patients and caregivers (63 people were contacted and 40 people agreed to participate).

XSL•FO
RenderX

Table . Demographics for interviewed adolescents and caregivers. Values may not total to 100% due to categorical overlap (eg, comorbidities). Frequencies and percentages reflect only reported information and are not intended to suggest any level of statistical significance.

	Adolescents (n=20, 50%)	Caregivers (n=20, 50%)	Total (N=40, 100%)
Sex, n (%)			
Male	12 (30)	2 (5)	14 (35)
Female	8 (20)	18 (45)	26 (65)
Race or ethnicity, n (%)			
American Indian or Alaska Native	0 (0)	1 (3)	1 (3)
Asian	1 (3)	1 (3)	2 (5)
Black or African American	5 (13)	4 (10)	9 (23)
Native Hawaiian or Other Pacific Islander	0 (0)	0 (0)	0 (0)
White	17 (43)	15 (37)	32 (80)
Hispanic or Latino	4 (10)	2 (5)	6 (15)
Not Hispanic or Latino	16 (40)	18 (45)	34 (85)
Marital status, n (%)			
Married and living with spouse	—[a]	13 (65)	13 (33)
Widowed	—	1 (5)	1 (3)
Divorced	—	4 (20)	4 (10)
Separated	—	1 (5)	1 (3)
Never married	—	1 (5)	1 (3)
Education, n (%)			
High school only or less	—	0 (0)	0 (0)
Trade school or associate's degree	—	2 (10)	2 (5)
Bachelor's degree	—	10 (50)	10 (25)
Master's degree	—	4 (20)	4 (10)
Doctoral degree	—	4 (20)	4 (10)
Parental status, n (%)			
Biological parent	—	18 (90)	18 (45)
Stepparent	—	0 (0)	0 (0)
Adoptive parent	—	2 (10)	2 (5)
Diagnosed condition, n (%)			
OCD[b]	4 (20)	—	4 (10)
Autism	5 (25)	—	5 (13)
ADHD[c]	3 (15)	—	3 (8)
Anxiety	4 (20; 1 self-reported)	—	4 (10)
Tourette	1 (5)	—	1 (3)
No clinical diagnosis or symptoms	9 (45)	—	9 (45)
Age, mean (SD)			
Average age	14.9 (2.2)	48.3 (6.4)	—

[a]Dashes indicate categories for which n=0 (0%).

[b]OCD: obsessive-compulsive disorder.

XSL•FO
RenderX

[C]ADHD: attention-deficit hyperactivity disorder.

Data Collection

We developed separate interview guides for adolescents and their caregivers (Multimedia Appendices 1 and 2), exploring the same constructs across both groups with appropriate variations in wording. Guides were informed by prior work exploring user perspectives in data privacy and governance [14,15,17,19,20,22,27,29-38] and with the guidance of experienced bioethicists, child mental health experts, and computational scientists who develop these algorithms. Constructs included data security and privacy issues, regarding perceived sensitivity of this type of data and data collection, beliefs about the potential for unintended uses and current data protections, thoughts on secondary usage of their data and access authorization, and transparency needs or preferences of their personal data and its transactions or security processes. Initial drafts of the interview guides were piloted with two psychologists specializing in adolescent mental health, resulting in minor clarifications in wording. Interviews were conducted by RAs (authors CD, AS, and MEH) with extensive training in qualitative interviewing and analysis. Interviews were held via a secure videoconferencing platform (Zoom for Healthcare [Zoom Communications, Inc]) and lasted an average of ≈45 minutes. Before beginning the interview, each participant was presented with a 7-minute primer video (Multimedia Appendix 3) explaining the nature and intended use of CP data in clinical research and health care (ie, a basic, 8th grade-level understanding of how CP technologies collect and analyze data and how they might be used during a clinical encounter), including the different types of CP technologies available and the types of data they collect (eg, different types of wearables, smart devices, and sensors that stakeholders would likely be familiar with and that can collect vocal, facial, accelerometry, GPS, heart rate data, and more). The primer video was developed to ensure informed responses from patients and caregivers who are likely to be less familiar with this suite of technologies. Additionally, the RA stopped the video at key predesignated intervals (eg, after the original explanation of how CP collects and analyzes patient data, after providing an example of what the integration of these tools into a traditional clinical assessment could look like, etc) to assess understanding and elicit clarifying questions from participants to ensure comprehension. Recruitment continued until we observed significantly diminished novel informational returns from each subsequent interview ("saturation"). Operationally, research team members reviewed transcripts to identify novel insights not present in prior interviews. Saturation was determined when successive interviews within a stakeholder group yielded no new insights, which occurred in around 20 interviews per stakeholder group.

Ethical Considerations

This study was reviewed and approved by the Baylor College of Medicine Institutional Review Board (H-52233), which also waived a requirement for written consent due to research procedures involving minimal risk to participating stakeholders; thus, participants provided verbal consent. Minors provided assent with parental consent. Identifying participant information was stored in a password-protected system with 2-factor authentication behind a university firewall. All results are reported in the aggregate and are not linked to identifiable participants. Each participant received a US $50 gift card as compensation for their time.

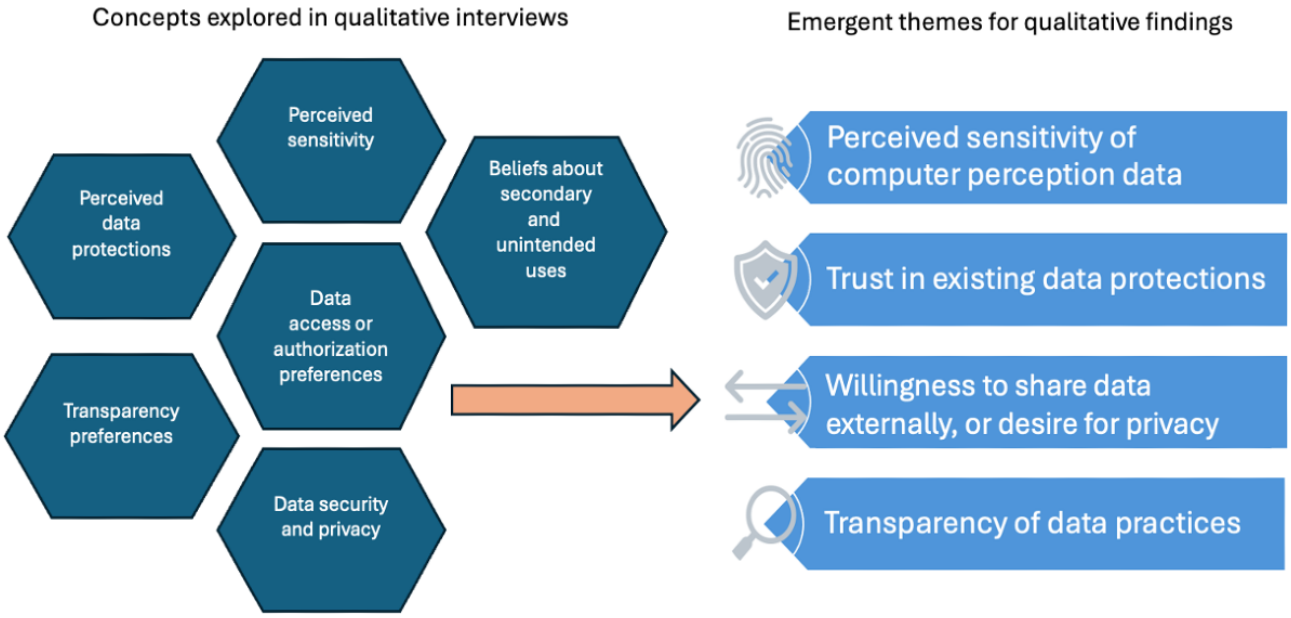
Data Analysis

Interviews were audio-recorded, transcribed verbatim, and analyzed using MAXQDA (VERBI GmbH) software. Led by a qualitative methods expert, team members developed a qualitative codebook to identify and label thematic patterns in adolescent and caregiver responses to questions addressing the topics above. Codebook development followed a hybrid inductive–deductive approach. We formed an initial set of broad analytic domains informed by prior literature on digital health ethics, data governance, trust, privacy, and informed consent [19,23-27,29-38], which served as sensitizing concepts rather than fixed categories. Within these domains, codes were developed inductively through our familiarity with the interview content (all team members listened to all interview recordings) and a close review of a subset of transcripts to capture emergent concepts. Codes and definitions were iteratively refined in collaborative team meetings, and subsequently used to identify patterns in the data using thematic analysis [49]. Specifically, outputs for each code were further analyzed by progressively abstracting relevant quotes, an inductive process that entails reading every quotation to which a given code was attributed, paraphrasing each quotation (primary abstraction), and further identifying which constructs were addressed by each quotation (secondary abstraction). These constructs mapped onto a core set of thematic findings across four emergent domains (see Figure 1). Each of the concepts in these four domains was further categorized along a scale of "high," "medium," or "low" (eg, "high" need for transparency) to examine relationships between beliefs and perspectives, informational needs and preferences, and trust. These designations reflect consensus-based qualitative judgments by the research team, grounded in close analysis of coded excerpts, rather than participant ratings or a formal scale. Each interview was coded by merging work from at least two separate coders to reduce interpretability bias and enhance reliability. To ensure the validity of our findings, all abstractions were validated by at least one other member of the research team before calculating thematic frequencies to characterize stakeholders' responses and obtain a sense of the proportion of stakeholders raising or discussing each concern/theme. In rare cases where abstractions reflected conflicting interpretations, members of our research team met to reach consensus. Frequencies were used as a descriptive aid to convey the relative salience of perspectives across participants, consistent with pattern-based thematic analysis, and were not intended to suggest statistical significance or quantitative generalizability [49]. Frequencies account for participants who expressed perspectives that were ambiguous or context-dependent (eg, perceiving CP data to be highly sensitive in some situations but not in others) and were calculated when respondents expressed views across the high, moderate, and low categories. Frequencies and percentages reflect only reported information and therefore do not always

XSL•FO
RenderX

sum to 100%, as some respondents did not express perspectives across all themes.

Figure 1. Transformation of qualitative inquiry domains into thematic findings.



Results

Overview

The results are organized around four analytic domains identified through thematic analysis of transcripts from 20 adolescents and 20 caregivers (total N=40; see Table 1). These domains include the following: (1) perceived sensitivity of CP data; (2) trust in existing data protections; (3) willingness to share data externally and desire for privacy; and (4) transparency of data practices (see Table 2). Within each domain, we describe the range of perspectives expressed by participants and use qualitative descriptors (eg, high, moderate, and low) to summarize relative emphasis and salience, illustrated with representative quotations (eg, higher to lower perceived sensitivity and higher to lower trust in existing data protections). Frequencies and percentages of expressed views across these domains are presented in Table 2, intended to aid interpretive comparison between and across groups. Taken together, our results indicate that most caregivers and adolescents view CP data as moderately to highly sensitive, are reluctant to share these data beyond their clinical teams, and expressed a desire to have visibility into or control over both primary and secondary uses of their data. Notably, adolescents expressed higher trust in existing data protections compared to caregivers. Most caregivers shared that they were less willing to share data externally compared to adolescents. The following sections elaborate on and summarize these findings, complemented by illustrative quotes.

XSL•FO
RenderX

Table . Frequencies and percentages of expressed views across thematic domains.

	Adolescents, n (%)	Caregivers, n (%)	Total, n (%)
Perceived sensitivity of CP[a] data			
High (only)	—[b]	1 (5)	1 (3)
Moderate to high	11 (55)	12 (60)	23 (58)
Moderate (only)	4 (20)	1 (5)	5 (13)
Moderate to low	5 (25)	1 (5)	6 (15)
Ambivalent (both high+low)	1 (5)	1 (5)	2 (5)
Significant ambivalence (high+moderate+low)[c]	2 (10)	4 (20)	6 (15)
Low (only)	—	—	—
Trust in existing data protections			
High (only)	9 (45)	5 (25)	14 (35)
Moderate to high	3 (15)	3 (15)	6 (15)
Moderate (only)	2 (10)	6 (30)	8 (20)
Moderate to low	—	—	—
Ambivalent (both high+low)	—	—	—
Low (only)	—	—	—
Willingness to share CP data			
High (only)	—	—	—
Moderate to high	—	—	—
Moderate (only)	—	1 (5)	1 (3)
Moderate to low	9 (45)	18 (90)	27 (68)
Ambivalent (both high+low)	—	—	—
Significant ambivalence (high+moderate+low)[c]	4 (20)	—	4 (10)
Low (only)	—	—	—
Desire for transparency in use or secondary use of CP data			
High (only)	11 (55)	12 (60)	23 (58)
Moderate to high	3 (15)	3 (15)	6 (15)
Moderate (only)	—	1 (5)	1 (3)
Moderate to low	—	—	—
Ambivalent (both high+low)	—	—	—
Low (only)	—	—	—

[a]CP: computer perception.

[b]Dashes throughout the table represent areas where no participant expressed perspectives that fall into those categories (eg, no adolescents or caregivers viewed computer perception data to be “low sensitivity”).

[c]This category was only present for these 2 themes but represents significant ambivalence among certain participants who expressed views across each level.

Perceived Sensitivity of Passive Continuous Monitoring Data

Most participants (15 adolescents and 14 caregivers) perceived CP data to be highly sensitive. Participants commonly understood CP data as revealing intimate aspects of identity, emotion, and daily life, leading many to view continuous monitoring as intrusive even when collected for clinical purposes. Some characterized it as an “invasion of privacy” (A16) because such data capture the “emotions and feelings of a person” (A16) or “who you are on the inside” (CG03; see Table 3). Some caregivers characterized continuous monitoring as “unnerving” (CG08) or “debilitating” (CG03), or likened it to “big brother” surveillance (CG19). More than half of all participants expressed a desire to pause or disable monitoring

XSL•FO
RenderX

during moments they considered private. These moments included private conversations, conflicts with family members, funerals, intimate situations, or emotionally vulnerable situations. Both stakeholder groups expressed a view that the potential negative impacts of collecting CP data may be especially pronounced in the context of physical and socioemotional changes associated with adolescence (eg, "where it would affect [their] self-image" CG07).

By contrast, a smaller number of participants (5 adolescents and 1 caregiver) viewed CP data as less sensitive, with one adolescent stating that it felt "sort of like wearing a Fitbit" (A07), particularly when perceived clinical benefits were salient (as noted by 6 adolescents and 9 caregivers).

Some stakeholders (3 adolescents and 5 caregivers) expressed ambivalence. For example, one caregiver shared that while the data itself did not seem invasive, since "our watches collect and our phones collect this kind of stuff all the time" (CG01), they would nevertheless be bothered if the data were shared with third parties (eg, insurance companies and future employers) without their knowledge (see also Willingness to Share CP Data section below).

XSL•FO
RenderX

Table . Quotes illustrating varying levels of perceived sensitivity of CP[a] data.

Perceived sensitivity: stated rationales	Adolescent quote	Caregiver quote
Highly sensitive		
CP data deserves greater protection than biological data	"It is a little bit more sensitive because it feels a lot more with the emotions and feelings of a person. So I do think they should be treated with a little bit of extra care." A16	"We need to have more protection [for CP data]. Because it's just different information. That's the voice, the facial expressions, again, that it's not going to be used for wrong purposes." CG16
CP data deserves the same high legal protections as other medical data	"It should be treated equal, because all medical information, unless you actually disclose it to somebody, it's not supposed to be known." A01	"Probably about the same [amount of protection]... Certainly not less." CG01
CP captures highly sensitive (personal) information: emotions, facial expressions, etc	"I think some people would maybe be reluctant to do things like this... Maybe people are afraid that something…may get exposed or… don't want someone examining their facial expressions [or] they don't want other people to know about something [a condition] they may or may not have." A01	"The physician has access to your naked body. But with emotions it's like having access to the naked naked because it's really who you are on the inside too. Things that you are unable to express. A machine that would be able to intimately know where you are, it's absolutely debilitating, it's scary." CG03
Passive continuous data collection is invasive and unsettling	"I wouldn't necessarily want it to monitor everything that I say. I would think that would be a bit invasive." A02 "It feels kind of invasive because I feel like it would just depend on the situation. You don't know, maybe there's a reason you aren't telling people about it and if they know without asking...it just seems invasive sometimes." A14	"So her DNA test says…she has a 40% higher chance than most people of developing, in the future, age related macular degeneration. That was helpful. So she takes lutein and ocuvite because it's progressive through time. But that was very easy to do. It was private. Blood test. She knows. I know. Her doctor knows, but it's not as invasive as what you're proposing." CG07
CP data collection is especially sensitive for adolescents	"If I was getting in a fight with my siblings or stuff like that, maybe I wouldn't want that to be recorded." A13	"When they hit high school. It gets very difficult. You've got all of those other complexities, the peers and society, I'm trying to figure out how I fit. It's really complex. When you're talking personal intrusiveness, I don't know. I'm a little sketchy on that." CG11
Moderately sensitive		
CP is invasive but justified by clinical benefits	"It's a little invasive. I personally wouldn't want my doctor to see like every single emotion and feeling that I'm having throughout the day. But at the same time, if there's someone who's unable to tell their feelings or emotions or what they're feeling properly, then it could be great help. It's helpful at the same time." A04	"If we come to the doctor and they explain, 'Hey, this is going to be better to help me, help you,' then sure, but unless we specifically come with a very specific issue... That's probably overkill. They deserve some privacy." CG04
Prefer ability to turn off continuous data collection in certain circumstances	"I feel actually most places I would probably just have it turned off because I don't like the idea of people knowing where I am at all times in case that data gets out… That just seems like a weird thing to me." A07 "When I'm going to the restroom or something like that." A17 "Probably when I'm taking a shower or something." A05 "Probably just my sleep." A04	"What if she wants to get intimate with her partner or to take a bath or she's having a really hard bowel movement that day. Those are certain things that might raise your heart rate that you don't necessarily want anybody to know about. It's very personal and doesn't necessarily need to be part of her diagnosis process... I think there should definitely be times you're allowed to take it off." CG05
CP data is more vulnerable to misinterpretation (needs context, can be subjective, etc)	"I guess the eye tracking and the listening to my voice**…**it's all being processed by maybe someone else or AI or something like that. And maybe someone else interprets what you're feeling differently than what you are actually feeling. Maybe you're trying to hide something and they just don't see it." A02	"I just worry that it could be wrong because sometimes computers don't know everything. They're not human. And it would be messed up for it to say, oh, well this person is feeling this and that person is not feeling that. It could mess up something. Oh, your child's feeling sad, to the point where they want to kill themselves. And [your kid is] like, 'No, that's not what I feel.' So I just worry about the error rate." CG08
Less sensitive		

XSL•FO
RenderX

Perceived sensitivity: stated rationales	Adolescent quote	Caregiver quote
No strong desire to turn off continuous monitoring in certain situations	"I mean there's nothing that's insanely private. I mean, [there aren't any moments where I would not want it to collect data in the moment], not that I can think of off the top of my head at least." A12	"Again, without being in that moment, no... I can't think of a thing where I would want total and utter privacy." CG10
Views CP as having a similar (low) sensitivity as other wearable or smartphone data	"I don't have that huge of a problem with it. I think it's sort of like wearing a Fitbit where it counts your steps and things like that. My only concern would be privacy, being shared with other people." A07	"Our watches collect and our phones collect this kind of stuff all the time. I think that if that's being transmitted to your insurance company or potential life insurance or disability insurance companies and they're going to rate you based on things like that, I guess that would probably bother me to some extent. But you don't want to get paranoid about it, you just want to be realistic about it." CG01

[a]CP: computer perception.

Trust in Existing Data Protections

Most participants (14 adolescents and 14 caregivers) expressed moderate to high trust in existing data protections governing CP data (see Table 4). This trust was largely grounded in confidence in clinicians and health care institutions, often accompanied by assumptions that CP data are protected under established health privacy laws. Several adolescents stated they had no concerns about CP data being shared beyond their doctor (eg, A11 and A13), while others emphasized that clinicians "are there to help you" and therefore could be trusted with sensitive data (A16). Some caregivers explicitly assumed that CP data were protected under HIPAA, stating that they "would expect the same type of privacy laws to apply" (CG03). This subset expressed more cautious trust, noting that while they had "no major concerns," they still wanted reassurance that data protections were "following HIPAA" (CG05). Notably, no respondents expressed low or no trust in existing data protections.

Table . Quotes illustrating varying levels of trust in existing data protections.

Trust: stated rationales	Adolescent quote	Caregiver quote
High trust		
Few, if any, concerns or perceived harm from sharing CP[a] data	"I think it depends on who it would be shared to, but I don't think for the majority I would have a problem with that being information being shared." A02	"I'm not sure what kind of problems could really arise from that [information collected potentially be shared with people other than their doctor]." CG09
Less concerned about unauthorized and harmful external data sharing because they trust or expect that it is kept safe/secure	"With every bit of information being locked down with multiple encrypted passwords, and everything's double encrypted, I have no concern." A09	"I know it's a very confidential doctor patient kind of relationship so there's no harm to my child's reputation." CG20
Expects/assumes HIPAA[b] already protects this type of user-generated data	"I guess it really depends on who it's getting shared to, but I guess it would be the same as what HIPAA is for. It would just be sent to the people that would be able to act upon it super fast..." A19	"I feel like it would be a part of their health information, so it would be private. I'm not any more worried about the technology specifically as opposed to overall somebody knowing he has an anxiety or OCD." CG18
Moderate trust		
Wants to be reassured of HIPAA protection	"What I would need though as a very firm understanding of the boundaries of the utilization of that data because HIPAA's a pretty big thing... I would want to know could anything in this data be not protected, why and how." CG12	"I want to know that it's HIPAA approved and although there's always a risk for a leak data because it's not a perfect world and there's a lot of hackers out there, just knowing that it's as safe as possible." CG05
Limited trust of technology developers to respect privacy (fear of technology)	"Even though you said that it wouldn't, I feel like I would always be nervous that it is tracking the things that I'm saying and what I'm actually doing, like even though it says it isn't, I feel like that would just be a fear." A14	"Right, the privacy aspect of it. Is it actually collecting texts or voicemails or recording conversations? We're always questioning whether our Alexa is actually listening to us." CG14
Low trust	—[c]	—

[a]CP: computer perception.

[b]HIPPA: Health Insurance Portability and Accountability Act.

[c]Not applicable.

XSL•FO
RenderX

Willingness to Share CP Data

Most participants were reluctant to share CP data beyond their immediate clinical teams. Reluctance to share reflected concerns about privacy, loss of control, and potential misuse of data, particularly among caregivers responsible for their children's information. Participants frequently emphasized ownership and control, with one adolescent stating, "It's my information, keep it that way" (A05). Caregivers expressed near-universal hesitancy about external sharing, particularly with commercial entities. Notably, twice as many caregivers (n=18, 90%) expressed this reluctance as adolescents (n=9, 45%) (Table 2).

Their rationales (see Table 5) centered around a common view of these data as invasive and sensitive. At the same time, some participants expressed ambivalence or conditional willingness, acknowledging that CP data could be "helpful" (A16) or valuable for research while simultaneously asserting that such data "should be mine and my doctor's, and that's it" (A07). Some adolescents and caregivers stated a preference to be able to vet or control what information their doctors could access. A small number expressed resignation or apathy, suggesting that data sharing "does not make much difference" given the broader data ecosystem (A02).

XSL•FO
RenderX

Table . Quotes illustrating varying willingness to share CP[a] data.

Desire for privacy and willingness to share: stated rationales	Adolescent quote	Caregiver quote
Most willing to share		
Positive attitudes toward secondary uses	"I think it's probably going to be really helpful for future researchers if they can better whatever they're trying to research and figure something out that could help a lot of people, then I'm all for it." A16	—[b]
Moderately willing to share		
Secondary uses must come with stipulations (informed consent, intended use, reputation, and compensation)	"Yes (I would be willing if I knew) just what they're using it for, how long they're going to hold onto it, who gets to see it and is the company or whatever who's accessing the data from a reputable source." A04	"I'm comfortable with it as long as, if there is a way to let the participants know that, 'Hey, there's a chance that your information is going to be used in another study.' And identify that study." CG04
Willing to share if given control over data-sharing decisions	"I would be more willing to share if I have complete control over it." A03	"If I have the ability to say no, I might be a little more open-minded about who wants to see it for why. But again, I would need to have that information." CG02
CP data should be kept private, on a need-to-know basis with medical personnel (and treated like other medical data—not shared widely)	"I do think they should be kept private for a good amount... Not everybody has to see it except for my doctor and other people who me, my doctor, give authorized access to. But other than that should be pretty private." A16	"I think that it should probably not go to future employers or colleges, it's personal information. And I think too that now in this day and age… [data] is it sold somewhere to advertisers to then manipulate you? I mean, you can get paranoid about it, but I think that there is some level of that concern about what happens to that data." CG01
Reluctant to share CP data when it oversteps relationship boundaries	"I feel like [I would] maybe [be worried about sharing this data with] my parents, I don't know. If I'm trying to deal with something and I don't want them on my back about it, I feel like I wouldn't want them to know." A14	"I could see that with the younger population. But then again, we have that crossover when they get to be young adults. It's so intrusive and you would probably produce battles that you wouldn't really want… there's a certain level of that's his right to not talk about it. It really sucks when I don't know…But you have to really be careful on when they get older, there's more permission that they have to give. It's been hard to balance that…" CG11
Prefer to vet what doctors can see	"I think I'd have to know that it would show me the results first and then ask if I'm okay with sending it to the doctor, instead of just sending it to the doctor first and then the doctor telling the things. You get what I'm saying?" A07	"Can the doctor themselves go in and listen to a conversation because I wouldn't like that. Now, if the data's being analyzed in a way that the doctor doesn't have access to and it's shooting off reports, I'm far less troubled by that." CG12
Least willing to share		
Would "not" influence willingness to share because control/privacy does not seem feasible, or they are not interested	"I don't think it [sharing] would make much of a difference. I mean, there's some stuff that's probably already out there... But if it was something that really influences my health or just my mental wellbeing, if it really matters that much to me, I would probably care about it. But just in general, probably not." A02	"I just don't see that there's really any safety valves in place that are 100% secure... So I say, yes, this person can have it… if they're on a research team with the other person is going to see the data anyways... You have the researchers starting public companies…and then that information that they received under a secure process has now moved from academia out into the public sector…I just don't see that this is ever going to be really safe. That's concerning." CG01

[a]CP: computer perception.

[b]Not applicable

Transparency of Data Sharing

Most participants expressed a desire for greater transparency regarding how CP data are collected, used, and shared. Transparency was valued as a means of maintaining trust and agency, though participants varied in whether transparency alone was sufficient without accompanying control. Participants (6 adolescents and 6 caregivers) sought clarity about what data were collected and when, asking "what information it's gathering, when it's gathering it, what it's tracking" (CG07; see

XSL•FO
RenderX

Table 6). Others focused on secondary use, questioning "Where is this going? Who are you sharing it with? What are we doing with this information?" (CG04). Some participants (8 adolescents and 7 caregivers) emphasized the importance of knowing how access was authorized, asking, "Who authorizes that sharing?" CG15; "How is it controlled? What happens if somebody who isn't supposed to see it, [sees it]?" A16), and how transactions will be tracked ("How are you going to document?" CG11).

XSL•FO
RenderX

Table . Quotes illustrating varying preferences for transparency in CP[a] data exchange practices.

Transparency preferences: stated rationales	Adolescent quote	Caregiver quote
Strong desire for transparency		
Desire for transparency of what data CP collects	"I think I would definitely want to know what information is being tracked and how or where the information is going and how it's being used." A02	"I'd want to know exactly what information it's gathering, when it's gathering it, what it's tracking." CG07
If given the chance, they would audit their data transactions frequently	"[I would check] Pretty often I suppose. Mainly because it is a lot about me [that] not a lot of people know about**,** so having a lot of people look at it, I would probably be like, oh, this is out there and let me make sure it's people I trust and approve of seeing it." A04 "I'd want to know just what it's [that data] going to. I wouldn't like say, no, you can't use this information… But I would want to know just who's using it. I would just check up every now and then to see okay. What's happening with it?" A02	"I would expect there would be an alert, some notification that that is happening. And maybe check it once a week anyway myself, kind of like what I do with my bank accounts, just go check and see if the charges are what I made. That's something that I would feel comfortable with… [then] I would be able to reach out that much sooner to the research institute and start asking questions." CG04
Desire for transparent data sharing policy (eg, at time of consent)	"…mainly just with the privacy, but how it's being stored and who gets to review it and also just the development. If that information is given out early on, then I'll be like, 'Oh, feel more secure using the tracking and such.'" A16	"I would want to know what they were collecting and for how long. If they're going to do it for six weeks, I would want to know some assurance that at six weeks it's no longer collecting data. And then how they were going to protect the information... Who had access to it, I guess. And that they had some steps in place to protect the information." CG17
Desire for transparency about data protections	"How is it controlled? Who gets to see it? When are they able to see it and do the person who gave the data have control over who gets to see it? And then…what happens if somebody who isn't supposed to see it sees it and stuff like that." A16	"[I would want to know] the privacy or the security of that data, that it wouldn't fall into the wrong hands, misused. I mean, collecting data for me, is wonderful. You get a different picture combined with the actual person's thoughts, ideas, it's more meaningful, but it's more about the security to that, I would always have a concern about." CG16
Moderate desire for transparency		
Value transparency, but only if accompanied by opportunities for control	"If I'm not able to do anything about it and I see that there's a company that's using my data in a way that I don't want, then I'd rather not know about it. But if there is something that I can do about it, then I'd like to be updated on it and know about it and check it often." A07	"I don't think there's any point in [transparency]…when you know something [that] you can't do anything about." CG01
Might have negative reactions to discovering unwanted data transactions	"I'd be pretty weirded out if I didn't know about it or I wasn't made aware of it. And I think I'd be pretty upset if it was used by a company to make a product that I don't support or something like that." A03	"Well, if they're getting access to it and I didn't give them consent to it, I'd be very upset. If I don't recognize it, I would like to be able to ask somebody who it is. But if it's just a company that just buys and sells data, no, I would not be very happy about that." CG02
Moderately interested in transparency and, if given the chance, would only audit their data transactions once in a while	"I think if I looked at [transparency of data transactions] too much, I'd be obsessed and paranoid about it. So I'd try to keep a limit on it." A03	"I think that would be interesting. I don't know that I would want it because I want keep track of who all has the information, but **just to kind of see, 'Oh, this company is using this and they're getting this out of it,**' kind of thing. Just out of curiosity, not out of need to know. [I would check] very randomly and probably pretty rarely." CG20
Low desire for transparency	—[b]	—

[a]CP: computer perception.

[b]Not applicable

XSL•FO
RenderX

Additionally, participants (6 caregivers and 2 adolescents) also desired reassurance through transparency of data security processes. For example, specific reassurances included wanting transparency around "how they were going to protect the information" (CG17), along with "how safe and secure the information is" (CG10) and also "some assurance [that] it's no longer collecting data" when it is physically turned off (CG17). Other participants (3 caregivers and 3 adolescents) shared that they would feel "weirded out" (A07), "upset" (CG09), "disappointed" (CG16), "angry" (CG07), and "violated" (A16) if shown transactions with their data that they did not explicitly consent to.

Other participants (n=4) expressed that transparency without the ability to intervene would be frustrating or distressing, noting they would "rather not know" if no action were possible (A03). As one caregiver put it, "I don't think there's any point… when you know something [that] you can't do anything about" (CG03). Another adolescent participant reasoned that "[they'd] rather not know about it" if they could not act on the knowledge (A03).

Discussion

Participant Perspectives Toward Current Data Protections

A majority of adolescents and caregivers view CP data, that is, data collected through wearable digital sensors, as sensitive information, characterizing clinical assessments using CP as "invasive" (a term used by multiple participants). Many felt that such data could enable insights into patients' emotions and overall well-being. This is, of course, the clinical rationale for collecting these data; however, many respondents expressed privacy concerns, particularly around the potential for stigma or discrimination in cases of unintended or unwanted disclosure, given that participation was connected to their mental health condition. Some respondents stated that CP data thus deserve as much as or greater protection than other personal health data, which should be kept private and deidentified before sharing, as previous studies have found. Consistent with other research findings [50-52], participants from our study said they would feel distressed upon discovering that their data were being used in ways they did not explicitly consent to, especially for commercial purposes. Their concerns are broadly echoed in the literature around CP technologies [53-55], as well as for nonhealth-related data [31,56-59].

Importantly, however, most participants believe that CP data already receives high levels of protection and security. Many stated they highly trust their clinicians and hospitals to protect CP data once collected and expressed confidence that CP data fall under HIPAA protections as with other medical information. Others said they would seek "reassurance" that this is the case, revealing that they consider HIPAA protection to be the default. Respondents were generally unaware that many, if not most, forms of CP data are not classified as protected health information (PHI). CP data often falls outside traditional health privacy protections for two reasons. First, many tools are developed by companies that are not covered health care entities under HIPAA. The data generated by these technology products is instead subject to other data protection laws at the federal level (eg, Gramm-Leach-Bliley Act, Federal Trade Commission Act, and Family Educational Rights and Privacy Act) and state level (eg, California Consumer Privacy Act, California Privacy Rights Act, and Virginia Consumer Data Protection Act). However, critics [60,61] have characterized these as a "patchwork" of data protections that are too specific to certain data types or entities or contain various loopholes that create data protection gaps for citizens, especially for consumer-facing health apps and wearables.

Second, many of these data types (eg, GPS, accelerometry, call logs, message logs, or content) are not directly related to a patient's health or diagnostic status, even if they may in some cases be collected in service of clinical objectives. As other scholars [18] have noted, these forms of NTH data occupy an ambiguous or gray area in health data protections, even more so when they are uncoupled from the 18 identifiers that define HIPAA-protected PHI, as HIPAA does not cover deidentified information. This gap in legal protections for deidentified data becomes increasingly problematic as the ability to reidentify data grows, or when certain types of data, such as facial expressions and voice, are increasingly hard to deidentify. These concerns around privacy have been widely debated in the context of insurance, genomics, and other domains of data governance [62]. A burgeoning industry devoted to reidentifying personal data using machine learning technologies (eg, for marketing purposes or other forms of profiling) increasingly enables unwanted disclosure of personal data. The potential for reidentification is especially high for NTH that is subject to data exchange decisions that are outside of patients' (and even many physicians' and researchers') control.

An important takeaway from our findings in this context is that most patients and caregivers erroneously believe that HIPAA protections guard them from unwanted disclosure of their PHI and may not recognize the limitations of existing data protection frameworks [63]. Importantly, participants rarely differentiated between research and clinical data-sharing frameworks, despite these contexts being governed by different standards for consent, oversight, and commercial use. This lack of differentiation reflects the reality that CP data are often collected and reused across translational pipelines in ways that obscure whether data are governed by research ethics approvals, clinical privacy regimes, or commercial data practices. This ambiguity contributes to epistemic vulnerability and complicates their ability to make informed decisions about whether to consent to the use of CP technologies in clinical care or research. When individuals later learn that protections are weaker or more fragmented than assumed, this mismatch can undermine trust, reduce willingness to share data, and negatively affect adoption and sustained engagement. These findings echo broader public concerns about digital privacy beyond health care, where individuals often misunderstand the extent of their data protections [64-66]. Similar trends exist in consumer data ecosystems, particularly in digital marketing, where data collection occurs without explicit user awareness [30]. To bridge the gap between perceived and actual protections, it is critical to educate individuals about the limitations of data protections across these domains and the circumstances under which their

XSL•FO
RenderX

data may be used without their consent. Education can help to promote greater reflection about potential threats to the security and privacy of the information that individuals agree to have collected and managed. The use of decision aids and other decision support approaches (eg, shared decision making) offers frameworks to ensure patients and caregivers receive sufficient risk/benefit information to engage in informed consent. However, as we discuss below (see "Clearly Communicating (and Knowing) Risk"), certain factors limit the efficacy of education as a primary approach to achieving informed consent.

Challenges to Patient Education and Informed Consent About CP Data Collection

Clearly Communicating (and Knowing) Risk

Several factors challenge the feasibility of effectively educating participants about health data protection and data-sharing. One major challenge is the elusive nature of data sharing. Hospitals, research institutes, data brokers, electronic health record companies, and other entities that comprise the HIE market often engage in transactions that are not made public or fully transparent to individuals who are the subjects of that data [67]. This lack of transparency endures despite findings, including those from the present study, that most individuals desire greater visibility into the nature and extent of secondary data sharing. Similar demands for transparency and accountability have been observed in nonhealth contexts, such as media personalization and digital marketing [33]. Other studies likewise suggest that people—especially young people—want an active, participatory role in decision-making about data sharing [50]. While most (deidentified) data exchanges are permitted under HIPAA, their notorious opacity (regarding with whom, why, how often, and for what profit) complicates the ability of researchers and clinicians to adequately inform participants about the risks associated with participating in research studies that capture sensitive data. Often, clinicians and researchers themselves are unaware of the nature and extent of potential risk in ways that would enable them to clearly communicate them to research participants, as required by most institutional review boards. Greater empirical insights are urgently needed into the impacts of collecting these data in practice. For example, what percentage of these data are deidentified and exchanged only to be reidentified? What percentage of these reidentified data, in turn, lead to patient harms, and what kinds of harm (eg, dignity-related, financial, physical, emotional, social, political, etc)? Answering such questions requires forensic analysis that traces data from its source as it moves across the HIE landscape. Such an endeavor could be facilitated by various technologies that enable transparent tracking of provenance and transaction histories, as we have argued elsewhere [68,69].

Currently, however, HIE transactions remain critically understudied. While new federal legislation has been proposed (eg, American Privacy Rights Act, American Data Privacy Protection Act, and Consumer Online Privacy Rights Act) to address consumer privacy concerns, to ensure greater control over personal information, and to establish robust enforcement mechanisms to hold companies accountable for data sharing practices, federal funding for empirical research into the nature and extent of secondary data exchange is still lacking. Critics argue that HIE remains largely unaddressed and unregulated due to fragmented state and federal laws [70,71], lobbying by powerful electronic health record companies seeking to maintain market dominance, and fears that regulation may stifle innovation in bioinformatics. The result is an enduring unawareness among most individuals, including experts, about what happens to their health (and other types of) data once collected.

One remedy might be to mandate greater visibility into health data transactions, such that patients and research participants can follow who sees their data, how often, and for what purposes (eg, data sharing between researchers, data exchange for profit, etc). We have described elsewhere some computational avenues (eg, distributed ledger technologies) for enhancing the transparency and auditability of data exchanges [68,69]. However, even if these mechanisms are put in place, our findings suggest that many participants have mixed feelings about being able to see transactions involving their health information; some express enthusiasm, while others point out that such transparency would only frustrate them if unaccompanied by the ability to intervene in or control those transactions. To date, limited mechanisms exist to empower such control. Approaches such as dynamic consent, which enable participants to continuously monitor, manage, and update their consent preferences over time, may be appealing to individuals who want to actively track transactions involving their data (in this study, three-quarters of our participants). Notably, dynamic consent mechanisms are still not common in clinical research outside of certain contexts, for example, genomic or long-term biobanking, where data may be repurposed numerous times for multiple research agendas. In most areas of research, traditional forms of consent prevail, particularly "broad" consent where patients consent once at the start of a study or treatment and cannot modify stated permissions thereafter. Despite affording participants almost no insight into secondary use applications, broad consent enjoys widespread support from the scientific community [72,73], as it can help to facilitate data sharing and drive scientific progress. However, it requires high trust in data controllers to make decisions that align with data subjects' preferences. Our findings show that trust is indeed high among most research participants, but perhaps it should not be, not necessarily because data controllers do not act in good faith when stewarding participant data, but because the current legal landscape offers limited protection for health and NTH data that may enter into a cascade of secondary data exchanges that are difficult for primary data controllers to forecast, trace, or control.

Avoiding Undue Aversion and Consent Refusal

Determining the appropriate level of risk information for clinician researchers to communicate to participants is not only challenged by a lack of clarity about the extent and nature of harms but also by researchers' awareness that highlighting such (uncertain) risks may generate risk aversion and lead to consent refusal. The inability to assuage potential participants' concerns is exacerbated by the difficulty of auditing or assessing when unwanted disclosure and associated harms do occur. Individuals may be reluctant to consent to secondary uses of their data if they fear potential harms will go unnoticed and continue to

XSL•FO
RenderX

perpetuate indefinitely. Balancing disclosure of potential risks while encouraging participation remains a complex challenge. Improved ethical frameworks are needed to communicate uncertain risk without "scaring off" potential research participants who may benefit from integrating CP into their clinical care or contribute significantly to scientific discovery and research. For example, in the context of genomic sequencing, Koplin et al [74] argue that appropriately informed consent may be a fitting alternative to fully informed consent in cases where risks are uncertain. Koplin et al [74] argue that focusing consent procedures on satisfying four morally important goals that informed consent is meant to achieve, including (1) respect for personal autonomy, (2) promotion of autonomy and (3) well - being, and (4) preservation of trust in medicine, may help individuals to make decisions that align with their values, even without a full picture of potential risks. Further work is needed to explore how each of these goals may be achieved in the context of participant consent to CP data collection. For example, what mechanisms may be put in place to ensure autonomy over decisions to share data within and/or beyond a participant's clinical team? What information is needed to clearly convey the potential benefit of CP data collection for participants' well-being, especially given that most CP technologies remain investigational? These questions deserve greater attention as CP and mHealth technologies continue to gain traction in health care and research.

Pathways Forward

Given these uncertainties, it remains unclear how clinicians and researchers should proceed with informed consent. On the one hand, they are ethically obligated to communicate risks associated with the use of CP in research or care, but they themselves lack sufficient insights into these risks. Further, they must be careful not to over- or understate risks, balancing awareness against the potential to affect individuals' willingness to participate in research or integrate technologies that may ultimately benefit patients. Greater visibility into the nature and extent of transactions may only partially solve the problem of informed consent, as it may lead to distress if accompanied by opportunities for intervention or control. However, data controllers (ranging from researchers, institutions, companies, and data brokers) currently have little incentive to offer such control if legal mandates are not in place (and there is no indication they will be anytime soon). A key arena to watch for developments around such mandates is the newly created federal Trusted Exchange Framework and Common Agreement (TEFCA; 2024), which aims to facilitate nationwide, query-based data sharing and to improve patient access to health information as mandated by the 21st Century Cures Act (2016) [75]. Mandel et al [76] recently argued that three principles should be put in place by the time TEFCA launches, including that patients can (1) query for data about themselves, (2) know when their data are queried and shared, and (3) configure what is shared about them. TEFCA launched in 2024, and so far, while it does appear to support patient access to their own health information, TEFCA does neither specifically mandate that patients be notified each time their data are queried or shared, nor do all constituent health information networks offer the capacity for patient-configurable data-sharing controls. Research to identify and validate procedures and technologies that may provide the necessary infrastructure for such controls should be prioritized at the federal level to evolve alongside emerging data privacy legislation.

Limitations

This study has several limitations. First, our qualitative sample was drawn from adolescents and caregivers already enrolled in a research study embedded within a clinical research setting. Participants may thus have higher baseline trust in clinicians and institutions than individuals engaging with direct-to-consumer digital health tools, potentially limiting generalizability to other populations or contexts. Second, participants' perspectives reflect experiences at the intersection of research and clinical care; while this overlap is increasingly common for CP technologies, findings may not fully capture attitudes in purely clinical or purely commercial settings. Third, although the primer video was designed to ensure a shared baseline understanding of CP technologies, it may have influenced participants' framing of risks and benefits. Finally, as with all qualitative research, our findings are not intended to quantify prevalence or establish causal relationships, but rather to identify salient themes and ethical concerns that warrant further empirical investigation.

Conclusions

Our findings highlight the need for clearer communication, stronger data governance, and consent models that better reflect how digital health data are used in practice. These data hold strong potential to provide ecologically valid insights into patients' health conditions beyond the clinic, but also pose significant privacy risks that remain understudied. While clinicians and researchers have a unique responsibility to safeguard participant and patient data, they cannot effectively do so without a better understanding about the nature and extent of these risks or how to effectively mitigate them. To promote greater visibility, understanding, and trust (when justified) in digital tools that have strong promise to promote patient health, we advocate for five key strategies, including: (1) educating patients on the limitations of existing data protections; (2) conducting targeted research, including forensic analyses, into secondary data exchanges and identifying privacy breaches or reidentification risks; (3) enacting regulations that mandate greater transparency in health data transactions; (4) implementing computational mechanisms, such as distributed ledger technologies, to enhance data traceability and auditability; and (5) adopting dynamic consent models that allow patients to continuously manage and update their consent preferences. These strategies may help to better balance the promise of CP technologies and insights from wearables and mHealth apps with the ethical imperative to protect patient privacy, fostering greater trust and responsible innovation in digital health care.

XSL•FO
RenderX

Acknowledgments

We would like to thank the patients and caregivers who took the time to interview with us and share their valuable perspectives. Additionally, we would like to thank the research coordinators from our “sister” study who helped recruit patients, including Rebecca Greenberg, BS, Jessica Foy, BS, and Yuen Yu, BSSc. No generative artificial intelligence was used in any portion of this paper's writing.

Funding

This study was supported by National Center for Advancing Translational Sciences (1R01TR004243) and National Institute of Mental Health (3R01MH125958). The views expressed here are solely those of the authors and do not necessarily reflect the official policies of the National Institutes of Health (NIH) or the US government.

Data Availability

The datasets presented in this paper are not readily available because full datasets must remain unavailable to ensure deidentification of interview participants. Requests to access the datasets should be directed to KK-Q (email: kristin.kostick@bcm.edu).

Authors' Contributions

Conceptualization: CD, BT, EAS, KK-Q, JDH
Formal analysis: CD, AS, BT, KK-Q, MEH
Funding acquisition: EAS, JDH, KK-Q
Investigation: CD, AS, MEH, KK-Q
Project administration: CD, AS, MEH, KK-Q
Resources: KK-Q
Supervision: EAS, JDH, JB-B, KK-Q
Visualization: CD, BT, KK-Q
Writing – original draft: CD, KK-Q
Writing – review & editing: CD, AS, MEH, BT, EAS, JDH, JB-B, KK-Q

Conflicts of Interest

EAS reports receiving research funding to his institution from the Ream Foundation, the International OCD Foundation, Wellcome Trust, and National Institutes of Health. He receives direct funding from the International OCD Foundation as well as Mental Health Network and Training Institute for providing trainings on treating obsessive-compulsive disorder with psychotherapy. He was formerly a consultant for Brainsway and Biohaven Pharmaceuticals in the past 12 months. He owns stock worth less than US $5000 in NView (NVIDIA Corporation) and receives royalties from OCD Scales LLC (for distribution of the Y-BOCS and CY-BOCS). He receives book royalties from Elsevier, Wiley, Oxford, American Psychological Association, Guildford, Springer, Routledge, and Jessica Kingsley. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Adolescent interview guide.
[DOCX File, 42 KB - mhealth_v14i1e78788_app1.docx]

Multimedia Appendix 2
Caregiver interview guide.
[DOCX File, 28 KB - mhealth_v14i1e78788_app2.docx]

Multimedia Appendix 3
Explainer video of the ethics of perceptive computing in health care study (EPICH).
[MP4 File, 10824 KB - mhealth_v14i1e78788_app3.mp4]

Checklist 1
COREQ Checklist.
[PDF File, 406 KB - mhealth_v14i1e78788_app4.pdf]

References

1. Bufano P, Laurino M, Said S, Tognetti A, Menicucci D. Digital phenotyping for monitoring mental disorders: systematic review. J Med Internet Res 2023 Dec 13;25:e46778. [doi: 10.2196/46778] [Medline: 38090800]

XSL•FO
RenderX

2. Oudin A, Maatoug R, Bourla A, et al. Digital phenotyping: data-driven psychiatry to redefine mental health. J Med Internet Res 2023 Oct 4;25:e44502. [doi: 10.2196/44502] [Medline: 37792430]
3. Perez-Pozuelo I, Spathis D, Clifton EAD, Mascolo C. Wearables, smartphones, and artificial intelligence for digital phenotyping and health. In: Syed-Abdul S, Zhu X, Fernandez-Luque L, editors. Digit Health: Elsevier; 2021:33-54. [doi: 10.1016/B978-0-12-820077-3.00003-1]
4. Torous J, Wisniewski H, Bird B, et al. Creating a digital health smartphone app and digital phenotyping platform for mental health and diverse healthcare needs: an interdisciplinary and collaborative approach. J Technol Behav Sci 2019 Jun;4(2):73-85. [doi: 10.1007/s41347-019-00095-w]
5. Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 2016 Jun;41(7):1691-1696. [doi: 10.1038/npp.2016.7] [Medline: 26818126]
6. Assabumrungrat R, Sangnark S, Charoenpattarawut T, et al. Ubiquitous affective computing: a review. IEEE Sensors J 2021;22(3):1867-1881. [doi: 10.1109/JSEN.2021.3138269]
7. Picard RW. Affective Computing: MIT press; 2000.
8. Monosov IE, Zimmermann J, Frank MJ, Mathis MW, Baker JT. Ethological computational psychiatry: challenges and opportunities. Curr Opin Neurobiol 2024 Jun;86:102881. [doi: 10.1016/j.conb.2024.102881] [Medline: 38696972]
9. Egnor SER, Branson K. Computational analysis of behavior. Annu Rev Neurosci 2016 Jul 8;39(1):217-236. [doi: 10.1146/annurev-neuro-070815-013845] [Medline: 27090952]
10. Sigfstead S, Jiang R, Avram R, Davies B, Krahn AD, Cheung CC. Applying artificial intelligence for phenotyping of inherited arrhythmia syndromes. Can J Cardiol 2024 Oct;40(10):1841-1851. [doi: 10.1016/j.cjca.2024.04.014] [Medline: 38670456]
11. Collins JT, Walsh DA, Gladman JRF, et al. The difficulties of managing pain in people living with frailty: the potential for digital phenotyping. Drugs Aging 2024 Mar;41(3):199-208. [doi: 10.1007/s40266-024-01101-4] [Medline: 38401025]
12. Bilal AM, Fransson E, Bränn E, et al. Predicting perinatal health outcomes using smartphone-based digital phenotyping and machine learning in a prospective Swedish cohort (Mom2B): study protocol. BMJ Open 2022 Apr 27;12(4):e059033. [doi: 10.1136/bmjopen-2021-059033] [Medline: 35477874]
13. Jaimini U, Thirunarayan K, Kalra M, Venkataraman R, Kadariya D, Sheth A. "How is my child's asthma?" Digital phenotype and actionable insights for pediatric asthma. JMIR Pediatr Parent 2018;1(2):e11988. [doi: 10.2196/11988] [Medline: 31008446]
14. Sivakumar CLV, Mone V, Abdumukhtor R. Addressing privacy concerns with wearable health monitoring technology. WIREs Data Min Knowl Discovery 2024 Jun;14(3):e1535. [doi: 10.1002/widm.1535]
15. Al Meslamani AZ. Gaps in digital health policies: an insight into the current landscape. J Med Econ 2023 Dec 31;26(1):1266-1268. [doi: 10.1080/13696998.2023.2266955]
16. Nahra KJ, Corbin BA. Digital health regulatory gaps in the United States. Compliance Elliance J 2018;4(2):21-34 [FREE Full text]
17. Suver C, Kuwana E. mHealth wearables and smartphone health tracking apps: a changing privacy landscape. Inf Serv Use 2021;41(1-2):71-79. [doi: 10.3233/ISU-210114]
18. Price WN, Cohen IG. Privacy in the age of medical big data. Nat Med 2019 Jan;25(1):37-43. [doi: 10.1038/s41591-018-0272-7] [Medline: 30617331]
19. Ostherr K, Borodina S, Bracken RC, Lotterman C, Storer E, Williams B. Trust and privacy in the context of user-generated health data. Big Data Soc 2017 Jun;4(1):1. [doi: 10.1177/2053951717704673]
20. Catapan SDC, Sazon H, Zheng S, et al. A systematic review of consumers' and healthcare professionals' trust in digital healthcare. NPJ Digit Med 2025 Feb 21;8(1):115. [doi: 10.1038/s41746-025-01510-8] [Medline: 39984678]
21. Perez-Pozuelo I, Spathis D, Gifford-Moore J, Morley J, Cowls J. Digital phenotyping and sensitive health data: implications for data governance. J Am Med Inform Assoc 2021 Aug 13;28(9):2002-2008. [doi: 10.1093/jamia/ocab012] [Medline: 33647989]
22. Deruelle T, Kalouguina V, Trein P, Wagner J. Designing privacy in personalized health: an empirical analysis. Big Data Soc 2023 Jan;10(1):20539517231158636. [doi: 10.1177/20539517231158636]
23. Alaattinoğlu D. Rethinking explicit consent and intimate data: the case of menstruapps. Fem Leg Stud 2022 Jul;30(2):157-179. [doi: 10.1007/s10691-021-09486-y]
24. Kollnig K, Binns R, Van Kleek M, et al. Before and after GDPR: tracking in mobile apps. Internet Policy Rev 2021 Dec 21;10(4):4-30. [doi: 10.14763/2021.4.1611]
25. Reijneveld MD. Quantified self, freedom, and the GDPR. SCRIPTed 2017 Dec;14(2):285-325. [doi: 10.2966/scrip.140217.285]
26. Rezaee R, Khashayar M, Saeedinezhad S, Nasiri M, Zare S. Critical criteria and countermeasures for mobile health developers to ensure mobile health privacy and security: mixed methods study. JMIR mHealth uHealth 2023 Mar 2;11(1):e39055. [doi: 10.2196/39055] [Medline: 36862494]
27. Hartman T, Kennedy H, Steedman R, Jones R. Public perceptions of good data management: findings from a UK-based survey. Big Data Soc 2020 Jan;7(1):205395172093561. [doi: 10.1177/2053951720935616]

XSL•FO
RenderX

28. Zadushlivy N, Biviji R, Williams KS. Exploration of reproductive health apps' data privacy policies and the risks posed to users: qualitative content analysis. J Med Internet Res 2025 Mar 5;27(1):e51517. [doi: 10.2196/51517] [Medline: 40053713]
29. Aitken M, de St Jorre J, Pagliari C, Jepson R, Cunningham-Burley S. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. BMC Med Ethics 2016 Nov 10;17(1):73. [doi: 10.1186/s12910-016-0153-x] [Medline: 27832780]
30. McStay A. Emotional AI, soft biometrics and the surveillance of emotional life: an unusual consensus on privacy. Big Data Soc 2020 Jan;7(1):205395172090438. [doi: 10.1177/2053951720904386]
31. Madden M. Public perceptions of privacy and security in the post-Snowden era. Pew Research Center. 2014 Nov 12. URL: https://www.pewresearch.org/internet/2014/11/12/public-privacy-perceptions/ [accessed 2026-06-11]
32. Wei D, Gao P, Zhai Y. The impact of data control and delayed discounting on the public's willingness to share different types of health care data: empirical study. JMIR Med Inform 2025 Jan 22;13(1):e66444. [doi: 10.2196/66444] [Medline: 39864171]
33. Wong YN, Jones R, Das R, Jackson P. Conditional trust: citizens' council on data-driven media personalisation and public expectations of transparency and accountability. Big Data Soc 2023 Jul;10(2):2. [doi: 10.1177/20539517231184892]
34. Kozyreva A, Lorenz-Spreen P, Hertwig R, Lewandowsky S, Herzog SM. Public attitudes towards algorithmic personalization and use of personal data online: evidence from Germany, Great Britain, and the United States. Humanit Soc Sci Commun 2021;8(1):1-11. [doi: 10.1057/s41599-021-00787-w]
35. Madden M, Rainie L. Americans' views about data collection and security. Pew Research Center. 2015 May 20. URL: https://www.pewresearch.org/internet/2015/05/20/americans-views-about-data-collection-and-security/ [accessed 2026-06-11]
36. Ryan K, Hogg J, Kasun M, Kim JP. Users' perceptions and trust in AI in direct-to-consumer mHealth: qualitative interview study. JMIR mHealth uHealth 2025 May 20;13(1):e64715. [doi: 10.2196/64715] [Medline: 40392584]
37. Alhammad N, Alajlani M, Abd-Alrazaq A, Epiphaniou G, Arvanitis T. Patients' perspectives on the data confidentiality, privacy, and security of mHealth apps: systematic review. J Med Internet Res 2024 May 31;26(1):e50715. [doi: 10.2196/50715] [Medline: 38820572]
38. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. JMIR mHealth uHealth 2019 Apr 16;7(4):e11223. [doi: 10.2196/11223] [Medline: 30990458]
39. Higgins B, Jones L, Devraj K, Kilduff C, Moosajee M. "It would help people to help me": acceptability of digital phenotyping among young people with visual impairment and their families. Digit Health 2024;10:20552076231220804. [doi: 10.1177/20552076231220804] [Medline: 38188864]
40. Nestor BA, Chimoff J, Koike C, et al. Adolescent and parent perspectives on digital phenotyping in youths with chronic pain: cross-sectional mixed methods survey study. J Med Internet Res 2024 Jan 11;26:e47781. [doi: 10.2196/47781] [Medline: 38206665]
41. Raugh IM, James SH, Gonzalez CM, et al. Digital phenotyping adherence, feasibility, and tolerability in outpatients with schizophrenia. J Psychiatr Res 2021 Jun;138:436-443. [doi: 10.1016/j.jpsychires.2021.04.022] [Medline: 33964681]
42. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
43. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572. [doi: 10.2196/17572] [Medline: 32348255]
44. Steare T, O'Hanlon P, Eskinazi M, et al. Smartphone-delivered self-management for first-episode psychosis: the ARIES feasibility randomised controlled trial. BMJ Open 2020 Aug 26;10(8):e034927. [doi: 10.1136/bmjopen-2019-034927] [Medline: 32847902]
45. Pratap A, Renn BN, Volponi J, et al. Using mobile apps to assess and treat depression in hispanic and latino populations: fully remote randomized clinical trial. J Med Internet Res 2018 Aug 9;20(8):e10130. [doi: 10.2196/10130] [Medline: 30093372]
46. Senapati B, Bin Naeem A, Khan TA, et al. A Study on Web User's Attitude and Knowledge Towards Data Security and Privacy Issues of Web Browser Extensions. In: 2024 4th International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME: IEEE; 2024:1-8. [doi: 10.1109/ICECCME62383.2024.10796625]
47. Arias-Cabarcos P, Khalili S, Strufe T. Surprised, shocked, worried'}: user reactions to Facebook data collection from third parties. PoPETs 2023;2023(1):384-399. [doi: 10.56553/popets-2023-0023]
48. Ur B, Leon PG, Cranor LF, Shay R, Wang Y. Smart, useful, scary, creepy: perceptions of online behavioral advertising. Presented at: Proceedings of the Eighth Symposium on Usable Privacy and Security, SOUPS 2012; Jul 11-13, 2012; New York, NY p. 1-15. [doi: 10.1145/2335356.2335362]
49. Boyatzis RE. Transforming Qualitative Information: Thematic Analysis and Code Development: Sage Publications, Inc; 1998.
50. Sieberts SK, Marten C, Bampton E, et al. Young people's data governance preferences for their mental health data: MindKind Study findings from India, South Africa, and the United Kingdom. PLoS One 2023;18(4):e0279857. [doi: 10.1371/journal.pone.0279857] [Medline: 37074995]

XSL•FO
RenderX

51. Cumyn A, Ménard JF, Barton A, Dault R, Lévesque F, Ethier JF. Patients' and members of the public's wishes regarding transparency in the context of secondary use of health data: scoping review. J Med Internet Res 2023 Apr 13;25:e45002. [doi: 10.2196/45002] [Medline: 37052967]
52. Cumyn A, Dault R, Barton A, Cloutier AM, Ethier JF. Citizens, research ethics committee members and researchers' attitude toward information and consent for the secondary use of health data: implications for research within learning health systems. J Empir Res Hum Res Ethics 2021 Jul;16(3):165-178. [doi: 10.1177/1556264621992214] [Medline: 33710932]
53. Martinez-Martin N, Kreitmair K. Ethical issues for direct-to-consumer digital psychotherapy apps: addressing accountability, data protection, and consent. JMIR Ment Health 2018 Apr 23;5(2):e32. [doi: 10.2196/mental.9423] [Medline: 29685865]
54. Sonig A, Deeney C, Hurley ME, et al. What patients and caregivers want to know when consenting to the use of digital behavioral markers. NPP Digit Psychiatry Neurosci 2024 Dec 6;2(1):19. [doi: 10.1038/s44277-024-00022-9] [Medline: 41361531]
55. Shen FX, Silverman BC, Monette P, Kimble S, Rauch SL, Baker JT. An ethics checklist for digital health research in psychiatry: viewpoint. J Med Internet Res 2022 Feb 9;24(2):e31146. [doi: 10.2196/31146] [Medline: 35138261]
56. Hemphill L, Schöpke-Gonzalez A, Panda A. Comparative sensitivity of social media data and their acceptable use in research. Sci Data 2022 Oct 22;9(1):643. [doi: 10.1038/s41597-022-01773-w] [Medline: 36273208]
57. Katsanis SH, Claes P, Doerr M, et al. U.S. adult perspectives on facial images, DNA, and other biometrics. IEEE Trans Technol Soc 2022 Mar;3(1):9-15. [doi: 10.1109/tts.2021.3120317] [Medline: 35360665]
58. Katsanis SH, Claes P, Doerr M, et al. A survey of U.S. public perspectives on facial recognition technology and facial imaging data practices in health and research contexts. PLOS ONE 2021 Oct;16(10):e0257923. [doi: 10.1371/journal.pone.0257923]
59. Milne GR, Pettinico G, Hajjat FM, Markos E. Information sensitivity typology: mapping the degree and type of risk consumers perceive in personal data sharing. J Consum Aff 2017 Mar;51(1):133-161. [doi: 10.1111/joca.12111]
60. Theodos K, Sittig S. Health information privacy laws in the digital age: HIPAA doesn't apply. Perspect Health Inf Manage 2021;18(Winter):1l. [Medline: 33633522]
61. Mandl KD, Perakslis ED. HIPAA and the leak of "Deidentified" EHR data. N Engl J Med 2021 Jun 10;384(23):2171-2173. [doi: 10.1056/NEJMp2102616] [Medline: 34110112]
62. Blasimme A, Vayena E, Van Hoyweghen I. Big data, precision medicine and private insurance: a delicate balancing act. Big Data Soc 2019 Jan;6(1):1. [doi: 10.1177/2053951719830111]
63. Cilliers L. Wearable devices in healthcare: privacy and information security issues. Health Inf Manag J 2020 May;49(2-3):150-156. [doi: 10.1177/1833358319851684]
64. Mcclain C, Faverio M, Anderson M, Park E. How Americans view data privacy. Pew Research Center. 2023 Oct 18. URL: https://www.pewresearch.org/internet/2023/10/18/how-americans-view-data-privacy/ [accessed 2026-06-11]
65. Korunovska J, Kamleitner B, Spiekermann S. The challenges and impact of privacy policy comprehension. arXiv. Preprint posted online on May 18, 2020. [doi: 10.48550/arXiv.2005.08967]
66. Schomakers EM, Lidynia C, Müllmann D, et al. Putting privacy into perspective -- comparing technical, legal, and users' view of data sensitivity. arXiv. Preprint posted online on Nov 15, 2019. [doi: 10.48550/arXiv.1911.06569]
67. Crain M. The limits of transparency: data brokers and commodification. New Media Soc 2018 Jan;20(1):88-104. [doi: 10.1177/1461444816657096]
68. Deeney CR, Kostick-Quenet K. Computational ethics tools to audit corporate self-governance in data processing. Am J Bioeth 2023 Nov;23(11):42-44. [doi: 10.1080/15265161.2023.2256258] [Medline: 37879005]
69. Kostick-Quenet K, Mandl KD, Minssen T, et al. How NFTs could transform health information exchange. Science 2022 Feb 4;375(6580):500-502. [doi: 10.1126/science.abm2004] [Medline: 35113709]
70. Vest JR, Gamm LD. Health information exchange: persistent challenges and new strategies. J Am Med Inform Assoc 2010;17(3):288-294. [doi: 10.1136/jamia.2010.003673] [Medline: 20442146]
71. Price WN, Kaminski ME, Minssen T, Spector-Bagdady K. Shadow health records meet new data privacy laws How will research respond to a changing regulatory space. Science 2019;363(6426):448-450. [doi: 10.1126/science.aav5133]
72. Final NIH policy for data management and sharing. National Institutes of Health. 2020 Oct 29. URL: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html [accessed 2026-06-11]
73. Rai AK. Risk regulation and innovation: the case of rights-encumbered biomedical data silos. Notre Dame Law Rev 2017 Apr 21;92(4):101-129. [doi: 10.2139/ssrn.2945595]
74. Koplin JJ, Gyngell C, Savulescu J, Vears DF. Moving from "fully" to "appropriately" informed consent in genomics: the PROMICE Framework. Bioethics 2022 Jul;36(6):655-665. [doi: 10.1111/bioe.13027] [Medline: 35390218]
75. 21st Century Cures Act. Congress.gov. 2016. URL: https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf [accessed 2026-06-22]
76. Mandel JC, Pollak JP, Mandl KD. The patient role in a federal national-scale health information exchange. J Med Internet Res 2022 Nov 4;24(11):e41750. [doi: 10.2196/41750] [Medline: 36331535]

XSL•FO
RenderX

Abbreviations

CP: computer perception
GDPR: General Data Protection Regulation
HIE: health information exchange
HIPAA: Health Insurance Portability and Accountability Act
mHealth : mobile health
NTH: nontraditional health
PHI: personal health information
RA: research assistant
TEFCA: Trusted Exchange Framework and Common Agreement

Edited by A Stone, L Buis; submitted 09.Jun.2025; peer-reviewed by I Schiering, SDC Catapan; revised version received 08.May.2026; accepted 11.May.2026; published 08.Jul.2026.

Please cite as:
Deeney C, Sonig A, Hurley ME, Tunç B, Storch EA, Herrington JD, Blumenthal-Barby J, Kostick-Quenet K
Perceived Sensitivity of Sensor-Based Digital Health Data: Qualitative Interview Study
JMIR Mhealth Uhealth 2026;14:e78788
URL: https://mhealth.jmir.org/2026/1/e78788
doi:10.2196/78788

© Christine Deeney, Anika Sonig, Meghan E Hurley, Birkan Tunç, Eric A Storch, John D Herrington, Jennifer Blumenthal-Barby, Kristin Kostick-Quenet. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 8.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Functionality Review of Mobile Apps for the Tracking and Self-Management of Fatigue: Systematic Search in App Stores and Content Analysis

Amr Diouf Abdulla[1], BSc; Corina Sas[1], PhD; Gavin Doherty[2], DPhil

[1]School of Computing and Communications, Lancaster University, Bailrigg, Lancaster, United Kingdom

[2]School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland

Corresponding Author:
Amr Diouf Abdulla, BSc
School of Computing and Communications, Lancaster University, Bailrigg, Lancaster, United Kingdom

Abstract

Background: Fatigue and chronic fatigue syndrome (CFS) have a considerable impact on quality of life, thus motivating people to develop skills for better management of their fatigue. While the number of commercial apps in this domain has increased, there has been limited exploration of their functionalities.

Objective: This paper aims to address this research gap through a functionality review of 17 top-rated iOS and Android apps for fatigue, with the aim to articulate design implications for technologies focused on supporting the management of fatigue.

Methods: We conducted a systematic search on the 2 most common app marketplaces, which resulted in the initial identification of 427 Apple apps and 1218 Google apps. From these, 17 apps were selected for review after applying a screening process to shortlist the top-rated apps. The functionalities of these apps were then coded through a week-long usage of each app for an expert evaluation leveraging authors' human-computer interaction (HCI) expertise. We looked for functionalities such as tracking and visualization seen in previous research on functionality reviews, in addition to interventional functionalities, which were informed by research on fatigue.

Results: Findings reveal the prevalence of functionalities for tracking fatigue (8/17, 47%), related symptoms (8/17, 47%), for visualizing tracked content (10/17, 59%), for assessing the user's condition (2/17, 12%), and for providing interventions for the management of fatigue (12/17, 71%). Functionalities providing interventions for self-management of fatigue are surprisingly limited, with the most relevant ones including pacing (2/17, 12%) alongside energy estimation (2/17, 12%).

Conclusions: The top-ranked apps for fatigue in the major marketplaces support 3 main functionalities under the scope of tracking fatigue along with related data, and visualizing such data, with limited provision of self-management interventions. Drawing from these findings, we articulate implications for the sensitive design of technologies to support the management of fatigue, including supporting hybrid tracking, combined visualizations to support sense-making of fatigue data with related factors, and supporting energy estimates and pacing interventions.

(JMIR Mhealth Uhealth 2026;14:e84755) doi:10.2196/84755

KEYWORDS

chronic fatigue syndrome; fatigue; mobile apps; tracking; self-management; interventions; pacing

Introduction

Background

Overview

Fatigue and particularly chronic fatigue syndrome (CFS) are conditions with a strong impact on quality of life. Crook et al [1] define fatigue as "more profound than being overtired; it is unrelenting exhaustion and a constant state of weariness that reduces a person's energy, motivation, and concentration." Within human-computer interaction (HCI) research on physical health, fatigue has been less explored, although recent work, focused on lived experiences of long COVID, has started to emerge [2]. In parallel to academic research, there has been a growing number of mobile apps targeting fatigue. Mobile apps for physical or mental health, and in particular, those highly rated by their users, are likely to reflect valuable design principles which benefit from being articulated as implications for design. Similar efforts have been made for apps for depression [3], digital well-being [4,5], and anxiety [6], but less so for fatigue. To address this gap, we report a functionality review of 17 top-rated apps for fatigue from the Apple App Store and Google Play Store, together with an analysis of their descriptions on marketplaces. Findings indicate that the main functionalities of these apps are tracking fatigue and its symptoms, visualizing tracked content, assessing symptom

XSL•FO
RenderX

impact on everyday life, and providing interventions. We conclude with 5 design implications for better support of people living with fatigue, which include supporting hybrid tracking to balance the validity of self-reported fatigue with low-burden automatic tracking of symptoms, contextualizing fatigue data with symptoms and associated factors through combined visualizations for sense-making, and supporting energy estimates and pacing interventions.

To support our exploration and understanding of the functionalities, we draw on the limited but emerging area of HCI work on fatigue, comprising the tracking of symptoms, the design of technologies supporting their management, and the emerging area of research that has engaged in the analysis of apps created for this purpose.

HCI Studies Around Technology Usage for Individuals With Chronic Conditions

A key area that we first needed to understand was the behaviors of individuals with CFS surrounding their use of technology to have a better idea of any ramifications that may arise from the condition, and consequently how they can be designed around. Paymal and Homewood [7] conducted a study to identify how CFS impacts technology usage in day-to-day life, in which they make use of their findings to present design recommendations. Their findings include the reduction of technology use on days where participants feel better, the increased usage of technology to help with tasks when symptoms were more noticeable, the avoidance of technology by some participants when their symptoms were more severe, and the usage of apps for tracking and self-management purposes. While the researchers identified how self-management technologies were used by this group, their reasons for turning to these technologies, and their criticisms regarding their usage experience, the researchers' findings could be further supported by detailing the different functionalities offered.

Virtual reality (VR) is also an area that has been the focus of research; Best and Butler [8] investigated the use of VR as a support space for individuals with CFS. They discovered that CFS symptoms would prevent users from engaging with the technology for extended periods of time; however, this led to users improving their energy management to use their limited reserves more efficiently, in addition to optimizing their physical spaces and the technology they use to minimize fatigue triggers. This highlights the interplay between their energy levels and their self-tracking efforts, and is a key area of concern when designing for this group of users.

Davies et al [9] investigated the experiences of patients with chronic fatigue regarding their usage of technology for disease management, and a key area of interest was the participants' usage of apps for self-management. In their study, they set forth design guidelines that are centered around the symptoms that patients want to manage the most, including features that assist users in identifying maximum and minimum energy thresholds, adding activity pacing functionality coupled with routine establishment, providing support for fatigue tracking by exploring automated logging methods to reduce effort expended, and finding ways to make self-management apps more engaging and motivating such as by gamification. Their findings can better inform our evaluation as the guidelines provide a frame of reference to examine the apps against.

Understanding Technologies for Individuals With CFS

Research has also delved into understanding how to support users with energy-limiting conditions (ELCs) through technology. For instance, Børsting and Culén [10] applied a sense-making process for complex domains in order to explore the design of technologies for supporting young individuals with CFS. The approach is a combination of a literature review, mapping techniques, and user research. The results of the user research, which was conducted with CFS experts, led to suggestions targeting self-management, namely, providing interventions discovered by other individuals with CFS, offering reminders for both resting and general purposes, providing functionalities to track and warn users of noise levels, providing social support technologies, and finally, providing the logging of symptoms coupled with tracking technologies suited for CFS.

Another strand of work has focused on technologies for supporting the management of chronic conditions. For instance, Sas et al [2] reported findings from workshops with individuals affected by long COVID, suggesting design implications for tracking and sense-making of fatigue and postexertional malaise (PEM), as well as self-management aimed at PEM prevention. Their findings stressed the importance of lightweight capture of data to prevent symptoms worsening among users, conjointly with reducing the burden of identifying both triggers and the PEM-approaching stage by making use of machine learning (ML) techniques or noninvasive items such as heart rate variability (HRV) tracking, and finally, supporting self-management of fatigue based on the stage of PEM that a user is at.

We also note the study of Homewood and colleagues [11] who co-designed through asynchronous communication with the aim of catering to the needs of individuals with ELCs. The researchers discovered that this method helped participants to only engage with the study when their energy levels were sufficient to partake; in addition, they were able to extract meaningful insights into improving the support for individuals with ELCs. These included designing technologies that improve communication with their social networks, implementing automatic tracking of their fatigue levels and associated symptoms, and using crip approaches in future studies dealing with ELCs.

Studies have also found that if participants' fatigue symptoms reached more severe levels, they may avoid interaction with technology to prevent overwhelming them [7,9,12]. For this reason, designing around the potential fluctuations in their energy levels is paramount [2,9,10,13,14], and one such way this can be achieved is by including automated logging methods, especially in scenarios where users track numerous types of data [2,9]. In the same vein, Mack et al [14] proposed tenets to improve the accessibility of technology, and from these, one tenet called for designs that account for rises in fatigue levels post exertion and consider the consequences of a task on an individual's energy reserves.

XSL•FO
RenderX

The findings detailed in this section can be used to better evaluate the current design space of commercial apps and provide more informed implications for future app designs.

Better Design of Online Resources for Individuals With CFS

Other areas of HCI research targeting chronic conditions have focused on the usage of online resources to support CFS, as in the study of Brigden et al [15]. This study, along with several others, found that social media could be of value to individuals with chronic conditions as a means of discussion and social support surrounding their condition [7,10,15-17]. These studies highlight the importance of observing for social media–based functionalities in the apps to be reviewed.

Another study is by Best and Butler [12], where they sought to identify the physiological effects experienced by individuals with CFS when socializing in a virtual world. Their findings also include the previously mentioned user behavior where they withdraw from technology as their fatigue levels worsen, so as to prevent further deterioration [7,9,12]. As such, the level at which users must interact with the apps could be a useful metric to evaluate the apps and think of further design implications.

HCI Research on Technologies for Long COVID

PEM is also known to be a characteristic of conditions such as long COVID [2]; therefore, the research in this field has value for the area of fatigue management. From the works in this field, we note the study of Homewood [13], who undertook an autoethnography where they used a Fitbit device for activity pacing purposes to manage their long COVID symptoms. This is also one of the studies that highlighted the importance of designing around the potential fluctuations in energy levels [2,9,10,13,14].

Another study of interest is that of Homewood et al [16], who conducted interviews with participants who had long COVID and made use of wearables for disease management, with findings highlighting the need for designs that are tailored to account for individual differences.

In addition, Pater and colleagues [17] carried out a 3-month-long cohort study on patients with long COVID, which involved the collection of both objective and subjective data. The authors emphasized that cognitive symptoms may make data gathering difficult, with participants finding it challenging to enter survey data regularly due to forgetfulness, which also made interviews burdensome [17]. They also posited that gaps in data tracked during their study were caused by participants forgetting to charge their wearable devices.

Finally, Mehdipour and Aharari [18] conducted a literature review along with case studies to analyze the marrying of the fields of AI, HCI, and Internet of Things (IoT)—which is defined by Kelly et al [19] as “a system of wireless, interrelated, and connected digital devices that can collect, send, and store data over a network without requiring human-to-human or human-to-computer interaction”—in regard to the analysis of data surrounding COVID. Their findings highlight that these fields aided the process of understanding large health datasets built from the data of numerous individuals.

These findings reveal the need for flexibility in tracking options to prevent overwhelming the user, and this exploration can be aided with HCI principles.

HCI Research on Technologies for Chronic Pain

We also draw from research on chronic pain, as it is a condition for which self-tracking is used for self-management. Adams and colleagues [20] investigated preferences regarding the self-assessment of pain for the purpose of self-managing chronic pain, through which they also discovered the previously mentioned importance of tailoring designs to suit individual differences [16,20].

Singh et al [21] carried out a study to explore the use of movement sensing and sonifying wearables by individuals with chronic pain in their undertaking of functional activity. The researchers found that the device challenged the user’s initial beliefs on their limits by improving their bodily awareness, and this led to users feeling more confident in trialing activity pacing methods. This highlights the importance of tracking and identifying limits for their self-management efforts.

Personal Informatics for the Support of Chronic Conditions

Personal informatics is an interdisciplinary research area within the broader HCI discipline, focusing on how people use personal technologies in life to collect, organize, and use information about their daily lives, in order to support personal goals in a range of domains such as productivity, learning, well-being, and health [22]. In the space of chronic conditions, this research area contributes design principles for digital health technologies such as mobile apps aimed at supporting fatigue. HCI work on chronic conditions has focused on tracking symptoms [23,24] or data related to physical activity [22,25,26], heart rate [27,28], sleep [22,29], and various other general health aspects [23,24,30].

Personal informatics also has value in supporting interventions; for instance, Felipe et al [29] found that it can greatly supplement the intervention of pacing, which is commonly undertaken by individuals with ELCs [2,9] and/or chronic pain [29,31]. In addition, individuals with chronic migraines also made use of personal informatics to identify if their condition is worsening in order to take preventative measures [32], which is known to be an important aspect of pacing [2].

Moreover, personal informatics can help users with chronic health conditions understand their condition better [24,29,32], offload the burden of monitoring their condition [29], improve communication and discussions with clinicians [24,32], help identify when their condition worsens, which is essential for taking preventative action in time [32], and finally, aid users in managing and monitoring their condition in the long term [32].

Objectives

Despite the numerous works focused on chronic conditions, limited research has drawn design implications for supporting fatigue by exploring the breadth of functionalities offered by top-rated apps in this space. In our research, we will highlight how our findings compare with existing apps for supporting the management of chronic fatigue syndrome, in terms of items

XSL•FO
RenderX

tracked and interventions. To this end, we put forth the following research objectives:

1. Conduct a systematic search on the top-rated commercial apps supporting fatigue
2. Explore and detail the range of functionalities offered by these apps
3. Identify whether the apps inform users of their theoretical or empirical underpinnings
4. Discern how the apps handle user data through the app's privacy policies and cookie notices
5. Inform design guidelines drawing from the functionalities discovered

We aim to address that gap by conducting a systematic search and a functionality review on the top-rated apps from today's marketplaces. To inform the methodology and paper structure, this paper draws on work done on app reviews in the areas of mindfulness [33], depression [34], goal setting [35], and digital wellbeing [4].

Methods

Overview

This section presents the process of selecting the apps and the description of functionality review, as well as analysis of app description and privacy terms.

Apps Selection Process

We conducted a systematic app search, for which we used a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) search strategy [36], similar to previous research on app reviews [4,33-35,37-39] including some in JMIR journals [4,34,38,39]. The search was conducted in the autumn of 2024 on the app's store descriptions, using the keywords "chronic fatigue syndrome," "chronic fatigue syndrome," "fatigue," "ME/CFS," "CFS," "myalgic encephalomyelitis," and "pace or pacing." This search was completed on the 2 most common marketplaces, which is also consistent with prior app reviews [4,33-35,37-39].

From the returned apps, we removed duplicates from both platforms, and for the remaining apps, we applied an inclusion criterion of having an average rating score higher than 3.0 and a number of ratings higher than 10. By screening the resulting apps, we further excluded apps that did not use any of the keywords or used them irrelevantly, as well as those not in English or requiring a subscription to use without even a free trial.

We chose to review apps that are top-rated (rating scores above 3.0) as they reflect positive user experiences of interacting with the apps, in addition to having a robust set of functionalities that the users appreciated and most likely found useful. From an HCI perspective, exploring these functionalities is beneficial as we can infer design principles and identify how we can improve the future design of such apps [4,33-35,37]. In addition, we chose to review apps that are free to access, similar to prior work [4,33,35,40], as they are more likely to provide better accessibility to the general population and fairness, as some users may not be able to afford the subscriptions [41], which is especially important in the health care context. In addition to being free, these apps might also have functionalities that are problematic, which our review aims to identify.

Functionality Review

As noted by Hossain et al [42], functionality review is an increasingly common HCI method used to explore current technologies and their main features, particularly commercial mobile apps [4,33-35], which authors identified as being descriptive, that is, merely describing the functionalities, or critical, that is, involving deeper analysis and integration [42].

For this, the first author engaged with the apps twice a day for at least 1 week, using both an Android 14 version phone and an iPhone iOS 18.1.1 version phone. This ensured that each app was used for a total cumulative time ranging from 70 to 140 minutes. The second author reviewed 3 out of 17, for a duration of 3 days. This was to add independent codes, ensure the codes are clear and consistently used by independent raters, and validate the coding process. The proportion of apps assigned to the second coder is consistent with similar studies such as the review of goal apps [35] and the review on depression apps [34].

The coding process was hybrid [43], involving both deductively identified functionalities such as tracking [33-35], visualizing [33-35], and interventions [33,34] for self-management, common in previously reported HCI work on app reviews, while the inductively identified ones emerging from the reviewed apps included subfunctionalities regarding the content being tracked, such as fatigue and its symptoms. We also coded characteristics such as the theoretical and scientific underpinnings of the app's functionalities plus their privacy policies and cookie notices. When coding the former, if we did not find the information pertaining to this while using the app, we sought it in the app's help pages and websites.

With regard to interventions for fatigue management, Davies et al [9] highlighted the importance of "pacing" in order to help people remain within limits of their energy level, promoting routinization, and tending to their mental health. Pacing has also been found to be used in complement with the participant's energy management efforts [9]. For this reason, the authors searched for functionalities specifically revolving around these concepts as part of their coding.

Interrater agreement—which was done only on the codes concerning tracking and interventional functionalities due to the large number of codes—using Cohen κ [44] showed almost perfect agreement for 149 out of 188 (79.3%) codes, fair agreement for 5 out of 188 (2.7%), and slight agreement on 34 out of 188 (18.1%) codes. The grouping and labeling of these codes into almost perfect, fair, and slight agreement were informed by the study of Landis and Koch [45]. From these, the authors revised the codes with fair or lower agreement.

The codes with slight agreement were mostly related to the tracked items captured along with their modalities. The main reason for the difference was due to a greater usage length being required to fully capture all the codes, especially with regard to the vast number of items tracked along with their recording

XSL•FO
RenderX

frequencies. As such, information gleaned from the longer usage duration was used to revise the codes into agreement.

Ethical Considerations

No human subjects participated in the research. For apps offering social media functionalities, we did not analyze any user-submitted content. Neither primary nor secondary human-related data were collected for this submission, which focuses exclusively on the description of the apps found on marketplaces and their functionalities.

Results

Overview

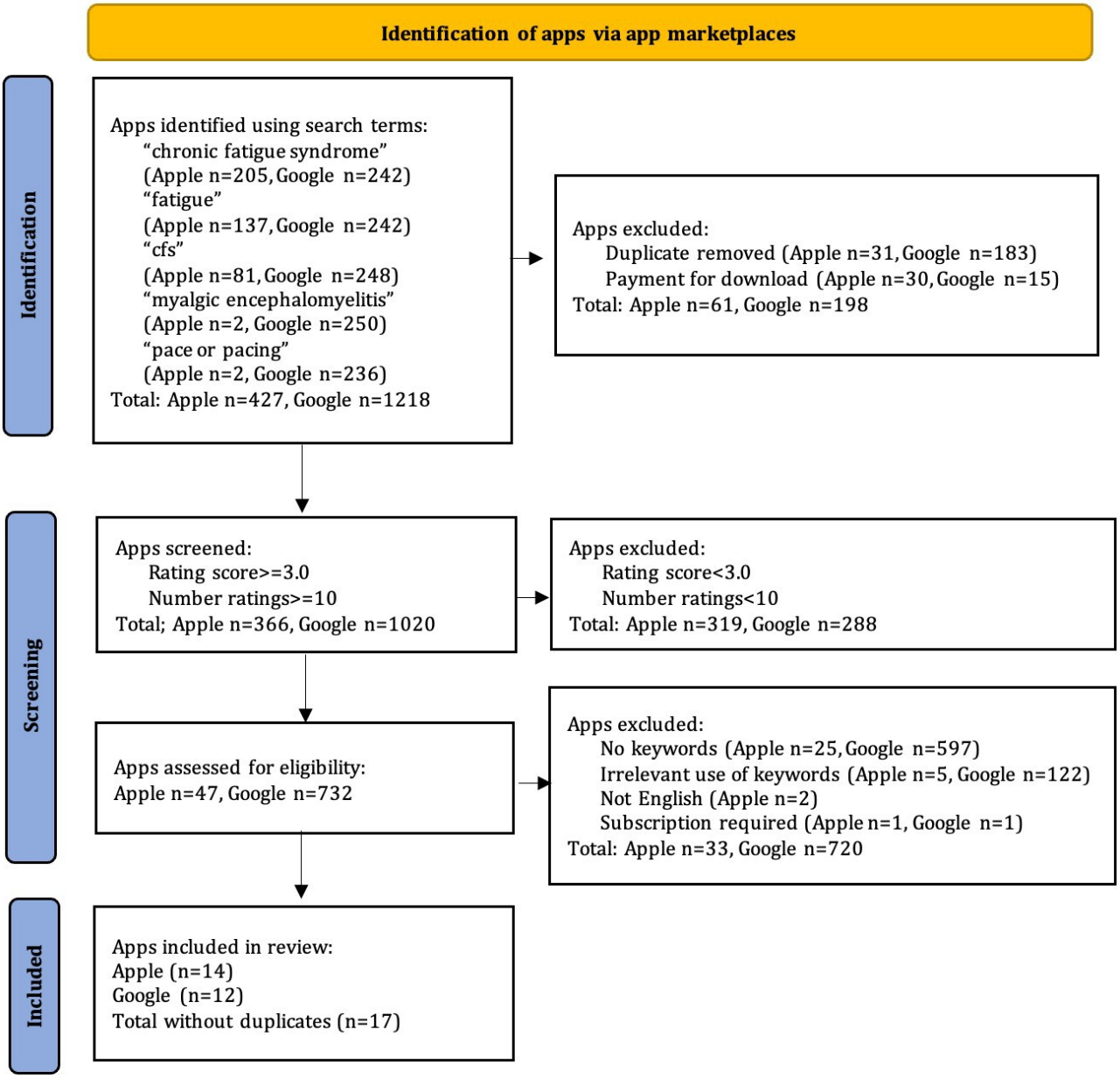
The app selection process resulted in a final set of 17 top-rated commercial apps from the Apple App Store and the Google App Store. Findings highlight that the identified apps not only include tracking functionalities but also support interventions for the management of fatigue, which involves managing energy levels and preventing crashes. The identified functionalities can be grouped into the main categories of tracking, visualizing, assessing, providing interventions, and ensuring privacy. Tracking functionalities comprise the collection of fatigue data and other related symptoms. Functionalities for visualization aim to represent the data in a manner that the user can understand and use toward the management of their condition. Assessment functionalities gauge the impact of the user's symptoms through validated scales. Intervention functionalities aim to help users manage their condition, and this can involve assisting with energy management and crash prevention. Finally, in the section highlighting the app's privacy measures, we will identify how apps handle users' data, the usage of cookies along with their purpose, and the age groups targeted. The full codebook has been supplied in the form of a Microsoft Excel sheet for viewing all codes and most of the subcodes (Multimedia Appendix 1).

App Selection Results

The initial search returned 427 apps on the Apple App Store and 1218 on the Google Play Store (Multimedia Appendix 2), and from these, we removed 214 duplicates from both platforms. Applying our inclusion criteria resulted in 366 apps from the Apple App Store and 1020 from the Google Play Store. By screening these, we further excluded apps that were less relevant, required a subscription, or were not in English, leading to 14 Apple apps and 12 Google apps. After accounting for duplicates across the 2 platforms, our final set comprised 17 apps: 5 from the Apple App Store, 3 from the Google Play Store, and 9 from both. The PRISMA diagram in Figure 1 [36] shows the selection process.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram for app extraction, detailing inclusion-exclusion criteria alongside app numbers included or excluded at each stage.



App Characteristics

A summarized view of the app characteristics has been presented in Multimedia Appendix 3.

While all selected apps had rating scores above 3, most apps from both stores had rating scores above 4.0 (13/14 for Apple apps and 10/12 for Google apps). Among the Apple apps, 4 apps had more than 100 reviews, and 2 apps had more than 1000 reviews. Most of the selected Google apps had more than 100 reviews (9/12 apps), with 5 apps having more than 1000 reviews. Our exclusion criteria removed apps that required either a one-time payment or a subscription; however, we included 3 apps with subscription models since they offered free trials.

Functionality Review Results

Overview

We now report on the findings from the functionality review, namely, the functionalities of tracking, visualizing, assessing, providing interventions, and ensuring privacy. Table 1 provides the main functionalities found in each app.

XSL•FO
RenderX

Table . Main functionalities that arose from the evaluation presented for each app.

Apps	Tracking	Visualizing	Providing interventions	Ensuring privacy
Bearable - Symptom Tracker	For tracking fatigue, symptoms, and associated factors	For visualizing tracked data	__[a]	For informing the user of their data handling
CareCircle - Health Platform	For tracking fatigue, symptoms, and associated factors	—	For social support	For informing the user of their data handling
Chronic Insights	For tracking fatigue, symptoms, and associated factors	For visualizing tracked data	—	For informing the user of their data handling
Curable: Chronic Pain Relief	—	—	For neuroeducation	For informing the user of their data handling
Elite HRV	For tracking associated factors	For visualizing tracked data	For energy estimation, psychoeducation	For informing the user of their data handling
Fibromyalgia Magazine	—	—	For psychoeducation	For informing the user of their data handling
Freeme: ME/CFS and Long COVID	—	—	For neuroeducation	For informing the user of their data handling
Guava: Health Tracker	For tracking fatigue, symptoms, and associated factors	For visualizing tracked data	—	For informing the user of their data handling
Gupta Program Brain Retraining	—	—	For neuroeducation, social support	For informing the user of their data handling
Turnto - Daily Breakthroughs	—	—	For social support, psychoeducation	For informing the user of their data handling
Visible: Pacing for illness	For tracking fatigue, symptoms, and associated factors	For visualizing tracked data	For energy estimation, psychoeducation	For informing the user of their data handling
Wave Health: Treatment Journal	For tracking fatigue, symptoms, and associated factors	For visualizing tracked data	—	For informing the user of their data handling
icompanion: understand your MS	For tracking fatigue, symptoms, and associated factors	For visualizing tracked data	For psychoeducation	For informing the user of their data handling
Ada - check your health	For tracking fatigue, symptoms, and associated factors	—	—	For informing the user of their data handling
SpoonieDay	For tracking associated factors	For visualizing tracked data	For pacing and energy estimation	For informing the user of their data handling
ME/CFS Pacing	For tracking associated factors	For visualizing tracked data	For pacing	—
Juva Stress and Migraine	For tracking associated factors	For visualizing tracked data	For biofeedback	For informing the user of their data handling

[a]Functionalities that are not available in the apps.

Tracking Fatigue, Symptoms, and Associated Factors

Overview

In this section, we explore the subfunctionalities that emerge under the main functionality of tracked content. At the first level of branching, we see fatigue tracking, symptom tracking, the tracking of associated factors and interventions, and finally modalities of the listed data as seen in Multimedia Appendix 4. The codes for the functionalities are also provided in Multimedia Appendix 5.

Tracking Fatigue

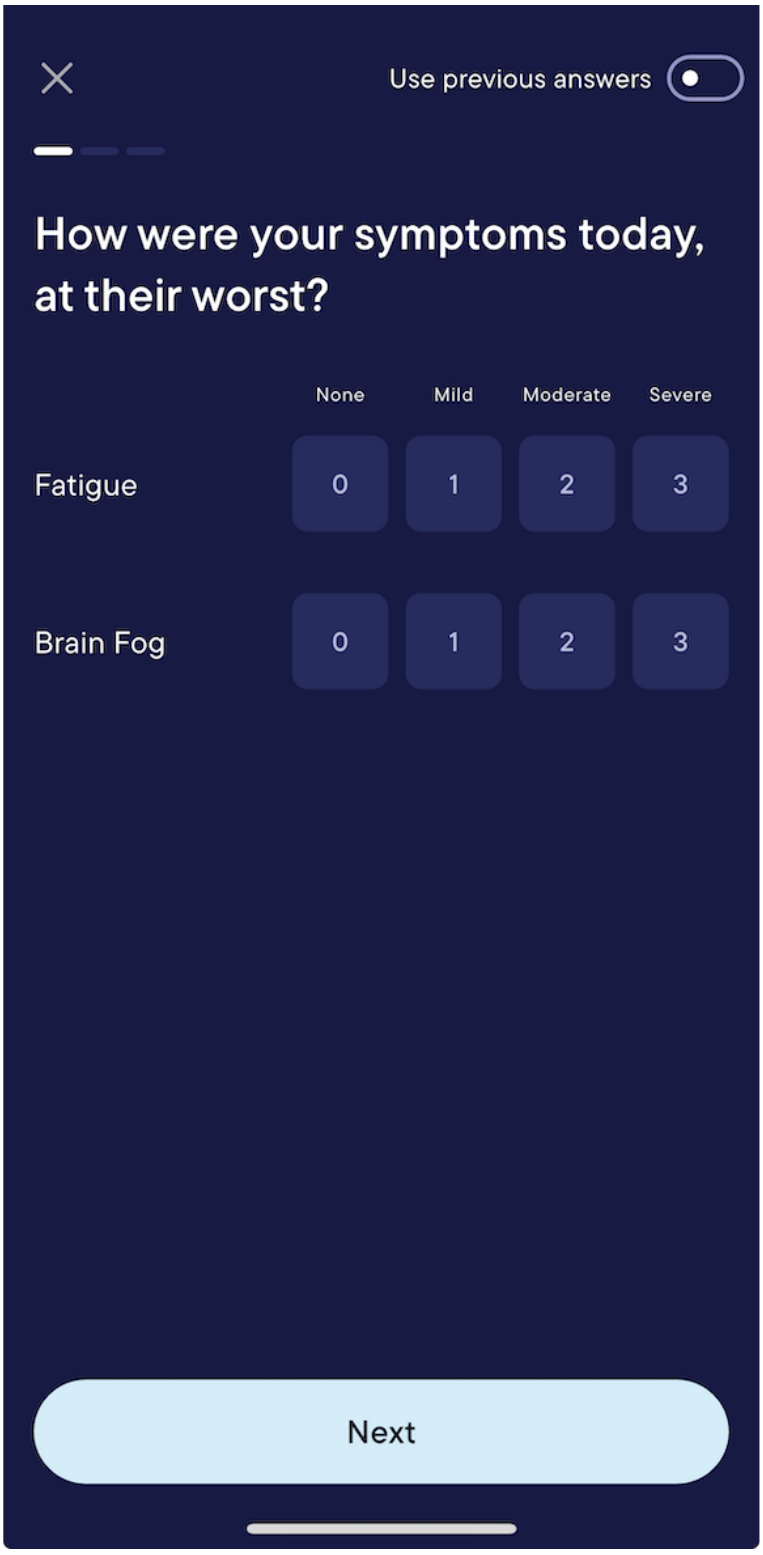
An important outcome is that from the 17 top-rated fatigue apps that we reviewed, less than half track fatigue (8 apps), and most of these (6 apps) track general fatigue, with the remaining 2 tracking physical fatigue. In addition, none of the apps offer a means of tracking fatigue automatically. This is important, indicating the potential of extending support for self-reporting of fatigue, and extending the measures of general fatigue to reflect other dimensions of fatigue: physical, cognitive, and mental or emotional. The apps that support self-reporting of fatigue do so through measures of trait fatigue or fatigue extended over time (5 apps), and state fatigue as fatigue momentarily experienced at the time of measurement (4 apps). The lack of usage regarding measures for state fatigue from half of the apps is also surprising, given the affordance of mobile apps for prompting its capture and thus tracking it through time.

Tracking fatigue is of great value to users with chronic fatigue, as this can be used to identify energy limits for the intervention of pacing [9] and thus, we believe the method of data capture and its granularity are important to be represented. To this end, we highlight that the few scales of fatigue used by the apps involve 1 item for each one, captured through 4-point (1 app) as seen in Figure 2 from the Visible: Pacing for illness app

XSL•FO
RenderX

(Visible Health Inc) [46], 5-point (2 apps), 7-point (1 app), 10-point (1 app), and 11-point (1 app) Likert scales, numerical scales (1 app), that is, 0.0 - 10.0 and semantic differential scales (1 app), that is, 3-point: mild, moderate, and severe level. From the apps that track fatigue only as a state, the scales can be completed at least once per day with no maximum (3 apps). No apps provide information on the validity of the used fatigue scales. Future app designs could better support ecological momentary assessment of fatigue states by using validated measures of state fatigue involving both general fatigue and its physical, cognitive, and mental aspects [2,47,48].

Figure 2. A screenshot showing a Likert scale used to capture data to track fatigue and other symptoms such as brain fog (image used with permission from the app's developer).



Besides fatigue, apps also track fatigue-related constructs such as energy level (4 apps), although there is surprisingly limited support for tracking exertion (2 apps) or crashes (1 app), which are essential for anticipating PEM and are much needed for

XSL•FO
RenderX

managing chronic fatigue [49]. Each of these constructs is captured manually, for energy levels through 5-point Likert scales (3 apps) and a semantic differential scale with 5 levels (1 app), exertion through 4-point Likert scales (2 apps), and crashes (an incidence of PEM) through a yes or no check augmented with a text entry for additional notes. Regarding the frequency option for daily manual entering of energy levels, this can be exactly once (1 app), at least once per day (2 apps), or 3 times per day (1 app). Interestingly, these self-reported measures of energy, exertion, and crashes can benefit from being contextualized with spatiotemporal and activity data [2], which may be automatically tracked. However, few apps (3 apps) supported such integration of self-reported and automatically tracked measures related to fatigue, and from the data mentioned, only activity data were leveraged.

Fatigue Symptoms

Findings indicate rich support for tracking 4 main types of fatigue symptoms: cognitive (8 apps), emotional (8 apps), physical (8 apps), and behavioral symptoms (5 apps). Multimedia Appendix 1 provides the top 40 most commonly tracked symptoms.

Most commonly tracked cognitive symptoms include brain fog (5 apps), difficulty with concentration (6 apps), or remembering (3 apps). Apps tracking emotional symptoms target mostly anxiety (8 apps), depression (7 apps), irritability (4 apps), pessimism (3 apps), sense of dread (3 apps), discouraged mood (1 app), disconnection or detachment from oneself and the surrounding world (ie, derealization or dissociation; 3 apps).

An important outcome is the prevalence of and range of bodily symptoms tracked by most apps. Such symptoms include most commonly pain (8 apps) and sensory symptoms (8 apps) such as numbness (6 apps) and sensitivity to light (6 apps), followed by specific ones concerning skin (6 apps), circulatory (7 apps), digestive (7 apps), muscular (5 apps), ocular (6 apps), respiratory (7 apps), reproductive (6 apps), or urinary systems (5 apps). Thus, most apps support a rich set of symptoms that could be tracked, which ranges between 76 and 155 different symptoms. While useful, tracking many symptoms, particularly manually, can be taxing [2], especially since many apps provide limited support for reducing such burden.

An interesting outcome is the various behavioral symptoms, with the most commonly tracked ones including nail biting (4 apps), restlessness (4 apps), fidgeting (3 apps), impulsivity (3 apps), skin picking (2 apps), and stimming (2 apps). A closer look at these indicates they are predominantly repetitive body movements [50], which previous work described as supporting self-regulation of sensory sensitivity in autism spectrum disorders [51], albeit also contributing to autistic fatigue [52].

Symptom Checkers

While the large number and diversity of the symptoms available to track is useful, it can also be overwhelming. Two apps use symptom checkers such as Ada's symptom assessment functionality and Isabel (used in the CareCircle – Health Platform app) for supporting users to report their symptoms. While developed to support clinician diagnosis, symptom checkers implemented on mobile apps have also started to be explored by researchers investigating user perspectives for usability and acceptability [53,54]. Recent work has also explored AI-augmented symptom checkers with findings showing effective diagnosis, but problematic triage, and thus their use by unsupervised patients may be harmful [53].

Factors Associated With Fatigue

The reviewed apps also support tracking of additional factors which may impact fatigue and thus may allow people to better understand their fatigue experiences. Such factors include health data (7 apps) such as manually entered medical information (4 apps), medication intake (6 apps), or acute conditions such as flu and cold (1 app). From these, medication is tracked either by automatic logging after a one-off entry (3 apps) or through manual logging with either daily entries (1 app) or recording in a list (2 apps). Other health data were also tracked through the same list style of recording (4 apps). The modality of data capture is interesting from an HCI perspective, as HCI research on ELCs has called for tracking apps with a low impact on users' energy levels [2,9]. In addition, since these users most likely have other symptoms alongside fatigue [55], this becomes particularly important. This opens up design opportunities toward implementing automatic tracking methods, as it would be the least taxing for this demographic. Health data were also automatically tracked, such as heart rate (6 apps), HRV (5 apps), resting heart rate (3 apps), oxygen saturation (2 apps), or breathing rate (2 apps). For the capture of health data, some apps leveraged the OS's health app (6 apps), while other apps made use of the phone camera (1 app) or additional wearables such as heart monitors or smartwatch sensors.

Apps tracking physical activity focused mostly on step count (6 apps) or exercise duration (4 apps), captured automatically (6 apps), directly from the OS's health app (6 apps), while 4 apps provide the option to manually enter step count. Mood was another important associated factor (6 apps), captured through a 5-point Likert scale (3 apps), augmented with emotion tags (1 app), a 10-point Likert scale with emotion tags (1 app), or merely a dropdown emotion list (1 app).

Another associated factor was sleep, captured through manually entered data using a 4- or 5-point Likert scale for sleep quality (1 app each), a dropdown list of factors impacting sleep (1 app), a time picker for daily entry of sleep duration (2 apps), or automatically tracked through the OS's health app (2 apps). Other associated factors include weight (4 apps) and lifestyle data such as stress, caffeine consumption, or alcohol consumption (3 apps).

Moreover, a total of 5 apps offer the option to track well-being activities such as journaling (3 apps), meditation (3 apps), mindfulness (3 apps), and yoga (3 apps).

The use of scales may result in the loss of contextual data [56], which may aid sense-making; however, some apps have attempted to remedy this by offering the option for notes (5 apps) with each recording of most or all data types. The optionality of this feature is important, as this lowers the burden on users by allowing them to keep their data collection simple if their energy reserves are low.

XSL•FO
RenderX

Visualizing Tracked Data

Overview

The daily use of the apps during the expert evaluation provided a sufficient amount of data for the assessment of their visualization functionalities. The findings from this reveal the provision of visualizations of tracked fatigue, symptoms, and associated data through charts representing different types of data, color-coded symptom severity, temporal aspects, and data summaries (Multimedia Appendices 6 and 7).

Types of Charts for Visualizing Fatigue

Findings show that the most common charts for fatigue visualization are line charts (4 apps) and bar charts (2 apps), although not all apps visualized tracked fatigue data (2 apps), despite users' interest in understanding their tracked fatigue for the purpose of self-management [9].

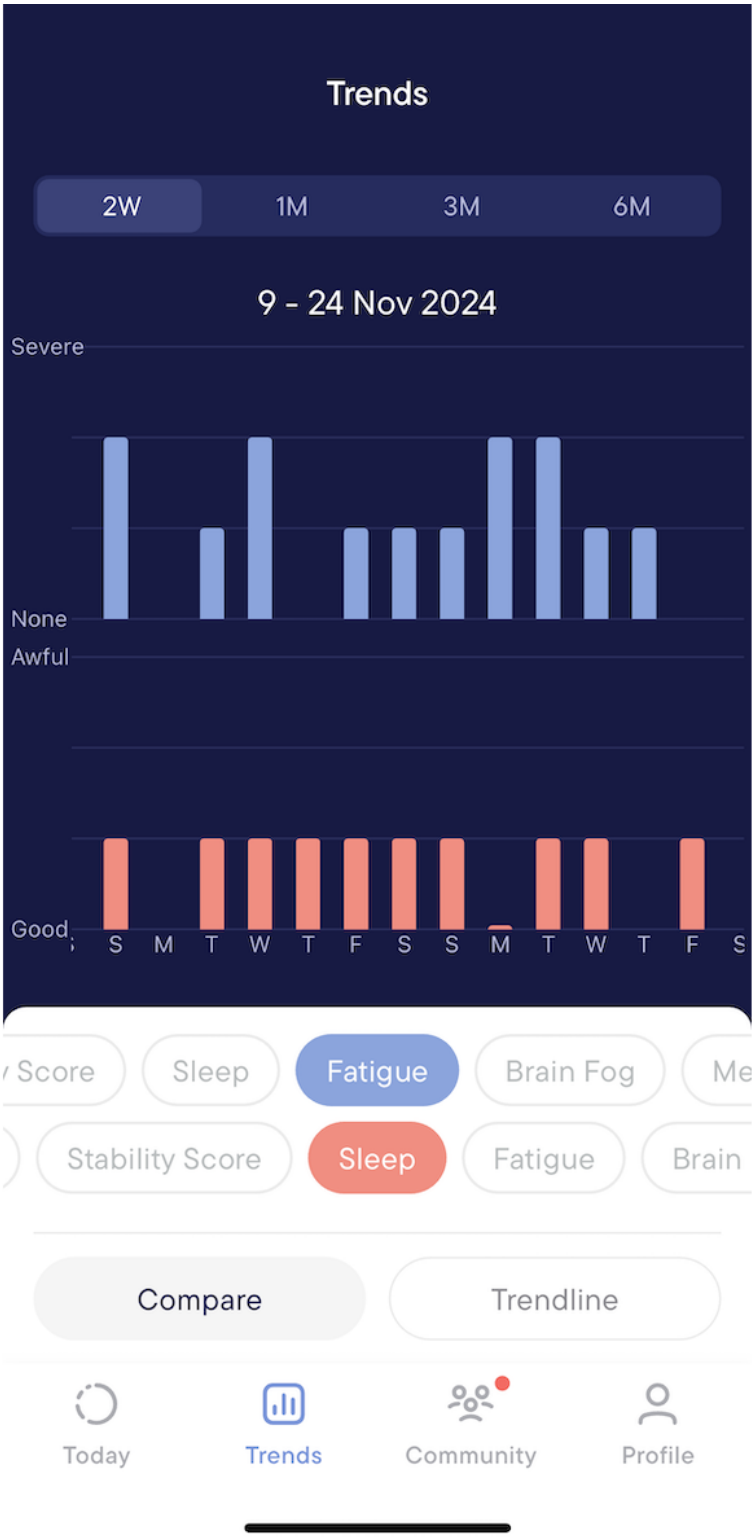
Types of Data Represented in Charts

Fatigue is known to have numerous factors [2,9], and for this reason, users will want to reflect on other data sources as well. Findings indicate that 9 apps provide visualizations of a single data type, such as fatigue scores (6 apps). In addition, as the factors are known to be interdependent [2,9], users would benefit from being able to compare the data on the same chart. To this end, apps also offer visualizations integrating different tracked data (7 apps): most often fatigue data + symptom data (4 apps), fatigue data + health data (4 apps), health data + health data (5 apps), symptom data + symptom data (4 apps), symptom data + health data (4 apps), fatigue data + mood data (3 apps), symptom data + mood data (3 apps), physical condition + mental condition (1 app), or morning + evening mood (1 app). It was also noted that 3 of these apps providing visualizations of combined data do not have the option of integrating fatigue, which may limit users' ability to better understand their fatigue by discerning factors that could affect it [9].

With respect to the number of different data used for such combined visualizations, most apps involve only 2 *y*-axis variables per visualization (6 apps), while 1 app allows an unlimited number of variables integrated in one app, which for numbers larger than 3 may be difficult to understand if not designed in an effective manner. Combined visualizations are often represented through multiple line charts, that is, one line for each *y*-axis variable (4 apps), combined bar and line charts (3 apps), combined scatter charts (1 app), parallel bar charts (1 app) as seen in Figure 3 from the Visible: Pacing for illness app [46], combined bar charts (1 app), and bar and dot combo charts (1 app). Of these, the most frequently used visualization was found to be combined line charts (4 apps). Given the interdependency of various aspects of fatigue [2,9], it is surprising that these visualizations do not look at them altogether in contrast to the isolated view, and this could be an area which future work could investigate. Finally, findings highlight that 1 app does not show combined visualizations, despite tracking different data.

XSL•FO
RenderX

Figure 3. Screenshot showing a combined visualization of 2 *y*-axis variables for comparison (image used with permission from the app's developer).



Use of Color-Coded Symptom Severity in Visualizations

From the 8 apps that provide visualizations for symptom tracking, half of the apps use warm colors such as red to indicate negative symptoms or high severity level (4 apps), while the others do not use such color codes (4 apps). The former highlights severity level, which can increase one's concerns and be perceived as judgmental [9].

Temporal Aspects of Visualizations: Timescale and Historic Versus Real-Time Data

All 9 apps providing visualizations for historic data do so with a range of timescale options such as the last 7 days (1 app), last 2 weeks (1 app), last month (3 apps), last 2 months (2 apps), last 3 months (3 apps), or a custom timescale ranging from the last 10 weeks to the last 7 years (1 app), any 7 days (3 apps), any 2 weeks (1 app), any month (2 apps), any year (1 app), as well as a custom scale with a start and end at any 2 points in

XSL•FO
RenderX

time (1 app). These various timescales are beneficial, with the most recent past supporting understanding while leveraging the memory of recent events impacting the displayed fatigue scores, and the far past supporting understanding broader patterns of how fatigue changes over time [57]. Moreover, findings show that 7 apps display only historic data, 1 app displays both historic and real-time data, and another app displays only real-time data. From the 2 apps that show real-time data, both apps use it for biofeedback interventions. In addition, the real-time data are measured with an HRV monitor for one app and with both a camera and an accelerometer for the other. Finally, one app visualizes HRV, and the other visualizes heart rate and breathing rate.

From the apps that visualized fatigue, we note that 2 apps represented all data points per day on the chart, while the other 4 apps represented one data point per day. While 2 of the latter only record 1 fatigue score per day, the other apps either average the score (Guava: Health Tracker) or show the highest score for the day (Bearable - Symptom Tracker). This may be either a benefit or a drawback to self-trackers who deal with chronic fatigue. On one hand, representing all the data points could take more mental effort to digest, which could be remedied by per-day summaries; however, they hide the changes in fatigue, diminishing the insights obtainable from their tracking efforts. This highlights the value of visualizing a minimum of 3 data points a day, with options to show fatigue data before and after activities, as users typically seek to understand the impacts of activities for pacing purposes [9].

Presentation of Distinct Summaries of Some or All Data

Findings reveal that 4 apps display data summaries. One app shows averages of any tracked variable within a chosen timescale. A second app shows the averages of symptoms and steps over a weekly or monthly period. A third app shows averages of morning and evening mood over 7, 30, or 90 days. The fourth app shows averages of daily scores within a yearly view visualization, and averages of monthly scores within a monthly view. These data summaries are important, as providing users a more condensed form of data could reduce the cognitive load of understanding their fatigue, which is important in preventing the apps from exacerbating their condition [2].

Assessment Tools

Regarding symptoms, 2 apps provide valid scales to assess the symptoms' impact on everyday life (Multimedia Appendices 8 and 9). One is the Functional Capacity in ME/CFS (FUNCAP) [58] (David Davies-Payne), whose development was informed by patients living with fatigue. This tool effectively accounts for the impact of PEM on completing activities. In its short version, it consists of 27 items covering 8 domains from personal hygiene to concentration. The other tool is Quality of Life in Neurological Disorders (Neuro QOL) [59,60], consisting of 8 or 9 items for each of 13 domains from physical to mental health. Both tools are retrospective reports as they require reflection on past experiences to complete [61].

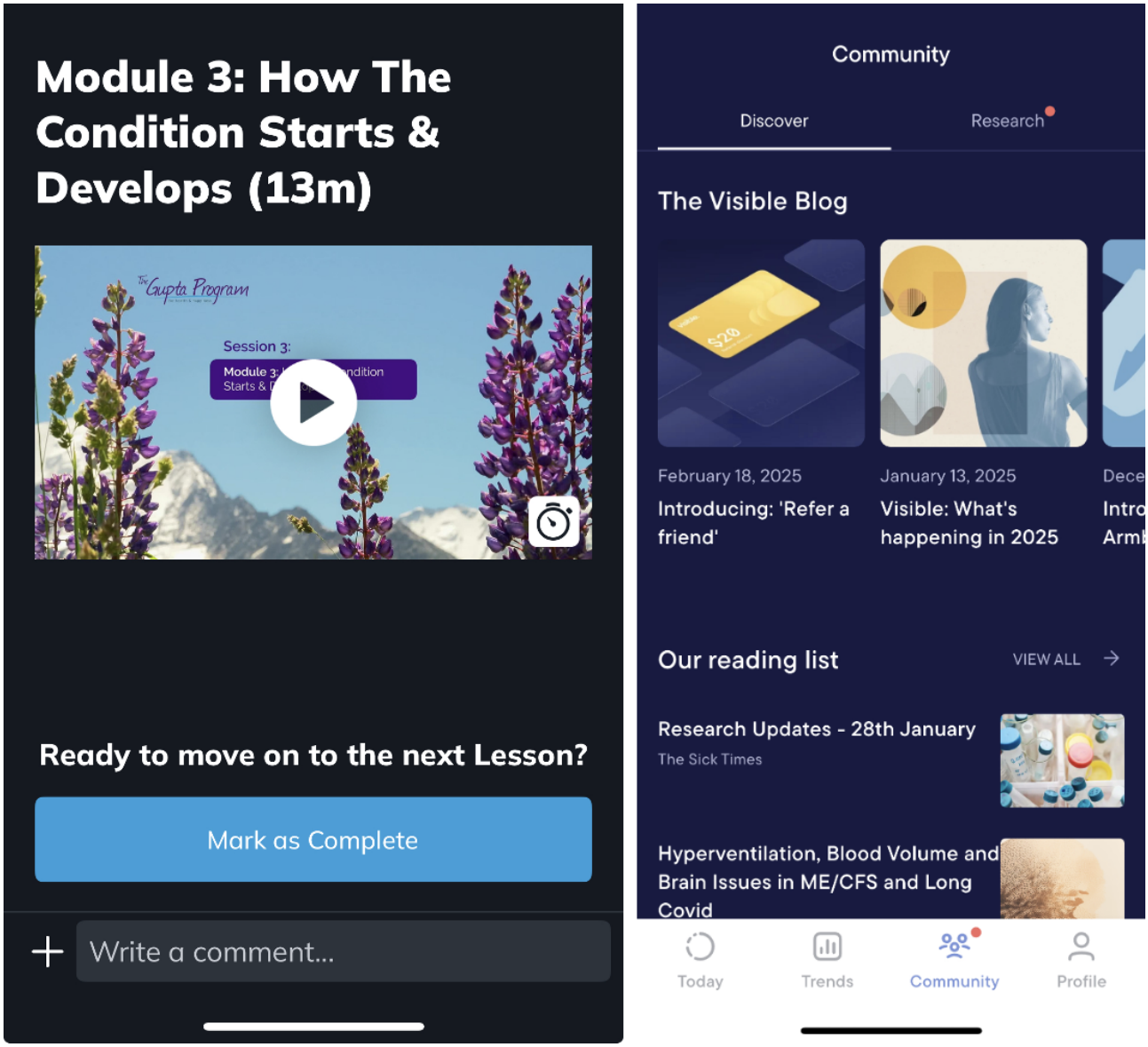
Interventions for Managing Fatigue

Overview

Despite the rich set of symptoms and tracked data, the top-rated apps for fatigue provide limited interventions for the self-management of fatigue, and in particular, those leveraging tracked data (Multimedia Appendices 10 and 11). Findings highlight a range of interventions provided by the apps, one of which is neuroeducation (3 apps); a treatment approach that aims to improve user well-being via educational lessons and exercises built around neuroscience concepts [62]. The apps in our review chose an approach centered on the concept of neuroplasticity and one such app is the Gupta Program Brain Retraining (Harley Street Solutions LTD) [63], as seen in Figure 4. Another intervention is psychoeducation (5 apps), which is delivered through articles informing users of their condition along with possible treatments and informing them of new advancements in research surrounding their condition, as seen in the Visible: Pacing for illness app [46] in Figure 4. Functionalities for social support (3 apps) were also implemented, in the form of community engagement through forum posts and social networking. These functionalities offer value to users with ELCs by providing them with a space for support while promoting discussion around their condition [7,10,15-17]. Moreover, 2 apps included biofeedback-based therapy, which aims to regulate the nervous system to reduce symptoms through the use of deep breathing (2 apps) or muscle relaxation (1 app) exercises, the latter of which has been known to reduce fatigue and may help individuals with CFS [64]. Finally, 2 apps were found to offer functionalities supporting pacing (2 apps).

XSL•FO
RenderX

Figure 4. Screenshots showcasing interventions provided by 2 different apps (psychoeducation on the left, and educational support on the right [images used with permission from the apps' developers]).



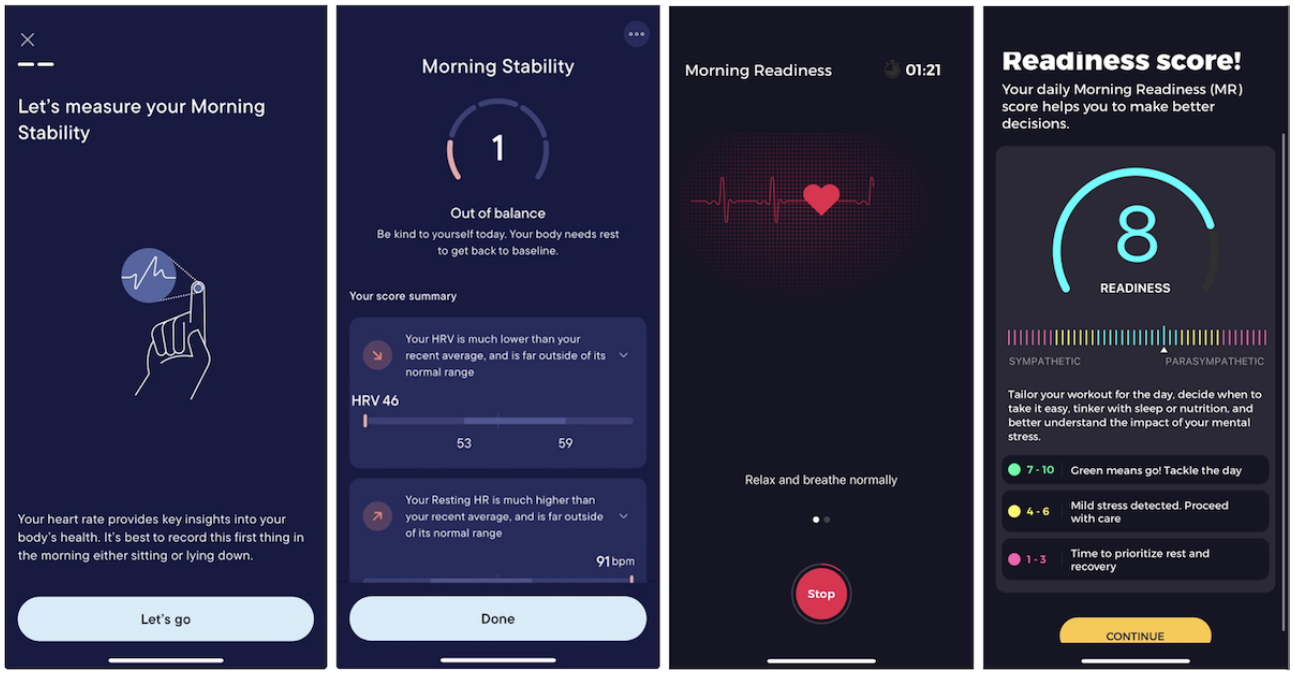
Pacing is an important intervention, tailored for fatigue [1], which involves planning and allocating one's energy toward activities to be completed based on their energy demands [9] through the so-called spoon management [9,65] or energy envelope approaches [9,66]. Pacing involves identifying what activities one can do, in addition to the extent that such activities can be carried out [67]; additionally, it entails stopping activities when breaching PEM and can also integrate the act of estimating energy levels [68]. Previous findings indicated the value of pacing interventions for managing fatigue in addition to preventing overexertion and PEM [1]. While not an intervention per se, 3 apps also provide support for energy estimates, which is important as it can provide input for pacing interventions.

Energy Estimation Using HRV and Pacing

Estimates of energy are manually tracked through self-reports (1 app), which is the traditional approach as a prerequisite for pacing. Interestingly, another 2 apps use a method of tracking daily energy estimates through HRV measurement in the morning, which is compared to the user's baseline to provide either a stability score by one app (Figure 5) [46], or a readiness score by the other (Figure 5) [69]. These scores are used to notify the user of potential flare-ups of their conditions and, as mentioned before, can inform their pacing decisions [68].

XSL•FO
RenderX

Figure 5. Screenshots showing energy estimation approaches by 2 different apps (images used with permission from the apps' developers).



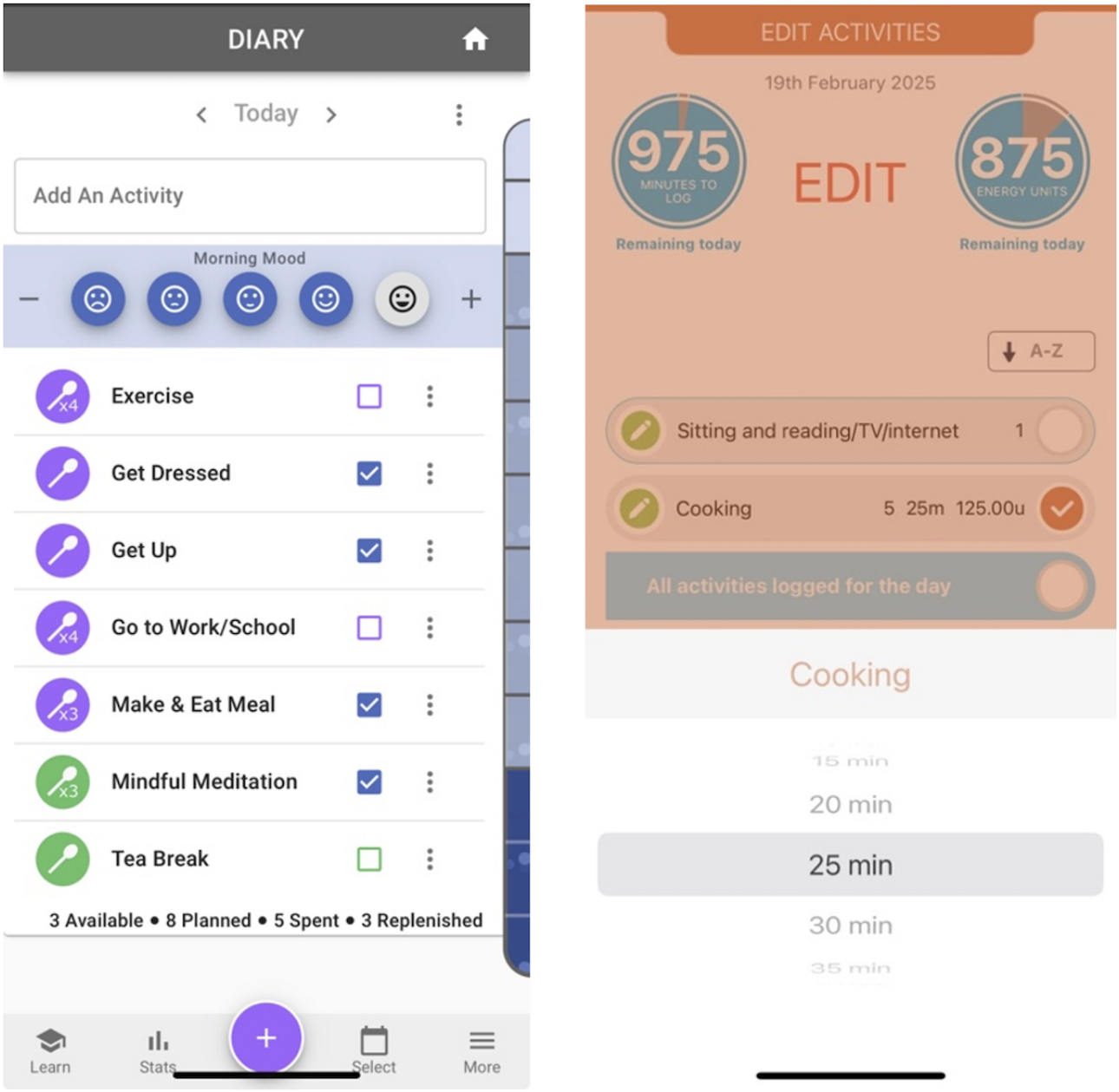
Previous work suggested the potential of HRV data as an indicator of fatigue [70,71], albeit its use to estimate energy has been less explored. Such approaches to estimate both fatigue and energy levels are crucial to lower the burden of tracking and its cognitive load, which is especially beneficial for an individual who may be approaching PEM [2]. Interestingly, however, not all apps estimating energy also provide pacing interventions.

Pacing Interventions

These interventions are provided only by 2 apps using different energy units. One app uses spoon management [9,65] as seen in Figure 6 [72], and the other leverages energy envelope approaches [9,66] as seen in Figure 6 [73]. The former has a maximum daily energy level set to 12 spoons, while the latter uses a less conventional and finer-grain maximum daily amount of 1000 energy units. Both apps provide predefined and customizable activity lists with predefined energy impacts for each of the listed activities, as well as gamification in the form of a limited-quantity source [74] to track and account for energy as a resource. Additionally, they both use traffic light–style red and green coding to visualize energy usage as positive and negative [75], and to deliver cautionary feedback as a consequence of extending one's energy beyond the daily threshold amount.

XSL•FO
RenderX

Figure 6. Screenshots from apps providing spoon management (left), and an energy envelope approach (right; images used with permission from the apps' developers).



Planning daily activities based on one's energy can support not just pacing, but also structure and order, or routinization, with previous findings indicating its benefits for people living with CFS [9]. Routinization could be further supported by automatically prefilling in activity fields on apps to further reduce the user's burden. This is important as a previous study has shown that fatigue symptoms can impede user engagement with self-management apps [9].

Theoretical and Empirical Underpinnings of Interventions

Our findings show that 6 of the apps offering interventions such as psychoeducation, neuroeducation, energy estimation, and biofeedback had provided either theoretical or empirical groundings of their methods and informed the user either through app help pages or the app's website. From these, 4 of the apps cite academic work to support the efficacy of the intervention that they provide neuroeducation (Gupta Program Brain Retraining [Harley Street Solutions LTD]); pacing (Visible: Pacing for illness [Visible Health Inc.]), an energy estimation app; biofeedback (Juva Stress and Migraine [Juva Health Inc.]); and HRV-based interventions (Elite HRV [Elite HRV Inc]), respectively. However, only 2 of these apps (Visible: Pacing for illness [Visible Health Inc.] and Gupta Program Brain Retraining [Harley Street Solutions LTD]) provide empirical evidence supporting the effectiveness of the app itself. Two other neuroeducation apps (Freeme: ME/CFS and Long COVID [Freeme Health Ltd], and Curable: Chronic Pain Relief [Curable Inc]) cite academic work supporting all or some of the information provided to users in their educational material.

XSL•FO
RenderX

Data Extracted From App Privacy Policies and Cookie Notices

Overview

We now report on data from app privacy policies and cookie notices through 3 main identified themes, namely the target age groups of the apps, information surrounding the handling of data, and the usage and purpose of cookies (Multimedia Appendices 12 and 13).

Target Age Group for App Use

Findings reveal that the majority of the apps (10 apps) restricted their use to nonadult individuals. From these, 7 apps mention that they do not provide services to users younger than 18 years, 2 apps do the same for users younger than 16 years, and 1 app restricted nonadult users unless the legal guardian or carer provides consent.

Information Surrounding the Handling of Data

Findings indicate that for 11 apps, their developers disclosed that they stored data gathered through the app on company servers, while 1 app stored it with a third-party site called Sentry Inc. The latter has stated in its privacy policy that this is done only for anonymized crash data.

As the apps store their data outside the user's device, it is important to reveal the parties with whom the data are being shared. Findings show that via the use of the app, users consent to sharing data with advertisers for 4 apps, 2 of which have revealed that personal data will be shared. Other parties with whom data will be shared in this manner include analytics (3 apps), anonymous crash data processing as mentioned earlier (1 app), and research (2 apps), but in an anonymized form.

Outside of the previously mentioned data, some apps also offered the option to opt-in to third-party data sharing with advertisers (1 app), analytics (1 app), the OS's health app (1 app), third-party backup storage providers (1 app), and 5 apps present the option to share data with researchers, of which 2 apps claimed that the data will be anonymized, and a third app revealed that the data will be used for product development and health research. It is worth noting that opting in with advertisers, the OS's health app, and backup storage providers involves the sharing of personal data.

Another important finding is the information surrounding the sale of user data; 3 of the apps were found to sell data, one of which revealed that they only sell personal information with the user's permission. Seven other apps offered no information surrounding the sale of data, and the remaining 6 apps revealed that they do not sell. Regarding the latter, 4 of them explicitly stated no, 1 app claimed that they do not sell the information covered by the policy, and the final app revealed that it does not sell data from Californian residents. Meanwhile, 2 of the 3 apps selling data informed that the data were being sold to advertisers (2 apps) and analytics providers (1 app). Data surrounding the sale of information were mostly easy to find, with the majority (6 apps) highlighting the information under clearly labeled headings and subheadings, while 2 apps detailed that they do not sell it without highlighting or obfuscating the information. However, one app that sold personal data did not state it on the main page of the privacy policy and was instead presented on a separate page for Californian and US residents. While this page was linked in the privacy policy, the document did not point to the information regarding the sales being there and thus appears hidden to users outside the United States, as they may not explore sections irrelevant to them. This emphasizes the importance of clearly highlighting information regarding sales to users through the use of subheadings to improve discoverability.

Regarding ownership of data, only 1 app's privacy policy offered information surrounding this, and it revealed that the developer is the data owner.

Usage and Purpose of Cookies

We also analyzed the app's privacy policies and cookie notices for information surrounding cookies and have found that 8 apps used cookies for analytics purposes, with one of the apps only enabling cookies outside of app and website areas requiring a log-in to use, to protect their users' health data. A total of 6 other apps stated that they made use of cookies for both advertisement and analytics, while the remaining 3 provided no information surrounding the use of cookies.

Discussion

Principal Results

Within the main tracking functionality, our findings highlight a rich set of tracked content on fatigue, its symptoms, and associated factors, whose understanding is supported by various visualizations. Regarding interventions, however, a limited number of apps leveraged this rich and diverse tracked data to support pacing of estimated energy. In light of these outcomes, we articulate 5 design implications capturing both the strengths of some of the top-rated apps for fatigue, as well as how such apps can be further improved, along with the broader design of technologies for fatigue.

Supporting Hybrid Tracking: Balancing Validity of Self-Reported Fatigue With Low Burden Automatic Tracking of Symptoms

Two key outcomes are the limited use of validated scales for capturing and tracking fatigue, which limits the validity of tracked data and its value for recommending interventions. Validated scales come with potential drawbacks regarding lower engagement due to their burden [56]. Information might also be less personally meaningful to the participant as the level of data collected is restricted to that elicited by the scale; however, the burden can be reduced by using scales that are short and simple [56]. In addition, the loss of meaningful contextual data can be remedied by offering the option to add notes with every entry, as 5 apps in this review have done. Therefore, we strongly recommend the use of validated fatigue scales, both for trait and state fatigue. The former can be used before and after an intervention program, while the latter can be used for frequent tracking of fatigue. To identify such valid scales, we can draw from a previous systematic review [76], showing that the most commonly used trait scale includes the Modified Fatigue Impact Scale (MFIS) [77], and for state scale we could use the 1-item

XSL•FO
RenderX

Samn-Perelli Fatigue Checklist [78], or the 4-item Fatigue State Questionnaire (FSQ) [79], which is also used in ecological momentary assessment with digital technologies [80]. The latter state scales are not only valid but also consist of a small number of items, representing an important consideration for lowering user burden for daily tracking. Moreover, we have seen emerging efforts to capture fatigue through HRV data both in some of the apps via commercial heart rate sensors (which include chest straps such as the Polar H10) capable of HRV capture by the Elite HRV app, and photoplethysmography [81] conducted through the smartphone camera by the Visible: Pacing for illness app (Visible Health Inc.), along with scholarly work [70,71] typically focusing on electrocardiogram (ECG) data. If these approaches can be scaled to an automatic process, this could offer substantial support to users seeking to lower the burden of tracking. In addition, the automatic tracking of fatigue has even been found to be desired by participants in co-design workshops tackling pacing technologies [11]. The implementation of a hybrid tracking system using low-effort or even automatic capture of fatigue through both self-reported measures of scales and HRV data can explore and ensure the validity of the latter, while lowering the burden of the former [2]. For this, we can also think of novel, low-cost tracking methods on smartwatch interfaces, possibly ML-powered [2].

Findings also indicate the vast number of symptoms available to be tracked, which can be much needed to inform the management of fatigue, albeit they may be taxing to track and understand. To mitigate this downside, we propose the use of free-form conversational interfaces supported by natural language processing (NLP); this can reduce the mental load on users while also allowing for richer data input and capture [82]. We further suggest hybrid tracking of such symptoms, leveraging low-burden self-reported data with lower frequency, augmented by automatically tracked data [2,9] through the app, or phone itself, as well as smartwatch sensors via the OS's health apps paired with heart rate monitors. We have also seen emerging efforts to integrate symptom checkers such as Ada and Isabel to support users in navigating their symptoms. This opens up new design opportunities for AI-augmented symptom checkers, although ethical implications need to be carefully considered, as despite their usability and acceptability [83], the value of AI checkers for unsupervised patients may be harmful [53].

Contextualizing Fatigue Data With Symptoms and Associated Factors: Combined Visualizations for Sense-Making

Findings reveal that while most apps support the tracking and visualizing of fatigue data, only half support combined visualizations, some of which show fatigue data shown alongside other tracked symptoms or associated factors, while others combine different symptoms with or without associated factors, albeit without fatigue data. Combined visualizations highlighting fatigue data against symptoms or associated factors are, however, crucial for understanding if and how the latter may have impacted changes in state fatigue over time. Integrating different data types in one ready-to-understand visualization is not trivial. For this, we suggest low complexity ensured by combining fewer data types, such as fatigue and steps, for instance, through visualizations with 1 or 2 line charts augmented with additional information such as one's location. The selection of data types, possibly powered by AI, would be best informed by data types with the strongest impact on the changes in fatigue scores. Visualizations may also include data summaries over time, some already supported by a small number of apps. Color-coded symptom severity levels may be avoided to lower users' emotional concerns, and more subtle ways of informing severity levels can be instead explored. With respect to timeline, we suggest providing options with complementary value: visualizations of the most recent past benefit from memory of recent events to support recall of key factors impacting fatigue and changes in daily scores, while those of the far past can highlight broader patterns of changes in fatigue over time [57].

Normalizing the Inclusion of Energy Estimates and Pacing Interventions in Apps Targeting Fatigue

While many of our reviewed apps support interventions, only a few provide support for the intervention of pacing. This is surprising, given the rich body of work indicating its effectiveness for the self-management of fatigue [2]. This opens up a design space for pacing interventions for which we suggest drawing from spoon management [9,65] or energy envelope approaches [9,66]. We can think of novel interfaces which leverage the spoons concept as energy units for energy estimates, scaffold users' planning of energy estimates to daily activities, and provide subtle rather than punitive guidance when an energy threshold is exceeded to help users prevent crashes. The latter is important, as crashes can cause people's energy levels to remain depleted over time, that is, days [2] or even months [49]. As shown by the apps "SpoonieDay" (Blackburn Labs) and "ME/CFS Pacing" (Emerge Australia Inc), one approach to mitigate this is for users to identify at the start of the day their available overall energy, prioritize the main activities that can be covered with it, and allocate the available energy toward core activities, while monitoring that their expenditure remains within the allocated limits. Although users can try to estimate this via methods such as spoon management [9,65] and energy envelope approaches [9,66], they may still underestimate their energy usage needs, and therefore, a warning system may be more useful.

Planning daily activities based on one's energy can support not just pacing, but also structure and order, or routinization, with previous findings indicating its benefits for people living with CFS [9]. Routinization could be further supported by automatically prefilling in activity fields on apps to further reduce the user's burden. This is important as a previous study has shown that fatigue symptoms can impede user engagement with self-management apps [9].

We have also noted the limited use of tracking crashes (PEM events), which can be highly useful, as these happen when participants go over their limits. Identifying one's own limits is one of the foundational aspects of pacing, and this can be further supplemented by contextualizing the data with activity data and spatiotemporal information.

XSL•FO
RenderX

Most importantly, findings revealed that the apps that offer pacing functionalities do not provide energy estimation functionalities, and vice versa. Users would benefit from these functionalities being put together, as energy estimation would give them more information to work with for pacing, and could especially support decision-making (when planning activities) on days when they are more cognitively fatigued.

Comparisons With Prior Personal Informatics Research in the Space of Chronic Conditions

We compared the findings to those of previous works in the field of personal informatics, as they can serve as valuable evidence given the common ground they share. The main difference that we noted is that the apps we reviewed track a much larger range of symptoms and health data compared to the works in the space of personal informatics. Due to fatigue being impacted by a wide range of interrelated factors [2,9], this level of tracking could be needed by users to understand their condition. While tracking a substantial number of symptoms will eventually overwhelm users [2], lightweight tracking solutions, as discussed earlier, could help negate this.

With regard to interventions, pacing is a unique approach that has received some focus from the field of personal informatics. For the purpose of pacing, while historic data are useful, Felipe et al [29] reported that users with chronic pain desired to see their data in real-time to identify if they were exceeding their limits. However, we have found that only a limited number of apps support this feature. Future apps could implement functionalities to show a live view of tracked data, as this could be useful for identifying if users are approaching PEM.

Limitations

The main limitations of the study are that our app review did not include non-English apps, apps outside the UK app stores, and finally, paid apps. Regarding the latter, as mentioned earlier, and aligned with previous work [4,33,35,40], we focused on apps that were either completely free or offered free trials because, in the health care context, accessibility and fairness are vital, especially since vulnerable users may not be able to afford subscriptions [41]. Nevertheless, other apps such as paid ones, non-English apps, and apps outside the United Kingdom are indeed likely to include additional valuable features, and future work could explore these to identify any further design insights on supporting the management of fatigue via apps.

Conclusions

This paper presents a functionality review of 17 fatigue apps selected as top-rated apps from the Google Play Store and the Apple App Store. Findings indicate 3 main functionalities pertaining to tracking fatigue, its symptoms, and associated factors, visualizing tracked content, and providing interventions for the self-management of fatigue. Our findings informed 3 implications for sensitive design of technologies for people living with fatigue, which include supporting hybrid tracking to balance the validity of self-reported fatigue with low-burden automatic tracking of symptoms, contextualizing fatigue data with symptoms and associated factors through combined visualizations for sense-making, and normalizing the inclusion of energy estimates and pacing interventions in apps targeting fatigue.

Acknowledgments

We did not use generative AI to produce any of the content in our manuscript. All our codes, findings, and synthesis are our own work.

Funding

The publishing of this paper was supported by the main author's host institution, through an APF fee waiver obtained from the institution's publishing agreement with JMIR. The funder had no involvement in the study design, data collection, analysis, interpretation, or the writing of the manuscript.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Codebook for the full set of codes, including main codes and all subcodes under the main functionalities of tracking, visualizing, assessing, providing interventions, and ensuring privacy.
[XLSX File, 44 KB - mhealth_v14i1e84755_app1.xlsx]

Multimedia Appendix 2
Apps obtained from searching the Apple App Store and the Google App Store with the keywords "chronic fatigue syndrome", "fatigue", "ME/CFS", "CFS," "myalgic encephalomyelitis," and "pace or pacing."
[XLSX File, 60 KB - mhealth_v14i1e84755_app2.xlsx]

XSL•FO
RenderX

Multimedia Appendix 3
Characteristics of apps selected for review.
[XLSX File, 10 KB - mhealth_v14i1e84755_app3.xlsx]

Multimedia Appendix 4
Frequency of codes for tracking functionality.
[DOCX File, 17 KB - mhealth_v14i1e84755_app4.docx]

Multimedia Appendix 5
Codes for tracking functionality.
[XLSX File, 19 KB - mhealth_v14i1e84755_app5.xlsx]

Multimedia Appendix 6
Frequency of codes for visualization functionality.
[DOCX File, 17 KB - mhealth_v14i1e84755_app6.docx]

Multimedia Appendix 7
Codes for visualization functionality.
[XLSX File, 11 KB - mhealth_v14i1e84755_app7.xlsx]

Multimedia Appendix 8
Frequency of codes for screening and assessment functionality.
[DOCX File, 15 KB - mhealth_v14i1e84755_app8.docx]

Multimedia Appendix 9
Codes for screening and assessment functionality.
[XLSX File, 11 KB - mhealth_v14i1e84755_app9.xlsx]

Multimedia Appendix 10
Frequency of codes for intervention functionality.
[DOCX File, 16 KB - mhealth_v14i1e84755_app10.docx]

Multimedia Appendix 11
Codes for intervention functionality.
[XLSX File, 11 KB - mhealth_v14i1e84755_app11.xlsx]

Multimedia Appendix 12
Frequency of codes for privacy features.
[DOCX File, 17 KB - mhealth_v14i1e84755_app12.docx]

Multimedia Appendix 13
Codes for privacy features.
[XLSX File, 11 KB - mhealth_v14i1e84755_app13.xlsx]

References

1. Crook H, Raza S, Nowell J, Young M, Edison P. Long covid-mechanisms, risk factors, and management. BMJ 2021 Jul 26;374:n1648. [doi: 10.1136/bmj.n1648] [Medline: 34312178]
2. Sas C, Lotankar Y, Adam R, et al. “A Switch Went off in my Whole Body”: lived experiences of fatigue and post-exertional malaise in long Covid. CHI EA ’23: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems 2023:1-7. [doi: 10.1145/3544549.3585846]
3. Bowie-DaBreo D, Sas C, Iles-Smith H, Sünram-Lea S. User perspectives and ethical experiences of apps for depression: a qualitative analysis of user reviews. Presented at: CHI ’22: CHI Conference on Human Factors in Computing Systems; Apr 29 to May 5, 2022; New Orleans, LA, USA p. 1-24. [doi: 10.1145/3491102.3517498]
4. Almoallim S, Sas C. Toward research-informed design implications for interventions limiting smartphone use: functionalities review of digital well-being apps. JMIR Form Res 2022 Apr 19;6(4):e31730. [doi: 10.2196/31730] [Medline: 35188897]

XSL•FO
RenderX

5. Roffarello AM, Russis LD. The race towards digital wellbeing: issues and opportunities. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems 2023:1-14. [doi: 10.1145/3290605.3300616] [Medline: 21775651]
6. Balaskas A, Schueller SM, Cox AL, Doherty G. The functionality of mobile apps for anxiety: systematic search and analysis of engagement and tailoring features. JMIR Mhealth Uhealth 2021 Oct 6;9(10):e26712. [doi: 10.2196/26712] [Medline: 34612833]
7. Paymal L, Homewood S. Good days, bad days: understanding the trajectories of technology use during chronic fatigue syndrome. Presented at: Proceedings of the CHI Conference on Human Factors in Computing Systems; May 11-16, 2024; Honolulu, HI, USA p. 1-10. [doi: 10.1145/3613904.3642553]
8. Best K, Butler S. The necessity of control for computer and internet users with myalgic encephalomyelitis/chronic fatigue syndrome. Inf Commun Soc 2013 Sep;16(7):1042-1060. [doi: 10.1080/1369118X.2012.656139]
9. Davies T, Jones SL, Kelly RM. Patient perspectives on self-management technologies for chronic fatigue syndrome. CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems :1-13. [doi: 10.1145/3290605.3300452]
10. Børsting J, Culén AL. Sense-making in complex healthcare domains: the role of technology in every day lives of youth suffering from ME/CFS. In: eTELEMED 2016, The Eighth International Conference on eHealth, Telemedicine and Social Medicine: International Academy, Research and Industry Association (IARIA); 2016. URL: https://nva.sikt.no/registration/0198cc963f65-bff0a66d-bac8-4653-9f70-4e390be77154 [accessed 2026-07-10]
11. Homewood S, Hinkle CA, Kaklopoulou I. Cripping the co-design of pacing technologies for energy-limiting conditions. Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems 2025:1-16. [doi: 10.1145/3706598.3713990]
12. Best K, Butler S. Second life avatars as extensions of social and physical bodies in people with myalgic encephalomyelitis/chronic fatigue syndrome. Continuum 2013 Dec;27(6):837-849. [doi: 10.1080/10304312.2013.794190]
13. Homewood S. Self-tracking to do less: an autoethnography of long COVID that informs the design of pacing technologies. CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems 2023:1-14. [doi: 10.1145/3544548.3581505]
14. Mack K, McDonnell EJ, Findlater L, Evans HD. Chronically under-addressed: considerations for HCI accessibility practice with chronically ill people. Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility 2022:1-15. [doi: 10.1145/3517428.3544803]
15. Brigden A, Barnett J, Parslow RM, Beasant L, Crawley E. Using the internet to cope with chronic fatigue syndrome/myalgic encephalomyelitis in adolescence: a qualitative study. BMJ Paediatr Open 2018;2(1):e000299. [doi: 10.1136/bmjpo-2018-000299] [Medline: 30167477]
16. Homewood S, Okholm Just K, Bramm Johansson O. The unanticipated use of fitness tracking technologies during post-COVID syndrome. DIS '24: Proceedings of the 2024 ACM Designing Interactive Systems Conference 2024:556-570. [doi: 10.1145/3643834.3661617]
17. Pater J, Chopra S, Carroll J, et al. Charting the COVID long haul experience - a longitudinal exploration of symptoms, activity, and clinical adherence. Presented at: Proceedings of the CHI Conference on Human Factors in Computing Systems; May 11-16, 2024; Honolulu,HI, USA p. 1-21. [doi: 10.1145/3613904.3642827]
18. Mehdipour F, Aharari A. Enhancing COVID-19 data analysis through HCI-driven AI and iot integration. Presented at: 2024 IEEE 13th Global Conference on Consumer Electronics (GCCE); Oct 29 to Nov 1, 2024; Kitakyushu, Japan p. 840-843. [doi: 10.1109/GCCE62371.2024.10760844]
19. Kelly JT, Campbell KL, Gong E, Scuffham P. The Internet of Things: impact and implications for health care delivery. J Med Internet Res 2020 Nov 10;22(11):e20135. [doi: 10.2196/20135] [Medline: 33170132]
20. Adams P, Murnane EL, Elfenbein M, Wethington E, Gay G. Supporting the self-management of chronic pain conditions with tailored momentary self-assessments. Proc SIGCHI Conf Hum Factor Comput Syst 2017:1065-1077. [doi: 10.1145/3025453.3025832]
21. Singh A, Bianchi-Berthouze N, Williams AC. Supporting everyday function in chronic pain using wearable technology. CHI ’17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems 2017:3903-3915. [doi: 10.1145/3025453.3025947]
22. Zhu J, Lolla S, Agnihotri M, et al. A systematic review and meta-analysis of research on goals for behavior change. Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems 2025:1-25. [doi: 10.1145/3706598.3714072]
23. Sankaran S, Dendale P, Coninx K. Evaluating the impact of the HeartHab App on motivation, physical activity, quality of life, and risk factors of coronary artery disease patients: multidisciplinary crossover study. JMIR Mhealth Uhealth 2019 Apr 4;7(4):e10874. [doi: 10.2196/10874] [Medline: 30946021]
24. Sefidgar YS, Castillo CL, Chopra S, et al. MigraineTracker: examining patient experiences with goal-directed self-tracking for a chronic health condition. Proc SIGCHI Conf Hum Factor Comput Syst 2024 May;2024:1-19. [doi: 10.1145/3613904.3642075] [Medline: 38741616]

XSL•FO
RenderX

25. Tang LM, Kay J. Harnessing long term physical activity data—how long-term trackers use data and how an adherence-based interface supports new insights. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Jun 30;1(2):1-28. [doi: 10.1145/3090091]
26. Geurts E, Geel FV, Feys P, Coninx K. WalkWithMe: personalized goal setting and coaching for walking in people with multiple sclerosis. UMAP '19: Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization :51-60. [doi: 10.1145/3320435.3320459]
27. Michael A, Lutteroth C. Race yourselves: a longitudinal exploration of self-competition between past, present, and future performances in a VR exergame. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems :1-17. [doi: 10.1145/3313831.3376256]
28. Xue M, Liang RH, Yu B, Funk M, Hu J, Feijs L. AffectiveWall: designing collective stress-related physiological data visualization for reflection. IEEE Access 2019;7:131289-131303. [doi: 10.1109/ACCESS.2019.2940866]
29. Felipe S, Singh A, Bradley C, Williams AC, Bianchi-Berthouze N. Roles for personal informatics in chronic pain. Presented at: 9th International Conference on Pervasive Computing Technologies for Healthcare; May 20-23, 2015; Istanbul, Turkey p. 161-168. [doi: 10.4108/icst.pervasivehealth.2015.259501]
30. Epstein DA, Caldeira C, Figueiredo MC, et al. Mapping and taking stock of the personal informatics literature. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020 Dec 17;4(4):1-38. [doi: 10.1145/3432231]
31. Ruehlman LS, Karoly P, Enders C. A randomized controlled evaluation of an online chronic pain self management program. PAIN 2012 Feb;153(2):319-330. [doi: 10.1016/j.pain.2011.10.025] [Medline: 22133450]
32. Schroeder J, Chung CF, Epstein DA, et al. Examining self-tracking by people with migraine: goals, needs, and opportunities in a chronic health condition. DIS '18: Proceedings of the 2018 Designing Interactive Systems Conference 2018:135-148.
33. Guluzade L, Sas C. Functionality and user review analysis of mobile apps for mindfulness eating and eating disorders. DIS '24: Proceedings of the 2024 ACM Designing Interactive Systems Conference :1350-1371. [doi: 10.1145/3643834.3661521]
34. Qu C, Sas C, Daudén Roquet C, Doherty G. Functionality of top-rated mobile apps for depression: systematic search and evaluation. JMIR Ment Health 2020 Jan 24;7(1):e15321. [doi: 10.2196/15321] [Medline: 32012079]
35. Lolla S, Sas C. Evaluating mobile apps targeting personal goals. CHI EA '23: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems 2023:1-7. [doi: 10.1145/3544549.3585725]
36. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
37. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016 Mar 10;13(1):35. [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
38. Richardson B, Dol J, Rutledge K, et al. Evaluation of mobile apps targeted to parents of infants in the neonatal intensive care unit: systematic app review. JMIR Mhealth Uhealth 2019 Apr 15;7(4):e11620. [doi: 10.2196/11620] [Medline: 30985282]
39. Giunti G, Guisado Fernández E, Dorronzoro Zubiete E, Rivera Romero O. Supply and demand in mHealth apps for persons with multiple sclerosis: systematic search in app stores and scoping literature review. JMIR Mhealth Uhealth 2018 May 23;6(5):e10512. [doi: 10.2196/10512] [Medline: 29792295]
40. Ferrara G, Kim J, Lin S, Hua J, Seto E. A focused review of smartphone diet-tracking apps: usability, functionality, coherence with behavior change theory, and comparative validity of nutrient intake and energy estimates. JMIR Mhealth Uhealth 2019 May 17;7(5):e9232. [doi: 10.2196/mhealth.9232] [Medline: 31102369]
41. Rezaei Nasab A, Dashti M, Shahin M, et al. Fairness concerns in app reviews: a study on AI-based mobile apps. ACM Trans Softw Eng Methodol 2025 Feb 28;34(2):1-30. [doi: 10.1145/3690633]
42. Hossain E, Ahmetoglu Y, Chen S, et al. Mapping the landscape of functionality reviews in HCI: trends, methods, and contributions. Interact Comput 2025 Apr. [doi: 10.1093/iwc/iwaf027]
43. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J Qual Methods 2006 Mar;5(1):80-92. [doi: 10.1177/160940690600500107]
44. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960 Apr;20(1):37-46. [doi: 10.1177/001316446002000104]
45. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [doi: 10.2307/2529310] [Medline: 843571]
46. Visible: pacing for illness. : Visible Health Inc. URL: https://apps.apple.com/gb/app/visible-pacing-for-illness/id1624474919 [accessed 2024-12-01]
47. Capuron L, Welberg L, Heim C, et al. Cognitive dysfunction relates to subjective report of mental fatigue in patients with chronic fatigue syndrome. Neuropsychopharmacology 2006 Aug;31(8):1777-1784. [doi: 10.1038/sj.npp.1301005] [Medline: 16395303]
48. Park JW, Park BJ, Lee JS, Lee EJ, Ahn YC, Son CG. Systematic review of fatigue severity in ME/CFS patients: insights from randomized controlled trials. J Transl Med 2024 Jun 3;22(1):529. [doi: 10.1186/s12967-024-05349-7] [Medline: 38831460]

XSL•FO
RenderX

49. Dafoe W. Extremely severe ME/CFS-a personal account. Healthcare (Basel) 2021 Apr 27;9(5):504. [doi: 10.3390/healthcare9050504] [Medline: 33925566]
50. Petty S, Ellis A. The meaning of autistic movements. Autism 2024 Dec;28(12):3015-3020. [doi: 10.1177/13623613241262151] [Medline: 38907717]
51. Leekam SR, Prior MR, Uljarevic M. Restricted and repetitive behaviors in autism spectrum disorders: a review of research in the last decade. Psychol Bull 2011 Jul;137(4):562-593. [doi: 10.1037/a0023341] [Medline: 21574682]
52. Keville S, Meek C, Ludlow AK. Mothers' perspectives of co-occurring fatigue in children with autism spectrum disorders. Fatigue: Biomed Health Behav 2021 Oct 2;9(4):209-226. [doi: 10.1080/21641846.2021.2008169]
53. Fraser H, Crossland D, Bacher I, Ranney M, Madsen T, Hilliard R. Comparison of diagnostic and triage accuracy of Ada Health and WebMD symptom checkers, ChatGPT, and Physicians for Patients in an emergency department: clinical data analysis study. JMIR Mhealth Uhealth 2023 Oct 3;11(1):e49995. [doi: 10.2196/49995] [Medline: 37788063]
54. Miller S, Gilbert S, Virani V, Wicks P. Patients' utilization and perception of an artificial intelligence-based symptom assessment and advice technology in a British primary care waiting room: exploratory pilot study. JMIR Hum Factors 2020 Jul 10;7(3):e19713. [doi: 10.2196/19713] [Medline: 32540836]
55. Komaroff AL, Buchwald D. Symptoms and signs of chronic fatigue syndrome. Rev Infect Dis 1991;13 Suppl 1(Supplement_1):S8-11. [doi: 10.1093/clinids/13.supplement_1.s8] [Medline: 2020806]
56. Doherty K, Balaskas A, Doherty G. The design of ecological momentary assessment technologies. Interact Comput 2020 May 17;32(3):257-278. [doi: 10.1093/iwcomp/iwaa019]
57. Sharpe MC, Archard LC, Banatvala JE, et al. A report--chronic fatigue syndrome: guidelines for research. J R Soc Med 1991 Feb;84(2):118-121. [doi: 10.1177/014107689108400224] [Medline: 1999813]
58. Sommerfelt K, Schei T, Seton KA, Carding SR. Assessing functional capacity in myalgic encephalopathy/chronic fatigue syndrome: a patient-informed questionnaire. J Clin Med 2024 Jun 14;13(12):3486. [doi: 10.3390/jcm13123486] [Medline: 38930014]
59. Ataman R, Alhasani R, Auneau-Enjalbert L, Quigley A, Michael HU, Ahmed S. The psychometric properties of the Quality of Life in Neurological Disorders (Neuro-QoL) measurement system in neurorehabilitation populations: a systematic review. J Patient Rep Outcomes 2024 Sep 18;8(1):106. [doi: 10.1186/s41687-024-00743-7] [Medline: 39292414]
60. Cella D, Lai JS, Nowinski CJ, et al. Neuro-QOL: brief measures of health-related quality of life for clinical research in neurology. Neurology 2012 Jun 5;78(23):1860-1867. [doi: 10.1212/WNL.0b013e318258f744] [Medline: 22573626]
61. Doherty K, Doherty G. The construal of experience in HCI: understanding self-reports. Int J Hum Comput Stud 2018 Feb;110:63-74. [doi: 10.1016/j.ijhcs.2017.10.006]
62. Miller R. Neuroeducation: integrating brain-based psychoeducation into clinical practice. J Ment Health Couns 2016 Apr 1;38(2):103-115. [doi: 10.17744/mehc.38.2.02]
63. Gupta program brain retraining. : Harley Street Solutions LTD URL: https://apps.apple.com/gb/app/gupta-program-brain-retraining/id6449983590 [accessed 2024-12-01]
64. Steidle A, Gonzalez-Morales MG, Hoppe A, Michel A, O'shea D. Energizing respites from work: a randomized controlled study on respite interventions. Eur J Work Organ Psychol 2017 Sep 3;26(5):650-662. [doi: 10.1080/1359432X.2017.1348348]
65. Miserandino C. The Spoon Theory. ButYouDontLookSick. URL: https://web.archive.org/web/20250330181714/https://butyoudontlooksick.com/articles/written-by-christine/the-spoon-theory/ [accessed 2025-04-03]
66. Jason LA. The PACE trial missteps on pacing and patient selection. J Health Psychol 2017 Aug;22(9):1141-1145. [doi: 10.1177/1359105317695801] [Medline: 28805518]
67. Wiener CL. The burden of rheumatoid arthritis: tolerating the uncertainty. Soc Sci Med 1975 Feb;9(2):97-104. [doi: 10.1016/0037-7856(75)90101-8] [Medline: 1129605]
68. Goudsmit EM, Nijs J, Jason LA, Wallman KE. Pacing as a strategy to improve energy management in myalgic encephalomyelitis/chronic fatigue syndrome: a consensus document. Disabil Rehabil 2012;34(13):1140-1147. [doi: 10.3109/09638288.2011.635746] [Medline: 22181560]
69. Elite HRV: Wellness & Fitness. : Elite HRV Inc URL: https://apps.apple.com/gb/app/elite-hrv-wellness-fitness/id868829970 [accessed 2024-12-01]
70. Gonzalez K, Sasangohar F, Mehta RK, Lawley M, Erraguntla M. Measuring fatigue through heart rate variability and activity recognition: a scoping literature review of machine learning techniques. Proc Hum Factors Ergon Soc Annu Meet 2017 Sep;61(1):1748-1752. [doi: 10.1177/1541931213601918]
71. Zhang C, Yu X. Estimating mental fatigue based on electroencephalogram and heart rate variability. Pol J Med Phys Eng 2010 Jan 1;16(2):67-84. [doi: 10.2478/v10013-010-0007-7]
72. SpoonieDay. : Blackburn Labs URL: https://apps.apple.com/gb/app/spoonieday/id1510762724 [accessed 2024-12-01]
73. ME/CFS pacing. : Emerge Australia Inc URL: https://apps.apple.com/gb/app/me-cfs-pacing/id1566816418 [accessed 2024-12-01]
74. Ašeriškis D, Damaševičius R. Gamification patterns for gamification applications. Procedia Comput Sci 2014;39:83-90. [doi: 10.1016/j.procs.2014.11.013]

XSL•FO
RenderX

75. Massung E, Coyle D, Cater KF, Jay M, Preist C. Using crowdsourcing to support pro-environmental community activism. Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; Apr 27 to May 2, 2013; Paris, France p. 371-380. [doi: 10.1145/2470654.2470708]
76. Sellitto G, Morelli A, Bassano S, et al. Outcome measures for physical fatigue in individuals with multiple sclerosis: a systematic review. Expert Rev Pharmacoecon Outcomes Res 2021 Aug;21(4):625-646. [doi: 10.1080/14737167.2021.1883430] [Medline: 33504225]
77. Fisk JD, Ritvo PG, Ross L, Haase DA, Marrie TJ, Schlech WF. Measuring the functional impact of fatigue: initial validation of the fatigue impact scale. Clin Infect Dis 1994 Jan;18 Suppl 1(Supplement_1):S79-S83. [doi: 10.1093/clinids/18.supplement_1.s79] [Medline: 8148458]
78. Samn S, Perelli L. Estimating aircrew fatigue: a technique with application to airlift operations. : US Air Force School OF Aerospace Medicine; 1982 Dec URL: https://www.researchgate.net/publication/235150215_Estimating_Aircrew_Fatigue_A_Technique_with_Application_to_Airlift_Operations [accessed 2026-07-10]
79. Greenberg S, Aislinn P, Kirsten D. Development and validation of the fatigue state questionnaire: preliminary findings. Open Psychol J 2016 Jun 30;9(1):50-65. [doi: 10.2174/1874350101609010050]
80. Adam R, Lotankar Y, Sas C, et al. Understanding patterns of fatigue in health and disease: protocol for an ecological momentary assessment study using digital technologies. BMJ Open 2024 May 27;14(5):e081416. [doi: 10.1136/bmjopen-2023-081416] [Medline: 38802273]
81. Gudi A, Bittner M, van Gemert J. Real-time webcam heart-rate and variability estimation with clean ground truth for evaluation. Appl Sci 2020;10(23):8630. [doi: 10.3390/app10238630]
82. Bowman R, Cooney O, Thieme A, Cowan BR, Doherty G. Pervasive therapy: designing conversation-based interfaces for ecological momentary intervention. IEEE Pervasive Comput 2022;21(4):55-63. [doi: 10.1109/MPRV.2022.3209068]
83. Sun Y, Sundar SS. Exploring the effects of interactive dialogue in improving user control for explainable online symptom checkers. CHI EA '22: CHI Conference on Human Factors in Computing Systems Extended Abstracts 2022:1-7. [doi: 10.1145/3491101.3519668]

Abbreviations

CFS: chronic fatigue syndrome
ECG: electrocardiogram
ELC: energy-limiting condition
FSQ: Fatigue State Questionnaire
HCI: human-computer interaction
HRV: heart rate variability
IoT: Internet of Things
MFIS: Modified Fatigue Impact Scale
ML: machine learning
Neuro QOL: Quality of Life in Neurological Disorders
NLP: natural language processing
PEM: postexertional malaise
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
VR: virtual reality

Edited by A Stone, L Buis; submitted 25.Sep.2025; peer-reviewed by G Varesco, R Marshall; revised version received 12.Apr.2026; accepted 23.Jun.2026; published 11.Aug.2026.

Please cite as:
Abdulla AD, Sas C, Doherty G
Functionality Review of Mobile Apps for the Tracking and Self-Management of Fatigue: Systematic Search in App Stores and Content Analysis
JMIR Mhealth Uhealth 2026;14:e84755
URL: https://mhealth.jmir.org/2026/1/e84755
doi:10.2196/84755

© Amr Diouf Abdulla, Corina Sas, Gavin Doherty. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,

XSL•FO
RenderX

provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Smartphone Apps for Preventing Adolescent Health Problems Among Health Care Professionals: Systematic Search and Quality Assessment

Elodie Million[1,2], MD, PhD; Marijn Kuijper[1], MD; Carole Bautista[1], MD; Lola Pirat[1], MD; Bruno Falissard[3], MD, PhD; Béatrice Lognos[1,2], MD, PhD; François Carbonnel[1,2], MD, PhD

[1]Department of General Practice, University of Montpellier, Campus Arnaud de Villeneuve, 641 Av. du Doyen Gaston Giraud, Montpellier, France
[2]Desbrest Institute of Epidemiology and Public Health (IDESP), UMR UA11 INSERM, University of Montpellier, Montpellier, France
[3]CESP (Centre for Research in Epidemiology and Population Health, INSERM), University Paris Sarclay, Paris, France

Corresponding Author:
Elodie Million, MD, PhD
Department of General Practice, University of Montpellier, Campus Arnaud de Villeneuve, 641 Av. du Doyen Gaston Giraud, Montpellier, France

Abstract

Background: Health care professionals must consider multiple dimensions of prevention when consulting with adolescents. Identifying risky behaviors early in adolescence is crucial for reducing both morbidity and mortality. General practitioners are increasingly eager to incorporate digital tools for prevention into their consultations with adolescents; however, the relevance and clinical validity of these digital tools are not always established or well-known. Consequently, primary care professionals require guidance and support in selecting relevant mobile health (mHealth) tools.

Objective: The aim of this study is to identify relevant and useful digital apps to help primary care professionals detect at-risk adolescents across all recommended areas of prevention: orthopedics, mental health, substance abuse, risk behaviors, sexual health, vaccinations, social relationships, and nutrition.

Methods: A systematic review of smartphone apps, with an analysis of content quality, was carried out by 4 researchers using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist. The App Store and Google Play Store platforms were surveyed. The inclusion criteria were as follows: free of charge, date of last update, availability in French or English, relevance of the preventive approach to adolescents, and scientific validation. Four health care professionals assessed the apps: 2 selected the apps relevant to health care professionals, then 3 analyzed these apps using the French version of the Mobile App Rating Scale (MARS-F). Intraclass correlation coefficient, model (2,1) (2-way random effects, absolute agreement, single measures); standard error of measurement; and mean absolute error were also calculated.

Results: A total of 976 apps were identified, 49 of which had disappeared from the platforms prior to analysis. Nine apps were retained. Seven (0.72%) were included after evaluation using the MARS-F: 2 on mental health and 5 on sexual health (including 3 on contraception only). The mean MARS-F interrater score ranged from 2.5/5 to 3.8/5. The global MARS-F score demonstrated a pooled SD of 0.60 and an intraclass correlation coefficient (2,1) of 0.0003, resulting in a calculated standard error of measurement of 0.60. The average discrepancy between raters was a mean absolute error of 0.53.

Conclusions: No similar studies have been identified in the literature that specifically focus on mobile apps designed to support health care professionals in delivering preventive care to adolescents. Of the 8 areas of prevention identified as relevant for adolescents, only 3 are addressed by the apps validated through our methodology (5 focus on sexual health). Consequently, current apps are insufficient to support health care professionals in their overall preventive work with adolescents. Such a review should be conducted systematically prior to the development of any new tool to prevent duplication and channel creative efforts toward truly innovative digital solutions. Furthermore, a thorough analysis of relevant, recommended websites is essential, as these resources complement the use of mobile apps designed for health care professionals.

(JMIR Mhealth Uhealth 2026;14:e83908) doi:10.2196/83908

KEYWORDS

adolescent; primary care; preventive medicine; mobile apps; mental health; sexual health

XSL•FO
RenderX

Introduction

Adolescence is a period of transition from childhood to adulthood. Adolescents are subject to many biopsychosocial changes, including the desire for autonomy and experimentation, the need to identify with peers, and risk-taking [1-6]. Health care professionals must consider a wide range of preventive domains during consultations with adolescents: sexual health, mental health, substance abuse and addiction, diet, physical activity, orthopedic disorders, vaccinations, screen use, and social relations [6-11]. Preventive actions are difficult for health care professionals to implement because of the complexity of consultations with adolescents [10-15]. Early identification of risky behaviors in adolescents is essential for reducing morbidity and mortality [11,16]. Although various recommendations and predominantly nondigital tools exist to support adolescent preventive care during consultations [8,9,14,17,18], health care professionals often find them difficult to implement effectively in practice.

The World Health Organization (WHO) defines eHealth as "digital services for human well-being," that is, the application of information and communication technologies to health and well-being [19,20]. Mobile health (mHealth) covers medical and public health practices based on mobile devices (mobile phones, patient monitoring systems, personal digital assistants, and other wireless devices) [19,21]. The development of health-focused websites and smartphone apps has contributed significantly to improving both the quality of care and patient health outcomes [20]. Digital tools are used in routine care by health care professionals and patients. In 2021, over 327,000 mHealth apps were listed in digital stores [21]. French general practitioners generally view mHealth apps and devices favorably and express readiness to integrate them into their clinical workflows [15,22]. In a qualitative study carried out among French general practitioners, the need for a preventive tool in consultations with adolescents was identified, with a preference for a practical digital tool [15]. Despite the desire of general practitioners to use these digital tools, the relevance of these tools is not always established or known [15,17].

Many apps are easy to use during consultations, and general practitioners often do not have the skills or time to select from the wide range on offer. General practitioners need guidance to choose relevant apps. The main aim of this study was to identify relevant and useful digital apps to help primary care professionals detect at-risk adolescents across all recommended areas of prevention: orthopedics, mental health, substance abuse, risk behaviors, sexual health, vaccinations, social relationships, and nutrition.

Methods

Study Design

A systematic review of smartphone apps with an analysis of content quality was carried out by 4 researchers. One general practitioner resident (CB) and 1 experienced general practitioner (EM) carried out the search for apps on search engines and the initial selection. Two general practitioner residents (MK and LP) and 1 experienced general practitioner (EM) evaluated the apps using the French version of the Mobile App Rating Scale (MARS-F) [23]. The study protocol was designed following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (items not relevant to a systematic search of apps were considered not applicable; Checklist 1) and the PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Literature Search Extension) guidelines (Checklist 2) [24,25]. We took into account the specialized reporting items suggested by the PASSR (Protocol for App Store Systematic Reviews) guidelines (Checklist 3) to address the unique challenges of searching and analyzing mobile app stores [26]. The complete and detailed study protocol is provided in Multimedia Appendix 1.

Search Strategy

A systematic search was performed on the 2 major digital platforms: the Apple App Store and the Google Play Store. These platforms are used by approximately 99% of smartphone users to download their apps [27]. Two similar searches, using the same method, were conducted between August 2022 and March 2023, and then between March 2023 and May 2023. The platforms were queried in French and English using keywords related to adolescent prevention areas derived from MeSH terms [28]. MeSH terms were preidentified in a literature review according to prevention themes (Table 1). MeSH terms were replaced with non-MeSH synonyms when a search did not find any relevant results. The choice of words was discussed between the 2 researchers, CB and EM. Each keyword was used individually on the app platforms. CB and EM carried out the search on an iPhone 11 smartphone running iOS 15 (Apple), an ASUS VivoBook computer (ASUSTeK Computer Inc), and a Google Pixel 7 smartphone with the S91xBXXS1AWD1 update. The search ended when the search engines no longer provided new apps: keywords were entered into the search tool, followed by an exhaustive review of the descriptions of the apps offered.

XSL•FO
RenderX

Table . English keywords.

Prevention areas	English keywords
Orthopedics	• Kyphosis • Spinal curvatures • Vertebral body • Lordosis • Spinal or vertebral column • Scoliosis • Back pain • Osgood-Schlatter disease[a] • Osteochondrosis • Scheuermann disease[a] • Patellofemoral pain syndrome
Substance abuse, screens, and addictions	• Alcohol • Nicotiana • Drug • Cannabis • Heroin prevention[b] • Cocaine • Ecstasy • Screen time
Mental health	• Sleep-wake disorders • Anxiety • Depression • Suicide • Attention-deficit/hyperactivity disorder • Psychosis • Autism
Nutrition	• Obesity • Deficiency diseases • Feeding and eating disorders
Sexual health and puberty	• Sexually transmitted infection • Unsafe sex • Contraception • Puberty
Social relations and violence	• School dropout • Harassment, bullying[a] • Friend • Violence
Vaccinations	• Vaccination

[a]Non-MeSH words.

[b]The term “heroin” was not retained because the search results yielded only games.

Selection Process and Eligibility Criteria

The selection by EM and CB followed a 2-stage process. First, EM and CB screened the apps based on their titles, descriptions, and metadata provided in the app stores. Second, the shortlisted apps were downloaded and tested for full eligibility. The initial selection of apps was based on inclusion and exclusion criteria from digital platforms. In cases of doubt about an exclusion criterion, the decision was made through discussions between CB and EM.

To meet our inclusion criteria, the apps had to (1) relate to one of the selected areas of prevention (orthopedics, mental health, substance abuse, risk behaviors, sexual health, vaccinations, social relations, nutrition); (2) be relevant to facilitating the preventive approach of health care professionals during consultations with adolescents; (3) be available in French or English; (4) have been scientifically validated (developed by health care professionals or a learned medical society and/or based on scientific or government recommendations); (5) be free (apps free for only a trial period were not selected); (6) have been updated within the 3 years preceding the end of the search, at least by 2020 (since 2022, the Google Play Store has removed apps that have not been updated for more than 2 years [29], and for the App Store, this period is 3 years for apps with few downloads in the last 12 months [30] but not all catalogs have yet been updated).

XSL•FO
RenderX

The exclusion criteria were as follows: (1) lack of information about the scientific data used to design the app's content, (2) games without medicoeducational content, (3) wallpaper apps, (4) religious apps, (5) apps that offer links to medical appointments, and (6) advertising apps.

Inclusion criteria were based on the data available on each app's presentation page: the title, description, images, and general information about the app. If any data were missing, the apps were opened to search for the missing information. Apps were excluded after download if they were no longer available in app stores at the time of evaluation by one of the evaluators. We chose to focus on free apps to ensure that the tools identified are accessible to all health care professionals, regardless of institutional funding or individual socioeconomic status, thereby promoting digital health equity regardless of which health care professional a patient consults [31]. Health care professionals have already invested in costly practice tools such as business software and telecare tools. We therefore wanted to offer them apps with no impact on their professional workloads. The high cost of an app is one of the main reasons why it is not downloaded [32,33]. Selecting free apps also limits the risk of commercial bias or conflicts of interest that can be associated with paid models.

As these digital tools are intended for use in clinical consultations, scientific rigor is nonnegotiable. Using unvalidated tools poses an ethical risk to the patient-provider relationship and the safety of care.

Quality Appraisal and Data Extraction

MK, LP, and EM evaluated apps using the MARS-F [23]. The MARS and its French version (MARS-F) have been validated for the overall quality assessment of apps in the field of mHealth [23,34,35]. The app evaluations took place between December 2023 and January 2024.

The MARS-F assesses four objective quality dimensions: (1) Engagement (entertainment, interest, customization, interactivity, and target group), (2) Functionality (performance, ease of use, navigation, and gestural design), (3) Esthetics (layout, graphics, and visual appeal), and (4) Information (description accuracy, goals, information quality or quantity, visual information, credibility, and evidence base). A subjective quality score and an app-specific section were also proposed. Each item is rated on a Likert scale from 1 ("inadequate") to 5 ("excellent").

When specific objectives were not delineated and information was lacking, the respective items were rated as "not attributable" (NA). Consequently, these items were excluded from consideration in the overall scoring process. In total, MARS-F is a total score out of 5, corresponding to the average of these 4 sections. For each evaluator, an average for each section (A, B, C, and D) from 1 to 5 was calculated, followed by an interevaluator average for each section. An overall average of the interrater MARS-F score was used to evaluate the apps. To ensure the selected apps were suitable for clinical practice and met the needs of health care professionals working with adolescents, specific quality cut-off scores were applied during the screening phase. Apps were excluded if they failed to meet the following minimum thresholds across the MARS-F sections (Multimedia Appendix 1). Apps were excluded if MARS-F ratings were below 3 for target group relevance (Item 5, Section A), below 2 for all technical usability items (Section B: Functionality), or if ratings were insufficient within the Information Quality domain (Section D: Item 15 <2 or NA, Item 16 <2, Item 17 <4). During the quality appraisal, if an exclusion criterion was found by only 1 of the 2 evaluators for the final evaluation score, the third researcher (BL) was asked to discuss the results, and consensus was sought.

From the apps selected on digital platforms by EM and CB, 3 researchers took part in evaluating them: the 2 general practitioner residents (MK and LP) divided up the prevention themes for analysis, and the experienced general practitioner (EM) analyzed all the apps in a blinded manner. All 3 completed the MARS-F [23] separately. A common table was then filled based on these MARS-F results. To ensure rating consistency, LP and MK underwent a structured calibration process led by the veteran expert (EM). During the 2-session calibration process, informal reliability checks were conducted to ensure rater alignment. The 3 researchers independently scored a pilot sample of test applications. Discrepancies and borderline cases were thoroughly discussed until a 100% consensus was reached on all criteria. However, no formal statistical reliability metrics, such as a training intraclass correlation coefficient (ICC), were computed during this preliminary phase." The training consisted of two 2-hour sessions. The first session involved a comprehensive review of each MARS-F item and its scoring criteria. In the second session, conducted 3 weeks later, the 3 researchers independently and blindly evaluated 2 apps randomly selected from the initial search results. Discrepancies were discussed collectively, and scoring rules were clarified under the supervision of the researcher EM. Throughout the formal evaluation period, raters had access to expert guidance to resolve technical queries. Specific apps were not discussed until the final consensus meeting to maintain blinded assessment integrity. While a formal training ICC was not calculated, this iterative calibration was designed to minimize subjective bias.

Statistical Analysis

MARS-F scores were presented as means and SDs. To assess the reliability of the MARS-F ratings (global and subscales), we used a multimetric approach. Interrater reliability was calculated using the ICC, model (2,1) (2-way random effects, absolute agreement, single measures), along with the SD. To address measurement precision, we calculated the standard error of measurement (SEM) using the formula $SEM=SD\times\sqrt{(1-ICC)}$. Descriptive statistics, including the mean and SD, were generated for each app and for each global scale or subscale to characterize the distribution of quality ratings. We analyzed the expert's ratings (EM) against the residents' ratings (MK or LP). Since the ICC is a ratio of intersubject variance to total variance, the lack of variability between subjects in a small sample could mathematically deflate the ICC, even when raters are in close agreement. To provide a more accurate assessment of rater reliability, we also calculated the mean absolute error (MAE). The MAE represents the actual average point difference between raters on the 5-point scale.

XSL•FO
RenderX

An ICC above 0.75 was considered excellent, good between 0.60 and 0.74, fair between 0.40 and 0.59, and poor if below 0.40. We have applied an interpretation threshold for MAE: excellent (<0.5), good (0.5 - 0.75), fair (0.75 - 1.0), and poor (>1.0).

ICC analyses were performed using JASP (version 0.96; 2026; University of Amsterdam), and MAE, SD, mean, and SEM were calculated in Microsoft Excel.

Ethical Considerations

A request for authorization has been made to the Data Protection Officer of the University of Montpellier. In addition, the study did not require authorization from the "*Comité de la Protection des Personnes*" (Personal Protection Committee), as it did not involve experimentation on individuals and did not fall within the scope of the General Data Protection Regulation and, therefore, of MR004. This study complies with the rules of the Helsinki Declaration [36].

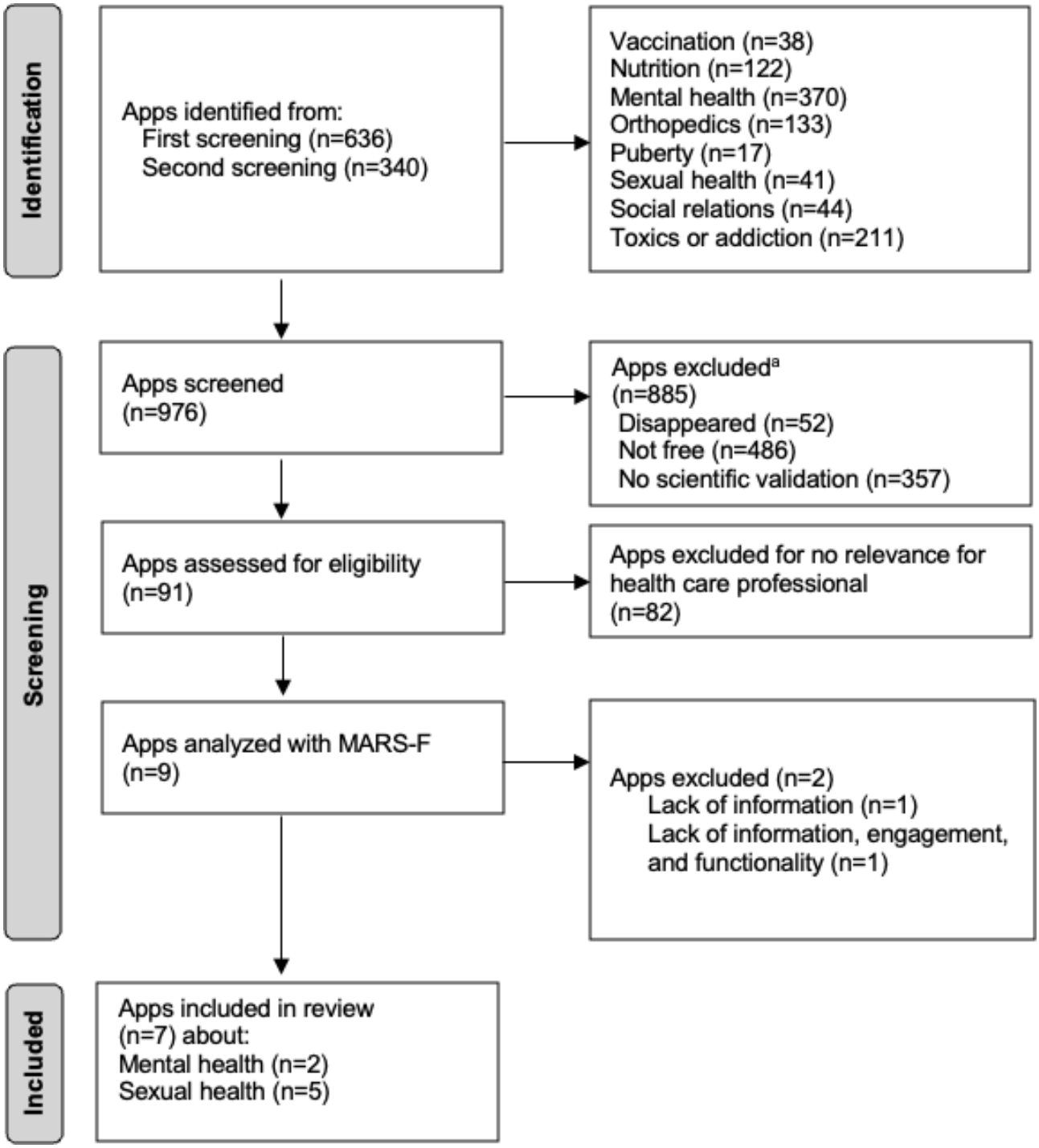
Results

App Selection

Using the preidentified keywords, the double screening identified 976 apps. Forty-seven apps had already disappeared from the platforms between the 2 screenings and at the time of the first analysis by CB and EM; 2 more disappeared during the final evaluations by LP, MK, and EM, giving a total of 52 apps that had disappeared during the search. In the end, 9 apps were retained. After evaluation with the MARS-F, 7 were included (Figure 1). The mean MARS-F interrater score ranged from 2.5/5 to 3.8/5 (Table 2).

XSL•FO
RenderX

Figure 1. Flowchart. [a]Apps can be excluded for several criteria. MARS-F: French version of the Mobile App Rating Scale.



XSL•FO
RenderX

Table . MARS-F[a] score for the 7 included apps.

App evaluated and rater	Section A	Section B	Section C	Section D	Average app score, mean (SD)	Interrater score, mean (SD)
Agenda du sommeil antifatigue [b]						3.5 (0.14)
EM	2.4	4.7	3.3	4.1	3.6 (0.86)	
MK	2.6	4.7	2.6	3.5	3.3 (0.86)	
mhGAP-IG 2.0 App (e-mhGAP)						3.6 (0.25)
EM	2.6	4	3	4	3.4 (0.61)	
MK	3.8	4.5	3.3	4	3.9 (0.43)	
WHO[c] Contraception Tool						3.8 (0.05)
EM	2.6	4.2	4.3	4.5	3.9 (0.75)	
LP	3.6	3.7	3.6	4	3.7 (0.16)	
Contraception Point-of-Care						2.5 (0.08)
EM	1.6	3.7	2	3.6	2.7 (0.93)	
LP	1.6	3.2	2.6	2.1	2.4 (0.59)	
Contraception Eligibility Tool						3.1 (0.15)
EM	2.6	4.5	3	4.2	3.5 (0.79)	
LP	2	3	3	3	2.7 (0.43)	
Practice Family Planning						3.0 (0.41)
EM	2.6	4.7	3.3	3.6	3.5 (0.75)	
LP	2.6	2.5	2	3.4	2.6 (0.50)	
Learn Family Planning						3.1 (0.47)
EM	2.6	4.5	3.3	4	3.6 (0.71)	
LP	2.6	2.5	2	3.4	2.6 (0.50)	

[a]MARS-F: French version of the Mobile App Rating Scale.

[b]App is available only in French.

[c]WHO: World Health Organization.

Quality of Apps Based on MARS-F

The evaluation scores for the 9 evaluated apps were listed in a table (Multimedia Appendix 2), and a summary of the content of the apps was produced for the apps not excluded from the analysis (Multimedia Appendix 3). We provide a summary of the MARS-F scores for the 7 included apps (Table 2) and a short summary of these apps.

Some items in MARS-F have been rated NA. We describe the NA items present in the apps evaluated by MARS-F in Table 3.

XSL•FO
RenderX

Table . Summary of quotes[a].

MARS-F[b] item	*Agenda du sommeil antifatigue*	mhGAP-IG 2.0 App (e-mhGAP)	WHO[c] Contraception Tool	Contraception Point-of-Care	Contraception Eligibility Tool	Practice Family Planning	Learn Family Planning	*Bariatrique Québec*	USPSTF Prevention Task-Force	NA[d]	NA, %
Item 14 (Goals)	0	1	1	0.5	1	1	1	0	0.5	6	66
Item 17 (Visual information)	0	0	1	0	1	0	0	0	0	2	22
Item 19 (Evidence base)	1	1	1	1	1	1	1	1	1	9	100

[a]"1" represents an NA rating by both evaluators independently, "0" represents no NA for this item, and "0.5" represents an NA rating by only 1 evaluator independently.

[b]MARS-F: French version of the Mobile App Rating Scale.

[c]WHO: World Health Organization.

[d]NA: not attributable.

Reliability Evaluation

The global MARS-F score demonstrated a pooled SD of 0.60 and an ICC (2,1) of 0.0003, resulting in a calculated SEM of 0.60. The average discrepancy between raters remained low, with an MAE of 0.53. Across subscales, the SEM ranged from 0.41 (Engagement) to 0.79 (Functionality) (Table 4).

Table . Reliability table[a].

Scale or subscale	ICC[b] (95% CI)	Global SD	Min-Max	SEM[c]	MAE[d]
Global MARS-F[e]	0.0003 (−0.51 to 0.61)	0.60	2.55 - 3.92	0.60	0.53
Engagement	0.54 (−0.12 to 0.88)	0.61	1.60 - 3.80	0.41	0.4
Functionality	0.00 (−0.46 to 0.61)	0.79	2.50 - 4.75	0.79	1.03
Esthetics	0.33 (−0.35 to 0.79)	0.70	1.66 - 4.33	0.57	0.75
Information	0.26 (−0.15 to 0.72)	0.59	2.66 - 4.50	0.51	0.66

[a]ICC (2.1) values are affected by range restriction (low intersubject variance) and low number of apps. MAE is provided to complement ICC analysis and clarify reliability.

[b]ICC: intraclass correlation coefficient.

[c]SEM: standard error of measurement

[d]MAE: mean absolute error.

[e]MARS-F: French version of the Mobile App Rating Scale.

The interrater reliability analysis comparing the veteran expert (EM) and general practitioner residents (MK and LP) yielded variable results across the MARS-F subscales (Table 4). As expected, the ICC (2,1) for the overall MARS-F score was poor at 0.0003 (95% CI −0.51 to 0.61). While ICC values for several subscales remained poor, ranging from 0.000 to 0.540, these findings are largely attributable to range restriction caused by the high homogeneity of app quality scores in our small sample (n=9). To account for this statistical artifact, we calculated an MAE to assess absolute agreement. The MAE for the global score was 0.53 (good), indicating that ratings differed by approximately half a point on the 5-point scale, which represents good clinical agreement.

Description of the Apps

Of the 976 apps identified, 37.9% (ie, 370) and 21.6% (ie, 211) were found using mood-related keywords and toxic consumption-related keywords, respectively, compared with only 13.6% (ie, 133) and 3.9% (ie, 38) for orthopedics and vaccination. Nine apps were subjected to double-blind analysis using the MARS-F, and 7 were included. The "*Bariatrique Québec*" app was excluded from Section D (Information); the "USPSTF Prevention TaskForce" app was excluded from 3 sections: A (Engagement), B (Functionality), and D (Information). Among the 7 apps selected from the 976 apps identified (0.72%), 71.4% (ie, 5) were related to sexual health, all of these sexual health apps (100%) dealt with contraception, and 60% of these apps (ie, 3) dealt very briefly with sexually transmitted infections (1 dealt with antiretrovirals and 2 with condom use). About 28.6% of the selected apps (ie, 2) were relevant to mental health. No apps were selected for the following themes: vaccination, nutrition, orthopedics, puberty, social or toxic relationships, and addictions.

XSL•FO
RenderX

We summarized the 7 apps selected for use by health care professionals to support prevention initiatives during consultation with adolescents as follows:

- *Agenda du sommeil antifatigue*: this French-language app allows patients to maintain a digital "sleep diary" to better understand their sleep habits and sleep efficiency, and potentially improve their sleep hygiene. The app serves as a clinical tool for physicians, enabling them to evaluate adolescents' sleep patterns, identify underlying sleep problems, and subsequently optimize the management of sleep disorders.
- mhGAP-IG 2.0 App: this app, available in several languages including English and French, is designed for health care providers around the world and is based on WHO recommendations. It serves as the electronic version of WHO's mhGAP Intervention Guide, version 2.0, for mental, neurological, and substance use disorders in nonspecialized health settings. The app provides information on mental, neurological, and substance-related disorders. It offers detailed descriptions of the diseases, questionnaires for identification and/or assessment, and recommendations for treatment. This app is highly valuable to physicians for detecting mental disorders and self-harming behaviors in adolescents. However, while the app met the minimum threshold for inclusion based on the MARS-F criteria, certain nonfunctional features may negatively affect its overall usability in clinical practice.
- WHO Contraception Tool: developed by the WHO, this app helps health care professionals select the contraceptive method that is best suited to the patient's medical history and characteristics and that complies with good-practice recommendations. Contraceptive methods are ranked from 1 to 4 according to the level of recommendation (1: "use possible," 2: "use possible with conditions," 3 and 4: "use not recommended"). This app uses the WHO 2015 recommendations and is available in several languages, including English and French. The app also allows providers to integrate patient preferences and delivers comprehensive information regarding emergency contraception, contraceptive efficacy, and antiretroviral drugs.
- Contraception Eligibility Tool: this app is the exclusive English version of the "WHO Contraception Tool" and is based on 2 WHO recommendations from 2015 to 2018. It has been developed for humanitarian professionals. It is a tool for selecting the recommended contraception according to the medical data and history of patients, with the same ranking key from 1 to 4 as the 2 previous apps. The app also provides additional clinical guidance on initiating contraception, infertility management, emergency contraception, contraceptive efficacy, the identification of victims of violence against women, and the management of missed pills.
- Contraception Point-of-Care: this app offers interactive clinical algorithms to assist health care providers in selecting the most appropriate contraceptive method for each patient. Updated in February 2025, it incorporates data on contraceptive costs in the United States, contraceptive adaptation, and natural contraceptive methods. This app uses a drop-down menu to aggregate static documents that cannot be navigated or interacted with.
- Practice Family Planning: this app was developed by the Global Health Media Project, a nonprofit organization specializing in the production of educational and preventive health videos. The app is available in French, English, and Spanish. The app offers approximately 40 videos tailored for both health care professionals and patients. Health care professionals can use them for self-training or to select and share with patients for information and education on sexual health and contraception. The videos were developed for developing countries or countries with special health conditions, but they can also be used in countries with a higher socioeconomic level. They teach professionals how to approach and present information on contraception. They also include educational videos for professionals on technical procedures such as intrauterine device insertion, implant insertion or removal, practice organization, and the prevention of nosocomial infections. The videos can be downloaded and viewed offline. This video directory is a useful tool to help professionals in their preventive actions, but some videos are of average quality and perform more reliably when downloaded locally.
- Learn Family Planning: this app, also developed by the Global Health Media Project, is very similar to the previous one. It offers videos to be shown directly to patients by professionals wishing to discuss contraceptive health. Developed for the same target audience as "Practice Family Planning," the videos can also be used in countries with higher socioeconomic levels. The topics covered are more targeted than those in "Practice Family Planning," as the videos are intended to be shared with patients: contraceptive methods, education on contraceptive use, method comparisons, pregnancy, and pelvic examinations.

Discussion

Main Results

This digital review of apps available on the Android and iOS platforms identified 976 apps addressing areas of prevention relevant to consultations between health care professionals and adolescents: vaccinations, orthopedics, drug use and addiction, mental health, nutrition, puberty, sexual health, social relations, and violence. Of these 976 apps, 7 (0.72%) were retained after analysis as recommendable to health care professionals for use in consultations to help them implement prevention measures with adolescents: 2 on mental health (including 1 sleep) and 5 on sexual health (including 3 contraception only). No apps specifically dedicated to adolescent prevention in any of the recommended areas were identified.

Method Discussion

In the MARS-F, NA ratings are described in great detail for each item within the subscales [23]. Since NA ratings are not included in the final scores, these ratings could artificially inflate the scores and lead to the inclusion of apps of insufficient quality. In our study, NA ratings were found only in subsection D (Information), items 14, 17, and 19, which could increase this risk. However, the researchers were rigorous in their

XSL•FO
RenderX

analyses, and this did not affect the selection of apps. We have analyzed these NA ratings found in the following:

- Item 14: Goals (Does the app have specific, measurable goals?). Only 2 apps (*Agenda du sommeil antifatigue* and *Bariatrique Québec*) received a numeric score on this item. The other 9 apps provide information but do not allow for any measurable goals.
- Item 17: Visual Information (NA=There is no visual information in the app; eg, it contains only audio or text). The NA rating is clear in this case.
- Item 19: Evidence Base (Scientific evidence: "Has the app been tested or evaluated, and is it supported by evidence in the published scientific literature?"). The NA is clear in this case. None of the apps has been evaluated in a published study. This item alone does not account for the scientific relevance of the app; item 15 (Quality of Information) supports the scientific quality of the information, as does the pre-evaluation screening of the apps (only apps that have been scientifically validated were included: inclusion criterion number 4).

Safety, Liability, and Ethical Considerations

The integration of Clinical Decision Support Systems, such as the WHO Contraception Tool or mhGAP-IG 2.0 App, into practice carries significant medicolegal and ethical implications. These tools are designed to support, not replace, professional expertise. The health care professional remains solely liable for the final clinical decision. Most apps include disclaimers stating that recommendations are based on global guidelines and may not account for individual patient complexities. A critical risk is "automation bias," in which a professional might overlook rare comorbidities or contraindications that are not programmed into the app's algorithms.

A scoping review published in 2023 examined studies of apps used by health care professionals in clinical practice [37]. This scoping review aimed to describe how physicians use smartphones and mobile apps in clinical settings. Ten studies were included, mainly using survey or focus group methods. Smartphones and mobile apps were used for communication, medical education and training, clinical decision-making, and accessing drug compendia in most studies. Health care professionals were concerned about patient privacy and confidentiality. In our study, only 1 app (*Agenda du sommeil antifatigue*) involved personal data, but it was nonnominative and confidentiality was guaranteed. In addition, this app stands out as a robust model, being managed by the Bordeaux University Hospital, validated by a Data Protection Officer, and hosted within a secure institutional framework ensuring General Data Protection Regulation compliance. While other apps lack transparency regarding data encryption and remote storage, they do not collect or store any personally identifiable information or nominative data within the app.

The absence of age-verification and crisis-response features represents a limitation in adolescent-focused care apps. However, as the apps in our study are professional-facing tools, they lack autonomous safety triggers by design. Consequently, the safety of vulnerable users relies on practitioners' expertise, with the clinical framework compensating for the absence of built-in safeguarding features.

Comparison With the Literature

mHealth Apps

No study similar to ours, focusing on the use of apps to facilitate preventive action by health care professionals among adolescents, has been found in the literature. One review of apps used a method similar to ours but focused on patient-facing medication management apps [35]. The aim of this 2024 study was to identify and evaluate the medication management smartphone apps available in France that send reminders to patients. Sorting apps using the MARS-F grid enables a meticulous and stringent selection of recommended apps. Our study selected 7 apps out of 976, while that of Toïgo et al [35] selected 49 out of 960. The authors stated that the quality of the apps was heterogeneous, with only 2 having been studied in a randomized controlled trial with positive results. In our study, none of the apps had been the subject of a randomized controlled trial. Choosing which apps are relevant for health care professionals is difficult, as the scientific data used to develop these tools are often lacking, as are analyses of real-life use. By using a rigorous and reproducible selection process, we have identified and recommended the most effective apps for daily clinical practice, tailored for both patients and health care professionals.

The existing literature primarily comprises scientific studies evaluating apps designed for either health care professionals or patients, typically using systematic, narrative, or scoping review methodologies. A 2012 systematic literature review examined articles addressing the design, development, evaluation, or use of smartphone apps for health care professionals, medical or nursing students, or patients [38]. The authors retained data on 57 apps for health care professionals focusing on, among other things, disease diagnosis, drug reference, medical calculators, clinical communication, medical training, and general health care apps. The disease diagnosis, drug reference, and medical calculator apps were reported as the most useful by health care professionals.

Apps specifically designed to assist health care professionals with adolescent preventive care during clinical consultations were rare in our study (7/976). Our search for apps on the subject of adolescent prevention for health care professionals enabled us to identify apps that were relevant to health care professionals. Of the 91 apps that met the inclusion criteria, only 7 were retained as relevant to health care professionals. The other 82 were intended for adolescents or their parents. Adolescents are connected and use their smartphones for an average of 4 to 8 hours a day [39-42]. Apps addressing prevention themes that concern them are more likely to be developed for them than for health care professionals. Our findings confirm a significant gap: while thousands of apps exist for patient self-management, there is a striking lack of tools designed to support providers during consultations (Decision Support Systems). This reflects a general lack of investment in "Professional mHealth," which is essential for preventive health care.

XSL•FO
RenderX

The small number of apps retained in our study was directly related to the specific inclusion and exclusion criteria established for our review. Notably, we chose to exclude paid apps to avoid imposing financial burdens on health care professionals and to preserve the absolute independence of our research, which is entirely free from conflicts of interest (ie, no recommendations of paid apps or of apps developed by the pharmaceutical industry). We would like all French and English-speaking health care professionals to be able to benefit, free of charge, from the results of our work. The exclusion criterion, which was also very selective, was the lack of explicit scientific backing for the development of an app. It was not appropriate to offer a health care professional a selection of apps that were not based on reliable scientific data. The MARS-F tool is widely used in the literature as a tool for assessing the quality of health apps: it is a simple, objective, and reliable tool for classifying and assessing the quality of mobile health apps [34,35]. The choice of this method for our work was, therefore, appropriate [34,43].

Many Areas of Prevention for Adolescents

There are many areas of prevention that need to be discussed with adolescents during a consultation with a health care professional. The development of apps designed to support health care professionals in their preventive work with adolescents does not always address current global public health challenges. Furthermore, the areas of prevention addressed by these digital tools are unevenly represented. A quantitative study carried out in France in 2021 confirmed that general practitioners mainly discussed orthopedics, sports activities, vaccinations, and contraception, but had more difficulty discussing the prevention of sexually transmitted infections or addiction when consulting an adolescent [7]. Two hundred eleven (21.6%) of the 976 apps were related to addictions and substance abuse, which could meet adolescents' needs. No apps were selected because they had to be paid for or because they lacked scientific validation. The development of these apps may have been guided by a marketing and financial objective (paid apps) rather than by a public health interest (no scientific validation). General practitioners in the French 2021 study readily and spontaneously discussed orthopedics, vaccination, and contraception [7]. Of the 976 apps found in our screening, only 133 (13.6%) and 38 (3.9%) concerned orthopedics and vaccination. No apps were selected for health care professionals. As general practitioners are already involved in preventive work on these topics, the development of relevant and scientifically valid apps on these subjects was probably not a priority. No apps related to vaccination were selected. However, vaccination coverage among adolescents remains insufficient today. Despite the need for digital interventions on vaccination, high development costs and the necessity for frequent updates could represent significant barriers for developers. General practitioners need tools to discuss sexual health, but seem to find it easier to discuss contraception with adolescents. Only 41 apps about sexual health (4.2%) were identified out of the 976. At the final inclusion, sexual health apps accounted for 5 (71.7%) of the 7 apps selected as relevant for health care professionals. These 5 apps all deal with contraception, but other areas of sexual health are not covered, or only to a limited extent. This may be due to historical market trends, whereby sexual health was an early adopter of digital formats, as well as the ongoing need to improve adolescent reproductive health worldwide. Despite constant progress in contraception, the rate of unwanted pregnancies worldwide has changed little. Already in 2015, among 4793 sexually active American students, 488 (10.2%) said they only used the withdrawal method, and 594 (12.4%) used no contraceptive method at all [44]. Two Canadian studies in 2017 and 2019 found a prevalence of adolescent pregnancy ranging from 3.9% to 14.1%, depending on the group studied [45,46]. A 2024 French study entitled "The context of sexuality in France," conducted from 2019 to 2023, reported some worrying results [47]. Among 18 - to 29-year-olds, use of the pill fell from 54.3% to 36.6% between 2016 and 2023, use of the intrauterine device increased from 10.9% to 19.3%, but the absence of contraception rose from 4.3% to 8.7% [47]. The mental health of adolescents is a constant and indeed growing concern in France and around the world. This situation has worsened since the COVID-19 pandemic [48]. However, no mental health apps were included in our study. Although 370 (37.9%) of the 976 apps were related to mental health, no apps were selected because they had to be paid for or because they lacked scientific validation. Developers have adapted to the market and to the needs of health care professionals in relation to public health issues. However, the cost of the apps and ethical concerns about offering an app that has not been scientifically validated mean that no tool was selected.

Implications for Policy and Practice

Analysis of the MARS-F subdomains reveals that the apps had higher scores in Information and Functionality (mean scores of 3.5 for both), while Engagement and Esthetics received lower ratings (mean scores of 2.0 and 2.5, respectively). This suggests that current apps are often developed as "digital brochures" rather than interactive tools. This situation is frequently observed in mHealth literature [49], where institution- or professional-facing apps prioritize clinical evidence and technical reliability over visual appeal or gamification. For a health care provider, a tool that is difficult to navigate during a consultation is unlikely to be adopted, regardless of the quality of its medical evidence. This implies the need for development frameworks. These frameworks have to integrate optimal user-centered design and esthetics to encourage users to engage with an app, so that tools can be more effective in clinical settings. Our findings align with the recent update by the High Authority for Health on the evaluation framework for mHealth apps [21]. The High Authority for Health confirmed that medical expertise and scientific evidence are mandatory for digital tools to be integrated into official health care pathways. However, our study shows that very few prevention-focused apps currently meet these high standards, particularly for the adolescent population. Medical societies and health authorities should establish certification labels or "official app libraries" to guide health care professionals. This would reduce the "search burden" for clinicians and ensure the quality of tools used with adolescents.

XSL•FO
RenderX

Strengths and Limitations

Method

This is a systematic digital review, as all the apps were scanned to exhaustion during the initial 2 screenings. A solid methodology involving 7 researchers and the drafting of the study based on the PRISMA, PRISMA-S, and PASSR checklists reinforce the methodological rigor [24-26]. Finally, the use of a blind analysis based on a validated tool (MARS-F), combined with a detailed analysis of reliability, ensures the validity of the results by limiting selection bias.

Fifty-two apps disappeared during the course of the study. The ephemeral nature of the apps partly explains the small number of results and must be taken into account when sharing the results of our study, as with similar research projects. The disappearance of 52 apps during our study period highlights the high volatility of the digital health market, a major challenge for health care professionals seeking stable clinical tools and a structural limitation for any systematic digital review. Even though the selected apps successfully advanced through all screening stages, several of them still exhibit technical deficiencies that may negatively impact their clinical utility and usability. Although the strict exclusion criteria may have omitted certain apps with potential use for health care professionals, the fact that the selected apps are free, commercially independent, and scientifically validated significantly strengthens the clinical relevance and validity of our findings.

Reliability

A key strength of this study is the transparency regarding measurement error. By reporting the SEM and MAE alongside the ICC, we provide a realistic context for the quality scores. The fact that the SEM (0.60) is often larger than the score gaps between top-rated apps indicates that these tools are of comparable quality within the limits of the MARS-F scale's precision. Since the ICC is a ratio of intersubject variance to total variance, the lack of variability between subjects in our small sample mathematically deflates the ICC, even when raters are in close agreement. The low ICC in some dimensions does not reflect poor rater performance but rather a "range restriction" effect, as the selected apps were homogeneously high in quality. This is confirmed by the good global MAE (0.53), indicating that, on average, raters differed by only half a point on a 5-point scale. This confirms that the ratings are clinically consistent and that the differences between the veteran expert and the residents represent an acceptable margin of measurement error rather than systematic disagreement.

The "Engagement" subscale had the highest statistical reliability (fair ICC=0.540 and excellent MAE=0.40), while the "Functionality" subscale exhibited the greatest rater divergence (poor MAE=1.03). Overall, the low absolute error across most dimensions confirms that the ratings provided by the residents are consistent with expert judgment for practical evaluation purposes.

While our rater calibration involved rigorous independent scoring and iterative consensus sessions, no formal statistical reliability metrics (such as a training ICC) were computed during this phase. This lack of a quantitative baseline check represents a limitation. It may explain why certain complex domains, specifically the Functionality subscale, demonstrated lower scoring homogeneity in the main evaluation, as reflected by an MAE of 1.03.

The global MARS-F score yielded an SEM of 0.60, providing essential context for comparing the evaluated apps. Given that the differences between the mean scores of several apps (eg, 3.5 vs 3.2) are smaller than this measurement error (Multimedia Appendix 2), this indicates that variations should be interpreted as qualitative trends rather than statistically significant differences in app quality (Table 4). The inclusion of the SD and range for each app further demonstrates the distribution of ratings and confirms that the observed consistency remains within acceptable clinical margins despite the small sample size.

Language

This study examines English and French apps evaluated by French health care providers. A selection was made to ensure the tools were applicable to clinical settings within these linguistic spheres. However, we acknowledge that perceptions of "clinical relevance" and "quality" are inherently shaped by national guidelines and cultural nuances. Consequently, while our findings offer robust insights for Western clinical contexts, they should be viewed as a foundation for further cross-cultural research across diverse linguistic and regulatory environments.

Perspective

Our study conducted a rigorous selection of existing apps covering all areas of prevention recommended for adolescents during consultations with a health care professional. This type of review should be carried out systematically before any new tool is developed to avoid duplication and focus creative energy on innovative digital tools. Another study will look at existing websites that are relevant to helping health care professionals with their prevention work with adolescents during consultations. Similar to this study, it will examine websites in all the recommended areas of prevention, and the sites will be analyzed using a validated tool, Netscoring [50]. These 2 research projects (reviews of apps and websites) could be used as models for similar digital reviews prior to the creation of any new app or website. Finally, whatever tool is identified or designed, feasibility and acceptability studies by the professionals themselves will be necessary.

While the MARS-F scale offers a robust evaluation of information quality and credibility, it lacks explicit criteria for assessing clinical safety features. Future research should integrate specific safety frameworks to determine whether apps incorporate essential emergency protocols or risk-warning systems, particularly for users presenting with severe mental health symptoms.

Pending assessment of the acceptability and feasibility in clinical practice of the 7 apps selected in our study, we recommend 4 of these apps: "*Agenda du sommeil antifatigue*" and "mhGAP-IG 2.0 App" (with certain limitations linked to the technical shortcomings of "mgGAP-IG 2.0 App"); of the 5 sexual health apps, we recommend "WHO Contraception Tool," available in several languages, and "Practice Family Planning," which is richer in video content for health care professionals.

XSL•FO
RenderX

These 5 apps could be included in the recommended tools for health care professionals on the planned global website.

To provide a comprehensive operational framework for the 4 recommended apps, Table 5 categorizes them based on the following 6 key implementation parameters:

- Clinical scenarios define the specific medical contexts (eg, diagnosis, prevention, or follow-up), where the app is most effective.
- Time-efficiency per consultation estimates the duration required to use the tool during a standard medical visit, ensuring that it remains compatible with professional time constraints.
- Technical information covers the hardware and software requirements (eg, operating system compatibility and offline access).
- Integration describes the clinical workflow: how the tool is incorporated into the patient-provider interaction or synchronized with health data management.
- Engagement refers to the strategies used to motivate health care professionals' participation (eg, visual aids).
- Barriers to adoption include linguistic limitations, technical bugs, and cognitive load for practitioners.
- Mitigation strategies represent the proactive measures or professional safeguards proposed to minimize identified risks and overcome barriers, ensuring the safe and effective use of the digital tool within a supervised clinical framework.

XSL•FO
RenderX

Table . Guide for practical clinical implementation of the 4 recommended apps.

App name	Clinical scenarios and time efficiency per consultation	TI[a] and integration	Engagement and barriers to implementation	Mitigation strategies
Agenda du sommeil antifatigue	Scenarios: • Diagnosis of insomnia • Sleep hygiene education Time: 5 min	TI: • iOS (Apple) or Android (Google) • French Integration: • HCP[b] reviews data entered by the patient	Engagement • Active self-monitoring by the patient Barrier: • Patient forgetfulness in logging data	• The patient uses automated app notifications • HCP validates data entry during each clinical visit
mhGAP-IG 2.0 App	Scenario: • Assessment of mental or neurological disorders in nonspecialist settings • Information on various neurological and mental disorders Time: 15 min (questionnaire and flowchart) or much more for information	TI: • iOS or Android • Multilanguage • E-version of a WHO[c] guide Integration: • Decision support tool used during the clinical interview	Engagement: • Standardized assessment tool Barrier: • Some nonfunctional features	• HCP should cross-reference with the physical or PDF WHO manual for nonfunctional app links
WHO Contraception Tool	Scenario: • Selection of evidence-based contraception • Checking for contraindications to current contraception Time: 3 min during the prescription phase	TI: • iOS or Android • Multilanguage • Regularly updated Integration: • Point-of-care tool to match medical history with contraception method	Engagement: • Inclusion of patient preferences Barrier: • Requires medical history data entry	• Prepare a patient history checklist prior to app use to speed up data entry (patient or HCP)
Practice Family Planning	Scenarios: • Professional self-training for women's health consultations • Clinical procedure preparation (eg, IUD[d] insertion) • Direct patient education on contraceptive health Time: 5 - 15 min/video	TI: • iOS or Android • Multilanguage • Offline mode available Integration: • Preprocedural refresher for women health providers • Shared screen education during consultation	Engagement: • Visual learning • Short videos Barrier: • Average video quality • Requires significant storage space	• Preselect videos relevant to the practice of HCP • Download videos prior to clinical sessions to ensure smooth playback offline

[a]TI: technical information.

[b]HCP: health care professional.

[c]WHO: World Health Organization.

[d]IUD: intrauterine device.

Conclusions

There are many areas of prevention that need to be addressed with adolescents during consultations. Only 7 apps were relevant for helping health care professionals in their prevention work with adolescents. Five deal with sexual health, particularly contraception, and 2 with mental health. No apps have been validated for orthopedics, social relations, addictions, or vaccinations. Health care professionals are connected, especially online during consultations. A general online tool to help health care professionals with prevention among adolescents during consultations is relevant but currently nonexistent. Our study will contribute to the selection of relevant content for the development of such a tool, but further research is needed to ensure that it meets the needs of health care professionals.

XSL•FO
RenderX

Acknowledgments

The authors thank Hervé Maisonneuve for his invaluable advice in writing this article. No generative AI tools were used at any stage in the preparation of this manuscript.

Funding

The authors declared no financial support was received for this work.

Data Availability

The datasets analyzed during this study are available from the corresponding author upon reasonable request and in Checklist 1 and Multimedia Appendices 1-3.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study protocol.
[DOCX File, 68 KB - mhealth_v14i1e83908_app1.docx]

Multimedia Appendix 2
MARS-F scores for the 9 evaluated apps.
[XLSX File, 86 KB - mhealth_v14i1e83908_app2.xlsx]

Multimedia Appendix 3
Abstracts of the 7 included apps.
[DOCX File, 27 KB - mhealth_v14i1e83908_app3.docx]

Checklist 1
PRISMA checklist.
[DOCX File, 30 KB - mhealth_v14i1e83908_app4.docx]

Checklist 2
PRISMA-S checklist.
[DOCX File, 17 KB - mhealth_v14i1e83908_app5.docx]

Checklist 3
PASSR checklist
[DOCX File, 31 KB - mhealth_v14i1e83908_app6.docx]

References

1. Sawyer SM, Azzopardi PS, Wickremarathne D, Patton GC. The age of adolescence. Lancet Child Adolesc Health 2018 Mar;2(3):223-228. [doi: 10.1016/S2352-4642(18)30022-1] [Medline: 30169257]
2. da Conceição Taborda-Simões M. L'adolescence: une transition, une crise ou un changement? [Article in French]. Bull Psychol 2005 May 1;479(5):521-534. [doi: 10.3917/bupsy.479.0521]
3. Basto-Pereira M, Miranda A, Ribeiro S, Maia Â. Growing up with adversity: from juvenile justice involvement to criminal persistence and psychosocial problems in young adulthood. Child Abuse Negl 2016 Dec;62:63-75. [doi: 10.1016/j.chiabu.2016.10.011] [Medline: 27794243]
4. Peeters M, Oldehinkel A, Veenstra R, Vollebergh W. Unique developmental trajectories of risk behaviors in adolescence and associated outcomes in young adulthood. PLoS One 2019;14(11):e0225088. [doi: 10.1371/journal.pone.0225088] [Medline: 31721787]
5. Duell N, Steinberg L, Icenogle G, et al. Age patterns in risk taking across the world. J Youth Adolesc 2018 May;47(5):1052-1072. [doi: 10.1007/s10964-017-0752-y] [Medline: 29047004]
6. Beck F, Richard JB, editors. Les Comportements de Santé Des Jeunes Analyses Du Baromètre Santé 2010 [Book in French]: Institut national de prévention et d'éducation pour la santé (INPES); 2013.
7. Vergne C, Bayard S, Rieu-Clotet L, Carbonnel F, Lognos B, Million E. Which prevention fields are most frequently forgotten during consultations with adolescents and young adults? An ECOGEN ancillary cross-sectional study. Exercer 2021 Mar 1;32(171):106-112. [doi: 10.56746/EXERCER.2021.171.106]

XSL•FO
RenderX

8. The Guide to Clinical Preventive Services 2014: Recommendations of the US Preventive Services Task Force: Agency for Healthcare Research and Quality (US); 2014.
9. Alderman EM, Breuner CC, Committee on Adolescence. Unique needs of the adolescent. Pediatrics 2019 Dec;144(6):e20193150. [doi: 10.1542/peds.2019-3150] [Medline: 31740496]
10. Yarnall KSH, Pollak KI, Østbye T, Krause KM, Michener JL. Primary care: is there enough time for prevention? Am J Public Health 2003 Apr;93(4):635-641. [doi: 10.2105/ajph.93.4.635] [Medline: 12660210]
11. Biglan A, Brennan PA, Foster SL, Holder HD. Helping Adolescents at Risk: Prevention of Multiple Problem Behaviors: Guilford Press; 2004.
12. Binder P, Heintz AL, Tudrej B, Haller DM, Vanderkam P. L'approche des adolescents en médecine générale Première partie. L'adolescent, cet inconnu [Article in French]. Exercer 2018;141:122-132. [doi: 10.56746/EXERCER.2018.141.122]
13. Tudrej BV, Heintz AL, Ingrand P, Gicquel L, Binder P. What do troubled adolescents expect from their GPs? Eur J Gen Pract 2016 Dec;22(4):247-254. [doi: 10.1080/13814788.2016.1216541] [Medline: 27594206]
14. Stheneur C, Alvin P, Boudaillez B, et al. La première consultation avec un adolescent [Article in French]. Arch Pediatr 2009 Sep;16(9):1309-1312. [doi: 10.1016/j.arcped.2009.06.003]
15. Million E, Herbreteau M, Bourrel G, et al. Adolescents report insufficient preventive health care in GP consultations: a qualitative study. BJGP Open 2025;9. [doi: 10.3399/BJGPO.2024.0290]
16. du Roscoät E, Beck F. Efficient interventions on suicide prevention: a literature review. Rev Epidemiol Sante Publique 2013 Aug;61(4):363-374. [doi: 10.1016/j.respe.2013.01.099] [Medline: 23849295]
17. Gelly J, Mentre F, Nougairede M, Duval X. Preventive services recommendations for adults in primary care settings: agreement between Canada, France and the USA: a systematic review. Prev Med 2013 Jul;57(1):3-11. [doi: 10.1016/j.ypmed.2013.03.012] [Medline: 23571186]
18. Just between us: how to initiate and implement a health education approach with adolescents? : Institut national de prévention et d'éducation pour la santé; 2009 URL: https://www.medecin-ado.org/addeo_content/documents_annexes/121-4-entrenousinpes.pdf.pdf [accessed 2026-07-28]
19. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. : World Health Organization; 2011 URL: https://iris.who.int/server/api/core/bitstreams/ad1b13c0-7c82-47b4-8dd5-f0a26c3a3cc3/content [accessed 2026-07-28]
20. Safon MO, Suhard V. E-health: telehealth, digital health, or connected health. : Institut de recherche et documentation en économie de la santé (IRDES); 2025 URL: https://www.irdes.fr/documentation/syntheses/e-sante.pdf [accessed 2026-07-28]
21. Évaluation des applications dans le champ de la santé mobile (mhealth) - état des lieux et critères de qualité du contenu médical pour le référencement des services numériques dans l'espace numérique de santé et le bouquet de services des professionnels [Report in French]. : Haute Autorité de Santé (HAS); 2021 URL: https://www.has-sante.fr/upload/docs/application/pdf/2021-06/criteres_de_qualite_du_contenu_medical_referencement_mhealth_ens_2021-06-30_10-58-28_773.pdf [accessed 2026-04-22]
22. Della Vecchia C, Leroy T, Bauquier C, et al. Willingness of French general practitioners to prescribe mHealth apps and devices: quantitative study. JMIR Mhealth Uhealth 2022 Feb 11;10(2):e28372. [doi: 10.2196/28372] [Medline: 35147508]
23. Saliasi I, Martinon P, Darlington E, et al. Promoting health via mHealth applications using a French version of the mobile app rating scale: adaptation and validation study. JMIR Mhealth Uhealth 2021 Aug 31;9(8):e30480. [doi: 10.2196/30480] [Medline: 34463623]
24. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. PLoS Med 2021 Mar;18(3):e1003583. [doi: 10.1371/journal.pmed.1003583] [Medline: 33780438]
25. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev 2021 Jan 26;10(1):39. [doi: 10.1186/s13643-020-01542-z] [Medline: 33499930]
26. Marshall JM, Dunstan DA, Bartik W. Apps with maps—anxiety and depression mobile apps with evidence-based frameworks: systematic search of major app stores. JMIR Ment Health 2020 Jun 24;7(6):e16525. [doi: 10.2196/16525] [Medline: 32579127]
27. Operating system market share worldwide. StatCounter Global Stats. URL: https://gs.statcounter.com/os-market-share/ [accessed 2025-04-22]
28. Medical subject headings. National Library of Medicine (NLM). 2026. URL: https://www.nlm.nih.gov/mesh/meshhome.html [accessed 2026-07-18]
29. Google Help. URL: https://support.google.com/googleplay/android-developer/answer/16070163?hl=en [accessed 2025-04-22]
30. Clarifying criteria & new timing extension for app store improvements process - latest news. Apple Developer. URL: https://developer.apple.com/news/?id=gi6npkmf [accessed 2025-04-22]
31. Crawford A, Serhal E. Digital health equity and COVID-19: the innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020 Jun 2;22(6):e19361. [doi: 10.2196/19361] [Medline: 32452816]
32. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 4;3(4):e101. [doi: 10.2196/mhealth.4924] [Medline: 26537656]

XSL•FO
RenderX

33. Folkvord F, Bol N, Stazi G, Peschke L, Lupiáñez-Villanueva F. Preferences in the willingness to download an mHealth app: discrete choice experimental study in Spain, Germany, and the Netherlands. JMIR Form Res 2023 Dec 25;7:e48335. [doi: 10.2196/48335] [Medline: 38145470]
34. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
35. Toïgo M, Marc J, Hayot M, Moulis L, Carbonnel F. Quality assessment of smartphone medication management apps in France: systematic search. JMIR Mhealth Uhealth 2024 Mar 18;12:e54866. [doi: 10.2196/54866] [Medline: 38498042]
36. WMA Declaration of Helsinki – ethical principles for medical research involving human participants. World Medical Association. 2013. URL: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ [accessed 2026-07-28]
37. Lee M, Mahmood ABSB, Lee ES, Smith HE, Car LT. Smartphone and mobile app use among physicians in clinical practice: scoping review. JMIR Mhealth Uhealth 2023 Mar 31;11:e44765. [doi: 10.2196/44765] [Medline: 37000498]
38. Mosa ASM, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012 Jul 10;12:67. [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
39. Kosola S, Mörö S, Holopainen E. Smartphone use and well-being of adolescent girls: a population-based study. Arch Dis Child 2024 Jun 19;109(7):576-581. [doi: 10.1136/archdischild-2023-326521] [Medline: 38772732]
40. Bae SM. Smartphone addiction of adolescents, not a smart choice. J Korean Med Sci 2017 Oct;32(10):1563-1564. [doi: 10.3346/jkms.2017.32.10.1563] [Medline: 28875596]
41. Kliesener T, Meigen C, Kiess W, Poulain T. Associations between problematic smartphone use and behavioural difficulties, quality of life, and school performance among children and adolescents. BMC Psychiatry 2022 Mar 18;22(1):195. [doi: 10.1186/s12888-022-03815-4] [Medline: 35300635]
42. Nagata JM, Al-Shoaibi AAA, Leong AW, et al. Screen time and mental health: a prospective analysis of the Adolescent Brain Cognitive Development (ABCD) Study. BMC Public Health 2024 Oct 7;24(1):2686. [doi: 10.1186/s12889-024-20102-x] [Medline: 39370520]
43. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the quality of mobile apps for the management of pain: systematic search and evaluation using the mobile app rating scale. JMIR Mhealth Uhealth 2018 Oct 25;6(10):e10718. [doi: 10.2196/10718] [Medline: 30361196]
44. Liddon N, O'Malley Olsen E, Carter M, Hatfield-Timajchy K. Withdrawal as pregnancy prevention and associated risk factors among US high school students: findings from the 2011 National Youth Risk Behavior Survey. Contraception 2016 Feb;93(2):126-132. [doi: 10.1016/j.contraception.2015.08.015] [Medline: 26363434]
45. Fortin-Langelier E, Daigneault I, Achim J, Vézina-Gagnon P, Guérin V, Frappier JY. A matched cohort study of the association between childhood sexual abuse and teenage pregnancy. J Adolesc Health 2019 Sep;65(3):384-389. [doi: 10.1016/j.jadohealth.2019.03.001] [Medline: 31196780]
46. Wall-Wieler E, Roos LL, Nickel NC. Adolescent pregnancy outcomes among sisters and mothers: a population-based retrospective cohort study using linkable administrative data. Public Health Rep 2018;133(1):100-108. [doi: 10.1177/0033354917739583] [Medline: 29262270]
47. Andro A, Bajos N, Bellamine R, Bergström M, Beaubatie E. Preliminary results of the national survey "Context of Sexualities in France 2023". ANRS Maladies infectieuses émergentes (ANRS MIE). 2024. URL: https://anrs.fr/fr/actualites/actualites/premiers-resultats-enquete-nationale-contexte-sexualites-france-2023/ [accessed 2026-07-23]
48. Jones EAK, Mitra AK, Bhuiyan AR. Impact of COVID-19 on mental health in adolescents: a systematic review. Int J Environ Res Public Health 2021 Mar 3;18(5):2470. [doi: 10.3390/ijerph18052470] [Medline: 33802278]
49. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567. [doi: 10.2196/14567] [Medline: 31573916]
1. D Darmoni SJ, Leroux V, Thirion B, Santamaria P, Gea M. Net Scoring ®: critères de qualité de l'information de santé sur l'Internet [Article in French]. 1999. URL: https://www.cismef.org/netscoring/ [accessed 2026-08-01]

Abbreviations

ICC: intraclass correlation coefficient
MAE: mean absolute error
MARS-F: French version of the Mobile App Rating Scale
mHealth: mobile health
NA: not attributable
PASSR: Protocol for App Store Systematic Reviews
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Literature Search Extension
SEM: standard error of measurement
WHO: World Health Organization

XSL•FO
RenderX

Edited by M DeJonckheere; submitted 10.Sep.2025; peer-reviewed by C Smith, H Balakrishna, J Odeniya, P Sengupta; revised version received 27.Jun.2026; accepted 17.Jul.2026; published 11.Sep.2026.

Please cite as:
Million E, Kuijper M, Bautista C, Pirat L, Falissard B, Lognos B, Carbonnel F
Smartphone Apps for Preventing Adolescent Health Problems Among Health Care Professionals: Systematic Search and Quality Assessment
JMIR Mhealth Uhealth 2026;14:e83908
URL: https://mhealth.jmir.org/2026/1/e83908
doi:10.2196/83908

© Elodie Million, Marijn Kuijper, Carole Bautista, Lola Pirat, Bruno Falissard, Béatrice Lognos, François Carbonnel. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobile Apps Addressing Feeding of Children With Autism: Systematic Search and Content and Quality Analysis

Fabrice Mukarage[1*], MSIT; Christopher Cofie Kuzagbe[1*], MSIT; Skye Nandi Adams[2*], PhD; N'guessan Yves-Roland Douha[1*], DEng; Edith Talina Luhanga[1*], DEng

[1]Carnegie Mellon University Africa, Special Economic Zone, Regional ICT Center of Excellence, Kigali, Rwanda

[2]University of the Witwatersrand, Johannesburg, Gauteng, South Africa

[*]all authors contributed equally

Corresponding Author:
Edith Talina Luhanga, DEng
Carnegie Mellon University Africa, Special Economic Zone, Regional ICT Center of Excellence, Kigali, Rwanda

Abstract

Background: Approximately 1 in 100 children worldwide are diagnosed with autism spectrum disorder (ASD), and 46% to 89% of those diagnosed experience significant feeding difficulties. Mobile health (mHealth) apps are a potential tool for scalable support. However, there is nascent literature on the quality and relevance of mHealth apps for managing ASD-related feeding challenges.

Objective: This study aimed at systematically identifying and reviewing the content, behavior change strategies, and quality of free English-language mHealth apps supporting the feeding of children with autism.

Methods: A systematic search was conducted on the Apple App Store and Google Play Store between September and October 2024 and updated between June and July 2026. Apps were included if they were free or freemium, in English, updated within the past year, and explicitly focused on the feeding of children with autism. Apps were excluded if they targeted clinicians rather than parents or children; could not be installed on the test devices; or primarily focused on learning or communication rather than feeding, despite descriptions suggesting relevance to feeding. We identified intervention strategies in the included apps using the behavior change wheel (BCW) framework and evaluated their quality using the Mobile App Rating Scale (MARS).

Results: The search returned 333 apps from both app stores. After removing duplicates and screening, only 2 iOS apps (EduKitchen-Toddlers Food Games and Autism Food Coach 2) met all inclusion criteria. 331 apps were excluded due to their focus on communication, early education, or other autism support that was not food related. EduKitchen-Toddlers Food Games provided child-centered interactive games and sensory-friendly visuals, while Autism Food Coach 2 provided structured caregiver tools, visual meal plans, and progress tracking. EduKitchen-Toddlers Food Games provided the BCW intervention functions of education, training, enablement, incentivization, and modeling, while Autism Food Coach 2 provided education, training, enablement, persuasion, incentivization, and environmental restructuring. The apps received MARS scores of 3.9 and 3.7, respectively, indicating acceptable quality. The apps' shortcomings included limited customization for diverse user needs and the absence of documented clinical validation.

Conclusions: There is a critical shortage of mHealth apps for feeding difficulties in children with autism that are both evidence-based and of high quality. Future development must integrate robust clinical validation and comprehensive, caregiver-centered support features to address this significant gap.

(JMIR Mhealth Uhealth 2026;14:e94908) doi:10.2196/94908

KEYWORDS

autism spectrum disorder; ASD; mobile health; mHealth; caregiver support; behavior change wheel; Mobile App Rating Scale; MARS

Introduction

Autism spectrum disorder (ASD) is a complex neurodevelopmental condition characterized by challenges in social interaction, communication, behavior regulation, and sensory processing. The global prevalence of ASD has increased significantly in recent decades; the US Centers for Disease Control and Prevention reports a 178% increase in autism prevalence since 2000 [1], and the World Health Organization estimates that approximately 1 in 100 children is diagnosed with ASD [2].

Among the various comorbidities observed in children with autism, feeding difficulties are especially common and underaddressed. Studies report prevalence rates ranging from

XSL•FO
RenderX

46% to 89% [3], with affected children exhibiting food selectivity, aversions to textures and flavors, rapid eating, and disruptive mealtime behaviors [4,5]. These challenges are linked to nutritional deficiencies, impaired growth, gastrointestinal disorders, and metabolic imbalances [6,7]. Children with autism experience feeding problems 5 times more frequently than their neurotypical peers [8]. This situation places a significant burden on caregivers and parents, who often serve as the primary facilitators of feeding interventions. They manage complex feeding routines while navigating children's behavioral and communication challenges. Given their central role in daily feeding routines and their intimate knowledge of their child's behaviors, parents represent a critical intervention point for addressing these challenges.

Parent-mediated interventions are a prominent model for supporting children with autism. Previous research indicates that parents can successfully learn and implement intervention strategies [9]. However, translating these interventions into community settings remains challenging due to persistent barriers in accessibility and scalability [10]. The emergence of mobile health (mHealth) apps presents a promising means to overcome these barriers. A growing body of research shows mHealth apps can effectively train parents and improve child outcomes in areas such as communication, socialization, and behavior management [11,12].

The success of these digital tools, however, is contingent upon their design and usability. Research indicates that parent engagement is directly linked to an app's perceived utility and its alignment with daily caregiving realities [13]. Poorly designed interventions that fail to account for parental needs, stress levels, and caregiving contexts may inadvertently exacerbate caregiver burden [14,15]. Indeed, systematic reviews of mHealth apps across domains such as child nutrition, child growth monitoring, and infant care show considerable variations in app quality [16,17]. Most apps have acceptable to good quality when evaluated with standardized tools such as the Mobile App Rating Scale (MARS) [18], largely due to lower engagement, usability, or quality of information scores. Other reviews have identified limited incorporation of evidence-based health information and behavior change practices in child health and parenting apps [16,17]. These findings highlight the necessity of systematically evaluating digital health interventions rather than assuming their effectiveness based on availability alone.

However, despite the growing interest in digital interventions for autism and the variable reported quality of mHealth apps, research has scarcely explored the range, support features, and quality of mHealth apps specifically designed to help parents manage feeding difficulties. The aim of this study was to conduct a systematic search and evaluation of the content, behavior change strategies, and quality of free English-language apps supporting the feeding of children with autism in the 2 major app stores (Google Play Store and Apple App Store). For included apps, we identified informational topics and app features and mapped them to the behavior change wheel (BCW) [19] to determine the scope and the supported behavior change intervention functions. The BCW provides a comprehensive theoretical framework for analyzing behavior change potential through intervention functions such as education, persuasion, and incentivization. Understanding the behavior change strategies embedded in an app is critical to evaluating its likely effectiveness. We then assessed the app quality using the MARS [18], a validated and widely used tool for evaluating mHealth app quality across 4 critical factors: engagement, functionality, aesthetics, and information quality.

Overall, this study answered the following research questions:

1. What is the availability of apps addressing feeding difficulties in children with autism in the 2 major commercial app stores?
2. What behavior change intervention functions are included in available apps?
3. What is the overall quality of available apps when assessed using the MARS?
4. What are the strengths and limitations of existing apps in supporting feeding-related interventions?
5. What actionable recommendations can guide the development of future mHealth apps that more effectively assist caregivers in managing ASD-related feeding difficulties?

Methods

Eligibility Criteria

Inclusion Criteria

We analyzed apps that met the following inclusion criteria: (1) the app is free to download (either fully free or freemium, ie, offering in-app purchases), (2) the content is written in English, (3) the app has been updated within the past year (to exclude defunct or outdated apps), and (4) the title or description explicitly addresses feeding in children with autism.

We limited inclusion to English-only apps, as it is the most widely spoken language worldwide when combining native and non-native speakers [20]. As researchers based in Africa, this choice also reflected local realities, as English is commonly used as an official or working language in many countries, with 26 countries in sub-Saharan Africa alone using English as an official language, either exclusively or alongside local languages [21].

Exclusion Criteria

Apps were excluded if they (1) targeted clinicians rather than caregivers or children; (2) primarily focused on learning or communication rather than feeding, despite descriptions suggesting relevance to feeding; or (3) could not be installed on the test devices (iPhone 12 Pro Max for iOS and Google Pixel 3a XL for Android) due to geographical restrictions, unsupported operating system versions, or other installation challenges.

Information Sources

Searches were conducted on the Google Play Store and Apple App Store. The Google Play Store holds more than 70% of smartphone apps market share [22], while the Apple App Store is considered the second largest app market store [23].

XSL•FO
RenderX

Search Strategy

We conducted a systematic search of the Apple App Store (iOS) and Google Play Store (Android) to identify mHealth apps related to feeding, food, meal planning, or nutrition for children with autism. The search was performed between September and October 2024 and updated between June and July 2026. All researchers were based in Rwanda, East Africa, during the study. The search terms used included: "child autism food," "kids autism food," "toddler autism food," "child autism nutrition," "kids autism nutrition," "toddler autism nutrition," "child autism meal prep planning," "kids autism meal prep planning," "toddler autism meal prep planning," "child autism feeding," "kids autism feeding," "toddler autism feeding," "child autism mealtime," "kids autism mealtime," and "toddler autism mealtime." These terms were entered directly into the Apple App Store and Google Play Store search interfaces with no filters or limits applied. The searches were conducted using an iPhone 12 Pro Max (iOS) and a Google Pixel 3a XL (Android) to simulate a typical end-user experience on each platform. All returned results were saved in a Microsoft Excel file for deduplication and screening.

Selection Process

After removing duplicates, 2 members of the research team independently screened the apps. We undertook a 2-phase screening process. In the first phase, the apps' information pages (screenshots and textual description) were screened against the inclusion and exclusion criteria. Apps with a clearly unrelated scope (eg, communication support only) were excluded. Apps that included child feeding as 1 of several features or suggested a broader scope of features than those listed on the page were moved to phase 2. In this phase, the apps were downloaded and installed on the test devices. They were then screened to determine if they contained feeding-related content. Apps that had feeding-related content were included in the final analysis. The screening results were compared, and disagreements were resolved by discussion or by consulting with a third member of the research team.

Data Collection Process and Data Items

Data was independently extracted by the same 2 members of the research team who conducted screening, with 1 reviewer focusing on the Apple App Store and the other on the Google Play Store. Reviewers independently navigated all freely accessible pages and recorded the informational content and features in a Microsoft Excel file and then validated each other's lists. The following data items were recorded per page: informational content (using article headings), interactive features, and whether each feature was free to access or paywalled. To ensure reliability, the reviewers cross-validated each other's assessments by switching platforms after the initial evaluations.

Deductive coding was then conducted using the BCW [19], with each feature mapped to 1 of the 9 intervention functions: education, persuasion, incentivization, coercion, training, restriction, environmental restructuring, modeling, and enablement. In a preparatory session, both reviewers reviewed the BCW framework and discussed descriptions of the intervention functions to ensure a shared understanding of the categories. They also created examples of app features to ensure consistent coding. Discrepancies were discussed until consensus was reached; when consensus was not possible, a third reviewer was consulted. A deductive approach was selected to assess the extent to which app content aligned with established behavior change theory, rather than to generate novel thematic categories.

App Quality Assessment

The quality of the apps was assessed using the MARS [18], a validated tool for the evaluation of mHealth apps. MARS evaluates apps on 4 dimensions: engagement (user interest and interaction), functionality (performance and usability), aesthetics (visual design), and quality of information (accuracy, relevance, and credibility of content). Each dimension is rated on a 5-point Likert scale. Two reviewers independently rated each app. Both reviewers have a background in mobile app design and development and had worked on mHealth projects over the past year. The reviewers also completed the official MARS training to ensure consistent interpretation of the scale. Following the independent rating phase, the 2 reviewers discussed their evaluations to reach a consensus score for each MARS item. These final consensus scores were used in the analysis.

Ethical Considerations

The study did not require any ethics review.

Results

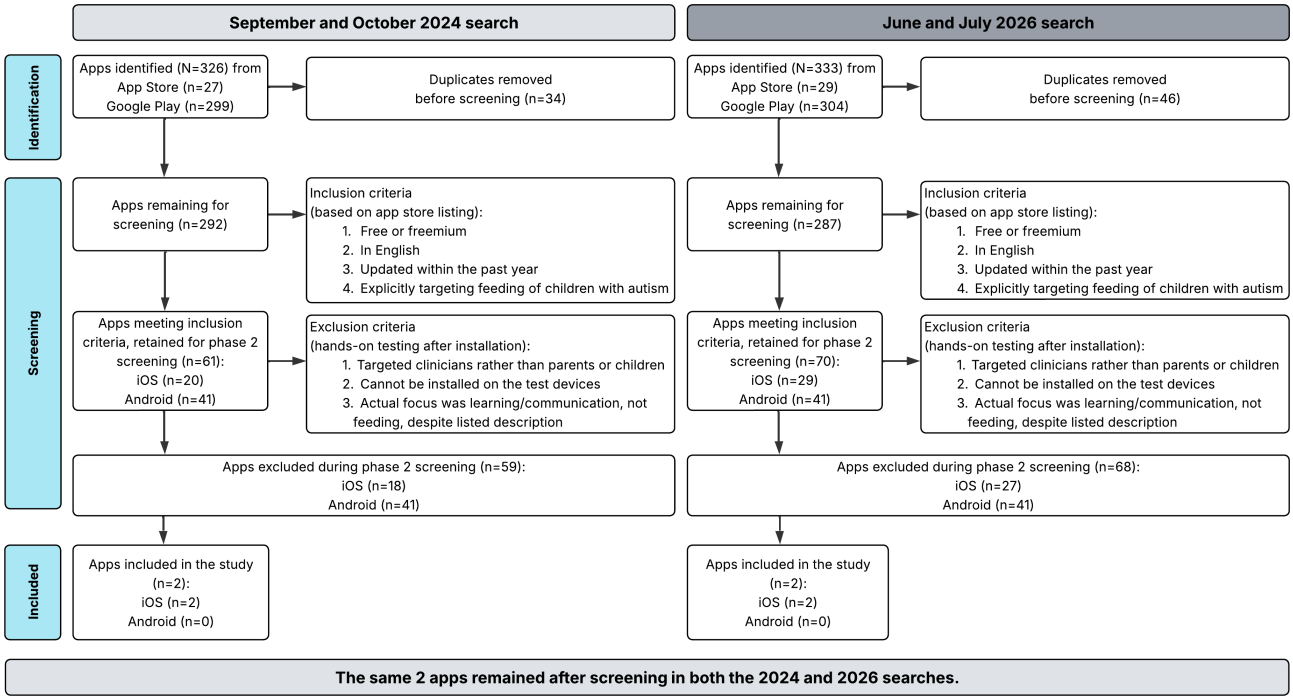
App Search and Selection

The initial search (conducted in September and October 2024) yielded 326 apps (27 iOS apps from the Apple App Store and 299 Android apps from the Google Play Store). From these 326 apps, 34 duplicates were removed, leaving 292 apps for phase 1 screening. After this stage, 61 apps were retained for phase 2 screening. The remaining apps were excluded because their focus was preschool education (eg, learning the alphabet, numbers, and animals), communication skills support, or emotional regulation through games or musical therapy. During phase 2 screening, a further 59 apps were excluded, as they had a similar focus on learning and communication. Overall, only 2 iOS apps (EduKitchen-Toddlers Food Games and Autism Food Coach 2) were found to address feeding challenges in children with autism, indicating a critical shortage on the number of available apps for supporting caregivers in the home management of feeding difficulties in children with autism.

Our updated search from June to July 2026 returned 333 apps, with 46 duplicates removed. After applying the inclusion and exclusion criteria, no additional apps were included in the study, showing no improvement in the availability of feeding support apps. We consider the implications of this in the Discussion section. A complete list of the screened apps, including their platform and primary focus, is provided in Table S1 of Multimedia Appendix 1. Figure 1 presents the app selection process following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [24].

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of mobile health (mHealth) app selection for the systematic review.



Description of the Selected Apps

Both EduKitchen-Toddlers Food Games and Autism Food Coach 2 are commercial apps, with the former developed by Cubic Frog Apps and the latter by Aidan's Apps. EduKitchen-Toddlers Food Games is a child-facing app. Its stated aim is to provide a set of kitchen games that encourage wiser food knowledge while also enhancing the cognitive skills emphasized in preschool curricula, such as sorting and matching. Meanwhile, Autism Food Coach 2 is a caregiver- and child-facing app that was developed by a parent to promote slower chewing behavior in their child. Neither of the apps reports any scientific studies conducted during app development or any evaluation studies conducted on efficacy. Similarly, there was no report on the evidence used to inform content and feature selection.

Evaluation of App Features and Content

EduKitchen-Toddlers Food Games targets early childhood learners and supports preschool cognitive skills of counting, sorting, sequencing, and related cognitive skills alongside building food familiarity through interactive play. The app (Table 1) includes 18 free mini-games focused on cooking and kitchen tasks, for example, frying and counting eggs (Figure 2A) and sorting food-related waste (Figure 2B). The foods covered in the app included whole foods such as eggs, fruits (eg, apples, bananas, and strawberries), vegetables, cereals, bread, and fruit juices, along with processed foods such as cookies, ice cream, popsicles, and sugar-sweetened soft drinks. Although there were no lessons around distinguishing healthy from unhealthy foods, we did note that healthy foods were more prominently represented.

XSL•FO
RenderX

Table . Core features and access models of the selected apps.

App name	Access model	Target audience	Core features
EduKitchen-Toddlers Food Games	One-time purchase	• Toddlers/preschool children including children with autism • Parents/caregivers who configure the app (secondary audience)	• Gamified food-related activities (eg, sorting, matching, and counting) • Reinforcement through animations and in-app rewards • Customizable settings for sensory preferences and learning levels • Engaging, child-friendly interface with cartoon visuals
Autism Food Coach 2	Subscription (£49.99/month)[a]	• Parents and caregivers of children with autism who have feeding difficulties • children with autism who participate in feeding routines (secondary audience)	• Structured mealtime routines with visual schedules • Meal tracking and progress dashboards for caregivers • Customizable feeding plans for behavioral and sensory needs • Motivational prompts, goal setting, and reward systems

[a] £1=US $1.32 as of October 5, 2026.

XSL•FO
RenderX

Figure 2. Screenshots from EduKitchen-Toddlers Food Games, developed by Cubic Frog Apps, obtained via the Apple App Store. The app features food-themed games intended to build early familiarity with food through sorting, counting, and categorization. Example games include (A) frying and counting eggs and (B) sorting food-related waste.



Children can interact with the app through touch-based activities guided by voice instructions, with multiple languages supported, while parents can customize the app to select the language, adjust the difficulty level of activities, and unlock the full set of educational games through the premium version. Since the app targets young children, there was no informational content provided.

Autism Food Coach 2 is designed for caregivers and children with autism, providing a structured approach to support feeding routines. As shown in Figure 3, the app includes motivational messages that encourage emotional reflection and sensory modulation during meals. For example, Figure 3A advises users to "notice the effects food has on your feelings," while Figure 3B suggests playing relaxing music during meals to enhance enjoyment. The app includes additional functionalities, such as progress tracking, visual schedules, and personalized meal planning tools, that allow systematic tracking of mealtime behaviors. Motivational prompts and sensory regulation suggestions (eg, relaxing music or emotional reflection) are also provided to create calmer mealtime environments.

Although both apps incorporate food-related content, they diverge considerably in functional purpose and clinical utility: one offers indirect exposure to foods and feeding support through child-led play, while the other provides direct support for caregiver-led interventions. This section evaluates the extent to which the apps support structured feeding interventions, promote adaptive mealtime behaviors, and guide caregiver decision-making.

We formalize this analytical comparison in Textbox 1, which categorizes app features based on their ability to provide actionable guidance to support caregivers during feeding interventions. The analysis confirms that only Autism Food Coach 2 includes features that align with caregiver-mediated feeding interventions.

XSL•FO
RenderX

Figure 3. Screenshots from Autism Food Coach 2, developed by Aidan's Apps, obtained via the Apple App Store. The app includes prompts that, for example, (A) encourage users to reflect on how food affects their feelings and (B) promote calm sensory environments by suggesting soft background music during meals.

XSL•FO
RenderX

Textbox 1. App features categorized by caregiver decision support capabilities.

Features with caregiver decision support

- *Instant feedback and rewards* reinforce positive child behaviors, allowing parents to adjust strategies in real time.
- *Parent support tools* offer structured feedback, coaching prompts, and meal tracking to help manage feeding routines.
- *Customizable progress monitoring* provides visual progress charts and behavior tracking to support informed parental decisions.

Features without caregiver decision support

- *Interactive child engagement* focuses on engaging the child through games and narratives but lacks direct caregiver guidance.

Intervention Function Mapping Using the BCW Framework

The BCW provides a systematic model for designing interventions based on behavioral determinants. Table 2 summarizes the BCW intervention functions identified in each app. We identified 7 of the 9 BCW functions across the 2 apps. It is worth noting that neither app incorporated coercion nor restriction.

Table . Behavior change wheel intervention function mapping for the 2 reviewed apps.

Intervention function	EduKitchen-Toddlers Food Games	Autism Food Coach 2
Education	✓	✓
Training	✓	✓
Enablement	✓	✓
Persuasion	✗	✓
Incentivization	✓	✓
Modeling	✓	✗
Environmental restructuring	✗	✓
Coercion	✗	✗
Restriction	✗	✗

EduKitchen-Toddlers Food Games mapped onto 5 BCW functions: education, training, enablement, incentivization, and modeling overall, but food-related content mapped onto only 3 interventions. These were education, delivered through interactive lessons introducing food types; incentivization, seen through gamified rewards (eg, praise animations); and modeling, in the use of animated characters demonstrating appropriate feeding behaviors. Features mapped to training and enablement primarily supported preschool learning outcomes. For instance, training is embedded in mini-games that reinforce skills such as sorting and categorization, and enablement is supported via customizable gameplay settings to accommodate diverse sensory or developmental needs.

Autism Food Coach 2 incorporated 6 intervention functions: education, training, enablement, persuasion, incentivization, and environmental restructuring. The app provides educational content through guidance on balanced nutrition and mindful eating. Training is facilitated through structured visual schedules that model consistent feeding routines. Enablement includes personalized meal planning and caregiver-facing tracking tools. Persuasion is operationalized through motivational messages and positive affirmations. Incentivization is achieved through visual progress tracking and behavioral goal setting. Environmental restructuring is implemented through tools that help caregivers create predictable low-stress mealtime environments (eg, use of music, calming messages, and visual schedules).

MARS Quality Assessment

We present the results of the MARS evaluation for the 2 selected apps. EduKitchen-Toddlers Food Games achieved an overall MARS score of 3.9, reflecting good usability and design quality. It performed strongest in functionality (4.0) due to its stable performance, intuitive navigation, and simple touch-based interactions that allowed young children to complete activities with minimal assistance. A similarly high score was achieved for aesthetics (4.0) because of its colorful, child-friendly interface, large interactive objects, high-contrast visuals, and consistent use of animated characters that maintained children's attention while reducing visual complexity. The app also incorporated clear audio instructions, simple layouts, and uncluttered screens, making interactions appropriate for preschool-aged children and potentially supportive for children with autism who may benefit from predictable visual environments and multimodal cues. Engagement (3.7) and information quality (3.9) also received strong ratings because the app used a variety of age-appropriate gamified activities, including cooking, counting, sorting, and matching games, to reinforce learning through repeated interaction rather than text-based instruction. However, the information score was slightly lower because educational content was embedded within gameplay and there were no caregiver resources, nutritional

XSL•FO
RenderX

guidance, or evidence-based informational materials. The app received a subjective quality score of 4.0, reflecting its suitability as an engaging educational app for young children.

Autism Food Coach 2 achieved an overall MARS score of 3.7, interpreted as acceptable to good usability and design quality. It received its highest ratings for information (3.95) and engagement (3.75). The information score reflected the inclusion of educational modules on mindful eating, digestion, eating pace, emotional awareness during meals, and strategies for improving mealtime experiences, together with interactive features such as personalized eating timers, bite pacing, and meal-type selection. These features encouraged active user participation and supported behavior change throughout the eating process. Functionality (3.8) was rated highly because of the app's intuitive navigation, responsive performance, and customizable meal timer that could be adapted to different food types or user preferences. Aesthetics (3.3) received the lowest score due to its relatively simple and utilitarian interface, which prioritized functionality over visual engagement. Although the app included clear graphics and a consistent layout, it offered limited animation, interactive visual elements, and child-oriented design features. The app received a subjective quality score of 4.0, reflecting its perceived usefulness as a behavioral support tool for improving eating pace and mindful eating habits.

Table 3 summarizes the scores for the apps across the 4 MARS quality dimensions, that is, engagement, functionality, aesthetics, and information, along with the overall mean and subjective quality ratings. The table highlights a significant difference in the aesthetics dimension, in which EduKitchen-Toddlers Food Games received a substantially higher score (4.0 compared to 3.3). In contrast, both apps performed comparably in engagement, functionality, and information, with differences of 0.2 points or less.

Table . Mobile App Rating Scale results for the 2 reviewed apps. Each dimension was scored on a 5-point scale.

Characteristic and dimension	EduKitchen-Toddlers Food Games	Autism Food Coach 2
Platform	iOS	iOS
Overall score (mean)	3.9	3.7
Engagement	3.7	3.75
Functionality	4.0	3.8
Aesthetics	4.0	3.3
Information	3.9	3.95
Subjective quality score	4.0	4.0

Discussion

Principal Findings

This review examined the Apple App Store and Google Play Store for mobile apps that address feeding challenges in children with autism. Our most significant finding is the scarcity of feeding support apps. Only 2 of more than 300 identified apps met our inclusion criteria, and the same 2 qualified in both search rounds (2024 and 2026). Most excluded apps focused on communication, early education, or other autism support that was not food related. Delayed speech, language impairments, and difficulty with reading and writing are common in autism. Around 30% of children with autism remain nonverbal [25], and many perform worse academically than their IQ would predict [26]. These challenges are serious [27], but feeding difficulties also affect a large share of children with autism, with prevalence estimates of 46% to 89% [3]. Poor nutrition affects physical growth and is linked to long-term cognitive impairment [28]. The emphasis on communication and learning among the autism apps we screened does not match the daily reality of raising a child with autism. Feeding is a daily need for many families, but almost no evidence-based digital support exists for it; mHealth should give feeding the same priority as communication and learning.

The 2 included apps differ in target users and intervention logic. Neither provided a theoretical basis for its features, and we found no documentation of efficacy. Both mapped to 4 core BCW functions: education, training, enablement, and incentivization. For EduKitchen-Toddlers Food Games, training and enablement served preschool learning rather than feeding. These functions match needs described in prior work. Education and training target the capability barriers of teachers and caregivers, who report limited knowledge of feeding strategies for children with autism [29-31]. Enablement helps reduce structural barriers to professional care, which are common in lower-resource settings [32]. Even so, broad BCW coverage did not produce broad support. EduKitchen-Toddlers Food Games used child-centered gamification to build food familiarity, and Autism Food Coach 2 offered basic caregiver tools for meal planning and mindful eating. Established programs such as BUFFET (Building Up Food Flexibility and Exposure Treatment) [33] and the Autism MEAL (Managing Eating Aversions and Limited Variety) Plan [34] combine family-based therapy, gradual food exposure, and sensory integration [35]. The apps offered far less. Limited evidence-based content is also common in nutrition and infant feeding apps more broadly [16].

Both apps also had acceptable but not high-quality scores on the MARS. EduKitchen-Toddlers Food Games scored 3.9 and Autism Food Coach 2 scored 3.7. Earlier MARS evaluations of child nutrition and feeding apps reported overall scores of about 3.0 to 4.0 out of 5 [16,17,36], so our evaluated apps fall in the upper part of that range. Those studies also reported limited engagement features, weak evidence-based content, poor aesthetic design, and little personalization. We found

XSL•FO
RenderX

similar limitations (Table 4). Neither app reported clinical validation. EduKitchen-Toddlers Food Games offered basic customization, and Autism Food Coach 2 had the lowest aesthetics score (3.3), reflecting a simple, utilitarian interface.

Table . Strengths and limitations of the 2 reviewed apps.

	EduKitchen-Toddlers Food Games	Autism Food Coach 2
Strengths	Provides an engaging and interactive learning experience with high-quality graphics and food-themed mini-games; includes basic caregiver customization	Offers tailored meal plans, progress tracking, and structured coaching tools that support caregiver-led feeding interventions
Limitations	No scientific validation; limited features for caregiver-led interventions or clinical feeding challenges	Limited visual design quality; lacks empirical validation through clinical trials

Closing this gap starts with bringing established feeding programs such as BUFFET and the Autism MEAL Plan into digital tools. Building an effective mHealth app requires several stakeholders. Clinicians, such as dietitians and occupational and speech therapists, can keep the clinical content faithful to the original programs. Caregivers know what fits into daily mealtimes, and poorly designed tools can add to their burden [14,15]. Developers can turn this input into features that families find easy and engaging to use. Researchers should then evaluate the tools.

We make 3 recommendations to support this work. First, developers should use structured co-design frameworks (eg, IDEAS [integrate, design, assess, and share] [37]) so that clinical experts and caregivers shape the tools from the start. This helps apps combine evidence-based strategies with informative, engaging features that serve caregivers and children. Second, quick reference guides and practical tutorials on behavioral theory would help developers and stakeholders during rapid ideation sessions, because comprehensive frameworks and training manuals can be hard to navigate. Third, researchers and developers should document feature efficacy through mixed methods evaluations. Quantitative measures, such as food variety and mealtime behavior, show whether a tool improves feeding outcomes. Qualitative methods show how caregivers and children use each feature and why some work better than others. Both are needed because clinical strategies do not always work the same way when delivered digitally [38].

Evidence to guide app design also remains limited, particularly across cultural contexts. Preferences for gamification vary widely with cultural values and social norms [39,40], but little is known about which gamification and engagement features work best in mHealth apps in different settings. Evidence is also thin on how to balance personalization with privacy for different users, such as privacy-conscious users and open sharers. Future research should address these questions. This work is most urgent in low-resource settings, where in-person feeding services are scarce [32] and digital tools could reach families who cannot access them [10-12]. This review provides a baseline against which to track how the app market changes.

Limitations

While this review provides a structured assessment of the current landscape, several limitations must be considered when interpreting these findings. App search results vary depending on the geographic location. As the research was conducted in Rwanda, the apps available are likely different from those accessible from other countries and regions. This search also used English keywords and only included apps available in English. To fully understand the scope of apps available to the African population, the search should be expanded to include French, Portuguese, Arabic, Spanish, and local languages. Although extensive search terms were used, some relevant apps may have been missed due to alternative labeling by developers. Furthermore, restricted features behind paywalls could not be fully evaluated. Lastly, while MARS provides a standardized assessment from a rater perspective, it may not fully capture caregiver preferences and usability. Future work could consider user-centered measures, such as the user version of the Mobile App Rating Scale (uMARS) [41], to assess app acceptability and perceived usefulness.

Future research should broaden the search strategy to mitigate geographic sampling bias, for example, by using multiregion app store accounts to more comprehensively capture global app availability. Co-design approaches that involve parents, clinicians, and speech or occupational therapists will be crucial in ensuring that future apps address both clinical needs and contextual realities. Particularly in resource-constrained settings, culturally tailored and clinically validated mHealth apps are urgently needed to support caregivers to manage ASD-related feeding challenges. Longitudinal studies evaluating behavioral outcomes and caregiver burden can further establish the real-world impact of these digital interventions.

Conclusions

This study highlights a critical shortage of mHealth apps specifically designed to support feeding challenges in children with autism. Using the BCW and MARS frameworks, we evaluated the functional strengths and theoretical underpinnings of the 2 eligible apps. Although both apps incorporated behavior change techniques and demonstrated moderate to high usability, neither was grounded in clinically validated feeding interventions. These findings underscore the need for user-centered, evidence-based app development that integrates behavioral, sensory, and skill-building supports tailored to the diverse needs of children with autism. Expanding access to effective, clinically validated, and culturally responsive mHealth tools is essential for advancing digital health equity, particularly in low-resource settings where traditional intervention services may be limited.

XSL•FO
RenderX

Acknowledgments

All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

Generative AI Use

Generative AI–assisted writing tools (eg, Grammarly) were used within the manuscript to suggest language improvements, including grammar correction, sentence clarity, and phrasing refinement.

Funding

This research was supported by the Consortium for Advanced Research Training in Africa (CARTA). CARTA is jointly led by the African Population and Health Research Center and the University of the Witwatersrand and funded by the Carnegie Corporation of New York (grants G-21-58722 and G-PS-23-60922), Sida (grant 16604), Norwegian Agency for Development Cooperation (Norad; grant QZA-21/0162), Oak Foundation (grant OFIL-24-091) and the Science for Africa Foundation to the Developing Excellence in Leadership, Training and Science in Africa (DELTAS Africa) program (Del-22-006) with support from Wellcome and the UK Foreign, Commonwealth & Development Office, and it is part of the European & Developing Countries Clinical Trials Partnership 2 (EDCTP2) program supported by the European Union. The funder had no involvement in the study design, data collection, analysis, interpretation, or the writing of the manuscript. The statements made and views expressed are solely the responsibility of the authors.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

Conceptualization: CCK, FM, SNA, NYD, ETL
Data curation: CCK (lead), FM (lead), ETL (supporting)
Formal analysis: CCK (lead), FM (lead), ETL (supporting)
Project administration: SNA, ETL
Visualization: CCK, FM, NYD
Writing – original draft: CCK, FM, NYD, SNA, ETL
Writing – review & editing: CCK, FM, NYD, ETL

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of screened apps.
[DOCX File, 30 KB - mhealth_v14i1e94908_app1.docx]

References

1. Data and statistics on autism spectrum disorder. CDC. 2025 May 27. URL: https://www.cdc.gov/autism/data-research/ [accessed 2026-10-05]
2. Autism. World Health Organization. 2025 Sep 17. URL: https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders [accessed 2026-10-05]
3. Nadon G, Feldman D, Gisel E. Feeding issues associated with the autism spectrum disorders. In: Fitzgerald M, editor. Recent Advances in Autism Spectrum Disorders - Volume 1: IntechOpen; 2013. [doi: 10.5772/53644]
4. Marshall J, Hill RJ, Ziviani J, Dodrill P. Features of feeding difficulty in children with autism spectrum disorder. Int J Speech Lang Pathol 2014 Apr;16(2):151-158. [doi: 10.3109/17549507.2013.808700] [Medline: 24001171]
5. Leader G, Tuohy E, Chen JL, Mannion A, Gilroy SP. Feeding problems, gastrointestinal symptoms, challenging behavior and sensory issues in children and adolescents with autism spectrum disorder. J Autism Dev Disord 2020 Apr;50(4):1401-1410. [doi: 10.1007/s10803-019-04357-7] [Medline: 31955310]
6. Liu X, Liu J, Xiong X, et al. Correlation between nutrition and symptoms: nutritional survey of children with autism spectrum disorder in Chongqing, China. Nutrients 2016 May 14;8(5):294. [doi: 10.3390/nu8050294] [Medline: 27187463]

XSL•FO
RenderX

7. Marinov D, Chamova R, Pancheva R. Micronutrient deficiencies in children with autism spectrum disorders compared to typically developing children – a scoping review. Res Autism Spectr Disord 2024 Jun;114:102396. [doi: 10.1016/j.rasd.2024.102396]
8. Sharp WG, Berry RC, McCracken C, et al. Feeding problems and nutrient intake in children with autism spectrum disorders: a meta-analysis and comprehensive review of the literature. J Autism Dev Disord 2013 Sep;43(9):2159-2173. [doi: 10.1007/s10803-013-1771-5] [Medline: 23371510]
9. Beaudoin AJ, Sébire G, Couture M. Parent training interventions for toddlers with autism spectrum disorder. Autism Res Treat 2014;2014(1):839890. [doi: 10.1155/2014/839890] [Medline: 24895534]
10. Rojas-Torres LP, Alonso-Esteban Y, Alcantud-Marín F. Early intervention with parents of children with autism spectrum disorders: a review of programs. Children (Basel) 2020 Dec 15;7(12):294. [doi: 10.3390/children7120294] [Medline: 33333900]
11. Lee J, Lim J, Kang S, et al. Mobile app-assisted parent training intervention for behavioral problems in children with autism spectrum disorder: pilot randomized controlled trial. JMIR Hum Factors 2024 Oct 28;11:e52295. [doi: 10.2196/52295] [Medline: 39466295]
12. Law GC, Neihart M, Dutt A. The use of behavior modeling training in a mobile app parent training program to improve functional communication of young children with autism spectrum disorder. Autism 2018 May;22(4):424-439. [doi: 10.1177/1362361316683887] [Medline: 29153002]
13. Bonnot O, Adrien V, Venelle V, Bonneau D, Gollier-Briant F, Mouchabac S. Mobile app for parental empowerment for caregivers of children with autism spectrum disorders: prospective open trial. JMIR Ment Health 2021 Sep 15;8(9):e27803. [doi: 10.2196/27803] [Medline: 34524101]
14. Houser L, McCarthy M, Lawer L, Mandell D. A challenging fit: employment, childcare, and therapeutic support in families of children with autism spectrum disorders. J Soc Serv Res 2014 Oct 20;40(5):681-698. [doi: 10.1080/01488376.2014.930944]
15. Bradshaw J, Wolfe K, Hock R, Scopano L. Advances in supporting parents in interventions for autism spectrum disorder. Pediatr Clin North Am 2022 Aug;69(4):645-656. [doi: 10.1016/j.pcl.2022.04.002] [Medline: 35934491]
16. Cheng H, Tutt A, Llewellyn C, et al. Content and quality of infant feeding smartphone apps: five-year update on a systematic search and evaluation. JMIR mHealth uHealth 2020 May 27;8(5):e17300. [doi: 10.2196/17300] [Medline: 32459187]
17. Francis L, Spaulding EM, Bloom I, Patel A, Perrin N. A systematic appraisal of the information, engagement, aesthetic and functional quality of nutrition-related smartphone apps for children and adolescents. Public Health Nutr 2023 Jul;26(7):1368-1379. [doi: 10.1017/S1368980023000526] [Medline: 36912110]
18. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
19. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42. [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
20. Eberhard DM, Simons GF, Robinson AJ. English. Ethnologue: languages of the world. 2026. URL: https://www.ethnologue.com/language/eng [accessed 2026-10-06]
21. Parry G. English language teaching in Africa (ELT): a comprehensive overview of ELT in Africa. Global Services in Education. 2025 Mar 9. URL: https://www.gsineducation.com/school-investment/english-language-teaching-in-africa-elt-a-comprehensive-overview-of-elt-in-africa [accessed 2026-10-06]
22. Vailshery LS. Google Play Store - statistics & facts. Statista. 2025 Dec 17. URL: https://www.statista.com/topics/9929/google-play-store [accessed 2026-10-05]
23. Slotta D. Apple App Store - statistics & facts. Statista. 2025 Dec 17. URL: https://www.statista.com/topics/9757/apple-app-store [accessed 2026-10-06]
24. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLOS Med 2009 Jul 21;6(7):e1000100. [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
25. Anderson DK, Lord C, Risi S, et al. Patterns of growth in verbal abilities among children with autism spectrum disorder. J Consult Clin Psychol 2007 Aug;75(4):594-604. [doi: 10.1037/0022-006X.75.4.594] [Medline: 17663613]
26. Kim SH, Bal VH, Lord C. Longitudinal follow-up of academic achievement in children with autism from age 2 to 18. J Child Psychol Psychiatry 2018 Mar;59(3):258-267. [doi: 10.1111/jcpp.12808] [Medline: 28949003]
27. Vogindroukas I, Stankova M, Chelas EN, Proedrou A. Language and speech characteristics in autism. Neuropsychiatr Dis Treat 2022 Oct 14;18:2367-2377. [doi: 10.2147/NDT.S331987] [Medline: 36268264]
28. Suryawan A, Jalaludin MY, Poh BK, et al. Malnutrition in early life and its neurodevelopmental and cognitive consequences: a scoping review. Nutr Res Rev 2022 Jun;35(1):136-149. [doi: 10.1017/S0954422421000159] [Medline: 34100353]
29. Adams SN. Feeding and swallowing issues in autism spectrum disorders. Neuropsychiatr Dis Treat 2022 Oct 14;18:2311-2321. [doi: 10.2147/NDT.S332523] [Medline: 36276431]
30. Adams SN, Verachia R, Coutts K. Mealtime challenges and food selectivity in children with autism spectrum disorder in South Africa. S Afr J Clin Nutr 2022 Jun 17;35(2):78-84. [doi: 10.1080/16070658.2021.1954794]

XSL•FO
RenderX

31. Adams SN, Matsimela N. Feeding children with autism in South Africa: the teachers' perspectives. Afr J Disabil 2023;12:1252. [doi: 10.4102/ajod.v12i0.1252] [Medline: 38059155]
32. Kakooza-Mwesige A, Bakare M, Gaddour N, Juneja M. The need to improve autism services in lower-resource settings. Lancet 2022 Jan 15;399(10321):217-220. [doi: 10.1016/S0140-6736(21)02658-1] [Medline: 34883050]
33. Kuschner ES, Morton HE, Maddox BB, de Marchena A, Anthony LG, Reaven J. The BUFFET program: development of a cognitive behavioral treatment for selective eating in youth with autism spectrum disorder. Clin Child Fam Psychol Rev 2017 Dec;20(4):403-421. [doi: 10.1007/s10567-017-0236-3] [Medline: 28534237]
34. Sharp WG, Burrell TL, Jaquess DL. The Autism MEAL Plan: a parent-training curriculum to manage eating aversions and low intake among children with autism. Autism 2014 Aug;18(6):712-722. [doi: 10.1177/1362361313489190] [Medline: 24101716]
35. Clancy AM, O'Driscoll K, Regan J, Walsh IP. Interventions for autistic children with behavioural feeding needs: a scoping review. Adv Autism 2025;11(4):348-360. [doi: 10.1108/AIA-03-2025-0029]
36. Martinon P, Saliasi I, Bourgeois D, et al. Nutrition-related mobile apps in the French app stores: assessment of functionality and quality. JMIR mHealth uHealth 2022 Mar 14;10(3):e35879. [doi: 10.2196/35879] [Medline: 35285817]
37. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016 Dec 16;18(12):e317. [doi: 10.2196/jmir.5927] [Medline: 27986647]
38. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232. [doi: 10.2196/jmir.7126] [Medline: 28663162]
39. Toda A, Klock ACT, Pereira FD, Rodrigues LA, Palomino PT, Lopes V. Towards the understanding of cultural differences in between gamification preferences: a data-driven comparison between the US and Brazil. Presented at: Proceedings of the 15th International Conference on Educational Data Mining; Jul 24-27, 2022; Durham, United Kingdom p. 560-564 URL: https://educationaldatamining.org/EDM2022/proceedings/2022.EDM-posters.60/index.html [accessed 2026-10-05]
40. Farinha J, de Fatima Pina M. Cross-cultural perspectives on gamification in business environments: how cultural differences influence design and reception. In: Evolving Strategies for Organizational Management and Performance Evaluation: IGI Global Scientific; 2025:77-88. [doi: 10.4018/979-8-3373-0149-5.ch005]
41. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR mHealth uHealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849] [Medline: 27287964]

Abbreviations

ASD: autism spectrum disorder
BCW: behavior change wheel
BUFFET: Building Up Food Flexibility and Exposure Treatment
IDEAS: integrate, design, assess, and share
MARS: Mobile App Rating Scale
MEAL: Managing Eating Aversions and Limited Variety
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
uMARS: user version of the Mobile App Rating Scale

Edited by L Buis; submitted 08.Mar.2026; peer-reviewed by M Esposito, N Admodisastro; revised version received 20.Sep.2026; accepted 28.Sep.2026; published 10.Nov.2026.

Please cite as:
Mukarage F, Cofie Kuzagbe C, Nandi Adams S, Yves-Roland Douha N, Talina Luhanga E
Mobile Apps Addressing Feeding of Children With Autism: Systematic Search and Content and Quality Analysis
JMIR Mhealth Uhealth 2026;14:e94908
URL: https://mhealth.jmir.org/2026/1/e94908
doi:10.2196/94908

© Fabrice Mukarage, Christopher Cofie Kuzagbe, Skye Nandi Adams, N'guessan Yves-Roland Douha, Edith Talina Luhanga. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR

XSL•FO
RenderX

mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobile App–Supported Self-Management for Chronic Low Back Pain: Realist Evaluation

Rebecca Hunter[1,2], PhD; Trish Gorely[1], PhD; Michelle Beattie[1], PhD

[1]Centre for Rural Health Sciences, University of the Highlands and Islands, Raigmore Hospital, Old Perth Road, Inverness, United Kingdom
[2]School of Communities and Education, Northumbria University, Newcastle, United Kingdom

Corresponding Author:
Michelle Beattie, PhD
Centre for Rural Health Sciences, University of the Highlands and Islands, Raigmore Hospital, Old Perth Road, Inverness, United Kingdom

Abstract

Background: As the world's population ages, the prevalence of chronic low back pain (CLBP) is increasing, placing a substantial burden on individuals and health care systems. Mobile health (mHealth) apps offer a potentially scalable solution to support self-management, but little is known about how, why, for whom, and under what circumstances such tools work in real-world settings.

Objectives: This study aimed to test and refine 3 program theories—developed through a previous realist review—on how mobile apps support CLBP self-management. The goal was to understand the key contextual factors and mechanisms that influence when and why a digital self-management intervention may succeed or fail.

Methods: A realist evaluation was conducted using one-on-one telephone interviews with 9 participants who had used the Curable app for 3 months to self-manage their CLBP. Realist interviews followed a teacher-learner cycle to explore, test, and refine the program theories. Abductive and retroductive analysis was used to develop context-mechanism-outcome configurations (CMOCs), which were synthesized into refined theories of digital self-management in chronic pain.

Results: A total of 20 CMOCs were constructed, supporting 3 overarching program theories centered on empowerment, self-management burden, and timing. First, the app was empowering when it offered credible and accessible knowledge that helped participants understand their pain, build confidence, and reduce reliance on health care providers. However, engagement depended on individual beliefs and expectations: those with strong biomedical views struggled to connect with the app's psychosocial framing. Second, while the app could ease the burden of self-management by offering support between appointments, it could also increase burden during flare-ups, when users lacked the capacity to engage. Features such as proactive content delivery and low-demand interfaces were viewed as essential for continued use. Third, timing emerged as a key factor. Early introduction was beneficial for some, but others needed to first accept the chronicity of their condition before they were ready to engage with self-management tools. Trust in the source recommending the app also influenced engagement. While clinician endorsement was often valued—especially early in the self-management journey—participants who had experienced unmet needs or disillusionment in clinical encounters reported that peer recommendations or nonclinical sources held greater weight. This highlights the importance of aligning recommendations with individuals' evolving relationships with authority and trust.

Conclusions: Mobile apps like Curable (Curable Inc) can support empowerment and continuity of care in CLBP, but their success depends on personalization, timing, and relational dynamics. To prevent feelings of abandonment, such tools should be introduced as an adjunct to—rather than a replacement for—ongoing clinical support.

(JMIR Mhealth Uhealth 2026;14:e66435) doi:10.2196/66435

KEYWORDS

chronic low back pain; back pain; realist evaluation; self-management; mHealth; mobile health; mobile app; smartphone; digital health; digital technology; digital intervention

Introduction

In 2020 nearly 10% of the world's population (approximately 619 million people) suffered from low back pain [1]. Studies estimate that between 5% - 15% of people will go on to develop chronic low back pain (CLBP; pain lasting longer than 3 months) [2,3]. People with CLBP are frequent health care users, which places a significant strain on medical services [3,4]. This demand is expected to rise as the global population grows older and the prevalence of CLBP increases with age [5]. Mobile health (mHealth) apps are becoming powerful tools in assisting health care services meet the rising demand of caring for people with chronic health conditions [6-8].

XSL•FO
RenderX

mHealth apps are software programs designed to be downloaded and run on smartphones or tablets with the aim of maintaining, improving, or managing the user's health [9]. mHealth apps have been recognized as a promising way of delivering timely, cost-effective, and individualized care [10]. It has been suggested that using mHealth apps to treat people with CLBP could reduce the number of general practitioner visits, provide quicker virtual consultations, promote self-management, and improve the psychological impact of pain [11]. Studies suggest that self-management apps for CLBP can be effective in reducing pain and disability [12-15].

However, recent reviews evaluating mHealth interventions for self-managing low back pain highlight mixed findings and persistent challenges. While some studies report improvements in pain and function, others point to inconsistent evidence, low methodological quality, and poor engagement [16,17]. Many apps reflect a narrow biomedical focus, lacking personalization and theoretical underpinning, and often fail to incorporate psychological and social dimensions of pain management. This limits their potential impact, particularly for people with complex, long-term conditions like CLBP [18]. These limitations underscore the need for theory-led research to better understand how, for whom, and in what contexts mHealth interventions work.

The Medical Research Council (MRC) encourages a theory-based approach to evaluation when reviewing complex interventions [19]. Theories can help identify behavioral mechanisms and contexts likely to bring about a desired outcome and, in so doing, potentially increase the effectiveness of both the intervention and its implementation [20]. This type of research is arguably of greater value to policy makers and program developers as the results go beyond reporting on effect sizes, which may be difficult to translate into real-world implementation [21,22]. Realist evaluations are a theory-led approach to evaluation designed to provide guidance on how to implement a program in a particular setting while avoiding potential obstacles along the way [23].

This study addressed the gap in theory-based research on self-management apps for CLBP. Aligned with the MRC's guidance on evaluating complex interventions, a realist evaluation was used to explore a commercially available self-management app. We selected Curable (Curable Inc.), a generic pain management app ranked among the top 3 in a systematic review of 19 chronic pain apps [24], notable for its embedded pain education—a key component of effective self-management [25]. Unlike apps specifically targeting musculoskeletal or low back pain, Curable's broader biopsychosocial approach aligns with current understanding that chronic pain, including CLBP, requires multifaceted management [26]. Rather than evaluating Curable's effectiveness, the app was used as a case study to test and refine program theories developed in a previous realist review [27]. This study aimed to extend their applicability by evaluating these theories in a real-world setting with individuals living with CLBP. Ultimately, the evaluation sought to refine existing theories and contribute to a more robust framework explaining who benefits from a self-management app for CLBP, why, and under what circumstances.

Methods

Ethical Considerations

The study received ethical approval from the University of the Highlands and Islands Research Ethics Committee (ETH2122-0819) and was conducted in accordance with the principles of the Declaration of Helsinki. Written informed consent was obtained from all participants before data collection. Participants were informed of the nature and potential consequences of the study and of their right to withdraw at any time without consequence. All data were anonymized before analysis and confidentiality was maintained in accordance with UK data protection legislation (UK GDPR). While no honorarium was provided, participants were offered a 12-month subscription to the Curable app as a noncoercive form of compensation, allowing continued access beyond the 12-week study period should they choose to continue using the app.

Theoretical Framework

Two substantive theories were used in this study to inform the testing and evaluation of the program theories: burden of treatment theory and empowerment theory. The burden of treatment theory, developed by May et al [28], focuses on the physical and emotional demands that patients face in managing chronic conditions, highlighting the work involved not only for the patients themselves but also for their relational networks (eg, family and health care providers). This theory was selected due to its emphasis on the complexity of managing CLBP within a health care system that increasingly expects patients to take responsibility for their care. It provided a useful lens for exploring how the structural, institutional, and professional dynamics interact with the experience of self-management in this context.

The empowerment theory used here draws from Lee and Koh's [29] model, which combines both the behavioral and psychological dimensions of empowerment. This theory was selected because it helps to explore the role of health care professionals (HCPs) in empowering patients to take control of their self-management, aligning with the goal of understanding how digital health tools, like the Curable app, might influence patient empowerment in managing CLBP. These theories guided the development of program theories by providing conceptual frameworks to understand the interactions between health care systems, patient behaviors, and technology. They also helped to identify key themes in the data, particularly around the concepts of support, control, and autonomy, which were central to understanding how participants engaged with the app. The use of substantive theories in this way is typical in realist research, as it allows for a richer, more contextually grounded exploration of program theories and the mechanisms at play [30].

Clarification of Key Terms

To ensure conceptual clarity and consistency throughout the review, we outline below how key terms have been defined and operationalized. This includes our understanding of context, mechanism, and outcome within a realist framework, as well

XSL•FO
RenderX

as how we have defined self-management, which is central to the intervention under study.

The following key terms are used consistently in this paper and are defined as follows:

1. Self-management: our definition of self-management was informed by a number of key papers in the field of self-management of chronic conditions, including Barlow et al [31] Bodenheimer et al [32], Lorig et al [25], and Wagner et al [33]. In this study, self-management was defined as the ability of an individual—supported by health care providers, family, and community—to acquire the skills and confidence to manage their chronic condition on a daily basis. This process was considered dynamic, allowing individuals to adapt to the fluctuating nature of CLBP and the changing intrinsic and extrinsic factors in their lives.
2. Context: the dynamic and influential conditions or circumstances that shape how an intervention operates. Contexts are not static backdrops but active forces that determine whether and how mechanisms are triggered. These may include individual factors (eg, psychological state and previous knowledge), social or environmental conditions, or institutional settings, all of which influence how participants engage with and respond to the intervention.
3. Mechanism: the underlying processes that explain how and why an intervention works (or does not). Mechanisms involve the interaction between the resources provided by the intervention (eg, app features) and participants' responses to these resources (eg, changes in confidence, engagement, or self-efficacy). Mechanisms are context-dependent and are activated based on the specific conditions in which they operate.
4. Outcome: the results or effects produced by the intervention, which occur because of the activation of mechanisms within a specific context. Outcomes may include changes in behavior, attitudes, or health status.

The definitions of context, mechanism, and outcome used in this study are drawn from our previously published work on realist methodology [34]

Study Design

A realist evaluation of a commercially available chronic pain self-management app was undertaken to test and refine program theories, derived from a previous realist review [27]. A realist evaluation is a systematic approach to studying how people respond to social programs delivered in real-life complex environments and how this influences program outcomes [35]. Realist logic is grounded in the concept of generative causation, which suggests that observable outcomes are brought about by underlying, often unseen, causal mechanisms that operate differently depending on context [36]. These mechanisms help explain why some people may benefit from an intervention and others may not. Context-mechanism-outcome configurations (CMOCs) are central to realist research, providing a way to understand how and why an intervention works (or does not) in specific contexts [37]. Another way to understand CMOCs is as testable propositions that explore the interaction between context, mechanisms, and outcomes, offering an evidence-based framework for explaining why different results might emerge in varying settings or populations [38]. Using a specific case study to test and refine theoretical understandings is a recognized and valuable application within realist research, particularly when evaluating complex interventions like mobile health apps for CLBP [39]. The aim of a realist evaluation is to develop theories as to who might benefit from an intervention or program, why, and in what circumstances.

The study was carried out remotely in Scotland with 9 participants living with CLBP. Participants were given licenses to use a commercially available app (Curable) for 12 weeks to help them self-manage their condition. No restrictions or limitations were placed on how the participants used the app. At the end of 12 weeks, participants engaged in one-to-one, semistructured realist interviews. The study adhered to the RAMESES II (Realist And Meta-narrative Evidence Syntheses: Evolving Standards II) reporting standards for realist evaluation [40] (refer to Multimedia Appendix 1)

Intervention

The Curable app, developed by the US-based company Curable Inc, is a commercially available digital health intervention designed to support individuals in the self-management of chronic pain. It delivers content via a smartphone or web browser through a combination of audio lessons, written articles, and interactive exercises. The app's program is grounded in pain neuroscience education and incorporates elements from cognitive behavioral therapy, mindfulness-based stress reduction, and expressive writing. Users are guided through the content by a virtual chatbot, which provides personalized recommendations and prompts based on user input. Interactive features include reflective journaling, symptom tracking, quizzes, and meditative or breathing exercises. The app encourages daily engagement, and its structure allows users to progress at their own pace through thematic content modules that address the psychological, emotional, and neurophysiological aspects of chronic pain [41].

Recruitment

Participants were recruited through 3 concurrent channels. In the first stream, individuals who had participated in the previous realist synthesis and consented to future contact were emailed directly by the lead researcher. In the second stream, an online advertisement was posted on a study-specific Twitter account (@RealBackstory), and interested individuals responded via email. In the third stream, several pain charities and third-sector organizations shared the study advertisement on their websites and/or in their digital newsletters. Recruitment was entirely conducted online, with participants self-identifying by emailing the lead researcher to express their interest. As such, most participants were self-selected. However, to ensure diversity in experience and explore rival theories, purposive sampling was used to include a small subset (n=3) of individuals identified by a third-sector organization as having had particularly positive experiences with in-person self-management programs for CLBP. This purposive sampling allowed us to examine potentially contrasting perspectives, enhancing the explanatory power and trustworthiness of the findings [42].

XSL•FO
RenderX

Eligibility Criteria

The inclusion and exclusion criteria are shown in Textbox 1.

Textbox 1. The inclusion and exclusion criteria.

Inclusion criteria:

- Any sex
- Aged 18 years or older
- Self-managing chronic low back pain (defined as pain in the lower back lasting >3 months)
- Have seen a medical practitioner about their back pain prior to being involved in this study
- Not receiving active medical treatment for back pain (surgery, physiotherapy, scans, etc)
- May still be taking prescribed analgesia for their low back pain
- Have a smartphone capable of running the Curable app
- Able to participate in a telephone interview

Exclusion criteria:

- Lacking capacity to provide informed consent
- Chronic pain condition that does not include low back pain
- Unable to commit to a 12-week study

Data Collection and Analysis

In line with realist methodology, the number of interviews was guided by the principle of providing rich and relevant data for theory testing rather than aiming for saturation [43]. Data collection ceased once participants had contributed sufficient conceptual depth to meaningfully explore and refine program theories. This approach is consistent with realist evaluation practice [44] and supported by the concept of information power [45], which recognizes that the more relevant and information-rich the data, the fewer participants are needed.

One-to-one, semistructured realist interviews were conducted via telephone by the lead researcher (RH), a chronic pain specialist physiotherapist with qualitative interviewing experience and training in realist methodology. The realist interviews used a teacher-learner cycle, which is a key approach in realist evaluation [44]. In the teacher-learner cycle, the interviewer initially adopts a “teacher” role by presenting a program theory for the interviewee to consider. The interviewee then engages with the theory, drawing on their lived experience to confirm, refute, refine, or elaborate on it. Through this interaction, the interviewer becomes the learner, and the interviewee becomes a co-constructor of knowledge, actively contributing to theory refinement [46]. This cyclical dynamic enhances theoretical sensitivity and contributes to the trustworthiness of the realist research process [46] Interviews were audio recorded with consent and transcribed by a third party. The interview guide was based on the theories from a previous realist review (refer to Multimedia Appendix 2).

Data from the transcripts were analyzed to test the 3 program theories developed from the preceding realist review. These program theories were underpinned by 16 CMOCs; CMOCs are used in realist evaluation to explain how and why particular outcomes occur in specific contexts [34]. Transcripts were coded via repeated rounds of direct, indirect, and holistic coding to identify key causal mechanisms and contextual factors. This coding framework was based on training received from the Center for Advancement in Realist Evaluation and Synthesis (CARES; Jagosh, unpublished data, 2020).

Direct coding was used sparingly and involved a “cut and paste” approach, where brief but pertinent passages from the transcripts were extracted and added to a mind mapping software (XMind; version 22.10). These excerpts typically served as key quotes to support or challenge the CMOCs from the realist review. Indirect coding involved annotating sections of the transcript in the margins, with these annotations extracted and added to the mind map under the relevant CMOC. Each annotation was linked to the original data with participant ID and line numbers for transparency. Holistic coding examined larger sections or entire transcripts to create broader inferences, which were similarly extracted and placed in the mind map.

Inferences were made from the data using retroductive and abductive logic [47,48]. Retroduction refers to reasoning backward from observed outcomes or phenomena to theorize about the underlying causal mechanisms, often combining both inductive and deductive reasoning [48]. Abduction has been described as a form of “reading between the lines” to imagine the underlying causal mechanisms that are typically not directly observable [49]. Jagosh [48] refers to abduction as the creative process of generating inferences to hypothesize what might be happening beneath the surface. It involves reasoning based on “educated guesswork” and “informed hunches,” where researchers draw on existing knowledge to propose plausible explanations for observed phenomena.

The analytical process involved testing the CMOCs with both confirming and disconfirming data extracted from the transcripts. Confirming data referred to instances where the findings aligned with the proposed theory, supporting the link between context, mechanism, and outcome. Disconfirming data presented cases that contradicted the expected outcomes, challenging the initial theory. These contrasting pieces of

XSL•FO
RenderX

evidence were placed side by side in the mind map to refine or adjust the CMOCs. Additionally, counterfactuals were explored—looking for instances where the absence of a certain context or mechanism led to a different outcome than expected. In cases where these contradictions or counterfactuals highlighted different theoretical possibilities, the CMOCs were split or revised to incorporate rival theories. This ensured the complexity of the data was preserved and not flattened out to create uniform themes [50]. The rigor of the final theories was judged using the principles of explanatory coherence [42,51] and was strengthened through an iterative comparison of evidence across participants. This involved assessing how well the findings aligned with existing literature and considering alternative explanations. Additional rigor was supported by ongoing discussions within the research team and peer review. While formal stakeholder engagement was not undertaken at this stage, informal discussions with a participant—who later became a Patient and Public Involvement (PPI) lead in follow-on research—also helped sense-check emerging interpretations.

Realist Review Overview

A realist review was conducted before this evaluation to develop program theories about how and for whom self-management apps might work for people with CLBP. The review, published elsewhere [27] drew on 57 sources, including peer-reviewed literature, UK government policy documents, national pain charity resources, and nontraditional materials such as blogs, social media, and book chapters to capture a broad range of perspectives. To further enhance the depth and relevance of the findings, 19 realist interviews were conducted with people with CLBP, clinicians, and digital health designers. These interviews were integrated directly into the analysis to ensure that the voices and lived experiences of people with CLBP were central to theory development. The review generated 16 CMOCs, which were grouped into 3 overarching program theories: empowerment, self-management burden, and timing. These are summarized in Table 1 and referenced throughout the results section, where they were further tested and refined.

XSL•FO
RenderX

Table . Context-mechanism-outcome refinement.

Realist synthesis CMOCs[a]	After testing	Realist evaluation CMOCs	Sources	Example quote
CMOC 1. Convenience, accessibility, and choice Traditional NHS-led[b] self-management programs for CLBP[c] provide participants with little choice in how, when, and where they are delivered (C). A self-management app for CLBP can be accessed at a time and location that is convenient to the user (M), which restores a person's sense of control and autonomy (O).	Refined	*CMOC 1. Choice and flexibility* When a person with CLBP has distinct preferences and varying levels of motivation for self-management (C), the app being adaptable and flexible enough to accommodate the user's wants (I) leads to the patient experiencing a sense of agency, choice, and control over their self-management program (M), ultimately resulting in sustained use of the app and adherence to their self-management program (O).	Participants: 2,5,7,9	***...****having time to go out for a walk and listen to it* [the app] *was quite useful. But some of the content wasn't great when you're actually on the move*. [Participant 7] *...it just didn't suit me and my lifestyle*. [Participant 5]
CMOC 2. Knowledge and self-reliance Many people with CLBP rely on HCPs for support because they do not know how to manage their symptoms (**C**). By providing the user with knowledge, advice, and strategies to self-manage CLBP, a mobile app enables the user to gain confidence and agency (**M**) to manage their condition on their own (**O**).	Split	*CMOC 2. Knowledge is empowering.* For people with CLBP who feel dependent on health care professionals due to a lack of confidence in managing their symptoms independently (C), being offered an app that provides accessible information and practical self-management strategies leads them to feel more empowered and confident in their ability to manage their condition (M). As a result, they become more self-reliant and reduce their dependence on healthcare professionals (O). *CMOC 3. Knowledge reduces fear* For people with CLBP who experience fear and anxiety related to the uncertain causes of their ongoing pain (C), an app that provides education about CLBP can increase their sense of understanding, control, and agency over their condition (M), ultimately leading to greater confidence and engagement in self-management (O).	Participants: 2,3,4,6,7 Participants: 1,4,6,7	***...****I've gone to them* [GP's] *looking for solutions an come away with nothing so I'm disappointed. Whereas when I found out about it* [pain] *myself, I could understand it more*. [Participant 3] *...managing that fear is absolutely critical to pain management itself and an app can help with that, without a doubt*. [Participant 4]

XSL•FO
RenderX

Realist synthesis CMOCs[a]	After testing	Realist evaluation CMOCs	Sources	Example quote
CMOC 3. Personalisation A person with CLBP needs to be able to recognize themselves in the advice and information the app provides (**C**) so that they can trust what they are being told (**M**) otherwise they unlikely to engage with the app because they do not consider it as being relevant to their situation (**O**).	No change	*CMOC 4. Personalisation* For a person with CLBP to engage with the app, they need to see the advice and information as personally relevant to their situation (C). This recognition fosters trust in the app's recommendations (M), which in turn leads to increased engagement because the individual perceives the app as a useful tool for managing their condition (O).	Participants: 1,4,5,6,7,9	*I'm not going to turn it* [the app] *on because I've started to hear more stuff that doesn't mean anything to me. It doesn't resonate with me*. [Participant 7]
CMOC 4. Hope If a mobile the app fails to provide the user with options that have not been tried before (**C**) then the initial hope they may have felt at being offered something that might alleviate their pain (**M**) can turn to bitterness, disappointment and sometimes anger (**O**).	Refined	*CMOC 5. Biomedical mindset* For people with CLBP who have a biomedical mindset (C), a self-management app that presents an alternative approach to understanding and managing CLBP may lead to frustration, annoyance, and disappointment (M), as the solution offered does not align with their expectations for managing their pain (O).	Participants: 1,5,6,8,9	*Pain is pain - the pain is not in my head – I have a diagnosis..I gave up on it because I just thought 'this is ridiculous.' I just felt that it wasn't teaching me anything I didn't already know. It wasn't helping me.* [Participant 6]
CMOC 5. Adjunct to care If a self-management app was used as an adjunct to care and not a replacement (**C**) then HCPs[d] are likely to welcome the tool as it helps them to deliver ongoing support remotely (**M**) thereby enabling them to treat more patients in their limited clinic time (**O**).	Refined	*CMOC 7. Enhancing Patient Confidence Between Appointments* Between healthcare appointments, when reduced contact can leave people with CLBP feeling vulnerable (C), giving them a self-management app can lead to the patient feeling reassured that they have timely and convenient access to ongoing support and answers (M), ultimately increasing their confidence to go longer before they need to see a HCP (O).	Participants: 8,9	***...****it* [the app] *helps until they maybe get an appointment.until maybe they go and have a onceover with a professional or a doctor – absolutely, great!* [Participant 8]

XSL•FO
RenderX

Realist synthesis CMOCs[a]	After testing	Realist evaluation CMOCs	Sources	Example quote
CMOC 6. Burden of care If people with CLBP lack the internal and or external resources to engage with a self-management app (**C**) then this can cause further stress and frustration (**M**) which adds to their burden of having to manage long term back pain (**O**).	Split	*CMOC 8. Proactive support* For individuals with CLBP experiencing severe symptoms that inhibit their ability to engage mentally with tasks (C), an app that proactively selects and plays content they have found helpful in the past reduces cognitive fatigue and the stress of decision-making (M), thereby making it easier to engage with the app, particularly during flare-ups (O) *CMOC 9. Voice activation* For individuals with CLBP experiencing severe flare-ups that impair their ability to engage physically or mentally with tasks (C), an app with voice activation (resource) reduces physical effort and fatigue (M), enabling continued access to support even during extreme pain (O).	Participants: 2,7,9 Participants: 2,7,9	*...at times when you are in so much pain.you just want some sort of reassurance or someone to come with an idea, because you can't think of anything for yourself.* [Participant 9] *Voice activation is probably the only chance you've got because sometimes the pain is so unbearable. the only way I could actually unlock it* [the app] *would be voice activation.* [Participant 7]

XSL•FO
RenderX

Realist synthesis CMOCs[a]	After testing	Realist evaluation CMOCs	Sources	Example quote
CMOC 7. Monitoring Monitoring their progress with a mobile app and sharing this data with a HCP (**C**) can help a person with CLBP convey more of a 'complete picture' of how they are managing their condition (**M**) and thereby improve the communication and quality of a healthcare consultation (**O**).	Refined New New	*CMOC 11. Improved communication* During healthcare consultations, when people with CLBP find it challenging to accurately recall and articulate the details of their condition and management (C), using an app to record and share this data with an HCP can lead to the patient feeling less stressed and more in control (M), which improves their ability to communicate and enhances the quality of the healthcare consultation (O). *CMOC 6. HCP buy-in needed* When a HCP does not see the value in the app's recorded data or dismisses its utility (**C**) then this disempowers the person with CLBP by devaluing their contribution to the consultation (**M**) which can adversely affect the therapeutic relationship from the patient's perspective (**O**). *CMOC 12. Monitoring is burdensome* For people with CLBP who are constantly aware of their pain and benefit from strategies that allow them to mentally switch off from it (C), repeated prompts by an app to record symptoms can disrupt this mental disengagement, adding an additional burden to self-managing CLBP and serving as a constant reminder of their ill health (M), which can worsen the pain experience (O).	Participants: 1,2,4,9 Participants: 1,4,7 Participants: 3,5	*...you have to make them* [healthcare professionals] *believe you, and it creates so much stress in that interaction, and if you were doing than on an app then you're taking the stress out of it.* [Participant 4] *...when I took that to my neurologist he just went 'but how do I know that that is genuinely like the right information?'"* [Participant 1] I found it [monitoring] very difficult to do and it actually made the pain worse because I was focusing on it. [Participant 5]
CMOC 8. Meaningful consultations Using an app to record and share their data with an HCP (either before or during the appointment) (**C**) can maximize limited consultation time (**M**) and thereby reduce the frustration felt by patients' valuable consultation is wasted bringing HCPs 'up to speed'(**O**).	Refined	*CMOC 10. Making consultations more effective/efficient* In healthcare settings with short patient appointments (eg, 10 min) where patients face the burden of using that time to provide background information (C), using an app to record and share pertinent data with an HCP ahead of time can lead to the patient feeling a sense of reduced frustration due to the expectation of a more efficient and focused consultation (M), ultimately leading to a more satisfactory consultation (O).	Participants: 4,5,6,7,9	…when you go and see him [the GP], he's got the information there [on the app]… so I think there's an opportunity for the GP to use it and save them a lot of time and a lot of work. [Participant 9]

XSL•FO
RenderX

Realist synthesis CMOCs[a]	After testing	Realist evaluation CMOCs	Sources	Example quote
CMOC 9. Abandonment By providing support as well as a means by which to contact a HCP if needed (**C**) a self-management app provides users with the reassurance of a 'safety net' should they feel they need additional support (**M**) thereby mitigating feelings of abandonment (**O**). *CMOC 10. Safety net* A self-management app that enabled a person with CLBP to maintain contact with a HCP (**C**) can provide a reassuring 'safety net' (**M**) leaving them more confident to be discharged from the healthcare service (**O**).	Merged	*CMOC 13. Ongoing support* After discharge, a person with CLBP needs occasional reviews from a HCP in addition to a self-management app (**C**) to reassure them that they have not been left to manage their condition on their own (**M**) and prevent feelings of abandonment (**O**).	Participants: 1,3,4,5,	A follow-up with a real live person is always a good idea in my book, just to see how they're getting on. I wouldn't want them to be given the app and say 'There you are, bye bye'. [Participant 3]
CMOC 11. Social Isolation In the absence of supportive 'real life' relationships (**C**) a self-management app with a chat bot feature that has been designed to communicate in a human-like way to offer comfort and reassurance (**M**) can help someone with CLBP feel less alone (**O**).	Split	*CMOC 14. Chatbot is supportive* For people with CLBP who experience social anxiety, fear of judgment, or find social interaction draining (C), the non-judgmental and socially undemanding nature of an app's chatbot (Intervention Feature) can reduce feelings of shame, guilt, and loneliness (Mechanism), ultimately helping to reduce feelings of social isolation (Outcome). *CMOC 15. Chatbot is poor proxy* For people with CLBP who find connecting with fellow human beings to be supportive and emotionally beneficial (C), the lack of warmth and empathy from an app's chatbot (Intervention Limitation) can lead to the patient experiencing the interaction as fake, cold, and devoid of meaning (Mechanism), making the chatbot an inadequate substitute for human support (O).	Participants: 1,3,4,7,8 Participants: 1,7,8	*...if you've been isolated for so long then the very act of meeting someone, getting to know them.can be quite intimidating.whereas I suppose something, you know, that is not quite real would be easier in some regards.* [Participant 3] *...it's the personal thing, that's the difference.it's* [the chatbot] *not real, you're not seeing facial expressions of somebody, how they react to you. The A.I* [artificial intelligence] *it's robotic basically.,it's cold*. [Participant 8]
CMOC 12. Acceptance If a person with CLBP remains steadfast in their search to find a cure for their back pain (**C**) and are unwilling to accept an active role in self-managing their condition (**M**) then they are likely to be disappointed with a self-management app because it does not rid them of their pain (**O**).	Refined	*CMOC 16. Expectations of a cure* For people with CLBP who are still seeking a cure for their back pain (C), a self-management app may not meet their expectations (M), leading to disappointment and a lack of engagement (O	Participants: 2,3,9	*Once you accept it and adjust your lifestyle around it and work with it not against it, any sort of information you can gather through the app* [is helpful]. *But it's acceptance. A lot of people won't accept that they've got a long-term issue*. [Participant 9]

XSL•FO
RenderX

Realist synthesis CMOCs[a]	After testing	Realist evaluation CMOCs	Sources	Example quote
CMOC 13. Believed Before a person with CLBP can accept the need to self-manage their condition they need to feel that HCPs believe them (**C**). Feeling believed triggers a sense of reassurance that they have been taken seriously (**M**) which makes them trust the HCP (**O**).	Refined	*CMOC 17. A proper diagnosis* For a person with CLBP, receiving a diagnosis that clearly explains their ongoing pain (C) can trigger feelings of validation and understanding of their pain experience, both to themselves and others (M), which helps them accept their condition and increases the likelihood of engaging with a self-management app (O)	Participants: 1,5,6	*I think if someone had no real reason and no answers that would be a bit more different. Because, like, I've got the answers.I've accepted it* [their diagnosis] *so I think that's the difference for me.* [Participant 1]
CMOC 14. Trust When a person with CLBP trusts a HCP (**C**) then they are likely to be more receptive (**M**) to HCP's recommendation of a self-management app (**O**).	Refined	*CMOC 18. Trust the messenger* When a person with CLBP is recommended a self-management app by a trusted healthcare professional or someone they view as credible and supportive (C), this endorsement can reassure them of the app's credibility (M), making them more willing to give it a try (O	Participants: 1,3,4,5	*When you have been written off effectively by the NHS and effectively abandoned by them, then you're skeptical and maybe that's held me back from benefiting from the app potentially.* [Participant 5]
CMOC 15. Fear/Reassurance When a person has been reassured that there is no serious spinal pathology, and they are not likely to do any harm to their backs (**C**) they become less fearful of movement (**M**) and are more likely to engage with the strategies offered by a self-management app (**O**).	Refined	*CMOC 19. Reassurance* For people with CLBP who have high levels of anxiety about potential serious underlying causes of their pain or who hold strong beliefs that imaging is necessary for diagnosis (C), an assessment from an HCP alongside radiographic imaging (Intervention/Resource) is sought to provide trust and reassurance that nothing sinister has been missed (Mechanism), before they feel comfortable and confident to use an app to self-manage their CLBP (Outcome).	Participants: 1,3,5,7,8	*..an app can't actually see what's going on inside your body. So if there is something else causing something, you many need to go for that scan or Xray.I think there's a time and a place for it* [an app] *and probably later on down the path* [Participant 8]

XSL•FO
RenderX

Realist synthesis CMOCs[a]	After testing	Realist evaluation CMOCs	Sources	Example quote
CMOC 16. Early in the Journey Introducing a self-management app early on in a patient's journey —whilst medical management and investigations are ongoing—(**C**) can provide reassurance and advice to help a person with CLBP return to everyday activities (**M**) and thereby reduce the risk of maladaptive behaviors developing (**O**).	Refined	*CMOC 20. Timing of introduction* When a self-management app is introduced after medical investigations are underway or completed — a point at which individuals with CLBP feel their condition is being taken seriously (C) — they are more open to engaging with self-management strategies (M), allowing them to begin managing their condition earlier and improve their quality of life sooner (O).	Participants: 1,2,3,4,6,9	*I think the way I would personally think to benefit from it* [the app] *most would be if someone said 'okay, we're starting the treatment, we're going to do this and this.but your homework in the meantime is try and make a tiny bit of progress yourself with the help of this app.* [Participant 2]

[a]CMOC: context-mechanism-outcome configuration.

[b]NHS: National Health Service.

[c]CLBP: chronic low back pain.

[d]HCP: health care professional.

Results

A total of 9 participants, aged between 35 and 70 years and all living with CLBP for a minimum of 3 years participated in the study. A total of 3 were female and 6 were male. The 16 CMOCs from the preceding realist synthesis were tested and refined by the realist evaluation to produce 20 CMOCs (see Table 1).

These 20 CMOCs were refined, but did not fundamentally alter, the 3 program theories identified in the preceding realist review [27]. The groupings of the CMOCs and subsequent program theories were informed by Burden of Treatment theory [28] and Lee and Koh's conceptualization of empowerment [29]. The evaluation revealed deeper insights into contextual factors and causal mechanisms that explain who might benefit from a self-management app for CLBP, why, and in what circumstances.

Program Theory 1: Empowerment

Program theory 1 (see Textbox 2), informed by Lee and Koh's [29] model of empowerment, was developed from 7 CMOCs: choice and flexibility (CMOC1), knowledge is empowering (CMOC2), knowledge reduces fear (CMOC3), personalization (CMOC4), biomedical mindset (CMOC5), HCP buy-in is needed (CMOC6), and reduce HCP workload (CMOC7).

Textbox 2. Program theory 1: empowerment

People with chronic low back pain can feel empowered by a self-management app if the app is personal and relevant to their situation, can be accessed when and where they need it, and is presented as an adjunct to ongoing care.

Participants highlighted the importance of having choice and flexibility in how they accessed self-management support. While the app was designed to be convenient (CMOC1), many felt its rigid content delivery did not meet their situational needs. This refined the original concept of "convenience" by suggesting that flexibility in interaction style—not just availability—is key to fostering a sense of control and agency.

The app's educational content was widely seen as a means of empowerment, particularly when it provided new insights or helped fill perceived gaps in HCP knowledge (CMOC2). Beyond information acquisition, knowledge also had an important emotional impact in that it reduced the fear associated with chronic pain and helped participants shift from passive coping to active self-management (CMOC3).

However, this process was moderated by the app's capacity to personalize its content. Participants expressed frustration at the generic feel of the Curable app, noting that it did not sufficiently tailor responses to individual experiences (CMOC4). This was particularly problematic for users who expected content to reflect their specific pain history or emotional context.

Engagement with the app was also shaped by participants' underlying pain beliefs. Those with a strong biomedical orientation struggled with the app's psychological framing, interpreting it as invalidating their pain (CMOC5). By contrast, participants who had already begun to reconceptualize their pain found the app aligned with their thinking and were more likely to benefit.

A recurring theme across CMOCs was the continued importance of HCPs. Participants expressed a need for HCP buy-in and endorsement to validate the app's credibility and reinforce its content (CMOC6). In addition, the app was perceived as potentially useful in reducing demand on HCPs by enabling patients to self-manage more confidently between appointments (CMOC7). However, this potential was conditional on the app's responsiveness and the willingness of HCPs to engage with its outputs.

XSL•FO
RenderX

Together, these CMOCs refined Program Theory One by demonstrating that empowerment in self-management is not solely a function of delivering knowledge. It depends on how that knowledge is delivered, who it is validated by, and how well the system supports both user autonomy and relational trust.

Program Theory 2: Self-Management Burden

Program theory 2 (see Textbox 3), developed from 9 CMOCs and informed by May colleagues' [28] burden of treatment theory, explores how self-management apps can either alleviate or intensify the burden of managing chronic low back pain (CLBP). The theory was refined to highlight that this burden is shaped not only by external resources, such as digital access, but more critically by internal capacities, especially during periods of heightened pain.

Textbox 3. Program theory 2: self-management burden

If people with chronic low back pain (CLBP) have the capacity to engage with a mobile app, then it can reduce the burden of having to self-manage CLBP by providing ongoing support, facilitating communication with health care professionals, and mitigating feelings of abandonment.

During acute flares, participants described a significant drop in cognitive and emotional bandwidth, making any self-management activity—even app use—challenging. The evaluation introduced two new CMOCs to address this context: voice activation (CMOC8) and proactive support (CMOC9). Participants suggested that voice control could reduce interaction effort, while a more proactive app—one that nudges rather than waits for user input—could better meet their needs during these episodes of intense pain.

Beyond these flare-specific adaptations, several CMOCs addressed how the app's design intersected with routine management and interaction with HCPs. A recurring theme was the desire for data collection and sharing functionality, which participants believed would enhance the efficiency and focus of medical consultations. This supported and refined the review's theory that digital tools can reduce the frustration of updating HCPs during limited appointment time (consultations, CMOC10). Participants felt that pre-shared data would allow consultations to center on interpretation and planning rather than recounting.

This aligns with CMOC11, which focused on improving communication. The app was perceived as a potential aid in memory and articulation, allowing users to convey details that might be difficult to express during face-to-face consultations. Importantly, some participants valued objective data as a way to counteract perceived skepticism from HCPs when symptoms were not outwardly visible—reducing the pressure to "look sick."

However, not all interactions with data were beneficial. For some, symptom monitoring (CMOC12) backfired. Participants who had used other apps with frequent tracking prompts described heightened focus on pain, emotional distress, and even symptom worsening. For these users, digital tracking increased—rather than reduced—the burden of self-management.

Participants also discussed the app's role as a source of ongoing support (CMOC13), particularly following National Health Service (NHS) discharge. While Curable provided a sense of continuity, it was not viewed as a replacement for professional input. Many still desired periodic reassurance and oversight from HCPs, highlighting that an app alone was insufficient to maintain confidence or alleviate feelings of abandonment.

The app's conversational agent, Clara, further divided opinion. For some, the chatbot was supportive (CMOC14)—a nonjudgmental, low-effort interaction partner that offered comfort without the social labor of human connection. In contrast, others saw Clara as a poor proxy (CMOC15) for genuine empathy. The absence of emotional nuance, spontaneity, and body language left these participants feeling disconnected, reinforcing the view that digital tools can offer support but rarely replace relational care.

Collectively, these CMOCs refine Program theory two by emphasizing that the burden of self-management is not only shaped by how much a person must do, but also how sensitively digital tools respond to their physical, cognitive, and emotional capacity. Apps must strike a careful balance: offering timely, tailored, and low-effort support while avoiding features that may inadvertently intensify distress or isolation.

Programme Theory 3: Timing

Program theory 3 (see Textbox 4), informed by both burden of treatment theory [28] and the empowerment model, emphasizes that the timing of introducing a self-management app is a key factor in determining whether it is embraced by users. The theory is underpinned by five key CMOCs: expectations of a cure (CMOC 16), a proper diagnosis (CMOC 17), trust the messenger (CMOC 18), reassurance (CMOC 19), and the early stage in the journey (CMOC 20).

Participants reflected on how the NHS, historically centered around finding cures and fixing ailments, shaped patients' expectations. This cultural backdrop led to an assumption that health issues like CLBP should be resolved, rather than managed over the long term. For many patients, this expectation made it difficult to accept self-management solutions like mobile apps. However, participants noted that once individuals moved beyond the expectation of a quick fix, typically after receiving a thorough diagnosis and reassurance, they were more receptive to long-term management strategies, such as using a self-management app (CMOC 16).

XSL•FO
RenderX

Textbox 4. Programme theory 3: timing

A person with chronic low back pain is likely to benefit from a self-management app early on in their patient journey but not before they feel believed and reassured by health care professionals and have accepted their condition cannot be cured (CMOCs 16-20).

This shift in mindset was facilitated by key moments in the care journey. First, receiving a "proper" diagnosis—a concrete explanation of their pain—was seen as vital for helping patients accept their condition (CMOC 17). However, participants stressed that this diagnosis alone was not enough; they also needed reassurance from a HCP to ensure their pain was not indicative of something serious (CMOC 19). Only when these needs for validation and reassurance were met did participants feel ready to consider self-management tools, such as the app.

The introduction of the app was considered most effective when it occurred after these key steps had been completed. Timing was seen as especially critical—if the app was introduced too early, before a proper diagnosis or reassurance, it could be perceived as a "brush-off" or a superficial solution. But when the app was introduced at the right moment—typically after a thorough clinical assessment and reassurance—it was viewed as a supportive tool that could help users build self-management skills earlier, improving their quality of life sooner (CMOC 20).

Trust also played a significant role in whether participants engaged with the app. As highlighted in CMOC 18, recommendations from a trusted individual—often, though not always, a HCP—were more likely to lead to acceptance. However, when trust in HCPs was eroded due to unmet needs for validation or reassurance, some participants sought advice elsewhere, including from peers or alternative sources, who then became the preferred recommenders. This underlines the importance of introducing the app at the right moment, and through the right source, based on the patient's journey and their current relationship with healthcare professionals.

Together, these CMOCs refine Program theory three by emphasizing that successful engagement with self-management apps is contingent on timing—not simply early in the journey, but at the right point in a person's clinical and emotional trajectory. This includes when expectations of a cure have softened, when the individual feels their pain has been properly explained and validated, and when reassurance has been offered by a trusted source. The refined theory shifts away from a generic call for early intervention and instead highlights that the effectiveness of app-based self-management depends on readiness, which is shaped by how the health system meets core psychological and relational needs.

Discussion

Principal Findings

The aim of this realist evaluation was to test and refine 3 program theories developed during a preceding realist review on the use of mobile apps for the self-management of CLBP. These program theories centered around the concepts of empowerment, self-management burden, and timing, and were each underpinned by a set of CMOCs. They were informed by two key substantive theories: May and colleagues' [28] burden of treatment theory and Lee and Koh's [29] model of empowerment. Using the Curable app as a case study, this evaluation explored how, for whom, and under what circumstances such a digital self-management tool might work—or not—in everyday life. This study found that mobile apps like Curable can empower individuals with CLBP by providing accessible knowledge and reducing reliance on HCPs. However, the effectiveness of these tools is contingent on personalization, timing, and relational dynamics with health care providers. The success of such tools also depends on their introduction as an adjunct to ongoing care rather than a replacement, with careful consideration given to users' evolving trust and readiness for self-management.

Our evaluation showed that digital self-management tools like the Curable app can support empowerment in people with CLBP by providing accessible and credible knowledge that fosters confidence, agency, and a reduced reliance on HCPs—a finding that echoes Lee and Koh's [29] model of empowerment and aligns with Lim and colleagues' [52] work showing that people with CLBP want information to manage their condition. Knowledge provided via the app helped participants understand their pain and reduce fear, confirming the empowering potential of education in chronic pain management [25]. However, empowerment was not universally experienced. Participants with a strong biomedical mindset struggled to engage with the app's psychosocial framing, perceiving it as invalidating or insufficient—a reflection of the enduring influence of the biomedical model in shaping patient expectations [53]. This supports findings by Stenner et al [54] and Van de Velde et al [55], who argue that acceptance of pain is a prerequisite for engaging with self-management. These results highlight the need for apps to be personalized, both in content and delivery, to match users' readiness for behavioral change—a key mechanism of engagement reflected in the technology acceptance model [56,57] and the unified theory of acceptance and use of technology [58].

Participants also believed that the app could help them rely less on HCPs between appointments, particularly when they felt the information was relevant and trustworthy—suggesting a role for digital tools in extending continuity of care. However, when HCPs failed to acknowledge app-related efforts or patient-generated data, participants felt dismissed, leading to disempowerment. This reinforces Lee and Koh's [29] assertion that empowerment is relational and shaped by the behavior of those in power. For digital tools to fully realize their empowering potential, HCPs must validate and incorporate these tools into consultations. This aligns with existing research, which demonstrates that apps designed to track and share health data have improved consultations in the management of conditions like irritable bowel syndrome [59], chronic pain [60], and heart failure [61]. Taken together, these findings challenge the assumption that digital tools are inherently empowering and instead suggest their success is conditional on user beliefs, professional endorsement, and contextual fit. App developers should prioritize personalization and create tools that support

XSL•FO
RenderX

a staged, user-centered journey toward empowerment, while health systems must formally recognize and accommodate digital self-management tools within routine care.

The evaluation highlighted that while self-management apps can ease the burden of care, they also risk amplifying it if poorly designed or insufficiently integrated with broader health care support. Findings align with May and colleagues' [28] burden of treatment theory, which asserts that the capacity to engage with self-management depends on an individual's personal resources and the demands placed upon them. Participants emphasized that during acute flare-ups of CLBP, the cognitive and physical effort required to interact with apps like Curable became a significant barrier. This underscores the importance of designing features that minimize interactional demands during times of distress—such as voice activation and proactive content delivery—echoing calls in the literature for "low-friction" digital health tools [62]. Although data-sharing capabilities were absent in Curable, participants voiced a strong desire for features that allow symptom tracking and integration into consultations, viewing the proactive sharing of health information with health care providers before consultations as beneficial. This aligns with Holt and colleagues' [63] research, which demonstrated that previsit data collection enhances patient-provider communication, particularly in areas like respect, care, and perceived physician time, suggesting a pathway for better integrated health care support.

However, symptom tracking also revealed potential harms: for some, it became a persistent reminder of their condition, reinforcing pain salience [64]—a finding echoed in pain psychology literature cautioning against hypervigilance [64]. The app's chatbot, Clara, was also polarizing. While some participants found its nonjudgmental, low-demand interaction helpful—aligning with its intended purpose—others felt it lacked the emotional depth of human connection, reflecting early observations in the literature on the limitations of artificial intelligence–driven relational agents in fully replicating therapeutic rapport [65]. However, more recent research suggests that humans can form meaningful emotional bonds with chatbots [66], which can help alleviate feelings of social isolation and loneliness [67]. This prompted participants to consider the potential for improvement: they envisioned that with more sophisticated programming and algorithms enabling greater empathy and responsiveness, a chatbot like Clara could potentially foster a more supportive relationship. Despite these possibilities, however, the majority of participants felt that the app alone was not sufficient for ongoing support after discharge from the health care service. While some found Clara helpful for building confidence between appointments, they agreed that a follow-up with an HCP was ultimately necessary. This perspective mirrors findings from a large cross-sectional online survey, which revealed that most people prefer using health apps as a complement to, rather than a substitute for, in-person doctor visits [68]. These results highlight that, although digital self-management tools can extend care between appointments, they are generally seen as supplementary to, not a replacement for, the ongoing expertise of health care professionals.

Program theory 3 focused on the role of timing in shaping a person's readiness to engage with a self-management app for CLBP. While this theory was not explicitly framed around a single substantive model, it was informed by both the burden of treatment and empowerment frameworks. These concepts helped illuminate how the shifting emotional, cognitive, and relational context of living with CLBP influences when and how individuals feel capable of self-managing.

Our findings align with previous research, highlighting that patient engagement with self-management is more likely when individuals feel validated, reassured, and well-informed about their condition. For example, Toye et al [69] found that people with CLBP often struggle to engage with self-management until they have redefined their identity in relation to pain, a process that takes time and often requires credible explanation and validation from a trusted source. Similarly, Ong et al [70] found that people with sciatica were unable to emotionally adjust or cope with their symptoms until they received a credible diagnosis and explanation, which helped them make sense of their suffering. These findings echo the importance our participants placed on receiving a "proper" diagnosis and reassurance from an HCP before engaging with the app.

The concept of "timing" also reflects broader discussions in the literature around readiness for behavior change. Prochaska and DiClemente's [71] transtheoretical model of change suggests that interventions are more effective when matched to a person's stage of readiness—something that emerged clearly in our evaluation. Offering a self-management app too early, before patients feel they have been taken seriously or ruled out serious pathology, can create resistance rather than engagement.

Trust in the person recommending the app emerged as a key factor influencing engagement, aligning with Greenhalgh and colleagues' [72] finding that patients are more likely to adopt health technologies when introduced by someone perceived as credible and caring. However, our study suggests that the timing of such a recommendation is equally important. For users early in their journey, a trusted clinician may be an ideal introducer. But for others who have experienced unmet needs, disillusionment, or frustration in clinical encounters, the same recommendation may be disregarded or even resisted. In these cases, peer recommendations or alternative sources carried more weight. This highlights that the right source of recommendation must come at the right time, tailored to the user's current relationship with health care authority and trust.

While Program theory 3 does not rest on a distinct theoretical framework, its development was shaped by the same constructs underpinning the burden of treatment and empowerment theories. Specifically, it adds a temporal dimension—showing that empowerment and the capacity to shoulder treatment burdens are not static traits but unfold over time. Understanding when individuals are ready to self-manage is, therefore, essential in aligning support tools, like self-management apps, with patients' evolving needs and expectations.

Limitations

This study was guided by the RAMESES quality and reporting standards for realist evaluation [40] to ensure the research was undertaken with rigor and transparency. However, there are some limitations to be noted.

XSL•FO
RenderX

The Curable app did not have some of the functionality that was initially theorized to be beneficial, which meant some theories from the realist review could not be tested. However, to address this gap, participants used their experience with other health apps to help refine and develop theories.

Participation in the study was voluntary, which introduces a risk of selection bias. To address this, purposive sampling was used to ensure a variety of views toward self-management were represented. In addition, the study had a small number of participants and only evaluated the Curable app, which limits the claims that can be made about how representative its findings are for other self-management apps. However, the results have been presented as middle-range theories [30]. That is, the theories are at a level of abstraction whereby readers can judge whether they might be transferable to their context.

Conclusions

mHealth apps have the potential to help people with CLBP self-manage their condition. This is important considering the growing number of people affected by this condition and the likelihood that the numbers will increase as the world's population ages. This realist evaluation identified several key contextual factors and causal mechanisms to determine who may benefit from a self-management app and why. First, patient acceptance of their condition is important. Second, HCPs buy-in and personalized, adaptable content are essential for promoting sustained user engagement. Finally, to fully address patient needs, self-management apps should ideally be used in conjunction with ongoing support from HCPs. This integrated approach can help alleviate feelings of abandonment that may arise when solely relying on a mobile app.

Acknowledgments

We sincerely thank all participants who generously shared their experiences for this study. We are also grateful to Pain Association Scotland, Let's Get On With It Together, and Affa Sair for their valuable support with participant recruitment. Additionally, we acknowledge the use of ChatGPT (ChatGPT-4) for light text editing and proofreading during the preparation of this manuscript. We confirm that no original content, figures, or data were generated by this artificial intelligence tool, and we retain full responsibility for the paper's content, accuracy, and integrity.

Funding

RH's work was supported by the Inverness and Highland City Region Deal Studentship Award (2018).

Data Availability

The datasets generated or analyzed during this study are not publicly available due to privacy and ethical restrictions but are available from the corresponding author on reasonable request.

Authors' Contributions

RH, TG, and MB conceptualized the study and developed the methodology. RH conducted the investigation, including data curation and formal analysis, while TG and MB supervised the project and validated the findings. RH drafted the manuscript, and TG and MB critically reviewed and edited it.

Conflicts of Interest

None declared.

Multimedia Appendix 1
RAMESES II reporting standards for realist evaluations.
[DOC File, 55 KB - mhealth_v14i1e66435_app1.doc]

Multimedia Appendix 2
Realist evaluation interview topic guide.
[DOC File, 52 KB - mhealth_v14i1e66435_app2.doc]

References

1. Ferreira ML, de Luca K, Haile LM, et al. Global, regional, and national burden of low back pain, 1990–2020, its attributable risk factors, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. The Lancet Rheumatology 2023 Jun;5(6):e316-e329. [doi: 10.1016/S2665-9913(23)00098-X]
2. Meucci RD, Fassa AG, Faria NMX. Prevalence of chronic low back pain: systematic review. Rev Saude Publica 2015;49:1. [doi: 10.1590/S0034-8910.2015049005874] [Medline: 26487293]
3. Gore M, Sadosky A, Stacey BR, Tai KS, Leslie D. The burden of chronic low back pain: clinical comorbidities, treatment patterns, and health care costs in usual care settings. Spine (Phila Pa 1976) 2012 May 15;37(11):E668-E677. [doi: 10.1097/BRS.0b013e318241e5de] [Medline: 22146287]

XSL•FO
RenderX

4. Clare A, Andiappan M, MacNeil S, Bunton T, Jarrett S. Can a pain management programme approach reduce healthcare use? Stopping the revolving door. Br J Pain 2013 Aug;7(3):124-129. [doi: 10.1177/2049463713484907] [Medline: 26516513]
5. Hartvigsen J, Hancock MJ, Kongsted A, et al. What low back pain is and why we need to pay attention. The Lancet 2018 Jun;391(10137):2356-2367. [doi: 10.1016/S0140-6736(18)30480-X]
6. Mahmood A, Kedia S, Wyant DK, Ahn S, Bhuyan SS. Use of mobile health applications for health-promoting behavior among individuals with chronic medical conditions. Digit Health 2019;5:2055207619882181. [doi: 10.1177/2055207619882181] [Medline: 31656632]
7. Fan K, Zhao Y. Mobile health technology: a novel tool in chronic disease management. Intelligent Medicine 2022 Feb;2(1):41-47. [doi: 10.1016/j.imed.2021.06.003]
8. Amdie F, Woo K. The use of mHealth technology for chronic disease management the challenges and opportunities for practical application. Wounds Int 2020;11(2):32-38 [FREE Full text]
9. Maaß L, Freye M, Pan CC, Dassow HH, Niess J, Jahnel T. The definitions of health apps and medical apps from the perspective of public health and law: qualitative analysis of an interdisciplinary literature overview. JMIR Mhealth Uhealth 2022 Oct 31;10(10):e37980. [doi: 10.2196/37980] [Medline: 36315221]
10. World Health Organization. "mHealth: New Horizons for Health through Mobile Technologies 2011.
11. Digital technologies for managing non-specific low back pain: early value assessment. 2024. URL: https://www.nice.org.uk/guidance/hte16#:~:text=NICE%20has%20assessed%20early%20evidence,of%20clinical%20and%20patient%20experts [accessed 2025-05-20]
12. Almhdawi KA, Obeidat DS, Kanaan SF, Oteir AO, Mansour ZM, Alrabbaei H. Efficacy of an innovative smartphone application for office workers with chronic non-specific low back pain: a pilot randomized controlled trial. Clin Rehabil 2020 Oct;34(10):1282-1291. [doi: 10.1177/0269215520937757] [Medline: 32602362]
13. Selter A, Tsangouri C, Ali SB, et al. An mHealth app for self-management of chronic lower back pain (Limbr): pilot study. JMIR Mhealth Uhealth 2018 Sep 17;6(9):e179. [doi: 10.2196/mhealth.8256] [Medline: 30224333]
14. Irvine AB, Russell H, Manocchia M, et al. Mobile-web app to self-manage low back pain: randomized controlled trial. J Med Internet Res 2015 Jan 2;17(1):e1. [doi: 10.2196/jmir.3130] [Medline: 25565416]
15. Chhabra HS, Sharma S, Verma S. Smartphone app in self-management of chronic low back pain: a randomized controlled trial. Eur Spine J 2018 Nov;27(11):2862-2874. [doi: 10.1007/s00586-018-5788-5] [Medline: 30324496]
16. Rintala A, Rantalainen R, Kaksonen A, Luomajoki H, Kauranen K. mHealth apps for low back pain self-management: scoping review. JMIR Mhealth Uhealth 2022 Aug 26;10(8):e39682. [doi: 10.2196/39682] [Medline: 36018713]
17. de Melo Santana B, Raffin Moura J, Martins de Toledo A, et al. Efficacy of mHealth interventions for improving the pain and disability of individuals with chronic low back pain: systematic review and meta-analysis. JMIR Mhealth Uhealth 2023 Nov 2;11(1):e48204. [doi: 10.2196/48204] [Medline: 37962085]
18. Zhou T, Salman D, McGregor A. mHealth apps for the self-management of low back pain: systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2024 Feb 1;12:e53262. [doi: 10.2196/53262] [Medline: 38300700]
19. Craig P, Dieppe P, Macintyre S, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655. [doi: 10.1136/bmj.a1655] [Medline: 18824488]
20. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71. [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
21. Fox DM. Evidence and health policy: using and regulating systematic reviews. Am J Public Health 2017 Jan;107(1):88-92. [doi: 10.2105/AJPH.2016.303485] [Medline: 27854522]
22. Tricco AC, Cardoso R, Thomas SM, et al. Barriers and facilitators to uptake of systematic reviews by policy makers and health care managers: a scoping review. Implementation Sci 2015 Dec;11(1):1-20. [doi: 10.1186/s13012-016-0370-1]
23. Boaz A, Pawson R. The perilous road from evidence to policy: five journeys compared. J Soc Pol 2005 Apr;34(2):175-194. [doi: 10.1017/S0047279404008530]
24. Devan H, Farmery D, Peebles L, Grainger R. Evaluation of self-management support functions in apps for people with persistent pain: systematic review. JMIR Mhealth Uhealth 2019 Feb 12;7(2):e13080. [doi: 10.2196/13080] [Medline: 30747715]
25. Lorig KR, Holman HR. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]
26. Alperovitch-Najenson D, et al. WHO Guideline for Non-Surgical Management of Chronic Primary Low Back Pain in Adults in Primary and Community Care Settings: World Health Organization; 2023.
27. Hunter R, Beattie M, O'Malley C, Gorely T. Mobile apps to self-manage chronic low back pain: a realist synthesis exploring what works, for whom and in what circumstances. PEC Innov 2023 Dec 15;3:100175. [doi: 10.1016/j.pecinn.2023.100175] [Medline: 38213759]
28. May CR, Eton DT, Boehmer K, et al. Rethinking the patient: using Burden of Treatment Theory to understand the changing dynamics of illness. BMC Health Serv Res 2014 Jun 26;14:1-11. [doi: 10.1186/1472-6963-14-281] [Medline: 24969758]
29. Lee M, Koh J. Is empowerment really a new concept? The International Journal of Human Resource Management 2001 Jan;12(4):684-695. [doi: 10.1080/713769649]

XSL•FO
RenderX

30. Pawson R. Evidence-Based Policy: A Realist Perspecitve: Sage Publications Ltd; 2006.
31. Barlow JH, Sturt J, Hearnshaw H. Self-management interventions for people with chronic conditions in primary care: examples from arthritis, asthma and diabetes. Health Educ J 2002 Dec;61(4):365-378. [doi: 10.1177/001789690206100408]
32. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. JAMA 2002 Nov 20;288(19):2469-2475. [doi: 10.1001/jama.288.19.2469] [Medline: 12435261]
33. Wagner EH, Austin BT, Davis C, Hindmarsh M, Schaefer J, Bonomi A. Improving chronic illness care: translating evidence into action. Health Aff (Millwood) 2001;20(6):64-78. [doi: 10.1377/hlthaff.20.6.64] [Medline: 11816692]
34. Hunter R, Gorely T, Beattie M, Harris K. Realist review. Int Rev Sport Exerc Psychol 2022 Dec 31;15(1):242-265. [doi: 10.1080/1750984X.2021.1969674]
35. Punton M, Vogel I, Lloyd R. Reflections from a realist evaluation in progress: scaling ladders and stitching theory. 2016 URL: https://www.ids.ac.uk/publications/reflections-from-a-realist-evaluation-in-progress-scaling-ladders-and-stitching-theory/ [accessed 2026-01-30] [Medline: 26838657]
36. Wong G, Westhorp G, Greenhalgh J, Manzano A, Jagosh J, Greenhalgh T. Quality and reporting standards, resources, training materials and information for realist evaluation: the RAMESES II project. In: Southampton (UK): NIHR Journals Library; 2017 Oct 2017, Vol. 5:1-108. [doi: 10.3310/hsdr05280] [Medline: 29072890]
37. Jagosh J, Pluye P, Wong G, et al. Critical reflections on realist review: insights from customizing the methodology to the needs of participatory research assessment. Res Synth Methods 2014 Jun;5(2):131-141. [doi: 10.1002/jrsm.1099] [Medline: 26052652]
38. Pawson R, Manzano-Santaella A. A realist diagnostic workshop. Evaluation (Lond) 2012 Apr;18(2):176-191. [doi: 10.1177/1356389012440912]
39. Paparini S, Papoutsi C, Murdoch J, et al. Evaluating complex interventions in context: systematic, meta-narrative review of case study approaches. BMC Med Res Methodol 2021 Oct 25;21(1):225. [doi: 10.1186/s12874-021-01418-3] [Medline: 34689742]
40. Wong G, Westhorp G, Manzano A, Greenhalgh J, Jagosh J, Greenhalgh T. RAMESES II reporting standards for realist evaluations. BMC Med 2016 Jun 24;14(1):96. [doi: 10.1186/s12916-016-0643-1] [Medline: 27342217]
41. Curable Inc. Curable. URL: https://www.curablehealth.com/ [accessed 2024-5-22]
42. Wong G. Data Gathering in Realist Reviews: Looking for Needles in Haystacks," Doing Realist Research: SAGE; 2018:131-145. [doi: 10.4135/9781526451729]
43. Emmel N. Sampling and Choosing Cases in Qualitative Research: A Realist Approach 2013.
44. Manzano A. The craft of interviewing in realist evaluation. Evaluation (Lond) 2016 Jul;22(3):342-360. [doi: 10.1177/1356389016638615]
45. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
46. Mukumbang FC, Marchal B, Van Belle S, van Wyk B. Using the realist interview approach to maintain theoretical awareness in realist studies. Qual Res 2020 Aug;20(4):485-515. [doi: 10.1177/1468794119881985]
47. Mukumbang FC, Kabongo EM, Eastwood JG. Examining the application of retroductive theorizing in realist-informed studies. Int J Qual Methods 2021 Jan;20:16094069211053516. [doi: 10.1177/16094069211053516]
48. Jagosh J. Retroductive theorizing in Pawson and Tilley's applied scientific realism. J Crit Realism 2020 Mar 14;19(2):121-130. [doi: 10.1080/14767430.2020.1723301]
49. Timmermans S, Tavory I. Theory construction in qualitative research: from grounded theory to abductive analysis. Sociol Theory 2012;30(3):167-186. [doi: 10.1177/0735275112457914]
50. Ravitch SM, Riggan M. Reason & Rigor: How Conceptual Frameworks Guide Research, 2nd edition: Sage Publications; 2016.
51. Thagard P. Analogy, explanation, and education. J Res Sci Teach 1992 Aug;29(6):537-544. [doi: 10.1002/tea.3660290603]
52. Lim YZ, Chou L, Au RT, et al. People with low back pain want clear, consistent and personalised information on prognosis, treatment options and self-management strategies: a systematic review. J Physiother 2019 Jul;65(3):124-135. [doi: 10.1016/j.jphys.2019.05.010] [Medline: 31227280]
53. Engel GL. The need for a new medical model: a challenge for biomedicine. Science 1977 Apr 8;196(4286):129-136. [doi: 10.1126/science.847460] [Medline: 847460]
54. Stenner P, Cross V, McCrum C, et al. Self-management of chronic low back pain: four viewpoints from patients and healthcare providers. Health Psychol Open 2015 Jul;2(2):2055102915615337. [doi: 10.1177/2055102915615337] [Medline: 28070378]
55. van der Velde M, van der Leeden M, Geleijn E, Veenhof C, Valkenet K. What moves patients to participate in prehabilitation before major surgery? A mixed methods systematic review. Int J Behav Nutr Phys Act 2023 Jun 21;20(1):75. [doi: 10.1186/s12966-023-01474-6] [Medline: 37344902]
56. Davis FD. Technology acceptance model: TAM. In: Al-Suqri MN, Al-Aufi AS, editors. Information Seeking Behavior and Technology Adoption: IGI Global; 1989, Vol. 205:205-219.

XSL•FO
RenderX

57. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep 1;13(3):319-340. [doi: 10.2307/249008]
58. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003 Sep 1;27(3):425-478. [doi: 10.2307/30036540]
59. Yin AL, Hachuel D, Pollak JP, Scherl EJ, Estrin D. Digital health apps in the clinical care of inflammatory bowel disease: scoping review. J Med Internet Res 2019 Aug 19;21(8):e14630. [doi: 10.2196/14630] [Medline: 31429410]
60. Whitney RL, Ward DH, Marois MT, Schmid CH, Sim I, Kravitz RL. Patient perceptions of their own data in mhealth technology-enabled N-of-1 trials for chronic pain: qualitative study. JMIR Mhealth Uhealth 2018 Oct 11;6(10):e10291. [doi: 10.2196/10291] [Medline: 30309834]
61. Bezerra Giordan L, Tong HL, Atherton JJ, et al. The use of mobile apps for heart failure self-management: systematic review of experimental and qualitative studies. JMIR Cardio 2022 Mar 31;6(1):e33839. [doi: 10.2196/33839] [Medline: 35357311]
62. Cho J, Park D, Lee HE. Cognitive factors of using health apps: systematic analysis of relationships among health consciousness, health information orientation, eHealth literacy, and health app use efficacy. J Med Internet Res 2014 May 9;16(5):e125. [doi: 10.2196/jmir.3283] [Medline: 24824062]
63. Holt JM, Cusatis R, Winn A, et al. Impact of pre-visit contextual data collection on patient-physician communication and patient activation: a randomized trial. J Gen Intern Med 2021 Nov;36(11):3321-3329. [doi: 10.1007/s11606-020-06583-7] [Medline: 33559067]
64. Eccleston C, Crombez G. Pain demands attention: a cognitive-affective model of the interruptive function of pain. Psychol Bull 1999 May;125(3):356-366. [doi: 10.1037/0033-2909.125.3.356] [Medline: 10349356]
65. Bickmore T, Gruber A. Relational agents in clinical psychiatry. Harv Rev Psychiatry 2010;18(2):119-130. [doi: 10.3109/10673221003707538] [Medline: 20235777]
66. Li J, Zhang C, Li X, Zhang C. Patients' emotional bonding with MHealth apps: an attachment perspective on patients' use of MHealth applications. Int J Inf Manage 2020 Apr;51:102054. [doi: 10.1016/j.ijinfomgt.2019.102054]
67. Dosovitsky G, Bunge EL. Bonding with bot: user feedback on a chatbot for social isolation. Front Digit Health 2021;3:735053. [doi: 10.3389/fdgth.2021.735053] [Medline: 34713203]
68. Rai A, Chen L, Pye J, Baird A. Understanding determinants of consumer mobile health usage intentions, assimilation, and channel preferences. J Med Internet Res 2013 Aug 2;15(8):e149. [doi: 10.2196/jmir.2635] [Medline: 23912839]
69. Toye F, Seers K, Allcock N, et al. A meta-ethnography of patients' experience of chronic non-malignant musculoskeletal pain. Osteoarthr Cartil 2013 Apr;21:S259-S260. [doi: 10.1016/j.joca.2013.02.539]
70. Ong BN, Konstantinou K, Corbett M, Hay E. Patients' own accounts of sciatica: a qualitative study. Spine (Phila Pa 1986) 2011;36(15):1251-1256. [doi: 10.1097/BRS.0b013e318204f7a2]
71. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 1983 Jun;51(3):390-395. [doi: 10.1037//0022-006x.51.3.390] [Medline: 6863699]
72. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367. [doi: 10.2196/jmir.8775] [Medline: 29092808]

Abbreviations

CARES: Center for Advancement in Realist Evaluation and Synthesis
CLBP: chronic low back pain
CMOC: context-mechanism-outcome configuration
HCP: health care professional
mHealth: mobile health
MRC: Medical Research Council
NHS: National Health Service
PPI: Patient and Public Involvement
RAMESES II: Realist And Meta-narrative Evidence Syntheses: Evolving Standards II

Edited by L Buis; submitted 12.Sep.2024; peer-reviewed by B Gilmore, M Stochkendahl; revised version received 20.May.2025; accepted 14.Oct.2025; published 17.Feb.2026.

Please cite as:
Hunter R, Gorely T, Beattie M
Mobile App–Supported Self-Management for Chronic Low Back Pain: Realist Evaluation
JMIR Mhealth Uhealth 2026;14:e66435
URL: https://mhealth.jmir.org/2026/1/e66435
doi:10.2196/66435

XSL•FO
RenderX

© Rebecca Hunter, Trish Gorely, Michelle Beattie. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 17.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Identifying and Evaluating Patient-Centered Mobile and Web Apps for Patients With Chronic Spontaneous Urticaria: Systematic Search and Content Analysis

Caroline Glatzel, MD; Bernadette Glatzel, MD; Tassilo Dege, MHA, MD; Lena Brückel, MD; Matthias Goebeler, MD; Astrid Schmieder, MD

Department of Dermatology, Venereology and Allergology, University Hospital Würzburg, Josef-Schneider-Straße 2, Würzburg, Germany

Corresponding Author:
Caroline Glatzel, MD
Department of Dermatology, Venereology and Allergology, University Hospital Würzburg, Josef-Schneider-Straße 2, Würzburg, Germany

Abstract

Background: Chronic spontaneous urticaria (CSU) is characterized by recurrent wheals or angioedema lasting for more than 6 weeks and substantially affecting the quality of life. Given its fluctuating course, accurate symptom monitoring is essential. Mobile health apps (MHAs) offer promising tools for real-time symptom tracking, patient education, and communication. Systematic evaluation of existing MHAs for CSU is critical to inform the development of effective, patient-centered digital solutions.

Objective: This study aimed to identify and evaluate publicly available MHAs for patients with CSU, assessing their quality, usability, and alignment with the needs of both patients and physicians to guide the development of future patient-centered apps.

Methods: A systematic search of app stores and the internet was conducted to identify MHAs for CSU. Inclusion required German or English language support and patient-centered content. Apps were excluded if they contained advertisements; lacked patient-centered content, designed to assist patients in the self-management and care of their condition; or were focused on clinical trials or health care professional use. After screening, 1 app, CRUSE Control, met all criteria and was evaluated by 23 physicians and 16 patients with CSU using the German versions of the Mobile Application Rating Scale (MARS and end-user version of MARS) and the mHealth App Usability Questionnaire. Participants' technical affinity was assessed using the affinity for technology interaction scale and the Mobile Device Proficiency Questionnaire. Additionally, they completed a custom questionnaire on their personal needs and expectations for CSU-specific MHAs.

Results: Fifteen MHAs were identified, with 12 available on both platforms. Eleven apps were excluded due to lack of specificity to CSU (n=10) or not being patient-centered (n=1). One app, CRUSE Control, met all inclusion criteria and was selected for final evaluation. CRUSE Control received similar mean (SD) quality ratings from physicians (MARS 4.03, SD 0.45) and patients (end-user version of MARS: 4.06, SD 0.40; *P*=.83). Among the MARS subcategories, functionality was rated significantly higher by patients than by physicians (4.75, SD 0.41 vs 4.47, SD 0.55; *P*=.04). Usability, measured using the German mHealth App Usability Questionnaire (assessing effectiveness, efficiency, and satisfaction), showed no significant difference between physicians (5.85, SD 0.71) and patients (5.76, SD 0.41; *P*=.64). Technology affinity was comparable between groups, with physicians scoring 3.50 (SD 0.66) and patients 4.00 (SD 0.88) on the affinity for technology interaction (*P*=.05). Proficiency with mobile devices, assessed via the Mobile Device Proficiency Questionnaire, also showed similar results (physicians: 4.81, SD 0.26; patients: 4.74, SD 0.45; *P*=.60).

Conclusions: Few high-quality MHAs for CSU are currently available, and only 1 met the inclusion criteria. Patients and physicians rated the app highly, though patients placed greater emphasis on functionality. High technology affinity in both groups supports adoption. Patients prioritized features that facilitate disease management. Although limited to a single app, these findings suggest that MHAs may support CSU care.

(JMIR Mhealth Uhealth 2026;14:e71435) doi:10.2196/71435

KEYWORDS

urticaria; mobile health apps; MHA; mobile apps; Mobile Application Rating Scale; MARS; user version of MARS; uMARS; German mHealth App Usability Questionnaire; G-MAUQ; Mobile Device Proficiency Questionnaire; MDPQ-16; affinity for technology interaction; ATI; skin

XSL•FO
RenderX

Introduction

Background

Urticaria is a common skin disorder characterized by the sudden onset of wheals and/or angioedema. Wheals (urticaria) typically presents with three key features: (1) sharply demarcated, superficial swellings of the skin of variable size and shape, usually surrounded by a reflex erythema; (2) intense pruritus or, less commonly, a burning sensation; and (3) a transient course, with lesions usually resolving within 30 minutes to 24 hours.

In around 40% of cases, episodes of urticaria are accompanied by angioedema, primarily affecting the face, but also the hands and feet [1]. Angioedema is defined by (1) rapid-onset, pronounced swelling in the deeper dermis, subcutis, or mucous membranes, often erythematous or skin-colored; (2) a sensation of tingling, burning, or tension and sometimes pain rather than itch; and (3) a slower resolution, typically lasting 72 hours. According to current guidelines, urticaria is classified by duration as either acute (≤6 wk) or chronic (>6 wk) and by trigger as spontaneous (no identifiable cause) or inducible (triggered by specific stimuli) [2,3].

The lifetime prevalence of urticaria is estimated to range between 10% to 20%, indicating that a substantial proportion of the population experiences at least one episode during their lifetime [2,4]. Chronic urticaria (CU) has a lifetime prevalence of approximately 1% to 2% among adults in Europe, the United States, and worldwide, with chronic spontaneous urticaria (CSU) being the most common subtype [4-6]. It most frequently manifests between the third and fifth decades of life [5,7,8] and affects women twice as often as men [8].

The impact of CSU on patients and their social environment is substantial and far-reaching [9]. Quality of life is markedly impaired, comparable to other severe dermatological conditions such as hidradenitis suppurativa or advanced atopic dermatitis [10]. The unpredictability of symptoms—such as itching, swelling, and wheals, which can lead to sleep disorders—contributes to emotional distress, particularly anxiety and depression [11-14].

CSU is primarily a clinical diagnosis, typically first assessed by general practitioners. A comprehensive diagnostic evaluation includes a detailed medical history, structured around the "7 C's" (Confirm, Cause, Cofactors, Comorbidities, Consequences, Components, and Course), as well as the assessment of disease activity, impact on patient's life, and disease management [2,15]. CU may also be triggered by circulating immune complexes and activation of the complement system. In cases persisting for several months or longer, autoimmune thyroiditis should be ruled out [16,17]. Standardized tools such as the urticaria activity score (UAS), Urticaria Control Test (UCT), and the Chronic Urticaria Quality of Life Questionnaire are recommended for quantifying disease burden and monitoring fluctuations over time [2,18].

Due to the episodic nature of CSU, which may persist for months or even years [19], symptom fluctuation is common and can make it difficult for health care providers to fully assess the disease's burden.

To bridge this gap, mobile technologies have gained traction in dermatology. Clinical photography has been used for documenting disease progression and supporting diagnostic accuracy [20], and messaging tools and teleconsultation already support remote assessment and communication [21]. This offers additional benefits in improving treatment monitoring and enhancing patient education [22]. A recent study has demonstrated the feasibility and high acceptance of teledermatologic disease management in a cohort of patients with CSU [23]. Building on these existing digital tools, mobile health apps (MHAs) offer promising opportunities for the management of CSU. As the demand for accessible and flexible health care grows, they are becoming increasingly relevant in disease management, enabling real-time symptom documentation, remote monitoring, and timely interventions [24-27]. Their portability, good availability, and widespread use make MHAs particularly valuable for managing health issues, as reflected in the growing digital health market. By the third quarter of 2020, the Apple App Store offered 48,608 medical mHealth apps (up from 44,384 in Q3 2019) and a total of 82,633 health and fitness apps [28].

In Germany, MHAs can be certified as digital health apps (DiGAs), making them eligible for prescription and reimbursement through public health insurance since 2019 [29,30]. As part of the Digital Healthcare Act—the first law worldwide to integrate digital therapeutics into public care—DiGAs support patient self-management, improve therapy adherence, and facilitate communication by enabling the sharing of patient-generated data [29].

Challenges such as reimbursement issues, digital literacy, and variable acceptance among health care providers and patients limit the use of DiGAs, and currently, no certified DiGA or high-quality app exists specifically for CSU [31,32]. Given the clinical need and patient demand, identifying and evaluating existing MHAs for CSU is essential to guide the development of effective, user-centered digital solutions.

Objective

The aim of this study was to identify MHAs currently available for patients with CSU, evaluate their quality and usability, and assess the technological preferences and needs of both patients and physicians. Based on these insights, recommendations for the development of future patient-centered CSU apps were derived.

Methods

Ethical Considerations

This analysis was approved by the Ethics Committee of the University of Würzburg (reference: 20230918 02) in accordance with the ethical standards established by the Declaration of Helsinki. All participants gave informed consent. Data was fully anonymized; no personal information was collected. No compensation was provided to the participants.

XSL•FO
RenderX

The reporting of this study followed established principles for transparent and structured reporting of research: STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines (Checklist 2).

App Screening

Between January 1 and March 31, 2024, 2 independent dermatologists with expertise in the management of urticaria conducted a systematic search of the Apple App Store, Google Play Store, and the internet using search engines (eg, Google and Safari) for mobile apps designed for patients with CU. Search terms used were "urticaria," "spontaneous urticaria," "hives," "itching," "angioedema," "wheal," and "swelling." Inclusion criteria were defined as (1) availability on at least 1 of the specified platforms (Apple App Store, Google Play Store, or internet-based) and (2) language support in either German or English. Exclusion criteria comprised the (1) presence of advertisements, (2) a lack of patient-centered content—meaning content that primarily focuses on supporting the care and self-management of patients, or (3) a primary focus on clinical trials or scientific conferences. Apps containing advertisements were excluded in order to avoid potential disruptions to user experience and to control for confounding effects on usability [33]. Each app was assessed individually and independently. Discrepancies between reviewers were resolved through discussion. No automation tools were used during the screening process.

The primary outcomes for which data were sought included app functionality (eg, symptom tracking, disease activity scores, and medication reminders), user engagement features (eg, personalization and notifications), and data protection or privacy measures. All available features and descriptions provided in app store listings or within the apps themselves were reviewed. All data reflected the app status at the time of assessment.

For each app, the following variables were collected, if available: app name, platform availability (Apple App Store, Google Play Store, Google, and Safari), search term used, app topic (primary focus or intended purpose), total number of hits per search term, supported languages (German and English), app store rating and number of user ratings, target user group (patients and health care professionals), cost (free or paid), presence of advertisements, date of access, and inclusion of specific features such as symptom scores or disease monitoring tools.

A total of 15 apps were identified. After applying the inclusion and exclusion criteria, only one app, the CRUSE Control app, met the required criteria. The app was available on both Apple App Store and Google Play Store and supported German and English languages.

This study focuses on comparing app quality and usability between patients and physicians, rather than providing absolute ratings of individual apps.

Raters' Characteristics

Patients were randomly selected from our outpatient clinic to reduce selection bias and obtain a diverse user perspective. Similarly, physicians were randomly selected among those working in the same outpatient clinics and regularly treating patients with CSU. As the primary aim of the study is to assess app quality and usability, detailed information on participants' social background, educational level, and clinical characteristics (eg, comorbidities or urticaria subtype) was not collected. This decision was made to minimize participant burden and maintain focus on the usability aspects of the app within the scope of this exploratory investigation.

Patients and physicians tested the CRUSE Control app version corresponding to the operating system available on their personal smartphones (iOS or Android). If they did not own a mobile device, they used a provided iPhone.

Evaluation of App Quality

The Mobile Application Rating Scale (MARS) is a widely used tool for evaluating the quality of mobile apps [34]. It consists of 19 items across four categories: engagement, functionality, aesthetics, and information. Each item is rated on a 5-point Likert scale, with higher scores indicating higher quality. The overall MARS score is calculated as the mean value of these four subscale scores. In addition, there is a separate "subjective" section to capture the user's personal impression of the app.

For this study, the validated German version of Mobile Application Rating Scale (MARS-G) [35] and the end-user version of Mobile Application Rating Scale (uMARS) were applied to assess the quality of MHAs. The uMARS represents a simplified version of the MARS designed for end users and has been shown to reliably assess the quality of MHAs [17]. The MARS-G includes an additional section, which evaluates therapeutic quality and safety across 4 dimensions: patient benefit, therapist benefit, potential risks, and transferability to routine care. It is referred to as the Psychotherapy Section in the questionnaire provided by Messner et al [35].

Although the German translation of the uMARS used in this study has not been formally published, its items are identical to those of the validated MARS-G. A comparative analysis between physicians' and patients' MARS and uMARS was conducted.

For clarity, the term "MARS" or "uMARS" is used throughout the manuscript to refer to this German version.

Evaluation of App Usability

The G-MAUQ (German mHealth App Usability Questionnaire), the German validated version of the MAUQ, is a usability assessment tool for MHAs [36]. With 18 questions in 3 subcategories, it assesses the user's perspective of the app's ease of use, interface, and overall satisfaction and usefulness.

A total of 23 physicians assessed the app, appropriate for patients with CSU, using MARS-G to evaluate quality and G-MAUQ to assess usability. Simultaneously, 16 patients with CSU rated the same app, applying the uMARS for quality assessment and the G-MAUQ for usability evaluation. Prior to completing the evaluations, all participants were instructed to test each app for at least 10 minutes. A comparative analysis between physicians' and patients' G-MAUQ was conducted.

XSL•FO
RenderX

Evaluation of Technical Affinity

The physicians' and patients' technology affinity was assessed using the affinity for technology interaction (ATI) scale and the Mobile Device Proficiency Questionnaire (MDPQ-16). The ATI is a well-established and widely used measure of an individual's general affinity for technology consisting of 9 questions [37]. The 16-question MDPQ-16 is also used to assess technology affinity specifically related to the use of mobile devices [38].

Spearman rank correlation coefficient and Pearson correlation coefficient were used to correlate the ATI scale and MDPQ-16 score with ages of physicians and patients and the MDPQ-16. A comparative analysis between physicians' and patients' ATI and MDPQ-16 was conducted.

Statistical Analysis

The calculations were performed using SPSS, version 23 (IBM Deutschland GmbH). Following the testing of the data for normal distribution with Shapiro-Wilk, either the Mann-Whitney *U* or *t* test was used. *P* values less than .05 were considered statistically significant. Correlation analysis was conducted using both the Spearman and Pearson correlation coefficients.

Evaluation of Patients' Need for a Patient App

Both physicians and patients completed a questionnaire specifically designed for this analysis. This questionnaire included open-ended questions regarding MHAs and their needs and expectations regarding such apps.

Results

App Screening and Selection

A total of 15 urticaria-related apps were identified: 14 in the Apple App Store and 13 in the Google Play Store. Of these, 12 were identified in both App Stores, and 11 of these apps were excluded for the following reasons: not specific to CSU (n=10) and not patient-centered (n=1; Figure 1).

Figure 1. App selection.



Only 1 currently available patient-centered urticaria app was identified: the CRUSE Control Urticaria app, a CU self-evaluation app available in both Apple App Store and Google Play Store. It was developed by Urticaria Centers of Reference and Excellence and is available in English and German. We tested version 1.0.69, last updated in November 2023.

The app offers patients with CU the opportunity to receive support in managing their condition. It provides a platform for daily documentation of symptoms and impact on quality of life within a few minutes. By sharing data with the treating physician, the efficacy of the current therapy and treatment can be optimized.

The app comprises 6 principal sections, each of which provides distinct functionality to assist patients with CU. In the "Daily Documentation" section, patients are required to evaluate the impact of urticaria over the preceding 24 hours. The questions are based on established scales, including the UAS and the Angioedema Activity Score. Additionally, patients are queried on a monthly basis regarding disease management, with questions based on the UCT and the Angioedema Control Test. The impact on the patient's quality of life is assessed using the Chronic Urticaria Quality of Life Questionnaire and the Angioedema Quality of Life Questionnaire. Furthermore, daily medication intake and dosage are recorded.

The "Results Review" section enables users to monitor their daily progress in relation to the UAS and Angioedema Activity Scores, as well as their long-term disease management via the UCT, Angioedema Control Test, and quality of life indicators. In the "Profile View" section, patients may view and update their personal information. The "Preparation for Doctor's Visit" section allows patients to prepare relevant information and questions for their upcoming appointment. Furthermore, patients

XSL•FO
RenderX

have the option of uploading images of their urticaria in this section. A reminder function is provided to ensure that important appointments and medication dosages are not overlooked. Ultimately, the "Information about Urticaria" section offers valuable, patient-oriented material to facilitate comprehension of the condition. The link to the "Hello Professor" function provides valuable medical information that expands patients' knowledge and increases their confidence in dealing with the disease.

The data collected can be summarized in a report for the physician, which may be transmitted via email or shared through a QR code. The integration of these elements facilitates uninterrupted and comprehensive observation of the condition, while also enhancing communication between patients and health care professionals (Table 1).

Table . Overview of patient-centered features in the CRUSE Control app.

Section/feature	Description/functionality	Notes/relevance to patient care
"Daily Documentation"	Daily symptom tracking (UAS[a], AAS[b]), monthly disease management (UCT[c], AECT[d]), quality of life (CU-Q2oL[e], AE-QoL[f]), medication intake	Enables monitoring of disease activity and therapy efficacy
"Results review"	Visualization of progress and long-term disease control	Supports self-management and gives insight into therapy outcomes
"Profile View"	View and update personal information	Personalization of patient data
"Preparation for Doctor's visit"	Upload images, prepare questions, reminders for appointments and medication	Enhances communication with physician and appointment readiness
"Information about Urticaria"	Educational content about the condition, "Hello Professor" cross-reference	Improves patient understanding and confidence
"Data Sharing"	Generate report for physician via email or QR code	Facilitates continuous monitoring and physician feedback

[a]UAS: urticaria activity score.

[b]AAS: angioedema activity score.

[c]UCT: urticaria control test.

[d]AECT: angioedema control test.

[e]CU-Q2oL: Chronic Urticaria Quality of Life Questionnaire.

[f]AE-QoL: Angioedema Quality of Life Questionnaire.

Raters' Characteristics

The app was evaluated by 23 dermatologists and 16 patients with CSU. Mean age was 34.0 (SD 9.7) years in physicians and 40.0 (SD 13.4) years in patients. The physicians' group was comprised of 6 men and 17 women, while the patients' group included 6 men and 10 women. All patients experienced multiple episodes of urticaria per month, lasting at least 12 weeks in total. Additionally, they reported sleep disorders and anxiety related to their condition.

Evaluation of App Quality, Including Comparative Analysis of Patients' and Physicians' Data

Only the CRUSE Control app met the inclusion criteria and was included in this analysis. The results of the MARS and uMARS scores for the app are shown in Tables 2 and 3.

Table . MARS[a] score of the CRUSE Control app assessed by 23 physicians. Maximum MARS score: 5 points. In addition to the validated German version of Mobile Application Rating Scale core sections, the table includes the psychotherapy section—an extension in the German MARS that evaluates therapeutic quality across 4 dimensions (patient benefit, therapist benefit, risks, and transferability to routine care) as described by Messner et al [35].

	CRUSE Control app, mean (SD)
MARS score	4.03 (0.45)
Engagement score	3.67 (0.71)
Functionality score	4.47 (0.55)
Esthetics score	4.10 (0.56)
Information score	3.86 (0.49)
Subjective quality	3.49 (0.42)
Psychotherapy score	3.83 (0.61)

[a]MARS: Mobile Application Rating Scale.

XSL•FO
RenderX

Table . uMARS[a] score of the CRUSE Control app assessed by 16 patients with urticaria. Maximum uMARS score: 5 points.

	CRUSE Control app, mean (SD)
uMARS score	4.06 (0.40)
Engagement score	3.55 (0.62)
Functionality score	4.75 (0.41)
Esthetics score	4.10 (0.45)
Information score	3.82 (0.81)
Subjective quality	3.35 (0.60)

[a]uMARS: end-user version of Mobile Application Rating Scale.

The quality of the CRUSE Control app was evaluated by both physicians and patients using the MARS or its user version (uMARS). Physicians and patients provided similar ratings, with an average score of 4.03 (SD 0.45) out of 5 for physicians and 4.06 (SD 0.40) for patients (*P*=.83), suggesting an overall high quality of the app. In most subcategories, including engagement (*P*=.58), esthetics (*P*=.97), information (*P*=.86), and subjective quality (*P*=.64), the ratings were closely aligned between the 2 groups. However, patients rated the functionality of the app significantly higher than physicians, with an average score of 4.75 (SD 0.41) compared to 4.47 (SD 0.55; *P*=.04). The subcategory measuring "Psychotherapy" was only included in the MARS for physicians. This section addresses the quality criteria for apps with respect to patient safety, the use of the app's presentation, and the quality of the therapeutic offerings. They scored it with an average of 3.83 (SD 0.61).

Evaluation of App Usability, Including Comparative Analysis of Patients' and Physicians' Data

The usability of the CRUSE Control app was assessed similarly by both physicians and patients. Using the G-MAUQ to measure usability, physicians gave an average score of 5.85 (SD 0.71) out of 7, while patients rated it slightly lower with an average score of 5.76 (SD 0.41) out of 7. No significant difference was found between the 2 groups (*P*=.64), as shown in Table 4. In the subcategories of G-MAUQ, both physicians and patients rated "Ease of Use" (6.27, SD 0.68 vs 6.45, SD 0.57), "Interface and Satisfaction" (6.04, SD 0.68 vs 6.00, SD 0.52), and "Usefulness" (5.30, SD 0.98 vs 4.91, SD 0.87) similarly. However, when looking at the absolute scores, patients rated "Usefulness" 0.39 points lower than physicians, though this difference was not statistically significant (*P*=.20).

Table . Evaluation of the CRUSE Control app with G-MAUQ[a] and G-MAUQ subscales by physicians and patients. Maximum G-MAUQ score: 7 points.

	Physicians (n=23), mean (SD)	Patients (n=16), mean (SD)	*P* value
G-MAUQ	5.85 (0.71)	5.76 (0.41)	.68
Ease of use	6.27 (0.68)	6.45 (0.57)	.51
Interface and satisfaction	6.04 (0.68)	6.00 (0.52)	.83
Usefulness	5.30 (0.98)	4.91 (0.87)	.20

[a]G-MAUQ: German mHealth App Usability Questionnaire.

Evaluation of Technical Affinity, Including Comparative Analysis of Patients' and Physicians' Data

The ATI score and the MDPQ-16 score, which evaluate the users' affinity for technology, were recorded for both patients and physicians.

Physicians achieved a mean ATI score of 3.50 (SD 0.66) out of 6 points and a MDPQ-16 score of 4.81 (SD 0.26) out of 5 points. Patients, on the other hand, had a mean ATI score of 4.00 (SD 0.88) and a MDPQ-16 score of 4.74 (SD 0.45; *P*=.05). Although there were no statistically significant differences between the 2 groups, both groups considered themselves highly experienced in mobile device usage, as indicated by their similar MDPQ-16 scores (physicians: 4.81, SD 0.26; patients: 4.74, SD 0.45; *P*=.60), as shown in Table 5.

Table . Evaluation of technical affinity using the ATI[a] scale and short version of MDPQ-16[b] score. Maximum affinity for technology interaction score: 6 points; maximum MDPQ-16 score: 5 points.

	Physicians (n=23), mean (SD)	Patients (n=16), mean (SD)	*P* value
ATI	3.50 (0.66)	4.00 (0.88)	.05
MDPQ-16	4.81 (0.26)	4.74 (0.45)	.60

[a]ATI: affinity for technology interaction.

[b]MDPQ-16: Mobile Device Proficiency Questionnaire.

XSL•FO
RenderX

The Spearman correlation coefficient revealed a slight negative correlation between the ATI score and age for both physicians (−0.224; *P*=.30) and patients (−0.290; *P*=.28). The younger the physicians and patients were, the more likely they were to consider themselves technology-savvy (ATI).

A small positive correlation was observed between the ATI score and MDPQ-16 score, with a coefficient of 0.256 (*P*=.24) for physicians and 0.322 (*P*=.22) for patients.

When analyzing both groups together, the Spearman correlation between ATI score and age was −0.206 (*P*=.21), and between ATI score and MDPQ-16 score, it was 0.309 (*P*=.06).

Patients' Expectations and Comments

The open-ended questionnaire provided valuable insights into patients' expectations and preferences regarding MHAs for managing urticaria. The 16 patients expressed a high level of openness to using MHAs to support their condition. Key functionalities that patients found particularly important included symptom tracking (n=14, 88%), photo documentation (n=15, 94%), medication management (n=14, 88%), and a note-taking and question feature (n=14, 88%). Almost half of the patients indicated a desire for an introductory guide to the app and technology training, as well as a digital companion to assist with navigation and provide encouragement. Additionally, 10 (63%) patients expressed interest in having a chat function with their doctor and a platform to connect with other individuals affected by urticaria.

A majority of the 16 patients (n=11, 69%) also requested a feature that would help organize and remind them of appointments. Other specific areas where patients sought support included information and educational content (n=11, 69%), tips for symptom management (n=11, 69%), psychosomatic support (n=5, 32%), and holistic treatment approaches (n=10, 63%).

To ensure regular use of the app, patients indicated that it would need to offer tangible improvements in their care, particularly by helping them manage their condition more effectively. Under these conditions, patients were open to using the app more than three times a week.

Additional patient wishes included "disease progression documentation," "symptom tracking," "support during acute flare-ups," "therapy suggestions," "features for identifying triggers," "emergency support features," "behavioral guidance for daily living," and "user-friendly interface."

Regarding the design of the CRUSE Control app, 14 (88%) out of 16 patients rated it at less than or equal to 7 out of 10 points. The information content was rated favorably by 12 (75%) out of 16 patients (≥7/10 points), while the usability and functionality were also rated positively by 12 (75%) and 10 (63%), respectively.

Discussion

Principal Findings

This study provides the first systematic analysis of MHAs for patients with CSU available in the App Stores and online. The patient cohort (mean age 40, SD 13.4 y; 6 men and 10 women) reflected the typical demographic pattern in CSU, with women being affected approximately twice as often as men [5,7,8]. The physician cohort (mean age 34, SD 9.7 y; 6 men and 17 women) represented the younger and predominantly female demographic commonly observed among university hospital staff in dermatology [39]. Despite broad inclusion criteria, only 1 out of 15 identified CSU apps met all predefined requirements. This highlights a substantial gap in the digital support landscape for CSU and underscores the need for targeted research and development of high-quality, evidence-based apps.

The episodic nature of CSU presents significant challenges for an accurate assessment of disease severity and activity [2]. Physicians rarely have the opportunity to observe patients during active flare-ups and rely on retrospective patient reports, photographs, and time-consuming interviews, which may result in misclassification and suboptimal treatment decisions. However, current guidelines recommend regular monitoring, ideally daily to weekly, using validated instruments such as the UAS and UCT to guide therapeutic adjustments [15]. MHAs have the potential to provide the solution. They constitute promising tools for continuous real-time symptom documentation, consolidating data into overviews to support both patients and physicians, and enhancing patient self-management and confidence. As their use in health care is becoming increasingly popular and steadily expanding, there is a growing need for their comprehensive and standardized evaluation [40]. A systematic review and evaluation of available medical apps, as provided by this study, using established and validated evaluation tools, is essential to gather feedback from both users (patients) and physicians (prescribers). This feedback will play a critical role in the ongoing development of MHAs [24,41]. The more closely an MHA is aligned with the needs of its users, the greater the likelihood of sustained engagement and long-term use [42].

The identified app, the CRUSE Control app, received high ratings across all MARS subdomains, including engagement, functionality, esthetics and information, with mean ratings of 4.03 (SD 0.45) on the MARS and 4.06 (SD 0.40) on the uMARS. However, high-quality ratings alone do not necessarily ensure sustained use [43]. Research has consistently shown that many MHAs, even those with favorable initial evaluations, experience a sharp decline in user engagement over time. In a mental health app, for instance, retention rates dropped to as low as 3.3% after 30 days [44]. Long-term adherence poses a major challenge in digital health [45] but is essential for achieving meaningful clinical outcomes. This is particularly important for chronic conditions such as CSU, which require ongoing monitoring of symptoms.

The positive evaluation of the CRUSE Control app is promising. Its high functionality and intuitive design facilitates rapid symptom documentation. This can reduce time burden and potentially support sustained user engagement, a factor consistently recognized as essential for long-term adherence in previous studies [46,47].

Although a user-friendly design may enhance initial engagement, sustained engagement ultimately depends on the app's ability to meet the specific needs of its target users [47].

XSL•FO
RenderX

In this study, the differing priorities of consumers were reflected in the evaluations.

Physicians rated the functionality of the CRUSE Control app significantly lower than patients (P=.04) [41,48]. They rather emphasized clinical relevance, accuracy, reliability, and integration into workflows. Their focus was on precise data collection, treatment support, and evidence-based decision-making to ensure that the app meets medical standards and functions without errors. Clinical quality was also confirmed by the physicians' rating of the psychotherapy subcategory (mean 3.83, SD 0.61). It was relatively high, indicating favorable assessments of patients' safety, the suitability of app-based delivery, and the quality of therapeutic content.

In contrast, patients valued ease of use, intuitive navigation, and practical features for daily self-management. Rapid symptom documentation, completed within minutes, was seen as a major advantage. It reduced user burden and likely promoted adherence and sustained engagement. In subcategory functionality, patients appreciated particularly its usability and its role in everyday health management.

The G-MAUQ usability score reflects this positive perception of both groups. Physicians and patients alike rated the CRUSE Control app highly for usability, with "ease of use" and "user satisfaction" scores nearing 6 out of 7 on the G-MAUQ. The ratings of the app's "usefulness" were slightly lower, though the differences between the groups were not significant.

Despite differences in priorities and usability needs, these high ratings indicate that both groups find significant value in the app: physicians for its clinical use and patients for its support in self-management.

Both patients with CSU and physicians demonstrated high technological affinity and competence in using mobile devices, as reflected by high ATI and MDPQ-16 scores. Hence, MHAs can be adopted and used effectively by this group [49,50]. A slight negative correlation between age and technological affinity was observed. This indicates that younger individuals tend to be more comfortable with digital tools. Older users may need additional support to ensure equitable access and sustained engagement.

Patient feedback in the open-ended questionnaire demonstrated a clear willingness to use MHAs for managing CSU, particularly if the app offered real benefits such as symptom tracking, photo documentation, and medication management. Patients wish for psychosocial support, education, and communication features. A future app could integrate community and peer support functionalities through health care professional–moderated forums or support groups, aiming to reduce patient isolation and encourage sharing personal experiences.

Limitations

Data collection focused on app ratings, without detailed clinical data such as disease severity or duration, which limits potential correlations with app use. The short-term evaluation of a single app without direct comparison to other CSU apps may have introduced some bias. The relatively young study population also limits the generalizability of the findings to older patients.

A major limitation of this study is that only 1 app met the predefined inclusion criteria at the time of the search, potentially introducing selection bias. The findings primarily reflect the characteristics of this single app and should be interpreted with caution regarding the broader CSU mHealth landscape. Because CRUSE Control is a well-developed and medically guided app, there is a possibility of performance or reporting bias, which could have led to an optimistic assessment of the potential of mHealth for CSU.

Future studies should include broader clinical data, longer follow-up, and direct comparisons across multiple CSU apps to reduce bias and enhance generalizability.

Conclusions

Despite broad search criteria, only 1 app met all inclusion requirements, highlighting a significant gap in digital support for patients with CSU. Both patients and physicians rated the app highly in terms of quality, usability, and information content. Patients rated functionality significantly higher, reflecting differing priorities between ease of daily use and clinical rigor. High technological affinity among both groups suggests favorable conditions for MHA adoption. Patients expressed clear interest in features supporting symptom tracking, medication management, and psychosocial support, emphasizing the need for patient-centered designs and onboarding assistance. Overall, MHAs show great potential to enhance CSU care.

While these findings are limited to a single app, they indicate that MHAs may have the potential to enhance CSU care. Future research should incorporate broader clinical parameters, longer follow-up, and diverse patient populations to further assess app effectiveness and optimize digital health tools for CSU management.

Acknowledgments

The authors sincerely thank all the patients and physicians who participated in this study. Their input was essential in advancing our understanding of mobile health apps for managing chronic spontaneous urticaria. No generative artificial intelligence tools were used in the writing or editing of this manuscript.

Funding

CG was supported by the TWINSIGHT Clinician Scientist Program (project: TWINSIGHT-09) at the Medical Faculty, University of Würzburg, that is funded by the Else-Kröner-Fresenius Foundation.

XSL•FO
RenderX

Data Availability

Data is presented in the main manuscript.

Authors' Contributions

Conceptualization: AS
Data curation: CG, BG, TD, LB
Formal analysis: CG
Investigation: CG, BG, TD, LB, MG
Methodology: CG, BG, TD, LB
Resources: MG
Supervision: MG, AS
Validation: CG, BG, TD, LB, MG, AS
Visualization: CG, BG, TD, LB, AS
Writing – original draft: CG, BG, TD, LB, AS
Writing – review & editing: MG, AS

Conflicts of Interest

None declared.

Checklist 1
PRISMA checklist.
[DOCX File, 31 KB - mhealth_v14i1e71435_app1.docx]

Checklist 2
STROBE checklist.
[DOCX File, 34 KB - mhealth_v14i1e71435_app2.docx]

References

1. Sussman G, Abuzakouk M, Bérard F, et al. Angioedema in chronic spontaneous urticaria is underdiagnosed and has a substantial impact: analyses from ASSURE-CSU. Allergy 2018 Aug;73(8):1724-1734. [doi: 10.1111/all.13430] [Medline: 29460968]
2. Zuberbier T, Altrichter S, Bauer S, et al. S3-Leitlinie Urtikaria. Teil 1: Klassifikation und Diagnostik der Urtikaria - deutschsprachige Adaptation der internationalen S3-Leitlinie [article in German]. J Deutsche Derma Gesell 2023 Jan;21(1):81-95. [doi: 10.1111/ddg.14906_g]
3. Vestergaard C, Deleuran M. Chronic spontaneous urticaria: latest developments in aetiology, diagnosis and therapy. Ther Adv Chronic Dis 2015 Nov;6(6):304-313. [doi: 10.1177/2040622315603951] [Medline: 26568807]
4. Fricke J, Ávila G, Keller T, et al. Prevalence of chronic urticaria in children and adults across the globe: systematic review with meta-analysis. Allergy 2020 Feb;75(2):423-432. [doi: 10.1111/all.14037] [Medline: 31494963]
5. Maurer M, Weller K, Bindslev-Jensen C, et al. Unmet clinical needs in chronic spontaneous urticaria. A GA²LEN task force report. Allergy 2011 Mar;66(3):317-330. [doi: 10.1111/j.1398-9995.2010.02496.x] [Medline: 21083565]
6. Kolkhir P, Bonnekoh H, Metz M, Maurer M. Chronic spontaneous urticaria: a review. JAMA 2024 Nov 5;332(17):1464-1477. [doi: 10.1001/jama.2024.15568] [Medline: 39325444]
7. Wertenteil S, Strunk A, Garg A. Prevalence estimates for chronic urticaria in the United States: a sex- and age-adjusted population analysis. J Am Acad Dermatol 2019 Jul;81(1):152-156. [doi: 10.1016/j.jaad.2019.02.064] [Medline: 30872154]
8. Liu X, Cao Y, Wang W. Burden of and trends in urticaria globally, regionally, and nationally from 1990 to 2019: systematic analysis. JMIR Public Health Surveill 2023 Oct 26;9:e50114. [doi: 10.2196/50114] [Medline: 37883176]
9. Gonçalo M, Gimenéz-Arnau A, Al-Ahmad M, et al. The global burden of chronic urticaria for the patient and society. Br J Dermatol 2021 Feb;184(2):226-236. [doi: 10.1111/bjd.19561] [Medline: 32956489]
10. Maurer M, Giménez-Arnau A, Ensina LF, Chu CY, Jaumont X, Tassinari P. Chronic urticaria treatment patterns and changes in quality of life: AWARE study 2-year results. World Allergy Organ J 2020 Sep;13(9):100460. [doi: 10.1016/j.waojou.2020.100460] [Medline: 32983330]
11. Zuberbier T, Abdul Latiff AH, Abuzakouk M, et al. The international EAACI/GA²LEN/EuroGuiDerm/APAAACI guideline for the definition, classification, diagnosis, and management of urticaria. Allergy 2022 Mar;77(3):734-766. [doi: 10.1111/all.15090] [Medline: 34536239]
12. Maurer M, Abuzakouk M, Bérard F, et al. The burden of chronic spontaneous urticaria is substantial: real-world evidence from ASSURE-CSU. Allergy 2017 Dec;72(12):2005-2016. [doi: 10.1111/all.13209] [Medline: 28543019]

XSL•FO
RenderX

13. Weller K, Koti I, Makris M, Maurer M. Anxiety and depression seem less common in patients with autoreactive chronic spontaneous urticaria. Clin Exp Dermatol 2013 Dec;38(8):870-873. [doi: 10.1111/ced.12190] [Medline: 23889120]
14. Staubach P, Dechene M, Metz M, et al. High prevalence of mental disorders and emotional distress in patients with chronic spontaneous urticaria. Acta Derm Venereol 2011 Sep;91(5):557-561. [doi: 10.2340/00015555-1109] [Medline: 21597672]
15. Zuberbier T, Altrichter S, Bauer S, et al. S3 - Leitlinie Urtikaria. Teil 2: Therapie der Urtikaria – deutschsprachige Adaption der internationalen S3 - Leitlinie [article in German]. J Deutsche Derma Gesell 2023 Feb;21(2):202-216. [doi: 10.1111/ddg.14932_g]
16. Netchiporouk E, Sasseville D, Moreau L, Habel Y, Rahme E, Ben-Shoshan M. Evaluating comorbidities, natural history, and predictors of early resolution in a cohort of children with chronic urticaria. JAMA Dermatol 2017 Dec 1;153(12):1236-1242. [doi: 10.1001/jamadermatol.2017.3182] [Medline: 28973060]
17. Kolkhir P, Altrichter S, Asero R, et al. Autoimmune diseases are linked to type IIb autoimmune chronic spontaneous urticaria. Allergy Asthma Immunol Res 2021 Jul;13(4):545-559. [doi: 10.4168/aair.2021.13.4.545] [Medline: 34212543]
18. Baiardini I, Braido F, Bindslev-Jensen C, et al. Recommendations for assessing patient-reported outcomes and health-related quality of life in patients with urticaria: a GA(2) LEN taskforce position paper. Allergy 2011 Jul;66(7):840-844. [doi: 10.1111/j.1398-9995.2011.02580.x] [Medline: 21385184]
19. van der Valk PGM, Moret G, Kiemeney L. The natural history of chronic urticaria and angioedema in patients visiting a tertiary referral centre. Br J Dermatol 2002 Jan;146(1):110-113. [doi: 10.1046/j.1365-2133.2002.04582.x] [Medline: 11841375]
20. Finnane A, Curiel-Lewandrowski C, Wimberley G, et al. Proposed technical guidelines for the acquisition of clinical images of skin-related conditions. JAMA Dermatol 2017 May 1;153(5):453-457. [doi: 10.1001/jamadermatol.2016.6214] [Medline: 28241182]
21. Altunisik N, Turkmen D, Calikoglu E, Sener S. Views and experiences of dermatologists in Turkey about teledermatology during the COVID - 19 pandemic. J Cosmet Dermatol 2020 Oct;19(10):2460-2463. [doi: 10.1111/jocd.13677]
22. Kashetsky N, Mar K, Liu C, Rivers JK, Mukovozov I. Photography in dermatology: a scoping review: practices, skin of color, patient preferences, and medical - legal considerations. J Deutsche Derma Gesell 2023 Oct;21(10):1102-1107. [doi: 10.1111/ddg.15129]
23. Hindelang M, Sitaru S, Fischer C, Biedermann T, Zink A. Bridging the gap through telemedicine: pilot study on the acceptance and use of teledermatology for urticaria. J Deutsche Derma Gesell 2025 Jan;23(1):40-52. [doi: 10.1111/ddg.15557]
24. Bol N, Helberger N, Weert JCM. Differences in mobile health app use: a source of new digital inequalities? Inf Soc 2018 May 27;34(3):183-193. [doi: 10.1080/01972243.2018.1438550]
25. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform 2012 Feb;45(1):184-198. [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]
26. Matsumura K, Nakagomi A, Yagi E, et al. Impact of an mHealth app (Kencom) on patients with untreated hypertension initiating antihypertensive medications: real-world cohort study. JMIR Cardio 2024 Nov 26;8:e52266. [doi: 10.2196/52266] [Medline: 39621938]
27. Kernebeck S, Busse TS, Böttcher MD, Weitz J, Ehlers J, Bork U. Impact of mobile health and medical applications on clinical practice in gastroenterology. World J Gastroenterol 2020 Aug 7;26(29):4182-4197. [doi: 10.3748/wjg.v26.i29.4182] [Medline: 32848328]
28. Anzahl der im apple app-store verfügbaren mhealth-apps vom 1 quartal 2015 bis zum 3 quartal 2020 [article in German]. Statista. URL: https://de.statista.com/statistik/daten/studie/1191205/umfrage/anzahl-der-bei-apple-verfuegbaren-mhealth-apps/ [accessed 2026-02-04]
29. Sauermann S, Herzberg J, Burkert S, Habetha S. DiGA: a chance for the German healthcare system. J Eur CME 2022;11(1):2014047. [doi: 10.1080/21614083.2021.2014047] [Medline: 34992948]
30. Gerke S, Stern AD, Minssen T. Germany's digital health reforms in the COVID-19 era: lessons and opportunities for other countries. NPJ Digit Med 2020;3(1):94. [doi: 10.1038/s41746-020-0306-7] [Medline: 32685700]
31. Reinders P, Augustin M, Otten M. General and dermatological population's use and acceptance of digital health in Germany: a representative survey. J Deutsche Derma Gesell 2024 Sep;22(9):1221-1231. [doi: 10.1111/ddg.15454]
32. Augustin M, Reinders P, Janke TM, et al. Attitudes toward and use of eHealth technologies among German dermatologists: repeated cross-sectional survey in 2019 and 2021. J Med Internet Res 2024 Feb 12;26:e45817. [doi: 10.2196/45817] [Medline: 38345855]
33. Haggag O, Grundy J, Abdelrazek M, Haggag S. A large scale analysis of mHealth app user reviews. Empir Softw Eng 2022;27(7):196. [doi: 10.1007/s10664-022-10222-6] [Medline: 36246486]
34. Rathner EM, Probst T, Terhorst Y, Barke A, Baumeister H. MARS - Mobile Application Rating Scale [video in German]. MARS-Rating Universität Ulm YouTube page. 2017 Jun 2. URL: https://www.youtube.com/watch?v=5vwMiCWC0Sc [accessed 2026-02-04]
35. Messner EM, Terhorst Y, Barke A, et al. The German version of the Mobile App Rating Scale (MARS-G): development and validation study. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e14479. [doi: 10.2196/14479] [Medline: 32217504]
36. Kopka M, Slagman A, Schorr C, et al. German mHealth App Usability Questionnaire (G-MAUQ) and short version (G-MAUQ-S): translation and validation study. Smart Health 2024;34:100517. [doi: 10.1016/j.smhl.2024.100517]

XSL•FO
RenderX

37. Franke T, Attig C, Wessel D. A personal resource for technology interaction: development and validation of the affinity for technology interaction (ATI) scale. Int J Hum–Comput Interact 2019 Apr 3;35(6):456-467. [doi: 10.1080/10447318.2018.1456150]
38. Roque NA, Boot WR. A new tool for assessing mobile device proficiency in older adults: the mobile device proficiency questionnaire. J Appl Gerontol 2018 Feb;37(2):131-156. [doi: 10.1177/0733464816642582] [Medline: 27255686]
39. Sakunchotpanit G, Trepanowski N, Awerman JL, Benlagha I, Nguyen B. Trends in gender representation and impact of practice setting on productivity and reimbursement gap for female Mohs micrographic surgery physicians: a cross-sectional study. Dermatol Surg 2024 Dec 1;50(12S):S207-S211. [doi: 10.1097/DSS.0000000000004457] [Medline: 39589848]
40. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3(1):14. [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]
41. Dege T, Glatzel B, Borst V, et al. Patient-centered chronic wound care mobile apps: systematic identification, analysis, and assessment. JMIR Mhealth Uhealth 2024 Mar 21;12:e51592. [doi: 10.2196/51592] [Medline: 38533818]
42. Biduski D, Bellei EA, Rodriguez JPM, Zaina LAM, De Marchi ACB. Assessing long-term user experience on a mobile health application through an in-app embedded conversation-based questionnaire. Comput Human Behav 2020 Mar;104:106169. [doi: 10.1016/j.chb.2019.106169]
43. Spethmann S, Hindricks G, Koehler K, et al. Telemonitoring for chronic heart failure: narrative review of the 20-year journey from concept to standard care in Germany. J Med Internet Res 2024 Dec 4;26:e63391. [doi: 10.2196/63391] [Medline: 39631073]
44. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567. [doi: 10.2196/14567] [Medline: 31573916]
45. Figueiredo T, Alcântara L, Carrilho J, Paúl C, Costa E. Understanding adherence to digital health technologies: systematic review of predictive factors. J Med Internet Res. 2025 Nov 17 p. e77362. [doi: 10.2196/77362] [Medline: 41248309]
46. Madeira RN, Germano H, Macedo P, Correia N. Personalising the user experience of a mobile health application towards patient engagement. Procedia Comput Sci 2018;141:428-433. [doi: 10.1016/j.procs.2018.10.173]
47. Gan DZQ, McGillivray L, Larsen ME, Christensen H, Torok M. Technology-supported strategies for promoting user engagement with digital mental health interventions: a systematic review. Digit Health 2022;8:20552076221098268. [doi: 10.1177/20552076221098268] [Medline: 35677785]
48. Gowin M, Cheney M, Gwin S, Franklin Wann T. Health and fitness app use in college students: a qualitative study. Am J Health Educ 2015 Jul 4;46(4):223-230. [doi: 10.1080/19325037.2015.1044140]
49. Hargittai E. From the “digital divide” to digital inequality: studying internet use as penetration increases. : Princeton Center for Arts and Cultural Policy Studies; 2001 URL: https://www.academia.edu/2802657/From_the_Digital_Divide_toDigital_Inequality_Studying_internet_use_as_penetration_increases [accessed 2026-02-14]
50. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125. [doi: 10.2196/jmir.5604] [Medline: 28428170]

Abbreviations

DiGA: Digital Health Applications/ Digitale Gesundheitsanwendung
ATI: affinity for technology interaction
CSU: chronic spontaneous urticaria
CU: chronic urticaria
DiGA: digitale gesundheitsanwendung (English: digital health apps)
G-MAUQ: German mHealth App Usability Questionnaire
MARS: Mobile Application Rating Scale
MDPQ-16: Mobile Device Proficiency Questionnaire
MHA: mobile health app
PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses
STROBE : Strengthening the Reporting of Observational Studies in Epidemiology
UAS: urticaria activity score
UCT: urticaria control test
uMARS: end-user version of Mobile Application Rating Scale

XSL•FO
RenderX

Edited by L Buis; submitted 17.Jan.2025; peer-reviewed by MA Tejedor-Alonso, M Bizjak-Suran, S Yang; revised version received 17.Dec.2025; accepted 17.Dec.2025; published 02.Mar.2026.

Please cite as:
Glatzel C, Glatzel B, Dege T, Brückel L, Goebeler M, Schmieder A
Identifying and Evaluating Patient-Centered Mobile and Web Apps for Patients With Chronic Spontaneous Urticaria: Systematic Search and Content Analysis
JMIR Mhealth Uhealth 2026;14:e71435
URL: https://mhealth.jmir.org/2026/1/e71435
doi:10.2196/71435

© Caroline Glatzel, Bernadette Glatzel, Tassilo Dege, Lena Brückel, Matthias Goebeler, Astrid Schmieder. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 2.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Clinical Usability of Exercise Prescription Apps for Professional Use: Systematic Review and Multidimensional Evaluation

Cheng-Hao Wu[1], MD; Che-Ning Chang[2], MD; Chu-Fang Chang[1], MSc; Ming-Hwai Lin[1], MD, PhD; Hsing-Yu Chen[3*], MD, PhD; Yu-Chun Chen[4*], MD, Dr med

[1]Department of Family Medicine, Taipei Veterans General Hospital, Taipei, Taiwan

[2]Department of Family Medicine, MacKay Memorial Hospital, Taipei, Taiwan

[3]Center for Traditional Chinese Medicine, Chang Gung Memorial Hospital, Guishan, Taiwan, Taiwan

[4]Taipei Veterans General Hospital, Yuli Branch, No. 91, Xinxing St., Yuli Township, Hualien County, Taiwan

[*]these authors contributed equally

Corresponding Author:
Yu-Chun Chen, MD, Dr med
Taipei Veterans General Hospital, Yuli Branch, No. 91, Xinxing St., Yuli Township, Hualien County, Taiwan

Abstract

Background: Exercise prescription is a structured and individualized intervention that requires appropriate progression, tailoring, and behavioral support to ensure safety and long-term effectiveness. With the expansion of mobile health technologies, exercise prescription apps are increasingly used to support the remote delivery of prescribed exercise programs. However, the extent to which widely adopted apps align with established clinical standards remains unclear.

Objective: This study aimed to evaluate the clinical usability of popular, no-cost exercise prescription apps from a professional perspective, focusing on clinical integrity, intervention fidelity, behavioral mechanisms, and clinician-assessed digital usability.

Methods: A systematic search of Google Play and the Apple App Store identified widely adopted apps that enable clinician-directed exercise prescription. Eligible apps were evaluated using established frameworks, including the frequency, intensity, time, and type (FITT) and FITT, volume, and progression (FITT-VP) principles; the Consensus on Exercise Reporting Template (CERT); the Behavior Change Technique Taxonomy version 1 (BCTTv1); and the Mobile App Rating Scale (MARS). Descriptive analyses and interrater reliability assessments were performed.

Results: Six apps met the inclusion criteria. All satisfied the basic FITT requirements; however, none incorporated explicit guidance on exercise progression or individualized adjustment consistent with the FITT-VP principles. CERT evaluation demonstrated comprehensive reporting of structural components but a consistent absence of progression logic, tailoring strategies, and adverse event documentation. Although multiple behavior change techniques were identified, several techniques considered important for graded progression and sustained adherence in unsupervised settings were infrequently implemented or absent. Overall app quality was moderate, characterized by strong functionality but limited engagement. Only 2 apps reported evidence of scientific evaluation.

Conclusions: Widely adopted exercise prescription apps meet fundamental structural requirements but do not fully support the progressive and individualized processes central to clinical exercise prescription. These findings highlight a gap between structural prescription delivery and independent clinical exercise management. Exercise prescription apps may therefore be most appropriately positioned as adjunctive tools within clinician-guided or hybrid care models. Future development should prioritize transparent progression mechanisms, individualized adjustment, and the implementation of clinically relevant behavior change strategies to enhance safety and long-term effectiveness.

(JMIR Mhealth Uhealth 2026;14:e77616) doi:10.2196/77616

KEYWORDS

exercise prescription; mHealth apps; behavior change techniques; FITT principle; digital health; app usability; clinical evaluation; mobile health; rehabilitation; exercise adherence

Introduction

Exercise prescription is a structured and individualized program of exercise, systematically designed to improve health or functional capacity, and these programs are commonly developed by specialists on the basis of patients' clinical status, goals, and needs [1]. Physicians in physical medicine and rehabilitation, as well as in primary care, routinely prescribe exercise as both treatment and secondary prevention for a broad range of conditions, including musculoskeletal, cardiovascular,

XSL•FO
RenderX

neurological, and pulmonary diseases [2-9]. Given this wide scope of application, precision is essential to ensure both safety and effectiveness. According to the American College of Sports Medicine (ACSM), exercise prescription should specify exercise frequency, intensity, time, and type (FITT), with additional consideration of volume and progression (FITT-VP) to support appropriate physiological adaptation and minimize the risk of adverse events [10]. When these elements are appropriately defined and individualized, exercise therapy can deliver evidence-based benefits while reducing the likelihood of exercise-related harm [11].

With the expansion of mobile health (mHealth) technologies, exercise prescription apps have emerged as tools to support the remote delivery of prescribed exercise programs. These systems typically consist of a clinician-facing prescription platform and a patient-facing app, enabling clinicians to design structured exercise programs that patients can access via mobile devices or computers. Prior studies indicate that digitally supported exercise interventions can be feasible, cost-effective, and minimally disruptive to daily life while maintaining patient satisfaction with outpatient services [12-15]. Randomized controlled trials have further demonstrated improvements in adherence to home exercise programs and increased confidence in exercising [16]. Since the COVID-19 pandemic, the growing adoption of digital health solutions has also expanded the use of online training platforms and home-based exercise apps [17-20]. Moreover, app-supported exercise prescription has shown potential benefits in specific clinical populations, including individuals with sarcopenia [21], myocardial infarction [22], or long COVID [23].

Despite growing evidence supporting the feasibility and effectiveness of digitally delivered exercise interventions, concerns remain regarding the professionalism and completeness of exercise prescriptions generated by mobile apps [24]. Compared with face-to-face care, mobile platforms may limit the delivery of detailed instructions, real-time biomechanical feedback, and hands-on correction, all of which are integral to safe and effective exercise prescription. From a professional perspective, the suitability of exercise prescription apps depends on their ability to support clinicians in prescribing, monitoring, and adjusting exercise programs in accordance with established clinical standards. Key considerations include whether exercise prescriptions adequately address essential ACSM components, whether intervention delivery is transparent and reproducible, and whether behavioral strategies are incorporated to support adherence in unsupervised or remote settings. In addition, the rapid development cycles of commercial apps raise concerns regarding usability, consistency, and quality from a clinical perspective [25]. While existing evaluations of health apps often focus on general users, single-disease outcomes, or basic activity tracking [26-30], evidence remains limited regarding the extent to which widely used exercise prescription apps meet professional requirements for clinical prescription and implementation.

To address these gaps, this study adopted a clinician-centered, multidimensional evaluation framework to assess whether popular, no-cost exercise prescription apps are suitable for professional clinical use. The evaluation focused on 4 complementary domains reflecting core requirements of exercise prescription in real-world practice: clinical integrity of exercise prescriptions, intervention fidelity and transparency, behavioral mechanisms supporting adherence, and digital usability and overall app quality. These domains were operationalized using established and validated instruments, including the FITT and FITT-VP principles, the Consensus on Exercise Reporting Template (CERT), the Behavior Change Technique Taxonomy version 1 (BCTTv1), and the Mobile App Rating Scale (MARS). By integrating these perspectives, this study extends beyond general app quality or user-centered evaluations and provides a structured professional assessment of exercise prescription apps intended for clinician-guided care.

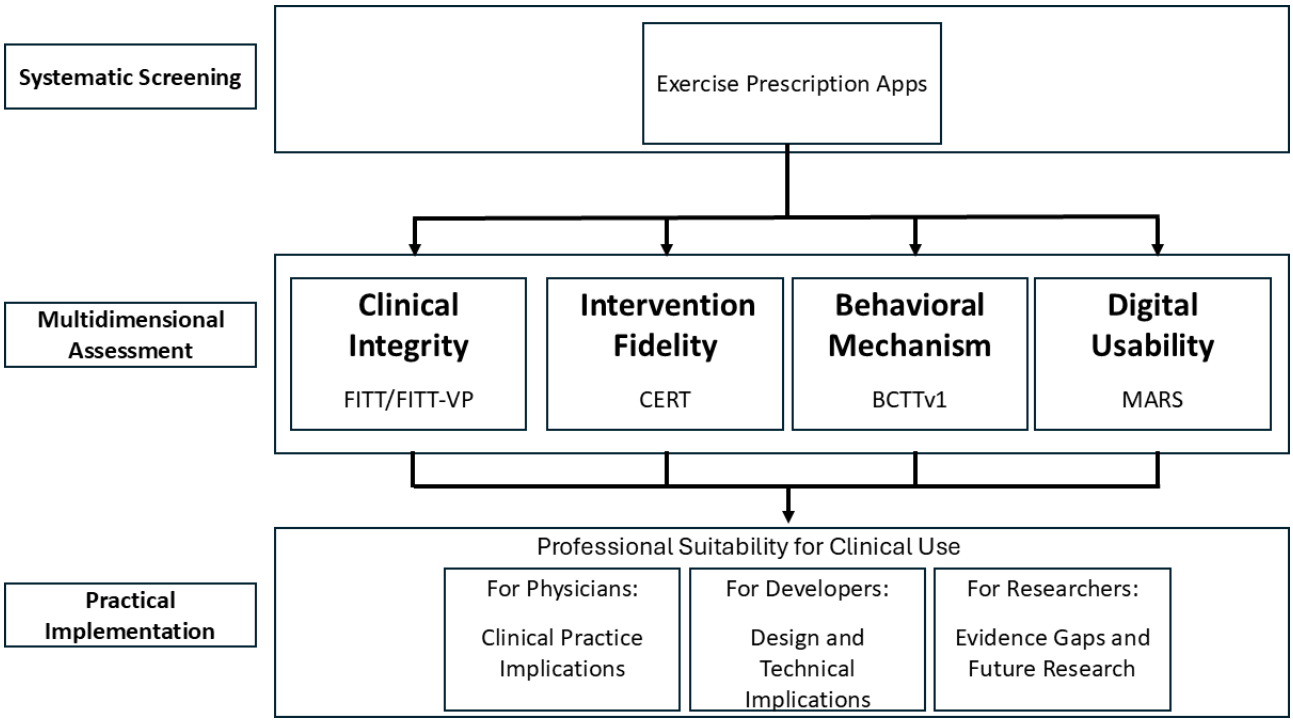
Methods

Study Design and Conceptual Framework

This study adopted a clinician-centered, cross-sectional evaluation design to assess exercise prescription apps intended for professional use. A multidimensional evaluation framework was applied to reflect the key components of clinical exercise prescription practice (Figure 1).

XSL•FO
RenderX

Figure 1. Conceptual framework for clinician-centered evaluation of exercise prescription apps. BCTTv1: Behavior Change Technique Taxonomy version 1; CERT: Consensus on Exercise Reporting Template; FITT: frequency, intensity, time, and type; FITT-VP: frequency, intensity, time, type, volume, and progression; MARS: Mobile App Rating Scale.



The framework included four domains: (1) clinical integrity, (2) intervention fidelity, (3) behavioral mechanism, and (4) digital usability. These domains were operationalized using established instruments: the FITT and FITT-VP principles, CERT, BCTTv1, and MARS.

Ethical Considerations

This study did not involve human participants, human data, or human biological materials. Therefore, institutional ethics board review and approval were not required.

App Identification and Search Strategy

This review adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) protocol, as detailed in the PRISMA 2020 statement (Checklist 1) [31]. A systematic search of Google Play and the Apple App Store was conducted on July 13, 2024, using Taiwanese regional settings. The following search terms were used: “exercise prescription AND doctor,” “exercise prescription AND therapist,” and “exercise prescription AND rehabilitation.” Duplicate apps were removed.

Apps were eligible if they were free, available in English or Chinese, designed for adult users, and not disease specific; did not require additional equipment; and allowed clinicians to prescribe exercise programs. To ensure market relevance, only apps available on both platforms with more than 10,000 downloads were included. A comprehensive and transparent description of this search process is provided in Multimedia Appendix 1.

Evaluation Framework and Domains

Clinical Integrity (FITT and FITT-VP)

Clinical integrity was assessed using the FITT and FITT-VP principles [10]. Apps were considered to meet the FITT criteria if all 4 core components were explicitly specified within the prescribed exercise programs. In addition, adherence to FITT-VP required explicit progression rules or defined mechanisms for adjusting the exercise volume or intensity over time.

Intervention Fidelity (CERT)

Intervention fidelity was evaluated using CERT, a 16-item checklist designed to assess the completeness and transparency of exercise interventions [32,33]. The CERT domains include materials, provider qualifications, delivery procedures, setting, dosage, tailoring, and adherence. Apps were assessed using a checklist approach to determine whether each CERT item was addressed. Achievement rates were calculated for individual items. The primary outcome was the achievement rate for individual CERT items.

Behavioral Mechanism (BCTTv1)

Behavioral mechanisms were evaluated using BCTTv1, a standardized framework for identifying discrete behavior change techniques (BCTs) within interventions [34]. The taxonomy includes 93 techniques and has been extended in mHealth apps to encompass 102 techniques [35]. Each app was systematically reviewed to identify the presence of BCTs according to BCTTv1 definitions. Achievement rates were calculated as the proportion of apps implementing each technique.

In this study, a predefined subset of BCTs within the BCTTv1 was designated as critical BCTs. These techniques were selected on the basis of prior studies examining strategies associated

XSL•FO
RenderX

with exercise initiation, maintenance, and physical activity promotion in digital or remote interventions [36-38]. The selected techniques were grouped into three functional categories: (1) behavior initiation, (2) behavior maintenance, and (3) promotion of physical activity (Multimedia Appendix 2). Achievement rates for this predefined subset were calculated separately.

Digital Usability (MARS)

Digital usability and overall app quality were evaluated using MARS, which assesses engagement, functionality, aesthetics, and information quality on a 5-point scale [39,40]. To assess the evidence base component of information quality, supporting scientific literature was identified through database searches in PubMed and Google Scholar, as well as through app descriptions and developer websites. Each item was independently rated by 2 reviewers.

Sample Assessment Procedure

All included apps were evaluated using their full range of patient-facing functions. Web-based prescription platforms were not assessed.

Each app was independently reviewed by 2 clinicians (CHW and CNC), both resident physicians with more than 3 years of clinical experience and extensive experience using Android and iOS systems. Prior to evaluation, both reviewers completed formal BCTTv1 training and certification and familiarized themselves with the CERT and MARS protocols. To calibrate scoring interpretations, 5 additional health and fitness apps not included in the final sample were jointly reviewed before data collection.

All apps were installed on an Apple iPhone 8 running iOS version 16.3.1 and a Samsung SM A5360 running Android version 13. Reviewers examined all available features for at least 30 minutes per app. Discrepancies in coding were resolved through discussion, and a third reviewer (YCC) was consulted when consensus could not be reached.

Data Extraction and Statistical Analysis

For our study, descriptive data were gathered from Google Play on July 29, 2024, using a Taiwanese account and regional settings. We collected information on app titles, developers, versions, and download numbers directly from the app listings. We chose not to include data from the App Store, as it closely mirrored the information available on Google Play but was less extensive. Ratings and review counts were also excluded from our analysis due to their significant variability across different countries, which complicates comparisons due to the apps' broad international user base. All statistical analyses were conducted using MedCalc Statistical Software (version 22.0.21; MedCalc Software Ltd). Given the exploratory design, descriptive statistics were primarily employed to present the findings. Interrater reliability was also assessed using Cohen κ or intraclass correlation coefficient (ICC) to determine the level of agreement between the 2 raters [41].

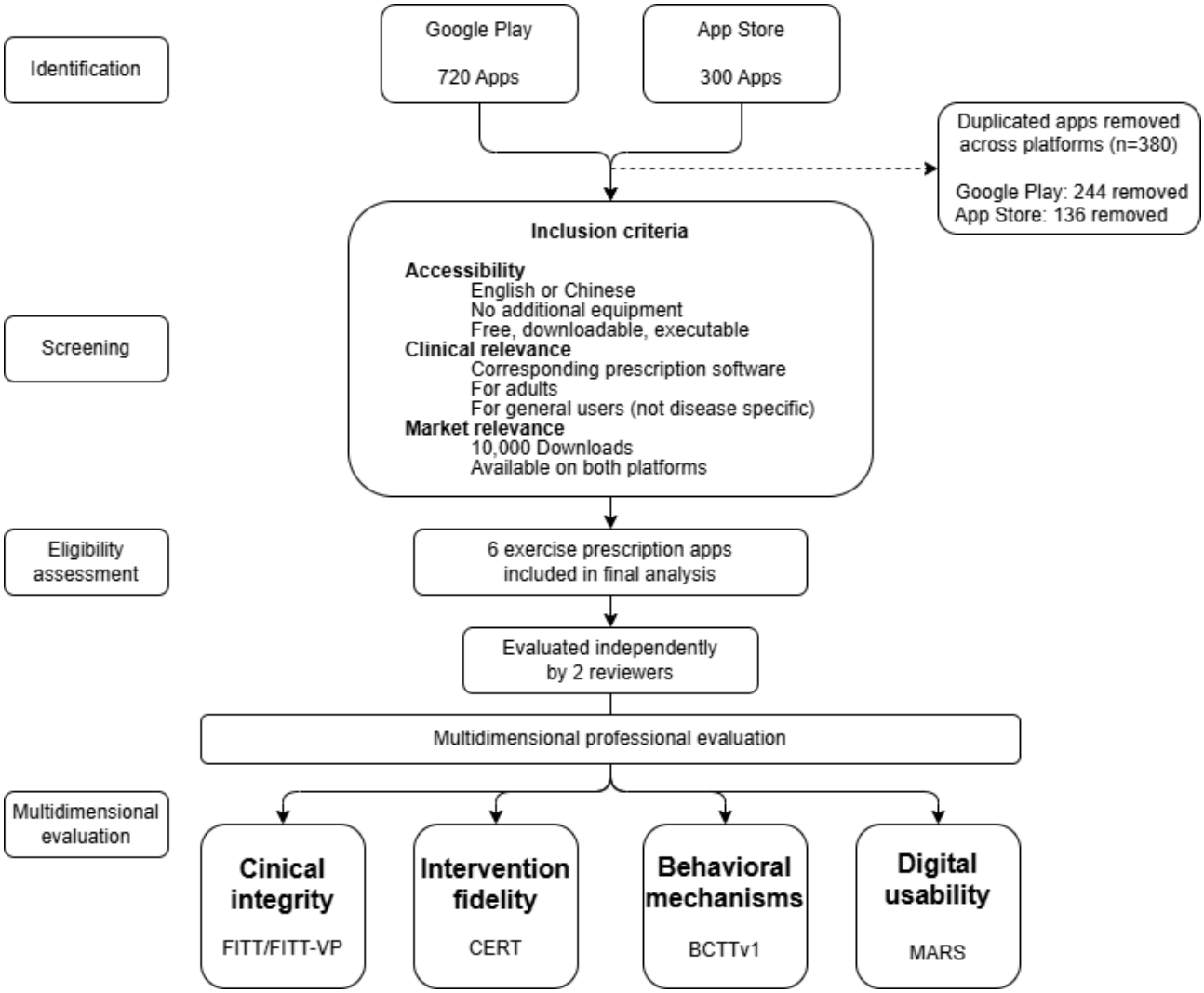
Results

App Identification and Selection

The initial search yielded a total of 1020 apps (each search term yielded 240 apps on Google Play and 100 on the App Store). After removing duplicates and applying the inclusion and exclusion criteria, 6 exercise prescription apps were included in the final analysis (Figure 2). All included apps were available on both platforms; were free to download; supported clinical exercise prescription; and had exceeded 10,000 downloads at the time of assessments.

XSL•FO
RenderX

Figure 2. Flow diagram of app identification, screening, eligibility assessment, and multidimensional professional evaluation. BCTTv1: Behavior Change Technique Taxonomy version 1; CERT: Consensus on Exercise Reporting Template; FITT: frequency, intensity, time, and type; FITT-VP: frequency, intensity, time, type, volume, and progression; MARS: Mobile App Rating Scale.



App Characteristics

As of July 29, 2024, the 6 apps included in our study demonstrated a wide range of installations, with numbers ranging from 10,000 to 1 million downloads (Table 1). The app developers were based in the United Kingdom, United States, Canada, and Australia. All apps were initially released between 2017 and 2020 and had all been updated in 2023 or 2024.

Table . Characteristics of included exercise prescription apps (N=6; July 29, 2024).

App characteristic	PhysiApp	MedBridge GO for Patients	Wibbi	TrackActive Pro - Patient App	Rehab Guru Client	Telehab
Number of installs	>1 million	>1 million	>100,000	>10,000	>10,000	>10,000
Developer	Physitrack PLC	MedBridge	Wibbi	Active Health Tech Pty Ltd	Rehab Guru Team	VALD
Country of developer	United Kingdom	United States	Canada	Australia	United Kingdom	Australia
Date of release	Oct 30, 2018	May 15, 2017	Feb 20, 2017	Feb 17, 2017	Aug 1, 2018	May 26, 2020
Version	4.20.0	4.6.3	1.3.1	1.6.7	3.1.0	1.3.0
Date of update	Jul 22, 2024	Apr 3, 2024	Jul 16, 2024	Feb 13, 2024	Oct 26, 2023	Mar 3, 2023

XSL•FO
RenderX

Clinical Integrity Assessment Based on FITT and FITT-VP Principles

All 6 apps met the FITT criteria, as they explicitly specified the exercise frequency, intensity, time, and type within prescribed exercise programs (Table 2). In contrast, none of the apps met the FITT-VP criteria. Across all apps, explicit guidance on progression of exercise volume or intensity over time was absent, and no app provided decision rules for adjusting exercise prescriptions based on user performance or program duration.

Table . Multidimensional evaluation results of included exercise prescription apps (N=6; July 29, 2024).

	PhysiApp	MedBridge GO for Patients	Wibbi	TrackActive Pro - Patient App	Rehab Guru Client	Telehab
Clinical integrity						
FITT[a]	Yes	Yes	Yes	Yes	Yes	Yes
FITT-VP[b]	No	No	No	No	No	No
Intervention fidelity						
CERT[c]	10	10	10	8	9	10
Behavioral mechanism						
BCTTv1[d]	18	16	15	14	15	19
Digital usability						
MARS[e]	3.82	3.85	3.36	3.36	3.55	3.99

[a]FITT: Frequency, intensity, type, time. “Yes” denotes full adherence to all 4 elements.

[b]FITT-VP: Frequency, intensity, type, time, volume, progression. “Yes” denotes full adherence to all 6 elements.

[c]CERT: Consensus on Exercise Reporting Template. Score indicates number of checklist items met (maximum=16).

[d]BCTTv1: Behavior Change Technique Taxonomy version 1. Number of identified behavior change techniques.

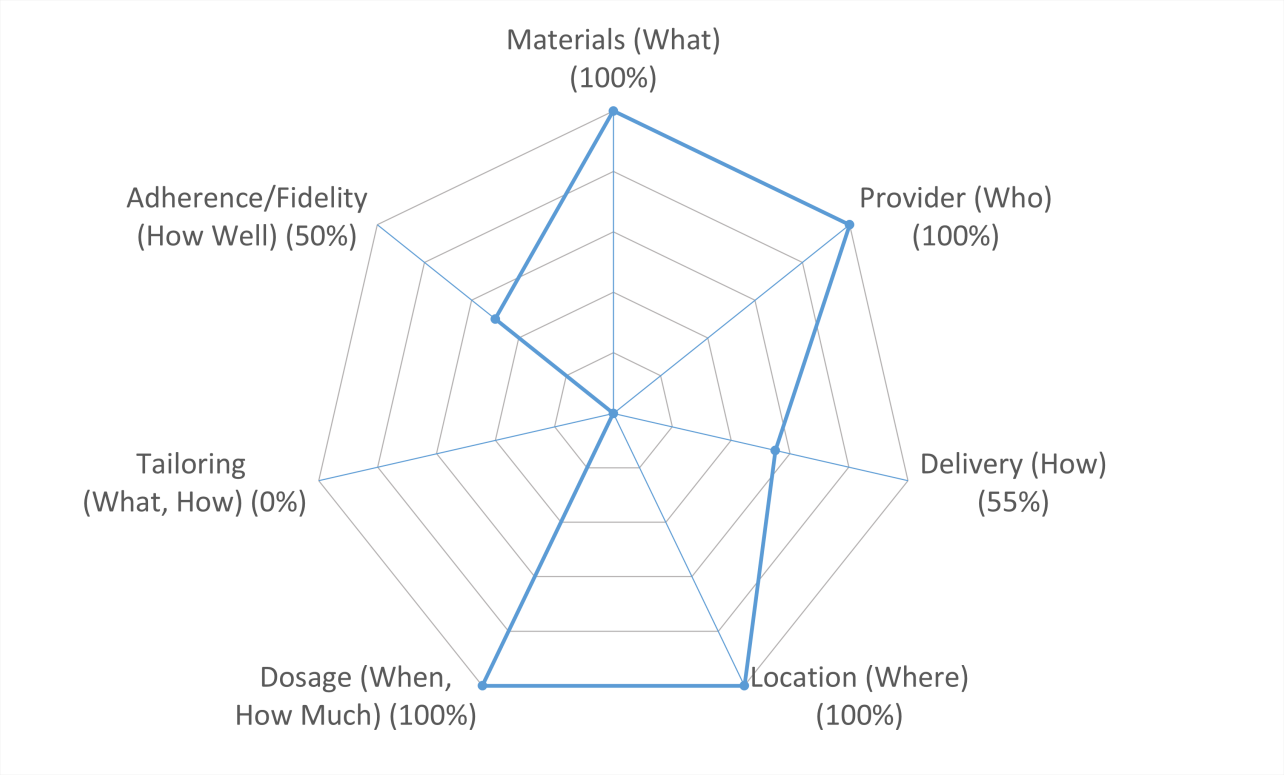
[e]MARS: Mobile App Rating Scale. Mean overall score on a 5-point scale.

Intervention Fidelity Assessment Based on CERT

Intervention fidelity varied across CERT domains (Table 3 and Figure 3). All 6 apps (100%) achieved complete reporting for core structural elements, including materials, provider qualifications, exercise setting, and dosage. Within the delivery domain, all apps specified whether exercises were performed individually or in groups and whether sessions were supervised or unsupervised. Of the 6 apps, 5 apps (83%) reported methods for assessing exercise adherence, and 4 apps (67%) included nonexercise components such as educational or dietary content.

XSL•FO
RenderX

Figure 3. Domain-level reporting patterns of the Consensus on Exercise Reporting Template across exercise prescription apps (N=6; 2024).



XSL•FO
RenderX

Table . Item-level achievement rates for the Consensus on Exercise Reporting Template (CERT) across included exercise prescription apps (N=6; 2024).

Abbreviated item description of CERT	Achievement rate
Materials (what)	
Detailed description of the type of exercise equipment	100%
Provider (who)	
Description of qualifications/expertise/training of instructor	100%
Delivery (how)	
Description of whether exercises are performed individually or in a group	100%
Description of whether exercises are supervised/unsupervised	100%
Description of the measurement/reporting of adherence to exercise	83%
Detailed description of motivation strategies	0%
Detailed description of the decision rule(s) of exercise progression	0%
Detailed description of how the exercise program was progressed	0%
Description of each exercise to enable replication (eg, illustrations and photographs)	0%
Description of any home programmed component	100%
Description of any nonexercise component	67%
Description of the type/number of adverse events that occurred during exercise	0%
Location (where)	
Description of exercise setting	100%
Dosage (when and how much)	
Description of exercise intervention and dosage	100%
Tailoring (what and how)	
Description of whether exercises are generic or tailored to the individual	0%
Detailed description of how exercises are tailored to the individual	0%
Decision rule for starting level of exercise	0%
Adherence/fidelity (how well)	
Description of how adherence or fidelity to the exercise intervention is assessed/measured	100%
Description of the extent to which the intervention was delivered as planned	0%

In contrast, no app (0%) reported motivational strategies, decision rules for exercise progression, or descriptions of how exercise programs were progressed over time. Similarly, none of the apps described whether or how exercises were tailored to individual users or how starting exercise levels were determined. Reporting of adverse events was absent across all apps, and none documented whether the prescribed intervention was delivered as planned. The comprehensive table with respective scoring can be found in Multimedia Appendix 3. Interrater reliability for CERT assessment was high (Cohen κ=0.88).

Behavioral Mechanism Assessment Based on BCTTv1

Overall Implementation of BCTs

Across the 6 included apps, a limited but heterogeneous set of BCTs was identified. When assessed against the original 93-item BCTTv1 taxonomy, the mean number of techniques implemented by apps was 13.2 (SD 1.77; range 11 - 16). When extended mHealth categories were included (102 techniques), the mean number of techniques increased to 16.2 (SD 1.77; range 14 - 19). The interrater reliability for BCT coding was high (Cohen κ=0.85).

Fourteen techniques were implemented across all apps (100% achievement), including goal setting (behavior), action planning, feedback on behavior, self-monitoring of behavior, instruction on how to perform the behavior, demonstration of the behavior,

XSL•FO
RenderX

and behavioral practice or rehearsal. Additional techniques were implemented in a subset of apps. The full list of identified BCTs and their achievement rates is presented in Table 4, and detailed coding results, including the extended 102-technique classification, are provided in Multimedia Appendix 4.

Table . Achievement rates of behavior change techniques (BCTs) identified from the full BCTTv1[a] taxonomy across included exercise prescription apps (N=6; 2024), ordered by frequency of implementation.

BCTTv1 code	BCT	Critical BCT[b]?	Achievement rate
1.1	Goal setting (behavior)	Yes	100%
1.4	Action planning	Yes	100%
1.5	Review behavior goal(s)	No	100%
2.1	Monitoring of behavior by others without feedback	No	100%
2.2	Feedback on behavior	Yes	100%
2.3	Self-monitoring of behavior	Yes	100%
4.1	Instruction on how to perform a behavior	Yes	100%
6.1	Demonstration of the behavior	Yes	100%
8.1	Behavioral practice/rehearsal	Yes	100%
9.1	Credible source	No	100%
12.6	Body changes	No	100%
17.1	Tailoring to demographic characteristics	No	100%
17.2	Tailoring to health status	No	100%
17.4	Adjusting intervention content to performance	No	100%
5.1	Information about health consequences	No	66%
10.4	Social reward	No	50%
1.3	Goal setting (outcome)	Yes	33%
1.7	Review outcome goal(s)	No	33%
2.5	Monitoring outcome(s) of behavior by others without feedback	No	33%

[a]BCTTv1: The Behavior Change Technique Taxonomy version 1. It comprises 102 distinct techniques. Only techniques identified in at least 1 app are presented; all remaining BCTTv1 items had an achievement rate of 0% across the evaluated exercise prescription apps.

[b]Critical BCTs were defined a priori as techniques with established relevance to exercise initiation, long-term maintenance, or physical activity promotion in digital or remote intervention contexts.

Predefined Clinically Relevant BCTs

Table 5 shows the predefined critical BCTs grouped into 3 functional categories: behavior initiation, behavior maintenance, and promotion of physical activity. Within behavior initiation, demonstration of the behavior and behavioral practice or rehearsal were incorporated by all apps (6/6, 100%), whereas biofeedback and graded tasks were not identified (0/6, 0%). For behavior maintenance, action planning and instruction on how to perform the behavior were universally implemented (6/6, 100%), while prompts or cues, graded tasks, and self-reward were absent (0/6, 0%). Among techniques associated with physical activity promotion, goal setting (behavior), action planning, feedback on behavior, and self-monitoring were consistently implemented, whereas social support, restructuring the physical environment, and framing or reframing were not identified.

XSL•FO
RenderX

Table . Achievement rates of predefined critical behavior change techniques across functional roles among included exercise prescription apps (N=6; 2024).

Functional category and behavior change techniques (BCTTv1[a] code)	Achievement rate
Behavior activation	
6.1 Demonstration of the behavior	100%
8.1 Behavioral practice/rehearsal	100%
2.6 Biofeedback	0%
8.7 Graded tasks	0%
Behavior maintenance	
1.4 Action planning	100%
4.1 Instruction on how to perform a behavior	100%
8.1 Behavioral practice/rehearsal	100%
7.1 Prompts/cues	0%
8.7 Graded tasks	0%
10.9 Self-reward	0%
Physical activity promotion	
1.1, 1.3 Goal setting (behavior and outcome)	100%
1.4 Action planning	100%
2.2 Feedback on behavior	100%
2.3, 2.4 Self-monitoring of behavior with its outcome(s)	100%
10.1 - 10.11 Items in group “Reward and threat”	50%
3.1 - 3.3 Items in group “Social support”	0%
12.1 Restructuring the physical environment	0%
13.2 Framing/reframing	0%

[a]BCTTv1: The Behavior Change Technique Taxonomy version 1.

Digital Usability Based on MARS

Assessment of overall app quality using MARS yielded a mean score of 3.66 (SD 0.25) across the 6 apps, with individual app scores ranging from 3.36 to 4.00 (Table 6 and Figure 4). Among MARS domains, engagement received the lowest ratings, with a mean score of 2.50 (SD 0.28; range 2.10 - 2.80), indicating limited use of interactive, motivational, or personalized features. In contrast, functionality was rated highest, achieving a mean score of 4.36 (SD 0.22; range 4.00 - 4.63), reflecting stable performance, ease of navigation, and technical reliability across all apps. Aesthetic quality was rated moderately to highly, with a mean score of 3.89 (SD 0.48; range 3.17 - 4.50), and information quality was rated similarly, with a mean score of 3.90 (SD 0.18; range 3.58 - 4.07). Comprehensive tables with both reviewer and respective scoring results are provided in Multimedia Appendices 5 and 6. The interrater reliability of MARS was assessed using the ICC (2-way random effects, absolute agreement, average measures). The resulting ICC was 0.71 (95% CI 0.59 - 0.81), indicating good reliability.

XSL•FO
RenderX

Figure 4. Radar chart illustrating the Mobile App Rating Scale scores of the 6 included exercise prescription apps (N=6; 2024).

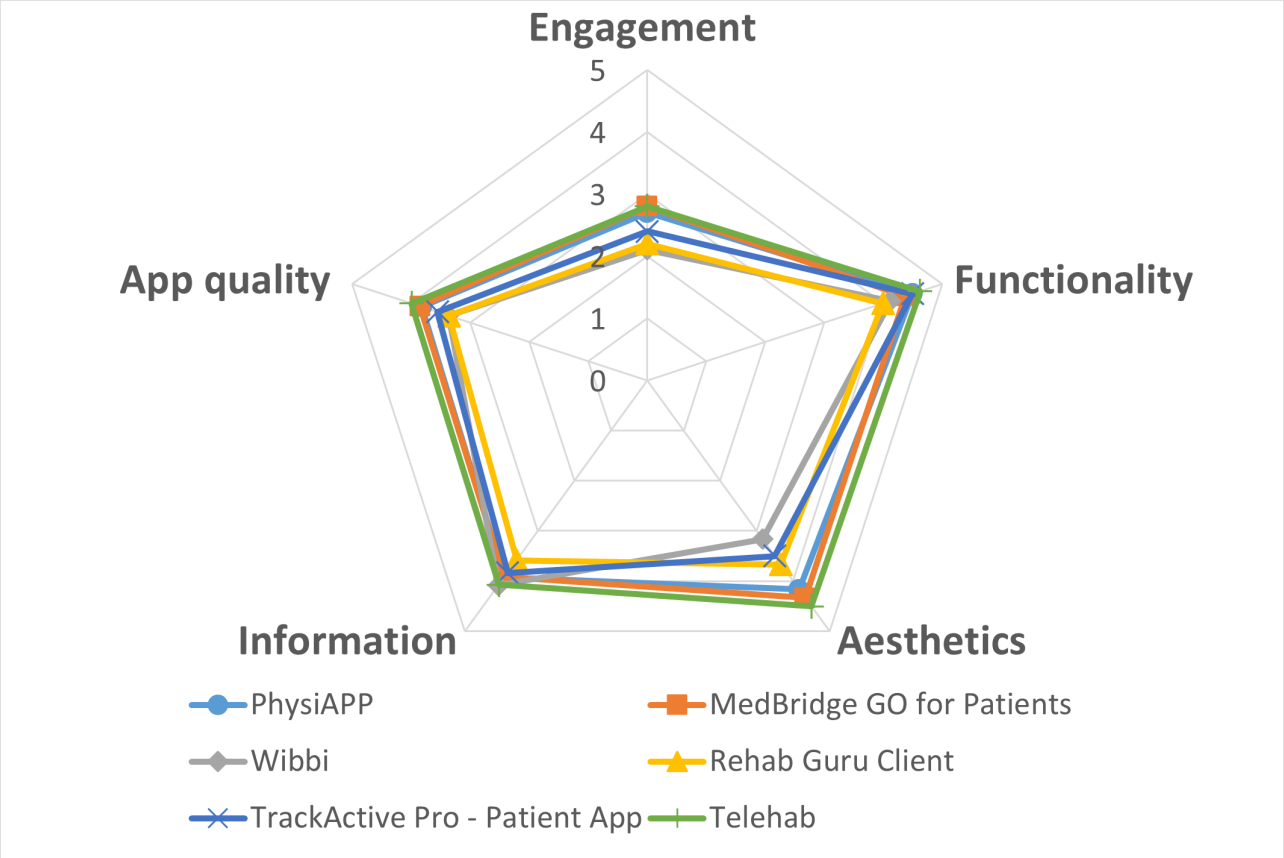


Table . Digital usability and quality scores of included exercise prescription apps (N=6; 2024) assessed using the Mobile App Rating Scale (MARS).

MARS domain	PhysiApp	MedBridge GO for Patients	Wibbi	TrackActive Pro - Patient App	Rehab Guru Client	Telehab	MARS score, mean (SD)[a]
Engagement	2.70	2.80	2.10	2.20	2.40	2.80	2.50 (0.28)
Functionality	4.50	4.38	4.13	4.00	4.50	4.63	4.36 (0.22)
Aesthetics	4.17	4.33	3.17	3.67	3.50	4.50	3.89 (0.48)
Information	3.93	3.92	4.08	3.58	3.83	4.07	3.90 (0.17)
Overall quality	3.82	3.86	3.37	3.36	3.56	4.00	3.66 (0.25)

[a]The maximum possible MARS score is 5.0, based on a 5-point Likert scale ranging from 1 (inadequate) to 5 (excellent).

Our review of the evidence item revealed that by July 2024, only 2 apps among the cohort had undergone scientific evaluation, suggesting a lack of empirically validated health apps in the current market.

Discussion

Principal Results

This study uses an integrated, clinician-centered, multidimensional evaluation framework to provide a novel professional perspective on how exercise prescription apps operate in real-world clinical contexts. Overall, popular, no-cost exercise prescription apps were found to meet the basic structural requirements for exercise prescription but fell short of key standards required for independent clinical use. Although prescriptions consistently specified core FITT elements, the absence of explicit progression and individualized adjustment limited alignment with established clinical practice. While multiple BCTs were identified, those critical for sustaining exercise behavior in unsupervised or remote settings were frequently lacking. Despite strong technical functionality and high information quality, low engagement scores suggest that usability alone is insufficient to support long-term adherence.

When considered collectively across domains, these apps nonetheless demonstrate operational strengths, including structured translation of prescriptions into actionable tasks, continuous visibility of adherence, low cognitive load instruction delivery, and delegation of routine monitoring. Taken together, these findings highlight a gap between clinical expectations for exercise prescription and the current capabilities of widely used digital tools, supporting the role of exercise prescription apps as adjuncts within clinician-guided care rather than as stand-alone prescribing solutions.

XSL•FO
RenderX

Conceptual Foundation and Innovation of the Multidimensional Evaluation Framework

This study introduces a clinician-centered evaluation framework that integrates clinical exercise prescription standards, behavior change theory, and digital usability into a unified, practice-oriented model. Unlike prior evaluations of physical activity and mHealth apps that have primarily examined usability, engagement, or behavior change features in isolation [26-28], this framework situates app assessment within the logic of real-world clinical exercise prescription and decision-making. By aligning established instruments with sequential stages of professional practice, including prescription structure and progression logic (FITT and FITT-VP), transparency and reproducibility of intervention delivery (CERT), support for exercise initiation and maintenance (BCTTv1), and interface quality and engagement (MARS), the framework enables clinicians to directly compare app features with clinical expectations. In doing so, it provides a shared reference to guide both informed clinical selection and future app development.

Adherence to FITT Principles and the Need for Enhanced Program Progression in Exercise Apps

Unlike general physical activity apps that often fail to meet the basic FITT criteria [24,42], the specialized apps in our study successfully incorporated these fundamental elements. However, the primary clinical limitation lies in their failure to satisfy the full FITT-VP framework, specifically regarding structured progression and individualized adjustment. In established exercise prescription practice, progression is essential for achieving physiological adaptation while minimizing injury risk, particularly by avoiding abrupt increases in exercise volume or intensity [43,44]. Clear guidance on how and when to advance exercises is also critical for informed consent and may facilitate patient engagement and adherence.

In addition, limited support for tailoring exercise programs to individual needs constrains the safe use of these tools in populations such as older adults and individuals with chronic diseases, for whom inappropriate exercise loading may increase the risk of adverse events [45]. The absence of explicit information regarding potential exercise-related risks further complicates informed decision-making by both clinicians and patients. To optimize the integration of exercise prescription apps into clinical practice, more comprehensive and transparent intervention descriptions are required to support individualized care and the effective translation of evidence-based exercise prescription into real-world settings.

Intervention Fidelity and Clinical Transparency in App-Based Exercise Prescription

The evaluation of intervention fidelity via the CERT domains revealed a stark disparity between the reporting of static structural elements and dynamic clinical processes. While all apps achieved 100% compliance in describing core components such as materials, provider qualifications, and settings, there was a universal failure (0%) to report motivational strategies, progression logic, or tailoring mechanisms. This deficiency mirrors findings from previous systematic reviews in pregnancy [46] and musculoskeletal rehabilitation [47], which similarly identified a pervasive lack of customization features in mHealth tools. Crucially, the absence of documented decision rules for exercise progression limits the transparency of the intervention. As highlighted by Hansford et al [48] and Conrado Ignacio et al [49], detailed reporting of intervention components—beyond simple dosage parameters—is essential for ensuring reproducibility and clearly defining the functional mechanisms of the prescribed therapy. Consequently, while digital platforms successfully deliver standardized content, the lack of explicit reporting on how exercises are adjusted obscures the clinical reasoning necessary to ensure safety and effectiveness throughout the rehabilitative trajectory.

Furthermore, the total lack of information regarding baseline determination and individual tailoring (0%) highlights a fundamental challenge in digital health. According to updated preparticipation health screening recommendations [50], a comprehensive evaluation of a patient's physical status and functional capacity is the indispensable prerequisite for ensuring safety before initiating any exercise program. The inability of current apps to document how starting levels are calibrated or adjusted to individual constraints suggests a reliance on generic protocols rather than precise clinical titration. Coupled with the complete omission of adverse event reporting, these findings indicate that current apps function as digital exercise libraries rather than sophisticated therapeutic tools, falling short of the safety and individualization standards required for clinical populations.

Enhanced BCTs With Challenges in Addressing Supervision and Adherence in Exercise Apps

The exercise prescription apps reviewed in this study incorporated a higher number of BCTs than reported in previous app evaluations [29,30]; the specific selection of strategies remains highly consistent with the broader mHealth literature [29,51-54]. The predominance of self-monitoring, feedback, goal setting, and action planning in our sample mirrors the most frequently identified features in recent content analyses, confirming that these specialized tools adhere to the standard design patterns observed in general physical activity interventions. However, consistent with established behavior change theory, the mere presence of multiple techniques does not ensure effective behavioral support if their selection and configuration are not aligned with the clinical requirements of long-term exercise adherence [55].

From a clinical perspective, several predefined critical BCTs for sustaining exercise behavior in unsupervised or remote settings were absent or infrequently implemented. These included graded tasks, biofeedback, prompts or cues, and self-reward—mechanisms closely associated with progression, self-regulation, and adaptive feedback over time. Their limited representation suggests that current exercise prescription apps are primarily optimized for exercise initiation and short-term compliance rather than for dynamic adjustment and sustained behavior change, which are central to clinical exercise prescription.

Conversely, several frequently implemented techniques that were not classified as critical were consistently present across all apps, including presentation of a credible source, routine

XSL•FO
RenderX

review of behavioral goals, monitoring without feedback, and repeated behavioral practice. Although these techniques may have limited independent effects on long-term adherence, they may provide complementary clinical value by reinforcing professional authority, standardizing instruction, and supporting structured follow-up [56,57]. Taken together, these findings indicate that while current exercise prescription apps insufficiently implement several clinically critical BCTs, their operational strengths may still augment clinician-guided or hybrid care models, although they may not effectively function as stand-alone behavioral interventions.

Digital Usability in Exercise Apps

In our analysis, the exercise prescription apps achieved an average MARS quality score of 3.66 (SD 0.25), consistent with prior studies by Paganini et al [58] and Simões et al [59], which reported moderate overall quality scores for physical activity apps. Paganini et al [58] found the average quality score to be 3.60, with variations across different subcategories: information averaged 3.24; engagement, 3.19; aesthetics, 3.65; and functionality, the highest at 4.35. Simões et al [59] reported a slightly higher overall MARS score of 3.88, with functionality again scoring the highest at 4.30, followed by aesthetics, information quality, and engagement.

Notably, in our study, the score for information quality was higher at 3.90, whereas the engagement score was considerably lower at 2.50. This discrepancy may reflect the more specific and task-oriented nature of our evaluated apps compared to the broader range of apps included in other studies, which might engage users differently. Despite substantial download figures, the apps we reviewed showed limited evidence of effectiveness, a finding that corroborates previous evaluations [60-62], which consistently highlighted a shortage of evidence-based content and professional guideline adherence in the mHealth landscape.

Clinical Implication

From a clinical perspective, the findings of this study indicate that currently available exercise prescription apps should not be used as stand-alone prescribing tools but rather incorporated into clinician-guided or hybrid care models. Although these apps consistently deliver structured exercise instructions and provide visibility of patient adherence, the absence of explicit progression logic, individualized adjustment, and several clinically critical BCTs limits their capacity to independently support safe and effective exercise prescription. Consequently, clinicians remain essential for determining appropriate starting levels, explaining progression criteria, and managing exercise-related risks, particularly in older adults and individuals with chronic diseases [63].

At the same time, the operational features that are consistently implemented across apps such as standardized exercise demonstrations, routine goal review, adherence tracking, and low cognitive load instruction delivery may still offer complementary clinical value. When used alongside professional judgment, these features can reduce instructional burden, support treatment consistency, and facilitate structured follow-up between clinical encounters. In this context, exercise prescription apps may function as digital extensions of clinician-led care rather than autonomous therapeutic systems.

For developers, these findings underscore the need to move beyond technically feasible features toward closer alignment with clinical exercise prescription standards. Enhancing transparency around progression rules, incorporating mechanisms for individualized adjustment, and prioritizing BCTs associated with long-term adherence may improve the clinical relevance of future applications. Most importantly, adopting a co-design approach—actively involving both clinical practitioners and end users in the development process—is essential to bridge the gap between technical implementation and real-world clinical utility [64,65].

Limitations

This study has several limitations. First, search results were inherently transient and geographically specific due to localized app store algorithms and the rapidly evolving mobile market [66]. While our semantic search strategy targeted popular, clinically relevant apps rather than an exhaustive catalog, the findings represent a 2024 snapshot, and the currency of these results may diminish, as apps undergo frequent updates. Second, while the interrater reliability was high for CERT and BCT assessments, the MARS evaluation yielded relatively lower consistency, reflecting the subjective nature of aesthetic and engagement metrics; however, all discrepancies were resolved through rigorous consensus-based discussion to ensure data validity.

Third, the CERT and BCTTv1 frameworks, originally designed for research interventions, may not fully capture modern app functionalities such as automated feedback, potentially creating a structural mismatch in scoring. Fourth, we identified the presence of BCTs but did not evaluate the quality of their implementation or their actual impact on user behavior. Furthermore, the lack of longitudinal adherence and clinical outcome data limits our assessment of real-world therapeutic impact.

Future Direction

While our study establishes fundamental professional quality and usability benchmarks for exercise prescription apps, it serves primarily as a foundational step mapping the current landscape. Future work should build upon these findings by transitioning from quality assessment to rigorous clinical verification.

Specifically, longitudinal randomized controlled trials are needed to determine if the high-scoring, professionally vetted apps identified in this study translate to better patient adherence and health outcomes compared to traditional methods. Furthermore, considering that the effectiveness of digital exercise prescription systems may vary significantly across conditions, these investigations should target specific clinical cohorts—such as patients with cardiac conditions, neurological disorders (eg, poststroke rehabilitation), or musculoskeletal issues (eg, sarcopenia)—rather than general users.

Crucially, to match the complexity of clinical needs, efficacy trials must move beyond self-reported measures and incorporate objective physiological outcomes, such as laboratory data (eg,

XSL•FO
RenderX

inflammatory markers and glycemic control) and body composition metrics assessed via bioelectrical impedance analysis. Finally, integrating wearable sensors for real-time biofeedback represents a promising frontier for ensuring that these digital prescriptions are delivered safely and effectively.

Conclusions

Using a clinician-centered, multidimensional evaluation framework, this study systematically assessed the clinical usability of popular, no-cost exercise prescription apps. Although these apps consistently met basic structural requirements and demonstrated strong technical functionality, they lacked essential features related to progression, individualization, intervention transparency, and sustained behavior change support. As a result, their capacity to independently deliver clinically appropriate exercise prescriptions remains limited.

Nevertheless, the standardized delivery of exercise instructions, structured translation of prescriptions into actionable tasks, and support for adherence monitoring suggest that these apps may still add value within clinician-guided or hybrid care models. Future development should prioritize a co-design approach involving both clinicians and users to bridge existing gaps. This collaborative strategy is essential to align apps with clinical standards and behavior change principles, thereby enhancing their safety, effectiveness, and clinical relevance.

Funding

The article processing fee was supported by intramural funding from Taipei Veterans General Hospital, Taiwan, and Chang Gung Memorial Hospital, Taiwan. These funding sources had no role in the interpretation of the study results. No external funding was received for this study.

Data Availability

The data underlying the analyses and results reported in this paper are available in the electronic appendix accompanying the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed search strategy and selection protocol.
[DOCX File, 23 KB - mhealth_v14i1e77616_app1.docx]

Multimedia Appendix 2
The behavior change techniques in prior studies coded in the Behavior Change Technique Taxonomy version 1 framework.
[DOCX File, 18 KB - mhealth_v14i1e77616_app2.docx]

Multimedia Appendix 3
The Consensus on Exercise Reporting Template checklist with detailed description across included exercise prescription apps (N=6; 2024).
[DOCX File, 22 KB - mhealth_v14i1e77616_app3.docx]

Multimedia Appendix 4
The behavior change techniques coded in Behavior Change Technique Taxonomy version 1 framework across included exercise prescription apps (N=6; 2024).
[DOCX File, 34 KB - mhealth_v14i1e77616_app4.docx]

Multimedia Appendix 5
The Mobile App Rating Scale scores of included exercise prescription apps on the Android system (N=6; 2024).
[DOCX File, 21 KB - mhealth_v14i1e77616_app5.docx]

Multimedia Appendix 6
The Mobile App Rating Scale scores of included exercise prescription apps on the iOS system (N=6; 2024).
[DOCX File, 21 KB - mhealth_v14i1e77616_app6.docx]

Checklist 1
PRISMA 2020 checklist.
[PDF File, 175 KB - mhealth_v14i1e77616_app7.pdf]

XSL•FO
RenderX

References

1. Luan X, Tian X, Zhang H, et al. Exercise as a prescription for patients with various diseases. J Sport Health Sci 2019 Sep;8(5):422-441. [doi: 10.1016/j.jshs.2019.04.002] [Medline: 31534817]
2. Hayden JA, Ellis J, Ogilvie R, Malmivaara A, van Tulder MW. Exercise therapy for chronic low back pain. Cochrane Database Syst Rev 2021 Sep 28;9(9):CD009790. [doi: 10.1002/14651858.CD009790.pub2] [Medline: 34580864]
3. Cools AM, Dewitte V, Lanszweert F, et al. Rehabilitation of scapular muscle balance: which exercises to prescribe? Am J Sports Med 2007 Oct;35(10):1744-1751. [doi: 10.1177/0363546507303560] [Medline: 17606671]
4. Billinger SA, Arena R, Bernhardt J, et al. Physical activity and exercise recommendations for stroke survivors: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2014 Aug;45(8):2532-2553. [doi: 10.1161/STR.0000000000000022] [Medline: 24846875]
5. Garvey C, Bayles MP, Hamm LF, et al. Pulmonary rehabilitation exercise prescription in chronic obstructive pulmonary disease: review of selected guidelines: an official statement from the American Association of Cardiovascular and Pulmonary Rehabilitation. J Cardiopulm Rehabil Prev 2016;36(2):75-83. [doi: 10.1097/HCR.0000000000000171] [Medline: 26906147]
6. Hansen D, Abreu A, Ambrosetti M, et al. Exercise intensity assessment and prescription in cardiovascular rehabilitation and beyond: why and how: a position statement from the Secondary Prevention and Rehabilitation Section of the European Association of Preventive Cardiology. Eur J Prev Cardiol 2022 Feb 19;29(1):230-245. [doi: 10.1093/eurjpc/zwab007] [Medline: 34077542]
7. Petrella RJ, Koval JJ, Cunningham DA, Paterson DH. Can primary care doctors prescribe exercise to improve fitness? The Step Test Exercise Prescription (STEP) project. Am J Prev Med 2003 May;24(4):316-322. [doi: 10.1016/s0749-3797(03)00022-9] [Medline: 12726869]
8. Booth FW, Roberts CK, Laye MJ. Lack of exercise is a major cause of chronic diseases. Compr Physiol 2012 Apr;2(2):1143-1211. [doi: 10.1002/cphy.c110025] [Medline: 23798298]
9. Moore SC, Lee IM, Weiderpass E, et al. Association of leisure-time physical activity with risk of 26 types of cancer in 1.44 million adults. JAMA Intern Med 2016 Jun 1;176(6):816-825. [doi: 10.1001/jamainternmed.2016.1548] [Medline: 27183032]
10. Garber CE, Blissmer B, Deschenes MR, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc 2011 Jul;43(7):1334-1359. [doi: 10.1249/MSS.0b013e318213fefb] [Medline: 21694556]
11. Hoffmann TC, Maher CG, Briffa T, et al. Prescribing exercise interventions for patients with chronic conditions. CMAJ 2016 Apr 19;188(7):510-518. [doi: 10.1503/cmaj.150684] [Medline: 26976965]
12. Lawford BJ, Delany C, Bennell KL, Hinman RS. "I was really pleasantly surprised": firsthand experience and shifts in physical therapist perceptions of telephone‐delivered exercise therapy for knee osteoarthritis–a qualitative study. Arthritis Care Res (Hoboken) 2019 Apr;71(4):545-557. [doi: 10.1002/acr.23618] [Medline: 29885026]
13. Negrini S, Donzelli S, Negrini A, Negrini A, Romano M, Zaina F. Feasibility and acceptability of telemedicine to substitute outpatient rehabilitation services in the COVID-19 emergency in Italy: an observational everyday clinical-life study. Arch Phys Med Rehabil 2020 Nov;101(11):2027-2032. [doi: 10.1016/j.apmr.2020.08.001] [Medline: 32800748]
14. Tenforde AS, Borgstrom H, Polich G, et al. Outpatient physical, occupational, and speech therapy synchronous telemedicine: a survey study of patient satisfaction with virtual visits during the COVID-19 pandemic. Am J Phys Med Rehabil 2020 Nov;99(11):977-981. [doi: 10.1097/PHM.0000000000001571] [Medline: 32804713]
15. Suso-Martí L, La Touche R, Herranz-Gómez A, Angulo-Díaz-Parreño S, Paris-Alemany A, Cuenca-Martínez F. Effectiveness of telerehabilitation in physical therapist practice: an umbrella and mapping review with meta–meta-analysis. Phys Ther 2021 May 4;101(5):pzab075. [doi: 10.1093/ptj/pzab075] [Medline: 33611598]
16. Bennell KL, Marshall CJ, Dobson F, Kasza J, Lonsdale C, Hinman RS. Does a web-based exercise programming system improve home exercise adherence for people with musculoskeletal conditions? A randomized controlled trial. Am J Phys Med Rehabil 2019 Oct;98(10):850-858. [doi: 10.1097/PHM.0000000000001204] [Medline: 31021823]
17. Kalgotra P, Raja U, Sharda R. Growth in the development of health and fitness mobile apps amid COVID-19 pandemic. Digit Health 2022;8:20552076221129070. [doi: 10.1177/20552076221129070] [Medline: 36211794]
18. Cottrell MA, Russell TG. Telehealth for musculoskeletal physiotherapy. Musculoskelet Sci Pract 2020 Aug;48:102193. [doi: 10.1016/j.msksp.2020.102193] [Medline: 32560876]
19. Thompson WR. Worldwide survey of fitness trends for 2021. ACSMs Health Fit J 2021;25(1):10-19. [doi: 10.1249/FIT.0000000000000631]
20. Thompson WR. Worldwide survey of fitness trends for 2022. ACSMs Health Fit J 2022;26(1):11-20. [doi: 10.1249/FIT.0000000000000732]
21. Bonato M, Marmondi F, Mastropaolo C, et al. A digital platform for home-based exercise prescription for older people with sarcopenia. Sensors (Basel) 2024 Jul 24;24(15):15. [doi: 10.3390/s24154788] [Medline: 39123834]
22. Jo HS, Kim HM, Go CH, Yu HY, Park HK, Han JY. Effectiveness of home-based cardiac rehabilitation with optimized exercise prescriptions using a mobile healthcare app in patients with acute myocardial infarction: a randomized controlled trial. Life (Basel) 2024 Sep 5;14(9):1122. [doi: 10.3390/life14091122] [Medline: 39337905]

XSL•FO
RenderX

23. Espinoza-Bravo C, Arnal-Gómez A, Martínez-Arnau FM, et al. Effectiveness of functional or aerobic exercise combined with breathing techniques in telerehabilitation for patients with long COVID: a randomized controlled trial. Phys Ther 2023 Nov 4;103(11):11. [doi: 10.1093/ptj/pzad118] [Medline: 37658773]
24. Chen S, Wu Y, Bushey EL, Pescatello LS. Evaluation of exercise mobile applications for adults with cardiovascular disease risk factors. J Cardiovasc Dev Dis 2023 Nov 28;10(12):477. [doi: 10.3390/jcdd10120477] [Medline: 38132645]
25. Gell NM, Smith PA, Wingood M. Physical therapist and patient perspectives on mobile technology to support home exercise prescription for people with arthritis: a qualitative study. Cureus 2024 Mar;16(3):e55899. [doi: 10.7759/cureus.55899] [Medline: 38601402]
26. De Santis KK, Jahnel T, Matthias K, Mergenthal L, Al Khayyal H, Zeeb H. Evaluation of digital interventions for physical activity promotion: scoping review. JMIR Public Health Surveill 2022 May 23;8(5):e37820. [doi: 10.2196/37820] [Medline: 35604757]
27. Bearne LM, Sekhon M, Grainger R, et al. Smartphone apps targeting physical activity in people with rheumatoid arthritis: systematic quality appraisal and content analysis. JMIR Mhealth Uhealth 2020 Jul 21;8(7):e18495. [doi: 10.2196/18495] [Medline: 32706727]
28. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046. [doi: 10.2196/17046] [Medline: 32186518]
29. Kebede M, Steenbock B, Helmer SM, Sill J, Möllers T, Pischke CR. Identifying evidence-informed physical activity apps: content analysis. JMIR Mhealth Uhealth 2018 Dec 18;6(12):e10314. [doi: 10.2196/10314] [Medline: 30563810]
30. Bondaronek P, Alkhaldi G, Slee A, Hamilton FL, Murray E. Quality of publicly available physical activity apps: review and content analysis. JMIR Mhealth Uhealth 2018 Mar 21;6(3):e53. [doi: 10.2196/mhealth.9069] [Medline: 29563080]
31. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Int J Surg 2021 Apr;88:105906. [doi: 10.1016/j.ijsu.2021.105906] [Medline: 33789826]
32. Slade SC, Dionne CE, Underwood M, Buchbinder R. Consensus on Exercise Reporting Template (CERT): explanation and elaboration statement. Br J Sports Med 2016 Dec;50(23):1428-1437. [doi: 10.1136/bjsports-2016-096651] [Medline: 27707738]
33. Slade SC, Dionne CE, Underwood M, et al. Consensus on Exercise Reporting Template (CERT): modified Delphi study. Phys Ther 2016 Oct;96(10):1514-1524. [doi: 10.2522/ptj.20150668] [Medline: 27149962]
34. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
35. Dugas M, Gao GG, Agarwal R. Unpacking mHealth interventions: a systematic review of behavior change techniques used in randomized controlled trials assessing mHealth effectiveness. Digit Health 2020;6:2055207620905411. [doi: 10.1177/2055207620905411] [Medline: 32128233]
36. Yang X, Ma L, Zhao X, Kankanhalli A. Factors influencing user's adherence to physical activity applications: a scoping literature review and future directions. Int J Med Inform 2020 Feb;134:104039. [doi: 10.1016/j.ijmedinf.2019.104039] [Medline: 31865054]
37. Howlett N, Trivedi D, Troop NA, Chater AM. Are physical activity interventions for healthy inactive adults effective in promoting behavior change and maintenance, and which behavior change techniques are effective? A systematic review and meta-analysis. Transl Behav Med 2019 Jan 1;9(1):147-157. [doi: 10.1093/tbm/iby010] [Medline: 29506209]
38. Sullivan AN, Lachman ME. Behavior change with fitness technology in sedentary adults: a review of the evidence for increasing physical activity. Front Public Health 2016;4:289. [doi: 10.3389/fpubh.2016.00289] [Medline: 28123997]
39. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
40. Terhorst Y, Philippi P, Sander LB, et al. Validation of the Mobile Application Rating Scale (MARS). PLoS One 2020;15(11):e0241480. [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
41. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994;6(4):284-290. [doi: 10.1037/1040-3590.6.4.284]
42. Elder A, Guillen G, Isip R, Zepeda R, Lewis ZH. A deeper look into exercise intensity tracking through mobile applications: a brief report. Technologies (Basel) 2023;11(3):66. [doi: 10.3390/technologies11030066]
43. Collado-Mateo D, Lavín-Pérez AM, Peñacoba C, et al. Key factors associated with adherence to physical exercise in patients with chronic diseases and older adults: an umbrella review. Int J Environ Res Public Health 2021 Feb 19;18(4):2023. [doi: 10.3390/ijerph18042023] [Medline: 33669679]
44. Lund ML, Tamm M, Br nholm IB. Patients' perceptions of their participation in rehabilitation planning and professionals' view of their strategies to encourage it. Occup Ther Int 2001;8(3):151-167. [doi: 10.1002/oti.143] [Medline: 11823880]
45. Franklin BA, Thompson PD, Al-Zaiti SS, et al. Exercise-related acute cardiovascular events and potential deleterious adaptations following long-term exercise training: placing the risks into perspective–an update: a scientific statement from

XSL•FO
RenderX

the American Heart Association. Circulation 2020 Mar 31;141(13):e705-e736. [doi: 10.1161/CIR.0000000000000749] [Medline: 32100573]
46. Hayman MJ, Alfrey KL, Waters K, et al. Evaluating evidence-based content, features of exercise instruction, and expert involvement in physical activity apps for pregnant women: systematic search and content analysis. JMIR Mhealth Uhealth 2022 Jan 19;10(1):e31607. [doi: 10.2196/31607] [Medline: 35044318]
47. Ryan S, Ní Chasaide N, O' Hanrahan S, Corcoran D, Caulfield B, Argent R. mHealth apps for musculoskeletal rehabilitation: systematic search in app stores and content analysis. JMIR Rehabil Assist Technol 2022 Aug 1;9(3):e34355. [doi: 10.2196/34355] [Medline: 35916688]
48. Hansford HJ, Wewege MA, Cashin AG, et al. If exercise is medicine, why don't we know the dose? An overview of systematic reviews assessing reporting quality of exercise interventions in health and disease. Br J Sports Med 2022 Jun;56(12):692-700. [doi: 10.1136/bjsports-2021-104977] [Medline: 35168956]
49. Conrado Ignacio A, Oliveira NL, Xavier Neves da Silva L, et al. Methodological rigor and quality of reporting of clinical trials published with physical activity interventions: a report from the Strengthening the Evidence in Exercise Sciences Initiative (SEES Initiative). PLoS One 2024;19(8):e0309087. [doi: 10.1371/journal.pone.0309087] [Medline: 39213281]
50. Riebe D, Franklin BA, Thompson PD, et al. Updating ACSM's recommendations for exercise preparticipation health screening. Med Sci Sports Exerc 2015 Nov;47(11):2473-2479. [doi: 10.1249/MSS.0000000000000664] [Medline: 26473759]
51. Patterson K, Davey R, Keegan R, Kunstler B, Woodward A, Freene N. Behaviour change techniques in cardiovascular disease smartphone apps to improve physical activity and sedentary behaviour: systematic review and meta-regression. Int J Behav Nutr Phys Act 2022 Jul 7;19(1):81. [doi: 10.1186/s12966-022-01319-8] [Medline: 35799263]
52. Yang CH, Maher JP, Conroy DE. Implementation of behavior change techniques in mobile applications for physical activity. Am J Prev Med 2015 Apr;48(4):452-455. [doi: 10.1016/j.amepre.2014.10.010] [Medline: 25576494]
53. Middelweerd A, Mollee JS, van der Wal CN, Brug J, Te Velde SJ. Apps to promote physical activity among adults: a review and content analysis. Int J Behav Nutr Phys Act 2014 Jul 25;11:97. [doi: 10.1186/s12966-014-0097-9] [Medline: 25059981]
54. Wang Y, Wang Y, Greene B, Sun L. An analysis and evaluation of quality and behavioral change techniques among physical activity apps in China. Int J Med Inform 2020 Jan;133:104029. [doi: 10.1016/j.ijmedinf.2019.104029] [Medline: 31759245]
55. Schroé H, Van Dyck D, De Paepe A, et al. Which behaviour change techniques are effective to promote physical activity and reduce sedentary behaviour in adults: a factorial randomized trial of an e- and m-health intervention. Int J Behav Nutr Phys Act 2020 Oct 7;17(1):127. [doi: 10.1186/s12966-020-01001-x] [Medline: 33028335]
56. Carraça E, Encantado J, Battista F, et al. Effective behavior change techniques to promote physical activity in adults with overweight or obesity: a systematic review and meta-analysis. Obes Rev 2021 Jul;22 Suppl 4(Suppl 4):e13258. [doi: 10.1111/obr.13258] [Medline: 33949778]
57. Spring B, Champion KE, Acabchuk R, Hennessy EA. Self-regulatory behaviour change techniques in interventions to promote healthy eating, physical activity, or weight loss: a meta-review. Health Psychol Rev 2021 Dec;15(4):508-539. [doi: 10.1080/17437199.2020.1721310] [Medline: 31973666]
58. Paganini S, Terhorst Y, Sander LB, et al. Quality of physical activity apps: systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2021 Jun 9;9(6):e22587. [doi: 10.2196/22587] [Medline: 34106073]
59. Simões P, Silva AG, Amaral J, Queirós A, Rocha NP, Rodrigues M. Features, behavioral change techniques, and quality of the most popular mobile apps to measure physical activity: systematic search in app stores. JMIR Mhealth Uhealth 2018 Oct 26;6(10):e11281. [doi: 10.2196/11281] [Medline: 30368438]
60. Higgins JP. Smartphone applications for patients' health and fitness. Am J Med 2016 Jan;129(1):11-19. [doi: 10.1016/j.amjmed.2015.05.038] [Medline: 26091764]
61. Knight E, Stuckey MI, Prapavessis H, Petrella RJ. Public health guidelines for physical activity: is there an app for that? A review of android and apple app stores. JMIR Mhealth Uhealth 2015 May 21;3(2):e43. [doi: 10.2196/mhealth.4003] [Medline: 25998158]
62. Modave F, Bian J, Leavitt T, Bromwell J, Harris III C, Vincent H. Low quality of free coaching apps with respect to the American College of Sports Medicine guidelines: a review of current mobile apps. JMIR Mhealth Uhealth 2015 Jul 24;3(3):e77. [doi: 10.2196/mhealth.4669] [Medline: 26209109]
63. Soto-Bagaria L, Eis S, Pérez LM, et al. Mobile applications to prescribe physical exercise in frail older adults: review of the available tools in app stores. Age Ageing 2023 Dec 1;52(12):afad227. [doi: 10.1093/ageing/afad227] [Medline: 38157286]
64. Mrklas KJ, Barber T, Campbell-Scherer D, et al. Co-design in the development of a mobile health app for the management of knee osteoarthritis by patients and physicians: qualitative study. JMIR Mhealth Uhealth 2020 Jul 10;8(7):e17893. [doi: 10.2196/17893] [Medline: 32673245]
65. Song T, Yu P, Bliokas V, et al. A clinician-led, experience-based co-design approach for developing mHealth services to support the patient self-management of chronic conditions: development study and design case. JMIR Mhealth Uhealth 2021 Jul 20;9(7):e20650. [doi: 10.2196/20650] [Medline: 34283030]
66. Grundy QH, Wang Z, Bero LA. Challenges in assessing mobile health app quality: a systematic review of prevalent and innovative methods. Am J Prev Med 2016 Dec;51(6):1051-1059. [doi: 10.1016/j.amepre.2016.07.009] [Medline: 27659122]

XSL•FO
RenderX

Abbreviations

ACSM: American College of Sports Medicine
BCT: behavior change technique
BCTTv1: Behavior Change Technique Taxonomy version 1
CERT: Consensus on Exercise Reporting Template
FITT: frequency, intensity, time, and type
FITT-VP: frequency, intensity, time, type, volume, and progression
ICC: intraclass correlation coefficient
MARS: Mobile App Rating Scale
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by G Badicu; submitted 16.May.2025; peer-reviewed by A Woodward, M Mcnarry; revised version received 13.Feb.2026; accepted 14.Feb.2026; published 25.Mar.2026.

Please cite as:
Wu CH, Chang CN, Chang CF, Lin MH, Chen HY, Chen YC
Clinical Usability of Exercise Prescription Apps for Professional Use: Systematic Review and Multidimensional Evaluation
JMIR Mhealth Uhealth 2026;14:e77616
URL: https://mhealth.jmir.org/2026/1/e77616
doi:10.2196/77616

© Cheng-Hao Wu, Che-Ning Chang, Chu-Fang Chang, Ming-Hwai Lin, Hsing-Yu Chen, Yu-Chun Chen. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 25.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobile Apps for Tinnitus: Systematic Search in App Stores and Review of Intervention Components and Behavior Change Techniques

Alina Rinn[1], MSc; Sarah Goetsch[2], MSc; Sandy Hannibal[2], MSc; Dirk Lehr[2*], PhD; Cornelia Weise[3*], PhD

[1]Clinical Psychology and Psychotherapy, Philipps-University of Marburg, Gutenbergstraße 18, Marburg, Germany

[2]Department of Health Psychology and Applied Biological Psychology, Leuphana University Lüneburg, Lüneburg, Germany

[3]Clinical Psychology and Behavioral Health Technology, Dept. of Psychology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

[*]these authors contributed equally

Corresponding Author:
Alina Rinn, MSc
Clinical Psychology and Psychotherapy, Philipps-University of Marburg, Gutenbergstraße 18, Marburg, Germany

Abstract

Background: Previous research suggests that 14.4% of the general population is affected by tinnitus. For some of those affected, the ear noise is bothersome or associated with severe distress. There are various treatment options such as cognitive behavioral therapy (CBT), sound therapy, or hearing aids. In addition to browser-based online interventions, mobile apps have been introduced as novel treatment approaches. Previous studies have identified several apps aimed at supporting users with tinnitus. Yet, knowledge about the content of tinnitus apps is limited.

Objective: This study aimed to provide an overview of apps specifically developed for tinnitus by analyzing general app characteristics, as well as app content, focusing on intervention components and behavior change techniques (BCTs).

Methods: A systematic search using 7 search terms (eg, tinnitus and ear noise) was conducted in the Google Play Store and the Apple App Store. Apps designed specifically for tinnitus and available in German or English met the inclusion criteria. Two independent trained raters assessed general app characteristics (eg, age group and costs) using the app description section of the German version of the Mobile App Rating Scale. In addition, raters analyzed app content using the BCT taxonomy (v1) and a list of typical intervention components in tinnitus treatment. Differences in ratings were discussed, and a third trained rater was consulted if no consensus was reached.

Results: A total of 1198 apps were identified in the systematic search. Of those, 69 apps were included in the final analysis. Fifty-two apps were available for free, 23 of which offered in-app purchases. Among the 17 paid apps, costs ranged between €0.69 (US $0.81) and €450 (US $527) per 12 months. Fifty-eight of 69 apps provided sounds (eg, white noise and nature sounds). Many apps assessed tinnitus characteristics (n=38) and provided information about tinnitus (n=27). The most frequently used BCTs were “instruction on how to perform the behavior” (n=25; eg, audio instructions for relaxation techniques), “feedback on behavior” (n=11), “behavioral practice/rehearsal” (n=11), “information about health consequences” (n=11), “information about emotional consequences” (n=11), and “prompts/cues” (n=11). The number of BCTs implemented varied widely across apps (0 - 18 per app).

Conclusions: Most tinnitus apps offer sound-based interventions (eg, white noise and nature sounds). Notably, CBT elements (eg, cognitive restructuring, attention training, and relaxation training) are implemented less frequently, despite CBT being recommended in tinnitus treatment guidelines. Further research on the efficacy of tinnitus apps is needed. Transparent reporting of intervention techniques may help clarify mechanisms of action and support the replication of effective interventions. Given the large number of readily accessible apps, this study provides an overview relevant to both researchers and health care professionals.

(JMIR Mhealth Uhealth 2026;14:e66151) doi:10.2196/66151

KEYWORDS

tinnitus; smartphone apps; behavior change techniques; intervention components; internet interventions

XSL•FO
RenderX

Introduction

Tinnitus is often described as a hissing, sizzling, or ringing in one or both ears or within the head and is defined as an auditory perception without corresponding external sound [1]. Recent meta-analytic findings point to a pooled prevalence of 14.4% for any tinnitus in adults [2]. For a smaller proportion, the ear noise is bothersome or related to severe distress. For example, 1.2% of the participants in a European Tinnitus Survey reported severe tinnitus [3,4]. Sleep disturbances and concentration difficulties, as well as depression and anxiety, can be associated with the ear noise [5-7].

Both patients seeking treatment and practitioners encounter numerous options with a varying base of evidence (eg, medication, hearing aids, cognitive behavioral therapy, or sound therapy) [1,8]. Digital innovation has paved the way for new methods to deliver treatment by providing internet- and mobile-based interventions (IMIs) [9]. IMIs have the potential to provide access to evidence-based interventions on a larger scale. They may accompany traditional interventions, potentially enhancing their effects or reducing costs [10]. Nevertheless, questions and potential risks arise when implementing IMIs, for example, concerning data security or the efficacy of stand-alone smartphone apps [10-12].

In recent years, especially internet-based interventions that provide cognitive behavioral strategies to people affected by tinnitus have been developed. These interventions target the reduction of tinnitus distress and have shown promising effects in several studies [9,13]. In addition to these web-based interventions, a range of smartphone apps aimed at alleviating burdensome tinnitus are easily accessible on app stores [14,15]. An understanding of what is being offered in these interventions may foster the replication of effective interventions, help identify potential mechanisms of action, and inform future intervention development [16]. However, to date, information and studies on the content and aims of tinnitus apps remain scarce.

A primary approach to providing an overview of app content is to categorize the intervention components used (eg, sounds, relaxation training, and information on tinnitus). Previous reviews of tinnitus apps have mostly focused on providing a brief summary on app content, for example, by sorting apps into categories (eg, “tinnitus education, awareness, and prevention,” “tinnitus relief,” “sound therapy,” or “individual therapy”) [14,15,17]. Two of these reviews additionally provide brief descriptions of what individual apps offer, without necessarily distinguishing the specific intervention components implemented (eg, offering sounds to mask tinnitus, providing individual therapy for tinnitus) [15,17]. Sereda et al [18] provided a detailed description of intervention components and functions of apps. However, the group did not systematically search the app stores but rather analyzed some of the apps that people affected by tinnitus reported using in their daily life [18]. We wanted to add to this line of research by using a systematic approach to differentiate individual intervention components (as described in treatment guidelines or handbooks on tinnitus therapy), thereby providing a detailed overview of the intervention components used.

A second approach for categorizing app content is to identify the behavior change techniques (BCTs) used [19]. Michie et al [19,20] developed a cross-domain taxonomy that includes 93 BCTs (eg, “self-monitoring of behavior”), allocated to 16 categories (eg, “feedback and monitoring”). This approach, rooted in health psychology, may be a useful second perspective to understand what might enhance active tinnitus management by fostering health behavior change. Prior studies on tinnitus have documented the BCTs used in an internet-based intervention by applying the behavior change wheel [21]. Furthermore, BCTs, alongside other measures, were assessed to describe the intervention content of a single internet-based intervention and several self-help interventions [16,22]. The authors concluded that intervention coding fostered the understanding of potential mechanisms of action and may be used to guide future developments [16,21,22]. BCTs have not yet been investigated in tinnitus apps.

Taken together, our study focused on the following research questions:

- How many German and English apps specifically developed for tinnitus are available on the Google Play Store and the Apple App Store?
- Which intervention components are provided in tinnitus apps?
- Which BCTs are used in tinnitus apps?

Findings on the quality and content of a subsample of the apps included in this study, namely 21 German apps, have previously been published in German to contribute to ongoing discussions specific to the German health care system [23]. Based on an updated systematic search, we aim to provide a broad overview of current app content in both English and German apps. To advance tinnitus-specific research and intervention development, and to inform researchers, practitioners, and future tinnitus app development, we focused exclusively on apps explicitly developed for people affected by tinnitus.

Methods

Systematic Search

The project was registered at the Open Science Framework [24]. The German Google Play Store and Apple App Store (iTunes) were searched to identify both paid and free Android and iOS apps. The following search terms, as well as their German equivalents, were used: “tinnitus,” “ear ringing,” “ear noise,” “ear buzzing,” “Ohrenklingeln,” “Ohrgeräusch,” and “Ohrensausen.” The search was performed by one of the authors (AR) from November 2020 to April 2021 and updated in February 2023 (conducted by MK and AR). The search strategy was designed to prioritize sensitivity over specificity to ensure a comprehensive capture of all potentially relevant apps. As our aim was to provide an overview of apps explicitly developed for tinnitus, we established strict inclusion and exclusion criteria detailed below to ensure the quality of the final sample.

Inclusion and Exclusion Criteria

In the first step, identified apps were screened based on information provided in the app store (title, description, user comments, and screenshots). The following inclusion criteria

XSL•FO
RenderX

were used: specifically developed for tinnitus and available in German or English. Apps were excluded if one of the following criteria applied: (1) duplicate; (2) app no longer available on the app store; (3) apps with limited functions with a full version of the app being included; (4) app bundles, that is, a group of apps offered together; and (5) app not developed for smartphones but for tablets. If an app was available on both stores, the iOS version was included. App screening for the initial search was performed by one of the authors (1073 apps, AR). In cases where the app screening was inconclusive, a second rater was involved (91 apps, LL; screening for updated search, 125 apps, by AR and MK). In the second step, the remaining apps were tested for eligibility by AR. For this purpose, apps were downloaded and tested again for inclusion criteria. Apps were excluded if either criteria (1) or (2) or one of the following criteria applied: (6) analysis not feasible due to technical shortcomings, (7) app has no app-specific functionality (eg, solely offers eBooks or articles), (8) app in development or test phase only, (9) no access to the app (eg, no purchase via the app store feasible, no user account provided by the app developer upon request), and (10) app offers no intervention itself but accompanies a therapy via earpiece.

Measures

General Characteristics

General app characteristics were assessed using the app description section of the German version of the Mobile App Rating Scale (MARS-G) [25]. We slightly adapted this section for the purpose of our study. The following factors were assessed: platform, cost, affiliations, age group, category in the app store, technical aspects of the app, embedding into routine care, type of use, type of treatment, support, certification, data safety, and barrier-free.

Intervention Components

To categorize app content, a list of typical intervention components in tinnitus treatment was developed by the authors (AR, DL, and CW). This list was built based on studies on the components of interventions for tinnitus [26], the German treatment guideline for chronic tinnitus [27], reviews on tinnitus treatment [9,28,29], as well as a handbook on cognitive behavioral therapy for tinnitus [30]. Furthermore, the evidence-based health information guideline was used [31]. The guideline, among other aspects, names criteria on how patients should be informed prior to the start of an intervention. The component "information on the intervention" was developed to capture several of the criteria described. For the list of typical intervention components in tinnitus treatment used to categorize app content, see Multimedia Appendix 1. Moreover, raters analyzed which of the identified intervention components were a key component in the app's concept (ie, important messages or key skills that are trained when using the app).

Behavior Change Techniques

To assess which techniques were used to encourage changes in behavior, the BCT taxonomy (v1) [19] was used. This taxonomy includes 93 techniques that might be used in behavior change interventions (eg, "self-monitoring of behavior" or "action planning"). These are allocated to 16 categories (eg, "feedback and monitoring" or "goals and planning"). For each app, raters analyzed which of the 93 techniques described in the BCT taxonomy (v1) were implemented.

Rater Training

Five raters with a bachelor's or master's degree in psychology were trained prior to the app ratings. The training consisted of several steps. First, the raters watched a video introducing the German version of the Mobile App Rating Scale rating, as provided by the authors [25]. As this video contains an example of how to rate a tinnitus app, this specific part was excluded, and instead, the raters viewed parts of the training video created by the authors of the original scale (which included an example of how to rate an app) [32]. Second, the raters used exercises offered on the BCT Taxonomy v1 online platform to train categorizing BCTs (all raters: "Practices"; rater 1: online-training) [33]. Third, to become familiar with the taxonomy, the raters developed and jointly discussed examples of how each BCT might be implemented in a tinnitus app. As a final step, the raters independently analyzed 5 apps developed for pain management and subsequently discussed their results until agreement was reached.

Procedure and Data Analysis

Two raters independently tested each app for at least 20 minutes and then proceeded to rate the app. The full versions of the apps (including all app features) were tested. Each rater was provided with a list of apps, which had to be rated in descending order. The order specified differed between raters to minimize potential order effects. Once all apps were analyzed, raters discussed the results based on differences in the ratings of general characteristics, intervention components, and BCTs. If no consensus was reached, a third rater participated in the discussion. Descriptive statistics (eg, frequencies of individual BCTs) were calculated using IBM SPSS Statistics version 28 [34].

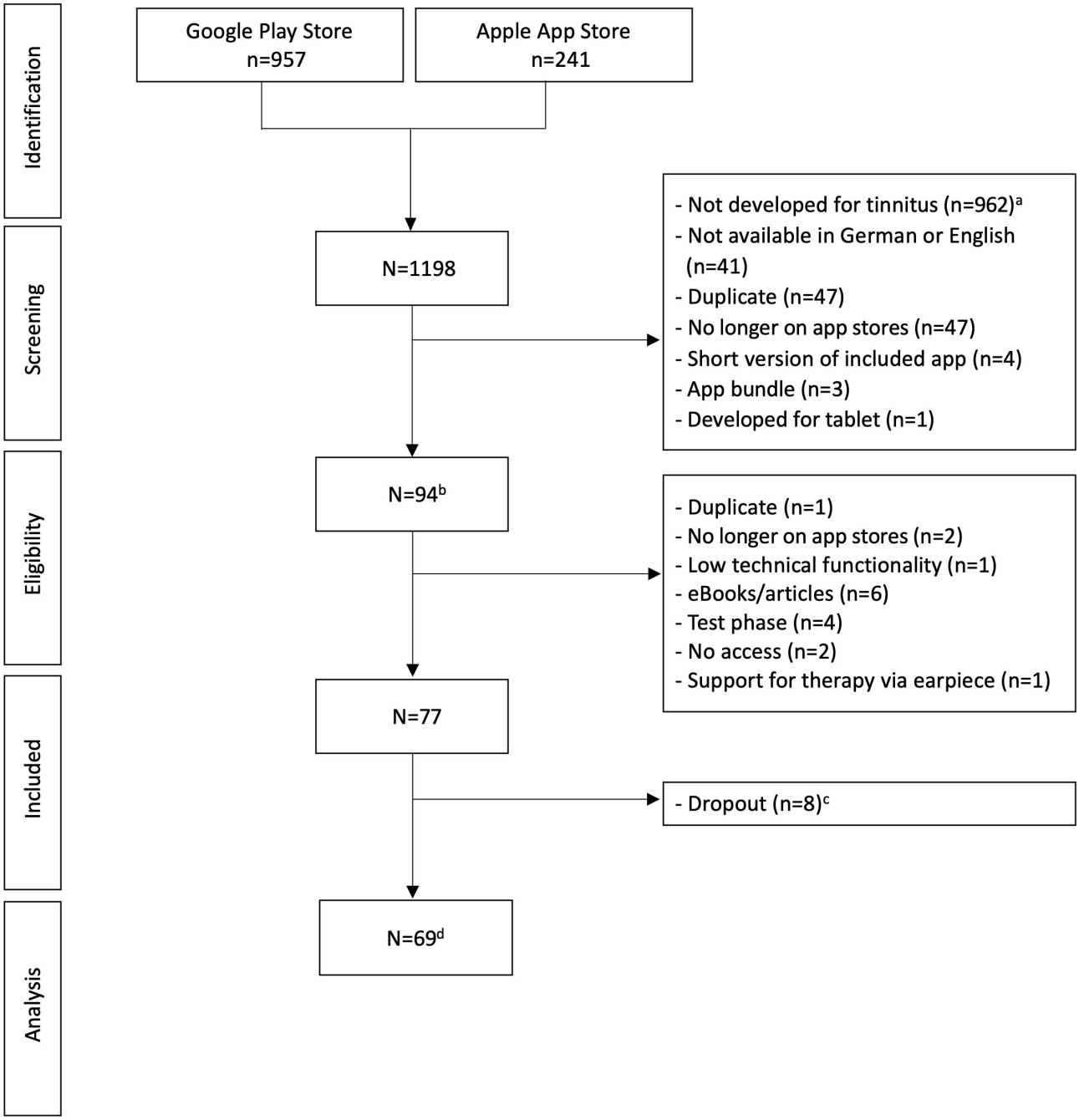
Results

Systematic Search

The initial search using the 7 search terms mentioned above yielded 1198 apps on the app stores (Google Play Store: n=957; Apple App Store: n=241). Of these, 1105 apps were excluded during screening. Most of the apps (n=962) were excluded as they were not explicitly developed for tinnitus (n=809 addressed topics other than tinnitus, eg, decibel meters; n=153 apps mentioned tinnitus but were not explicitly developed for it, eg, apps accompanying hearing aid devices; Figure 1). The remaining 94 apps were downloaded and tested for eligibility. Of these, 17 apps were excluded (eg, due to low technical functionality), leaving 77 apps that fulfilled all inclusion criteria. During the app testing, a further 8 apps had to be excluded (eg, no longer available on the app stores). Accordingly, 69 apps were included in the final analysis. Figure 1 displays the flowchart of the systematic search and inclusion process for tinnitus apps, and the caption provides further detailed information on the selection process. The list of all apps included in the final analysis can be found in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 1. Systematic search and inclusion of tinnitus apps. [a]809 apps were not developed for tinnitus, for example, apps offering a decibel meter; 153 apps did mention tinnitus but were not explicitly designed for tinnitus: 115 apps offering sounds for sleep, to foster concentration, to accompany meditation or yoga praxis, to reduce stress, for tinnitus; 17 apps for hearing aid devices offering tinnitus programs; 21 apps mentioning tinnitus in the app title or description (eg, information on different diseases of the ear, nose, or throat; frequency generator; and listening to a tinnitus tone). [b]Inclusion of the full version of an app following a suggestion of the app developer. [c]Six apps no longer on app stores, 1 app with low technical functionality, 1 link to website, and 1 duplicate. [d]One app was no longer available on the Apple App Store, but still offered on the Google Play Store, so the Android version was tested instead; one app was tested with limited functions (no full version or access to the app store description available).



App Characteristics

A total of 45 iOS and 24 Android apps were analyzed. Fifty-two apps were free, with 23 of them offering in-app purchases. The price for 17 paid apps ranged between €0.69 (US $0.81) and €450 (US $527) per 12 months. The majority of tinnitus apps were offered in the app store categories medical (n=29 apps) or health and fitness (n=28 apps). For 29 apps, we were not able to identify the developers or providers. Sixty-six apps offered a treatment, with 1 app additionally including content for prevention. One app aimed to track symptoms and 1 app offered an assessment of the tinnitus tone. Three apps had an app community (eg, motivational notifications can be sent to other users), and 3 apps offered an exchange with others via social media or messenger apps. Almost all apps were not by design embedded into routine care (n=63). Only 3 apps allowed for the sharing of information (eg, user statistics) with a health care professional. Few apps provided support for users (n=8), most of them regarding potential technical issues (n=6). For a full overview of app characteristics, see Table 1.

XSL•FO
RenderX

Table . App characteristics for all apps included[a].

Characteristics	Values, n (%)
Platform	
iOS	45 (65.2)
Android	24 (34.8)
Cost[b]	
Free	52 (75.4)
In-app purchases available (range: €109 [US $1.28]-€6799 [US $79.59]/mo)	23 (33.3)
Paid (range: €0.69 [US $0.81]-€450 [US $527]/12 mo)	17 (24.6)
Affiliations	
Unknown	29 (42)
Commercial	30 (43.5)
Nonprofit organization	1 (<0.1)
Hospital	1 (<0.1)
Other: private individual	5 (0.1)
Other: foundation	3 (<0.1)
Age group	
18+ y	1 (<0.1)
USK[c] rating	24 (34.8)
Age unspecified	1 (<0.1)
Other: age rating Apple	44 (63.8)
Category in the app store[d]	
Lifestyle	1 (<0.1)
Medical	29 (42.0)
Health and fitness[e]	28 (40.6)
Other: music and audio	5 (0.1)
Other: social networking	1 (<0.1)
Other: entertainment	1 (<0.1)
Other: tools	1 (<0.1)
Other: utilities	2 (<0.1)
Technical aspects of app	
Exchange with others (eg, social media)	3 (<0.1)
Has an app community	3 (<0.1)
Needs web access to function	22 (31.9)
Other: headphones needed	1 (<0.1)
Embedding into routine care	
None	63 (91.3)
Sharing of content (eg, user statistics) with health care professional	3 (<0.1)
Other[f]	4 (0.1)
Type of use	
Prevention	1 (<0.1)
Treatment	66 (95.7)

XSL•FO
RenderX

Characteristics	Values, n (%)
Unspecified	1 (<0.1)
Other[g]	3 (<0.1)
Type of treatment	
Stand-alone	69 (100)
Support	
Guided synchron	1 (<0.1)
Guided asynchron	2 (<0.1)
Technically guided	6 (0.1)
Unguided	61 (88.4)
Certification	
None	61 (88.4)[h]
Accordance with German medical device law	4 (0.1)
Other: CE[i] marked	6 (0.1)
Other: certified in accordance with 93/42/EEC for Medical Devices	1 (<0.1)
Data safety	
Allows password protection	13 (18.8)
Requires login	13 (18.8)
Has a privacy policy	53 (76.8)
Barrier-free	
No	69 (100)

[a]Characteristics of 69 apps as assessed by the app description section Mobile App Rating Scale–German (adapted).

[b]n=2 cost coverage by German health insurance company.

[c]USK: German Entertainment Software Self-Regulation Body.

[d]For 1 app, categories were not assessed, as the app store description was no longer available.

[e]One app listed “mental well-being” as a second app category.

[f]n=2 usage primarily within the context of study participation; n=2 further guidance, customized therapy, or online training are offered on the app`s website.

[g]n=1 tracking symptoms, n=1 assessment of tinnitus tone, and n=1 additionally assessing sleep quality and problematic internet use.

[h]n=1 developed in accordance with norms IEC 62304 and IEC 82304 as well as the “Good Automated Manufacturing Practice 5” (GAMP 5).

[i]CE: Conformité Européenne.

Intervention Components

Intervention components were assessed using a list of components in tinnitus treatment developed by the authors. Most frequently used intervention components were sounds (eg, white noise, nature sounds, and music), which were implemented in 58 of 69 apps. In 52 apps, sounds were a key element. An assessment of tinnitus characteristics was offered by 38 apps (key element in 22 apps). Furthermore, information on tinnitus (n=27, key element in 16 apps) and the provision of information on, or assessment of risk factors for tinnitus (n=23) were implemented in several apps. Less than a third of the apps assessed tinnitus distress (n=18) or provided information on or assessed further otological symptoms (n=16). Further intervention components typically used in CBT were implemented in a smaller number of apps: relaxation training (n=16), techniques to improve sleep (n=12), identification of maladaptive thoughts (n=8), restructuring maladaptive thoughts (n=8), attention training (n=4), and exposure to tinnitus (n=1). Besides, a larger part of the apps (n=39) informed users about the aim of the intervention, while only a few apps offered information on uncertainties and missing evidence (n=7). Across all apps, we identified 41 different intervention components. Apps highly differed in the number of intervention components they included, ranging from 2 to 20 intervention components per app. For full details on the identified intervention components, see Table 2.

XSL•FO
RenderX

Table . Intervention components used in 69 analyzed apps.[a]

Intervention component	Implemented in n apps	Key element in n apps
Assessing tinnitus characteristics	38	22
Assessing tinnitus distress	18	4
Analysis of risk factors for tinnitus	23	1
Provision of information regarding risk factors	21	
Assessment of risk factors	4	
Analysis of factors influencing the development and maintenance of tinnitus distress	16	1
Provision of information regarding influencing factors	14	
Assessment of influencing factors	4	
Information on or examination of further otological symptoms (eg, hearing loss and hyperacusis)	16	—[b]
Information on the intervention regarding[c]		
Aim of the intervention	39	—
Treatment options (including nonintervention)	1	—
Information on uncertainties and missing evidence	7	—
Probabilities of success, failure, and side effects of the intervention and of treatment options	1[d]	—
Costs	69	—
Information on tinnitus or psychoeducation	27	16
Relaxation training	16	9
Mindfulness exercises	12	3
Identification of maladaptive thoughts	8	3
Restructuring of maladaptive thoughts	8	3
Attention training	4	1
Exposure to tinnitus	1	—
Emotional regulation	1	1
Social support (eg, contact with others)	5	—
Hearing strategies for hearing impairment	2	—
Techniques to improve sleep	12	2
Techniques to improve concentration and work efficacy	2	—
Sounds (eg, white noise, nature sounds, and music)	58	52
Auditory discrimination training	1	1
Tailor-made notched music therapy	10	9
Acoustic neuromodulation	5	5
Relapse prevention	2	—
Other [e]		
Behavioral activation	4	—
Resource-oriented or acceptance-based exercises[f]	4	1
Exercises focusing on residual inhibition	3	—

XSL•FO
RenderX

Intervention component	Implemented in n apps	Key element in n apps
Tracking of tinnitus and other symptoms (eg, mood, stress, concentration, or sleep)	2	2
Exercises to improve auditory perception and directional hearing	2	1

[a]Intervention components were assessed using a list of typical intervention components in tinnitus treatment developed by the authors (AR, DL, and CW; Multimedia Appendix 1). Components identified in at least 1 app are depicted.

[b]Not applicable.

[c]Based on the guideline for evidence-based health information, namely requirements regarding the content of information for patients about a treatment [31].

[d]Developed by app community, no information provided by the app developers.

[e]For details on further intervention components identified in only one app, see Multimedia Appendix 1.

[f]For example, focus on strengths and successes, gratitude exercises, assessment of important motives and needs—setting life goals.

Behavior Change Techniques

BCTs were categorized using the taxonomy by Michie et al [19]. In total, 36 out of 93 BCTs were identified in at least 1 app. The most frequently implemented BCTs were "instruction on how to perform the behavior" (n=25, eg, audio instructions on how to perform a relaxation technique), "feedback on behavior" (n=11), "behavioral practice/rehearsal" (n=11, eg, prompting users to practice mindfulness-based exercises), "information about health consequences" (n=11), "information about emotional consequences" (n=11), and "prompts/cues" (n=11, eg, app reminders to engage in pleasant activities). Apps highly differed in the number of implemented BCTs (none to 18 BCTs per app), with 32 apps including no BCTs at all. See Table 3 for a full overview of BCTs identified.

XSL•FO
RenderX

Table . Behavior change techniques (BCTs) identified in 69 analyzed apps[a].

Category and BCT	Implemented in n apps
1. Goals and planning	
1.1 Goal setting (behavior)	5
1.2 Problem solving	2
1.3 Goal setting (outcome)	2
1.4 Action planning	1
1.6 Discrepancy between current behavior and goal	1
1.9 Commitment	1
2. Feedback and monitoring	
2.2 Feedback on behavior	11
2.3 Self-monitoring of behavior	2
2.4 Self-monitoring of outcome of behavior	7
2.7 Feedback on outcome of behavior	3
3. Social support	
3.1 Social support (unspecified)	4
4. Shaping knowledge	
4.1 Instruction on how to perform the behavior	25
4.3 Reattribution	1
5. Natural consequences	
5.1. Information about health consequences	11
5.4 Monitoring of emotional consequences	1
5.6 Information about emotional consequences	11
6. Comparison of behavior	
6.1 Demonstration of the behavior	3
7. Associations	
7.1 Prompts/cues	11
7.7 Exposure	1
7.8 Associative learning	1
8. Repetition and substitution	
8.1 Behavioral practice or rehearsal	11
8.3 Habit formation	2
8.7 Graded tasks	1
9. Comparison of outcomes	
9.1 Credible source	3
10. Reward and threat	
10.3 Nonspecific reward	5
10.4 Social reward	1
10.6 Nonspecific incentive	1
10.9 Self-reward	1
11. Regulation	
11.1 Pharmacological support	1
11.2 Reduce negative emotions	5
12. Antecedents	

XSL•FO
RenderX

Category and BCT	Implemented in n apps
12.1 Restructuring the physical environment	2
13. Identity	
13.2 Framing or reframing	8
13.4 Valued self-identity	2
14. Scheduled consequences	—
15. Self-belief	
15.1 Verbal persuasion about capability	1
15.3 Focus on past success	1
15.4 Self-talk	1
16. Covert learning	—

[a]BCTs were assessed using the behavior change technique taxonomy v1 [19]. Only those BCTs that were identified at least once are displayed.

Discussion

Principal Findings

In this systematic review, we assessed intervention components and BCTs [19] used in tinnitus apps and provided a detailed summary of the content of available apps. We analyzed a total of 69 apps available in English or German, all of which were easily accessible via 2 major app stores.

Within the 69 apps included in our study, numerous intervention components were identified (eg, relaxation, assessment of tinnitus distress, or techniques to improve sleep). The most frequently implemented intervention component was sounds (eg, white noise and nature sounds). Furthermore, many apps offered an assessment of tinnitus characteristics or tinnitus distress and provided information on tinnitus, risk factors, and the basics of the intervention offered (eg, its aim). One possible reason for focusing on these components is that they are relatively easy to implement. Our findings are consistent with previous research reporting the frequent use of sound therapy in tinnitus apps [14,15] and patients' preference for such apps [18]. In a recent health service evaluation, sound therapy (including hearing aids) was found to be the most frequently available intervention in both northern and southern Europe [35]. At the same time, clinical guidelines differ on whether they recommend sound therapy, given the lack of evidence or the risk of bias in previous studies [8,36]. Although meta-analytic evidence clearly supports the efficacy of CBT for treating tinnitus distress [8,36,37], CBT intervention components (eg, cognitive restructuring, relaxation training, and attention shifting) were implemented in only a small number of apps. In contrast, when analyzing CBT apps for depression, Martinengo and colleagues identified a variety of apps offering, for example, cognitive restructuring [38]. Thus, it seems possible to realize even more complex CBT components in smartphone apps. Interestingly, a closer look at the intervention components revealed considerable differences in how certain components were implemented. For example, 1 app addressed typical maladaptive thoughts by naming several maladaptive thoughts, briefly elaborating on their potential negative effects, and suggesting alternative thoughts. In another app, users were asked to name individual debilitating thoughts and were then provided with techniques to develop a more helpful thought.

Our second aim was to assess the BCTs used in available tinnitus apps, as up-to-date knowledge on this topic is limited. In our study, we identified 36 of the 93 BCTs defined by Michie et al [19]. The 6 most frequently identified BCTs were "instruction of how to perform the behavior," "information about emotional consequences," "information about health consequences," "feedback on behavior," "behavioral practice/rehearsal," and "prompts/cues." Only a few other studies have assessed BCTs used in tinnitus interventions. Greenwell et al [16] analyzed BCTs in 5 studies on self-help interventions for tinnitus, and 2 studies reported on BCTs used in individual online interventions [21,22]. The techniques most frequently identified in our study are likely common because they are straightforward to implement in an app. For example, to promote relaxation behavior, users were instructed on how to perform a breathing exercise ("instruction on how to perform a behavior"), motivated to practice relaxation with audio recordings ("behavioral practice/ rehearsal"), or sent reminders to complete an exercise ("prompts/cues"). Interestingly, social support techniques were identified in only 4 out of 69 apps. This aligns with Greenwell et al [16] findings that social support components were largely absent in self-help interventions for tinnitus. In previous research on tinnitus support groups and group therapy, features such as the exchange of information, a sense of belonging, and validation of the tinnitus experience were identified as central helpful elements [39,40]. Including peer support in apps could therefore benefit tinnitus users, and smartphone apps offer various ways to provide this digitally. Moreover, goal setting was implemented in only a few apps. From a clinical perspective, prompting users to set goals and review them ("goal setting behavior," "goal setting outcome," "review behavior goals," and "review outcome goals") could enhance motivation. Reviewing these goals, according to Michie et al [19], may lead to adapting goals or modifying the behavior change strategy based on prior achievements. This process would allow users to evaluate whether they have met their goals using the app or if adjustments to the goal are needed. Additionally, normalizing the adaptation of goals or strategies may further support user motivation.

XSL•FO
RenderX

In addition to numerous apps focusing on the provision of sounds and a small number of apps offering various CBT strategies, we identified apps offering, for example, acoustic neuromodulation or tailor-made notched music therapy. This highlights the general problem that tinnitus sufferers searching for an app are confronted with treatment approaches supported by varying levels of evidence [29,37,41]. It is likely that most app users are unaware of this. Furthermore, while a few tinnitus apps have been tested for efficacy, many apps lack scientific evaluation [42-47]. This aligns with our findings from a subsample of the apps included in this study [23]. In the current study, we also found only a few apps providing information on uncertainties or missing evidence regarding the intervention provided. As a basic requirement, app developers should inform users about the evidence base of the treatment approach and whether the app has been scientifically evaluated. Additionally, information on alternative treatment approaches should be provided to enable informed decision-making. This would also provide guidance to users if they are unable to reach their goals using the app. Similarly, Sander et al [48], in their review on posttraumatic stress disorder apps, proposed that apps categorized as medical or health apps should inform users that the app is not a replacement for standard treatment and provide guidance on accessing other treatment options. Correspondingly, other groups have called for evidence-based information in medical and health-related apps [49,50]. In our subsample analysis, we identified a few intervention components that could have negative effects if used without guidance from a clinical expert (eg, insufficient instructions for a training designed for users with hyperacusis) [23]. This is consistent with previous research highlighting potential negative effects when certain exercises are used without support [18]. Remarkably, none of the apps were categorized as "barrier-free." However, we only screened for app features aimed at reducing barriers using a single yes or no item. In light of our findings and the large number of available tinnitus apps, health care professionals should ask patients whether they are using tinnitus apps and evaluate app content together. This may help patients assess whether the interventions offered are evidence based. In addition, potential shortcomings and negative effects can be identified, and professionals can provide guidance on the proper use of certain strategies. This aligns with recommendations by Brown et al [50], who call for increased awareness of potential negative effects among users and health care professionals. Assessing apps used in daily life also offers the opportunity to integrate exercises into a broader treatment plan. For example, a patient may use relaxation training offered in an app but struggle to practice it regularly. In this case, a health care professional can help identify barriers and develop strategies to overcome them.

This review provided several key insights into the current landscape of tinnitus apps. Using a broad and systematic search strategy, we identified a large number of tinnitus-specific apps and evaluated each app in detail, rather than focusing only on the most popular ones. This approach allowed us to provide a comprehensive overview of the intervention components and BCTs currently implemented in tinnitus apps. A key strength of this review is the high sensitivity of the initial search. By using a broad search strategy, we accounted for nonclinical terminology often used by developers in app store metadata. During the screening phase, a significant number of hits had to be excluded because the apps were not related to tinnitus (eg, offering a decibel meter). Probably due to the app store search algorithms using a multitude of criteria (eg, screening title, description, and keywords), 153 apps (eg, accompanying hearing aid devices or generic sound or relaxation apps that lacked tinnitus-specific content) were further retrieved. While these apps might have used "tinnitus relief" as a keyword in their metadata to, for example, increase visibility, they were not specifically designed for tinnitus and were, therefore, excluded from the final analysis. Although the broad search strategy resulted in a high exclusion rate of apps, it ensured that no dedicated tinnitus-specific interventions were overlooked, providing a more reliable overview of the current digital health landscape for tinnitus.

Our findings show that sounds were by far the most frequently implemented intervention component, whereas evidence-based CBT components were present in only a small proportion of apps, despite strong evidence for CBT in the treatment of tinnitus distress. Furthermore, only 36 of the 93 BCTs defined by Michie et al [19] were identified across all apps, indicating a limited use of established BCTs. By systematically mapping these components, we highlight which BCTs are currently under-represented and discuss their potential relevance for future app development. In addition, our results underscore the lack of transparent information on the evidence base of the treatment approaches offered in many apps and the need for scientific evaluation. Regular assessment of tinnitus apps by health care professionals may help identify potential shortcomings, support informed use of apps, and facilitate their integration into a broader treatment plan.

Limitations and Future Directions

Besides these strengths, several limitations should be considered. First, the systematic search was last updated in 2023. Accordingly, newly developed apps may not be included in the analysis. Furthermore, some of the included apps may no longer be available on the app stores, or their content might have changed substantially since the time of data extraction, limiting the temporal validity of our findings. Moreover, more recently developed apps might include more complex CBT components or a greater variety of BCTs. We searched 2 major app stores (Google Play Store and Apple App Store) but did not search app libraries (eg, ORCHA App Library) or other app stores (eg, Windows Store). Thus, the coverage of available apps may be incomplete. It is also possible that additional tinnitus apps were unavailable on the German app store but accessible to users in other countries. Second, we only analyzed apps available in English or German. Therefore, our findings may not be generalizable to apps in other languages and potential cultural or regional differences in app design and content are not captured. Third, as we used tinnitus-specific search terms and focused solely on apps explicitly developed for tinnitus, this review does not cover other apps that might potentially be beneficial for people suffering from tinnitus. In particular, transdiagnostic or mental health apps (eg, targeting stress or sleep) that may indirectly address tinnitus-related distress were not included, which may limit the scope of the review. Fourth,

XSL•FO
RenderX

although we provide an overview of intervention components currently used in apps, no conclusions about the quality of their digital implementation can be drawn based on our data. Accordingly, we did not assess the mode of delivery for individual BCTs or how often they were used in each app. Our coding was based on the presence or absence of components rather than their depth or user engagement, which may vary substantially between apps. Fifth, based on our results, we cannot investigate associations between the implementation of a particular BCT or intervention component and the efficacy of an app. We did not incorporate user-level data (eg, adherence, engagement, or outcomes), which limits conclusions regarding the practical effectiveness of the apps and the identified components.

Future research is needed to determine which intervention components, BCTs, or combination of BCTs are most effective in promoting behavior change and reducing tinnitus-related distress, ideally by linking specific components to user engagement and clinical outcomes. Moreover, subsequent studies may clarify how complex CBT components, such as cognitive restructuring, can be realized in an app to be beneficial for the user. Such research should also include questions of usability, personalization, and appropriate guidance. Future research should also examine how intervention components need to be implemented to be used safely without professional supervision, and under which conditions additional human support may be required to ensure both safety and effectiveness.

Conclusions

In this review, we provide a detailed and systematic analysis of tinnitus app content by categorizing intervention components and BCTs. By applying an established BCT framework, this study offers a structured and theory-informed overview of current app-based interventions. A variety of intervention components were identified at least once (eg, relaxation training, sound therapy, and information on tinnitus). However, the component most frequently implemented was sounds (eg, white noise and nature sounds), highlighting a discrepancy between commonly implemented features and evidence-based treatment recommendations. Core CBT elements, which are recommended in tinnitus treatment guidelines, were offered less frequently. Patients with tinnitus searching for support in app stores are confronted with a large number of easily accessible apps, yet the content and theoretical grounding of these apps vary substantially. In light of the limited scientific evaluation of these apps, future research on their efficacy as well as potential negative effects is highly needed. Importantly, this review provides a transparent and reproducible foundation for such evaluations by systematically mapping existing intervention components and BCTs. Overall, this study provides an overview for research and practice by summarizing current tinnitus app content and identifying key gaps that can inform evidence-based app development and future research priorities.

Acknowledgments

The authors thank Marc Philipp Scheib and Sophie Helene Eg Pakleppa for their help in data collection, and Malin Keller as well as Lea Liebchen for support during the conduct of this study. No artificial intelligence tools were used for the purposes of information gathering, analyses, content creation, manuscript writing, and/or editing.

Funding

Open Access funding provided by the Open Access Publishing Fund of Philipps-Universität Marburg.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: AR (lead), DL (equal), and CW (equal)
Formal analysis: AR
Investigation: AR (lead), SG (equal), SH (equal), MPS (supporting), SHEP (supporting), MK (supporting), and LL (supporting)
Project administration: AR (lead), CW (equal), and DL (supporting)
Supervision: CW (lead) and DL (equal)
Writing—original draft: AR
Writing—review and editing: CW (lead), DL (equal), SH (supporting), and SG (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of typical intervention components in tinnitus treatment, list of apps analyzed and further intervention components identified.
[DOCX File, 20 KB - mhealth_v14i1e66151_app1.docx]

XSL•FO
RenderX

References

1. Baguley D, McFerran D, Hall D. Tinnitus. The Lancet 2013 Nov;382(9904):1600-1607. [doi: 10.1016/S0140-6736(13)60142-7]
2. Jarach CM, Lugo A, Scala M, et al. Global prevalence and incidence of tinnitus: a systematic review and meta-analysis. JAMA Neurol 2022 Sep 1;79(9):888-900. [doi: 10.1001/jamaneurol.2022.2189] [Medline: 35939312]
3. Biswas R, Lugo A, Akeroyd MA, Schlee W, Gallus S, Hall DA. Tinnitus prevalence in Europe: a multi-country cross-sectional population study. Lancet Reg Health Eur 2022 Jan;12:100250. [doi: 10.1016/j.lanepe.2021.100250] [Medline: 34950918]
4. Gallus S, Lugo A, Garavello W, et al. Prevalence and determinants of tinnitus in the Italian adult population. Neuroepidemiology 2015;45(1):12-19. [doi: 10.1159/000431376] [Medline: 26182874]
5. Bhatt JM, Bhattacharyya N, Lin HW. Relationships between tinnitus and the prevalence of anxiety and depression. Laryngoscope 2017 Feb;127(2):466-469. [doi: 10.1002/lary.26107] [Medline: 27301552]
6. Manchaiah V, Beukes EW, Granberg S, et al. Problems and life effects experienced by tinnitus research study volunteers: an exploratory study using the ICF classification. J Am Acad Audiol 2018;29(10):936-947. [doi: 10.3766/jaaa.17094] [Medline: 30479266]
7. McCormack A, Edmondson-Jones M, Fortnum H, et al. Investigating the association between tinnitus severity and symptoms of depression and anxiety, while controlling for neuroticism, in a large middle-aged UK population. Int J Audiol 2015;54(9):599-604. [doi: 10.3109/14992027.2015.1014577] [Medline: 25766493]
8. Fuller TE, Haider HF, Kikidis D, et al. Different teams, same conclusions? A systematic review of existing clinical guidelines for the assessment and treatment of tinnitus in adults. Front Psychol 2017;8:206. [doi: 10.3389/fpsyg.2017.00206] [Medline: 28275357]
9. Andersson G. Internet-delivered psychological treatments for tinnitus: a brief historical review. Am J Audiol 2022 Sep 21;31(3S):1013-1018. [doi: 10.1044/2022_AJA-21-00245] [Medline: 35442718]
10. Ebert DD, Van Daele T, Nordgreen T, et al. Internet- and mobile-based psychological interventions: applications, efficacy, and potential for improving mental health. Eur Psychol 2018 May;23(2):167-187. [doi: 10.1027/1016-9040/a000318]
11. Grundy Q, Chiu K, Held F, Continella A, Bero L, Holz R. Data sharing practices of medicines related apps and the mobile ecosystem: traffic, content, and network analysis. BMJ 2019 Mar 20;364:l920. [doi: 10.1136/bmj.l920] [Medline: 30894349]
12. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019;2(1):118. [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
13. Beukes EW, Manchaiah V, Allen PM, Baguley DM, Andersson G. Internet-based interventions for adults with hearing loss, tinnitus, and vestibular disorders: a systematic review and meta-analysis. Trends Hear 2019;23:2331216519851749. [doi: 10.1177/2331216519851749] [Medline: 31328660]
14. Deshpande AK, Shimunova T. A comprehensive evaluation of tinnitus apps. Am J Audiol 2019 Sep 13;28(3):605-616. [doi: 10.1044/2019_AJA-18-0135] [Medline: 31330114]
15. Mehdi M, Stach M, Riha C, et al. Smartphone and mobile health apps for tinnitus: systematic identification, analysis, and assessment. JMIR Mhealth Uhealth 2020 Aug 18;8(8):e21767. [doi: 10.2196/21767] [Medline: 32808939]
16. Greenwell K, Sereda M, Coulson N, El Refaie A, Hoare DJ. A systematic review of techniques and effects of self-help interventions for tinnitus: application of taxonomies from health psychology. Int J Audiol 2016 Jul;55(sup3):S79-S89. [doi: 10.3109/14992027.2015.1137363] [Medline: 27146036]
17. Mehdi M, Riha C, Neff P, et al. Smartphone apps in the context of tinnitus: systematic review. Sensors (Basel) 2020 Mar 19;20(6):1725. [doi: 10.3390/s20061725] [Medline: 32204540]
18. Sereda M, Smith S, Newton K, Stockdale D. Mobile apps for management of tinnitus: users' survey, quality assessment, and content analysis. JMIR Mhealth Uhealth 2019 Jan 23;7(1):e10353. [doi: 10.2196/10353] [Medline: 30672739]
19. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
20. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: The development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). Health Technol Assess 2015 Nov;19(99):1-188. [doi: 10.3310/hta19990] [Medline: 26616119]
21. Beukes EW, Manchaiah V, Andersson G, Maidment DW. Application of the behavior change wheel within the context of internet-based cognitive behavioral therapy for tinnitus management. Am J Audiol 2022 Jun 2;31(2):433-444. [doi: 10.1044/2022_AJA-21-00160] [Medline: 35436419]
22. Greenwell K, Featherstone D, Hoare DJ. The application of intervention coding methodology to describe the tinnitus e-programme, an internet-delivered self-help intervention for tinnitus. Am J Audiol 2015 Sep;24(3):311-315. [doi: 10.1044/2015_AJA-14-0089] [Medline: 26649537]
23. Rinn A, Hannibal S, Goetsch S, Weise C, Lehr D. Mit Apps gegen den Tinnitus? Ein systematisches Review zu Qualität, Interventionselementen und Techniken der Verhaltensänderung [Article in German]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2024 Feb;67(2):203-214. [doi: 10.1007/s00103-023-03805-1]

XSL•FO
RenderX

24. Mobile applications for tinnitus: a review of quality, intervention components and behavior change techniques. OSF (Open Science Framework). URL: https://osf.io/jsv9z/overview [accessed 2026-05-04]
25. Messner EM, Terhorst Y, Barke A, et al. The German version of the Mobile App Rating Scale (MARS-G): development and validation study. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e14479. [doi: 10.2196/14479] [Medline: 32217504]
26. Thompson DM, Hall DA, Walker DM, Hoare DJ. Psychological therapy for people with tinnitus: a scoping review of treatment components. Ear Hear 2017;38(2):149-158. [doi: 10.1097/AUD.0000000000000363] [Medline: 27541331]
27. German S3 guideline: 017/064: chronic tinnitus. : Association of the Scientific Medical Societies in Germany (AWMF); 2015.
28. Weise C. Tinnitus. Psychotherapeut 2011 Jan;56(1):61-78. [doi: 10.1007/s00278-010-0791-2]
29. Zenner HP, Delb W, Kröner-Herwig B, et al. A multidisciplinary systematic review of the treatment for chronic idiopathic tinnitus. Eur Arch Otorhinolaryngol 2017 May;274(5):2079-2091. [doi: 10.1007/s00405-016-4401-y] [Medline: 27995315]
30. Weise C, Kleinstäuber M, Kaldo V, Andersson G. Mit Tinnitus Leben Lernen: Ein Manual Für Therapeuten Und Betroffene [Book in German]: Springer; 2016. [doi: 10.1007/978-3-642-54763-8]
31. Lühnen J, Albrecht M, Mühlhauser I, Steckelberg A. Leitlinie evidenzbasierte Gesundheitsinformation. : EbM-Netzwerk; 2017 URL: https://www.ebm-netzwerk.de/de/medien/pdf/leitlinie-evidenzbasierte-gesundheitsinformation-fin.pdf [accessed 2026-04-23]
32. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
33. BCT Taxonomy. 2024. URL: https://www.bct-taxonomy.com [accessed 2026-04-23]
34. IBM SPSS Statistics for Windows: IBM Corp; 2021. URL: https://www.scirp.org/reference/referencespapers?referenceid=3235210 [accessed 2026-04-23]
35. Cima RFF, Kikidis D, Mazurek B, et al. Tinnitus healthcare: a survey revealing extensive variation in opinion and practices across Europe. BMJ Open 2020 Jan 21;10(1):e029346. [doi: 10.1136/bmjopen-2019-029346] [Medline: 31969359]
36. Meijers S, Stegeman I, van der Leun JA, Assegaf SA, Smit AL. Analysis and comparison of clinical practice guidelines regarding treatment recommendations for chronic tinnitus in adults: a systematic review. BMJ Open 2023 Sep 15;13(9):e072754. [doi: 10.1136/bmjopen-2023-072754] [Medline: 37714675]
37. Hesser H, Weise C, Westin VZ, Andersson G. A systematic review and meta-analysis of randomized controlled trials of cognitive-behavioral therapy for tinnitus distress. Clin Psychol Rev 2011 Jun;31(4):545-553. [doi: 10.1016/j.cpr.2010.12.006] [Medline: 21237544]
38. Martinengo L, Stona AC, Griva K, et al. Self-guided cognitive behavioral therapy apps for depression: systematic assessment of features, functionality, and congruence with evidence. J Med Internet Res 2021 Jul 30;23(7):e27619. [doi: 10.2196/27619] [Medline: 34328431]
39. Pryce H, Moutela T, Bunker C, Shaw R. Tinnitus groups: a model of social support and social connectedness from peer interaction. Br J Health Psychol 2019 Nov;24(4):913-930. [doi: 10.1111/bjhp.12386] [Medline: 31449732]
40. Thompson P, Pryce H, Refaie E. Group or individual tinnitus therapy: what matters to participants? Audiol Med 2011 Sep;9(3):110-116. [doi: 10.3109/1651386X.2011.604470]
41. Sereda M, Xia J, El Refaie A, Hall DA, Hoare DJ. Sound therapy (using amplification devices and/or sound generators) for tinnitus. Cochrane Database Syst Rev 2018 Dec 27;12(12):CD013094. [doi: 10.1002/14651858.CD013094.pub2] [Medline: 30589445]
42. Abouzari M, Goshtasbi K, Sarna B, et al. Adapting personal therapies using a mobile application for tinnitus rehabilitation: a preliminary study. Ann Otol Rhinol Laryngol 2021 Jun;130(6):571-577. [doi: 10.1177/0003489420962818] [Medline: 33030042]
43. Mehdi M, Dode A, Pryss R, Schlee W, Reichert M, Hauck FJ. Contemporary review of smartphone apps for tinnitus management and treatment. Brain Sci 2020 Nov 17;10(11):867. [doi: 10.3390/brainsci10110867] [Medline: 33212928]
44. Nagaraj MK, Prabhu P. Internet/smartphone-based applications for the treatment of tinnitus: a systematic review. Eur Arch Otorhinolaryngol 2020 Mar;277(3):649-657. [doi: 10.1007/s00405-019-05743-8] [Medline: 31807891]
45. Oron Y, Ben David BM, Doron G. Brief cognitive-behavioral training for tinnitus relief using a mobile application: a pilot open trial. Health Informatics J 2022;28(1):14604582221083483. [doi: 10.1177/14604582221083483] [Medline: 35349777]
46. Schlee W, Neff P, Simoes J, et al. Smartphone-guided educational counseling and self-help for chronic tinnitus. J Clin Med 2022 Mar 25;11(7):1825. [doi: 10.3390/jcm11071825] [Medline: 35407432]
47. Walter U, Pennig S, Kottmann T, Bleckmann L, Röschmann-Doose K, Schlee W. Randomized controlled trial of a smartphone-based cognitive behavioral therapy for chronic tinnitus. PLOS Digit Health 2023 Sep;2(9):e0000337. [doi: 10.1371/journal.pdig.0000337] [Medline: 37676883]
48. Sander LB, Schorndanner J, Terhorst Y, et al. “Help for trauma from the app stores?” A systematic review and standardised rating of apps for post-traumatic stress disorder (PTSD). Eur J Psychotraumatol 2020;11(1):1701788. [doi: 10.1080/20008198.2019.1701788] [Medline: 32002136]

XSL•FO
RenderX

49. Schoeppe S, Alley S, Rebar AL, et al. Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: a review of quality, features and behaviour change techniques. Int J Behav Nutr Phys Act 2017 Jun 24;14(1):83. [doi: 10.1186/s12966-017-0538-3] [Medline: 28646889]
50. Brown HM, Bucher T, Collins CE, Rollo ME. A review of pregnancy iPhone apps assessing their quality, inclusion of behaviour change techniques, and nutrition information. Matern Child Nutr 2019 Jul;15(3):e12768. [doi: 10.1111/mcn.12768] [Medline: 30569549]

Abbreviations

BCT: behavior change technique
CBT: cognitive behavioral therapy
IMI: internet- and mobile-based intervention
MARS-G: German version of the Mobile App Rating Scale
MARS-G: German version of the Mobile App Rating Scale
SPSS: Statistical Package for the Social Sciences

Edited by L Buis; submitted 05.Sep.2024; peer-reviewed by A Eisingerich, KL Alfrey, R Pryss; revised version received 28.Mar.2026; accepted 30.Mar.2026; published 19.May.2026.

Please cite as:
Rinn A, Goetsch S, Hannibal S, Lehr D, Weise C
Mobile Apps for Tinnitus: Systematic Search in App Stores and Review of Intervention Components and Behavior Change Techniques
JMIR Mhealth Uhealth 2026;14:e66151
URL: https://mhealth.jmir.org/2026/1/e66151
doi:10.2196/66151

© Alina Rinn, Sarah Goetsch, Sandy Hannibal, Dirk Lehr, Cornelia Weise. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 19.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Patient-Centered Lupus Erythematosus Mobile Apps: Systematic Search and Cross-Sectional Evaluation by Patients and Physicians

Tassilo Dege[1,2,3], MD; Antonia Ullmann[1], DMD; Caroline Glatzel[1], MD; Janik Fleißner[1], MD; Vanessa Borst[2], MS; Patrick-Pascal Strunz[3], MD; Marc Schmalzing[3], MD; Matthias Goebeler[1], Prof Dr; Astrid Schmieder[1], Prof Dr

[1]Department of Dermatology, Venereology and Allergology, Universitätsklinikum Würzburg, Josef-Schneider-Str. 2, Würzburg, Germany

[2]Department of Computer Science, University of Würzburg, Würzburg, Germany

[3]Department of Rheumatology, Universitätsklinikum Würzburg, Würzburg, Germany

Corresponding Author:
Tassilo Dege, MD
Department of Dermatology, Venereology and Allergology, Universitätsklinikum Würzburg, Josef-Schneider-Str. 2, Würzburg, Germany

Abstract

Background: Lupus erythematosus (LE) is a chronic autoimmune disease that significantly impacts patients' quality of life. Photosensitivity is a key impairment that severely limits the quality of life, especially in cutaneous lupus erythematosus (CLE), where exposure to sunlight can lead to rashes, exacerbations, and pain. In systemic lupus erythematosus (SLE), other manifestations such as joint pain, fatigue, and organ damage may contribute to decreased physical function and emotional distress. Mobile health apps (MHA) offer potential support for comprehensive disease management for the symptoms mentioned above. However, there is a lack of systematic analysis of available lupus management apps.

Objective: This study aims to systematically identify publicly available German or English MHA for lupus management as well as to assess their quality by surveying both patients and physicians.

Methods: A systematic search and assessment of German or English mobile apps for patients with lupus, available in the Google Play Store and Apple App Store, was conducted independently by two reviewers. The two apps that met all relevant criteria were then reviewed independently by seven physicians using the German Mobile Application Rating Scale (MARS) and the System Usability Scale (SUS). Subsequently, they were reviewed by five patients (three with SLE and two with CLE), using the user version of MARS (uMARS) and SUS. Additionally, the Affinity for Technology Interaction (ATI) scale was collected from both patients and physicians to evaluate the technical affinity in both groups.

Results: In total, 29 apps were available on the Apple Store and 26 on the Google Store, with 18 apps being present and downloadable on both platforms. Of the 18 apps, 16 were excluded because they did not meet the inclusion and exclusion criteria. Only two apps, *Lupus Log* and *Lupus Minder* met all the required criteria and were included in the study. The mean MARS scores varied from 2.61/5 to 4.17/5 and mean SUS from 17.5/100 to 100/100 between physicians. The app with the highest mean overall MARS score was *Lupus Log*, which was rated with 3.91/5 on average by the physicians. Patients evaluated the app with a comparably mean uMARS score (3.95/5). Technical affinity, objectified by ATI, was higher in patients than physicians (3.9 vs 3.68).

Conclusions: Systematic identification and evaluation showed high-quality apps for patient-centered lupus MHA as indicated by MARS and uMARS scores greater than 3 for both *Lupus Log* and *Lupus Minder*.

(JMIR Mhealth Uhealth 2026;14:e73019) doi:10.2196/73019

KEYWORDS

lupus erythematosus; systemic lupus erythematosus; cutaneous lupus erythematosus; mobile applications; evaluation; mental health; mars; sus; ati; teledermatology; disease management; health app; skin; eHealth; mHealth; telemedicine

Introduction

Background

Lupus erythematosus (LE) is a chronic autoimmune disorder with great heterogeneity of symptoms. It can affect multiple organs, including skin, kidney, heart, and the nervous system [1]. In cutaneous lupus erythematosus (CLE), inflammation is typically confined to the skin and does not affect internal organs. CLE is often worsened or even triggered by sun exposure [2]. Systemic lupus erythematosus (SLE) can present with many clinical symptoms of varying severity, including lupus nephritis, which can lead to kidney failure if left untreated. Therefore, SLE patients are treated with systemic corticosteroids and other immunosuppressants or immunomodulators, which can again

XSL•FO
RenderX

cause side effects and result in additional morbidity. In contrast, for CLE, topical treatments are often sufficient to relieve symptoms [3].

Due to chronic inflammation, patients experience a variety of physical limitations, including joint pain, fatigue, sleep disorders, photosensitivity as well as cardiovascular and pulmonary symptoms. Mental and emotional health is further impaired with patients showing high prevalence of depression, anxiety, and cognitive disorders [4]. Furthermore, SLE can also directly affect the central nervous system causing a wide variety of psychiatric symptoms [5]. Some patients experience a so called “lupus fog” with memory loss and difficulty concentrating, which further negatively impacts daily life and social interactions [6]. Corticosteroids and immunosuppressive drugs can cause weight gain, diabetes, osteoporosis, and increased risk of infection that are detrimental to patients’ quality of life [7].

In people living with LE, symptoms such as chronic fatigue, joint pain, and cognitive difficulties often occur unnoticed by others, making it hard for patients to feel understood and to receive empathy. This invisibility can lead to several emotional and social struggles such as frustration and a feeling of stigmatization [8]. The unpredictability of lupus flare-ups makes this disease more challenging. Even when the disease appears to be controlled, symptoms can suddenly return and cause fear and anxiety [9].

Since patients with lupus often experience fluctuating symptoms, it can complicate treatment adherence and monitoring. Therefore, mobile health apps (MHA) present an opportunity to support patients in managing their condition by providing accessible information, symptom tracking, and direct communication with health care providers [10]. However, there is a lack of robust evidence regarding the effectiveness of these apps for lupus patients, and high-quality studies are essential to assess their impact [11]. To date, comprehensive reviews of MHA for lupus management have been scarce [12], as they were restricted to evaluations conducted exclusively by health care professionals (with small reviewer samples) and without incorporating patient perspectives. To complement existing reviews, a comparative analysis using validated scoring systems and involving both health professionals and patients is needed to generate more robust data on the usability and quality of MHAs for lupus self-management. The Affinity for Technology Interaction (ATI) scale provides a reliable method for quantifying an individual’s technology affinity, providing important context for interpreting their app evaluations [13]. To subjectively assess app quality, the Mobile App Rating Scale (MARS) was developed, evaluating aspects such as engagement, functionality, esthetics, and information [14-16]. MARS has been applied in studies evaluating mobile apps related to various chronic diseases, including chronic wounds [17]. A user version (uMARS) allows app quality to be assessed from the patient perspective [16]. The System Usability Scale (SUS), a 10-item questionnaire, has been used broadly to assess the usability of systems, including mobile apps for conditions like dementia, depression, pediatric obesity, and smoking cessation [18].

Aim of This Study

This study aims to systematically identify and assess mobile apps available for lupus management. By using validated tools, such as the MARS and the SUS, this research seeks to provide a comprehensive evaluation of these apps from both professional and CLE as well as SLE patient perspectives.

Methods

Study Design and Setting

This was a cross-sectional observational study conducted at the Department of Dermatology, University Hospital Würzburg, Germany. The study involved a systematic search and comparative evaluation of MHAs for patients with LE by both patients and physicians.

Participants

Physicians were dermatologists involved in direct lupus patient care. Patients were adults (≥18 y) with a confirmed diagnosis of cutaneous or systemic lupus erythematosus, identified during routine outpatient visits.

Ethical Considerations

This study was conducted in accordance with ethical guidelines and standards, with approval from the ethics committee at the University of Würzburg (reference number 2022062501). All participants provided written informed consent. No personal identifiers were collected on the forms, and responses were manually entered by the study team. All data were deidentified at the point of analysis. No financial or material compensation was provided to participants for taking part in the study.

App Identification and Selection

Two independent reviewers systematically searched the Apple App Store and the Google Play Store in June and July of 2023 to identify patient-centered mobile health applications related LE. The search terms used were: “lupus,”, “lupus erythematosus,” “lupus erythematodes,” “systemic lupus,” “cutaneous lupus”. German and English apps were included in order to reflect the languages most used by the study population, and to ensure relevance to the target user group. To guarantee cross-platform usability and broader accessibility, only apps available in both app stores were considered. Additionally, it was required that apps were specifically designed for patients with LE, were free of charge, did not require user registration or community interaction, and did not contain advertisements. We excluded apps with advertisements to prevent disruptions in users’ experiences and to avoid the confounding effects advertisements can have on usability [19]. Advertisements can frustrate users, reduce the clarity and intuitiveness of an app’s interface, and make it difficult to compare apps’ performance or users’ experiences [20]. As part of the inclusion criteria, only apps with a publicly accessible privacy policy and evidence of basic data protection measures were included. This ensured alignment with good practice guidelines for mobile health evaluations.

XSL•FO
RenderX

Data Extraction and App Characteristics

The following information available in the app stores and on app websites was collected

- App name
- Target consumer group (eg, patients, medical personnel)
- Cost
- Platform
- Advertisement
- Features
- Search term used to identify the app

Evaluation

Mobile Application Rating Scale

The quality of the identified and included apps was evaluated by seven physicians using the MARS, which assesses the four domains: engagement, functionality, esthetics, and information. Each item is rated on a 5-point Likert scale, where 1 indicates "inadequate" and 5 indicates "excellent" performance. Higher scores reflect better app quality. The overall MARS score is calculated as the mean of the subdomain scores, with a maximum possible score of 5; a score above 3 is considered acceptable or good app quality. The apps were then further evaluated by five patients (three with systemic lupus erythematosus and two with cutaneous lupus erythematosus) using the user version of the scale (ie, uMARS), which uses the same scoring logic and dimensions but is adapted for the user perspective.

System Usability Scale

In addition, the SUS was applied to assess overall usability. SUS consists of 10 items rated on a 5-point Likert scale, with the total score converted to a 0 - 100 scale. Higher SUS scores indicate better usability. A SUS score ≥70 is generally considered acceptable, scores between 50 and 70 indicate marginal usability, and scores below 50 are interpreted as poor usability. Including both MARS and SUS allowed for a comprehensive assessment of app quality and usability from both the health care professional and the patient viewpoint.

Affinity for Technology Interaction

The ATI scale was used to gather information on the physicians' and patients' technical affinity. ATI consists of 9 items rated on a 6-point Likert scale, resulting in a score range from 1 (very low affinity) to 6 (very high affinity). Higher scores indicate a greater interest in interacting with technology. A score above three reflects average or moderate technology affinity, while a score above four is typically interpreted as high. Assessing ATI in both user groups provided context for interpreting their app ratings, particularly in terms of potential differences in expectations, familiarity, and digital literacy.

Evaluation Procedures

Each physician and patient received access to the same apps and tested them on their own smartphones or tablets (iOS or Android). Participants were instructed to test each app for at least ten minutes, explore all relevant features, and assess its usability, content, and overall functionality. To assist participants in understanding the rating process, an optional explanatory video describing the MARS/uMARS tool was made available via a QR code. Following app testing, participants completed the appropriate paper-based evaluation tools. All questionnaires were filled out on paper and collected by the study team for manual data entry and analysis.

Statistical Analysis

Descriptive statistics were used to summarize app quality scores and demographic data. The Mann-Whitney *U* test was used to compare scores between patients and physicians (eg, MARS vs. uMARS, SUS, ATI). The Wilcoxon signed-rank test was applied to compare differences between apps. *P* values below .05 were considered significant. The data analysis was performed using Python, with libraries such as Pandas, SciPy [21], NumPy [22], Matplotlib, and Seaborn [23].

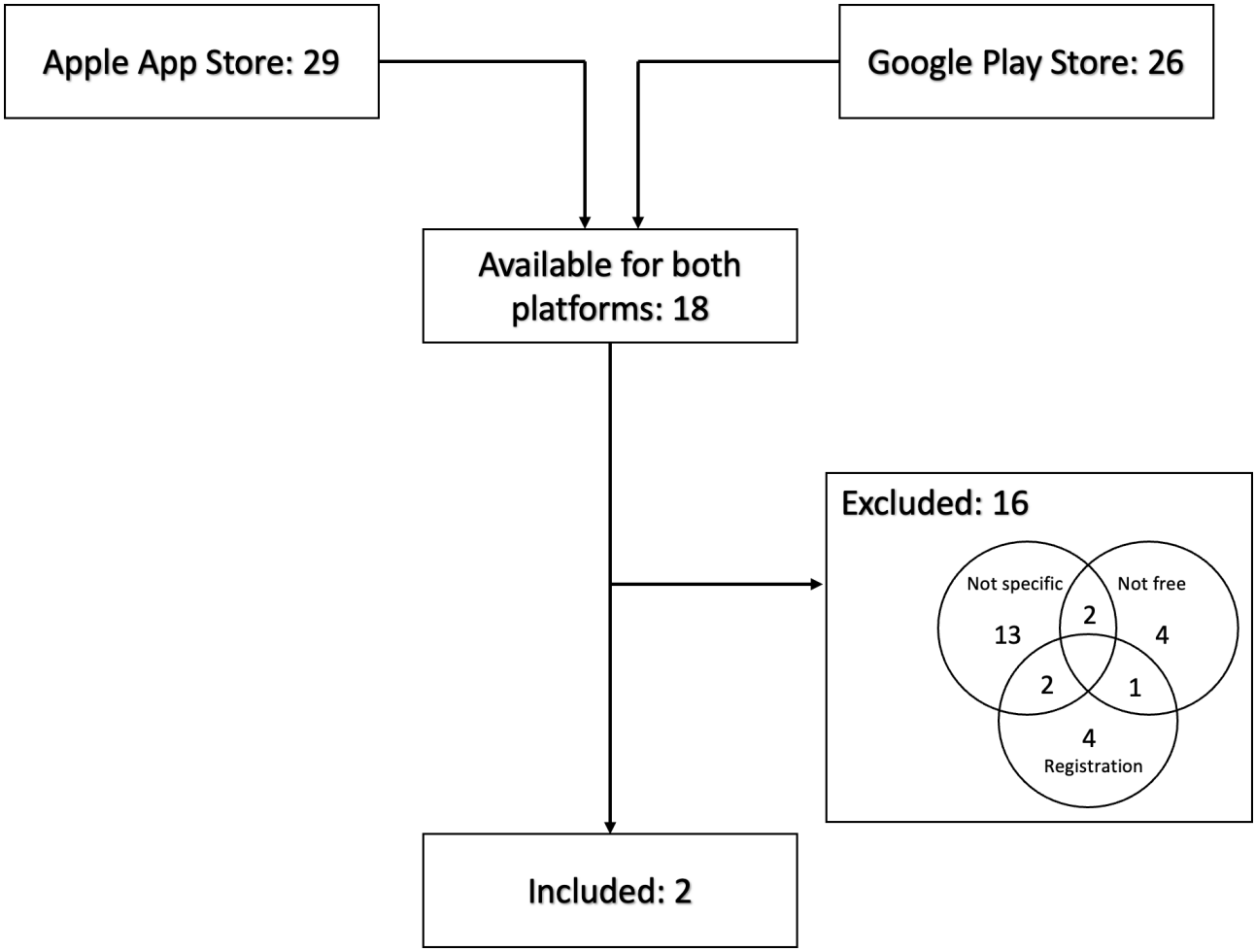
Results

App Selection

In total, 37 potentially relevant mobile health apps targeting lupus patients were initially identified through the app store search. These apps were then reviewed independently by two authors to assess eligibility according to predefined inclusion and exclusion criteria. Of the 37, 18 apps were available in both the Apple App Store and Google Play Store (see Figure 1). Of these 18 apps, 10 were targeted at patients. Strict criteria were established for app inclusion: the apps needed to be specifically designed for LE, free of charge, and did not require registration/community connections, or advertising. Of the 18 apps, 13 were excluded because they were not specifically designed for LE. Additionally, 4 apps required payment after download or for full access, and 4 apps asked for registration, included ads, or had community connections. Some apps met multiple exclusion criteria. Ultimately, only 2 apps, *Lupus Log* and *Lupus Minder*, met all the required criteria and were included in the study.

XSL•FO
RenderX

Figure 1. Flowchart illustrating the screening process for identifying suitable mobile apps.



Evaluation of App Quality

A total of two apps —*Lupus Log* and *Lupus Minder* —met all inclusion and exclusion criteria and were evaluated by 7 physicians. *Lupus Log* and *Lupus Minder* are designed specifically for lupus patients, offering features tailored to disease management and symptom tracking. Both apps provided a user-friendly interface and allow for the creation of a personalized medication plan. *Lupus Minder* features a structured design with an intuitive layout. It offers general lupus information and a planner for medical appointments, which can be printed or exported. A daily symptom tracker is available, but it only includes a free-text field without predefined options. The app ensures data security through password protection. *Lupus Log* stood out for its well-structured design and includes a real-time, location-based UV index. Its daily symptom tracker allows users to select predefined symptoms, add custom categories (eg, fatigue, lesions, ulcers), and to attach notes or photos. The app also provides lupus-related tips, news updates, and access to personal health statistics. Additionally, push notifications can remind users to complete the symptom tracker, though medication reminders are not included.

The mean MARS scores (and SDs) of the physicians, including subcategory scores and SUS scores, are provided in Table 1, while the corresponding uMARS and SUS results from patient evaluations are shown in Table 2.

XSL•FO
RenderX

Table . MARS[a] and SUS[b] scores of the *Lupus Log* and *Lupus Minder* app, as assessed by 7 physicians.

Scores	Lupus Log	Lupus Minder
MARS, mean (SD)	3.91 (0.42)	3.29 (0.38)
A: Engagement score, mean (SD)	3.69 (0.58)	2.86 (0.65)
B: Functionality score, mean (SD)	4.46 (0.37)	3.75 (0.46)
C: Esthetics score, mean (SD)	4.19 (0.72)	3.24 (0.66)
D: Information score, mean (SD)	3.48 (0.65)	3.43 (0.26)
SUS, mean (SD)	87.14 (4.43)	65.36 (27.59)

[a]MARS: Mobile Application Rating Scale.

[b]SUS: System Usability Score.:

Table . uMARS[a] and SUS[b] scores of the *Lupus Log* and *Lupus Minder* app, as assessed by 5 patients.

Scores	Lupus log	Lupus minder
uMARS, mean (SD)	3.95 (0.62)	3.32 (0.52)
A: Engagement score, mean (SD)	3 (0.94)	2.4 (0.52)
B: Functionality score, mean (SD)	4.35 (0.93)	3.8 (1)
C: Esthetics score, mean (SD)	4.41 (0.44)	3.74 (0.91)
D: Information score, mean (SD)	4 (0.5)	3.56 (0.43)
SUS	76 (29.18)	49.38 (34.18)

[a]uMARS: User version of Mobile Application Rating Scale.

[b]SUS: System Usability Score.

Lupus Log had the highest mean MARS (mean 3.91, SD 0.42 out of 5) and mean uMARS (mean 3.95, SD 0.62 out of 5). When Wilcoxon signed-rank test was applied *Lupus Log* had a significantly higher MARS score (P=.03*)* but not significantly higher uMARS scores (P=.125), possibly due to small size (see detailed information on demographics, uMARS, SUS and ATI scores of CLE and SLE patients in Multimedia Appendix 1). *Lupus Minder* had a lower mean MARS (mean 3.29, SD 0.38) and mean uMARS (mean 3.31, SD 0.52 out of 5). Mean SUS by physicians and patients scores were 87% and 76% for *Lupus Minder* and 65% and 49% for *Lupus Log*, respectively.

Comparative Analysis of Patients and Physicians Data

Lupus Log MARS (see Table 1) and uMARS (see Table 2) were almost identical and differences were not statistically significant (P=.76). In the subcategory engagement and information, mean scores differed most. Patients rated engagement with 3.0 lower than physicians 3.69 (P=.16). In contrast, information was rated higher by patients with a mean score of 4.0 than by physicians with a mean score of 3.48 (P=.19).

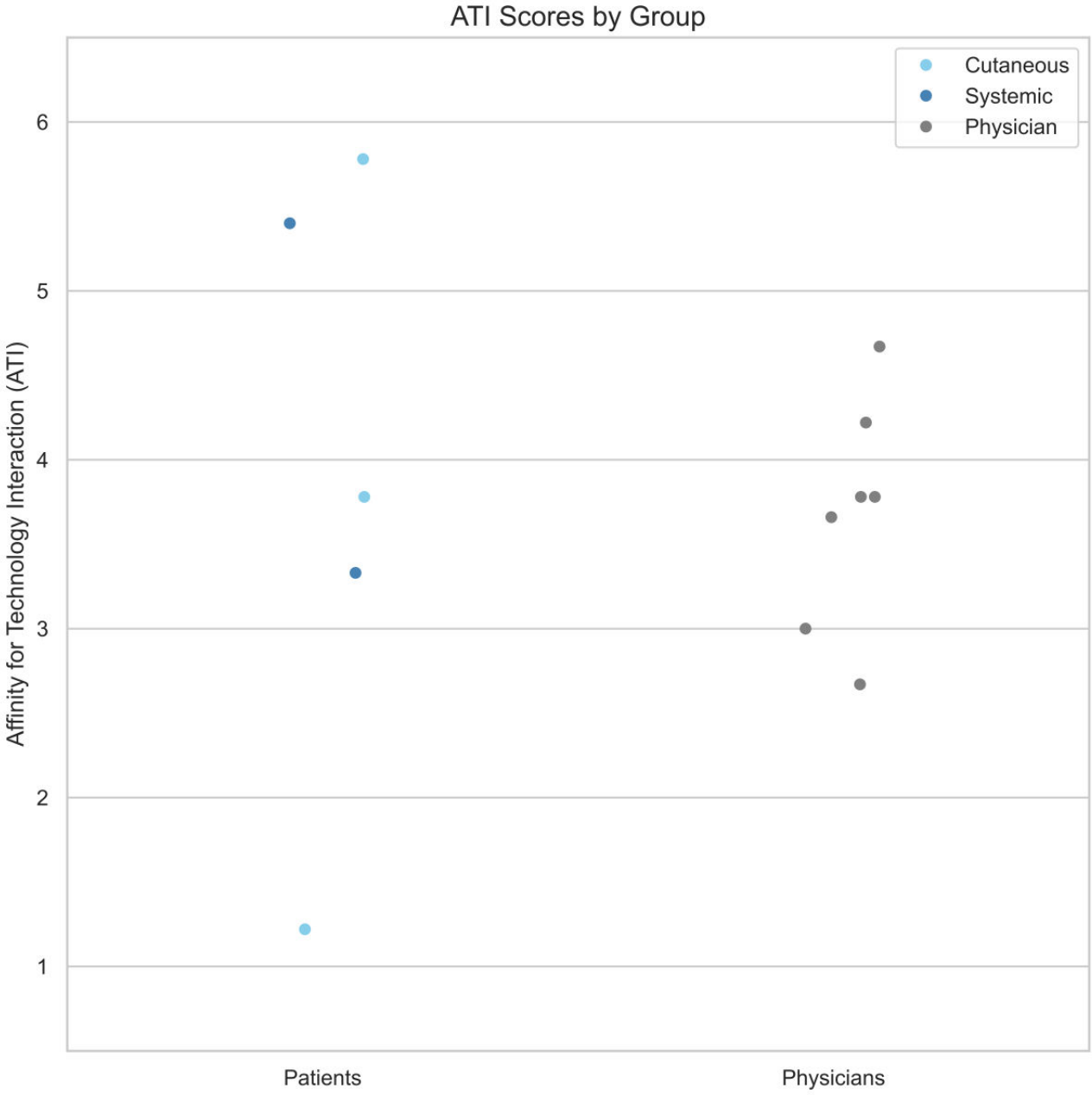
Lupus Minder's MARS (see Table 1) score and uMARS (see Table 2) were almost identical (P=.70). In the subcategory engagement and esthetics, mean scores differed most. Patients rated engagement with 2.4 lower than physicians with 2.86 (P=.30). In contrast, esthetics was rated higher by patients with a mean score of 3.74 than by physicians with a mean score of 3.24 (P=.40).

Evaluation of Technical Affinity

The physician group consisted of 5 females and 2 males, with an age range of 27 - 42 years (median: 30 y). The patient group included 3 patients with systemic lupus erythematosus (SLE) and 2 with cutaneous lupus erythematosus (CLE). Four patients were female, and one was male, with ages ranging from 35 to 67 years (median: 52 y). ATI scores ranged from 1.22 to 5.78 out of 6. An ATI score of >3 indicates average technology affinity, and a score of >4 indicates high technology affinity. Technical affinity was lower in physicians than in patients (3.68, SD 0.68 vs 3.9, SD 1.82) but this difference was not statistically significant (P=.74). A visualization of ATI scores in both groups is shown in Figure 2.

XSL•FO
RenderX

Figure 2. Individual ATI scores (range: 1 - 6) for patients with cutaneous (light blue) or systemic (dark blue) lupus erythematosus and physicians (gray). Patients showed slightly higher scores than physicians, though not statistically significant (P=.74, Mann-Whitney U test).



Discussion

Principal Findings

In this study, we first conducted a systematic search to identify smartphone apps specifically designed for patients with LE, following approaches used in previous evaluations of mobile health apps [24-26]. Among the included apps, *Lupus Log* received a significantly higher MARS score than *Lupus Minder* (P=.03), indicating a clear preference among physicians regarding overall app quality. However, no statistically significant differences were observed in the uMARS scores among patients (P=.*12*), nor between patients and physicians across most subdomains, possibly due to the small sample size and the limited power to detect subtle differences. Although several subgroup comparisons (eg, ATI scores between groups, SUS scores between apps) showed meaningful trends, they did not reach statistical significance.

High agreement among assessors, even for subjective measures, is an encouraging indicator of the reliability of the assessment. While the Enlight [27] framework is another feasible and validated tool for digital health app evaluation, MARS and uMARS were chosen for this study due to its established use in comparative assessments by both health professionals and patients [25], aligning with the study's focus on multi-stakeholder perspectives.

This study identified a limited number of high-quality lupus management apps that met our strict inclusion criteria. Both *Lupus Log* and *Lupus Minder* received decent ratings when rated by raters who can be considered affine to technology based on mean ATI scores greater than 3 for physicians and patients.

XSL•FO
RenderX

With a mean MARS score of 3.91 (SD 0.42), *Lupus Log* was rated highest by physicians. Likewise, it achieved a mean uMARS score of 3.95 (SD 0.62). In the subcategory ‚information', *Lupus Log* was rated higher by patients overall and in the subcategory 'engagement,' possibly due to offering additional information on UV index.

Previous studies have assessed the quality of lupus-related mobile health apps but were limited to evaluations conducted exclusively by healthcare professionals, often with only two raters per app. Ramasamy et al [11] evaluated 19 apps for systemic lupus erythematosus (SLE) and lupus nephritis, reporting that *Lupus Minder* received an overall MARS score of 2.96, with domain-specific scores of 2.9 (engagement), 3.5 (functionality), 3.1 (esthetics), and 2.35 (information). In a separate systematic review, Dantas et al [12] evaluated *Lupus Log*, which scored 2.7 overall, including 3.3 (engagement), 3.2 (functionality), 2.3 (esthetics), and 2.1 (information). While informative, both studies relied on assessments from only two health care professionals per app, without incorporating patient perspectives. In contrast, our study included a broader group of raters—seven physicians and five patients—and uniquely combines both clinical and user-centered evaluations. This dual-stakeholder approach provides a more comprehensive evaluation of app quality and usability. Compared to the previous reviews by Dantas et al and Ramasamy et al, which identified a larger number of lupus apps, our study applied stricter inclusion criteria focused on real-world usability; we included only apps that were free, available on both major platforms, and did not require registration or contain advertising. In contrast, both prior reviews allowed paid apps, iOS-only apps, and apps with limited accessibility.

The episodic nature of lupus makes assessing disease activity and severity challenging, as flare-ups are unpredictable and often go unobserved by physicians. Current evaluations rely on patients' memories, subjective reports, and occasional photos, which can lead to misjudgments and suboptimal treatment. Regular monitoring patient-reported outcome measures is therefore essential, as recommended by lupus management guidelines [28], which advocate for validated tools like LupusQoL [29] and SLAQ (Systemic Lupus Activity Questionnaire) [30] to assess disease control. A mobile health app can simplify this process by enabling continuous tracking of symptoms, severity, and flare-ups while providing an overview for both patients and physicians. This improves support, builds confidence, and enhances treatment precision for physicians and patients.

However, patient involvement in the app development process remains limited [31]. As the primary end users, patients should play a more active role in the co-design and development process to ensure usability, relevance, and engagement [32,33] Both patient-centered apps evaluated in this study, provided a diary function which serves both physicians and patients. While no prior studies have evaluated mobile apps for lupus erythematosus using both MARS and uMARS, comparisons with evaluations in other disease areas, such as mental health, are informative. A recent study using MARS to assess French-language mental health apps [34] reported a top MARS score of 3.85, which is comparable to the scores observed in our study. These findings suggest that the quality of our top-rated apps is on par with apps in better-studied domains, such as mental health.

Limitations

The study's limitations include a small sample size and the subjective nature of the evaluation tools. Future research should involve larger patient cohorts and consider randomized controlled trials to assess app effectiveness more robustly over a longer period. Other app platforms (eg, Huawei AppGallery, Samsung Galaxy Store, Windows Phone Store and Blackberry World) were excluded from the search to focus on the most used platforms—Apple App Store and Google Play Store—which dominate the mobile health app market in Western countries. This exclusion is acknowledged as a limitation of the study. By excluding apps with advertisements, some potentially relevant apps may have been omitted. This narrower inclusion strategy, while intentionally designed, likely contributed to the lower number of included apps compared to earlier reviews.

Conclusions

Lupus Log and *Lupus Minder* offer practical tools for tracking disease activity in both CLE and SLE. These apps have received positive feedback from patients and physicians, supporting disease control and promoting patient self-management. Their acceptance among patients highlights the need for further studies to evaluate long-term use and adapt the apps to meet the needs of both groups. Involving patients in the app design process can enhance usability and relevance, ultimately leading to better disease management outcomes.

Acknowledgments

Tassilo Dege is supported by the Clinician Scientist Program, project number Z2/CSP-25, of the Interdisciplinary Center for Clinical Research (IZKF) at the Medical Faculty of the University of Würzburg.

Funding

The authors declared no financial support was received for this work.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Patient (A) and physician (B) characteristics (age and sex) and the Affinity for Technology Interaction scale, System Usability Scale, Mobile App Rating Scale (MARS), and user version of the MARS scores (including subcategory scores) of all 5 patients and 7 physicians who participated.
[DOC File, 56 KB - mhealth_v14i1e73019_app1.doc]

References

1. Barber MRW, Drenkard C, Falasinnu T, et al. Global epidemiology of systemic lupus erythematosus. Nat Rev Rheumatol 2021 Sep;17(9):515-532. [doi: 10.1038/s41584-021-00668-1] [Medline: 34345022]
2. Niebel D, de Vos L, Fetter T, Brägelmann C, Wenzel J. Cutaneous lupus erythematosus: an update on pathogenesis and future therapeutic directions. Am J Clin Dermatol 2023 Jul;24(4):521-540. [doi: 10.1007/s40257-023-00774-8] [Medline: 37140884]
3. Rothfield N, Sontheimer RD, Bernstein M. Lupus erythematosus: systemic and cutaneous manifestations. Clin Dermatol 2006;24(5):348-362. [doi: 10.1016/j.clindermatol.2006.07.014] [Medline: 16966017]
4. Schmeding A, Schneider M. Fatigue, health-related quality of life and other patient-reported outcomes in systemic lupus erythematosus. Best Pract Res Clin Rheumatol 2013 Jun;27(3):363-375. [doi: 10.1016/j.berh.2013.07.009]
5. Carrión-Barberà I, Salman-Monte TC, Vílchez-Oya F, Monfort J. Neuropsychiatric involvement in systemic lupus erythematosus: A review. Autoimmun Rev 2021 Apr;20(4):102780. [doi: 10.1016/j.autrev.2021.102780] [Medline: 33609799]
6. Cleanthous S, Strzok S, Haier B, Cano S, Morel T. The patient experience of fatigue in systemic lupus erythematosus: a conceptual model. Rheumatol Ther 2022 Feb;9(1):95-108. [doi: 10.1007/s40744-021-00374-0] [Medline: 34622426]
7. Wirz-Justice A, Skene DJ, Münch M. The relevance of daylight for humans. Biochem Pharmacol 2021 Sep;191(114304):114304. [doi: 10.1016/j.bcp.2020.114304] [Medline: 33129807]
8. Sehlo MG, Bahlas SM. Perceived illness stigma is associated with depression in female patients with systemic lupus erythematosus. J Psychosom Res 2013 Mar;74(3):248-251. [doi: 10.1016/j.jpsychores.2012.09.023] [Medline: 23438717]
9. Zhang L, Fu T, Yin R, Zhang Q, Shen B. Prevalence of depression and anxiety in systemic lupus erythematosus: a systematic review and meta-analysis. BMC Psychiatry 2017 Feb 14;17(1):70. [doi: 10.1186/s12888-017-1234-1] [Medline: 28196529]
10. Karlson EW, Liang MH, Eaton H, et al. A randomized clinical trial of a psychoeducational intervention to improve outcomes in systemic lupus erythematosus. Arthritis Rheumatism 2004 Jun;50(6):1832-1841. [doi: 10.1002/art.20279]
11. Ramasamy A, Dugyala P, Mohan C. Mobile health apps for systemic lupus erythematosus and lupus nephritis: a critical appraisal. Arthritis Res Ther 2022 Dec;24(1):110. [doi: 10.1186/s13075-022-02791-0]
12. Dantas LO, Weber S, Osani MC, Bannuru RR, McAlindon TE, Kasturi S. Mobile health technologies for the management of systemic lupus erythematosus: a systematic review. Lupus (Los Angel) 2020 Feb;29(2):144-156. [doi: 10.1177/0961203319897139] [Medline: 31924145]
13. Franke T, Attig C, Wessel D. A personal resource for technology interaction: development and validation of the Affinity for Technology Interaction (ATI) scale. Int J Hum-Comput Int 2019 Apr 3;35(6):456-467. [doi: 10.1080/10447318.2018.1456150]
14. Knitza J, Tascilar K, Messner EM, et al. German mobile apps in rheumatology: review and analysis using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2019 Aug 5;7(8):e14991. [doi: 10.2196/14991] [Medline: 31381501]
15. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth 2015;3(1):e27. [doi: 10.2196/mhealth.3422]
16. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016;4(2):e72. [doi: 10.2196/mhealth.5849]
17. Dege T, Glatzel B, Borst V, et al. Patient-Centered chronic wound care mobile apps: systematic identification, analysis, and assessment. JMIR Mhealth Uhealth 2024 Mar 21;12:e51592. [doi: 10.2196/51592] [Medline: 38533818]
18. Hyzy M, Bond R, Mulvenna M, et al. System Usability Scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022 Aug 18;10(8):e37290. [doi: 10.2196/37290] [Medline: 35980732]
19. Haggag O, Grundy J, Abdelrazek M, Haggag S. A large scale analysis of mHealth app user reviews. Empir Softw Eng 2022;27(7):196. [doi: 10.1007/s10664-022-10222-6] [Medline: 36246486]
20. Lobo EH, Abdelrazek M, Frølich A, et al. Detecting user experience issues from mHealth apps that support stroke caregiver needs: an analysis of user reviews. Front Public Health 2023;11:1027667. [doi: 10.3389/fpubh.2023.1027667] [Medline: 37304084]
21. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020 Mar;17(3):261-272. [doi: 10.1038/s41592-019-0686-2] [Medline: 32015543]
22. Harris CR, Millman KJ, van der Walt SJ, et al. Array programming with NumPy. Nature New Biol 2020 Sep 17;585(7825):357-362. [doi: 10.1038/s41586-020-2649-2]
23. Waskom ML. seaborn: statistical data visualization. JOSS 2021;6(60):3021. [doi: 10.21105/joss.03021]

XSL•FO
RenderX

24. Glatzel C, Dege T, Glatzel B, Goebeler M, Presser D, Schmieder A. Identifying and evaluating mobile and web apps for patients to manage hidradenitis suppurativa: systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2025 Aug 29;13:e69030. [doi: 10.2196/69030] [Medline: 40882617]
25. Lull C, von Ahnen JA, Gross G, et al. German mobile apps for patients with psoriasis: systematic search and evaluation. JMIR Mhealth Uhealth 2022 May 26;10(5):e34017. [doi: 10.2196/34017] [Medline: 35617014]
26. Mandracchia F, Llauradó E, Tarro L, Valls RM, Solà R. Mobile phone apps for food allergies or intolerances in app stores: systematic search and quality assessment using the Mobile App Rating Scale (MARS). JMIR Mhealth Uhealth 2020 Sep 16;8(9):e18339. [doi: 10.2196/18339] [Medline: 32936078]
27. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based ehealth interventions. J Med Internet Res 2017 Mar 21;19(3):e82. [doi: 10.2196/jmir.7270] [Medline: 28325712]
28. Fanouriakis A, Kostopoulou M, Andersen J, et al. EULAR recommendations for the management of systemic lupus erythematosus: 2023 update. Ann Rheum Dis 2024 Jan 2;83(1):15-29. [doi: 10.1136/ard-2023-224762] [Medline: 37827694]
29. McElhone K, Abbott J, Shelmerdine J, et al. Development and validation of a disease‐specific health‐related quality of life measure, the LupusQol, for adults with systemic lupus erythematosus. Arthritis Rheumatism 2007 Aug 15;57(6):972-979. [doi: 10.1002/art.22881]
30. Karlson EW, Daltroy LH, Rivest C, et al. Validation of a Systemic Lupus Activity Questionnaire (SLAQ) for population studies. Lupus (Los Angel) 2003;12(4):280-286. [doi: 10.1191/0961203303lu332oa] [Medline: 12729051]
31. Frey AL, Baines R, Hunt S, Kent R, Andrews T, Leigh S. Association between the characteristics of mHealth apps and user input during development and testing: secondary analysis of app assessment data. JMIR Mhealth Uhealth 2023 Nov 22;11:e46937. [doi: 10.2196/46937] [Medline: 37991822]
32. Or CK, Holden RJ, Valdez RS. Human factors engineering and user-centered design for mobile health technology: enhancing effectiveness, efficiency, and satisfaction. In: Human-Automation Interaction: Mobile Computing: Springer; 2022:97-118. [doi: 10.1007/978-3-031-10788-7_6]
33. Mao JY, Vredenburg K, Smith PW, Carey T. The state of user-centered design practice. Commun ACM 2005 Mar;48(3):105-109. [doi: 10.1145/1047671.1047677]
34. Carrouel F, du Sartz de Vigneulles B, Bourgeois D, et al. Mental health mobile apps in the French app store: assessment study of functionality and quality. JMIR Mhealth Uhealth 2022 Oct 12;10(10):e41282. [doi: 10.2196/41282] [Medline: 36223178]

Abbreviations

ATI: affinity for technology interaction
CLE: cutaneous lupus erythematosus
LE: lupus erythematosus
MARS: Mobile Application Rating Scale
MHA: Mobile health apps
SLAQ: Systemic Lupus Activity Questionnaire
SLE: systemic lupus erythematosus
SUS: System Usability Scale
uMARS: user version of Mobile Application Rating Scale

Edited by L Buis; submitted 23.Feb.2025; peer-reviewed by F Carrouel, J Grosser, N Yajima; revised version received 29.Sep.2025; accepted 12.Mar.2026; published 29.May.2026.

Please cite as:
Dege T, Ullmann A, Glatzel C, Fleißner J, Borst V, Strunz PP, Schmalzing M, Goebeler M, Schmieder A
Patient-Centered Lupus Erythematosus Mobile Apps: Systematic Search and Cross-Sectional Evaluation by Patients and Physicians
JMIR Mhealth Uhealth 2026;14:e73019
URL: https://mhealth.jmir.org/2026/1/e73019
doi:10.2196/73019

© Tassilo Dege, Antonia Ullmann, Caroline Glatzel, Janik Fleißner, Vanessa Borst, Patrick-Pascal Strunz, Marc Schmalzing, Matthias Goebeler, Astrid Schmieder. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 29.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,

XSL•FO
RenderX

provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Digital Health Apps and Web-Based Platforms to Support the Prevention and Management of Snakebite Envenoming: Scoping Review

Deborah Hosemann[1*], MSc; Oliver Gries[1*], MD; Jade Dean Rae[1], PhD; Thao Vi Tran[2], PhD; Philipp Sprengholz[3], PhD; Lars Korn[4,5], PhD; Thi Thien Thanh Pham[2], MPH; Thi Anh Thu Dang[2], PhD; Benno Kreuels[1], MD

[1]Research Group Neglected Diseases and Envenoming, Bernhard Nocht Institute for Tropical Medicine, Bernhard-Nocht-Straße 74, Hamburg, Germany

[2]Faculty of Public Health, University of Medicine and Pharmacy, Hue, Vietnam

[3]Institute of Psychology, University of Bamberg, Bamberg, Germany

[4]Research Group Health Communication, Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany

[5]Institute for Planetary Health Behaviour, University of Erfurt, Erfurt, Germany

[*]these authors contributed equally

Corresponding Author:
Benno Kreuels, MD
Research Group Neglected Diseases and Envenoming, Bernhard Nocht Institute for Tropical Medicine, Bernhard-Nocht-Straße 74, Hamburg, Germany

Abstract

Background: Neglected tropical diseases disproportionately affect underserved populations, with snakebite envenoming (SBE) remaining one of the most overlooked, despite its significant global burden. Digital health applications (DHAs) offer potential to improve prevention, care, and resource management, especially when integrated into digital health interventions. However, despite growing interest, evidence and structured evaluations are limited, making it difficult to assess their impact without a clear overview of existing tools.

Objective: This scoping review aims to provide the first comprehensive mapping of DHAs for SBE, highlighting their potential to strengthen the World Health Organization (WHO) strategy while underscoring the urgent need for structured evaluation, improved quality, and strategic integration to enhance prevention, treatment, and coordination efforts.

Methods: This review followed the Joanna Briggs Institute and PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines, with a protocol registered on the Open Science Framework. We searched the PubMed database, app stores, and Google for DHAs between September 24 and 26, 2024, addressing snakebite prevention or treatment. To be included, the DHA had to be accessible via the recorded link, contain a description with snakebite-related features (eg, identification, first aid, and treatment), and allow user interaction. Descriptions had to appear in abstracts, app store listings, or website text. Results were grouped by type (mobile- or web-based) and by WHO region. Furthermore, we examined the 2 most common features: first aid and snake identification. First aid content was benchmarked against global guidelines, while identification methods were categorized, and selected artificial intelligence (AI)–based identification apps were exploratively tested using images of medically significant snakes.

Results: A total of 52 eligible results were included, of which 94.2% (49/52) were mobile apps and 5.8% (3/52) were web-based. Regional focus varied, with most apps targeting South-East Asia (n=11, 21.2%), the Americas (n=9, 17.3%), and the Western Pacific (n=5, 9.6%). However, these numbers largely reflect concentration in just a few countries, namely India (n=10, 19.2%), the United States (n=5, 9.6%), and Australia (n=5, 9.6%). The most frequent feature was snake identification support, for example, through photo upload and algorithm-based recognition. However, AI-driven identification often lacked clarity and performed inconsistently in testing. First aid guidance was also common, but only a handful of apps offered comprehensive, evidence-based advice, while others omitted key steps or recommended unsafe practices.

Conclusions: This review provides the first structured evaluation of DHAs for SBE and offers a reproducible framework for assessing digital solutions across neglected tropical diseases. By highlighting key gaps and proposing a foundation for integration into national strategies, it supports the development of equitable, evidence-based digital health innovation in underserved areas.

Trial Registration: OSF Registries osf.io/2zsfu; https://osf.io/2zsfu

(JMIR Mhealth Uhealth 2026;14:e83744) doi:10.2196/83744

XSL•FO
RenderX

KEYWORDS

snakebite; neglected tropical diseases; digital health; mobile health; mHealth; digital health application; artificial intelligence; AI

Introduction

Snakebite envenoming (SBE) is currently classified by the World Health Organization (WHO) as one of the 21 neglected tropical diseases (NTDs). Even within the NTD portfolio, SBE receives limited attention and resources, despite global estimates of 2.7 million SBE cases annually and an estimated 81,000 to 138,000 deaths each year [1-3]. The WHO's snakebite prevention and control strategy aims to halve snakebite deaths and morbidity by 2030, by focusing on four priority areas: (1) empowering and engaging communities; (2) ensuring safe and effective treatment; (3) strengthening health systems; and (4) increasing partnerships, coordination, and resources [1]. Progress has been made since this goal was set in 2019, but it remains slow [4].

Digital health interventions (DHIs), which are specifically designed to address health functions through digital means, can enhance access to treatment, health care provider training, collaboration, and resource distribution [5]. DHIs often use digital health applications (DHAs), such as telemedicine, mobile apps, or web-based platforms [6]. Despite the recognized value of DHIs in advancing universal health coverage, evidence of successful implementation for NTDs, including SBE, remains limited [7]. One of the few well-documented examples is in skin NTDs, where mHealth apps, teledermatology, and artificial intelligence (AI)–driven detection tools have improved diagnosis and management in resource-limited settings [7-9]. Additionally, the WHO's "Skin NTDs app" demonstrates how targeted mHealth apps can support health care workers in making informed clinical decisions, which improve patient care [10].

In SBE, DHIs have the potential to strengthen all 4 WHO priority areas. First, they could empower communities by providing easily accessible first aid and prevention advice. Second, an easy-to-access digital guide outlining safe and effective treatment protocols can enhance the knowledge and capabilities of health care workers. Third, digital reporting applications could be used to collect data on cases that present to the health facilities, helping to fill existing data gaps. Fourth, DHIs could facilitate collaboration among hospitals, nongovernmental organizations, researchers, and patients, improving coordination, treatment, and data sharing. These advancements could streamline prevention, care, and resource management for better outcomes. Despite the rise of DHAs for SBE, there is a lack of evidence, integration within broader DHIs, and structured evaluations. In the absence of a clear overview of existing applications, assessing their impact remains a challenge. Therefore, a comprehensive review of the current landscape for SBE is essential to inform and guide future efforts.

This scoping review systematically identifies currently available DHAs for SBE management, maps their distribution by WHO region, and summarizes their functions. With this review, we aim to provide an overview of the existing DHAs for SBE, laying the foundation to reduce redundancy, foster collaboration, optimize priorities, and improve resource allocation in line with the WHO's goals.

Methods

This review was structured according to the Joanna Briggs Institute's *Manual for Evidence Synthesis on Scoping Reviews*, the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist (Checklist 1), and design guidelines for scoping reviews, as outlined in the protocol registered with the Open Science Framework on August 2, 2024 [11-13].

Search Strategy

We searched PubMed database for literature on DHAs for SBE, as well as the Google Play Store, Apple App Store (using fnd.io [14]), and Google search engine for digital applications related to SBE. Search terms for PubMed database were ("snakebite" OR "snake bite") AND ("digital health" OR "telemedicine" OR "telehealth" OR "mHealth" OR "m-Health" OR "eHealth" OR "E-Health" OR "health app" OR "digital health intervention" OR "digital health technology" OR "healthcare app" OR "mobile health app" OR "remote healthcare" OR "digital diagnosis" OR "digital treatment" OR "digital tool" OR "digital surveillance" OR "digital monitoring" OR "AI"). As Boolean searching did not work in app stores, these were searched by each search term individually ("snakebite," "snake bite," "snake identification," "snake surveillance," "snakebite surveillance," "snakebite diagnosis," "snakebite treatment," "snakebite first aid," "snakebite education," "snakebite prevention"). The internet search was kept broad to try to capture as much as possible ("app" OR "digital health" OR "eHealth" OR "mHealth" AND "snakebite" OR "snake bite"). The complete search strategies and translations are detailed in Multimedia Appendix 1 (Search Queries). All searches were first conducted in English. The Google search was repeated in 13 languages. The first 50 nonsponsored results were selected from each Google search. Non-English results were translated using Google Translate or DeepL [15,16]. Applications known to the author team that were not retrieved through the user-centric search were subsequently added as expert knowledge and screened against the same eligibility criteria.

The searches were performed between September 24 and 26, 2024, from a German IP address. Further details are provided in the published protocol [13].

Selection Criteria

To be eligible for inclusion, the application had to be accessible through the recorded link and include a description that included features relevant to the treatment or prevention of snakebites in humans (eg, snake identification, snake information, first aid, and treatments). These descriptions had to be available in the abstract (for published literature), app store description or images (Google Play Store and Apple App Store), or website

XSL•FO
RenderX

text (Google search). Users had to be able to perform actions in the application through buttons, forms, or menus.

Study Selection

All search results were exported to a data extraction template in Microsoft Excel (Multimedia Appendix 2) with the following information: publisher or source name, app title (app name, website, or article title), link to the result, search type, and language. After removing duplicates within the searches and results with invalid links, the remaining results were screened for inclusion according to the selection criteria. If an application appeared in multiple searches, it was included only once. When identified in both an app store and through a Google search, the app store version was included in the results. If an app was available in both the Apple App Store and Google Play Store, only the latter was included. Excluded results were documented with exclusion reasons. Two reviewers conducted screening and extraction in duplicate.

Data Extraction

For the included applications, we extracted the data policy availability, associated access costs, features offered (grouped according to those with a human focus [prevention, first aid, and advanced treatment], snake focus [snake identification and habitat information], or other focus [eg, snake rescue]), the region and/or country if applicable, the platform type (mobile- or web-based), and whether the application was snake-specific or not. In addition to the data included in the original study protocol, we decided during the search to further record user ratings (count and stars), downloads (available for Google Play Store), app store category, device availability (iOS/Android), release date, and the last update before November 18, 2024, if available.

Data were extracted iteratively from online sources for more than 2 months. This resulted in missing values for some of the apps that became unavailable later in the extraction period. While all core variables, such as features and region, had already been recorded before the apps became inaccessible, assessments of AI and first aid information could not be performed for all apps.

Discrepancies were resolved through discussion and cross-verification. Data extraction was completed by November 19, 2024. Any deviations from the original protocol, along with their rationale, are detailed in Multimedia Appendix 3.

Ethical Considerations

This study synthesized information from publicly accessible sources (including app store listings and publicly available websites) and did not involve human participants, interventions, or individual-level identifiable data. Therefore, ethics approval, informed consent, and participant compensation were not applicable.

Synthesis of Results and Content Evaluation

Because of regional download restrictions and the lack of a standardized evaluation framework, we were unable to assess all app features systematically. Instead, we focused on the 2 most common features: first aid and snake identification.

First aid content was compared with globally applicable guidelines from Health Action International and the Global Snakebite Initiative [17]. The results can be found in Multimedia Appendix 4. For snake identification features, we classified apps into different categories based on the methodology provided by the app provider. In cases where apps reported AI-based identification features, we reviewed publisher descriptions and user feedback.

We then conducted exploratory testing of 3 different AI-based identification apps to assess their performance. We selected and tested 3 apparently credible apps: the single available web-based app, one general reptile identification app, and one snakebite-specific identification app selected at random from the existing options. Apps with unrealistic claims or consistently poor user ratings were excluded, ensuring only reasonably credible applications were included. To cover a wide range of species from different regions, we used 2 images for each of the 12 most medically significant snake species (as defined by the WHO [18]), resulting in 24 test images. These species were (1) *Naja kaouthia*, (2) *Crotalus durissus*, (3) *Calloselasma rhodostoma*, (4) *Dendroaspis polylepis*, (5) *Daboia russelii*, (6) *Bitis arietans*, (7) *Echis ocellatus*, (8) *Naja haje*, (9) *Bothrops asper*, (10) *Pseudonaja textilis*, (11) *Oxyuranus scutellatus*, and (12) *Bungarus candidus*. We used verified photos from iNaturalist (with the appropriate use rights) that reflect what a typical user might upload [19]. Responses for each image were recorded as shown in Multimedia Appendix 5. We divided the results into 5 categories. The first category, “correctly identified,” was used when the species in the image was correctly named (eg, when an image of *Bitis arietans* was identified as *Bitis arietans*). The second category, “misidentified as a nonvenomous snake,” was used when a venomous species was incorrectly labeled as nonvenomous (eg, when *Bitis arietans* was misidentified as *Aspidites ramsayi*). The third category, “misidentified as another venomous snake,” included cases where the species was confused with another venomous species (eg, when *Bitis arietans* was identified as *Daboia russelii*). The fourth category, “misidentified as nonsnake,” was assigned when the image was classified as something other than a snake (eg, a worm). The fifth category, “NA,” was used when no response was generated due to technical issues.

The included applications were grouped according to their platform type (mobile- or web-based), regional focus (WHO regions), and features offered.

Results

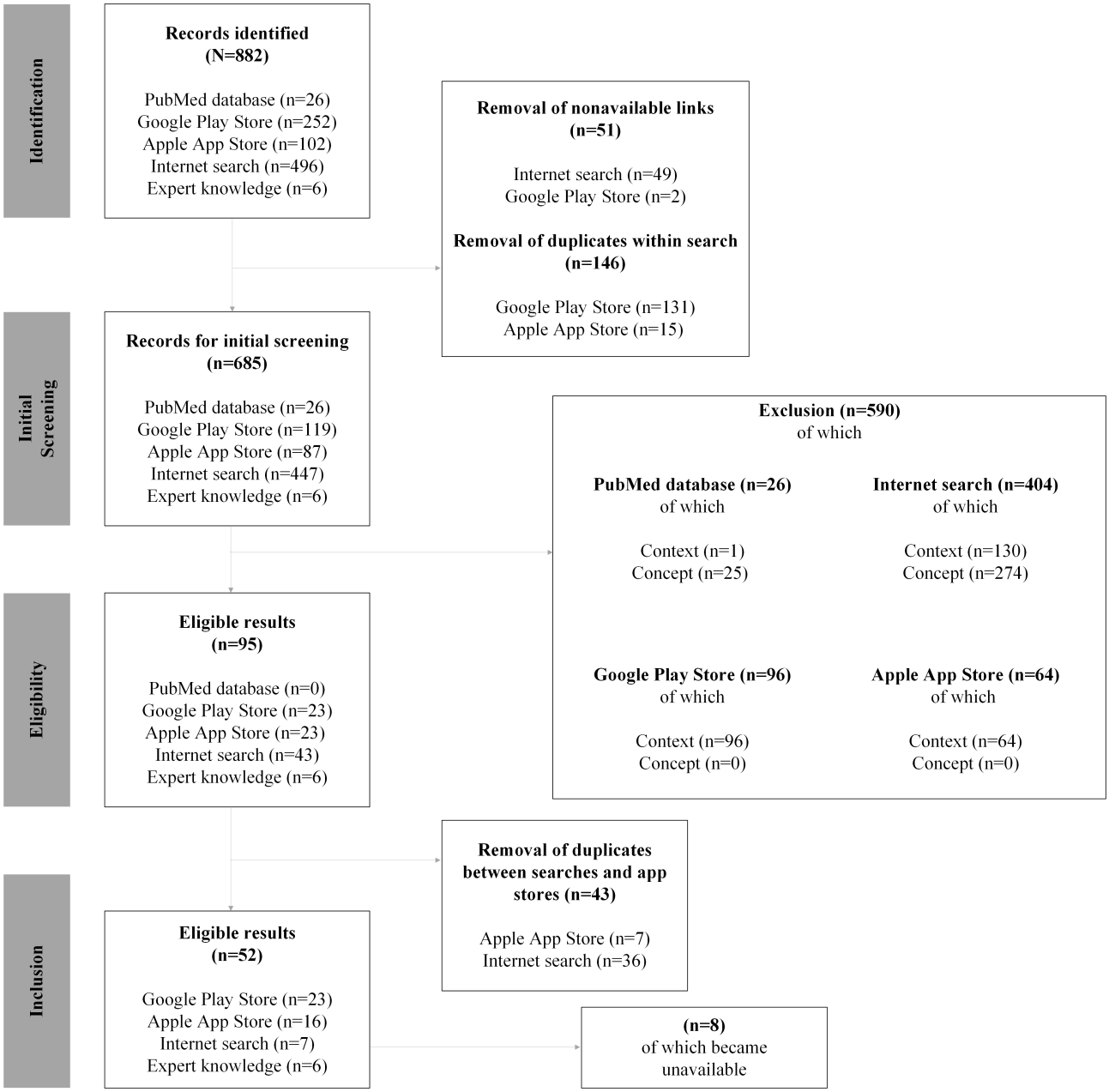
Overview

A total of 882 results were identified, including those from the PubMed database (n=26, 2.9%), Google Play Store (n=252, 28.6%), Apple Store (n=102, 11.6%), Google Search (n=496, 56.2%), and expert knowledge (n=6, 0.7%; Figure 1). After removing duplicates within single searches (n=146, 16.6%) and results with nonfunctioning links (n=51, 5.8%), 685 (77.7%) records were screened using the eligibility criteria. All PubMed database (n=26, 2.9%) results were excluded either because they were not SBE-related or did not describe a DHA (eg, general publications about snakebites). App store results were

XSL•FO
RenderX

excluded for lacking snake or SBE-related content (n=160, 18.1%; eg, snake games), and internet search results were excluded for not being SBE-related or not being interactive DHAs (n=404, 45.8%; eg, news on snakebite victims). After removing cross-search duplicates (n=43, 4.9%), 52 (5.9%) results remained: 39 (39/52, 75%) from app stores (23/39, 59% Google Play Store and 16/39, 41% Apple App Store), 7 (13.5%) from the internet search, and 6 (11.5%) from expert knowledge (Figure 1). Of these, 94.2% (49/52) were mobile apps and 5.8% (3/52) were web-based platforms. Of the 49 mobile apps, the oldest was released in 2015, while 28 (57.1%) apps were released since 2019.

Figure 1. Flowchart of the digital health applications (DHAs) identified during screening. Reasons for exclusion were categorized as (1) context: not relevant for snakebite envenoming; (2) concept: not an interactive DHA; or (3) unavailable: not (or no longer) available (eg, broken link, 404 error, or security warning).



Of the 52 applications, 32 (61.5%) were snakebite-specific, while 20 (38.5%) were general applications that included snake-related content. Key features identified in these apps included snake identification (n=41, 78.8%), first aid advice (n=27, 51.9%), habitat information (n=18, 34.6%), prevention advice (n=8, 15.4%), and advanced treatment advice (n=7, 13.5%). A total of 18 (34.6%) applications included additional features, such as snake removal contacts or lists of hospitals with antivenom.

Mobile Apps

Most mobile apps targeted a specific WHO region: South-East Asia (11/49, 22.4%), the Americas (9/49, 18.4%), Western Pacific (5/49, 10.2%), Africa (3/49, 6.1%), and Europe (1/49, 2.0%); 20 (40.8%) were nonregion specific or the region was unclear. We divided these 20 apps into two categories: (1) worldwide region, which includes apps with a clear global focus as indicated by the provided information or images, and (2) unclear region, which includes apps for which the regional scope

XSL•FO
RenderX

could not be determined from the provided information or images. App details are summarized in Table 1.

XSL•FO
RenderX

Table . Description of mobile apps.

Region and app name	Prevention	First aid	Advanced treatment	Snake identification	Habitat	Other
South-East Asia						
Indian Snakes				✓	✓	
SARPA	✓	✓		✓		✓
SERPENT by Indiansnakes		✓	✓	✓	✓	✓
Snake Bite Awareness App		✓		✓	✓	✓
Snake bite prevention and rescue	✓	✓		✓		✓
Snake Bite Treatment						✓
Snake Helpline		✓		✓		✓
SnakeHub		✓		✓	✓	✓
Snakelens[a]		✓		✓[b]		
Snakepedia		✓	✓	✓	✓	✓
The Snakebite Assistant	✓	✓	✓	✓	✓	
Americas						
Asp Snake Identifier - USA	✓	✓		✓	✓	
GOES Health: Outdoor First Aid[c]		✓		✓		
Redtox App	✓			✓	✓	✓
SERPENT Brazil		✓		✓		✓
Snake Patrol Suriname				✓	✓	
SnakeBite911	✓	✓	✓	✓		
Snakes of North Carolina				✓	✓	
SnakeSnap![c]				✓		
Suriname Snakebite Initiative[a]		✓	✓			
Western Pacific						
Australian Bites and Stings	✓	✓		✓		
Australian Snake ID[c]				✓	✓	
Field Guide to Victorian Fauna				✓		
First Aid		✓				
iFirstAid		✓				
Africa						
ASI Snakes		✓		✓	✓	✓

XSL•FO
RenderX

Region and app name	Prevention	First aid	Advanced treatment	Snake identification	Habitat	Other
eSnakes Southern Africa[c]	✓	✓		✓	✓	
First Aid Africa		✓				
Europe						
Reptiles and Amphibians of Sweden		✓		✓	✓	
Worldwide						
Animal & Plant ID EarthSnap[c]				✓[b]		
EveryScan: Identify Anything[c]				✓[b]		
iNaturalist				✓	✓	
Seek by iNaturalist				✓		✓
Unclear						
Army First Aid[c]		✓				
Automatic Snake Identifier				✓[b]		
Emergency Techniques & Guides[a]		✓				
First Aid and Emergency Techni		✓				
First Aid Guide Offline[c]		✓				
Frog Identifier Reptile ID[c]				✓[b]	✓	
Frst Aid | Emergency Med Aid[a]		✓				
Monosha[a]						✓
Picture Nature: Animal ID[c]				✓[b]	✓	
PIdentify[a]				✓[b]		
Snake ID - reptile identifier[c]				✓[b]		
Snake Identifier[c]				✓[b]		
Snake Identifier[a]				✓[b]	✓	
Snake Identifier: AI Scanner[c]				✓[b]		
Snake name - identify snakes[c]				✓		

XSL•FO
RenderX

Region and app name	Prevention	First aid	Advanced treatment	Snake identification	Habitat	Other
Snake Species Identifier[a]				✓[b]		

[a]Not available anymore.

[b]Artificial intelligence–based photo identification.

[c]Cost involved.

South-East Asia

Eleven apps were identified for South-East Asia (India n=10, 90.9% and Bangladesh n=1, 9.1%), and all were freely accessible. The app for Bangladesh included information on snake species, first aid advice, hospital and snake removal contacts, and educational content [20].

For India, 5 (50%) apps were designed for nationwide use, while 3 (30%) were specific to Kerala, 2 (20%) to Odisha, and 1 (10%) to West Bengal. The countrywide apps featured AI-assisted snake identification [21], field guides [22,23], and apps for snakebite management, including hospital and snake removal contacts [24,25]. The Kerala apps included a field guide with expert support and resource lists [26], first aid advice and emergency contacts [27], and a hospital search app with information on antivenom availability [28]. The apps for Odisha [29] and West Bengal [30] offered local snake information, snakebite symptoms, first aid advice, and snake removal contacts. The West Bengal app also listed hospitals with antivenom. Download numbers were available for all apps, ranging from 500 [28] to 100,000 [26].

Americas

Nine apps were identified for the Americas (United States: n=5, Suriname: n=2, Mexico: n=1, and Brazil: n=1). Seven apps were free, and 2 for the United States had in-app purchases [31,32].

For the United States, 4 apps were designed for nationwide use, and 1 was specific to North Carolina. Three supported snake identification [31,33,34]: 2 field guides that also contained first aid advice [33,34], and 1 that claimed to offer AI-based snake identification with expert verification [31]. One app focused on prevention and first aid specifically for North American pit vipers [35], while the remaining app offered general first aid advice and information on snakebite prevention [32].

The apps for Mexico, Brazil, and Suriname provided first aid advice, expert contacts, and antivenom information. The Mexican app focused on venomous scorpions, spiders, and snakes, as well as expert contacts [36]. The Brazilian app offered SBE victims and medical personnel first aid advice and snake reporting [37]. One Suriname app offered first aid advice and treatment information for SBE victims and medical professionals [38], while the other was a field guide with first aid advice [39]. Download numbers were available for 4 of these apps, ranging from 10 [39] to 100,000 [31].

Western Pacific

All 5 apps for the Western Pacific were specific to Australia, with 1 specific to the state of Victoria. Most were free, with 1 snakebite-specific app available for US $6.99 [40].

Three apps offered snake identification features: 2 field guides [40,41], and 1 with additional educational resources, first aid advice, and prevention [42]. Two others offered general first aid advice, including specific advice for snakebites [43,44]. Download numbers were available for 4 of these apps, ranging from 100 [40] to 50,000 [40,42].

Africa

Three apps were identified for Africa: 2 were free [45,46], and 1 cost US $15.99 [47]. All offered first aid advice. Two apps were designed for Southern Africa (without defining specific countries). These included an interactive, comprehensive field guide [47] and a basic snake identification guide with expert identification and contacts for snake removal [45]. The third was a general first aid app for Africa, which included specific advice for snakebites [46]. Download numbers were available for 2 of these apps, ranging from 1000 [46] to 100,000 [45].

Europe

One app was identified for Europe, specifically designed for Sweden. This free app serves as a field guide offering snake identification, habitat information, and basic first aid advice [48].

Worldwide

All 4 apps identified for worldwide use offered snake identification features. Two were general AI-based identification apps with optional in-app purchases [49,50]; the other 2 were free and used crowdsourced identification, 1 included habitat information [51], and the other used gamification [52]. Download numbers were available for the 2 crowdsourced apps, ranging from 1 million [52] to 5 million [51].

Unclear

Sixteen apps had an unclear regional focus. Most were free, while 6 (37.5%) apps involved a cost. A total of 10 (62.5%) offered snake identification features: 8 (80%) were snake-specific [53-60], and 2 (20%) were general animal apps [61,62]. Five (31.2%) were general first aid apps, which included specific advice for snakebites [63-67], and 1 (6.2%) collected snakebite data and linked snakebite victims to hospital care [68]. Download numbers were available for 6 of these apps, ranging from 100 [68] to 100,000 [65,66].

Web-Based Platforms

Only 1 web-based platform was identified: “AI.Snakes,” which offered advanced treatment information, AI photo-based and clinical assessment–based identification, and regional risk assessments [69].

XSL•FO
RenderX

Two global platforms were added through author team expert knowledge [70,71], as they were not retrieved though the search strategy: the WHO's "Snakebite Information and Data Platform" with information for up to 373 snake species, including distribution maps, antivenom, and clinical guidance; and the "VAPAGuide," a free expert resource with emergency protocols, clinical flow charts, and species data. The web-based platforms and their functions are summarized in Table 2.

Table . Dedicated web-based platforms.

Region and platform name	Prevention	First aid	Advanced treatment	Snake identification	Habitat	Other
Eastern Mediterranean						
AI.Snakes			✓	✓[a]		
Worldwide						
Snakebite Information and Data Platform				✓		✓
VAPAGuide		✓	✓	✓		

[a]Artificial intelligence–based photo identification.

Snake Identification

Of the 41 DHAs offering snake identification features, 17 (17/41, 41.5%) enabled photo uploads for identification, of which 2 (2/17, 11.8%) relied on expert identification [31,45], 2 (2/17, 11.8%) on identification by online communities [51,52], 13 (13/17, 76.5%) claimed to offer AI identification [21,49,50,53-59,61,62,69], 15 (15/41, 36.6%) displayed snake images or photos, 5 (12.2%) were field guides, and 1 (2.4%) used algorithms based on clinical symptoms and described snake morphology (eg, scale patterns) [24]. The identification process was not described in 3 apps (7.3%) [22,30,60].

For 4 AI-based DHAs, developer credentials and algorithm accuracy were unclear [21,54-56]. One claimed accuracy of more than 100% for each identification result [58], 4 apps were no longer available when we attempted to evaluate the feature [21,54-56], and some received user criticism in reviews based on the algorithm's randomness in the results [53].

On the basis of our explorative assessments of 3 selected AI-based snake identification applications, the performance of this feature was highly variable. The "AI.Nature" application failed to return a result for any of the 24 photos used. The "Snake ID – reptile identifier" correctly identified 2 images but failed to detect a snake in 5 images, misidentified the snake as nonvenomous in 10 images, and suggested the wrong venomous species in 7 images. The "Frog Identifier Reptile ID" app correctly identified 13 images but failed to detect a snake in 1 image, misidentified the snake as nonvenomous in 1 image, and suggested the wrong venomous species in 9 images. Performance results and images used are shown in detail in Multimedia Appendix 5.

First Aid

A total of 27 applications featured first aid advice for snakebites. These included 15 (55.6%) snake-specific apps, 9 (33.3%) general first aid apps with snakebite content, 2 (7.4%) general reptile guides, and 1 (3.7%) app on bites and stings. During the evaluation period, 5 (18.5%) apps were not accessible, and 2 (7.4%) were no longer available in the app store. Therefore, we were unable to evaluate them. This resulted in 20 (74.1%) apps being available for download and assessment. Of those, 3 (15%) covered all recommended first aid steps [23,24,26]. Most advised reassuring the bitten person, remaining calm, minimizing movement, and seeking medical care, but many omitted warnings about traditional or unsafe treatment (n=10, 50%) or the importance of moving away from the snake (n=12, 60%). Three (15%) apps incorrectly recommended washing the wound, and the "Army First App" advised killing the snake for identification and using a constricting band. Three (15%) apps provided minimal or no information [64,66,67]. Full details are provided in Multimedia Appendix 4.

Discussion

Principal Findings

This scoping review presents the first structured overview of DHAs for SBE and represents one of the earliest systematic efforts to characterize DHAs for NTDs with a breakdown of available features [13]. We identified 52 DHAs relevant to snakebite treatment or prevention, most of which were offered as snakebite-specific mobile apps. The most common features included snake identification and first aid advice.

Despite the number of apps available, their distribution does not align with the global burden of SBE. High-incidence regions, such as sub-Saharan Africa, remain underrepresented, with only 3 apps available for the region, while low-incidence countries, such as the United States and Australia have 5 apps each. This reflects the broader disparities in funding, research capacity, treatment infrastructure, and epidemiological data [72]. Most African countries lack reliable data on SBE, limiting effective policymaking and resource allocation [73]. This underrepresentation of the African continent extends to the nondigital space. Despite promising initiatives, snakebite management in Africa remains largely underfunded and underresourced [72]. Access to smartphones and reliable internet is increasing across the African continent, making DHAs increasingly useful, particularly for health care workers [74]. Nevertheless, access remains limited, especially in rural populations, which continues to restrict the utility of DHAs for those populations. This contrasts with better rural connectivity

XSL•FO
RenderX

observed in South Asia and the Western Pacific [75]. However, beyond India and Australia, these regions still face limited availability of snakebite-related DHAs.

First aid and snake identification were the most common features. Most apps' first aid features were aligned with established guidelines, and only one app included several contraindicated practices [63]. However, transparency around authorship, clinical validation, and accountability was often lacking, raising safety concerns. These apps could be improved by referencing international guidelines, linking sources, and warning against harmful practices. Because many people self-treat using unsafe methods, validated content could improve outcomes and support the WHO's pillar to empower communities with better knowledge and applications.

Despite reported advances, the availability of AI-based snake identification apps remains limited and user-facing performance is variable. Challenges, such as interspecies similarities, species variation, and poor image quality, make accurate identification difficult. Nevertheless, research prototypes trained on large image datasets have reported high classification accuracy across many species and in real-world images [76-78]. However, even occasional misidentifications pose serious risks if used for medical decisions after snakebites. In our exploratory assessment, using verified images loaded into 3 apps which claimed to use AI for snake identification, outputs were inconsistent, with frequent misidentifications or no results (Multimedia Appendix 5), suggesting limited real-world usability of current AI-claimed snake identification apps. None of the reviewed apps offering AI-based snake identification disclosed details on development or evaluation details (eg, training data, validation datasets, and versioning), and most relied on internally derived confidence scores rather than externally validated outcomes. These gaps highlight the mismatch between research-level AI and published apps [79].

Few apps disclosed their development purpose or process, and many were developed outside health systems, lacked regulatory oversight, and included liability disclaimers, indicating the need for stronger regulation. Several apps also lacked emergency-friendly interfaces, relying on long, unstructured PDFs that pose risks during time-sensitive situations. Despite these limitations, apps with expert input and clear scope hold promise, especially if further developed within regulatory frameworks. Only 1 app was described in a peer-reviewed publication, highlighting the general lack of scientific documentation or transparency in app development [80]. This manuscript was published after our analysis and was therefore not identified through our initial search.

Recommendations

The review identifies 4 priority actions to improve the quality, accessibility, and alignment of DHAs for SBE. Responsibility should be considered on multiple levels: national health authorities and professional societies can define minimum clinical content standards and integration pathways, app marketplaces can require baseline transparency fields as listing criteria, and global actors can support shared taxonomies and registries to improve discoverability and interoperability.

First, enforcing basic transparency and documentation standards, such as publishing development histories, data sources, and use context, would enable consistent assessment. Co-designing apps with target user groups, particularly underserved groups, would ensure relevance and usability. Second, a global digital health registry for SBE and other NTDs could enhance visibility, reduce duplication, support interoperability, and promote uptake. Third, incentivizing open collaboration (eg, co-design, open-source codebases, stakeholder involvement, or version control) would improve usability and trust. Fourth, aligning app development with national digital health strategies and systems would strengthen sustainability and institutional support.

Strengths and Limitations

This study has several limitations. First, reliance on real-time data extraction rather than app downloads may have led to the extraction of incomplete data. However, this only applies to data from apps no longer accessible, illustrating the app marketplace's volatility, particularly in niche or fast-growing areas. Second, the user-centric search strategy may have missed less visible resources. Notably, 2 expert platforms known to the author team were not identified in the search and were only included via expert knowledge [70,71]. The inability to find these platforms, despite using a very broad search strategy and searching in multiple languages, probably reflects challenges that typical users may also face. Visibility of apps and platforms is influenced by algorithmic ranking and varying levels of search engine optimization, and not by the quality of the DHA itself. End users and frontline health workers operating under time pressure and resource constraints are unlikely to find these platforms with a simple search and have even fewer cues to judge whether tools are trustworthy and fit for purpose, an especially acute challenge in neglected disease settings.

Furthermore, our search strategy may have overlooked digital tools without a direct web presence, such as chatbots running on platforms such as WhatsApp [81]. It also excluded other service-based digital solutions, such as remote consultation platforms, including the Remote Envenomation Consultancy Services in Malaysia [82]. Although this is a limitation, the lack of an online presence meant that this type of DHA could not be directly analyzed. Its content was not readily accessible and therefore falls outside the scope of this review, as this was an inclusion criterion. Finally, we did not download, evaluate, or compare all available apps, as currently no standardized framework exists for assessing snakebite apps, and any such comparison would have been inherently subjective. Despite these constraints, the gray-inclusive strategy and registered protocol support reproducibility. Findings align with WHO priorities on transparency, interoperability, and equity in digital health [5,83,84].

This review identified critical gaps in transparency, metadata completeness, standardized evaluation, efficacy metrics, and alignment of digital snakebite apps with health governance frameworks. Many apps lacked documented development processes, references to clinical standards, or integration with national health strategies. These deficits undermine long-term sustainability, impede user trust, and hinder integration into structured care pathways.

XSL•FO
RenderX

Geographic misalignment, particularly underrepresentation in high-burden regions such as sub-Saharan Africa, along with persistent digital inequities in low- and middle-income countries, further highlight the disconnection between needs and app availability. Without targeted efforts to strengthen DHA and address usability barriers in low- and middle-income countries, such apps risk reinforcing exclusion rather than improving access [7]. Infrastructure challenges, regulatory gaps, and poor interoperability further complicate integration into formal care systems.

Despite pervasive gaps in transparency, quality, and integration, the findings from this review provide a replicable foundation for future assessment and strategic planning of digital applications in the field of SBE. Scaling this framework across NTDs and embedding shared standards can support equitable, evidence-based digital innovation in neglected disease areas.

Acknowledgments

The authors would like to thank Anaïs Ludolph, Johanna Holm, Rangarirai Makuku, Ana Castiella García, Roselyn Kang'ara, Quy Quoc Bao Truong, Ahmad Junaedi, Ángela Duarte Pardo, Atishay Mathur, Helen Merin-Sostmann, Mohamed El Houboub, Jarntrah Sappayabanphot, and Jessica Lee for their support with the translations of the search terms. Generative artificial intelligence (AI) tools were used solely for language editing and condensation support under full author supervision. No sections of the manuscript were generated autonomously by AI, and all wording and content were reviewed and approved by the authors.

Funding

The review was supported with core funding from the Bernhard Nocht Institute for Tropical Medicine.

Data Availability

Detailed search terms can be found in Multimedia Appendix 2, and inclusion and exclusion criteria are available in the protocol. The main dataset and the final results generated from this review are available at Open Science Framework [85].

Authors' Contributions

Conceptualization: DH, OG, BK, PS, TATD
Data curation: DH, OG, TVT, TTTP
Formal analysis: DH, OG, JDR
Investigation: DH, OG
Methodology: DH, OG, LK, BK
Project administration: DH, BK
Resources: BK
Software: BK
Supervision: BK
Validation: DH, OG
Visualization: DH, OG, JDR
Writing – original draft: DH, OG, BK, JDR
Writing – review & editing: DH, OG, JDR, TVT, PS, LK, TTTP, TATD, BK
All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication. DH and OG contributed equally as joint first authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search queries.
[PDF File, 117 KB - mhealth_v14i1e83744_app1.pdf]

Multimedia Appendix 2
Data extraction sheet.
[XLSX File, 19 KB - mhealth_v14i1e83744_app2.xlsx]

Multimedia Appendix 3
Adaptations from protocol.
[PDF File, 77 KB - mhealth_v14i1e83744_app3.pdf]

XSL•FO
RenderX

Multimedia Appendix 4
First aid testing.
[PDF File, 122 KB - mhealth_v14i1e83744_app4.pdf]

Multimedia Appendix 5
Artificial intelligence picture testing.
[PDF File, 2508 KB - mhealth_v14i1e83744_app5.pdf]

Checklist 1
PRISMA-ScR checklist.
[PDF File, 747 KB - mhealth_v14i1e83744_app6.pdf]

References

1. Snakebite envenoming -- a strategy for prevention and control. World Health Organization. 2019. URL: https://www.who.int/publications/i/item/9789241515641 [accessed 2026-05-18]
2. Chippaux JP. Snake-bites: appraisal of the global situation. Bull World Health Organ 1998;76(5):515-524. [Medline: 9868843]
3. Gutiérrez JM, Calvete JJ, Habib AG, Harrison RA, Williams DJ, Warrell DA. Snakebite envenoming. Nat Rev Dis Primers 2017 Sep 14;3:17063. [doi: 10.1038/nrdp.2017.63] [Medline: 28905944]
4. Gopalakrishnan M, Tianyi FL, Pinfield T. Uniting against the snakebite challenge: are we there yet? Trans R Soc Trop Med Hyg 2025 Apr 1;119(4):398-399. [doi: 10.1093/trstmh/traf025] [Medline: 40048590]
5. Recommendations on digital interventions for health system strengthening. World Health Organization. 2019. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2026-05-18]
6. Classification of digital interventions, services and applications in health: a shared language to describe the uses of digital technology for health. World Health Organization. 2023. URL: https://www.who.int/publications/i/item/9789240081949 [accessed 2026-05-18]
7. Technology for world elimination of neglected tropical diseases. Lancet Digit Health 2023 Feb;5(2):e51. [doi: 10.1016/S2589-7500(23)00003-1] [Medline: 36707185]
8. Barnowska EJ, Fastenau A, Penna S, Bonkass AK, Stuetzle S, Janssen R. Diagnosing skin neglected tropical diseases with the aid of digital health tools: a scoping review. PLOS Digit Health 2024 Oct;3(10):e0000629. [doi: 10.1371/journal.pdig.0000629] [Medline: 39374195]
9. Carrion C, Robles N, Sola-Morales O, Aymerich M, Ruiz Postigo JA. Mobile health strategies to tackle skin neglected tropical diseases with recommendations from innovative experiences: systematic review. JMIR Mhealth Uhealth 2020 Dec 31;8(12):e22478. [doi: 10.2196/22478] [Medline: 33382382]
10. Skin NTDs app. Google Play. URL: https://play.google.com/store/apps/details?id=com.universaldoctor.skin_ntds&hl=en [accessed 2026-05-18]
11. Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z. JBI Manual for Evidence Synthesis: JBI; 2024. URL: https://jbi-global-wiki.refined.site/space/MANUAL [accessed 2026-05-18]
12. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
13. Hosemann D, Kreuels B, Gries O. Protocol: digital health applications to support the management of snakebite envenoming: a scoping review. OSF (ewms6/overview). URL: https://osf.io/ewms6/overview [accessed 2026-05-18]
14. fnd. URL: https://www.fnd.io/ [accessed 2026-05-18]
15. Google Translate. URL: https://translate.google.com/?hl=de&sl=auto&tl=en&op=translate [accessed 2026-05-18]
16. DeepL. URL: https://www.deepl.com/de/translator [accessed 2026-05-18]
17. Snakebite first aid. Health Action International. URL: https://snakebiteawareness.org/wp-content/uploads/Prevention-Poster-English-Health-Action-International.pdf [accessed 2026-05-18]
18. Control of neglected tropical diseases. World Health Organization. URL: https://www.who.int/teams/control-of-neglected-tropical-diseases/snakebite-envenoming/snakes-gallery [accessed 2025-05-19]
19. iNaturalist. URL: https://www.inaturalist.org [accessed 2025-05-19]
20. . Google Play. URL: https://play.google.com/store/apps/details?id=com.smartsoftware.snakebite&hl=en [accessed 2025-05-13]
21. Snakelens. Google Play. 2024. URL: https://play.google.com/store/apps/details?id=com.laresai.snakelens&hl=en [accessed 2024-11-19]
22. Indian snakes. Google Play. URL: https://play.google.com/store/apps/details?id=net.thenatureweb.snakes.india [accessed 2025-05-13]
23. SnakeHub. Google Play. URL: https://play.google.com/store/apps/details?id=in.snakehub.snakehub [accessed 2025-05-13]

XSL•FO
RenderX

24. The Snakebite Assistant. Google Play. URL: https://play.google.com/store/apps/details?id=com.wowlabz.vapa_guide [accessed 2025-05-13]
25. SERPENT by Indiansnakes. Google Play. URL: https://play.google.com/store/apps/details?id=leopard.tech.lab.big4mapper [accessed 2025-05-13]
26. Snakepedia. Google Play. URL: https://play.google.com/store/apps/details?id=app.snakes [accessed 2025-05-13]
27. SARPA. Google Play. URL: https://play.google.com/store/apps/details?id=ltl.kfdsr&hl=en [accessed 2025-05-13]
28. Snake bite treatment. Google Play. URL: https://play.google.com/store/apps/details?id=com.cobratech.kurian.snakebitehospital&hl=ar [accessed 2025-05-13]
29. Kanti N. Snake Helpline. Google Play. URL: https://play.google.com/store/apps/details?id=com.snakehelpline.app [accessed 2025-05-13]
30. Snakebite Prevention & Rescue. Google Play. URL: https://play.google.com/store/apps/details?id=com.wbhealth.snakebiteprevention&hl=en [accessed 2025-05-13]
31. SnakeSnap!. Google Play. URL: https://play.google.com/store/apps/details?id=co.snakesnap.application [accessed 2025-05-13]
32. GOES: Outdoor Health & Safety. App Store. URL: https://apps.apple.com/us/app/goes-outdoor-health-safety/id1576527901 [accessed 2025-05-13]
33. Asp Snake Identifier - USA. Google Play. URL: https://play.google.com/store/apps/details?id=com.vanillastudio.aspsnakeidentifierusa [accessed 2025-05-13]
34. Snakes of North Carolina. App Store. URL: https://apps.apple.com/us/app/snakes-of-north-carolina/id940870626 [accessed 2025-05-13]
35. SnakeBite911. App Store. URL: https://apps.apple.com/us/app/snakebite911/id1067383282 [accessed 2025-05-13]
36. Redtox App. Google Play. URL: https://play.google.com/store/apps/details?id=com.nmn.redtox&hl=es_MX&gl=US [accessed 2025-05-13]
37. SERPENT-Brazil. Google Play. URL: https://play.google.com/store/apps/details?id=leopard.tech.lab.serpent&hl=gsw [accessed 2025-05-13]
38. Mugjooz NV. Suriname snakebite initiative. App Store. 2024. URL: https://apps.apple.com/us/app/suriname-snakebite-initiative/id1450645507?uo=4&at=11l6hc&ct=fnd [accessed 2025-05-13]
39. Snake Patrol Suriname. Google Play. URL: https://play.google.com/store/apps/details?id=com.sps.snakepatrol&hl=en [accessed 2025-05-13]
40. Australian Snake ID. Google Play. URL: https://play.google.com/store/apps/details?id=com.lucidcentral.mobile.snake_id&hl=en&gl=US [accessed 2025-05-13]
41. Field Guide to the Victorian Fauna. App Store. URL: https://apps.apple.com/us/app/field-guide-to-victorian-fauna/id423945031 [accessed 2025-05-13]
42. Australian Bites and Stings. Google Play. URL: https://play.google.com/store/apps/details?id=com.healthy.AustralianBitesAndStings&hl=en,%20https://apps.apple.com/au/app/australian-bites-and-stings/id765162888 [accessed 2025-05-13]
43. First Aid. Google Play. URL: https://play.google.com/store/apps/details?id=com.viveogroup.firstaidkitselectronicguide&hl=en_IN [accessed 2025-05-13]
44. iFirstAid. Google Play. URL: https://play.google.com/store/apps/details?id=com.ifirstaid [accessed 2025-05-13]
45. Marais J. ASI snakes. Google Play. URL: https://play.google.com/store/apps/details?id=com.app.p4397JF [accessed 2025-05-13]
46. First Aid Africa. Google Play. URL: https://play.google.com/store/apps/details?id=org.mednav.firstaidafrica [accessed 2025-05-13]
47. eSnakes Southern Africa. App Store. URL: https://apps.apple.com/us/app/esnakes-southern-africa/id1490225566 [accessed 2025-05-13]
48. Macdonald S. Reptiles and Amphibians of Sweden. App Store. 2025. URL: https://apps.apple.com/us/app/reptiles-and-amphibians-of-sweden/id777495883 [accessed 2025-05-13]
49. Plant & Animal ID EarthSnap. App Store. URL: https://apps.apple.com/us/app/plant-animal-id-earthsnap/id1590178460 [accessed 2025-05-13]
50. Lens AI & Identifier EveryScan. App Store. URL: https://apps.apple.com/us/app/everyscan-identify-anything/id1663862037?uo=4&at=11l6hc&ct=fnd [accessed 2025-05-13]
51. iNaturalist. Google Play. URL: https://play.google.com/store/apps/details?id=org.inaturalist.android [accessed 2025-05-13]
52. Seek by iNaturalist. Google Play. URL: https://play.google.com/store/apps/details?id=org.inaturalist.seek [accessed 2025-05-13]
53. Automatic Snake Identifier. Google Play. URL: https://play.google.com/store/apps/details?id=identifier.neural.snakes [accessed 2025-05-13]
54. Sudhakar RM. PIdentify. App Store. 2024. URL: https://apps.apple.com/us/app/pidentify/id1630169767?uo=4&at=11l6hc&ct=fnd [accessed 2024-11-19]

XSL•FO
RenderX

55. Diamond detector. Snake IDentifier. Google Play. 2024. URL: https://play.google.com/store/apps/details?id=com.snakegame.snakeid [accessed 2024-11-29]
56. Snake species identifier. Google Play. 2024. URL: https://play.google.com/store/apps/details?id=com.bostanji.snakespeciesidentifier [accessed 2024-11-19]
57. Snake ID - reptile identifier. App Store. URL: https://apps.apple.com/us/app/snake-id-reptile-identifier/id6581482039 [accessed 2025-05-13]
58. Fabre P. Snake Identifier.: App Store: Apple; 2025. URL: https://apps.apple.com/us/app/snake-identifier/id1547206441 [accessed 2025-05-13]
59. Snake Detector: AI Scanner. App Store. URL: https://apps.apple.com/us/app/snake-detector-ai-scanner/id6557079975 [accessed 2025-05-13]
60. Snake name - identify snakes. Google Play. URL: https://play.google.com/store/apps/details?id=name.snake_id [accessed 2025-05-13]
61. Huyen TD. Frog Identifier Frog Sound ID. App Store. URL: https://apps.apple.com/us/app/frog-identifier-reptile-id/id6499125621 [accessed 2025-05-13]
62. Picture Nature: Animal ID. App Store. URL: https://apps.apple.com/us/app/picture-nature-animal-id/id1474585677 [accessed 2025-05-13]
63. Army First Aid. App Store. URL: https://apps.apple.com/us/app/army-first-aid/id326842780 [accessed 2025-05-13]
64. Emergency Techniques & Guides. Google Play. URL: https://play.google.com/store/apps/details?id=appscodevalley.firstaid.emergency.techniques.medicines88 [accessed 2026-05-18]
65. Fatbelly. First Aid and Emergency Techni. Google Play Store: Google. 2025. URL: https://play.google.com/store/apps/details?id=com.fatbelly.firstaidsandemergencytechniques [accessed 2025-05-13]
66. First Aid Guide Offline. Google Play. URL: https://play.google.com/store/apps/details?id=ferdari.primeros_auxilios.english [accessed 2025-05-13]
67. Alpha Z studio. Frst Aid | Emergency Med Aid. Google Play Store. 2024. URL: https://play.google.com/store/apps/details?id=com.alphazstudio.firstaid [accessed 2025-05-13]
68. Monosha. Google Play. 2024. URL: https://play.google.com/store/apps/details?id=com.rmspl.snake.wb.data.ui&hl=en [accessed 2025-05-13]
69. AI Snake. URL: https://app.ai-nature.com [accessed 2025-05-13]
70. Snakebite information and data platform. World Health Organization. URL: https://www.who.int/teams/control-of-neglected-tropical-diseases/snakebite-envenoming/snakebite-information-and-data-platform [accessed 2025-05-13]
71. VAPAGuide. URL: https://www.vapaguide.info [accessed 2026-05-18]
72. Berg P, Theart F, van Driel M, Saaiman EL, Mavoungou LB. Snakebite envenoming in Africa remains widely neglected and demands multidisciplinary attention. Nat Commun 2024 Nov 6;15(1):9598. [doi: 10.1038/s41467-024-54070-y] [Medline: 39505859]
73. GBD 2019 Snakebite Envenomation Collaborators. Global mortality of snakebite envenoming between 1990 and 2019. Nat Commun 2022 Oct 25;13(1):6160. [doi: 10.1038/s41467-022-33627-9] [Medline: 36284094]
74. The state of broadband in Africa. International Telecommunication Union and United Nations Educational, Scientific and Cultural Organization. 2025. URL: https://www.itu.int/dms_pub/itu-s/opb/pol/S-POL-BROADBAND.32-2025-PDF-E.pdf [accessed 2026-05-19]
75. Bahia K. The state of mobile internet connectivity 2024. : GSMA Intelligence; 2024 URL: https://www.gsmaintelligence.com/research/the-state-of-mobile-internet-connectivity-2024 [accessed 2026-05-19]
76. Bolon I, Picek L, Durso AM, Alcoba G, Chappuis F, Ruiz de Castañeda R. An artificial intelligence model to identify snakes from across the world: opportunities and challenges for global health and herpetology. PLoS Negl Trop Dis 2022 Aug;16(8):e0010647. [doi: 10.1371/journal.pntd.0010647] [Medline: 35969634]
77. Norouzzadeh MS, Nguyen A, Kosmala M, et al. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proc Natl Acad Sci U S A 2018 Jun 19;115(25):E5716-E5725. [doi: 10.1073/pnas.1719367115] [Medline: 29871948]
78. Tabak MA, Norouzzadeh MS, Wolfson DW, et al. Machine learning to classify animal species in camera trap images: applications in ecology. Methods Ecol Evol 2019 Apr;10(4):585-590. [doi: 10.1111/2041-210X.13120]
79. Zhang J, Chen X, Song A, Li X. Artificial intelligence-based snakebite identification using snake images, snakebite wound images, and other modalities of information: a systematic review. Int J Med Inform 2023 May;173:105024. [doi: 10.1016/j.ijmedinf.2023.105024] [Medline: 36848781]
80. Sachett JA, Rocha GD, Silva E, et al. SERPENT-Brasil: a technological tool for snakebite management. Rev Panam Salud Publica 2025;49:e39. [doi: 10.26633/RPSP.2025.39] [Medline: 40330807]
81. WhatsApp chatbot for snakebite awareness launched. The Hindu. 2024. URL: https://www.thehindu.com/news/national/karnataka/whatsapp-chatbot-for-snakebite-awareness-launched/article68236141.ece [accessed 2024-05-31]
82. Remote Envenomation Consultancy Services. Malaysian Society on Toxinology. URL: http://mstoxinology.blogspot.com/p/recs.html [accessed 2025-09-08]

XSL•FO
RenderX

83. Regional Action Plan for prevention and control of snakebite envenoming in the South-East Asia 2022–2030. World Health Organization. 2023. URL: https://www.who.int/publications/i/item/9789290209829 [accessed 2026-05-19]
84. Ending the neglect to attain the Sustainable Development Goals: a road map for neglected tropical diseases 2021–2030. World Health Organization. 2021. URL: https://www.who.int/publications/i/item/9789240010352 [accessed 2026-05-19]
85. Hosemann D, Kreuels B, Gries O. Protocol: digital health applications to support the management of snakebite envenoming: a scoping review. OSF (2zsfu/overview). URL: https://osf.io/2zsfu/overview [accessed 2026-05-22]

Abbreviations

AI: artificial intelligence
DHA: digital health application
DHI: digital health intervention
NTD: neglected tropical disease
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
SBE: snakebite envenoming
WHO: World Health Organization

Edited by A Stone, L Buis; submitted 08.Sep.2025; peer-reviewed by C Anyaegbuna, DP Pandey; revised version received 03.Feb.2026; accepted 30.Apr.2026; published 02.Jun.2026.

Please cite as:
Hosemann D, Gries O, Rae JD, Tran TV, Sprengholz P, Korn L, Pham TTT, Dang TAT, Kreuels B
Digital Health Apps and Web-Based Platforms to Support the Prevention and Management of Snakebite Envenoming: Scoping Review
JMIR Mhealth Uhealth 2026;14:e83744
URL: https://mhealth.jmir.org/2026/1/e83744
doi:10.2196/83744

© Deborah Hosemann, Oliver Gries, Jade Dean Rae, Thao Vi Tran, Philipp Sprengholz, Lars Korn, Thi Thien Thanh Pham, Thi Anh Thu Dang, Benno Kreuels. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 2.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Quality and Multifunctionality in Mobile Apps for Gestational Diabetes: Systematic App Review

Qimeng Zhao[1], MSc; Alison Cooke[1,2,3], PhD; Lishan Huang[1], MSc; Yimin Tang[1], MSc; Dawn Dowding[1], PhD

[1]Division of Nursing, Midwifery and Social Work, Faculty of Biology, Medicine and Health, University of Manchester, Room 6.306, Jean McFarlane Building, Oxford Road, Manchester, United Kingdom

[2]Centre for Research & Education Excellence (CeNREE), Royal Stoke University Hospital, University Hospitals of North Midlands NHS Trust, Stoke on Trent, United Kingdom

[3]School of Nursing and Midwifery, Keele University, Keele, United Kingdom

Corresponding Author:
Qimeng Zhao, MSc
Division of Nursing, Midwifery and Social Work, Faculty of Biology, Medicine and Health, University of Manchester, Room 6.306, Jean McFarlane Building, Oxford Road, Manchester, United Kingdom

Abstract

Background: The use of mobile health (mHealth) apps can assist with the management of gestational diabetes (GDM). Although a number of studies have demonstrated their efficacy in improving maternal-fetal outcomes, opinions differ regarding their usability and overall quality. Poorly designed apps, with ill-conceived features or inappropriate content, may pose a threat to patient safety. Nevertheless, very few studies provide in-depth evaluations of app design quality, and the diversity of features and techniques used remains insufficiently explored.

Objective: We aimed to evaluate the quality and multifunctionality of commercially available mHealth apps for GDM.

Methods: This is a systematic app review guided by the TECH (target user, evaluation focus, connectedness, and health domain) framework and the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist. Searches were conducted on the Apple App Store and Google Play. Apps were screened by name, description, and full navigation to identify inclusions. The quality of the apps was evaluated using the Mobile App Rating Scale and IMS Institute for Healthcare Informatics Functionality Score. Multifunctionality of the apps was evaluated using the GDM-adapted features and techniques list developed from the App Behavior Change Scale, NICE (National Institute for Health and Care Excellence) 2015 guidelines, and previous studies. The general features list, which contains instruction, data security, customization, and technical issues, was derived from previous studies.

Results: The search (June 2024) identified 23 commercially available apps from UK app stores. The overall app quality was evaluated to be satisfactory (Mobile App Rating Scale: mean 4.0, SD 0.36; IMS Institute for Healthcare Informatics Functionality Score: mean 5.83, SD 3.03). The multifunctionality evaluation found that the apps had a mean of 17.95 and SD of 7.31 across all 45 items. Overall, our findings suggested that mHealth apps for GDM achieved a certain level of multifunctionality. However, their feature types and supporting digital techniques are relatively basic. The apps focused on education and managing blood glucose control rather than integrating other self-monitoring data and pregnancy-relevant management into their design. The digital techniques used to achieve these features included text and manual operation, rather than other automated features.

Conclusions: This is the first app review to consider the relationship between app features and usability for women with GDM. Future app development should integrate a wide range of pregnancy-relevant information and more automated features and use advanced digital techniques to enable a holistic digital solution for women with GDM.

(JMIR Mhealth Uhealth 2026;14:e76862) doi:10.2196/76862

KEYWORDS

diabetes, gestational; pregnant woman; smartphone; mobile health; mHealth; app; behaviour change; mobile phone

Introduction

Challenges in Gestational Diabetes Self-Management

The self-management of gestational diabetes (GDM) usually includes a series of complex lifestyle behavioral changes such as diet, exercise, and continuous self-testing of blood glucose [1]. This systematic approach aims to help women maintain optimal blood glucose levels and minimize complications for both mother and baby. However, previous research highlights the challenges for women with GDM, who find it hard to understand or follow the instructions from doctors [2]. Women have to make decisions on food choices daily, and this is done through effective cooking strategies [3]. It also challenges

XSL•FO
RenderX

women to understand the labels and nutritional values of food [4]. Guidelines across countries emphasize the importance of strengthening physical activities and regard aerobic activities as acceptable approaches [5]. However, this requires carefully choosing the level of intensity and the types of exercise [5].

Mobile Health Apps Help With Self-Management of GDM

GDM digital technologies have covered various areas of GDM management, providing features including aspects of educational information, health behavior coaching, data recording, and communication interfaces for women [6]. By analyzing the overall effectiveness, previous literature found that GDM digital technologies could provide comparable quality of care to face-to-face visits with health care professionals or midwives, manifesting in heightened maternal health conditions and increased rates of natural childbirth, alongside diminished occurrences of adverse maternal and neonatal outcomes [7-9].

There are various forms of digital technologies, such as mobile apps, webpages, and digital devices, among which mobile apps are one of the most common types of technology used [9]. The rise in global smartphone penetration has led to the evolution of mobile apps, which are now one of the most representative forms of GDM digital technologies [8-10]. Mobile phones possess advantages in terms of accessibility as they are portable devices, thereby facilitating women's access to digital services irrespective of their geographical locations [11,12].

Variation in Perceived Usability of Mobile Health Apps for GDM

However, unlike the overall optimistic results from the evaluation of effectiveness, our previous review and other studies found a variation in pregnant women's views on the app's usability [13,14]. Women had positive views of apps when they supported self-monitoring and viewed them as beneficial tools for obtaining information about GDM, making informed dietary decisions, and facilitating exercise regimens [13,15-17]. The technologies also led to increased levels of adherence, concentration, and satisfaction among women [7,14]. The apps offered platforms for the women and clinicians to keep continuous interaction, fostering the client-clinician relationship [16,18]. This mitigated various concerns that women encounter in managing their condition, demonstrating efficacy in facilitating women's self-management practices [7-9]. In the meantime, apps added to women's emotional burden with problematic features and content [13]. For example, redundant processes and frequent failures in data recording reduced women's motivation for using the technology [18,19]. The differences between the information provided by apps and clinicians confused the women [16]. Evidence suggests that mobile health (mHealth) apps might unintentionally lead to increased unhealthy behavior [20]. mHealth apps designed inappropriately may contain inaccurate or incomplete information that misleads users or provide insufficient support for managing high-risk behaviors, which could affect users' safety [21].

Challenges to Conducting a Comprehensive Evaluation of GDM Apps

To ensure that women are offered apps of the best quality, health care practices face challenges in evaluating and selecting suitable options. The current era is facing the results of unregulated mHealth markets, which flourish in the number of emerging mHealth apps but vary in quality [22]. The increasing study of app quality evaluation has triggered a discussion on the definition of app quality. Studies found that the criteria used for evaluating app quality were heterogeneous [23,24]. In addition, established criteria lack precise definitions, which hinders the assessment of app quality.

Two previous reviews have evaluated the functionality and usability of GDM mHealth apps [6,25]. Both reviews primarily relied on a single quality criterion: the Mobile App Rating Scale (MARS). However, neither study addressed how well app features are adapted to GDM self-management content. Similarly, they offered no practical insights to assist health care professionals in selecting apps that are both functional and user-friendly. What is missing from the current literature is an evaluation of how these apps contain features suitable for GDM care, and guidance that helps health care professionals when selecting usable apps.

Objectives

This study reviewed commercially available mobile apps to explore app quality and multifunctionality, targeted at antenatal care for women with GDM or pre-existing diabetes during pregnancy, available in the UK app market, and in the English language. This systematic evaluation only includes free-download apps and analyzes the available free-of-charge features, providing insights into the types of features that are most accessible for women with GDM. We assessed the quality of the apps using standardized evaluation scales. We evaluated the app's multifunctionality based on app features and content tailored specifically to the GDM population, as well as general features that affect the normal use of apps.

Methods

Study Design

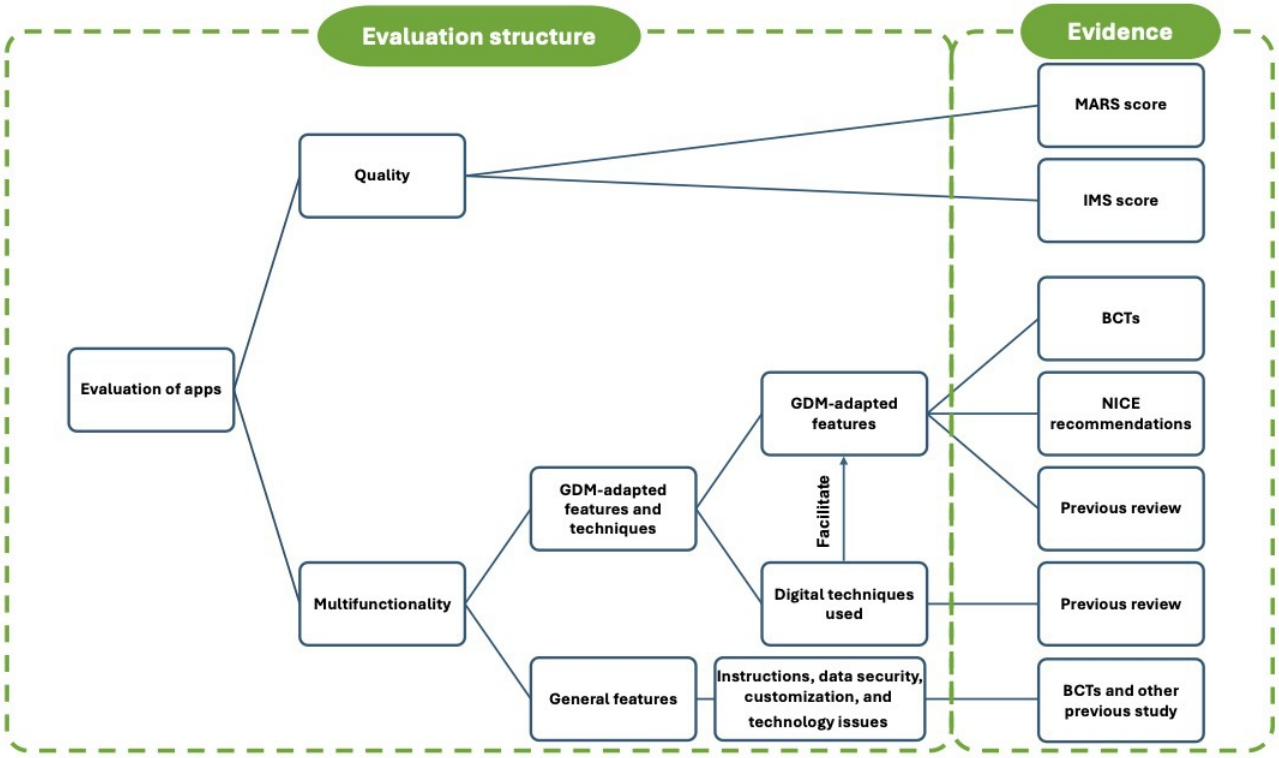
This study followed the TECH (target user, evaluation focus, connectedness, and health domain) framework proposed by Gasteiger et al [26] (2023), which is tailored for the context of reviewing and evaluating the quality of commercially available mHealth apps as opposed to the systematic literature review methodology. Informed by the traditional systematic review process, the TECH framework aims to help conduct reviews on mobile apps by establishing search strategies, the use of eligibility criteria, and a comprehensive quality appraisal. It provides guidance to conduct a search, and has been applied in recent studies to evaluate mHealth apps, including risk assessment and self-management tools, by examining both their functionality and content [27-29].

This study consists of 2 major domains of evaluation: app quality and app multifunctionality. Figure 1 shows the structure of the evaluation elements and the evidence supporting the evaluation. The app quality evaluation in this study aimed to

XSL•FO
RenderX

explore the type of features being used in the included apps and used generalized tools to quantify the level of quality. The measurement was informed by the TECH framework and guidance by introducing both the MARS and the IMS Institute for Healthcare Informatics Functionality Score (IMS). The aggregation of the 2 quality appraisal tools helps to evaluate mobile apps more comprehensively.

Figure 1. Evaluation structure. BCT: behavior change technique; GDM: gestational diabetes; IMS: The IMS Institute for Healthcare Informatics Functionality Score; NICE: National Institute for Health and Care Excellence; MARS: Mobile App Rating Scale.



The multifunctionality evaluation aimed to conduct a GDM context-tailored assessment on the behavior change techniques (BCT) used in the apps, as well as other essential features needed to be enabled for normal app usage. Therefore, there were two elements in the multifunctionality evaluation: (1) the GDM-adapted features and techniques, which refer to the app features aligned with BCTs, along with the digital techniques used to enable the features; and (2) the general features, which refers to the essential features that enable the normal usage of apps, including instructions, data security, customization, and technical issues.

Eligibility Criteria

Following the guidance of the TECH framework, a scoping search on Apple App Store and Google Play in the UK context was run in March 2024 to identify eligibility criteria [26]. This study aims to identify apps that include GDM as one of the targeted user groups and that eligible apps must refer to GDM in their description to qualify for inclusion in this study. From this process, three potential types of mobile apps were identified: (1) mobile apps for GDM management, (2) mobile apps for pregnancy management with additional features for blood glucose control, and (3) mobile apps tailored for diabetes patients but claimed to be suitable for women with GDM. Additionally, the mobile apps were available in the UK context and the English language, free to download, running on Android (Google LLC) or iOS (Apple Inc). The eligibility criteria are listed as follows (Multimedia Appendix 1):

1. Target user (T): mobile apps tailored for GDM, pregnancy, and diabetes users, with GDM as one of the targeted user groups.
2. Evaluation focus (E): content of information, quality, and functionality of the mobile apps that are suitable for users with GDM.
3. Connectedness (C): mobile apps with external devices such as glucose meters and smartwatches.
4. Health domain (H): mobile apps suitable for antenatal behavioral change for women with GDM rather than prevention or postpartum care.

Search

The search keywords "gestational diabetes," "GDM," and "pregnancy diabetes" were applied in the search strategy. The final search was conducted by 3 reviewers on the Apple App Store and Google Play in June 2024. Due to the Apple App Store's online page design, it was not possible to export the search results into screening software to remove duplicates. Additionally, the mobile app search results were subject to frequent changes over time. To avoid unnecessary duplication in the search, all 3 reviewers searched and performed screening on the same day.

Screening

The screening process was guided by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist and the TECH methodology [26,30]. Two steps of

XSL•FO
RenderX

screening were applied, including title screening and full mobile app screening. The mobile app name and description screening process was conducted on the search results pages to identify potential mobile apps for inclusion. We contacted the app companies to access those apps with limited access during the retrieval process. A subsequent full mobile app screening was conducted to check the relevance and feature operability. To include the mobile apps which were usable at the technological level as well as suitable for the targeted population, the full mobile app screening process made a decision of inclusion when (1) the screened mobile app was available in English; (2) no constant technical glitch or crash occurred when operating; and/or (3) the mobile app mentioned GDM at least once in options, information, or features. A snowballing strategy was used after the full mobile app screening process to check the relevant-recommended apps generated within the search results pages of the final-included mobile apps, such as the "you might also like" and "similar apps" pages, to reduce missing mobile apps.

During the screening process, 2 pairs of authors independently screened the search results in Google Play and Apple App Store. The level of agreement between the reviewers was determined using Cohen κ to calculate interrater reliability. To minimize the potential for duplication inherent in app store searches, the analysis was based on the results from a single search term. Discrepancies between the paired authors were discussed and resolved. The results of eligible mobile apps were checked and agreed upon by the other authors. The devices used to search and screen the apps were iPhone 11 Pro Max (iOS 17.5), iPhone 11 (iOS 17.5), iPhone X (iOS 16.7.8), Samsung Galaxy J5 (Android v8.1 Oreo), and Huawei Mate 10 (HarmonyOS 4.0).

Quality Appraisal

Overview

Two quality appraisal tools, which are widely applied for mHealth evaluation, the MARS [26,31] and the IMS tool [32], were used to evaluate the mHealth apps. The MARS (26 items) was used to evaluate the technical design, and the IMS (11 items) to evaluate the number of functions.

The MARS was scored from inadequate to excellent with a corresponding score from 1 to 5 in the 5-point Likert scale. This review followed the recommendation of MARS to calculate MARS scores using the 16 items of app quality rating. The subjective quality and perceived impact items were combined as subjective quality (10 items) for evaluation in this review. The IMS [32] items were coded as 1 per item if app features were presented and otherwise 0 per item, generating a functionality score ranging from 0 to 11 for each app. The overall mapping scores representing the frequencies of applied features were presented in a radar graph as recommended [32].

Quality appraisal was conducted by 2 pairs of reviewers. The agreement rate between the reviewers using different scales was calculated, with interclass correlation coefficients (ICCs) for MARS results [33] and Cohen κ for the IMS results. The level of agreement in ICCs, which ranges between 0 (no agreement) and 1 (perfect agreement), was reported in the value and followed the level categorization as either poor (<0.40), fair (0.40 - 0.59), good (0.60 - 0.74), or excellent (0.75 - 1.0) [34]. The calculation was conducted using IBM SPSS (version 29; IBM Corp).

Data Extraction

Data extraction focused on two categories: (1) app characteristics, which included basic information such as the app operating system, version, and size; and (2) multifunctionality, which included GDM-adapted features and techniques, as well as general features. A list of uncategorized app features and techniques was initially extracted and used to inform the GDM-adapted features and techniques list and the general features list. The two lists contributed to the evaluation of the app's multifunctionality (Multimedia Appendix 2). After the development of the evaluation lists, the apps were reviewed again by 1 author and extracted for the items on the lists. The extracted features and techniques of individual apps were checked and discussed to reach an agreement in the author group.

The GDM-adapted features and techniques lists included types of app features (GDM-adapted features) and digital techniques (GDM-adapted techniques). GDM-adapted features refer to the app feature content, which served as a digital solution to clinical interventions for GDM management. For example, blood glucose data recording served as a digital solution to a paper diary for blood glucose documentation. GDM-adapted techniques are the digital techniques used to enable the app features. For example, Bluetooth (Bluetooth Special Interest Group) techniques and manual recording are both digital techniques used to enable the data recording app feature. General features referred to instructions, data security, and customization, which were extracted based on other studies [35,36].

The development of the GDM-adapted features and techniques list included app features aligned with BCTs [36], NICE (National Institute for Health and Care Excellence) 2015 recommendations [37], and our previous review on women's preferences [13]. BCTs include strategies such as education, goal setting, self-monitoring, feedback, and reinforcement that encourage healthy behaviors [36]. The 2015 NICE guideline recommends managing GDM through individualized diet and exercise, along with blood glucose monitoring [37]. It also includes pregnancy management guidance by including frequent antenatal visits and monitoring of fetal growth and well-being [37]. Our previous review found that women with GDM valued digital tools with education, personalized guidance, and easy data management [13]. It highlights the need for more usable technologies that support self-monitoring of GDM [13]. Integrating these sources created a clear framework for assessing app features, making sure the GDM-adapted features and techniques list reflects behavior change, clinical care, and women's needs when managing GDM. Additionally, some basic digital techniques, such as text-based educational information and data recording approaches, were added to the final list of app features without direct indications from literature or recommendations.

XSL•FO
RenderX

Data Analysis and Presentation

The extracted features and techniques of apps were analyzed and presented in aggregated data. Categorical data, including app characteristics items, GDM-adapted features, GDM-adapted techniques, and general features, were summarized as frequencies and percentages. Continuous data, including app size, MARS score, and IMS score, were presented as means with SDs. App characteristics, MARS score, GDM-adapted features and techniques, and general features were summarized in tables to provide a clear presentation of results. The IMS score was presented in a radar chart to illustrate the distribution of IMS functionalities across 10 domains.

Ethical Considerations

This study did not involve human participants, animals, or any collection of personally identifiable information. All data analyzed were publicly available in the Apple App Store and Google Play Store. Accordingly, institutional review board approval was not required. As no human subjects were enrolled, informed consent was not applicable.

Results

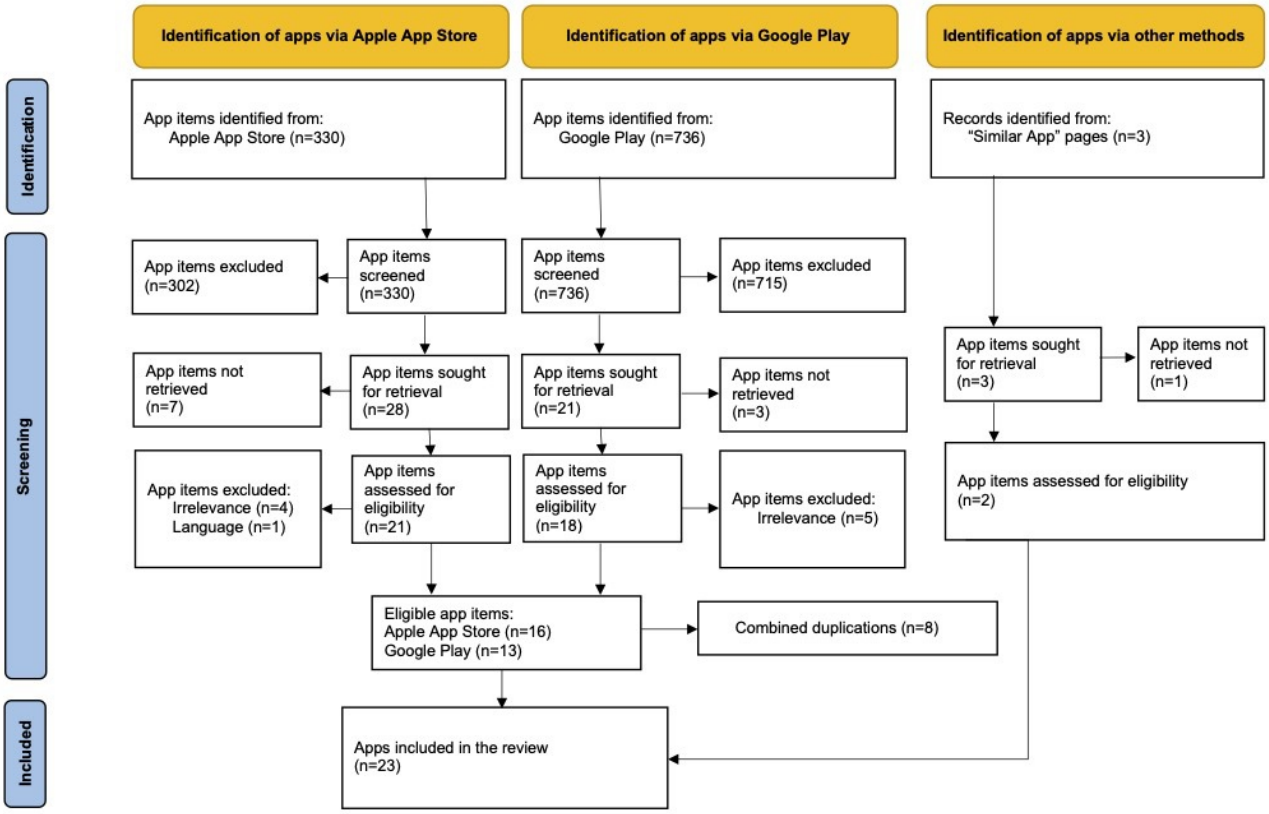
Search Results

The search identified 1066 apps, 330 from the Apple App Store and 736 from Google Play. Two paired authors independently screened the search results in both app stores, the Apple App Store, resulting in 28 eligible mobile apps from the Apple App Store, and 21 in Google Play. The interrater reliability in both pairs indicates a minimum of substantial agreement, with Cohen κ values of 0.62 and 0.91. The title and description screening process resulted in 49 potentially eligible mobile apps in the two app stores. The mobile apps available in both operating systems were not combined at this step.

The 49 eligible mobile apps were retrieved and installed before the full mobile app screening process. This process led to 10 exclusions due to activation (n=7), subscription (n=1), and technical crash issues (n=2). The retrieving, installing, and checking process led to 39 (n=21 in iOS, n=18 in Android) inclusions for the full mobile app screening. The full app screening led to 10 exclusions (n=5 in iOS, n=5 in Android). The reasons for exclusion included irrelevance (n=9) and language (n=1). Perfect agreement was identified between the 2 pairs of authors. Mobile apps available in both operating systems (n=8) were then combined in number, resulting in 21 eligible mobile apps.

A subsequent snowballing process generated the inclusion of 2 mobile apps (available in both operating systems). The search and screening process contributed to the eventual inclusion of 23 eligible mobile apps for this study. Figure 2 illustrates the PRISMA diagram for the screening process (Multimedia Appendix 3 for the enlarged PRISMA diagram).

Figure 2. PRISMA diagram. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



App Characteristics

Of the 23 reviewed apps, 56% (n=13) were adapted to both operating systems (iOS and Android), followed by 22% (n=5) of the apps which were exclusively for iOS, and 22% (n=5) of the apps which were exclusively for Android. Most (n=16, 70%) of the apps were updated in 2024. On average, the apps were of mean 66.43 and SD 53.46 MB, ranging between 6.2 and

XSL•FO
RenderX

223.4 MB. All the apps were free to download, with 48% (n=11) free to access all features, and 52% (n=12) required a subscription. The major targeted user groups of the apps were people with diabetes (n=12, 52%), women with GDM (n=7, 30%), followed by women who are pregnant (n=3, 13%) and women with diabetes who are pregnant (n=1, 4%).

We categorized the apps into 4 major types based on the characteristics of features: multipurpose, self-monitoring, educational, and communicational. Self-monitoring apps were identified by their ability to assist women's behavior for self-monitoring, such as recording data and reminding them of self-tests, with a total of 5 apps. Educational apps were identified by their ability to offer women information, with a total of 5 apps. Communicational apps were identified by their ability to provide a platform for women to communicate with their peers and health care professionals, with a total of 2 apps. The apps that used more than one of the mentioned characteristics of features were identified as multipurpose apps (n=11). All apps (n=23) were free to download, with all apps that were tailored for GDM or pregnant users (n=11) being free to access all features; meanwhile, the other 12 apps for general diabetes users needed to subscribe for full access. Table 1 shows the characteristics of the reviewed apps.

Table . App characteristics.

Evaluations	Values
Operating systems, n (%)	
Both	13 (56)
iOS	5 (22)
Android	5 (22)
Last update, n (%)	
2024	16 (70)
2023	4 (17)
2022 and before	3 (13)
Size (MB), n (%)	
0 - 50	11 (48)
50 - 100	7 (30)
>100	5 (22)
Targeted population, n (%)	
Diabetes	12 (52)
GDM[a]	7 (30)
Pregnancy	3 (13)
Diabetes in pregnancy	1 (4)
App type, n (%)	
Multipurpose	11 (48)
Educational	5 (22)
Self-monitoring	5 (22)
Communicational	2 (9)
Price, n (%)	
Subscription for full access	12 (52)
Free full access	11 (48)
MARS[b] quality score, mean (SD)	4.0 (0.37)
IMS[c] score, mean (SD)	5.83 (3.10)

[a]GDM: gestational diabetes.

[b]MARS: Mobile App Rating Scale.

[c]IMS: The IMS Institute for Healthcare Informatics Functionality Score.

XSL•FO
RenderX

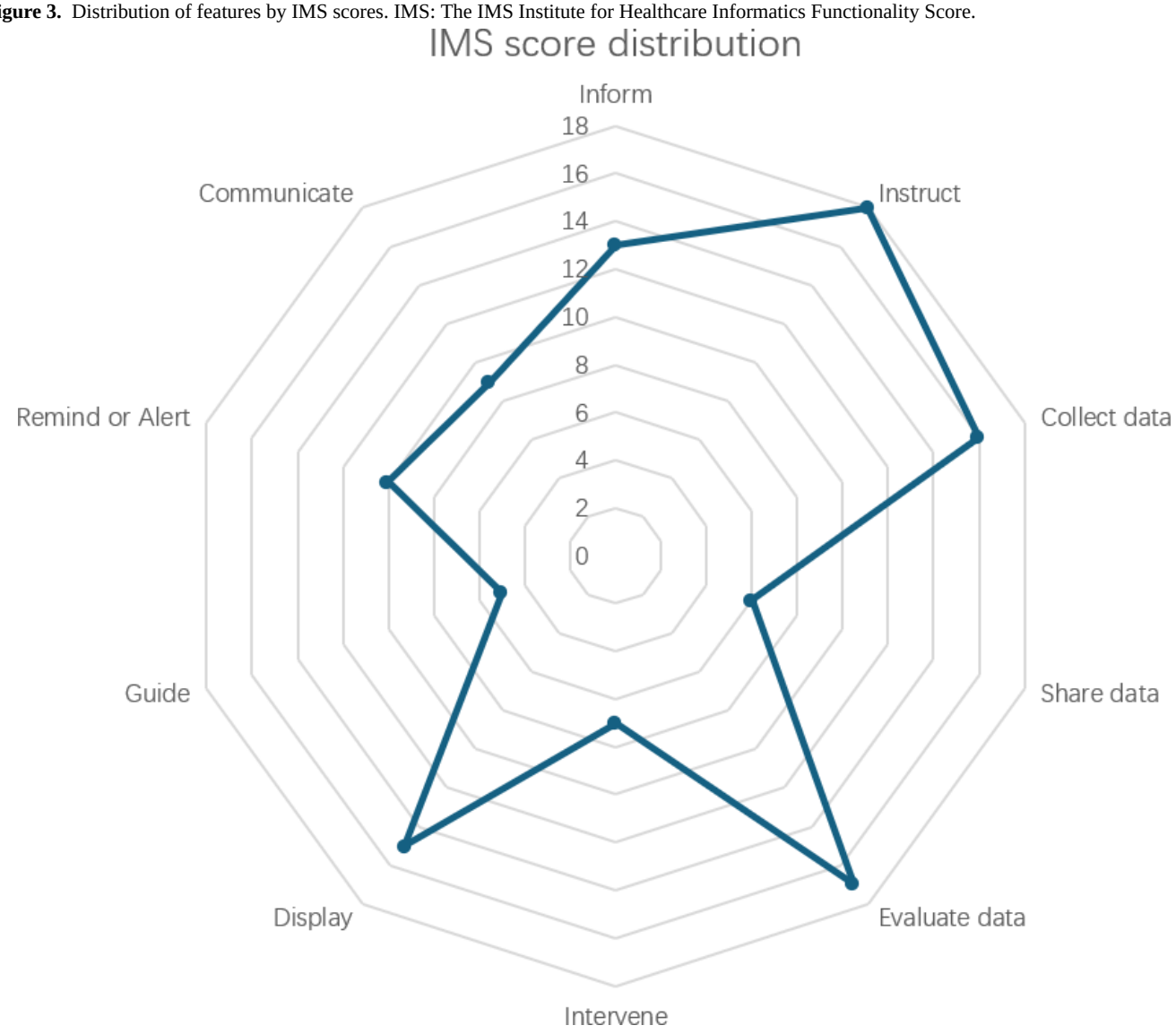
Quality Appraisal

Overview

Overall, the mean MARS score was 4.0 and SD 0.36, and the mean IMS score was 5.83 and SD 3.03 (see details in Multimedia Appendix 4). The mean MARS quality score indicates a satisfactory level of quality, which appears to be higher than the average MARS quality score reported for mHealth apps in the literature (mean 4.0, SD 0.36 vs mean 3.51, SD 0.71) [38]. Regarding individual MARS quality score domains, the MARS engagement domain, which is relevant to coaching and personalization features, scored the lowest (mean 3.56, SD 0.60) compared to other MARS domains (MARS functionality: mean 4.48, SD 0.72, MARS esthetics: mean 4.12, SD 0.48, MARS information: mean 4.02, SD 0.44). Apps were also evaluated based on their different types of functionality using the MARS quality score: the multipurpose apps, educational apps, self-monitoring apps, and communicational apps. Multipurpose (mean 4.06, SD 0.44) and self-monitoring (mean 4.02, SD 0.25) apps appeared to score higher than other types of apps (educational apps: mean 3.92, SD 0.37, communicational apps: mean 3.84, SD 0.47).

The mean IMS score indicated a moderate level of the number of app features. Multipurpose (mean 7.91, SD 2.12) and self-monitoring (mean 6.60, SD 1.14) apps tended to score higher than the other two types of apps. Figure 3 shows the radar graph of the distribution of features scored by the IMS. Among all the 10 IMS features, instructions (n=18), data management features (n=18), and information (n=13) were the predominant features used by the 23 reviewed apps. The guide feature scored lowest compared to other IMS features (n=5).

Figure 3. Distribution of features by IMS scores. IMS: The IMS Institute for Healthcare Informatics Functionality Score.



Reviewers had substantial agreement on their evaluation of iOS apps; IMS (κ=0.834, 95% CI 0.74-0.93, *P*<.001) and MARS (ICC=0.808, 95% CI 0.742 - 0.858) scores. A similar level of agreement was observed for Android apps, with IMS (κ=0.851, 95% CI 0.76-0.95, *P*<.001) and MARS (ICC=0.755, 95% CI 0.67 - 0.82).

Multifunctionality

The analysis of features and techniques was based on the exploration of features that would need to be integrated into an app to assist specifically with GDM. We incorporated the BCTs [36], NICE recommendations [37], and our previous review on women's preferred features [13]. The process generated a list

XSL•FO
RenderX

of GDM-adapted features and techniques (Table 2, details in Multimedia Appendix 5). The features were categorized into 5 domains: education, data management, coaching features, communication features, and pregnancy management.

XSL•FO
RenderX

Table . GDM[a]-adapted features and techniques (N=23).

GDM-adapted features and techniques	Evidence[b]	Frequencies of applied features
Education (n=18)		
Educational content features		
Knowledge about GDM diagnosis	N, P	10
Treatment types for GDM	N	7
Low GI[c] index food choices	B, N	15
Food recipes	B	4
Evidence-based information	B, P	11
Self-monitoring skill	B, N, P	8
Educational information delivery techniques		
Text	__[d]	18
Pictures or graphs	P	9
Videos	P	4
Data management (n=16)		
Data management content features		
Blood glucose	N, P	14
Food	N, P	8
Exercise	N, P	7
Abnormal symptom reporting	N	4
Medication	N	6
Data recording techniques		
Manual (number or text)	—	14
Automatic glucose data transmission	P	4
Automatic exercise data transmission	P	3
Food or exercise database	P	7
Image recognition techniques	P	3
Data visualization techniques		
Tables with text	—	11
Graphs	P	15
Combination charts	P	3
Other data management features		
Digital feedback	B, N, P	14
Data export	B	8
Personal-tailored information or suggestions	B, N, P	3
Coaching features (n=11)		
Reminders	B, P	9
Compliance motivator	B, P	5
Rewarding for milestone achievements	B	3
Communication features (n=9)		
Communication with HCPs[e]	B, N, P	6
Communication with peers	B, P	6
Pregnancy management features (n=5)		

XSL•FO
RenderX

GDM-adapted features and techniques	Evidence[b]	Frequencies of applied features
Pregnancy progression education and management	N, P	4
Fetal monitoring	N, P	3

[a]GDM: gestational diabetes.

[b]Evidence: B: behavior change techniques; N: NICE (National Institute for Health and Care Excellence) recommendations; P: women's preferred features.

[c]GI: glycemic index.

[d]Not applicable.

[e]HCP: health care professional.

Table 3 summarizes the predominant GDM-adapted features and techniques applied by the reviewed apps. The analysis reveals a strong emphasis on educational content (n=18) and data management (n=16) within the mobile apps. The educational components of the apps focus on delivering content critical for GDM management, mainly including knowledge about GDM diagnosis (n=10) and low GI index food choices (n=15). Data management predominantly focuses on collecting blood glucose data, with over 50% (n=14) of apps, rather than food (n=8) and exercise tracking (n=7).

XSL•FO
RenderX

Table . Predominant applied GDM[a]-adapted features and techniques.

App names	Predominant educational features		Predominant data management features	Predominant techniques		
	GDM diagnosis	Food choices	Blood glucose	Text information delivery	Manual data recording	Text data visualization
GDm-Health	✓	✓	✓	✓	✓	✓
Malama	✓	✓	✓	✓	✓	✓
MyManis	✓	✓	✓	✓	✓	✓
Flora	✓	✓		✓		
hug + u			✓	✓	✓	✓
myFetalLife			✓	✓		
Blood Sugar - Diabetes App		✓	✓	✓	✓	✓
Blood Sugar Tracker - Diabetes	✓		✓	✓	✓	✓
Diabetes:M	✓	✓	✓	✓	✓	✓
DiabTrend		✓	✓	✓	✓	
MyNetDiary			✓	✓	✓	✓
Gestational Diabetes Diet		✓		✓		
glukoGest	✓			✓		
Pregnant with diabetes	✓	✓		✓		
NewToType2	✓	✓		✓		
Pingoo	✓	✓		✓		
Gestational Diabetes Tracker			✓		✓	✓
Carbs & Cals		✓				✓
Diabetic diary-Glucose tracker			✓		✓	✓
forDiabetes			✓		✓	✓
mySugr			✓		✓	
Hlgedi		✓		✓		
Diabetes Forum		✓		✓		

[a]GDM: gestational diabetes.

Text-based educational information was nearly ubiquitous, and core tracking features—particularly for blood glucose—were widely implemented. Text-based information delivery was the most common mode, used in over 50% (n=18) of apps. Manual data recording was predominant (used in 14 apps), and 12 used tables (text) to visualize the data. Among all 23 apps, Malama (a multifunctional, GDM-focused app) and Diabetes:M (a multifunctional, diabetes-focused app) cover all the predominant features and techniques.

We identified the features and techniques applied by fewer than 4 apps as the least applied features. The least addressed features included the provision of food recipes (n=4), abnormal symptom reporting (n=4), rewarding for milestone achievements (n=3), and pregnancy progression and education management (n=4). The least addressed techniques included video information delivery (n=4), automatic blood glucose data transmission (n=4), automatic exercise data transmission (n=3), image recognition techniques (n=3), combination charts (n=3), and personal-tailored information or suggestions (n=3).

General Features

Overall, most reviewed apps used general features which ensured normal app usage, with 23 apps addressing data security, 22 apps addressing instructions, and 19 apps addressing personalization. There were 5 apps identified as having technical issues that affect normal use, or distractions that disrupt user interactions. General features and technical issues are summarized in Multimedia Appendix 6.

XSL•FO
RenderX

Some general features were applied universally by the apps. These include consenting to privacy policies (n=23) and asking permission to access data (n=23), providing a quick step-by-step introduction (n=16) when users launch the apps. For personalization, collecting users' demographic data (n=16), asking their preferences for measurement units (n=14), and enabling customized features (n=16), such as reminder timing, were used. Very few apps mentioned the location of data storage (n=5) and the techniques (n=4) used to protect user data.

Additionally, of the 23 evaluated apps, there were 7 with issues that could affect normal user experience, including 5 multipurpose apps (MyManis, hug + u, myFetalLife, Blood Sugar - Diabetes App, and Diabetes:M) and 2 self-monitoring apps (Carbs & Cals and Diabetic diary-Glucose tracker). There were three major issues among these apps: (1) frequent technical glitches, (2) data management issues, and (3) frequent advertisements. Data management issues existed in 4 multipurpose apps, with frequent failure or errors of data recording (MyManis and Diabetes:M), limitations in the data recording feature when inputting decimals and time points (hug + u), and failure of displaying the recorded data in 1 app (myFetalLife). Frequent advertisements took place in 3 apps (Blood Sugar - Diabetes App, Carbs & Cals, and Diabetic diary-Glucose tracker), with advertisements showing up frequently when moving between screens or inputting data, leading to a sense of distraction. Technical glitches and crashes happened to 1 multipurpose app (myFetalLife). While this app contained various features suitable for the management of GDM, technical issues took place frequently and caused crashes and failures when recording the data.

Usable Apps

This review evaluated the app functionality based on the generated lists of GDM-adapted features and techniques, as well as general functionality, collectively with 45 evaluation items on features and content. On average, the apps applied a mean of 17.95 and SD of 7.31 of all 45 items across domains. For GDM-adapted features and techniques, the apps applied a mean of 10.82 and SD of 5.31 of the 32 items. For general functionality evaluation, the apps applied a mean of 7.13 and SD of 2.58 of the 13 items.

We considered apps usable when they contained features and content higher than the average value in each domain and the total average value, with no distractions or technical issues (Table 4). These 6 apps were either categorized in the multipurpose (n=4) or the self-monitoring (n=2) app type in this review.

Table . Usable apps and evaluation.

	App type	Health care domain	MARS[a] quality	IMS[b] score	Multifunctionality and content	General functionality
Full score or number of features	N/A[c]	N/A	5	11	32	13
Average score or addressed features, mean (SD)	N/A	N/A	4.0 (0.36)	5.83 (3.03)	10.82 (5.31)	7.13 (2.58)
Malama	Multipurpose	GDM[d]	4.63	9	24	9
GDm-Health	Multipurpose	GDM	4.47	8	13	11
MyNetDiary	Multipurpose	Diabetes	4.40	10	16	12
DiabTrend	Multipurpose	Diabetes	4.03	9	20	9
MySugr	SMBG[e] app	Diabetes	4.13	8	14	11
forDiabetes	SMBG app	Diabetes	4.3	6	11	9

[a]MARS: Mobile App Rating Scale.

[b]IMS: The IMS Institute for Healthcare Informatics Functionality Score.

[c]N/A: not applicable.

[d]GDM: gestational diabetes.

[e]SMBG: self-monitoring of blood glucose.

Discussion

Overview

This study aimed to provide a comprehensive evaluation of the quality and multifunctionality of commercially available apps on their cash-free features for GDM self-management. By evaluating app quality using the MARS quality and IMS scales, the results showed that the overall app quality is at a satisfactory level. However, regarding individual MARS domains, the MARS engagement domain, which contains coaching and personalized content, was scored lowest among other domains. Following this, the multifunctionality of apps was evaluated by two elements: (1) GDM-adapted features and techniques, namely the app features that fitted into GDM management context, and the corresponding digital techniques used to enable the features, explored the variety of features used by the reviewed apps; (2) general features, which included instruction, data security, customization, and technical issues, explored the app features regarded as essential for basic app usage.

XSL•FO
RenderX

Primary Findings: Basic Features, Limited Content, and the Need for Advanced Design

The reviewed apps predominantly focused on GDM educational content and blood glucose data recording. In addition, the digital techniques used to enable such features were relatively basic. Most apps relied on text-based approaches for content presentation and required manual operation for data management. Moreover, their content reflected only the most fundamental requirement for GDM, focusing primarily on blood glucose management rather than addressing the broader pregnancy experience.

One of our primary findings is the narrow focus on blood glucose monitoring in the design of many apps. Many of the apps included in this review offered features that could support blood glucose self-management, such as educational resources on self-test skill education and dietary guidance, as well as data recording, digital feedback, and graphic visualization. These features were also identified in our previous study as being preferred by women with GDM, as they facilitated self-management [13]. For example, having educational information in a single app improved easy access and helped women recall information [13]. Initial digital feedback using colored labels, combined with graphic visualization, helped the women interpret their data and adjust their diet according to blood glucose trends [13]. However, most apps lacked a multidimensional approach to GDM management. Even though diet and physical activity are both essential for stabilizing blood glucose levels, very few apps in this review included relevant content and features.

Our findings suggest that features relevant to pregnancy progression management were lacking in available apps. In this study, only a small number of apps provided information or features to support pregnancy management and fetal monitoring. There were only 5 apps that included features or information relevant to pregnancy, such as informing about baby growth over time. The features and information about pregnancy and fetal well-being were more likely to be available in those apps designed for pregnancy care. However, the functionality of these pregnancy-tailored apps in blood sugar control was relatively underdeveloped, typically due to technical issues related to data recording and display. These findings indicated an imbalance in focus in app design, with limited integration of broader pregnancy-relevant needs.

In addition, automated features, including automated data transmission and app-generated behavior change suggestions, were used by only a few apps. Similar patterns are also observed in other mHealth apps for pregnancy care, where basic features and techniques still dominate the feature composition [39-41]. These apps tended to focus on either collecting users' personal and physiological data or providing normative and comprehensive information, rather than proactively facilitating women's behavior changes [39-41]. Automated features, such as rewarding on milestone achievements, identifying barriers for behavior change, or providing feedback for behavioral adjustment, are rarely incorporated into the design of pregnancy-related apps [39,40]. Instead, these tasks are sometimes managed through interactions with health care professionals and handled manually [41].

This indicates that the apps currently available and free to download appear to lag behind the evolving needs of women and health care professionals. Women seek tailored guidance for managing their blood glucose, which enhances women's self-awareness and supports their autonomy in self-management [13]. This is usually achieved through effective communication with health care professionals, where the women can receive one-to-one advice and personalized feedback on their blood glucose levels, diet, and lifestyle adjustment [42,43]. However, limited opportunities for antenatal consultations and the heavy workload of health care professionals have often resulted in limited advice on continuous behavior changes [42,43].

Implications for Designing Guidelines and Recommendations

Our findings highlight a gap between the potential of technology and its real-world application. Frontier automated techniques are advancing at an astonishing pace, such as deep learning, decision trees, and reinforcement learning, which have been explored for predicting blood glucose levels and providing personalized treatment recommendations [44,45]. These tools can continuously process data such as blood glucose levels, ketonuria, and dietary adherence, identifying patterns and making real-time adjustments [44]. However, our review revealed that many frontier technologies remain largely inaccessible to women with GDM.

To address this gap, it is essential to integrate broader pregnancy progression management into a condition-specific management app that requires thoughtful design to ensure relevance across different stages of pregnancy and to balance general maternal health management with condition-focused support. As recommended by the 2015 NICE guidance, pregnancy progression management is an important component of GDM management [37]. This requires additional attention on fetal growth and well-being, indicating that additional screening test results and fetal self-observation data recording should be considered [37]. Additional evidence found that women with GDM preferred mHealth technologies that were closely aligned with their pregnancy needs, such as reminders for antenatal visits and pregnancy stages, and they particularly valued educational information related to their pregnancies [16,46].

The findings emphasize the potential of digital tools to provide support for individualized care. It calls for greater integration of automated technologies into app design. This requires co-design approaches involving women and health care professionals to ensure that future apps move from basic features to personalized, interactive, and context-based support for GDM management. Additionally, further exploration is needed to evaluate the automated app feature designs regarding their usability, effectiveness, and safety in real-world contexts.

Recommendations for Development

Given the wide variation in features and techniques available in current GDM apps, and the range of factors that may influence app development, we adopted an evidence-based approach to summarize broadly applicable recommendations

XSL•FO
RenderX

for the development of apps for women with GDM (Textbox 1). These recommendations outline options for core functional design, while also identifying optional advanced features that may be incorporated where appropriate.

XSL•FO
RenderX

Textbox 1. App development recommendations.

Education

Information about gestational diabetes (GDM):

- Include information on the mechanism of GDM, its relevant risks, and impact on pregnancy. The information should aim to help women recognize the importance of monitoring their blood glucose levels.

Dietary management information:

- The rationale for dietary management and basic dietary strategies should be explained, with links to practical and reliable resources, such as guidance on low-glycemic index food choices.

Physical activity information:

- Provide information about the appropriate types of exercise for pregnant women, optimal timing for exercise, the importance of physical activity, and considerations for ensuring safety.
- Information should be provided on appropriate exercise duration and intensity, conditions in which exercise is not suitable, and when exercise should be stopped if certain symptoms occur.

Self-monitoring skills:

- Provide information about how to perform self-monitoring, including capillary blood glucose testing skills, appropriate timing of testing (for example, fasting, before or after meals, and before bedtime), and evidence-based target ranges for blood glucose levels.

Pregnancy-relevant information:

- Provide information on pregnancy progression, including the different stages of pregnancy and corresponding information about the baby, such as fetal growth.

Basic techniques:

- Text information delivery.

Advanced techniques:

- Graphs, pictures, and videos in information delivery.

Development considerations:

- The primary aim of providing information is to increase women's awareness of self-monitoring and, in turn, improve their skills. Information should be clear, easy to understand, and sufficient to support understanding.
- Aside from basic information, consider practical resources, such as food recipes or exercise tutorial videos, that support behavior changes.
- Resources provided should be evidence-based and reliable.

Data management

Types of data:

- Ensure blood glucose data is accurately recorded and clearly visualized in the app. Consider other essential data, including dietary intake, physical activity, symptoms, medication usage, fetal growth, and fetal well-being, to be recorded within the app.

Basic techniques:

- Manual data recording using text, data visualization using tables, and initial data evaluation by preset normal values.

Advanced techniques:

- Automated data transmission (eg, Bluetooth and image recognition techniques), graphic data visualization to show trends, combination charts to cross-check different types of data, and personalized suggestions based on recorded data.

Development considerations:

- Enable easy data recording, prioritizing automated digital techniques over manual text entry, and provide appropriate indicators and options to improve efficiency.
- Considering digital data analysis techniques to evaluate patients' recorded data, and where appropriate, to offer initial suggestions to support patients.

Communication

XSL•FO
RenderX

Features:

- Communication with health care professionals and peers.

Basic techniques:

- Easy-access contact buttons to start phone calls or send text messages.

Advanced techniques:

- Built-in chat box, online forum.

Development considerations:

- Ensure access to health care professionals within the app.
- The features should be designed to help address the women's questions in real time.
- Specific staffing roles should be considered to support and manage communication modules within the app.

Coaching, customization, and technical issues

Reminding and adherence:

- Basic reminders should be provided for blood glucose testing and other relevant activities.
- More advanced features may include reminders when tests are missed, motivational messages to acknowledge achievements, and rewards linked to milestones.
- Development considerations: the aim of developing such features should be to enable dynamic and intelligent tracking of women's behaviors, using reminders and notifications to help prevent unhealthy behaviors, support lifestyle adjustment, and encourage women's behavior change.

Customization:

- Recommended features: demographic data recording, personalized measurement units, customizable normal data ranges, multilanguage options, customizable self-monitoring goals, and a customizable data visualization dashboard.
- Development considerations: the diverse needs, demographic background, and levels of literacy of the women should be considered in app development.

Technical issues:

- Ensure the data details are accurately captured in the app, including numerical precision (for example, decimals), measurement units, and time of data entry.
- Ensure data visualizations are clear and well-organized, and that every recorded data point can be easily viewed.
- Minimize or eliminate connectivity issues and app crashes.

Strengths and Limitations

The strengths of this review include its standardized process and comprehensiveness of evaluation. Standardized evaluation tools were used to measure the quality and multifunctionality of mHealth apps for GDM. In addition, it extends the TECH framework by Gasteiger et al [26] by introducing a more context-based, user-informed, and evidence-informed approach, integrating BCTs, NICE recommendations, and women's preferences, enabling a detailed evaluation of app multifunctionality.

However, several limitations should be considered when interpreting the findings. First, the search was conducted only once on a single day. While this minimizes duplication of included apps, it may have resulted in an incomplete capture of all available apps due to the dynamic nature of app store listings, which can be influenced by advertising and other factors. Second, the decision to restrict the search to UK app stores, while ensuring consistency with UK-based NICE guidelines, may have limited the scope and breadth of the review by excluding insights from apps available in other regions. Third, our findings reflect the status of apps in June 2024. Given the high turnover and frequent updates of apps, their scores, features, and technical issues may have changed postreview, which may affect reproducibility.

Conclusion

This study explored the quality and multifunctionality of commercially available apps for GDM self-management, focusing on the cash-free features, following the TECH and PRISMA frameworks. Our findings highlight that GDM apps should provide more interactive features and support comprehensive pregnancy care, rather than being limited to glucose management and using basic digital techniques. Valuable insights can be drawn from other models of remote pregnancy care, especially those implemented under unavoidable circumstances, such as in rural settings or during the COVID-19 pandemic, where integration of data was required to collect from remote monitoring approaches [47,48]. On one hand, when designing digital technologies for a specific pregnancy complication, it is important to integrate general

XSL•FO
RenderX

pregnancy-related data, such as self-reported symptoms, so that they can be managed alongside the target condition [47]. On the other hand, efforts should be made to enable the in-person services and advice to be available online, ensuring that women receive adequate support while engaging in self-management [48]. Taken together, the evidence highlights an emerging research and design agenda for future mHealth app design—one that prioritizes the integration of comprehensive pregnancy progression management and data interconnectivity to better meet the complex requirements of GDM management.

Acknowledgments

We would like to thank Dr Helen Hawley-Hague for generously lending us the required device.

Authors' Contributions

QZ designed this study, with the support of DD and AC. QZ, YT, and LH reviewed the apps. QZ analyzed the data and wrote the first draft of this paper. QZ, DD, and AC participated in the interpretation of this study's findings and the preparation of this paper. All of the authors approved the final paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Eligibility criteria guided by the TECH framework. TECH: target user, evaluation focus, connectedness, and health domain.
[DOCX File, 19 KB - mhealth_v14i1e76862_app1.docx]

Multimedia Appendix 2
Evidence of evaluation lists.
[XLSX File, 17 KB - mhealth_v14i1e76862_app2.xlsx]

Multimedia Appendix 3
Enlarged PRISMA diagram. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.
[PDF File, 60 KB - mhealth_v14i1e76862_app3.pdf]

Multimedia Appendix 4
Quality appraisal.
[XLSX File, 12 KB - mhealth_v14i1e76862_app4.xlsx]

Multimedia Appendix 5
GDM-adapted features and techniques. GDM: gestational diabetes.
[XLSX File, 20 KB - mhealth_v14i1e76862_app5.xlsx]

Multimedia Appendix 6
General features.
[XLSX File, 14 KB - mhealth_v14i1e76862_app6.xlsx]

References

1. Martis R, Crowther CA, Shepherd E, Alsweiler J, Downie MR, Brown J. Treatments for women with gestational diabetes mellitus: an overview of Cochrane systematic reviews. Cochrane Database Syst Rev 2018 Aug 14;8(8):CD012327. [doi: 10.1002/14651858.CD012327.pub2] [Medline: 30103263]
2. Tzotzis L, Hooper ME, Douglas A, et al. The needs and experiences of women with gestational diabetes mellitus from minority ethnic backgrounds in high-income nations: a systematic integrative review. Women Birth 2023 Mar;36(2):205-216. [doi: 10.1016/j.wombi.2022.08.006] [Medline: 36038477]
3. Karavasileiadou S, Almegwely W, Alanazi A, Alyami H, Chatzimichailidou S. Self-management and self-efficacy of women with gestational diabetes mellitus: a systematic review. Glob Health Action 2022 Dec 31;15(1):2087298. [doi: 10.1080/16549716.2022.2087298] [Medline: 35867537]
4. Doran F, Davis K. Gestational diabetes mellitus in Tonga: insights from healthcare professionals and women who experienced gestational diabetes mellitus. NZ Med J 2010;123(1326):59-67 [FREE Full text]
5. Evenson KR, Barakat R, Brown WJ, et al. Guidelines for physical activity during pregnancy: comparisons from around the world. Am J Lifestyle Med 2014 Mar;8(2):102-121. [doi: 10.1177/1559827613498204] [Medline: 25346651]

XSL•FO
RenderX

6. Nguyen M, Hossain N, Tangri R, et al. Systematic evaluation of canadian diabetes smartphone applications for people with type 1, type 2 and gestational diabetes. Can J Diabetes 2021 Mar;45(2):174-178. [doi: 10.1016/j.jcjd.2020.07.005] [Medline: 33127288]
7. Eberle C, Stichling S. Effects of telemetric interventions on maternal and fetal or neonatal outcomes in gestational diabetes: systematic meta-review. JMIR Diabetes 2021 Aug 27;6(3):e24284. [doi: 10.2196/24284] [Medline: 34448717]
8. Li SY, Ouyang YQ, Qiao J, Shen Q. Technology-supported lifestyle interventions to improve maternal-fetal outcomes in women with gestational diabetes mellitus: a meta-analysis. Midwifery 2020 Jun;85:102689. [doi: 10.1016/j.midw.2020.102689] [Medline: 32193015]
9. Xie W, Dai P, Qin Y, Wu M, Yang B, Yu X. Effectiveness of telemedicine for pregnant women with gestational diabetes mellitus: an updated meta-analysis of 32 randomized controlled trials with trial sequential analysis. BMC Pregnancy Childbirth 2020 Apr 6;20(1):198. [doi: 10.1186/s12884-020-02892-1] [Medline: 32252676]
10. Global smartphone penetration rate as share of population from 2016 to 2024. Statista. URL: https://www.statista.com/statistics/203734/global-smartphone-penetration-per-capita-since-2005 [accessed 2026-01-20]
11. Khalil C. Understanding the adoption and diffusion of a telemonitoring solution in gestational diabetes mellitus: qualitative study. JMIR Diabetes 2019 Nov 28;4(4):e13661. [doi: 10.2196/13661] [Medline: 31778118]
12. Skar JB, Garnweidner-Holme LM, Lukasse M, Terragni L. Women's experiences with using a smartphone app (the Pregnant+ app) to manage gestational diabetes mellitus in a randomised controlled trial. Midwifery 2018 Mar;58:102-108. [doi: 10.1016/j.midw.2017.12.021] [Medline: 29329023]
13. Zhao Q, Cooke A, Aurizki G, Dowding D. Women's experiences and needs in the use of digital technologies for the management of gestational diabetes: an integrative systematic review. Midwifery 2025 Feb;141:104262. [doi: 10.1016/j.midw.2024.104262] [Medline: 39662131]
14. Safiee L, Rough DJ, Whitford H. Barriers to and facilitators of using eHealth to support gestational diabetes mellitus self-management: systematic literature review of perceptions of health care professionals and women with gestational diabetes mellitus. J Med Internet Res 2022 Oct 27;24(10):e39689. [doi: 10.2196/39689] [Medline: 36301613]
15. Al Hashmi I, Alsabti H, Al Omari O, Al Nasseri Y, Khalaf A. Development, feasibility and acceptability of a self-efficacy-enhancing smartphone application among pregnant women with gestational diabetes mellitus: single-arm pilot clinical trial. BMC Pregnancy Childbirth 2022 Apr 23;22(1):358. [doi: 10.1186/s12884-022-04684-1] [Medline: 35461221]
16. Edwards KJ, Bradwell HL, Jones RB, Andrade J, Shawe JA. How do women with a history of gestational diabetes mellitus use mHealth during and after pregnancy? Qualitative exploration of women's views and experiences. Midwifery 2021 Jul;98:102995. [doi: 10.1016/j.midw.2021.102995] [Medline: 33784541]
17. Safiee L, Rough D, George P, Mudenha R. Baseline perceptions of women with gestational diabetes mellitus and health care professionals about digital gestational diabetes mellitus self-management health care technologies: interview study among patients and health care professionals. JMIR Hum Factors 2023 Dec 19;10:e51691. [doi: 10.2196/51691] [Medline: 38113070]
18. Hirst JE, Mackillop L, Loerup L, et al. Acceptability and user satisfaction of a smartphone-based, interactive blood glucose management system in women with gestational diabetes mellitus. J Diabetes Sci Technol 2015 Jan;9(1):111-115. [doi: 10.1177/1932296814556506] [Medline: 25361643]
19. Given JE, Bunting BP, O'Kane MJ, Dunne F, Coates VE. Tele-mum: a feasibility study for a randomized controlled trial exploring the potential for telemedicine in the diabetes care of those with gestational diabetes. Diabetes Technol Ther 2015 Dec;17(12):880-888. [doi: 10.1089/dia.2015.0147] [Medline: 26394017]
20. Gajecki M, Berman AH, Sinadinovic K, Rosendahl I, Andersson C. Mobile phone brief intervention applications for risky alcohol use among university students: a randomized controlled study. Addict Sci Clin Pract 2014 Jul 2;9(1):11. [doi: 10.1186/1940-0640-9-11] [Medline: 24985342]
21. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020 Feb 1;27(2):330-340. [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
22. Chan S, Torous J, Hinton L, Yellowlees P. Towards a framework for evaluating mobile mental health apps. Telemed J E Health 2015 Dec;21(12):1038-1041. [doi: 10.1089/tmj.2015.0002] [Medline: 26171663]
23. Nouri R, Kalhori SRN, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inform Assoc 2018 Aug 1;25(8):1089-1098. [doi: 10.1093/jamia/ocy050] [Medline: 29788283]
24. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422]
25. Kalhori SRN, Hemmat M, Noori T, Heydarian S, Katigari MR. Quality evaluation of english mobile applications for gestational diabetes: app review using Mobile Application Rating Scale (MARS). Curr Diabetes Rev 2021;17(2):161-168. [doi: 10.2174/1573399816666200703181438] [Medline: 32619173]
26. Gasteiger N, Dowding D, Norman G, et al. Conducting a systematic review and evaluation of commercially available mobile applications (apps) on a health-related topic: the TECH approach and a step-by-step methodological guide. BMJ Open 2023 Jun 12;13(6):e073283. [doi: 10.1136/bmjopen-2023-073283] [Medline: 37308269]

XSL•FO
RenderX

27. Cargnin ZA, Schneider DG, Souza MD, Vargas MDO, Tourinho FSV. Low back pain self-management mobile applications: a systematic review on digital platforms. Rev Esc Enferm USP 2024;58:e20230326. [doi: 10.1590/1980-220X-REEUSP-2023-0326en] [Medline: 38875500]
28. Chavez-Ecos FA, Chavez-Ecos R, Vergara Sanchez C, Chavez-Gutarra MA, Agarwala A, Camacho-Caballero K. Mobile health apps for cardiovascular risk assessment: a systematic review. Front Cardiovasc Med 2024;11:1420274. [doi: 10.3389/fcvm.2024.1420274] [Medline: 39376625]
29. Safeer V S M, Gupta P, Behl S, Bansal D, Sahu JK. Mobile health applications for epilepsy in Indian app stores: a systematic review and content analysis using the mobile app rating scale. Epilepsy Res 2024 Mar;201:107331. [doi: 10.1016/j.eplepsyres.2024.107331] [Medline: 38442549]
30. Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021 Mar 29;372:n160. [doi: 10.1136/bmj.n160]
31. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR mHealth uHealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849]
32. Aitken M, Gauntlett C. Patient apps for improved healthcare: from novelty to mainstream. : IMS Institute for Healthcare Informatics; 2013 URL: https://ignacioriesgo.es/wp-content/uploads/2014/03/iihi_patient_apps_report_editora_39_2_1.pdf [accessed 2026-01-20]
33. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86(2):420-428. [doi: 10.1037/0033-2909.86.2.420]
34. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol 2012;8(1):23-34. [doi: 10.20982/tqmp.08.1.p023] [Medline: 22833776]
35. Alfawzan N, Christen M, Spitale G, Biller-Andorno N. Privacy, data sharing, and data security policies of women's mHealth apps: scoping review and content analysis. JMIR mHealth uHealth 2022 May 6;10(5):e33735. [doi: 10.2196/33735] [Medline: 35522465]
36. McKay FH, Slykerman S, Dunn M. The app behavior change scale: creation of a scale to assess the potential of apps to promote behavior change. JMIR mHealth uHealth 2019 Jan 25;7(1):e11130. [doi: 10.2196/11130]
37. Diabetes in pregnancy: management from preconception to the postnatal period. NICE. 2015. URL: https://www.nice.org.uk/guidance/ng3 [accessed 2026-01-20]
38. Lef H, Swoboda W, Schobel J, Hieber D, Holl F. Insights into the quality of mobile health apps: preliminary results of an analysis of MARS scores. Stud Health Technol Inform 2024 Aug 22;316:420-421. [doi: 10.3233/SHTI240438] [Medline: 39176767]
39. Tassone C, Keshavjee K, Paglialonga A, Moreira N, Pinto J, Quintana Y. Evaluation of mobile apps for treatment of patients at risk of developing gestational diabetes. Health Informatics J 2020 Sep;26(3):1983-1994. [doi: 10.1177/1460458219896639] [Medline: 31912754]
40. Hayman MJ, Alfrey KL, Waters K, et al. Evaluating evidence-based content, features of exercise instruction, and expert involvement in physical activity apps for pregnant women: systematic search and content analysis. JMIR mHealth uHealth 2022 Jan 19;10(1):e31607. [doi: 10.2196/31607] [Medline: 35044318]
41. Evans K, Donelan J, Rennick-Egglestone S, Cox S, Kuipers Y. Review of mobile apps for women with anxiety in pregnancy: maternity care professionals' guide to locating and assessing anxiety apps. J Med Internet Res 2022 Mar 23;24(3):e31831. [doi: 10.2196/31831]
42. Sahu B, Babu GR, Gurav KS, et al. Health care professionals' perspectives on screening and management of gestational diabetes mellitus in public hospitals of South India - a qualitative study. BMC Health Serv Res 2021 Feb 12;21(1):133. [doi: 10.1186/s12913-021-06077-0] [Medline: 33579259]
43. Utz B, Assarag B, Lekhal T, Van Damme W, De Brouwere V. Implementation of a new program of gestational diabetes screening and management in Morocco: a qualitative exploration of health workers' perceptions. BMC Pregnancy Childbirth 2020 May 24;20(1):315. [doi: 10.1186/s12884-020-02979-9] [Medline: 32448233]
44. Albert L, Capel I, García-Sáez G, Martín-Redondo P, Hernando ME, Rigla M. Managing gestational diabetes mellitus using a smartphone application with artificial intelligence (SineDie) during the COVID-19 pandemic: much more than just telemedicine. Diabetes Res Clin Pract 2020 Nov;169:108396. [doi: 10.1016/j.diabres.2020.108396] [Medline: 32890548]
45. Lu HY, Ding X, Hirst JE, et al. Digital health and machine learning technologies for blood glucose monitoring and management of gestational diabetes. IEEE Rev Biomed Eng 2024;17:98-117. [doi: 10.1109/RBME.2023.3242261] [Medline: 37022834]
46. Yee LM, Leziak K, Jackson J, et al. Patient and provider perspectives on a novel mobile health intervention for low-income pregnant women with gestational or type 2 diabetes mellitus. J Diabetes Sci Technol 2021 Sep;15(5):1121-1133. [doi: 10.1177/1932296820937347] [Medline: 32627582]
47. Kariman SS, van den Heuvel JFM, Adriaanse BME, Oepkes D, Bekker MN. The potential of tele-ultrasound, handheld and self-operated ultrasound in pregnancy care: a systematic review. Prenat Diagn 2025 Jun;45(7):906-920. [doi: 10.1002/pd.6679] [Medline: 39390612]

XSL•FO
RenderX

48. Moise IK, Ivanova N, Wilson C, Wilson S, Halwindi H, Spika VM. Lessons from digital technology-enabled health interventions implemented during the coronavirus pandemic to improve maternal and birth outcomes: a global scoping review. BMC Pregnancy Childbirth 2023 Mar 20;23(1):195. [doi: 10.1186/s12884-023-05454-3] [Medline: 36941565]

Abbreviations

BCT: behavior change technique
GDM: gestational diabetes
ICC: interclass correlation coefficient
IMS: The IMS Institute for Healthcare Informatics Functionality Score
MARS: Mobile App Rating Scale
mHealth : mobile health
NICE: National Institute for Health and Care Excellence
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
TECH: target user, evaluation focus, connectedness, health domain

Edited by M DeJonckheere; submitted 09.May.2025; peer-reviewed by C Obianyo, DO Alabi, F Carrouel, M Gasmi, PD Haghighi, S Mohanadas; revised version received 03.Dec.2025; accepted 03.Dec.2025; published 05.Feb.2026.

Please cite as:
Zhao Q, Cooke A, Huang L, Tang Y, Dowding D
Quality and Multifunctionality in Mobile Apps for Gestational Diabetes: Systematic App Review
JMIR Mhealth Uhealth 2026;14:e76862
URL: https://mhealth.jmir.org/2026/1/e76862
doi:10.2196/76862

© Qimeng Zhao, Alison Cooke, Lishan Huang, Yimin Tang, Dawn Dowding. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 5.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Mobile Apps for Oncology Health Care Professionals: Mapping and Assessment Study

David Liñares[1,2,3], PhD; Iolie Nicolaidou[4], PhD; Andreas Charalambous[4,5], PhD; Daniela Cabutto[6], MSc; Deborah Moreno-Alonso[7], MD; Clara Madrid Alejos[8], MSc; Norbert Couespel[9], MSc; Noemí López-Rey[1], MD, PhD; María José Fernández-Domínguez[1], MD, PhD; Carme Carrion[3,6,10], MPH, Prof Dr; Ana Clavería[1,3], MD, PhD

[1]Área Sanitaria de Vigo, I-Saúde Group, Instituto de Investigación Sanitaria Galicia Sur, Servizo Galego de Saúde, Centro de Salud de Rosalía de Castro, Vigo, Spain

[2]Galician Agency for Health Technology Assessment, Avalia-t, Galician Agency for Health Knowledge Management, Santiago de Compostela, Galicia, Spain

[3]Research Network in Chronicity, Primary Care and Health Promotion (Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud), Barcelona, Catalonia, Spain

[4]Nursing Department, Cyprus University of Technology, Limassol, Cyprus

[5]University of Turku, Turku, Finland

[6]eHealth Lab Research Group, School of Health Science, Universitat Oberta de Catalunya, Barcelona, Spain

[7]Àrea suport integran en càncer, Institut del Càncer i Malalties de la Sang, Hospital Clínic de Barcelona, Barcelona, Catalonia, Spain

[8]e-Oncologia and Cancer Education Research Unit, Institut Català d'Oncologia, Barcelona, Spain

[9]European CanCer Organisation, Brussels, Belgium

[10]School of Medicine, Universitat de Girona, Girona, Catalonia, Spain

Corresponding Author:
Carme Carrion, MPH, Prof Dr
School of Medicine
Universitat de Girona
Emili Grahit 77
Girona, Catalonia, 17071
Spain
Phone: 34 630983328
Email: mcarrionr@uoc.edu

Abstract

Background: The use of mobile apps in oncology has been expanding rapidly, encompassing prevention, treatment, and patient support. These technologies hold significant potential to improve care delivery and enhance the efficiency of health care services. However, their integration into clinical practice faces important challenges. A key issue lies in the difficulties health care professionals (HCPs) encounter when selecting apps that adequately meet their specific needs and comply with appropriate standards of quality and clinical effectiveness. This lack of robust evidence on the availability, adoption, and evaluation of mobile apps designed for cancer care professionals not only hinders their wider adoption but also restricts their potential to serve as reliable tools in oncology practice.

Objective: This study aims to map the landscape of free mobile apps for cancer prevention, treatment, therapy, or support for HCPs, and assess the quality of the apps identified.

Methods: A systematic search was conducted on Google Play and Apple App Store in June 2023 and December 2024 using predefined oncology- and professional-related keywords, following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. Two independent reviewers (DL and AC) assessed the selected apps using the Mobile App Rating Scale (MARS), which evaluates engagement, functionality, aesthetics, information quality, and subjective quality on a 5-point Likert scale. Discrepancies in ratings were resolved by a third reviewer. Descriptive statistics summarized the app quality and characteristics.

Results: Out of 221 apps initially identified, 20 met the inclusion criteria and were evaluated. Most apps (15/20, 75%) supported both Android and iOS platforms, with 90% (18/20) commercially developed. The mean overall MARS score was 3.51 (SD 0.54), indicating moderate quality but with room for improvement. Only 2 apps, ONCOassist (Portable Medical Technology Ltd.) (mean

XSL•FO
RenderX

4.25, SD 0.26) and Oncology Board Review (mean 4.03, SD 0.39), surpassed the threshold of 4.0, considered good quality. ONCOassist stood out for its comprehensive functionality and high information quality, offering clinical decision support tools such as treatment protocols, prognostic calculators, and toxicity grading aligned with professional oncology practice. Prevention and support apps generally scored lower, particularly in engagement and interactive features. No app achieved a high score across all MARS domains.

Conclusions: The study highlights a fragmented landscape of free mobile apps for cancer care professionals, with predominantly low to moderate quality and limited evidence to support clinical effectiveness. ONCOassist emerges as a promising tool warranting further investigation. This underscores the urgent need for standardized evaluation frameworks, regulatory oversight, and sustainable development strategies to ensure the creation and adoption of reliable, evidence-based digital health tools in oncology.

(JMIR Mhealth Uhealth 2026;14:e71203) doi:10.2196/71203

KEYWORDS

mobile apps; neoplasms; technology assessment; biomedical; telemedicine; quality of health care

Introduction

Cancer is the second leading cause of premature mortality and morbidity in Europe, with 4.7 million new cases and 2.1 million deaths annually in the European Union [1]. The European Cancer Control Plan reflects this political priority, further emphasized by the inclusion of cancer as one of the 5 missions in the Horizon Europe program [2,3].

In recent years, partly accelerated by the COVID-19 pandemic, telehealth apps have grown rapidly, including mobile health (mHealth) apps designed for oncology [4,5]. These apps leverage the convenience of smartphones to support cancer prevention and diagnosis, symptom monitoring, and communication between health care providers and patients [4,6]. Although available to diverse audiences, nearly 45% are designed for health care professionals (HCPs), making them the most prominent target group [7]. Their primary functions are educational, with secondary roles in supporting clinical decision-making and patient care [7]. Mobile apps may be used as stand-alone interventions or in combination with additional support material as part of broader interventions [8].

There is growing evidence that digital health technologies, such as mobile apps, can support self-management, reduce symptom burden, improve cancer care management, and enhance quality of life for patients with cancer [4,9,10]. These apps, used on smartphones or tablets, may facilitate medication adherence, symptom tracking, and disease management [11], as well as promote healthy lifestyles and access to information [6]. Despite this potential, and although HCPs generally express positive attitudes toward their use [12], actual adoption in cancer care remains limited. Evidence from the United States and several European countries indicates a mismatch between the rapid proliferation of cancer-related apps and their real-world use among both professionals and patients [5].

The adoption of mHealth apps by HCPs is hindered by several barriers. First, HCPs face difficulties in identifying appropriate apps for their needs and in evaluating the extent to which scientific evidence supports their use [9]. Moreover, no tools are currently available to help cancer patients or professionals search for high-quality cancer apps [13], making it difficult for both digital laypersons and HCPs to identify reliable options [14]. The quality of existing apps for HCPs also varies and is generally low: systematic searches have found that almost half of the tested apps were rated as deficient or insufficient [15]. Another limitation is that most cancer care apps are not developed with the involvement of health providers, which complicates their implementation [13]. Finally, cost represents an additional barrier, as 77% of health app users rely exclusively on free apps [16,17]. Consequently, the most recent evaluations of health apps have focused only on those that are freely available [18].

Given these challenges, app evaluations can help clarify the landscape. There are several approaches to evaluating apps, including content analyses, usability testing, observational studies, and efficacy trials [19]. Among these, questionnaires and scales are the most common for assessing usability in eHealth apps, although they do not always identify the specific issues that need to be addressed [20]. However, despite the increasing number of cancer-related apps, evidence about the availability and quality of free apps specifically designed for HCPs remains scarce.

To address this gap, this cross-sectional study aims to map the landscape of free mobile apps for cancer prevention, treatment, therapy, or support for HCPs, and assess the quality of the apps identified. This analysis focuses on the characteristics and overall quality of these apps, thereby contributing to the evidence base and informing future development.

Methods

Study Design

This study was designed to evaluate the availability and quality of apps developed for the prevention, treatment, therapy, or support of cancer for HCP during the years 2023 and 2024. The decision to conduct the study over this period was informed by the dynamic nature of mHealth technologies, characterized by rapid advancements and frequent updates.

To achieve this, a systematic search was conducted across major app download platforms, followed by a peer evaluation of each identified app. It is part of the Digital Transition and Digital resilience in Oncology (TRANSiTION) project [21], which aims to develop advanced training to equip both clinical and non-clinical professionals (ie, health managers) with essential digital skills.

XSL•FO
RenderX

Search Strategy

A systematic search of mHealth apps related to cancer prevention, treatment, therapy, and support for HCPs was conducted in Google Play and the App Store. A completed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist is provided in Multimedia Appendix 1. For the purpose of this study, a cancer HCP is defined as any clinical professional (eg, oncologist and radiation therapist) or non-clinical professional (eg, manager and administrative staff) involved in cancer care.

The search strategy was based on methods used in previous studies [22,23]. The search was conducted using a Spain and Cyprus IP addresses. The initial search was performed in June 2023, with a follow-up search in December 2024, to compare the evolution of apps over 1.5 years. The search terms used were "oncology or cancer prevention" or "cancer treatment or cancer therapy" or "cancer support"; "health care professional*" or "doctor*" or "nurse*", "nursing*"; "health manager*". The terms and parameters used were identical to those used in the systematic review conducted by Cloconi et al [24], which aimed to review the digital health technologies used by HCPs and health managers in the academic literature on cancer care.

The inclusion and exclusion criteria used for the selection of apps are listed in Textbox 1.

Textbox 1. Inclusion and exclusion criteria for app selection.

Inclusion criteria

- Apps targeted at health care professionals providing cancer prevention, treatment, therapy, or support.
- Apps available on Google Play and the Apple App Store currently.

Exclusion criteria

- Apps that were not available in English.
- Apps that were not freely available.
- Apps with nonfunctional links.
- Apps not related to oncology.
- Apps addressing the needs of patients, but not of health care professionals.
- Apps with nonfunctional links. Apps not related to oncology.

To be included in the data extraction, apps had to meet all inclusion criteria and none of the exclusion criteria. Google Play apps were searched using a laptop (Windows 11) by systematically entering predefined keywords. The same process was applied for the App Store, through a desktop computer (macOS 13). This approach ensured the selection of apps that are directly accessible and functional for end users.

Data Extraction

Two independent researchers (DL and AC) collected data from descriptions of online apps (including app features and narrative text) available in the App Store and Google Play Store. The variables collected for each app included the following: app name and target group, category, objective, rating (out of 5), number of downloads, development company, cost, creation date, last update, availability, and whether an official source was cited. The technical aspects of the apps were also documented, including features such as the ability to share on social media, the requirement to log in, and the need for web access to function. Additionally, information was collected regarding potential commercial interests, including the presence of advertisements or options for purchasing services. The purposes of the apps were classified using an open-ended question that provided a brief description of the digital tool.

App Quality Evaluation

To assess app quality, the Mobile Apps Rating Scale (MARS) was used [25]. MARS is a tool specifically designed to evaluate the quality and content of mHealth apps. MARS consists of a total of 23 items covering five dimensions: (1) engagement (5 items: fun, interest, individual adaptability, interactivity, and target group), (2) functionality (5 items: performance, usability, navigation, and gestural design), (3) aesthetics (3 items: layout, graphics, and visual appeal), (4) information (7 items: accuracy of app, description, goals, quality of information, quantity of information, quality of visual information, credibility, and evidence base), and (5) subjective quality (4 items: recommendation, times of use, pay-per-use, and overall rating). Items are assessed on a 5-point Likert scale. Items assessing information quality can also be rated as not applicable (eg, in case of missing evidence or missing visual information). Each dimension's score is the average of its items. Therefore, the score for each dimension can range from 1 to 5 (1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent). MARS has excellent psychometric properties [26]. In this study, all the dimensions showed adequate internal consistency (Cronbach α=0.78-0.91).

The app evaluation procedure used was consistent with that of previous studies [27]. Apps were reviewed by 2 trained reviewers (DL and AC), selected randomly from a group of experts engaged in the European project TRANSiTION on e-skills training for cancer HCPs, as mentioned previously. The reviewer team consisted of professionals trained in family medicine, psychology, and nursing, with additional specialization in health technology assessment and digital education. Before conducting the MARS evaluation, reviewers piloted a review of one app and discussed each MARS item to

XSL•FO
RenderX

ensure an appropriate level of interrater consensus ratings in a hybrid session. In case the concordance between the reviewers was not reached (ie, a difference greater than one point in A, B, C, or D MARS dimensions), a third reviewer assessed the app. The lead researcher, an expert in digital health with extensive expertise in biotechnology, oversaw the MARS qualification process.

A descriptive analysis was developed, with continuous variables presented as means. Interrater reliability was examined using the intraclass correlation coefficient (ICC), calculated with a 2-way random-effects model and absolute agreement definition. The data generated were analyzed using IBM SPSS Statistics (version 25.0; IBM Corp).

Ethical Considerations

This study did not involve human participants or the collection of personal data. It was reviewed by the Pontevedra-Vigo-Ourense Research Ethics Committee (ref: 2023/309), which indicated that no formal evaluation was required.

Results

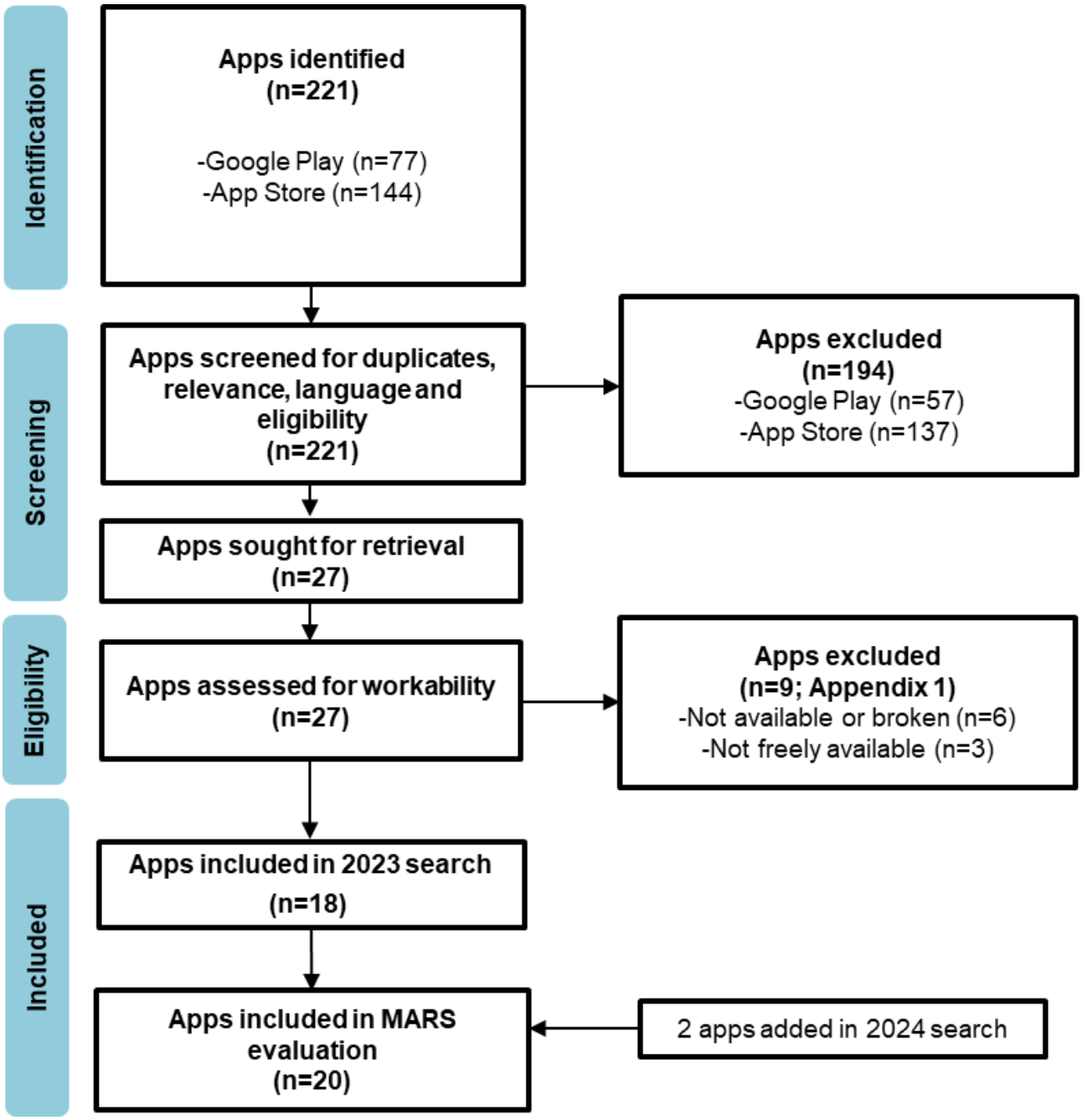
Description

Within the scope of this study, searches were conducted in June 2023 and December 2024. In the initial search, 221 apps were screened, of which 18 met the inclusion criteria. Among these 18 apps, 16 underwent version updates in 2024, with the exceptions being the European Society for Medical Oncology (ESMO) Interactive Guidelines and TNM Cancer Staging Manual.

The subsequent search in December 2024 identified 20 eligible apps, 18 of which had been initially detected in the June 2023 search and remained available. Consequently, this search marked the introduction of 2 new apps that were unavailable in the preceding year (Figure 1). The list of excluded apps can be found in Multimedia Appendix 2.

XSL•FO
RenderX

Figure 1. PRISMA flowchart of the identification, screening, and selection of free oncology mobile applications apps for health care professionals, based on searches conducted in Google Play and Apple App Store in June 2023 and December 2024.



Of the total, 75% (15/20) of the apps are available for Android and iOS. All apps had commercial affiliations, except for 2 (ESMO Interactive Guidelines and "NCCN Guidelines" [National Comprehensive Cancer Network]), which are affiliated with non-governmental organizations. Only 45% (9/20) of the apps were created after the COVID-19 pandemic.

The apps, while related to cancer, exhibited considerable heterogeneity with respect to their focus. Several were focused on teaching users about the treatment of the disease, while others intended to prepare users for an examination. Some cancer guidelines apps were also included. Other apps even extended beyond cancer to address a range of physical or mental illnesses. None of these apps has conclusive evidence of their validity and reliability for their purpose. The final list of analyzed apps is shown in Table 1, where it is specified the app's name, cancer type, platform, and year of creation are specified.

XSL•FO
RenderX

Table 1. Oncology mobile apps for health care professionals were included for evaluation in this study.

Apps	Developer	Focus: cancer type	Platform	Created
All Diseases Treatments	My Apps Studio	Learning disease treatment. Includes most types of cancer	Android	2021
All Diseases Treatments 2023	Soft Solutions	Learning disease treatment. Includes most types of cancer	Android	2022
Cervical Cancer Guide	Apps For Everyone	Cervical Cancer Guide: symptoms, causes, prevention, and treatment	Android	2024
Colon Cancer Guide	Apps For Everyone	Colon Cancer Guide: symptoms, causes, prevention, and treatment	Android	2024
Current Medical Diagnosis and Treatment	Skyscape Medpresso Inc	Diagnosis and treatment information. Includes most types of cancer	Android and iOS	2021
ESMO Interactive Guidelines	European Society for Medical Oncology	Cancer diagnosis and treatment recommendations guidelines. Includes most types of cancer	Android and iOS	2015
Fitzpatrick's Color Atlas	Skyscape Medpresso Inc	Guideline update for dermatological disease. Includes precancerous lesions and cutaneous carcinomas	Android and iOS	2021
Harrison's Manual of Medicine	Skyscape Medpresso Inc	Guideline update in different pathologies. Includes most types of cancer	Android and iOS	2018
Hematology & Oncology Consult	Skyscape Medpresso Inc	Hematology and oncology information. Includes most types of cancer	Android and iOS	2017
NCCN Guidelines	National Comprehensive Cancer Network	Guideline update in oncology. Includes most types of cancer	Android and iOS	2013
ONCOassist	Portable Medical Technology Ltd	Decision support by providing a range of tools (eg, treatment protocols and prognostic scores). Includes most types of cancer	Android and iOS	2015
Oncology Board Review	Higher Learning Technologies Inc	Oncology exam training. Includes most types of cancer	Android and iOS	2017
Oncology Nursing Drug Handbook	Skyscape Medpresso Inc	Cancer Treatment and drug therapy, symptom management cancer. Includes most types of cancer	Android and iOS	2018
Pediatric Disease & Treatment	Patrikat Softech	Information on children's diseases, pediatric treatments, and medical calculators. Includes bone, eye, and hematological cancers	Android	2018
Pharmacology for Nursing 2023	Pass Your Exam	Oncology exam training. Includes most types of cancer	Android and iOS	2022
Radiation Oncology Exam Review	Higher Learning Technologies Inc	Radiation exam training. Includes most types of cancer	Android and iOS	2018
Radiation Oncology Q&A Review	Skyscape Medpresso Inc	Learning radiation oncology. Includes most types of cancer	Android and iOS	2021
TNM Cancer Staging Manual	Skyscape Medpresso Inc	Classification and staging of cancer. Includes most types of cancer	Android and iOS	2021
Washington Manual of Oncology	Skyscape Medpresso Inc	Guideline update in oncology. Includes for most types of cancer	Android and iOS	2017
Williams Manual of Hematology	Skyscape Medpresso Inc	Guideline update in hematology. Includes for most types of cancer	Android and iOS	2018

Interrater reliability indicated a moderate overall level of agreement (ICC=0.66). Reliability varied across MARS dimensions, ranging from excellent for engagement (ICC=0.90) to good for functionality (ICC=0.75), moderate for aesthetics (ICC=0.64) and information (ICC=0.52), and poor for subjective quality (ICC=0.38).

The combined results of the objective assessments (ie, engagement, functionality, aesthetics, and information evaluations) and the subjective assessments are presented in Table 2.

XSL•FO
RenderX

Table 2. Mobile Apps Rating Scale assessment of the included oncology mobile apps: dimension scores and overall quality.

Apps	A. Engagement (SD)	B. Functionality (SD)	C. Aesthetics (SD)	D. Information (SD)	App subjective quality (SD)	App quality, mean (SD) (A+B+C+D)/4
ONCOassist	4.10 (0.42)	4.87 (0.18)	3.66 (0.47)	4.37 (0.05)	3.75 (0.35)	4.25 (0.26)
Oncology Board Review	3.47 (0.31)	4.33 (0.38)	4.11 (0.51)	4.24 (0.41)	3.17 (0.76)	4.03 (0.39)
Hematology & Oncology Consult	3.07 (0.83)	4.50 (0.43)	4.22 (0.69)	4.12 (0.40)	2.92 (0.58)	3.98 (0.58)
Radiation Oncology Exam Review	3.73 (1.10)	4.17 (0.29)	3.55 (1.02)	4.25 (0.69)	3.08 (0.95)	3.92 (0.75)
NCCN Guidelines	3.20 (0)	3.62 (0.18)	3.83 (0.24)	4.75 (0.12)	3.87 (0.53)	3.85 (0.07)
ESMO Interactive Guidelines	3.60 (0.56)	4.62 (0.18)	3 (0.47)	4.10 (0.42)	3.87 (0.18)	3.83 (0.04)
Current Medical Diagnosis and Treatment	3.50 (0.14)	4.12 (0.53)	2.83 (0.71)	4.13 (0.38)	3.12 (0.88)	3.65 (0.37)
TNM Cancer Staging Manual	2.93 (0.12)	4.25 (0.25)	3.22 (1.02)	4.17 (0.60)	3.08 (1.01)	3.64 (0.33)
Pediatric Disease & Treatment	2.60 (0.57)	4.25 (0.35)	4.17 (0.24)	3.39 (0.15)	3.12 (0.88)	3.60 (0.25)
Radiation Oncology Q&A Review	2.93 (0.42)	4.17 (0.29)	3.67 (0.88)	3.59 (0.54)	2.66 (0.52)	3.59 (0.31)
Williams Manual of Hematology	3.20 (0.28)	4.37 (0.53)	3 (0.94)	3.58 (0.59)	2.62 (1.24)	3.53 (0.59)
All Diseases Treatments	2.50 (0.42)	4.62 (0.18)	4 (0.47)	2.95 (0.64)	2.50 (0.35)	3.52 (0.13)
Oncology Nursing Drug Handbook	3.07 (0.31)	3.75 (0.75)	3.78 (0.69)	3.11 (0.63)	2.66 (0.95)	3.43 (0.41)
Pharmacology for Nursing 2023	3.73 (0.76)	3.17 (1.42)	3.22 (1.07)	3.25 (0.66)	3.33 (1.38)	3.34 (0.91)
Washington Manual of Oncology	2.55 (0.66)	3.75 (0.54)	3 (0.82)	4.05 (0.71)	2.50 (1.43)	3.34 (0.59)
All Diseases Treatments 2023	1.90 (0.14)	4.37 (0.18)	3 (0.47)	3.20 (0.28)	2.37 (0.18)	3.12 (0.11)
Harrison's Manual of Medicine	2.60 (0.20)	3.67 (0.38)	3 (1)	3.15 (0.53)	1.83 (0.38)	3.11 (0.27)
Fitzpatrick's Color Atlas	2.60 (0.20)	3.33 (0.88)	2.78 (0.19)	3.12 (0.62)	2.58 (0.38)	2.96 (0.37)
Cervical Cancer Guide	2.10 (0.14)	3.50 (0)	2.67 (0.47)	3.25 (0.83)	2.13 (0.53)	2.87 (0.29)
Colon Cancer Guide	2.40 (0.28)	3.37 (0.18)	2.17 (0.24)	2.92 (0.35)	2.25 (0.71)	2.71 (0.17)
Overall mean	2.99 (0.68)	4.04 (0.65)	3.34 (0.80)	3.71 (0.69)	2.87 (0.86)	3.51 (0.54)

Notably, the scores were generally low, with an average of 3.51. Only 2 apps achieved an objective score higher than 4 ("good"), while in the subjective evaluations, none reached this threshold. The highest-rated app was ONCOassist (mean 4.25, SD 0.26). Oncology Board Review ranked second (mean 4.03, SD 0.39), followed by Hematology & Oncology Consult in third place (mean 3.98, SD 0.58). Reviewers noted that areas for potential improvement include aesthetics for ONCOassist and engagement for both Oncology Board Review and Hematology & Oncology Consult. The lowest-scoring apps were Colon Cancer Guide (2.71, SD 0.17), Cervical Cancer Guide (2.87, SD 0.29), and Fitzpatrick's Color Atlas (mean 2.96, SD 0.37). The domain with the lowest scores for these 3 apps was engagement and subjective quality evaluation. The results are discussed below according to the different dimensions of the MARS. It is noteworthy that none of the evaluated apps attained a score of 4 in all dimensions.

Apps' Engagement

The apps with the highest scores in engagement were ONCOassist (mean 4.10, SD 0.42), Radiation Oncology Exam Review (mean 3.73, SD 1.10), and Pharmacology for Nursing 2023 (mean 3.73, SD 0.76), while the lowest scoring apps were All Diseases Treatments 2023 (mean 1.90, SD 0.14), Cervical Cancer Guide (mean 2.10, SD 0.14), and Colon Cancer Guide (mean 2.40, SD 0.28).

XSL•FO
RenderX

Apps' Functionality

The apps with the highest scores for functionality were ONCOassist (mean 4.87, SD 0.18), ESMO Interactive Guidelines (mean 4.62, SD 0.17), and All Diseases Treatments (mean 4.62, SD 0.17). Conversely, the apps with the lowest scores were Pharmacology for Nursing 2023 (mean 3.17, SD 1.42), Colon Cancer Guide (mean 3.37, SD 0.17), and Cervical Cancer Guide (mean 3.50, SD 0).

Apps' Aesthetics

The apps with the highest aesthetics scores were Hematology & Oncology Consult (mean 4.22, SD 0.69), Pediatric Disease & Treatment (mean 4.17, SD 0.23), and Oncology Board Review (mean 4.11, SD 0.51). The apps with the lowest scores were Colon Cancer Guide (mean 2.17, SD 0.23) and Cervical Cancer Guide (mean 2.67, SD 0.47).

Apps' Information

The apps with the highest scores in information were NCCN Guidelines (mean 4.75, SD 0.12), ONCOassist (mean 4.37, SD 0.05), and Radiation Oncology Exam Review (mean 4.25, SD 0.69). Conversely, the apps with the lowest scores were Colon Cancer Guide (mean 2.92, SD 0.35), All Diseases Treatments (mean 2.95, SD 0.64), and Oncology Nursing Drug Handbook (mean 3.11, SD 0.63).

Apps Subjective Quality

The apps with the highest scores in this section are NCCN Guidelines and ESMO Interactive Guidelines (mean 3.87, SD 0.53, and SD 0.18). Next is ONCOassist (mean 3.75, SD 0.35), followed by Pharmacology for Nursing 2023 (mean 3.33, SD 1.38). In contrast, the app with the lowest subjective rating is Harrison's Manual of Medicine (mean 1.83, SD 0.38).

Discussion

Principal Findings

In this study, the app review identified 221 apps between June 2023 and December 2024 following the search strategy, keywords, and parameters outlined by Cloconi et al [24]. After the screening phase, 20 apps for cancer prevention, treatment, therapy, or support for HCPs were selected in the systematic search. To explore them further, this study evaluated these apps with the MARS, a scientifically validated instrument that assesses 5 key dimensions: engagement, functionality, aesthetics, information, and subjective quality.

As a result, the ONCOassist app achieved the highest score among all the apps. Endorsed by the European Scientific Organisations such as the ESMO [28] and the European Oncology Nursing Society [29], it obtained particularly strong ratings in functionality and information. Its feature set is closely aligned with daily oncology practice, offering toxicity criteria, staging tools, prognostic calculators, and adjunctive functions that streamline clinical workflows and reduce time spent switching between resources [30]. These tools are clinically relevant, for example, in grading adverse events during treatment, estimating prognosis through staging systems, or calculating chemotherapy dosing at the point of care, which may explain their advantage over guideline-oriented, exam-oriented, or single-purpose apps. ONCOassist has also been recognized as a best practice model for incorporating user feedback in its development [30] and is valued for enhancing performance by providing guideline-based recommendations that act as a safety net for clinical decision-making [31], as well as showing potential utility in frailty assessment [32]. Its integration into oncology practice is already being discussed within professional networks [33], and its uptake is evident in some regions. As an instance, Devnani et al [34] reported that nearly 60% of a sample of 155 Indian oncology professionals had used it for over a year. Nonetheless, the available evidence should be interpreted with caution due to small samples, non-probabilistic recruitment, and potential conflicts of interest, highlighting the need for future research with more robust designs to confirm its clinical effectiveness and real-world impact.

Despite the increasing relevance of mobile apps for cancer care and management, the number of apps identified in this study is lower than that reported in previous research. Comparatively, Lima et al [4] identified 34 mobile apps published between 2015 and 2022 aimed at patients with any type of cancer for diagnosis, prognosis, monitoring, and treatment. In any case, this study focused exclusively on free apps designed for use by health professionals. In this regard, although the study by Charbonneau et al [13] identified 123 apps for cancer intended for the general public, only 3 of them are aimed at professionals. Consequently, it is to be expected that the number of apps identified would be relatively limited. Importantly, this limited number is not only consistent with prior studies but also reflects a broader structural scarcity of oncology-specific mobile apps for professionals. Much of digital oncology innovation currently takes place through web-based platforms or tools integrated into electronic health records rather than stand-alone mobile apps, which narrows the availability of dedicated mobile solutions in app stores.

The assessment of apps indicates that the quality of the identified apps is generally poor. The mean score was 3.51 (SD 0.54), with only 2 apps (the aforementioned ONCOassist app and Oncology Board Review) exceeding the 4-point threshold, which is considered to indicate a "good" level of quality. These findings align with the existing literature. Amor-García et al [35] used the same tool to evaluate 46 apps designed for patients with genitourinary tumors, obtaining a mean score of 2.98 (SD 0.77). In their study, only 13% (6 out of 46) of the apps achieved a score exceeding 4 on the MARS scale [36]. In a similar vein, Böhme et al [15] used a tool based on the MARS scale, which indicated that 60% of the apps for cancer were rated as insufficient or deficient. Similarly, the mean MARS score for the overall quality of the 88 apps addressing hematological conditions (including hematological cancers among other pathologies) in the review by Narrilos et al [37] was 3.03 (SD 1.14). They concluded that more than half of the apps do not meet acceptable criteria for quality and content, with most providing only information about the conditions and lacking interactivity and personalization features. Scholz and Teetz [14], who analyzed the quality of 33 identified mHealth apps focusing on breast cancer, agreed with the findings of previous studies

XSL•FO
RenderX

as they also found problems with app quality. Additionally, results of studies that focused on the evaluation of mobile apps intended for cancer patients are similar to the results of studies that evaluated the quality of apps intended for HCPs [13]. This consistency in lack of quality in apps was also observed in this study, as none of the evaluated apps achieved a score of at least 4 across all dimensions of the MARS scale. The assurance of high-quality apps (eg, through the provision of user-friendly interfaces, the presentation of evidence to support them, or the involvement of users in their creation) would serve as a facilitator for the implementation of mHealth [38].

Beyond individual app performance, our findings also reveal differences across categories. Apps classified under prevention tended to focus on awareness and lifestyle recommendations, but they generally scored lower on functionality, reflecting a lack of interactive or decision-support features. In contrast, treatment-related apps offered more sophisticated functionalities such as toxicity grading, staging tools, or guideline-based recommendations, which explains their higher ratings in functionality and information. Finally, support apps showed strengths in accessibility and patient-oriented resources but often lacked the depth of evidence-based content required for professional use. Taken together, these patterns suggest that while treatment-oriented apps currently provide the most comprehensive toolsets for oncology professionals, prevention and support apps remain underdeveloped for clinical practice and represent important areas for future improvement. However, these comparative insights should be interpreted with caution, since the small number of apps included (n=20) restricts both the representativeness and the statistical power of the findings, as the limited sample constrains subgroup comparisons and increases the influence of outliers, making the results more descriptive than inferential. Taken together, these category patterns highlight the inherently heterogeneous nature of the digital oncology landscape. This heterogeneity can be understood as a structural feature of current app ecosystems, where tools differ sharply in purpose, evidence base, and alignment with clinical workflows.

This study is not without limitations. First, there may be additional apps not captured in this review. The keywords used for the review might not identify other relevant apps for cancer HCPs. Furthermore, while Google Play and the Apple Store account for approximately two-thirds of the mobile apps downloaded annually [39], there are other, less commonly used download platforms that were not included in this review. In addition, the review focused exclusively on free apps. While this approach reflects real-world usage patterns and aligns with the scope of the TRANSiTION project [16,17], it may also reduce the representativeness of the findings, since paid apps, which could differ in quality or functionality, were excluded. Limiting the search to English-language apps also excludes numerous potentially relevant apps in other languages, which may affect the generalizability of our findings. In particular, this restriction could underrepresent apps developed for and widely used within non–English-speaking European health care systems. However, this choice was consistent with the primary aim of the study, which is embedded in the TRANSiTION project, where the apps may be used as common resources for participants from 14 European countries. English was selected as the working language of the project and provided a common ground to ensure accessibility, comparability, and feasibility of evaluation across countries. Moreover, the number of apps evaluated after the screening process (n=20) is relatively small. This limits the representativeness of the sample and reduces the statistical power of the findings, as subgroup analyses could not be conducted, and outliers may disproportionately affect the results. Therefore, the study should be viewed primarily as descriptive. Future research with larger and more diverse samples, ideally through systematic surveillance and international collaborations, is needed to provide a more comprehensive and robust overview of oncology apps. Second, although the evaluation of the apps was conducted using a standardized procedure and a validated tool, the evaluators may be susceptible to response biases, as the assessment relies on subjective ratings. In addition, the MARS assesses the app's evidence base with a single item (item 19), which appears insufficient for determining whether an app should be included in an HCP's resource portfolio. These limitations imply that the validity and reliability of the quality scores may be constrained, since rater bias and the restricted appraisal of evidence can influence results. Consequently, MARS ratings should be understood as standardized indicators of app quality rather than definitive evidence of clinical effectiveness or robustness. Nevertheless, MARS is the most widely used among all app evaluation scales and is suitable for general public use [40]. Future research should therefore combine standardized quality assessments with empirical studies to determine the actual impact of oncology apps on professional performance, efficiency, and well-being.

Health apps are emerging technologies that handle sensitive personal health data, which can significantly impact health outcomes and are susceptible to cybersecurity threats [41]. This study highlights that while some apps have limited supporting evidence from previous studies, none have provided evidence of their validity and reliability for their intended purpose. In the study of Berger-Groch et al [42] that focused on identifying mobile apps that can be used in the treatment of patients with musculoskeletal tumors, only 3 scientific articles on mobile apps in orthopedic oncology were present, yet several more apps were available without scientific medical evaluation. Some of the justifications provided in the literature are that apps tend to become unavailable in app stores once the studies have concluded, or that a significant proportion of the papers do not reveal the specific apps that were used [11]. The lack of regulatory standards or frameworks for assessing the quality of these apps is, therefore, of serious concern [43] and has been described as the "Wild West of healthcare" [44]. There are some initiatives to regulate the quality of apps, such as those in the National Health Service [45], the Spanish Ministry of Health and Agència de Qualitat i Avaluació Sanitàries de Catalunya [46], or the ongoing European project ASSESS DHT [47], which focuses precisely on filling this gap.

In addition to these regulatory challenges, the predominance of commercially developed apps in our sample (18 out of 20) raises concerns regarding both content quality and sustainability. On the one hand, commercial investment may provide resources

XSL•FO
RenderX

for more advanced design and maintenance. On the other hand, it can introduce potential conflicts of interest, with content that may prioritize promotional goals over evidence-based practice. Sustainability is also uncertain, as the long-term availability of commercially driven apps depends on market viability, a factor that may explain why many oncology apps become unavailable after only a few years. Similarly, apps developed within the framework of publicly funded research projects may encounter comparable sustainability challenges. Although grounded in scientific evidence, their maintenance and updates are often limited to the project's duration, after which support may be discontinued, restricting their long-term utility. Prior research has shown that such discontinuity contributes to poor adoption rates and a lack of long-term sustainability, highlighting the need for evidence-informed methodological approaches that extend beyond funding cycles to ensure continuity and adoption in real-world practice [48]. These findings underscore the importance of regulatory oversight and sustainable development models to guarantee that oncology professionals can rely on digital tools with transparent, evidence-based, and durable content. Overall, the scarcity and heterogeneity identified in this review underscore the fragmented nature of the current digital oncology ecosystem and highlight the need for coordinated development and robust evaluation frameworks.

From a clinical perspective, our findings highlight the need for oncology professionals to exercise caution when selecting mobile apps, giving priority to those that demonstrate transparent evidence and functionalities that support clinical workflows. In this sense, the implementation of regulatory frameworks and validated quality standards will be crucial for transforming our results into practical guidance, ensuring that HCPs can safely rely on digital tools that enhance efficiency and patient care. In this context, this study, as part of the TRANSiTION project, which incorporates multiprofessional and patient perspectives from several European Union Member States into its design and implementation, as well as in the dissemination and communication of its outcomes, will contribute to initiating new research lines and developing apps with enhanced methodologies [49].

Conclusions

This study identified mobile apps available to cancer care HCPs and highlighted their limitations. Currently, there is insufficient evidence in the literature to support the effectiveness of any specific downloadable app. However, this review offers a structured evaluation of app quality and a comparative overview of each app's strengths and limitations, providing a baseline for future research and development in this field. While overall app quality was poor, the findings point to the potential usefulness of the ONCOassist app, which emerged as the highest-rated app in our assessment and merits further investigation through robust empirical studies. Continued monitoring, evidence-informed development, and the implementation of quality standards will be essential to ensure that mobile apps can reliably support oncology practice in the future.

Acknowledgments

The authors used DeepL Write to improve grammar and language clarity. The authors reviewed, edited, and took full responsibility for the final content of the manuscript.

Data Availability

Data supporting the results reported in the article are available on request from the authors.

Funding

This study has been elaborated in the framework of the European project TRANSiTION (GA 101101261). Co-funded by the European Union (EU). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the EU or European Health and Digital Executive Agency. Neither the EU nor the granting authority can be held responsible for them.

Authors' Contributions

Conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, writing - original draft, and writing - review and editing: DL
Conceptualization, data curation, formal analysis, investigation, methodology, supervision, validation, and writing - original draft: IN
Funding acquisition, investigation, project administration, resources, supervision, and validation: AC
Investigation, methodology, and project administration: DC
Funding acquisition, methodology, resources, supervision, and validation: DMA
Investigation, methodology, and project administration: CMA
Funding acquisition, resources, supervision, and validation: NC
Investigation, methodology, resources, validation, visualization, and writing - review and editing: NLR
Investigation, supervision, and validation: MJFD
Conceptualization, funding acquisition, methodology, project administration, supervision, validation, writing - original draft, and writing - review and editing: CC

XSL•FO
RenderX

Conflicts of Interest

None declared

Multimedia Appendix 1
PRISMA checklist.
[DOCX File , 270 KB - mhealth_v14i1e71203_app1.docx]

Multimedia Appendix 2
Excluded apps.
[DOCX File , 17 KB - mhealth_v14i1e71203_app2.docx]

References

1. Dyba DT, Randi G, Bray F, Martos C, Giusti F, Nicholson N, et al. The European cancer burden in 2020: Incidence and mortality estimates for 40 countries and 25 major cancers. Eur J Cancer 2021;157:308-347 [FREE Full text] [doi: 10.1016/j.ejca.2021.07.039] [Medline: 34560371]
2. Communication from the commission to the European parliament and the council: Europe's beating cancer plan. European Commission. 2021. URL: https://eur-lex.europa.eu/resource.html?uri=cellar:8dec84ce-66df-11eb-aeb5-01aa75ed71a1.0002.02/DOC_1&format=PDF [accessed 2024-12-26]
3. EU missions in horizon Europe. European Commission. 2025. URL: https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe_en [accessed 2025-09-03]
4. Lima LA, Danilo MDM, Vaz CB, Pereira AI. Using m-health apps in oncology : a review from 2015 to 2022. : IEEE; 2022 Presented at: 18th Iberian Conference on Information Systems and Technologies (CISTI); 2023 June 20-23; Aveiro, Portugal URL: https://doi.org/10.23919/CISTI58278.2023.10211467 [doi: 10.23919/cisti58278.2023.10211467]
5. Tarricone TR, Cucciniello M, Armeni P, Petracca F, Desouza KC, Hall LK, et al. Mobile health divide between clinicians and patients in cancer care: results from a cross-sectional international survey. JMIR Mhealth Uhealth 2019;7(9):e13584 [FREE Full text] [doi: 10.2196/13584] [Medline: 31493318]
6. Saevarsdottir SR, Gudmundsdottir SL. Mobile apps and quality of life in patients with breast cancer and survivors: systematic literature review. J Med Internet Res 2023;25:e42852 [FREE Full text] [doi: 10.2196/42852] [Medline: 37494111]
7. Upadhyay VA, Landman AB, Hassett MJ. Landscape analysis of oncology mobile health applications. JCO Clin Cancer Inform 2021;5:579-587. [doi: 10.1200/CCI.20.00156] [Medline: 34033510]
8. Iribarren SJ, Akande TO, Kamp KJ, Barry D, Kader YG, Suelzer E. Effectiveness of mobile apps to promote health and manage disease: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2021;9(1):e21563 [FREE Full text] [doi: 10.2196/21563] [Medline: 33427672]
9. Dhar E, Bah AN, Chicchi Giglioli IA, Quer S, Fernandez-Luque L, Núñez-Benjumea FJ, et al. A scoping review and a taxonomy to assess the impact of mobile apps on cancer care management. Cancers (Basel) 2023;15(6):1775 [FREE Full text] [doi: 10.3390/cancers15061775] [Medline: 36980661]
10. Wanchai A, Anderson EA, Armer JM. A systematic review of m-health apps on managing side effects of breast cancer treatment. Support Care Cancer 2022;31(1):86. [doi: 10.1007/s00520-022-07464-x] [Medline: 36574048]
11. Ana FA, Loreto M, José LMM, Pablo S, María Pilar MJ, Myriam S. Mobile applications in oncology: a systematic review of health science databases. Int J Med Inform 2020;133:104001. [doi: 10.1016/j.ijmedinf.2019.104001] [Medline: 31706229]
12. Kessel KA, Vogel MME, Schmidt-Graf FS, Combs SE. Mobile apps in oncology: a survey on health care professionals' attitude toward telemedicine, mhealth, and oncological apps. J Med Internet Res 2016;18(11):e312 [FREE Full text] [doi: 10.2196/jmir.6399] [Medline: 27884810]
13. Charbonneau DH, Hightower S, Katz A, Zhang K, Abrams J, Senft N, et al. Smartphone apps for cancer: a content analysis of the digital health marketplace. Digit Health 2020;6:2055207620905413 [FREE Full text] [doi: 10.1177/2055207620905413] [Medline: 32110428]
14. Scholz S, Teetz L. Smart health via mHealth? Potentials of mobile health apps for improving prevention and adherence of breast cancer patients. Digit Health 2022;8:20552076221074127 [FREE Full text] [doi: 10.1177/20552076221074127] [Medline: 35096411]
15. Böhme C, von Osthoff MB, Frey K, Hübner J. Qualitative evaluation of mobile cancer apps with particular attention to the target group, content, and advertising. J Cancer Res Clin Oncol 2018;144(1):173-181. [doi: 10.1007/s00432-017-2533-0] [Medline: 29052039]
16. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. JMIR Mhealth Uhealth 2019;7(4):e11223 [FREE Full text] [doi: 10.2196/11223] [Medline: 30990458]
17. Peng W, Kanthawala S, Yuan S, Hussain S. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]

XSL•FO
RenderX

18. Báez Gutiérrez N, Rodríguez Ramallo H, Mendoza-Zambrano E, Brown Arreola B, Santos Ramos B, Abdel-Kader Martín L, et al. Smartphone apps for pulmonary hypertension: systematic search and content evaluation. JMIR Mhealth Uhealth 2024;12:e57289 [FREE Full text] [doi: 10.2196/57289] [Medline: 39476376]
19. Jake-Schoffman DE, Silfee V, Waring M, Boudreaux E, Sadasivam R, Mullen S, et al. Methods for evaluating the content, usability, and efficacy of commercial mobile health apps. JMIR Mhealth Uhealth 2017;5(12):e190 [FREE Full text] [doi: 10.2196/mhealth.8758] [Medline: 29254914]
20. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
21. TRANSiTION project. European Cancer Organisation. 2024. URL: https://www.europeancancer.org/eu-projects/impact/resource/transition [accessed 2024-12-20]
22. Carrion C, Robles N, Sola-Morales O, Aymerich M, Ruiz Postigo JA. Mobile health strategies to tackle skin neglected tropical diseases with recommendations from innovative experiences: systematic review. JMIR Mhealth Uhealth 2020;8(12):e22478 [FREE Full text] [doi: 10.2196/22478] [Medline: 33382382]
23. Eis S, Solà-Morales O, Duarte-Díaz A, Vidal-Alaball J, Perestelo-Pérez L, Robles N, et al. Mobile applications in mood disorders and mental health: systematic search in apple app store and google play store and review of the literature. Int J Environ Res Public Health 2022;19(4) [FREE Full text] [doi: 10.3390/ijerph19042186] [Medline: 35206373]
24. Cloconi C, Protogiros D, Tsitsi T, Christodoulou M, Christofi M, Papoui E. A systematic review of digital health technologies utilized by hcp and health managers in the context of cancer care. 2025. URL: https://www.crd.york.ac.uk/PROSPERO/view/CRD42023425898 [accessed 2025-09-03]
25. Stoyanov SR, Hides L, Kavanagh D, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
26. Terhorst Y, Philippi P, Sander L, Schultchen D, Paganini S, Bardus M, et al. Validation of the mobile application rating scale (MARS). PLoS One 2020;15(11):e0241480 [FREE Full text] [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
27. Lu DJ, Girgis M, David J, Chung E, Atkins K, Kamrava M. Evaluation of mobile health applications to track patient-reported outcomes for oncology patients: a systematic review. Adv Radiat Oncol 2021;6(1):100576 [FREE Full text] [doi: 10.1016/j.adro.2020.09.016] [Medline: 33073061]
28. ESMO community oncologists task force: major achievements. ESMO. URL: https://www.esmo.org/about-esmo/organisational-structure/educational-committee/practising-oncologists-working-group/major-achievements [accessed 2025-09-03]
29. European oncology nursing society (EONS). Other Educational Resources. Brussels: EONS URL: https://cancernurse.eu/education/other-educational-resources/ [accessed 2024-11-07]
30. Jacob C, Sanchez-Vazquez A, Ivory C. Clinicians' role in the adoption of an oncology decision support app in Europe and its implications for organizational practices: Qualitative case study. JMIR Mhealth Uhealth 2019;7 [FREE Full text] [doi: 10.2196/preprints.13555]
31. Jacobson A, Macfarlane C, Pozniak E. Feasibility of integrating a mobile decision-support app into a multicomponent CME initiative: developing clinician competence at the point of care. J Eur CME 2020;9(1):1834762 [FREE Full text] [doi: 10.1080/21614083.2020.1834762] [Medline: 33178490]
32. Banna GL, Cantale O, Haydock M, Battisti N, Bambury K, Musolino N, et al. International survey on frailty assessment in patients with cancer. Oncologist 2022;27(10):e796-e803 [FREE Full text] [doi: 10.1093/oncolo/oyac133] [Medline: 35905085]
33. Gandur-Quiroga MN, Krasnapolski MA, Yekeduz E, Izzo M, Bambury KD, Morgan G. Impact of digital tools in oncology: a global analysis of the adoption of ONCOassist and NCCN chemotherapy order templates (NCCN Templates) in collaboration with OncoAlert. J Clin Oncol 2024;42(16_suppl):e13662. [doi: 10.1200/jco.2024.42.16_suppl.e13662]
34. Devnani B, Noronha V, Sirohi B, Bambury K, George R, Pareek V, et al. 429P cancer and COVID-19 in India: assessing the impact in a nationwide survey. Annals of Oncology 2022;33:S1608 [FREE Full text] [doi: 10.1016/j.annonc.2022.10.460]
35. Böhme C, von Osthoff MB, Frey K, Hübner J. Development of a Rating Tool for Mobile Cancer Apps: Information Analysis and Formal and Content-Related Evaluation of Selected Cancer Apps. J Cancer Educ 2019 Feb;34(1):105-110. [doi: 10.1007/s13187-017-1273-9] [Medline: 28819928]
36. Amor-García M, Collado-Borrell R, Escudero-Vilaplana V, Melgarejo-Ortuño A, Herranz-Alonso A, Arranz Arija J, et al. Assessing apps for patients with genitourinary tumors using the mobile application rating scale (MARS): systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2020;8(7):e17609 [FREE Full text] [doi: 10.2196/17609] [Medline: 32706737]
37. Narrillos-Moraza Á, Gómez-Martínez-Sagrera P, Amor-García M, Escudero-Vilaplana V, Collado-Borrell R, Villanueva-Bueno C, et al. Mobile apps for hematological conditions: review and content analysis using the mobile app rating scale. JMIR Mhealth Uhealth 2022;10(2):e32826 [FREE Full text] [doi: 10.2196/32826] [Medline: 35171109]
38. Ardito V, Golubev G, Ciani O, Tarricone R. Evaluating barriers and facilitators to the uptake of mhealth apps in cancer care using the consolidated framework for implementation research: scoping literature review. JMIR Cancer 2023;9:e42092 [FREE Full text] [doi: 10.2196/42092] [Medline: 36995750]

XSL•FO
RenderX

39. Grundy QH, Wang Z, Bero L. Challenges in assessing mobile health app quality: a systematic review of prevalent and innovative methods. Am J Prev Med 2016;51(6):1051-1059. [doi: 10.1016/j.amepre.2016.07.009] [Medline: 27659122]
40. Azad-Khaneghah P, Neubauer N, Miguel Cruz A, Liu L. Mobile health app usability and quality rating scales: a systematic review. Disabil Rehabil Assist Technol 2021;16(7):712-721. [doi: 10.1080/17483107.2019.1701103] [Medline: 31910687]
41. Shuren J, Patel B, Gottlieb S. FDA regulation of mobile medical apps. JAMA 2018;320(4):337-338. [doi: 10.1001/jama.2018.8832] [Medline: 29971339]
42. Berger-Groch J, Keitsch M, Reiter A, Weiss S, Frosch K, Priemel M. The use of mobile applications for the diagnosis and treatment of tumors in orthopaedic oncology - a systematic review. J Med Syst 2021;45(11):99 [FREE Full text] [doi: 10.1007/s10916-021-01774-z] [Medline: 34628540]
43. Grundy Q. A review of the quality and impact of mobile health apps. Annu Rev Public Health 2022;43:117-134 [FREE Full text] [doi: 10.1146/annurev-publhealth-052020-103738] [Medline: 34910582]
44. Anthes E. Mental health: there's an app for that. Nature 2016;532(7597):20-23. [doi: 10.1038/532020a] [Medline: 27078548]
45. Standards for NHS app integration. NHS England Digital n.d. URL: https://digital.nhs.uk/services/nhs-app/how-to-integrate-with-the-nhs-app/standards-for-nhs-app-integration [accessed 2025-09-03]
46. Segur-Ferrer J, Moltó-Puigmartí C, Pastells-Peiró R, Vivanco-Hidalgo R. Health Technology Assessment Framework: adaptation for digital health technology assessment: user guide. Madrid: ministry of health. Barcelona: Agency for Health Quality and Assessment of Catalonia; Dec 14, 2023:A.
47. NA. ASSESS DHT. 2025. URL: https://assess-dht.eu/ [accessed 2025-09-03]
48. An Q, Kelley M, Hanners A, Yen PY. Sustainable development for mobile health apps using the human-centered design process. JMIR Form Res 2023;7:e45694 [FREE Full text] [doi: 10.2196/45694] [Medline: 37624639]
49. Linares D, Charalambous A. Addressing digital skills gaps in cancer care: a call for targeted training programs. Eur J Oncol Nurs 2025;75:102842. [doi: 10.1016/j.ejon.2025.102842] [Medline: 40010018]

Abbreviations

ESMO: European Society for Medical Oncology
HCP: health care professional
ICC: intraclass correlation coefficient
MARS: Mobile Apps Rating Scale
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
TRANSiTION: Digital Transition and Digital resilience in Oncology

Edited by A Stone; submitted 12.Jan.2025; peer-reviewed by F Al Dhabbari, A Cabanes, Y Chu, J Chow; comments to author 15.Aug.2025; revised version received 30.Nov.2025; accepted 01.Dec.2025; published 11.Feb.2026.

Please cite as:
Liñares D, Nicolaidou I, Charalambous A, Cabutto D, Moreno-Alonso D, Madrid Alejos C, Couespel N, López-Rey N, Fernández-Domínguez MJ, Carrion C, Clavería A
Mobile Apps for Oncology Health Care Professionals: Mapping and Assessment Study
JMIR Mhealth Uhealth 2026;14:e71203
URL: https://mhealth.jmir.org/2026/1/e71203
doi:10.2196/71203
PMID:

©David Liñares, Iolie Nicolaidou, Andreas Charalambous, Daniela Cabutto, Deborah Moreno-Alonso, Clara Madrid Alejos, Norbert Couespel, Noemí López-Rey, María José Fernández-Domínguez, Carme Carrion, Ana Clavería. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Safer Sleep Guidance in Standalone UK Smartphone Apps for Parents of Newborns and Infants: Systematic mHealth App Review

Senna Yousef[1,2]; Niamh Meaney[1]; Dawn Dowding[1], PhD; Norina Gasteiger[1], PhD

[1]Division of Nursing, Midwifery and Social Work, University of Manchester, Manchester, England, United Kingdom

[2]The School of Health and Society, University of Salford, Salford, England, United Kingdom

Corresponding Author:
Norina Gasteiger, PhD
Division of Nursing, Midwifery and Social Work
University of Manchester
Jean McFarlane Building, 176 Oxford Rd
Manchester, England, M13 9PY
United Kingdom
Phone: 44 0161 306 7798
Email: norina.gasteiger@manchester.ac.uk

Abstract

Background: Parents are increasingly using smartphone apps to help guide decision-making around aspects of caregiving, including safer sleep. However, it remains unclear whether the guidance provided aligns with best practice guidelines.

Objective: This review aimed to identify smartphone apps that provide safer sleep guidance and to assess their content, technical features, functionality, and quality. A secondary aim was to make recommendations on the improvements that existing and future apps can make to provide safer sleep guidance.

Methods: In January 2026, we searched the UK Google Play Store and Apple App Store for safer sleep guidance apps for parents and caregivers of newborns and infants (babies aged ≤12 months). Information regarding content, technical features, and functional features was extracted and summarized descriptively. Each app was evaluated by 2 raters using the Mobile App Rating Scale (MARS) and the IMS Institute for Healthcare Informatics functionality score. Adherence to best practice sleep guidance was assessed using bespoke criteria. Interrater reliability was calculated.

Results: A total of 1345 apps were identified, with 12 meeting the inclusion criteria. All were free to download, with 42% (5/12) containing in-app purchases. Only 17% (2/12) were affiliated with a professional health, medical, clinical, or government body. All 12 apps contained some information regarding safer sleep practices, although this was addressed and named differently (eg, calling it cot death, sudden infant death syndrome, or safe sleep). Of the 9 criteria relating to safer sleep best practice guidance, the apps mentioned an average of 6.75 (SD 1.71; range 3-9) criteria. Some promoted unsafe sleep practices (ie, around swaddling and cosleeping). The apps included an average of 5.4 of the 11 IMS functionality criteria (SD 2.5; range 2-9). The mean MARS score was 3.7 out of 5 (SD 0.43; range 2.92-4.13), and most (11/12, 92%) had a minimum acceptability score of 3 or above. Two had been formally trialed or tested. Although the language used was considered accessible (easy to understand), none of the apps allowed the language to be changed (beyond English), and very few (3/12, 25%) enabled users to change to night/dark mode, which would be considered essential when using these apps at night.

Conclusions: Several apps about safer baby sleep exist. Few discuss safer sleep guidance comprehensively or raise awareness about sudden infant death syndrome and its links with unsafe sleep. All the reviewed apps met the minimum acceptability criteria for quality; however, not many were formally trialed or tested. More long-term research should be conducted on the effectiveness of safer sleep apps on adherence to safer sleep guidance. We recommend that apps relating to infant and baby safety in general include information on realistic safer sleep practices that are informed by best practice guidelines and take a risk minimization approach.

(JMIR Mhealth Uhealth 2026;14:e95642) doi:10.2196/95642

KEYWORDS

babies; digital health; guidance; infants; mHealth; mobile health; newborn; review; safer sleep

XSL•FO
RenderX

Introduction

Safer sleep refers to a range of evidence-based practices that reduce the risk of sudden infant death syndrome (SIDS), an ongoing and significant public health concern. SIDS is defined as the sudden, unexpected, and unexplained death of an otherwise healthy infant occurring in babies aged 12 months or younger and most commonly within the first 6 months [1]. SIDS is listed in the *ICD-11* (*International Classification of Diseases, 11th Revision*) under the code MH11: sudden infant death syndrome and includes terms such as "cot death," "infantile sudden infant death syndrome," and related variants [2]. As the exact etiology of SIDS remains unknown and is believed to result from a combination of factors, current guidance focuses on implementing measures, such as safer sleep practices, that mitigate and reduce associated risks [1]. In the United Kingdom, safer sleep recommendations are primarily informed by guidance from The Lullaby Trust, the National Health Service (NHS), and the National Institute for Health and Care Excellence (NICE), all of which aim to reduce the incidence of SIDS.

The safest sleep environment for an infant is a separate sleep space, located within the same room as their carer [3]. Infants should be placed on their backs on a firm, flat, waterproof mattress, with their feet positioned closest to the foot of the cot or Moses basket. The cot or Moses basket should be kept free of objects and contain only the mattress and lightweight bedding [3-5]. They should be kept smoke-free, and the room temperature should be maintained between 16 °C and 20 °C, as overheating is a contributing factor to SIDS [3,6]. Current guidance recommends room sharing with their carer for at least the first 6 months [4]. This is central advice provided by professionals, and all parents should be given risk-reduction information on bed-sharing to support informed decision-making [6].

Parents should be advised not to bed-share if their baby has a low birth weight, weighs less than 2.5 kg, was born prematurely, or if either parent has consumed 2 or more units of alcohol, smokes, is taking medication that causes drowsiness, or has used recreational drugs [3,6]. Cosleeping should never occur on a sofa or armchair, as this significantly increases the risk of fatal sleep accidents and SIDS [7]. Parents should also assess the sleep environment for potential entrapment risks, including any gaps between the mattress and the bedframe or adjacent walls [8].

According to provisional data from the Office for National Statistics [9], a total of 164 unexplained infant deaths occurred in England and Wales in 2023, representing 7.1% of all infant mortality. Of these, 52% were classified as SIDS, equating to 0.14 sudden infant deaths per 1000 live births in 2023. Evidence indicates that hazardous sleep environments are present in most unexplained infant deaths in the United Kingdom, which is a key focus of national guidance [10].

While safer sleep campaigns, notably the "Back to Sleep" initiative launched in the 1990s, contributed to a 90% reduction in SIDS in England, this decline has plateaued in recent years [10,11]. Although the campaign was successful in targeting position-related SIDS, it has had less impact on deaths associated with cosleeping, which now accounts for approximately 30%-50% of all SIDS cases, mostly due to hazardous cosleeping environments, such as parental alcohol or drug consumption (which may cause drowsiness), smoking, preterm infants, or sofa or armchair use [3,10]. Additionally, didactic approaches that rely on fear, judgment, or inflexible ideal rules are often ineffective. Many parents report struggling to follow advice they perceive as unrealistic, particularly during periods of routine disruption [11]. In contrast, risk-reduction approaches acknowledge real-world circumstances and provide practical strategies that explain how and why safer sleep practices protect infants [11].

Messaging of safer sleep campaigns must be adapted and reinforced for the current generation of parents, as gaps remain in parental knowledge and adherence to safer sleep guidance, particularly in higher-risk groups. A cross-sectional survey conducted by Pease et al [12] in deprived areas of Bristol, United Kingdom, found that mothers in higher-risk groups were less likely to breastfeed, had more difficulty recalling 2 or more unprompted correct SIDS risk reduction strategies, and scored lower overall on prompted safer sleep scenarios. Notably, only 52% of the 400 mothers surveyed identified infant sleep position as a risk reduction strategy for SIDS [12].

In May 2024, UK adult smartphone users averaged 38 apps on their devices, and 32.5 million visited one of the top 10 health and well-being websites or apps [13]. Of the adult visitors to NHS websites that month, 57% were women [13]. Mobile health (mHealth) apps, which leverage the widespread use of smartphones, provide a cost-effective way to deliver accessible information and influence specific, health-related behaviors within targeted populations [14]. Additionally, with the increase in artificial intelligence (AI), many people, including parents, are turning to AI for information surrounding a multitude of topics; however, there are concerns about AI models such as ChatGPT because they draw their information from what is available on the internet, leading to the potential for misleading or incorrect data [15].

Technology, specifically mHealth apps, is increasingly influencing how parents make decisions about infant sleep. Lambert et al [11] found that most of their participants (28 families) relied on online resources, including sleep tracking apps, social media, and search engines, to guide their understanding of infant sleep, particularly when they perceived contact with health visitors was limited. Younger parents reported a stronger reliance on online sources and influencer content, preferring social media platforms over the NHS website, which they described as overly wordy or jargon-heavy. mHealth apps offer an opportunity to bridge the gap left by traditional services by providing resources that are collaboratively developed, accessible, and succinct, which can help parents plan for safer sleep and navigate routine disruptions, rather than relying solely on traditional didactic approaches [11].

Research surrounding systematic reviews of publicly and commercially available mHealth apps specifically targeting safer sleep for newborns and infants has not yet been conducted in the United Kingdom, although some research does exist on similar topics. For example, Simon et al [16] conducted a

XSL•FO
RenderX

content analysis of smartphone apps for children's sleep; however, they reviewed apps that offer sleep improvement strategies rather than safer sleep guidance for babies and infants. Virani et al [17] reviewed mHealth apps targeting new parents, specifically parents of newborns and infants; however, the study focused on apps that provided sleep tracking, sleeping aids, and photo sharing, and their target sleeping aid applications presented white noise and lullabies to soothe infants to sleep. They did not provide education on safer sleep guidance. Another review by Lee-Tobin et al [18] examined the educational information provided in sleep mHealth apps; however, it was not solely targeted toward newborns and infants. A final review by Grigsby-Toussaint et al [19] excluded apps targeting babies and included those used as sleep alarms. This suggests a gap in the availability of systematic evaluations of mHealth apps targeting parents of infants and newborns that provide guidance on safer sleep practices.

Safer sleep guidance is widely available on the internet, although it is difficult to access evidence-based, reliable guidance via an app without knowing the names of specific organizations or apps that provide this information. This is further emphasized by the increase in misinformation surrounding safer sleep on various social media platforms [20], making it difficult for parents of newborns, especially first-time parents, to access trustworthy information. This systematic mHealth app review will therefore seek to answer the question of what stand-alone safer sleep apps are available for parents and caregivers of newborns and infants in the United Kingdom and what their quality and functionality are. How accessible are they, and do they adhere to best practice guidelines?

Methods

Review Design

A systematic review of safer sleep mHealth apps available on the UK Google Play (Android) and Apple App Store (iOS) was undertaken. This review was conducted using the 7-step process, and the review question was informed by the TECH (target user, evaluation focus, connectedness, and health domain) framework [21] (Textbox 1). The protocol is registered on the Open Science Framework [22], and the review was reported using an initial version of the CAPPRRI (Consensus for APP Review Reporting Items) guidance [23], which was under development at the time of the review. Multimedia Appendix 1 contains the completed CAPPRRI checklist.

Textbox 1. The TECH (target user, evaluation focus, connectedness, and health domain) framework used to determine the review question.

Target user

- Parents and caregivers of newborns up to infancy

Evaluation focus

- Quality, functionality, accessibility, and adherence to best practice guidelines

Connectedness

- Stand-alone

Health domain

- Safer sleep

This review was conducted by a UK pediatric nursing student in her second year of study (SY) and a UK midwifery student in her second year of study (NM), under the guidance of 2 researchers with extensive experience in conducting systematic mHealth app reviews and digital health evaluation (NG and DD). One author is also a registered nurse (DD). The clinical training backgrounds of the reviewers (SY and NM) may have positively influenced the appraisal of app content, particularly with respect to infant safety, clinical credibility, and alignment with evidence-based guidance. To mitigate any potential biases (eg, favoring mHealth interventions), all stages of the review followed a predefined protocol, used explicit inclusion criteria, and applied standardized and reproducible appraisal methods.

Eligibility Criteria, Search, and Selection

As seen in Table 1, the eligibility criteria were determined through application of the TECH framework [21]. The target users were parents or caregivers of newborns and infants (0-12 months), including individuals and families, while apps targeting health care professionals, nurseries, care providers, or organizations were excluded. For the purpose of this review, infants were defined as children aged 12 months or younger, including the neonatal period (0-28 days) and postneonatal period (aged ≤12 months) [9]. The evaluation focus included app quality, functionality, accessibility, and adherence to best practice guidelines; privacy features were excluded because personal data were not expected to be collected. Eligible apps were required to be stand-alone and specifically address safer sleep by providing educational or informational content. Apps that functioned solely as sleep trackers, feeding or milestone trackers, or sound machines were excluded. Apps that cost to download were also excluded.

XSL•FO
RenderX

Table 1. Inclusion and exclusion criteria applied to the TECH (target user, evaluation focus, connectedness, and health domain) framework.

Domain	Inclusion criteria	Exclusion criteria
Target user	• Parents or caregivers of newborns up to infancy (aged 0-12 months) • Individuals or families	Health care professionals, nursery or care providers, or organizations
Evaluation focus	• Quality, functionality, accessibility, and adherence to best practice guidelines	Privacy (do not expect personal data to be collected)
Connectedness	• Stand-alone	Connected to wearable devices, EHRs[a], and health care services
Health domain	• Safer sleep apps (eg, sleep positions, bed set-up, sleep environment, and cosleeping) • Educational or informational	Sleep trackers, feeding, milestones, or sound machines

[a]EHR: electronic health record.

A preliminary search was conducted using no limits or filters and the following key terms: safe sleep, safe sleep guidance for babies, and safe sleep guidance for infants, on the UK Apple App Store and Google Play Store. This confirmed that the number of apps available was manageable for the review and ensured that relevant apps were not missed. These same search terms were used for the final search conducted between January 6 and 13, 2026. The search was done manually, without the use of automated tools.

The name and developer of each app were manually recorded in a Google Sheets spreadsheet. Three lists of apps were produced following the 3 key term searches, and deduplication of apps was conducted separately by both researchers across the keyword searches. One author then screened each app's eligibility based on its name and the written and image descriptions on the respective app store to assess whether it met the eligibility criteria. A second screening and deduplication stage then took place to identify duplicates between the stores and to screen eligibility after each app had been downloaded. Depending on platform availability, researchers downloaded each app onto 1 of 2 devices: an iPhone 11 (iOS system) and a Samsung S22 (Android system). Two researchers (NM and SY) then conducted independent eligibility assessments and reached a consensus. A third researcher (NG) helped resolve any disagreements. The screening of downloaded mHealth apps was conducted by both researchers on the same day to avoid changes in app store search results, which can vary from day to day. A flow diagram was developed to depict the search and screening process.

Apps available on both iOS and Android platforms were retained only once, to determine their multiplatform availability and to confirm which version was the most recent. Only 1 multiplatform app was included in the review, and this was evaluated on the Android device.

Data Extraction

Data were extracted into a predefined data extraction (coding) sheet created within Google Sheets. Descriptive characteristics were extracted, including app name, developer, version number, the app stores where they were available, cost of in-app purchases, size, average user rating, number of user ratings, and affiliation (name of professional health, medical, clinical, or government body affiliated with the app).

Quality was assessed using the 19-item Mobile App Rating Scale (MARS), which evaluates 4 dimensions: engagement, functionality, aesthetics, and information quality [24]. Each item was rated using a 5-point Likert scale: inadequate (1 point), poor (2 points), acceptable (3 points), good (4 points), and excellent (5 points). The authors evaluating the apps (SY and NM) used the step-by-step training video on YouTube to aid their use of MARS and received additional training from the supervising author (NG). The subjective quality scale was excluded to ensure that all evaluations remained fully objective.

Functionality was assessed using the IMS Institute for Healthcare Informatics Functionality Scale, with each of the following functions rated 1 (present) or 0 (absent): inform; instruct; record, collect data, share data, evaluate, intervene; display; guide; remind or alert; and communicate [25]. The researchers also recorded the presence of AI as present (n=1) or absent (n=0). A written summary was made of any AI features.

Readability metrics were used to evaluate accessibility features, including the Flesch-Kincaid metrics integrated within Microsoft Word. The Flesch Reading Ease score provided a value between 0 and 100, where higher scores reflect that the content is easier to read [26]. Alongside this, the Flesch-Kincaid Grade Level offered an approximate US school grade required to understand the text [27]. It was recorded whether the app enabled users to change the language and whether dark or night mode features were available, either by reflecting the default phone settings or allowing in-app adjustments. These features were considered important for parents using the app during nighttime care, as they may help to minimize sleep disruption.

Each app's adherence to best practice guidelines for safer sleep and cosleeping, defined as any shared sleep between an adult and a baby (eg, on a bed, sofa, or chair) [8], was evaluated using bespoke criteria informed by evidence-based guidance [4,5,28]. The presence of the following information was rated as present (n=1) or absent (n=0) and then summed for a total score:

- Lie your baby on their back
- Keep their cot clear

XSL•FO
RenderX

- Use a firm, flat, and waterproof mattress
- Recommend against bed-sharing (shared sleep between an adult and a baby on an adult bed for the majority of the night) if the baby has a low birth weight, or either parent has had 2 or more units of alcohol, smokes, is taking medication that causes drowsiness, or has used recreational drugs
- Avoid your baby getting too hot
- Sleep your baby in the same room as you for at least the first 6 months
- Do not sleep on a sofa or chair with the baby
- If bed-sharing: do not have pillows or duvets near the baby
- If bed-sharing: do not have other children or pets in the bed

Data Synthesis and Analysis

Descriptive statistics were calculated in Microsoft Excel for all relevant items related to app characteristics, quality, functionality, accessibility, and adherence to best-practice guidelines. Text-based summaries (eg, description of any AI features) were summarized using a content synthesis approach. Results were presented using tables and figures (eg, a radar graph), where applicable. The apps were cross-compared using criteria related to quality, functionality, and adherence to best-practice guidelines to identify the highest-scoring apps.

Interrater reliability was also calculated to determine levels of agreement between the researchers conducting the quality and functionality evaluations using MARS and the IMS Institute for Healthcare Informatics functionality scores. These were calculated using IBM SPSS. We calculated an intraclass correlation coefficient for all MARS items using an absolute-agreement 2-way mixed-effects, average-measures model [29]. Interpretations were as follows: <0.5 (poor), 0.5-0.75 (moderate), 0.75-0.90 (good), and >0.90 (excellent) [30]. As the IMS Institute for Healthcare Informatics functionality score uses a binary measure (categories: present=1 and absent=0), we calculated Cohen κ. Interpretations of agreement were <0 (poor), 0.01-0.20 (slight), 0.21-0.40 (fair), 0.41-0.60 (moderate), 0.61-0.80 (substantial), and 0.81-1.00 (almost perfect) [31].

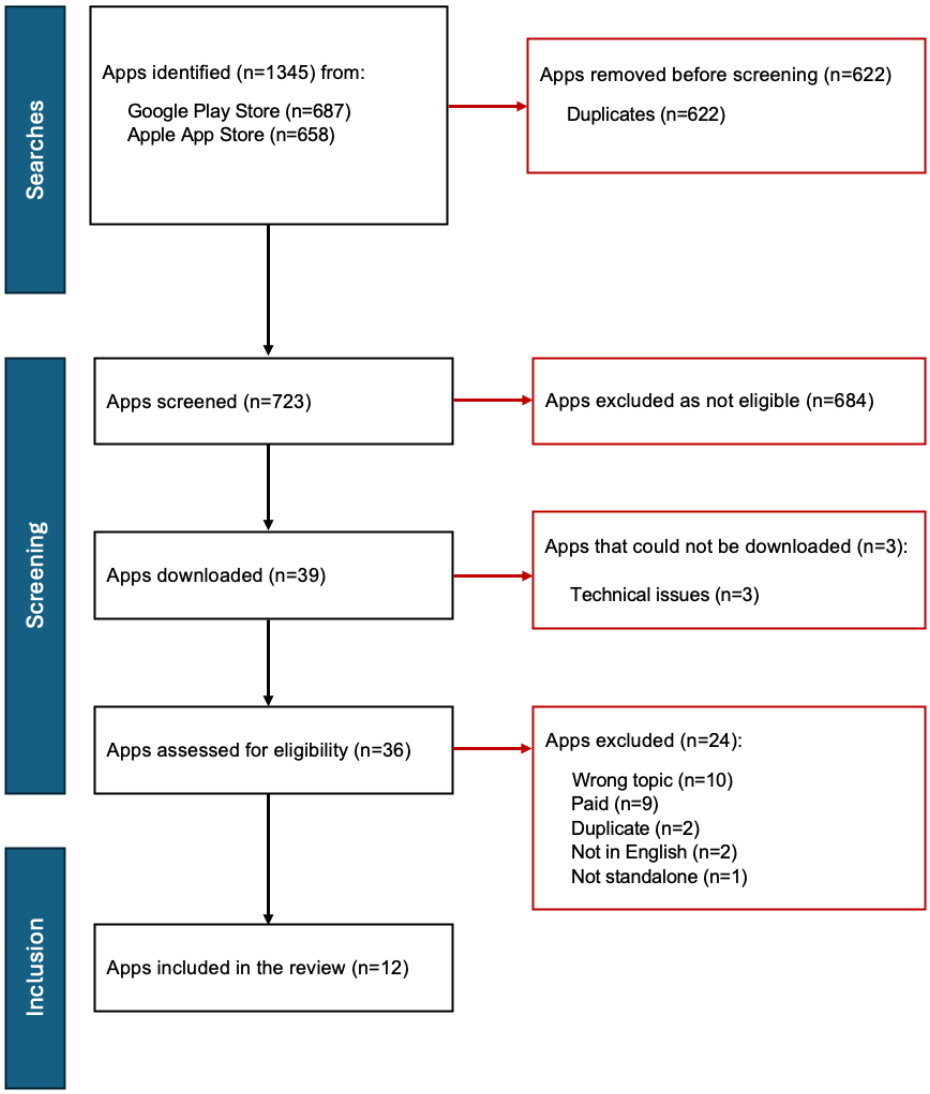
Results

App Market Search and Screening

The initial search yielded 1345 apps. A total of 622 duplicates were removed: 297 from the Google Play Store, 286 from the Apple App Store, and an additional 39 identified between the 2 stores. A total of 723 apps were screened based on their names and the text and graphic descriptions in the app stores, leading to the exclusion of 684 apps as they did not meet the eligibility criteria. After this, 39 apps were downloaded for the second screening stage; however, 3 did not open or could not be accessed. Apps were excluded for the following reasons: addressing the wrong topic (n=10), requiring paid subscriptions or payments to access in-app content and features (n=9), not being in English (n=2), and not being stand-alone (n=1). Two others were duplicates; however, because the apps and developers had different names, they were not identified as duplicates during the initial screening stage. As a result, 12 apps were included in the review. Figure 1 presents the CAPPRRI flowchart summarizing the app search and screening process for the review.

XSL•FO
RenderX

Figure 1. CAPPRRI (Consensus for APP Review Reporting Items) flowchart presenting the app search and screening process.



Descriptive Characteristics

Of the 12 apps included in the review, 50% (n=6) were available on the Apple App Store, and 42% (n=5) were available on the Google Play Store; only 1 (8%) was available on both. All were free to download, although 5 (42%) contained in-app purchases to unlock premium or additional content. Two (17%) apps were affiliated with a professional health, medical, clinical, or government body. The average app size was 99.34 (SD 54.10) MB, ranging from 17.70 MB to 232.70 MB. Of the 10 apps that had been rated by users, the total number of user ratings was 571,016, ranging from 1 to 340,000. An average user rating score of 4.11 out of 5 was calculated. Table 2 provides a summary of the app characteristics, and Multimedia Appendix 2 provides a list of all 12 apps and a plain language summary of the review.

XSL•FO
RenderX

Table 2. Key characteristics and content of the reviewed apps (N=12).

Characteristic	Value
Market availability, n (%)	
Apple Store	6 (50)
Google Play Store	5 (42)
Both	1 (8)
Cost, n (%)	
Free to download	12 (100)
In-app purchases	5 (42)
Affiliated with a professional health, medical, clinical, or government body, n (%)	
Yes	2 (17)
No	10 (83)
The user can change the language, n (%)	
Yes	0 (0)
No	12 (100)
The user can change to night/dark mode, n (%)	
Yes	3 (25)
No	9 (75)
Sleep guidance mentioned, n (%)	
Lie your baby on their back	12 (100)
Keep their cot clear	10 (83)
Use a firm, flat, and waterproof mattress	11 (92)
Recommend against bed-sharing (shared sleep between an adult and a baby on an adult bed for the majority of the night) if the baby has a low birth weight, or either parent has had 2 or more units of alcohol, smokes, is taking medication that causes drowsiness, or has used recreational drugs	11 (92)
Avoid your baby getting too hot	12 (100)
Sleep your baby in the same room as you for at least the first 6 months	10 (83)
Do not sleep on a sofa or chair with the baby	8 (67)
If bed-sharing: do not have pillows or duvets near the baby	4 (33)
If bed-sharing: do not have other children or pets in the bed	3 (25)
Readability	
Reading ease score, mean (SD)	76.6 (SD 7.98)
Reading age (US grade level)	6.9

Sleep Guidance

While all 12 apps contained some information regarding safer sleep practices, this was addressed and named differently (eg, calling it cot death, SIDS, or safe sleep). Of the 9 criteria pertaining to safer sleep best practice guidance, the apps mentioned an average of 6.75 criteria (SD 1.71; range 3-9). Only 2 apps mentioned all the criteria.

All 12 apps mentioned placing the infant on their back as well as avoiding the baby getting too hot. Additionally, 92% (n=11) of the apps addressed the importance of using a firm, flat, and waterproof mattress and 92% (n=11) recommended against bed-sharing (shared sleep between an adult and a baby on an adult bed for the majority of the night [8]) if the baby has a low birth weight or either parent has had 2 or more units of alcohol, smokes, is taking medication that causes drowsiness, or has used recreational drugs. Most (10/12, 83%) mentioned keeping the cot clear and sleeping in the same room as the baby for at least the first 6 months. However, only 33% (n=4) mentioned not having pillows or duvets in the bed if parents choose to bed-share, and only 25% (n=3) highlighted the importance of not having any other person or pet in the bed if parents choose to bed-share. Table 2 presents the criteria and how many apps mentioned each one.

Two (17%) of the apps contained information that contradicts safer sleep guidance. Both highlighted practices such as swaddling and cosleeping with no additional information on how to do so safely, for example, not swaddling above the

XSL•FO
RenderX

shoulders, keeping swaddles loose around the hips, and only cosleeping with 1 adult in the bed and no pillows or blankets.

Accessibility

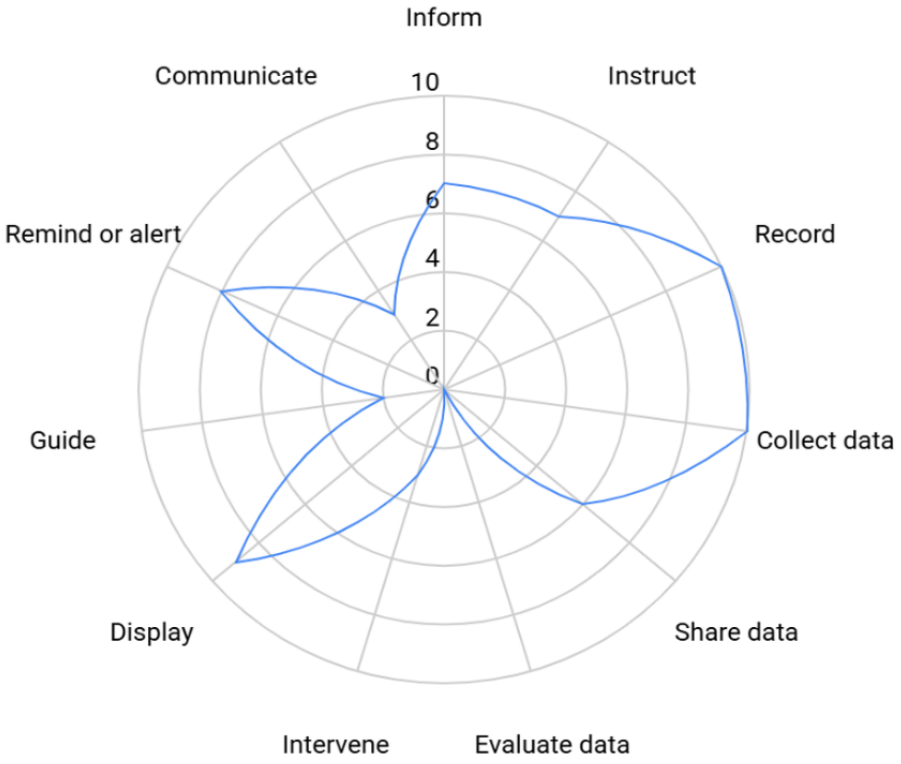
The average reading ease score was 76.6 (SD 7.98; range 62.90-91.40). All apps scored above 60, meaning that most adults should be able to understand the content. The average reading age was US grade level 6.9, ranging from the fifth grade to the ninth grade. However, 1 app exclusively provided content in video format, and therefore, readability scores could not be generated. None of the apps provided the ability to change the language (beyond English), and only 3 (25%) could be switched to night/dark mode either through app settings or by automatically adapting to the device's settings.

Functionality

The IMS Institute for Healthcare Informatics functionality score was used to identify functions within each app. There was moderate agreement between the 2 raters (κ=0.494, 95% CI 0.353-0.635; *P*<.001). Consensus was met on all items.

The apps had an average of 5.4 (SD 2.5) functions. The range of functions was from 2 to 9. The 3 most common functions were record (10/12, 83%), collect data (10/12, 83%), and display (9/12, 75%). The 4 least common functions were evaluate data (0/12, 0%), guide (2/12, 17%), and communicate and intervene (both 3/12, 25%; Figure 2).

Figure 2. Radar graph showing app functionality using the IMS Institute for Healthcare Informatics functionality score for 11 items.



A total of 2 apps (Baby Buddy: Pregnancy & Parent and Pregnancy Tracker & Baby App) had 9 out of the 11 functions, and 3 apps (Little One, Baby+ Your Baby Tracker, and Infant sleep info) had 7 functions. There were 2 apps (SIDS Info and Baby Sleep Magic) that only had 2 out of the 11 functions. The remaining apps ranged from having 3 to 6 functions.

Only 1 (8%) app contained an AI assistant. The application, "Bounty: Pregnancy and Baby App," featured a "friendly AI Midwife Assistant" that was available to answer questions until a human midwife could answer between Monday and Friday from 10 AM to 2 PM.

Quality

All apps underwent an independent evaluation carried out by 2 raters using MARS. There was excellent agreement between the raters (intraclass correlation coefficient 0.96, 95% CI 0.948-0.969; *P*<.001).

The mean overall quality score was 3.7 (SD 0.43; range 2.92-4.13). Almost all applications, 92% (11/12) met the minimum acceptability score of 3. The apps "Little Ones" and "Baby Buddy: Pregnancy and Parent" had the highest overall rating (4.1), and "Asianparent: Pregnancy & Baby" had the worst rating (2.9). Mean scores for engagement, functionality, aesthetics, and information quality were 3.3 (SD 0.56; range 2.4-4.3), 4.1 (SD 0.58; range 2.88-5.00), 3.6 (SD 0.73; range 2.67-4.67), and 3.8 (SD 0.68; range 2.38-4.80), respectively. Half of the apps (6/12, 50%) scored the highest on the functionality dimension, 4 apps (33%) scored the highest on the aesthetics dimension, 2 (17%) scored the highest on the information quality dimension, and none scored the highest on the engagement dimension. Three apps (25%) scored the lowest on the engagement dimension, 1 (8%) scored the lowest on the functionality dimension, 4 (33%) scored the lowest on the aesthetics dimension, and 3 scored the lowest on the information quality dimension.

XSL•FO
RenderX

Only 2 (17%) apps, "Baby Buddy: Pregnancy & Parent" and "Bounty: Pregnancy and Baby App," had been formally trialed or tested. However, one yielded positive or partially positive results [32,33] while the other received critical results for its content as it provided contradictory information around avoiding alcohol consumption but also limiting this to 1-2 units once or twice per week [34]. It is important to note that we did not verify whether this contradiction has been amended or whether the app has been updated. None of the apps had been tested in randomized controlled trials.

Cross-Comparison of Highest-Scoring Apps

A cross-comparison of the apps with the highest IMS functionality and MARS scores was undertaken. The highest performing apps and their safer sleep guidance score (out of 9) are presented in Table 3.

Table 3. The highest performing apps included in the review.

App; developer (version)	Cost; platform	IMS score[a]	Total MARS[b] score	Purpose	Safer sleep guidance criteria[c]
Pregnancy Tracker & Baby App; What to Expect (version 7.81)	Free; Google Play Store	9	4.13	Baby growth tracker that provides general advice and guidance around pregnancy and infancy	7
Baby Buddy: Pregnancy & Parent; Babyzone (version 1.2.0)	Free; Google Play Store	9	4.09	Provides personalized information and guidance through pregnancy and infancy	8
Little Ones; Little Ones (version 7.6.4)	GBP £59- GBP £99[d] (in-app); Apple App Store	7	4.09	Provides a comprehensive sleep program for babies and infants	6

[a]IMS Institute for Healthcare Informatics functionality score (maximum score=11).

[b]MARS: Mobile App Rating Scale; maximum score=5.

[c]Safer sleep guidance criteria score (maximum score=9).

[d]GBP £1=USD $1.32 as of June 26, 2026.

Discussion

Overview

This systematic mHealth app review found that there were 12 stand-alone safer sleep apps free to download for parents and caregivers of newborns up to infancy in the United Kingdom. Almost all met the minimum acceptability criteria for quality. The apps ranged in functionality, with an average of 5.4 functions each, although only 1 had an AI feature. A total of 11 apps had written content and could be considered to contain accessible language, as demonstrated by reading ease scores of more than 60. However, none of the apps provided the option to change the language, creating a barrier for people who cannot speak or read English, and very few could change to night/dark mode, limiting usability in dark rooms. Regarding alignment to best practice guidelines, the apps mostly did not provide comprehensive content—most met at least half of the criteria, but only 2 presented all the criteria. Some apps also promoted unsafe practices.

This review showed that there is a limited number of safer sleep guidance apps available across both stores. It was challenging to find apps that contained safer sleep guidance because of the vast number of apps catering to babies, infants, and sleep, many of which were simply sleep trackers, white noise apps, alarms, or baby growth trackers. A large percentage of the apps reviewed were not limited to safer sleep guidance apps, with only 1 focusing solely on providing safer sleep guidance. The others also included general baby and infant safety advice, and many were growth and sleep trackers with guidance sections focusing on safer sleep as a small component. This finding is consistent with a previous review of the app market in Germany, finding that the content of most sleep apps was lullabies, music, or songs, and sleep information was very rare [35]. Similarly, Ejezie et al [36] concluded that digital health strategies for SIDS prevention remain limited, and mostly include online interventions (eg, web-based modules).

Although almost all of the apps we reviewed (11/12, 92%) did meet the minimum acceptability criterion for quality, only 2 of them managed to fully adhere to best practice guidelines as outlined by established UK health bodies, including NICE [28], the NHS [4], and the Lullaby Trust [5]. A few also mentioned the use of techniques that did not align with best practice, such as swaddling, and some did not include further guidance on how to do so safely. The aforementioned review conducted by Schlarb et al [35] also found that there was only a small percentage of apps available in Germany related to sleep information for parental behavior. An analysis of web search results by Chung et al [37] related to safer sleep recommendations found that only 43.5% of the 1300 websites provided accurate information. This further emphasizes the gap in accurate, safer sleep guidance and recommendations observed through this review.

Furthermore, it is important to note that although only 1 of the apps reviewed in this study used AI, this feature is increasingly being incorporated into mHealth apps [38], which raises additional considerations around data privacy and the uncertainty of AI-generated results (especially when delivering personalized medical care) [39].

Not many of the apps had been trialed or tested, and so it remains unclear whether the reviewed apps are effective in promoting safer sleep behaviors and in reducing SIDS. Previous research has shown that although families are often aware of

XSL•FO
RenderX

safer sleep guidance, this does not always translate into practice, especially when the guidance is impractical to implement [40-42]. Information-only resources (including those delivered via smartphone apps) may therefore have limited effectiveness if not provided with contextual support. There has been 1 randomized controlled trial conducted by Moon et al [43] with 1263 mothers in the United States, finding that the intervention group receiving the mHealth intervention with regular text and email messages and videos had significantly higher rates of safer sleep practice compared to the control group. However, the authors point to further research needed to assess long-term impact in order to establish whether adherence to best practice sleep guidance contributes to a reduction in SIDS and whether widespread implementation is possible [43]. The requirement for further research on whether engagement with mHealth interventions correlates with changes in safer sleep practices is also emphasized by Krishnamurti et al [44], who integrated safer sleep information into an existing app, and Aggelou et al [45], who systematically reviewed 23 studies on educational safer sleep interventions targeting caregivers.

It is also important to explore the impact of these interventions in the context of parents receiving alternative advice, as well as on their effects on parental attitudes and norms regarding safer sleep recommendations. This review recognizes that adherence to safer sleep guidance can become secondary when dealing with the responsibilities of caring for a newborn, and parents who opt to bed-share do so for a multitude of reasons, among them to provide comfort, for bonding, and for better/more sleep for both the infant and parent [46]. Additionally, previous research from Australia, New Zealand, and internationally has highlighted that cosleeping (including bed-sharing) is common [42,47,48], either intentionally or unintentionally, during feeding or periods of exhaustion [47,49]. As a result, many families develop their own safer sleep strategies, such as choosing to sleep on a chair rather than in a bed [40,49]. Our research reflects this broader issue, as three-quarters of the apps did not provide strategies for safer bed-sharing, despite the likelihood that many families will encounter these situations. Current apps available in the United Kingdom, therefore, do not adequately support parents across the full spectrum of real-world caregiving scenarios, including unintentional shared sleep and culturally normalized bed-sharing contexts.

Consequently, it is essential to provide context-specific support to families and assess alternative advice given to parents who cannot or do not follow safer sleep guidance, for example, how to safely cosleep or swaddle [43]. Ball and Keegan [50] also highlight that parents often use social media to share information around sleep and this, along with infant sleep apps, may influence cultural narratives around nighttime care. They caution that costly digital sleep technologies may undermine responsive parenting practices and capitalize on parental anxiety. Evidence also suggests that mHealth interventions can enhance adherence to safer sleep guidance by positively influencing maternal attitudes and social norms related to specific recommendations (eg, supine sleeping and room-sharing without cosleeping) [51]. Research is warranted in these areas.

Implications

As outlined above, research is needed to determine the long-term impacts of mHealth apps on preventing SIDS and to explore the context in which parents receive and follow safer sleep guidance. However, there are also several implications for app developers and established safer sleep organizations.

As already discussed, many of the apps were not limited to providing only safer sleep advice. It is therefore important to acknowledge that many people will not be interested in downloading an app that simply contains information on safer infant sleep, as this can already be found online. Rather than developing new apps, we therefore recommend that established organizations such as The Lullaby Trust [5,8] incorporate safer sleep information into the apps they have developed and manage. This is because the organization is a recognizable and trusted name in the United Kingdom that many health care workers already signpost new parents to. For example, their app “Lullaby Trust Baby Check” is already used by parents to determine whether their baby needs to see a health care professional. In a review of apps for managing acute childhood illness, this app was among the highest rated for aligning with end user requirements and was found to already support parents in the United Kingdom [52]. Another service evaluation found that most health visitors were already familiar with it and recommended it to parents of infants up to 6 months old [53]. Interestingly, and consistent with our findings, the evaluation found that health visitors wanted the app to be translated into other languages [53]. It is vital that any content for diverse communities be co-designed with them, given that there are deep-rooted cultural and traditional sleep practices that must be considered. A brief section on safer sleep (including how best to support families who are already bed-sharing) within the app could also leverage The Lullaby Trust’s reputation and help reduce digital waste by reducing the need for a new app to be developed.

Although non–app-based tools such as dedicated websites exist, it can be more convenient to have a smartphone app that already contains evidence-based information regarding safer sleep, particularly as this removes the burden of finding information and may be used offline. Developers who are considering creating apps about infancy or improving their existing apps should include engaging features that provide advice and guidance around safer sleep practices. For example, an American project leveraged simple push notifications in order to increase engagement with safer sleep and breastfeeding education modules within a pregnancy app, which was found to be particularly effective in engaging parents living in high-risk zip codes where SIDS education is most likely to have an impact [44]. When producing this content, developers should partner with health and care workers who have expertise in safer sleep practices and can translate the best practice guidance into easy-to-understand (and reliable) content. Information could be presented in a variety of ways to accommodate different visual and auditory preferences, for example, diagrams, pictures, videos, audio recordings, and writing. This, along with the ability to change languages and to switch to nighttime mode, is important to ensure maximum accessibility and engagement.

XSL•FO
RenderX

It is important to acknowledge that the criteria used to evaluate safer sleep apps in this review take a primarily risk elimination approach, strongly advising against cosleeping. A review by Kruse et al [54] on safer sleep messaging in Australia also highlighted that less than half of the organizations they reviewed provided strategies to engage in safer shared sleep and acknowledged that shared sleep may occur unintentionally or be considered a familial, cultural, or logistical preference. Families often do not disclose cosleeping to health care practitioners [49], potentially due to anticipating judgment or disapproval. A shift toward sensitive and consistent messaging that better reflects the realities of baby sleeping is needed, and this should ideally employ a risk minimization strategy, acknowledging that cosleeping does occur and that risks can be minimized [54,55]. This approach should be used in the future to evaluate safer sleep apps and the content within these apps.

Strengths and Limitations

This review followed a structured and transparent 7-step process informed by appropriate frameworks, while validated measures were used to evaluate quality, functionality, and readability. Two raters were involved in the screening and evaluations, with strong agreement. The review was further strengthened by examining adherence to evidence-based UK safer sleep guidance.

Limitations included the restriction to free-to-download apps available on the UK Apple App Store and Google Play Store at the time of the search, which may not capture apps available in other regions, other stores, or that have since become available. Additionally, conclusions regarding the impact of these apps on safer sleep practices cannot be drawn, as the review did not assess changes in parental behavior or downstream outcomes. It is therefore unclear whether the highest-scoring apps are the most effective.

Conclusion

This systematic mHealth app review identified a limited number of safer sleep guidance apps for parents and caregivers of newborns and infants across the 2 app stores in the United Kingdom. Most of these met the minimum acceptability criteria for quality and contained accessible written content. However, most of the reviewed apps did not provide comprehensive safer sleep guidance, and it was concerning to note that some encouraged practices that are against best practice guidelines. Further long-term research is needed on the effectiveness of mHealth interventions in improving adherence to safer sleep guidance and their ability to reduce SIDS. Information related to best practice guidance around safer sleep and preventing SIDS should be included in established newborn and infant apps, in partnership with trusted organizations, and should be available in several languages. This may result in the development of safer sleep guidance applications that are both trusted and accessible.

Funding

This work is funded by the National Institute for Health and Care Research (NIHR) Undergraduate Internship Programme (NIHR305965) and was supported by the National Institute for Health and Care Research Applied Research Collaboration Greater Manchester. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Authors' Contributions

NG acquired funding and provided supervision. SY and NM conceptualized the study. SY, NM, NG, and DD developed the study protocol. SY and NM conducted the app market searches, screening, data extraction, and analysis. SY, NM, and NG drafted the manuscript. SY, NM, NG, and DD reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CAPPRRI checklist.
[PDF File (Adobe PDF File), 1662 KB - mhealth_v14i1e95642_app1.pdf]

Multimedia Appendix 2
Additional table and the plain English summary.
[DOCX File , 22 KB - mhealth_v14i1e95642_app2.docx]

References

1. Sudden infant death syndrome (SIDS). National Health Service. 2021. URL: https://www.nhs.uk/conditions/sudden-infant-death-syndrome-sids/ [accessed 2026-06-18]
2. International Classification of Diseases 11th Revision (ICD-11) for mortality and morbidity statistics. World Health Organization. 2025. URL: https://icd.who.int/browse/2025-01/mms/en#56255971%2Funspecified [accessed 2026-06-18]

XSL•FO
RenderX

3. Caring for your baby at night: a guide for parents. UNICEF UK Baby Friendly Initiative. 2022. URL: https://www.unicef.org.uk/babyfriendly/wp-content/uploads/sites/2/2018/08/Caring-for-your-baby-at-night-web.pdf [accessed 2026-06-18]
4. Baby safer sleep advice. Best Start in Life. 2023. URL: https://www.nhs.uk/best-start-in-life/baby/baby-basics/newborn-and-baby-sleeping-advice-for-parents/safe-sleep-advice-for-babies/ [accessed 2026-06-18]
5. Safer sleep overview. The Lullaby Trust. 2025. URL: https://www.lullabytrust.org.uk/baby-safety/safer-sleep-information/safer-sleep-overview/ [accessed 2026-06-18]
6. Postnatal care of the baby. NICE guideline NG194. URL: https://www.nice.org.uk/guidance/ng194/chapter/Recommendations#postnatal-care-of-the-baby [accessed 2026-06-18]
7. Rechtman LR, Colvin JD, Blair PS, Moon RY. Sofas and infant mortality. Pediatrics 2014;134(5):e1293-e1300 [FREE Full text] [doi: 10.1542/peds.2014-1543] [Medline: 25311597]
8. Co-sleeping with your baby. The Lullaby Trust. 2025. URL: https://www.lullabytrust.org.uk/baby-safety/safer-sleep-information/co-sleeping/ [accessed 2026-06-18]
9. Unexplained deaths in infancy, England and Wales: 2023. Office for National Statistics. 2025. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/unexplaineddeathsininfancyenglandandwales/2023 [accessed 2026-06-18]
10. Blair PS, Sidebotham P, Pease A, Fleming PJ. Bed-sharing in the absence of hazardous circumstances: is there a risk of sudden infant death syndrome? An analysis from two case-control studies conducted in the UK. PLoS One 2014;9(9):e107799 [FREE Full text] [doi: 10.1371/journal.pone.0107799] [Medline: 25238618]
11. Lambert B, Keegan AA, Ingram J, Blair PS, Fleming PJ, Pease A. "Okay in theory": a qualitative study of safer sleep advice in families with infants at risk and families reporting risky sleep practices. BMJ Paediatr Open 2025;9(1):e003620 [FREE Full text] [doi: 10.1136/bmjpo-2025-003620] [Medline: 40707137]
12. Pease AS, Blair PS, Ingram J, Fleming PJ. Mothers' knowledge and attitudes to sudden infant death syndrome risk reduction messages: results from a UK survey. Arch Dis Child 2018;103(1):33-38. [doi: 10.1136/archdischild-2017-312927] [Medline: 28814422]
13. Online Nation 2024. Ofcom. 2024. URL: https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/online-research/online-nation/2024/online-nation-2024-report.pdf?v=386238 [accessed 2026-06-18]
14. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
15. Khlaif ZN, Mousa A, Hattab MK, Itmazi J, Hassan AA, Sanmugam M, et al. The potential and concerns of using AI in scientific research: ChatGPT performance evaluation. JMIR Med Educ 2023;9:e47049 [FREE Full text] [doi: 10.2196/47049] [Medline: 37707884]
16. Simon SL, Kaar JL, Talker I, Reich J. Evidence-based behavioral strategies in smartphone apps for children's sleep: content analysis. JMIR Pediatr Parent 2022;5(1):e32129 [FREE Full text] [doi: 10.2196/32129] [Medline: 35238787]
17. Virani A, Duffett-Leger L, Letourneau N. Parenting apps review: in search of good quality apps. Mhealth 2019;5:44 [FREE Full text] [doi: 10.21037/mhealth.2019.08.10] [Medline: 31620471]
18. Lee-Tobin PA, Ogeil RP, Savic M, Lubman DI. Rate my sleep: Examining the information, function, and basis in empirical evidence within sleep applications for mobile devices. J Clin Sleep Med 2017;13(11):1349-1354 [FREE Full text] [doi: 10.5664/jcsm.6814] [Medline: 28992831]
19. Grigsby-Toussaint DS, Shin JC, Reeves DM, Beattie A, Auguste E, Jean-Louis G. Sleep apps and behavioral constructs: a content analysis. Prev Med Rep 2017;6:126-129 [FREE Full text] [doi: 10.1016/j.pmedr.2017.02.018] [Medline: 28316907]
20. Jawed A, Ehrhardt C, Rye M. Handle with care: a narrative review of infant safe sleep practices across clinical guidelines and social media to reduce SIDS. Children (Basel) 2023;10(8):1365 [FREE Full text] [doi: 10.3390/children10081365] [Medline: 37628364]
21. Gasteiger N, Dowding D, Norman G, McGarrigle L, Eost-Telling C, Jones D, et al. Conducting a systematic review and evaluation of commercially available mobile applications (apps) on a health-related topic: the TECH approach and a step-by-step methodological guide. BMJ Open 2023;13(6):e073283 [FREE Full text] [doi: 10.1136/bmjopen-2023-073283] [Medline: 37308269]
22. Open Science Framework. URL: https://osf.io/5p3tn [accessed 2026-06-24]
23. Gasteiger N, Norman G, Grainger R, Eost-Telling C, Jones D, Ali MS, et al. Consensus for APP Review Reporting Items (CAPPRRI). Reporting guidelines under development for systematic reviews. 2025. URL: https://www.equator-network.org/library/reporting-guidelines-under-development/reporting-guidelines-under-development-for-systematic-reviews/#APP [accessed 2026-06-18]
24. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
25. Aitken M, Gauntlett C. Patient apps for improved healthcare: from novelty to mainstream. IMS Institute for Healthcare Informatics. 2013. URL: http://ignacioriesgo.es/wp-content/uploads/2014/03/iihi_patient_apps_report_editora_39_2_1.pdf [accessed 2026-06-18]
26. Flesch R. How to Write Plain English: A Book for Lawyers and Consumers. New York: Harper & Row; 1979.

XSL•FO
RenderX

27. Kincaid J, Fishburne RP, Rogers RL, Chissom BS. Derivation of new readability formulas (Automated Readability Index, Fog Count And Flesch Reading Ease Formula) for navy enlisted personnel. Research Branch Report 8-75. 1975. URL: https://stars.library.ucf.edu/cgi/viewcontent.cgi?article=1055&context=istlibrary [accessed 2026-06-18]
28. Quality statement 5: safer practices for bed sharing. Postnatal Care. Quality Standard QS37. 2022. URL: https://www.nice.org.uk/guidance/qs37/chapter/Quality-statement-5-Safer-practices-for-bed-sharing [accessed 2026-06-18]
29. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86(2):420-428. [doi: 10.1037//0033-2909.86.2.420] [Medline: 18839484]
30. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
31. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159-174. [Medline: 843571]
32. Deave T, Ginja S, Goodenough T, Bailey E, Piwek L, Coad J, et al. The bumps and BaBies longitudinal study (BaBBLeS): a multi-site cohort study of first-time mothers to evaluate the effectiveness of the Baby Buddy app. Mhealth 2019;5:42 [FREE Full text] [doi: 10.21037/mhealth.2019.08.05] [Medline: 31620469]
33. Bailey E, Nightingale S, Thomas N, Coleby D, Deave T, Goodenough T, et al. First-time mothers' understanding and use of a pregnancy and parenting mobile app (The Baby Buddy App): qualitative study using appreciative inquiry. JMIR Mhealth Uhealth 2022;10(11):e32757 [FREE Full text] [doi: 10.2196/32757] [Medline: 36409530]
34. Bland C, Dalrymple KV, White SL, Moore A, Poston L, Flynn AC. Smartphone applications available to pregnant women in the United Kingdom: an assessment of nutritional information. Matern Child Nutr 2020;16(2):e12918 [FREE Full text] [doi: 10.1111/mcn.12918] [Medline: 31833237]
35. Schlarb AA, Kater M, Werner A, Hertel D, Lollies F, Landwehr J, et al. Sleep apps for children—a critical view. Somnologie (Berl) 2021;25(1):4-10. [doi: 10.1007/s11818-021-00297-6]
36. Ejezie CL, Sacca L, Burgoa S, Zerrouki Y, Oladoyin OO, Anunobi PC, et al. Digital health strategies for sudden infant death syndrome prevention: a scoping review. Prev Med 2025;200:108412. [doi: 10.1016/j.ypmed.2025.108412] [Medline: 40934978]
37. Chung M, Oden RP, Joyner BL, Sims A, Moon RY. Safe infant sleep recommendations on the internet: let's Google it. J Pediatr 2012;161(6):1080-1084 [FREE Full text] [doi: 10.1016/j.jpeds.2012.06.004] [Medline: 22863258]
38. Morelli S, Giansanti D. Recent advances in AI-driven mobile health enhancing healthcare-narrative insights into latest progress. Bioengineering (Basel) 2025;13(1) [FREE Full text] [doi: 10.3390/bioengineering13010054] [Medline: 41595987]
39. Deniz-Garcia A, Fabelo H, Rodriguez-Almeida AJ, Zamora-Zamorano G, Castro-Fernandez M, Alberiche Ruano MDP, et al. Quality, usability, and effectiveness of mHealth apps and the role of artificial intelligence: current scenario and challenges. J Med Internet Res 2023;25:e44030 [FREE Full text] [doi: 10.2196/44030] [Medline: 37140973]
40. Cole R, Young J, Kearney L, Thompson JMD. Challenges parents encounter when implementing infant safe sleep advice. Acta Paediatr 2021;110(11):3083-3093. [doi: 10.1111/apa.16040] [Medline: 34297875]
41. Cole R, Young J, Kearney L, Thompson JMD. Awareness of infant safe sleep messages and associated care practices: findings from an Australian cohort of families with young infants. BMJ Paediatr Open 2021;5(1):e000972 [FREE Full text] [doi: 10.1136/bmjpo-2020-000972] [Medline: 33718628]
42. Cole R, Young J, Kearney L, Thompson JMD. Infant care practices and parent uptake of safe sleep messages: a cross-sectional survey in Queensland, Australia. BMC Pediatr 2020;20(1):27 [FREE Full text] [doi: 10.1186/s12887-020-1917-5] [Medline: 31964354]
43. Moon RY, Hauck FR, Colson ER, Kellams AL, Geller NL, Heeren T, et al. The effect of nursing quality improvement and mobile health interventions on infant sleep practices: a randomized clinical trial. JAMA 2017;318(4):351-359 [FREE Full text] [doi: 10.1001/jama.2017.8982] [Medline: 28742913]
44. Krishnamurti T, Moon R, Richichi R, Berger R. Integrating infant safe sleep and breastfeeding education into an app in a novel approach to reaching high-risk populations: prospective observational study. JMIR Pediatr Parent 2025;8:e65247 [FREE Full text] [doi: 10.2196/65247] [Medline: 39810395]
45. Aggelou M, Metallinou D, Dagla M, Vivilaki V, Sarantaki A. Evaluating educational patterns and methods in infant sleep care: trends, effectiveness, and impact in home settings-A systematic review. Children (Basel) 2024;11(11):1337 [FREE Full text] [doi: 10.3390/children11111337] [Medline: 39594912]
46. Barry ES, McKenna JJ. Reasons mothers bedshare: a review of its effects on infant behavior and development. Infant Behav Dev 2022;66:101684. [doi: 10.1016/j.infbeh.2021.101684] [Medline: 34929477]
47. Mileva-Seitz VR, Bakermans-Kranenburg MJ, Battaini C, Luijk MPCM. Parent-child bed-sharing: the good, the bad, and the burden of evidence. Sleep Med Rev 2017;32:4-27. [doi: 10.1016/j.smrv.2016.03.003] [Medline: 27107752]
48. Carter ML, Paine SJ, Sweeney BM, Taylor JE, Leigh Signal T. Characterizing the sleep location, patterns, and maternally perceived sleep problems of the infants of Māori and non-Māori mothers in Aotearoa New Zealand. Sleep Health 2025;11(2):158-165 [FREE Full text] [doi: 10.1016/j.sleh.2024.10.010] [Medline: 39755499]
49. Grubb C, Young J, Downer T, D'Souza L. Beyond the rules: an integrative review of parental perspectives on safer infant sleep in shared environments. Front Public Health 2025;13:1629678 [FREE Full text] [doi: 10.3389/fpubh.2025.1629678] [Medline: 41030357]

XSL•FO
RenderX

50. Ball HL, Keegan AA. Digital health tools to support parents with parent-infant sleep and mental well-being. NPJ Digit Med 2022;5(1):185 [FREE Full text] [doi: 10.1038/s41746-022-00732-4] [Medline: 36543920]
51. Moon RY, Corwin MJ, Kerr S, Heeren T, Colson E, Kellams A, et al. Mediators of improved adherence to infant safe sleep using a mobile health intervention. Pediatrics 2019;143(5):e20182799 [FREE Full text] [doi: 10.1542/peds.2018-2799] [Medline: 31015374]
52. Rathbone AL, Neill S, Prime S, Thomas N, Everett K. Supporting parents with acutely ill children: environment scan and user evaluation of mobile applications (the SuPa kids project). Int J Med Inform 2024;189:105504. [doi: 10.1016/j.ijmedinf.2024.105504] [Medline: 38833841]
53. Maramba IDC, Menzies JC, Lim E, Nijman RG, Zhou SM, Latour JM. Healthcare professionals' experiences of a symptom checker app for young infants: A mixed methods service evaluation. J Fam Child Health 2025;2(9):398-405. [doi: 10.12968/jfch.2025.2.9.398]
54. Kruse SP, D'Souza L, Young J, Tuncer HGG. A systematic evaluation and comparison of the consistency of infant safer sleep messaging in Australia. Front Commun 2025;10. [doi: 10.3389/fcomm.2025.1527164]
55. Safer infant sleeping. Queensland Clinical Guideline No. MN22.71-V1-R27. 2022. URL: https://www.health.qld.gov.au/__data/assets/pdf_file/0025/1166353/g-safer-sleep.pdf [accessed 2026-06-18]

Abbreviations

AI: artificial intelligence
CAPPRRI: Consensus for APP Review Reporting Items
ICD-11: International Classification of Diseases, 11th Revision
MARS: Mobile App Rating Scale
mHealth: mobile health
NHS: National Health Service
NICE: National Institute for Health and Care Excellence
SIDS: sudden infant death syndrome
TECH: target user, evaluation focus, connectedness, and health domain

Edited by G Badicu; submitted 19.Mar.2026; peer-reviewed by A Pease, J Young; comments to author 14.May.2026; accepted 04.Jun.2026; published 10.Jul.2026.

Please cite as:
Yousef S, Meaney N, Dowding D, Gasteiger N
Safer Sleep Guidance in Standalone UK Smartphone Apps for Parents of Newborns and Infants: Systematic mHealth App Review
JMIR Mhealth Uhealth 2026;14:e95642
URL: https://mhealth.jmir.org/2026/1/e95642
doi:10.2196/95642
PMID:

©Senna Yousef, Niamh Meaney, Dawn Dowding, Norina Gasteiger. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Improving Pre-Exposure Prophylaxis Adherence in People at Risk for HIV: Secondary Analysis of a Digital Health Intervention to Enhance User Engagement

Lisa Hightow-Weidman[1], MPH, MD; Seul-Ki Choi[2], PhD; Ibrahim Yigit[1], PhD; Crissi Rainer[1], MPH; Victoria Phillips[3], PhD; Kathryn Muessig[1], PhD

[1]Institute on Digital Health and Innovation, College of Nursing, Florida State University, Innovation Park, Research Building B, 2010 Levy Avenue, RM B3400, Tallahassee, FL, United States

[2]Department of Family and Community Health, College of Nursing, University of Pennsylvania, Philadelphia, PA, United States

[3]Department of Health Policy and Management, Rollins School of Public Health, Emory University, Atlanta, GA, United States

Corresponding Author:
Lisa Hightow-Weidman, MPH, MD
Institute on Digital Health and Innovation, College of Nursing, Florida State University, Innovation Park, Research Building B, 2010 Levy Avenue, RM B3400, Tallahassee, FL, United States

Abstract

Background: Although highly effective HIV pre-exposure prophylaxis (PrEP) is available, its usage and adherence among young men who have sex with men and young transgender women remain low, reducing its overall effectiveness. The study included a 3-arm randomized clinical trial of Prepared, Protected, emPowered (P3), a comprehensive PrEP adherence digital health intervention, compared to an enhanced version, P3+, which incorporates in-app adherence coaching.

Objective: This study aims to analyze data from study participants in the P3/P3+ intervention arms to understand how different levels of user engagement with the app's features were associated with adherence to PrEP as well as the costs of each intervention and their relative cost-effectiveness.

Methods: Descriptive statistics for study variables at baseline were calculated. To examine the differences in intervention engagement and acceptability by arm, independent samples 2-tailed *t* tests for continuous variables and a chi-square analysis for categorical variables were conducted. To examine the effect of arm and engagement categories on PrEP adherence at 3 months, three logistic regression analyses were conducted: (1) the effect of arm on PrEP adherence, (2) the effect of predefined engagement categories (high vs moderate and low) on PrEP adherence, and (3) the interaction effect of arm and predefined engagement categories on PrEP adherence, along with the main effects of arm and predefined engagement categories. The study team calculated the average cost per participant and the incremental cost-effectiveness for PrEP adherence and engagement measures.

Results: A total of 163 participants were randomized to the P3 intervention (82 to the P3 arm and 81 to the P3+ arm). Participants in the P3+ arm earned higher incentives (US $90.6 vs $75.4; *P*=.04), had more app log-ins (96.6 vs 76.1; *P*=.01), used the app on more days (63.3 d vs 53.2 d; *P*=.04), and spent more time in the intervention (378.8 min vs 186.66 min; *P*<.001) compared to those in the P3 arm. There was no significant association between intervention arm and PrEP adherence at 3 months (*P*=.99). Engagement category (high vs moderate or low) was significantly associated with PrEP adherence at 3 months (*P*=.003). The overall average total monthly cost of P3 was US $1118 (SD $305.1). Average total monthly cost per P3 participant was $280 (SD $118.5), with an additional cost for P3+ of $72.

Conclusions: This study highlights the critical role of user engagement in enhancing PrEP adherence among young individuals at high risk for HIV. While the P3+ intervention led to increased engagement, this did not translate into significantly better adherence compared to the standard P3 arm. This, coupled with the increased cost and complexity of P3+ delivery, indicates that further studies are necessary to determine whether this intensified intervention is the appropriate fit.

Trial Registration: ClinicalTrials.gov NCT03320512; https://clinicaltrials.gov/study/NCT03320512

International Registered Report Identifier (IRRID): RR2-10.2196/10448

(JMIR Mhealth Uhealth 2026;14:e66837) doi:10.2196/66837

KEYWORDS

pre-exposure prophylaxis; PrEP; digital health intervention; DHI; engagement; gamification; applications; user engagement; users; chi-square analysis; adherence; engagement; coaching; costs; HIV; young adults; young individuals; youth; prophylaxis; optimizing; optimization; apps

XSL•FO
RenderX

Introduction

The introduction of pre-exposure prophylaxis (PrEP) has been a groundbreaking advancement in the drive to eliminate HIV [1]. Despite its efficacy, the overall impact of PrEP is heavily influenced by adherence levels among users [2]. High adherence to PrEP is crucial for its effectiveness in preventing HIV infection, yet PrEP adherence over time is poor, particularly among youth [3,4]. To reap the potential benefits, interventions for young men who have sex with men and young transgender women who are disproportionately impacted by HIV in the United States that can enhance and maintain PrEP adherence over time are needed [5,6].

In recent years, digital health interventions (DHIs) have emerged as promising tools in promoting health behaviors, including medication adherence [7]. These interventions offer innovative ways to engage users, provide information, and support adherence through interactive and personalized digital platforms. Engagement in DHIs has been shown to play a critical role in their efficacy in supporting behavior change across a variety of health domains [8-11]. Research has consistently demonstrated that increased user interaction with digital platforms is associated with improved outcomes, particularly in the context of HIV prevention and care [12-15].

The mechanisms through which engagement enhances efficacy are multifaceted. Active participation in DHIs can lead to increased knowledge, stronger motivation, and better self-management skills among users [16-18]. These interventions often use reminders, motivational messages, and educational content, which help sustain users' commitment to their health regimens [16,17]. Promotion of engagement through incentives and gamification has also shown promise, though studies documenting their ability to maintain engagement over time are limited [19-22].

The Adolescent Medicine Trials Network for HIV Interventions (ATN) conducted a study using the Prepared, Protected, emPowered (P3) app, a comprehensive PrEP adherence DHI, and its enhanced version, P3+, which incorporates in-app adherence coaching [23]. The P3 app included multiple features aimed at increasing user engagement, including a social wall, daily quests, and medication tracking, as well as both in-app "virtual rewards" in the form of unlocking video series produced specifically for the study and real-world rewards in the form of monetary incentives awarded based on principles of behavioral economics [24-26]. This study provided a rich dataset to explore how digital engagement correlates with PrEP adherence among a young demographic at high risk for HIV. This paper seeks to analyze data from a subset of the P3 study participants—those who were part of the intervention arms using P3 and P3+. The focus is on understanding how different levels of user engagement with the app's features were associated with adherence to PrEP as well as the costs. This analysis is crucial, as it contributes to optimizing digital health strategies to support sustained PrEP use.

Methods

Parent Study

P3 is a comprehensive PrEP adherence app which incorporates a variety of content in multiple formats including articles, videos, quizzes, and a social wall in which participants can share experiences, from success stories to challenges. P3 core components included: social prompts to foster community among participants; medication tracking and adherence support; brain builders, which included goal setting and knowledge building; and the knowledge center. P3+ is an extension of P3, in which participants receive all components in P3 and the addition of in-app adherence coaching using an adapted curriculum based on Next Step Counseling adherence counseling [27]. The study included a 3-arm randomized clinical trial conducted from March 2019 to September 2021 and followed each participant for 6 months. Participants were recruited through the ATN from 9 participant recruitment venues in Bronx, New York; Chicago, Illinois; Atlanta, Georgia; Houston, Texas; Boston, Massachusetts; Philadelphia, Pennsylvania; Tampa, Florida; Charlotte, North Carolina; and Chapel Hill, North Carolina. Recruitment methods included in-person, venue-based (including recruitment from local PrEP clinics or providers) and web-based recruitment on social media platforms. Eligible participants were (1) aged 16 to 24 years; (2) assigned male sex at birth; (3) reported sex with or intentions to have sex with men or transgender women; (4) had reliable daily access to an Android or iOS smartphone with a data plan; (5) able to speak and read English; (6) self-reported HIV-negative; and (7) were either not currently on PrEP but planned to initiate in the next 7 days and had an active PrEP prescription (prescription confirmed by study staff) or currently on PrEP and had an active PrEP prescription (prescription confirmed by study staff). Additional information about the P3 study has been previously published [23]. The original study had a baseline survey with 3- and 6-month follow-up. However, this secondary analysis only included those participants randomized to the 2 intervention arms (n=163) and data from baseline (demographic characteristics), 3-month follow-up (PrEP adherence and acceptability), and app data (engagement).

Measures

PrEP Adherence

The study team developed a composite PrEP adherence score to balance data accuracy and completeness across participants by combining biological and self-reported measures: (1) objective estimates of drug levels from dried blood spots, specifically tenofovir-diphosphate (TFV-DP) and (2) self-reported PrEP adherence in the past month. The composite score first used the participants' TFV-DP levels in the blood consistent with more than 4 pills per week considered as adherent (0="nonadherent" and 1="adherent"). For those without TFV-DP measurements (ie, missing laboratory data point), self-reported PrEP adherence from the 3-month survey was used, as prior studies have demonstrated a strong correlation between self-report and biomarker-based adherence measures [28,29]. Self-reported PrEP adherence in the past month was assessed with a single question: "In the last month, what percent

XSL•FO
RenderX

of the time did you take your PrEP as prescribed [once a day]?" with response options ranging from 0% (none of the time) to 100% (all of the time). Following previous research, participants without a TFV-DP measurement but with self-reported adherence were considered adherent if they reported taking PrEP 60% or more of the time in the past month [2]. Among participants with both TFV-DP and self-reported adherence data (n=97), the 2 measures demonstrated moderate agreement (Cohen κ=0.405), with no significant difference detected by the McNemar test (test statistic=2.25 *P*=.14), supporting the reliability of using both measures to assess adherence.

Although in-app daily pill-taking data were collected, these data were not used as the primary adherence measure due to inconsistencies in app usage and reporting, which varied widely across participants. In contrast, TFV-DP provides an objective biomarker of cumulative adherence over several weeks, and self-report offers a retrospective summary measure when biomarker data are unavailable. This dual approach allowed us to capture a more complete picture of adherence while reducing potential bias from missing or irregular app data. While the study team recognizes the value of real-time app data, the priority was to develop an adherence outcome that maximized both validity and data availability across the sample.

In summary, participants were classified as adherent to PrEP if their TFV-DP levels indicated usage of more than 4 pills per week, and as nonadherent if levels indicated fewer than 4 pills per week. For participants without TFV-DP measurements, self-reported survey data were used to minimize missingness. Those who reported taking PrEP at least 60% of the time were categorized as adherent, while those reporting less than 60% adherence were considered nonadherent.

Engagement

Engagement in P3 was defined as using at least 1 of the 3 key app features each day: medication tracker, social wall, or daily quest. Financial incentives, based on behavioral economic principles, were awarded [24,25]. Participants who interacted with any of the 3 features received US $0.50 deposited in their virtual bank account each day; those who did not interact had US $1.00 deducted. Virtual accounts were seeded with US $90 at the start of the study. Thus, participants who used the app every day for the entire 90-day study could amass US $135 at study completion; those who did not use the app at all would end the study with US $0. Three classifications of use were defined a priori based on financial incentives awarded at the end of the intervention: high use (>US $90 correlating to usage on >60 d), moderate use (US $67-$89.50 correlating to usage on 45 - 59 d), and low use (<US $67 correlating to usage on <44 d). These values were chosen to align with prior studies of app engagement and literature regarding how long it takes to form a habit [16,30]. Additional paradata were collected, including total app log-ins, total days logged in, total time spent in the app (min), and total number of days participants tracked their PrEP use (taken or not taken). For those in the P3+ arm, the total number of in-app counseling sessions was also recorded.

Acceptability

Participants were asked acceptability-related questions at the 3 months. Three questions (eg, "I trust the information in P3," "The information in P3 is easy to understand," and "The information in P3 is accurate") were adapted from Horvath et al [31] to assess end users' perceptions of the quality of the information contained in the P3 app. These questions were assessed using a 4-point Likert scale, ranging from "strongly disagree" to "strongly agree," where a higher score indicated better information quality. Then a dichotomized measure was created for each question, where responses of "agree" and "strongly agree" were interpreted as indicating high acceptability.

An adapted version (eg, questions altered slightly to align with mobile app intervention) of the Client Satisfaction Questionnaire-8 which included 8 items (eg, "How would you rate the quality of the help you have received from P3?" "Did you get the kind of help and/or support you wanted from P3?" "To what extent has P3 met your needs related to improving taking your meds?") rated on 4-point Likert scale was used to assess satisfaction with the app [32]. The scale scores were computed by summing the responses to the 8 items (range 8 - 32) with a higher score indicating a higher level of satisfaction (α=.89).

Two questions assessed participants' likelihood of future use and willingness to pay to use the app. The first item, "I would use P3 in the future if available" was rated on a 4-point Likert scale ("strongly disagree" to "strongly agree"), with higher scores indicating higher likelihood of use. Then, a dichotomized measure was created, where responses of "agree" and "strongly agree" were interpreted as an intention to use P3 in the future if it becomes available. The second item, "If P3 were available in the app store, how much would you be willing to pay for it?" Response options included "US $0.99," "US $1.99," "US $2.99," "more than $2.99," and "I would not be willing to pay for it." Responses were dichotomized into willing and unwilling to pay.

Statistical Analysis

Engagement Analysis

Descriptive statistics for study variables at baseline were calculated. To examine the differences in intervention engagement and acceptability by arm, independent samples 2-tailed *t* tests for continuous variables and a chi-square analysis for categorical variables were conducted. To examine the effect of arm and engagement categories on PrEP adherence at 3 months, three logistic regression analyses were conducted: (1) the effect of arm on PrEP adherence, (2) the effect of engagement categories (high vs moderate or low) on PrEP adherence, and (3) the interaction effect of arm and engagement categories on PrEP adherence, along with the main effects of arm and engagement categories. We also included a single demographic variable in each of the 3 models to adjust for potential confounding effects. Due to imbalanced data, dichotomized engagement categories were used instead of 3 engagement categories. Finally, adherence data, including TFV-DP levels and self-reported adherence, were collected at

XSL•FO
RenderX

a single time point during the 3-month follow-up visit, and thus the data are cross-sectional in nature. As such, analyses were conducted using cross-sectional methods, and any inferences about adherence patterns are based on a snapshot of behavior at that specific time.

Costing Analysis

The study team adopted a provider perspective and estimated the costs of P3/P3+ implementation in a community setting. This included the use of activity-based costing and, based on staff input, created an inventory of tasks associated with the management and implementation of P3 and P3+ [33]. The study team developed a survey in Qualtrics whereby each staff person at all sites reported time spent per week on the itemized task for a 2-week period. The survey was administered 4 times: October 2019, December 2019, February 2020, and November 2020 (post-COVID).

Cost per task was calculated by multiplying the median wage for the job category of the respondent staff member by the time spent on the activity. Costs per activity across staff were summed; extrapolated a monthly cost per activity based on the 2-week survey values and then averaged the monthly values by site and for all sites. Although each site provided enrollees for P3+, costs for P3+ were based on data from the North Carolina site only. Costs for P3 were based on all sites. The number of participants per month was estimated and based on the number enrolled in a given month, along with those participating in the intervention from the prior 2 months. Then, the average cost per enrollee per month was calculated. Costs for P3 app development were not included. Activities and materials specifically associated with grant requirements, such as laboratory tests, were not included in the calculations. The incremental cost-effectiveness ratio (ICER) was also calculated with respect to PrEP adherence and engagement in terms of number of app sessions, frequency, and time of app use (intensity), where the ICER is calculated as $(Cost_{P3+} - Cost_{P3})/(Outcome_{P3+} - Outcome_{P3})$ [34].

Ethical Considerations

The ATN P3 protocol, informed assent or consent documents, and related modifications were reviewed and approved by the University of North Carolina at Chapel Hill Institutional Review Board (IRB number 17 - 1951). Informed consent was obtained from all participants, and no study activities were completed prior to giving informed consent. A waiver of parental consent to participate in this research study for youth participants aged 16 to 17 years was granted by the University of North Carolina at Chapel Hill Institutional Review Board. All data and records were stored on password-protected servers. This study is registered at ClinicalTrials.gov (NCT03320512). Participants were incentivized to use the P3 app and started with a baseline amount of US $ in a virtual bank account (US $90) and gained (+US $0.50) or lost (–US $1.00) money from this initial amount each day corresponding to use and nonuse of the app, respectively. The final amount in the virtual bank account (total possible ranged from US $0.00, representing no app use, to US $135, representing perfect daily app use) was paid out as part of the study remuneration at the 3-month study follow-up. Participants were compensated with US $50 upon completion of their enrollment visit, and US $50 each at the 3- and 6-month visits (US $150 total). Participants also received US $50 for each remote specimen collection. All study data were deidentified prior to analysis.

Results

Study Participant Demographic Characteristics

A total of 163 participants were randomized to the P3 intervention (n=82, 50% to the P3 arm; n=81, 50% to the P3+ arm). Retention was 84% (n=69) and 89% (n=72) at 3 months in the P3 and P3+ arms, respectively. The mean age of participants was 21.4 (SD 2.0) years, 91.4% (n=149) identified as male, 52% (n=85) identified as non-White, and 27% (n=44) of the sample identified as Hispanic. Of the total sample, 150 (92%) participants were actively taking PrEP at the time of the study, 6 (3.7%) participants were preparing to initiate PrEP for the first time, and 7 (4.3%) participants had recently resumed PrEP after a prior discontinuation. There were no significant differences in these descriptive statistics at baseline between the P3 and P3+ arms (Table 1).

XSL•FO
RenderX

Table . Descriptive statistics for study variables at baseline by intervention arm (N=163).

Variables	Total (N=163)	Arm		*P* value
		P3 intervention (n=82)	P3+ intervention (n=81)	
Age (y), mean (SD)	21.4 (2.0)	21.5 (2.0)	21.4 (2.1)	.74
Ethnicity, n (%)				.17
Hispanic	44 (27)	26 (59.1)	18 (40.9)	
Non-Hispanic	119 (73)	56 (47.1)	63 (52.9)	
Race				.31
Others	85 (52.2)	46 (56.1)	39 (48.1)	
White	78 (47.8)	36 (43.9)	42 (51.9)	
Currently attending school				.45
No	59 (36.2)	32 (54.2)	27 (45.8)	
Yes	104 (63.8)	50 (48.1)	54 (51.9)	
Currently employed				.58
No	41 (25.2)	20 (48.8)	21 (51.2)	
Yes	121 (74.2)	62 (51.2)	59 (48.8)	
Gender identity				.09
Man	149 (91.4)	78 (52.3)	71 (47.7)	
Other	14 (8.6)	4 (28.6)	10 (71.4)	

Intervention Engagement and Acceptability

Overall engagement was high in P3+ arm compared to the P3 arm (Table 2). Participants in the P3+ arm earned higher incentives ($90.6 vs $75.4; *P*=.04), had more app log-ins (96.6 vs 76.1; *P*=.01), used the app on more days (63.3 d vs 53.2 d; *P*=.04), and spent more time in the intervention (378.8 min vs 186.66 min; *P*<.001) compared to those in the P3 arm. Overall acceptability was high in both study arms, and there was no significant difference between P3 and P3+ arms.

XSL•FO
RenderX

Table . Intervention engagement and acceptability at 3 months (N=163).

Variables	Total (N=163)	Arm		*P* value
		P3 intervention (n=82)	P3+ intervention (n=81)	
Engagement				
Total incentives earned (US $), mean (SD)	83.0 (46.6)	75.4 (48.4)	90.6 (43.6)	.04
Engagement category, n (%)				.15
High	95 (58.3)	42 (51.2)	53 (65.4)	
Moderate	12 (7.4)	6 (7.3)	6 (7.4)	
Low	56 (34.3)	34 (41.5)	22 (27.2)	
Number of sessions, mean (SD)	86.3 (52.6)	76.1 (51.0)	96.6 (52.5)	.01
Days used, mean (SD)	58.2 (31.1)	53.2 (32.1)	63.3 (29.3)	.04
Time spent in app (min), mean (SD)	282.1 (242.0)	186.6 (166.5)	378.8 (267.9)	<.001
Days PrEP[a] tracked, mean (SD)	65.9 (33.5)	61.7 (34.7)	70.1 (31.9)	.11
Coaching sessions attended, mean (SD)	—[b]	—	3.50 (1.6)	—
Acceptability				
Information quality, n (%)				
Trust information	132 (98.5)	67 (100)	65 (97)	.15
Information easy to understand	132 (98.5)	67 (100)	65 (97)	.15
Information is accurate	130 (97)	66 (98.5)	64 (95.5)	.31
Satisfaction (CSQ-8)[c], mean (SD)	24.7 (5.0)	24.4 (4.9)	25.0 (5.0)	.50
Future use, n (%)				
Use in future if available	97 (72.4)	50 (73.5)	47 (71.2)	.76
Willingness to pay	52 (38.5)	25 (36.8)	27 (40.3)	.67

[a]PrEP: pre-exposure prophylaxis.

[b]Not applicable.

[c]CSQ-8: Client Satisfaction Questionnaire-8.

The Effect of Intervention Arm and Engagement on PrEP Adherence

Using the composite endpoint which included both self-report and TFV-DP levels, adherence was 78.6% (n=55) in the P3 arm and 78.7% (n=59) in the P3+ arm. As shown in Table 3, there was no significant association between intervention arm and PrEP adherence at 3 months (*P*=.989). Engagement category (high vs moderate or low) was significantly associated with PrEP adherence at 3 months (*P*=.003); however, the interaction effect between intervention status and engagement categories on PrEP adherence was not significant (*P*=.577). We tested models including a single demographic variable to adjust for potential confounding, but the results remained unchanged (Table S1 in Multimedia Appendix 1).

XSL•FO
RenderX

Table . Relationship between intervention arm and engagement categories on pre-exposure prophylaxis (PrEP) adherence at 3 months.

	PrEP adherence[a]		
	Estimate	SE	*P* value
Intervention arm			
P3+ intervention vs P3 intervention	0.01	0.40	.99
Engagement category			
High vs moderate/low	1.27	0.42	.003
Interaction between arm and engagement category			
P3+ intervention vs P3 intervention	−0.38	0.59	.52
High vs moderate/low	1.04	0.60	.08
P3+ intervention vs P3 intervention and high vs moderate/low	0.47	0.84	.58

[a]Composite outcome.

Intervention Cost

Table S2 in Multimedia Appendix 1 shows the average monthly costs per activity. Intervention implementation and management costs were, on average, US $617 (SD $270.9) per month across sites with staff meetings accounting for over half of the monthly total. Average monthly costs associated with participant interactions were US $502 (SD $78.6). Participant recruitment, at US $204 (SD $100.0), was the primary cost driver, accounting for 41% of the total, and the largest reported expense across sites. Troubleshooting was the second with Chapel Hill, the central study site, reporting relatively higher expense than other sites.

Taken together, the overall average total monthly cost of P3 was US $1118 (SD $305.1) and ranged from US $797 to $1368. Sites averaged 4.5 (SD 3.9) participants per month resulting in an average total monthly cost per participant of US $280 (SD $118.5). Table S3 in Multimedia Appendix 1 shows the additional costs associated with the P3+ intervention. Average monthly costs for ongoing management of P3+ were US $655 (SD $263.0), while those for counseling-related issues were US $565 (SD $45.0), dominated by counseling itself. An estimated 2.3 participants received counseling per month at an average cost of US $72 (2.7).

Given the difference in P3+ and P3 adherence was small and not significant, the ICER was not calculated. The ICER for the number of app sessions was US $7.06 ([US $352-US $280]/[US $86.3 - US $76.1]) per additional session used, while the ICER for time in session was US $45 per additional hour of use.

Discussion

The findings from this analysis underscore the critical role that user engagement plays in the effectiveness of DHIs aimed at enhancing adherence to PrEP among young individuals at high risk for HIV. A key finding from this study is the significant association between higher levels of engagement with the P3 intervention and increased adherence to PrEP. This aligns with prior research, which has consistently demonstrated that user engagement is a pivotal factor in the success of DHIs across various health domains, including HIV care and prevention [13,14]. Despite the high levels of engagement observed in this study, it is challenging to draw direct comparisons with other DHIs due to the lack of standardized metrics for measuring engagement. A recent scoping review focused on engagement in DHIs for sexual and gender minorities noted that only a small fraction of trials reported engagement metrics, and even fewer studies defined an “optimal” level of engagement necessary to achieve significant health outcomes [35].

The P3 app employed several strategies to boost user engagement, including financial incentives and gamification [19,23]. These methods are supported by a growing body of evidence that suggests such approaches can effectively maintain user interaction over extended periods. For example, Looyestyn et al [20] conducted a systematic review that concluded gamification often increases engagement with online health programs, although the long-term effects on health outcomes remain uncertain. However, enhancing engagement strategies through more robust and personalized approaches has the potential to further improve adherence rates [36]. Moreover, conducting longer-term studies could offer valuable insights into the lasting impact of these strategies, thereby enabling the refinement and adaptation of interventions for sustained long-term outcomes [37].

Interestingly, this study observed that while the P3+ intervention, which included in-app adherence coaching, led to increased engagement (eg, more time spent in the app and higher incentive earnings), this did not translate into significantly better adherence compared to the standard P3 arm. This, coupled with the increased cost and complexity of P3+ delivery, indicates that further studies are necessary to determine whether this intensified intervention is the appropriate fit. This finding may indicate a potential ceiling effect where additional engagement beyond a certain threshold does not necessarily result in proportional improvements in adherence. Using sequential multiple assignment randomized trials could be particularly useful in this context, as they would allow for the systematic evaluation and optimization of the intervention’s components

XSL•FO
RenderX

(and human resources) to better tailor the approach to individual needs [35]. Our results here show that increasing frequency of use was of relatively low cost, while increasing intensity of use cost US $45 an hour. As these are interim outcomes, we cannot declare whether they are cost-effective as we have no standard against which to measure the results as the literature in this area is extremely limited. Although the P3+ compared ot P3 did not increase adherence, it may have produced other health benefits that were not measured as part of the trial.

Given the challenges of sustaining engagement in standalone digital health interventions, future research may explore whether adherence support features could be effectively delivered through widely used social and dating platforms. These platforms attract substantial daily engagement from populations at high risk for HIV, suggesting potential reach. However, our study did not collect data on participants' platform preferences, usage patterns, or privacy concerns, and thus any discussion of integrating interventions into these environments remains hypothetical. Importantly, any such approach would require additional research to assess user acceptability, data security considerations, and potential risks, as well as partnerships with platform developers. If found feasible and acceptable, embedding adherence support within high-traffic commercial apps could reduce the costs of sustaining standalone interventions and may leverage existing infrastructure for personalization and scalability.

This study has several limitations that must be acknowledged. Collection of biologic outcome data was impacted by the COVID-19 pandemic, potentially affecting the accuracy of adherence measurements and the robustness of the findings. The cost analysis presented is based on data collected over a limited period, with potential disruptions also caused by the COVID-19 pandemic. Additionally, due to the small number of participants who were initiating PrEP for the first time, the study team was unable to conduct meaningful comparisons between novice and experienced PrEP users. This limited the ability to explore whether patterns of adherence or engagement differed by prior PrEP experience. Future studies with a more balanced distribution of new and experienced users are needed to examine these potential differences. Moreover, this study was unable to adjust for behavioral confounders such as sexual activity frequency, PrEP experience, and partner status, as well as demographic confounders that are known to influence the relationship between engagement and adherence. This may result in an estimated association that reflects the influence of confounders rather than the true effect of the exposure. Due to the limited sample size, we were unable to explicitly examine how these confounders affect engagement and adherence, or how they interact with one another. Future research with larger samples is warranted to enable comprehensive adjustment for these factors and to better understand their combined impact.

Furthermore, variability in site-specific factors, such as participant recruitment and staff composition, likely contributed to the wide range of cost estimates. While staff were asked to report time for non–study-related tasks, separating the two may have been challenging, particularly in relation to staff meetings. In addition, some tasks conducted in a study setting, such as recruitment, may not be a good proxy for real-world costs and are likely an overestimate. A notable limitation of this cost analysis is the exclusion of initial app development expenses, which are often among the most substantial costs for digital health interventions. This decision was made to provide a clearer estimate of the ongoing operational and maintenance costs associated with scaling the P3 and P3+ interventions in community settings, as these are the costs most likely to be absorbed by public health agencies or clinical partners. Including development costs would significantly increase the apparent total cost, potentially skewing the analysis away from the more immediate, recurring expenses that directly impact long-term sustainability. Moreover, app development costs can vary widely based on factors such as platform selection (eg, proprietary vs open-source), feature complexity, integration with existing health care systems, and ongoing user feedback. For instance, the P3 app incorporated multiple interactive elements, including social walls, daily quests, and personalized adherence coaching, each of which adds to the initial investment. Given this variability, including these costs would complicate direct cost comparisons across different digital health interventions. Nonetheless, future studies should consider a more comprehensive economic evaluation that includes both fixed and variable costs, as this will be critical for assessing the true scalability and long-term viability of such digital health solutions in real-world settings. Finally, this study provided financial incentives for app engagement which may not be sustainable outside of research contexts.

In conclusion, this study highlights the critical role of user engagement in enhancing PrEP adherence among young individuals at high risk for HIV. While increased engagement through DHIs like the P3 app correlates with higher adherence rates, the findings suggest that simply intensifying interventions, such as with the P3+ model, may not necessarily lead to proportionate improvements. The additional costs and complexities associated with more intensive approaches require careful consideration and further investigation. Future research should prioritize the development of more personalized and adaptive engagement strategies, leveraging methodologies such as sequential multiple assignment randomized trials designs to optimize intervention components in real-time. By focusing on long-term adherence and the sustainability of engagement, these efforts can better inform the design of scalable, cost-effective digital health solutions that are both effective and responsive to the evolving needs of at-risk populations.

Acknowledgments

The authors would like to collectively acknowledge and thank the Prepared, Protected, emPowered (P3) study team, P3 study participants, iTech, and the ATN for their continued support and contributions to this project.

XSL•FO
RenderX

Funding

This work was supported by the National Institutes of Health (U19HD089881). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability

The datasets generated and analyzed during this study are provided in the NICHD DASH (National Institute of Child Health and Human Development Data and Specimen Hub) repository and will be available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: LH-W; data collection: CR; data analysis: SKC, IY, KM, VP, and LH-W; manuscript preparation: LH-W, SKC, IY, VP, and KM; revision and approval of the final manuscript: all authors. All authors have agreed on the journal to which the article will be submitted, gave final approval of the version to be published, and agreed to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Relationship between intervention arm and engagement categories on PrEP adherence at 3 months, adjusted for demographics.
[DOCX File, 37 KB - mhealth_v14i1e66837_app1.docx]

References

1. Grant RM, Lama JR, Anderson PL, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med 2010 Dec 30;363(27):2587-2599. [doi: 10.1056/NEJMoa1011205] [Medline: 21091279]
2. Anderson PL, Glidden DV, Liu A, et al. Emtricitabine-tenofovir concentrations and pre-exposure prophylaxis efficacy in men who have sex with men. Sci Transl Med 2012 Sep 12;4(151):151ra125. [doi: 10.1126/scitranslmed.3004006] [Medline: 22972843]
3. Hosek SG, Landovitz RJ, Kapogiannis B, et al. Safety and feasibility of antiretroviral preexposure prophylaxis for adolescent men who have sex with men aged 15 to 17 years in the United States. JAMA Pediatr 2017 Nov 1;171(11):1063-1071. [doi: 10.1001/jamapediatrics.2017.2007] [Medline: 28873128]
4. Hosek S, Celum C, Wilson CM, Kapogiannis B, Delany‐Moretlwe S, Bekker L. Preventing HIV among adolescents with oral PrEP: observations and challenges in the United States and South Africa. J Int AIDS Soc 2016 Oct;19(7S6). [doi: 10.7448/IAS.19.7.21107] [Medline: 27760684]
5. Cambou MC, Landovitz RJ. Challenges and opportunities for preexposure prophylaxis. Top Antivir Med 2021;29(4):399-406. [Medline: 34856093]
6. Ezennia O, Geter A, Smith DK. The PrEP care continuum and Black men who have sex with men: a scoping review of published data on awareness, uptake, adherence, and retention in PrEP care. AIDS Behav 2019 Oct;23(10):2654-2673. [doi: 10.1007/s10461-019-02641-2] [Medline: 31463711]
7. Sun L, Qu M, Chen B, Li C, Fan H, Zhao Y. Effectiveness of mHealth on adherence to antiretroviral therapy in patients living with HIV: meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2023 Jan 23;11:e42799. [doi: 10.2196/42799] [Medline: 36689267]
8. Bauermeister J, Choi SK, Bruehlman-Senecal E, et al. An identity-affirming web application to help sexual and gender minority youth cope with minority stress: pilot randomized controlled trial. J Med Internet Res 2022 Aug 1;24(8):e39094. [doi: 10.2196/39094] [Medline: 35916700]
9. Forbes A, Keleher MR, Venditto M, DiBiasi F. Assessing patient adherence to and engagement with digital interventions for depression in clinical trials: systematic literature review. J Med Internet Res 2023 Aug 11;25:e43727. [doi: 10.2196/43727] [Medline: 37566447]
10. Regmi K, Kassim N, Ahmad N, Tuah NA. Effectiveness of mobile apps for smoking cessation: a review. Tob Prev Cessat 2017;3(12):12. [doi: 10.18332/tpc/70088] [Medline: 32432186]
11. Couper MP, Alexander GL, Zhang N, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010 Nov 18;12(4):e52. [doi: 10.2196/jmir.1430] [Medline: 21087922]
12. Choi SK, Golinkoff J, Michna M, Connochie D, Bauermeister J. Correlates of engagement within an online HIV prevention intervention for single young men who have sex with men: randomized controlled trial. JMIR Public Health Surveill 2022 Jun 27;8(6):e33867. [doi: 10.2196/33867] [Medline: 35759333]
13. Hightow-Weidman LB, LeGrand S, Muessig KE, et al. A randomized trial of an online risk reduction intervention for young Black MSM. AIDS Behav 2019 May;23(5):1166-1177. [doi: 10.1007/s10461-018-2289-9] [Medline: 30269231]

XSL•FO
RenderX

14. Hightow-Weidman L, Muessig KE, Egger JR, Vecchio A, Platt A. Epic allies: a gamified mobile app to improve engagement in HIV care and antiretroviral adherence among young men who have sex with men. AIDS Behav 2021 Aug;25(8):2599-2617. [doi: 10.1007/s10461-021-03222-y] [Medline: 33740213]
15. Fan X, Ning K, Liu C, et al. Uptake of an app-based case management service for HIV-positive men who have sex with men in China: process evaluation study. J Med Internet Res 2023 Apr 26;25:e40176. [doi: 10.2196/40176] [Medline: 37099367]
16. Hightow-Weidman LB, Bauermeister JA. Engagement in mHealth behavioral interventions for HIV prevention and care: making sense of the metrics. Mhealth 2020;6(7):7. [doi: 10.21037/mhealth.2019.10.01] [Medline: 32190618]
17. Hightow-Weidman LB, Horvath KJ, Scott H, Hill-Rorie J, Bauermeister JA. Engaging youth in mHealth: what works and how can we be sure? Mhealth 2021;7(23):23. [doi: 10.21037/mhealth-20-48] [Medline: 33898592]
18. Hightow-Weidman L, Muessig K, Knudtson K, et al. A gamified smartphone app to support engagement in care and medication adherence for HIV-positive young men who have sex with men (AllyQuest): development and pilot study. JMIR Public Health Surveill 2018 Apr 30;4(2):e34. [doi: 10.2196/publichealth.8923] [Medline: 29712626]
19. Tran S, Smith L, El-Den S, Carter S. The use of gamification and incentives in mobile health apps to improve medication adherence: scoping review. JMIR Mhealth Uhealth 2022 Feb 21;10(2):e30671. [doi: 10.2196/30671] [Medline: 35188475]
20. Looyestyn J, Kernot J, Boshoff K, Ryan J, Edney S, Maher C. Does gamification increase engagement with online programs? A systematic review. PLoS ONE 2017;12(3):e0173403. [doi: 10.1371/journal.pone.0173403] [Medline: 28362821]
21. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Jul;71:31-48. [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]
22. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013 Dec 12;1(1):e3. [doi: 10.2196/games.3139] [Medline: 25658754]
23. LeGrand S, Knudtson K, Benkeser D, et al. Testing the efficacy of a social networking gamification app to improve pre-exposure prophylaxis adherence (P3: Prepared, Protected, emPowered): protocol for a randomized controlled trial. JMIR Res Protoc 2018 Dec 18;7(12):e10448. [doi: 10.2196/10448] [Medline: 30563818]
24. Kimmel SE, Troxel AB. Novel incentive-based approaches to adherence. Clin Trials 2012 Dec;9(6):689-695. [doi: 10.1177/1740774512453520] [Medline: 22848071]
25. Haff N, Patel MS, Lim R, et al. The role of behavioral economic incentive design and demographic characteristics in financial incentive-based approaches to changing health behaviors: a meta-analysis. Am J Health Promot 2015;29(5):314-323. [doi: 10.4278/ajhp.140714-LIT-333] [Medline: 25928816]
26. Vlaev I, King D, Darzi A, Dolan P. Changing health behaviors using financial incentives: a review from behavioral economics. BMC Public Health 2019 Aug 7;19(1):1059. [doi: 10.1186/s12889-019-7407-8] [Medline: 31391010]
27. Amico KR, Miller J, Balthazar C, et al. Integrated next step counseling (iNSC) for sexual health and PrEP use among young men who have sex with men: implementation and observations from ATN110/113. AIDS Behav 2019 Jul;23(7):1812-1823. [doi: 10.1007/s10461-018-2291-2] [Medline: 30315429]
28. Blumenthal J, Pasipanodya EC, Jain S, et al. Comparing self-report pre-exposure prophylaxis adherence questions to pharmacologic measures of recent and cumulative pre-exposure prophylaxis exposure. Front Pharmacol 2019;10(721):721. [doi: 10.3389/fphar.2019.00721] [Medline: 31333454]
29. Zimmerman RS, Mehrotra P, Madden T, Paul R. The value of assessing self-reported and biological indicators of outcomes in evaluating HIV programs. Curr HIV/AIDS Rep 2021 Aug;18(4):365-376. [doi: 10.1007/s11904-021-00560-3] [Medline: 33993397]
30. Lally P, van Jaarsveld CHM, Potts HWW, Wardle J. How are habits formed: Modelling habit formation in the real world. Euro J Social Psych 2010 Oct;40(6):998-1009. [doi: 10.1002/ejsp.674]
31. Horvath KJ, Oakes JM, Rosser BRS, et al. Feasibility, acceptability and preliminary efficacy of an online peer-to-peer social support ART adherence intervention. AIDS Behav 2013 Jul;17(6):2031-2044. [doi: 10.1007/s10461-013-0469-1] [Medline: 23553347]
32. Kelly PJ, Kyngdon F, Ingram I, Deane FP, Baker AL, Osborne BA. The Client Satisfaction Questionnaire-8: Psychometric properties in a cross-sectional survey of people attending residential substance abuse treatment. Drug Alcohol Rev 2018 Jan;37(1):79-86. [doi: 10.1111/dar.12522] [Medline: 28480521]
33. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the Economic Evaluation of Health Care Programmes: Oxford University Press; 2002.
34. Brandão SMG, Brunner-La Rocca HP, de Lima ACP, Bocchi EA. A review of cost-effectiveness analysis: from theory to clinical practice. Medicine (Baltimore) 2023 Oct 20;102(42):e35614. [doi: 10.1097/MD.0000000000035614] [Medline: 37861539]
35. Choi SK, Muessig KE, Hightow-Weidman LB, Bauermeister JA. Paradata: measuring engagement in digital HIV interventions for sexual and gender minorities. Curr HIV/AIDS Rep 2023 Dec;20(6):487-501. [doi: 10.1007/s11904-023-00679-5] [Medline: 37930613]
36. Saleem M, Kühne L, De Santis KK, Christianson L, Brand T, Busse H. Understanding engagement strategies in digital interventions for mental health promotion: scoping review. JMIR Ment Health 2021 Dec 20;8(12):e30000. [doi: 10.2196/30000] [Medline: 34931995]

XSL•FO
RenderX

37. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]

Abbreviations

ATN: Adolescent Medicine Trials Network for HIV Interventions
DHI: digital health intervention
ICER: incremental cost-effectiveness ratio
P3: Prepared, Protected, emPowered
PrEP: pre-exposure prophylaxis
TFV-DP: tenofovir-diphosphate

Edited by L Buis; submitted 24.Sep.2024; peer-reviewed by HJ Gilmore, J Liu; revised version received 14.Nov.2025; accepted 29.Jan.2026; published 10.Mar.2026.

Please cite as:
Hightow-Weidman L, Choi SK, Yigit I, Rainer C, Phillips V, Muessig K
Improving Pre-Exposure Prophylaxis Adherence in People at Risk for HIV: Secondary Analysis of a Digital Health Intervention to Enhance User Engagement
JMIR Mhealth Uhealth 2026;14:e66837
URL: https://mhealth.jmir.org/2026/1/e66837
doi:10.2196/66837

© Lisa Hightow-Weidman, Seul-Ki Choi, Ibrahim Yigit, Crissi Rainer, Victoria Phillips, Kathryn Muessig. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evidence of Efficacy of the My Personal Health Guide Mobile Phone App on Antiretroviral Therapy Adherence Among Young African American Men Who Have Sex With Men at 1 Month: Randomized Controlled Trial

Mark S Dworkin[1], MD, MPH&TM; Kara Herrera[1], MPH, PhD; Sierra Upton[1], MS, MPH; Casey M Luc[1], MPH, PhD; Jeb Jones[2], PhD; Paul Burns[3], PhD; Li Liu[1], PhD; Antonio Jimenez[1], PhD; Ruiqi Ren[1], MS; Meaghan Woody[2], MPH; Robert Garofalo[4], MD, MPH; Sangyoon Lee[5], PhD

[1]Community Outreach Intervention Projects, University of Illinois Chicago, Chicago, IL, United States

[2]Rollins School of Public Health, Emory University, Atlanta, United States

[3]John D. Bower School of Population Health, University of Mississippi Medical Center, Jackson, MS, United States

[4]Adolescent and Young Adult Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, United States

[5]Department of Computer Science, Connecticut College, New London, CT, United States

Corresponding Author:
Kara Herrera, MPH, PhD
Community Outreach Intervention Projects
University of Illinois Chicago
1603 West Taylor Street
Chicago, IL, 60607
United States
Phone: 1 (312) 996 2028
Email: knitti2@uic.edu

Abstract

Background: Young African American men who have sex with men (AAMSM) experience disproportionately high HIV incidence and are less likely to achieve viral suppression compared to White men who have sex with men, an outcome that relies on antiretroviral therapy (ART) adherence. We created My Personal Health Guide, a talking relational agent–based mobile health app to improve ART adherence among young AAMSM.

Objective: The objective was to determine the efficacy of My Personal Health Guide on improving ART adherence among young AAMSM living with HIV.

Methods: We implemented a randomized controlled trial among young (aged 18-34 years) AAMSM with nonoptimal ART adherence throughout the United States between February 2020 and September 2023, predominantly through social media and by word of mouth, provider referral, and fliers in selected health care settings. Participants were randomized in a 1:1 ratio using permuted blocks of 8 to the intervention, My Personal Health Guide, or the attention control arm. ART adherence was assessed with Wilson's 3-item self-reported adherence measurement and dichotomized at ≥80%. Logistic regression models using backward selection were used to evaluate the efficacy of My Personal Health Guide on ≥80% ART adherence at 1-month follow-up.

Results: Among the 253 AAMSM at baseline, most (n=180, 71.1%) self-reported being ≥80% adherent to ART, over half (n=145, 57.3%) resided in the Southern United States, but all US regions were represented, nearly half (n=175, 42.3%) had some college education, over one-third (n=96, 37.9%) had less than optimal literacy, and approximately one-quarter (n=61, 24.1%) experienced housing insecurity in the past 6 months. The sample for analysis of the My Personal Health Guide app efficacy was 131 (intervention=76 and control=55). The odds of being ≥80% adherent to ART at 1-month follow-up were 3.97 (95% CI 1.26-12.55) times greater among participants randomized to the My Personal Health Guide app compared to the controls, after adjusting for ART adherence at baseline, treatment adherence self-efficacy, and ever being incarcerated. Additionally, for every 1-point increase in the HIV Treatment Adherence Self-Efficacy Scale, the odds of ≥80% ART adherence increased by 3% (odds ratio 1.03, 95% CI 1.00-1.06).

XSL•FO
RenderX

Conclusions: Participants randomized to receive My Personal Health Guide reported nearly 4 times greater odds of being ≥80% adherent to ART compared to the attention control group at 1-month follow-up. To our knowledge, this is the first randomized controlled trial demonstrating improved medication adherence using a relational agent–based behavioral intervention. These findings provide evidence of short-term efficacy of My Personal Health Guide to improve ART adherence among young AAMSM. We recommend further research on the inclusion of relational agents in behavioral research, especially in populations affected by stigma and nonoptimal health literacy, where this nonhuman supportive and educational approach may be complementary to health care systems.

Trial Registration: ClinicalTrials.gov NCT04217174; https://clinicaltrials.gov/study/NCT04217174

(JMIR Mhealth Uhealth 2026;14:e75005) doi:10.2196/75005

KEYWORDS

men who have sex with men; African American men who have sex with men; avatar-based mobile phone intervention; Black or African American; HIV; medical adherence; mobile apps; sexual and gender minorities; treatment adherence

Introduction

African American men who have sex with men (AAMSM) have disproportionate HIV infection rates, retention in care [1], mortality, and rates of viral nonsuppression [2-8], an outcome that relies on antiretroviral therapy (ART) adherence. Additionally, health literacy is a factor associated with ART adherence [9] that is lower among African American people than White people [10]. Therefore, interventions that bolster health literacy might improve healthy HIV-related behaviors like ART adherence, potentially resulting in personal and public health benefits.

Relational agents are computerized images that a person may react to as if they are in a relationship. A relational agent can be designed to promote healthy behavioral change. We developed My Personal Health Guide, an innovative theory-driven relational agent–based app with improvement of ART adherence as one of its objectives [11]. During iterative app development focus groups with young AAMSM living with HIV, participants reported universal acceptability for the African American realistic avatar we developed that spoke about, taught, explained, and encouraged behavior to promote healthy HIV-related outcomes [11]. Responding to meeting their adherence-related needs and interests, the app includes clear and simple language, an audiovisual format, and gamification. A pilot study performed in Chicago provided evidence that My Personal Health Guide has the potential to improve adherence [12]. Here, we present preliminary short-term follow-up results from a randomized controlled trial (RCT) of My Personal Health Guide that demonstrate evidence of improvement of ART adherence among young AAMSM living with HIV. We hypothesized that those randomized to the My Personal Health Guide app would have greater ART adherence compared to controls at 1-month follow-up post randomization. Evidence of short-term improvement in ART adherence provides scientific rationale for further research of this realistic avatar and relational agent approach as an adherence intervention and informs efforts to refine the approach to produce longer-term effects.

Methods

Intervention

My Personal Health Guide features a realistic talking avatar with supportive functions that were designed to inform users of HIV-related health information, encourage participants to take their medication, and facilitate adherence-related behavioral skills with reminders and push notifications to take medication and promotion of using weekly pill boxes. It also includes recordings of peers and health care providers offering personal and motivational messages. The app provides a private space, in one's own mobile phone, for hearing and seeing information relevant to healthy living with HIV. Users may replay information, which can be especially helpful to people who are uncomfortable asking questions of their health care provider, including asking for a repeat of what was said or an explanation in simple terms.

Design, Recruitment, Eligibility, and Enrollment

This RCT was performed among young AAMSM living with HIV. Participants were recruited throughout the United States between February 2020 and September 2023, predominantly through social media (eg, the dating app Jack'd and Facebook advertisements) and also by word of mouth, provider referral, and fliers in health care settings of the 3 enrollment sites: University of Illinois at Chicago, Emory University, and University of Mississippi Medical Center. Those who were eligible were assigned male sex at birth, currently identified as a man, African American or Black, aged 18-34 years, had at least one male sexual partner in their lifetime, English speaking, owned a smartphone, were initiating or prescribed oral ART, and had a history of nonoptimal ART adherence defined as having a detectable viral load within the past 4 weeks, self-reported nonoptimal ART adherence within the past 30 days, or were referred by a provider for adherence concerns. Those who were on injectable ART, participating in another HIV treatment study, or did not have a stable mailing address at the time of enrollment were ineligible for participation. This study was determined to be minimal risk, with no adverse events reported.

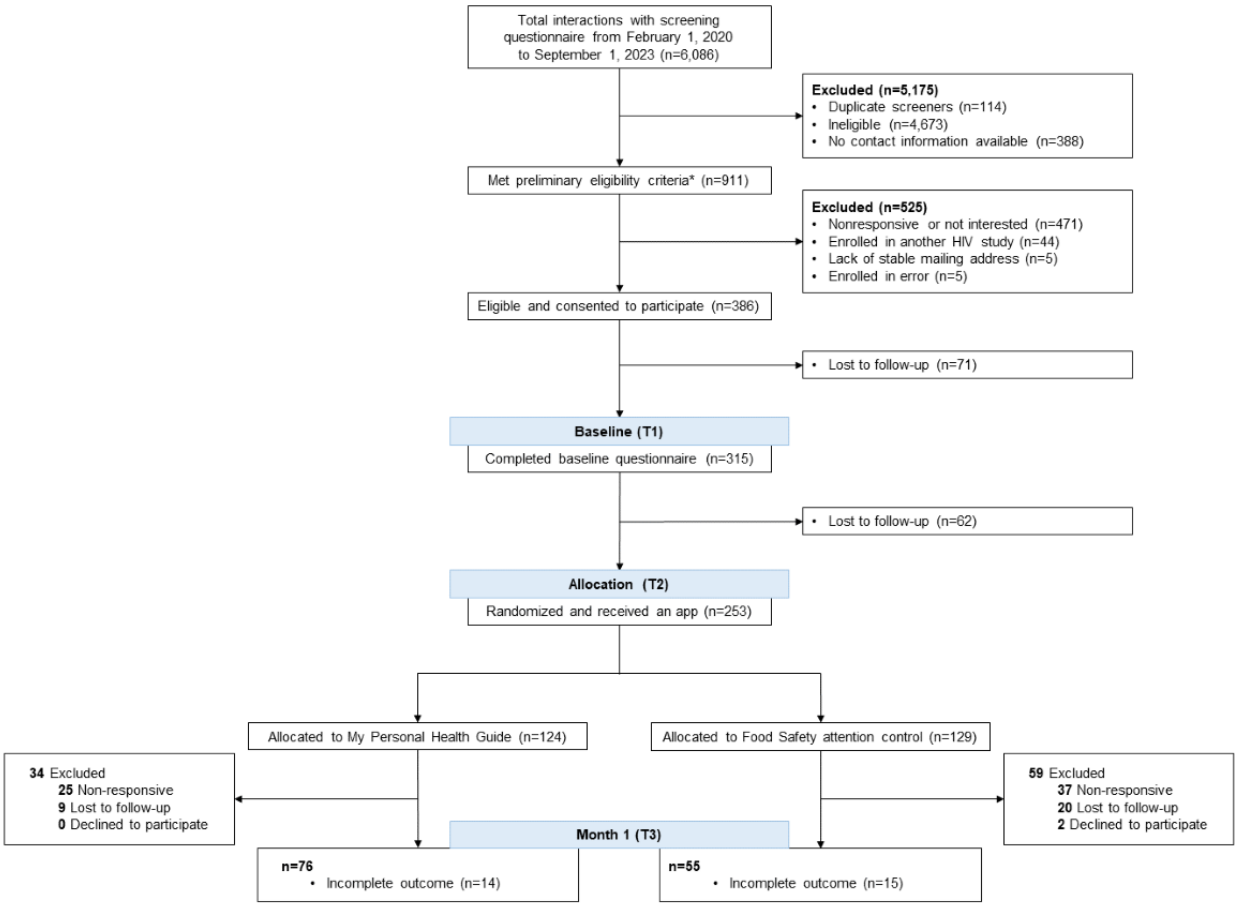
Data Collection, Sample Size, and Randomization

Figure 1 summarizes participant interactions with the study through the first month of follow-up. A structured,

XSL•FO
RenderX

interviewer-administered baseline questionnaire included collection of demographics, baseline knowledge of HIV (items taught in the app), self-reported adherence, incarceration history, housing stability, substance use history, self-efficacy (HIV Treatment Adherence Self-Efficacy Scale [HIV-ASES]) [13], literacy (adapted Rapid Estimate of Adult Literacy in Medicine-Short Form [REALM-SF]) [14], perceived social support (Multidimensional Scale of Perceived Social Support [MSPSS]) [15], and depression symptomatology (Patient Health Questionnaire for Depression [PHQ-9]) [16]. If a participant exhibited moderately severe or severe depressive symptomatology, they were encouraged to reach out to their provider for mental health resources. One month after the baseline interview, participants were scheduled for the app orientation and download that included explaining how to download, setting up a password, briefly pointing out features, and encouraging app use. Additionally, a check-in call was made monthly during 6 months of follow-up that asked troubleshooting questions about the app and collected self-reported ART adherence. Participants received compensation of US $40 for the baseline interview, US $75 for an appointment to download the app, and US $10 for responding to each monthly check-in call. Sample size calculations were carried out using Repeated Measures and Sample Size (RMASS2), the power calculation software for repeated measures design that allows attrition over time. The study was designed with a proposed sample size of 250 at the end of the study duration, or 295 at baseline, which would provide 80% statistical power to detect a 10% group difference in HIV care retention at the end of the study, considering a 15% attrition rate. Permuted randomization in blocks of 8 was used to allocate research participants in a 1:1 ratio to the intervention or control arm, and allocation concealment from the principal investigator, research staff, and research participants was carried out by the senior statistician. Participants randomized to the intervention arm received the My Personal Health Guide app, and those randomized to the attention control arm received a food safety app, which featured an avatar and functions related to food and nutrition, narrated illustrated stories about food safety, an educational game about sugar content in food, and information promoting oral health. The principal investigator and research participants were blinded after assignment to the intervention. Research coordinators oriented the participants to their respective app after randomization.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram. Preliminary eligibility did not include verifying age and active HIV prescription.



Outcome Measurement

ART adherence was assessed with Wilson's 3-item self-reported adherence measurement, which has been validated against electronic drug monitoring for HIV medication (Cronbach α=0.83) [17]. These three questions include (1) "In the last 30 days, on how many days did you miss at least one dose of any of your HIV medication?" (with a response range from 0 to 30 days); (2) "In the last 30 days, how often did you take your HIV medication in the way you were supposed to?" (with response

XSL•FO
RenderX

options including Never, Rarely, Sometimes, Usually, Almost Always, and Always); and (3) "In the last 30 days, how good a job did you do at taking your HIV medications in the way you were supposed to?" (with response options including Very Poor, Poor, Fair, Good, Very Good, and Excellent). Answers were averaged together and converted to a percentage score on a 100-point scale. Consistent with other studies [18,19], 80% was defined as the cut point for nonoptimal ART adherence. HIV viral load was the intended primary outcome for this research. However, the trial was disturbed by the COVID-19 pandemic, and significant difficulty occurred collecting viral load from this population. Therefore, self-reported ART adherence was used.

Statistical Analysis

The primary statistical goal for this analysis was to determine the efficacy of the My Personal Health Guide app on ART adherence at 1-month follow-up. Participant demographic and study characteristics are presented as means and SDs for continuous variables and as absolute numbers (n) and proportions (%) for categorical variables stratified by randomization arm. The success of randomization was assessed using 2-tailed *t* tests for continuous baseline characteristics or chi-square tests for categorical characteristics. Baseline characteristics between individuals that completed both the baseline meeting and 1-month follow-up versus those who had missing data at 1-month follow-up were compared to ensure comparability between the samples. Bivariate comparisons were made between participant characteristics and ART adherence at baseline (≥80% vs <80%) using chi-square tests for categorical data, Cochran-Armitage tests for ordinal data, and independent *t* tests for continuous data. Logistic regression models with the backward model selection method were used to evaluate the effect of the My Personal Health Guide app, adjusted for significant factors that were associated with ART adherence at month 1, with a $P \leq .05$ considered statistically significant. Baseline adherence was adjusted for a priori. The baseline characteristics that were considered in the logistic regression models included age, ethnicity, remote participation, residence, education, employment status, sexual orientation, currently in a committed relationship, health literacy (REALM-SF), HIV-ASES, depression symptomatology (PHQ-9), perceived social support (MSPSS), ever incarcerated, housing insecurity, and substance use seriousness (serious includes cocaine, methamphetamines, crack, heroin, and opiates; less serious includes alcohol, marijuana, ecstasy, speedballs, and gamma-hydroxybutyrate; none includes no use of substances or alcohol) at baseline. Adjusted model results are presented as odds ratios (ORs) and 95% CI.

Ethical Considerations

This research was reviewed and approved by the institutional review boards of the University of Illinois Chicago, Emory University, and the University of Mississippi Medical Center (protocol #2019-1184). The research involving human data was conducted in accordance with institutional guidelines of the University of Illinois Chicago, Emory University, and the University of Mississippi Medical Center. All participants were administered written informed consent. All data were deidentified prior to analysis and dissemination. Participants were compensated up to US $125 by the 1-month follow-up by their preferred method of check, cash, or e-code.

Results

Sample Characteristics

We attempted to contact the 911 men that met preliminary eligibility criteria, responded to recruitment methods, and provided contact information. Among those, 471 were either not responsive or not interested in the study, 44 were enrolled in another HIV treatment study, and 5 did not have a stable mailing address in order to receive study materials. Five participants were erroneously enrolled in the study due to either being unable to verify having a current HIV prescription (n=3) or having been previously enrolled in the study (n=2). A total of 386 men were consented. Among these, the baseline questionnaire was completed by 315 (81.6%), and 253 (65.5%) had the study apps installed on their mobile phones 1 month later (Figure 1), including 124 (49%) randomized to My Personal Health Guide and 129 (51%) randomized to the Food Safety attention control app. After 1 month, 93 participants did not respond to contact attempts to complete the month 1 check-in call and were excluded from the current analysis. The remaining 160 participants completed the first check-in phone call, and of those, 131 had complete self-reported adherence data at month 1.

The mean age of the 253 participants at baseline was 29.4 (SD 3.7; range18-35) years; most (n=221, 87.4%) were 25-34 years of age. Few (n=18, 7%) identified as Hispanic. Most participated remotely (n=202, 79.8%). Most participants were enrolled by the University of Illinois at Chicago (n=131, 51.8%) and Emory University (n=106, 41.9%), and 16 (6%) were enrolled by the University of Mississippi Medical Center. Over half (n=145, 57.3%) of participants resided in the Southern United States, but all US regions were represented in the sample. Nearly half (n=175, 42.3%) had some college education, and over half (n=146, 57.7%) reported being employed. Three-quarters (n=180, 71.1%) identified as homosexual or gay and were not currently in a committed relationship with a male partner (n=198, 78.3%). Over one-third (n=96, 38%) had less than optimal literacy (adapted REALM-SF). Over half reported mild-severe depressive symptomatology, with 26% (n=65) reporting mild (PHQ-9=5.0-9.9), 14% (n=35) reporting moderate (PHQ-9=10.0-14.9), 9% (n=23) reporting moderately severe (PHQ-9=15.0-19.9), and 4% (n=11) reporting severe (PHQ-9>20.0) depressive symptomatology. The median HIV-ASES score was 104 (IQR 89-114). Although most (n=150, 59.3%) reported high perceived social support (MSPSS>5.0), 9% (n=23) had low (MSPSS<3.0) perceived social support, and 32% (n=80) had moderate (MSPSS=3.0-5.0) perceived social support. Approximately one-quarter experienced housing insecurity in the past 6 months (n=61, 24%), and nearly one-third were ever incarcerated (n=77, 30%). Some (n=39, 15%) reported serious substance use in the past 2 weeks. Nearly three-quarters (n=178, 70.4%) reported less serious substance use, including alcohol, marijuana, ecstasy, speedballs, or gamma-hydroxybutyrate. Most (n=180, 71.1%) self-reported

XSL•FO
RenderX

being ≥80% adherent to ART at baseline. Group comparisons for all baseline characteristics and ART adherence at baseline (Table 1) revealed that randomizations were successful.

XSL•FO
RenderX

Table 1. Baseline demographic characteristics and antiretroviral therapy (ART) adherence of young African American men who have sex with men living with HIV stratified by randomization arm (n=253), 2020-2024.

	Overall (n=253)	Intervention (n=124)	Control (n=129)	*P* value
Age in years, mean (SD)	29.4 (3.7)	29.6 (3.5)	29.1 (4)	.46[a]
Age group (years), n (%)				.41[b]
18-24	32 (12.6)	13 (10.5)	19 (14.7)	
25-34	221 (87.4)	111 (89.5)	110 (85.3)	
Ethnicity, n (%)				>.99[b]
Hispanic	18 (7.1)	9 (7.3)	9 (7)	
Not Hispanic	235 (92.9)	115 (92.7)	120 (93)	
Enrollment site, n (%)				.18[b]
University of Illinois at Chicago	131 (51.8)	59 (47.6)	72 (55.8)	
Emory University	106 (41.9)	54 (43.5)	52 (40.3)	
University of Mississippi Medical Center	16 (6.3)	11 (8.9)	5 (3.9)	
Remote participation, n (%)				.87[b]
Yes	202 (79.8)	98 (79)	104 (80.6)	
No	51 (20.2)	26 (21)	25 (19.4)	
Residence, n (%)				.42[c]
Midwest	77 (30.4)	34 (27.4)	43 (33.3)	
Northeast	27 (10.7)	12 (9.7)	15 (11.6)	
South	145 (57.3)	77 (62.1)	68 (52.7)	
West	4 (1.6)	1 (0.8)	3 (2.3)	
Education, n (%)				.59[b]
College	68 (26.9)	30 (24.2)	38 (29.5)	
Some college	107 (42.3)	51 (41.1)	56 (43.4)	
High school or GED[d]	65 (25.7)	36 (29)	29 (22.5)	
Less than high school	13 (5.1)	7 (5.6)	6 (4.7)	
Employment, n (%)				.45[b]
Active duty	1 (0.4)	0 (0)	1 (0.8)	
Employed	146 (57.7)	75 (60.5)	71 (55)	
Student	18 (7.1)	10 (8.1)	8 (6.2)	
Unable to work or unemployed	83 (32.8)	36 (29)	47 (36.4)	
Missing	5 (2)	3 (2.4)	2 (1.6)	
Sexual orientation, n (%)				.81[b]
Homosexual or gay	180 (71.1)	85 (68.5)	95 (73.6)	
Bisexual	49 (19.4)	27 (21.8)	22 (17.1)	
Heterosexual or straight	4 (1.6)	2 (1.6)	2 (1.6)	
Other	20 (7.9)	10 (8.1)	10 (7.8)	
Currently in a committed relationship with a male partner, n (%)				.66[b]
Yes	55 (21.7)	25 (20.2)	30 (23.3)	
No	198 (78.3)	99 (79.8)	99 (76.7)	

XSL•FO
RenderX

	Overall (n=253)	Intervention (n=124)	Control (n=129)	*P* value
Marital status, n (%)				.69[b]
Legally married	10 (4)	4 (3.2)	6 (4.7)	
Registered domestic partnership	1 (0.4)	1 (0.8)	0 (0)	
Widowed	1 (0.4)	0 (0)	1 (0.8)	
Divorced	8 (3.2)	5 (4)	3 (2.3)	
Separated	5 (2)	2 (1.6)	3 (2.3)	
Never married	228 (90.1)	112 (90.3)	116 (89.9)	
Health literacy (adapted REALM-SF[e]), n (%)				>.99[b]
Nonoptimal (at least 1 wrong)	96 (37.9)	47 (37.9)	49 (38)	
Optimal (none wrong)	157 (62.1)	77 (62.1)	80 (62)	
PHQ-9[f] score, mean (SD)	6.67 (5.81)	6.71 (5.96)	6.62 (5.69)	.90[a]
PHQ-9 score, n (%)				.72[g]
None or minimal (<5.0)	117 (46.2)	55 (44.4)	62 (48.1)	
Mild (5.0-9.9)	65 (25.7)	35 (28.2)	30 (23.3)	
Moderate (10.0-14.9)	35 (13.8)	14 (11.3)	21 (16.3)	
Moderately severe (15.0-19.9)	23 (9.1)	11 (8.9)	12 (9.3)	
Severe (≥20.0)	11 (4.3)	7 (5.6)	4 (3.1)	
Missing	2 (0.8)	2 (1.6)	0 (0)	
HIV-ASES[h], n (%)				.29[a]
Mean (SD)	98.1 (20.2)	99.5 (18.1)	96.8 (22)	
Median (IQR)	104 (89.0-114)	105 (88.3-114)	103 (89.0-115)	
Missing, n (%)	3 (1.2)	2 (1.6)	1 (0.8)	
MSPSS[i] score, mean (SD)	5.08 (1.36)	5.09 (1.45)	5.08 (1.28)	.98[a]
MSPSS score, n (%)				
Low (<3.0)	23 (9.1)	13 (10.5)	10 (7.8)	.68[b]
Moderate (3.0-5.0)	80 (31.6)	37 (29.8)	43 (33.3)	.81[d]
High (≥5.1)	150 (59.3)	74 (59.7)	76 (58.9)	
Ever incarcerated, n (%)				.07[b]
Yes	77 (30.4)	45 (36.3)	32 (24.8)	
No	176 (69.6)	79 (63.7)	97 (75.2)	
Housing insecurity in previous 6 months, n (%)				.64[b]
Yes	61 (24.1)	32 (25.8)	29 (22.5)	
No	192 (75.9)	92 (74.2)	100 (77.5)	
Substance use seriousness in previous 2 weeks[j], n (%)				.32[b]
None	36 (14.2)	17 (13.7)	19 (14.7)	
Less serious	178 (70.4)	92 (74.2)	86 (66.7)	
Serious	39 (15.4)	15 (12.1)	24 (18.6)	
Adherence to ART (baseline), n (%)				.14[b]
Yes (≥80)	180 (71.1)	94 (75.8)	86 (66.7)	
No (<80)	73 (28.9)	30 (24.2)	43 (33.3)	

XSL•FO
RenderX

	Overall (n=253)	Intervention (n=124)	Control (n=129)	*P* value
Adherence to ART (month 1), n (%)				.04[b]
Yes (≥80)	111 (43.9)	69 (55.6)	42 (32.6)	
No (<80)	20 (7.9)	7 (5.6)	13 (10.1)	
Missing data (month 1), n (%)				.005[b]
Yes	122 (48.2)	48 (38.7)	74 (57.4)	
No	131 (51.8)	76 (61.3)	55 (42.6)	

[a]Independent samples *t* test.

[b]Chi-square test for comparison.

[c]Fisher exact test.

[d]GED: General Educational Development.

[e]REALM-SF: Rapid Estimate of Adult Literacy in Medicine-Short Form.

[f]PHQ-9: Patient Health Questionnaire for Depression.

[g]Cochran-Armitage test.

[h]HIV-ASES: HIV Treatment Adherence Self-Efficacy Scale.

[i]MSPSS: Multidimensional Scale of Perceived Social Support.

[j]Serious includes cocaine, methamphetamines, crack, heroin, opiates; less serious includes alcohol, marijuana, ecstasy, speedballs, gamma-hydroxybutyrate.

Primary Outcome

Among the 253 participants who were randomized and downloaded an app, 131 (51.8%) completed the 1-month follow-up for ART adherence, and 122 (48.2%) did not. A significantly greater proportion of those randomized to the Food Safety app were missing 1-month follow-up data compared to the My Personal Health Guide app (74/129, 57.4% vs 48/124, 38.7%; *P*=.005; Table 2). Additionally, remote participation, employment, substance use, and ART adherence at baseline differed by complete 1-month follow-up data and were therefore considered during multivariable model construction. Bivariate analysis (Table 3) revealed the mean HIV-ASES score (mean 103, SD 15.8 vs mean 86.4, SD 28.5; *P*=.02) was significantly higher in those who had ≥80% adherence to ART at 1 month as compared to those who had <80% adherence at 1 month. The mean depression symptomatology scale (PHQ-9) was moderately significantly lower among those who were ≥80% adherent to ART at 1 month as compared to those who were <80% adherent to ART at 1 month (5.70, SD 4.96 vs 9.00, SD 7.43; *P*=.07). There was also a significantly higher prevalence of being in a committed relationship with a male partner among those who had ≥80% adherence to ART at 1 month as compared to those who had <80% adherence at 1 month (25/111, 22.5% vs 0/20, 0%; *P*=.01).

In multivariable analysis (Table 4), the odds of being ≥80% adherent to ART at 1-month follow-up were almost 4 times greater (OR 3.97, 95% CI 1.26-12.55) among participants randomized to the My Personal Health Guide app as compared to the Food Safety app after adjusting for other characteristics that were also associated with ART adherence at 1 month, including HIV-ASES and having a history of ever being incarcerated. Every 1-point increase in the HIV-ASES increased the odds of ≥80% ART adherence by 3% (OR 1.03, 95% CI 1.00-1.06). A history of ever being incarcerated reduced the odds of ≥80% ART adherence (OR 0.31, 95% CI 0.09-1.06). ART adherence at baseline was marginally associated with ≥80% ART at 1-month follow-up (OR 2.97, 95% CI 0.88-9.99).

XSL•FO
RenderX

Table 2. Baseline demographic characteristics and antiretroviral therapy (ART) adherence of young African American men who have sex with men living with HIV stratified by missing outcome (self-reported ART adherence at 1 month) data (n=253), 2020-2024.

	Overall (n=253)	Missing outcome data (month 1; n=122)	Complete outcome data (month 1; n=131)	*P* value
Age in years, mean (SD)	29.4 (3.7)	29.2 (3.6)	29.5 (3.6)	.51[a]
Age group (years), n (%)				.98[b]
18-24	32 (12.6)	16 (13.1)	16 (12.2)	
25-34	221 (87.4)	106 (86.9)	115 (87.8)	
Ethnicity, n (%)				.37[b]
Hispanic	18 (7.1)	11 (9)	7 (5.3)	
Not Hispanic	235 (92.9)	111 (91)	124 (94.7)	
Enrollment site, n (%)				.28[b]
University of Illinois at Chicago	131 (51.8)	66 (54.1)	65 (49.6)	
Emory University	106 (41.9)	46 (37.7)	60 (45.8)	
University of Mississippi Medical Center	16 (6.3)	10 (8.2)	6 (4.6)	
Remote participation, n (%)				<.001[b]
Yes	202 (79.8)	83 (68)	119 (90.8)	
No	51 (20.2)	39 (32)	12 (9.2)	
Residence, n (%)				
Midwest	77 (30.4)	44 (36.1)	33 (25.2)	.16[b]
Northeast	27 (10.7)	15 (12.3)	12 (9.2)	.14[c]
South	145 (57.3)	61 (50)	84 (64.1)	
West	4 (1.6)	2 (1.6)	2 (1.5)	
Education, n (%)				.29[b]
College	68 (26.9)	26 (21.3)	42 (32.1)	
Some college	107 (42.3)	56 (45.9)	51 (38.9)	
High school or GED[d]	65 (25.7)	33 (27)	32 (24.4)	
Less than high school	13 (5.1)	7 (5.7)	6 (4.6)	
Employment, n (%)				.005
Active duty	1 (0.4)	0 (0)	1 (0.8)	
Employed	146 (57.7)	58 (47.5)	88 (67.2)	
Student	18 (7.1)	12 (9.8)	6 (4.6)	
Unable to work or unemployed	83 (32.8)	49 (40.2)	34 (26)	
Missing	5 (2)	3 (2.5)	2 (1.5)	
Sexual orientation, n (%)				.87[b]
Homosexual or gay	180 (71.1)	84 (68.9)	96 (73.3)	
Bisexual	49 (19.4)	25 (20.5)	24 (18.3)	
Heterosexual or straight	4 (1.6)	2 (1.6)	2 (1.5)	
Other	20 (7.9)	11 (9)	9 (6.9)	
Currently in a committed relationship with a male partner, n (%)				.36[b]
Yes	55 (21.7)	30 (24.6)	25 (19.1)	
No	198 (78.3)	92 (75.4)	106 (80.9)	
Marital status, n (%)				

XSL•FO
RenderX

	Overall (n=253)	Missing outcome data (month 1; n=122)	Complete outcome data (month 1; n=131)	*P* value
Legally married	10 (4)	4 (3.3)	6 (4.6)	.71[b]
Registered domestic partnership	1 (0.4)	0 (0)	1 (0.8)	.83[c]
Widowed	1 (0.4)	0 (0)	1 (0.8)	
Divorced	8 (3.2)	5 (4.1)	3 (2.3)	
Separated	5 (2)	2 (1.6)	3 (2.3)	
Never married	228 (90.1)	111 (91)	117 (89.3)	
Health literacy (adapted REALM-SF[e]), n (%)				.41[b]
Nonoptimal (at least 1 wrong)	96 (37.9)	50 (41)	46 (25.1)	
Optimal (None wrong)	157 (62.1)	72 (59)	85 (64.9)	
PHQ-9[f] score, mean (SD)	6.67 (5.81)	7.16 (6.10)	6.21 (5.51)	.20[a]
PHQ-9 score, n (%)				.78[b]
None or minimal (<5.0)	117 (46.2)	55 (45.1)	62 (47.3)	
Mild (5.0-9.9)	65 (25.7)	29 (23.8)	36 (27.5)	
Moderate (10.0-14.9)	35 (13.8)	18 (14.8)	17 (13)	
Moderately severe (15.0-19.9)	23 (9.1)	12 (9.8)	11 (8.4)	
Severe (≥20.0)	11 (4.3)	7 (5.7)	4 (3.1)	
Missing	2 (0.8)	1 (0.8)	1 (0.8)	
HIV-ASES[g] score				.11[a]
Mean (SD)	98.1 (20.2)	96 (21.1)	100 (19.2)	
Median (IQR)	104 (89-114)	105 (90-115)	104 (89-114)	
Missing, n (%)	3 (1.2)	0 (0)	3 (2.3)	
MSPSS[h] score, mean (SD)	5.08 (1.36)	4.96 (1.39)	5.19 (1.33)	.18[a]
MSPSS score, n (%)				
Low (<3.0)	23 (9.1)	11 (9)	12 (9.2)	.12[b]
Moderate (3.0-5.0)	80 (31.6)	46 (37.7)	34 (26)	.17[i]
High (≥5.1)	150 (59.3)	65 (53.3)	85 (64.9)	
Ever incarcerated, n (%)				.14[b]
Yes	77 (30.4)	43 (35.2)	34 (26)	
No	176 (69.6)	79 (64.8)	97 (74)	
Housing insecurity in previous 6 months, n (%)				.57[b]
Yes	61 (24.1)	27 (22.1)	34 (26)	
No	192 (75.9)	95 (77.9)	97 (74)	
Substance use seriousness in previous 2 weeks[j] , n (%)				.02[b]
None	36 (14.2)	25 (20.5)	11 (8.4)	
Less serious	178 (70.4)	77 (63.1)	101 (77.1)	
Serious	39 (15.4)	20 (16.4)	19 (14.5)	
Adherence to ART (baseline), n (%)				.05[b]
Yes (≥80%)	180 (71.1)	79 (64.8)	101 (77.1)	

XSL•FO
RenderX

	Overall (n=253)	Missing outcome data (month 1; n=122)	Complete outcome data (month 1; n=131)	*P* value
No (<80%)	73 (28.9)	43 (35.2)	30 (22.9)	

[a]Independent samples *t* test.

[b]Chi-square test for comparison.

[c]Fisher exact test.

[d]GED: General Educational Development.

[e]REALM-SF: Rapid Estimate of Adult Literacy in Medicine-Short Form.

[f]PHQ-9: Patient Health Questionnaire for Depression.

[g]HIV-ASES: HIV Treatment Adherence Self-Efficacy Scale.

[h]MSPSS: Multidimensional Scale of Perceived Social Support.

[i]Cochran-Armitage test.

[j]Serious includes cocaine, methamphetamines, crack, heroin, opiates; less serious includes alcohol, marijuana, ecstasy, speedballs, gamma-hydroxybutyrate.

XSL•FO
RenderX

Table 3. Bivariate analysis of demographic characteristics of young African American men who have sex with men living with HIV by antiretroviral therapy (ART) adherence at 1 month (n=131), 2020-2024.

	Overall (n=131)	<80% ART adherence (n=20)	≥80 ART adherence (n=111)	*P* value
Randomization, n (%)				.03[a]
My Personal Health Guide (intervention)	76 (58)	7 (35)	69 (62.2)	
Food safety (control)	55 (42)	13 (65)	42 (37.8)	
Age (years), mean (SD)	29.5 (3.62)	29.9 (2.60)	29.4 (3.78)	.53[b]
Age group (years), n (%)				.13[a]
25-34	115 (87.8)	20 (100)	95 (85.6)	
18-24	16 (12.2)	0 (0)	16 (14.4)	
Ethnicity, n (%)				.59[a]
Hispanic	7 (5.3)	0 (0)	7 (6.3)	
Not Hispanic	124 (94.7)	20 (100.0)	104 (93.7)	
Remote participation, n (%)				.39[a]
Yes	119 (90.8)	17 (85.0)	102 (91.9)	
No	12 (9.2)	3 (15.0)	9 (8.1)	
Residence, n (%)				.63[a]
Midwest	33 (25.2)	7 (35.0)	26 (23.4)	
Northeast	12 (9.2)	2 (10.0)	10 (9.0)	
South	84 (64.1)	11 (55.0)	73 (65.8)	
West	2 (1.5)	0 (0.0)	2 (1.8)	
Education, n (%)				.75[a]
College	68 (26.9)	18 (26.5)	50 (73.5)	
Some college	107 (42.3)	30 (28.0)	77 (72.0)	
High school or GED[c]	65 (25.7)	22 (33.8)	43 (66.2)	
Less than high school	13 (5.1)	3 (23.1)	10 (76.9)	
Employment, n (%)				.27[a]
Employed, student, active duty	95 (72.5)	12 (60.0)	83 (74.8)	
Unemployed or unable to work	34 (26.0)	7 (35.0)	27 (24.3)	
Missing	2 (1.5)	1 (5)	1 (0.9)	
Sexual orientation, n (%)				.44[a]
Homosexual or gay	96 (73.3)	18 (90)	78 (70.3)	
Bisexual	24 (18.3)	2 (10)	22 (19.9)	
Heterosexual or straight	2 (1.5)	0 (0)	2 (1.9)	
Other	9 (6.9)	0 (0)	9 (8.1)	
Currently in a committed relationship with a male partner, n (%)				.01[a]
Yes	25 (19.1)	0 (0)	25 (22.5)	
No	106 (80.9)	20 (100)	86 (77.5)	
Health literacy (adapted REALM-SF[d]), n (%)				>.99[a]
Nonoptimal (at least 1 wrong)	46 (35.1)	7 (35)	39 (35.1)	
Optimal (none wrong)	85 (64.9)	13 (65)	72 (64.9)	

XSL•FO
RenderX

	Overall (n=131)	<80% ART adherence (n=20)	≥80 ART adherence (n=111)	*P* value
PHQ-9[e] score, mean (SD)	6.21 (5.51)	9 (7.43)	5.70 (4.96)	.07[b]
PHQ-9 (missing), n (%)	1 (0.8)	0 (0)	1 (0.9)	.07[b]
HIV-ASES[f] score, mean (SD)	100 (19.2)	86.4 (28.5)	103 (15.8)	.02[b]
HIV-ASES (missing), n (%)	3 (2.3)	0 (0)	3 (2.7)	02[b]
MSPSS[g] score, mean (SD)	5.19 (1.33)	4.81 (1.36)	5.26 (1.32)	.18[b]
Ever incarcerated				.41[a]
Yes	34 (26)	7 (35)	27 (24.3)	
No	97 (74)	13 (65)	84 (75.7)	
Housing insecurity in previous 6 months				.41[a]
Yes	34 (26)	7 (35)	27 (24.3)	
No	97 (74)	13 (65)	84 (75.7)	
Substance use seriousness in previous 2 weeks[h]				.12[a]
Serious	19 (14.5)	6 (30)	13 (11.7)	
Less serious	101 (77.1)	13 (65)	88 (79.3)	
None	11 (8.4)	1 (5)	10 (9)	

[a]Fisher exact test.

[b]Independent samples *t* test.

[c]GED: General Educational Development.

[d]REA LM-SF: Rapid Estimate of Adult Literacy in Medicine-Short Form.

[e]PHQ-9: Patient Health Questionnaire for Depression.

[f]HIV-ASES: HIV Treatment Adherence Self-Efficacy Scale.

[g]MSPSS: Multidimensional Scale of Perceived Social Support.

[h]Serious includes cocaine, methamphetamines, crack, heroin, opiates; less serious includes alcohol, marijuana, ecstasy, speedballs, gamma-hydroxybutyrate

Table 4. Multivariable logistic regression for the association between My Personal Health Guide and ≥80% antiretroviral therapy (ART) adherence at 1-month follow-up among young African American men who have sex with men living with HIV (n=131), 2020-2024.

	Unadjusted odds ratio	Adjusted odds ratio (95% CI)	*P* value
Randomization			.02
My Personal Health Guide (intervention)	4.55	3.97 (1.26-12.55)	
Food Safety (control)	Reference	Reference	
ART adherence at baseline			.08
≥80%	3.05	2.97 (0.88-9.99)	
<80%	Reference	Reference	
HIV-ASES[a]	1.04	1.03 (1.00-1.06)	.04
Ever incarcerated			.06
Yes	0.60	0.31 (0.09-1.06)	
No	Reference	Reference	

[a]HIV-ASES: HIV Treatment Adherence Self-Efficacy Scale.

Discussion

Principal Findings

Participants randomized to receive My Personal Health Guide demonstrated nearly 4 times greater odds of being ≥80% adherent to ART compared to participants who received the Food Safety control app at 1-month follow-up. This is a substantial magnitude, and although these are short-term follow-up results, these data provide strong evidence that a

XSL•FO
RenderX

relational agent approach to improving medication adherence is promising.

Our study is the first of its kind in the field of HIV medication adherence. We are aware of only 2 other published RCTs of a relational agent for behavior change. Rubin et al [20] performed an RCT of a relational agent to deliver a brief alcohol intervention and referral to treatment to veterans within VA primary care clinics. Their approach involved 2 meetings with the relational agent, who requested a commitment to change. They reported that their relational agent elicited concern about the consequences of drinking and bolstered motivation to change. Although they did not observe accelerated decline in drinking during 3 months with their intervention compared to their control group, they noted that their intervention group was more likely to receive referrals for follow-up and that this approach allowed for more screening and brief intervention without increasing the burden on clinic staff. Another study by Prochaska et al [21] studied a smartphone intervention called Woebot for substance use disorders that used a therapeutic relational agent delivering cognitive behavioral therapy. At 8 weeks, they reported reduced substance use occasions compared to a waitlist group. These studies demonstrate the potential for a relational agent to influence behavior and underscore the scientific rationale for their application to the field of medication adherence. Relational mobile health interventions may be particularly beneficial for younger (18-24 years) AAMSM because they are more likely to still be developing medication-taking routines and may be especially receptive to interactive mobile technologies that can bolster daily ART adherence (Hightow-Weidman et al [22]). Given that these studies and ours are RCTs, they provide evidence for consideration of relational agents in the design of behavioral interventions, especially involving conditions where stigma about diagnosis or treatment could create hesitancy with face-to-face health care provider education or counseling.

In addition to our study demonstrating a large magnitude efficacy, there are several other strengths and considerations. First, our study began shortly before the COVID-19 pandemic spread throughout the United States, disrupting in-person research with lockdowns and other restrictions. This led to a nearly entirely remote rather than facility-based administration and may have made the meeting with research staff for orientation to the app less interactive than it might have been otherwise. The remote approach could have diminished the participant's feeling of accountability to the study or facilitated participation by reducing the barrier of travel to a study activity. Second, unlike medication or immunization, our app intervention involves a passive approach. In other words, patients are encouraged to use it, but it is their choice how much of the app they experience and how often they open it. The app was designed to have new functions and experiences occur with repeated use to encourage return to the app. However, we have no control to ensure regular app exposure. Despite these issues, we observed a significant effect in achieving the adherence goal. This suggests that the effect might increase further and benefit more patients with additional app refinement and involvement of health care providers or case managers in implementation who could refer the app to their patients and check in with them at future appointments. Third, our study sample was diverse, including a large proportion who experienced housing insecurity, screened for moderate or severe depression, and reported a history of incarceration. Additionally, our sample included an overrepresentation of men living in the Southern United States, a high HIV incidence region where many experience structural barriers to HIV care [23] and where a low-cost, scalable intervention like ours could contribute to achieving higher prevalence of viral suppression and help end the HIV epidemic. Finally, our study also demonstrated an independent effect of self-efficacy on being ≥80 adherent. This finding is consistent with other studies, such as those by Dworkin et al [24], Barclay et al [25], and Waldrop-Valverde et al [26]. However, this is the first study to report this specifically in a population of young AAMSM living with HIV.

Limitations

Limitations of this study included that many participants were recruited using a social media dating app and the responsiveness of participants to the 1-month call was moderate, both of which may limit generalizability. Notably, we struggled to recruit younger (18-24 years) AAMSM; therefore, we recommend future studies use a multimodal recruitment approach. Additionally, missingness of self-reported ART adherence was high, and we observed a higher proportion of those randomized to the control with missing data at month 1. Study procedures were disrupted by the COVID-19 pandemic, necessitating the change from a facility-based trial to online. While this pivot allowed for further recruitment and retention, it compromised the collection of viral load data from participants' clinics, which was initially an outcome of interest. This change also impacted the collection of dried blood spots at the 6-month follow-up, which was initially intended to be conducted on-site and changed to be self-administered in the participant's home to accommodate COVID-19 stay-at-home orders. This resulted in a high proportion of missingness. Additionally, the laboratory methodology for the threshold of detection changed during the study period and used a higher threshold (<839 copies/mL) than the clinical definition of undetectable viral load (<200 copies/mL). Similarly, the usage of Wisepill devices to monitor ART adherence was low, so self-report was used as the primary outcome of interest. Using self-reported ART adherence also had limitations. Self-report can overestimate individual medication-taking behavior and therefore may introduce recall and social desirability bias. However, self-reported adherence is often used to assess ART adherence in both routine care and research settings and has been correlated with viral load and other objective markers of adherence [26]. Last, these findings are limited to the first month of observation. Additional follow-up analysis from this trial is planned to determine evidence of long-term effect.

Conclusion

These findings provide evidence that My Personal Health Guide has short-term efficacy in improving ART adherence among young AAMSM living with HIV and contribute to the emerging field of technology-based behavioral interventions promoting health using relational agents. We recommend further research on the inclusion of relational agents in behavioral research,

XSL•FO
RenderX

especially in populations affected by stigma and nonoptimal health literacy, where this nonhuman supportive and educational approach may be complementary to health care systems.

Acknowledgments

We thank the National Institute of Mental Health for funding this research (R01MH116721). We acknowledge Allison Huske, BS; Autumn Smith, MPH; and Elise Dressel, BS, from the University of Illinois at Chicago, School of Public Health; and Paige Jones, BS, and Blake Acebo, BS, from the University of Illinois at Chicago, School of Medicine, for their contributions to data collection.

The National Institute of Mental Health had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The authors were not precluded from accessing data in the study, and they accept responsibility to submit for publication.

Data Availability

A deidentified dataset, a data dictionary defining each field in the set, the study protocol including the statistical analysis plan, and informed consent forms will be made available upon request and with publication acceptance. Per university protocol, all sensitive and high-risk data will be transmitted via an encrypted protocol, or Secure File Transfer Protocol, to individuals outside of the university with investigator support, approval from the data safety and monitoring board, and a signed data access agreement.

Authors' Contributions

MSD and SL contributed to the conceptualization. MSD, RG, and SL contributed to the funding acquisition. MSD, JJ, AJ, RG, and SL contributed to the methodology. MSD, KH, SU, JJ, and PB contributed to project administration. MSD, JJ, PB, RG, and SL contributed to the resources. MSD, JJ, and PB contributed to the supervision. KH, SU, CML, MW, and SL contributed to the data curation. KH, SU, CML, LL, and RR contributed to the formal analysis. KH, SU, and CML contributed to the investigation. KH, SU, CML, and SL contributed to the software. MSD, KH, SU, CML, JJ, LL, and RR contributed to the validation. KH, SU, and RR contributed to the visualization. MSD, KH, SU, CML, LL, and RR contributed to writing the original draft. All authors reviewed and edited the final manuscript.

Conflicts of Interest

The authors report no conflicts of interest. The authors attest that there was no use of generative artificial intelligence technology in the generation of the text, figures, or other informational content of this manuscript.

Multimedia Appendix 1
CONSORT 2010 checklist.
[PDF File (Adobe PDF File), 1410 KB - mhealth_v14i1e75005_app1.pdf]

References

1. Singh S, Bradley H, Hu X, Skarbinski J, Hall HI, Lansky A, Centers for Disease ControlPrevention (CDC). Men living with diagnosed HIV who have sex with men: progress along the continuum of HIV care--United States, 2010. MMWR Morb Mortal Wkly Rep 2014;63(38):829-833 [FREE Full text] [Medline: 25254559]
2. HIV among African Americans. Centers for Disease Control and Prevention. URL: http://www.cdc.gov/hiv/pdf/risk_HIV_AAA.pdf [accessed 2024-08-22]
3. Estimates on new HIV infections in the United States. Centers for Disease Control and Prevention. URL: http://www.cdc.gov/nchhstp/newsroom/docs/fact-sheet-on-hiv-estimates.pdf [accessed 2024-08-22]
4. HIV among Black/African American gay, bisexual, and other men who have sex with men. Centers for Disease Control and Prevention. URL: http://www.cdc.gov/hiv/risk/racialethnic/bmsm/facts/ [accessed 2024-08-22]
5. Prejean J, Song R, Hernandez A, Ziebell R, Green T, Walker F, HIV Incidence Surveillance Group. Estimated HIV incidence in the United States, 2006-2009. PLoS One 2011;6(8):e17502 [FREE Full text] [doi: 10.1371/journal.pone.0017502] [Medline: 21826193]
6. Back of the line: the state of AIDS among Black gay men in America 2012. Black AIDS Institute. 2012. URL: http://www.blackaids.org/images/reports/back.pdf [accessed 2024-08-22]
7. Diagnoses of HIV infection in the United States and dependent areas. Centers for Disease Control and Prevention. Atlanta, GA: National Center for HIV/AIDS, Viral Hepatitis, STD. and TB Prevention; 2012. URL: http://www.cdc.gov/hiv/library/reports/surveillance/ [accessed 2025-12-06]

XSL•FO
RenderX

8. Beer L, Oster A, Mattson C, Skarbinski J, Medical Monitoring Project. Disparities in HIV transmission risk among HIV-infected black and white men who have sex with men, United States, 2009. AIDS 2014;28(1):105-114 [FREE Full text] [doi: 10.1097/QAD.0000000000000021] [Medline: 23942058]
9. Drainoni M, Rajabiun S, Rumptz M, Welles SL, Relf M, Rebholz C, et al. Health literacy of HIV-positive individuals enrolled in an outreach intervention: results of a cross-site analysis. J Health Commun 2008;13(3):287-302. [doi: 10.1080/10810730801985442] [Medline: 18569359]
10. The health literacy of America's adults: results from the 2003 national assessment of adult literacy. National Center for Education Statistics. URL: https://nces.ed.gov/pubs2006/2006483.pdf [accessed 2018-02-23]
11. Dworkin M, Chakraborty A, Lee S, Monahan C, Hightow-Weidman L, Garofalo R, et al. A realistic talking human embodied agent mobile phone intervention to promote HIV medication adherence and retention in care in young HIV-positive african american men who have sex with men: qualitative study. JMIR Mhealth Uhealth 2018;6(7):e10211 [FREE Full text] [doi: 10.2196/10211] [Medline: 30064971]
12. Dworkin MS, Lee S, Chakraborty A, Monahan C, Hightow-Weidman L, Garofalo R, et al. Acceptability, feasibility, and preliminary efficacy of a theory-based relational embodied conversational agent mobile phone intervention to promote HIV medication adherence in young HIV-Positive African American MSM. AIDS Educ Prev 2019;31(1):17-37. [doi: 10.1521/aeap.2019.31.1.17] [Medline: 30742481]
13. Johnson MO, Neilands TB, Dilworth SE, Morin SF, Remien RH, Chesney MA. The role of self-efficacy in HIV treatment adherence: validation of the HIV treatment adherence self-efficacy scale (HIV-ASES). J Behav Med 2007;30(5):359-370. [doi: 10.1007/s10865-007-9118-3] [Medline: 17588200]
14. Arozullah A, Yarnold P, Bennett C, Soltysik R, Wolf M, Ferreira R, et al. Development and validation of a short-form, rapid estimate of adult literacy in medicine. Med Care 2007;45(11):1026-1033. [doi: 10.1097/MLR.0b013e3180616c1b] [Medline: 18049342]
15. Zimet GD, Powell SS, Farley GK, Werkman S, Berkoff KA. Psychometric characteristics of the Multidimensional Scale of Perceived Social Support. J Pers Assess 1990;55(3-4):610-617. [doi: 10.1080/00223891.1990.9674095] [Medline: 2280326]
16. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
17. Hightow-Weidman LB, Fowler B, Kibe J, McCoy R, Pike E, Calabria M, et al. HealthMpowerment.org: development of a theory-based HIV/STI website for young black MSM. AIDS Educ Prev 2011;23(1):1-12 [FREE Full text] [doi: 10.1521/aeap.2011.23.1.1] [Medline: 21341956]
18. Ito KE, Kalyanaraman S, Brown JD, Miller WC. Factors affecting avatar use in a STI prevention CD-ROM. J Adolesc Health 2008;42(2 (Suppl 1)):19. [doi: 10.1016/j.jadohealth.2007.11.052]
19. LeGrand S, Muessig KE, McNulty T, Soni K, Knudtson K, Lemann A, et al. Epic Allies: development of a gaming app to improve antiretroviral therapy adherence among young HIV-positive men who have sex with men. JMIR Serious Games 2016;4(1):e6 [FREE Full text] [doi: 10.2196/games.5687] [Medline: 27178752]
20. Rubin A, Livingston NA, Brady J, Hocking E, Bickmore T, Sawdy M, et al. Computerized relational agent to deliver alcohol brief intervention and referral to treatment in primary care: a randomized clinical trial. J Gen Intern Med 2022;37(1):70-77 [FREE Full text] [doi: 10.1007/s11606-021-06945-9] [Medline: 34145518]
21. Prochaska JJ, Vogel EA, Chieng A, Baiocchi M, Maglalang DD, Pajarito S, et al. A randomized controlled trial of a therapeutic relational agent for reducing substance misuse during the COVID-19 pandemic. Drug Alcohol Depend 2021;227:108986 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.108986] [Medline: 34507061]
22. Hightow-Weidman LB, LeGrand S, Muessig KE, Simmons RA, Soni K, Choi SK, et al. A randomized trial of an online risk reduction intervention for young black MSM. AIDS Behav 2019;23(5):1166-1177 [FREE Full text] [doi: 10.1007/s10461-018-2289-9] [Medline: 30269231]
23. Carter JW, Flores SA. Improving the HIV prevention landscape to reduce disparities for black MSM in the South. AIDS Behav 2019;23(Suppl 3):331-339 [FREE Full text] [doi: 10.1007/s10461-019-02671-w] [Medline: 31541391]
24. Dworkin MS, Chakraborty A, Zychowski D, Donenberg G, Novak R, Garofalo R. Self-efficacy and ability to read as factors associated with antiretroviral therapy adherence in an HIV-infected population. Int J STD AIDS 2018;29(12):1154-1164. [doi: 10.1177/0956462418776073] [Medline: 29890903]
25. Barclay TR, Hinkin CH, Castellon SA, Mason KI, Reinhard MJ, Marion SD, et al. Age-associated predictors of medication adherence in HIV-positive adults: health beliefs, self-efficacy, and neurocognitive status. Health Psychol 2007;26(1):40-49 [FREE Full text] [doi: 10.1037/0278-6133.26.1.40] [Medline: 17209696]
26. Waldrop-Valverde D, Dong C, Ownby R. Medication-taking self-efficacy and medication adherence among HIV-infected cocaine users. J Assoc Nurses AIDS Care 2013;24(3):198-206 [FREE Full text] [doi: 10.1016/j.jana.2012.05.005] [Medline: 23122904]

Abbreviations

AAMSM: African American men who have sex with men

XSL•FO
RenderX

ART: antiretroviral therapy
HIV-ASES: HIV Treatment Adherence Self-Efficacy Scale
MSPSS: Multidimensional Scale of Perceived Social Support
OR: odds ratio
PHQ-9: Patient Health Questionnaire for Depression
RCT: randomized controlled trial
REALM-SF: Rapid Estimate of Adult Literacy in Medicine-Short Form
RMASS2: Repeated Measures and Sample Size

Edited by L Buis; submitted 28.Mar.2025; peer-reviewed by A Adeoye, ME Heidari; comments to author 13.Jun.2025; revised version received 28.Nov.2025; accepted 01.Dec.2025; published 03.Feb.2026.

Please cite as:
Dworkin MS, Herrera K, Upton S, Luc CM, Jones J, Burns P, Liu L, Jimenez A, Ren R, Woody M, Garofalo R, Lee S
Evidence of Efficacy of the My Personal Health Guide Mobile Phone App on Antiretroviral Therapy Adherence Among Young African American Men Who Have Sex With Men at 1 Month: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e75005
URL: https://mhealth.jmir.org/2026/1/e75005
doi:10.2196/75005
PMID:

©Mark S Dworkin, Kara Herrera, Sierra Upton, Casey M Luc, Jeb Jones, Paul Burns, Li Liu, Antonio Jimenez, Ruiqi Ren, Meaghan Woody, Robert Garofalo, Sangyoon Lee. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 03.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Use of Software as a Medical Device to Improve Therapeutic Adherence in Patients With Hematological Malignancies: Prospective Interventional MargheRITA Study

Serena Camilla Dalto[1], MD; Antonella Romanelli[1], MD; Martina Violati[1], MD; Carlotta Galeone[2], PhD; Davide Gaudesi[3], PhD; Francesca Sacchi[3], MSc; Massimo Beccaria[3], MEng; Mauro Moroni[4], MD; Alessandro Ferri[3], EMBA, MSc; Vittorio Montefusco[1], MD

[1]Hematology Division, ASST Santi Paolo e Carlo, Milano, Italy

[2]Biostatistics & Outcome Research, Statinfo, Bicocca Applied Statistics Center (B-ASC), Universita degli Studi di Milano-Bicocca, Milano, Italy

[3]Advice Pharma Group S.r.l. (Italy), Milano, Italy

[4]Oncology Division, ASST Santi Paolo e Carlo, Milano, Italy

Corresponding Author:
Alessandro Ferri, EMBA, MSc
Advice Pharma Group S.r.l. (Italy)
Via G. Durando 38
Milano, 20158
Italy
Phone: 39 3286656325
Email: ferri11settembre@gmail.com

Abstract

Background: Hematological malignancies are a global health challenge, with a substantial number of deaths each year. Treatment adherence is crucial for improving patient outcomes in patients with hematological malignancies, but resource limitations and logistical challenges hinder optimal outpatient management. Digital health solutions such as the Remote Intelligence for Therapeutic Adherence (RITA) software as a medical device (SaMD) offer potential solutions by facilitating telemedicine visits and supporting patients in managing their treatment.

Objective: The aim of this study was to evaluate the performance and safety of RITA SaMD in improving patient adherence to treatment protocols for hematological malignancies.

Methods: This prospective clinical investigation enrolled patients with hematological malignancies at Azienda Socio Sanitaria Territoriale Santi Paolo e Carlo in Milan, Italy. The RITA SaMD group used the RITA platform, while the control group comprised historical patients. The primary end point was average therapeutic adherence to the prescribed drug treatment, measured as at least the 80% of the relative dose intensity, at month 3. Secondary end points were based on comparisons between the RITA SaMD group and the control group and included the average therapeutic adherence to the prescribed drug treatment at months 1 and 2, number of emergency room visits for minor and severe complications, and the average hospital stay. Multivariable logistic regression models were used to evaluate the effectiveness of RITA.

Results: Between July and December 2022, 119 patients were included in the analysis (57 in the RITA SaMD group and 62 in the control group). At multivariable analysis, the probability of being adherent to treatment at month 3 in the RITA SaMD group was significantly higher than that in the control group (odds ratio 3.0, 95% CI 1.0-8.8; P=.04). A total of 1476 self-reported adverse events (AEs) were collected through RITA SaMD usage, the majority (N=1080) being grade 1 events. During the study visits, 20 AEs were recorded by the study physician (16 in the RITA SaMD group and 4 in the control group). Of the recorded AEs during study visits, 14 were serious AEs (11 in the RITA SaMD group and 3 in the control group). None of the reported AEs was considered related to RITA SaMD usage.

Conclusions: The MargheRITA clinical investigation showed that after 3 months of using RITA SaMD, patients with hematological malignancies had 3 times higher odds ratio of being adherent to the prescribed treatment than the control group. The use of RITA SaMD facilitated the reporting of AEs, reinforcing the role of mobile health apps and software in optimizing patient outcomes. Further research is needed to fully understand its interdisciplinary potential and long-term impact on patient outcomes.

XSL•FO
RenderX

Trial Registration: ClinicalTrials.gov CTI05260203; https://clinicaltrials.gov/study/NCT05260203

(JMIR Mhealth Uhealth 2026;14:e59662) doi:10.2196/59662

KEYWORDS

medical device software; digital health; therapy adherence; mobile health; patient outcome; mHealth; medical device; adherence; hematological malignancies; telemedicine; cancer

Introduction

Global incident cases of hematologic malignancies have been increasing since 1990, surpassing 1,343,000 in 2019. Still, the age-standardized death rate for all types of hematologic malignancies has been declining [1]. This implies more extended patient management for the health care system.

Treatment adherence is a critical factor in the management of patients with hematological malignancies since poor adherence can result in the emergence of resistance [2,3]. Often, patients underestimate the mild drug- or disease-related symptoms during the treatment, leading to a more severe symptomatology later on that could further impair the patient's adherence to treatment [4]. In addition, many antitumor therapies are now continuous treatments, making therapeutic adherence particularly relevant for patient outcomes. For instance, in a real-world study comparing ibrutinib and acalabrutinib, the two drugs used to treat chronic lymphocytic leukemia, patients were more likely to be adherent to treatment (defined as ≥80% adherence) if they were being treated with ibrutinib, which is administered once daily than if they were being treated with acalabrutinib, which needs to be administered twice daily [5]. Similarly, a study assessing adherence to imatinib treatment in a resource-constrained environment showed that adherence during continuous treatment was related to molecular responses in patients with chronic myeloid leukemia [6]. These data underline the importance of developing innovative approaches to improve therapeutic adherence in patients with hematological malignancies. Therefore, outpatient care models have been developed to offer patients the advantages of home-based treatments, but their implementation and optimization in clinical practice are often challenging as the health care system must deal with resource constraints, exacerbated by factors such as the aging population and improvements in survival rates [1].

In response to these challenges, innovative solutions based on digital health systems such as the use of software as a medical device (SaMD) have recently been designed to complement traditional outpatient care [7]. SaMD can potentially facilitate access to oncological treatments for a wider audience of patients, overcoming geographical and logistical barriers [7].

The onset of the COVID-19 pandemic underscored the urgency of streamlining outpatient management processes [8-10]. Among the most promising strategies to address this need is the development of specific SaMD for oncology, which facilitates telemedicine visits and provides ongoing support for patients in their daily management of the disease, its associated complications, and treatment regimens [8-11]. Compared to nonmedical device software and health care apps, SaMD is developed following the International Organization for Standardization (ISO) 13485 and the European Medical Devices Regulation (EU MDR) 2017/745 (where it is known as medical device software) ensures patient safety and is an essential tool for improving health care quality [12,13].

With this in mind, an SaMD called Remote Intelligence for Therapeutic Adherence (RITA) was developed to serve as a communication platform between doctor and patient and to support the patient in his or her therapeutic journey.

This manuscript briefly describes RITA SaMD and reports the findings of the MargheRITA clinical investigation, which evaluates the performance and safety of the RITA SaMD in improving therapeutic adherence to treatment protocols in patients with hematological malignancies.

Methods

The CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist version 1.6.1 [14] was used to report this study (Multimedia Appendix 1).

Study Design

This was a premarket, pivotal, confirmatory, prospective, and interventional clinical investigation in patients with hematological malignancies. Patients were screened in a single center in Italy from July to September 2022; the period from the end of enrollment on September 22, 2022, to December 16, 2022, needed to match the RITA SaMD patient to the control group patient.

Ethical Considerations

This study was conducted under the clinical investigation plan and international guidelines, including the most recent versions of the Declaration of Helsinki, the International Conference on Harmonization guidelines for Good Clinical Practice, EU MDR 2017/745 [13], ISO 14155, as well as all applicable local laws and regulations. The clinical investigation plan was approved by the local ethics committee (Milano Area 1 IEC 0019486; April 21, 2022) and the Italian Ministry of Health (protocol 2028, May 3, 2022). Signed informed consent was obtained from all patients, starting from July 4, 2022 (start of enrollment) to September 22, 2022 (end of enrollment); the period from end of enrollment to December 16, 2022, was used to match the RITA SaMD patients to the control group patients. No compensation was offered to the participants. All data were anonymized to avoid any possible identification and collected in a validated electronic case report form. The clinical investigation is registered at ClinicalTrials.gov (NCT05260203).

XSL•FO
RenderX

Patient Recruitment

The investigation enrolled individuals who were followed up as outpatients at Azienda Socio Sanitaria Territoriale Santi Paolo e Carlo in Milan who met the following inclusion criteria: comprehension of the investigation's objectives and procedures as well as acceptance to voluntarily sign an informed consent form before any investigation-related activities. In addition, participants needed to be at least 18 years of age and have a diagnosis of a hematological malignancy (encompassing conditions such as symptomatic multiple myeloma, solitary plasmacytoma, amyloidosis, chronic myeloid leukemia, chronic lymphocytic leukemia, lymphocytic lymphoma, Hodgkin lymphoma, B-cell non-Hodgkin lymphoma, T-cell non-Hodgkin lymphoma, acute myeloid leukemia, myelodysplasia, and chronic myeloproliferative disorder). Eligibility also required prior receipt of standard-of-care therapy, irrespective of the administration method, and was open to patients at various therapy stages (in their first or subsequent line of therapy and at the beginning of treatment or not), with a minimum life expectancy of more than 6 months. Exclusion criteria were established to ensure investigation integrity and participant well-being, excluding patients who had solely received radiotherapy, those with clinical conditions hindering treatment adherence, individuals unable to use smartphones or computers, those with major psychopathological conditions or cognitive impairments potentially affecting participation, and patients enrolled in another clinical investigation or in a clinical trial at the time of recruitment and patients with no available information on inclusion/exclusion criteria, pathology, treatments (therapeutic treatments, changes in prescribed regimens, missed doses), and outcomes, thus enabling the evaluation of study end points. Participants had the ability to opt out of the study.

RITA SaMD

RITA SaMD (Advice Pharma Group S.r.l) was designed and developed as a class I medical device in accordance with ISO 13485 and EU MDR 2017/745 and intended to increase treatment adherence in oncohematological therapeutic regimens. The RITA SaMD system includes the clinical panel, a mobile app, and the backend. The clinical panel is a web platform managed by the health care professional on which doctors access through a dedicated medical dashboard, after authentication and authorization, to record the patient’s anamnesis and diagnosis, type of drugs, and corresponding posology, both prescribed and effectively taken during follow-up (information verified and recorded at each study visit). The mobile app allows the patient to receive and communicate information about the quality of life and treatment-related adverse events (AEs), providing support in therapy management. The RITA app is designed as a monitoring and support tool for patients with hematological malignancies, allowing clinical professionals to have a clear and complete view of the patient’s state of health. The backend is part of SaMD dedicated to exchanging communications between the RITA mobile app and the clinical panel.

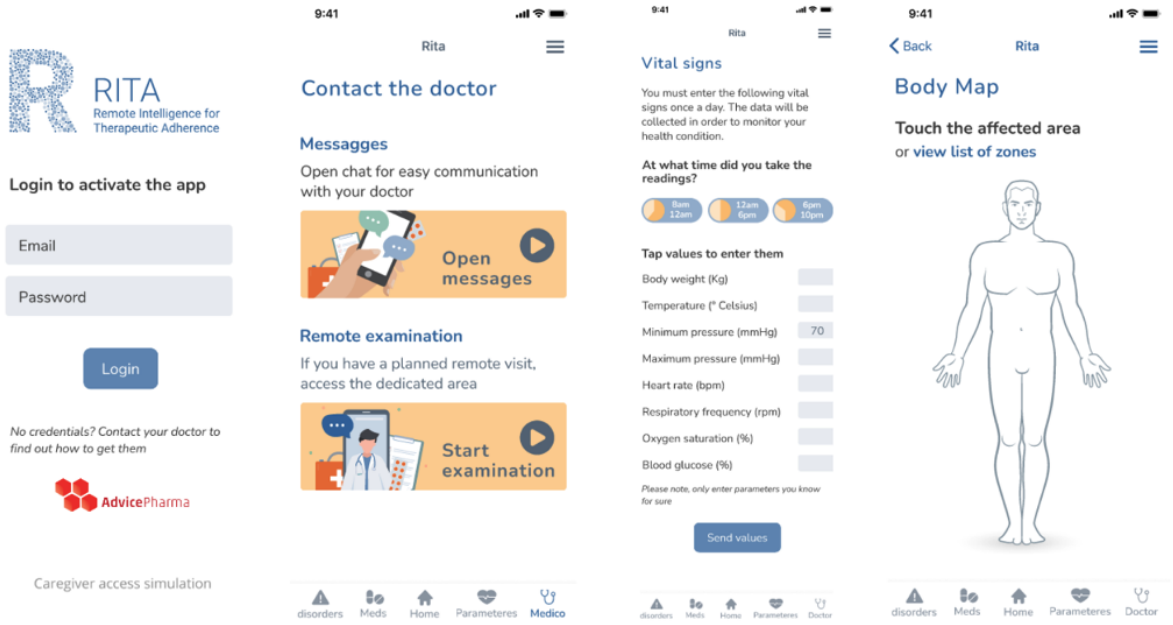
This software is intended to be used by oncohematologists in health care structures (clinical panel module) and by patients in a home care setting (mobile app module). It was designed on the doctor's side as a support for the management of the oncohematological patient and, on the patient's side, as a first aid tool for the management of the most common problems, as well as an efficient and noninvasive means of communication with their doctor. The system also includes the caregiver and general practitioner in the communication group.

RITA SaMD is also designed to communicate and receive information on the quality of life and side effects, thus making the medical doctor aware of the health status of his or her patient population.

Using RITA SaMD, patients could manually notify their daily physical condition as well as disorders, vital parameters (minimum blood pressure, max blood pressure, temperature, heart rate, breath frequency, oxygen saturation, blood sugar, and body weight), and therapy (drug’s name and assumption’s time). RITA SaMD collects patient complaints with a proprietary grading system, which is modeled on Common Terminology Criteria for Adverse Events (version 5.0) [15] (Figure 1). In this way, physicians could manage and resolve mild to moderate symptoms during treatment.

XSL•FO
RenderX

Figure 1. Example screenshots of Remote Intelligence for Therapeutic Adherence (RITA) software as a medical device (SaMD). Physicians were able to access RITA SaMD directly from a computer by using a web-based interface.



The RITA SaMD database has been developed on an electronic data capture technology (ICE, Advice Pharma), compliant with the regulations for clinical data management to facilitate the study data analysis.

Study Interventions

The principal investigator granted access to RITA SaMD to each patient prospectively enrolled in the study (RITA SaMD group). The follow-up visits occurred 1, 2, and 3 months after enrollment. The follow-up visits at 3 months coincided with the end of the investigation visit. Each follow-up visit was a routine standard-of-care visit; no investigation-specific assessment was performed. During this visit, the physician paid particular attention to the patient's reported AE/serious AEs (SAEs). At each follow-up visit, the physician evaluated the self-reported delivered dose intensity.

A control group was also established; this group included patients with hematological malignancies treated in the same hospital in the past 3 years. Each patient in the RITA SaMD group was paired in a 1:1 ratio with a patient in the control group of the same gender, pathology, treatment, and of the closest possible age.

End Points

The primary end point was the average therapeutic adherence to the prescribed drug treatment, measured as at least 80% of the relative dose intensity, at month 3. Mean treatment adherence was evaluated at the end of months 1, 2, and 3; the primary end point considered the 3-month mean treatment adherence, accounting for the adherence data over 3 months of follow-up. Dose intensity was defined as the ratio of self-reported delivered dose intensity to the prescribed referenced dose intensity (expressed as a percentage) during the investigation period. Data used to calculate the therapeutic adherence were recorded in the clinical panel by the referring physicians; the therapeutic adherence calculation was performed during the analysis. The effectiveness of RITA SaMD use was evaluated by comparing the results of therapeutic adherence with the control group.

At the baseline visit, only for oral medications, routine nursing interventions were provided, as nonoral medications were administered at the clinical site (hospital) by nurses or physicians. Physicians and nurses educated patients on the intake of prescribed therapy in terms of amount and modality, basic nursing care for symptoms, and related complications. Furthermore, for each oral treatment, they verbally instructed patients on how to record information about their prescribed treatments. The physicians asked to document the number of pills assumed according to the assumption scheduling, that is, daily for daily drug assumption, every 7 days for weekly assumption at each visit. No specific questionnaire or standardized tools of recording adherence of oral medications were used, allowing each patient to self-report information as preferred. This mode of recording adherence was used because it is the standard of care in use at the clinical center and because it avoided bias of data collection of this information with respect to matched control group.

During each visit of the study, clinicians reported the therapy administration information reported by patients. The answers were recorded in the electronic case report form. Doses effectively assumed, reason for the dose reductions, percentage of dose reductions, and delays in the administration (both the number of the days and the reason of the delay) were also documented. Only physicians talked with participants during the study visits to gather data. The patients daily tracked the

XSL•FO
RenderX

self-reported delivered dose intensity and then reported these data to the physician at study visits at months 1, 2, and 3.

Secondary end points were based on comparisons between the RITA SaMD group and the control group and included the average therapeutic adherence to the prescribed drug treatment at months 1 and 2, number of emergency room (ER) visits for minor and severe complications, and the average hospital stay. Safety secondary end points included physician-reported AEs during the study visits in the RITA SaMD group vs the control group. Other safety secondary end points were only evaluated in the RITA SaMD group and included self-reported AEs and adverse device effects through RITA SaMD usage.

Patient-reported outcomes (PROs) were collected at months 1, 2, and 3 in the RITA SaMD group. Among the questionnaires used were the EQ-5D-5L descriptive system [16], the EuroQol visual analog scale [17], the Instrumental Activities of Daily Living scale [18], and the Activities of Daily Living scale [19].

Statistical Analysis

The clinical investigation planned a sample size of 112 patients (56 patients in the RITA SaMD group and 56 in the control group) to provide 80% power to detect 20% improvement in treatment adherence after 3 months of RITA SaMD usage, with a 1-sided type I error rate of .05. The enrollment of 124 patients (62 patients per group) was set as the goal to account for a dropout rate of 10%.

The primary analysis included all patients observed until the end of the investigation. The patient demographics were reported using descriptive analyses by tabulating frequencies and percentages for categorical variables and mean and median values, SD, quartiles, and extreme values for continuous variables. Continuous data between RITA SaMD and control groups were compared using a 2-sided Student *t* test, after verifying that the data were normally distributed (based on the Shapiro-Wilk test) and a 2-sided Wilcoxon's rank-sum test otherwise. For categorical data, comparisons between groups were performed using the chi-square or Fisher exact test, as appropriate.

The effectiveness of RITA SaMD was compared between the RITA SaMD group and the control group by using multivariable logistic regression models, including the following factors (chosen a priori) as the main confounders: age, gender, comorbidities, and disease severity. Adjusted odds ratios (ORs) and the corresponding 95% CIs were estimated and reported. AEs occurring during the investigation follow-up were also described.

The statistical significance limit was accepted as 5%, and the results below this value ($P<.05$) were considered statistically significant. Statistical analyses were performed using SAS system software (v. 9.4; SAS Institute Inc).

Results

Patient Characteristics

The first RITA SaMD patient signed the informed consent on July 4, 2022 (start of enrollment), and the last RITA SaMD patient signed the informed consent on September 22, 2022. A total of 62 patients were enrolled in the RITA SaMD group and 62 were, therefore, matched in the control group. Among patients in the RITA SaMD group, 5 did not complete the study: 1 patient was lost to follow-up, 3 patients died from causes not related to the investigation, and 1 patient discontinued treatment after an SAE not related to the investigational device. Thus, 119 patients were analyzed for the primary end point: 57 patients in the RITA SaMD group and 62 in the control group. Baseline patient characteristics are reported in Table 1.

XSL•FO
RenderX

Table 1. Baseline patient characteristics.

	RITA[a] SaMD[b] group (n=62)	Control group (n=62)	*P* value
Gender, n (%)			.21
Male, n (%)	37 (59.7)	30 (48.4)	
Female, n (%)	25 (40.3)	32 (51.6)	
Age (y), mean (SD)	69.3 (15.1)	73.57 (9.9)	.16
Pathology, n (%)			.99
Monoclonal gammopathy	22 (35.5)	20 (32.3)	
Chronic myeloid leukemia	4 (6.5)	22 (35.5)	
Chronic lymphocytic leukemia/lymphocytic lymphoma	9 (14.5)	9 (14.5)	
Hodgkin lymphoma	3 (4.8)	3 (4.8)	
B-cell non-Hodgkin lymphoma	12 (19.4)	13 (21)	
T-cell non-Hodgkin lymphoma	3 (4.8)	3 (4.8)	
Acute myeloid leukemia	2 (3.2)	2 (3.2)	
Myelodysplasia	3 (4.8)	4 (4.5)	
Chronic myeloproliferative syndrome	4 (4.5)	4 (4.5)	
Treatment phase, n (%)			.005
1st line	52 (83.8)	38 (61.3)	
2nd line or more	10 (16.1)	24 (38.7)	
Method of treatment administration, n (%)			.37
Intravenous	35 (56.5)	30 (51.6)	
Oral/other	27 (43.5)	32 (48.4)	

[a]RITA: Remote Intelligence for Therapeutic Adherence.

[b]SaMD: software as a medical device.

Performance

Primary End Point

In the primary end point analysis, 4 patients in the RITA SaMD group had missing data at the follow-up study visits. Therefore, the analyses included 115 patients at month 3. At univariable analysis, the proportion of treatment-adherent patients, that is, ≥80% adherent to treatment, in the RITA SaMD group tended to be slightly higher than that in the control group at all time points (46/57, 82.1% vs 50/62, 80.6% at month 1, 43/57, 81.1% vs 45/62, 76.3% at month 2, and 46/57, 85.2% vs 47/62, 77% at month 3) (Table 2).

Table 2. Treatment adherence to therapy.

	RITA[a] SaMD[b] group (n=57), n (%)	Control group (n=62), n (%)	*P* value
Month 1	46 (82.1)	50 (80.6)	.84
Month 2	43 (81.1)	45 (76.3)	.53
Month 3	46 (85.2)	47 (77)	.27

[a]RITA: Remote Intelligence for Therapeutic Adherence.

[b]SaMD: software as a medical device.

n terms of treatment adherence, univariable analyses revealed that differences between the two groups were not statistically significant, even when considering the mode of treatment administration (intravenous vs oral): treatment adherence was higher in the RITA SaMD Group than in the Control Group [intravenous: 22/3, 71% vs 22/30, 73.3% at Month 1 (*P*=.84), 20/29, 69% vs 16/27, 59.2% at month 2 (*P*=.45), and 23/27, 85.1% vs 18/29, 62% at month 3 (*P*=.09); oral: 24/25, 96% vs 28/32, 87.5% at month 1 (*P*=.26), 23/24, 95.8% vs 29/32, 90.6% at month 2 (*P*=.45), and 23/26, 88.5% vs 29/32, 90.6% at month 3 (*P*=.79)].

A multivariable analysis, adjusted for potential confounders (gender, age, number of concomitant pathologies, and treatment line), was performed. At month 3, the OR of the RITA SaMD group to be adherent to treatment was 3 times higher than that of the control group (adjusted OR 3.0, 95% CI 1.0-8.8; *P*=.04) (Table 3).

XSL•FO
RenderX

Table 3. Multivariable logistic regression analysis on the risk of being adherent to therapy over 3 months.

	Adjusted odds ratio (95% CI)	*P* value
Use of RITA[a] SaMD[b] app at month 3 (yes vs no)	3.0 (1.0-8.8)	.04
Sex (female vs male)	1.4 (0.5-3.8)	.53
Age (60-79 y vs 60 y)	2.5 (0.6-10.6)	.23
Age (≥80 y vs 60 y)	2.2 (0.5-10.3)	.30
Number of concomitant pathologies (1-2 vs 0)	0.9 (0.1-5.3)	.88
Number of concomitant pathologies (≥3 vs 0)	0.5 (0.1-2.3)	.35
Number of treatment lines (≥1 vs 1)	6.8 (1.4-33.3)	.02

[a]RITA: Remote Intelligence for Therapeutic Adherence.

[b]SaMD: software as a medical device.

Secondary End Points

Treatment adherence was analyzed in 118 patients at month 1 and 112 patients at month 2. The multivariable analysis with adjustment for potential confounders showed that the differences between the two groups in terms of treatment adherence were not statistically significant in the first 2 months, with an OR of 1.7 (95% CI 0.61-4.9) at month 1 and an OR of 2.0 (95% CI 0.8-5.6) at month 2.

Regarding ER visits, there were no visits in either group for minor complications. At the end of the study period, there were 9 ER visits in the RITA SaMD group requiring hospitalization and 3 in the control group, totaling 12 ER visits for severe complications. The average hospital stay was 25.3 (SD 22.7) days in the RITA SaMD group and 8.3 (SD 2.1) days in the control group. Due to the small number of events, no statistical test was performed for these end points.

Safety (Secondary End Points)

In the RITA SaMD Group, there were 16 AEs reported by the physicians during the study visits (Table 4), most of them grade 2, while in the control group, there were 4 AEs reported (Table 5). During the study, 14 SAEs occurred, 11 of which were in the RITA SaMD group. None of the recorded AEs was related to RITA SaMD usage; therefore, there were no adverse device effects.

XSL•FO
RenderX

Table 4. Adverse events recorded by the physician during the study visits—RITAa SaMDb group (n=16).

Body system organ class	Grade 1 (n=1), n (%)	Grade 2 (n=8), n (%)	Grade 3 (n=3), n (%)	Grade 4 (n=2), n (%)	Grade 5 (n=2), n (%)
Cardiac disorders (n=4)	0 (0)	3 (75)	1 (25)	0 (0)	0 (0)
Cardiac failure congestive (n=1)	0 (0)	0 (0)	1 (100)	0 (0)	0 (0)
Cardiac failure (n=1)	0 (0)	1 (100)	0 (0)	0 (0)	0 (0)
Atrial fibrillation (n=2)	0 (0)	2 (100)	0 (0)	0 (0)	0 (0)
General disorders and administration site conditions (n=5)	0 (0)	3 (60)	0 (0)	0 (0)	2 (40)
Asthenia (n=2)	0 (0)	2 (100)	0 (0)	0 (0)	0 (0)
Disease progression (n=2)	0 (0)	0 (0)	0 (0)	0 (0)	2 (100)
Pyrexia (fever) (n=1)	0 (0)	1 (100)	0 (0)	0 (0)	0 (0)
Infections and infestations (n=4)	0 (0)	2 (50)	1 (25)	1 (25)	0 (0)
COVID-19 infection (n=2)	0 (0)	1 (50)	1 (50)	0 (0)	0 (0)
Infection (not specified) (n=1)	0 (0)	1 (100)	0 (0)	0 (0)	0 (0)
Septic shock (n=1)	0 (0)	0 (0)	0 (0)	1 (100)	0 (0)
Respiratory, thoracic, and mediastinal disorders (n=1)	0 (0)	0 (0)	1 (100)	0 (0)	0 (0)
Interstitial lung disease (interstitial pneumonia) (n=1)	0 (0)	0 (0)	1 (100)	0 (0)	0 (0)
Vascular disorders (n=2)	1 (50)	0 (0)	0 (0)	1 (50)	0 (0)
Hypotensive syncope (anemia lipothymia) (n=1)	1 (100)	0 (0)	0 (0)	0 (0)	0 (0)
Cardiovascular disorder (recurrence of circulatory decompensation) (n=1)	0 (0)	0 (0)	0 (0)	1 (100)	0 (0)

[a]RITA: Remote Intelligence for Therapeutic Adherence.

[b]SaMD: software as a medical device.

Table 5. Adverse events recorded by the physician during the study visits—control group (n=4).

Body system organ class	Grade 1 (n=0), n (%)	Grade 2 (n=1), n (%)	Grade 3 (n=3), n (%)	Grade 4 (n=0), n (%)	Grade 5 (n=0), n (%)
Blood and lymphatic system disorders (n=1)	0 (0)	0 (0)	1 (100)	0 (0)	0 (0)
Neutropenia (n=1)	0 (0)	0 (0)	1 (100)	0 (0)	0 (0)
General disorders and administration site conditions (n=1)	0 (0)	1 (100)	0 (0)	0 (0)	0 (0)
Asthenia (n=1)	0 (0)	1 (100)	0 (0)	0 (0)	0 (0)
Hepatobiliary disorders (n=1)	0 (0)	0 (0)	1 (100)	0 (0)	0 (0)
Hyperbilirubinemia (n=1)	0 (0)	0 (0)	1 (100)	0 (0)	0 (0)
Respiratory, thoracic, and mediastinal disorders (n=1)	0 (0)	0 (0)	1 (100)	0 (0)	0 (0)
Dyspnea (persistent fever and difficulty breathing) (n=1)	0 (0)	0 (0)	1 (100)	0 (0)	0 (0)

Regarding self-reported AEs, a total of 1476 AEs were recorded through RITA SaMD usage (Table 6). There were 1080 minor (grade 1) events, with tiredness being the most frequently reported grade 1 event (n=720).

XSL•FO
RenderX

Table 6. Self-reported adverse events in RITAa SaMDb group.c

Body system organ class	Grade 1 (n=1080), n (%)	Grade 2 (n=370), n (%)	Grade 3 (n=26), n (%)
Cardiac disorders (n=5)	3 (60)	2 (40)	0 (0)
Palpitations (n=5)	3 (60)	2 (40)	0 (0)
Eye disorders (n=99)	0 (0)	90 (90.9)	9 (9.1)
Vision decreased (decrease of vision) (n=99)	0 (0)	90 (90.9)	9 (9.1)
Gastrointestinal disorders (n=50)	41 (82)	7 (14)	2 (4)
Diarrhea (n=32)	27 (84.4)	4 (12.5)	1 (3.1)
Constipation (n=10)	7 (70)	3 (30)	0 (0)
Stomatitis (oral cavity mucositis) (n=6)	6 (100)	0 (0)	0 (0)
Dyspepsia (difficult digestion) (n=1)	0 (0)	0 (0)	1 (100)
Gastric dilatation (stomach distension) (n=1)	1 (100)	0 (0)	0 (0)
General disorders and administration site conditions (n=955)	773 (80.9)	172 (18)	10 (1.1)
Fatigue (tiredness) (n=858)	720 (83.9)	129 (15)	9 (1.1)
Peripheral swelling (swelling in the legs) (n=42)	12 (28.6)	30 (71.4)	0 (0)
Pain (ache) (n=30)	17 (56.7)	12 (40)	1 (3.3)
Pyrexia (fever) (n=25)	24 (96)	1 (4)	0 (0)
Infections and infestations (n=41)	36 (87.8)	5 (12.2)	0 (0)
Flu symptoms (n=39)	35 (89.7)	4 (10.3)	0 (0)
Conjunctivitis (n=2)	1 (50)	1 (50)	0 (0)
Musculoskeletal and connective tissue disorders (n=3)	3 (100)	0 (0)	0 (0)
Gait disturbance (walking disorders) (n=3)	3 (100)	0 (0)	0 (0)
Nervous system disorders (n=103)	44 (42.7)	59 (57.3)	0 (0)
Insomnia (n=55)	6 (10.9)	49 (89.1)	0 (0)
Anxiety (n=29)	21 (72.4)	8 (27.6)	0 (0)
Taste disorder (n=14)	13 (92.9)	1 (7.1)	0 (0)
Tremors (n=3)	3 (100)	0 (0)	0 (0)
Cognitive disorder (n=2)	1 (50)	1 (50)	0 (0)
Psychiatric disorders (n=14)	13 (92.9)	1 (7.1)	0 (0)
Nervousness (n=14)	13 (92.9)	1 (7.1)	0 (0)
Renal and urinary disorders (n=14)	14 (100)	0 (0)	0 (0)
Pollakiuria (urinate often) (n=13)	13 (100)	0 (0)	0 (0)
Urinary incontinence (n=1)	1 (100)	0 (0)	0 (0)
Respiratory, thoracic, and mediastinal disorders (n=38)	17 (44.7)	17 (44.7)	4 (10.5)
Cough (n=36)	16 (44.4)	17 (47.2)	3 (8.3)
Dyspnea (difficulty breathing) (n=1)	0 (0)	0 (0)	1 (100)
Hiccups (n=1)	1 (100)	0 (0)	0 (0)
Skin and subcutaneous tissue disorders (n=146)	129 (88.4)	16 (11)	1 (0.7)
Pruritus (itching) (n=122)	110 (90.2)	11 (9)	1 (0.8)
Paresthesia (tingling feeling) (n=21)	17 (81)	4 (19)	0 (0)
Skin alterations (n=3)	2 (66.7)	1 (33.3)	0 (0)
Vascular disorders (n=8)	7 (87.5)	1 (12.5)	0 (0)
Epistaxis – nosebleed (n=8)	7 (87.5)	1 (12.5)	0 (0)

XSL•FO
RenderX

[a]RITA: Remote Intelligence for Therapeutic Adherence.

[b]SaMD: software as a medical device.

[c]Percentages are based on the total number of cases for each row.

PROs in Questionnaires

In the RITA SaMD group, 1 patient had complete questionnaires at all time points and 17 patients had complete questionnaires at month 1 and month 2. For each questionnaire, some patients were not considered in the analysis because their assessments were done out of the specified time frame, they reported incorrect values, or data were missing (Table 7).

Table 7. Patient-reported outcomes.

	No change, n	Improved, n	Worsened, n
EQ-5D-5L			
Mobility item	6	2	3
Self-care item	9	1	1
Usual activities item	9	1	1
Pain/discomfort	2	3	6
Anxiety/depression	4	1	6
Other			
EuroQol Visual Analog Scale	2	5	3
Instrumental Activities of Daily Living scale	—[a]	1	2
Activities of Daily Living scale	8	—	1

[a]Not applicable.

Discussion

Principal Findings

In this clinical investigation of patients with hematological malignancies, patients using the RITA SaMD device had a 3-fold higher risk of being adherent to treatment compared with nonusers after 3 months of use, with patients who received more than 1 line of therapy having a higher likelihood to be adherent. A high number of self-reported AEs of grade 1 was collected in the RITA SaMD group.

Given that patients with hematological malignancies often must adhere to strenuous medication regimens, manage symptoms associated with treatment, and, eventually, experience late effects associated with survivorship, they would benefit greatly from digital technology aimed at decreasing these burdens [20]. Indeed, previous research in the area of mobile health has shown that the use of information and communication platforms can increase the patient's sense of autonomy and facilitate disease management, resulting in higher quality of life and treatment adherence [21].

The results of a pragmatic pilot trial comparing the use of the digital solution Safety and Adherence to Medications and Self-Care Advice in Oncology (SAMSON) with usual care in patients with hematological malignancies who started oral cancer medicines showed that the use of SAMSON was acceptable, feasible, and helpful in improving medication adherence [22].

Another experience with an online eHealth system for patients to self-report symptoms during cancer treatment, the Electronic Patient Self-Reporting of Adverse-events: Patient Information and Advice (eRAPID) demonstrated how real-time monitoring with PROs ameliorated physical well-being in terms of symptom control measured using the Functional Assessment of Cancer Therapy Scale General Physical Well-Being subscale at 6 weeks. Clinically meaningful benefits in patients' physical well-being were observed during the first 6 and 12 weeks of treatment when it is expected to see challenges controlling side effects; symptoms stabilized by week 18, when patients reach a steady state of supportive medications or chemotherapy dose adjustments [23]. Based on these observations, the predefined period of observations of 3 months that we adopted in our study was appropriate.

In our study, the number of previous treatment lines affected the odds of being adherent to treatment, with patients who underwent more than 1 line of treatment having a much higher likelihood of adhering to the prescribed drug treatment. This is somewhat in contradiction to what was found by Seal et al [24], who described how adherence declines with increasing lines of treatment but only for oral medications and not intravenous ones. Given that the number of available oral formulations of both chemotherapy and biological therapies is rising and that patients tend to prefer oral therapy, it is worth further exploring the role of mobile health and SaMD in the factors influencing adherence, such as forgetfulness or loss of motivation [25,26].

A total of 1476 AEs were self-reported in the RITA SaMD group with SaMD usage occurrence; the majority of them (73.2%) was of grade 1, showing that app usage facilitated collection of such events and may improve the recording of low-grade AEs compared with traditional methods (namely, AE recording only during medical visits). During the study visits, 20 AEs were recorded by the physician: 16 in the RITA SaMD group and 4 in the control group. Of the collected AEs

XSL•FO
RenderX

during the study visits, 14 were considered SAEs (11 in the RITA SaMD group and 3 in the control group). None of the collected AEs was considered to be related to RITA SaMD usage. In a recent study evaluating the factors impacting medication adherence in patients with hematologic malignancies who have undergone allogeneic hematopoietic stem cell transplants, although AEs associated with medication did not directly impact treatment adherence, they increased the patients' psychological distress [27]. Interestingly, the same study reported that caregiver and clinician support as well as tools to aid medication management were factors that facilitated treatment adherence [27]. RITA SaMD can, therefore, help alleviate stress by providing a way for patients to report their ailments and feel they have clinician support.

RITA SaMD allowed the collection of PROs, including quality of life data. Although the collected data were insufficient to analyze differences in therapeutic adherence relative to self-reported changes in the questionnaires' items, the fact that RITA SaMD enables the collection of PROs is an added value and may be further explored in the future. Unrecorded PROs are still sadly familiar in the literature; any SaMD that strives to help both patients and physicians work better with the health care system cannot overlook PRO recording and the real-world data it generates [17,28].

Previous research has identified necessary features in mobile health apps and software, such as symptoms and medication tracking [29]. Patients themselves also state that features that facilitate information exchange with their clinicians are important [30]. Our results in patients with hematological malignancies illustrate how RITA SaMD can help any oncology setting to promote patients' therapeutic adherence while limiting health care resources used in patient management and contribute to optimal patient care.

Strengths and Limitations

Our clinical investigation has some limitations. This study was undertaken in a single center, which might limit the generalizability of the results. Another limitation is related to the lack of randomization and the unblinded nature of the study, potentially resulting in some biases such as selection bias, recall bias, and bias due to missing data, which were addressed using appropriate methodological approaches during the study design. To overcome the selection bias, the controls originated from the same population used to enroll patients in the intervention group: they had been treated in the same hospital and followed the same treatment protocols in the 3 years before the start of enrollment of patients in the intervention arm. In addition, control patients were identified based on the availability in the hospital's database of information on inclusion/exclusion criteria, pathology, treatments, and outcomes, thus enabling the evaluation of study end points. Finally, control patients were matched to patients of the intervention group based on the same characteristics, potentially influencing the primary outcome of the study.

As a further limitation, we acknowledge that no specific questionnaire or standardized tools of recording adherence of oral medications were used, allowing each patient to self-report information as preferred. However, this mode was used because it is the standard of care in use at the clinical center and because it avoided bias of data collection of this information with respect to the matched control group. Participants were instructed to report on all the prescribed oncological therapies; however, we could not ensure that their reporting was correct and unbiased, as no methods were in place to verify participants' self-reporting. Data cleaning verified only if the number of taken doses did not exceed the number of prescribed doses. No imputation was performed; the investigators recorded the available quantities of the drugs effectively assumed in the electronic case report form.

The short follow-up time precludes any conclusion regarding long-term adherence to treatment, despite the promising results. However, the patients considered in the study had different stages of progress in cancer therapy (patients in first-line or later lines, with therapy to be started and with therapies already started); this different composition of the sample implies that the included patients might have different "levels of therapeutic fatigue" and certainly included patients with longer exposure to cancer treatments. The high level of differentiation of the patients included in the study increased the generalization of the results of this study.

Our study also has important strengths. To our knowledge, this is the first clinical investigation showing the effectiveness of an SaMD application in improving therapeutic adherence in patients with hematological malignancies. High-quality methods were implemented, including prospective registration in a publicly available clinical trial database and the use of a control group.

Conclusion

The MargheRITA clinical investigation revealed that after 3 months of using RITA SaMD, patients with hematological malignancies were 3 times more likely to be adherent to the prescribed treatment than the control group. The use of RITA SaMD facilitated the reporting of AEs and (to a minor extent) of PROs, reinforcing the role of mobile health apps and software in optimizing patient outcomes. This initial clinical investigation has confirmed the performance and safety of RITA SaMD; future studies are needed to prove its interdisciplinary application.

Acknowledgments

Francesca Paoloni (senior biostatistician) ran the statistical analysis. Laura C. Collada Ali (senior medical writing consultant), Diana Ribeiro (medical writer), and Elisa Sala (medical writer) provided editorial assistance in preparing this article. The investigators thank Laura Galli for her support, the researchers, and the patients in this clinical investigation for their involvement.

XSL•FO
RenderX

Funding

This clinical investigation has been funded and sponsored by Advice Pharma Group S.r.l. (Italy).

Authors' Contributions

Conceptualization: all authors
Writing – original draft preparation: VM, MM, CG, DG, FS, AF
Review and editing: VM, MM, CG, DG, FS, AF, MB
All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

AF (President of Advice Pharma Group), DG, MB, and FS are employees of Advice Pharma Group S.r.l. (Italy). All remaining authors have no potential conflicts of interest concerning this article's research, authorship, and/or publication.

Multimedia Appendix 1
CONSORT-EHEALTH checklist.
[PDF File (Adobe PDF File), 855 KB - mhealth_v14i1e59662_app1.pdf]

References

1. Zhang N, Wu J, Wang Q, Liang Y, Li X, Chen G, et al. Global burden of hematologic malignancies and evolution patterns over the past 30 years. Blood Cancer J 2023 May 17;13(1):82 [FREE Full text] [doi: 10.1038/s41408-023-00853-3] [Medline: 37193689]
2. Collins J, Stump SE, Heiling H, Muir M, Deal A, Proco D, et al. Impact of adherence to ibrutinib on clinical outcomes in real-world patients with chronic lymphocytic leukemia. Leuk Lymphoma 2022 Aug;63(8):1823-1830. [doi: 10.1080/10428194.2022.2045597] [Medline: 35249442]
3. Williams AM, Baran AM, Casulo C, Reagan P, Friedberg JW, Helber M, et al. Ibrutinib dose adherence and therapeutic efficacy in non-hodgkin lymphoma: a single-center experience. Clin Lymphoma Myeloma Leuk 2019 Jan;19(1):41-47 [FREE Full text] [doi: 10.1016/j.clml.2018.10.005] [Medline: 30409718]
4. Solano M, Daguindau E, Faure C, Loriod P, Pain C, Maes A, et al. Oral therapy adherence and satisfaction in patients with multiple myeloma. Ann Hematol 2021 Jul 03;100(7):1803-1813. [doi: 10.1007/s00277-021-04543-7] [Medline: 33938996]
5. Lu X, Emond B, Morrison L, Kinkead F, Lefebvre P, Lafeuille M, et al. Real-world comparison of first-line treatment adherence between single-agent ibrutinib and acalabrutinib in patients with chronic lymphocytic leukemia. Patient Prefer Adherence 2023;17:2073-2084 [FREE Full text] [doi: 10.2147/PPA.S417180] [Medline: 37641660]
6. Yanamandra U, Malhotra P, Sahu KK, Sushma Y, Saini N, Chauhan P, et al. Variation in adherence measures to imatinib therapy. J Glob Oncol 2018 Sep;4:1-10 [FREE Full text] [doi: 10.1200/JGO.2016.007906] [Medline: 30241224]
7. Raffenaud A, Gurupur V, Fernandes SL, Yeung T. Utilizing telemedicine in oncology settings: patient favourability rates and perceptions of use analysis using chi-square and neural networks. Technol Health Care 2019;27(2):115-127. [doi: 10.3233/THC-181293] [Medline: 30664510]
8. Shirke MM, Shaikh SA, Harky A. Implications of telemedicine in oncology during the COVID-19 pandemic. Acta Biomed 2020 Sep 07;91(3):e2020022 [FREE Full text] [doi: 10.23750/abm.v91i3.9849] [Medline: 32921719]
9. Lewis GD, Hatch SS, Wiederhold LR, Swanson TA. Long-term institutional experience with telemedicine services for radiation oncology: a potential model for long-term utilization. Adv Radiat Oncol 2020;5(4):780-782 [FREE Full text] [doi: 10.1016/j.adro.2020.04.018] [Medline: 32391444]
10. Portnoy J, Waller M, Elliott T. Telemedicine in the era of COVID-19. J Allergy Clin Immunol Pract 2020 May;8(5):1489-1491 [FREE Full text] [doi: 10.1016/j.jaip.2020.03.008] [Medline: 32220575]
11. Efficace F, Breccia M, Fazi P, Cottone F, Holzner B, Vignetti M. The GIMEMA-ALLIANCE digital health platform for patients with hematologic malignancies in the COVID-19 pandemic and postpandemic era: protocol for a multicenter, prospective, observational study. JMIR Res Protoc 2021 Jun 01;10(6):e25271 [FREE Full text] [doi: 10.2196/25271] [Medline: 33890580]
12. Alotaibi YK, Federico F. The impact of health information technology on patient safety. Saudi Med J 2017 Dec;38(12):1173-1180 [FREE Full text] [doi: 10.15537/smj.2017.12.20631] [Medline: 29209664]
13. Regulation (EU) 2017/745. European Union. URL: https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng [accessed 2026-03-12]
14. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
15. Common terminology criteria for adverse events (CTCAE) version 5. US Department of Health and Human Services. 2017 Nov 27. URL: https://dctd.cancer.gov/research/ctep-trials/for-sites/adverse-events/ctcae-v5-5x7.pdf [accessed 2026-03-12]

XSL•FO
RenderX

16. EQ-5D-5L. EuroQol Group. URL: https://euroqol.org/information-and-support/euroqol-instruments/eq-5d-5l/ [accessed 2026-03-12]
17. Feng Y, Parkin D, Devlin NJ. Assessing the performance of the EQ-VAS in the NHS PROMs programme. Qual Life Res 2014 Apr;23(3):977-989 [FREE Full text] [doi: 10.1007/s11136-013-0537-z] [Medline: 24081873]
18. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist 1969;9(3):179-186. [Medline: 5349366]
19. Katz S, Ford AB, Moskowitz RW, et al. Studies of illness in the aged, the index of ADL: a standardized measure of biological and psychosocial function. JAMA 1963 Sep 21;185:914-919. [doi: 10.1001/jama.1963.03060120024016] [Medline: 14044222]
20. Werk RS, Heneghan MB, Badawy SM. Use of patient-centered technology and digital interventions in pediatric and adult patients with hematologic malignancies. Curr Hematol Malig Rep 2024 Aug;19(4):153-162. [doi: 10.1007/s11899-024-00732-z] [Medline: 38806875]
21. Collado-Borrell R, Escudero-Vilaplana V, Ribed A, Gonzalez-Anleo C, Martin-Conde M, Romero-Jimenez R, et al. Effect of a mobile app for the pharmacotherapeutic follow-up of patients with cancer on their health outcomes: quasi-experimental study. JMIR Mhealth Uhealth 2020 Oct 16;8(10):e20480 [FREE Full text] [doi: 10.2196/20480] [Medline: 33064100]
22. Dang TH, Wickramasinghe N, Jayaraman PP, Burbury K, Alexander M, Whitechurch A, et al. Digital solution to support medication adherence and self-management in patients with cancer (SAMSON): pilot randomized controlled trial. JMIR Form Res 2025 Feb 19;9:e65302 [FREE Full text] [doi: 10.2196/65302] [Medline: 39969972]
23. Absolom K, Warrington L, Hudson E, Hewison J, Morris C, Holch P, et al. Phase III randomized controlled trial of eRAPID: eHealth intervention during chemotherapy. J Clin Oncol 2021 Mar 01;39(7):734-747 [FREE Full text] [doi: 10.1200/JCO.20.02015] [Medline: 33417506]
24. Seal BS, Anderson S, Shermock KM. Factors associated with adherence rates for oral and intravenous anticancer therapy in commercially insured patients with metastatic colon cancer. J Manag Care Spec Pharm 2016 Mar;22(3):227-235 [FREE Full text] [doi: 10.18553/jmcp.2016.22.3.227] [Medline: 27003552]
25. Hess LM, Louder A, Winfree K, Zhu YE, Oton AB, Nair R. Factors associated with adherence to and treatment duration of erlotinib among patients with non-small cell lung cancer. J Manag Care Spec Pharm 2017 Jun;23(6):643-652 [FREE Full text] [doi: 10.18553/jmcp.2017.16389] [Medline: 28530522]
26. O'Neill VJ, Twelves CJ. Oral cancer treatment: developments in chemotherapy and beyond. Br J Cancer 2002 Oct 21;87(9):933-937 [FREE Full text] [doi: 10.1038/sj.bjc.6600591] [Medline: 12434279]
27. Amonoo HL, Deary EC, Wang A, Newcomb RA, Daskalakis E, Weber D, et al. Medication adherence in patients with hematologic malignancies who are hematopoietic stem cell transplantation survivors: a qualitative study. Transplant Cell Ther 2023 Oct;29(10):620.e1-620.e11 [FREE Full text] [doi: 10.1016/j.jtct.2023.07.019] [Medline: 37516379]
28. Au H, Ringash J, Brundage M, Palmer M, Richardson H, Meyer RM, NCIC CTG Quality of Life Committee. Added value of health-related quality of life measurement in cancer clinical trials: the experience of the NCIC CTG. Expert Rev Pharmacoecon Outcomes Res 2010 Apr;10(2):119-128. [doi: 10.1586/erp.10.15] [Medline: 20384559]
29. Vaffis S, Whaley S, Axon DR, Hall-Lipsy E, Hincapie A, Slack M, et al. Features of cancer mHealth apps and evidence for patient preferences: scoping literature review. JMIR Cancer 2023 Apr 28;9:e37330 [FREE Full text] [doi: 10.2196/37330] [Medline: 37115587]
30. Richards R, Kinnersley P, Brain K, Staffurth J, Wood F. The preferences of patients with cancer regarding apps to help meet their illness-related information needs: qualitative interview study. JMIR Mhealth Uhealth 2019 Jul 31;7(7):e14187 [FREE Full text] [doi: 10.2196/14187] [Medline: 31368446]

Abbreviations

AE: adverse event
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
eRAPID: Electronic Patient Self-Reporting of Adverse-events: Patient Information and Advice
EU MDR: European Medical Devices Regulation
ISO: International Organization for Standardization
OR: odds ratio
PRO: patient-reported outcome
RITA: Remote Intelligence for Therapeutic Adherence
SAE: serious adverse event
SaMD: software as a medical device
SAMSON: Safety and Adherence to Medications and Self-Care Advice in Oncology

XSL•FO
RenderX

Edited by L Buis; submitted 18.Apr.2024; peer-reviewed by TH Dang, A Mussetti, A Cruz; comments to author 10.Jan.2025; revised version received 29.Oct.2025; accepted 27.Jan.2026; published 27.Mar.2026.

Please cite as:
Dalto SC, Romanelli A, Violati M, Galeone C, Gaudesi D, Sacchi F, Beccaria M, Moroni M, Ferri A, Montefusco V
Use of Software as a Medical Device to Improve Therapeutic Adherence in Patients With Hematological Malignancies: Prospective Interventional MargheRITA Study
JMIR Mhealth Uhealth 2026;14:e59662
URL: https://mhealth.jmir.org/2026/1/e59662
doi:10.2196/59662
PMID:

©Serena Camilla Dalto, Antonella Romanelli, Martina Violati, Carlotta Galeone, Davide Gaudesi, Francesca Sacchi, Massimo Beccaria, Mauro Moroni, Alessandro Ferri, Vittorio Montefusco. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 27.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Health Support for Cataract Surgery With the Sharp Health Companion CareKit App: Randomized Controlled Trial

Tommy S Korn[1], MD; Armand Assissini[1]; Lesli Ann Slater[2]; Janet Holland[1]; Lauren Doone[2]; Elvis Lam[2]; Joy Fox[1]; Kristina L Greenwood[2], PhD; Eric Linebarger[1], MD

[1]Department of Ophthalmology, Sharp HealthCare, Sharp Rees-Stealy Medical Group, San Diego, CA, United States

[2]Sharp HealthCare, San Diego, CA, United States

Corresponding Author:
Tommy S Korn, MD
Department of Ophthalmology
Sharp HealthCare
Sharp Rees-Stealy Medical Group
3075 Health Center Drive, Suite 401
San Diego, CA, 92123
United States
Phone: 1 858 939 5400
Fax: 1 858 939 5419
Email: oppeye@sharp.com

Abstract

Background: Cataract surgery is the most frequently performed surgery worldwide, crucial for restoring sight in millions. The COVID-19 pandemic and an aging population have increased barriers to timely surgery. Missed preoperative instructions and poor adherence to postoperative care contribute to surgery cancellations, delays, and potential complications. Mobile digital health interventions could enhance adherence and reduce cancellations.

Objective: This study assessed the effectiveness of the Sharp Health Companion smartphone app, built on the CareKit health platform, compared with printed instructions. The aim was to evaluate its impact on medication adherence, surgery delays and cancellations, visual outcomes, and patient experience among older adults undergoing cataract surgery.

Methods: In this randomized controlled trial, 200 patients aged 41-87 (mean 69, SD 8.2) years were enrolled at a high-volume ophthalmology practice between December 2022 and January 2024. Most participants (145/200, 73%) were 65 years or older. Patients were randomized to group 1 (printed instructions with phone reminders, n=104) or group 2 (Sharp Health Companion app supplemented with backup printed instructions, n=96). Both groups received identical perioperative instructions and medications. Data included demographics, visual acuity, medication adherence (self-reported by paper checklist or Sharp Health Companion app care-card checklist and eye medication bottle weights), surgery cancellations and delays, and satisfaction surveys administered preoperatively, on postoperative day 1, and at 30 days.

Results: Sharp Health Companion app users had fewer same-day surgery delays (1/96, 1%) than those receiving printed instructions (10/104, 10%; *P*=.01), while cancellation rates were similar (*P*=.33). Patient-reported preparedness for surgery was high and comparable between groups on postoperative day 1 (group 1: mean 9.56, SD 1.19; group 2: 9.77, SD 0.73; *P*=.16). Preparedness for recovery at postoperative month 1 was similarly high (mean 9.92, SD 0.37 vs mean 9.85, SD 0.47; *P*=.28). At 30 days postoperatively, visual acuity improvement was similar (mean 0.14, SD 0.17 vs mean 0.11, SD 0.13 logarithm of the minimum angle of resolution; *P*=.13), and complications were rare (iritis 2/98, 2% vs 1/87,1%; cystoid macular edema 1/98, 1% vs 1/87, 1%). Self-reported medication adherence favored printed instruction users (66/68, 97% vs 47/64, 73%; *P*<.001), whereas objective antibiotic-drop adherence favored app users (mean 5.36, SD 1.17 g vs mean 5.67, SD 1.00 g; *P*=.046).

Conclusions: The Sharp Health Companion app reduced same-day surgery delays and improved patient experience while supporting objective medication adherence in predominantly older adults undergoing cataract surgery. These findings suggest mobile health interventions can enhance perioperative care and efficiency, even in populations less familiar with technology.

Trial Registration: ClinicalTrials.gov NCT07028359; https://clinicaltrials.gov/ct2/show/NCT07028359

(JMIR Mhealth Uhealth 2026;14:e78710) doi:10.2196/78710

XSL•FO
RenderX

KEYWORDS

CareKit; cataract surgery; digital health; medication adherence; mHealth; mobile health; older adults; patient education; perioperative care; surgical outcomes; telemedicine

Introduction

Background

Cataracts are the leading cause of preventable blindness worldwide, affecting millions annually, significantly impacting vision, and placing a considerable strain on health care systems as the global population ages. Cataract surgery, the most commonly performed surgical procedure globally, restores eyesight and significantly improves quality of life, with over 30 million surgeries performed annually [1-3]. At Sharp HealthCare in San Diego, California, over 5000 cataract surgeries are conducted annually. However, rising global demand due to an aging population, combined with surgical backlogs exacerbated by the COVID-19 pandemic, has created substantial barriers to timely care, negatively impacting health care accessibility and patient morbidity [4-6].

Delays in surgery extend beyond administrative issues. Prolonged wait times can result in denser, more complicated cataracts that necessitate longer and technically challenging surgeries, increasing the risk of complications and resource burden [3,7]. Patient adherence to perioperative instructions, both preoperative and postoperative, is vital for optimal outcomes. Missed, forgotten, or poorly understood instructions such as dietary restrictions, medication schedules, self-quarantine requirements, and postoperative care frequently result in surgery delays and compromised surgical outcomes [8,9]. Cancellation rates increased notably during the COVID-19 pandemic, posing financial challenges to health care systems and negatively affecting patient experiences [10-12]. Adherence to postoperative medication regimens, specifically eye drops, is often problematic, potentially leading to complications and suboptimal visual recovery [13-15]. Furthermore, patients undergoing cataract surgery frequently receive intravenous sedation, which may impair memory and hinder recall of postoperative care instructions provided in the recovery room, a challenge further exacerbated by isolation during the COVID-19 pandemic [16,17].

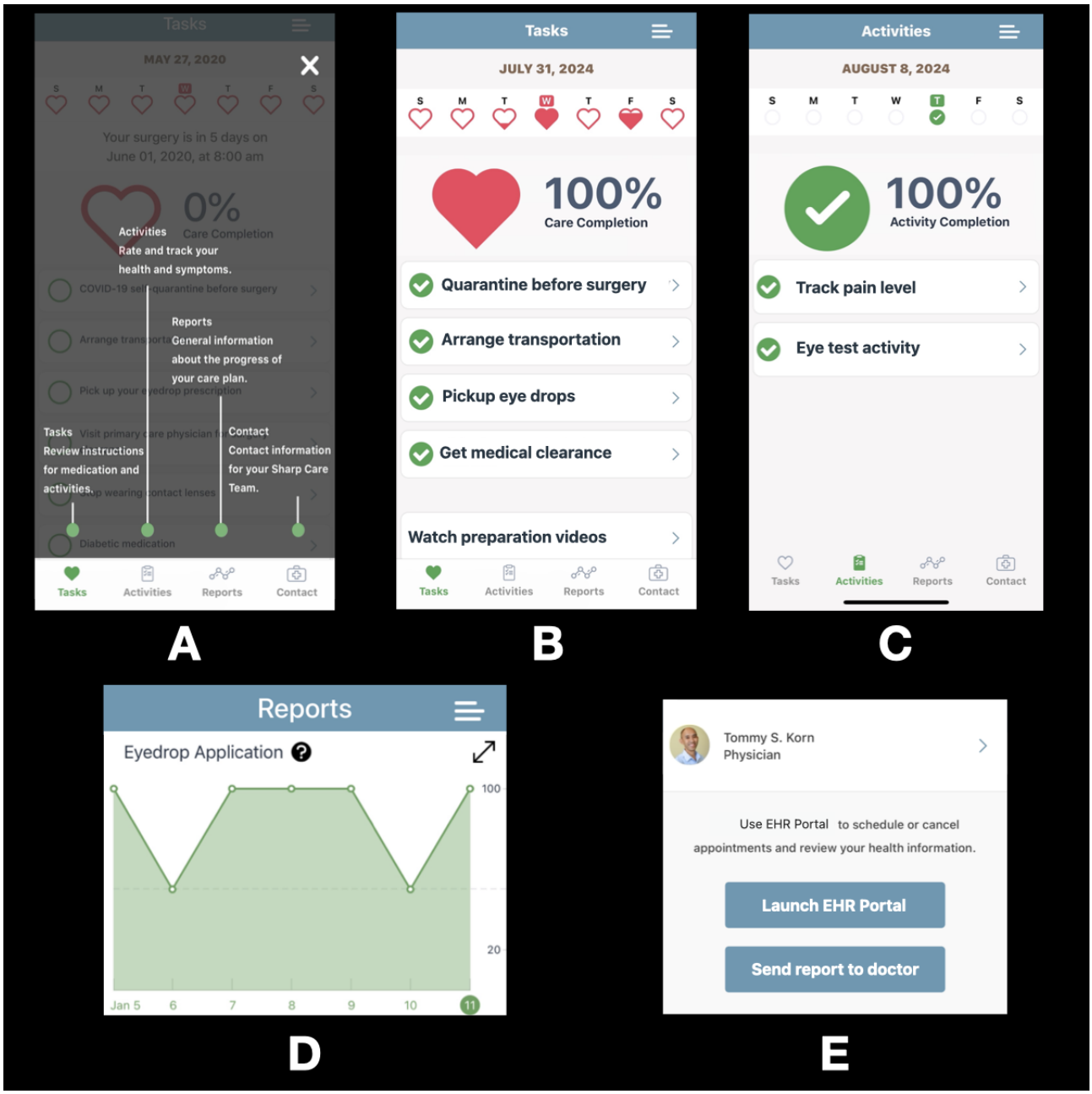
Historically, printed instructions have guided patients through perioperative processes; however, studies indicate inconsistent adherence, with instructions frequently lost or misunderstood, leading to cancellations and adverse outcomes [18,19]. This issue intensified during the COVID-19 pandemic, highlighting significant limitations in the flexibility and responsiveness of printed materials when protocols rapidly changed [20]. Studies have also identified biases among health care providers, who may question older adults' ability to effectively use digital technology, emphasizing the importance of addressing these perceptions [21]. These barriers underscore the need to better understand the role of mobile digital health interventions to consistently support patients throughout their surgical journey.

Related Work

Recognizing the emerging trend toward increasing mobile technology adoption among seniors, Sharp HealthCare proactively explored digital health solutions prior to the COVID-19 pandemic [22]. Introduced in 2016, CareKit is an open-source platform facilitating rapid, cost-effective development of health care–focused mobile apps designed to empower patients in managing their health under clinical guidance [23]. Use of an open-source digital health platform provides critical flexibility for health care systems by creating scalable solutions that effectively address the diverse surgical needs of an ever-increasing geriatric population. Leveraging this opportunity, a multidisciplinary Sharp HealthCare team (ophthalmologists, nurses, developers, and designers) produced and piloted the Sharp Health Companion app in 2017 to support older cataract surgery patients (Figure 1 and Multimedia Appendix 1).

XSL•FO
RenderX

Figure 1. Screenshots of the Sharp Health Companion app illustrating key perioperative features: (A) interactive task overview, (B) care task completion and medication adherence checklist, (C) postoperative symptom tracking (eye pain and visual acuity), (D) patient-reported medication adherence graph, and (E) integrated care team communication portal.



Strategically, the app was initially created as a stand-alone solution that was not integrated directly with the health system's multiple electronic health record (EHR) platforms, due to an ongoing organizational transition toward a unified EHR system at the time. This approach later provided valuable flexibility, enabling straightforward future integration with the unified EHR. Initial pilot results of the Sharp Health Companion app showed improved medication adherence and reduced surgery cancellations [24].

Objective

This study examined the role of digital health tools in helping the senior population with their preparation and recovery from cataract surgery. Consistent with systematic reviews indicating that the COVID-19 pandemic accelerated seniors' adoption of digital health tools and telehealth, this study examined the capability and willingness of older patients to engage with these technologies for remote care [25]. Supported by an administrative portal, the Sharp Health Companion app allowed care teams to instantly disseminate updated surgical care instructions and rapidly evolving COVID-19 safety protocols (self-quarantine and polymerase chain reaction testing reminders) to hundreds of patients simultaneously. This real-time communication and remote monitoring capability ensured the continuity of cataract surgeries at Sharp HealthCare even as elective procedures in other specialties faced significant disruptions.

Previous research has largely focused on digital educational interventions limited to preoperative preparation, lacking comprehensive assessment throughout the full perioperative cataract surgery journey, including the day of surgery and

XSL•FO
RenderX

postoperative recovery phases [26,27]. The aim of this randomized controlled trial was to evaluate whether the Sharp Health Companion smartphone app could improve medication adherence and perioperative preparedness compared with traditional paper instructions among predominantly older adults undergoing cataract surgery.

Methods

Study Design and Participants

This was a prospective, randomized, masked comparative study conducted over 18 months with the objective of evaluating the effectiveness of a digital health intervention in enhancing perioperative adherence, reducing surgery cancellations, and improving patient experience and outcomes in cataract surgery. Participants were recruited from a high-volume, 2-surgeon cataract ophthalmology practice in a metropolitan setting. This study was reported in accordance with the CONSORT (Consolidated Standards of Reporting Trials) guidelines (Multimedia Appendix 2). Eligible participants were English-speaking adults aged 18 years or older, deemed appropriate surgical candidates for cataract surgery (visual symptoms such as blurred vision and/or daytime or nighttime glare not improved with refraction and prescription eyeglasses or contact lenses, best-corrected visual acuity 20/40 or worse) by their ophthalmologists. Participants were required to verbally confirm basic technical proficiency with their iPhone during the preoperative consultation prior to randomization. Clinic staff informally assessed this by asking whether participants could operate their smartphone, use secure login (eg, Face ID, Touch ID, or passcode), send SMS text messages or email, and access the institution's EHR portal app. No demonstration was required on the smartphone, ensuring patient privacy and safety. This step ensured that participants had sufficient skills to independently use the study app. Eligibility criteria included participant capacity to understand study requirements, provide informed consent, and authorize the use of their protected health information. Exclusion criteria included visual, auditory, cognitive, or motor impairments that could prevent effective use of the app. Participants with a history of prior cataract surgery were also excluded, as their previous experience would confer an advantage in navigating the perioperative process, potentially biasing the study outcome. No formal a priori power calculation was conducted. The planned target enrollment was approximately 300 participants based on feasibility estimates; however, recruitment of smartphone-capable older adults was challenging.

Randomization

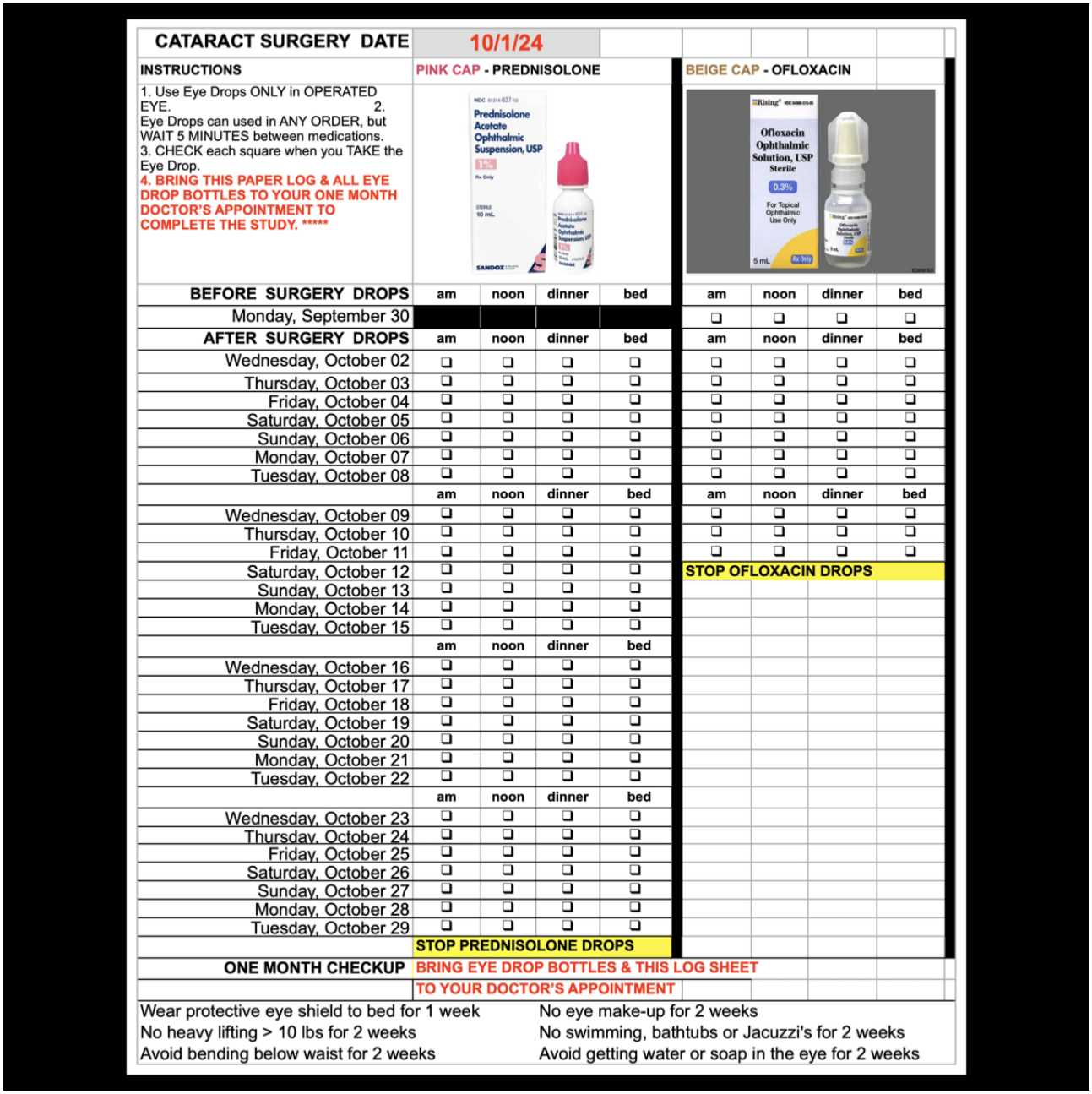
Eligible participants who elected to undergo cataract surgery and consented to study participation were randomized via a digital randomization app to either group 1 or group 2. Each participant received a unique study identifier to protect personal information and ensure confidentiality. The randomization assignments were masked from the principal investigators, ophthalmologists, and research team, including ophthalmic technicians, administrative staff, and surgery schedulers, who remained masked to group allocation throughout the study.

Interventions

Both groups 1 and 2 received the standard printed instructions, and group 2 additionally used the Sharp Health Companion app; the perioperative content was identical across groups, covering preoperative, day of surgery, and postoperative care. Group 1 participants received the standard printed instructions provided at the doctor's office (Multimedia Appendix 3), which included essential presurgery protocols and a paper-based medication adherence checklist for postoperative use (Figure 2). In addition, group 1 received a 1-time automated telephone reminder the day before surgery. The brief prerecorded message confirmed the scheduled surgery date, time, and check-in location and instructions; it did not include dietary restriction instructions.

XSL•FO
RenderX

Figure 2. Printed medication adherence checklist provided to group 1 participants using traditional printed instructions, used to track antibiotic and steroid eye drop usage before and after cataract surgery.

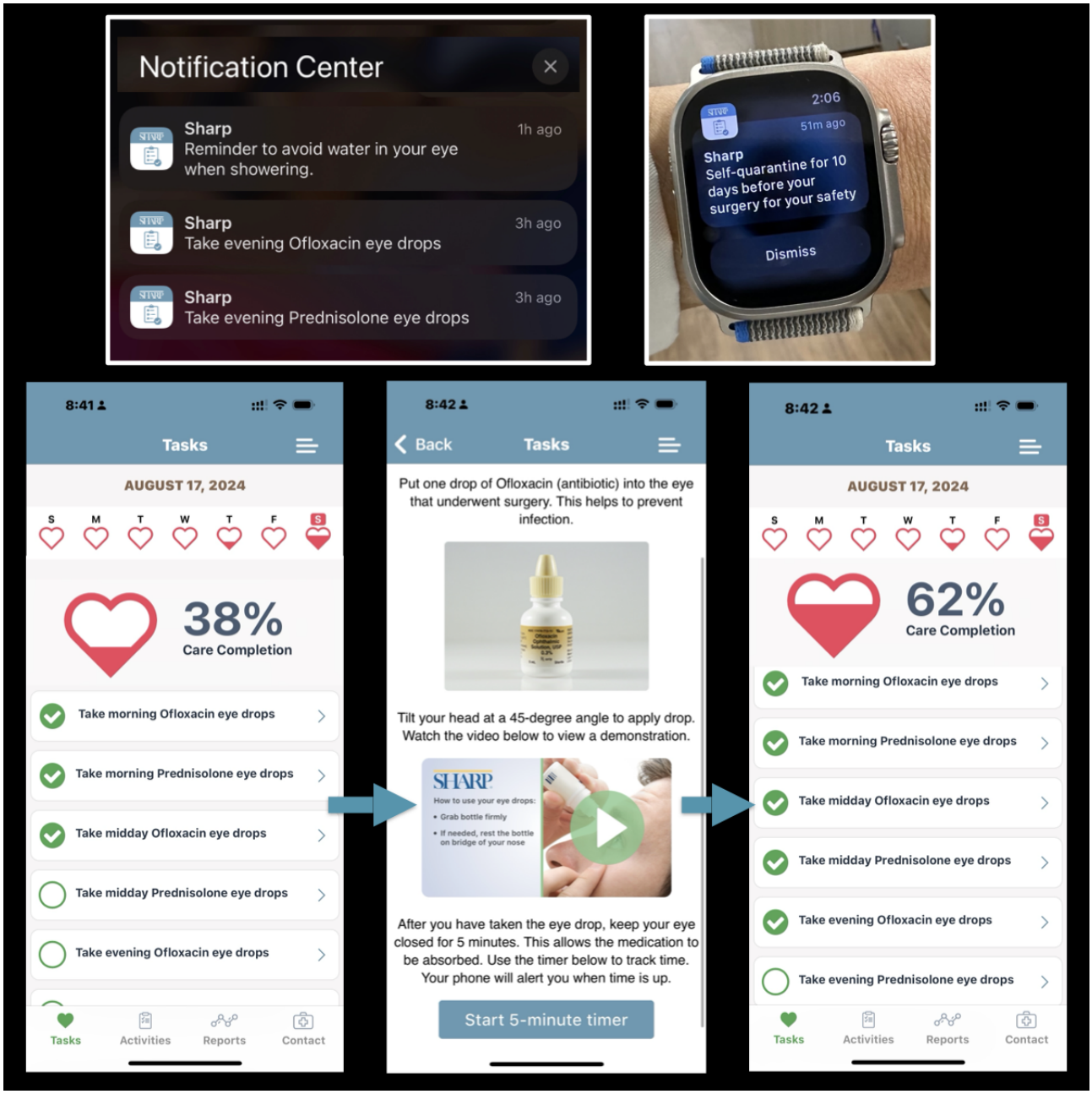


CATARACT SURGERY DATE	10/1/24							
INSTRUCTIONS	PINK CAP - PREDNISOLONE				BEIGE CAP - OFLOXACIN			
1. Use Eye Drops ONLY in OPERATED EYE. 2. Eye Drops can used in ANY ORDER, but WAIT 5 MINUTES between medications. 3. CHECK each square when you TAKE the Eye Drop. 4. BRING THIS PAPER LOG & ALL EYE DROP BOTTLES TO YOUR ONE MONTH DOCTOR'S APPOINTMENT TO COMPLETE THE STUDY. *****								
BEFORE SURGERY DROPS	am	noon	dinner	bed	am	noon	dinner	bed
Monday, September 30					❑	❑	❑	❑
AFTER SURGERY DROPS	am	noon	dinner	bed	am	noon	dinner	bed
Wednesday, October 02	❑	❑	❑	❑	❑	❑	❑	❑
Thursday, October 03	❑	❑	❑	❑	❑	❑	❑	❑
Friday, October 04	❑	❑	❑	❑	❑	❑	❑	❑
Saturday, October 05	❑	❑	❑	❑	❑	❑	❑	❑
Sunday, October 06	❑	❑	❑	❑	❑	❑	❑	❑
Monday, October 07	❑	❑	❑	❑	❑	❑	❑	❑
Tuesday, October 08	❑	❑	❑	❑	❑	❑	❑	❑
	am	noon	dinner	bed	am	noon	dinner	bed
Wednesday, October 09	❑	❑	❑	❑	❑	❑	❑	❑
Thursday, October 10	❑	❑	❑	❑	❑	❑	❑	❑
Friday, October 11	❑	❑	❑	❑	❑	❑	❑	❑
Saturday, October 12	❑	❑	❑	❑	STOP OFLOXACIN DROPS			
Sunday, October 13	❑	❑	❑	❑				
Monday, October 14	❑	❑	❑	❑				
Tuesday, October 15	❑	❑	❑	❑				
	am	noon	dinner	bed				
Wednesday, October 16	❑	❑	❑	❑				
Thursday, October 17	❑	❑	❑	❑				
Friday, October 18	❑	❑	❑	❑				
Saturday, October 19	❑	❑	❑	❑				
Sunday, October 20	❑	❑	❑	❑				
Monday, October 21	❑	❑	❑	❑				
Tuesday, October 22	❑	❑	❑	❑				
	am	noon	dinner	bed				
Wednesday, October 23	❑	❑	❑	❑				
Thursday, October 24	❑	❑	❑	❑				
Friday, October 25	❑	❑	❑	❑				
Saturday, October 26	❑	❑	❑	❑				
Sunday, October 27	❑	❑	❑	❑				
Monday, October 28	❑	❑	❑	❑				
Tuesday, October 29	❑	❑	❑	❑				
	STOP PREDNISOLONE DROPS							
ONE MONTH CHECKUP	BRING EYE DROP BOTTLES & THIS LOG SHEET TO YOUR DOCTOR'S APPOINTMENT							
Wear protective eye shield to bed for 1 week					No eye make-up for 2 weeks			
No heavy lifting > 10 lbs for 2 weeks					No swimming, bathtubs or Jacuzzi's for 2 weeks			
Avoid bending below waist for 2 weeks					Avoid getting water or soap in the eye for 2 weeks			

Group 2 participants used the Sharp Health Companion app, developed on the CareKit platform, which provided digital notifications for preoperative requirements such as self-quarantine and polymerase chain reaction testing (during the COVID-19 pandemic), dietary restrictions, surgery appointment details (date, time, and location), and medication reminders. The app featured interactive care checklists, educational videos, and timely eye drop medication adherence reminders delivered via iPhone and/or Apple Watch notifications (Figure 3 and Multimedia Appendix 1). Additionally, the app offered tools for visual acuity self-monitoring, easy access to care team contact information, and postoperative care instructional videos (Figure 1). As a precaution, group 2 also received the same printed instructions (Multimedia Appendix 3) as group 1, serving as a backup in the event of app malfunction or patient technical difficulties. Unlike group 1, participants in group 2 did not receive any automated telephone reminders or a printed medication adherence checklist; the app served as their primary tool for perioperative support, supplemented only by backup printed instructions.

XSL•FO
RenderX

Figure 3. Sharp Health Companion app workflow: smartphone and smartwatch notifications prompt patients to administer eye drops, mark adherence tasks as complete, review instructional videos, and use an integrated 5-minute timer, tracking daily progress toward 100%.



Both groups received identical, standardized eye drop medications at no cost: 1 bottle of prednisolone acetate 1% (10 mL; Sandoz Pharmaceuticals) and 1 bottle of ofloxacin 0.3% (5 mL; Rising Pharmaceuticals), dispensed directly by the institution's pharmacy to ensure consistency.

Data Collection

Participant experience was assessed using an adapted version of the Consumer Assessment of Healthcare Providers and Systems-2.0 survey, evaluating clarity of information, preparedness, and overall surgical satisfaction (Multimedia Appendix 4). At their initial cataract consultation, eligible participants from both groups completed the survey to report demographic and technology use characteristics prior to undergoing surgery. Subjective data were collected at three time points: (1) the baseline survey (preoperative office consultation), (2) the postoperative day 1 survey, and (3) the postoperative month 1 survey (Multimedia Appendix 4). Patient confidentiality was ensured by administering surveys on a HIPAA (Health Insurance Portability and Accountability Act)-compliant, enterprise-grade SurveyMonkey platform accessed via secure health system iPads configured in kiosk mode. Participants entered only their assigned study identifier number for the surveys; no personally identifiable information was collected.

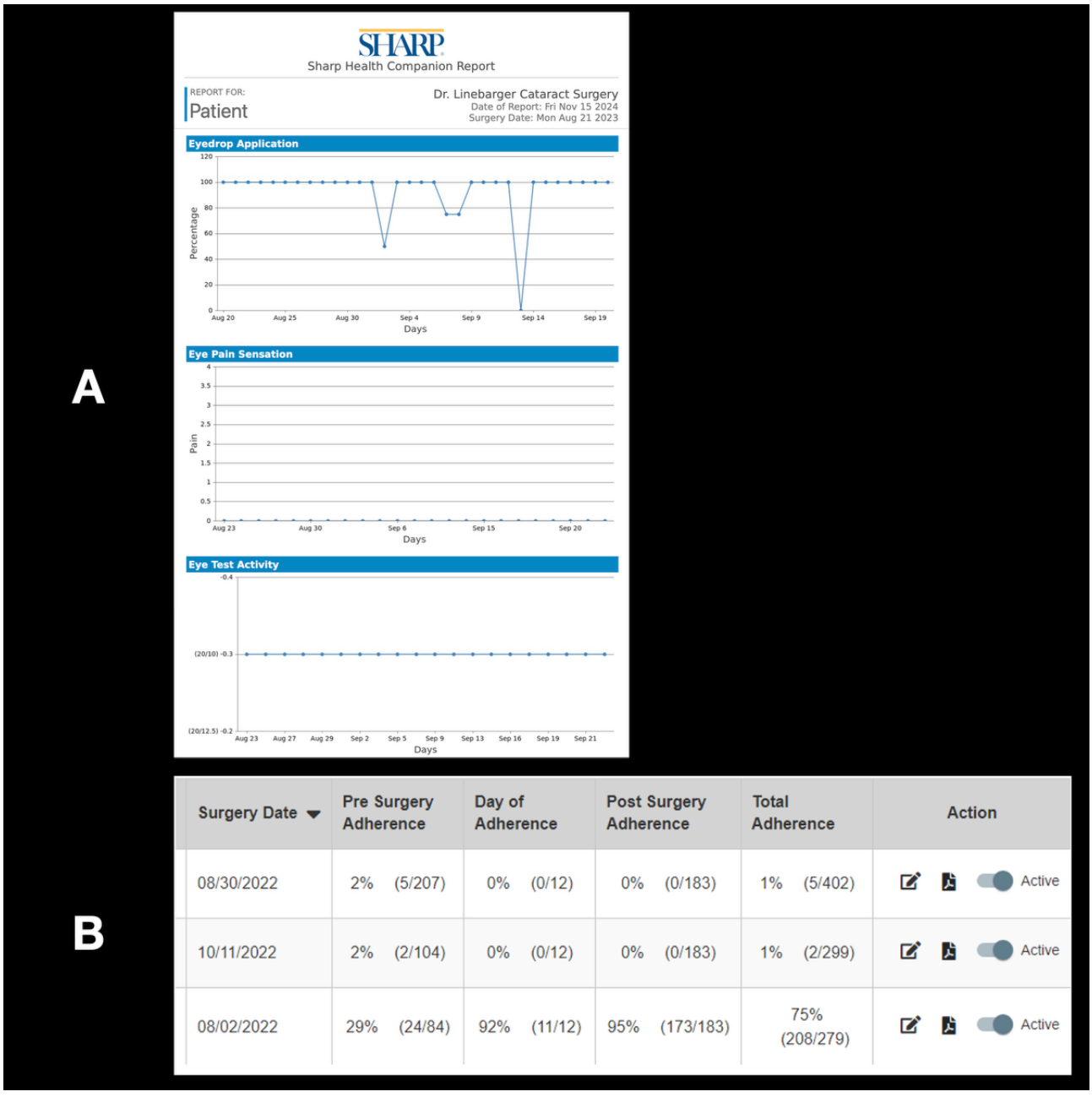
Objective medication adherence was determined by weighing antibiotic and steroid eye drop bottles returned at the 30-day postoperative visit. Returned bottles were weighed using an analytical balance (Apollo GX-603A; A&D Weighing), accurate to ±0.001 g, with all exterior packaging removed. To establish a baseline, the average weight of 5 unopened bottles of each medication was recorded: prednisolone acetate 1% (10 mL) weighed 17.223 g, and ofloxacin 0.3% (5 mL) weighed 8.466

XSL•FO
RenderX

g. Each participant was prescribed exactly 1 bottle of each medication for the postoperative course, and all medications were obtained from the same manufacturer to ensure consistency. Bottle weights at 30 days were compared to these baselines to determine the amount of medication consumed. All participants from both groups were instructed to return their bottles to precisely quantify medication usage.

In addition, group 1 participants were asked to return their printed medication adherence checklists at the final visit (Figure 2). Group 2 adherence was monitored through a HIPAA-secure, encrypted back-end portal that restricted retrospective reporting beyond 24 hours, ensuring an accurate reflection of daily medication adherence (Figure 4).

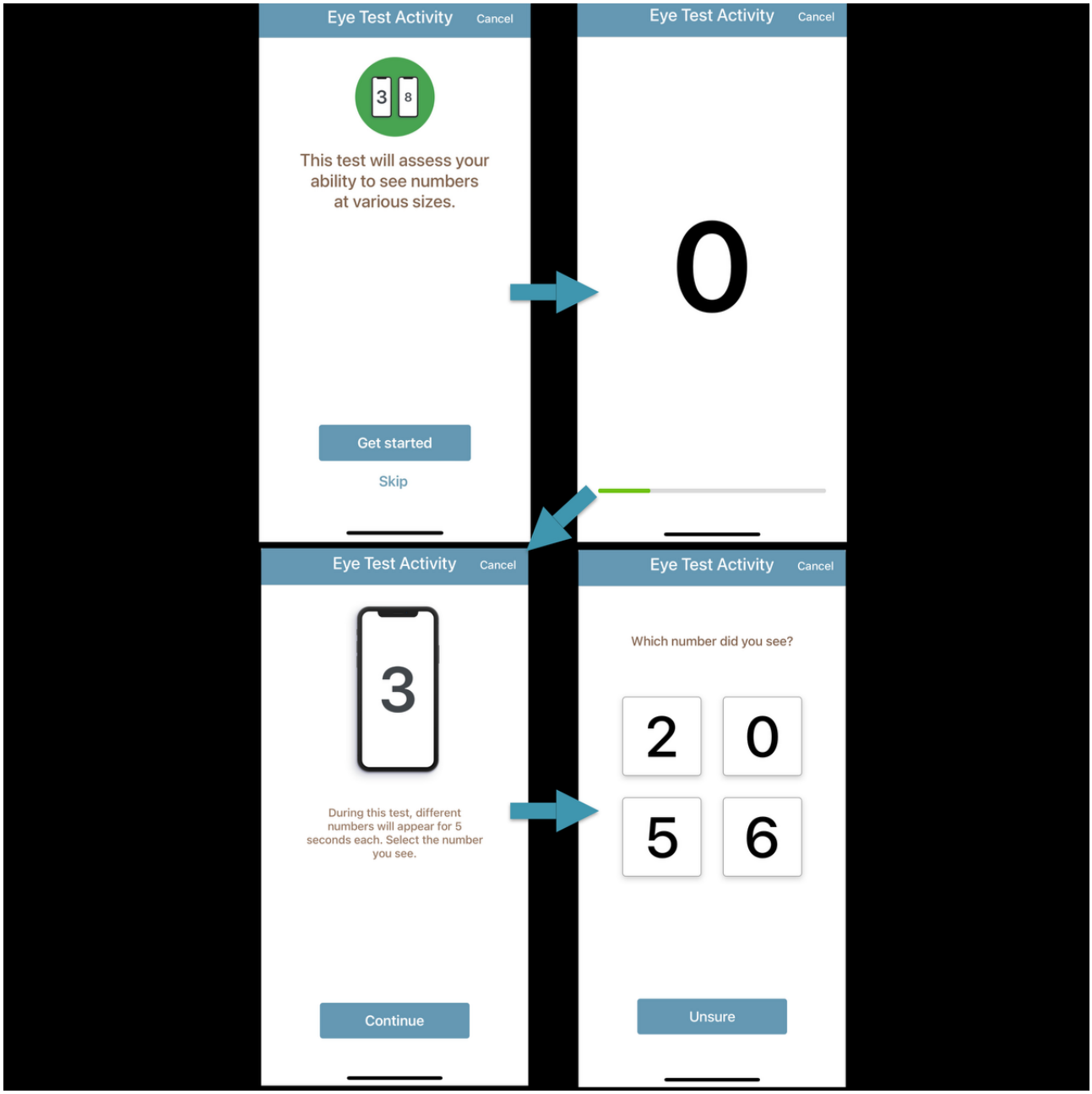
Figure 4. (A) Patient-generated report from the Sharp Health Companion app, showing self-reported medication adherence (eye drops), pain levels, and visual acuity assessments. Patients could review this summary and securely share it with their care team before and after cataract surgery. (B) Care team view from the HIPAA (Health Insurance Portability and Accountability Act)–secure back-end portal, displaying objective eye drop adherence data linked to each participant's study identifier, enabling clinicians to monitor adherence trends and provide timely support.



Visual acuity was assessed by ophthalmic technicians using an Early Treatment Diabetic Retinopathy Study chart on the logarithm of the minimum angle of resolution (logMAR) scale. Measurements were recorded preoperatively and at 1 day and 30 days postoperatively, allowing comparison of surgical effectiveness between groups. Additionally, group 2 participants had optional access to a visual acuity self-assessment activity within the Sharp Health Companion app, enabling independent self-monitoring of vision recovery following cataract surgery (Figure 5). The care team could remotely access and review data from this vision activity throughout the postoperative period if participants chose to use this optional feature.

XSL•FO
RenderX

Figure 5. Sharp Health Companion app's integrated vision self-assessment workflow, guiding patients through instructions and sequential vision testing. The ability to correctly identify smaller numbers indicates better visual acuity (eye performance).



Outcomes and Measurements

The primary outcomes included patient satisfaction with the care experience, surgery delays and cancellations, and subjective and objective medication adherence. A surgery delay was prespecified as a postponement of >2 hours with the procedure still completed on the scheduled day; a surgery cancellation was defined as a same-day procedure not performed. Reasons were abstracted from the surgery schedule and chart notes by masked staff at the conclusion of the study.

Subjective medication adherence was defined as the number of prescribed eye drop doses marked as completed divided by the total number of prescribed doses. For both groups, antibiotic drops were initiated 1 day before surgery and administered 4 times daily (4 doses preoperatively), then continued at the same frequency for 10 days postoperatively (40 doses in total). Steroid drops were initiated on the day of surgery and taken 4 times daily for 30 days (120 total doses). Group 1 participants manually checked off each completed dose on printed checklists (Figure 2), while group 2 participants recorded each administered dose digitally within the Sharp Health Companion app (Figures 3 and 4B). The app required daily reporting and did not allow retrospective entries beyond 24 hours, capturing real-time adherence data.

Study completion was a composite outcome that included cataract surgery performed, attendance at all scheduled postoperative visits, completion of all 3 surveys, and return of both antibiotic and steroid bottles at 30 days. In addition, group 1 participants were required to return the printed medication adherence checklist, while group 2 participants' medication adherence was verified through the Sharp Health Companion app portal monitored by the research team (Figure 4). Secondary

XSL•FO
RenderX

outcomes encompassed postoperative visual acuity and postsurgical complications, such as cystoid macular edema and uveitis at the final 30-day appointment after cataract surgery. Postoperative iritis was determined by slit-lamp examination, documenting anterior chamber cells/flare requiring treatment by the ophthalmologist. Cystoid macular edema was diagnosed clinically and confirmed with optical coherence tomography when indicated. The likelihood of recommending the health care organization item was measured on a 1-10 scale (1=not at all likely to 10=extremely likely).

Statistical Analysis

Statistical analyses were conducted using SPSS Statistics for Mac (version 31.0; IBM Corp). Continuous outcomes were compared with independent *t* tests, and categorical outcomes with chi-square tests. All analyses adhered to the intention-to-treat principle, with participants analyzed in the groups to which they were randomized, regardless of adherence or protocol deviations. Statistical significance was set at $P<.05$.

Ethical Considerations

This study was approved by the Sharp HealthCare Institutional Review Board (approval number 2209803) and adhered to institutional guidelines and the principles of the Declaration of Helsinki. All participants provided informed consent via DocuSign and received a digital and/or printed copy of the California Experimental Subject's Bill of Rights. Participants were informed of their ability to withdraw at any time during the study without affecting their clinical care. To protect privacy, participants were assigned unique identifiers, surveys were administered via HIPAA-compliant software (SurveyMonkey Enterprise, kiosk mode), and app data were encrypted on secure servers. Participants did not receive financial compensation. Prescribed eye medications (prednisolone acetate 1% and ofloxacin 0.3%) were provided at no cost as part of study participation.

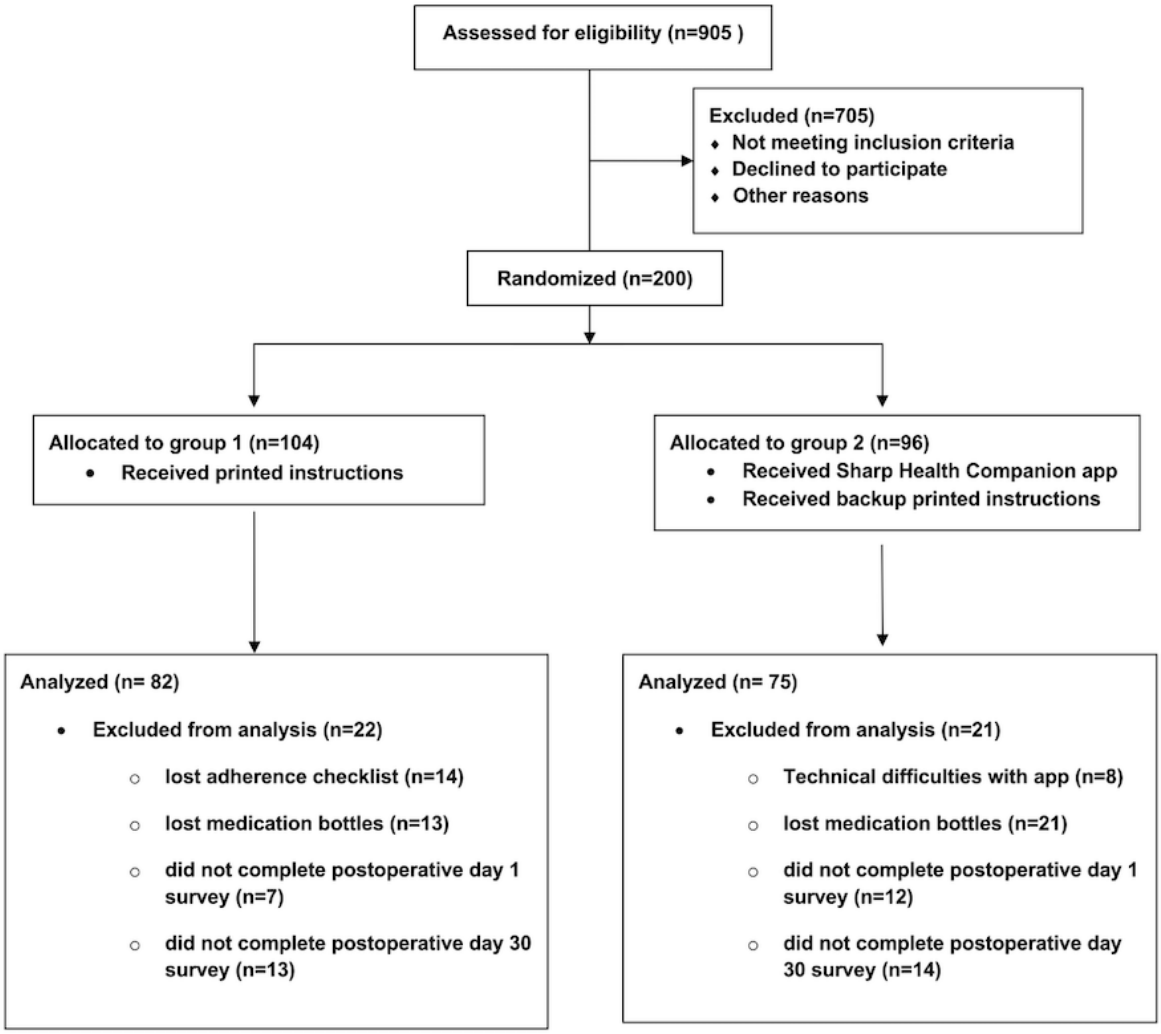
Results

Participant Demographics

Between December 2022 and January 2024, a total of 905 patients were referred for cataract surgery, and 200 were enrolled and randomized into 2 groups: group 1 (printed instruction participants; n=104) and group 2 (Sharp Health Companion app participants; n=96; Figure 6). Participants ranged in age from 41 to 87 years, with an overall mean age of 69 (SD 8.2) years. The app group was slightly younger than the printed instruction group (mean 67.1, SD 8.7 years vs mean 69.9, SD 7.8 years; $P=.02$). Most participants were aged ≥65 years (145/200, 73%), with no difference between groups ($P=.25$). The majority were female (140/200, 70%), and gender distribution did not differ between groups ($P=.50$). Racial and ethnic distributions were also comparable ($P=.39$). Nearly all participants reported Wi-Fi access (185/200, 93%), slightly higher in the printed instruction group (100/104, 96% vs 85/96, 89%; $P=.04$). Comfort with smartphone apps (1=least comfortable to 5=most comfortable) was higher among app participants (mean 3.67, SD 0.68 vs mean 3.36, SD 0.82; $P<.001$). Smartwatch ownership was more common among app users (55/96, 57% vs 38/104, 37%; $P=.003$). Prior experience with smartphone EHR portal apps was similar between groups ($P=.15$; Table 1).

XSL•FO
RenderX

Figure 6. CONSORT (Consolidated Standards of Reporting Trials) flow diagram. Flow of participants through the randomized controlled trial, including numbers assessed for eligibility, excluded, randomized, allocated to group 1 (printed instructions) or group 2 (Sharp Health Companion app), and included in the final analyses. Some participants met more than 1 exclusion criterion (eg, lost both the adherence checklist and medication bottles or experienced technical difficulties and lost bottles).



XSL•FO
RenderX

Table 1. Demographics of study participants (n=200).

Characteristic	Group 1 (printed instructions; n=104)	Group 2 (Sharp Health Companion app; n=96)	*P* value
Age (years), mean (SD)	69.9 (7.8)	67.1 (8.7)	.02[a,b]
Age (≥65 years), n (%)	79 (76)	66 (69)	.25[c]
Gender, n (%)			.50[c]
Male	29 (28)	31 (32)	
Female	75 (72)	65 (68)	
Ethnicity, n (%)			.39[c]
White	78 (75)	71 (74)	
Hispanic/Latino	13 (12)	7 (7)	
Asian or Pacific Islander	8 (8)	12 (13)	
Black or African American	2 (2)	4 (4)	
American Indian or Alaskan Native	0 (0)	1 (1)	
Other	2 (2)	0 (0)	
Prefer not to answer	1 (1)	1 (1)	
Primary home internet connection, n (%)			
Wi-Fi	100 (96)	85 (89)	.04[b,c]
Mobile cellular plan	4 (4)	11 (11)	.07[d]
Comfort with smartphone apps (1=least comfortable to 5=most comfortable), mean (SD)	3.36 (0.82)	3.67 (0.68)	<.001[a,b]
Prior use of smartphone medical apps (eg, EHR[e] portal apps), n (%)	78 (75)	80 (83)	.15[c]
Uses smartwatch (Apple Watch, Fitbit, etc), n (%)	38 (37)	55 (57)	.003[b,c]

[a]Independent *t* test (2-tailed).

[b]Statistically significant (*P*<.05).

[c]Chi-square test.

[d]Fisher exact test.

[e]EHR: electronic health record.

Preparedness and Clarity of Instructions: Postoperative Day 1 Survey

Participants completed an in-office survey 1 day after cataract surgery to assess preparedness, clarity, and usability of preoperative instructions (Table 2 and Multimedia Appendix 4). Most participants reviewed their instructions within 2-7 days before surgery (100/181, 55%), with no significant difference in timing between groups (*P*=.21). Among app participants, 56% (47/84) used both the Sharp Health Companion app and backup printed instructions before surgery.

XSL•FO
RenderX

Table 2. Patient-reported preparedness and clarity of preoperative instructions (postoperative day 1 survey; n=181).

Preparedness for surgery	Group 1 (printed instructions; n=97)[a]	Group 2 (Sharp Health Companion app; n=84)[a]	*P* value
Q10: Before your surgery, what did you use to help prepare for surgery, n (%)			—[b]
Printed instructions	97 (100)	10 (12)	
App	N/A[c]	27 (32)	
App + printed instructions backup	N/A	47 (56)	
Q11: When did you review your information (paper instructions and/or Sharp Health Companion app) before surgery, n (%)			.21[d]
Over 31 days before surgery	6 (6)	5 (6)	
8-30 days before surgery	16 (15)	26 (31)	
2-7 days before surgery	59 (57)	41 (49)	
1 day before surgery	13 (13)	9 (11)	
On the day of surgery	2 (2)	3 (4)	
Q12: Before your surgery, did anyone in this surgeon's office give you all the information you needed about your eye surgery, n (%)			.63[d]
Yes, definitely	88 (92)	80 (95)	
Yes, somewhat	6 (6)	3 (4)	
No	2 (2)	1 (1)	
Q13: Before your surgery, did anyone in this surgeon's office give you easy to understand instructions about getting ready for your eye surgery, n (%)			.02[d,e]
Yes, definitely	86 (90)	83 (99)	
Yes, somewhat	8 (8)	0 (0)	
No	2 (2)	1 (1)	
Q14: Before your surgery, on a scale from 1 to 10, how would you rate your preparedness for surgery given the information you received from the surgeon's office (1-10 scale, 1=least prepared, 10=most prepared), mean (SD)	9.56 (1.19)	9.77 (0.73)	.16[f]
Q15: How would you rate your surgery experience so far with the information you've been provided (1-10 scale, 1=least favorable, 10=most favorable), mean (SD)	9.92 (0.37)	9.96 (0.24)	.32[f]
Q16: If you used the Sharp Health Companion app, did you have to use the backup paper instructions at any time, n (%)			—
Yes, definitely used the paper instructions	N/A	16 (19)	
Yes, somewhat used the paper instructions	N/A	37 (44)	
No, I did not have to use my backup paper instructions	N/A	31 (37)	
Q17: How does the Sharp Health Companion app compare to other medical apps you have used, n (%)			—
Easier and more valuable	N/A	44 (54)	
About as easy and valuable	N/A	37 (46)	
Not as easy or valuable	N/A	0 (0)	
Q18: I would use the Sharp Health Companion app for other surgeries or medical procedures if available, n (%)			—
Yes, definitely	N/A	68 (87)	
Yes, somewhat	N/A	10 (13)	
No	N/A	0 (0)	
Patients who initiated communication (phone calls or emails) with staff regarding surgery logistics, n (%)	9 (9)	6 (7)	.62[d]

XSL•FO
RenderX

[a]Sample sizes differ from the total group because not all participants completed this survey item.

[b]Not available.

[c]N/A indicates response options that were not applicable to a group (eg, group 1 did not have access to the Sharp Health Companion app).

[d]Pearson chi-square test.

[e]Statistically significant (P<.05).

[f]Independent t test.

Regarding clarity of instructions, response distributions ("yes, definitely," "yes, somewhat," or "no") did not differ significantly between groups (P=.63). Most participants rated their instructions as "definitely easy to understand," with slightly higher ratings among app users compared with printed instruction users (83/84, 99% vs 86/96, 90%; P=.02). Preparedness ratings on a 1-10 scale (1=least prepared to 10=most prepared) were high and similar in both groups (mean 9.56, SD 1.19 vs mean 9.77, SD 0.73). Surgery experience ratings were likewise high (P=.32).

Among app users, 63% (53/84) reported using backup printed instructions to some extent, while the minority 37% (31/84) did not need them. When comparing the Sharp Health Companion app to other medical apps, 54% (44/84) rated it as easier and more valuable, and 46% (37/84) as about as easy and valuable; none rated it as less valuable. Most app participants (68/84, 87%) indicated they would "definitely" use the Sharp Health Companion app for future medical or surgical procedures, and the remainder (10/84, 13%) responded, "yes, somewhat." Finally, a small number of participants contacted the surgeon's office for clarification about surgery logistics, with no significant difference between groups (9/97, 9% vs 6/84, 7%; P=.62).

Cataract Surgery Delays and Cancellations

Surgery delays were significantly more frequent among participants receiving printed instructions than those using the Sharp Health Companion app (10/104, 10% vs 1/96, 1%; P=.01). Delays were defined as procedures postponed by >2 hours but completed on the scheduled day, most commonly due to participants arriving at the incorrect surgery location or time. Surgery cancellations occurred in 15 participants overall, with similar proportions in the printed instruction and app groups (6/104, 6% vs 9/96, 9%; P=.33). The primary reasons included acute illness, participant-initiated withdrawal, and financial or insurance issues (Table 3).

Table 3. Cataract surgery delays and cancellations (n=200).

Outcome measure	Group 1 (printed instructions; n=104), n (%)	Group 2 (Sharp Health Companion app; n=96), n (%)	P value
Delayed surgery	10 (10)	1 (1)	.01[a,b]
Missed dietary restrictions	2 (2)	0 (0)	
Wrong location/time	6 (6)	1 (1)	
Other reasons	2 (2)	0 (0)	
Canceled surgery	6 (6)	9 (9)	.33[c]
Illness	3 (3)	2 (2)	
Missed dietary restrictions	1 (1)	0 (0)	
Wrong location/time	1 (1)	0 (0)	
Self-canceled	1 (1)	5 (5)	
Insurance/financial reasons	0 (0)	1 (1)	
Not a candidate	0 (0)	1 (1)	

[a]Statistically significant (P<.05).

[b]Fisher exact test.

[c]Chi-square test.

Study Completion and Adherence to Protocol

Study completion rates were high and comparable between groups (82/104, 79% vs 75/96, 78%). Among app users, engagement with specific features varied. Most participants used both the Sharp Health Companion app and backup printed instructions (42/75, 56%), while smaller subsets used the app alone (24/75, 32%) or printed instructions only (9/75, 12%). About half completed the vision self-assessment (38/75, 51%) and eye pain self-assessment (37/75, 49%) activities within the app. Loss of medication bottles was slightly higher in the app group (21/75, 28%) compared to the printed instruction group (13/82, 16%) but the difference was not statistically significant (P=.07). Some participants (8/75, 11%) in the app group reported technical issues, such as forgotten login credentials or intermittent software errors that interfered with adherence tracking. Overall, adherence and study completion were comparable between groups, indicating that the addition of the Sharp Health Companion app did not adversely affect protocol compliance (Table 4).

XSL•FO
RenderX

Table 4. Study completion and adherence to protocol (n=200).

Study completion details	Group 1 (printed instructions; n=104), n (%)	Group 2 (Sharp Health Companion app; n=96), n (%)	*P* value
Participants who fully completed the study	82 (79)	75 (78)	.90[a]
Used backup printed instructions only	N/A[b]	9 (12)	
Used the Sharp Health Companion app only	N/A	24 (32)	
Used the Sharp Health Companion app and backup printed instructions	N/A	42 (56)	
Used the vision self-assessment test activity in the app	N/A	38 (51)	
Used the eye pain self-assessment activity in the app	N/A	37 (49)	
Participants with incomplete data who were excluded from the analysis	22 (27)	21 (28)	
Lost printed medication adherence checklist (Figure 2)	14 (17)	N/A	.90[a]
Lost medication bottles	13 (16)	21(28)	.07[a]
Technical difficulties with the app	N/A	8 (11)	.90[a]

[a]Chi-square test.

[b]N/A indicates response options that were not applicable to a group (eg, group 1 did not have access to the Sharp Health Companion app).

Preparedness and Clarity of Instructions: Postoperative Month 1 Survey

At the month 1 follow-up visit, nearly all participants in both groups reported receiving complete and clear preoperative information (Table 5 and Multimedia Appendix 4). Almost all participants indicated that someone in the surgeon's office provided all the information they needed about their eye surgery, and nearly all rated the preoperative instructions as easy to understand. Mean preparedness ratings before and after surgery were uniformly high and comparable between groups, and no statistically significant differences were noted. All participants in both groups reported clear recovery instructions, and overall surgery experience ratings remained similar. Among app users, the majority (49/82, 60%) relied solely on the Sharp Health Companion app, while the remainder also used backup printed instructions. Most app users rated the app as easier or more valuable than other medical apps (49/82, 60%) and nearly all indicated they would use a similar app again for other surgeries or medical procedures (78/82, 95%). Participants from both groups rated their likelihood of recommending Sharp HealthCare to a friend or colleague based on their experience highly (mean 9.83, SD 0.59 vs mean 9.74, SD 0.91), with no significant difference between groups (*P*=.43).

XSL•FO
RenderX

Table 5. Patient-reported preparedness and clarity of preoperative instructions: postoperative month 1 survey (n=174).

Outcome measure	Group 1 (printed instructions; n=92)[a]	Group 2 (Sharp Health Companion app; n=82)[a]	*P* value
Q20: Before your surgery, did anyone in this surgeon's office give you all the information you needed about your eye surgery, n (%)			.94[b]
Yes, definitely	90 (99)	81 (99)	
Yes, somewhat	1 (1)	1 (1)	
No	0 (0)	0 (0)	
Q21: Before your surgery, did anyone in this surgeon's office give you easy to understand instructions about getting ready for your eye surgery, n (%)			.34[b]
Yes, definitely	91 (99)	82 (100)	
Yes, somewhat	1 (1)	0 (0)	
No	0 (0)	0 (0)	
Q22: Before your surgery, on a scale from 1 to 10, how would you rate your preparedness for surgery given the information you received from the surgeon's office, (1-10 scale: 1=least prepared to 10=most prepared), mean (SD)	9.86 (0.38)	9.88 (0.40)	.74[c]
Q23: After your surgery, on a scale from 1 to 10, how would you rate your preparedness for surgery given the information you received from the surgeon's office, (1-10 scale: 1=least prepared to 10=most prepared), mean (SD)	9.92 (0.37)	9.85 (0.47)	.28[c]
Q24: Did anyone in this surgeon's office give you easy to understand instructions about what to do during your recovery period?, n (%)			—[d]
Yes, definitely	92 (100)	82 (100)	
Yes, somewhat	0 (0)	0 (0)	
No	0 (0)	0 (0)	
Q25: How would you rate your overall surgery experience from preparation to recovery using the information you were provided (1-10 scale: 1=least favorable to 10=most favorable), mean (SD)	9.92 (0.45)	9.78 (1.04)	.23[c]
Q26: If you used the Sharp Health Companion app, did you have to use the backup paper instructions at any time, n (%)			—
Yes, definitely used the paper instructions	N/A[e]	6 (7)	
Yes, somewhat used the paper instructions	N/A	27 (33)	
No, I did not have to use my backup paper instructions	N/A	49 (60)	
Q27: How does the Sharp Health Companion app compare to other medical apps you have used, n (%)			—
Easier and more valuable	N/A	49 (60)	
About as easy and valuable	N/A	27 (33)	
Not as easy or valuable	N/A	3 (4)	
Q28: I would use the Sharp Health Companion app for other surgeries or medical procedures if available, n (%)			—
Yes, definitely	N/A	72 (88)	
Yes, somewhat	N/A	6 (7)	
No	N/A	1 (1)	
Q29: How likely is it that you would recommend Sharp HealthCare to a friend or colleague based on your experience (1-10 scale: 1=no at all likely to 10=extremely likely), mean (SD)	9.83 (0.59)	9.74 (0.91)	.43[c]

[a]Sample sizes differ from the total group n because not all participants completed this survey item.

XSL•FO
RenderX

[b]Pearson chi-square test.

[c]Independent *t* test.

[d]Not available.

[e]N/A indicates response options that were not applicable to a group (eg, group 1 did not have access to the Sharp Health Companion app).

Clinical Outcomes Following Cataract Surgery: Month 1 Visit

Clinical outcome data were available for 185 participants, including 98 in the printed instruction group and 87 in the app group. This accounted for 15 participants whose surgeries were canceled (6 in the printed instruction group and 9 in the app group). At the month 1 postoperative visit, mean visual acuity (logMAR) was comparable between the printed instruction and app groups (mean 0.14, SD 0.17 vs mean 0.11, SD 0.13; *P*=.13; Table 6). Postoperative inflammation and macular edema were uncommon, with no significant differences noted. Iritis occurred in 2/98 (2%) in the printed instruction group and 1/87 (1%) in the app group, and cystoid macular edema occurred in 1/98 (1%) and 1/87 (1%), respectively. No other postoperative complications were identified.

Table 6. Clinical outcomes following cataract surgery: month 1 visit (n=185).

Outcome Measure	Group 1 (printed instructions; n=98)	Group 2 (Sharp Health Companion app; n=87)	*P* value
Visual acuity (ETDRS[a]; logMAR[b]), mean (SD)	0.14 (0.17)	0.11 (0.13)	.13[c]
Postoperative iritis incidence, n (%)	2 (2)	1 (1)	>.99[d]
Postoperative cystoid macular edema incidence, n (%)	1 (1)	1 (1)	>.99[d]

[a]ETDRS: Early Treatment Diabetic Retinopathy Study.

[b]logMAR: logarithm of the minimum angle of resolution.

[c]Independent *t* test.

[d]Fisher exact test.

Eye Medication Adherence: Patient-Reported vs Objective Measures

Medication adherence data were available for 132 participants, including 68 in the printed instruction group and 64 in the app group. This total reflects participants who returned their medication bottles and completed adherence documentation; others were excluded because printed checklists were misplaced or bottles were lost or not returned at the end of the study.

Self-reported medication adherence was significantly higher among participants using printed checklists than those using the app (66/68, 97% vs 47/64, 73%; *P*<.001; Table 7). Objective medication adherence, assessed by returned eye medication bottle weights, showed broadly similar findings between groups. Residual antibiotic bottle weights were higher in the printed instruction group (mean 5.67, SD 1.00 g vs mean 5.36, SD 1.17 g; *P*=.046), while steroid bottle weights were comparable (mean 11.80, SD 2.00 g vs mean 11.39, SD 1.95 g; *P*=.14).

Table 7. Eye medication adherence: self-reported vs objective measures (n=132).

Measure	Group 1 (printed instructions; n=68)	Group 2 (Sharp Health Companion app; n=64)	*P* value
Self-reported eye drop medication adherence, mean (%)[a]	97	73	<.001[b,c]
Final antibiotic bottle weight (g), mean (SD)	5.67 (1.00)	5.36 (1.17)	.05[a,b]
Final steroid bottle weight (g), mean (SD)	11.80 (2.00)	11.39 (1.95)	.14[a]

[a]Adherence represents the mean percentage of prescribed doses completed per participant (161 total doses per patient in both groups).

[b]Independent *t* test.

[c]Statistically significant (*P*<.05).

Discussion

Principal Findings

This randomized controlled trial found that use of the Sharp Health Companion smartphone app improved medication adherence, increased patient confidence, and reduced cataract surgery delays compared with printed instructions.

This study uniquely evaluated how a perioperative app during the COVID-19 pandemic could help predominantly older patients receive seamless care, highlighting digital health's potential to sustain patient engagement during evolving health care conditions. A key finding was the significantly lower rate of surgery delays in the app group 2 compared with the printed group 1, suggesting improved patient readiness and adherence to preoperative instructions. Patient noncompliance with preoperative guidelines, such as dietary restrictions, commonly

XSL•FO
RenderX

leads to surgery cancellations and delays, incurring significant hospital and societal costs [12]. The app's automated reminders and checklists likely mitigated these oversights, aligning with previous findings from a spine surgery mobile app study where app interventions reduced delays and cancellations [28].

Similar postoperative visual acuity improvements were observed in both groups, underscoring that both digital and printed instruction–based methods can support successful cataract surgery outcomes. The comparable, low rates of complications (postoperative iritis and cystoid macular edema) in both groups suggest that the app did not increase the risk of adverse outcomes. These data support using an app alongside printed instructions in routine cataract surgery care.

Comparison With Prior Work

This study is among the first randomized, controlled trials evaluating a smartphone app vs standard printed instructions specifically in ophthalmology and cataract surgery. A recent European study also explored an app-based perioperative approach with cataract surgery but focused on the app's user-friendliness and patient acceptance of digital education [27]. In contrast, this study comprehensively evaluated the full perioperative experience of predominantly older adults undergoing cataract surgery, incorporating subjective patient experiences, objective adherence measures, clinical outcomes, and relevant health system metrics.

Prior studies across various surgical domains, including neurosurgery, general surgery, and orthopedic surgery, have consistently demonstrated improvements in adherence, patient satisfaction, and reductions in cancellations and unplanned visits with digital interventions [29,30]. This study's focused approach on older patients undergoing cataract surgery addresses a critical gap in digital health research, complementing broader telehealth adoption trends during the COVID-19 pandemic [31].

Adherence and Patient Behavior

We observed notable differences in medication adherence: participants using printed checklists reported higher adherence compared with app users, yet objective measures (eye drop bottle weights) indicated superior adherence among app users. This discrepancy likely reflects a combination of factors. Patients often overreport adherence on printed checklists due to social desirability and recall biases [32], whereas the app required real-time tracking and prevented retrospective entries, potentially capturing a more accurate representation of daily use. Bottle weights also reflect unavoidable variability, such as wastage from repeated instillation attempts, differences in patient technique, or clinical variation in dosing intensity (eg, additional drops prescribed for postoperative inflammation). Multiple studies in ophthalmology and other surgical disciplines have similarly demonstrated that self-reported adherence overestimates true use compared with objective electronic monitoring [33-35]. Importantly, improved objective adherence did not negatively affect clinical outcomes, as visual acuity improvements and postoperative complication rates were comparable between groups, confirming the effectiveness and safety of digital perioperative guidance.

Patient Experience and Acceptability

Patient experience is a critical aspect of perioperative care, and the Sharp Health Companion app received overwhelmingly positive feedback from participants. Despite advanced age, many app users reported ease of use, comfort after initial onboarding, and appreciation for automatic reminders on eye drops and fasting. Many participants noted feeling calmer and more prepared, highlighting the app's psychological benefits. Our findings align with broader evidence from similar studies where predominantly older adults demonstrated high acceptance and usability of perioperative apps [36]. Importantly, tailored digital tools have shown value for patients with limited health literacy, providing standardized, clear instructions that may surpass traditional printed instruction methods in effectiveness [37-39].

Limitations

This study had several limitations. The sample was limited to English-speaking iPhone users. Although an Android version of the Sharp Health Companion app was initially developed, extensive software and hardware fragmentation among Android devices resulted in persistent technical issues and inconsistent user experiences. To ensure a uniform and reliable user experience and accurately address the study objectives, we chose to exclusively enroll iPhone users in this prospective study. Limiting the study to English speakers may also restrict generalizability, and future research should test multilingual app versions to improve accessibility in more diverse populations. Consequently, this potentially limited the generalizability of our findings across diverse patient populations and other smartphone platforms. Additionally, this study was conducted at a single, high-volume urban ophthalmology practice, and our findings may not reflect experiences in rural or lower-income settings that can affect mobile health adoption [40].

There was an unexpected statistically significant difference in participants' age, comfort with smartphone apps between groups, and usage of smartwatches, raising the possibility of inadvertent selection bias or unequal participant enthusiasm for digital technology despite randomization. Future studies could address this limitation through stratified randomization based on participants' baseline technology comfort.

Additionally, significantly more participants in group 2 regularly used smartwatches, potentially providing an inherent advantage through increased familiarity with digital notifications and convenient adherence prompts that may have influenced study outcomes. Additionally, group 2 (app users) received backup printed instructions for safety reasons, potentially conferring an advantage by providing supplementary resources unavailable to group 1 participants.

The predominance of female participants, although consistent with broader trends in digital health research among predominantly older adults reflecting higher health care engagement among women, may limit the generalizability of results to a broader population [41].

The study excluded patients with prior cataract surgeries, focusing only on first-time cataract surgery patients. This control

XSL•FO
RenderX

helped standardize patient experience and outcomes, but may limit the generalizability of findings to patients with previous surgical experience. All participants were provided with free eye medications, potentially influencing adherence rates positively compared with typical real-world experiences where medication costs could impact adherence.

A notable proportion of participants in both groups did not complete all study procedures, resulting in missing adherence data due to lost medication bottles or incomplete postoperative surveys from missed follow-up appointments. This finding is consistent with inherent challenges noted in previous clinical studies involving predominantly older adults, who may experience difficulties with medication management, memory, or adherence documentation [39].

Furthermore, subjective survey responses regarding patient experience and satisfaction are vulnerable to social desirability and recall bias, potentially leading to inflated satisfaction or adherence reporting. Although we used deidentified objective adherence measures (medication bottle weights), the complete elimination of subjective bias in self-report metrics was not feasible.

Finally, the follow-up duration of this study was limited to immediate postoperative outcomes (up to 30 days), and we did not evaluate long-term medication adherence or sustained engagement with the app beyond the initial postoperative period. Future research should include longer-term assessments to fully capture sustained benefits and adherence behaviors.

Conclusions and Future Directions

Cataract surgery is the most commonly performed surgical procedure globally. With an aging population, demand for this essential vision-restoring surgery continues to rise. This study demonstrates that the Sharp Health Companion app is a feasible, acceptable, and effective mobile digital tool to support patients, predominantly older adults, throughout their cataract surgery journey. As one of the first randomized controlled trials in ophthalmology directly comparing a digital health intervention against traditional printed instructions, it contributes robust evidence that digital platforms can safely and effectively enhance geriatric surgical care. Specifically, our study suggests the role of digital apps in geriatric surgical care may help reduce delays, increase medication adherence, and enhance patient experience.

Leveraging the CareKit open-source platform enabled rapid, cost-effective development of the Sharp Health Companion app, progressing from concept to minimum-viable product in just 8 months, substantially quicker than typical app development cycles. This accelerated timeline proved especially beneficial when the COVID-19 pandemic unexpectedly emerged, allowing our institution to swiftly adapt perioperative care protocols and continue elective cataract surgeries even as other elective procedures faced significant disruption. Our study also demonstrates older patients' capability and willingness to adopt digital health tools, emphasizing the importance of agile digital health tools in rapidly evolving and challenging health care scenarios.

Our findings suggest that perioperative app companions for older adults could soon become standard tools across various surgical disciplines. Future research should expand these digital care apps to broader populations supporting multiple languages and explore integration with advanced technologies, including artificial intelligence, spatial computing, and cost-effective electronic medication monitoring. Patient-centered digital solutions hold promise for optimizing outcomes, enhancing care experiences, supporting clinician efficiency, and improving overall health system performance.

Acknowledgments

The authors thank the Sharp HealthCare Foundation for supporting this study. We are also grateful to colleagues across Sharp HealthCare and Sharp Rees-Stealy Medical Group, as well as developers from Thread Research, Mobile Programming, and Apple for their technical input. We also thank Kelly J Ko for his counsel. Finally, we thank the patients from the San Diego community whose participation made this study possible.

Generative artificial intelligence tools (Microsoft 365 Copilot) were used for minor language editing, specifically word count, minor grammar corrections, and formatting consistency. All substantive study content, analysis, interpretation, final edits, and approval were performed by the authors.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to patient privacy and Health Insurance Portability and Accountability Act regulations. Deidentified summary data supporting the findings of this study are available from the corresponding author upon reasonable request with approval by the Sharp HealthCare Institutional Review Board.

Funding

This study was funded by the Sharp HealthCare Foundation.

Authors' Contributions

Conceptualization: TSK
Data curation: AA, JF, JH
Formal analysis: TSK, KLG

XSL•FO
RenderX

Funding acquisition: TSK, KLG
Investigation: TSK, AA, LAS, JH, JF, E Linebarger
Methodology: TSK, AA
Project administration: TSK, E Linebarger, LD, AA, LAS
Resources: LD, E Lam, LAS
Software: LD, E Lam, LAS
Supervision: TSK, E Linebarger
Validation: TSK, KLG
Visualization: TSK
Writing—original draft: TSK
Writing—review and editing: All authors

Conflicts of Interest

None declared.

Multimedia Appendix 1
Demonstration video of the Sharp Health Companion CareKit app.
[MP4 File (MP4 Video), 14834 KB - mhealth_v14i1e78710_app1.mp4]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 315 KB - mhealth_v14i1e78710_app2.pdf]

Multimedia Appendix 3
Backup printed instructions with perioperative logistics for cataract surgery for both study groups.
[PDF File (Adobe PDF File), 64 KB - mhealth_v14i1e78710_app3.pdf]

Multimedia Appendix 4
Complete list of patient survey questions administered at three time points: baseline (preoperative office consultation), postoperative day 1, and postoperative month 1.
[DOCX File , 31 KB - mhealth_v14i1e78710_app4.docx]

References

1. Lee CM, Afshari NA. The global state of cataract blindness. Curr Opin Ophthalmol 2017;28(1):98-103. [doi: 10.1097/ICU.0000000000000340] [Medline: 27820750]
2. Rossi T, Romano MR, Iannetta D, Romano V, Gualdi L, D'Agostino I, et al. Cataract surgery practice patterns worldwide: a survey. BMJ Open Ophthalmol 2021;6(1):e000464 [FREE Full text] [doi: 10.1136/bmjophth-2020-000464] [Medline: 33501377]
3. Han X, Zhang J, Liu Z, Tan X, Jin G, He M, et al. Real-world visual outcomes of cataract surgery based on population-based studies: a systematic review. Br J Ophthalmol 2023;107(8):1056-1065 [FREE Full text] [doi: 10.1136/bjophthalmol-2021-320997] [Medline: 35410876]
4. Wang W, Yan W, Fotis K, Prasad NM, Lansingh VC, Taylor HR, et al. Cataract surgical rate and socioeconomics: a global study. Invest Ophthalmol Vis Sci 2016;57(14):5872-5881. [doi: 10.1167/iovs.16-19894] [Medline: 27802517]
5. Aggarwal S, Jain P, Jain A. COVID-19 and cataract surgery backlog in medicare beneficiaries. J Cataract Refract Surg 2020;46(11):1530-1533 [FREE Full text] [doi: 10.1097/j.jcrs.0000000000000337] [Medline: 32694309]
6. Balas M, Vasiliu D, Austria G, Felfeli T. The impact of the COVID-19 pandemic on wait-times for ophthalmic surgery in Ontario, Canada: a population-based study. Clin Ophthalmol 2023;17:1823-1831 [FREE Full text] [doi: 10.2147/OPTH.S409479] [Medline: 37397955]
7. Lundström M, Goh PP, Henry Y, Salowi MA, Barry P, Manning S, et al. The changing pattern of cataract surgery indications: a 5-year study of 2 cataract surgery databases. Ophthalmology 2015;122(1):31-38. [doi: 10.1016/j.ophtha.2014.07.047] [Medline: 25234011]
8. De Oliveira GS, McCarthy RJ, Wolf MS, Holl J. The impact of health literacy in the care of surgical patients: a qualitative systematic review. BMC Surg 2015;15:86 [FREE Full text] [doi: 10.1186/s12893-015-0073-6] [Medline: 26182987]
9. Bheemidi AR, Kailar R, Valentim CCS, Kalur A, Singh RP, Talcott KE. Baseline factors and reason for cancellation of elective ophthalmic surgery. Eye (Lond) 2023;37(13):2788-2794. [doi: 10.1038/s41433-023-02421-2] [Medline: 36750587]
10. Koushan M, Wood LC, Greatbanks R. Evaluating factors associated with the cancellation and delay of elective surgical procedures: a systematic review. Int J Qual Health Care 2021;33(2):mzab092. [doi: 10.1093/intqhc/mzab092] [Medline: 34100548]

XSL•FO
RenderX

11. Al Talalwah N, McIltrot KH. Cancellation of surgeries: integrative review. J Perianesth Nurs 2019;34(1):86-96. [doi: 10.1016/j.jopan.2017.09.012] [Medline: 29678319]
12. Koh WX, Phelan R, Hopman WM, Engen D. Cancellation of elective surgery: rates, reasons and effect on patient satisfaction. Can J Surg 2021;64(2):E155-E161 [FREE Full text] [doi: 10.1503/cjs.008119] [Medline: 33666393]
13. Dadak R, Hatamnejad A, Patil NS, Qiu H, Chan TYB, Rayat J. Eyedrop instillation techniques, difficulties, and currently available solutions: a literature review. J Curr Ophthalmol 2023;35(3):226-230 [FREE Full text] [doi: 10.4103/joco.joco_308_22] [Medline: 38681691]
14. Poojeri U, Babu M, Bhavya H, Shruthi K. Eyedrop instillation technique by medical interns: assessment and impact of video-assisted training. TNOA J Ophthalmic Sci Res. 2024 Jul?Sep 2024;62(3):316-320. [doi: 10.4103/tjosr.tjosr_51_24]
15. An JA, Kasner O, Samek DA, Lévesque V. Evaluation of eyedrop administration by inexperienced patients after cataract surgery. J Cataract Refract Surg 2014;40(11):1857-1861. [doi: 10.1016/j.jcrs.2014.02.037] [Medline: 25248295]
16. Kumar CM, Seet E. Cataract surgery in dementia patients-time to reconsider anaesthetic options. Br J Anaesth 2016;117(4):421-425 [FREE Full text] [doi: 10.1093/bja/aew301] [Medline: 28077527]
17. Moss SJ, Stelfox HT, Krewulak KD, Ahmed S, Anglin MC, Bagshaw SM, et al. Impact of restricted visitation policies in hospitals on patients, family members and healthcare providers during the COVID-19 pandemic: a scoping review protocol. BMJ Open 2021;11(9):e048227 [FREE Full text] [doi: 10.1136/bmjopen-2020-048227] [Medline: 34556510]
18. Vetter TR, Downing ME, Vanlandingham SC, Noles KM, Boudreaux AM. Predictors of patient medication compliance on the day of surgery and the effects of providing patients with standardized yet simplified medication instructions. Anesthesiology 2014;121(1):29-35. [doi: 10.1097/ALN.0000000000000175] [Medline: 24518232]
19. Aronson S, Grocott MPW, Mythen MG. Preoperative patient preparation, programs, and education in the United States: state of the art, state of the science, and state of affairs. Adv Anesth 2019;37:127-143. [doi: 10.1016/j.aan.2019.08.005] [Medline: 31677653]
20. National Academies of Sciences, Engineering, and Medicine, Division of Behavioral and Social Sciences and Education, Health and Medicine Division; Board on Behavioral, Cognitive, and Sensory Sciences, Board on Health Sciences Policy, Committee on the Health and Medical Dimensions of Social Isolation and Loneliness in Older Adults. Social Isolation and Loneliness in Older Adults: Opportunities for the Health Care System. Washington DC: National Academies Press; 2020.
21. Mace RA, Mattos MK, Vranceanu AM. Older adults can use technology: why healthcare professionals must overcome ageism in digital health. Transl Behav Med 2022;12(12):1102-1105 [FREE Full text] [doi: 10.1093/tbm/ibac070] [Medline: 36073770]
22. Faverio M. Share of those 65 and older who are tech users has grown in the past decade. Pew Research Center. 2022. URL: https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/ [accessed 2025-11-08]
23. Apple Inc. Press release - Apple advances health apps with CareKit. Apple. 2016. URL: https://www.apple.com/newsroom/2016/03/21Apple-Advances-Health-Apps-with-CareKit/ [accessed 2025-11-08]
24. Korn T, Slater LA. Seniors can do mHealth apps: tale of inclusivity and innovation. : Healthcare Information and Management Systems Society; 2022 Presented at: HIMSS 2022 Annual Meeting; 2022 Mar 15; Orlando, Florida, Chicago.
25. Abbaspur-Behbahani S, Monaghesh E, Hajizadeh A, Fehresti S. Application of mobile health to support the elderly during the COVID-19 outbreak: a systematic review. Health Policy Technol 2022;11(1):100595 [FREE Full text] [doi: 10.1016/j.hlpt.2022.100595] [Medline: 35018280]
26. Herrera-Usagre M, Santana V, Burgos-Pol R, Oliva JP, Sabater E, Rita-Acosta M, et al. Effect of a mobile app on preoperative patient preparation for major ambulatory surgery: protocol for a randomized controlled trial. JMIR Res Protoc 2019;8(1):e10938 [FREE Full text] [doi: 10.2196/10938] [Medline: 30664480]
27. Gerbutavicius R, Merle DA, Wolf A, Dimopoulos S, Kortuem KU, Kortuem FC. User friendliness and perioperative guidance benefits of a cataract surgery education app: randomized controlled trial. JMIR Form Res 2024;8:e55742 [FREE Full text] [doi: 10.2196/55742] [Medline: 38551619]
28. Stewart JJ, Fayed I, Henault S, Kalantar B, Voyadzis JM. Use of a smartphone application for spine surgery improves patient adherence with preoperative instructions and decreases last-minute surgery cancellations. Cureus 2019;11(3):e4192 [FREE Full text] [doi: 10.7759/cureus.4192] [Medline: 31106092]
29. Grygorian A, Montano D, Shojaa M, Ferencak M, Schmitz N. Digital health interventions and patient safety in abdominal surgery: a systematic review and meta-analysis. JAMA Netw Open 2024;7(4):e248555 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.8555] [Medline: 38669018]
30. van der Horst AY, Kelders SM, Bohlmeijer ET, Schreurs KMG, Jukema JS. Digital health interventions for spinal surgery patients: a systematic scoping review. Digit Health 2025;11:20552076251328549 [FREE Full text] [doi: 10.1177/20552076251328549] [Medline: 40297368]
31. Doraiswamy S, Abraham A, Mamtani R, Cheema S. Use of telehealth during the COVID-19 pandemic: scoping review. J Med Internet Res 2020;22(12):e24087 [FREE Full text] [doi: 10.2196/24087] [Medline: 33147166]
32. Stirratt MJ, Dunbar-Jacob J, Crane HM, Simoni JM, Czajkowski S, Hilliard ME, et al. Self-report measures of medication adherence behavior: recommendations on optimal use. Transl Behav Med 2015;5(4):470-482 [FREE Full text] [doi: 10.1007/s13142-015-0315-2] [Medline: 26622919]

XSL•FO
RenderX

33. Dériot JB, Albertini E, nconnectdrop Consortium. Is CONNECTDROP®, a medication event monitoring system, a viable solution for monitoring adherence to eye drops? Ophthalmol Sci 2024;4(6):100541 [FREE Full text] [doi: 10.1016/j.xops.2024.100541] [Medline: 39310045]
34. Kinast RM, Sanchez FG, Rees JP, Yeh K, Gardiner SK, Dawes J, et al. Eye drop adherence with an eye drop bottle cap monitor. J Glaucoma 2023;32(5):369-373 [FREE Full text] [doi: 10.1097/IJG.0000000000002166] [Medline: 37053080]
35. Tabuchi H, Nishimura K, Akada M, Ishikami T, Shirakami T, Kamiura N, et al. Real-world evaluation of novel eye drop bottle sensors: cloud-based AI support for eye drop adherence. Heliyon 2024;10(14):e34167 [FREE Full text] [doi: 10.1016/j.heliyon.2024.e34167] [Medline: 39092249]
36. Mitzner TL, Savla J, Boot WR, Sharit J, Charness N, Czaja SJ, et al. Technology adoption by older adults: findings from the PRISM trial. Gerontologist 2019;59(1):34-44 [FREE Full text] [doi: 10.1093/geront/gny113] [Medline: 30265294]
37. Park Y, Kim E, Park S, Lee M. Digital health intervention effect on older adults with chronic diseases living alone: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2025;27:e63168 [FREE Full text] [doi: 10.2196/63168] [Medline: 40163849]
38. Heart T, Kalderon E. Older adults: are they ready to adopt health-related ICT? Int J Med Inform 2013;82(11):e209-e231. [doi: 10.1016/j.ijmedinf.2011.03.002] [Medline: 21481631]
39. Lemos M, Henriques AR, Lopes DG, Mendonça N, Victorino A, Costa A, et al. Usability and utility of a mobile app to deliver health-related content to an older adult population: pilot noncontrolled quasi-experimental study. JMIR Form Res 2024;8:e46151 [FREE Full text] [doi: 10.2196/46151] [Medline: 38758585]
40. Alsahli S, Hor SY, Lam M. Factors influencing the acceptance and adoption of mobile health apps by physicians during the COVID-19 pandemic: systematic review. JMIR Mhealth Uhealth 2023;11:e50419. [doi: 10.2196/50419]
41. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act
logMAR: logarithm of the minimum angle of resolution

Edited by L Buis; submitted 07.Jun.2025; peer-reviewed by F Yeh, R Gerbutavicius; comments to author 22.Jul.2025; revised version received 09.Nov.2025; accepted 05.Feb.2026; published 29.Apr.2026.

Please cite as:
Korn TS, Assissini A, Slater LA, Holland J, Doone L, Lam E, Fox J, Greenwood KL, Linebarger E
Digital Health Support for Cataract Surgery With the Sharp Health Companion CareKit App: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e78710
URL: https://mhealth.jmir.org/2026/1/e78710
doi:10.2196/78710
PMID:

©Tommy S Korn, Armand Assissini, Lesli Ann Slater, Janet Holland, Lauren Doone, Elvis Lam, Joy Fox, Kristina L Greenwood, Eric Linebarger. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 29.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Support Strategies to Enhance Adherence to a Prescription Digital Therapeutic for Erectile Dysfunction: Retrospective Quasi-Experimental Cohort Study

Leo Dieter[1], MSc; Mara Haschke[2], MSc; Kurt Miller[1], MD; Laura Elisa Wiemer[1], MD

[1]Department of Urology, Charité - Universitätsmedizin Berlin, Berlin, Germany

[2]Technical University of Munich, Munich, Germany

Corresponding Author:
Leo Dieter, MSc
Department of Urology
Charité - Universitätsmedizin Berlin
Charitépl. 1
Berlin, 10117
Germany
Phone: 49 30 45050
Email: leo.dieter@charite.de

Abstract

Background: Sustained engagement is a central challenge for digital therapeutics. In routine operations, inactivity-triggered supports (eg, SMS reminders and telephone follow-up) are used to nudge patients back after lapses, but evidence in digital erectile dysfunction (ED) therapy is limited.

Objective: This study aims to evaluate whether low-threshold support in the form of SMS reminders and structured telephone outreach, both triggered after 7 or more days of inactivity, improves engagement and patient-centered outcomes in a certified German digital health application (Digitale Gesundheitsanwendung [DiGA]) for ED.

Methods: In a pragmatic, calendar-time (rotating-week) quasi-experimental allocation, 470 men with physician-diagnosed ED entered 1 of 3 groups: control (no additional support), SMS (automated reminders), or call (structured telephone contact). Analyses were conducted as assigned. The primary end point was active weeks (number of weeks with ≥1 completed session) during the 12-week study period, derived from app logs with built-in completion rules. A prespecified mechanistic end point captured any reactivation after an inactive week (yes or no). Secondary end points included Clinical Global Impression-Improvement (CGI-I), week-12 5-item International Index of Erectile Function (IIEF-5), intention to continue therapy, and documented continuation ("conversion") within 3 months. Covariate-adjusted models included age, BMI, smoking, and baseline pharmacotherapy; calendar-week sensitivity was planned.

Results: Participants completed a mean of 6.34 (SD 4.44) active weeks. Unadjusted means favored both support groups (control: 5.79; call: 6.57; SMS: 6.80); the prespecified SMS-vs-control contrast was nominally significant (P=.049, small effect). In covariate-adjusted models, planned pairwise contrasts were not significant. In a calendar-adjusted sensitivity model, the omnibus group term reached significance, but adjusted pairwise contrasts remained nonsignificant; stricter adherence definitions led to the same overall inference. By contrast, any reactivation after inactivity was more common in the intervention groups than in the control group (omnibus chi-square test P=.008), consistent with the intended mechanism of breaking inactivity spells. The call group showed a higher intention to continue therapy (49% vs 38% in the control arm; P=.04); conversion did not differ. CGI-I and IIEF-5 showed no arm-wise differences over 12 weeks; follow-up availability was limited in routine care.

Conclusions: Inactivity-triggered supports in a real-world ED DiGA reactivated use after lapses, and telephone outreach increased motivation to continue, while effects on weekly dose were small and not significant after adjustment, and clinical outcomes did not differ over 12 weeks. Programs may consider a stepped approach (SMS first-line nudge and call as escalation) and target system bottlenecks that decouple motivation from realized continuation. Further work should test longer-term outcomes, targeting, and generalizability.

(JMIR Mhealth Uhealth 2026;14:e76724) doi:10.2196/76724

XSL•FO
RenderX

KEYWORDS

adherence; behavioral therapy; DiGA; digital health; digital therapy; Digitale Gesundheitsanwendung; erectile dysfunction; mHealth; mobile health; nudging; SMS; telephone intervention

Introduction

The increasing digitization of the health care system, particularly through the Digital Healthcare Act (Digitale-Versorgung-Gesetz [DVG]), has allowed physicians in Germany to prescribe digital health applications (Digitale Gesundheitsanwendung [DiGA]) since December 2019. These applications offer innovative approaches to improving patient care and are specifically designed to close outpatient care gaps and increase treatment quality.

A crucial factor for the success of digital therapy approaches is therapy adherence, defined as the extent to which patients consistently and sustainably follow medical and therapeutic recommendations [1]. In the case of erectile dysfunction (ED), a condition involving both physical and psychological components, continuous engagement with the therapeutic content is essential for successful treatment outcomes [2,3]. In practice, however, many patients discontinue digital therapies prematurely, often due to a lack of motivation, insufficient personal support, or technical barriers [4].

To address this challenge, nudging strategies, such as automated SMS reminders or personal telephone calls, are gaining increasing attention. Studies have shown their effectiveness in enhancing adherence in chronic conditions [5-8]. Nevertheless, there is still a lack of studies specifically investigating the benefits of such interventions in the context of digital ED therapy.

The DiGA Kranus Edera, which has been permanently listed in the DiGA directory since 2023 [9] based on its randomized controlled trial results, follows a holistic treatment approach that combines physical exercises with psychological modules in a mobile app [10,11]. Building on this foundation, the present retrospective evaluation investigates whether digital support measures, such as SMS reminders or telephone calls, can improve adherence and motivation during the course of this digital therapy.

This study aims to analyze the effect of such strategies on therapy adherence (primary end point: active training weeks) as well as on other patient-related outcomes (eg, Clinical Global Impression-Improvement [CGI-I], continuation intent, and actual therapy continuation). Furthermore, the study investigates whether certain sociodemographic or clinical factors, such as age, BMI, or preexisting conditions, affect adherence. The findings aim to contribute to the optimization of digital therapy concepts and more personalized patient engagement in the digital health care setting [12].

Methods

Ethical Considerations

This retrospective real-world analysis of the Kranus Edera digital therapy program was reviewed by the Charité - Universitätsmedizin Berlin Ethics Committee (Ethikausschuss am Campus Benjamin Franklin) and approved under protocol EA4/104/24 (committee vote: December 9, 2024; confirmation letter: December 16, 2024). The study analyzed secondary, deidentified usage data exported from routine operations; no research procedures were conducted by the academic team. At program enrollment, users agreed to the provider's terms and privacy policy permitting the use of deidentified data for quality improvement and research. The ethics committee confirmed that the original consent covers this secondary analysis; therefore, no additional consent was required. Participants could opt out of SMS or telephone contact during routine operations; such preferences were honored by the provider. Before transfer to the academic team, all records were deidentified (ie, no names, telephone numbers, or other direct identifiers). Operational messaging and call logs containing identifiers were not shared. Analyses were conducted under a data use arrangement compliant with the General Data Protection Regulation (GDPR) and institutional policies; only deidentified aggregate data are reported. No compensation was provided for this retrospective secondary analysis; participants received standard program communications (SMS messages and calls) as part of routine care.

Study Design

This evaluation is a quasi-experimental retrospective cohort analysis based on anonymized usage data from the Kranus Edera DiGA, exported from the provider's internal database. Data from patients who began therapy between May 28, 2023, and October 20, 2023, were included; the observation window covered 12 weeks from therapy initiation. The support strategies (telephone calls and SMS reminders) were implemented by the provider as part of routine operations; the academic team did not design or assign these interventions and analyzed their real-world use post hoc. The aim was to compare adherence and clinical outcomes among the call, SMS, and control (no additional support) groups. The study protocol was approved by the Ethics Committee of Charité - Universitätsmedizin Berlin (EA4/104/24; Multimedia Appendix 1). All data were fully anonymized prior to analysis and processed in accordance with the Declaration of Helsinki and the GDPR.

Hypotheses

Based on existing literature on digital health interventions and adherence behavior, the following hypotheses were prespecified for this study:

- Patients receiving support interventions (telephone calls or SMS reminders) will demonstrate significantly higher therapy adherence, operationalized as the number of active weeks during the 12-week program, compared with patients in the control group without support.
- Patients in the support groups will report higher subjective treatment satisfaction, measured using the CGI-I scale, and greater improvement in erectile function, measured by the

XSL•FO
RenderX

change in 5-item International Index of Erectile Function (IIEF-5) scores, compared with patients in the control group.
- Patients receiving support interventions will show a higher intention to continue the therapy after the initial 12-week program compared with patients without support.
- Higher therapy adherence (number of active weeks) will be positively associated with subjective treatment improvement (CGI-I) and real-world therapy continuation behavior (conversion rate).

Group Assignment

Participants were allocated at enrollment by the provider's prespecified weekly rotation used in routine operations. The unit of allocation was the calendar week (no individual randomization by the research team):

- Week W: telephone support
- Week W+1: SMS support
- Week W+2: usual care or control (no proactive outreach)
- Week W+3: telephone support, and the sequence repeats

Assignment depended only on each participant's enrollment date (therapy initiation) and was independent of individual characteristics. This pragmatic scheme minimized operational burden and contamination between groups while creating quasi-experimental variation. Because assignment was tied to calendar time, primary analyses adjusted for prespecified baseline covariates and were complemented by calendar-adjusted sensitivity models that included a standardized enrollment-time covariate. We also examined robustness to calendar time using alternative time codings and exploratory interaction tests. All analyses were conducted as assigned (intention-to-treat).

Clinical Eligibility

To align the cohort with the product's intended population in routine care, we included:

- Male adults (aged ≥18 years) with physician-confirmed ED
- Baseline IIEF-5 at therapy initiation with an IIEF-5 score of ≤21 (at least mild ED)
- First-time Kranus Edera use within the study window (May 28, 2023-October 20, 2023)

Definition and rationale: the IIEF-5 is a validated questionnaire assessing erectile function (total score: 5-25, higher scores indicate better erectile function; 4-week recall period). A cutoff score of ≤21 corresponds to at least mild ED (no ED, 22-25; mild, 17-21; mild-to-moderate, 12-16; moderate, 8-11; severe, 5-7) and is commonly used to identify clinically relevant ED.

Analytic Exclusions

From the eligible pool, we excluded:

- Repeat users within the study window
- Records lacking the minimum variables required for analysis (eg, all weekly usage data missing)

Comorbidities (eg, hypertension and diabetes) were not exclusion criteria; they were described and, where appropriate, used as covariates.

Prescribing Context and Contraindications

Kranus Edera is prescribed in Germany through the DiGA pathway. Prescribers are responsible for indication and contraindication screening according to the product's instructions for use. Because this was a retrospective, deidentified data export, individual prescription records were not available; eligibility was operationalized using the variables above. A full list of contraindications and precautions (eg, unstable angina, recent myocardial infarction, New York Heart Association class III-IV heart failure, high-risk arrhythmias, etc, with *ICD-10* [*International Statistical Classification of Diseases, Tenth* Revision] codes where applicable) is provided in Multimedia Appendix 2.

App Structure and Modules

Overview

Kranus Edera delivers a 12-week, guided digital therapy for ED, combining (1) basic pelvic floor muscle training (PFMT), (2) pelvic floor physiotherapy, (3) cardiovascular endurance sessions, (4) brief mindfulness and mental exercises (eg, breathing and stress regulation), and (5) educational and sexual therapy content (knowledge modules on ED etiology, lifestyle, medication, and partner communication). The program is progressive: difficulty and training load increase across weeks, and the app adapts suggested tasks based on completion and self-reported ease.

- Pelvic floor (PFMT): short, technique-focused basic drills (approximately 3-7 minutes) emphasizing anterior pelvic floor activation, relaxation, and coordination.
- Cardiovascular training: time-based walking or aerobic sessions with targets expressed in seconds or minutes (approximately 10-30 minutes), aligned with guideline-consistent aerobic activity to support endothelial function.
- Physiotherapy: longer guided physiotherapy videos (approximately 5-10 minutes) to facilitate PFMT and physical activity.
- Mental exercises: brief (approximately 5 minutes) audio-guided breathing and mindfulness exercises to reduce performance anxiety and support adherence.
- Knowledge and education: lessons of approximately 3-7 minutes on ED mechanisms, risk factors, medication classes, lifestyle, and communication strategies.

Session Cadence and Examples

Typical weekly plans include multiple short sessions (eg, 3-5 PFMT or physiotherapy tasks, 1-3 cardiovascular sessions totaling approximately 20-90 minutes across the week, 1-2 mental exercises, and 1-7 knowledge modules). Example exercises include supine pelvic contractions with breath cueing, bridges, hip openers, interval running cardio, 4-7-8 breathing, and myth-busting lessons on ED. A week-by-week module map with example tasks and nominal durations is provided in Multimedia Appendix 3.

Standards and Prior Evidence

The app's multidomain design (PFMT, aerobic activity, education, and psychological support) mirrors guideline

XSL•FO
RenderX

recommendations for multimodal ED management and demonstrated clinically meaningful improvements in IIEF-5 in a randomized, single-blind controlled trial of this app (EDDIG) [10].

Description of the Support Interventions

We evaluated 2 inactivity-triggered, weekly reactivation strategies implemented in routine operations.

Telephone support (call arm): when a participant had ≥7 consecutive inactive days (no logged activity in any domain), an automated flag prompted outreach. Calls were delivered by technical support staff from the provider's digital therapeutics support team (nonclinical), trained in basic troubleshooting and motivational interviewing and working according to a standardized standard operating procedure. Trained staff attempted up to 3 calls per episode (same day, next day, and the following day), between 9 AM and 7 PM local time. Calls followed a standardized opener and brief problem-solving script (technical troubleshooting or motivational and organizational support), aiming to set 1 concrete activity within 48 hours; red-flag symptoms prompted escalation. Participants could opt out of telephone contact. In routine operations, outreach cycles were repeated if inactivity persisted and were typically capped after 3 consecutive weeks of nonengagement.

SMS support (SMS arm): a weekly automated check sent 1 prescripted, week-specific SMS message when ≥7 inactive days were present. Messages were delivered in the morning (7 AM-11 AM), normalized short lapses, and prompted 1 short, week-appropriate module. The channel was one-way by design; replies were handled ad hoc during routine operations. Participants could opt out of SMS messages. As in the call arm, rechecks occurred weekly, and SMS outreach was typically capped after 3 consecutive weeks of persistent inactivity.

Operational rationale for the 7-day trigger: the ≥7-day threshold was defined by the provider for standard operating procedure simplicity and operational feasibility (weekly check and weekly nudge) and to target early reactivation after a full week of nonuse. This retrospective study did not set that threshold; we analyzed its real-world use. Notably, the same 7-day lapse key performance indicator has been used in prior digital adherence work (eg, medication refill gaps and SMS nudging frameworks) and in broader digital health reengagement strategies [6-8].

Design intent and end point alignment: both support strategies were designed to reactivate participants once per week rather than escalate intensity daily. Accordingly, the primary adherence end point was active weeks (≥1 activity per week), and the mechanism-aligned end point was reactivation after prior inactivity (active in week t given inactivity in week t–1). Stricter intensity and dose metrics are reported as sensitivity or exploratory outcomes.

Measurement of the Primary End Point

Overview

The prespecified primary end point was the number of active weeks (range 0-12 weeks) during the 12-week program. A week was considered active if the user completed ≥1 training session; multiple sessions in the same week still counted as 1 active week. Session completion required passing app-enforced criteria (completion of the full exercise flow and confirmation screens); merely opening the app or brief taps did not qualify. Participants remained in their assigned group according to the intention-to-treat principle; weeks with no telemetry were counted as inactive.

Link to the Intervention's Reactivation Mechanism

Outreach (telephone or SMS) was triggered only after ≥7 consecutive inactive days, with the goal of weekly reactivation. Completion of any session paused further outreach until the next inactivity episode. Thus, active weeks directly capture the intended proximal effect (breaking inactivity within the program's weekly cadence).

This operationalization is consistent with established practice in digital health evaluations that derive adherence from objective usage traces rather than self-report [13] and aligns with payer reporting in Germany, where weekly engagement is a core indicator in DiGA evaluations (eg, Techniker Krankenkasse annual reports [14]).

Robustness Checks

We additionally report (1) the reactivation rate, defined as the proportion of weeks t that were active given that week t–1 was inactive (denominator = all person-weeks following an inactive week); (2) active-weeks version 2 (≥2 sessions or ≥30 minutes of cardiovascular exercise); (3) multidomain engagement (≥2 of 5 weekly goals met); and continuous dose measures (total sessions and median sessions per active week). These thresholds reflect the program's recommended cadence (3-5 physiotherapy or PFMT sessions, 1-3 cardiovascular sessions, and mental exercises and education). Adherence derivation steps (pseudocode) are provided in Multimedia Appendix 3.

Measurement of Secondary End Points

CGI-I

To evaluate patients' subjective impression of change, a modified version of the CGI-I scale was used. At the end of the 12-week program, participants were asked via an email-based questionnaire, "How would you rate your current erectile function compared to before the therapy?"

Responses were given on a 7-point scale from 1=very much worse to 7=very much improved. For consistency with the original CGI-I interpretation, scores were recoded so that lower values indicated greater improvement. The scale was used in a self-rated, ED-specific format, in line with prior validation studies [15,16].

IIEF-5 (Erectile Function) and ΔIIEF-5

Erectile function was measured with the IIEF-5, a validated patient-reported outcome assessing erectile function over the past 4 weeks (items: erection confidence, firmness, penetration ability, maintenance, and intercourse satisfaction) [17]. Each item is scored from 1 to 5; the total score ranges from 5 to 25, with higher scores indicating better erectile function. Common severity bands were as follows: no ED, 22-25; mild, 17-21; mild-to-moderate, 12-16; moderate, 8-11; and severe, 5-7. Baseline IIEF-5 was collected at therapy initiation; a follow-up

XSL•FO
RenderX

IIEF-5 was offered voluntarily in the app during week 12. Change in erectile function (ΔIIEF-5) was computed as the week-12 score minus the baseline score, so positive values reflected improvement. Baseline IIEF-5 was also used as a covariate in adjusted outcome models.

Intention to Continue Therapy

Patients' intention to continue the digital intervention was assessed at the end of the 12-week program via a standardized email-based survey. Participants were asked the following dichotomous question: "Would you like to continue the therapy?" (response options: "yes" or "no").

Although this single-item question is not part of a validated instrument, it serves as a pragmatic proxy for patient satisfaction and perceived utility of the intervention. In line with previous adherence research, stated intention has been interpreted as an early indicator of behavioral engagement and treatment acceptance, particularly in digital health contexts where user motivation is critical to sustained use [12]. The variable was coded as binary (1=yes and 0=no).

Therapy Continuation Rate

Actual therapy continuation, also referred to as conversion, was operationalized based on patients' real-world therapy behavior. Using the manufacturer's internal database, it was determined whether a participant continued the same therapy within 3 months after completion of the initial 12-week therapy cycle.

This binary outcome (1=therapy continued and 0=no continuation) reflects objective, externally validated treatment behavior beyond self-report. As a behavioral end point, it offers insight into long-term therapeutic adherence and real-world uptake of the intervention under routine care conditions.

Statistical Analysis

Overview

All analyses followed an intention-to-treat principle based on calendar-week operational allocation (control, call, and SMS). Because assignment followed calendar-time operational changes, we adjusted for baseline covariates (age, BMI, smoking status, and baseline IIEF-5 where applicable) and acknowledge the possibility of residual time-trend confounding. We prespecified a calendar-time sensitivity analysis in which the primary adherence model additionally included a standardized enrollment-time covariate; we report the omnibus group test and Bonferroni-adjusted pairwise contrasts from that model. Exploratory analyses assessed alternative time codings and group-by-time interactions.

Descriptive Statistics

We summarized the total sample and the groups (control, call, and SMS) with means and SDs. Distributional diagnostics (skewness and kurtosis, histograms, and probability-probability plots) informed model choice. With n>30 per group, 2-tailed *t* tests and general linear models (GLM) were considered robust to moderate nonnormality; BMI showed slight positive skewness but remained within acceptable bounds.

Primary Adherence End Point (Mechanism-Aligned)

The prespecified primary end point was active weeks (number of weeks with ≥1 completed session, 0-12). Group differences were tested with independent-samples *t* tests (call vs control; SMS vs control) and one-way GLM or analysis of covariance (ANCOVA) with group as the factor and age, BMI, and smoking status as covariates. We report adjusted means (estimated marginal means), 95% CIs, and Bonferroni-adjusted pairwise comparisons.

Clinical Outcomes (IIEF-5 and CGI-I)

The primary model for IIEF-5 was an ANCOVA of week-12 IIEF-5 with baseline IIEF-5 (and age, BMI, and smoking status) as covariates. This approach is preferred to testing change scores because it typically offers greater precision when baseline and follow-up are correlated and helps mitigate baseline imbalance and regression to the mean in nonrandomized cohorts.

- Sensitivity: we additionally report ΔIIEF-5 = week 12 − baseline (*t* tests or GLM) for comparability with prior work; interpretation prioritizes the ANCOVA results.
- CGI-I: analyzed directly using a GLM or ANCOVA with the same covariates (CGI-I is itself an improvement scale).

For the ANCOVA models, we checked homogeneity of regression slopes, residual normality, and influential observations; conclusions were unchanged.

Intent to Continue and Conversion

Group differences in intent to continue (implementation outcome) were tested using chi-square tests and logistic regression with group as the main factor and age, BMI, and smoking status as covariates (reporting adjusted odds ratios [ORs] and 95% CIs). Conversion (documented continuation) was analyzed analogously. Because responses were voluntary, these were treated as complete-case outcomes, and response rates are reported; we do not infer clinical efficacy from intention.

Additional Adherence Measures

Threshold-based adherence was assessed as the proportion of participants achieving ≥6, ≥8, and ≥9 active weeks using chi-square tests, with adjusted risk differences and ORs reported with 95% CIs. Adherence intensity was evaluated as the median number of sessions per active week using nonparametric tests alongside GLMs when the data were skewed. Domain-specific goal attainment was assessed as the number of weeks participants met weekly goals in pelvic floor training, pelvic floor physiotherapy, cardiovascular exercise, mental exercises, and knowledge modules, as well as the percentage of all weekly goals achieved across the 12-week intervention, using GLMs or chi-square tests as appropriate. Additional adherence metrics included the number of "stricter" active weeks (version 2), defined as ≥2 sessions per week or ≥30 minutes of cardiovascular exercise per week, and the number of multidomain weeks, defined as weeks in which participants met goals in at least 2 of the 5 intervention domains.

XSL•FO
RenderX

Mechanism-Specific Reactivation and Persistence

Reactivation was defined as being active in week t given inactivity in week t−1. Among participants with ≥1 opportunity, we compared any reactivation (yes or no) using chi-square tests and modeled the reactivation rate (reactivations and opportunities) with ANCOVA using the same covariates.

Persistence, defined as time to the first inactive week, was analyzed using Kaplan-Meier analysis with log-rank tests; participants who remained active throughout the 12 weeks were censored at week 12. We also computed the longest active streak.

Missingness, Nonresponder, and Bias Checks

Primary models used complete cases for the end point of interest; descriptive analyses used pairwise deletion. We compared baseline characteristics and adherence between responders and nonresponders (follow-up completion) using *t* tests and chi-square tests and fitted a logistic regression predicting follow-up completion (covariates plus group) to assess differential missingness by group.

Sensitivity Analyses

We repeated key analyses excluding participants reporting baseline pharmacotherapy (any phosphodiesterase type 5 [PDE-5] inhibitors, Schwellkörper-Autoinjektionstherapie [SKAT], or medicated urethral system for erection [MUSE]). When assumptions were questionable, we added nonparametric companion analyses (Mann-Whitney or Kruskal-Wallis).

All tests were 2-sided with α=.05. We report effect sizes (mean differences, adjusted mean differences, ORs, and hazard ratios where applicable) with 95% CIs. Analyses were conducted using SPSS (version 30; IBM Corp) with the GLM, univariate ANOVA, Kaplan-Meier survival, and logistic regression procedures.

Results

Description of the Study Population

The study population consisted of 470 men with physician-diagnosed ED. The mean baseline IIEF score was 13.1 (SD 4.65), indicating moderate to severe ED. Thus, this cohort represents a clinically relevant population with typical ED characteristics.

The average BMI was 27.3 (SD 4.33) kg/m^2, which falls within the preobesity range according to the World Health Organization (WHO) classification [18]. This value is comparable to the average BMI of German men aged 45-64 years (approximately 27.5-27.7 kg/m^2) [19].

Participants were, on average, aged 48 (SD 14.2) years, with a broad age distribution centered in middle adulthood, an age group in which ED is increasingly common and often overlaps with metabolic risk factors [20].

Functional capacity was high, with 83% (390/470) of participants able to perform 40 minutes of cardiovascular exercise and 96% (451/470) able to move independently.

This high level of functional independence suggests that, despite comorbidities, the cohort was predominantly physically active.

These prevalences align with known risk factors for ED and are partially elevated compared with the general German male population, consistent with the clinical nature of this sample [21]. Nearly half of the men used PDE-5 inhibitors; other invasive or pharmacological therapies, such as SKAT or MUSE, were rare.

Baseline Group Comparability

Baseline demographic and clinical characteristics were broadly similar across groups (usual care [control], telephone support, and SMS). Continuous variables are summarized as mean (SD) and compared descriptively; where hypothesis tests are reported, one-way ANOVA was used for continuous variables and chi-square tests (or Fisher exact tests where expected cell counts were sparse) for categorical variables. To quantify balance independent of sample size, we report absolute standardized mean differences alongside *P* values.

The mean age was 48 (SD 14.2) years, and the mean BMI was 27.3 (SD 4.3) kg/m^2, with no between-group differences (ANOVA *P*=.74 and *P*=.89, respectively). Functional capacity and independence were high across groups (able to perform 40 minutes of cardiovascular exercise: 390/470, 83%; independent mobility: 451/470, 96%). The prevalence of common comorbidities was comparable, including hypertension (129/470, 27.4%), dyslipidemia (93/470, 19.8%), and diabetes mellitus (51/470, 10.9%). Baseline pharmacotherapy also did not differ significantly between groups (PDE-5 inhibitor use: 230/470, 48.9% overall; SKAT or MUSE: 6/470, 1.3%; any pharmacotherapy approximately 50%; all chi-square tests: not significant). Key factors relevant to outcomes and missingness (current smoking: 111/470, 23.6%; history of prostatectomy: 34/470, 7.2%) were similarly distributed.

Between-group differences at baseline were small by ANOVA or chi-square testing. For the prespecified covariates (age, BMI, current smoking, baseline pharmacotherapy, and baseline IIEF-5), no statistically significant differences were detected (Table 1). Primary models adjusted for age, BMI, smoking status, and baseline pharmacotherapy.

XSL•FO
RenderX

Table 1. Baseline characteristics by group. Values are mean (SD) or n (%). *P* values are from one-way ANOVA (continuous variables) or chi-square tests (categorical variables). Denominators vary because of item-level missingness; percentages are based on the nonmissing denominator for each variable.

Characteristic	Total (N=470)	Control (n=180)	Call (n=150)	SMS (n=140)	*P* value
Age (years), mean (SD)	48 (14.2)	48.3 (14.5)	48.4 (13.6)	47.2 (14.6)	.74
BMI (kg/m^2), mean (SD)	27.3 (4.3)	27.4 (4.4)	27.4 (4.6)	27.2 (4.0)	.89
Able to do 40 minutes of cardiovascular exercise, n (%)	390 (83)	144 (80)	123 (82)	123 (87.9)	.17
Able to move independently, n (%)	451 (96)	171 (95)	144 (96)	136 (97.1)	.63
History of spinal injury, n (%)	28 (6)	12 (6.7)	9 (6)	7 (5)	.82
Circulatory disorder, n (%)	14 (3)	9 (5)	2 (1.3)	3 (2.1)	.12
Diabetes, n (%)	51 (10.9)	21 (11.7)	19 (12.7)	11 (7.9)	.38
Dizziness, n (%)	33 (7)	11 (6.1)	13 (8.7)	9 (6.4)	.63
Elevated blood lipids, n (%)	93 (19.8)	45 (25)	26 (17.3)	22 (15.7)	.08
Hypertension, n (%)	129 (27.4)	48 (26.7)	40 (26.7)	41 (29.3)	.84
Pulmonary issues, n (%)	47 (10)	16 (8.9)	17 (11.3)	14 (10)	.76
Mental health burden, n (%)	112 (23.8)	42 (23.3)	38 (25.3)	32 (22.9)	.87
History of spinal surgery, n (%)	52 (11.1)	18 (10)	21 (14)	13 (9.3)	.37
History of prostatectomy, n (%)	34 (7.2)	14 (7.8)	11 (7.3)	9 (6.4)	.90
Knee instability, n (%)	121 (25.7)	55 (30.6)	38 (25.3)	28 (20)	.01
Smoker, n (%)	111 (23.6)	46 (25.6)	38 (25.3)	27 (19.3)	.35
Takes PDE-5[a] inhibitors, n (%)	230 (48.9)	93 (51.7)	67 (44.7)	70 (50)	.43
Uses SKAT[b] or MUSE[c], n (%)	6 (1.3)	2 (1.1)	3 (2)	1 (0.7)	.60

[a]PDE-5: phosphodiesterase type 5.

[b]SKAT: Schwellkörper-Autoinjektionstherapie.

[c]MUSE: medicated urethral system for erection.

Primary Outcome: Therapy Adherence

Definition and Hypothesis

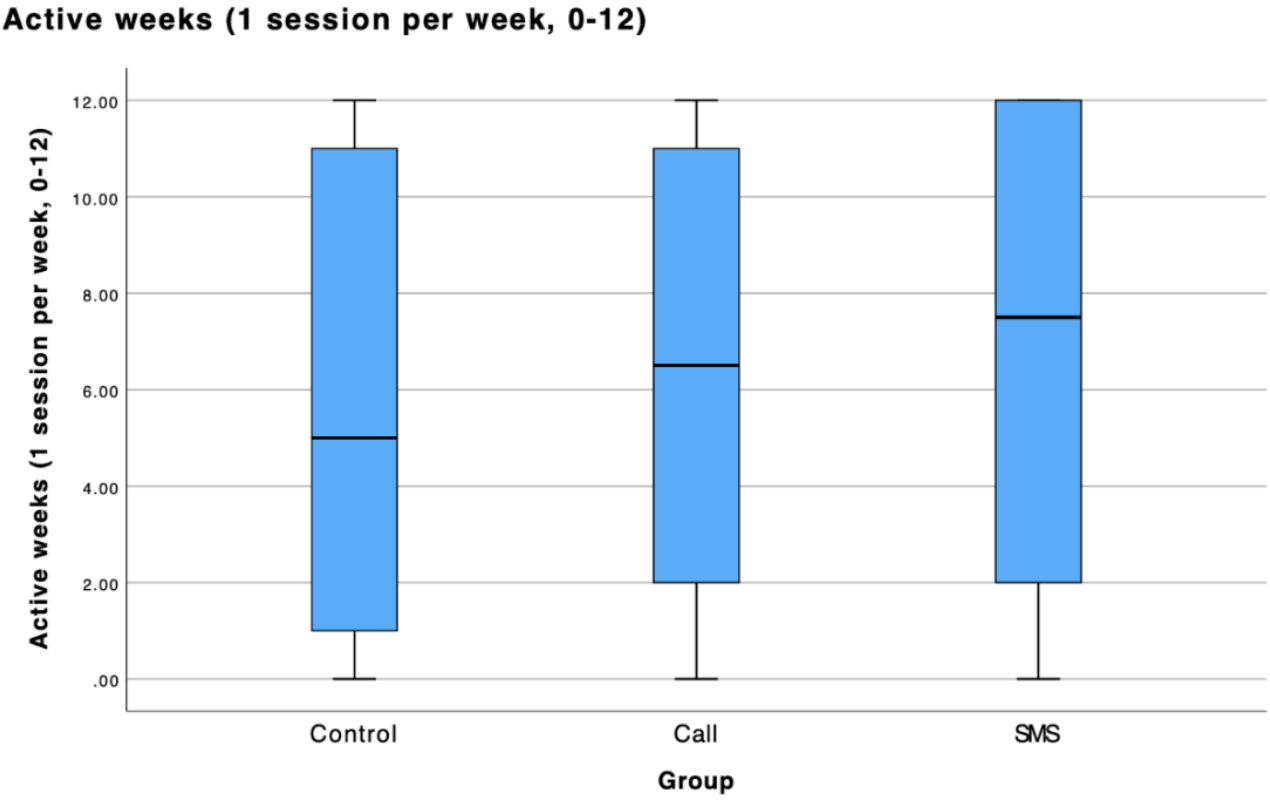
Therapy adherence was prespecified as the number of active weeks (range 0-12 weeks) during the 12-week program; a week was “active” if the user completed ≥1 training session in the app (completion enforced by app logic). The confirmatory hypothesis specified 2 planned contrasts, call vs control and SMS vs control, with α=.05 (2-sided) for each contrast.

Descriptives

Across the cohort (N=470), participants completed a mean of 6.34 (SD 4.44) active weeks. Group distributions were approximately symmetric, with small skewness in each group. The median number of active weeks was 5.0 weeks in the control group, 6.5 weeks in the call group, and 7.5 weeks in the SMS group. Figure 1 shows the unadjusted distribution of active weeks by group.

XSL•FO
RenderX

Figure 1. Boxplot of active training weeks by group. The plot displays the distribution of weekly exercise adherence across the control, call, and SMS groups. Median adherence was highest in the SMS group, with slightly lower variability than in the control group.



Unadjusted Planned Contrasts

Group means were 5.79 (SD 4.54) for the control group, 6.57 (SD 4.22) for the call group, and 6.80 (SD 4.48) for the SMS group. The omnibus one-way ANOVA suggested a trend ($F_{2,467}$=2.333; *P*=.09). Prespecified pairwise tests vs the control group were as follows: call vs control, Δ=0.78 weeks (95% CI −0.17 to 1.73; *P*=.11, not significant); SMS vs control, Δ=1.01 weeks (95% CI 0.01-2.00; *P*=.049); Cohen *d* for SMS vs control was 0.22 (small). Bonferroni-adjusted GLM pairwise comparisons were nonsignificant.

Adjusted analyses

GLM models adjusted for age, BMI, smoking status, and baseline pharmacotherapy showed no significant planned pairwise contrasts in active weeks; the unadjusted SMS advantage was attenuated after adjustment. In a calendar-adjusted sensitivity that additionally included a standardized enrollment-time covariate, the omnibus group effect was small but statistically significant, whereas Bonferroni-adjusted pairwise contrasts remained nonsignificant. Corresponding adjusted estimates are reported in Multimedia Appendix 4.

Sensitivity Analyses

Using stricter adherence definitions (≥2 sessions per week or ≥30 minutes of cardiovascular exercise per week), a multidomain engagement definition (≥2 of 5 weekly goals), and intensity measures (total sessions and median sessions per active week) yielded the same conclusion, with no significant adjusted group differences.

Mechanism-Aligned Proximal Signal

Unadjusted, the call and SMS groups showed higher mean reactivation rates than the control group (control group 0.167; call group 0.223; SMS group 0.235). In a GLM adjusted for age, BMI, smoking status, and baseline pharmacotherapy, between-group differences in the reactivation rate were small and not significant ($F_{2,467}$=1.438; *P*=.24; partial η^2=0.008). By contrast, the probability of any reactivation differed across groups on an unadjusted 3 × 2 test (χ^2_2=9.7; *P*=.008), with higher proportions in call and SMS groups than in the control group.

Persistence and Clinical Outcomes

Time to first inactivity did not differ between groups (log-rank *P*=.36). In ANCOVA of week-12 IIEF-5 adjusted for baseline IIEF-5, age, BMI, smoking status, and baseline pharmacotherapy, the group effect was not significant; adjusted pairwise comparisons were also nonsignificant. Table 2 summarizes the unadjusted group means for the primary and secondary outcomes for transparency; inferential conclusions rely on the adjusted models described above.

XSL•FO
RenderX

Table 2. Primary and secondary outcomes by group (N=470).

Outcome	Total (N=470)	Control group	Call group	SMS group	*P* value (omnibus across groups)	*P* value (control vs call)	*P* value (control vs SMS)
Active training weeks (adherence), mean (SD)	6.34 (4.44)	5.79 (4.54)	6.57 (4.22)	6.80 (4.48)	.09[a]	.11	.049
CGI-I[b] score, mean (SD)	4.85 (1.61)	4.68 (1.91)	4.95 (1.36)	5.00 (1.37)	.80[a]	.29	.22
ΔIIEF-5[c], mean (SD)	3.20 (3.35)	3.37 (3.65)	3.49 (2.96)	2.80 (3.34)	.61[a]	.88	.44
Intention to continue therapy, n/N (%)	192/470 (40.9)	68/180 (37.8)	74/150 (49.3)	50/140 (35.7)	.04[d]	—[e]	—
Actual continuation of therapy, n/N (%)	128/470 (27.2)	47/180 (26.1)	48/150 (32)	33/140 (23.6)	.25[d]	—	—

[a]One-way ANOVA was used for continuous outcomes.

[b]CGI-I: Clinical Global Impression-Improvement.

[c]ΔIIEF-5: change in the 5-item International Index of Erectile Function.

[d]Chi-square tests were used for categorical outcomes.

[e]Not applicable.

Secondary Outcomes

CGI-I

Posttherapy CGI-I data were available for 48.3% (227/470) of participants. Group means were similar (control 4.93, SD 1.60; call 4.95, SD 1.36; SMS 5.08, SD 1.22), corresponding on average to a rating of "minimally improved," and no between-group differences were detected in unadjusted contrasts (control vs call $P>.99$; control vs SMS $P>.99$; omnibus $F_{2,224}=0.220$; $P=.80$). In models adjusted for age, BMI, smoking status, baseline pharmacotherapy, and enrollment-week fixed effects, the group term remained nonsignificant. Further details are available under "Clinical: CGI-I (Adjusted with Covariates)" in Multimedia Appendix 4.

ΔIIEF

ΔIIEF data were available for 27% (127/470) of participants, limiting statistical power. The mean change was 3.20 (SD 3.35; range −7 to +13). Group means did not differ (omnibus ANOVA $F_{2,124}=0.503$; $P=.61$; Bonferroni-adjusted pairwise comparisons, all $P\geq.99$). Consistent with this finding, an ANCOVA of week-12 IIEF-5 adjusted for baseline IIEF-5, age, BMI, smoking status, and baseline pharmacotherapy showed no group effect. Further details are available under "Clinical: IIEF-5 Week 12 (ANCOVA)" in Multimedia Appendix 4.

Intention to Continue Therapy (Binary)

Overall, 40.9% (192/470) of participants indicated an intention to continue. By group, 37.8% (68/180) were in the control group, 49.3% (74/150) were in the call group, and 35.7% (50/140) were in the SMS group. An unadjusted omnibus chi-square test showed a group difference in intention to continue ($\chi^2_2=6.7$; $P=.04$). In the adjusted logistic regression model, the call group vs the control group showed higher odds of intention to continue (OR 1.61, 95% CI 1.03-2.50; $P=.04$), whereas SMS vs control was not significant. Overall model fit was modest (model $\chi^2_6=11.7$; $P=.07$; low pseudo-R^2), so this result should be interpreted as a small adjusted difference. Exploratory adjusted logistic regression models are reported in the section "Intention to Continue Therapy - Logistic Regression (Adjusted)" in Multimedia Appendix 4.

Actual Continuation of Therapy (Conversion)

Overall, 27.2% (128/470) of participants continued therapy. By group, 26.1% (47/180) were in the control group, 32% (48/150) were in the call group, and 23.6% (33/140) were in the SMS group. An unadjusted omnibus chi-square test showed no group difference ($\chi^2_2=2.8$; $P=.25$).

Exploratory Correlations

Exploratory bivariate correlations are reported in the section "Correlations: Active Weeks With Baseline and Clinical Variables" in Multimedia Appendix 4. Adherence correlated negatively with BMI ($r=-0.123$; $P=.007$) and current smoking ($r=-0.112$; $P=.02$). Correlations with age, baseline pharmacotherapy, and baseline IIEF-5 were small, and the correlation with ΔIIEF was not significant ($r=0.020$; $P=.82$).

Discussion

Principal Findings

For the primary adherence end point (active weeks, defined as ≥1 completed session per week), unadjusted means favored both support groups (control 5.79; call 6.57; SMS 6.80). The prespecified SMS vs control contrast reached $P=.049$ with a small effect (Cohen $d=0.22$), whereas the call vs control contrast was not significant. In models adjusted for age, BMI, smoking status, and baseline pharmacotherapy, between-group differences were not statistically significant, and the SMS advantage was attenuated. In a calendar-adjusted sensitivity model, the omnibus group term reached significance, yet Bonferroni-adjusted pairwise contrasts vs the control group did not; effect sizes were small. Together with the null findings for the clinical end points, this supports our interpretation that the support strategies primarily aid reactivation after lapses rather than materially increasing the weekly dose. Sensitivity analyses using stricter

XSL•FO
RenderX

adherence definitions (≥2 sessions per week or ≥30 minutes of cardiovascular exercise; multidomain weekly goals; thresholds of ≥6, ≥8, and ≥9 active weeks) led to the same conclusion: small or null adjusted differences across groups.

For the mechanistic outcomes, any reactivation after an inactive week occurred more often in the intervention groups than in the control group (omnibus chi-square P=.008), indicating that outreach was associated with a higher likelihood of at least 1 return to use after inactivity. Persistence (time to first inactivity) did not differ by group (log-rank P=.36).

Clinical secondary outcomes were null overall: adjusted analyses of CGI-I and week-12 IIEF-5 showed no group effects, and the SMS group did not differ from the control group on these end points. Concurrent pharmacotherapy neither explained the engagement signals (active weeks and any reactivation) nor modified the intervention effects; interaction tests by baseline pharmacotherapy status were nonsignificant. Given the self-reported nature of the pharmacotherapy data and the lack of dose and timing information, these analyses are exploratory.

Taken together, these end point results suggest that the mechanics of the support strategy worked as intended at the event level: nudges helped break inactivity spells (ie, reactivation) while producing only small changes in week-to-week adherence totals. In addition, the telephone intervention increased stated intention to continue therapy vs the control group (n/N, 49% vs n/N, 38%; P=.04; small effect). Prior evidence indicates that personal contact can strengthen perceived support and reduce uncertainty, which may explain the greater stated willingness to continue [22]. In the real-world pathway studied, continuation typically requires renewed physician involvement after 3 months; thus, stated intention to continue therapy functions as a pragmatic implementation outcome and a proximal indicator of readiness to pursue maintenance treatment. The absence of between-group differences in documented therapy continuation within 3 months likely reflects system-level constraints (eg, access and scheduling) that can decouple motivation from realized continuation [23]. Finally, covariate associations aligned with prior work: higher BMI and current smoking were independently associated with fewer active weeks, underscoring the relevance of lifestyle factors for digital engagement [24].

Overall, the interventions modestly influenced weekly adherence, successfully reactivated some users after inactivity, and, especially for telephone contact, increased motivation to continue therapy [22-27].

Hypotheses Revisited

We prespecified that inactivity-triggered supports would increase weekly engagement (H1), that telephone support might increase stated intention to continue therapy (H2), and that clinical outcomes could differ between groups (H3, exploratory). The results supported H1 in a mechanism-aligned way: the support groups, particularly the SMS group, showed higher odds of any reactivation after ≥7 days of inactivity, with small gains in active weeks; however, within-week dose did not materially increase. H2 was supported: the telephone group reported a higher intention to continue therapy at week 12. H3 was not supported: IIEF-5 and CGI-I did not differ between groups. Because IIEF-5 was collected only at the end of the intervention among users who remained engaged, we consider the clinical outcomes exploratory and include a nonresponder analysis (baseline responder vs nonresponder contrasts and a logistic regression model of follow-up completion with covariates and study group; the group term was not associated with completion). Given the short 12-week follow-up and the quasi-experimental design, we interpret the observed effects as engagement-focused (preventing dropout through weekly reactivation) rather than evidence of therapeutic efficacy and therefore avoid causal claims.

Comparison With Prior Work

Prior research has shown that digital reminders, especially SMS messages, can support adherence across chronic conditions, including type 2 diabetes and cardiovascular disease [6-8]. Meta-analytic and trial evidence similarly point to mobile prompts as practical “nudges” for health behavior change [24,26-29]. This study extends these principles to a sexual medicine setting, where adherence involves completing exercise-based therapeutic sessions rather than taking medication. In our data, the SMS group showed a nominal improvement in active weeks in the unadjusted analysis, but adjusted analyses indicated small, nonsignificant differences vs the control group. This pattern is consistent with the broader literature, in which effect sizes for digital nudges are often modest and context-dependent [6-8,28,29].

At the same time, not all large, pragmatic evaluations have found benefits. A recent randomized trial by Ho et al [30] testing reminder and nudge strategies and personalized data feedback for chronic cardiovascular medications did not improve adherence (proportion of days covered) relative to usual care, underscoring how clinical context, target behavior, and implementation details shape effectiveness.

Finally, our observation that telephone support increased stated intention to continue therapy complements prior work on the motivational value of personal contact, while the absence of a corresponding increase in documented therapy continuation highlights a well-described intention-behavior gap in health care [31]. Together, these findings suggest that scalable digital nudges may help restart engagement after lapses, whereas relational outreach can shift motivation. Both are relevant but may be insufficient on their own without system-level supports that translate motivation into sustained behavior.

Strengths and Limitations

This evaluation was conducted within routine operations, using a rotating calendar-week schedule to deliver SMS and telephone outreach in the same workflow patients encounter. That pragmatic embedding enhances external validity and supports scalability. We also aligned the outcomes with the intervention’s mechanism of action, prioritizing the proximal ability of outreach to restart use after a ≥7-day lapse and summarizing week-to-week participation as “active weeks” (≥1 completed session). Because adherence was derived from app telemetry with enforced completion rules, recorded sessions reflect purposeful therapeutic activity rather than passive app use.

XSL•FO
RenderX

Finally, the cohort of 470 men with physician-diagnosed ED was clinically relevant and well characterized; baseline profiles were broadly comparable across groups, and prespecified covariates (age, BMI, smoking status, and baseline pharmacotherapy) were accounted for in the adjusted analyses.

This study used a quasi-experimental allocation by calendar week rather than individual randomization. Although this approach facilitated implementation in a real-world setting and minimized contamination between participants, it introduces potential biases, including unmeasured confounding and time-related effects. Although we adjusted for covariates and examined calendar-time sensitivity, residual and time-varying confounding cannot be excluded, and the design was unblinded. Industry collaboration was limited to the design of reminders and, despite safeguards (independent statistical cross-checks, governance procedures, conflict-of-interest disclosures, and planned code and data sharing), had no influence on data analysis, interpretation, or reporting. Clinical end points were limited by voluntary follow-up in routine care (IIEF-5 and CGI-I), which reduced statistical power and raised the possibility of missing-not-at-random mechanisms; nonresponder analyses did not indicate between-group differences in follow-up, but the clinical findings are interpreted as exploratory. Log-based adherence measures, while objective, do not capture training quality or within-week dose beyond completion thresholds, and passive engagement is not observed; active weeks therefore summarize participation but may underestimate intensity. The observation window was 12 weeks, so durability is unknown. Notably, the higher stated intention to continue therapy after telephone outreach did not translate into higher documented therapy continuation within 3 months, plausibly reflecting system-level constraints (eg, access, scheduling, and represcription). The single-item intention measure is susceptible to response and social desirability bias and should be viewed as an implementation indicator rather than a clinical outcome. Baseline pharmacotherapy was self-reported without dose or timing verification; interaction tests did not suggest effect modification by medication status, but these analyses remain exploratory. Finally, generalizability may be limited because the findings arise from a single DiGA within Germany's reimbursement and 12-week prescription context and may not extend to other health care systems, payment models, or digital therapeutics with different content or support workflows; replication with longer follow-up is warranted.

Implications for Clinical Practice and Future Research

Taken together, our findings suggest that inactivity-triggered outreach can reliably restart engagement after lapses, whereas the sustained week-to-week dose increased only modestly and did not remain significant after covariate adjustment. The interventions were associated with a higher probability of reactivation following an inactive week, and telephone outreach was additionally associated with a greater stated intention to continue therapy at 12 weeks; by contrast, adjusted differences in total active weeks were small, and clinical end points (IIEF-5 and CGI-I) did not differ between groups. In practice, automated SMS reminders are best used as a scalable first-line nudge to break ≥7-day inactivity spells, aligning support with the mechanism most clearly affected (ie, reactivation). Telephone outreach should be reserved for escalation, targeting patients who remain inactive after SMS or who are at higher risk of disengagement, where it may enhance motivation to continue therapy. A pragmatic approach is to combine and sequence support strategies (SMS followed by telephone outreach if the patient remains inactive), integrate them with appointment scheduling or represcription workflows, and continuously monitor barriers to therapy continuation (eg, access and logistics) that can decouple motivation from realized continuation. Expectations for clinical impact should be calibrated accordingly: within the 12-week observation period, we observed no between-group differences in clinical outcomes; therefore, these support strategies should be positioned primarily as engagement tools. Any downstream clinical benefits will likely require longer follow-up and/or stronger changes in dose intensity. In short, support strategies modestly influenced weekly adherence totals but were most effective at reactivating use after lapses, with telephone contact additionally increasing motivation; programs should therefore emphasize preventing dropout at the event level and apply stepped, risk-targeted escalation while avoiding overinterpretation of short-term adherence changes as clinical efficacy.

Conclusions

In this pragmatic, calendar-time–allocated evaluation of a digital ED therapy, inactivity-triggered support showed mechanism-aligned effects on engagement. Across groups, outreach was associated with a higher probability of any reactivation after an inactive week (ie, breaking inactivity spells). In addition, the telephone intervention yielded a higher stated intention to continue therapy at 12 weeks vs the control group, a pragmatic implementation outcome in settings where continuation typically requires renewed prescription. By contrast, differences in weekly adherence totals (active weeks) were small and not statistically significant after covariate adjustment, and clinical end points (IIEF-5 and CGI-I) did not differ between groups over 12 weeks.

Taken together, these findings support a practical division of labor:

- Automated SMS reminders as a scalable first-line nudge to prevent dropout by reactivating users after lapses; and
- Telephone outreach as a more resource-intensive escalation that may shift motivation (higher intention to continue).

These are engagement tools, not stand-alone efficacy treatments. Any downstream clinical benefit will likely depend on sustained exposure and system factors (eg, access to follow-up care) that translate motivation into realized continuation. The promise of simple, low-cost digital communication in routine care remains noteworthy [32], but effect sizes were modest, and causal inference is limited by the quasi-experimental design.

Future work should test stepped, adaptive support (eg, SMS reminders followed by telephone outreach for persistent inactivity), optimize timing and targeting, address conversion bottlenecks (scheduling and represcription), and evaluate longer-term clinical outcomes and cost-effectiveness. Replication in other health systems and with alternative digital therapeutics will clarify generalizability and help define where,

XSL•FO
RenderX

for whom, and at what intensity digital nudges deliver the most value.

Acknowledgments

The authors thank the technical support team for implementing the company's call/SMS support procedures during routine care and the participants who used Kranus Edera as part of their treatment. The authors also thank the Charité - Universitätsmedizin Berlin ethics office for guidance.

Generative AI tools (ChatGPT [GPT 5] and Claude [Sonnet 4.5]) were used for language editing and for assistance with SPSS syntax refactoring and debugging. All statistical analyses, model specifications, and interpretation were designed, executed, and verified by the authors. No AI systems were used to generate, analyze, or interpret the data, and no identifiable data were entered into public AI systems. The authors critically reviewed all AI-assisted text and code suggestions and take full responsibility for the content.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The provider of Kranus Edera (Kranus Health) supplied a deidentified analysis dataset under a data-use arrangement.

The provider implemented the operational outreach workflow and supplied the deidentified dataset. The academic investigators designed the retrospective evaluation, performed the analyses independently, and interpreted the findings.

Data Availability

The participant-level dataset analyzed during the current study is owned by Kranus Health GmbH and is available from Zenodo under restricted access [33]. Access requires a data-use agreement with the rights holder and is limited to research and teaching; no reidentification or redistribution is permitted. The full analysis code (SPSS version 30) and variable codebook are openly available at Zenodo [34].

Authors' Contributions

Conceptualization: LD, LEW
Methodology: LD, LEW
Software: LD, MH
Formal analysis: LD
Investigation/data curation: LD
Resources: KM
Data visualization: LD
Writing - original draft: LD
Writing - review & editing: LD, MH, KM, LEW
Supervision: LEW, KM
Project administration: KM
Guarantor: LD

Conflicts of Interest

LD was employed by Kranus Health until August 2025 and reports no equity or current financial interest; LD is also affiliated with Charité Universitätsmedizin Berlin. MH was employed by Kranus Health until May 2025 and reports no equity or current financial interest. KM is an equity holder and stockholder in Kranus Health and is also affiliated with Charité Universitätsmedizin Berlin. LEW is an employee of Kranus Health, reports no role in statistical analysis or interpretation beyond the contributions listed under CRediT, and is also affiliated with Charité Universitätsmedizin Berlin.

Multimedia Appendix 1
Ethics approval document.
[PDF File (Adobe PDF File), 59 KB - mhealth_v14i1e76724_app1.pdf]

Multimedia Appendix 2
Contraindications and precautions: with ICD-10 (International Classification of Diseases, Tenth Revision) codes.
[PDF File (Adobe PDF File), 84 KB - mhealth_v14i1e76724_app2.pdf]

XSL•FO
RenderX

Multimedia Appendix 3
Week-by-week overview of module content in the app.
[PDF File (Adobe PDF File), 94 KB - mhealth_v14i1e76724_app3.pdf]

Multimedia Appendix 4
SPSS Syntax for Analysis.
[PDF File (Adobe PDF File), 229 KB - mhealth_v14i1e76724_app4.pdf]

References

1. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
2. Burnett AL, Nehra A, Breau RH, Culkin DJ, Faraday MM, Hakim LS, et al. Erectile dysfunction: AUA guideline. J Urol 2018;200(3):633-641. [doi: 10.1016/j.juro.2018.05.004] [Medline: 29746858]
3. Salonia A, Bettocchi C, Boeri L, Capogrosso P, Carvalho J, Cilesiz NC, et al. European association of urology guidelines on sexual and reproductive health-2021 update: male sexual dysfunction. Eur Urol 2021;80(3):333-357. [doi: 10.1016/j.eururo.2021.06.007] [Medline: 34183196]
4. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
5. Anglada-Martinez H, Riu-Viladoms G, Martin-Conde M, Rovira-Illamola M, Sotoca-Momblona JM, Codina-Jane C. Does mHealth increase adherence to medication? Results of a systematic review. Int J Clin Pract 2015;69(1):9-32. [doi: 10.1111/ijcp.12582] [Medline: 25472682]
6. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]
7. Vervloet M, van Dijk L, Santen-Reestman J, van Vlijmen B, van Wingerden P, Bouvy M, et al. SMS reminders improve adherence to oral medication in type 2 diabetes patients who are real time electronically monitored. Int J Med Inform 2012;81(9):594-604. [doi: 10.1016/j.ijmedinf.2012.05.005] [Medline: 22652012]
8. Luong P, Glorioso TJ, Grunwald GK, Peterson P, Allen LA, Khanna A, et al. Text message medication adherence reminders automated and delivered at scale across two institutions: testing the nudge system: pilot study. Circ Cardiovasc Qual Outcomes 2021;14(5):e007015 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.120.007015] [Medline: 33993727]
9. Kranus Edera Information für Fachkreise. DiGA-Verzeichnis. URL: https://diga.bfarm.de/de/verzeichnis/01282/fachkreise [accessed 2025-04-14]
10. Kliesch S, Cremers J, Krallmann C, Epplen R, Scheffer B, Schubert T, et al. App-based therapy of erectile dysfunction using a digital health application (EDDIG Study): a randomized, single-blind, controlled trial. Eur Urol Focus 2024;10(6):1003-1010 [FREE Full text] [doi: 10.1016/j.euf.2024.05.020] [Medline: 38853028]
11. Wiemer L, Bartelheimer T, Raschke R, Miller K. Erste Daten aus einer digitalen Gesundheits-App für Erektionsstörungen [First data from a digital health app for erectile dysfunction]. Urologie 2022;61(9):971-981 [FREE Full text] [doi: 10.1007/s00120-022-01872-x] [Medline: 35925101]
12. Napolitano L, Fusco GM, Cirillo L, Abate M, Mirone C, Barone B, et al. Erectile dysfunction and mobile phone applications: quality, content and adherence to European Association guidelines on male sexual dysfunction. Arch Ital Urol Androl 2022;94(2):211-216 [FREE Full text] [doi: 10.4081/aiua.2022.2.211] [Medline: 35775349]
13. Kernebeck S, Busse TS, Ehlers JP, Vollmar HC. Adhärenz digitaler Interventionen im Gesundheitswesen: Definitionen, Methoden und offene Fragen [Adherence to digital health interventions: definitions, methods, and open questions]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021;64(10):1278-1284 [FREE Full text] [doi: 10.1007/s00103-021-03415-9] [Medline: 34559252]
14. DiGA-Report 2022. Techniker Krankenkasse. URL: https://www.tk.de/resource/blob/2126090/778e6135918696524cccc1b7be39fd1b/diga-report-data.pdf [accessed 2025-03-30]
15. Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. Psychiatry (Edgmont) 2007;4(7):28-37 [FREE Full text] [Medline: 20526405]
16. Guy W. ECDEU Assessment Manual for Psychopharmacology: Revised. Rockville: US Department of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration; 1976.
17. Rosen R, Cappelleri J, Smith M, Lipsky J, Peña BM. Development and evaluation of an abridged, 5-item version of the International Index of Erectile Function (IIEF-5) as a diagnostic tool for erectile dysfunction. Int J Impot Res 1999;11(6):319-326. [doi: 10.1038/sj.ijir.3900472] [Medline: 10637462]
18. Obesity: Preventing and Managing the Global Epidemic: Report of a WHO Consultation. Geneva, Switzerland: World Health Organization; 2000.

XSL•FO
RenderX

19. Height, weight and body mass index of the male population. Federal Statistical Office (Statistisches Bundesamt). URL: https://www.destatis.de/EN/Themes/Society-Environment/Health/Health-Status-Behaviour-Relevant-Health/Tables/body-mass-index-male.html [accessed 2025-04-14]
20. Kaya E, Sikka SC, Gur S. A comprehensive review of metabolic syndrome affecting erectile dysfunction. J Sex Med 2015;12(4):856-875. [doi: 10.1111/jsm.12828] [Medline: 25675988]
21. Braun M, Klotz T, Reifenrath B, Mathers M, Wassmer G, Schoenenberger A, et al. Die Prävalenz von männlichen Erektionsstörungen in Deutschland heute und in der Zukunft [The prevalence of male erectile dysfunction in Germany today and in the future]. Aktuelle Urol 2000;31(5):302-307. [doi: 10.1055/s-2000-7200]
22. Lin H, Wu X. Intervention strategies for improving patient adherence to follow-up in the era of mobile information technology: a systematic review and meta-analysis. PLoS One 2014;9(8):e104266 [FREE Full text] [doi: 10.1371/journal.pone.0104266] [Medline: 25100267]
23. Conner M, Norman P. Understanding the intention-behavior gap: the role of intention strength. Front Psychol 2022;13:923464 [FREE Full text] [doi: 10.3389/fpsyg.2022.923464] [Medline: 35992469]
24. Jakicic JM, Rogers RJ, Davis KK, Collins KA. Role of physical activity and exercise in treating patients with overweight and obesity. Clin Chem 2018;64(1):99-107. [doi: 10.1373/clinchem.2017.272443] [Medline: 29158251]
25. Lang S, McLelland C, MacDonald D, Hamilton DF. Do digital interventions increase adherence to home exercise rehabilitation? A systematic review of randomised controlled trials. Arch Physiother 2022;12(1):24 [FREE Full text] [doi: 10.1186/s40945-022-00148-z] [Medline: 36184611]
26. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
27. Zheng X, Spatz ES, Bai X, Huo X, Ding Q, Horak P, et al. Effect of text messaging on risk factor management in patients with coronary heart disease: the CHAT randomized clinical trial. Circ Cardiovasc Qual Outcomes 2019;12(4):e005616. [doi: 10.1161/CIRCOUTCOMES.119.005616] [Medline: 30998400]
28. Chu L, Jin X, Wu S, Tong X, Li H, Chen X. Effect of pelvic floor muscle training with smartphone reminders on women in the postpartum period: a randomized controlled trial. Urogynecology (Phila) 2024;30(2):138-146. [doi: 10.1097/SPV.0000000000001401] [Medline: 37556387]
29. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of lifestyle-focused text messaging on risk factor modification in patients with coronary heart disease: a randomized clinical trial. JAMA 2015;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
30. Ho PM, Glorioso TJ, Bull S. Personalized patient data and behavioral nudges for medication adherence-reply. JAMA 2025;333(18):1639-1640. [doi: 10.1001/jama.2025.0694] [Medline: 40202730]
31. Sheeran P. Intention—behavior relations: a conceptual and empirical review. Eur Rev Soc Psychol 2011;12(1):1-36. [doi: 10.1080/14792772143000003]
32. Hibbard JH, Greene J, Tusler M. Improving the outcomes of disease management by tailoring care to the patient's level of activation. Am J Manag Care 2009;15(6):353-360 [FREE Full text] [Medline: 19514801]
33. INTERADIG participant-level dataset (restricted access), v1.0.0. Zenodo. URL: https://zenodo.org/records/17329318 [accessed 2026-07-06]
34. INTERADIG analysis code and documentation (no participant data). Zenodo. URL: https://zenodo.org/records/17564279 [accessed 2026-07-06]

Abbreviations

ANCOVA: analysis of covariance
CGI-I: Clinical Global Impression-Improvement
DiGA: Digitale Gesundheitsanwendung
DVG: Digitale-Versorgung-Gesetz
ED: erectile dysfunction
GDPR: General Data Protection Regulation
GLM: general linear model
ICD-10: International Statistical Classification of Diseases, Tenth Revision
IIEF-5: 5-item International Index of Erectile Function
MUSE: medicated urethral system for erection
OR: odds ratio
PDE-5: phosphodiesterase type 5
PFMT: pelvic floor muscle training
SKAT: Schwellkörper-Autoinjektionstherapie
WHO: World Health Organization

XSL•FO
RenderX

Edited by L Buis; submitted 29.Apr.2025; peer-reviewed by JH Song, Y Sapanel; comments to author 23.Sep.2025; accepted 19.Mar.2026; published 14.Jul.2026.

Please cite as:
Dieter L, Haschke M, Miller K, Wiemer LE
Support Strategies to Enhance Adherence to a Prescription Digital Therapeutic for Erectile Dysfunction: Retrospective Quasi-Experimental Cohort Study
JMIR Mhealth Uhealth 2026;14:e76724
URL: https://mhealth.jmir.org/2026/1/e76724
doi:10.2196/76724
PMID:

©Leo Dieter, Mara Haschke, Kurt Miller, Laura Elisa Wiemer. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Real-World Meditation App Engagement: Longitudinal Study of the Medito Meditation App

Julia Adams[1], MPhil; Jonathan Davies[2], PhD; Prai Wattanatakulchat[1], BA; Julieta Galante[1], MD, PhD; Simon D'Alfonso[3], PhD; Nicholas T Van Dam[1], PhD

[1]Contemplative Studies Centre, Melbourne School of Psychological Sciences, The University of Melbourne, 700 Swanson St, Melbourne, Australia

[2]Centre for Innovation in Pain and Health Research, The University of Queensland, Brisbane, Australia

[3]School of Information and Computing Systems, Melbourne School of Psychological Sciences, The University of Melbourne, Melbourne, Australia

Corresponding Author:
Julia Adams, MPhil
Contemplative Studies Centre, Melbourne School of Psychological Sciences, The University of Melbourne, 700 Swanson St, Melbourne, Australia

Abstract

Background: Meditation apps are increasingly popular but face significant engagement challenges. Most research does not meaningfully capture real-world engagement or associated user characteristics. Engagement patterns and reasons for engaging or disengaging remain relatively unexplored.

Objective: This study aimed to examine Medito app user engagement over the first 30 days after download and how intended use, demographics, user traits, and mental health factors predict engagement.

Methods: A prospective online survey was conducted among 668 Medito app users from 30 countries. Factors assessed included demographic factors (eg, age, sex, education, employment, and country of residence); user factors (eg, number of apps tried, hours of experience, meditation-related adverse events, expectations, readiness to change, and personality); and mental health factors (eg, quality of life, perceived stress, psychological distress, well-being, and satisfaction with life). Detailed engagement data included days of use, meditations completed, app opens, and minutes of use obtained via a data-sharing agreement with Medito. Minutes of use in the first 30 days after download served as the main outcome variable.

Results: App use was relatively low, with 50% (328/655) of users engaging for a total of 16 minutes or less in the first month after download (median 16.11, IQR 0 - 74.51 min). Fewer than 20% (124/655, 18.86%) of users continued using the app after 14 days. Intended use (mean 418.56, SD 472.5) significantly exceeded actual use (mean 70.02, SD 176.81; d=0.710; P<.001). In terms of user factors, expectation match (ie, extent to which outcomes from the app matched initial expectations; ρ=0.214; P=.002), expectations for anxiety (ρ=0.102; P=.01), expectations for attention or focus (ρ=0.091; P=.02), and conscientiousness (ρ=0.124; P=.003) were associated with higher engagement. Neuroticism was negatively associated with engagement (ρ=−0.103; P=.010). For mental health factors, satisfaction with life (ρ=0.123; P=.002) and well-being (ρ=0.135, P<.001) were associated with higher engagement, while perceived stress (ρ=−0.107; P=.007), psychological distress (ρ=−0.138, P<.001), and quality of life (ρ=−0.100; P=.011) were associated with lower engagement. Only readiness to change showed unique associations with higher engagement (semipartial r=0.156; P<.001). Regression analysis showed that only perceived stress predicted higher engagement (β=.020; P=.04). However, when mental health was included as a single component, expectations for anxiety (β=.015; P=.049) and readiness to change (β=.011; P=.048) predicted greater engagement, and mental ill health predicted lower engagement (β=−0.008; P=.049).

Conclusions: Overall, app engagement is generally quite low. Acute stress motivated meditation app use, while chronic stress disrupted it. Engagement is optimal when experiences match expectations and users are prepared to make a change. More transparency is necessary in the promotion of meditation apps so that users have a realistic understanding of the time and effort required to achieve benefits.

Trial Registration: OSF Registries osf.io/3v897; https://osf.io/3v897/

(JMIR Mhealth Uhealth 2026;14:e79366) doi:10.2196/79366

KEYWORDS

meditation; mindfulness; engagement; app; meditation apps; digital mental health interventions; behavior change

XSL•FO
RenderX

Introduction

Background

Mental health is a global issue, with 1 billion people living with a mental health disorder—a 50% increase since 1990 [1]. Mental health interventions require 10 to 15 times more resources (eg, medicines, personnel, and equipment) than other health interventions to impact disability-adjusted life years (a measure of the burden of disease through years lost to death and disability), thereby straining limited resources [2]. Smartphone-based meditation apps offer a scalable solution to provide general mental health support [3]. However, with fewer than 10% of individuals who download an app regularly using it beyond 1 week [4] and dose-response examinations suggesting high minimum practice requirements [5,6], most users are unlikely to achieve clinically significant improvements as a result of using a meditation app. A better understanding of what predicts engagement with meditation apps would help support overall use and the targeting of users most likely to benefit. While engagement rates in meditation apps and associated factors have been explored [4,7], to date, limited research has examined how baseline characteristics predict meditation app use over time, which could inform our understanding of which individuals may benefit most from these apps.

Mindfulness-Based Programs, Apps, and Efficacy

Meditation apps make up the majority of mental health apps [8], with the 2 leading apps, Headspace (Headspace Inc) and Calm (Calm Inc), collectively accounting for 70% of the market share in the mental health app sector [9]. Several evidence-based not-for-profit apps have emerged in recent years, such as the Medito app (Medito Foundation) [10]. Unlike traditional mindfulness-based programs (MBPs), which involve group and individual sessions that average 6 hours per week [11,12], most meditation apps involve no human support and have a lower suggested time commitment [5]. While MBPs show moderate improvements in a variety of mental health outcomes [8,13], apps produce significant but smaller effects compared with passive controls [14,15]. App efficacy is undermined by higher attrition rates in randomized controlled trials of apps than those of face-to-face MBPs (see section "*Attrition, Adherence, and Engagement*"), with most users abandoning apps within a week of download [4].

Attrition, Adherence, and Engagement

A challenge in digital meditation research is that engagement is not reported in one-quarter of studies [16] and, where noted, is largely attrition rather than meaningful app use. Engagement measures the amount, frequency, and duration of use, while attrition refers to abandonment or noncompletion rates in an intervention or study. Traditional in-person MBPs have low attrition from study measures (13% - 20%) [17], while attrition from study measures in meditation app randomized controlled trials is approximately 42%, with 90% attrition from application use after a week in naturalistic populations [4,14]. Adherence refers to compliance with an intervention that has a clinically effective "dose" [18]. Adherence captures engagement beyond attrition but is less relevant in nonclinical meditation apps that have no empirically ascertained minimal dose [6]. In our previous analyses of retrospective use patterns, meditation apps were only used, on average, 2.5 times per month.

A further challenge is that digital meditation studies often rely on self-reported adherence or attrition, offering limited insight into use patterns, reasons for disengaging, or real-world use with lower motivation and minimal guidance. Intended and self-reported engagement are typically higher than actual engagement levels and reflect distinct constructs, with some findings suggesting no relationship between reported and actual smartphone use [4,19-21]. Previous work has explored real-world use patterns [4,19,20], predictors of use [7,14], and associations with engagement as well as the relationship between well-being and app use [22,23]. Despite this work exploring factors and use, little is known about which user characteristics predict sustained engagement during the critical first month of use. Understanding prospective indicators of early use patterns may reveal user trajectories, distinguishing the features of those who engage from those who disengage.

Our previous study examined factors associated with meditation app use in general [7], but the retrospective nature of the study limited our ability to identify causality and may have been affected by inaccurate recall [24-26]. Given the importance of consistent engagement in achieving meaningful outcomes [27] and gaps in engagement reporting, examining predictors of engagement with accurate, objective data during crucial habit-building stages [28] could provide valuable insights into what drives meaningful engagement in meditation apps. Prior work has captured retention rates for the first 30 days of use based on publicly available data. A 30-day period provides insight across typical large early drop-offs in app use. This time frame supports an understanding of both individuals who engage and those who disengage.

Reasons for Engaging or Disengaging

Factors predicting app use often focus extensively on app characteristics alone [29] but have recently been extended to include sociodemographics, behavior change, and mental health factors. User factors that affect engagement are explorative but broadly reflect theories of habit formation and planned behavior [30,31].

User Factors: Sociodemographics

Sociodemographic factors may influence engagement in meditation apps. Older age and higher levels of education have been associated with greater engagement in meditation or mobile health apps [23,32], which may be due to poorer health outcomes or health literacy, although use to address health-related issues appears to be common [33,34]. Lower engagement has been found among minority groups in the United States [35,36]. Comparisons of engagement across countries are limited, although one cross-sectional survey found that residing in the United Kingdom was positively associated with engagement [7]. Most meditation app users have relatively high levels of income, although lower-income users may be more engaged in US samples [37,38]. Being female was found to be associated with lower engagement [7]. Barriers include difficulty finding a suitable space for practice and challenging thoughts [39,40]. Spiritual motivations for meditation also

XSL•FO
RenderX

correlate with engagement [5]. Personality factors also predict engagement or cause meditation-related challenges [41], and the latter may result in discontinuation of practice [41]. Agreeableness, conscientiousness, openness, and neuroticism have all been found to predict engagement in meditation [37,42-45], although neuroticism has been linked to more perceived barriers to meditation [46,47].

User Factors: Expectations and Experiences of Progress in Behavior Change

Behavior change factors may also be relevant to meditation app engagement, such as the alignment of expectations and experience. In a prior cross-sectional study, expectations of efficacy (eg, for sleep, stress, and other outcomes) correlated with engagement [7]. Ongoing engagement is promoted by experiences of progress [39,48]. Failure of an intervention to meet expectations predicts (dis-)engagement more than matched expectations [49,50]. Unmet or low expectations reduce perceived usefulness, while matched or exceeded expectations positively influence opinion and behavior [51]. App design can influence expectations, as features such as goal setting and feedback increase perceived efficacy [52]. Fewer perceived barriers and a more positive perception of meditation predict engagement in both meditation in general and meditation apps [32,37,53]. Other factors linked to engagement include higher perceived effectiveness and a higher growth mindset [37].

Mental Health Factors

The relationship between baseline mental health and meditation app engagement remains unclear. Our preceding cross-sectional study found no clear association of engagement with mental health factors [7]. Prior studies suggest that higher anxiety and motivations for mental health are associated with lifetime use, although the relationship with engagement remains unclear [37]. Experiences of mental ill health can lead to more barriers or difficulty engaging in behavior change [54]. As people often use meditation apps for mental health reasons, research should further investigate how baseline mental health factors may contribute to engagement.

Prospective Study Designs

Prospective study designs allow researchers to assess causality in engagement directly and measure baseline factors in real time. Prospective work is also beneficial for determining factors that fluctuate over time. Apps may improve mental health outcomes [8], which could positively or negatively reinforce app use. Additionally, as continual app users are likely to provide higher app ratings over time, understanding baseline expectations will provide insight into whether expectations for efficacy are significant in predicting use.

This Study

This prospective study sought to examine the pattern of engagement seen in new users of the free not-for-profit meditation app, Medito, over the first 30 days of download. The study addressed the following questions, with the following hypotheses stated in the preregistration. Note that question 3 and hypothesis 3 are question 4 and hypothesis 4, respectively, in the preregistration. Preregistration question 3 and hypothesis 3 were not addressed in this manuscript.

- Question 1. What is the *pattern* of engagement in the Medito meditation app in the first 30 days after download?
- Hypothesis 1. Engagement rates will decrease over time.
- Question 2. What is the relationship between *intended use and actual use* of the Medito meditation app?
- Hypothesis 2. Intended use of the Medito app will be associated with actual use (ie, engagement level per primary outcome).
- Question 3. Are there baseline *user factors* that will predict engagement levels over a 30-day period?
- Hypothesis 3. Prospective ratings of readiness to change, expectations for sleep efficacy, expectations for anxiety, expectations for efficacy, and expectations for thriving will predict engagement levels over the 30-day period.

Deviations and Clarifications

Note that the order of preregistration questions has been adjusted to improve logical flow. In addition, question 1 originally referred to the "Medito mindfulness app," which has been changed to "Medito meditation app" for clarity and consistency. Question 4 in the preregistration is question 2 in the manuscript, and question 2 is now question 3. "Use" in hypothesis 4 is conceptualized as engagement level as shown by the primary outcome variable. While we initially obtained follow-up data, the number of people who returned follow-up data was relatively small. Therefore, question 3 in the preregistration is not addressed in this manuscript.

Methods

Ethical Considerations

Ethical Approval and Participant Consent

This study was conducted according to ethical guidelines and was approved by the Office of Research Ethics and Integrity, University of Melbourne (2024-29651-52025-2).

Participants consented to participate in the study via the online Qualtrics survey. Consent information was provided in the online survey, and a plain language statement was available for download. The plain language statement can be found in Multimedia Appendix 1. The consent information included explicitly clarifying the participant's understanding of their right to withdraw at any time without explanation. Participants consented to secondary analyses. The survey flow prevented anyone who did not provide consent from completing the survey, and all responses included were double-checked for consent provision.

Participant Compensation

Our consenting sample before exclusions consisted of 1214 participants. Participants were compensated with one entry into a gift card draw for each survey completed, for 2 possible entries across both surveys. One AUD $250 (US $176.05) gift card was drawn per 250 sequential survey completions, resulting in 4 gift cards for the baseline survey and 1 for the follow-up. This equated to a total prize pool of AUD $1250 (US $825.25).

XSL•FO
RenderX

Privacy and Confidentiality

Participant data with identifying information (such as first name, first initial of last name, and email address) were retained for compensation purposes but stored securely according to relevant privacy guidelines and encrypted with Transport Layer Security encryption (also known as HTTPS).

Study Design and Setting

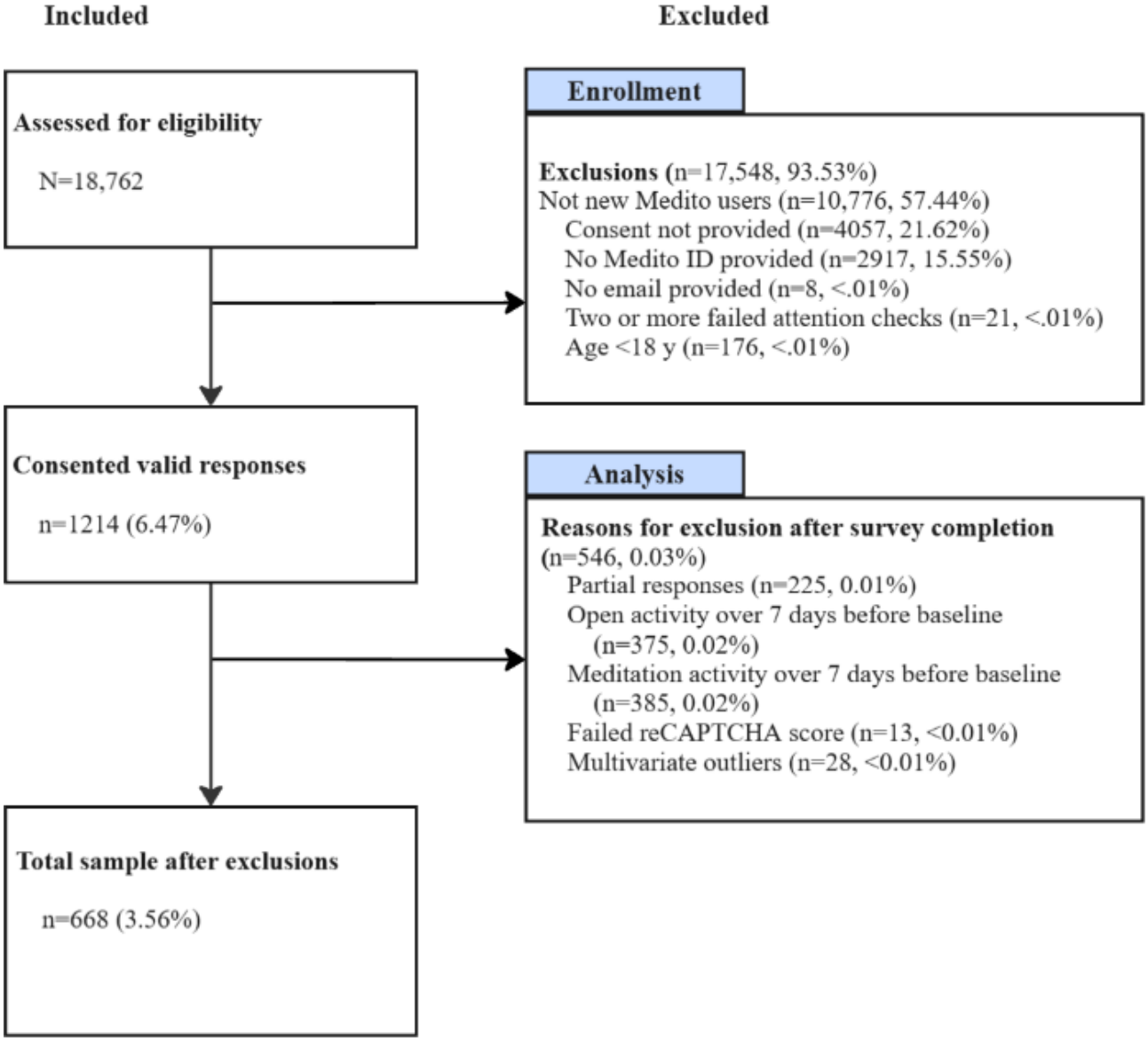
Medito App

Medito is a free meditation app offering guided audio meditations, available on the iOS and Android app stores. Meditators can choose individual sessions or follow structured programs, such as a 30-day introductory challenge.

Data Collection

The 2 methods of data collection were via a data sharing agreement with the Medito app developers and through a Qualtrics survey. Users provided their Medito ID number when consenting to the research team analyzing their app use. The baseline survey included a screening question, "Did you download Medito for the first time within the past day?" to target new users. The baseline survey was completed at the point of sign-up, and the follow-up survey was distributed for completion 30 days after baseline completion. Participant flow is discussed in the following section and displayed in Figure 1.

Figure 1. Participant flow diagram. The exclusion groups within the enrollment and analysis stages are not mutually exclusive.



Eligibility and Recruitment

Inclusion and Exclusion

Participants had to have downloaded Medito within the previous 24 hours to be eligible. Survey access was restricted to major countries that are predominantly English speaking, along with members of the European Union: Australia, Austria, Belgium, Bulgaria, Canada, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, New Zealand, Poland, Portugal, Romania, Spain, Sweden, the United States, and the United Kingdom. These countries were selected due to the predominance of English speakers and

XSL•FO
RenderX

approximate financial parity to ensure ethical compensation and consistency in privacy laws.

Participant Flow

See Figure 1 and Supplementary 1 in Multimedia Appendix 1 for a detailed description of participant flow.

Procedure

Participants were sourced directly from Medito to complete a survey hosted by Qualtrics and advertised on the Medito app platform. The survey was live on the Medito platform from July 20 to August 30, 2024. Follow-up surveys were automatically sent 30 days after baseline completion. Surveys could be completed online using a laptop or mobile device.

Outcome Variables

Engagement Data

Medito provided detailed app usage data, including app opens count, meditation tracks completed, track length, track title, time of activity, meditation tracks started, and streaks (days of consecutive use). Information provided by Medito and used in analysis included Medito IDs for linking data to survey responses, activity time stamps for opens and track listens, and duration of meditation tracks. The total minutes of engagement in the first 30 days were calculated by summing the duration of each track completed in the 30 days following the user's first app open.

Study Outcomes

Total minutes of use in the first 30 days were the primary outcome variable to address research questions 1, 2, and 4. Minutes of use were calculated as a total over the 30-day period following the first time the app was opened by adding the durations of tracks completed. Days of use, meditations completed, and app opens were also collected to provide a holistic view of engagement in addressing question 1. App opens allowed for a comparison between daily app activity in our sample compared with Baumel et al [4]. Intended use patterns were collected to compare with actual use and to address question 4. The follow-up survey addressed the relationship between app factors and engagement outlined in question 3, which will be published elsewhere.

Data Sources and Measurement

Overview

Participants reported their age, gender (eg, male, female, nonbinary, genderfluid, or other), income, and education level. Participants reported the number of apps they had tried, including Medito, ranging from 1 (Medito only) to ≥5. Participants also reported hours of meditation experience in the following ranges: beginner (0 - 100 h), intermediate (101 - 1000 h), and expert (≥1001 h). Mental health factors included satisfaction with life, meditation-related adverse event impairment level and length, well-being, perceived stress, distress, and quality of life. Expectations were assessed for various domains (eg, sleep, stress, anxiety, attention, happiness, thriving, and performance enhancement), as well as whether the app matched expectations, how confident the user was in improvement, what degree of improvement was expected, and how ready they were to change. Personality factors were based on the Big Five Inventory [55].

Meditation-Related Adverse Events Abbreviated 3-item

Part 1 of the Meditation-Related Adverse Events [56], a 2-part scale that assesses meditation-related events, was completed. The first part of the scale consists of 3 items assessing the frequency of perceived adverse effects, the extent of related functional impairment, and the length of functional impairment. Each item has a Likert-style response format, ranging from never to frequently for frequency, not at all to severely for impairment, and 1 day or less to 1 year or longer for length of impairment.

The EuroQoL Health and Wellbeing Assessment–Short Form

The EuroQoL [57] assessment is a newly formulated quality of life measure. The scale consists of 9 items on a 5-point scale, where 1 represents "no difficulty," "none of the time," or "no physical pain" and 5 represents "unable," "most or all of the time," or "very severe physical pain," depending on the context. In this study, the scale had good internal consistency (α=.86; ω=0.72).

Perceived Stress Scale

The Perceived Stress Scale [58] is a brief measure of subjective stress. It ranges from 0 (never) to 4 (almost always). In this study, the scale had reasonable internal consistency (α=.71; ω=0.69).

The Kessler Psychological Distress Scale

The Kessler Psychological Distress Scale (K10) [59] was used to assess psychological distress in the last 30 days. The scale consists of 10 items assessing anxiety and depressive symptoms on a 5-point scale, ranging from 1 (none of the time) to 5 (all of the time). In this study, the scale had excellent consistency (α=.91; ω=0.78).

The Short Warwick-Edinburgh Mental Wellbeing Scale

The Short Warwick-Edinburgh Mental Wellbeing Scale [60] was used to assess positive elements of mental health and well-being. This scale consists of 7 items on a scale of 1 to 5, ranging from 1 (none of the time) to 5 (all of the time). In this study, the scale showed good consistency (α=.85; ω=0.75).

The Satisfaction With Life Survey (1-item)

The Satisfaction With Life Survey [61] Single Item is a 1-item version of the more established Satisfaction with Life Survey. The scale has reasonable criterion validity with the full Satisfaction With Life Survey (zero-order r=0.62 - 0.64; disattenuated r=0.78 - 0.80 [61]). The single-item measure asks participants to rate life satisfaction from 1 (extremely dissatisfied) to 7 (extremely satisfied).

Expectancy for Various Domains

Medito was rated in general for how effective the user thought it was for the following: anxiety, attention or focus, happiness, thriving, and performance enhancement. The rating scale ranged from 1 (not at all effective) to 7 (extremely effective).

XSL•FO
RenderX

Readiness to Change 1-item

The Readiness to Change 1-item [62] assessment is a 10-point scale that assesses an individual's preparedness to enact a behavioral change. The scale ranges from 0 (not prepared to change) to 10 (already changing). It is based on the transtheoretical model of behavior change stages [63,64]. The readiness ruler captures a construct related to actual readiness [63,65].

The Big Five Inventory Short Form 2

The Big Five Inventory Short Form 2 is a 30-item [55] questionnaire that assesses 5 domains of personality as subscales: extraversion, agreeableness, conscientiousness, negative emotionality, and open-mindedness. Each subscale contains 6 items. The subscales representing various personality domains were assessed for internal consistency using Cronbach α and McDonald ω. Cronbach α and ω consistency ranged from poor to excellent (extraversion: ω=0.57; α=.74; agreeableness: ω=0.56, α=.67; conscientiousness: ω=0.59, α=.73; neuroticism: ω=0.69, α=.85; and openness: ω=0.44, α=.63).

Statistical Analysis

The target sample size was 1000 participants at baseline, providing 90% power to detect a small effect size of r=0.102 at α=.05 [14]. The effect size of r=0.102 was hypothesized based on prior meta-analyses effect size estimates. The target sample size was determined using G*Power (Heinrich Heine University Düsseldorf) software [66]. Due to unexpected exclusions based on preexisting app use (~30% - 32% of the eligible sample), the final baseline sample size was 668 participants. With this reduced sample, the study retained 90% power to detect an effect of r=0.177.

Intended use was individually reported by users across sessions per day, minutes per session, and days per week. The indicated levels of intended use per person were used to calculate the intended total minutes of use over the first 30 days.

A paired-samples 2-tailed t test was conducted to compare intended use and actual use (in minutes). Zero-order correlations were calculated to investigate the overall relationship between the factor and the dependent variable. Semipartial correlations were calculated to investigate the unique relationship between a given factor and engagement, controlling for other factors. A regression model was used to investigate which factors showed significant associations with engagement, accounting for multiple predictors. The main outcome variable was total minutes of use over the first 30 days of app use. All factors that approached a significant zero-order association with engagement (P<.10) were included in the regression model.

A principal component analysis (PCA) was used to address multicollinearity among predictors and reduce dimensionality in our data. The PCA identified single components within the mental health and expectation outcomes. These were used in additional regression models to further investigate contributions among correlated predictors.

Results

Participant Characteristics: Descriptives

Participants ranged from age 18 to 74 (mean 32.2, SD 11.99) years and were close to evenly split between male (n=322, 48.20%) and female (n=318, 47.60%) participants. A small proportion reported nonbinary, gender fluid, or other gender (n=28, 4.20%). Most participants lived in the European Union (n=240, 35.96%) or the United States (n=239, 35.78%). Approximately two-thirds of our sample had completed education above high school (n=458, 68.59%). Most participants were employed full time (n=310, 46.41%) or part time (n=114, 17.08%). Approximately one-fifth were unemployed (n=123, 18.41%). Income was adjusted within countries into 5 brackets, with each bracket covering approximately 20% of the income range for that country. People were evenly distributed across the 4 lower income brackets adjusted for income levels across countries, with fewer (n=77, 11.53%) people in the top income bracket for their country. Notably, participants reported relatively high levels of perceived stress [67] and moderate-to-high psychological distress levels [68] compared to normative English and Australian samples (Table 1).

XSL•FO
RenderX

Table . Meditation experience and mental health descriptives.

Categories	Values
Meditation experience, n (%)	
Apps tried	
1 (Medito only)	209 (31.25)
2	183 (27.49)
3	171 (25.67)
4	53 (7.93)
No response	52 (77.90)
Hours of meditation experience	
0 - 100	545 (81.66)
101 - 1000	114 (17.03)
1000+	9 (1.35)
Mental health, mean (SD)	
Perceived stress (PSS-4[a])	7.20 (3.11)
Psychological distress (K10[b])	23.87 (8.29)

[a]PSS-4: Perceived Stress Scale-4.

[b]K10: 10-item Kessler Psychological Distress Scale.

Statistical Recording Issues

From July 20 to August 13, Medito indicated a technical anomaly in the recording of statistics for tracks completed and streaks. However, analyses indicated the glitch was unlikely to have seriously affected the data (Supplementary 2 and Supplementary 3 in Multimedia Appendix 1).

Overall Engagement

To research app engagement, we assessed user engagement levels via total minutes. We further clarified patterns of use by assessing weekly minutes and days of use (Table 2). Engagement was highly positively skewed (mean 70, SD 176.81; median 16.11, IQR 0-74.51 during the first 30 days), suggesting that most participants had low engagement. Most sessions lasted 10 to 11 minutes; approximately half of the users in our sample did not engage beyond approximately one guided meditation. The majority (520/655, 79.4%) of participants engaged for less than 100 minutes in the first 30 days of use (Figure 2 and Table 3), despite the majority of participants intending to engage for 10 minutes per day over the 30-day period (Figures 3 and 4). Engagement dropped quickly over time; on day 7, only 27.68% (n=181) of participants used the app. Approximately one-fifth (n=124, 18.86%) were still engaged on day 14, while 16.26% (n=107) were still engaged at day 30. Overall, a higher proportion of users were engaged each day than in the sample of Baumel et al [4], with a median daily use of 17.47%, compared with 4.10% (see Figure 5 for comparison). No differences in engagement were found by age, sex, education, religion, income, or country (Figures 3 and 6 and Supplementary 4 and Supplementary 5 in Multimedia Appendix 1).

Table . App use descriptives.

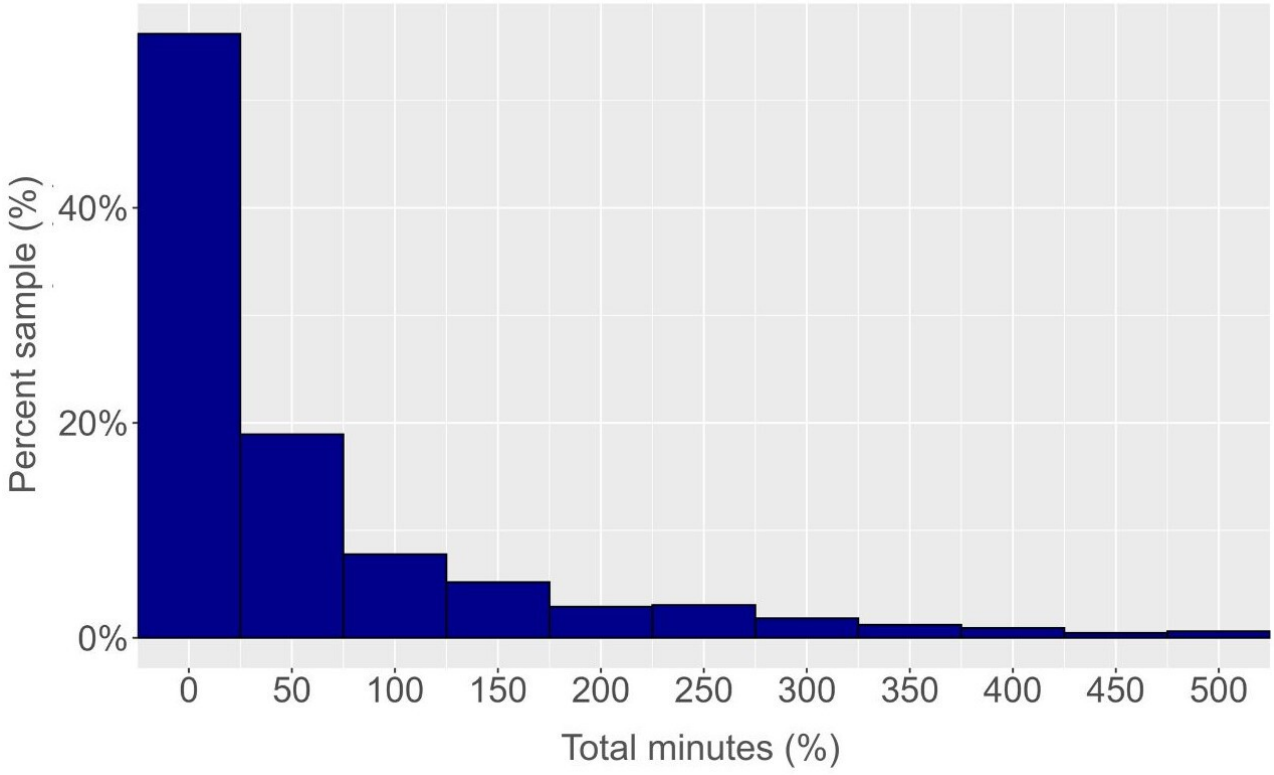
Engagement variable	Values, n (%)	Values, mean (SD)	Percentile				
			5%	25%	50%, median (IQR)	75%	95%
Total minutes	655 (98.05)	70.02 (176.81)	0.00	0.00	16.11 (0-74.51)	74.51	292.49
Total sessions completed	668 (100)	10.29 (11.12)	1.00	2.00	6.00 (2-14)	14	34
Tracks started	518 (77.54)	12.48 (14.96)	1.00	3.00	8.00 (3-19)	19	45.15
Duration (d)[a]	455 (68.11)	13.80 (11.80)	0.01	1.43	11.96 (1.43-27)	27	29.75
Minutes per session[b]	468 (70.06)	11.34 (11.57)	2.68	5.01	10.00 (5.01-11.75)	11.75	29.95

[a]Duration in days equals the number of days between first opening the app and the most recent use within the 30-day period.

[b]Minutes per session were calculated by taking the statistic for minutes and dividing it by sessions completed for the mean and percentile ranges.

XSL•FO
RenderX

Figure 2. Engagement pattern—total minutes of use in the first 30 days, expressed as a percentage of the sample. This figure is truncated for visualization, excluding 21 outliers with values above 1000 min.



XSL•FO
RenderX

Table . Sociodemographic features.

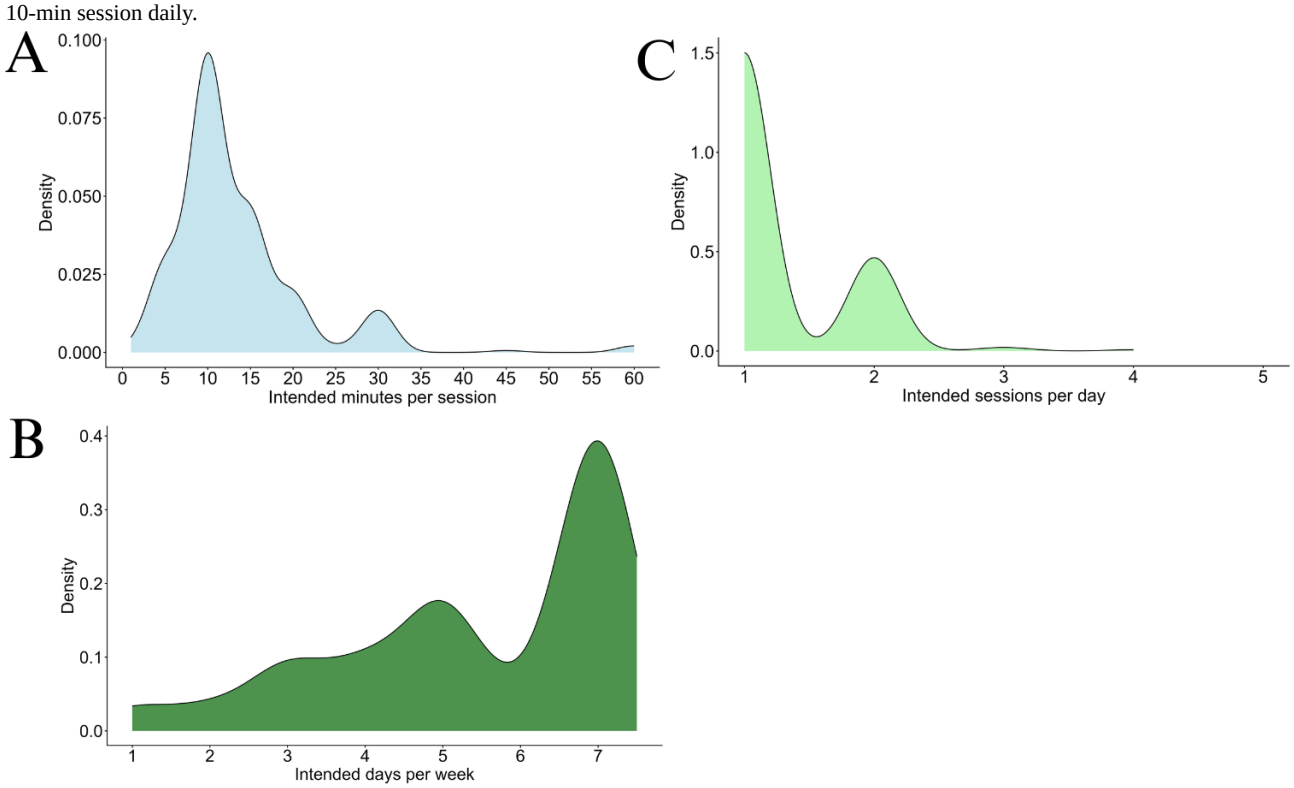
Categories	Values
Age (y), mean (SD)	32.20 (11.99)
Gender, n (%)	
Male	318 (47.60)
Female	322 (48.20)
Nonbinary, genderfluid, or other	28 28 (4.20)
Income level[a], n (%)	
1	159 (23.80)
2	161 (24.10)
3	106 (15.87)
4	77 (11.53)
5	132 (19.76)
Country/continent[b], n (%)	
European Union	240 (35.96)
United States	239 (35.78)
United Kingdom	53 (7.93)
Australia	57 (8.53)
Canada	64 (9.58)
New Zealand	15 (2.24)
Occupation, n (%)	
Employed part-time	26 (3.89)
Employed full-time	56 (8.38)
Due to start a job within the next month	67 (10.02)
Not in paid work (eg, homemaker, retired, or disabled)	95 (14.22)
Unemployed (and job seeking)	114 (17.07)
Other	310 (46.41)
Religion, n (%)	
No religion	8 (1.20)
Christian	10 (1.50)
Hindu	11 (1.65)
Jewish	14 (2.10)
Muslim	21 (3.14)
Buddhist	47 (7.04)
Sikh	53 (7.93)
Spiritual but not religious	121 (18.11)
Atheist	152 (22.75)
Agnostic	231 (34.58)

[a]Income level is reported in quintiles standardized for local currency and income distribution, with 1 representing the lowest quintile and 5 the highest. Frequency in each quintile represents the distribution of participants across income brackets.

[b]See Supplementary Material 5 for full country of residence details. European Union countries have been reported as a frequency together here for brevity.

XSL•FO
RenderX

Figure 3. Smoothed density plots showing intended use. Panel A shows intended minutes per session, panel B shows sessions per day, and panel C shows days per week. Panels A to C show intended use by Medito users at the time of download, reported in sessions per day, minutes per session, and days per week. Participants reported intended sessions per day on a scale of 1 to 5 and intended minutes per session from 0 to 60 in approximately 5-min increments (excluding between 30 and 45 min). Intended days per week ranged from 1 to 7 d. Most users (>50%) reported intending to use the app for one 10-min session daily.



XSL•FO
RenderX

Figure 4. Density curves showing intended and actual minutes of use within the first 30 days, with minutes on the x-axis and probability density on the y-axis. The blue dotted line represents the median intended minutes, and the green dotted line represents the median actual minutes.

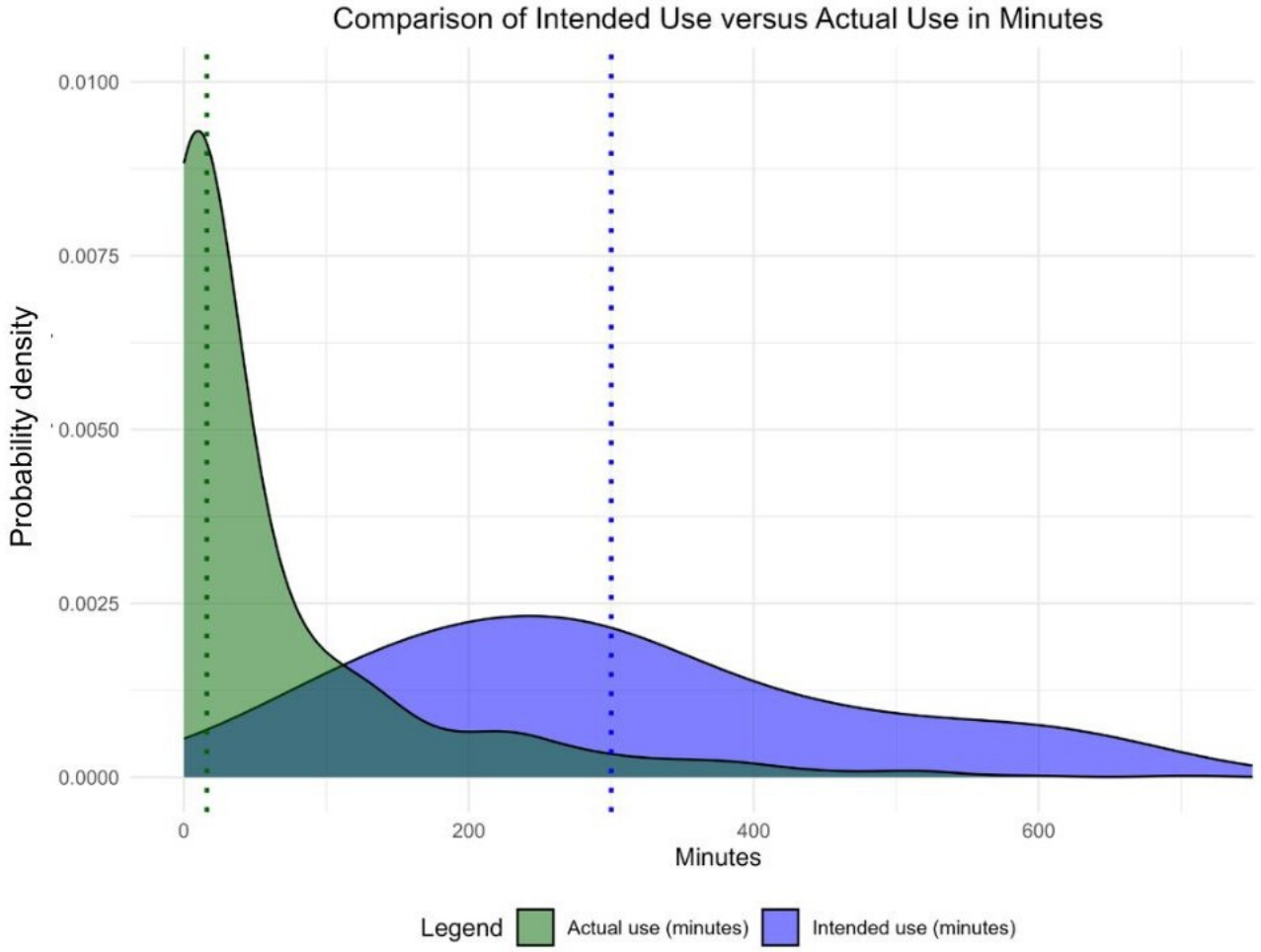
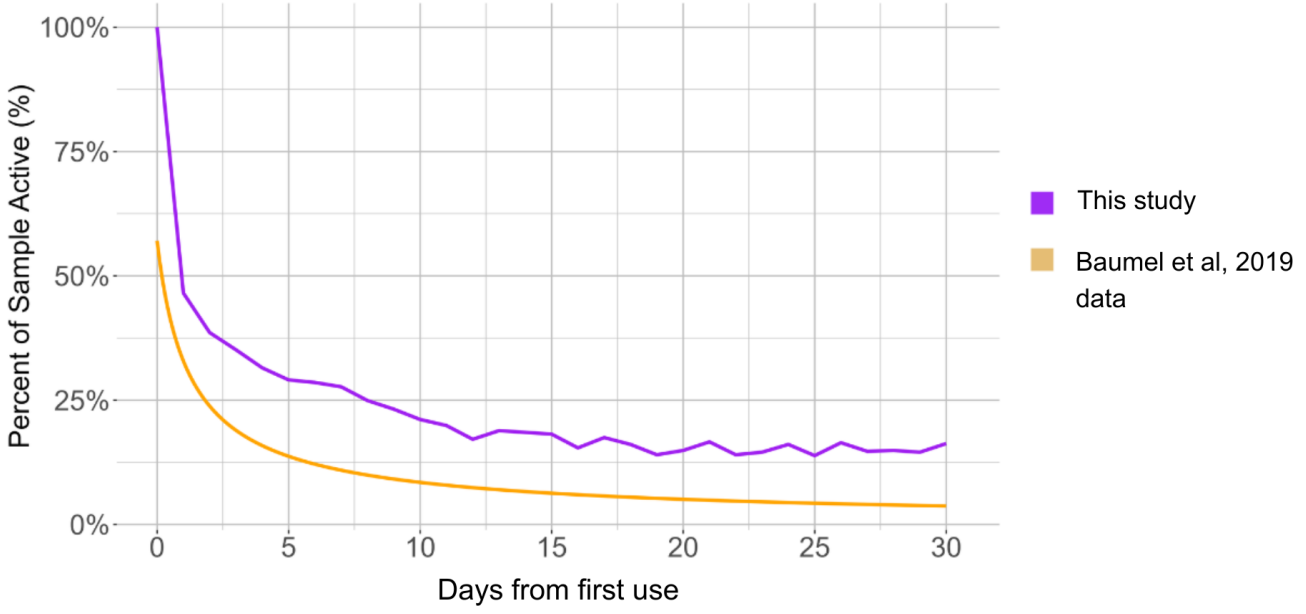
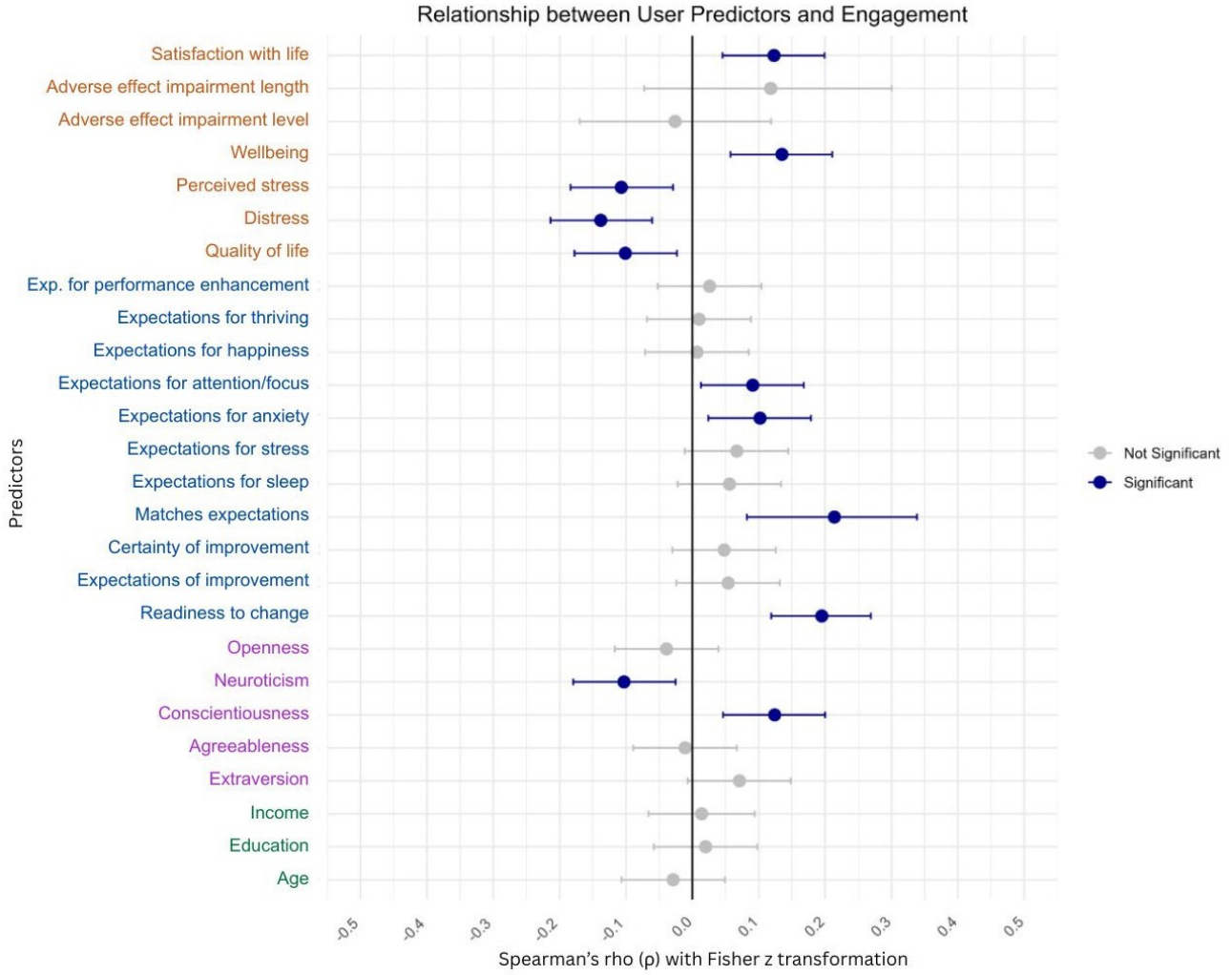


Figure 5. Curve showing the percentage of the sample actively using the app each day after download. App use on a given day represents the percentage of the sample who used the app on that day. The percentage of the sample active per day was used to provide a direct comparison with the findings of Baumel et al [4]. The yellow line represents an estimated curve from Baumel et al [4] and does not reflect the variation in app opens on each given day. The purple line represents the trend line of active use each day in the first 30 days of use in this study.



XSL•FO
RenderX

Figure 6. Spearman ρ coefficients of association between predictor factors and total minutes of use over the first 30 days. Whiskers represent 95% CI generated using Fisher z transformation. Dark blue represents significant association at $P<.05$; gray represents nonsignificant associations. Orange represents mental health factors, dark blue represents expectations factors, purple represents personality factors, and green represents sociodemographic factors. Quality of life is reverse coded; therefore, a negative association indicates that higher quality of life relates to higher engagement.



Patterns of Engagement

Figure 5 shows the percentage of participants actively using the app each day following download.

Intended Versus Actual Use

Figure 3 illustrates participants' intended app use at the time of download across 3 dimensions: minutes per session, sessions per day, and days per week.

Intended Use in Comparison With Actual Use

To address question 2, we investigated the relationship between intended and actual use. There was a weak associative relationship (r=0.094; P=.02). Figure 4 shows the density curve and median intended use stated as baseline (purple) compared with actual use over the first 30 days (green). Users intended a median of 300 (IQR 192.86-482.14) minutes over the first 30 days but only engaged for a median of 16 (IQR 0-74.51) minutes. In other words, median intended use was nearly 19 times greater than actual use. This suggests that users who intended to meditate more were more likely to do so (ρ=0.106; P=.007).

However, as expected, there was a large difference whereby intended use was significantly higher than actual use (mean difference 350.61, SD 493.53, 95% CI 312.72-388.51 min; Cohen d=0.710; t_{653}=18.168; $P<.001$; Figure 4).

Predictors of Engagement: User Factors (Question 2, Hypothesis 2)

Zero-order correlations for user factors are displayed in the forest plot below (Figure 6). Zero-order correlations, semipartial correlations, and regression beta weights are reported in Table 4.

XSL•FO
RenderX

Table . Zero-order correlation, semipartial correlations, and regression beta weights for predictors. * <.05, ** <.01 and *** <.001

Variable	ρ	sr[a]	β[b]
Mental health factors			
Satisfaction with life (SWLS[c])	0.123**	0.038	.008
Well-being (WEMWBS[d])	0.135**,***	0.020	.010
Perceived stress (PSS-4[e])	−0.107**	0.027	.020
Psychological distress (K10[f])	−0.138***	−0.070	−0.022
Quality of life (EQ-HWB-S[g])	−0.10*	0.037	<.001
Expectation factors			
Expectations for sleep	0.056	0.026	.001
Expectations for stress	0.067	−0.050	−0.012
Expectations for anxiety	0.102*	0.063	.014
Expectations for attention/focus	0.091*	0.014	.003
Personality factors			
BFI2[h]—conscientiousness	0.124**	0.055	.002
BFI2—neuroticism	−0.103*	0.033	.010
Readiness to change	0.195***	0.146	.012

[a]sr: semipartial correlation.

[b]R^2=0.041.

[c]SWLS: Satisfaction with Life Survey.

[d]WEMWBS: Warwick-Edinburgh Mental Wellbeing Scale.

[e]PSS-4: Perceived Stress Scale-4.

[f]K10: 10-item Kessler Psychological Distress Scale.

[g]EQ-HWB-S: EuroQol Health and Wellbeing Scale.

[h]BFI2: Big Five Inventory Short Form 2 (30-item).

Mental Health Factors

All mental health factors were negatively associated with engagement in the zero-order correlations, indicating that poorer mental health, regardless of scale direction, was linked to lower engagement (Table 4). Only psychological distress showed a negative semipartial relationship with engagement that approached significance (sr=−0.070; *P*=.09). Perceived stress was the only significant mental health factor in the multiple regression, showing a positive association to engagement (β=.02; *P*=.04), contrasting with its zero-order correlation.

User Factors

Readiness to change showed the largest zero-order and semipartial correlation (ρ=0.194, *P*<.001; and sr=0.156; *P*<.001). Of the personality factors, conscientiousness and neuroticism (positive relationship) and neuroticism (negative relationship) correlated with engagement (ρ=0.125, *P*=.002; and ρ=−0.102, *P*=.009). Expectations for anxiety (ρ=0.102; *P*=.01) and attention (ρ=0.091; *P*=.02) were weakly positively associated with engagement. No user factors were significant in the regression model.

Regression Models With PCA Components

Multicollinearity was observed within the regression (Figure S2 in Multimedia Appendix 1). As a result, we undertook PCAs for mental health and expectations factors. Both mental health and expectations each showed one central factor with high loadings from contributing factors that accounted for a majority of the variance (Tables S9-S12 in Multimedia Appendix 1). Regression models were conducted using the mental health and expectations components found in the PCA (Tables S6-S8 in Multimedia Appendix 1).

With both components included, no factor or component achieved significance, and the amount of variance was reduced substantially (R^2=0.02; Table S7 in Multimedia Appendix 1). In a regression containing the expectations component, only perceived stress was significant (β=.022; *P*=.03), and the expectations component was nonsignificant, in line with prior results (Table S8 in Multimedia Appendix 1). In a regression containing the mental health component, the mental health component (β=−0.008; *P*=.049), readiness to change (β=.011; *P*=.048), and expectations for anxiety (β=.015; *P*=.049) were significant (Table 5). Deviating from the main model, readiness to change and expectations for anxiety were significant positive predictors when controlling for overlap among measures of

XSL•FO
RenderX

mental health. Furthermore, the mental health component, representing mental ill health, also negatively predicted engagement. The negative relationship between mental ill health and engagement suggests that worse mental health results in lower engagement.

Table . Regression model with mental health component only. *<.05 and *** <.001

Variable	Standardized estimate[a,b]	SE	*t* value	Pr (>|t|)
Intercept	−0.313***	0.012	−26.832	<0.001
User factors				
Mental health component	−0.008*	0.003	−1.976	0.049
Conscientiousness (BFI)[c]	0.001	0.006	0.241	0.81
Neuroticism (BFI)	0.008	0.008	0.998	0.32
Readiness to change	0.011*	0.006	1.983	0.048
Expectations for sleep	0.001	0.008	0.137	0.89
Expectations for stress	−0.013	0.009	−1.441	0.15
Expectations for anxiety	0.015*	0.008	1.976	0.049
Expectations for attention/focus	0.002	0.007	0.352	0.73

[a]R^2=0.027.

[b]R=0.015.

[c]BFI: Big Five Inventory.

Discussion

Principal Findings

This study investigates low rates of meditation app engagement in prior literature by examining prospective predictors of use. Minutes of use in the first 30 days after download were compared with user characteristics among 668 participants. While our findings align with prior works [4], our data showed a more gradual decline in use than has previously been observed. Actual use departed from intended use, with participants meditating about 350 minutes less (on average) than planned. Most participants aimed to meditate for 10 minutes daily but only engaged in approximately one 10-minute session. Nevertheless, there was a correlation between intended and actual use, suggesting that those who intended to meditate more practiced slightly more. The gap between intended and actual meditation time reinforces prior findings regarding the intention-behavior gap and engagement challenges with digital interventions [4,19-21,69].

Contrary to hypotheses, perceived stress was the only significant predictor of engagement in the regression model. Expectations for anxiety (positive) and readiness to change (positive) reached significance when modeling mental health as a single component. Higher stress may motivate behavioral change, reinforced by beliefs in meditation's efficacy for anxiety to manage acute stress. Our findings suggest that baseline characteristics, including stress, expectations, and readiness to change, predict initial meditation app use, implicating these factors in habit-building as well as drop-off.

Mental Health

As expected, measures of mental ill health, such as distress and perceived stress, were associated with lower engagement, while measures of positive mental health, such as satisfaction with life and well-being, were associated with greater engagement. Depression is related to low adherence to, and attrition from, meditation and health interventions [54]. This may reflect a broader message about the ability for people experiencing mental ill health to engage in meditation apps. People with significant psychological distress and psychoticism may disengage prematurely due to the increased likelihood of meditation-related adverse events [70]. Mental ill health also often reduces executive functioning, cognitive functioning, and affective functioning, which can interfere with one's ability to set goals or follow through on health-related goals or behaviors [71]. A reduction from normal functioning is itself a key criterion for many psychological disorders. This is not a new issue for engagement in mental health interventions but underscores that mental health difficulties hinder the meaningful engagement required to experience change. Having good mental well-being, on the other hand, may relate to greater energy, motivation, and self-efficacy when making behavior change, all of which are either further promoted and reinforced by consistent meditation app use. People scoring higher on well-being and satisfaction may also face less difficulty with the act of meditation, as they may have less difficulty dealing with negative thoughts or experience fewer negative thoughts overall [53].

While not mental health per se, consistent with this pattern, lower quality of life is associated with lower engagement. This means health issues in general may mean more difficulty in engaging, although barriers to health care motivate initial engagement [33]. Low quality of life may result in a lower capacity to self-regulate, plan, or cope with challenges [72].

The relationship between perceived stress shifted direction from negative (on its own) to positive (controlling for other variables),

XSL•FO
RenderX

suggesting that there may be a unique contribution of acute stress unrelated to mental health. The perceived stress scale potentially reflects different latent variables when presented on its own (acute stress) versus as part of a mental health component variable (chronic stress) [73]; therefore, controlling for mental health may remove shared variance related to chronic stress, leaving the remaining variance to reflect more acute experiences. While mental health and chronic stress may present a barrier to engagement [54,74], our most robust finding shows that higher perceived stress drives higher engagement. Meditation apps have shown preliminary efficacy for perceived stress [75], and population-level work shows that people may seek out meditation to address unmet needs in mental health [33].

Our nonclinical population reported high levels of perceived stress and psychological distress relative to normative samples. Individuals may have downloaded the meditation app as a step toward alleviating their distress but may have found it harder to engage with significant mental ill health or chronic stress, reflected in the negative relationship between the mental health component and engagement. Given the moderate effects of meditation apps on perceived stress [14], some level of perceived relief may be felt early, promoting further engagement. Preliminary evidence suggests one session of meditation may be enough to produce at least a temporary reduction in perceived stress [76].

Readiness and Expectations

Readiness to change accounted for the most unique variance and approached significance in the regression model, showing the strongest zero-order correlation with engagement. However, robust regression downweights outliers, which contribute to discrepancies between semipartial correlations and regression beta weights, may have reduced the contributions of readiness to change in the regression model (Table 5). Readiness was no longer significantly associated when controlling for other factors but significant with a single mental health component. When mental ill health and/or poor well-being were modeled as one component, variance potentially capturing desire to address life dissatisfaction and/or acute stress might then best be captured by readiness to change. The meaningful positive association of readiness to change is consistent with prior work showing that readiness to change is a central facet of behavior change [7,65] and that intrinsic motivation predicts subsequent behavior [32], and that intrinsic motivation predicts subsequent behavior [32].

Out of the expectations domains, only expectations for anxiety accounted for substantial unique variance and remained significant when modeling one mental health component. This may indicate that alleviating anxiety motivates app-based meditation and aligns with mental health as a common motivator for commencing meditation [6]. While it could not be included in the main regression model due to limited responses (n=221), expectation match showed the strongest correlation with engagement. This indicates that while expectations are important, users may discontinue practice if benefits are not observed or anxiety is potentially exacerbated [77], as app-based meditation shows smaller effect sizes for anxiety than depression or perceived stress [14,77]. In other words, expectations interact with experience to influence behavior.

Personality

Conscientiousness and neuroticism were both significantly associated with engagement but were not significant in the regression model. This is in contrast to prior work that suggests agreeableness, neuroticism, and openness are related to meditation engagement [43]. It is logical that conscientiousness would predict higher engagement, given that conscientious traits are focused on self-discipline [55]. Previously, longitudinal studies found that higher baseline conscientiousness predicts higher adherence to meditation, which in turn predicts higher stress reduction [78]. However, one might also logically expect openness to be important in trying a new habit. It may be that since buy-in for meditation apps is much lower than, for example, signing up for an extended in-person course, openness matters less than simply the diligence of continuing (ie, conscientiousness).

Neuroticism negatively predicting engagement is consistent with our finding that mental health factors are negatively associated with engagement and the mental health component negatively predicts engagement in the regression. Higher neuroticism, or the tendency to experience negative emotion, relates to a higher perception of barriers to meditation, likely due to the impact of rumination on self-efficacy, readiness, and the subjective experience of meditation [42,46,79]. Difficult or negative thoughts are common challenges in beginning meditation [48], and people high in neuroticism may face more negative thoughts or perceive these thoughts as more of an obstacle. Sitting with one's thoughts may be more difficult for those with a more negative predisposition. Neuroticism is distinct from low mood or mental ill health, as it reflects an overall underlying tendency to negativity; however, neuroticism is a strong predictor of mental ill health, as neuroticism exacerbates stress and interferes with coping [80]. However, interventions for mental health can modestly reduce neuroticism [81,82], suggesting that persistence has positive outcomes, although it is important to differentiate between challenges that are part of the learning process and harmful adverse events [56].

While the prior study by Khwaja et al [43] investigated meditation app use via self-report usage logs, and our previous survey used objective app data but retrospective factors, this study uses objective app engagement data with a larger sample, which may be more reliable. Personality may play a smaller role in engagement than perceived stress and readiness to change during relatively short periods of monitoring use.

Overall, people are commencing meditation app use when they believe the app is effective and are cognizant of their stress. The significance of the mental health component, readiness, expectations, and perceived stress suggests a connection between engagement and awareness of mental ill health and/or acute stress, making a conscious decision to address it and believing that the intervention will work.

Limitations

This study was limited by a reduced sample size with power to detect effects equivalent to r=0.177. While it is possible that

XSL•FO
RenderX

we were underpowered, it is notable that other meditation and mental health app engagement studies have observed larger effect sizes, suggesting that the true effects are smaller than what has been previously suggested [14,17,37,53]. Generalizability is limited, as the study focused on a single meditation app, although this is a common approach [19,20]. This ensured design and quality consistency but restricts applicability to other apps. Future work could compare engagement levels across apps.

A technical error in the recording of tracks completed occurred during the study period. No significant difference was found in the relationship between completed and started tracks. Nonetheless, the glitch may have introduced additional noise, affecting our ability to detect relationships. The inclusion of started tracks in the engagement variable could have partially accounted for unrecorded tracks, but without knowing the percentage of each track completed, engagement duration would remain unclear.

Our model only accounted for a small amount of variance (R^2=0.041, 4%), compared to 13% to 24% found in other behavior change studies [45,83]. This suggests key factors are missing. However, prior work largely relied on cross-sectional reports or recall, relating more to baseline or intended behavior rather than objective engagement level. Other studies examining out-of-class meditation practice predicted 21% variance [45]. Unconsidered factors may include app preferences, unconsidered demographic differences, psychological traits, or contextual variables. Predicting meditation may also be harder than other health-related behaviors due to less short-term tangible reinforcement, higher perceived effort, and the intentional nature of the behavior as opposed to the habitual or automatic nature of other behavior change. Future research could build on the current and preceding studies to further refine predictors and improve model power.

Future Directions

Future work could compare 30-day follow-up data with initial findings and detect whether changes over time predict engagement. This could explore differences between expectations and perceptions after initial versus ongoing use. A greater gap between baseline and follow-up ratings might suggest that exceeded expectations may promote engagement, whereas expectations not met or matched may have a lesser influence. Notably, however, steep discontinuation rates may necessitate large sample sizes to make such work possible. If only 10% to 20% of users are still actively engaged after 30 days, samples would need to be substantial to include a large enough proportion of continuing users.

Additionally, follow-up data may allow for mobile app ratings such as the user Mobile Application Rating Scale, which would provide clearer insights into users' app perceptions after 3 days of use. Digital therapeutic alliance or the relationship between the user and the digital intervention [84] and user Mobile Application Rating Scale [85] engagement (ie, appeal) merits further investigation as they have some of the strongest zero-order correlations with engagement [7].

Conclusions

This study reveals that half of new meditation app users disengage within the first 2 weeks of use, and the majority disengage by the end of 30 days. While most users plan daily engagement, they tend to complete only a single 10-minute session. Perceived stress is the strongest predictor of engagement, while mental health, readiness to change, and expectations for anxiety management show predictive use. While most users disengage, those who are under stress, do not have pronounced mental ill health, are motivated to make changes, and/or have reasonably high expectations that a meditation app can reduce anxiety are more likely to continue using the app. Given the growing role of meditation apps in public mental health [3,4], understanding the motivations and expectations of meditation app users at baseline provides valuable insight into how and why meditation apps are being used in nonclinical populations.

Acknowledgments

This research was supported by the Contemplative Studies Centre, established by a Philanthropic Gift from the Three Springs Foundation, Pty, Ltd.

The authors would like to acknowledge the coding contributions of Alex Burger and Karen Trapani as well as the administrative support of Cathleen Benevento.

This study would not have been possible without the support of the Contemplative Studies Centre and the Melbourne School of Psychological Sciences community at the University of Melbourne.

No original content in the manuscript was generated by artificial intelligence. The authors occasionally used the free version of Grammarly for content flow.

Figure 1 was created using Microsoft Word and Miro workflow, and Figure 2 through 8 were created in R with additional editing completed in Canva.

Funding

Funding was provided for the Contemplative Studies Centre via a philanthropic gift from the Three Springs Foundation, Pty, Ltd.

XSL•FO
RenderX

Data Availability

Data and analytic code are uploaded to OSF [86]. Datasets and materials related to this study will be made available on the Open Science Framework.

Authors' Contributions

Conceptualization – JA, JD, NTVD, JG
Data curation – JA, JD, PW, NTVD
Formal analysis – JA, JD, NTVD
Funding acquisition – NTVD
Methodology – JA, JD, JG, NTVD
Resources – JA, JD, JG, NTVD
Visualization – JA, NTVD
Writing – original draft – JA, JG, NTVD
Writing – review & editing – JA, JG, SD'A, NTVD

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials that support the main analyses.
[DOCX File, 1394 KB - mhealth_v14i1e79366_app1.docx]

References

1. Collaborators GBDMD. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry 2022 Feb;9(2):137-150. [doi: 10.1016/S2215-0366(21)00395-3]
2. McBain R, Salhi C, Morris JE, Salomon JA, Betancourt TS. Disease burden and mental health system capacity: WHO Atlas study of 117 low- and middle-income countries. Br J Psychiatry 2012 Dec;201(6):444-450. [doi: 10.1192/bjp.bp.112.112318] [Medline: 23137730]
3. Lattie EG, Stiles-Shields C, Graham AK. An overview of and recommendations for more accessible digital mental health services. Nat Rev Psychol 2022 Feb;1(2):87-100. [doi: 10.1038/s44159-021-00003-1] [Medline: 38515434]
4. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567. [doi: 10.2196/14567] [Medline: 31573916]
5. Bowles NI, Davies JN, Strohmaier S, Galante J, Van Dam NT. Too much or not enough of a good thing? Reconciling differing perspectives on the optimal dose of mindfulness meditation and advancing a dose-response research agenda. JCS 2026:1-37. [doi: 10.57010/CAXS3076]
6. Bowles NI, Davies JN, Van Dam NT. Dose–response relationship of reported lifetime meditation practice with mental health and wellbeing: a cross-sectional study. Mindfulness (N Y) 2022 Oct;13(10):2529-2546. [doi: 10.1007/s12671-022-01977-6]
7. Adams J, Davies J, Wattanatakulchat P, et al. Engagement with meditation apps: cross-sectional survey of use and associations. J Med Internet Res 2026 Feb 2;28:e71960. [doi: 10.2196/71960] [Medline: 41627883]
8. Goldberg SB, Baldwin SA, Riordan KM, et al. Alliance with an unguided smartphone app: validation of the digital working alliance inventory. Assessment 2022 Sep;29(6):1331-1345. [doi: 10.1177/10731911211015310] [Medline: 34000843]
9. Mindfulness meditation apps market – latest report on the current trends in the industry. BioSpace. 2020. URL: https://www.biospace.com/mindfulness-meditation-apps-market-latest-report-on-the-current-trends-in-the-industry [accessed 2025-01-17]
10. Remskar M, Western MJ, Ainsworth B. Mindfulness improves psychological health and supports health behaviour cognitions: evidence from a pragmatic RCT of a digital mindfulness-based intervention. Br J Health Psychol 2024 Nov;29(4):1031-1048. [doi: 10.1111/bjhp.12745] [Medline: 39169217]
11. Strohmaier S. The relationship between doses of mindfulness-based programs and depression, anxiety, stress, and mindfulness: a dose-response meta-regression of randomized controlled trials. Mindfulness (N Y) 2020 Jun;11(6):1315-1335. [doi: 10.1007/s12671-020-01319-4]
12. Carmody J, Baer RA. How long does a mindfulness-based stress reduction program need to be? A review of class contact hours and effect sizes for psychological distress. J Clin Psychol 2009 Jun;65(6):627-638. [doi: 10.1002/jclp.20555] [Medline: 19309694]
13. Galante J, Friedrich C, Collaboration of Mindfulness Trials (CoMinT) 3, et al. Individual participant data systematic review and meta-analysis of randomised controlled trials assessing adult mindfulness-based programmes for mental health promotion in non-clinical settings. Nat Ment Health 2023 Jul 10;1(7):462-476. [doi: 10.1038/s44220-023-00081-5] [Medline: 37867573]

XSL•FO
RenderX

14. Gál É, tefan S, Cristea IA. The efficacy of mindfulness meditation apps in enhancing users' well-being and mental health related outcomes: a meta-analysis of randomized controlled trials. J Affect Disord 2021 Jan 15;279:131-142. [doi: 10.1016/j.jad.2020.09.134] [Medline: 33049431]
15. Sommers-Spijkerman M, Austin J, Bohlmeijer E, Pots W. New evidence in the booming field of online mindfulness: an updated meta-analysis of randomized controlled trials. JMIR Ment Health 2021 Jul 19;8(7):e28168. [doi: 10.2196/28168] [Medline: 34279240]
16. Liptáková S, Světlák M, Matis J, Slezáčková A, Rastislav R, Faculty of Social Studies, Masaryk University, Brno. Signing up is not yet mindfulness practice: a systematic review of adherence to eHealth and mHealth mindfulness-based programs in the pre-pandemic period. CESK PSYCHOL 2022;66(3):233-254. [doi: 10.51561/cspsych.66.3.233]
17. Lam SU, Kirvin-Quamme A, Goldberg SB. Overall and differential attrition in mindfulness-based interventions: a meta-analysis. Mindfulness (N Y) 2022 Nov;13(11):2676-2690. [doi: 10.1007/s12671-022-01970-z]
18. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009;11(2):e13. [doi: 10.2196/jmir.1194]
19. Huberty J, Vranceanu AM, Carney C, Breus M, Gordon M, Puzia ME. Characteristics and usage patterns among 12,151 paid subscribers of the calm meditation app: cross-sectional survey. JMIR Mhealth Uhealth 2019;7(11):e15648. [doi: 10.2196/15648]
20. Cearns M, Clark SR. The effects of dose, practice habits, and objects of focus on digital meditation effectiveness and adherence: longitudinal study of 280,000 digital meditation sessions across 103 countries. J Med Internet Res 2023 Sep 19;25:e43358. [doi: 10.2196/43358] [Medline: 37725801]
21. Mahalingham T, McEvoy PM, Clarke PJF. Assessing the validity of self-report social media use: evidence of no relationship with objective smartphone use. Comput Human Behav 2023 Mar;140:107567. [doi: 10.1016/j.chb.2022.107567]
22. Antezana G, Venning A, Smith D, Bidargaddi N. Do young men and women differ in well-being apps usage? Findings from a randomised trial. Health Informatics J 2022;28(1):14604582211064825. [doi: 10.1177/14604582211064825] [Medline: 35128952]
23. Jakob R, Harperink S, Rudolf AM, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022;24(5):e35371. [doi: 10.2196/35371]
24. Stull DE, Leidy NK, Parasuraman B, Chassany O. Optimal recall periods for patient-reported outcomes: challenges and potential solutions. Curr Med Res Opin 2009 Apr 1;25(4):929-942. [doi: 10.1185/03007990902774765]
25. Popov V, Reder LM. Frequency effects on memory: a resource-limited theory. Psychol Rev 2020 Jan;127(1):1-46. [doi: 10.1037/rev0000161] [Medline: 31524424]
26. Aarts H, Dijksterhuis A. How often did I do it? Experienced ease of retrieval and frequency estimates of past behavior. Acta Psychol (Amst) 1999 Nov;103(1-2):77-89. [doi: 10.1016/S0001-6918(99)00035-9]
27. Chaudri NA. Adherence to long-term therapies evidence for action. Ann Saudi Med 2004;24(3):221-222. [doi: 10.5144/0256-4947.2004.221]
28. Lally P, Gardner B. Promoting habit formation. Health Psychol Rev 2013 May;7(sup1):S137-S158. [doi: 10.1080/17437199.2011.603640]
29. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572. [doi: 10.2196/17572] [Medline: 32348255]
30. Gardner B, Lally P, Wardle J. Making health habitual: the psychology of "habit-formation" and general practice. Br J Gen Pract 2012 Dec;62(605):664-666. [doi: 10.3399/bjgp12X659466] [Medline: 23211256]
31. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
32. Osin EN, Turilina II. Mindfulness meditation experiences of novice practitioners in an online intervention: trajectories, predictors, and challenges. Appl Psychol Health Well Being 2022 Feb;14(1):101-121. [doi: 10.1111/aphw.12293] [Medline: 34268871]
33. Davies JN, Faschinger A, Galante J, Van Dam NT. Prevalence and 20-year trends in meditation, yoga, guided imagery and progressive relaxation use among US adults from 2002 to 2022. Sci Rep 2024 Jul 1;14(1):14987. [doi: 10.1038/s41598-024-64562-y] [Medline: 38951149]
34. van der Heide I, Wang J, Droomers M, Spreeuwenberg P, Rademakers J, Uiters E. The relationship between health, education, and health literacy: results from the Dutch Adult Literacy and Life Skills Survey. J Health Commun 2013;18 Suppl 1(Suppl 1):172-184. [doi: 10.1080/10810730.2013.825668] [Medline: 24093354]
35. Jiwani Z, Tatar R, Dahl C, et al. Examining equity in access and utilization of a freely available meditation app. Npj Ment Health Res 2023;2(1):5. [doi: 10.1038/s44184-023-00025-y] [Medline: 37159797]
36. Olano HA, Kachan D, Tannenbaum SL, Mehta A, Annane D, Lee DJ. Engagement in mindfulness practices by U.S. adults: sociodemographic barriers. J Altern Complement Med 2015 Feb;21(2):100-102. [doi: 10.1089/acm.2014.0269] [Medline: 25685958]
37. Lam SU, Xie Q, Goldberg SB. Situating meditation apps within the ecosystem of meditation practice: population-based survey study. JMIR Ment Health 2023 Apr 28;10:e43565. [doi: 10.2196/43565] [Medline: 37115618]

XSL•FO
RenderX

38. Bhuiyan N, Puzia M, Stecher C, Huberty J. Associations between rural or urban status, health outcomes and behaviors, and COVID-19 perceptions among meditation app users: longitudinal survey study (preprint). JMIR Mhealth Uhealth. Preprint posted online on 2020. [doi: 10.2196/preprints.26037]
39. Laurie J, Blandford A. Making time for mindfulness. Int J Med Inform 2016 Dec;96:38-50. [doi: 10.1016/j.ijmedinf.2016.02.010] [Medline: 26965526]
40. Lomas T, Cartwright T, Edginton T, Ridge D. A qualitative analysis of experiential challenges associated with meditation practice. Mindfulness (N Y) 2015 Aug;6(4):848-860. [doi: 10.1007/s12671-014-0329-8]
41. Altizer CC, Ferrell BT, Natale AN. Mindfulness and personality: more natural for some than others and how it matters. Consulting Psychology Journal: Practice and Research 2021;73(1):51-64. [doi: 10.1037/cpb0000189]
42. Borghouts J, Eikey E, Mark G, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387. [doi: 10.2196/24387] [Medline: 33759801]
43. Khwaja M, Pieritz S, Faisal AA, Matic A. Personality and engagement with digital mental health interventions. Presented at: UMAP '21: Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization; Jun 21-25, 2021; Utrecht Netherlands p. 235-239. [doi: 10.1145/3450613.3456823]
44. Alqahtani F, Meier S, Orji R. Personality-based approach for tailoring persuasive mental health applications. User Model User-Adap Inter 2022 Jul;32(3):253-295. [doi: 10.1007/s11257-021-09289-5]
45. Canby NK, Eichel K, Peters SI, Rahrig H, Britton WB. Predictors of out-of-class mindfulness practice adherence during and after a mindfulness-based intervention. Psychosom Med 2021;83(6):655-664. [doi: 10.1097/PSY.0000000000000873] [Medline: 33038188]
46. Whitford S, Warren K. Perceived barriers to meditation among college students: the role of personality traits. BHAC 2019;3(1):23-33. [doi: 10.18061/bhac.v3i1.6678]
47. Hunt CA, Hoffman MA, Mohr JJ, Williams AL. Assessing perceived barriers to meditation: the Determinants of Meditation Practice Inventory-Revised (DMPI-R). Mindfulness (N Y) 2020 May;11(5):1139-1149. [doi: 10.1007/s12671-020-01308-7] [Medline: 33664878]
48. Banerjee A, Banerji R, Berry J, et al. From proof of concept to scalable policies: challenges and solutions, with an application. Journal of Economic Perspectives 2017 Nov 1;31(4):73-102. [doi: 10.1257/jep.31.4.73]
49. Armitage CJ, Norman P, Alganem S, Conner M. Expectations are more predictive of behavior than behavioral intentions: evidence from two prospective studies. Ann Behav Med 2015 Apr;49(2):239-246. [doi: 10.1007/s12160-014-9653-4]
50. Michalco J, Simonsen JG, Hornbæk K. An exploration of the relation between expectations and user experience. Int J Hum Comput Interact 2015 Sep 2;31(9):603-617. [doi: 10.1080/10447318.2015.1065696]
51. Bhattacherjee A. Understanding information systems continuance: an expectation-confirmation model. MIS Q 2001 Sep 1;25(3):351-370. [doi: 10.2307/3250921]
52. Cheng VWS, Davenport T, Johnson D, Vella K, Hickie IB. Gamification in apps and technologies for improving mental health and well-being: systematic review. JMIR Ment Health 2019 Jun 26;6(6):e13717. [doi: 10.2196/13717] [Medline: 31244479]
53. Crandall A, Cheung A, Young A, Hooper AP. Theory-based predictors of mindfulness meditation mobile app usage: a survey and cohort study. JMIR Mhealth Uhealth 2019 Mar 22;7(3):e10794. [doi: 10.2196/10794] [Medline: 30900992]
54. Prince M, Patel V, Saxena S, et al. No health without mental health. The Lancet 2007 Sep;370(9590):859-877. [doi: 10.1016/S0140-6736(07)61238-0]
55. Soto CJ, John OP. The next Big Five Inventory (BFI-2): developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. J Pers Soc Psychol 2017 Jul;113(1):117-143. [doi: 10.1037/pspp0000096] [Medline: 27055049]
56. Britton WB, Lindahl JR, Cooper DJ, Canby NK, Palitsky R. Defining and measuring meditation-related adverse effects in mindfulness-based programs. Clin Psychol Sci 2021 Nov;9(6):1185-1204. [doi: 10.1177/2167702621996340]
57. Brazier J, Peasgood T, Mukuria C, et al. The EQ-HWB: overview of the development of a measure of health and wellbeing and key results. Value Health 2022 Apr;25(4):482-491. [doi: 10.1016/j.jval.2022.01.009]
58. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [doi: 10.2307/2136404] [Medline: 6668417]
59. Kessler RC, Andrews G, Colpe LJ, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002 Aug;32(6):959-976. [doi: 10.1017/S0033291702006074]
60. Tennant R, Hiller L, Fishwick R, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Dec;5(1):63. [doi: 10.1186/1477-7525-5-63]
61. Cheung F, Lucas RE. Assessing the validity of single-item life satisfaction measures: results from three large samples. Qual Life Res 2014 Dec;23(10):2809-2818. [doi: 10.1007/s11136-014-0726-4]
62. Zimmerman GL, Olsen CG, Bosworth MF. A "stages of change" approach to helping patients change behavior. Am Fam Physician 2000 Mar 1;61(5):1409-1416. [Medline: 10735346]
63. Hesse M. The Readiness Ruler as a measure of readiness to change poly-drug use in drug abusers. Harm Reduct J 2006 Dec;3(1):3. [doi: 10.1186/1477-7517-3-3]

XSL•FO
RenderX

64. McConnaughy EA, Prochaska JO, Velicer WF. Stages of change in psychotherapy: measurement and sample profiles. Psychotherapy: Theory, Research & Practice 1983;20(3):368-375. [doi: 10.1037/h0090198]
65. Moyers TB, Martin T, Houck JM, Christopher PJ, Tonigan JS. From in-session behaviors to drinking outcomes: a causal chain for motivational interviewing. J Consult Clin Psychol 2009 Dec;77(6):1113-1124. [doi: 10.1037/a0017189] [Medline: 19968387]
66. Kang H. Sample size determination and power analysis using the G*Power software. J Educ Eval Health Prof 2021;18:17. [doi: 10.3352/jeehp.2021.18.17] [Medline: 34325496]
67. Warttig SL, Forshaw MJ, South J, White AK. New, normative, English-sample data for the Short Form Perceived Stress Scale (PSS-4). J Health Psychol 2013 Dec;18(12):1617-1628. [doi: 10.1177/1359105313508346]
68. Slade T, Grove R, Burgess P. Kessler Psychological Distress Scale: normative data from the 2007 Australian National Survey of Mental Health and Wellbeing. Aust N Z J Psychiatry 2011 Apr;45(4):308-316. [doi: 10.3109/00048674.2010.543653]
69. Agudo-Peregrina Á, Hernández-García Á, Pascual-Miguel FJ. Behavioral intention, use behavior and the acceptance of electronic learning systems: differences between higher education and lifelong learning. Comput Human Behav 2014 May;34:301-314. [doi: 10.1016/j.chb.2013.10.035]
70. Van Dam NT, Targett J, Davies JN, Burger A, Galante J. Incidence and predictors of meditation-related unusual experiences and adverse effects in a representative sample of meditators in the United States. Clin Psychol Sci 2025 May;13(3):632-648. [doi: 10.1177/21677026241298269]
71. Galderisi S, Heinz A, Kastrup M, Beezhold J, Sartorius N. Toward a new definition of mental health. World Psychiatry 2015 Jun;14(2):231-233. [doi: 10.1002/wps.20231] [Medline: 26043341]
72. Jiwani Z, Lam SU, Richard JD, Goldberg SB. Motivation for meditation and its association with meditation practice in a national sample of internet users. Mindfulness (N Y) 2022 Oct;13(10):2641-2651. [doi: 10.1007/s12671-022-01985-6] [Medline: 36506892]
73. Harris KM, Gaffey AE, Schwartz JE, Krantz DS, Burg MM. The perceived stress scale as a measure of stress: decomposing score variance in longitudinal behavioral medicine studies. Ann Behav Med 2023 Sep 13;57(10):846-854. [doi: 10.1093/abm/kaad015] [Medline: 37084792]
74. Creswell D, Brown KW, Cohen S, et al. Does high perceived stress over the past month alter cortisol reactivity to the Trier Social Stress Test? Psychoneuroendocrinology 2025 Feb;172:107256. [doi: 10.1016/j.psyneuen.2024.107256] [Medline: 39732087]
75. Prasad K, Wahner-Roedler DL, Cha SS, Sood A. Effect of a single-session meditation training to reduce stress and improve quality of life among health care professionals: a “dose-ranging” feasibility study. Altern Ther Health Med 2011;17(3):46-49. [Medline: 22164812]
76. Johnson S, Gur RM, David Z, Currier E. One-session mindfulness meditation: a randomized controlled study of effects on cognition and mood. Mindfulness (N Y) 2015 Feb;6(1):88-98. [doi: 10.1007/s12671-013-0234-6]
77. Farias M, Maraldi E, Wallenkampf KC, Lucchetti G. Adverse events in meditation practices and meditation-based therapies: a systematic review. Acta Psychiatr Scand 2020 Nov;142(5):374-393. [doi: 10.1111/acps.13225] [Medline: 32820538]
78. de Vibe M, Solhaug I, Tyssen R, et al. Does personality moderate the effects of mindfulness training for medical and psychology students? Mindfulness (N Y) 2015 Apr;6(2):281-289. [doi: 10.1007/s12671-013-0258-y]
79. Ryakhovskaya Y, Van Dam NT, Smillie LD. Who wants to be mindful? Personality predictors of meditation practice. Mindfulness (N Y) 2025 Feb;16(2):525-543. [doi: 10.1007/s12671-025-02526-7]
80. Ormel J, Jeronimus BF, Kotov R, et al. Neuroticism and common mental disorders: meaning and utility of a complex relationship. Clin Psychol Rev 2013 Jul;33(5):686-697. [doi: 10.1016/j.cpr.2013.04.003] [Medline: 23702592]
81. Aichele SR, Sahdra BK, Zanesco AP, et al. Meditation training and enduring changes in psychological functioning: a 7-year longitudinal study. J Posit Psychol 2025 May 4;20(3):544-557. [doi: 10.1080/17439760.2024.2378700]
82. Norris CJ, Creem D, Hendler R, Kober H. Brief mindfulness meditation improves attention in novices: evidence from ERPs and moderation by neuroticism. Front Hum Neurosci 2018;12. [doi: 10.3389/fnhum.2018.00315]
83. McEachan RRC, Conner M, Taylor NJ, Lawton RJ. Prospective prediction of health-related behaviours with the Theory of Planned Behaviour: a meta-analysis. Health Psychol Rev 2011 Sep;5(2):97-144. [doi: 10.1080/17437199.2010.521684]
84. Henson P, Wisniewski H, Hollis C, Keshavan M, Torous J. Digital mental health apps and the therapeutic alliance: initial review. BJPsych Open 2019 Jan;5(1):e15. [doi: 10.1192/bjo.2018.86] [Medline: 30762511]
85. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016;4(2):e72. [doi: 10.2196/mhealth.5849]
86. Adams J, Davies J, Wattanatakulchat P, et al. User and app related factors associated with engagement with mindfulness apps: a longitudinal study of Medito app users. OSF. 2024. URL: https://osf.io/3v897/overview [accessed 2026-04-21]

Abbreviations

K10: Kessler Psychological Distress Scale
MBP: mindfulness-based programs

XSL•FO
RenderX

PCA: principal component analysis

Edited by L Buis; submitted 22.Jun.2025; peer-reviewed by C Callahan, S LAU, S Kunkle; revised version received 24.Feb.2026; accepted 25.Feb.2026; published 08.Jun.2026.

Please cite as:
Adams J, Davies J, Wattanatakulchat P, Galante J, D'Alfonso S, Van Dam NT
Real-World Meditation App Engagement: Longitudinal Study of the Medito Meditation App
JMIR Mhealth Uhealth 2026;14:e79366
URL: https://mhealth.jmir.org/2026/1/e79366
doi:10.2196/79366

© Julia Alice Adams, Jonathan Davies, Prai Wattanakulchat, Julieta Galante, Simon D'Alfonso, Nicholas T Van Dam. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 8.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Managing BMI and Emotional Distress Using mHealth: Nationally Representative Survey Study

Ranran Z Mi[1], PhD; Fei Shen[2], PhD; Julia Strugala[2], BA

[1]Department of Communication, Media and Journalism, Kean University, 1000 Morris Ave, Union, NJ, United States

[2]School of Psychology, Kean University, Union, NJ, United States

Corresponding Author:
Ranran Z Mi, PhD
Department of Communication, Media and Journalism, Kean University, 1000 Morris Ave, Union, NJ, United States

Abstract

Background: Mobile health (mHealth) technologies, including smartphone health apps and wearable trackers, are increasingly used to promote health behaviors. However, their impact on physical and mental well-being remains complex, with both benefits and potential unintended negative consequences.

Objective: This study aimed to examine the relationship between mHealth use (ie, health app and wearable tracker) and 2 health outcomes (BMI and emotional distress), as well as the mediating roles of healthy eating, sleep, and physical activity based on a representative sample.

Methods: We analyzed data from a nationally representative sample of US adults aged 33 to 43 years (N=1931). Chi-square tests and 1-way ANOVA were used to compare demographic differences between mHealth users and nonusers. A path model examined the relationship between mHealth use (ie, smartphone health apps and wearable trackers) and health outcomes (ie, BMI and emotional distress), with lifestyle factors (ie, healthy eating, physical activity, and sleep) as mediators. Mediation analyses tested indirect effects through these lifestyle factors.

Results: mHealth users were more likely to be female, married, have higher levels of education and income, and have health insurance. The primary use of mHealth was the management of physical activity. Smartphone health app use positively correlated with wearable tracker use (β=.394; *P*<.001). Smartphone health app use predicted greater BMI (β=.068; *P*=.006), whereas wearable tracker use did not significantly predict BMI. Smartphone health app use was unrelated to emotional distress, while wearable tracker use was associated with lower emotional distress (β=–.074; *P*=.003). Mediation analyses showed that physical activity negatively mediated the relationships between both types of mHealth use and health outcomes, indicating that mHealth users were more physically active, which was linked to lower BMI and emotional distress. Sleep hours mediated only the association between wearable tracker use and health outcomes, such that greater tracker use was related to fewer sleep hours, predicting higher BMI and emotional distress. Finally, healthy eating mediated only the associations between mHealth use and emotional distress, suggesting that healthier dietary behaviors among mHealth users contributed to lower emotional distress.

Conclusions: mHealth technologies can potentially promote healthier behaviors, but their effectiveness depends on users taking the initiative to sustain lifestyle changes. While wearable trackers may aid in mental well-being, their association with reduced sleep warrants further investigation.

(JMIR Mhealth Uhealth 2026;14:e75499) doi:10.2196/75499

KEYWORDS

mobile health; wearable trackers; eating; dietary; physical activity; sleep; emotional distress; body mass index

Introduction

Background

Mobile health (mHealth) technologies, including smartphone apps and wearable trackers, have become increasingly popular tools for managing personal health and wellness. Approximately half of the US population has at least 1 mHealth app on their mobile devices [1], and 21% of US adults regularly use a wearable tracker or smartwatch, a number that has increased due to the COVID-19 pandemic and the subsequently growing fitness culture [2]. While previous studies have highlighted socioeconomic disparities in mHealth adoption [3], these technologies continue to expand in popularity due to their convenience, personalized nature, and growing integration with health care systems [4,5]. Research increasingly supports mHealth's potential for managing health outcomes, such as weight management [6] and emotional regulation [7].

Despite the growing body of evidence, a significant gap remains: studies that aim to assess the effectiveness of mHealth technology either focus on behavioral lifestyle as the end

XSL•FO
RenderX

outcome or examine physical or mental health outcomes without exploring the underlying mechanism. It is critical to empirically investigate whether short-term behavioral changes induced by mHealth technology are linked to long-term health outcomes. Therefore, a more comprehensive analysis is needed to clarify the direct and indirect associations between mHealth use and health outcomes. By examining lifestyle factors such as diet, sleep, and physical activity as mediators, this study aims to provide deeper insights into the potential mechanisms driving these relationships. On the basis of a nationally representative sample of middle-aged US adults, this study explores the relationships between mHealth use (ie, health apps and wearables) and both physical and mental health outcomes (ie, BMI and emotional distress), while also examining the mediating roles of lifestyle factors (ie, diet, sleep, and physical activity).

mHealth Technology Use

Smartphone health apps and wearable trackers have seen increased adoption, with distinct demographic patterns among their users. Research shows that age and socioeconomic status play a critical role in mHealth adoption rates. For instance, these technologies are predominantly used by younger individuals with higher levels of education and income who come from more affluent households [8]. A gender difference also stands out in mHealth use, as research has found that women are 2.3 times more likely to use wearable health trackers, a trend further substantiated by data from the Pew Research Center [2,9].

Over a decade of research has reported that health apps and wearable trackers are used for various health purposes such as weight management, stress coping, substance use disorder recovery support, chronic condition monitoring, telemedicine support, and electronic medical records [10-13]. The types of health aspects managed through these technologies are expanding to lifestyle domains, including documenting dietary habits, monitoring physical activity, and tracking sleep patterns [14,15]. By providing real-time feedback, mHealth tools encourage users to reflect on their behaviors and make informed decisions, leading to better management of various health aspects.

mHealth and Weight Management

Weight management is a critical focus of mHealth technologies due to its profound implications for physical well-being. A systematic review found that mHealth technologies had a small but statistically significant effect on BMI reduction in children and adolescents [16]. While research supports the association between mHealth technology use and weight management, this relationship is likely mediated by lifestyle factors such as eating habits [17], sleep patterns [18], and physical activities [19,20]. For instance, app use is associated with healthy eating motives, decisions, and behaviors [21,22]. Wearable technologies further enhance self-monitoring by automatically detecting eating behaviors and tracking dietary intake [23,24]. Additionally, mHealth tools help reduce sedentary behavior and encourage physical activity [9]. A meta-analysis found that the use of activity trackers was associated with an average daily increase of 1800 steps, 40 minutes of walking, and 1 kg weight loss [25]. Sleep tracking is also designed as a key feature of mobile devices, aiming to improve sleep quality. Sleep-tracking features in mHealth tools enable users to identify patterns, address sleep issues, and access actionable insights that improve sleep and related health outcomes [26,27]. Evidence, although limited, shows wearable tracker use improves sleep quality among the health population [28].

mHealth and Emotional Distress

Previous research has shown that mHealth technologies can monitor mental health and alleviate emotional distress [29,30], with users reporting a greater sense of responsibility for their well-being and confidence in the accuracy of app-generated assessments [31]. Additional studies have suggested that engagement with mood-tracking apps can potentially improve users' ability to cope with unpleasant emotions and thoughts [32].

Additional studies suggest that lifestyle factors, especially physical activity and dietary habits, mediate these effects [33-36]. For example, a US national survey found that wearable users reported lower psychological distress, attributing this to longer workouts [37]. A randomized trial found that mindfulness app users, compared to nonusers, reported reduced emotional eating and stress levels, suggesting mHealth contributes to stress reduction by reducing emotional eating [38]. Additionally, mobile devices can track meal timing, location, and social context to encourage communal eating and reduce skipping meals, which correlates with elevated stress, anxiety, and depression [39]. Research on the mediating role of sleep in the relationship between mHealth use and emotional well-being remains limited. While one study suggests that sleep app use is positively associated with perceived well-being [40], it remains unclear whether actual sleep serves as a mediating factor in this process.

Taken together, the literature provides consistent evidence for meaningful relationships between mHealth and health (ie, BMI and emotional distress), as well as how lifestyle behaviors might play roles in such relationships. Detangling the lifestyle mechanisms will inform strategies for optimizing mHealth tools to improve overall health and guide future interventions. In tackling the gaps, we posed four hypotheses (H): (1) H1: the use of (a) smartphone health apps and (b) wearable trackers is associated with lower BMI; (2) H2: the use of (a) smartphone health apps and (b) wearable trackers is associated with lower emotional distress; (3) H3: healthy eating, sleep hours, and physical activities mediate the relationships between (a) the use of smartphone health apps and BMI and (b) the use of wearable trackers and BMI; and (4) H4: healthy eating, sleep hours, and physical activities mediate the relationships between (a) the use of smartphone health apps and emotional distress and (b) the use of wearable trackers and emotional distress.

Methods

Participants

Data were retrieved from the National Longitudinal Study of Adolescent to Adult Health (Add Health), a 5-wave nationally representative sample of adolescents' health and health-related behaviors in the United States. Add Health uses a complex,

XSL•FO
RenderX

stratified, and cluster sampling design [41]. More than 20,000 adolescents in grades 7 to 12 (aged 12 - 19 years) from more than 80 high schools and 52 feeder middle schools were interviewed at school and home in wave I (1994-1995). The participants' backgrounds were diverse in terms of locations, school types, regions, and ethnicities. Participants were followed up in 1996 (wave II; aged 13 - 20 years), 2001 to 2002 (wave III; aged 18 - 26 years), 2008 (wave IV; aged 24 - 32 years), and 2016 to 2018 (wave V; aged 33 - 43 years).

Ethical Considerations

Permission to conduct a secondary analysis was approved by Kean University's Institutional Review Board. The original consent and institutional review board authorization included provisions for secondary data analyses; thus, additional consent was not required. In this study, we conducted analyses using the deidentified public-use data from wave V of the full Add Health sample (2016 - 2018). The final analytic sample consisted of 1931 participants who provided valid responses to health technology questions.

Measures

Smartphone Health App Use

Participants indicated whether they used smartphone apps to track or manage their health. Responses were coded as 1 for "yes" and 0 for "no."

Wearable Tracker Use

Participants reported whether they had ever used a fitness or activity tracker. Responses were coded as 1 for "yes" and 0 for "no."

Healthy Eating

Participants reported the frequency of (1) consuming food from fast-food restaurants and (2) drinking nondiet sweetened beverages over the past 7 days. The frequencies were summed and reverse-coded so that higher scores indicated healthier eating behaviors.

Sleep Hours

Participants reported their average number of sleep hours per day/night.

Physical Activity

Participants reported how often they engaged in various physical activities (eg, biking, weightlifting, swimming, and golfing) over the past 7 days. These frequencies were summed to create a composite physical activity score such that higher scores indicate a higher level of physical activity.

BMI

BMI was calculated based on participants' self-reported height and weight, following the formula BMI=weight (in kilograms)/height (in meters)^2 [42].

Emotional Distress

Emotional distress was measured using a set of items assessing participants' feelings over the past 7 days. Participants indicated how often each of the following statements was true for them on a 4-point Likert scale: (1) "I felt that I could not shake off the blues, even with help from my family and friends," (2) "I felt depressed," (3) "I was happy" (reverse-coded), (4) "I felt sad," and (5) "I felt that life was not worth living." Responses were scored on a scale ranging from 1="Never or rarely" to 4="Most of the time or all of the time." A composite score was calculated, with higher scores indicating greater emotional distress (α=.82).

Covariates

Participants' age at survey completion, biological sex, education, household income, race, and ethnicity were included as covariates in all analyses.

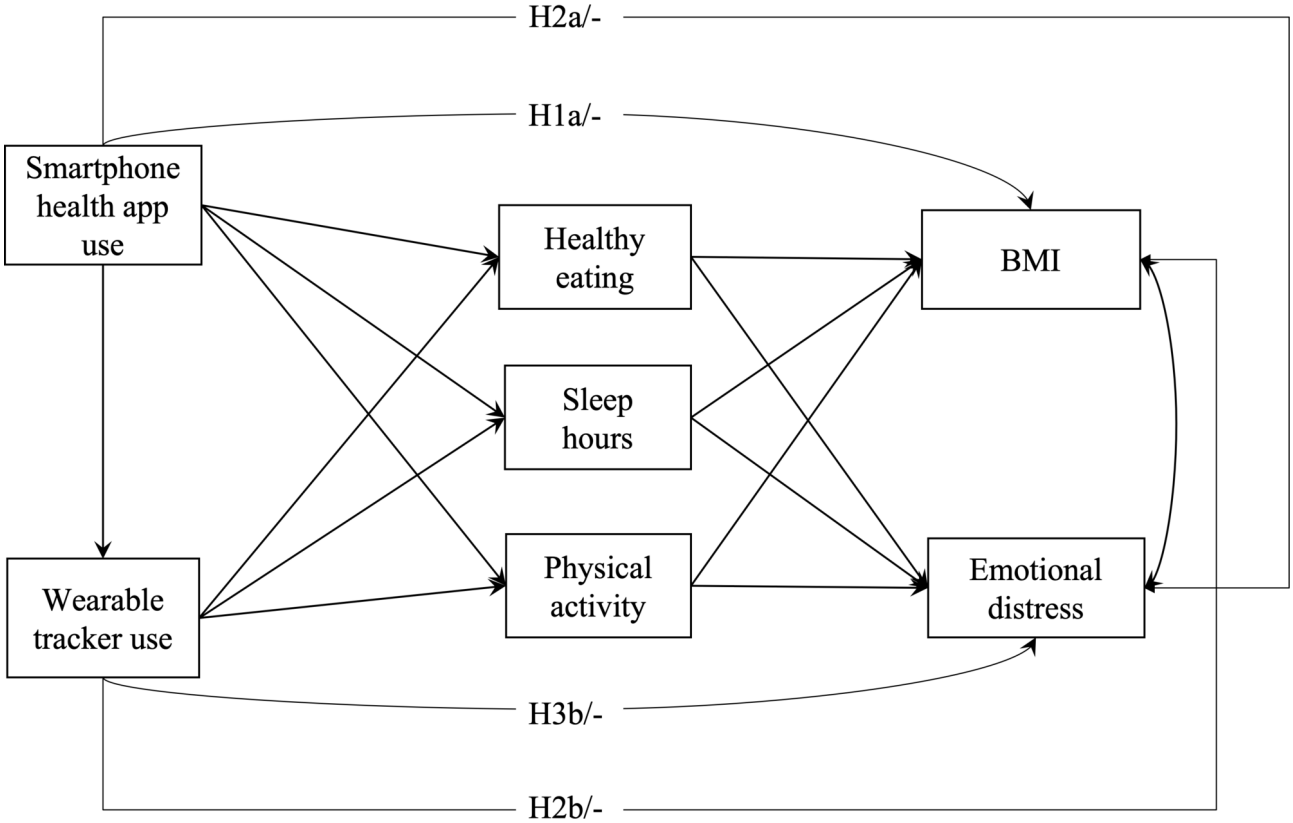
Analytic Approach

We used R (R Foundation for Statistical Computing) to perform all analyses. The "tableone" package was used to provide descriptive statistics and perform statistical comparisons between users and nonusers of mHealth [43]. More specifically, chi-square tests were performed on categorical variables, and 1-way ANOVA was used for continuous variables to compare differences.

To test the hypotheses, path analysis was performed using "lavaan" package [44]. The hypothesized paths are depicted in Figure 1. In the model, mobile technology use variables (ie, smartphone health app use and wearable tracker use) were specified as predictors of health outcomes (ie, BMI and emotional distress), with 3 lifestyle variables (ie, healthy eating, sleep hours, and physical activity) included as mediators. Mediation analyses were conducted to assess the indirect effects of mobile technology use on health outcomes through lifestyle mediators. All path coefficients were estimated using maximum likelihood with robust SEs to address nonnormality, and full information maximum likelihood estimation was used to handle missing data. Indirect effects were assessed using bias-corrected bootstrapped 95% CIs based on 5000 resamples. Age, sex, education, household income, race, and ethnicity were controlled for as covariates.

XSL•FO
RenderX

Figure 1. Hypothesized paths testing the relationship among mobile technology use, lifestyle mediators, and health outcomes. Note: H1: the use of (a) smartphone health apps and (b) wearable trackers is associated with lower BMI; H2: the use of (a) smartphone health apps and (b) wearable trackers is associated with lower emotional distress; H3: healthy eating, sleep hours, and physical activities mediate the relationships between (a) the use of smartphone health apps and BMI and (b) the use of wearable trackers and BMI; H4: healthy eating, sleep hours, and physical activities mediate the relationships between (a) the use of smartphone health apps and emotional distress and (b) the use of wearable trackers and emotional distress. H3 and H4 hypothesize that healthy eating, sleep hours, and physical activity mediate the relationship between mobile health use (smartphone health app use and wearable tracker use) and health outcomes (BMI and emotional distress).



As the public-use version of Add Health Wave V does not include design variables (weights, primary sampling units, or strata), all analyses were conducted without applying survey design variables. Given that the primary objective was to examine internal associations rather than to produce nationally representative estimates, unweighted path analyses were deemed appropriate, and the findings should be interpreted accordingly [45].

Results

Descriptives

Our analysis, based on a nationally representative sample, revealed distinct demographic trends among smartphone health app users and wearable tracker users. Specifically, smartphone health app users and wearable tracker users share the following characteristics in common: they are more likely to be female, married, have higher levels of education and income, and possess health insurance, compared with their nonuser counterparts. In addition, smartphone health app users are more likely to be younger, identify as liberal, and have 1 or no children compared to nonusers. Wearable tracker users are more likely to own the place where they live compared to their nonuser counterparts. Detailed participants' characteristics can be found in Table 1.

Participants use smartphones to manage a range of personal health aspects, with physical activity (714/842, 84.8%) being the most common, followed by diet management (399/838, 47.6%) and other aspects (Table 2). Female users (534/851, 62.7%) also reported using smartphones for menstrual cycle tracking (182/534, 34.1%) and pregnancy tracking (57/534, 11%). See Table 2 for the full list of health management aspects. Descriptive statistics and a correlation matrix for endogenous variables in the path model were reported in Table 3.

XSL•FO
RenderX

Table . Participant characteristics (N=1931)[a].

	Overall	Smartphone health app use			Wearable tracker use		
		No (n=1080)	Yes (n=851)	*P* value	No (n=1251)	Yes (n=665)	*P* value
Sex, n (%)				<.001[b]			.001[c]
Female	1106 (57.3)	572 (53.0)	534 (62.7)		680 (54.4)	415 (62.4)	
Age (years) mean (SD)	37.37 (1.90)	37.46 (1.87)	37.25 (1.93)	.01[d]	37.42 (1.89)	37.26 (1.91)	.06
Race, n (%)				.51			.16
White	1232 (68.5)	699 (69.2)	533 (67.6)		788 (67.5)	436 (70.8)	
Ethnicity, n (%)				.09			.19
Hispanic	179 (9.3)	89 (8.2)	90 (10.6)		108 (8.6)	70 (10.6)	
Marital status, n (%)				.01[d]			<.001[b]
Married	1140 (59.2)	611 (56.7)	529 (62.2)		702 (56.2)	431 (65.0)	
Education, n (%)				<.001[b]			<.001[b]
Some college or below	1132 (58.6)	697 (64.5)	435 (51.1)		807 (64.5)	317 (47.7)	
Living arrangement, n (%)				.21			.003[c]
Own place	1721 (89.3)	953 (88.5)	768 (90.4)		1095 (87.8)	614 (92.3)	
Household income (US $), n (%)				<.001[b]			<.001[b]
<50,000	426 (25.9)	289 (31.3)	137 (18.9)		346 (32.6)	72 (12.6)	
50,000-74,999	286 (17.4)	156 (16.9)	130 (18.0)		186 (17.5)	98 (17.1)	
75,000-99,999	295 (17.9)	161 (17.5)	134 (18.5)		179 (16.9)	115 (20.1)	
100,000-149,999	329 (20.0)	172 (18.7)	157 (21.7)		179 (16.9)	150 (26.2)	
150,000-199,999	164 (10.0)	80 (8.7)	84 (11.6)		93 (8.8)	69 (12.0)	
>200,000	146 (8.9)	64 (6.9)	82 (11.3)		77 (7.3)	69 (12.0)	
Health insurance, n (%)				<.001[b]			<.001[b]
Uninsured	158 (8.2)	117 (10.9)	41 (4.8)		134 (10.8)	22 (3.3)	
Ideology, n (%)				.01[d]			.58
Conservative	521 (28.0)	295 (28.8)	226 (27.0)		339 (28.3)	175 (26.9)	
Middle	837 (45.0)	480 (46.8)	357 (42.7)		541 (45.2)	290 (44.5)	
Liberal	504 (27.1)	250 (24.4)	254 (30.3)		316 (26.4)	186 (28.6)	
Pregnancy, n (%)				.83			.98
Yes	44 (3.0)	26 (3.2)	18 (2.8)		28 (3.0)	16 (3.2)	
Number of children, n (%)				.01[d]			.10
0	106 (7.2)	59 (7.1)	47 (7.4)		74 (7.8)	32 (6.3)	
1	342 (23.3)	172 (20.7)	170 (26.7)		210 (22.2)	131 (25.6)	
2	593 (40.5)	338 (40.7)	255 (40.1)		372 (39.4)	216 (42.3)	
3	285 (19.4)	164 (19.8)	121 (19.0)		184 (19.5)	98 (19.2)	

XSL•FO
RenderX

	Overall	Smartphone health app use			Wearable tracker use		
		No (n=1080)	Yes (n=851)	*P* value	No (n=1251)	Yes (n=665)	*P* value
4	98 (6.7)	65 (7.8)	33 (5.2)		74 (7.8)	23 (4.5)	
5	28 (1.9)	20 (2.4)	8 (1.3)		21 (2.2)	7 (1.4)	
6	14 (1.0)	12 (1.4)	2 (0.3)		10 (1.1)	4 (0.8)	

[a]Statistical comparisons between users and nonusers were performed using chi-square tests for categorical variables and 1-way ANOVA for continuous variables.

[b]*P*<.001.

[c]*P*<.01.

[d]*P*<.05.

Table . Different health aspects managed by smartphones (N=851).

	Values, n (%)
Physical activities or routines	714 (84.8)
Diet, food, or calories	399 (47.6)
Weight	359 (42.7)
Heart rate	347 (41.6)
Sleep	276 (33)
Blood pressure	97 (12)
Mood	56 (7)
Medication	54 (7)
Blood sugar or diabetes	24 (3)
Period or menstrual cycle[a]	182 (34.1)
Pregnancy[a]	57 (11)

[a]Responses were collected from female participants only. Percentage is calculated based on the number of female participants included in the sample (n=534).

Table . Mean and SD for continuous endogenous variables and correlation matrix (N=1931).

	Mean (SD)	1 (healthy eating)	2 (sleep hours)	3 (physical activity)	4 (BMI)
Healthy eating	66.9 (8.7)	1.00	—[a]	—	—
Sleep hours	6.7 (1.2)	0.12[b]	1.00	—	—
Physical activity	6.5 (5.5)	0.12[b]	0.04	1.00	—
BMI	29.9 (7.2)	−0.10[b]	−0.13[b]	−0.15[b]	1.00
Emotional distress	1.5 (0.5)	−0.15[b]	−0.14[b]	−0.11[b]	0.02

[a]Not applicable.

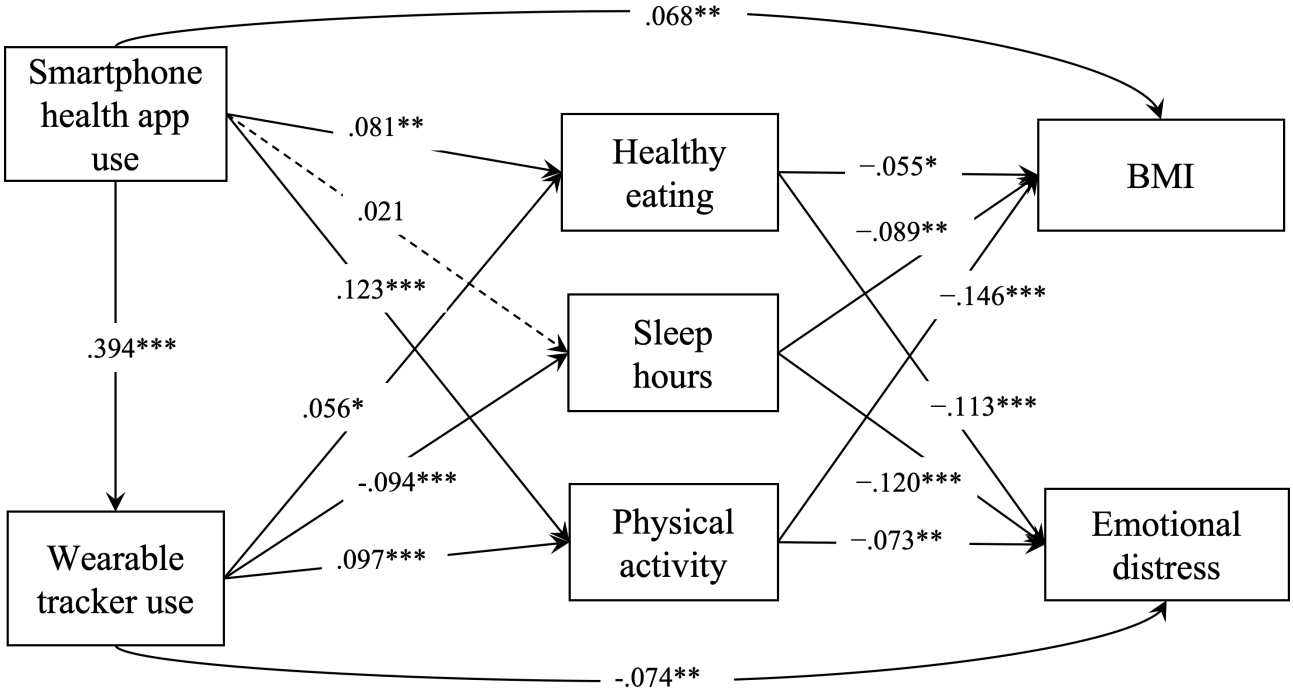
[b]*P*<.001.

Path Model (H1 and H2)

The final path model demonstrated a good fit with the data (χ^2_1=0.9, *P*=.02; root mean square error of approximation=0.048, 90% CI 0.02-0.09; comparative fit index=0.998; standardized root mean square residual=0.005). Results showed that the use of smartphone health apps positively predicted the use of wearable trackers (β=.394; *P*<.001). The use of smartphone health apps predicted higher BMI (β=.068; *P*=.006); thus, H1a was not supported. However, the use of wearable trackers did not significantly predict BMI (β=.047; *P*=.05); thus, H1b was also not supported. Results showed that the use of smartphone health apps did not predict emotional distress (β=−.016; *P*=.49), while the use of wearable trackers significantly predicted lower emotional distress (β=−.074; *P*=.003); thus, H2a was not supported, while H2b was supported. See Figure 2 for the full path model.

XSL•FO
RenderX

Figure 2. Path model: the relationships among mobile health (smartphone health app use, wearable tracker use), lifestyle mediators (healthy eating, sleep hours, and physical activity), BMI, and emotional distress. Note: solid lines indicate that the path coefficients were significantly different from 0, whereas path coefficients that were not significantly different from 0 were removed for presentation simplicity. Covariates controlled include age, sex, education, household income, race, and ethnicity. Unweighted n=1786. *P<.05, **P<.01, ***P<.001.



Mediation Analyses (H3 and H4)

Table 4 presents the indirect path coefficients from the mediation analyses to examine H3 and H4. To summarize the pattern, physical activity negatively mediated the associations between both types of mHealth use (health app and wearable tracker) and health outcomes (BMI and emotional distress). Nevertheless, sleep hours positively mediated only the relationship between wearable tracker use and health outcomes. Finally, healthy eating negatively mediated only the associations between both types of mHealth use and emotional distress.

XSL•FO
RenderX

Table . Indirect path coefficients for mediation analyses.

Indirect path	Estimate	SE	95% CI	β
Smartphone health app → eating → BMI	−0.064	0.034	−0.140 to 0.009	−.004
Smartphone health app → sleep → BMI	−0.027	0.035	−0.105 to 0.037	−.002
Smartphone health app → physical activity → BMI	−0.259	0.068	−0.401 to −0.136	−.018[a]
Wearable tracker → eating → BMI	−0.046	0.027	−0.108 to −0.003	−.003
Wearable tracker → sleep → BMI	0.127	0.051	0.042 to 0.241	.008[b]
Wearable tracker → physical activity → BMI	−0.213	0.068	−0.355 to −0.091	−.014[c]
Smartphone health app → eating → emotional distress	−0.009	0.004	−0.018 to −0.003	−.009[b]
Smartphone health app → sleep → emotional distress	−0.003	0.004	−0.010 to 0.003	−.003
Smartphone health app → physical activity → emotional distress	−0.009	0.004	−0.017 to −0.003	−.009[b]
Wearable tracker → eating → emotional distress	−0.007	0.003	−0.013 to −0.001	−.006[b]
Wearable tracker → sleep → emotional distress	0.012	0.004	0.005 to 0.021	.011[c]
Wearable tracker → physical activity → emotional distress	−0.008	0.004	−0.015 to −0.007	−.007[b]

[a]$P<.001$.

[b]$P<.05$.

[c]$P<.01$.

Discussion

Principal Results

This study, based on a nationally representative sample of adults in their late 30s to early 40s, found that mHealth users are more likely to be female, married, and of higher socioeconomic status. These users primarily engage with mobile technologies for tracking physical activity, followed by a range of other personal health aspects. Overall, our findings reveal a nuanced relationship between mHealth use and health outcomes: contrary to expectations, the use of smartphone health apps was associated with higher BMI, and the use of wearable trackers did not significantly predict BMI. However, the use of wearable trackers, not smartphone health apps, was linked to lower emotional distress. Mediation analyses further highlighted the complex mechanisms underlying these associations. Physical activity negatively mediated the relationships between both types of mHealth use and adverse health outcomes. This indicates that mHealth users were more likely to engage in physical activities, which predicted lower BMI and less emotional distress. Sleep hours mediated only the association between wearable tracker use and both health outcomes. This suggests that wearable tracker use was associated with fewer sleep hours, which predicted higher BMI and greater emotional distress. Finally, healthy eating mediated only the relationship between mHealth use and emotional distress. This suggests that mHealth users were more likely to adopt healthy dietary change, which was related to lower emotional distress.

Comparison With Prior Work

Our findings regarding mHealth use and BMI highlight the mediating role played by lifestyle factors in shaping physical health. Aligned with previous studies, on a positive note, individuals who used smartphones or wearables for health management reported engaging in increased physical activity [9,25], which was associated with lower BMI [17,19,20]. This finding supports the idea that mHealth technologies can potentially serve as effective tools for promoting lifestyle changes that lead to successful weight control. However, while these apps provide valuable information and motivation, their potential effectiveness depends on users taking actionable steps toward improving their health. It is possible that some individuals rely on health apps for tracking or goal setting but struggle with adherence to recommended behaviors, resulting in limited or even counterproductive health outcomes [46]. It is worth noting that the positive direct association between smartphone health app use and BMI can be interpreted in the other way around; people who are of higher BMI tend to use

XSL•FO
RenderX

health apps more because they see the need [47]. Future research should explore the psychological and behavioral mechanisms underlying this association.

The relationship between mHealth use and emotional distress also reveals important implications. Using wearable trackers is directly and indirectly associated with lower emotional distress, supporting their potential for mental health intervention [29,31,37]. While smartphone health app use did not show a direct connection to emotional distress, healthy eating and physical activity emerged as key mediators [37-39]. This finding underscores the importance of promoting both a healthy diet and physical activity for emotional well-being rather than relying solely on physical activity [21,33,39]. Future mHealth interventions could enhance their impact by incorporating personalized dietary and workout recommendations to better support emotional well-being.

Finally, an unexpected yet important finding is that wearable tracker use is negatively associated with sleep, which, in turn, is linked to both higher BMI and emotional distress. While our data support the well-established link between sleep and BMI [18], as well as sleep and emotional distress [34,35], they did not provide evidence that mHealth tools designed to improve sleep actually increase the absolute number of sleep hours. This may explain the limited evidence we found regarding the effectiveness of mHealth tools in improving sleep. Even in 1 study that reported wearable tracker use as positively associated with improved sleep, the focus was on perceived sleep quality rather than sleep duration [28]. Additionally, another study cited found that individuals using the wearable tracker had higher levels of well-being before than after its use [40].

There were several speculations on why sleep plays such an alarming mediating role in the relationship between wearable tracker use and health. While wearable devices provide valuable insights into sleep patterns, some users may develop a reliance on these tools, leading to obsessive tendencies where self-worth becomes tied to achieving predefined health metrics, potentially exacerbating emotional distress [37]. Misinterpretation of health data or false alarms may also contribute to heightened anxiety [37], and frequent notifications from these devices can also create a sense of urgency or fear as users may associate them with potential health concerns [48]. Notably, only the use of wearable trackers red-flagged this pattern, while smartphone health app use does not correlate with sleep. This raises the question of whether wearing a physical device itself might cause some sleep disturbances [26]. However, our results do not establish causality; it is also possible that individuals who already experience sleep deprivation and struggle with physical and mental health issues turn to wearable trackers for assistance. Future research is needed to explore these dynamics in greater depth and consider using a tracker to objectively measure sleep or consider sleep quality rather than mere quantity.

Practical Implications

The findings from this study offer several key implications for the design, development, and implementation of mHealth technologies. First, mHealth designers should prioritize integrating behavioral support features that guide users from self-monitoring to sustained lifestyle changes, given that simply using a smartphone health app without following corresponding lifestyle guidance may be associated with counterproductive outcomes. Health care providers and public health professionals should also avoid an overemphasis on outcome-oriented obsessive-tracking behaviors. Second, mHealth holds promise for mental health intervention, and mHealth solutions could incorporate evidence-based dietary and physical activity programs to further enhance emotional resilience and mental health benefits. Finally, developers should prioritize user-centered design approaches that minimize sleep disruption, such as optimizing comfort, reducing unnecessary notifications, and providing clearer guidance on interpreting sleep data. By addressing these challenges, mHealth technologies can be better optimized to support long-term health benefits while mitigating potential downsides, ultimately improving both physical and emotional well-being among users.

Limitations

There are several limitations of this study. First, measurement limitations should be acknowledged. For example, the healthy eating measure captures only a subset of dietary behaviors and does not reflect broader aspects of diet quality, the physical activity measure combines frequencies of heterogeneous activities without accounting for intensity, and the sleep variable measures only duration and does not capture quality, disturbances, or restorative aspects of sleep. And the reliance on self-reports may be subject to recall, response, and social desirability biases. Future research could use more comprehensive instruments, include validated indices of behavioral constructs, and incorporate objective health metrics (eg, biometrics from wearables) to enhance data reliability. Second, while this study identifies associations between mHealth use and health outcomes, causal relationships cannot be inferred. Longitudinal studies or experimental designs would provide stronger evidence of causality. Finally, given the growing diversity of mHealth apps, future research should explore how specific app features (eg, gamification and artificial intelligence–driven recommendations) influence user engagement and health behaviors.

Conclusions

This study underscores the potential of mHealth technologies to improve health outcomes through their influence on lifestyle behaviors. However, their impact is not uniform, and unintended negative consequences such as reduced sleep from wearable tracker use warrant further investigation. By addressing these challenges, mHealth technologies can better fulfill their promise of supporting individuals in achieving healthier lifestyles and advancing population health.

XSL•FO
RenderX

Acknowledgments

This manuscript would not have been possible without Kean University's support to provide research release time awarded to RZM and FS and the graduate research assistantship to JS. This research uses public data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill. The authors attest that there was no use of generative artificial intelligence technology in the generation of text, figures, or other informational content of this manuscript.

Funding

This study was funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. No direct support was received from grant P01-HD31921 for this study.

Data Availability

The data supporting the findings of this study are publicly available through the Inter-university Consortium for Political and Social Research (ICPSR). The dataset analyzed in this study (ICPSR 21600) is deposited in a publicly accessible repository and can be accessed.

Authors' Contributions

Conceptualization: RZM (lead), FS (equal)
Data curation: RZM (lead), FS (equal), JS (supporting)
Formal analysis: RZM
Funding acquisition: RZM (lead), FS (equal)
Methodology: RZM (lead), FS (equal)
Project administration: RZM (lead), JS (supporting)
Resources: FS
Software: RZM
Visualization: RZM (lead), JS (supporting)
Writing – original draft: RZM (lead), JS (supporting)
Writing – review & editing: RZM (lead), FS (equal), JS (supporting)

Conflicts of Interest

None declared.

References

1. Johnson C, Richwine C, Patel V. Individuals' access and use of patient portals and smartphone health apps, 2020. In: ASTP Health IT Data Brief [Internet]: Office of the Assistant Secretary for Technology Policy; 2021.
2. Vogels EA, Perrin A, Rainie L, Anderson M. 53% of Americans say the internet has been essential during the COVID-19 outbreak. Pew Research Center. 2020. URL: https://www.pewresearch.org/internet/2020/04/30/53-of-americans-say-the-internet-has-been-essential-during-the-covid-19-outbreak/ [accessed 2026-05-18]
3. Ernsting C, Dombrowski SU, Oedekoven M, et al. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017 Apr 5;19(4):e101. [doi: 10.2196/jmir.6838] [Medline: 28381394]
4. Silva BM, Rodrigues JJ, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272. [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
5. Pires IM, Marques G, Garcia NM, Flórez-Revuelta F, Ponciano V, Oniani S. A research on the classification and applicability of the mobile health applications. J Pers Med 2020 Feb 27;10(1):11. [doi: 10.3390/jpm10010011] [Medline: 32120849]
6. Dounavi K, Tsoumani O. Mobile health applications in weight management: a systematic literature review. Am J Prev Med 2019 Jun;56(6):894-903. [doi: 10.1016/j.amepre.2018.12.005] [Medline: 31003801]
7. Eisenstadt M, Liverpool S, Infanti E, Ciuvat RM, Carlsson C. Mobile apps that promote emotion regulation, positive mental health, and well-being in the general population: systematic review and meta-analysis. JMIR Ment Health 2021 Nov 8;8(11):e31170. [doi: 10.2196/31170] [Medline: 34747713]
8. Carroll K, Kennedy RA, Koutoulas V, Werake U, Bui M, Kraan CM. Comparability between wearable inertial sensors and an electronic walkway for spatiotemporal and relative phase data in young children aged 6-11 years. Gait Posture 2024 Jun;111:30-36. [doi: 10.1016/j.gaitpost.2024.04.003] [Medline: 38615566]
9. Tricás-Vidal HJ, Lucha-López MO, Hidalgo-García C, Vidal-Peracho MC, Monti-Ballano S, Tricás-Moreno JM. Health habits and wearable activity tracker devices: analytical cross-sectional study. Sensors (Basel) 2022 Apr 12;22(8):2960. [doi: 10.3390/s22082960] [Medline: 35458945]

XSL•FO
RenderX

10. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3. [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
11. Friel CP, Garber CE. An examination of the relationship between motivation, physical activity, and wearable activity monitor use. J Sport Exerc Psychol 2020;42(2):153-160. [doi: 10.1123/jsep.2019-0170] [Medline: 32150727]
12. Free C, Phillips G, Galli L, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362. [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
13. Gagnon MP, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
14. Romeo A, Edney S, Plotnikoff R, et al. Can smartphone apps increase physical activity? Systematic review and meta-analysis. J Med Internet Res 2019 Mar 19;21(3):e12053. [doi: 10.2196/12053] [Medline: 30888321]
15. Harvey J, Krukowski R, Priest J, West D. Log often, lose more: electronic dietary self-monitoring for weight loss. Obesity (Silver Spring) 2019 Mar;27(3):380-384. [doi: 10.1002/oby.22382] [Medline: 30801989]
16. Darling KE, Sato AF. Systematic review and meta-analysis examining the effectiveness of mobile health technologies in using self-monitoring for pediatric weight management. Child Obes 2017 Oct;13(5):347-355. [doi: 10.1089/chi.2017.0038] [Medline: 28471699]
17. O'Reilly GA, Huh J, Schembre SM, Tate EB, Pentz MA, Dunton G. Association of usual self-reported dietary intake with ecological momentary measures of affective and physical feeling states in children. Appetite 2015 Sep;92:314-321. [doi: 10.1016/j.appet.2015.05.032] [Medline: 26032196]
18. Cooper CB, Neufeld EV, Dolezal BA, Martin JL. Sleep deprivation and obesity in adults: a brief narrative review. BMJ Open Sport Exerc Med 2018;4(1):e000392. [doi: 10.1136/bmjsem-2018-000392] [Medline: 30364557]
19. Silveira EA, Mendonça CR, Delpino FM, et al. Sedentary behavior, physical inactivity, abdominal obesity and obesity in adults and older adults: a systematic review and meta-analysis. Clin Nutr ESPEN 2022 Aug;50:63-73. [doi: 10.1016/j.clnesp.2022.06.001] [Medline: 35871953]
20. Rippe JM, Hess S. The role of physical activity in the prevention and management of obesity. J Am Diet Assoc 1998 Oct;98(10 Suppl 2):S31-S38. [doi: 10.1016/s0002-8223(98)00708-1] [Medline: 9787734]
21. McCarroll R, Eyles H, Ni Mhurchu C. Effectiveness of mobile health (mHealth) interventions for promoting healthy eating in adults: a systematic review. Prev Med 2017 Dec;105:156-168. [doi: 10.1016/j.ypmed.2017.08.022] [Medline: 28882743]
22. Modrzejewska J, Modrzejewska A, Czepczor-Bernat K, Matusik P. The role of body mass index, healthy eating-related apps and educational activities on eating motives and behaviours among women during the COVID-19 pandemic: a cross sectional study. PLoS One 2022;17(3):e0266016. [doi: 10.1371/journal.pone.0266016] [Medline: 35344563]
23. Bell BM, Alam R, Alshurafa N, et al. Automatic, wearable-based, in-field eating detection approaches for public health research: a scoping review. NPJ Digit Med 2020;3:38. [doi: 10.1038/s41746-020-0246-2] [Medline: 32195373]
24. Bedri A, Li R, Haynes M, et al. EarBit: using wearable sensors to detect eating episodes in unconstrained environments. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Sep;1(3):37. [doi: 10.1145/3130902] [Medline: 30135957]
25. Ferguson T, Olds T, Curtis R, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022 Aug;4(8):e615-e626. [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
26. Griffiths C, Walker K, Willis A, Pollard L. Treatment resistant depression (TRD) service outpatient's experience of sleep, activity, and using a Fitbit wearable activity and sleep tracker. Ment Health Rev J 2022 May 2;27(2):158-174. [doi: 10.1108/MHRJ-04-2021-0036]
27. Robbins R, Seixas A, Masters LW, et al. Sleep tracking: a systematic review of the research using commercially available technology. Curr Sleep Med Rep 2019;5(3):156-163. [doi: 10.1007/s40675-019-00150-1] [Medline: 33134038]
28. Berryhill S, Morton CJ, Dean A, et al. Effect of wearables on sleep in healthy individuals: a randomized crossover trial and validation study. J Clin Sleep Med 2020 May 15;16(5):775-783. [doi: 10.5664/jcsm.8356] [Medline: 32043961]
29. Adil M, Atiq I, Younus S. Effectiveness of the Apple Watch as a mental health tracker. J Glob Health 2024 Feb 9;14:03010. [doi: 10.7189/jogh.14.03010] [Medline: 38332682]
30. Bakker D, Rickard N. Engagement in mobile phone app for self-monitoring of emotional wellbeing predicts changes in mental health: MoodPrism. J Affect Disord 2018 Feb;227:432-442. [doi: 10.1016/j.jad.2017.11.016] [Medline: 29154165]
31. Funnell EL, Spadaro B, Martin-Key N, Metcalfe T, Bahn S. mHealth solutions for mental health screening and diagnosis: a review of app user perspectives using sentiment and thematic analysis. Front Psychiatry 2022;13:857304. [doi: 10.3389/fpsyt.2022.857304] [Medline: 35573342]
32. Bush NE, Smolenski DJ, Denneson LM, Williams HB, Thomas EK, Dobscha SK. A virtual hope box: randomized controlled trial of a smartphone app for emotional regulation and coping with distress. Psychiatr Serv 2017 Apr 1;68(4):330-336. [doi: 10.1176/appi.ps.201600283] [Medline: 27842473]
33. Veenhoven R. Will healthy eating make you happier? A research synthesis using an online findings archive. Appl Res Qual Life 2021 Feb;16:221-240. [doi: 10.1007/s11482-019-09748-7]
34. Benca RM, Okawa M, Uchiyama M, et al. Sleep and mood disorders. Sleep Med Rev 1997 Nov;1(1):45-56. [doi: 10.1016/s1087-0792(97)90005-8] [Medline: 15310523]

XSL•FO
RenderX

35. Neckelmann D, Mykletun A, Dahl AA. Chronic insomnia as a risk factor for developing anxiety and depression. Sleep 2007 Jul;30(7):873-880. [doi: 10.1093/sleep/30.7.873] [Medline: 17682658]
36. Camiletti-Moirón D, Aparicio VA, Aranda P, Radak Z. Does exercise reduce brain oxidative stress? A systematic review. Scand J Med Sci Sports 2013 Aug;23(4):e202-e212. [doi: 10.1111/sms.12065] [Medline: 23495801]
37. Choudhury A, Asan O. Impact of using wearable devices on psychological distress: analysis of the Health Information National Trends Survey. Int J Med Inform 2021 Dec;156:104612. [doi: 10.1016/j.ijmedinf.2021.104612] [Medline: 34649113]
38. Lyzwinski LN, Caffery L, Bambling M, Edirippulige S. The mindfulness app trial for weight, weight-related behaviors, and stress in university students: randomized controlled trial. JMIR Mhealth Uhealth 2019 Apr 10;7(4):e12210. [doi: 10.2196/12210] [Medline: 30969174]
39. Morshed MB, Kulkarni SS, Saha K, et al. Food, mood, context: examining college students' eating context and mental well-being. ACM Trans Comput Healthcare 2022 Oct;3(4):1-26. [doi: 10.1145/3533390]
40. Attie E, Meyer-Waarden L. How do you sleep? The impact of sleep apps on Generation Z's well-being. J Interact Mark 2023;58(2-3):222-247. [doi: 10.1177/10949968221142806]
41. Hummer RA, Aiello AE, Kathleen Mullan H, Richard UJ. National Longitudinal Study of Adolescent to Adult Health (Add Health), 1994-2025 [Public Use] (ICPSR 21600). ICPSR. 2025. URL: https://www.icpsr.umich.edu/web/ICPSR/studies/21600/datadocumentation [accessed 2026-05-29]
42. Keys A, Fidanza F, Karvonen MJ, Kimura N, Taylor HL. Indices of relative weight and obesity. J Chronic Dis 1972 Jul 1;25(6):329-343. [doi: 10.1016/0021-9681(72)90027-6] [Medline: 4650929]
43. Yoshida K, Bartel A, Chipman JJ, et al. Package 'tableone'. The Comprehensive R Archive Network. 2026. URL: https://cran.r-project.org/web/packages/tableone/tableone.pdf [accessed 2026-05-18]
44. Rosseel Y, Oberski D, Byrnes J, Vanbrabant L, Savalei V, Merkle E, et al. Package 'lavaan'. The Comprehensive R Archive Network. URL: https://cran.r-project.org/web/packages/lavaan/lavaan.pdf [accessed 2026-05-18]
45. Korn EL, Graubard BI. Examples of differing weighted and unweighted estimates from a sample survey. Am Stat 1995 Aug;49(3):291-295. [doi: 10.1080/00031305.1995.10476167]
46. Giebel GD, Speckemeier C, Abels C, et al. Problems and barriers related to the use of digital health applications: scoping review. J Med Internet Res 2023 May 12;25:e43808. [doi: 10.2196/43808] [Medline: 37171838]
47. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 4;3(4):e101. [doi: 10.2196/mhealth.4924] [Medline: 26537656]
48. Rosman L, Gehi A, Lampert R. When smartwatches contribute to health anxiety in patients with atrial fibrillation. Cardiovasc Digit Health J 2020;1(1):9-10. [doi: 10.1016/j.cvdhj.2020.06.004] [Medline: 34386784]

Abbreviations

mHealth: mobile health

Edited by L Buis; submitted 04.Apr.2025; peer-reviewed by J Igboanugo, S Mohanadas, S Sahai; revised version received 18.Nov.2025; accepted 18.Feb.2026; published 03.Jun.2026.

Please cite as:
Mi RZ, Shen F, Strugala J
Managing BMI and Emotional Distress Using mHealth: Nationally Representative Survey Study
JMIR Mhealth Uhealth 2026;14:e75499
URL: https://mhealth.jmir.org/2026/1/e75499
doi:10.2196/75499

© Ranran Z Mi, Fei Shen, Julia Strugala. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 3.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Impact of Telehealth Use Experience on Recruitment of Underserved Populations Into a Telehealth-Delivered Mindfulness-Based Trial for Patients With Chronic Low Back Pain: Cross-Sectional Survey

Marva V Foster[1,2], PhD; Phuong Tra Nguyen[3], MSc, MPH; Janice Weinberg[4], SCD; Keturah R Faurot[5], MPH, PhD; Natalia Morone[1,3], MD, MS

[1]Department of Medicine, Section of General Internal Medicine, Boston University Chobanian and Avedisian School of Medicine, 801 Massachusetts Ave., Boston, MA, United States

[2]Center for Health Optimization and Implementation Research (CHOIR), VA Bedford and VA Boston Healthcare Systems, Boston, MA, United States

[3]Boston Medical Center, Boston, MA, United States

[4]Department of Biostatistics, Boston University School of Public Health, Boston, MA, United States

[5]Program on Integrative Medicine, Department of Physical Medicine and Rehabilitation, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Corresponding Author:
Marva V Foster, PhD
Department of Medicine, Section of General Internal Medicine, Boston University Chobanian and Avedisian School of Medicine, 801 Massachusetts Ave., Boston, MA, United States

Abstract

Background: Mindfulness-based stress reduction (MBSR) is recommended as a noninvasive, evidence-based treatment for chronic low back pain (CLBP); however, access and uptake are often constrained among medically underserved populations. Telehealth-delivered MBSR may scale access, yet little is known about the characteristics of underserved patients who have experience using telehealth and are willing to enroll in pragmatic trials evaluating MBSR.

Objective: This study examined individual- and county-level social vulnerability characteristics associated with telehealth use experience among people living with CLBP who enrolled in OPTIMUM (Optimizing Pain Treatment in Medical Settings Using Mindfulness), a pragmatic randomized clinical trial of telehealth-delivered MBSR.

Methods: We conducted a cross-sectional analysis of baseline measures from the OPTIMUM trial, focusing on participants from a safety-net health institution and a federally qualified health center. The primary outcome was telehealth use experience (having had a prior telehealth visit at or before enrollment). Guided by Andersen's Behavioral Model of Health Services Use, we evaluated predisposing, enabling, and need factors. Prespecified measures included the following: demographics; employment and education; internet access and comfort using Zoom; neighborhood indices (Area Deprivation Index [ADI], Social Vulnerability Index [SVI]); rurality; back-pain duration; and validated patient-reported outcomes. Multivariable logistic regression was used to estimate adjusted odds ratios (AORs) and 95% CIs.

Results: Of 244 enrolled participants (mean age 50.6, SD 13.5 y; n=162, 66.7% female; n=124, 50.8% non-Hispanic Black), 174 (71.3%) reported previous telehealth use. Predisposing factors associated with lower telehealth use included race or ethnicity (non-Hispanic Black: AOR 0.38, 95% CI 0.2 - 0.8; Hispanic: AOR 0.21, 95% CI 0.06 - 0.7) and part-time employment (AOR 0.33, 95% CI 0.1 - 0.9) vs full-time employment, whereas educational attainment above or below a high school degree was associated with higher odds of telehealth use (AORs 2.7 - 3.2). Enabling factors showed that lower comfort with Zoom was associated with lower telehealth use (AOR 0.5, 95% CI 0.2 - 0.9), whereas home internet access was common and not independently significant in adjusted models. Neighborhood socioeconomic disadvantages, social vulnerability, and rurality did not demonstrate significant associations. Needs factors indicated that longer CLBP duration was positively associated with telehealth experience (>5 y: AOR 3.5, 95% CI 1.4 - 8.8). Other patient-reported outcomes, comorbidity burden, and surgical history were not significantly associated.

Conclusions: In a safety-net and federally qualified health center context, prior telehealth experience was common among patients with CLBP willing to enroll in a telehealth-delivered MBSR trial. Although certain predisposing factors (race or ethnicity, employment status) were associated with lower odds of telehealth experience, these factors did not prevent trial enrollment. Targeted outreach to improve digital comfort and inclusive engagement strategies may further reduce disparities and support broader implementation of telehealth for CLBP.

XSL•FO
RenderX

Trial Registration: ClinicalTrials.gov NCT04129450; https://clinicaltrials.gov/study/NCT04129450

(JMIR Mhealth Uhealth 2026;14:e82093) doi:10.2196/82093

KEYWORDS

telehealth; mindfulness-based stress reduction; low back pain; safety net; pragmatic trial; digital literacy; social determinants of health

Introduction

Chronic low back pain (CLBP) is a costly problem that results in US $100 to 200 billion per year in expenditures [1,2]. Many medical approaches used to treat CLBP are associated with adverse events and negative side effects [3,4]. Current treatment guidelines therefore recommend evidence-based, noninvasive approaches as first-line treatments for CLBP [5,6]. Mindfulness-based stress reduction (MBSR) is a recommended evidence-based treatment that focuses on increasing participants' awareness and acceptance of moment-to-moment experiences, including physical discomfort and difficult emotions [7]. Prior research has found that medically underserved populations have an increased prevalence of chronic pain, less access to pain care, and poorer outcomes [8-11]. Telehealth-delivered MBSR offers a low-risk, effective, and viable alternative to conventional in-person delivery for patients with CLBP [12].

Telehealth refers to the use of electronic information and telecommunications (eg, videoconferencing) to provide clinical services [13]. Although prior to the COVID-19 pandemic, telehealth equipment was available, there was low usage due to a myriad of factors related to billing practices, patient and physician comfort, and familiarity with using the technology [14,15]. Use of telehealth leading up to the COVID-19 pandemic was more likely to be used by affluent, privately insured populations [16]. The COVID-19 pandemic created a unique and sudden need for virtual health visits for both urgent and nonurgent health visits, leading to an unprecedented expansion of telehealth access among underserved populations [17-19]. Expanded telehealth accessibility enhanced the potential to scale MBSR treatment to underserved populations.

In 2021, we launched OPTIMUM (Optimizing Pain Treatment In Medical Settings Using Mindfulness), a pragmatic randomized clinical trial comparing MBSR with usual care within primary care clinics for patients with CLBP at a safety-net hospital, federally qualified health centers, and a large academic health system. The COVID-19 pandemic created an urgent need for non–face-to-face care, prompting the conversion of MBSR from in-person to telehealth-delivered MBSR. While telehealth may improve access to MBSR, little data existed regarding how attractive this treatment would be to patients at safety net hospitals and federally qualified health centers. Safety-net hospitals and federally qualified health centers provide essential inpatient and outpatient services to low-income, uninsured people residing in rural areas and people with limited access to health care, that is, medically underserved populations [20].

Given the disparities and emerging telehealth opportunities, this study aimed to explore the impact of prior telehealth use on enrollment in the OPTIMUM trial. Therefore, the purpose of the study was to examine the individual- and county-level social vulnerability characteristics of people living with CLBP who were willing to enroll in OPTIMUM, a telehealth-delivered MBSR randomized pragmatic trial.

Methods

Theoretical Framework

Andersen's Behavioral Model of Health Services Use has been widely used to analyze factors associated with health service use. The results suggest that health service usage is primarily motivated by individual illness, but the quality and quantity of health service usage vary significantly based on socioeconomic factors, such as income or health-insurance status [21-23]. The model posits 3 dimensions: predisposing factors (eg, age, education), enabling factors (eg, income, hospital density), and need factors (eg, health status) that influence health care usage [24-26]. Examining these dimensions may also be a suitable model when exploratory research is needed due to lack of previous studies on telehealth use among underserved patients with CLBP, as in this study.

Study Design and Setting

This cross-sectional analysis was conducted on the baseline study measures data from the OPTIMUM trial, a multisite pragmatic randomized trial of telehealth-delivered MBSR for patients with CLBP. Data collection occurred between April 2021 and August 2023, with 12-month outcome data collection completed in November 2024. All variables were collected upon enrollment. OPTIMUM was embedded into primary care clinics at 3 health systems (Boston Medical Center, Massachusetts; University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania; and Piedmont Health Services, in partnership with the University of North Carolina [UNC] Chapel Hill). The 3 health systems serve predominantly underserved patients: Boston Medical Center is a safety-net institution serving an urban, low-income, racially and ethnically diverse population. UNC Chapel Hill has 2 sites: an academic center–based family medicine clinic and a federally qualified health center network serving many rural patients and those underrepresented in research. The University of Pittsburgh serves an urban or suburban population. Because we were interested in the factors involved in telehealth use experience for underserved populations, we restricted our analysis to the baseline measurements of participants from the safety-net health setting (Boston Medical Center) and the federally qualified health centers (UNC at Chapel Hill, Piedmont Health Services).

XSL•FO
RenderX

Recruitment

Research staff at each participating clinic approached potentially eligible patients during clinic visits, through telephone outreach, or via clinician referrals. Staff verified eligibility, provided study information, and obtained informed consent. Participants were screened to ensure they met inclusion criteria: age ≥18 years, diagnosis of CLBP, willing and able to provide telephone informed consent, able to speak and read English, absence of markers suggesting a serious condition, and a Pain, Enjoyment of Life and General Activity (PEG) scale score ≥3. Participants were excluded if they had an unexplained fever or weight loss, pregnancy, metastatic cancer, or cohabitation with another study participant.

Participants of the OPTIMUM Study

Patients who met eligibility criteria were randomly assigned to medical group visits plus 90-minute, instructor-led MBSR treatment for 8 consecutive weeks via telehealth or to usual care alone. At the beginning of the MBSR session, for approximately 30 minutes, the primary care provider met briefly and privately with each participant through a telehealth visit (in a breakout room in the HIPAA [Health Insurance Portability and Accountability Act] compliant Zoom platform) to review medical aspects of treatment, focused specifically on the patient's back pain. Participants were administered several measures at baseline, 8-week, 6-month, and the 12-month follow-up [27]. Further information about OPTIMUM is published elsewhere [27].

Because this analysis focused on understanding telehealth use experience among underserved populations, we restricted the analytic sample to participants enrolled from Boston Medical Center (safety-net setting) and UNC or Piedmont Health Services (federally qualified health centers).

Outcome

Our outcome of interest in this paper was telehealth use experience as measured at enrollment, that is, telehealth use experience. Telehealth use experience includes the use of telehealth prior to or during the COVID-19 pandemic.

Predictor Variables

The online baseline data included measures of demographics, back-related pain, treatments, the impact of pain on quality of life, pain beliefs, medical history, health habits, and traumatic experiences. Dimensions of Andersen's Behavioral Model of Health Services Use and variables are as follows:

1. Predisposing factors: Age in 2021, sex, race or ethnicity, educational level, marital status, employment status, and health beliefs or psychological predispositions (Pain Catastrophizing Questionnaire [28], CAMS-R [Cognitive and Affective Mindfulness Scale-Revised] [29])
2. Enabling factors: Annual income, access to internet at home, comfort levels with Zoom or virtual meetings (confidence in their ability to use the platform), Area Deprivation Index (ADI), Social Vulnerability Index (SVI; both SVI and ADI are well-established indices that have been used to assess socioeconomic disadvantage and vulnerability [30,31]), and rural-urban commuting area
3. Needs factors: Number of years with low back pain; receiving disability or workers' compensation due to low back pain or unemployed for 1 month or more due to low back pain [32]; Charlson Comorbidity Index [33]; number of low back surgeries; PROMIS-29 (Patient-Reported Outcomes Measurement Information System-29) domains, including sleep disturbance, fatigue, physical function, pain interference, ability to participate in social roles or activities, cognitive function, anxiety, and depression [34]; pain impact score; PEG [35]; TAPS (tobacco, alcohol, prescription medication, and other substance use) [36]

Statistical Analyses

This study is reported in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines (Checklist 1) [37] for cross-sectional studies. For descriptive analysis, we assessed the characteristics based on telehealth use experience (had previous telehealth visit or never had telehealth visit). Chi-square or Fisher exact tests were used for categorical variables, and Wilcoxon rank-sum tests were used for continuous variables.

Missing data were minimal (<5% for most variables). The low level of missingness did not warrant multiple imputation, which would have been unlikely to materially change the results. For continuous variables, histograms and Q-Q plots were used to assess normality. Nonnormally distributed variables were analyzed using nonparametric tests (Wilcoxon rank-sum). Categorical variables were analyzed using chi-square or Fisher exact tests depending on cell counts.

Bivariate analyses served as a screening step for multivariable modeling. A liberal threshold (α=.10) was used to avoid excluding potential predictors. Multivariable logistic regression retained variables with P<.05 in unadjusted models. Formal multiple-comparison correction was not applied because the primary goal was to build an adjusted model to identify factors associated with telehealth use experience, rather than to conduct multiple independent hypothesis tests.

The analytic sample was determined by available baseline data from the safety-net and federally qualified health center sites. These sites were selected based on the study objective of understanding telehealth use among underserved populations; thus, sample size was dictated by enrolled participants meeting eligibility criteria within these settings. We report both unadjusted odds ratios (ORs) and adjusted odds ratios (AORs) and 95% CIs as applicable. All analyses were conducted in SAS (version 9.4) with P<.05 considered statistically significant unless otherwise specified.

Ethical Considerations

All study procedures were approved by the single Institutional Review Board at the University of Pittsburgh (STUDY20110378). All participants provided informed consent before the study and agreed to the use of their data for research purposes, including secondary analyses. Participation in the OPTIMUM study was voluntary, and participants could decline without consequences. To protect participants' privacy and confidentiality, only anonymized data were used for this analysis, and results are reported in aggregate so that no

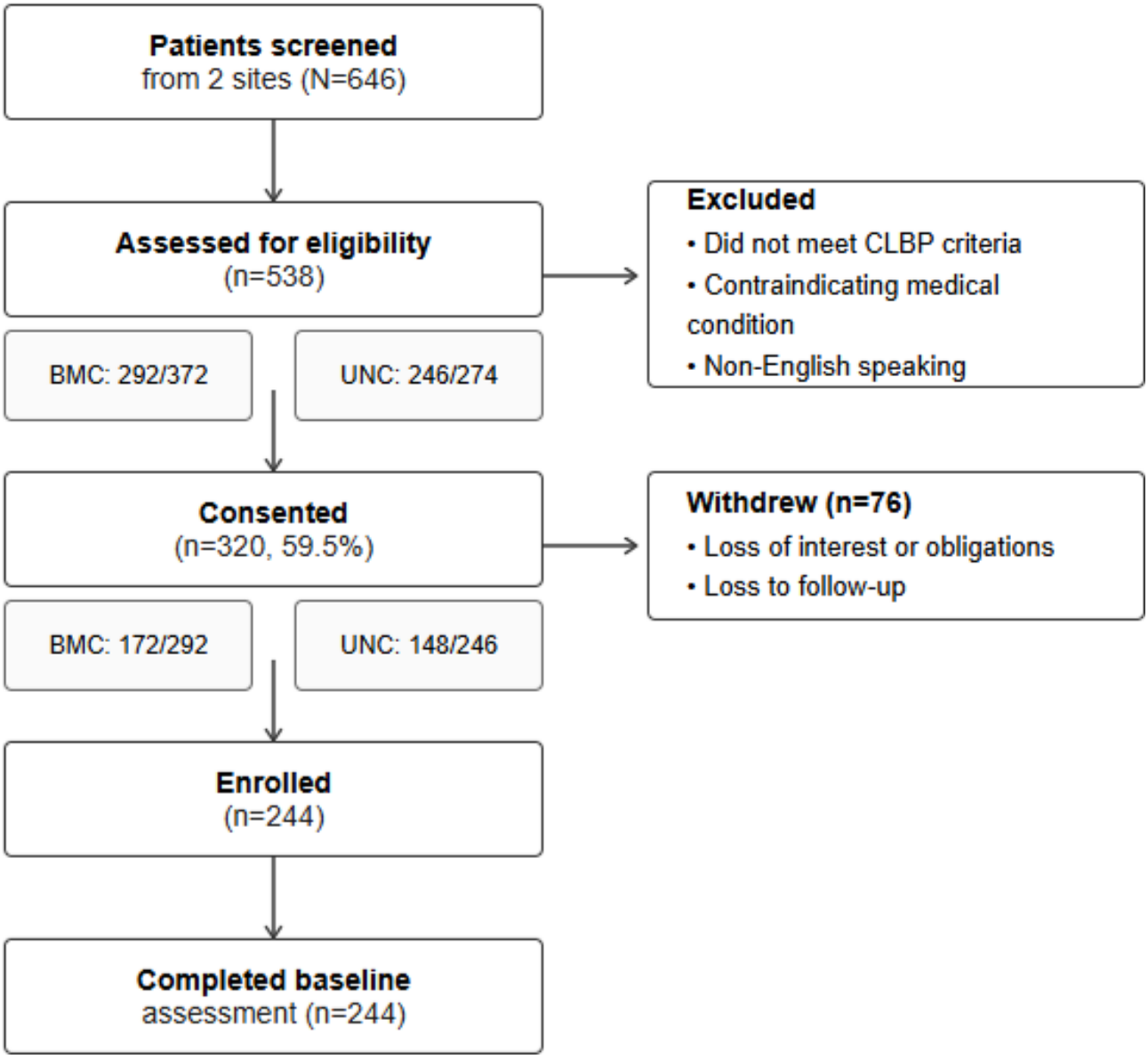
XSL•FO
RenderX

individual participant can be identified. Participants were compensated US $30 for completing study measures administered at baseline and follow-up.

Results

Out of 646 patients with CLBP from the safety-net health setting (Boston Medical Center) and the federally qualified health centers (UNC at Chapel Hill, Piedmont Health Services), 538 were assessed for eligibility, of whom 320 (59.5%) were eligible and consented. Of the 320 patients who consented, 244 completed baseline measures. Among those who initially enrolled but later withdrew, reasons included loss of interest, work or family obligations, or loss to follow-up. Patients who declined or were excluded were typically ineligible due to not meeting CLBP criteria, having contraindicated medical conditions, or being unable to complete consent in English (Figure 1). Analytic sample sizes were as follows: race or ethnicity (n=238), employment status (n=243), internet access (n=240), comfort with Zoom (n=241), and sex at birth (n=243). We performed a complete case analysis for multivariable modeling, yielding a final sample of 233.

Figure 1. Participant flow diagram. CLBP: chronic low back pain; BMC: Boston Medical Center; UNC: University of North Carolina.



Several factors were associated with having telehealth use experience. *Predisposing factors* such as race or ethnicity and employment status showed significant differences. Non-Hispanic White participants had the highest telehealth uptake at 84.1% (58/69), while Hispanic participants had the lowest at 56.5% (13/23; Table 1). In both the unadjusted and adjusted models (Tables 2 and 3), non-Hispanic Black participants had lower odds of telehealth use (OR 0.38, 95% CI 0.2 - 0.8). Hispanic participants also had lower odds in both models, with an unadjusted OR of 0.25 (95% CI 0.09 - 0.7) and an AOR of 0.21 (95% CI 0.06 - 0.7). Employment status was significantly associated with telehealth use experience (P=.03). Those who were employed part-time were less likely to have used telehealth (AOR 0.33, 95% CI 0.1 - 0.9) in the adjusted model compared to those who worked full-time. Education demonstrated a positive trend, ranging from 62.3% (71/114) among those with a high school degree to 85.7% (12/14) among those with doctoral education, but the overall P=.07 was not significant.

XSL•FO
RenderX

The pain catastrophizing questionnaire and CAMS-R were not significant (P=.40 and P=.87, respectively). Age, sex, and marital status were not significant predictors in the adjusted model (Table 3).

XSL•FO
RenderX

Table . Characteristics of OPTIMUM (Optimizing Pain Treatment In Medical Settings Using Mindfulness) study participants by telehealth group experience[a].

Characteristics[b,c]	All (N=244)	Never had telehealth visit (n=70)	Had telehealth visit (n=174)	*P* value
Age (y), mean (SD)	50.6 (13.5)	50.8 (13.8)	50.5 (13.4)	.74[d]
Sex, n (%)				.09[e]
Male	81 (33.2)	29 (35.8)	52 (64.2)	
Female	162 (66.4)	41 (25.3)	121 (74.7)	
Race or ethnicity, n (%)				.04[f,g]
Non-Hispanic White	69 (28.3)	11 (15.9)	58 (84.1)	
Non-Hispanic Black	124 (50.8)	41 (33.1)	83 (66.9)	
Hispanic	23 (9.4)	10 (43.5)	13 (56.5)	
Non-Hispanic Other	13 (5.3)	5 (38.5)	8 (61.5)	
Unknown ethnicity, Black/African American	9 (3.7)	2 (22.2)	7 (77.8)	
Unknown or not reported	6 (2.5)	1 (16.7)	5 (83.3)	
Education, n (%)				.07[g]
Less than high school	6 (2.5)	1 (16.7)	5 (83.3)	
Some high school education	25 (10.2)	8 (32)	17 (68)	
High school diploma or equivalent	114 (46.7)	43 (37.7)	71 (62.3)	
Associate's or technical degree	41 (16.8)	8 (19.5)	33 (80.5)	
College or baccalaureate degree	44 (18.0)	8 (18.2)	36 (81.8)	
Doctoral or postgraduate education	14 (5.7)	2 (14.3)	12 (85.7)	
Employment status, n (%)				.03[g]
Full-time	69 (28.3)	18 (26.1)	51 (73.9)	
Not employed	138 (56.6)	35 (25.4)	103 (74.6)	
Part-time	36 (14.8)	17 (47.2)	19 (52.8)	
Marital status, n (%)				.39[e]
Never married	98 (40.2)	35 (35.7)	63 (64.3)	
Married	44 (18.0)	12 (27.3)	32 (72.7)	
Divorced	46 (18.9)	11 (23.9)	35 (76.1)	
Separated	24 (9.8)	4 (16.7)	20 (83.3)	
Domestic partner	19 (7.8)	4 (21.1)	15 (78.9)	
Widowed	13 (5.3)	4 (30.8)	9 (69.2)	
Annual income (US $), n (%)				.40[g]
Less than 10,000	40 (16.4)	13 (32.5)	27 (67.5)	
10,000-24,999	56 (22.9)	14 (25)	42 (75)	
25,000-34,999	23 (9.4)	9 (39.1)	14 (60.9)	
35,000-49,999	29 (11.8)	6 (20.7)	23 (79.3)	
50,000-74,999	19 (7.8)	6 (31.6)	13 (68.4)	

XSL•FO
RenderX

Characteristics[b,c]	All (N=244)	Never had telehealth visit (n=70)	Had telehealth visit (n=174)	*P* value
75,000-99,999	6 (2.5)	3 (50)	3 (50)	
100,000-149,999	9 (3.7)	0 (0.0)	9 (100)	
150,000-199,999	2 (0.8)	0 (0.0)	2 (100)	
200,000 or more	1 (0.4)	0 (0.0)	1 (100)	
Prefer not to answer	56 (22.9)	17 (30.4)	39 (69.6)	
Years of low back pain, n (%)				<.001[f]
Less than 5 years	68 (27.9)	33 (48.5)	35 (51.5)	
5 - 15 years	115 (47.1)	27 (23.5)	88 (76.5)	
More than 15 years	61 (25)	10 (16.4)	51 (83.6)	
Off work ≥1 month due to low back pain, n (%)				.68
Agree	112 (45.9)	30 (26.8)	82 (73.2)	
Disagree	78 (31.9)	22 (28.2)	56 (71.8)	
Does not apply	54 (22.1)	18 (33.3)	36 (66.7)	
Receives or has applied for disability or workers' compensation due to low back pain, n (%)				.55
Agree	80 (32.8)	20 (25)	60 (75)	
Disagree	117 (47.9)	34 (29.1)	83 (70.9)	
Does not apply	47 (19.3)	16 (34.0)	31 (65.9)	
Home internet access with Zoom capability, n (%)				.04[g]
Yes	228 (93.4)	62 (27.2)	166 (72.8)	
No	12 (4.9)	7 (58.3)	5 (41.7)	
Comfort level with Zoom or virtual meetings, n (%)				.002[f]
Not at all	23 (9.4)	14 (60.9)	9 (39.1)	
A little bit	22 (9.0)	7 (31.8)	15 (68.2)	
Somewhat	24 (9.8)	8 (33.3)	16 (66.7)	
Quite a bit	67 (27.5)	11 (16.4)	56 (85.6)	
Very much	108 (44.3)	30 (27.8)	78 (72.2)	
Prior low back surgery, n (%)				.57[g]
Yes, 1 operation	26 (10.7)	7 (26.9)	19 (73.1)	
Yes, more than 1 operation	8 (3.3)	1 (12.5)	7 (87.5)	
No	210 (86.0)	62 (29.5)	148 (70.5)	
TAPS[h] part 1: possible substance use, n (%)				.68[e]
Yes	166 (68.0)	49 (29.5)	117 (70.5)	
No	78 (31.9)	21 (26.9)	57 (73.1)	
PEG[i] score, mean (SD)	6.7 (2.1)	7.1 (1.9)	6.5 (2.2)	.08[d]
Pain impact score, mean (SD)	32.2 (3.8)	32.8 (3.6)	32.0 (3.9)	.14[d]

XSL•FO
RenderX

Characteristics[b,c]	All (N=244)	Never had telehealth visit (n=70)	Had telehealth visit (n=174)	*P* value
PROMIS-29[j] depression, mean (SD)	55.7 (9.4)	55.1 (9.7)	56.0 (9.3)	.57[d]
PROMIS-29 anxiety, mean (SD)	59.0 (9.1)	59.3 (10.3)	58.9 (8.6)	.80[d]
PROMIS-29 cognitive function, mean (SD)	47.1 (7.1)	46.7 (7.3)	47.1 (7.0)	.61[d]
PROMIS-29 ability to participate in social roles or activities, mean (SD)	44.2 (8.7)	44.8 (8.6)	43.9 (8.8)	.60[d]
PROMIS-29 pain interference, mean (SD)	63.7 (6.8)	63.4 (7.4)	63.8 (6.5)	.66[k]
PROMIS-29 physical function, mean (SD)	36.8 (5.6)	37.6 (5.4)	36.5 (5.7)	.27[d]
PROMIS-29 fatigue, mean (SD)	56.6 (9.7)	54.9 (9.5)	57.2 (9.7)	.10[k]
PROMIS-29 sleep disturbance, mean (SD)	58.0 (8.1)	58.5 (8.4)	57.8 (8.0)	.56[k]
Pain Catastrophizing Questionnaire, mean (SD)	13.8 (5.8)	14.3 (5.5)	13.7 (6.0)	.40[d]
CAMS-R[l] mindfulness, mean (SD)	32.7 (6.8)	32.9 (7.5)	32.7 (6.6)	.87[d]
Charlson Comorbidity Index (weighted sum), mean (SD)	2.4 (3.3)	2.3 (2.7)	2.5 (3.5)	.64[d]
ADI[m] national percentile (block group), quintile, n (%)				.27[e]
Quintile 1 (1 - 20, least deprived)	90 (36.9)	31 (34.4)	59 (65.6)	
Quintile 2 (21-40)	69 (29.3)	21 (30.4)	48 (69.6)	
Quintile 3 (41-60)	22 (9.0)	3 (13.6)	19 (86.4)	
Quintile 4 (61-80)	25 (10.2)	8 (32)	17 (68)	
Quintile 5 (81 - 100, most deprived)	30 (12.3)	6 (20)	24 (80)	
RUCA[n] classification, n (%)				.88[e]
Metropolitan	222 (90.9)	64 (28.8)	158 (71.2)	

XSL•FO
RenderX

Characteristics[b,c]	All (N=244)	Never had telehealth visit (n=70)	Had telehealth visit (n=174)	*P* value
Nonmetropolitan or rural	22 (9.0)	6 (27.3)	16 (72.7)	

[a]Unless otherwise noted, percentages reflect row totals (ie, the percentage of participants with each characteristic who did vs did not have a telehealth visit).

[b]Percentages for age, PEG score, and other continuous measures are not applicable.

[c]Percentages are based on no missing data for each characteristic. The following variables had missing data (N=244 unless noted): sex (n=1 missing), employment status (n=1 missing), home internet access with Zoom capability (n=4 missing), ADI national percentile (n=8 missing).

[d]Wilcoxon rank-sum test.

[e]Chi-square test.

[f]Statistically significant at $P<.05$.

[g]Fisher exact test.

[h]TAPS: tobacco, alcohol, prescription medication, and other substance use.

[i]PEG: Pain, Enjoyment of Life, General Activity.

[j]PROMIS-29: Patient-Reported Outcomes Measurement Information System-29.

[k]Pooled 2-sample *t* test.

[l]CAMS-R: Cognitive and Affective Mindfulness Scale-Revised.

[m]ADI: Area Deprivation Index.

[n]RUCA: rural-urban commuting area.

XSL•FO
RenderX

Table . Univariate logistic regression of predictors of telehealth use (n=244).

Characteristic	Unadjusted	
	OR[a] (95% CI)	*P* value
Sex at birth (n=243; reference: female)		
Male	0.6 (0.3 - 1.1)	.09
Race or ethnicity (n=238; reference: Non-Hispanic White)		
Non-Hispanic Black	0.38 (0.2 - 0.8)	.01[b]
Hispanic	0.25 (0.09 - 0.7)	.008[b]
Other	0.41 (0.1 - 1.2)	.11
Education (n=244; reference: high school degree)		
Less than high school	1.5 (0.6 - 3.5)	.37
More than high school degree	2.7 (1.4 - 5.1)	.002[b]
Employment status (n=243; reference: full-time)		
Not employed	1.03 (0.5 - 2)	.91
Part-time	0.39 (0.2 - 0.9)	.03[b]
Years of back pain (n=244; reference: less than 5 y)		
5 to 15 years	3.1 (1.6 - 5.8)	<.001[b]
More than 15 years	4.8 (2.1 - 11)	<.001[b]
Internet status: do you have internet at home and are you able to access programs like Zoom? (n=240; reference: no)		
Yes	3.7 (1.1 - 12.2)	.02[b]
Comfort levels with Zoom or virtual meetings (n=241; reference: high comfort level)		
Low comfort level	0.4 (0.2 - 0.7)	.004[b]
PEG[c] (n=244)	0.8 (0.7 - 1.01)	.08
PROMIS[d]-29 fatigue (n=244)	1.02 (0.9 - 1.05)	.10

[a]OR: odds ratio.

[b]Statistically significant at $P<.05$.

[c]PEG: Pain, Enjoyment of Life, and General Activity.

[d]PROMIS-29: Patient-Reported Outcomes Measurement Information System-29.

XSL•FO
RenderX

Table . Adjusted logistic regression of predictors of telehealth use (N=233).

Characteristic	Adjusted OR[a] (95% CI)	*P* value
Race or ethnicity (reference: Non-Hispanic White)		
Non-Hispanic Black	0.38 (0.2 - 0.8)	.02[b]
Hispanic	0.21 (0.06 - 0.7)	.01[b]
Other	0.32 (0.1 - 1.1)	.07
Education (reference: high school degree)		
Less than high school	3.2 (1.1 - 8.9)	.03[b]
More than high school degree	2.7 (1.3 - 5.1)	.01[b]
Employment status (reference: full-time)		
Not employed	1.11 (0.5 - 2.4)	.91
Part-time	0.33 (0.1 - 0.9)	.03[b]
Years of back pain (reference: less than 5 y)		
5 to 15 years	3.1 (1.6 - 6.5)	<.001[b]
More than 15 years	3.5 (1.4 - 8.8)	<.001[b]
Internet status: do you have internet at home and are you able to access programs like Zoom? (reference: no)		
Yes	2.2 (0.6 - 8.7)	.25
Comfort levels with Zoom or virtual meeting (reference: high comfort level)		
Low comfort level	0.5 (0.2 - 0.9)	.04[b]

[a]OR: odds ratio.

[b]Statistically significant at *P*<.05.

Among *enabling factors*, access and comfort with technology were strongly associated with telehealth use. Participants who had internet at home and had the ability to use Zoom had a telehealth usage rate of 72.8% compared to 41.7% among those without access (*P*=.04). Comfort with Zoom ranged from "not at all" 39.1% (9/23), to "quite a bit" 83.6% (56/67), to "very much" comfortable 72.2% (78/108); lower comfort levels were associated with lower odds of telehealth use experience (AOR 0.5, 95% CI 0.2 - 0.9). Most of the sample, 169 of 236 (71.6%) participants, were in the low ADI (≤50) national percentile block group, indicating neighborhoods with lower levels of socioeconomic disadvantage. The mean SVI score was 0.7 (SD 0.3), indicating a moderate-to-high level of vulnerability. See Multimedia Appendix 1 for ADI and SVI quartiles, quintiles, mean scores, and specific cutoff points (ie, >50th percentile vs <50th percentile). Overall, annual income, ADI, SVI, and rurality (metro vs nonmetro or rural) were not significant factors associated with telehealth use experience in both the unadjusted and the adjusted models.

Among *needs factors*, years with back pain strongly predicted telehealth use experience: 51.5% (35/68) among those with less than 5 years, 76.5% (88/115) among those with 5 to 15 years, and 83.6% among those with more than 15 years of pain (*P*<.001). In the adjusted model, more than 5 years of pain was associated with higher odds of telehealth use experience (AOR 3.5, 95% CI 1.4 - 8.8). Having received or applying for disability or workers' compensation benefits due to low back pain was not a significant factor in telehealth use. Other need indicators such as PEG, pain impact score, PROMIS-29 domains (sleep disturbance, fatigue, physical function, pain interference, ability to participate in social roles or activities, cognitive function, anxiety, and depression), comorbidity index, and lower back operations were not significantly associated with telehealth use experience.

Discussion

Principal Findings

This cross-sectional analysis examined the individual- and county-level social vulnerability characteristics of those who had and those who never had a telehealth visit at enrollment among participants from a safety-net hospital and a network of federally qualified health centers in the OPTIMUM trial. We found that 174 of 244 (71.3%) participants had prior experience with telehealth. While a large portion of our sample had previously used telehealth, there were many who had not, thus indicating that while telehealth may be well received by many patients, it remains more accessible to certain groups of patients than others [17,38].

Similar to previous research, we found that females reported a higher experience with telehealth usage compared to males [19]. Our analysis indicated that compared to non-Hispanic White participants, non-Hispanic Black and Hispanic participants had

XSL•FO
RenderX

lower odds of telehealth use experience. Given the range of health disparities across the United States, it is reasonable to assume that not all people benefit from using telehealth. In the pre–COVID-19 period, telehealth visits were significantly lower among Hispanic patients than among non-Hispanic White patients and non-Hispanic Black patients [39]. Although telehealth use increased among these groups during the COVID-19 period, disparities persist with reliable internet connectivity, digital literacy, and user comfort with technology among these populations [40,41].

In the adjusted models, we did not find a statistically significant association in our data between internet access and telehealth use experience, which may be explained by factors unique to our sample, such as high number of participants with internet access (228/244, 93.4%) and a large portion of our sample (213/244, 87.3%) having a high school degree or higher, which may have impacted significance. Previous research has found that while internet access is a foundational requirement for telehealth, it is not the only or even the most significant factor determining telehealth use when compared to, or in combination with, education, employment status, and digital literacy [42,43].

Unsurprisingly, lower comfort with Zoom and other virtual meeting platforms was associated with markedly lower odds of telehealth use experience. Prior research has found that the strongest predictor for telehealth use among participants was their comfort with technology rather than access alone [44]. Additionally, we found that in comparison to a high school degree, having less than a high school degree or more than a high school degree was associated with greater odds of previous telehealth use after adjustment. Our findings add to the mixed results of previous studies, which have found either no differences associated with telehealth use or that lower education and income were associated with lower telehealth use [45,46].

Our findings indicated that participants who had 5 or more years of CLBP had higher telehealth use experience, which may have influenced their desire to participate in the telehealth trial. Previous research has shown that patients with chronic pain are more likely to use telehealth because it offers easy access to specialists and reduces the amount of time taken off work to travel to an appointment [47-49].

We found no statistically significant association between ADI, SVI ranking, and the odds of previous telehealth use. Characteristics of our sample may explain possible reasons why ADI and SVI ranking was not significantly associated with previous telehealth. Among 244 participants, a large portion of our sample (n=213, 87.3%) had a high school degree or higher, lived in the metro area (n=222, 90.9%), and had access to the internet at home (n=228, 93.4%). Other research on the association between ADI, SVI, and telehealth usage shows mixed results. While some studies link higher disadvantage (higher ADI or SVI) to lower telehealth usage, others report no significant association between these indices and telehealth, suggesting factors such as age or technology access may play a larger role [50-52].

Strengths and Limitations

Our study has several strengths and limitations. The strengths of our study include a large sample size of an understudied population with CLBP. Additional strengths include the use of reliable and valid measurements. A limitation of this study is that all participants voluntarily enrolled in a telehealth trial and were required to have internet access in some form, which may indicate a higher comfort level with technology and limit the generalizability of the findings. Additionally, participants in our sample were more likely to receive a telehealth visit because they lived in communities with higher telehealth use (eg, urban communities served by large health systems).

Another limitation was the comfort with Zoom or virtual meeting platforms measure, which captured participants' confidence in using these platforms. The measure did not discern whether participants had adequate digital literacy skills, just if they felt confident using those platforms. Other limitations include the following: (1) the cross-sectional nature of the study design, which precludes any ability to draw causal inferences; (2) self-reported variables, introducing the likelihood of recall bias; and (3) the recruitment of most participants from metro areas. Prior research has found that while both metro and rural underserved populations own and use digital health technology, rural residents are less likely to communicate with their health providers using that technology [53].

Implications and Recommendations

The results of our study have important implications. First, given the high economic burden of CLBP and the growing emphasis on evidence-based treatments such as MBSR, telehealth offers a promising alternative for medically underserved populations. Factors that served as enablers included home internet access and greater comfort with Zoom. Targeted outreach and culturally tailored engagement strategies are needed to address disparities related to race or ethnicity and employment status. Policymakers and health care stakeholders must take proactive measures to mitigate these disparities, enhance access, and ensure that telehealth fulfills its potential to improve outcomes for all patients with CLBP.

Conclusions

In a safety-net and federally qualified health center context, prior telehealth experience was common among patients with CLBP willing to enroll in a telehealth-delivered MBSR trial. Although certain predisposing factors (race or ethnicity, employment status) were associated with lower odds of telehealth experience, these factors did not prevent trial enrollment. Thus, targeted outreach to improve digital comfort and inclusive engagement strategies may further reduce disparities and support broader implementation of telehealth for CLBP.

XSL•FO
RenderX

Acknowledgments

The National Center for Complementary and Integrative Health and the National Institutes of Health (NIH) Pragmatic Trials Collaboratory through the NIH HEAL (Helping to End Addiction Long-term) Initiative. The authors used Claude (Anthropic) to assist with generating and formatting Figure 1.

Funding

This work was supported within the National Institutes of Health (NIH) Pragmatic Trials Collaboratory through the NIH HEAL (Helping to End Addiction Long-term) Initiative under award number UG3/UH3AT010621 administered by the National Center for Complementary and Integrative Health (NCCIH). This work also received logistical and technical support from the PRISM (Pragmatic and Implementation Studies for the Management of Pain to Reduce Opioid Prescribing) Resource Coordinating Center under award numbers U24AT010961 and UG3AT010621 from the NIH through the NIH HEAL Initiative. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NCCIH, the NIH, or its HEAL Initiative.

Data Availability

The datasets generated in this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: MVF, NM, PTN
Data curation: PTN
Formal analysis: PTN
Funding acquisition: NM
Methodology: PTN, JW, KRF
Supervision: NM
Validation: MVF, NM, JW, PTN
Visualization: PTN
Writing – original draft: MVF
Writing – review & editing: MVF, PTN, NM, JW, KRF

Conflicts of Interest

None declared.

Multimedia Appendix 1
Different ways to interpret Area Deprivation Index and Social Vulnerability Index.
[DOCX File, 18 KB - mhealth_v14i1e82093_app1.docx]

Checklist 1
STROBE checklist.
[DOCX File, 34 KB - mhealth_v14i1e82093_app2.docx]

References

1. Goode AP, Cleveland RJ, Schwartz TA, et al. Relationship of joint hypermobility with low back pain and lumbar spine osteoarthritis. BMC Musculoskelet Disord 2019 Apr 9;20(1):158. [doi: 10.1186/s12891-019-2523-2] [Medline: 30967130]
2. 57 back pain statistics-understanding the prevalence and impact of back pain: a statistical overview. April ABA. 2025 Feb 18. URL: https://www.aprilaba.com/resources/57-back-pain-statistics [accessed 2026-07-03]
3. Morlion B. Chronic low back pain: pharmacological, interventional and surgical strategies. Nat Rev Neurol 2013 Aug;9(8):462-473. [doi: 10.1038/nrneurol.2013.130] [Medline: 23817349]
4. Deyo RA, Mirza SK, Turner JA, Martin BI. Overtreating chronic back pain: time to back off? J Am Board Fam Med 2009;22(1):62-68. [doi: 10.3122/jabfm.2009.01.080102] [Medline: 19124635]
5. Day MA, Ciol MA, Mendoza ME, et al. The effects of telehealth-delivered mindfulness meditation, cognitive therapy, and behavioral activation for chronic low back pain: a randomized clinical trial. BMC Med 2024 Apr 12;22(1):156. [doi: 10.1186/s12916-024-03383-2] [Medline: 38609994]
6. Qaseem A, Wilt TJ, McLean RM, et al. Noninvasive treatments for acute, subacute, and chronic low back pain: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2017 Apr 4;166(7):514-530. [doi: 10.7326/M16-2367] [Medline: 28192789]
7. Kabat-Zinn J. Full Catastrophe Living Using the Wisdom of Your Body and Mind to Face Stress, Pain, and Illness: Delta Trade Paperbacks; 2009.

XSL•FO
RenderX

8. Baker MB, Liu EC, Bully MA, et al. Overcoming barriers: a comprehensive review of chronic pain management and accessibility challenges in rural America. Healthcare (Basel) 2024 Sep 4;12(17):1765. [doi: 10.3390/healthcare12171765] [Medline: 39273789]
9. Nguyen LH, Dawson JE, Brooks M, Khan JS, Telusca N. Disparities in pain management. Anesthesiol Clin 2023 Jun;41(2):471-488. [doi: 10.1016/j.anclin.2023.03.008] [Medline: 37245951]
10. Cyr ME, Etchin AG, Guthrie BJ, Benneyan JC. Access to specialty healthcare in urban versus rural US populations: a systematic literature review. BMC Health Serv Res 2019 Dec 18;19(1):974. [doi: 10.1186/s12913-019-4815-5] [Medline: 31852493]
11. Burke CA, Fillipo R, Epplein M, Brookhart MA, Bosworth HB, Goode AP. Racial and ethnic disparities in the incidence and prevalence of low back pain in the United States: a systematic review. Arthritis Care Res (Hoboken) 2026 Jan;78(1):141-153. [doi: 10.1002/acr.25665] [Medline: 41083919]
12. Chou R, Qaseem A, Snow V, et al. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. Ann Intern Med 2007 Oct 2;147(7):478-491. [doi: 10.7326/0003-4819-147-7-200710020-00006] [Medline: 17909209]
13. What is telehealth? How is telehealth different from telemedicine? HealthIT.gov. 2019. URL: https://www.healthit.gov/faq/what-telehealth-how-telehealth-different-telemedicine [accessed 2026-07-03]
14. Grigsby B, Brega AG, Bennett RE, et al. The slow pace of interactive video telemedicine adoption: the perspective of telemedicine program administrators on physician participation. Telemed J E Health 2007 Dec;13(6):645-656. [doi: 10.1089/tmj.2007.0090] [Medline: 18092926]
15. Al Mohaya MA, Almaziad MM, Al-Hamad KA, Mustafa M. Telemedicine among oral medicine practitioners during COVID-19 pandemic and its future impact on the specialty. Risk Manag Healthc Policy 2021;14:4369-4378. [doi: 10.2147/RMHP.S325777] [Medline: 34707420]
16. Barnhill JL, Castro G, Lathren C, et al. The hidden complexity of virtual mindfulness-based group medical visits: comfort, challenge, and the influence of social determinants of health. Glob Adv Integr Med Health 2025;14:27536130251326938. [doi: 10.1177/27536130251326938] [Medline: 40162193]
17. Franciosi EB, Tan AJ, Kassamali B, et al. The impact of telehealth implementation on underserved populations and no-show rates by medical specialty during the COVID-19 pandemic. Telemed J E Health 2021 Aug;27(8):874-880. [doi: 10.1089/tmj.2020.0525] [Medline: 33826411]
18. Williams C, Shang D. Telehealth usage among low-income racial and ethnic minority populations during the COVID-19 pandemic: retrospective observational study. J Med Internet Res 2023 May 12;25:e43604. [doi: 10.2196/43604] [Medline: 37171848]
19. Xu P, Hudnall M, Zhao S, Raja U, Parton J, Lewis D. Pandemic-triggered adoption of telehealth in underserved communities: descriptive study of pre- and postshutdown trends. J Med Internet Res 2022 Jul 15;24(7):e38602. [doi: 10.2196/38602] [Medline: 35786564]
20. Popescu I, Fingar KR, Cutler E, Guo J, Jiang HJ. Comparison of 3 safety-net hospital definitions and association with hospital characteristics. JAMA Netw Open 2019 Aug 2;2(8):e198577. [doi: 10.1001/jamanetworkopen.2019.8577] [Medline: 31390034]
21. Kim HK, Lee M. Factors associated with health services utilization between the years 2010 and 2012 in Korea: using Andersen's behavioral model. Osong Public Health Res Perspect 2016 Feb;7(1):18-25. [doi: 10.1016/j.phrp.2015.11.007] [Medline: 26981338]
22. Andersen R, Lewis SZ, Giachello AL, Aday LA, Chiu G. Access to medical care among the Hispanic population of the southwestern United States. J Health Soc Behav 1981 Mar;22(1):78-89. [Medline: 7240708]
23. Gelberg L, Andersen RM, Leake BD. The behavioral model for vulnerable populations: application to medical care use and outcomes for homeless people. Health Serv Res 2000 Feb;34(6):1273-1302. [Medline: 10654830]
24. Andersen RM. National health surveys and the behavioral model of health services use. Med Care 2008 Jul;46(7):647-653. [doi: 10.1097/MLR.0b013e31817a835d] [Medline: 18580382]
25. Babitsch B, Gohl D, von Lengerke T. Re-revisiting Andersen's Behavioral Model of Health Services Use: a systematic review of studies from 1998-2011. Psychosoc Med 2012;9:Doc11. [doi: 10.3205/psm000089] [Medline: 23133505]
26. Andersen RM, Davidson PL, Baumeister SE. Improving access to care in America. In: Changing the US Health Care System: Key Issues in Health Services Policy and Management, 3rd edition: Jossey-Bass; 2007:3-31.
27. Greco CM, Gaylord SA, Faurot K, et al. The design and methods of the OPTIMUM study: a multisite pragmatic randomized clinical trial of a telehealth group mindfulness program for persons with chronic low back pain. Contemp Clin Trials 2021 Oct;109:106545. [doi: 10.1016/j.cct.2021.106545] [Medline: 34455111]
28. McWilliams LA, Kowal J, Wilson KG. Development and evaluation of short forms of the Pain Catastrophizing Scale and the Pain Self-efficacy Questionnaire. Eur J Pain 2015 Oct;19(9):1342-1349. [doi: 10.1002/ejp.665] [Medline: 25766681]
29. Feldman G, Hayes A, Kumar S, Greeson J, Laurenceau JP. Mindfulness and emotion regulation: the development and initial validation of the Cognitive and Affective Mindfulness Scale-Revised (CAMS-R). J Psychopathol Behav Assess 2007;29(3):177-190. [doi: 10.1007/s10862-006-9035-8]

XSL•FO
RenderX

30. Spielman SE, Tuccillo J, Folch DC, et al. Evaluating social vulnerability indicators: criteria and their application to the Social Vulnerability Index. Nat Hazards 2020;100(1):417-436. [doi: 10.1007/s11069-019-03820-z]
31. Kind AJH, Jencks S, Brock J, et al. Neighborhood socioeconomic disadvantage and 30-day rehospitalization: a retrospective cohort study. Ann Intern Med 2014 Dec 2;161(11):765-774. [doi: 10.7326/M13-2946] [Medline: 25437404]
32. Beneciuk JM, Bishop MD, Fritz JM, et al. The STarT back screening tool and individual psychological measures: evaluation of prognostic capabilities for low back pain clinical outcomes in outpatient physical therapy settings. Phys Ther 2013 Mar;93(3):321-333. [doi: 10.2522/ptj.20120207] [Medline: 23125279]
33. Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. J Clin Epidemiol 1994 Nov;47(11):1245-1251. [doi: 10.1016/0895-4356(94)90129-5] [Medline: 7722560]
34. Dewitt B, Feeny D, Fischhoff B, et al. Estimation of a preference-based summary score for the Patient-Reported Outcomes Measurement Information System: the PROMIS®-Preference (PROPr) scoring system. Med Decis Making 2018 Aug;38(6):683-698. [doi: 10.1177/0272989X18776637] [Medline: 29944456]
35. Krebs EE, Lorenz KA, Bair MJ, et al. Development and initial validation of the PEG, a three-item scale assessing pain intensity and interference. J Gen Intern Med 2009 Jun;24(6):733-738. [doi: 10.1007/s11606-009-0981-1] [Medline: 19418100]
36. McNeely J, Wu LT, Subramaniam G, et al. Performance of the Tobacco, Alcohol, Prescription Medication, and Other Substance Use (TAPS) tool for substance use screening in primary care patients. Ann Intern Med 2016 Nov 15;165(10):690-699. [doi: 10.7326/M16-0317] [Medline: 27595276]
37. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med 2007 Oct 16;4(10):e296. [doi: 10.1371/journal.pmed.0040296] [Medline: 17941714]
38. Shaver J. The state of telehealth before and after the COVID-19 pandemic. Prim Care 2022;49(4):517-530. [doi: 10.1016/j.pop.2022.04.002] [Medline: 36357058]
39. White-Williams C, Liu X, Shang D, Santiago J. Use of telehealth among racial and ethnic minority groups in the United States before and during the COVID-19 pandemic. Public Health Rep 2023;138(1):149-156. [doi: 10.1177/00333549221123575] [Medline: 36113138]
40. Campos-Castillo C, Anthony D. Racial and ethnic differences in self-reported telehealth use during the COVID-19 pandemic: a secondary analysis of a US survey of internet users from late March. J Am Med Inform Assoc 2021 Jan 15;28(1):119-125. [doi: 10.1093/jamia/ocaa221] [Medline: 32894772]
41. Iasiello JA, Rajan A, Zervos E, Parikh AA, Snyder RA. Racial differences in patient-reported access to telehealth: an important and unmeasured social determinant of health. JCO Oncol Pract 2023 Dec;19(12):1215-1223. [doi: 10.1200/OP.23.00006] [Medline: 37844269]
42. Labadorf J, Nichols M, Williams T, Cunanan C, D'Anza B. A smartphone is not enough: telehealth attendance and the digital divide. Health Care Sci 2025 Aug;4(4):259-268. [doi: 10.1002/hcs2.70033] [Medline: 40861513]
43. Kemp M, Rising KL, Laynor G, et al. Barriers to telehealth uptake and use: a scoping review. JAMIA Open 2025;8(2):ooaf019. [doi: 10.1093/jamiaopen/ooaf019] [Medline: 40109500]
44. Yamazaki KG, Hewitt L, Torres L, Colon-Vazquez K, Rogers P, Wang G. Technology access, digital literacy, and enrollment support preferences in a federally qualified health center: cross-sectional study. JMIR Form Res 2026 Jan 5;10:e78850. [doi: 10.2196/78850] [Medline: 41490221]
45. Thomson MD, Mariani AC, Williams AR, Sutton AL, Sheppard VB. Factors associated with use of and satisfaction with telehealth by adults in rural Virginia during the COVID-19 pandemic. JAMA Netw Open 2021 Aug 2;4(8):e2119530. [doi: 10.1001/jamanetworkopen.2021.19530] [Medline: 34351404]
46. Kini S, Duluk D, Weinstein J. Modeling the impact of digital readiness in recruiting and sustaining underrepresented groups: data from the All of Us research program. Front Digit Health 2023;4:1082098. [doi: 10.3389/fdgth.2022.1082098] [Medline: 37124163]
47. Tauben DJ, Langford DJ, Sturgeon JA, et al. Optimizing telehealth pain care after COVID-19. Pain 2020 Nov;161(11):2437-2445. [doi: 10.1097/j.pain.0000000000002048] [Medline: 32826752]
48. Pritzlaff SG, Singh N, Sanghvi C, Schatman ME. Telehealth is crucial for pain medicine: patients and doctors are at the brink, and Medicare must act now. J Pain Res 2025;18:1057-1059. [doi: 10.2147/JPR.S524840] [Medline: 40046186]
49. McLaughlin KH, Fritz JM, Minick KI, et al. Examining the relationship between individual patient factors and substantial clinical benefit from telerehabilitation among patients with chronic low back pain. Phys Ther 2024 Mar 1;104(3):pzad180. [doi: 10.1093/ptj/pzad180] [Medline: 38157307]
50. Bose S, Dun C, Zhang GQ, Walsh C, Makary MA, Hicks CW. Medicare beneficiaries in disadvantaged neighborhoods increased telemedicine use during the COVID-19 pandemic. Health Aff (Millwood) 2022 May;41(5):635-642. [doi: 10.1377/hlthaff.2021.01706] [Medline: 35500186]
51. Boudreau E, Sutherland A, Bozzi D, Canterberry M, Sylwestrzak G. Primary care telehealth utilization by access-challenged populations in Medicare Advantage. Health Aff Sch 2024;2(10):qxae120. [doi: 10.1093/haschl/qxae120] [Medline: 39391688]

XSL•FO
RenderX

52. Ostovari M, Zhang Z, Patel V, Jurkovitz C. Telemedicine and health disparities: association between the area deprivation index and primary care telemedicine utilization during the COVID-19 pandemic. J Clin Transl Sci 2023;7(1):e168. [doi: 10.1017/cts.2023.580] [Medline: 37588680]
53. Okobi E, Adigun AO, Ozobokeme OE, et al. Examining disparities in ownership and use of digital health technology between rural and urban adults in the US: an analysis of the 2019 Health Information National Trends Survey. Cureus 2023 May;15(5):e38417. [doi: 10.7759/cureus.38417] [Medline: 37273368]

Abbreviations

ADI: Area Deprivation Index
AOR: adjusted odds ratio
CAMS-R: Cognitive and Affective Mindfulness Scale-Revised
CLBP: chronic low back pain
HIPAA: Health Insurance Portability and Accountability Act
MBSR: mindfulness-based stress reduction
OPTIMUM: Optimizing Pain Treatment in Medical Settings Using Mindfulness
OR: odds ratio
PEG: Pain, Enjoyment of Life, and General Activity
PROMIS-29: Patient-Reported Outcomes Measurement Information System-29
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
SVI: Social Vulnerability Index
TAPS: tobacco, alcohol, prescription medication, and other substance use
UNC: University of North Carolina

Edited by A Stone, L Buis; submitted 08.Aug.2025; peer-reviewed by L Benjamin, M Justice, Y Mallisetty; revised version received 29.Apr.2026; accepted 29.Apr.2026; published 29.Jul.2026.

Please cite as:
Foster MV, Nguyen PT, Weinberg J, Faurot KR, Morone N
The Impact of Telehealth Use Experience on Recruitment of Underserved Populations Into a Telehealth-Delivered Mindfulness-Based Trial for Patients With Chronic Low Back Pain: Cross-Sectional Survey
JMIR Mhealth Uhealth 2026;14:e82093
URL: https://mhealth.jmir.org/2026/1/e82093
doi:10.2196/82093

© Marva V Foster, Phuong Tra Nguyen, Janice Weinberg, Keturah R Faurot, Natalia Morone. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 29.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Unlimited Smartphone Data Plans in Older Adults With Data Deprivation: Quasi-Experimental Study

Seoyeong Choi[1], BS; Eunjeong Choi[1], PhD; Suk-Yong Jang[2,3], JD, MD, PhD

[1]Department of Public Health, Graduate School, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, Republic of Korea

[2]Department of Biohealth Industry, Policy Analysis Division, Graduate School of Transdisciplinary Health Science, Yonsei University, 50 Yonsei-ro, Seodaemun-gu,, Seoul, Republic of Korea

[3]Department of Health Policy and Management, Graduate School of Public Health, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, Republic of Korea

Corresponding Author:
Suk-Yong Jang, JD, MD, PhD
Department of Biohealth Industry, Policy Analysis Division, Graduate School of Transdisciplinary Health Science, Yonsei University, 50 Yonsei-ro, Seodaemun-gu,, Seoul, Republic of Korea

Abstract

Background: The growth of mobile health has underscored the critical importance of equitable internet access in promoting healthy aging. Among older adults, particularly those in digitally underserved populations, access to mobile data is often limited due to affordability and technological barriers, leading to a phenomenon known as “data deprivation.” This form of digital inequality limits the older adults’ ability to participate in social, recreational, and community-based activities, which are protective against isolation and decline in later life. In South Korea, where unlimited smartphone data plans have become increasingly accessible, a unique opportunity exists to examine the real-world association of improved data accessibility on older adults’ social lives and digital engagement.

Objective: This study aimed to examine whether switching to unlimited smartphone data plans enhances social participation among older adults in South Korea. It also explored whether this relationship differs by demographic and socioeconomic characteristics, such as age, gender, region, household composition, and income level. The study focused on offline domains of social participation, including hobby gatherings, religious services, volunteering, and routine activities such as shopping.

Methods: Data were drawn from the Korea Media Panel Survey (2016 - 2022), a nationally representative longitudinal dataset. The final sample included 5021 individuals. Social participation was measured using a self-reported 8-item scale (8 - 64 points) covering outdoor activities, volunteering, and more. A difference-in-differences approach was used to assess the association of switching from limited to unlimited smartphone data plans on social activity scores. Subgroup analyses examined heterogeneity by gender, household composition, income, and region.

Results: Overall, among individuals aged 60 years and older, switching to an unlimited data plan was not associated with a statistically significant change in social activity scores. However, within this group, those aged 70 years and older showed a more notable—though not statistically significant—improvement (differential of 1.54, 95% CI −0.41 to 3.50). In contrast, older men living alone experienced a significant differential improvement of 6.44 (95% CI 3.39 to 9.50) points, compared with those who remained on limited plans.

Conclusions: Although the overall association of unlimited data plans was limited, certain vulnerable subgroups—particularly older men living alone—experienced meaningful gains. These findings suggest that improving mobile data accessibility may enhance social engagement among digitally underserved older adults. Compared with more complex or resource-intensive interventions, expanding access to unlimited smartphone data plans may offer a relatively simple and scalable strategy to support healthy aging and reduce social isolation.

(JMIR Mhealth Uhealth 2026;14:e68930) doi:10.2196/68930

KEYWORDS

data accessibility; data deprivation; social participation; older adults; unlimited data plans; wellness

Introduction

Across the Organization for Economic Cooperation and Development (OECD) countries, the population aged 65 years and older has significantly increased from under 9% in 1960 to 18% by 2021 [1]. The average of old age–related social expenditure (including pensions, health care, and long-term care) as a percentage of gross domestic product for OECD countries increased almost 1.5 times from 2000 to 2022, highlighting the growing economic burdens of an aging

XSL•FO
RenderX

population [2]. In South Korea, the proportion of individuals aged 65 years and older is anticipated to reach approximately 40% by 2050, making it the fastest-aging nation among OECD countries [3]. Furthermore, South Korea has the highest older adult poverty rate among the 37 OECD member countries [4]. This significant aging issue necessitates effective policies to improve the lives of older adults. South Korea's approach to this trend could provide valuable insights for other countries grappling with aging-related challenges.

The public health objective for older adults is to promote and maintain functional well-being, known as healthy aging. Successful aging requires preventing diseases, maintaining physical and mental function, and actively engaging in social participation, a principle endorsed by the World Health Organization (WHO) for national public health policies amidst population aging challenges [5]. Social scientists have long emphasized the importance of the relationship between social networks (active social participation) and health, arguing that social networks influence human health through 5 pathways: provision of social support, social influence, social participation and attachment, access to material goods and resources, and negative social interactions [6]. Social participation fostering healthy relationships is gaining attention as a concept of wellness and geriatric health [7]. The classification of social participation varies depending on the criteria, but generally aligns with the framework proposed by Berkman et al [8]. They noted that examples of social activity, known as social participation or engagement, may involve meeting friends, attending events, volunteering, fulfilling occupational duties, or participating in group recreational activities [8].

Previous studies have demonstrated a relationship between the quality of life, health outcomes of older adults, and social participation [9-11]. Specifically, reports indicate that older adult social participation is influenced by gender [12], region of residence [13], and household members [14]. However, traditional methods to enhance older adult social participation must evolve as society transitions into a digital era. Digital health is being highlighted as a mediator for this [15]. Studies have shown that higher smartphone use among older adults correlates with greater overall life satisfaction [16,17].

Older adults face a digital divide due to the lack of an environment that allows them to freely access the internet compared with other generations [18]. While prior research has largely focused on digital literacy, less attention has been paid to foundational access barriers—particularly mobile data affordability and availability [19,20]. Smartphone usage involves multiple dimensions, including device ownership, connectivity, and digital skills. However, access to stable and affordable data remains a prerequisite for all subsequent digital engagement. In this context, we introduce the concept of "data deprivation," defined as the lack of consistent and affordable access to mobile data services. Unlike digital exclusion, which refers to general access to digital technologies, or "digital literacy," which relates to skills and competencies, data deprivation emphasizes material constraints—such as limited data plans or high costs—that restrict older adults' ability to participate online. We argue that this form of deprivation is a foundational driver of digital inequality, particularly among socioeconomically vulnerable older adults, and must be addressed before other digital inclusion efforts can be effective. Furthermore, data accessibility is not only essential for general social engagement but also for promoting digital health participation among older adults. In the context of mobile health (mHealth), which emphasizes the use of mobile devices to support public health and clinical care, access to mobile data accessibility serves as a bridge linking digital health technologies with real-world social engagement. Recent studies have shown that mHealth tools increasingly integrate features for mental wellness, social connection, and lifestyle tracking [21,22].

Without affordable and stable mobile internet accessibility, older adults may be unable to use health-related mobile apps, participate in online health communities, or benefit from remote monitoring technologies that support chronic disease management [23]. While previous discussions around older adult health have focused on economic or geographic access, this study brings attention to mobile data accessibility as a foundational but often overlooked determinant [24]. In this regard, our study aims to extend the discourse in digital equity by highlighting how mobile data accessibility, a prerequisite for mHealth use, may influence health-related behaviors such as social participation.

The relationship between internet use and social activity has long been debated through 2 opposing theoretical perspectives. The social reward hypothesis suggests that internet use enhances psychological well-being and social connectedness, especially among those with limited offline interactions [25]. In contrast, the displacement hypothesis argues that increased time online can replace face-to-face interactions, ultimately reducing social engagement and well-being [26]. While these theories have primarily been applied to younger populations, limited research has examined their relevance to older adults, who may have distinct digital usage patterns and social needs. While the social reward hypothesis has been the dominant perspective for adolescents prone to internet addiction [27,28], it may not be directly applicable to older adults. Most existing research has focused on reducing smartphone or internet use [29,30], with limited attention given to how these technologies might benefit older adults. As a result, it remains unclear which theoretical perspective—reward or displacement—better explains digital engagement among older adults, highlighting the need for empirical investigation in this population.

A unique aspect of the South Korean context is the widespread availability of unlimited smartphone data plans. Unlike expensive limited plans or Wi-Fi, which may pose cost or spatial restrictions, unlimited plans allow users to access the internet freely, regardless of time or place [31]. This feature offers a rare opportunity to examine whether removing mobile data constraints can enhance digital engagement and social participation among older adults. The observed effects in vulnerable subgroups suggest that improving mobile data accessibility could complement existing policies aimed at supporting social participation among older adults. Compared with complex and resource-intensive interventions, enhancing access to unlimited data plans could offer a relatively simple and scalable strategy. South Korea's unique context—characterized by rapid population aging, high older

XSL•FO
RenderX

adult poverty, and the widespread availability of unlimited data plans—provides an appropriate setting to examine the association between mobile data accessibility and social participation. Additionally, we examined how gender, residence, and household composition moderated the relationship between data accessibility and social participation in this context.

Methods

Data

In this study, we used data from the Korea Media Panel Survey (2016-2022). The Korea Media Panel Survey, constructed by the Korea Information Society Development Institute and funded by the Korean government, targets the same sample over an extended period to accumulate data [32]. This dataset aims to track changes in media environments and usage behaviors of households and individuals, providing fundamental data for in-depth research and policy development. We conducted face-to-face interviews with over 5000 households and individuals annually across 17 regions nationwide. The survey is divided into two main parts: one investigating media device ownership, subscription to broadcasting and communication services, and expenditure status; and the other examining life satisfaction, self-esteem, and similar factors.

Study Population

In this study, data from 9788 participants in the 2016 Korea Media Panel Survey were used as baseline data. Participants who were already using a smartphone unlimited data plan in 2016 (n=1854) and those not present in both 2016 and 2022 (n=2396) were excluded. Participants with missing values in key variables such as smartphone plan type, social activity score, or baseline covariates (n=517) were also excluded. The final analytic sample consisted of 5021 participants.

Variables

Social Participation

The dependent outcome in this study was social participation, for which the Korea Media Panel Survey used an 8-item self-report questionnaire [32]. Participants rated the frequency of their engagement in 8 types of activities on an 8-point Likert scale, ranging from “almost never” to “almost every day.” The activities included outdoor sports, social gatherings, volunteer and political activities, religious participation, hobbies, attending performances or exhibitions, watching sports, and outdoor shopping. These categories reflect the domains of social activity proposed by Berkman et al [8]. Total scores ranged from 8 to 64, with higher scores indicating more frequent social participation. This variable was only measured in 2016 and 2022.

Assessment of Unlimited Smartphone Data Plan

A smartphone unlimited data plan allows users to consume mobile data without any restrictions. Therefore, users can freely engage in activities such as web browsing, streaming, gaming, and social media. In this study, we use “data accessibility” to refer to the extent to which users can use mobile data without limitations such as data caps or additional charges. This is conceptually distinct from “data access,” which indicates the basic physical or technical availability of an internet connection or device. Based on changes in the unlimited smartphone data plan, which was the primary independent variable, the participants were divided into 2 groups: those who switched to an unlimited data plan in 2022 from a nonunlimited plan in 2016 (coded as “1”), and those who did not switch (coded as “0”).

Other Covariates

To address the potential confounders that might influence both the independent and dependent variables, covariates were selected based on Berkman et al’s [8] social integration framework, which emphasizes how demographic and socioeconomic factors shape health through pathways such as social participation, access to resources, and social support. Additionally, variable selection was informed by prior empirical studies on social participation in later life, including Finlay and Kobayashi [33]. Specifically, we controlled for demographic characteristics (age, gender, educational attainment, marital status, employment status, household composition, and region) and socioeconomic status (household income quartile). Variables were categorized as follows: age groups (under 20, 20 - 29, 30 - 39, 40 - 49, 50 - 59, 60 - 69, and ≥70 years); educational attainment (below middle school, high school, and college or higher); marital status (married, divorced or separated, and unmarried); employment status (unemployed, salaried employee, and self-employed or unpaid family worker); region (capital area, metropolitan, and city or province); household income quartiles; and household composition (living alone, living with a spouse only, and living with other household members excluding a spouse).

Statistical Analysis

The data analysis was performed in 3 parts. First, to evaluate the covariate balance between the unlimited data plan and limited data plan groups, baseline characteristics were compared using standardized differences, with a difference of less than 0.1 (10%) typically considered insignificant. All variables were measured in 2016. For the subsequent analysis, only participants who did not use unlimited smartphone data in 2016 were selected and analyzed. Second, we applied a difference-in-differences (DID) regression model to estimate the effect of adopting unlimited smartphone data plans on social activity scores. The model included an interaction term between a treatment indicator (1=adopted unlimited plan by 2022; 0=remained on limited plan) and a time indicator (1=2022; 0=2016), while adjusting for baseline covariates such as age, gender, education, marital status, employment, income, household composition, region, and Wi-Fi access. SEs were clustered at the individual level. Third, a subgroup analysis was conducted for individuals aged 60 years and older, stratifying the sample by household composition, household income quartiles, and region, to explore whether the effects of adopting unlimited smartphone data plans differed across these demographic and household characteristics. Sampling weights were constructed following the procedures described in the Korea Media Panel Survey user guide, including initial design weights, nonresponse adjustment, and population benchmarking through raking ratio, as outlined in the panel survey’s

XSL•FO
RenderX

methodological documentation. These weights were applied in all descriptive and regression analyses [34]. For baseline characteristics, we reported descriptive statistics and assessed between-group balance using standardized mean differences. CIs for regression estimates were obtained using robust standard errors clustered at the individual level. All statistical analyses were performed using SAS (Statistical Analysis System; version 9.4 M6; SAS Institute Inc).

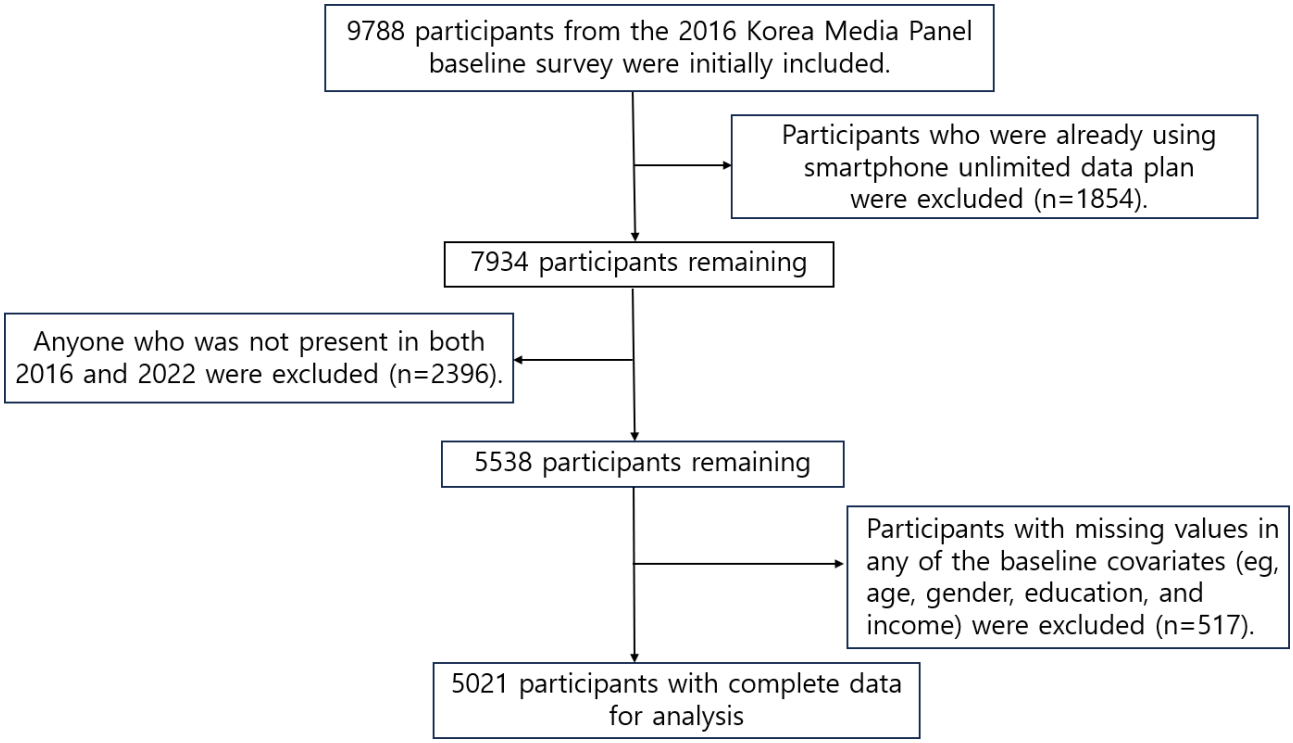
Ethical Considerations

This study used data from the Korea Media Panel Survey, a resource openly available for academic research. As the data do not contain identifiable private information, ethical approval was exempted by the institutional review board of Severance Hospital, South Korea (4-2024-0212). Informed consent was obtained from participants at the time of the original data collection, and the dataset was fully anonymized prior to analysis. No additional consent was required for this secondary data analysis. Participants' privacy and confidentiality were protected through the use of deidentified data. No compensation was provided, as this study involved secondary analysis of an existing dataset.

Results

The final analytic sample consisted of 5021 participants with complete data (Figure 1). Table 1 presents the general characteristics of the study populations at baseline according to the smartphone unlimited data plan. Men outnumbered women among those who used unlimited data plans, who were mostly in their 40s and under 20s.

Figure 1. Flow diagram for study sample.



XSL•FO
RenderX

Table . General characteristics of study participants who did not use unlimited data plans at baseline (2016).[a]

Variables	Unlimited data plan[b]			
	Total (n=5021), n (%)	No→Yes (n=1404), n (%)	No→No (n=3617), n (%)	Standardized difference
Age (y)				0.907
<20	761 (15.16)	297 (21.15)	464 (12.83)	
20 - 29	291 (5.80)	148 (10.54)	143 (3.95)	
30 - 39	444 (8.84)	217 (15.46)	227 (6.28)	
40 - 49	941 (18.74)	364 (25.93)	577 (15.95)	
50 - 59	893 (17.79)	247 (17.59)	646 (17.86)	
60 - 69	786 (15.65)	97 (6.91)	689 (19.05)	
≥70	905 (18.02)	34 (2.42)	871 (24.08)	
Gender				0.190
Man	2184 (43.50)	706 (50.28)	1478 (40.86)	
Woman	2837 (56.50)	698 (49.72)	2139 (59.14)	
Education				0.731
Under middle school	2073 (41.29)	294 (20.94)	1779 (49.18)	
High school	1707 (34)	562 (40.03)	1145 (31.66)	
College and higher	1241 (24.72)	548 (39.03)	693 (19.16)	
Marriage status				0.485
Married	3175 (63.23)	809 (57.62)	2366 (65.41)	
Divorced or separated	601 (11.97)	66 (4.70)	535 (14.79)	
Unmarried	1245 (24.80)	529 (37.68)	716 (19.80)	
Employment status				0.432
Unemployed	2578 (51.34)	645 (45.94)	1933 (53.44)	
Salaried employee	1619 (32.24)	614 (43.73)	1005 (27.79)	
Self-employed or unpaid family worker	824 (16.41)	145 (10.33)	679 (18.77)	
Household income				0.550
Low	2322 (46.25)	460 (32.76)	1862 (51.48)	
Mid-low	2000 (39.83)	664 (47.29)	1336 (36.94)	
Mid-high	587 (11.69)	233 (16.60)	354 (9.79)	
High	112 (2.23)	47 (3.35)	65 (1.80)	
Region[c]				0.238
Capital region	1612 (32.11)	540 (38.46)	1072 (29.64)	
Metropolitan	1288 (25.65)	377 (26.85)	911 (25.19)	
City or province	2121 (42.24)	487 (34.69)	1634 (45.18)	
Wi-Fi installation status[d]				0.455
Yes	2699 (53.75)	926 (65.95)	1773 (49.02)	
No	2322 (46.25)	478 (34.05)	1844 (50.98)	

XSL•FO
RenderX

Variables	Unlimited data plan[b]			
	Total (n=5021), n (%)	No→Yes (n=1404), n (%)	No→No (n=3617), n (%)	Standardized difference
Household composition				0.559
Single-person household	347 (6.91)	46 (3.28)	301 (8.32)	
Couple-only household	1226 (24.42)	166 (11.82)	1060 (29.31)	
Multigenerational household	3448 (68.67)	1192 (84.90)	2256 (62.37)	

[a]Values with unweighted frequency (weighted %) are presented. Standardized differences were used to compare baseline characteristics between groups. A value <0.1 was considered indicative of negligible imbalance.

[b]Unlimited smartphone data refers to a mobile plan offering unrestricted access to data usage without caps or limitations.

[c]The capital region of South Korea, encompassing Seoul and its surrounding metropolitan areas, serves as the political, economic, and cultural hub of the country, while metropolitan cities represent major urban areas and their environments, functioning as significant administrative and economic centers. City or province divisions in South Korea act as regional governmental bodies responsible for local governance and administration.

[d]Indicates whether Wi-Fi (wireless internet) is installed within the household.

Table 2 presents DID estimates of changes in social activity score before and after switching to an unlimited smartphone data plan by age group. Among people aged 60 years and older, the social activity score increased by 0.43 (95% CI −0.65 to 1.50; *P*=.44) points for those who switched from limited plans to unlimited plans between 2016 and 2022, though this change was not significant. When further stratified by age, individuals aged 70 years and older showed a numerically larger differential improvement of 1.54 (95% CI −0.41 to 3.50; *P*=.12) points. While suggestive, this finding was not statistically significant and should be interpreted with caution.

Table . Difference-in-differences estimates of changes in social activity score before and after using smartphone unlimited data plan.

Outcome	Unadjusted means (social activity score[a])						Absolute difference in differences		
	Limited[b,c]→Limited			Limited[c]→Unlimited[d]			Adjusted[e] (95% CI)	SE	*P* value
	Before[f]	After[f]	Difference	Before	After	Difference			
Age (y)									
<20	21.55	21.58	0.03	21.55	22.78	1.23	1.24 (–0.58 to 3.07)	0.93	.18
20-39	22.32	24.32	2.00	22.32	23.97	1.65	0.12 (–1.00 to 1.23)	0.57	.84
40-59	20.53	21.73	1.20	20.53	21.70	1.17	–0.44 (–1.19 to 0.32)	0.38	.25
≥60	17.16	17.83	0.67	17.16	20.06	2.90	0.43 (–0.65 to 1.50)	0.55	.44
60-69	17.99	19.65	1.66	17.99	20.16	2.17	–0.29 (–1.60 to 1.03)	0.67	.67
≥70	16.43	16.91	0.48	16.43	19.78	3.35	1.54 (–0.41 to 3.50)	1.00	.12

[a]Social activity score refers to the total combined score of 8 social participation and the maximum score is 64 points.

[b]Limited: smartphone limited data plan.

[c]The baseline excludes smartphone users with unlimited data plans.

[d]Unlimited: smartphone unlimited data plan.

[e]Adjusted for age, gender, education, marriage status, employment status, household income, region, Wi-Fi installation status, and family members.

[f]"Before" and "After" indicate the mean social activity score in the pre- and post-periods, respectively.

Gender-stratified subgroup analyses (Tables 3 and 4) indicated an increase in the estimated social activity score among the older adult group that switched to unlimited smartphone plans, stratified by independent variables (household composition, household income, and region). Particularly for older adult men living alone, the differential improvement was 6.44 (95% CI 3.39 to 9.50; *P*<.001) points higher for those who switched to unlimited data plans than those who remained on limited plans

XSL•FO
RenderX

from 2016 to 2022. The estimate for socially isolated older men was higher in magnitude; however, as subgroup analyses are exploratory, these results should be interpreted with caution. Similarly, older adult men living in metropolitan areas were 3.47 (95% CI 0.15 to 6.78; *P*=.04) points higher than those who remained on limited plans, suggesting that differences in digital access across regions may affect the potential effectiveness of mobile interventions. For household income, differential improvement was observed in the low-income (DID change 1.54 points, 95% CI −0.05 to 3.60; *P*=.14) and high-income (DID change 7.67 points, 95% CI −3.03 to 18.38; *P*=.16) groups (suggesting that the change to unlimited plans had an association with social participation), but it was not significant. As shown in Table 4, the effects of several covariates were not as evident for older adult women as they were for older adult men. Among older adult women, the effects of unlimited data access were generally smaller and less consistent across subgroups. However, women living alone still showed a notable improvement of 3.38 (95% CI −0.55 to 7.30; *P*=.09) points, highlighting that living arrangement may be a more influential factor than gender alone.

Table . Gender-stratified subgroup analysis results by independent variables in male adults aged 60 years and older (n=1594)

Variables	Adults, n	Unadjusted means (social activity score[a])						Absolute difference-in-differences		
		Limited[b,c]→Limited			Limited[c]→Unlimited[d]			Adjusted (95% CI)[e]	SE	*P* value
		Before[f]	After[f]	Difference	Before[f]	After[f]	Difference			
Household composition										
Single-person household (n=94)	94	17.70	17.11	−0.59	17.70	20.06	2.36	6.44 (3.39 to 9.50)	1.55	<.001
Multiperson household (n=1500)	1500	20.00	20.09	0.09	20.00	22.02	2.02	0.61 (−0.91 to 2.13)	0.77	.43
Household income										
Low	1034	18.16	17.46	−0.69	18.16	20.44	2.28	1.54 (−0.50 to 3.60)	1.04	.14
Middle	466	21.28	21.46	0.19	21.28	22.07	0.80	−0.05 (−2.74 to 2.63)	1.37	.97
High	94	22.62	21.24	−1.38	22.62	22.43	−0.19	7.67 (−3.03 to 18.38)	5.46	.16
Region[g]										
Capital region	371	20.09	20.94	0.84	20.09	23.27	3.17	3.47 (0.15 to 6.78)	1.69	.04
Metropolitan	367	20.76	21.08	0.32	20.76	21.55	0.79	0.03 (−2.64 to 2.71)	1.37	.98
City/province	856	19.28	18.67	-0.60	19.28	20.97	1.70	−0.31 (−2.20 to 1.56)	0.95	.75

[a]Social activity score refers to the total combined score of 8 social participation, and the maximum score is 64 points.

[b]Limited: smartphone-limited data plan.

[c]The baseline excludes smartphone users with unlimited data plans.

[d]Unlimited: smartphone unlimited data plan.

[e]Adjusted for age, gender, education, marriage status, employment status, household income, region, Wi-Fi installation status, family members.

[f]"Before" and "After" indicate the mean social activity score in the pre- and post-periods, respectively.

[g]The capital region of South Korea, encompassing Seoul and its surrounding metropolitan areas, serves as the political, economic, and cultural hub of the country, while metropolitan cities represent major urban areas and their environments, functioning as significant administrative and economic centers. City or province divisions in South Korea act as regional governmental bodies responsible for local governance and administration.

XSL•FO
RenderX

Table . Gender-stratified subgroup analysis results by independent variables in female adults aged 60 years and older (n=2321)

Variables	Adults, n	Unadjusted means (social activity score[a])						Absolute difference in differences		
		Limited[b,c]→Limited			Limited[c]→Unlimited[d]			Adjusted (95% CI)[e]	SE	*P* value
		Before[f]	After[f]	Difference	Before[f]	After[f]	Difference			
Household composition										
Single-person household (n=567)	567	17.58	17.12	−0.47	17.58	21.25	3.67	3.38 (−0.55 to 7.30)	2.00	.09
Multiperson household (n=1754)	1754	19.95	20.27	0.31	19.95	22.24	2.29	−0.59 (−2.36 to 1.18)	0.91	.52
Household income										
Low	1671	18.16	17.71	−0.44	18.16	19.77	1.61	0.79 (−1.59 to 3.17)	1.21	.51
Middle	560	21.13	21.07	−0.06	21.13	21.95	0.83	−0.75 (−3.52 to 2.02)	1.41	.60
High	90	21.85	21.92	0.07	21.85	23.75	1.91	—[g]	—	—
Region[h]										
Capital region	542	19.99	21.14	1.15	19.99	23.61	3.62	−0.62 (−4.62 to 3.37)	2.04	.76
Metropolitan	517	20.75	20.33	−0.42	20.75	21.68	0.93	−0.03 (−2.82 to 2.77)	1.42	.99
City or province	1262	18.91	18.59	−0.32	18.91	20.87	1.97	0.70 (−1.47 to 2.86)	1.10	.53

[a]Social activity score refers to the total combined score of 8 social participation, and the maximum score is 64 points.

[b]Limited: smartphone-limited data plan

[c]The baseline excludes smartphone users with unlimited data plans.

[d]Unlimited: smartphone unlimited data plan.

[e]Adjusted for age, gender, education, marriage status, employment status, household income, region, Wi-Fi installation status, and family members.

[f]"Before" and "After" indicate the mean social activity score in the pre- and postperiods, respectively.

[g]Not applicable.

[h]The capital region of South Korea, encompassing Seoul and its surrounding metropolitan areas, serves as the political, economic, and cultural hub of the country, while metropolitan cities represent major urban areas and their environments, functioning as significant administrative and economic centers. City or province divisions in South Korea act as regional governmental bodies responsible for local governance and administration.

Discussion

Principal Findings

This study used a DID design to examine the potential association between unlimited smartphone data plans and changes in social participation of older adults. Notably, the findings were numerically larger among men and older adults living alone. These results suggest that unlimited data plans could be explored as a potential digital inclusion strategy for specific vulnerable subgroups.

Prior studies have primarily examined the relationship between improved digital literacy and increased social engagement among this population [16,17]. In contrast, our findings suggest that providing unrestricted smartphone access—regardless of digital literacy level—may be explored as a potentially meaningful intervention. This highlights the unique challenges faced by older adults who cannot freely use smartphones in the digital age. For instance, we found that living alone was significantly associated with lower social activity scores, indicating limited opportunities for social interaction compared with those living with others [35,36]. As a result, these individuals may depend more on digital tools for connection. The provision of unlimited data plans—by removing spatiotemporal restrictions on internet access—may reduce barriers to social participation for these individuals. While the small sample size limits the significance of the results, a notable differential change was observed in the incomes of those living

XSL•FO
RenderX

in metropolitan areas. A previous study identified community-based activities as central to older adults' social participation [30]. In rural or smaller communities, older adults often benefit from closer-knit social networks and community activities. In contrast, those living in metropolitan areas may face greater social isolation and thus rely more heavily on digital means for connection. This could explain why the provision of unlimited smartphone data showed a suggestive, numerically larger association with their social participation [37].

While no statistically significant overall effect was observed among individuals aged 60 years and older, subgroup analyses revealed a significant improvement in social activity scores for specific populations—most notably, older men living alone. These findings suggest that older adults' ability to participate in social participation increasingly depends on access to digital devices and mobile data. For those with limited financial resources, the cost of data plans or smartphones may pose significant barriers, reflecting the broader economic inequality in digital participation. Gender-stratified analysis revealed that older men living alone experienced a numerically larger improvement in social activity following the adoption of unlimited smartphone data plans. This finding is particularly meaningful given that older men are generally less socially active than women [38], suggesting that enhanced digital access may help narrow gender disparities in social engagement. The effect was even more evident among male residents of metropolitan areas, indicating that digital connectivity plays a critical role in enabling social participation for this subgroup. The unlimited smartphone data plan differs from Wi-Fi. It reduces spatiotemporal constraints on internet use and encourages older adults to engage in social participation. Compared with traditional interventions such as community-based programs or digital literacy education [16,17], providing unlimited smartphone data plans offers a lower-cost and scalable alternative to enhance social participation among older adults. By reducing structural barriers to digital access, this approach could help address digital inclusion more directly and efficiently.

Implications

These subgroup findings suggest that policymakers could consider subsidizing mobile data plans for vulnerable older populations, such as those living alone or in metropolitan areas. Conducted in South Korea—where smartphone penetration is high and unlimited data plans are relatively affordable—this study highlights how improved digital access can enhance social participation among older adults. The findings of this study may be relevant to other settings with high smartphone penetration, although their applicability will depend on local infrastructure, affordability, and broader socioeconomic context.

Limitations

This study has some limitations. First, due to data source limitations, we analyzed data only from 2016 and 2022. Second, our analysis included only the covariates calibrated in this study, potentially omitting other unmeasured covariates that could influence the relationship between unlimited data use and social activity scores. Third, we lack precise information on the timing of participants' transition to unlimited data plans. Therefore, we cannot pinpoint the exact time of participants' switch to unlimited data and its subsequent association with their social activity scores over the 6-year period. To address this, we assumed an average switch point sometime in the middle years. In addition, we conducted multiple subgroup analyses to explore heterogeneous effects across demographic and household characteristics. While these analyses provide valuable insight, they also introduce the possibility of spurious findings due to multiple comparisons. Thus, subgroup results should be interpreted with caution and seen as exploratory.

In addition, we did not formally assess whether the assumptions of linear regression—such as normality of residuals, homoscedasticity, and absence of multicollinearity—were met. This may limit the robustness of our estimates and should be addressed in future studies. Finally, the data only included 2 time points (2016 and 2022), which limited our ability to verify the parallel trends assumption—an essential prerequisite for the validity of DID analyses. Although we aimed to minimize baseline imbalance and carefully selected covariates, the inability to test parallel trends remains a critical methodological limitation. Future research using panel data with multiple time points would allow for formal testing of this assumption and a more robust causal inference.

Conclusions

This is the first study to use unlimited smartphone data plans to examine changes in social activity among older adults. By using longitudinal data with repeated measures of the same participants, we effectively captured changes in social activity over time. Therefore, our findings can inform policy discussions and contribute to the important public health goal of promoting social engagement among older adults.

This study unveils a significant association between older adults' adoption of unlimited smartphone plans and their social engagement, suggesting that switching to an unlimited plan may be associated with increased social participation. Ensuring older adults have unrestricted internet access to electronic devices is crucial before initiating digital literacy initiatives. These findings highlight the potential of improving data accessibility as a supportive lever for promoting social participation in older adults, offering a comparatively simple and scalable approach alongside existing policy measures.

Acknowledgments

The authors express sincere appreciation to the Korea Information Society Development Institute for providing access to the Korea Media Panel data.

XSL•FO
RenderX

Funding

This research received no external funding and was conducted using institutional resources available to the authors.

Data Availability

The data used in this study are publicly available through the Korea Media Panel [39]. The dataset is deidentified and accessible to researchers upon request to the Korea Information Society Development Institute.

Authors' Contributions

Conceptualization: SC, SYJ
Data curation: SC
Formal analysis: SC, SYJ
Investigation: SC, EC
Methodology: SC, EC, SYJ
Supervision: SYJ
Validation: SC, EC, SYJ
Writing – original draft: SC
Writing – review & editing: SC, EC, SYJ

Conflicts of Interest

None declared.

References

1. Raleigh VS. Trends in life expectancy in EU and other OECD countries: why are improvements slowing? : Organization for Economic Co-Operation and Development; 2019 URL: https://www.oecd.org/content/dam/oecd/en/publications/reports/2019/02/trends-in-life-expectancy-in-eu-and-other-oecd-countries_93f16d1f/223159ab-en.pdf [accessed 2026-01-16]
2. Social expenditure database. Organization for Economic Co-Operation and Development. URL: https://www.oecd.org/en/data/datasets/social-expenditure-database-socx.html?utm_source=chatgpt.com [accessed 2024-04-13]
3. OECD health statistics. Organization for Economic Co-operation and Development. 2023. URL: https://www.oecd.org/en/data/datasets/oecd-health-statistics.html?utm_source=chatgpt.com [accessed 2026-01-16]
4. Pensions at a glance 2023: OECD and G20 indicators. : Organization for Economic Co-operation and Development; 2023 URL: https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/12/pensions-at-a-glance-2023_4757bf20/678055dd-en.pdf [accessed 2024-04-13]
5. Active ageing: a policy framework. : World Health Organization; 2002 URL: https://extranet.who.int/agefriendlyworld/wp-content/uploads/2014/06/WHO-Active-Ageing-Framework.pdf [accessed 2026-01-16]
6. Berkman LF, Kawachi I, Glymour MM, editors. Social Epidemiology, 2nd edition: Oxford University Press; 2014. [doi: 10.1093/med/9780195377903.001.0001]
7. Edlin G, Golanty E. Health & Wellness: Jones & Bartlett Publishers; 2012.
8. Berkman LF, Glass T, Brissette I, Seeman TE. From social integration to health: Durkheim in the new millennium. Soc Sci Med 2000 Sep;51(6):843-857. [doi: 10.1016/s0277-9536(00)00065-4] [Medline: 10972429]
9. Blazer DG. Social support and mortality in an elderly community population. Am J Epidemiol 1982 May;115(5):684-694. [doi: 10.1093/oxfordjournals.aje.a113351] [Medline: 7081200]
10. Levasseur M, Desrosiers J, Noreau L. Is social participation associated with quality of life of older adults with physical disabilities? Disabil Rehabil 2004 Oct 21;26(20):1206-1213. [doi: 10.1080/09638280412331270371] [Medline: 15371021]
11. Sayin Kasar K, Karaman E. Life in lockdown: social isolation, loneliness and quality of life in the elderly during the COVID-19 pandemic: a scoping review. Geriatr Nurs 2021;42(5):1222-1229. [doi: 10.1016/j.gerinurse.2021.03.010] [Medline: 33824008]
12. Tomioka K, Kurumatani N, Hosoi H. Age and gender differences in the association between social participation and instrumental activities of daily living among community-dwelling elderly. BMC Geriatr 2017 Apr 28;17(1):99. [doi: 10.1186/s12877-017-0491-7] [Medline: 28454521]
13. Liu AQ, Besser T. Social capital and participation in community improvement activities by elderly residents in small towns and rural communities. Rural Sociol 2003 Sep;68(3):343-365. [doi: 10.1111/j.1549-0831.2003.tb00141.x]
14. Rabelo DF, Neri AL. The household arrangements, physical and psychological health of the elderly and their satisfaction with family relationships. Rev Bras Geriatr Gerontol 2015;18(3):507-519. [doi: 10.1590/1809-9823.2015.14120]
15. Bernard M, Phillips J. The challenge of ageing in tomorrow's Britain. Ageing Soc 2000 Jan;20(1):33-54. [doi: 10.1017/S0144686X99007576]
16. Lelkes O. Happier and less isolated: internet use in old age. J Poverty Soc Justice 2013;21(1):33-46. [doi: 10.1332/175982713X664047]

XSL•FO
RenderX

17. Bobillier Chaumon ME, Michel C, Tarpin Bernard F, Croisile B. Can ICT improve the quality of life of elderly adults living in residential home care units? From actual impacts to hidden artefacts. Behav Inform Technol 2014 Jun 3;33(6):574-590. [doi: 10.1080/0144929X.2013.832382]
18. Wu YH, Damnée S, Kerhervé H, Ware C, Rigaud AS. Bridging the digital divide in older adults: a study from an initiative to inform older adults about new technologies. Clin Interv Aging 2015;10:193-200. [doi: 10.2147/CIA.S72399] [Medline: 25624752]
19. Castilla D, Botella C, Miralles I, et al. Teaching digital literacy skills to the elderly using a social network with linear navigation: a case study in a rural area. Int J Hum Comput Stud 2018 Oct;118:24-37. [doi: 10.1016/j.ijhcs.2018.05.009]
20. Oh SS, Kim KA, Kim M, Oh J, Chu SH, Choi J. Measurement of digital literacy among older adults: systematic review. J Med Internet Res 2021 Feb 3;23(2):e26145. [doi: 10.2196/26145] [Medline: 33533727]
21. Kawaguchi K, Nakagomi A, Ide K, Kondo K. Effects of a mobile app to promote social participation on older adults: randomized controlled trial. J Med Internet Res 2024 Sep 30;26:e64196. [doi: 10.2196/64196] [Medline: 39348180]
22. Althoff T, Jindal P, Leskovec J. Online actions with offline impact: how online social networks influence online and offline user behavior. Presented at: WSDM '17: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining; Feb 6-10, 2017. [doi: 10.1145/3018661.3018672]
23. Cao L, Chongsuvivatwong V, McNeil EB. The sociodemographic digital divide in mobile health app use among clients at outpatient departments in Inner Mongolia, China: cross-sectional survey study. JMIR Hum Factors 2022 May 19;9(2):e36962. [doi: 10.2196/36962] [Medline: 35587367]
24. Lin T, Guo W, Li Y, Guo X, Bai X, Min R. Geographical accessibility of medical resources, health status, and demand of integrated care for older people: a cross-sectional survey from Western China. BMC Geriatr 2024 May 20;24(1):440. [doi: 10.1186/s12877-024-04987-2] [Medline: 38769502]
25. Shaw LH, Gant LM. In defense of the internet: the relationship between internet communication and depression, loneliness, self-esteem, and perceived social support. Cyberpsychol Behav 2002 Apr;5(2):157-171. [doi: 10.1089/109493102753770552] [Medline: 12025883]
26. Morahan-Martin J, Schumacher P. Loneliness and social uses of the internet. Comput Human Behav 2003 Nov;19(6):659-671. [doi: 10.1016/S0747-5632(03)00040-2]
27. Parker JDA, Taylor RN, Eastabrook JM, Schell SL, Wood LM. Problem gambling in adolescence: relationships with internet misuse, gaming abuse and emotional intelligence. Pers Individ Dif 2008 Jul;45(2):174-180. [doi: 10.1016/j.paid.2008.03.018]
28. Ballarotto G, Volpi B, Marzilli E, Tambelli R. Adolescent internet abuse: a study on the role of attachment to parents and peers in a large community sample. Biomed Res Int 2018;2018:5769250. [doi: 10.1155/2018/5769250] [Medline: 29707572]
29. Gencer SL, Koc M. Internet abuse among teenagers and its relations to internet usage patterns and demographics. J Educ Technol Soc 2012;15(2):25-36 [FREE Full text]
30. Greydanus DE, Greydanus MM. Internet use, misuse, and addiction in adolescents: current issues and challenges. Int J Adolesc Med Health 2012;24(4):283-289. [doi: 10.1515/ijamh.2012.041] [Medline: 23183727]
31. Unlimited (LTE) plan [Web page in Korean]. Korea Telecom (KT) Corporation. URL: https://product.kt.com/wDic/productDetail.do?ItemCode=909 [accessed 2024-06-12]
32. Korea Media Panel user guide, edition 2023 [Web page in Korean]. Korea Information Society Development Institute. 2023. URL: https://stat.kisdi.re.kr/kor/contents/ContentsList.html?subject=USER&sub_div=U [accessed 2024-03-05]
33. Finlay JM, Kobayashi LC. Social isolation and loneliness in later life: a parallel convergent mixed-methods case study of older adults and their residential contexts in the Minneapolis metropolitan area, USA. Soc Sci Med 2018 Jul;208:25-33. [doi: 10.1016/j.socscimed.2018.05.010] [Medline: 29758475]
34. 2020 Korea Media Panel Survey [Report in Korean]. : Korea Information Society Development Institute; 2021 URL: https://www.kisdi.re.kr/report/view.do?arrMasterId=3934581&artId=558016&key=m2101113024973&masterId=3934581 [accessed 2026-01-16]
35. Dehi Aroogh M, Mohammadi Shahboulaghi F. Social participation of older adults: a concept analysis. Int J Community Based Nurs Midwifery 2020 Jan;8(1):55-72. [doi: 10.30476/IJCBNM.2019.82222.1055] [Medline: 32039280]
36. Townsend BG, Chen JTH, Wuthrich VM. Barriers and facilitators to social participation in older adults: a systematic literature review. Clin Gerontol 2021;44(4):359-380. [doi: 10.1080/07317115.2020.1863890] [Medline: 33393443]
37. Levasseur M, Cohen AA, Dubois MF, et al. Environmental factors associated with social participation of older adults living in metropolitan, urban, and rural areas: the NuAge study. Am J Public Health 2015 Aug;105(8):1718-1725. [doi: 10.2105/AJPH.2014.302415] [Medline: 25689194]
38. Katagiri K, Kim JH. Factors determining the social participation of older adults: a comparison between Japan and Korea using EASS 2012. PLoS ONE 2018;13(4):e0194703. [doi: 10.1371/journal.pone.0194703] [Medline: 29624588]
39. Media statistics portal [Web page in Korean]. Korea Information Society Development Institute. URL: https://stat.kisdi.re.kr [accessed 2026-01-05]

Abbreviations

DID: difference-in-differences

XSL•FO
RenderX

mHealth: mobile health
OECD: Organization for Economic Cooperation and Development
SAS: Statistical Analysis System
WHO: World Health Organization

Edited by L Buis; submitted 19.Nov.2024; peer-reviewed by A Muftah, I Iyamu, P Joshi; revised version received 14.Nov.2025; accepted 22.Dec.2025; published 23.Feb.2026.

Please cite as:
Choi S, Choi E, Jang SY
Unlimited Smartphone Data Plans in Older Adults With Data Deprivation: Quasi-Experimental Study
JMIR Mhealth Uhealth 2026;14:e68930
URL: https://mhealth.jmir.org/2026/1/e68930
doi:10.2196/68930

© Seoyeong Choi, Eunjeong Choi, Suk-Yong Jang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 23.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Impact of Different Onboarding Strategies on Low Adoption and Engagement With a Self-Monitoring and Management App for Chronic Musculoskeletal Pain: Prospective Study

Cinja Nadana Koller[1], BA, MSc; Marc Blanchard[1], BA, MSc, DrPH; Johanna Mettler[1], Dr med; Tiffany Prétat[1], Dr med; Pedro Ming Azevedo[1], Dr med; Thomas Hügle[1], Prof Dr Med

Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

Corresponding Author:
Cinja Nadana Koller, BA, MSc
Lausanne University Hospital and University of Lausanne
Avenue Pierre-Decker 4
Lausanne, 1011
Switzerland
Phone: 41 795564438
Email: cinja.koller@unil.ch

Abstract

Background: Suboptimal adoption and engagement rates of digital health applications present challenges to their effectiveness, particularly in chronic disease management such as fibromyalgia. Up to half of patients do not download the prescribed digital health applications or actively engage with them, making effective onboarding a critical opportunity for improvement.

Objective: We aimed to investigate the impact of 3 different patient onboarding strategies on adoption, adherence, and engagement with a digital health application for the symptom management of chronic pain syndromes.

Methods: We conducted a 4-week nonrandomized prospective study comprising patients (aged ≥18 years) with chronic musculoskeletal pain who fulfilled the fibromyalgia criteria and were using a new self-monitoring and management application after onboarding. The pain organizer and companion system consists of symptom reporting, symptom monitoring, and an advice and exercise section. Participants in group 1 (standard) received an email-based onboarding with detailed instructions on application download, registration, use, usability data capture, data security, and study procedures without any in-person assistance. Group 2 (video) received the same email plus a video tutorial explaining the application's use also without any in-person assistance. Group 3 (assisted) underwent an enhanced in-person onboarding, which included hands-on support for application setup and guided instructions from a health care professional. Primary outcomes included adoption, engagement, and adherence, measured respectively by successful application download with at least 1 log-in, cumulative activations, and connection at least once per week over 4 weeks.

Results: A total of 48 patients were recruited (mean age 45.7, SD 12.2; range 19-79 years). In the assisted group, all 15 participants in the assisted onboarding group downloaded the application compared to 63% (10/16) in the standard group and 77% (13/17) in the video group. A significant difference in adoption between the in-person onboarding group (group 3) and the remote groups (groups 1 and 2) with P=.009 was observed. Overall engagement, measured by cumulative app activation, ranged between 0 and 28 log-ins, with no significant difference (P=.18) in mean log-ins across the 3 groups. Overall adherence was low with 27% (13/48) of participants meeting adherence criteria and a retention rate of 46% (22/48) at week 4. Responses to the System Usability Scale survey were limited to 8 responses but achieved a score of 70.31, indicating good usability of the application.

Conclusions: Unassisted downloading and account creation might be barriers to adopting a digital health app among patients with chronic musculoskeletal pain.

(JMIR Mhealth Uhealth 2026;14:e78827) doi:10.2196/78827

KEYWORDS

onboarding strategies; adoption; mHealth; adherence; engagement; chronic musculoskeletal pain

XSL•FO
RenderX

Introduction

Digital health solutions, including mobile applications and digital therapeutics, have emerged as promising tools in modern health care delivery. A growing body of evidence suggests that digital health solutions, when effectively adopted and integrated into patient care, can result in positive health outcomes. These technologies offer the potential to enhance patient engagement, improve treatment adherence, and empower individuals to actively manage their health, particularly in the context of chronic conditions [1-3]. However, the active use of digital interventions influences their effectiveness. It was found that increased user engagement with digital interventions is linked to greater improvements in outcomes [4]. However, despite the increasing availability and potential benefits of digital health solutions, achieving high levels of adoption and engagement remains a significant challenge [5-7]. Studies consistently report that only about half of prescribed digital health solutions are downloaded by patients, leading to low use rates and missed opportunities for improved health management [8,9]. This discrepancy between the potential benefits of digital health solutions and their limited adoption and engagement rates raises critical questions about the factors influencing adoption and the potential strategies for enhancing adoption [10]. Implementation science examines whether an intervention is delivered as intended and investigates strategies to enhance its integration into real-world settings, using outcomes such as adoption, engagement, and retention to assess implementation success [11].

Increased adoption rates not only facilitate the effective use of these technologies but also hold the potential to positively impact patient outcomes, reduce health care costs, and improve overall health care delivery [3,12]. Among the various factors affecting adoption and engagement, the onboarding process plays a pivotal role [6]. Onboarding refers to the initial introduction of a patient to a digital health solution including, for example, downloading, registering, and device pairing [13,14]. A major reason for the low adoption of mobile health (mHealth) applications is concerns about data security and privacy [15,16]. Another barrier to adoption is the lack of support from physicians [17]. However, according to Dahlhausen et al [18], health care professionals (HCPs) hold significant potential to enhance adherence to digital therapeutics. Moreover, Hernandez-Ramos et al [14] found low self-reported digital health literacy among study participants reporting difficulties with new tools such as mHealth apps. They addressed this issue by providing onboarding support (eg, downloading and installing the application) to digital health platforms.

Therefore, onboarding serves as a critical opportunity to engage and motivate patients, establish trust, and familiarize them with the functionalities and benefits of the technology. A well-designed onboarding process can potentially address patient concerns, mitigate barriers to adoption, and optimize the user experience, ultimately increasing the likelihood of sustained engagement and use of the digital health solution. Given the importance of adoption and the critical role of onboarding strategies, it is essential to explore how different onboarding approaches influence adoption rates, user engagement, and retention of digital health solutions. Despite the importance of onboarding, many studies focus only on engagement strategies or barriers to adoption after the initiation of the digital intervention [10,15,16,19-24]. In this study, we aim to address this gap by investigating the effect of patient onboarding strategies, used as an implementation strategy, on the adoption, engagement, and retention of a digital health solution for chronic conditions, specifically for the management of chronic musculoskeletal pain. By evaluating the impact of tailored onboarding approaches compared with a standard onboarding process, we seek to identify effective strategies that can increase adoption, enhance retention, and optimize the overall benefits of digital health interventions.

Methods

Study Participants

The study included individuals who had been diagnosed with chronic musculoskeletal pain resistant to conventional treatment. Participants had to be aged ≥18 years and have the capability to use a mobile app–based intervention. Additionally, participants needed to possess an email account and a smartphone capable of downloading and using the study-specific self-monitoring and self-management application. Individuals with insufficient proficiency in French to provide informed consent or complete study procedures were excluded. Participants were recruited from the University Hospital Lausanne after they underwent a 2-week in-house multimodal pain program due to being refractory to outpatient treatment. These patients typically present with severe pain as well as high levels of anxiety, depression, catastrophizing, and cognitive impairments, including deficits in attention and memory.

Ethical Considerations

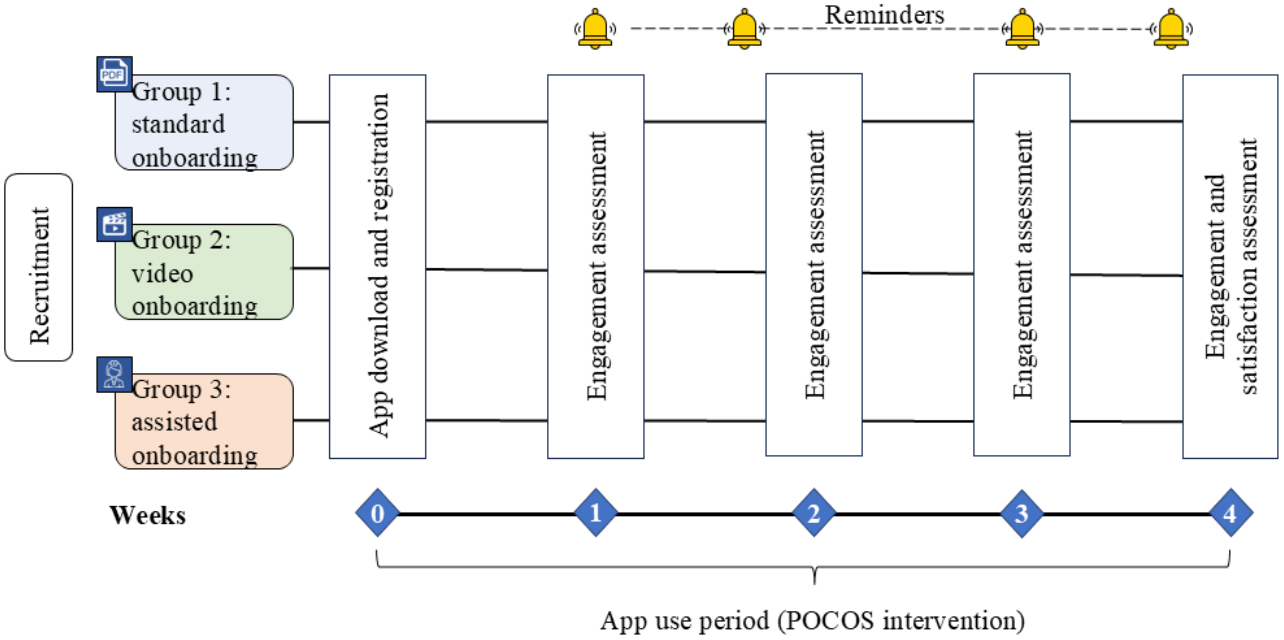
Ethics approval was obtained from the Commission Cantonale d'éthique de la Recherche sur l'être Humain (CER-VD 2021-01680). Informed written consent was given by all participants prior to their inclusion in the study. The study was conducted ethically in accordance with the World Medical Association Declaration of Helsinki. Data were deidentified to protect participant privacy and maintain confidentiality. Participants did not receive any compensation for taking part in this study.

Study Design

This comparative, nonrandomized 4-week study examined the effect of different onboarding strategies on the adoption of an mHealth application and retention of use in patients with chronic musculoskeletal pain (Figure 1). Eligible participants were enrolled sequentially into the following three groups (each representing a different onboarding strategy):

XSL•FO
RenderX

Figure 1. Study design. Patient-reported outcomes were collected through the pain organizer and companion system (POCOS) app. Engagement was assessed through usability data, and satisfaction was assessed using the System Usability Scale. Reminders to use the POCOS app were delivered via SMS text messages.



1. Group 1 (standard onboarding group): participants in this group received a standard onboarding process that was delivered entirely remotely via email. This onboarding included step-by-step instructions for downloading and registering an account, information on application use and usability data collection, details regarding data security, an overview of study procedures related to the self-monitoring and self-management application, and a concise summary of the application's core features and functionalities. The manual can be accessed in Multimedia Appendix 1 (in French).
2. Group 2 (video onboarding group): participants in this group underwent the same onboarding process as group 1 but were also provided with an onboarding video. This video, of 3 minutes duration, walked the participants through a comprehensive demonstration of how to use various features and functionalities of the application. The video can be accessed [25].
3. Group 3 (assisted onboarding group): participants in this group received onboarding assistance from a dedicated study HCP via a 30-minute in-person session at the University Hospital Lausanne. The HCP guided each participant individually through the onboarding process and assisted in downloading, registering, logging in, and entering initial data into the application. Furthermore, the HCP provided personalized support and explanations, addressing any questions or concerns raised by the participants during the onboarding session. To minimize bias, the HCP followed a script.

Participants in all 3 groups had access to the same user manual to guide them through application use. The onboarding procedures in groups 1 and 2 were fully remote, while group 3 was in person. Blinding of HCPs and participants was not feasible due to the nature of the intervention.

All participants had the possibility to reach out for support via email or phone for assistance with any step during the study time.

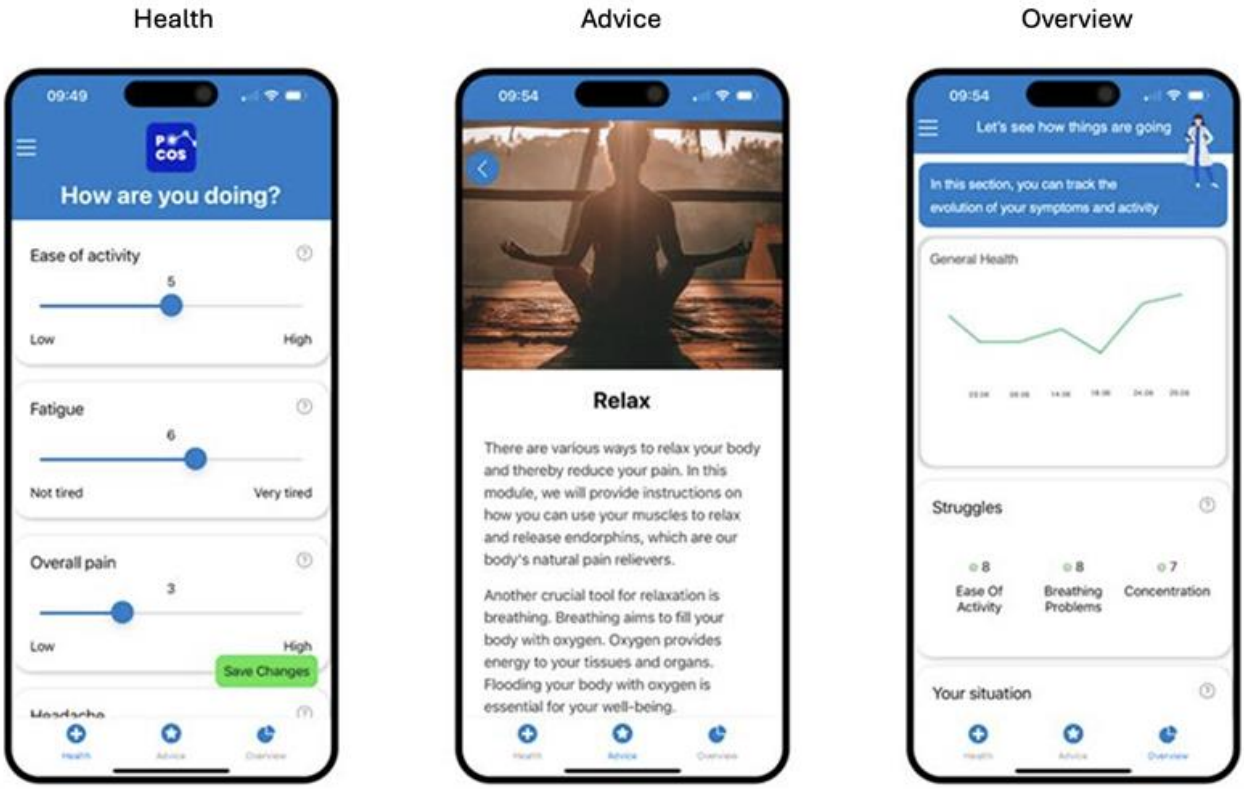
Adoption and engagement are well-documented challenges in mHealth applications, with several studies identifying key barriers and influencing factors. The literature emphasizes the importance of reminders, user-friendly and robust technical design, data privacy and security, tailored content, and personal support [19,20,26]. To address these factors and enhance adherence and engagement, we incorporated them into the study design alongside the specific onboarding strategies. All participants received weekly SMS text message reminders to engage with the pain organizer and companion system (POCOS) application and to provide usability data. To mitigate concerns around data privacy and security, information on secure data handling was explicitly addressed during onboarding. The POCOS application was co-designed with patients and HCPs and features a user-friendly interface with reliable technical performance [27].

POCOS User Interface

POCOS is a self-monitoring and management application, which was tested in a previous study [27,28]. This application consists of 3 components: symptom reporting, symptom monitoring, and an advice and exercise section (Figure 2). Moreover, the content can be tailored by patients, as they can freely navigate through the application. Following the onboarding phase, participants in all 3 groups were asked to use POCOS 2 to 3 times a week for 1 month. During this period, patient-reported outcomes related to chronic pain (fatigue, concentration, sleep quality, memory, pain, ease of activity, and anxiety) were asked to be reported by patients via the application at baseline, after 2 weeks, and after 4 weeks of using the application. At the end of the study, participants received a link via text message to fill out a System Usability Scale (SUS) survey.

XSL•FO
RenderX

Figure 2. Pain organizer and companion system user interface. Navigation is provided via the bottom menu with 3 main sections: Health, Advice, and Overview.



Outcomes Measures

Primary Outcomes

The primary outcomes of this study were adoption, adherence, and engagement with the application. Adoption was defined as the uptake or intention to try or use a new intervention [29]. The primary measure was the adoption rate of the self-monitoring and self-management application among participants. Adoption rates were assessed by calculating the percentage of participants who successfully downloaded the application and logged in at least once.

Adherence was measured as the percentage of participants who continued using the application after the initial download and remained engaged with at least 1 connection per week throughout the 1-month period.

Engagement was assessed over the 4-week period using the number of log-ins and time spent with the application, two commonly used application engagement metrics [30].

Secondary Outcome

The secondary outcome of this study was patient satisfaction, assessed at the end of the study using the SUS survey. In the survey, participants were asked how they liked the application and how satisfied they were with the onboarding support that they had received.

Data Collection and Management

Data collection for this study was fully decentralized and involved a combination of participant-reported data and application-collected use data. Participants were provided with secure log-in credentials to access the self-monitoring and self-management application. All data were coded (deidentified) and stored in a secure, password-protected database. The application use and engagement metrics, such as account registration and frequency of application use, were recorded. These engagement measures were obtained passively via the database or actively through screenshots of application use that had to be sent biweekly by participants to the research team. At the end of the study, participants received a link for a follow-up SUS survey, and all responses were fully anonymized.

Statistical Analysis

Descriptive statistics were calculated for all variables, including means and SDs for continuous variables and frequencies and percentages for categorical variables. Differences in continuous outcomes between the 3 onboarding groups were assessed using the Kruskal-Wallis *H* test, as the data did not meet the assumption of normality. Group differences and associations between categorical variables (adoption and adherence) were assessed using the chi-square test or Fisher exact test (if <5 observations per category). Generalized linear modeling (linear or logistic, as appropriate) was used to examine the associations between user characteristics and engagement. To analyze survival rate and time among the different groups, we used Kaplan-Meier survival functions with a log-rank (Mantel-Cox) test. Two-sided *P* values <.05 were considered statistically significant, and 95% CIs were used. All analyses were performed with SPSS software (version 29.0.2.0; IBM Corp). We analyzed with an intention-to-treat approach.

XSL•FO
RenderX

Results

Overview

A total of 48 patients with chronic musculoskeletal pain were recruited from 2023 to 2025 and sequentially distributed into 3 groups: 16 (33%; standard), 17 (35%; video), and 15 (31%; assisted) participants per group. Mean age was 45.7 (SD 12.2) years, and 40 (83%) participants were female. Overall, 39 (81%) participants had depression, 12 (25%) had anxiety, and 18 (38%) were opioid users. All participants fulfilled the fibromyalgia criteria. No clinically meaningful differences in baseline characteristics were observed. Detailed clinical information per group is shown in Table 1.

Table 1. Clinical and demographic characteristics of the study participants (N=48).

Characteristics	Standard onboarding[a]	Video onboarding	Assisted onboarding
Comorbidities, n (%)			
Fibromyalgia criteria fulfilled	16 (100)	17 (100)	15 (100)
Depression			
Yes	13 (81)	14 (82)	12 (80)
No	2 (13)	3 (18)	3 (20)
Anxiety			
Yes	2 (12)	6 (35)	4 (27)
No	13 (81)	11 (65)	11 (73)
Opioid use			
Yes	5 (31)	6 (35)	7 (47)
No	11 (69)	11 (65)	8 (53)
Sex, n (%)			
Female	16 (100)	13 (76)	11 (73)
Male	0 (0)	4 (24)	4 (27)
Age (y), mean (SD)	46.3 (14)	45.9 (12)	44.9 (10)

[a]n=1 is missing for depression and anxiety in the standard onboarding group.

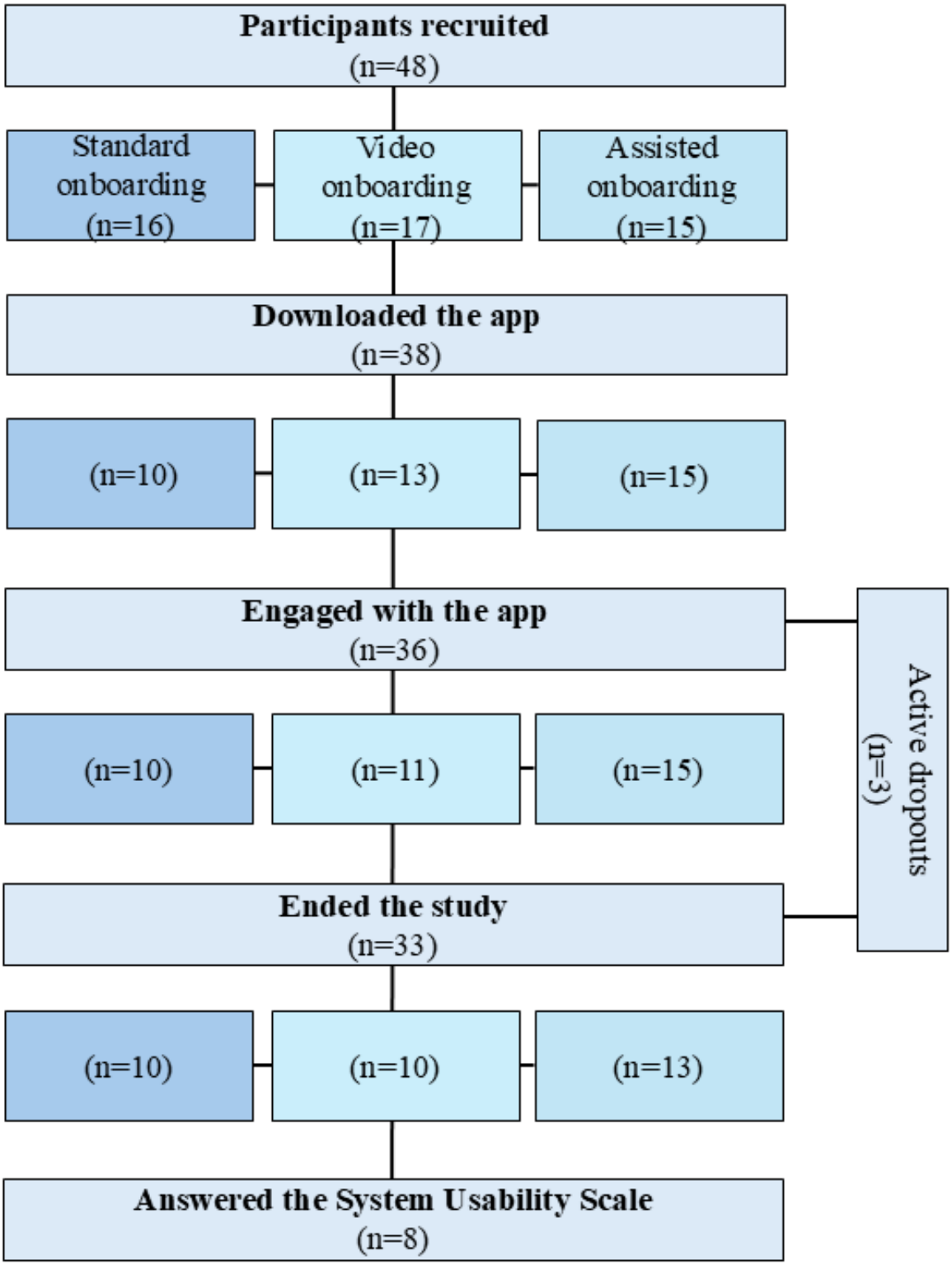
Adoption: Downloads and First Engagement

Overall, 81% (39/48) of participants downloaded the application after they received the instructions. All 15 participants in the assisted onboarding group downloaded and registered the application, while 63% (10/16) of participants registered an account in the standard onboarding without assistance. In the video group, 76% (13/17) of participants downloaded and registered an account. In total, 24% (4/17) of participants in the video group contacted the research team for assistance as they struggled with completing the onboarding independently. Support was given via phone or email by the research team. Of these 15 assisted participants, 1 (7%) had additional help from a relative to complete the downloading. Of them, 3 (20%) participants successfully managed to create their accounts, while 1 (7%) was lost in follow-up. Overall, 75% (36/48) of participants adopted the application, meaning that they successfully downloaded and tested the application at least once. About 69% (33/48) completed the study, meaning that they did not actively drop out of the study. Figure 3 shows the flow of participants in the different groups.

In a combined analysis of remote onboarding (standard and video onboarding) vs assisted in-person onboarding, adoption was significantly higher in the in-person onboarding group (*df*=1, Cramer *V*=0.389; *P*=.009; Figure 4).

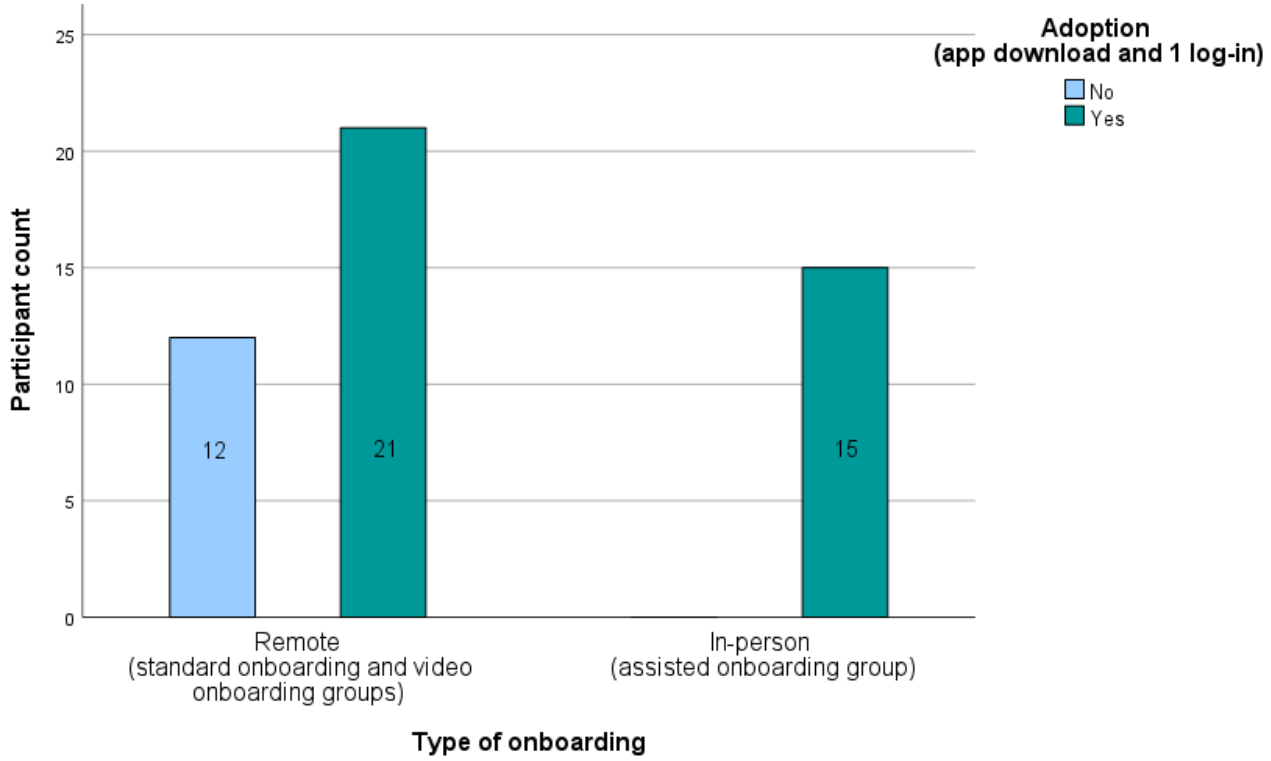
XSL•FO
RenderX

Figure 3. Participant flow per onboarding type.



XSL•FO
RenderX

Figure 4. Adoption of the application in the remote onboarding vs the in-person onboarding group.



Dropouts

Two participants in the assisted onboarding actively terminated the study prematurely after 14 days and 16 days. Reasons for dropout were a broken phone and poor health status. Another dropout occurred in the video group after 2 weeks. The participant mentioned feeling that they are not useful in contributing to the study. Baseline characteristics were comparable between participants who actively dropped out and those retained.

Time-to-Account Creation

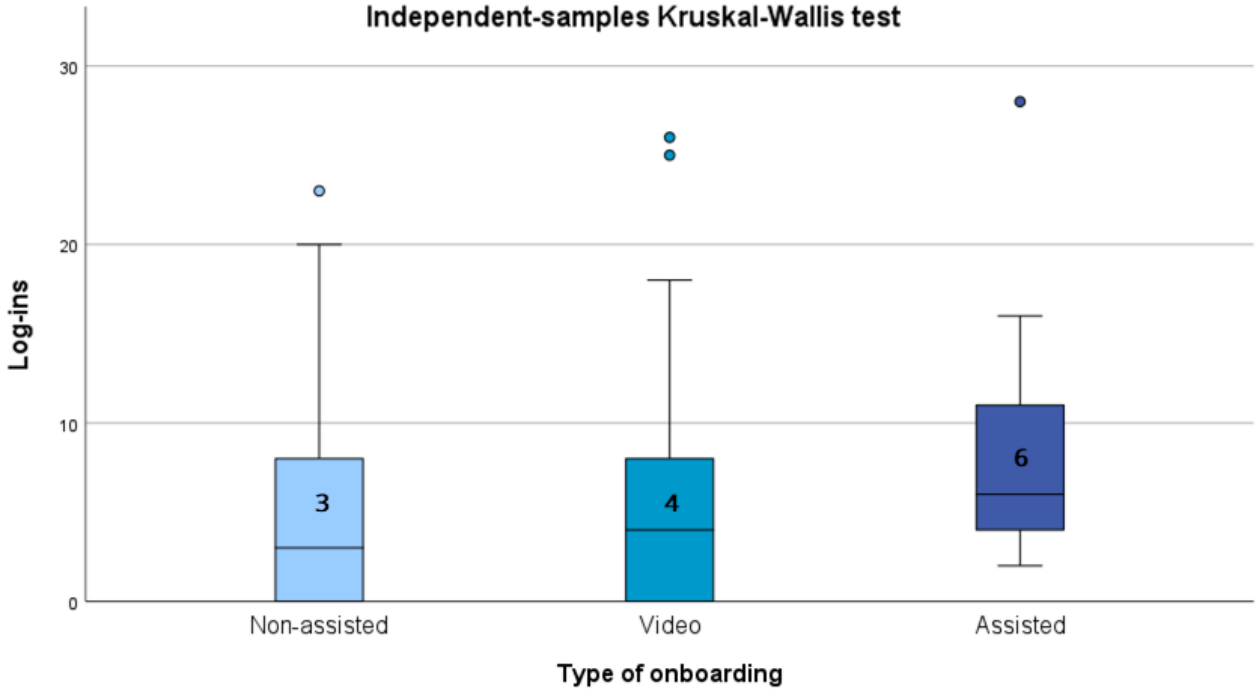
The time between account creation and the onboarding date was calculated for all participants who downloaded the application in the standard and video onboarding groups. The mean time to download was 5.61 (SD 6) days, with a range of 0 to 23 days. In total, 36% (12/33) of participants downloaded the application before the first reminder (at the end of the first week) and 64% (21/33) after the first reminder. There was no significant difference observed between the standard and the video group (P=.63).

Engagement: Minutes, Log-Ins, and Survival

In total, 96 screenshots of usability data were expected to be provided by participants. About 25% (12/48) of participants submitted usability data; however, these data were mostly incomplete (eg, only providing the minutes spent on the application for a single day). Out of the usability data obtained in minutes, participants used the application for 14 to 68 minutes during a week. Cumulative log-ins were passively collected. The number of log-ins for the duration of the study ranged from 0 to 28, with an overall mean of 7 (SD 7.68). A Kruskal-Wallis test showed no significant differences between the 3 groups (Kruskal-Wallis test df=2, median 3, IQR 8 vs median 4, IQR 11vs median 6, IQR 8; P=.18), as displayed in Figure 5. As no group differences were observed for onboarding type, we additionally ran a linear regression including the available covariates (sex, age, depression, anxiety, and opioid use) to examine their influence on engagement (number of log-ins during the 1-month period). Anxiety was a significant predictor of log-ins (β=−5.148, 95% CI −10.2 to −0.09; P=.046).

XSL•FO
RenderX

Figure 5. Median number of log-ins per onboarding group.



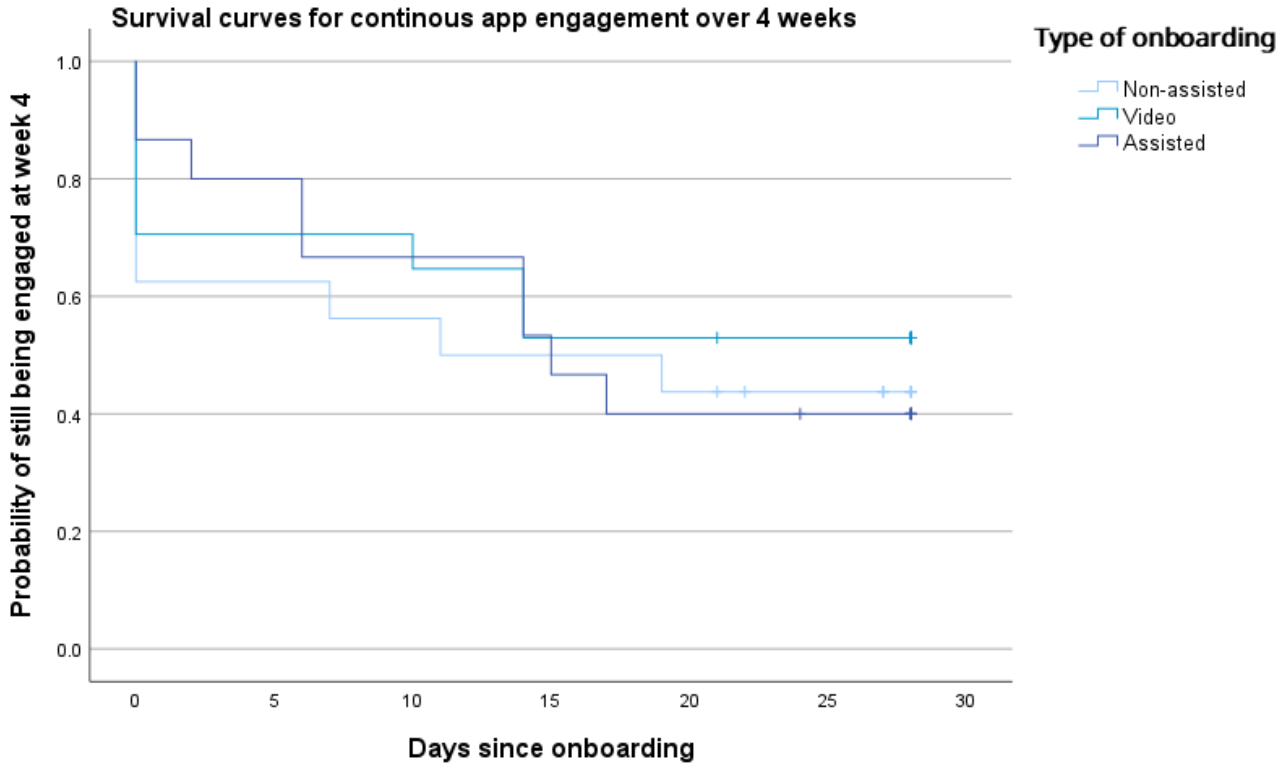
For the Kaplan-Meier survival analysis, we examined how many participants were still using the application during the final week of the 4-week study period. Time-to-event was defined as the number of days between onboarding and the participant's last log-in. An event was considered to have occurred if a participant did not use the application during the final study week, meaning that their last log-in occurred before week 4. A log-rank (Mantel-Cox) test revealed no significant difference in survival distributions between the onboarding groups (χ^2_2=0.4; *P*=.82; Table 2). The overall survival rate was 46% (22/48), with a mean survival time of 15.9 days. Among the groups, the video-onboarded participants showed the highest probability of continued application use, with a survival rate of 53% (9/17) and a mean survival time of 17.1 (SE 3.0) days (Figure 6).

Table 2. Survival data stratified by onboarding type.

Type of onboarding	Participants, n (%)	Events, n (%)	Survival rate (%)	Survival duration (days), mean (SE)
Standard	16 (33)	9 (56)	44	14.6 (3.2)
Video	17 (35)	8 (47)	53	17.1 (3.0)
Assisted	15 (31)	9 (60)	40	16.1 (2.8)

XSL•FO
RenderX

Figure 6. Survival curves for the probability of still engaging with the application in the last week of the study per onboarding type.



Adherence

Out of the 48 participants, only 13 (27%) were adherent with at least 1 connection per week for 4 weeks. Adherence was not influenced by onboarding group membership.

Participant Satisfaction: SUS

About 17% (8/48) of participants answered the SUS survey, leading to an overall score of 70.31. This score suggests a "good" usability level, as SUS scores typically range as follows: >78.8 (excellent usability), 70-85= (good usability), 50-69= (acceptable usability), and <51.6 =(poor usability) [31]. Only 50% (3/6) of participants provided feedback on their satisfaction with the onboarding type they received (standard and assisted groups); they were all satisfied and mentioned that the instructions that they received were clear and sufficient. Suggestions to improve the application by participants included fixing technical bugs, adding other language options, expanding content, integrating goal-setting features, enabling chat options with peers and HCPs, improving the design, providing links to scientific evidence and statistics about the disease, adding subtitles to videos, including a place to take notes, incorporating exercises or advice from ergotherapists, and implementing nudging strategies.

Discussion

Principal Findings

In this study, we found that in-person assisted onboarding significantly increased the adoption of a newly developed self-management application among patients with chronic musculoskeletal pain. While 81% (39/48) of participants downloaded the application and 75% (36/48) used it at least once (ie, adopted it), sustained use declined notably after approximately 2 weeks, with only 46% (22/48) still using the application by the fourth week. Mean and median log-in frequencies were highest in the assisted onboarding group and progressively lower with decreasing levels of support. However, onboarding type did not significantly influence overall engagement, as measured by median log-ins throughout the study period, nor adherence to regular use. The only significant predictive factor for app engagement we found was anxiety. Usability and satisfaction with onboarding were also assessed; however, due to the limited number of participants who provided detailed usability metrics in minutes, no statistical comparisons between onboarding groups could be performed for these outcomes.

Assisted in-person onboarding was effective for the adoption of the POCOS application and facilitated the timely initiation of application use, in contrast to the other two groups where account creation generally occurred a week later, aligning with the first reminder they received. Studies using health applications should consider implementing assisted onboarding to guarantee a seamless start to the study. On average, participants logged into the application 7 times during the study, and the time spent using the app ranged from 14 to 68 minutes per week. Our results regarding adoption and engagement are comparable to those reported in other mHealth studies in populations with chronic pain [32,33]. For example, in one study evaluating a chronic pain management app, 45% of participants remained engaged after 1 month [33], while in another study, 39% of participants completed the 3-month self-management program of a chronic pain app [32]. In addition, mHealth studies primarily focusing on depression and anxiety [7-9,34] may also be considered, as these conditions

XSL•FO
RenderX

are highly prevalent among patients with chronic pain. Compared to our study, fewer participants downloaded the app in these studies, which may be explained by the fact that one group in our study received tailored support. In one of these studies, higher baseline depression and anxiety scores were associated with lower app use; similarly, anxiety was negatively associated with engagement in our study. This relationship warrants further investigation.

However, overall adherence in these studies was higher than in our study. This difference may be partly explained by the fact that patients with chronic pain often experience, in addition to depression and anxiety, other comorbidities and persistent pain, which may further complicate engagement and adherence. In particular, our sample included patients with complex chronic pain conditions, all of whom were refractory to treatment. Besides measuring engagement, it is important to define what meaningful engagement is [35]. For example, more log-ins do not automatically mean that a patient engaged more with the application. To analyze this, the minutes spent on the application would need to be structurally captured. Moreover, meaningful engagement not only is time related but also can include engagement with certain parts of the application, regularity of use, or sustained use for a certain period. Another aspect is that patients' symptoms might improve over time, and therefore, the application becomes less needed, which could be shown in less use [35]. Therefore, it is important to investigate which part or combination of the application is effective or how much time patients have to spend on the application to achieve an improvement of their symptoms when designing studies with therapeutic applications. A systematic review of mHealth interventions for chronic pain self-management reported study durations ranging from 4 to 24 weeks [36]. In comparison, the survival time observed in our study with the POCOS application was relatively short. If the effectiveness of the application is to be assessed in future studies, extending survival time by implementing strategies to foster user engagement will likely be necessary. The successful phone support that has been provided for participants who reached out for help with downloading and account registration suggests that remote but assisted onboarding could be another option to improve adoption. Future research should investigate whether this approach could be an option on a larger scale, as this would offer more flexibility for patients and HCPs than in-person assisted onboarding. It remains unclear why none of the participants in the standard onboarding group reached out to the research team for additional support to download and create an account, while in the video group, 4 participants reached out. We assume that after watching the video showing and explaining the application, participants from video group got more motivated and incentivized to use the application than the participants from the standard group.

Application engagement was consistent across onboarding groups, likely because the application was user-friendly and the user manual was clear and sufficient. The SUS score of 70.31 supports the hypothesis that the application has good usability. However, the score must be interpreted with care due to the limited response rate. Feedback from participants showed that there is room for improvement in the technical functionality and content of the application. This could be addressed with another round of cocreation and design workshops with patients, as well as input from other HCPs.

Although engagement strategies were considered during the application's development, other factors may also play a role and warrant further investigation. For example, previous studies have identified several patient-level factors that negatively influence engagement with mHealth applications, including higher pain levels, depression, longer disease duration, male sex, and lower sociodemographic status or education level [20,37]. Participants in our cohort were characterized by high levels of pain and depression and had a long-standing, treatment-refractory condition, which may have negatively affected adherence and sustained engagement with the application. Patients with chronic pain are considered particularly challenging with regard to adherence compared to other patient groups. This is partly due to their long-term experience with limited symptom relief. The complexity and multimodal nature of their treatments result in a high treatment burden. Additionally, emotional factors such as depression, anxiety, and frustration are more prevalent in this population and can negatively affect motivation and confidence in managing treatment, further reducing adherence [38-40]. Studies investigating adherence to medication among patients with chronic pain have shown, for example, that patients often report taking their medication as prescribed, while in reality they self-adjust their intake. Medication nonadherence in this population is a known issue, although it varies across countries, ranging from 8% to 60%. A systematic review found that age, pain intensity, and the patient-caregiver relationship were negatively associated with adherence to analgesic medication in patients with chronic pain [38,41-43]. While there is extensive literature on medication adherence, less is known about adherence to mHealth interventions, particularly in populations with chronic pain. It is possible that behaviors or factors that negatively influence medication adherence may also translate into reduced adherence when using digital health applications. For example, a study investigating predictors of patient engagement with an app for chronic pain patients reported that those experiencing higher levels of pain showed better engagement with the app than patients whose pain was less severe [44]. Identifying more characteristics in patients with chronic musculoskeletal pain associated with higher application use could enable the development of patient profiles, recognizing that mHealth solutions may not be equally suitable for all individuals.

Beyond the limited return of SUS survey responses, participants provided minimal insight into their application use behavior (eg, minutes spent on the application). The active provision of usability data seems to be too demanding for participants and should be switched fully to passive collection, as has been done with the log-ins. To improve adherence, measures such as increased follow-up, support and monitoring by study personnel, or reimbursement for study tasks could be an option [45].

Limitations

Although our study provided insights into the barriers of managing the onboarding process without assistance, there are

XSL•FO
RenderX

some limitations. We did not track whether participants of group 2 watched the video or how often; however, we were able to track the views and downloads as a total. Therefore, we cannot say with certainty that every participant watched the video. However, on the basis of 17 views and 33 downloads, it is most likely that each of the 17 participants in group 2 had watched the video at least once. The application used in this study is still under development, which led to the occurrence of a technical bug during the 2-week period. The section with advice had disappeared for 3 patients, which might have demotivated some patients from using the application regularly. Although the usability of the application was assessed in this study, no conclusion can be drawn from the SUS score due to the small amount of feedback that has been obtained. A formal sample size calculation was not undertaken, as this study was designed as a feasibility study. Therefore, the sample size was based on practical considerations rather than on estimating or detecting an effect size. Therefore, this study had a small sample size and should be replicated with a larger cohort to assess the robustness and replicability of the findings. Future studies should also include more detailed patient characteristics and demographics to provide a more comprehensive understanding of the factors influencing adoption and engagement. In particular, variables such as depression and anxiety should be measured on a continuous scale rather than only as binary indicators, to capture the full range of symptom severity and their potential impact on engagement. Moreover, as patients were together undergoing a 2-week in-house multimodal chronic pain program before the study and commonly maintained contact beyond its completion, randomization was not feasible owing to a substantial risk of cross-group contamination.

Conclusions

Onboarding without in-person assistance appears to potentially hinder the adoption of digital health applications, even among relatively younger users (mean age 45.7, SD 12.2 years). While these apps are designed to support remote access and use, HCPs should still consider offering guidance during the initial setup, particularly for downloading and registration.

Acknowledgments

The authors thank all participants who took part in the study. ChatGPT (version 5.3; OpenAI) was used to check grammar and sentence structure but not to write the manuscript.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Funding

The authors declared that no financial support was received for this study.

Authors' Contributions

Conceptualization: CNK, MB, TH
Data curation: CNK, MB
Formal analysis: CNK
Investigation: CNK, MB, JM, TP, TH
Methodology: CNK, TH
Project administration: CNK, PMA, TH
Resources: CNK, TH
Supervision: PMA, TH
Validation: CNK, TH
Visualization: CNK, MB, TH
Writing – original draft: CNK, MB
Writing – review and editing: CNK, MB, JM, TP, PMA, TH

Conflicts of Interest

MB and TH are shareholders and board members of Atreon SA. TH is the scientific advisor of Vtuls. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
User manual for the study and pain organizer and companion system (POCOS) app.
[PDF File (Adobe PDF File), 1966 KB - mhealth_v14i1e78827_app1.pdf]

References

XSL•FO
RenderX

1. Hewitt S, Sephton R, Yeowell G. The effectiveness of digital health interventions in the management of musculoskeletal conditions: systematic literature review. J Med Internet Res 2020 Jun 05;22(6):e15617 [FREE Full text] [doi: 10.2196/15617] [Medline: 32501277]
2. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
3. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
4. Gan DZ, McGillivray L, Han J, Christensen H, Torok M. Effect of engagement with digital interventions on mental health outcomes: a systematic review and meta-analysis. Front Digit Health 2021;3:764079 [FREE Full text] [doi: 10.3389/fdgth.2021.764079] [Medline: 34806079]
5. Wouters F, Gruwez H, Vranken J, Vanhaen D, Daelman B, Ernon L, et al. The potential and limitations of mobile health and insertable cardiac monitors in the detection of atrial fibrillation in cryptogenic stroke patients: preliminary results from the REMOTE trial. Front Cardiovasc Med 2022 Apr 13;9:848914 [FREE Full text] [doi: 10.3389/fcvm.2022.848914] [Medline: 35498000]
6. Cruz-Ramos NA, Alor-Hernández G, Colombo-Mendoza LO, Sánchez-Cervantes JL, Rodríguez-Mazahua L, Guarneros-Nolasco LR. mHealth apps for self-management of cardiovascular diseases: a scoping review. Healthcare (Basel) 2022 Feb 08;10(2):322 [FREE Full text] [doi: 10.3390/healthcare10020322] [Medline: 35206936]
7. Pratap A, Renn BN, Volponi J, Mooney SD, Gazzaley A, Arean PA, et al. Using mobile apps to assess and treat depression in Hispanic and Latino populations: fully remote randomized clinical trial. J Med Internet Res 2018 Aug 09;20(8):e10130 [FREE Full text] [doi: 10.2196/10130] [Medline: 30093372]
8. Labinsky H, Gupta L, Raimondo MG, Schett G, Knitza J. Real-world usage of digital health applications (DiGA) in rheumatology: results from a German patient survey. Rheumatol Int 2023 Apr;43(4):713-719 [FREE Full text] [doi: 10.1007/s00296-022-05261-7] [Medline: 36543961]
9. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
10. Gan DZ, McGillivray L, Larsen ME, Christensen H, Torok M. Technology-supported strategies for promoting user engagement with digital mental health interventions: a systematic review. Digit Health 2022;8:20552076221098268 [FREE Full text] [doi: 10.1177/20552076221098268] [Medline: 35677785]
11. Lipschitz JM, Pike CK, Hogan TP, Murphy SA, Burdick KE. The engagement problem: a review of engagement with digital mental health interventions and recommendations for a path forward. Curr Treat Options Psychiatry 2023 Sep;10(3):119-135 [FREE Full text] [doi: 10.1007/s40501-023-00297-3] [Medline: 38390026]
12. Viswanathan M, Golin CE, Jones CD, Ashok M, Blalock SJ, Wines RC, et al. Interventions to improve adherence to self-administered medications for chronic diseases in the United States: a systematic review. Ann Intern Med 2012 Dec 04;157(11):785-795 [FREE Full text] [doi: 10.7326/0003-4819-157-11-201212040-00538] [Medline: 22964778]
13. Ismaila H, Blandford A, Sunkersing D, Stevenson F, Goodfellow H. Comparative insights into clinic onboarding and interaction practices for patient engagement in long COVID digital health care. Digit Health 2024 Nov 26;10:20552076241294101 [FREE Full text] [doi: 10.1177/20552076241294101] [Medline: 39600394]
14. Hernandez-Ramos R, Aguilera A, Garcia F, Miramontes-Gomez J, Pathak LE, Figueroa CA, et al. Conducting internet-based visits for onboarding populations with limited digital literacy to an mHealth intervention: development of a patient-centered approach. JMIR Form Res 2021 Apr 29;5(4):e25299 [FREE Full text] [doi: 10.2196/25299] [Medline: 33872184]
15. Lipschitz J, Miller CJ, Hogan TP, Burdick KE, Lippin-Foster R, Simon SR, et al. Adoption of mobile apps for depression and anxiety: cross-sectional survey study on patient interest and barriers to engagement. JMIR Ment Health 2019 Jan 25;6(1):e11334 [FREE Full text] [doi: 10.2196/11334] [Medline: 30681968]
16. Tudor AI, Nichifor E, Litră AV, Chi u IB, Brătucu TO, Brătucu G. Challenges in the adoption of eHealth and mHealth for adult mental health management-evidence from Romania. Int J Environ Res Public Health 2022 Jul 27;19(15):9172 [FREE Full text] [doi: 10.3390/ijerph19159172] [Medline: 35954526]
17. Zakerabasali S, Ayyoubzadeh SM, Baniasadi T, Yazdani A, Abhari S. Mobile health technology and healthcare providers: systemic barriers to adoption. Healthc Inform Res 2021 Oct;27(4):267-278 [FREE Full text] [doi: 10.4258/hir.2021.27.4.267] [Medline: 34788907]
18. Dahlhausen F, Zinner M, Bieske L, Ehlers JP, Boehme P, Fehring L. There's an app for that, but nobody's using it: insights on improving patient access and adherence to digital therapeutics in Germany. Digit Health 2022 Jul 03;8:20552076221104672 [FREE Full text] [doi: 10.1177/20552076221104672] [Medline: 35811758]
19. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022 Apr 26;24(4):e35120 [FREE Full text] [doi: 10.2196/35120] [Medline: 35471414]

XSL•FO
RenderX

20. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371 [FREE Full text] [doi: 10.2196/35371] [Medline: 35612886]
21. Steinberg JR, Yeh C, Jackson J, Saber R, Niznik CM, Leziak K, et al. Optimizing engagement in an mHealth intervention for diabetes support during pregnancy: the role of baseline patient health and behavioral characteristics. J Diabetes Sci Technol 2022 Nov 22;16(6):1466-1472. [doi: 10.1177/19322968211035441] [Medline: 34423677]
22. Sarraju A, Seninger C, Parameswaran V, Petlura C, Bazouzi T, Josan K, et al. Pandemic-proof recruitment and engagement in a fully decentralized trial in atrial fibrillation patients (DeTAP). NPJ Digit Med 2022 Jun 28;5(1):80 [FREE Full text] [doi: 10.1038/s41746-022-00622-9] [Medline: 35764796]
23. Luderer H, Chiodo L, Wilson A, Brezing C, Martinez S, Xiong X, et al. Patient engagement with a game-based digital therapeutic for the treatment of opioid use disorder: protocol for a randomized controlled open-label, decentralized trial. JMIR Res Protoc 2022 Jan 26;11(1):e32759. [doi: 10.2196/32759] [Medline: 35080499]
24. Simmons LA, Phipps JE, Whipps M, Smith P, Carbajal KA, Overstreet C, et al. From hybrid to fully remote clinical trial amidst the COVID-19 pandemic: strategies to promote recruitment, retention, and engagement in a randomized mHealth trial. Digit Health 2022 Sep 25;8:20552076221129065 [FREE Full text] [doi: 10.1177/20552076221129065] [Medline: 36185388]
25. POCOS - instructions d'utilisation - FR. POCOS YouTube Page. URL: https://www.youtube.com/watch?v=UgKa-yyxL7I [accessed 2026-03-23]
26. Rennie KL, Lawlor ER, Yassaee A, Booth A, Westgate K, Sharp SJ, et al. Engagement with mHealth COVID-19 digital biomarker measurements in a longitudinal cohort study: mixed methods evaluation. J Med Internet Res 2023 Jan 13;25:e40602 [FREE Full text] [doi: 10.2196/40602] [Medline: 36194866]
27. Blanchard M, Koller CN, Azevedo PM, Prétat T, Hügle T. Development of a management app for postviral fibromyalgia-like symptoms: patient preference-guided approach. JMIR Form Res 2024 Apr 19;8:e50832. [doi: 10.2196/50832] [Medline: 38639986]
28. Blanchard M, Backhaus L, Ming Azevedo P, Hügle T. An mHealth app for fibromyalgia-like post-COVID-19 syndrome: protocol for the analysis of user experience and clinical data. JMIR Res Protoc 2022 Feb 04;11(2):e32193 [FREE Full text] [doi: 10.2196/32193] [Medline: 34982039]
29. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar 19;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
30. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]
31. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008;24(6):574-594. [doi: 10.1080/10447310802205776]
32. Selter A, Tsangouri C, Ali SB, Freed D, Vatchinsky A, Kizer J, et al. An mHealth app for self-management of chronic lower back pain (Limbr): pilot study. JMIR Mhealth Uhealth 2018 Sep 17;6(9):e179 [FREE Full text] [doi: 10.2196/mhealth.8256] [Medline: 30224333]
33. Bhatia A, Kara J, Janmohamed T, Prabhu A, Lebovic G, Katz J, et al. User engagement and clinical impact of the Manage My Pain app in patients with chronic pain: a real-world, multi-site trial. JMIR Mhealth Uhealth 2021 Mar 04;9(3):e26528 [FREE Full text] [doi: 10.2196/26528] [Medline: 33661130]
34. McCloud T, Jones R, Lewis G, Bell V, Tsakanikos E. Effectiveness of a mobile app intervention for anxiety and depression symptoms in university students: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 31;8(7):e15418 [FREE Full text] [doi: 10.2196/15418] [Medline: 32735221]
35. Strauss G, Flannery JE, Vierra E, Koepsell X, Berglund E, Miller I, et al. Meaningful engagement: a crossfunctional framework for digital therapeutics. Front Digit Health 2022 Aug 11;4:890081 [FREE Full text] [doi: 10.3389/fdgth.2022.890081] [Medline: 36052316]
36. Moreno-Ligero M, Moral-Munoz JA, Salazar A, Failde I. mHealth intervention for improving pain, quality of life, and functional disability in patients with chronic pain: systematic review. JMIR Mhealth Uhealth 2023 Feb 02;11:e40844 [FREE Full text] [doi: 10.2196/40844] [Medline: 36729570]
37. Abrams CM, Davila N, Badawy SM, Marks S, Kepper MM, Gibson RW, et al. Factors influencing engagement with adherence-enhancing mobile health app in sickle cell disease. BMC Digit Health 2024 Dec 03;2:81. [doi: 10.1186/s44247-024-00138-0]
38. Timmerman L, Stronks DL, Groeneweg JG, Huygen FJ. Prevalence and determinants of medication non-adherence in chronic pain patients: a systematic review. Acta Anaesthesiol Scand 2016 Apr 09;60(4):416-431. [doi: 10.1111/aas.12697] [Medline: 26860919]
39. Rosser BA, McCracken LM, Velleman SC, Boichat C, Eccleston C. Concerns about medication and medication adherence in patients with chronic pain recruited from general practice. Pain 2011 May;152(5):1201-1205 [FREE Full text] [doi: 10.1016/j.pain.2011.01.053] [Medline: 21376463]

XSL•FO
RenderX

40. Pagès-Puigdemont N, Tuneu L, Masip M, Valls P, Puig T, Mangues MA. Determinants of medication adherence among chronic patients from an urban area: a cross-sectional study. Eur J Public Health 2019 Jun 01;29(3):419-424. [doi: 10.1093/eurpub/cky259] [Medline: 30521035]
41. da Costa FA, Pedro AR, Teixeira I, Bragança F, da Silva JA, Cabrita J. Primary non-adherence in Portugal: findings and implications. Int J Clin Pharm 2015 Aug;37(4):626-635. [doi: 10.1007/s11096-015-0108-1] [Medline: 25832675]
42. Broekmans S, Dobbels F, Milisen K, Morlion B, Vanderschueren S. Pharmacologic pain treatment in a multidisciplinary pain center: do patients adhere to the prescription of the physician? Clin J Pain 2010 Feb;26(2):81-86. [doi: 10.1097/AJP.0b013e3181b91b22] [Medline: 20090432]
43. Ortega-Jiménez P, De Sola H, Salazar A, Dueñas M, Del Reguero L, Failde I. A nationwide cross-sectional study of self-reported adherence and factors associated with analgesic treatment in people with chronic pain. J Clin Med 2020 Nov 14;9(11):3666 [FREE Full text] [doi: 10.3390/jcm9113666] [Medline: 33202632]
44. Ross EL, Jamison RN, Nicholls L, Perry BM, Nolen KD. Clinical integration of a smartphone app for patients with chronic pain: retrospective analysis of predictors of benefits and patient engagement between clinic visits. J Med Internet Res 2020 Apr 16;22(4):e16939 [FREE Full text] [doi: 10.2196/16939] [Medline: 32297871]
45. Zweben A, Fucito LM, O'Malley SS. Effective strategies for maintaining research participation in clinical trials. Drug Inf J 2009 Jul 30;43(4):459-467 [FREE Full text] [doi: 10.1177/009286150904300411] [Medline: 24311825]

Abbreviations

HCP: health care professional
mHealth: mobile health
POCOS: pain organizer and companion system
SUS: System Usability Scale

Edited by L Buis; submitted 26.Jun.2025; peer-reviewed by S Goodday, K Becerro de Bengoa Losa; comments to author 16.Dec.2025; accepted 26.Feb.2026; published 30.Mar.2026.

Please cite as:
Koller CN, Blanchard M, Mettler J, Prétat T, Ming Azevedo P, Hügle T
Impact of Different Onboarding Strategies on Low Adoption and Engagement With a Self-Monitoring and Management App for Chronic Musculoskeletal Pain: Prospective Study
JMIR Mhealth Uhealth 2026;14:e78827
URL: https://mhealth.jmir.org/2026/1/e78827
doi:10.2196/78827
PMID:

©Cinja Nadana Koller, Marc Blanchard, Johanna Mettler, Tiffany Prétat, Pedro Ming Azevedo, Thomas Hügle. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 30.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effect of Teledermatologist Supervision in AI Skin Cancer Apps on Patients' Trust and Adherence Intentions: Randomized Controlled Trial

Ritwik Swain[1], BSc, MSc, PhD; Mihaela Masic[1,2], BSc; Justine Staal[1], BSc, MSc, PhD; Natalie Dangerman[3]; Elise Barendse[3], BSc, MSc; Marlies Wakkee[4], MD, PhD; Laura Zwaan[1], BSc, MSc, PhD

[1]Department of Public Health, Erasmus MC, Dr Molewaterplein 40, Rotterdam, South Holland, The Netherlands

[2]Department of Social and Behavioural Sciences, Erasmus University Rotterdam, Rotterdam, South Holland, The Netherlands

[3]Department of Quality and Patient Care, Erasmus MC, Rotterdam, South Holland, The Netherlands

[4]Deparment of Dermatology, Erasmus MC Cancer Institute, University Medical Center, Rotterdam, South Holland, The Netherlands

Corresponding Author:
Ritwik Swain, BSc, MSc, PhD
Department of Public Health, Erasmus MC, Dr Molewaterplein 40, Rotterdam, South Holland, The Netherlands

Abstract

Background: Skin cancer incidence is rising globally, increasing pressure on dermatology services and making efficient triage increasingly important. AI-driven skin cancer mobile health apps may help address this need, but their usefulness depends on whether users trust the advice and intend to follow it. Teledermatologist supervision has been proposed to combine AI efficiency with human clinical oversight, but dermatologist shortages make it critical to understand whether such supervision improves patient trust and adherence at the population level, or whether AI-only advice suffices for most users.

Objective: This study tests whether teledermatologist supervision in an AI-driven skin cancer app affects patient trust and adherence intentions, compared with AI-only advice, and whether these effects differ by motivational context and risk assessment level.

Methods: The randomized controlled, online vignette experiment with repeated measures was conducted from May to June 2025 as a single-center study among Dutch adults via an academic hospital-based patient panel. Eligible participants were adult members of the Erasmus MC Digital Patient Panel invited by email; those not completing all 4 vignette trials were excluded from the analysis. Participants completed 4 simulated mobile app scenarios varying in advice source (AI-only vs hybrid teledermatologist and AI) and risk assessment level (high vs low). The primary outcomes were trust in the advice (0 - 100 scale) and adherence intention (recoded 3-level ordinal scale). Mixed-effects models were used to analyze the data.

Results: Of the 2707 patient panel members, 879 participated (response rate: 32.5%). Participants were aged 18 to 93 (mean 62.5, SD 12.64) years; 50.4% (n=443) were female, and 21.4% (n=188) had a history of skin cancer. Participants were randomly allocated to motivation conditions (concern: n=438 vs curiosity: n=441). Trust was higher for hybrid advice than AI-only advice (mean 76.9, SE 0.73 vs mean 65.7, SE 0.73; estimated difference 11.98; partial η^2=0.22; *P*<.001). The odds of adherence were 1.7 times greater with hybrid vs AI-only (odds ratio 1/0.59=1.7, 95% CI 1.28-2.22; *P*<.001). High-risk assessments increased adherence (odds ratio 2.9, 95% CI 2.2-3.8; *P*<.001), with a moderated effect by motivation (*P*=.03), showing stronger adherence with concern vs curiosity. The effect of advice source on adherence was fully mediated by trust.

Conclusions: Teledermatologist supervision of AI-based advice robustly increased patient trust and adherence intentions, particularly for high-risk assessments and concerned users. Our experiment is innovative in testing teledermatologist supervision in a population-level AI skin cancer app and in examining how motivational context shapes responses to advice. Unlike studies focused on postreferral teledermatology or chatbot-based advice, this study isolates the effect of hybrid AI-human advice in a store-and-forward screening context. These findings suggest that developers and health care implementers may consider integrating teledermatologist supervision to maximize patient trust of and adherence to advice, potentially improving triaging efficiency and safety.

Trial Registration: Open Science Framework 64pjn; https://osf.io/64pjn

(JMIR Mhealth Uhealth 2026;14:e93991) doi:10.2196/93991

KEYWORDS

mobile health; mHealth apps; health behavior; patient motivation; user experience; human-AI collaboration; dermatology; teledermatology; mHealth

XSL•FO
RenderX

Introduction

Background

Globally, the incidence of both melanoma [1-3] and other skin cancer [4,5] is increasing, with related deaths projected to increase by 68% from 2020 to 2040 [6]. The increased incidence compounds the already significant strain on dermatology services, causing delays in diagnosis and treatment in some regions [7]. Teledermatology [8], whereby a dermatologist makes their assessment remotely using digital technologies, has helped address this challenge by postreferral triaging [9,10], achieving improved accessibility and cost savings of 6.3% to 45.3% per patient [11,12] by potentially avoiding unnecessary appointments with the general practitioner [13]. However, widescale teledermatology is likely unrealistic, given global dermatologist shortages [14-16].

Using AI in place of teledermatology has been proposed [17-19], as experiments found AI to diagnose lesion images more accurately than both nonspecialists [20] and dermatologists [21]. Certain use cases for AI in dermatology are promising (Du‐Harpur et al [22]), for example, as an aid for diagnosing [23-25] or in triaging [26]. In fact, both AI alone and AI with a dermatologist check were found to be more cost-effective in postreferral triage than face-to-face assessments or teledermatology [26]. Combining teledermatology and AI postreferral enhances care quality, streamlines processes, and reduces costs [27], and consequently, hybrid models are recommended to help expand access [28,29] and are often preferred by patients [23,24]. However, less research has been conducted on AI-teledermatology apps at the population level, which can face practical challenges, such as users not being sufficiently reassured by low-risk assessments and seeking doctor appointments anyway, as found in a Dutch study [30]. The app in the study [30] increased premalignant or malignant lesion detection by 32%, while also tripling to quadrupling benign lesion claims, indicating population-level overdiagnosis. Furthermore, patients receiving a low-risk assessment from the app had twice as many claims for benign skin lesions, suggesting that they were neither reassured by the app nor their general practitioner [30]. Therefore, adherence to low-risk advice for population-level apps needs to be high to mitigate the risk of increased health care demand. Using AI-based apps to prevent unnecessary appointments appears achievable, as a feasibility study reported that 54% of patients who visited the general practitioner with benign lesions and had low-risk app ratings would have felt reassured enough to cancel their general practitioner visit [13].

People's adherence may differ based on their reason for using the app. For highly concerned users, the perceived threat [31] of skin cancer cannot be alleviated by an algorithm [32], so they will still seek follow-up appointments regardless of the app's risk assessment. In contrast, users motivated by curiosity, which is often a reason to seek health information [33], may be less inclined to follow-up with their general practitioner even if advised by their app. How users' adherence is influenced by motivational context (ie, concern or curiosity) and risk assessment (ie, low or high risk) is unexplored, yet it is critical to understand if AI-based apps in dermatology are able to alleviate health care demand at the population level. Trust is key [34] as patients who trust their clinicians show greater confidence in clinicians' decisions [35] and better adherence [36,37], and presumably the same would apply to health apps. An online experiment [38] found that trust in advice is highest when given by a doctor, followed by human-AI collaboration, and least when offered by AI alone. However, the experiment [38] involved real-time involvement of a doctor, but in teledermatology, a store-and-forward format [39] is common, where patients upload photos [28] for later assessment. Furthermore, the experiment's [38] AI-only option was a chatbot, not an AI photo analysis.

Although population-level AI skin cancer apps have shown feasibility [13], it remains unclear whether teledermatologist supervision improves users' trust and intended adherence when advice is delivered in a patient-facing, store-and-forward format. This study aims to address this gap by assessing user responses to store-and-forward–style, AI-based dermatology apps using a vignette-based experimental design. Vignette experiments have been established as a valid, reliable, and practical method in health care research, offering strong criterion validity while bypassing the ethical and practical constraints of real-world experiments [40].

Research Question and Hypotheses

The current study poses the question: How do advice source (AI alone or hybrid) and user motivation (curiosity vs concern) shape patient trust and adherence intentions to the provided advice within the context of AI-supported teledermatology? Three preregistered hypotheses are posed:

1. H1: Trust (H1a) and adherence intentions (H1b) will be higher for hybrid advice vs AI-only (main effect of advice source on trust and adherence intentions).
2. H2: High-risk assessments will increase adherence intentions (H2a), particularly in the concern condition (H2b) (a main effect of risk assessment and an interaction of motivation and risk assessment on adherence intentions).
3. H3: Trust will mediate the effect of advice source on adherence intentions.

Methods

Trial Design

Our randomized controlled, online vignette experiment with repeated measures uses a 2×2×2 design, with motivational context (concern vs curiosity) as a between-participants factor and advice source (AI-only vs hybrid) and risk assessment (high vs low) as within-participants factors. The dependent variables include trust and adherence to medical advice. Figure 1 presents a study design flowchart. The public was not involved in the design, conduct, or reporting of the trial.

XSL•FO
RenderX

Figure 1. Experimental design flowchart for a 2×2×2 randomized vignette experiment among Dutch adults recruited from the Erasmus MC Patient Panel, Rotterdam, the Netherlands, May to June 2025. Participants were randomized to curiosity or concern, and then completed 4 simulated app trials varying in advice source (AI-only vs hybrid teledermatologist + AI) and risk assessment (low vs high). Participants were assigned to motivation conditions between participants, then received both advice source types and risk assessment levels in randomized order (within participants). Each completed 4 trials of risk assessments, with trust and adherence intentions measured after each trial.

XSL•FO
RenderX

Sample Size

An a priori power analysis was conducted in G*Power 3.1.9.6 [41] for the planned within-between interaction in a repeated measures ANOVA (test family: *F* tests; statistical test: ANOVA: repeated measures, within-between interaction). Because no directly comparable prior studies were available, we conservatively assumed a small-to-medium effect size of f(V)=0.25 [42], α=.05, power=0.80, 2 groups (motivation factor), 4 measurements (2×2 within-participants factors), and nonsphericity correction (=1). This yielded a minimum total sample size of 168 [41].

Materials

The experiment was created using Qualtrics and conducted in Dutch. No personnel were required to conduct the study as it was automated and online. The experiment included an introduction to AI, experiment instructions (Multimedia Appendix 1), 2 motivation texts, and image stimuli for the motivation conditions (Multimedia Appendix 2). Additionally, 4 simulated mobile health app trials, varying in combinations between advice source and risk assessment, were included (see Figure 2 and Multimedia Appendix 3).

XSL•FO
RenderX

Figure 2. Simulated medical risk assessment outputs used in a randomized vignette experiment among Dutch adults recruited from the Erasmus MC Patient Panel, Rotterdam, the Netherlands, May to June 2025. Each risk assessment also reminds participants of the advice source and provides follow-up advice. App images were created using Canva and AI.

XSL•FO
RenderX

Ethical Considerations

This study was approved by the Medical Ethics Committee of Erasmus MC (MEC-2025 - 0371) and conducted in accordance with the procedures of the Erasmus MC Digital Patient Panel [43]. Informed consent was obtained through panel membership registration, by which members agree to be approached for research participation and may opt out of any individual study at any time; no separate consent procedure was required within the study itself. All data were collected anonymously through the panel infrastructure; no personally identifiable information was linked to experimental responses, ensuring participant confidentiality. Participants received no financial or material compensation for participation. No harms were posed on participants as it was an online study, and no concomitant care was provided. All efforts have been made to ensure that individual participants cannot be identified in any images of the manuscript or supplementary material.

Procedure

As this study involved a hypothetical vignette-based experiment assessing patient-reported trust and adherence intentions rather than a health-related intervention delivered to patients or a health-related outcome, prospective trial registration was not required under International Committee of Medical Journal Editors guidelines [44]. The study was nonetheless preregistered on the Open Science Framework [45] to enhance transparency of our research on May 26, 2025, after recruitment had begun but before any data inspection, cleaning, or analysis. The experiment was emailed to 2707 panel members between May 22 and June 10, 2025, with the first response being on May 22. The data collection period was defined by the panel administrators. After demographics and an AI attitudes scale, participants were randomly assigned by Qualtrics to motivation conditions (concern or curiosity; see Multimedia Appendix 2). They completed 4 trials in 2 randomized blocks, varying advice source (AI-only vs hybrid) and risk assessment (high vs. low). Each trial presented 1 advice output, followed by trust and adherence intention measures.

Consequently, each participant saw 4 assessment outputs, 1 high-risk and 1 low-risk per advice source. After each assessment, participants answered 1 item measuring trust and another measuring their intention to adhere to the advice. Finally, participants were debriefed about the aim of the study.

Outcomes

In a preexperiment questionnaire, participants answered questions on demographic data (age, gender, and skin cancer history) and their attitudes toward AI in health care, measured using a Dutch version of the validated 4-item AI Attitude Scale (AIAS-4) [46], with participants indicating their level of agreement on scales from 1 (“Not at all”) to 10 (“Completely agree”). After each trial, trust was rated on a continuous slider (0=no trust, 100=full trust) with the question “How much trust do you have in this result?” This single-item format was chosen to minimize respondent burden across 4 repeated trials. While single-item measures cannot capture all dimensions of trust, they can provide reliable and valid indicators of overall trust judgments [47].

Adherence was assessed by asking “What would you do next regarding your mole?” Participants selected 1 of 5 response options (Table 1), which were recoded, with input from a clinician, into a 3-level ordinal adherence variable for both risk assessment levels. The recoding scheme was designed to reflect medically meaningful levels of adherence. Because the clinical implications of the same response differ depending on whether the AI advice indicated high or low risk, the classification of the responses varied by risk condition.

XSL•FO
RenderX

Table . Recoding of postvignette adherence intention responses into a 3-level ordinal outcome in a randomized vignette experiment on an AI-supported skin cancer app among Dutch adults recruited from the Erasmus MC Patient Panel, Rotterdam, the Netherlands, May to June 2025[a].

Response option	High-risk advice	Low-risk advice	Justification
"I would make an appointment with my GP the same day."	High adherence	Neglect	It is high adherence for high risk because there is same-day intention to make an appointment. Conversely, it is neglect for low risk because an appointment is made.
"I would make an appointment within a week."	Moderate adherence	Neglect	For high risk, this option is moderate adherence because an appointment is still made but with suboptimal urgency. Conversely, it is neglect for low risk because an appointment is made.
"I would make an appointment when I have time; I am not immediately worried."	Neglect	Moderate adherence	For high risk, this option is clearly neglect because there is a lack of clear intention to make a doctor's appointment. For low risk, it is moderate adherence because they might make an appointment but there is not a clear intention.
"I would monitor the mole and only make an appointment if it changes."	Neglect	High adherence	For high risk, this option is considered neglect because the user intends to wait for further signs before making an appointment. In contrast, for low risk, it is desirable that an appointment is not made, unless the mole changes.
"I would not seek any further medical assessment."	Neglect	High adherence	For high risk, this decision is clear neglect because no intention to make an appointment is present. In contrast, for low risk, it is high adherence because there is no sign that the user intends to make an appointment.

[a]Participants were asked "What would you do next regarding your mole?"

Statistical Methods

Analyses were conducted using both SPSS (version: 28.0.1.0(142); IBM Statistics) and RStudio (2025.05.1 Build 513). For H1, a linear mixed-effects model (LMM) was used with trust as the continuous outcome, advice source and risk assessment as within-participants factors, motivation as a between-participants covariate, and participant as a random intercept. Although the Shapiro-Wilk test was significant, the residuals appeared normally distributed and no other assumptions were violated, and we proceeded with the LMM. Effect size was calculated using the following equation: $\eta p2 = F \times dfeffect F \times dfeffect + dferror$.

For H1b and H2, a cumulative link mixed model (CLMM) [48,49] was used given the ordinal nature of adherence intentions. Category counts were sufficient. Logit plots supported proportional odds assumption for advice source and motivation, but risk assessment violated it. To address this, 2 binary generalized linear mixed model sensitivity analyses were conducted: 1 contrasting high adherence vs moderate adherence/neglect combined and 1 contrasting neglect vs high/moderate adherence combined. The latter reflects the theoretically motivated distinction between engagement (high and moderate adherence) and complete disengagement (neglect). The results of both sensitivity analyses are reported alongside the primary CLMM findings. A CLMM was conducted with adherence intentions, predicted by the independent variables: advice source, risk assessment, motivation, and their interactions, with a random intercept for participant ID.

For H3, the main effects of trust, advice source, and their interaction on adherence intentions were tested using a CLMM, with no proportional odds assumption violations for any predictor. A mediation analysis was conducted to test the indirect effect of advice source on adherence intentions via trust as the product of 2 paths: source → trust and trust → adherence intentions, assessed through a Monte Carlo simulation with 10,000 draws. To obtain unbiased estimates, both the CLMM for trust's effect on adherence intentions (path b) and the LMM for source's effect on trust (path a) were rerun without any interaction terms.

Reporting Guidelines

We reported this randomized vignette experiment in accordance with [50] the CONSORT (Consolidated Standards of Reporting Trials) abstract extension [51], CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and

XSL•FO
RenderX

Mobile Health Applications and Online Telehealth) guidelines (Checklist 1) [52], and TIDieR (Template for Intervention Description and Replication) guidelines [53].

Changes to Trial Protocol

We made the following four changes compared to the preregistration [45]:

1. A wording clarification for H2
2. Substitution of mixed-factor ANOVA with LMM for trust
3. Substitution of mixed-factor ANOVA with CLMM for adherence
4. Replacement of the regression-based mediation with Monte Carlo mediation using mixed-model path estimates

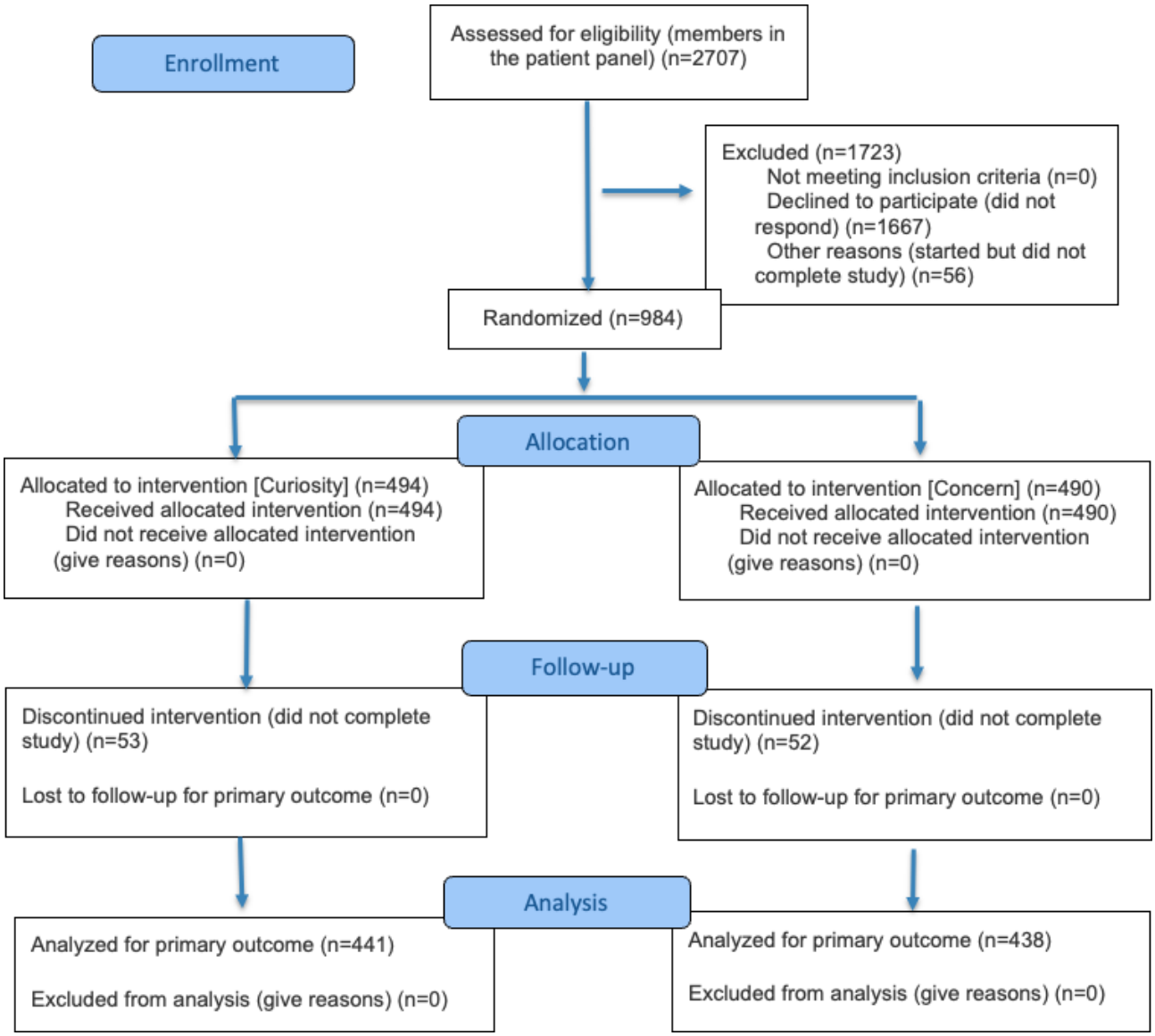
Points 2 and 3 allow for random intercepts for each participant, enhancing the modeling of individual variability [54]. A full account of deviations from the preregistration, including justification for each change, is documented in a Transparent Changes file on the OSF project page [45] and in Multimedia Appendix 4.

Results

Participant Flow

The study received 1040 responses in total. Of the 2707 patient panel members, there were 879 full completions, resulting in a response rate of 32.5% (Figure 3).

Figure 3. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of participant screening, randomization, exclusions, and analysis in the randomized vignette experiment among Dutch adults recruited from the Erasmus MC Patient Panel, Rotterdam, the Netherlands, May to June 2025.



Excluded participants were slightly older than included participants (mean 64.9, SD 12.31 y vs mean 62.5, SD 12.62 y; t_{1030}=2.19, d=0.19; P=.03). There were no statistically significant differences in baseline AI attitudes (mean 5.95, SD 2.33 vs mean 6.17, SD 2.08; P=.30), gender distribution (P=.22), personal skin cancer history (P=.82), or motivation assignment (P=.95).

Baseline Data

The median completion time was 4 minutes 56 seconds (IQR 2:55-3:45; range 1:28-175:26). Participants had a mean age of

XSL•FO
RenderX

62.5 (SD 12.64, range 18 - 93) years with balanced gender distribution (female: n=443, 50.4%; male: n=433, 49.3%). A total of 188 (21%) reported a prior skin cancer diagnosis. The average AIAS-4 score was 6.17 (SD 2.08; n=879), indicating a slightly positive attitude toward AI.

Effects on Primary Outcomes of Trust and Adherence Intentions

H1a Influence of Advice Source on Trust

The LMM revealed a significant main effect of advice source on trust ($F_{1,2643}$=758.50; *P*<.001); trust was higher for hybrid (estimated marginal mean 76.88, SE 0.73) than AI-only (estimated marginal mean 65.68, SE 0.73), with an estimated difference of 11.98 (95% CI 10.39-13.58). This relatively narrow interval indicates a precise estimate of the source effect on trust. Advice source's effect size was large (partial η^2=0.22), supporting hypothesis H1a that participants trust hybrid-based assessments more than AI-only assessments. Table S1 in Multimedia Appendix 5 shows the tests of fixed effects, while Table S2 in Multimedia Appendix 5 shows the estimates of fixed effects.

H1b and H2: Influence of Advice Source, Risk Assessment, and Motivational Context on Adherence Intentions

In all CLMMs, odds ratios (ORs) reflect the cumulative odds of being in a higher adherence intention category, with values greater than 1 indicating increased odds of stronger adherence intentions. The CLMM found a significant main effect of advice source (*b*=–0.52, SE 0.14; *z*=–3.84; *P*<.001) on adherence intentions (see the full table in Multimedia Appendix 6). The OR for hybrid compared to AI was approximately 0.59 (OR 1/0.59=1.7, 95% CI 1.28-2.22; *P*<.001), indicating that participants were about 1.7 times more likely to be in a higher adherence intention category under the hybrid advice source condition compared to the AI-only advice, supporting hypothesis H1b. The CI suggests a reasonably precise estimate of this effect.

The CLMM also found a significant main effect of risk assessment on adherence intentions (*b*=–1.07, SE =0.14; *z*=–7.86; *P*<.001). Because lower scores indicate stronger adherence intentions, this means that participants who received a high-risk assessment were about 3 times more likely (OR 2.9, 95% CI 2.2-3.8) to report higher adherence intentions compared to those with low-risk diagnoses. The relatively narrow CI indicates a precise estimate of the risk effect.

Importantly, a significant interaction between risk and motivation was found (*riskHigh* × *motivationCuriosity; b*=0.40, SE =0.19; *z*=2.13; *P*=.03). This indicates that the boost in adherence intentions associated with high risk was weaker among curiosity-motivated participants compared to those motivated by concern, which can also be seen in Table 2.

Table . Distribution of adherence intention responses by motivation condition, advice source, and risk assessment in a randomized vignette experiment among Dutch adults recruited from the Erasmus MC Patient Panel, Rotterdam, the Netherlands, May to June 2025[a].

Motivation	Source	Risk	Total, n	High adherence intention, n (%)	Moderate adherence intention, n (%)	Neglect, n (%)
Concern	AI	High	438	264 (60.3)	142 (32.4)	32 (7.3)
Concern	AI	Low	438	192 (43.8)	89 (20.3)	157 (35.8)
Concern	Hybrid	High	438	297 (67.8)	120 (27.4)	21 (4.8)
Concern	Hybrid	Low	438	246 (56.2)	65 (14.8)	127 (29)
Curiosity	AI	High	441	247 (56)	141 (32)	53 (12)
Curiosity	AI	Low	441	206 (46.7)	94 (21.3)	141 (32)
Curiosity	Hybrid	High	441	281 (63.7)	130 (29.5)	30 (6.8)
Curiosity	Hybrid	Low	441	255 (57.8)	64 (14.5)	122 (27.7)

[a]Percentages were calculated within each motivation × source × risk condition.

In practical terms, individuals driven by concern showed a larger increase in adherence intentions when receiving a high-risk assessment than those driven by curiosity.

Sensitivity Analysis

Overview

To address the proportional odds assumption violation for risk assessment, 2 binary generalized linear mixed model sensitivity analyses were conducted with the same predictors and participant random intercepts. The first contrasted high adherence intentions vs moderate/neglect; the second contrasted neglect vs high/moderate adherence, reflecting the theoretically motivated distinction between engagement (high and moderate) and disengagement (neglect).

Main effects replicated across both models: hybrid source increased odds of high adherence intentions (OR 1.45, 95% CI 1.08-1.94; *P*=.01), and high-risk assessment strongly reduced odds of neglect (OR 9.40, 95% CI 6.04-14.63; *P*<.001). Critically, the risk × motivation interaction replicated in the neglect model (OR 0.44, 95% CI 0.25-0.78; *P*=.005), confirming that concern-motivated participants were significantly less likely to neglect a high-risk assessment compared to curiosity-motivated participants—consistent with the CLMM finding that the boost in adherence intentions from high risk

XSL•FO
RenderX

was larger for concern than curiosity. The interaction did not reach significance in the high adherence intentions model (P=.11), indicating that the effect is specifically concentrated at the disengagement threshold rather than the high engagement threshold. These analyses confirm that the CLMM findings are robust to the proportional odds violation. The relatively narrow CIs indicate that these effects were estimated with good precision.

H3: Trust Will Mediate the Effect of Advice Source on Adherence Intentions

CLMM on Adherence Intentions by Trust and Advice Source

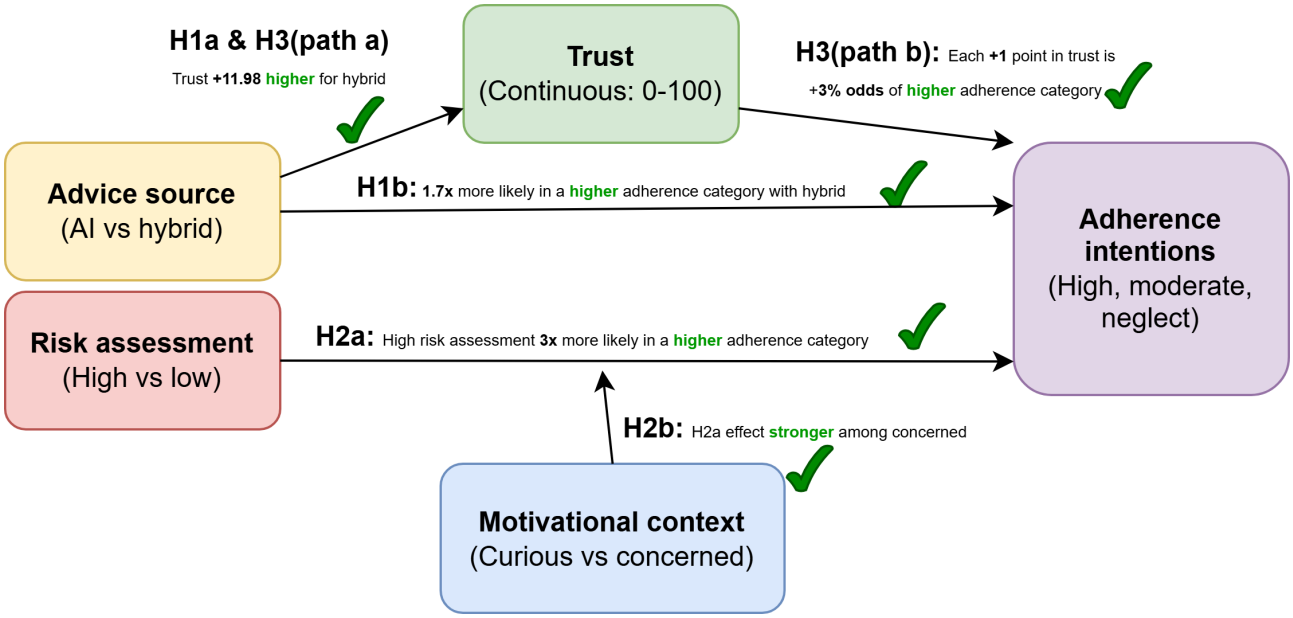
The CLMM—with participant as a random intercept—showed that greater trust significantly increased odds of stronger adherence intentions. Advice source had no main (P=.29) or interaction effect (P=.17), indicating trust's influence on adherence intentions was consistent across sources (see Multimedia Appendix 7). These results suggest that trust mediates advice source's effect on adherence intentions, supporting hypothesis H3, contrasting with H1b where advice source affected adherence when trust was not accounted for.

Mediation Analyses via Monte Carlo Simulation

The indirect effect of advice source on adherence intentions via trust was estimated as the product of 2 paths: source → trust and trust → adherence intentions, assessed through a Monte Carlo simulation with 10,000 draws. To obtain unbiased estimates, both CLMM for trust's effect on adherence (path b) and LMM for source's effect on trust (path a) were rerun without any interaction terms. The CLMM showed a statistically significant negative effect of trust on adherence intentions (estimate=–0.030, SE =0.002; P<.001), indicating that higher trust predicted better adherence. The LMM revealed a highly significant and negative effect of AI source compared to the reference hybrid source (estimate=–11.20, SE =0.41; P<.001), indicating lower trust under AI-only advice. Combining these path coefficients, the indirect effect was estimated as 0.34 with a 95% CI (0.29-0.39), excluding zero and indicating good precision as well as full mediation (H3). The relatively narrow CI indicates that this mediated effect was estimated with good precision. This positive indirect effect suggests that hybrid advice sources increase adherence intentions by enhancing trust.

Figure 4 provides an overview of the results of the main hypothesis tests.

Figure 4. Summary of preregistered hypotheses and main results from the randomized vignette experiment among Dutch adults recruited from the Erasmus MC Patient Panel, Rotterdam, the Netherlands, May to June 2025. Ticks (✔) indicate hypotheses supported by the evidence.



Discussion

Principal Findings

This randomized controlled, vignette experiment aimed to answer the question "How do advice source (AI alone or Hybrid) and user motivation (curiosity vs concern) shape patient trust and adherence intentions to the provided advice within the context of AI-supported teledermatology?" The 879 participating adults reported trusting risk assessments from a hybrid source (AI plus a teledermatologist) more than AI-only assessments and reported higher intentions to follow the advice. Mediation analysis showed that trust fully explained this effect, meaning that hybrid advice increased adherence intentions by increasing the participants' trust. A high risk of a malignancy (rather than a low risk) further strengthened the intention to follow the advice, particularly among participants who use the app out of concern rather than out of curiosity. Notably, the sensitivity analysis revealed that the interaction between motivational context and risk assessment level was specific to the disengagement threshold: concern-motivated participants were significantly less likely to neglect a high-risk assessment compared to curiosity-motivated participants, while the 2 groups did not differ substantially in rates of high adherence intentions. This suggests that users motivated by curiosity are more likely to neglect when receiving high-risk assessments, compared to users who are already concerned. From a health behavior perspective, this is consistent with protection motivation theory

XSL•FO
RenderX

[55], which posits that threat appraisal motivates protective behavior most strongly when perceived vulnerability is high and personally relevant. Curiosity-driven users—who may perceive themselves as less personally at risk—may be more prone to dismissing unexpected high-risk findings as a false alarm.

The results also identified trust as the central mechanism through which hybrid advice improves adherence intentions, highlighting the value of integrating teledermatologist supervision into AI-driven skin cancer apps to maximize triaging efficiency. An overview of the main results can be found in Figure 4. Our findings differ slightly from those of Soellner and Koenigstorfer [56], who found that the mediators between increased patient compliance and human physician supplementing AI were perceived social presence and innovativeness, suggesting multiple mediating factors are induced by human clinician presence. Similarly, Reis et al [57] found patients preferred human-only chat responses for reliability, empathy, and willingness to follow advice, compared with an AI chatbot or a hybrid chat. Soellner and Koenigstorfer [56] presented AI in a positive manner and communicated a clear purpose, mentioning AI's capability in diagnosing skin lesion photos. In contrast, Reis et al [57] used chatbot transcripts without clarifying AI's role, possibly explaining the preference for human-only advice rather than hybrid. These findings suggest hybrid advice works best when users understand the AI's purpose, not as a full substitute for human care [58]. Therefore, it is recommended to ensure patients understand the purpose of AI's involvement, alleviating patients' concerns about losing human involvement in health care [59], and for making human-AI collaboration more effective.

Practical Implications

Although including a human teledermatologist was found to increase the likelihood of higher adherence intentions by a factor of 1.7, and users' trust in the assessment from 66% to 77%, always including human teledermatologist supervision may be increasing workload too much [26]. Both AI-alone and hybrid models are more cost-effective postreferral than face-to-face assessments without AI [26], but only offering teledermatology to cases where the AI assessment is high risk, or shifting the usage of AI to postreferral by a primary health care provider, may be an effective compromise, depending on the intervention requirements and goals. This is particularly relevant given that Smak Gregoor et al [30] found that low-risk app assessments failed to prevent unnecessary appointments, tripling benign lesion claims. Offering teledermatology review to concerned users with low-risk assessments may be a strategy to prevent in-person appointments, depending on the health care system. To avoid users neglecting high-risk assessments, strategies are required to ensure that those who use the app out of curiosity take the assessment seriously. High-risk assessments may require tailored messaging for users who do not present with preexisting concern, to prevent disengagement in precisely the cases where action is most warranted.

Strengths and Limitations

Our study provides insights on intended adherence to low-risk/benign assessments—reflecting most real-world cases—rather than comparing stage 0 skin cancer with stage 2 as did Soellner and Koenigstorfer [56], which represents the minority of cases. Thus, our results provide insights into drivers of patients' adherence intentions prior to triaging, which is crucial for avoiding overdiagnosis [9,30]. It is worthy of noting that the vignettes in our study altered the motivation (curiosity vs concern), which is intertwined with perceived lesion risk. This was done intentionally to create credible vignettes. However, it is plausible that a user of an app has low perceived lesion risk and is not curious but is using the app for other reasons. Further research could investigate whether using the app because it was encouraged by a family doctor, or family member, or for any other reason influences adherence to the app's advice.

Our study did not consider costs of teledermatology integration, so AI-only solutions may be more cost-effective despite lower adherence intentions and trust although hybrid solutions appear to be most effective postreferral [26]. The study sample was drawn from the Erasmus MC Patient Panel, representing highly engaged health care users with typically higher digital literacy and education levels than in the general population, which may limit generalizability.

The vignette-based experiment involved participants experiencing hypothetical scenarios, and stated trust and adherence intentions may not fully reflect behavior during actual app use. The use of predetermined simulated app outputs rather than participant-uploaded images prioritizes internal validity by isolating the effects of advice source and risk level, though it may reduce perceived personal relevance compared to a real app scenario. However, vignette responses have been shown to correspond closely with real clinical behavior [60], and relative comparisons between conditions are generally robust to hypothetical bias [61,62]. The mixed design—with motivational context as a between-participants factor and advice source and risk level as within-participants factors—allowed for controlled manipulation of key variables while accounting for individual differences. The crossover design used randomized trial order, meaning each condition appeared first equally often across participants, so any contrast effects would be unsystematic and unlikely to bias results in a particular direction. Nonetheless, the possibility that participants compared conditions rather than evaluating each independently remains a limitation inherent to within-participants designs, and trust and adherence may further differ when participants submit an actual image and perceive the assessment as personally relevant to their own lesion. Trust was measured using a single-item slider, which, while practical for repeated trial-level measurement, cannot capture the multidimensional nature of trust. Although single-item trust measures have been shown to produce reliable and valid results comparable to multi-item scales [47], a validated instrument such as the 11-item Trust in Physician Scale by Thom et al [63] may have revealed differential effects of advice source on specific facets of trust—such as perceived competence or benevolence—rather than on overall trust judgments alone [34]. Administering such a scale after each of 4 trials would, however, have imposed considerable respondent burden. Future research could use multi-item trust scales in a between-participants design to examine this more comprehensively.

XSL•FO
RenderX

Conclusions

The findings of this preregistered online experiment support the hypothesis that adding hybrid (teledermatologist and AI) advice increases trust, which, in turn, improves intended adherence, particularly for high-risk assessments and concerned users. This study is innovative in testing teledermatologist supervision in a population-level AI skin cancer app and in examining how motivational context shapes responses to advice. It differs from existing studies focused on postreferral teledermatology [26] or chatbot-based advice [38] by isolating the effect of hybrid AI-human advice in a store-and-forward screening context. These findings can inform the design of AI-based health care apps such that maximal triaging efficiency is achieved. Practically, these results support consideration of teledermatologist supervision in AI-based triage systems where maximizing trust, adherence, and safe follow-through are priorities.

Acknowledgments

The authors would like to thank Mr Christos Diamantidis, MD, for his advice on the coding of the adherence intention answer options, as described in Table 1. The data were obtained through the Erasmus MC Patient Panel [43] (quality instrument), organized by the Directorate of Quality & Patient Care of Erasmus Medical Centre.

Generative AI Use

According to the GAIDeT taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: code generation, code optimization, data cleaning, data curation and organization, data analysis, visualization, proofreading and editing, formulation of conclusions, and translation. The GenAI tool used was Perplexity, which uses a variety of models, and ChatGPT 5.6 Terra. Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes. Declaration submitted by RS, MM

Funding

This work was supported by the Netherlands Organisation for Health Research and Development (ZonMw) through a Vidi grant (grant number 09150172210015) awarded to LZ.

Data Availability

Data used in this study are available upon request—please contact secretariaat.mgz@erasmusmc.nl. The available data include (1) the Qualtrics printout PDF, (2) R and SPSS analysis codes, and (3) the anonymized dataset.

Authors' Contributions

Conceptualization: RS, MM, LZ
Data curation: RS, MM
Formal analysis: RS, MM
Funding acquisition: LZ
Methodology: RS, MM, LZ
Project administration: ND, EB
Resources: ND, EB
Supervision: JS, MW, LZ
Visualization: RS
Writing – original draft: RS, MM
Writing – review & editing: RS, JS, MW, LZ

Conflicts of Interest

The Department of Dermatology of the Erasmus MC Cancer Institute, where MW is affiliated, has received an unrestricted research grant from SkinVision. The remaining authors declare no conflicts of interest.

Multimedia Appendix 1
Instruction before starting the main experiment.
[DOCX File, 27 KB - mhealth_v14i1e93991_app1.docx]

Multimedia Appendix 2
Motivation condition stimuli.
[DOCX File, 340 KB - mhealth_v14i1e93991_app2.docx]

Multimedia Appendix 3

XSL•FO
RenderX

Loading pages of the mobile health app.
[DOCX File, 149 KB - mhealth_v14i1e93991_app3.docx]

Multimedia Appendix 4
Document of changes to preregistration.
[DOCX File, 16 KB - mhealth_v14i1e93991_app4.docx]

Multimedia Appendix 5
Tests and estimates of fixed effects.
[DOCX File, 20 KB - mhealth_v14i1e93991_app5.docx]

Multimedia Appendix 6
Coefficients from the cumulative link mixed model examining advice source, risk assessment, and motivational context.
[DOCX File, 15 KB - mhealth_v14i1e93991_app6.docx]

Multimedia Appendix 7
Coefficients from the cumulative link mixed model examining trust and advice source.
[DOCX File, 15 KB - mhealth_v14i1e93991_app7.docx]

Checklist 1
CONSORT-EHEALTH checklist.
[PDF File, 3812 KB - mhealth_v14i1e93991_app8.pdf]

References

1. Whiteman DC, Green AC, Olsen CM. The growing burden of invasive melanoma: projections of incidence rates and numbers of new cases in six susceptible populations through 2031. J Invest Dermatol 2016 Jun;136(6):1161-1171. [doi: 10.1016/j.jid.2016.01.035] [Medline: 26902923]
2. Roky AH, Islam MM, Ahasan AMF, et al. Overview of skin cancer types and prevalence rates across continents. Cancer Pathog Ther 2025;3(2):89-100. [doi: 10.1016/j.cpt.2024.08.002] [Medline: 40182119]
3. Saginala K, Barsouk A, Aluru JS, Rawla P, Barsouk A. Epidemiology of melanoma. Med Sci (Basel) 2021 Oct 20;9(4):63. [doi: 10.3390/medsci9040063] [Medline: 34698235]
4. Lomas A, Leonardi-Bee J, Bath-Hextall F. A systematic review of worldwide incidence of nonmelanoma skin cancer. Br J Dermatol 2012 May;166(5):1069-1080. [doi: 10.1111/j.1365-2133.2012.10830.x] [Medline: 22251204]
5. Urban K, Mehrmal S, Uppal P, Giesey RL, Delost GR. The global burden of skin cancer: a longitudinal analysis from the Global Burden of Disease Study, 1990-2017. JAAD Int 2021 Mar;2:98-108. [doi: 10.1016/j.jdin.2020.10.013] [Medline: 34409358]
6. Arnold M, Singh D, Laversanne M, et al. Global burden of cutaneous melanoma in 2020 and projections to 2040. JAMA Dermatol 2022 May 1;158(5):495-503. [doi: 10.1001/jamadermatol.2022.0160] [Medline: 35353115]
7. Zhou L, Zhong Y, Han L, Xie Y, Wan M. Global, regional, and national trends in the burden of melanoma and non-melanoma skin cancer: insights from the global burden of disease study 1990-2021. Sci Rep 2025 Feb 18;15(1):5996. [doi: 10.1038/s41598-025-90485-3] [Medline: 39966563]
8. Tommasino N, Megna M, Cacciapuoti S, et al. The past, the present and the future of teledermatology: a narrative review. Clin Cosmet Investig Dermatol 2024;17:717-723. [doi: 10.2147/CCID.S462799] [Medline: 38529172]
9. Zakaria A, Miclau TA, Maurer T, Leslie KS, Amerson E. Cost minimization analysis of a teledermatology triage system in a managed care setting. JAMA Dermatol 2021 Jan 1;157(1):52-58. [doi: 10.1001/jamadermatol.2020.4066] [Medline: 33206146]
10. Drabarek D, Habgood E, Ackermann D, et al. Perspectives and experiences of patient-led melanoma surveillance using digital technologies from clinicians involved in the MEL-SELF pilot randomized controlled trial: qualitative interview study. JMIR Dermatol 2022 Dec 20;5(4):e40623. [doi: 10.2196/40623] [Medline: 37632906]
11. Ijaz M. BT01 The effectiveness of telemedicine in dermatology services: a systematic review. Br J Dermatol 2025;193(Supplement_1). [doi: 10.1093/bjd/ljaf085.526]
12. Vidal-Alaball J, Garcia Domingo JL, Garcia Cuyàs F, et al. A cost savings analysis of asynchronous teledermatology compared to face-to-face dermatology in Catalonia. BMC Health Serv Res 2018 Aug 22;18(1):650. [doi: 10.1186/s12913-018-3464-4] [Medline: 30134891]
13. Smak Gregoor AM, Sangers TE, Eekhof JA, et al. Artificial intelligence in mobile health for skin cancer diagnostics at home (AIM HIGH): a pilot feasibility study. EClinicalMedicine 2023;60:102019. [doi: 10.1016/j.eclinm.2023.102019] [Medline: 37261324]

XSL•FO
RenderX

14. Trakatelli M, Siskou S, Proby C, et al. The patient journey: a report of skin cancer care across Europe. Br J Dermatol 2012 Aug;167 Suppl 2:43-52. [doi: 10.1111/j.1365-2133.2012.11086.x] [Medline: 22881587]
15. Balboul S, Gronbeck C, Feng H. Dermatology workforce projections in the United States, 2021 to 2036. Arch Dermatol Res 2024 May 22;316(5):192. [doi: 10.1007/s00403-024-02924-2] [Medline: 38775980]
16. Shah M, Burshtein J, Zakria D, Rigel D. Analysis of trends in US dermatologist density and geographic distribution. J Am Acad Dermatol 2024 Aug;91(2):338-341. [doi: 10.1016/j.jaad.2024.03.037] [Medline: 38574771]
17. Talebi-Liasi F, Markowitz O. Is artificial intelligence going to replace dermatologists? Cutis 2020 Jan;105(1):28-31. [Medline: 32074153]
18. Li Z, Koban KC, Schenck TL, Giunta RE, Li Q, Sun Y. Artificial intelligence in dermatology image analysis: current developments and future trends. J Clin Med 2022 Nov 18;11(22):6826. [doi: 10.3390/jcm11226826] [Medline: 36431301]
19. Lim BCW, Flaherty G. Artificial intelligence in dermatology: are we there yet? Br J Dermatol 2019 Jul;181(1):190-191. [doi: 10.1111/bjd.17899] [Medline: 30912580]
20. Salinas MP, Sepúlveda J, Hidalgo L, et al. A systematic review and meta-analysis of artificial intelligence versus clinicians for skin cancer diagnosis. NPJ Digit Med 2024 May 14;7(1):125. [doi: 10.1038/s41746-024-01103-x] [Medline: 38744955]
21. Biswas S, Achar U, Hakim B, Achar A. Artificial intelligence in dermatology: a systematized review. Int J Dermatol Venereol 2025;8(1):33-39. [doi: 10.1097/JD9.0000000000000404]
22. Du-Harpur X, Watt FM, Luscombe NM, Lynch MD. What is AI? Applications of artificial intelligence to dermatology. Br J Dermatol 2020 Sep;183(3):423-430. [doi: 10.1111/bjd.18880] [Medline: 31960407]
23. Goessinger EV, Niederfeilner JC, Cerminara S, et al. Patient and dermatologists' perspectives on augmented intelligence for melanoma screening: a prospective study. J Eur Acad Dermatol Venereol 2024 Dec;38(12):2240-2249. [doi: 10.1111/jdv.19905] [Medline: 38411348]
24. Kuppanda PM, Janda M, Soyer HP, Caffery LJ. What are patients' perceptions and attitudes regarding the use of artificial intelligence in skin cancer screening and diagnosis? Narrative review. J Invest Dermatol 2025 Aug;145(8):1858-1865. [doi: 10.1016/j.jid.2025.01.013] [Medline: 40019459]
25. Mevorach L, Farcomeni A, Pellacani G, Cantisani C. A comparison of skin lesions' diagnoses between AI-based image classification, an expert dermatologist, and a non-expert. Diagnostics (Basel) 2025 Apr 28;15(9):1115. [doi: 10.3390/diagnostics15091115] [Medline: 40361933]
26. Zhelev Z, Puttergill D, Kalsi D, Mullarkey D, Hyde C. BT29 Modelling the cost-effectiveness of an artificial intelligence as a medical device. Br J Dermatol 2024 Jun 28;191(Supplement_1):i202. [doi: 10.1093/bjd/ljae090.426]
27. Giansanti D. The artificial intelligence in teledermatology: a narrative review on opportunities, perspectives, and bottlenecks. Int J Environ Res Public Health 2023 May 12;20(10):5810. [doi: 10.3390/ijerph20105810] [Medline: 37239537]
28. Sangers T, Reeder S, van der Vet S, et al. Validation of a market-approved artificial intelligence mobile health app for skin cancer screening: a prospective multicenter diagnostic accuracy study. Dermatology 2022;238(4):649-656. [doi: 10.1159/000520474] [Medline: 35124665]
29. Greis C, Maul LV, Hsu C, Djamei V, Schmid-Grendelmeier P, Navarini AA. Artificial intelligence to support telemedicine in Africa [Article in German]. Hautarzt 2020 Sep;71(9):686-690. [doi: 10.1007/s00105-020-04664-6] [Medline: 32761386]
30. Smak Gregoor AM, Sangers TE, Bakker LJ, et al. An artificial intelligence based app for skin cancer detection evaluated in a population based setting. NPJ Digit Med 2023 May 20;6(1):90. [doi: 10.1038/s41746-023-00831-w] [Medline: 37210466]
31. Champion VL, Skinner CS. The health belief model. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior and Health Education: Theory, Research, and Practice, 4th edition: Jossey-Bass; 2008:45-65.
32. Chen C, Liao M, Walther JB, Sundar SS. When an AI doctor gets personal: the effects of social and medical individuation in encounters with human and AI doctors. Communic Res 2024 Oct;51(7):747-781. [doi: 10.1177/00936502241263482]
33. Horn S, Litovsky Y, Loewenstein G. Using curiosity to counter health information avoidance. Soc Sci Med 2024 Jan;340:116383. [doi: 10.1016/j.socscimed.2023.116383] [Medline: 38039766]
34. Ozawa S, Sripad P. How do you measure trust in the health system? A systematic review of the literature. Soc Sci Med 2013 Aug;91:10-14. [doi: 10.1016/j.socscimed.2013.05.005] [Medline: 23849233]
35. Martin RW, Head AJ, René J, et al. Patient decision-making related to antirheumatic drugs in rheumatoid arthritis: the importance of patient trust of physician. J Rheumatol 2008 Apr;35(4):618-624. [Medline: 18278840]
36. Wu D, Lowry PB, Zhang D, Tao Y. Patient trust in physicians matters-understanding the role of a mobile patient education system and patient-physician communication in improving patient adherence behavior: field study. J Med Internet Res 2022 Dec 20;24(12):e42941. [doi: 10.2196/42941] [Medline: 36538351]
37. Feng Y, Guan S, Xu Y, et al. Effects of the two-dimensional structure of trust on patient adherence to medication and non-pharmaceutical treatment: a cross-sectional study of rural patients with essential hypertension in China. Front Public Health 2022;10:818426. [doi: 10.3389/fpubh.2022.818426] [Medline: 35309202]
38. Riedl R, Hogeterp SA, Reuter M. Do patients prefer a human doctor, artificial intelligence, or a blend, and is this preference dependent on medical discipline? Empirical evidence and implications for medical practice. Front Psychol 2024;15:1422177. [doi: 10.3389/fpsyg.2024.1422177] [Medline: 39188871]

XSL•FO
RenderX

39. Wang RH, Barbieri JS, Nguyen HP, et al. Clinical effectiveness and cost-effectiveness of teledermatology: where are we now, and what are the barriers to adoption? J Am Acad Dermatol 2020 Jul;83(1):299-307. [doi: 10.1016/j.jaad.2020.01.065] [Medline: 32035106]
40. Evans SC, Roberts MC, Keeley JW, et al. Vignette methodologies for studying clinicians' decision-making: validity, utility, and application in ICD-11 field studies. Int J Clin Health Psychol 2015;15(2):160-170. [doi: 10.1016/j.ijchp.2014.12.001] [Medline: 30487833]
41. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
42. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition: Lawrence Erlbaum Associates; 1988.
43. Digitale patiëntpanel [Article in Dutch]. Erasmus MC. URL: https://www.erasmusmc.nl/nl-nl/uw-mening-telt/patientpanel [accessed 2026-10-27]
44. Clinical trials. ICMJE (International Committee of Medical Journal Editors). 2026. URL: https://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html [accessed 2026-10-27]
45. Swain R, Zwaan L, Masic M. Influence of teledermatologist supervision on patient trust and adherence in an AI-driven dermatology app: a randomised-controlled experiment. OSF (Open Science Framework). 2025. URL: https://osf.io/64pjn/overview [accessed 2026-10-27]
46. Grassini S. Development and validation of the AI Attitude Scale (AIAS-4): a brief measure of general attitude toward artificial intelligence. Front Psychol 2023;14:1191628. [doi: 10.3389/fpsyg.2023.1191628] [Medline: 37554139]
47. Castro MS, Bahli B, Ferreira JJ, Figueiredo R. Comparing single-item and multi-item trust scales: insights for assessing trust in project leaders. Behav Sci (Basel) 2023 Sep 21;13(9):786. [doi: 10.3390/bs13090786] [Medline: 37754064]
48. Christensen RHB. Cumulative link models for ordinal regression with the r package ordinal. R Project. 2018. URL: https://cloud.r-project.org/web/packages/ordinal/vignettes/clm_article.pdf [accessed 2026-11-12]
49. Hedeker D, Gibbons RD. A random-effects ordinal regression model for multilevel analysis. Biometrics 1994 Dec;50(4):933-944. [Medline: 7787006]
50. Hopewell S, Chan AW, Collins GS, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. Lancet 2025 Apr 14;405(10489):1633-1640. [doi: 10.1016/S0140-6736(25)00672-5] [Medline: 40245901]
51. Hopewell S, Clarke M, Moher D, et al. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. PLoS Med 2008 Jan 22;5(1):e20. [doi: 10.1371/journal.pmed.0050020] [Medline: 18215107]
52. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
53. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014 Mar 7;348:g1687. [doi: 10.1136/bmj.g1687] [Medline: 24609605]
54. Oberg AL, Mahoney DW. Linear mixed effects models. Methods Mol Biol 2007;404:213-234. [doi: 10.1007/978-1-59745-530-5_11] [Medline: 18450052]
55. Rogers RW. A protection motivation theory of fear appeals and attitude change1. J Psychol 1975 Sep;91(1):93-114. [doi: 10.1080/00223980.1975.9915803] [Medline: 28136248]
56. Soellner M, Koenigstorfer J. Compliance with medical recommendations depending on the use of artificial intelligence as a diagnostic method. BMC Med Inform Decis Mak 2021 Aug 6;21(1):236. [doi: 10.1186/s12911-021-01596-6] [Medline: 34362359]
57. Reis M, Reis F, Kunde W. Influence of believed AI involvement on the perception of digital medical advice. Nat Med 2024 Nov;30(11):3098-3100. [doi: 10.1038/s41591-024-03180-7] [Medline: 39054373]
58. Wu A, Ngo M, Thomas C. Assessment of patient perceptions of artificial intelligence use in dermatology: a cross-sectional survey. Skin Res Technol 2024 Mar;30(3):e13656. [doi: 10.1111/srt.13656] [Medline: 38481072]
59. Jones OT, Calanzani N, Scott SE, Matin RN, Emery J, Walter FM. User and developer views on using AI technologies to facilitate the early detection of skin cancers in primary care settings: Qualitative semistructured interview study. JMIR Cancer 2025 Jan 28;11:e60653. [doi: 10.2196/60653] [Medline: 39874580]
60. Peabody JW, Luck J, Glassman P, Dresselhaus TR, Lee M. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. JAMA 2000 Apr 5;283(13):1715-1722. [doi: 10.1001/jama.283.13.1715] [Medline: 10755498]
61. Hainmueller J, Hangartner D, Yamamoto T. Validating vignette and conjoint survey experiments against real-world behavior. Proc Natl Acad Sci U S A 2015 Feb 24;112(8):2395-2400. [doi: 10.1073/pnas.1416587112] [Medline: 25646415]
62. Stone-Romero EF. The relative validity and usefulness of various empirical research designs. In: Rogelberg SG, editor. Handbook of Research Methods in Industrial and Organizational Psychology: Blackwell; 2002:77-98. [doi: 10.1002/9780470756669]
63. Thom DH, Wong ST, Guzman D, et al. Physician trust in the patient: development and validation of a new measure. Ann Fam Med 2011;9(2):148-154. [doi: 10.1370/afm.1224] [Medline: 21403142]

XSL•FO
RenderX

Abbreviations

AIAS-4: 4-item AI Attitude Scale
CLMM: cumulative link mixed model
CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
LMM: linear mixed-effect model
OR: odds ratio
TIDieR: Template for Intervention Description and Replication

Edited by S Brini; submitted 23.Feb.2026; peer-reviewed by A Hartanto, S Kakraba, S McRoy; revised version received 26.Aug.2026; accepted 01.Sep.2026; published 17.Nov.2026.

Please cite as:
Swain R, Masic M, Staal J, Dangerman N, Barendse E, Wakkee M, Zwaan L
Effect of Teledermatologist Supervision in AI Skin Cancer Apps on Patients' Trust and Adherence Intentions: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e93991
URL: https://mhealth.jmir.org/2026/1/e93991
doi:10.2196/93991

© Ritwik Swain, Mihaela Masic, Justine Staal, Natalie Dangerman, Elise Barendse, Marlies Wakkee, Laura Zwaan. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 17.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Insights on Postimplementation and Delivery of Telemedicine in Remote Areas of Rajasthan: Qualitative Study

Ramesh Kumar Huda[1*], PhD; Rahul Singh Chowhan[1*], ME; Sumit Kalra[2], PhD; Ramesh Kumar Sangwan[1], PhD; Mukti Khetan[1], PhD

[1]National Institute for Implementation Research on Non-Communicable Diseases, New Pali Road, Bhagat ki Kothi, Jodhpur, Rajasthan, India

[2]Indian Institute of Technology Jodhpur, Jodhpur, Rajasthan, India

[*]these authors contributed equally

Corresponding Author:
Ramesh Kumar Huda, PhD
National Institute for Implementation Research on Non-Communicable Diseases, New Pali Road, Bhagat ki Kothi, Jodhpur, Rajasthan, India

Abstract

Background: Telemedicine has emerged as a critical strategy to address health care access gaps in hard-to-reach areas, particularly in resource-constrained settings. While telemedicine services have expanded across India, evidence on post-implementation experiences, specifically regarding the integration of portable diagnostic tools, and the assisting role of frontline health care workers, remains limited. Understanding these sociotechnical dynamics is essential for optimizing digital health interventions in remote areas.

Objective: This study aimed to explore perceptions, experiences, and implementation processes of an assisted telemedicine model where frontline health care workers facilitate real-time, doorstep patient-assisted-to-doctor consultations using portable diagnostics and a customized portal among health care workers and patients in hard-to-reach villages of Rajasthan, India.

Methods: A qualitative descriptive study was conducted in 5 remote villages of Balotra district, Rajasthan, between December 2025 and January 2026, following a 10-month telemedicine intervention. In-depth, semistructured interviews were conducted with 28 participants, including medical officers (n=3, 10.71%), nurses, and accredited social health activists or auxiliary nurse midwives (n=10, 35.71%), and patients (n=15, 53.57%). Cadre-specific semistructured interview guides were developed using the Consolidated Framework for Implementation Research and Technology Acceptance Model. Interviews were analyzed using a reflexive thematic analysis approach, incorporating both inductive and deductive coding approaches in NVivo 15 (Lumivero). Triangulation across participant groups and analytic perspectives was used to enhance credibility and trustworthiness.

Results: The analysis revealed that the assisted telemedicine model fundamentally reconfigured treatment-seeking behavior by providing doorstep care, significantly reducing travel burdens, and out-of-pocket expenditures. Frontline health care workers acted as crucial sociotechnical mediators, bridging the digital divide and fostering patient trust, particularly among women and older adults. Health care providers reported increased clinical confidence and workflow efficiency, supported by real-time vital monitoring and cross-verification via digital chatbots. However, implementation was challenged by intermittent network connectivity, diagnostic limitations for conditions requiring physical palpation, and occasional stock-outs of essential medicines. Despite these barriers, the gradual acceptance of the system was heavily facilitated by the continuous engagement and cultural mediation of local health workers.

Conclusions: Telemedicine was perceived as a feasible and acceptable model for extending primary health care services in hard-to-reach areas, provided it is embedded within an assisted model supported by trained frontline workers and responsive technical systems. The findings emphasize the need to shift health care policy toward mediated-digital models, formally recognizing community health workers as digital health navigators. Future scale-up must prioritize strengthening rural digital infrastructure and ensuring a reliable supply chain for essential medicines to sustain long-term engagement and health equity.

Trial Registration: Clinical Trials Registry—India (CTRI) CTRI/2024/10/076070; https://tinyurl.com/yr952847

International Registered Report Identifier (IRRID): RR2-10.2196/92996

(JMIR Mhealth Uhealth 2026;14:e93146) doi:10.2196/93146

KEYWORDS

medical informatics; telemedicine; implementation research; patient perspectives; digital health; rural health care

XSL•FO
RenderX

Introduction

Telemedicine has emerged as a critical tool for bridging health care access gaps, particularly in hard-to-reach areas with limited medical resources. Over the past two decades, it has evolved from experimental pilot projects into a core component of modern health delivery, extending specialist and primary care to remote populations [1]. This technology uses information and communication technologies to deliver health care services remotely, thereby enhancing accessibility and improving patient outcomes in remote areas [2]. It significantly reduces patient transportation costs and improves the efficiency of medical services in areas with scarce medical resources [3]. Nevertheless, the adoption of telemedicine worldwide remains uneven, with infrastructural, organizational, and regulatory barriers consistently limiting its scale-up in resource-constrained settings [4]. Telemedicine can lower overall health care costs by streamlining service delivery and minimizing the need for physical consultations. Moreover, it can empower individuals in these regions, who often face greater challenges in accessing health care, by improving inclusivity and participation in health care services [5,6].

The integration of telemedicine is especially pertinent in lower-middle-income countries such as India, where vast populations and significant geographical distances between urban centers and rural towns create substantial barriers to accessing specialized health care services. In addition, especially in regions such as Rajasthan, where geographical barriers and a sparse distribution of health care facilities often hinder timely access to essential medical care, telemedicine can promote equitable health care by extending specialist consultations and ongoing care to remote populations. Moreover, telemedicine implementation in Health and Wellness Centers facilitates consultation, screening, and remote monitoring, enabling individuals to connect with health care providers and receive medical advice. This is particularly beneficial for patients with chronic conditions, enabling continuous monitoring and timely interventions without extensive travel [7,8].

To contextualize this implementation, India's national telemedicine initiatives, such as the eSanjeevani platform (developed by the Centre for Development of Advanced Computing, Ministry of Health and Family Welfare, Government of India), have largely focused on facility-based models (eg, doctor-to-doctor consultations at Health and Wellness Centres or subcenters) to bridge the urban-rural health care divide [9]. While these facility-based models address specialist shortages, they still require patients to travel to a local health facility, which remains a barrier in geographically isolated terrains. The intervention evaluated in this study represents a distinct piloting and evidence generation stage of an adapted mobile health model [8]. Specifically, it introduces an assisted telemedicine model that extends care beyond the subcenter level directly to the patient's doorstep. In this workflow, frontline health care workers (accredited social health activists [ASHAs], auxiliary nurse midwives [ANMs], and nurses) conduct home visits or respond to patient calls, capture real-time vital parameters using portable digital diagnostic tools, and upload the data to a customized telemedicine portal. The health worker then facilitates a live video consultation between the patient and a remote medical officer, who generates an e-prescription that the health worker downloads and delivers to the patient. The added value of this specific intervention lies in its mobility and real-time diagnostic capability, effectively overcoming the geographical barriers inherent in standard facility-based models by bringing the digital health hub directly to the patient's home [10,11]. This implementation-focused perspective is consistent with iCHECK-DH (Guidelines and Checklist for the Reporting on Digital Health Implementations; Checklist 1), which emphasizes documenting the contextual factors, implementation processes, and adaptations that shape digital health delivery in real-world settings [12].

This study aimed to address the gap by presenting in-depth qualitative findings from Balotra district of Rajasthan, which marks its geographical remoteness and limited health infrastructure on the post-implementation delivery of telemedicine. The insights on perception of acceptability, operational challenges, and actionable recommendations have been drawn from interviews with ASHAs, ANMs, nursing officers (NOs), medical officers (MOs), and patients who used the telemedicine facilities.

Methods

Study Design

A qualitative descriptive study design was used to understand the experiences and perceptions of frontline health care workers and patients involved in the implementation and delivery of telemedicine services in hard-to-reach areas of Balotra district of Rajasthan. The study used in-depth interviews (IDIs) to document lived experiences, perceived benefits, implementation challenges, acceptability, and recommendations for improvement among those who used telemedicine services during the intervention group's implementation phase. This manuscript is reported in accordance with the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist (Checklist 2). Additionally, the reporting of the digital health implementation context, intervention characteristics, and structural barriers was guided by the iCHECK-DH mentioned in Checklist 1 [12].

Study Setting

The study was conducted in 5 randomly chosen hard-to-reach areas of the Balotra district of Rajasthan between December 2025 and January 2026, where health care is severely constrained due to a lack of proper health care accessibility, geographical isolation, limited access to specialist care, a health care worker shortage, and minimal existing health care infrastructure. These areas were facilitated with telemedicine services, in which teleconsultations were operational from 8 AM to 5 PM for 5 working days, supported by trained frontline health care workers.

Study Context

As part of the telemedicine intervention (parent project), a team of frontline health care workers, ASHAs or ANMs, and nurses were trained to operate digital diagnostic devices and a customized telemedicine portal and were deployed with medical kits and smartphones to facilitate teleconsultations with doctors.

XSL•FO
RenderX

Five teams were assigned to the 5 randomly selected villages, and their contact information was shared across the community through an initial door-to-door survey. During operational hours, frontline workers visited patients at their doorsteps or patients visited them, measured vital parameters such as blood pressure (BP), oxygen saturation (SpO_2), and heart rate, conducted electrocardiography (ECG) if required, and provided real-time video consultations with the MOs sitting remotely. The MOs used patients' data and interactions with patients to provide medical advice, e-prescriptions, and referrals, as needed. The telemedicine portal was developed on the MERN stack (MongoDB, Express, React, and Node.js) with WebRTC integration to allow secure transmission and storage of patient data, live e-prescription generation, and follow-up management.

Study Participants

The study included 28 participants from multiple stakeholder groups who were directly involved in or had experience with telemedicine services during the intervention period, enabling a comprehensive understanding of telemedicine implementation and delivery in hard-to-reach areas. Purposive sampling was used to recruit both frontline health care workers and patients to ensure representation of diverse experiences and perspectives. Frontline health care workers (n=13) were purposively recruited through their respective primary health centers and project headquarters and comprised 5 (38.5%) ASHAs or ANMs, 5 (38.5%) nurses, and 3 (23.1%) MOs. Patients were purposively selected from the telemedicine intervention registry (n=4401) with assistance from local ASHA workers, who identified individuals with varied demographic characteristics, frequencies of telemedicine use, health conditions, and levels of engagement with the system. Fifteen patients, representing both genders and different age groups, were selected, with 3 participants recruited from each of the 5 intervention villages. This purposive selection ensured variation in participants' experiences, symptoms, and patterns of telemedicine use, thereby providing diverse perspectives on the implementation and use of telemedicine services.

Data Collection

Qualitative data were collected through face-to-face IDIs conducted post-implementation of the intervention project for 10 months within the intervention group villages. RS Chowhan, RKH, and project staff, holding ME, PhD, and MHA, respectively, conducted the IDIs. All received training in qualitative research methods, including thematic analysis and conducting IDIs, at Indian Council of Medical Research (ICMR)-National Institute for Implementation Research on Non-Communicable Diseases (NIIRNCD), Jodhpur. Separate semistructured interview guides were developed for ASHAs, ANMs, or nurses, MOs, and patients based on the Consolidated Framework for Implementation Research (CFIR) and technology acceptance model (TAM) frameworks after a literature review and through multiple rounds of discussion with the research team and domain experts to ensure questionnaire alignment with study objectives and to reflect differences in roles, experiences, and interaction with the telemedicine system.

Trained qualitative researchers with no prior relationship with participants conducted interviews in the local language (Hindi or Marwari). Each interview lasted approximately 30 to 35 minutes and was audio recorded with written consent in a quiet, private space, ensuring participants' comfort, anonymity, and confidentiality. Semistructured interview guides for patients, MOs, and nurses were used to facilitate data collection (Multimedia Appendix 1).

Data Analysis

The study was grounded in a critical realist epistemological perspective, acknowledging the objective realities of the rural health care infrastructure while recognizing that participants' experiences and perceptions of the telemedicine system are socially constructed and context dependent. The audio recordings were transcribed into English by bilingual researchers and verified by the principal investigator for consistency, accuracy, and data integrity, without distorting participant intent throughout the translation process. Transcribed files and audio recordings were securely stored in password-protected folders. Data saturation was achieved when no new themes were obtained. The thematic analysis incorporated both deductive (literature-driven themes) and inductive (field-driven themes) coding approaches as described by Braun and Clarke using NVivo 15 software (Lumivero) [13]. This study used an integrated theoretical approach combining the CFIR and the TAM. CFIR guided the deductive analysis of multilevel implementation determinants, while TAM served as an interpretive framework to explore individual perceptions of telemedicine, including ease of use, usefulness, and intention to use.

Thematic saturation was assessed concurrently during data collection and analysis by the qualitative research team (RS Chowhan and RK). Saturation was assessed independently for each of the 3 stakeholder groups (MOs, frontline health care workers, and patients) to ensure diverse perspectives were adequately captured. Regarding the MOs (n=3, 10.71%), the intervention specifically deployed exactly 3 (10.71%) MOs to cover the 5 remote villages. Therefore, n=3 represents a complete census of the MO cadre involved in this specific pilot, not a sample. Their interviews were extensive (30 - 40 min), and thematic saturation for their specific operational context was achieved within this bounded group. For frontline health care workers (n=10) and patients (n=15), recruitment continued until no new themes or codes emerged from successive interviews.

Triangulation was applied during data analysis to enhance credibility. Data source triangulation was achieved by comparing perspectives across MOs, frontline health care workers, and patients involved in the telemedicine intervention. Investigator triangulation was ensured through the involvement of multiple trained researchers in data collection and coding, with iterative discussions to reach consensus. Theoretical triangulation was further applied by integrating CFIR to examine implementation contexts and processes, alongside TAM to interpret individual perceptions of telemedicine use. Rigor and trustworthiness criteria, as described by Lincoln and Guba [14], were applied to enhance credibility, dependability, confirmability, and contextual transferability through data saturation, systematic

XSL•FO
RenderX

coding, the use of verbatim quotations, and detailed descriptions of the study setting and participants.

In alignment with the COREQ checklist, the research team maintained reflexive journals during data collection and analysis to document personal assumptions and biases regarding digital health interventions. Field notes were taken during and immediately after each interview to capture nonverbal cues, environmental context, and initial analytical memos. There were no refusals or dropouts among the eligible participants approached for the study (0% nonparticipation rate). Owing to the low literacy levels among some patient participants and logistical constraints in remote field settings, transcripts were not returned to participants for formal member checking; however, preliminary findings were shared and discussed with local community health leaders to ensure contextual accuracy.

A detailed codebook outlining codes, categories, and thematic hierarchies for patients, MOs, and ASHAs or ANMs and nurses is provided in Multimedia Appendix 2.

Ethical Considerations

This study is part of the telemedicine intervention project titled "Acceptance of Digital Medical Tools and Telemedicine in Hard-to-Reach Areas as Augmentation of Conventional Physical Clinics and Face-to-Face Consultation," funded under grant IIRP-2023 - 0000151. Ethical approval was obtained from the Institutional Ethics Committee of ICMR-NIIRNCD, Jodhpur, India (approval IEC-NIIRNCD/2023/DEC/FR/013; approved on December 22, 2023). The study was prospectively registered with the Clinical Trials Registry-India (CTRI/2024/10/076070). The study was conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants prior to data collection. Participants were assured confidentiality and anonymity, and all identifying information was removed from transcripts before analysis. Participants did not receive financial compensation for their participation in the study; however, standard refreshments were offered during the interview sessions as a gesture of appreciation for their time.

Results

Overview

A total of 28 participants were included in the study, comprising 3 (10.7%) MOs, 10 (35.7%) nurses and ASHAs or ANMs, and 15 (53.6%) patients (3 from each of 5 villages). Data were collected through cadre-specific semistructured interviews, which were audio recorded and transcribed verbatim. Recruitment and data collection continued until thematic saturation was achieved within each participant group. Findings are presented cadre wise according to 3 participant categories. Participant characteristics and thematic tables are mentioned in Multimedia Appendix 3, and the COREQ checklist is presented in Checklist 2. An overview of key themes emerging from each cadre's interviews is illustrated in Figures 1-3. The conceptual framework has been added to showcase the perspectives integrated across stakeholders in Figure 4. The final participant composition and characteristics are presented in Table 1.

XSL•FO
RenderX

Figure 1. Thematic diagram of key themes and subthemes derived from interviews with medical officers.



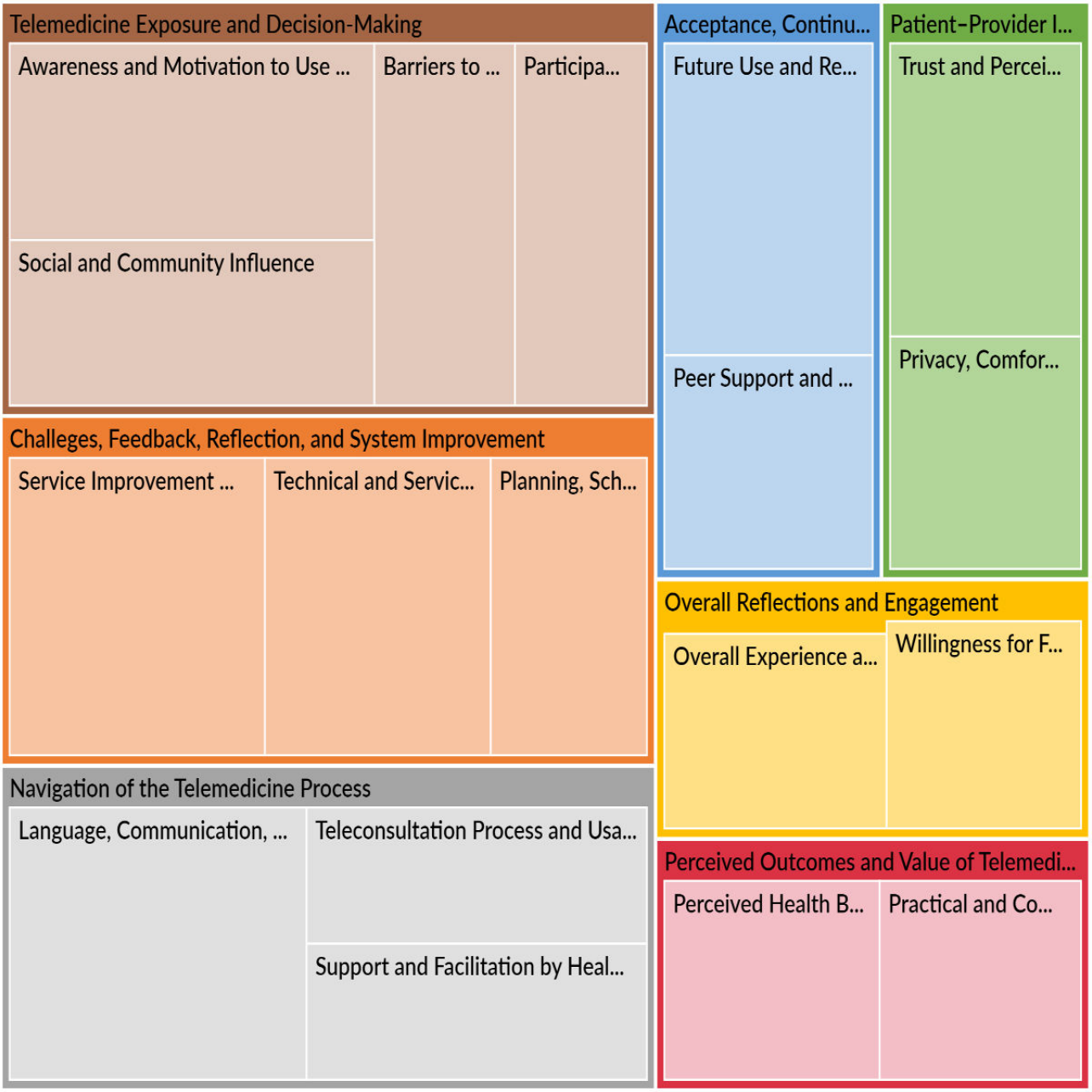
XSL•FO
RenderX

Figure 2. Thematic diagram illustrating key themes and subthemes derived from interviews with accredited social health activists or auxiliary nurse midwife and nursing officers.



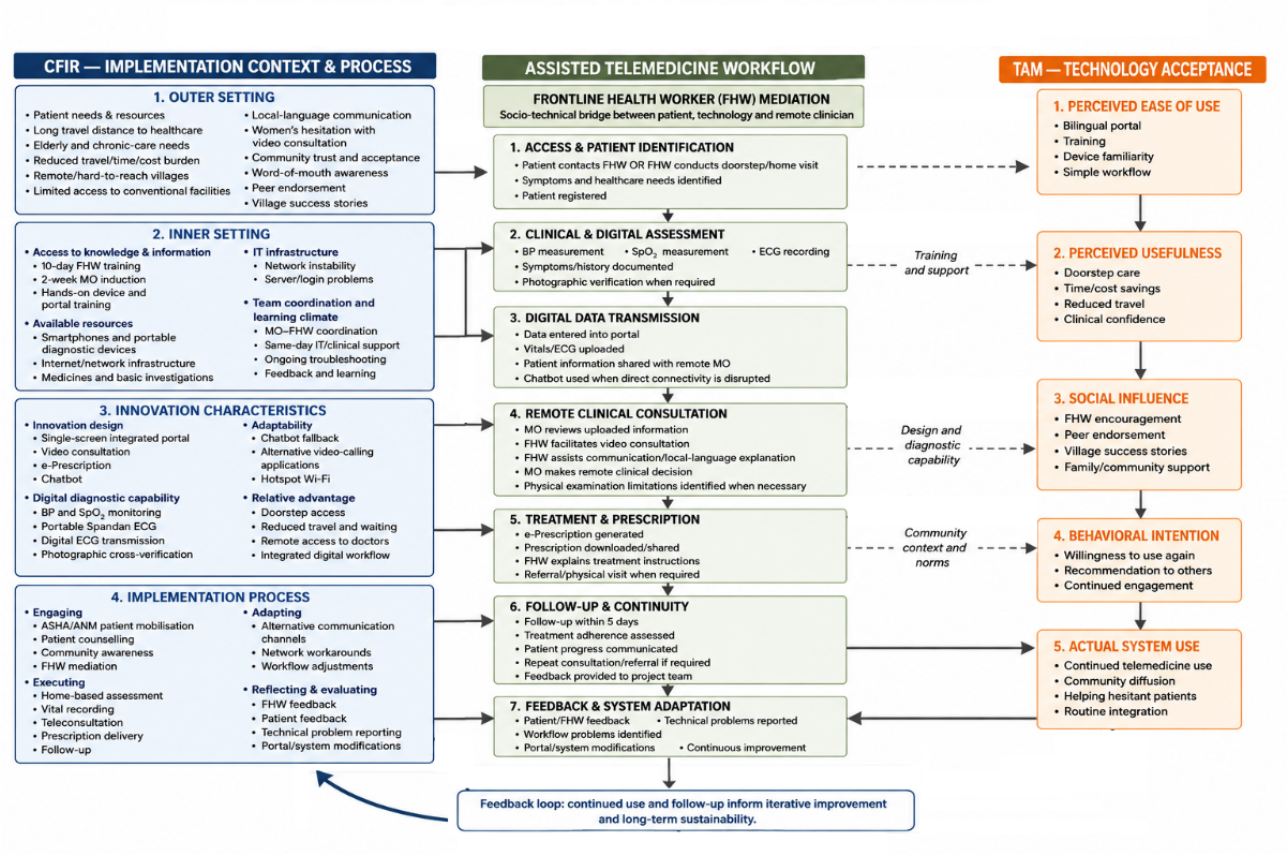
XSL•FO
RenderX

Figure 3. Thematic diagram illustrating key themes and subthemes derived from interviews with patients as beneficiaries of telemedicine intervention during the study period.



XSL•FO
RenderX

Figure 4. Conceptual framework of Consolidated Framework for Implementation Research and technology acceptance model in assisted telemedicine implementation and acceptance. ANM: auxiliary nurse midwife; ASHA: accredited social health activist; ECG: electrocardiogram; FHW: frontline health care worker; MO: medical officer; SpO_2: oxygen saturation.



XSL•FO
RenderX

Table . Participant characteristics by cadre (N=28).

Characteristics	Medical officers (n=3)	Nurses (n=5)	ASHA[a] or ANM[b] (n=5)	Patients (n=15)	Total (n=28)
Age (years), mean (SD)	30.3 (2.1)	25.2 (1.5)	38.4 (7.3)	39.1 (12.8)	35.6 (12.2)
Sex, n (%)					
Male	3 (100)	3 (60)	0 (0)	8 (53.3)	14 (50)
Female	0 (0)	2 (40)	5 (100)	7 (46.7)	14 (50)
Religion, n (%)					
Hindu	3 (100)	5 (100)	5 (100)	14 (93.3)	27 (96.4)
Muslim	0 (0)	0 (0)	0 (0)	1 (6.7)	1 (3.6)
Marital status, n (%)					
Married	2 (66.7)	3 (60)	5 (100)	10 (66.7)	20 (71.4)
Unmarried	1 (33.3)	2 (40)	0 (0)	5 (33.3)	8 (28.6)
Languages spoken, n (%)[c]					
Marwari	3 (100)	5 (100)	5 (100)	15 (100)	28 (100)
Hindi	3 (100)	5 (100)	5 (100)	12 (80)	25 (89.3)
English	3 (100)	0 (0)	0 (0)	0 (0)	3 (10.7)
Education level, n (%)					
Uneducated	0 (0)	0 (0)	0 (0)	4 (26.7)	4 (14.3)
Primary (≤5th standard)	0 (0)	0 (0)	1 (20)	5 (33.3)	6 (21.4)
Secondary (10th-12th)	0 (0)	0 (0)	3 (60)	3 (20)	6 (21.4)
Graduate and above	0 (0)	3 (60)	1 (20)	3 (20)	7 (25)
Professional degree	3 (100)	2 (40)	0 (0)	0 (0)	5 (17.9)
Monthly income (INR), n (%)[d]	__[e]	—	—	—	15 (100)
<INR 5000 (US $52.21).	—	—	—	8 (53.3)	8 (53.3)
INR 5000 (US $52.21) to INR 10,000 (US $104.42)	—	—	—	1 (6.7)	1 (6.7)
INR 10,001 (US $104.42) to INR 20,000 (US $208.85)	—	—	—	5 (33.3)	5 (33.3)
>INR 20,000 (US $208.85)	—	—	—	1 (6.7)	1 (6.7)
Family type, n (%)[d]	—	—	—	—	15 (100)
Nuclear	—	—	—	3 (20)	3 (20)
Joint	—	—	—	12 (80)	12 (80)
Distance from health care center (km), n (%)[d]	—	—	—	—	15 (100)
<5	—	—	—	6 (40)	6 (40)
5 - 10	—	—	—	4 (26.7)	4 (26.7)
10 - 20	—	—	—	3 (20)	3 (20)
>20	—	—	—	2 (13.3)	2 (13.3)

XSL•FO
RenderX

Characteristics	Medical officers (n=3)	Nurses (n=5)	ASHA[a] or ANM[b] (n=5)	Patients (n=15)	Total (n=28)
Years of professional experience, n (%)[f]					
≤5	3 (100)	5 (100)	1 (20)	—	9 (69.2)
>5	0 (0)	0 (0)	4 (80)	—	4 (30.7)

[a]ASHA: accredited social health activist.

[b]ANM: auxiliary nurse midwife.

[c]Participants could speak more than 1 language; percentages for languages are calculated based on the total sample (n=28).

[d]Monthly income, family type, and distance from health care center are calculated only for patients (n=15).

[e]Not applicable.

[f]Professional experience is calculated only for health care workers (n=13), excluding patients.

MOs' Insights

Theme 1: Roles and Responsibilities—From Individual Readiness to Routine Telemedicine Practice

MOs described telemedicine adoption as a progressive process shaped by prior clinical experience, digital exposure, and evolving professional responsibilities. Their accounts reflected the CFIR domain of Individuals—Capability, as MOs drew on previous clinical training and project experience to operationalize remote care within their routine practice. One MO described his professional pathway and current responsibility within the telemedicine model:

> *I have been working as a medical officer in the Telemedicine Project. Before this, I did my internship at my medical college, and after that, I joined here as a Medical Officer. My responsibility is to provide teleconsultation & I have 2 years of experience in the same project.* [MO02]

Prior familiarity with digital health platforms also influenced the ease with which MOs transitioned to virtual consultations. In terms of the TAM construct of perceived ease of use, previous exposure to existing telemedicine systems reduced the learning burden associated with the new portal. One MO explained:

> *I have used portals like e-Sanjeevani portals before this telemedicine project, so It has been comfortable enough for me to upload data, make video calls.* [MO03]

At the same time, the introduction of new portable diagnostic devices required adaptation. This again reflected TAM's perceived ease of use, as MOs compared earlier clinical tools with the simplified but unfamiliar portable equipment used in the project. One MO stated:

> *Before the project, pulse oximeter, manual BP apparatus, and ECG machine of nine leads were part of my workflow, while in this project portable Spandan ECG machine with three leads has been a new experience for me.* [MO03]

Together, these accounts suggest that MOs' readiness for telemedicine was not uniform but was shaped by cumulative clinical experience, prior digital exposure, and gradual adaptation to new diagnostic and consultation workflows.

Theme 2: Training and System Onboarding—Building Confidence Through Structured Support

Structured training and ongoing system-level support were central to MOs' confidence in delivering telemedicine. This aligned with the CFIR Inner Setting domain, particularly Access to Knowledge and Information, because formal induction training provided the knowledge required to use the portal, conduct video consultations, and generate e-prescriptions. One MO described the initial training process:

> *Two weeks of induction training were provided on how to use the telemedicine portal, how to connect with nurses and patients through video calls, and how to generate e-prescriptions.* [MO01]

In addition to formal training, MOs emphasized the importance of real-time collaboration with nursing and field staff. This reflected the CFIR construct of Available Resources within the Inner Setting because the presence of trained support staff during consultations made remote clinical workflows more manageable. One MO explained:

> *The nursing staff also assisted in positioning the patient and sharing vital signs during the video call. This coordination made teleconsultation easier.* [MO03]

MOs also reported that continuous technical and clinical support reduced early hesitation and enabled them to use the system more efficiently over time. This again highlighted the CFIR Inner Setting—Available Resources domain, as accessible support mechanisms helped convert initial uncertainty into routine use. One MO stated:

> *Initially, we were not very confident about video consultations, but because technical and clinical support was always available, our hesitation reduced. Over time, it became more efficient and comfortable to use.* [MO02]

These findings indicate that MOs' adoption of telemedicine was supported not only by individual capability but also by structured onboarding, team-based assistance, and responsive technical support.

XSL•FO
RenderX

Theme 3: System Use and Workflow Integration—Platform Design and Adaptive Communication

Once MOs became familiar with the telemedicine system, platform design and adaptive communication strategies played an important role in integrating teleconsultations into routine clinical workflows. The centralized design of the portal was appreciated because it brought multiple consultation-related functions into a single interface. This reflected the CFIR construct of Innovation—Innovation Design, as a well-organized system reduced clinical friction and supported efficient decision-making. One MO explained:

> *The telemedicine portal was well-designed and user-friendly. All patient details, vitals, doctor notes, video consultations, and e-prescriptions were available on a single screen, which reduced errors and saved time.* [MO02]

However, network instability sometimes disrupted communication during consultations. MOs described using alternative communication mechanisms, such as the chatbot, when symptoms could not be clearly conveyed through video. This reflected the CFIR domain of Implementation Process—Tailoring Strategies because clinicians and field teams adapted the consultation process according to local connectivity constraints. One MO stated:

> *There were network issues at patients' side, because of which at times the symptoms were not clearly stated and understood, and then with the help of a chatbot, it was communicated.* [MO03]

Technical difficulties such as server downtime and login problems were also reported, although MOs noted that these issues were generally resolved quickly. This again highlighted the importance of the CFIR Inner Setting—Information Technology Infrastructure, as the reliability of technical support influenced whether workflow disruptions became minor interruptions or major barriers to care. One MO explained:

> *We faced server downtime, login issues due to poor network connectivity, and occasional portal crashes. However, these technical problems were resolved promptly.* [MO01]

These accounts suggest that workflow continuity depended not only on the usability of the digital platform but also on the availability of fallback mechanisms and responsive technical support when connectivity or system problems occurred.

Theme 4: Clinical Decision-Making and Diagnostic Confidence—Remote Evidence and Clinical Boundaries

MOs reported that digital recording of vital signs and photographic verification strengthened their confidence in remotely obtained clinical information. This aligned with the CFIR construct of Innovation—Innovation Evidence Base because the ability to review entered vitals and cross-check photographic evidence enhanced trust in the quality of remote clinical data. One MO described the verification process:

> *The field staff would send the data by typing in the vital signs and uploading photos of the digital devices on the chatbot, allowing us to cross-verify the data. If there was any issue with the readings or if the patient's symptoms differed from those recorded, we would ask to recheck.* [MO02]

At the same time, MOs recognized that telemedicine had clinical boundaries. Although remote consultations were sufficient for many routine conditions, certain clinical situations still required physical examination. This reflected the CFIR domain of Individuals—Capability, as MOs exercised professional judgment about when teleconsultation was appropriate and when in-person care was necessary. One MO stated:

> *I was confident to treat patients as per the telemedicine guidelines, but certain cases of dermatological conditions, emergencies, and ANC visits needed in-person examination.* [MO01]

Thus, telemedicine was perceived as a valuable model for routine primary care delivery, but MOs maintained clear clinical boundaries and referred patients to face-to-face examination when physical assessment was essential.

Theme 5: Communication and Trust—Frontline Mediation and Continuity of Care

MOs described patient acceptance as developing through continued communication, follow-up, and positive experiences with the service. This reflected the CFIR Inner Setting—Communication domain because sustained interaction between patients, frontline workers, and remote clinicians helped build trust in the telemedicine model. One MO reported:

> *Patients reported high satisfaction with teleconsultation services and maintained regular contact with ASHA and ANM. Many continued follow-ups for chronic conditions and sought timely advice for acute illnesses.* [MO01]

Communication during teleconsultations was also shaped by local language use and cultural expectations. This aligned with the CFIR Outer Setting—Local Attitudes domain because rural patients' comfort with video-based care was influenced by language, gender norms, and familiarity with digital interaction. One MO explained:

> *People from rural areas use local language, and a few patients, especially women, were shy or hesitant to speak on video.* [MO03]

Continued information exchange between MOs and field workers supported follow-up and allowed clinicians to monitor patient progress. This reflected the CFIR construct of Innovation—Innovation Recipient Impact, as the telemedicine model influenced not only individual consultations but also ongoing care relationships. One MO stated:

> *We took follow-ups when patients revisited for check-ups. Sometimes, field workers also updated us about the patient's progress.* [MO02]

These findings indicate that acceptance of telemedicine was supported not simply by the availability of remote consultation but by sustained interpersonal mediation and continuity between remote clinicians and community-based health workers.

XSL•FO
RenderX

Theme 6: Collaboration and Coordination—Team Learning and Feedback-Driven Improvement

MOs emphasized that telemedicine delivery depended on coordination among clinical, nursing, technical, and community teams. This reflected the CFIR Inner Setting—Learning Centeredness domain because collective preparation and shared understanding were important before the service was deployed in the field. One MO described this process:

> *Before the field deployment of telemedicine services, we met, trained and explained he project details to all the ASHA, ANM, nursing staff, and peers.* [MO02]

Feedback from clinical staff also contributed to iterative improvements in the telemedicine workflow. This aligned with the CFIR domain of Implementation Process—Reflecting and Evaluating because monthly review and practical feedback led to modifications that simplified the consultation process. One MO explained:

> *Earlier, workflow required multiple systems for teleconsultations, but monthly feedback led to integration of video calls, chatbot, and e-prescription generation on a single screen.* [MO02]

These accounts show that implementation was not a one-time deployment but an evolving process in which team coordination, training, and continuous feedback shaped the usability and effectiveness of the telemedicine system.

Theme 7: Sustainability, Integration, and Reflective Evaluation—Adaptation, Adoption, and Long-Term Value

MOs described several practical adaptations that helped sustain telemedicine delivery when digital infrastructure was unstable. This reflected the CFIR domain of Implementation Process—Adapting because clinicians and field teams developed workarounds to maintain service continuity. One MO stated:

> *Network connectivity sometimes led to video call disruptions and portal lag, so we shifted to alternative modes such as video call through a hotspot Wi-Fi.* [MO03]

MOs also viewed telemedicine as a valuable long-term addition to primary health care delivery. This aligned with the CFIR Inner Setting—Mission Alignment domain because telemedicine was seen as complementary to existing services and consistent with broader goals of improving access in remote areas. One MO explained:

> *Telemedicine is a valuable long-term support tool with improvements in features such as chatbots and video calls to benefit patients in real-time.* [MO01]

The commitment of MOs extended beyond routine working hours in some situations, reflecting both individual motivation and the adoptability of the model. This aligned with the CFIR construct of Implementation Outcome—Adoptability, as clinicians were willing to use personal devices when necessary to ensure continuity of care. One MO stated:

> *Consultations were not done on weekends, but if needed, we used phones for consultations from our homes.* [MO03]

MOs also acted as internal champions for telemedicine by discussing complex cases with colleagues and sharing positive experiences with peers. This again supported the CFIR domain of Implementation Outcome—Adoptability because professional endorsement and peer communication can strengthen broader acceptance of digital health innovations. One MO explained:

> *We discussed critical cases and shared experiences to consult with our fellow specialists. I have shared the usefulness of telemedicine among rural populations with my peers.* [MO01]

Finally, MOs reflected on the long-term benefits of telemedicine for patients, particularly older patients, while also identifying the need for privacy-sensitive physical spaces. This reflected the CFIR domain of Implementation Outcome—Reflecting and Evaluating because participants evaluated both the benefits of the service and the improvements required for equitable and sustainable implementation. One MO stated:

> *Telemedicine enabled access to medical reports, e-prescriptions, and teleconsultations from home, reducing hospital visits, particularly for elderly patients. Additionally, the provision of a separate examination room for female patients or patients with sensitive conditions should be made.* [MO01]

These reflections suggest that while MOs broadly endorsed the telemedicine model, they recognized that its long-term sustainability and equity depend on technological refinement, reliable infrastructure, peer support, and attention to culturally appropriate physical spaces at the point of care.

Nurses' Insights

Theme 1: Background and Digital Experience

Nurses and ASHA or ANM workers described their roles as extending beyond basic facilitation of teleconsultations to include patient assessment, vital sign measurement, digital data entry, coordination with remote doctors, patient education, and follow-up. This reflected the CFIR domain of Individuals—Capability, as frontline workers took on an expanded clinical and digital role within the assisted telemedicine model. One nurse described this end-to-end role:

> *I accompanied ASHA and ANM workers to villages, shared contact details, responded to calls by visiting households, assessed patients' conditions, conducted teleconsultations, and shared e-prescriptions.* [N04]

Similarly, an ASHA worker described simultaneously assessing patients, recording BP, SpO_2, and ECG, communicating findings to doctors, explaining prescriptions, and creating community awareness (ASHA01). These accounts indicate that the assisted telemedicine model fundamentally expanded the scope of frontline practice, positioning these workers as the primary clinical and digital interface between patients and remote doctors.

XSL•FO
RenderX

Prior clinical and technical exposure appeared to facilitate adaptation to the digital workflow. In terms of the TAM construct of perceived ease of use, existing familiarity with clinical monitoring equipment eased the transition to portable diagnostic tools, while structured hands-on training provided the skills required to perform the expanded role. A nurse noted previous experience with ventilators, patient monitoring systems, and 12-lead ECG, suggesting that existing familiarity with clinical monitoring equipment eased the transition to portable diagnostic tools (NO10). Structured onboarding further reinforced this readiness; one nurse described 10 days of practical training covering vital sign measurement on dummy patients, ECG operation, portal use, video consultations, chatbot functions, e-prescriptions, and monitoring doctor availability (NO04).

Together, these findings suggest that frontline workers' initial readiness for telemedicine was shaped by prior clinical experience, digital familiarity, and the expanded professional responsibilities they assumed within the telemedicine workflow.

Theme 2: Onboarding and Training

Structured hands-on training provided the skills required to perform the expanded frontline role. This aligned with the CFIR Inner Setting—Access to Knowledge and Information domain because formal training equipped workers with the practical knowledge needed to operate devices, use the portal, and coordinate teleconsultations. One nurse described the training process:

> *We received 10 days of training on taking BP, SpO_2 , and ECG on a dummy patient, uploading data to the portal, conducting video calls with doctors, and logging into the nurse profile. We also learned to use the chatbot to enter patient details, upload medical reports, connect with doctors, receive e-prescriptions, and monitor doctors' availability status.* [N04]

Continued access to technical assistance further supported implementation. This reflected the CFIR Inner Setting—Available Resources domain, as responsive technical support helped frontline workers resolve operational problems quickly. One ASHA worker reported:

> *We contacted the IT team of our department for any issues and usually received solutions the same day.* [ASHA03]

Together, these findings suggest that frontline workers' capacity to deliver telemedicine was built through a combination of prior clinical experience, structured practical training, and responsive post-deployment support.

Theme 3: Usability and Technical Functionality

Frontline workers generally found routine devices and the telemedicine portal manageable after training, although the portable ECG initially required additional familiarization. This aligned with the TAM construct of perceived ease of use, as workers' confidence improved through training and repeated use. An ASHA worker explained:

> *ECG was newer and more difficult at first, but after training, I understood it better.* [ASHA04]

The portal and app were also considered accessible, with bilingual instructions in Hindi and English and clearly structured data entry fields facilitating use among workers with varying levels of digital familiarity. An ASHA worker stated:

> *The portal and app were easy to understand with instructions in both Hindi and English and clear columns for details, so there was no problem using it.* [ASHA03]

Connectivity remained an important operational constraint, particularly during video consultations. This reflected the CFIR Inner Setting—Information Technology Infrastructure domain because network limitations directly affected the reliability of real-time consultations. Workers reported using the chatbot and other communication channels when network availability was inadequate. As an ANM explained:

> *Sometimes there was no network, then we used the chatbot or other alternatives for consultation.* [ANM01]

Similarly, when the portal was unavailable, workers shifted consultations to alternative video calling apps and simultaneously reported technical failures to the IT team while maintaining direct communication with doctors. This reflected the CFIR domain of Implementation Process—Adapting, as frontline workers developed practical workarounds to maintain service continuity. One nurse stated:

> *If the portal did not work, then we made the patient talk to the doctor through video call on another app.* [N03]

Another nurse explained:

> *We conveyed technical failure to the IT team and talked to the doctors directly on the video call.* [N04]

Taken together, these accounts suggest that continuity of telemedicine depended not only on system usability but also on frontline workers' ability to adapt workflows dynamically and access timely technical support when digital infrastructure failed.

Theme 4: Communication Practices and Patient Support

Nurses and community health workers played an important intermediary role in introducing telemedicine, addressing hesitation, and making remote consultations understandable to patients. This aligned with the CFIR Outer Setting—Local Conditions domain because frontline workers used existing community relationships to improve access and encourage use of the service. ASHA workers shared their phone numbers during awareness meetings and health camps so that patients could contact them directly when unwell. One ASHA worker reported:

> *We shared our phone numbers during awareness meetings and camps, and patient called us directly whenever they fell sick so that we could visit them for teleconsultations.* [ASHA02]

Local examples of patients who had benefited from telemedicine were used to build confidence among hesitant community members. This reflected the CFIR domain of Implementation Process—Engaging because trust was generated through local

XSL•FO
RenderX

evidence and repeated interpersonal communication. One ASHA worker stated:

> *We used examples of treated patients of the village itself to build trust.* [ASHA01]

Continued counseling and motivation were also considered essential for patients accustomed to conventional face-to-face care. Another ASHA worker noted:

> *With proper counselling and motivation, it is easier to make them believe and trust the telemedicine services.* [ASHA02]

During consultations, frontline workers mediated linguistic and health literacy barriers by translating doctors' instructions and explaining digital prescriptions. This reflected the CFIR domain of Implementation Process—Doing because frontline workers actively facilitated the consultation process and ensured that medical advice was understood. A nurse stated:

> *During teleconsultation, if the patient did not understand the doctors' instructions, then I explained the advice in their local language.* [N02]

Another nurse described explaining prescription abbreviations and translating e-prescription instructions into the local language:

> *We clarified instructions like BD (twice daily) or SOS (only when needed), and explained the e-prescription in local language to patients.* [N03]

Gender and privacy concerns also influenced acceptance. This aligned with the CFIR construct of Innovation—Innovation Recipient Impact because patients' acceptance of telemedicine was shaped by reassurance, confidentiality, and observed benefits. An ANM reported that some female patients initially hesitated about video consultations but became more accepting after reassurance about data confidentiality, discussion of other patients' recovery, and explanation of the benefits of telemedicine:

> *Female patients hesitated and doubted video calls, but after assurance of data confidentiality, citing fellow patients' recovery and explaining the benefits of telemedicine, they trusted and accepted the services.* [ANM01]

The frontline role also extended beyond the initial teleconsultation to monitoring treatment adherence and assessing patient improvement. This reflected the TAM construct of perceived usefulness because follow-up reinforced the practical value of telemedicine for patients. A nurse described:

> *During teleconsultation, if the patient is provided with a 7-day course, then we do a follow-up within 5 days to see whether they have taken the medicines on time, or any health improvement is noted or arrange another consultation.* [N01]

Patients also returned with positive feedback and reported recovery, reinforcing frontline workers' confidence in the service and validating the model's clinical utility at the community level. An ASHA worker stated:

> *Patients came back with positive feedback and reported recovery.* [ASHA03]

These findings indicate that frontline workers functioned as clinical, linguistic, and cultural intermediaries, connecting the digital system with patients' social, linguistic, and practical needs.

Theme 5: Feedback and System Improvement

Although frontline workers perceived clear benefits from telemedicine, they identified several conditions required for sustained implementation. This reflected the CFIR domain of Implementation Process—Assessing Context because participants highlighted gaps in medicines, physical infrastructure, connectivity, and devices that needed to be addressed to strengthen service delivery. One nurse summarized the importance of integrating medicines into the model:

> *If the medicines were made available to us, we could support the patient in a better way.* [N02]

An ASHA worker advocated for dedicated consultation spaces, noting:

> *A proper place for teleconsultation should be set up with a medical facility where the patient can easily reach and be examined.* [ASHA04]

An ANM further emphasized the need for technology adapted to rural connectivity constraints, stating:

> *Low-internet-compatible devices and tablets are needed for smoother data entry and better video calls, along with simple video/user guides for troubleshooting.* [ANM01]

Ongoing feedback from frontline workers additionally contributed to system improvement. This aligned with the CFIR domain of Implementation Process—Reflecting and Evaluating because technical and operational problems were reported, reviewed, and resolved during implementation. Participants described technical and operational problems being reported to the project team and resolved within a few days. One nurse stated:

> *The issue reported to the team and that get solved next time or in 2-3 days.* [N02]

Online training was also provided when additional support was required. One nurse explained:

> *We were provided online training whenever needed to do things correctly in the field.* [N05]

Coordination between frontline workers and remote doctors also provided an important safety mechanism during technical disruptions. This reflected the CFIR domain of Individuals—Capability, as clinicians and frontline workers used alternative communication methods to maintain care when the primary system failed. One nurse stated:

> *When the portal was down, doctors supported us through alternate video calls, guiding patients with diagnosis and treatment so continuity of care was never disrupted.* [N03]

XSL•FO
RenderX

Thus, continuity was supported by a combination of structured follow-up practices, interprofessional communication, responsive technical support, and iterative learning rather than by the digital platform alone.

Theme 6: Reflections and Future Engagement

Despite the constraints identified, frontline workers expressed strong support for continuation because telemedicine reduced travel, waiting time, and costs, particularly for older and geographically remote patients. This reflected the CFIR construct of Innovation—Relative Advantage and the TAM construct of perceived usefulness because frontline workers judged telemedicine to be beneficial compared with conventional care-seeking pathways. An ANM stated:

> *Earlier we had to travel far and wait in queues; now telemedicine gives us treatment at home, saving time and money.* [ANM01]

An ASHA worker similarly advocated for the service's continuation, noting:

> *Telemedicine should continue as it saves time and money and helps elderly and remote patients get treatment at home.* [ASHA03]

At the same time, participants recognized that further improvements were necessary for long-term sustainability. This reflected the CFIR domain of Implementation Process—Reflecting and Evaluating because frontline workers evaluated both the benefits of the service and the system-level improvements required for continued delivery. One nurse summarized:

> *Telemedicine gave timely care and relief at home, but better connectivity, medicines, and extended hours are needed.* [N05]

These reflections suggest that while frontline workers strongly endorsed the telemedicine model, its long-term sustainability depends on addressing systemic gaps in physical infrastructure, medicine supply, digital connectivity, and service availability.

Patients' Insights

Theme 1: Telemedicine Addressed Important Barriers to Conventional Health Care Access

Patients described telemedicine primarily in relation to the practical difficulties they previously experienced when seeking conventional health care. This reflected the CFIR domain of the Outer Setting—Local Conditions, as participants reported long travel distances, transportation costs, accompanying family members, waiting times, and disruption of daily activities. One participant explained the logistical burden:

> *With no nearby hospital, we had to travel 40 km or use private care, leading to an accompanying person, high costs and loss of time.* [PT08]

Awareness of telemedicine was largely generated through ASHA workers, project staff, and community members, which aligned with the CFIR construct of Individuals—Motivation, as interpersonal networks drove the initial desire to try the service. As one participant described:

> *I came to know about telemedicine through the villagers and the project staff, about the direct consultation with doctors on video call.* [PT03]

Furthermore, positive experiences shared within villages encouraged its use, reflecting the TAM construct of social influence. Another participant noted:

> *People in my village encouraged me to use telemedicine because of its benefits, but I haven't seen it used like this in nearby villages.* [PT02]

These accounts indicate that initial engagement was driven not by institutional marketing but by organic, community-level diffusion through trusted personal networks. Additionally, patients' baseline digital readiness was limited; reflecting the CFIR domain of Individuals—Capability, 1 patient recalled having seen a BP machine at the anganwadi but never at home (PT07), highlighting the absolute necessity of the assisted telemedicine model.

Theme 2: Local Health Workers Enabled Patients to Navigate the Digital Care Pathway

Patients did not generally experience telemedicine as a fully self-directed digital service. Instead, local health workers provided the practical bridge between patients and remote doctors. This reflected the CFIR domain of Implementation Process—Engaging, as frontline workers registered patients, recorded symptoms and vital signs, facilitated video consultations, and explained prescriptions. One participant described the structured workflow:

> *Team visits homes daily to register patients, record symptoms, connect with doctors online, and note e-prescriptions.* [PT15]

Another stated:

> *The telemedicine staff and ASHA assisted us in teleconsultation with the doctor, explaining symptoms and e-prescription.* [PT09]

This active facilitation compensated for the patients' limited direct experience with digital devices. Patients' confidence also developed through repeated exposure and communication in familiar language, which addressed the CFIR Outer Setting—Local Attitudes regarding cultural and linguistic comfort. One participant explained that the consultation language was easy to understand and that continued support could enable educated family members to assist with future use:

> *The consultation language was easy to understand. With better systems and continued support, my educated family members can use it, and I can encourage them.* [PT11]

Thus, patients' engagement with telemedicine appeared to depend substantially on human mediation and the Innovation Design of the assisted delivery model rather than digital access alone.

Theme 3: Trust, Communication, and Privacy Shaped Acceptance of Remote Consultations

Patients described trust as developing through familiarity with local health workers, clear communication, and positive

XSL•FO
RenderX

experiences with treatment. This aligned with the CFIR construct of Innovation—Innovation Evidence Base, as observable clinical effectiveness validated the remote care model. One participant stated:

> *The team members have earned our trust and act as our trusted link to the doctors, especially because the treatment works.* [PT07]

Privacy and confidentiality were also important for patients' comfort during remote consultations. This reflected the CFIR Inner Setting—Recipient Centeredness, as respectful interactions and data protection ensured patient dignity in a close-knit community setting. One participant reported:

> *My privacy was respected, and my health information was kept confidential.* [PT01]

These accounts suggest that acceptance of telemedicine was influenced not only by access to a remote doctor but also by the quality of interpersonal mediation surrounding the consultation, including reassurance, understandable communication, and confidence that personal health information would remain private.

Theme 4: Convenience and Perceived Benefits Reinforced Continued Use

Patients consistently emphasized the practical value of receiving care closer to home. This strongly reflected the TAM construct of perceived usefulness, as telemedicine was perceived to reduce travel, time, financial expenditure, and disruption to daily activities. One participant reported:

> *My health improved, and worries reduced as I do not have to go to the doctor physically and can get treatment at the doorstep.* [PT10]

Another emphasized the comparative benefits:

> *Telemedicine helps in avoiding long travelling distances, saves money, time and effort with sufficient care.* [PT08]

These perceived advantages contributed to a strong TAM behavioral intention to continue using the service and to recommend it to others. One participant stated:

> *We will definitely use these services because they come to our doorstep at lower cost and save our time and money.* [PT12]

Furthermore, acceptance extended beyond individual users through organic peer support, reflecting actual system use and community-level diffusion. One patient reported actively supporting hesitant peers:

> *I have helped women and elderly patients who were hesitant to speak on video calls by explaining and supporting them.* [PT01]

This suggests that satisfied patients assumed informal advocacy roles that reinforced community-level adoption.

Theme 5: Technical and Service Limitations Remained Important Constraints

Despite generally positive experiences, patients identified limitations that restricted the completeness and reliability of telemedicine. Network instability sometimes resulted in interrupted audio and blurred video during consultations, highlighting constraints within the CFIR Inner Setting—Information Technology Infrastructure. One participant noted:

> *Sometimes we faced network issues like unstable connectivity, voice breaks, and video blurring during video consultations.* [PT09]

Participants also reported that medicines and certain diagnostic tests were not always available through the telemedicine service, requiring visits to government hospitals. This reflected gaps in local resource availability and the Outer Setting. One participant stated:

> *They do not provide medicines or perform tests for dengue, malaria, blood infection and allergies, for which we have to visit nearby government hospitals, which is a disadvantage.* [PT12]

Patients therefore suggested that the service would be more useful if medicines, injections, and basic investigations were available locally with trained staff:

> *The service would improve if medicines, injections, and basic tests were available locally with trained staff.* [PT08]

Importantly, patients also provided a balanced assessment of the service's operational reliability, reflecting the CFIR Inner Setting—Work Infrastructure. One participant reported:

> *We never faced any major planning or scheduling problems; only occasional delays happened when the network was poor, or doctors were busy.* [PT05]

Patients further recognized the workload pressures on frontline workers and advocated for enhanced institutional support, noting:

> *With proper training, enough staff, and support like medicines and tools, the health workers can manage the work smoothly and help everyone better.* [PT09]

They also valued the responsiveness of the project team to feedback, an essential component of the CFIR Implementation Process—Reflecting and Evaluating, although quicker responses and follow-up were suggested as areas for improvement:

> *The team listens to suggestions and responds well, but a quicker response and follow-up would help.* [PT15]

Theme 6: Patients Viewed Telemedicine as a Useful Addition to Community Health Care

Overall, patients' accounts reflected a positive assessment of telemedicine particularly because it brought medical consultation closer to their homes and reduced the logistical burden of conventional health care seeking. Engaging in the CFIR Implementation Process—Reflecting and Evaluating, one participant summarized this experience:

> *Telemedicine has been very helpful for us as it gives good treatment in the village, saves time and money,*

XSL•FO
RenderX

> *avoids long travel, and makes us feel confident that care is available when we need it.* [PT11]

At the same time, patients' reflections indicate that continued engagement depended on maintaining the human support surrounding the technology and addressing practical limitations, such as connectivity, medicine availability, diagnostic services, staffing, and follow-up. Participants also expressed willingness to remain involved in future discussions and service improvement, reflecting the CFIR Implementation Outcome—Implementability and a sense of shared ownership. One participant stated:

> *I am ready to talk again, and to help if needed in the future.* [PT02]

This willingness to contribute suggests that patients had transitioned from passive recipients to active stakeholders in the telemedicine model.

Discussion

Principal Findings

A primary innovation of this study lies in the empirical validation of the assisted telemedicine model in hard-to-reach terrains. Unlike traditional direct-to-consumer telehealth models that assume a baseline of digital literacy, this study highlights a novel sociotechnical mediation framework where frontline workers (ASHAs, ANMs, and NOs) act as digital health navigators. By combining human cultural mediation with portable, specialized diagnostic tools (such as the Spandan ECG by Sunfox Technologies Pvt. Ltd.), this model overcomes the critical barriers of the digital divide and the lack of physical examination, offering a highly scalable blueprint for universal health coverage in resource-constrained settings. This study provides a comprehensive evaluation of telemedicine implementation within rural, hard-to-reach districts, specifically identifying the sociotechnical factors that facilitate digital adoption among underserved populations. By assessing the efficacy of an assisted model, using frontline health care workers and portable diagnostic tools for real-time vital screening, the findings suggest a potential shift in treatment-seeking behavior compared to traditional face-to-face consultations. This analysis offers valuable insights for optimizing digital health interventions.

Sociotechnical Intermediation and Assisted Adoption Model

The persistent shortage of skilled health care providers in rural India remains a primary obstacle to achieving universal health coverage, with approximately 74% of medical practitioners serving only 28% of the population [15,16]. While the National Rural Health Mission has introduced financial incentives and contractual appointments, a significant gap remains in reaching isolated terrains. A distinguishing feature of this study is the identification of the assisted telemedicine model, where frontline workers act as essential technical and cultural mediators. Unlike previous models that rely on direct patient-to-provider interfaces, our findings reveal an increase in acceptance driven by ASHAs, ANMs, and nurses who bridge the digital divide for patients unfamiliar with screens. This is particularly crucial for women and older patients in hard-to-reach areas who may hesitate to consult male specialists or navigate complex portals [17]. By providing a human connection, the study demonstrates that acceptance is not merely a product of software functionality but of a structured human touch in a system that builds trust and overcomes cultural barriers [18,19].

Clinical Efficacy of Digital Medical Tools and Decision-Making in Remote Care

A key contribution of this study is the evaluation of specialized digital tools, including the portable Spandan ECG (Sunfox Technologies Pvt. Ltd.) for real-time first-level cardiac screening, along with the measurement of vital parameters, such as BP, SpO_2, and heart rate. While telemedicine is often criticized for a lack of physical examination [20], our study reveals a robust workflow where field staff record vitals (BP, SpO_2, and heart rate) and upload photographic evidence of digital devices via chatbots for clinician cross-verification. This protocol significantly enhances diagnostic confidence and data accuracy, a finding that moves beyond generic teleconsultation literature.

Despite these advancements, diagnostic limitations remain in dermatology and emergency care where physical palpation is irreplaceable. However, our findings align with those of large-scale health care systems, showing that sustained digital interaction can improve glycemic control and continuity of care for chronic conditions [21,22]. The efficacy of this model is therefore contingent on the robustness of rural digital infrastructure, such as BharatNet, and the hands-on training of health care workers to manage these sophisticated tools [23,24]. In such environments, the burden of managing chronic conditions shifts to these intermediaries, necessitating robust educational interventions to prevent implementation fatigue and ensure the quality of remote clinical evidence [25,26].

Reconfiguring Treatment-Seeking Behavior and Health Equity

Our study provides empirical evidence of a fundamental shift in treatment-seeking behavior among rural populations. Traditionally, patients in remote Rajasthan or Maharashtra faced prohibitive logistical and financial barriers, often traveling up to 850 km for specialist care. By bringing health care to the "doorstep," this study shows that telemedicine reduces attrition in traditional testing cascades and allows patients to reinvest the time saved into essential economic activities such as farming [27].

This study emphasizes that while telemedicine enhances equity and reduces financial hardship, its success is conditional. Trust building is a critical determinant; repeated technical failures or doctor unavailability can lead to a rapid loss of faith in the system. When functioning reliably, telemedicine acts as a peer-support and counseling platform, fostering treatment adherence. Ultimately, this research suggests that while digital health can bridge the health care divide, it requires addressing underlying gaps in connectivity, staffing, and trust to ensure long-term sustainability and systemic integration [28,29].

XSL•FO
RenderX

Policy Implications for Scalable Digital Health Systems

The findings of this study underscore the need to shift health care policy from "provider-centric" to "mediated-digital" models. For telemedicine to be effectively integrated into the existing public health framework, policy interventions must prioritize formalizing the ASHA and ANM roles as "digital health navigators." Furthermore, to sustain the positive shift in treatment-seeking behavior, future initiatives should focus on strengthening the rural digital infrastructure (BharatNet) and providing standardized, hands-on training for portable diagnostic tools.

Strengths and Limitations

This study's multistakeholder design provides a 360-degree view of the telemedicine ecosystem by integrating the perspectives of MOs, frontline health care workers (nurses and ASHAs or ANMs), and patients themselves. This approach enabled a holistic understanding of how the telemedicine model operates across different levels of the health care hierarchy. Furthermore, the use of the CFIR and TAM frameworks provided a robust theoretical foundation for moving beyond anecdotal evidence, enabling a structured analysis of how capability, motivation, and system design influence digital adoption in resource-limited settings. The inclusion of specific evaluations of portable diagnostic tools, such as the Spandan ECG (Sunfox Technologies Pvt. Ltd.), adds a layer of clinical relevance often missing from broader qualitative digital health studies.

While the study captures in-depth, real-world experiences of patients and frontline health care workers who were actively involved in telemedicine services, the findings are context specific to this implementation setting. The extended implementation period and inclusion of multiple stakeholder groups strengthen the relevance of the findings. Additionally, patients who opted to visit primary health Centerscenters despite the availability of telemedicine services were not included in this study.

Conclusions

This study provides rich qualitative insights into the post-implementation delivery and use of telemedicine in remote areas of Rajasthan. Telemedicine was perceived as acceptable, useful, and supportive of continuity of care by both patients and frontline health care workers when embedded within existing primary health care systems. The shift in treatment-seeking behavior observed offers valuable evidence to support the further integration of digital health into the National Rural Health Mission.

Telemedicine should be viewed not as a secondary substitute for in-person care, but as a robust, complementary system that, when supported by trained mediators and reliable infrastructure, can significantly advance the goal of universal health coverage for India's most vulnerable populations.

Acknowledgments

The authors declare that no generative AI or large language models were used in the generation of this manuscript, including the writing, analysis, or interpretation of data.

All authors declared that they had insufficient or no funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided APF support for the publication of this article.

Funding

This study was supported by the Indian Council of Medical Research (grant IIRP-2023- - 0000151).

Data Availability

Due to ethical restrictions and participant confidentiality agreements, data from this study are not publicly available but can be made available in aggregated or deidentified form upon request to the corresponding author.

Authors' Contributions

RK Huda conceptualized the study and developed the methodology. SK was responsible for software development. SK and RS Chowhan performed the validation. RS Chowhan and SK conducted the formal analysis. RS Chowhan, RK Sangwan, and MK conducted the investigation. RK Huda was responsible for the resources, supervision, project administration, and funding acquisition. RK Sangwan and MK conducted data curation. RK Huda prepared the original draft of the manuscript. RS Chowhan, SK, RK Sangwan, and MK contributed to the review and editing of the manuscript. RS Chowhan was responsible for visualization. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

Interview guides used for in-depth qualitative interviews with study participants.
[DOCX File, 22 KB - mhealth_v14i1e93146_app1.docx]

Multimedia Appendix 2
Codebooks.
[DOCX File, 16 KB - mhealth_v14i1e93146_app2.docx]

Multimedia Appendix 3
Participant characteristics, study extraction, and thematic table derived from the qualitative interviews.
[DOCX File, 24 KB - mhealth_v14i1e93146_app3.docx]

Checklist 1
iCHECK-DH guidelines and checklist for the reporting on digital health implementations.
[DOCX File, 21 KB - mhealth_v14i1e93146_app4.docx]

Checklist 2
COREQ checklist.
[DOCX File, 16 KB - mhealth_v14i1e93146_app5.docx]

References

1. Dorsey ER, Topol EJ. State of telehealth. N Engl J Med 2016 Jul 14;375(2):154-161. [doi: 10.1056/NEJMra1601705] [Medline: 27410924]
2. Ndwabe H, Basu A, Mohammed J. Post pandemic analysis on comprehensive utilization of telehealth and telemedicine. Clinical eHealth 2024 Dec;7:5-14. [doi: 10.1016/j.ceh.2023.12.002]
3. Mittal AK, Patel M, Dwivedi R, et al. Economic evaluation of telemedicine services provided at Satellite Centre for Tribal Health and Research, Abu Road Sirohi Rajasthan India. Research Square. Preprint posted online on Oct 28, 2024. [doi: 10.21203/rs.3.rs-5310762/v1]
4. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: a systematic review. J Telemed Telecare 2018 Jan;24(1):4-12. [doi: 10.1177/1357633X16674087] [Medline: 29320966]
5. Ravikumar H, Govindarajan D, G V. The impact of online doctor consultations on healthcare access and utilization among rural and underserved populations: an observational study in India. IJGMED 2025;3(1):15-20. [doi: 10.34218/IJGMED_03_01_002]
6. M A Dr, C.m C Dr. A qualitative study of community health officers' experiences with telemedicine in a rural block of Madhya Pradesh. PIJR 2023 Aug 15:15-17. [doi: 10.36106/paripex/5800656]
7. Huda RK, Chowhan RS, Arora S, et al. Perceptions of healthcare workers and patients on the implementation of telemedicine in hard-to-reach areas: a qualitative study from Rajasthan, India. Digit Health 2025 May;11. [doi: 10.1177/20552076251331874]
8. Bhagavathula AS, Aldhaleei WA. Bridging the telehealth divide: racial and ethnic disparities in medicare telehealth usage highlights the need for equity-focused approaches. Telemed J E Health 2024 May;30(5):1272-1278. [doi: 10.1089/tmj.2023.0536] [Medline: 38215268]
9. Haimi M. The tragic paradoxical effect of telemedicine on healthcare disparities- a time for redemption: a narrative review. BMC Med Inform Decis Mak 2023 May 16;23(1):95. [doi: 10.1186/s12911-023-02194-4] [Medline: 37193960]
10. Siwach P, Pawaria S, Bajaj R, Sachdeva S. Patient adherence in telehealth: a comparative analysis with conventional outpatient services. Musculoskeletal Care 2025 Jun;23(2):e70137. [doi: 10.1002/msc.70137] [Medline: 40540309]
11. Ezenwaji CO, Alum EU, Ugwu OPC. Bridging the gap: telemedicine as a solution for HIV care inequities in rural and vulnerable communities. Int J Equity Health 2025 Jul 14;24(1):205. [doi: 10.1186/s12939-025-02584-2] [Medline: 40660269]
12. Perrin Franck C, Babington-Ashaye A, Dietrich D, et al. iCHECK-DH: guidelines and checklist for the reporting on digital health implementations. J Med Internet Res 2023 May 10;25:e46694. [doi: 10.2196/46694] [Medline: 37163336]
13. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
14. Lincoln YS, Guba EG. Naturalistic Inquiry: Sage Publications; 1985.
15. Sundararaman T, Gupta G. Indian approaches to retaining skilled health workers in rural areas. Bull World Health Organ 2011 Jan 1;89(1):73-77. [doi: 10.2471/BLT.09.070862] [Medline: 21346894]
16. Dasgupta A, Deb S. Telemedicine: a new horizon in public health in India. Indian J Community Med 2008 Jan;33(1):3-8. [doi: 10.4103/0970-0218.39234] [Medline: 19966987]
17. Kamble S, Gawde N, Goel N, et al. Access, timeliness and retention for HIV testing under early infant diagnosis (EID) program, India. Sci Rep 2023 Apr 6;13(1):5638. [doi: 10.1038/s41598-023-32056-y] [Medline: 37024531]

XSL•FO
RenderX

18. Mashoudy KD, Perez SM, Nouri K. From diagnosis to intervention: a review of telemedicine's role in skin cancer care. Arch Dermatol Res 2024 May 2;316(5). [doi: 10.1007/s00403-024-02884-7]
19. Patel R, Huang J, Hsueh L. Telemedicine's impact on diabetes care during the COVID-19 pandemic: a cohort study in a large integrated healthcare system. medRxiv. Preprint posted online on Feb 27, 2024. [doi: 10.1101/2024.02.25.24303335]
20. Tenderich A. Virtual nation: telemedicine's breakout moment. J Diabetes Sci Technol 2020 Jul;14(4):799-800. [doi: 10.1177/1932296820929359] [Medline: 32429751]
21. Parthasarathi A, George T, Kalimuth MB, et al. Exploring the potential of telemedicine for improved primary healthcare in India: a comprehensive review. Lancet Reg Health Southeast Asia 2024 Aug;27:100431. [doi: 10.1016/j.lansea.2024.100431] [Medline: 38957222]
22. Sood S, Lal K, Bhatia M, et al. Adoption and utilization of India's eSanjeevani national telemedicine service. Oxford Open Digital Health 2025 Jan 18;3. [doi: 10.1093/oodh/oqaf025]
23. Jacob SA, Daas R, Feliciano A, LaMotte JE, Carroll AE. Caregiver experiences with accessing sickle cell care and the use of telemedicine. BMC Health Serv Res 2022 Feb 22;22(1):239. [doi: 10.1186/s12913-022-07627-w] [Medline: 35193570]
24. Manglani M, Lala MM, Gabhale Y, et al. Attitudes and acceptability of children, caregivers, and healthcare providers about using telemedicine for pediatric HIV care in a resource-limited setting. In: M VLC, editor. PLoS One 2022;17(5):e0268740. [doi: 10.1371/journal.pone.0268740] [Medline: 35613106]
25. Akintunde TY, Akintunde OD, Musa TH, et al. Expanding telemedicine to reduce the burden on the healthcare systems and poverty in Africa for a post-coronavirus disease 2019 (COVID-19) pandemic reformation. Glob Health J 2021 Sep;5(3):128-134. [doi: 10.1016/j.glohj.2021.07.006] [Medline: 36338822]
26. Demant MN, Jensen RG, Bhutta MF, Laier GH, Lous J, Homøe P. Smartphone otoscopy by non-specialist health workers in rural Greenland: a cross-sectional study. Int J Pediatr Otorhinolaryngol 2019 Nov;126:109628. [doi: 10.1016/j.ijporl.2019.109628]
27. Drake C, Lian T, Cameron B, Medynskaya K, Bosworth HB, Shah K. Understanding telemedicine's "new normal": variations in telemedicine use by specialty line and patient demographics. Telemed J E Health 2022 Jan;28(1):51-59. [doi: 10.1089/tmj.2021.0041] [Medline: 33769092]
28. Thapa SB, Gandhi A. Exploring telemedicine and organizational challenges in the healthcare system: a qualitative analysis using Grounded Theory. J Health Organ Manag 2025 Apr 10;39(3):402-419. [doi: 10.1108/JHOM-04-2024-0157] [Medline: 39601257]
29. Orrange S, Patel A, Mack WJ, Cassetta J. Patient satisfaction and trust in telemedicine during the COVID-19 pandemic: retrospective observational study. JMIR Hum Factors 2021 Apr 22;8(2):e28589. [doi: 10.2196/28589] [Medline: 33822736]

Abbreviations

ANM: auxiliary nurse midwife
ASHA: accredited social health activist
BP : blood pressure
CFIR: Consolidated Framework for Implementation Research
COREQ: Consolidated Criteria for Reporting Qualitative Research
ECG: electrocardiogram
iCHECK-DH: Guidelines and Checklist for the Reporting on Digital Health Implementations
ICMR: Indian Council of Medical Research
IDI: in-depth interview
MERN: MongoDB, Express, React, Node.js
MO: medical officer
NIIRNCD: National Institute for Implementation Research on Non-Communicable Diseases
NO: nursing officer
SpO2: oxygen saturation
TAM: technology acceptance model

Edited by A Stone; submitted 09.Feb.2026; peer-reviewed by C Baxter, S Hassan; revised version received 28.Aug.2026; accepted 28.Aug.2026; published 16.Oct.2026.

Please cite as:
Huda RK, Chowhan RS, Kalra S, Sangwan RK, Khetan M
Insights on Postimplementation and Delivery of Telemedicine in Remote Areas of Rajasthan: Qualitative Study
JMIR Mhealth Uhealth 2026;14:e93146
URL: https://mhealth.jmir.org/2026/1/e93146
doi:10.2196/93146

XSL•FO
RenderX

© Ramesh Kumar Huda, Rahul Singh Chowhan, Sumit Kalra, Ramesh Kumar Sangwan, Mukti Khetan. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 16.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effectiveness of a Multicomponent Digital Therapeutic Program on Inhaled Medication Adherence in Chronic Obstructive Pulmonary Disease: Pilot Randomized Controlled Trial

Hanna Wu[1,2], BMed; Fen Dong[2], PhD; Yanan Cui[2], PhD; Danjing Hu[2], PhD; Yushi Zhang[2,3], BMed; Tianqi Peng[1], BMed; Yaodie Peng[1], BMed; Xiang Li[1], BMed; Yong Li[2], BSc; Xiaopan Li[2], BSc; Yongxia Chen[4], BSc; Qingjie Yu[4], BSc; Hongtao Niu[1,2], PhD; Ting Yang[1,2], MD

[1]Peking University China-Japan Friendship School of Clinical Medicine, Yinghua Street, Beijing, China

[2]National Center for Respiratory Medicine; State Key Laboratory of Respiratory Health and Multimorbidity; National Office for Chronic Respiratory Disease Prevention and Control; National Clinical Research Center for Respiratory Diseases; Institute of Respiratory Medicine, Chinese Academy of Medical Sciences & Peking Union Medical College; Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine, China-Japan Friendship Hospital, Beijing, China

[3]Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[4]Tianjin Center for Medical Devices Evaluation and Inspection, Tianjin, China

Corresponding Author:
Hongtao Niu, PhD
Peking University China-Japan Friendship School of Clinical Medicine, Yinghua Street, Beijing, China

Abstract

Background: Poor medication adherence is a critical impediment to effective chronic obstructive pulmonary disease (COPD) management and remains challenging to track in routine care. Digital therapeutic interventions may improve chronic disease management through remote monitoring and patient education. However, most existing digital interventions focus on individual components of disease management, and evidence from longer-term pragmatic randomized controlled trials evaluating integrated digital therapeutics in routine clinical practice remains limited.

Objective: This study aimed to evaluate the effectiveness of a multicomponent digital therapeutic program combining electronic inhaler monitoring (EIM) with behavioral support in improving medication adherence and symptom-related outcomes among patients with COPD prescribed pressurized metered-dose inhalers, and to assess patient engagement and implementation in routine clinical practice.

Methods: In this prospective, open-label, unblinded, randomized controlled trial, patients with COPD prescribed pressurized metered-dose inhalers were enrolled and randomized in a 2:1 ratio to either the intervention group or the control group. All participants received an EIM device to objectively record inhaler use throughout the 24-week study period. Patients in the intervention group also received personalized medication reminders from the EIM, adherence feedback, educational content, and clinician-supported follow-up through an integrated digital management platform based on EIM-derived data, whereas the control group received usual care. The primary outcome was medication adherence over 24 weeks. Secondary outcomes included exacerbations, COPD symptoms, quality of life, and lung function.

Results: A total of 158 participants were randomized, including 106 in the intervention group and 52 in the control group. The median monitoring coverage was 161.5 (IQR 86.3 - 168.0) days during the follow-up period. The self-reported adherence measured using the Test of Adherence to Inhalers showed no significant between-group differences at any follow-up time point. However, the electronically monitored adherence was significantly higher in the intervention group than in the control group (59.5% vs 33.1%, *P*=.009). Adherence declined over time in both groups, decreasing from 85.3% during weeks 0 to 4 to 43.0% during weeks 12 to 24 in the intervention group and from 76.8% to 11.9% in the control group over the same intervals, resulting in a progressively larger between-group difference over follow-up. No significant differences were observed in exacerbations, health-related quality of life, symptom burden, lung function, or mental health outcomes. In an exploratory analysis, a significantly greater proportion of participants in the intervention group achieved the minimal clinically important difference in the COPD Assessment Test score than in the control group (odds ratio 2.21, 95% CI 1.11 - 4.41; *P*=.02).

Conclusions: The multicomponent digital therapeutic program significantly improved objective medication adherence in patients with COPD over 24 weeks. While most clinical outcomes did not differ significantly between groups, the higher proportion of participants achieving a clinically meaningful improvement in the COPD Assessment Test score warrants further investigation.

Trial Registration: ClinicalTrials.gov NCT05667363; https://clinicaltrials.gov/study/NCT05667363

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2026;14:e98870) doi:10.2196/98870

KEYWORDS

digital therapeutics; chronic obstructive pulmonary disease; COPD; adherence; electronic inhaler monitoring

Introduction

Chronic obstructive pulmonary disease (COPD) is a leading cause of morbidity and mortality worldwide and imposes a substantial burden on health care systems [1-3]. Inhaled pharmacotherapy remains the cornerstone of COPD management. However, suboptimal adherence to inhaled therapy remains highly prevalent and is a major barrier to effective disease control, contributing to an increased risk of exacerbations and a higher symptom burden [4,5]. Adherence is influenced by multiple interrelated factors at the patient, inhaler, and health care provider levels, including age, cognitive and physical abilities, health literacy, behavioral determinants, inhaler complexity and usability, and health care professionals' competency and capacity to provide appropriate inhaler education. The multidimensional nature of these factors highlights the complexity of addressing medication nonadherence in routine COPD care [6,7]. Monitoring adherence to inhaled therapy is also challenging in routine clinical practice. Traditional approaches such as patient self-report and pharmacy refill records are prone to substantial inaccuracies and may overestimate adherence due to recall bias or demonstration of idealized inhaler technique, failing to capture real-world medication use [8].

Recent advances in digital health have introduced digital therapeutic interventions designed to improve chronic disease management through behavioral support, remote monitoring, and patient education. In respiratory diseases, digital therapeutic programs typically include digital inhaler platforms and mobile health apps [9,10]. Digital inhaler monitoring devices attached to inhalers, which can objectively record medication use, provide automated reminders, and enable personalized adherence feedback, have been shown to improve adherence, enhance disease control, and reduce health care burden in chronic airway diseases in previous trials. However, most existing studies have focused on asthma and have evaluated individual components, such as adherence reminders, inhaler monitoring, or telehealth consultations. In COPD, few interventions have combined objective inhaler medication monitoring with patient self-management tools and remote clinician monitoring within a single platform [11-17]. Evidence from randomized trials evaluating integrated digital therapeutic programs in COPD remains limited, particularly in real-world clinical settings.

To address this gap, we developed a multicomponent digital therapeutic program integrating an electronic inhaler monitoring (EIM) device with a mobile management app and conducted the DIGI-COPD (Effectiveness of a Multi-Component Digital Therapeutic Program on Inhaled Medication Adherence in COPD) trial to evaluate whether this intervention could improve inhaled medication adherence and clinical outcomes in patients with COPD.

Methods

Study Design

This prospective, open-label, unblinded pilot randomized trial was conducted at the respiratory outpatient clinic of China-Japan Friendship Hospital, Beijing, China. Patients were recruited through physician recommendation during outpatient visits or self-referral in response to advertisements posted in the outpatient area. All potential participants underwent eligibility screening by the research team, which included review of medical records and spirometry results obtained within the previous 6 months, and were invited to an introductory meeting to review the study details and ask questions. Those who met the eligibility criteria and agreed to participate provided written informed consent, and baseline information was then collected.

At baseline, demographic and clinical characteristics, medical and smoking history, and exacerbation history of the participants were collected. All participants completed the modified Medical Research Council (mMRC), COPD Assessment Test (CAT), St George's Respiratory Questionnaire for Chronic Obstructive Pulmonary Disease (SGRQ-C), Patient Health Questionnaire-9 (PHQ-9), Generalized Anxiety Disorder-7 (GAD-7), and Test of Adherence to Inhalers (TAI) questionnaires [18]. Trained assessors evaluated participants' inhaler technique using the metered-dose inhaler score [19] and performed physical examination (height, weight, waist and hip circumference, blood pressure, resting pulse rate, and oxygen saturation). Follow-up assessments were conducted at weeks 4, 12, and 24. This study was reported in accordance with the CONSORT (Consolidated Standards of Reporting Trials) guidelines (Checklist 1).

Participants and Randomization

Patients were eligible if they (1) fully understood and were willing to sign the written informed consent, (2) were local residents diagnosed with COPD by a physician, (3) were aged 40 years or older, (4) were prescribed a metered-dose inhaler as long-term treatment, and (5) were able to use a smartphone app after training. COPD diagnosis was confirmed during screening by reviewing medical records and spirometry results obtained within the previous 6 months, including a documented postbronchodilator FEV_1/FVC (forced expiratory volume in 1 second/forced vital capacity) <0.70.

Participants were randomized in a 2:1 ratio (intervention:control) stratified into 2 strata according to baseline CAT score (<10 vs ≥10) using a computer-generated randomization sequence. Neither researchers nor participants were blinded to group assignment.

Sample Size

The initial sample size was estimated based on a previous questionnaire conducted by our research group, in which 258 of 737 (35.01%) patients with COPD were classified as having good adherence. Assuming a 15% absolute increase in the

XSL•FO
RenderX

proportion of patients with good adherence in the intervention group at week 24, with an α of 0.05 and 90% power, the estimated sample size was 400 participants in the intervention group and 200 in the control group, allowing for approximately 20% loss to follow-up. However, as this was a pilot study, recruitment was conducted within a predefined study period and was constrained by available funding and research personnel.

Intervention

The multicomponent digital therapeutic program comprised an EIM and a remote management platform for both patients and clinicians, supporting both patient self-management and clinician oversight (Figure 1A).

The same EIM was compatible with any pressurized metered-dose inhaler (pMDI) device regardless of whether the prescribed regimen was dual bronchodilation therapy, consisting of a long-acting beta-agonist and a long-acting muscarinic antagonist, or triple therapy, consisting of a long-acting beta-agonist, a long-acting muscarinic antagonist, and inhaled corticosteroids, as it was externally attached to the pMDI. It passively recorded each inhalation event with a time stamp and transmitted the data via a cellular network to the patient's mobile phone. The adherence data were then synchronized with the patient app and, with patient consent, transmitted to a cloud-based data center. A built-in buzzer delivered audible reminders at scheduled dosing times for the intervention group. The device emitted a daily "heartbeat" signal to confirm proper functioning (Figure 1B). The safety of the device has been tested by the China Mobile Electronic Component Testing Center.

The management app enabled patients to visualize adherence data through graphs and tables, generate monthly adherence reports, complete a daily COPD Exacerbation Recognition Tool (CERT) diary [20], and record vital signs. Additional features included accessing step counts and weather information, setting medication reminders, self-reporting of CAT and mMRC scores, and viewing educational content. Clinicians could remotely monitor patient data via mobile or desktop devices, identify high-risk individuals, and initiate timely, individualized communication (Figure 1C).

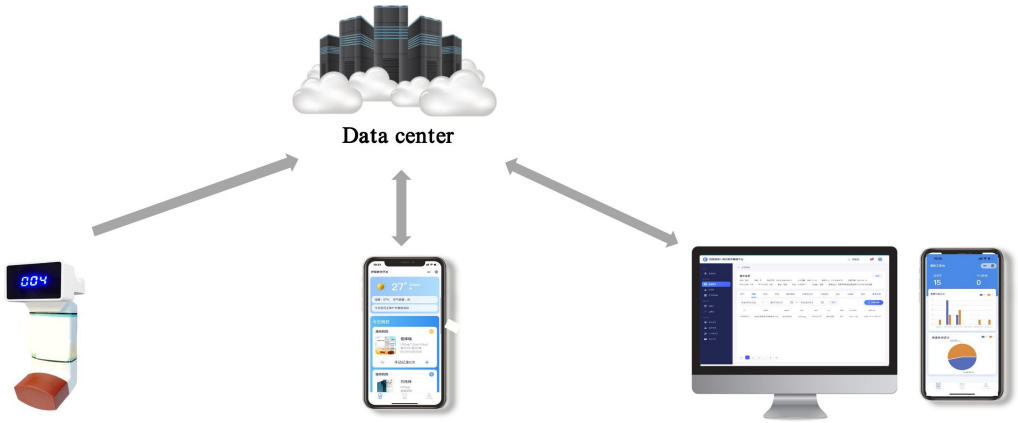
Participants in the intervention group received an EIM with a reminder paired with a mobile management app, which delivered personalized reminders and adherence feedback based on EIM data. They were instructed to complete a daily CERT diary and record vital signs in the app, which displayed trends and generated monthly reports on adherence, physiological parameters, and symptoms. Participants in the control group received an EIM without reminders, solely for recording inhaler use. They also had the app installed solely to relay EIM data, without access to adherence feedback or educational content, or clinician access for review. Both groups continued to receive usual care from their physicians at each scheduled follow-up visit.

XSL•FO
RenderX

Figure 1. Multicomponent digital therapeutic program. (A) Framework of the program. (B) Appearance of the electronic inhaler monitoring (EIM) device. (C) User interface of the management app. SN: serial number.

XSL•FO
RenderX

A

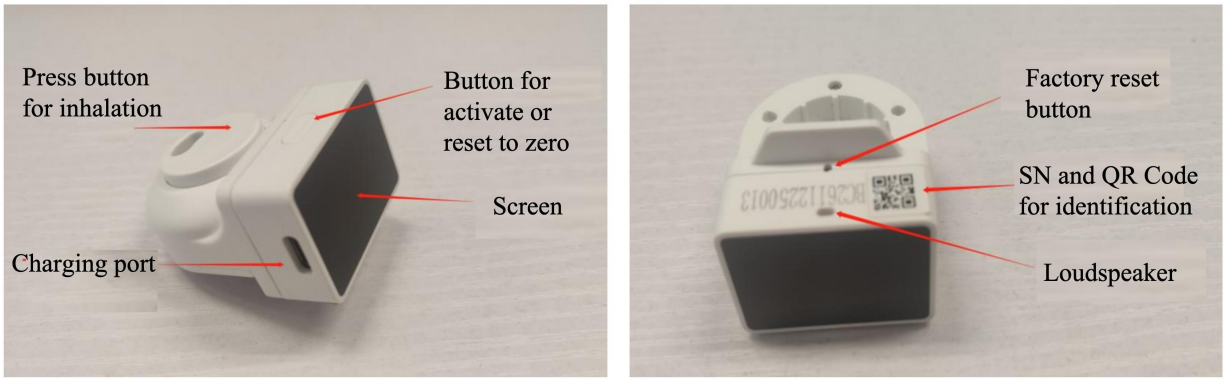


- Detect medication use
- Record medication use
- Display press actions numbers
- Emit reminders

- Visualize medication use data
- Generate adherence reports
- Keep daily symptom diaries
- Log vital signs and activities
- Recommend weather and patient education information

- Create patient files
- Record prescriptions
- Review adherence and symptom reports
- Identify low adherence and high-risk patients
- Monitor and follow-up

B



C



App interface for clinicians

App interface for patients

XSL•FO
RenderX

Outcome Measures

The primary outcomes were the proportion of participants with good adherence (TAI score ≥46) and the average daily adherence (calculated as the ratio of actual inhalations over prescribed inhalations) recorded by EIM during the study period. The TAI questionnaire assessed self-reported adherence during the preceding week and was administered at baseline and each scheduled follow-up visit. Adherence was continuously recorded by the EIM and summarized as the average daily adherence over the corresponding follow-up intervals (weeks 0 - 4, 4 - 12, and 12 - 24). The secondary outcomes included CAT score, mMRC score, PHQ-9 score, GAD-7 score, lung function, and the number of exacerbations between groups.

Statistical Analysis

All 158 randomized participants were included in the intention-to-treat analysis for primary and secondary outcomes. Continuous data were presented as mean with SD or median with IQR and analyzed using 2-tailed Student *t* tests or Mann-Whitney *U* tests as appropriate, and categorical data were presented as numbers with percentages and analyzed using the chi-square test or Fisher exact test. Two-sided tests were used throughout, and a value of $P<.05$ was considered statistically significant. Statistical analyses were conducted using IBM SPSS (version 29.0.1.0) for data management and descriptive statistics, R (version 4.4.2; R Foundation for Statistical Computing) for linear mixed-effects modeling and advanced analyses, and GraphPad Prism (version 10.4.1) for data visualization.

The primary outcome was analyzed using a linear mixed-effects model, with follow-up adherence measurements as the dependent variable, and group, time, and their interaction as fixed effects. A random intercept was included to account for within-participant correlation, and an unstructured covariance matrix was used for repeated measures over unequally spaced visits. Secondary outcomes were analyzed as follows: acute exacerbations using negative binomial regression; lung function and SGRQ scores using linear mixed models (after the Levene test indicated heterogeneity of variance, $P=.046$); and PHQ-9, GAD-7, mMRC, and CAT scores using linear mixed models. All models were adjusted for corresponding baseline values.

Subgroup analyses for mMRC and CAT scores examined potential effect modification by including group, time, subgroup, and their 3-way interaction terms in linear mixed models, with a significance threshold of $P<.10$ for interaction terms and not adjusted for multiple comparisons because it was considered exploratory. For other secondary outcomes, potential effect modification was examined using linear regression models that included group, subgroup, and their interaction term.

Ethical Considerations

The study was approved by the China-Japan Friendship Hospital Ethics Committee (approval number 2022-KY-129) and was conducted in accordance with the Declaration of Helsinki (1975, as revised in 2008) and applicable Chinese regulations. Written informed consent was obtained from all participants before enrollment. Participant data were deidentified prior to analysis and stored on secure password-protected servers accessible only to authorized study personnel. Participants did not receive any financial compensation for participation. As appreciation for their time and contribution, participants were offered 1 complimentary lung function test at the 24-week follow-up and permanent access to the EIM device and integrated digital management platform after study completion, regardless of study group.

Results

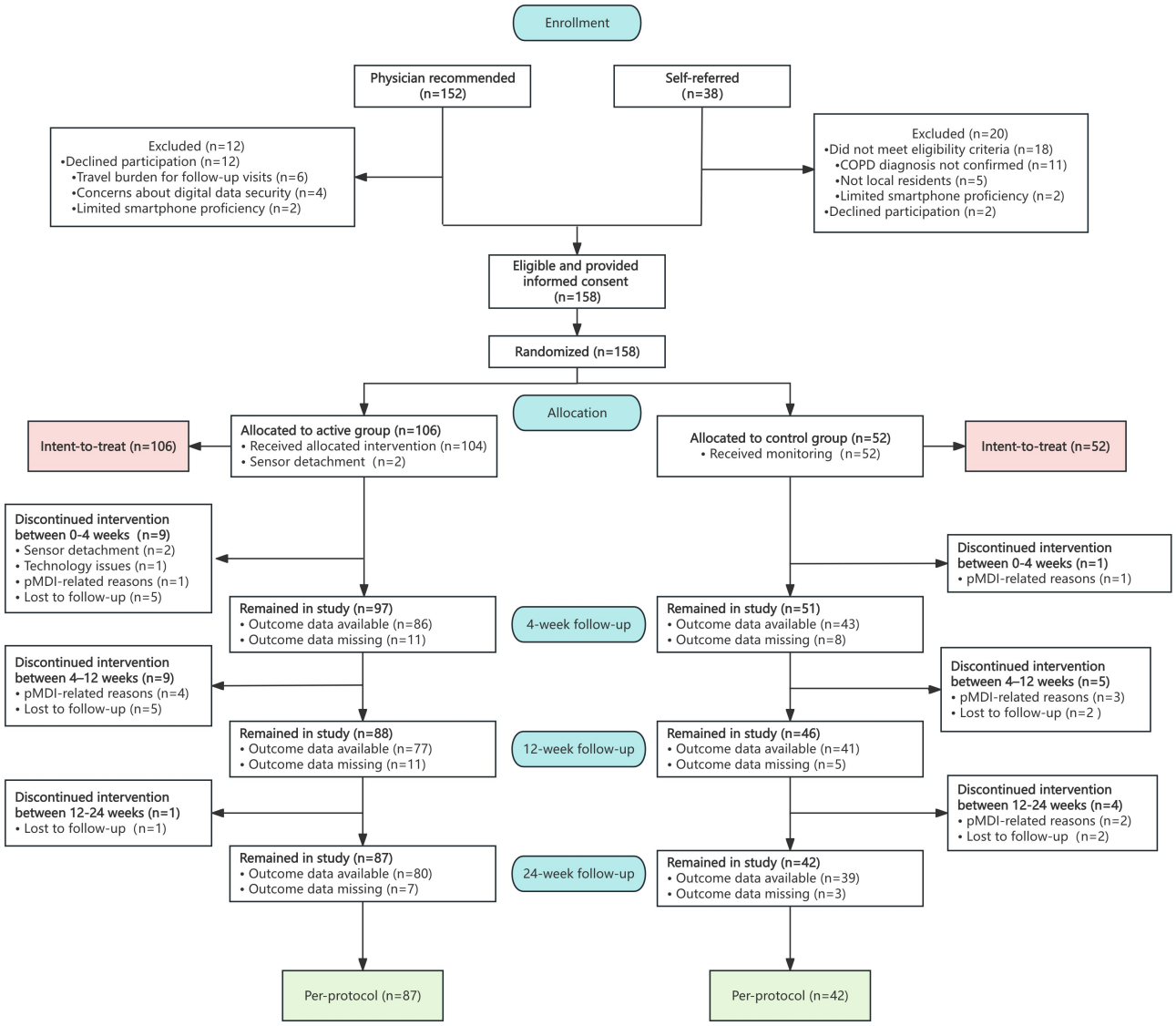
Participant Characteristics

Participant recruitment was conducted between March 6, 2023, and January 31, 2024, and follow-up was completed on July 17, 2024. In the context of the COVID-19 pandemic during the recruitment period, a total of 158 eligible participants provided written informed consent by the end of the predefined study period, with 106 assigned to the intervention group and 52 to the control group. Participants continued their physician-prescribed inhaled therapy throughout the study. A total of 19 participants in the intervention group and 10 in the control group withdrew during the study, of whom 11 discontinued due to pMDI cessation (Figure 2). Among the remaining participants, 2 in the intervention group (1 switched from dual bronchodilation to triple therapy and 1 switched from triple therapy to dual bronchodilation) and 1 in the control group (switched from dual bronchodilation to triple therapy) had their inhaled regimens switched by their physicians during the study, without changing the pMDI device. No other changes in inhaled therapy were documented.

Baseline characteristics were similar between groups. A total of 129 (81.6%) participants were male, and the mean age was 64 (SD 8.30) years. The mean number of acute exacerbations in the previous year (including mild, moderate, and severe) was 0.95 (SD 1.34), and 72.2% (n=114) of participants were classified as having good adherence based on the self-reported TAI questionnaire (Table 1).

XSL•FO
RenderX

Figure 2. Flow diagram of the study. COPD: chronic obstructive pulmonary disease; pMDI: pressurized metered-dose inhaler.



XSL•FO
RenderX

Table . Baseline characteristics of intervention and control participants (N=158).

Variables	All participants (N=158)	Intervention group (n=106)	Control group (n=52)	*P* value
Age (y), mean (SD)	64.08 (8.30)	63.30 (8.74)	65.65 (7.16)	.09
Sex (male), n (%)	129 (81.6)	85 (80.2)	44 (84.6)	.66
BMI (kg/m^2), mean (SD)	24.33 (4.03)	24.49 (4.03)	23.98 (4.02)	.45
Education level, n (%)				.90
No formal education	6 (3.7)	5 (4.7)	1 (1.9)	
Primary school	16 (10.1)	10 (9.4)	6 (11.5)	
Junior high school	52 (32.9)	35 (33.0)	17 (32.6)	
Senior high school	43 (27.2)	27 (25.4)	16 (30.7)	
College/associates degree	29 (18.3)	20 (18.8)	9 (17.3)	
Postgraduate degree	12 (7.5)	9 (8.4)	3 (5.7)	
Health insurance type, n (%)				.73
Urban employee basic	86 (54.4)	56 (52.8)	30 (57.6)	
New rural cooperative	35 (22.1)	22 (20.7)	13 (25.0)	
Urban resident basic	24 (15.1)	19 (17.9)	5 (9.6)	
Urban and rural resident	3 (1.8)	2 (1.8)	1 (1.9)	
Private	10 (6.3)	7 (6.6)	3 (5.7)	
Smoking status, n (%)				.40
Never	42 (26.6)	31 (29.2)	11 (21.2)	
Former	65 (41.1)	40 (37.7)	25 (48.1)	
Current	51 (32.3)	35 (33.0)	16 (30.8)	
Smoking pack-year history for smokers, mean (SD)	19.49 (5.30)[a]	19.49 (5.67)[b]	19.49 (4.64)[c]	.10
FVC[d] (L), mean (SD)	3.07 (0.90)	3.09 (0.90)	3.03 (0.88)	.71
FEV_1[e] (L), mean (SD)	1.68 (0.70)	1.71 (0.71)	1.63 (0.68)	.51
FEV_1 predictive (%), mean (SD)	60.66 (22.96)	60.59 (8.74)	60.78 (23.86)	.96
FEV_1/FVC (%), mean (SD)	53.88 (13.08)	54.24 (12.39)	53.14 (14.45)	.63
PEF[f] (L/s), mean (SD)	4.70 (2.17)	4.79 (2.11)	4.52 (2.28)	.48
GOLD[g] stage, n (%)				.71
I	30 (18.9)	18 (16.9)	12 (23.0)	
II	66 (41.7)	47 (44.3)	19 (36.5)	
III	41 (25.9)	27 (25.4)	14 (26.9)	
IV	13 (8.2)	8 (7.5)	5 (9.6)	
Prior year exacerbation history				
Mean (SD)	0.95 (1.34)	1.00 (1.39)	0.85 (1.26)	.42
At least 1, n (%)	70 (44.3)	49 (46.2)	21 (40.4)	.49
Inhaled prescription, n (%)				.92
ICS[h]/LAMA[i]/LABA[j]	128 (81.0)	85 (80.2)	42 (80.8)	
LAMA/LABA	30 (19.0)	21 (19.8)	10 (19.2)	
mMRC[k] score ≥2, n (%)	76 (48.1)	51 (48.1)	25 (48.1)	.10
CAT[l] score ≥10, n (%)	59 (37.3)	40 (37.7)	19 (36.5)	.88

XSL•FO
RenderX

Variables	All participants (N=158)	Intervention group (n=106)	Control group (n=52)	*P* value
TAI[m] score				
Mean (SD)	45.96 (6.84)	45.74 (7.09)	45.60 (6.36)	.90
≥46, n (%)	114 (72.2)	78 (73.6)	36 (69.2)	.57
SGRQ-C[n] total score, mean (SD)	38.70 (19.18)	38.83 (18.34)	38.42 (20.98)	.90
PHQ-9[o] score				
Mean (SD)	1.56 (2.95)	1.68 (3.25)	1.33 (2.23)	.48
≤4, n (%)	143 (90.5)	96 (90.6)	47 (90.4)	>.99
GAD-7[p] score				
Mean (SD)	0.87 (2.16)	0.79 (2.09)	1.04 (2.31)	.50
≤4, n (%)	149 (94.3)	101 (95.3)	48 (92.3)	.69
Comorbidity, n (%)				
None	27 (17.0)	17 (16.0)	10 (19.2)	.62
Tuberculosis	11 (6.9)	8 (7.5)	3 (5.7)	.71
Cardiovascular disease	25 (15.8)	18 (16.9)	7 (13.4)	.62
Hypertension	61 (38.6)	42 (39.6)	19 (36.5)	.74
Diabetes	18 (11.3)	14 (13.2)	4 (7.6)	.29
Osteoporosis	14 (8.8)	9 (8.4)	5 (9.6)	.74
Anxiety or depressive disorder	3 (1.8)	2 (1.8)	1 (1.9)	>.99

[a]n=116.
[b]n=75.
[c]n=41.
[d]FVC: forced vital capacity.
[e]FEV_1: forced expiratory volume in 1 second.
[f]PEF: peak expiratory flow rate.
[g]GOLD: Global Initiative for Chronic Obstructive Lung Disease.
[h]ICS: inhaled corticosteroid.
[i]LAMA: long-acting muscarinic antagonist.
[j]LABA: long-acting beta-agonist.
[k]mMRC: modified Medical Research Council.
[l]CAT: COPD Assessment Test.
[m]TAI: Test of Adherence to Inhalers.
[n]SGRQ-C: St George's Respiratory Questionnaire for Chronic Obstructive Pulmonary Disease.
[o]PHQ-9: Patient Health Questionnaire–9.
[p]GAD-7: Generalized Anxiety Disorder–7.

Adherence (Primary Outcome)

Except for 2 participants in the intervention group with sensor detachment, 156 participants had at least one valid daily adherence record. The median number of valid monitoring days was 161.5 (IQR 86.3 - 168.0). TAI scores were collected at 3 follow-up visits, with response rates of 82.9% (n=131; 88 intervention vs 43 control), 75.3% (n=119; 77 vs 42), and 79.1% (n=125; 85 vs 40). No statistically significant differences were observed between groups in the proportion of participants classified as having good TAI adherence at any visits (Table 2). Linear mixed-effects modeling indicated that neither the main effect of group (*P*=.64), time within group (*P*=.051), nor the group × time interaction (*P*=.66) reached statistical significance.

XSL•FO
RenderX

Table . Self-reported and electronically measured adherence at weeks 0 to 4, 4 to 12, and 12 to 24 (N=158).

Adherence	0 to 4 weeks			4 to 12 weeks			12 to 24 weeks		
	Intervention group (n=106)	Control group (n=52)	*P* value	Intervention group (n=106)	Control group (n=52)	*P* value	Intervention group (n=106)	Control group (n=52)	*P* value
TAI[a] score ≥46, n (%)	92 (86.8)	41 (78.8)	.20	88 (83.0)	42 (80.0)	.73	77 (72.6)	38 (73.1)	.95
Electronic adherence[b] (%), median (IQR)	85.27 (55.36 - 95.54)	76.79 (46.65 - 85.49)	.04[c]	74.11 (31.70 - 89.96)	31.47 (8.59 - 76.00)	.004[c]	43.01 (2.53 - 87.57)	11.90 (0.00 - 71.73)	.03[c]

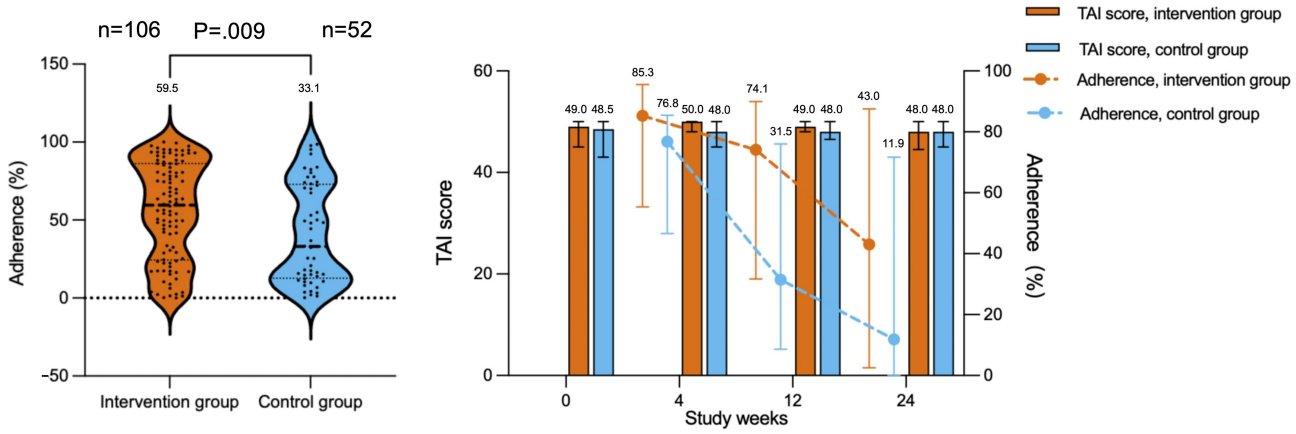
[a]TAI: Test of Adherence to Inhalers. TAI was assessed at weeks 4, 12, and 24, with each assessment reflecting self-reported adherence during the preceding 1 week.

[b]Electronic adherence was continuously recorded by the electronic inhaler monitoring and averaged over the corresponding intervals (weeks 0 - 4, 4 - 12, and 12 - 24).

[c]Statistically significant values (*P*<.05).

Over the 24-week study period, the electronically measured medication adherence was significantly higher in the intervention group than the control group (59.52% vs 33.11%; *P*=.009; Figure 3A). The between-group difference increased over time: 85.27% vs 76.79% during weeks 0 to 4 (*P*=.04), 74.11% vs 31.47% during weeks 4 to 12 (*P*=.004), and 43.01% vs 11.90% during weeks 12 to 24 (*P*=.03; Table 2 and Figure 3B). The proportion of days with 100% of prescribed inhalations taken was also higher in the intervention group than in the control group (31.5% vs 16.7%; *P*=.007). Linear mixed-effects modeling revealed a significant main effect of group (*P*=.004) and time within groups (*P*<.001), whereas the group × time interaction did not reach statistical significance (*P*=.24).

Figure 3. Adherence (median and IQR) over 24 weeks. (A) Violin plot of electronically measured adherence averaged over 24 weeks. (B) Test of Adherence to Inhalers (TAI) score at baseline, weeks 4, 12, and 24 and electronically measured adherence averaged over weeks 0 to 4, 4 to 12, and 12 to 24.



COPD Assessments (Secondary Outcomes)

No statistically significant between-group differences were observed for SGRQ-C scores, PHQ-9 score, GAD-7 score, mMRC grade, CAT score, or lung function at week 24 (Table 3).

No significant between-group differences were detected for continuous CAT and mMRC scores either; however, longitudinal analyses showed significant changes over time in both groups (main effect of time, both *P*<.001). Compared with baseline, the mean change in CAT score was −7.09 (95% CI −10.55 to −3.62) at week 4,−8.60 (95% CI −12.03 to −5.17) at week 12, and −8.72 (95% CI −12.18 to −5.25) at week 24 (all *P*<.001). The mean change in mMRC grade was −0.59 (95% CI −0.97 to −0.22), −0.87 (95% CI −1.25 to −0.50), and −0.93 (95% CI −1.32 to −0.54) at weeks 4, 12, and 24 (Figure 4). Exploratory generalized estimating equation analysis showed that the intervention group had higher odds of achieving the minimal clinically important difference in CAT scores over follow-up (odds ratio 2.21, 95% CI 1.11 - 4.41; *P*=.02), with a significant group × time interaction (*P*<.001). However, no individual follow-up time point remained statistically significant after Bonferroni correction for multiple comparisons (Figure 5).

XSL•FO
RenderX

Table . Secondary outcomes at week 24.

Variables	Intervention group, β[a] (95% CI)	Control group, β (95% CI)	*P* value
SGRQ-C[b] total score	20.29 (18.34 to 22.23)	23.38 (20.51 to 26.25)	.08
SGRQ-C symptoms score	34.89 (31.46 to 38.33)	40.10 (35.03 to 45.17)	.10
SGRQ-C activity score	29.02 (26.01 to 32.02)	31.49 (27.06 to 35.93)	.36
SGRQ-C impact score	10.07 (8.41 to 11.74)	12.84 (10.38 to 15.30)	.07
PHQ-9[c] score	0.75 (0.29 to 1.20)	0.22 (−0.46 to 0.90)	.21
GAD-7[d] score	0.50 (0.06 to 0.94)	0.09 (−0.56 to 0.75)	.31
mMRC[e] score	0.987 (0.86 to 1.12)	0.926 (0.73 to 1.12)	.31
CAT[f] score	6.36 (5.57 to 7.15)	7.11 (5.95 to 8.28)	.29
FVC[g] (L)	3.25 (3.13 to 3.38)	3.16 (2.98 to 3.33)	.37
FEV_1[h] (L)	1.88 (1.78 to 1.98)	1.77 (1.63 to 1.91)	.22
FEV_1 predictive (%)	65.25 (61.93 to 68.58)	62.26 (57.51 to 67.02)	.31
FEV_1/FVC (%)	57.5 (55.0 to 60.0)	54.6 (51.0 to 58.2)	.19
PEF[i] (L/s)	5.53 (5.21 to 5.85)	5.24 (4.78 to 5.69)	.30

[a]β: regression coefficient.

[b]SGRQ-C: St George's Respiratory Questionnaire for Chronic Obstructive Pulmonary Disease.

[c]PHQ-9: Patient Health Questionnaire-9.

[d]GAD-7: Generalized Anxiety Disorder-7.

[e]mMRC: modified Medical Research Council.

[f]CAT: COPD Assessment Test.

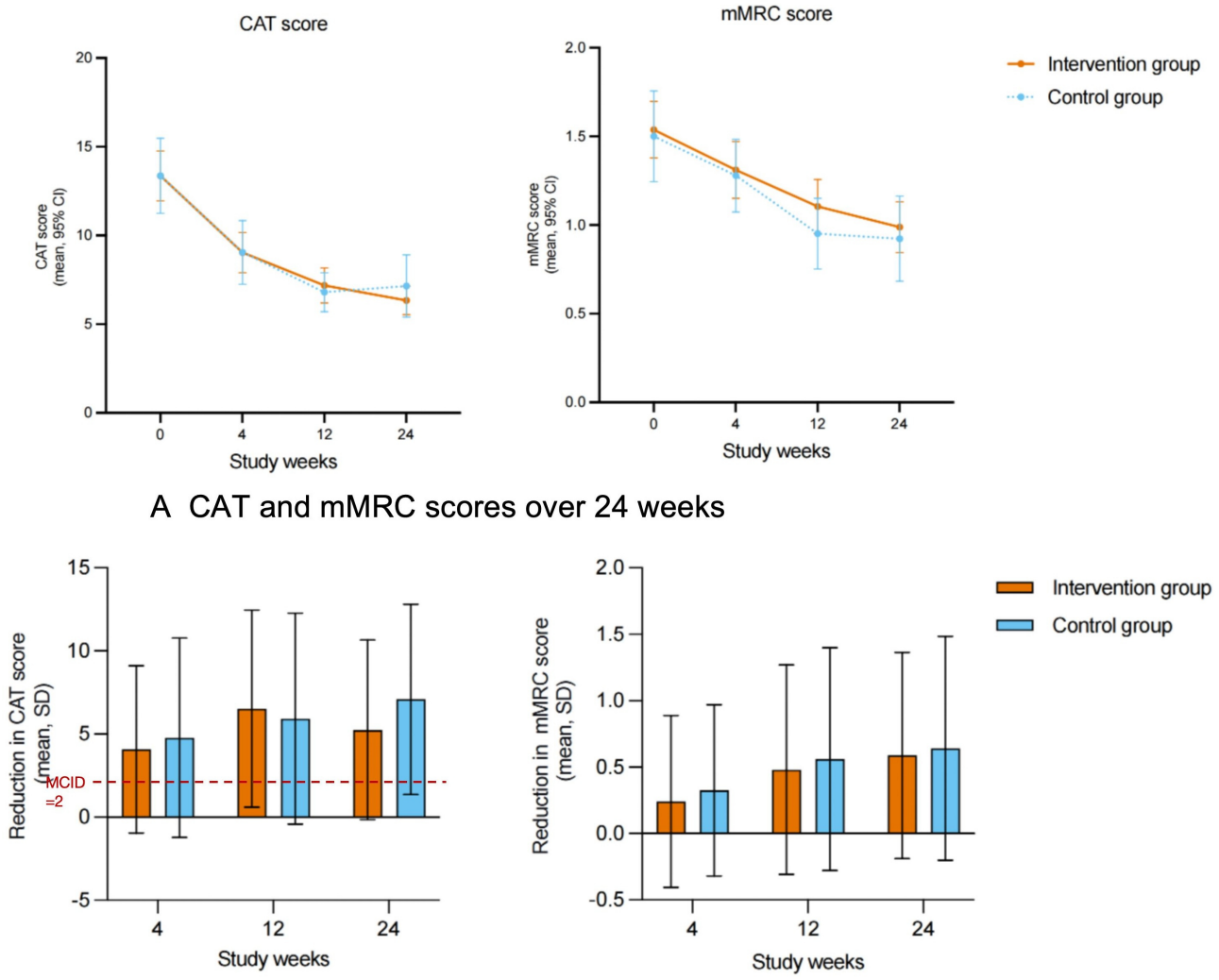
[g]FVC: forced vital capacity.

[h]FEV_1: forced expiratory volume in 1 second.

[i]PEF: peak expiratory flow rate.

XSL•FO
RenderX

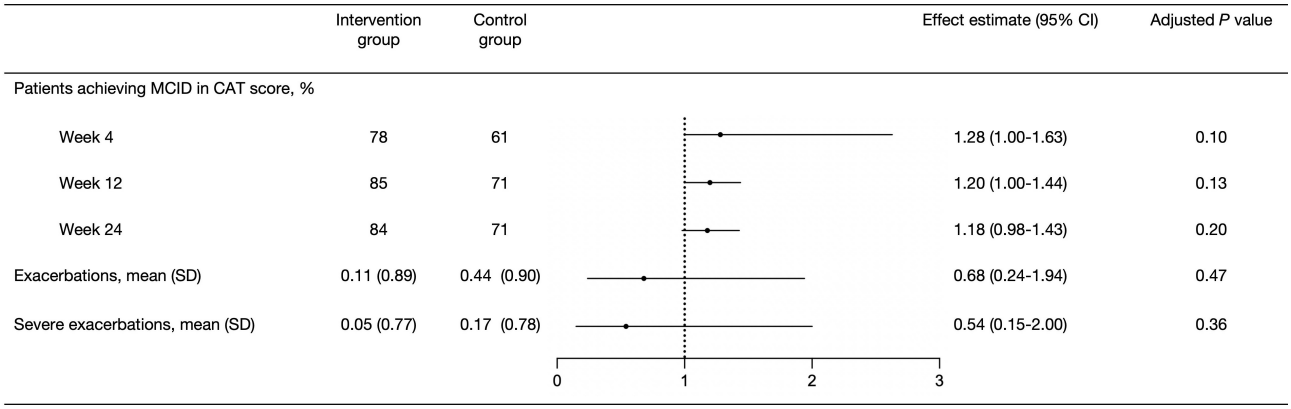
Figure 4. (A) COPD Assessment Test (CAT) and modified Medical Research Council (mMRC) scores over 24 weeks. Mean scores at weeks 4, 12, and 24. (B) Mean reduction from baseline at weeks 4, 12, and 24. COPD: chronic obstructive pulmonary disease; MCID: minimal clinically important difference.



During the 24-week follow-up, the mean number of moderate-to-severe exacerbations was 0.11 (SD 0.89) in the intervention group and 0.44 (SD 0.90) in the control group, with 8 (7.5%) and 10 (19.2%) participants, respectively, experiencing at least one event. After adjustment for baseline exacerbation history, negative binomial regression showed no statistically significant between-group differences in moderate-to-severe exacerbations (incidence rate ratio 0.68, 95% CI 0.24 - 1.94; *P*=.47) or severe exacerbations (incidence rate ratio 0.54, 95% CI 0.15 - 2.00; *P*=.36; Figure 5). Baseline exacerbation history was a strong predictor of subsequent events, with each additional exacerbation at baseline associated with a 3.69-fold higher risk of exacerbation at 24 weeks (risk ratio 3.69, 95% CI 2.37 - 5.75; *P*<.001).

XSL•FO
RenderX

Figure 5. Effect of the intervention on COPD Assessment Test (CAT) and exacerbations. COPD: chronic obstructive pulmonary disease; MCID: minimal clinically important difference.



Post Hoc Analyses (Exploratory Outcomes)

To further explore potential factors associated with adherence and clinical outcomes, we conducted post hoc exploratory correlation analyses, subgroup analyses, and multivariable regression.

Exploratory Pearson correlation analysis of the main outcome and baseline information showed that the average adherence of the intervention group was positively correlated with age (Pearson correlation coefficient, r=0.31; P=.03) and the presence of an acute exacerbation in the past year (r=0.225, P=.02), whereas no significant correlation was observed in the control group (r=0.049; P=.28).

Exploratory subgroup analyses for secondary outcomes were conducted according to age (<60 vs≥60 y), sex, GOLD (Global Initiative for Chronic Obstructive Lung Disease) stage, baseline TAI score (<46 vs ≥ 46), baseline mMRC grade (<2 vs≥2), history of exacerbations (0 vs ≥1), and electronic adherence (<80% vs ≥80%). For acute exacerbations at week 24, potential effect modification was observed by mMRC grade (P=.02), GOLD stage (P=.03), and adherence (P=.045). Patients with mMRC ≥2, GOLD stage 3, or adherence ≥80% showed greater reductions with the intervention (mean differences −0.68, 95% CI −1.10 to −0.26; P=.002; −1.75, 95% CI −2.72 to −0.78; P<.001; and −0.79, 95% CI −1.32 to −0.26; P=.004) compared with the respective comparator subgroups. No significant differential response pattern was observed for the other subgroups.

To identify characteristics associated with high adherence (adherence ≥80%), we performed univariate logistic regression analyses using baseline variables. Eleven variables were associated with high adherence (P<.20 or area under the curve [AUC]>0.6). Among these, GOLD group (AUC=0.631; P=.03) and BMI (AUC=0.614; P=.03) exhibited the highest predictive ability. A multivariable logistic model incorporating these 2 variables yielded a mean cross-validated AUC of 0.611 (SD 0.046) after 5-fold cross-validation, with a sensitivity of 31.7% and a specificity of 83.3%.

Engagement and Satisfaction

The EIM demonstrated satisfactory performance in tracking medication adherence, but the completion rates of actively reported measures via the patient mobile app were notably low. During the 168-day follow-up period, the average completion rate for the daily CERT diary was only 7.0% (SD 23.7%), accompanied by a 2.3% (SD 16.5%) completion rate for body temperature logs and 5.3% (SD 23.2%) for oxygen saturation recordings. This finding may be attributed to the predominantly older adult study population, who generally exhibit lower proficiency and frequency in smartphone use. These results underscore the need for future digital therapeutic tools to incorporate more automated and user-friendly symptom monitoring modalities that minimize reliance on active patient input. Despite the low engagement with active app features, the end-of-study user satisfaction survey revealed favorable attitudes toward the digital therapeutic software. Among the 158 participants, 101 (63.9%) reported being “very satisfied” with the device, and 98 (62.0%) rated it as very convenient to use.

Adverse Events

No adverse events were reported during the study.

Discussion

This randomized controlled study provides evidence supporting the feasibility and preliminary effectiveness of a digital therapeutic intervention in Chinese patients with COPD. Participants receiving the digital therapeutic intervention integrating EIM with a digital management platform demonstrated significantly higher objectively measured inhaler adherence compared with electronic monitoring alone over the 24-week study period. However, improvements in clinical outcomes, including exacerbations, health-related quality of life, symptoms, and lung function, did not reach statistical significance. These findings suggest that digital therapeutics may be effective in improving medication-taking behavior, whereas larger and longer studies are needed to determine their impact on clinical outcomes.

Our results are consistent with previous studies in populations with asthma, in which electronic monitoring revealed adherence gaps not captured by patient self-report, emphasizing the value of objective, technology-enabled assessment in chronic airway disease management [21-24]. Despite the older age profile of patients with COPD relative to typical asthma cohorts, our results suggest that older adult patients can effectively engage with digital inhaler technologies. However, we observed a gradual decline in adherence over time despite digital support,

XSL•FO
RenderX

suggesting that maintaining long-term behavioral change remains challenging. This finding highlights the value of regular, in-person clinical follow-up alongside digital tools to maintain long-term adherence [25]. Thus, we conducted exploratory telephone follow-ups in December 2025 to provide preliminary information regarding the longer-term implementation of the intervention. A total of 128 patients participated in this follow-up, of whom 91 continued to use inhaled medications, and 87 reported daily adherence in the past week. Future studies with predefined long-term follow-up protocols are needed to determine whether digital therapeutics can produce sustained behavioral and clinical benefits.

Notably, electronically measured adherence and self-reported adherence showed marked divergence. This discrepancy may be partly attributable to differences in the assessment windows: TAI assessed self-reported adherence during the preceding 1 week at prespecified follow-up visits, whereas EIM continuously captured inhaler use and summarized adherence over longer follow-up intervals. However, the magnitude of the observed difference suggests that the assessment window alone does not fully account for the divergence. Self-reported adherence is susceptible to recall bias and social desirability bias, as patients may overestimate their medication use during clinic visits. In addition, TAI includes items assessing medication-taking behavior as well as behavioral and perceptual factors related to nonadherence, whereas EIM primarily captures only the occurrence and timing of inhaler use. These findings highlight the complementary value of both objective monitoring and subjective self-report in assessing adherence and evaluating adherence-promoting strategies.

After adjusting for baseline values, the intervention group showed numerically better outcomes in exacerbation frequency, symptom scores, and lung function, though none reached statistical significance. These findings likely reflect the study being powered for adherence rather than clinical outcomes, moderate baseline disease severity, low exacerbation rates, and a relatively short 24-week follow-up. Additionally, CAT and mMRC scores improved over time in both groups. Given that both groups received regular follow-up and study participation itself may increase disease awareness and self-management behaviors, these improvements may partly reflect nonspecific effects, including increased attention or Hawthorne effects, rather than the isolated effect of the digital therapeutic intervention. Future studies incorporating attention-matched control conditions, factorial intervention designs, and longer follow-up are needed to better determine the independent contribution of specific digital therapeutic components. Although no statistically significant differences were observed in mental health measures, slightly higher PHQ-9 and GAD-7 scores in the intervention group suggest a need to monitor psychological burden associated with digital interventions. Previous literature indicates that higher eHealth literacy can improve quality of life, sleep quality, and resilience but may also be associated with increased anxiety and depression, highlighting the need for careful monitoring of potential adverse effects on mental health [26].

Subgroup analysis suggested that the effect of the digital therapeutic on acute exacerbations may be more pronounced in patients with greater symptom burden, more severe disease, and higher adherence. The observed reduction in acute exacerbations among patients with higher adherence suggests that improved medication adherence may represent one potential pathway through which digital therapeutics could influence clinical outcomes [27]. However, these findings should be interpreted cautiously, given the exploratory nature of the analyses, and require confirmation in future adequately powered studies. Multivariable regression analysis showed that the baseline GOLD grade and BMI modestly predicted high adherence. This indicates that, in clinical practice, adherence cannot be reliably predicted based solely on basic physiological or disease severity indicators. Dynamic factors during the intervention, such as patient engagement, social or family support, digital literacy, or behavioral patterns, may be stronger determinants of adherence than static baseline characteristics. Therefore, future research should integrate multidimensional assessments to better capture the complexity of medication adherence behavior and to elucidate the mechanisms underlying responses to digital therapeutics, including how specific functional designs influence patient behavioral change and how factors such as patients' behavioral patterns, cognitive characteristics, and technology acceptance modulate intervention outcomes. By integrating behavioral science theories with digital technologies to develop more targeted intervention modules, we may further enhance the clinical efficacy of digital therapeutics and facilitate their evolution from a "one-size-fits-all" approach toward precision and personalized health management [28,29].

This study has several limitations. First, this was a single-center study recruiting through physician recommendation and patient self-referral. This may have introduced selection bias as individuals who agreed to participate were predominantly urban and educated and may have been more motivated and receptive to digital health technologies. In addition, the open-label design resulted in unequal levels of participant interaction between groups, and improvements in subjective outcomes may have been partly influenced by increased attention, behavioral reinforcement, or Hawthorne effects. These factors may limit the generalizability of the findings and may also have attenuated between-group differences in secondary clinical outcomes, as participants may have represented a more engaged subgroup of the population with COPD. Second, the EIM used is compatible only with pMDI devices and provides temporal reminders without guidance on inhalation technique. Third, the reliance on smartphone use among mostly older adult participants may exacerbate health inequities in underserved populations. Fourth, low completion rates of actively reported symptom diaries limited early detection of exacerbations. In addition, because the CERT diary was available only to the intervention group, the 2 groups differed with respect to the availability of systematic symptom data between scheduled visits. This differential access reflects the intended design of the digital therapeutic program; however, the very low completion rates in the intervention group suggest that the actual impact of this differential monitoring on clinical management may have been limited. Nonetheless, we cannot exclude the possibility that even sporadic symptom reports contributed to clinical awareness and earlier intervention in some cases. Fifth, the 6-month follow-up may be insufficient to capture long-term behavioral

XSL•FO
RenderX

or clinical outcomes. Finally, health care resource use and cost-effectiveness data were not collected, precluding economic evaluation.

In summary, this study demonstrates that a digital therapeutic program can significantly improve inhaler adherence among Chinese patients with COPD, even in older adults, while the translation into short-term clinical benefit remains uncertain. These findings support the feasibility of digital therapeutics and highlight the need for longer-term studies incorporating automated, user-friendly monitoring and targeted behavioral strategies to maximize clinical impact. Future research should focus on identifying patient subgroups most likely to benefit and integrating behavioral science with digital health design to optimize personalized COPD management [7,28,29].

Generative AI Use

The authors declare that no generative AI tools were used during manuscript preparation.

Funding

This study was funded by the National High Level Hospital Clinical Research Funding, Elite Medical Professionals Project of China-Japan Friendship Hospital (number ZRJY2023-QM16), Noncommunicable Chronic Diseases–National Science and Technology Major Project (numbers 2023ZD0516300, 2023ZD0506000, and 2024ZD0528501), National Natural Science Foundation of China (number 82400043), and National Key Research and Development Program of China (number 2022YFC3702704).

Data Availability

Deidentified participant data will be available upon request from the authors, including a data dictionary.

Authors' Contributions

TY and HN initiated the study and obtained funding for this study. HW, HN, and FD developed the initial study protocol. DH, Y Cui, Y Chen, and QY contributed to the design and implementation of the study. HW, YZ, TP, YP, Xiang Li, YL, and Xiaopan Li conducted the study. HW conducted the analyses and drafted the manuscript. HN and TY provided critical feedback and revised sections of the manuscript. All authors provided critical revision of the paper and approved the final manuscript.

Conflicts of Interest

None declared.

Checklist 1
CONSORT checklist.
[PDF File, 1290 KB - mhealth_v14i1e98870_app1.pdf]

References

1. GBD 2023 Causes of Death Collaborators. Global burden of 292 causes of death in 204 countries and territories and 660 subnational locations, 1990-2023: a systematic analysis for the Global Burden of Disease Study 2023. Lancet 2025 Oct 18;406(10513):1811-1872. [doi: 10.1016/S0140-6736(25)01917-8] [Medline: 41092928]
2. Wang C, Xu J, Yang L, et al. Prevalence and risk factors of chronic obstructive pulmonary disease in China (the China Pulmonary Health [CPH] study): a national cross-sectional study. Lancet 2018 Apr 28;391(10131):1706-1717. [doi: 10.1016/S0140-6736(18)30841-9] [Medline: 29650248]
3. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: 2024 report. : Global Initiative for Chronic Obstructive Lung Disease (GOLD); 2024 URL: https://goldcopd.org/wp-content/uploads/2024/02/GOLD-2024_v1.2-11Jan24_WMV.pdf [accessed 2026-10-17]
4. Moradkhani B, Mollazadeh S, Niloofar P, Bashiri A, Oghazian MB. Association between medication adherence and health-related quality of life in patients with chronic obstructive pulmonary disease. J Pharm Health Care Sci 2021 Nov 15;7(1):40. [doi: 10.1186/s40780-021-00222-x] [Medline: 34775992]
5. Vestbo J, Anderson JA, Calverley PMA, et al. Adherence to inhaled therapy, mortality and hospital admission in COPD. Thorax 2009 Nov;64(11):939-943. [doi: 10.1136/thx.2009.113662] [Medline: 19703830]
6. Bhattarai B, Walpola R, Mey A, Anoopkumar-Dukie S, Khan S. Barriers and strategies for improving medication adherence among people living with COPD: a systematic review. Respir Care 2020 Nov;65(11):1738-1750. [doi: 10.4187/respcare.07355] [Medline: 32576706]
7. Ma J, Sun X, Wang X, Liu B, Shi K. Factors affecting patient adherence to inhalation therapy: an application of SEIPS model 2.0. Patient Prefer Adherence 2023;17:531-545. [doi: 10.2147/PPA.S395327] [Medline: 36896268]
8. Kini V, Ho PM. Interventions to improve medication adherence: a review. JAMA 2018 Dec 18;320(23):2461-2473. [doi: 10.1001/jama.2018.19271] [Medline: 30561486]

XSL•FO
RenderX

9. Dhruve H, Jackson DJ. Assessing adherence to inhaled therapies in asthma and the emergence of electronic monitoring devices. Eur Respir Rev 2022 Jun 30;31(164):210271. [doi: 10.1183/16000617.0271-2021] [Medline: 35613744]
10. Watson A, Wilkinson TMA. Digital healthcare in COPD management: a narrative review on the advantages, pitfalls, and need for further research. Ther Adv Respir Dis 2022;16:17534666221075493. [doi: 10.1177/17534666221075493] [Medline: 35234090]
11. Dierick BHJ, Achterbosch M, Been-Buck S, et al. Can electronic monitoring with a digital smart spacer support personalised medication adherence and inhaler technique education in patients with asthma?: Protocol of the randomised controlled OUTERSPACE trial. BMJ Open 2022 Jun 13;12(6):e059929. [doi: 10.1136/bmjopen-2021-059929] [Medline: 35697450]
12. Dierick BJH, Been-Buck S, Klemmeier T, et al. Digital spacer data driven COPD inhaler adherence education: the OUTERSPACE proof-of-concept study. Respir Med 2022 Sep;201:106940. [doi: 10.1016/j.rmed.2022.106940] [Medline: 35933835]
13. Merchant R, Szefler SJ, Bender BG, et al. Impact of a digital health intervention on asthma resource utilization. World Allergy Organ J 2018;11(1):28. [doi: 10.1186/s40413-018-0209-0] [Medline: 30524644]
14. O'Dwyer S, Greene G, MacHale E, et al. Personalized biofeedback on inhaler adherence and technique by community pharmacists: a cluster randomized clinical trial. J Allergy Clin Immunol Pract 2020 Feb;8(2):635-644. [doi: 10.1016/j.jaip.2019.09.008] [Medline: 31568927]
15. Gregoriano C, Dieterle T, Breitenstein AL, et al. Does a tailored intervention to promote adherence in patients with chronic lung disease affect exacerbations? A randomized controlled trial. Respir Res 2019 Dec 3;20(1):273. [doi: 10.1186/s12931-019-1219-3] [Medline: 31796013]
16. Rumi G, Canonica GW, Foster JM, et al. Digital coaching using smart inhaler technology to improve asthma management in patients with asthma in Italy: community-based study. JMIR Mhealth Uhealth 2022 Nov 2;10(11):e25879. [doi: 10.2196/25879] [Medline: 36322120]
17. Charles T, Quinn D, Weatherall M, Aldington S, Beasley R, Holt S. An audiovisual reminder function improves adherence with inhaled corticosteroid therapy in asthma. J Allergy Clin Immunol 2007 Apr;119(4):811-816. [doi: 10.1016/j.jaci.2006.11.700] [Medline: 17320942]
18. Plaza V, Fernández-Rodríguez C, Melero C, et al. Validation of the "Test of the Adherence to Inhalers" (TAI) for asthma and COPD patients. J Aerosol Med Pulm Drug Deliv 2016 Apr;29(2):142-152. [doi: 10.1089/jamp.2015.1212] [Medline: 26230150]
19. Yildiz E, Çeçen S, Lafci D, Eker A. The effect of repeated video-training session on metered doseinhaler skills among individuals with COPD: prospective randomised controlled study. J Clin Nurs 2022 May;31(9-10):1339-1347. [doi: 10.1111/jocn.15990] [Medline: 34352135]
20. Jones PW, Wang C, Chen P, et al. The development of a COPD Exacerbation Recognition Tool (CERT) to help patients recognize when to seek medical advice. Int J Chron Obstruct Pulmon Dis 2022;17:213-222. [doi: 10.2147/COPD.S337644] [Medline: 35087270]
21. Foster JM, Usherwood T, Smith L, et al. Inhaler reminders improve adherence with controller treatment in primary care patients with asthma. J Allergy Clin Immunol 2014 Dec;134(6):1260-1268. [doi: 10.1016/j.jaci.2014.05.041] [Medline: 25062783]
22. Mosnaim GS, Stempel DA, Gonzalez C, et al. The impact of patient self-monitoring via electronic medication monitor and mobile app plus remote clinician feedback on adherence to inhaled corticosteroids: a randomized controlled trial. J Allergy Clin Immunol Pract 2021 Apr;9(4):1586-1594. [doi: 10.1016/j.jaip.2020.10.064] [Medline: 33212239]
23. Kaye L, Gondalia R, Thompson A, Stempel DA, Barrett MA. The relationship between objective app engagement and medication adherence in asthma and COPD: a retrospective analysis. Sci Rep 2021 Dec 21;11(1):24343. [doi: 10.1038/s41598-021-03827-2] [Medline: 34934164]
24. Mosnaim G, Safioti G, Brown R, et al. Digital health technology in asthma: a comprehensive scoping review. J Allergy Clin Immunol Pract 2021 Jun;9(6):2377-2398. [doi: 10.1016/j.jaip.2021.02.028] [Medline: 33652136]
25. Wu R, Zhou J, Mariakakis A, de Lara E, Jeyaparan J, Gershon A. Determining what patients admitted with a chronic obstructive pulmonary disease (COPD) exacerbation will use for remote clinical monitoring: a patient engagement survey. BMJ Open Respir Res 2025 Jul 21;12(1):e002841. [doi: 10.1136/bmjresp-2024-002841] [Medline: 40695531]
26. Li J, Wu X, Wu Y, et al. Physical, mental, and health empowerment disparities across chronic obstructive pulmonary disease, asthma, and combined groups and the moderating role of eHealth literacy: cross-sectional study. J Med Internet Res 2025;27:e70822. [doi: 10.2196/70822] [Medline: 40324764]
27. Aksu K, Vural Solak GT, Mutlu LC, et al. Demographic and asthma-related characteristics of asthmatics using pressurized metered dose inhalers and dry powder inhalers. J Aerosol Med Pulm Drug Deliv 2024 Dec;37(6):346-350. [doi: 10.1089/jamp.2024.0006] [Medline: 39075039]
28. Adejumo I, Patel M, McKeever TM, Shaw DE. Does inhaler technology improve adherence and asthma control? A pilot randomized controlled trial. Ann Allergy Asthma Immunol 2022 Jun;128(6):727-729. [doi: 10.1016/j.anai.2022.02.023] [Medline: 35257874]

XSL•FO
RenderX

29. Drummond MB, Hemphill CC, Hill T, Boe A, Yu D, Ohar JA. Use of a digital inhaler to assess COPD disease variability and identify impending acute COPD exacerbations: a pilot study. Chronic Obstr Pulm Dis 2025 May 27;12(3):250-259. [doi: 10.15326/jcopdf.2024.0555] [Medline: 40459430]

Abbreviations

AUC: area under the curve
CAT: COPD Assessment Test
CERT: COPD Exacerbation Recognition Tool
CONSORT: Consolidated Standards of Reporting Trials
COPD: chronic obstructive pulmonary disease
DIGI-COPD: Effectiveness of a Multi-Component Digital Therapeutic Program on Inhaled Medication Adherence in COPD
EIM: electronic inhaler monitoring
FEV_1/FVC: forced expiratory volume in 1 second/forced vital capacity
GAD-7: Generalized Anxiety Disorder-7
GOLD: Global Initiative for Chronic Obstructive Lung Disease
mMRC: modified Medical Research Council
PHQ-9: Patient Health Questionnaire-9
pMDI: pressurized metered-dose inhaler
SGRQ-C: St George's Respiratory Questionnaire for Chronic Obstructive Pulmonary Disease
TAI: Test of Adherence to Inhalers

Edited by A Stone; submitted 20.Apr.2026; peer-reviewed by C Kapatais, J Edwards; revised version received 04.Oct.2026; accepted 05.Oct.2026; published 06.Nov.2026.

Please cite as:
Wu H, Dong F, Cui Y, Hu D, Zhang Y, Peng T, Peng Y, Li X, Li Y, Li X, Chen Y, Yu Q, Niu H, Yang T
Effectiveness of a Multicomponent Digital Therapeutic Program on Inhaled Medication Adherence in Chronic Obstructive Pulmonary Disease: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e98870
URL: https://mhealth.jmir.org/2026/1/e98870
doi:10.2196/98870

© Hanna Wu, Fen Dong, Yanan Cui, Danjing Hu, Yushi Zhang, Tianqi Peng, Yaodie Peng, Xiang Li, Yong Li, Xiaopan Li, Yongxia Chen, Qingjie Yu, Hongtao Niu, Ting Yang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 6.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Home-Based Digital Exercise Program for Patients After Open Repair of Acute Achilles Tendon Rupture: Noninferiority Randomized Controlled Trial

Hui Wang[1], PhD; Tianyi Wu[2], PhD; Jiye He[1], PhD; Lihua Huang[3], MSc; Yanhong Ma[3], PhD; Manrong Xu[4], MD; Shengdi Lu[2*], MD; Jian Zou[2*], PhD

[1]Department of Orthopedics, Shanghai Xinhua Hospital, Shanghai, China

[2]Department of Orthopedics, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, 600 Yishan Rd, Xuhui District, Shanghai, China

[3]Department of Rehabilitation, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China

[4]Department of Endocrinology, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Jian Zou, PhD
Department of Orthopedics, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, 600 Yishan Rd, Xuhui District, Shanghai, China

Abstract

Background: Achilles tendon rupture significantly affects patient mobility and quality of life. Postoperative rehabilitation is critical to regain ankle function, strength, and return to activity. Home-based digital rehabilitation has emerged as an accessible alternative to traditional clinic-based physiotherapy; however, high-quality evidence comparing these approaches after Achilles tendon repair remains scarce.

Objective: This randomized controlled trial aimed to evaluate whether a home-based digital exercise program (DEP) is noninferior to clinic-based face-to-face physiotherapy for restoring ankle range of motion (ROM), strength, patient-reported outcomes, and cost-effectiveness following open Achilles tendon repair.

Methods: Between August 2020 and June 2023, 200 adult patients (mean age 36.5, SD 5.0 years; n=135, 67.5% male) undergoing open Achilles tendon repair at 2 trauma centers were randomly assigned (1:1) to either a 12-week home-based DEP group (n=100) or traditional 12-week clinic-based physiotherapy (CP group, n=100) in Shanghai. The digital program was delivered through an online app, providing instructional videos, real-time telecommunication, daily exercise tasks, and monitoring of adherence. The clinic-based group received weekly physiotherapist-supervised sessions. The primary outcome was ankle ROM at 12 weeks postoperatively. Secondary outcomes included plantarflexion strength at 0° and 12°, heel-rise index and height, Achilles Tendon Total Rupture Score (ATRS), Victorian Institute of Sport Assessment–Achilles scores, adherence, patient satisfaction, cost-effectiveness, and adverse events assessed at 6, 12, and 24 weeks. Mixed-effects models were applied for intention-to-treat analyses, and sensitivity analyses were performed in the per-protocol population.

Results: At 12 weeks postoperatively, the DEP group was noninferior to the CP group for ankle ROM: mean difference (DEP–CP=−0.08°) with a 1-sided 95% CI lower bound of −2.16°, exceeding the noninferiority margin of −5° (P<.001). Secondary clinical outcomes at 12 weeks showed no material between-group differences, and after Holm-Bonferroni adjustment for multiplicity across secondary endpoints, no comparison reached statistical significance. Standardized effects at 12 weeks were trivial to small (|Cohen d|≤0.16). Patients rated DEP higher for convenience and ease of access in unadjusted analyses, but these did not remain significant after multiplicity correction (adjusted P=.08). DEP significantly reduced total rehabilitation costs vs CP (CNY 59,260.77 vs 68,432.80), with economic analyses favoring DEP. Adverse events were comparable between groups (DEP: 18, 18% patients and CP: 22, 22% patients; P>.05). The exchange rate used in this study was 1 US $=6.90 Chinese Yuan.

Conclusions: The home-based DEP following Achilles tendon repair provided comparable clinical outcomes to traditional physiotherapy while significantly improving cost-effectiveness. These findings support the broader adoption of tele-supervised digital rehabilitation as a safe and effective alternative.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2400093415; https://www.chictr.org.cn/showprojEN.html?proj=240693

(JMIR Mhealth Uhealth 2026;14:e78100) doi:10.2196/78100

XSL•FO
RenderX

KEYWORDS

achilles tendon rupture; digital exercise program; home-based rehabilitation; physiotherapy; cost-effectiveness

Introduction

Achilles tendon rupture is one of the most common tendon injuries of the lower extremity, typically affecting physically active adults in midlife during sports activities [1]. This injury carries a substantial burden, as it often causes sudden pain, loss of function, and prolonged disability, with a profound impact on patients' mobility and quality of life [2]. Management of acute ruptures can be surgical (open or minimally invasive repair) or nonsurgical; however, in either case, a structured rehabilitation program is critical to optimize outcomes. Early functional rehabilitation has been shown to reduce complications and improve recovery speed compared with prolonged immobilization [1,3]. Evidence from meta-analyses indicates that rehabilitative strategies emphasizing early weight-bearing and ankle mobilization after Achilles repair lead to higher patient satisfaction and faster return to sport without increasing the risk of adverse events [3]. As a result, current best practice encourages supervised, progressive rehabilitation to restore ankle range of motion (ROM), muscle strength, and functional performance after tendon healing.

Despite advances in surgical techniques, patients often experience significant deficits for months or even years after Achilles tendon injury. Studies have found that even at 1 - 2 years post surgery, many individuals exhibit persistent deficits in calf muscle strength and endurance and incomplete recovery of ankle biomechanics [2,4]. These long-term impairments underscore the importance of effective rehabilitation programs to maximize functional outcomes. Standard postoperative rehabilitation for Achilles repair usually involves in-person physiotherapy sessions focusing on incremental tendon loading, gait retraining, and exercises to improve plantarflexion strength and ankle ROM. Outcomes are typically tracked using objective measures, such as ankle ROM goniometry, isometric and isokinetic plantarflexion strength testing, and single-leg heel-rise tests, as well as patient-reported outcome scores. Common validated instruments include the Achilles Tendon Total Rupture Score (ATRS) and the Victorian Institute of Sport Assessment–Achilles (VISA-A) questionnaire, which assess symptoms and functional abilities in daily activities and sports. These metrics frequently reveal that without comprehensive rehabilitation, patients may experience persistent functional deficits and lower satisfaction [2,4]. Thus, there is a strong rationale for interventions that enhance rehabilitation adherence and effectiveness after Achilles tendon repair.

In recent years, digital health and telemedicine have opened new avenues for delivering rehabilitation outside of the traditional clinic setting. Mobile health technologies, such as smartphone apps, online platforms, and videoconferencing tools, allow patients to perform prescribed exercises at home under remote guidance. Telerehabilitation programs have grown particularly prominent, accelerated by the COVID-19 pandemic, which necessitated alternatives to face-to-face therapy. Such programs typically provide instructional exercise videos, progress tracking, and sometimes real-time feedback or virtual coaching. These home-based digital exercise solutions aim to increase accessibility and patient engagement in rehabilitation, especially for those facing barriers to frequent clinic visits. Early studies in other orthopedic populations have demonstrated the potential of telerehabilitation: after knee and hip replacements, home-based telerehabilitation has yielded recovery outcomes comparable to conventional in-person physiotherapy [5]. In a 2017 systematic review of postorthopedic surgery rehabilitation, strong evidence supported telerehabilitation as an effective alternative in total joint arthroplasty, with similar improvements in function and pain as standard care [5]. Telehealth approaches can also confer practical advantages, including convenience, reduced travel time, and greater cost-effectiveness. A recent meta-analysis reported that telerehabilitation can provide equivalent or improved health outcomes while lowering overall costs and enhancing patient satisfaction and adherence to exercise programs [6]. These findings suggest that well-designed digital rehab interventions can meet the therapeutic needs of patients without the logistical constraints of clinic-based therapy.

However, despite the growing adoption of digital rehabilitation, there remains a lack of high-quality evidence specifically for Achilles tendon rupture rehabilitation. Rehabilitation after Achilles repair poses unique challenges, balancing tendon protection with early mobilization, and it is unclear whether a purely home-based app-guided regimen can match the effectiveness of hands-on physiotherapist supervision. No large randomized controlled trials have definitively compared remote digital rehabilitation to traditional in-person rehabilitation after Achilles tendon repair. Some related research in chronic Achilles conditions is promising [7], but the postoperative scenario has distinct demands and outcome measures. Given the importance of restoring ankle ROM, plantarflexion strength, and functional performance after Achilles tendon surgery, it is crucial to determine whether at-home digital programs can deliver these results on par with standard care. This study was designed to address this gap by evaluating a 12-week home-based digital exercise program (DEP) in comparison with a conventional 12-week physiotherapist-supervised rehabilitation program. By using a noninferiority randomized controlled trial, the aim is to rigorously test whether the digital home program is not worse than clinic-based rehabilitation in improving ankle mobility and functional outcomes, while also examining differences in cost-effectiveness.

Methods

Study Design and Setting

We conducted a 2-center, parallel-group randomized controlled trial with a noninferiority design to compare a 12-week home-based DEP with a clinic-based physiotherapy program for patients after open repair of Achilles tendon rupture. The trial took place at Shanghai Sixth People's Hospital and Shanghai Xinhua Hospital, 2 tertiary academic centers, from August 2020 to June 2023. Eligible patients were randomized in a 1:1 ratio to the digital or conventional rehabilitation arm.

XSL•FO
RenderX

We adhered to CONSORT (Consolidated Standards of Reporting Trials) [8] guidelines for trial reporting.

Participants

Participants were recruited on the day following their emergency admission to the 2 participating hospitals. One of the coordinators screened the eligibility of participants and explained all potential harms and benefits to eligible participants before written informed consent was obtained. The inclusion and exclusion criteria are provided in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria:

- Adults (18-65 years) who were scheduled to undergo open surgical repair for an acute, complete Achilles tendon rupture.
- Surgery was planned within 2 weeks after injury.
- Participants had to commit to the rehabilitation program, follow-up visits, and be willing to provide written informed consent.
- Access to a smartphone and basic proficiency with the app.

Exclusion criteria:

- Participants with a history of prior Achilles tendon rupture, chronic corticosteroid use, or uncontrolled diabetes (due to potential impaired healing), or any significant comorbidity affecting the lower limb (eg, neuropathy or severe arthritis).
- Coexisting inflammatory arthritis, neurological conditions, or other lower-limb musculoskeletal disorders.
- Uncontrolled comorbidities (eg, diabetes and heart disease) that could interfere with recovery.
- Participants unable to participate in exercise (due to other injuries or contraindications) or unable to use the smartphone app or attend follow-ups.

Intervention

All participants received standard postoperative care in the initial phase: a posterior splint or boot with the ankle in plantar flexion for 2 weeks, followed by protected weight-bearing as tolerated in a walking boot up to 6 weeks. At approximately 2 - 3 weeks postsurgery (after suture removal and when acute pain subsided), participants began the assigned intervention program.

Home-Based DEP Group

Participants randomized to the experimental group received a 12-week exercise program delivered via the “Joymotion” (Shanghai Medmotion Medical Management Co, Ltd) smartphone app. This app provides daily interactive exercise modules with video demonstrations and audio instructions for each exercise. Participants were assigned step-by-step daily rehabilitation tasks, which they performed at home while following a Physiotherapeutic Achilles-Specific Exercise program (details in Multimedia Appendix 1). The app included features for logging completed exercises and tracking progress over time. Participants could receive feedback and motivational prompts through the app, and the system flagged any missed sessions. A licensed physiotherapist remotely monitored each participant’s adherence and performance through the app’s dashboard and supervised the progression of exercises weekly, adjusting the difficulty or adding new exercises as appropriate. The therapist and participant communicated via in-app messaging or phone or video calls if issues arose, ensuring expert oversight despite the home setting. Physiotherapists maintained a high level of supervision intensity for the DEP group. They responded to participant queries within approximately 2 hours and provided feedback after each exercise session, with additional weekly summary feedback to guide the participant’s progression. We monitored intervention fidelity using the app’s dashboard metrics, including the average response time to participant messages and the frequency of feedback provided, to ensure the digital program was delivered consistently as designed. This interactive telerehabilitation approach was designed to mirror the content of standard physiotherapy in a convenient home format.

Clinic-Based Physiotherapist Rehabilitation (CP Group)

Patients in the control group underwent a conventional rehabilitation regimen of supervised physical therapy at participating hospitals and 2 other assigned physical therapy clinics in Shanghai. This consisted of weekly in-clinic physiotherapy sessions (Physiotherapeutic Achilles-Specific Exercise program) for 12 weeks postoperative. In each session, a physiotherapist guided the patient through a standardized Achilles tendon rehab protocol, including joint mobilization, stretching, strengthening of the gastrocsoleus, balance and proprioception exercises, and functional training. The program was progressive: early sessions emphasized a gentle ROM and isometric strength exercises, advancing to resistance exercises, tendon loading drills, and functional activities in later weeks, in line with typical Achilles rehab guidelines. Between these weekly appointments, control patients were instructed in a home exercise program to perform on their own at least 5 days per week. However, unlike the digital group, they did not receive daily interactive guidance or monitoring between visits.

We recorded any deviations from the protocol or cointerventions. Compliance was encouraged in both groups, and adherence data were collected via app logs for the digital group and attendance records for the clinic group. Both groups were managed by experienced physiotherapists. We monitored adverse events throughout the program in both groups.

XSL•FO
RenderX

Outcomes

The primary outcome was the ankle ROM of the affected side at 12 weeks postoperatively. We defined this as the active ROM in dorsiflexion (movement from the neutral ankle position to maximum dorsiflexion) measured in degrees with a goniometer. For consistency, measurements were taken with the patient supine and knee extended (to standardize gastrocnemius tension). We also recorded plantar flexion ROM, but dorsiflexion limitation was of primary interest given its functional importance after Achilles repair. At the 12-week endpoint, a blinded assessor measured the ankle ROM of both the injured and contralateral limbs; the primary analysis focused on the between-group difference in injured-side ROM.

Secondary outcomes included functional, strength, patient-reported, and economic measures, assessed at baseline (at the beginning of the intervention) and at 6-, 12-, and 24-week post surgery:

1. Plantarflexion muscle strength: isometric plantarflexion strength of the affected calf was measured at 2 ankle positions, 0° (neutral ankle) and 12° of dorsiflexion, using a dynamometer. Strength was recorded in Newton-meters (Nm) as the peak torque the patient could generate in a 3-second maximal voluntary contraction [9]. These 2 positions were chosen to assess strength across the ankle's range (neutral vs a dorsiflexed, more elongated tendon position). Any strength deficit compared to the contralateral side was calculated as a percentage.
2. Heel-rise index and height: we evaluated calf muscle endurance and power using the single-leg heel-rise test. Patients performed single-leg heel raises to exhaustion on both the injured and uninjured sides. The heel-rise index was calculated as the ratio of the affected limb's performance relative to the healthy side [10]. Performance was quantified as the total work or height × repetition product achieved; specifically, we counted the number of consecutive heel raises and measured the maximum heel-rise height (in cm) achieved by the affected limb and then compared it to the contralateral limb. A heel-rise index of 100% indicates full recovery (equal performance to the uninjured side), whereas lower percentages indicate residual deficits [10]. We also reported heel-rise height (the maximum height of a single heel rise on the injured side) as a separate outcome, since it reflects calf muscle function and tendon elasticity.
3. The ATRS: a validated patient-reported outcome for Achilles tendon rupture recovery. The ATRS consists of 10 items assessing symptoms and activity limitations [11]. Each item is scored 0 - 10, and total scores range from 0 to 100, with 100 indicating no symptoms or functional limitations (best outcome) and 0 indicating maximal impairment. ATRS is reliable, valid, and responsive for patients with Achilles rupture and is a common primary outcome in Achilles rehabilitation studies [11].
4. The VISA-A questionnaire: originally developed for Achilles tendinopathy, we included this 8-item instrument to gauge Achilles tendon pain and function as another patient-reported metric. VISA-A scores also range from 0 to 100 (100=asymptomatic, full function). It covers pain during activity, function in daily living, and the ability to participate in sports [12]. While VISA-A is not specific to acute rupture, we hypothesized that it would capture complementary information on tendon function and any tendinopathy-like symptoms during recovery.
5. Cost-effectiveness: it was assessed by calculating the incremental cost-effectiveness ratio (ICER) between the DEP and clinic-based physiotherapy (CP) groups. Costs included intervention-related costs, health care resource usage, and productivity losses due to absenteeism and presenteeism. Data were collected from hospital records and patient-reported logs. Cost-effectiveness analysis was conducted based on the outcomes at 12 weeks of follow-up (at the end of the intervention).

All outcomes were measured by assessors blinded to group allocation (described below in the "Randomization, Allocation, and Blinding" subheading). The primary endpoint for efficacy was at 12 weeks postoperative (end of the program), with longer-term outcomes at 24 weeks used to observe maintained effects or any late differences.

Sample Size Calculation

This trial was powered to evaluate the noninferiority of the digital rehabilitation program compared to standard physiotherapy in terms of 12-week ankle ROM. No definitive minimal clinically important difference for ankle dorsiflexion ROM after Achilles repair has been established in the literature [13]. Based on clinical judgment and prior data on variability, we set a noninferiority margin (Δ) of 5 degrees of ankle ROM. A difference of 5° was deemed the largest loss of ROM that would be considered clinically unimportant to patients. Notably, postoperative studies have found that by 6 months after Achilles repair, side-to-side dorsiflexion differences are on average near 0° (with a SD of 3°). Thus, a 5° deficit is larger than normal variability yet small enough to be functionally negligible. This choice is also consistent with the principle that the noninferiority margin should be smaller than the expected minimal clinically important difference for the outcome. Using a 1-sided type I error of .05 (α) and power (1–β) of .80, we calculated the required sample size per group for 2 independent means in a noninferiority design and estimated approximately 50 patients per group to be needed to detect noninferiority [8,14,15]. To account for potential dropouts or loss to follow-up (we anticipated 40% attrition by 24 weeks), we aimed to recruit 85 patients in each group.

Randomization, Allocation, and Blinding

Participants were randomized in a 1:1 ratio to either the digital rehab group or the clinic-based rehab group. An independent researcher (not involved in recruitment or assessment) generated the random allocation sequence using a computer program. We used variable block sizes of 4 and 6 to ensure concealment and balanced group sizes. The allocations were sealed in opaque, sequentially numbered envelopes. After a participant was enrolled and completed all baseline assessments, the next envelope in sequence was opened to reveal the assignment. This ensured allocation concealment until the point of assignment. Randomization was not stratified (since all participants had

XSL•FO
RenderX

similar baseline characteristics) and thus was purely random within blocks.

Due to the nature of the interventions, it was not possible to blind the participants or treating physiotherapists to group assignments. Participants obviously knew whether they were using an app at home or attending in-person therapy sessions. However, we implemented measures to blind outcome assessments and analyses to reduce bias. All outcome evaluations at 6, 12, and 24 weeks were performed by a research assistant blinded to the patient's group at the outpatient department of 2 participating hospitals. This assessor was not involved in delivering the rehabilitation to either group. Participants were instructed not to reveal details of their rehabilitation program or discuss their therapy during assessment visits. The assessor recorded the outcomes without knowledge of which intervention was received. Furthermore, the trial statistician was kept blinded to group labels when analyzing primary outcomes; groups were coded anonymously until after the primary analysis was complete. By maintaining assessor and analyst blinding, we aimed to minimize measurement and analytical bias. Participants themselves were aware of their treatment, so participant blinding was not possible, but they were informed of the importance of unbiased assessment.

Statistical Analysis

The primary outcome was the between-group difference in ankle ROM at 12 weeks. We prespecified a noninferiority margin $\Delta=-5°$ and tested $H: \mu^{DEP-CP} \leq -5°$ vs $H_1: \mu^{DEP-CP} > -5°$ using a 1-sided test at $\alpha=.05$. Mixed-effects models with fixed effects for group, time point at 6, 12, and 24 weeks, and group × time, and a random intercept for participants were used. Analyses followed a prespecified intention-to-treat (ITT) framework as primary and a per-protocol (PP) framework as sensitivity. ITT used linear mixed-effects models that incorporated all available observations under a missing-at-random assumption and retained participants in their randomized groups regardless of adherence. PP used a complete-case approach that included only participants who completed all scheduled follow-ups. Missingness occurred only at the visit level because a missed follow-up meant all variables for that visit were absent. There was no item-level missingness within attended visits.

For secondary outcomes at Week 12, we present 2-sided 95 % CIs and *P* values from the same mixed-effects model, and Holm-Bonferroni control for multiplicity across the Week 12 clinical outcomes (familywise $\alpha=.05$); 6 and 24 weeks are presented descriptively. For the effect size, Cohen *d* at Week 12 was used as the standardized mean difference using the pooled SD from the ITT group. Patient satisfaction and acceptability items were treated as a separate hypothesis family; their *P* values were adjusted for multiplicity using the Holm-Bonferroni procedure (familywise $\alpha=.05$). PP analyses served as sensitivity checks.

For cost-effectiveness analysis, total costs for each participant were calculated, and the difference in mean cost between groups was evaluated. In a cost-effectiveness plane, we will compare cost vs effectiveness for the 2 strategies. If the digital program were cheaper and with no worse outcome, it would be a dominant strategy. Otherwise, an ICER (cost per unit of benefit) will be computed. Nonparametric bootstrap methods (10,000 resamples) will be used to estimate uncertainty in the ICER and to construct cost-effectiveness acceptability curves.

We aimed to minimize missing data by thorough follow-up, but any remaining missing outcome data were handled under the ITT framework. The mixed-effects models can accommodate missing values under a missing-at-random assumption by using all available data for each participant. Baseline variables and other outcomes would be used in a chained equations imputation model to predict missing values, and analyses would be repeated on imputed datasets to ensure results are robust.

Data analysis was conducted using SPSS (version 28.0; IBM Corp) and R (version 4.2.2; the R Core Team). The trial results will be reported according to the CONSORT guidelines [8], including the noninferiority margin and CIs for primary outcomes.

Ethical Considerations

Ethical approval was obtained from the Institutional Review Board of Shanghai Sixth People's Hospital (IRB no 2020-KY-089[K]), and this approval was also acknowledged by Shanghai Xinhua Hospital. All procedures were conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants before enrollment, and participants were informed of their right to withdraw from the study at any time. All personal data were anonymized (deidentified) to ensure participant privacy and confidentiality. No financial or other compensation was provided to participants for their participation in this study.

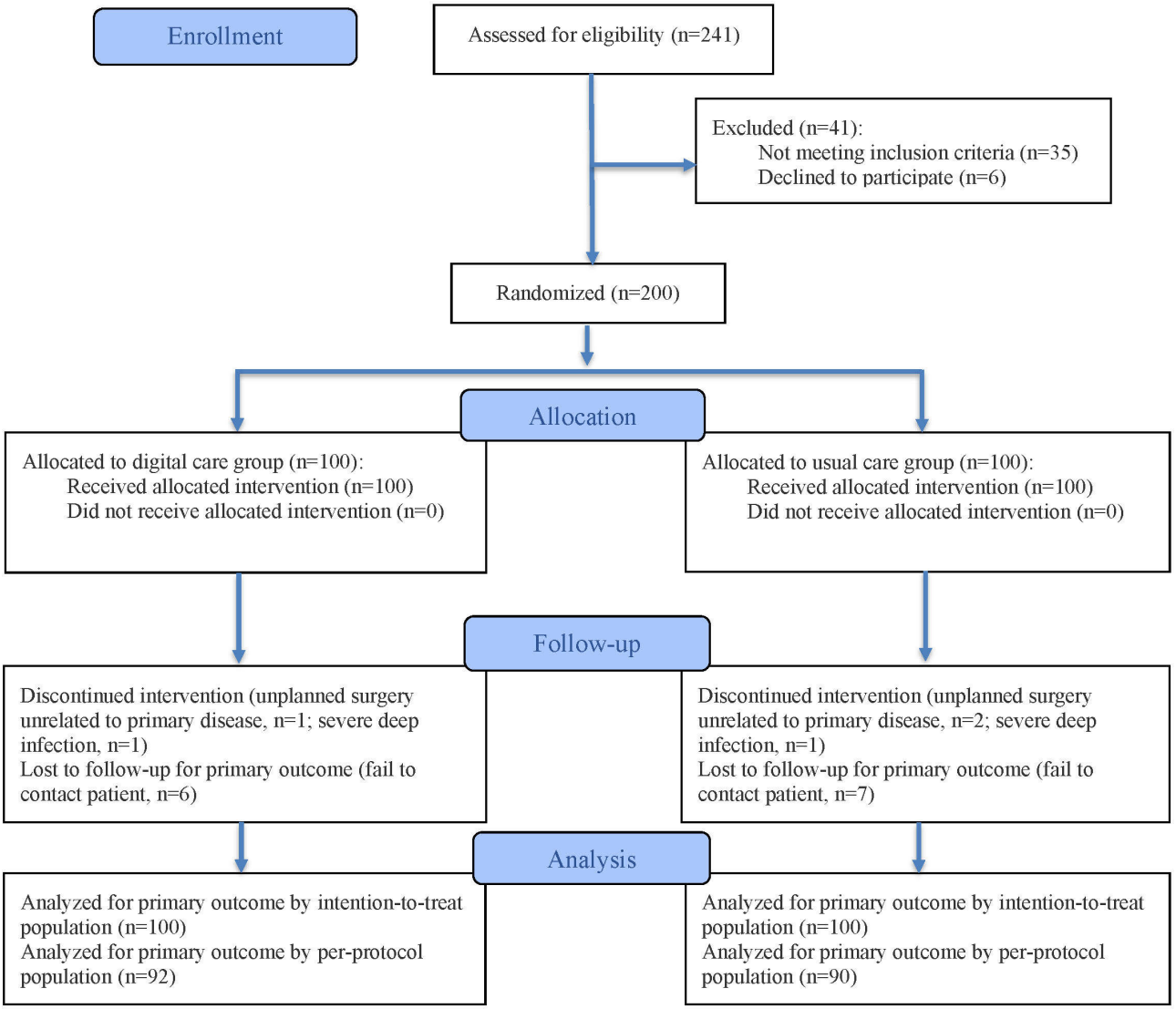
Results

Participants

We evaluated 241 patients with acute Achilles tendon injury, and 200 were randomized. The flowchart of our study is specified in Figure 1 (92 participants in the DEP group and 90 participants in the CP group completed the entire follow-up investigation). The baseline characteristics were comparable between the DEP and CP groups. Most participants were male, with 71 out of 100 (71%) in the DEP group and 64 out of 100 (64%) in the CP group (*P*=.21), with a mean age of approximately 37 and 36 years, respectively (*P*=.08). Average BMI was similar between groups (DEP: 25.0 kg/m^2; CP: 25.0 kg/m^2; *P*=.25). Education levels and insurance types were evenly distributed between the groups (*P*=.55 and *P*=.66, respectively). Smoking status did not significantly differ between groups; 17 out of 100 (17%) in the DEP group and 21 out of 100 (21%) in the CP group were smokers (*P*=.47). Injury distribution (right vs left side) was also balanced (*P*=.57). Time from rupture to surgery averaged around 3.5 days in both groups (DEP: mean 3.55, SD 0.63; CP: mean 3.59, SD 0.58; *P*=.60; Table 1).

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of the study.



XSL•FO
RenderX

Table . Baseline characteristics of the digital exercise program (DEP) and clinic-based physiotherapy (CP) groups.

Patient characteristic	DEP[a] group (n=100)	CP[b] group (n=100)	*P* value
Male patients, n (%)	71 (71.00)	64 (64.00)	.21
Age (years), mean (SD)	37 (4.93)	36 (5.14)	.08
BMI (kg/m^2), mean (SD)	25 (0.60)	25 (0.52)	.25
Education level, n (%)			.55
Lower than high school	16 (16.00)	13 (13.00)	
Equal to or higher than high school	84 (84.00)	87 (87.00)	
Insurance type, n (%)			.66
Government	77 (77.00)	74 (74.00)	
Commercial	14 (14.00)	13 (13.00)	
Self-financed	9 (9.00)	13 (13.00)	
Current smoker, n (%)	17 (17.00)	21 (21.00)	.47
Ruptured side, n (%)			.57
Right side	55 (55.00)	52 (52.00)	
Left side	45 (45.00)	50 (50.00)	
Days from rupture to surgery, mean (SD)	3.55 (0.63)	3.59 (0.58)	.60

[a]DEP: digital exercise program.

[b]CP: clinic-based physiotherapy.

Primary and Secondary Outcomes

The primary and secondary outcomes changes (Table 2) and mixed-effects model estimates across time (Table 3) are presented. The primary analysis supported noninferiority of DEP for ankle ROM at 12 weeks. The estimated mean difference was −0.081°(DEP – CP); the 1-sided 95% CI lower bound was −2.16°, which is above the noninferiority margin of −5° (*P*<.001). In the PP population, results were consistent (difference −0.218°; 1-sided 95% CI lower bound −2.38°; *P*<.001; Table 4).

XSL•FO
RenderX

Table . Changes in outcomes at weeks 6, 12, and 24 after surgery.

Outcome	6 weeks post surgery		12 weeks post surgery				24 weeks post surgery	
	DEP[a] group (n=100)	CP[b] group (n=100)	DEP group (n=100)	CP group (n=100)	Cohen *d*	Adjusted *P* value[c]	DEP[a] group (n=100)	CP[b] group (n=100)
Intention-to-treat population, mean (SD)								
Range of motion	16.76 (5.30)	16.40 (6.37)	31.73 (6.47)	32.73 (5.90)	__[d]	—	65.74 (6.40)	65.04 (6.03)
Plantarflexion strength 0	20.73 (3.25)	21.34 (2.90)	41.80 (3.03)	41.87 (3.31)	−0.02	1	85.02 (2.87)	85.13 (3.18)
Plantar flexion strength 12	18.47 (4.14)	18.10 (4.07)	35.95 (3.93)	36.07 (3.95)	−0.03	1	71.71 (3.83)	71.33 (4.05)
Heel-rise index	10.92 (4.53)	11.06 (3.72)	22.23 (3.69)	22.10 (3.52)	0.04	1	44.60 (4.32)	44.48 (3.81)
Heel-rise height	15.05 (4.55)	15.85 (5.47)	29.72 (5.35)	29.76 (4.74)	−0.01	1	60.51 (4.97)	61.18 (4.45)
ATRS[e]	12.06 (2.92)	11.86 (2.99)	33.33 (3.27)	32.93 (3.12)	0.13	1	53.59 (4.03)	53.75 (3.95)
VISA-A[f]	10.20 (2.85)	9.98 (3.32)	43.82 (3.78)	44.07 (4.23)	−0.06	1	65.09 (3.30)	65.63 (3.84)
Per-protocol population, mean (SD)								
Range of motion	16.71 (5.22)	16.38 (6.25)	31.54 (6.42)	32.53 (5.80)	—	—	65.93 (6.44)	64.98 (6.07)
Plantarflexion strength 0	20.63 (3.19)	21.44 (2.96)	41.75 (3.08)	41.99 (3.32)	—	—	85.04 (2.88)	85.23 (3.20)
Plantar flexion strength 12	18.35 (4.21)	18.14 (4.06)	35.92 (3.97)	35.92 (3.87)	—	—	71.68 (3.87)	71.33 (4.04)
Heel-rise index	10.92 (4.51)	10.95 (3.80)	22.34 (3.68)	22.01 (3.53)	—	—	44.52 (4.37)	44.68 (3.82)
Heel-rise height	14.96 (4.24)	15.80 (5.40)	29.48 (5.21)	29.78 (4.80)	—	—	60.62 (4.99)	61.21 (4.52)
ATRS	12.10 (2.91)	11.75 (2.86)	33.28 (3.32)	32.94 (3.11)	—	—	53.61 (3.96)	53.80 (3.98)
VISA-A	10.22 (2.91)	10.04 (3.29)	43.87 (3.81)	44.29 (4.20)	—	—	65.12 (3.20)	65.76 (3.84)

[a]DEP: digital exercise program.

[b]CP: clinic-based physiotherapy.

[c]Adjusted *P* value: Holm-Bonferroni adjustment.

[d]Not applicable.

[e]ATRS: Achilles tendon Total Rupture Score.

[f]VISA-A: Victorian Institute of Sport Assessment–Achilles.

XSL•FO
RenderX

Table . Effectiveness estimates from linear mixed-effects models.

Outcome	6 weeks post surgery		12 weeks post surgery		24 weeks post surgery	
	Coefficient (95% CI)	*P* value	Coefficient (95% CI)	*P* value	Coefficient (95% CI)	*P* value
Intention-to-treat population						
Range of motion	0.204 (−0.614 to 1.023)	.62	−0.081 (−2.563 to 2.400)	.95	0.163 (−5.254 to 5.580)	.95
Plantarflexion strength 0	−0.301 (−0.734 to 0.132)	.17	−0.039 (−3.135 to 3.058)	.98	0.03 (−5.718 to 5.777)	.99
Plantar flexion strength 12	0.214 (−0.518 to 0.947)	.56	0.205 (−2.490 to 2.899)	.88	0.325 (−4.525 to 5.174)	.90
Heel-rise index	−0.051 (−0.635 to 0.532)	.86	0.069 (−1.634 to 1.773)	.94	0.142 (−3.557 to 3.841)	.94
Heel-rise height	−0.38 (−1.083 to 0.323)	.29	−0.145 (−2.418 to 2.128)	.90	−0.209 (−5.224 to 4.806)	.94
ATRS[a]	0.117 (−0.356 to 0.59)	.62	0.289 (−2.198 to 2.775)	.82	0.231 (−3.543 to 4.006)	.90
VISA-A[b]	0.12 (−0.303 to 0.544)	.58	0.093 (−3.300 to 3.486)	.96	−0.024 (−4.828 to 4.779)	.99
Per-protocol population						
Range of motion	0.162 (−0.673 to 0.997)	.70	−0.218 (−2.789 to 2.353)	.87	0.059 (−5.639 to 5.757)	.98
Plantarflexion strength 0	−0.407 (−0.855 to 0.041)	.08	−0.329 (−3.559 to 2.901)	.84	−0.303 (−6.324 to 5.719)	.92
Plantarflexion strength 12	0.106 (−0.658 to 0.870)	.78	0.053 (−2.748 to 2.853)	.97	0.136 (−4.934 to 5.205)	.96
Heel-rise index	−0.018 (−0.626 to 0.590)	.95	0.089 (−1.690 to 1.867)	.92	0.025 (−3.866 to 3.916)	≥.99
Heel-rise height	−0.421 (−1.112 to 0.27)	.23	−0.358 (−2.716 to 1.999)	.77	−0.43 (−5.702 to 4.841)	.87
ATRS[a]	0.179 (−0.316 to 0.674)	.48	0.217 (−2.379 to 2.814)	.87	0.129 (−3.818 to 4.076)	.95
VISA-A[b]	0.091 (−0.355 to 0.537)	.70	−0.083 (−3.646 to 3.480)	.96	−0.241 (−5.267 to 4.786)	.93

[a]ATRS: Achilles tendon Total Rupture Score.

[b]VISA-A: Victorian Institute of Sport Assessment–Achilles.

Table . Primary outcome (ankle range of motion [ROM]) noninferiority analysis at 12 weeks. Difference is digital exercise program (DEP) – clinic-based physiotherapy (CP); noninferiority margin (Δ)=−5°.

Population	Mean difference (°)	One-sided 95% CI lower bound (°)	Noninferiority *P* value
Intention-to-treat population	−0.081	−2.16	<.001
Per-protocol population	−0.218	−2.38	<.001

At Week 12, unadjusted model estimates showed no significant between-group differences for plantarflexion strength (0° and 12°), heel-rise index, heel-rise height, ATRS, or VISA-A. After Holm-Bonferroni correction across these 6 secondary clinical endpoints, none was significant (all adjusted *P*≥.99).

To contextualize magnitudes, Week 12 Cohen *d* values indicated trivial to small effects: ankle ROM d=−0.16; plantarflexion strength 0° d=−0.02; plantarflexion strength 12° d=−0.03; heel-rise index d=+0.04 ; heel-rise height d=−0.01; ATRS d=+0.13; VISA-A d=−0.06 (positive d favors DEP; Table 2). We summarized participant-level and visit-level missingness by group and time point in Table S1 in Multimedia Appendices 1 and 2.

Cost-Effectiveness Analysis

The average total cost per participant was significantly lower in the DEP group compared to the CP group during the 24-week postoperative period (CNY 59,260.77 vs CNY 68,432.80;

XSL•FO
RenderX

difference=−9172.03; *P*<.001). The exchange rate used in this study was 1 US $=6.90 Chinese Yuan. Major cost savings for the DEP group arose from rehabilitation costs (difference=−3932.65; *P*<.001), primary care expenses (difference=−1820.62; *P*<.001), secondary care costs (difference=−2245.19; *P*<.001), medication (difference=−1073.66; *P*<.001), and transportation (difference=−960.10; *P*<.001). Incremental cost-effectiveness analysis showed negative ICER values for primary and secondary outcomes, indicating cost savings without clinically relevant differences between groups. Specifically, the ICER for ankle ROM was CNY 113,234.94 per additional unit improvement, highlighting significant cost savings without sacrificing clinical effectiveness. Similar patterns were observed for secondary outcomes, including ATRS (ICER=−31,737.13) and VISA-A (ICER=−98,623.98), reinforcing the economic advantage of the digital rehabilitation program (Tables 5 and 6).

Table . Average total cost per patient in the digital exercise program (DEP) group and clinic-based physiotherapy (CP) group during the 24 weeks after the postoperative period.

Cost category (CNY[a])	DEP[b] group (n=100), mean (SD)	CP[c] group (n=100), mean (SD)	*P* value
Intention-to-treat population			
Rehabilitation cost	6980.00 (0)	10912.65 (1509.25)	<.001
Hospital stay cost	6903.08 (962.78)	6809.41 (998.34)	.51
Primary care cost	121.13 (48.32)	1941.75 (268.55)	<.001
Secondary care cost	4886.38 (948.22)	7131.57 (2160.98)	<.001
Paid home cost	550.04 (203.51)	782.28 (282.86)	<.001
Medication	6941.91 (1416.06)	8015.57 (1588.23)	<.001
Transportation cost	518.80 (199.38)	1478.90 (268.81)	<.001
Nutrition cost	2945.16 (493.63)	3039.46 (456.52)	.17
Lost wages for patients	26383.63 (14552.00)	25727.87 (12775.52)	.74
Lost wages for families	3030.64 (3397.40)	2583.35 (3313.99)	.35
Total cost	59260.77 (15099.91)	68432.80 (13750.30)	<.001
Per-protocol population			
Rehabilitation cost	6980.00 (0)	10921.53 (1482.25)	<.001
Hospital stay cost	6943.15 (924.93)	6848.47 (990.64)	<.001
Primary care cost	120.11 (47.50)	1943.33 (263.75)	.51
Secondary care cost	4873.82 (953.44)	7009.79 (2143.54)	<.001
Paid home cost	555.10 (205.82)	783.13 (274.64)	<.001
Medication	6889.73 (1412.89)	7986.04 (1574.05)	<.001
Transportation cost	516.24 (200.37)	1476.30 (274.05)	<.001
Nutrition cost	2908.00 (474.83)	3037.04 (463.03)	<.001
Lost wages for patients	27061.77 (14540.13)	25610.86 (12444.39)	.07
Lost wages for families	3073.00 (3423.92)	2434.28 (3179.87)	.47
Total cost	59920.91 (15162.77)	68050.76 (13462.49)	.19

[a]CNY: Chinese YUAN.

[b]DEP: digital exercise program.

[c]CP: clinic-based physiotherapy.

XSL•FO
RenderX

Table . Incremental cost-effectiveness ratio (ICER).

Population and analysis	Incremental cost (CNY[a])	Range of motion, mean (95% CI)	Plantarflexion strength 0 , mean (95% CI)	Plantarflexion strength 12 , mean (95% CI)	Heel-rise index, mean (95% CI)	Heel-rise height, mean (95% CI)	ATRS[b], mean (95% CI)	VISA-A[c], mean (95% CI)
Intention-to-treat population								
Main analysis (mixed effects)	−9172.03	−0.081 (−2.563 to 2.400)	−0.039 (−3.135 to 3.058)	0.205 (−2.490 to 2.899)	0.069 (−1.634 to 1.773)	−0.145 (−2.418 to 2.128)	0.289 (−2.198 to 2.775)	0.093 (−3.300 to 3.486)
ICER[d]	—	113,234.94	235,180.26	−44,741.61	−132,927.97	63,255.38	−31,737.13	−98,623.98
Per-protocol population								
Main analysis (mixed effects)	−8129.85	−0.081 (−2.563 to 2.400)	−0.039 (−3.135 to 3.058)	0.205 (−2.490 to 2.899)	0.069 (−1.634 to 1.773)	−0.145 (−2.418 to 2.128)	0.289 (−2.198 to 2.775)	0.093 (−3.300 to 3.486)
ICER	—	100,368.52	208,457.69	−39,657.80	−117,823.91	56,067.93	−28,130.97	−87,417.74

[a]CNY: Chinese Yuan.

[b]ATRS: Achilles tendon Total Rupture Score.

[c]VISA-A: Victorian Institute of Sport Assessment–Achilles.

[d]ICER: incremental cost-effectiveness ratio.

Patients' Adherence

In the DEP group, participants completed a mean of 28.5 (SD 2.6) of 30 educational modules and attended 5.7 (SD 1.1) of 6 scheduled video consultations (metrics specific to the digital program; not applicable to the CP group). At 24 weeks, patient ratings were high in both groups. Although mean scores favored DEP for convenience (mean 4.6, SD 0.5 vs mean 4.4, SD 0.7) and ease of use and access (mean 4.4, SD 0.7 vs mean 4.2, SD 0.5), these differences were not significant after Holm-Bonferroni adjustment within the satisfaction or acceptability family (both adjusted P=.08). Helpfulness (mean 4.7, SD 0.4 vs mean 4.7, SD 0.5; adjusted P≥.99) and communication with the physiotherapist (mean 4.8, SD 0.4 vs mean 4.7, SD 0.4; adjusted P=.16) were similarly high and comparable between groups (Table 7).

Table . Patients' adherence to treatment (intention-to-treat population).

Outcome measure	DEP[a] group (n=100), mean (SD)	CP[b] group (n=100), mean (SD)	Adjusted P value
Educational modules completed (maximum of 30 modules)	28.5 (2.6)	N/A[c]	—
Number of video consultations attended (maximum of 6 times)	5.7 (1.1)	N/A	—
Patients' scoring for intervention at 24 weeks			
Convenience (1-5)	4.6 (0.5)	4.4 (0.7)	.08
Ease of use and access (1-5)	4.4 (0.7)	4.2 (0.5)	.08
Helpfulness (1-5)	4.7 (0.4)	4.7 (0.5)	≥.99
Communication with the physiotherapist (1-5)	4.8 (0.4)	4.7 (0.4)	.16

[a]DEP: digital exercise program.

[b]CP: clinic-based physiotherapy.

[c]N/A: not applicable.

Adverse Events

The incidence of adverse events was similar between the DEP and CP groups: 18 out of 100 (18%) patients in the DEP group vs 22 out of 100 (22%) patients in the CP group experienced an adverse event. Most adverse events were related to therapy, primarily ankle pain (DEP: 8 and CP: 10) and swelling (DEP: 6 and CP: 5). Muscle strain occurred equally (2 per group). Serious adverse events were rare, occurring in 3 out of 100 (3%) patients in each group, including deep infections (DEP: 1 and CP: 1) and unrelated surgeries (DEP: 1 and CP: 2). One patient in the DEP group withdrew because of relocation. No substantial differences in safety profiles emerged, suggesting comparable tolerability and safety of both rehabilitation programs (Table 8).

XSL•FO
RenderX

Table . Adverse events and serious adverse events.

Adverse events	DEP[a] group (n=100)	CP[b] group (n=100)
Patients with adverse events, n (%)	18 (18)	22 (22)
Events related to study therapy, n (%)	17 (17)	17 (17)
Events unrelated to study therapy, n (%)	4 (4)	8 (8)
Type of event, n (%)		
Involved ankle		
Pain	8 (8)	10 (10)
Swelling	6 (6)	5 (5)
Muscle strain[c]	2 (2)	2 (2)
Signs of infection[d] (swelling, redness, heat, or pus)	1 (1)	2 (2)
Mobilization under anesthesia	1 (1)	0 (0)
Other		
Nausea and dizziness	0 (0)	2 (2)
Insomnia	1 (1)	1 (1)
Anxiety about ankle recovery	2 (2)	3 (3)
Serious adverse events[e], n (%)		
Patients with serious adverse events	3 (3)	3 (3)
Events related to study therapy	1 (1)	1 (1)
Events unrelated to study therapy	2 (2)	2 (2)
Unplanned surgery, which is unrelated to primary injury	1 (1)	2 (2)
Moving to another city	1 (1)	0 (0)
Deep infection	1 (1)	1 (1)

[a]DEP: digital exercise program.

[b]CP: clinic-based physiotherapy.

[c]Four patients sustained gastrocnemius muscle strain during exercise with minor consequent symptoms.

[d]Three patients developed superficial infection and eventually healed with conservative treatment.

[e]Patients with serious adverse events were automatically withdrawn from the study.

Discussion

Principal Findings

This randomized controlled trial demonstrated that the home-based DEP was noninferior to traditional clinic-based physiotherapy for patients following open Achilles tendon repair. At the primary endpoint, ankle ROM showed no significant differences between groups, affirming the effectiveness of the digital rehabilitation approach. Similarly, no meaningful differences emerged across secondary outcomes, including plantarflexion strength, heel-rise tests, and patient-reported outcomes, at all time points assessed (6, 12, and 24 weeks). Importantly, the DEP strategy achieved these clinical results at significantly lower costs, with average savings of approximately CNY 9172 per patient compared to the conventional approach. Incremental cost-effectiveness analyses further indicated that DEP delivered equivalent or improved cost-effectiveness across all measured outcomes. Additionally, adherence to rehabilitation was notably high in the digital group, where participants reported greater convenience and ease of use compared to face-to-face rehabilitation. Adverse events and serious adverse events were infrequent and evenly distributed between the 2 groups, suggesting comparable safety profiles. Collectively, these findings suggest that a structured digital home rehabilitation program can effectively substitute clinic-based physiotherapy after Achilles tendon repair, offering a safe, cost-effective, and patient-friendly alternative.

Our noninferiority analysis using a 1-sided 95% CI lower bound confirmed that DEP is not inferior to CP for restoring ankle ROM at 12 weeks. Importantly, standardized effect sizes were uniformly small ($|d| \leq 0.16$), emphasizing that between-group differences were not only statistically nonsignificant after multiplicity control but also clinically negligible in magnitude. Furthermore, multiple-adjusted analyses of secondary endpoints showed no clinically relevant advantages for CP over DEP. While unadjusted patient-reported convenience and access ratings favored DEP, these did not meet family-wise significance after correction and should be interpreted as exploratory.

XSL•FO
RenderX

A high-quality randomized controlled trial on acute Achilles tendon rupture rehabilitation found that a home-based digital rehabilitation program produced clinical outcomes equivalent to conventional in-clinic physiotherapy. Patients in the digital rehab group achieved similar ankle ROM, calf muscle strength, and patient-reported outcome measures as those attending standard physiotherapy, indicating noninferiority of the telehealth approach. This finding is consistent with prior evidence in related conditions, for example, a noninferiority trial in chronic Achilles tendinopathy reported no worse pain or functional outcomes with telehealth-only or hybrid care compared to all in-person treatment [7]. Such results suggest that when rehab protocols are well-designed and supervised remotely, patients can attain comparable functional recovery to traditional face-to-face therapy.

Our trial's results align with broader Achilles rupture rehabilitation evidence, emphasizing that rehabilitation protocols, rather than the setting, drive outcomes. Past research shows that early functional rehabilitation is critical for optimal recovery after Achilles repair. For example, whether managed operatively or nonoperatively, patients who undergo an accelerated rehab program (early weight-bearing and mobilization) attain excellent strength, ROM, and return-to-activity outcomes [16]. Willits et al [16] demonstrated that nonoperative treatment with an aggressive rehab protocol yielded strength and functional scores equivalent to surgical repair, highlighting rehab's importance in tendon healing. In practice, many components of Achilles rehab (joint mobilizations, progressive loading exercises, and balance training) can be instructed and progressed remotely. A recent small study even found that an unsupervised, home-based exercise program after Achilles repair achieved high VISA-A scores and no increase in complications compared to traditional supervised therapy [17]. These findings reinforce that, as long as evidence-based rehabilitation principles are followed, including protection in early healing and progressive loading to restore tendon capacity, outcomes after Achilles rupture can be excellent with either clinic or home-based guidance. The digital model simply provides a different delivery mechanism for the same key interventions, and the trial confirms it can do so without compromising efficacy.

Growing evidence in other musculoskeletal conditions bolsters the trial's implications. In total joint arthroplasty, multiple studies have shown telerehabilitation to be comparable to standard outpatient rehabilitation for restoring function. A comprehensive meta-analysis of 11 randomized controlled trials in posttotal knee replacement patients found no significant differences in pain reduction or functional improvement between telehealth and in-person rehabilitation, with telerehabilitation yielding similar gains in walking ability, strength, and ROM [18]. Notably, remote programs often come with lower health care usage and cost. Tsang et al [18] reported that telerehabilitation after knee replacement significantly reduced hospital visits and was more cost-effective than usual care. Likewise, technology-assisted rehabilitation after hip or knee arthroplasty has been shown to improve pain and mobility outcomes on par with or better than usual care, partly by enabling more frequent exercise and immediate feedback through apps or video coaching [19]. Beyond postsurgical care, tele-physiotherapy has proven effective in chronic orthopedic conditions. For instance, in knee and hip osteoarthritis and tendinopathies, real-time video- or phone-based rehabilitation produces pain relief and functional gains equivalent to face-to-face therapy [7,19]. Systematic reviews consistently conclude that telerehabilitation is a safe and effective modality for musculoskeletal rehabilitation, with outcomes in pain, strength, and quality of life similar to traditional care across diverse conditions [7,19].

Importantly, some studies suggested that telehealth models can confer additional advantages, such as greater exercise adherence and empowerment. Patients often become more actively involved in self-management when guided remotely, which can enhance long-term exercise habits. However, clinicians implementing telerehabilitation should ensure proper patient selection and education. Key success factors include an initial in-person assessment when needed, clear exercise instructions (via videos or mobile apps), and regular follow-up to maintain accountability. When these conditions are met, the literature strongly supports that digital rehabilitation matches conventional physiotherapy in outcomes for Achilles tendon repair and beyond, offering a convenient alternative without loss of efficacy [18,19].

Interactive telerehabilitation may improve exercise adherence. In a randomized trial after total knee arthroplasty, patients using a guided telehealth platform had significantly higher adherence to prescribed exercises than the usual-care group [20]. The engaging digital interface and at-home convenience can motivate patients to complete rehabilitation exercises regularly.

Overall, patient satisfaction with telerehabilitation is high when implemented properly. During the COVID-19 pandemic, both patients and physiotherapists reported positive experiences with telehealth musculoskeletal care [21]. Patients often appreciate the flexibility and reduced travel burden, and many find remote coaching effective for guidance and feedback. That said, some individuals may miss hands-on contact; addressing patient education and ensuring a user-friendly technology experience are important for maintaining satisfaction.

Clinical Implication of the Study

Our cost-effectiveness analysis confirmed that the DEP intervention was a dominant strategy, providing equivalent effectiveness at a significantly lower cost compared to the clinic-based physiotherapy program. This study has significant clinical implications, indicating that home-based digital rehabilitation can effectively substitute traditional in-clinic physiotherapy following Achilles tendon repair. Clinicians can confidently implement structured telerehabilitation programs, which provide comparable clinical outcomes with additional benefits such as cost savings, convenience, and high patient adherence. Particularly, patients with geographic, logistical, or time constraints may benefit substantially from remotely supervised rehabilitation. Health care systems could leverage digital rehabilitation solutions to enhance resource allocation efficiency and reduce health care costs without compromising clinical quality. Moreover, physiotherapists can expand their service delivery through digital platforms, improving patient

XSL•FO
RenderX

engagement, continuity of care, and accessibility to expert guidance. However, careful patient selection and initial training remain essential to optimize telehealth outcomes, underscoring the necessity for ongoing professional education in digital rehabilitation practices.

Strength and Limitation

This study's strengths include its rigorous randomized controlled trial design, adequate sample size based on a clear noninferiority margin, and comprehensive assessment of both clinical and economic outcomes. Adherence and satisfaction were meticulously monitored, enhancing practical insights into patient acceptability of digital rehabilitation. However, limitations must be acknowledged. First, blinding participants and therapists was inherently impossible, potentially introducing performance bias. Second, follow-up was relatively short; thus, longer-term outcomes remain uncertain. Third, generalizability may be restricted, as the study design focused predominantly on younger, technology-proficient patients. Fourth, the use of the VISA-A score, a scale originally validated for chronic tendinopathy, has inherent limitations in acute rupture patients [13,22]. Some VISA-A items (eg, 30-minute walking endurance) are not feasible in early postoperative rehabilitation, making the score insensitive to early improvements [13,22]. Moreover, VISA-A changes do not necessarily reflect structural tendon healing, and the instrument is generally less responsive to recovery progress in acute ruptures compared to the ATRS [11]. We included the VISA-A as a complementary patient-reported measure of tendon-related symptoms, but we interpret its findings alongside objective outcomes (ankle ROM and calf strength) for a more comprehensive assessment [11]. Given our cohort's relatively young mean age and requirement for smartphone use, older adults and those with limited digital literacy were underrepresented. These groups often face barriers to telerehabilitation (eg, limited internet access and difficulty using smartphone interfaces), so the generalizability of DEP to these populations is uncertain. Finally, despite rigorous adherence monitoring, actual exercise intensity and fidelity to prescribed techniques in home settings were not objectively quantified, possibly influencing rehabilitation outcomes.

Conclusion

Home-based digital exercise after Achilles tendon repair was clinically noninferior to traditional clinic-based physiotherapy, providing similar functional outcomes, safety, greater patient convenience, adherence, and significant cost savings. These findings support broader adoption of structured telerehabilitation programs, offering effective, accessible, and patient-centered postoperative care. Further research should evaluate long-term outcomes and applicability across diverse patient populations.

Acknowledgments

We sincerely thank all the patients who participated in this study. The authors also confirm that no generative artificial intelligence (AI) tools (such as ChatGPT) were used in the writing of this manuscript.

Funding

No external financial support or grants were received for this work.

Data Availability

The datasets analyzed during this study are not publicly available due to a confidential policy by Shanghai Sixth People's Hospital and Xinhua Hospital, but are available from the corresponding author upon reasonable request.

Authors' Contributions

HW contributed to writing the original draft, software development, methodology, and formal analysis. TW contributed to writing the original draft, methodology, and data curation. JH, LH, and YM contributed to the investigation. SL contributed to writing (review and editing), supervision, and methodology. JZ contributed to writing (review and editing), conceptualization, and project administration. SL and JZ are cocorresponding authors of this manuscript, and JZ is the primary corresponding author responsible for the majority of the correspondence and inquiries regarding this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Physiotherapeutic Achilles-Specific Exercise (PASE) program aligned with early accelerated rehabilitation protocols.
[DOCX File, 46 KB - mhealth_v14i1e78100_app1.docx]

Multimedia Appendix 2
Follow-up attendance by group and scheduled visit.
[XLSX File, 9 KB - mhealth_v14i1e78100_app2.xlsx]

XSL•FO
RenderX

Checklist 1
CONSORT-eHEALTH checklist (version 1.6.1).
[PDF File, 1145 KB - mhealth_v14i1e78100_app3.pdf]

References

1. Lemme NJ, Li NY, DeFroda SF, Kleiner J, Owens BD. Epidemiology of Achilles tendon ruptures in the United States: athletic and nonathletic injuries from 2012 to 2016. Orthop J Sports Med 2018 Nov;6(11):2325967118808238. [doi: 10.1177/2325967118808238] [Medline: 30505872]
2. Briggs-Price S, Mangwani J, Houchen-Wolloff L, et al. Incidence, demographics, characteristics and management of acute Achilles tendon rupture: an epidemiological study. PLoS ONE 2024;19(6):e0304197. [doi: 10.1371/journal.pone.0304197] [Medline: 38905182]
3. Zhao JG, Meng XH, Liu L, Zeng XT, Kan SL. Early functional rehabilitation versus traditional immobilization for surgical Achilles tendon repair after acute rupture: a systematic review of overlapping meta-analyses. Sci Rep 2017 Jan 5;7:39871. [doi: 10.1038/srep39871] [Medline: 28054658]
4. Olsson N, Nilsson-Helander K, Karlsson J, et al. Major functional deficits persist 2 years after acute Achilles tendon rupture. Knee Surg Sports Traumatol Arthrosc 2011 Aug;19(8):1385-1393. [doi: 10.1007/s00167-011-1511-3] [Medline: 21533539]
5. Pastora-Bernal JM, Martín-Valero R, Barón-López FJ, Estebanez-Pérez MJ. Evidence of benefit of telerehabitation after orthopedic surgery: a systematic review. J Med Internet Res 2017 Apr 28;19(4):e142. [doi: 10.2196/jmir.6836] [Medline: 28455277]
6. Shambushankar AK, Jose J, Gnanasekaran S, Kaur G. Cost-effectiveness of telerehabilitation compared to traditional in-person rehabilitation: a systematic review and meta-analysis. Cureus 2025 Feb;17(2):e79028. [doi: 10.7759/cureus.79028] [Medline: 40099085]
7. Post AA, Rio EK, Sluka KA, et al. Efficacy of telehealth for movement-evoked pain in people with chronic Achilles tendinopathy: a noninferiority analysis. Phys Ther 2023 Mar 3;103(3):pzac171. [doi: 10.1093/ptj/pzac171] [Medline: 37172125]
8. Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG, CONSORT Group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. JAMA 2012 Dec 26;308(24):2594-2604. [doi: 10.1001/jama.2012.87802] [Medline: 23268518]
9. Wagner E, Gould J, Bilen E, Fleisig GS, Wilk K, Fowler R. Change in plantarflexion strength after complete detachment and reconstruction of the Achilles tendon. Foot Ankle Int 2004 Nov;25(11):800-804. [doi: 10.1177/107110070402501108] [Medline: 15574239]
10. Brorsson A, Willy RW, Tranberg R, Grävare Silbernagel K. Heel-rise height deficit 1 year after Achilles tendon rupture relates to changes in ankle biomechanics 6 years after injury. Am J Sports Med 2017 Nov;45(13):3060-3068. [doi: 10.1177/0363546517717698] [Medline: 28783473]
11. Dams OC, Reininga IHF, Zwerver J, Diercks RL, van den Akker-Scheek I. The Achilles tendon total rupture score is a responsive primary outcome measure: an evaluation of the Dutch version including minimally important change. Knee Surg Sports Traumatol Arthrosc 2020 Oct;28(10):3330-3338. [doi: 10.1007/s00167-020-05924-7] [Medline: 32146524]
12. de Mesquita GN, de Oliveira MNM, Matoso AER, de Moura Filho AG, de Oliveira RR. Cross-cultural adaptation and measurement properties of the Brazilian Portuguese version of the Victorian Institute of Sport Assessment-Achilles (VISA-A) questionnaire. J Orthop Sports Phys Ther 2018 Jul;48(7):567-573. [doi: 10.2519/jospt.2018.7897] [Medline: 29690827]
13. Young R, Nix S, Wholohan A, Bradhurst R, Reed L. Interventions for increasing ankle joint dorsiflexion: a systematic review and meta-analysis. J Foot Ankle Res 2013 Nov 14;6(1):46. [doi: 10.1186/1757-1146-6-46] [Medline: 24225348]
14. Axelsson E. Equivalence and non-inferiority trials in the evaluation of non-pharmacological interventions: rationale, challenges and recommendations. BMJ Open 2025 Aug 31;15(8):e102996. [doi: 10.1136/bmjopen-2025-102996] [Medline: 40887112]
15. Mauri L, D'Agostino RB Sr. Challenges in the design and interpretation of noninferiority trials. N Engl J Med 2017 Oct 5;377(14):1357-1367. [doi: 10.1056/NEJMra1510063] [Medline: 28976859]
16. Willits K, Amendola A, Bryant D, et al. Operative versus nonoperative treatment of acute Achilles tendon ruptures: a multicenter randomized trial using accelerated functional rehabilitation. J Bone Joint Surg Am 2010 Dec 1;92(17):2767-2775. [doi: 10.2106/JBJS.I.01401] [Medline: 21037028]
17. Mitrogiannis L, Mitrogiannis G, Melaki K, Karamitros A, Karantanas A, Paxinos O. An early functional unsupervised rehabilitation protocol allows safe return to function after Achilles tendon repair. Cureus 2024 Jan;16(1):e52477. [doi: 10.7759/cureus.52477] [Medline: 38371156]
18. Tsang MP, Man GCW, Xin H, Chong YC, Ong MTY, Yung PSH. The effectiveness of telerehabilitation in patients after total knee replacement: a systematic review and meta-analysis of randomized controlled trials. J Telemed Telecare 2024 Jun;30(5):795-808. [doi: 10.1177/1357633X221097469] [Medline: 35549756]
19. Wang X, Hunter DJ, Vesentini G, Pozzobon D, Ferreira ML. Technology-assisted rehabilitation following total knee or hip replacement for people with osteoarthritis: a systematic review and meta-analysis. BMC Musculoskelet Disord 2019 Nov 3;20(1):506. [doi: 10.1186/s12891-019-2900-x] [Medline: 31679511]

XSL•FO
RenderX

20. Nuevo M, Rodríguez-Rodríguez D, Jauregui R, et al. Telerehabilitation following fast-track total knee arthroplasty is effective and safe: a randomized controlled trial with the ReHub® platform. Disabil Rehabil 2024 Jun;46(12):2629-2639. [doi: 10.1080/09638288.2023.2228689] [Medline: 37403684]
21. Bennell KL, Lawford BJ, Metcalf B, et al. Physiotherapists and patients report positive experiences overall with telehealth during the COVID-19 pandemic: a mixed-methods study. J Physiother 2021 Jul;67(3):201-209. [doi: 10.1016/j.jphys.2021.06.009] [Medline: 34147399]
22. Robinson JM, Cook JL, Purdam C, et al. The VISA-A questionnaire: a valid and reliable index of the clinical severity of Achilles tendinopathy. Br J Sports Med 2001 Oct;35(5):335-341. [doi: 10.1136/bjsm.35.5.335] [Medline: 11579069]

Abbreviations

ATRS: Achilles Tendon Total Rupture Score
CONSORT: Consolidated Standards of Reporting Trials
CP: clinic-based physiotherapy
DEP: digital exercise program
ICER: incremental cost-effectiveness ratio
ITT: intention-to-treat
PP: per-protocol
ROM: range of motion
VISA-A: Victorian Institute of Sport Assessment–Achilles

Edited by L Buis; submitted 26.May.2025; peer-reviewed by G Si, W LaMendola; revised version received 27.Oct.2025; accepted 04.Feb.2026; published 19.Mar.2026.

Please cite as:
Wang H, Wu T, He J, Huang L, Ma Y, Xu M, Lu S, Zou J
Home-Based Digital Exercise Program for Patients After Open Repair of Acute Achilles Tendon Rupture: Noninferiority Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e78100
URL: https://mhealth.jmir.org/2026/1/e78100
doi:10.2196/78100

© Hui Wang, Tianyi Wu, Jiye He, Lihua Huang, Yanhong Ma, Manrong Xu, Shengdi Lu, Jian Zou. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 19.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Remote Patient Monitoring Program Components and Short-Term Hypertension Control: Retrospective Cohort Study

Sarah LaPointe[1,2], PhD; Michael Merrill[2], MS, MD, MBA

[1]Department of Epidemiology, Rollins School of Public Health, Emory University, 1518 Clifton Rd, Atlanta, GA, United States

[2]Research Department, Brook Health, Seattle, WA, United States

Corresponding Author:
Sarah LaPointe, PhD
Department of Epidemiology, Rollins School of Public Health, Emory University, 1518 Clifton Rd, Atlanta, GA, United States

Abstract

Background: Remote patient monitoring (RPM) is recommended for hypertension control. However, less is known about short-term responses of hypertension to RPM and which program components are most important for hypertension control.

Objective: This study aimed to evaluate the association between frequency of blood pressure monitoring, nurse monitoring, and their combination and hypertension control within 3 months.

Methods: This retrospective cohort study was conducted among a convenience sample of 1464 patients with hypertension enrolled in the Brook Remote Care RPM program who sought care at any of the 68 participating primary care clinics in New York or Massachusetts, United States, between 2021 and 2023. Patients with at least 3 blood pressure measures for 70% of the weeks they were in the program were defined as having adequate readings. Brook nurses monitored patients from certain clinics, whereas clinic nurses monitored other patients. Hypertension control was defined as a weekly average blood pressure of less than 140/90 mm Hg. Generalized linear models with a binomial specification and log link were used to estimate the prevalence ratio (PR) and 95% CI of the mutually adjusted associations between adequate blood pressure readings and Brook nurse monitoring, as well as their combination with hypertension control, at 4, 8, and 12 weeks, adjusted for patient age and sex.

Results: At weeks 4, 8, and 12, patients with adequate readings had 10% (PR=1.10, 95% CI 1.02-1.20), 12% (PR=1.12, 95% CI 1.04-1.20), and 15% (PR=1.15, 95% CI 1.07-1.24) higher prevalence of hypertension control than patients with lower frequencies of readings, respectively. Brook nurse monitoring was associated with 16% (PR=1.16, 95% CI 1.07-1.27), 6% (PR=1.06, 95% CI 0.99-1.15), and 8% (PR=1.08, 95% CI 1.00-1.16) higher prevalence of hypertension control at weeks 4, 8, and 12, respectively, compared to clinical staff monitoring. The combination of adequate readings plus Brook nurse monitoring was associated with 26% (PR=1.26, 95% CI 1.11-1.44), 17% (PR=1.17, 95% CI 1.05-1.31), and 27% (PR=1.27, 95% CI 1.12-1.43) higher prevalence of hypertension control at weeks 4, 8, and 12, respectively, compared to an inadequate number of blood pressure readings and not receiving Brook nurse monitoring.

Conclusions: This is the first study to compare outcomes for RPM patients using clinical nurse monitoring and independent staff monitoring. This study represents a contribution to the literature on clinical outcomes for RPM patients with a focus on short-term, rather than longer-term, hypertension control, evaluation of nurse monitoring conducted by Brook Health compared to clinical staff, and the joint association of monitoring frequency and Brook nurse monitoring on hypertension control. Our study findings suggest that both frequency of monitoring and specialized nurse monitoring improve hypertension control within 3 months of program enrollment. The combination of higher monitoring frequency and Brook nurse monitoring may serve as a cost-effective approach to hypertension control in high-risk populations that overcomes the limitations of clinical care staff.

(JMIR Mhealth Uhealth 2026;14:e69546) doi:10.2196/69546

KEYWORDS

remote patient monitoring; RPM; hypertension control; nurse monitoring; hypertension; engagement

Introduction

In the United States, the prevalence of hypertension in the adult population is 48%, and only 22% achieve hypertension control with a blood pressure of less than 140/90 mm Hg [1]. Patients with hypertension incur approximately US $2000 higher medical expenditures than their counterparts without hypertension, accounting for a total health care cost of US $130 billion [2]. While hypertension may contribute to premature mortality, patients can be treated, and progression can be prevented to achieve hypertension control.

The World Health Organization's guidelines on the treatment of hypertension in adults recommend telemonitoring and community- or home-based self-care to improve blood pressure

XSL•FO
RenderX

control as part of an integrated system, although the certainty of the evidence supporting these recommendations has been rated as low [3]. Patients with chronic conditions, including hypertension, have reported improved knowledge of their health status, more timely care, and better self-management of their conditions using remote monitoring [4]. A meta-analysis of 18 randomized controlled trials found that remote patient monitoring (RPM) lowered systolic blood pressure by 4.2 mm Hg and diastolic blood pressure by 2.4 mm Hg and increased the proportion of patients with hypertension control by 11% at 1 year of follow-up [5].

Beyond home monitoring, clinical guidelines recommend lifestyle changes and medication adherence for hypertension management and prevention [6]. Both lifestyle and pharmacological interventions have been shown to be cost-effective approaches for the reduction of blood pressure [7,8]. Thus, interest in the potential clinical benefits of joint programs that offer remote monitoring and lifestyle management for hypertension control has grown. In remote settings, nurses generally coordinate care for RPM and provide essential services for patients, including medication titration and lifestyle modification consultations [9-11].

Studies have shown that RPM programs coupled with care coordination are more effective for blood pressure reduction and hypertension control than RPM alone [12-14]. Previous studies have compared differences in average blood pressure and hypertension control between RPM and usual care groups at 3, 6, and 12 months [15,16]. However, results from a non–physician-led, entirely remote hypertension management program suggest that hypertension control is achievable at 7 weeks on average [17]. Moreover, an information and communications technology–based integrated care model showed clinically meaningful reductions in systolic blood pressure at 4 weeks [18].

A recent systematic review found that nurse-led telehealth interventions reduced blood pressure, more specifically, systolic blood pressure, based on randomized controlled trials and quasi-experimental studies [19]. Among the studies that were included in this review, few assessed the effectiveness of RPM and care coordination independently for hypertension control. Choi et al [13] found that patients who conducted remote blood pressure monitoring and received weekly virtual consultations had significantly lower systolic and diastolic blood pressure after 8 weeks than patients conducting blood pressure monitoring alone. Brennan et al [20] showed that privately insured African American individuals with hypertension who received a high-intensity, multimodal disease management program with nurse support had lower systolic blood pressure on average, had higher odds of achieving optimal blood pressure, and were more likely to record blood pressure at least once a week compared to a control group with a light-support educational program after 12 months. Hebert et al [21] compared a blood pressure monitoring intervention alone and nurse management in concert with blood pressure monitoring to usual care and found that the nurse management with blood pressure monitoring group had significantly lower systolic blood pressure at 9 months than the usual care group. No differences in systolic or diastolic blood pressure or hypertension control were observed for the blood pressure monitoring alone group at 9 or 18 months. While these studies demonstrate the effectiveness of nurse-led hypertension telehealth interventions, the specific components that affect these changes remain unclear. Moreover, all studies included in the aforementioned review by Kappes et al [19] assessed outcomes at 2 to 18 months. Therefore, there remains a gap in knowledge on the short-term independent and joint impacts of blood pressure monitoring and nurse monitoring among a diverse population of patients with hypertension.

Importantly, existing studies have explored whether nurse-led programs are effective for hypertension. There is a paucity of evidence on whether independent RPM companies with full-time nursing staff dedicated to monitoring patients offer more benefits than monitoring provided by clinical nursing staff. Given the costs of remote care [22], more research is needed to assess the short-term responses of hypertension to components of RPM and care coordination and whether RPM companies offer cost-effective programming. Therefore, the objective of this study was to examine the associations between both frequency of blood pressure readings and nurse monitoring and hypertension control at 4, 8, and 12 weeks for a population of patients with hypertension enrolled in an ongoing RPM program with nurse monitoring. We hypothesized that patients who recorded their blood pressure more frequently and who received monitoring from Brook nurses were more likely to achieve hypertension control within 3 months of program enrollment.

Methods

Study Design and Settings

The data used for this retrospective cohort study were drawn from a convenience sample of patients with hypertension enrolled in Brook Health's RPM program between December 2020 and December 2023. Brook Health is a commercial digital health company that offers continuous telemonitoring programs that include a mobile app with educational content and tracking tools, telemonitoring devices, health coaching, and remote nurse monitoring and care coordination to individuals with various chronic medical conditions (eg, diabetes, hypertension, obesity, chronic obstructive pulmonary disease, and congestive heart failure) who sought care at any of the 68 (36 as of December 2023) participating primary care clinics in New York or Massachusetts, United States.

Patient eligibility for Brook RPM referral is clinic specific; however, only patients with low or no copays were enrolled in the Brook RPM program. If interested and eligible, patients were then provided with remote monitoring instruments based on their diagnoses (ie, a blood pressure cuff, scale, pulse oximeter, and/or glucose monitor) and were provided with a health coach and nurse monitor. Duration of patient monitoring, coaching, and nursing is jointly determined by health care providers and patients, but patients are in the program for 12 to 15 months on average. The independent variables in this analysis included Brook nurse monitoring compared to clinical staff monitoring and frequency of blood pressure monitoring. The dependent variable was hypertension control within 12 weeks of program initiation.

XSL•FO
RenderX

Ethical Considerations

Approval for this secondary analysis of existing data without additional consent was provided by the University at Buffalo Institutional Review Board (00007012). Waiver of informed consent for primary and secondary data analysis was granted by the University at Buffalo Institutional Review Board as the study presents no more than minimal risk to patients. A deidentified, limited dataset was used for this study, and data access was limited to the research team to protect participant information. No compensation was provided to study participants as we used deidentified, secondary data. No identification of individual participants is possible in any images included in the manuscript or supplementary materials.

Variables

Patients enrolled in the Brook RPM hypertension program were provided with remote cellular blood pressure cuffs (BodyTrace and Tellihealth). Patients with obesity diagnoses were also provided with remote scales (BodyTrace and Withings) to monitor weight (in kilograms) and physical activity trackers (Fitbit) to monitor physical activity (in seconds). Weight was converted to pounds, and physical activity was converted to minutes.

Brook RPM Program Components

Brook conducted nurse monitoring for most participating clinics during the study years, whereas the remainder of the clinics conducted monitoring using their own nursing staff. Brook RPM nurse monitors and health coaches provided personalized recommendations to patients based on patient need and diagnosis. Generally, nurse monitors encouraged patients to record at least one blood pressure reading per day, although other patients were advised to record blood pressure more frequently depending on hypertension severity and presence of other comorbidities. Patients were instructed to record their blood pressure 1 to 2 hours after taking medication, before activity, or after 15 to 30 minutes of postactivity rest.

Each patient on the monitoring service received a phone call from a nurse once a month. These monthly calls were used to check patients' status, as well as to reinforce medication compliance and adherence to device use. The monthly phone call interaction was also used to reinforce signs and symptoms for which they should call their physician or 911 to ensure that urgent and emergent care were obtained when necessary. Patients also received calls as needed based on routine readings for symptom checks and validations for any readings that were outside of the designated parameters; these calls offered opportunities for early intervention.

Patient systolic and diastolic blood pressure measured using remote blood pressure cuffs was directly transmitted to the Brook data server. The first date with a blood pressure recording was considered a patient's activation date. We calculated baseline values of systolic and diastolic blood pressure using the average of all blood pressures recorded in the first week following the activation date. To determine whether patients adhered to blood pressure monitoring protocols, we first counted the total number of weeks with at least 3 blood pressure measurements as Brook Health patients recorded blood pressure 3 days a week on average. We calculated the proportion of weeks with at least 3 blood pressure measurements with the total number of weeks in the program as the denominator for each patient. Finally, we defined patients with at least 70% of their weeks in the program with at least 3 blood pressure measurements as patients with adequate blood pressure readings. On average, Brook Health patients submitted blood pressure readings 3 days a week for more than half of the weeks they were enrolled in the program; therefore, we selected 70% of weeks in the program to capture the patients with highest monitoring activity.

Blood Pressure Outcomes

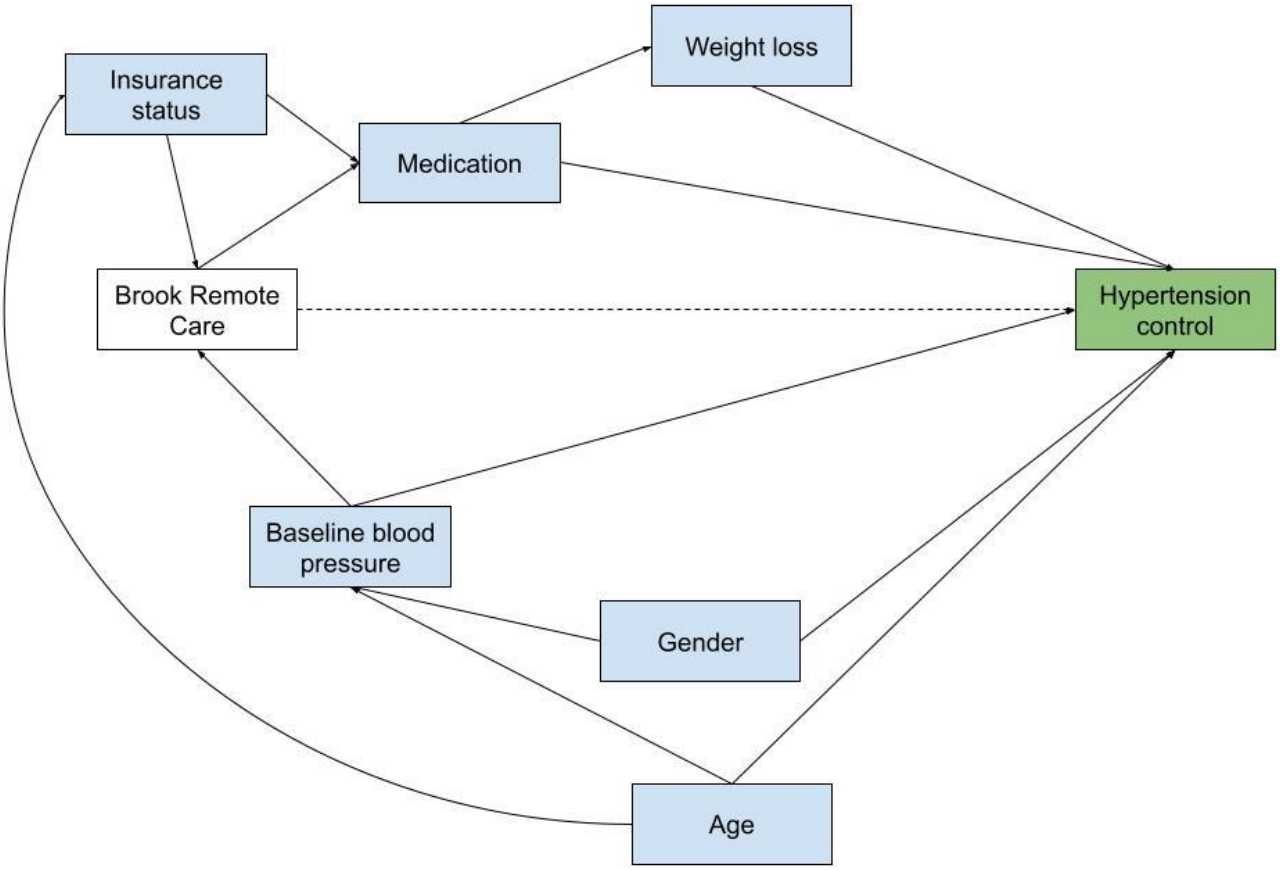
The primary outcome for this study was hypertension control, which we defined as an average weekly systolic blood pressure of less than 140 mm Hg and diastolic blood pressure of less than 90 mm Hg in weeks 4, 8, and 12 of the Brook RPM program.

Covariates

Potential confounders in this study were identified a priori using directed acyclic graphs (Figure 1). Patient age, sex, and health insurance type were collected from patient electronic health records. Diagnoses of patient comorbidities that informed patient care and follow-up were reported as binary indicators. Baseline weight (in pounds) and physical activity (in minutes) included the first measures for each patient.

XSL•FO
RenderX

Figure 1. Directed acyclic graph of the association between Brook Remote Care and hypertension control.



Statistical Analysis

Patient demographic and health characteristics were presented using counts and percentages for binary variables and medians and IQRs for continuous variables. We applied chi-square tests to assess the bivariate associations between blood pressure control at 4, 8, and 12 weeks and both adequate blood pressure readings and Brook nurse monitoring. To estimate the association between both adequate blood pressure readings and Brook nurse monitoring and hypertension control, we used generalized linear models with a log link and binomial family (*epitools* package) [23] mutually adjusted for Brook nurse monitoring and adequate blood pressure readings, respectively; age; and sex. All analyses were conducted using R (version 4.3.1; R Foundation for Statistical Computing) [24]. This study was conducted in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) reporting guidelines (Checklist 1).

Sensitivity Analysis

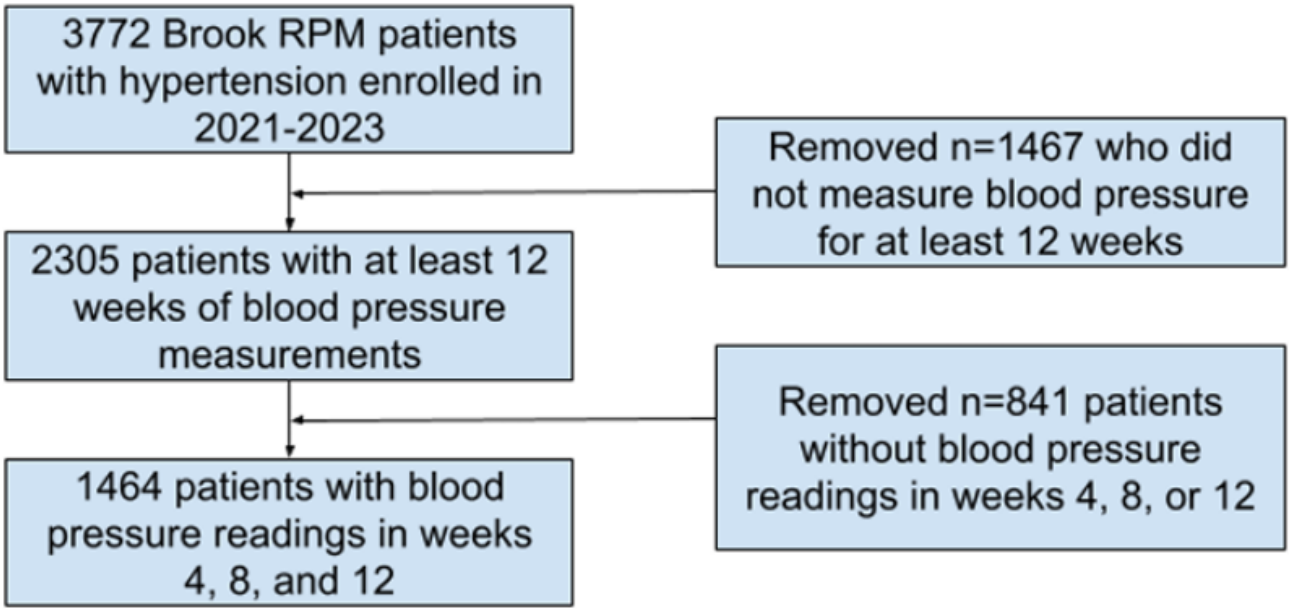
To control for the potential influence of other comorbidities on these associations, we conducted an analysis among a subset of patients diagnosed with hypertension only at 4, 8, and 12 weeks using adjusted generalized linear models with a log link and binomial family.

Results

Our study sample included 38.8% (1464/3772) of the patients enrolled in the Brook RPM program for hypertension (Figure 2). The median age of the patients was 71 (IQR 64-78) years (Table 1). In total, 52.9% (775/1464) of the patients were female, and 67.7% (991/1464) were enrolled in the Brook RPM program in 2023. At baseline, median systolic and diastolic blood pressure were 136.8 (IQR 126.6-149.0) mm Hg and 80.8 (IQR 73.0-88.4) mm Hg, respectively. At baseline, 12.2% (179/1464) of the patients were diagnosed with diabetes, 25.1% (368/1464) were diagnosed with obesity, 1.6% (23/1464) were diagnosed with chronic obstructive pulmonary disease, and 13.6% (199/1464) were diagnosed with congestive heart failure. Patient characteristics were similar when comparing the group of patients who recorded adequate blood pressure readings and received Brook nurse monitoring to those with neither (Multimedia Appendix 1).

XSL•FO
RenderX

Figure 2. Diagram of analytic sample selection of Brook remote patient monitoring (RPM).



XSL•FO
RenderX

Table . Characteristics of patients enrolled in the Brook remote patient monitoring program for hypertension overall, by presence of adequate readings, and by receipt of Brook nurse monitoring (N=1464).

Characteristic	Overall	Adequate blood pressure readings		Brook nurse monitoring	
		No (n=606)	Yes (n=858)	No (n=528)	Yes (n=936)
Age (y), median (IQR)	71.0 (64.0-78.0)	70.0 (62.0-77.0)	72.0 (66.0-79.0)	71.0 (60.0-80.0)	72.0 (66.0-78.0)
Sex, n/N (%)					
Female	775/1464 (52.9)	339/606 (55.9)	436/858 (50.8)	272/528 (51.5)	503/936 (53.7)
Male	689/1464 (47.1)	267/606 (44.1)	422/858 (49.2)	256/528 (48.5)	433/936 (46.3)
Health insurance type, n/N (%)					
Commercial	19/435 (4.4)	13/194 (6.7)	6/241 (2.5)	13/299 (4.3)	6/136 (4.4)
Medicaid	48/435 (11.0)	28/194 (14.4)	20/241 (8.3)	47/299 (15.7)	1/136 (0.7)
Medicare	368/435 (84.6)	153/194 (78.9)	215/241 (89.2)	239/299 (79.9)	129/136 (94.9)
Participants with missing data	1029/1464 (70.3)	412/606 (68.0)	617/858 (71.9)	229/528 (43.4)	800/936 (85.5)
Activation y, n/N (%)					
2021	77/1464 (5.3)	53/606 (8.7)	24/858 (2.8)	0/528 (0.0)	77/936 (8.2)
2022	396/1464 (27.0)	184/606 (30.4)	212/858 (24.7)	209/528 (39.6)	187/936 (20.0)
2023	991/1464 (67.7)	369/606 (60.9)	622/858 (72.5)	319/528 (60.4)	672/936 (71.8)
Baseline health metrics, median (IQR)					
Systolic blood pressure (mm Hg)	136.8 (126.6-149.0)	136.4 (126.5-149.0)	137.0 (126.7-148.5)	139.2 (127.8-150.3)	135.9 (126.0-148.1)
Diastolic blood pressure (mm Hg)	80.8 (73.0-88.4)	81.8 (74.0-90.0)	80.0 (72.5-87.8)	82.7 (73.8-91.1)	79.9 (72.8-87.0)
Weight (lbs)[a]	195.2 (163.1-236.9)	200.4 (169.3-238.3)	192.9 (159.8-232.1)	182.3 (150.6-216.2)	202.3 (170.4-240.6)
Physical activity (h)[b]	0.3 (0.1-0.6)	0.3 (0.1-0.7)	0.3 (0.0-0.5)	0.4 (0.3-0.7)	0.3 (0.0-0.5)
Baseline comorbidities, n/N (%)					
Diabetes					
No	1285/1464 (87.8)	527/606 (87.0)	758/858 (88.3)	513/528 (97.2)	772/936 (82.5)
Yes	179/1464 (12.2)	79/606 (13.0)	100/858 (11.7)	15/528 (2.8)	164/936 (17.5)
Obesity					
No	1096/1464 (74.9)	457/606 (75.4)	639/858 (74.5)	491/528 (93.0)	605/936 (64.6)
Yes	368/1464 (25.1)	149/606 (24.6)	219/858 (25.5)	37/528 (7.0)	331/936 (35.4)
COPD[c]					
No	1441/1464 (98.4)	599/606 (98.8)	842/858 (98.1)	525/528 (99.4)	916/936 (97.9)
Yes	23/1464 (1.6)	7/606 (1.2)	16/858 (1.9)	3/528 (0.6)	20/936 (2.1)
Congestive heart failure					
No	1265/1464 (86.4)	522/606 (86.1)	743/858 (86.6)	414/528 (78.4)	851/936 (90.9)
Yes	199/1464 (13.6)	84/606 (13.9)	115/858 (13.4)	114/528 (21.6)	85/936 (9.1)

[a]Missing 774 observations.

[b]Missing 1188 observations.

[c]COPD: chronic obstructive pulmonary disease.

In unadjusted bivariate analyses, Brook RPM patients who measured their blood pressure at least 3 times a week for at least 70% of the weeks they were enrolled in the program (ie, “adequate” readings) had higher proportions of hypertension control at 4 weeks (557/858, 64.9% vs 358/606, 59.1%; *P*=.03), 8 weeks (613/858, 71.4% vs 388/606, 64.0%; *P*=.003), and 12 weeks (622/858, 72.5% vs 379/606, 62.5%; *P*<.001) than patients who did not submit adequate readings (Table 2). A higher proportion of patients receiving Brook nurse monitoring achieved hypertension control at week 4 (616/936, 65.8% vs

XSL•FO
RenderX

299/528, 56.6%; *P*<.001), week 8 (653/936, 69.8% vs 348/528, 65.9%; *P*=.14), and week 12 (659/936, 70.4% vs 342/528, 64.8%; *P*=.03) than patients receiving nurse monitoring from nursing staff at their primary care clinics. Among the group of patients who recorded adequate blood pressure readings and received Brook nurse monitoring, 68.8% (379/551), 73.5% (405/551), and 74.2% (409/551) achieved hypertension control at 4, 8, and 12 weeks, respectively (Multimedia Appendix 2).

In multivariable analyses, patients with adequate blood pressure readings had higher prevalence of hypertension control at week 4 (prevalence ratio [PR]=1.10, 95% CI 1.02-1.20), week 8 (PR=1.12, 95% CI 1.04-1.20), and week 12 (PR=1.15, 95% CI 1.07-1.24) than those who did not have adequate readings (Table 3). In models mutually adjusted for adequate readings, patients who received Brook nurse monitoring had higher prevalence of hypertension control at week 4 (PR=1.16, 95% CI 1.07-1.27), week 8 (PR=1.06, 95% CI 0.99-1.15), and week 12 (PR=1.08, 95% CI 1.00-1.16) than patients monitored by staff at primary care clinics. Patients who recorded adequate blood pressure readings and received Brook nurse monitoring (772/1464, 52.7%) had higher prevalence of hypertension control at week 4 (PR=1.26, 95% CI 1.11-1.46), week 8 (PR=1.17, 95% CI 1.05-1.31), and week 12 (PR=1.27, 95% CI 1.12-1.43) than patients who recorded blood pressure less frequently and who received remote monitoring from clinical staff. These associations among adequate readings, nurse monitoring, and blood pressure control were generally similar, though slightly higher in magnitude, in the sensitivity analyses (Multimedia Appendix 3).

Table . Bivariate associations between both adequate blood pressure readings and Brook nurse monitoring and hypertension control at 4, 8, and 12 weeks (N=1464).

	Adequate readings			Brook nurse monitoring		
	No (n=606), n (%)	Yes (n=858), n (%)	*P* value[a]	No (n=528), n (%)	Yes (n=936), n (%)	*P* value[a]
Hypertension control at 4 wk			.03			<.001
No	248 (40.9)	301 (35.1)		229 (43.4)	320 (34.2)	
Yes	358 (59.1)	557 (64.9)		299 (56.6)	616 (65.8)	
Hypertension control at 8 wk			.003			.14
No	218 (36.0)	245 (28.6)		180 (34.1)	283 (30.2)	
Yes	388 (64.0)	613 (71.4)		348 (65.9)	653 (69.8)	
Hypertension control at 12 wk			<.001			.03
No	227 (37.5)	236 (27.5)		186 (35.2)	277 (29.6)	
Yes	379 (62.5)	622 (72.5)		342 (64.8)	659 (70.4)	

[a]Pearson chi-square test.

Table . Associations between Brook Remote Care program components and hypertension control at 4, 8, and 12 weeks in the program. Models were adjusted for patient age and sex.

	Hypertension control prevalence ratio (95% CI)		
	4 wk	8 wk	12 wk
Adequate readings (N=1464)			
No	Reference	Reference	Reference
Yes	1.10 (1.02-1.20)	1.12 (1.04-1.20)	1.15 (1.07-1.24)
Brook nurse monitoring (N=1464)			
No	Reference	Reference	Reference
Yes	1.16 (1.07-1.27)	1.06 (0.99-1.15)	1.08 (1.00-1.16)
Adequate readings and Brook nurse monitoring combined (n=772)			
No	Reference	Reference	Reference
Yes	1.26 (1.11-1.44)	1.17 (1.05-1.31)	1.27 (1.12-1.43)

Discussion

Principal Findings

This longitudinal retrospective cohort study among a convenience sample of patients at any of the 68 participating primary care clinics across New York and Massachusetts, United States, enrolled in the Brook RPM program for hypertension examined the associations between 2 Brook RPM program components—Brook nurse monitoring relative to clinical nurse monitoring and frequency of blood pressure readings—and hypertension control at 4, 8, and 12 weeks. Patients who

XSL•FO
RenderX

recorded blood pressure more than 70% of their time in the program had higher prevalence of hypertension control than those who did not. Brook nurse monitoring was associated with greater prevalence of hypertension control than clinical nursing staff monitoring. The group of patients with high frequency of blood pressure readings and who received Brook nurse monitoring had the highest prevalence of hypertension control compared to all other groups.

Evaluation of the frequency of blood pressure readings recorded by patients is important to guide clinical recommendations for RPM. Frazier et al [25] observed that mean arterial pressure reductions were greatest for patients with higher frequencies of measurements per month over the course of a 6-month RPM program. Despite the Centers for Medicare and Medicaid Services stipulation that Current Procedural Terminology codes for initial equipment setup and monthly data collection are billed only for patients who record at least 16 measurements per month [26], there is no evidence documenting the clinical relevance of these suggested monitoring frequencies. Therefore, more studies evaluating the clinical implications of monitoring intensity are needed.

Team-based care has been recommended as a cost-effective approach for patients' hypertension control and management [27]. Our study observed that, compared to nurse monitoring provided by clinical staff, Brook nurse monitoring was consistently associated with higher hypertension control among patients, highlighting that the quality of nurse interactions may be more impactful for hypertension control than simply having a nurse monitor present. A pragmatic observational cohort study among Medicare patients with uncontrolled hypertension who sought care at primary care clinics in Chicago, Illinois (age range 65-85 years), and the general hypertension population found that RPM plus care coordination provided by nurses showed clinically meaningful improvements in hypertension control at 3 and 6 months compared to usual care and RPM alone [12]. Additionally, a cluster randomized trial of the HyperLink RPM program that provided pharmacist-led care coordination to patients attending primary care in Minnesota showed that the telemonitoring intervention had 25% to 30% higher absolute hypertension control at 6 and 12 months than usual care [28]. An experimental longitudinal study among African American individuals with uncontrolled hypertension found that a 12-month nurse-managed telemonitoring program that provided RPM and nurse calls to patients led to statistically and clinically significant reductions in systolic blood pressure compared to usual care, with the greatest reductions observed in the first 3 months of the program [9]. Moreover, in a mediation analysis, secure messaging with a pharmacist accounted for 96% of the effects of an RPM program on hypertension control after 12 months [29].

A unique contribution of our Brook RPM study is the short-term hypertension control response examined at 4, 8, and 12 weeks in relation to program components. Most studies have evaluated digital health interventions in relation to hypertension control and blood pressure outcomes at longer periods, such as 6 and 12 months of program participation [19,30]. However, there is evidence that RPM shows promise for improving blood pressure outcomes at earlier time points [17,18,31]. Naqvi et al [31] observed an 18.4 mm Hg reduction in systolic blood pressure among those using RPM after 3 months compared to controls. A non–physician-led hypertension management program that included a Bluetooth-enabled home blood pressure device was shown to lead to hypertension control for 91% of patients after approximately 7 weeks of consistent measurement [17]. A study among 20 patients with difficult-to-control hypertension enrolled in an information and communications technology–based program that included telemonitoring and weekly access to nurses showed that systolic blood pressure measured at home decreased by 15 mm Hg on average after 4 weeks [18]. In concert with these study findings, our study showed improved blood pressure control among patients who recorded blood pressure more frequently and who received Brook nurse monitoring at 4, 8, and 12 weeks.

Limitations

Our study findings have limited generalizability as Brook RPM patients are older, reside in western New York or eastern Massachusetts, and are generally enrolled in Medicare health insurance plans that cover RPM costs with low to no patient cost sharing. Patients who received Brook nurse monitoring were identified based on primary care clinics, as Brook nurses conduct monitoring for patients at certain clinics. This definition masks the variability in individual patient experiences as monitoring protocols are tailored to patient needs. Moreover, the level of monitoring provided by clinical staff and the differences between staff within the same clinic as well as between clinics are unclear as clinical monitoring protocols are not made available to Brook Health. Additionally, most Brook patients were encouraged to monitor their blood pressure at least once a day. However, we used a more data-driven estimate of monitoring intensity to define adequate blood pressure readings in this study (ie, 70% of weeks in the program with at least 3 days of blood pressure readings) as most patients did not consistently record their blood pressure every day for all weeks in the program. The use of a complete-case analysis may introduce bias into our study, and we encourage cautious interpretation of study findings. Additionally, uncontrolled confounding may bias our study findings as we lacked information on important demographic and clinical characteristics that likely confound these associations, and health insurance information was incomplete and not accounted for in our analysis due to a high degree of missingness. For instance, baseline blood pressure used to determine eligibility for the Brook Health program was not available for this study as it was based on clinical medical records, and patient medication adherence was only available through nurse notes, which were not complete. Therefore, our study may be subject to residual confounding that may bias our results. However, it is more likely that these variables do not act as true confounders but rather serve as mediators between the program's nurse monitoring component and hypertension control.

Conclusions

This is the first study to compare short-term hypertension control outcomes for RPM patients using clinical nurse monitoring and independent RPM staff monitoring. Our study findings suggest that both frequency of monitoring and specialized nurse

XSL•FO
RenderX

monitoring improve hypertension control within 3 months of program enrollment. The combination of higher monitoring frequency and Brook nurse monitoring may represent a cost-effective approach to improving hypertension control in high-risk populations by addressing limitations in clinical care staffing. Professional RPM companies with full-time staff may augment the benefits of RPM provided by clinical care providers who have other responsibilities. Future studies should further evaluate the impact of individual and collective program components on health outcomes and assess these potential impacts at various time points of program participation.

Acknowledgments

The authors are grateful to Kelsea Hoover and Alex Foley for their contributions to the program description and details.

Funding

This research was funded by Brook Health.

Data Availability

The datasets generated or analyzed during this study are not publicly available to protect patient information but are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: SL, MM
Formal analysis: SL
Methodology: SL
Writing—original draft: SL
Writing—review and editing: SL, MM

Conflicts of Interest

SL is employed as a part-time researcher at Brook Health, and MM serves as the chief medical officer at Brook Health.

Multimedia Appendix 1
Characteristics of patients enrolled in the Brook remote patient monitoring program for hypertension overall, by presence of adequate readings, and by receipt of Brook nurse monitoring.
[PDF File, 94 KB - mhealth_v14i1e69546_app1.pdf]

Multimedia Appendix 2
Bivariate associations between adequate blood pressure readings and Brook nurse monitoring combined and hypertension control at 4, 8, and 12 weeks (n=772).
[PDF File, 79 KB - mhealth_v14i1e69546_app2.pdf]

Multimedia Appendix 3
Associations between Brook Remote Care program components and hypertension control at 4, 8, and 12 weeks in the program among patients diagnosed with hypertension only (n=864).
[PDF File, 75 KB - mhealth_v14i1e69546_app3.pdf]

Checklist 1
STROBE checklist.
[PDF File, 155 KB - mhealth_v14i1e69546_app4.pdf]

References

1. Hypertension cascade: hypertension prevalence, treatment and control estimates among US adults aged 18 years and older applying the criteria from the American College of Cardiology and American Heart Association's 2017 hypertension guideline—NHANES 2017–2020. : Centers for Disease Control and Prevention; 2023 URL: https://millionhearts.hhs.gov/data-reports/hypertension-prevalence.html [accessed 2026-03-03]
2. Kirkland EB, Heincelman M, Bishu KG, et al. Trends in healthcare expenditures among US adults with hypertension: national estimates, 2003-2014. J Am Heart Assoc 2018 May 30;7(11):e008731. [doi: 10.1161/JAHA.118.008731] [Medline: 29848493]

XSL•FO
RenderX

3. Guideline for the pharmacological treatment of hypertension in adults. World Health Organization. 2021. URL: https://www.who.int/publications/i/item/9789240033986 [accessed 2026-03-03]
4. Walker RC, Tong A, Howard K, Palmer SC. Patient expectations and experiences of remote monitoring for chronic diseases: systematic review and thematic synthesis of qualitative studies. Int J Med Inform 2019 Apr;124:78-85. [doi: 10.1016/j.ijmedinf.2019.01.013] [Medline: 30784430]
5. Cappuccio FP, Kerry SM, Forbes L, Donald A. Blood pressure control by home monitoring: meta-analysis of randomised trials. BMJ 2004 Jul 17;329(7458):145. [doi: 10.1136/bmj.38121.684410.AE] [Medline: 15194600]
6. Whelton PK, Carey RM, Aronow WS, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension 2018 Jun;71(6):e13-e115. [doi: 10.1161/HYP.0000000000000065] [Medline: 29133356]
7. Rubinstein A, Colantonio L, Bardach A, et al. Estimation of the burden of cardiovascular disease attributable to modifiable risk factors and cost-effectiveness analysis of preventative interventions to reduce this burden in Argentina. BMC Public Health 2010 Oct 20;10:627. [doi: 10.1186/1471-2458-10-627] [Medline: 20961456]
8. Neter JE, Stam BE, Kok FJ, Grobbee DE, Geleijnse JM. Influence of weight reduction on blood pressure: a meta-analysis of randomized controlled trials. Hypertension 2003 Nov;42(5):878-884. [doi: 10.1161/01.HYP.0000094221.86888.AE] [Medline: 12975389]
9. Artinian NT, Flack JM, Nordstrom CK, et al. Effects of nurse-managed telemonitoring on blood pressure at 12-month follow-up among urban African Americans. Nurs Res 2007;56(5):312-322. [doi: 10.1097/01.NNR.0000289501.45284.6e] [Medline: 17846552]
10. Bosworth HB, Olsen MK, McCant F, et al. Hypertension Intervention Nurse Telemedicine Study (HINTS): testing a multifactorial tailored behavioral/educational and a medication management intervention for blood pressure control. Am Heart J 2007 Jun;153(6):918-924. [doi: 10.1016/j.ahj.2007.03.004] [Medline: 17540191]
11. Laurant M, Reeves D, Hermens R, Braspenning J, Grol R, Sibbald B. Substitution of doctors by nurses in primary care. Cochrane Database Syst Rev 2005 Apr 18(2):CD001271. [doi: 10.1002/14651858.CD001271.pub2] [Medline: 15846614]
12. Persell SD, Petito LC, Anthony L, et al. Prospective cohort study of remote patient monitoring with and without care coordination for hypertension in primary care. Appl Clin Inform 2023 May;14(3):428-438. [doi: 10.1055/a-2057-7277] [Medline: 36933552]
13. Choi WS, Kim NS, Kim AY, Woo HS. Nurse-coordinated blood pressure telemonitoring for urban hypertensive patients: a systematic review and meta-analysis. Int J Environ Res Public Health 2021 Jun 27;18(13):6892. [doi: 10.3390/ijerph18136892] [Medline: 34199019]
14. Bosworth HB, Olsen MK, Grubber JM, et al. Two self-management interventions to improve hypertension control: a randomized trial. Ann Intern Med 2009 Nov 17;151(10):687-695. [doi: 10.7326/0003-4819-151-10-200911170-00148] [Medline: 19920269]
15. McManus RJ, Mant J, Bray EP, et al. Telemonitoring and self-management in the control of hypertension (TASMINH2): a randomised controlled trial. Lancet 2010 Jul 17;376(9736):163-172. [doi: 10.1016/S0140-6736(10)60964-6] [Medline: 20619448]
16. Omboni S. Connected health in hypertension management. Front Cardiovasc Med 2019 Jun 13;6:76. [doi: 10.3389/fcvm.2019.00076] [Medline: 31263703]
17. Fisher ND, Fera LE, Dunning JR, et al. Development of an entirely remote, non-physician led hypertension management program. Clin Cardiol 2019 Feb;42(2):285-291. [doi: 10.1002/clc.23141] [Medline: 30582181]
18. Visco V, Finelli R, Pascale AV, et al. Difficult-to-control hypertension: identification of clinical predictors and use of ICT-based integrated care to facilitate blood pressure control. J Hum Hypertens 2018 Jul;32(7):467-476. [doi: 10.1038/s41371-018-0063-0] [Medline: 29713051]
19. Kappes M, Espinoza P, Jara V, Hall A. Nurse-led telehealth intervention effectiveness on reducing hypertension: a systematic review. BMC Nurs 2023 Jan 17;22(1):19. [doi: 10.1186/s12912-022-01170-z] [Medline: 36650463]
20. Brennan T, Spettell C, Villagra V, et al. Disease management to promote blood pressure control among African Americans. Popul Health Manag 2010 Apr;13(2):65-72. [doi: 10.1089/pop.2009.0019] [Medline: 20415618]
21. Hebert PL, Sisk JE, Tuzzio L, et al. Nurse-led disease management for hypertension control in a diverse urban community: a randomized trial. J Gen Intern Med 2012 Jun;27(6):630-639. [doi: 10.1007/s11606-011-1924-1] [Medline: 22143452]
22. American Diabetes Association. Economic costs of diabetes in the U.S. in 2017. Diabetes Care 2018 May;41(5):917-928. [doi: 10.2337/dci18-0007] [Medline: 29567642]
23. Aragon TJ, Fay MP, Wollschlaeger D, Omidpanah A. epitools: epidemiology tools. The Comprehensive R Archive Network. 2020. URL: https://cran.r-project.org/web/packages/epitools/index.html [accessed 2026-03-03]
24. R Core Team. R: a language and environment for statistical computing. The Comprehensive R Archive Network. 2025. URL: https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf [accessed 2026-03-03]
25. Frazier WD, Beins M, DaVanzo J, Heath S, Dobson A. Six months of remote patient monitoring is associated with blood pressure reduction in hypertensive patients: an uncontrolled observational study. Telemed J E Health 2023 Aug;29(8):1164-1170. [doi: 10.1089/tmj.2022.0418] [Medline: 36576990]

XSL•FO
RenderX

26. Telehealth and remote patient monitoring. Telehealth.HHS.gov. URL: https://telehealth.hhs.gov/providers/best-practice-guides/telehealth-and-remote-patient-monitoring/billing-remote-patient#:~:text=Remote%20physiologic%20monitoring%2C%20but%20not,be%20medically%20reasonable%20and%20necessary [accessed 2026-01-31]
27. Jacob V, Reynolds JA, Chattopadhyay SK, et al. Economics of team-based care for blood pressure control: updated community guide systematic review. Am J Prev Med 2023 Oct;65(4):735-754. [doi: 10.1016/j.amepre.2023.04.013] [Medline: 37121447]
28. Margolis KL, Asche SE, Bergdall AR, et al. Effect of home blood pressure telemonitoring and pharmacist management on blood pressure control: a cluster randomized clinical trial. JAMA 2013 Jul 3;310(1):46-56. [doi: 10.1001/jama.2013.6549] [Medline: 23821088]
29. Ralston JD, Cook AJ, Anderson ML, et al. Home blood pressure monitoring, secure electronic messaging and medication intensification for improving hypertension control: a mediation analysis. Appl Clin Inform 2014;5(1):232-248. [doi: 10.4338/ACI-2013-10-RA-0079] [Medline: 24734136]
30. Katz ME, Mszar R, Grimshaw AA, et al. Digital health interventions for hypertension management in US populations experiencing health disparities: a systematic review and meta-analysis. JAMA Netw Open 2024 Feb 5;7(2):e2356070. [doi: 10.1001/jamanetworkopen.2023.56070] [Medline: 38353950]
31. Naqvi IA, Strobino K, Kuen Cheung Y, et al. Telehealth after stroke care pilot randomized trial of home blood pressure telemonitoring in an underserved setting. Stroke 2022 Dec;53(12):3538-3547. [doi: 10.1161/STROKEAHA.122.041020] [Medline: 36314123]

Abbreviations

PR: prevalence ratio
RPM: remote patient monitoring
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by S Brini; submitted 02.Dec.2024; peer-reviewed by V Alikhani, Y Wang; revised version received 31.Jan.2026; accepted 02.Feb.2026; published 24.Mar.2026.

Please cite as:
LaPointe S, Merrill M
Remote Patient Monitoring Program Components and Short-Term Hypertension Control: Retrospective Cohort Study
JMIR Mhealth Uhealth 2026;14:e69546
URL: https://mhealth.jmir.org/2026/1/e69546
doi:10.2196/69546

© Sarah LaPointe, Michael Merrill. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 24.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effects of the mHealth Supportive Care Program for Family Caregivers of Individuals With Dementia and Diabetes: Pilot Randomized Controlled Trial

Xinying Lin[1*], MM; Miaoqian Huang[1,2*], MM; Tuantuan Guo[1,3], MM; Rong Lin[1], PhD; Chenshan Huang[1], PhD; Yuanjiao Yan[1], PhD; Hong Li[1], PhD

[1]The School of Nursing, Fujian Medical University, No.1, Xuefu North Road, Minhou County, Fuzhou, Fujian, China

[2]Xiangping Street Community Health Service Center, Tong'an District, XiaMen, China

[3]Department of General Practice, The Second Affiliated Hospital of Fujian Medical University, Quanzhou, China

[*]these authors contributed equally

Corresponding Author:
Hong Li, PhD
The School of Nursing, Fujian Medical University, No.1, Xuefu North Road, Minhou County, Fuzhou, Fujian, China

Abstract

Background: The comorbidity of dementia and type 2 diabetes mellitus exacerbates the burden on family caregivers (FCGs). Mobile health (mHealth) technology offers a promising alternative to overcome the spatiotemporal limitations of traditional interventions, but evidence for its efficacy in supporting dementia–type 2 diabetes mellitus caregivers remains scarce.

Objective: This study aimed to evaluate the effectiveness of an mHealth supportive care program for FCGs of individuals with dementia and diabetes, focusing on caregiver burden, social support, and dementia care knowledge.

Methods: A 2-arm, parallel-group randomized controlled trial was conducted. Between September 2022 and January 2023, FCGs were recruited from 5 urban and 10 rural communities under a community health center in Xiamen, China. Eligible caregivers were legally related to the patient, providing care for more than 8 hours per day for at least 1 month, conscious adults with basic literacy, owning and able to use a smartphone, and willing to provide informed consent. Their care recipients met diagnostic criteria for both dementia and type 2 diabetes, aged more than 60 years. Participants were randomly allocated (1:1) to intervention (n=30) or wait-list control (n=30). The intervention group received a 12-week mHealth supportive care program via the "Xiamen i-Health" platform, comprising 6 core modules (updated biweekly) and on-demand teleconsultation, in addition to conventional offline health education. The control group received conventional monthly 1-hour home-visit health education only. The primary outcome was caregiver burden measured by the Caregiver Burden Inventory (CBI). Secondary outcomes included social support (Social Support Rating Scale; SSRS) and dementia care knowledge (Dementia Care Knowledge Scale; DCKS). Assessments were performed at baseline (T0) and 3-month postintervention (T1). Only data collectors and statistical analysts were blinded.

Results: Of 108 potential participants, 60 were randomly assigned. Per-protocol analysis included 55 participants (intervention group n=28 and control n=27). Postintervention, the intervention group showed a significantly greater reduction in CBI scores compared to the control group (between-group difference, Z=−3.534, P<.001, r=0.477). The intervention group also demonstrated significantly greater improvements in SSRS (Z=2.494, P=.01) and DCKS scores (Z=−4.233, P<.001, r=0.570). Subgroup analyses revealed that the reduction in caregiver burden was more pronounced among male, younger (60 y), physical labor, and lower-income caregivers. No intervention-related adverse events were reported.

Conclusions: This theoretically grounded mHealth supportive care program effectively reduced burden and improved outcomes for FCGs managing dementia-diabetes comorbidity. The integrated online-offline delivery model shows particular promise for male, younger, manual labor, and lower-income caregivers, suggesting mHealth's potential to address health equity. These findings provide a feasible model for scalable caregiver support in resource-limited primary care settings. Future research should involve multicenter trials with longer follow-up and cost-effectiveness analyses.

Trial Registration: ClinicalTrials.gov NCT06785857; https://clinicaltrials.gov/study/NCT06785857

(JMIR Mhealth Uhealth 2026;14:e72012) doi:10.2196/72012

KEYWORDS

senile dementia; diabetes; family caregivers; mobile health; supportive care; randomized controlled trial

XSL•FO
RenderX

Introduction

Dementia care remains a significant public health challenge for health systems worldwide. By 2019, the economic cost of dementia had escalated to $1.3 trillion globally [1], posing a substantial financial strain. In China, the older population (age ≥60 y) living with dementia has reached 15.07 million, with incidence rates increasing markedly with age [2]. While pharmacological treatments delay disease progression to some extent, the absence of definitive curative measures leads to patients progressively losing their self-care, work, and social abilities [3]. This dramatically increases the demand for dementia care, requiring round-the-clock care for advanced disease [4]. Notably, in China, approximately 70% of patients with dementia rely on family caregivers (FCGs), who bear a disproportionately heavy burden and face numerous challenges. The caregiving challenge becomes markedly more complex when dementia coexists with type 2 diabetes mellitus (T2DM). Epidemiological studies demonstrate that T2DM elevates Alzheimer disease risk by 1.5 to 2 times [5], driven by shared pathophysiology including insulin resistance and oxidative stress that accelerate cognitive decline [6]. The progression from mild cognitive impairment to dementia occurs in 6% to 25% of patients with T2DM, markedly higher than the 0.2% to 3.9% observed in age-matched controls [5]. This comorbidity leads to compounded cognitive deterioration [7] and severely impairs self-care management [8] and long-term quality of life [9]. Patients often struggle with medication adherence, frequently forgetting or duplicating doses, which results in dangerous glycemic fluctuations that may further exacerbate cognitive dysfunction [10]. China's metabolic disease epidemic further amplifies this issue, with over 140.9 million diabetes cases [11], 95% being T2DM [12]. The clinical complexity of dementia-T2DM comorbidity generates substantially greater caregiver burdens, as FCGs must concurrently manage dementia-related behavioral symptoms and diabetes-specific care requirements, including meticulous medication supervision, dietary control, and hypoglycemia prevention. These patients exhibit highly variable symptomatology [13,14] and extremely high care dependency [15], yet existing research predominantly focuses on patient self-management rather than critical external determinants like social support, caregiving contexts, or health care resource availability [16-18].

As a vital extension of the health care system, FCGs play an indispensable role in caring for patients with dementia and diabetes [19,20]. Enhancing FCG care skills and addressing their care needs are pivotal to mitigating public health pressures and improving patient quality of life. In developing countries, the paucity of primary health care resources shifts a greater proportion of caregiving responsibilities to FCGs, thereby augmenting their burden. While FCGs, as unpaid caregivers [21], may derive emotional fulfillment and personal growth from caregiving [22], they face numerous challenges, including psychological, emotional, physical, and financial burdens [23], which often lead to social isolation and reduced quality of life and physical health [24-27]. In China, strong family ties and a profound sense of responsibility socially obligate family members to undertake caregiving duties, placing an emotional burden on them when caring for frail relatives [28]. Studies indicate that caregivers are more likely to choose institutionalization when experiencing high emotional stress and burden [29], and those under greater stress face a 63% higher mortality rate compared to others [30]. Consequently, caregivers are frequently referred to as "hidden patients" [31].

Furthermore, as FCGs are often instructed to prioritize managing patients' symptoms [32], their needs are frequently overlooked and unacknowledged by health and social care professionals [33]. Patients with dementia and diabetes require intricate and meticulous care due to cognitive decline and metabolic abnormalities, increasing caregiving challenges and placing higher demands on FCGs. According to a survey by the Alzheimer's Disease Society of China, the primary challenges faced by the FCGs of patients with dementia include inadequate caregiving skills, limited care resources, and restricted treatment services. The level of care often exceeds the FCGs' capabilities, potentially increasing their burden and worsening mental health issues [34]. Developing effective strategies to support caregiving is imperative. Targeted interventions to train and support FCGs can enhance their competence [35], reducing caregiver stress and burden.

Social support and caregiving knowledge are critical for dementia-combined diabetes caregivers. Studies have shown that educational interventions can significantly reduce caregiver anxiety and burden and improve patient quality of life [36,37], while unmet information needs can lead to reduced caregiving outcomes [38]. Werner et al [39] noted that caregivers need timely, personalized, and accessible information support. Effective interventions (eg, providing education and practical information) can improve knowledge, reduce stress, and enhance caregiving [40-42]. In addition, caregiver needs span emotional social support, knowledge and skills training, and self-management [43], but existing resources often lack credibility or have mismatched learning styles [44]. Digital support tools (eg, teleconsultation) can provide flexible and professional guidance that can improve caregiver mental health [45] and reduce caregiver stress and burden [46-48], and multicomponent interventions are particularly effective [49,50].

The World Health Organization emphasizes the importance of supporting families caring for older adults in its guidelines for community-based integrated care [51]. Advancements in mobile health (mHealth) technology, which overcome time and space constraints and expand the intervention scope, have led to its widespread application in chronic disease management and home care [52,53]. Internet-based supportive interventions have been implemented in countries such as the United Kingdom [54], Germany [55], and India [56], offering convenient and effective support to alleviate the physical and mental burdens of caregiving. However, few studies in China have examined such interventions, and most lack theoretical framework support and standardization. The following key gaps remain in existing caregiver support research. First, the majority of current studies only address caregivers of individuals with dementia only [57,58] or diabetes only [59,60], while interventions for dementia-T2DM co-morbidities caregivers remain scarce. Second, less than 30% of digital intervention studies have articulated their theoretical underpinnings, limiting

XSL•FO
RenderX

reproducibility [61,62]. Third, systematic reviews have found incomplete reporting of the implementation process of home and community-based interventions in existing studies (only 28% of studies reported on implementation outcomes) [63], which limits the generalized use of interventions. Fourth, there are fewer relevant studies in China, and most lack theoretical framework support and standardized protocols.

Roy's Adaptation Model and Fitch's Theory of Supportive Care are widely used theoretical frameworks for constructing care programs. Roy's Adaptation Model explains how individuals cope with environmental changes through adaptation, focusing on their adaptation level, stimulating factors, and adaptive mechanisms. While this model provides valuable insights into individual coping dynamics, its emphasis on adaptive processes rather than structured care delivery makes it less suitable for guiding standardized intervention design. In contrast, Fitch's Theory of Supportive Care addresses and optimizes patients' physiological, psychological, social, and environmental care needs, emphasizing the importance of personalized care planning, which is more appropriate for this study. Since its inception, this theory has been extensively applied in research addressing supportive care needs for patients and FCGs both domestically and internationally [64-68], yielding notable outcomes. Traditional intervention programs often rely on in-person or telephone formats, focusing mainly on health education while neglecting supportive care needs. Additionally, limited human resources in primary health care services constrain traditional offline interventions, creating significant implementation challenges. While some research platforms have explored online interventions, these often lack transparency in service fees and operate on a small scale, limiting their generalization and promotion. In this context, mHealth supportive care programs offer unique advantages by overcoming the resource and implementation challenges of traditional interventions and providing a novel pathway to meet supportive care needs with promising development prospects.

Building upon Fitch's Theory of Supportive Care [69] and evidence from cross-sectional surveys, we developed an integrated online-offline mHealth supportive care program targeting FCGs of individuals with comorbid dementia and diabetes in urban-rural grassroots communities. This study aims to clarify the intervention effects of the mHealth supportive care program, with the primary outcome being caregiver burden measured by the Caregiver Burden Inventory (CBI), and secondary outcomes including social support assessed by the Social Support Rating Scale (SSRS) and dementia care knowledge measured by the Dementia Care Knowledge Scale (DCKS). Additionally, subgroup analyses were conducted to evaluate potential heterogeneity of intervention effects across different demographic characteristics such as gender, age, careers, and monthly family income. We hypothesize that, at 3-month follow-up, (a) FCGs receiving the mHealth intervention will demonstrate a significantly greater reduction in CBI scores compared with routine-care controls, and (b) the intervention group will show significantly larger improvements in SSRS and DCKS scores; furthermore, (c) intervention effects may differ across demographic subgroups. Findings from this study will provide evidence for scalable supportive care strategies to address caregiving disparities in resource-limited settings.

Methods

Study Design and Objectives

This 2-arm, parallel-group randomized controlled trial evaluated the efficacy of an mHealth supportive care program for FCGs of individuals with comorbid dementia and diabetes. The trial was designed and reported in accordance with the CONSORT (Consolidated Standards of Reporting Trials) 2025 statement [70] (see Checklist 1) and its extension for reporting trials of electronic and mobile health applications (CONSORT-EHEALTH [Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth]) [71] (see Checklist 2).

Sample Size

Simple Size Calculation

The sample size was calculated based on a 1:1 ratio between the intervention and control groups, using the sample size calculation method for comparing the means of two samples. The calculation formula is as follows:n1=n2=2[(μα+μβ)δ/σ]+142μα 2

Based on a literature review, the caregiver burden was selected as the primary outcome measure for this study. n_1 and n_2 represent the required sample sizes for the two groups, respectively. Due to the lack of prior data from identical designs, the sample size calculation was based on the SD (σ=6.85) and clinically important difference (δ=6.87) reported in a quasi-experimental study [72], which tested a dementia caregiver intervention under comparable conditions. Power analysis (α=.05, β=.1, 2-tailed) indicated 22 participants per group. During the recruitment process, participants showed great enthusiasm in signing up, with the number of applicants exceeding the actual required number by 16. In consideration of ethical principles and the fact that a larger sample size would enhance reliability, we ultimately enrolled 30 per group (total N=60) to further strengthen robustness. Given the exploratory nature of this study, the sample size was estimated based on prior effect sizes, but statistical power may still be limited.

Participants

Participate Enrollment

This study was conducted at a community health center in Xiamen, Fujian Province, from September 2022 to January 2023. Participants were recruited from 5 urban neighborhood committees and 10 rural villages under the jurisdiction of the selected health center to ensure urban-rural diversity. Due to COVID-19 pandemic restrictions and limited research resources, a single community health center was chosen as the study site.

Inclusion Criteria

The inclusion criteria for care recipients are as follows:

- meeting the diagnostic criteria for dementia
- meeting the diagnostic criteria for type 2 diabetes
- aged ≥60 years and permanent resident in the study area

XSL•FO
RenderX

The inclusion criteria for caregivers are as follows:

- legally related to the patient (spouse, child, or sibling), excluding paid caregivers
- providing care ≥8 hours per day, ≥5 days per week for ≥1 month [51]
- conscious adult with basic literacy and comprehension skills
- owning a smartphone with internet access and proficiency in its use
- willing to provide informed consent and participate voluntarily

Providing care ≥8 hours/day, ≥5 days/week for ≥1 month [51]

Exclusion Criteria

Patients or caregivers meeting any of the following conditions were excluded:

- having severe physical illness, major diseases, or terminal conditions
- having substance or alcohol dependence
- currently participating in similar intervention studies

Withdrawal Criteria

Participation would be terminated in the following conditions:

- if the care recipient was hospitalized for severe illness or passed away
- if the participant voluntarily withdrew from the study for any reason

Randomization, Allocation Concealment, and Blinding

Random allocation was set at a 1:1 ratio. Participants who were ultimately enrolled in the study were numbered in the order of enrollment. A digital custodian, who did not directly participate in this study, used a computer to generate a random number sequence from 1 to 60. The custodian placed each random number in sequentially numbered, opaque envelopes, which were then sealed and stored to reduce selection and information bias. Researchers inquired about the group assignment of participants by calling the digital custodian. The custodian opened the envelopes in numerical order and informed the researchers of the group assignments. Participants with random numbers 1 - 30 were assigned to the intervention group, and those with 31 - 60 were assigned to the control group. Since this study was based on an mHealth-supported care intervention, blinding the intervention implementers and participants was challenging. Therefore, blinding was only implemented for data collectors and statistical analysts.

Interventions

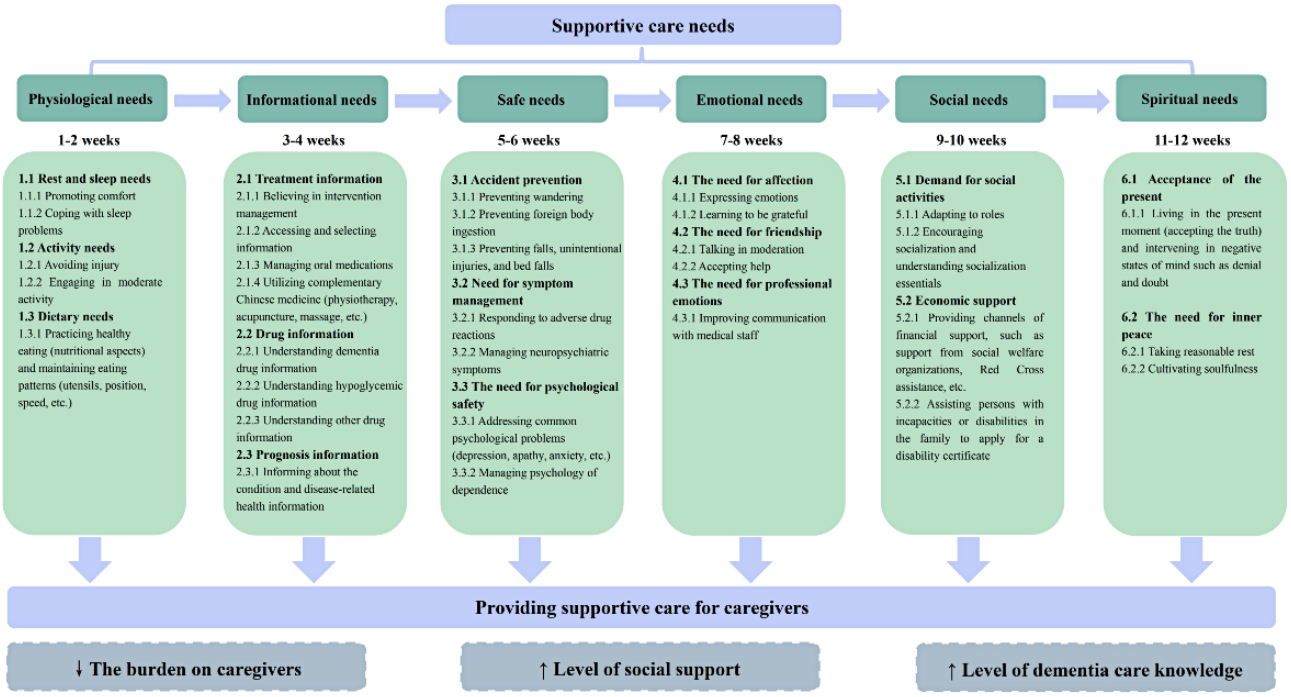
All routine care for patients in both groups was independently performed by FCGs. The control group received offline basic health services and health education, whereas the intervention group received an mHealth supportive care program in addition to the services provided to the control group. The intervention was developed and reported following the TIDieR (Template for Intervention Description and Replication) checklist (see Checklist 3).

Following standard community geriatric chronic disease management protocols, control group participants received monthly home visits (1 hour per session for 12 wk) where implementers delivered targeted health education and addressed caregiver-initiated queries regarding dementia-diabetes comorbidity management. The educational program comprised the following: (a) Diabetes management training: systematic instruction on essential care components, including nutritional regulation, sleep optimization, and exercise regimens; pharmacological guidance included drug categorization, therapeutic mechanisms, dosage administration, and compliance enhancement techniques; training on glycemic monitoring protocols focused on accurate fasting and 2-hour postprandial blood glucose assessment; instruction included early detection of hypoglycemic episodes, ophthalmologic complication signs, crisis management, and dermatological protection measures. (b) Dementia management training: Comprehensive instruction on basic caregiving competencies, including nutritional support, sleep cycle regulation, and physical activity facilitation; educational content encompassed dementia etiology, prognostic trajectories, stage-dependent symptom management, and prevention protocols for elopement, falls, and scald injuries.

The mHealth supportive care intervention was developed based on Fitch's theoretical model [69], which systematically addresses caregivers' multidimensional needs. The intervention's theoretical structure is illustrated in Figure 1; the 6 core intervention modules directly operationalize Fitch's domains. This intervention program was further informed by baseline epidemiological surveys quantifying caregiving burdens, needs analyses of FCGs, and evidence synthesis through a systematic literature review and Delphi expert consensus. Implementation utilized a robust, municipally administered public health mobile infrastructure to guarantee program fidelity and continuity. The intervention group accessed supplementary weekly mHealth interventions through the "Xiamen i-Health" portal (12 structured sessions across 12 weeks with fortnightly module updates), complemented by asynchronous consultation capabilities. Preintervention preparations and specific steps for implementation are documented in Multimedia Appendix 1. The implementation team comprised professionals with decade-long gerontological nursing practice and 6-year specialized intervention management tenure. Rigorous centralized training was conducted to maintain intervention fidelity, mitigate operational variability, and standardize delivery protocols.

XSL•FO
RenderX

Figure 1. The mHealth supportive care program based on Fitch's supportive care model for caregivers of patients with dementia and type 2 diabetes: design and intervention components.



Data Collection and Outcome Measures

Self-Designed Questionnaire

The self-designed questionnaire, developed based on the research objectives and a review of relevant literature [64,72-74], gathered demographic and clinical information, including age, gender, careers, educational levels, monthly family income, living with patient, relationship with patients, as well as current medical history and family history. The questionnaire was evaluated and refined by an expert panel of 5 clinical nursing specialists, all with more than a decade of experience.

Primary Outcome

The CBI was used to evaluate the burden experienced by the FCGs of individuals with dementia and diabetes. Zhang [75] adapted this scale from the Chinese version of the CBI developed by Taiwanese scholar Chen et al. [73], with modifications to suit the linguistic characteristics of mainland China. Each item is scored using a Likert 5-point scale ranging from strongly agree (4 points) to strongly disagree (0 points). The total score of the scale ranges from 0 to 96, with higher scores indicating a heavier caregiver burden. Scores ≥24 suggest the need for therapeutic intervention, whereas scores ≥36 indicate a risk of severe burden. The overall Cronbach α coefficient for this scale [75] is 0.77, with an average content validity index (CVI) of 0.95 per item, indicating good reliability and validity. It has been widely used in similar studies [76,77].

Secondary Outcomes

The SSRS is a 10-item rating scale that evaluates the psychological support an individual receives in social life and their utilization of such support. The Chinese scholar Shuiyuan [78] developed this scale based on the sociocultural context and familial structure characteristic of Chinese society, ensuring its relevance and validity for use in Chinese populations. The objective support score is the sum of items 2, 4, and 6; the subjective support score is the sum of items 1, 3, 4, and 5; and utilization of support score is the sum of items 8, 9, and 10. Generally, scores below 20 indicate low social support, 20 - 30 indicate moderate social support, and 30 - 40 indicate satisfactory social support. The test-retest reliability of this scale [78] is 0.92, with item consistency ranging from 0.89 to 0.94 and a Cronbach α coefficient of 0.89. The scale demonstrates good reliability and validity, accurately reflecting an individual's level of social support. It has been widely used in similar studies [79,80].

The DCKS was used to assess the FCGs' extent of knowledge about dementia caring [81]. This questionnaire, based on Maslow's hierarchy of needs, covers personal hygiene, diet, excretion, sleep, disease and rehabilitation, medication, environmental and travel safety, social relationships and family ties, self-esteem, and self-actualization. A Likert scale was used, with 1 point for a correct answer and 0 points for an incorrect answer or "don't know." The DCKS scores range from 0 to 22, with higher scores reflecting a higher level of dementia care knowledge. The Cronbach α coefficient for the overall scale [81] is 0.626, with CVI ranging from 0.86 to 1 and an average CVI index of 0.95. It has been widely used in similar studies [82].

Adverse Events Monitoring

As this trial evaluated a low-risk supportive care intervention, no systematic, active surveillance for specific adverse events was prespecified. However, any unintended effects or concerns raised by participants during home visits or via the mHealth platform were documented. Furthermore, all instances of participant withdrawal were tracked and reviewed for any potential link to the intervention.

XSL•FO
RenderX

Statistical Analysis

Data analysis was conducted using IBM SPSS Statistics for Windows version 25.0. Quantitative data with a normal distribution were presented as mean (SD), whereas nonnormally distributed data were presented as median with IQR. Categorical data were presented as frequency and percentage. For comparisons between 2 independent samples, the independent-sample *t* test was used for normally distributed data, whereas the Mann-Whitney *U* test was used for nonnormally distributed data. Comparisons pre- and post-intervention were conducted using a paired *t* test for normally distributed differences or Wilcoxon signed-rank test for nondistributed differences. The χ^2 test or Fisher exact test was used for comparisons of categorical data between groups, and rank-sum tests were used for comparisons of ordinal data. Effect sizes were calculated as Cohen *d* (continuous) or rank correlation r (nonparametric), with 95% CIs reported where applicable. Given the 2-time-point design of this pilot RCT and the completeness of the per-protocol dataset, repeated measures ANOVA was selected as the primary framework for testing group × time interactions and the critical 3-way (group × time × subgroup) interactions in preplanned subgroup analyses. Where significant interactions were detected ($P<.05$), simple effects analyses were performed. The dataset for the final per-protocol analysis was complete, with no missing data at the item level for any outcome measures among participants who completed the follow-up.

Ethical Considerations

This study was approved by the Ethics Committee of Fujian Provincial Hospital (approval K2021-05-024). Due to delays in the institutional ethics approval process during the COVID-19 pandemic, this trial was retrospectively registered. The trial is registered at ClinicalTrials.gov (NCT06785857; Date: Jan 21, 2025). No changes were made to the trial methods or outcomes after the trial commenced.

Prior to enrollment, all participating caregivers provided written informed consent after being fully informed of the study objectives, procedures, and their right to withdraw at any time without penalty. To ensure confidentiality, all collected data were deidentified and stored securely, with access limited to authorized researchers only. Although no monetary compensation was offered, small gifts (soap, tissue, shopping bags) valued at approximately 20 RMB (equivalent to approximately US $2.97 based on the average exchange rate during the intervention period in 2022) were provided as tokens of appreciation during home visits. We further confirm that no identifiable images, audio, video, or text of participants are included in this manuscript or its supplementary materials. No deviations from the prespecified trial protocol occurred after the commencement of the study.

No patients or members of the public were involved in the design, conduct, reporting, or dissemination of this research.

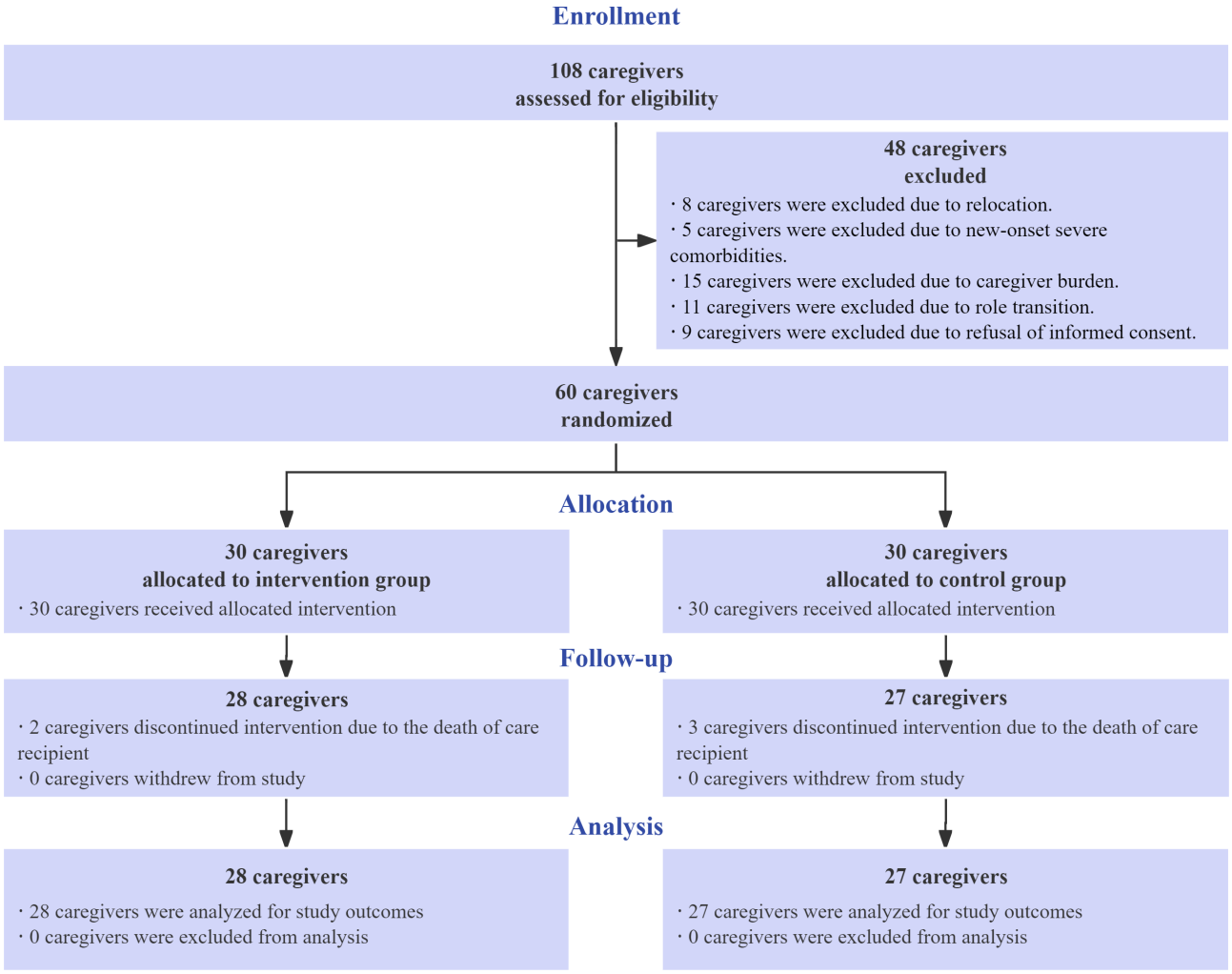
Results

Participants

A total of 108 potential participants were assessed for eligibility in this study. Out of the 60 who were eligible, 48 were excluded due to scheduling conflicts and a lack of interest. Consequently, 60 individuals agreed to participate and were randomly assigned to either the intervention group (n=30) or the control group (n=30) after providing informed consent. Herein, due to the death of care recipients during intervention, the number of remaining participants was 28 in the intervention group and 27 in the control group. Ultimately, a total of 55 out of 60 (91.67%) participants completed the follow-up assessments (Figure 2).

XSL•FO
RenderX

Figure 2. Consolidated Standards of Reporting Trials (CONSORT) flow diagram of participant enrollment, allocation, follow-up, and analysis in a 2-arm, parallel-group, pilot randomized controlled trial evaluating an mHealth supportive care program for family caregivers of individuals with dementia and type 2 diabetes in Xiamen, China (September 2022 to January 2023).



Demographic and Baseline Characteristics

Table 1 presents the sociodemographic characteristics of the caregivers. No significant differences were observed in the characteristics between the intervention and control groups (P>.05), indicating that no significant differences between the groups at baseline. Over half (27/55, 49.09%) of the FCGs were aged ≥60. Most caregivers were married (47/55, 85.45%), and 72.73% (40/55) had completed education at the junior high school and above level. Notably, 90.91% (50/55) of caregivers lived with the patients, and 76.36% (42/55) were the patients' children.

XSL•FO
RenderX

Table . Comparison of basic demographic information of two groups of caregivers in a pilot randomized controlled trial: caregivers of patients with dementia and type 2 diabetes in Xiamen, China, 2022 - 2023.

Items and categories	Control group, n (%)	Intervention group, n (%)	Chi-square (*df*)	*P* value
Gender			0.022 (1)	.88[a]
Male	12 (44.4)	13 (46.4)		
Female	15 (55.6)	15 (53.6)		
Age (y)			1.480 (1)	.22[a]
<60	11 (40.74)	16 (57.14)		
≥60	16 (59.26)	12 (42.86)		
Educational level			1.438 (1)	.16[b]
Primary and below	9 (33.33)	6 (21.43)		
Junior high school and above	18 (66.67)	22 (78.57)		
Marriage			<0.001 (1)	≥.99[a]
Married	23 (85.19)	24 (88.89)		
Unmarried, divorced, widowed, etc	4 (14.81)	4 (14.81)		
Careers			0.379 (1)	0.71[b]
Physical labor	20 (29.63)	20 (32.14)		
Mental labor	7 (14.82)	8 (17.86)		
Living with patient			0.262 (1)	.61[b]
Yes	24 (88.89)	26 (92.86)		
No	3 (11.11)	2 (7.14)		
Relationship with patients			0.059 (1)	.81[a]
Spouse	6 (22.22)	7 (25)		
Children	21 (78.78)	21 (75)		
Monthly family income (RMB)[c]			0.165 (2)	.92[b]
1000 - 3000	6 (22.22)	5 (17.86)		
3000 - 5000	9 (33.33)	10 (35.71)		
>5000	12 (44.45)	13 (46.43)		

[a]Refers to the χ^2 test.

[b]Refers to Fisher exact test.

[c]All income values are in RMB. Based on the 2022 exchange rate (US $1≈6.731 RMB), 1,000–3,000 RMB≈US $149–446, 3,000–5,000 RMB≈US $446–743, and >5,000 RMB≈US $>743.

Comparison of Outcome Measures Pre-intervention

The primary outcome measure in this study was caregiver burden, assessed using the CBI. Secondary outcome measures included caregiver social support level and knowledge level, which were evaluated using the SSRS and DCKS, respectively. Preintervention, no significant differences were detected between the intervention and control groups in CBI (z=−0.675, P=.50), SSRS (t_{53}=−0.949, P=.35), and DCKS (t_{53}=−1.256, P=.22), indicating that the groups were not statistically different (Table 2).

XSL•FO
RenderX

Table . Comparison of outcome measures (pre-intervention) between two groups of caregivers in a randomized controlled trial of mHealth intervention for caregivers of patients with dementia and type 2 diabetes in Xiamen, China, 2022 - 2023.

Items	Total (N=55)	Control group (n=27)	Intervention group (n=28)	Statistic	*P* value	Cohen *d*/*r*
SSRS[a], median (IQR)	24.00 (22.00 to 26.00)	24.00 (22.00 to 26.00)	25.00 (23.25 to 26.00)	−0.949 (53)[b]	.35[c]	Cohen *d*=0.26 (−0.28 to 0.79)
CBI[d], median (IQR)	57.00 (53.00 to 60.00)	56.00 (53.00 to 60.00)	57.00 (54.00 to 60.75)	−0.675[e]	.50[f]	*r*=0.09 (0.01 to 0.38)
DCKS[g], median (IQR)	11.00 (10.00 to 13.00)	12.00 (10.00 to 16.00)	11.00 (10.00 to 12.75)	−1.256 (53)[b]	.22[c]	Cohen *d*=0.34 (−0.19 to 0.87)

[a]SSRS: Social Support Rating Scale.

[b]*t* test value (*df*)

[c]Normally distributed variables are presented as mean (SD) (independent *t* test).

[d]CBI: Caregiver Burden Inventory.

[e]*z* score.

[f]Nonnormal variables are represented as median (IQR) (Mann-Whitney *U* test).

[g]DCKS: Dementia Caring Knowledge Scale.

Comparison of Outcome Measures Post-intervention

Postintervention, significant differences were observed between the intervention and control groups in SSRS (z=2.494, P=.01), CBI (z=−3.534, P<.001), and DCKS scores (z=−4.233, P<.001). The results are presented in Table 3.

Table . Comparison of outcome measures (postintervention) between two groups of caregivers in a randomized controlled trial of mHealth intervention for caregivers of patients with dementia and type 2 diabetes in Xiamen, China, 2022 - 2023.

Items	Total (N=55)	Control group (n=27)	Intervention group (n=28)	*z*	*P* value	*r*
SSRS[a], median (IQR)	31.00 (29.00-33.00)	31.00 (28.00-32.00)	32.00 (30.00-34.00)	2.494	.01[b]	0.34 (0.09-0.55)
CBI[c], median (IQR)	50.00 (48.00-52.00)	52.00 (49.00-55.00)	48.00 (46.00-50.75)	−3.534	<.001[b]	0.48 (0.24-0.68)
DCKS[d], median (IQR)	18.00 (16.00-19.00)	16.00 (15.00-18.00)	19.00 (18.00-19.00)	−4.233	<.001[b]	0.57 (0.35-0.75)

[a]SSRS: Social Support Rating Scale.

[b]Nonnormal variables are represented as median (IQR) (Mann-Whitney *U* test).

[c]CBI: Caregiver Burden Inventory.

[d]DCKS: Dementia Caring Knowledge Scale.

Comparison of Outcome Measures in the Intervention Group Pre- and Post-intervention

Within the intervention group, statistically significant differences were observed in SSRS (t_{27}=−14.959, P<.001), CBI (t_{27}=11.415, P<.001), and DCKS scores (z=4.637, P<.001) pre- and postintervention (Table 4).

XSL•FO
RenderX

Table . Comparison of outcome measures (pre- and post-intervention) in the intervention group of a randomized controlled trial of mHealth intervention for caregivers of patients with dementia and type 2 diabetes in Xiamen, China, 2022 - 2023.

Items	Preintervention (n=28)	Postintervention (n=28)	Statistic	*P* value	Cohen *d*/*r*
SSRS[a], median (IQR)	25.00 (23.25-26.00)	32.00 (30.00-34.00)	−14.959[b](27)	<.001[c]	Cohen *d*=2.83 (1.99-3.66)
CBI[d], median (IQR)	57.00 (54.00-60.75)	48.00 (46.00-50.75)	11.415[b](27)	<.001[c]	Cohen *d*=2.16 (1.47-2.83)
DCKS[e], median (IQR)	11.00 (10.00-12.75)	19.00 (18.00-19.00)	4.637[f]	<.001[g]	*r*=0.86[h]

[a]SSRS: Social Support Rating Scale.

[b]*t* test value (*df*).

[c]Normally distributed variables are presented as mean (SD) (paired *t* test).

[d]CBI: Caregiver Burden Inventory.

[e]DCKS: Dementia Caring Knowledge Scale.

[f]z score.

[g]Nonnormal variables are represented as median (IQR) (Mann-Whitney *U* test).

[h]Regarding the Wilcoxon signed-rank test, there is currently no consensus in the statistical literature on a standardized method for calculating confidence intervals for its effect size (eg, *r*), and commonly used statistical software (eg, SPSS, R) does not provide this functionality.

Subgroup Analysis of Intervention Effects by Caregiver Characteristics

The 3-way interaction analyses indicated that changes in SSRS and DCKS scores were not significantly moderated by the examined demographic variables (all 3-way interactions *P*>.05; see Multimedia Appendix 2). For CBI score changes, 3-way interaction tests revealed significant time×group×gender ($F_{1,51}$=6.712, *P*=.01), time×group×age ($F_{1,51}$=7.784, *P*=.007), time×group×careers ($F_{1,51}$=8.327, *P*=.006), time×group×educational level ($F_{1,51}$=8.325, *P*=.006), and time×group×monthly family income ($F_{1,51}$=3.793, *P*=.03); therefore, subgroup analyses were conducted by sex, age, careers, educational level, and monthly family income.

As shown in Table 5, the gender subgroup analysis indicated that male participants in the control group had significantly higher CBI scores postintervention compared to those in the intervention group (t_{23}=5.289, *P*=.01), whereas no significant differences were observed between the groups for females (t_{28}=0.427, *P*=.59). In the age subgroup analysis, caregivers under 60 years in the control group exhibited higher CBI scores (t_{25}=4.088, *P*=.02), with no differences found among those aged 60 years and older (t_{26}=0.363, *P*=.64). The careers subgroup analysis revealed that participants engaged in physical labor had higher scores in the control group (t_{38}=4.249, *P*=.02), while no differences were noted for those in mental labor (t_{13}=0.208, *P*=.72). Regarding the educational level subgroup analysis, caregivers with primary education or below in the control group showed significantly higher CBI scores post-intervention (t_{13}=4.494, *P*=.04), whereas no significant differences were observed among those with junior high school education and above (t_{38}=1.273, *P*=.16). In the monthly family income subgroup, caregivers with incomes between 1000 and 3000 RMB (approximately US $149-446) showed a marginally significant trend toward higher scores in the control group (t_9=6.422, *P*=.06), and no statistical differences were observed in the higher income groups.

XSL•FO
RenderX

Table . Subgroup analysis of intervention effects on CBI[a] by caregiver characteristics in a randomized controlled trial of mHealth intervention for caregivers of patients with dementia and type 2 diabetes in Xiamen, China, 2022 - 2023.

Variable, categories, and time	Mean difference (SE; 95% CI)	*t* value	*P* value
Gender			
Male			
Preintervention	−1.99 (2.834; −7.68 to 3.70)	−1.139 (23)	.49
Postintervention	5.04 (1.889; 1.25 to 8.83)	5.289 (23)	.01
Female			
Preintervention	−0.80 (2.585; −5.99 to 4.39)	−0.252 (28)	.76
Postintervention	0.93 (1.723; −2.53 to 4.39)	0.427 (28)	.59
Age (y)			
<60			
Preintervention	−2.93 (2.751; −8.46 to 2.59)	−1.676 (25)	.29
Postintervention	4.32 (1.856; 0.59 to 8.04)	4.088 (25)	.02
≥60			
Preintervention	−0.73 (2.713; −6.18 to 4.72)	−0.214 (26)	.78
Postintervention	0.85 (1.830; −2.82 to 4.53)	0.363 (26)	.64
Careers			
Physical labor			
Preintervention	−2.36 (2.266; −6.91 to 2.19)	−1.543 (38)	.30
Postintervention	3.60 (1.516; 0.56 to 6.64)	4.249 (38)	.02
Mental labor			
Preintervention	1.57 (3.704; −5.87 to 9.01)	0.264 (13)	.67
Postintervention	0.89 (2.478; −4.08 to 5.87)	0.208 (13)	.72
Educational level			
Primary and below			
Preintervention	-0.36 (3.920; -11.47 to 4.27)	−1.516 (13)	.36
Postintervention	5.50 (2.646; 0.19 to 10.81)	4.494 (13)	.04
Junior high school and above			
Preintervention	-0.03 (2.274; -4.59 to 4.54)	−0.010 (38)	.99
Postintervention	2.20 (1.536; -0.88 to 5.29)	1.273 (38)	.16
Monthly family income (RMB)[b]			
1000 - 3000			
Preintervention	−3.18 (4.559; 12.34 to 5.98)	−1.209 (9)	.49
Postintervention	6.00 (3.048; 0.163 to 12.13)	6.422 (9)	.06
3000 - 5000			
Preintervention	−1.56 (3.342; −8.27 to 5.16)	−0.700 (17)	.64
Postintervention	3.40 (2.234; −1.09 to 7.89)	2.365 (17)	.13
>5000			
Preintervention	−0.37 (2.912; −6.22 to 5.49)	−0.097 (23)	.90
Postintervention	1.12 (1.947; −2.79 to 5.03)	0.431 (23)	.57

[a]CBI: Caregiver Burden Inventory.

[b]All income values are in RMB. Based on the 2022 exchange rate (1 US $≈6.731 RMB), 1,000–3,000 RMB≈US $149–446, 3,000–5,000 RMB≈US $446–743, and >5,000 RMB≈>US $743.

XSL•FO
RenderX

Adverse Events

No serious adverse events or intervention-related harms were reported among caregivers in either group during the trial. All participant withdrawals (2 in the intervention group and 3 in the control group) were attributable to the death of their care recipients. No predefined adverse outcomes or deaths occurred among the caregiver participants.

Discussion

Principal Results

This study evaluated an mHealth supportive care program based on Fitch's theoretical framework for community-recruited caregivers of people with dementia and diabetes. Screening, recruitment, and participation outcomes indicated that although the number of participants was limited, the vast majority were willing to participate and complete the program. Over the 12-week intervention period, under the supervision and guidance of community health workers, the program exerted positive effects on FCGs of individuals with dementia and diabetes. Specifically, the intervention group demonstrated a statistically significant and substantial reduction in caregiver burden, along with considerable improvements in dementia care knowledge and perceived social support. These robust findings align with existing literature [83,84], supporting the efficacy of structured digital health interventions in alleviating caregiver burden while enhancing caregiving knowledge and perceived support.

Comparison With Prior Work

The observed intervention effects can be attributed to the program's systematic approach to meeting the multidimensional needs of FCGs. The significant reduction in caregiver burden is consistent with the Stress Process Model [85], which suggests that interventions enhancing caregivers' coping resources can alleviate their burden. The platform facilitated health care professionals' delivery of accurate, comprehensible information to FCGs, potentially increasing caregivers' confidence in dementia management, which has a positive effect on reducing caregiver burden [86]. Furthermore, the emotional, social, and spiritual needs module also encouraged caregivers to seek help from family, friends, and health care professionals, activating social support networks and thereby alleviating stress. This buffering effect aligns with the social support model [87], emphasizing the role of social networks in reducing caregiving stress. It is noteworthy that, despite a clinically meaningful reduction (approximately 15%), postintervention burden scores remained above the high-threshold level (36). This underscores the persistent, severe strain inherent in dementia caregiving [88] and highlights that while mHealth support can provide significant relief, it may need to be part of longer-term, adaptive, or multicomponent strategies to address profound, chronic burden fully.

The significant improvement in social support scores within the intervention group indicates the program's efficacy in enhancing caregivers' perception and utilization of social support, aligning with the findings of Tremont et al [89], which underscores the importance of strengthening social support for caregivers. The mechanism of action for this change is consistent with the social support buffer model [87]. On the one hand, the design of the social demand module not only facilitates communication between FCGs and health care workers to reduce social isolation [90] but also provides practical advice for FCGs to cope with stressors and improve their social support capacity. On the other hand, the online platform provided FCGs facing stigmatization, social isolation, or resource deprivation with key alternative pathways for emotional recognition and practical advice, as evidenced in other stigmatized health domains [91-94]. This dual-support model proved effective in alleviating negative emotions such as loneliness and helplessness [95,96], thereby improving caregivers' quality of life [97,98].

Acquiring relevant knowledge is a prerequisite and foundation for improving skill levels [99]. The median dementia care knowledge score in the intervention group increased significantly from 11 to 19, with a significant between-group difference also observed. This was reflected in caregivers' enhanced ability to recognize early complications more accurately, perform more standardized medication management (eg, avoiding incorrect and missed doses), and use nonpharmacological intervention techniques to address common behavioral symptoms. This design aligns with the "goal-oriented" principle of adult learning by providing learners with a clear modular knowledge structure. This improvement (approximately 73%) far exceeded the 20% to 30% knowledge gain achieved by home care workers through online training reported in the studies by Yeh et al [100,101], demonstrating the effectiveness of the mHealth supportive care program designed based on the supportive care framework in this study. This improvement can be attributed to the combination of online learning modules, hands-on demonstration videos, and offline follow-up coaching, which together reinforced caregivers' competence and confidence in managing dementia care [102,103].

The differential intervention effects on caregiver burden across demographic subgroups are a critical finding. Greater burden reduction among male, younger (<60 y), manual labor, caregivers with primary education or below, and lower-income caregivers suggests that the mHealth program's value is particularly pronounced for those who may face barriers to traditional support. This aligns with literature noting higher technology acceptance among males [104,105] and the greater reliance on accessible, free resources among economically disadvantaged groups [106-109], and the finding that individuals with lower educational attainment, who often have limited access to diverse health information sources [110,111], may rely more on structured digital tools that lower the barrier to information acquisition [112]. Conversely, the attenuated effects observed among female caregivers may stem from a misalignment between the intervention's primarily informational and instrumental support and their often more pronounced needs for emotional coping strategies and skills to manage role conflict and intra-family dynamics [113-115]. The text-based, asynchronous nature of the mHealth platform might have been less effective in addressing these immediate emotional and relational stressors. For older caregivers (≥60 y), lower digital literacy and potential discomfort with mobile technology [116,117] could have hindered full engagement with the digital

XSL•FO
RenderX

modules, thereby diluting the intervention's potential impact [118]. Among knowledge workers and caregivers with higher educational attainment (junior high school and above), their preexisting proficiency in information-seeking and likely higher access to alternative educational resources may have diminished the marginal benefit provided by our structured program. Similarly, higher-income caregivers likely have established access to private care services or consultations [119], making a supplementary public mHealth program less critical for burden alleviation. Importantly, the consistent improvements in social support and knowledge across all subgroups confirm that the program's core educational and connective functions delivered universal value, irrespective of these moderating factors for burden reduction.

Limitations

This study has several limitations that warrant consideration. First, although the study achieved statistical power, the geographical restriction to a single urban-rural cluster in Xiamen (5 neighborhood committees and 10 villages) may limit the generalizability of the findings. Additionally, the relatively small sample size could further constrain the population representativeness. Second, this study did not evaluate the cost-effectiveness of the mHealth supportive care program, which is critical for real-world implementation. Third, the COVID-19 pandemic disrupted in-person follow-ups, necessitating alternative methods (eg, phone interviews) that may introduce bias. Then, the DCKS used in this study demonstrated a Cronbach α of 0.626, which, while acceptable for a pilot study, suggests that future large-scale studies should consider using more established scales or further refining this instrument. The scale's reliability may be influenced by item quantity, cultural adaptation factors, and participant heterogeneity. Finally, although the 12-week intervention duration is consistent with comparable studies [120-124], the intervention period was insufficient to assess long-term effects on patient quality of life or program sustainability.

Conclusions

This study examined the efficacy of a supportive care framework-based mHealth intervention for FCGs of individuals with dementia and diabetes. Its primary innovation lies in its theory-driven design that addresses the complex, multidimensional needs of this comorbid caregiving population, distinguishing it from interventions targeting single conditions. The program significantly improved caregiver burden, social support, and dementia care knowledge, contributing novel evidence to a sparsely researched area. Importantly, subgroup analyses revealed differential effectiveness across demographic groups, suggesting that mHealth interventions can be tailored to enhance equity in access to support resources. In practical terms, by offering convenient, low-cost services, integrating health care resources, and enhancing personalized education, the program demonstrates the substantial potential of mHealth technology to improve caregiving outcomes. Furthermore, integrating structured mHealth support with existing community health resources presents a viable and scalable strategy to alleviate caregiver burden in resource-constrained settings. This approach not only helps FCGs cope with practical challenges but also provides a new perspective for promoting health equity at the grassroots level. Future research and implementation should incorporate adaptive or tailored features to meet diverse caregiver needs and advance health equity. Subsequent research should prioritize multicenter trials with longer follow-up, formal cost-effectiveness analyses, and the development of strategies to personalize digital supportive care.

Acknowledgments

We thank all the caregivers of people with dementia combined with diabetes and community health workers who participated in this study.

Funding

This study was supported by the Science and Technology Innovation Joint Fund Project of Fujian Province (grant 2020Y9021).

Data Availability

The deidentified data, trial protocol, and statistical analysis plan that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: HL (lead), MH (supporting)
Data curation: MH (lead), TG (supporting)
Formal analysis: XL (lead), MH (supporting)
Funding acquisition: HL
Investigation: MH (lead), TG (supporting)
Methodology: MH (lead), HL (supporting), RL (supporting), CH (supporting), YY (supporting)
Project administration: HL (lead), RL (supporting), CH (supporting), YY (supporting)
Supervision: HL (lead), RL (supporting), CH (supporting), YY (supporting)
Validation: XL (lead), TG (supporting), YY (supporting)
Visualization: XL

XSL•FO
RenderX

Writing – Original Draft: XL (lead), MH (supporting)
Writing – Review & Editing: XL (lead), TG (supporting), YY (supporting), HL (supporting)
YY and HL are co-corresponding authors (emails: yuanjiao1994@126.com and leehong99@126.com, respectively), with HL as the senior corresponding author.

Conflicts of Interest

None declared.

Editorial Notice

This randomized study was only retrospectively registered because of substantial delays in the institutional ethics review process during the COVID-19 pandemic, which prevented completion of registration prior to participant enrollment. The study commenced promptly upon receiving expedited ethics approval to address the urgent support needs of caregivers in locked-down communities. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials because the risk of bias appears low and because the primary purpose was to examine the intervention's effect on the caregivers (as providers) rather than on the patients. The majority of outcomes—social support and dementia care knowledge—directly assess the providers' own psychosocial state and skill acquisition and are not patient-level health outcomes. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
The process and content of the mHealth supportive care intervention program. mHealth: mobile health.
[DOCX File, 25 KB - mhealth_v14i1e72012_app1.docx]

Multimedia Appendix 2
Results of repeated-measures ANOVA for different demographic variables on outcome indicators.
[DOCX File, 24 KB - mhealth_v14i1e72012_app2.docx]

Checklist 1
CONSORT 2025 checklist.
[DOCX File, 31 KB - mhealth_v14i1e72012_app3.docx]

Checklist 2
CONSORT-EHEALTH checklist (V1.6.1).
[PDF File, 18221 KB - mhealth_v14i1e72012_app4.pdf]

Checklist 3
The TIDieR checklist.
[DOCX File, 29 KB - mhealth_v14i1e72012_app5.docx]

References

1. Wang C, Song P, Niu Y. The management of dementia worldwide: a review on policy practices, clinical guidelines, end-of-life care, and challenge along with aging population. Biosci Trends 2022 May 17;16(2):119-129. [doi: 10.5582/bst.2022.01042] [Medline: 35466154]
2. Jia L, Du Y, Chu L, et al. Prevalence, risk factors, and management of dementia and mild cognitive impairment in adults aged 60 years or older in China: a cross-sectional study. Lancet Public Health 2020 Dec;5(12):e661-e671. [doi: 10.1016/S2468-2667(20)30185-7] [Medline: 33271079]
3. Liao X, Huang Y, Zhang Z, et al. Factors associated with health-related quality of life among family caregivers of people with Alzheimer's disease. Psychogeriatrics 2020 Jul;20(4):398-405. [doi: 10.1111/psyg.12528] [Medline: 32027776]
4. Tabrizi LB, Navab E, Afshar PF, Noghabi AAA, Haghani H. Effect of cognitive-behavioral intervention on burden of family caregivers of patients with Alzheimer's disease. Hayat J Nurs Midwifery Sci ;21(1):94-102 [FREE Full text]
5. Chen S, Kuper H. Tracing the temporal trends of modifiable risk factors in dementia: insights from the English Longitudinal Study of Ageing (2004-2019). Lancet 2023 Nov;402 Suppl 1:S34. [doi: 10.1016/S0140-6736(23)02078-0] [Medline: 37997075]
6. Park JC, Lim H, Byun MS, et al. Sex differences in the progression of glucose metabolism dysfunction in Alzheimer's disease. Exp Mol Med 2023 May;55(5):1023-1032. [doi: 10.1038/s12276-023-00993-3] [Medline: 37121979]
7. Shang Y, Fratiglioni L, Vetrano DL, Dove A, Welmer AK, Xu W. Not only diabetes but also prediabetes leads to functional decline and disability in older adults. Diabetes Care 2021 Mar;44(3):690-698. [doi: 10.2337/dc20-2232] [Medline: 33446522]

XSL•FO
RenderX

8. Al-Khaledi M, Al-Dousari H, Al-Dhufairi S, et al. Diabetes self-management: a key to better health-related quality of life in patients with diabetes. Med Princ Pract 2018;27(4):323-331. [doi: 10.1159/000489310] [Medline: 29669330]
9. John R, Pise S, Chaudhari L, Deshpande PR. Evaluation of quality of life in type 2 diabetes mellitus patients using quality of life instrument for Indian diabetic patients: a cross-sectional study. J Midlife Health 2019;10(2):81-88. [doi: 10.4103/jmh.JMH_32_18] [Medline: 31391757]
10. Sjöholm Å, Bennet L, Nilsson PM. Cognitive dysfunction in diabetes - the "forgotten" diabetes complication: a narrative review. Scand J Prim Health Care 2025 Jun;43(2):448-454. [doi: 10.1080/02813432.2025.2455136] [Medline: 39876043]
11. IDF Diabetes Atlas. URL: https://diabetesatlas.org/ [accessed 2025-07-23]
12. Koekkoek PS, Kappelle LJ, van den Berg E, Rutten GEHM, Biessels GJ. Cognitive function in patients with diabetes mellitus: guidance for daily care. Lancet Neurol 2015 Mar;14(3):329-340. [doi: 10.1016/S1474-4422(14)70249-2] [Medline: 25728442]
13. Kautzky-Willer A, Harreiter J, Pacini G. Sex and gender differences in risk, pathophysiology and complications of type 2 diabetes mellitus. Endocr Rev 2016 Jun;37(3):278-316. [doi: 10.1210/er.2015-1137] [Medline: 27159875]
14. Diniz Pereira J, Gomes Fraga V, Morais Santos AL, Carvalho MDG, Caramelli P, Braga Gomes K. Alzheimer's disease and type 2 diabetes mellitus: a systematic review of proteomic studies. J Neurochem 2021 Mar;156(6):753-776. [doi: 10.1111/jnc.15166] [Medline: 32909269]
15. Alnaim L, Altuwaym RA, Aldehan SM, Alquraishi NM. Assessment of knowledge among caregivers of diabetic patients in insulin dosage regimen and administration. Saudi Pharm J 2021 Oct;29(10):1137-1142. [doi: 10.1016/j.jsps.2021.08.010] [Medline: 34703367]
16. Alaofè H, Amoussa Hounkpatin W, Djrolo F, Ehiri J, Rosales C. Factors associated with quality of life in patients with type 2 diabetes of South Benin: a cross-sectional study. Int J Environ Res Public Health 2022 Feb 18;19(4):2360. [doi: 10.3390/ijerph19042360] [Medline: 35206551]
17. Gaffari-Fam S, Lotfi Y, Daemi A, et al. Impact of health literacy and self-care behaviors on health-related quality of life in Iranians with type 2 diabetes: a cross-sectional study. Health Qual Life Outcomes 2020 Nov 4;18(1):357. [doi: 10.1186/s12955-020-01613-8] [Medline: 33148266]
18. Barua L, Faruque M, Chowdhury HA, Banik PC, Ali L. Health-related quality of life and its predictors among the type 2 diabetes population of Bangladesh: a nation-wide cross-sectional study. J Diabetes Investig 2021 Feb;12(2):277-285. [doi: 10.1111/jdi.13331] [Medline: 32564501]
19. Zan H, Meng Z, Li J, Zhang X, Liu T. Factors associated with quality of life among elderly patients with type 2 diabetes mellitus: the role of family caregivers. BMC Public Health 2024 Feb 21;24(1):539. [doi: 10.1186/s12889-024-17917-z] [Medline: 38383369]
20. Thongduang K, Boonchieng W, Chautrakarn S, Ong-Artborirak P. The influence of family caregiver knowledge and behavior on elderly diabetic patients' quality of life in Northern Thailand. Int J Environ Res Public Health 2022 Aug 17;19(16):10216. [doi: 10.3390/ijerph191610216] [Medline: 36011847]
21. Bailes CO, Kelley CM, Parker NM. Caregiver burden and perceived health competence when caring for family members diagnosed with Alzheimer's disease and related dementia. J Am Assoc Nurse Pract 2016 Oct;28(10):534-540. [doi: 10.1002/2327-6924.12355] [Medline: 27073005]
22. Amtmann D, Liljenquist KS, Bamer A, et al. Development and validation of the University of Washington caregiver stress and benefit scales for caregivers of children with or without serious health conditions. Qual Life Res 2020 May;29(5):1361-1371. [doi: 10.1007/s11136-019-02409-0] [Medline: 31902052]
23. Rodakowski J, Skidmore ER, Rogers JC, Schulz R. Role of social support in predicting caregiver burden. Arch Phys Med Rehabil 2012 Dec;93(12):2229-2236. [doi: 10.1016/j.apmr.2012.07.004] [Medline: 22824248]
24. Möller-Leimkühler AM, Wiesheu A. Caregiver burden in chronic mental illness: the role of patient and caregiver characteristics. Eur Arch Psychiatry Clin Neurosci 2012 Mar;262(2):157-166. [doi: 10.1007/s00406-011-0215-5] [Medline: 21538092]
25. Dickson A, O'Brien G, Ward R, Flowers P, Allan D, O'Carroll R. Adjustment and coping in spousal caregivers following a traumatic spinal cord injury: an interpretative phenomenological analysis. J Health Psychol 2012 Mar;17(2):247-257. [doi: 10.1177/1359105311411115] [Medline: 21708870]
26. Grant M, Sun V, Fujinami R, et al. Family caregiver burden, skills preparedness, and quality of life in non-small cell lung cancer. Oncol Nurs Forum 2013 Jul;40(4):337-346. [doi: 10.1188/13.ONF.337-346] [Medline: 23803267]
27. Yazdanmanesh M, Esmaeili R, Nasiri M, Vasli P. Relieving care burden and promoting health-related quality of life for family caregivers of elderly people with Alzheimer's disease via an empowerment program. Aging Clin Exp Res 2023 Jan;35(1):73-83. [doi: 10.1007/s40520-022-02277-8] [Medline: 36282451]
28. Sadat Hejazi S, Hosseini M, Ebadi A, Alavi Majd H. Family caregiver: the forgotten savior. Int J Community Based Nurs Midwifery 2021 Oct;9(4):346-347. [doi: 10.30476/IJCBNM.2021.90118.1673] [Medline: 34604403]
29. Gaugler JE, Yu F, Krichbaum K, Wyman JF. Predictors of nursing home admission for persons with dementia. Med Care 2009 Feb;47(2):191-198. [doi: 10.1097/MLR.0b013e31818457ce] [Medline: 19169120]
30. Schulz R, Beach SR. Caregiving as a risk factor for mortality: the Caregiver Health Effects Study. JAMA 1999 Dec 15;282(23):2215-2219. [doi: 10.1001/jama.282.23.2215] [Medline: 10605972]

XSL•FO
RenderX

31. Jafari H, Ebrahimi A, Aghaei A, Khatony A. The relationship between care burden and quality of life in caregivers of hemodialysis patients. BMC Nephrol 2018 Nov 12;19(1):321. [doi: 10.1186/s12882-018-1120-1] [Medline: 30419837]
32. Jain FA, Chernyak SV, Nickerson LD, et al. Four-week mentalizing imagery therapy for family dementia caregivers: a randomized controlled trial with neural circuit changes. Psychother Psychosom 2022;91(3):180-189. [doi: 10.1159/000521950] [Medline: 35287133]
33. Liu S, Li C, Shi Z, et al. Caregiver burden and prevalence of depression, anxiety and sleep disturbances in Alzheimer's disease caregivers in China. J Clin Nurs 2017 May;26(9-10):1291-1300. [doi: 10.1111/jocn.13601] [Medline: 27681477]
34. Beinart N, Weinman J, Wade D, Brady R. Caregiver burden and psychoeducational interventions in Alzheimer's disease: a review. Dement Geriatr Cogn Dis Extra 2012 Jan;2(1):638-648. [doi: 10.1159/000345777] [Medline: 23341829]
35. Chan EY, Glass G, Chua KC, Ali N, Lim WS. Relationship between mastery and caregiving competence in protecting against burden, anxiety and depression among caregivers of frail older adults. J Nutr Health Aging 2018;22(10):1238-1245. [doi: 10.1007/s12603-018-1098-1] [Medline: 30498832]
36. Chandler MJ, Locke DE, Crook JE, et al. Comparative effectiveness of behavioral interventions on quality of life for older adults with mild cognitive impairment: a randomized clinical trial. JAMA Netw Open 2019 May 3;2(5):e193016. [doi: 10.1001/jamanetworkopen.2019.3016] [Medline: 31099860]
37. Amofa PA, Locke DEC, Chandler M, et al. Comparative effectiveness of behavioral interventions to prevent or delay dementia: one-year partner outcomes. J Prev Alzheimers Dis 2021;8(1):33-40. [doi: 10.14283/jpad.2020.59] [Medline: 33336222]
38. Steiner V, Pierce LL, Salvador D. Information needs of family caregivers of people with dementia. Rehabil Nurs 2016;41(3):162-169. [doi: 10.1002/rnj.214] [Medline: 25858031]
39. Werner NE, Stanislawski B, Marx KA, et al. Getting what they need when they need it. Identifying barriers to information needs of family caregivers to manage dementia-related behavioral symptoms. Appl Clin Inform 2017 Feb 22;8(1):191-205. [doi: 10.4338/ACI-2016-07-RA-0122] [Medline: 28224163]
40. Whitlatch CJ, Orsulic-Jeras S. Meeting the informational, educational, and psychosocial support needs of persons living with dementia and their family caregivers. Gerontologist 2018 Jan 18;58(suppl_1):S58-S73. [doi: 10.1093/geront/gnx162] [Medline: 29361068]
41. Christie HL, Bartels SL, Boots LMM, Tange HJ, Verhey FJJ, de Vugt ME. A systematic review on the implementation of eHealth interventions for informal caregivers of people with dementia. Internet Interv 2018 Sep;13:51-59. [doi: 10.1016/j.invent.2018.07.002] [Medline: 30206519]
42. Egan KJ, Pinto-Bruno Á, Bighelli I, et al. Online training and support programs designed to improve mental health and reduce burden among caregivers of people with dementia: a systematic review. J Am Med Dir Assoc 2018 Mar;19(3):200-206. [doi: 10.1016/j.jamda.2017.10.023] [Medline: 29306605]
43. Bressan V, Visintini C, Palese A. What do family caregivers of people with dementia need? A mixed-method systematic review. Health Soc Care Community 2020 Nov;28(6):1942-1960. [doi: 10.1111/hsc.13048] [Medline: 32542963]
44. Ringer TJ, Wong-Pack M, Miller P, et al. Understanding the educational and support needs of informal care-givers of people with dementia attending an outpatient geriatric assessment clinic. Ageing Soc 2020 Jan;40(1):205-228. [doi: 10.1017/S0144686X18000971]
45. Jackson D, Roberts G, Wu ML, Ford R, Doyle C. A systematic review of the effect of telephone, internet or combined support for carers of people living with Alzheimer's, vascular or mixed dementia in the community. Arch Gerontol Geriatr 2016;66:218-236. [doi: 10.1016/j.archger.2016.06.013] [Medline: 27372903]
46. Leng M, Zhao Y, Xiao H, Li C, Wang Z. Internet-based supportive interventions for family caregivers of people with dementia: systematic review and meta-analysis. J Med Internet Res 2020 Sep 9;22(9):e19468. [doi: 10.2196/19468] [Medline: 32902388]
47. Teahan Á, Lafferty A, McAuliffe E, et al. Psychosocial interventions for family carers of people with dementia: a systematic review and meta-analysis. J Aging Health 2020 Oct;32(9):1198-1213. [doi: 10.1177/0898264319899793] [Medline: 31957536]
48. Zhao Y, Feng H, Hu M, et al. Web-based interventions to improve mental health in home caregivers of people with dementia: meta-analysis. J Med Internet Res 2019 May 6;21(5):e13415. [doi: 10.2196/13415] [Medline: 31066680]
49. Etxeberria I, Salaberria K, Gorostiaga A. Online support for family caregivers of people with dementia: a systematic review and meta-analysis of RCTs and quasi-experimental studies. Aging Ment Health 2021 Jul;25(7):1165-1180. [doi: 10.1080/13607863.2020.1758900] [Medline: 32363901]
50. Deeken F, Rezo A, Hinz M, Discher R, Rapp MA. Evaluation of technology-based interventions for informal caregivers of patients with dementia-a meta-analysis of randomized controlled trials. Am J Geriatr Psychiatry 2019 Apr;27(4):426-445. [doi: 10.1016/j.jagp.2018.12.003] [Medline: 30642650]
51. Integrated care for older people: guidelines on community-level interventions to manage declines in intrinsic capacity. : World Health Organization; 2017 URL: https://iris.who.int/server/api/core/bitstreams/4639a0b9-f77e-48db-9617-987aabb143f2/content [accessed 2026-02-25]
52. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3. [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]

XSL•FO
RenderX

53. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 1;3(1):e7. [doi: 10.2196/mental.4984] [Medline: 26932350]
54. Windle G, Flynn G, Hoare Z, et al. Effects of an e-health intervention "iSupport" for reducing distress of dementia carers: protocol for a randomised controlled trial and feasibility study. BMJ Open 2022 Sep 21;12(9):e064314. [doi: 10.1136/bmjopen-2022-064314] [Medline: 36130751]
55. Meichsner F, Theurer C, Wilz G. Acceptance and treatment effects of an internet-delivered cognitive-behavioral intervention for family caregivers of people with dementia: a randomized-controlled trial. J Clin Psychol 2019 Apr;75(4):594-613. [doi: 10.1002/jclp.22739] [Medline: 30597537]
56. Baruah U, Varghese M, Loganathan S, et al. Feasibility and preliminary effectiveness of an online training and support program for caregivers of people with dementia in India: a randomized controlled trial. Int J Geriatr Psychiatry 2021 Apr;36(4):606-617. [doi: 10.1002/gps.5502] [Medline: 33491811]
57. Pritchard E, Cussen A, Delafosse V, Swift M, Jolliffe L, Yeates H. Interventions supporting caregiver readiness when caring for patients with dementia following discharge home: a mixed-methods systematic review. Australas J Ageing 2020 Sep;39(3):e239-e250. [doi: 10.1111/ajag.12765] [Medline: 31944506]
58. Moore KJ, Lee CY, Sampson EL, Candy B. Do interventions that include education on dementia progression improve knowledge, mental health and burden of family carers? A systematic review. Dementia (London) 2020 Nov;19(8):2555-2581. [doi: 10.1177/1471301219831530] [Medline: 30786759]
59. Zupa MF, Lee A, Piette JD, et al. Impact of a dyadic intervention on family supporter involvement in helping adults manage type 2 diabetes. J Gen Intern Med 2022 Mar;37(4):761-768. [doi: 10.1007/s11606-021-06946-8] [Medline: 34240285]
60. Peiris D, Sun L, Patel A, et al. Systematic medical assessment, referral and treatment for diabetes care in China using lay family health promoters: protocol for the SMARTDiabetes cluster randomised controlled trial. Implement Sci 2016 Aug 17;11(1):116. [doi: 10.1186/s13012-016-0481-8] [Medline: 27535128]
61. Sztramko R, Levinson AJ, Wurster AE, et al. Online educational tools for caregivers of people with dementia: a scoping literature review. Can Geriatr J 2021 Dec;24(4):351-366. [doi: 10.5770/cgj.24.506] [Medline: 34912490]
62. Scerbe A, O'Connell ME, Astell A, et al. Digital tools for delivery of dementia education for caregivers of persons with dementia: a systematic review and meta-analysis of impact on caregiver distress and depressive symptoms. PLoS One 2023;18(5):e0283600. [doi: 10.1371/journal.pone.0283600] [Medline: 37196022]
63. Zhu EM, Buljac-Samardžić M, Ahaus K, Sevdalis N, Huijsman R. Implementation and dissemination of home- and community-based interventions for informal caregivers of people living with dementia: a systematic scoping review. Implement Sci 2023 Nov 8;18(1):60. [doi: 10.1186/s13012-023-01314-y] [Medline: 37940960]
64. Drummond N, McCleary L, Garcia L, et al. Assessing determinants of perceived quality in transitions for people with dementia: a prospective observational study. Can Geriatr J 2019 Mar;22(1):13-22. [doi: 10.5770/cgj.22.332] [Medline: 31501679]
65. Spigelmyer PC, Schreiber JB. A pilot study: resistive behavior in the context of informal caregiver-assisted activities of daily living. Geriatr Nurs 2019;40(4):399-404. [doi: 10.1016/j.gerinurse.2019.01.005] [Medline: 30878279]
66. Huis In Het Veld JG, van Asch IFM, Willemse BM, et al. Process evaluation of nurse-led online self-management support for family caregivers to deal with behavior changes of a relative with dementia (Part 1): mixed methods study. J Med Internet Res 2019 Oct 11;21(10):e13002. [doi: 10.2196/13002] [Medline: 31605517]
67. Baharudin AD, Din NC, Subramaniam P, Razali R. The associations between behavioral-psychological symptoms of dementia (BPSD) and coping strategy, burden of care and personality style among low-income caregivers of patients with dementia. BMC Public Health 2019 Jun 13;19(Suppl 4):447. [doi: 10.1186/s12889-019-6868-0] [Medline: 31196141]
68. Zwingmann I, Hoffmann W, Michalowsky B, et al. Supporting family dementia caregivers: testing the efficacy of dementia care management on multifaceted caregivers' burden. Aging Ment Health 2018 Jul;22(7):889-896. [doi: 10.1080/13607863.2017.1399341] [Medline: 29156941]
69. Fitch MI. Supportive care framework. Can Oncol Nurs J 2008;18(1):6-24. [doi: 10.5737/1181912x181614] [Medline: 18512565]
70. Hopewell S, Chan AW, Collins GS, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. BMJ 2025 Apr 14;389:e081123. [doi: 10.1136/bmj-2024-081123] [Medline: 40228833]
71. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
72. Wei WJ. Construction and empirical study of an intervention program for care skills of senile dementia caregivers [Master's thesis in Chinese]. : Zhengzhou University; 2019 URL: https://d-wanfangdata-com-cn-s-119.lib1-fx.fjmu.edu.cn/thesis/Ch1UaGVzaXNOZXdTb2xyOVMyMDI2MDExNzA4NTkxNhIIWTM1NzAwNjIaCDQ4cXpmMXVi [accessed 2025-06-30]
73. Chou KR, Jiann-Chyun L, Chu H. The reliability and validity of the Chinese version of the caregiver burden inventory. Nurs Res 2002;51(5):324-331. [doi: 10.1097/00006199-200209000-00009] [Medline: 12352781]
74. Gao FF, Shang SM. Study on the changes of quality of life and its influencing factors during one year among dementia patients [Article in Chinese]. J Nurs Adm 2012;12(3):153-155. [doi: 10.3969/j.issn.1671-315X.2012.03.001]

XSL•FO
RenderX

75. Zhang R. A study on the caregiving experience of caregivers for patients with senile dementia [Master's thesis in Chinese]. : Peking Union Medical College; 2007 URL: https://d-wanfangdata-com-cn-s-119.lib1-fx.fjmu.edu.cn/thesis/Ch1UaGVzaXNOZXdTb2xyOVMyMDI2MDExNzA4NTkxNhIIWTExMjcxMTAaCG91ejY3Ynhs [accessed 2025-06-30]
76. Giulietti MV, Fabbietti P, Spatuzzi R, Vespa A. Effects of mindfulness based interventions in adults and older adults caregivers of patients with early stage Alzheimer's disease: a randomized pilot study. J Alzheimers Dis 2024;97(4):1923-1930. [doi: 10.3233/JAD-230284] [Medline: 38306028]
77. Biliunaite I, Kazlauskas E, Sanderman R, Truskauskaite-Kuneviciene I, Dumarkaite A, Andersson G. Internet-based cognitive behavioral therapy for informal caregivers: randomized controlled pilot trial. J Med Internet Res 2021 Apr 7;23(4):e21466. [doi: 10.2196/21466] [Medline: 33825687]
78. Xiao SY. Theoretical basis and research application of social support rating scale [Article in Chinese]. J Clin Psychiatry 1994;4:98-100 [FREE Full text]
79. Wang Y, Tian S, Yi J, Cai C, Zhu Y, Shi C. The impact of family burden on the psychological distress of informal caregivers for older adults living with dementia: the role of internal and external coping resources. Geriatr Nurs 2024 Nov;60:92-98. [doi: 10.1016/j.gerinurse.2024.08.017] [Medline: 39236371]
80. Yu H, Wang X, He R, Liang R, Zhou L. Measuring the caregiver burden of caring for community-residing people with Alzheimer's disease. PLoS One 2015;10(7):e0132168. [doi: 10.1371/journal.pone.0132168] [Medline: 26154626]
81. Wei YP. Investigation and intervention study on quality of life, social support, and knowledge-attitude-practice among caregivers of senile dementia patients in Fujian Province [Master's thesis in Chinese]. : Fujian Medical University; 2015 URL: http://dx.chinadoi.cn/10.7666/d.D738528 [accessed 2025-06-30] [doi: 10.7666/d.D738528]
82. Han HF, Zhao YL, Ge QJ, Sun CL, Ji YN, Wang YB. Study on the influence of knowledge, attitude, and practice of family caregivers on the quality of life of patients with senile dementia. World Latest Med Inf 2018;18(76):235-236. [doi: 10.19613/j.cnki.1671-3141.2018.76.131]
83. Xie Y, Shen S, Liu C, et al. Internet-based supportive interventions for family caregivers of people with dementia: randomized controlled trial. JMIR Aging 2024 Oct 4;7:e50847. [doi: 10.2196/50847] [Medline: 39365914]
84. Torkamani M, McDonald L, Saez Aguayo I, et al. A randomized controlled pilot study to evaluate a technology platform for the assisted living of people with dementia and their carers. J Alzheimers Dis 2014;41(2):515-523. [doi: 10.3233/JAD-132156] [Medline: 24643137]
85. Pearlin LI, Mullan JT, Semple SJ, Skaff MM. Caregiving and the stress process: an overview of concepts and their measures. Gerontologist 1990 Oct;30(5):583-594. [doi: 10.1093/geront/30.5.583] [Medline: 2276631]
86. Kaliappan S, Liu C, Jain Y, Karkar R, Saha K. Online communities as a support system for Alzheimer disease and dementia care: large-scale exploratory study. JMIR Aging 2025 May 5;8:e68890. [doi: 10.2196/68890] [Medline: 40324770]
87. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychol Bull 1985 Sep;98(2):310-357. [doi: 10.1037/0033-2909.98.2.310] [Medline: 3901065]
88. Hong YA, Shen K, Han HR, et al. Two-year follow-up of dementia caregivers after a digital health intervention WECARE: a mixed-method study. Aging Ment Health 2025 Apr;29(4):631-638. [doi: 10.1080/13607863.2024.2449141] [Medline: 39757693]
89. Tremont G, Davis JD, Ott BR, et al. Randomized trial of the family intervention: telephone tracking-caregiver for dementia caregivers: use of community and healthcare resources. J Am Geriatr Soc 2017 May;65(5):924-930. [doi: 10.1111/jgs.14684] [Medline: 28008609]
90. Saragih ID, Tonapa SI, Porta CM, Lee BO. Effects of telehealth intervention for people with dementia and their carers: a systematic review and meta-analysis of randomized controlled studies. J Nurs Scholarsh 2022 Nov;54(6):704-719. [doi: 10.1111/jnu.12797] [Medline: 35769007]
91. Taggart T, Grewe ME, Conserve DF, Gliwa C, Roman Isler M. Social media and HIV: a systematic review of uses of social media in HIV communication. J Med Internet Res 2015 Nov 2;17(11):e248. [doi: 10.2196/jmir.4387] [Medline: 26525289]
92. De Choudhury M, Kıcıman E. The language of social support in social media and its effect on suicidal ideation risk. Presented at: Proceedings of the Eleventh International AAAI Conference on Web and Social Media; May 15-18, 2017. [doi: 10.1609/icwsm.v11i1.14891]
93. Andalibi N, Haimson OL, De Choudhury M, Forte A. Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. Presented at: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI 2016); May 7-12, 2016. [doi: 10.1145/2858036.2858096]
94. Yang D, Yao Z, Seering J, Kraut R. The channel matters: self-disclosure, reciprocity and social support in online cancer support groups. Proc SIGCHI Conf Hum Factor Comput Syst 2019 May;2019:31. [doi: 10.1145/3290605.3300261] [Medline: 31448374]
95. Wang Q, Xiao X, Zhang J, et al. The experiences of East Asian dementia caregivers in filial culture: a systematic review and meta-analysis. Front Psychiatry 2023;14:1173755. [doi: 10.3389/fpsyt.2023.1173755] [Medline: 37151975]
96. Lin XY, Zhang L, Yoon S, Zhang R, Lachman ME. A social exergame intervention to promote physical activity, social support, and well-being in family caregivers. Gerontologist 2023 Oct 17;63(9):1456-1466. [doi: 10.1093/geront/gnad028] [Medline: 36916022]

XSL•FO
RenderX

97. Ganapathy SS, Sooryanarayana R, Ahmad NA, et al. Prevalence of dementia and quality of life of caregivers of people living with dementia in Malaysia. Geriatr Gerontol Int 2020 Dec;20 Suppl 2(S2):16-20. [doi: 10.1111/ggi.14031] [Medline: 33370858]
98. Nemcikova M, Katreniakova Z, Nagyova I. Social support, positive caregiving experience, and caregiver burden in informal caregivers of older adults with dementia. Front Public Health 2023;11:1104250. [doi: 10.3389/fpubh.2023.1104250] [Medline: 36761127]
99. Farran CJ, McCann JJ, Fogg LG, Etkin CD. Developing a measurement strategy for assessing family caregiver skills: conceptual issues. Alzheimers care today 2009;10(3):129-139. [doi: 10.1097/ACQ.0b013e3181b15d82] [Medline: 20179779]
100. Yeh J, Beld M, Pond B, et al. Competency-based training boosts dementia knowledge and skills in home care workers. Alzheimers Dement 2025 Jun;21(6):e70323. [doi: 10.1002/alz.70323] [Medline: 40538020]
101. Yeh J, Pond B, Beld M, et al. Enhancing dementia knowledge and self-efficacy of in-home supportive services caregivers through online training. J Appl Gerontol 2023 Apr;42(4):617-626. [doi: 10.1177/07334648221144023] [Medline: 36482678]
102. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037/0033-295X.84.2.191] [Medline: 847061]
103. Boyt N, Ho AK, Morris-Bankole H, Sin J. Internet-facilitated interventions for informal caregivers of patients with neurodegenerative disorders: Systematic review and meta-analysis. Digit Health 2022;8:20552076221129069. [doi: 10.1177/20552076221129069] [Medline: 36276187]
104. Huffman AH, Whetten J, Huffman WH. Using technology in higher education: the influence of gender roles on technology self-efficacy. Comput Human Behav 2013 Jul;29(4):1779-1786. [doi: 10.1016/j.chb.2013.02.012]
105. Xiong C, Ye B, Mihailidis A, et al. Sex and gender differences in technology needs and preferences among informal caregivers of persons with dementia. BMC Geriatr 2020 May 18;20(1):176. [doi: 10.1186/s12877-020-01548-1] [Medline: 32423406]
106. Bundorf MK, Wagner TH, Singer SJ, Baker LC. Who searches the internet for health information? Health Serv Res 2006 Jun;41(3 Pt 1):819-836. [doi: 10.1111/j.1475-6773.2006.00510.x] [Medline: 16704514]
107. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009 Feb 20;11(1):e4. [doi: 10.2196/jmir.1035] [Medline: 19275980]
108. Information and its outcomes interplay of negative emotion and health self-efficacy on the use of health. Semantic Scholar. URL: https://www.semanticscholar.org/paper/Information-and-Its-Outcomes-Interplay-of-Negative-Lee-Hwang/c05b4b0754375302a14ce6afeeb0c303338aa039 [accessed 2025-08-09]
109. Hassan AYI, Lamura G, Hagedoorn M. Predictors of digital support services use by informal caregivers: a cross-sectional comparative survey. BMJ Open 2022 Apr 29;12(4):e059897. [doi: 10.1136/bmjopen-2021-059897] [Medline: 35487716]
110. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011 Jul 19;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
111. Wang W, Zhang Y, Lin B, Mei Y, Ping Z, Zhang Z. The urban-rural disparity in the status and risk factors of health literacy: a cross-sectional survey in central China. Int J Environ Res Public Health 2020 May 29;17(11):3848. [doi: 10.3390/ijerph17113848] [Medline: 32485790]
112. Jacobs RJ, Lou JQ, Ownby RL, Caballero J. A systematic review of eHealth interventions to improve health literacy. Health Informatics J 2016 Jun;22(2):81-98. [doi: 10.1177/1460458214534092] [Medline: 24916567]
113. Zurlo MC, Vallone F, Smith AP. Work-family conflict and psychophysical health conditions of nurses: gender differences and moderating variables. Jpn J Nurs Sci 2020 Jul;17(3):e12324. [doi: 10.1111/jjns.12324] [Medline: 31930657]
114. Sharma N, Chakrabarti S, Grover S. Gender differences in caregiving among family - caregivers of people with mental illnesses. World J Psychiatry 2016 Mar 22;6(1):7-17. [doi: 10.5498/wjp.v6.i1.7] [Medline: 27014594]
115. de Graaff MLA, van der Heide I, Rademakers JJDJM, et al. Is caregiver burden associated with sex and gender-related characteristics? A large-scale survey study among family caregivers of people with dementia. BMC Geriatr 2025 Mar 14;25(1):171. [doi: 10.1186/s12877-025-05795-y] [Medline: 40087581]
116. Fisher E, Proctor D, Perkins L, Felstead C, Stott J, Spector A. Is virtual cognitive stimulation therapy the future for people with dementia? An audit of UK NHS memory clinics during the COVID-19 pandemic. J Technol Behav Sci 2023 Feb 24:1-8. [doi: 10.1007/s41347-023-00306-5] [Medline: 36855467]
117. Cuffaro L, Di Lorenzo F, Bonavita S, Tedeschi G, Leocani L, Lavorgna L. Dementia care and COVID-19 pandemic: a necessary digital revolution. Neurol Sci 2020 Aug;41(8):1977-1979. [doi: 10.1007/s10072-020-04512-4] [Medline: 32556746]
118. Cui S, Crowe-Cumella H, Fortuna KL, Jain FA. A state-of-the-art narrative review of peer support for family caregivers of people with dementia: from in-person to digital delivery. mHealth 2024;11:9. [doi: 10.21037/mhealth-24-19] [Medline: 39944854]
119. Hassan AYI, Cucculelli M, Lamura G. Caregivers' willingness to pay for digital support services: comparative survey. Health Policy 2023 Apr;130:104751. [doi: 10.1016/j.healthpol.2023.104751] [Medline: 36857837]

XSL•FO
RenderX

120. Su HF, Koo M, Lee WL, Sung HC, Lee RP, Liu WI. A dementia care training using mobile e-learning with mentoring support for home care workers: a controlled study. BMC Geriatr 2021 Feb 16;21(1):126. [doi: 10.1186/s12877-021-02075-3] [Medline: 33593287]
121. Williams K, Blyler D, Vidoni ED, et al. A randomized trial using telehealth technology to link caregivers with dementia care experts for in-home caregiving support: FamTechCare protocol. Res Nurs Health 2018 Jun;41(3):219-227. [doi: 10.1002/nur.21869] [Medline: 29504666]
122. Li FY, Xin HJ, Lou GB, Tang XF. Influence of continuing care based on WeChat on mental health status of the dementia caregivers [Article in Chinese]. Chin J Gen Pract 2017;15(6):1069-1071. [doi: 10.16766/j.cnki.issn.1674-4152.2017.06.047]
123. Carbone G, Barreca F, Mancini G, et al. A home assistance model for dementia: outcome in patients with mild-to-moderate Alzheimer's disease after three months. Ann Ist Super Sanita 2013;49(1):34-41. [doi: 10.4415/ANN_13_01_07] [Medline: 23535128]
124. Zang WP, Jiang CY. Study for the effects of community home visiting service on the positive aspects of care-giving among community-dwelling family of senile dementia patients [Article in Chinese]. J Nurses Train 2017;32(23):2115-2118. [doi: 10.16821/j.cnki.hsjx.2017.23.002]

Abbreviations

CBI: Caregiver Burden Inventory
CONSORT: Consolidated Standards of Reporting Trials
CONSORT-eHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
CVI: content validity index
DCKS: Dementia Caring Knowledge Scale
FCG: family caregiver
mHealth: mobile health
SSRS: Social Support Rating Scale
T2DM: type 2 diabetes mellitus
TIDieR: Template for Intervention Description and Replication

Edited by S Brini; submitted 01.Feb.2025; peer-reviewed by N Chaimongkol, WQ Chen; accepted 12.Feb.2026; published 24.Mar.2026.

Please cite as:
Lin X, Huang M, Guo T, Lin R, Huang C, Yan Y, Li H
Effects of the mHealth Supportive Care Program for Family Caregivers of Individuals With Dementia and Diabetes: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e72012
URL: https://mhealth.jmir.org/2026/1/e72012
doi:10.2196/72012

© Xinying Lin, Miaoqian Huang, Tuantuan Guo, Rong Lin, Chenshan Huang, Yuanjiao Yan, Hong Li. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 24.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effectiveness of a Telecare-Based Intervention Program in Supporting Informal Caregivers of Community-Dwelling Older Adults in Reducing Stress Levels: Randomized Controlled Trial

Arkers Kwan Ching Wong[1], PhD; Jonathan Bayuo[2], PhD; Nga Ping Ng[1], BSN; Matthew Yau[3], MSW; Ka Kit Simon Yu[3], MSW; Rose Sin Yi Lin[4], PhD; Jing Jing Su[5], PhD; Vivian Hui[1], PhD; Jed Montayre[1], PhD

[1]School of Nursing, Faculty of Health and Social Science, The Hong Kong Polytechnic University, Room 502, Block GH, 1 Cheong Wan Road, Hung Hom, Hong Kong, China (Hong Kong)

[2]School of Nursing and Midwifery, University of Southern Queensland, Brisbane, Australia

[3]The Salvation Army, Hong Kong, China (Hong Kong)

[4]School of Nursing, University of Rochester, Rochester, NY, United States

[5]School of Nursing, Tung Wah College, Homantin, China (Hong Kong)

Corresponding Author:
Arkers Kwan Ching Wong, PhD
School of Nursing, Faculty of Health and Social Science, The Hong Kong Polytechnic University, Room 502, Block GH, 1 Cheong Wan Road, Hung Hom, Hong Kong, China (Hong Kong)

Abstract

Background: Informal caregivers are essential in supporting community-dwelling older adults, especially as global populations age. However, caregiving responsibilities often result in high stress, depressive symptoms, and diminished quality of life. Traditional support services are fragmented, and many programs overlook caregivers' emotional and informational needs. Telecare—the remote delivery of health and social care via digital technologies—offers a promising approach to provide accessible, personalized, and timely support.

Objective: This study evaluated the effectiveness of a telecare-based intervention program, developed by a health-social partnership team, in improving psychological well-being and caregiving outcomes among informal caregivers of community-dwelling older adults. We hypothesized that caregivers in the intervention group would show greater improvements in stress, self-efficacy, depression, quality of life, and caregiving burden compared to those receiving usual care.

Methods: A single-blind randomized controlled trial was conducted from January to December 2023, with 75 informal caregivers providing care to older adults (aged ≥60 y) for at least 4 hours per week. Participants were randomized to an intervention group (n=38, 51%) or a control group (n=37, 49%). The 3-month intervention included (1) biweekly online nurse case management supported by a multidisciplinary health-social team, (2) personalized WhatsApp videos on caregiving skills, and (3) a password-protected caregiver website offering resources and peer discussion. The control group attended 6 in-person educational sessions at a community center. Outcomes such as stress, self-efficacy, depression, quality of life, and caregiving burden were measured at baseline and postintervention.

Results: Participants had a mean age of 65.5 (SD 9.617) years, and 81.3% (61/74) were female. Baseline characteristics were similar between groups. Compared to controls, the intervention group showed significantly higher self-efficacy (t=1.98; P=.04) and lower depression (t=−2.24; P=.03) at follow-up. Within-group improvements in the intervention group were also observed for stress, depression, self-efficacy, mental quality of life, and caregiving burden. No significant between-group differences were found for stress, quality of life, or caregiving burden. No significant within-group improvements were observed in the control group.

Conclusions: This study provides novel evidence that a digitally delivered, nurse-led telecare program—integrating health and social services—can improve psychological outcomes among informal caregivers. These findings have practical implications for health systems seeking scalable, low-barrier solutions to support caregivers. Telecare models may be a valuable complement to traditional services, enhancing caregiver resilience and promoting aging in place while alleviating long-term strain on institutional care systems.

Trial Registration: ClinicalTrials.gov NCT05636982; https://clinicaltrials.gov/study/NCT05636982

International Registered Report Identifier (IRRID): RR2-10.3389/fpsyt.2023.1167479

(JMIR Mhealth Uhealth 2026;14:e70791) doi:10.2196/70791

XSL•FO
RenderX

KEYWORDS

older adults; informal caregivers; caregiving; care; stress; telecare; community

Introduction

Background

With declining fertility and increasing life expectancy, global populations are aging rapidly. In Hong Kong, the proportion of older adults is projected to rise from 17.9% in 2018 to 31.9% by 2038, placing increasing pressure on health care and welfare systems [1,2]. Most older adults prefer aging in place over institutional care, even when frail, making informal caregivers central to community-based elderly care [3].

These caregivers assist with daily tasks, medical appointments, and decision-making, often acting as crucial advocates for their loved ones [4,5]. While their contributions help reduce health care use and improve outcomes for older adults, caregiving is physically and emotionally demanding. Many caregivers face limited access to support, inadequate training, and social isolation, leading to feelings of guilt, fatigue, and mental distress. Studies show that more than 60% of caregivers experience high caregiving burden, 55% report depressive symptoms, and 40% face poor family functioning [6-8].

Despite their vital role, informal caregivers often receive far less support than care recipients. In Hong Kong, the lack of an integrated health and social care system contributes to fragmented services, making it difficult for caregivers to seek timely and coordinated assistance [9]. A local study in 2022 found that many caregivers expressed a strong need for a centralized platform offering information, caregiving resources, and service referrals [10,11].

Advances in digital health offer promising solutions. Telecare, which leverages information and communication technologies, enables remote monitoring and timely support from health care professionals [12,13]. It is especially useful for working caregivers, offering both practical information and emotional support without the constraints of time and location [3,14].

Several systematic reviews have demonstrated that telehealth or telecare interventions—often delivered via phone or videoconferencing—can moderately reduce caregiver stress, burden, and depressive symptoms, particularly among those caring for individuals with chronic conditions or dementia [14,15]. However, most studies have focused on specific clinical populations and lacked integration with broader community-based health and social services [14]. Few programs have incorporated real-time digital case management, asynchronous tailored messaging, and interactive web-based peer support in a comprehensive model. Furthermore, although some telecare interventions have been tested among older adults with multimorbidity and young populations [16], there remains a significant gap in research targeting informal caregivers of community-dwelling older adults with diverse needs. Addressing these gaps, this study evaluates the effectiveness of a multicomponent telecare program, co-developed by a health-social partnership team, in improving caregivers' psychological well-being and caregiving outcomes. We hypothesized that participants in the intervention group would demonstrate greater reductions in stress and depression and greater improvements in self-efficacy, quality of life, and caregiving burden compared to those receiving usual care.

Conceptual Framework

The design of this program was informed by Lazarus and Folkman's [17] transactional theory of stress and coping and the COPE (Creativity, Optimism, Planning, and Expert Information) model. In line with the transactional theory, the intervention aimed to reduce caregiver stress by enhancing their appraisal and coping capacity. Caregivers were supported to adopt problem-focused strategies (eg, goal setting and seeking solutions) and emotion-focused strategies (eg, stress management and emotional support) to manage caregiving challenges [18].

The COPE model provided a practical framework for empowering caregivers through the intervention components. For instance, creativity was encouraged through guided problem-solving discussions; optimism was fostered by reframing challenges and sharing success stories; planning was embedded in structured goal-setting and action plans; and expert information was delivered via health-social professionals and curated digital content [19]. Together, these models guided the development of a multicomponent program that aimed to enhance caregivers' psychological well-being and self-efficacy in managing care responsibilities.

Methods

Overview

The details can be found in the published protocol [20]. The reporting of the trial adhered to the CONSORT (CONsolidated Standards Of Reporting Trials) 2025 checklist (Checklist 1).

Study Design and Setting

This is a single-blinded, 2-armed randomized controlled trial. The study was conducted in collaboration with one of the largest nongovernmental organizations in the world, which has more than 1.7 million older members.

Participants, Recruitment Strategy, and Randomization

The criteria for the inclusion of participants were as follows: (1) being aged ≥18 years; (2) having no cognitive impairment; (3) being able to understand and communicate in Cantonese; (4) providing care for older adults aged ≥60 years for a minimum of 4 hours per week over a period of at least 3 months; (5) being a current smartphone user; (6) having knowledge of how to access the internet; (7) committing to attend biweekly online meetings with the program providers, lasting 15 to 30 minutes, over a period of 3 months; and (8) being willing to receive individual-specific video messages covering caregiving skills via WhatsApp (Meta Platforms Inc.). Exclusion criteria included (1) being an alcoholic or using psychiatric drugs, (2) being illiterate (inability to read and write), (3) having a diagnosis of

XSL•FO
RenderX

chronic mental problems, and (4) having previously participated in health or social programs offered by other organizations.

Participants were recruited between January 2023 and September 2024. The intervention was delivered from January 2023 to December 2024. In community centers, a trained research assistant assessed and recruited eligible participants by sharing details of the study with them and inviting them to join. Verbal consent was obtained from those who agreed to join the program. Following the collection of baseline data, the research assistant contacted the principal investigator of the research team to proceed with randomization. The principal investigator used a software program called "Research Randomizer" to generate numbers, which were placed in sealed envelopes. Each envelope corresponded to either the intervention group ("1") or the control group ("2"). The envelopes were opened sequentially upon receiving a call from the research assistant. In this study, the research assistant was blinded to group assignments, whereas the principal investigator and health care providers were not blinded.

Intervention Group

Overview

The participants in the telecare-based intervention group received a 3-month program consisting of three intervention components: (1) online nurse case management supported by a health-social partnership team, (2) individual-specific video messages covering caregiving skills via WhatsApp, and (3) an online information center and discussion forum accessed via a password-protected, newly developed caregiver website.

Online Nurse Case Management Supported by a Health-Social Partnership Team

Online case management was facilitated by a registered nurse who played a leading role, along with a team of health care professionals from various disciplines, including physiotherapists, Chinese medicine practitioners, and social workers. The registered nurse conducted individual Zoom meetings with each participant in the intervention group every 2 weeks, for a total of 6 Zoom meetings per participant throughout the intervention period. During the initial meeting, the nurse encouraged caregivers to share their current situation and explored any needs or difficulties they were facing, using the Omaha system. The Omaha system comprises the Problem Classification Scheme, the Problem Rating Scale for Outcomes, and the Intervention Scheme [20,21]. This structured approach assisted the nurse in identifying the health needs of the caregivers, setting intervention goals, and evaluating the effectiveness of the intervention. Subsequently, personalized care plans were developed for each participant. The nurse provided recommendations and consulted other health care professionals from relevant disciplines as needed. Follow-up meetings involved ongoing communication with caregivers to discuss any updated situations and provide tailored physical, social, and emotional support based on their specific needs. The nurses offered additional resources and made referrals to the health-social partnership team whenever necessary. This approach leveraged technology to facilitate the provision of coordinated care tailored to the unique needs of each individual. It involved coordinating the efforts of health care professionals to ensure that the clients' requirements were effectively met.

Individual-Specific Video Messages Covering Caregiving Skills Via WhatsApp

With the widespread popularity of WhatsApp, the registered nurse used this platform to offer additional support to the caregivers. The nurse sent individualized messages or videos weekly to the participants through WhatsApp, focusing on their needs as identified from Zoom meetings. The information that was shared encompassed caregiving knowledge and skills, self-care recommendations, health information, and tips for managing emotions. Using this convenient communication tool, the nurse ensured the timely delivery of information while also providing an opportunity for participants to engage in asynchronous communication with the nurse. In the event that participants had any questions or concerns, they could easily respond to the nurse's messages through WhatsApp, allowing the nurse to provide prompt responses and support.

Online Information Center and Discussion Forum Via a Password-Protected, Newly Developed Caregiver Website

A new website was created specifically for informal caregivers, consisting of two components: an information center and a discussion forum. At the beginning of the intervention period, each participant was randomly assigned a preset username and password to log into the website. The usernames were kept anonymous to encourage participants to take part without hesitation, and the passwords were securely protected.

The research assistant played a role in maintaining the website by posting 2 to 3 pieces of useful information in the information center on a weekly basis for caregivers to read. Caregivers were able to comment on these posts and raise questions related to caregiving. Members of the health-social partnership team promptly replied to these questions, providing valuable insights and guidance.

Regarding the discussion forum, participants had the freedom to initiate any caregiving-related questions or discuss the challenges they were facing at any time. The members of the health-social team were available to respond to these questions, and participants were also encouraged to interact with each other, sharing their perspectives and providing suggestions. To foster engagement, the research assistant occasionally initiated new discussion topics, encouraging participants to share their own experiences.

The website served as more than just a source of knowledge and information for caregivers; it also provided them with a platform to express their feelings, receive encouragement, and connect with other caregivers who shared similar experiences.

Control Group

In addition to the telecare-based intervention group, the control group also received the usual services from community centers. The participants in the control group attended 6 face-to-face educational talks during the intervention period, which took place at the community center. These talks were conducted by members of the health-social partnership team, including a

XSL•FO
RenderX

registered nurse, a Chinese medicine practitioner, a physiotherapist, and a mental health nurse.

Each talk covered 1 or 2 specific topics and lasted approximately 2 hours. The health care professionals provided informal caregivers with knowledge and live demonstrations related to self-care and elderly care. The topics addressed during the talks included the management of chronic diseases, acupressure techniques, eye and oral care for older adults, pain management strategies, lifting and transfer techniques, fall prevention measures, swallowing issues, and the management of insomnia. During the questions and answers session of each talk, participants had the opportunity to ask questions and seek clarification on the topics that had been discussed. Upon completion of the follow-up data collection, participants in the control group were offered access to the online caregiver platform and its educational resources to ensure equitable support and knowledge dissemination.

Data Collection

The data were collected at two time points: at baseline (T1) and at the follow-up after the interventions (T2). Data collection for the study involved the use of Qualtrics, a survey platform, where participants were provided with a link to complete the questionnaires themselves. However, for participants who encountered difficulties in filling out the questionnaires online, alternative methods were used. In such cases, data collection took place face-to-face with the assistance of the research assistant and staff from the nongovernmental organizations. They provided support and guidance to the participants, ensuring that the questionnaires were completed accurately and comprehensively. This approach allowed for flexibility in data collection, accommodating participants who may have experienced challenges with completing the online survey.

Outcome Measures

Primary Outcome

The 14-Item Chinese Version of the Perceived Stress Scale

The primary outcome of the study was the stress level of the caregivers. The 14-item Chinese version of the Perceived Stress Scale was used to evaluate the participants' perception of stress and the degree to which they perceived their lives to be uncontrollable [22]. All items are rated on a 5-point Likert scale, ranging from 0 (never) to 4 (very often), and 7 of them (4, 5, 6, 7, 9, 10, and 13) are reverse-scored, with higher scores representing lesser perceived stress. The overall Cronbach α of the Chinese version of the Perceived Stress Scale was 0.906, indicating good reliability [23].

Secondary Outcomes

The secondary outcomes included self-efficacy, depression, quality of life, and caregiving burden of the caregivers.

Self-Efficacy—The Chinese Version of the General Self-Efficacy Scale

The self-efficacy of the individuals was assessed using the Chinese version of the General Self-Efficacy Scale. It is a scale that is widely used to assess an individual's belief in his or her own capacity to cope with stressful situations or difficulties [24,25]. It contains 10 items, and scores for each item range from 1 (not at all true) to 4 (exactly true). The higher the total score, the greater the individual's self-efficacy. The reliability of the Chinese version of the General Self-Efficacy Scale is good, with a Cronbach α of >0.91, indicating internal consistency [26]. The scale has been shown to have good criterion validity and good psychometric properties [27].

Depression—The Chinese Version of the Center for Epidemiologic Studies Depression Scale

The Chinese version of the Center for Epidemiologic Studies Depression Scale is a widely used self-reported questionnaire consisting of 20 items that assess depression symptomatology. It is designed to measure the frequency and severity of depressive symptoms experienced by individuals. The scale has been widely used in research and clinical settings to assess symptoms of depression in Chinese-speaking populations [28-30]. The Chinese version of the Center for Epidemiologic Studies Depression Scale has been validated, with a Cronbach α and intraclass correlation coefficient of the total scale and subscales ranging from 0.82 to 0.88 and from 0.73 to 0.81, respectively [31].

Quality of Life—The Chinese Short Form-12 Health Survey Version 2

The Chinese Short Form-12 Health Survey version 2 is a measure used to assess health-related quality of life. It evaluates 8 domains, namely physical functioning, role physical, bodily pain, general health, vitality, social functioning, role emotional, and mental health [32]. It consists of 12 items divided into a physical component summary score and a mental component summary score [33]. Higher scores indicate that an individual has a better health-related quality of life. The reliability of the Chinese Short Form-12 Health Survey version 2 was good, with a Cronbach α of 0.81 [34].

Caregiving Burden—The Short Form of the Zarit Burden Interview

The Chinese version of the Zarit Burden Interview is a 22-item scale widely used to assess the level of caregiving burden of informal caregivers [35]. The response to each item is rated on a 5-point scale, ranging from 0 (never) to 4 (always). The total score ranges from 0 to 88, with higher scores indicating a greater caregiving burden. The Chinese version of the short form of the Zarit Burden Interview demonstrates very good internal consistency, with a Cronbach α of 0.875 [36,37].

Background Demographic Data

The background demographic data were collected at baseline (T1). They included age, gender, marital status, educational level, employment, economic status, caregiving experience, time of providing care to the older adult per week, relationship with the care recipient, and experience using a smartphone.

Sample Size

A 2-tailed alpha of .05, a power of 80% (a beta error probability of .2), and an effect size of 0.403 were used to determine the size of the sample. This effect size was based on a recent telephone-based mobile health program that assessed the same

XSL•FO
RenderX

primary outcome measure as in this study, namely perceived stress level [38]. On the basis of these considerations, a sample size of 34 individuals per group was necessary.

Considering an attrition rate of 10% to 15% reported in a previous program [38], a dropout rate of 15% was assumed for this study. Taking into account this dropout rate, the total sample size required was 80 participants, with 40 participants allocated to each group.

Statistical Analysis

The statistical analysis was conducted using the Statistical Package for the Social Sciences (version 29.0). To compare the demographic data of the intervention and control groups at baseline, the Mann-Whitney *U* test, *t* tests, and chi-square tests were used. Normality was tested using frequency distribution histograms, and all data were found to conform to a normal distribution. Therefore, independent *t* tests (two-tailed) were used to analyze the mean differences in primary and secondary outcomes between the two groups, and paired *t* tests (two-tailed) were used to compare within-group changes before and after the intervention. A primary intention-to-treat analysis was performed. Missing outcome data for participants lost to follow-up were addressed using multiple imputation based on the assumption of missing completely at random. A significant result was determined when the *P* value (level of significance) was .04 for a 2-tailed test, indicating a statistically significant difference.

Ethical Considerations

This study was approved by the Human Subjects Ethics Sub-Committee of the Hong Kong Polytechnic University (reference number HSEARS20210907001). All participants were provided with detailed information about the study's purpose, procedures, risks, and benefits. Written informed consent was obtained from all participants before data collection, and they were informed of their right to withdraw at any time without consequences. All collected data were anonymized, and identifying information was removed before analysis to ensure confidentiality. Secure storage procedures were implemented to safeguard digital and paper records. No personal identifiers were used in reporting results. Participants received a supermarket coupon valued at HK $50 (approximately US $6.50) as a token of appreciation for their time and effort.

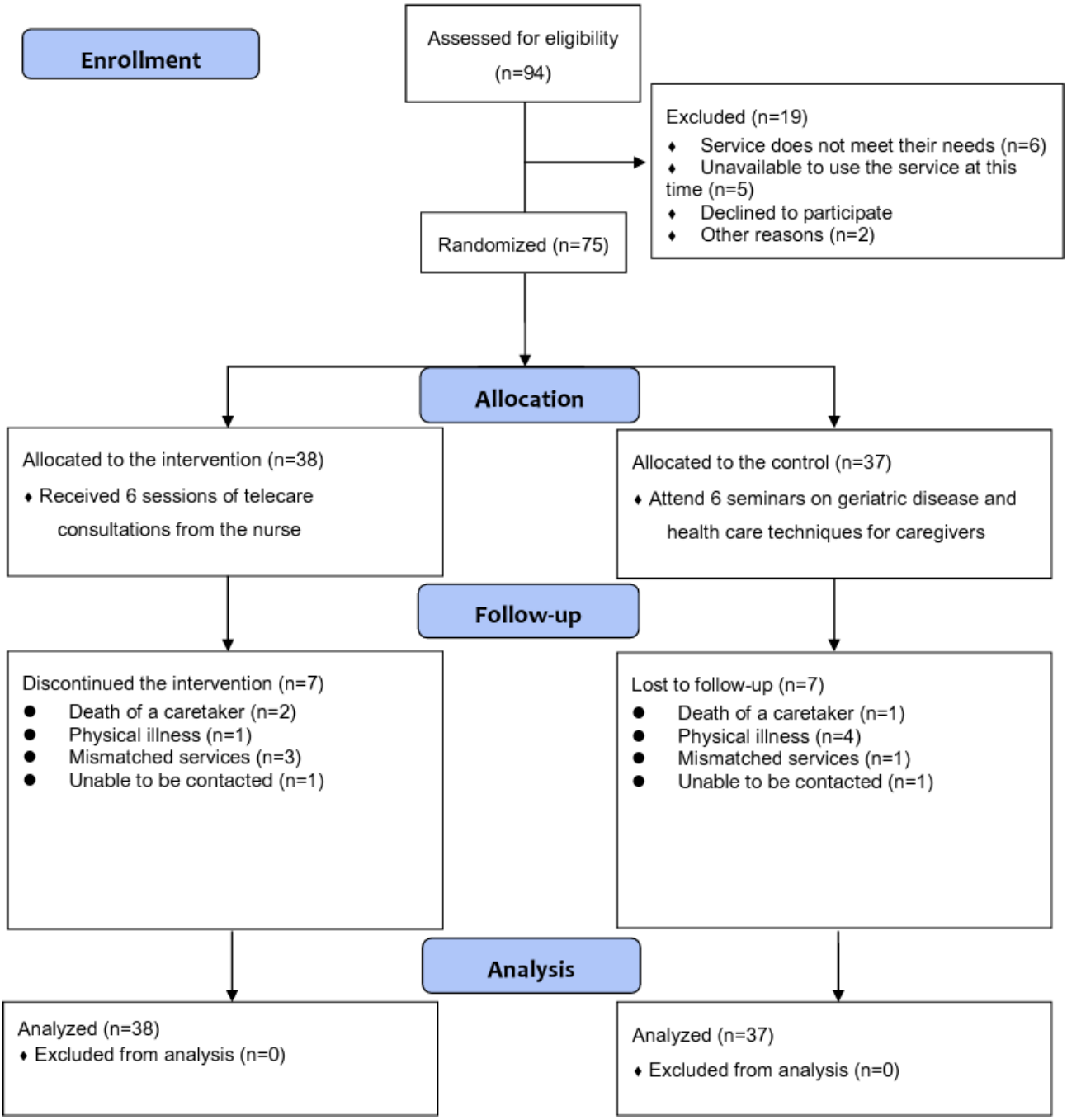
Results

Baseline Characteristics

A total of 94 participants were assessed for eligibility. Of these, 75 (79.8%) individuals (n=38, 51%, in the intervention group; and n=37, 49%, in the control group) were eventually recruited and randomized for the study. Seven participants from both groups dropped out for various reasons, such as the death of a caretaker, physical illness, mismatched services, or they were unable to be contacted (Figure 1). Among the 75 participants, 14 (18.7%) were male and 61 (81.3%) were female, with a mean age of 65.5 years. Approximately 62.7% (47/75) were retired, and 52.0% (39/75) were caregivers for their parents. A significant proportion (44/75, 58.6%) had been taking care of older adults for 5 years or longer. All participants were experienced in using smartphones. There were no statistically significant differences in participant characteristics or outcome measures between the groups at baseline. Table 1 presents the demographic characteristics of the participants at baseline.

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards Of Reporting Trials) flow diagram.



XSL•FO
RenderX

Table . Baseline characteristics of both the intervention and control groups.

Characteristics	Total (n=75)	Intervention group (n=38)	Control group (n=37)	*P* value
Gender, n (%)				.59
Female	61 (81.3)	30 (78.9)	31 (83.8)	
Male	14 (18.7)	8 (21.1)	6 (16.2)	
Age (y)				.22
Mean (SD)	65.5 (9.617)	64.2 (9.519)	66.7 (9.686)	
Median (range)	67 (44 - 84)	65.5 (44-81)	70 (46 - 84)	
Age range (y), n (%)				.22
40-50	6 (8.0)	4 (10.5)	2 (5.4)	
51-60	17 (22.7)	9 (23.7)	8 (21.6)	
61-70	29 (38.7)	16 (42.1)	13 (35.1)	
71-80	20 (26.7)	8 (21.1)	12 (32.4)	
81-84	3 (4.0)	1 (2.6)	2 (5.4)	
Marital status, n (%)				.14
Not married	53 (70.7)	23 (60.5)	30 (81.1)	
Married	19 (25.3)	13 (34.2)	6 (16.2)	
Divorced	3 (4.0)	2 (5.3)	1 (2.7)	
Education level, n (%)				.10
Primary	9 (12.0)	4 (10.5)	5 (13.5)	
Secondary	42 (56.0)	19 (50.0)	23 (62.2)	
Tertiary or above	22 (29.3)	15 (39.5)	7 (18.9)	
Others	2 (2.7)	0 (0.0)	2 (5.4)	
How many hours do you spend caring for your care recipient? (on a weekly basis) (h), n (%)				.21
1 - 10	14 (18.7)	5 (13.2)	9 (24.3)	
11 - 20	7 (9.3)	5 (13.2)	2 (5.4)	
21 - 30	13 (17.3)	6 (15.8)	7 (18.9)	
31 - 40	5 (6.7)	3 (7.9)	2 (5.4)	
41 - 50	2 (2.7)	2 (5.3)	0 (0.0)	
51 - 60	7 (9.3)	3 (7.9)	4 (10.8)	
61 - 70	3 (4.0)	3 (7.9)	0 (0.0)	
71 - 80	2 (2.7)	1 (2.6)	1 (2.7)	
81 - 90	1 (1.3)	0 (0.0)	1 (2.7)	
91 - 100	0 (0.0)	0 (0.0)	0 (0.0)	
101 - 110	1 (1.3)	0 (0.0)	1 (2.7)	
111 - 120	2 (2.7)	2 (5.3)	0 (0.0)	
121 - 130	2 (2.7)	1 (2.6)	1 (2.7)	
131 - 140	0 (0.0)	0 (0.0)	0 (0.0)	
141 - 150	1 (1.3)	1 (2.6)	0 (0.0)	
151 - 160	1 (1.3)	0 (0.0)	1 (2.7)	
161 - 170	14 (18.7)	6 (15.8)	8 (21.6)	

XSL•FO
RenderX

Characteristics	Total (n=75)	Intervention group (n=38)	Control group (n=37)	*P* value
Relationship with the care recipient, n (%)				.10
Spouse	33 (44.0)	13 (34.2)	20 (54.1)	
Parents	39 (52.0)	22 (57.9)	17 (45.9)	
Sibling	1 (1.3)	1 (2.6)	0 (0.0)	
Others	2 (2.7)	2 (5.3)	0 (0.0)	
Experience of taking care of the care recipient, n (%)				.20
<6 mo	6 (8.0)	2 (5.3)	4 (10.8)	
6 mo to 1 y	6 (8.0)	5 (13.2)	1 (2.7)	
2-5 y	19 (25.3)	11 (28.9)	8 (21.6)	
5-10 y	31 (41.3)	16 (42.1)	15 (40.5)	
>10 y	13 (17.3)	4 (10.5)	9 (24.3)	
Occupation, n (%)				.06
Full time	11 (14.7)	10 (26.3)	1 (2.7)	
Part time	1 (1.3)	0 (0.0)	1 (2.7)	
Unemployed	1 (1.3)	0 (0.0)	1 (2.7)	
Retired	47 (62.7)	23 (60.5)	24 (64.9)	
No intention to find a job	15 (20.0)	5 (13.2)	10 (27.0)	
Living (more than one option), n (%)				
Alone	5 (5.6)	3 (7.1)	3 (6.1)	.62
With parents	12 (13.3)	7 (16.7)	5 (10.2)	.56
With spouse	41 (45.5)	16 (38.1)	25 (51.0)	.03
With children	15 (16.7)	6 (14.3)	9 (18.4)	.36
With domestic helper	12 (13.3)	7 (16.7)	5 (10.2)	.53
With sibling	5 (5.6)	3 (7.1)	2 (4.1)	1.00
Experience of using a smartphone (y), n (%)				.11
3 - 6	18 (24.0)	10 (26.3)	8 (21.6)	
7 - 10	40 (53.3)	18 (47.4)	22 (59.5)	
11 - 14	5 (6.7)	2 (5.3)	3 (8.1)	
15 - 18	7 (9.3)	4 (10.5)	3 (8.1)	
>18	5 (6.7)	4 (10.5)	1 (2.7)	.

Perceived Stress Level

There were no statistically significant differences in the means between the 2 groups at T2. However, a statistically significant within-group difference was found in the intervention group, where the perceived stress level of the participants was lower at postintervention than preintervention (t=−2.21; P=.04). No within-group difference was observed in the control group.

Self-Efficacy

The intervention group had statistically significant higher self-efficacy scores (t=1.98; P=.04) when compared to the control group. There was also a within-group effect on self-efficacy level in the intervention group (t=−2.28; P=.03).

Depression

The participants in the intervention group had statistically lower depression scores than those in the control group (t=−2.24; P=.03). The intervention group showed a significant within-group effect (t=2.98; P=.009), whereas no significant within-group effect was observed in the control group.

Quality of Life

There were no statistically significant within-group effects in either the intervention group or the control group from T1 to

XSL•FO
RenderX

T2 in the physical components of quality of life. A statistically significant within-group difference was observed in the intervention group from T1 to T2 in the mental components of quality of life. There was no between-group difference at T2 in either the physical or mental components of quality of life.

Caregiving Burden

Similar to perceived stress levels, there was no statistically significant between-group difference at T2 for caregiving burden. However, a statistically significant within-group difference was found in the intervention group from T1 to T2 (t=2.89; P=.009), where participants had a lower caregiving burden postintervention than preintervention.

Comparisons within and between groups in the pretest and posttest mean scores for all primary and secondary outcomes can be found in Table 2.

XSL•FO
RenderX

Table . Within- and between-group comparisons of the pretest and posttest mean scores for primary and secondary outcomes.

Variables	Intervention group (n=38), mean (SD)	Control group (n=37), mean (SD)	*t*	*P* value
Perceived stress level				
Pretest	30.35 (7.08)	31.47 (8.10)	−0.47[a]	.64
Posttest	32.58 (8.01)	33.47 (6.84)	−0.46[a]	.64
t	−2.21[b]	−1.12[b]	N/A[c]	N/A
P value	.04[d]	.27	N/A	N/A
Self-efficacy				
Pretest	24.00 (5.01)	24.93 (4.86)	−0.77[a]	.44
Posttest	26.90 (6.53)	23.27 (6.50)	1.98[a]	.04[d]
t	−2.28[b]	0.90[b]	N/A	N/A
P value	.03[d]	.84	N/A	N/A
Depression				
Pretest	16.16 (7.93)	15.03 (12.57)	.11[a]	.91
Posttest	10.07 (9.05)	14.16 (8.66)	−2.24[a]	.03[d]
t	2.98[b]	1.11[b]	N/A	N/A
P value	.009[d]	.29	N/A	N/A
Quality of life				
Physical component				
Pretest	43.46 (8.86)	42.11 (9.15)	.81[a]	.42
Post-test	44.77 (7.30)	42.45 (9.20)	1.09[a]	.28
t	−0.81[b]	−0.30[b]	N/A	N/A
P value	.42	.77	N/A	N/A
Mental component				
Pretest	43.95 (11.51)	47.39 (12.61)	−0.74[a]	.32
Posttest	49.48 (8.20)	48.11 (9.37)	.61[a]	.55
t	−1.99[b]	−0.96[b]	N/A	N/A
P value	.04[d]	.58	N/A	N/A
Caregiving burden				
Pretest	40.90 (16.47)	32.20 (15.27)	1.80[a]	.075
Posttest	30.83 (18.40)	35.32 (16.99)	−.99[a]	.33
t	2.89[b]	−0.89[b]	N/A	N/A
P value	.009[d]	.60	N/A	N/A

[a] Independent *t* test.

[b] Paired *t* test.

[c]N/A: not applicable.

[d]*P*<.05; statistical tests used: chi-square test/Kruskal-Wallis test.

XSL•FO
RenderX

Discussion

Principal Findings

As populations increasingly age, the role of informal caregivers remains critical in promoting aging in place and ensuring continuity of care. This randomized controlled trial evaluated the effectiveness of a theory-informed, telecare-based intervention in improving caregiver outcomes. The findings demonstrate that caregivers who received the intervention reported significantly higher self-efficacy and lower levels of depression compared to those receiving usual care. Additionally, significant within-group improvements were observed in stress levels, mental quality of life, and caregiving burden in the intervention group, although not reflected in between-group comparisons.

The findings of this randomized controlled trial provide evidence that a telecare-based intervention, grounded in psychological and behavioral theories, can significantly improve depression and self-efficacy among informal caregivers of older adults. These findings align with prior studies demonstrating that accessible, technology-enabled interventions can offer caregivers both emotional support and practical guidance, leading to enhanced psychological resilience and self-management skills [39,40]. Notably, the individualized approach in this intervention—through nurse-led case management and tailored video content—may have allowed caregivers to address their unique caregiving challenges, which supports findings from prior personalized eHealth models [41].

While between-group differences were not observed for perceived stress and caregiving burden, significant within-group improvements in these domains suggest that the intervention had a positive psychosocial effect. This reinforces the importance of early and sustained caregiver support, particularly in settings where community resources are fragmented. Given the rising demand for informal care and global efforts to enable aging in place, these results contribute to a growing body of literature supporting the integration of digital care into community-based services.

The observed reduction in depression among participants may also be attributed to the intervention's multidimensional support system. Caregivers had multiple channels for engagement, including structured video consultations, asynchronous communication via WhatsApp, and peer interaction on a dedicated web platform. This combination not only improved knowledge and problem-solving capabilities but may also have addressed the social isolation frequently reported by caregivers [42]. The inclusion of personalized content likely enhanced relevance and motivation, 2 factors known to influence behavior change in caregiver populations [43].

The lack of significant improvement in the physical components of quality of life may be attributed to the short duration of the intervention, as changes in physical functioning often require longer follow-up or more intensive caregiver support services. Additionally, variations in care recipients' health status and caregiving intensity could have influenced these outcomes. Future studies should consider stratifying analyses by caregiving complexity and incorporating longer-term follow-up, cost-effectiveness analysis, and implementation evaluation to better understand differential effects and inform wider adoption.

Limitations

Several limitations should be noted. First, this trial was prospectively powered for the primary outcome and not powered for subgroup analyses or for detecting smaller effects on secondary outcomes; these estimates should therefore be interpreted with caution. Second, the intervention could not be blinded to participants or providers, potentially introducing response bias. Third, some participants encountered technical difficulties with WhatsApp and the online forum, suggesting that digital literacy and access may affect scalability. Fourth, the duration and content of nurse case management sessions were not formally tracked, limiting fidelity assessment. Fifth, only short-term effects were measured; long-term follow-up is needed to assess sustained impact.

Additionally, we did not collect data on care recipients' functional status, digital literacy, or the caregivers' relationship to the care recipient and caregiving network size. These contextual factors may influence intervention effectiveness and should be considered in future research to enable subgroup analyses and more tailored support.

Conclusions

This study demonstrates that integrating health-social partnership teams with telecare services can significantly improve caregivers' psychological outcomes. Such community-based interventions offer accessible, tailored support to informal caregivers, helping to reduce emotional burden and promote mental well-being—both of which are essential to sustaining long-term caregiving capacity and ensuring quality of care.

Acknowledgments

The authors would like to thank the Salvation Army for their collaboration and express our gratitude to the telecommunications company for their technical support.

Ho Ching Lau was originally the sixth author in the byline but has been removed from the authorship list due to uncontactability during the production process as they were unable to approve revisions or the final, accepted version of the manuscript. Additionally, the Salvation Army and Ho Ching Lau's immediate supervisor at the Salvation Army approve the removal on the ground of uncontactability and confirm Ho Ching Lau was unable to approve the final version of the manuscript or provide a licence to publish agreement. All remaining authors approve of this change and the assessment of the Salvation Army.

XSL•FO
RenderX

Funding

This work was supported by a grant from the Undergraduate Research and Innovation Scheme of the Hong Kong Polytechnic University (reference number 1-TA46). The funder had no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: AKCW, NPN
Funding acquisition: AKCW
Writing original draft: AKCW, NPN, JB
Writing – review & editing: AKCW, NPN, MY, KKSY, RSYL, JJS, VH, and JM

Conflicts of Interest

None declared.

Checklist 1
CONSORT checklist.
[PDF File, 148 KB - mhealth_v14i1e70791_app1.pdf]

References

1. Population ageing trend of hong kong. : Office of the Government Economist; 2019 URL: https://www.hkeconomy.gov.hk/en/pdf/el/el-2019-02.pdf [accessed 2025-01-12]
2. Thematic household survey report no 74 health status of hong kong residents. : Hong Kong SAR: Census and Statistics Department; 2021 URL: https://www.censtatd.gov.hk/en/data/stat_report/product/B1130201/att/B11302742021XXXXB0100.pdf [accessed 2025-01-12]
3. Wong AKC, Wong FKY, Wong MCS, Chow KKS, Kwan DKS, Lau DYS. A community-based health-social partnership program for community-dwelling older adults: a hybrid effectiveness-implementation pilot study. BMC Geriatr 2022 Oct 7;22(1):789. [doi: 10.1186/s12877-022-03463-z] [Medline: 36207685]
4. Amiresmaili M, Emrani Z. Informal caregivers in hospitals: opportunities and threats. Int J Health Plann Manage 2018 Oct;33(4):880-889. [doi: 10.1002/hpm.2543] [Medline: 29781090]
5. Wang Y, Yang W, Avendano M. Does informal care reduce health care utilisation in older age? Evidence from China. Soc Sci Med 2022 Aug;306:115123. [doi: 10.1016/j.socscimed.2022.115123] [Medline: 35724586]
6. The challenging roles of informal carers. : UNECE; 2019 URL: https://unece.org/DAM/pau/age/Policy_briefs/ECE_WG1_31.pdf [accessed 2025-01-16]
7. Sarabia-Cobo C, Sarriá E. Satisfaction with caregiving among informal caregivers of elderly people with dementia based on the salutogenic model of health. Appl Nurs Res 2021 Dec;62:151507. [doi: 10.1016/j.apnr.2021.151507] [Medline: 34815003]
8. Research on the profile and service needs of elderly carers: 25% elderly caregivers found as high risk level having high caring burden, depressive symptoms and poor family functioning. : The Hong Kong Council of Social Service; 2018 URL: https://www.hkcss.org.hk/research-on-the-profile-and-service-needs-of-elderly-carers/?lang=en [accessed 2025-01-16]
9. Wong AKC, Wong FKY, Ngai JSC, Hung SYK, Li WC. Effectiveness of a health-social partnership program for discharged non-frail older adults: a pilot study. BMC Geriatr 2020 Sep 10;20(1):339. [doi: 10.1186/s12877-020-01722-5] [Medline: 32912218]
10. Lo SKW. Informal caregivers in Hong Kong: a review and an agenda for policy development. Hong Kong J Social Work 2022 Jan;56(1n02):135-161. [doi: 10.1142/S0219246222000110]
11. Consultancy study on needs and support required of carers of elderly persons and of persons with disabilities in hong kong: main report. : Hong Kong Polytechnic University Consulting Team; 2022 URL: https://www.lwb.gov.hk/en/other_info/1.%20Carer%20Study-Main%20Report%20(Eng)_2022.05-r.pdf [accessed 2025-01-19]
12. Shirk S, Moo LR, McLaren JE, Pugh K, Nguyen AH, O'Connor MK. TeleCARE: a telehealth group to teach dementia caregivers how to manage neuropsychiatric symptoms. Alzheimer's & Dementia 2023 Dec;19(S19):e077540 [FREE Full text] [doi: 10.1002/alz.077540]
13. Wong AKC, Kwok VWY, Wong FKY, et al. Improving post-acute stroke follow-up care by adopting telecare consultations in a nurse-led clinic: study protocol of a hybrid implementation-effectiveness trial. J Adv Nurs 2024 Mar;80(3):1222-1231. [doi: 10.1111/jan.15960] [Medline: 37950400]

XSL•FO
RenderX

14. Graven LJ, Glueckauf RL, Regal RA, Merbitz NK, Lustria MLA, James BA. Telehealth interventions for family caregivers of persons with chronic health conditions: a systematic review of randomized controlled trials. Int J Telemed Appl 2021;2021:3518050. [doi: 10.1155/2021/3518050] [Medline: 34093704]
15. Zhai S, Chu F, Tan M, Chi NC, Ward T, Yuwen W. Digital health interventions to support family caregivers: an updated systematic review. Digit Health 2023;9:20552076231171967. [doi: 10.1177/20552076231171967] [Medline: 37223775]
16. Kirakalaprathapan A, Oremus M. Efficacy of telehealth in integrated chronic disease management for older, multimorbid adults with heart failure: a systematic review. Int J Med Inform 2022 Jun;162:104756. [doi: 10.1016/j.ijmedinf.2022.104756] [Medline: 35381436]
17. Lazarus RS, Folkman S. Stress, Appraisal, and Coping: Springer Pub. Co; 1984. URL: https://books.google.com.hk/books?id=i-ySQQuUpr8C [accessed 2025-07-11]
18. Ghaffari M, Morowatisharifabad MA, Jadgal MS, Mehrabi Y, Alizadeh S. The effectiveness of intervention based on the transactional model on improving coping efforts and stress moderators in hemodialysis patients in Tehran: a randomized controlled trial. BMC Nephrol 2021 Nov 11;22(1):377. [doi: 10.1186/s12882-021-02592-8] [Medline: 34763652]
19. Tofthagen C, Chesak SS. Creativity, optimism, planning, and expert advise (COPE): a problem-solving intervention for supporting cancer patients and their family caregivers. In: Holtslander L, Peacock S, Bally J, editors. Hospice Palliative Home Care and Bereavement Support: Springer; 2019:169-179. [doi: 10.1007/978-3-030-19535-9_13]
20. Wong AKC, Ng NP, Hui VCC, Montayre J. Effect of a telecare-based intervention on stress levels in informal caregivers of older adults: protocol for a randomized controlled trial. Front Psychiatry 2023;14:1167479. [doi: 10.3389/fpsyt.2023.1167479] [Medline: 37377468]
21. Martin KS. The Omaha System: A Key to Practice, Documentation, and Information Management, 2nd edition: Elsevier Saunders; 2005. URL: https://books.google.com.mt/books?id=89Rjdd-ulIwC&hl=it&source=gbs_citations_module_r&cad=5 [accessed 2025-07-11]
22. Meng R, Li J, Wang Z, et al. The Chinese version of the Perceived Stress Questionnaire: development and validation amongst medical students and workers. Health Qual Life Outcomes 2020 Mar 13;18(1):70. [doi: 10.1186/s12955-020-01307-1] [Medline: 32169070]
23. Shi C, Guo Y, Ma H, Zhang M. Psychometric validation of the 14-item perceived stress scale in Chinese medical residents. Curr Psychol 2019 Dec;38(6):1428-1434. [doi: 10.1007/s12144-019-00285-y]
24. Schwarzer R, Jerusalem M. Generalized self-efficacy scale. In: Weinman J, Wright S, Johnston M, editors. Measures in Health Psychology: A User's Portfolio Windsor: NFER-Nelson; 1995:35-37 URL: https://userpage.fu-berlin.de/health/selfscal.htm [accessed 2025-07-11]
25. Schwarzer R, Bäßler J, Kwiatek P, Schröder K, Zhang JX. The assessment of optimistic self‐beliefs: comparison of the German, Spanish, and Chinese versions of the general self‐efficacy scale. Applied Psychology 1997 Jan;46(1):69-88. [doi: 10.1111/j.1464-0597.1997.tb01096.x]
26. Leung DYP, Leung AYM. Factor structure and gender invariance of the Chinese General Self-Efficacy Scale among soon-to-be-aged adults. J Adv Nurs 2011 Jun;67(6):1383-1392. [doi: 10.1111/j.1365-2648.2010.05529.x] [Medline: 21129011]
27. Zeng G, Fung SF, Li J, Hussain N, Yu P. Evaluating the psychometric properties and factor structure of the general self-efficacy scale in China. Curr Psychol 2022 Jun;41(6):3970-3980. [doi: 10.1007/s12144-020-00924-9]
28. Niu L, He J, Cheng C, Yi J, Wang X, Yao S. Factor structure and measurement invariance of the Chinese version of the Center for Epidemiological Studies Depression (CES-D) scale among undergraduates and clinical patients. BMC Psychiatry 2021 Sep 23;21(1):463. [doi: 10.1186/s12888-021-03474-x] [Medline: 34556088]
29. Radloff LS. The CES-D scale: a self report depression scale for research in the general population. Appl Psychol Meas 1977;1:385-401. [doi: 10.1177/014662167700100306]
30. Weissman MM, Sholomskas D, Pottenger M, Prusoff BA, Locke BZ. Assessing depressive symptoms in five psychiatric populations: a validation study. Am J Epidemiol 1977 Sep;106(3):203-214. [doi: 10.1093/oxfordjournals.aje.a112455] [Medline: 900119]
31. Li J, Zhao F, Bai H, et al. Psychometric properties of the Chinese version of the center for epidemiologic studies depression scale-revised in patients with cancer: a cross-sectional study. Int J Nurs Stud 2019 Sep;97:14-20. [doi: 10.1016/j.ijnurstu.2019.04.008] [Medline: 31129444]
32. Fong DYT, Wong JYH, Choi EPH, Lam KF, Kwok C. The English and Chinese language versions of the Short Form 12-item Health Survey are equivalent. Health Qual Life Outcomes 2021 Jan 7;19(1):14. [doi: 10.1186/s12955-020-01653-0] [Medline: 33413452]
33. Self-rated health status and quality of life. : Centre for Health Protection; 2015 URL: https://www.chp.gov.hk/files/pdf/chapter_2_eng.pdf [accessed 2025-02-02]
34. Lam ETP, Lam CLK, Fong DYT, Huang WW. Is the SF-12 version 2 Health Survey a valid and equivalent substitute for the SF-36 version 2 Health Survey for the Chinese? J Eval Clin Pract 2013 Feb;19(1):200-208. [doi: 10.1111/j.1365-2753.2011.01800.x] [Medline: 22128754]
35. Lai DWL. Validation of the zarit burden interview for Chinese Canadian caregivers. Soc Work Res 2007 Mar 1;31(1):45-53. [doi: 10.1093/swr/31.1.45]

XSL•FO
RenderX

36. Lu L, Wang L, Yang X, Feng Q. Zarit Caregiver Burden Interview: development, reliability and validity of the Chinese version. Psychiatry Clin Neurosci 2009 Dec;63(6):730-734. [doi: 10.1111/j.1440-1819.2009.02019.x] [Medline: 19781014]
37. Ko KT, Yip PK, Liu SI, Huang CR. Chinese version of the Zarit caregiver Burden Interview: a validation study. Am J Geriatr Psychiatry 2008 Jun;16(6):513-518. [doi: 10.1097/JGP.0b013e318167ae5b] [Medline: 18515696]
38. Sarabia-Cobo C, Pérez V, de Lorena P, Sáenz-Jalón M, Alconero-Camarero AR. Effectiveness of a telephone intervention based on motivational health coaching for improving the mental health of caregivers of people with dementia: a randomised controlled trial. Int J Older People Nurs 2021 Sep;16(5):e12398. [doi: 10.1111/opn.12398] [Medline: 34337869]
39. Biliunaite I, Kazlauskas E, Sanderman R, Truskauskaite-Kuneviciene I, Dumarkaite A, Andersson G. Internet-based cognitive behavioral therapy for informal caregivers: randomized controlled pilot trial. J Med Internet Res 2021 Apr 7;23(4):e21466. [doi: 10.2196/21466] [Medline: 33825687]
40. Longacre ML, Valdmanis VG, Handorf EA, Fang CY. Work impact and emotional stress among informal caregivers for older adults. J Gerontol B Psychol Sci Soc Sci 2017 May 1;72(3):522-531. [doi: 10.1093/geronb/gbw027] [Medline: 27048567]
41. Wong AKC, Bayuo J, Wong FKY, Chow KKS, Wong SM, Lau ACK. The synergistic effect of nurse proactive phone calls with an mHealth app program on sustaining app usage: 3-arm randomized controlled trial. J Med Internet Res 2023 May 1;25:e43678. [doi: 10.2196/43678] [Medline: 37126378]
42. Chakraborty R, Jana A, Vibhute VM. Caregiving: a risk factor of poor health and depression among informal caregivers in India- a comparative analysis. BMC Public Health 2023 Jan 6;23(1):42. [doi: 10.1186/s12889-022-14880-5] [Medline: 36609312]
43. Cai Y, Simons A, Toland S, Zhang J, Zheng K. Informal caregivers' quality of life and management strategies following the transformation of their cancer caregiving role: a qualitative systematic review. Int J Nurs Sci 2021 Apr 10;8(2):227-236. [doi: 10.1016/j.ijnss.2021.03.006] [Medline: 33997139]

Abbreviations

CONSORT: CONsolidated Standards Of Reporting Trials
COPE: Creativity, Optimism, Planning, and Expert Information

Edited by L Buis; submitted 02.Jan.2025; peer-reviewed by J Kim, R Fang, W Tang; revised version received 16.Oct.2025; accepted 16.Oct.2025; published 07.Apr.2026.

Please cite as:
Wong AKC, Bayuo J, Ng NP, Yau M, Yu KKS, Lin RSY, Su JJ, Hui V, Montayre J
Effectiveness of a Telecare-Based Intervention Program in Supporting Informal Caregivers of Community-Dwelling Older Adults in Reducing Stress Levels: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e70791
URL: https://mhealth.jmir.org/2026/1/e70791
doi:10.2196/70791

© Arkers Kwan Ching Wong, Jonathan Bayuo, Nga Ping Ng, Matthew Yau, Ka Kit Simon Yu, Rose Sin Yi Lin, Jing Jing Su, Vivian Hui, Jed Montayre. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 7.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Role of Videoconferencing Teleconsultation in Improving Transfer Efficiency and Functional Outcomes in Rural Stroke Care: Retrospective Cohort Study

Chi Sheng Wang[1,2,3], MD; Yi-Ju Chen[2,4], MD, PhD; Tzu-Chieh Lin[5,6], MD, PhD; Hui-Mei Huang[2], MS; Pei-Ru Tu[2], MS; Po-Lin Chen[1,3,7*], MD; Jin-An Huang[1*], MD, PhD

[1]Division of Neurology, Neurological Institute, Taichung Veterans General Hospital, No. 1650, Taiwan Boulevard, Sec. 4, Taichung City, Taichung, Taiwan

[2]Telehealth Center, Taichung Veterans General Hospital, Taichung, Taiwan

[3]Institute of Brain Science, College of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan

[4]Graduate Institute of Clinical Medicine, National Chung Hsing University, Taichung, Taiwan

[5]Department of Emergency Medicine, Taichung Veterans General Hospital, Taichung, Taiwan

[6]College of Fine Arts and Creative Design, Tunghai University, Taichung, Taiwan

[7]Department of Post-Baccalaureate Medicine, College of Medicine, National Chung Hsing University, Taichung, Taiwan

[*]these authors contributed equally

Corresponding Author:
Po-Lin Chen, MD
Division of Neurology, Neurological Institute, Taichung Veterans General Hospital, No. 1650, Taiwan Boulevard, Sec. 4, Taichung City, Taichung, Taiwan

Abstract

Background: Interhospital transfer delays remain a major barrier to timely reperfusion therapy and are associated with worse functional outcomes in acute ischemic stroke (AIS), particularly in rural regions.

Objective: This study evaluated whether videoconferencing teleconsultation, compared with the standard referral process, was associated with improved transfer efficiency, treatment delivery, and functional outcomes for patients with AIS requiring interhospital transfer in a hub-and-spoke model.

Methods: We conducted a retrospective cohort study of patients with AIS identified as potential candidates for endovascular thrombectomy (EVT) who were transferred from a primary stroke center (PSC) to a comprehensive stroke center (CSC) between January 2022 and December 2024. Patients were managed using either videoconferencing teleconsultation or the standard referral process, defined as telephone-based consultation between emergency physicians at the PSC and CSC, in which clinical evaluation and thrombolysis decisions were made primarily by the PSC emergency physicians. Group allocation was determined via institutional workflow. The primary outcome was door-in-door-out time, with additional analyses on its components. Secondary outcomes included intravenous thrombolysis rate at the PSC, EVT rates at the CSC, door-to-puncture time, reperfusion rates, and 90-day functional outcomes assessed via modified Rankin Scale shift analysis. Safety outcomes included all-cause mortality within 90 days and symptomatic intracranial hemorrhage after intravenous thrombolysis and/or EVT.

Results: A total of 83 patients were included, with 41 (49.4%) in the teleconsultation group and 42 (50.6%) in the standard referral process group (mean age 73.3, SD 12.9 years), and baseline characteristics were comparable. Teleconsultation was associated with a significant reduction in door-in-door-out time (mean 95.2, SD 22.9 vs 132.3, SD 41.5 minutes; $P<.001$) by shortening computed tomography angiography–to-ambulance notification time (mean 44.6, SD 17.4 vs 79.5, SD 37.6 minutes; $P<.001$). The teleconsultation group had higher intravenous thrombolysis rates at the PSC (26/41, 63.4% in the teleconsultation group vs 17/42, 40.5% in the standard referral process group; $P=.04$), higher EVT rates (14/41, 34.1% in the teleconsultation group vs 6/42, 14.3% in the standard referral process group; $P=.03$), and shorter door-to-puncture time (mean 83.0, SD 35.5 vs 118.5, SD 25.9 minutes; $P=.04$) at the CSC. Patients who received teleconsultation demonstrated a greater shift toward better functional outcomes at the 90th day (27/41, 65.9%; odds ratio 4.55, 95% CI 1.96-11.11; $P<.001$) than patients who did not (13/42, 31.0%; odds ratio 1.35, 95% CI 0.63-2.94; $P=.07$). Safety outcomes were comparable between groups.

Conclusions: Videoconferencing teleconsultation was associated with improved transfer efficiency and higher use of reperfusion therapies and was potentially associated with better functional outcomes. This model may represent a feasible strategy for optimizing stroke care pathways in rural settings. Future studies are warranted to assess its applicability in broader stroke populations beyond conventional EVT eligibility criteria across multicenter networks.

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2026;14:e86436) doi:10.2196/86436

KEYWORDS

acute ischemic stroke; rural; teleconsultation; videoconferencing; transfer efficiency

Introduction

Acute ischemic stroke (AIS) treatment is highly time sensitive as delays in reperfusion therapies such as intravenous thrombolysis and endovascular thrombectomy (EVT) significantly reduce their efficacy [1]. The hub-and-spoke model, widely used in AIS management, connects primary stroke centers (PSCs) with comprehensive stroke centers (CSCs) to provide timely access to advanced care [2]. However, this model often encounters delays, particularly at the PSC level. Current guidelines recommend door-in-door-out (DIDO) times within 120 minutes, but many institutions struggle to meet this target [3,4]. These delays significantly impact EVT initiation and worsen clinical outcomes [5].

In rural regions lacking stroke neurologists, telestroke networks have emerged as a promising solution by facilitating remote neurologic expertise [3]. However, despite prior studies demonstrating improved thrombolysis rates with teleconsultation [6,7], its effects on components of DIDO time and long-term functional outcomes remain less understood. Moreover, existing research has largely focused on telephone-based consultation, with limited integration of videoconferencing-enabled real-time assessment, including clinical evaluation, supervision of thrombolysis administration, imaging interpretation, and treatment planning, which may further optimize stroke-related decision-making and enhance interhospital coordination [8,9].

We hypothesized that videoconferencing teleconsultation would shorten DIDO time, optimize stroke-related decision-making, and improve long-term functional outcomes in patients with AIS transferred from a PSC to a CSC.

Methods

Study Design and Participants

This retrospective cohort study (December 2024) included patients with AIS identified as potential EVT candidates transferred from a PSC in a rural region of central Taiwan (Puli Branch of Taichung Veterans General Hospital) to a CSC (Taichung Veterans General Hospital). Patients were categorized into teleconsultation and standard referral process groups based on whether real-time videoconferencing for neurologic consultation was used. Demographic characteristics and ancillary tests were assessed upon CSC arrival; the National Institutes of Health Stroke Scale (NIHSS) and modified Rankin Scale (mRS) at presentation to the PSC were scored by emergency physicians in the standard referral process group or by neurologists in the teleconsultation group at the PSC.

The primary outcome was DIDO time, defined as the duration from the patient's arrival at the PSC to their departure. To identify specific areas of improvement, DIDO time was subdivided into 3 components, as described in a previously validated study [10]: door to computed tomography angiography (CTA; time from arrival to completion of CTA), CTA to notification (time from completion of CTA to notifying the CSC), and notification to departure (time from notification to ambulance departure). Secondary outcomes included intravenous thrombolysis (tissue-type plasminogen activator [tPA], the only thrombolytic agent in our stroke network) rates at the PSC; EVT rates at the CSC; EVT-related factors at the CSC, including door-to-puncture time, reperfusion rate (defined by a modified treatment in cerebral infarction score of >2b), and rate of tPA administration prior to EVT; and 90-day functional outcomes assessed via the mRS. The safety outcomes included all-cause mortality within 90 days and symptomatic intracranial hemorrhage (sICH) after tPA and/or EVT, defined via follow-up brain computed tomography performed 22 to 36 hours after tPA and/or EVT, which was validated by a previous study [11].

Ethical Considerations

This retrospective cohort study was approved by the Institutional Review Board of Taichung Veterans General Hospital (CE25005B), which granted a waiver of informed consent due to the retrospective nature of the study and the use of existing clinical data as the study posed no more than minimal risk to participants. A deidentified dataset was used for all analyses, and access to the data was restricted to the research team to protect participant privacy and confidentiality. No compensation was provided to study participants. No identifiable participant information or images are included in this manuscript or the supplementary materials.

Study Inclusion and Exclusion Criteria

The inclusion criteria for this study were as follows: (1) patients clinically diagnosed with cerebral infarction by emergency physicians at the PSC at the time of patient arrival and (2) patients identified as potential candidates for EVT by neurologists in the teleconsultation group or by emergency physicians at the PSC in the standard referral process group.

The exclusion criteria were (1) patients diagnosed with hemorrhagic infarction or other stroke mimics based on brain CTA at the PSC; (2) patients exceeding the EVT time window, defined as 6 hours from onset for anterior circulation symptoms or 24 hours from onset for posterior circulation symptoms, as assessed by emergency physicians in the standard referral process group or neurologists in the teleconsultation group based on international guidelines in place at the time [12] and the reimbursement framework of the Taiwan National Health Insurance system; (3) follow-up period of less than 90 days; (4) patients with missing data; and (5) patients not transferred to the CSC via ambulance.

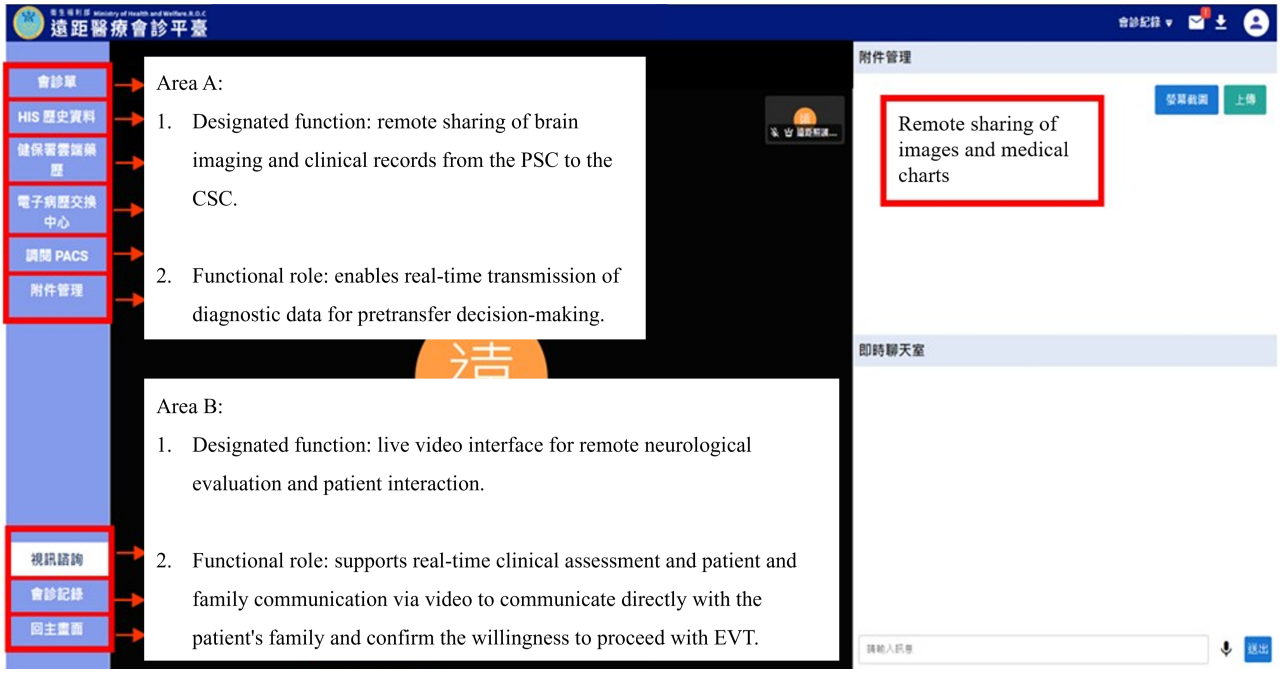
The Process of Teleconsultation and Standard Referral Process

In the teleconsultation group, the videoconferencing process was initiated upon patient arrival at the PSC. Using a 5G-enabled online consultation system, the neurologist at the CSC accessed

XSL•FO
RenderX

shared medical images and charts remotely before or during the brain CTA process. The structural layout and key functional components of the 5G-enabled teleconsultation interface are shown in Figure 1. Following completion of the brain CTA, the neurologist used videoconferencing to directly interact with the patient and their family. This process involved the following six steps: (1) confirming the diagnosis of AIS, (2) assessing the patient and calculating the NIHSS score, (3) verifying the indications and contraindications for intravenous thrombolysis, (4) supervising tPA administration if indicated, (5) communicating directly with the patient's family to explain the indications and risks of EVT, and (6) precoordinating with the neurointerventionalist at the CSC to prepare the operating room after obtaining consent.

Figure 1. The 5G-enabled teleconsultation interface. CSC: comprehensive stroke center; EVT: endovascular thrombectomy; PSC: primary stroke center.



In contrast, the standard referral process involved direct telephone communication between emergency physicians at the PSC and CSC. In this process, the PSC emergency physician independently decided on the administration of tPA, managed the thrombolysis procedure, and determined whether the patient was a potential EVT candidate.

Patients were allocated according to institutional workflow, with the standard referral process used during regular working hours and teleconsultation used during off-hours, including weekday nights and weekends.

Statistical Analyses

Descriptive statistics were reported as means and SDs or medians and IQRs. Group differences were analyzed using the independent 2-tailed *t* test for normally distributed continuous variables, the Mann-Whitney *U* test for nonnormally distributed continuous variables, and the chi-square test for categorical variables. A *P* value of less than .05 was considered statistically significant.

We used independent *t* tests for the primary efficacy outcome, DIDO time and its components. The secondary outcomes, except for the functional outcomes, were analyzed using independent *t* tests or Mann-Whitney *U* tests for continuous variables and chi-square tests for categorical variables. Safety outcomes were compared using the chi-square test. To evaluate functional outcomes, we compared the shift in mRS from presentation to the PSC to the 90th day between groups using proportional odds logistic regression to calculate the odds ratio (OR) for mRS improvement, indicating the likelihood of shifting toward better functional outcomes. The distribution of mRS scores at presentation to the PSC and on the 90th day for each group was visualized using stacked bar charts.

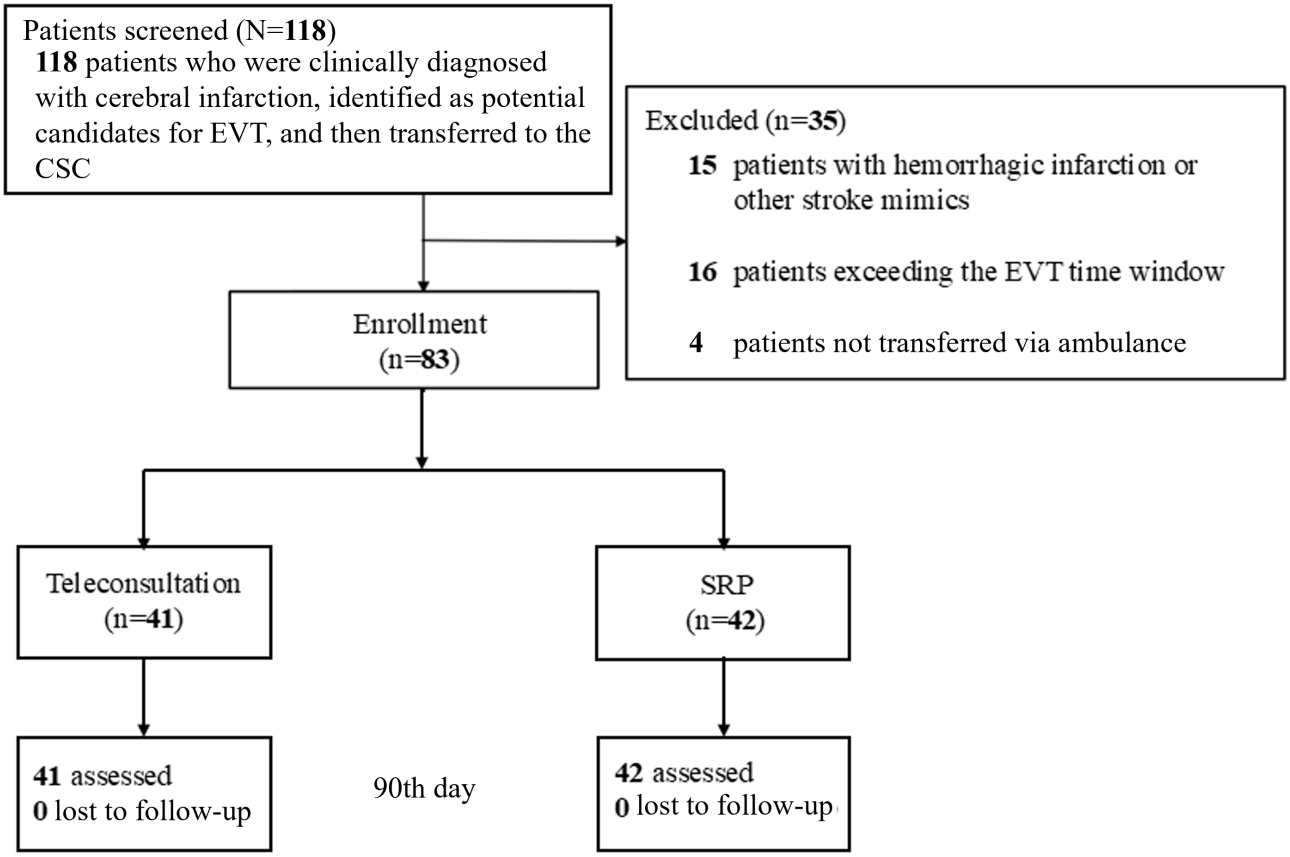
Results

Participants and Baseline Characteristics

A total of 83 patients were enrolled, with 41 (49.4%) in the teleconsultation group and 42 (50.6%) in the standard referral process group (Figure 2). The mean age was 73.3 years, and the baseline characteristics were comparable between the 2 groups. There were no significant differences in gender, comorbidities, smoking, mRS at presentation to the PSC, or NIHSS scores (Table 1).

XSL•FO
RenderX

Figure 2. Study flow diagram. CSC: comprehensive stroke center; EVT: endovascular thrombectomy; SRP: standard referral process.



XSL•FO
RenderX

Table . Baseline characteristics.

	Teleconsultation (n=41)	SRP[a] (n=42)	*P* value
Age (y), mean (SD)	75.2 (12.4)	71.4 (13.5)	.19[b]
Male sex, n (%)	18 (43.9)	23 (54.8)	.32[c]
Hypertension, n (%)	33 (80.5)	31 (73.8)	.47[c]
LDL[d] (mg/dL), mean (SD)	93.5 (31.9)	95.3 (42.7)	.83[b]
Hyperlipidemia, n (%)	21 (51.2)	27 (64.3)	.59[c]
DM[e], n (%)	17 (41.5)	17 (40.5)	.93[c]
HbA_{1c}[f] (mg/dL), mean (SD)	6.6 (1.5)	6.5 (1.3)	.70[b]
Smoking, n (%)	13 (31.7)	15 (35.7)	.70[c]
PAOD[g], n (%)	3 (7.3)	3 (7.1)	.98[c]
NIHSS[h] score (0-42), median (IQR)	15 (10-21)	15 (7-26)	.65[i]
mRS[j] score (0-6) at presentation to the PSC[k], median (IQR)	5 (4-5)	5 (3-5)	.37[i]
AF[l], n (%)	13 (31.7)	14 (33.3)	.88[c]

[a]SRP: standard referral process.
[b]*P* value analyzed using the independent *t* test.
[c]*P* value analyzed using the Mann-Whitney *U* test.
[d]LDL: low-density lipoprotein.
[e]DM: diabetes mellitus.
[f]HbA_{1c}: hemoglobin A_{1c}.
[g]PAOD: peripheral arterial occlusive disease.
[h]NIHSS: National Institutes of Health Stroke Scale.
[i]*P* value analyzed using the chi-square test.
[j]mRS: modified Rankin Scale.
[k]PSC: primary stroke center.
[l]AF: atrial fibrillation.

Primary Outcome

The DIDO time was significantly shorter in the teleconsultation group than in the standard referral process group (mean 95.2, SD 22.9 vs 132.3, SD 41.5 minutes; *P*<.001). Among the DIDO components, the CTA-to-notification time was significantly reduced in the teleconsultation group (mean 44.6, SD 17.4 vs 79.5, SD 37.6 minutes; *P*<.001). Other components, including door-to-CTA time and notification-to-departure time, showed no significant differences except for a borderline significance in the notification-to-departure time (mean 29.2, SD 3.3 vs 30.9, SD 3.9 minutes; *P*=.05; Table 2).

Table . Primary outcome: the door-in-door-out (DIDO) time and the DIDO components between groups.

	Teleconsultation (n=41), mean (SD)	SRP[a] (n=42), mean (SD)	*P* value
DIDO time (min)	95.2 (22.9)	132.3 (41.5)	<.001[b]
Door-to-CTA[c] time (min)	21.2 (5.4)	21.9 (4.2)	.54[b]
CTA-to-notification time (min)	44.6 (17.4)	79.5 (37.6)	<.001[b]
Notification-to-departure time (min)	29.2 (3.3)	30.9 (3.9)	.05[b]

[a]SRP: standard referral process.
[b]*P* value analyzed using the independent *t* test.
[c]CTA: computed tomography angiography.

XSL•FO
RenderX

Secondary Outcomes

For stroke-related decision-making, the rates of tPA at the PSC (26/41, 63.4% vs 17/42, 40.5%; *P*=.04) and EVT at the CSC (14/41, 34.2% vs 6/42, 14.3%; *P*=.03) were higher in the teleconsultation group than in the standard referral process group. Among EVT-related factors in the CSC, the door-to-puncture time was shorter in the teleconsultation group (mean 83.0, SD 35.5 vs 118.5, SD 25.9 minutes; *P*=.04). The rate of tPA administration prior to EVT revealed no significant differences (9/14, 64.3% vs 4/6, 66.7%; *P*=.92). The reperfusion rate was higher in the teleconsultation group (13/14, 92.9% vs 3/6, 50%; *P*=.03; Table 3).

Table . Secondary outcomes and safety outcomes.

	Teleconsultation (n=41)	SRP[a] (n=42)	*P* value
Stroke-related decision-making, n/N (%)			
Intravenous thrombolysis at the PSC[b]	26/41 (63.4)	17/42 (40.5)	.04[c]
EVT[d] at the CSC[e]	14/41 (34.1)	6/42 (14.3)	.03[c]
EVT-related factors			
Door-to-puncture time (min), mean (SD)	83.0 (35.5)	118.5 (25.9)	.04[f]
Intravenous thrombolysis before EVT, n/N (%)	9/14 (64.3)	4/6 (66.7)	.92[c]
Reperfusion rate (mTICI[g] score of >2b), n/N (%)	13/14 (92.9)	3/6 (50)	.03[c]
Safety outcomes, n/N (%)			
All-cause mortality	5/41 (12.2)	7/42 (16.7)	.56[c]
sICH[h] after intravenous thrombolysis and/or EVT	3/31 (9.7)	2/19 (10.5)	.92[c]

[a]SRP: standard referral process.

[b]PSC: primary stroke center.

[c]*P* value analyzed using the Mann-Whitney *U* test.

[d]EVT: endovascular thrombectomy.

[e]CSC: comprehensive stroke center.

[f]*P* value analyzed using the independent *t* test.

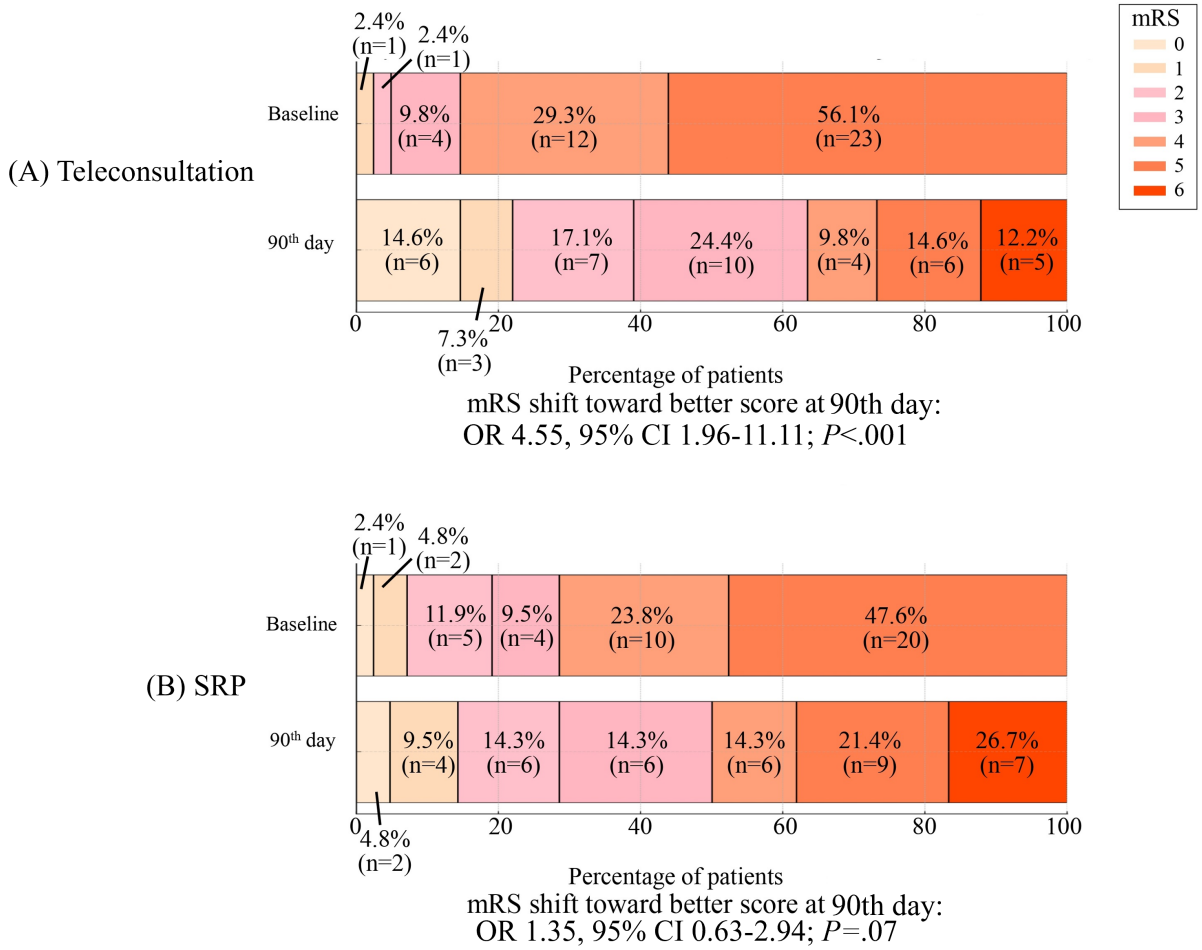
[g]mTICI: modified treatment in cerebral infarction.

[h]sICH: symptomatic intracranial hemorrhage.

Among functional outcomes, at the 90th day, the teleconsultation group demonstrated a greater shift toward better mRS scores (27/41, 65.9% of the participants; OR 4.55, 95% CI 1.96-11.11; *P*<.001) compared to the standard referral process group (13/42, 31.0% of the participants; OR 1.35, 95% CI 0.63-2.94; *P*=.07; Figure 3).

XSL•FO
RenderX

Figure 3. The modified Rankin Scale (mRS) score at presentation to the primary stroke center and the functional outcomes at the 90th day. OR: odds ratio; SRP: standard referral process.



Safety Outcomes

There was no significant difference in all-cause mortality within 90 days between groups (5/41, 12.2% vs 7/42, 16.7%; *P*=.56) or in the incidence of sICH after tPA and/or EVT (3/31, 9.7% vs 2/19, 10.5%; *P*=.92; Table 3).

Discussion

Principal Findings

Our study suggests that videoconferencing teleconsultation is associated with improved interhospital transfer efficiency and functional outcomes for patients with AIS within a hub-and-spoke model. The DIDO time, a critical metric for transfer efficiency, was significantly reduced in the teleconsultation group, driven primarily by a shorter CTA-to-notification time. During the study period, EVT eligibility was determined based on national health insurance reimbursement policies in Taiwan, which adopted extended time windows for anterior circulation stroke only after November 2023. To maintain consistency in inclusion criteria, patients beyond the previous time window were excluded.

Prior research, including a retrospective cohort study in a Dutch ambulance region, identified CTA-to-notification time as the largest contributor to total DIDO time, often exceeding 40% [10]. This delay is primarily attributed to the sequential nature of imaging interpretation, decision-making, and notifying the receiving center. Traditionally, these steps occur in isolation, resulting in cumulative delays before a final transfer decision is made. In our study, the significant reduction in CTA-to-notification time observed in the teleconsultation group may be related to the integration of real-time teleconsultation and 5G-enabled patient assessment, both of which likely contributed to a more streamlined workflow.

First, real-time teleconsultation allowed CSC neurologists to engage in the diagnostic process earlier, either while CTA was being performed or during the interval awaiting the formal CTA report. This parallel approach eliminated the traditional stepwise delays, allowing neurologists to simultaneously interpret imaging results, assess stroke severity, and discuss treatment plans with PSC emergency physicians. The ability to directly support PSC physicians in evaluating tPA eligibility and communicating with patients and their families may have facilitated faster decision-making and contributed to the higher tPA administration rate. This aligns with prior studies demonstrating that teleconsultation improves thrombolysis rates, particularly in facilities lacking on-site neurologists [13,14]. Second, the 5G-enabled patient assessment likely played a role in improving workflow efficiency by integrating real-time video and clinical data sharing. Unlike conventional workflows in which imaging review and patient assessment occur in separate sequential steps, this system enabled CSC neurologists to

XSL•FO
RenderX

conduct simultaneous neurological evaluations and imaging interpretation, ensuring that NIHSS scoring and stroke severity assessments were completed without unnecessary delays [15]. By minimizing gaps between imaging interpretation and transfer notification, decision-making efficiency was likely enhanced [16]. Third, neurologists provided supervision of thrombolysis administration, ensuring protocol adherence and reducing procedural uncertainties, which may have contributed to optimizing treatment delivery and alleviating the workload of PSC physicians [17]. The combined effect of these improvements underscores the potential of videoconferencing teleconsultation in enhancing acute stroke management in resource-limited settings where on-site stroke expertise is unavailable.

Beyond facilitating initial treatment decisions at the PSC, teleconsultation was also associated with higher EVT rates and improved procedural efficiency at the CSC. Traditional telephone-based handovers are often fragmented, leading to delays in aligning treatment plans and preparing intervention teams [18]. In contrast, the video-based teleconsultation model enabled synchronized discussions between PSC and CSC teams, ensuring immediate procedural planning and early mobilization of neurointerventionalists [19]. This streamlined coordination likely contributed to the higher EVT rates, consistent with prior studies showing that pretransfer consultation increases EVT use. Furthermore, the reduction in door-to-puncture time suggests that early procedural activation played a crucial role in optimizing workflow efficiency. Teleconsultation enabled neurologists to assess imaging remotely before patient arrival, confirm EVT candidacy in advance, and initiate angiography suite preparations, leading to a more time-efficient transition. These findings align with evidence suggesting that early procedural planning and pretransfer activation of interventional teams are key determinants in reducing door-to-puncture time [19-21]. The improvement in reperfusion rates likely reflects the combined impact of increased EVT use and shorter procedural delays, which is highly time dependent [1,22].

Although teleconsultation has been well established to improve stroke-related decision-making, its effect on long-term functional outcomes remains controversial [8,23]. Prior studies have shown that hospitals with telestroke capacity achieve higher rates of reperfusion treatment and lower 30-day mortality than hospitals without telestroke capacity in rural settings [23]. As our PSC was located in a rural region, our findings reinforced the role of telestroke networks in expanding access to acute stroke therapies in rural stroke care. However, prior studies have not consistently demonstrated a direct association between telestroke network implementation and long-term functional improvement [5,8,24-27]. Notably, a previous study reported that telestroke network implementation did not significantly alter hospital transfer decisions, suggesting its primary role in improving treatment efficiency rather than hospital selection [28]. However, our current PSC-to-CSC model ensured that all eligible patients were transferred to the same CSC, eliminating confounding effects from variable hospital destinations. This allowed for a more precise evaluation of functional outcomes without interhospital transfer variability.

Our study provides additional evidence suggesting that videoconferencing teleconsultation may be associated with improved functional outcomes. Compared with prior studies that primarily focused on PSC-level decision-making, our study evaluated an integrated workflow spanning both a PSC and a CSC. By incorporating real-time video-based patient assessment, this model enables simultaneous clinical evaluation, imaging interpretation, and treatment planning. In addition to improving transfer efficiency, the observed association between workflow optimization and functional outcomes suggests a potential downstream clinical benefit of teleconsultation beyond conventional process measures. These findings are consistent with those of the TEMPiS (Telemedical Project for Integrative Stroke Care) trial [25], which demonstrated that telestroke networks were associated with a lower proportion of poor outcomes, reinforcing the benefits of telemedicine in acute stroke management. However, while prior research has primarily emphasized reducing severe disability and mortality, our study further explored functional outcome improvements. Taken together, our findings suggest that the potential benefit of teleconsultation may extend beyond isolated improvements in treatment rates to encompass system-wide optimization of stroke care pathways. Future research should investigate whether similar benefits can be observed in more diverse stroke networks.

Notably, while this study used a single hub–and–single spoke design, the streamlined and well-defined teleconsultation workflow described herein suggests that the model is highly feasible and potentially scalable. We have provided a detailed account of the process, encompassing early identification at the PSC, interhospital communication, and pretransfer decision-making. Given that implementation requires only a stable internet connection, standard videoconferencing tools, and a dedicated stroke team on call, this approach could be adapted in other underserved regions aiming to strengthen acute stroke care delivery through telemedicine.

Safety outcomes, including rates of sICH and all-cause mortality, were comparable between groups, supporting the safety of integrating teleconsultation into stroke care. In addition, the comparable rates of tPA administration prior to EVT suggest that the process does not compromise standard protocols but, instead, enhances patient selection and coordination for advanced therapies, which is in line with prior studies demonstrating that remote supervision of intravenous tPA via telemedicine is both feasible and safe [29].

Importantly, beyond clinical outcome measures, our study highlights the role of videoconferencing teleconsultation as a system-level digital health intervention that enables real-time, synchronized decision-making across geographically separated care settings. Unlike conventional telestroke models that primarily support decision-making at the PSC, our approach integrates pretransfer evaluation, interhospital communication, and procedural preparation at the CSC into a continuous workflow. This end-to-end coordination represents a key advancement in telemedicine-enabled stroke care, shifting from isolated decision support toward comprehensive workflow integration. Such a model may be particularly relevant for rural

XSL•FO
RenderX

and resource-limited regions, where timely access to specialized stroke care remains a major challenge.

The strengths of this study include a detailed analysis of DIDO time, allowing for precise identification of workflow inefficiencies. The integration of real-time videoconferencing facilitated real-time patient assessment, ensuring direct NIHSS scoring and verification of tPA eligibility. Furthermore, the PSC in our study serves as the primary stroke referral center for multiple rural townships in central Taiwan, including Puli, Guoxing, Ren'ai, Yuchi, and Zhongliao, with a total service population of approximately 150,000. Our findings may offer applicable insights for enhancing rural stroke care.

However, several limitations should be considered. First, the relatively small sample size limits the statistical power of this study and may affect the precision of effect estimates, particularly for secondary and safety outcomes. As a result, this study may be underpowered to detect modest differences between groups, and the observed associations, especially regarding functional outcomes, should be interpreted with caution. Second, the single hub–and–single spoke design, chosen to minimize variability in transfer times by ensuring that all patients were transported via ambulance, may further limit the generalizability of our findings to more complex, multicenter stroke networks. Third, the retrospective design might introduce potential selection bias as treatment decisions were based on real-world clinical practice rather than on randomized allocation. Fourth, the inclusion criteria were aligned with the reimbursement policy of the Taiwan National Health Insurance system, which covers EVT within 6 hours for anterior circulation stroke and within 24 hours for posterior circulation stroke. This policy-driven framework may have resulted in the inclusion of patients who were transferred but ultimately did not undergo EVT. In addition, group allocation was determined by institutional workflow, which may introduce selection bias. Notably, workflow efficiency remained improved in the teleconsultation group despite being implemented during off-hours.

Conclusions

Videoconferencing teleconsultation was associated with improved interhospital transfer efficiency, facilitated timely reperfusion therapy, and may be associated with better functional outcomes in patients with AIS within a hub-and-spoke model while maintaining patient safety. Future research should evaluate this model's scalability in multicenter networks and its impact on broader stroke populations beyond standard EVT eligibility criteria.

Acknowledgments

The authors appreciate all the participants in this study and the superintendents of Taichung Veterans General Hospital, Professor Shih-Ann Chen and Professor Yun-Ching Fu, for establishing the Telehealth Center, whose advanced insights contributed to this manuscript. The authors also sincerely thank Professor I-Hui Lee, the mentor of the first author (CSW), for her valuable guidance in study design.

Funding

This research did not receive any specific grants from public, commercial, or not-for-profit funding agencies.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: CSW (lead), PLC (equal), YJC (supporting), TCL (supporting), HMH (supporting), JAH (supporting)
Data curation: YJC (lead), CSW (equal), PLC (equal), HMH (supporting), PRT (supporting), JAH (supporting)
Formal analysis: CSW (lead), YJC (supporting)
Funding acquisition: PLC (lead), YJC (equal), JAH (supporting)
Investigation: PLC (lead), JAH (equal), CSW (supporting), YJC (supporting), TCL (supporting), HMH (supporting)
Methodology: CSW (lead), YJC (supporting)
Project administration: JAH (lead), PLC (equal)
Resources: PLC (lead), PRT (supporting), JAH (supporting)
Supervision: PLC (lead), JAH (equal)
Validation: CSW (lead), YJC (supporting)
Visualization: CSW (lead)
Writing—original draft: CSW (lead)
Writing—review and editing: CSW (lead), PLC (supporting)

Conflicts of Interest

None declared.

References

XSL•FO
RenderX

1. Tsivgoulis G, Saqqur M, Sharma VK, et al. Timing of recanalization and functional recovery in acute ischemic stroke. J Stroke 2020 Jan;22(1):130-140. [doi: 10.5853/jos.2019.01648] [Medline: 32027798]
2. Morris S, Ramsay AI, Boaden RJ, et al. Impact and sustainability of centralising acute stroke services in English metropolitan areas: retrospective analysis of hospital episode statistics and stroke national audit data. BMJ 2019 Jan 23;364:l1. [doi: 10.1136/bmj.l1] [Medline: 30674465]
3. Stamm B, Royan R, Giurcanu M, Messe SR, Jauch EC, Prabhakaran S. Door-in-door-out times for interhospital transfer of patients with stroke. JAMA 2023 Aug 15;330(7):636-649. [doi: 10.1001/jama.2023.12739] [Medline: 37581671]
4. Alberts MJ, Wechsler LR, Jensen ME, et al. Formation and function of acute stroke-ready hospitals within a stroke system of care recommendations from the Brain Attack Coalition. Stroke 2013 Dec;44(12):3382-3393. [doi: 10.1161/STROKEAHA.113.002285] [Medline: 24222046]
5. Royan R, Stamm B, Giurcanu M, et al. Door-in-door-out times and outcomes in patients with acute ischaemic stroke transferred for endovascular therapy in the USA: a retrospective cohort study. Lancet Neurol 2026 Feb;25(2):160-169. [doi: 10.1016/S1474-4422(25)00478-8] [Medline: 41579902]
6. Demaerschalk BM, Berg J, Chong BW, et al. American Telemedicine Association: telestroke guidelines. Telemed J E Health 2017 May;23(5):376-389. [doi: 10.1089/tmj.2017.0006] [Medline: 28384077]
7. Worthmann H, Winzer S, Schuppner R, Gumbinger C, Barlinn J. Telestroke networks for area-wide access to endovascular stroke treatment. Neurol Res Pract 2023 Mar 3;5(1):9. [doi: 10.1186/s42466-023-00237-9] [Medline: 36864498]
8. Mohamed A, Elsherif S, Legere B, Fatima N, Shuaib A, Saqqur M. Is telestroke more effective than conventional treatment for acute ischemic stroke? A systematic review and meta-analysis of patient outcomes and thrombolysis rates. Int J Stroke 2024 Mar;19(3):280-292. [doi: 10.1177/17474930231206066] [Medline: 37752674]
9. Bergrath S, Reich A, Rossaint R, et al. Feasibility of prehospital teleconsultation in acute stroke--a pilot study in clinical routine. PLoS One 2012;7(5):e36796. [doi: 10.1371/journal.pone.0036796] [Medline: 22629331]
10. van de Wijdeven RM, Duvekot MH, van der Geest PJ, et al. Determinants of door-in-door-out time in patients with ischaemic stroke transferred for endovascular thrombectomy. Eur Stroke J 2023 Sep;8(3):667-674. [doi: 10.1177/23969873231177768] [Medline: 37248995]
11. Mazya M, Egido JA, Ford GA, et al. Predicting the risk of symptomatic intracerebral hemorrhage in ischemic stroke treated with intravenous alteplase: Safe Implementation of Treatments in Stroke (SITS) symptomatic intracerebral hemorrhage risk score. Stroke 2012 Jun;43(6):1524-1531. [doi: 10.1161/STROKEAHA.111.644815] [Medline: 22442178]
12. Powers WJ, Rabinstein AA, Ackerson T, et al. 2018 Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2018 Mar;49(3):e46-e110. [doi: 10.1161/STR.0000000000000158] [Medline: 29367334]
13. Janßen A, Pardey N, Zeidler J, et al. Support by telestroke networks is associated with increased intravenous thrombolysis and reduced hospital transfers: a German claims data analysis. Health Econ Rev 2024 Nov 28;14(1):100. [doi: 10.1186/s13561-024-00577-5] [Medline: 39604598]
14. Rubin MN, Demaerschalk BM. The use of telemedicine in the management of acute stroke. Neurosurg Focus 2014 Jan;36(1):E4. [doi: 10.3171/2013.11.FOCUS13428] [Medline: 24380481]
15. Weiss D, Rubbert C, Kaschner M, et al. Prehospital telemedicine support for urban stroke care: analysis of current state of care and conceptualization. BMC Emerg Med 2024 Nov 27;24(1):224. [doi: 10.1186/s12873-024-01142-3] [Medline: 39604854]
16. Bladin CF, Cadilhac DA. Effect of telestroke on emergent stroke care and stroke outcomes. Stroke 2014 Jun;45(6):1876-1880. [doi: 10.1161/STROKEAHA.114.003825] [Medline: 24763931]
17. Wechsler LR, Demaerschalk BM, Schwamm LH, et al. Telemedicine quality and outcomes in stroke: a scientific statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2017 Jan;48(1):e3-e25. [doi: 10.1161/STR.0000000000000114] [Medline: 27811332]
18. Solet DJ, Norvell JM, Rutan GH, Frankel RM. Lost in translation: challenges and opportunities in physician-to-physician communication during patient handoffs. Acad Med 2005 Dec;80(12):1094-1099. [doi: 10.1097/00001888-200512000-00005] [Medline: 16306279]
19. Yang S, Yao W, Siegler JE, et al. Shortening door-to-puncture time and improving patient outcome with workflow optimization in patients with acute ischemic stroke associated with large vessel occlusion. BMC Emerg Med 2022 Jul 26;22(1):136. [doi: 10.1186/s12873-022-00692-8] [Medline: 35883030]
20. Mehta BP, Leslie-Mazwi TM, Chandra RV, et al. Reducing door-to-puncture times for intra-arterial stroke therapy: a pilot quality improvement project. J Am Heart Assoc 2014 Nov 11;3(6):e000963. [doi: 10.1161/JAHA.114.000963] [Medline: 25389281]
21. Kim DH, Kim B, Jung C, et al. Consensus statements by Korean Society of Interventional Neuroradiology and Korean Stroke Society: hyperacute endovascular treatment workflow to reduce door-to-reperfusion time. Korean J Radiol 2018;19(5):838-848. [doi: 10.3348/kjr.2018.19.5.838] [Medline: 30174472]
22. Saver JL. Time is brain--quantified. Stroke 2006 Jan;37(1):263-266. [doi: 10.1161/01.STR.0000196957.55928.ab] [Medline: 16339467]

XSL•FO
RenderX

23. Wilcock AD, Schwamm LH, Zubizarreta JR, et al. Reperfusion treatment and stroke outcomes in hospitals with telestroke capacity. JAMA Neurol 2021 May 1;78(5):527-535. [doi: 10.1001/jamaneurol.2021.0023] [Medline: 33646272]
24. Almallouhi E, Holmstedt CA, Harvey J, et al. Long-term functional outcome of telestroke patients treated under drip-and-stay paradigm compared with patients treated in a comprehensive stroke center: a single center experience. Telemed J E Health 2019 Aug;25(8):724-729. [doi: 10.1089/tmj.2018.0137] [Medline: 30256734]
25. Audebert HJ, Schenkel J, Heuschmann PU, Bogdahn U, Haberl RL, Telemedic Pilot Project for Integrative Stroke Care Group. Effects of the implementation of a telemedical stroke network: the Telemedic Pilot Project for Integrative Stroke Care (TEMPiS) in Bavaria, Germany. Lancet Neurol 2006 Sep;5(9):742-748. [doi: 10.1016/S1474-4422(06)70527-0] [Medline: 16914402]
26. Johansson T, Mutzenbach SJ, Ladurner G. Telemedicine in acute stroke care: the TESSA model. J Telemed Telecare 2011;17(5):268-272. [doi: 10.1258/jtt.2011.101213] [Medline: 21824968]
27. Dutta D, Kendall J, Holmes C, et al. Evaluation of a telephone advice system for remote intravenous thrombolysis in ischemic stroke: data from a United Kingdom network. Stroke 2015 Mar;46(3):867-869. [doi: 10.1161/STROKEAHA.114.008190] [Medline: 25604248]
28. Zachrison KS, Richard JV, Wilcock A, et al. Association of hospital telestroke adoption with changes in initial hospital presentation and transfers among patients with stroke and transient ischemic attacks. JAMA Netw Open 2021 Sep 1;4(9):e2126612. [doi: 10.1001/jamanetworkopen.2021.26612] [Medline: 34554236]
29. Pervez MA, Silva G, Masrur S, et al. Remote supervision of IV-tPA for acute ischemic stroke by telemedicine or telephone before transfer to a regional stroke center is feasible and safe. Stroke 2010 Jan;41(1):e18-e24. [doi: 10.1161/STROKEAHA.109.560169] [Medline: 19910552]

Abbreviations

AIS: acute ischemic stroke
CSC: comprehensive stroke center
CTA: computed tomography angiography
DIDO: door-in-door-out
EVT: endovascular thrombectomy
mRS: modified Rankin Scale
NIHSS: National Institutes of Health Stroke Scale
OR: odds ratio
PSC: primary stroke center
sICH: symptomatic intracranial hemorrhage
TEMPiS: Telemedical Project for Integrative Stroke Care
tPA: tissue-type plasminogen activator

Edited by L Buis; submitted 24.Oct.2025; peer-reviewed by B Stamm, DL Isenberg; revised version received 20.May.2026; accepted 21.May.2026; published 04.Jun.2026.

Please cite as:
Wang CS, Chen YJ, Lin TC, Huang HM, Tu PR, Chen PL, Huang JA
The Role of Videoconferencing Teleconsultation in Improving Transfer Efficiency and Functional Outcomes in Rural Stroke Care: Retrospective Cohort Study
JMIR Mhealth Uhealth 2026;14:e86436
URL: https://mhealth.jmir.org/2026/1/e86436
doi:10.2196/86436

© Chi Sheng Wang, Yi-Ju Chen, Tzu-Chieh Lin, Hui-Mei Huang, Pei-Ru Tu, Po-Lin Chen, Jin-An Huang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 4.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Influence of Online Sessions via a Deep Brain Stimulation Device: Prospective, Single-Arm, Longitudinal, Nonrandomized Self-Controlled Cohort Study

Ryoma Morigaki[1,2,3], MD, PhD, Prof Dr Med; Hideo Mure[4,5], MD, PhD; Kouki Takasuka[4]; Joji Fujikawa[1], PhD; Toshiro Shinmen[4]; Shin Tarui[4]; Masataka Hatanaka[4]; Keisuke Shinohara[1], PhD; Taku Matsuda[3], MD, PhD; Kazuhisa Miyake[3], MD, PhD; Yasushi Takagi[3], MD, PhD, Prof Dr Med

[1]Department of Advanced Brain Research, Graduate School of Biomedical Sciences, Tokushima University, 3-18-15 Kuramoto-chou, Tokushima, Japan

[2]Parkinson's Disease and Dystonia Research Center, Tokushima University Hospital, Tokushima, Japan

[3]Department of Neurosurgery, Graduate School of Biomedical Sciences, Tokushima University, Tokushima, Japan

[4]Center for Neuromodulation, Kurashiki Heisei Hospital, Kurashiki, Japan

[5]Department of Neurological Surgery, Kurashiki Heisei Hospital, Kurashiki, Japan

Corresponding Author:
Ryoma Morigaki, MD, PhD, Prof Dr Med
Department of Advanced Brain Research, Graduate School of Biomedical Sciences, Tokushima University, 3-18-15 Kuramoto-chou, Tokushima, Japan

Abstract

Background: Deep brain stimulation (DBS) is widely performed in patients with advanced Parkinson disease (PD). Recent advances in technology have facilitated remote programming of DBS devices, reflecting an emerging trend in neuromodulation approaches, and offering a potential framework for patient-centered care. These online sessions for patients with PD who underwent de novo implantation of DBS devices have been reported to be safe and effective, similar to in-clinic sessions. Currently, evidence for patients with chronically implanted DBS devices remains limited.

Objective: This study aimed to evaluate the safety, feasibility, and preliminary efficacy of online sessions for patients with PD who had chronically implanted DBS devices, and to determine whether this approach could reduce patient burden without compromising clinical management.

Methods: A prospective, single-arm, longitudinal, nonrandomized, self-controlled cohort study was conducted at 2 centers in Japan. Eighteen patients with PD who had chronically implanted DBS devices were enrolled. Two in-clinic sessions were substituted with online sessions. The Movement Disorder Society Unified Parkinson's Disease Rating Scale Part III scores were assessed at baseline, after in-clinic sessions, and online sessions. The Multidimensional Evaluation Scale for Patient Impression Change—Japanese version, Patient's Global Impression of Change, and Clinical Global Impression of Change were each administered after both in-clinic and online sessions. The Telehealth Usability Questionnaire and the Online Usability Questionnaire for Programming were evaluated once after online sessions.

Results: The Movement Disorder Society Unified Parkinson's Disease Rating Scale Part III scores did not differ significantly across the 3 assessments. The Multidimensional Evaluation Scale for Patient Impression Change—Japanese version scores revealed that no items but item 3, "overall sleep," significantly worsened after online sessions. The Patient's Global Impression of Change and Clinical Global Impression of Change scores were comparable between in-clinic and online sessions. The Telehealth Usability Questionnaire scores indicated that online sessions were generally favorable, although the "Reliability" was rated comparatively lower. The mean total time and cost saved per visit were 280 minutes and 7974 yen (US $1≈155 JPY), respectively. One patient experienced a fall and lumbar compression fracture during the online visit period and dropped out of the study because of hospitalization.

Conclusions: Online sessions may be a feasible option for a subset of patients with PD carrying chronically implanted DBS devices. Although online sessions cannot fully replace in-person assessments, they may serve as a practical alternative when adequate caregiver support, sufficient digital literacy, and a reliable connection are ensured. Clinicians should note that sleep-related symptoms may worsen in some patients. While time and cost savings are not direct indicators of care quality, these benefits still provide meaningful support for patients and caregivers. Overall, online sessions may complement routine follow-up in stable patients undergoing chronic DBS therapy.

Trial Registration: Japan Registry of Clinical Trials jRCT1062230055; https://jrct.mhlw.go.jp/latest-detail/jRCT1062230055

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2026;14:e80223) doi:10.2196/80223

KEYWORDS

Parkinson disease; telemedicine; deep brain stimulation; remote consultation; motor disorders; global burden of disease; caregiver burden

Introduction

Online Sessions as an Emerging Modality in Parkinson Disease Care

Parkinson disease (PD) is a multisystem neurodegenerative disorder characterized by motor and nonmotor symptoms [1]. Due to its progressive nature, the burden on patients and their caregivers increases over time. Recent studies have demonstrated the effectiveness of online sessions including remotely supervised exercise and telerehabilitation for patients with PD [2-6]. The advancement and prevalence of remote technologies, in part owing to the COVID-19 pandemic, have promoted remote programming and surveillance in patients who have undergone deep brain stimulation (DBS) [7-18]. Several studies have shown that online sessions are highly effective, particularly in reducing the cost and time burden for patients with PD who underwent DBS, as well as their caregivers [8-11,15-18].

Current Evidence on Online Sessions for DBS

In patients with de novo DBS, 2 randomized studies and several observational studies have shown that remote programming is safe, effective, and comparable with in-clinic adjustments, with high satisfaction rates [7,11,12,15,16,18]. These findings suggest that online sessions can support early postoperative management and reduce the need for frequent hospital visits. In contrast, evidence in patients who had chronically implanted DBS devices is more limited. Retrospective and cross-sectional studies indicate that remote programming may reduce travel distance, time, and financial burden, and that patient satisfaction is generally favorable [8,10,13,14,17]. However, these studies are heterogeneous, mostly observational, and lack prospective evaluation.

Rationale and Objective for This Study

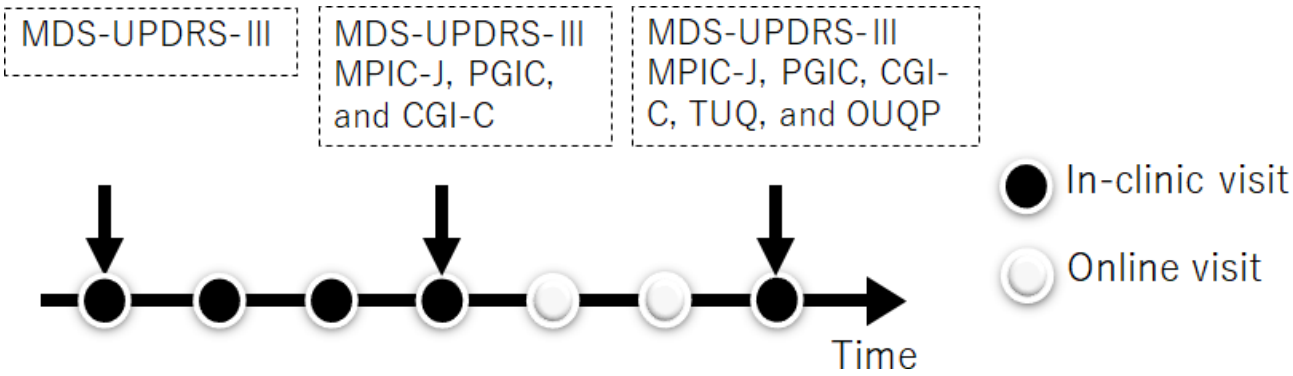
The clinical needs of patients who had undergone chronic DBS differ from those of de novo cases, as long-term follow-up typically involves clinical review, medication management, and device checks rather than frequent parameter adjustments. Thus, despite accumulating evidence, prospective, population-specific data on the safety and feasibility of online sessions in chronic DBS care remain insufficient. To date, no prospective study has evaluated online sessions in patients with PD who underwent chronic implantation of DBS devices, and detailed data are still lacking. Given these gaps, we aimed to explore whether online sessions could be implemented safely and feasibly in patients with chronic DBS devices, while assessing their preliminary efficacy in maintaining clinical stability and reducing patient burden—particularly travel-related demands—without compromising routine clinical management.

Methods

Study Design

A prospective, single-arm, longitudinal, multicenter, nonblinded, nonrandomized, and self-controlled cohort trial was conducted at Tokushima University Hospital and Kurashiki Heisei Hospital in Japan between 2023 and 2024. All participants had been treated with in-clinic sessions, followed by 2 consecutive online sessions (Figure 1).

Figure 1. Schematic figure explaining the planning of the experiment. CGI-C: Clinical Global Impression of Change; MDS-UPDRS-III: Movement Disorder Society Unified Parkinson's Disease Rating Scale Part III; MPIC-J: Multidimensional Evaluation Scale for Patient Impression Change—Japanese Version; OUQP: Online Usability Questionnaire for Programming; PGIC: Patient's Global Impression of Change; TUQ: Telehealth Usability Questionnaire.



Ethical Considerations

This study was approved by the Ethics Committee of Tokushima University (approval number 4409, date of approval August 28, 2023) and the Ethics Committee of Kurashiki Heisei Hospital (approval number R05-008, date of approval July 4, 2023) in accordance with the Declaration of Helsinki. The participants received detailed study information and signed informed consent forms before enrollment. To protect participant privacy, all data were deidentified, stored on secure institutional servers, and accessible only to authorized study personnel. No audio or video recordings were collected during online sessions, and all telehealth communications were conducted through encrypted,

XSL•FO
RenderX

institution-approved platforms. Participants did not receive financial compensation for study participation.

Participants

All participants met the following inclusion criteria: (1) adult patients (aged 18 years and older) diagnosed with PD, (2) those who had received bilateral subthalamic nucleus (STN) DBS using the Infinity DBS system with the NeuroSphere Virtual Clinic remote care (Abbott), (3) those who had a follow-up period of at least 3 years postsurgery, (4) those who had never experienced online sessions, and (5) those who are able to operate the online system with or without caregivers and consented to be included in this study. For the sample size calculation, we estimated that 15 participants would provide 80% power (with a 5% probability of type I error) and a large effect size using G*Power 3.1. Therefore, we set the target sample size to at least 17 participants and estimated a dropout rate of 10%.

Assessments

Primary Outcomes

Clinical Motor Assessments

The Movement Disorder Society Unified Parkinson's Disease Rating Scale Part III (MDS-UPDRS-III) scores were recorded before and after each of the 2 in-clinic sessions and 2 online sessions, using identical in-person, face-to-face assessment protocols conducted by the same evaluators with active DBS stimulation and on medications (Figure 1).

Safety Assessments

Safety was evaluated throughout the study period by monitoring adverse events. Participants were reported any new or worsening medical conditions during both in-clinic and online sessions. All adverse events, including falls, injuries, or any clinical deterioration, were recorded by the investigators, regardless of their perceived relationship to the online sessions.

Secondary Outcomes

Patient-Reported Assessments

Three self-reported numerical rating scales, including the Multidimensional Evaluation Scale for Patient Impression Change—Japanese version (MPIC-J) [19,20], Patient's Global Impression of Change (PGIC) [21], and Clinical Global Impression of Change (CGI-C) [22,23], were obtained before and after the 2 online sessions (Figure 1). Two self-reported numerical rating scales, the Telehealth Usability Questionnaire (TUQ) [24] and Online Usability Questionnaire for Programming (OUQP), were evaluated after the 2 online sessions (Figure 1).

The MPIC assesses 8 domains (overall status, pain, sleep, mood, physical functioning, coping with pain, managing pain flare-ups, and medication effectiveness) with 7 ratings: 1 = very much improved, 2 = much improved, 3 = minimally improved, 4 = no change, 5 = minimally worse, 6 = much worse, and 7 = very much worse [19]. PGIC was rated on the 7-point scale: 1 = no change (or condition has got worse); 2 = almost the same, hardly any change at all; 3 = a little better but no noticeable change; 4 = somewhat better, but the change has not made any real difference; 5 = moderately better, and a slight but noticeable change; 6 = better, and a definite improvement that has made a real and worthwhile difference; and 7 = a great deal better, and a considerable improvement that has made all the difference [25]. The CGI-C was also rated on the 7-point scale: 1 = very much improved, 2 = much improved, 3 = minimally improved, 4 = no change from baseline (the initiation of treatment), 5 = minimally worse, 6 = much worse, and 7 = very much worse since the initiation of treatment [22,23]. The TUQ has 6 subdomains, with each consisting of 3 - 4 questionnaires with 5 answers: "Strongly disagree," "Disagree," "Neutral," "Agree," and "Strongly Agree." Subdomains are "Usefulness," "Ease of use and learnability," "Interface quality," "Interaction quality," "Reliability," and "Satisfaction and future use" [24]. The OUQP was developed by modifying the assessment items originally proposed by Deer et al [9], a 6-item questionnaire with cost and time savings. Patient satisfaction was assessed using 3 criteria: answering yes/no or neither. Patients were additionally invited to comment on any challenges or difficulties they experienced during the online sessions.

Clinician-Reported Assessments

Clinician satisfaction with the online sessions was assessed using 5 criteria in the OUQP. Clinicians also provided qualitative feedback regarding any operational or clinical difficulties encountered during remote programming.

Cost and Time Burden Evaluation

Patients were also asked to report the cost and time burdens associated with their in-clinic sessions. The OUQP included items assessing perceived time and cost savings during online sessions. Patient satisfaction with online sessions was assessed using 3 yes/no/neither questions, and patients were asked to describe any difficulties encountered. The Japanese version of PGIC, CGI-C, TUQ, and OUQP is provided in Multimedia Appendix 1.

Statistics

All measured data are presented as mean (SD). The 2-tailed *t* test (matched pairs) was used to analyze the 2 groups when the normality of distribution was verified by the Shapiro-Wilk test. If normality was not confirmed, the Wilcoxon rank-sum test was used. Significance and effect size were determined using Hedges *g* when normality was observed, or Cliff δ when it was not. The Friedman test was applied to compare repeated measures of MDS-UPDRS-III across the 3 assessment points and to evaluate seasonal variation in exposure days across 4 seasons (spring, summer, autumn, and winter).

For any MPIC-J items that showed statistically significant pre-post differences, additional exploratory analyses were conducted to evaluate the clinical relevance of the observed changes. Specifically, the minimally important change (MIC) was estimated using the MPIC-J 7-point global change scale as an anchor; receiver operating characteristic (ROC) analysis was performed to identify the optimal cutoff discriminating subjective worsening, and distribution-based estimates (0.5 SD) were also calculated. Correlation analyses were performed to explore associations between changes in clinical outcomes and candidate explanatory variables. Point-biserial correlation

XSL•FO
RenderX

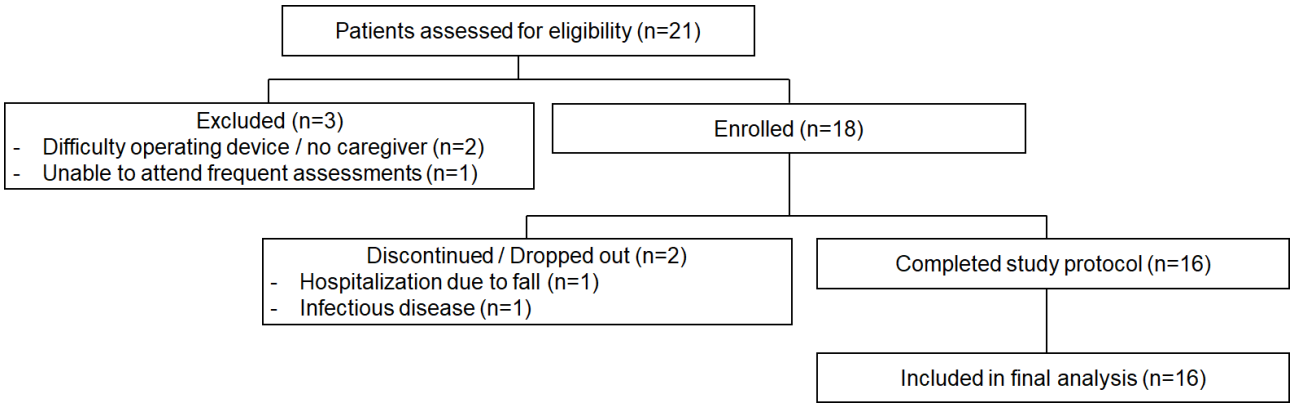
coefficients (*r*) were used for sex, which was treated as a dichotomous variable. For all other continuous or ordinal variables, Spearman rank correlation coefficients (ρ) were calculated. All statistical analyses were performed using R (version 4.2.1; R Core Team) [26], and the significance level was set at $P<.05$.

Results

Overview of Participants

Between August 2023 and April 2024, 21 patients were screened for eligibility from 2 hospitals. Three patients were excluded from the study for the following reasons: 2 patients had difficulty operating the equipment and no familial support, and 1 patient found it challenging to attend frequent assessment appointments. Eighteen patients were enrolled, of whom 2 discontinued participation because of hospitalization and infectious disease. The dropout rate was 11% (2/18). All data points were complete, and no missing values were identified throughout the dataset (Figure 2).

Figure 2. Participant flow diagram. A total of 21 patients were screened; 3 were excluded due to difficulty operating the device or inability to attend frequent assessments. Eighteen patients were enrolled, of whom 2 were discontinued because of hospitalization or infectious disease. Sixteen patients completed all assessments and were included in the final analysis.



The demographic profile was as follows: 50% (8/16) were female, average age was 67.1 (SD 8.4) years (95% CI 62.6 - 71.5), average Hoehn-Yahr Scale score was 3.7 (SD 1.0; 95% CI 3.1 - 4.2), average duration of illness was 15.8 (SD 4.4) years (95% CI 13.4 - 18.1), average duration after bilateral STN-DBS was 6.6 (SD 1.9) years (95% CI 5.5 - 7.6), and the average distance from the resident to the clinic was 42.8 (SD 49.7) kilometers (95% CI 16.3 - 69.3; Table 1). There was no difference in the duration between each assessment visit of in-clinic and online sessions (mean 138.4, SD 64.5 days, 95% CI 104.1 - 172.8 vs mean 152.6, SD 57.2 days, 95% CI 122.2 - 209.8; $P=.53$; Hedges $g=0.220$). Friedman tests revealed no significant seasonal variation in exposure days across the 4 seasons for in-clinic sessions (mean 56.8, SD 33.2 days, 95% CI 39.0 - 74.6; $\chi^2_3=6.73$; $P=.081$; Kendall $W=0.140$) or online sessions (mean 54.7, SD 36.2 days, 95% CI 35.8 - 73.6; $\chi^2_3=6.83$; $P=.078$; Kendall $W=0.143$). No adjustments were made to antiparkinsonian medications or to any sleep-affecting drugs throughout all study periods.

XSL•FO
RenderX

Table . Characteristics of 16 participants.

Number	Sex	Age (years; mean 67.1, SD 8.4 years)	Hoehn-Yahr grade (mean 3.7, SD 1.0)	Disease duration (years; mean 15.8, SD 4.4 years)	Duration after surgery (years; mean 6.6, SD 1.9 years)	Distance from the residence (km; mean 42.8, SD 49.7 km)	Travel time to the clinic (minutes; mean 148.4, SD 81.1 minutes)	Travel expenses for medical visits (JPY[a]; US $1≈155 JPY; mean JPY 4365.0, SD JPY 7472.9)	Caregivers (mean 0.9, SD 0.6), n
1	F[b]	67	3	14	5	39.9	120	2000	1
2	M[c]	56	2	18	10	48.1	120	3000	0
3	F	76	5	11	5	71.6	240	10,000	1
4	F	65	4	9	5	15.1	80	300	1
5	F	72	3	10	5.5	62.5	180	1000	0
6	F	78	5	19	6.5	38.7	120	30,000	1
7	M	75	4	21	4.5	3.0	25	3000	1
8	F	50	2	25	8	13.5	60	390	0
9	F	66	4	18	5	25.6	50	600	1
10	M	68	3	12	5.5	135.0	240	10,000	0
11	M	57	3	12	5.5	24.2	150	750	1
12	M	65	3	13	5.5	17.0	180	2000	1
13	M	60	5	17	10	178.0	360	2000	1
14	M	66	5	17	9	2.0	150	2000	2
15	F	80	4	17	6	1.8	180	300	2
16	M	72	4	19	9	8.6	120	2500	1

[a]JPY: Japanese yen.

[b]F: female.

[c]M: male.

Primary Outcome

Motor Assessments

All 16 patients completed all the assessments. Motor symptoms were assessed by using MDS-UPDRS-III. The average total scores were 24.9 (SD 21.1; 95% CI 13.7 - 36.2), 28.6 (SD 22.5; 95% CI 16.6 - 40.6), and 28.8 (SD 22.0; 95% CI 17.1 - 40.6) at baseline, after in-clinic session, and after online session. Friedman test revealed no differences between the groups (P=.27, Friedman χ^2_2=2.63). Comparison of the rate of change (%) in total MDS-UPDRS-III scores between in-clinic and online sessions revealed n significant difference (mean −19.3%, SD 33.4% vs mean −3.2%, SD 35.3%; P=.31; Hedges g=0.474). These findings indicate no statistically or clinically meaningful worsening of motor symptoms in online sessions compared with in-clinic sessions.

Safety

One patient experienced a fall and lumbar compression fracture during the online visits period. and dropped out from the study because of hospitalization. No other adverse events were observed.

Secondary Outcomes

Patient-Reported Assessments

In comparison of MPIC-J scores of in-clinic and online sessions, no items but MPIC-J item 3 (overall sleep) showed significant difference (mean 3.31, SD 0.85 vs mean 4.19, SD 0.88; P=.013, Cliff δ=−0.49) (Table 2). To further examine the clinical relevance of this change, exploratory analyses were performed for MPIC-J item 3. The mean change of 0.88 points exceeded the distribution-based threshold of 0.5 SD (=0.43), indicating a moderate magnitude of change. Using the MPIC-J 7-point global change scale as an anchor, participants who reported “minimally worse” sleep (MPIC-J score=5; n=4) showed a mean increase of 2.0 points in MPIC-J item 3, representing a reference value for the MIC in the worsening direction. ROC analysis, in which MPIC-J scores were dichotomized into “worsened” (5-7) and “not worsened” (1-4), identified a 1-point increase in MPIC-J item 3 as the optimal cutoff for discriminating subjective worsening (sensitivity 1.00 and specificity 0.73). Taken together, although the observed mean change exceeded the distribution-based threshold, it did not reach the ROC-derived cutoff (1 point) or the anchor-based MIC estimate

XSL•FO
RenderX

(2 points), suggesting that the magnitude of sleep deterioration may be limited in its clinical significance.

Correlation analyses were conducted to evaluate whether the change in MPIC-J item 3 was associated with baseline characteristics, exposure-related factors such as the duration of in-person and online sessions, or changes in other clinical measures (Tables S1-S3 in Multimedia Appendix 2). No baseline characteristics showed significant correlations with the change in MPIC-J item 3. The duration of in-person sessions was not associated with ΔMPIC-J item 3 (ρ=0.0077, *P*=.98). In contrast, the duration of online sessions showed a strong positive correlation with ΔMPIC-J item 3 (ρ=0.71, *P*=.0021), and this association remained significant after Bonferroni correction. Regarding exposure-related and change-based variables, ΔMDS-UPDRS-III during the in-person period showed a positive correlation with ΔMPIC-J item 3 (ρ=0.52, *P*=.038), and ΔMPIC-J item 7 (coping with pain) was also correlated with ΔMPIC-J item 3 (ρ=0.54, *P*=.031). ΔMPIC-J item 2 (pain) showed a similar trend (ρ=0.44, *P*=.086). This implies that patients with greater clinical instability were more likely to experience sleep deterioration when they spent a longer proportion of the follow-up period in online sessions. However, none of these correlations remained significant after Bonferroni correction, indicating that these associations should be interpreted as exploratory and may reflect covarying deterioration across multiple domains rather than robust predictors of sleep worsening.

There was no difference between in-clinic and online sessions for both PGIC and CGI-C (mean 1.88, SD 1.05, 95% CI 1.31 - 2.44 vs mean 1.81, SD 0.95, 95% CI 1.31 - 2.32, *P*=.73, Cliff δ=0.01 for PGIC-1; mean 5.19, SD 0.81, 95% CI 4.76 - 5.62 vs mean 5.06, SD 1.25, 95% CI 4.40 - 5.73, *P*≥.99, Cliff δ=−0.10 for PGIC-2; and mean 3.81, SD 0.73, 95% CI 3.43 - 4.20 vs mean 3.63, SD 0.70, 95% CI 3.25 - 4.00, *P*=.53, Cliff δ=0.15 for CGI-C).

Table . MPIC-J[a] scores on in-clinic and online sessions.

Items	In-clinic		Online		*P* value	Cliff δ
	Mean (SD)	95% CI	Mean (SD)	95% CI		
Overall status	3.94 (0.97)	3.42 - 4.45	4.31 (0.98)	3.79 - 4.85	.27	−0.22
Overall pain	3.94 (0.66)	3.59 - 4.29	4.25 (1.03)	3.70 - 4.80	.13	−0.13
Overall sleep	3.31 (0.85)	2.86 - 3.76	4.19 (0.88)	3.72 - 4.66	.01[b]	−0.49
Overall mood	3.94 (1.30)	3.25 - 4.63	4.19 (0.88)	3.72 - 4.66	.45	−0.14
Overall physical functioning	4.19 (1.38)	3.45 - 4.92	4.19 (0.88)	3.72 - 4.66	.≥.99	0.02
Overall ability to cope with my pain	3.94 (0.90)	3.46 - 4.42	4.00 (0.71)	3.62 - 4.38	.65	−0.05
Overall ability to manage pain flare-ups	3.63 (0.70)	3.25 - 4.00	3.88 (0.86)	3.42 - 4.33	.16	−0.20
Overall effectiveness of my medication	3.75 (0.97)	3.23 - 4.27	4.19 (0.95)	3.68 - 4.69	.10	−0.18

[a]MPIC-J: Multidimensional Evaluation Scale for Patient Impression Change—Japanese version.

[b]*P*<.05.

Usability

The TUQ revealed that more than 70% of the patients responded with “Neutral,” “Agree,” or “Strongly agree” to all items except item 15, “I think the visits provided over the telehealth system are the same as in-person visits.” For this item, 25% (4/16) of patients selected “Disagree,” and 6.25% (1/16) of patients selected “Strongly disagree” (Table 3). There were 6 subdomains in the TUQ: “usefulness” (items 1 - 3), “ease of use and learnability” (items 4 - 6), “interface quality” (items 7 - 10), “interaction quality” (items 11 - 14), “reliability” (items 15 - 17), and “satisfaction and future use” (items 18 - 21). In favor of online sessions, selecting “Strongly agree” or “Agree” was 70.8% (34/48) in “usefulness” domain, 85.4% (41/48) in “ease of use and learnability” domain, 53.1% (34/64) in “interface quality” domain, 62.5% (40/64) in “interaction quality” domain, 37.5% (18/48) in “reliability” domain, and 62.5% (40/64) in “satisfaction and future use” domain (Tables 3 and 4). Overall, patients found the online sessions useful, feasible, and satisfactory, although some concerns about reliability were noted.

XSL•FO
RenderX

Table . Telehealth Usability Questionnaire.

Number	Items	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Usefulness, n (%)						
1	Telehealth improves my access to health care services	0 (0.0)	3 (18.8)	4 (25.0)	3 (18.8)	6 (37.5)
2	Telehealth saves me time traveling to a hospital or specialist clinic	0 (0.0)	1 (6.25)	0 (0.0)	3 (18.8)	12 (75.0)
3	Telehealth provides for my health care needs	1 (6.25)	0 (0.0)	5 (31.3)	6 (37.5)	4 (25.0)
Ease of use and learnability, n (%)						
4	It was simple to use this system	0 (0.0)	2 (12.5)	0 (0)	7 (43.8)	7 (43.8)
5	It was easy to learn to use the system	0 (0.0)	1 (6.25)	2 (12.5)	6 (37.5)	7 (43.8)
6	I believe I could become productive quickly using this system	0 (0.0)	0 (0.0)	2 (12.5)	7 (43.8)	7 (43.8)
Interface quality, n (%)						
7	The way I interact with this system is pleasant	0 (0.0)	0 (0.0)	8 (50.0)	3 (18.8)	5 (31.3)
8	I like using the system	2 (12.5)	2 (12.5)	3 (18.8)	4 (25.0)	5 (31.3)
9	The system is simple and easy to understand	1 (6.3)	0 (0.0)	5 (31.3)	4 (25.0)	6 (37.5)
10	This system is able to do everything I would want it to be able to do	0 (0.0)	4 (25.0)	5 (31.3)	3 (18.8)	4 (25.0)
Interaction quality, n (%)						
11	I could easily talk to the clinician using the telehealth system	0 (0.0)	0 (0.0)	5 (31.3)	3 (18.8)	8 (50.0)
12	I could hear the clinician clearly using the telehealth system	1 (6.3)	0 (0.0)	2 (12.5)	6 (37.5)	7 (43.8)
13	I felt I was able to express myself effectively	0 (0.0)	1 (6.25)	7 (43.8)	4 (25.0)	4 (25.0)
14	Using the telehealth system, I could see the clinician as well as if we met in person	1 (6.3)	2 (12.5)	5 (31.3)	3 (18.8)	5 (31.3)
Reliability, n (%)						

XSL•FO
RenderX

Number	Items	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
15	I think the visits provided over the telehealth system are the same as in-person visits	1 (6.3)	4 (25.0)	3 (18.8)	3 (18.8)	5 (31.3)
16	Whenever I made a mistake using the system, I could recover easily and quickly	2 (12.5)	0 (0.0)	9 (56.3)	1 (6.3)	4 (25.0)
17	The system gave error messages that clearly told me how to fix problems	0 (0.0)	0 (0.0)	11 (68.8)	1 (6.3)	4 (25.0)
Satisfaction and future use, n (%)						
18	I feel comfortable communicating with the clinician using the telehealth system	0 (0.0)	4 (25.0)	4 (25.0)	3 (18.8)	5 (31.3)
19	Telehealth is an acceptable way to receive health care services	0 (0.0)	0 (0.0)	5 (31.3)	4 (25.0)	7 (43.8)
20	I would use telehealth services again	0 (0.0)	2 (12.5)	4 (25.0)	3 (18.8)	7 (43.8)
21	Overall, I am satisfied with this telehealth system	0 (0.0)	0 (0.0)	5 (31.3)	4 (25.0)	7 (43.8)

Table . Subdomains in Telehealth Usability Questionnaire.

Subdomains	In favor of in-clinic sessions	Neutral	In favor of online sessions
Usefulness, %	10.4	18.8	70.8
Ease of Use and Learnability, %	6.3	8.3	85.4
Interface Quality, %	14.1	32.8	53.1
Interaction Quality, %	7.8	29.7	62.5
Reliability, %	14.6	47.9	37.5
Satisfaction and Future Use, %	9.4	28.1	62.5

The OUQP revealed that 2 clinicians experienced no difficulty in using online devices and felt a reduced necessity for in-clinic sessions in 62.5% (10/16) of the patients; however, sometimes it was not easy to provide rapid treatment resolution (5/16, 31.3%; Table 5). The reasons were, in part, the inability of clinicians to touch patients directly (eg, inability to check the rigidity of extremities) and the anxiety of the patients for not consulting in person. Approximately 70% of patients with PD were satisfied with online sessions; however, the preference for online sessions was 37.5% (6/16), which was the same percentage as the preference for an in-clinic session.

XSL•FO
RenderX

Table . Online Usability Questionnaire for Programming.

Items	Yes	No	Neither
For physician, %			
Ability to establish an audiovisual connection	100.0	0.0	0.0
Achieving the clinical goals of a programming session	100.0	0.0	0.0
Ability to complete a session	100.0	0.0	0.0
Reduction in the need for an in-person follow-up after a remote session	62.5	18.8	18.8
Providing rapid treatment resolution	68.8	0.0	31.3
For patient, %			
Satisfied with a remote session	68.8	12.5	18.8
Preference for a remote session	37.5	37.5	25.0
Providing rapid treatment resolution	56.2	37.5	6.3

Reduction in Time and Cost Burdens

The mean time and cost per patient visit were 148.4 (SD 81.1 minutes; 95% CI 105.2 - 191.6) and 4365.0 (SD 7472.9 Japanese yen [US $1≈155 JPY]; 95% CI 383.0 - 8,347.0), respectively. Seventy-five percent of patients with PD required caregivers for their visit, irrespective of in-clinic or online sessions. The average number of caregivers attending the patients was 0.88 (SD 0.60; 95% CI 0.56 - 1.19) per visit (Table 1). The average total time and cost saved per visit for patients and caregivers were calculated as 280.0 (SD 181.5 minutes; 95% CI 183.3 - 376.7) and 7974.4 (SD 14,235.9 JPY [US $1≈155 JPY]; 95% CI 388.6 - 15,560.1), respectively (Table 1).

Discussion

Main Findings in Patients Under Chronic DBS

Our study demonstrated the feasibility of online sessions using the NeuroSphere Virtual Clinic remote care system compared with the conventional in-clinic sessions. Motor symptoms did not change in either group, as assessed using the MDS-UPDRS-III. Motor stability in patients under chronic DBS is expected over short intervals, regardless of visit modality; therefore, the absence of deterioration might not be interpreted as evidence of clinical effectiveness. This study provides a prospective evaluation of the effect of online sessions on patients who have already been implanted with and under chronic DBS. To our knowledge, this is the first study to investigate the effects of online sessions using DBS devices in Japanese patients. While the prospective design improves data quality compared with retrospective reports, the single-arm nature of this study precludes definitive causal inferences.

The only adverse event occurred between the last in-clinic visit and the first online session, making it unlikely that the event was related to the online sessions. Overall, online sessions appeared to be feasible and clinically acceptable from a safety perspective for patients with PD who have chronically implanted DBS devices. Although the fall was not clearly attributable to online sessions, the inability to directly assess rigidity, balance, and gait remains an inherent limitation of remote care. As expected, there was a substantial reduction in both cost and time burden for patients and caregivers [8,9,11,15-18].

Patient-Reported Outcomes and Usability

In this study, approximately half of the patients thought that online sessions were equivalent to in-clinic sessions, as reflected in TUQ item 15 and the OUQP. Around 70% of patients reported satisfaction with the online sessions (TUQ item 21 and OUQP; Tables 3 and 5). Given that TUQ subdomain "Usefulness" refers to "the positive effects on clinical outcomes" or "reduction in clinical cost" [24], our results indicated that 70.8% (34/48) of the patients felt that online sessions had positive effects on clinical outcomes. Previous retrospective studies assessing patients with chronic DBS implantation indicated that 62% of patients preferred online visits [17]. A cross-sectional survey of teleprogramming after each online session for patients with chronic pain equipped with the NeuroSphere Virtual Clinic remote care showed that 93.8% (15/16) of the patients prefer the online session [9]. Similarly, a cross-sectional survey of teleprogramming following online sessions in patients with PD who had undergone de novo DBS implantation found that 100% of patients were satisfied with the online sessions and expressed a willingness to participate in future [11]. PD is progressive, and the burden of in-clinic visits for both patients and caregivers increases as symptoms worsen. In this context, online systems may help reduce that burden without negatively affecting patients' core motor symptoms, although this requires confirmation in controlled studies.

XSL•FO
RenderX

Sleep Deterioration and Its Clinical Interpretation

One unexpected finding was a potential signal of deterioration in sleep status in patients who received online sessions, which has not been previously reported (Table 2), although the observed change in sleep status consistently fell below the established MIC thresholds, indicating limited clinical relevance. Exploratory correlations suggested that patients with preexisting clinical instability—such as motor worsening or fluctuations in pain-related domains—were more likely to show sleep deterioration.

Interestingly, the duration of online sessions showed a strong association with ΔMPIC-J item 3. Because the online period varied among patients, this association is unlikely to reflect intentional differences in exposure to online sessions. Instead, it may imply that patients with greater instability were more susceptible to sleep deterioration when they spent a longer proportion of the follow-up period in online sessions. These exploratory correlations could imply a potential interaction between patient instability and the extent of online follow-up; however, these findings should be interpreted with caution. Taken together, these findings suggest a need for further investigation into whether sleep deterioration during online-based DBS management may not be solely attributable to the online modality itself but rather an emergence in clinically unstable patients who may instead benefit from shorter online follow-up and earlier return to in-person assessment.

Comparison With Previous Studies and Added Value of This Study

According to a randomized controlled trial, online programming sessions for de novo implanted DBS systems resulted in approximately 38% shorter treatment duration compared with in-clinic optimization and increased patient access to therapy adjustments by nearly 1.9 times [18]. Given the strong therapeutic impact of the initial stimulation, there is concern that the added value of online sessions may be overlooked or inadequately assessed. In addition, Chen et al [15] reported the possibility that physicians may adjust medications less frequently during remote programming than during in-hospital programming for de novo DBS-implanted patients with PD. This raises the possibility of bias due to differences in dopamine levels. In our study, we targeted patients with PD with an average duration of 6.6 (SD 1.9, range 4.5 - 10, median 5.5) years, representing the longest follow-up period reported in the literature. In contrast to the evidence in patients who have undergone de novo DBS, studies involving chronically implanted DBS systems have been retrospective or cross-sectional (Table 6). These reports suggest that remote programming can substantially reduce travel time and financial burden and is generally well accepted by patients [10,17] but lacked prospective evaluation of safety, usability, or clinical stability in long-term DBS management. Our study partially addresses this gap by providing preliminary prospective evaluation in patients who had chronically implanted DBS devices, offering a baseline for future comparative trials. It also offers the first population-specific data from Japan, using a device platform not previously examined in this context. By focusing on long-term DBS users, we examined the specific population for whom online sessions are presently feasible in real-world Japanese clinical settings. The inclusion of detailed usability measures (TUQ and OUQP) and the observation of a potential signal in sleep-related symptoms further add domain-specific insights not highlighted in earlier reports. Together, these findings suggest the feasibility and acceptable short-term safety profile of integrating online sessions into DBS care. Although time and cost savings are not direct care quality measures, our descriptive estimates—based on a typical frequency of 6 clinic visits per year—suggested that patients and caregivers might save approximately 28 hours and 47,844 JPY (US $1≈155 JPY).

XSL•FO
RenderX

Table . Summary of key studies on teleprogramming in deep brain stimulation for Parkinson disease.

Study (year)	Population type	Study design	Values, n[a]	Main outcomes	Key findings
PD[b] with chronic DBS[c]					
Present study (2026)	PD with chronic STN-DBS[d] (mean 6.6 years postoperatively)	Prospective, single-arm, multicenter, and self-controlled cohort	16	MDS-UPDRS-III[e], MPIC-J[f], PGIC[g], CGI-C[h], TUQ[i], OUQP[j], and time/cost savings	No motor worsening. Modest sleep deterioration. Saved 280 minutes and 7974 JPY[k] (US $1≈155 JPY) per visit. 70% satisfaction.
Ma et al (2021) [10]	PD with STN-DBS (mean 27.0 months postoperatively)	Retrospective, single-center database review	90	Motor symptoms (UPDRS-III) and time/cost savings	Saved >12 hours (for 90% of patients) and >2000 CNY[l] (US $1≈155 JPY) (for 76.7%) per visit. Feasible for long-term management.
Wan et al (2024) [17]	Patients with chronic DBS	Single-center cross-sectional survey	132	Patient burden, attitudes, and satisfaction	Teleprogramming reduces burden; 62% of chronic patients preferred online visits.
Pintér et al (2022) [14]	Movement disorders (PD, dystonia, and ET) with DBS (mean 5.8 years postoperatively)	Retrospective, single-center registry-based modeling	319	Modeled travel distance, time, cost, and caregiver burden	Saved 415.2 kilometers and 342.1 minutes per visit. Estimated to save 17 days of caregiver leave over 10 years.
Zhang et al (2021) [13]	Movement disorders (PD and dystonia) with DBS (median 6.7 months postoperatively)	Retrospective, national multicenter observational study	196	Frequency of requests, safety, and satisfaction	89% of 909 sessions were satisfactory. Essential for care during pandemic without serious adverse events.
Zhang et al (2020) [8]	Movement disorders (PD, dystonia, and ET[m]) with DBS (median 6.7 months)	Retrospective, single-center observational study	589	Implementation of Bluetooth technology	Safe and efficient parameter adjustments. Average travel distance saved: 1141 kilometers.
PD with de novo DBS					
Gharabaghi et al (2025) [18]	PD with de novo STN-DBS implantation	Prospective, multicenter, randomized controlled trial	38	Time to symptom improvement (PGI-C), QoL[n], and safety	Remote group achieved clinical benefit 15.1 days faster than in-clinic group. Safe postoperative optimization.
Wan et al (2024) [16]	PD with de novo STN-DBS	Retrospective, single-center controlled study	18	ΔUPDRS-III, ΔPDQ-8[o], and ΔLEDD[p] (1-year follow-up)	No significant difference in motor improvement or LEDD reduction between in-clinic and online groups.
Chen et al (2022) [15]	PD with de novo STN-DBS	Retrospective, single-center controlled study	58	ΔMDS-UPDRS-III, LEDD reduction, and sessions	Improvement in motor scores was similar across remote, hospital, and combined groups.
Xu et al (2021) [12]	PD with de novo DBS	Retrospective, single-center observational study	28	Motor symptoms (UPDRS-III) and satisfaction	UPDRS-III scores significantly improved. 93.75% were satisfied with the procedure convenience.
Li et al (2017) [7]	PD with de novo bilateral STN-DBS	Prospective, single-center, randomized controlled trial	31	Safety and efficacy of remote programming	Remote adjustment was found effective compared with sham stimulation. Early safety verification.

XSL•FO
RenderX

Study (year)	Population type	Study design	Values, n[a]	Main outcomes	Key findings
Xu et al (2021) [11]	PD with de novo DBS	Prospective, single-center pilot study	18	Remote video-based outcome measures	Smartphone-based remote motor assessment (video) is a feasible pilot approach for post-DBS care.

[a]Number of patients who underwent remote programming and successfully completed the study assessments.
[b]PD: Parkinson disease.
[c]DBS: deep brain stimulation.
[d]STN-DBS: subthalamic nucleus–deep brain stimulation.
[e]MDS-UPDRS-III: Movement Disorder Society Unified Parkinson's Disease Rating Scale Part Ⅲ.
[f]MPIC-J: Multidimensional Evaluation Scale for Patient Impression Change—Japanese version.
[g]PGIC: Patient's Global Impression of Change.
[h]CGI-C: Clinical Global Impression of Change.
[i]TUQ: Telehealth Usability Questionnaire.
[j]OUQP: Online Usability Questionnaire for Programming.
[k]JPY: Japanese yen.
[l]CNY: Chinese yuan.
[m]ET: essential tremor.
[n]QoL: quality of life.
[o]PDQ-8: Parkinson's Disease Questionnaire-8.
[p]LEDD: levodopa equivalent daily dose.

Barriers to Implementation and Practical Implications for Chronic DBS Care

Two patients were excluded from this study because of insufficient caregiver support and lack of digital literacy. Caregiver support is essential for patients with PD at these stages. Patients with PD may experience difficulty using small screens on tablet computers or smartphones due to motor impairments and cognitive dysfunction. These limitations may partly explain the relatively low agreement rates in the "Reliability" and "Interface quality" subdomains of the TUQ. These subdomains assess, "how easily the user can recover from an error and how the system provides guidance to the user in the event of an error" and "how pleasant the telemedicine technology or computer system was to use for the consumer," respectively (Tables 3 and 4) [24]. It remains difficult for clinicians to establish online consultation systems in hospitals because of limited medical resources. Administrative challenges include difficulties obtaining official prescriptions and a lack of medical insurance coverage [17]. The OUQP showed that it was sometimes difficult to provide rapid treatment resolution (31.3%; Table 5). This was, in part, due to the inability to touch the patients directly (eg, the inability to check the rigidity of the extremities). These challenges should be addressed to improve feasibility. Despite these concerns, it is surprising to see many patients satisfied with online sessions and inclined to use in the future as shown in the TUQ subdomain "Satisfaction and future use" (Table 4).

Limitations

This study has some limitations. First, this was a single-arm longitudinal study without a comparator group, which limits causal inference and the ability to contextualize the observed changes. Long-term patients with PD equipped with DBS devices with no experience of online sessions are rare, making recruitment challenging; therefore, a single-arm design was the most feasible approach for this population. Second, the small sample size may have reduced the statistical power. Although it was determined a priori using conservative assumptions, the study may still have been underpowered to detect smaller effects. Third, the observation period for the online sessions was short and longer online sessions may yield different effects. Fourth, the study population was necessarily selective, as patients without caregiver support or with low digital literacy were unable to participate. This introduces a potential selection bias and limits the generalizability of our findings to more vulnerable groups. At the same time, patients who can realistically engage in remote DBS care in routine clinical practice are typically those with caregiver support and sufficient digital literacy; therefore, the present cohort also reflects the subset of patients who had undergone chronic DBS for whom online sessions are currently feasible. Finally, the findings apply specifically to patients with bilateral STN-DBS using the Infinity/NeuroSphere remote care system within a Japanese health care setting with adequate infrastructure, which further limits the applicability of our findings to other devices, stimulation targets, or health care systems. The absence of randomization, blinding, or parallel controls also raises the possibility of observer, performance, detection, and expectation bias. Thus, we must evaluate our results carefully because of these limitations.

Conclusions

Sharma et al [27] described ideal candidates for online sessions as those who have a supportive caregiver, do not have cognitive or psychiatric impairment, or have impediments to in-person visits. Given that establishing a relationship with the DBS care team is essential for patients who had undergone de novo DBS

XSL•FO
RenderX

[27] and that the postoperative levodopa equivalent daily dose reduction rate may be lower in patients managed online [15], online sessions might be most suitable for patients with chronically implanted DBS devices. The absence of statistically or clinically meaningful worsening in all aspects—with only a modest and clinically limited change in sleep status—in patients with PD and chronic DBS implantations supports the feasibility of incorporating online sessions, or a hybrid of in-clinic and online sessions, into DBS care. Such approaches could be considered as a potential means to alleviate the burden on patients and caregivers. Future studies incorporating validated sleep-specific instruments, such as the Pittsburgh Sleep Quality Index, objective measures including actigraphy, and longer follow-up periods, will be essential to determine whether the sleep changes identified here represent transient fluctuations or clinically relevant deterioration. Despite some limitations, the online sessions showed potential feasibility and an acceptable safety profile in this preliminary evaluation, suggesting their potential role in reducing the time and cost burden for patients.

Acknowledgments

The authors thank all the members of the clinical research support team for movement disorders in Tokushima University and Kurashiki Heisei Hospital.

The authors attest that there was no use of generative artificial intelligence (AI) technology in the generation of text, figures, or other informational content of this manuscript.

Funding

No financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Data Availability

The original data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

RM contributed to conceptualization, data curation, formal analysis, writing—original draft, and writing—review & editing. HM contributed to conceptualization, data curation, formal analysis, and writing—review & editing. KT contributed to conceptualization, data curation, and writing—review & editing. JF participated in data curation, formal analysis, and writing—review & editing. TS, ST, MH, and KS participated in data curation and writing—review & editing. TM and KM participated in writing—review & editing. YT participated in supervision and writing—review & editing.

Conflicts of Interest

The Department of Advanced Brain Research is funded by the Beauty Life Corporation.

Multimedia Appendix 1
Japanese version of Patient’s Global Impression of Change, Clinical Global Impression of Change, Telehealth Usability Questionnaire, and Online Usability Questionnaire for Programming.
[DOCX File, 28 KB - mhealth_v14i1e80223_app1.docx]

Multimedia Appendix 2
Factors associated with changes in Multidimensional Evaluation Scale for Patient Impression Change—Japanese version item 3 (Sleep).
[DOCX File, 19 KB - mhealth_v14i1e80223_app2.docx]

References

1. Tanner CM, Ostrem JL. Parkinson’s disease. N Engl J Med 2024 Aug 1;391(5):442-452. [doi: 10.1056/NEJMra2401857] [Medline: 39083773]
2. Gandolfi M, Geroin C, Dimitrova E, et al. Virtual reality telerehabilitation for postural instability in Parkinson’s disease: a multicenter, single-blind, randomized, controlled trial. Biomed Res Int 2017;2017:7962826. [doi: 10.1155/2017/7962826] [Medline: 29333454]
3. van der Kolk NM, de Vries NM, Kessels RPC, et al. Effectiveness of home-based and remotely supervised aerobic exercise in Parkinson’s disease: a double-blind, randomised controlled trial. Lancet Neurol 2019 Nov;18(11):998-1008. [doi: 10.1016/S1474-4422(19)30285-6] [Medline: 31521532]
4. Bianchini E, Onelli C, Morabito C, et al. Feasibility, safety, and effectiveness of telerehabilitation in mild-to-moderate Parkinson’s disease. Front Neurol 2022;13:909197. [doi: 10.3389/fneur.2022.909197] [Medline: 35785358]

XSL•FO
RenderX

5. Silva-Batista C, Wilhelm JL, Scanlan KT, et al. Balance telerehabilitation and wearable technology for people with Parkinson's disease (TelePD trial). BMC Neurol 2023 Oct 13;23(1):368. [doi: 10.1186/s12883-023-03403-3] [Medline: 37833645]
6. Nakanishi H, Morigaki R, Fujikawa J, et al. Online training program maintains motor functions and quality of life in patients with Parkinson's disease. Front Digit Health 2024;6:1486662. [doi: 10.3389/fdgth.2024.1486662] [Medline: 39606179]
7. Li D, Zhang C, Gault J, et al. Remotely programmed deep brain stimulation of the bilateral subthalamic nucleus for the treatment of primary Parkinson disease: a randomized controlled trial investigating the safety and efficacy of a novel deep brain stimulation system. Stereotact Funct Neurosurg 2017;95(3):174-182. [doi: 10.1159/000475765] [Medline: 28571034]
8. Zhang J, Hu W, Chen H, Meng F, Li L, Okun MS. Implementation of a novel Bluetooth technology for remote deep brain stimulation programming: the pre- and post-COVID-19 Beijing experience. Mov Disord 2020 Jun;35(6):909-910. [doi: 10.1002/mds.28098] [Medline: 32478886]
9. Deer TR, Esposito MF, Cornidez EG, Okaro U, Fahey ME, Chapman KB. Teleprogramming service provides safe and remote stimulation options for patients with DRG-S and SCS implants. J Pain Res 2021;14:3259-3265. [doi: 10.2147/JPR.S332966] [Medline: 34703301]
10. Ma Y, Miao S, Zhou R, Zhang Q, Chen H, Liang Y. Application of remote deep brain stimulation programming for Parkinson's disease patients. World Neurosurg 2021 Mar;147:e255-e261. [doi: 10.1016/j.wneu.2020.12.030]
11. Xu X, Zeng Z, Qi Y, et al. Remote video-based outcome measures of patients with Parkinson's disease after deep brain stimulation using smartphones: a pilot study. Neurosurg Focus 2021 Nov;51(5):E2. [doi: 10.3171/2021.8.FOCUS21383] [Medline: 34724646]
12. Xu J, Wang J, Keith S, et al. Management of Parkinson's disease patients after DBS by remote programming: preliminary application of single center during quarantine of 2019-nCoV. J Neurol 2021 Apr;268(4):1295-1303. [doi: 10.1007/s00415-020-10273-z] [Medline: 33104873]
13. Zhang C, Zhu K, Lin Z, et al. Utility of deep brain stimulation telemedicine for patients with movement disorders during the COVID-19 outbreak in China. Neuromodulation 2021 Feb;24(2):337-342. [doi: 10.1111/ner.13274] [Medline: 33006811]
14. Pintér D, Járdaházi E, Janszky J, Kovács N. Potential clinical and economic benefits of remote deep brain stimulation programming. Sci Rep 2022 Oct 19;12(1):17420. [doi: 10.1038/s41598-022-22206-z] [Medline: 36261678]
15. Chen S, Xu SJ, Li WG, et al. Remote programming for subthalamic deep brain stimulation in Parkinson's disease. Front Neurol 2022 Nov 24;13:1061274. [doi: 10.3389/fneur.2022.1061274]
16. Wan X, Duan C, Lin Z, Zeng Z, Zhang C, Li D. Motor improvement of remote programming in patients with Parkinson's disease after deep brain stimulation: a 1-year follow-up. Front Neurol 2024;15:1398929. [doi: 10.3389/fneur.2024.1398929] [Medline: 38962477]
17. Wan X, Lin Z, Duan C, Zeng Z, Zhang C, Li D. Comparison of the burdens and attitudes between standard and web-based remote programming for deep brain stimulation in Parkinson disease: survey study. JMIR Aging 2024 Oct 23;7:e57503. [doi: 10.2196/57503] [Medline: 39441208]
18. Gharabaghi A, Groppa S, Navas-Garcia M, et al. Accelerated symptom improvement in Parkinson's disease via remote internet-based optimization of deep brain stimulation therapy: a randomized controlled multicenter trial. Commun Med (Lond) 2025 Jan 31;5(1):31. [doi: 10.1038/s43856-025-00744-7] [Medline: 39890864]
19. Gagnon CM, Scholten P, Atchison J. Multidimensional patient impression of change following interdisciplinary pain management. Pain Pract 2018 Nov;18(8):997-1010. [doi: 10.1111/papr.12702] [Medline: 29679452]
20. Wakaizumi K, Yamada K, Mibu A, Kogo H, Kosugi S, Morisaki H. Development of the multidimensional evaluation scale for patient impression change—Japanese version (MPIC-J). J Jpn Soc Pain Clin 2021;28(5):59-63. [doi: 10.11321/jjspc.20-0044]
21. Farrar JT, Young JP, LaMoreaux L, Werth JL, Poole MR. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001 Nov;94(2):149-158. [doi: 10.1016/S0304-3959(01)00349-9] [Medline: 11690728]
22. Guy W. ECDEU Assessment Manual for Psychopharmacology, Revised: U.S. Dept. of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, National Institute of Mental Health, Psychopharmacology Research Branch; 1976:218-222 URL: urn:oclc:record:1043039853 [accessed 2026-05-19]
23. Schneider LS, Olin JT. Clinical global impression of change clinical global impressions in Alzheimer's clinical trials. Int Psychogeriatr 1996 Jun;8(2):277-290. [doi: 10.1017/S1041610296002645]
24. Parmanto B, Lewis, Jr. AN, Graham KM, Bertolet MH. Development of the telehealth usability questionnaire (TUQ). Int J Telerehab 2016;8(1):3-10. [doi: 10.5195/ijt.2016.6196]
25. Hurst H, Bolton J. Assessing the clinical significance of change scores recorded on subjective outcome measures. J Manipulative Physiol Ther 2004 Jan;27(1):26-35. [doi: 10.1016/j.jmpt.2003.11.003] [Medline: 14739871]
26. Ihaka R, Gentleman R. R: a language for data analysis and graphics. J Comput Graph Stat 1996 Sep;5(3):299-314. [doi: 10.1080/10618600.1996.10474713]
27. Sharma VD, Safarpour D, Mehta SH, et al. Telemedicine and deep brain stimulation - current practices and recommendations. Parkinsonism Relat Disord 2021 Aug;89:199-205. [doi: 10.1016/j.parkreldis.2021.07.001] [Medline: 34274215]

XSL•FO
RenderX

Abbreviations

CGI-C: Clinical Global Impression of Change
DBS: deep brain stimulation
MDS-UPDRS-III: Movement Disorder Society Unified Parkinson's Disease Rating Scale Part III
MIC: minimally important change
MPIC-J: Multidimensional Evaluation Scale for Patient Impression Change—Japanese version
OUQP: Online Usability Questionnaire for Programming
PD: Parkinson disease
PGIC: Patient's Global Impression of Change
ROC: receiver operating characteristic
STN: subthalamic nucleus
TUQ: Telehealth Usability Questionnaire

Edited by J Zhou; submitted 19.Sep.2025; peer-reviewed by F Asci, I Haq, N Modugno; revised version received 14.Apr.2026; accepted 16.Apr.2026; published 09.Jun.2026.

Please cite as:
Morigaki R, Mure H, Takasuka K, Fujikawa J, Shinmen T, Tarui S, Hatanaka M, Shinohara K, Matsuda T, Miyake K, Takagi Y
Influence of Online Sessions via a Deep Brain Stimulation Device: Prospective, Single-Arm, Longitudinal, Nonrandomized Self-Controlled Cohort Study
JMIR Mhealth Uhealth 2026;14:e80223
URL: https://mhealth.jmir.org/2026/1/e80223
doi:10.2196/80223

© Ryoma Morigaki, Hideo Mure, Kouki Takasuka, Joji Fujikawa, Toshiro Shinmen, Shin Tarui, Masataka Hatanaka, Keisuke Shinohara, Taku Matsuda, Kazuhisa Miyake, Yasushi Takagi. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 9.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Inclusive Contactless Monitoring for Older Adults From Diverse Backgrounds: Mixed Methods Study

Titilola Yakubu[1], MBBS, MSc; Nooshin Jafari[1], MSc; Samya Torres[1], MBA; Michael Lim[1], BCom; David Rivest-Henault[2], PhD; Thomas Vaughan[2], PhD; Catherine Proulx[2], MSc; Linda Pecora[2], BEng; Di Jiang[2], PhD; Kendall Ho[1], MD

[1]Faculty of Medicine, University of British Columbia, 6200 University Blvd, Vancouver, BC, Canada

[2]Medical Devices Research Centre, National Research Council Canada, Boucherville, QC, Canada

Corresponding Author:
Kendall Ho, MD
Faculty of Medicine, University of British Columbia, 6200 University Blvd, Vancouver, BC, Canada

Abstract

Background: As Canada's population ages, accessible tools for chronic health monitoring are increasingly needed. Traditional and contact-based devices pose barriers for underserved populations due to cost, maintenance, and usability. Contactless sensing technologies offer a promising alternative, but equitable development requires inclusive engagement and diverse data collection.

Objective: This study aimed, first, to gather perspectives from culturally and health-diverse older adults on innovative contactless vital sign monitoring technology and, second, to build on a diverse dataset for developing and testing the contactless sensing software VitalSeer by the National Research Council of Canada.

Methods: A mixed methods study was conducted with older adults from diverse backgrounds across 3 sites. Participants were asked to respond to a questionnaire that had closed- and open-ended questions regarding their perspectives on contactless sensing technology. Video and reference vital sign data were collected from each participant using a portable system designed specifically for gathering controlled data.

Results: We collected data from 48 participants (mean age 70, SD 8 years), of whom 98% (n=47) expressed a positive perception of the usefulness of a contactless sensing system. We also identified four themes from the qualitative analysis of the open-ended questions: (1) perceived value—system potential and clinical relevance; (2) ease of use—noninvasiveness and comfort; (3) trust and transparency—data security and clarity of design; and (4) inclusion and improvement—accessibility, functionality, and feature expansion. Finally, the collected data, 288 minutes of concurrent video and reference vital sign data, will be used to test and enhance contactless sensing software for diverse older adult populations.

Conclusions: This work advances the goal of inclusive medical device research and development. It highlights the potential for contactless sensing to be adopted to support independent living for older, diverse adults. Research is ongoing to adapt the technology for widespread adoption.

(JMIR Mhealth Uhealth 2026;14:e79892) doi:10.2196/79892

KEYWORDS

contactless monitoring; vital signs; digital intervention; remote monitoring; older adults; telemedicine

Introduction

Background

As Canada's population ages, the demand for health care services, particularly those aimed at monitoring and managing chronic health conditions, is growing [1]. Traditional monitoring solutions require in-person checkups, which can be physically exhausting and disruptive [2,3]. Current contact-based solutions, such as using a pulse oximeter, might be undesirable or less suitable for some populations. Limiting factors include the cost of purchase, the need to manage an extra device, and the need to properly maintain and use it regularly. Older adults, particularly those living independently in underserved populations, may not have access to adequate tools for monitoring their health, putting them at greater risk [4].

Research has found that well-structured community engagement efforts, developed through meaningful consultation and active participation, can lead to positive changes in health behaviors, better access to health care services, enhanced health literacy, and improved health outcomes [5]. Integrating diverse perspectives in health technology development is crucial to address disparities in health outcomes, barriers to service access, and representation gaps [6]. Historically, marginalized and underserved populations have been excluded from health technology development, leading to poor uptake and even

XSL•FO
RenderX

technological bias within important health assistive devices [7,8]. Additional barriers, such as cost, maintenance, and unclear added value, further deter widespread use among these demographics [9].

VitalSeer [10,11] is a software technology under development by the National Research Council of Canada for contactless, noninvasive vital signs assessment. It uses camera-based remote photoplethysmography (rPPG) to measure heart rate, respiratory rate, and oxygen saturation (SpO_2) [12]. This technology could enable an easy-to-use monitoring solution that uses ubiquitous cameras, for example, mobile phone cameras, webcams, or home surveillance cameras. Sensing based on rPPG relies on the analysis of the light reflected by superficial tissues through the skin, so it can be affected by several extrinsic and intrinsic factors, such as the quality of the camera, skin type and tone, and environmental lighting conditions. Therefore, representative data acquired from a diverse population are critical to ensuring applicability.

The purpose of this study was to engage diverse populations for the development of contactless sensing technology [10] and to gather pertinent perceptions for technology improvement and eventual deployment. The contributions of this work include (1) a report on the perceptions of contactless sensing technologies from diverse populations in real-world community settings and (2) the acquisition of diverse high-quality data for VitalSeer development and testing.

Objective

The primary objective of this study was to gather perceptions about technology adoption from older adults with diverse health and cultural backgrounds. In parallel, a secondary objective was to collect high-quality diverse data to support the development of the technology.

Methods

Study Design

We conducted a mixed methods study using a concurrent triangulation design, in which closed-ended questions and open-ended questions were administered to the same convenience sample of older adults living in British Columbia, Canada. Quantitative and qualitative data were collected concurrently during the same study sessions, analyzed independently, and then integrated during the interpretation phase. Quantitative findings were used to identify patterns and trends, while qualitative responses were used to contextualize and explain those themes, with convergence across both datasets used to strengthen our conclusions.

Ethical Considerations

Ethical approval for the study was obtained from the University of British Columbia Ethics Board (H23-02571) and the National Research Council of Canada (NRC 2020 - 134).

All participants provided informed consent before participating in the study. Initial consent discussions were conducted via telephone calls and in person to explain the study objectives and address participants' questions.

All study data were deidentified before analysis by replacing identifiable information, such as names, with participant IDs. Study data were stored securely, and only study team members could access them.

Participants received a CAD $10 (CAD $1=US $0.71 as of June 17, 2026) e-gift card as a token of appreciation for their time and participation.

Sampling and Recruitment

Participant recruitment was facilitated through REACH BC, an online research platform for the Canadian province of British Columbia. Additionally, the research team conducted extensive outreach to multicultural older adult home and community centers in the Lower Mainland region of Vancouver, British Columbia. Out of 49 centers approached, 2 (4.1%) were selected for facilitating on-site participant recruitment and study conduct: Unique Get Together Society and South Granville Seniors Centre. The centers were chosen based on their representation of a diverse range of ethnic groups, including Black, Latino, White, and Chinese communities. Furthermore, interest and capacity to participate, proximity to the research lab, and the availability of suitable space were also factors considered for selection.

Eligibility criteria prioritized older adults aged 60 years and older, who could attend a 1-hour in-person session at participating community centers. Participants younger than 60 years were allowed to participate for broader inclusion across cultural and ethnic backgrounds (people of color and Indigenous communities). Participants were required to remain still for approximately 2 minutes at a time. Proficiency in English was required unless a multilingual member of the research team was available to assist with instructions. Furthermore, participants needed to be in generally good health and capable of sitting still. The study emphasized the inclusion of individuals from diverse backgrounds, with varying skin tones and health conditions, to ensure a representative sample. Individual sessions were conducted at either a laboratory or one of 2 community centers.

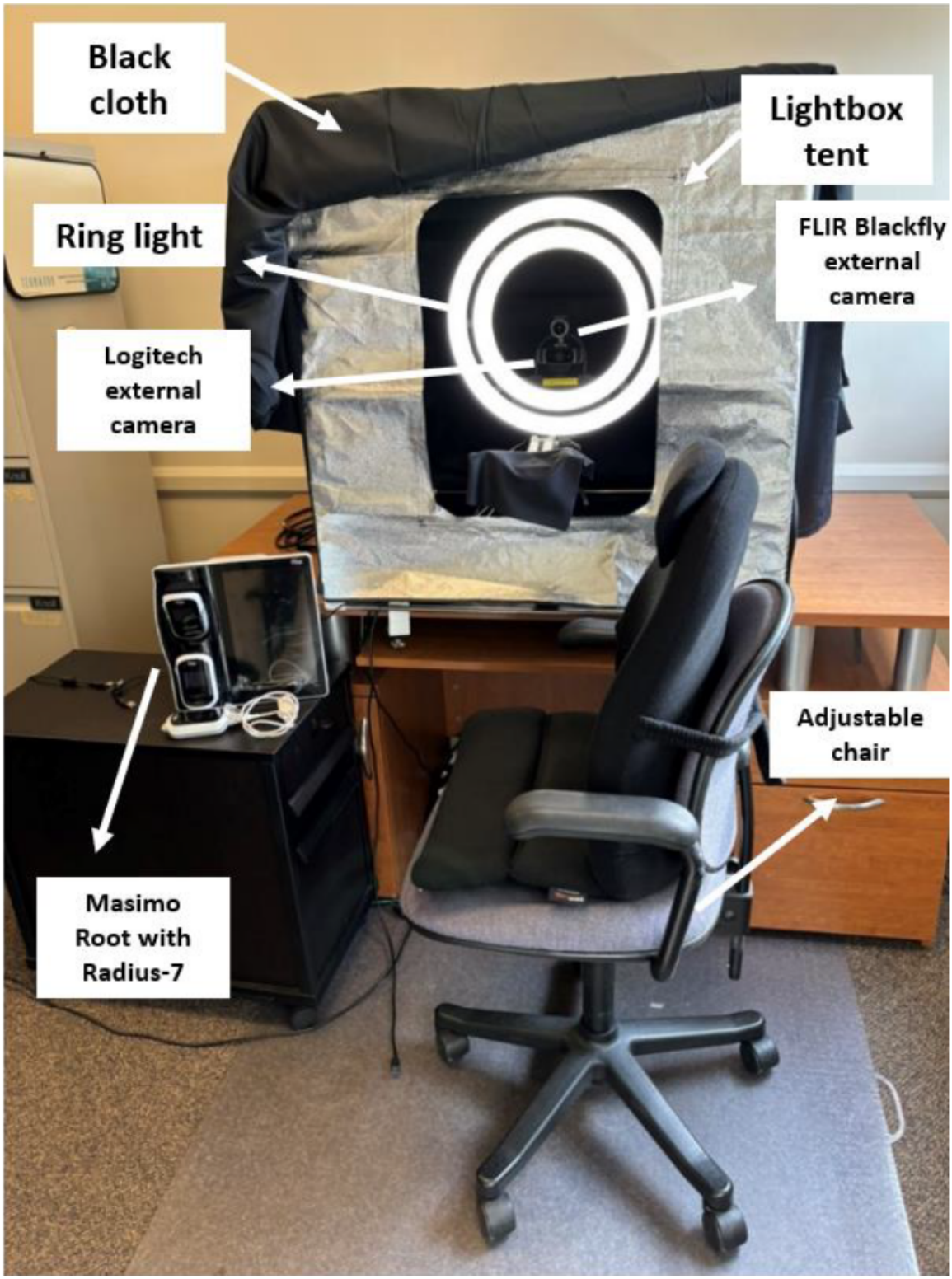
Data Collection

Recordings

A deployable system, henceforth referred to as a "lightbox," was designed specifically for controlled data collection in this study. The lightbox featured an LED ring-light enclosure for uniform illumination, a chin rest to help participants remain still and minimize motion during video acquisitions, a metal support structure to mount 2 cameras (a consumer-grade Logitech webcam and a research-grade FLIR Blackfly), both directed at the participant's face, and a black cloth to cover the participant and block external light from entering. The lightbox was specifically designed to help collect data under optimal and reproducible conditions, ensuring minimal influence by external variables, such as fluctuations in ambient light and participant movement. Adjacent to the lightbox was a Lenovo T470s laptop (Windows 10) running the VitalSeer software. The cameras were connected via USB, and the laptop was connected to Ethernet for optimal bandwidth, as depicted in Figure 1.

XSL•FO
RenderX

Figure 1. Hardware configuration of the data collection system at the research team's laboratory.



Reference physiological parameters were measured from Food and Drug Administration–cleared and Health Canada–approved Masimo medical contact-based patient monitoring devices. These devices included a reference pulse oximeter, which provides noninvasive measurements of heart rate, respiratory rate, and SpO_2, and the Masimo Root monitor equipped with the Radius-7 continuous monitoring system (Masimo Corp). The Masimo Root system uses advanced signal extraction technology to deliver highly accurate vital sign measurements even in challenging clinical environments, such as low perfusion or motion artifacts.

Participants underwent three 2-minute recordings. Simultaneous readings from the Masimo Root system served as the gold standard for comparison, as these data were later used for validation.

XSL•FO
RenderX

Questionnaire

The questionnaire assessed user perceptions, acceptance, and usability of the remote health monitoring system using Likert-scale and open-ended questions. It evaluated perceived usefulness for self-monitoring and remote health care, willingness to share data, comfort with prolonged use, and concerns about wearables. Open-ended questions gathered feedback on user experience, system improvements, and overall satisfaction. This approach provided insights into user needs and barriers, informing system refinements for real-world application.

The questionnaire was administered at the end of the session after the VitalSeer recordings were completed. The survey contained a mix of quantitative and qualitative elements, including 6 rating scale questions and 4 open-ended questions to allow participants to provide more detailed feedback (Table 1). Skin tone scale data were also collected during study sessions using the Fitzpatrick skin type classification system [13]. Participants self-identified their skin type using the Fitzpatrick scale (I-VI), and a research team member (TY, NJ, or ST) also independently categorized each participant based on visible facial skin characteristics. In cases of discrepancy, the team revisited the classification and assigned the closest matching Fitzpatrick skin type.

Table . Participants' postsession questionnaire.

Question text	Type
I feel that the system that collects my vital signs would be useful to me in self-monitoring my health by frequently tracking of my vital signs at home	Likert scale (5-point)
I feel that the system that collects vital signs would be helpful for healthcare providers to monitor their patients remotely (*e.g.*, for patients living in rural/remote areas)	Likert scale (5-point)
In the future, I would be willing to send my health-related data and share my results from the system to my healthcare provider online	Likert scale (5-point)
In the future, I would be willing to use the system for a longer duration (*e.g.*, one week) to track my health-related data, after my visit to emergency department or hospital	Likert scale (5-point)
In the future, I think the system will be helpful to me for managing the health-related concerns related to wearable sensors that touch skin and are used between multiple users	Likert scale (5-point)
In the future, I think the system at home will be helpful for me to self-manage my health and live more independently	Likert scale (5-point)
After testing the system, what do you think about it? (What did/did not like about it)?	Open-ended
Do you have any suggestions for improving the system?	Open-ended
Do you have any suggestions for us to help improve your experience?	Open-ended
On a scale of 1-10, how was your overall test experience with the system today? Would you like to elaborate?	Open-ended

Data Processing and Analysis

Quantitative data were analyzed descriptively using SPSS Statistics for Mac (Version 29.0; IBM Corp). Frequencies and percentages were calculated for Likert-scale responses, and demographic variables were summarized using descriptive statistics (frequencies and proportions).

Qualitative feedback from the open-ended questions was analyzed using NVivo 12 (Lumivero) following reflexive thematic analysis. The analysis involved familiarization with the data through repeated reading of responses, inductive code generation, theme development, theme review and refinement, and final theme definition and naming. Codes were generated inductively from participant responses rather than a predefined coding framework, allowing themes to emerge directly from the data. Two coders (TY and NJ) independently reviewed and coded the responses, comparing findings throughout to enhance consistency and credibility. In cases of disagreement, coders collaboratively reviewed the relevant data and reassessed their interpretations. Where consensus could not be reached, KH reviewed the codes within the context of the research questions and made the final decision.

Ensuring Rigor and Trustworthiness

We ensured the rigor and trustworthiness of our research through triangulation, combining closed-ended responses and open-ended responses to capture and explore deeper perspectives. Peer debriefing by 2 independent coders (TY and NJ) validated themes, ensuring consistency and accuracy.

We also documented each research step to maintain transparency and reduce bias, enhancing credibility. The team's positionality also strengthened the process: KH, with more than 25 years of experience in mixed methodologies, provided theoretical expertise; TY, who is experienced in qualitative methods and has an MBBS, contributed clinical insights and methodological knowledge; and NJ, a researcher with an engineering background, added relevant academic and practical experience.

XSL•FO
RenderX

Results

Participants

We conducted a focused community recruitment phase over 34 days (March 2-April 4, 2024), following several months of engagement with community stakeholders. A total of 48 participants were recruited, with 26 sessions held at community centers. Most participants were aged between 66 and 70 years. The mean age of participants was 70 (SD 8) years. Detailed demographic information, including age, sex, data collection location, and Fitzpatrick skin tone [13] (as observed by the study team), is presented in Table 2 and Figure 2. The study also included 5 participants aged 60 or younger, comprising 4 Black individuals and 1 Indigenous participant.

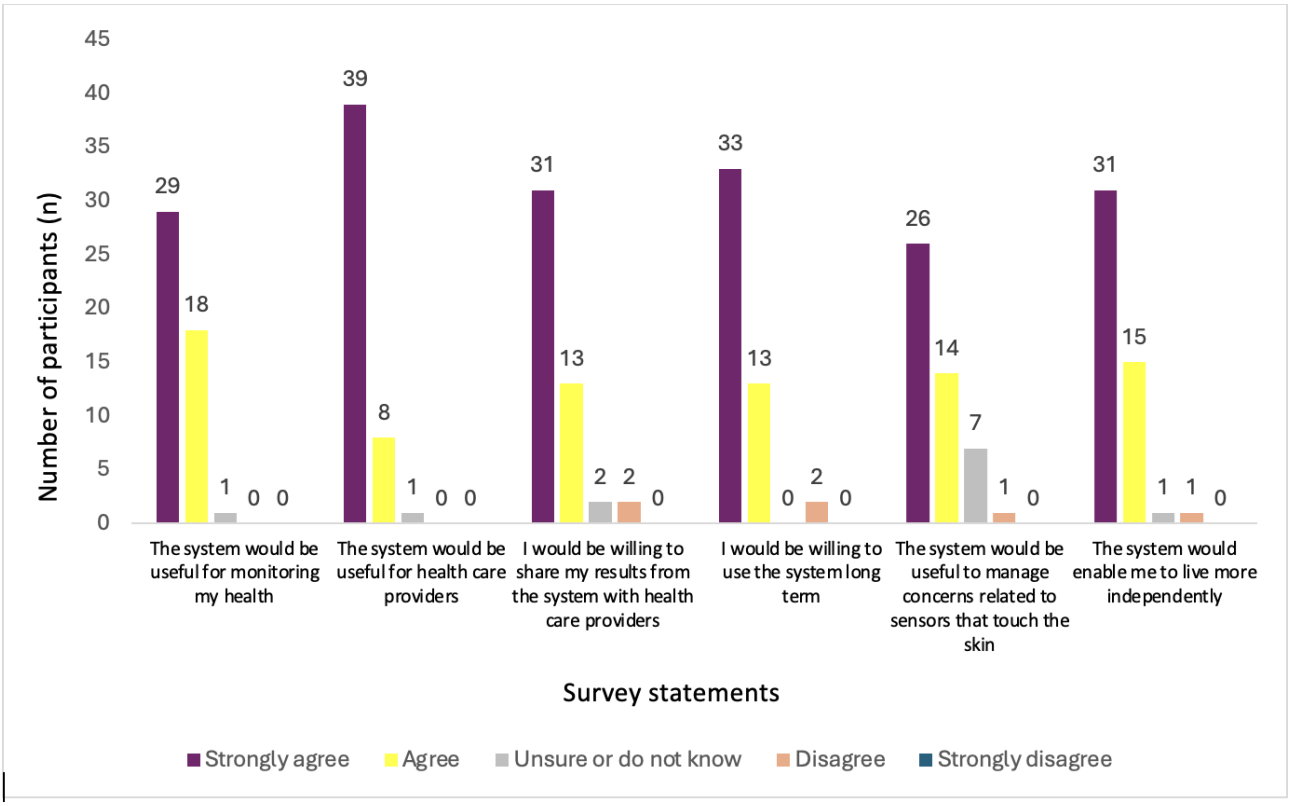
Table . Participant demographics (N=48).

Demographics	Participants, n (%)
Age group (y)	
<55	1 (2.1)
56 - 60	4 (8.3)
61 - 65	8 (16.7)
66 - 70	14 (29.2)
71 - 75	5 (10.4)
76 - 80	8 (16.7)
≥81	8 (16.7)
Sex	
Female	29 (60.4)
Male	19 (39.6)
Location	
UBC[a] Digital Emergency Medicine Laboratory	22 (45.8)
Unique Get Together Society	13 (27.1)
South Granville Senior Center	13 (27.1)
Fitzpatrick skin tone scale, type I	3 (6.25)

[a]UBC: The University of British Columbia.

XSL•FO
RenderX

Figure 2. Responses to Likert-scale questionnaire items regarding perceptions of the contactless monitoring system.



Recordings

Each of the 48 participants contributed 3 rounds of recordings, resulting in 144 distinct recordings. Each of these recordings consisted of paired video data (from 2 cameras) over 2 minutes, yielding 288 minutes of high-quality, controlled video data for each of the 2 cameras. In addition, reference vital sign data were collected from Food and Drug Administration–cleared and Health Canada–approved Masimo medical contact-based patient monitoring devices that correspond to the videos.

Questionnaire

Likert-Scale Questions

Figure 2 shows that most (47/48, 98%) participants agreed that contactless monitoring would be useful for independently monitoring their vital signs. Participants largely agreed on several other points, including the willingness to share results with health care providers, long-term use, and benefits for independent living.

However, 8% (4/48) of participants disagreed or were unsure about sharing their results with health care providers, 4% (2/48) disagreed with using the system long term, and 2% (1/48) felt that the system would not help manage concerns related to contact monitoring or support independent living.

Open-Ended Questions

Overview

We asked participants open-ended questions to elicit further perceptions of the technology. We identified four overarching themes that reflect participants' perceptions of the system: (1) perceived value—system potential and clinical relevance; (2) ease of use—noninvasiveness and comfort; (3) trust and transparency—data security and clarity of design; and (4) inclusion and improvement—accessibility, functionality, and feature expansion.

Each theme is broken down into subthemes, with participant quotes used to illustrate key points.

Theme 1: Perceived Value—System Potential and Clinical Relevance

Value for Remote Monitoring in Chronic Conditions

Participants emphasized the potential of VitalSeer to support at-home health monitoring, especially for individuals with chronic or respiratory conditions. Many saw it as a valuable tool that could allow early detection of symptoms and continuous monitoring without the need for frequent clinic visits. They believed this would be particularly useful for patients with long-term conditions who often require close surveillance:

> *This system has a lot of potential, especially for people with breathing problems or conditions like COPD. It could help doctors catch issues early.* [Participant 12]

> *This could be a very useful tool for patients to use and for doctors to gain data on their patients. It is very promising even in this test configuration.* [Participant 31]

Anticipated Benefits for Patients and Clinicians

In addition to monitoring individual health, participants highlighted the broader benefit of the system in supporting clinical decision-making. They felt that providing real-time access to vital signs could enhance health care providers' ability

XSL•FO
RenderX

to respond quickly to changes in a patient's health, enabling more efficient and personalized care. This was viewed as a win-win for both patients and clinicians:

> *If this works as intended, it'll reduce my visits to the clinic and still help my doctor stay updated on my health.* [Participant 3]

> *The idea that doctors can see what's going on in real-time from home—that's huge.* [Participant 19]

Brief Positive Feedback Without Detailed Commentary

While the majority offered detailed responses, a small proportion (12.5%) simply described the system as "fine," "good," or "satisfactory" without elaboration. This may reflect satisfaction without specific points of critique, or it may indicate a neutral experience with no strong positive or negative reactions. Although not descriptive, this feedback suggests the system was at least acceptable in its current form:

> *Good.* [Participant 38]

> *Fine. It worked okay.* [Participant 40]

Theme 2: Ease of Use—Noninvasiveness and Comfort

Simple and Fast to Use

A key feature appreciated by participants was the system's simplicity. Many emphasized how quick and effortless it felt to use, with a minimal learning curve. This ease of use was seen as crucial for ensuring long-term engagement, especially for users with limited time or low interest in technology. The process was described as unobtrusive and easy to fit into everyday life:

> *It was easy and fast; I didn't feel like it took much time out of my day.* [Participant 5]

> *Convenient and easy to navigate.* [Participant 13]

Low Technical Barriers

Even participants with limited prior experience using digital health tools commented on how straightforward the system was. They appreciated not having to deal with complicated instructions or setup steps. This accessibility was framed as especially important for broad adoption, including among older adults or individuals less familiar with technology:

> *Not complicated.* [Participant 37]

> *Effortless.* [Participant 29]

Noninvasive Experience

Participants highlighted that the system's noncontact approach was one of its most appealing features. Unlike wearables or devices that require physical contact or attachment, for many, this hands-off approach appeared to be convenient for daily use:

> *The fact that it's non-invasive makes it less stressful to use daily.* [Participant 43]

> *I like that it doesn't require me to wear anything or have anything attached to my body.* [Participant 30]

Theme 3: Trust and Transparency—Data Security and Clarity of Design

Concerns About Data Privacy and Protection

Several participants raised concerns about the safety of personal health data collected by the system. Given the sensitivity of the information involved, they wanted clear assurances that their data would be securely transmitted and stored. Participants highlighted this as a critical factor in determining their willingness to use the system regularly:

> *The big concern would be data security during transmission.* [Participant 26]

> *If I'm going to be using this regularly, I need to know that my data is safe.* [Participant 11]

Lack of Transparency About Data Use

Participants also expressed a desire for more transparency about how the data collected would be used. Specifically, they wanted to understand what types of data were being recorded, who would have access to them, and how they would inform their care. This was seen as important for building trust and encouraging users to engage with the system fully:

> *I want to know exactly what data is being collected and how it will help in the long run.* [Participant 30]

> *More explanation would help. It's not very clear right now how this ties into my overall care.* [Participant 41]

Perception of an Unfinished Design

Some participants described the system as still in development and wanted clearer communication about its future direction. These participants noted a lack of polish in the current interface and requested updates on what improvements could be expected. They felt that greater transparency in the development process would improve their confidence in the system:

> *The design feels a bit unfinished. If we could get more information about where it's heading, it would help us trust it more.* [Participant 21]

> *It looks like it's still in progress. I'd like to see more polish before using it daily.* [Participant 8]

Theme 4: Inclusion and Improvement—Accessibility, Functionality, and Feature Expansion

Barriers for Older Adults and Low-Tech Users

Participants raised concerns about the system's accessibility for older adults or users unfamiliar with technology. They emphasized that many older adults might lack smartphones, tablets, or reliable internet access. These barriers could prevent widespread adoption unless targeted solutions, such as simplified interfaces or alternative access methods, were implemented:

> *This is not going to be feasible for adults who do not have access to a computer or cellphone. How are you going to solve that problem?* [Participant 9]

> *It's a great idea, but I'm worried about how people with little tech knowledge will manage.* [Participant 24]

XSL•FO
RenderX

Need for Training and Visual Support

To support wider usability, participants recommended features such as color-coded feedback and audio instructions. These additions were viewed as particularly important for older adults, many of whom may struggle to interpret numerical health data. Suggestions included multilingual audio guidance and visuals that clearly indicate whether a measurement is within a healthy range:

> *Color coding for vital signs (numbers) so seniors will understand if the numbers are good.* [Participant 17]
>
> *Voice guidance in the app use—especially in multiple languages—would really help.* [Participant 23]

Desire for Expanded Monitoring Features

Several participants suggested that the system include other vital signs, such as blood pressure, SpO_2, and temperature. These metrics were viewed as essential for people managing chronic illnesses—particularly in communities where conditions like hypertension are common:

> *Blood pressure in the Black community is important… hypertension is a typical problem, so consider that one in VitalSeer.* [Participant 38]
>
> *What would be useful would be a warning for low O_2 saturation.* [Participant 46]

Suggestions for Practical Design Considerations

Participants also made comments about day-to-day use cases—such as whether the system would function properly if someone was wearing glasses or makeup. Others noted that low internet connectivity in rural or remote areas might prevent some users from accessing the system altogether:

> *At this time, it is OK to remove the glasses, but later, you need to consider not removing them.* [Participant 2]
>
> *This needs to work for rural users with slow or no internet.* [Participant 36]

Discussion

Principal Findings

This study demonstrates the feasibility of contactless sensing research and development in real-world community settings beyond the laboratory, engaging older adults from diverse backgrounds. The study participants included an older adult population comprising Black, Latino, White, Indigenous, and Chinese communities. Contactless sensing technology was well received, making it a promising solution for monitoring vital signs among older adults. Feedback from the participants provided insights into technology adoption.

Comparison With Previous Work

Our findings also show interest in technological solutions to support independent living for older adults. Noninvasive features helped increase acceptance by overcoming a major barrier to technology adoption in older adults compared with previous studies [14-17]. Although several studies have explored the use of contactless health monitoring systems [18], few have advanced to real-world testing. Most research on rPPG has largely been limited to laboratory settings [19-21]. Our study implemented a field-based approach and revealed important insights into the practical challenges and adoption of the technology when used by diverse older adults. While participants raised comments related to ease of use and data privacy, security, and monitoring, these findings were consistent with findings from previous studies [16,22]. It will be important to identify, implement, and communicate measures that can help build trust in these technologies.

The data collection in this research also contributes to the development and robustness of the VitalSeer software through the inclusion of participants with varying skin tones, a known variable that affects the accuracy of rPPG-based systems [18,23]. The inclusion of diverse participants also reflects a commitment to equity, diversity, and inclusion in health care technology development. By ensuring representation across different skin tones, genders, and health conditions, this work promotes solutions that are accessible, effective, and beneficial to all.

Novelty of This Clinical Study

One of the most important contributions of this study is its focus on diverse and underserved populations, addressing a significant issue in both health care and research: the underrepresentation of marginalized groups. By including a wide range of participants from various backgrounds, this study supports equitable health care solutions. Remote monitoring technologies have the potential to bridge gaps in health care access by providing individuals in underserved communities with tools to manage their health autonomously, particularly those who face barriers to frequent in-person medical visits. To our knowledge, this study is the first to elicit open-ended feedback from diverse participants to understand the perceived benefits and limitations of contactless sensing technology.

The use of contactless sensing technology in this study also shows that it can be implemented outside laboratory and traditional clinical environments, suggesting it can be made accessible to individuals who may not regularly interact with the health care system.

Limitations

Camera-based vital sign monitoring is susceptible to poor lighting conditions [23,24]. In this study, we used the lightbox to control the environment and ensure high-quality data collection. The effect of lighting on rPPG quality is an area of continued research; the methods used for signal acquisition and analysis are expected to evolve and improve with time.

We cannot discount the possibility that the presence and use of the lightbox may have altered some participants' perceptions of contactless sensing technology. If future researchers evaluate perceptions of a technological concept, it may be advisable to administer questionnaires before exposing participants to development or testing measures (in this study, the presence of the lightbox) that could influence their responses.

Implications for Practice

The findings from this study have several implications for both health care practice and policy. First, the deployment of the

XSL•FO
RenderX

contactless sensing technology in this study in community settings suggests that it could eventually be scaled up for broader use in home health monitoring programs, particularly for older populations. Such systems have the potential to ease the strain on health care facilities by empowering older adults to independently monitor and manage their health from home. This reduces the need for frequent hospital visits or long-term care stays, which is particularly beneficial in Canada, where rising emergency department wait times have been documented [18,23].

From a policy perspective, the results suggest that with the right design and implementation strategies, contactless monitoring systems can help close some of the gaps in health care access that disproportionately affect underserved populations. Future policy initiatives should focus on integrating these technologies into public health strategies, particularly in remote areas.

Future Directions

Moving forward, several avenues for future research emerge from this study. First, insights and video data from this study will assist in the ongoing development of the VitalSeer software, including its current extension to the mobile environment. Future research could also focus on expanding participant pools to target, for example, individuals with severe health conditions. Participants with mobility impairments might provide valuable insights into how these technologies can be more effectively integrated. Finally, studies should explore the long-term impact of using contactless sensing technologies in home settings, particularly how their use influences health outcomes, quality of life, and health care use over time.

Conclusions

In conclusion, contactless sensing represents a promising advancement in health monitoring. The development and testing we conducted in real-world settings, with a diverse sample population, highlight its potential to be successfully adopted to improve health outcomes and support aging in place for older adults. Further research and development are ongoing to ensure applicability under varying lighting conditions and the inclusion of additional features. These refinements will help facilitate its widespread adoption in Canada for both home and hospital use.

Acknowledgments

The authors thank the participants for their time and for sharing their insights. The authors also thank the Unique Get Together Society and the South Granville Seniors Centre for referring participants to the study and providing space to carry out the study. Generative artificial intelligence tools were not used to generate or edit any portion of the scientific content, analysis, results, or interpretation in this manuscript. All content, study design, data analysis, interpretation, and conclusions were fully developed and verified by the authors.

Funding

This study was supported by the 2022 National Research Council Aging in Place Program.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to the individually identifiable nature of the video recordings (ie, full-face images) and the associated physiological measurements. These data constitute personal health information and cannot be openly shared in accordance with the University of British Columbia Research Ethics Board requirements (H23-02571) and the National Research Council of Canada ethics approval (NRC 2020-134). Deidentified quantitative survey data and aggregated qualitative themes are available within the manuscript. The full questionnaire instrument is provided in Multimedia Appendix 1. Researchers who wish to request restricted access to the controlled video and physiological datasets for scientific inquiry or validation studies may contact the corresponding author. Requests will be reviewed in consultation with the University of British Columbia and the National Research Council of Canada data governance teams to ensure compliance with privacy, ethical, and data-handling requirements. A data access agreement will be required.

Authors' Contributions

Conceptualization: ML, DJ, KH
Data curation: TY, NJ, ST
Formal analysis: TY
Funding acquisition: DJ, KH
Investigation: TY, NJ, ST, LP
Methodology: TY, NJ, DJ, KH
Project administration: NJ, ML, LP
Resources: DJ, KH
Software: DR-H, TV, CP, DJ
Supervision: NJ, DJ, KH
Validation: DR-H, TV, KH
Writing—original draft: TY

XSL•FO
RenderX

Writing—review and editing: NJ, ML, DR-H, TV, CP, LP, DJ, KH

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire.
[DOCX File, 141 KB - mhealth_v14i1e79892_app1.docx]

Checklist 1
SRQR checklist.
[PDF File, 138 KB - mhealth_v14i1e79892_app2.pdf]

References

1. Aging and chronic diseases: a profile of Canadian seniors. Government of Canada. URL: https://publications.gc.ca/site/eng/9.888658/publication.html [accessed 2026-06-04]
2. Legramante JM, Morciano L, Lucaroni F, et al. Frequent use of emergency departments by the elderly population when continuing care is not well established. PLoS One 2016;11(12):e0165939. [doi: 10.1371/journal.pone.0165939] [Medline: 27973563]
3. Carpenter CR, Shelton E, Fowler S, et al. Risk factors and screening instruments to predict adverse outcomes for undifferentiated older emergency department patients: a systematic review and meta-analysis. Acad Emerg Med 2015 Jan;22(1):1-21. [doi: 10.1111/acem.12569] [Medline: 25565487]
4. Fulmer T, Reuben DB, Auerbach J, Fick DM, Galambos C, Johnson KS. Actualizing better health and health care for older adults. Health Aff (Millwood) 2021 Feb;40(2):219-225. [doi: 10.1377/hlthaff.2020.01470] [Medline: 33476185]
5. Cyril S, Smith BJ, Possamai-Inesedy A, Renzaho AM. Exploring the role of community engagement in improving the health of disadvantaged populations: a systematic review. Glob Health Action 2015;8:29842. [doi: 10.3402/gha.v8.29842] [Medline: 26689460]
6. Wallerstein N, Duran B. Community-based participatory research contributions to intervention research: the intersection of science and practice to improve health equity. Am J Public Health 2010 Apr 1;100 Suppl 1(Suppl 1):S40-S46. [doi: 10.2105/AJPH.2009.184036] [Medline: 20147663]
7. Rodriguez JA, Clark CR, Bates DW. Digital health equity as a necessity in the 21st Century Cures Act era. JAMA 2020 Jun 16;323(23):2381-2382. [doi: 10.1001/jama.2020.7858] [Medline: 32463421]
8. Yao R, Zhang W, Evans R, Cao G, Rui T, Shen L. Inequities in health care services caused by the adoption of digital health technologies: scoping review. J Med Internet Res 2022 Mar 21;24(3):e34144. [doi: 10.2196/34144] [Medline: 35311682]
9. Moore K, O'Shea E, Kenny L, et al. Older adults' experiences with using wearable devices: qualitative systematic review and meta-synthesis. JMIR Mhealth Uhealth 2021 Jun 3;9(6):e23832. [doi: 10.2196/23832] [Medline: 34081020]
10. Ho K, Rivest-Henault D, Jafari N, et al. VitalSeer: the development of a contactless sensing technology based on a user-centric data-driven clinical approach. AMIA Jt Summits Transl Sci Proc 2022;2022:244-253. [Medline: 35854744]
11. Rivest-Hénault D, Proulx C, Ho K, et al. Influential factors in rPPG: insights from a diverse and inclusive empirical study. In: IEEE BSN 2024 Conference Proceedings: IEEE; 2024. [doi: 10.1109/BSN63547.2024.10780670]
12. Rivest-Hénault D, Pagiatakis C, Bernhardt R, et al. Quasi real-time contactless physiological sensing using consumer-grade cameras. In: 2021 IEEE 9th International Conference on Healthcare Informatics (ICHI): IEEE; 2021. [doi: 10.1109/ICHI52183.2021.00038]
13. Gupta V, Sharma VK. Skin typing: Fitzpatrick grading and others. Clin Dermatol 2019;37(5):430-436. [doi: 10.1016/j.clindermatol.2019.07.010] [Medline: 31896400]
14. Peek ST, Luijkx KG, Rijnaard MD, et al. Older adults' reasons for using technology while aging in place. Gerontology 2016;62(2):226-237. [doi: 10.1159/000430949] [Medline: 26044243]
15. Kim D, Bian H, Chang CK, Dong L, Margrett J. In-home monitoring technology for aging in place: scoping review. Interact J Med Res 2022 Sep 1;11(2):e39005. [doi: 10.2196/39005] [Medline: 36048502]
16. Wang S, Bolling K, Mao W, et al. Technology to support aging in place: older adults' perspectives. Healthcare (Basel) 2019 Apr 10;7(2):60. [doi: 10.3390/healthcare7020060] [Medline: 30974780]
17. Boldy D, Grenade L, Lewin G, Karol E, Burton E. Older people's decisions regarding "ageing in place": a Western Australian case study. Australas J Ageing 2011 Sep;30(3):136-142. [doi: 10.1111/j.1741-6612.2010.00469.x] [Medline: 21923707]
18. Molinaro N, Schena E, Silvestri S, et al. Contactless vital signs monitoring from videos recorded with digital cameras: an overview. Front Physiol 2022;13:801709. [doi: 10.3389/fphys.2022.801709] [Medline: 35250612]
19. Novita D, Adikusuma FW, Rohadi N, et al. Development of contactless human vital signs monitoring device with remote-photoplethysmography using adaptive region-of-interest and hybrid processing methods. Intell Based Med 2024;10:100160. [doi: 10.1016/j.ibmed.2024.100160]

XSL•FO
RenderX

20. Karthick R, Dawood MS, Meenalochini P. Analysis of vital signs using remote photoplethysmography (RPPG). J Ambient Intell Human Comput 2023 Dec;14:16729-16736. [doi: 10.1007/s12652-023-04683-w]
21. Ma X, Tobón DP, El SA. Remote photoplethysmography (rPPG) for contactless heart rate monitoring using a single monochrome and color camera. In: McDaniel T, Berretti S, Curcio I, Basu A, editors. Smart Multimedia: Springer; 2020:248-262. [doi: 10.1007/978-3-030-54407-2_21]
22. Chen C, Ding S, Wang J. Digital health for aging populations. Nat Med 2023 Jul;29(7):1623-1630. [doi: 10.1038/s41591-023-02391-8] [Medline: 37464029]
23. Lee RJ, Sivakumar S, Lim KH. Review on remote heart rate measurements using photoplethysmography. Multimed Tools Appl 2024;83:44699-44728. [doi: 10.1007/s11042-023-16794-9]
24. Castaneda D, Esparza A, Ghamari M, Soltanpur C, Nazeran H. A review on wearable photoplethysmography sensors and their potential future applications in health care. Int J Biosens Bioelectron 2018;4(4):195-202. [doi: 10.15406/ijbsbe.2018.04.00125] [Medline: 30906922]

Abbreviations

rPPG: remote photoplethysmography
SpO_2: oxygen saturation

Edited by A Stone, L Buis; submitted 30.Jun.2025; peer-reviewed by FVD Ouderaa, WYH Lam; revised version received 15.May.2026; accepted 18.May.2026; published 03.Jul.2026.

Please cite as:
Yakubu T, Jafari N, Torres S, Lim M, Rivest-Henault D, Vaughan T, Proulx C, Pecora L, Jiang D, Ho K
Inclusive Contactless Monitoring for Older Adults From Diverse Backgrounds: Mixed Methods Study
JMIR Mhealth Uhealth 2026;14:e79892
URL: https://mhealth.jmir.org/2026/1/e79892
doi:10.2196/79892

© Titilola Yakubu, Nooshin Jafari, Samya Torres, Michael Lim, David Rivest-Henault, Thomas Vaughan, Catherine Proulx, Linda Pecora, Di Jiang, Kendall Ho. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 3.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Postoperative Telerehabilitation in Patients With Hip Fracture: Systematic Review and Meta-Analysis

Can-xin Cai[1,2*], MSc; Penny Ping Qin[3*], PhD; Chuan-yao Liu[2], BSc; Peng Cai[2], MSc; Xijun Wei[1,2,4], PhD

[1]Rehabilitation Laboratory of Mixed Reality, Shenzhen Hospital, Southern Medical University, Shenzhen, China

[2]Department of Rehabilitation Medicine, Shenzhen Hospital, Southern Medical University, 13 Xinhu Road, Bao'an District, Shenzhen, China

[3]Department of Rehabilitation Sciences, Faculty of Health and Social Sciences, Hong Kong Polytechnic University, Hong Kong Special Administrative Region, China (Hong Kong)

[4]Department of Occupational Therapy, School of Rehabilitation Sciences, Southern Medical University, Shenzhen, China

[*]these authors contributed equally

Corresponding Author:
Xijun Wei, PhD
Rehabilitation Laboratory of Mixed Reality, Shenzhen Hospital, Southern Medical University, Shenzhen, China

Abstract

Background: Telerehabilitation programs use information and communications technologies to facilitate exercise training, self-management education, and health behavior modifications for nonhospitalized patients. In recent years, 2 systematic reviews have examined the effectiveness of telerehabilitation in the recovery of patients with hip fractures but have yielded inconsistent results. This is a significant gap because tools to assess clinical domains such as pain, range of motion, and deformity are crucial for patient outcomes. Moreover, the long-term effects of telerehabilitation on postoperative functional recovery in patients with hip fractures have not been investigated.

Objective: To address the abovementioned research gap, this systematic review aimed to evaluate the short- and long-term effects of telerehabilitation on postoperative functional recovery in patients with hip fractures.

Methods: We searched the PubMed, Cochrane Library, Embase, and Web of Science databases from inception to March 31, 2025. Randomized controlled trials investigating the effect of postoperative telerehabilitation in patients with hip fractures were included in this systematic review. Outcomes of interest included hip function measured using the Harris Hip Score; functional mobility measured using the Short Physical Performance Battery and the timed up and go test; and ability to perform basic activities of daily living measured using the Functional Independence Measure, Barthel index, or modified Barthel index. Meta-analyses were performed using the RevMan software (version 5.3).

Results: A thorough literature search conducted in April 2025 yielded a total of 8 studies involving 740 patients for inclusion in this systematic review. Meta-analysis showed that telerehabilitation was effective for improving the Harris Hip Score at the intervention end point (2 included studies involving 156 participants; mean difference [MD] 7.42, 95% CI 5.61-9.23; P<.001; I^2=3%), the Short Physical Performance Battery score at the end point (4 included studies involving 430 participants; MD 1.34, 95% CI 1.14-1.55; P<.001; I^2=33%) and at follow-up (2 included studies involving 292 participants; MD 1.17, 95% CI 1.00-1.34; P<.001; I^2=0%), the timed up and go score at the end point (4 included studies involving 156 participants; MD –8.45, 95% CI −11.28 to −5.62; P<.001; I^2=0%) and at follow-up (2 included studies involving 63 participants; MD −7.66, 95% CI −13.78 to −1.53; P=.01; I^2=0%), and ability to perform basic activities of daily living at the end point (5 included studies involving 354 participants; standardized MD 1.65, 95% CI 0.78-2.51; P<.001; I^2=91%) and at follow-up (standardized MD 0.43, 95% CI 0.05-0.81; P=.03; I^2=48%).

Conclusions: Our review found that postoperative telerehabilitation may benefit short- and long-term functional recovery in patients with hip fractures compared to conventional rehabilitation. However, the evidence was weak due to the limited number and insufficient quality of the included studies and the heterogeneity across the studies.

Trial Registration: PROSPERO CRD42024498569; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024498569

(JMIR Mhealth Uhealth 2026;14:e77341) doi:10.2196/77341

KEYWORDS

hip fracture; postoperation; telerehabilitation; systematic review; meta-analysis

XSL•FO
RenderX

Introduction

Hip fractures are a major global public health challenge, with an annual incidence of over 10 million cases [1]. They are most prevalent among adults aged 65 years and older, a population in which fractures are correlated with high mortality rates and a substantial burden on patients and their families [2]. The standard management for hip fractures involves surgical intervention followed by rehabilitation to restore walking ability [3,4]. Despite this, most older patients fail to recover their prefracture level of function, and this persistent impairment may lead to an increased risk of falls [5]. Therefore, long-term postoperative management for older adults with hip fractures is essential to enhance their functional recovery and overall quality of life.

Since the COVID-19 pandemic, telerehabilitation has gained popularity, with physical therapists increasingly willing to incorporate it into their routine practices, prompting health care facilities to adopt innovative methods for engaging with patients. As a branch of telemedicine, telerehabilitation is defined as the use of remote health care methodologies, specifically information and communications technologies, to deliver therapeutic rehabilitation services outside of medical institutions [6]. Telerehabilitation programs use information and communications technologies such as telephones and videoconferencing to facilitate exercise training, self-management, and lifestyle modifications for nonhospitalized patients [7,8].

In recent years, several systematic reviews have been published regarding the effectiveness of telerehabilitation programs for patients with specific chronic diseases, including cardiac, respiratory, and neurological disorders [9,10]. Of note, 2 previous systematic reviews have examined the effectiveness of postoperative telerehabilitation in the recovery of patients with hip fractures, yielding inconsistent results. One review found that, compared to conventional intervention, telerehabilitation was not effective in improving postoperative mobility (as measured using the timed up and go [TUG] test and gait speed) or in alleviating pain but facilitated the ability to perform basic activities of daily living (BADLs) with a clinically irrelevant improvement [11]. Another recently published systematic review found that, compared to traditional rehabilitation, digital health interventions significantly improved postoperative mobility and the ability to perform BADLs, as indicated by the improvement in the TUG performance, the Short Physical Performance Battery (SPPB) score, and the Functional Independence Measure (FIM) score [12]. However, specific measurement tools for hip function, such as the Harris Hip Score (HHS), have not been examined in any systematic review. This is a significant research gap because these tools assess crucial clinical domains, including pain, range of motion, and deformity, that are key outcomes for patients with hip fractures. Moreover, the long-term effects of telerehabilitation on postoperative functional recovery in patients with hip fractures have not been investigated.

Therefore, this systematic review aimed to evaluate the short- and long-term effects of telerehabilitation on postoperative functional recovery in patients with hip fractures. The outcomes of interest included hip function as measured using the HHS [13]; functional mobility as measured using the TUG test [14,15] and SPPB [16-18]; and ability to perform BADLs as measured using the Barthel index (BI), modified BI (MBI), or FIM. We hypothesized that telerehabilitation would be effective for improving postoperative short- and long-term functional recovery in patients with hip fractures compared to conventional rehabilitation.

Methods

This systematic review and meta-analysis adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines [19], with a registration in PROSPERO (CRD42024498569).

Literature Search

Two researchers (CC and PPQ) independently searched the PubMed, Embase, Web of Science, and CENTRAL databases from inception to March 31, 2025. The search was limited to English-language literature. Randomized controlled trials investigating the effectiveness of postoperative telerehabilitation treatment for patients with hip fractures were included (Multimedia Appendix 1).

Additional articles were identified by reviewing the reference lists of relevant reviews. The following keywords and the corresponding MeSH (Medical Subject Headings) terms were used: (telerehabilitation OR telerehabilitations OR tele-rehabilitation OR tele-rehabilitations OR remote rehabilitation OR remote rehabilitations OR rehabilitation, remote OR rehabilitations, remote OR virtual rehabilitation OR virtual rehabilitations OR rehabilitation, virtual OR rehabilitations, virtual) AND (hip fractures OR fractures, hip OR intertrochanteric fracture OR intertrochanteric fractures OR trochanteric fractures OR fractures, trochanteric OR femur trochlear fracture OR femur trochlear fractures OR trochlear fracture, femur OR trochlear fractures, femur OR fractures, femur trochlear OR femoral trochlear fracture OR femoral trochlear fractures OR fracture, femoral trochlear OR fractures, femoral trochlear OR trochlear fracture, femoral OR trochlear fractures, femoral OR subtrochanteric fracture OR subtrochanteric fractures).

Eligibility Criteria

The inclusion criteria were as follows: (1) studies involving participants diagnosed with a hip fracture who were in a stable condition (no further fractures were observed or reported) after surgical intervention; (2) studies involving telerehabilitation as the primary intervention following surgery, including the use of SMS text messaging, telephones, smartphones, and internet-based and software-based methods; (3) studies that reported at least one outcome of interest, including the HHS, TUG test, SPPB score, BI, MBI, and FIM; and (4) randomized controlled trials.

The exclusion criteria were (1) only study protocol or conference abstract available and (2) unavailability of the full text.

XSL•FO
RenderX

Study Selection

The search results were imported into EndNote X9 (Clarivate Analytics) for duplicate removal. To ensure the accuracy of this process, 2 authors (CC and PPQ) manually verified the deleted duplicates. The titles and abstracts of the remaining records were then screened independently by these authors (CC and PPQ). Full texts were retrieved for further review if deemed eligible by either author. Any disagreements between the reviewers were resolved through discussion with a third author (CL).

Data Extraction

Two authors (CC and PPQ) independently extracted data from the included studies. The extracted data encompassed the following: basic study information (the first author, publication year, and country), participant baseline characteristics (sample size, age, and gender), details of the telerehabilitation and control interventions, and outcome measures (including follow-up period and measurement tools).

Quality Evaluation

Two authors (CC and PPQ) independently evaluated the quality of the included studies using version 2 of the Cochrane risk-of-bias tool for randomized trials [20]. The risk-of-bias assessment consisted of the following domains: (1) randomization process, (2) deviations from intended interventions, (3) missing outcome data, (4) measurement of the outcome, and (5) selection of the reported results. The overall risk of bias was classified as low risk of bias, some concerns, and high risk of bias.

Data Analysis

The Cochrane Collaboration's Review Manager software (version 5.3) was used for all statistical analyses. The overall estimate of the treatment effect was calculated using the means and SDs of outcome scores for continuous data in the telerehabilitation intervention and control groups. Short- and long-term effects were analyzed by comparing the differences in outcome scores between the 2 groups at the intervention end point and at the last follow-up time point, respectively. For studies using the same measurement tool, we calculated a pooled estimate of the mean differences (MDs) with 95% CIs. In cases in which different measurement tools were used, we used standardized MDs (SMDs).

To investigate potential clinical and methodological heterogeneity across the included studies, the following factors were summarized and compared: (1) patient characteristics (age and baseline scores of the outcomes of interest), (2) characteristics of telerehabilitation (duration, frequency, and delivery mode), (3) type of control intervention, (4) follow-up period for the outcomes of interest, and (5) risk of bias. Where low clinical and methodological heterogeneity was found across the studies in a given meta-analysis, a fixed-effects model was used; otherwise, a random-effects model was applied. Statistical heterogeneity was quantified using the I^2 statistic. I^2 values greater than 50% were considered to indicate substantial heterogeneity [21]. When substantial statistical heterogeneity was present and a sufficient number of studies were available, subgroup analyses were performed based on the abovementioned factors contributing to clinical or methodological heterogeneity. Furthermore, a sensitivity analysis was planned to assess the impact on the pooled results provided that a sufficient number of studies were available. This was performed by excluding studies with a high risk of bias or those that differed substantially from others in the factors considered for subgroup analyses.

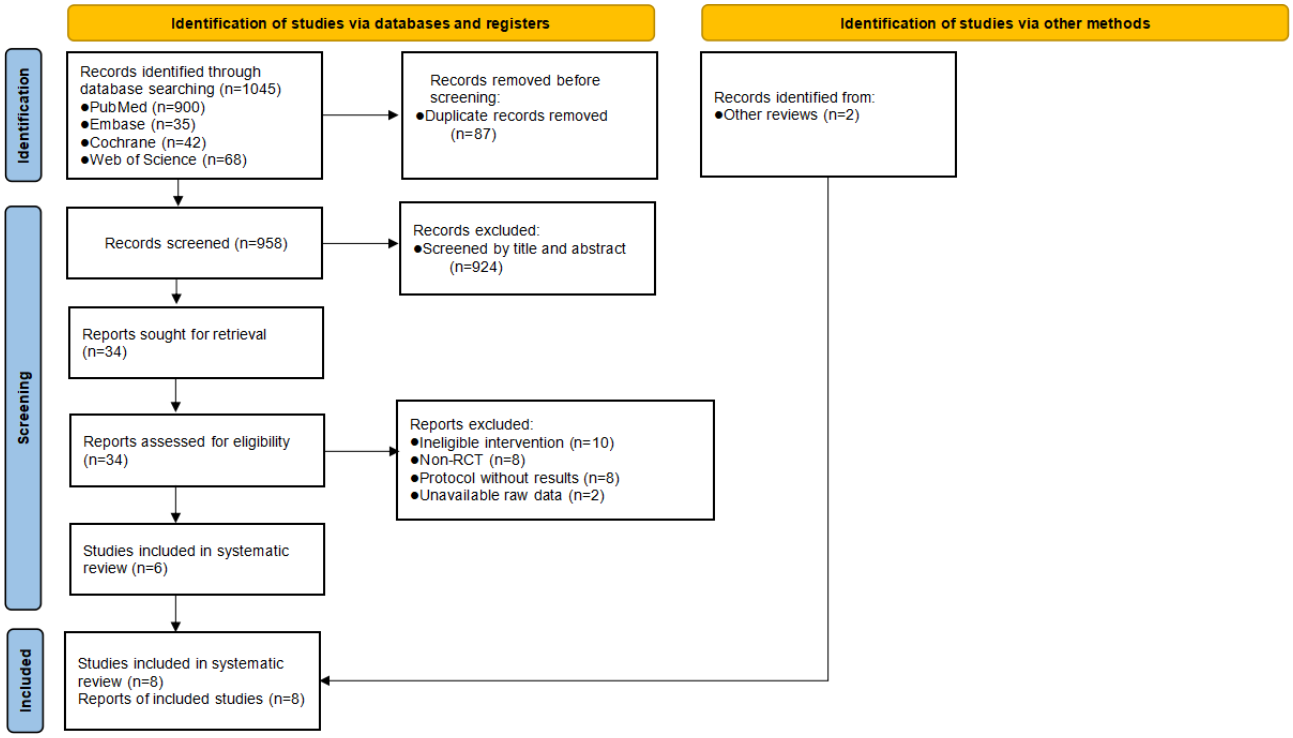
Results

Study Identification

The initial database search yielded a total of 1045 records, of which 6 (0.6%) met the eligibility criteria and were included. A further 2 studies were identified by checking the reference lists of relevant reviews [11], resulting in a total of 8 studies for the systematic review and meta-analysis. The study selection process is illustrated in the PRISMA flow diagram (Figure 1).

XSL•FO
RenderX

Figure 1. Flow diagram for the systematic search. RCT: randomized controlled trial.



Study Characteristics

The 8 included studies comprised 740 patients, including 569 female and 161 male individuals. The sample sizes varied from 31 to 232 in the individual studies.

The treatment strategies for postoperative telerehabilitation for patients with hip fractures included strengthening, balance, coordination, stretching exercises, BADL training, introduction of walking aids, and safety education. Most included studies provided telerehabilitation through mobile apps or software [5,22-28] supervised by professionals remotely, one study used a DVD supervised via telephone calls [27], and one study used a telephone call as the primary strategy [28]. The control interventions comprised home-based exercise through written materials [5,25,26], in-person rehabilitation [23,24], follow-up assessment through telephone [22] and health nutrition education [27], and no intervention [28]. The duration of telerehabilitation varied from 35 minutes to 6 months. The follow-up periods varied from 3 weeks to 9 months [23,25-28]. Table 1 presents the main characteristics of the included studies [5,22-28].

XSL•FO
RenderX

Table . Characteristics of the included studies.

Study and country	Participants	Telerehabilitation	Control intervention	Outcome measures	Time points of outcome measurement
Kalron et al [26], 2018, Israel	Sample size: 15 in telerehabilitation vs 17 in control group (female: 10/15 vs 9/17, respectively); age: mean 65.7, SD 7.8 y vs mean 67.3, SD 9.5 y	• Home-based training program through software, including exercises for movement, strengthening the lower limbs, and balance; 40-50 min per session, 3 times per week for 6 weeks • Supervised by professionals through the software	Home-based exercises through an exercise booklet (exercises were similar to those in the telerehabilitation group)	TUG[a] test	Intervention end point: 6 weeks; follow-up: 4 weeks after the intervention
Li et al [25], 2022, China	Sample size: 15 in telerehabilitation vs 16 in control group (female: 14/15 vs 11/16, respectively); age: mean 76.5, SD 8.6 y vs mean 82.1, SD 9.7 y	• Home-based training program through mobile app, including trunk and lower-extremity strengthening and stretching, coordination, balance, and functional exercises related to ADLs[b], for 3 weeks • Supervised by professionals through the mobile app	Home-based training through written sheets (exercises were equivalent to those in the telerehabilitation group)	TUG test and MBI[c]	Intervention end point: 3 weeks; follow-up: 3 weeks after the intervention
Ortiz-Piña et al [24], 2021, Spain	Sample size: 35 in telerehabilitation vs 36 in control group (female: 26/35 vs 27/36, respectively); age: mean 76.71, SD 6.04 y vs mean 80.72, SD 5.59 y	• Home-based training through software, including lower- and upper-body strengthening, balance, cardiovascular exercises, and occupational therapy (ADL training and introduction to walking aids); 50-60 min per session for 12 weeks • Supervised by caregivers	Home-based in-person rehabilitation, including 5-15 sessions of physiotherapy and occupational therapy	TUG test, SPPB[d], and FIM[e]	Intervention end point: 12 weeks; follow-up: N/A[f]

XSL•FO
RenderX

Study and country	Participants	Telerehabilitation	Control intervention	Outcome measures	Time points of outcome measurement
Wu et al [5], 2022, China	Sample size: 43 in telerehabilitation vs 42 in control group (female: 30/43 vs 31/42, respectively); age: mean 74.28, SD 5.06 y vs mean 72.0, SD 6.77 y	• Home-based training through software, including cardiopulmonary function exercise and balance exercises, ADL training, and introduction to walking aids, for 6 months • Supervised by professionals through the software and communication with professionals through videoconferencing	Home-based rehabilitation through a written text and communication with professionals through phone calls and outpatient service	HHS[g] and FIM	Intervention end point: 6 months; follow-up: N/A
Zhang et al [22], 2022, China	Hip replacement group sample size: 17 in telerehabilitation vs 14 in control group; internal fixation group sample size: 10 in telerehabilitation vs 10 in control group	• Home-based training through a mobile app, including education and walking, balance, and stair-walking exercises, for 3 months • Supervised by professionals through the mobile app	Follow-up assessment through telephone	HHS, FIM, TUG test, and SPPB	During the intervention period: at 2 weeks and 2 months; intervention end point: 3 months; follow-up: N/A
Prieto-Moreno et al [23], 2024, Portugal	Sample size: 51 in telerehabilitation vs 54 in control group (female: 37/51 vs 38/54, respectively); age: mean 79.55, SD 7.11 y vs mean 80.07, SD 7.74 y	• Home-based training through software, including lower- and upper-body strengthening, balance, cardiovascular exercises, and occupational therapy (ADL training and introduction to walking aids); 3 sessions per week for 12 weeks • Supervised by professionals through the software	Home-based in-person rehabilitation, including 5-15 sessions of physiotherapy and occupational therapy	SPPB and FIM	Intervention end point: 3 months; follow-up: 9 months after the intervention
Latham et al [27], 2014, United States	Sample size: 120 in telerehabilitation vs 112 in control group (female: 83/120 vs 77/112, respectively); age: mean 77.2, SD 10.2 y vs mean 78.9, SD 9.4 y	• Home-based training through DVD, including strengthening and standing exercises; 3 times per week for 6 months • Supervised by professionals through telephone	Nutrition education for cardiovascular health through telephone	SPPB	Intervention end point: 6 months; follow-up: 3 months after the intervention
			No intervention		

XSL•FO
RenderX

Study and country	Participants	Telerehabilitation	Control intervention	Outcome measures	Time points of outcome measurement
Di Monaco et al [28], 2015, Italy	Sample size: 78 in telerehabilitation vs 75 in control group (female: 78/78 vs 75/75, respectively); age: mean 78.7, SD 7.2 y vs mean 79.3, SD 8.0 y	• A telephone call to check environmental hazards, behaviors in ADLs, and use of assistive devices and reinforced targeted modifications to prevent falls for 35 minutes		BI[h]	Intervention end point: N/A; follow-up: 5.5 months after the intervention

[a]TUG: timed up and go.

[b]ADL: activity of daily living.

[c]MBI: modified Barthel index.

[d]SPPB: Short Physical Performance Battery.

[e]FIM: Functional Independence Measure.

[f]N/A: not applicable.

[g]HHS: Harris Hip Score.

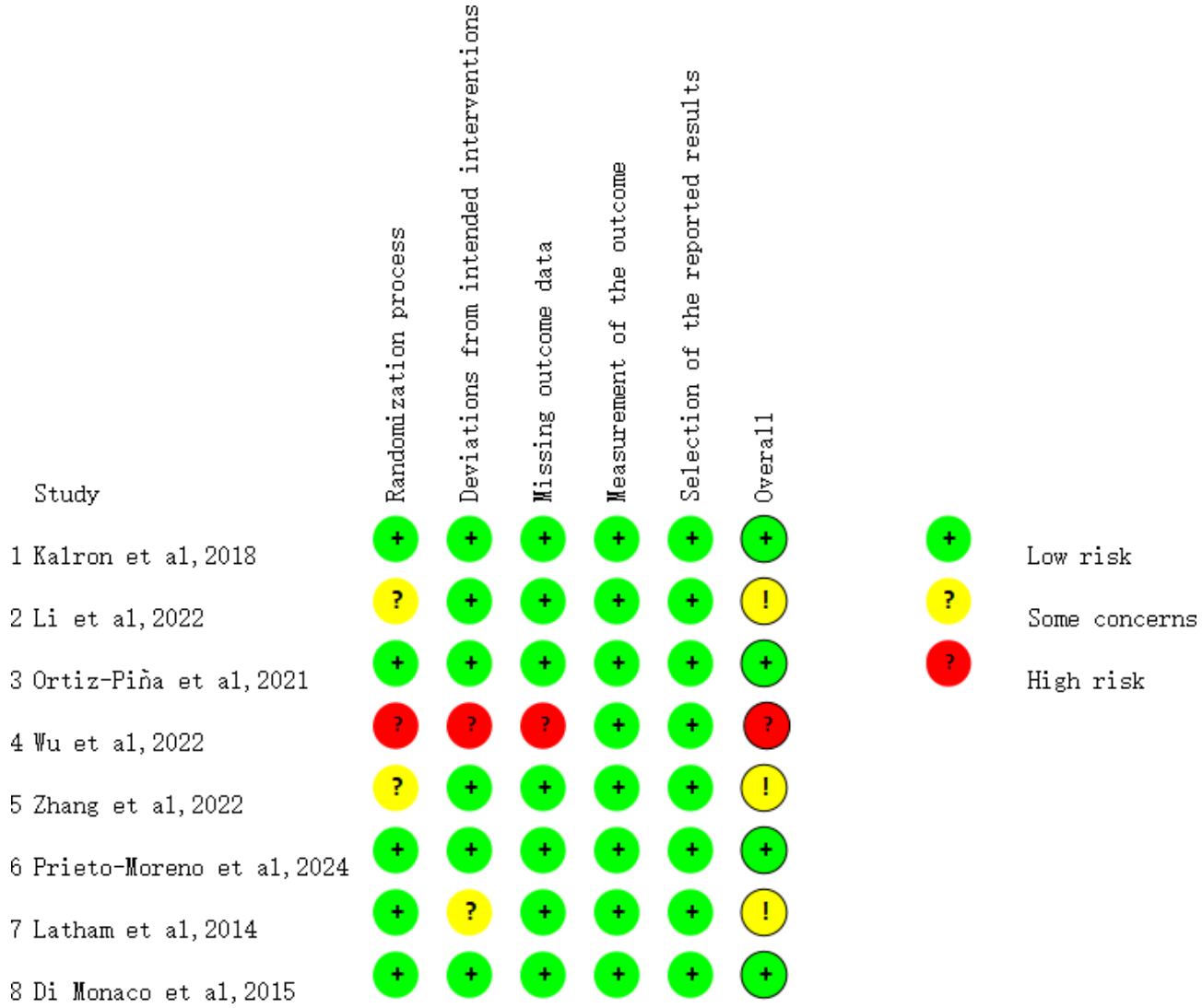
[h]BI: Barthel index.

Risk of Bias in the Included Studies

Four studies [23,24,26,28] had a low risk of bias, 3 [22,25,27] had some concerns, and 1 [5] had a high risk of bias. Overall, the randomization process was the most problematic domain as 3 studies poorly described the method used [5,22,25]. The results of the methodological quality assessment are provided in Figure 2 [5,22-28].

XSL•FO
RenderX

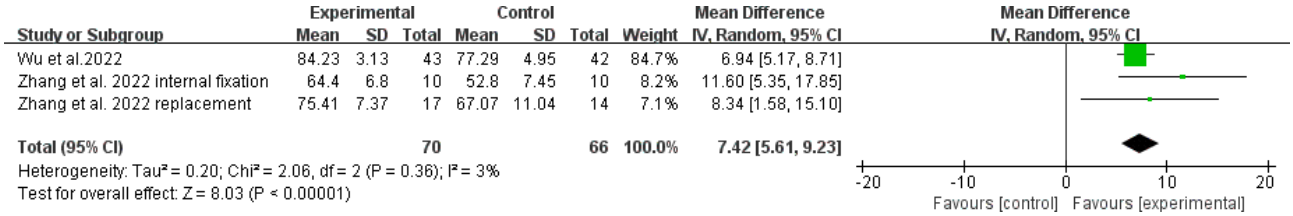
Figure 2. A detailed breakdown of the methodological quality assessment for the included studies. Overall, 4 studies exhibited a low risk of bias, 3 raised some concerns, and 1 exhibited a high risk of bias [5,22-28].



Effectiveness of Postoperative Telerehabilitation on Functional Recovery at the Intervention End Point and Follow-Up

HHS was measured in 2 studies involving a total of 156 participants at the intervention end point [5,22]. Low clinical heterogeneity was observed in patient characteristics (mean age range 72-77 years; comparable baseline mean HHS between 34.9 and 43.6). However, substantial clinical heterogeneity was found regarding the duration of telerehabilitation (6 months [5] vs 3 months [22]). A random-effects meta-analysis yielded a statistically significant, positive result (MD 7.42, 95% CI 5.61-9.23; *P*<.001; Figure 3 [5,22]) with low statistical heterogeneity (I^2=3%; *P*=.36; Figure 3 [5,22]).

Figure 3. Forest plot of Harris Hip Score (HHS) results at the intervention end point. The analysis yielded a positive result with a mean difference of 7.42 (95% CI 5.61-9.23; *P*<.001; I^2=3%) [5,22]. IV: inverse variance.

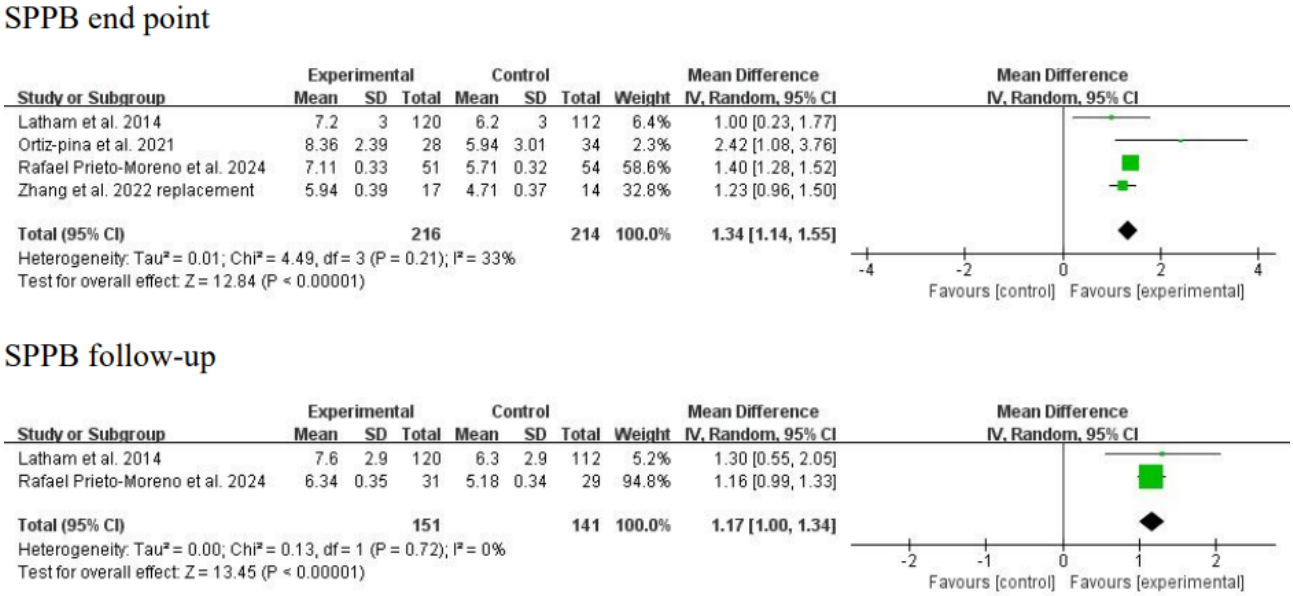


SPPB was measured in 4 studies (n=430 participants) at the intervention end point [22-24,27]. Considerable clinical heterogeneity was observed in baseline SPPB scores (mean scores 4.71-5.9 [22,27] vs 2.69-3.21 [23,24]). Furthermore, both the duration of telerehabilitation (3 months [22-24] vs 6 months [27]) and the type of control intervention (telephone assessment [22,27] vs in-person rehabilitation [23,24]) varied across studies. A random-effects meta-analysis showed a significant positive result at the intervention end point (MD 1.34, 95% CI 1.14-1.55; *P*<.001; Figure 4 [22-24,27]) with low to moderate statistical heterogeneity (I^2=33%; *P*=.21; Figure 4 [22-24,27]). SPPB was

XSL•FO
RenderX

measured in 2 studies (n=292 participants) at follow-up [23,27]. In addition to the aforementioned clinical heterogeneity, great methodological heterogeneity was noted in the follow-up period (3 months [27] vs 9 months after the intervention [23]). The random-effects analysis remained significant at follow-up (MD 1.17, 95% CI 1.00-1.34; *P*<.001; Figure 4 [22-24,27]) with low statistical heterogeneity (I^2=0%; *P*=.72; Figure 4 [22-24,27]).

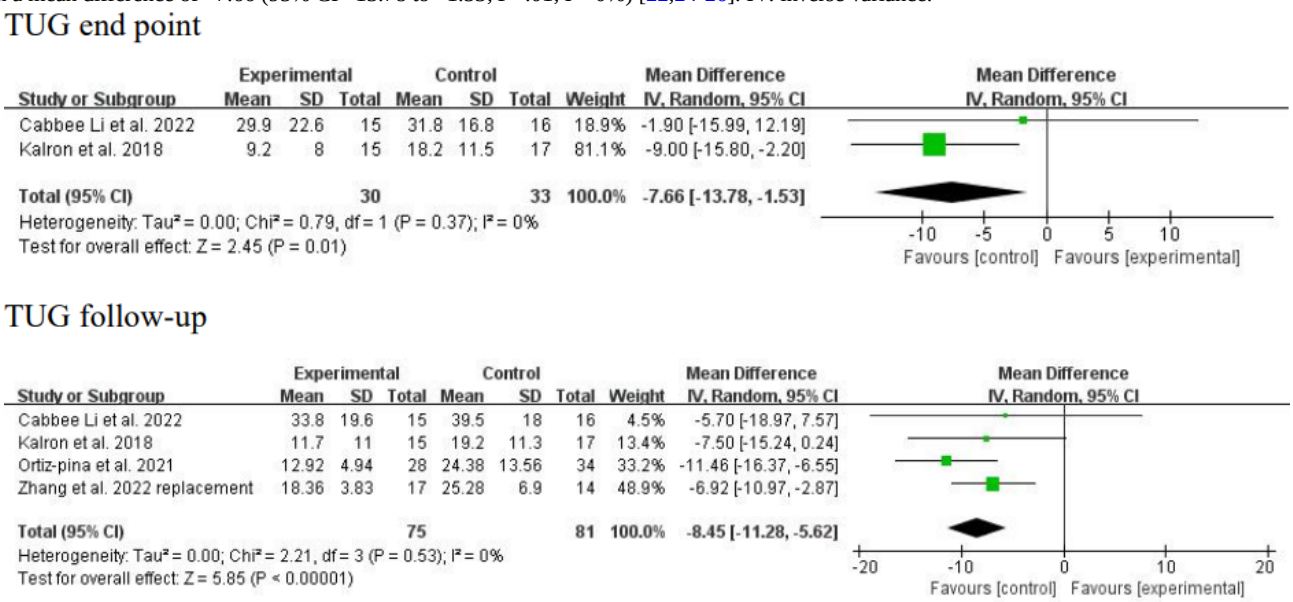
Figure 4. Forest plot of Short Physical Performance Battery (SPPB) results at the intervention end point and follow-up. The analysis yielded a positive result at the intervention end point with a mean difference of 1.34 (95% CI 1.14-1.55; *P*<.001; I^2=33%). The analysis yielded a positive result at follow-up with a mean difference of 1.17 (95% CI 1.00-1.34; *P*<.001; I^2=0%) [22-24,27]. IV: inverse variance.



The TUG test was measured in 4 studies (n=156 participants) at the intervention end point [22,24-26]. Considerable clinical heterogeneity was found regarding the patient characteristics (mean age 65.7-67.3 years [26] vs 75.17-82.1 years [22,24,25]; mean baseline TUG score 21.4-25.6 [26] vs 39.7-45.2 [25] vs 66.53-99.72 [24]; data unavailable for the study by Zhang et al [22]). The duration of telerehabilitation (3-6 weeks [25,26] vs 3 months [22,24]) and type of control intervention (written materials [25,26] vs telephone assessment [22] vs in-person rehabilitation [24]) also varied. A random-effects meta-analysis demonstrated a significant positive result at the intervention end point (MD –8.45, 95% CI −11.28 to −5.62; *P*<.001; Figure 5) with low statistical heterogeneity (I^2=0%; *P*=.53; Figure 5 [22,24-26]). The TUG test was measured in 2 studies (n=63 participants) at follow-up [25,26]. Low methodological heterogeneity was observed in the follow-up period (3-4 weeks after the intervention [25,26]) and risk of bias (low to moderate risk of bias [25,26]). However, due to the considerable clinical heterogeneity noted above (in patient age and baseline function), a random-effects model was applied. This analysis yielded a significant positive result at follow-up (MD −7.66, 95% CI −13.78 to −1.53; *P*=.001; I^2=0%; *P*=.37; Figure 5 [22,24-26]).

XSL•FO
RenderX

Figure 5. Forest plot of timed up and go (TUG) test results at the intervention end point and follow-up. The analysis yielded a positive result at the intervention end point with a mean difference of –8.45 (95% CI –11.28 to –5.62; *P*<.001; I^2=0%). The analysis yielded a positive result at follow-up with a mean difference of –7.66 (95% CI –13.78 to –1.53; *P*=.01; I^2=0%) [22,24-26]. IV: inverse variance.



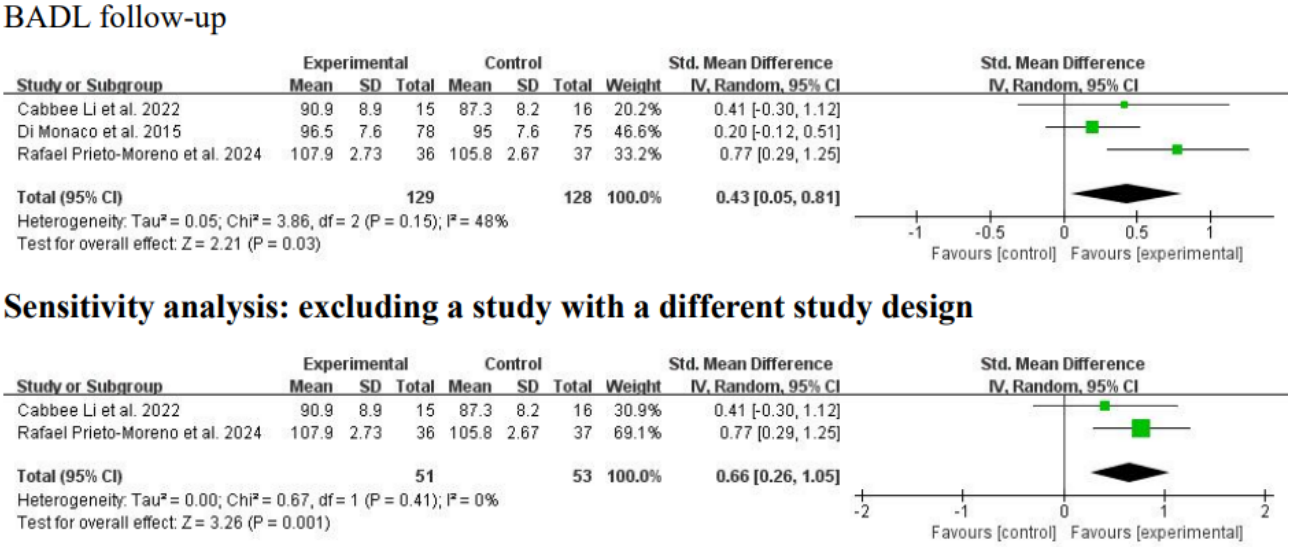
Because various BADL measurement tools were used across the included studies, SMDs were used as the effect size measure. Zhang et al [22] reported BADL outcomes separately for the hip replacement and internal fixation groups. BADLs were measured in 5 studies (n=354 participants) at the intervention end point [5,22-25]. Considerable clinical heterogeneity was observed regarding patient characteristics (baseline dependence level: maximal dependence [5,22] vs mild to moderate dependence [23-25]), duration of telerehabilitation (3 weeks [25] vs 3 months [22-24] vs 6 months [5]), and type of control intervention (written materials [25] vs telephone assessment [5,22] vs in-person rehabilitation [23,24]). In addition, substantial methodological heterogeneity was observed in the risk of bias (high [5] vs some concerns [22,25] vs low [23,24]). A random-effects meta-analysis yielded a significant positive result with high heterogeneity (SMD 1.65, 95% CI 0.78-2.51; *P*<.001; I^2=91%; *P*<.001; Figure 6 [5,22-25]). Subgroup analyses were conducted based on baseline dependence level, duration of telerehabilitation, and type of control intervention. For studies involving patients with a high baseline dependence level [5,22], the analysis showed a positive result with high heterogeneity (SMD 1.87, 95% CI 0.59-3.16; *P*=.004; I^2=89%; *P*<.001; Figure 6 [5,22-25]). For studies with a lower baseline dependence level [23-25], the result was also positive with high heterogeneity (SMD 1.42, 95% CI 0.04-2.81; *P*=.04; I^2=94%; *P*<.001; Figure 6 [5,22-25]). Subgroup analysis by duration of telerehabilitation was not possible due to an insufficient number of studies. For studies with telephone-based assessment as a control intervention [5,22], the result was positive with high heterogeneity (SMD 1.87, 95% CI 0.59-3.16; *P*=.004; I^2=89%; *P*<.001; Figure 6 [5,22-25]). For studies with in-person rehabilitation as the control [23,24], the result was similarly positive with high heterogeneity (SMD 1.90, 95% CI 0.16-3.63; *P*=.03; I^2=95%; *P*<.001; Figure 6 [5,22-25]). A sensitivity analysis was conducted by (1) excluding 2 studies with intervention durations shorter [25] or longer [5] than 3 months and (2) excluding 1 study with a high risk of bias [5]. Pooling the 3 remaining studies with a 3-month intervention duration [22-24] using a random-effects model yielded a positive result with low statistical heterogeneity (SMD 1.13, 95% CI 0.76-1.50; *P*<.001; I^2=0%; *P*=.82; Figure 6 [5,22-25]). Pooling the 4 studies with low to moderate risk of bias [22-25] yielded a positive result with high heterogeneity (SMD 1.36, 95% CI 0.52-2.21; *P*=.002; I^2=88%; *P*<.001). BADLs were measured in 3 studies (n=257 participants) at follow-up [23,25,28]. Considerable clinical heterogeneity was observed regarding baseline dependence level (mild to moderate dependence [23,25] vs minimal dependence [28]), duration of telerehabilitation (2.5-3 weeks [25,28] vs 12 weeks [23]), delivery mode of telerehabilitation (telephone based [28] vs software based [23,25]), and type of control intervention (no intervention [28] vs in-person rehabilitation [23] vs written materials [25]). Moreover, substantial methodological heterogeneity was found in the follow-up period (3 weeks [25] vs 5.5 months [28] vs 9 months [23] after the intervention). A random-effects model yielded a positive result with moderate heterogeneity at follow-up (SMD 0.43, 95% CI 0.05-0.81; *P*=.03; I^2=48%; *P*=.15; Figure 7 [23,25,28]). A sensitivity analysis excluded 1 study [28] that differed substantially from the others in design (baseline dependence level, type of telerehabilitation, and type of control intervention). A random-effects model for the remaining 2 studies [23,25] yielded a positive result with low heterogeneity (SMD 0.66, 95% CI 0.26-1.05; *P*=.001; I^2=0%; *P*=.41; Figure 7 [23,25,28]).

XSL•FO
RenderX

Figure 6. Forest plot of basic activity of daily living (BADL) results at the intervention end point. The analysis yielded a positive result at the intervention end point with a standardized mean difference of 1.65 (95% CI 0.78-2.51; P<.001) and high heterogeneity (I^2=91%). Subgroup analyses were conducted based on baseline dependence level and type of control intervention. A sensitivity analysis was conducted by excluding a study with a high risk of bias [5,22-25]. IV: inverse variance.

XSL•FO
RenderX

Figure 7. Forest plot of basic activity of daily living (BADL) results at follow-up. The analysis yielded a positive result at follow-up with a standardized mean difference of 0.43 (95% CI 0.05-0.81; P=.03; I^2=48%). A sensitivity analysis was conducted by excluding a study with a different study design from the others [23,25,28]. IV: inverse variance.



Discussion

Principal Findings

Our review found that postoperative telerehabilitation was effective in improving short- and long-term hip function, functional mobility, and ability to perform BADLs in patients with hip fractures compared to conventional rehabilitation. However, current evidence was weak due to the limited number and insufficient quality of the included studies and the heterogeneity across studies.

Our findings align with previous evidence that telerehabilitation is beneficial to short-term recovery of functional mobility and ability to perform BADLs at posttreatment compared to usual care [12]. Moreover, our review found that hip function was also improved after telerehabilitation. In addition, our findings revealed the effect of telerehabilitation on long-term recovery (3 weeks to 9 months after the intervention) in functional mobility and the ability to perform BADLs. The effectiveness in improving BADLs showed substantial heterogeneity across the included studies in the postintervention analysis. Although subgroup analyses were conducted based on the type of control intervention and the baseline dependence level, they failed to reduce the heterogeneity. A potential source of heterogeneity may be the duration of telerehabilitation as the statistical heterogeneity decreased dramatically when 2 studies with telerehabilitation shorter (3 weeks) [25] and longer (6 months) [5] than 3 months were excluded. Another potential source of heterogeneity may be differences in adherence to the telerehabilitation programs. Although only 2 of the 5 studies in the BADL analysis reported adherence rates for telerehabilitation, the substantial difference (87% [25] vs 15% [24]) suggests considerable variability in compliance across studies.

This systematic review revealed that telerehabilitation strategies, including exercises for strengthening the lower extremities, balance, coordination, stretching, and BADL training, introduction of walking aids, and safety education are commonly used for patients with hip fractures. According to clinical practice guidelines, postoperative progressive resistance exercises and balance training are strongly recommended, and BADL training and complication prevention are suggested for older patients with hip fractures [29]. However, it is essential to standardize these training programs to mitigate the risk of adverse events [30,31].

Regarding the delivery mode of telerehabilitation, most studies used software- or mobile app–based telerehabilitation with remote supervision by professionals, and only 2 studies conducted before 2015 used DVD-based or telephone-based telerehabilitation. Recently, the application of virtual reality has become an increasingly common approach in telerehabilitation, with superior long-term benefits for patients with total knee replacements compared with traditional rehabilitation [32]. Although all these remote rehabilitation approaches could facilitate adherence to postoperative management, reduce costs, and decrease the burden on therapists and caregivers, the efficacy of different delivery strategies could vary. Future studies are needed to compare the cost-effectiveness ratio between different delivery modes of telerehabilitation.

Our findings may have several clinical implications. First, substantial clinically important improvements in functional mobility, as measured using the SPPS [33] and TUG test [34,35], were observed following telerehabilitation compared to usual care at the postintervention and follow-up time points. However, the overall improvement in HHS (7.35 points) fell below the minimal clinically important improvement threshold (15.9-18 points) [36]. For BADLs, a large short-term effect size (SMD 1.65) was found, but this decreased to a small to moderate effect at the long-term follow-up (SMD 0.43). Furthermore, no included studies demonstrated a clinically relevant improvement for BADLs (21 points on the FIM [37] and 9.8 points on the BI [38]). These findings indicate that telerehabilitation may be more beneficial for functional mobility than for hip function or BADLs. Second, we noted that most included studies were conducted in countries with established research environments

XSL•FO
RenderX

(Israel, Spain, Portugal, and the United States), with China being the only included country with an emerging research landscape. Therefore, the generalizability of telerehabilitation requires further investigation, specifically in terms of its feasibility and effectiveness in low- and middle-income countries. Third, the included studies exhibited variable follow-up periods (3 weeks to 9 months), which impedes the investigation of long-term effects. Future studies could adopt a more standardized follow-up schedule (eg, every 3 months after the intervention). Fourth, adherence to telerehabilitation was poorly reported in the included studies, leaving its influence on the effect size unclear. As a key proposed benefit of telerehabilitation is the facilitation of prolonged rehabilitation training—and studies have demonstrated higher compliance in telerehabilitation groups, potentially leading to greater improvement [22,26]—detailed adherence data are crucial for interpreting effectiveness. Future studies should report adherence and compliance rates in detail and develop effective supervision approaches to ensure the efficacy of telerehabilitation.

Limitations

This study has several limitations. First, it only included literature published in English, which may have omitted relevant evidence from articles published in other languages, potentially introducing language bias and limiting the generalizability of the findings. Second, the small number of included studies and the substantial heterogeneity observed influenced the strength of the evidence. Third, no statistical analysis for publication bias was conducted due to the small number of included studies.

Conclusions

Telerehabilitation is effective for short- and long-term functional recovery in postoperative patients with hip fractures compared to conventional rehabilitation. However, the evidence is weak due to the limited number of included studies and the high heterogeneity across studies. Future high-quality studies with larger sample sizes are needed to investigate the effectiveness of telerehabilitation. These studies should report the intervention strategies, delivery modes, and adherence rates of interventions in detail; incorporate comprehensive outcome measures for hip function (eg, HHS), functional mobility (eg, SPPS and TUG test), and ability to perform BADLs (eg, BI, MBI, and FIM); and apply a relatively standard follow-up period to investigate the long-term effect of telerehabilitation.

Funding

This work was supported by the Shenzhen Science and Technology Program (JCYJ20240813145301003 awarded to XW).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy for PubMed.
[DOCX File, 25 KB - mhealth_v14i1e77341_app1.docx]

Checklist 1
PRISMA checklist.
[PDF File, 99 KB - mhealth_v14i1e77341_app2.pdf]

References

1. Sing CW, Lin TC, Bartholomew S, et al. Global epidemiology of hip fractures: secular trends in incidence rate, post-fracture treatment, and all-cause mortality. J Bone Miner Res 2023 Aug;38(8):1064-1075. [doi: 10.1002/jbmr.4821] [Medline: 37118993]
2. Tebé C, Martínez-Laguna D, Carbonell-Abella C, et al. The association between type 2 diabetes mellitus, hip fracture, and post-hip fracture mortality: a multi-state cohort analysis. Osteoporos Int 2019 Dec;30(12):2407-2415. [doi: 10.1007/s00198-019-05122-3] [Medline: 31444526]
3. Handoll HH, Sherrington C, Mak JC. Interventions for improving mobility after hip fracture surgery in adults. Cochrane Database Syst Rev 2011 Mar 16;9(3):CD001704. [doi: 10.1002/14651858.CD001704.pub4] [Medline: 21412873]
4. Handoll HH, Cameron ID, Mak JC, Panagoda CE, Finnegan TP. Multidisciplinary rehabilitation for older people with hip fractures. Cochrane Database Syst Rev 2021 Nov 12;11(11):CD007125. [doi: 10.1002/14651858.CD007125.pub3] [Medline: 34766330]
5. Wu WY, Zhang YG, Zhang YY, Peng B, Xu WG. Clinical effectiveness of home-based telerehabilitation program for geriatric hip fracture following total hip replacement. Orthop Surg 2023 Feb;15(2):423-431. [doi: 10.1111/os.13521] [Medline: 36448261]
6. Duruturk N. Telerehabilitation intervention for type 2 diabetes. World J Diabetes 2020 Jun 15;11(6):218-226. [doi: 10.4239/wjd.v11.i6.218] [Medline: 32547696]

XSL•FO
RenderX

7. Peretti A, Amenta F, Tayebati SK, Nittari G, Mahdi SS. Telerehabilitation: review of the state-of-the-art and areas of application. JMIR Rehabil Assist Technol 2017 Jul 21;4(2):e7. [doi: 10.2196/rehab.7511] [Medline: 28733271]
8. Wong AKC, Wong FKY, Chow KKS, Wong SM, Lee PH. Effect of a telecare case management program for older adults who are homebound during the COVID-19 pandemic: a pilot randomized clinical trial. JAMA Netw Open 2021 Sep 1;4(9):e2123453. [doi: 10.1001/jamanetworkopen.2021.23453] [Medline: 34499135]
9. Lee AYL, Wong AKC, Hung TTM, Yan J, Yang S. Nurse-led telehealth intervention for rehabilitation (telerehabilitation) among community-dwelling patients with chronic diseases: systematic review and meta-analysis. J Med Internet Res 2022 Nov 2;24(11):e40364. [doi: 10.2196/40364] [Medline: 36322107]
10. Truijen S, Abdullahi A, Bijsterbosch D, et al. Effect of home-based virtual reality training and telerehabilitation on balance in individuals with Parkinson disease, multiple sclerosis, and stroke: a systematic review and meta-analysis. Neurol Sci 2022 May;43(5):2995-3006. [doi: 10.1007/s10072-021-05855-2] [Medline: 35175439]
11. Tsuge T, Yamamoto N, Taito S, Miura T, Shiratsuchi D, Yorifuji T. Efficacy of telerehabilitation for patients after hip fracture surgery: a systematic review and meta-analysis. J Telemed Telecare 2025 Feb;31(2):174-183. [doi: 10.1177/1357633X231181632] [Medline: 37416946]
12. Pliannuom S, Pinyopornpanish K, Buawangpong N, et al. Characteristics and effects of home-based digital health interventions on functional outcomes in older patients with hip fractures after surgery: systematic review and meta-analysis. J Med Internet Res 2024 Jun 12;26:e49482. [doi: 10.2196/49482] [Medline: 38865706]
13. Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. J Bone Joint Surg Am 1969 Jun;51(4):737-755. [Medline: 5783851]
14. Podsiadlo D, Richardson S. The timed “Up & Go”: a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991 Feb;39(2):142-148. [doi: 10.1111/j.1532-5415.1991.tb01616.x] [Medline: 1991946]
15. Bohannon RW. Reference values for the timed up and go test: a descriptive meta-analysis. J Geriatr Phys Ther 2006;29(2):64-68. [doi: 10.1519/00139143-200608000-00004] [Medline: 16914068]
16. Volpato S, Cavalieri M, Sioulis F, et al. Predictive value of the short physical performance battery following hospitalization in older patients. J Gerontol A Biol Sci Med Sci 2011 Jan;66(1):89-96. [doi: 10.1093/gerona/glq167] [Medline: 20861145]
17. Soubra R, Chkeir A, Novella JL. A systematic review of thirty-one assessment tests to evaluate mobility in older adults. Biomed Res Int 2019;2019:1354362. [doi: 10.1155/2019/1354362] [Medline: 31321227]
18. Pavasini R, Guralnik J, Brown JC, et al. Short physical performance battery and all-cause mortality: systematic review and meta-analysis. BMC Med 2016 Dec 22;14(1):215. [doi: 10.1186/s12916-016-0763-7] [Medline: 28003033]
19. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev 2021 Mar 29;10(1):89. [doi: 10.1186/s13643-021-01626-4] [Medline: 33781348]
20. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
21. Higgins JP, Thomas J, Chandler J, et al. Cochrane Handbook for Systematic Reviews of Interventions Version 65: Cochrane; 2024.
22. Zhang YY, Zhang YG, Li Z, Li SH, Xu WG. Effect of home-based telerehabilitation on the postoperative rehabilitation outcome of hip fracture in the aging population. Orthop Surg 2022 Aug;14(8):1768-1777. [doi: 10.1111/os.13293] [Medline: 35819099]
23. Prieto-Moreno R, Mora-Traverso M, Estévez-López F, et al. Effects of the ActiveHip+ mHealth intervention on the recovery of older adults with hip fracture and their family caregivers: a multicentre open-label randomised controlled trial. EClinicalMedicine 2024 Jul;73:102677. [doi: 10.1016/j.eclinm.2024.102677] [Medline: 38911836]
24. Ortiz-Piña M, Molina-Garcia P, Femia P, et al. Effects of tele-rehabilitation compared with home-based in-person rehabilitation for older adult’s function after hip fracture. Int J Environ Res Public Health 2021 May 20;18(10):5493. [doi: 10.3390/ijerph18105493] [Medline: 34065523]
25. Li CT, Hung GK, Fong KN, Gonzalez PC, Wah SH, Tsang HW. Effects of home-based occupational therapy telerehabilitation via smartphone for outpatients after hip fracture surgery: a feasibility randomised controlled study. J Telemed Telecare 2022 May;28(4):239-247. [doi: 10.1177/1357633X20932434] [Medline: 32594832]
26. Kalron A, Tawil H, Peleg-Shani S, Vatine JJ. Effect of telerehabilitation on mobility in people after hip surgery: a pilot feasibility study. Int J Rehabil Res 2018 Sep;41(3):244-250. [doi: 10.1097/MRR.0000000000000296] [Medline: 29794545]
27. Latham NK, Harris BA, Bean JF, et al. Effect of a home-based exercise program on functional recovery following rehabilitation after hip fracture. JAMA 2014 Feb 19;311(7):700-708. [doi: 10.1001/jama.2014.469] [Medline: 24549550]
28. Di Monaco M, De Toma E, Gardin L, Giordano S, Castiglioni C, Vallero F. A single postdischarge telephone call by an occupational therapist does not reduce the risk of falling in women after hip fracture: a randomized controlled trial. Eur J Phys Rehabil Med 2015 Feb;51(1):15-22. [Medline: 25184799]
29. Min K, Beom J, Kim BR, et al. Clinical practice guideline for postoperative rehabilitation in older patients with hip fractures. Ann Rehabil Med 2021 Jun;45(3):225-259. [doi: 10.5535/arm.21110] [Medline: 34233406]
30. Rosner BI, Gottlieb M, Anderson WN. Effectiveness of an automated digital remote guidance and telemonitoring platform on costs, readmissions, and complications after hip and knee arthroplasties. J Arthroplasty 2018 Apr;33(4):988-996. [doi: 10.1016/j.arth.2017.11.036] [Medline: 29229238]

XSL•FO
RenderX

31. Nelson M, Russell T, Crossley K, Bourke M, McPhail S. Cost-effectiveness of telerehabilitation versus traditional care after total hip replacement: a trial-based economic evaluation. J Telemed Telecare 2021 Jul;27(6):359-366. [doi: 10.1177/1357633X19869796] [Medline: 31530065]
32. Gazendam A, Zhu M, Chang Y, Phillips S, Bhandari M. Virtual reality rehabilitation following total knee arthroplasty: a systematic review and meta-analysis of randomized controlled trials. Knee Surg Sports Traumatol Arthrosc 2022 Aug;30(8):2548-2555. [doi: 10.1007/s00167-022-06910-x] [Medline: 35182172]
33. Perera S, Mody SH, Woodman RC, Studenski SA. Meaningful change and responsiveness in common physical performance measures in older adults. J Am Geriatr Soc 2006 May;54(5):743-749. [doi: 10.1111/j.1532-5415.2006.00701.x] [Medline: 16696738]
34. Wright AA, Cook CE, Baxter GD, Dockerty JD, Abbott JH. A comparison of 3 methodological approaches to defining major clinically important improvement of 4 performance measures in patients with hip osteoarthritis. J Orthop Sports Phys Ther 2011 May;41(5):319-327. [doi: 10.2519/jospt.2011.3515] [Medline: 21335930]
35. Yuksel E, Unver B, Kalkan S, Karatosun V. Reliability and minimal detectable change of the 2-minute walk test and timed up and go test in patients with total hip arthroplasty. Hip Int 2021 Jan;31(1):43-49. [doi: 10.1177/1120700019888614] [Medline: 31928090]
36. Singh JA, Schleck C, Harmsen S, Lewallen D. Clinically important improvement thresholds for Harris Hip Score and its ability to predict revision risk after primary total hip arthroplasty. BMC Musculoskelet Disord 2016 Jun 10;17:256. [doi: 10.1186/s12891-016-1106-8] [Medline: 27286675]
37. Arcolin I, Godi M, Giardini M, Guglielmetti S, Bellotti L, Corna S. Minimal clinically important difference of the functional independence measure in older adults with hip fracture. Disabil Rehabil 2024 Feb;46(4):812-819. [doi: 10.1080/09638288.2023.2175386] [Medline: 36750763]
38. Unnanuntana A, Jarusriwanna A, Nepal S. Validity and responsiveness of Barthel index for measuring functional recovery after hemiarthroplasty for femoral neck fracture. Arch Orthop Trauma Surg 2018 Dec;138(12):1671-1677. [doi: 10.1007/s00402-018-3020-z] [Medline: 30094561]

Abbreviations

BADL: basic activity of daily living
BI: Barthel index
FIM: Functional Independence Measure
HHS: Harris Hip Score
MBI: modified Barthel index
MD: mean difference
MeSH: Medical Subject Headings
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SMD: standardized mean difference
SPPB: Short Physical Performance Battery
TUG: timed up and go

Edited by J Zhou; submitted 12.May.2025; peer-reviewed by B Cedeno-Veloz, YIP Chi Bun, PC Lin, WJ He; revised version received 09.Jan.2026; accepted 19.Jan.2026; published 09.Mar.2026.

Please cite as:
Cai CX, Qin PP, Liu CY, Cai P, Wei X
Postoperative Telerehabilitation in Patients With Hip Fracture: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2026;14:e77341
URL: https://mhealth.jmir.org/2026/1/e77341
doi:10.2196/77341

© Can-xin Cai, Penny Ping Qin, Chuan-yao Liu, Peng Cai, Xijun Wei. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 9.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Acceptability of Telehealth as the Default Modality for Multiple Sclerosis Care in Switzerland: Cross-Sectional Study

Sintieh Nchinda Ngek Ekongefeyin[1,2,3], MSc, MD; Paola Daniore[4], PhD; Vasileios Nittas[3], PhD; Stefania Iaquinto[3], MSc; Enriqueta Vallejo-Yagüe[3,5], PhD; Christian P Kamm[6,7], Prof Dr Med; Pasquale Calabrese[8], Prof Dr Med; Claudia Baum[9], Dr med; Claudio Gobbi[10,11], Prof Dr Med; Chiara Zecca[10,11], Prof Dr Med; Andrew Chan[7], Prof Dr Med; Milo Puhan[3], Prof Dr; Viktor von Wyl[1,2,3], Prof Dr

[1]Institute for Implementation Science in Health Care, University of Zurich, Universitatstrasse 84, Zurich, Switzerland

[10]Department of Neurology, Neurocenter of Southern Switzerland, Ospedale Regionale di Lugano, EOC, Lugano, Switzerland

[11]Faculty of Biomedical Sciences, Università della Svizzera Italiana, Lugano, Switzerland

[2]Digital Society Initiative, University of Zurich, Zurich, Switzerland

[3]Epidemiology, Biostatistics, and Prevention Institute, University of Zurich, Zurich, Switzerland

[4]Center for Digital Trust, Swiss Federal Institute of Technology Lausanne, Lausanne, Switzerland

[5]Institute of Primary Health Care (BIHAM), Faculty of Medicine, University of Bern, Bern, Switzerland

[6]Neurocenter, Luzerner Kantonsspital, Lucerne, Switzerland

[7]Department of Neurology, Inselspital, University Hospital Bern, University of Bern, Bern, Switzerland

[8]Neuropsychology and Behavioral Neurology Unit, Division of Cognitive and Molecular Neuroscience, University of Basel, Basel, Switzerland

[9]Rehaklinik Zihlschlacht AG, Neurological Rehabilitation Center, Part of the VAMED Group, Zihlschlacht, Switzerland

Corresponding Author:
Sintieh Nchinda Ngek Ekongefeyin, MSc, MD
Institute for Implementation Science in Health Care, University of Zurich, Universitatstrasse 84, Zurich, Switzerland

Abstract

Background: Telehealth can improve access to care for people living with multiple sclerosis (MS), but information on its acceptance is limited in Switzerland.

Objective: This study aimed to determine the proportion of people living with MS willing to accept telehealth as a new default and the factors associated with their acceptance.

Methods: We conducted a cross-sectional analysis using survey data from the Swiss Multiple Sclerosis Registry. We defined "telehealth as a default" as a health care model where remote consultations (telephone and/or video calls) are the primary mode of interaction between patients and their physicians, with in-person visits based on clinical necessity. Multivariable logistic regression was performed to evaluate the association between telehealth acceptance and sociodemographic and health-related factors. Telehealth acceptance was described in relation to 3 survey variables that mirrored key constructs from the Non-Adoption, Abandonment, Scale-Up, Spread, and Sustainability (NASSS) framework. The variables were digital communication preferences, internet use for health provider searches, and experience with telemedicine.

Results: Among 427 respondents, 15.5% (66/427) reported a willingness to accept telehealth as their default. In this group, only 21.2% (14/66) had experience using telemedicine. A descriptive analysis of our 3 NASSS-derived key constructs showed that among the 78.5% (335/427) respondents who generally agreed to digital access to health data, only 17.0% (57/335) accepted telehealth as a default. Notably, 30.7% (129/427) of participants stated a wish for support for using devices or the internet. Among those 129 individuals, 17.1% (22/129) were willing to accept telehealth as a default. Of the 89 people with prior telehealth experience, 15.7% (14/89) were willing to accept telehealth. In multivariable analysis, digital communication with health care providers (adjusted odds ratio [aOR] 14.56, 95% CI 6.18 - 39.04; *P*<.001*)*, current internet use for health care provider search (aOR 7.78, 95% CI 1.34 - 45.32; *P*=.021), and a secondary progressive MS diagnosis (aOR 0.22, 95% CI 0.05 - 0.72; *P*=.021) were independently associated with accepting telehealth as a default.

Conclusions: Our findings suggest a low acceptance of telehealth as a default among people living with MS in Switzerland. While our 3 postulated NASSS-derived key constructs were not associated with telehealth acceptance, we noted additional behavioral factors, including previous digital communication with health care providers and using the internet to search for health care provider information, which were associated with telehealth acceptance. Moreover, advanced disease states like secondary progressive MS were negatively associated with telehealth acceptance. Thus, telehealth as a default will be most acceptable in

XSL•FO
RenderX

people living with MS who already use the internet for their health, and those with less severe disease. Future research should explore provider perspectives and evaluate long-term strategies for the acceptance of telehealth in MS care.

(JMIR Mhealth Uhealth 2026;14:e84447) doi:10.2196/84447

KEYWORDS

acceptance; chronic disease management; digital health; multiple sclerosis; patient preferences; telehealth as a default; telemedicine

Introduction

Telehealth, the use of information and communication technology to provide remote health care, is a relatively novel approach in chronic disease management [1]. If widely adopted, it has the potential to improve access, continuity, and quality of care while reducing costs for populations with chronic diseases [2,3]. Evidence of patient benefits through its application in chronic disease care is promising. When used among people with cardiovascular disease (CVD), half of them had improved CVD risk scores [4]. When used among people living with diabetes, the quality of care was comparable to standard care while improving access and reducing the cost of care [5]. Among people living with cancer, it improved the quality of life and psychological well-being [6]. The COVID-19 pandemic also accelerated the use of telehealth in patient care, demonstrating its feasibility and acceptability for both patients and clinicians in chronic disease care, including multiple sclerosis (MS) [7,8].

In MS care, telehealth has been used for follow-up consultations, cognitive assessments, rehabilitation, and the management of symptoms that may limit mobility or access to specialist care. Altman et al [9] reported that both people living with MS and health care providers had a high level of satisfaction with telehealth. A review reported that people living with MS and their health care providers generally viewed telehealth as an acceptable and effective alternative to traditional in-person visits [10]. Also, telehealth has been reported to improve self-management [11,12], anxiety and depression [13-15], and activities of daily living in people living with MS [16]. Finally, no adverse events were reported with telerehabilitation for people living with MS [17].

The successful implementation of telehealth in MS and other chronic diseases is influenced by a range of enablers and barriers. Enablers include the perceived value proposition of the technology, such as improved access, convenience, and patient outcomes, alongside supportive organizational structures and positive attitudes among both patients and clinicians [18-20]. Conversely, barriers often arise from the severity of the disease, organizational complexity, technological challenges, digital literacy gaps, and concerns regarding data security or the adequacy of remote clinical assessments [16,21-23].

To systematically assess enabling and inhibiting factors, several frameworks have been proposed [24-26]. A widely used framework is the Non-Adoption, Abandonment, Scale-Up, Spread, and Sustainability (NASSS) framework [27]. The NASSS framework provides a comprehensive approach to understanding the adoption and sustained use of technologies in health care and has been used for a wide range of technologies and disease areas [27], including the use of telehealth in chronic care for conditions like CVDs, cancers, and neurological disorders [26]. The framework evaluates the complexity of technology acceptance within 7 constructs included in NASSS, which aim to identify and address factors influencing technology acceptance, anticipate barriers to implementation, and design interventions that are more likely to achieve scale-up and sustainability.

In MS care, few studies examine under what conditions people living with MS would accept telehealth as a default. In our study, we define “telehealth as a default” as a health care delivery model where remote consultations are the primary mode of interaction between patients and their physicians, with in-person visits occurring based on clinical necessity [28]. Previous studies report high patient and provider satisfaction [29], variable confidence in using telehealth in MS care [19], barriers limiting telehealth adoption [23]; but fail to report what conditions would enable or inhibit telehealth as a default among people living with MS.

Overall, many preconditions for successful telehealth implementation in Switzerland should be in place, including the presumed (and partially demonstrated) advantages of telehealth, the generally positive health-seeking attitude of people living with MS in Switzerland [30,31], and the push from global and national authorities for digital infrastructure to reduce diagnostic and therapeutic delays [1,32,33]. However, it remains unclear to what extent the Swiss population of people living with MS is capable and willing to engage in telehealth as a default mode for routine consultations. Therefore, our research seeks to assess the proportion of people living with MS in Switzerland potentially willing to accept telehealth as a default and the factors influencing their acceptance. Our findings are contextualized using the NASSS framework [34] and will provide information on how to target people living with MS with telehealth and consequently improve the quality of care for people living with MS.

Methods

Study Design and Setting

This study was a nested cross-sectional analysis within the Swiss Multiple Sclerosis Registry (SMSR), an ongoing observational cohort initiated in 2016. The SMSR recruits adults with MS residing or receiving treatment in Switzerland. Participation is voluntary and requires written informed consent, alongside confirmation of MS or clinically isolated syndrome (CIS) diagnosis. Upon enrollment, participants complete a baseline questionnaire and are invited to provide self-reported data via follow-up questionnaires approximately twice a year, available both online and in paper format.

XSL•FO
RenderX

Ethical Considerations

Ethical approval for the SMSR was granted by the ethics committee of the Canton of Zurich (PB–2016–00894; BASEC-NR 2019–01027). Further details on the SMSR have been previously published [35,36]. All participants had previously provided informed consent to be contacted and the use of their data for research purposes, including secondary analyses. Participation in this survey was voluntary, and individuals could decline without consequences. All data provided to the study team were fully anonymized. To protect participants' privacy and confidentiality, only anonymized data were used for this analysis, and results are reported in aggregate such that no individual participant can be identified. Participants received no compensation for completing the survey.

For this analysis, we used data from an anonymous, complementary online survey conducted between April 2023 and June 2023. This survey is a follow-up investigation of an earlier study on patterns of digital device use and digital literacy that was launched in October 2020 [37]. The 2023 survey used for this analysis was conducted online, in German, and offered independently of routine SMSR follow-ups (n=1720 eligible; 431 respondents; Multimedia Appendix 1). The rationale for offering the follow-up survey to a limited subsample had several reasons. First, we aimed to inform ongoing debates on telehealth and electronic health records promptly and therefore decided to offer the survey only in one language and online, also to facilitate data analysis. Second, our decision to offer the survey online and anonymously was also designed to attract inactive participants (ie, those who had not completed surveys in the past 2 years but remained in the cohort). Including these participants was intended to mitigate nonresponse bias and to capture perspectives from people living with MS who may be less engaged with research or digital tools. Third, the well-described SMSR source population and the existence of results from the 2020 study, which was offered to the full SMSR study population in all languages and participation modes (online and paper-pencil), enabled us to examine the generalizability and projections of key findings based on the limited, second survey (Multimedia Appendix 1).

Using the NASSS Framework for Hypothesis Development and Interpretation of Findings

The NASSS framework provides a structured approach to understanding the facilitators and barriers to technology uptake [34]. In our study, we used the NASSS framework to generate hypotheses, given the available data, for which factors may determine the acceptance of telehealth as a default by people living with MS in Switzerland. It comprises 7 domains: *the condition*—the nature of the health condition; *the technology*—the technology itself (eg, usability and reliability); *the value proposition*—the value of the technology to health care stakeholders; *the adopters*—the individuals expected to adopt the technology (patients, clinicians, and carers); *the organization*—the health care organizational context; *the wider system*—the wider system including policy and regulatory factors; and *embedding over time*—how the technology and its use may evolve through processes of embedding and adaptation. To describe the complexity of technology acceptance in the aforementioned domains, the NASSS framework defines a "simple technology" as one that has few components and its successful implementation is predictable, a "complicated technology" as one that has multiple components and its success is largely predictable, and a "complex technology" as one that has numerous interacting components with limited chances of succeeding and hence, less likely to be adopted [27]. Systematic reviews applying the NASSS framework have identified that barriers are most frequently encountered within the domains of organization and adopter system, while enablers are linked to clear value propositions and alignment with patient needs [26,34,38].

Main Outcomes

The primary study outcome was the proportion of participants who accepted telehealth as a default method for consultations. The acceptance of telehealth as a default was assessed using the following question: "Agreement to Digital Default (Would you find it beneficial for your medical care if medical consultations were through telemedicine?)," collected as a Likert scale from 0 (completely disagree) to 4 (completely agree). For further analysis, a binary variable was created where "Yes" included Likert levels (3,4) and "No" included levels (0,1,2). When describing the prevalence of acceptance in the population, we refer to the observed proportion as "low acceptance" descriptively; we did not define formal quantitative thresholds for low versus high acceptance.

Main Explanatory Variables of Interest

We hypothesized 3 main drivers of telehealth acceptance that are included in our data: (1) knowledge of and prior experience with telehealth, (2) general openness toward digital health tools and electronic health records, and (3) needs and wishes for digital support.

Knowledge of and experience with telehealth was assessed using the following question: "Do you have knowledge of telemedicine offerings, and have you already gathered experiences with telemedicine?" Answer options were "neither knowledge of nor experience with telemedicine," "knowledge of but no experience with telemedicine," or "knowledge of and experience with telemedicine."

Openness for use of other digital health tools was assessed through items on willingness to accept other digital health services (eg, file sharing, health data use, and provider communication) and experience with telemedicine. Each variable was collected as a Likert scale ranging from 0 (completely disagree with digital default) to 4 (completely agree with specific digital default).

Digital support needs were assessed with two items: (1) "Do you require support in using digital devices (eg, smartphones, laptops)?" and (2) "Do you need assistance with internet-related tasks (eg, security, navigation, program management)?" Each was rated on a Likert scale from 0 (no support needed) to 2 (extensive support needed). Participants reporting moderate (1) or extensive support (2) needs for both items were classified as requiring digital support.

XSL•FO
RenderX

Other Explanatory Variables

Covariates included age (categorized as 18‐30, 31‐40, 41‐50, 51‐60, 61‐70, >70 y), sex (male or female), MS type (CIS or relapsing-remitting MS, primary progressive MS [PPMS], secondary progressive MS [SPMS]), and frequency of internet use for health-related purposes (eg, appointment scheduling and provider communication). Prior or current telemedicine use was included as a binary variable. Additional covariates included concerns related to internet use (eg, data security and information reliability). Internet use frequency was categorized as never, less than monthly, monthly, weekly, daily, or several times daily. A composite "use score" (range 0‐5) was calculated by summing the number of online activities performed at least monthly.

Statistical Analysis

We first described the study population using MS type, sociodemographic characteristics, and experience with telemedicine. We then compared key variables based on previous telehealth experience (Multimedia Appendix 2) and openness to default telehealth consultations (see Table 1). In this study, a complete case analysis was undertaken. We excluded persons where key variables were missing from the analysis, leading to the exclusion of 4 participants.

XSL•FO
RenderX

Table . Comparison of preferences by default telehealth consultation status (N=427).

Variables	Overall, n (%)	Telehealth consultation as default =no (n=361), n (%)	Telehealth consultation as default=yes (n=66), n (%)
Descriptive variables			
Age group (y)			
18 - 30	11 (2.6)	9 (2.5)	2 (3.0)
31 - 40	56 (13.1)	48 (13.3)	8 (12.1)
41 - 50	108 (25.3)	90 (24.9)	18 (27.3)
51 - 60	143 (33.5)	121 (33.5)	22 (33.3)
61 - 70	81 (19.0)	73 (20.2)	8 (12.1)
>70	28 (6.6)	20 (5.5)	8 (12.1)
Sex			
Female	298 (69.8)	256 (70.9)	42 (63.6)
Male	129 (30.2)	105 (29.1)	24 (36.4)
MS[a] types			
CIS[b] and RRMS[c]	285 (66.7)	240 (66.5)	45 (68.2)
PPMS[d]	51 (11.9)	44 (12.2)	7 (10.6)
SPMS[e]	91 (21.3)	77 (21.3)	14 (21.2)
Hypothesis 1: Have you ever used a telemedicine service?			
No knowledge, no experience	169 (39.6)	143 (39.6)	26 (39.4)
Knowledge, no experience	169 (39.6)	143 (39.6)	26(39.4)
Knowledge, experience	89 (20.9)	75 (20.8)	14 (21.2)
Hypothesis 2: acceptability to digital default (Would you find any of the following functionalities beneficial for your medical care?)			
Sell your health data (agree)	102 (24.2)	72 (20.2)	30 (45.5)
Default digital access to health data (agree)	335 (78.5)	277 (76.7)	58 (87.9)
Default digital document exchange (agree)	236 (55.7)	187 (52.2)	49 (74.2)
Default digital communication with a health care provider (agree)	167 (39.1)	114 (31.6)	53 (80.3)
Hypothesis 3: In which areas do you wish for more support in the “digital” world?			
Need support when using the internet (yes)	104 (24.8)	85 (24.0)	19 (28.8)
Need support when using hardware (yes)	109 (25.8)	86 (24.2)	23 (34.8)
Needs support for both	129 (30.2)	105 (29.7)	24 (36.4)
Other potentially explanatory variables			
Which concerns do you have related to internet use?			
Data security (agree)	271 (64.2)	235 (66.0)	36 (54.5)
Reliability of information “invalid data” (agree)	252 (60.0)	220 (62.1)	32 (48.5)
Isolation from real life (agree)	97 (23.1)	80 (22.6)	17 (25.8)

XSL•FO
RenderX

Variables	Overall, n (%)	Telehealth consultation as default =no (n=361), n (%)	Telehealth consultation as default=yes (n=66), n (%)
Health effects of internet use (agree)	70 (16.6)	54 (15.2)	16 (24.2)
For which activity do you go online when related to the topic of medicine and health (components for use score)? Percentage with at least 1 monthly activity			
Exchange with other patients	41 (10.5)	33 (9.9)	8 (13.6)
Self-tracking	43 (11.2)	32 (9.7)	11 (20.0)
Communication with a health care provider	57 (14.0)	44 (12.8)	13 (21.0)
Health care provider search	30 (7.5)	21 (6.2)	9 (15.0)
MS information search	195 (48.6)	157 (46.3)	38 (61.3)
Health information search	249 (62.7)	209 (61.5)	40 (70.2)
Private appointments (yes)	250 (58.5)	208 (57.6)	42 (63.6)
Professional appointments (yes)	188 (44.0)	157 (43.5)	31 (47.0)
All appointment types (yes)	299 (70.0)	250 (69.3)	49 (74.2)
How often do you use electronic devices (at least monthly)			
Frequency of smartwatch use	57 (18.6)	48 (18.0)	9 (22.0)
Frequency of smartphone use	365 (95.3)	307 (94.8)	58 (98.3)
Frequency of tablet use	146 (45.8)	126 (45.7)	20 (46.5)
Frequency of PC use	294 (89.1)	246 (88.8)	48 (90.6)

[a]MS: multiple sclerosis.

[b]CIS: clinically isolated syndrome.

[c]RRMS: relapsing-remitting multiple sclerosis.

[d]PPMS: primary progressive multiple sclerosis.

[e]SPMS: secondary progressive multiple sclerosis.

Univariable logistic regression models were used to identify factors associated with willingness to accept default telehealth consultations (Multimedia Appendix 3). All prespecified variables relating to the use of digital tools and attitudes toward digitalization were included in the full multivariable logistic regression model, which also included predefined adjustments for potential confounders (age, sex, MS type, and telemedicine experience). A 2-sided *P* value <.05 was considered statistically significant. All analyses were conducted using R (version 4.4.2; R Foundation for Statistical Computing).

Reporting Framework

This paper was written following the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines (Checklist 1) for cross-sectional studies [39].

Results

Sociodemographic Data

Of 431 survey respondents, 4 had missing information regarding age. Therefore, a total of 427 participants with complete data were included in the analysis. The study sample consisted of 30.2% (129/427) men and 69.8% (298/427) women, with 41% (175/427) of participants aged between 18 and 50 years. Of the respondents, 66.7% (285/427) reported having relapsing-remitting MS or CIS, while 11.9% (51/427) had PPMS and 21.3% (91/427) SPMS. A total of 20.9% (89/427) reported previous experience with telemedicine consultations (Table 2).

XSL•FO
RenderX

Table . Characteristics of participants (N=427).

Variable	All participants, n (%)
Age group (y)	
18 - 30	11 (2.6)
31 - 40	56 (13.1)
41 - 50	108 (25.3)
51 - 60	143 (33.5)
61 - 70	81 (19)
70+	28 (6.6)
Sex	
Female	298 (69.8)
Male	129 (30.2)
MS[a] types	
CIS[b] and RRMS[c]	285 (66.7)
PPMS[d]	51 (11.9)
SPMS[e]	91 (21.3)
Experience or knowledge with telemedicine	
No experience or knowledge	169 (39.6)
Telemedicine knowledge, no experience	169 (39.6)
Telemedicine knowledge and experience	89 (20.9)

[a]MS: multiple sclerosis.

[b]CIS: clinically isolated syndrome.

[c]RRMS: relapsing-remitting multiple sclerosis.

[d]PPMS: primary progressive multiple sclerosis.

[e]SPMS: secondary progressive multiple sclerosis.

Acceptance of Telehealth as a Default Method of Consultation

A total of 66 out of 427 (15.5%) participants were open to a default telehealth consultation (Table 1). Among this group, 21.2% (14/66) had previous telemedicine experience. Although 78.5% (335/427) agreed to default digital access to health data, only 17.3 (58/335) agreed to default to telehealth. Of the 30.2% (129/427) participants needing hardware and software support, 81.4% (105/129) did not agree to telehealth as a default. Older age groups greater than or equal to 60 years 90.1% (73/81,) and those concerned with data security 86.7% (235/271,) and information validity 87.3%(220/252,) rarely accepted telehealth as a default.

Factors Associated With Acceptance of Telehealth as a Default

Multivariable Regression

In the multivariable regression model (Table 3), default communication with health care providers (adjusted odds ratio [aOR] 14.56, 95% CI 6.18 - 39.04; *P*<.001), internet use for health care provider search (aOR 7.78, 95% CI 1.34 - 45.32; *P*=.021) and a SPMS diagnosis (aOR 0.22, 95% CI 0.05 - 0.72; *P*=.021) were independently associated with an agreement to default telehealth consultation. Other factors, such as concerns about invalid information, data security, and telemedicine experience, did not show significant associations when adjusted for other predictors. The multivariable model showed low multicollinearity (variance inflation factor <1.5), suggesting that the predictors are not highly correlated with each other and that the regression coefficients can be reliably interpreted.

XSL•FO
RenderX

Table . Multivariable logistic regression for telehealth consultations.

Variable	aOR[a] (95% CI)	*P* value
Descriptive variables		
Sex (male)	0.88 (0.35-2.11)	.778
Age (y)		.271[b]
18 - 30 (reference)	—[c]	
31 - 40	0.49 (0.07-4.60)	
41 - 50	0.60 (0.10-5.12)	
51 - 60	0.67 (0.10-5.89)	
61 - 70	0.55 (0.07-5.41)	
>70	4.15 (0.40-54.92)	
MS[e] type		.041[d]
MS type: CIS[f] or RRMS[g] (reference)	—	
MS type: PPMS[h]	0.67 (0.17-2.24)	
MS type: SPMS[i]	0.22 (0.05-0.72)	
Hypothesis 1: Have you ever used a telemedicine service?		
Telemedicine experience (yes or no; reference)	0.94 (0.37-2.23)	.888
Hypothesis 2: agreement to digital default (Would you find any of the following functionalities beneficial for your medical care?)		
Default digital communication with HCP[j]	14.56 (6.18-39.04)	<.001
Default digital access to health data	0.65 (0.19-2.34)	.492
Default digital document exchange	1.60 (0.66-4.15)	.311
Hypothesis 3: In which areas do you wish for more support in the “digital” world?		
Needs support (yes or no)	1.98 (0.84-4.60)	.112
Other explanatory variables		
Concern about invalid information	0.96 (0.42-2.20)	.920
Concerned about data security	0.85 (0.37-1.97)	.702
Uses the internet for health info search	0.86 (0.37-2.05)	.737
Uses internet for MS info search	1.68 (0.73-3.93)	.225
Uses the internet for HCP search	7.78 (1.34-45.32)	.021

XSL•FO
RenderX

Variable	aOR[a] (95% CI)	*P* value
Uses the internet for communication with an HCP	0.62 (0.19-1.82)	.409
Uses the internet for self-tracking	1.38 (0.45-3.95)	.560
Uses the internet for patient exchange	1.02 (0.28-3.21)	.980

[a]aOR: adjusted odds ratio.

[b]Likelihood ratio test: χ^2_5=5.49; *P*=.36.

[c]Not applicable.

[d]Likelihood ratio test: χ^2_2=6.44; *P*=.040.

[e]MS: multiple sclerosis.

[f]CIS: clinically isolated syndrome.

[g]RRMS: relapsing-remitting multiple sclerosis.

[h]PPMS: primary progressive multiple sclerosis.

[i]SPMS: secondary progressive multiple sclerosis.

[j]HCP: health care professional.

Discussion

Summary of Findings

This cross-sectional study determined the proportion of people living with MS who are willing to accept telehealth as a default and the factors influencing their acceptance. We found that only 66 out of 427 (15.5%) participants were willing to accept telehealth as a default, which was positively associated with a general openness for digital communication with health care providers or using the internet for health care provider searches. We also found a negative association of telehealth acceptance with having SPMS.

The Proportion of People Living With MS Who Accepted Telehealth as a Default

In our study, less than one-fifth of people living with MS accepted telehealth as a default. However, this finding contrasts with other studies reporting over 70% acceptance of telehealth among people living with MS compared to in-person visits [23,29,40]. These earlier studies examined acceptance and satisfaction with telehealth during the COVID-19 pandemic, a period in which the extensive use of telemedicine ensured continuity of care, which probably contributed to increased acceptance. Hence, the difference in rates of acceptance might be explained by the forced global shift to digital services to curb the pandemic. Also, these earlier investigations did not specifically assess accepting telehealth as a default. Our low acceptance rate could also be due to the complexity of MS and factors like comorbidities and the older age of our study sample [41]. MS severity is often linked with more physical impairments, comorbidities, and aging [42]. In our study, persons with PPMS and SPMS, as well as persons aged over 60 years, were overrepresented compared with the projected population of people living with MS in Switzerland (Multimedia Appendix 1). These sample differences may explain the low percentage of telehealth acceptance as a default. Indeed, studies have repeatedly shown that people aged above 60 years are least accepting telehealth as a default; that corroborates with research on technology acceptance in older populations [23,27,43-45]. Nevertheless, some research highlights that older adults may still be receptive to telehealth when it addresses access barriers, reduces the burden of travel, or is perceived as enhancing continuity of care [46,47]; hence, addressing these preconditions may increase acceptance of telehealth.

The contrast between the high acceptance of default digital access to health data (78.5%) and the low acceptance of telehealth as default MS care (17.1%) suggests that people living with MS may favor asynchronous over synchronous digital modalities. Asynchronous tools such as secure messaging, patient portals, and digital care paths enable patients to access information, contact their care team, and receive tailored feedback without the need for real-time interaction and have been shown to support chronic disease management and self-management, including in MS [48-50]. Our findings most likely suggest the use of asynchronous telehealth as a complementary, but not necessarily default, component of chronic MS care.

Factors Associated With Accepting Telehealth as a Default

In our multivariable regression analysis, neither of the 3 prepostulated NASSS constructs as possible drivers of telehealth acceptance (hypotheses 1 - 3) showed independent associations. For example, although prior telemedicine experience has been linked to greater acceptance in other settings, it was not significantly associated with acceptance in our cohort. Only 20% of participants reported any prior telemedicine experience—consistent with the generally low acceptance of telehealth among people living with MS in Switzerland [32]—which likely limited its influence in our analysis.

However, we found that a general acceptance toward digital communication with health care providers, as well as regular internet use to search for health care providers, was positively associated with a willingness to accept telehealth as a default.

XSL•FO
RenderX

A general acceptance of digital communication may indeed ease a transition toward telehealth, accelerated by the COVID-19 pandemic, when tools for digital communication through voice and video became more widely adopted, also among older populations [8,10,29,51]. The positive association of prior experience with online searches of health care providers is more challenging to interpret but may signal a general interest in using the internet as a tool for personal health management [7,11]. In the first digital survey, frequent online searches for health care providers distinguished a subgroup of people living with MS with greater health awareness who regularly used digital tools for managing their condition [37]. In general, individuals familiar with digital tools may perceive a lower cognitive and practical barrier in using telehealth [52]. This observation could also be true for those who use the web to search for care providers. Most likely explaining the positive association between telehealth and prior use of the internet for health care provider search found in our study, because the user's perceived effort with technology will be relatively low [45].

Individuals with SPMS were less likely to endorse telehealth as a default. This may reflect the greater physical and cognitive burden associated with SPMS—which can limit mobility, manual dexterity, and cognitive processing—and thereby reduce patients' ability or willingness to use the internet-based devices for medical consultations [23,41].

In Switzerland, telehealth expanded during the COVID-19 pandemic [8]. Remote consultations (mainly telephone and video) are reimbursed via existing tariff positions, and some basic insurance products use telemedicine "gatekeeper" models that require initial contact with a telemedicine provider. Despite this structural role, data on outpatient MS telehealth use remain limited, and our findings suggest that people living with MS do not yet view telehealth as an acceptable default for specialist care.

Contextualizing Our Findings Using the NASSS Framework

We further applied the NASSS framework to our findings to contextualize the results and to understand the complexity of implementing telehealth as a default in MS care [27]. Our review of the results identified four NASSS domains that are reflected by our study as follows:

1. Condition (MS complexity): MS is a heterogeneous disease with varying degrees of physical and cognitive impairment. Our finding of a negative association between SPMS and telehealth acceptance suggests that those with more severe or chronic forms of MS are less likely to accept telehealth by default, which is consistent with the literature [23,26,41].
2. Technology (telehealth use): the acceptance of telehealth as a default depends on prior digital literacy and technology exposure. In our cohort, only 20% had prior telemedicine experience, and over half expressed concerns about data security. These are moderating factors in digital health acceptance [44]. While global health bodies and the Swiss federal government promote digital health, acceptance may likely depend on how people living with MS engage with technology [1,32].
3. Value proposition (perceived benefits of telehealth)*:* over 70% of participants supported digital access to health data, but 80% declined telehealth as a default. Therefore, while most survey respondents are generally open toward provider-facing digital health tools in MS care, they seem more reluctant to actively engage with digital health technologies themselves. This finding could highlight a need to better communicate the potential benefits of telehealth to people living with MS [24,45]. While people with chronic diseases may accept certain aspects of digital health care [23], efforts are needed to better inform people living with MS on the value of telehealth to their health care.
4. Adopter system (integration into existing health care services)*:* one-third of participants who declined telehealth by default expressed the wish for support to use computer hardware or software. The presence of structured support systems is a critical factor in technology acceptance [7,44]. Addressing digital literacy gaps, easing the onboarding of persons with MS onto telehealth, and providing ongoing technical support and assistance could enhance the feasibility of telehealth as a default option for people living with MS [7,18].

While our analysis focused on 4 NASSS domains, we did not capture factors related to the organizational context, wider system, or long-term adaptation (Multimedia Appendix 4). This limits our understanding of how systemic factors—such as provider readiness, policy frameworks, and reimbursement models—may influence the acceptance of telehealth by default. These omissions are important, as they likely shape how people living with MS perceive and engage with telehealth beyond individual factors. Future research should prioritize collecting data on these broader domains to enable a more comprehensive use of the NASSS framework.

Strengths and Limitations

Our study has several limitations. First, given our recruitment methods and mode of survey delivery, our study population is not intended to be representative of the Swiss population of persons with MS at large (details provided in Iaquinto et al)[53]. Overall, our study sample seems to be older and more frequently of a PPMS or SPMS type. Therefore, we may have underestimated the general acceptability of telehealth (even after bias mitigation through reweighting our estimates based on our earlier, more inclusive digital health survey). However, the deviations of our (reweighted) samples from the Swiss population of persons with MS are unlikely to materially alter our main conclusion that the general openness to accept telehealth is presently quite low. Second, because the survey was administered online and focused on eHealth, it is possible that more digitally interested people living with MS were more likely to participate, which could in principle bias acceptance estimates upward, although the very low observed acceptance makes substantial overestimation unlikely.

Furthermore, since our study is a secondary data analysis, omitting some NASSS-relevant items, we were unable to assess certain NASSS domains, such as organizational factors, the wider health care system, and long-term sustainability. Also,

XSL•FO
RenderX

within the value proposition and adopter system domains, we were unable to consider the perspectives of health care providers and institutions. However, by assessing the perspective of people living with MS, our study offers important insights into the real-world acceptance of technology by the population of interest. Lastly, this study was conducted post-COVID-19 era, when the pressure to use telehealth is relatively low, and there is an increasing need for things to return to how they were before the pandemic, a possible reason for the low uptake.

However, our study has the following strengths. To our knowledge, this is the first study to specifically assess the acceptance of telehealth as a default in MS care. We also identified subgroups requiring targeted support to improve telehealth acceptance. Taken together, our study has identified some additional aspects that might pave the way toward an individualized telehealth setting suited for people with chronic diseases.

Implications and Recommendations

The low acceptance of telehealth as a default highlights the need for measures that address the barriers affecting acceptance. These measures may include targeted communication campaigns or having technical staff available to assist people living with MS with technical questions during onboarding. Our findings suggest that individuals who already engage with digital health tools for health management-related tasks, such as provider searches, are more receptive to telehealth. This suggests that furthering positive experiences with digital communication and provider interactions may also enhance the acceptance of telehealth. Furthermore, given that approximately one-third of participants required support with digital devices, implementing technical support and training programs could improve uptake. Policymakers and health care providers should prioritize these measures to ensure equitable digital health care access for people living with MS.

Clinically, our data argue against telehealth as a universal default model for MS care. Since participants with more severe disease were less likely to accept telehealth as the default modality. A more nuanced approach could involve triaging people living with MS according to disease severity, stability, and complexity: telehealth could be preferentially offered to those with stable, less severe disease for routine follow-up and counseling, whereas individuals with higher disability, new or complex symptoms, or substantial care coordination needs would continue to receive in-person visits by default. Such a stratified approach may optimize resource use while respecting individual preferences and clinical safety.

Conclusions

The acceptance and scaling of telehealth as a default is complex. Our study highlights the need for targeted implementation and testing of telehealth among a diverse cohort of people living with MS. These findings hint at potential considerations for the acceptance of telehealth as a default among people living with chronic diseases. Acceptance may be potentially improved if support is provided, and the value of these technologies is discussed with target patient groups. Such information should be accompanied by trust-building measures in data security and privacy, as well as offers for improving digital literacy for those in need. We conclude that acceptance of telehealth as a default needs to be targeted to the group of people living with chronic diseases most likely to benefit from it, and future research should explore provider perspectives and evaluate long-term strategies for integrating telehealth in MS care and chronic disease care.

Acknowledgments

The authors are grateful to Irene Rapold for her critical review of this manuscript. This manuscript was written without using generative artificial intelligence. The manuscript was drafted and reviewed in several rounds by all authors.

Funding

SNNE is funded by the Digital Society Initiative at the University of Zurich, Switzerland. The Swiss Multiple Sclerosis Registry is funded by the Swiss Multiple Sclerosis Society. These organizations were not involved in any study decisions.

Data Availability

The underlying data for this analysis are human research participant data, and in combination, are potentially identifiable. The data that supports the findings of this study are available upon reasonable request. Requests should be directed to the Swiss Multiple Sclerosis Registry at the University of Zurich; Epidemiology, Biostatistics & Prevention Institute; Hirschengraben 84, CH-8001 Zurich (ms-register@ebpi.uzh.ch).

Authors' Contributions

Conceptualization: SNNE, VVW
Data curation: SNNE, CPK, PC, CB, CG, CZ, AC, MP, VVW
Formal analysis: SNNE, VN, VVW
Funding acquisition: VVW
Methodology: SNNE, VVW
Supervision: VVW
Validation: SNNE, PD, VN, SI, EV-Y, CPK, PC, CB, CG, CZ, AC, MP, VVW
Writing – original draft: SNNE, PD, VN, VVW

XSL•FO
RenderX

Writing – review & editing: SNNE, PD, VN, SI, EV-Y, CPK, PC, CB, CG, CZ, AC, MP, VVW

Conflicts of Interest

AC has served on advisory boards for, and received funding for travel or speaker honoraria from Advisis, Alexion, Amgen, Biogen, BMS (Celgene), Dresden International Uni, Horizon Therapeutics, Janssen (J&J), Merck, Novartis, Roche, Sandoz, Sanofi-Genzyme, Teva, UCB, Uni Leipzig, Wiley, all for hospital research funds; and research support from Biogen, CSL Behring, Genzyme, Roche, UCB, and European Union and Swiss National Research Foundation. AC is the associate editor of the European Journal of Neurology and serves on the editorial board for Clinical and Translational Neuroscience and as topic editor for the Journal of International Medical Research.
CG: Ente Ospedaliero Cantonale (employer) received compensation for CG's speaking activities, consulting fees, or grants from Abbvie, Almirall, Biogen, Bristol Meyer Squibb, Lundbeck, Merck, Novartis, Sandoz, Sanofi, Teva Pharma, and Roche.
CZ: Ente Ospedaliero Cantonale (employer) received compensation for CZ's speaking activities, consulting fees, or grants from Abbvie, Alexion, Almirall, Biogen, Bristol Meyer Squibb, Eisai, Lilly, Lundbeck, Merck, Merz, Novartis, Organon, Pfizer, Sandoz, Sanofi, Teva Pharma, and Roche. CZ is a recipient of a grant for senior researchers provided by the Area Formazione accademica, Ricerca e Innovazione, EOC.
SNNE, PD, VN, SI, EV-Y, CK, PC, CB, MP, and VVW have no conflict of interest.

Multimedia Appendix 1
Swiss Multiple Sclerosis Registry (SMSR) participants in our digital survey.
[DOCX File, 18 KB - mhealth_v14i1e84447_app1.docx]

Multimedia Appendix 2
Use of technology and telemedicine experience.
[DOCX File, 21 KB - mhealth_v14i1e84447_app2.docx]

Multimedia Appendix 3
Univariable logistic regression analysis.
[DOCX File, 21 KB - mhealth_v14i1e84447_app3.docx]

Multimedia Appendix 4
Contextualizing our findings using the NASSS (Non-Adoption, Abandonment, Scale-Up, Spread, and Sustainability) framework.
[DOCX File, 20 KB - mhealth_v14i1e84447_app4.docx]

Checklist 1
STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist.
[PDF File, 116 KB - mhealth_v14i1e84447_app5.pdf]

References

1. Going digital for noncommunicable diseases: the case for action. : World Health Organization; 2024 URL: https://iris.who.int/bitstream/handle/10665/378478/9789240089921-eng.pdf [accessed 2025-12-23]
2. Ganguli I, Lim C, Daley N, Cutler D, Rosenthal M, Mehrotra A. Telemedicine adoption and low-value care use and spending among fee-for-service medicare beneficiaries. JAMA Intern Med 2025 Apr 1;185(4):440-449. [doi: 10.1001/jamainternmed.2024.8354] [Medline: 39992684]
3. Xiao Z, Han X. Evaluation of the effectiveness of telehealth chronic disease management system: systematic review and meta-analysis. J Med Internet Res 2023 Apr 27;25(1):e44256. [doi: 10.2196/44256] [Medline: 37103993]
4. Salisbury C, O'Cathain A, Thomas C, et al. Telehealth for patients at high risk of cardiovascular disease: pragmatic randomised controlled trial. BMJ 2016 Jun 1;353:i2647. [doi: 10.1136/bmj.i2647] [Medline: 27252245]
5. Hong YR, Xie Z, Nguyen OT, Turner K, Walker AF. Telehealth service use and quality of care among US adults with diabetes: a cross-sectional study of the 2022 health information national trends survey. BMJ Open 2024 Nov 3;14(10):e086418. [doi: 10.1136/bmjopen-2024-086418] [Medline: 39489536]
6. Shaffer KM, Turner KL, Siwik C, et al. Digital health and telehealth in cancer care: a scoping review of reviews. Lancet Digit Health 2023 May;5(5):e316-e327. [doi: 10.1016/S2589-7500(23)00049-3] [Medline: 37100545]
7. Learmonth YC, Galna B, Laslett LL, van der Mei I, Marck CH. Improving telehealth for persons with multiple sclerosis—a cross-sectional study from the Australian MS longitudinal study. Disabil Rehabil 2024 Oct;46(20):4755-4762. [doi: 10.1080/09638288.2023.2289594] [Medline: 38088346]
8. Nittas V, von Wyl V. COVID-19 and telehealth: a window of opportunity and its challenges. Swiss Med Wkly 2020 May 4;150(1920):w20284. [doi: 10.4414/smw.2020.20284] [Medline: 32400882]

XSL•FO
RenderX

9. Altmann P, Leutmezer F, Ponleitner M, et al. Remote visits for people with multiple sclerosis during the COVID-19 pandemic in Austria: the TELE MS randomized controlled trial. Digit Health 2022;8:20552076221112154. [doi: 10.1177/20552076221112154] [Medline: 35847524]
10. Yeroushalmi S, Maloni H, Costello K, Wallin MT. Telemedicine and multiple sclerosis: a comprehensive literature review. J Telemed Telecare 2020;26(7-8):400-413. [doi: 10.1177/1357633X19840097] [Medline: 31042118]
11. Ehde DM, Arewasikporn A, Alschuler KN, Hughes AJ, Turner AP. Moderators of treatment outcomes after telehealth self-management and education in adults with multiple sclerosis: a secondary analysis of a randomized controlled trial. Arch Phys Med Rehabil 2018 Jul;99(7):1265-1272. [doi: 10.1016/j.apmr.2017.12.012] [Medline: 29337024]
12. Turner AP, Wallin MT, Sloan A, et al. Clinical management of multiple sclerosis through home telehealth monitoring: results of a pilot project. Int J MS Care 2013;15(1):8-14. [doi: 10.7224/1537-2073.2012-012] [Medline: 24453757]
13. Gromisch ES, Turner AP, Neto LO, Haselkorn JK, Raskin SA. Improving prospective memory in persons with multiple sclerosis via telehealth: a randomized feasibility study. Mult Scler Relat Disord 2024 Aug;88:105718. [doi: 10.1016/j.msard.2024.105718] [Medline: 38878624]
14. Kever A, Aguerre IM, Vargas W, et al. Feasibility trial of a telehealth support group intervention to reduce anxiety in multiple sclerosis. Clin Rehabil 2022 Oct;36(10):1305-1313. [doi: 10.1177/02692155221107077] [Medline: 35673256]
15. Leavitt VM, Riley CS, De Jager PL, Bloom S. eSupport: Feasibility trial of telehealth support group participation to reduce loneliness in multiple sclerosis. Mult Scler 2020 Nov;26(13):1797-1800. [doi: 10.1177/1352458519884241] [Medline: 31668134]
16. Van Denend T, Mathiowetz V, Preissner K, et al. Adapting the multiple sclerosis functional composite for telehealth administration using videoconference delivery: methodological considerations and interrater reliability. Arch Rehabil Res Clin Transl 2024 Jun;6(2):100337. [doi: 10.1016/j.arrct.2024.100337] [Medline: 39006110]
17. Khan F, Amatya B, Kesselring J, Galea M, editor. Telerehabilitation for persons with multiple sclerosis. Cochrane Database Syst Rev 2015 Apr 9;2015(4):CD010508. [doi: 10.1002/14651858.CD010508.pub2] [Medline: 25854331]
18. Landi D, Ponzano M, Nicoletti CG, et al. Patient's point of view on the use of telemedicine in multiple sclerosis: a web-based survey. Neurol Sci 2022 Feb;43(2):1197-1205. [doi: 10.1007/s10072-021-05398-6] [Medline: 34283343]
19. Roth EG, Minden SL, Maloni HW, Miles ZJ, Wallin MT. A qualitative, multiperspective inquiry of multiple sclerosis telemedicine in the United States. Int J MS Care 2022;24(6):275-281. [doi: 10.7224/1537-2073.2021-117] [Medline: 36545645]
20. Uribe Guajardo MG, Baillie A, Louie E, et al. The evaluation of the role of technology in the pathways to comorbidity care implementation project to improve management of comorbid substance use and mental disorders. J Multimorb Comorb 2022;12:26335565221096977. [doi: 10.1177/26335565221096977] [Medline: 35586033]
21. Jafri RF, Jafri NF, Sayeed M, Musthafa M, Riyas S. Telehealth in multiple sclerosis: fostering quality of life. Mult Scler Relat Disord 2023 Dec;80:105280. [doi: 10.1016/j.msard.2023.105280]
22. Leal Neto O, Von Wyl V. Digital transformation of public health for noncommunicable diseases: narrative viewpoint of challenges and opportunities. JMIR Public Health Surveill 2024 Jan 25;10:e49575. [doi: 10.2196/49575] [Medline: 38271097]
23. Marrie RA, Kosowan L, Cutter G, Fox R, Salter A. Disparities in telehealth care in multiple sclerosis. Neur Clin Pract 2022 Jun;12(3):223-233. [doi: 10.1212/CPJ.0000000000001167]
24. Blut M, Chong AYL, Tsigna Z, et al. Meta-analysis of the unified theory of acceptance and use of technology (UTAUT): challenging its validity and charting a research agenda in the Red Ocean. J Assoc Inf Syst 2022 Jan;23(1):13-95. [doi: 10.17705/1jais.00719]
25. Salisbury C, Thomas C, O'Cathain A, et al. Telehealth in chronic disease: mixed-methods study to develop the TECH conceptual model for intervention design and evaluation. BMJ Open 2015 Feb 6;5(2):e006448. [doi: 10.1136/bmjopen-2014-006448] [Medline: 25659890]
26. Shin HD, Hamovitch E, Gatov E, et al. The NASSS (Non-Adoption, Abandonment, Scale-Up, Spread and Sustainability) framework use over time: a scoping review. PLOS Digit Health 2025 Mar;4(3):e0000418. [doi: 10.1371/journal.pdig.0000418] [Medline: 40096260]
27. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367. [doi: 10.2196/jmir.8775] [Medline: 29092808]
28. Greenhalgh T, Rosen R. Remote by default general practice: must we, should we, dare we? Br J Gen Pract 2021 Apr;71(705):149-150. [doi: 10.3399/bjgp21X715313] [Medline: 33771790]
29. Li V, Roos I, Monif M, et al. Impact of telehealth on health care in a multiple sclerosis outpatient clinic during the COVID-19 pandemic. Mult Scler Relat Disord 2022 Jul;63:103913. [doi: 10.1016/j.msard.2022.103913] [Medline: 35661564]
30. Barin L, Kaufmann M, Salmen A, et al. Patterns of care for multiple sclerosis in a setting of universal care access: a cross-sectional study. Mult Scler Relat Disord 2019 Feb;28:17-25. [doi: 10.1016/j.msard.2018.11.033] [Medline: 30530118]
31. Stanikić M, Twomey E, Puhan MA, et al. Experiences of persons with multiple sclerosis with the Covid-19 vaccination: a cross-sectional study of the Swiss Multiple Sclerosis Registry. Mult Scler Relat Disord 2023 Jun;74:104707. [doi: 10.1016/j.msard.2023.104707] [Medline: 37068368]

XSL•FO
RenderX

32. Zimmer A. Transformation numérique du système de santé: gare aux chantiers [Article in French]. Bull Med Suisses 2022;103(48):26-27. [doi: 10.4414/bms.2022.21300]
33. Xiang XM, Bernard J. Telehealth in multiple sclerosis clinical care and research. Curr Neurol Neurosci Rep 2021 Feb 28;21(4):14. [doi: 10.1007/s11910-021-01103-4] [Medline: 33646409]
34. Greenhalgh T, Abimbola S. The NASSS framework—a synthesis of multiple theories of technology implementation. Stud Health Technol Inform 2019 Jul 30;263:193-204. [doi: 10.3233/SHTI190123] [Medline: 31411163]
35. Steinemann N, Kuhle J, Calabrese P, et al. The Swiss Multiple Sclerosis Registry (SMSR): study protocol of a participatory, nationwide registry to promote epidemiological and patient-centered MS research. BMC Neurol 2018 Aug 13;18(1):111. [doi: 10.1186/s12883-018-1118-0] [Medline: 30103695]
36. Puhan MA, Steinemann N, Kamm CP, et al. A digitally facilitated citizen-science driven approach accelerates participant recruitment and increases study population diversity. Swiss Med Wkly 2018;148(1920):w14623. [doi: 10.4414/smw.2018.14623] [Medline: 29767828]
37. Nittas V, Zecca C, Kamm CP, Kuhle J, Chan A, von Wyl V. Digital health for chronic disease management: an exploratory method to investigating technology adoption potential. PLOS ONE 2023;18(4):e0284477. [doi: 10.1371/journal.pone.0284477] [Medline: 37053272]
38. Abimbola S, Patel B, Peiris D, et al. The NASSS framework for ex post theorisation of technology-supported change in healthcare: worked example of the TORPEDO programme. BMC Med 2019 Dec 30;17(1):233. [doi: 10.1186/s12916-019-1463-x] [Medline: 31888718]
39. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007 Oct 16;147(8):573-577. [doi: 10.7326/0003-4819-147-8-200710160-00010] [Medline: 17938396]
40. Chen MH, Goverover Y, Botticello A, DeLuca J, Genova HM. Healthcare disruptions and use of telehealth services among people with multiple sclerosis during the COVID-19 pandemic. Arch Phys Med Rehabil 2022 Jul;103(7):1379-1386. [doi: 10.1016/j.apmr.2021.12.028] [Medline: 35093328]
41. Özden F, Özkeskin M, Ekici E, Tümtürk İ, Ekmekci Ö, Yüceyar N. Opinions, satisfaction and expectations of individuals with multiple sclerosis about telerehabilitation services. Clin Neurol Neurosurg 2024 Feb;237:108162. [doi: 10.1016/j.clineuro.2024.108162] [Medline: 38325037]
42. Ward M, Goldman MD. Epidemiology and pathophysiology of multiple sclerosis. Continuum (Minneap Minn) 2022 Aug 1;28(4):988-1005. [doi: 10.1212/CON.0000000000001136] [Medline: 35938654]
43. Lee EC, Grigorescu V, Enogieru I, et al. Updated national survey trends in telehealth utilization and modality (2021-2022). : Office of the Assistant Secretary for Planning and Evaluation (ASPE); 2022 Mar URL: https://www.ncbi.nlm.nih.gov/books/NBK604393/ [accessed 2025-12-23]
44. Venkatesh V, Thong J, Xu X, University of Arkansas, Hong Kong University of Science and Technology, The Hong Kong Polytechnic University. Unified theory of acceptance and use of technology: a synthesis and the road ahead. JAIS 2016 May;17(5):328-376. [doi: 10.17705/1jais.00428]
45. Zhao Y, Ni Q, Zhou R. What factors influence the mobile health service adoption? A meta-analysis and the moderating role of age. Int J Inf Manage 2018 Dec;43:342-350. [doi: 10.1016/j.ijinfomgt.2017.08.006]
46. Raja M, Bjerkan J, Kymre IG, Galvin KT, Uhrenfeldt L. Telehealth and digital developments in society that persons 75 years and older in European countries have been part of: a scoping review. BMC Health Serv Res 2021 Oct 26;21(1):1157. [doi: 10.1186/s12913-021-07154-0] [Medline: 34696789]
47. Bhatia R, Gilliam E, Aliberti G, et al. Older adults' perspectives on primary care telemedicine during the COVID-19 pandemic. J Am Geriatr Soc 2022 Dec;70(12):3480-3492. [doi: 10.1111/jgs.18035] [Medline: 36169152]
48. Nguyen OT, Alishahi Tabriz A, Huo J, Hanna K, Shea CM, Turner K. Impact of asynchronous electronic communication-based visits on clinical outcomes and health care delivery: systematic review. J Med Internet Res 2021 May 5;23(5):e27531. [doi: 10.2196/27531] [Medline: 33843592]
49. Sadeghi N, Eelen P, Nagels G, et al. Innovating care in multiple sclerosis: feasibility of synchronous internet-based teleconsultation for longitudinal clinical monitoring. J Pers Med 2022 Mar 10;12(3):433. [doi: 10.3390/jpm12030433] [Medline: 35330433]
50. Vesinurm M, Maunula A, Olli P, et al. Effects of a digital care pathway for multiple sclerosis: observational study. JMIR Hum Factors 2024 Aug 7;11(1):e51872. [doi: 10.2196/51872] [Medline: 39110966]
51. D'Haeseleer M, Eelen P, Sadeghi N, D'Hooghe MB, Van Schependom J, Nagels G. Feasibility of real time internet-based teleconsultation in patients with multiple sclerosis: interventional pilot study. J Med Internet Res 2020 Aug 13;22(8):e18178. [doi: 10.2196/18178] [Medline: 32447274]
52. Cimperman M, Makovec Brenčič M, Trkman P. Analyzing older users' home telehealth services acceptance behavior-applying an extended UTAUT model. Int J Med Inform 2016 Jun;90:22-31. [doi: 10.1016/j.ijmedinf.2016.03.002] [Medline: 27103194]
53. Iaquinto S, Chan A, Manjaly ZM, et al. Rising prevalence of multiple sclerosis in Switzerland: results from the Swiss Multiple Sclerosis Registry. Neuroepidemiology 2025;59(6):623-632. [doi: 10.1159/000542632] [Medline: 39557017]

XSL•FO
RenderX

Abbreviations

aOR: adjusted odds ratio
CIS: clinically isolated syndrome
CVD: cardiovascular disease
HCP: health care professional
MS: multiple sclerosis
NASSS: Non-Adoption, Abandonment, Scale-Up, Spread, and Sustainability
PPMS: primary progressive multiple sclerosis
RRMS: relapsing-remitting multiple sclerosis
SMSR: Swiss Multiple Sclerosis Registry
SPMS: secondary progressive multiple sclerosis
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by L Buis; submitted 19.Sep.2025; peer-reviewed by ES Yoon, R Haase; revised version received 16.Dec.2025; accepted 16.Dec.2025; published 23.Jan.2026.

Please cite as:
Ekongefeyin SNN, Daniore P, Nittas V, Iaquinto S, Vallejo-Yagüe E, Kamm CP, Calabrese P, Baum C, Gobbi C, Zecca C, Chan A, Puhan M, von Wyl V
Acceptability of Telehealth as the Default Modality for Multiple Sclerosis Care in Switzerland: Cross-Sectional Study
JMIR Mhealth Uhealth 2026;14:e84447
URL: https://mhealth.jmir.org/2026/1/e84447
doi:10.2196/84447

© Sintieh Nchinda Ngek Ekongefeyin, Paola Daniore, Vasileios Nittas, Stefania Iaquinto, Enriqueta Vallejo-Yagüe, Christian P Kamm, Pasquale Calabrese, Claudia Baum, Claudio Gobbi, Chiara Zecca, Andrew Chan, Milo Puhan, Viktor von Wyl. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 23.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Effects of Telehealth Interventions for People With Parkinson Disease: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Minyue Sun[1*], MSN; Fuyou Tang[2*], MSN; Luo min[1], BS; Shiyu Wen[1], BS; Shuang Wang[3], BS; Huiping Jiang[3], BS

[1]Department of Neurology, Mianyang Central Hospital, School of Medicine, University of Electronic Science and Technology of China, Mianyang, China

[2]Mianyang Key Laboratory of Anesthesia and Neuroregulation, Department of Anesthesiology, Mianyang Central Hospital, School of Medicine, University of Electronic Science and Technology of China, Mianyang, Sichuan, China

[3]Department of Neurology, Mianyang Central Hospital, School of Medicine, University of Electronic Science and Technology of China, Mianyang, Sichuan, China

[*]these authors contributed equally

Corresponding Author:
Luo min, BS
Department of Neurology
Mianyang Central Hospital, School of Medicine
University of Electronic Science and Technology of China
Mianyang
China
Phone: 86 18324126071
Email: 410244386@qq.com

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2026/1/e108495

Abstract

Background: The global integration of telehealth into the management of Parkinson disease (PD) addresses critical gaps in health care access, especially for patients with limited mobility in underserved regions. Despite accelerated adoption during the COVID-19 pandemic, evidence regarding telehealth's multidimensional efficacy remains inconsistent. Previous meta-analyses reported conflicting outcomes for quality of life (QOL), motor symptoms, and neuropsychiatric comorbidities.

Objective: This study aimed to quantitatively synthesize the effects of telehealth interventions across six core PD domains: (1) QOL, (2) depression, (3) anxiety, (4) motor symptoms, (5) activities of daily living (ADL), and (6) cognition.

Methods: PubMed, Embase, Cochrane Library, Scopus, and Web of Science were systematically searched until June 21, 2024. In adherence to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, English-language randomized controlled trials evaluating telehealth interventions for PD were included. Study quality was assessed using the Cochrane Risk of Bias tool. A dual analytical approach using random-effects models was applied to address heterogeneity. Studies reporting a single effect size were analyzed using the Hartung-Knapp-Sidik-Jonkman correction. Studies with multiple dependent effect sizes were analyzed using a 3-level random-effects meta-analysis with t-distribution inference, accounting for sampling, within-study, and between-study variance. Effect sizes were expressed as standardized mean differences (SMD) with 95% CIs. Heterogeneity was quantified using the τ^2; prediction intervals were not calculated due to the limited number of studies. Prespecified subgroup analyses examined intervention types (digital vs traditional telehealth) and follow-up durations. Sensitivity analyses and assessments for small-study effects (multilevel Egger tests, funnel plots) were conducted.

Results: A total of 15 randomized controlled trials (765 participants) demonstrated significant telehealth benefits: QOL significantly improved on the Medical Outcomes Study 36-Item Short Form Health Survey and Brunnsviken Brief Quality of Life Scale (SMD 0.39, 95% CI 0.06-0.72; *P*=.03), with marginal improvement on the Parkinson Disease Questionnaire-8 (SMD –0.42, 95% CI –0.88 to 0.03; *P*=.07). Telephone-based interventions outperformed digital approaches (*P*=.002). Depression symptoms were significantly reduced (SMD –0.64, 95% CI –0.93 to 0.34; *P*<.001), particularly with traditional telehealth (*P*<.001).

XSL•FO
RenderX

Anxiety also decreased significantly (SMD –0.64, 95% CI –0.92 to 0.35; *P*=.003) with negligible heterogeneity (I^2=0%). Motor symptoms improved (SMD –0.46, 95% CI –0.69 to 0.24; *P*=.001), and ADL showed substantial impairment reduction (SMD –0.79, 95% CI –1.04 to –0.54; *P*=.002). Cognition was significantly enhanced (SMD 1.12, 95% CI 0.03 to 2.20; *P*=.045) though with moderate heterogeneity (I^2=52.3%) and significant publication bias (*P*<.001). Follow-up duration did not significantly moderate effects.

Conclusions: Telehealth interventions significantly enhance multiple PD domains, with traditional (telephone/tablet-based) approaches demonstrating particular advantages for QOL and depression. Digital interventions showed more limited efficacy. These findings support telehealth as a multifaceted management tool for PD, although cognition outcomes require further investigation.

Trial Registration: PROSPERO CRD42024520169; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024520169

(JMIR Mhealth Uhealth 2026;14:e70994) doi:10.2196/70994

KEYWORDS

depression; meta-analysis; motor symptoms; Parkinson disease; quality of life; systematic review; telehealth intervention

Introduction

Parkinson disease (PD), currently the second-most common neurodegenerative disorder after Alzheimer disease, affects approximately 1% of individuals aged 55 years or older, typically manifesting around the age of 60 years [1]. The incidence of PD increases with advancing age, particularly among individuals older than 60 years [1]. Currently, PD affects nearly 6 million people globally [2]. By 2030, China is expected to have approximately 4.94 million patients with PD, representing nearly half of global cases [3]. This demographic shift could significantly strain national economies and health care systems [3].

Patients with PD experience a range of motor and nonmotor symptoms. Motor symptoms include tremors, rigidity, bradykinesia, postural instability, and gait freezing [4]. Nonmotor symptoms encompass cognitive decline, pain, fatigue, psychiatric conditions such as depression and anxiety, and sleep disturbances [5,6]. These symptoms substantially reduce patients' quality of life (QOL) and increase psychological stress and physical demands on caregivers [7]. Traditionally, management and evaluation for patients with PD require visits to outpatient clinics or hospitals for advanced diagnostics and assessments [8]. However, many patients encounter significant barriers to accessing such facilities, including limited mobility, fear of falling, depressive symptoms, fatigue, and time constraints. These barriers can worsen their symptoms, delay treatment, lead to potentially life-threatening complications, and increase overall disease burden [9]. Consequently, a growing number of researchers advocate telehealth as a means to improve PD diagnosis, treatment, and rehabilitation. Telehealth aims to overcome challenges associated with in-person consultations and geographical disparities in health care resources, thereby reducing delays in treatment, decreasing morbidity and mortality, and improving QOL among patients with PD [10].

Telehealth uses digital information and communication technologies to connect patients with health care providers and deliver medical services [11]. These technologies include internet-connected desktop computers, tablets, smartphones, and wearable devices [12,13]. The COVID-19 pandemic significantly accelerated the adoption of telehealth among patients with PD, improving health care accessibility. Research supports the practicality of telehealth and underscores its perceived effectiveness by patients with PD and neurologists [10,14-16]. Neurologists can independently conduct comprehensive assessments using the Movement Disorders Society-Unified Parkinson's Disease Rating Scale Part III (MDS-UPDRS-III), enabling more accurate evaluations in patients' usual settings rather than clinical environments. This method provides a more accurate reflection of the patients' actual condition [10,17-19]. Advantages of telehealth, such as time and cost efficiency, have resulted in high patient satisfaction [14-16,20-22]. Additionally, telehealth facilitates virtual monitoring for rehabilitation, psychotherapy, and advanced PD treatments [23-25]. It promotes interdisciplinary collaboration and provides education and training opportunities for physicians and health care workers in developing regions, overcoming geographic, travel, and financial barriers [10].

Numerous studies have confirmed the practicality and effectiveness of telehealth in managing patients with PD; yet, findings regarding QOL, anxiety, depression, motor function, activities of daily living (ADL), and cognitive function have been variable [8,26,27]. A meta-analysis by Chen et al [26] in 2020 demonstrated that telehealth interventions effectively reduced motor symptoms compared with traditional care. However, these interventions showed no substantial improvements in QOL, depression, cognitive functions, or balance abilities, contrasting with more recent findings [28-31]. A systematic review by Leon-Salas et al [8] published in 2023 indicated limited and inconclusive data regarding telehealth services for patients with PD. Following the COVID-19 pandemic, an influx of new studies necessitates an updated synthesis and analysis to clarify the effects of telehealth interventions on patients with PD. Additionally, a 2024 systematic review by Federico et al [27] reported telerehabilitation outcomes comparable to in-person therapy for patients with PD. However, the review focused solely on telerehabilitation for various neurological disorders without extensively evaluating broader telehealth effects in patients with PD.

XSL•FO
RenderX

Thus, the objective of this research was to systematically compile and analyze the most recent data on telehealth intervention effects in patients with PD.

Methods

Overview

This systematic review and meta-analysis was conducted in strict accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Multimedia Appendix 1) [32]. To ensure transparency and reproducibility, the study protocol was registered with PROSPERO (registration number CRD42024520169).

Inclusion and Exclusion Criteria

The selection criteria followed the PICOS framework: (1) Population: individuals officially diagnosed with PD; (2) Interventions: telehealth or telemedicine interventions delivered via telephone, internet, or other digital communication technologies [11]; (3) Comparison/Control: studies with an experimental group compared to control groups using standard care, routine care, conventional care, or waitlist control; (4) Outcomes: assessment of intervention effects on overall health or specific behavioral and psychological symptoms associated with Parkinsonism; (5) Study type: only randomized controlled trials (RCTs).

The exclusion criteria were as follows: (1) studies published in non-English languages, (2) incomplete studies, such as research protocols or ongoing studies, (3) studies where interventions were exclusively telehealth-based without a comparison group, or where no telehealth intervention was applied, (4) studies lacking sufficient details on relevant outcome measures, and (5) studies without adequate statistical data for analysis. No publication date limitations were applied.

Search Strategy

Two researchers (MS and FT) systematically searched 5 English-language databases (PubMed, Embase, Cochrane Library, Scopus, and Web of Science) from database inception until June 21, 2024. The comprehensive search strategy combined subject headings and keywords related to three main topics: (1) PD, (2) telehealth, and (3) RCTs. There were no limits on publication status or dates. The detailed search methods are provided in Multimedia Appendix 2. Additionally, references of included studies were manually reviewed.

Study Selection and Data Extraction

The reference management tool EndNote X9 (Clarivate Analytics) was used for data management. After removing duplicates, 2 reviewers (MS and FT) independently screened titles and abstracts based on predefined inclusion and exclusion criteria. Potentially relevant articles underwent full-text assessment to determine eligibility. Any discrepancies were resolved through discussion, and when consensus was unattainable, a third reviewer (Luomin) was consulted for a final decision.

Data extraction used a specifically designed form. Extracted information included authors' names, publication dates, country, study design, sample size, demographic and clinical characteristics (average age, gender distribution, and disease attributes), type and duration of interventions and control groups, outcome measures, and key study findings.

Risk of Bias

Two researchers (MS and FT) independently assessed the methodological quality and potential bias of included studies using the Cochrane Collaboration's Risk of Bias tool [33]. Disagreements were resolved through discussion with a third reviewer (Luomin) to achieve consensus. The tool systematically evaluates several dimensions of bias, including random sequence generation, allocation concealment, blinding of participants and personnel, handling of incomplete outcome data, and selective outcome reporting. Each study was assessed for potential selection, performance, detection, attrition, and reporting biases. Risks in each domain were categorized as low (unlikely to significantly affect results), high (likely to significantly undermine confidence), or unclear.

Statistical Analysis

All meta-analyses were conducted using random-effects models, based on the conceptual assumption that true effects vary across studies due to inherent differences in populations, interventions, and settings, rather than on statistical metrics such as I^2. A dual analytical approach incorporating the Hartung-Knapp-Sidik-Jonkman (HKSJ) adjustment was used to obtain more accurate and conservative interval estimates: for studies providing a single effect size, the HKSJ method was used to calculate 95% CIs; for studies with multiple dependent effect sizes, a 3-level random-effects meta-analysis was performed in R software using the *metafor* package (version 4.5.1; R Foundation for Statistical Computing) using t-distribution–based inference. This model accounts for sampling variance (level 1), within-study variance (level 2), and between-study variance (level 3), with t-distribution inference being analogous to the HKSJ correction for multilevel models. Heterogeneity was quantified with the τ^2 statistic. Although 95% prediction intervals were initially planned to illustrate the expected range of true effects in similar future studies, they were not calculated due to the limited number of studies for each outcome (all $k<10$), as such intervals are unreliable with small samples. The I^2 statistic is reported for descriptive purposes only, acknowledging its limited pragmatic use in conveying the magnitude of true effect variation across settings. Prespecified subgroup analyses examined intervention type (digital vs traditional telehealth) and follow-up duration. Sensitivity analyses were conducted using leave-one-out elimination. Small-study effects (for which publication bias is one possible explanation) were assessed with funnel plots and multilevel Egger tests. Statistical significance was set at 2-tailed $P<.05$. Complete analysis code is provided in Multimedia Appendix 3.

Results

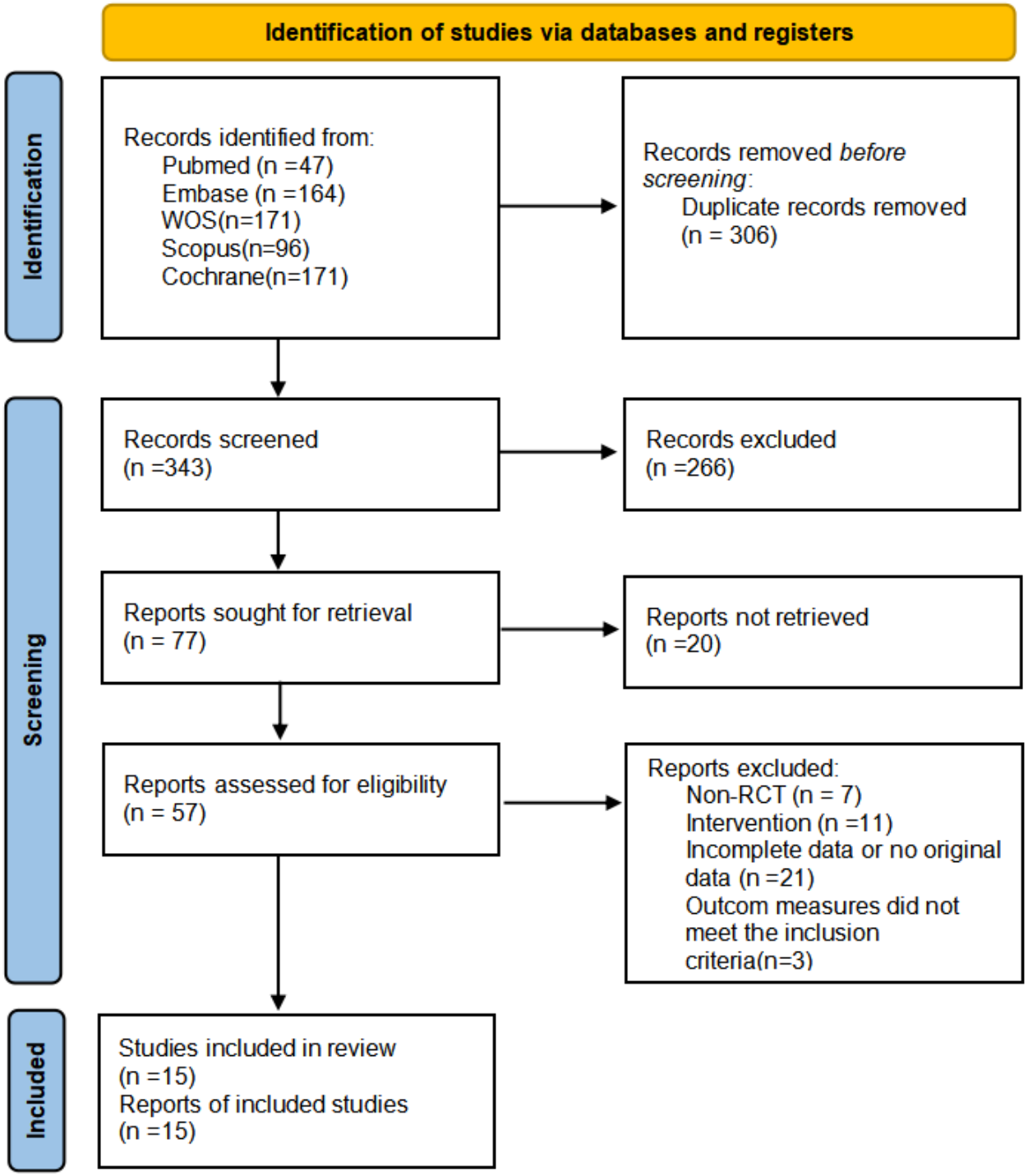
Study Selection

Figure 1 shows the PRISMA flowchart detailing the study selection process for this systematic review and meta-analysis.

XSL•FO
RenderX

The initial electronic database search yielded 649 records. After removing duplicates, 343 articles remained. Following title and abstract screening, 266 articles were excluded, leaving 77 articles eligible for full-text review. Among these, 20 articles (conference abstracts, reviews, or research protocols) were unavailable in full, leaving 57 articles for detailed evaluation. Ultimately, 15 articles met the inclusion criteria and were included in the final analysis.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart detailing the identification and selection of randomized controlled trials (RCTs) on telehealth interventions for people with Parkinson disease. WOS: Web of Science.



Study Characteristics

This systematic review included 15 RCTs [20,28-31,34-43] with a total of 765 participants. The studies were conducted across various countries: 2 in Spain, 6 in the United States, 1 in Japan (Tokyo), 3 in Italy, 1 in Sweden, 1 in Brazil, and 1 in Australia. These studies were published between 2016 and 2024. All participants were diagnosed with PD and aged 18 years or older. Sample sizes ranged from 9 to 49 participants. All trials used telehealth technologies, including telehealth systems, telephones, applications, and tablets. Follow-up durations ranged

XSL•FO
RenderX

from 1 to 12 months post intervention. Detailed study characteristics are provided in Table 1.

XSL•FO
RenderX

Table 1. Characteristics of randomized controlled trials evaluating telehealth interventions for people with Parkinson disease (N=15 studies, 2016-2024).

Author, year	Country	Sample size, T/C[a]	Age (years), T/C[b], mean (SD)	Participant	Diagnostic criteria	Telehealth technology	Follow-up time	Outcomes (with measure scales)
Cubo et al, 2017 [34]	Spain	17/18	66.44 (7.09)/ 66.05 (9.76)	PD[c]	Medical records	Telehealth system (Kinesia system included a tablet software app, a wireless finger-worn motion sensor unit, and automated web-based symptom reporting.)	12 months	PD severity (UPDRS[d], parts I-IV), the severity of nonmotor symptoms (Non-Motor Quest), QOL[e] (EQ-5D), depression and anxiety (HADS[f]), caregiver burden (Zarit Burden)
Del Pino et al, 2023 [35]	Spain	10/10	64.5 (7.9)/ 69.1 (3.5)	PD	UK Brain Bank	Telemedicine system (vCare system: a virtual training platform created based on an intelligent ICT[g] environment for rehabilitation of neurological and cardiac diseases related to aging)	4 months	QOL (Euroqol 5D), cognitive general status (MoCA[h]), PD severity (UPDRS, parts I-IV), functional disability (H&Y[i]), England ADL[j]
Dobkin et al, 2020 [36]	United States	37/35	65.62 (9.76)/ 64.80 (9.62)	PD and caregivers	National Institute of Neurological Disorders and Stroke research criteria	Telephone (using the telephone for CBT[k])	End of treatment/6 months	Depression (HAM-D[l], BDI[m]), anxiety (HAM-A[n]), QOL (SF-36[o])
Dobkin et al, 2021 [30]	United States	45/45	67.27 (7.79)/66.42 (9.51)	PD	Medical records	Telephone (using the telephone for CBT)	End of treatment/6 months	Depression (HAM-D, BDI), anxiety (HAM-A), QOL (SF-36)
Duffley et al, 2021 [37]	United States	23/19	65.0 (10.9)/64.1 (10.0)	PD and caregivers	Medical records	Telephone (using the telephone for health guidance and follow-up)	6 months	Motor symptoms (UPDRS, part III), QOL (PDQ-39[p]), caregiver burden (MCSI[q]),
Eldemir et al, 2023 [38]	Turkey	15/15	57.87 (9.79)/61.40 (7.29)	PD	UK Brain Bank	Telephone (carrying out rehabilitation training courses through telephone videoconferences)	1.5 months	ADL (UPDRS, part II), motor symptoms (UPDRS, part III), QOL (PDQ-8)
Ellis et al, 2018 [39]	United States	23/21	64.8 (8.5)/63.3 (10.6)	PD	UK Brain Bank	App (Wellpepper app: this health app provides a detailed exercise plan, including what, how, when, and where to perform exercises. Push notifications motivate users to complete their exercise and walking programs. A physical therapist remotely adjusts the regimen based on user progress. Visual progress tracking helps users monitor their performance throughout the program.)	12 months	QOL (PDQ-39), walking capacity (6-MWT[r])

XSL•FO
RenderX

Author, year	Country	Sample size, T/C[a]	Age (years), T/C[b], mean (SD)	Participant	Diagnostic criteria	Telehealth technology	Follow-up time	Outcomes (with measure scales)
Gandolfi et al, 2017 [40]	Italy	36/34	67.45 (7.18)/69.84 (9.41)	PD	UK Brain Bank	App (TeleWii-Lab: contains the Nintendo Wii console for motion control input, the Wii Fit game system, and the balance board. A laptop computer connected to a high-definition webcam was used to establish real-time remote video communication between the rehabilitation unit and the patient's residence via Skype software.)	End of treatment/1 month	QOL (PDQ-8, walking capacity (10-MWT[s]), balance (BBS[t], ABC[u])
GoffrEdo et al, 2023 [41]	Italy	49/48	67.8 (6.6)/68.2 (5.8)	PD	UK Brain Bank	Tablet (carry out motor and cognitive rehabilitation training through the VRRS[v] tablet system based on nonimmersive virtual reality.)	1.5-2.5 months	Walking capacity (6-MWT, TUG[w]), motor symptoms (UPDRS, part III), balance (mini-Balance Evaluation)
Heldman et al, 2017 [42]	United States	9/9	65.2 (10.1)/68.6 (10.2)	PD	Medical records	Telephone (conducting videoconferences or providing telephone guidance for implementing interventions via telephone)	7 months	PD severity (UPDRS, part I-IV), QOL (PDQ-39), the Patient Assessment of Chronic Illness Care
Kraepelien et al, 2020 [29]	Sweden	38/39	65.9 (8.5)/66.1 (9.8)	PD	Medical records	Telephone (using the telephone for CBT)	1.25 months/2.5 months	Depression and anxiety (HADS), QOL (PDQ-8, BBQ[x])
Maggio et al, 2024 [31]	Italy	T1: 12; T2: 12/10	T1: 59.7 (9.7); T2: 63.8 (8.3)/66.8 (6.5)	PD	Medical records	App (nonimmersive VR[y] app: NeuroNation Brain Training by Synaptikon GMBH, Berlin, offering science-based mental training to enhance various cognitive abilities with personalized data reports, and Train Your Brain by Grove FX, focusing on specific skills like concentration, spatial thinking, and reasoning. Additionally, there's a social cognitive app known as Sims Mobile.)	1.5 months/3 months	Cognitive general status (MMSE[z], MoCa), depression and anxiety (HAM-D)
Pastana Ramos et al, 2023 [28]	Brazil	8/11	60.7 (17.04)/58.6 (8.15)	PD	UK Brain Bank	Tablet\telephone (individualized telerehabilitation sessions were conducted using a tablet or mobile phone through videoconferencing and verbal guidance.)	1 month/2 months	Walking capacity (TUG, 5STS[aa]), balance (ABC), QOL (PDQ-8), motor symptoms (MDS-UPDRS[ab], part III)
Theodoros et al, 2016 [43]	Australia	15/16	71.62 (7.77)/72.86 (9.99)	PD	Medical records	Telemedicine system (eHAB: a mobile multimedia telerehabilitation system that offers real-time videoconferencing and transmits treatment data to the user's computer as images and texts. It also records high-definition live video and audio.)	1 month	QOL (PDQ-39)

XSL•FO
RenderX

Author, year	Country	Sample size, T/C[a]	Age (years), T/C[b], mean (SD)	Participant	Diagnostic criteria	Telehealth technology	Follow-up time	Outcomes (with measure scales)
Wilkinson et a, 2016 [20]	United States	Arm 1: 26/24; Arm 2: 18/18	Arm 1: 76.1 (8.4)/76.1 (7.9); Arm 2: 67.2 (9.8)/70.9 (8.4)	PD	UK Brain Bank	Telemedicine system (a global health care telemedicine specialist cart and Cisco webcam provide real-time high-definition audio-visual connectivity between patients and health care providers.)	6 months/12 months	PD severity (Arm 1: UPDRS, part I-IV; Arm 2: H&Y), QOL (PDQ-8), depression and anxiety (GDS[ac])

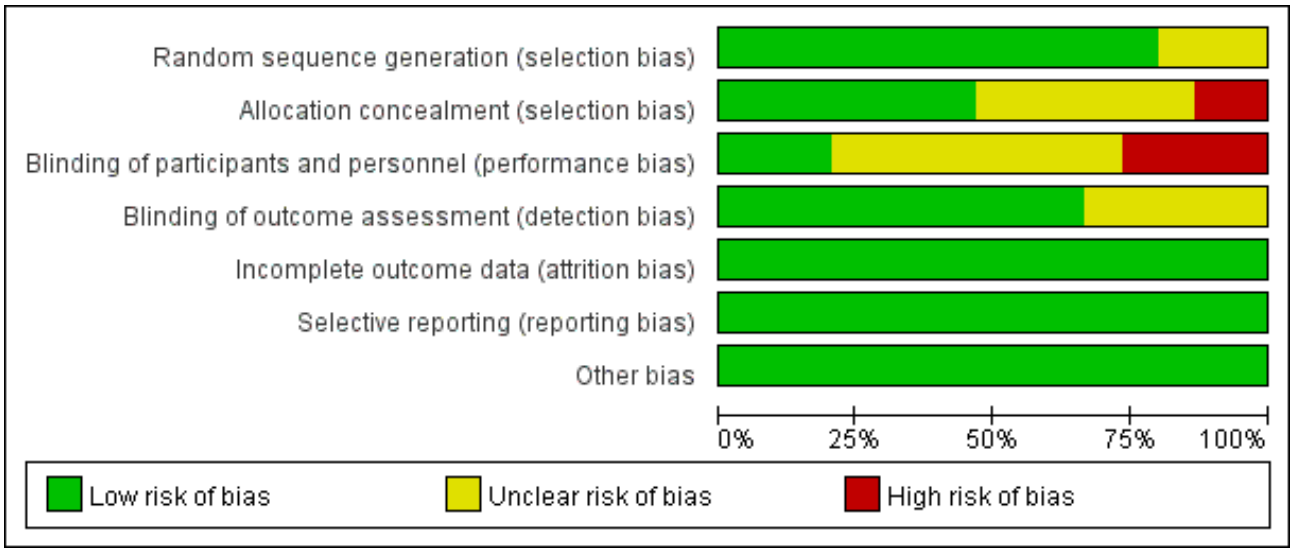
[a]T/C: sample size in treatment group/control group.
[b]T/C: mean (SD) values in treatment group/control group.
[c]PD: Parkinson disease.
[d]UPDRS: Unified Parkinson's Disease Rating Scale.
[e]QOL: quality of life.
[f]HADS: Hospital Anxiety and Depression Scale.
[g]ICT: digital information and communication technologies.
[h]MoCA: Montreal Cognitive Assessment.
[i]H&Y: Hoehn and Yahr Scale.
[j]ADL: activities of daily living.
[k]CBT: cognitive behavioral therapy.
[l]HAM-D: Hamilton Depression Rating Scale.
[m]BDI: Beck Depression Inventory.
[n]HAM-A: Hamilton Anxiety Rating Scale.
[o]SF-36: Medical Outcomes Study 36-Item Short Form Health Survey.
[p]PDQ: Parkinson Disease Questionnaire.
[q]MCSI: Michigan Consumer Sentiment Index.
[r]6-MWT: 6-minute walk test.
[s]10-MWT: 10-minute walk test.
[t]BBS: Berg Balance Scale.
[u]ABC: Activities-specific Balance Confidence scale.
[v]VRRS: Virtual Router Redundancy Service.
[w]TUG: Timed Up and Go Test.
[x]BBQ: Brunnsviken Brief Quality of Life Scale.
[y]VR: virtual reality.
[z]MMSE: Mini-Mental State Examination.
[aa]5STS: 5 Times Sit-to-Stand Test.
[ab]MDS-UPDRS: Movement Disorders Society-Unified Parkinson's Disease Rating Scale.
[ac]GDS: Geriatric Depression Scale.

Risk of Bias

The methodological quality and potential biases of the included studies were assessed using the Cochrane Risk of Bias tool [33]. Figure 2 presents detailed findings. All 15 studies were confirmed as randomized, and 12 studies [20,28,30,31,34,36-41,43] clearly described methods of random sequence generation. Only 7 studies [28-30,36,38,39,41] detailed allocation concealment methods, categorizing them as having a low risk of selection bias. Regarding the blinding of participants and intervention providers, only 3 studies [36,39,41] reported adequate blinding procedures and thus had a low risk of performance bias. In contrast, 4 studies [28,37,40,43] without blinding were classified as having a high risk of performance bias. The remaining 8 studies [20,29-31,34,35,38,42] lacked sufficient information and were classified as unclear. Blinding of outcome assessors was reported in 10 studies [28,30,36-43], indicating a low risk of detection bias, while 5 studies [20,29,31,34,35] did not report this information. All studies provided complete outcome data, indicating low attrition bias. Assessment for other potential biases generally indicated low risk. Funnel plot analysis (Multimedia Appendix 4) revealed no significant publication bias among included studies.

XSL•FO
RenderX

Figure 2. Risk of bias assessment (using the Cochrane Risk of Bias tool) for the included randomized controlled trials evaluating telehealth interventions for Parkinson disease.



Meta-Analysis

Quality of Life

A total of 11 studies [20,28-30,36-40,42,43] evaluated QOL following telehealth interventions, using 5 different measurement tools. Overall, 9 studies used Parkinson Disease Questionnaire (PDQ) scales (PDQ-39 [37,39,42,43] and PDQ-8 [20,28,29,38,40]), with higher scores indicating worse QOL. Additionally, 2 studies used the Medical Outcomes Study 36-Item Short Form Health Survey (SF-36) [30,36], and 1 study incorporated both PDQ-8 and Brunnsviken Brief Quality of Life Scale (BBQ) [29], with higher scores reflecting better QOL. Due to measurement variability, a standardized mean difference (SMD)–based meta-analysis was conducted following Cochrane handbook guidelines.

A 3-level random-effects meta-analysis of 8 studies (13 effect sizes) assessing QOL using PDQ scales revealed a marginal trend toward improvement after telehealth interventions (SMD –0.42, 95% CI –0.88 to 0.03; P=.07). Heterogeneity was present (τ^2=0.258; I^2=56.2%; Q_{12}=27.37; P=.007), with between-study variance accounting for 65.5% of total variability. Prediction intervals were not calculated due to the limited number of studies (k<10; Table 2; Figures S1-S3 in Multimedia Appendix 4). Subgroup analysis demonstrated significant moderation by intervention type (test of moderators=9.84; P=.002). Telephone-based interventions significantly improved QOL (SMD –0.83, 95% CI –1.22 to –0.44; P<.001), whereas digital interventions had minimal effect (SMD –0.05, 95% CI –0.58 to 0.49; P=.86). The between-group difference was 0.78 (95% CI 0.29-1.27; P=.002) points. Follow-up duration did not significantly moderate outcomes (categorical: P=.58; continuous: β=.01/month; P=.83; Table 3; Figures S4-S7 in Multimedia Appendix 4). Sensitivity analysis identified substantial influence from Ramos et al [28], whose exclusion reduced the effect magnitude by 52.2%. Multilevel Egger test suggested possible small-study effects (P=.02; Figures S8-S9 in Multimedia Appendix 4).

XSL•FO
RenderX

Table 2. Results of 3-level random-effects meta-analyses for primary outcomes in patients with Parkinson disease receiving telehealth interventions.

Outcome	*k* [a]	*#ES* [b]	Mg[c]	95% CI	*P* value	σ^2_{level2}[d]	σ^2_{level3}[e]	Variance level 1 (%)	Variance level 2 (%)	Variance level 3 (%)
QOL[f,g]	8	13	–0.42	–0.88 to 0.034	.07	0.000	0.258	34.5	0	65.5
QOL[h]	3	5	0.39	0.06 to 0.72	.03	0.019	0.000	72.3	27.7	0
Anxiety	3	5	–0.64	–0.92 to –0.35	.003	0.000	0.000	100	0	0
Cognitive function	2	9	1.12	0.03 to 2.20	.045	0.192	0.275	33.9	27.2	38.7
Depression	5	17	–0.64	–0.93 to –0.34	<.001	0.016	0.062	55.4	8.9	35.6
Motor symptoms	7	11	–0.46	–0.69 to –0.24	<.001	0.000	0.000	100	0	0

[a]*k*=number of studies.

[b]*#ES*=number of effect sizes.

[c]Mg=mean effect size (g).

[d]σ^2_{level2}=variance between effect sizes extracted from the same study.

[e]σ^2_{level3}=variance between studies.

[f]QOL: quality of life.

[g]The 3-level meta-analysis of 8 studies (13 effect sizes) assessing QOL with Parkinson Disease Questionnaire.

[h]The 3-level meta-analysis of 3 studies (5 effect sizes) assessing QOL with Medical Outcomes Study 36-Item Short Form Health Survey and Brunnsviken Brief Quality of Life Scale.

XSL•FO
RenderX

Table 3. Moderator analyses examining the effects of intervention type and follow-up duration on primary outcomes in the meta-analysis.

	k [a]	*#ES* [b]	B_0/g [c]	t_0 [d]	B_1 [e]	t_1 [f]	*F* test (*df*) [g]
QOL[h,i]							
Intervention type	8	13	–0.83	–4.179[j]	0.78	3.137[j]	9.842[j] (1, 11)
Follow duration[k]	8	13	–0.53	–1.852	0.26	0.550	0.302 (1, 11)
Follow duration[l]	8	13	–0.46	–0.167	0.01	0.219	0.048 (1, 11)
QOL[m]							
Follow duration[k]	3	5	0.73	3.033[j]	–0.42	–0.157	2.470 (1, 3)
Follow duration[l]	3	5	0.69	3.369[n]	–0.06	–1.713	2.937 (1, 3)
Anxiety							
Follow duration[k]	3	5	–0.51	–2.206[o]	–0.16	–0.605	0.366 (1, 3)
Follow duration[l]	3	5	–0.66	–3.179[j]	0.004	0.124	0.015 (1, 3)
Cognitive function							
Follow duration[k]	2	9	1.39	2.876[j]	–0.31	–0.654	0.427 (1, 7)
Follow duration[l]	2	9	2.07	3.125[j]	–0.30	–0.186	1.407 (1, 7)
Depression							
Intervention type	5	17	–0.84	–9.933[n]	0.53	3.416[n]	11.669[n] (1, 15)
Follow duration 1[k]	5	17	–0.54	–1.979[o]	–0.13	–0.419	0.175 (1, 15)
Follow duration 2[l]	5	17	—0.87	–4.798[n]	0.04	1.860	0.346 (1, 15)
Motor symptoms							
Intervention type	7	11	—0.54	—3.542[n]	0.13	0.660	0.436 (1, 9)
Follow duration[k]	7	11	–0.58	–3.624[n]	0.20	0.949	0.901 (1, 9)
Follow duration[l]	7	11	–0.55	–3.037[j]	0.015	0.595	0.354 (1, 9)

[a]k=number of independent studies.

[b]*#ES*=number of effect sizes.

[c]B_0/mean g=intercept/mean effect size (g).

[d]t_0=*t* value for mean g.

[e]B_1=estimated regression coefficient.

[f]t_1=*t* value for regression coefficient.

[g]Omnibus *F* test.

[h]QOL: quality of life.

[i]The 3-level meta-analysis of 8 studies (13 effect sizes) assessing QOL with Parkinson Disease Questionnaire.

[j]*P*<.01.

[k]Results of analysis based on categorical variables.

[l]Results of analysis based on continuous variables.

[m]The 3-level meta-analysis of 3 studies (5 effect sizes) assessing QOL with the Medical Outcomes Study 36-Item Short Form Health Survey and Brunnsviken Brief Quality of Life Scale.

[n]*P*<.001.

[o]*P*<.05.

For 3 studies assessing QOL using SF-36 and BBQ (5 effect sizes), the 3-level random-effects meta-analysis indicated significant improvement after telehealth interventions (SMD 0.39, 95% CI 0.06-0.72; *P*=.03). Minimal heterogeneity was observed (τ^2=0.020; I^2=25.9%; Q_4=5.40; *P*=.25), with within-study variance contributing 27.7% of total variability. Prediction intervals were not calculated (*k*<10; Table 2; Figures S11-S12 in Multimedia Appendix 4). Subgroup analyses indicated greater effects in short-term follow-ups (<3 months: SMD 0.73, 95% CI 0.26-1.19) compared with longer-term

XSL•FO
RenderX

follow-ups (≥3 months: SMD 0.31, 95% CI –0.40 to 0.43; β=–.42; *P*=.12). A marginal negative correlation between effect magnitude and follow-up duration was observed (β=–.06/month; *P*=.09; Table 3; Figures S13-S15 in Multimedia Appendix 4). Sensitivity analysis confirmed robustness; no single-study exclusion altered significance substantially (maximum change: +33.2% when excluding Dobkin et al [36]). The Egger test indicated no significant small-study effects (*P*=.28; Figures S16-S17 in Multimedia Appendix 4).

Depression

A total of 5 studies [20,29-31,36] evaluated depression levels in patients with PD following telehealth interventions, using 4 distinct assessment tools: Hamilton Depression Rating Scale (HAM-D) [30,31,36], Beck Depression Inventory (BDI) [30,36], Hospital Anxiety and Depression Scale-Depression (HADS-D) [29], and Geriatric Depression Scale [20]. Higher scores on these scales indicate increased depression severity. Two studies used both the HAM-D and BDI to assess depression. A 3-level random-effects meta-analysis (5 studies, 17 dependent effect sizes) with t-distribution–based inference revealed a significant reduction in depression symptoms after telehealth interventions (SMD –0.64, 95% CI –0.93 to –0.34; *P*<.001). Moderate heterogeneity was observed (I^2=45.9%; Q_{16}=29.60; *P*=.02). Variance component analysis indicated that between-study differences accounted for 35.6% of total variability (τ^2=0.062), within-study variability explained 8.9% (τ^2=0.016), and sampling error contributed 55.4% (Table 2; Figures S17-S19 in Multimedia Appendix 4). Prediction intervals were not calculated because the number of studies was below the recommended threshold (*k*<10). Subgroup analysis showed significantly greater improvements with traditional telehealth (SMD –0.84, 95% CI –1.00 to –0.67; *P*<.001) compared with digital interventions (SMD –0.31, 95% CI –0.54 to –0.08; *P*=.008; β=.53; *P*<.001). Follow-up duration was not significant in categorical analysis (β=–.13; *P*=.68), but continuous analysis showed a marginal positive association (β=.04/month; *P*=.06; Table 3; Figures S20-S23 in Multimedia Appendix 4). Sensitivity analyses confirmed robustness; effects remained significant after excluding each study (SMD range –0.55 to –0.79). The Egger test indicated no significant small-study effects (*P*=.82; Figures S24-S25 in Multimedia Appendix 4).

Anxiety

A total of 3 studies [29,30,36] assessed anxiety levels in patients with PD using 2 measurement scales: Hamilton Anxiety Rating Scale (HAM-A) [30,36] and Hospital Anxiety and Depression Scale-Anxiety (HADS-A) [29]. Higher scores represent greater anxiety severity. A 3-level random-effects meta-analysis (3 studies, 5 dependent effect sizes) with t-distribution–based inference indicated significant anxiety reduction following telehealth interventions (SMD –0.64, 95% CI –0.92 to –0.35; *P*=.003). Negligible heterogeneity was observed (I^2=0%; Q_4=0.90; *P*=.92), with variance component analysis indicating that sampling error explained all variability (τ^2=0 for both inter- and intrastudy variance; Table 2; Figures S26-S28 in Multimedia Appendix 4). Prediction intervals were not calculated due to the limited number of studies (*k*<10). All studies used telephone-based interventions, thus subgroup analyses examined follow-up durations. Follow-up duration was not significantly associated with outcomes in categorical (β=–.16, 95% CI –0.66 to 0.35; *P*=.55) or continuous analyses (β=.004/month, 95% CI –0.06 to 0.07; *P*=.90; Table 3; Figures S29-S31 in Multimedia Appendix 4). Sensitivity analyses confirmed robustness, with effects remaining significant upon exclusion of individual studies (SMD range –0.67 to –0.60), and maximum deviation of –5.4% when excluding Dobkin et al [30]. Funnel plots were symmetrical, and the Egger test revealed no significant small-study effects (*P*=.68; Figures S32-S33 in Multimedia Appendix 4).

Motor Symptoms

A total of 7 studies [20,28,34,35,38,41,42] used the MDS-UPDRS-III to evaluate motor symptoms in patients with PD post telehealth interventions, with higher scores indicating greater severity. A 3-level random-effects meta-analysis (7 studies, 11 dependent effect sizes) with t-distribution–based inference demonstrated significant improvements in motor symptoms after telehealth interventions (SMD –0.46, 95% CI –0.69 to –0.24; *P*=.001). Heterogeneity among studies was negligible (I^2=0%; Q_{10}=7.85; *P*=.64). Variance component analysis showed that sampling error accounted for all variability, with negligible between-study (τ^2=0) and within-study variance (τ^2=0; Table 2; Figures S34-S36 in Multimedia Appendix 4). Prediction intervals were not calculated because the number of studies was below the recommended threshold (*k*<10). Subgroup analysis by intervention type revealed no significant difference between digital and other telehealth interventions (β=.13, 95% CI –0.26 to 0.53; *P*=.51). Follow-up duration did not significantly moderate outcomes in categorical (β=.20, 95% CI –0.21 to 0.60; *P*=.34) or continuous analyses (β=.02/month, 95% CI –0.03 to 0.06; *P*=.55; Table 3; Figures S37-S39 in Multimedia Appendix 4). Sensitivity analyses confirmed robustness, with significant effects maintained after each study's exclusion (SMD range –0.52 to –0.42), and the largest change being 11.5% when excluding Wilkinson et al [20]. Funnel plots showed symmetry, and the Egger test indicated no significant small-study effects (*P*=.88; Figures S40-S41 in Multimedia Appendix 4).

Activities of Daily Living

A total of 4 studies [34,35,38,42] evaluated the impact of telehealth interventions on daily activities in patients with PD using the MDS-UPDRS-II scale. Higher scores represent greater impairment. A random-effects meta-analysis with HKSJ correction demonstrated that telehealth interventions significantly reduced impairment in daily activities compared with controls (SMD –0.79, 95% HKSJ-adjusted CI –1.04 to –0.54; *P*=.002). Heterogeneity was negligible (τ^2=0.000; I^2=0.0%; Q_3=0.43; *P*=.93). Prediction intervals were not calculated because the number of studies was below the recommended threshold (*k*<10). Sensitivity analysis (leave-one-out) confirmed the robustness of the findings. The largest change in effect size occurred after excluding Del Pino et al [35] (–7.9% change; SMD range across exclusions –0.83 to –0.73). Egger test indicated no significant small-study effects

XSL•FO
RenderX

(intercept=–1.35; *P*=.31; Figure S42 in Multimedia Appendix 4).

Cognition

A total of 2 studies [31,35] examined cognitive outcomes in patients with PD after telehealth interventions using the Montreal Cognitive Assessment (MoCA) and Mini-Mental State Examination (MMSE), where higher scores indicate better cognitive function. One study used both the MoCA and MMSE scales to comprehensively assess cognition. A 3-level random-effects meta-analysis (2 studies, 9 dependent effect sizes) with t-distribution–based inference indicated significant cognitive improvement following telehealth interventions (SMD 1.12, 95% CI 0.03-2.20; *P*=.045). Moderate heterogeneity was detected (I^2=52.3%; Q_8=16.77; *P*=.03). Variance component analysis showed that between-study differences accounted for 38.9% (τ^2=0.275) of total variability, within-study differences accounted for 27.2% (τ^2=0.192), and sampling error explained 33.9% (Table 2; Figures S43-S45 in Multimedia Appendix 4). Prediction intervals were not calculated due to the limited number of studies (*k*<10). A follow-up duration did not significantly moderate outcomes in categorical (β=–.31, 95% CI –1.25 to 0.62; *P*=.51) or continuous analysis (β=–.30/month, 95% CI –0.79 to 0.20; *P*=.24; Table 3; Figures S46-S48 in Multimedia Appendix 4). The Egger test indicated significant small-study effects (intercept=15.17; *P*<.001). Sensitivity analysis was not feasible due to the limited number of studies (Figure S49 in Multimedia Appendix 4).

Discussion

Overview

This systematic review evaluated the effects of telehealth interventions on the QOL and associated health outcomes in patients with PD. Findings demonstrated that telehealth interventions significantly enhanced various dimensions of patient well-being, including QOL, depressive symptoms, anxiety levels, motor function, ADL, and cognitive abilities. These results differ from earlier reviews and meta-analyses [8,26,27]. The discrepancies might be due to accelerated advancements in telehealth and remote neurology after the epidemic, alongside improvements in telehealth service quality and increased research volume [22,44]. Although the effectiveness of telehealth interventions appears promising, additional studies are needed to establish more conclusive evidence.

Effectiveness of Telehealth Interventions on QOL in Patients With PD

Our meta-analysis identified complex patterns regarding telehealth's effect on QOL in patients with PD. Interventions assessed by the SF-36/BBQ demonstrated a significant improvement, whereas those assessed using PDQ scales indicated only marginal benefit. This difference likely stems from the fundamental distinctions between scales: PDQ scales specifically measure PD-related deficits, whereas SF-36/BBQ assess general well-being [45,46]. Notably, telephone-based interventions substantially improved PDQ-based QOL, while digital interventions showed negligible effects. This distinction explains inconsistencies in prior meta-analyses [1,26,27] that did not account for intervention modality. Significant heterogeneity in PDQ analyses was primarily attributed to between-study differences, suggesting methodological variations. Sensitivity analysis highlighted substantial influence from Ramos et al [28], and potential small-study effects were noted. These factors necessitate cautious interpretation of PDQ outcomes. Conversely, SF-36/BBQ analyses exhibited minimal heterogeneity and robust sensitivity. Differences in results may also reflect variations in intervention type, therapeutic intensity, assessment timing, methodological quality, and sample sizes [27]. Follow-up duration did not moderate effects, contradicting assumptions that longer interventions yield superior outcomes. Instead, the intervention modality emerged as the critical moderator, highlighting the importance of direct human interaction in managing PD-specific QOL concerns. This aligns with telehealth's capacity for dynamic therapeutic engagement [47,48], particularly beneficial for isolated patients [10], although our findings indicate that these advantages depend on modality.

Telehealth provides more dynamic, immersive methods for treatment, education, and counseling compared to traditional medical approaches, enhancing patient engagement and interaction [47,48]. Such enhancements assist patients with PD and their families in comprehensively understanding and managing disease-related challenges, thus promoting independence, motivation for self-care, and improved life quality [49]. Patients with PD require consistent engagement with health care teams for effective management of disease progression and treatment complexity [8]. A primary advantage of telehealth lies in serving patients in isolated or underserved areas, addressing health care provider shortages, and offering timely, high-quality care to improve patient outcomes [10]. Additionally, economic burdens and logistical difficulties substantially reduce patients with PD's QOL [50,51]. By reducing health care–related costs and travel demands, telehealth can expand home-based medical services, further enhancing life quality for patients with PD [10,22].

Effectiveness of Telehealth Interventions on Depression in Patients With PD

Our meta-analysis demonstrated a significant antidepressant effect of telehealth interventions. Traditional approaches (telephone or cognitive behavioral therapy [CBT]–based) showed nearly 3 times the efficacy compared to digital interventions. This advantage aligns with neurobiological evidence linking depression in PD to dysfunction in serotonergic pathways and frontostriatal circuits [52], suggesting human-mediated therapies more effectively modulate emotional processing compared to automated digital tools. This finding notably diverges from earlier studies, which primarily emphasized motor symptoms and physical rehabilitation, often neglecting depressive symptoms [10,27]. The distinct focus of 3 specific studies [29,30,36] included in this meta-analysis may explain this difference. These studies emphasized cognitive and behavioral aspects of patient care, integrating telehealth with CBT, a method recognized for effectively reducing depression levels [53,54]. Moderate heterogeneity mainly resulted from

XSL•FO
RenderX

between-study methodological differences, likely reflecting variations in measurement tools and intervention protocols. Contrary to expectations, categorical follow-up duration showed no significant moderating effect, although continuous analysis revealed a marginal positive association. This result suggests sustained engagement—particularly through telephone-based CBT [53,54]—might progressively reinforce neuroplastic changes in emotion-regulation networks. These findings reconcile previous contradictions in the literature [1,26,27]. Whereas earlier reviews primarily targeted motor symptoms, our analysis confirms telehealth's antidepressant benefits when including behavioral interventions tailored to PD-related psychopathology.

Our findings align with those of Dou et al [55], indicating that telehealth interventions (tele-CBT and telerehabilitation training) significantly improve depressive symptoms in patients with PD. Feasibility has been verified internationally; for example, a cross-sectional study in Brazil showed effective telehealth implementation in resource-limited settings with high patient satisfaction [56,57]. Regarding specific methods, tele-motor training significantly enhanced patients' motor function and indirectly alleviated depressive symptoms [55]. In contrast, tele-CBT directly targeted depressive and anxiety symptoms, showing greater effectiveness compared to other teleinterventions [31]. Teleconsultation had relatively limited efficacy in alleviating depressive symptoms but significantly improved access to medical resources [58]. From a neurobiological standpoint, PD and depression share common pathological mechanisms, including gut microbiota dysregulation, neuroinflammation, and reward-processing dysfunction [52]. Telehealth, especially through behavioral interventions such as CBT, may modulate these pathological processes and consequently alleviate depressive symptoms [59]. However, existing evidence suggests telehealth's effectiveness may be weaker for chronic, nonepisodic mental disorders (eg, depression in PD) compared to primary depression [60]. To optimize telehealth potential, future research should investigate long-term outcomes, standardization of techniques, and cybersecurity considerations [61,62]. In summary, telehealth effectively reduces depressive symptoms in patients with PD, especially via tele-CBT, which overcomes geographical barriers and improves treatment accessibility. Nevertheless, individualized plans and sustained follow-up are necessary to achieve optimal therapeutic outcomes.

Effectiveness of Telehealth Interventions on Anxiety in Patients With PD

Our meta-analysis demonstrated robust anxiolytic effects of telephone-based telehealth interventions, with remarkable consistency across studies. This homogeneity suggests that telephone-delivered CBT provides a reliably standardized approach for managing PD-related anxiety. Notably, these benefits remained stable irrespective of follow-up duration, indicating sustained therapeutic effects without attenuation over 3-9 months. These findings resolve previous contradictions [14,26,27] by demonstrating that structured tele-CBT can effectively address PD-specific anxiety mechanisms, including fear-avoidance cycles and "off"-period distress resistant to conventional treatments. The negligible heterogeneity, with variance entirely attributable to sampling error, likely reflects 3 factors. First, interventions used standardized CBT protocols targeting PD-specific anxiety mechanisms such as hypervigilance toward motor fluctuations. Second, the uniform application of validated and sensitive scales (HAM-A/HADS-A) ensured measurement precision. Third, telephone delivery strengthened therapeutic alliances through real-time emotional interaction absent in purely digital interfaces. The integration of standardized protocols, precise assessments, and person-centered delivery resulted in methodological consistency across studies.

Although considerable evidence supports telehealth for anxiety in patients with PD, its exact mechanism and broader applicability require further investigation. Previous studies [63,64] showed comparable efficacy of telehealth and face-to-face interventions in reducing anxiety, depression, and stress scores, alongside improved heart rate variability. Anxiety reductions persisted long-term after telehealth interventions, confirming their noninferiority to in-person care. This effectiveness largely stems from multimodal interventions; for example, remote CBT overcomes movement-related barriers and, combined with exercise and biomarker monitoring, allows personalized care beneficial to underserved populations [55,64,65]. However, some research highlights intervention heterogeneity. A small study [66] indicated that telephone CBT effectively alleviated depression but not anxiety symptoms in patients with PD, suggesting anxiety may require more tailored strategies. Our analysis, in contrast, supports the long-term feasibility, effectiveness, and durability of telephone CBT effects. Earlier discrepancies might stem from small sample sizes or limitations of measurement tools. Although the revised Parkinson Anxiety Scale improved cultural adaptability, general scales (eg, HADS) may underestimate actual effectiveness due to limited sensitivity [29,67]. Future research should expand sample sizes, develop PD-specific anxiety interventions, and integrate multidimensional biomarker monitoring to improve telehealth precision and applicability.

Effectiveness of Telehealth Interventions on Motor Symptoms of Patients With PD

Telehealth interventions significantly improved motor symptoms in patients with PD. This refined estimate may reflect advancements in methodological rigor involving multilevel analyses that account for independent effect sizes, an approach not consistently used in previous meta-analyses [1,26]. Due to the standardized use of MDS-UPDRS-III assessments and similar intensities of interventions, we observed remarkably low heterogeneity among studies. Notably, digital and traditional telehealth approaches demonstrated comparable effectiveness, indicating that essential motor rehabilitation components, such as amplitude training and balance exercises, effectively translated across different treatment platforms. The temporal stability of benefits further supported telehealth as a sustainable management option, with sensitivity analyses confirming robustness to study exclusion.

A primary therapeutic objective in PD involves improving motor symptoms, wherein treatment adjustments frequently depend on accurate motor assessments [68]. The telehealth framework

XSL•FO
RenderX

enables improved and timely interactions between patients and health care providers compared to traditional face-to-face consultations, allowing for more individualized rehabilitation strategies tailored specifically to patients with movement disorders [69]. Multiple studies have confirmed the significant impact of telehealth interventions in alleviating motor symptoms in patients with PD. For instance, structured telerehabilitation programs, such as the Lee Silverman Voice Treatment BIG rehabilitation method, have effectively enhanced motor function, alleviated nonmotor symptoms, and improved the QOL for patients with PD [70]. Compared to teleconsultations alone, tele-motor interventions demonstrate superior efficacy in motor function improvement [55]. From a neuromechanism perspective, cueing techniques activate the motor cortex, thereby enhancing the stability of motor output, which provides scientific justification for using cue-based strategies in telerehabilitation [71]. Moreover, telerehabilitation is particularly suitable for patients with restricted mobility or those residing in medically underserved regions. Real-time video guidance ensures continuous rehabilitation training, effectively overcoming geographical limitations. Its safety and potential effectiveness in improving balance and functional activities have been confirmed by existing research [28,41,72]. Telehealth facilitates comprehensive monitoring of treatment effects through standardized scales (such as MDS-UPDRS) for assessing motor symptoms, combined with evaluation of nonmotor symptoms and QOL questionnaires [73,74]. Additionally, tele-motor interventions based on live-streaming have been proven feasible and safe, demonstrating high patient adherence (eg, twice a week) and thus confirming their practical use for continuous management of motor symptoms in PD [75]. Therefore, telehealth effectively enhances motor functions in patients with PD, offering advantages in personalized program design, activation of neural plasticity, and overcoming limitations in medical resource availability. With ongoing advancements in assessment instruments and technological integration, telehealth is anticipated to further improve long-term intervention outcomes.

Effectiveness of Telehealth Interventions on ADL in Patients With PD

The results of this study showed that telehealth interventions significantly improved ADL among patients with PD. This result aligns with previous research, reinforcing that remote health care interventions significantly enhance both ADL performance and motor symptoms in individuals with PD. Our analysis suggests that the significant improvements in ADL resulting from telehealth are due to multidimensional intervention strategies addressing the core symptoms of PD.

Relevant studies have shown that structured remote rehabilitation programs, delivered through real-time video instruction, enhance functional mobility and directly improve basic ADL tasks such as walking and dressing [57]. Simultaneously, high-intensity remote exercise interventions reduce motor sluggishness and freezing of gait, indirectly enhancing instrumental ADLs, such as complex daily activities like shopping and meal preparation [76]. The simultaneous improvements observed in ADLs and motor symptoms share clear pathophysiological connections; enhanced motor functions directly alleviate limitations in physical activity, enabling patients to execute daily routines more effectively [77]. Additionally, remote CBT improves executive functions, mitigating motor-related restrictions on complex ADL performance [77,78]. Telehealth frameworks achieve these synergistic effects by integrating 3 primary components: real-time video supervision ensures adherence to exercise regimens; home-based cognitive training modules restructure the prefrontal-limbic circuitry; and wearable sensors provide immediate feedback regarding movement quality [57,79,80]. Therefore, telehealth interventions positively and synergistically influence both motor symptoms and ADL performance in patients with PD. Future research should focus on optimizing intervention strategies and integrating motor and ADL training components comprehensively to further enhance the overall QOL for patients.

Effectiveness of Telehealth Interventions on Cognition in Patients With PD

Preliminary evidence shows that telehealth interventions may enhance cognitive function in patients with PD; however, these findings should be interpreted cautiously. Although statistically significant, effect sizes exhibited substantial variability, ranging from negligible to considerable clinical improvement. This observed heterogeneity primarily stems from methodological differences among studies, potentially reflecting (1) the use of varied cognitive assessments (MoCA vs MMSE), each with differing sensitivities to PD-specific cognitive deficits, and (2) distinct intervention protocols within the limited scope of available evidence. Additionally, significant publication bias and insufficient data for sensitivity analyses further limit definitive conclusions.

Overall, the efficacy of telehealth interventions for enhancing cognitive functions in patients with PD has been established. These interventions significantly improve cognitive status, particularly executive functions and memory, as well as emotional and behavioral disorders, consequently enhancing the QOL for both patients and caregivers [78,81]. Among specific intervention methods, computer-assisted cognitive training has shown potential benefits for patients with PD accompanied by mild cognitive impairment, with feasibility confirmed for home-based training modalities [82,83]. Moreover, remote virtual reality applications (telehealth virtual reality) have shown promising results for improving cognitive task performance [31]. A recent network meta-analysis further supports the beneficial effects of remote interventions, including remote cognitive training, on cognition and other nonmotor symptoms [55]. The primary advantages of telehealth interventions include high accessibility (especially beneficial for patients with limited mobility or those residing in remote areas) and flexibility, with the patient's cognitive reserve potentially enhancing treatment effect [84]. However, considerable heterogeneity exists within current evidence, aligning with our findings. This heterogeneity is largely attributed to variations in study design, inconsistencies in cognitive assessment tools, and diverse responses among patient subtypes [55,82,85]. In addition, the efficacy of telehealth interventions differed across cognitive domains. Therefore, these findings should be considered exploratory and interpreted

XSL•FO
RenderX

cautiously. In conclusion, telehealth represents a promising cognitive management approach for PD with substantial potential; nevertheless, implementation barriers must be considered, strategies tailored to individual patient needs, and larger standardized trials conducted to further substantiate effectiveness.

Strengths and Limitations

This systematic review benefits from a rigorous methodological approach, using a meta-analysis grounded in RCTs and strictly adhering to established guidelines for systematic reviews. All analyses were conducted using random-effects models based on conceptual considerations, with HKSJ or t-distribution–based corrections applied to provide more accurate and conservative CIs. This significantly enhances the credibility of the findings. Furthermore, the review assesses the impact of telehealth interventions not only on QOL but also on multiple health-related domains such as depression, anxiety, motor function, ADL, and cognitive function in patients with PD, rather than restricting its focus solely to treatment modalities.

Nonetheless, several limitations of this review should be acknowledged. First, the limited number of RCTs included in the analysis may constrain the generalizability of these conclusions to broader populations. Second, due to the small number of studies per outcome (all $k<10$), prediction intervals were not calculated, which limits the interpretation of how the true effect may vary across different settings. Third, funnel plots and Egger tests were used to assess small-study effects, but these methods have reduced accuracy when fewer than 10 studies are analyzed per outcome, and they do not specifically measure publication bias. Finally, although we applied multilevel modeling to account for dependent effect sizes, residual heterogeneity, and variations in intervention protocols may still influence the results. Therefore, additional RCTs must be incorporated into future research to enhance the robustness and reliability of the findings.

Implications for Practice

Global disparities in medical resource distribution present substantial challenges to health care service advancement. This issue is particularly pronounced in neurological care, where specialist availability is limited, notably in suburban and rural regions. Consequently, many individuals with PD struggle to receive continuous medical support, resulting in significant declines in their QOL as the disease progresses. This situation places considerable strain not only on patients and their families but also on societal resources. The emergence of telehealth, however, offers a promising solution by providing innovative avenues for managing and treating PD. Telehealth has the potential to bridge existing gaps, enabling patients with PD who previously had limited or no access to receive essential health care services.

Implications for Further Research

PD exerts substantial impacts on public health, prompting significant attention from the health care community toward preventative, diagnostic, and therapeutic strategies. As an innovative product of rapid technological advancement, telehealth represents a cost-effective, real-time, and secure platform for collecting patient data, significantly facilitating the diagnosis, monitoring, and rehabilitation of PD. Nonetheless, the efficacy of telehealth requires further validation through comprehensive and rigorous RCTs. Future research should not only evaluate functional recovery, cognitive enhancement, and health-related QOL but also examine aspects such as cost-effectiveness, patient satisfaction, and digital health literacy among older adults. Such investigations will facilitate more informed decisions and optimal tailoring of telehealth interventions for patients with PD.

Conclusion

Telehealth interventions have demonstrated the potential to significantly enhance various aspects of life among patients with PD, including alleviating symptoms of depression and anxiety, improving motor function, facilitating ADL, and enhancing cognitive performance. Despite these encouraging findings, there remains an urgent need for meticulously designed, large-scale RCTs to comprehensively evaluate telehealth's effectiveness across the full spectrum of PD management.

Acknowledgments

We would like to express our heartfelt thanks to all those involved in this study.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

Conceptualization: MS, FT, Luomin
Data curation: MS, FT
Formal analysis: MS, FT
Investigation: MS, FT
Methodology: FT, Luomin
Project administration: Luomin
Software: MS
Validation: S Wang, S Wen

XSL•FO
RenderX

Visualization: HJ
Writing - original draft: MS
Writing - review & editing: MS, FT, Luomin, S Wen, S Wang, HJ

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA 2020 checklist.
[DOCX File , 29 KB - mhealth_v14i1e70994_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOC File , 45 KB - mhealth_v14i1e70994_app2.doc]

Multimedia Appendix 3
R language code.
[DOCX File , 25 KB - mhealth_v14i1e70994_app3.docx]

Multimedia Appendix 4
Supplementary figures.
[DOC File , 6938 KB - mhealth_v14i1e70994_app4.doc]

References

1. Lee A, Gilbert RM. Epidemiology of parkinson disease. Neurol Clin 2016;34(4):955-965. [doi: 10.1016/j.ncl.2016.06.012] [Medline: 27720003]
2. Bisaglia M, Filograna R, Beltramini M, Bubacco L. Are dopamine derivatives implicated in the pathogenesis of Parkinson's disease? Ageing Res Rev 2014;13:107-114. [doi: 10.1016/j.arr.2013.12.009] [Medline: 24389159]
3. Li G, Ma J, Cui S, He Y, Xiao Q, Liu J, et al. Parkinson's disease in China: a forty-year growing track of bedside work. Transl Neurodegener 2019;8:22 [FREE Full text] [doi: 10.1186/s40035-019-0162-z] [Medline: 31384434]
4. Engelender S, Isacson O. The threshold theory for Parkinson's disease. Trends Neurosci 2017;40(1):4-14. [doi: 10.1016/j.tins.2016.10.008] [Medline: 27894611]
5. Sekirina TP, Voronkova TL, Tsutsu'lkovskaia MI, Abramova LI. Disorders of interleukin 2 biosynthesis by peripheral blood lymphocytes in patients with schizophrenia. Zh Nevropatol Psikhiatr Im S S Korsakova 1989;89(5):95-97. [Medline: 2789461]
6. Przedborski S. The two-century journey of Parkinson disease research. Nat Rev Neurosci 2017;18(4):251-259. [doi: 10.1038/nrn.2017.25] [Medline: 28303016]
7. Macchi ZA, Koljack CE, Miyasaki JM, Katz M, Galifianakis N, Prizer LP, et al. Patient and caregiver characteristics associated with caregiver burden in Parkinson's disease: a palliative care approach. Ann Palliat Med 2020;9(Suppl 1):S24-S33 [FREE Full text] [doi: 10.21037/apm.2019.10.01] [Medline: 31735048]
8. León-Salas B, González-Hernández Y, Infante-Ventura D, de Armas-Castellano A, García-García J, García-Hernández M, et al. Telemedicine for neurological diseases: a systematic review and meta-analysis. Eur J Neurol 2023;30(1):241-254. [doi: 10.1111/ene.15599] [Medline: 36256522]
9. Schootemeijer S, van der Kolk NM, Ellis T, Mirelman A, Nieuwboer A, Nieuwhof F, et al. Barriers and motivators to engage in exercise for persons with Parkinson's disease. J Parkinsons Dis 2020;10(4):1293-1299 [FREE Full text] [doi: 10.3233/JPD-202247] [Medline: 32925106]
10. Shalash A, Spindler M, Cubo E. Global perspective on telemedicine for Parkinson's disease. J Parkinsons Dis 2021;11(s1):S11-S18 [FREE Full text] [doi: 10.3233/JPD-202411] [Medline: 33579872]
11. Bashshur R, Shannon G, Krupinski E, Grigsby J. The taxonomy of telemedicine. Telemed J E Health 2011;17(6):484-494 [FREE Full text] [doi: 10.1089/tmj.2011.0103] [Medline: 21718114]
12. Telemedicine: opportunities and developments in member states: report on the second global survey on EHealth. World Health Organization (Global Observatory for eHealth). 2010. URL: https://apps.who.int/iris/handle/ [accessed 2024-12-26]
13. Lindeman D. Position Paper: mHealth technologies: applications to beneft older adults (Discussion Draft). Center for Technology and Aging. Oakland, CA; 2011. URL: https://www.phi.org/thoughtleadership/position-paper-mhealth-technologies-applications-tobenefit-older-adults-discussion-draft/ [accessed 2024-12-26]
14. Samii A, Ryan-Dykes P, Tsukuda RA, Zink C, Franks R, Nichol WP. Telemedicine for delivery of health care in Parkinson's disease. J Telemed Telecare 2006;12(1):16-18. [doi: 10.1258/135763306775321371] [Medline: 16438773]

XSL•FO
RenderX

15. Beck CA, Beran DB, Biglan KM, Boyd CM, Dorsey ER, Schmidt PN, Connect Parkinson Investigators. National randomized controlled trial of virtual house calls for Parkinson disease. Neurology 2017;89(11):1152-1161 [FREE Full text] [doi: 10.1212/WNL.0000000000004357] [Medline: 28814455]
16. Hanson RE, Truesdell M, Stebbins GT, Weathers AL, Goetz CG. Telemedicine vs office visits in a movement disorders clinic: comparative satisfaction of physicians and patients. Mov Disord Clin Pract 2019;6(1):65-69 [FREE Full text] [doi: 10.1002/mdc3.12703] [Medline: 30746418]
17. Schneider RB, Biglan KM. The promise of telemedicine for chronic neurological disorders: the example of Parkinson's disease. Lancet Neurol 2017;16(7):541-551. [doi: 10.1016/S1474-4422(17)30167-9] [Medline: 28566190]
18. Kumar A. Experience of video consultation during the COVID-19 pandemic in elderly population for Parkinson's disease and movement disorders. Postgrad Med J 2021;97(1144):117-118 [FREE Full text] [doi: 10.1136/postgradmedj-2020-138846] [Medline: 33008959]
19. Sibley KG, Girges C, Hoque E, Foltynie T. Video-based analyses of Parkinson's disease severity: a brief review. J Parkinsons Dis 2021;11(s1):S83-S93 [FREE Full text] [doi: 10.3233/JPD-202402] [Medline: 33682727]
20. Wilkinson JR, Spindler M, Wood SM, Marcus SC, Weintraub D, Morley JF, et al. High patient satisfaction with telehealth in Parkinson disease: a randomized controlled study. Neurol Clin Pract 2016;6(3):241-251 [FREE Full text] [doi: 10.1212/CPJ.0000000000000252] [Medline: 27347441]
21. Peacock D, Baumeister P, Monaghan A, Siever J, Yoneda J, Wile D. Perception of healthcare access and utility of telehealth among Parkinson's disease patients. Can J Neurol Sci 2020;47(5):700-704. [doi: 10.1017/cjn.2020.99] [Medline: 32450924]
22. Feeney MP, Xu Y, Surface M, Shah H, Vanegas-Arroyave N, Chan AK, et al. The impact of COVID-19 and social distancing on people with Parkinson's disease: a survey study. NPJ Parkinsons Dis 2021;7(1):10 [FREE Full text] [doi: 10.1038/s41531-020-00153-8] [Medline: 33479241]
23. Siegert C, Hauptmann B, Jochems N, Schrader A, Deck R. ParkProTrain: an individualized, tablet-based physiotherapy training programme aimed at improving quality of life and participation restrictions in PD patients - a study protocol for a quasi-randomized, longitudinal and sequential multi-method study. BMC Neurol 2019;19(1):143 [FREE Full text] [doi: 10.1186/s12883-019-1355-x] [Medline: 31238908]
24. Dobkin RD, Interian A, Durland JL, Gara MA, Menza MA. Personalized telemedicine for depression in Parkinson's disease: a pilot trial. J Geriatr Psychiatry Neurol 2018;31(4):171-176. [doi: 10.1177/0891988718783274] [Medline: 29945467]
25. Jitkritsadakul O, Rajalingam R, Toenjes C, Munhoz RP, Fasano A. Tele-health for patients with deep brain stimulation: the experience of the Ontario Telemedicine Network. Mov Disord 2018;33(3):491-492. [doi: 10.1002/mds.27230] [Medline: 29119600]
26. Chen YY, Guan BS, Li ZY, Yang QH, Xu TJ, Li HB, et al. Application of telehealth intervention in Parkinson's disease: a systematic review and meta-analysis. J Telemed Telecare 2018;26(1-2):3-13. [doi: 10.1177/1357633x18792805]
27. Federico S, Cacciante L, Cieślik B, Turolla A, Agostini M, Kiper P, RIN_TR_Group. Telerehabilitation for neurological motor impairment: a systematic review and meta-analysis on quality of life, satisfaction, and acceptance in stroke, multiple sclerosis, and Parkinson's disease. J Clin Med 2024;13(1) [FREE Full text] [doi: 10.3390/jcm13010299] [Medline: 38202306]
28. Pastana Ramos LF, Vilacorta-Pereira TDCS, Duarte JDS, Yamada ES, Santos-Lobato BL. Feasibility and effectiveness of a remote individual rehabilitation program for people with Parkinson's disease living in the Brazilian Amazon: a randomized clinical trial. Front Neurol 2023;14:1244661 [FREE Full text] [doi: 10.3389/fneur.2023.1244661] [Medline: 37693755]
29. Kraepelien M, Schibbye R, Månsson K, Sundström C, Riggare S, Andersson G, et al. Individually tailored internet-based cognitive-behavioral therapy for daily functioning in patients with Parkinson's disease: a randomized controlled trial. J Parkinsons Dis 2020;10(2):653-664 [FREE Full text] [doi: 10.3233/JPD-191894] [Medline: 32176657]
30. Dobkin RD, Mann SL, Weintraub D, Rodriguez KM, Miller RB, St Hill L, et al. Innovating Parkinson's care: a randomized controlled trial of telemedicine depression treatment. Mov Disord 2021;36(11):2549-2558. [doi: 10.1002/mds.28548] [Medline: 33710659]
31. Maggio MG, Luca A, Cicero CE, Calabrò RS, Drago F, Zappia M, et al. Effectiveness of telerehabilitation plus virtual reality (Tele-RV) in cognitive e social functioning: a randomized clinical study on Parkinson's disease. Parkinsonism Relat Disord 2024;119:105970. [doi: 10.1016/j.parkreldis.2023.105970] [Medline: 38142630]
32. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009;339:b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]
33. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
34. Cubo E, Mariscal N, Solano B, Becerra V, Armesto D, Calvo S, et al. Prospective study on cost-effectiveness of home-based motor assessment in Parkinson's disease. J Telemed Telecare 2017;23(2):328-338. [doi: 10.1177/1357633X16638971] [Medline: 27000142]
35. Del Pino R, Ortiz de Echevarría A, Díez-Cirarda M, Ustarroz-Aguirre I, Caprino M, Liu J. Virtual coach and telerehabilitation for Parkinson´s disease patients: vCare system. J Public Health 2017;33:1583-1596 [FREE Full text] [doi: 10.1371/journal.pone.0326705] [Medline: 40880394]

XSL•FO
RenderX

36. Dobkin RD, Mann SL, Gara MA, Interian A, Rodriguez KM, Menza M. Telephone-based cognitive behavioral therapy for depression in Parkinson disease: a randomized controlled trial. Neurology 2020;94(16):e1764-e1773 [FREE Full text] [doi: 10.1212/WNL.0000000000009292] [Medline: 32238507]
37. Duffley G, Lutz BJ, Szabo A, Wright A, Hess CW, Ramirez-Zamora A, et al. Home health management of Parkinson disease deep brain stimulation: a randomized clinical trial. JAMA Neurol 2021;78(8):972-981 [FREE Full text] [doi: 10.1001/jamaneurol.2021.1910] [Medline: 34180949]
38. Eldemir S, Guclu-Gunduz A, Eldemir K, Saygili F, Yilmaz R, Akbostancı MC. The effect of task-oriented circuit training-based telerehabilitation on upper extremity motor functions in patients with Parkinson's disease: a randomized controlled trial. Parkinsonism Relat Disord 2023;109:105334. [doi: 10.1016/j.parkreldis.2023.105334] [Medline: 36917914]
39. Ellis TD, Cavanaugh JT, DeAngelis T, Hendron K, Thomas CA, Saint-Hilaire M, et al. Comparative effectiveness of mHealth-supported exercise compared with exercise alone for people with Parkinson disease: randomized controlled pilot study. Phys Ther 2019;99(2):203-216. [doi: 10.1093/ptj/pzy131] [Medline: 30715489]
40. Gandolfi M, Geroin C, Dimitrova E, Boldrini P, Waldner A, Bonadiman S, et al. Virtual reality telerehabilitation for postural instability in Parkinson's disease: a multicenter, single-blind, randomized, controlled trial. Biomed Res Int 2017;2017:7962826 [FREE Full text] [doi: 10.1155/2017/7962826] [Medline: 29333454]
41. Goffredo M, Baglio F, DE Icco R, Proietti S, Maggioni G, Turolla A, RIN_TR_Group. Efficacy of non-immersive virtual reality-based telerehabilitation on postural stability in Parkinson's disease: a multicenter randomized controlled trial. Eur J Phys Rehabil Med 2023;59(6):689-696 [FREE Full text] [doi: 10.23736/S1973-9087.23.07954-6] [Medline: 37847247]
42. Heldman DA, Harris DA, Felong T, Andrzejewski KL, Dorsey ER, Giuffrida JP, et al. Telehealth management of Parkinson's disease using wearable sensors: an exploratory study. Digit Biomark 2017;1(1):43-51 [FREE Full text] [doi: 10.1159/000475801] [Medline: 29725667]
43. Theodoros DG, Hill AJ, Russell TG. Clinical and quality of life outcomes of speech treatment for Parkinson's disease delivered to the home via telerehabilitation: a noninferiority randomized controlled trial. Am J Speech Lang Pathol 2016;25(2):214-232. [doi: 10.1044/2015_AJSLP-15-0005] [Medline: 27145396]
44. Esper CD, Valdovinos BY, Schneider RB. The importance of digital health literacy in an evolving Parkinson's disease care system. J Parkinsons Dis 2024;14(s1):S181-S189 [FREE Full text] [doi: 10.3233/JPD-230229] [Medline: 38250786]
45. Ruotolo I, Sellitto G, Berardi A, Simeon R, Panuccio F, Amadio E, et al. Psychometric properties of the Parkinson's Disease Questionnaire-39 and its short form Parkinson's disease Questionnaire-8: a systematic review and meta-analysis. J Clin Neurosci 2024;123:100-117 [FREE Full text] [doi: 10.1016/j.jocn.2024.03.032] [Medline: 38564966]
46. Tuba BK, Elif DE, Ozgur O, Fatos ES, Emine KAE, Ozden TO. Reliability and validity of the Turkish version of the 39-item Parkinson Disease Questionnaire. Ideggyogy Sz 2023;76(5-6):181-188 [FREE Full text] [doi: 10.18071/isz.76.0181] [Medline: 37294025]
47. Diwakar V, Ertmer PA, Nour AYM. Developing interactive course Web sites for distance education and characteristics of students enrolled in distance learning courses. J Vet Med Educ 2003;30(4):351-357. [doi: 10.3138/jvme.30.4.351] [Medline: 14976621]
48. Li J, Liu Y, Jiang J, Peng X, Hu X. Effect of telehealth interventions on quality of life in cancer survivors: a systematic review and meta-analysis of randomized controlled trials. Int J Nurs Stud 2021;122:103970. [doi: 10.1016/j.ijnurstu.2021.103970] [Medline: 34303269]
49. Xu Y, Feeney MP, Surface M, Novak D, Troche MS, Beck JC, et al. Attitudes toward telehealth services among people living with Parkinson's disease: a survey study. Mov Disord 2022;37(6):1289-1294 [FREE Full text] [doi: 10.1002/mds.28990] [Medline: 35338664]
50. Chaudhuri KR, Azulay J, Odin P, Lindvall S, Domingos J, Alobaidi A, et al. Economic burden of Parkinson's disease: a multinational, real-world, cost-of-illness study. Drugs Real World Outcomes 2024;11(1):1-11 [FREE Full text] [doi: 10.1007/s40801-023-00410-1] [Medline: 38193999]
51. Brock P, Oates LL, Gray WK, Henderson EJ, Mann H, Haunton VJ, et al. Driving and Parkinson's disease: a survey of the patient's perspective. J Parkinsons Dis 2022;12(1):465-471 [FREE Full text] [doi: 10.3233/JPD-212686] [Medline: 34542030]
52. Zhang P, Jin W, Lyu Z, Lyu X, Li L. Study on the mechanism of gut microbiota in the pathogenetic interaction between depression and Parkinson 's disease. Brain Res Bull 2024;215:111001 [FREE Full text] [doi: 10.1016/j.brainresbull.2024.111001] [Medline: 38852651]
53. Tolin DF, Lord KA, Knowles KA. Cognitive-behavioral therapy enhancement strategies. Psychiatr Clin North Am 2024;47(2):355-365. [doi: 10.1016/j.psc.2024.02.005] [Medline: 38724125]
54. Xiang X, Kayser J, Turner S, Ash S, Himle JA. Layperson-supported, web-delivered cognitive behavioral therapy for depression in older adults: randomized controlled trial. J Med Internet Res 2024;26:e53001 [FREE Full text] [doi: 10.2196/53001] [Medline: 38437013]
55. Dou J, Wang J, Gao X, Wang G, Bai Y, Liang Y, et al. Effectiveness of telemedicine interventions on motor and nonmotor outcomes in Parkinson disease: systematic review and network meta-analysis. J Med Internet Res 2025;27:e71169 [FREE Full text] [doi: 10.2196/71169] [Medline: 40460428]

XSL•FO
RenderX

56. Santos DT, Camelo DMF, Strelow MZ, Silva MTS, Führ P, Marins LW, et al. Feasibility of telemedicine for patients with parkinsonism in the Brazilian public health system. Arquivos de neuro-psiquiatria 2022;80(9):914-921 [FREE Full text] [doi: 10.1055/s-0042-1755323] [Medline: 36261128]
57. Lima DP, Gomes VC, Viana Júnior AB, Assis F, Oliveira P, Cunha L, Chaves, et al. Telehealth for Parkinson disease patients during the COVID-19 pandemic: the TeleParkinson study. Arq Neuropsiquiatr 2022;80(10):1026-1035 [FREE Full text] [doi: 10.1055/s-0042-1758751] [Medline: 36535287]
58. Joo JY, Yun JY, Kim YE, Jung YJ, Kim R, Yang H, et al. A survey of perspectives on telemedicine for patients with Parkinson's disease. J Mov Disord 2024;17(1):89-93 [FREE Full text] [doi: 10.14802/jmd.23130] [Medline: 37604653]
59. Perskaudas R, Myers CE, Interian A, Gluck MA, Herzallah MM, Baum A, et al. Reward and punishment learning as predictors of cognitive behavioral therapy response in Parkinson's disease comorbid with clinical depression. J Geriatr Psychiatry Neurol 2024;37(4):282-293. [doi: 10.1177/08919887231218753] [Medline: 38158704]
60. Rohrmann T, Praus P, Proctor T, Benedyk A, Tost H, Hennig O, et al. Patients with affective disorders profit most from telemedical treatment: evidence from a naturalistic patient cohort during the COVID-19 pandemic. Front Psychiatry 2022;13 [FREE Full text] [doi: 10.3389/fpsyt.2022.971896] [Medline: 36532188]
61. Angelopoulou E, Papageorgiou SG. Telemedicine in Alzheimer's disease and other dementias: where we are? J Alzheimers Dis 2025;103(1):3-18. [doi: 10.1177/13872877241298295] [Medline: 39639574]
62. Lau TK, Tse M, Liu Y, Leung AYM. Effectiveness of technological interventions on psychosocial well-being and perception of technological interventions among people with Parkinson's disease: a systematic review. Australas J Ageing 2025;44(2):e70034. [doi: 10.1111/ajag.70034] [Medline: 40317851]
63. Lee D, Erande A, Christodoulou G, Malik S. Addressing mental health symptoms among COVID-19 healthcare workers: a heart rate variability biofeedback pilot study. Stress Health 2024;40(6):e3502 [FREE Full text] [doi: 10.1002/smi.3502] [Medline: 39513426]
64. García-Bustillo Á, Ramírez-Sanz JM, Garrido-Labrador JL, Olivares-Gil A, Valiñas-Sieiro F, Allende-Río M, et al. A multidisciplinary telemedicine approach for managing frailty in Parkinson's disease. A longitudinal, case-control study. Parkinsonism Relat Disord 2024;130:107215 [FREE Full text] [doi: 10.1016/j.parkreldis.2024.107215] [Medline: 39586130]
65. Roper A, Brooks D, Mitchell LK, Pachana NA, Au TR, Byrne GJ, et al. Feasibility and acceptability of a videoconferencing CBT intervention for anxiety in people with Parkinson's disease. Clin Gerontol 2025;48(4):828-843 [FREE Full text] [doi: 10.1080/07317115.2024.2306861] [Medline: 38277135]
66. Wuthrich VM, Rapee RM. Telephone-delivered cognitive behavioural therapy for treating symptoms of anxiety and depression in Parkinson's disease: a pilot trial. Clin Gerontol 2019;42(4):444-453. [doi: 10.1080/07317115.2019.1580811] [Medline: 30821649]
67. Poon S, Tan C, Hong W, Chen KC, Yu R. Tailoring anxiety assessment for Parkinson's disease: the Chinese Parkinson anxiety scale with cultural and situational anxiety considerations. Soc Sci Med 2025;381 [FREE Full text] [doi: 10.1016/j.socscimed.2025.118284] [Medline: 40513505]
68. What is Parkinson?s? 2017. Association EPsD. URL: http://www.epda.eu.com/about-parkinsons/what-is-parkinsons/ [accessed 2026-01-06]
69. Barbour PJ, Arroyo J, High S, Fichera LB, Staska-Pier MM, McMahon MK. Telehealth for patients with Parkinson's disease: delivering efficient and sustainable long-term care. Hosp Pract (1995) 2016;44(2):92-97. [doi: 10.1080/21548331.2016.1166922] [Medline: 26982525]
70. Ekmekyapar Fırat Y, Turgay T, Soğan SS, Günel Karadeniz P. Effects of LSVT-BIG via telerehabilitation on non-motor and motor symptoms and quality of life in Parkinson's disease. Acta Neurol Belg 2023;123(1):207-214 [FREE Full text] [doi: 10.1007/s13760-022-02104-x] [Medline: 36175786]
71. Mustile M, Kourtis D, Ladouce S, Edwards MG, Volpe D, Pilleri M, et al. Investigating the brain mechanisms of externally cued sit-to-stand movement in Parkinson's disease. Mov Disord 2024;39(9):1556-1566. [doi: 10.1002/mds.29889] [Medline: 38984716]
72. D'Souza AF, Jasti DB, Rao RR, Natarajan M. Feasibility of a tele-assisted home exercise program for balance and functional mobility in persons with Parkinson's disease (TELEPORT-PD). Int J Telemed Appl 2025;2025:9936329 [FREE Full text] [doi: 10.1155/ijta/9936329] [Medline: 40519980]
73. Kumar A, Patil S, Singh VK, Pathak A, Chaurasia RN, Mishra VN, et al. Assessment of non-motor symptoms of Parkinson's disease and their impact on the quality of life: an observational study. Ann Indian Acad Neurol 2022;25(5):909-915 [FREE Full text] [doi: 10.4103/aian.aian_647_21] [Medline: 36561034]
74. Diaconu, Irincu L, Ungureanu L, Ciopleia B, în D, Falup-Pecurariu C. Restless legs syndrome in Parkinson's disease. J Pers Med 2023;13(6) [FREE Full text] [doi: 10.3390/jpm13060915] [Medline: 37373904]
75. Ha J, Park JH, Lee JS, Kim HY, Song JO, Yoo J, et al. Effectiveness of live-streaming tele-exercise intervention in patients with Parkinson's disease: a pilot study. J Mov Disord 2024;17(2):189-197 [FREE Full text] [doi: 10.14802/jmd.23251] [Medline: 38419488]
76. Fleisher JE, Hess SP, Klostermann EC, Lee J, Myrick E, Mitchem D, et al. IN-HOME-PD: the effects of longitudinal telehealth-enhanced interdisciplinary home visits on care and quality of life for homebound individuals with Parkinson's

XSL•FO
RenderX

disease. Parkinsonism Relat Disord 2022;102:68-76 [FREE Full text] [doi: 10.1016/j.parkreldis.2022.07.017] [Medline: 35963046]
77. Bode M, Kalbe E, Liepelt-Scarfone I. Cognition and activity of daily living function in people with Parkinson's disease. J Neural Transm (Vienna) 2024;131(10):1159-1186. [doi: 10.1007/s00702-024-02796-w] [Medline: 38976044]
78. Latella D, Maresca G, Formica C, Sorbera C, Bringandì A, Di Lorenzo G, et al. The role of telemedicine in the treatment of cognitive and psychological disorders in Parkinson's disease: an overview. Brain Sci 2023;13(3) [FREE Full text] [doi: 10.3390/brainsci13030499] [Medline: 36979309]
79. Johnson JK, Longhurst JK, Gevertzman M, Jefferson C, Linder SM, Bethoux F, et al. The use of telerehabilitation to improve movement-related outcomes and quality of life for individuals with Parkinson disease: pilot randomized controlled trial. JMIR Form Res 2024;8:e54599 [FREE Full text] [doi: 10.2196/54599] [Medline: 39083792]
80. Putzolu M, Manzini V, Gambaro M, Cosentino C, Bonassi G, Botta A, et al. Home-based exercise training by using a smartphone app in patients with Parkinson's disease: a feasibility study. Front Neurol 2023;14:1205386 [FREE Full text] [doi: 10.3389/fneur.2023.1205386] [Medline: 37448748]
81. Cubo E, Delgado-López PD. Telemedicine in the management of Parkinson's disease: achievements, challenges, and future perspectives. Brain Sci 2022;12(12) [FREE Full text] [doi: 10.3390/brainsci12121735] [Medline: 36552194]
82. Kotsimpou S, Liampas I, Dastamani M, Marogianni C, Stamati P, Tsika A, et al. Evaluation of computer-based cognitive training on mild cognitive impairment in Parkinson's disease (PD-MCI): a review. J Clin Med 2025;14(9) [FREE Full text] [doi: 10.3390/jcm14093001] [Medline: 40364031]
83. Tagliente S, Minafra B, Aresta S, Santacesaria P, Buccoliero A, Palmirotta C, et al. Effectiveness of a home-based computerized cognitive training in Parkinson's disease: a pilot randomized cross-over study. Front Psychol 2024;15 [FREE Full text] [doi: 10.3389/fpsyg.2024.1531688] [Medline: 39850970]
84. Isernia S, Di Tella S, Rossetto F, Borgnis F, Realdon O, Cabinio M, et al. Exploring cognitive reserve's influence: unveiling the dynamics of digital telerehabilitation in Parkinson's disease resilience. NPJ Digit Med 2024;7(1):116 [FREE Full text] [doi: 10.1038/s41746-024-01113-9] [Medline: 38710915]
85. D'Iorio A, Aiello E, Vitale C, Amboni M, Verde F, Silani V, et al. Diagnostics and ecological validity of the Italian version of the Parkinson's disease cognitive rating scale. Dement Geriatr Cogn Disord 2025;54(5):347-351. [doi: 10.1159/000545090] [Medline: 40068648]

Abbreviations

ADL: activities of daily living
BBQ: Brunnsviken Brief Quality of Life Scale
BDI: Beck Depression Inventory
CBT: cognitive behavioral therapy
HADS-A: Hospital Anxiety and Depression Scale-Anxiety
HADS-D: Hospital Anxiety and Depression Scale-Depression
HAM-A: Hamilton Anxiety Rating Scale
HAM-D: Hamilton Depression Rating Scale
HKSJ: Hartung-Knapp-Sidik-Jonkman
MDS-UPDRS: Movement Disorders Society-Unified Parkinson’s Disease Rating Scale
MMSE: Mini-Mental State Examination
MoCA: Montreal Cognitive Assessment
PD: Parkinson disease
PDQ: Parkinson Disease Questionnaire
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QOL: quality of life
RCT: randomized controlled trial
SF-36: Medical Outcomes Study 36-Item Short Form Health Survey
SMD: standard mean difference

XSL•FO
RenderX

Edited by A Stone; submitted 07.Jan.2025; peer-reviewed by F Asci, S Tonapa, K Elechi; comments to author 05.May.2025; revised version received 29.Dec.2025; accepted 30.Dec.2025; published 28.Jan.2026.

Please cite as:
Sun M, Tang F, min L, Wen S, Wang S, Jiang H
Effects of Telehealth Interventions for People With Parkinson Disease: Systematic Review and Meta-Analysis of Randomized Controlled Trials
JMIR Mhealth Uhealth 2026;14:e70994
URL: https://mhealth.jmir.org/2026/1/e70994
doi:10.2196/70994
PMID:

©Minyue Sun, Fuyou Tang, Luo min, Shiyu Wen, Shuang Wang, Huiping Jiang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 28.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Clinical Improvements From Telemedicine Interventions for Managing Type 2 Diabetes Compared With Usual Care: Systematic Review, Meta-Analysis, and Meta-Regression

Shujie Jiang[1], MD; Xianru Gao[2], MD; Haiqing Diao[1], MD; Yang Zhang[1], MD; Guangyu Lu[3], PhD; Xiaoguang Liu[4], PhD; Yuping Li[4], PhD

[1]School of Nursing, Faculty of Medicine, Yangzhou University, Yangzhou, China

[2]Jining Public Health Medical Center, Jining, China

[3]School of Basic Medical Sciences & School of Public Health, Faculty of Medicine, Yangzhou University, Yangzhou, China

[4]Neuro-Intensive Care Unit, Department of Neurosurgery, Northern Jiangsu People's Hospital Affiliated to Yangzhou University, Yangzhou, China

Corresponding Author:
Yuping Li, PhD
Neuro-Intensive Care Unit, Department of Neurosurgery
Northern Jiangsu People's Hospital Affiliated to Yangzhou University
Nantong West Road 98
Yangzhou, 225001
China
Phone: 86 514 873 7563
Fax: 86 514 873 7560
Email: yupingli@yzu.edu.cn

Abstract

Background: Type 2 diabetes mellitus (T2DM) is a prevalent chronic metabolic disorder that poses substantial challenges to global health care systems and patient management. Telemedicine, defined as the use of information and communication technologies to enhance health care delivery, has emerged as a potential tool to improve access to care and facilitate the management of T2DM.

Objective: This systematic review and meta-analysis aimed to evaluate the clinical effectiveness of various telemedicine interventions compared with usual care in glycemic control, and cardiovascular health in adults with T2DM.

Methods: A comprehensive literature search was conducted across databases such as PubMed, Cochrane Library, and Web of Science for randomized controlled trials (RCTs) published up to August 23, 2024. Eligible RCTs compared telemedicine interventions with usual care in adults with T2DM. The primary outcome assessed was hemoglobin A_{1c} (HbA_{1c}) levels, while the secondary outcomes included mean glucose, fasting blood glucose, BMI, weight, systolic blood pressure, diastolic blood pressure, high-density lipoprotein cholesterol, and low-density lipoprotein cholesterol. The quality of the included studies was examined via the Cochrane risk-of-bias tool. Data were extracted and analyzed using a random-effects model, and meta-regression was performed to explore potential moderators. The quality of the evidence was assessed via the Grading of Recommendations, Assessment, Development, and Evaluation approach.

Results: A total of 58 RCTs, encompassing 13,942 participants, were included in the analysis. Our findings showed that telemedicine interventions significantly improved HbA_{1c} levels compared with usual care (mean difference [MD] –0.38, 95% CI –0.49 to –0.27; Z=6.94; P<.001), despite high heterogeneity (I^2=96%). Significant effects were also found for fasting blood glucose (MD –11.29, 95% CI –17.65 to –4.93; Z=3.48; P<.001), weight (MD –1.33, 95% CI –2.23 to –0.44; Z=2.91; P=.004), BMI (MD –0.43, 95% CI –0.72 to –0.13; Z=2.84; P=.004), systolic blood pressure (MD –2.14, 95% CI –3.02 to –1.26; Z=4.76; P<.001), and diastolic blood pressure (MD –1.24, 95% CI –2.02 to –0.46; Z=1.10; P=.002). No significant between-group differences were found in high-density lipoprotein cholesterol and low-density lipoprotein cholesterol improvement. Subgroup analyses revealed that telemedicine delivered by physicians, dietitians, and researchers achieved the most significant reductions in HbA_{1c} levels. Short-term and long-term interventions showed significant HbA_{1c} improvements, while medium-term interventions did not achieve statistical significance. Meta-regression analysis did not identify any statistically significant moderators.

XSL•FO
RenderX

Conclusions: This review highlights telemedicine's superior effectiveness over usual care in improving HbA_{1c} levels in patients with T2DM, regardless of the type of intervention. Telemedicine led by physicians, dietitians, and researchers showed the greatest efficacy in managing blood glucose levels. Furthermore, telemedicine interventions show promise for monitoring weight and cardiovascular health in patients with T2DM.

Trial Registration: PROSPERO CRD42024608130; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=608130

(JMIR Mhealth Uhealth 2026;14:e70429) doi:10.2196/70429

KEYWORDS

telemedicine; type 2 diabetes mellitus; systematic review; meta-analysis; metaregression

Introduction

Diabetes is a chronic metabolic disorder with increasing prevalence worldwide, placing a significant burden on global health care systems [1]. Currently, it affects more than 537 million adults, a number projected to increase to 783.2 million by 2045 [2]. Type 2 diabetes mellitus (T2DM) accounts for 90% of all diabetes cases [3]. T2DM not only poses a significant burden on individuals and society but also leads to reduced life expectancy and impaired quality of life [3,4]. Poor glycemic control in T2DM increases the risk of complications such as retinopathy, nephropathy, neuropathy, and cardiovascular diseases, leading to disability and premature mortality [5,6].

The management of T2DM is challenging and requires personalized, lifelong care [7,8]. Key aspects of T2DM management include glycemic control, weight management, and cardiovascular health monitoring, all of which are critical for preventing and managing the condition [9-11]. Researchers have identified 5 essential components of diabetes care: nutrition, physical activity, glycemic control, medical care, and patient education [12]. As such, continuous management and regular follow-up are imperative [13]. Global digitalization offers innovative digital opportunities for intensive diabetes management [14]. Telemedicine, which has proven effective in managing chronic diseases [15], holds particular promise for improving health care access for underserved populations. First conceptualized in the 1970s, telemedicine refers to the "use of [information and communication technologies] to improve patient outcomes by increasing access to care and medical information" [16]. In diabetes management specifically, telemedicine interventions have shown promise in enhancing glycemic control outcomes [14,17]. However, a recent review indicated that mobile health (mHealth) tools within telemedicine have shown only modest effectiveness [18].

Existing reviews often focus on specific types of telemedicine tools [19-21], settings [22], or providers [23], which limit generalizability across broader diabetes care contexts. This highlights the need for a more comprehensive and inclusive systematic review that synthesizes findings across various contexts [20,22,24]. With the onset of the COVID-19 pandemic, outpatient services for patients with diabetes have been limited, creating an opportunity for health care providers to implement telemedicine for diabetes management. Consequently, a substantial number of recent studies may have been published, warranting an updated review. This study aims to evaluate the impact of various telemedicine interventions on clinical outcomes in patients with T2DM compared with usual care through a systematic review and meta-analysis, providing valuable insights for future clinical practice and research.

Methods

This systematic review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; checklist provided in Multimedia Appendix 1) guidelines and was registered in the PROSPERO database (CRD 42024608130).

Search Strategy

We searched the following databases from inception to August 2024 to identify eligible randomized controlled trials (RCTs): Embase, PubMed, Scopus, CINAHL, and Web of Science. The search strategy was optimized to capture studies in telemedicine, people with T2DM, and glycemic control. Telemedicine is a broad and evolving field encompassing various aspects of remote health care delivery. While our search strategies primarily focused on the MeSH (Medical Subject Headings) term "telemedicine," it is important to acknowledge that this field includes a wide range of related terms, such as "telemonitoring," "telehealth," "mobile health," "mHealth," "eHealth," "teleconsultation," and "telemetry." The full search strategy is shown in the Multimedia Appendix 2 [25-82].

Eligibility Criteria

RCTs that compared telemedicine interventions with usual care in adults (aged 18 years and older) with T2DM were included. The primary outcomes include an assessment of effectiveness based on clinical indicators, such as changes in hemoglobin A_{1c} (HbA_{1c}) levels, mean glucose, fasting blood glucose (FBG), BMI, weight, systolic blood pressure (SBP), diastolic blood pressure (DBP), high-density lipoprotein cholesterol (HDL-c), and low-density lipoprotein cholesterol (LDL-c).

Here, telemedicine is defined as a form of health care delivery using electronic information and telecommunications, telemedicine facilitates information exchange, education, counseling, monitoring, and management between health care professionals and patients [83,84]. The usual care refers to face-to-face care, standard care, or traditional care.

Studies were excluded if they (1) reported findings from pregnant patients, patients with type 1 diabetes mellitus, and patients with prediabetes or other comorbidities; (2) reported populations at high risk only for diabetes or prediabetes; (3) not reported primary outcomes (changes of HbA_{1c} levels); (4) were

XSL•FO
RenderX

published as conference abstracts, case reports, reviews, posters, comments, letters, and research protocols; or (5) were published in languages other than English.

Outcomes

The primary outcome measure was a change in HbA_{1c} levels, while the secondary outcomes measure included changes in mean glucose, FBG, BMI, weight, SBP, DBP, HDL-c, and LDL-c. Telemedicine interventions were categorized into four types based on the mode of delivery [85]: (1) synchronous, involving real-time communication (eg, video or telephone consultations); (2) asynchronous, involving "store-and-forward" technologies such as messaging or email; (3) hybrid, defined as interventions that combined 2 or more telemedicine modalities, for example, in the study by Yang et al [25], patients uploaded daily records and received feedback and reminders through an application (asynchronous), in addition to receiving monthly telephone consultations (synchronous), thus meeting the criteria for hybrid telemedicine; and (4) unspecified, referring to studies that did not clearly report the telemedicine modality used and could not be classified as synchronous, asynchronous, or hybrid telemedicine.

In addition, the telemedicine interventions were categorized into 9 distinct content types: monitoring, counseling, education, reminders, training, feedback, medication management, treatment, and supervision [86]. The duration of interventions was also classified into 3 categories based on their length: short-term interventions (lasting up to 3 months), midterm interventions (lasting between 3 and 6 months), and long-term interventions (lasting 6 months or more) [87].

Study Selection and Data Extraction

Duplicate studies were identified and removed using EndNote. The remaining studies were screened independently by 2 authors (JSJ and GXR) in a sequential manner of title, abstract, and full-text screening. Conflicts were settled by consulting a third author (LGY). Our data extraction included publication details (title, author, and year), study characteristics (country, purpose, blinding and randomization method, and year of publication), participant demographics, intervention details, comparison details, and results (primary and secondary outcomes with their SDs, SEs of the mean, and 95% CIs).

Quality Assessment

The quality of the included RCTs was evaluated using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach [88]. This method is widely used to assess the certainty of evidence and the strength of recommendations. It provides a structured framework for judging the quality of evidence in systematic reviews. The GRADE system categorizes the certainty of evidence into 4 levels: high, moderate, low, and very low (Table S1 in Section 4 in Multimedia Appendix 2). Using the Cochrane risk-of-bias tool for RCTs, we evaluated the risk of bias and classified each trial as having a low, high, or unclear risk of bias for each area. Six bias domains are included in the tool: selection (random sequence generation and allocation concealment), performance, detection, attritions, reporting, and other biases [89]. Each trial's risk of bias was evaluated independently by 2 reviewers (JSJ and XRG). In the event of a disagreement, consensus decision-making was used to achieve the most agreeable decision to all.

Data Synthesis and Analysis

All outcomes were reported as means and SDs. If SDs were not reported and could not be obtained from study authors, they were estimated using available information such as SEs, 95% CIs, or *P* values, in accordance with the Cochrane Handbook. HbA_{1c} values were presented as percentages in accordance with the National Glycohemoglobin Standardization Program [90]; therefore, the values from studies that reported HbA_{1c} in millimoles per mole were converted to HbA_{1c} %. Likewise, mean glucose, FBG, HDL-c, and LDL-c reported in millimoles per liter were converted to milligrams per deciliter.

All statistical analyses were performed in Review Manager (version 5.4.1; Cochrane Collaboration) and Stata 16 (Stata Statistical Software: Release 16, StataCorp 2019; StataCorp LLC). Reported medians, IQRs, ranges, and CIs were transformed to means and SDs by traditional methods [91,92]. An overall treatment effect was estimated with a meta-analysis of the pool of included studies based on the mean difference (MD). Heterogeneity was assessed statistically using I^2 tests. The results were combined with a random-effects model (due to heterogeneity, ie, an I^2 statistic >50%). Subgroup analyses were conducted based on the characteristics of the intervention, including the type of telemedicine, the telemedicine provider, and the duration of the intervention. Univariate a priori subgroup analyses based on meta-regression of the telemedicine characteristics were conducted in Stata and combined with post hoc analyses of the association of study and patient characteristics with the treatment effect of telemedicine. Publication bias was evaluated using visual inspection of the funnel plot and the Egger test, both performed in Stata.

Ethical Considerations

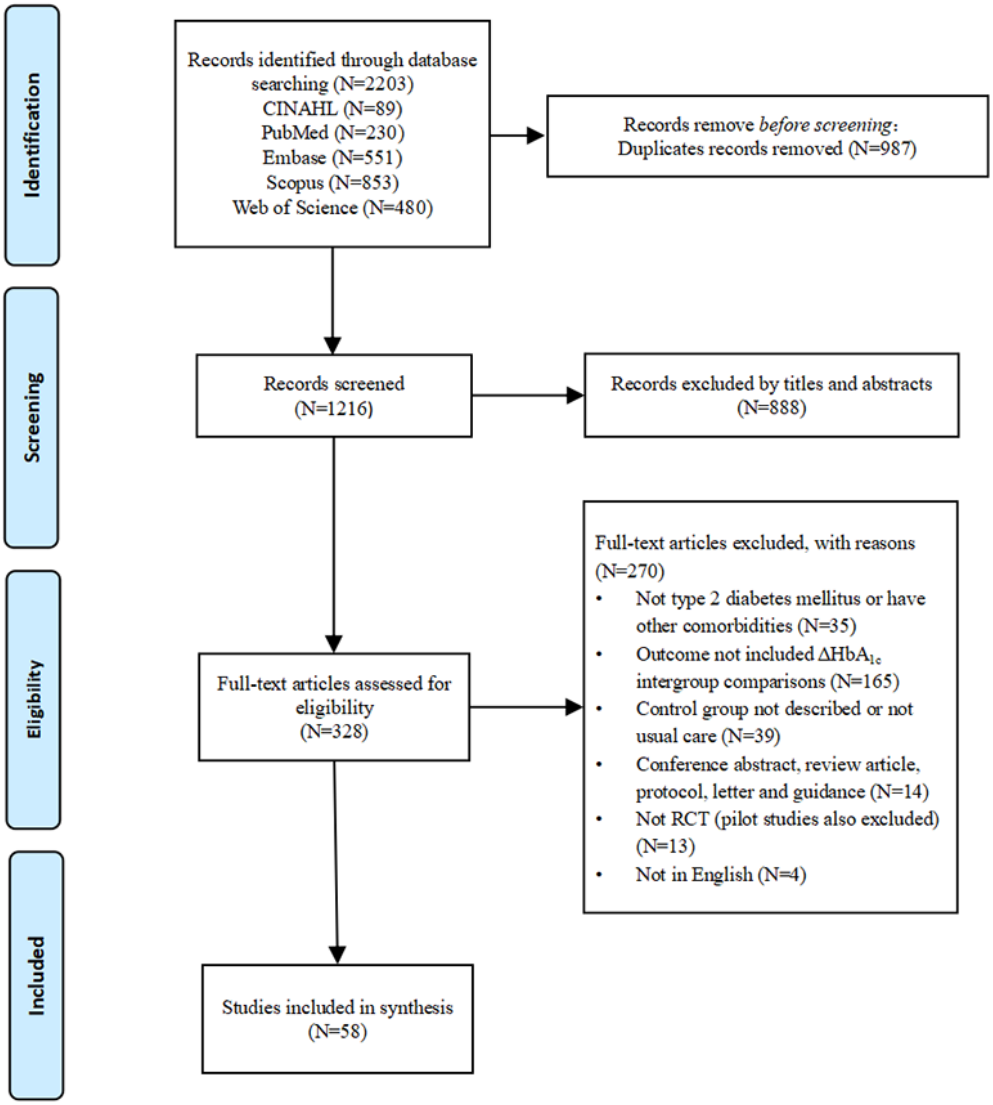
This study is a systematic review and meta-analysis based on previously published RCTs. No new human participants were involved, and no new data were collected. Therefore, ethics approval and informed consent were not required. All data used in this review were extracted from publicly available articles, and no identifiable personal information was involved.

Results

A total of 2203 studies were retrieved from 5 databases. After removing 987 duplicates, 1216 studies remained. Titles and abstracts were screened to determine eligibility, resulting in the exclusion of 888 studies. The full text of the remaining 328 studies was reviewed (Section 3 in Multimedia Appendix 2), and 58 studies were ultimately deemed eligible for analysis (Figure 1).

XSL•FO
RenderX

Figure 1. Flowchart summarizing the process of study selection. HbA_{1c}: glycated hemoglobin; RCT: randomized controlled trial.



Basic Characteristics of the Included Studies

A total of 58 RCTs, encompassing 13,942 participants, were included in the analysis. The detailed basic information, sample characteristics, type of telemedicine used, and clinical outcomes of the included studies are summarized in Table S2 in Section 4 in Multimedia Appendix 2.

The number of published studies has shown a steady increase over recent decades, with the most studies in the last 5 years accounting for 41% (24/58). Geographically, the studies were primarily conducted in Asia (18/58, 31%), Europe (18/58, 31%), and North America (17/58, 29%). Additional studies were conducted in Africa (1/58, 2%), South America (1/58, 2%), and Oceania (2/58, 3%; Figure S1 in Section 5 in Multimedia Appendix 2). Notably, 72% (42/58) of the studies originated from the high-income countries (Table 1). The most common settings for these studies were hospitals (20/58, 34%) and primary health care facilities (20/58, 34%).

XSL•FO
RenderX

Table 1. Characteristics of the randomized controlled trial studies (N=58).

Characteristics	Values
Year of publication, n (%)	
2005-2009	3 (5)
2010-2014	11 (19)
2015-2019	20 (34)
2020-2024	24 (41)
Study location, n (%)	
Africa	2 (3)
Asia	18 (31)
Europe	18 (31)
North America	17 (29)
Oceania	2 (3)
South America	1 (2)
Whether the study site is in a high-income country, n (%)	
Yes	42 (72)
No	16 (28)
Study setting, n (%)	
Hospital	20 (34)
Primary care	20 (34)
Not reported	6 (10)
Others	12 (21)
Total number of participants in included studies, n	13,942
Mean age of participants, median (range)	
Intervention group	55.99 (33.00-73.05)
Control group	56.70 (32.40-73.04)
Proportion of male participants in percentage, median (range)	
Intervention group	51.7 (22.73-80.2)
Control group	53.00 (0-81.30)
Trial length in months, median (range)	6 (3-24)

The median age of participants in the intervention group was 55.99 (range 33.00-73.05) years, while the control group's median age was 56.70 (range 32.40-73.04) years. Male participants constituted a median of 51.7% in the intervention group and 53.0% in the control group (Table 2).

XSL•FO
RenderX

Table 2. Summary of descriptive characteristics of included papers.

Author (year)	Country	Setting	Sample size, I[a]/C[b] (n)	Type of telemedicine	Intervention duration
Anzaldo-Campos et al (2016) [30]	Mexico	Family medical unit	102/100	Remote monitoring	10 months
Arora et al (2014) [31]	United States	Hospital	64/64	Asynchronous	6 months
Azelton et al (2021) [32]	United States	Family medical unit	16/14	Remote monitoring	12 weeks
Basudev et al (2016) [33]	United Kingdom	NR[c]	93/115	NR	12 months
Bentley et al (2016) [34]	United Kingdom	NR	9/9	Remote monitoring	12 weeks
Capozza et al (2015) [72]	United States	Primary care	58/35	Asynchronous	6 months
Cho et al (2017) [35]	South Korea	Hospital	244/240	Remote monitoring	24 weeks
Christensen et al (2022) [36]	Denmark	Hospital	81/55	Asynchronous	24 months
Christensen et al (2022) [37]	Denmark	Primary care	100/70	Synchronous and asynchronous	6 months
Dale et al (2009) [38]	United Kingdom	Primary care	44/97	Synchronous	6 months
Dario et al (2017) [39]	Italy	Local health authority	208/91	Remote monitoring	12 months
Dunkel et al (2023) [40]	Germany	NR	86/65	Synchronous and remote monitoring	12 months
Eakin et al (2014) [41]	Australia	Primary care	151/151	Synchronous	18 months
Farmer et al (2021) [42]	Southern Africa	Primary care	558/561	Asynchronous	12 months
Franc et al (2019) [73]	France	Hospital	C-G1: 62; I-G2: 64; I-G3: 63	Synchronous	13 months
Gerber et al (2023) [74]	United States	Primary care	109/112	Synchronous	12 months
Gong et al (2020) [43]	Australia	NR	93/94	Synchronous and remote monitoring	12 months
Greenwood et al (2015) [76]	United States	Primary care	45/45	Remote monitoring	6 months
Haghighinejad et al (2022) [26]	Iran	Hospital	I-G1: 50; I-G2: 50; C-G3: 46	Asynchronous	3 months
Hee-Sung (2007) [44]	South Korea	Hospital	12/15	Synchronous	12 weeks
Hoda et al (2023) [45]	India	Hospital	50/50	Synchronous and asynchronous	3 months
Holmen et al (2014) [27]	Norway	Hospital	I-G1: 50; I-G2: 50; C-G3: 50	Remote monitoring; synchronous and remote monitoring	12 months
Hsu et al (2016) [46]	United States	Hospital	20/20	Synchronous and asynchronous	12 ±2 weeks
Jantraporn et al (2019) [71]	Thailand	Primary care	26/27	Synchronous	12 weeks
Jarab et al (2012) [47]	Jordan	Hospital	85/86	Synchronous	6 months
Jeong et al (2018) [28]	South Korea	Hospital	I-G1: 113; I-G2: 112; C-G3: 113	Remote monitoring; synchronous and remote monitoring	24 weeks
Kempf et al (2017) [48]	Germany	Institute	93/74	Synchronous and remote monitoring	12 weeks
Kempf et al (2023) [49]	Germany	Institute	192/275	Synchronous and asynchronous	12 months
Khanna et al (2014) [77]	United States	Primary care	38/37	Synchronous	12 weeks
Kitazawa et al (2024) [50]	Japan	Institute	86/82	Remote monitoring	12 weeks
Kleinman et al (2016) [78]	India	NR	45/45	NR	6 months

XSL•FO
RenderX

Author (year)	Country	Setting	Sample size, I[a]/C[b] (n)	Type of telemedicine	Intervention duration
Klingeman et al (2017) [79]	United States	Primary care	30/30	Synchronous and asynchronous	1 year
Kooiman et al (2018) [51]	United States	Health care organizations	40/32	Remote monitoring	12 weeks
Lauffenburger et al (2019) [52]	United States	Health insurer	700/700	Synchronous	12 months
Lee et al (2022) [29]	South Korea	Hospital	I-G1: 91; I-G2: 91; C-G3: 87	Remote monitoring; synchronous and remote monitoring	26 weeks
Lee et al (2020) [53]	Malaysia	Primary care	120/120	Remote monitoring	6 months
Leong et al (2022) [54]	China	Hospital	91/90	Synchronous and asynchronous	3 months
Lim et al (2021) [55]	Singapore	Primary care	99/105	Asynchronous and remote monitoring	6 months
Liou et al (2014) [56]	China	Primary care	54/41	Synchronous	6 months
Lorig et al (2008) [57]	Spain	Primary care	219/198	Synchronous	6 months
Luley et al (2011) [58]	Germany	Hospital	35/35	Remote monitoring	6 months
María Gómez et al (2022) [75]	Colombia	Hospital	41/45	Synchronous and asynchronous	3 months
Mitchell et al (2023) [59]	United States	Primary care	158/151	Synchronous	6 months
Orsama et al (2013) [60]	Finland	NR	24/24	Remote monitoring	10 months
Oseran et al (2022) [61]	United States	Primary care	130/130	Synchronous	6 months
Parsons et al (2019) [62]	United Kingdom	Hospital	148/151	Synchronous	12 months
Quinn et al (2011) [63]	United States	Primary care	62/56	Remote monitoring	12 months
Sachmechi et al (2023) [64]	United States	Hospital	39/39	Synchronous	12 weeks
Sarayani et al (2018) [65]	Iran	Hospital	50/50	Synchronous	3 months
Sun et al (2019) [66]	China	Hospital	44/47	Synchronous and remote monitoring	3 months
Tang et al (2013) [80]	United States	Institute	202/213	Asynchronous and remote monitoring	12 months
Torbjørnsen et al (2014) [67]	Norway	Study center	50/50	Synchronous and asynchronous	4 months
Turnin et al (2021) [81]	France	Hospital	128/135	Remote monitoring	1 year
Vaughan et al (2021) [82]	United States	Clinic	44/45	Synchronous	6 months
Yin et al (2022) [68]	China	Clinic	52/47	Remote monitoring	6 months
Zhang et al (2024) [69]	China	Primary care	1038/1034	Remote monitoring	24 months
Yang et al (2022) [25]	China	Primary care	50/50	Synchronous and asynchronous	12 months
Yang et al (2020) [70]	South Korea	Primary care	150/97	Asynchronous	3 months

[a]I: intervention group.

[b]C: control group.

[c]NR: not reported.

[d]G: group.

Description of Telemedicine Interventions

Among the included studies, synchronous telemedicine, remote monitoring, and hybrid telemedicine were equally prevalent, with each accounting for 29% (18/63) of studies. Asynchronous telemedicine was used in 11% (7/63), and 3% (2/63) of studies did not specify the type of telemedicine used (Figure S3A in Section 5 in Multimedia Appendix 2). The most frequently implemented intervention components were monitoring (29 studies), counseling (29 studies), and education (24 studies; Figure S3B in Section 5 in Multimedia Appendix 2). Tools commonly used in telemedicine included telephone calls (42%),

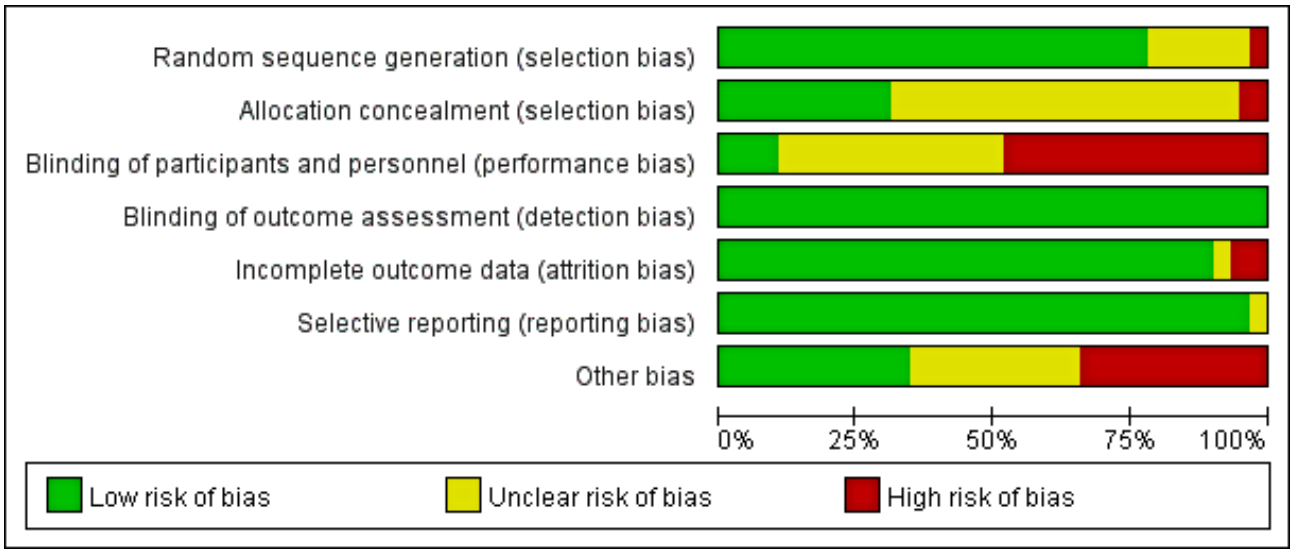
XSL•FO
RenderX

remote monitoring devices (39%), apps (36%), and text messages (32%; Figure S3C in Section 5 in Multimedia Appendix 2).

Risk of Bias in the Included Studies

The risk-of-bias assessment for the 58 included studies is summarized in Figure 2. All studies reported random sequence generation, with 47 (81%) studies assessed as low risk and 11 (19%) as unclear risk. For allocation concealment, 18 (31%) studies were judged as low risk, while 37 (64%) were unclear and 3 (5%) were rated as high risk. Blinding of participants and personnel was the most frequently identified source of bias. Only 6 (10%) studies were considered low risk, whereas 22 (38%) were unclear and 30 (51%) were at high risk. In contrast, all studies were rated as low risk for blinding of outcome assessment, as most primary outcomes were objective laboratory measures such as HbA_{1c}. With respect to incomplete outcome data, 52 (90%) studies had a low risk of bias, 2 (3%) were unclear, and 4 (7%) were considered high risk due to missing data without adequate explanation. Selective reporting bias was generally low across the studies, with 56 (97%) studies judged to be low risk and 2 (3%) as unclear. For other sources of bias, 20 (34%) studies were considered low risk, 18 (31%) were unclear, and 20 (34%) were rated as high risk. Common issues included potential conflicts of interest, such as sponsorship from device companies or external funding sources. Overall, the methodological quality of the included studies was considered moderate. Details of the evaluation process are shown in Table S4 in Section 4 in Multimedia Appendix 2.

Figure 2. RoB assessment: summary plot. RoB: risk of bias.



Effects of Telemedicine on the Clinical Indicators of Patients With T2DM

Among the 58 included studies, 4 [26-29] used 2 types of telemedicine interventions, leading to a total of 62 intervention groups. Detailed intergroup comparisons for primary and secondary outcomes are shown in Table 3. Of the 58 studies, 47 provided sufficient quantitative data (eg, MDs with SDs, CIs, or SEs) to be included in the meta-analysis of HbA_{1c} outcomes. The remaining 11 studies, although included in the systematic review, did not report complete statistical information and were therefore not included in the quantitative synthesis for HbA_{1c}. Overall, all trials demonstrated a positive impact of telemedicine on glycemic control (including changes in HbA_{1c}, mean glucose, and FBG), weight management (change in weight and BMI), and cardiovascular health monitoring (including changes in SBP, DBP, LDL-c, and HDL-c).

XSL•FO
RenderX

Table 3. Effect of telemedicine on primary outcome (change of HbA_{1c}[a]) and secondary outcome.

Clinical indicators outcome	HbA_{1c} (N=62)	Mean glucose (n=2)	FBG[b] (N=15)	Weight (N=20)	BMI (N=20)	SBP[c] (N=23)	DBP[d] (N=22)	HDL-c[e] (N=12)	LDL-c[f] (N=14)
☆[g]	23	—[h]	6	6	7	4	2	—	1
+[i]	8	1	4	2	2	4	2	2	1
=[j]	31	1	5	11	11	14	18	9	12
×[k]	—	—	—	1	—	1	—	1	—

aHbA_{1c}: hemoglobin A_{1c}.

[b]FBG: fasting blood glucose.

[c]SBP: systolic blood pressure.

[d]DBP: diastolic blood pressure.

[e]HDL-c: high-density lipoprotein cholesterol.

[f]LDL-c: low-density lipoprotein cholesterol.

[g]Significant difference results in the intervention group compared with the control group; $P<.01$.

[h]Not available.

[i]Statistically significant results in the intervention group compared with the control group; $P<.05$.

[j]No significant difference results in the intervention group compared with the control group; $P<.05$.

[k]Statistically significant results in the control group compared with the intervention group.

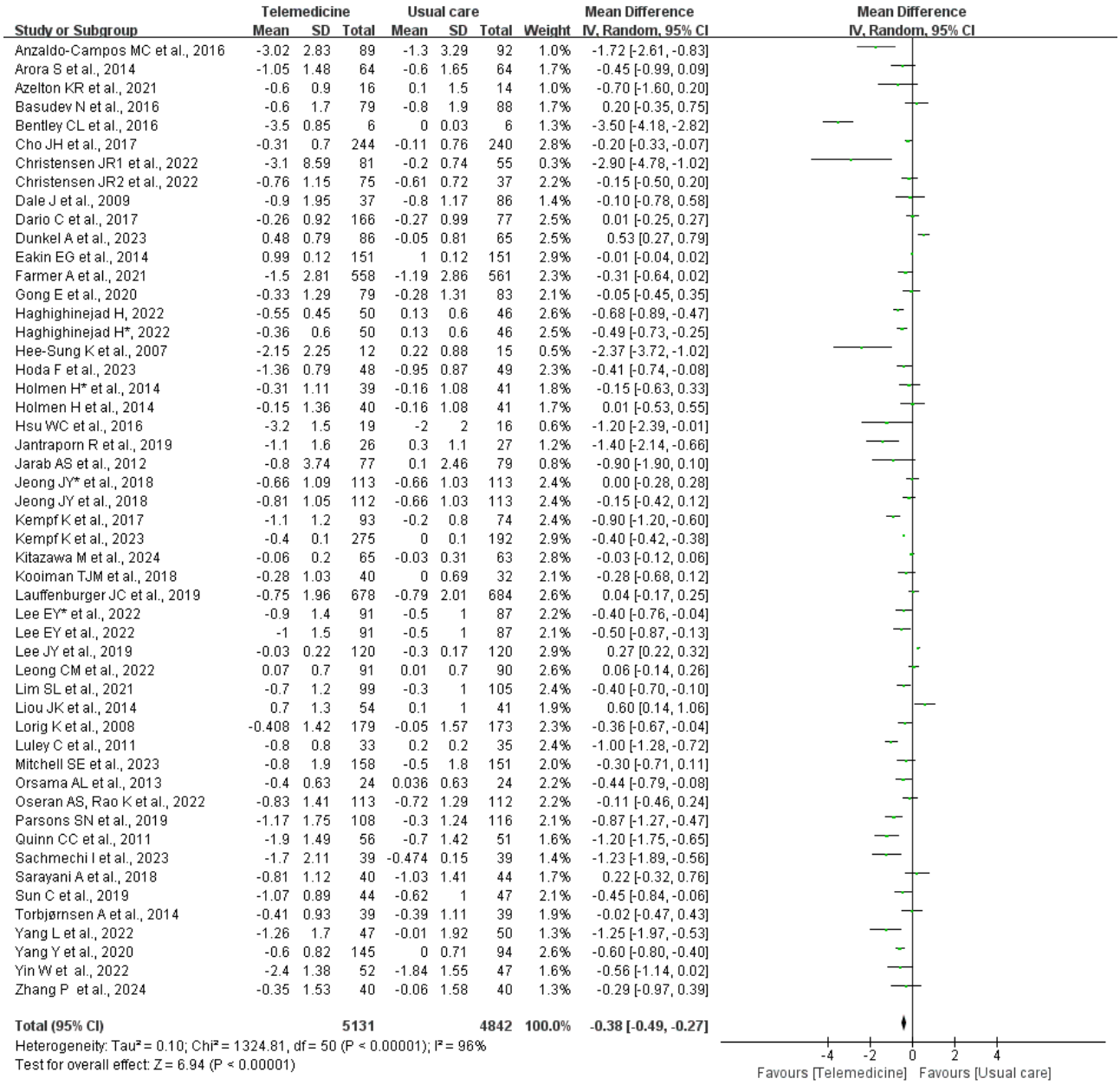
Meta-Analysis of the Effects of Telemedicine on Primary Outcome ($\triangle HbA_{1c}$)

Overall Meta-Analysis of $\triangle HbA_{1c}$ Between Intervention and Control Groups

A total of 47 studies [25-71] reported data on the change in HbA_{1c} levels between groups following telemedicine interventions, with 4 [26-29] studies including 2 types of telemedicine. The analysis revealed a high level of heterogeneity (I^2=96%). Consequently, a random-effects model was used. The pooled results showed a significant reduction in HbA_{1c} levels in the telemedicine group compared with the control group (MD –0.38, 95% CI –0.49 to –0.27; Z=6.94; $P<.001$; Figure 3 [25-71]).

XSL•FO
RenderX

Figure 3. Forest plot of the overall meta-analysis of the change in HbA_{1c} between the telemedicine intervention group and the control group (47 studies [28,36-81]). *A study used 2 different telemedicine tools for the intervention, which are used here to differentiate. HbA_{1c}: glycated hemoglobin.

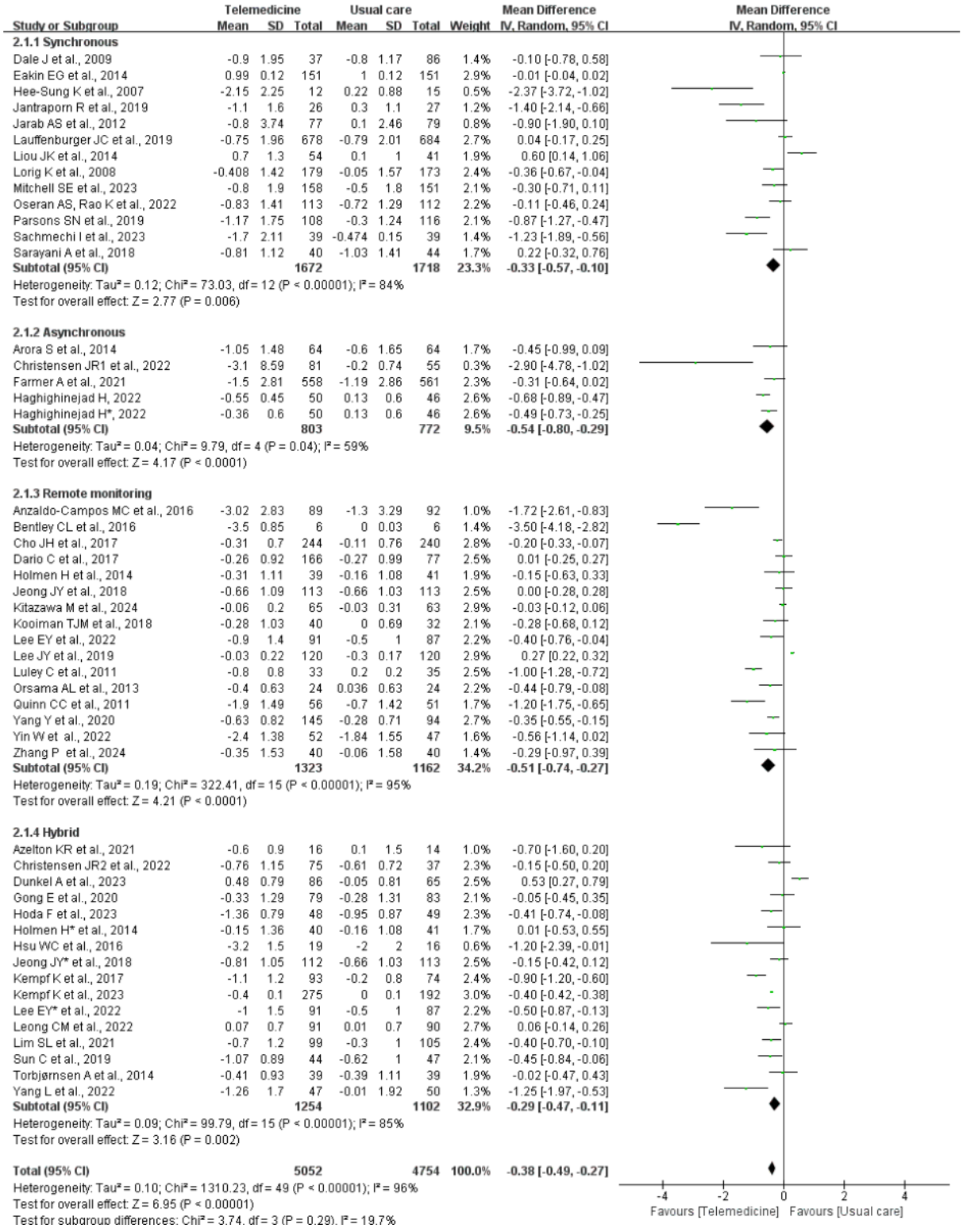


Subgroup Meta-Analysis of $\triangle HbA_{1c}$ by Type of Telemedicine, Duration of Telemedicine Intervention, and Telemedicine Provider

For subgroups based on telemedicine type—synchronous, asynchronous, remote monitoring, and hybrid—telemedicine interventions consistently showed a greater reduction in HbA_{1c} compared with usual care. However, substantial heterogeneity was observed in all subgroups (Figure 4 [25-32,34-71]). The test for subgroup differences was not statistically significant ($^2_3$=3.74, *P*=.29; I^2=19.7%).

XSL•FO
RenderX

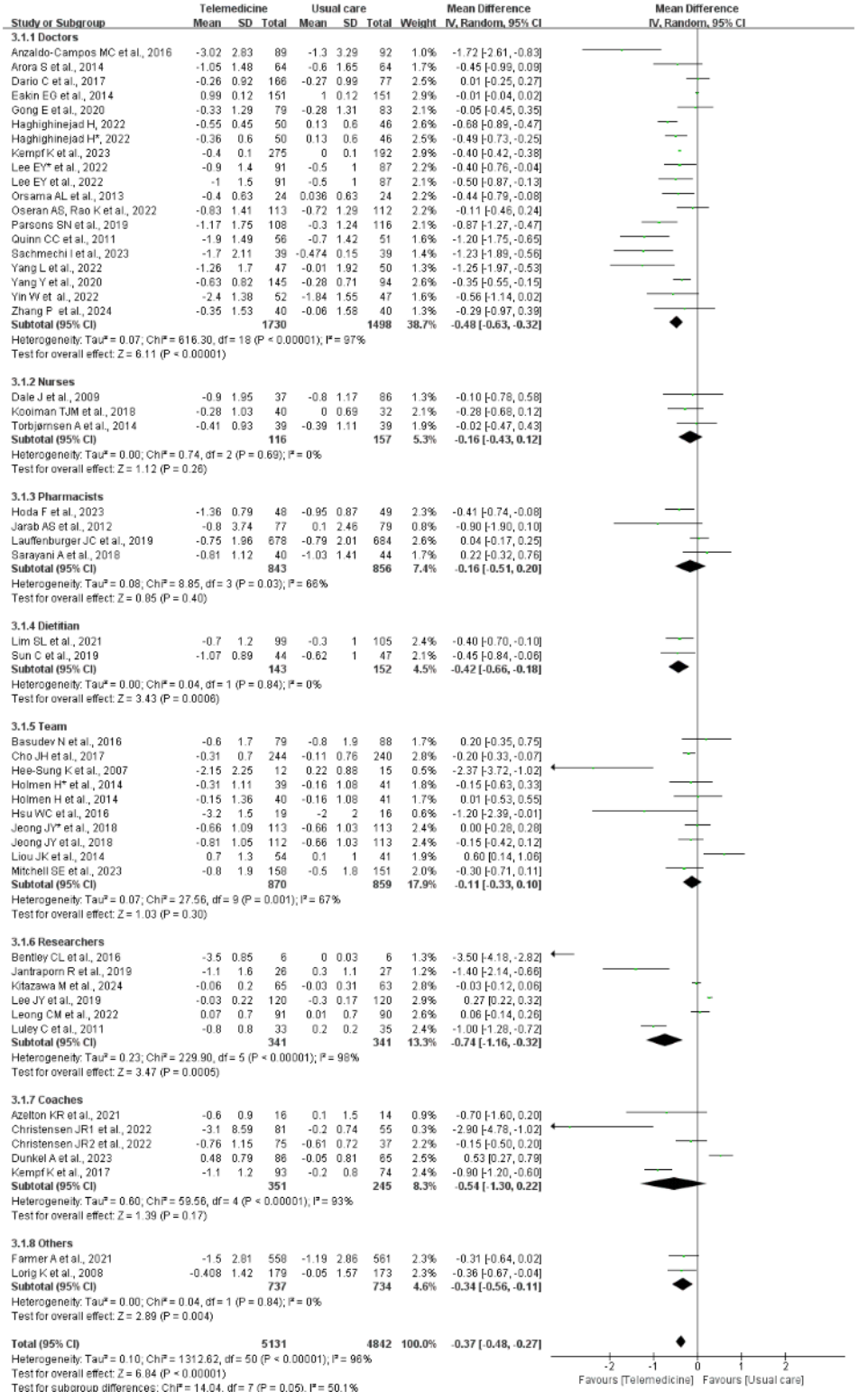
Figure 4. Forest plot of the subgroup meta-analysis of the change in HbA_{1c} by type of telemedicine intervention (46 studies [28,36-42,44-81]). *A study used 2 different telemedicine tools for the intervention, which are used here to differentiate. HbA_{1c}: glycated hemoglobin.



Eight subgroups were identified based on telemedicine providers (Figure 5 [25-71]). Telemedicine delivered by physicians, dietitians, and researchers showed a significantly greater reduction in HbA_{1c} levels compared with usual care but with very high heterogeneity in the physician and researcher subgroups. However, telemedicine delivered by nurses, pharmacists, medical teams, and coaches did not show statistically significant differences between the intervention and control groups. The between-subgroup difference approached statistical significance ($^2_7$=14.04, P=.05; I^2=50.1%).

XSL•FO
RenderX

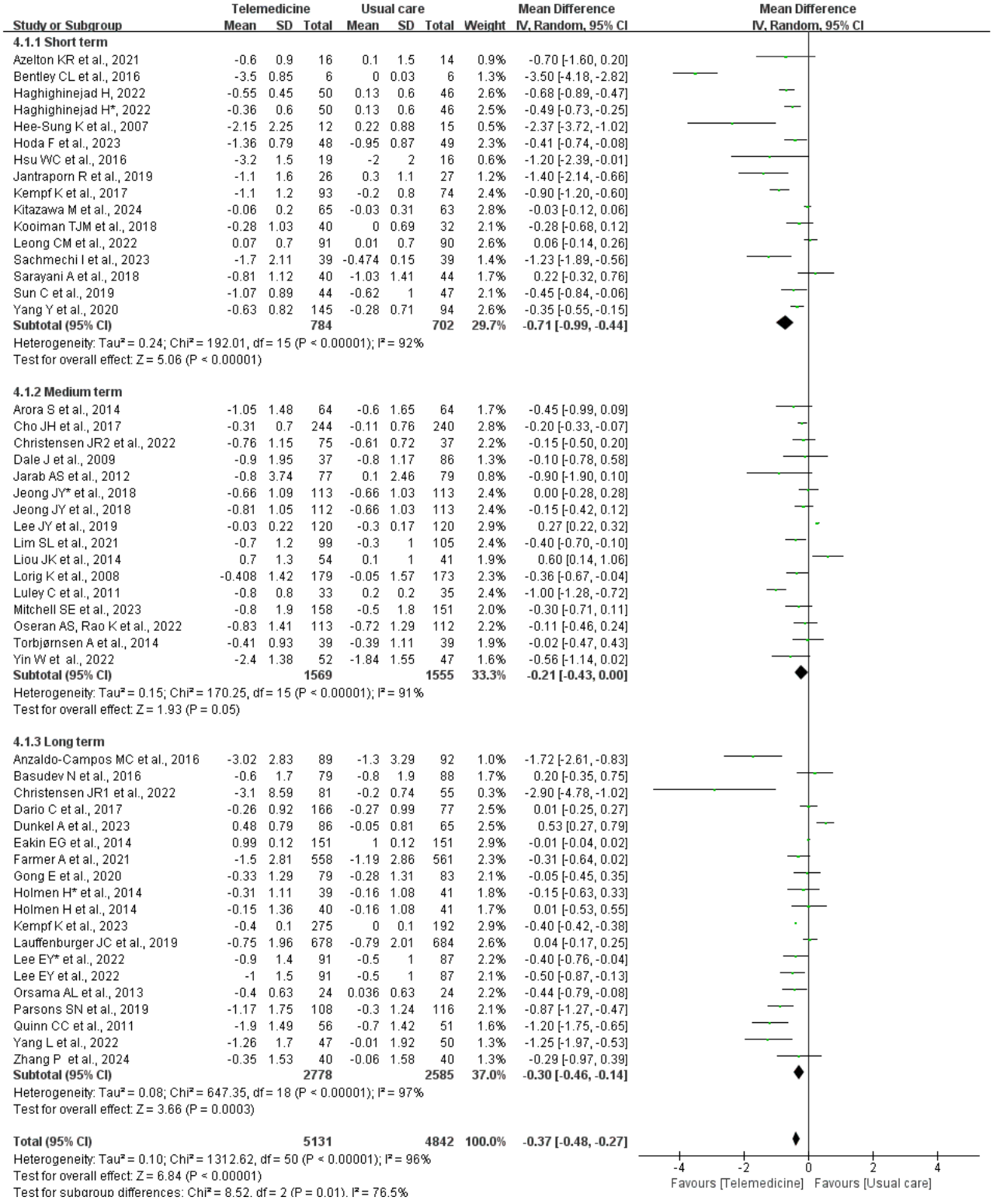
Figure 5. Forest plot of the subgroup meta-analysis of the change in HbA_{1c} by telemedicine provider (47 studies [28,36-81]). *A study used 2 different telemedicine tools for the intervention, which are used here to differentiate. HbA_{1c}: glycated hemoglobin.



Subgroup analysis by intervention duration revealed significant reductions in HbA_{1c} levels for both short-term and long-term interventions. In contrast, medium-term interventions showed no statistically significant difference in HbA_{1c} changes between telemedicine and usual care (Figure 6 [25-71]). The test for subgroup differences indicated significant variation across durations ($^2_2$=8.52, *P*=.01; I^2=76.5%).

XSL•FO
RenderX

Figure 6. Forest plot of the subgroup meta-analysis of the change in HbA_{1c} by duration of telemedicine intervention (47 studies [28,36-81]). *A study used 2 different telemedicine tools for the intervention, which are used here to differentiate. HbA_{1c}: glycated hemoglobin.



Meta-Analysis of the Effects of Telemedicine in Secondary Outcomes

Meta-Analysis of △FBG Between Intervention and Control Groups

A total of 13 studies [25,26,28,29,35,44,47,48,53,55,68-70] provided data on changes in FBG between groups, comprising 16 group comparisons (Figure S4A in Section 5 in Multimedia Appendix 2). A high level of heterogeneity was observed (I^2=99%), prompting the use of a random-effects model. The analysis revealed a significant reduction in FBG levels in the intervention group (MD –11.29, 95% CI –17.65 to –4.93; Z=3.48; P<.001).

XSL•FO
RenderX

Meta-Analysis of △Weight Between Intervention and Control Groups

Nineteen studies [25,27-29,34-37,41,48-51,55,58,60,62,70] reported data related to weight change between groups (Figure S4B in Section 5 in Multimedia Appendix 2). The results showed high heterogeneity (I^2=97%), and a random-effects model was applied. The pooled analysis indicated a significant reduction in weight in the intervention group (MD –1.33, 95% CI –2.23 to –0.44; Z=2.91; P=.004).

Meta-Analysis of △BMI Between Intervention and Control Groups

The BMI data (Figure S4C in Section 5 in Multimedia Appendix 2) were derived from 20 group comparisons across included studies [25,28-30,33,37,40,42,47-51,55,56,58,62,68,70], indicating high heterogeneity (I^2=93%). Using a random-effects model, the intervention group exhibited a significant reduction in BMI (MD –0.43, 95% CI –0.72 to –0.13; Z=2.84; P=.004).

Meta-Analysis of △SBP Between Intervention and Control Groups

Data from 20 group comparisons on SBP changes [7,25,28,30,32,33,35,37,41,42,49,50,53,55,56,60,63,68-70] showed moderate heterogeneity (I^2=64%; Figure S4D in Section 5 in Multimedia Appendix 2). Analyses showed more significant changes in SBP levels in the intervention group (MD –2.14, 95% CI –3.02 to –1.26; Z=4.76; P<.001).

Meta-Analysis of △DBP Between Intervention and Control Groups

Nineteen studies [25, 28, 30, 32, 33, 35, 37, 41, 47, 49, 50, 53, 55, 56, 60, 63, 68-70] provided data on DBP changes (Figure S4E in Section 5 in Multimedia Appendix 2). The results showed a moderate level of heterogeneity with an I^2 value of 50%. Analyses showed more significant changes in DBP levels in the intervention group (MD –1.24, 95% CI –2.02 to –0.46; Z=1.10; P=.002).

Meta-Analysis of △LDL-c Between Intervention and Control Groups

A total of 8 studies [28-30,47,55,68-70] reported data related to LDL-c changes between groups (Figure S4F in Section 5 in Multimedia Appendix 2). The results showed a very high degree of heterogeneity (I^2=100%). Analyses showed more significant changes in LDL-c levels in the intervention group (MD –0.69, 95% CI –11.69 to 10.31; Z=0.12; P=.90).

Meta-Analysis of △HDL-c Between Intervention and Control Groups

A total of 7 studies [28,30,47,55,68-70] reported data related to HDL-c changes between groups (Figure S4G in Section 5 in Multimedia Appendix 2). The results showed very high heterogeneity with an I^2 value of 99%. The analysis showed more significant changes in HDL-c levels in the control group (MD –3.41, 95% CI –2.67 to 9.49; Z=1.10; P=.27).

Meta-Rregression Analysis of Telemedicine

In this meta-regression analysis, none of the variables, including study location, whether the study was conducted in a high-income country, study setting, type of telemedicine, provider, and duration of intervention, significantly influenced the effect of telemedicine on HbA_{1c} levels (P>.05). Although the provider variable approached statistical significance (P=.084), no clear associations were found overall (Table 4).

Table 4. Association between study covariates and effect of telemedicine on HbA_{1c}[a]% (meta-regression).

Variable	Coefficient	SE	Two-tailed *t* test (*df*)	*P* value > |t|	95% CI
Study location (continent)	0.031	0.141	0.220 (44)	.826	–0.254 to 0.317
Whether the study site is in a high-income country	0.112	0.275	0.410 (44)	.687	–0.442 to 0.666
Study setting	–0.072	0.101	–0.710 (44)	.480	–0.274 to 0.131
Type of telemedicine	–0.124	0.116	–1.070 (44)	.291	–0.357 to 0.110
Provider of telemedicine	0.084	0.048	1.770 (44)	.084	–0.012 to 0.180
Duration of telemedicine	0.142	0.134	1.070 (44)	.293	–0.127 to 0.412
Intercept	–0.815	0.836	–0.970 (44)	.335	–2.500 to 0.871

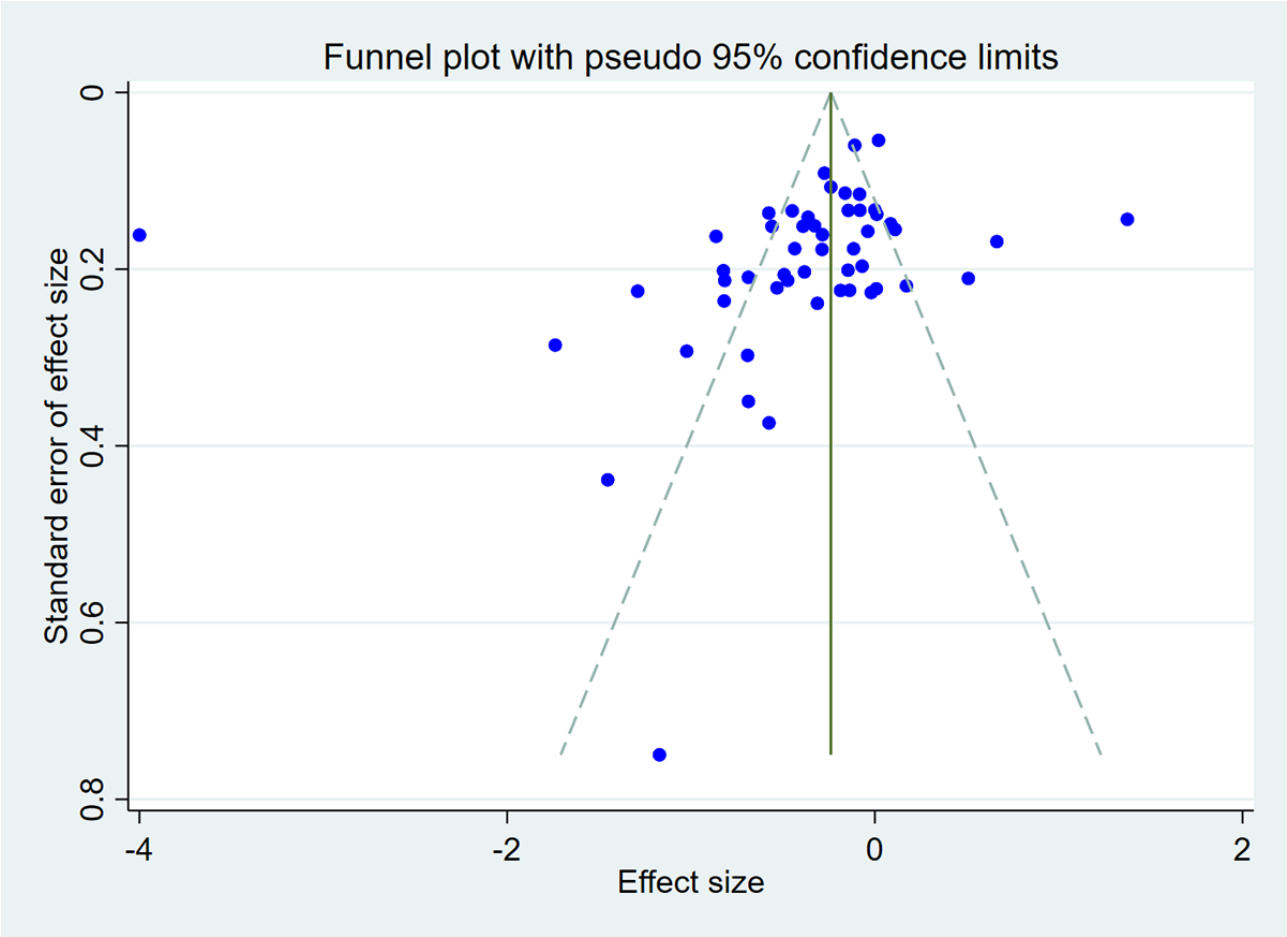
[a]HbA_{1c}: hemoglobin A_{1c}.

Publication Bias

Publication bias was assessed through both visual inspection of the funnel plot and formal statistical methods (Figure 7). Egger regression analysis showed borderline significance (t=–1.98; P=.054), suggesting potential asymmetry. To further assess this, we performed a trim-and-fill analysis using a random-effects model. When imputing missing studies on the left side, the method suggested 14 potentially missing studies, resulting in an adjusted effect size of –0.607 (95% CI –0.800 to –0.414), compared with the observed –0.379 (95% CI –0.579 to –0.180). In contrast, no studies were imputed on the right, and the effect estimate remained unchanged. This asymmetry indicates that publication bias may be present, possibly underestimating the true effect. Therefore, findings should be interpreted with caution.

XSL•FO
RenderX

Figure 7. Funnel plot of effect size versus standard error for change in HbA_{1c}.



Sensitivity Analysis

In the primary meta-analysis of HbA_{1c} outcomes, significant heterogeneity (I^2=96%) was observed. Despite this, sensitivity analysis indicated that the overall effect remained robust, as no single study had a disproportionate impact on the results (Figure S5 in Section 5 in Multimedia Appendix 2).

Discussion

Principal Findings

This systematic review and meta-analysis synthesized data from 58 RCTs to evaluate the effectiveness of telemedicine interventions in managing glycemic control, weight, and cardiovascular health in patients with T2DM. The results showed that telemedicine was more effective than usual care, especially in improving HbA_{1c} levels. All types of telemedicine—synchronous, asynchronous, remote monitoring, and hybrid—significantly reduced HbA_{1c} levels. Interventions delivered by physicians and dietitians had a greater impact on HbA_{1c} than those delivered by other providers, suggesting that provider expertise may affect outcomes. In addition, telemedicine demonstrated greater effectiveness in short- and long-term interventions, while mid-term interventions showed no statistically significant difference compared with usual care.

This review found that the use of telemedicine in patients with T2DM has been more extensively studied in high-income countries. Consistent with previous findings [87,93], the majority of the evidence (72%) originates from high-income countries, with the United States contributing the highest proportion (28%). Telemedicine remains in the early stages in low-income countries. Several factors may explain this trend, including the digital divide, legal barriers, and reimbursement challenges. The digital divide, characterized by differences in access to telecommunications technologies based on geographic and socioeconomic factors (eg, rural areas, low income, and low education levels), is a major barrier to the adoption of telemedicine [94]. In addition, legal barriers such as national licensing, practice laws, certification requirements, and liability issues also limit the use of telemedicine [94]. Legal and policy changes regarding reimbursement can ensure that the convenience and benefits of telemedicine are equitably accessible and enjoyed by the public, especially vulnerable groups. Limited reimbursement policies constrain the adoption of telemedicine, especially for underserved populations. Legal and policy reforms that ensure equitable reimbursement can enhance the accessibility and benefits of telemedicine for vulnerable groups [94,95]. National and local-level advocacy is needed to support legislation that can address reimbursement challenges and improve access to telemedicine.

The included studies exhibited substantial heterogeneity (I^2=94%), a common finding in telemedicine reviews, where heterogeneity typically ranges from 65% to 99% [17,22,86,93]. Hence, this level of large heterogeneity found in the present

XSL•FO
RenderX

review is not unexpected when compared with similar reviews. Several factors may account for the observed heterogeneity. First, differences in intervention characteristics—including provider type, modality (eg, synchronous, asynchronous, and remote monitoring), and intensity—likely contributed to variability in outcomes. Previous evidence has suggested that moderate- and high-intensity interventions tend to be more effective than low-intensity ones [96]. Second, differences in patient characteristics, such as comorbidities, digital health literacy, or engagement levels, may have influenced the effectiveness of interventions [97]. Only a few included studies explicitly assessed digital literacy, with most only including basic criteria such as the ability to read or operate a device. Finally, the inclusion of studies conducted in the early 2000s may have added to the heterogeneity, as telemedicine technologies and approaches have evolved significantly over the past 2 decades [98].

Based on the pooled analysis of the included trials, telemedicine emerges as a promising approach for providing effective consultation, monitoring, and management for patients with T2DM. The finding demonstrated that telemedicine interventions positively affect glycemic management, consistent with previous studies [14,99]. HbA_{1c} is an important indicator of long-term blood glucose control, and lower HbA_{1c} levels are linked to reduced risks of diabetes-related complications. Evidence suggests that a 1% decrease in HbA_{1c} is associated with a 37% reduction in microvascular complications, a 21% reduction in diabetes-related mortality, and a 14% decrease in the risk of myocardial infarction [100]. In this meta-analysis, telemedicine interventions were associated with a greater reduction in HbA_{1c} compared with usual care (MD –0.38, 95% CI –0.49 to –0.27; P<.001). These findings indicate that telemedicine has the potential not only to match, but in some aspects exceed, the clinical benefits of usual care in the management of T2DM.

Telemedicine may also provide broader benefits for managing metabolic and cardiovascular risk factors in patients with T2DM. Compared with usual care, telemedicine interventions showed greater improvements in FBG, body weight, BMI, SBP, and DBP. Although slight improvements in LDL-c and HDL-c levels were observed in the intervention group, no significant differences were found. This finding is consistent with a previous meta-analysis conducted in primary care settings, which also reported no significant differences between telemedicine and usual care in lipid-related outcomes [22]. However, the absence of significant effects on lipid profiles does not necessarily imply a lack of cardiovascular benefit. One possible explanation is that most telemedicine interventions for T2DM primarily focused on glycemic control, with fewer targeting lipid management specifically. Cardiovascular disease remains the leading cause of death among patients with T2DM [101]. This highlights that, beyond glycemic control, managing cardiovascular risk factors is also essential in T2DM population [102]. This review suggests that telemedicine plays an important role in improving glycemic control and cardiovascular-related health indicators in patients with T2DM. They also underscore the need for more standardized, comprehensive intervention protocols and better integration of multidisciplinary teams to enhance the overall effectiveness of telemedicine in future research and practice [103].

This study observed that different intervention modalities—including synchronous, asynchronous, remote monitoring, and hybrid models—were significantly more effective in improving HbA_{1c} levels compared with usual care, aligning with findings from prior reviews [20,104,105]. However, some studies suggested that synchronous and asynchronous interventions did not consistently outperform usual care. For instance, an earlier review indicated that synchronous and asynchronous teleconsultations did not lead to statistically significant reductions in HbA_{1c} levels [106]. Furthermore, while synchronous teleconsultations showed more substantial HbA_{1c} improvements, no significant benefits were observed in BMI, blood pressure, or cardiovascular health [20,105]. These discrepancies may be attributed to advancements in telemedicine technology, the enhancement of platform functionality, and the increasing standardization of service content. Differences in study design, including intervention frequency, methodology standardization, and varying content, could also play a significant role in influencing the outcomes observed across studies. Further research is needed to better understand these variations and optimize telemedicine interventions for more consistent and comprehensive clinical benefits.

The effectiveness of telemedicine interventions may vary depending on the type of health care provider involved. This review found that interventions led by doctors, dietitians, and researchers were significantly effective in improving HbA_{1c} levels compared with usual care. However, no significant differences were observed in the subgroups involving nurses, pharmacists, and coaches when compared with usual care. This discrepancy may be attributed to the limited ability of these providers to adjust treatment and management based on patients' evolving conditions, highlighting the critical role of doctor involvement, guidance, and supervision in telemedicine interventions. Moreover, remote monitoring requires more complex operations, relying on the accuracy of devices and patients' skills [107]. A previous study emphasized the importance of pretraining both patients and health care providers to enhance the reliability of remote monitoring [108]. Therefore, health care professionals need appropriate training, technical support, and clear guidelines to seamlessly integrate telemedicine into existing electronic health systems.

Although meta-regression analysis did not identify any statistically significant moderators of effect size, some variables showed trends that may have practical implications and warrant further investigation. For example, interventions delivered by physicians, dietitians, or researchers showed a trend toward greater improvements in HbA_{1c} (coefficient=0.084; P=.084), suggesting that provider expertise may influence outcomes. This is consistent with findings from subgroup analyses. Similarly, longer intervention durations (coefficient=0.142; P=.293) also demonstrated a nonsignificant positive trend, aligning with the notion that sustained engagement may enhance the effectiveness of disease management. A previous review has suggested that low-intensity telemedicine interventions are

XSL•FO
RenderX

less effective than moderate- or high-intensity ones in supporting self-management among patients with type 2 diabetes [96]. The lack of statistically significant associations may be due to the complex nature of telemedicine interventions, residual confounding, and the predominantly categorical format of moderator variables, which may limit the sensitivity of meta-regression models to detect subtle effects. These observations highlight the importance of designing tailored telemedicine strategies that consider provider qualifications, intervention duration, and delivery context—that is, the health care setting, technological infrastructure, and patient engagement environment where telemedicine is implemented [103].

Implications for Telemedicine Use on Future Public Health Emergencies

Beyond routine disease management, telemedicine also holds substantial potential in maintaining continuity of care during public health emergencies. It has already been used to manage disruptions in medical services caused by disasters such as the wildfires in Australia and Hurricane Harvey [109,110]. While the exact timing of natural disasters or infectious disease pandemics may be unpredictable, disruptions to medical services are certain to recur in the future [108]. This review focused on comparing telemedicine with usual care and demonstrated that telemedicine interventions achieved superior outcomes in patients with T2DM. These findings highlight the potential of telemedicine not only as a complementary strategy but also as an effective alternative to conventional care in situations where usual services are disrupted. To optimize the implementation of telemedicine in clinical practice, careful coordination with in-person health care systems is essential. This includes developing structured workflows involving health care professionals (eg, physicians, pharmacists, and dietitians) and support services (eg, scheduling departments, laboratories, and pharmacies). Patient stratification based on clinical needs is also crucial, as certain individuals—such as newly diagnosed patients or those starting therapies requiring specific training (eg, insulin injections)—may benefit more from hybrid care models that combine virtual assessment with in-person visits. Finally, establishing a robust, well-integrated telemedicine management system is critical to ensuring the continuous delivery of high-quality care for individuals with chronic conditions such as diabetes, particularly during emergencies.

Strengths and Limitations

This meta-analysis has several advantages. First, to our knowledge, it is the first systematic review and meta-analysis of the management of T2DM that does not restrict the types of telemedicine interventions, tools, providers, or settings, and includes a wide range of important clinical outcomes. Second, we performed several important subgroup analyses based on the type of telemedicine, intervention content, and providers. Third, our study includes a sufficient number of long-term studies. In addition, we did not impose a time limit on published studies to ensure that earlier studies were not overlooked.

However, several limitations must be considered. First, a high level of heterogeneity was observed in the results, likely due to variations in the content of telemedicine interventions, types of media used, providers, and intervention durations. While heterogeneity is common in large meta-analyses, it remains a factor that could impact interpretation [111]. Second, due to the nature of telemedicine, blinding was not possible, which may have influenced the quality of the included studies and introduced potential bias. Third, in our subgroup analyses, the number of studies was limited and displayed high heterogeneity, so the findings should be interpreted cautiously. Further research is needed to strengthen the reliability of these results. Finally, the exclusion of non-English studies could be a potential source of publication bias.

Conclusions

The results of this review demonstrate that telemedicine interventions, regardless of the type, are more effective than usual care in improving HbA_{1c} levels in patients with T2DM. In addition, telemedicine interventions led by physicians, dietitians, and researchers showed greater effectiveness in managing blood glucose levels. While the evidence for the impact of telemedicine on FBG, mean glucose, BMI, weight, blood pressure, LDL-c, and HDL-c is not as robust as for HbA_{1c}, some studies have already highlighted the potential benefits of telemedicine for these outcomes.

Acknowledgments

The authors would like to express their sincere gratitude to Professor Guangyu Lu for her valuable guidance and technical assistance with the comprehensive literature search. The authors also thank the researchers from their research team for their constructive input during study design, data extraction, and analysis discussions. The authors confirm that no generative artificial intelligence tools were used in the writing or preparation of this manuscript.

Data Availability

All data generated or analyzed during this study are included in this published article and its Multimedia Appendices.

Funding

This study was supported by Jiangsu Provincial Health Commission New Technology Introduction and Evaluation Project (grant M2022044), the Open Project Program of Key Laboratory of Big Data Analysis and Knowledge Services, Yangzhou City, Yangzhou University (grant YBK202202), Jiangsu Provincial Science and Technology Programme Special Funds (Key R&D Programme for Social Development) (BE2022775), and Yangzhou Science and Technology Development Project (YZ2024068).

XSL•FO
RenderX

Authors' Contributions

GYL and YPL were involved in the design and concept of this study. JSJ and XRG conducted the systematic literature search and data extraction and wrote this paper. JSJ and HQD conducted the analyses. JSJ, YZ, XGL, and HQD contributed to drafting the final paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[DOCX File , 25 KB - mhealth_v14i1e70429_app1.docx]

Multimedia Appendix 2
Detailed list of included and excluded studies and supplementary figures and tables related to the meta-analysis.
[DOCX File , 588 KB - mhealth_v14i1e70429_app2.docx]

References

1. Zimmet P, Alberti KGMM, Shaw J. Global and societal implications of the diabetes epidemic. Nature 2001;414(6865):782-787. [doi: 10.1038/414782a] [Medline: 11742409]
2. Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF diabetes atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022;183:109119 [FREE Full text] [doi: 10.1016/j.diabres.2021.109119] [Medline: 34879977]
3. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat Rev Endocrinol 2018;14(2):88-98. [doi: 10.1038/nrendo.2017.151] [Medline: 29219149]
4. Rapoport M, Chetrit A, Cantrell D, Novikov I, Roth J, Dankner R. Years of potential life lost in pre-diabetes and diabetes mellitus: data from a 40-year follow-up of the Israel study on glucose intolerance, obesity and hypertension. BMJ Open Diabetes Res Care 2021;9(1):e001981 [FREE Full text] [doi: 10.1136/bmjdrc-2020-001981] [Medline: 33692115]
5. Fasil A, Biadgo B, Abebe M. Glycemic control and diabetes complications among diabetes mellitus patients attending at University of Gondar Hospital, Northwest Ethiopia. Diabetes Metab Syndr Obes 2019;12:75-83 [FREE Full text] [doi: 10.2147/DMSO.S185614] [Medline: 30613158]
6. Zheng B, Su B, Price G, Tzoulaki I, Ahmadi-Abhari S, Middleton L. Glycemic control, diabetic complications, and risk of dementia in patients with diabetes: results from a large U.K. cohort study. Diabetes Care 2021;44(7):1556-1563. [doi: 10.2337/dc20-2850] [Medline: 34035076]
7. Long C, Tang Y, Huang J, Liu S, Xing Z. Association of long-term visit-to-visit variability of HbA1c and fasting glycemia with hypoglycemia in type 2 diabetes mellitus. Front Endocrinol (Lausanne) 2022;13:975468 [FREE Full text] [doi: 10.3389/fendo.2022.975468] [Medline: 36034445]
8. Yamaoka K, Nemoto A, Tango T. Comparison of the effectiveness of lifestyle modification with other treatments on the incidence of type 2 diabetes in people at high risk: a network meta-analysis. Nutrients 2019;11(6):1373 [FREE Full text] [doi: 10.3390/nu11061373] [Medline: 31248094]
9. Vijan S, Sussman JB, Yudkin JS, Hayward RA. Effect of patients' risks and preferences on health gains with plasma glucose level lowering in type 2 diabetes mellitus. JAMA Intern Med 2014;174(8):1227-1234 [FREE Full text] [doi: 10.1001/jamainternmed.2014.2894] [Medline: 24979148]
10. Kaur N, Majumdar V, Nagarathna R, Malik N, Anand A, Nagendra HR. Diabetic yoga protocol improves glycemic, anthropometric and lipid levels in high risk individuals for diabetes: a randomized controlled trial from Northern India. Diabetol Metab Syndr 2021;13(1):149 [FREE Full text] [doi: 10.1186/s13098-021-00761-1] [Medline: 34949227]
11. Cardoso CRL, Leite NC, Salles GF. Prognostic impact of changes in aortic stiffness for cardiovascular and mortality outcomes in individuals with type 2 diabetes: the Rio de Janeiro cohort study. Cardiovasc Diabetol 2022;21(1):76 [FREE Full text] [doi: 10.1186/s12933-022-01514-8] [Medline: 35568947]
12. AkbariRad M, Dehghani M, Sadeghi M, Torshizian A, Saeedi N, Sarabi M, et al. The effect of telenursing on disease outcomes in people with type 2 diabetes mellitus: a narrative review. J Diabetes Res 2023;2023:4729430 [FREE Full text] [doi: 10.1155/2023/4729430] [Medline: 38098964]
13. Raesi R, Shaye ZA, Saghari S, Sheikh Beig Goharrizi MA, Raei M, Hushmandi K. The impact of education through nurse-led telephone follow-up (telenursing) on the quality of life of COVID-19 patients. J Egypt Public Health Assoc 2021;96(1):30 [FREE Full text] [doi: 10.1186/s42506-021-00093-y] [Medline: 34748085]
14. Eberle C, Stichling S. Clinical improvements by telemedicine interventions managing type 1 and type 2 diabetes: systematic meta-review. J Med Internet Res 2021;23(2):e23244 [FREE Full text] [doi: 10.2196/23244] [Medline: 33605889]

XSL•FO
RenderX

15. Timpel P, Oswald S, Schwarz PEH, Harst L. Mapping the evidence on the effectiveness of telemedicine interventions in diabetes, dyslipidemia, and hypertension: an umbrella review of systematic reviews and meta-analyses. J Med Internet Res 2020;22(3):e16791 [FREE Full text] [doi: 10.2196/16791] [Medline: 32186516]
16. Ryu S. Telemedicine: opportunities and developments in member states: report on the second global survey on eHealth 2009 (global observatory for eHealth series, volume 2). Healthc Inform Res 2012;18(2):153-155. [doi: 10.4258/hir.2012.18.2.153]
17. Tchero H, Kangambega P, Briatte C, Brunet-Houdard S, Retali G, Rusch E. Clinical effectiveness of telemedicine in diabetes mellitus: a meta-analysis of 42 randomized controlled trials. Telemed J E Health 2019;25(7):569-583. [doi: 10.1089/tmj.2018.0128] [Medline: 30124394]
18. Shan R, Sarkar S, Martin SS. Digital health technology and mobile devices for the management of diabetes mellitus: state of the art. Diabetologia 2019;62(6):877-887. [doi: 10.1007/s00125-019-4864-7] [Medline: 30963188]
19. Hood M, Wilson R, Corsica J, Bradley L, Chirinos D, Vivo A. What do we know about mobile applications for diabetes self-management? A review of reviews. J Behav Med 2016;39(6):981-994. [doi: 10.1007/s10865-016-9765-3] [Medline: 27412774]
20. Zhang J, Ji X, Xie J, Lin K, Yao M, Chi C. Effectiveness of synchronous teleconsultation for patients with type 2 diabetes mellitus: a systematic review and meta-analysis. BMJ Open Diabetes Res Care 2023;11(1):e003180 [FREE Full text] [doi: 10.1136/bmjdrc-2022-003180] [Medline: 36822665]
21. Eberle C, Löhnert M, Stichling S. Effectiveness of disease-specific mhealth apps in patients with diabetes mellitus: scoping review. JMIR Mhealth Uhealth 2021;9(2):e23477 [FREE Full text] [doi: 10.2196/23477] [Medline: 33587045]
22. Alfarwan N, Hodkinson A, Panagioti M, Hassan L, Kontopantelis E. Clinical and cost-effectiveness of telemedicine among patients with type 2 diabetes in primary care: a systematic review and meta-analysis. Diabet Med 2024;41(8):e15343. [doi: 10.1111/dme.15343] [Medline: 38780107]
23. Niu XD, Chi JT, Guo JB, Ruan HH, Zhang J, Tao HX, et al. Effects of nurse-led web-based interventions on people with type 2 diabetes mellitus: a systematic review and meta-analysis. J Telemed Telecare 2021;27(5):269-279. [doi: 10.1177/1357633X211010019] [Medline: 33966522]
24. Fu H, McMahon SK, Gross CR, Adam TJ, Wyman JF. Usability and clinical efficacy of diabetes mobile applications for adults with type 2 diabetes: a systematic review. Diabetes Res Clin Pract 2017;131:70-81. [doi: 10.1016/j.diabres.2017.06.016] [Medline: 28692830]
25. Yang L, Xu J, Kang C, Bai Q, Wang X, Du S, et al. Effects of mobile phone-based telemedicine management in patients with type 2 diabetes mellitus: a randomized clinical trial. Am J Med Sci 2022;363(3):224-231. [doi: 10.1016/j.amjms.2021.09.001] [Medline: 34534510]
26. Haghighinejad H, Liaghat L, Malekpour F, Jafari P, Taghipour K, Rezaie M, et al. Comparing the effects of SMS-based education with group-based education and control group on diabetes management: a randomized educational program. BMC Prim Care 2022;23(1):209 [FREE Full text] [doi: 10.1186/s12875-022-01820-w] [Medline: 35986262]
27. Holmen H, Torbjørnsen A, Wahl AK, Jenum AK, Småstuen MC, Arsand E, et al. A mobile health intervention for self-management and lifestyle change for persons with type 2 diabetes, part 2: one-year results from the Norwegian randomized controlled trial RENEWING HEALTH. JMIR Mhealth Uhealth 2014;2(4):e57 [FREE Full text] [doi: 10.2196/mhealth.3882] [Medline: 25499872]
28. Jeong JY, Jeon J, Bae K, Choi Y, Park K, Kim J, et al. Smart care based on telemonitoring and telemedicine for type 2 diabetes care: multi-center randomized controlled trial. Telemed J E Health 2018;24(8):604-613. [doi: 10.1089/tmj.2017.0203] [Medline: 29341843]
29. Lee EY, Cha S, Yun J, Lim S, Lee J, Ahn Y, et al. Efficacy of personalized diabetes self-care using an electronic medical record-integrated mobile app in patients with type 2 diabetes: 6-month randomized controlled trial. J Med Internet Res 2022;24(7):e37430 [FREE Full text] [doi: 10.2196/37430] [Medline: 35900817]
30. Anzaldo-Campos MC, Contreras S, Vargas-Ojeda A, Menchaca-Díaz R, Fortmann A, Philis-Tsimikas A. Dulce Wireless Tijuana: a randomized control trial evaluating the impact of project dulce and short-term mobile technology on glycemic control in a family medicine clinic in Northern Mexico. Diabetes Technol Ther 2016;18(4):240-251 [FREE Full text] [doi: 10.1089/dia.2015.0283] [Medline: 26914371]
31. Arora S, Peters AL, Burner E, Lam CN, Menchine M. Trial to examine text message-based mHealth in emergency department patients with diabetes (TExT-MED): a randomized controlled trial. Ann Emerg Med 2014;63(6):745-54.e6. [doi: 10.1016/j.annemergmed.2013.10.012] [Medline: 24225332]
32. Azelton KR, Crowley AP, Vence N, Underwood K, Morris G, Kelly J, et al. Digital health coaching for type 2 diabetes: randomized controlled trial of healthy at home. Front Digit Health 2021;3:764735 [FREE Full text] [doi: 10.3389/fdgth.2021.764735] [Medline: 34901926]
33. Basudev N, Crosby-Nwaobi R, Thomas S, Chamley M, Murrells T, Forbes A. A prospective randomized controlled study of a virtual clinic integrating primary and specialist care for patients with type 2 diabetes mellitus. Diabet Med 2016;33(6):768-776. [doi: 10.1111/dme.12985] [Medline: 27194175]

XSL•FO
RenderX

34. Bentley CL, Otesile O, Bacigalupo R, Elliott J, Noble H, Hawley MS, et al. Feasibility study of portable technology for weight loss and HbA1c control in type 2 diabetes. BMC Med Inform Decis Mak 2016;16:92 [FREE Full text] [doi: 10.1186/s12911-016-0331-2] [Medline: 27418275]
35. Cho JH, Kim H, Yoo SH, Jung CH, Lee WJ, Park CY, et al. An internet-based health gateway device for interactive communication and automatic data uploading: clinical efficacy for type 2 diabetes in a multi-centre trial. J Telemed Telecare 2017;23(6):595-604. [doi: 10.1177/1357633X16657500] [Medline: 27381040]
36. Christensen JR, Hesseldal L, Olesen TB, Olsen MH, Jakobsen PR, Laursen DH, et al. Long-term weight loss in a 24-month primary care-anchored telehealth lifestyle coaching program: randomized controlled trial. J Telemed Telecare 2022;28(10):764-770. [doi: 10.1177/1357633X221123411] [Medline: 36346936]
37. Christensen JR, Laursen DH, Lauridsen JT, Hesseldal L, Jakobsen PR, Nielsen JB, et al. Reversing type 2 diabetes in a primary care-anchored eHealth lifestyle coaching programme in Denmark: a randomised controlled trial. Nutrients 2022;14(16):3424 [FREE Full text] [doi: 10.3390/nu14163424] [Medline: 36014930]
38. Dale J, Caramlau I, Sturt J, Friede T, Walker R. Telephone peer-delivered intervention for diabetes motivation and support: the telecare exploratory RCT. Patient Educ Couns 2009;75(1):91-98. [doi: 10.1016/j.pec.2008.09.014] [Medline: 19013741]
39. Dario C, Toffanin R, Calcaterra F, Saccavini C, Stafylas P, Mancin S, et al. Telemonitoring of type 2 diabetes mellitus in Italy. Telemed J E Health 2017;23(2):143-152. [doi: 10.1089/tmj.2015.0224] [Medline: 27379995]
40. Dunkel A, von Storch K, Hochheim M, Zank S, Polidori MC, Woopen C. Long-term effects of a telemedically-assisted lifestyle intervention on glycemic control in patients with type 2 diabetes—a two-armed randomised controlled trial in Germany. J Diabetes Metab Disord 2023;23(1):519-532 [FREE Full text] [doi: 10.1007/s40200-023-01290-6] [Medline: 38932898]
41. Eakin EG, Winkler EA, Dunstan DW, Healy GN, Owen N, Marshall AM, et al. Living well with diabetes: 24-month outcomes from a randomized trial of telephone-delivered weight loss and physical activity intervention to improve glycemic control. Diabetes Care 2014;37(8):2177-2185 [FREE Full text] [doi: 10.2337/dc13-2427] [Medline: 24658390]
42. Farmer A, Bobrow K, Leon N, Williams N, Phiri E, Namadingo H, et al. Digital messaging to support control for type 2 diabetes (StAR2D): a multicentre randomised controlled trial. BMC Public Health 2021;21(1):1907 [FREE Full text] [doi: 10.1186/s12889-021-11874-7] [Medline: 34674688]
43. Gong E, Baptista S, Russell A, Scuffham P, Riddell M, Speight J, et al. My diabetes coach, a mobile app-based interactive conversational agent to support type 2 diabetes self-management: randomized effectiveness-implementation trial. J Med Internet Res 2020;22(11):e20322 [FREE Full text] [doi: 10.2196/20322] [Medline: 33151154]
44. Hee-Sung K. Impact of web-based nurse's education on glycosylated haemoglobin in type 2 diabetic patients. J Clin Nurs 2007;16(7):1361-1366. [doi: 10.1111/j.1365-2702.2007.01506.x] [Medline: 17584355]
45. Hoda F, Arshad M, Khan MA, Kohli S, Kareem S, Akhtar M, et al. Impact of a mHealth intervention in type 2 diabetes mellitus patients: a randomized clinical trial. SN Compr. Clin. Med 2023;5(1). [doi: 10.1007/s42399-023-01564-3]
46. Hsu WC, Lau KHK, Huang R, Ghiloni S, Le H, Gilroy S, et al. Utilization of a cloud-based diabetes management program for insulin initiation and titration enables collaborative decision making between healthcare providers and patients. Diabetes Technol Ther 2016;18(2):59-67. [doi: 10.1089/dia.2015.0160] [Medline: 26645932]
47. Jarab AS, Alqudah SG, Mukattash TL, Shattat G, Al-Qirim T. Randomized controlled trial of clinical pharmacy management of patients with type 2 diabetes in an outpatient diabetes clinic in Jordan. J Manag Care Pharm 2012;18(7):516-526 [FREE Full text] [doi: 10.18553/jmcp.2012.18.7.516] [Medline: 22971205]
48. Kempf K, Altpeter B, Berger J, Reuß O, Fuchs M, Schneider M, et al. Efficacy of the telemedical lifestyle intervention program telipro in advanced stages of type 2 diabetes: a randomized controlled trial. Diabetes Care 2017;40(7):863-871. [doi: 10.2337/dc17-0303] [Medline: 28500214]
49. Kempf K, Dubois C, Arnold M, Amelung V, Leppert N, Altin S, et al. Effectiveness of the telemedical lifestyle intervention program TeLIPro for improvement of HbA in type 2 diabetes: a randomized-controlled trial in a real-life setting. Nutrients 2023;15(18) [FREE Full text] [doi: 10.3390/nu15183954] [Medline: 37764738]
50. Kitazawa M, Takeda Y, Hatta M, Horikawa C, Sato T, Osawa T, et al. Lifestyle intervention with smartphone app and iscgm for people at high risk of type 2 diabetes: randomized trial. J Clin Endocrinol Metab 2024;109(4):1060-1070 [FREE Full text] [doi: 10.1210/clinem/dgad639] [Medline: 37931069]
51. Kooiman TJM, de Groot M, Hoogenberg K, Krijnen WP, van der Schans CP, Kooy A. Self-tracking of physical activity in people with type 2 diabetes: A randomized controlled trial. Comput Inform Nurs 2018;36(7):340-349. [doi: 10.1097/CIN.0000000000000443] [Medline: 29742550]
52. Lauffenburger JC, Ghazinouri R, Jan S, Makanji S, Ferro CA, Lewey J, et al. Impact of a novel pharmacist-delivered behavioral intervention for patients with poorly-controlled diabetes: The ENhancing outcomes through goal assessment and generating engagement in diabetes mellitus (ENGAGE-DM) pragmatic randomized trial. PLoS One 2019;14(4):e0214754 [FREE Full text] [doi: 10.1371/journal.pone.0214754] [Medline: 30939143]
53. Lee JY, Chan CKY, Chua SS, Ng CJ, Paraidathathu T, Lee KKC, et al. Telemonitoring and team-based management of glycemic control on people with type 2 diabetes: a cluster-randomized controlled trial. J Gen Intern Med 2020;35(1):87-94 [FREE Full text] [doi: 10.1007/s11606-019-05316-9] [Medline: 31512187]

XSL•FO
RenderX

54. Leong CM, Lee T, Chien Y, Kuo L, Kuo Y, Chen H. Social media-delivered patient education to enhance self-management and attitudes of patients with type 2 diabetes during the covid-19 pandemic: randomized controlled trial. J Med Internet Res 2022;24(3):e31449 [FREE Full text] [doi: 10.2196/31449] [Medline: 35319478]
55. Lim SL, Ong KW, Johal J, Han CY, Yap QV, Chan YH, et al. Effect of a smartphone app on weight change and metabolic outcomes in asian adults with type 2 diabetes: a randomized clinical trial. JAMA Netw Open 2021;4(6):e2112417 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.12417] [Medline: 34081137]
56. Liou JK, Soon MS, Chen CH, Huang TF, Chen YP, Yeh YP, et al. Shared care combined with telecare improves glycemic control of diabetic patients in a rural underserved community. Telemed J E Health 2014;20(2):175-178. [doi: 10.1089/tmj.2013.0037] [Medline: 24320193]
57. Lorig K, Ritter P, Villa F, Piette J. Spanish diabetes self-management with and without automated telephone reinforcement: two randomized trials. Diabetes Care 2008;31(3):408-414. [doi: 10.2337/dc07-1313] [Medline: 18096810]
58. Luley C, Blaik A, Reschke K, Klose S, Westphal S. Weight loss in obese patients with type 2 diabetes: effects of telemonitoring plus a diet combination - the active body control (ABC) program. Diabetes Res Clin Pract 2011;91(3):286-292. [doi: 10.1016/j.diabres.2010.11.020] [Medline: 21168231]
59. Mitchell S, Bragg A, De La Cruz BA, Winter M, Reichert M, Laird L, et al. Effectiveness of an immersive telemedicine platform for delivering diabetes medical group visits for african american, black and hispanic, or latina women with uncontrolled diabetes: the women in control 2.0 noninferiority randomized clinical trial. J Med Internet Res 2023;25:e43669 [FREE Full text] [doi: 10.2196/43669] [Medline: 37163341]
60. Orsama AL, Lähteenmäki J, Harno K, Kulju M, Wintergerst E, Schachner H, et al. Active assistance technology reduces glycosylated hemoglobin and weight in individuals with type 2 diabetes: results of a theory-based randomized trial. Diabetes Technol Ther 2013;15(8):662-669. [doi: 10.1089/dia.2013.0056] [Medline: 23844570]
61. Oseran AS, Rao K, Chang Y, He W, Sikora CE, Wexler DJ, et al. HbA1c-triggered endocrinology electronic consultation for type 2 diabetes management. J Gen Intern Med 2022;37(5):1081-1087 [FREE Full text] [doi: 10.1007/s11606-021-07157-x] [Medline: 34608564]
62. Parsons SN, Luzio SD, Harvey JN, Bain SC, Cheung WY, Watkins A, et al. Effect of structured self-monitoring of blood glucose, with and without additional TeleCare support, on overall glycaemic control in non-insulin treated Type 2 diabetes: the SMBG Study, a 12-month randomized controlled trial. Diabet Med 2019;36(5):578-590 [FREE Full text] [doi: 10.1111/dme.13899] [Medline: 30653704]
63. Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. Diabetes Care 2011;34(9):1934-1942 [FREE Full text] [doi: 10.2337/dc11-0366] [Medline: 21788632]
64. Sachmechi I, Salam S, Amini M, Khan R, Spitznogle A, Belen T. Frequent monitoring of blood glucose levels via a remote patient monitoring system helps improve glycemic control. Endocr Pract 2023;29(6):441-447. [doi: 10.1016/j.eprac.2023.03.270] [Medline: 36965657]
65. Sarayani A, Mashayekhi M, Nosrati M, Jahangard-Rafsanjani Z, Javadi M, Saadat N, et al. Efficacy of a telephone-based intervention among patients with type-2 diabetes; a randomized controlled trial in pharmacy practice. Int J Clin Pharm 2018;40(2):345-353. [doi: 10.1007/s11096-018-0593-0] [Medline: 29435911]
66. Sun C, Sun L, Xi S, Zhang H, Wang H, Feng Y, et al. Mobile phone-based telemedicine practice in older chinese patients with type 2 diabetes mellitus: randomized controlled trial. JMIR Mhealth Uhealth 2019;7(1):e10664 [FREE Full text] [doi: 10.2196/10664] [Medline: 30609983]
67. Torbjørnsen A, Jenum AK, Småstuen MC, Arsand E, Holmen H, Wahl AK, et al. A low-intensity mobile health intervention with and without health counseling for persons with type 2 diabetes, part 1: baseline and short-term results from a randomized controlled trial in the norwegian part of RENEWING HEALTH. JMIR Mhealth Uhealth 2014;2(4):e52 [FREE Full text] [doi: 10.2196/mhealth.3535] [Medline: 25499592]
68. Yin W, Liu Y, Hu H, Sun J, Liu Y, Wang Z. Telemedicine management of type 2 diabetes mellitus in obese and overweight young and middle-aged patients during COVID-19 outbreak: A single-center, prospective, randomized control study. PLoS One 2022;17(9):e0275251 [FREE Full text] [doi: 10.1371/journal.pone.0275251] [Medline: 36174028]
69. Zhang P, Tao X, Ma Y, Zhang Y, Ma X, Song H, et al. Improving the management of type 2 diabetes in China using a multifaceted digital health intervention in primary health care: the SMARTDiabetes cluster randomised controlled trial. Lancet Reg Health West Pac 2024;49:101130 [FREE Full text] [doi: 10.1016/j.lanwpc.2024.101130] [Medline: 39056088]
70. Yang Y, Lee EY, Kim H, Lee S, Yoon K, Cho J. Effect of a mobile phone-based glucose-monitoring and feedback system for type 2 diabetes management in multiple primary care clinic settings: cluster randomized controlled trial. JMIR Mhealth Uhealth 2020;8(2):e16266 [FREE Full text] [doi: 10.2196/16266] [Medline: 32130172]
71. Jantraporn R, Pichayapinyo P, Lagampan S, Mayurasakorn K. Effects of carbohydrate reduction program and telemonitoring on glycosylated hemoglobin in patients with poorly controlled type 2 diabetes: A randomized controlled trial. J Med Assoc Thailand 2019;102(5):523-529.
72. Capozza K, Woolsey S, Georgsson M, Black J, Bello N, Lence C, et al. Going mobile with diabetes support: a randomized study of a text message-based personalized behavioral intervention for type 2 diabetes self-care. Diabetes Spectr 2015 May;28(2):83-91 [FREE Full text] [doi: 10.2337/diaspect.28.2.83] [Medline: 25987806]

XSL•FO
RenderX

73. Franc S, Joubert M, Daoudi A, Fagour C, Benhamou P, Rodier M, et al. Efficacy of two telemonitoring systems to improve glycaemic control during basal insulin initiation in patients with type 2 diabetes: the TeleDiab - 2 randomized controlled trial. Diabetes Obes Metab 2019 Jul 03;21(10):2327-2332. [doi: 10.1111/DOM.13806]
74. Gerber BS, Biggers A, Tilton JJ, Smith Marsh DE, Lane R, Mihailescu D, et al. Mobile health intervention in patients with type 2 diabetes: a randomized clinical trial. JAMA Netw Open 2023 Sep 05;6(9):e2333629 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.33629] [Medline: 37773498]
75. María Gómez A, Cristina Henao D, León Vargas F, Mauricio Muñoz O, David Lucero O, García Jaramillo M, et al. Efficacy of the mHealth application in patients with type 2 diabetes transitioning from inpatient to outpatient care: a randomized controlled clinical trial. Diabetes Res Clin Pract 2022 Jul;189:109948 [FREE Full text] [doi: 10.1016/j.diabres.2022.109948] [Medline: 35700926]
76. Greenwood DA, Blozis SA, Young HM, Nesbitt TS, Quinn CC. Overcoming clinical inertia: a randomized clinical trial of a telehealth remote monitoring intervention using paired glucose testing in adults with type 2 diabetes. J Med Internet Res 2015 Jul 21;17(7):e178 [FREE Full text] [doi: 10.2196/jmir.4112] [Medline: 26199142]
77. Khanna R, Stoddard PJ, Gonzales EN, Villagran-Flores M, Thomson J, Bayard P, et al. An automated telephone nutrition support system for Spanish-speaking patients with diabetes. J Diabetes Sci Technol 2014 Nov;8(6):1115-1120 [FREE Full text] [doi: 10.1177/1932296814550186] [Medline: 25239122]
78. Kleinman NJ, Shah A, Shah S, Phatak S, Viswanathan V. Impact of the gather mHealth system on A1C: primary results of a multisite randomized clinical trial among people with type 2 diabetes in India. Diabetes Care 2016 Oct;39(10):e169-e170. [doi: 10.2337/dc16-0869] [Medline: 27493133]
79. Klingeman H, Funnell M, Jhand A, Lathkar-Pradhan S, Hodish I. Type 2 diabetes specialty clinic model for the accountable care organization era. J Diabetes Complications 2017 Oct;31(10):1521-1526. [doi: 10.1016/j.jdiacomp.2017.05.011] [Medline: 28793967]
80. Tang PC, Overhage JM, Chan AS, Brown NL, Aghighi B, Entwistle MP, et al. Online disease management of diabetes: engaging and motivating patients online with enhanced resources-diabetes (EMPOWER-D), a randomized controlled trial. J Am Med Inform Assoc 2013 May 01;20(3):526-534 [FREE Full text] [doi: 10.1136/amiajnl-2012-001263] [Medline: 23171659]
81. Turnin MC, Gourdy P, Martini J, Buisson JC, Chauchard MC, Delaunay J, Educ@dom Study Group. Impact of a remote monitoring programme including lifestyle education software in type 2 diabetes: results of the Educ@dom randomised multicentre study. Diabetes Ther 2021 Jul;12(7):2059-2075 [FREE Full text] [doi: 10.1007/s13300-021-01095-x] [Medline: 34160791]
82. Vaughan EM, Hyman DJ, Naik AD, Samson SL, Razjouyan J, Foreyt JP. A telehealth-supported, integrated care with CHWs, and MEdication-access (TIME) program for diabetes improves HbA1c: a randomized clinical trial. J Gen Intern Med 2021 Feb;36(2):455-463 [FREE Full text] [doi: 10.1007/s11606-020-06017-4] [Medline: 32700217]
83. Viswanathan M, Golin CE, Jones CD, Ashok M, Blalock SJ, Wines RCM, et al. Interventions to improve adherence to self-administered medications for chronic diseases in the United States: a systematic review. Ann Intern Med 2012;157(11):785-795 [FREE Full text] [doi: 10.7326/0003-4819-157-11-201212040-00538] [Medline: 22964778]
84. Telemedicine and telehealth. Assistant Secretary for Technology Policy. URL: https://healthit.gov/public-health/telemedicine-and-telehealth/ [accessed 2026-01-27]
85. Mechanic OJ, Persaud Y, Kimball AB. Telehealth systems. In: Treasure Island (Fl) Ineligible Companies. Disclosure: Yudy Persaud Declares No Relevant Financial Relationships with Ineligible Companies. Orlando, Florida: StatPearls Publishing LLC; 2024.
86. Hangaard S, Laursen SH, Andersen JD, Kronborg T, Vestergaard P, Hejlesen O, et al. The effectiveness of telemedicine solutions for the management of type 2 diabetes: a systematic review, meta-analysis, and meta-regression. J Diabetes Sci Technol 2023;17(3):794-825 [FREE Full text] [doi: 10.1177/19322968211064633] [Medline: 34957864]
87. Zhai Y, Zhu W, Cai Y, Sun D, Zhao J. Clinical- and cost-effectiveness of telemedicine in type 2 diabetes mellitus: a systematic review and meta-analysis. Medicine (Baltimore) 2014;93(28):e312 [FREE Full text] [doi: 10.1097/MD.0000000000000312] [Medline: 25526482]
88. GRADEpro GDT. Cochrane Methods GRADEing. URL: https://methods.cochrane.org/gradeing/gradepro-gdt [accessed 2026-01-15]
89. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
90. Little RR, Rohlfing C, Sacks DB. The national glycohemoglobin standardization program: over 20 years of improving hemoglobin a measurement. Clin Chem 2019;65(7):839-848 [FREE Full text] [doi: 10.1373/clinchem.2018.296962] [Medline: 30518660]
91. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol 2014;14:135 [FREE Full text] [doi: 10.1186/1471-2288-14-135] [Medline: 25524443]

XSL•FO
RenderX

92. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. BMC Med Res Methodol 2005;5:13 [FREE Full text] [doi: 10.1186/1471-2288-5-13] [Medline: 15840177]
93. Faruque LI, Wiebe N, Ehteshami-Afshar A, Liu Y, Dianati-Maleki N, Hemmelgarn BR, Alberta Kidney Disease Network. Effect of telemedicine on glycated hemoglobin in diabetes: a systematic review and meta-analysis of randomized trials. CMAJ 2017;189(9):E341-E364 [FREE Full text] [doi: 10.1503/cmaj.150885] [Medline: 27799615]
94. Dorsey ER, Topol EJ. State of telehealth. N Engl J Med 2016;375(2):154-161. [doi: 10.1056/NEJMra1601705] [Medline: 27410924]
95. Bashshur RL, Shannon G, Krupinski EA, Grigsby J. Sustaining and realizing the promise of telemedicine. Telemed J E Health 2013;19(5):339-345 [FREE Full text] [doi: 10.1089/tmj.2012.0282] [Medline: 23289907]
96. Kerr D, Ahn D, Waki K, Wang J, Breznen B, Klonoff DC. Digital interventions for self-management of type 2 diabetes mellitus: systematic literature review and meta-analysis. J Med Internet Res 2024;26:e55757 [FREE Full text] [doi: 10.2196/55757] [Medline: 39037772]
97. Butayeva J, Ratan ZA, Downie S, Hosseinzadeh H. The impact of health literacy interventions on glycemic control and self-management outcomes among type 2 diabetes mellitus: A systematic review. J Diabetes 2023;15(9):724-735 [FREE Full text] [doi: 10.1111/1753-0407.13436] [Medline: 37407516]
98. Noel K, Ellison B. Inclusive innovation in telehealth. NPJ Digit Med 2020;3:89 [FREE Full text] [doi: 10.1038/s41746-020-0296-5] [Medline: 32613084]
99. Chiaranai C, Chularee S, Saokaew S, Bhatarasakoon P, Umnuaypornlert A, Chaomuang N, et al. Effectiveness of telehealth on the glycemic control of patients with type 2 diabetes mellitus during the COVID-19 pandemic: A systematic review and meta-analysis of randomised controlled trials. Int J Nurs Stud Adv 2024;6:100169 [FREE Full text] [doi: 10.1016/j.ijnsa.2023.100169] [Medline: 38746799]
100. Stratton IM, Adler AI, Neil HA, Matthews DR, Manley SE, Cull CA, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. BMJ 2000;321(7258):405-412 [FREE Full text] [doi: 10.1136/bmj.321.7258.405] [Medline: 10938048]
101. Einarson TR, Acs A, Ludwig C, Panton UH. Prevalence of cardiovascular disease in type 2 diabetes: a systematic literature review of scientific evidence from across the world in 2007-2017. Cardiovasc Diabetol 2018;17(1):83. [doi: 10.1186/s12933-018-0728-6] [Medline: 29884191]
102. Ma C, Ma X, Guan C, Li Y, Mauricio D, Fu S. Cardiovascular disease in type 2 diabetes mellitus: progress toward personalized management. Cardiovasc Diabetol 2022;21(1):74 [FREE Full text] [doi: 10.1186/s12933-022-01516-6] [Medline: 35568946]
103. M Tourkmani A, J ALHarbi T, Rsheed AMB, Alrasheedy AA, ALMadani W, ALJuraisi F, et al. The impact of telemedicine on patients with uncontrolled type 2 diabetes mellitus during the COVID-19 pandemic in Saudi Arabia: findings and implications. J Telemed Telecare 2023;29(5):390-398 [FREE Full text] [doi: 10.1177/1357633X20985763] [Medline: 33525952]
104. Lee PA, Greenfield G, Pappas Y. The impact of telehealth remote patient monitoring on glycemic control in type 2 diabetes: a systematic review and meta-analysis of systematic reviews of randomised controlled trials. BMC Health Serv Res 2018;18(1):495. [doi: 10.1186/s12913-018-3274-8] [Medline: 29940936]
105. Mabeza RMS, Maynard K, Tarn DM. Influence of synchronous primary care telemedicine versus in-person visits on diabetes, hypertension, and hyperlipidemia outcomes: a systematic review. BMC Prim Care 2022;23(1):52 [FREE Full text] [doi: 10.1186/s12875-022-01662-6] [Medline: 35313804]
106. Verhoeven F, Tanja-Dijkstra K, Nijland N, Eysenbach G, van Gemert-Pijnen L. Asynchronous and synchronous teleconsultation for diabetes care: a systematic literature review. J Diabetes Sci Technol 2010;4(3):666-684 [FREE Full text] [doi: 10.1177/193229681000400323] [Medline: 20513335]
107. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
108. Smith AC, Thomas E, Snoswell CL, Haydon H, Mehrotra A, Clemensen J, et al. Telehealth for global emergencies: Implications for coronavirus disease 2019 (COVID-19). J Telemed Telecare 2020;26(5):309-313 [FREE Full text] [doi: 10.1177/1357633X20916567] [Medline: 32196391]
109. A multinational telemedicine systems for disaster response: opportunities and challenges. North Atlantic Treaty Organization. 2017. URL: https://www.nato.int/cps/en/natohq/topics_168039.htm [accessed 2024-10-20]
110. Caffery LJ, Bradford NK, Smith AC, Langbecker D. How telehealth facilitates the provision of culturally appropriate healthcare for indigenous Australians. J Telemed Telecare 2018;24(10):676-682. [doi: 10.1177/1357633X18795764] [Medline: 30343658]
111. Kontopantelis E, Springate DA, Reeves D. A re-analysis of the cochrane library data: the dangers of unobserved heterogeneity in meta-analyses. PLoS One 2013;8(7):e69930 [FREE Full text] [doi: 10.1371/journal.pone.0069930] [Medline: 23922860]

XSL•FO
RenderX

Abbreviations

DBP: diastolic blood pressure
FBG: fasting blood glucose
GRADE: Grading of Recommendations Assessment, Development, and Evaluation
HbA_{1c}: hemoglobin A_{1c}
HDL-c: high-density lipoprotein cholesterol
LDL-c: low-density lipoprotein cholesterol
MD: mean difference
MeSH: Medical Subject Headings
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analyses
RCT: randomized controlled trial
SBP: systolic blood pressure
T2DM: type 2 diabetes mellitus

Edited by L Buis; submitted 23.Dec.2024; peer-reviewed by AN Ali, S Mukherjee; comments to author 24.Jun.2025; revised version received 16.Aug.2025; accepted 04.Sep.2025; published 18.Feb.2026.

Please cite as:
Jiang S, Gao X, Diao H, Zhang Y, Lu G, Liu X, Li Y
Clinical Improvements From Telemedicine Interventions for Managing Type 2 Diabetes Compared With Usual Care: Systematic Review, Meta-Analysis, and Meta-Regression
JMIR Mhealth Uhealth 2026;14:e70429
URL: https://mhealth.jmir.org/2026/1/e70429
doi:10.2196/70429
PMID:

©Shujie Jiang, Xianru Gao, Haiqing Diao, Yang Zhang, Guangyu Lu, Xiaoguang Liu, Yuping Li. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 18.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Telemedicine App for Nonrigid Facial Rehabilitation Training Enhanced by Efficient Fully Convolutional Neural Network With Residual Network (EffiFCNN-ResNet) to Improve Accessibility for Patients With Nasopharyngeal Carcinoma Cancer: Randomized Controlled Trial

Tong Wu[1], PhD; Ting Han[1,2,3], Prof Dr; Xiaoju Zhang[4,5], PhD; Yumei Dai[4,5], MSN; Xiaoyan Meng[4,5], BSN

[1]School of Design, Shanghai Jiao Tong University, Shanghai, China

[2]Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai, China

[3]Innovation Center of Yangtze River Delta, Zhejiang University, Jiaxing, China

[4]Department of Nursing, Fudan University Shanghai Cancer Center, Shanghai, China

[5]Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China

Corresponding Author:
Ting Han, Prof Dr
School of Design
Shanghai Jiao Tong University
800 Dongchuan Road
Minhang District
Shanghai, 200240
China
Phone: 86 13916099300
Email: hanting@sjtu.edu.cn

Abstract

Background: Resource limitations in public hospitals may hinder timely monitoring and management of rehabilitation in patients with nasopharyngeal carcinoma (NPC) after radiotherapy.

Objective: This study developed and evaluated the telemedicine app "Open Care," which integrates the Efficient Fully Convolutional Neural Network with Residual Network (EffiFCNN-ResNet) model and computer vision to monitor facial training exercises and provide real-time feedback, aiming to improve outcomes in patients with restricted mouth opening.

Methods: Initially, the EffiFCNN-ResNet model underwent 5-fold cross-validation, expert validation, and robustness testing to assess its reliability and clinical applicability in complex real-world environments. Subsequently, to evaluate the telemedicine app, a parallel-group, 2-arm randomized controlled trial was conducted with 109 patients, who were randomly assigned to either the intervention group (n=55) or the control group (n=54). The intervention group performed mouth-opening exercises under the supervision and guidance of the telemedicine app, whereas the control group followed traditional video-based instructions. Primary outcome measures included maximum mouth opening, mouth-opening symmetry, exercise frequency, and rehabilitation-related health beliefs. Secondary outcomes included fatigue (Brief Fatigue Inventory), health-related quality of life (Assessment of Quality of Life—6 Dimensions), and system usability scores. Data were analyzed using 2-tailed (unpaired) independent-samples *t* tests and chi-square tests, and the Mann-Whitney U test was used to assess intra- and inter-group differences before and after the intervention.

Results: The "Open Care" system leverages a lightweight fully convolutional neural network (FCNN) depth model integrated with network communication to enable real-time capture, recognition, and correction of nonrigid facial training movements. It also provides visual feedback and supports automated rehabilitation assessment. The model demonstrated strong generalization ability (macro-averaged F1-score, mean 0.96, SD 0.01) and clinical-grade stability (performance degradation: mean 5.2%, SD 0.6%, under lighting disturbances and challenging pathological cases; n=160 video segments). Compared with the control group, the intervention group showed significant improvements in maximum mouth opening (*P*=.04), exercise frequency (*P*=.001), perceived severity (*P*=.007), perceived benefits (*P*=.04), perceived barriers (*P*=.001), self-efficacy (*P*=.04), cues to action (*P*=.001),

XSL•FO
RenderX

health behavior (*P*=.03), and fatigue (*P*=.04). Participants also reported favorable training experiences, with a mean system usability score of 74.3 out of 100.

Conclusions: This telemedicine approach was more effective than traditional methods, improving patient engagement and rehabilitation outcomes while providing a more objective and precise monitoring tool. Future apps may benefit patients with NPC and other head and neck cancers.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2400090305; https://www.chictr.org.cn/showprojEN.html?proj=235073

(JMIR Mhealth Uhealth 2026;14:e72560) doi:10.2196/72560

KEYWORDS

nonrigid facial action tracking; fully convolutional neural network; telemedicine app; behavioral intervention

Introduction

Background

Nasopharyngeal carcinoma (NPC) is a common malignant tumor of the head and neck, with a high prevalence in Eastern and Southeast Asian countries. Comprehensive treatment, primarily based on radiotherapy, is the preferred therapeutic approach for NPC [1]. Trismus is a common complication of radiotherapy [2,3], with an incidence ranging from 12.0% to 58.5% [4]. This complication is primarily caused by radiotherapy-induced fibrosis of the masticatory muscles and temporomandibular joint, muscle atrophy, and neural injury. It may also act synergistically with dysphagia, posing long-term challenges to patients' nutritional intake and psychosocial well-being [5] and significantly affecting patients' quality of life [6]. Current research on mouth-opening training and intervention for patients with NPC primarily focuses on the content, timing, methods, and tools of such exercises. Traditional clinical interventions commonly employ cork resistance training; however, patients often report significant pain perception [7]. To address this issue, a recent literature review [8] proposed a novel therapy combining hardware-assisted passive mouth-opening traction with transcutaneous neuromuscular electrical stimulation. However, contact-based training may exacerbate symptoms in patients with radiation-induced dermatitis or skin fibrosis [9]. Therefore, there is an urgent need to explore alternative approaches, particularly noncontact, visually interactive supportive interventions for patients with NPC after radiotherapy, to improve prognosis. Studies have demonstrated that early implementation of active mouth-opening exercises can effectively prevent the onset of trismus [10,11]. Moreover, Samarah et al [12] confirmed that targeted mouth-opening training interventions play a significant role in reducing the incidence of trismus, improving patient adherence, and enhancing quality of life in patients with NPC undergoing radiotherapy and chemotherapy [13]. Despite the proven efficacy of mouth-opening exercises, long-term treatment outcomes remain inconsistent [14], likely due to the high health care burden associated with training, movement correction, and supervision [15].

Digital health interventions can provide scalable health education and self-management support for patients with NPC through apps or web-based platforms [16,17], offering high coverage, low cost, and easy accessibility to mitigate constraints related to time and resources [18-21]. However, the effectiveness of some digital interventions varies. For instance, although physiotherapy programs and online apps have demonstrated efficacy in improving physical function [22], their long-term effectiveness declines due to a lack of precise, individualized support [23]. Additionally, a physical activity counseling program based on wearable devices reported no prognostic benefits, as it lacked accurate monitoring and feedback on training movements [24,25]. This limitation reduces the effectiveness of telemedicine apps in facilitating patient rehabilitation behaviors. A meta-analysis further highlighted that precise movement monitoring and targeted feedback are critical factors in enhancing the effectiveness of digital health interventions for patient health management. Therefore, future digital health interventions should prioritize advancements in personalized, intelligent, real-time monitoring and feedback technologies to enhance intervention efficacy [26,27].

In clinical practice, evaluation of jaw training in patients with limited mouth opening mainly relies on direct observation and manual measurements [28], which are limited in objectivity and real-time applicability. Recently, machine learning has shown promise in intelligent assessment and quantitative rehabilitation [29], integrating neural networks with medical imaging, treatment planning, patient simulation, quality assurance, and radiation dose delivery to provide objective monitoring in clinical care [30]. However, patients with NPC undergoing radiotherapy may experience skin fibrosis, inflammation, swelling, and mandibular deformities, which complicate facial motion analysis [31]. Additionally, factors such as head movement and lighting variations exacerbate nonrigid facial deformations, making accurate motion tracking difficult [32]. Existing methods that rely on contact-based markers are prone to causing facial trauma and patient discomfort and provide insufficient real-time tracking accuracy. Thus, developing a deep learning–based, noncontact, real-time monitoring system is crucial for enhancing NPC rehabilitation [33,34]. Recent advancements in artificial intelligence have enabled machines to automatically analyze and interpret complex data, supporting more personalized treatment strategies. For instance, He et al [35] explored dense facial tracking sequences using convolutional neural network (CNN) networks combined with nonrigid iterative closest point algorithms to address frame-to-frame relationships in 3D facial analysis. Durga et al [36] employed a hybrid deep learning algorithm, including MtCNN (multitask cascaded convolutional neural network) and DeepFace, to overcome nonrigid facial changes, such as

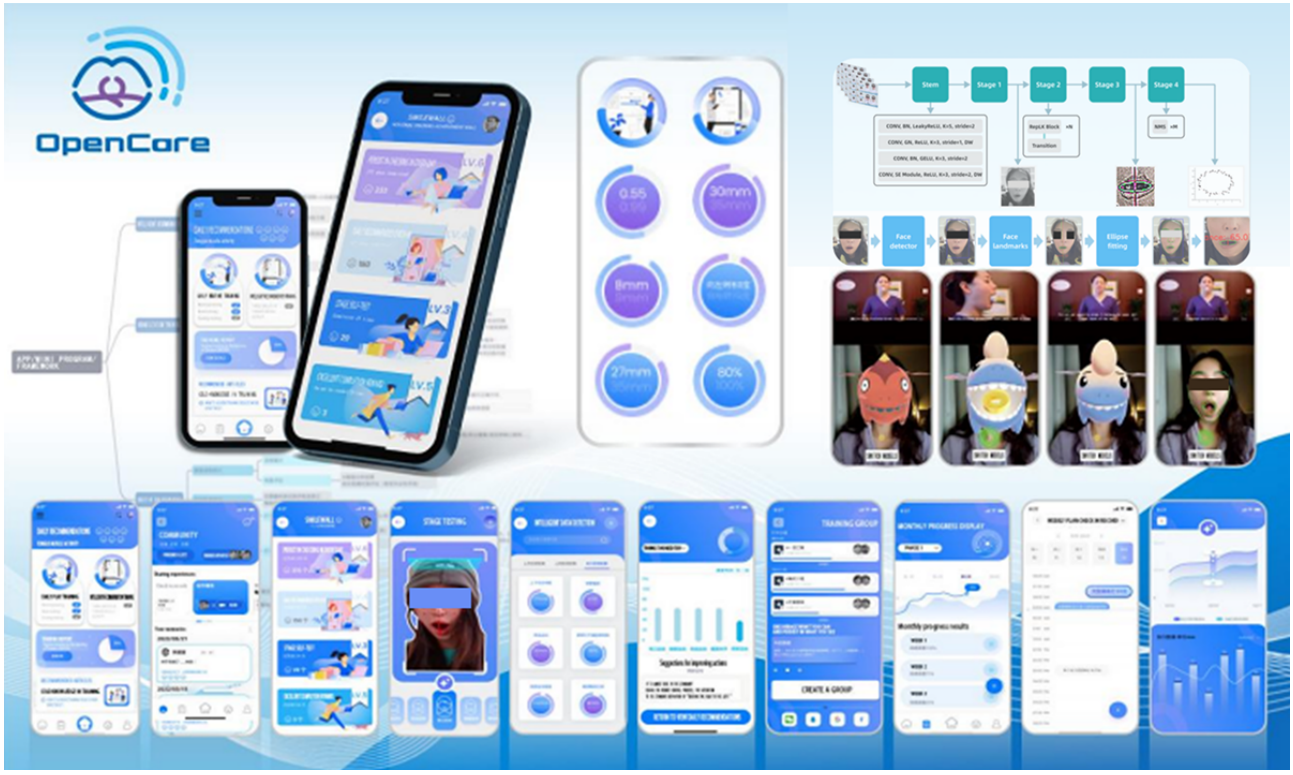
XSL•FO
RenderX

variations in size, shape, and color. The authors proposed the Tiefes fully convolutional neural network (FCNN) model, achieving high accuracy in microexpression recognition for nonrigid facial movements, demonstrating the feasibility of CNN-based deep networks. Relatedly, Le et al [37] applied deep learning for human segmentation and tracking, whereas Singh et al [38] used a deep network to reduce the joint image–motion parameter search to a search limited to rigid motion parameters.

Objectives

To address the challenges in rehabilitation training for patients with NPC, we developed Open Care, a remote health care app designed specifically for NPC jaw rehabilitation (Figure 1). By utilizing an Efficient Fully Convolutional Neural Network with Residual Network (EffiFCNN-ResNet) model, Open Care enables noncontact, real-time facial motion tracking, measurement, and automated feedback. This system offers a novel solution for personalized and efficient NPC rehabilitation. A randomized controlled trial was conducted to evaluate its effectiveness in improving patient adherence and rehabilitation outcomes.

Figure 1. The overall design of Open Care.



Methods

Tasks and Process: Involvement

All participants, regardless of group assignment, received initial standardized, equally timed exercise instruction from researchers to ensure consistent foundational guidance. Researchers provided detailed instructions on the key components of the training video and informed participants of the benefits of regular mouth-opening exercises.

The core component of standard care for the control group consisted of routine mouth-opening exercise guidance delivered through a 15-minute educational video, meticulously designed by radiation oncology professionals and validated by a multidisciplinary expert team. The video provided step-by-step demonstrations of active mouth-opening exercises, and participants were instructed to perform them twice daily, consistently within a pain-free range. An ancillary app was provided to the control group solely for video playback and for recording daily maximum mouth opening, symmetry, and training frequency.

Conversely, the experimental group used the OpenCare health app, which, while integrating the identical instructional videos, was augmented with an EffiFCNN-ResNet model. This advanced functionality enabled real-time facial movement recognition, monitoring, and automated feedback, providing users with intuitive visualizations of exercise completion, rehabilitation progress, and maximum mouth-opening values through graphical representations. The OpenCare app further incorporated structured, gamified guidance, featuring an interactive module that displayed both standard demonstrations and the participant's live facial video for comparison. It provided immediate feedback, verbal cues for insufficient movement (eg, "Please try to increase the range of motion, hold longer"), and visual reinforcement, such as completion percentages and gamified animations. Beyond these advanced functionalities, OpenCare's video materials and data-recording features were strictly identical to those used by the control group.

To clearly delineate the intervention's impact, participants in both groups received weekly phone follow-ups of consistent content and duration to assess progress and address questions.

XSL•FO
RenderX

Crucially, participants in both groups were strictly prohibited from engaging in any additional training beyond the prescribed regimen, thereby minimizing confounding factors.

Procedure

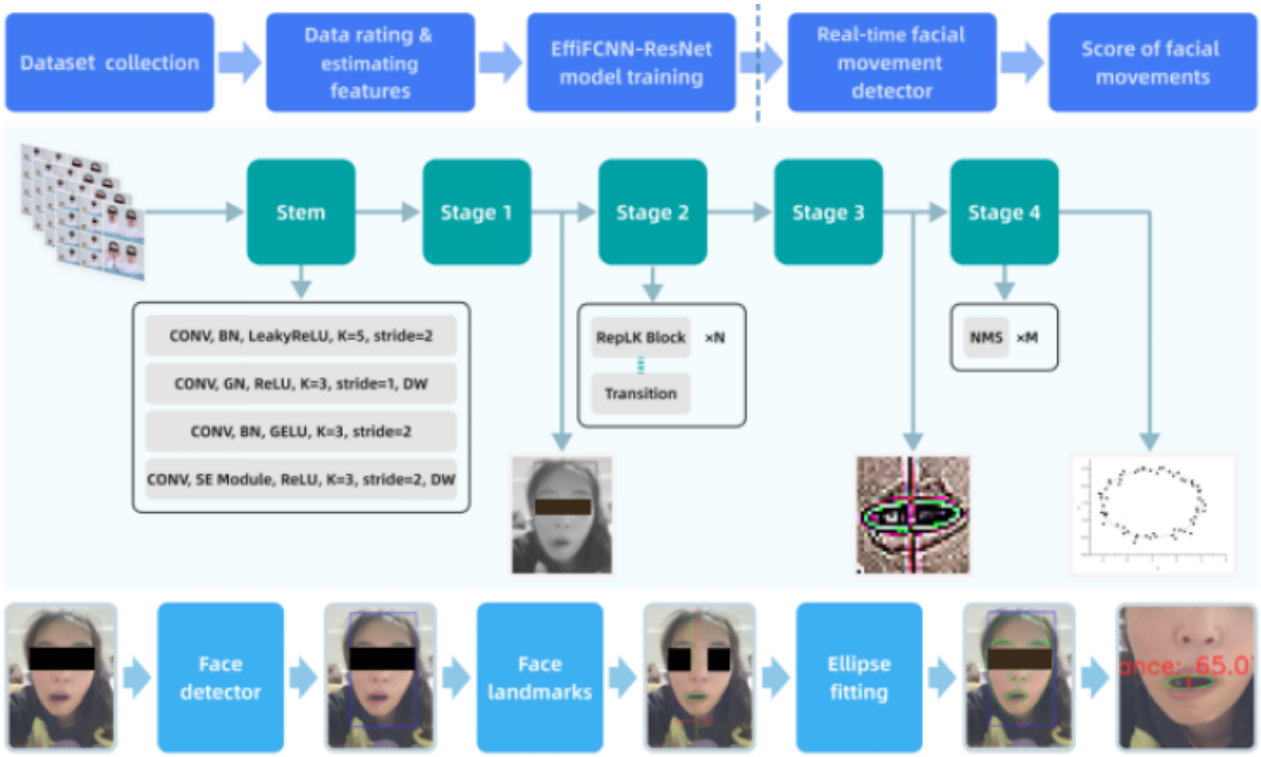
Overview of the EffiFCNN-ResNet-Based System for Precise Training Monitoring and Assessment

We developed the EffiFCNN-ResNet architecture to address the complex task of recognizing a large number of nonrigid facial landmarks [39-41]. Typically, increasing network depth or input image tensors is a common approach to overcoming challenges posed by strong environmental conditions and substantial facial structural changes during facial landmark monitoring [36]. However, this strategy often increases computational complexity and execution time. Our backbone network leverages the inherent advantages of FCNN in spatial data processing and incorporates a lightweight EffiResNet backbone designed for real-time applications requiring rapid and accurate facial detection and keypoint localization [42]. The underlying principle is rooted in a single-stage object detection framework, integrating a carefully designed detector with a lightweight backbone network to achieve fast and accurate facial recognition [40], as shown in Figure 2. The advantages and contributions of our model are summarized as follows:

- Custom training on the large-scale public movie dataset Acted Facial Expressions in the Wild (AFEW) using FCNN to detect facial landmarks in complex scene settings [43].
- Replacing the FCNN backbone with a rescaled EffiResNet backbone and utilizing large-kernel depthwise convolutions to expand the receptive field of the output feature map [44].
- Integrating the backbone network with channel separation and multinetwork landmark detection modules, followed by a linear spatial channel attention module combined with nonmaximum suppression to further optimize and filter features [45].

Specifically, our model consists of the components listed in the following subsections.

Figure 2. Architecture overview of the EffiFCNN-ResNet model for real-time facial landmark detection. BN: batch normalization; DW: depthwise; GELU: Gaussian error linear unit; EffiFCNN-ResNet: Efficient Fully Convolutional Neural Network with Residual Network; GN: group normalization; LK: large kernel; NMS: nonmaximum suppression; ReLU: rectified linear unit.



Backbone Network

A lightweight backbone network based on ResNet was used to extract salient features from input images. Its core component is the residual block, which addresses the vanishing gradient problem in deep networks by introducing skip connections, thereby improving training efficiency and overall model performance. The ResNet architecture demonstrates strong hierarchical feature extraction capability, enabling efficient capture of multiscale features from low to high levels. Designed according to the structural reparameterization principle, the backbone network consists of multiple residual blocks, each incorporating convolutional layers, batch normalization, and rectified linear unit activation functions. Skip connections allow each layer to directly receive input features from the preceding layer, enhancing training stability. The primary objective is to decouple the multibranch topology used during training from the simplified structure applied during inference. This backbone design achieves a balance between computational efficiency and feature expressiveness by separating the multibranch topology at training time from the streamlined architecture used at inference time.

XSL•FO
RenderX

Neck Network

The model integrates the classic spatial pyramid pooling (SPP) module and an improved path aggregation network (PAN), enhancing feature fusion capability and multiscale perception. The SPP layer performs multiscale pooling operations to integrate features across different scales, overcoming the limitation of fixed input image sizes and enabling the capture of multiscale contextual information to improve target region perception. After the SPP module, the feature map is transformed into a fixed-length output, providing richer contextual information for the subsequent PAN. The improved PAN uses a bottom-up routing fusion strategy, effectively reducing information loss during the transfer of deep-layer features to shallow layers. Adaptive feature pooling dynamically integrates feature maps, ensuring that information from different levels is fused across all feature scales. As a result, the features passed to the detection head contain comprehensive and semantically rich information, contributing to improved target detection accuracy.

Anchor Generation Module

This module generates dense anchors (proposed bounding boxes) at different scales and aspect ratios to detect nonrigid facial structures of varying sizes and shapes within the image. It is responsible for merging semantic features from deep layers with texture features from shallow layers and relies on an efficient decoupled head with an appropriate loss function.

Face Detection Head and Keypoint Localization Head

The detection head performs classification and regression of candidate boxes to optimize their positions and sizes. Meanwhile, the keypoint localization head accurately identifies facial landmarks, enhancing the algorithm's applicability in tasks requiring detailed facial analysis while maintaining a balance between inference time and accuracy.

Dilated Convolution and Basic Block

These components are used to expand the receptive field during feature extraction while preserving spatial resolution.

Multiscale Detection

A key feature of the model is its multiscale detection strategy, which utilizes anchors generated on different feature maps to accommodate facial structures of varying sizes.

Concatenate (C) and Upsampling (U)

By combining low- and high-level features through feature fusion and upsampling, the model integrates features from different resolutions, ensuring robustness across a wide range of facial scales. This capability is critical for real-world deployment scenarios.

Joint Loss Function and FCOS Head

As the final detection head, the fully convolutional one-stage object detection (FCOS) head combines focal loss and generalized intersection over union (GIoU) loss to achieve efficient object classification and bounding box regression. The objective of the joint loss function is to simultaneously minimize classification error (facial region classification) and regression error (bounding box regression and keypoint localization), thereby optimizing network learning and improving the accuracy of both facial detection and keypoint localization. During deployment, the predicted probability labels for each sample are input to compute the focal loss, which is then averaged across all samples as the classification loss. GIoU is calculated to evaluate the intersection over union (IoU) between the predicted and ground truth bounding boxes. Subsequently, keypoint bounding box errors are removed, squared errors between predicted and true coordinates are calculated, and the results are averaged to obtain the keypoint localization loss. Self-monitoring of both losses constitutes a self-distillation technique, in which knowledge is transferred from the teacher model to the student model during training. The overall loss can be explained as follows:



where $-\alpha_t(1-p_t)^{\gamma}\log(p_t)$ represents the action recognition loss, $1-\mathrm{GIoU}(B_p, B_g)$ represents the facial region localization loss, and  represents the anatomical keypoint precision loss. λ_1, λ_2, and λ_3 are the corresponding adjustable weight coefficients for each term, used to balance the contribution of different tasks to the total loss, thereby adapting to varying optimization priorities during the rehabilitation training process. The term λ_1 is designed based on an improved focal loss, primarily addressing the inherent challenge of severe class imbalance in rehabilitation training data, specifically between positive and negative samples. The modulating factor γ strengthens the weight assigned to hard-to-classify samples. Specifically, the $(1-p_t)^{\gamma}$ in the formula markedly amplifies the loss contribution of low-confidence samples (ie, when the predicted probability p_t for the correct class approaches 0), while conversely reducing the contribution of easily classified examples (where p_t approaches 1). Furthermore, α_t serving as a class-balancing factor mitigates training bias arising from the disparity in the number of mouth-opening action frames versus nonaction frames. The term λ_2 employs the GIoU loss, replacing the traditional IoU loss, with the aim of significantly enhancing the bounding box regression accuracy for dynamic facial regions, particularly the perioral area. The calculation of GIoU requires the simultaneous input of the coordinate parameters for both the predicted bounding box B_p and the ground truth bounding box B_g, as detailed in equation 2:

$$\mathrm{GLoU} = \mathrm{LoU} - [|C\backslash(B_p \cup B_g)|]/|C| \quad \textbf{(2)}$$

Even in cases where the predicted and ground-truth bounding boxes are nonoverlapping, the GIoU loss can provide informative gradients by penalizing the difference between the smallest enclosing convex object (C) and the $B_p \cup B_g$. For instance, when minor rotations of a patient's head or variations in camera angle cause the predicted bounding box for the oral region to deviate from the ground-truth box without intersecting, GIoU still yields a directional gradient. This characteristic enhances the model's robustness to changes in camera pose, distance, and subtle patient head movements. The term λ_3 directly constrains the spatial deviation of facial anatomical keypoints through the mean squared error (MSE) formulation,

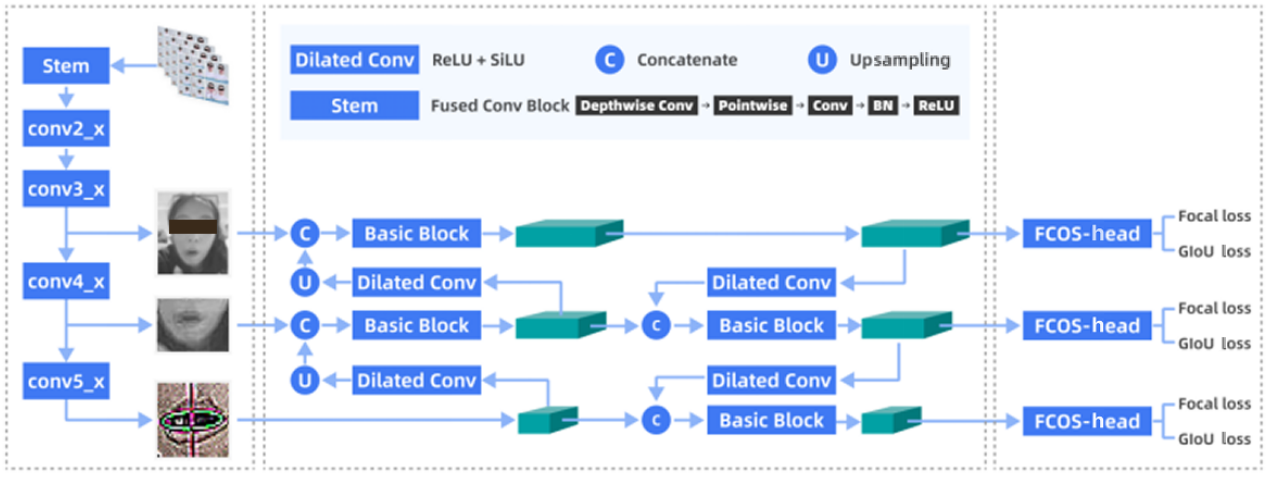
XSL•FO
RenderX

ensuring high-precision landmark localization. Specifically, for the *i*th keypoint, the squared Euclidean distance between the predicted coordinate vector P_i and the expert-annotated ground truth coordinate vector G_i is computed, with the squared Euclidean distance calculated as  for each keypoint. Subsequently, the error is averaged over N keypoints using the formula . This process ensures the model's simultaneous and consistent optimization of all annotated points, including the upper lip, lower lip, and other crucial landmarks. This loss term directly prompts the model to precisely optimize the spatial localization consistency of all annotated points, including the upper lip, lower lip, and mouth corners.

Real-Time Capture

To enable rapid inference, the model employs multiscale preprocessing and multiscale detection, supplemented by optimization techniques such as nonmaximum suppression, making it suitable for real-time applications. Through the architecture described above, FCNN preserves spatial relationships across the entire network, enabling the generation of dense prediction maps. The lightweight ResNet extends convolutional layers throughout the network, facilitating direct correspondence between input pixel positions and their feature representations. Compared with traditional landmark detection methods, this approach captures a greater number of facial landmarks with higher accuracy while mitigating disturbances in texture and shape caused by nonrigid facial deformations, as illustrated in Figure 3. For rapid inference, the model adopts a lightweight ResNet-based object detection network enhanced with dilated convolutions. The core architecture uses multilevel dilated convolutions to expand receptive fields and capture contextual information, while integrating a stem module with lightweight pointwise convolutions and hybrid sigmoid linear unit/rectified linear unit activations to reduce computational complexity. Cross-level feature fusion is achieved through concatenation and upsampling operations to enhance detail retention. The FCOS head addresses class imbalance using focal loss and optimizes bounding box regression with GIoU loss, making it suitable for dense scenarios and small-object detection. Overall, the unified framework emphasizes 3 critical aspects: receptive field expansion through hierarchical dilation, multiscale feature interaction, and lightweight deployment capability, thereby balancing detection accuracy with computational efficiency.

Figure 3. Real-time facial landmark detection with multiscale preprocessing and nonmaximum suppression optimization. BN: batch normalization; FCOS: fully convolutional one-stage object detection; GIoU: generalized intersection over union; ReLU: rectified linear unit; SiLU: sigmoid linear unit.

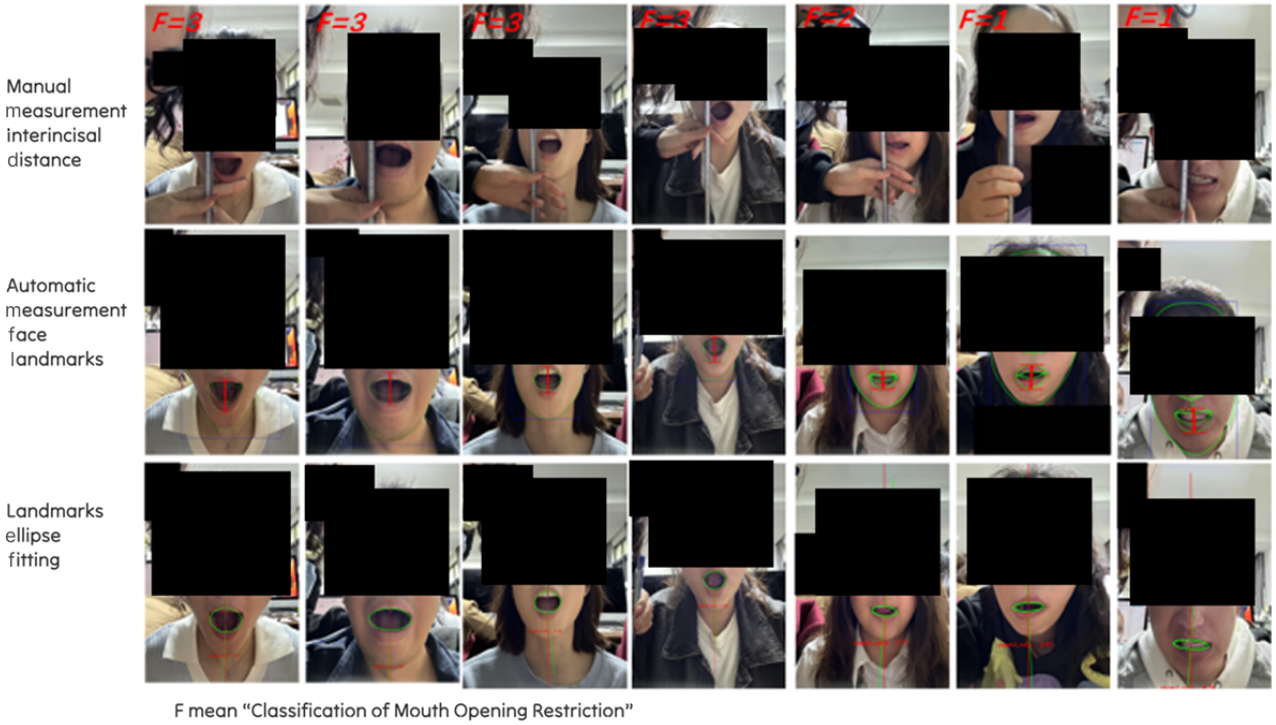


Training Action Classification and Evaluation

Taking the mouth-opening action as an example, a more precise evaluation of movement is achieved by introducing a 4-level grading system that incorporates both the degree of mouth opening and its symmetry. This grading system simultaneously considers 2 key aspects: maximum mouth opening and mouth shape symmetry. Maximum mouth opening is defined as the vertical distance between the highest point of the upper semicircle and the lowest point of the lower semicircle, derived from the elliptical fitting of the mouth region. Symmetry is assessed based on the rotation angle of the fitted ellipse. To compute the elliptical equation of the oral region, we use a facial landmark detection and oral-region keypoint elliptical fitting algorithm based on the aforementioned model, utilizing more than 10,000 keypoints. This algorithm provides intelligent, automated evaluation of mouth-opening performance for each patient, as illustrated in Figure 4.

XSL•FO
RenderX

Figure 4. Action evaluation output results for a subset of the patient dataset.



Through these steps, the keypoints obtained using the deep learning–based facial landmark localization algorithm can be further analyzed to assess the morphological characteristics of the oral region, thereby quantifying the degree of mouth opening and symmetry. The elliptical equation of the oral region can be expressed in the following general form:

$$(x - x_0)^2/a^2 + (y - y_0)^2/b^2 \textbf{(3)}$$

where (x_0, y_0) are the coordinates of the center of the ellipse, and a and b denote the lengths of the major and minor axes, respectively. Subsequently, the vertical distance between the highest point (u) of the upper semicircle and the lowest point (d) of the lower semicircle is calculated. The grading criteria were defined by oncologic medical experts. F4 indicates the normal range (35-50 mm), F3 represents mild restriction (25-35 mm), F2 represents moderate restriction (15-25 mm), and F1 represents severe restriction (<15 mm).

At the same time, the rotation angle of the ellipse is used to assess mouth symmetry. The rotation angle of the ellipse can be calculated using the following formula.

$$\theta=\arctan(b\sin[\Phi]/a\cos[\Phi]) \textbf{(4)}$$

where Φ represents the inclination angle of the ellipse. The grading criteria for symmetry are defined as follows: F4, rotation angle <5° (good symmetry); F3, rotation angle 5°-10° (mild asymmetry); F2, rotation angle 10°-15° (moderate asymmetry); and F1, rotation angle >15° (severe asymmetry).

To correct for variations in the distance between the user and the camera, the 2 feature points at the inner corners of the eyes are connected to form a reference line. A standard value of 3.5 cm is defined, and a real-time scaling factor is derived by calculating the ratio between this reference distance and the actual measured value to perform distance correction, thereby facilitating subsequent data analysis. The corresponding code is shown below.

"# Input: Facial landmarks detected via EffiFCNN-ResNet

leftEye = (landmarks.part(39).x, landmarks.part(39).y) # Left eye corner (point 39)

rightEye = (landmarks.part(42).x, landmarks.part(42).y) # Right eye corner (point 42)

Calculate Euclidean distance between eyes

eyeDistance = np.linalg.norm(np.array(leftEye) - np.array(rightEye)) # $||p_left - p_right||_2$

Compute scale factor using reference eye distance

scale = 3.5 / eyeDistance # d_reference = 3.5 (anthropometric mean)"

Intelligent recognition and evaluation of other training actions follow a similar approach.

Dataset

General Facial Representation Pretraining Phase

This phase used the AFEW dataset for training [46]. The AFEW dataset comprises 1200 video segments from real movies and clips, from which approximately 48,000 representative facial image frames were extracted. These frames include a wide range of individuals displaying diverse expressions under varying lighting conditions, poses, and background environments. The data modality consisted of RGB (red-green-blue) video frames and their corresponding emotional category labels. Optimization was performed using the AdamW optimizer with an initial learning rate of 1×10^{-4}. A cosine annealing learning rate schedule, coupled with a warm-up phase, was applied. The model was trained for 80 epochs with a batch size of 32,

XSL•FO
RenderX

incorporating gradient accumulation and various data augmentation techniques. Hyperparameters were determined through a combination of grid search and Bayesian optimization. This foundational training aimed to enhance the model's generalization capability and prepare it for subsequent fine-tuning, particularly to achieve robust facial feature extraction in patient-specific scenarios.

Dataset Training Phase

The formal training dataset used in this study was derived from video data of real patients with NPC undergoing postradiotherapy rehabilitation training. The dataset comprised 800 video segments, from which approximately 12,000 keyframe images were extracted. Together with their corresponding annotations, this process generated 85,000 image-annotation pairs and 65,000 video-annotation pairs. To ensure robustness and generalizability of the developed model (EffiFCNN) in real-world clinical applications, the dataset was designed with a strong emphasis on diversity. Environmental diversity was achieved through simulated or real recordings of patients performing exercises at home using smartphones or tablets. These recordings encompassed a wide range of lighting conditions (eg, strong light, low light, side light), background complexities, camera angles (eg, frontal, top-down, bottom-up, side view), and shooting distances, thereby enhancing adaptability to complex daily environments. In terms of action content diversity, the dataset focused not only on the core mouth-opening rehabilitation exercise, comprehensively documenting maximal active mouth opening, sustained holding, and relaxation, but also included samples reflecting different degrees of mouth opening classified according to clinical F1-F4 grades and varying levels of symmetry quantified by the elliptical rotation angle θ. In addition, auxiliary rehabilitation exercises, such as neck movements, ocular movements, and cheek puffing (quantification methods are detailed in Multimedia Appendix 1), were incorporated to support a comprehensive rehabilitation assessment. Patient state diversity was a central characteristic of the dataset, with particular attention to challenges unique to postradiotherapy patients with NPC. The dataset covered all levels of mouth-opening restriction, ranging from severe restriction (F1, <15 mm) to normal function (F4, 35-50 mm). Special emphasis was placed on collecting samples from patients with moderate to severe restrictions (F1-F3), as this population represents the primary target group for intervention. The dataset also included samples exhibiting common postradiotherapy sequelae, such as skin hyperpigmentation, fibrosis, edema, facial asymmetry, and subtle jaw deformities, all of which pose significant challenges for nonrigid facial deformation analysis. To address class imbalance, the synthetic minority over-sampling technique was applied to augment minority class samples within the feature space, thereby improving recognition performance for rare categories. Within ethical and privacy constraints, efforts were made to include patients across diverse age groups, genders, and body types to enhance model generalizability. During data preprocessing, all patient-identifiable information was deidentified (Figure 4), and strict adherence to ethical guidelines was maintained, with approval obtained from the Institutional Review Board (approval number 2404294-Exp8). Finally, the dataset was partitioned into training, validation, and test sets (70%:15%:15%) using stratified sampling based on key clinical features. This approach ensured sample independence and distributional consistency, providing a robust foundation for model training and evaluation.

Despite concerted efforts to ensure diversity in its design, the dataset remains susceptible to potential biases. These are explained in the following sections.

Sampling Bias

This bias may arise from the participant recruitment process, which primarily involved voluntary patients capable of using smart devices. Such an approach may lead to the underrepresentation of patients who are critically ill, severely mobility-impaired, digitally illiterate, or socioeconomically disadvantaged. In addition, the concentration of data sources within specific hospitals or regions may introduce bias related to variations in medical practice patterns.

Annotation Bias

Annotation bias may result from the inherent subjectivity in clinical experts' assessments of mouth-opening grades (F1-F4) and symmetry (rotation angle θ). Moreover, precise keypoint annotation becomes substantially more challenging in patients with severe postradiotherapy facial sequelae, such as pronounced edema or deformity.

Temporal Bias

This type of bias reflects variations associated with different postradiotherapy recovery phases (eg, acute, subacute, chronic) or heterogeneous baseline conditions, such as initial mouth-opening capacity.

Validation of Precision

To ensure rigorous and objective evaluation of the proposed EffiFCNN-ResNet model, we implemented a comprehensive validation strategy combining stratified 5-fold cross-validation with an independent test set. The dataset, comprising 800 video segments from patients with NPC, was first stratified by patient ID and mouth-opening restriction grade (F1-F4) into 560 (70%) segments for training, 120 (15%) segments for validation, and 120 (15%) segments for testing. The cross-validation procedure employed Scikit-learn's StratifiedGroupKFold method to ensure that all 448 (80% of the training set) segments were used for model training in each fold, while 112 (20%) segments served as validation data. This approach prevented patient-level data leakage across folds and maintained balanced-grade distributions throughout the validation process.

The model demonstrated consistent performance across all validation metrics. In classification tasks, it achieved a mean accuracy of 89.2% (SD 2.4%; 95% CI 86.8%-91.6%) for mouth-opening restriction grades (as detailed in equation 5), with a macro-averaged F_1-score of 0.875 (SD 0.031) and an area under the receiver operating characteristic curve of 0.934 (SD 0.018). For regression tasks, the mean absolute error for mouth-opening distance was 1.23 mm (SD 0.15 mm; 95% CI 1.08-1.38 mm), as detailed in equation 6, while the rotation angle θ showed a mean absolute error of 2.15° (SD 0.32°; 95%

XSL•FO
RenderX

CI 1.83°-2.47°). Keypoint detection achieved a normalized mean error of 0.043 (SD 0.006), with a failure rate of 3.8% (SD 1.2%; 95% CI 2.6%-5.0%). On the independent test set comprising 120 video segments from 45 patients, the model achieved an accuracy of 98.3% (118/120 correctly classified segments; 95% CI 97.6%-98.8%), with particularly strong performance on severely restricted F1-grade samples (96.4% recall; 27/28 cases correctly identified). In robustness testing involving patients with significant postradiotherapy sequelae, the model maintained 94.4% accuracy (151/160 challenging samples correctly classified). Expert annotations established a reliable gold standard with excellent interrater reliability (intraclass correlation coefficient 0.942; 95% CI 0.925-0.956) among 20 clinical oncologists. The Bland-Altman analysis demonstrated excellent agreement with expert measurements (mean difference −0.14 mm; 95% limits of agreement −1.86 to 1.58 mm), and the intraclass correlation coefficient reached 0.968 (95% CI 0.953-0.979), confirming the model's clinical reliability and readiness for deployment in diverse rehabilitation settings.

$$\text{Accuracy} = (TP + TN)/(TP + TN + FP + FN) \quad \textbf{(5)}$$

where TP is true positive, which represents the number of actual positive samples correctly predicted as positive; TN is true negative, which represents the number of actual negative samples correctly predicted as negative; FP is false positive, which represents the number of negative samples incorrectly predicted as positive; and FN is false negative, which represents the number of positive samples incorrectly predicted as negative.



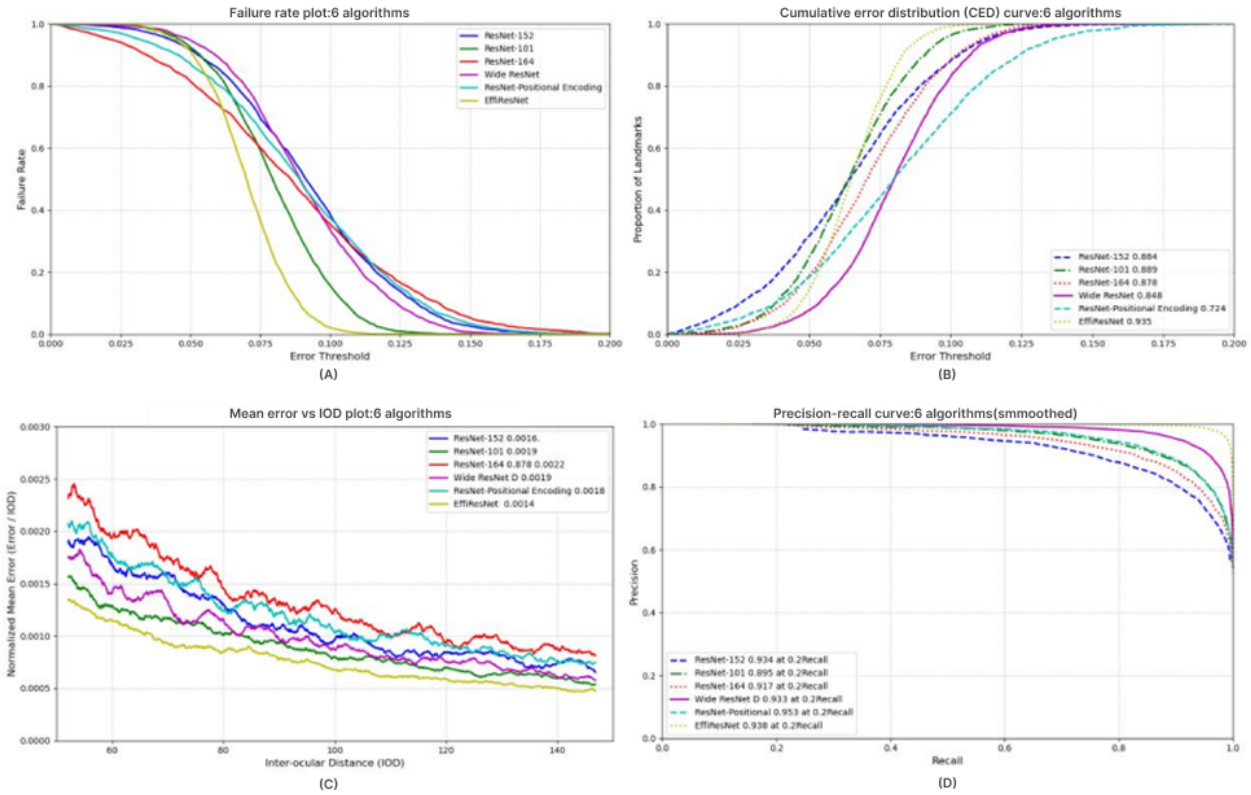
where *C* denotes the number of classes (F1-F4), emphasizing the model's balanced evaluation of performance across all individual categories.

To rigorously evaluate the clinical applicability of the EffiFCNN-ResNet system, comprehensive robustness testing was conducted under simulated challenging conditions encompassing both environmental and patient-specific variations. Environmental perturbations were introduced through structured image transformations, including +30% or –30% linear and gamma-adjusted illumination shifts in the hue-saturation-value color space and affine transformations simulating +15° or –15° viewpoint deviations in pitch, yaw, and roll angles. A dedicated subset containing 160 video samples (20% of the test set; N=800), consisting exclusively of postradiotherapy patients with NPC exhibiting pronounced sequelae such as skin fibrosis, edema, and facial asymmetry, was used to assess performance under clinically relevant morphologic distortions. The model demonstrated exceptional stability, with an overall performance degradation of 5.2% (SD 0.6%) under combined disturbances. Specifically, there was a 0.007 increase in normalized mean error for landmark detection (SD 0.002), a 0.4 mm increase in mouth-opening distance mean absolute error (SD 0.1), and 94.5% accuracy retention across F1-F3 restriction grades (SD 2.1%). These findings confirm the model's ability to maintain high accuracy and reliability under real-world variability, underscoring its suitability for deployment in clinical and home-based rehabilitation settings.

In the final validation stage, the superiority of the proposed method was objectively evaluated through systematic comparative experiments against multiple ResNet architecture variants, including ResNet-101, ResNet-152, ResNet-164, Wide ResNet, and ResNet-Positional Encoding. These comparisons were conducted on both the public AFEW dataset and a self-collected mouth-opening training dataset from patients with NPC, with standardized environments and data preprocessing to ensure fairness. Four evaluation metrics were employed: failure rate, cumulative error distribution, average error relative to interocular distance, and precision-recall curves. All models were assessed using identical hardware configurations and datasets. The performance of the proposed model is presented in Figure 5. The comparison of EffiResNet with other mainstream ResNet variants demonstrates its superior ability to balance accuracy, robustness, and computational efficiency. In terms of failure rate, EffiResNet shows a markedly faster decline as the error threshold increases compared with other models, indicating stronger robustness and adaptability when handling tasks with larger error tolerances. This characteristic enables stable performance in complex and dynamic computational environments. Regarding cumulative error distribution, the EffiResNet curve is smooth and approaches the cumulative error distribution performance of ResNet-152, suggesting stable behavior across varying error magnitudes. Compared with other ResNet variants, EffiResNet more effectively balances the influence of different error scales during accumulation, thereby ensuring overall performance stability. From a spatial awareness perspective, the relationship between average error and interocular distance demonstrates that the EffiResNet curve remains relatively stable, highlighting its advantage in spatial perception. This enables more accurate capture and analysis of spatial information in tasks involving spatial positioning. In classification tasks, EffiResNet also exhibits strong performance. Precision-recall curve comparisons indicate that, relative to ResNet-Positional Encoding, EffiResNet maintains high recall while achieving comparatively high precision. This reflects its ability to accurately identify target categories while minimizing misclassification.

XSL•FO
RenderX

Figure 5. Model performance test results. IOD: interocular distance.



Different performance indicators :The performance in computational tasks is compared and displayed through various charts that (A) encompass failure rates, (B) cumulative error distributions, (C) average error in relation to inter-ocular distance, and (D) precision-recall curves.

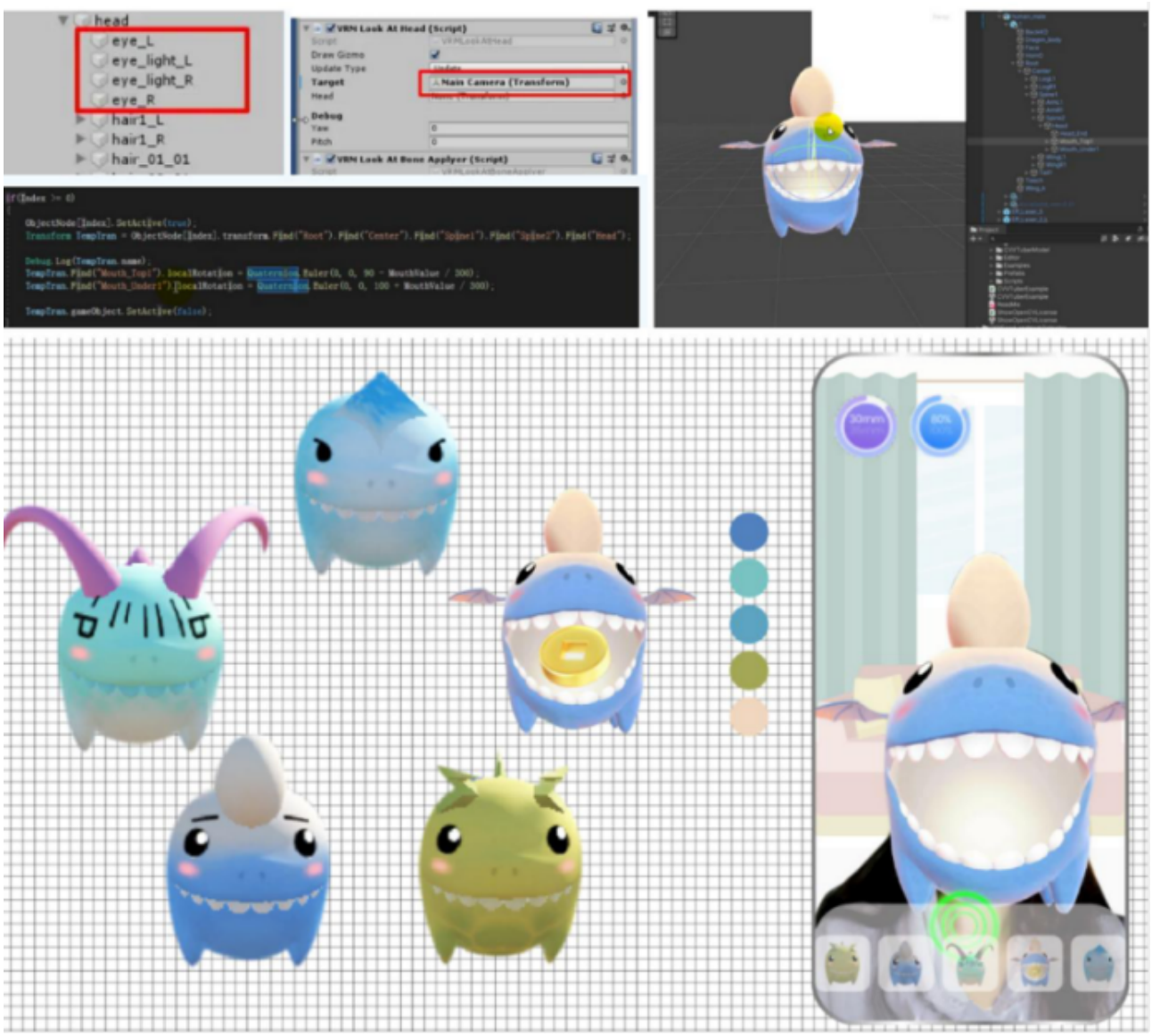
Personalized Real-Time Feedback

In designing the virtual avatar for patients with nasopharyngeal cancer, we considered potential deficiencies in the sensory integration system while avoiding the uncanny valley effect [47]. A series of brightly colored, soft, rounded dinosaur-themed virtual headgear avatars with simple structures was developed [48]. These avatars are linked to mouth-opening training movements through gamification based on skeletal keypoint capture, providing real-time visual and auditory feedback (eg, swallowing coins) when the required movement is achieved (Figure 6). As training scores accumulate, users can unlock additional virtual avatars as rewards. This design aims to stimulate patients' intrinsic motivation during training and create an immediate, supportive, and personalized rehabilitation environment. The remote mouth-opening data tracking function was developed using the Unity3D engine (Unity Technologies) and Python (Python Foundation), with communication implemented through SocketTools.cs. Upon user authorization, the mobile device camera is used for facial landmark detection, enabling the system to collect mouth-opening data over a specified period, including maximum mouth opening, opening frequency, and symmetry. These data are transmitted to the backend via transmission control protocol/IP sockets for processing, with serial communication facilitating data exchange between devices and enabling real-time monitoring of patient movements. Data transmission is secured using encryption protocols to ensure privacy protection. The collected data support long-term daily monitoring of mouth-opening rehabilitation progress and provide actionable feedback to both patients and health care professionals regarding training patterns and behavioral habits.

Furthermore, personalized feedback is provided according to each patient's functional level. Before each training session, an initial mouth-opening test is conducted to categorize patients into specific levels. Based on this classification, the performance thresholds required to complete the prescribed training tasks are individualized and progressively adjusted according to the severity of the condition, enabling a more flexible and targeted rehabilitation approach. This strategy ensures that the training protocol is tailored to the unique needs of each patient, thereby facilitating more effective recovery and sustained progress.

XSL•FO
RenderX

Figure 6. Training virtual headgear and its development program demonstration.



Individuals Involved

The study participants comprised patients with nasopharyngeal cancer who had completed radiotherapy and were preparing for discharge from the oncology department. Participants were recruited through posters and questionnaire screening.

The inclusion criteria were as follows: (1) a pathology or histologic diagnosis of nasopharyngeal cancer with a prescribed course of radiotherapy; (2) age 18-65 years; (3) awareness of their medical condition; (4) provision of informed consent and willingness to participate in the study; (5) access to a smartphone and related apps; and (6) Functional Status Score (Karnofsky performance status) of 80-100 [49] and Eastern Cooperative Oncology Group performance status of 0-5 [50].

The exclusion criteria were as follows: (1) a history of prior radiotherapy; (2) inability to participate in physical activity due to underlying cardiac, neurological, muscular, or joint disorders; and (3) previous engagement in relevant mouth-opening exercises (eg, as part of routine daily care, follow-up video guidance, or other structured rehabilitation programs).

The withdrawal criteria were defined as follows: (1) voluntary decision to discontinue participation in the trial; (2) development of a severe illness that precluded continuation of the study; (3) occurrence of an adverse event related to mouth-opening exercises or routine physical activity; and (4) requirement for additional surgical interventions, such as mandibular resection.

Following Testing

After the 4-week intervention period, all participants underwent a comprehensive assessment, including measurement of maximal mouth opening as well as other predefined outcome measures and relevant clinical indicators. All data were collected using the same measurement procedures as those employed during the baseline assessment, ensuring methodological consistency. This approach enabled reliable quantitative analysis for both within- and between-group comparisons.

Size of Sample

In our previous comparative study, the sample size was calculated based on the change in perceived benefit scores measured 1 week after the intervention. Pilot data demonstrated that the intervention group showed a mean increase of 0.8 points

XSL•FO
RenderX

(from mean 10.5, SD 2.6 to mean 11.3, SD 2.4), corresponding to a Cohen *d* effect size of 0.33. Using G*Power (Heinrich Heine University Düsseldorf) for sample size estimation, we determined that a total sample of 108 participants (54 per group) would be required to achieve 80% statistical power [51] at a 2-tailed significance level of 5% [52].

Blind Method, Randomization, and Allocation Concealing

This study rigorously employed a stratified block randomization approach to allocate eligible participants to either the experimental or control group, ensuring balanced distribution of key baseline characteristics while preserving allocation unpredictability. An independent statistician generated the randomization sequence using R (version 4.4.x; R Foundation). Stratification variables included age (<45, 45-60, and >60 years), gender (male or female), and cancer stage (stage II and below, stage III, and stage IV). These factors were selected based on their clinical relevance and prognostic significance, while avoiding excessive stratification that could result in sparse subgroup sizes [53]. To further enhance allocation concealment, varying block sizes (4 or 6) were applied [54]. Each group assignment was placed in an opaque, sealed, and consecutively numbered envelope prepared and safeguarded by the independent statistician. Envelopes were opened sequentially by unblinded study personnel only after completion of participant screening and informed consent procedures. This process ensured that recruiting staff, intervention providers, and outcome assessors remained unaware of group allocation prior to assignment, thereby effectively minimizing selection bias [55]. Additionally, a triple-blind design was implemented, whereby participants, outcome assessors, and statisticians were blinded to group allocation. This strategy reduced the risk of performance and ascertainment biases, substantially enhancing the internal validity, objectivity, and reliability of the study findings.

Documents of Safety

No adverse events were reported in association with this app during the study period. Participants were instructed that if they experienced significant respiratory distress, palpitations, or any other form of physical discomfort during the intervention, they must immediately discontinue the training session and return to a resting position. If symptoms were severe or persisted, participants were advised to seek prompt medical evaluation. All serious adverse events were reported to the institutional ethics committee in accordance with regulatory requirements, while any unexpected or nonserious adverse events were documented and reported to the research team for further evaluation and follow-up.

Outcomes Measures

Questionnaire

A validated questionnaire, administered either online or in person, was completed by each participant at baseline and again 4 weeks after the intervention. No financial compensation or other incentives were provided for participation in the study. The collection of baseline demographic data and related variables is described in detail later.

Primary Outcomes

Maximum Mouth Aperture Measurement

To evaluate the effectiveness of the intervention app on mouth-opening function, participants' maximum mouth aperture was measured and compared before and after the intervention. This comparison enabled assessment of functional improvement attributable to the training program. During measurement, participants were instructed to open their mouths as widely and steadily as possible to reach their individual maximal range. Once the maximum opening position was achieved, a 3-second stabilization period was maintained. Five data points were collected during the middle 1-second interval to minimize transient fluctuations. The average of these 5 values was calculated and recorded as the final maximum mouth-opening measurement.

Evaluation of Mouth Aperture Symmetry

To assess the impact of the intervention on oral motor function, the symmetry of participants' mouth-opening movements was documented and compared before and after the intervention. Symmetry was evaluated by analyzing bilateral movement patterns during mouth opening, allowing identification of deviations or asymmetries in mandibular motion. Changes in symmetry measurements between baseline and postintervention assessments were used to determine the effectiveness of the app in improving coordinated oral function.

Number of Exercises Per Week

Adherence to the prescribed training behavior was quantitatively evaluated by recording and comparing the participants' weekly exercise frequency before and after the intervention. The total number of completed mouth-opening training sessions per week was documented to assess behavioral implementation and engagement. Changes in weekly training frequency served as an indicator of intervention adherence and behavioral improvement.

Health Belief Model Questionnaire Inquiry Form

The Health Belief Model Questionnaire is widely used to assess health-related behaviors [56]. The version developed by social psychologists in the United States comprises 7 core constructs of the Health Belief Model: perceived susceptibility (3 items), perceived severity (7 items), perceived benefits (3 items), perceived barriers (7 items), self-efficacy (5 items), cues to action (3 items), and behavior (3 items) [57]. For all subscales except the behavior domain, responses were rated on a 5-point Likert scale ranging from "completely disagree" (0) to "completely agree" (4). Participants were asked to evaluate statements reflecting their health beliefs across the 6 domains. The behavior subscale, designed with strong psychometric properties, assesses the frequency of relevant health behaviors (eg, mouth-opening exercises) using a scale ranging from "never" (0) to "always" (4). Higher scores indicate greater engagement in the targeted health behavior.

Secondary Cognitive Outcomes

Borg Rating of Perceived Exertion Scale

The Borg Rating of Perceived Exertion Scale [58] is used to quantify self-reported physical exertion during exercise. The

XSL•FO
RenderX

scale score ranges from 6 (no exertion at all) to 20 (maximal exertion). Participants rate their perceived intensity of effort during mouth-opening training, providing a subjective measure of exercise load.

Assessment of Quality of Life—6 Dimensions

Health-related quality of life was assessed using the Assessment of Quality of Life—6 Dimensions (AQoL-6D) [59] instrument. Scores range from −0.04 to 1.00, with higher scores indicating better quality of life. This measure captures multiple domains of well-being to evaluate the broader impact of the intervention.

System Usability Scale

Perceived usability of the app was evaluated using the System Usability Scale (SUS) [60]. After completing the designated training tasks, participants completed the SUS questionnaire to obtain an overall usability score. Higher SUS scores indicate better perceived system usability and user satisfaction.

Data Analysis

Data were analyzed using SPSS (IBM Corp.). Descriptive statistics were presented as means with SDs for continuous variables, and as frequencies (n) with percentages (%) for categorical variables. Baseline characteristics of participants who completed the study were compared with those who did not using 2-tailed (unpaired) independent-samples *t* tests or chi-square tests, as appropriate. Given the distribution characteristics of the variables, the Mann-Whitney *U* test was used to evaluate differences between groups, and within-group comparisons before and after the intervention were also conducted accordingly. Comparisons were performed between the intervention and control groups to determine the effectiveness of the intervention. For all statistical analyses, inter- and intragroup differences were expressed as means with 95% CIs. The significance level (α) was set at .05, and a *P* value ≤.05 was considered statistically significant.

Ethical Considerations

This study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki and received approval from the institutional Human Research Ethics Committee (protocol number: 2404294-Exp8; reference number: H20240007I). All participants provided written informed consent before enrollment after receiving a detailed explanation of the study's objectives and procedures. They were explicitly informed that participation was entirely voluntary and that they could withdraw from the study at any time without any negative consequences. To ensure privacy and confidentiality, all collected data were fully anonymized before analysis, and no personally identifiable information was collected or retained. Additionally, no financial compensation or other incentives were provided, thereby maintaining a transparent and voluntary recruitment process.

Results

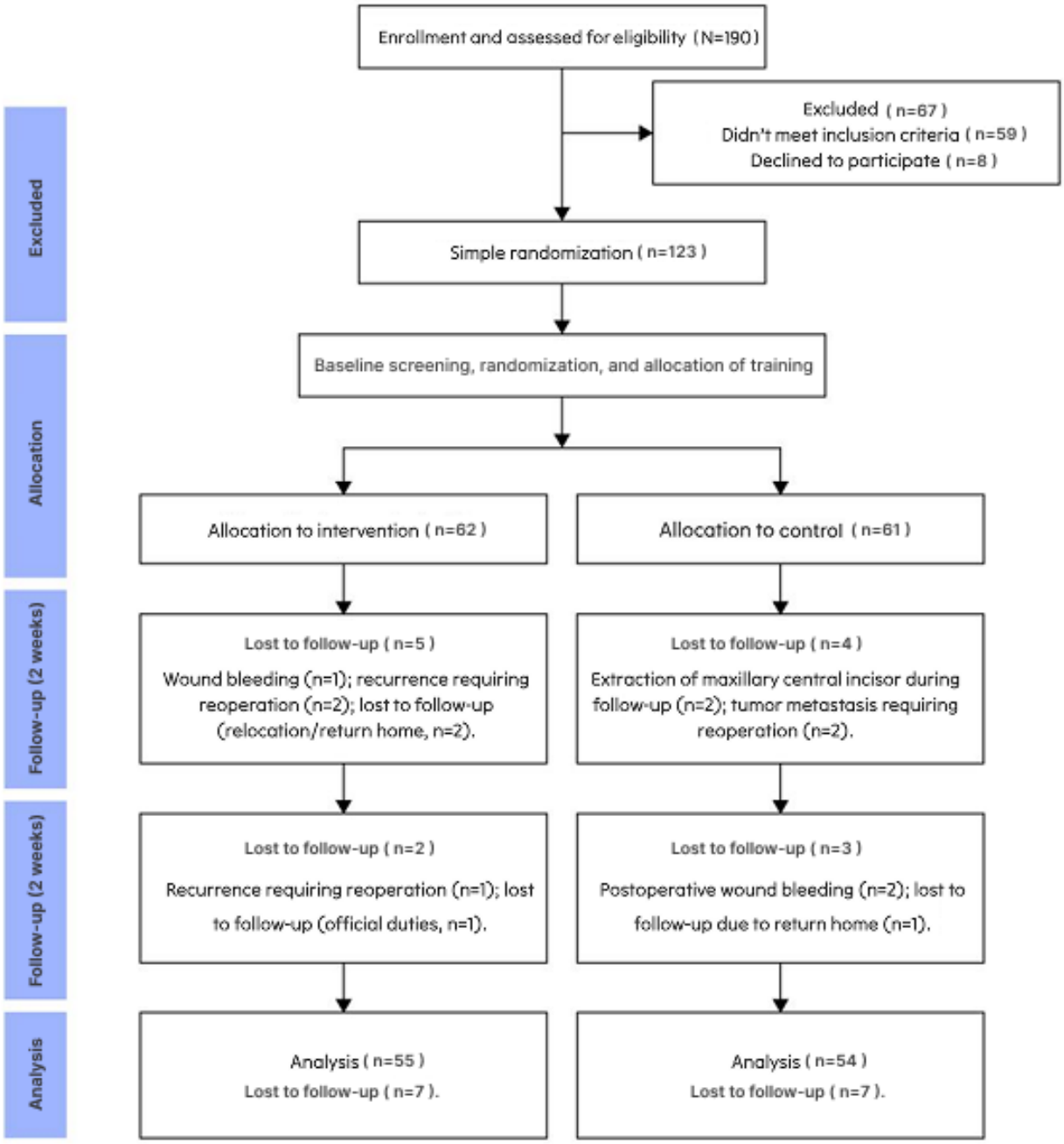
Characteristics at Baseline

A total of 190 patients were screened for eligibility, of whom 67 did not meet the inclusion criteria and were excluded. Written informed consent was obtained from all participants. During the study, 7 participants in the app-based intervention group were excluded for the following reasons: postoperative wound bleeding (n=1); postoperative recurrence requiring reoperation (n=3); and loss to follow-up due to relocation, official duties, or return to hometown (n=3). In the routine care control group, 7 participants were excluded for the following reasons: extraction of the maxillary central incisor during follow-up (n=2), tumor metastasis requiring reoperation (n=2), postoperative wound bleeding (n=2), and loss to follow-up due to return to hometown (n=1). Ultimately, 109 valid cases were included. In accordance with the CONSORT (Consolidated Standards of Reporting Trials) flow diagram (Figure 7; see also Multimedia Appendix 2) [61], participants in both groups were analyzed according to their original allocation (intervention group: n=55; control group: n=54).

Before beginning the intervention, disease-related data (cancer type, treatment method, course of disease, number of mouth-opening recovery exercises per week, and household residence), as well as name, gender, age, and education level, were collected from all participants. Baseline measurements were conducted 1 week before the commencement of the intervention, focusing on maximum mouth opening and other scores (Table 1). Males comprised the majority of the sample (75/109, 68.8%), with a mean age of 47.31 (SD 8.47) years, which was consistent with the prevalence of NPC. Table 1 presents outcomes related to NPC-related indicators, maximal mouth opening, mouth-opening symmetry, weekly training frequency, gender, age, education level, and the Shapiro-Wilk test. There were no significant differences between groups (see Table 1), indicating comparability and suitability for subsequent experiments and analyses. At baseline, based on the evaluation of participants' cancer-related function and the potential impact on quality of life after radiotherapy, mouth-opening training was supported.

XSL•FO
RenderX

Figure 7. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.



XSL•FO
RenderX

Table 1. Baseline characteristics.

Variables	Intervention group (n=55)	Control group (n=54)	*P* value
Age	46.55 (7.98)	48.06 (8.94)	.43
Gender, n (%)			.53
Male	41 (74.5)	34 (63.0)	
Female	14 (25.5)	20 (37.0)	
Education level			.93
Lower secondary and below	11 (20.0)	9 (16.7)	
High school	18 (32.7)	18 (33.3)	
Tertiary and above	26 (47.3)	27 (50.0)	
Family residence			.77
City	32 (58.2)	36 (66.7)	
Countryside	23 (41.8)	18 (33.3)	
Marital status			.69
Unmarried	7 (12.7)	9 (16.7)	
Married	48 (87.3)	45 (83.3)	
Monthly income (Yuan[a])			.81
<3000	7 (12.7)	4 (7.4)	
3001-5000	25 (45.5)	29 (53.7)	
>5001	23 (41.8)	21 (38.9)	
Cancer stage, n (%)			.74
Ⅱ and below	7 (12.7)	11 (20.4)	
Ⅲ	18 (32.7)	16 (29.6)	
Ⅳ	30 (54.5)	27 (50.0)	
Number of chemotherapy sessions, mean (SD)	3.13 (1.98)	3.33 (1.60)	.75
Number of radiotherapy sessions, mean (SD)	22.43 (1.60)	22.55 (1.95)	.11
Duration of disease (months), n (%)			.84
≤5	11 (20.0)	16 (29.6)	
6-11	18 (32.7)	14 (25.9)	
12-17	16 (29.1)	18 (33.3)	
≥18	10 (18.2)	6 (11.1)	
Number of exercises per week, n (%)			.81
≤4	34 (61.8)	38 (70.4)	
5-8	18 (32.7)	14 (25.9)	
≥9	3 (5.5)	2 (3.7)	
Maximum value of mouth opening (mm), mean (SD)	36.98 (1.61)	37.05 (1.63)	.88
Evaluation of mouth opening symmetry, mean (SD)	0.836 (0.086)	0.841 (0.079)	.67
Karnofsky performance status, mean (SD)	82.16 (10.80)	81.85 (16.45)	.44
Eastern Cooperative Oncology Group, mean (SD)	2.55 (1.85)	2.65 (1.65)	.48

[a]1 yuan=US $0.14.

XSL•FO
RenderX

Outcomes

Primary Outcomes

Table 2 shows changes in the Health Belief Model subscales and clinical indicators of improved jaw clenching from baseline to the 4-week trial period after the intervention. For perceived susceptibility, although changes in both the intervention and control groups did not reach statistical significance (*P*=.07), the intervention group score increased from 6.0 at baseline to 7.5, indicating a certain degree of positive change. This contrasted with the small change observed in the control group (from 5.8 to 6.1). For perceived severity, the mean value in the intervention group increased from 23.3 to 27.8, with a difference of 4.5 (95% CI 3.6-5.4; *P*=.007), indicating a significant difference, whereas the control group showed little change, increasing only from 24.6 to 25.4. For perceived benefits, the mean score in the intervention group increased from 11.2 to 11.9 (difference 0.7; 95% CI 0.2-1.2), which was statistically significant (*P*=.04). At the same time, the reduction in perceived impairment was particularly significant (from 15.5 to 8.5; difference −7.0; 95% CI −8.1 to −5.9; *P*=.001). In addition, self-efficacy increased from 5.4 to 6.4 (difference 1.0; 95% CI 0.6-1.4; *P*=.04), and action cues increased from 7.5 to 11.2 (difference 3.7; 95% CI 3.2-4.2; *P*=.001), indicating that the intervention successfully improved individuals' confidence and motivation to adopt healthy behaviors. The increase in maximum mouth opening (from 36.98 to 39.05; difference 2.07; 95% CI 1.49-2.61; *P*=.04) and improvements in physical indicators related to oral health demonstrated the potential benefits of the intervention in promoting specific health behaviors. The intervention group performed significantly more exercises per week (mean 10.0, SD 3.5) than the control group (mean 6.7, SD 2.3; *P*=.001). However, small changes in mouth-opening symmetry (−0.014; 95% CI −0.024 to 0.005; *P*=.15) suggested that the intervention was not equally effective across all outcomes.

Table 2. Primary and secondary intervention results.

Results	Intervention (n=55)			Control (n=54)			*P* value
	Baseline, mean (SD)	Follow-up, mean (SD)	Difference (95% CI)	Baseline, mean (SD)	Follow-up, mean (SD)	Difference (95% CI)	
Primary outcome							
Perceived susceptibility	6.0 (3.4)	7.5 (1.8)	1.5 (0.4 to 2.6)	5.8 (3.4)	6.1 (2.9)	0.3 (–1.2 to 1.8)	.07
Perceived severity	23.3 (3.8)	27.8 (2.6)	4.5 (3.6 to 5.4)	24.6 (3.6)	25.4 (2.7)	0.8 (0.3 to 1.9)	.007
Perceived benefits	11.2 (1.2)	11.9 (1.4)	0.7 (0.2 to 1.2)	10.9 (1.3)	10.8 (1.6)	–0.1 (–0.7 to 0.5)	.04
Perceived barriers	15.5 (5.5)	8.5 (5.2)	–7.0 (–8.1 to –5.9)	15.4 (6.4)	15.5 (5.9)	0.1 (–3.5 to 3.7)	.001
Self-efficacy	5.4 (1.1)	6.4 (1.2)	1.0 (0.6 to 1.4)	5.9 (1.1)	5.5 (1.2)	–0.4 (–0.9 to –0.1)	.04
Cues to action	7.5 (1.8)	11.2 (2.9)	3.7 (3.2 to 4.2)	7.7 (1.6)	7.2 (1.4)	–0.5 (–1.1 to 0.1)	.001
Action (behavior)	6.6 (2.7)	8.1 (2.8)	1.5 (0.9 to 2.1)	6.4 (2.6)	5.9 (2.8)	–0.5 (–1.3 to 0.3)	.03
Number of exercises per week	4.2 (2.6)	10.0 (3.5)	5.8 (4.8 to 6.8)	3.4 (2.6)	6.7 (2.3)	3.3 (1.9 to 4.6)	.001
Maximum value of mouth opening (mm)	36.98 (1.61)	39.05 (1.38)	2.07 (1.49 to 2.61)	37.05 (1.63)	38.02 (1.54)	0.97 (0.49 to 1.41)	.04
Evaluation of mouth opening symmetry	0.836 (0.086)	0.822 (0.048)	–0.014 (–0.024 to 0.005)	0.841 (0.079)	0.838 (0.076)	–0.003 (–0.023 to 0.017)	.15
Secondary cognitive outcomes							
Borg Rating of Perceived Exertion Scale score	11.85 (1.11)	12.98 (0.76)	1.13 (0.81 to 1.45)	11.65 (0.66)	13.43 (1.13)	1.78 (1.33 to 2.23)	.04
Assessment of Quality of Life—6 Dimensions	0.46 (0.19)	0.48 (0.18)	0.02 (–0.04 to 0.08)	0.55 (0.17)	0.52 (0.21)	–0.03 (–0.09 to 0.03)	.18

Secondary Cognitive Outcomes

The control group experienced a significant increase in perceived exertion, with a mean difference in Borg Rating of Perceived Exertion Scale score of 1.78 (95% CI 1.33-2.23; *P*=.04), indicating a notable rise in fatigue levels compared with the intervention group. By contrast, changes in quality of life, as measured by AQoL-6D, were not significant in either group (*P*=.18). System availability was evaluated using the SUS, and the experimental group achieved a mean usability score of 74.3 out of 100. Scores exceeding 70 were classified as good, indicating high system availability.

Discussion

Principal Findings

This study evaluated Open Care, a telemedicine app based on a deep learning model integrated with computer vision, for real-time monitoring and feedback of facial training exercises. The app aims to enhance rehabilitation outcomes in patients with nasopharyngeal cancer following radiation therapy. By leveraging the lightweight EffiFCNN-ResNet deep learning architecture and integrated network communication technologies, Open Care effectively captures and corrects

XSL•FO
RenderX

nonrigid facial training movements in real time. It also provides intuitive visual feedback and automated rehabilitation assessments through computer vision. The results indicate that, following use of the app, patients with nasopharyngeal cancer showed significant improvements in clinical metrics, including maximum mouth opening, training frequency for mouth opening, perceived severity, perceived benefits, perceived barriers, self-efficacy, action cues, behavior, and perceived fatigue, compared with the control group. These findings suggest that the Open Care app holds substantial potential for replicating successful rehabilitation interventions in patients with nasopharyngeal cancer. Consistent with existing literature, this study reinforces the critical role of structured, feedback-driven rehabilitation in alleviating trismus and improving quality of life for patients with NPC [62,63]. Our findings align with recent remote intervention studies, which highlight that integrating motivational elements and structured feedback significantly enhances patient exercise adherence and functional recovery outcomes [64-66]. This research further validates the positive impact of guided, regular training on symptom amelioration.

However, the "Open Care" methodology and technological pathway employed in this study diverge significantly from existing work. Most current related research predominantly relies on contact sensors or hardware devices (eg, wearable sensors or specialized rehabilitation equipment) coupled with periodic remote health education for intervention [67,68]. Outcome assessment in these approaches is largely based on manual measurements (eg, caliper measurement of mouth opening) and patient self-report questionnaires. While these traditional methods possess certain value, they generally exhibit inherent limitations: self-reported measurements are susceptible to subjective bias [69], real-time continuous monitoring of movement quality is challenging [70], intervention in daily patient training is often delayed [71], and hardware dependence elevates the barrier to entry and cost [72]. By contrast, "Open Care" facilitates entirely contactless, computer vision–based real-time motion analysis and immediate corrective feedback. It leverages high-efficiency deep learning models to automatically extract facial movement features and quantify similarity to standard movements for scoring, thereby mitigating human measurement error, minimizing intervention invasiveness, and significantly enhancing patient engagement and enjoyment [73]. This is likely a key reason why the experimental group in this study demonstrated more pronounced improvements in objective metrics [74]. Thus, unlike previous studies solely focused on remote follow-up education or single-assessment dimensions, the assessment methodology of "Open Care" demonstrates distinct innovation and advantages. It enables multimodal data fusion assessment, integrating objective training data (eg, movement accuracy and adherence) with subjective patient-reported outcomes, providing a more scientific and comprehensive quantification of rehabilitation efficacy [75]. Furthermore, Open Care represents an advancement in technological integration and interaction paradigms, innovatively incorporating real-time computer vision feedback into the clinical practice of remote rehabilitation for NPC. By offering an immersive and gamified training experience, the system effectively enhances patients' long-term adherence [73], a feat that is challenging to achieve with traditional remote education or simple video guidance [76].

Further in-depth analysis of the intervention data indicates that, in terms of objective clinical indicators, our clinical trial reported significant differences in maximum mouth opening and frequency of mouth-opening exercises, a promising finding that demonstrates better clinical outcomes than those reported in previous remote support studies [77-79]. After the intervention, maximum mouth opening increased from 36.98 to 39.05, corresponding to improvements in physical indicators related to oral health. In the intervention group, the mean increase in maximum mouth opening was 2.07 mm, with a median increase ranging from 1.5 to 2.6 mm. In addition, within 4 weeks after discharge, the number of patients with clenched jaws decreased, and the magnitude of this change was comparable to findings from previous studies involving patients with head and neck cancer [80,81], indicating the potential benefits of the intervention in improving specific health behaviors. Furthermore, treatment adherence is a critical determinant of the success of any preventive exercise program [15,82]. We observed that the intervention group performed significantly more exercises per week (mean 10.0, SD 3.5) than the control group (mean 4.2, SD 2.6), demonstrating the significant impact of our developed app on improving patient adherence to home rehabilitation exercises. However, despite the significant improvement in maximum mouth opening, 11 of 109 (10.1%) participants in the experimental group still exhibited clenched jaws during follow-up, suggesting that a longer intervention period may be required [66]. In addition, our study did not find that the app was associated with improvement in mouth-opening symmetry, which contrasts with previous studies. This discrepancy may be partly attributable to differences in study duration and in the type and level of functional support provided [83]. A comprehensive strengthening of the facial nerve structure and musculature is required for symmetry to improve, and short-term interventions are ineffective [84].

In terms of improvements in rehabilitation training behaviors, the remote health care intervention group in this study, compared with the standard care control group, demonstrated significant enhancements in patients' perceived benefits and self-efficacy. These findings are consistent with previous studies [85-87], including one by Schlieter et al [88], which examined the impact of remote health care–based educational programs on improving oral health training behaviors in pregnant women. The results of that randomized controlled trial showed significant improvements in oral health beliefs and clinical training indicators, similar to our study, suggesting that remote health care interventions can effectively improve individuals' health cognition and behavior change [89]. This supports the rational use of such models to improve health behaviors across a wide range of environments and populations [90,91]. The remote health care tool developed in this study provides patients with real-time supervision of training movements, corrective prompts, and periodic visual training score reports. This semisupervised, freely accessible intervention appears to be an effective approach for enhancing perceived benefits and self-efficacy [92]. Compared with traditional educational methods, it is more cost-effective and offers greater penetration [93]. The significant

XSL•FO
RenderX

improvements in Perceived Severity and Perceived Benefits further validate the effectiveness of the intervention. Our app, using the EffiFCNN-ResNet deep learning model and computer vision, tracks participants' mouth-opening movements to facilitate appropriate exercise and training plans while providing personalized feedback for patients at different levels. This may also effectively promote the growth of self-efficacy among patients [94], thereby contributing to clinical evidence for personalized health care design indicators in remote health care, as reported in a meta-analysis [95].

This study yielded more substantial changes in action cues and perceived barriers than previous research. In addition, there was a significant increase in the frequency of weekly mouth-opening training behavior. This confirms the promising potential of health apps that use computer vision for personalized training correction and feedback [96]. These apps overcome the spatial limitations of traditional methods and provide more effective interaction and facilitation for monitoring, feedback, and motivation in rehabilitation training [73]. Integrating the activation of behavioral cues into digital intervention tools aims to guide patients with nasopharyngeal cancer to implement training as a health behavior by prompting and facilitating modifications to their decision-making framework [97]. This finding is consistent with previous studies [98-100]. The implementation of multiple strategies can effectively bridge the gap between patients and professional rehabilitation knowledge, reduce training costs, diminish perceived barriers to mouth training as a health behavior, narrow the gap between training cognition and practice, and encourage patients to demonstrate commitment to health behaviors through mouth training. Consistent with the findings of Tore et al [101], telerehabilitation yields a considerably higher standard of physical therapy than self-management. This may be attributable to the duration and quality of training tasks [101,102]. However, there was no statistically significant increase in perceived susceptibility. This may be attributable to the fact that participants in the sample shared a common disease group, which tends to result in more consistent perceptions of specific health hazards [103]. Nonetheless, perceived susceptibility continued to increase after the intervention, which may indicate the intervention's potential.

Furthermore, we propose a reliable and objective intelligent assessment method for mouth-opening rehabilitation in clinical practice. A deep learning model, EffiFCNN-ResNet, was developed to address the needs of nonrigid, markerless, real-time facial capture, tracking correction, and automated assessment for patients with nasopharyngeal cancer. By combining the spatial processing strengths of FCNNs with a lightweight EffiResNet backbone [104], our model achieves fast and accurate facial detection, addressing the trade-offs between computational complexity and execution time. Performance analysis shows that EffiResNet demonstrates strong robustness, quickly reducing failure rates as the error threshold increases, in contrast to other ResNet variants [105,106], which are less adaptable in error-prone environments. The model shows stable performance across varying error magnitudes and excels in dynamic spatial changes [107]. In classification tasks, EffiResNet outperforms other ResNet variants in recall without sacrificing precision. Overall, EffiFCNN-ResNet excels in robustness, stability, and spatial awareness, offering an efficient solution for real-time facial landmark detection in complex environments. Its design enhancements, including the rescaled ResNet backbone and large-kernel depthwise convolutions, significantly contribute to its performance, positioning it as a promising tool for advanced facial recognition tasks. However, it is important to note that, due to certain objective factors such as limited data availability, the model may occasionally exhibit errors in specific situations. These errors primarily result from insufficient diversity in the training data or a lack of specific facial action samples, which can affect the model's accuracy when encountering rare or abnormal actions. Nonetheless, these occasional errors are not substantial enough to notably degrade overall model performance. With continued data collection and model training, these errors are expected to diminish, thereby further enhancing the model's precision.

It is noteworthy that participants reported positive feedback regarding their training experience with the Open Care app, along with a significant reduction in fatigue levels. This positive user experience not only attests to the app's utility and user-friendliness but also highlights its important role in promoting sustained patient engagement in rehabilitation training. Because of the potential presence of uncontrolled external factors during the intervention period, such as other treatments received by participants and changes in lifestyle habits, these elements may have interfered with the study outcomes, particularly for indicators closely related to daily quality of life, such as AQoL-6D [108]. Nevertheless, based on responses from the user experience questionnaire, the researchers confirmed that the app successfully achieved its objectives of precise training movement monitoring and targeted feedback. The majority of respondents regarded the electronic exercise intervention as beneficial. They believed that incorporating interactive design and real-time feedback into exercise programs enhances enjoyment while also encouraging individuals to take the initiative to engage in this healthy behavior. Regarding comfort, most participants reported that the electronic exercise intervention was more personalized and better aligned with their individual rehabilitation needs. Thus, patients with NPC who have undergone radiotherapy may benefit from telemedicine apps that use deep learning and computer vision technologies [109], and this potential should not be overlooked.

Furthermore, Open Care offers substantial advantages in terms of comprehensive benefits and clinical utility. Compared with conventional rehabilitation models that primarily rely on outpatient visits and manual supervision, this app, through its automated real-time monitoring and feedback mechanisms [110], significantly reduces dependence on specialized rehabilitation therapists. This, in turn, minimizes long-term human resource investment and associated costs, aligning with current health care policy orientations toward efficient, accessible, and intelligent services [111], and serving as an important decision-making factor for health care providers when considering new therapeutic interventions [112]. At the same time, because training and assessment can be completed remotely, the frequency of patient hospital visits is substantially reduced. This not only saves transportation, escort, and

XSL•FO
RenderX

time-related costs for patients but also alleviates patient load on health care institutions [113]. From a broader health system perspective, this standardized and scalable intervention approach is particularly well-suited for resource-limited regions. It enables expanded service coverage and equity while maintaining rehabilitation quality by leveraging digital means to enhance high-quality service delivery and decentralize health care resources, thereby optimizing overall resource allocation efficiency [114]. More importantly, through continuous and objective data collection, Open Care offers the potential for early intervention and dynamic adjustment of rehabilitation plans. This proactively mitigates the risk of secondary health complications and additional medical expenditures that may arise from improper or delayed rehabilitation [115]. From the perspective of health care payers, such preventive, data-driven health management tools help control overall health care expenditures, consistent with the ongoing policy shift from treatment-centric to health-centric approaches [116]. Consequently, Open Care not only provides immediate cost-saving benefits but also, in the long term, strengthens the resilience and value proposition of health care services, addressing the shared objectives of multiple stakeholders, including patients, health care institutions, payers, and governmental bodies.

Limitations and Future

However, our investigation is not without limitations. First, the sample size was limited and comprised individuals from several nearby tumor departments. This may have resulted in population clustering, which is not ideal for deriving generalizable conclusions. Furthermore, during the pre-experiment phase, some patients with limited digital health proficiency were unable to independently troubleshoot the equipment used in the experimental procedure. In addition, as an intervention app based on an augmented reality environment, the system may be unsuitable for patients who are sensitive to sensory stimulation at the design level. Moreover, although the trial demonstrated short-term user engagement, the long-term sustainability of engagement and activity remains uncertain. The advancement of tools for capturing, monitoring, and evaluating nonrigid facial models must contend with a wide range of human facial expressions and movements, including complex mouth movements and subtle eye and eyebrow motions. The recognition accuracy of the model we developed still has room for improvement. Precisely capturing and tracking these subtle variations is particularly challenging under low-resolution or real-time conditions. Resource limitations may also affect system response time and user experience [117]. The experiment lasted only 4 weeks; moving forward, it will be essential to extend the intervention duration and further refine the app design to optimize outcomes. In conclusion, future research should investigate the applicability of Open Care for patients with NPC living in diverse and complex environments and at different stages of rehabilitation. Furthermore, in light of ongoing technological advances and integration of user feedback, continuous refinement and expansion of app functionalities will be critical to ensuring sustained effectiveness and broad applicability.

Conclusion

This study successfully developed “Open Care,” a novel telemedicine app using the EffiFCNN-ResNet deep learning model and computer vision for real-time facial motion tracking and assessment in NPC rehabilitation. Our randomized controlled trial demonstrated significant efficacy in improving objective clinical metrics, such as maximum mouth opening and training frequency, as well as a comprehensive range of patient-reported outcomes, including perceived severity, perceived benefits, perceived barriers, self-efficacy, action cues, enhanced training behavior, and reduced perceived fatigue. These findings highlight the substantial potential of “Open Care” to enhance patient adherence, functional recovery, and quality of life in NPC rehabilitation, offering an accessible, cost-effective, and scalable digital solution for clinical practice, particularly in resource-limited settings. Future research will focus on refining model accuracy, extending intervention duration, and exploring applicability to broader head and neck cancer populations, further strengthening the role of artificial intelligence–driven telemedicine in oncological supportive care.

Acknowledgments

The authors thank the anonymous reviewers and the Human Subjects Research Ethics Committee for granting access to patients, as well as all participants for their collaboration in this study. This study was supported by the National Natural Science Foundation of China (72471140), the National Key R&D Program (grant 2022YFB3303303, First label), and the Fundamental Research Funds for the Central Universities (grant YG2023ZD10). In addition, the authors gratefully acknowledge the valuable support provided by Master Yang Yange, Master Wu Hantian, Master Liu Yongzhi, Master Yuan Yuyi, and Dr. Wang Luyao during the software interface design stage. During the preparation of this work, the author(s) conducted all research and writing manually, ensuring the accuracy and quality of the content. After completion, the author(s) reviewed and edited the manuscript as needed and take full responsibility for the content of the publication.

Data Availability

The datasets used or analyzed during this study are available from the corresponding author (TH) on reasonable request.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Specific fitting functions.
[DOCX File , 20 KB - mhealth_v14i1e72560_app1.docx]

Multimedia Appendix 2
CONSORT eHEALTH checklist (v 1.6.1).
[PDF File (Adobe PDF File), 1513 KB - mhealth_v14i1e72560_app2.pdf]

References

1. Qian S, Tan G, Lei G, Zhang X, Xie Z. Programmed cell death in nasopharyngeal carcinoma: mechanisms and therapeutic targets. Biochim Biophys Acta Rev Cancer 2025 Feb;1880(1):189265-189272. [doi: 10.1016/j.bbcan.2025.189265] [Medline: 39809344]
2. Bhatia KS, King AD, Paunipagar BK, Abrigo J, Vlantis AC, Leung SF, et al. MRI findings in patients with severe trismus following radiotherapy for nasopharyngeal carcinoma. Eur Radiol 2009 Nov 6;19(11):2586-2593. [doi: 10.1007/s00330-009-1445-z] [Medline: 19504110]
3. Chen S, Duan X, Li H, Peng L, Wang Z, Xu G, et al. Efficacy of sequential chemoradiotherapy combined with toripalimab in de novo metastatic nasopharyngeal carcinoma: a phase II trial. Cell Rep Med 2023 Nov 21;4(11):101279-101285 [FREE Full text] [doi: 10.1016/j.xcrm.2023.101279] [Medline: 37951218]
4. Zhu Y, Zheng D, Xu S, Chen J, Wen L, Zhang Z, et al. Intratumoral habitat radiomics based on magnetic resonance imaging for preoperative prediction treatment response to neoadjuvant chemotherapy in nasopharyngeal carcinoma. Jpn J Radiol 2024 Dec;42(12):1413-1424. [doi: 10.1007/s11604-024-01639-8] [Medline: 39162780]
5. Cheng J, Su W, Wang Y, Zhan Y, Wang Y, Yan S, et al. Magnetic resonance imaging based on radiomics for differentiating T1-category nasopharyngeal carcinoma from nasopharyngeal lymphoid hyperplasia: a multicenter study. Jpn J Radiol 2024 Jul 27;42(7):709-719. [doi: 10.1007/s11604-024-01544-0] [Medline: 38409300]
6. Yan C, Wang R, Zhang C, Lin G, Du Y, Chen M, et al. Induction chemotherapy for locally advanced nasopharyngeal carcinoma: efficacy and safety of the TPC regimen compared to GP and TPF. Oral Oncol 2025 Jan;160:107119-107126. [doi: 10.1016/j.oraloncology.2024.107119] [Medline: 39591693]
7. Yang L, Hao G, Yang W, Hou L. The impact of different timing of mouth opening exercises on trismus in postoperative radiotherapy patients with oral cancer. J Stomatol Oral Maxillofac Surg 2025 Jun;126(3):102104. [doi: 10.1016/j.jormas.2024.102104] [Medline: 39366485]
8. Shao CH, Chiang CC, Huang TW. Exercise therapy for cancer treatment-induced trismus in patients with head and neck cancer: a systematic review and meta-analysis of randomized controlled trials. Radiother Oncol 2020 Oct 22;151(1):249-255. [doi: 10.1016/j.radonc.2020.08.024] [Medline: 32890607]
9. Li Y, Hou L, Gan X, Yang X. Meta-analysis of the effect of non-pharmacological interventions on the development of dysgeusia in patients undergoing radiotherapy for head and neck squamous cell carcinoma. Sci Rep 2025 Jan 07;15(1):1094 [FREE Full text] [doi: 10.1038/s41598-024-83081-4] [Medline: 39774979]
10. Tan X, Liu J, Deng Y, Li X, Yang W, Hu X, et al. Analysis of factors related to radiation-induced oral mucositis in patients with head and neck tumors undergoing radiotherapy. J Stomatol Oral Maxillofac Surg 2025 Mar;126(2):102042. [doi: 10.1016/j.jormas.2024.102042] [Medline: 39244031]
11. Yarlagadda S, Rohe RM, McNeill V, Nemec JR, Tohtz CP, Fellows ZW, et al. Radiother Oncol 2025 Jun;207:110882. [doi: 10.1016/j.radonc.2025.110882] [Medline: 40194703]
12. Samarah H, Crawley M, Young G. Treatment modalities, surgical principles, and rehabilitation considerations in head and neck cancer. In: Head and Neck Cancer Rehabilitation. Amsterdam, The Netherlands: Elsevier; 2025:392-362.
13. Hapid MH, Sufiawati I, Kusumadjati A, Denta E. Evaluation of radiation planning in the development of acute radiation oral mucositis in head and neck cancer patients: a case series. Journal of Oral and Maxillofacial Surgery, Medicine, and Pathology 2025 Mar;37(2):356-360. [doi: 10.1016/j.ajoms.2024.08.006]
14. Martínez-Borba V, Lorente-Debón B, Osma J. A systematic review of telemedicine solutions to provide psychological interventions for women receiving fertility treatments. Telemed J E Health 2025 May;31(5):555-568. [doi: 10.1089/tmj.2024.0446] [Medline: 39841539]
15. Walha R, Koubaa N, Chagnon M, Lortie-Milner E, Aubertin-Leheudre M, Levasseur M, et al. e-Health interventions for promoting physical activity in aging adults: a scoping review. Telemed J E Health 2025 May;31(5):531-539. [doi: 10.1089/tmj.2024.0414] [Medline: 39757866]
16. Hwang S, Agarwal P, Dakov M, Downes MH, Kummer BR. Health care utilization patterns associated with econsults for headache: insights from an urban academic medical center. Telemed J E Health 2025 May;31(5):614-623. [doi: 10.1089/tmj.2024.0483] [Medline: 39778898]
17. Chow JC. Internet-based computer technology on radiotherapy. Rep Pract Oncol Radiother 2017;22(6):455-462. [doi: 10.1016/j.rpor.2017.08.005] [Medline: 28932174]

XSL•FO
RenderX

18. Wilson R, Chua J, Briggs AM, Abbott JH. The cost-effectiveness of recommended adjunctive interventions for knee osteoarthritis: results from a computer simulation model. Osteoarthr Cartil Open 2020 Dec;2(4):100123 [FREE Full text] [doi: 10.1016/j.ocarto.2020.100123] [Medline: 36474885]
19. Kostic AM, Leifer VP, Gong Y, Robinson MK, Collins JE, Neogi T, et al. Cost-effectiveness of surgical weight-loss interventions for patients with knee osteoarthritis and class III obesity. Arthritis Care Res (Hoboken) 2023 Mar;75(3):491-500 [FREE Full text] [doi: 10.1002/acr.24967] [Medline: 35657632]
20. Zhu A, Tailor P, Verma R, Zhang I, Schott B, Ye C, et al. Implementation of deep learning artificial intelligence in vision-threatening disease screenings for an underserved community during COVID-19. J Telemed Telecare 2023 Mar 13;30(10):1590-1597. [doi: 10.1177/1357633x231158832]
21. Snoswell CL, Chelberg G, De Guzman KR, Haydon HH, Thomas EE, Caffery LJ, et al. The clinical effectiveness of telehealth: a systematic review of meta-analyses from 2010 to 2019. J Telemed Telecare 2023 Oct;29(9):669-684. [doi: 10.1177/1357633X211022907] [Medline: 34184580]
22. Lee K, Kim S, Kim SH, Yoo S, Sung JH, Oh EG, et al. Digital health interventions for adult patients with cancer evaluated in randomized controlled trials: scoping review. J Med Internet Res 2023 Jan 06;25:e38333 [FREE Full text] [doi: 10.2196/38333] [Medline: 36607712]
23. Vinolo-Gil MJ, García-Campanario I, Estebanez-Pérez MJ, Góngora-Rodríguez J, Rodríguez-Huguet M, Martín-Valero R. Telehealth assessment of diagnostic and therapeutic efficacy in peripheral vestibular symptoms: a systematic review and meta-analysis. Telemed J E Health 2025 May;31(5):540-554. [doi: 10.1089/tmj.2024.0301] [Medline: 39757883]
24. Moschonis G, Siopis G, Jung J, Eweka E, Willems R, Kwasnicka D, DigiCare4You Consortium. Effectiveness, reach, uptake, and feasibility of digital health interventions for adults with type 2 diabetes: a systematic review and meta-analysis of randomised controlled trials. Lancet Digit Health 2023 Mar;5(3):e125-e143 [FREE Full text] [doi: 10.1016/S2589-7500(22)00233-3] [Medline: 36828606]
25. Alenezi EM, Veselinović T, Tao KF, Altamimi AA, Tran TT, Herbert H, et al. Ear Portal: an urban-based ear, nose, and throat, and audiology referral telehealth portal to improve access to specialist ear health services for children. J Telemed Telecare 2023 Mar 14;30(10):1581-1589. [doi: 10.1177/1357633x231158839]
26. Siopis G, Moschonis G, Eweka E, Jung J, Kwasnicka D, Asare B, DigiCare4You Consortium. Effectiveness, reach, uptake, and feasibility of digital health interventions for adults with hypertension: a systematic review and meta-analysis of randomised controlled trials. Lancet Digit Health 2023 Mar;5(3):e144-e159 [FREE Full text] [doi: 10.1016/S2589-7500(23)00002-X] [Medline: 36828607]
27. Kene MV, Sax DR, Bhargava R, Somers MJ, Warton EM, Zhang JY, et al. Post-telemedicine acute care for undifferentiated high-acuity conditions: is a picture worth a thousand words? Telemed J E Health 2025 May;31(5):569-578. [doi: 10.1089/tmj.2024.0425] [Medline: 39791218]
28. Petersson K, Finizia C, Pauli N, Tuomi L. Preventing radiation-induced dysphagia and trismus in head and neck cancer-a randomized controlled trial. Head Neck 2025 Jan;47(1):159-174. [doi: 10.1002/hed.27886] [Medline: 39091121]
29. Chow J. Artificial intelligence in radiotherapy and patient care. In: Lidströmer N, Ashrafian H, editors. Artificial Intelligence in Medicine. Cham, Switzerland: Springer International Publishing; Feb 18, 2022:1-13.
30. Siddique S, Chow JC. Artificial intelligence in radiotherapy. Rep Pract Oncol Radiother 2020;25(4):656-666 [FREE Full text] [doi: 10.1016/j.rpor.2020.03.015] [Medline: 32617080]
31. Tai DT, Oanh LT, Son ND, Loan TTH, Chow JC. Dosimetric and Monte Carlo verification of jaws-only IMRT plans calculated by the Collapsed Cone Convolution algorithm for head and neck cancers. Rep Pract Oncol Radiother 2019;24(1):105-114 [FREE Full text] [doi: 10.1016/j.rpor.2018.11.004] [Medline: 30532658]
32. Justribó-Manion C, Mesa-Jiménez J, Bara-Casaus J, Zuil-Escobar J, Wachowska K, Álvarez-Bustins G. Additional effects of therapeutic exercise and education on manual therapy for chronic temporomandibular disorders treatment: a randomized clinical trial. Physiother Theory Pract 2025 Jan;41(1):12-27. [doi: 10.1080/09593985.2024.2316305] [Medline: 38353484]
33. Morris G, Ye J. Innovations in physical and mental education recovery programs in China: patient initiatives, supervised rehabilitation, and future implications. In: Global Innovations in Physical Education and Health. New York, NY: IGI Global; Sep 1, 2024:61-84.
34. Xu L, Sanders L, Li K, Chow JCL. Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review. JMIR Cancer 2021 Nov 29;7(4):e27850 [FREE Full text] [doi: 10.2196/27850] [Medline: 34847056]
35. He Z, Zhang K, Zhao N, Wang Y, Hou W, Meng Q, et al. Deep learning for real-time detection of nasopharyngeal carcinoma during nasopharyngeal endoscopy. iScience 2023 Oct 20;26(10):107463 [FREE Full text] [doi: 10.1016/j.isci.2023.107463] [Medline: 37720094]
36. Durga BK, Rajesh V, Jagannadham S, Kumar PS, Rashed ANZ, Saikumar K. Deep learning-based micro facial expression recognition using an adaptive Tiefes FCNN model. Trait Signal 2023 Jun 28;40(3):1035-1043. [doi: 10.18280/ts.400319]
37. Le V. Deep learning-based for human segmentation and tracking, 3D human pose estimation and action recognition on monocular video of MADS dataset. Multimed Tools Appl 2022 Oct 25;82(14):20771-20818. [doi: 10.1007/s11042-022-13921-w]

XSL•FO
RenderX

38. Singh NM, Dey N, Hoffmann M, Fischl B, Adalsteinsson E, Frost R, et al. Data consistent deep rigid MRI motion correction. Proc Mach Learn Res 2024;227:368-381. [Medline: 39415845]
39. Sadiq M, Shi D. Attentive occlusion-adaptive deep network for facial landmark detection. Pattern Recognition 2022 May;125:108510. [doi: 10.1016/j.patcog.2021.108510]
40. Elsayed M, Reda M, Mashaly AS, Amein AS. LERFNet: an enlarged effective receptive field backbone network for enhancing visual drone detection. Vis Comput 2024 Jul 01;41(4):2219-2232. [doi: 10.1007/s00371-024-03527-8]
41. Wang Y, Wang W, Li Y, Jia Y, Xu Y, Ling Y, et al. An attention mechanism module with spatial perception and channel information interaction. Complex Intell Syst 2024 May 06;10(4):5427-5444. [doi: 10.1007/s40747-024-01445-9]
42. Wen A, Zhu Y, Xiao N, Gao Z, Zhang Y, Wang Y, et al. Comparison Study of Extraction Accuracy of 3D Facial Anatomical Landmarks Based on Non-Rigid Registration of Face Template. Diagnostics (Basel) 2023 Mar 13;13(6):1086 [FREE Full text] [doi: 10.3390/diagnostics13061086] [Medline: 36980394]
43. Kossaifi J, Tzimiropoulos G, Todorovic S, Pantic M. AFEW-VA database for valence and arousal estimation in-the-wild. Image and Vision Computing 2017 Sep;65:23-36. [doi: 10.1016/j.imavis.2017.02.001]
44. Ding X, Zhang X, Han J, Ding G. Scaling up your kernels to 31x31: revisiting large kernel design in CNNs. New York, NY: IEEE; 2022 Apr 2 Presented at: Conference on Computer Vision and Pattern Recognition; March 13, 2022; New Orleans, LA p. 18-24. [doi: 10.48550/arXiv.2203.06717]
45. Tang X, Xu W, Li K, Han M, Ma Z, Wang R. PIAENet: pyramid integration and attention enhanced network for object detection. Information Sciences 2024 Jun;670:120576. [doi: 10.1016/j.ins.2024.120576]
46. Li S, Deng Y, Zhu Z, Hua H, Tao Z. A comprehensive review on radiomics and deep learning for nasopharyngeal carcinoma imaging. Diagnostics (Basel) 2021 Aug 24;11(9):1523 [FREE Full text] [doi: 10.3390/diagnostics11091523] [Medline: 34573865]
47. Aurelian S, Aurelian J, Stoica S, Vlădulescu-Trandafir A, Zamfirescu A, Onose G. How do physicians and older adults perceive the use of E-health in the management of chronic pathology? A pilot study. Balneo and PRM Research Journal 2024;15(4):761.
48. Inoo Y, Iida H, Nakada H, Tezuka K, Kikuchi Y, Fujimura E, et al. Retrospective observational study on telemedicine in sublingual immunotherapy for patients with Japanese cedar pollinosis and house dust mite allergic rhinitis. Telemed J E Health 2025 Jan;31(1):107-118. [doi: 10.1089/tmj.2024.0065] [Medline: 39258769]
49. Lavanya Y, Rajalakshmi N, Sumanth K, Gowrishankar S. A novel approach for developing inclusive real-time yoga pose detection for health and wellness using Raspberry pi. 2023 Presented at: 7th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS); November 2-4, 2023; Bangalore, Karnataka, India. [doi: 10.1109/CSITSS60515.2023.10334109]
50. Mollica V, Rizzo A, Marchetti A, Tateo V, Tassinari E, Rosellini M, et al. The impact of ECOG performance status on efficacy of immunotherapy and immune-based combinations in cancer patients: the MOUSEION-06 study. Clin Exp Med 2023 Dec;23(8):5039-5049. [doi: 10.1007/s10238-023-01159-1] [Medline: 37535194]
51. Kang H. Sample size determination and power analysis using the G*Power software. J Educ Eval Health Prof 2021;18:17 [FREE Full text] [doi: 10.3352/jeehp.2021.18.17] [Medline: 34325496]
52. Lan L, Lian Z. Application of statistical power analysis – how to determine the right sample size in human health, comfort and productivity research. Building and Environment 2010 May;45(5):1202-1213. [doi: 10.1016/j.buildenv.2009.11.002]
53. Roque DR, Cronin B, Robison K, Lopes V, Rizack T, Dizon DS. The effects of age on treatment and outcomes in women with stage IB-IIB cervical cancer. JCO 2012 May 20;30(15_suppl):5100-5100. [doi: 10.1200/jco.2012.30.15_suppl.5100]
54. Nunan D, Heneghan C, Spencer EA. Catalogue of bias: allocation bias. BMJ Evid Based Med 2018 Feb;23(1):20-21. [doi: 10.1136/ebmed-2017-110882] [Medline: 29367320]
55. Schulz KF, Grimes DA. Blinding in randomised trials: hiding who got what. Lancet 2002 Feb 23;359(9307):696-700. [doi: 10.1016/S0140-6736(02)07816-9] [Medline: 11879884]
56. Huang J, Lin S. Exploring the key factors in the choice of home telehealth by using the health belief model. Telemed J E Health 2009 Jan;15(1):87-92. [doi: 10.1089/tmj.2008.0069] [Medline: 19199852]
57. Sanaeinasab H, Saffari M, Taghavi H, Karimi Zarchi A, Rahmati F, Al Zaben F, et al. An educational intervention using the health belief model for improvement of oral health behavior in grade-schoolers: a randomized controlled trial. BMC Oral Health 2022 Mar 27;22(1):94 [FREE Full text] [doi: 10.1186/s12903-022-02132-2] [Medline: 35346148]
58. Moon I, An Y, Min S, Park C. Therapeutic effects of metaverse rehabilitation for cerebral palsy: a randomized controlled trial. Int J Environ Res Public Health 2023 Jan 15;20(2):1578 [FREE Full text] [doi: 10.3390/ijerph20021578] [Medline: 36674332]
59. Bulamu NB, Mpundu-Kaambwa C, O'Callaghan M, Kaambwa B. Responsiveness and construct validity of EPIC-26, AQoL-6D and SF-6D following treatment in prostate cancer. BMC Cancer 2023 Apr 01;23(1):297 [FREE Full text] [doi: 10.1186/s12885-023-10732-6] [Medline: 37005587]
60. Amadio Viceré MG, Sus M. Differentiated cooperation as the mode of governance in EU foreign policy. Contemporary Security Policy 2023 Feb 01;44(1):4-34. [doi: 10.1080/13523260.2023.2168854]

XSL•FO
RenderX

61. Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P, CONSORT NPT Group. CONSORT statement for randomized trials of nonpharmacologic treatments: a 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. Ann Intern Med 2017 Jul 04;167(1):40-47 [FREE Full text] [doi: 10.7326/M17-0046] [Medline: 28630973]
62. Shaban MM, Sharaa HM, Amer FGM, Shaban M. Effect of digital based nursing intervention on knowledge of self-care behaviors and self-efficacy of adult clients with diabetes. BMC Nurs 2024 Feb 20;23(1):130. [doi: 10.1186/s12912-024-01787-2] [Medline: 38378505]
63. Zhu J, Wang X, Chen S, Du R, Zhang H, Zhang M, et al. Improving compliance with swallowing exercise to decrease radiotherapy-related dysphagia in patients with head and neck cancer. Asia Pac J Oncol Nurs 2023 Jan;10(1):100169 [FREE Full text] [doi: 10.1016/j.apjon.2022.100169] [Medline: 36583099]
64. Sitjar PHS, Tan SY, Wong M, Li J, Jalil RBA, Aw H, et al. Combined aerobic and strength exercise training on biological ageing in Singaporean breast cancer patients: protocol for the Breast Cancer Exercise Intervention (BREXINT) pilot study. Geroscience 2024 Dec;46(6):6029-6038. [doi: 10.1007/s11357-024-01145-9] [Medline: 38546907]
65. Hussey C, Hanbridge M, Dowling M, Gupta A. Cancer survivorship: understanding the patients' journey and perspectives on post-treatment needs. BMC Sports Sci Med Rehabil 2024 Apr 12;16(1):82 [FREE Full text] [doi: 10.1186/s13102-024-00864-y] [Medline: 38605386]
66. Charters E, Ricketts V, Sharman AR, Clark J. Systematic review of adherence to swallow and trismus exercises during radiation therapy for head and neck cancer. Head Neck 2024 Sep 19;46(9):2348-2362. [doi: 10.1002/hed.27849] [Medline: 38895913]
67. Deng B, Chen Y, Meng Y, Zhang Y, Tan X, Zhou X, et al. A self-efficacy-enhancing intervention for Chinese patients after total hip arthroplasty: study protocol for a randomized controlled trial with 6-month follow-up. J Orthop Surg Res 2022 Jan 04;17(1):1 [FREE Full text] [doi: 10.1186/s13018-021-02689-8] [Medline: 34983553]
68. Brown SE, Shah A, Czuber-Dochan W, Bench S, Stayt L. Fatigue after CriTical illness (FACT): co-production of a self-management intervention to support people with fatigue after critical illness. Intensive Crit Care Nurs 2024 Jun;82:103659 [FREE Full text] [doi: 10.1016/j.iccn.2024.103659] [Medline: 38401405]
69. Ortega-Martorell S, Olier I, Ohlsson M, Lip GY, TARGET Consortium. Advancing personalised care in atrial fibrillation and stroke: the potential impact of AI from prevention to rehabilitation. Trends Cardiovasc Med 2025 May;35(4):205-211 [FREE Full text] [doi: 10.1016/j.tcm.2024.12.003] [Medline: 39653093]
70. Charters E, Loy J, Sharman AR, Cheng K, Dunn M, Clark J. Exercise adherence for patients with trismus after head and neck cancer treatment. Head Neck 2024 Nov;46(11):2717-2724. [doi: 10.1002/hed.27799] [Medline: 38708897]
71. Yang L, Hao G, Hou L, Yang W. Rehabilitation strategies for trismus post oral cancer treatment: progress in the study of mouth opening exercises. J Stomatol Oral Maxillofac Surg 2024 Dec;125(6):101796. [doi: 10.1016/j.jormas.2024.101796] [Medline: 38331216]
72. Nayiga BK, Abrams SW, Rhayel A, Edward H, Tang A, Kho ME, et al. Exploring the use of rehabilitation in individuals with head and neck cancer undergoing treatment: a scoping review. Disabil Rehabil 2024 Dec;46(26):6302-6322. [doi: 10.1080/09638288.2024.2328810] [Medline: 38494954]
73. Carlozzi NE, Troost J, Lombard WL, Miner JA, Graves CM, Choi SW, et al. Completion and compliance rates for an intensive mHealth study design to promote self-awareness and self-care among care partners of individuals with traumatic brain injury: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2025 Aug 21;13:e73772. [doi: 10.2196/73772] [Medline: 40840460]
74. Rekola H, Tolmunen T, Mattila E, Strömmer J, Lakka TA, Länsimies H, et al. User archetypes of a well-being-promoting mobile app among adults: cross-sectional study and cluster analysis of usage patterns. JMIR Mhealth Uhealth 2025 Aug 18;13:e68982 [FREE Full text] [doi: 10.2196/68982] [Medline: 40825235]
75. Rajan M, Furgal A, Kadri R, Arman O, Panzer K, Wicker D, et al. Exploring the utility of digital voice assistants for primary care patients, including those with physical and visual disabilities: cross-sectional study. JMIR Mhealth Uhealth 2025 Aug 14;13:e66185 [FREE Full text] [doi: 10.2196/66185] [Medline: 40811816]
76. Schier de Fraga F, Narita MM, Schreiner M, Belli F, Leonel Celestino J, Braz Pereira K, et al. Mobile health adoption in high-risk pregnancies using cluster analysis of biopsychosocial outcomes: observational longitudinal cohort study. JMIR Hum Factors 2025 Aug 21;12:e67680 [FREE Full text] [doi: 10.2196/67680] [Medline: 40840461]
77. Bragante K, Groisman S, Carboni C, Baiocchi J, da Motta NW, Silva M, et al. Efficacy of exercise therapy during radiotherapy to prevent reduction in mouth opening in patients with head and neck cancer: a randomized controlled trial. Oral Surg Oral Med Oral Pathol Oral Radiol 2020 Jan;129(1):27-38. [doi: 10.1016/j.oooo.2019.09.011] [Medline: 31685429]
78. Liu Y, Hou R, Yu Q, Niu Q. Effect of voice training intervention on swallowing function in patients with head and neck cancer undergoing radiotherapy: a randomized controlled trial. Eur J Oncol Nurs 2024 Jun;70:102551. [doi: 10.1016/j.ejon.2024.102551] [Medline: 38642524]
79. Shinn E, Garden A, Chen M, Basen-Engquist K, Fellman B, Hutcheson K, et al. Self-management intervention improves patient adherence to swallowing exercises during radiation for head and neck cancer. Head Neck 2024 Nov;46(11):2878-2889 [FREE Full text] [doi: 10.1002/hed.27832] [Medline: 38873861]
80. Wang T, Su J, Leung K, Liang S, Wu S, Wang H. Corrigendum to ˙Effects of a mouth-opening intervention with remote support on adherence, the maximum interincisal opening, and mandibular function of postoperative oral cancer patients:

XSL•FO
RenderX

A randomized clinical trial' [Eur. J. Oncol. Nurs. 40 (2019) 111-119]. Eur J Oncol Nurs 2019 Aug;41:195. [doi: 10.1016/j.ejon.2019.06.009] [Medline: 31358252]
81. Jr D, Jr C. Trismus therapy devices: a systematic review. Oral Oncol 2022 Mar;126:105728. [doi: 10.1016/j.oraloncology.2022.105728] [Medline: 35104753]
82. Pamplin JC, Quinn MT, Little JR, Passman D, Scott BK. Experience of The National Emergency Tele-Critical Care Network. Telemed J E Health 2025 May;31(5):643-650. [doi: 10.1089/tmj.2024.0585] [Medline: 39804399]
83. Zhao F, Dou Z, Xie C, Yang C, Wan G, Zhang Y, et al. Effect of intensive oropharyngeal training on radiotherapy-related dysphagia in nasopharyngeal carcinoma patients. Dysphagia 2022 Dec;37(6):1542-1549. [doi: 10.1007/s00455-022-10419-3] [Medline: 35175419]
84. Chen H, Zheng X, Huang H, Liu C, Wan Q, Shang S. The effects of a home-based exercise intervention on elderly patients with knee osteoarthritis: a quasi-experimental study. BMC Musculoskelet Disord 2019 Apr 09;20(1):160 [FREE Full text] [doi: 10.1186/s12891-019-2521-4] [Medline: 30967131]
85. Hauth F, Bizu V, App R, Lautenbacher H, Tenev A, Bitzer M, et al. Electronic patient-reported outcome measures in radiation oncology: initial experience after workflow implementation. JMIR Mhealth Uhealth 2019 Jul 24;7(7):e12345 [FREE Full text] [doi: 10.2196/12345] [Medline: 31342906]
86. Pandit JA, Pawelek JB, Leff B, Topol EJ. The hospital at home in the USA: current status and future prospects. NPJ Digit Med 2024 Feb 27;7(1):48 [FREE Full text] [doi: 10.1038/s41746-024-01040-9] [Medline: 38413704]
87. Romalee W, Tsai F, Hsu Y, Hsu M, Wang D. Effectiveness of mobile augmented reality-integrated oral health education for community-dwelling older adults: a randomized controlled trial. Arch Gerontol Geriatr 2024 Feb;117:105277 [FREE Full text] [doi: 10.1016/j.archger.2023.105277] [Medline: 37995647]
88. Schlieter H, Gand K, Weimann TG, Sandner E, Kreiner K, Thoma S, et al. Designing virtual coaching solutions. Bus Inf Syst Eng 2024 May 29;66(3):377-400. [doi: 10.1007/s12599-024-00871-w]
89. Ali O, Abdelbaki W, Shrestha A, Elbasi E, Alryalat MAA, Dwivedi YK. A systematic literature review of artificial intelligence in the healthcare sector: benefits, challenges, methodologies, and functionalities. Journal of Innovation & Knowledge 2023 Jan;8(1):100333. [doi: 10.1016/j.jik.2023.100333]
90. Jönsson T, Dell'Isola A, Lohmander LS, Wagner P, Cronström A. Comparison of face-to-face vs digital delivery of an osteoarthritis treatment program for hip or knee osteoarthritis. JAMA Netw Open 2022 Nov 01;5(11):e2240126 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.40126] [Medline: 36326763]
91. Cenamor J. Use of health self-management platform features: the case of a specialist ehealth app. Technological Forecasting and Social Change 2022 Dec;185:122066 [FREE Full text] [doi: 10.1016/j.techfore.2022.122066]
92. Foo MM, Roberts Z, Roden RC. Use of telemedicine services is associated with longer time to initiation of hormone therapy in transgender teens and young adults. Telemed J E Health 2025 Jan;31(1):119-123. [doi: 10.1089/tmj.2024.0302] [Medline: 39291313]
93. Hood K, Bergenstal RM, Cushman T, Gal RL, Raghinaru D, Kruger D, et al. Patient-reported outcomes improve with a virtual diabetes care model that includes continuous glucose monitoring. Telemed J E Health 2025 Jan;31(1):75-84. [doi: 10.1089/tmj.2024.0093] [Medline: 39166322]
94. Santos ATD, Machado CMS, Adamatti DF. Circadian rhythm and pain: mathematical model based on multiagent simulation. J Med Syst 2020 Aug 17;44(10):173. [doi: 10.1007/s10916-020-01622-6] [Medline: 32803513]
95. Teti SD, Murray LL, Orange JB, Kankam KS, Roberts AC. Telepractice assessments for individuals with aphasia: a systematic review. Telemed J E Health 2025 Jan;31(1):37-49. [doi: 10.1089/tmj.2024.0268] [Medline: 39178126]
96. Rakhshani T, Dolatkhah SM, Kashfi SM, Khani Jeihooni A. The effect of a self-learned virtual learning package on knowledge, attitude, and self-care behaviors of COVID-19 in people referred to health and treatment centers. BMC Public Health 2024 Jun 26;24(1):1710 [FREE Full text] [doi: 10.1186/s12889-024-19233-y] [Medline: 38926819]
97. Cadellans Arroniz A, Romero Rodríguez D, Zárate V, Dantony F, Madruga Parera M, Ortega Cebrian S, et al. Effectiveness of iso-inertial resistance training on muscle power in middle-older adults: randomized controlled trial. JMIR Aging 2025 Aug 21;8:e66414. [doi: 10.2196/66414] [Medline: 40840466]
98. Bennell KL, Campbell PK, Egerton T, Metcalf B, Kasza J, Forbes A, et al. Telephone coaching to enhance a home-based physical activity program for knee osteoarthritis: a randomized clinical trial. Arthritis Care Res (Hoboken) 2017 Jan;69(1):84-94. [doi: 10.1002/acr.22915] [Medline: 27111441]
99. Bennell K, Nelligan RK, Schwartz S, Kasza J, Kimp A, Crofts SJ, et al. Behavior change text messages for home exercise adherence in knee osteoarthritis: randomized trial. J Med Internet Res 2020 Sep 28;22(9):e21749 [FREE Full text] [doi: 10.2196/21749] [Medline: 32985994]
100. Brady TJ. Measures of self-efficacy: Arthritis Self-Efficacy Scale (ASES), Arthritis Self-Efficacy Scale-8 Item (ASES-8), Children's Arthritis Self-Efficacy Scale (CASE), Chronic Disease Self-Efficacy Scale (CDSES), Parent's Arthritis Self-Efficacy Scale (PASE), and Rheumatoid Arthritis Self-Efficacy Scale (RASE). Arthritis Care Res (Hoboken) 2011 Nov;63 Suppl 11:S473-S485 [FREE Full text] [doi: 10.1002/acr.20567] [Medline: 22588769]
101. Tore NG, Oskay D, Haznedaroglu S. The quality of physiotherapy and rehabilitation program and the effect of telerehabilitation on patients with knee osteoarthritis. Clin Rheumatol 2023 Mar;42(3):903-915 [FREE Full text] [doi: 10.1007/s10067-022-06417-3] [Medline: 36279075]

XSL•FO
RenderX

102. Batrakoulis A, Jamurtas AZ, Metsios GS, Perivoliotis K, Liguori G, Feito Y, et al. Comparative efficacy of 5 exercise types on cardiometabolic health in overweight and obese adults: a systematic review and network meta-analysis of 81 randomized controlled trials. Circ: Cardiovascular Quality and Outcomes 2022 Jun;15(6):109203. [doi: 10.1161/circoutcomes.121.008243]
103. Rangraz Jeddi F, Nabovati E, Mobayen M, Akbari H, Feizkhah A, Osuji J, et al. Health care needs, eHealth literacy, use of mobile phone functionalities, and intention to use it for self-management purposes by informal caregivers of children with burns: a survey study. BMC Med Inform Decis Mak 2023 Oct 23;23(1):236 [FREE Full text] [doi: 10.1186/s12911-023-02334-w] [Medline: 37872538]
104. Chen Y, Wang L, Ding B, Shi J, Wen T, Huang J, et al. Automated Alzheimer's disease classification using deep learning models with Soft-NMS and improved ResNet50 integration. Journal of Radiation Research and Applied Sciences 2024 Mar;17(1):100782. [doi: 10.1016/j.jrras.2023.100782]
105. Friji R, Chaieb F, Drira H, Kurtek S. Geometric deep neural network using rigid and non-rigid transformations for landmark-based human behavior analysis. IEEE Trans Pattern Anal Mach Intell 2023 Nov 1:1-13. [doi: 10.1109/tpami.2023.3291663]
106. Peng Z, Jiang B, Xu H, Feng W, Zhang J. Facial optical flow estimation via neural non-rigid registration. Comp Visual Med 2023 Mar;9(1):109-122. [doi: 10.1007/s41095-021-0267-z]
107. Li H, Zhang J, Li R, Zhang H, Zou L, Liu S. Swin-RGC: Swin-transformer with recursive gated convolution for substation equipment non-rigid defect detection. IEEE Access 2023;11:72655-72664. [doi: 10.1109/access.2023.3289874]
108. Wang X, Yang M, Ge Y, Tang M, Rao B, Chen Y, et al. Association of systemic inflammation and malnutrition with survival in nasopharyngeal carcinoma undergoing chemoradiotherapy: results from a multicenter cohort study. Front Oncol 2021;11:766398 [FREE Full text] [doi: 10.3389/fonc.2021.766398] [Medline: 34765561]
109. Stringer E, Rio LH, Leitz L, Prisman E, Borycki E, Kushniruk A, et al. Information needs of patients with head and neck cancer and their supports in relation to treatment management choices: scoping review. JMIR Cancer 2025 Aug 21;11:e64108 [FREE Full text] [doi: 10.2196/64108] [Medline: 40840481]
110. Hosseini SH, Ho N, Huntley S, Piper S, Crosland P, Skinner A, et al. Examining the impact of youth mental health services capacity growth trajectories and digital interventions on youth mental health outcomes: system dynamics modeling analysis. J Med Internet Res 2025 Aug 21;27:e71256 [FREE Full text] [doi: 10.2196/71256] [Medline: 40840438]
111. Joo J, Lee M, Yoon J, Cho H, Warrier G, Thrul J, et al. Boosting physical activity among individuals with low engagement through double-point incentives in a community-based mHealth intervention: retrospective observational study. JMIR Mhealth Uhealth 2025 Aug 21;13:e66227 [FREE Full text] [doi: 10.2196/66227] [Medline: 40840364]
112. Mbithi F, Donovan-Hall M, Bramley J, Steer J, Rossides C, Worsley P, et al. Evidence-generated sockets for transtibial prosthetic limbs compared with conventional computer-aided designs: a multiple-methods study from the patient's perspective. JMIR Rehabil Assist Technol 2025 Aug 21;12:e69962 [FREE Full text] [doi: 10.2196/69962] [Medline: 40840190]
113. Salam A, Dhurjati R, Vidyasagar K, Kaistha P, Esam H, Haghdoost F, et al. Development of the Double-Blind, Randomized Trials of Effects of Antihypertensive Medicines (DREAM) Database and characteristics of the included trials: protocol for an umbrella review and meta-analyses. JMIR Res Protoc 2025 Aug 21;14:e65205 [FREE Full text] [doi: 10.2196/65205] [Medline: 40840872]
114. Wang D, Liu Y. Medical science data value evaluation model: mixed methods study. JMIR Med Inform 2025 Aug 21;13:e63544 [FREE Full text] [doi: 10.2196/63544] [Medline: 40839787]
115. DeCamp M, Barnard JG, Ritger C, Helmkamp LJ, Begum A, Garcia-Hernandez S, et al. Engagement methods in brain tumor genomic research: multimethod comparative study. J Particip Med 2025 Aug 21;17:e68852 [FREE Full text] [doi: 10.2196/68852] [Medline: 40839870]
116. Philpot LM, Singla A, Dugani SB, Canning RE, Smith CM, DeZutter MA, et al. Patient-derived design principles for technology-enabled healing at home following hospital discharge: mixed methods study. JMIR Hum Factors 2025 Aug 20;12:e72913 [FREE Full text] [doi: 10.2196/72913] [Medline: 40835402]
117. Pargaonkar S. A comprehensive review of performance testing methodologies and best practices: software quality engineering. IJSR 2023 Aug 5;12(8):2008-2014. [doi: 10.21275/sr23822111402] [Medline: 269911398]

Abbreviations

AFEW: Acted Facial Expressions in the Wild
AQoL-6D: Assessment of Quality of Life—6 Dimensions
CNN: convolutional neural network
CONSORT: Consolidated Standards of Reporting Trials
EffiFCNN-ResNet: Efficient Fully Convolutional Neural Network with Residual Network
FCNN: fully convolutional neural network
FCOS: fully convolutional one-stage object detection
FN: false negative
FP: false positive
GIoU: generalized intersection over union

XSL•FO
RenderX

IoU: intersection over union
MtCNN: multitask cascaded convolutional neural network
NPC: nasopharyngeal carcinoma
PAN: path aggregation network
RGB: red-green-blue
SPP: spatial pyramid pooling
SUS: System Usability Scale
TN: true negative
TP: true positive

Edited by L Buis; submitted 12.Feb.2025; peer-reviewed by JCL Chow, A Zhang; comments to author 02.Jul.2025; accepted 09.Feb.2026; published 10.Mar.2026.

Please cite as:
Wu T, Han T, Zhang X, Dai Y, Meng X
A Telemedicine App for Nonrigid Facial Rehabilitation Training Enhanced by Efficient Fully Convolutional Neural Network With Residual Network (EffiFCNN-ResNet) to Improve Accessibility for Patients With Nasopharyngeal Carcinoma Cancer: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e72560
URL: https://mhealth.jmir.org/2026/1/e72560
doi:10.2196/72560
PMID:

©Tong Wu, Ting Han, Xiaoju Zhang, Yumei Dai, Xiaoyan Meng. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Tailored and Interactive Mobile Telehealth Contraceptive Counseling Compared With In-Person Care: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Maja Weinryb[1,2], MD; Wilhelmus JA Grooten[3,4], PhD; Andrea Koris[5], MSc; Margit Endler[1,2], MD, PhD

[1]Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Stockholm, Sweden

[2]Division of Social and Behavioural Sciences, School of Public Health, University of Cape Town, Cape Town, Western Cape, South Africa

[3]Department of Neurobiology, Care Sciences and Society, Division of Physiotherapy, Karolinska Institutet, Stockholm, Sweden

[4]School of Health and Welfare, Dalarna University, Dalarna, Sweden

[5]Global Women's Institute, George Washington University, Washington, DC, United States

Corresponding Author:
Margit Endler, MD, PhD
Department of Women's and Children's Health
Karolinska Institutet
Tomtebodavägen 18
Stockholm, Stockholm
Sweden
Phone: 46 8 524 800
Email: margit.endler@ki.se

Abstract

Background: Use of effective contraceptive methods (ECMs) reduces maternal mortality. Person-centered counseling increases uptake, but barriers to high-quality counseling persist. Telehealth may improve access to comprehensive contraceptive care, but its effectiveness remains unclear.

Objective: This study assesses the effectiveness and acceptability of tailored, interactive telehealth contraceptive counseling (TECC) compared with in-person counseling.

Methods: We conducted a systematic review and meta-analysis of English-language randomized controlled trials (RCTs) comparing TECC with in-person counseling for women and girls of any age and setting. We searched MEDLINE, Embase, Web of Science, and the Cochrane Library from inception through October 15, 2025. Outcomes were use of ECM less than 6 months from intervention (primary outcome); use of ECM at 6-12 months; use of long-acting reversible contraception (LARC); choice of ECM; choice of LARC; satisfaction with counseling; and method switching. Two researchers assessed risk of bias (RoB) using Cochrane RoB 2, certainty of evidence using GRADE (Grading of Recommendations Assessment, Development, and Evaluation), and performed meta-analysis using a random-effects model. The protocol was registered with PROSPERO (International Prospective Register of Systematic Reviews) a priori (CRD42023404402). No specific funding was received.

Results: Eight RCTs (RoB: low, n=4; moderate, n=2; and high, n=2) and 1 cluster RCT (moderate-to-high RoB) were included in the review, with 5353 participants across all included studies. Eight studies evaluated TECC as an adjunct to in-person care (hybrid model) and 1 as a standalone model, of which 7 and 1, respectively, contributed outcome data for the meta-analyses. Certainty of evidence was low to very low. The pooled effect of 4 studies of TECC on ECM use less than 6 months showed no clear evidence of an effect (relative risk [RR] 1.10; 95% CI 0.95-1.29). The pooled effect of 4 studies of TECC on choice of ECM (RR 1.07; 95% CI 0.96-1.18) likewise showed no clear effect. The pooled effect of 4 studies of TECC on ECM use at 6-12 months showed a small but statistically significant (*P*=.04) effect in favor of TECC, narrowly excluding the null (RR 1.07; 95% CI 1.002-1.130). Because of high statistical and clinical heterogeneity (I2=87%-96%), results for LARC choice (2 studies) and LARC use at 0-6 months (2 studies) and 6-12 months (2 studies) were narratively described. The evidence for these outcomes was very uncertain. Narrative analysis of satisfaction across 2 studies showed no difference in effect. There were no data on method switching to support the analysis.

Conclusions: Current evidence suggests, with low certainty, that adjunct TECC, when delivered alongside in-person care, shows little to no effect on contraceptive use compared with in-person care. For use at 12 months, method choice, LARC use, and satisfaction compared with in-person care, the evidence is very uncertain. Future research should prioritize adequately powered

XSL•FO
RenderX

evaluations of standalone models and assess how tailoring, timing, and delivery influence effectiveness, including long-term use and method switching.

(JMIR Mhealth Uhealth 2026;14:e88887) doi:10.2196/88887

KEYWORDS

contraceptive counseling; systematic review; telemedicine; randomized controlled trials; meta-analysis; long-acting reversible contraception; mobile health

Introduction

Contraception is one of the most cost-effective public health strategies for reducing maternal mortality and improving reproductive, perinatal, and child health. Many unintended pregnancies and maternal deaths globally could be averted by satisfying the unmet need for modern contraceptives [1-3]. Women and girls continue to face barriers to access due to misconceptions and concerns about side effects, social stigma, provider bias, geographic barriers, and health system constraints [4]. Barriers disproportionately affect already vulnerable populations. Person-centered contraceptive counseling (CC) aims to support informed contraceptive decision-making by addressing misconceptions and side effects and tailoring counseling to the individual's preferences, values, needs, medical eligibility, and reproductive life goals. High-quality counseling has been shown to improve the use and continuation of effective modern contraception [5].

The core components of person-centered CC, such as eliciting individual preferences and addressing misconceptions and side effect concerns, could in principle be delivered remotely through interactive telehealth formats. The WHO defines telehealth as the use of electronic information and communication technologies for the "delivery of health care services, where patients and providers are separated by distance" [6]. Telehealth can offer a convenient, low-cost, and private alternative to in-person care. It can also expand access to person-centered care in settings where geographic or system constraints limit access through messaging, web-based tools, mobile apps, or virtual consultations [7].

Telehealth contraception services have been used for more than a decade, with commercial services expanding in response to technological advances and the COVID-19 pandemic. These digital services range from prescription and delivery services, 1-way messaging, and decision aids to more comprehensive counseling delivered by telephone or video. Several US-based observational studies report high acceptability, but evidence of effectiveness remains limited and heterogeneous [8-12]. Previous reviews have grouped heterogeneous digital contraceptive services, including 1-way messaging and nontailored models. However, none have specifically examined tailored and interactive telehealth contraceptive counseling (TECC) [11-15].

Our definition of TECC entails bidirectional communication and individualized content aligned with users' preferences, reproductive goals, and medical eligibility. Although TECC can conceptually function as a standalone counseling modality, most existing telehealth contraception trials have evaluated hybrid models delivered as an adjunct to in-person care, and the extent to which TECC can substitute for in-person counseling remains unknown. We aimed to systematically assess the effectiveness and acceptability of standalone and hybrid TECC compared with any form of standard in-person contraceptive care, including verbal or written information delivered individually or in a group setting.

Methods

Search Strategy

We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (see Multimedia Appendix 1) [16]. The protocol was registered a priori with PROSPERO (International Prospective Register of Systematic Reviews; registration number CRD42023404402) [17]. We searched MEDLINE, Web of Science, Embase, the Cochrane Library from inception to October 15, 2025, using MeSH (Medical Subject Headings) and free-text terms, with a validated randomized controlled trial (RCT) filter. A snowball search was conducted to identify references cited by and citations of eligible studies identified through the database searches. The full search strategies for all databases are available in Multimedia Appendix 2.

Eligibility Criteria

We included single and cluster RCTs comparing TECC with routine in-person CC among women and girls of any age or setting. Only full, published, English-language reports were eligible; abstracts and study protocols were excluded. Intervention eligibility required meeting 3 core domains: telehealth delivery modality, interactivity, and tailoring. Telehealth delivery modalities included phone or video calls, voice or SMS text messaging, mobile apps, and websites. Interactivity implied bidirectional communication to some degree, excluding, for example, unidirectional information delivery or reminders. Tailoring meant that interventions were individualized to users' input, generating adapted output based on at least one of the following domains: contraceptive preferences, reproductive goals, medical history, or method-specific information on benefits, side effects, and effectiveness. Interventions could be delivered in real time or asynchronously, with or without human involvement. Interventions could be delivered before, during, or after a clinical visit and could complement or substitute for in-person clinical care. We categorized models in which TECC was delivered as an adjunct to in-person care as hybrid models. Hybrid models were eligible if the telehealth component independently met the tailoring and interactivity criteria. Shared decision-making tools used in provider-led consultations were excluded. Our

XSL•FO
RenderX

comparator was any form of routine CC delivered in person as verbal or written information, provided individually or in groups. Studies using telehealth as a comparator were excluded. The rationale for excluding other digital interventions as comparators was to assess the effect of the telehealth component on CC. An overview of the concepts in our definition of TECC is provided in Multimedia Appendix 3.

Outcomes

Our primary outcome was the use of any effective contraceptive method (ECM) up to 6 months after the intervention. We distinguished use from continuation or adherence to a specific method. ECM was defined as long-acting reversible contraception (LARC), including the progestin implant and hormonal or copper intrauterine devices; combined oral contraceptives; progestin-only pills; progestin injections; the contraceptive patch; the contraceptive ring; and sterilization—methods with typical-use failure rates of less than 10% per year [18]. Additional outcomes were (1) ECM use at 6-12 months; (2) LARC use at less than 6 months; (3) LARC use at 6-12 months; (4) choice of ECM at the time of counseling; (5) choice of LARC at the time of counseling; (6) satisfaction with counseling at less than 6 months; and (7) method switching.

"Use" was assessed based on self-reported or recorded use in follow-up surveys or clinical records. Where multiple time points were reported, we prioritized the time point that enabled comparison across the greatest number of studies or, secondarily, the latest time point within the specified range. "Choice" was defined as documented intention to use a specific method immediately after counseling. We defined "switching" as a documented change from no method or a less effective method to a more effective ECM. "Satisfaction" included any reported measure of satisfaction with counseling up to 6 months after the intervention.

Study Selection

After deduplication of the final search results, we imported records into Rayyan for screening. Three reviewers (MW, AK, and ME) independently screened titles and abstracts. We unblinded the selection process after the initial screening and resolved discrepancies through discussion. We retrieved and assessed full-text articles against the inclusion criteria, and disagreements were resolved through discussion until consensus was reached. Records excluded at the full-text review stage and the reasons for exclusion are provided in Multimedia Appendix 4.

Data Collection and Quality Assessment

Data on publication year, study design, setting, participant characteristics, intervention, comparator, and raw data for each outcome were extracted by one reviewer (MW) from full-text articles, supported by GPT-4o (OpenAI, May 2024) as a preliminary extraction tool for nonoutcome study characteristics (population, setting, intervention, and comparator descriptions). All AI-assisted outputs were reviewed and corrected against the source articles by MW and then verified by ME. Studies in which data relevant to our outcomes were grouped in a way that did not permit the extraction of stratified data, or in which outcome data were missing, were excluded from the analysis.

Risk of bias (RoB) was assessed independently by reviewers MW and ME for each outcome using the Cochrane RoB 2 tools for RCTs and cluster RCTs, respectively [19]. Five domains were evaluated, including (1) bias arising from the randomization process, (2) bias due to deviations from intended interventions, (3) bias due to missing outcome data, (4) bias in outcome measurement, and (5) bias in the selection of the reported result. Discrepancies were resolved through discussion until consensus was reached.

Assessment of the overall certainty of evidence was performed using the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) approach for each outcome. GRADE ratings were independently assessed by MW and ME; discrepancies were resolved through discussion until consensus was reached [20]. As we included only RCTs, the certainty of evidence started as high and was downgraded based on RoB, inconsistency, imprecision, indirectness, and publication bias. Statistical heterogeneity was quantified using the I^2 statistic in accordance with the Cochrane Handbook for Systematic Reviews of Interventions, with 50%-90% potentially representing substantial heterogeneity and 75%-100% representing considerable heterogeneity. Our interpretation was informed by the magnitude and direction of effects, the number and size of contributing studies, and the plausibility of clinical or methodological diversity. We considered downgrading for inconsistency if I^2 exceeded 50%. Imprecision was assessed using the optimal information size (OIS) criterion, in accordance with GRADE guidance for systematic reviews [21]. For each outcome eligible for meta-analysis, OIS was determined using the reference values reported by Guyatt et al [21], applying the observed pooled control event rate and the GRADE-recommended default relative risk (RR) reduction of 25%, with α=.05 and power of 80%. Certainty of evidence was downgraded by 1 level for imprecision when the total sample size contributing to the pooled estimate fell below the corresponding OIS threshold. Where the OIS threshold was met but the 95% CI of the pooled effect crossed or narrowly excluded the line of no effect, we additionally considered whether the CI excluded an appreciable effect, using the GRADE default RR reduction threshold of 25% for appreciable benefit or harm. For outcomes not eligible for meta-analysis (LARC use at 6-12 months, choice of LARC, and satisfaction with counseling), imprecision was judged by comparing the combined sample size across contributing studies with the OIS reference and by inspecting the width of individual study CIs in relation to the line of no effect. Imprecision judgments for each outcome are documented in the footnotes to Table 3.

Data Synthesis

We grouped studies for each synthesis according to outcome and follow-up time point. Studies reporting data in a comparable format were included in random-effects meta-analyses when at least 2 comparable studies were available. We calculated RRs with 95% CIs for all outcomes based on available data from intention-to-treat study groups, deriving them from raw data or alternative effect measures, and presented pooled results as forest plots. Heterogeneity was assessed using the I^2 statistic and interpreted in relation to the direction and consistency of

XSL•FO
RenderX

effects across studies. For outcomes for which data quality or heterogeneity did not support meta-analysis, we reported results descriptively and interpreted them exploratorily. Prespecified exploratory subgroup analyses were performed by country income level (low- and middle-income countries [LMICs] versus high-income countries [HICs]). One included cluster RCT (Dehlendorf et al [22]) was randomized at the provider level (28 providers; mean cluster size of ≈27). Raw event counts per study arm were extracted from the published article. As the reported provider-level intracluster correlation coefficient was approximately 0, the corresponding design effect was ≈1.0, and no design-effect adjustment was applied. Sensitivity analyses examined the effect of publication year (before vs after 2015), which stratifies the approximately 20-year period during which telemedicine has been used for reproductive health, and low versus moderate or high RoB. Where meta-analysis was not feasible, we summarized findings narratively using incidence ranges. We conducted analyses using SPSS (version 30; IBM Corp) and Review Manager (RevMan 5, desktop version).

Ethics Considerations

This review used only publicly available data and required no ethical approval. There was no funding source for this study. GPT-4o (OpenAI, May 2024) was used as a preliminary tool for data extraction of setting, population, intervention, and comparator characteristics from full-text articles; for the generation of a conceptual figure; and to support language editing. All AI-assisted data extraction outputs were verified and corrected against the source articles by reviewer MW and subsequently by ME. All analyses, interpretations, and conclusions were independently conducted by the authors. Outcomes related to potential harms were not prespecified, which we acknowledge as a limitation of the protocol.

The initial protocol was amended in the following ways before data extraction: outcomes were reorganized to specify a single primary outcome; the population eligibility criterion was clarified to restrict inclusion to women and girls of any age (including adolescents); men and couples were excluded; and subgroup and sensitivity analyses were added. The revised protocol also clarified our a priori intent to restrict the search to English-language publications.

Patient and Public Involvement

Patients or members of the public were not involved in the design, conduct, reporting, or dissemination of this systematic review.

Results

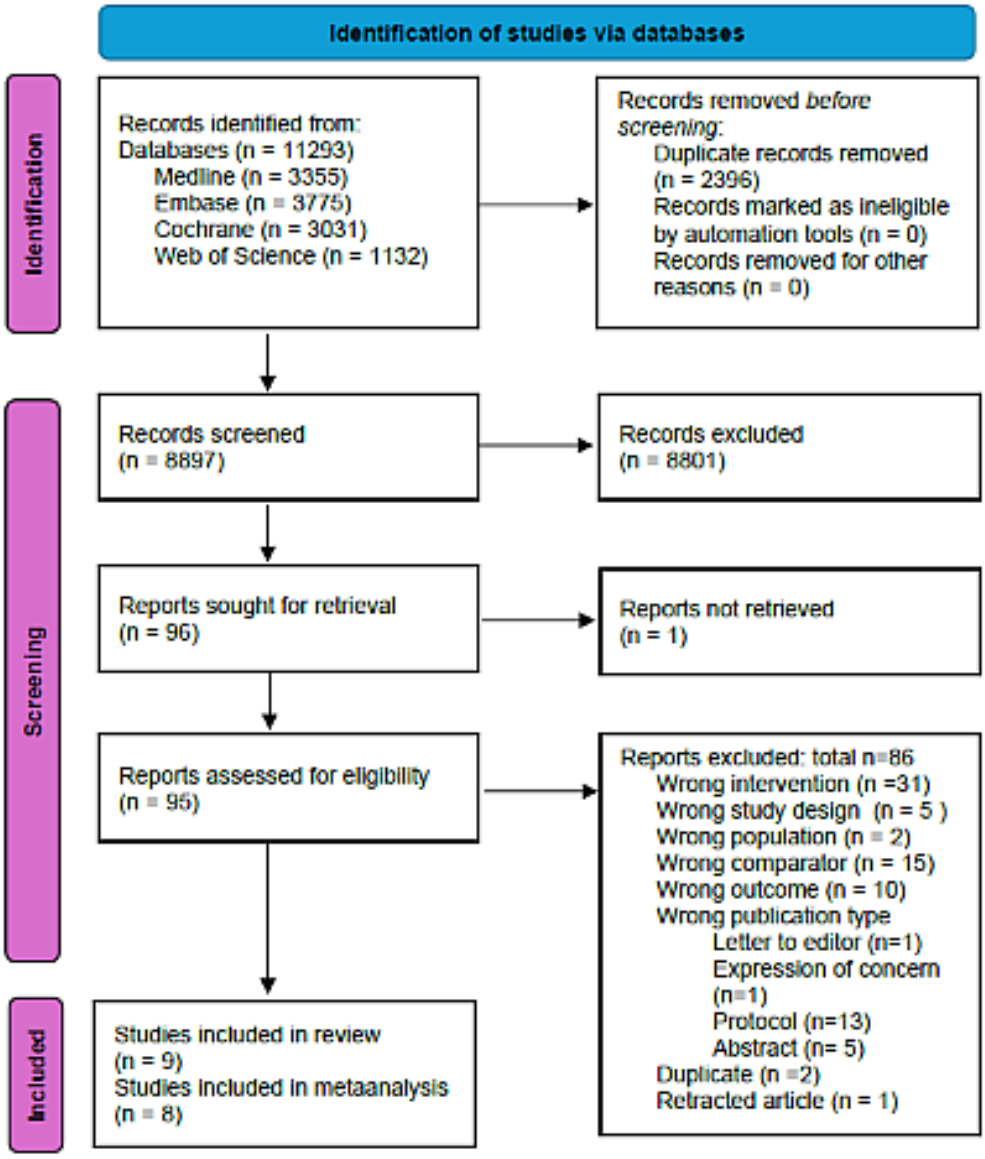
Study Selection and Characteristics

From 8897 records screened, 8 individual RCTs and 1 cluster RCT met the inclusion criteria, involving 5353 participants from 5 HICs and 4 middle-income countries. The PRISMA flowchart illustrates the selection process (Figure 1), and Table 1 presents the study characteristics and sample sizes. Included studies were conducted in antenatal care [23,24], postabortion care [25-27], and routine contraceptive services [22,28-30]. Sample sizes ranged from 86 to 1475 participants.

All interventions fulfilled the 3 core components of our definition of TECC—telehealth delivery modality, interactivity, and tailoring—but the degree of interactivity and tailoring varied. Eight were hybrid models, delivered in addition to, before, during, or after an in-person clinical visit. Only one study [27] tested a standalone model that substituted for in-person care using telephone calls, but recruitment was terminated after only 125 of the 1222 (10.23%) planned enrollment due to changes in the comparator condition in response to the COVID-19 pandemic. The TECC modality and degree of interactivity ranged from potentially interactive voice [25,26] or SMS text messaging [23] with interaction on demand; asynchronous, self-guided computer- [28,29] or tablet-based modules [22,30]; to video or telephone calls delivered in real time [24,27]. Provider involvement in the TECC component, therefore, ranged from minimal to high. Tailoring also varied, ranging from SMS text or voice messages tailored to an already chosen method [23,25,26] to decision aids limited to life preferences and excluding medical criteria [29], comprehensive tailored decision aids [22,28,30], and real-time individualized assessments [24,27].

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.



XSL•FO
RenderX

Table 1. Study characteristics.

Study[a]	Country (income classification)[b]	Setting	Population	Study design	Total sample size, n	Participants per arm, n
Reiss et al [25]	Bangladesh (lower middle income)	Public postmenstrual regulation services	Women aged 18-49 following menstrual regulation, with access to a personal mobile phone	RCT[c] (2 arms)	969	Intervention: 485 and control: 484
Stephenson et al [29]	United Kingdom (high income)	Sexual and reproductive health clinics and online booking system	Women aged 15-30 with current or future contraceptive needs	RCT (2 arms)	927	Intervention: 464 and control: 463
Harrington et al [23]	Kenya (lower middle income)	Kenyan public hospitals (antenatal to postpartum)	Women and girls aged >14 years, pregnant ≥28 weeks' gestation, and HIV negative, with daily phone access (partners optional)	RCT (2 arms)	260	Intervention: 130 and control: 130
Garbers et al [28]	United States (high income)	Urban family planning clinics	Women aged >16 years at risk of unintended pregnancy and attending family planning visits	RCT (3 arms)[d]	1475	Intervention: 985 and control: 490
Saglam Aksut and Inam [24]	Turkey (upper middle income)	Antenatal primary care	Pregnant women registering for antenatal care and using smartphones	RCT (2 arms)	86	Intervention: 43 and control: 43
Smith et al [26]	Cambodia (lower middle income)	Marie Stopes abortion clinics in peri-urban and rural settings	Women aged >17 years seeking abortion services, with access to a personal or predominantly private smartphone	RCT (2 arms)	500	Intervention: 249 and control: 251
Dehlendorf et al [22]	United States (high income)	Urban family planning centers, public health centers, college student health centers, and hospital outpatient clinics	Women aged 15-45 seeking contraception and not planning pregnancy within the next 7 months	Cluster RCT (2 arms)	Providers: 28; participants: 758	Providers—intervention: 15 and control: 13; participants—intervention: 407 and control: 351
Madden et al [30]	United States (high income)	Obstetrics/gynecology clinics at academic centers	Women aged 18-45 at risk of unintended pregnancy and planning to discuss contraception	RCT (2 arms)	253	Intervention: 167 and control: 86
Reynolds-Wright et al [27]	United Kingdom (high income)	Public abortion care center in Edinburgh	Pregnant individuals aged >16 years seeking medical abortion at home	RCT (2 arms)	125	Intervention: 63 and control: 62

[a]Studies are listed in the order in which the data were first extracted.

[b]Country income classes according to The World Bank classification.

[c]RCT: randomized controlled trial.

[d]Only 1 intervention arm met the inclusion criteria and is reported in this review.

Comparators were standard in-person CC, with or without printed materials. In 1 trial [29], controls were offered delayed access to TECC after study completion. Table 2 outlines the intervention and comparator characteristics. Studies grouped by outcome and RoB assessment are presented in Figure 2 (also see [22-30]). Multimedia Appendix 5 details study outcomes and findings, and Multimedia Appendix 6 details how each intervention met the tailoring and interactivity criteria, as well as additional intervention characteristics that may have modified effects. GRADE-assessed certainty of evidence ranged from low to very low across outcomes. The GRADE assessment is presented in the summary of findings (Table 3), and a breakdown of each GRADE assessment is provided in Multimedia Appendix 7; 7 hybrid models and 1 standalone model contributed outcome data to the meta-analyses. One hybrid model contributed data for the outcome of satisfaction, which was reported descriptively [24].

XSL•FO
RenderX

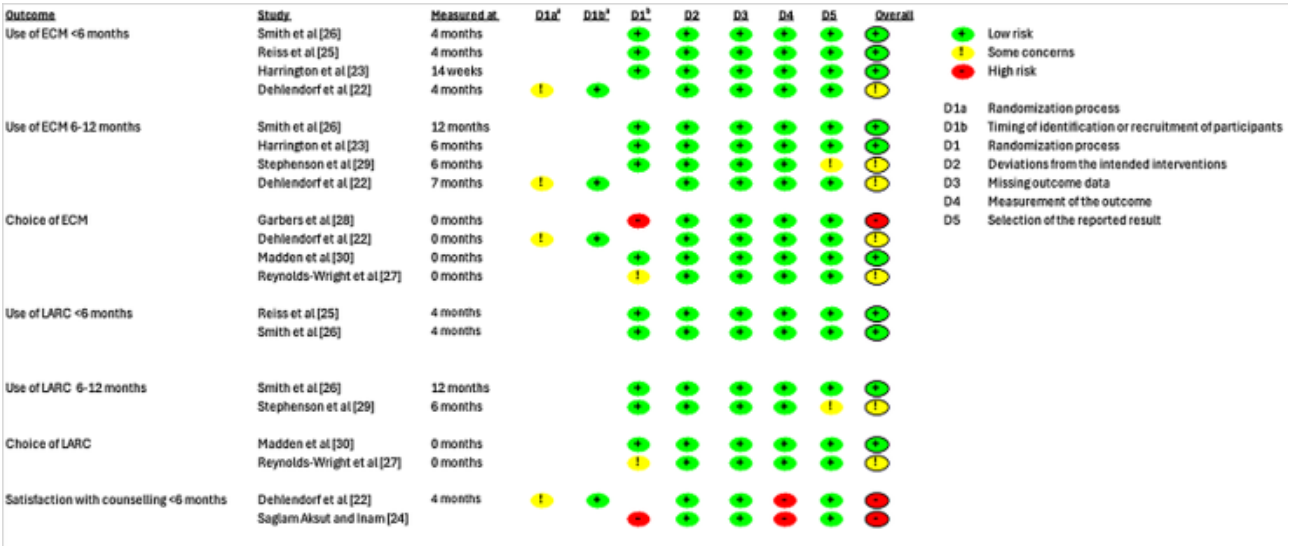
Table 2. Intervention and comparator characteristics.

Study[a]	Relation to clinical care	Intervention	Timing and duration	Comparator
Reiss et al [25]	Hybrid/adjunct	Automated interactive voice messages delivered to mobile phones providing contraceptive method information and addressing concerns; messages tailored to the method chosen at menstrual regulation and updated if the method is switched; option to request callbacks for additional individual counseling	Delivered during the 4 months following in-person menstrual regulation; included 7 weekly messages and 4 fortnightly messages over 4 months	Routine in-person menstrual regulation counseling care without digital follow-up
Stephenson et al [29]	Hybrid/adjunct	Mobile-optimized website with videos, frequently asked questions, and a tailored decision aid generating 3 matched contraceptive methods based on user-reported preferences; designed to support informed choice before in-person counseling	Self-directed use before a clinical appointment	Routine in-person contraceptive counseling; website access offered after study follow-up
Harrington et al [23]	Hybrid/adjunct	SMS text messages tailored to user language and gestational age/postpartum stage, prompting user responses; nurses engaged in asynchronous 2-way communication; optional partner enrollment included	Weekly automated messages delivered from 28 weeks of gestation to 6 months postpartum	Routine in-person antenatal and postpartum care with no SMS text message follow-up
Garbers et al [28]	Hybrid/adjunct	Audio-assisted touchscreen decision aid eliciting user preferences and medical history, producing a traffic-light classification of contraceptive methods (green=most suitable; yellow=less effective or less suitable; and red=medically contraindicated methods); designed to support precounseling decision-making	Self-directed, 1-time use on a touchscreen laptop in the waiting room before a clinical appointment	Routine in-person contraceptive counseling and attention-matched control (demographic survey on laptop and generic printed handout)
Saglam Aksut and Inam [24]	Hybrid/adjunct	Video counseling calls using a structured handbook and visual demonstration of contraceptive methods before in-person visits; real-time tailored information provided with opportunity for user questions; partners included where possible	Provided in the third trimester of pregnancy as 2 sessions of approximately 50 minutes each, delivered 2 weeks apart	Routine in-person contraceptive counseling during antenatal visit at 24 weeks of gestation
Smith et al [26]	Hybrid/adjunct	Automated interactive voice messages providing tailored contraceptive information based on current method use; optional reminders for method adherence and option to request counseling callbacks	Initiated within 1 week after abortion and continued up to 3 months after abortion; fortnightly messages	Routine in-person postabortion care; no digital follow-up
Dehlendorf et al [22]	Hybrid/adjunct	Web-based interactive decision aid incorporating educational content, user preferences, reproductive life goals, and medical eligibility criteria; generates a personalized summary report with recommendations for shared decision-making during consultation	Self-directed, 1-time use on a tablet in the waiting room before a clinic visit	Routine in-person contraceptive counseling
Madden et al [30]	Hybrid/adjunct	Tablet-based interactive decision aid assessing user preferences, reproductive and medical history, prior contraceptive use, and preferences; generates top 3 tailored contraceptive recommendations and provides printed personalized information	Self-directed, 1-time use on a tablet in the waiting room before a clinic visit	Routine in-person contraceptive counseling with attention-matched control (demographic and reproductive survey on tablet plus generic handout)
Reynolds-Wright et al [27]	Standalone	Telephone counseling providing abortion-related counseling and postabortion contraceptive information; tailored real-time discussion during the call	Delivered at booking or another chosen time; replaced in-person counseling; followed by in-person ultrasound and medication provision; 1-time session	Routine in-person counseling

[a]Studies are listed in the order in which the data were first extracted.

XSL•FO
RenderX

Figure 2. Studies grouped by outcome and Risk of Bias (RoB) related to outcome. [a]: Cluster-randomized trial domain. [b]: Individual randomized trial domain.



Risk of Bias

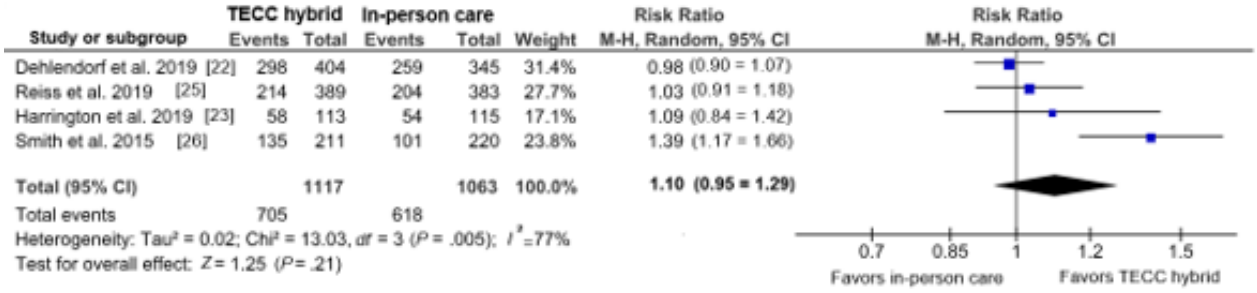
RoB assessments for each outcome are presented in Figure 2. Four studies [23,25,26,30] had low RoB across all domains and outcomes. The most frequent concerns related to the randomization process, including sequential allocation to control and intervention arms, imbalanced distribution of study groups across study sites, changes to the randomization procedure [28], and small but potentially meaningful baseline imbalances [22,24,27,28].

Stephenson et al [29] had moderate RoB in domain 5 (selection of the reported result) due to post hoc changes in study outcomes when the trial expanded from a feasibility study to an effectiveness trial without a prespecified analysis plan. We therefore judged this outcome to have moderate RoB. We classified the studies by Saglam Aksut and Inam [24] and Dehlendorf et al [22] as having high RoB for satisfaction with counseling outcomes because outcome measurement was conducted by study staff, creating a risk of social desirability bias.

Use of ECMs

Four studies (n=2180) reported on our primary outcome, use of ECM less than 6 months after the intervention. The meta-analysis indicated a small positive but nonsignificant (P=.21) effect of TECC compared with standard care (SC; RR 1.10; 95% CI 0.95-1.29; I^2=77%; Figure 3; see also [22,23,25,26]). The certainty of evidence was low, downgraded for inconsistency and imprecision (Table 3).

Figure 3. Forest plot of four studies comparing hybrid telehealth contraceptive counseling (TECC) with in-person care for effective contraceptive method use within 6 months of the intervention. M-H: Mantel–Haenszel method.



XSL•FO
RenderX

Table 3. Summary of findings and GRADE[a].

Outcome and follow-up	Patients (studies), N	Relative effect, risk ratio (95% CI)	Absolute effects (95% CI)			Certainty	What happens
			Standard care	TECC[b]	Difference		
Use of ECM[c] <6 months	2180 (4 RCTs[d])[e]	1.10 (0.95-1.29)	582 per 1000	646 per 1000 (553-751)	64 more per 1000 (from 29 fewer to 169 more)	Low[f,g]	TECC may result in little to no difference in use of any effective contraceptive method at 0-6 months.
Use of ECM at 6-12 months	2043 (4 RCTs)[e]	1.070 (1.002-1.130)	646 per 1000[e]	685 per 1000 (648-730)[e]	39 more per 1000 (from 1 more to 84 more)	Low[h,i]	TECC may result in a slight increase in use of any effective contraceptive method at 6-12 months.
Use of LARC[j] <6 months	1203 (2 RCTs)	1.62	129 per 1000	210 per 1000	81 more per 1000	Very low[k,l]	The evidence is very uncertain about the effect of TECC on use of LARC at 0-6 months.
Use of LARC 6-12 months	1042 (2 RCTs)[e]	Not estimable	N/A	N/A[m]	N/A	Very low[n,o,p]	The evidence is very uncertain about the effect of TECC on use of LARC at 6-12 months.
Choice of ECM	2590 (4 RCTs)	1.07 (0.96-1.18)	564 per 1000	603 per 1000 (547-660)	39 more per 1000 (from 17 fewer to 96 more)	Very low[g,q,r]	The evidence is very uncertain about the effect of telemedicine on the choice of effective contraceptive methods.
Choice of LARC	366 (2 RCTs)	Not estimable	N/A	N/A	N/A	Very low[n,s]	The evidence is very uncertain about the effect of telemedicine on the choice of LARC.
Satisfaction with counseling <6 months	749 (2 RCTs)	Not estimable	N/A	N/A	N/A	Very low[t,u,v]	The evidence is very uncertain about the effect of TECC on satisfaction with counseling up to 6 months.
Use of ECM <6 months in LMIC[w] or low RoB[x]	580 (2 RCTs)[e]	1.20 (1.03-1.30)	500 per 1000[e]	600 per 1000 (515-695)[e]	100 more per 1000 (from 15 more to 195 more)	Very low[y,z]	The evidence is very uncertain about the effect of TECC on use of any effective contraceptive method at 0-6 months in LMICs.

[a]GRADE: Grading of Recommendations Assessment, Development, and Evaluation.

[b]TECC: telehealth contraceptive counseling.

[c]ECM: effective contraceptive method.

[d]RCT: randomized controlled trial.

[e]Numbers calculated from percentages, rounded to whole numbers.

[f]Heterogeneity I^2=0.77.

[g]Optimal information size adequate. Downgraded for imprecision because the effect estimate CI overlaps with no effect.

[h]Moderate RoB in 2/4 studies.

[i]Optimal information size is adequate, but the lower bound of the effect estimate CI (1002) fails to reliably exclude no effect.

[j]LARC: long-acting reversible contraception.

[k]Heterogeneity I^2=0.96.

[l]Optimal information size not met. Very wide CI of the effect estimate. Lower bound below 1; fails to exclude no effect.

[m]N/A: not applicable.

[n]Moderate RoB in 1/2 studies

[o]Heterogeneity I^2=0.87.

[p]Results not pooled. Smith et al [26] showed a very wide CI of effect estimate, whereas for Stephenson et al [29], the lower bound of the CI fails to exclude no effect.

[q]High and moderate RoB in 2/2 studies, respectively.

[r]Heterogeneity I^2=0.53.

[s]Total sample sizes are small. The lower bound of the effect estimate CI fails to exclude no effect.

XSL•FO
RenderX

[t]High RoB in 2/2 studies

[u]Data cannot be pooled due to different measurement methods. One study showed no difference, and the other study suggested higher satisfaction in the intervention group.

[v]Data cannot be pooled. The lower bound of the effect estimate CI (Dehlendorf et al [22]) fails to exclude no effect.

[w]LMIC: low- and middle-income country.

[x]RoB: risk of bias.

[y]Heterogeneity I^2=0.7.

[z]Wide CI of effect estimate. Lower bound below 1.

Four studies (n=2043) reported data on ECM use at 6-12 months after the intervention. The meta-analysis showed a small positive effect that reached statistical significance (*P*=.04), although the CI only narrowly excluded the null (RR 1.07; 95% CI 1.002-1.130; I^2=4%; Figure 4; see also [22,23,26,29]). The certainty of evidence was low, downgraded for RoB and imprecision (Table 3).

Our subgroup and sensitivity analyses assessed whether the effect of TECC on ECM choice and use varied by country income level, RoB, and publication period. Two subgroups overlapped: studies evaluating ECM use in LMICs also had low RoB. These analyses included a small number of studies, and the results should be interpreted with caution. Pooled results for ECM use at 6-12 months among studies conducted in LMICs with low RoB (n=580) showed that TECC significantly (*P*=.02) increased use (RR 1.20; 95% CI 1.03-1.30; I^2=0%; Figure 5; see also [23,26]). The certainty of evidence was low, downgraded for serious imprecision (Table 3). We found no evidence of an effect on ECM choice or ECM use at 6-12 months in studies from HICs with moderate-to-high RoB, or on ECM use less than 6 months in LMICs. Comparisons by publication period (pre- versus post-2015) showed no differences in ECM choice or use. Subgroup analyses based on geographic region and ECM use less than 6 months in HICs were not possible because of the small number of studies.

Figure 4. Forest plot of four studies comparing hybrid telehealth contraceptive counseling (TECC) with in-person care for effective contraceptive method use at 6-12 months after intrevention. M-H: Mantel–Haenszel method.

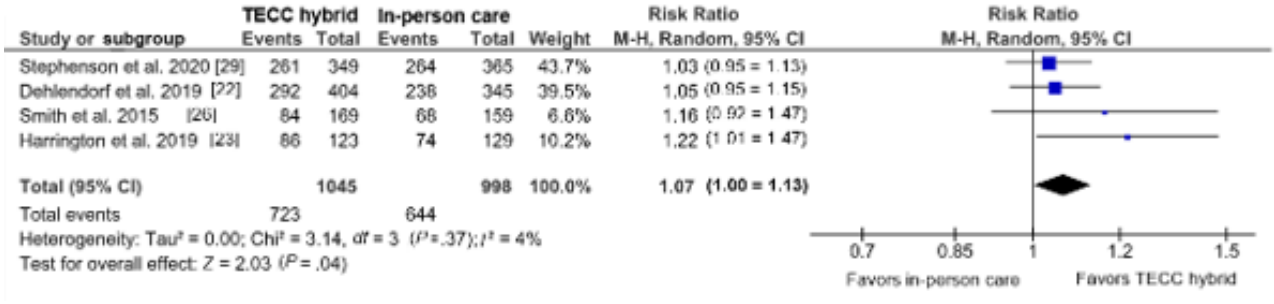
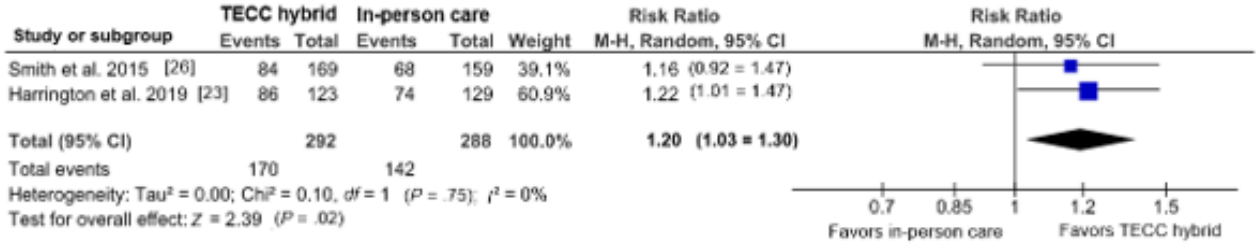


Figure 5. Forest plot of two low-risk-of-bias studies conducted in low- and middle-income countries, comparing hybrid telehealth contraceptive counseling (TECC) with in-person care for effective contraceptive method use at 6-12 months after intervention. M-H: Mantel–Haenszel method.



LARC Use

Two studies (n=1203) reported on LARC use at 0-6 months. Both studies were conducted in low-income settings in a postabortion context and used interactive voice messaging for TECC. However, statistical heterogeneity was high (I^2=96%), and pooled estimation was not considered appropriate. Reiss et al [25] reported no significant difference between groups (RR 0.80; 95% CI 0.56-1.14), whereas Smith et al [26] found an effect in favor of the telehealth intervention (RR 3.35; 95% CI 2.07-5.40). The certainty of evidence was very low, downgraded for moderate RoB and serious imprecision (Table 3).

Two studies (n=1042) reported LARC use at 6-12 months (Smith et al [26] and Stephenson et al [29]). Statistical heterogeneity was very high (I^2=87%) across 2 studies with different populations (postabortion care in Cambodia vs routine sexual health clinic attendees in the United Kingdom) and delivery modalities. Given these differences, pooled estimation was not considered appropriate. Smith et al [26] reported higher LARC use in the intervention group at 6 months compared with in-person care (RR 2.08; 95% CI 1.27-3.42), whereas Stephenson et al [29] reported no difference at 12 months (RR 0.98; 95% CI 0.79-1.22). The certainty of evidence was very

XSL•FO
RenderX

low, downgraded for RoB, inconsistency, and serious imprecision (Table 3).

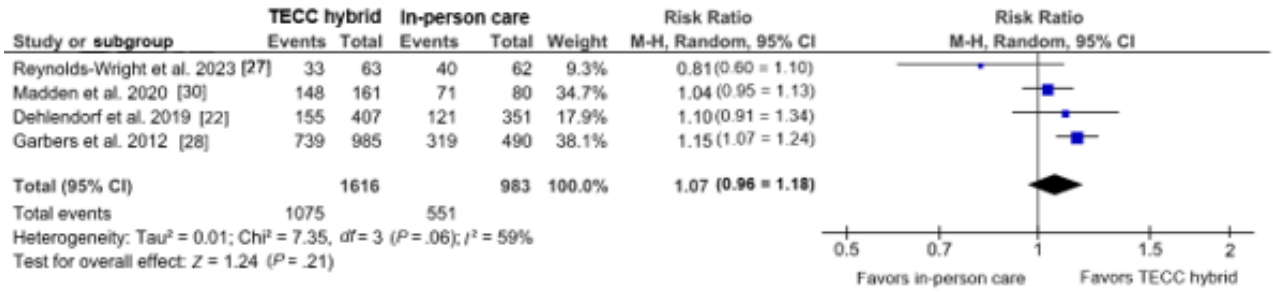
Method Choice

Four studies (n=2599) assessed the choice of ECM immediately after the intervention. The meta-analysis showed no clear evidence (*P*=.21) of an effect of TECC compared with SC (RR 1.07; 95% CI 0.96-1.18; I^2=59%; Figure 6; see also [22,27,28,30]). The certainty of evidence was very low, downgraded for RoB, inconsistency, and imprecision (Table 3).

Two small studies (n=366) assessed the choice of LARC: 1 was conducted in a postabortion setting in the United Kingdom and was terminated early, and the other was conducted in general gynecologic clinics in the United States. Pooled analysis was not considered appropriate because the total sample size did not meet the OIS threshold and clinical heterogeneity was high across study populations, interventions, and contexts. Neither study showed increased choice of LARC at the time of counseling after TECC compared with in-person care: Reynolds-Wright et al [27] (RR 0.77; 95% CI 0.30-1.93) and Madden et al [30] (RR 1.04; 95% CI 0.74-1.47). The certainty of evidence was very low, downgraded for RoB and imprecision (Table 3).

Subgroup and sensitivity analyses examining whether the effect of TECC on ECM choice varied by publication period (*P*=.52) or RoB (*P*=.55) showed no significant effects. Subgroup analyses of LARC choice could not be performed because of the small number of studies.

Figure 6. Forest plot of four studies comparing hybrid telehealth contraceptive counseling (TECC) with in-person care for choice of effective contraceptive method at the time of counseling. M-H: Mantel-Haenszel method.



Satisfaction With Counseling

Two studies (n=749) reported on satisfaction with counseling (Saglam Aksut and Inam [24] and Dehlendorf et al [22]). Madden et al [30] noted no difference in satisfaction but did not report data suitable for extraction. Meta-analysis was not possible because of differences in outcome measurement and reporting. Saglam Aksut and Inam [24] reported significantly higher satisfaction scores for TECC compared with SC (mean 174.86, SD 7.71 vs mean 154.63, SD 26.54; *P*<.001). However, both groups scored well above the predefined threshold for high satisfaction (142 on the 185-point scale used). Dehlendorf et al [22] found no difference in overall satisfaction between groups (odds ratio [OR] 1.11; 95% CI 0.74-1.67), but a higher proportion of TECC users assigned the highest quality ratings (266/404, 65.8% vs 198/345, 57.4%; OR 1.45; 95% CI 1.03-2.05), suggesting that TECC improved perceived counseling quality. The certainty of evidence was very low, downgraded for high RoB, inconsistency, and imprecision (Table 3).

Method Switching

One study reported on method switching following TECC, but did not provide sufficient detail on the methods switched from and to; therefore, the outcome could not be analyzed [29].

Adverse Events and Safety

Only 2 studies explicitly evaluated potential harms as prespecified outcomes. Reiss et al [25] found a statistically significant increase in self-reported physical intimate partner violence (IPV) in the intervention arm compared with the control arm after menstrual regulation (11% vs 7%; adjusted OR 1.97; 95% CI 1.12-3.46). Menstrual regulation is a pragmatic term used for abortion care in Bangladesh, where abortion is illegal. This finding was detectable only when closed questions naming specific acts of violence were used; the open adverse event question (Did anything happen to you as a result of being in this study?) yielded no reports of violence in either arm. Smith et al [26] found no evidence of adverse outcomes, including domestic abuse or road traffic accidents. Other studies did not prespecify or systematically report unintended consequences, representing a gap in the evidence.

Heterogeneity and Publication Bias

We observed 2 forms of heterogeneity. Clinical heterogeneity was considerable, arising from differences in intervention design, including timing, delivery modality, provider involvement, and study setting. Studies were conducted across a range of care settings, including antenatal, postpartum, abortion, family planning, and general gynecologic clinics. Given the small number of studies available for each outcome, stratified analyses by clinical setting were not feasible. Although all interventions met our inclusion criteria for tailoring and interactivity, the extent of user engagement and integration with clinical care varied. Second, statistical heterogeneity was considerable for some outcomes (I^2>85%), as reflected by the wide CIs around pooled effect estimates. The scarcity of positive effect sizes did not suggest publication bias; however, the small number of studies per outcome limited the ability to detect asymmetry, and imputation-based analyses were not conducted.

XSL•FO
RenderX

Discussion

Principal Findings

This review identified 9 RCTs comparing TECC models with in-person care, 8 of which were hybrid models, and the remaining 1 was a standalone model that was terminated early. We found no clear evidence of an effect of TECC on ECM use less than 6 months. We found a small, statistically significant increase in ECM use at 6-12 months (RR 1.07; 95% CI 1.002-1.130). However, this finding is fragile: the lower bound of the CI only narrowly excludes no effect, meaning that a small shift in events could render the result nonsignificant, and the certainty of evidence is low. Given that the pooled estimates for method use were derived from trials with heterogeneous interventions and low-certainty evidence, the small effect observed at 6-12 months should be interpreted with caution. For other outcomes, including choice of ECM, choice of LARC, LARC use at less than 6 months and 6-12 months, and satisfaction with counseling, the evidence was very uncertain compared with SC. Method switching was assessed in 1 trial [29] but was not reported in sufficient detail to permit analysis. Across outcomes, the certainty of evidence ranged from low to very low. Hybrid interventions dominated the included studies; therefore, no conclusions could be drawn regarding the effectiveness of standalone TECC models.

The included trials were designed to evaluate the superiority of TECC over standard in-person care, not its noninferiority. In contexts where geographic, financial, or structural barriers limit access to in-person CC, a telehealth alternative that achieves equivalent outcomes could still represent a clinically meaningful advance by enabling access to care that would otherwise be unavailable. The current evidence base cannot exclude the possibility that TECC is noninferior to standard in-person care; however, formally demonstrating noninferiority would require adequately powered noninferiority trials with prespecified margins.

Comparison With Prior Work

In this review, we included only RCTs of tailored, interactive TECC, a definition that has not been used in any previous systematic review. Three previous systematic reviews, all published before 2020 and before the rapid expansion of telehealth during the COVID-19 pandemic, have assessed the effect of telehealth interventions on contraceptive use or choice [11,12,15].

Our primary outcome, ECM use, has been investigated in 1 previous review [11], which included 7 RCTs conducted in LMICs, 3 of which overlapped with our review [23,25,26]. Their meta-analysis showed no overall benefit of telehealth on ECM use compared with SC, but found a small effect in favor of telehealth in a subgroup analysis of 4 studies that evaluated ECM use as a primary outcome. Unlike our review, they included noninteractive interventions and excluded telephone-based interventions. Broader contraceptive use and continuation outcomes have been investigated in a review of telehealth decision aids [12], which included 11 RCTs, 4 of which overlapped with our review [22, 28-31]. Of the remaining 7 studies, 6 did not use SC comparators, and 1 intervention was not tailored. Their meta-analysis showed a small increase in use of any contraceptive method following telehealth contraceptive services compared with controls. Additionally, a recent cluster RCT among Latinx adolescents in the United States showed increased contraceptive uptake with a tailored TECC intervention, but the study was excluded from our review because not all controls received CC [32].

Contraceptive method choice has been evaluated in 1 review that included 5 RCTs, none of which were included in our review because they evaluated noninteractive interventions, such as video-based interventions, and defined their outcome as choice of any contraceptive method [15]. Consistent with our findings, their narrative synthesis showed that telehealth had no effect on method choice in 4 of the 5 trials.

Evidence on satisfaction with and acceptability of telehealth contraceptive services comes predominantly from observational studies conducted in the United States during the COVID-19 pandemic, which consistently report high acceptability, with users valuing convenience and rating counseling quality as comparable to in-person care [8-10,33]. However, a recent evidence map of telehealth sexual and reproductive health studies in high-income settings found that all observational and quasi-experimental studies demonstrated a very high susceptibility to confounding and selection bias, with many relying on convenience samples likely to overestimate acceptability and effectiveness [34]. These findings may also not reflect the experiences of populations with limited digital access or those who prefer in-person care. In LMICs, promising results regarding acceptability have been reported in pilot studies of telehealth communication for contraception among postabortion women in South Africa and among men and their spouses in Uganda [35,36].

Strengths and Limitations

This review has several strengths, including adherence to a preregistered protocol, rigorous application of RoB 2 and GRADE, and an exclusive focus on RCTs of interactive and tailored interventions to examine the causal effect of the telehealth counseling modality compared with in-person care. One limitation is the restriction to English-language publications, which may have introduced language bias and excluded relevant non-English-language trials. A further limitation is the exclusion of observational evidence. Prespecification of RCT-only inclusion strengthens causal inference regarding the effect of the telehealth counseling modality but constrains external validity and excludes a substantial body of real-world implementation data that have grown since the expansion of telehealth during the COVID-19 pandemic. Observational studies provide important insights into who accesses telehealth contraceptive services and how these services are experienced, but they are less suited to answering the specific effectiveness question addressed in this review because patients who select telehealth differ systematically from those who attend in-person care with respect to age, language, insurance status, income, and digital literacy. In addition, observational designs in this field have been shown to be highly susceptible to confounding and selection bias [34,37].

XSL•FO
RenderX

We used the standard DerSimonian-Laird random-effects approach in our meta-analysis. The Hartung-Knapp-Sidik-Jonkman method should be considered in future systematic reviews, particularly when the risk of spurious positive findings (type I error) is greater than it was in our analysis. The evidence has important limitations. We did not contact the study authors for additional or clarifying information. Interventions differed widely in content, delivery, timing, degree of provider involvement, and clinical context, potentially contributing to substantial heterogeneity in the results. Outcomes were often self-reported, and group allocation was unblinded, increasing RoB for subjective outcomes such as satisfaction. Some studies were excluded because outcome data did not distinguish between any contraceptive method and ECM as defined in our review. The small number of trials limited subgroup analyses and reduced power to explore sources of heterogeneity through sensitivity analyses. In line with Cochrane guidance, we did not present funnel plots to assess publication bias because fewer than 10 studies were included. We do not suspect publication bias, as most trials showed no effect in favor of the intervention.

Despite our focus on comprehensive TECC, tailoring and interactivity were only partially implemented in most included trials, highlighting a gap in TECC intervention design that may have limited effectiveness. Our review did not assess which components of TECC were most influential in determining outcomes. Previous reviews have attempted to identify the most effective features and have suggested that the incorporation of behavior change techniques and motivational interviewing may increase effectiveness [11,13,14]. However, none were able to determine which specific components were most effective.

The included trials were conducted across antenatal, postabortion, and routine contraceptive care settings, and in both high- and middle-income countries, supporting the generalizability of our findings to some extent. However, important gaps remain: no trial was conducted in a low-income country, and adolescents were represented in only 2 trials [22,28]. These gaps limit the generalizability of the findings to populations with a high unmet need for contraception. Furthermore, only 1 trial [27], which was terminated early, evaluated standalone TECC, highlighting the evidence gap regarding standalone TECC.

Safety outcomes were seldom assessed. One trial reported increased IPV following a mobile-based intervention in Bangladesh [25]. The intervention explicitly encouraged women to discuss contraception with their partners and included content aimed at increasing partner engagement in contraceptive decision-making. Mobile-based interventions that generate records or notifications may inadvertently increase the risk of IPV. This finding underscores the critical importance of prospective safety monitoring, harm-informed design, and safeguard protocols in all TECC evaluations, particularly in settings with a high prevalence of IPV. Although this finding may reflect context-specific factors related to postmenstrual regulation services in a setting with a high prevalence of IPV, it should not be dismissed as an isolated observation. Future TECC trials should consider incorporating routine baseline IPV risk screening, prespecified adverse event monitoring protocols, and context-sensitive content design that avoids prompting partner communication when doing so may increase risk.

Observational studies conducted in the United States have shown that food and housing insecurity, language barriers, lack of insurance, and race are associated with reduced access to telehealth contraceptive care [37,38]. Equity dimensions were not systematically addressed in the included trials, underscoring the need for equity-focused implementation and evaluation in diverse and underserved populations, including individuals with low literacy and limited digital access.

Implications

Although the evidence on adjunct TECC is inconclusive, it does not appear to substantially improve use, method choice, or satisfaction in settings where in-person counseling is already available. Our review assessed hybrid TECC in comparison with in-person counseling, which is not available in all contexts. If TECC were noninferior to SC in whatever form that takes, it may offer benefits in settings with limited trained providers, geographical or social barriers to care, or in overburdened health systems. In such settings, even modest effects might translate into population-level impact. The effectiveness of standalone TECC remains unknown, and its clinical relevance would depend on demonstrating that TECC is no less effective, safe, or acceptable than SC.

Future Research

Future research should prioritize adequately powered evaluations of standalone TECC, including low-income settings and adolescents, and assess long-term outcomes such as continuation, switching, and unintended pregnancy. Noninferiority designs could establish clinical relevance in settings with limited access to in-person care. Mixed-methods approaches are needed to capture user experiences, equity implications, and unintended harms. Standardized and validated measures of contraceptive use, satisfaction, and decision quality would improve comparability. Complementary real-world effectiveness evidence could be generated through rigorous quasi-experimental designs, such as propensity score–matched cohort studies in routine care populations and interrupted time-series analyses that treat system-level changes, such as the COVID-19 telehealth expansion, as natural experiments. The rapid growth of online contraceptive platforms and AI-driven communication requires rigorous evaluation of their effectiveness, safety, and implementation. Evidence is needed to guide decision makers in integrating TECC into routine contraceptive services amid ongoing digital health system transformations.

Conclusions

Current evidence suggests, with low certainty, that adjunct TECC delivered alongside in-person care shows little to no effect on contraceptive use compared with in-person care. For method choice, LARC uptake, and satisfaction compared with in-person care, the evidence was very uncertain. The effectiveness of standalone TECC remains unknown. High acceptability and scalability warrant further rigorous and equity-focused evaluations to inform integration into routine contraceptive services.

XSL•FO
RenderX

Data Sharing

All data used in this article are from publicly available published sources. Extracted data are provided in the article and its multimedia appendices. Additional outcome data, risk-of-bias assessments, and statistical calculations for meta-analyses will be made available on reasonable request to the corresponding author (ME). The PROSPERO protocol (CRD42023404402) is publicly available.

Acknowledgments

We thank Lovisa Liljegren and Emma-Lotta Säätelä, librarians at Karolinska Institutet University Library, for assistance with developing and documenting the search strategies. During the preparation of this work, the authors used GPT-4o (OpenAI, May 2024 release) to perform preliminary data extraction from full-text articles, generate a conceptual figure, and support language editing. After using this tool, 2 authors (MW and ME) reviewed and edited the content as needed and take full responsibility for the content of the published article.

Authors' Contributions

MW and ME conceived the study and designed the protocol. MW, AK, and ME screened titles and abstracts. Initial outcome data were extracted using GPT-4o (OpenAI, May 2024 release) by MW, and the data were verified and corrected if needed by MW and ME. MW and ME independently assessed the risk of bias. MW conducted the data analyses with input from WJAG and ME. WJAG provided statistical guidance and interpretation of the meta-analyses. MW drafted the initial manuscript. ME supervised the project. ME and WJAG contributed to the interpretation of findings and provided critical revisions. MW, ME, and WJAG had full access to all study data. All authors had final responsibility for the decision to submit for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 113 KB - mhealth_v14i1e88887_app1.pdf]

Multimedia Appendix 2
Search strategies.
[DOCX File , 124 KB - mhealth_v14i1e88887_app2.docx]

Multimedia Appendix 3
Artificial intelligence–generated conceptual framework of telehealth contraceptive counseling (TECC). The framework includes 3 core domains: (1) telehealth delivery via a user interface on a phone, laptop, or tablet; (2) interactive, bidirectional communication; and (3) personalization based on users' preferences, reproductive goals, or medical eligibility. Additional characteristics may vary by degree of automation and by whether communication is delivered and responded to in real time or asynchronously. TECC may be integrated before, during, or after clinical visits and may serve either as a standalone substitute for, or a hybrid adjunct to, in-person counseling.
[DOCX File , 1565 KB - mhealth_v14i1e88887_app3.docx]

Multimedia Appendix 4
Records excluded at full-text review with reasons.
[DOCX File , 40 KB - mhealth_v14i1e88887_app4.docx]

Multimedia Appendix 5
Outcomes and findings.
[DOCX File , 22 KB - mhealth_v14i1e88887_app5.docx]

Multimedia Appendix 6
Characteristics of study interventions categorized according to the 3 domains of our telehealth contraceptive counseling definition: modality, tailoring, and interactivity.
[DOCX File , 4450 KB - mhealth_v14i1e88887_app6.docx]

Multimedia Appendix 7

XSL•FO
RenderX

Detailed breakdown of GRADE (Grading of Recommendations Assessment, Development, and Evaluation) assessment. [DOCX File , 22 KB - mhealth_v14i1e88887_app7.docx]

References

1. Cleland J, Conde-Agudelo A, Peterson H, Ross J, Tsui A. Contraception and health. The Lancet 2012 Jul;380(9837):149-156. [doi: 10.1016/s0140-6736(12)60609-6]
2. Ahmed S, Li Q, Liu L, Tsui AO. Maternal deaths averted by contraceptive use: an analysis of 172 countries. The Lancet 2012 Jul;380(9837):111-125. [doi: 10.1016/s0140-6736(12)60478-4]
3. Sully E, Biddlecom A, Darroch J, Riley T, Ashford L, Lince-Deroche N. Adding it up: investing in sexual and reproductive health 2019. Guttmacher Institute. 2020. URL: https://www.guttmacher.org/report/adding-it-up-investing-in-sexual-reproductive-health-2019 [accessed 2025-11-15]
4. Sedgh G, Hussain R. Reasons for contraceptive nonuse among women having unmet need for contraception in developing countries. Stud Fam Plann 2014 Jun 14;45(2):151-169. [doi: 10.1111/j.1728-4465.2014.00382.x] [Medline: 24931073]
5. Emtell Iwarsson K, Envall N, Bizjak I, Bring J, Kopp Kallner H, Gemzell-Danielsson K. Increasing uptake of long-acting reversible contraception with structured contraceptive counselling: cluster randomised controlled trial (the LOWE trial). BJOG 2021 Aug 07;128(9):1546-1554. [doi: 10.1111/1471-0528.16754] [Medline: 33988917]
6. World Health Organization (WHO). WHO-ITU global standard for accessibility of telehealth services. WHO. Geneva, Switzerland: World Health Organization and International Telecommunication Union; 2022. URL: https://www.who.int/publications/i/item/9789240050464 [accessed 2025-11-18]
7. World Health Organization (WHO). WHO guideline: recommendations on digital interventions for health system strengthening. WHO. Geneva, Switzerland: WHO; 2019. URL: https://www.who.int/publications/i/item/9789241550505/ [accessed 2025-11-15]
8. Stifani BM, Smith A, Avila K, Boos EW, Ng J, Levi EE, et al. Telemedicine for contraceptive counseling: patient experiences during the early phase of the COVID-19 pandemic in New York City. Contraception 2021 Sep;104(3):254-261 [FREE Full text] [doi: 10.1016/j.contraception.2021.04.006] [Medline: 33861981]
9. Lindberg LD, Mueller J, Haas M, Jones RK. Telehealth for contraceptive care during the COVID-19 pandemic: results of a 2021 national survey. Am J Public Health 2022 Jun;112(S5):S545-S554. [doi: 10.2105/ajph.2022.306886]
10. Merz-Herrala AA, Kerns JL, Logan R, Gutierrez S, Marshall C, Diamond-Smith N. Contraceptive care in the United States during the COVID-19 pandemic: a social media survey of contraceptive access, telehealth use and telehealth quality. Contraception 2023 Jul;123:110000 [FREE Full text] [doi: 10.1016/j.contraception.2023.110000] [Medline: 36871620]
11. Zulu EM, Sukwa T. Impact of mHealth on contraceptive use among women and men of reproductive age in low- and middle-income countries: a systematic review and meta-analysis. Trop Med Int Health 2020 Oct 11;25(10):1182-1197 [FREE Full text] [doi: 10.1111/tmi.13464] [Medline: 33463875]
12. Goueth RC, Maki KG, Babatunde A, Eden KB, Darney BG. Effects of technology-based contraceptive decision aids: a systematic review and meta-analysis. Am J Obstet Gynecol 2022 Nov;227(5):705-713.e9 [FREE Full text] [doi: 10.1016/j.ajog.2022.06.050] [Medline: 35779590]
13. Smith C, Gold J, Ngo TD, Sumpter C, Free C. Mobile phone-based interventions for improving contraception use. Cochrane Database Syst Rev 2015 Jun 26;2015(6):CD011159 [FREE Full text] [doi: 10.1002/14651858.CD011159.pub2] [Medline: 26115146]
14. Aung B, Mitchell JW, Braun KL. Effectiveness of mHealth interventions for improving contraceptive use in low- and middle-income countries: a systematic review. Glob Health Sci Pract 2020 Nov 12;8(4):813-826. [doi: 10.9745/ghsp-d-20-00069]
15. Dewart CM, Serpico J, Steiner MJ, Gallo MF. Electronic interventions for changing knowledge, attitudes or practices regarding contraception: a systematic review. Contraception 2019 Jul;100(1):10-25. [doi: 10.1016/j.contraception.2019.04.004] [Medline: 30998928]
16. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021 Mar 29;372:n160 [FREE Full text] [doi: 10.1136/bmj.n160] [Medline: 33781993]
17. Weinryb M, Endler M, Grooten W. How effective are telehealth interventions for contraceptive counseling compared to standard care? International Prospective Register of Systematic Reviews (PROSPERO). 2023 Jun 01. URL: https://www.crd.york.ac.uk/PROSPERO/view/CRD42023404402 [accessed 2026-05-18]
18. Trussell J. Contraceptive failure in the United States. Contraception 2004 Aug;70(2):89-96. [doi: 10.1016/j.contraception.2004.03.009] [Medline: 15288211]
19. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
20. Schünemann H, Brożek J, Guyatt G, Oxman A. The GRADE handbook. World Health Organization (WHO). Geneva, Switzerland: WHO; 2013. URL: https://evidence-impact.org/storage/85/GRADE-Handbook.pdf [accessed 2025-10-22]

XSL•FO
RenderX

21. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence--imprecision. J Clin Epidemiol 2011 Dec;64(12):1283-1293. [doi: 10.1016/j.jclinepi.2011.01.012] [Medline: 21839614]
22. Dehlendorf C, Fitzpatrick J, Fox E, Holt K, Vittinghoff E, Reed R, et al. Cluster randomized trial of a patient-centered contraceptive decision support tool, My Birth Control. American Journal of Obstetrics and Gynecology 2019 Jun;220(6):565.e1-565.e12. [doi: 10.1016/j.ajog.2019.02.015] [Medline: 30763545]
23. Harrington EK, Drake AL, Matemo D, Ronen K, Osoti AO, John-Stewart G, et al. An mHealth SMS intervention on postpartum contraceptive use among women and couples in Kenya: a randomized controlled trial. Am J Public Health 2019 Jun;109(6):934-941. [doi: 10.2105/ajph.2019.305051]
24. Saglam Aksut R, Inam O. Evaluation of telemedicine for contraceptive counseling given to pregnant women during the COVID-19 pandemic: A randomized controlled trial. Health Care Women Int 2025 Jun 28;46(2):115-133. [doi: 10.1080/07399332.2024.2371386] [Medline: 38940954]
25. Reiss K, Andersen K, Pearson E, Biswas K, Taleb F, Ngo TD, et al. Unintended consequences of mHealth interactive voice messages promoting contraceptive use after menstrual regulation in Bangladesh: intimate partner violence results from a randomized controlled trial. Glob Health Sci Pract 2019 Sep 26;7(3):386-403. [doi: 10.9745/ghsp-d-19-00015]
26. Smith C, Ngo TD, Gold J, Edwards P, Vannak U, Sokhey L, et al. Effect of a mobile phone-based intervention on post-abortion contraception: a randomized controlled trial in Cambodia. Bull World Health Organ 2015 Dec 01;93(12):842-50A [FREE Full text] [doi: 10.2471/BLT.15.160267] [Medline: 26668436]
27. Reynolds-Wright JJ, Norrie J, Cameron ST. Using telemedicine to improve early medical abortion at home (UTAH): a randomised controlled trial to compare telemedicine with in-person consultation for early medical abortion. BMJ Open 2023 Sep 13;13(9):e073630 [FREE Full text] [doi: 10.1136/bmjopen-2023-073630] [Medline: 37709327]
28. Garbers S, Meserve A, Kottke M, Hatcher R, Ventura A, Chiasson MA. Randomized controlled trial of a computer-based module to improve contraceptive method choice. Contraception 2012 Oct;86(4):383-390. [doi: 10.1016/j.contraception.2012.01.013] [Medline: 22402258]
29. Stephenson J, Bailey JV, Gubijev A, D'Souza P, Oliver S, Blandford A, et al. An interactive website for informed contraception choice: randomised evaluation of contraception choices. Digit Health 2020;6:2055207620936435 [FREE Full text] [doi: 10.1177/2055207620936435] [Medline: 32704380]
30. Madden T, Holttum J, Maddipati R, Secura GM, Nease RF, Peipert JF, et al. Evaluation of a computerized contraceptive decision aid: a randomized controlled trial. Contraception 2020 Nov;102(5):339-345 [FREE Full text] [doi: 10.1016/j.contraception.2020.08.002] [Medline: 32771369]
31. Dehlendorf C, Fox E, Sharma AE, Zhang J, Yang S, Centola D. Birth Control Connect: a randomized trial of an online group to disseminate contraceptive information. Contraception 2020 Jun;101(6):376-383 [FREE Full text] [doi: 10.1016/j.contraception.2020.01.014] [Medline: 32032641]
32. Tebb KP, Rodriguez F, Pollack LM, Adams S, Rico R, Renteria R, et al. Improving contraceptive use among Latina adolescents: a cluster-randomized controlled trial evaluating an mHealth application, Health-E You/Salud iTu. Contraception 2021 Sep;104(3):246-253 [FREE Full text] [doi: 10.1016/j.contraception.2021.03.004] [Medline: 33744300]
33. Shin RJ, Yao M, Akesson C, Blazel M, Mei L, Brant AR. An exploratory study comparing the quality of contraceptive counseling provided via telemedicine versus in-person visits. Contraception 2022 Aug;112:86-92. [doi: 10.1016/j.contraception.2022.02.004] [Medline: 35247368]
34. Parikh R, Shetty N, Singhal C, Patel P, Manghani P, Pillai A, et al. Telehealth for sexual and reproductive healthcare: evidence map of effectiveness, patient and provider experiences and preferences, and patient engagement strategies. Clin Pract 2026 Jan 09;16(1):14 [FREE Full text] [doi: 10.3390/clinpract16010014] [Medline: 41590208]
35. Weinryb M, Knight L, Larsson EC, Endler M. Telemedicine contraceptive counselling through WhatsApp in South African abortion care: a mixed-methods acceptability study. Eur J Contracept Reprod Health Care 2025 Aug 11;30(4):190-199 [FREE Full text] [doi: 10.1080/13625187.2025.2500637] [Medline: 40497321]
36. Kamulegeya LH, Bwanika J, Banonya J, Atuhaire J, Musinguzi D, Nakate V, et al. Feasibility and acceptability of a Ugandan telehealth engagement platform for informational messaging on modern contraception: pilot cross-sectional study. JMIR Form Res 2022 Jun 28;6(6):e34424 [FREE Full text] [doi: 10.2196/34424] [Medline: 35763336]
37. Meurice ME, Mody SK, Nodora J, Marengo A, Averbach S. Social determinants of choosing telemedicine for contraceptive care: a retrospective cohort study. Contraception 2024 Jun;134:110414 [FREE Full text] [doi: 10.1016/j.contraception.2024.110414] [Medline: 38431258]
38. Yarger J, Hopkins K, Elmes S, Rossetto I, Van Liefde D, De La Melena S, et al. Use of telemedicine to obtain contraception among young adults: inequities by health insurance. Contraception 2024 Jun;134:110419 [FREE Full text] [doi: 10.1016/j.contraception.2024.110419] [Medline: 38467325]

Abbreviations

CC: contraceptive counseling
ECM: effective contraceptive method

XSL•FO
RenderX

GRADE: Grading of Recommendations Assessment, Development, and Evaluation
HIC: high-income country
IPV: intimate partner violence
LARC: long-acting reversible contraception
LMIC: low- and middle-income country
MeSH: Medical Subject Headings
OIS: optimal information size
OR: odds ratio
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
RoB: risk of bias
RR: relative risk
SC: standard care
TECC: telehealth contraceptive counseling

Edited by L Buis; submitted 08.Dec.2025; peer-reviewed by PA Okooboh, R Parikh, S Berendes; comments to author 23.Mar.2026; revised version received 03.Jun.2026; accepted 04.Jun.2026; published 16.Jul.2026.

Please cite as:
Weinryb M, Grooten WJA, Koris A, Endler M
Tailored and Interactive Mobile Telehealth Contraceptive Counseling Compared With In-Person Care: Systematic Review and Meta-Analysis of Randomized Controlled Trials
JMIR Mhealth Uhealth 2026;14:e88887
URL: https://mhealth.jmir.org/2026/1/e88887
doi:10.2196/88887
PMID:

©Maja Weinryb, Wilhelmus JA Grooten, Andrea Koris, Margit Endler. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 16.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Home-Based, Digitally Assisted Dyadic Tai Chi Intervention for Patients With Chronic Heart Failure and Their Caregivers: Development of an Intervention and Pilot Randomized Controlled Trial

Yan Zhang[1,2], PhD; Fei Wan Ngai[2], PhD; Lihua Cao[1], BSN; Xiaohong Lu[1], MSN; Wenzhong Zhang[1], PhD; Lili Wei[3], PhD; Yao Jie Xie[2,4,5], PhD

[1]Department of Cardiology, Affiliated Hospital of Qingdao University, Qingdao, Shandong, China

[2]School of Nursing, Hong Kong Polytechnic University, Hong Kong, China

[3]Office of the Dean, Affiliated Hospital of Qingdao University, Qingdao, Shandong, China

[4]Research Centre for Chinese Medicine Innovation, Hong Kong Polytechnic University, Hong Kong, China

[5]Joint Research Centre for Primary Health Care, Hong Kong Polytechnic University, Hong Kong, China

Corresponding Author:
Yao Jie Xie, PhD
School of Nursing
Hong Kong Polytechnic University
No. 11 Yuk Choi Road, Hung Hom, Kowloon
Hong Kong
China
Phone: 852 34003798
Email: grace.yj.xie@polyu.edu.hk

Abstract

Background: Chronic heart failure (CHF) imposes a dual and dynamically interrelated burden on patients and family caregivers, whose health and illness management behaviors are interdependent. Exercise interventions that actively engage the patient-caregiver dyad as the unit of rehabilitation in the home setting remain underdeveloped.

Objective: This study aimed to develop a home-based, digitally assisted, dyadic tai chi (HDTC) intervention and evaluate its feasibility, acceptability, safety, and preliminary effects among dyads of patients with CHF and their caregivers.

Methods: This sequential, multiphase study comprised intervention development followed by a single-center, assessor-blinded, parallel-group pilot randomized controlled trial. Phase 1 integrated literature review, theoretical frameworks, and stakeholder focus group interviews to develop and refine the HDTC intervention. In phase 2, eligible adults with CHF (aged 18-75 years; New York Heart Association class I-III) and their primary caregivers were recruited from a tertiary medical center in Qingdao, China, from October 14 to November 12, 2024, and followed for 12 weeks. Forty dyads were allocated 1:1 to the 12-week HDTC intervention plus usual care or usual care alone (n=20 per group). Outcome assessors and the data analyst were blinded to group allocation. Feasibility, acceptability, and safety were the primary focus of the pilot trial. Patient, caregiver, and dyadic health outcomes were preliminarily explored at baseline and 4 and 12 weeks. All 40 randomized dyads were included in the intention-to-treat analyses.

Results: The HDTC intervention comprised an online educational seminar, nurse-led dyadic coaching, preliminary face-to-face tai chi training, and supervised HDTC practice. The recruitment rate was 83.3% (40/48), mean session attendance was 73.7% (SD 8.6%), and 70% (14/20) of participants achieved high adherence rate (≥75% attendance). All intervention dyads rated the intervention as acceptable. No intervention-related adverse events were identified. Exploratory generalized estimating equation analyses suggested greater improvements from baseline to 12 weeks in the intervention group than in the control group in functional capacity (β=38.858; *P*=.001), together with greater reductions in Minnesota Living With Heart Failure Questionnaire scores (β=−1.850; *P*=.002), sedentary time (β=−1.025; *P*=.01), and fear of activity (β=−6.550; *P*=.01). Caregiver burden also decreased (β=−4.350; *P*<.001), while mutuality improved more for both patients (β=10.508; *P*<.001) and caregivers (β=6.450; *P*<.001).

Conclusions: The HDTC intervention was feasible and acceptable, with preliminary improvements across patient, caregiver, and dyadic outcomes. By integrating shared patient-caregiver participation with digitally supported home-based tai chi, the HDTC

XSL•FO
RenderX

intervention extends patient-centered rehabilitation toward a dyad-oriented model. If confirmed in a fully powered trial, this approach may offer a practical strategy for expanding home-based cardiac rehabilitation.

Trial Registration: ClinicalTrials.gov NCT06876441; https://clinicaltrials.gov/study/NCT06876441

(JMIR Mhealth Uhealth 2026;14:e95689) doi:10.2196/95689

KEYWORDS

heart failure; caregivers; home-based rehabilitation; tai chi; dyadic intervention

Introduction

Heart failure (HF) remains a prevalent and burdensome chronic condition worldwide [1,2]. Patients with HF commonly exhibit high levels of sedentary behavior, which are associated with increased risks of hospitalization and mortality [3]. Although exercise-based cardiac rehabilitation is recommended to improve functional outcomes and prognosis [4], participation in center-based programs remains low because of travel, scheduling, and service capacity barriers [5]. Digitally supported home-based exercise rehabilitation offers a potential alternative, although evidence regarding its safety, feasibility, and effects in HF remains limited [6].

Within the home setting, exercise is embedded in the family context, with caregivers often providing encouragement, safety monitoring, scheduling support, and reinforcement of continued participation [7]. Meanwhile, caregivers experience substantial caregiving burden that can adversely affect their own physical and psychological health [8,9], which may in turn influence patient outcomes [10]. This reciprocal influence supports conceptualizing the patient-caregiver dyad, rather than the patient alone, as a target of intervention [11,12]. However, dyadic intervention research in HF remains limited [13], and existing interventions have focused primarily on self-care and psychosocial support, with relatively little attention to structured exercise jointly undertaken by patients and caregivers [14,15].

Dyadic exercise extends conventional patient-centered rehabilitation by engaging both members of the dyad in a shared health behavior [16]. Consistent with the theory of dyadic illness management, which emphasizes reciprocal influences within the dyad [17], this approach moves beyond targeting patient behavior alone. It also differs from caregiver-supported self-care, in which caregivers primarily facilitate patient disease management; shared exercise actively involves both members and may benefit both while strengthening mutual support and dyadic relationship quality [11].

Tai chi, a moderate-intensity aerobic exercise, is recommended as an adjunctive therapy in international HF management guidelines [18,19]. Prior studies suggest that it may enhance exercise capacity, psychological well-being, and quality of life (QoL) in patients with chronic HF (CHF) [20,21], but practical barriers, including travel demands, time constraints, self-discipline challenges, and limited access to qualified masters, hinder patients' access to center-based practice [22], underscoring the need for home-based, digitally assisted tai chi. Previous evidence has revealed that digitally assisted tai chi, in which all components are delivered through an online platform without face-to-face instruction, is feasible and effective among other frail populations [23]. To date, however, the integration of tai chi into dyadic management models remains largely unexplored.

To address these gaps, this study aimed to systematically develop a home-based, digitally assisted, dyadic tai chi (HDTC) intervention for patients with CHF and their family caregivers. Then, a pilot randomized controlled trial (RCT) was conducted to evaluate its feasibility, acceptability, and safety and explore preliminary effects on patient, caregiver, and dyadic outcomes.

Methods

Study Design and Setting

This sequential multimethod study guided by the Medical Research Council framework comprised 2 stepwise phases [24]. Phase 1 developed the HDTC intervention for dyads through (1) iterative evidence synthesis (literature review, systematic review of digitally assisted interventions, and scoping review of caregiver roles), (2) application of theoretical frameworks to inform a stage transition model for sustained dyadic exercise, and (3) focus group interviews with key stakeholders (physicians, nurses, patients with CHF, caregivers, and tai chi master) to refine the digital implementation of the intervention. Phase 2 was a prospective, 2-arm pilot RCT conducted at a tertiary medical center in Qingdao, China, evaluating the feasibility, acceptability, and preliminary effects of the HDTC intervention. This pilot RCT was reported in accordance with the CONSORT (Consolidated Standards of Reporting Trials) 2010 extension for randomized pilot and feasibility trials [25] (Multimedia Appendix 1).

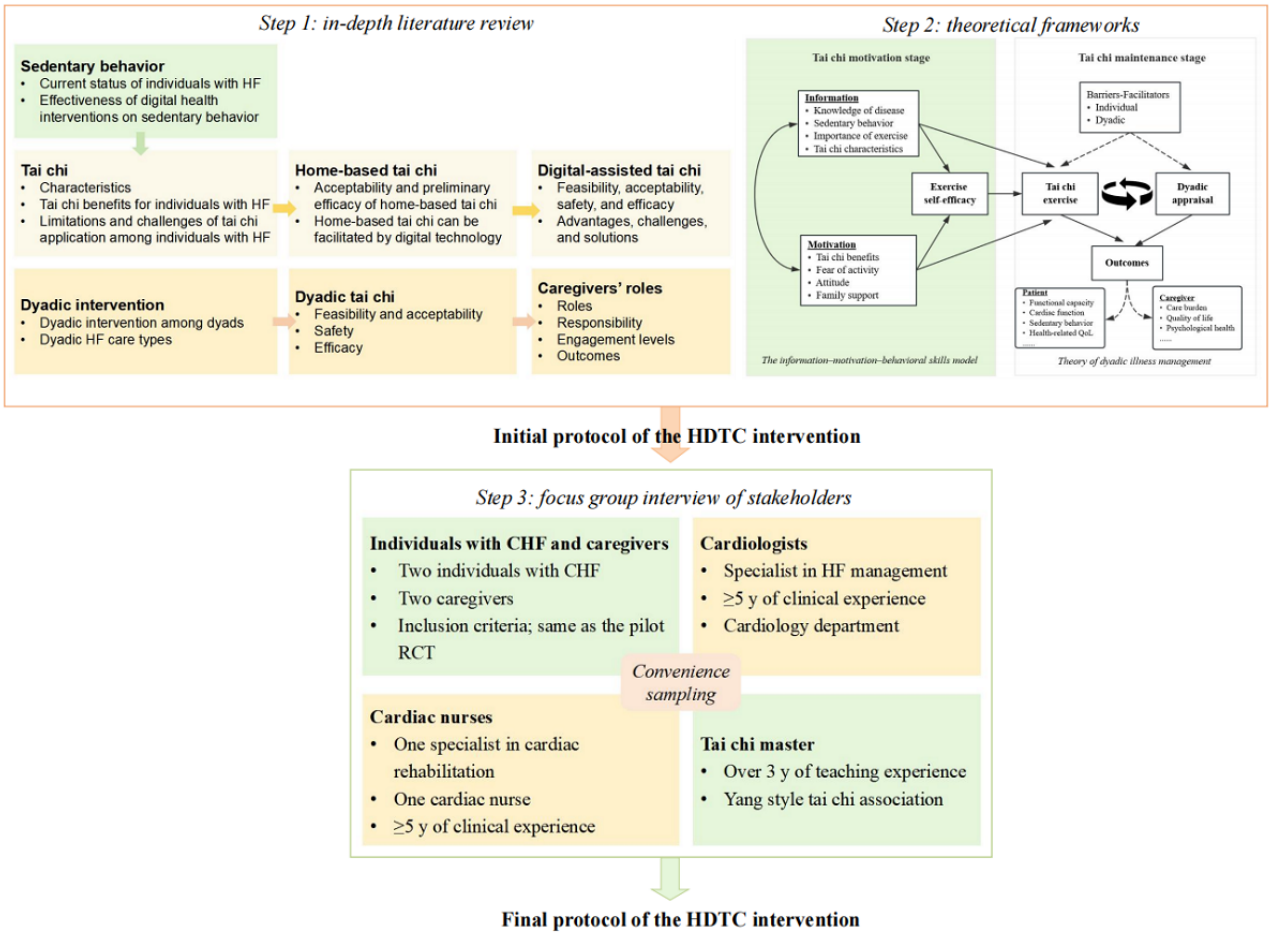
Phase 1: Development of the HDTC Intervention

The intervention was developed through 3 sequential steps. First, an iterative evidence synthesis via literature review identified key components informing the HDTC intervention protocol: sedentary behavior, digital health interventions, home-based tai chi, dyadic tai chi, and caregiver roles. Second, 2 theoretical frameworks guided the transition from sedentary behavior to sustained tai chi exercise: the information-motivation–behavioral skills model and theory of dyadic illness management. The information-motivation–behavioral skills model was used primarily to structure the initial transition from sedentary behavior to exercise participation (*tai chi motivation stage*) [26,27]. The educational seminar addressed exercise-related information, the dyadic coaching session explored motivation and shared goals, and introductory tai chi training was used to develop the behavioral skills required for safe home practice. The theory of dyadic illness management informed the

XSL•FO
RenderX

subsequent maintenance phase (*tai chi maintenance stage*) [17]. In this phase, patients and caregivers were encouraged to view exercise as a shared rehabilitation task, coordinate participation and support, monitor progress, and address exercise-related challenges collaboratively. Digital communication, shared reminders, and professional feedback were incorporated to support continued dyadic engagement. Third, focus group interviews were conducted with key stakeholders, comprising 2 dyads (2 patients and their respective caregivers), 2 cardiologists and 2 nurses with more than 5 years of experience in HF management, and 1 tai chi master with 11 years of teaching experience from the Yang Style Tai Chi Association. The interview guide was designed based on the nonadoption, abandonment, scale-up, spread, and sustainability framework (Multimedia Appendix 2) to enhance the practicality and adaptability of the digital intervention [28]. The intervention was subsequently refined based on stakeholder feedback. The flow diagram of the intervention development is shown in Figure 1.

Figure 1. The home-based, digital-assisted, dyadic tai chi (HDTC) was systematically developed through 3 steps: literature review, theory-guided intervention design, and stakeholder consultation. CHF: chronic heart failure; HF: heart failure; RCT: randomized controlled trial.



Phase 2: Pilot-Testing of the HDTC Intervention

Participants and Sample Size

A convenience sample of 40 dyads of patients with CHF and their caregivers was recruited from the Affiliated Hospital of Qingdao University. The inclusion criteria for patients with CHF were (1) age of 18 to 75 years with a diagnosis of CHF [29]; (2) New York Heart Association class I, II, or III and 6-minute walk test distance of more than 150 m [29]; (3) ability to perform the prescribed tai chi movements without physical assistance, as assessed during pre-enrollment screening; (4) smartphone ownership; and (5) voluntary participation and informed consent. Exclusion criteria were as follows: planned surgery within 6 months; severe comorbidities such as cancer, liver or kidney failure, and diagnosed psychiatric disorders; planned relocation to care facilities; regular exercise in the prior 3 months; or concurrent trial participation. Caregiver inclusion criteria were age of 18 years or older, being primary caregivers for patients with CHF [30], ability to practice tai chi without help, and voluntary participation and informed consent.

On the basis of pilot trial recommendations, at least 15 participants per group were required to detect a medium effect size assuming a 90% power with a 2-sided 5% significance level [31]. Accounting for 20% attrition, 18 dyads per group were targeted.

Recruitment of Dyads of Patients With CHF and Their Caregivers

Potential patients with CHF were recruited through electronic advertisements with QR codes displayed in the HF clinic and cardiology department and via electronic questionnaires disseminated to 30 WeChat groups of discharged patients. After

XSL•FO
RenderX

initial prescreening, eligible patients underwent a comprehensive assessment followed by screening of their caregivers. Informed consent was obtained from all dyads prior to enrollment.

Randomization, Allocation Concealment, and Blinding

Randomization was performed by an independent individual using the Sealed Envelope online tool [32]. Allocation used a 1:1 ratio with stratification by New York Heart Association class (I, II, and III) and block sizes of 4 and 6. Allocation concealment was ensured through sequentially numbered, opaque, sealed envelopes opened only after baseline assessments. Owing to the nature of the behavioral intervention, blinding of intervention providers and participants was not feasible [33]; only outcome assessors and the data analyst were blinded to group allocation.

Interventions

HDTC Intervention

Dyads in the intervention group participated in the HDTC intervention plus usual care. The intervention was delivered over 12 weeks, consistent with the duration commonly used in previous tai chi studies among patients with CHF [16]. The contents of the intervention are detailed in the Results section.

Control Group

Dyads in the control group received usual care, including online comprehensive education and nurse-led telephone consultations. Consultations addressed symptom management, medication adherence, dietary and exercise suggestions, psychosocial support, and regular follow-ups. Caregivers served as both partners and facilitators, assisting patients with CHF in implementing these recommendations.

Intervention Fidelity

The Behavior Change Consortium framework was adopted to guide fidelity assurance [34]. First, the initial HDTC protocol was developed through a rigorous 3-stage process to ensure design integrity. Second, a tai chi master received standardized training from a cardiac rehabilitation nurse on intervention delivery, safety protocols, and participant characteristics, with ongoing researcher supervision throughout all sessions to maintain protocol adherence. Third, all tai chi sessions were delivered by a single master from the Qingdao Yang Style Tai Chi Association with 11 years of teaching experience. The master was scheduled to deliver all sessions throughout the 12-week intervention to ensure consistency in intervention delivery. If the master was unavailable, the affected session was rescheduled to maintain continuity of the intervention. Upon completion of the tai chi form learning, the master assessed participants' skill mastery. Additionally, all online classes were recorded, with an independent expert conducting a random audit of 20% of recorded classes to evaluate intervention fidelity.

Outcomes

Assessments were conducted at baseline (T0) and 4 weeks (T1), and 12 weeks (T2) of the intervention to evaluate the feasibility, acceptability, and preliminary effects of the HDTC intervention.

Feasibility

Feasibility was assessed through the recruitment period, recruitment rate, attendance rate, adherence rate, and attrition rate. Adherence was calculated as the proportion of completed tai chi classes to total prescribed classes, with high adherence defined as attendance to at least 75% of classes, consistent with previous studies [35].

Acceptability

Acceptability was assessed after the intervention using separate patient and caregiver questionnaires based on the theoretical framework of acceptability, covering affective attitude, burden, ethicality, perceived effectiveness, intervention coherence, self-efficacy, opportunity costs, and overall acceptability [36]. Each domain was rated using domain-specific ordered response options. Responses were dichotomized as favorable or unfavorable according to the positive response categories for each domain; for overall acceptability, "acceptable" and "completely acceptable" were classified as favorable. The intervention was considered acceptable if 80% or more of the participants provided a favorable overall rating [37].

Adverse Events

Consistent with the CONSORT Harms statement [38], adverse events were defined as any possible adverse consequences attributable to the intervention. Potential adverse events, including dyspnea, palpitations, chest pain, arrhythmia, myocardial infarction, falls, and other unexpected symptoms, were recorded by a nurse and subsequently assessed by a physician to determine their relationship to the intervention.

Preliminary Effects

Outcomes of Patients With CHF

The functional capacity of patients with CHF was assessed using the 6-minute walk test through the Walker Doctor Intelligent 6-Minute Walk Test Monitoring and Analysis System [39,40]. Sedentary behavior and fear of activity were evaluated using the International Physical Activity Questionnaire–Short Form [41] and Fear of Activity in Patients With Chronic Heart Failure [42], respectively. Confidence in exercising was assessed using the Self-Efficacy for Exercise Scale [43]. The Minnesota Living With Heart Failure Questionnaire was used to measure patients' health-related QoL [44].

Outcomes of Caregivers

The Chinese version of the Zarit Caregiver Burden Interview was used to measure caregivers' perceived caring burden [45], and their QoL was assessed using the Chinese Family Caregiver Quality of Life Scale [46].

Outcomes of Dyads

Mutuality, the relationship between patients with CHF and their caregivers, was measured using the Chinese Mutuality Scale [47], with higher scores denoting better mutuality.

Data Analysis

Statistical analyses were performed using SPSS Statistics (version 26; IBM Corp). Descriptive statistics summarized participant characteristics and assessed feasibility and

XSL•FO
RenderX

acceptability outcomes. Normality was evaluated using the Shapiro-Wilk test. Preliminary effects were examined using 2-tailed paired *t* tests or Wilcoxon signed-rank tests for within-group comparisons and independent *t* tests or Mann-Whitney *U* tests for between-group comparisons. Generalized estimating equation models with an independence working correlation structure and robust SEs were used to assess between-group differences across multiple time points. Missing data were assessed using the Little missing completely at random (MCAR) test to evaluate the missingness pattern. All analyses followed the intention-to-treat principle. Effect sizes were calculated using the Cohen *d* (small effect size: $d \leq 0.2$; medium effect size: $d \leq 0.5$; large effect size: $d \geq 0.8$) [48].

Ethical Considerations

This study was approved by the ethics committees of Hong Kong Polytechnic University (HSEARS20240904005) and the Affiliated Hospital of Qingdao University (QYFYWZLL29400). Written informed consent was obtained from all participating patients and caregivers before enrollment. To protect participants' privacy and confidentiality, study data were deidentified and stored securely, with access restricted to authorized research personnel. Participants received no financial compensation for taking part in the study. No identifiable images of participants are included in the manuscript or supplementary materials.

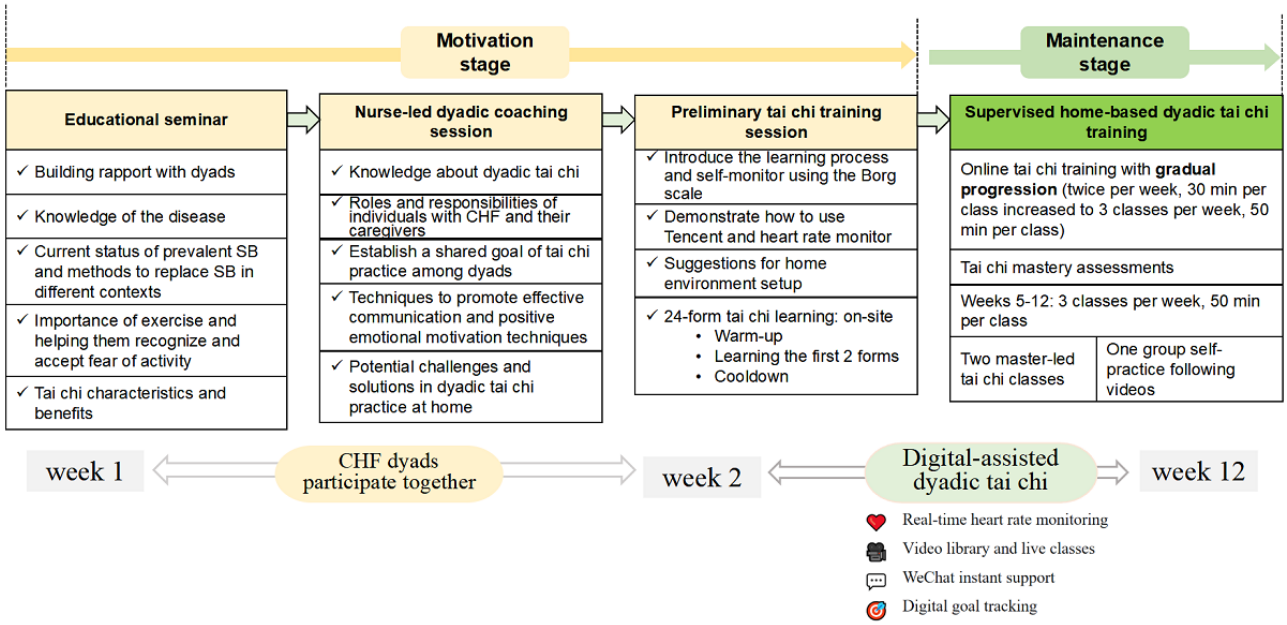
Results

HDTC Intervention Protocol

Intervention Components and Delivery

The HDTC intervention was informed by multiple sources: a systematic review on digital health interventions for sedentary behavior [49], a scoping review of caregiver roles and engagement levels [50], and focus group interviews with stakeholders (Multimedia Appendix 3). The final intervention comprised 4 components delivered in a structured sequence: an online educational seminar and a nurse-led online dyadic coaching session in week 1, an introductory face-to-face training session of 24-form Yang-style tai chi in week 2, and nurse-supervised HDTC classes (Figure 2). The tai chi sessions were delivered by a master from the Qingdao Yang Style Tai Chi Association with 11 years of teaching experience. During weeks 2 to 4, dyads attended live online tai chi classes, with the frequency and duration progressively increasing from two 30-minute sessions to three 50-minute sessions per week. During weeks 5 to 12, dyads participated in 2 master-led online sessions and 1 video-guided group practice session each week, with each session lasting 50 minutes. During supervised online sessions, a cardiac nurse monitored participants' exercise-related symptoms and heart rate and provided safety guidance as needed.

Figure 2. Protocol for the home-based, digital-assisted, dyadic tai chi intervention. The intervention comprised 4 components delivered across 2 stages: the motivation stage and the maintenance stage. Digital support included real-time heart rate monitoring, instructional videos, live online classes, WeChat-based support, and goal tracking. CHF: chronic heart failure; SB: sedentary behavior.



Digitally Assisted Features

The HDTC intervention was supported by a multifunctional digital infrastructure. Synchronous live video sessions and an on-demand video library delivered structured, progressive tai chi instruction. Real-time heart rate monitoring enhanced safety by enabling continuous oversight of patients during home-based practice. Asynchronous communication via WeChat offered ongoing access to nursing support and peer interaction, thereby reinforcing adherence. Meanwhile, digital goal-tracking features, such as shared progress updates and reminders in the online community via WeChat group, fostered shared accountability and motivation within the dyad. Together, these digitally enabled components created a supportive and interactive remote environment that effectively extends clinical supervision into the home and enhances both engagement and safety.

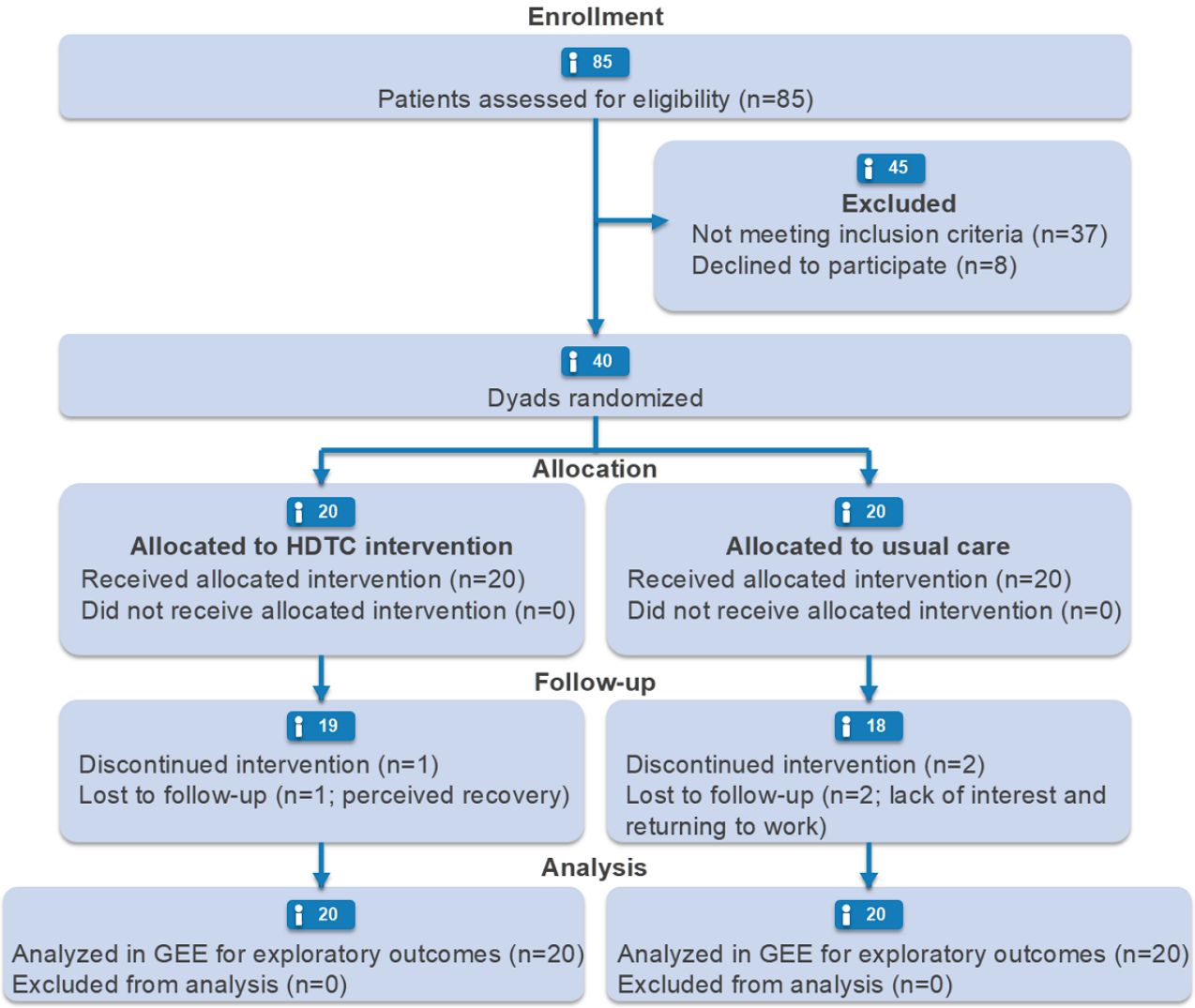
XSL•FO
RenderX

Feasibility, Acceptability, and Effects of the Intervention

Baseline Characteristics of the Dyads

The dyads were recruited from a tertiary medical center in Qingdao, China, from October 14 to November 12, 2024. A total of 85 patients (62 online; 23 in person) were referred to this study. During initial screening, 37 patients were excluded. Of the 48 eligible dyads, 8 declined participation due to scheduling conflicts (caregivers: n=4; patients with CHF: n=2) or lack of interest (n=2). The remaining 40 dyads were randomly assigned to the intervention or control group. All dyads completed the 4-week assessment. However, 3 dyads (1 in the intervention group and 2 in the control group) did not complete the 12-week follow-up because of perceived recovery, lack of interest, or returning to work (Figure 3) [51]. No item-level missing data were identified among completed questionnaires. The Little MCAR test was not statistically significant (χ^2_{16}=13.1; *P*=.66), providing no evidence against the MCAR assumption. Baseline characteristics of dyads were comparable between groups (Table 1).

Figure 3. CONSORT (Consolidated Standards of Reporting Trials) 2025 flow diagram of recruitment, randomization, follow-up, and analysis in the pilot randomized controlled trial [51]. GEE: generalized estimating equations; HDTC: home-based, digital-assisted, dyadic tai chi.



XSL•FO
RenderX

Table 1. Baseline characteristics of patients with chronic heart failure (CHF) and caregivers (N=40).

Characteristics	Total	IG[a] (n=20)	CG[b] (n=20)	χ^2/*t* test	*P* value
Patients with CHF					
Age (y), mean (SD)	61.43 (8.55)	60.65 (9.71)	62.20 (7.39)	0.568[c]	.57
Sex, n (%)				1.129[d]	.29
Male	29 (72.5)	13 (65)	16 (80)		
Female	11 (27.5)	7 (35)	4 (20)		
Educational level, n (%)				0.128[d]	.94
Elementary school	2 (5)	1 (5)	1 (5)		
Secondary school	27 (67.5)	13 (65)	14 (70)		
College or higher	11 (27.5)	6 (30)	5 (25)		
Family monthly income (CN¥; CN¥ 1=US $0.15 as of October 16, 2026), n (%)				2.349[d]	.50
<3000	15 (37.5)	8 (40)	7 (35)		
3000-5000	18 (45)	7 (35)	11 (55)		
5000-10,000	4 (10)	3 (15)	1 (5)		
>10,000	3 (7.5)	2 (10)	1 (5)		
Time since diagnosis (mo), n (%)				1.933[d]	.38
≤6	28 (70)	12 (60)	16 (80)		
6-12	3 (7.5)	2 (10)	1 (5)		
>12	9 (22.5)	6 (30)	3 (15)		
NYHA[e] class, n (%)				1.010[d]	.60
I	12 (30)	7 (35)	5 (25)		
II	13 (32.5)	7 (35)	6 (30)		
III	15 (37.5)	6 (30)	9 (45)		
Etiology of HF[f], n (%)				2.371[g]	.67
Coronary artery disease	30 (75)	14 (70)	16 (80)		
Cardiomyopathies	2 (5)	1 (5)	1 (5)		
Hypertension	21 (52.5)	8 (40)	13 (65)		
Valvular heart disease	6 (15)	4 (20)	2 (10)		
Arrhythmias	7 (17.5)	2 (10)	5 (25)		
Comorbidity types, n (%)				<0.001[d]	>.99
0	4 (10)	2 (10)	2 (10)		
1-2	16 (40)	8 (40)	8 (40)		
≥3	20 (50)	10 (50)	10 (50)		
Caregivers					
Age (y), mean (SD)	54.80 (10.25)	54.15 (9.99)	55.45 (10.71)	0.397[c]	.69
Sex, n (%)				<0.001[d]	>.99
Male	12 (30)	6 (30)	6 (30)		
Female	28 (70)	14 (70)	14 (70)		
Educational level, n (%)				0.451[g]	.80
Elementary school	3 (7.5)	1 (5)	2 (10)		
Secondary school	28 (70)	14 (70)	14 (70)		

XSL•FO
RenderX

Characteristics	Total	IG[a] (n=20)	CG[b] (n=20)	χ^2/*t* test	*P* value
College or higher	9 (22.5)	5 (25)	4 (20)		
Caregiver type, n (%)				3.584[d]	.06
Spouse	31 (77.5)	18 (90)	13 (65)		
Child	9 (22.5)	2 (10)	7 (35)		
Caring h per d, mean (SD)	3.18 (1.89)	3.20 (1.51)	3.15 (2.25)	−0.082[c]	.94

[a]IG: intervention group.

[b]CG: control group.

[c]Independent *t* test.

[d]Pearson chi-square test.

[e]NYHA: New York Heart Association.

[f]HF: heart failure.

[g]Likelihood ratio.

Feasibility Outcomes

Of the 85 patients referred for screening (average recruitment rate: 2.8 dyads per day; n=85, 100% response rate), 48 were eligible, and 40 dyads were enrolled and randomized, corresponding to a recruitment rate of 83.3% (40/48). Among interested patients (n=85), primary motivators for participation were perceived benefits (68/85, 80%), convenience of home-based delivery (47/85, 55.3%), and insufficient postdiagnosis exercise knowledge (38/85, 44.7%; Table S1 in Multimedia Appendix 4). Retention rates were 100% (40/40) at 4 weeks and 92.5% (37/40) at 12 weeks. Mean attendance to online tai chi classes was 80.6% (SD 14.9%) at 4 weeks and 73.7% (SD 8.6%) at 12 weeks, with 70% (14/20) of participants demonstrating high adherence (≥75% attendance).

Acceptability Outcomes

The detailed responses from the dyadic perspective were shown in Table S2 in Multimedia Appendix 4. Notably, dyads reported positive appraisals of the HDTC intervention. Overall, 100% (20/20) of dyads rated the intervention as acceptable, with 60% (12/20) deeming it completely acceptable.

Safety of the Intervention

During each tai chi class, the dyads' physiological responses were closely monitored and documented by a cardiology nurse. Classes began with warm-up exercises and concluded with cooldown stretches to minimize musculoskeletal injury risk. No intervention-related adverse events were reported during the 12-week intervention, suggesting that it was well tolerated in this pilot sample.

Preliminary Effects of the HDTC Intervention

Effects of the HDTC Intervention for Patients With CHF

Compared to the control group, patients with CHF in the intervention group exhibited significant improvement in 6-minute walk test distance at both 4 weeks (β=12.300, 95% CI 8.733-15.867; *P*<.001) and 12 weeks (β=38.858, 95% CI 15.691-62.025; *P*=.001; Table S3 and Figure S1 in Multimedia Appendix 4). Significant reductions were observed in sedentary time (T1: β=−0.975, 95% CI −1.338 to −0.612; *P*<.001; T2: β=−1.025, 95% CI −1.617 to −0.433; *P*=.01), fear of activity (T1: β=−2.050, 95% CI −3.865 to −0.235; *P*=.03; T2: β=−6.550, 95% CI −11.548 to −1.552; *P*=.01), and Minnesota Living With Heart Failure Questionnaire scores (T2: β=−1.850, 95% CI −3.045 to −0.655; *P*=.002). Effect sizes for these outcomes ranged from moderate to large (Table 2).

XSL•FO
RenderX

Table 2. Patient, caregiver, and dyadic outcome comparisons at baseline (T0), 4 weeks (T1), and 12 weeks (T2; N=40).

Outcomes and time points	IG[a] (n=20), mean (SD)	CG[b] (n=20), mean (SD)	Effect size, Cohen *d*	*t* test (*df*)	*P* value
Patients with CHF[c]					
6MWT[d] (m)					
T0	415.35 (94.18)	413.70 (81.76)	—[e]	−0.059 (38)	.95
T1	435.95 (93.38)	422.00 (79.80)	0.16	−0.508 (38)	.62
T2	474.16 (75.52)	433.65 (74.50)	0.54	−1.719 (38)	.09
Exercise self-efficacy (0-90)					
T0	41.50 (12.34)	39.95 (13.28)	—	−0.383 (38)	.70
T1	45.00 (10.88)	40.75 (12.13)	0.37	−1.167 (38)	.25
T2	49.45 (10.27)	41.05 (11.61)	0.77	−2.423 (38)	.02
Sedentary time (h)					
T0	9.30 (1.67)	9.40 (1.25)	—	0.215 (38)	.83
T1	8.28 (1.37)	9.35 (1.13)	0.85	2.710 (38)	.01
T2	7.58 (0.80)	8.70 (1.07)	1.18	3.770 (38)	.001
Fear of activity (0-60)					
T0	32.35 (7.85)	32.80 (7.90)	—	0.181 (38)	.86
T1	27.30 (7.80)	29.80 (9.06)	0.30	0.935 (38)	.36
T2	22.50 (8.31)	29.50 (9.74)	0.77	2.370 (38)	.02
MLHFQ[f] (0-105)					
T0	59.05 (9.59)	60.05 (12.14)	—	0.289 (38)	.77
T1	50.85 (10.51)	56.80 (11.98)	0.53	1.670 (38)	.10
T2	40.70 (10.76)	53.05 (11.99)	1.08	3.428 (38)	.001
Mutuality (0-60)					
T0	37.45 (5.87)	37.05 (8.24)	—	−0.177 (38)	.86
T1	43.05 (9.45)	36.95 (7.70)	0.71	−2.239 (38)	.03
T2	48.16 (8.90)	37.25 (7.59)	1.32	−4.124 (38)	<.001
Caregivers					
Care burden (0-88)					
T0	26.35 (7.76)	26.10 (7.69)	—	0.102 (38)	.92
T1	24.25 (7.22)	25.75 (7.89)	0.20	−0.627 (38)	.54
T2	22.60 (6.27)	26.70 (6.45)	0.64	−2.039 (38)	.048
Personal burden (0-48)					
T0	9.50 (3.58)	9.95 (3.14)	—	−0.423 (38)	.68
T1	8.90 (3.11)	9.70 (3.37)	0.25	−0.780 (38)	.44
T2	8.25 (2.40)	9.90 (2.79)	0.63	−2.004 (38)	.05
Role burden (0-36)					
T0	16.85 (5.20)	16.15 (5.39)	—	0.418 (38)	.68
T1	15.35 (5.01)	16.05 (5.34)	0.14	−0.428 (38)	.67
T2	14.35 (4.51)	16.80 (5.00)	0.51	−1.627 (38)	.11
Quality of life (16-80)					
T0	49.75 (6.08)	47.60 (4.56)	—	−1.266 (38)	.21
T1	52.50 (5.52)	49.55 (3.25)	0.60	−2.059 (38)	.048

XSL•FO
RenderX

Outcomes and time points	IG[a] (n=20), mean (SD)	CG[b] (n=20), mean (SD)	Effect size, Cohen *d*	*t* test (*df*)	*P* value
T2	52.55 (5.03)	49.10 (3.13)	0.82	−2.605 (38)	.01
Mutuality (0-60)					
T0	37.75 (10.28)	37.40 (10.28)	—	−0.109 (38)	.91
T1	42.50 (12.01)	38.30 (9.14)	0.39	−1.245 (38)	.22
T2	44.95 (12.43)	38.20 (7.35)	0.66	−2.090 (38)	.04

[a]IG: intervention group.

[b]CG: control group.

[c]CHF: chronic heart failure.

[d]6MWT: 6-minute walk test.

[e]Not applicable.

[f]MLHFQ: Minnesota Living With Heart Failure Questionnaire.

Effects of the HDTC Intervention for Caregivers

Caregivers in the intervention group demonstrated significant reduction in Zarit Caregiver Burden Interview scores compared to controls at both 4 weeks (β=−1.750, 95% CI −2.499 to −1.001; *P*<.001) and 12 weeks (β=−4.350, 95% CI −5.765 to −2.935; *P*<.001), with effect sizes ranging from small to moderate (Table 2). However, no significant between-group differences were observed in Family Caregiver Quality of Life scores (β=1.300, 95% CI −1.644 to 4.244; *P*=.39; Table S3 and Figure S2 in Multimedia Appendix 4).

Effects of the HDTC Intervention for Dyads

Both patients with CHF and caregivers in the intervention group showed significant improvements in mutuality scores at 4 weeks (patients with CHF: β=5.700, 95% CI 1.126-10.274; *P*=.02; caregivers: β=3.850, 95% CI 1.961-5.739; *P*<.001) and 12 weeks (patients with CHF: β=10.508, 95% CI 5.534-15.482; *P*<.001; caregivers: β=6.450, 95% CI 3.652-9.248; *P*<.001; Table S3 and Figure S3 in Multimedia Appendix 4).

Discussion

Principal Findings

The potential therapeutic benefits of tai chi have been increasingly investigated [52,53], but dyadic tai chi interventions among patients with CHF and their caregivers remain under-studied [13]. This study developed and evaluated an HDTC intervention that engaged patients and caregivers as active participants in a shared rehabilitation process, thereby extending conventional patient-centered exercise toward a dyad-oriented model. The intervention demonstrated satisfactory recruitment, retention, attendance, and acceptability, providing initial evidence that this approach is feasible in the home setting. No intervention-related adverse events were reported during the 12-week intervention, although the small sample size limits conclusions regarding safety. Exploratory analyses also indicated preliminary improvements across patient, caregiver, and dyadic outcomes. However, this study was not powered to establish clinical effectiveness. If confirmed in a fully powered trial, this model may offer a practical strategy for expanding home-based cardiac rehabilitation while addressing the needs of both patients and caregivers.

Feasibility of the HDTC Intervention Among Dyads

The HDTC intervention demonstrated encouraging recruitment, with enrollment completed within 1 month. The overall recruitment rate was 83.3% (40/48), comparable to the 75.1% reported in a previous dyadic study [14]. Although dyadic interventions often face recruitment challenges because of the need to coordinate participation by both members [54,55], recruitment in this study may have been facilitated by multiple recruitment pathways and close collaboration with the Department of Cardiology at the Affiliated Hospital of Qingdao University.

Notably, the relatively low refusal rate may have been due to nurses' extensive experience with patients with CHF [54], stakeholder-informed tailoring of the intervention, and its home-based delivery [56]. Participant feedback also suggested that dyads perceived the intervention as beneficial, and more than half valued the home-based convenience. Feasibility was further supported by complete retention at 4 weeks and a satisfactory retention rate at 12 weeks. Free access, convenient home-based practice, and social support from caregivers and nurses may have contributed to sustained engagement and retention [57]. Mean attendance was consistent with findings from a previous tai chi study [58], while adherence was comparable to that reported in a previous home-based cardiac rehabilitation study (14/20, 70% vs 73%) [59]. Collectively, these findings support the feasibility and acceptability of the HDTC intervention in this pilot setting.

Acceptability of the HDTC Intervention Among Dyads

All dyads in the intervention group rated the HDTC intervention as acceptable, with 60% (12/20) rating it as completely acceptable. This high acceptability may reflect the intervention's dyadic tailoring and home-based delivery, which may also have promoted participant engagement and adherence [60]. Findings align with evidence that tai chi has been reported to be feasible and well tolerated among patients with frailty [61-62]. Its acceptability may have been further enhanced by caregiver involvement, which could provide encouragement and reinforce patients' motivation to engage in regular exercise [63].

Regarding engagement, 45% (9/20) of patients with CHF reported that the intervention required substantial effort, whereas 70% (14/20) of caregivers described it as moderately effortful.

XSL•FO
RenderX

Informal telephone follow-up suggested that difficulties in learning tai chi forms contributed to the perceived effort, a challenge also reported in previous studies involving older adults [64]. This highlights the need for refinements in a full-scale RCT. Suggested strategies include (1) enhanced instructional support with clearer, more detailed demonstrations; (2) incremental learning of new forms; and (3) pretraining videos coupled with regular review sessions. Together, these approaches aim to ease skill acquisition, foster independent practice, and support long-term retention.

Preliminary Effects of the HDTC Intervention

For patients with CHF, the intervention showed favorable preliminary effects on exercise tolerance, sedentary time, fear of activity, exercise self-efficacy, and health-related QoL, generally consistent with previous evidence on tai chi and home-based rehabilitation [21,49]. These changes may reflect the gradual, low-impact nature of tai chi and the intervention's emphasis on replacing sedentary behavior with structured home exercise, while supervised practice and progressive skill acquisition may also have supported confidence and reduced fear of activity [65,66]. Given the pilot design, these findings should be interpreted as preliminary signals rather than definitive evidence of efficacy.

For caregivers, the intervention was associated with reduced caregiver burden, whereas QoL did not show a significant improvement, possibly because broader QoL is influenced by multiple factors, including caregiving distress and competing social and family demands, and may therefore be less responsive over a short intervention period [13,67,68]. At the dyadic level, both patients and caregivers showed improved mutuality, which may reflect the intervention's emphasis on joint decision-making, shared exercise participation, and mutual encouragement, in contrast to dyadic interventions focused primarily on patient self-care [69]. These findings suggest that dyadic exercise may influence not only individual outcomes but also the relational processes within the patient-caregiver pair.

Strengths and Limitations

A key strength of this study was its explicit focus on dyads of patients with CHF and their caregivers throughout intervention conceptualization, development, and evaluation, addressing an important gap in previous research [13]. The HDTC intervention was developed through an evidence-based approach incorporating a comprehensive literature review, theoretical frameworks, and stakeholder consultation, thereby enhancing its conceptual foundation and contextual relevance.

Several limitations should be considered. First, the pilot sample was small and was not powered for definitive between-group testing of clinical outcomes; consequently, effect estimates are imprecise and exploratory. Second, the 12-week follow-up does not establish whether engagement or outcome changes are sustained after intervention completion. Third, participants were recruited from a single tertiary medical center and through voluntary recruitment pathways, which may limit generalizability to other health care settings and may have preferentially enrolled dyads who were more motivated to exercise or use digital technology. Fourth, requiring both members of the dyad to participate may have further selected families with greater caregiver availability and relational capacity. Fifth, although both patient and caregiver outcomes were measured, the small sample limited the ability to reliably estimate dyadic interdependence. Finally, feasibility estimates such as recruitment and adherence are context dependent and require confirmation in larger, multicenter settings.

Conclusions

The HDTC intervention demonstrated feasibility and acceptability among dyads, with preliminary evidence suggesting potential to improve dyadic health outcomes. These findings support further evaluation of this dyad-oriented, digitally supported home rehabilitation model in a sufficiently powered multicenter RCT with longer follow-up.

Acknowledgments

The authors express their sincere gratitude to the physicians and nurses of the Department of Cardiology, Affiliated Hospital of Qingdao University, for their professional support and to the tai chi master for providing expert instruction. They further extend their deep appreciation to the patients with chronic heart failure and caregivers for their participation in this study. The authors declare the use of generative AI in the manuscript preparation and revision process. According to the Generative AI Delegation Taxonomy (2026), the following tasks were delegated to generative AI tools under full human supervision: proofreading and editing, adapting and adjusting emotional tone, and reformatting. The generative AI tool used was ChatGPT (GPT-6; OpenAI). Responsibility for the final manuscript lies entirely with the authors. Generative AI tools are not listed as authors and do not bear responsibility for the final outcomes.

Funding

This work was supported by the General Research Fund, Research Grants Council of the University Grants Committee of Hong Kong (grant 15100822); the Affiliated Hospital of Qingdao University (grant QDFY+X2024206); and Shandong Province Traditional Chinese Medicine Science and Technology Project (grant M20244004).

Data Availability

The data that support the findings of this study are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Generative AI Use

The generative AI tool used was ChatGPT (GPT-6; OpenAI). Responsibility for the final manuscript lies entirely with the authors.

Authors' Contributions

Conceptualization: YZ, FWN, YJX
Formal analysis: YZ
Methodology: YZ, LC, XL, YJX
Project administration: YZ, LC, XL, WZ
Supervision: FWN, WZ, LW, YJX
Validation: YJX
Writing—original draft: YZ
Writing—review and editing: YZ, FWN, LC, LW, YJX

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist.
[PDF File (Adobe PDF File), 3062 KB - mhealth_v14i1e95689_app1.pdf]

Multimedia Appendix 2
Interview guide for the focus group interviews.
[DOCX File , 13 KB - mhealth_v14i1e95689_app2.docx]

Multimedia Appendix 3
Results of the focus group interviews.
[DOCX File , 22 KB - mhealth_v14i1e95689_app3.docx]

Multimedia Appendix 4
Supplementary analyses and longitudinal changes in patient, caregiver, and dyadic outcomes.
[DOCX File , 418 KB - mhealth_v14i1e95689_app4.docx]

References

1. Savarese G, Becher PM, Lund LH, Seferovic P, Rosano GM, Coats AJ. Global burden of heart failure: a comprehensive and updated review of epidemiology. Cardiovasc Res 2023 Jan 18;118(17):3272-3287 [FREE Full text] [doi: 10.1093/cvr/cvac013] [Medline: 35150240]
2. Wang H, Liu Y, Yang J. Epidemiology of heart failure. J Clin Cardiol 2023;39(4):243-247. [doi: 10.13201/j.issn.1001-1439.2023.04.001]
3. Chen MF, Ke SR, Liu CL, Wu TC, Yu YM, Chiou AF. Associated factors and impacts of sedentary behaviour in patients with heart failure: a longitudinal study. Eur J Cardiovasc Nurs 2020 Oct;19(7):609-618 [FREE Full text] [doi: 10.1177/1474515120912381] [Medline: 32338530]
4. Dibben GO, Faulkner J, Oldridge N, Rees K, Thompson DR, Zwisler AD, et al. Exercise-based cardiac rehabilitation for coronary heart disease: a meta-analysis. Eur Heart J 2023 Feb 07;44(6):452-469 [FREE Full text] [doi: 10.1093/eurheartj/ehac747] [Medline: 36746187]
5. Molloy C, Long L, Mordi IR, Bridges C, Sagar VA, Davies EJ, et al. Exercise-based cardiac rehabilitation for adults with heart failure. Cochrane Database Syst Rev 2024 Mar 07;3(3):CD003331. [doi: 10.1002/14651858.CD003331.pub6] [Medline: 38451843]
6. Hwang HF, Tseng KC, Chen SJ, Yu WY, Chen CY, Lin MR. Effects of home-based computerized cognitive training and Tai Chi exercise on cognitive functions in older adults with mild cognitive impairment. Aging Ment Health 2023;27(11):2170-2178. [doi: 10.1080/13607863.2023.2225430] [Medline: 37365961]
7. Buck HG, Howland C, Stawnychy MA, Aldossary H, Cortés YI, DeBerg J, et al. Caregivers' contributions to heart failure self-care: an updated systematic review. J Cardiovasc Nurs 2024;39(3):266-278 [FREE Full text] [doi: 10.1097/JCN.0000000000001060] [Medline: 38306302]
8. Jackson JD, Cotton SE, Bruce Wirta S, Proenca CC, Zhang M, Lahoz R, et al. Burden of heart failure on caregivers in China: results from a cross-sectional survey. Drug Des Devel Ther 2018 Jun 08;12:1669-1678 [FREE Full text] [doi: 10.2147/DDDT.S148970] [Medline: 29922041]

XSL•FO
RenderX

9. Dirikkan F, Baysan Arabacı L, Mutlu E. The caregiver burden and the psychosocial adjustment of caregivers of cardiac failure patients. Turk Kardiyol Dern Ars 2018 Dec;46(8):692-701 [FREE Full text] [doi: 10.5543/tkda.2018.10.5543/tkda.2018.69057] [Medline: 30516527]
10. Suksatan W, Tankumpuan T, Davidson PM. Heart failure caregiver burden and outcomes: a systematic review. J Prim Care Community Health 2022;13:21501319221112584 [FREE Full text] [doi: 10.1177/21501319221112584] [Medline: 35938489]
11. Yu DS, Qiu C, Li PW, Lau J, Riegel B. Effects of dyadic care interventions for heart failure on patients' and caregivers' outcomes: a systematic review, meta-analysis and meta-regression. Int J Nurs Stud 2024 Sep;157:104829. [doi: 10.1016/j.ijnurstu.2024.104829] [Medline: 38901123]
12. Di Maio S, Villinger K, Knoll N, Scholz U, Stadler G, Gawrilow C, et al. Compendium of dyadic intervention techniques (DITs) to change health behaviours: a systematic review. Health Psychol Rev 2024 Sep;18(3):538-573 [FREE Full text] [doi: 10.1080/17437199.2024.2307534] [Medline: 38437798]
13. Bidwell JT, Conway C, Babicheva V, Lee CS. Person with heart failure and care partner dyads: current knowledge, challenges, and future directions: state-of-the-art review. J Card Fail 2023 Aug;29(8):1187-1206 [FREE Full text] [doi: 10.1016/j.cardfail.2023.02.017] [Medline: 36958392]
14. Yang W, Sun L, Hao L, Zhang X, Lv Q, Xu X, et al. Effects of the family customised online FOCUS programme on patients with heart failure and their informal caregivers: a multicentre, single-blind, randomised clinical trial. EClinicalMedicine 2024 Feb 12;69:102481 [FREE Full text] [doi: 10.1016/j.eclinm.2024.102481] [Medline: 38370538]
15. Zhao Y, Zhang S, Lu M, Zhen Q, Shi X. Patient and informal caregiver-centered nursing interventions for adults with heart failure: a systematic review and meta-analysis. Intensive Crit Care Nurs 2025 Jun;88:103943. [doi: 10.1016/j.iccn.2025.103943] [Medline: 40049039]
16. Doyle KL, Toepfer M, Bradfield AF, Noffke A, Ausderau KK, Andreae S, et al. Systematic review of exercise for caregiver-care recipient dyads: what is best for spousal caregivers-exercising together or not at all? Gerontologist 2021 Aug 13;61(6):e283-e301 [FREE Full text] [doi: 10.1093/geront/gnaa043] [Medline: 32614050]
17. Lyons KS, Lee CS. The theory of dyadic illness management. J Fam Nurs 2018 Feb;24(1):8-28 [FREE Full text] [doi: 10.1177/1074840717745669] [Medline: 29353528]
18. Chow SL, Bozkurt B, Baker WL, Bleske BE, Breathett K, Fonarow GC, et al. Complementary and alternative medicines in the management of heart failure: a scientific statement from the American Heart Association. Circulation 2023 Jan 10;147(2):e4-30 [FREE Full text] [doi: 10.1161/CIR.0000000000001110] [Medline: 36475715]
19. Professional Committee of Cardiovascular Prevention and Rehabilitation of Chinese Rehabilitation Medical Association. Expert consensus on cardiac rehabilitation for chronic heart failure in China [Article in Chinese]. Zhonghua Nei Ke Za Zhi 2020 Dec 01;59(12):942-952. [doi: 10.3760/cma.j.cn112138-20200309-00210] [Medline: 33256335]
20. Huang J, Qin X, Shen M, Xu Y, Huang Y. The effects of Tai Chi exercise among adults with chronic heart failure: an overview of systematic review and meta-analysis. Front Cardiovasc Med 2021 Mar 18;8:589267 [FREE Full text] [doi: 10.3389/fcvm.2021.589267] [Medline: 33816570]
21. Hui J, Wang Y, Zhao J, Cong W, Xu F. Effects of Tai Chi on health status in adults with chronic heart failure: a systematic review and meta-analysis. Front Cardiovasc Med 2022 Sep 9;9:953657 [FREE Full text] [doi: 10.3389/fcvm.2022.953657] [Medline: 36158796]
22. Du Y, Roberts P, Liu W. Facilitators and barriers of Tai Chi practice in community-dwelling older adults: qualitative study. Asian Pac Isl Nurs J 2023 Jan 23;7:e42195 [FREE Full text] [doi: 10.2196/42195] [Medline: 36720122]
23. Teo JL, Bird SR, Wang X, Zheng Z. Using telehealth to deliver Qi Gong and Tai Chi programs: a mixed-methods systematic review on feasibility, acceptability and participant engagement factors. Arch Gerontol Geriatr 2024 Feb;117:105203 [FREE Full text] [doi: 10.1016/j.archger.2023.105203] [Medline: 37741135]
24. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. Framework for the development and evaluation of complex interventions: gap analysis, workshop and consultation-informed update. Health Technol Assess 2021 Sep;25(57):1-132 [FREE Full text] [doi: 10.3310/hta25570] [Medline: 34590577]
25. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016 Oct 24;355:i5239 [FREE Full text] [doi: 10.1136/bmj.i5239] [Medline: 27777223]
26. Ferrari M, Speight J, Beath A, Browne JL, Mosely K. The information-motivation-behavioral skills model explains physical activity levels for adults with type 2 diabetes across all weight classes. Psychol Health Med 2021 Mar;26(3):381-394. [doi: 10.1080/13548506.2020.1749292] [Medline: 32266821]
27. Fisher WA, Fisher JD, Harman J. The information-motivation-behavioral skills model: a general social psychological approach to understanding and promoting health behavior. In: Suls J, Wallston KA, editors. Social Psychological Foundations of Health and Illness. Oxford, UK: Blackwell Publishing; Jan 01, 2003.
28. Greenhalgh T, Abimbola S. The NASSS framework - a synthesis of multiple theories of technology implementation. Stud Health Technol Inform 2019 Jul 30;263:193-204. [doi: 10.3233/SHTI190123] [Medline: 31411163]
29. Chinese Society of Cardiology, Chinese Medical Association, Chinese College of Cardiovascular Physician, Chinese Heart Failure Association of Chinese Medical Doctor Association, Editorial Board of Chinese Journal of Cardiology. Chinese

XSL•FO
RenderX

guidelines for the diagnosis and treatment of heart failure 2024 [Article in Chinese]. Zhonghua Xin Xue Guan Bing Za Zhi 2024 Mar 24;52(3):235-275. [doi: 10.3760/cma.j.cn112148-20231101-00405] [Medline: 38514328]
30. Castro AR, Arnaert A, Moffatt K, Kildea J, Bitzas V, Tsimicalis A. "Informal caregiver" in nursing: an evolutionary concept analysis. ANS Adv Nurs Sci 2023;46(1):E29-E42. [doi: 10.1097/ANS.0000000000000439] [Medline: 36006014]
31. Whitehead AL, Julious SA, Cooper CL, Campbell MJ. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. Stat Methods Med Res 2016 Jun;25(3):1057-1073 [FREE Full text] [doi: 10.1177/0962280215588241] [Medline: 26092476]
32. Sealed Envelope. URL: https://www.sealedenvelope.com/ [accessed 2026-10-19]
33. Boutron I, Tubach F, Giraudeau B, Ravaud P. Blinding was judged more difficult to achieve and maintain in nonpharmacologic than pharmacologic trials. J Clin Epidemiol 2004 Jun;57(6):543-550. [doi: 10.1016/j.jclinepi.2003.12.010] [Medline: 15246122]
34. Bellg AJ, Borrelli B, Resnick B, Hecht J, Minicucci DS, Ory M, et al. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. Health Psychol 2004 Sep;23(5):443-451. [doi: 10.1037/0278-6133.23.5.443] [Medline: 15367063]
35. Lackinger C, Grabovac I, Haider S, Kapan A, Winzer E, Stein KV, et al. Adherence is more than just being present: example of a lay-led home-based programme with physical exercise, nutritional improvement and social support, in prefrail and frail community-dwelling older adults. Int J Environ Res Public Health 2021 Apr 15;18(8):4192 [FREE Full text] [doi: 10.3390/ijerph18084192] [Medline: 33920981]
36. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
37. Barr A, Coates E, Kingsley E, de la Cuesta GG, Biggs K, Le Couteur A, et al. A mixed methods evaluation of the acceptability of therapy using LEGO® bricks (LEGO® based therapy) in mainstream primary and secondary education. Autism Res 2022 Jul;15(7):1237-1248 [FREE Full text] [doi: 10.1002/aur.2725] [Medline: 35396821]
38. Junqueira DR, Zorzela L, Golder S, Loke Y, Gagnier JJ, Julious SA, et al. CONSORT Harms 2022 statement, explanation, and elaboration: updated guideline for the reporting of harms in randomised trials. BMJ 2023 Apr 24;381:e073725 [FREE Full text] [doi: 10.1136/bmj-2022-073725] [Medline: 37094878]
39. Giannitsi S, Bougiakli M, Bechlioulis A, Kotsia A, Michalis LK, Naka KK. 6-minute walking test: a useful tool in the management of heart failure patients. Ther Adv Cardiovasc Dis 2019;13:1753944719870084 [FREE Full text] [doi: 10.1177/1753944719870084] [Medline: 31441375]
40. WOCaring. URL: http://www.wocaring.com/ [accessed 2026-10-19]
41. Macfarlane DJ, Lee CCY, Ho EYK, Chan KL, Chan DTS. Reliability and validity of the Chinese version of IPAQ (short, last 7 days). J Sci Med Sport 2007 Feb;10(1):45-51. [doi: 10.1016/j.jsams.2006.05.003] [Medline: 16807105]
42. Ozyemisci-Taskiran O, Demirsoy N, Atan T, Yuksel S, Coskun O, Aytur YK, et al. Development and validation of a scale to measure fear of activity in patients with coronary artery disease (Fact-CAD). Arch Phys Med Rehabil 2020 Mar;101(3):479-486. [doi: 10.1016/j.apmr.2019.09.001] [Medline: 31562874]
43. Liu T, Chan AW, Chair SY. Group- plus home-based Tai Chi program improves functional health among patients with coronary heart disease: a randomized controlled trial. Eur J Cardiovasc Nurs 2022 Aug 29;21(6):597-611 [FREE Full text] [doi: 10.1093/eurjcn/zvab126] [Medline: 35022661]
44. Mogle J, Buck H, Zambroski C, Alvaro R, Vellone E. Cross-validation of the Minnesota Living With Heart Failure Questionnaire. J Nurs Scholarsh 2017 Sep;49(5):513-520. [doi: 10.1111/jnu.12318] [Medline: 28755434]
45. Ko KT, Yip PK, Liu SI, Huang CR. Chinese version of the Zarit Caregiver Burden Interview: a validation study. Am J Geriatr Psychiatry 2008 Jun;16(6):513-518. [doi: 10.1097/jgp.0b013e318167ae5b]
46. Qian H, Wang J, Shen W, Miao A, Yu Y. The Chinese version of the Family Caregiver-Specific Quality of Life Scale: validation among community family caregivers of chronic heart failure patients [Article in Chinese]. J Nurs Sci 2012;27(21):85-87. [doi: 10.3870/hlxzz.2012.21.085]
47. Dellafiore F, Buck HG, Pucciarelli G, Barbaranelli C, Paturzo M, Alvaro R, et al. Psychometric characteristics of the mutuality scale in heart failure patients and caregivers. Heart Lung 2018 Nov;47(6):553-561. [doi: 10.1016/j.hrtlng.2018.05.018] [Medline: 29887127]
48. Sullivan GM, Feinn R. Using effect size-or why the P value is not enough. J Grad Med Educ 2012 Sep;4(3):279-282 [FREE Full text] [doi: 10.4300/JGME-D-12-00156.1] [Medline: 23997866]
49. Zhang Y, Ngai FW, Yang Q, Xie YJ. Effectiveness of digital health interventions on sedentary behavior among patients with chronic diseases: systematic review and meta-analysis. JMIR Mhealth Uhealth 2025 Jun 24;13:e59943 [FREE Full text] [doi: 10.2196/59943] [Medline: 40554804]
50. Zhang Y, Ngai FW, Liang J, Yang Q, Zhao M, Wei L, et al. Roles and engagement levels of caregivers during dyadic Tai Chi interventions: a scoping review. Complement Ther Med 2026 Jun;98:103350 [FREE Full text] [doi: 10.1016/j.ctim.2026.103350] [Medline: 41839354]

XSL•FO
RenderX

51. Hopewell S, Chan AW, Collins GS, Hróbjartsson A, Moher D, Schulz KF, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. BMJ 2025 Apr 14;389:e081123 [FREE Full text] [doi: 10.1136/bmj-2024-081123] [Medline: 40228833]
52. Szabo SM, Whitlatch CJ, Orsulic-Jeras S, Johnson JD. Recruitment challenges and strategies: lessons learned from an early-stage dyadic intervention (innovative practice). Dementia (London) 2018 Jul;17(5):621-626. [doi: 10.1177/1471301216659608] [Medline: 27407077]
53. Gu Q, Wu SJ, Zheng Y, Zhang Y, Liu C, Hou JC, et al. Tai Chi exercise for patients with chronic heart failure: a meta-analysis of randomized controlled trials. Am J Phys Med Rehabil 2017 Oct;96(10):706-716. [doi: 10.1097/PHM.0000000000000723] [Medline: 28234634]
54. Jiao Q, Meng C, He H, Li S, Xu F, Cui W, et al. Safety and effects of a home-based Tai Chi exercise rehabilitation program in patients with chronic heart failure: study protocol for a randomized controlled trial. Front Cardiovasc Med 2023 Nov 29;10:1237539 [FREE Full text] [doi: 10.3389/fcvm.2023.1237539] [Medline: 38094121]
55. Fruhauf C, Schmid A, Prabhu N, Swink L, Portz J, Leach H, et al. Challenges of recruiting caregiver and care receiver dyads for a randomized clinical trial. Innov Aging 2020 Dec 16;4(Suppl 1):356-357. [doi: 10.1093/geroni/igaa057.1147]
56. Cheung DS, Tang SK, Ho KH, Jones C, Tse MM, Kwan RY, et al. Strategies to engage people with dementia and their informal caregivers in dyadic intervention: a scoping review. Geriatr Nurs 2021;42(2):412-420. [doi: 10.1016/j.gerinurse.2021.02.002] [Medline: 33639545]
57. Yang Z, Jia H, Zhang F, Huang H, Hao X, Wang A. A behavioural driving model of adherence to home-based cardiac rehabilitation exercise among patients with chronic heart failure: a mixed-methods study. J Clin Nurs 2024 Feb;33(2):531-542. [doi: 10.1111/jocn.16901] [Medline: 37881110]
58. Yeh GY, McCarthy EP, Wayne PM, Stevenson LW, Wood MJ, Forman D, et al. Tai Chi exercise in patients with chronic heart failure: a randomized clinical trial. Arch Intern Med 2011 Apr 25;171(8):750-757 [FREE Full text] [doi: 10.1001/archinternmed.2011.150] [Medline: 21518942]
59. Nagatomi Y, Ide T, Higuchi T, Nezu T, Fujino T, Tohyama T, et al. Home-based cardiac rehabilitation using information and communication technology for heart failure patients with frailty. ESC Heart Fail 2022 Aug;9(4):2407-2418 [FREE Full text] [doi: 10.1002/ehf2.13934] [Medline: 35534907]
60. Liang IJ, Francombe-Webb J, McGuigan PM, Perkin OJ, Thompson D, Western MJ. The acceptability of homebased exercise snacking and Tai-Chi snacking amongst high and low function UK and Taiwanese older adults. Front Aging 2023 Aug 1;4:1180939 [FREE Full text] [doi: 10.3389/fragi.2023.1180939] [Medline: 37593678]
61. Wang C, Schmid CH, Fielding RA, Harvey WF, Reid KF, Price LL, et al. Effect of Tai Chi versus aerobic exercise for fibromyalgia: comparative effectiveness randomized controlled trial. BMJ 2018 Mar 21;360:k851 [FREE Full text] [doi: 10.1136/bmj.k851] [Medline: 29563100]
62. Takemura N, Cheung DS, Fong DY, Lin CC. Impact of adherence to exercise interventions on effectiveness in a randomized controlled trial in improving sleep in advanced lung cancer patients: a comparison between aerobic exercise and Tai Chi. Palliat Support Care 2023 Jul 28:1-9. [doi: 10.1017/S1478951523000962] [Medline: 37503844]
63. van Beurden SB, Cross R, McDonagh ST, Clark L, Thomas C, Greaves CJ, et al. Digital adaptation of the clinically effective REACH-HF home-based cardiac rehabilitation programme for people living with heart failure (D:REACH-HF). Digit Health 2025 Dec 18;11:20552076251406545 [FREE Full text] [doi: 10.1177/20552076251406545] [Medline: 41425279]
64. Sadeghian F, Zolaktaf V, Shigematsu R. A comparison between effects of square-stepping exercise and Tai Chi Chuan on functional fitness and fear of falling in older women. Aging Clin Exp Res 2023 Apr;35(4):827-834. [doi: 10.1007/s40520-023-02367-1] [Medline: 36853504]
65. Liu L, Yang Q, Li T, Xie H, Zeng B, Zha L, et al. Prevalence and influencing factors of kinesiophobia in patients with heart disease: a meta-analysis and systematic review. Sci Rep 2024 Aug 15;14(1):18956 [FREE Full text] [doi: 10.1038/s41598-024-69929-9] [Medline: 39147837]
66. Lüke T, Brandenburg VM, Partetzke TM, Voss F, Spaderna H. Accelerometer-measured daily physical activity in adults with chronic heart failure and associations with fear of physical activity and coping dispositions. J Cardiovasc Nurs 2026;41(3):165-173. [doi: 10.1097/JCN.0000000000001242] [Medline: 40719556]
67. Piamjariyakul U, Werkowitch M, Wick J, Russell C, Vacek JL, Smith CE. Caregiver coaching program effect: reducing heart failure patient rehospitalizations and improving caregiver outcomes among African Americans. Heart Lung 2015;44(6):466-473. [doi: 10.1016/j.hrtlng.2015.07.007] [Medline: 26307539]
68. Alaei A, Babaei S, Farzi S, Hadian Z. Effect of a supportive-educational program, based on COPE model, on quality of life and caregiver burden of family caregivers of heart failure patients: a randomized clinical trial study. BMC Nurs 2024 Jan 26;23(1):72 [FREE Full text] [doi: 10.1186/s12912-024-01709-2] [Medline: 38279135]
69. Pucciarelli G, Occhino G, Locatelli G, Baricchi M, Ausili D, Rebora P, et al. The effectiveness of a motivational interviewing intervention on mutuality between patients with heart failure and their caregivers: a secondary outcome analysis of the MOTIVATE-HF randomized controlled trial. J Cardiovasc Nurs 2024;39(2):107-117. [doi: 10.1097/JCN.0000000000000991] [Medline: 37074953]

XSL•FO
RenderX

Abbreviations

CHF: chronic heart failure
CONSORT: Consolidated Standards of Reporting Trials
HDTC: home-based, digitally assisted, dyadic tai chi
HF: heart failure
MCAR: missing completely at random
QoL: quality of life
RCT: randomized controlled trial

Edited by S Brini; submitted 22.Mar.2026; peer-reviewed by YJ Son, W Cai; comments to author 17.Sep.2026; revised version received 13.Oct.2026; accepted 14.Oct.2026; published 11.Nov.2026.

Please cite as:
Zhang Y, Ngai FW, Cao L, Lu X, Zhang W, Wei L, Xie YJ
Home-Based, Digitally Assisted Dyadic Tai Chi Intervention for Patients With Chronic Heart Failure and Their Caregivers: Development of an Intervention and Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e95689
URL: https://mhealth.jmir.org/2026/1/e95689
doi:10.2196/95689
PMID:

©Yan Zhang, Fei Wan Ngai, Lihua Cao, Xiaohong Lu, Wenzhong Zhang, Lili Wei, Yao Jie Xie. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 11.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

mHealth–Supported Perioperative Care for Family Caregivers of Children Undergoing Tonsillectomy and/or Adenoidectomy (TONAPP): Randomized Controlled Trial

Raffaella Dobrina[1], MES, MSN, PhD; Chiara De Vita[2], PhD; Laura Brunelli[3], MD, PhD; Giulia Galvani[1], MSc; Margherita Dal Cin[4], MD; Manuela Giangreco[1], MSTAT, PhD; Milena Ciampechini[1], MSTAT; Silvana Schreiber[1], MSN; Giada Ferrari[1], MSN; Paola Di Rocco[1], MSN; Sara Zaccariotto[1], BSN; Maria Lucrezia Saija[1], BSN; Ilaria del Giorno[1], BSN; Sara Zanchiello[2,5], MEng; Anja Starec[2], MSc; Andrea Cassone[1], MSN

[1]Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", via dell'Istria 65/1, Trieste, Friuli Venezia Giulia, Italy

[2]Area Science Park, Trieste, Friuli Venezia Giulia, Italy

[3]Department of Medical, Surgical and Health Science, University of Udine, Udine, Italy

[4]Department of Health Prevention, Azienda Sanitaria Universitaria Giuliano isontina, Trieste, Friuli Venezia Giulia, Italy

[5]Central Directorate for Health, Social Policies and Disability, Trieste, Autonomous Region of Friuli Venezia Giulia, Italy

Corresponding Author:
Giulia Galvani, MSc
Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", via dell'Istria 65/1, Trieste, Friuli Venezia Giulia, Italy

Abstract

Background: Pediatric ear, nose, and throat (ENT) surgery is common, but generates perioperative anxiety for caregivers and distress in children. Limited time for perioperative education and reliance on unverified online information can reduce family preparedness and increase stress. Few studies have evaluated co-designed mobile health (mHealth) apps to support and engage families in the perioperative ENT journey.

Objective: This study aimed to compare caregiver anxiety between an mHealth app-supported care pathway and standard supportive and educational care alone in the perioperative ENT context. Secondary objectives explored between-group differences in caregiver anxiety at follow-up, family preparation, child distress, and social-impact indicators.

Methods: A 2-arm, parallel-group, open-label randomized controlled trial (RCT) enrolled caregivers of children undergoing ENT surgery (tonsillectomy, adenoidectomy, tympanostomy tube insertion). The intervention was an mHealth app co-designed through a user-centered participatory approach and developed following Schnall and colleagues' Information Systems Research Framework, with content based on caregivers' informational needs. RCT participants were recruited at the hospital during their presurgery visit, when a health care provider introduced the study and provided instructions on how to use the app. No additional human support was scheduled thereafter. A sample size of 180 participants (90 per group) was estimated to detect the expected between-group difference in caregiver anxiety. Participants were randomly assigned in a 1:1 ratio to app use or standard care alone. The primary outcome was the between-group difference in caregiver state anxiety (State-Trait Anxiety Inventory [STAI-Y]). Secondary outcomes included between-group differences in child distress (modified version of the Yale Preoperative Anxiety Scale [mYPAS]), child preparation for surgery, family preparation for hospital admission and surgery, and social impact indicators. Outcomes were assessed online through questionnaires, which included both self-reported measures and evaluations completed by a nurse on the day of surgery. App engagement metrics were also collected. Reporting followed the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines.

Results: The study enrolled 227 caregivers, with 111 allocated to the control group (CG) and 116 to the experimental group (EG), achieving the target sample size. No statistically significant differences were observed between the groups for the primary or secondary outcomes (all $P>.05$). In the EG, 75% (n=87) of the participants accessed at least 1 item of in-app content. Higher baseline anxiety was linked to lower app use ($\rho=-0.22$, 95% CI –0.39 to –0.04; $P=.02$), while greater use was linked to lower child distress ($\rho=-0.23$, 95% CI –0.40 to –0.04; $P=.02$).

Conclusions: Although the hypotheses were not confirmed, these findings provide valuable insights for future perioperative mHealth research. The lack of effectiveness may reflect limited exposure to the intervention, outcome selection and timing, and contextual factors such as caregivers' independent information-seeking. These findings support a greater focus on implementation processes and on identifying the caregivers most likely to benefit from mHealth-supported education.

Trial Registration: Clinicaltrials.gov NCT05460689; https://clinicaltrials.gov/study/NCT05460689

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-023-07376-z

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2026;14:e93989) doi:10.2196/93989

KEYWORDS

family-centered care; mHealth; app; pediatric surgery; caregiver education; caregiver information; anxiety

Introduction

Background

Pediatric ear, nose, and throat (ENT) surgery has undergone substantial advances over recent decades, with the introduction of innovative surgical techniques and standardized clinical pathways. In this context, elective procedures, such as tonsillectomy and/or adenoidectomy, with or without tympanostomy tube insertion, remain among the most frequently performed surgical interventions in children [1-5].

These procedures are typically delivered in high-pressure clinical settings characterized by short hospital stays and limited time for structured education. For children and their caregivers, the perioperative pathway is often unfamiliar and information-dense, leaving little opportunity to adequately address the different phases of the process, potential intraoperative risks, and common postoperative complications [6]. Postoperative sequelae may include pain, bleeding, nausea, vomiting, infection, swallowing difficulties, and prolonged discomfort after discharge. These complications can reduce oral intake, lead to dehydration, disrupt sleep, and, in more severe cases, result in hospital readmission and repeated school absences with potential academic consequences, thus placing considerable emotional and organizational demands and burden on families [5-8]. In addition, emotional factors play a key role in shaping perioperative outcomes in children, influencing both preoperative preparation and postoperative management and recovery [9]. In particular, perioperative anxiety has been consistently associated with adverse postoperative outcomes. Evidence shows that children who experience higher levels of anxiety before surgery are more likely to report increased postoperative pain and face a higher risk of complications [10,11]. Elevated preoperative anxiety in young children is also associated with increased sleep disturbances and higher levels of postoperative anxiety during recovery [10]. From a behavioral perspective, children may also exhibit problematic behavioral changes and temporary limitations in daily activities in the immediate postoperative period following ENT surgery [12].

Consistent with the family-centered approach, which considers family caregivers as key partners throughout the perioperative pathway, parental emotional responses represent another critical component of the pediatric perioperative experience. In fact, parental anxiety has been consistently identified as a significant predictor of children's anxiety prior to surgery [13] and appears to play a decisive role in shaping postoperative outcomes. Several studies have shown that higher levels of preoperative parental anxiety are associated with increased postoperative pain in children, a longer duration of analgesic use, and a higher number of analgesic administrations following tonsillectomy with or without adenoidectomy [4]. Similarly, children undergoing adenotonsillectomy whose parents report elevated anxiety levels have been found to experience higher rates of postoperative pain and complications [5]. Therefore, addressing parental anxiety before and during hospitalization is crucial not only to improve children's postoperative recovery but also to reduce emotional distress in both children and caregivers throughout the perioperative process. Interventions aimed at supporting parents' coping strategies and self-efficacy, including the provision of clear, timely, and tailored information, have been recognized as particularly relevant in pediatric surgical care [14-17].

Considering the complexity of the ENT perioperative process and its multiple clinical, emotional, and organizational implications, the provision of effective preoperative and postoperative education for both children and their family caregivers represents an essential component of care. In this context, parents' health literacy (HL), defined as "the knowledge, motivation, and competencies to access, understand, appraise, and apply health information in order to take judgements and take decisions in everyday life concerning health care, disease prevention and health promotion" [18], plays a pivotal role in how families navigate the perioperative surgical pathway. Evidence indicates that a substantial proportion of parents and caregivers of pediatric surgical patients have limited HL, which may negatively affect their understanding of surgical conditions, perceived risks and benefits, adherence to preoperative and postoperative instructions, and the quality of informed consent, while also contributing to higher levels of anxiety during surgical consultations [19-21]. These difficulties may be exacerbated by time constraints during consultations and the use of medical terminology that caregivers may find difficult to understand or feel uncomfortable questioning [22].

Moreover, the literature has revealed that parents tend to self-educate by relying on available online health care information, quickly seeking answers to their questions directly and anonymously from home [23,24]. These practices expose them to the risk of accessing information that, coming from very heterogeneous sources, is unreliable and not scientifically validated [25].

Within this framework, addressing these educational and communicative challenges requires structured perioperative preparation strategies that support both children and families throughout the ENT surgical pathway. Evidence suggests that nonpharmacological and digital interventions, including educational and immersive approaches aimed at familiarizing children with the hospital environment, equipment, and planned procedures, can reduce anxiety and emotional distress in pediatric surgical patients, particularly when integrated into perioperative preparation [15-17].

Building on these findings, perioperative digital interventions have been increasingly implemented across pediatric surgery. These interventions range from apps to text-messaging systems that support education, symptom monitoring, and reminders for medications or follow-up appointments. Overall, the available evidence suggests potential benefits in outcomes such as

XSL•FO
RenderX

follow-up adherence, postoperative monitoring, and user satisfaction, although study quality is variable and several trials show limitations (eg, inadequate controls, lack of blinding, and unequal attention to the experimental group [EG] and control group [CG]) [26].

Evidence syntheses also indicate that digital and distraction-based strategies, particularly virtual reality and other interactive formats, can reduce children's preoperative anxiety and may improve immediate perioperative outcomes, supporting their broader adoption in practice [15-17,27]. However, much of the literature focuses on single time windows (eg, preoperative anxiety reduction, postoperative follow-up or monitoring) rather than supporting families across the full perioperative pathway in a coherent way [15,26]. Recent evidence indicates that perioperative programs integrating multiple interventions across different time points, rather than focusing on a single moment of care, may achieve greater effectiveness in reducing parental and child distress [28].

Moreover, parental anxiety has been recognized as a relevant and modifiable target, with reviews describing multiple approaches (eg, education, audiovisual tools, preparation programs, and parental presence strategies), but with heterogeneity in content and delivery and limited clarity on which components work best for which families [28]. Furthermore, prior randomized controlled trials (RCTs) and meta-analyses include heterogeneous surgical populations, limiting conclusions specific to tonsillectomy and/or adenoidectomy with or without tympanostomy tube insertion [15].

Importantly, parents report valuing information provided directly by the hospital, suggesting that institution-endorsed digital tools may be particularly acceptable and useful for addressing practical issues such as pain management and fasting discussions at home [29]. Within this framework, there is a need for a condition-specific, hospital-based mobile health (mHealth) intervention designed to support family caregivers across milestones of the pediatric ENT perioperative process, using a user-centered participatory approach, with outcomes assessed through a rigorous trial design [30].

The Study: Rationale and Objectives

Considering the gaps in the literature, the aim of the current study was to evaluate the effectiveness of an mHealth app in supporting family caregivers of pediatric patients undergoing tonsillectomy and/or adenoidectomy with or without tympanostomy tube insertion, compared with supportive and educational methods related to standard care, in an RCT. Building on existing evidence, the present study focuses on the implementation of an mHealth system tailored to a specific pediatric surgical context and a clearly defined group of end users, considering the multiple phases of the ENT perioperative process. The mHealth app used in the intervention was a custom-designed, disease-specific, accessible tool, capable of providing validated scientific information, complying with current safety and privacy regulations, and designed to accompany children and their family caregivers throughout the perioperative process [31].

In line with the general aim, we intended to pursue 2 objectives. The primary objective was to explore the effect of using the mHealth app on the level of anxiety experienced by caregivers compared with standard care. The secondary objectives of the RCT were to evaluate the impact of using the mHealth app, compared with standard care, on 3 educational-emotional-organizational aspects, including (1) child preparation for hospitalization and surgery, (2) child distress before surgery, and (3) social impact such as postdischarge management. The study is based on the hypothesis that the possibility of using the mHealth app could help both reduce caregivers' state anxiety (primary outcome) and improve child preparation for hospitalization and surgery, child distress, and social impact related to postdischarge management at home (secondary outcomes).

Methods

Study Design and Setting

A full description of the design and methods used is contained in the study protocol [32]. The study used a 2-arm, parallel, open-label RCT involving family caregivers of pediatric patients aged between 2 and 10 years, who were scheduled for tonsillectomy and/or adenoidectomy with or without insertion of a tympanostomy tube. The study was conducted within a superiority framework, aiming to determine whether the EG yields better outcomes compared to the CG.

Blinding of participants, nurses, and researchers involved in the study was not feasible due to the nature of the intervention. Caregivers allocated to the EG were required to download and actively use the mHealth app and could seek support from nurses throughout the perioperative pathway, whereas those in the CG received standard care alone. In addition, nurses and researchers were involved in participant enrollment and data collection activities within the clinical setting, and app use could become apparent during routine clinical interactions, making concealment of group allocation impracticable.

The monocentric study was conducted at a 136-bed maternal and child health hospital (MCHH) in Northern Italy, where approximately 450 tonsillectomies or adeno-tonsillectomies are performed each year. The participant enrollment process, which took place in the consulting rooms of the hospital's pediatric ENT surgery department, began in December 2022 and concluded in December 2024. The random allocation sequence was generated independently by the hospital's expert statistician from the Clinical Epidemiology and Public Health Research Unit using computer-generated randomization with a block size of 4. Allocation concealment was ensured through sequentially numbered, opaque, sealed envelopes. Personnel involved in participant enrollment did not have access to the allocation sequence.

Nursing staff responsible for enrollment were in charge of providing eligible family caregivers with information about the study and obtaining signed informed consent from those wishing to participate in the RCT. Participation was voluntary, and caregivers could withdraw from the study at any time. The nurses, previously adequately trained in the mHealth app

XSL•FO
RenderX

content, features, and functionalities, were also responsible for instructing the EG caregivers both on how to download the app onto their smartphones or tablets and on how to use its contents and functionalities. During this training session, the nurses also showed the EG participants a 3-minute video presentation of the app on a tablet.

This trial is reported in accordance with the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (Checklist 1) [33].

Participants

Eligible participants were caregivers of children aged 2 to 10 years scheduled for tonsillectomy and/or adenoidectomy, with or without tympanostomy tube insertion. Caregivers were required to be able to communicate in spoken and written Italian and to have access to a smartphone and an internet connection. Exclusion criteria included cognitive or visual impairment in caregivers, cognitive impairment and/or chronic pain conditions in the child, previous surgery within the last month, and lack of previous experience with smartphone apps among caregivers. A formal assessment of reading comprehension was not performed during recruitment. However, eligibility was verified through direct interaction with caregivers, including confirmation of their ability to communicate in spoken and written Italian. HL was assessed as a study variable rather than used as an eligibility criterion, as the aim was to evaluate the intervention in a caregiver population representative of routine clinical practice.

After enrollment, participants were recruited in person at the hospital, on the day of their pre-surgery visit. Participants were not required to pay to access the app. Participants were randomly assigned in a 1:1 ratio by opening a sequentially numbered opaque and sealed envelope to either the EG, which received access to the mHealth app in addition to standard care, or the CG, which received standard care only. Standard care comprised oral information and education on the ENT perioperative pathway, supplemented by printed brochures provided by physicians and nurses during preoperative visits and throughout the hospitalization period.

Intervention Description

The intervention tested in this study was a purely app-based intervention. It allowed participants in the EG to access and use an mHealth app on their smartphones to acquire information and education about the ENT perioperative process related to 2 specific surgical procedures, namely tonsillectomy and/or adenoidectomy with or without insertion of a tympanostomy, in addition to receiving information and education on standard perioperative care provided by the ward's health care professionals. The EG was allowed to access the mHealth app from the day of the presurgery visit or preadmission consultation—when the enrollment took place—until the day of the scheduled follow-up visit, about 7 days after surgery. There was no mandatory minimum number of content views.

The contents, features, and functionalities of the mHealth app have been designed and developed by adopting a user-centered participatory design approach, in line with the Information System Research (ISR) framework [34], through the exploration of the informational or educational needs and preferences of both family caregivers (ie, primary end-users) and health care providers (ie, secondary end users) toward the mHealth app itself. A detailed description of the participatory design approach has already been published [30]. This co-design and codevelopment process involved the MCHH, which hosted the RCT and a public research institution, both situated in northern Italy.

Regarding content, the mHealth app contained 116 PDF documents covering the following six different thematic areas corresponding to 6 key informational or educational moments in the ENT perioperative process identified by primary and secondary end users who were directly involved in the participatory development process of the mHealth app: (1) the first inpatient ENT surgical consultation; (2) the phone call from the ENT surgical planning office to the family caregiver to communicate the date of the planned surgery and a list of necessary documents to bring on the day of preadmission consultations; (3) the day of preadmission consultations when caregivers are responsible for delivering the documents necessary for hospital admission and nursing, during which the anesthesiological and surgical consultations and examinations are also conducted; (4) the day of the family's admission to the hospital and where the surgery is performed; (5) discharge from the hospital and preparation for postoperative care at home, generally on the same day as the surgery (in the absence of complications); and (6) the follow-up visit in the hospital which takes place approximately 7 days after surgery. The app's contents were validated by the research team and the MCHH Scientific Director. Screenshots of the app have been published elsewhere [30].

The app's content was delivered in accordance with a timing principle, ensuring that information relevant to the specific stage in the ENT perioperative process was presented to the caregiver at the appropriate moment. In particular, depending on the child's stage in the perioperative process (eg, the day of preadmission consultation, the day of hospital admission and surgery, or postoperative recovery at home), the relevant content was suggested to the family caregiver through pop-up notifications as newly available content to consult within the app.

All the proposed content included validated medical information and reliable scientific references. All documents were provided in Italian and in text format. Depending on the specific topic covered, documents could include images, hyperlinks, graphical representations, and/or links to in-depth videos designed to facilitate caregivers' understanding. To access specific information content, caregivers had to open the corresponding PDF document in the app. Although the educational content was designed to be conveyed to the study participants progressively, reflecting the evolution of the ENT perioperative process, family caregivers still had access to all the documents throughout the period between enrollment and the follow-up visit, since they were available to end users at any time in a dedicated section of the mHealth app called Library in a consumable format and in line with the "just in time" principle [35].

XSL•FO
RenderX

As anticipated in the *Study Design and Setting* section, both participants in the EG and CG benefited from the supportive educational methods of standard care provided by health care personnel (ie, physicians and nurses), both orally and through the delivery of printed brochures. This education on the ENT perioperative process was comparable to that conveyed to the EG via the mHealth app, including content on strategies for preparing children for the experience of hospitalization, the specific surgical procedure selected, the type of anesthesia chosen, or managing postoperative pain in children at home after discharge. During the RCT, neither the EG nor the CG was prevented from seeking and using further information from books or other print materials, consulting other health care professionals, or web or social media sources. This approach was intentionally adopted to reflect routine clinical practice, where caregivers are exposed to and may actively seek information from multiple sources. The intervention was therefore evaluated as an adjunct to usual information-seeking behaviors rather than as a replacement for them.

Measures Used to Evaluate the Outcomes

The measures were taken at 4 different times, including T0 (ie, the enrollment), T1 (ie, the day of preadmission consultations), T2 (ie, the day when the surgery was performed), and T3 (ie, the follow-up visit). The timing, administration, and scoring methods for each measure are indicated below.

Sociodemographic Questionnaire

A sociodemographic questionnaire was self-assessed by participants at T0 to collect information on the family caregiver's age, gender, education level, occupational status, type of relationship with the child (ie, mother, father, other guardian), possible employment as a health professional, and any previous experience in caring for others with health problems. The questionnaire also asked caregivers about the child's age and sex, the surgical procedure planned during the current hospitalization, and any previous surgical experiences the child had undergone.

HL

Caregivers' level of HL was assessed at T0 using the 16-item self-assessment version of the European Health Literacy Survey Questionnaire (HLS-EU-Q16) [36]. This tool provides 5 possible answers on a 4-point Likert scale ranging from 1 (very difficult) to 4 (very easy), with a fifth possible answer equivalent to "I don't know." The expected score range is 0 - 16 and, depending on the total score obtained, 3 levels of HL can be distinguished, that is inadequate (0 - 8), problematic (9–12), and sufficient HL (13–16).

Caregiver's Preparation for Hospital Admission and Surgery

At T1, administrative nurses assessed caregivers' preparation for hospital admission and surgery using a checklist to verify the completeness of the necessary documentation provided by parents when they arrived at the hospital. The number of missing documents (eg, identity card or health care card) was recorded.

Child's Preparation for Surgery

At T2, a nurse in the surgical department assessed whether all the instructions provided by the staff for standard child surgical preparation had been followed, including proper hygiene, fasting, and not wearing nail polish or jewelry.

Caregiver's Anxiety

Caregivers' anxiety was evaluated using the Italian version of the State-Trait Anxiety Inventory questionnaire [37], consisting of 2 self-report scales for assessing state anxiety and trait anxiety, respectively: (1) the STAI Form Y-1 (STAI-S), including 20 statements aimed at measuring how the family caregiver feels "now, at this moment" on a 4-point Likert scale ranging from "not at all" to "very much", and (2) the STAI Form Y-2 (STAI-T), comprising a further 20 statements aimed at assessing how the participant feels "generally" on a 4-point Likert scale ranging from "almost never" to "almost always." The total score ranged from 20 to 80 points, and lower scores corresponded to higher levels of trait and state anxiety. Family caregivers were administered both scales at T0 to establish baseline levels of state and trait anxiety. The STAI-S scale was then further administered to caregivers in the hospital's ENT surgical department at both T2, before the surgery was performed, and T3, at the follow-up visit, approximately 7 days after surgery.

Child's Distress

To assess the child's distress at T2, a modified version of the Yale Preoperative Anxiety Scale (m-YPAS) [38] was used. This is an observational tool consisting of 27 items that explore five specific areas: (1) child activity, (2) child emotional expressiveness, (3) child arousal state, (4) child vocalization, and (5) caregiver engagement. The total score ranged from 23.33 to 100 points, and higher scores corresponded to higher levels of distress. Since a validated Italian version of the scale was not available, the tool was administered only by English-speaking nurses, in line with Liguori et al [39].

Social Impact Indicators

This area was self-assessed by family caregivers at T3, by answering questions about aspects such as how often caregivers called the hospital for further information, experienced and managed complications at home after discharge (eg, child pain or bleeding), or accessed the pediatric emergency department of the hospital.

mHealth App Usage

Data relating to the use of the mHealth app by family caregivers in the EG (eg, number of app accesses, the number and type of documents consulted, time spent exploring the app) were collected.

mHealth App Satisfaction

A satisfaction survey regarding the app was administered at T3. Participants were asked (1) to rate the app's completeness, usefulness, and ease of understanding; (2) to indicate whether they had sought additional information online due to dissatisfaction with the app's content; (3) to report any technical

XSL•FO
RenderX

issues experienced while using the app; and (4) to provide an overall satisfaction rating.

The STAI Forms Y-1 and Y-2, the HLS-EU-Q16, and the m-YPAS are validated instruments. However, the questionnaires administered online have not been specifically validated for online use.

The questionnaires administered at T0 and T2 were completed online in the ENT surgical department via a tablet provided by nurses to the EG and CG family caregivers. The T1 assessment was completed instead in paper form. Furthermore, the T3 evaluation was conducted online, after sending caregivers an email with a link to complete the relevant questionnaires. The latter approach was chosen since not all caregivers bring their children for follow-up at the MCHH promoting the study, sometimes preferring to refer and consult their general pediatrician.

Before the study began, nurses in the ENT surgical department involved in enrollment and subsequent measurements were adequately trained on the use of the assessment tools through the provision of standardized instructions. The training was conducted by a nurse from the research team and took place in a meeting room of the department.

Sample Size

The sample size was estimated considering preoperative anxiety as the primary outcome. The sample size was estimated considering parental preoperative anxiety as the primary outcome, assessed using the Italian version of the State-Trait Anxiety Inventory (STAI-Y) [37]. The expected mean anxiety score, informed by data previously reported in an Italian pediatric surgical population [40], was 50 (SD 13) in the CG and 45 (SD 10) in the EG, corresponding to an effect size of 0.43. Based on these assumptions, a total sample of 180 participants was required, with 90 participants allocated to each group. The calculation was performed using G*Power for a 2-group Wilcoxon-Mann-Whitney test, assuming a significance level of 0.05 and a power of 80% (β=.20).

Data Analysis

Description of the sample was based on frequency and percentage for categorical variables and on median and IQR for continuous variables. The standard statistical analysis was for intention-to-treat, and the balancing of baseline characteristics of the sample between the 2 groups of intervention was tested. The chi-square test or the exact Fisher test, as appropriate, was used to assess the association between 2 categorical variables. The nonparametric Wilcoxon-Mann-Whitney test was applied to evaluate the difference in the distribution of a continuous variable between 2 modalities of a categorical variable. The nonparametric Wilcoxon signed-rank test was calculated to test the difference between 2 paired continuous variables. Correlation was evaluated with the Spearman coefficient; a *P* value <.05 was considered statistically significant. SAS (version software 9.4 (SAS Institute Inc.) was used to conduct all statistical analyses. All randomized participants were analyzed in the group to which they had originally been assigned, regardless of intervention adherence. Reasons for study withdrawal were documented and recorded. Missing data were not imputed, and analyses were conducted using available data only. Data from the EG were analyzed in detail using the Wilcoxon rank-sum test and Spearman's rank correlation coefficient.

Ethical Considerations

This study received approval from the local ethics committee (CEUR-2022-Sper-26). As the intervention consisted exclusively of providing information and educational content through a smartphone app, it was intended as a low-risk intervention, and no serious adverse events were expected. Written informed consent was obtained from all caregivers through an offline, paper-based procedure, and no compensation was provided. Participants indicated their choice by selecting either a box confirming their willingness to participate or a box declining consent. Participants were informed that they could withdraw from the study at any time without the need to provide a reason. Children over 6 years of age were also asked to assent to participate in the study, signing a simplified, age-appropriate document that explained the study and what their agreement entailed. All personal data were processed in accordance with applicable data-protection regulations, and anonymization procedures were implemented prior to analysis.

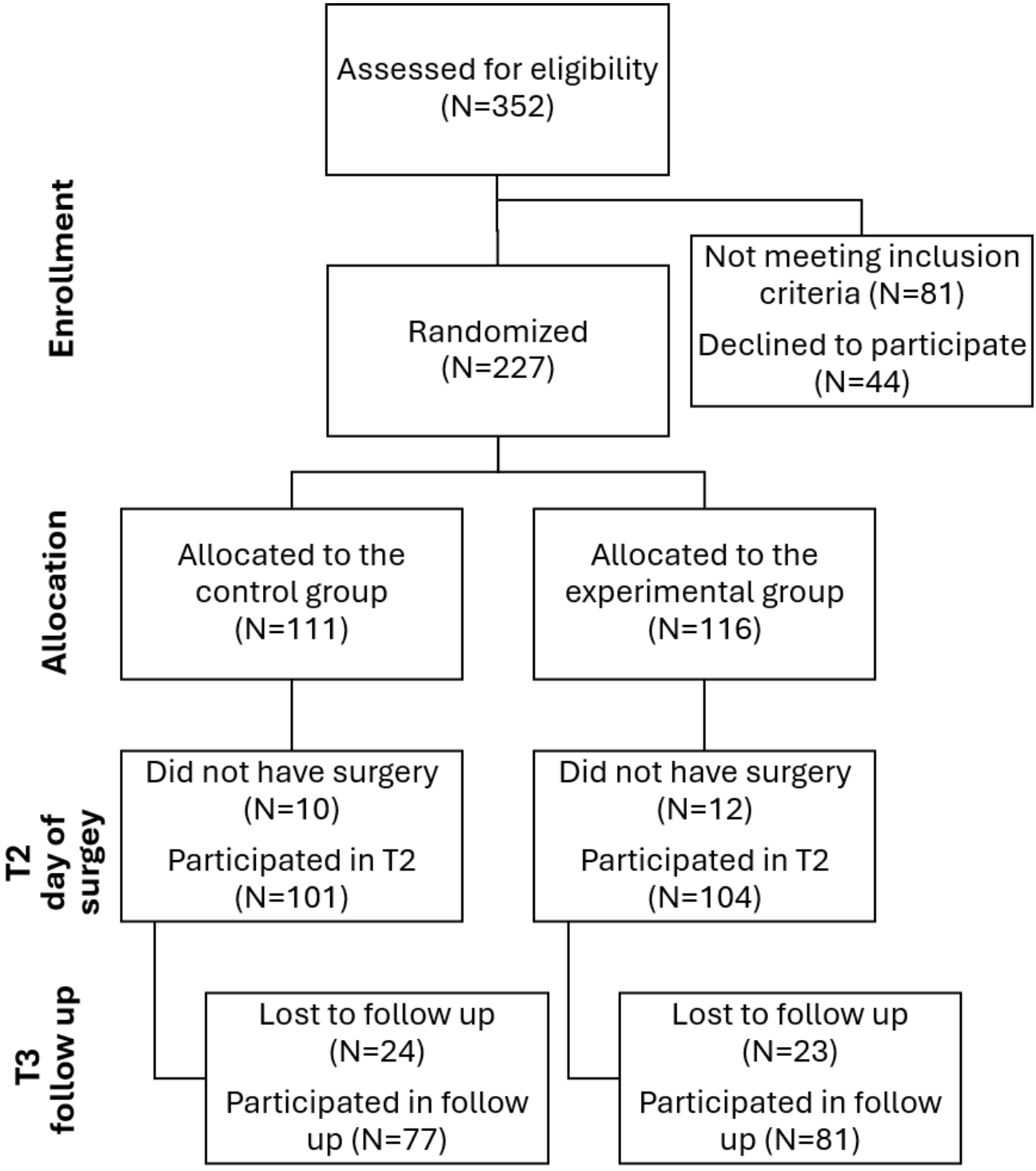
Results

Participant Characteristics

A total of 352 caregivers were screened between December 2022 and December 2024. Of these, 125 were excluded due to ineligibility or refusal to participate, resulting in the enrollment of 227 family caregivers of pediatric ENT patients, which met the target sample size. The participant flow diagram is detailed in Figure 1. As some questionnaires were not completed by all participants, only the available data were included in the analyses.

XSL•FO
RenderX

Figure 1. Participants' flow diagram.



Participants were randomly assigned to either the EG (n=116) or the CG (n=111). In both groups, women represented the clear majority (CG: n=90, 82.57%; EG: n=92, 79.31%). Their educational backgrounds spanned a wide range, from lower-secondary to postgraduate levels, with upper-secondary schooling being the most common (CG: n=61, 55.96%; EG: n=59, 50.86%). The majority were employed (CG: n=90, 82.57%; EG: n=94, 81.03%), often in clerical or office-based roles (CG: n=50, 45.87%; EG: n=57, 49.14%), though various occupational categories were represented. Caregivers were generally in mid-adulthood, with most falling between 30 and 49 years of age. The most common surgical procedures among children in both groups were adenoidectomy alone and tonsillectomy combined with adenoidectomy, which together accounted for 82.56% (n=90) of surgeries in the CG and 80.17% (n=93) in the EG.

At baseline, caregivers showed a distribution of HL levels ranging from inadequate to sufficient, with the majority classified in the sufficient category (CG: n=69, 62.73%; EG: n=72, 62.07%). The initial anxiety median scores in the CG were 37 (IQR 30 - 44) for the STAI-S and 36 (IQR 32 - 43) for the STAI-T scores, while the initial anxiety median scores

XSL•FO
RenderX

were 36 (IQR 31 - 44) for the STAI-S and 36 (IQR 31 - 42) for the STAI-T in the EG.

A complete set of participants' characteristics is reported in Table 1.

XSL•FO
RenderX

Table . Baseline characteristics, health literacy (HL) levels and state (STAI Form Y-1 [STAI-S]) and trait (STAI Form Y-2 [STAI-T]) anxiety scores of family caregivers and surgical procedures by study group.

Characteristics	Control group (n=111)[a]	Experimental group (n=116)
Caregiver sex, n (%)		
Female	90 (82.57)	92 (79.31)
Male	19 (17.43)	24 (20.69)
Intersex	0 (0)	0 (0)
Caregiver country of birth, n (%)		
Italy	92 (84.40)	97 (83.62)
Foreign country	17 (15.60)	19 (16.38)
Caregiver native language, n (%)		
Italian	95 (87.16)	102 (87.93)
Non-Italian language	14 (12.84)	14 (12.07)
Caregiver highest educational attainment, n (%)		
Primary education	0 (0.00)	1 (0.86)
Lower secondary education	13 (11.93)	13 (11.21)
Upper secondary education	61 (55.96)	59 (50.86)
Vocational education	1 (0.92)	0 (0.00)
University degree (bachelor's or master's degree)	24 (22.02)	33 (28.45)
Postgraduate education	9 (8.26)	10 (8.62)
Other	1 (0.92)	0 (0.00)
Employment status, n (%)		
Unemployed	19 (17.43)	22 (18.97)
Employed	90 (82.57)	94 (81.03)
Current occupation, n (%)		
Manager or entrepreneur	6 (5.50)	10 (8.62)
Clerical or office worker	50 (45.87)	57 (49.14)
Self-employed professional	8 (7.34)	11 (9.48)
Manual laborer	26 (23.85)	16 (13.79)
Unemployed	19 (17.43)	22 (18.97)
Caregiver age (y), n (%)		
<30	5 (4.63)	6 (5.17)
30 - 39	48 (44.44)	47 (40.52)
40 - 49	51 (47.22)	54 (46.55)
≥50	4 (3.70)	9 (7.76)
Surgical procedures, n (%)		
Adenoidectomy alone	45 (41.28)	36 (31.03)
Tonsillectomy alone	10 (9.17)	14 (12.07)
Adenoidectomy + tympanostomy tube placement	4 (3.67)	9 (7.76)
Tonsillectomy + adenoidectomy	45 (41.28)	57 (49.14)
Tonsillectomy + tympanostomy tube placement	1 (0.92)	0 (0.00)

XSL•FO
RenderX

Characteristics	Control group (n=111)[a]	Experimental group (n=116)
Combined adenoidectomy, tonsillectomy, and tympanostomy tube placement	4 (3.67)	0 (0.00)
Children gender, n (%)		
Female	44 (40)	47 (41)
Male	65 (60)	69 (59)
Children age, median (IQR)	5 (4-6)	6 (4-7)
T0 HL Scores, n (%)		
Inadequate	7 (6.36)	11 (9.48)
Problematic	34 (30.91)	33 (28.45)
Sufficient	69 (62.73)	72 (62.07)
Score T0 STAI-S, median (IQR)	37 (30 - 44)	36 (31 - 44)
Score STAI-T, median (IQR)	36 (32 - 43)	36 (31 - 42)

[a]Some questionnaires were not completed by all participants. In the control group, 3 sociodemographic questionnaires and 1 STAI Form Y-2 or State-Trait Anxiety Inventory questionnaire were not submitted. Furthermore, 1 control group participant did not provide a valid age value.

Caregiver Anxiety (Distribution of STAI Scores)

Total STAI-S scores changed significantly over time (P<.0001). The median score increased from 36 (IQR 31 - 42) at T0 to 38 (IQR 33 - 46) at T2, indicating a temporary rise in anxiety following baseline, since at T3, the median score decreased to 32 (IQR 28 - 39), reflecting a longer-term reduction in anxiety. All changes were statistically significant (P<.0001).

No difference in the distribution of caregiver state anxiety was found between T2 and T0 when comparing the EG and CG. The same result emerged between T3 and T0 and between T3 and T2. Table 2 presents the distribution of changes in caregivers' state anxiety levels (STAI-S) at the different measurement time points by study group.

Table . Distribution of the difference in the caregiver state anxiety levels (STAI Form Y-1 [STAI-S]) between T0, T2, and T3 by study group[a].

Difference in STAI-S between T0, T2, and T3	CG[b]		EG[c]		
	Participants, n	Median (IQR; min-max)	Participants, n	Median (IQR; min-max)	*P* value
STAI-S at T2—STAI-S at T0	98	2 (−2 to 7; −19 to 22)	96	3 (−1 to 7; −24 to 25)	.38
STAI-S at T3—STAI-S at T0	77	−2 (−11 to 1; −33 to 14)	81	−1 (−8 to 3; −25 to 24)	.08
STAI-S at T3—STAI-S at T2	75	−5 (−15 to 0; −43 to 12)	76	−6 (−12 to 0; −35 to 23)	.82

[a]Group differences were assessed using the nonparametric Wilcoxon Mann-Whitney test. Statistical significance was defined as P<.05.

[b]CG: control group.

[c]EG: experimental group.

From T0 to T2, both the CG and the EG showed a statistically significant increase in anxiety. In the CG, the median change was 2 (IQR −2 to 7), while in the EG it was 3 (IQR −1 to 7). From T0 to T3, the CG showed a median decrease of –2 (IQR −11 to 1), and the EG reported a median change of –1 (IQR −8 to 3). Nonetheless, the EG reported significantly more favorable scores than the CG for the statement "I feel comfortable" (item 10) of the STAI-S at T2 (P=.04; Table 3).

XSL•FO
RenderX

Table . Distribution of responses of item 10 of STAI Form Y-1 (STAI-S) at T2.

STAI-S at T2	CG[a], n (%)	EG[b], n (%)	*P* value
Item 10: "I feel confident"			.04
1	17 (17.35)	12 (12.50)	
2	60 (61.22)	57 (59.38)	
3	13 (13.27)	25 (26.04)	
4	8 (8.16)	2 (2.08)	

[a]CG: control group.

[b]EG: experimental group.

Caregiver's Preparation for Hospital Admission and Surgery

The data collected at T1 were lost due to the renovation of the hospital's administrative Welcome Point, which resulted in the displacement of the stored materials.

Child Distress (M-YPAS)

No statistical difference in child distress m-YPAS score at T2 emerged between the groups (*P*=.10), with equal median values (CG: 23.40; EG: 23.40) and overlapping IQRs (CG: 23.40 - 38.40; EG: 23.40 - 29.20).

Child's Preparation for Surgery

Both groups demonstrated consistently high compliance with preoperative instructions, such as arriving fasting (CG: n=97, 100%; EG: n=97, 99.0%), maintaining hygiene (100% in both groups, n=97 and n=98, respectively), and removing nail polish (CG: n=97, 100%; EG: n=93, 94.9%) or jewelry (CG: n=95, 97.9%; EG: n=97, 99.0%). Overall, no statistically significant differences were observed between the CG and EG regarding any aspect of the child's preparation for surgery (all *Ps*<.05).

Social Impact Indicators

Most social impact indicators did not show significant differences between the CG and EG. For the item "On a scale from 1 to 10, how adequate was the in-person time allocated by the medical and nursing staff at the MCHH throughout your child's surgical care pathway?", however, a significant difference emerged in the score distribution (*P*=.03). A larger proportion of parents in the CG assigned the maximum score of 10 (n=30, 39.0% vs n=17, 21% for the CG and EG, respectively), whereas the EG ratings were more concentrated around intermediate values (6 - 7: n=8, 10.4% vs n=21, 25.9% for the CG and EG, respectively).

mHealth App Usage and Satisfaction (EG Only)

The median duration of app use in the EG was 6 (IQR 5 - 7) days. Overall, 87 (75%) out of 116 caregivers in the EG viewed at least 1 content item, with a minimum of 13 contents opened and a maximum of 120. In particular, 35.3% (n=41) viewed fewer than 5 items. The 10 topics that attracted the most caregivers were, in decreasing order of number of views, the following: "Postoperative pain management (n=171)," "Postoperative diet at home (n=161)," "Postoperative home-care instructions (n=125)," "Items to bring to the hospital on the day of admission and surgery (n=117)," "How to prepare my child for outpatient visits, hospital admission, and surgery (n=102)," "The operating room (n=102)," "Can my child eat whatever they want? (n=102)," "From the ward to the operating room (n=100)," "On the ward after surgery (n=89)," and "Can my child bathe/wash? (n=87)."

A detailed analysis of the data from the EG, conducted using the Wilcoxon rank-sum test and Spearman's rank correlation coefficient, did not reveal a significant association between HL score and the number of viewed contexts (ρ=0.13, 95% CI –0.05 to 0.31; *P*=.16). A significant negative correlation emerged between the number of app contents opened and both the mYPAS child distress score (ρ=–0.23, 95% CI –0.40 to –0.04; *P*=.02) and the STAI-S caregiver state anxiety score at T0 (ρ=–0.22, 95% CI –0.39 to –0.04; *P*=.02), indicating that higher engagement with app contents was associated with lower child distress at T2 and with lower caregiver state anxiety levels at T0 in the EG, respectively.

The app was evaluated through a satisfaction questionnaire completed by 46.55% (n=56) of the users. Overall, respondents rated the app as "very" or "completely" complete (n=46, 85.18%), useful (n=48, 88.89%), and easy to understand (n=29, 53%). Most participants reported "never" having searched for additional information online due to dissatisfaction with the app's content (n=36, 66.67%) and indicated that they did not experience technical issues while using the app (n=47, 87.04%). Regarding overall satisfaction, the majority of users were "completely" (n=14, 25.93%) or "very" (n=26, 48.15%) satisfied with the app. Users' overall satisfaction was not significantly associated with baseline characteristics, HLS-EU-Q16, STAI, or m-YPAS scores.

Discussion

Principal Findings

This RCT evaluated whether a co-designed mHealth app could reduce caregiver anxiety during the pediatric ENT perioperative pathway. Contrary to our hypothesis, no significant differences in anxiety were found between caregivers who received the mHealth app in addition to standard care and those who received standard care alone. Likewise, no significant between-group differences were observed for child distress, child preparation for hospitalization and surgery, and postoperative management at home. This lack of detectable effects of using the mHealth app may appear disappointing, yet it aligns with a growing body of evidence showing limited effectiveness of app-based

XSL•FO
RenderX

interventions in pediatric perioperative care, particularly regarding caregiver anxiety [15]. Overall, while mHealth tools have shown promise in reducing distress among children [15,41], evidence of benefits for caregivers' anxiety remains weak [15,42]. In this light, the null findings reinforce the need for critical reflection on both methodological and contextual factors that may have influenced the effectiveness of the intervention. First, even if the baseline characteristics of the CG and EG overlapped, over one-third of the sample of family caregivers in the EG had a low HL level**,** which may have reduced their ability to understand, appraise, or use the health information provided through the app [18]. Moreover, although the language of the content was simplified, 1 in 6 of caregivers in the EG spoke a different native language. Therefore, the extent to which linguistic or literacy-related barriers influenced the EG's ability to use the app remains unknown [43].

A further factor that may have influenced the findings is that participants in the CG may have independently sought information online, consistent with the phenomenon known as "contamination by informal information-seeking," which is now well documented in the literature [44], thereby blurring the differences between the 2 groups.

Another contributing factor may relate to the timing of anxiety measurement, which occurred shortly before surgery, a moment that may reflect an expected physiological peak in parental stress rather than a modifiable psychological state [40]. This calls into question whether state anxiety is the most sensitive or appropriate outcome to capture the effects of digital tools that may operate more subtly, by improving knowledge, emotional preparedness, or perceived control. Notably, anxiety levels increased for both groups between baseline and the day of surgery, reflecting expected surgery anticipatory stress [44]. Following surgery, anxiety levels decreased in both groups, consistent with previous findings [41]. This pattern suggests that perioperative anxiety may evolve over time according to the natural course, regardless of the specific support tool used.

Finally, the time available to engage caregivers in using the app before surgery may have been insufficient. With a median usage period of only 6 days, caregivers may have had limited opportunities to explore and interact with the tool and related content and functionalities, given competing responsibilities and the demands of daily life.

Regarding app use, a notable share of caregivers either did not access the app or used it only minimally. Several factors may account for this limited use. One plausible factor is variation in individual preferences for how educational content was delivered. A systematic review of pediatric pretonsillectomy education programs found that interventions using smartphone apps, brochures, text messages, and videos all showed variable levels of engagement and efficacy, depending on the modality used and the target audience selected, suggesting that engagement may be influenced more by preferred learning styles than content quality alone [13]. Consistent with previous literature, another contributing factor may have been that some caregivers preferred direct human interaction and information delivered by health care professionals [45,46]. This interpretation is also supported by the findings on social impact indicators, as only 1 item showed a significant difference between the groups: parents in the CG rated the adequacy of in-person time provided by staff more highly than those in the EG. This pattern may suggest that reliance on face-to-face communication increased the perceived value of direct interaction, whereas access to the app may have shifted expectations toward digital support, influencing perceptions of in-person contact rather than actual staff time. This interpretation is consistent with evidence that technology-based preparatory tools can modify caregivers' expectations about interpersonal communication and shape how they appraise and value face-to-face interactions [47].

Another barrier to app use may have been that some family caregivers perceived it as difficult to navigate and engage with it because of the volume and density of the content provided, which may have led to information overload [47]. Although overall satisfaction with the app was high among approximately half of the participants who completed the app satisfaction evaluation survey, our results show that caregivers with higher baseline anxiety were less likely to engage with the app. This finding is consistent with Zhuo et al [48], who distinguish "monitors" (active seekers) from "blunters" or "avoiders." These different information-seeking behaviors suggest that those experiencing higher anxiety at preadmission, and who presumably would have needed the most support in managing it, were less inclined to use the app. Such behavioral heterogeneity, typical in nonstratified samples, may have diluted observable group differences.

Usage patterns also led researchers to hypothesize that usability barriers may have emerged from the document format: materials were indeed provided as PDFs and did not adapt to screen size, resulting in a text font size that was too small to read, necessitating frequent zooming and panning on mobile devices.

Despite the general null findings, a dose-response effect was observed in the EG: caregivers who accessed more than 5 pieces of content had children who appeared less stressed preoperatively, supporting the findings of Wang et al [44].

Moreover, a slightly higher level of preoperative comfort was found in the EG. This small advantage may reflect the reassuring effect of clearer procedural and orientation information provided through the app, consistent with evidence that preoperative education can reduce parental anxiety [28].

Beyond the specific findings of this study, it is important to acknowledge that null results can provide valuable contributions to the evidence base. Reporting studies that do not demonstrate significant effects helps prevent publication bias and offers a more balanced understanding of the effectiveness of digital health interventions [49,50]. Such results can also inform the refinement of app design and implementation strategies, as well as help recalibrate unrealistic expectations regarding the potential impact of digital health solutions in complex clinical contexts. More broadly, making these findings available supports research transparency, improves the accuracy of future systematic reviews and meta-analyses, and ultimately contributes to cumulative scientific progress [49,50].

XSL•FO
RenderX

Limitations

This study has several limitations. First, caregivers and nurses were not blinded to group allocation, which may have introduced expectation- and assessment-related biases, respectively. Furthermore, the RCT findings may have been affected by contamination bias resulting from caregivers' informal information-seeking from multiple additional sources [25], including health care professionals, printed materials, websites, and social media. Future studies could assess information-seeking behaviors more systematically (and stratify the sample accordingly) to explore their influence on intervention effectiveness [48].

Another limitation concerns the variability in engagement with the intervention. Many caregivers only accessed a limited portion of the app content and therefore did not receive the intended intervention "dose", which may have affected the ability to detect significant effects. Future studies should explore alternative and more engaging content delivery formats, such as videos, podcasts, and gamification features (eg, feedback, challenges, or rewards), which may enhance accessibility, usability, and user engagement with digital interventions [51,52].

Moreover, caregivers' reading fluency and comprehension competencies as well as document or digital literacy were not assessed in this study, although these skills affect the ability to navigate written materials, particularly when delivered in PDF files that are not optimized for mobile devices.

Furthermore, the timing of the measurement of anxiety, taken shortly before surgery, is another limitation, as it coincides with an expected peak of anticipatory stress. This may have reduced the sensitivity for detecting short-term changes attributable to app use.

Moreover, the app's navigability and information architecture, such as the length or text density of written materials, may not have fully supported ease of use for all users. These considerations are addressed in the Guide to Implementing the Health Literacy Universal Precautions Toolkit and its communication-design guidance, although this resource was not incorporated into the app's design and development process [53].

A further limitation was the loss of data collected at T1 regarding caregivers' preparation for hospital admission. However, this did not affect the study's primary or key secondary outcomes.

Finally, the single-site study design limits the generalizability of the findings, as participants may not be representative of broader linguistic, cultural, or sociodemographic populations.

Conclusions

The findings of this RCT suggest that adding an mHealth app to standard care did not significantly reduce caregiver anxiety. Beyond its clinical outcomes, this study provides a methodological contribution through the development and rigorous evaluation of a user-centered mHealth intervention for family caregivers of pediatric surgical patients. The trial also underscores key challenges associated with the implementation of digital health interventions in real-world clinical settings, including variability in user engagement, the timing and duration of intervention exposure, and the potential influence of information-seeking behaviors. These factors should be carefully considered when interpreting intervention effectiveness and when designing future trials.

Rather than questioning the value of mHealth interventions, these findings highlight that future research should shift its focus from developing and evaluating digital tools alone to understanding how they can be effectively implemented, integrated into routine care, and how engagement with digital interventions can be optimized across the perioperative pathway. This will require implementation strategies that promote meaningful and sustained engagement, while identifying which caregivers are most likely to benefit from mHealth-supported education. Optimizing caregiver support throughout the perioperative pathway remains an important objective, given its potential to positively influence the perioperative experience of both caregivers and children.

Acknowledgments

The authors would like to express their sincere gratitude to the nursing staff who contributed to the data collection process: Bartoli Araya, Cantoro Marco, Colautti Marzia, Culqui Betty Roxana, Cvetkovic Valentina, Damato Giulia, De Cesco Beatrice, Ferluga Rosita, Iacobelli Giulia, Ieracintano Nicole, La Macchia Renata, Mauri Tanja, Mele Francesco, Migliozzi Maria, Moretti Annamaria, Muzzachi Antonio, Pertot Federica, Rinaldi Lucia, Vatta Mascia, Zariali Metka, and Zecchin Alice.

The research team is grateful to the staff of CGM TELEMEDICINE S.R.L. (a CompuGroup Medical company) for developing the app that was tested in this study, Fablab S.R.L. (a CompuGroup Medical Italia company) for improving the legibility of the app's content, and Quality S.R.L. (a start-up spin-off of the University of Trieste, Italy) for developing the questions for the social impact survey.

The authors used ChatGPT to improve the readability and language of the manuscript. All outputs were reviewed and edited by the authors, who take full responsibility for the final content.

Funding

The development of the mHealth app experimented in this study was funded by "PROTOCOLLO DI INTESA TRA IL MINISTERO DELL'UNIVERSITA' E DELLA RICERCA E IL MINISTERO DELLO SVILUPPO ECONOMICO E LA REGIONE

XSL•FO
RenderX

AUTONOMA FRIULI VENEZIA GIULIA—Progetto complesso “SISTEMA ARGO” signed on March 1, 2018 and renewed on July 13, 2021 until December 31, 2023. Moreover, this work was supported by the Ministry of Health, through the contribution given to the Institute for Maternal and Child Health IRCCS Burlo Garofolo, Trieste – Italy. [RC 3/22].

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: RD, AC, LB, CDV, MDC, AS, S Zanchiello
Data curation: MG, MC, RD, GG
Formal analysis: MG, MC
Funding acquisition: AS, S Zanchiello, MDC, AC
Investigation: GG, S Zaccariotto, PDR, GF, MLS, IdG
Methodology: RD, AC, LB, CDV, MDC, AS, S Zanchiello, MG
Project administration AC, RD, S Zanchiello
Resources: PDR, MG, GF, AC
Software: MG, AS, S Zanchiello
Supervision: RD, AC, AS, S Zanchiello, PDR
Validation: RD, MG, MC, LB, CDV
Visualization: RD, GG, MG, MC, LB, CDV
Writing – original draft: RD, LB, CDV, MG, MC, AC
Writing – review and editing: RD, CDV, LB, GG, MDC, MG, MC, S Schreiber, GF, PDR, S Zaccariotto, MLS, IdG, S Zanchiello; AS, AC

Conflicts of Interest

None declared.

Checklist 1
CONSORT-EHEALTH (version 1.6) Checklist.
[PDF File, 10210 KB - mhealth_v14i1e93989_app1.pdf]

References

1. Mitchell RB, Archer SM, Ishman SL, et al. Clinical practice guideline: tonsillectomy in children (update)-executive summary. Otolaryngol Head Neck Surg 2019 Feb;160(2):187-205. [doi: 10.1177/0194599818807917] [Medline: 30921525]
2. Han CJ, Bergman M, Harley RJ, Harley EH. The pediatric indications for tonsillectomy and adenotonsillectomy: race/ethnicity, age, and gender. Laryngoscope Investig Otolaryngol 2023 Apr;8(2):577-583. [doi: 10.1002/lio2.1017] [Medline: 37090873]
3. Heward E, Rocke J, McNally G, et al. The post-operative tonsillectomy (POPT) study: a multi-centre prospective paediatric cohort study. Clin Otolaryngol 2024 Mar;49(2):176-184. [doi: 10.1111/coa.14110] [Medline: 37915294]
4. Alves CS, Dias J, Azevedo S, et al. Keep calm: does parental preoperative anxiety affect post-tonsillectomy pain scores in children? Eur Arch Otorhinolaryngol 2024 Aug;281(8):4385-4392. [doi: 10.1007/s00405-024-08683-0] [Medline: 38739185]
5. Doluoglu S, Gazeloglu AZ, Kocyigit Y, Vural Camalan B, Ozlugedik S. Effects of parental anxiety on the postoperative pain and complications of children undergoing adenotonsillectomy. Laryngoscope 2025 Sep;135(9):3422-3429. [doi: 10.1002/lary.32217] [Medline: 40323129]
6. Richards J, Lang M, Andresen E, et al. Impact of paediatric tonsillectomy perioperative management on pain, nausea and recovery: a prospective cohort study. J Paediatr Child Health 2020 Jan;56(1):114-122. [doi: 10.1111/jpc.14505] [Medline: 31144404]
7. Alm F, Lööf G, Blomberg K, Ericsson E. Establishment of resilience in a challenging recovery at home after pediatric tonsil surgery-Children’s and caregivers’ perspectives. Paediatr Neonatal Pain 2021 Jun;3(2):75-86. [doi: 10.1002/pne2.12051] [Medline: 35547595]
8. Dias JI, Alves CS, Castro SSE, Meireles L, Coutinho MB. Does parental stress improve after pediatric tonsillectomy? Int J Pediatr Otorhinolaryngol 2024 Jan;176:111830. [doi: 10.1016/j.ijporl.2023.111830] [Medline: 38141550]
9. Lööf G, Lönnqvist PA. Role of information and preparation for improvement of pediatric perioperative care. Paediatr Anaesth 2022 May;32(5):600-608. [doi: 10.1111/pan.14419] [Medline: 35167154]
10. Chow CHT, Schmidt LA, Buckley DN. The role of anxiety and related states in pediatric postsurgical pain. Can J Pain 2020 Dec 30;4(4):26-36. [doi: 10.1080/24740527.2020.1847600] [Medline: 33987517]
11. Mathias EG, Pai MS. Anxiety and pain in children undergoing surgery: a scoping review. J Perianesth Nurs 2022 Aug;37(4):545-550. [doi: 10.1016/j.jopan.2021.10.002] [Medline: 35279386]

XSL•FO
RenderX

12. Lao BK, Kain ZN, Khoury D, et al. A comprehensive examination of the immediate recovery of children following tonsillectomy and adenoidectomy. Int J Pediatr Otorhinolaryngol 2020 Aug;135:110106. [doi: 10.1016/j.ijporl.2020.110106] [Medline: 32422367]
13. Levin M, Seligman NL, Hardy H, Mohajeri S, Maclean JA. Pediatric pre-tonsillectomy education programs: a systematic review. Int J Pediatr Otorhinolaryngol 2019 Jul;122:6-11. [doi: 10.1016/j.ijporl.2019.03.024] [Medline: 30921630]
14. Rosenberg RE, Clark RA, Chibbaro P, et al. Factors predicting parent anxiety around infant and toddler postoperative and pain. Hosp Pediatr 2017 Jun;7(6):313-319. [doi: 10.1542/hpeds.2016-0166] [Medline: 28512138]
15. Rantala A, Pikkarainen M, Miettunen J, He HG, Pölkki T. The effectiveness of web-based mobile health interventions in paediatric outpatient surgery: a systematic review and meta-analysis of randomized controlled trials. J Adv Nurs 2020 Aug;76(8):1949-1960. [doi: 10.1111/jan.14381] [Medline: 32281673]
16. Chen H, Zhang J, Li S, Zhang H, Wei L. Non-pharmacological interventions for preoperative anxiety in children: a systematic review and network meta-analysis. J Clin Nurs 2025 Apr;34(4):1493-1507. [doi: 10.1111/jocn.17582] [Medline: 39763216]
17. Luo Z, Zhou R, Nong K, et al. Digital health interventions in pediatric perioperative care: a network meta-analysis. JAMA Pediatr 2025 Nov 1;179(11):1153-1161. [doi: 10.1001/jamapediatrics.2025.3099] [Medline: 40952714]
18. Sørensen K, Van den Broucke S, Fullam J, et al. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80. [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
19. Wasserzug O, Fishman G, Sternbach D, et al. Informed consent for tonsillectomy: do parents comprehend the information we provide? Int J Pediatr Otorhinolaryngol 2016 Sep;88:163-167. [doi: 10.1016/j.ijporl.2016.07.005] [Medline: 27497406]
20. Kampouroglou G, Velonaki VS, Pavlopoulou I, et al. Parental anxiety in pediatric surgery consultations: the role of health literacy and need for information. J Pediatr Surg 2020 Apr;55(4):590-596. [doi: 10.1016/j.jpedsurg.2019.07.016] [Medline: 31431293]
21. Kampouroglou G, Velonaki VS, Pavlopoulou ID, et al. Health literacy of parents/caregivers of paediatric surgical patients: a study on 1000 individuals. Afr J Paediatr Surg 2021;18(2):85-89. [doi: 10.4103/ajps.AJPS_42_20] [Medline: 33642404]
22. Walsh AM, Hamilton K, White KM, Hyde MK. Use of online health information to manage children's health care: a prospective study investigating parental decisions. BMC Health Serv Res 2015 Apr 2;15:131. [doi: 10.1186/s12913-015-0793-4] [Medline: 25889493]
23. Pehora C, Gajaria N, Stoute M, Fracassa S, Serebale-O'Sullivan R, Matava CT. Are parents getting it right? A survey of parents' internet use for children's health care information. Interact J Med Res 2015 Jun 22;4(2):e12. [doi: 10.2196/ijmr.3790] [Medline: 26099207]
24. Mohamed H, Kittle E, Nour N, et al. An integrative systematic review on interventions to improve layperson's ability to identify trustworthy digital health information. PLOS Digit Health 2024 Oct;3(10):e0000638. [doi: 10.1371/journal.pdig.0000638] [Medline: 39453891]
25. Risk communication, community engagement and infodemic management at the core of health security in humanitarian emergencies: operational tool designed from experiences in the WHO European region. : World Health Organization; 2026 URL: https://www.who.int/europe/publications/i/item/WHO-EURO-2026-13295-53069-82872 [accessed 2026-08-19]
26. Nguyen N, Leveille E, Guadagno E, Kalisya LM, Poenaru D. Use of mobile health technologies for postoperative care in paediatric surgery: a systematic review. J Telemed Telecare 2022 Jun;28(5):331-341. [doi: 10.1177/1357633X20934682] [Medline: 32605411]
27. Chen YJ, Wang CJ, Chen CW. Effects of virtual reality on preoperative anxiety in children: a systematic review and meta-analysis of randomised controlled trials. J Clin Nurs 2023 Jun;32(11-12):2494-2504. [doi: 10.1111/jocn.16394] [Medline: 35672942]
28. Santapuram P, Stone AL, Walden RL, Alexander L. Interventions for parental anxiety in preparation for pediatric surgery: a narrative review. Children (Basel) 2021 Nov 20;8(11):1069. [doi: 10.3390/children8111069] [Medline: 34828782]
29. Kennedy M, Howlin F. Preparation of children for elective surgery and hospitalisation: a parental perspective. J Child Health Care 2022 Dec;26(4):568-580. [doi: 10.1177/13674935211032804] [Medline: 34275370]
30. Dobrina R, Starec A, Brunelli L, et al. Applying the participatory slow design approach to a mhealth application for family caregivers in pediatric ear, nose, and throat surgery. Healthcare (Basel) 2024;12(4):442. [doi: 10.3390/healthcare12040442] [Medline: 38391818]
31. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mHealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817. [doi: 10.2196/13817] [Medline: 31293246]
32. Dobrina R, Cassone A, Dal Cin M, et al. Study protocol for a randomised controlled trial to determine the effectiveness of a mHealth application as a family supportive tool in paediatric otolaryngology perioperative process (TONAPP). Trials 2023 May 25;24(1):355. [doi: 10.1186/s13063-023-07376-z] [Medline: 37231477]
33. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
34. Schnall R, Rojas M, Bakken S, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251. [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]

XSL•FO
RenderX

35. Cook DJ, Moradkhani A, Douglas KSV, Prinsen SK, Fischer EN, Schroeder DR. Patient education self-management during surgical recovery: combining mobile (iPad) and a content management system. Telemed J E Health 2014 Apr;20(4):312-317. [doi: 10.1089/tmj.2013.0219] [Medline: 24443928]
36. Lorini C, Santomauro F, Grazzini M, et al. Health literacy in Italy: a cross-sectional study protocol to assess the health literacy level in a population-based sample, and to validate health literacy measures in the Italian language. BMJ Open 2017 Nov 14;7(11):e017812. [doi: 10.1136/bmjopen-2017-017812] [Medline: 29138204]
37. Pedrabissi L, Santinello M. Manual for the New Italian Version of the STAI - Form Y [Book in Italian]: Joint OS Special Organizations; 1989. URL: https://www.research.unipd.it/handle/11577/2523749 [accessed 2026-08-19]
38. Kain ZN, Mayes LC, Cicchetti DV, Bagnall AL, Finley JD, Hofstadter MB. The Yale Preoperative Anxiety Scale: how does it compare with a "gold standard"? Anesth Analg 1997 Oct;85(4):783-788. [doi: 10.1097/00000539-199710000-00012] [Medline: 9322455]
39. Liguori S, Stacchini M, Ciofi D, Olivini N, Bisogni S, Festini F. Effectiveness of an app for reducing preoperative anxiety in children: a randomized clinical trial. JAMA Pediatr 2016 Aug 1;170(8):e160533. [doi: 10.1001/jamapediatrics.2016.0533] [Medline: 27294708]
40. Pomicino L, Maccacari E, Buchini S. Levels of anxiety in parents in the 24 hr before and after their child's surgery: a descriptive study. J Clin Nurs 2018 Jan;27(1-2):278-287. [doi: 10.1111/jocn.13895] [Medline: 28544343]
41. Pan Y, Wan L, Xue Q, Ran G, Chen K, Shen X. Evaluating the efficacy of a smartphone application in reducing preoperative anxiety in children: a randomised controlled trial. Indian J Anaesth 2025 Mar;69(3):308-314. [doi: 10.4103/ija.ija_526_24] [Medline: 40161921]
42. Kerimaa H, Hakala M, Haapea M, et al. Effectiveness of a mobile app intervention for preparing preschool children and parents for day surgery: randomized controlled trial. J Med Internet Res 2023 Sep 29;25:e46989. [doi: 10.2196/46989] [Medline: 37773624]
43. Health literacy. World Health Organization. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/health-literacy [accessed 2025-08-04]
44. Wang J, Wu Z, Choi SW, et al. The dosing of mobile-based just-in-time adaptive self-management prompts for caregivers: preliminary findings from a pilot microrandomized study. JMIR Form Res 2023 Sep 14;7:e43099. [doi: 10.2196/43099] [Medline: 37707948]
45. Aranha PR, Dsouza SN. Preoperative information needs of parents: a descriptive survey. J Res Nurs 2019 Aug;24(5):305-314. [doi: 10.1177/1744987118821708] [Medline: 34394540]
46. Yahya Al-Sagarat A, Al-Oran HM, Obeidat H, Hamlan AM, Moxham L. Preparing the family and children for surgery. Crit Care Nurs Q 2017;40(2):99-107. [doi: 10.1097/CNQ.0000000000000146] [Medline: 28240692]
47. Kim J, Chiesa N, Raazi M, Wright KD. A systematic review of technology-based preoperative preparation interventions for child and parent anxiety. Can J Anaesth 2019 Aug;66(8):966-986. [doi: 10.1007/s12630-019-01387-8] [Medline: 31098960]
48. Zhuo Q, Ma F, Cui C, et al. Effects of pre-operative education tailored to information-seeking styles on pre-operative anxiety and depression among patients undergoing percutaneous coronary intervention: a randomized controlled trial. Int J Nurs Sci 2023 Apr;10(2):174-181. [doi: 10.1016/j.ijnss.2023.03.015] [Medline: 37128491]
49. Matosin N, Frank E, Engel M, Lum JS, Newell KA. Negativity towards negative results: a discussion of the disconnect between scientific worth and scientific culture. Dis Model Mech 2014 Feb;7(2):171-173. [doi: 10.1242/dmm.015123] [Medline: 24713271]
50. Sharma H, Verma S. Is positive publication bias really a bias, or an intentionally created discrimination toward negative results? Saudi J Anaesth 2019;13(4):352-355. [doi: 10.4103/sja.SJA_124_19] [Medline: 31572081]
51. Kakhki SK, Aghebati N, Moonaghi HK. Exploring the impact, challenges, and integration of podcasts in patient education: a systematic review. BMC Med Educ 2025 May 12;25(1):690. [doi: 10.1186/s12909-025-07217-4] [Medline: 40355902]
52. Chen J, Ying Y, Pang M, et al. The application of gamification in children's oral health management: systematic review. J Med Internet Res 2025 Nov 4;27:e75541. [doi: 10.2196/75541] [Medline: 41187330]
53. AHRQ health literacy universal precautions toolkit. : Agency for Healthcare Research and Quality (AHRQ); 2025 URL: https://www.ahrq.gov/health-literacy/improve/precautions/index.html [accessed 2026-02-25]

Abbreviations

CG: control group
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
EG: experimental group
ENT: ear, nose, and throat
HL: health literacy
HLS-EU-Q16: 16-item European Health Literacy Survey Questionnaire
ISR: Information System Research

XSL•FO
RenderX

m-YPAS: modified version of the Yale Preoperative Anxiety Scale
MCHH: maternal and child health hospital
mHealth: mobile health
RCT: randomized controlled trial
STAI-S: STAI Form Y-1
STAI-T: STAI Form Y-2
STAI-Y: State-Trait Anxiety Inventory

Edited by JN Siebert; submitted 23.Feb.2026; peer-reviewed by DV Dantas, E Hamilton; revised version received 15.Jul.2026; accepted 17.Jul.2026; published 01.Sep.2026.

Please cite as:
Dobrina R, De Vita C, Brunelli L, Galvani G, Dal Cin M, Giangreco M, Ciampechini M, Schreiber S, Ferrari G, Di Rocco P, Zaccariotto S, Saija ML, del Giorno I, Zanchiello S, Starec A, Cassone A
mHealth–Supported Perioperative Care for Family Caregivers of Children Undergoing Tonsillectomy and/or Adenoidectomy (TONAPP): Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e93989
URL: https://mhealth.jmir.org/2026/1/e93989
doi:10.2196/93989

© Raffaella Dobrina, Chiara De Vita, Laura Brunelli, Giulia Galvani, Margherita Dal Cin, Manuela Giangreco, Milena Ciampechini, Silvana Schreiber, Giada Ferrari, Paola Di Rocco, Sara Zaccariotto, Maria Lucrezia Saija, Ilaria del Giorno, Sara Zanchiello, Anja Starec, Andrea Cassone. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 1.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Families Moving Forward Connect mHealth Intervention for Caregivers of Children With Fetal Alcohol Spectrum Disorders: Randomized Controlled Trial

Christie L M Petrenko[1,2], PhD; Madeline N Rockhold[1], MA; Julianne M Myers[1,2], PhD; Cristiano Tapparello[3], PhD; Carson Kautz-Turnbull[1], MA; Emily Speybroeck[1], BA; Zhi Li[1], PhD; Lynn L Cole[2], DNP, RN, PPCNP-BC; Heather Carmichael Olson[4,5], PhD

[1]Mt. Hope Family Center, University of Rochester, 187 Edinburgh St., Rochester, NY, United States

[2]Division of Developmental and Behavioral Pediatrics, University of Rochester, Rochester, NY, United States

[3]Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, United States

[4]Department of Psychiatry and Behavioral Sciences, Center on Child Health, Behavior and Development, School of Medicine, University of Washington, Seattle, WA, United States

[5]Seattle Children's Research Institute, Seattle, WA, United States

Corresponding Author:
Christie L M Petrenko, PhD
Mt. Hope Family Center, University of Rochester, 187 Edinburgh St., Rochester, NY, United States

Abstract

Background: Fetal alcohol spectrum disorders (FASD) affect 1.1% to 5% of the general population. Yet, most children with FASD and their families cannot access evidence-based interventions. Mobile health (mHealth) interventions have the potential to increase access to care on a broad scale. While numerous self-directed parenting apps exist, none have been tested for FASD. The FMF (Families Moving Forward) Connect app is a self-directed intervention derived from an empirically supported intervention for caregivers raising children with FASD. FMF Connect is the first self-directed parenting app for FASD, and also one of the first parenting apps to be systematically developed and tested.

Objective: This study aimed to test the efficacy of FMF Connect for caregivers raising children with FASD on targeted primary (child behavior, caregiver attributions, parenting efficacy and satisfaction, FASD knowledge, and family needs met) and secondary (child adaptive behavior, caregiver self-care, and app satisfaction) outcomes.

Methods: This study involved a 3-arm randomized controlled trial with equal allocation to groups (1) FMF Connect+coaching, (2) FMF Connect, or (3) waitlist control. Participants from the United States were recruited online through an open access website. Recruitment materials were distributed by the Collaborative Initiative on FASD, FASD listserves, and social media. In total, 129 caregivers of children (aged 3 - 12 y) with FASD or prenatal alcohol exposure (PAE) were enrolled. Online surveys were administered at baseline, 6 weeks, and 12 weeks. Data were analyzed with linear mixed modeling, linear regressions, and structural equation modeling using SPSS (version 29.0; IBM) and Mplus 8 (Muthén & Muthén).

Results: A total of 43 participants were randomized to each group. Caregivers were predominantly White adoptive mothers. Of the total, 64% (n=83) of participants were retained through the 12-week follow-up. Groups did not differ in terms of demographic characteristics, baseline levels of functioning, or attrition. Usage patterns were similar across groups, suggesting coaching did not increase engagement. Given a few differences, app intervention groups were combined for analyses. Relative to the waitlist group, caregivers in the FMF Connect group evidenced greater improvements in FASD knowledge, child behavior attributions, family needs met, and self-care after 12 weeks (P=.01-.048). After controlling for multiple comparisons, differences in FASD knowledge, self-care, and family needs met approached significance (P=.06-.07). Groups did not differ in parenting satisfaction, child behavior problems, or adaptive functioning. More app usage is related to greater changes in parenting efficacy. Caregiver behavior attributions at 6 weeks did not mediate intervention effects.

Conclusions: This study demonstrated initial efficacy of the FMF Connect app for targeted caregiver outcomes, with small to medium effect sizes. As an mHealth app, the FMF Connect intervention has potential for scalability and accessibility. This could lead to a substantial public health impact, particularly for families who face challenges accessing evidence-based resources or encounter other barriers to care.

Trial Registration: ClinicalTrials.gov NCT05028517; https://clinicaltrials.gov/NCT05028517

(JMIR Mhealth Uhealth 2026;14:e73647) doi:10.2196/73647

XSL•FO
RenderX

KEYWORDS

fetal alcohol spectrum disorders; prenatal alcohol exposure; mobile health; intervention; treatment; parenting; families moving forward; neurodevelopmental disabilities; mobile phone

Introduction

Fetal alcohol spectrum disorders (FASD) are highly prevalent developmental disabilities. Studies estimate that FASD affects 1.1% - 5% of the population [1]. Yet, most people with FASD remain undiagnosed and have inadequate access to services [2]. Diagnosis of FASD focuses on evidence of prenatal alcohol exposure (PAE) and neurobehavioral impairment. Additional symptoms, such as microcephaly, seizures, subtle characteristic facial features, and growth delays, are present in approximately 10% of individuals with FASD [1,3-5]. Individuals with FASD often require lifelong, multidisciplinary, community-based services in order to receive appropriate diagnosis, treatment planning, and daily-living supports [6-9]. Caregivers of children with FASD experience high rates of stress, feelings of isolation, and relationship strain between child and caregiver [7,10,11]. Despite higher needs, individuals with FASD and their families report substantial barriers to services due to limited availability of knowledgeable providers, societal stigma, lack of policy-based supports, and insufficient availability of appropriate, evidence-based services [9,12]. These systematic barriers emphasize the need for evidence-based interventions that are both accessible and tailored to the needs of families managing FASD, presenting an opportunity to apply innovative methods and new implementation strategies.

One method to surmount these barriers is mobile delivery of services or app-based interventions. There has been an increase in technology-augmented parenting programs that have been shown to support intervention engagement, caregiver skill building, and positive child behavior outcomes [13-15]. While promising, these programs require additional live therapeutic services and many are web-based. This trend does not reflect current accessibility and preference patterns, which indicate higher preference for mobile devices compared with desktop services [16]. Most adults have access to smartphones and demonstrate consistent understanding and use of phone-based apps [17]. Mobile health (mHealth) apps provide self-paced access to material, services, and supports that can be more easily scaled to larger populations than in-person or telehealth programs [18]. A recent meta-analysis documented that health behavior apps can result in significant improvement in behavior change across a variety of settings [19]. This suggests that app-based interventions are a promising way to solve problems of intervention accessibility.

Despite the promise of app-based interventions, there have been very few mHealth apps with evidence-based review that follows recommended mental health app guidelines [20]. A meta-review of parenting-based technological interventions by David and colleagues [13] indicated that of the 53 usable parenting-focused apps available in app stores, few offered personalized content or focused on caregiver functioning, and none have yet been evaluated in a randomized controlled trial (RCT).

As mHealth apps offer a promising solution to many barriers faced by families and providers when trying to access FASD-informed care, the FMF (Families Moving Forward) Connect app was developed to provide accessible, evidence-based support for caregivers of children with FASD. The FMF Connect app is the first known stand-alone, self-directed parenting app to be empirically evaluated for efficacy for caregiver and child outcomes [21-23]. The app was derived from the FMF Program, a 14- to 17-session program for caregivers of children aged 3 - 12 years with FASD and challenging behavior problems delivered by a trained specialist [24-26]. This program was developed and tested by Olson and colleagues [24-26] at Seattle Children's Research Institute and the University of Washington, with funding from the Centers for Disease Control and Prevention. The FMF Program was designed to address the needs of diverse family structures, socioeconomic status, and racial and ethnic identities, in clinic, home-based, or, recently, telehealth settings. Derived from developmental and family systems theories, the FMF Program integrates positive behavior support, motivational interviewing, and cognitive behavioral techniques. The FMF Program helps caregivers to understand their child's strengths and challenges and alter cognitions and attitudes toward child behavior from that of willful disobedience to a neurodevelopmental "reframed" approach. Using this neurodevelopmental viewpoint, families are encouraged to focus on accommodations to support their children's needs and are also helped to understand behavior management. Research has shown that the FMF Program offers benefits for both caregivers and children [24,27,28] (personal communication by HCO, March 5, 2016, and March 2, 2017). Among other findings, caregivers experience improvements in their knowledge of FASD and advocacy skills, parenting efficacy, and some specialized parenting practices. Additionally, they report better self-care and a greater sense of their family's needs being met. For these developmentally complex children, the program leads to positive changes in behavior immediately post treatment. A growing number of innovative programs are being derived from the FMF Program to meet the diverse needs of families, affected children and youth of different ages, and various settings [21-23,29,30]. This includes the FMF Connect caregiver app, which is the focus of this study.

The FMF Connect app incorporates core content, principles, and methods of the standard FMF Program while also adding new features and content to support caregivers using the app in a self-directed, digital format. FMF Connect consists of 5 primary components, as illustrated in Figure 1, Learning Modules, Library, Notebook, Family Forum, and Dashboard. A coaching component was also trialed in this study to see if text-based coaching could increase engagement with the app and improve family outcomes.

XSL•FO
RenderX

Figure 1. The 5 main components of the Families Moving Forward (FMF) Connect app: Learning Modules, Library, Notebook, Family Forum, and Dashboard.



The systematic user-centered design approach to develop and test the FMF Connect app has been relatively rare in the field of mHealth apps. This approach, detailed in the study by Petrenko et al [22], emphasizes iterative feedback at each stage of development from community members who represent target users for the app. This process has included focus groups with caregivers on the initial design and functionalities of the proposed app [23], interviews with caregivers and providers following beta testing of initial app prototypes [22], and surveys and interviews with caregivers during a larger-scale feasibility trial [21]. This study represents the next critical phase of our systematic evaluation, which is to test the efficacy of the app in an RCT. The primary objective of this study was to determine if the FMF Connect app was related to improved targeted caregiver and child outcomes over a 12-week period. This study was conducted in the United States and involved a 3-arm parallel RCT with equal allocation to 3 study groups who received (1) FMF Connect app plus text-based coaching to support app engagement, (2) FMF Connect app without coaching, and (3) waitlist control (WLC) group who received the app after study conclusion. Relative to the waitlist control group, we hypothesized that assignment to the FMF Connect groups would be associated with increases in caregiver FASD knowledge, improvements in parenting efficacy and satisfaction, reductions in caregiver attributions of willful child behavior, increases in family needs met, and improved child behavior. We hypothesized that the addition of text-based coaching would result in greater app usage and, in turn, greater change across outcomes. A 6-week time point was also included to test whether improvements in caregiver attributions of child behavior would mediate intervention effects on child behavior and caregiver self-efficacy and satisfaction. We also assessed caregiver perceptions of app quality, consistent with previous feasibility trials [21].

Methods

Ethical Considerations

This study was registered at ClinicalTrials.gov on August 24, 2021, before study initiation. The full study protocol and study consent document are publicly available within the ClinicalTrials.gov study record. All study procedures and materials were reviewed and approved by the University Institutional Review Board (STUDY00006555). Deidentified data were coded by participant number and identifying information was securely stored separately. There were no costs to participants for participating in the study. They received payments for completing surveys (US $40 at baseline, US $20 at 6 weeks, and US $40 at 12 weeks).

Trial Design

This study involved a 3-arm parallel RCT with equal allocation to the following conditions: (1) FMF Connect+Coaching, (2) FMF Connect without coaching, and (3) waitlist control. Quantitative survey data were collected at 3 time points—baseline, 6 weeks, and 12 weeks. The 6-week time point involved an abbreviated battery designed to allow for tests of statistical mediation hypotheses. App usage was recorded through Amazon Web Services (AWS) during the study period. After 12 weeks, study participants in the waitlist control group

XSL•FO
RenderX

received the FMF Connect app. No changes were made to the trial design after initiation.

Participants

The participant population included caregivers of children (aged 3 - 12 y) with FASD or PAE who live in the United States. To be eligible, caregivers needed to be a biological parent or other primary caregiver (eg, foster or adoptive parent, relative, or legal guardian) who was at least 18 years old, living in the United States, and had a smartphone or iPad (Apple Inc) with iOS operating system. Caregivers needed to have a child between the ages of 3 and 12 years with a diagnosis of FASD or confirmed PAE. As part of study registration, caregivers completed a survey about the child's PAE and uploaded diagnostic records, if available. If there was more than 1 child in the family with FASD or PAE in this age range, the caregiver was instructed during screening to choose 1 child as the target of study surveys. The child needed to have lived with the caregiver for at least 4 months and was expected to remain in the home for at least 1 year. Exclusion criteria included the caregiver not being fluent in English (the FMF Connect app and measures were only available in English). Caregivers were also excluded if there was another caregiver of the same child or living in the home who was already enrolled in the study (couples are excluded to prevent dependence in the data), the caregiver participated in a previous trial of the FMF Connect app as part of earlier development phases of this project [21-23], or the family had previously received or was currently receiving the specialist-led standard FMF Program.

Recruitment, informed consent, and data collection procedures were conducted online. Study recruitment strategy and materials were developed through consultation with the University's Clinical and Translational Science Institute, with considerations to engage a diverse participant pool. These considerations included materials representing diverse caregivers and families across race, ethnicity, age, and family structures and strategies to leverage partnerships and social media (ie, Facebook [Meta] and Twitter [subsequently rebranded as X]) outlets across the United States. Participants were recruited through a variety of mechanisms by our study team and collaborators within the Collaborative Initiative on FASD (CIFASD; [31]) and community partners. These included disseminating recruitment materials on social media, through national and regional FASD listserves (eg, FASD United's Weekly Round-up), at relevant conferences, and to participants or patients with whom the study team or collaborators had routine access and/or had agreed to contact about new research opportunities. We aimed to enroll about 150 participants in this study. Power analyses were conducted before recruitment using GPower3. This analysis determined that with a sample size of 120 and assuming α=.05, there will be a power of 0.97 to detect medium effects (f^2=0.15) and power >0.99 to detect large effects (f^2=0.35). We anticipated about 20% attribution based on the previous feasibility trial [21].

Study recruitment materials included links to the study website [32], where potential participants could access study information and link to the study's consenting and screening module. All surveys were administered via REDCap (Research Electronic Data Capture), hosted by the University of Rochester [33,34], and REDCap's eConsent package was used to document participants' informed consent. Participants had the option to review study consent information by video or text and were required to answer several questions to ensure comprehension. Participants were also offered the opportunity to ask questions via direct contact with the study coordinator before proceeding to other parts of the survey.

After completing informed consent, participants were directed in REDCap to answer screening questions to assess for study eligibility, including information about PAE, FASD diagnostic records, and demographic information. The study team determined eligibility using a study-created eligibility verification form in REDCap to carefully review the criteria and records provided. No evidence of bots or fraudulent participants was detected. A globally unique identifier [35] was then assigned to each eligible caregiver participant and child to facilitate data sharing within CIFASD or future studies. Eligible participants then received an email inviting them to complete baseline study surveys.

Outcomes

Table 1 lists all study surveys administered per timepoint. Primary study outcomes included child behavior problems (the Eyberg Child Behavior Inventory [ECBI]), attributions of child behavior (the Reasons for Children's Behavior Scale [RCB]), parenting self-efficacy and satisfaction (the Parenting Sense of Competence Scale [PSOC]), family needs met (the Family Needs Met [FNM] Questionnaire), and caregiver FASD knowledge (the FASD Knowledge and Advocacy [K&A] Questionnaire). Secondary outcomes included child adaptive functioning (the Everyday Life Scale [ELS]) and caregiver use of self-care. We also examined participant ratings of app quality (the User Version of the Mobile Health Application Rating Scale [uMARS]) and patterns of app usage (from AWS). Additional measures of family background and services were included to characterize the sample. There were no changes in measurement after study initiation.

XSL•FO
RenderX

Table . Study assessment battery.

Measure	Baseline	6 weeks	12 weeks
Eyberg Child Behavior Inventory (ECBI)	✓		✓
Reasons for Children's Behavior Scale (RCB)	✓	✓	✓
Parenting Sense of Competence Scale (PSOC)	✓	✓	✓
Family Needs Questionnaire (FNM)	✓		✓
Knowledge & Advocacy Questionnaire (K&A)	✓		✓
Everyday Life Scale (ELS)	✓	✓	✓
Self-Care and Support Assessment	✓		✓
Family Background & Services Survey	✓		✓
User Version of the Mobile Health Application Rating Scale (uMARS)			✓

ECBI

The ECBI [36] is a 36-item caregiver-report scale measuring frequency (Intensity scale) and perception (Problem scale) of disruptive behavior in children aged 2-16 years. Items are rated on a 7-point scale indicating frequency of the behavior. Scores are presented as T-scores (mean 50, SD 10), with higher scores indicating more frequent behavior problems (Intensity) or greater caregiver perception that behavior is problematic. T-scores of ≥60 are viewed as clinically elevated. Internal consistency in the current sample was high (baseline: α=.90 for Intensity and α=.88 for Problem, follow-up: α=.92 for Intensity and α=.91 for Problem).

RCB

The RCB [37] is a 30-item scale assessing caregiver attributions for children's behavior. The RCB includes 7 subscales measuring neurodevelopmental and willful attributions across multiple areas of behavior, including task completion, sensory seeking and avoiding, and dysregulated or disruptive behavior. Items are rated on a 6-point Likert scale ranging from strongly disagree to strongly agree. Higher scores indicate greater agreement with the attribution. Internal consistency across subscales in the current sample was adequate to high (baseline ranging from α=.68 to α=.92, follow-up ranging from α=.67 to α=.95).

PSOC

The PSOC [38] is a 16-item self-report measure of the caregiver's sense of parenting efficacy (ie, perceived competence, problem-solving ability, and capability) and satisfaction (ie, the extent to which the individual enjoys the parenting role or experiences parenting frustration and anxiety). The items in the PSOC are answered on a 6-point scale ranging from "strongly disagree" to "strongly agree." Higher scores on the Efficacy scale indicate lower efficacy, and higher scores on the Satisfaction scale indicate higher satisfaction. Internal consistency in the current sample was high (baseline: α=.80 for satisfaction and α=.80 for Efficacy, follow-up: α=.84 for Satisfaction and α=.80 for Efficacy).

FMN Questionnaire

The FMN Questionnaire [39] includes 19 items relating to potential needs of caregivers of children with FASD or PAE and assesses the degree to which caregivers perceive these needs have been met. It is based on a measure developed for traumatic brain injury [40] and has been used in several previous studies testing FMF and derivative products [21,24,27,28]. Each need is rated on a 4-point scale (1=not at all met and 4=a great deal met) and is scored by computing the average response across items. Higher scores reflect caregivers' perceptions of more needs being met. Internal consistency in the current sample was high (baseline α=.90 and follow-up α=.93).

FASD Questionnaire

The K&A Questionnaire [39] was developed for use in the FMF Program. The version of the K&A used in this study was adapted to emphasize the content most relevant to the FMF Connect app and was informed by response patterns in the larger feasibility trial [21]. It includes 28 core items testing knowledge of FASD, advocacy, and key FMF Program principles. It is scored by calculating the number of items answered correctly. Furthermore, 8 items are rated on a true or false scale, and 20 items are multiple choice. For this study, a total score was computed by summing correct answers. Higher scores indicate higher levels of knowledge of FASD and advocacy. Internal consistency was not calculated for this measure as it measures content mastery targeted by the app versus a unified construct.

ELS

The ELS was developed for this study to measure caregiver-reported child adaptive functioning. The ELS includes 20 items that are each rated twice by the caregiver, yielding 2 scales, Responsibility and Smoothness. Sample items include "getting ready in the morning," "making a transition between tasks," and "calming down when upset." The Responsibility scale, rated on a 5-point Likert scale, assesses to what degree

XSL•FO
RenderX

the child and/or adult is responsible for the behavior. Lower scores reflect that the adult takes full or most responsibility for the behavior, whereas higher scores reflect that the child takes greater responsibility and independence in the behavior. The Smoothness scale, rated on a 5-point Likert scale, assesses how smoothly the behavior goes, ranging from "rarely goes well" to "almost always goes well." Higher scores reflect greater smoothness. Internal consistency in this study's sample was high (baseline: α=.92 for Responsibility and α=.88 for Smoothness, follow-up: α=.94 for Responsibility and α=.90 for Smoothness).

Self-Care Assessment

The Self-care Assessment (SCA) [24,41] was administered at follow-up. Caregivers were asked how their self-care changed over the past 3 months, rated on a 5-point scale (1=a lot less self-care, 2=somewhat less self-care, 3=no change in level of self-care, 4=somewhat more self-care, and 5=a lot more self-care).

uMARS

The uMARS [42] is a 26-item survey measuring user perception of the quality of mHealth apps. It includes 6 subscales—Engagement, Functionality, Aesthetics, Information, Subjective Quality, and Perceived Impact, and a total App Quality score (summary of first 4 subscales). The uMARS was completed at follow-up only. Caregivers rated the items on a 5-point scale that was specific to each question; higher scores indicate more positive impressions of the app. Internal consistency in this study's sample was adequate (α=.62 for Engagement, α=.76 for Functionality, α=.82 for Aesthetics, α=.66 for Information, α=.72 for Subjective Quality, α=.93 for Perceived Impact, and α=.93 for App Quality).

Interventions

FMF Connect Intervention

FMF Connect includes cloud infrastructure and an innovative, multilayered mHealth app. It incorporates tailored content for caregivers of children (aged 3 - 12 y) with FASD or PAE. The target age group for the FMF Connect app was selected to be the same as the standard FMF Program, from which the app was derived. Content and methods of these programs were designed with caregivers of preschool and school-aged children in mind. The FMF Connect app integrates five main components, (1) Dashboard, (2) Learning Modules, (3) Family Forum, (4) Library, and (5) Notebook (Figure 1). Weekly emails were also sent to support motivational engagement.

App education, derived from the standard FMF Program, is packaged in easily digestible learning modules, including text, exercises, and videos tailored for the child's age and presenting behaviors. The FMF Connect app includes videos and examples of diverse children across ages 3 - 12 years. In the first module, families are asked to provide the age of their child and the main behaviors that concern them. Videos throughout the app are selected from a larger library to be most relevant to the age and primary behavior concerns of the caregiver. The Learning Modules are organized into three levels: (1) Getting Started, (2) Taking Action, and (3) Taking It to the Next Level. The Library provides access to additional optional videos and content that can be downloaded and printed to use in advocacy and skill development. The Notebook is a personalized section that organizes caregivers' responses to learning activities in a convenient location for later easy access.

The Family Forum serves to (1) engage and facilitate caregivers' continued usage of the app, (2) support and extend caregivers' implementation of new knowledge and skills, and (3) provide peer support, encouragement, and understanding of the caregiver's own needs. Trained peer moderators moderated discussion in the Family Forum to ensure safety. They also provided support and prompt skills or content areas to consider.

The Dashboard displays caregiver progress in Learning Module completion, use of the Family Forum, and daily and weekly ratings inputted by the caregiver. It also includes the "Tip of the Day" feature.

Complementing the app is a Cloud infrastructure to transparently but securely distribute information, including storing and retrieving data, managing notification and messaging, and synchronizing data on all devices. A local datastore protects the mobile app against network failures and connectivity issues. This local datastore is transparently synchronized with the Cloud database so that all querying and security features are constantly available regardless of network connectivity. Updates are automatically pushed so users have the latest version.

Furthermore, 4 minor updates were made to the app during the first 2 months of the trial. These included small wording changes, optimization of the connection with the backend system, adding banners to the Family Forum to easily identify moderators, fixing an issue with the behavior tracker tool affecting some users, adding a new notification management system, and minor bug fixes for performance improvements.

Coaching Module

A coaching module was developed within the FMF Connect app to help engage caregivers and support continued use and individualized goal setting within the app. The coaching module was activated for caregivers in the FMF Connect+Coaching group only. This module allowed an assigned coach to exchange text-based messages with caregivers through the FMF Connect app's notification system. Coaches had access to a dashboard of their participants' progress and use of the app, which aided in tailoring messages and supporting caregivers in their identified goals. The coach's interface also provided a secure way to provide text-based messages to the participant within the app.

Coaches were graduate students on the project familiar with FMF Program principles and knowledgeable about FMF Connect content and features. They had training in motivational interviewing (MI), which is a key therapeutic technique used in the FMF Program. Coaches used the spirit and skills of MI to evoke the participant's own motivation to learn more about FASD and skills taught in FMF Connect. They also aimed to support caregiver goals by helping link them with content and features within FMF Connect.

XSL•FO
RenderX

Randomization

Randomization was conducted using REDCap's Randomization package. The allocation sequence was developed by a biostatistician at the University not affiliated with the project to randomly allocate participants equally across the 3 study arms. No stratification or blocking was used. The University Research IT department implemented the allocation table in the REDCap Randomization package, and the research team did not have access to the sequence. Once the study coordinator determined a participant was eligible and baseline measures were complete, the study randomization form would determine the participant's group status. The coordinator then informed the participant of their group status and sent instructions on next steps (ie, how to install the app or wait time) via email. Participants in the intervention groups were given a username and password and were given instructions on how to install the app from the Apple App Store. The research assistant processing survey data and payments could not access the REDCap Randomization package and was blind to group status.

Statistical Methods

Data analyses were conducted using SPSS (version 29.0; IBM Corp) and Mplus 8 (Muthén & Muthén). All variables were assessed for outliers, skew, and kurtosis. No extreme outliers or skew or kurtosis were detected. Missing data in SPSS were handled using multiple imputation. Missing data in Mplus used Full Information Maximum Likelihood (FIML).

Demographic variables were analyzed using descriptive statistics in SPSS. Group differences in demographic variables were assessed using 1-way ANOVAs for continuous variables and Pearson chi-square analysis for categorical variables.

Attrition analysis involved analyzing demographic differences at baseline, 6 weeks, and 12 weeks between those who did complete each timepoint and those who did not. For demographics, these differences were conducted using 1-way ANOVAs including all 3 timepoints. Pearson correlations (r) were used to detect relationships between baseline variables, demographic variables (caregiver age and child age), and attrition at all timepoints. App usage and perceived app quality were assessed using descriptive statistics and independent samples t tests grouped by FMF Connect+Coaching and FMF Connect.

Intervention change analyses were intent-to-treat, including all participants assigned to their respective randomly assigned groups whether or not they completed the FMF Connect program, and involved 2-sided hypothesis testing. Analyses used linear mixed modeling in SPSS for variables collected at the 3 timepoints (RCB, ELS, and PSOC). False discovery rate (FDR) corrections for all P values are reported. Simultaneous linear regression was used in Mplus for variables collected at baseline and 12-week follow-up. For regressions, group status and baseline levels were entered in step 1. All regressions were predicting outcomes at 12-week follow-up. FDR correction was used and reported for all regressions. Cohen d effect sizes were calculated for each outcome to demonstrate the magnitude of the intervention change. Additionally, simultaneous linear regressions were carried out to assess whether app usage within the FMF group predicted 6-week change scores and 12-week change scores.

Point-biserial correlations between group (FMF Connect vs WLC) and change scores (6-wk timepoint−baseline) were conducted to see if mediation was plausible. For those significant correlations, mediation analyses were conducted in Mplus using change scores as the mediator and the ECBI (child behavior) and PSOC (parenting efficacy or satisfaction) as the outcomes, controlling for baseline levels.

Results

Sample Characteristics

Recruitment for this study began on January 6, 2022, with enrollment ending on August 31, 2022. The study was completed on December 31, 2022, after all participants had the opportunity to complete follow-up surveys. Baseline sample demographics can be found in Table 2. No harms or unintended effects were identified for participants.

XSL•FO
RenderX

Table . Sample demographics at baseline by group.

	FMF[a] Connect+Coaching (n=43)	FMF Connect (n=43)	WLC[b] (n=43)	Group difference, *P* value
Age, mean (SD)				
Child	7.3 (2.8)	7.9 (2.6)	8.5 (2.4)	.43
Caregiver	45.2 (7.0)	43.1 (7.6)	43.7 (7.6)	.11
Gender, n (%)				
Child				.31
Boy	21 (49)	28 (65)	24 (56)	
Girl	22 (51)	15 (35)	19 (44)	
Caregiver				.23
Man	0 (0)	2 (5)	3 (7)	
Woman	43 (100)	41 (95)	40 (93)	
Race, n (%)				
Child				.07
White	28 (65)	26 (61)	21 (49)	
Black or African American	6 (14)	0 (0)	8 (19)	
American Indian, Native, or Aboriginal	1 (2)	3 (7)	1 (2)	
Asian	2 (5)	1 (2)	1 (2)	
More than one race	4 (9)	13 (30)	10 (23)	
Unknown	2 (5)	0 (0)	2 (5)	
Caregiver				.42
White	40 (93)	41 (95)	39 (91)	
Black or African American	2 (5)	0 (0)	2 (5)	
American Indian, Native, Aboriginal	0 (0)	0 (0)	0 (0)	
Asian	0 (0)	0 (0)	0 (0)	
More than one race	0 (0)	2 (5)	1 (2)	
Unknown	1 (2)	0 (0)	1 (2)	
Ethnicity, n (%)				
Child				.92
Non-Hispanic or Latine	33 (77)	34 (79)	36 (84)	
Hispanic or Latine	7 (16)	6 (14)	4 (9)	
Unknown	3 (7)	3 (7)	3 (7)	
Caregiver				.23
Non-Hispanic or Latine	41 (95)	43 (100)	40 (93)	
Hispanic or Latine	1 (2)	0 (0)	3 (7)	
Unknown	1 (2)	0 (0)	0 (0)	
Marital status of caregiver, n (%)				.01
Married	32 (74)	36 (84)	33 (77)	
Living with partner	4 (9)	1 (2)	1 (2)	
Separated	0 (0)	0 (0)	3 (7)	

XSL•FO
RenderX

	FMF[a] Connect+Coaching (n=43)	FMF Connect (n=43)	WLC[b] (n=43)	Group difference, *P* value
Divorced	1 (2)	6 (14)	4 (9)	
Never married	6 (14)	0 (0)	2 (5)	
Education of caregiver, n (%)				.28
Did not complete high school	1 (2)	0 (0)	1 (2)	
High school degree	1 (2)	0 (0)	1 (2)	
Partial college	4 (9)	7 (16)	2 (5)	
Undergraduate degree	16 (37)	22 (51)	13 (30)	
Partial graduate	5 (12)	2 (5)	6 (14)	
Graduate degree	16 (37)	12 (30)	20 (47)	
Relationship to child, n (%)				.21
Biological mother	1 (2)	0 (0)	0 (0)	
Biological father	0 (0)	0 (0)	2 (5)	
Grandparent	2 (5)	3 (7)	1 (2)	
Other relative	1 (2)	2 (5)	3 (7)	
Adoptive parent	37 (86)	39 (91)	35 (81)	
Foster parent	5 (12)	1 (2)	3 (7)	
Stepparent	2 (5)	0 (0)	1 (2)	
Income (in thousands; US $), mean (SD)	118.2 (76.0)	110.9 (66.2)	108.8 (65.0)	.80
Household structure, mean (SD)				
People in home	4.9 (1.7)	5.0 (1.5)	5.0 (1.6)	.99
Children in home	3.1 (1.6)	3.1 (1.5)	3.1 (1.6)	.99

[a]FMF: Families Moving Forward.

[b]WLC: waitlist control group.

Baseline correlations by group can be found in Multimedia Appendix 1.

Demographic Group Differences

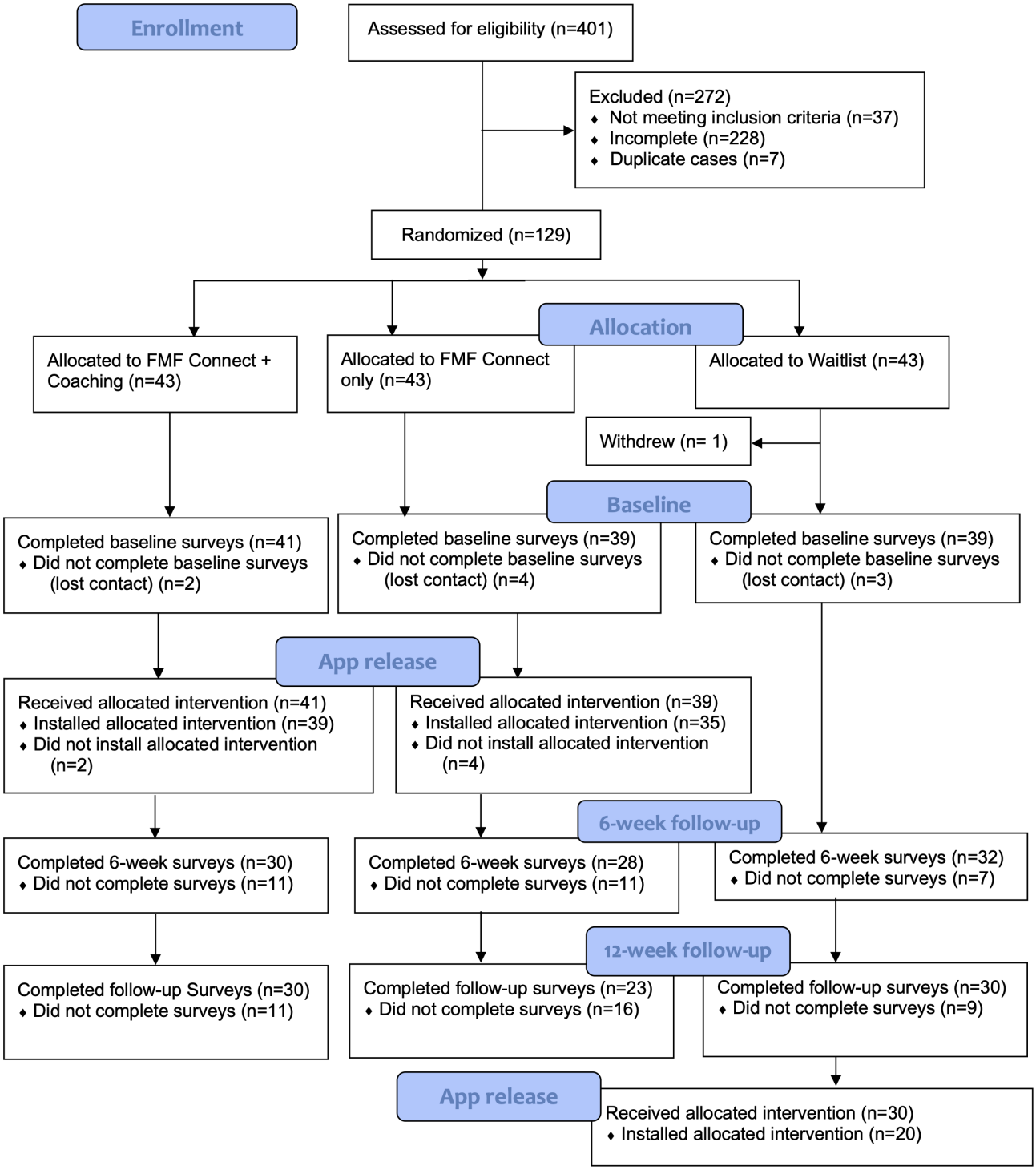
There were no significant differences between the FMF Connect+Coaching, FMF Connect, and WLC groups on any continuous variables (age, income, and household structure), nor were there significant group differences among the majority of categorical variables (gender, race, ethnicity, and relationship to child). There was a significant group difference on variables characterizing caregiver marital status. However, given the small size of marital status subgroups, this variable was not controlled for in further analyses. Given the absence of group differences, no other demographic factors were controlled for.

Attrition Analysis

The trial CONSORT (Consolidated Standards of Reporting Trials) diagram is presented in Figure 2. There were no significant differences in demographic characteristics (caregiver's gender, caregiver's race or ethnicity, caregiver's education level, comfort with technology, knowledge of FASD, income, child's gender, child's race or ethnicity, and caregiver type) between those who did and did not complete data collection at the 6-week or 12-week timepoints. Caregiver's age and child's age were not correlated with attrition. *t*-tests resulted in no significant difference in attrition by number of children in the home.

XSL•FO
RenderX

Figure 2. Study CONSORT (Consolidated Standards of Reporting Trials) diagram for the FMF Connect mobile health intervention randomized controlled trial. FMF: Families Moving Forward.



Differential attrition was also assessed for in baseline levels of functioning across groups. Greater attrition was identified among caregivers who, at baseline, reported lower child responsibility for daily tasks on the ELS for completion at 6 weeks (t_{117}=−2.85; P=.003) and 12 weeks (t_{117}=−2.69; P=.004). No other baseline scores were found to correlate significantly to attrition.

Perceived App Quality

Table 3 provides means, SDs, and score ranges on the uMARS measure of perceived app quality by users in the 2 intervention groups with data at 12 weeks. Consistent with a past feasibility trial [21], users reported high perceived quality and satisfaction with the app, with nearly all scores averaging at or above 4 (maximum 5). The quality of information provided remains the highest rated domain across groups, consistent with the goal of the FMF Connect app. This scale alone was significantly different across groups. The FMF Connect+Coaching group reported slightly lower mean ratings on quality of information than did the FMF Connect app group.

XSL•FO
RenderX

Table . App quality assessment by intervention group at 12 weeks.

	FMF Connect+Coaching (n=29[a]) mean (SD; range)	FMF Connect (n=22[a]) mean (SD; range)	Group differences, *P* value
Engagement	3.8 (0.5; 3.0-4.8)	3.9 (0.3; 3.2-4.4)	.49
Functionality	4.2 (0.5; 3.3-5.0)	4.2 (0.48; 3.0-5.0)	.53
Aesthetics	4.1 (0.6; 3.0-5.0)	4.2 (0.6; 2.7-5.0)	.59
Information	4.5 (0.4; 3.8-5.0)	4.7 (0.3; 4.3-5.0)	.03
App Quality	4.1 (0.4; 3.4-5.0)	4.2 (0.3; 3.6-4.8)	.26
Subjective Quality	4.1 (0.6; 2.8-5.0)	4.1 (0.8; 2.5-5.0)	.99
Perceived Impact	4.0 (0.8; 1.3-5.0)	4.3 (0.6; 3.0-5.0)	.26

[a]Number corresponds to all participants with data on the respective measures at that time point. Participants are included in their assigned group irrespective of app usage patterns. Some participants used the app but did not complete the uMARS at 12-week surveys.

App Usage

App usage in the FMF Connect groups was measured through number of app openings, number of modules completed, number of fact sheet openings, and number of videos watched. Table 4 provides means, SDs, and ranges for participants in each of the intervention groups who registered usage in the app (which is higher than the number who completed follow-up surveys).

Table . App usage by intervention group.

	FMF Connect+Coaching (n=39[a]), mean (SD; range)	FMF Connect (n=35[a]), mean (SD; range)	Group differences, *P* values
App openings	27.18 (26.53; 2-111)	42.23 (143.60; 1-861[b])	.52
Learning Module sections opened (out of 131 maximum[c])	67.54 (54.46; 0-165)	55 (47.47; 0-175)	.30
Learning Module sections completed	61.23 (50.52; 0-148)	49.14 (42.98; 0-138)	.27
Fact sheets opened	7.69 (11.03; 0-50)	7.37 (12.15; 0-65)	.91
Videos started	1.46 (6.78; 0-41)	2.8 (10.21; 0-53)	.51
Videos completed	1.15 (5.95; 0-37)	2.31 (9.26; 0-49)	.52

[a]Number corresponds to all participants with data on the respective measures at that time point. Participants are included in their assigned group irrespective of app usage patterns.

[b]The participant with 861 app openings in the FMF Connect group was an extreme outlier. When this outlier is removed, app opening descriptives are as follows: 18.15 (18.28), 1-75. Group differences are still non-significant (*P*=.10).

[c]While there were 131 maximum learning module sections, modules could have been initiated or completed more than once resulting in a range larger than 131.

For those assigned to coaching, participants messaged their coach an average of 2.95 (SD 2.72) times. The number of messages sent from the participant to their coach ranged from 0 to 10, with 10 participants never using the coaching feature. Coaches messaged participants an average of 6.55 (SD 3.61) times. The number of messages sent from coaches to the participant ranged from 1 to 17.

App usage was not significantly different between the FMF Connect+Coaching and FMF Connect only groups on any of the usage measures, which suggests coaching did not significantly impact app engagement as hypothesized. Pearson *r* correlations were used to see if demographic and baseline data were related to app usage across FMF Connect groups. In terms of demographic factors, older caregivers opened the app (r=0.39, P<.001) and fact sheets (r=0.31, P=.009) more than younger caregivers. Caregivers who had less comfort with technology opened fewer PDFs in the Library (r=−0.23, P=.048). No other demographic factors were associated with app usage.

In terms of baseline measures, caregivers with lower parenting efficacy opened the app (r=−0.31, P=.007) and fact sheets (r=−0.30, P=.01) more. Caregivers with fewer family needs met opened the app (r=−0.25, P=.03) and fact sheets (r=−0.25, P=.03) more than families with greater needs met at baseline. Caregivers with lower baseline FASD knowledge (K&A) finished fewer Learning Modules (r=0.24, P=.04). Caregivers who rated everyday life tasks as less smooth (ELS) opened the app (r=−0.27, P=.02) and fact sheets (r=−0.27, P=.02) more. Caregivers with higher sensory seeking attributions (RCB) opened the app (r=−0.26, P=.03) and fact sheets (r=−0.26, P=.03) less often. Those with higher ability-based task attributions watched fewer videos (r=−0.23, P=.046). In summary, these score patterns suggest that caregivers reporting greater difficulties (eg, less smooth daily tasks, lower efficacy,

XSL•FO
RenderX

and fewer needs met) generally engaged with the app more than those with relatively fewer difficulties.

Intervention Change

The FMF Connect+Coaching and FMF Connect groups were aggregated for analyses examining intervention change, given minimal group differences across demographic characteristics, baseline and follow-up measures, and app usage. Analyses conducted with 3 separate groups, as documented on clinicaltrials.gov, can be found in Multimedia Appendix 2. These analyses did not use intent-to-treat or multiple imputation methodologies; therefore, they should be interpreted with caution.

Table 5 reports linear mixed models for all intervention outcomes with 3 time points (RCB, PSOC, and ELS). Pairwise comparisons for significant models are further discussed in this study.

Table . Linear mixed modeling for intervention change.

Measure	F test (*df*)	*P* value	FDR[a]
Primary outcomes			
RCB[b] Sensation Avoiding attributions	0.65 (2)	.52	0.60
RCB Sensation Seeking attributions	0.61 (2)	.55	0.89
RCB Task Willful attributions	4.60 (2)	.02	0.14
RCB Task Ability attributions	0.78 (2)	.46	0.89
RCB Disruptive attributions	2.54 (2)	.08	0.46
RCB Emotion Seeking attributions	0.04 (2)	.96	0.96
RCB Dysregulated attributions	1.70 (2)	.19	0.52
PSOC[c] Parenting Efficacy	0.38 (2)	.68	0.82
PSOC Parenting Satisfaction	2.08 (2)	.13	0.48
Secondary outcomes			
ELS[d] child or caregiver Responsibility of everyday tasks	0.61 (2)	.54	0.75
ELS Smoothness of everyday tasks	0.30 (2)	.74	0.82

[a]FDR: false discovery rate correction.

[b]RCB: Reasons for Child Behavior.

[c]PSOC: Parenting Self-Efficacy and Satisfaction.

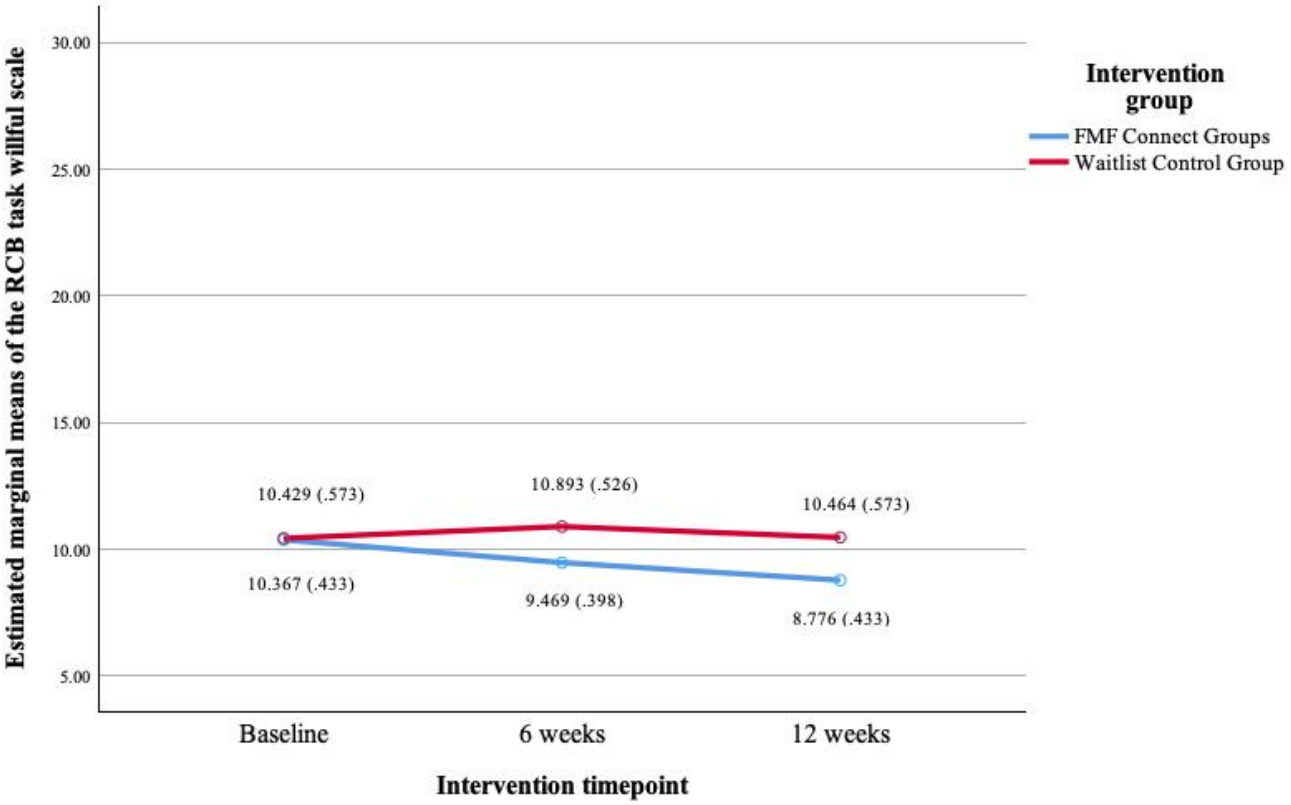
[d]ELS: Everyday Life Scale.

RCB (Primary Outcome)

There was a significant group×timepoint interaction on the RCB Task Willful measure (F_2=4.60, *P*=.01; Figure 3). Pairwise comparisons indicate significant group differences at the 12-week timepoint (mean difference=0.86, *P*=.002), with the FMF Connect group demonstrating lower endorsement of task willful attributions (mean 9.63, SE 0.30) compared with the control group (mean 10.13, SE 0.43). Interactions are no longer statistically significant when applying FDR correction.

XSL•FO
RenderX

Figure 3. Intervention group differences by timepoint on the Reasons for Child Behavior Scale Task Willful subscale (FMF Connect vs waitlist control). FMF: Families Moving Forward; RCB: Reasons for Child Behavior Scale.



No differences were detected for parenting efficacy and satisfaction or child adaptive behavior.

Linear regression analyses were conducted for all outcome measures collected at baseline and 12-week follow-up timepoints only (ECBI, FNM, K&A, and SCA). Baseline and FMF Connect group status were entered into the regression model simultaneously. The dependent variable entered was the measure at the 12-week follow-up to determine treatment effects (Table 6). Covariates were not included in analyses as there were minimal correlations or group differences on any of the demographic variables assessed (Table 2). Significant findings are summarized further in this study. FDR corrections for all *P* values are reported.

XSL•FO
RenderX

Table . Linear regression and Cohen *d* effect sizes for intervention change.

Measure	B (SE)	b	*P* value	FDR[a]	Cohen *d* (95% CI)
Primary outcomes					
ECBI[b] Child Behavior Intensity					−0.22 (−0.69 to 0.23)
Baseline	0.81 (0.07)	0.80	<.001	<0.001	
Group	−1.18 (1.24)	−0.06	.33	0.33	
ECBI Child Behavior Problem					−0.35 (−0.80 to 0.11)
Baseline	0.80 (0.10)	0.67	<.001	<0.001	
Group	−2.82 (1.87)	−0.12	.13	0.14	
Family Needs Met					0.49 (−0.02 to 0.94)
Baseline	0.45 (0.12)	0.38	<.001	<0.001	
Group	4.91 (2.38)	0.21	.04	0.06	
K&A[c] FASD[d] Knowledge					0.47 (0.02 to 0.93)
Baseline	.64 (.10)	0.57	<.001	<0.001	
Group	1.18 (0.57)	0.18	.04	0.06	
Secondary outcomes					
SCA[e] Caregiver Self-Care					0.46 (0.00 to 0.91)
Baseline	0.64 (0.10)	0.57	<.001	<0.001	
Group	0.93 (0.47)	0.17	.048	0.07	

[a]FDR: false discovery rate correction.

[b]ECBI: Eyberg Child Behavior Inventory.

[c]K&A: Knowledge & Advocacy Questionnaire.

[d]FASD: fetal alcohol spectrum disorders.

[e]SCA: Self-Care Assessment.

FMF (Primary Outcome)

Group status (FMF Connect: n=80 and control: n=39) significantly predicted the FNM summary score at 12 weeks (*b*=0.21, *P*=.04), when accounting for scores at baseline and accounting for the interaction. The FMF Connect group showed significantly higher scores on the FNM post intervention (mean 47.56, SD 11.06) compared with WLC (mean 42.30, SD 11.29), indicating more family needs met. This prediction approaches significance when adjusting for multiple comparisons (*P*=.06).

K&A (Primary Outcome)

Group status (FMF Connect: n=80 and control: n=39) significantly predicted the K&A summary score at 12 weeks (*b*=0.18, *P*=.04), when accounting for scores at baseline. The FMF Connect group showed significantly higher scores on the K&A post intervention (mean 21.28, SD 3.28) compared with the WLC group (mean 20.33, SD 2.69), meaning they acquired greater knowledge about FASD, advocating for their child, and core FMF Program concepts. This prediction approaches significance when adjusting for multiple comparisons (*P*=.06).

Self-Care (Secondary Outcome)

Group status (FMF Connect: n=80 and control: n=39) significantly predicted the SCA score at 12 weeks (*b*=0.17, *P*=.048), when accounting for scores at baseline. The FMF Connect group showed significantly higher scores on the SCA post intervention (mean 5.38, SD 2.46) compared with the WLC group (mean 4.401, SD 2.59), meaning they acquired more self-care skills compared with waitlist controls. This prediction approaches significance when adjusting for multiple comparisons (*P*=.07).

Group status did not predict child behavior (ECBI) at 12 weeks when controlling for scores at baseline.

App Usage and Intervention Change (Secondary Outcome)

Learning module completion and app opening were entered into a simultaneous regression model to predict change scores at the 6-week and 12-week timepoints (Multimedia Appendix 3). Learning module completion (*b*=−0.20, *P*=.02) and app openings (*b*=0.79, *P*=.02) had an effect on PSOC Efficacy change scores over 12 weeks, in that those who completed more learning modules and opened the app more had greater levels of parenting

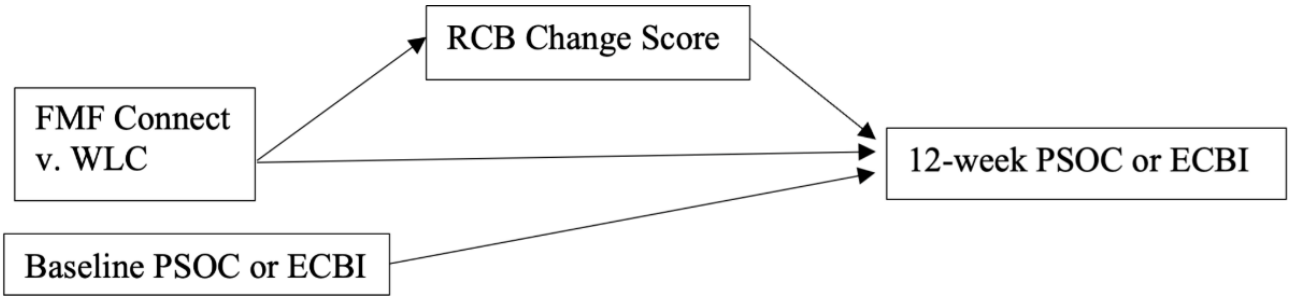
XSL•FO
RenderX

efficacy after 12 weeks. Caregivers who opened the app more showed more improvement on RCB Emotion Seeking at the 12-week timepoint (b=0.76, P=.03) and PSOC Efficacy at the 6-week timepoint (b=0.35, P=.002).

Mediation Hypotheses

Point-biserial correlations between group (FMF Connect vs WLC) and change scores (6-week timepoint−baseline) were conducted to see if mediation was plausible. Group status was correlated with 6-week change scores on the RCB Disruptive Behavior (r=−0.24, P=.03) and RCB Task Willful (r=−0.26, P=.02) scales. Mediation analyses were conducted in Mplus using the model illustrated (Figure 4) looking at both the RCB Disruptive Behavior and Task Willful scores as the mediator and the ECBI and PSOC as the outcome. No significant mediation was found.

Figure 4. Mediation model testing whether caregiver attributions at 6 weeks (measured by the Reasons for Child Behavior scale) mediate the intervention effect of receiving the Families Moving Forward Connect app versus the waitlist control on parenting self-efficacy or satisfaction at 12 weeks (measured by the Parenting Sense of Competence scale). ECBI: Eyberg Child Behavior Inventory; FMF: Families Moving Forward; PSOC: Parenting Sense of Competence scale; RCB: Reasons for Child Behavior scale; WLC: waitlist control.



Discussion

Primary Findings

Families raising children with FASD experience poor access to evidence-based care in their communities given multiple systems-level barriers. These include limited professional training in FASD, long waitlists and provider shortages, and pervasive stigma [9,12,43]. Complementing efforts to build workforce capacity, this study tested the efficacy of FMF Connect, a first of its kind—a caregiver self-directed mHealth intervention. Results from this RCT provide initial support for the efficacy of the FMF Connect app for caregivers raising preschool and school-aged children with FASD or PAE. Relative to the waitlist group, groups that received the FMF Connect app had greater increases in FASD knowledge, reductions in willful attributions of child behavior, improved self-care practices, and more family needs met. These changes in caregiver outcomes are notable given the low intensity, self-directed nature of the intervention. Contrary to hypotheses, no significant group differences were found for child outcomes assessed. Consistent with our previous feasibility trial [21], caregivers gave high ratings for app quality, with the quality of information rated highest among indicators. App usage was relatively greater for caregivers reporting more difficulties with parenting efficacy, completing daily tasks with their children, and meeting family needs. Higher app usage was also related to greater changes in parenting efficacy scores across the 12-week intervention period.

The systematic development and testing of the FMF Connect app is particularly notable given that a 2024 review of parenting apps found that none of the available apps in app stores had been empirically tested [13]. The FMF Connect app is also part of a growing continuum of products, all derived from the standard FMF Program to meet the varying needs of children and youth of different ages, the families who care for them, and different settings in which they are treated. Similar to the standard specialist-led FMF Program, the FMF Connect app is related to important changes in caregiver and family functioning [24,27,28]. The magnitude of effect sizes for the FMF Connect app (small-medium) is generally smaller than the standard FMF Program (many medium-large), which is not surprising given the difference in duration (12 wk vs 9 mo) and format (self-directed vs specialist implemented) of the interventions. Dosage of content is also much more variable with the FMF Connect app (intervention groups averaging 56 of 131 core Learning Module sections; range 0 - 148, including repeated content); past trials with the FMF Program have had nearly all families complete the full program. Across studies, the FMF Program and FMF Connect app result in consistent improvements in FASD knowledge and family needs met. Changes in caregiver attributions of child behavior have also been found across studies, although measurement approaches have been variable. Interestingly, studies testing the standard FMF Program have found medium to large effects in parenting efficacy, whereas this study did not find significant group differences after access to the FMF Connect app. However, higher levels of app usage did relate to greater change in parenting efficacy scores, suggesting greater dosage of content is important for this outcome. Caregivers across studies have reported a high level of satisfaction with both interventions. While the magnitude of effects for the FMF Connect app is more modest, the potential for scalability creates the possibility of a broad public health impact. This impact and reach could potentially be leveraged as additional products in the FMF continuum of care are developed, continuing to increase accessibility and capacity to provide evidence-based care.

Unsupported Hypotheses

In contrast to our hypothesis, text-based coaching did not result in improved app usage or outcomes for families. Across a broad range of mHealth apps designed to manage health conditions in adults, a positive correlation has been found between app engagement and the level of personal support provided [44].

XSL•FO
RenderX

However, there is a high level of heterogeneity in methods and outcomes for mHealth interventions using coaching, and many studies did not directly assess the effect of coaching on outcomes [45]. Another possible explanation for the pattern of findings in this study is that user engagement in coaching was lower than anticipated. Study team observations noted that a few users reported difficulty finding or navigating the coaching messaging component. It is also possible that the users turned off app notifications that would have alerted them to new messages or were not interested in this feature. A future direction could explore whether peer coaches (vs graduate students) increase engagement.

We also predicted that more app usage would relate to greater improvement in outcomes. However, this hypothesis was not generally supported; usage did not correlate with greater effects for the majority of outcomes, with the exception of parenting efficacy noted above. Possibly, there is a threshold effect after accessing a certain amount (or type) of content. Alternatively, change may be related to more individualized patterns of usage or needs based on caregiver and family circumstances. Future larger-scale investigations may allow for greater power and ability to use machine learning techniques to explore usage patterns in more nuanced ways.

Finally, based on the theorized logic model for the FMF Program, this study hypothesized that caregiver attributions of behavior would mediate changes in caregiver efficacy and satisfaction and child behavior outcomes. Results of this study did not support this hypothesis. While the intervention group was correlated with 6-week change on 2 scales, mediation was not significant. In addition, the app was not associated with change in parenting efficacy, satisfaction, or child behavior over the 12-week evaluation period. Given intervention effects are seen on these outcomes in the longer duration follow-ups for the specialist-led FMF program [24,27,28], it is possible that more intensive intervention and/or follow-up is needed for these outcomes.

Strengths and Limitations

Results from this study should be interpreted in the context of several notable strengths and limitations. First, this study empirically tested one of the first stand-alone, self-directed mHealth interventions targeting parenting. Although the FMF Connect app is tailored to caregivers raising children with FASD, findings from this study have broader relevance for parenting programs and use of mHealth interventions more generally. Much of the content and form of FMF Connect could be adapted and trialed with other related populations. The FMF Connect app was rigorously developed and tested using a systematic user-centered design approach [21-23]. The app was derived from an evidence-based intervention with a strong theoretical foundation and adapted for self-directed administration on a mobile platform. Our approach is consistent with best practices in mental health app development and broader principles of user-centered design [20,46].

The generalizability of study findings may well be impacted by characteristics of study participants who enrolled in the study and other design features. Although careful attention was paid to participant reach in previous feasibility trials [21] and efforts were made to recruit a diverse participant pool, enrolled participants were largely White, relatively wealthy, adoptive mothers. Caregiver education and income were higher than national averages. Our study was also limited to participants with Apple iOS devices, which may have influenced patterns of participant demographics. For example, commercial survey data indicate iOS users tend to have higher average income, longer duration of smartphone use per day, and be younger in age than Android users [47]. The decision to limit the study to iOS devices was made given poorer stability of our Android prototypes across user devices in previous feasibility studies [21,22]. This decision likely eliminated about 40% of potential participants based on rates of Android users in our past studies and current market share in the United States [47]. Expanding the app to the Android platform and further engagement or co-design work with demographic groups not well-represented in this study are important future steps to address this limitation. Since this trial, we have subsequently rebuilt the app on Flutter (Google), which allows seamless development and refinement across iOS, Android, and web platforms. We are also in the process of commercializing the app for broad distribution. Some families may not be interested in participating in research due to barriers or mistrust but might be more likely to download the app from the app store and use it outside of the research context. It is possible these steps will increase the diversity of users. Additionally, targeted outreach and collaboration with community organizations serving underrepresented populations could help address this generalizability gap.

As with most behavioral intervention studies, participants were aware of their treatment condition, which could have potentially biased follow-up responses. FMF Connect targeted multiple child and family outcomes. As a result, multiple analyses were run to assess the range of possible intervention effects targeted. The number of analyses may have increased the likelihood of false positive intervention effects, as suggested by FDR corrections.

Conclusions and Future Directions

This study provided initial evidence for the efficacy of the FMF Connect app for targeted caregiver outcomes, with effect sizes in the small to medium range. As an mHealth app, the FMF Connect intervention has potential for scalability and accessibility. Even with modest effect sizes, this potential for scalability could result in a sizeable public health impact, especially for families who may have difficulty accessing evidence-based resources in their communities or have other barriers to care. The FMF Connect app is not intended to replace diagnostic and treatment services delivered by providers. Yet, given the severe lack of FASD-informed providers in the United States, this resource can help provide families with easy access to evidence-based information and ideas for how to access and advocate for services in their communities. Beyond that, through the Family Forum, the app can provide connections with other families with similar experiences, providing peer support.

Building on this work, our team is developing and testing a new program called FMF Connect Pro to further expand the continuum of care for families based on the FMF Program. This effort aims to train mental health providers in how to screen for

XSL•FO
RenderX

PAE, diagnose FASD (using Neurobehavioral Disorder Associated with Prenatal Alcohol Exposure in the *DSM-5* [*Diagnostic and Statistical Manual of Mental Disorders* {Fifth Edition}] [3]), and support families using the FMF Connect app within their clinical practice. This middle tier between the self-directed FMF Connect app and the standard specialist-led FMF Program has the potential to increase identification of FASD in settings with increased prevalence [48,49] and enhance access to FASD-informed treatment for families. It could also provide clinicians with easy-to-use tools and a lower training investment option to gain competencies with this population across a range of mental health settings. It is also possible that caregiver-professional relationships may increase engagement and use with families, especially those less likely to use technology on their own. Engagement with FMF Connect Pro could also increase future provider uptake of the standard specialist-led FMF Program, which may often be the best fit for higher-acuity children and families.

Acknowledgments

The authors appreciate the Collaborative Initiative on Fetal Alcohol Spectrum Disorders (CIFASD) collaborators who assisted with recruitment, as well as FASD United. Seattle Children's Research Institute (SCRI) and the University of Washington are recognized for their contributions to intellectual property from the standard Families Moving Forward (FMF) Program. The standard FMF Program was developed, tested, and materials refined by a team led by Heather Carmichael Olson, PhD, based at and sponsored by these institutions (led by SCRI), with funding via multiple grants from the Centers for Disease Control and Prevention.

We would also like to thank Alicia Roth, Shuo Zhang, Molly Arnold, and Alesia Richardson for their contributions to this study and additional staff at Mt. Hope Family Center for their support throughout the project. The authors appreciate all the families who made time in their busy schedules to participate in this study. This research would not have been possible without their valuable insights.

All or part of this work was done in conjunction with the Collaborative Initiative on Fetal Alcohol Spectrum Disorders (CIFASD, [30]), which is funded by grants from the National Institute on Alcohol Abuse and Alcoholism (NIAAA). Additional information about CIFASD, including information describing available data, can be found at [30]. The data presented in this manuscript was specifically funded under grant U01AA026104 "Development and Evaluation of an Evidence-Based Mobile Health Caregiver Intervention for FASD." Data from this study are part of CIFASD Phase 4 (2017-2022).

Authors' Contributions

All authors contributed to the preparation of this study. CP conceived of the study, oversaw the study team, led app content development, obtained necessary approvals, and led manuscript preparation and revision. MR and JM contributed to data analysis, manuscript preparation, and revision. CT led technological development of the app and contributed to manuscript revisions. CKT assisted in data analysis and contributed to manuscript revision. ES assisted in manuscript revision. ZL assisted in data analysis. LC contributed to app content review and manuscript revision. HCO is the developer of the standard FMF Program, from which the app and original assessment battery are derived. She made significant contributions to intellectual property exchange, app content adaptation and development, and assisted with manuscript revisions. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Baseline variable correlations by group (FMF Connect/Waitlist).
[DOCX File, 22 KB - mhealth_v14i1e73647_app1.docx]

Multimedia Appendix 2
FMF Connect app usage predicting intervention change scores.
[DOCX File, 16 KB - mhealth_v14i1e73647_app2.docx]

Multimedia Appendix 3
Clinical Trials registration three-group (FMFa Connect vs FMF Connect+Coaching vs Control) ANOVA analyses (baseline and 12-week timepoints).
[DOCX File, 22 KB - mhealth_v14i1e73647_app3.docx]

Checklist 1

XSL•FO
RenderX

CONSORT checklist.
[PDF File, 1242 KB - mhealth_v14i1e73647_app4.pdf]

References

1. May PA, Chambers CD, Kalberg WO, et al. Prevalence of fetal alcohol spectrum disorders in 4 US communities. JAMA 2018 Feb 6;319(5):474-482. [doi: 10.1001/jama.2017.21896] [Medline: 29411031]
2. Chasnoff IJ, Wells AM, King L. Misdiagnosis and missed diagnoses in foster and adopted children with prenatal alcohol exposure. Pediatrics 2015 Feb;135(2):264-270. [doi: 10.1542/peds.2014-2171] [Medline: 25583914]
3. Diagnostic and Statistical Manual of Mental Disorders, 5th edition: American Psychiatric Association; 2022. [doi: 10.1176/appi.books.9780890425787]
4. Hemingway SJ. Diagnostic Guide for Fetal Alcohol Spectrum Disorders: The 4-Digit Code, 4th edition: University of Washington; 2024.
5. Hoyme HE, Kalberg WO, Elliott AJ, et al. Updated clinical guidelines for diagnosing fetal alcohol spectrum disorders. Pediatrics 2016 Aug;138(2):e20154256. [doi: 10.1542/peds.2015-4256] [Medline: 27464676]
6. Flannigan K, Coons‐Harding KD, Anderson T, et al. A systematic review of interventions to improve mental health and substance use outcomes for individuals with prenatal alcohol exposure and fetal alcohol spectrum disorder. Alcoholism Clin Exp Res 2020 Dec;44(12):2401-2430 [FREE Full text] [doi: 10.1111/acer.14490]
7. Reid N, Moritz KM. Caregiver and family quality of life for children with fetal alcohol spectrum disorder. Res Dev Disabil 2019 Nov;94:103478. [doi: 10.1016/j.ridd.2019.103478] [Medline: 31476724]
8. Paley B, O'Connor MJ. Intervention for individuals with fetal alcohol spectrum disorders: treatment approaches and case management. Dev Disabil Res Rev 2009;15(3):258-267. [doi: 10.1002/ddrr.67] [Medline: 19731383]
9. Petrenko CLM, Tahir N, Mahoney EC, Chin NP. Prevention of secondary conditions in fetal alcohol spectrum disorders: identification of systems-level barriers. Matern Child Health J 2014 Aug;18(6):1496-1505. [doi: 10.1007/s10995-013-1390-y] [Medline: 24178158]
10. Domeij H, Fahlström G, Bertilsson G, et al. Experiences of living with fetal alcohol spectrum disorders: a systematic review and synthesis of qualitative data. Dev Med Child Neurol 2018 Aug;60(8):741-752. [doi: 10.1111/dmcn.13696] [Medline: 29479676]
11. Petrenko CLM, Alto ME, Hart AR, Freeze SM, Cole LL. "I'm Doing My Part, I Just Need Help From the Community": intervention implications of foster and adoptive parents' experiences raising children and young adults with FASD. J Fam Nurs 2019 May;25(2):314-347. [doi: 10.1177/1074840719847185] [Medline: 31079560]
12. McCormack JC, Chu JTW, Marsh S, Bullen C. Knowledge, attitudes, and practices of fetal alcohol spectrum disorder in health, justice, and education professionals: a systematic review. Res Dev Disabil 2022 Dec;131:104354. [doi: 10.1016/j.ridd.2022.104354] [Medline: 36375286]
13. David OA, Iuga IA, Miron IS. Parenting: there is an app for that. A systematic review of parenting interventions apps. Child Youth Serv Rev 2024 Jan;156:107385. [doi: 10.1016/j.childyouth.2023.107385]
14. Jones DJ, Loiselle R, Zachary C, et al. Optimizing engagement in behavioral parent training: progress toward a technology-enhanced treatment model. Behav Ther 2021 Mar;52(2):508-521. [doi: 10.1016/j.beth.2020.07.001] [Medline: 33622517]
15. Rabbitt SM, Carrubba E, Lecza B, McWhinney E, Pope J, Kazdin AE. Reducing therapist contact in parenting programs: evaluation of internet-based treatments for child conduct problems. J Child Fam Stud 2016 Jun;25(6):2001-2020. [doi: 10.1007/s10826-016-0363-3] [Medline: 27453678]
16. Mobile vs desktop usage in 2020. Perficient. URL: https://www.perficient.com/insights/research-hub/mobile-vs-desktop-usage [accessed 2025-03-15]
17. The growing value of digital health: evidence and impact on human health and the healthcare system. IQVIA institute for human data science. 2017 Nov. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports-and-publications/reports/the-growing-value-of-digital-health [accessed 2025-03-09]
18. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20. [doi: 10.2196/mhealth.3335] [Medline: 25803705]
19. Yang Q, Van Stee SK. The comparative effectiveness of mobile phone interventions in improving health outcomes: meta-analytic review. JMIR Mhealth Uhealth 2019 Apr 3;7(4):e11244. [doi: 10.2196/11244] [Medline: 30942695]
20. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 1;3(1):e7. [doi: 10.2196/mental.4984] [Medline: 26932350]
21. Petrenko CLM, Kautz-Turnbull C, Rockhold MN, et al. Larger-scale feasibility trial of the Families Moving Forward (FMF) Connect mobile health intervention for caregivers raising children with fetal alcohol spectrum disorders. Digit HEALTH 2024;10:20552076241242328. [doi: 10.1177/20552076241242328] [Medline: 38550260]
22. Petrenko CLM, Kautz-Turnbull CC, Roth AR, et al. Initial feasibility of the "Families Moving Forward Connect" mobile health intervention for caregivers of children with fetal alcohol spectrum disorders: mixed method evaluation within a

XSL•FO
RenderX

systematic user-centered design approach. JMIR Form Res 2021 Dec 2;5(12):e29687. [doi: 10.2196/29687] [Medline: 34860661]
23. Petrenko CL, Parr J, Kautz C, Tapparello C, Olson HC. A mobile health intervention for fetal alcohol spectrum disorders (Families Moving Forward Connect): development and qualitative evaluation of design and functionalities. JMIR Mhealth Uhealth 2020 Apr 6;8(4):e14721. [doi: 10.2196/14721] [Medline: 32250274]
24. Bertrand J, Interventions for Children with Fetal Alcohol Spectrum Disorders Research Consortium. Interventions for children with fetal alcohol spectrum disorders (FASDs): overview of findings for five innovative research projects. Res Dev Disabil 2009;30(5):986-1006. [doi: 10.1016/j.ridd.2009.02.003] [Medline: 19327965]
25. Families Moving Forward Program – FASD Intervention, Training & Research. URL: https://familiesmovingforwardprogram.org/ [accessed 2025-03-19]
26. Olson HC, Montague RA. An innovative look at early intervention for children affected by prenatal alcohol exposure. In: Prenatal Alcohol Use and Fetal Alcohol Spectrum Disorders: Diagnosis, Assessment and New Directions in Research and Multimodal Treatment Bentham Science: Bentham Science Publishers; 2011:64-107. [doi: 10.2174/978160805031411101010064]
27. Petrenko CLM, Demeusy EM, Alto ME. Six-month follow-up of the families on track intervention pilot trial for children with fetal alcohol spectrum disorders and their families. Alcohol Clin Exp Res 2019 Oct;43(10):2242-2254. [doi: 10.1111/acer.14180] [Medline: 31408192]
28. Petrenko CLM, Pandolfino ME, Robinson LK. Findings from the families on track intervention pilot trial for children with fetal alcohol spectrum disorders and their families. Alcohol Clin Exp Res 2017 Jul;41(7):1340-1351. [doi: 10.1111/acer.13408] [Medline: 28440861]
29. Olson HC, Kuhn M, Trevino CO, et al. Bringing FASD-informed care to child welfare: an efficient community model for FASD assessment and consultation. Child Welfare Child Welfare League of America 2023;101(3):77-106 [FREE Full text]
30. Kautz-Turnbull C, Carmichael Olson H, Coles C, et al. The families moving forward connect teacher companion website: development and initial pilot randomized trial of an FASD-informed resource for teachers. Alcohol Clin Exp Res (Hoboken) 2025 Dec;49(12):2796-2809. [doi: 10.1111/acer.70189] [Medline: 41225288]
31. Collaborative Initiative on Fetal Alcohol Spectrum Disorders – CIFASD. URL: https://cifasd.org/ [accessed 2025-03-09]
32. Home. FMF Connect. URL: https://www.fmfconnect.com/ [accessed 2025-03-09]
33. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
34. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
35. Johnson SB, Whitney G, McAuliffe M, et al. Using global unique identifiers to link autism collections. J Am Med Inform Assoc 2010;17(6):689-695. [doi: 10.1136/jamia.2009.002063] [Medline: 20962132]
36. Eyberg S, Pincus D. Eyberg Child Behavior Inventory & Sutter-Eyberg Student Behavior Inventory-Revised: Professional Manual.: Psycholgoical Assessment Resources; 1999. URL: https://www.nctsn.org/measures/eyberg-child-behavior-inventory [accessed 2023-01-16]
37. Kautz-Turnbull C, Rogge RD, Petrenko CLM. Child ability and parental attributions: development and validation of the Reasons for Children's Behavior Scale. Child Psychiatry Hum Dev 2024 May 20. [doi: 10.1007/s10578-024-01713-2] [Medline: 38767735]
38. Johnston C, Mash EJ. A measure of parenting satisfaction and efficacy. J Clin Child Psychol 1989 Jun;18(2):167-175. [doi: 10.1207/s15374424jccp1802_8]
39. Olson HC, Oti R, Gelo J, Beck S. "Family matters:" fetal alcohol spectrum disorders and the family. Dev Disabil Res Rev 2009;15(3):235-249. [doi: 10.1002/ddrr.65] [Medline: 19731388]
40. Kreutzer J, Marwitz J. The Family Needs Questionnaire. In: The National Resource Center for Traumatic Brain Injury 1989. URL: https://www.tbims.org/fnq/fnqref.html
41. Kautz C, Parr J, Petrenko CLM. Self-care in caregivers of children with FASD: how do caregivers care for themselves, and what are the benefits and obstacles for doing so? Res Dev Disabil 2020 Apr;99:103578. [doi: 10.1016/j.ridd.2020.103578] [Medline: 32014800]
42. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849] [Medline: 27287964]
43. Roozen S, Stutterheim SE, Bos AER, Kok G, Curfs LMG. Understanding the social stigma of fetal alcohol spectrum disorders: from theory to interventions. Found Sci 2022 Jun;27(2):753-771. [doi: 10.1007/s10699-020-09676-y]
44. Jakob R, Harperink S, Rudolf AM, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371. [doi: 10.2196/35371] [Medline: 35612886]
45. Meyer A, Wisniewski H, Torous J. Coaching to support mental health apps: exploratory narrative review. JMIR Hum Factors 2022 Mar 8;9(1):e28301. [doi: 10.2196/28301] [Medline: 35258468]

XSL•FO
RenderX

46. Baek E, Cagiltay K, Frick T. User-centered design and development. In: Handbook of Research on Educational Communications and Technology: A Project of the Association for Educational Communications and Technology: Routledge; 2008:659-671.
47. iPhone vs android user stats (2024 data). Exploding topics. 2023. URL: https://explodingtopics.com/blog/iphone-android-users [accessed 2025-03-09]
48. O'Connor M, Best A, McCracken J, McCracken J. Under recognition of prenatal alcohol exposure in a child inpatient psychiatric setting. Mental Health Aspects of Developmental Disabilities 2006;9:105-109.
49. Popova S, Lange S, Shield K, Burd L, Rehm J. Prevalence of fetal alcohol spectrum disorder among special subpopulations: a systematic review and meta-analysis. Addiction 2019 Jul;114(7):1150-1172. [doi: 10.1111/add.14598] [Medline: 30831001]

Abbreviations

AWS: Amazon Web Services
CIFASD: Collaborative Initiative on Fetal Alcohol Spectrum Disorders
CONSORT: Consolidated Standards of Reporting Trials
***DSM-5*:** *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition)
ECBI: Eyberg Child Behavior Inventory
ELS: Everyday Life Scale
FASD: fetal alcohol spectrum disorders
FDR: false discovery rate
FIML: Full Information Maximum Likelihood
FMF: Families Moving Forward
FNM: Family Needs Questionnaire
K&A: Knowledge & Advocacy Questionnaire
mHealth: mobile health
MI: motivational interviewing
PAE: prenatal alcohol exposure
PSOC: Parenting Sense of Competence Scale
RCB: Reasons for Children's Behavior Scale
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
SCA: Self-care Assessment
uMARS: User Version of the Mobile Health Application Rating Scale
WLC: waitlist control

Edited by M Waring; submitted 19.Mar.2025; peer-reviewed by A Gibbs, J Chu, R Mccarthy; revised version received 15.Jan.2026; accepted 22.Jan.2026; published 20.Mar.2026.

Please cite as:
Petrenko CLM, Rockhold MN, Myers JM, Tapparello C, Kautz-Turnbull C, Speybroeck E, Li Z, Cole LL, Olson HC
Families Moving Forward Connect mHealth Intervention for Caregivers of Children With Fetal Alcohol Spectrum Disorders: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e73647
URL: https://mhealth.jmir.org/2026/1/e73647
doi:10.2196/73647

© Christie L M Petrenko, Madeline N Rockhold, Julianne M Myers, Cristiano Tapparello, Carson Kautz-Turnbull, Emily Speybroeck, Zhi Li, Lynn L Cole, Heather Carmichael Olson. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 20.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Long-Term Effectiveness of a Game-Based Mobile App for Training in Cardiopulmonary Resuscitation and Automated External Defibrillator Use: Nonrandomized Controlled Trial

Ruting Gu[1*], MSN; Yueshuai Pan[1*], MSN; Jing Han[1], MSN; Min Wang[2], MD; Xiaomin Liu[1], MSN; Xia Liu[1], MBBSMD; Qianqian Li[1], MSN; Jingyuan Wang[1], MSN; Shanshan Ji[2], MSN; Changfang Shi[1], MBBSMD; Haiqing Zhou[1], MSN; Lili Wei[1], MD

[1]Affiliated Hospital of Qingdao University, 16 Jiangsu Road, Qingdao, China

[2]Qingdao University, Qingdao, China

[*]these authors contributed equally

Corresponding Author:
Lili Wei, MD
Affiliated Hospital of Qingdao University, 16 Jiangsu Road, Qingdao, China

Abstract

Background: Bystander cardiopulmonary resuscitation (CPR) and automated external defibrillator (AED) use are critical for improving survival after out-of-hospital cardiac arrest. Although conventional training methods are initially effective, they are often hampered by rapid skill decay over time. Game-based mobile apps have emerged as a promising and scalable alternative for CPR and AED education; however, evidence of their long-term efficacy remains scarce.

Objective: This study aimed to evaluate the integration of a game-based mobile app into traditional CPR and AED training. We assessed its impact on university students' theoretical knowledge, practical skills, and theoretical knowledge retention, as well as their willingness to perform CPR and their awareness of disseminating these skills.

Methods: A nonrandomized controlled trial was conducted among university students in China from March 21 to September 21, 2024. Participants were assigned to either an experimental group, which received game-based mobile app training supplemented with traditional training, or a control group, which received traditional training only. The game-based app featured a simulated scenario that required users to execute the correct sequence of resuscitation procedures and operate a virtual AED under time constraints. The intervention period lasted for 6 months. Participants' theoretical knowledge and practical skills were assessed immediately after training (baseline) and at the 7-day follow-up. Long-term retention of knowledge, willingness to perform CPR, and dissemination awareness were evaluated at the 6-month follow-up. Data were analyzed using SPSS software (IBM Corp), employing the chi-square test, Mann-Whitney *U* test, and Wilcoxon signed-rank test.

Results: A total of 481 participants completed the entire survey (n=241 in the experimental group and n=240 in the control group). In the short-term (7-d) assessment, the experimental group demonstrated significantly higher scores in both theoretical knowledge (P=.02) and practical skills (P<.001) compared to the control group. This advantage was maintained in the long term, with the experimental group showing superior knowledge retention at the 6-month follow-up (median score: 9/10 vs 8/10; P<.001). Furthermore, a majority of all participants expressed willingness to perform CPR on strangers (70.9%, 341/481) and to disseminate first-aid knowledge (92.1%, 443/481). However, no significant intergroup differences were observed for these latter 2 outcomes (P=.85 and P=.97, respectively).

Conclusions: Despite the methodological limitations inherent in this nonrandomized study, our findings indicate that supplementing traditional training with the game-based mobile app significantly enhanced short-term acquisition of theoretical knowledge and practical skills and promoted sustained knowledge retention. This supports the app's potential as an effective and promising complement to conventional CPR and AED training programs.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2500102813; https://www.chictr.org.cn/hvshowproject.html?id=277054&v=1.0

(JMIR Mhealth Uhealth 2026;14:e78689) doi:10.2196/78689

KEYWORDS

game-based; mobile app; cardiopulmonary resuscitation; automated external defibrillator; training

XSL•FO
RenderX

Introduction

Out-of-hospital cardiac arrest (OHCA) is an important global public health problem with poor outcomes [1]. The worldwide incidence of OHCA among apparently healthy adults younger than 40 years of age ranges from 4 to 14 per 100,000 person-years [2]. Of an estimated 350,000 to 450,000 total annual OHCA cases in the United States, only approximately 10% survive [3]. Similarly, in Europe, the annual incidence of OHCA is between 67 and 170 per 100,000 inhabitants, with a mean survival-to-hospital discharge rate of 8% [4]. The crude incidence of emergency medical service–assessed OHCA in China is 95.7 per 100,000 population, with a low rate of resuscitation attempts (31.8%). The outcomes of emergency medical service–treated nontraumatic OHCA are unsatisfactory, with only 1.2% surviving to hospital discharge or 30 days [5]. However, the population in mainland China is approximately 1.41 billion [6], representing a substantial public health burden. Early bystander cardiopulmonary resuscitation (CPR) and rapid public access defibrillation are critical determinants of survival. Unfortunately, the bystander CPR rate in China is suboptimal compared with rates reported in the United States, Europe, and Japan [7,8]. One contributing factor may be the limited coverage and insufficient dissemination of bystander CPR training and retraining in China [9,10]. Therefore, extensive public education in basic life support is essential to improve OHCA outcomes [11].

Recently, technological advancements have become increasingly pivotal in promoting awareness and training for CPR and automated external defibrillator (AED) use [12]. Most studies focus on the initial acquisition and retention of CPR and AED knowledge and skills following training [13]. Initiatives such as the International Liaison Committee on Resuscitation World Restart a Heart Initiative are instrumental in fostering a culture of preparedness and encouraging widespread layperson CPR training, thereby helping to save countless lives [14]. Despite these efforts, key barriers—including uncertainty regarding correct CPR and AED procedures, fear of causing harm, and concerns about legal liability—continue to hinder public willingness to act during cardiac emergencies [15,16]. This underscores a critical need not only to raise public awareness and disseminate information on the importance of CPR and AED but also to develop and provide more innovative and accessible training methodologies [17].

In response to this need, we designed and developed a game-based mobile app for CPR and AED training. This study aimed to investigate whether integrating this app with traditional training could improve the acquisition and retention of CPR and AED knowledge and skills compared with traditional training alone. A nonrandomized controlled trial involving 481 participants from 2 Chinese universities was conducted to evaluate this hypothesis.

Methods

Study Design

We conducted a prespecified posttraining survey to compare the effectiveness of incorporating a game-based mobile app into traditional training methods on CPR and AED theoretical knowledge, practical skills, willingness, and dissemination awareness of CPR and AED among university students. The participants received unified theoretical and operational training. Participants in the experimental group used a game-based mobile app for simulation exercises, whereas those in the control group received no additional intervention. In this study, we strictly followed the Consolidated Standards of Reporting Trial (CONSORT)-EHEALTH (V 1.6.1) guidelines for reporting mobile health intervention trials, as detailed in Checklist 1.

Participants

Using a convenience sampling method, 2 universities in the researcher's city with comparable curricula were selected as research sites. Based on a natural grouping design, School A—which had similar curriculum systems, student compositions, and teaching resources—was assigned as the control group, while School B was designated as the experimental group. The inclusion criteria were as follows: adults aged 18 years or older, ownership of a smartphone (iOS or Android), and no prior participation in any CPR and AED training course. The exclusion criteria included physical limitations that would prevent the performance of CPR and AED techniques (including back or joint conditions).

Ethical Considerations

This study was approved by the Medical Ethics Committee of the Affiliated Hospital of Qingdao University (QYFYWZLL29327). All participants took part voluntarily and provided written informed consent. Participants were informed of their right to withdraw at any time. To ensure privacy, all data were anonymized such that no individuals could be identified in any published results. In addition, no financial compensation was provided to participants.

Features of the CPR and AED Gamification App

The gamified app was designed based on the CPR and AED procedure guidelines provided by the Red Cross Society of China Training Center. Game scenarios were developed accordingly by the research team. Software engineers used the Cocos2D game engine and the C++ programming language for development, which helped reduce costs while maintaining professional quality and achieving high visual realism in operational scenes. The app is compatible with multiple operating systems, including Windows, Android, and macOS, and can be installed on devices such as laptops, tablets, and smartphones via a downloadable installation package. This enables users to operate the app anywhere, even without an internet connection.

The CPR and AED gamified app is divided into 2 portals: a student portal and an administrator portal. The student portal comprises 4 modules: Personal Information, Operational Practice Guidance, Game Module, and System Settings. The administrator portal includes 2 modules: Information Review and Data Statistics. By integrating CPR and AED operational content with game elements, the app is designed to enhance student engagement and learning interest.

XSL•FO
RenderX

Students activated the game by registering with their mobile phone numbers. During gameplay, users interact directly with on-screen elements—such as the shoulder, neck, center of the chest, AED storage compartment button, and AED electrode pad switch—by clicking on the relevant parts and following prompts to complete each step. When moving objects, such as a battery or an electrode plug, users can either click the object and then the target location or drag the object directly to the desired position. Students were required to correctly perform the CPR and AED procedure within a specified time limit. The practice scenario consisted of 6 sequential stages that had to be completed in the correct order. Screenshots illustrating the practice process in the CPR and AED gamification app are provided in Multimedia Appendix 1 (screenshots of the CPR and AED gamification app).

1. Ensure scene safety
   - Approach the scene with caution and confirm that the environment is safe for both rescuers and the person.
2. Check for responsiveness
   - Assess responsiveness by tapping the person's shoulders and shouting (eg, "Are you okay?"). If no response is obtained, immediately check the person's breathing.
3. Check breathing
   - Check for absence of breathing or only gasping while simultaneously checking the carotid pulse. Confirm whether a pulse is definitely palpable within 10 seconds.
4. Activate emergency response system and retrieve AED
   - Send someone for help and to retrieve an AED. If alone, activate the emergency response system (eg, call the local emergency number) and get the AED if immediately available.
5. Begin high-quality CPR: follow the C-A-B sequence
   - C—Chest compressions
     - Position: Heel of one hand on the center of the chest.
     - Technique: Arms straight, shoulders directly over the hands.
     - Depth: 5 to 6 cm.
     - Rate: 100 to 120 compressions per minute.
     - Recoil: Allow complete chest recoil.
     - Minimize interruptions in compressions.
   - A—Open the airway
     - After 30 compressions, open the airway using the head-tilt or chin-lift maneuver.
   - B—Rescue breaths
     - Give 2 breaths, each delivered over about 1 second, ensuring visible chest rise.
     - Compression-to-ventilation ratio: 30:2.
     - Note: If unable or unwilling to give rescue breaths, perform compression-only CPR.
6. Early defibrillation with AED
   - Once the AED arrives:
     - Turn on the device and attach the electrode pads to the person's bare chest.
     - Ensure that no one is touching the person during rhythm analysis.
     - If a shock is advised, clear all individuals from contact with the person and deliver the shock promptly.
     - Resume CPR immediately after shock delivery, beginning with chest compressions.
     - Continue repeated cycles of approximately 2 minutes of CPR followed by AED rhythm analysis as instructed.

Intervention

This study was conducted between March 21 and September 21, 2024. The training sessions were delivered by instructors certified by the Red Cross Society of China. To ensure standardization, all instructors first underwent a researcher-led briefing to ensure a comprehensive understanding of the CPR and AED program and its key learning objectives. On March 21, 2024, a 30-minute theoretical training session on CPR and AED was conducted, followed by a theoretical test. Subsequently, a 30-minute practical demonstration was performed by the researchers in accordance with the standard procedure. Each instructor was assigned to train 10 students, allowing 1 hour for independent practice and question-and-answer interaction. Finally, a training supervisor carried out the skill assessment based on unified evaluation criteria.

Subsequently, the researchers provided the installation package to participants in the experimental group and ensured that each student successfully completed the download and installation process. They also explained the app's usage procedures and module functions to ensure that all students were familiar with its features. Students were permitted to use the game program freely at any time over the following 6 months, without needing to perform any additional operations. Each student was required to spend a cumulative duration of at least 30 minutes per week using the app and to complete at least 1 in-game level during that period. If a student failed to meet the level-completion requirement, the instructor was responsible for identifying the reasons and providing relevant guidance—addressing difficulties and questions raised during gameplay—until the student mastered the corresponding knowledge point and successfully passed the level. Instructors could access students' learning data by searching for their names directly through the management portal.

No additional intervention was provided to the control group during the study period. However, these participants were offered access to the game-based mobile app after they completed the postintervention survey.

Data Collection

Prior to the commencement of the study, all data collectors and CPR and AED skill assessors received standardized training from the investigators. An electronic questionnaire was administered immediately after the theoretical training session

XSL•FO
RenderX

to evaluate participants' knowledge acquisition. Following practical skill training, participants were invited in batches to a unified testing center for CPR and AED skills assessment. Assessors were explicitly instructed that they would evaluate participants from any study group and were required not to inquire about or speculate on group assignments throughout the process. In line with the study protocol and consistent with the structure of our team's first-aid courses at the university, both theoretical and practical tests were repeated 7 days later. Six months after completing the CPR and AED training, all participants were asked to complete a follow-up electronic questionnaire. This survey aimed to compare the 2 training methods in terms of long-term retention of theoretical knowledge, willingness to perform CPR, and dissemination of CPR-AED awareness.

Questionnaires

The study used several questionnaires. First, a general information questionnaire was used to collect demographic data, including age, sex, and major. Second, the CPR and AED Theoretical Knowledge Questionnaire, developed by the Red Cross Society of China, was used. This instrument comprises 10 items covering key aspects of CPR and AED application, with total scores ranging from 0 to 100. The next section assessed participants' attitudes and willingness to perform CPR and AED in cardiac arrest scenarios. The Willingness to Perform CPR and AED Questionnaire was developed by the research team based on a review of relevant literature [18,19], and the final version contained 6 items. Additionally, participants completed a modified version of the Dissemination of CPR Awareness Questionnaire, originally developed by Nas et al [20]. This modified instrument included 4 items specifically designed to evaluate the dissemination of CPR and AED awareness and was used to assess participants' knowledge sharing in this domain.

Sample Size and Power

The sample size calculation was based on a previous study [21], which reported significant differences in CPR compression depth and frequency compliance rates between a VR training group and a traditional training group (51% vs 75%; 50% vs 63%). Using these proportions (p_1 =0.50, p_2 =0.63), a 2-sided α of .05, and a statistical power (1 – β) of 80% (β=.20), the calculated sample size was 227 participants per group. To account for a potential 5% invalid response or attrition rate, the minimum required sample size was adjusted to 239 per group. Accordingly, this study enrolled 241 participants in the control group and 240 in the experimental group.

The formula used to calculate the sample size is as follows:

$$(1)N=(Z1-\alpha/22\bar{p}(1-\bar{p})+Z1-\beta p1(1-p1)+p2(1-p2))2(p1-p2)2$$

Statistical Analysis

All statistical analyses were performed using SPSS version 27.0 (IBM Corp). Descriptive statistics were presented as means with SDs or as numbers and percentages, as appropriate. The normality of the data distribution was assessed using the one-sample Kolmogorov-Smirnov test. Since the skill scores deviated significantly from a normal distribution (*df*=481, *P*<.001), nonparametric tests were used for subsequent comparisons. The Mann-Whitney *U* test was used to compare outcomes between the 2 independent groups, whereas the Wilcoxon signed-rank test was applied for within-group comparisons across different time points. Categorical variables, expressed as numbers and percentages, were compared using the chi-square test or Fisher exact test, depending on the expected frequencies. All *P* values were 2-sided, and statistical significance was set at *P*<.05.

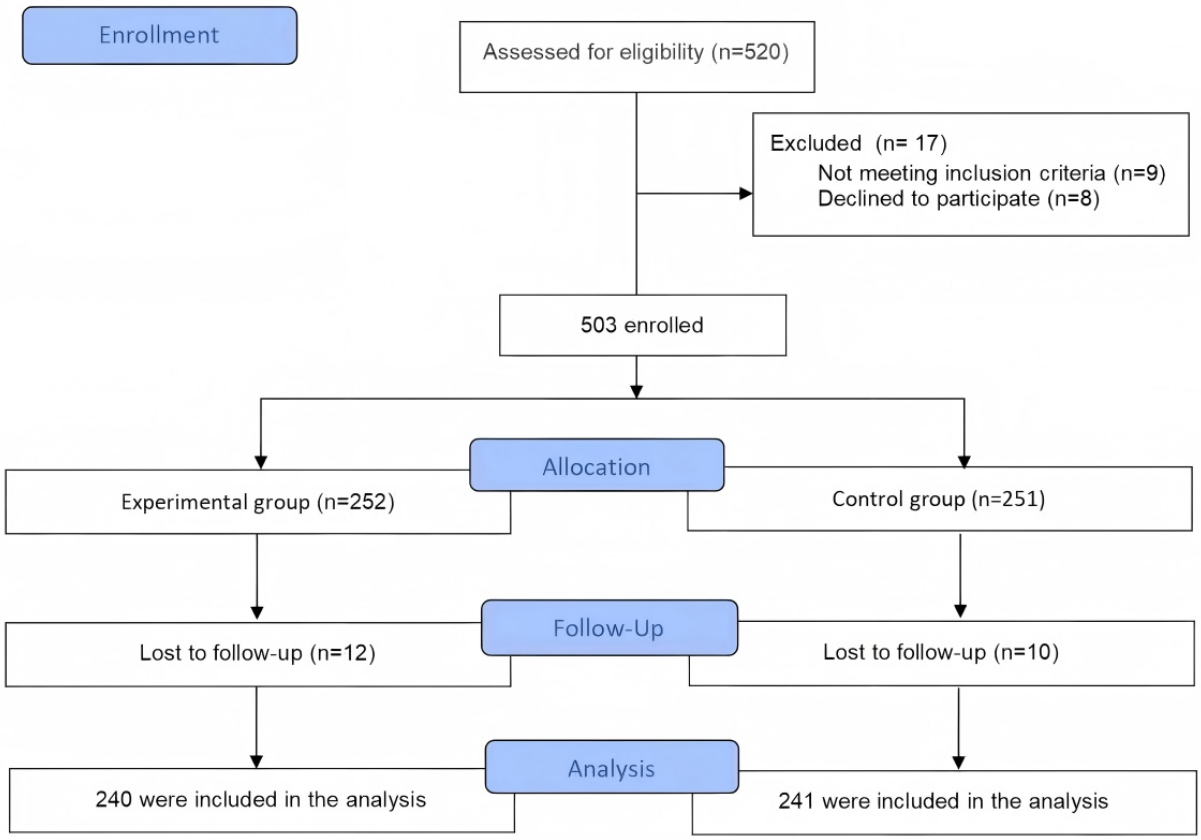
Results

Analytical Sample

A total of 503 participants were enrolled in this study (n=252 in the experimental group and n=251 in the control group). After the 6-month intervention, which included 2 follow-up assessments, 4.8% (12/252) of participants in the experimental group and 4% (10/251) of participants in the control group were lost to follow-up or withdrew. Consequently, the final analysis included 240 and 241 participants in the experimental and control groups, respectively (Figure 1).

XSL•FO
RenderX

Figure 1. Flowchart of included participants.



Participant Characteristics

Table 1 presents the baseline characteristics of the participants. A total of 481 university students were included in this study, comprising 253 men and 228 women, with a median age of 20 (IQR 19-20) years. Regarding academic major, the proportion of medical students was comparable between the control group (38/241, 15.8%) and the experimental group (32/240, 13.3%). No statistically significant differences were observed in any baseline characteristics between the 2 groups (all $P>.05$).

Table . Participants' characteristics at baseline.

Variable	Total (N=481)	Control group (n=241)	Experimental group (n=240)	Z/χ^2 (*df*)[a]	*P* value
Age (y), median (IQR)	20 (1)	20 (1)	20 (1)	1.598	.11
Sex, n (%)				0.051 (1)	.82
Female	228 (47.4)	113 (46.9)	115 (47.9)		
Male	253 (52.6)	128 (53.1)	125 (52.1)		
Major, n (%)				0.573 (1)	.45
Medical specialty	70 (14.6)	38 (15.8)	32 (13.3)		
Nonmedical specialty	411 (85.4)	203 (84.2)	208 (86.7)		

[a](df) indicates the degrees of freedom of the chi-square test.

Theoretical and Practical Skills Tests

At baseline, no significant differences were found in the theoretical and practical skills tests for CPR and AED between the control and experimental groups ($P=.50$ and $P=.63$, respectively; Tables 2 and 3). After 7 days, the experimental group achieved significantly higher scores than the control group in both theoretical knowledge and practical skills ($P=.02$ and $P<.001$, respectively).

XSL•FO
RenderX

Table . Difference in theoretical scores between the control and experimental groups.[a]

Time point	Control group, median (IQR)	Experimental group, median (IQR)	Z	*P* value
First theoretical scores	90 (10)	90 (10)	0.678	.50
7-day theoretical scores	80 (10)	90 (10)	2.323	.02

[a]Within-group comparisons (Wilcoxon signed-rank test): control group: $Z=-6.214$, $P<.001$; experimental group: $Z=-6.9$, $P<.001$.

Table . Difference in skill performance scores between the control and experimental groups.[a]

Time point	Control group, median (IQR)	Experimental group, median (IQR)	Z	*P* value
First practical skill scores	84 (6)	85 (8)	0.481	.63
7-day practical skill scores	84 (7)	87 (4)	7.703	＜.001

[a]Within-group comparisons (Wilcoxon signed-rank test): control group: $Z=-1.241$, $P=.22$; experimental group: Z=5.261, $P<.001$.

In the control group, the theoretical score significantly decreased from the immediate posttest to the 7-day follow-up ($P<.001$; Table 2). In contrast, the observed decline in operational skills was not statistically significant ($P=.22$; Table 3).

In the experimental group, theoretical scores exhibited a statistically significant change between the immediate posttest and the 7-day follow-up ($Z=-6.9$, $P<.001$). Although the median scores remained constant at 90, the distribution of the data changed. This indicates a substantial redistribution in knowledge retention that is not reflected by the median alone (Table 2). Furthermore, practical skills demonstrated a significant improvement at the 7-day follow-up compared with the initial assessment ($P<.001$; Table 3).

Theoretical Knowledge Retention

The responses to individual CPR and AED theoretical knowledge questions are detailed in Table 4. The experimental group achieved a significantly higher total theoretical knowledge score than the control group (9 vs 8 points; $P<.001$). Notably, the experimental group also demonstrated superior performance on several key items: a higher proportion of the experimental group correctly identified the initial step as assessing scene safety (204/240, 85% vs 184/241, 76.3%; $P=.02$), the correct chest compression rate (200/240, 83.3% vs 174/241, 72.2%; $P=.003$), the appropriate compression depth (186/240, 77.5% vs 166/241, 68.9%; $P=.03$), and the correct compression-to-ventilation ratio (189/240, 78.8% vs 169/241, 70.1%; $P=.03$). No significant differences were found for the remaining items (Table 4).

XSL•FO
RenderX

Table . Theoretical knowledge retention.

Questions and answers	Participants, n (%)			χ^2 (*df*)[a]	*P* value[b]
	All (N=481)	Control group (n=241)	Experimental group (n=240)		
Q1. When you see a person lying on the ground, what do you do first?				5.771 (1)	.02
Correct: Check for safety	388 (80.7)	184 (76.3)	204 (85)		
Incorrect	93 (19.3)	57 (23.7)	36 (15)		
Check for response	—[c]	26 (10.8)	17 (7.1)		
Check for breathing	—	31 (12.9)	19 (7.9)		
Q2. The golden time for rescuing person with cardiac arrest is:				0.806 (1)	.37
Correct: 4-6 min	433 (90)	214 (88.8)	219 (91.3)		
Incorrect	48 (10)	27 (11.2)	21 (8.7)		
7-8 min	—	22 (9.1)	19 (7.9)		
More than 10 min	—	5 (2.1)	2 (0.8)		
Q3. To check if the person is conscious, you have to do the following:				0.460 (1)	.50
Correct: Gently shake his (her) shoulders and ask loudly, “Sir (lady) can you hear me?”	422 (87.7)	209 (86.7)	213 (88.8)		
Incorrect	59 (12.3)	32 (13.3)	27 (11.2)		
Gently shake his (her) shoulders, pinch his arm, and scream, “Sir (lady) can you hear me?”	—	22 (9.1)	18 (7.5)		
Gently shake his head and ask loudly, “Sir (lady) can you hear me?”	—	10 (4.2)	9 (3.7)		
Q4. How fast should you perform chest compressions on adults?				8.619 (1)	.003
Correct: 100-120 beats per minute	374 (77.8)	174 (72.2)	200 (83.3)		
Incorrect	107 (22.2)	67 (27.8)	40 (16.7)		
60-80 beats per minute	—	27 (11.2)	21 (8.8)		
80-100 beats per minute	—	40 (16.6)	19 (7.9)		
Q5. How deep should you perform chest compressions when performing CPR[h]?				4.553 (1)	.03
Correct: 5-6 cm	352 (73.2)	166 (68.9)	186 (77.5)		

XSL•FO
RenderX

Questions and answers	Participants, n (%)			χ^2 (*df*)[a]	*P* value[b]
	All (N=481)	Control group (n=241)	Experimental group (n=240)		
Incorrect	129 (26.8)	75 (31.1)	54 (22.5)		
About 4 cm	—	39 (16.2)	31 (12.9)		
About 5 cm	—	36 (14.9)	23 (9.6)		
Q6. Which of the following is not the method of chest compression?				0.075 (1)	.78
Correct: Use the hip joint as the fulcrum and press vertically downward with the strength of the shoulder	344 (71.5)	171 (71)	173 (72.1)		
Incorrect	137 (28.5)	70 (29)	67 (27.9)		
Overlap the bases of the 2 palms	—	34 (14.1)	27 (11.2)		
The wrist, elbow, and shoulder joints are in a straight line without any curvature	—	36 (14.9)	40 (16.7)		
Q7. The ratio of chest compression to artificial respiration on adult is:				4.701 (1)	.03
Correct: 30:2	358 (74.4)	169 (70.1)	189 (78.8)		
Incorrect	123 (25.6)	72 (29.9)	51 (21.2)		
15:1	—	43 (17.9)	28 (11.6)		
15:2	—	29 (12)	23 (9.6)		
Q8. When should you stop chest compressions to give mouth-to-mouth ventilations?				0.438 (1)	.51
Correct: After 30 compressions	385 (80)	190 (78.8)	195 (81.3)		
Incorrect	96 (20)	51 (21.2)	45 (18.7)		
After 10 compressions	—	28 (11.6)	20 (8.3)		
After about 1 min of compressions	—	23 (9.6)	25 (10.4)		
Q9. When rescuing a person with cardiac arrest, if someone brings in an AED[e], what should the rescuer do?				3.418 (1)	.06
Correct: Persist in chest compression, after the AED is installed, defibrillation should be carried out quickly	391 (81.3)	188 (78)	203 (84.6)		
Incorrect	90 (18.7)	53 (22)	37 (15.4)		
To increase the success rate of defibrillation, chest compressions for about 2 min should be performed before electric shock, followed by rapid defibrillation	—	35 (14.5)	28 (11.7)		

XSL•FO
RenderX

Questions and answers	Participants, n (%)			χ^2 (*df*)[a]	*P* value[b]
	All (N=481)	Control group (n=241)	Experimental group (n=240)		
Suspend chest compressions and prioritize the installation and use of AED	—	18 (7.5)	9 (3.7)		
Q10. How do you place the AED pads (Figure 2)?				3.025 (1)	.08
Correct: Figure 2C	418 (86.9)	203 (84.2)	215 (89.6)		
Incorrect	63 (13.1)	38 (15.8)	25 (10.4)		
Figure 2A	—	17 (7.1)	14 (5.8)		
Figure 2B	—	21 (8.7)	11 (4.6)		
Overall					
Total score (each correctly answered question is 1 point)[f]	—	8 (2)	9 (2)	4.034	<.001

[a](df) indicates the degrees of freedom of the chi-square test.

[b]The chi-square test of independence was used to examine the association between group membership (experimental vs. control) and response outcome (correct vs. incorrect). No distinction was made among different types of errors.

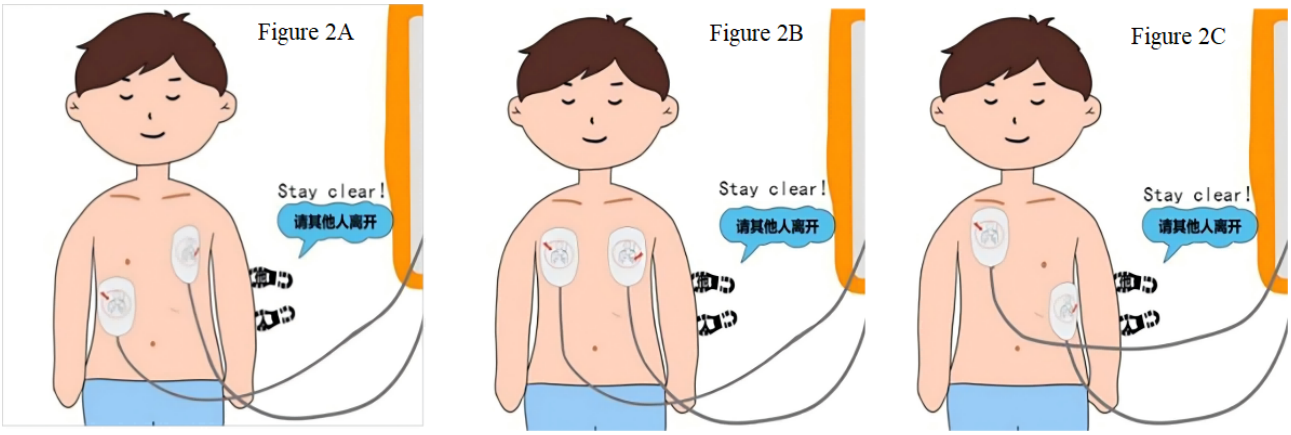
[c]Not applicable.

[d]CPR: cardiopulmonary resuscitation.

[e]AED: automated external defibrillator.

[f]Total scores are median values (IQRs).

Figure 2. Position for placing the automated external defibrillator pads.



Willingness to Perform CPR and AED

Participants' self-reported willingness to perform CPR and use an AED is summarized in Table 5. Overall, 34.1% (164/481) of participants felt capable of performing both CPR and using an AED. Of these participants, 29.9% (144/481) attributed this capability directly to the training received in the study, and this proportion did not differ significantly between groups (*P*=.53). When asked about assisting a stranger, a strong majority of participants (341/481, 70.9%) expressed willingness, with no significant difference observed between the control group (168/241, 69.7%) and the experimental group (173/240, 72.1%) (*P*=.85).

XSL•FO
RenderX

Table . Self-reported attitude and willingness.

Question and answers	Participants, n (%)			χ^2 (*df*)[a]	*P* value
	All (N=481)	Control group (n=241)	Experimental group (n=240)		
1. Do you feel capable to perform CPR[b] after study participation?				1.270 (2)	.53
Yes, because of study participation	164 (34.1)	88 (36.5)	76 (31.7)		
Yes, because I did receive CPR training after the study	144 (29.9)	69 (28.6)	75 (31.2)		
No	173 (36)	84 (34.9)	89 (37.1)		
2. If you were to witness a cardiac arrest of a stranger, would you start CPR?				0.335 (2)	.85
Yes	341 (70.9)	168 (69.7)	173 (72.1)		
No	81 (16.8)	31 (12.9)	28 (11.7)		
Do not know	59 (12.3)	42 (17.4)	39 (16.2)		
3. If you were to witness a cardiac arrest of a family member or friend, would you start CPR?				0.220 (2)	.90
Yes	461 (95.8)	232 (96.3)	229 (95.4)		
No	11 (2.3)	5 (2)	6 (2.5)		
Do not know	9 (1.9)	4 (1.7)	5 (2.1)		
4. I have the feeling that I am capable of helping somebody experiencing cardiac arrest				0.643 (4)	.96
Strongly agree	71 (14.7)	34 (14.1)	37 (15.4)		
Agree	200 (41.6)	101 (41.9)	99 (41.3)		
Neutral	106 (22)	55 (22.8)	51 (21.3)		
Disagree	72 (15)	34 (14.1)	38 (15.8)		
Strongly disagree	32 (6.7)	17 (7.1)	15 (6.2)		
5. Can you accurately recall and apply CPR and AED[c] skills in an emergency?				0.803 (4)	.94
Definitely	61 (12.7)	29 (12)	32 (13.3)		
Basically	188 (39.1)	93 (38.6)	95 (39.6)		
Neutral	139 (28.9)	73 (30.3)	66 (27.5)		
Unlikely	48 (10)	25 (10.4)	23 (9.6)		
Impossible	45 (9.3)	21 (8.7)	24 (10)		
6. I would be scared to perform CPR				0.167 (1)	.68
Yes	202 (42)	99 (41.1)	103 (42.9)		
No	279 (58)	142 (58.9)	137 (57.1)		

[a](df) indicates the degrees of freedom of the chi-square test.

XSL•FO
RenderX

[b]CPR: cardiopulmonary resuscitation.

[c]AED: automated external defibrillator.

The vast majority of participants (461/481, 95.8%) reported that they would perform CPR and use an AED on a relative or friend, and no significant difference was found between the control group (232/241, 96.3%) and the experimental group (229/240, 95.4%) (*P*=.90). Overall, 56.3% (271/481) of participants felt capable of performing these skills, and 51.8% (249/481) reported confidence in recalling and applying first-aid skills when needed; neither measure differed significantly between groups (*P*=.96 and *P*=.94, respectively). In contrast, 42% (202/481) of participants reported fear of performing CPR and using an AED, with comparable levels between the control group (99/241, 41.1%) and experimental group (103/240, 42.9%) (*P*=.68).

Dissemination of Awareness

The willingness to disseminate first-aid knowledge was high among participants, with 92.1% (443/481) expressing this intention.

Overall, 66.3% (319/481) of participants indicated that they would tell their family or friends about the importance of CPR, and this proportion did not differ significantly between the control group (66.8%, 161/241) and the experimental group (65.8%, 158/240) (*P*=.85). Similarly, 61.3% (295/481) of participants had already communicated with relatives or friends about the importance of CPR and AED training, with comparable rates between the control group (147/241, 61%) and the experimental group (148/240, 61.7%) (*P*=.94).

In total, 7.5% (36/481) of participants reported that their family members or acquaintances took concrete action, such as searching for CPR and AED training information, applying for a course, or attending a course. Although the proportion was higher in the experimental group (20/240, 8.3%) than in the control group (16/241, 6.6%), this difference was not statistically significant (*P*=.77). Detailed responses regarding the dissemination of CPR and AED awareness are provided in Table 6.

Table . Dissemination of cardiopulmonary resuscitation and automated external defibrillator awareness.

Question and answers	Participants, n (%)			χ^2 (*df*)[a]	*P* value
	All (N=481)	Control group (n=241)	Experimental group (n=240)		
1. Would you be willing to spread knowledge related to CPR[b] and AED[c] skills to others?				0.253 (3)	.97
Strongly agree	364 (75.7)	183 (75.9)	181 (75.4)		
Agree	79 (16.4)	39 (16.2)	40 (16.7)		
Neutral	33 (6.9)	17 (7.1)	16 (6.7)		
Disagree	5 (1)	2 (0.8)	3 (1.2)		
2. Did you tell family or acquaintances about the importance of CPR and AED?				0.274 (2)	.87
Do not know	32 (6.7)	17 (7.1)	15 (6.3)		
No	130 (27)	63 (26.1)	67 (27.9)		
Yes	319 (66.3)	161 (66.8)	158 (65.8)		
3. Did you tell family or acquaintances about the importance of CPR and AED training?				0.318 (2)	.85
Do not know	29 (6.1)	16 (6.6)	13 (5.4)		
No	157 (32.6)	78 (32.4)	79 (32.9)		
Yes	295 (61.3)	147 (61.0)	148 (61.7)		
4. The reaction of your family members or friends to your participation in the emergency rescue training				0.517 (2)	.77
Do not know	230 (47.8)	117 (48.5)	113 (47.1)		
No response	215 (44.7)	108 (44.9)	107 (44.6)		
Positive reaction	36 (7.5)	16 (6.6)	20 (8.3)		

[a](df) indicates the degrees of freedom of the chi-square test.

[b]CPR: cardiopulmonary resuscitation.

[c]AED: automated external defibrillator.

XSL•FO
RenderX

Discussion

Principal Findings

Integration of the game-based mobile app into training led to greater short-term mastery of both theoretical knowledge and practical CPR and AED skills compared with traditional methods. These findings are consistent with our previous research on other skill training [22] and are supported by preliminary studies indicating that gamification enhances learning outcomes in CPR training [23,24]. The app integrates game design elements into animated procedural simulations, which likely helps maintain learner engagement. Furthermore, its accessibility on mobile phones and computers allows for flexible, repeated practice in various settings. While these features position gamified mobile learning as a flexible, practical, and scalable tool—particularly when integrated with high-fidelity simulation—further rigorously designed studies are needed to confirm its long-term effectiveness and potential for optimizing CPR and AED training.

The follow-up assessment prioritized the evaluation of CPR and AED knowledge retention over practical skills, taking into account some participants' upcoming internships or graduation. Results from the 6-month follow-up indicated a high level of knowledge retention, which is consistent with the findings reported by Nas et al [20]. In this study, the experimental group achieved marginally higher scores across the 10 CPR and AED knowledge items, and this difference was statistically significant. This outcome may be attributed to the game-based app's emphasis on key procedural elements through a clear and visually engaging interface, which likely facilitates comprehension and long-term memory encoding. These observations suggest that a gamified mobile app could play a valuable role in supporting the sustained acquisition of CPR and AED knowledge. It is widely recognized that CPR and AED practical skills are prone to decay over time—a concern highlighted in the European Resuscitation Council Guidelines 2021 [25], which emphasize the need for regular retraining to sustain competency. These guidelines recommend distributing resuscitation training over time and maintaining competencies through frequent retraining. In this context, game-based mobile apps, given their accessibility and self-paced format, could effectively complement traditional training approaches, potentially extending reach and supporting ongoing skill maintenance.

Willingness to perform CPR and use an AED was generally high in both study groups. Compared with the control group, a greater proportion of participants in the experimental group expressed willingness to initiate CPR and AED procedures, and a slightly higher percentage reported feeling capable of assisting a person with cardiac arrest. This may be attributed to the game-based mobile app, which enabled repeated, low-stakes practice in diverse settings, thereby enhancing learners' familiarity and self-efficacy. The self-paced learning feature likely reduced anxiety and fostered incremental confidence—advantages not typically offered by single-session training. Furthermore, the gamified environment may have contributed to a more engaging learning experience, potentially reducing the perceived dread and complexity associated with emergency scenarios [26]. Additionally, the app's design, which breaks down the rescue process into clear, sequential steps with immediate feedback, may have helped users develop a structured mental model for responding in an actual cardiac arrest situation. Such clarity could alleviate the fear of making errors, a commonly reported barrier among lay responders [27]. However, the specific mechanisms underlying these effects require confirmation through studies specifically designed to establish causality. Notably, over 42% of participants across the study reported fear of performing CPR and using an AED, primarily due to concerns about causing secondary harm and facing legal consequences. These are widely documented barriers among laypersons, as also reflected in the study by Daud et al [15]. Together, these findings underscore that improving bystander willingness to act remains a critical challenge for future training initiatives.

Bystander-initiated CPR and rapid public access defibrillation are known to significantly improve survival outcomes after OHCA, with their combination exhibiting a synergistic effect [28]. A primary goal of resuscitation education for the public is, therefore, to raise awareness, train as many laypersons as possible, and develop innovative strategies to save more lives [29]. As is well known, the most common location for OHCA is the home or residence [30]; therefore, training that influences household members is of particular importance. In the present study, over 60% of the participants reported sharing information on the importance of CPR and AED with their family members or friends, indicating a potential diffusion of training awareness beyond the immediate learners. Some of these family members or friends took further active steps, such as searching for related materials, watching demonstration videos, or enrolling in formal courses, in response to participants' encouragement.

This gamified mobile app was developed based on standardized CPR and AED training scenarios. Unlike traditional approaches such as on-site training or video-based instruction, it offers more flexible and accessible learning opportunities. The app incorporates game-like elements to promote interactive and intuitive engagement, as well as active participation among learners and their peers [31]. It also enables repeated, self-directed practice in diverse environments, which may strengthen familiarity with operational procedures. Furthermore, by breaking down the rescue process into sequential steps supported by real-time feedback, the app helps users establish a clear and actionable mental model for responding to an actual cardiac arrest. We propose that integrating this gamified CPR and AED training tool with conventional methods could serve as an effective complementary strategy for broadening the dissemination of CPR and AED knowledge and skills.

Limitations and Future Study

This study has several limitations. First, the use of a nonrandomized controlled trial design represents a key methodological constraint. The lack of random assignment to the experimental or control group may have introduced selection bias, thereby limiting the strength of causal inferences that can be drawn. Future research should employ randomized controlled trials to validate these findings. Second, although high

XSL•FO
RenderX

theoretical knowledge retention was observed, the study did not assess the retention of practical skills. Thus, it remains unclear how well retained knowledge translates into actual CPR and AED performance in real-world settings. Subsequent studies should incorporate longitudinal evaluations of both knowledge and hands-on competencies in simulated or real cardiac arrest scenarios. Third, the absence of a preintervention survey prevented comparisons between baseline and postintervention outcomes, which similarly applies to measures such as willingness to perform CPR and dissemination of awareness. Finally, the generalizability of the findings may be limited by the study population, which consisted predominantly of young, highly educated participants with relatively high information technology literacy. The effectiveness of this game-based mobile app may vary among older adults, individuals with lower digital literacy, or those in different cultural settings. Future studies involving more diverse populations are needed to assess the broader applicability and effectiveness of this approach.

Conclusion

This study evaluated the integration of a game-based mobile app with traditional CPR and AED training. We assessed participants' short-term acquisition of theoretical knowledge and practical skills, as well as long-term knowledge retention, willingness to act, and dissemination awareness after 6 months. The results indicate that supplementing traditional training with the game-based app may improve short-term mastery of both knowledge and practical skills and sustain knowledge retention over 6 months. Overall willingness to perform CPR and disseminate related knowledge was high, with no significant differences between groups. Despite the methodological limitations inherent in this nonrandomized study, the findings indicate that this accessible, interest-stimulating mobile learning tool may serve as a useful complement to existing CPR and AED training paradigms.

Acknowledgments

The authors would like to thank Professor Dejun Lin for his assistance in developing the game-based mobile app. The authors also thank all participants in this study for their contributions. No generative artificial intelligence tools were used in the writing, analysis, revision, or any other process of this manuscript preparation. All content of the paper was independently completed by the authors. The authors take full responsibility for the accuracy, originality, and integrity of the entire manuscript.

Funding

This project was supported by the Qingdao Outstanding Health Professional Development Fund, the Qingdao Philosophy and Social Science Planning Project (grant QDSKL2401066), the Chinese Medical Association Press Nursing Research and Development Project 2025 (grant CMAPH-NRD2025003), and the Affiliated Hospital of Qingdao University "Clinical Medicine+X" Research Program (grant QDFY+X2023205).

Data Availability

The datasets generated and analyzed in this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: RG, YP
Formal analysis: SJ, QL
Investigation: HZ, YP, JH
Methodology: Xiaomin Liu, Xia Liu
Project administration: LW, MW
Software: JH, JW, CS
Supervision: LW, HZ
Writing – original draft: RG
Writing – review & editing: LW, YP

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the cardiopulmonary resuscitation and automated external defibrillator gamification application.
[DOCX File, 6973 KB - mhealth_v14i1e78689_app1.docx]

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 1253 KB - mhealth_v14i1e78689_app2.pdf]

XSL•FO
RenderX

References

1. Lazzeri C, Grippo A, Feltrin G, Peris A, Quatrale R. Out-of-hospital cardiac arrest patients: different donor pathways for an existing donor pool still underestimated-perspective. J Clin Med 2025 Sep 30;14(19):6946. [doi: 10.3390/jcm14196946] [Medline: 41096025]
2. Tseng ZH, Nakasuka K. Out-of-hospital cardiac arrest in apparently healthy, young adults. JAMA 2025 Mar 18;333(11):981-996. [doi: 10.1001/jama.2024.27916] [Medline: 39976933]
3. Tsao CW, Aday AW, Almarzooq ZI, et al. Heart disease and stroke statistics-2023 update: a report from the American Heart Association. Circulation 2023 Feb 21;147(8):e93-e621. [doi: 10.1161/CIR.0000000000001123] [Medline: 36695182]
4. Latsios G, Sanidas E, Velliou M, et al. Cardiac arrest: pre-hospital strategies to facilitate successful resuscitation and improve recovery rates. World J Cardiol 2025 Jan 26;17(1):100782. [doi: 10.4330/wjc.v17.i1.100782] [Medline: 39866210]
5. Zheng J, Lv C, Zheng W, et al. Incidence, process of care, and outcomes of out-of-hospital cardiac arrest in China: a prospective study of the BASIC-OHCA registry. Lancet Public Health 2023 Dec;8(12):e923-e932. [doi: 10.1016/S2468-2667(23)00173-1] [Medline: 37722403]
6. Dai K, Shen S, Cheng C. Evaluation and analysis of the projected population of China. Sci Rep 2022 Mar 7;12(1):3644. [doi: 10.1038/s41598-022-07646-x] [Medline: 35256676]
7. Hou L, Chen G, Kan H, et al. Legislative status of public-access cardiopulmonary resuscitation - China, 2024. China CDC Wkly 2025 May 23;7(21):737-742. [doi: 10.46234/ccdcw2025.122] [Medline: 40620847]
8. Chen Z, Zheng B, Lu Q, Huang S, Lai K. Enhancing bystander cardiopulmonary resuscitation training in China: a narrative review of current situation and strategic recommendations to improve survival outcomes in out-of-hospital cardiac arrest. Resusc Plus 2025 Sep;25:101019. [doi: 10.1016/j.resplu.2025.101019] [Medline: 40688257]
9. Yan S, Gan Y, Wang R, Song X, Zhou N, Lv C. Willingness to attend cardiopulmonary resuscitation training and the associated factors among adults in China. Crit Care 2020 Jul 23;24(1):457. [doi: 10.1186/s13054-020-03165-1] [Medline: 32703294]
10. Mao J, Chen F, Xing D, Zhou H, Jia L, Zhang Y. Knowledge, training and willingness to perform bystander cardiopulmonary resuscitation among university students in Chongqing, China: a cross-sectional study. BMJ Open 2021 Jun 23;11(6):e046694. [doi: 10.1136/bmjopen-2020-046694] [Medline: 34162648]
11. Marijon E, Narayanan K, Smith K, et al. The Lancet Commission to reduce the global burden of sudden cardiac death: a call for multidisciplinary action. Lancet 2023 Sep 9;402(10405):883-936. [doi: 10.1016/S0140-6736(23)00875-9] [Medline: 37647926]
12. Greif R, Bray JE, Djärv T, et al. 2024 international consensus on cardiopulmonary resuscitation and emergency cardiovascular care science with treatment recommendations: summary from the basic life support; advanced life support; pediatric life support; neonatal life support; education, implementation, and teams; and first aid task forces. Resuscitation 2024 Dec;205:110414. [doi: 10.1016/j.resuscitation.2024.110414] [Medline: 39549953]
13. Chien CY, Tsai SL, Huang CH, et al. Effectiveness of blended versus traditional refresher training for cardiopulmonary resuscitation: prospective observational study. JMIR Med Educ 2024 Apr 29;10:e52230. [doi: 10.2196/52230] [Medline: 38683663]
14. Rott N, Reinsch L, Böttiger BW, Lockey A, WRAH Collaborators. ILCOR World Restart a Heart - spreading global CPR awareness and empowering communities to save lives since 2018. Resusc Plus 2025 Jan;21:100853. [doi: 10.1016/j.resplu.2024.100853] [Medline: 39868345]
15. Daud A, Nawi AM, Aizuddin AN, Yahya MF. Factors and barriers on cardiopulmonary resuscitation and automated external defibrillator willingness to use among the community: a 2016-2021 systematic review and data synthesis. Glob Heart 2023;18(1):46. [doi: 10.5334/gh.1255] [Medline: 37649652]
16. Dalle H, Trisyani Y, Nur'aeni A. Innovative training strategies for public response to out-of-hospital cardiac arrest: a scoping review. Risk Manag Healthc Policy 2025;18:1635-1650. [doi: 10.2147/RMHP.S519931] [Medline: 40416834]
17. Uny I, Angus K, Duncan E, Dobbie F. Barriers and facilitators to delivering bystander cardiopulmonary resuscitation in deprived communities: a systematic review. Perspect Public Health 2023 Jan;143(1):43-53. [doi: 10.1177/17579139211055497] [Medline: 35100885]
18. Wang YM, Lin LT, Jiang JH, Jiang Y, Jin XQ. Public knowledge and attitudes toward automated external defibrillators use among first aid eLearning course participants: a survey. J Cardiothorac Surg 2022 May 16;17(1):119. [doi: 10.1186/s13019-022-01863-1] [Medline: 35578261]
19. Cui S, Ye D, Yang H, Zhang L, Chen L. Knowledge, attitude, and willingness to perform on-site cardiopulmonary resuscitation among individuals trained in public CPR: a cross-sectional survey. PLoS ONE 2025;20(5):e0319884. [doi: 10.1371/journal.pone.0319884] [Medline: 40434980]
20. Nas J, Thannhauser J, Konijnenberg LSF, et al. Long-term effect of face-to-face vs virtual reality cardiopulmonary resuscitation (CPR) training on willingness to perform CPR, retention of knowledge, and dissemination of CPR awareness: a secondary analysis of a randomized clinical trial. JAMA Netw Open 2022 May 2;5(5):e2212964. [doi: 10.1001/jamanetworkopen.2022.12964] [Medline: 35587346]

XSL•FO
RenderX

21. Hubail D, Mondal A, Al Jabir A, Patel B. Comparison of a virtual reality compression-only cardiopulmonary resuscitation (CPR) course to the traditional course with content validation of the VR course - a randomized control pilot study. Ann Med Surg (Lond) 2022 Jan;73:103241. [doi: 10.1016/j.amsu.2022.103241] [Medline: 35079374]
22. Gu R, Wang J, Zhang Y, et al. Effectiveness of a game-based mobile application in educating nursing students on flushing and locking venous catheters with pre-filled saline syringes: a randomized controlled trial. Nurse Educ Pract 2022 Jan;58:103260. [doi: 10.1016/j.nepr.2021.103260] [Medline: 34864483]
23. Phungoen P, Promto S, Chanthawatthanarak S, et al. Precourse preparation using a serious smartphone game on advanced life support knowledge and skills: randomized controlled trial. J Med Internet Res 2020 Mar 9;22(3):e16987. [doi: 10.2196/16987] [Medline: 32149711]
24. Cheng P, Huang Y, Yang P, et al. The effects of serious games on cardiopulmonary resuscitation training and education: systematic review with meta-analysis of randomized controlled trials. JMIR Serious Games 2024 Feb 6;12:e52990. [doi: 10.2196/52990] [Medline: 38319697]
25. Greif R, Lockey A, Breckwoldt J, et al. European Resuscitation Council guidelines 2021: education for resuscitation. Resuscitation 2021 Apr;161:388-407. [doi: 10.1016/j.resuscitation.2021.02.016] [Medline: 33773831]
26. Gutiérrez-Puertas L, García-Viola A, Márquez-Hernández VV, Garrido-Molina JM, Granados-Gámez G, Aguilera-Manrique G. Guess it (SVUAL): an app designed to help nursing students acquire and retain knowledge about basic and advanced life support techniques. Nurse Educ Pract 2021 Jan;50:102961. [doi: 10.1016/j.nepr.2020.102961] [Medline: 33421681]
27. Murk W, Goralnick E, Brownstein JS, Landman AB. Quality of layperson CPR instructions from artificial intelligence voice assistants. JAMA Netw Open 2023 Aug 1;6(8):e2331205. [doi: 10.1001/jamanetworkopen.2023.31205] [Medline: 37639274]
28. Luo Y, Wu J. Considerations on witnessed status, bystander CPR, and statistical methods in OHCA defibrillation dose-outcome study. Resusc Plus 2025 Mar;22:100925. [doi: 10.1016/j.resplu.2025.100925] [Medline: 40165804]
29. Horriar L, Rott N, Semeraro F, Böttiger BW. A narrative review of European public awareness initiatives for cardiac arrest. Resusc Plus 2023 Jun;14:100390. [doi: 10.1016/j.resplu.2023.100390] [Medline: 37128626]
30. Martin SS, Aday AW, Allen NB, et al. 2025 heart disease and stroke statistics: a report of US and global data from the American Heart Association. Circulation 2025 Feb 25;151(8):e41-e660. [doi: 10.1161/CIR.0000000000001303] [Medline: 39866113]
31. Alnuaim A. The impact and acceptance of gamification by learners in a digital literacy course at the undergraduate level: randomized controlled trial. JMIR Serious Games 2024 Aug 23;12:e52017. [doi: 10.2196/52017] [Medline: 39177662]

Abbreviations

AED: automated external defibrillator
CONSORT: Consolidated Standards of Reporting Trials
CPR: cardiopulmonary resuscitation
OHCA: out-of-hospital cardiac arrest

Edited by L Buis; submitted 07.Jun.2025; peer-reviewed by S Dhillon, THT Kamauzaman; revised version received 06.Nov.2025; accepted 23.Mar.2026; published 06.May.2026.

Please cite as:
Gu R, Pan Y, Han J, Wang M, Liu X, Liu X, Li Q, Wang J, Ji S, Shi C, Zhou H, Wei L
Long-Term Effectiveness of a Game-Based Mobile App for Training in Cardiopulmonary Resuscitation and Automated External Defibrillator Use: Nonrandomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e78689
URL: https://mhealth.jmir.org/2026/1/e78689
doi:10.2196/78689

© Ruting Gu, Yueshuai Pan, Jing Han, Min Wang, Xiaomin Liu, Xia Liu, Qianqian Li, Jingyuan Wang, Shanshan Ji, Changfang Shi, Haiqing Zhou, Lili Wei. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 6.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

WhatsApp-Supported Teledentistry to Reinforce Oral Health Promotion Among Older Adults Residing in Rural and Urban Areas: Randomized Controlled Trial

Fernanda Muñoz-Sepúlveda[1,2,3], MSc; Víctor Beltrán[1,2], PhD; Alfredo von Marttens[4], PhD; Pablo Navarro[5,6], PhD; Claudia Acevedo[1], PhD; Rodrigo Mariño[7,8], PhD; Alejandra Chaparro[9], PhD; Leonardo López[10], PhD

[1]Clinical Investigation and Dental Innovation Center (CIDIC), Dental School and Center for Translational Medicine (CEMT-BIOREN), Universidad de La Frontera, Temuco, Chile

[2]Interuniversity Center for Healthy Aging (CIES), Talca, Chile

[3]Program of Master in Dental Science, Dental School, Universidad de La Frontera, Temuco, Araucanía, Chile

[4]Department of Prosthodontics , Faculty of Dentistry, University of Chile, Santiago, Chile

[5]Research Center for Dental Sciences (CICO), Dental School, Universidad de La Frontera, Temuco, Chile

[6]Facultad de Ciencias de la Salud, Universidad Autónoma de Chile, Temuco, Chile

[7]Melbourne Dental School, University of Melbourne, Melbourne, Australia

[8]Center for Research in Epidemiology, Economics and Oral Public Health (CIEESPO), Faculty of Dentistry, Universidad de La Frontera, Temuco, Chile

[9]Department of Periodontology, Centre for Biomedical Research, Faculty of Dentistry, Universidad de Los Andes, Chile, Santiago, Chile

[10]Information and Communication Technology Institute, Universidad de La Frontera, Temuco, Chile

Corresponding Author:
Víctor Beltrán, PhD
Clinical Investigation and Dental Innovation Center (CIDIC), Dental School and Center for Translational Medicine (CEMT-BIOREN)
Universidad de La Frontera
Av Francisco Salazar 01145
Temuco
Chile
Phone: 56 45 2325000
Email: victor.beltran@ufrontera.cl

Abstract

Background: Access to oral health promotion for older adults is globally limited, especially in rural, low- and middle-income settings. Digital research often lacks theoretical foundation and focuses primarily on younger cohorts, yielding few randomized trials evaluating accessible tools for oral health education in older adults.

Objective: This study aimed to develop a telehealth reinforcement strategy for oral health promotion to improve knowledge, attitudes, and self-efficacy in community-dwelling older adults.

Methods: A single-center, parallel-group randomized controlled trial was conducted in 4 municipalities (2 urban and 2 rural) in La Araucanía, Chile. Eligible participants were functionally independent adults aged ≥60 years with smartphone and internet access; those with cognitive impairment, complete edentulism, or inability to use WhatsApp were excluded. Participants were recruited from regional databases and assessed using the Geriatric Dental Specialties Tele-platform, a teledentistry tool for older adults. Participants were randomized (1:1) to face-to-face instruction (comparator) or the same instruction plus 2 weeks of social cognitive theory–informed telehealth reinforcement (4 validated videos via WhatsApp). Clinicians and statistical advisors were blinded. Primary outcomes (oral health knowledge, attitudes, and self-efficacy) were measured via telephone-administered questionnaires at baseline and 6 weeks post intervention. Secondary outcomes included acceptability and self-reported behaviors. Analyses included hypothesis testing, multiple correspondence analysis, and k-means clustering.

Results: A total of 120 older adults were randomized (comparator: n=59; telehealth: n=61), with 103 analyzed (comparator: n=51; telehealth: n=52). Both groups showed substantial within-group improvements in oral health knowledge (comparator: Cohen *d*=0.93, 95% CI 0.52-1.34; *P*<.001; telehealth: Cohen *d*=1.07, 95% CI 0.66-1.48; *P*<.001) and self-efficacy (comparator: *r*=0.59, 95% CI 0.38-0.74; *P*<.001; telehealth: *r*=0.62, 95% CI 0.43-0.77; *P*<.001). In per-protocol analysis, telehealth improved

XSL•FO
RenderX

dental caries knowledge (*P*=.03) and attitudes (*P*=.004), with no between-group differences in other domains (*P*>.05). In intention-to-treat analysis, telehealth showed a significant between-group difference for attitudes only (adjusted mean difference=0.91, 95% CI 0.34-1.48; *P*=.002), with no differences for overall oral health knowledge (*P*=.11) or self-efficacy (*P*=.59). Exploratory analyses indicated only the rural telehealth subgroup showed significant gains in attitudes (*P*=.003) and flossing (*P*<.001). Clustering suggested greater improvements among participants with higher baseline needs, predominantly rural, with fewer teeth. Telehealth demonstrated acceptability across multiple indicators (>80% for most measures) with no clinical adverse events; minor video-access issues occurred.

Conclusions: Telehealth reinforcement provided significant advantages in oral health attitudes compared with face-to-face instruction. The intervention was acceptable and showed benefits among older adults with higher preventive needs, commonly seen in rural settings. By integrating theory-informed strategies into a familiar digital platform, this study adds evidence from rural and urban contexts, extending prior work on mobile oral health. It offers insights to address service gaps in underserved areas and highlights potential for feasible, context-aligned implementation. Future research should evaluate long-term effects, adaptability, and cost-effectiveness.

Trial Registration: ClinicalTrials.gov NCT05917548; https://clinicaltrials.gov/study/NCT05917548

(JMIR Mhealth Uhealth 2026;14:e71251) doi:10.2196/71251

KEYWORDS

e-health; geriatric dentistry; health education; health inequities; health promotion; healthy aging; m-oral health; social media; teledentistry; telemedicine

Introduction

Oral Health Context in Older Adults

As global life expectancy continues to rise, promoting healthy aging has become a critical public health priority, with the older population projected to more than double by 2050 [1]. However, significant oral health disparities persist among this population, who are disproportionately affected by preventable conditions—dental caries, periodontal disease, dry mouth, and tooth loss, among others [2-4].

Despite global efforts, significant disparities in oral health promotion remain, especially in low- and middle-income countries, where older adults face unique barriers such as limited access to health care and lower health literacy [5]. Additional barriers include insufficient health coverage, scarcity of tailored programs, declining overall health, and social determinants such as income, educational level, and rural residence [6-10]. Rural communities, home to 43% of the world's population, are especially underserved [11,12]. In Chile, 14.1% of older adults live in rural areas—a proportion that continues to grow—highlighting the urgent need to address territorial inequities in oral health and to develop more equitable strategies [12,13].

Oral Health Promotion and Strategies Based on Behavioral Change Theories

Health promotion empowers individuals to improve their well-being; health education is its core instrument for enhancing knowledge, motivation, and self-management skills [14]. Reinforced educational interventions can prevent disease, support well-being, and foster informed decision-making [15]. In oral health, educational interventions have shown to improve plaque control, gingival health, and salivary function by targeting knowledge, attitudes, and self-efficacy—key predictors of behavioral change [16-18].

By addressing psychosocial determinants, interventions based on psychological theories, such as Bandura's social cognitive theory (SCT), can support behavioral change [19-21]. SCT emphasizes the role of self-efficacy, observational learning, and reinforcement in promoting behavior change [22]. Among older adults, fostering self-efficacy, the belief in one's ability to perform a specific behavior to achieve a desired outcome, is key to promoting autonomy and long-term well-being, as it strengthens their physical abilities and fosters a sense of active participation in addressing daily challenges [23-25]. This psychological resource plays a pivotal role in influencing long-term oral health behaviors, linking knowledge and healthy actions such as toothbrushing and overall oral care, which are associated with healthy aging lifestyles [26-31].

Innovations in Oral Health for Older Adults: The Role of Digital Technologies

Persistent inequities in health demand a fundamental shift in how care is conceptualized and delivered. Information and communications technologies (ICTs) play a central role in advancing several Sustainable Development Goals (3, 4, 5, 9, and 17) by strengthening health systems and promoting universal health coverage [32-34]. For older adults, ICTs offer notable benefits—such as improved access to care, fewer in-person visits, lower costs, and enhanced overall quality of life [35-39]. However, adoption remains uneven due to age-related barriers, including physical limitations, resistance to change, low digital literacy, and privacy concerns—challenges known collectively as the "gray digital divide" [37,40,41].

In dentistry, ICT-based solutions have improved knowledge, promoted self-care, facilitated early disease detection, and enabled timely referrals to specialized care [41]. Some oral health promotion strategies, known as telehealth education [42], have been implemented among community-dwelling older adults through web platforms, text messages, and mobile apps [31,43-45]. A promising innovation within this field is mobile oral health (m-oral health), which uses mobile devices to

XSL•FO
RenderX

promote and maintain oral health [42]. This strategy could prove particularly effective as technology use among older adults is increasing, using it not only for communication but also to seek health-related information—such as diseases, symptoms, treatments, nutrition, and health promotion [46-48]. Current internet usage among this population ranges from 32% to 99%, with higher rates observed among individuals with greater educational attainment and those living in urban settings (91% vs 76% in rural areas) [13,49]. Since the COVID-19 pandemic, smartphone adoption has also grown, reaching 60% among adults aged 60 years and older, although adoption remains significantly lower (20%) among individuals aged 80 years and older [50,51]. Additionally, 55.5% of individuals aged 60 years and older report using social media platforms—such as WhatsApp, Facebook, YouTube, and Instagram—for both information and interaction [51]. Although still underexplored, mobile health (mHealth) and social media are gaining traction and hold great potential to support older adults, particularly as many seek additional health information after clinical encounters but often lack guidance on where to find it [52,53].

When integrated with telehealth education, social media allow on-demand access to instructional content, enhancing their comfort and independence [52,54]. WhatsApp in particular has been increasingly used in health education interventions, demonstrating positive outcomes and highlighting its potential for supporting chronic disease management and health promotion [55]. It has shown effectiveness in reinforcing behaviors related to medication adherence [56], chronic disease management [57], and increasing the uptake of cervical cancer screening [58]. In oral health, social media-based interventions have improved knowledge and reduced gingival index scores, mainly in younger cohorts [46]; extending these approaches to older adults offers a promising way to integrate tailored oral health promotion into daily routines, potentially enhancing their well-being and benefiting their families and communities [36].

Therefore, this randomized controlled trial aims to develop and evaluate a theory-informed telehealth educational reinforcement intervention delivered via WhatsApp, assessing its impact on oral health knowledge, attitudes, and self-efficacy among community-dwelling older adults living in rural and urban areas of the La Araucanía Region, Chile. We hypothesize that participants receiving the intervention will demonstrate greater improvements in these outcomes compared with the comparator group.

Methods

Ethical Considerations

Ethics approval was obtained from the Research Ethics Committee of the Universidad de La Frontera, Temuco, Chile (Folio 082/22), prior to study initiation. The study was conducted in accordance with institutional guidelines, national regulations, and the Declaration of Helsinki. Study activities began on March 3, 2023; participant enrollment occurred between March and August 2023, and the trial concluded on October 14, 2023. The trial was retrospectively registered at ClinicalTrials.gov (NCT05917548) on June 6, 2023, due to an administrative delay in completing the registry process. Primary and secondary outcomes and the statistical analysis plan were prespecified prior to participant enrollment and were not modified thereafter. No deviations from the original study protocol were made during the conduct of the trial. Although these documents were not publicly posted at that time, all prespecified methodological and analytical procedures are transparently and comprehensively reported in this manuscript.

Participants first provided verbal consent by phone to be contacted and screened for eligibility. All participants then signed written informed consent after explaining study procedures (educational activities, survey administration at 3 points), voluntary participation, the right to withdraw at any time, and potential benefits and harms of participation. The anticipated benefits included improvements in oral health knowledge, attitudes, and self-efficacy; however, to minimize bias, the specific study aims were withheld in accordance with the Council for International Organizations of Medical Sciences guideline 10 [59]. No physical harm was anticipated due to the noninvasive nature of the intervention; potential risks were limited to minor discomfort from survey completion or digital content delivery. Data were anonymized prior to analysis, and no financial or material incentives were provided. This study did not collect or include any identifiable images of participants; therefore, no image-related consent was required.

Patient and Public Involvement

Participants were not involved in study design, conduct, or reporting; they participated only in the recruitment, consent, and data collection phases. The intervention procedures and outcomes were developed by the research team based on evidence and expert consensus in geriatric oral health and telehealth.

Trial Design

This single-center, parallel-group, randomized controlled clinical trial used a 1:1 allocation ratio to compare the impact of 2 health education interventions: one delivered face-to-face and the other through ICT as an educational reinforcement. The trial followed a superiority framework to evaluate whether the ICT-based reinforcement produced greater improvements in oral health knowledge, attitudes, and self-efficacy than standard face-to-face education. The report follows the CONSORT (Consolidated Standards of Reporting Trials) 2025 statement and CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (Multimedia Appendix 1) [60,61].

Trial Setting

The study was conducted in the La Araucanía Region in southern Chile, selected for its high proportion of population aged ≥60 years (18.9%) [62], and its combination of rurality and social vulnerability, which contribute to health inequities and low frequency of routine and preventive oral health care [9,63]. Through a presidential social grant, our team provided oral health care to 480 older adults in the Region via a teledentistry workflow [64,65]. Participants were recruited by telephone from the regional National Senior Citizen Service (Chile) branch and local community organization databases and invited if eligible to attend appointments at a mobile dental

XSL•FO
RenderX

clinic in the municipalities of Temuco (the regional capital, urban), Angol (urban), Freire (rural), and Los Sauces (rural).

Eligibility Criteria

The inclusion criteria were adults aged 60 years or older, of any gender, functionally independent (Barthel index ≥90), cognitively able to follow instructions, and with access to a smartphone with internet connectivity. Exclusion criteria included cognitive impairment (Mini-Mental State Examination score of <25), complete edentulism, or inability to use WhatsApp.

Sample Size

Sample size was calculated using G*Power (version 3.1.9.7; Heinrich Heine University Düsseldorf) for Windows, to assess differences between 2 independent means with unequal variances. The calculation was based on a prior study comparing oral health knowledge gains in adolescents who received face-to-face health promotion (SD 0.66) versus those who received additional video reinforcement via mobile app (SD 0.52), showing a knowledge difference of 0.42 points and an effect size of 0.7 [66]. Given the differences between adolescents and older adults in terms of cognitive processing, health literacy, and digital engagement, a more conservative effect size of 0.5 was selected. With a significance level of 0.05 and a power of 0.80, the minimum required sample size was 102 participants. Accounting for potential attrition, and in line with average withdrawal rates in clinical trials [67], the final sample included 120 participants randomly assigned to each group.

Stratified random sampling was conducted based on age categories (60-69 years, 70-79 years, and ≥80 years) and geographic sector (urban/rural by province), adjusted based on the percentage of older adults with access to a smartphone within each age group [51].

Randomization and Blinding

Participants were randomly allocated to the comparator or the telehealth education group in a 1:1 ratio using the EPIDAT random number generator (version 4.2) [68], stratified by age group. Randomization was managed by an independent researcher not involved in recruitment or intervention delivery. After confirming eligibility and obtaining written consent, the principal investigator contacted the independent researcher to receive each participant's group allocation.

To prevent contamination and crossover, the general dentist, outcome assessor, and data analysts were blinded to group assignments. Although participants were aware of the intervention during the reinforcement phase, they signed an agreement committing not to share the videos with family or friends if they received them.

Development of Educational Materials

A pilot study in rural La Araucanía (n=76) identified knowledge gaps that informed the development of 4 educational modules: self-care, dental caries, periodontal disease, and oral cancer. Guided by SCT, materials were designed to model desired behaviors, build behavioral capability, offer verbal persuasion from professionals, and present positive outcomes to foster self-efficacy [22]. Content followed clinical guidelines [69-72] and focused on oral hygiene practices (toothbrushing and flossing), disease prevention, and symptom recognition.

Quality was evaluated by 8 academic reviewers using the Item Content Validity Index (I-CVI≥0.83) and the Patient Education Materials Assessment Tool for Audiovisual Materials (adequacy threshold ≥70% for both understandability and actionability) [73,74]. Existing videos on oral hygiene and oral cancer prevention already met these criteria (understandability: 91% and 87%; actionability: 100% for both). Revisions improved readability, clarity of clinical terms, and the use of informative headings. Newly developed scripts on dental caries and periodontal disease achieved full content validation (I-CVI=1.0), except for the "periodontal disease risk factors" item (I-CVI=0.88). Final videos scored 97% and 98% in understandability and 100% in actionability for dental caries and periodontal disease, respectively [74].

Videos combined peer narratives and demonstrations by older adults with expert guidance from a dental specialist. A standardized Microsoft PowerPoint presentation was also developed to align with video content.

Intervention and Comparator

Comparator Group (Face-to-Face Only)

Participants received a 15-minute individual session delivered by a trained general dentist. The session incorporated key SCT components including behavioral capability (clear instruction and live demonstrations of toothbrushing, flossing, and recognition of early oral disease signs), observational learning (dentist modeling and hands-on practice using a dental model and participants' own mouths), verbal persuasion (personalized encouragement), self-efficacy (personalized feedback), and outcome expectations (discussion of benefits of preventive behaviors and potential consequences of untreated oral conditions). Environmental facilitation was provided through a standardized Microsoft PowerPoint presentation, protected instructional time, and direct professional support.

Telehealth Education Group

In addition to the face-to-face session, participants received 4 educational videos sent via WhatsApp over 2 weeks (2 videos per week, approximately 5 minutes each, linked to YouTube). Participants were asked to watch the videos and informed that their opinions would be requested later; however, no specific viewing instructions were provided. Videos reinforced SCT components by offering environmental facilitation (on-demand access to educational content), behavioral capability (clear demonstrations of practical oral hygiene skills, explanations of early signs and symptoms of oral diseases, and guidance on preventive behaviors and risk factor management), observational learning (peer modeling—older adults performing target behaviors such as toothbrushing, flossing, seeking dental care, inspecting the oral mucosa for signs of disease, choosing healthy foods, maintaining hydration, and avoiding risk factors like smoking). The intervention also emphasized mastery experience (indirectly by encouraging participants to try the behaviors themselves after viewing the demonstrations), verbal persuasion (motivational messages and actionable recommendations from dental professionals), and outcome expectations (positive results

XSL•FO
RenderX

of maintaining oral health and the benefits associated with the recommended behaviors).

Procedures

Participants were enrolled and managed using the Geriatric Dental Specialties Tele-platform (TEGO), which supported data collection and clinical evaluations. The study followed four phases: (1) enrollment and baseline (pretest) assessment conducted via telephone, (2) clinical evaluation and face-to-face education at a mobile dental clinic by a trained general dentist, (3) telehealth educational intervention via WhatsApp for the intervention group, and (4) follow-up assessments conducted by telephone 6 weeks post intervention.

TEGO System

Overview

The TEGO is a web-based system supporting oral health assessment and education for older adults, integrating an electronic dental record and interactive 3D models. Its protocol and workflow are detailed in prior publications [64,75,76].

Module 1: Enrollment

Collects sociodemographic information, including age, sex, place of residence [77], socioeconomic status via the Social Registry of Households [78], educational level, and smartphone usage. It also serves as the entry point for administering questionnaires, detailed in the "Outcomes" section.

Module 2: Medical-Geriatric Anamnesis

Conducted by a dentist, assessing cognitive status (Mini-Mental State Examination), psychological well-being (Short Geriatric Depression Scale), functional capacity (Barthel Index and Fried's Frailty Phenotype), and medical conditions (eg, multimorbidity, defined as the presence of 2 or more comorbidities [79]). All instruments are validated in Spanish and commonly used in primary care [80-83].

Module 3: Odontogeriatric Anamnesis

Collects self-reported oral health behaviors (eg, toothbrushing, flossing, sugar consumption, and dental service use) and performs clinical evaluation using a standardized 3D dental model to record the number and condition of teeth [76].

Outcomes

Overview

Assessments were conducted at baseline and 6 weeks post intervention. Some secondary outcomes (eg, acceptability) were measured only at follow-up, while oral hygiene behaviors were first collected during the clinical evaluation and at follow-up. The independent variable was the type of educational intervention.

Primary Outcomes

The primary outcomes were changes in oral health knowledge, attitudes, and self-efficacy, measured using the 53-item Oral health Knowledge, Attitudes, and Self-Efficacy Questionnaire, previously used in older adults' populations [17,30,31,84]. The instrument includes 3 components:

1. Knowledge (39 items): assessed understanding of dental caries, periodontal disease, and oral cancer. Items included dichotomous (yes/no) questions on signs and symptoms, and a 5-point Likert scale rating the importance of risk and preventive factors ("Very important" to "Not important at all"). Correct answers scored 1 point (maximum=39 points; subscores: caries 12, periodontal disease 14, oral cancer 13).
2. Attitudes toward oral health (7 items): evaluated beliefs about the inevitability of oral diseases in older age, perceptions of prevention, and general oral health beliefs. Responses were recorded on a 4-point Likert scale ranging from "Strongly agree" to "Strongly disagree," with higher scores indicating more positive attitudes.
3. Oral Health Self-Efficacy Scale (OHSES): The original English version of the OHSES includes 8 items divided into 2 domains: oral health self-efficacy (6 items) and general health self-efficacy (2 items) [85]. The Spanish version of OHSES, validated in older Chilean adults [86], uses a 5-point Likert scale ("Not at all confident" to "Strongly confident"). It assesses the perceived ability to prevent tooth loss (3 items) and general self-care (4 items). The scale has shown high internal consistency (Cronbach α=0.82) and acceptable test-retest reliability (intraclass correlation coefficient=0.41) [86].

Change scores were compared between intervention groups and further stratified by place of residency (urban vs rural, based on national classification [77]).

Secondary Outcomes

1. Self-reported oral hygiene behaviors: frequency of toothbrushing and dental flossing.
2. Smartphone use: categorized into four increasing levels of complexity: (1) basic communication (calls and WhatsApp), (2) informative search (adds internet browsing), (3) extended social interaction (adds other social media), and (4) advanced use (adds email).
3. Acceptability of telehealth education (intervention group only): defined as the perceived appropriateness of the intervention [87], and assessed through: (1) burden: evaluated as the perceived effort required to participate, approximated through behavioral proxies: the proportion of eligible individuals who agreed to enroll (willingness to participate) and the dropout rate (number of withdrawals divided by the total randomized × 100). (2) Affective attitude: assessed participants' feelings about the intervention, measured via a dichotomous posttest question on willingness to recommend the video-based program. (3) Perceived effectiveness: this construct examines the extent to which the intervention is perceived to achieve its purpose and was measured through the perceived use of each video using a 5-point Likert scale ("Extremely useful" to "Not useful at all").

Harms

Given the noninvasive nature of the intervention, harms were defined as any adverse events or burdens related to participation in study procedures, including difficulties with digital content delivery, frustration associated with technology use, or

XSL•FO
RenderX

discomfort during questionnaire completion. No physical harm was expected.

Harms were assessed nonsystematically, through participants' reports during follow-up contact and at each assessment point. Study staff also monitored indications of discomfort or technical burdens during phone interactions.

Statistical Methods

The per-protocol analysis was established as the primary approach, including only participants who completed all study procedures. Noncompleters were excluded from this analysis.

Descriptive statistics were computed for sociodemographic variables, clinical data, self-reported behaviors, and questionnaire variables and are presented as percentages, means, SDs, and 95% CIs. The oral health knowledge, attitudes, and self-efficacy questionnaire was also examined by domain. Normality was assessed using the Anderson-Darling and D'Agostino tests. Within-group changes were assessed using 2-tailed paired *t* tests or Wilcoxon signed-rank tests. Between-group posttest comparisons were analyzed via independent *t* tests or Mann-Whitney *U* tests, and further examined using analysis of covariance (ANCOVA), adjusting for baseline scores. Effect sizes (Cohen *d* or *r*), 95% CIs, and *P* values are reported.

Three supplementary post hoc analyses were conducted:

1. Sensitivity analysis: an intention-to-treat (ITT) analysis was performed to preserve the integrity of randomization and minimize attrition bias. Missing data were handled using multiple imputation based on fully conditional specification, with linear regression models applied to scale variables. Five imputed datasets were generated with 10 iterations each, assuming data were missing at random. Both within-group and between-group ANCOVA analyses were conducted. A complementary Worst-Case Reasonable Scenario analysis assessed robustness under potential missing not at random (MNAR) mechanisms, assigning noncompleters in the telehealth group their baseline score (assuming no benefit) and comparator noncompleters the 75th percentile score of completers (assuming maximal benefit). This dual approach evaluates the stability of primary findings across different missing data assumptions.
2. Exploratory subgroup analysis: post hoc subgroup analysis was conducted by place of residence (urban/rural), introduced to enhance the interpretability of differential responses across settings. This includes the baseline percentage of correct responses by disease category and within-group analysis.
3. A post hoc multiple correspondence analysis and k-means clustering were performed within the telehealth group to explore associations between pre-post score changes and sociodemographic (age, educational level, place of residence, and smartphone use) and clinical variables (functional dentition, defined as ≥20 teeth vs <20 teeth). Score differences were categorized into 3 levels based on thresholds reported in previous studies [30,31,84]: significant improvement (≥50% increase), moderate improvement (20%-49% increase), and no change or deterioration (<20% improvement or decrease).

Acceptability of the telehealth intervention was assessed descriptively using percentages for perceived usefulness, willingness to recommend, and dropout rates.

Harms were assessed descriptively, including any difficulties related to digital engagement with digital content or survey procedures.

Analyses were performed using IBM SPSS Statistics (version 23.0) with supplementary visualizations in GraphPad Prism (version 8; GraphPad Software). A *P* value of <.05 was considered statistically significant.

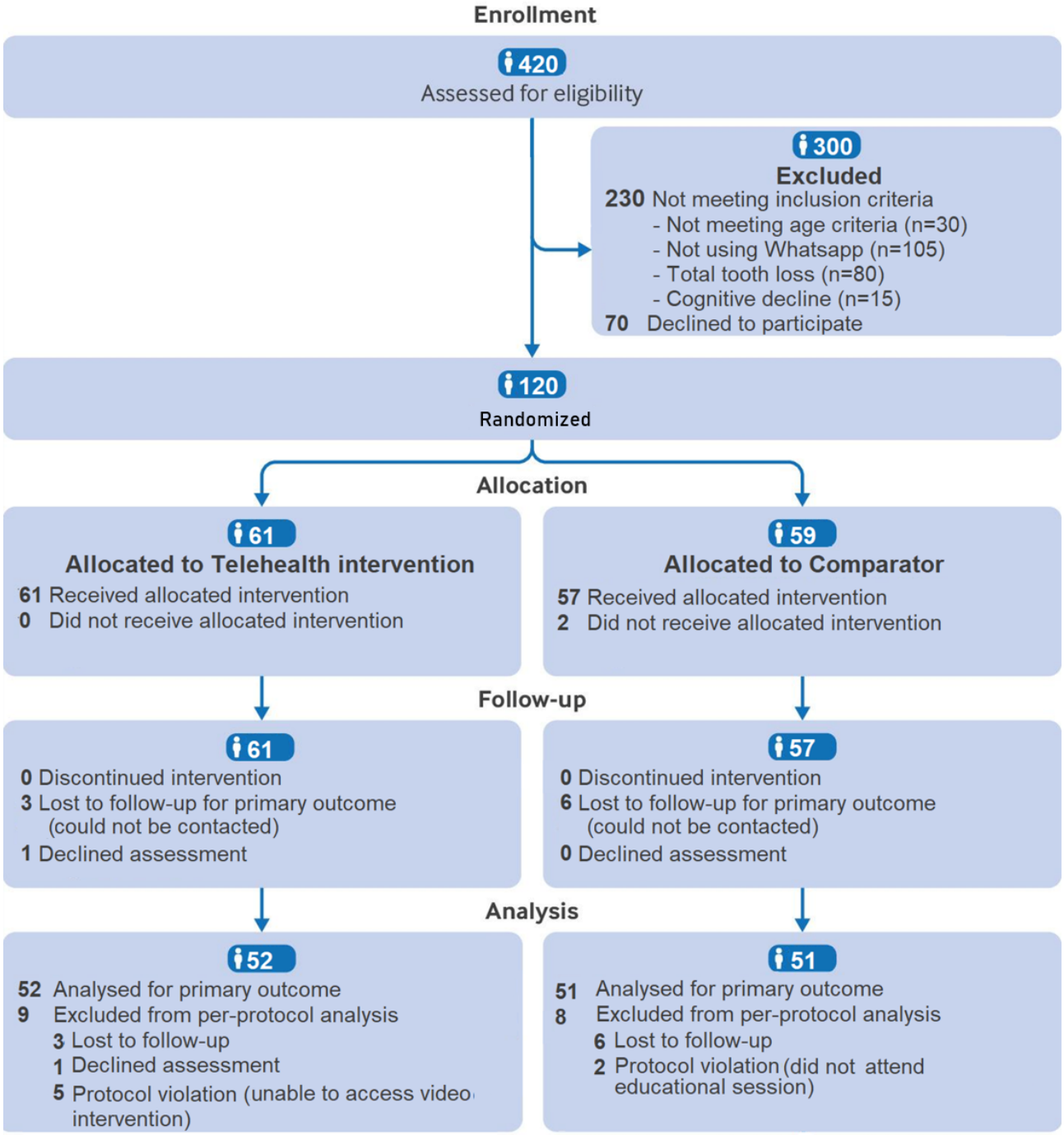
Results

Participant Flow and Recruitment

Between March 3 and August 4, 2023, a total of 120 participants were randomly assigned to the telehealth education group (n=61) or the comparator group (n=59; Figure 1). The last follow-up was completed on September 15, 2023. All participants provided informed consent prior to enrollment. Most participants received the intervention as intended; losses and reasons for exclusion after randomization are detailed in Figure 1. Both interventions were delivered as planned, with standardized procedures to ensure fidelity. Analyses followed a per-protocol approach and included only those who completed the study procedures (n=103), with 52 in the telehealth education group and 51 in the comparator group. No additional oral-health educational activities were provided during the trial period.

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) 2025 flow diagram of the study showing participant enrollment, group allocation, follow-up, and analysis.



Baseline Data

The median age of participants was 68 (IQR 63-73) years, and 64% (77/103) were female. Most had 9-12 years of education (54/103, 52%) and belonged to the most vulnerable socioeconomic category according to the Social Registry of Households classification (67/103, 65%). Overall, both groups had similar baseline characteristics (P>.05; Multimedia Appendix 2).

There were no significant baseline differences between those participants randomized and those included in the per-protocol analysis (P>.05).

Primary Outcomes

Within-Group Effects

Both the comparator and telehealth groups showed significant pre-post improvements across most oral health outcomes (Table 1).

XSL•FO
RenderX

Table 1. Baseline and postintervention scores by study group.

Variable	Comparator group (n=51)			Telehealth group (n=52)			*P* value[a]
	Pretest, mean (95% CI)	Posttest, mean (95% CI)	*P* value[b]	Pretest, mean (95% CI)	Posttest, mean (95% CI)	*P* value[b]	
Dental caries knowledge	9.6 (9.11-10.09)	10.1 (9.65-10.58)	.08	9.5 (8.96-10.03)	10.7 (10.29-11.12)	<.001	.046
Periodontal disease knowledge	8.6 (7.67-9.54)	10.2 (9.89-10.65)	.002	8.0 (6.91-9.08)	10.5 (10.11-10.99)	<.001	.31
Oral cancer knowledge	5.4 (4.25-6.56)	8.7 (8.04-9.48)	<.001	6.0 (4.87-7.12)	9.1 (8.35-9.91)	<.001	.59
Total knowledge	23.6 (21.59-25.65)	29.1 (27.93-30.37)	<.001	23.5 (21.30-25.69)	30.4 (29.12-31.68)	<.001	.14
Oral health attitudes	4.0 (3.59-4.44)	3.9 (3.55-4.36)	.80	3.9 (3.59-4.36)	4.9 (4.44-5.40)	.002	.002
OHSES[c]	5.2 (4.65-5.93)	6.7 (6.51-6.89)	<.001	5.0 (4.54-5.53)	6.5 (6.25-6.74)	<.001	.44

[a]Between-group comparisons at posttest (adjusted for baseline).

[b]Within-group comparisons (pre vs post).

[c]OHSES: Oral Health Self-Efficacy Scale.

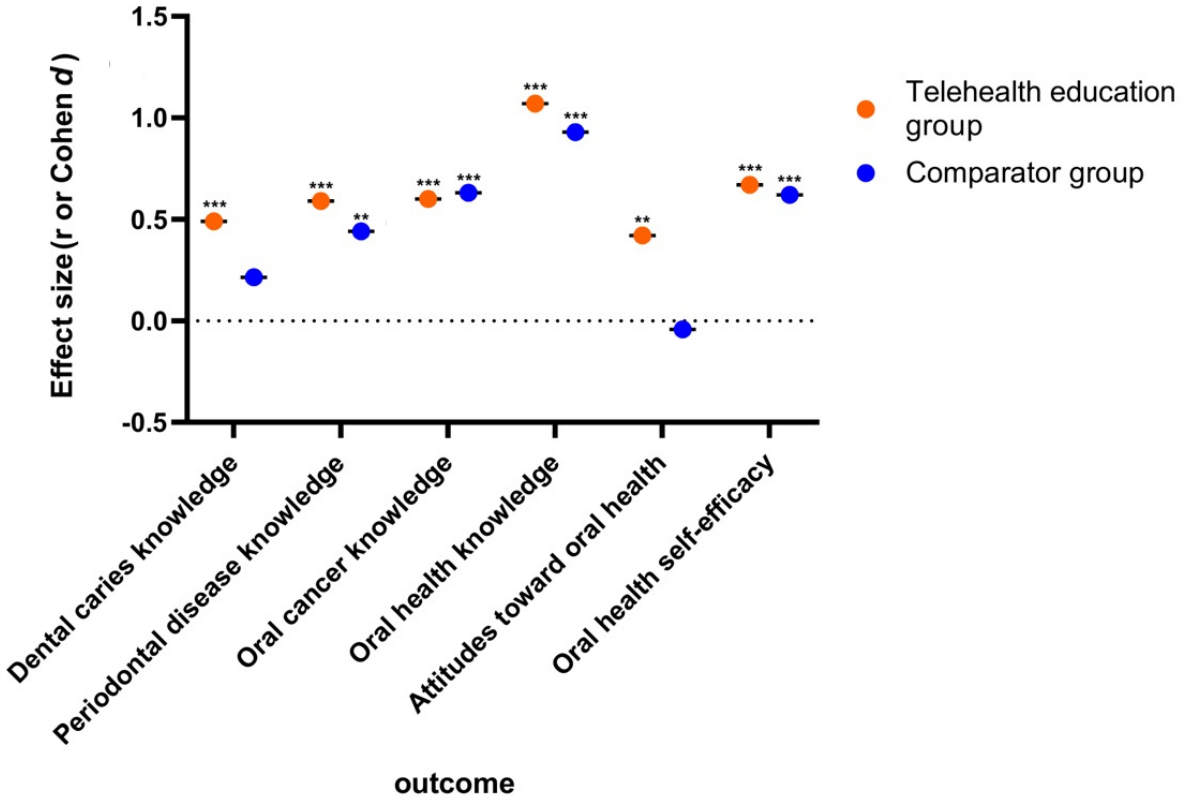
Participants in the comparator group (face-to-face education; n=51) showed large effects in oral health knowledge (Cohen *d*=0.93, 95% CI 0.52-1.34; *P*<.001), oral cancer knowledge (*r*=0.63, 95% CI 0.36-0.90; *P*<.001), oral health self-efficacy (*r*=0.59, 95% CI 0.38-0.74; *P*<.001), and toothbrushing behavior (*r*=0.51, 95% CI 0.24-0.79; *P*<.001). Additionally, it had a moderate effect on periodontal disease knowledge (*r*=0.44, 95% CI 0.17- 0.71; *P*=.002). However, no significant changes were observed in attitudes (*P*=.76) or dental caries knowledge (*P*=.12).

In the telehealth education group (n=52), which combined face-to-face education with telehealth, improvements were even more marked across multiple domains. This group showed a large effect on dental caries knowledge (*r*=0.49, 95% CI 0.23-0.77; *P*<.001), attitudes toward oral health (*r*=0.42, 95% CI 0.14-0.69; *P*=.003), and periodontal disease knowledge (*r*=0.59, 95% CI 0.31-0.86; *P*<.001). Similar to the comparator group, the telehealth education group also exhibited large effects in oral health knowledge (Cohen *d*=1.07, 95% CI 0.66-1.48; *P*<.001), oral health self-efficacy (*r*=0.62, 95% CI 0.43-0.77; *P*<.001), toothbrushing behavior (*r*=0.54, 95% CI 0.26-0.81; *P*<.001), and oral cancer knowledge (*r*=0.60, 95% CI 0.33-0.87; *P*<.001).

Figure 2 visualizes the magnitude of effects across groups and outcomes, with asterisks indicating significance levels.

XSL•FO
RenderX

Figure 2. Effect sizes for preintervention to postintervention differences in oral health knowledge, attitudes toward oral health, and oral health self-efficacy, stratified by intervention group. **P<.01, ***P<.001.



Detailed Changes in Knowledge Subdomains

Total Knowledge Score

Total knowledge scores were comparable at baseline (*P*=.93; Table 1). Following the intervention, both groups demonstrated substantial within-group improvements: the comparator group increased by 5.5 (95% CI 3.37-7.68) points (*P*<.001), and the telehealth group improved by 6.9 (95% CI 4.70-9.10) points (*P*<.001). Notably, no participant achieved the maximum possible score of 39 points.

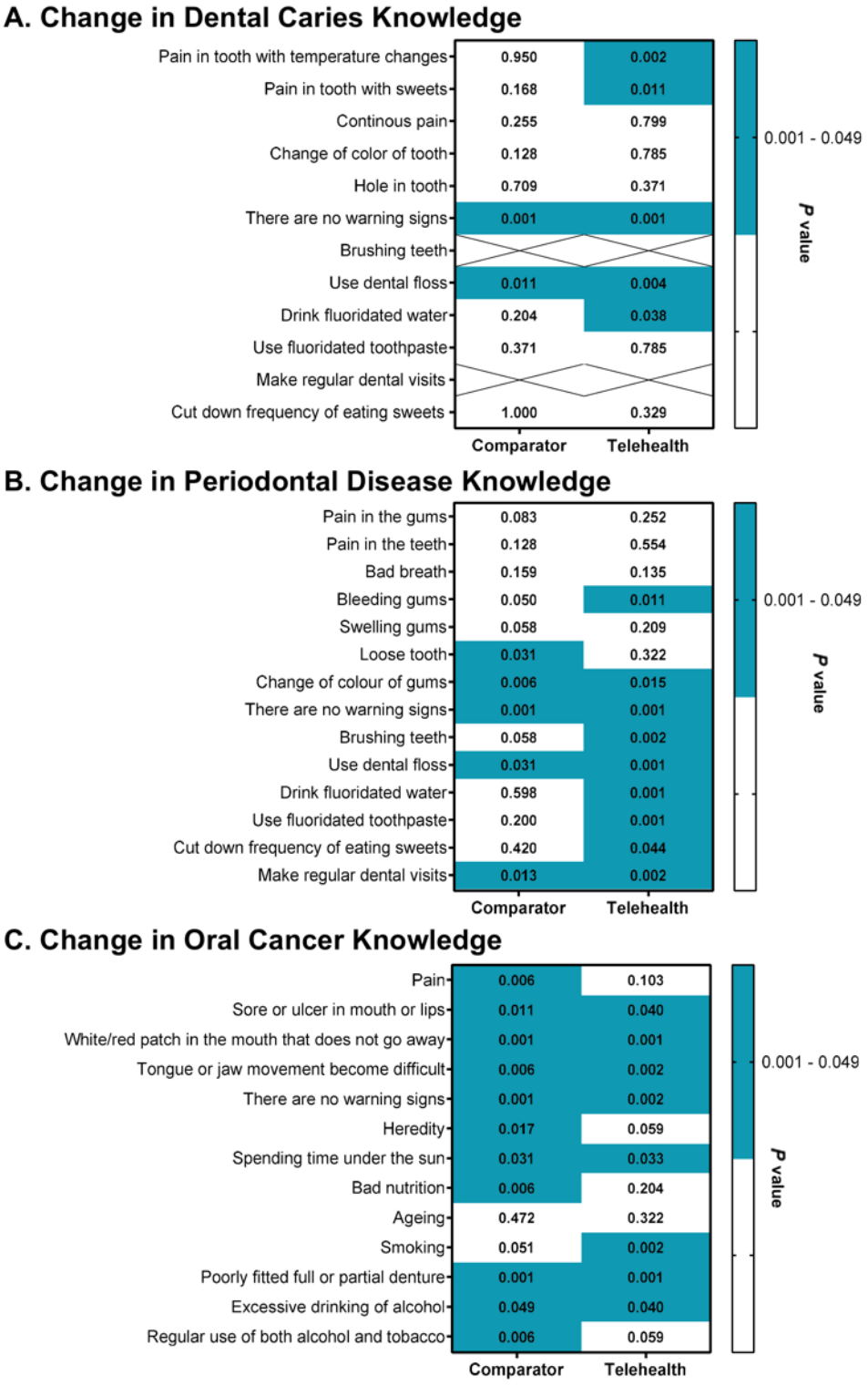
Dental Caries Knowledge

Baseline knowledge of caries signs and symptoms was high across the sample (mean score 4.4 out of 6, SD 1.38), with no significant differences between groups (*P*=.12). The most frequently recognized signs included the presence of a cavity (92/103, 89%), and persistent pain (85/103, 82%), while tooth pain when consuming sweets was the least recognized (68/103, 66%).

Similarly, awareness of preventive measures was high overall (mean score 5.1 out of 6, SD 0.9), with no significant differences between intervention groups (comparator: mean 5.0, 95% CI 4.73-5.26; telehealth: mean 5.30, 95% CI 5.05-5.55; *P*=.09). However, the largest knowledge deficits were concentrated in 2 specific preventive measures, fluoride in water (61/103, 59%) and flossing (77/103, 79%). After the intervention, both items showed significant increases in the telehealth education group, reaching 83% (43/52; *P*=.03) and 100% (52/52; *P*=.004), respectively (Figure 3).

XSL•FO
RenderX

Figure 3. Preintervention to postintervention changes in participants' knowledge related to dental caries, periodontal disease, and oral cancer, stratified by group. Statistically significant improvements are highlighted in blue.



Periodontal Disease Knowledge

At baseline, the mean score for identifying signs and symptoms of periodontal disease was 5.5 (SD 2.6) out of 8 points. Scores were similar between groups (comparator: mean 5.6, 95% CI 4.99-6.37; telehealth: mean 5.5, 95% CI 4.70-6.33; *P*=.75). Recognition was below 80% for all signs and symptoms, except for gum inflammation (87/103, 85%). Notably, approximately 70% (72/103) of the overall sample believed that periodontal disease has no signs or symptoms.

Baseline recognition of preventive measures was generally low, mean 2.7 (SD 1.2) out of 5 points, without differences across groups (comparator: mean 2.9, 95% CI 2.54-3.29; telehealth: mean 2.48, 95% CI 2.15-2.80; *P*=.07). Most participants (77%-95%), held misconceptions about the role of fluoride in preventing periodontal disease, and these misconceptions

XSL•FO
RenderX

persisted even after the intervention. However, the telehealth group demonstrated partial improvement, with a significant decrease specifically in the misconception associating fluoride in toothpaste with periodontitis prevention (*P*<.001; Figure 3).

Following the intervention, participants showed significant improvements in identifying signs, symptoms, and preventive measures (*P*<.001). These gains were particularly evident in the telehealth education group, which demonstrated significantly better recognition of preventive measures (comparator: mean 3.3, 95% CI 3.15-3.63; telehealth: mean 3.7, 95% CI 3.52-4.04; *P*=.01).

Oral Cancer Knowledge

Baseline knowledge of oral cancer signs, symptoms, and risk factors was low across the sample, with mean scores of 1.7 (SD 1.7) out of 5 and 3.8 (SD 2.8) out of 8, respectively. Overall, 69% (71/103) of participants were unable to identify signs or symptoms, and 24% (25/103) were unaware of risk factors. The most frequently recognized sign was a persistent oral ulcer or wound (53/103, 51%), and smoking was the most identified risk factor (74/103, 72%).

At pretest, there were no significant differences between groups in the identification of signs and symptoms (comparator: mean 1.8, 95% CI 1.29-2.31; telehealth: mean 1.7, 95% CI 1.26-2.27; *P*=.92) or in awareness of risk factors (comparator: mean 3.6, 95% CI 2.86-4.34; telehealth: mean 4.2, 95% CI 3.44-5.01; *P*=.24).

Following the intervention, both groups demonstrated a significant increase in the identification of signs, symptoms, and risk factors *(P*<.001). Posttest comparisons showed no significant between-group differences in identification (comparator: mean 3.5, 95% CI 3.15-3.90; telehealth: mean 3.3, 95% CI 2.98-3.78; *P*=.60) or in risk-factor awareness (comparator: mean 5.2, 95% CI 4.72-5.74; telehealth: mean 5.5, 95% CI 5.24-6.25; *P*=.057). Across the sample, 90% (93/103) of participants could identify at least 2 signs or symptoms of oral cancer, and 96% (99/103) recognized at least 2 risk factors. Nearly all individual indicators showed statistically significant improvement (*P*<.05; Figure 3).

Attitudes Toward Oral Health

No statistically significant differences in baseline attitude scores were observed between groups (comparator: mean 4.0, 95% CI 3.59-4.44; telehealth: mean 3.9, 95% CI 3.59-4.36; *P*=.89). Following the intervention, only the telehealth group demonstrated statistically significant improvement (mean difference 0.9, 95% CI 0.37-1.51, *P*=.002; Table 1).

Notable changes were observed in the reduction of agreement with negative attitudes, with all improvements favoring the telehealth education group. Specifically, telehealth participants were less likely to agree with statements such as “As people become older, they can expect to lose their teeth” (*P*=.01), “Only people with teeth need to visit the dentist” (*P*=.01), and “Older adults must learn to live with the condition of their teeth and gums” (*P*=.02).

However, a highly prevalent negative baseline attitude—the belief that “Having dental problems is a normal part of aging” (87/103, 84%)—did not significantly change following the intervention in either group (*P*=.18).

Oral Health Self-Efficacy

At baseline, OHSES scores were comparable between groups (comparator: mean 5.2, 95% CI 4.65-5.93; telehealth: mean 5.0, 95% CI 4.54-5.53; *P*=.52). Most items were rated positively (>50%), except confidence in using dental floss (49/103, 47%), which was significantly lower in the telehealth group (*P*=.03)

Following the intervention, total OHSES scores increased significantly (*P*<.001) across all individual items (*P*<.01) and intervention groups (*P*<.001; Table 1). No between-group posttest differences were detected (*P*=.44). The proportion of participants achieving the maximum score rose from 32% (33/103) to 75.7% (78/103), and all items exceeded 80% at follow-up.

Between-Group Comparisons

When comparing the postintervention scores adjusted for baseline (Table 1), the telehealth group was significantly superior to the comparator group in 2 primary outcomes. Specifically, at 6 weeks, participants receiving the telehealth intervention (52/103) were significantly more likely than those in the comparator group (51/103) to demonstrate improved knowledge about dental caries (*r*=0.21, 95% CI 0.11-0.3; *P*=.03) and more positive attitudes toward oral health (*r*=0.28, 95% CI 0.18-0.38; *P*=.004), corresponding to small and moderate effect sizes, respectively. For all other primary outcomes, no significant between-group differences were detected at the posttest assessment (*P*>.05; Table 1).

Secondary Outcomes

Sensitivity Analysis: ITT

The ITT analysis was conducted to assess the effect of the intervention on the full randomized sample (n=120). Missing data were handled using multiple imputation, where all prespecified sociodemographic, clinical, pretest, posttest, and group assignment variables were included in the imputed model.

Between-group analysis showed that the effect on posttest dental caries knowledge was no longer statistically significant (B=0.60, SE 0.32, 95% CI -0.03-1.23; *P*=.06), whereas the telehealth advantage for oral health attitudes remained significant (B=0.91, SE 0.29, 95% CI 0.34-1.48; *P*=.002; Table 2).

XSL•FO
RenderX

Table 2. Intention-to-treat analysis results by group using multiple imputation.

Variable	Comparator group (n=59)			Telehealth group (n=61)			*P* value[a]
	Pretest, mean (95% CI)	Posttest, mean (95% CI)	*P* value[b]	Pretest, mean (95% CI)	Posttest, mean (95% CI)	*P* value[b]	
Dental caries knowledge	9.6 (9.11-10.09)	10.1 (9.66-10.60)	.05	9.5 (8.96-10.03)	10.7 (10.33-11.12)	<.001	.06
Periodontal disease knowledge	8.6 (7.67-9.54)	10.2 (9.90-10.66)	<.001	8.0 (6.91-9.08)	10.5 (10.16-10.98)	<.001	.30
Oral cancer knowledge	5.4 (4.25-6.56)	9.0 (8.29-9.75)	<.001	6.0 (4.87-7.12)	9.1 (8.42-9.81)	<.001	.75
Total knowledge	23.6 (21.59-25.65)	29.3 (27.95-30.69)	<.001	23.5 (21.30-25.69)	30.3 (28.96-31.66)	<.001	.11
Oral health attitudes	4.0 (3.59-4.44)	3.9 (3.58-4.38)	.87	3.9 (3.59-4.36)	4.8 (4.45-5.33)	<.001	.002
OHSES[c]	5.2 (4.65-5.93)	6.6 (6.45-6.92)	<.001	5.0 (4.54-5.53)	6.4 (6.27-6.70)	<.001	.59

[a]Between-group comparisons at posttest (adjusted for baseline).

[b]Within-group comparisons (pre vs post).

[c]OHSES: Oral Health Self-Efficacy Scale.

To assess the stability of the primary conclusion regarding attitudes against potential MNAR mechanisms, a Worst-Case Reasonable Scenario sensitivity analysis was conducted. Between-group difference in oral health attitudes remained significant under the MNAR mechanisms assumption (*P*=.02). This finding confirms the stability of the telehealth intervention's effect on attitudes, despite potential attrition bias.

Subgroup Analyses by Place of Residence

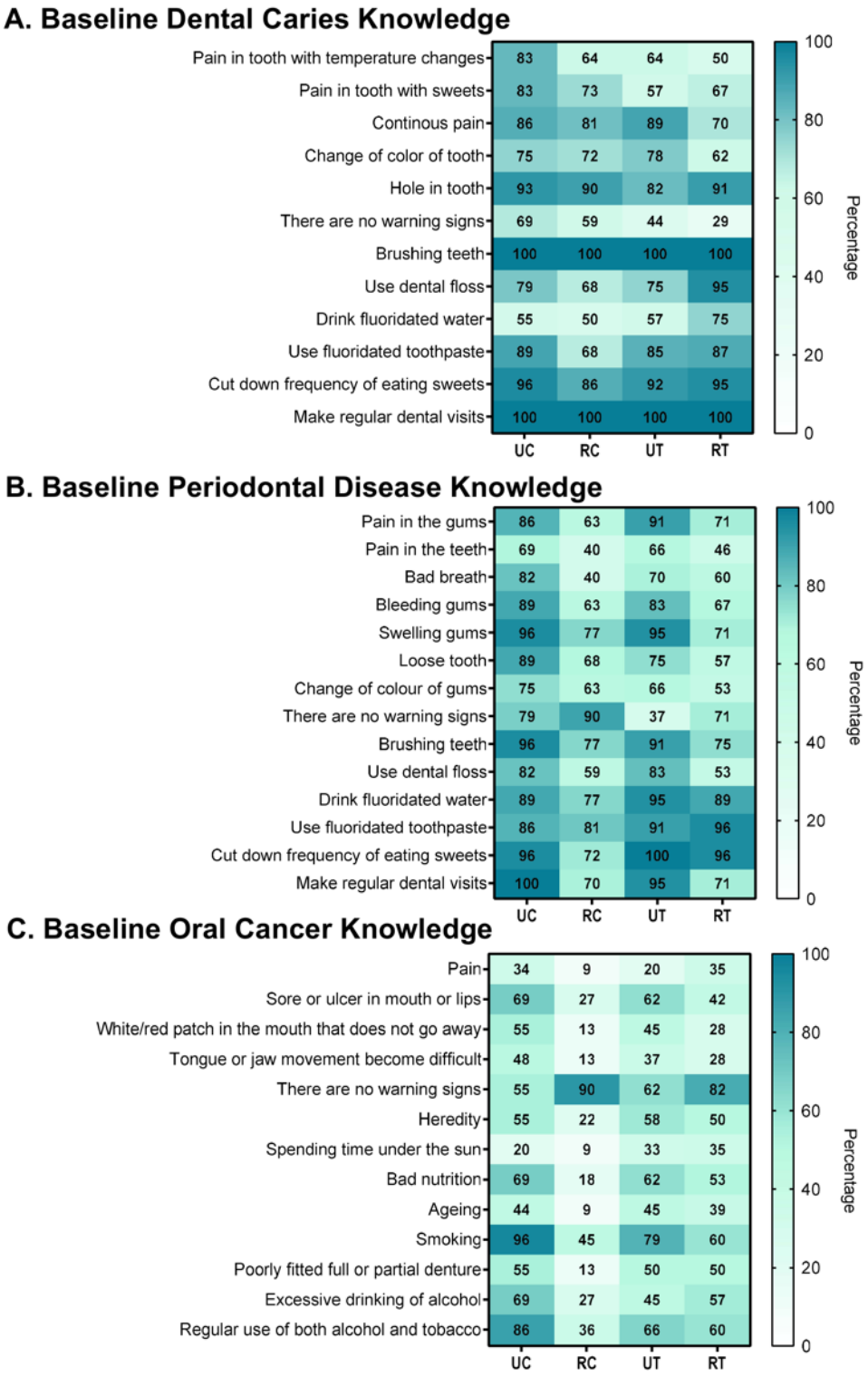
Secondary exploratory analyses examined response patterns by place of residence (rural vs urban). Rural participants accounted for 50% (60/120) of the randomized sample, with no significant differences between intervention groups (*P*=.85) or in the final analyzed sample (*P*=.27).

At baseline, rural participants showed greater vulnerability, including differences in smartphone use (*P*=.002), higher multimorbidity (*P*=.01), and fewer number of teeth present (rural: mean 16.24, SD 7.15; urban: mean 19.20, SD 7.16; *P*=.03), and consistently lower knowledge across all domains (all *P*<.05; Multimedia Appendix 3).

Urban participants demonstrated significantly higher baseline knowledge of dental caries (*P*=.04), periodontal disease (*P*<.001), and oral cancer knowledge (*P*<.001). Item-level percentages illustrating these baseline disparities are presented in Figure 4, while Table 3 summarizes the corresponding pre-post differences by group and place of residence. At baseline, a substantial proportion of rural participants were unable to identify key signs of periodontal disease (12/50, 24% rural vs 1/53, 1.8% urban; *P*=.01) or oral cancer; specifically 86% (43/50) of rural participants could not identify any signs or symptoms of oral cancer, and 40% (20/50) were unaware of risk factors, compared with 9% (5/53) among urban participants (*P*=.003).

XSL•FO
RenderX

Figure 4. Heatmap of baseline oral health knowledge related to dental caries, periodontal disease, and oral cancer, expressed as the percentage of participants responding "yes" to each question. Results are stratified by group and place of residence. The color gradient ranges from white (lower percentage) to blue (higher percentage) of affirmative responses. RC: rural comparator; RT: rural telehealth education; UC: urban comparator; UT: urban telehealth education.



XSL•FO
RenderX

Table 3. Comparison of scores in oral health knowledge, attitudes, self-efficacy, and self-reported oral hygiene behaviors before and after intervention, by group and place of residence.

Variable, group, and place of residence	Pretest	Posttest	*P* value
Dental caries knowledge, mean (95% CI)			
Telehealth education			
Rural (n=28)	9.3 (8.53-10.17)	10.9 (10.28-11.50)	<.001[a]
Urban (n=24)	9.7 (8.93-10.39)	10.5 (9.91-11.08)	.08[a]
Comparator			
Rural (n=22)	8.9 (8.03-9.87)	10.0 (9.31-10.77)	.04[a]
Urban (n=29)	10.1 (9.62-10.58)	10.2 (9.53-10.80)	.85[a]
Periodontal disease knowledge, mean (95% CI)			
Telehealth education			
Rural (n=28)	7.1 (5.38-8.97)	10.8 (10.17-11.46)	<.001[a]
Urban (n=24)	8.9 (7.86-10.04)	10.2 (9.63-10.86)	.02[a]
Comparator			
Rural (n=22)	7.1 (5.19-8.98)	10.5 (9.86-11.13)	.004[a]
Urban (n=29)	9.7 (9.12-10.39)	10.1 (9.62-10.58)	.20[a]
Oral cancer knowledge, mean (95% CI)			
Telehealth education			
Rural (n=28)	5.6 (3.96-7.24)	10.2 (9.12-11.23)	<.001[a]
Urban (n=24)	6.4 (4.83-8.08)	7.9 (6.89-8.93)	.19[a]
Comparator			
Rural (n=22)	2.5 (1.02-4.06)	8.2 (6.95-9.50)	<.001[a]
Urban (n=29)	7.5 (6.38-8.78)	9.2 (8.30-10.04)	.02[a]
Total knowledge, mean (95% CI)			
Telehealth education			
Rural (n=28)	22.1 (18.69-25.59)	31.9 (30.01-33.77)	<.001[b]
Urban (n=24)	25.1 (22.40-27.75)	28.7 (27.12-30.20)	.02[b]
Comparator			
Rural (n=22)	18.5 (15.46-21.71)	28.8 (26.87-30.67)	<.001[b]
Urban (n=29)	27.4 (25.70-29.18)	29.4 (27.76-31.13)	.05[b]
Oral health attitudes, mean (95% CI)			
Telehealth education			
Rural (n=28)	3.8 (3.31-4.40)	5.2 (4.53-5.82)	.004[a]
Urban (n=24)	4.1 (3.55-4.69)	4.6 (3.87-5.37)	.25[a]
Comparator			
Rural (n=22)	3.7 (2.97-4.38)	3.6 (3.16-4.01)	.79[a]
Urban (n=29)	4.3 (3.73-4.81)	4.2 (3.60-4.88)	.85[a]
OHSES [c], mean (95% CI)			
Telehealth education			

XSL•FO
RenderX

Variable, group, and place of residence	Pretest	Posttest	*P* value
Rural (n=28)	5.3 (4.80-5.83)	6.5 (6.24-6.89)	<.001[a]
Urban (n=24)	4.7 (3.78-5.62)	6.4 (6.00-6.82)	.002[a]
Comparator			
Rural (n=22)	5.1 (4.01-6.25)	6.8 (6.65-7.07)	.004
Urban (n=29)	5.4 (4.62-6.20)	6.5 (6.28-6.88)	.002[a]
Toothbrushing frequency, mean (95% CI)			
Telehealth education			
Rural (n=28)	2.3 (2.06-2.58)	2.9 (2.69-3.16)	<.001[a]
Urban (n=24)	2.5 (2.08-2.91)	2.9 (2.56-3.18)	.03[a]
Comparator			
Rural (n=22)	2.5 (2.24-2.84)	3.0 (2.69-3.39)	.008[a]
Urban (n=29)	2.6 (2.36-2.94)	3.0 (2.64-3.42)	.01[a]
Flossing, n (%)			
Telehealth education			
Rural (n=28)	8 (29)	21 (75)	<.001[d]
Urban (n=24)	12 (50)	15 (62)	.37[d]
Comparator			
Rural (n=22)	8 (36)	11 (50)	.45[d]
Urban (n=29)	10 (34)	17 (59)	.06[d]

[a]Wilcoxon signed-rank test.

[b]Paired *t* test.

[c]OHSES: Oral Health Self-Efficacy Scale.

[d]McNemar test.

Across both study arms, rural participants exhibited larger within-group pre-post improvements than urban participants. In the telehealth group, rural participants showed significant gains across all knowledge domains (all *P*<.001), OHSES (*P*<.001), and were the only subgroup to show significant improvements in attitudes (*P*=.004) and flossing (*P*<.001). Urban participants improved in fewer outcomes (Table 3). A similar pattern was observed in the comparator group, where rural participants improved across all knowledge components and OHSES, whereas several pre-post comparisons among urban participants were nonsignificant. Overall, improvements were more consistent and pronounced among rural subgroups, particularly in the telehealth arm. However, these findings are exploratory and reflect within-group changes only; limited sample sizes across the 4 intervention-residence subgroups do not allow for formal comparative analyses.

Self-Reported Oral Hygiene Behaviors

At baseline, nearly all participants reported brushing their teeth regularly (102/103, 99%), with 54% (56/103) brushing once or twice daily. At baseline, there were no significant differences between groups (comparator: mean 2.6, 95% CI 2.40-2.81; telehealth: mean 2.4, 95% CI 2.17-2.63; *P*=.18). Following the intervention, 75% (77/103) reported brushing 3 or more times per day, with significant within-group increases (both *P*<.001). However, baseline-adjusted posttest comparisons showed no significant differences between groups (comparator: mean 3.0, 95% CI 2.78-3.29; telehealth: mean 2.9, 95% CI 2.71-3.08; *P*=.90). The change in toothbrushing frequency showed a weak yet significant correlation with perceived self-efficacy in thoroughly brushing all tooth surfaces (*r*=0.196; *P*=.04).

At baseline, 63% (65/103) of the participants reported not using dental floss, with no significant differences between groups (*P*=.73). Pretest flossing behavior was significantly associated with baseline attitudes toward oral health (*P*=.01), oral health self-efficacy (*P*=.04), educational level (*P*=.007), and number of teeth present (*P*=.01). A moderate positive correlation was found between baseline flossing behavior and baseline flossing self-efficacy (*r*=0.41; *P*<.001), while postintervention flossing demonstrated a weaker yet significant correlation (*r*=0.29; *P*=.003).

The frequency of self-reported flossing significantly increased following the intervention (38/103, 36.9% vs 64/103, 62.1%; *P*<.001). Participants in the telehealth group showed a higher likelihood of reporting flossing compared to those in the comparator group (risk ratio=1.31, 95% CI 0.96-1.78), corresponding to an absolute increase in flossing prevalence of

XSL•FO
RenderX

16.7% (risk difference=0.16, 95% CI –0.02 to 0.35); however, this between-group difference was not statistically significant (*P*=.08).

Smartphone Usage and Telehealth Education Acceptability

Among those assessed for eligibility, 25% (105/420) reported not using WhatsApp (Figure 1), with this pattern more pronounced in rural areas. No significant differences in WhatsApp use were found between study groups (*P*=.86). However, rural residents were significantly less likely to use online search tools (*P*=.04), other social media apps (*P*=.04), or email (*P*<.001). Advanced smartphone use—defined as use of WhatsApp, online search, social media, and email—was reported by 45% (24/53) of urban participants (Multimedia Appendix 3). Limited smartphone users (WhatsApp only) were significantly more likely to have ≤8 years of formal education and live in rural areas (*P*=.008).

Of the eligible participants, 63.1% (120/190) agreed to enroll in the intervention (willingness to participate; Figure 1). Overall, the study experienced a 14.2% (17/120) attrition rate. At 6 weeks, dropout rates did not differ significantly between the telehealth education group (9/61, 15%) and the comparator group (8/59, 13%; *P*=.85).

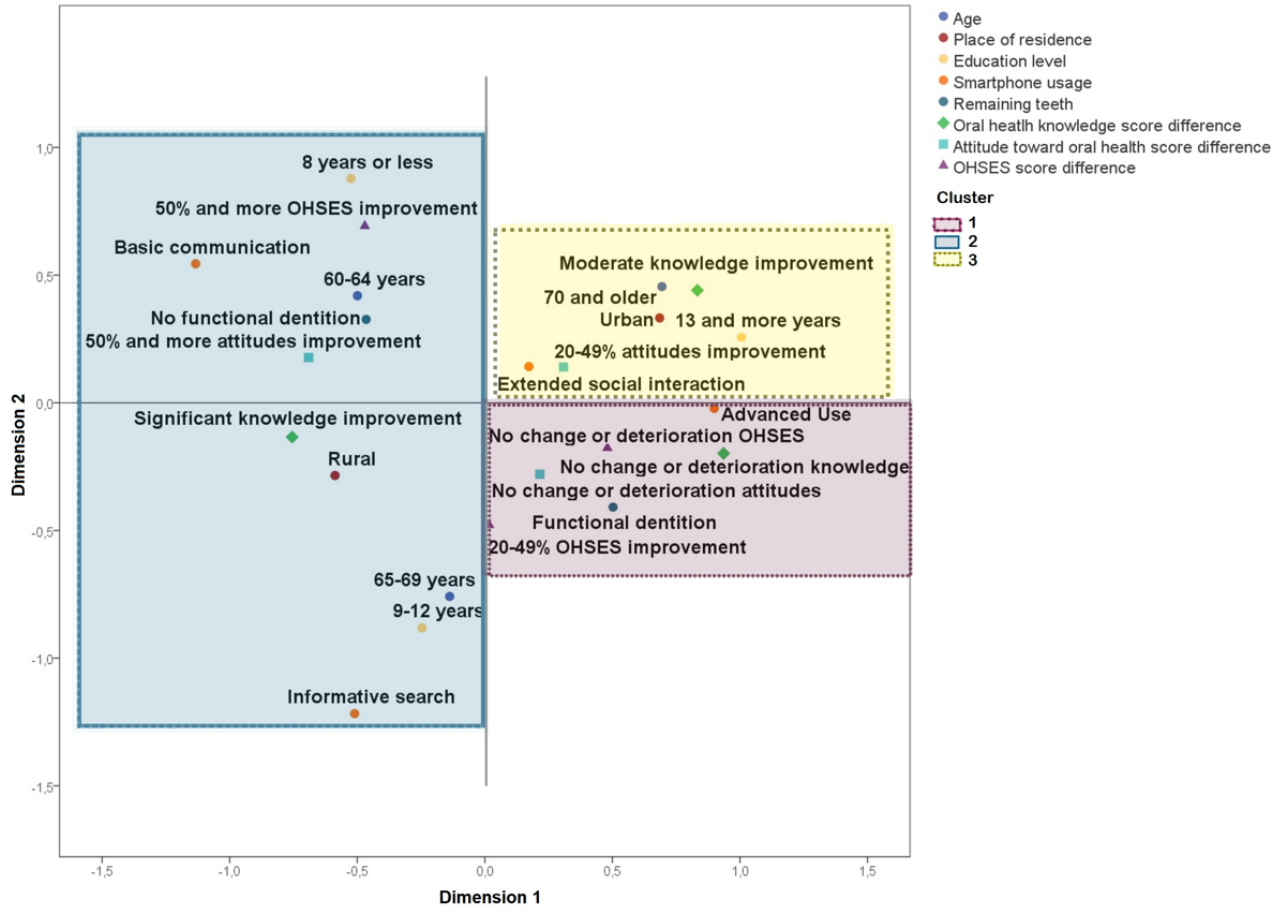
Most participants of the telehealth education group (48/52, 92%) stated they would recommend the use of telehealth education videos, indicating a positive attitude toward the intervention. Recommendation rates did not significantly differ by place of residence (*P*=.84), educational level (*P*=.29), or smartphone usage (*P*=.27). Perceived effectiveness of the instructional videos was high, with all participants finding the videos on dental hygiene, prevention of dental caries, and oral cancer prevention useful (52/52, 100%), and 90% (47/52) finding the periodontal disease prevention video useful.

Cluster Analysis of Demographic- and Intervention-Related Variables in the Telehealth Education Group

The analysis identified 2 main dimensions, explaining 56.5% of the total variance. Dimension 1 primarily captured the relationships between smartphone use, place of residence, and changes in knowledge and attitude scores, while dimension 2 reflected educational level and changes in oral health self-efficacy.

Subsequent k-means clustering identified 3 distinct participant profiles (Figure 5):

Figure 5. Subgroup profiles within the telehealth education group were identified using multiple correspondence analysis followed by k-means clustering. Each cluster (cluster 1: pink, square-dotted line; cluster 2: blue, solid line; cluster 3: yellow, round-dotted line) corresponds to distinct patterns of sociodemographic and clinical characteristics, including age, place of residence, education level, smartphone use, and number of remaining teeth. These clusters show differential improvements in oral health knowledge, attitudes, and Oral Health Self-Efficacy Scale scores.



XSL•FO
RenderX

- Cluster 1: participants who derived little to no benefit from the intervention. This cluster included individuals with functional dentition (12/22, 86%; χ^2_2=11.65; *P*=.003; post hoc Dwass-Steel-Critchlow-Fligner: C1>C2,*P*=.002; C1>C3, *P*=.04), who showed minimal or negative changes in knowledge, attitudes, and oral health self-efficacy.
- Cluster 2: participants who benefited the most, with most members showing high improvement (≥50%) in oral health knowledge (20/22, 91%; χ^2_2=23.24; *P*<.001; post hoc Dwass-Steel-Critchlow-Fligner: C2>C1, *P*=.01; C2>C3, *P*<.001), particularly in oral cancer knowledge (13/22, 59%; χ^2_2=17.56; *P*<.001; post hoc Dwass-Steel-Critchlow-Fligner: C2>C1, *P*=.01; C2>C3, *P*=.04), and in self-efficacy (12/22, 54%;χ^2_2=9.33; *P*=.009; post hoc Dwass-Steel-Critchlow-Fligner: C2>C1, *P*=.01; C2>C3, *P*=.04) This cluster was characterized by rural residence (16/22, 73%; χ^2_2=8.57; *P*=.01; post hoc Dwass-Steel-Critchlow-Fligner: C2>C3, *P*=.02), and limited smartphone use (11/22, 50%; χ^2_2=15.60; *P*<.001; post hoc Dwass-Steel-Critchlow-Fligner: C2>C3, *P*<.001).
- Cluster 3: participants with moderate improvements. This group consisted mainly of urban adults (12/16, 75%; *P*=.01) aged ≥70 years (13/22, 81.3%; χ^2_2=0.25.00; *P*<.001; post hoc Dwass-Steel-Critchlow-Fligner: C3>C1, *P*<.001; C3>C2, *P*<.001) and smartphone use oriented toward extended social interaction (9/16, 56.3%; *P*<.001).

Harms and Unintended Events

No clinical harms or adverse health-related events were observed. However, one unintended operational issue arose in the telehealth group: 5 (8.2%) participants of the 61 assigned to telehealth—despite initially reporting confidence using WhatsApp—were unable to access or view the educational videos, 4 of whom lived in rural areas. This usability-related limitation led to their exclusion from the per-protocol analysis but did not pose any health risks.

Discussion

Principal Findings

This randomized trial tested a WhatsApp-supported telehealth reinforcement strategy among community-dwelling older adults in the La Araucanía Region of Chile. Both study arms demonstrated significant improvements in oral health knowledge and self-efficacy. Between-group differences favored telehealth only for attitudes toward oral health and dental caries knowledge in the primary per-protocol analysis, with attitudes showing the strongest effect. No significant differences between groups were identified for total knowledge, periodontal disease recognition, or oral cancer knowledge. These findings underscore the potential of mHealth interventions in addressing disparities in preventive care, aligning with World Health Organization recommendations [88].

Intention-to-treat sensitivity analysis attenuated the between-group difference in dental caries knowledge, while superiority for attitudes remained. This outcome suggests that the added benefit of telehealth reinforcement, while present, was modest and more evident among participants with full adherence to the intervention. A plausible explanation is the high effectiveness of the face-to-face education delivered to the comparator group, which produced substantial baseline improvements, thereby reducing the margin for further short-term gains. Similar patterns have been reported in prior education trials, where reinforcement strategies yielded clearer benefits over longer follow-up periods rather than immediately post intervention [15].

From an SCT perspective, learning and behavior change are supported by reinforcement and repeated opportunities to observe, rehearse, and apply learned behaviors in supportive environments [22,25]. While face-to-face education appeared sufficient for most knowledge domains, educational research indicates that single-session interventions are vulnerable to declining retention over time in the absence of reinforcement [89,90]. In this study, telehealth operated as a structured extension of the in-person education, allowing repeated exposure to key concepts and practical application through a familiar digital platform—an approach aligned with older adults' preference for reliable postconsultation resources [46,52,53]. Although absolute knowledge gains were slightly larger in the telehealth group, no statistically significant between-group differences were observed. Notably, the magnitude of improvement observed in the telehealth arm falls at the upper range reported in prior digital interventions for older adults (3.1-5.4 points) [31,45,91], although comparisons should be interpreted cautiously given heterogeneity in study designs and outcome measures. Taken together, this combined mechanism may help explain the between-group gains observed in specific knowledge domains—particularly dental caries and periodontal prevention—among fully adherent participants.

Beyond knowledge acquisition, effective preventive behavior change requires additional mechanisms, such as self-efficacy [92]. Prior research suggests that gains in knowledge can positively influence attitudes and self-efficacy, supporting the adoption of healthier behaviors, and, over time, improvements in oral-health related quality of life [16]. Both groups demonstrated substantial improvements in self-efficacy, which were positively associated with self-reported improvements in hygiene behaviors. Exploratory subgroup analysis further suggested that rural participants receiving telehealth reinforcement experienced additional behavioral gains, specifically increased flossing. Although exploratory, this pattern is consistent with SCT-based mechanisms—such as observational learning (video-based demonstrations), verbal persuasion (via simple and supportive messaging), environmental facilitation (home-based, flexible access to content), and repetition for mastery (allowing participants to review and practice skills introduced during the face-to-face session) [22,25]. Together, these elements may have strengthened self-regulation, reinforced self-efficacy, and supported the internalization of health recommendations, thereby promoting skill development and behavior adoption [21].

The improvement in prevention-oriented attitudes may reflect changes in outcome expectancies, a core SCT construct referring to beliefs about the anticipated benefits of a behavior [22,25].

XSL•FO
RenderX

Repeated, self-paced engagement with preventive content may have strengthened participants' expectations that preventive oral health behaviors are effective and worthwhile, thereby supporting more positive attitudes toward oral health in older age. However, the persistence of fatalistic beliefs—such as viewing oral deterioration as an inevitable part of aging—highlights the resistance of deeply-rooted social narratives, particularly in socioeconomically vulnerable groups [93]. Addressing normative beliefs may require culturally sensitive messaging, accessible education, and stronger community engagement strategies to build collective efficacy and shift these deeply rooted perceptions [94].

Exploratory subgroup analyses suggested that the intervention may have helped reduce baseline knowledge disparities among rural participants. These differences aligned with underlying oral health inequalities, as rural participants had fewer teeth on average—an indicator commonly associated with limited access to preventive services and delayed diagnosis among higher-risk populations [2,95]. Findings from the multiple correspondence analysis and k-means clustering further refined this interpretation. Although the cluster showing the greatest improvements was predominantly rural, another cluster with mixed urban-rural composition and higher rates of functional dentition displayed limited responsiveness. This pattern suggests that baseline preventive need, proxied by dentition status, may play a more influential role than residence alone in shaping intervention response. Accordingly, delivering educational content through accessible digital platforms may be particularly beneficial for older adults with greater unmet preventive needs—many of whom reside in rural settings [95]—by helping mitigate structural barriers in high-need populations. This approach may be especially relevant in settings with limited access to oral health care and quality educational resources, while also representing a potentially cost-effective strategy to narrow rural-urban disparities in oral health promotion [12,95,96].

The study also highlights persistent challenges related to the "gray digital divide" [40], as a notable proportion of otherwise eligible participants were excluded due to limited WhatsApp skills. To address this, the intervention incorporated inclusive design principles by using familiar social media tools, providing basic digital guidance, and videos tailored and validated for older adults with simple language, visual modeling, and accessibility features such as large fonts, high-contrast images, clear audio, appropriate pacing, and mobile-friendly formats. Despite these actions, some telehealth participants still faced access difficulties. Although this represents a minor unintended event rather than a clinical harm, it underscores how digital competencies—unequally distributed, especially in rural areas—can limit engagement [97,98]. Combining social media use with digital literacy training can enhance eHealth engagement in older adults [99,100].

Overall, the reinforcement was well received and perceived as useful, with high willingness to recommend the intervention, regardless of sociodemographic characteristics. Incorporating older adults' perspectives in co-design may further improve relevance, usability, and equity of digital health tools [53,101]. Recent qualitative research using the Unified Theory of Acceptance and Use of Technology framework highlights the importance of understanding older adults' intentions, attitudes, and concerns to enhance telehealth uptake [102]. The TEGO enabled continuous monitoring and identification of subgroups—such as participants with functional dentition or higher baseline knowledge—who may benefit from tailored content, aligning with prior teledentistry recommendations [43].

From a public health perspective, these findings support the use of smartphone-based telehealth reinforcement as a complementary component of in-person oral health promotion programs, particularly in settings facing structural barriers to preventive care. Although effects were modest, incremental gains—especially among rural older adults—suggest that targeted telehealth interventions can improve access and reinforce learning where traditional services are limited, potentially preventing more costly oral and systemic complications over time [103-106]. Overall, telehealth reinforcement appeared feasible, acceptable, and safe; however, its impact was shaped by digital access, literacy, place of residence, and oral health status, which may limit its universal applicability.

Evidence on the cost-effectiveness of telehealth interventions in aging populations remains limited [107,108]. While economic evaluation was beyond the scope of this trial, findings highlight the importance of context-specific cost analyses, particularly for older adults with heterogeneous needs and patterns of health care use. Existing evidence indicates mixed results, with simpler mHealth approaches showing some benefit, unlike more complex smartphone-based apps [109]. Notably, cost-effectiveness varies across populations and settings, with greater savings reported in rural contexts, and more favorable outcomes for in-person care in some urban minority groups [110]. These insights underscore the need for tailored economic evaluations that prioritize high-need settings, define resource use clearly, use appropriate comparators, and adopt a societal perspective to capture both direct and indirect costs and benefits [111].

Scaling up WhatsApp-based or similar mHealth interventions in primary care will require investment in digital literacy training for health care professionals, integration of mobile-compatible tools into routine workflow, and strengthening technological infrastructure through reliable internet connectivity (eg, satellite or mobile broadband) and ensuring device availability, especially in rural areas [112,113]. Policy support is essential to promote adoption among patients, caregivers, and health teams, recognizing the role of mHealth in education, prevention, and continuity of care [114]. Moreover, safeguarding data privacy and controlling misinformation remain critical, as highlighted by recent qualitative evidence from health system decision-makers emphasizing coordinated policies across governments, clinicians, digital platforms, and the public [115].

Limitations

While the study demonstrates promising results, it is important to recognize its limitations when interpreting the findings. First, the short follow-up period precluded assessment of the sustainability of changes in knowledge, attitudes, and self-efficacy. Second, although the intervention was guided by

XSL•FO
RenderX

SCT, only self-efficacy was directly measured, whereas other constructs were embedded within the intervention design and assessed indirectly through participants' engagement, knowledge acquisition, attitudes, and behavior change outcomes. Engagement with the educational videos was not systematically monitored, limiting evaluation of dose-response relationships or specific mechanisms of change. The absence of qualitative data—such as interviews or open-ended feedback—further restricted insight into participants' subjective experiences and perceived usefulness of the intervention.

Self-reported measures were limited to secondary behavioral outcomes and may be subject to social desirability bias, warranting cautious interpretation. The relatively small sample size reduced statistical power for subgroup analyses and limited generalizability across geographic settings and levels of smartphone use. Subgroup findings, including differences by place of residence, were observational and should be interpreted as descriptive rather than causal. Finally, selection bias may have occurred at the eligibility stage due to unequal access to technology or differential motivation to participate, potentially limiting applicability to populations with lower digital proficiency.

Implications and Future Directions

Despite these limitations, this study contributes to oral health promotion for older adults in underserved regions by demonstrating a structured application of a theory-informed intervention design. While individual components are not novel [17,18,21,31,43,116], their integration into a coherent strategy explicitly informed by SCT across face-to-face and telehealth delivery provides a concrete example of theory-informed intervention design, addressing documented gaps in explicit theory operationalization in oral health interventions targeting older adults [18,117,118]. Unlike most prior m-oral health studies, which have focused predominantly on urban populations [31,43-45,91,119], this study evaluates a geographically balanced rural-urban sample and offers suggestive evidence of differential intervention response across settings, a perspective rarely examined in previous studies. Overall, the findings indicate that theory-informed telehealth reinforcement may help address service gaps in underserved settings. The intervention demonstrated consistently high acceptability across multiple indicators, together with the use of accessible digital platforms—aligned with the digital context of older adults [50,51]—supporting its feasibility and suggesting potential for scalable, context-aligned implementation in resource-limited health systems [95,96]. By generating evidence in a local context, this study offers insights relevant to context-specific policy development and contributes to the growing literature on digital health for aging populations.

Looking ahead, future research should build upon these findings by using larger and more diverse samples, extending follow-up periods, and including clinical outcomes. Given the improvements observed, particularly among rural older adults, future studies should also explore how tailoring interventions to participants' digital literacy, sociodemographic profiles, or baseline health status may enhance outcomes. Incorporating qualitative methods to capture participants' experiences, alongside in-depth analysis of their interaction with digital content (eg, process measures), will be essential to better understand the mechanisms underlying behavioral change. Furthermore, emerging technologies—such as artificial intelligence, gamification, or virtual reality—offer opportunities to enhance engagement, but must be coupled with digital literacy strategies to ensure equitable adoption.

Conclusions

This study highlights the effectiveness of both face-to-face education and telehealth reinforcement, each grounded in SCT, in improving key oral health outcomes among older adults. Per-protocol analyses showed modest benefits of telehealth in specific domains—caries knowledge and attitudes—whereas ITT analyses confirmed significance only for attitudes. Exploratory subgroup findings suggest that individuals from rural settings and those with greater baseline preventive needs may derive added benefits, though these patterns should be interpreted cautiously. The acceptability of the telehealth approach further supports the feasibility of m-oral health and social media as a scalable strategy. Future research should examine long-term effects, impact on clinical outcomes, and explore adaptation for diverse populations and settings, ensuring their relevance and cost-effectiveness for broader implementation.

By examining how telehealth can reinforce high-quality face-to-face promotion, this study adds evidence supporting digital health strategies to promote oral health, particularly among vulnerable populations, and underscores their potential to advance healthy aging and equitable access to preventive care, encouraging further work to refine and expand this approach.

Acknowledgments

The authors thank Universidad de La Frontera for logistical support and the Inter-University Center of Healthy Aging (CIES) of the Council of Rectors of Chilean Universities (CRUCH) for their collaboration. FM-S acknowledges the National Commission for Scientific and Technological Research of Chile (Comisión Nacional de Investigación Científica y Tecnológica de Chile [CONICYT]) support (Magíster Nacional/2022-22220928). Special thanks are given to collaborators Miss Javiera Muñoz-Sepúlveda, Ms María Inés Carileo, Miss Catalina Salgado, and Mr Pablo Uribe, as well as senior adult clubs leaders and participants.

This study was supported by the FRO2495 MINEDUC-Chile project, "Integrated Technological Ecosystem of Tele-Dentistry to Provide General and Specialized Dental Care for Older Adults," funded by the Ministry of Education of Chile. The project was awarded to the Universidad de La Frontera under the Plan for Strengthening State Universities, Regional Higher Education (ESR) line.

XSL•FO
RenderX

ChatGPT (OpenAI) was used exclusively to assist with grammar and language editing, specifically to review the manuscript for grammatical and stylistic issues. All content and scientific writing were developed by the authors.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Funding

This research was funded by the Presidential Grant Project Clinical Center for Innovation in Oral Health of the Elderly in La Araucanía region (exempt resolution number 472).

Authors' Contributions

Conceptualization: FM-S, VB
Data curation: FM-S
Formal analysis: PN (lead), FM-S
Funding acquisition: VB (lead), CA
Investigation: CA, FM-S
Methodology: FM-S, VB, AvM, RM
Project administration: VB (lead), CA, FM-S
Resources: CA, VB
Software: LL
Supervision: AvM
Validation: FM-S
Visualization: FM-S (lead), PN
Writing – original draft: FM-S (lead), AC, CA
Writing – review & editing: VB (lead), RM, AC, AvM, PN, LL

Conflicts of Interest

None declared.

Editorial Notice

This randomized study was only retrospectively registered because of an administrative delay related to internal institutional processing and coordination of clinical trial registry submission. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials because the risk of bias appears low and because the authors provided an official letter from the IRB confirming that the submitted study design is identical to the protocol assessed and approved before trial initiation. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1048 KB - mhealth_v14i1e71251_app1.pdf]

Multimedia Appendix 2
Baseline sociodemographic, clinical, and oral health characteristics of participants by intervention group. Values are presented as mean (standard deviation) or percentage (number).
[PDF File (Adobe PDF File), 132 KB - mhealth_v14i1e71251_app2.pdf]

Multimedia Appendix 3
Baseline sociodemographic, clinical, and oral health characteristics of participants by intervention group and place of residence. Values are presented as mean (SD) or percentage (number). P values correspond to comparisons across the 4 subgroups (rural telehealth education, urban telehealth education, rural comparator, and urban comparator). The Kruskal-Wallis test was used for continuous variables.
[PDF File (Adobe PDF File), 146 KB - mhealth_v14i1e71251_app3.pdf]

References

1. World Health Organization. UN decade of healthy ageing: plan of action 2021-2030. 2020. URL: https://cdn.who.int/media/docs/default-source/decade-of-healthy-ageing/decade-proposal-final-apr2020-en.pdf

XSL•FO
RenderX

2. World Health Organization. Global oral health status report: towards universal health coverage for oral health by 2030. 2022. URL: https://www.who.int/publications/i/item/9789240061569
3. Patel J, Wallace J, Doshi M, Gadanya M, Ben Yahya I, Roseman J, et al. Oral health for healthy ageing. Lancet Healthy Longev 2021;2(8):e521-e527 [FREE Full text] [doi: 10.1016/S2666-7568(21)00142-2] [Medline: 36098001]
4. FDI World Dental Federation. FDI policy statement on oral health for healthy ageing: adopted by the FDI General Assembly: 24 September 2015, Bangkok, Thailand. Int Dent J 2016;66(1):7-8 [FREE Full text] [doi: 10.1111/idj.12231] [Medline: 26803941]
5. Petersen PE, Baez RJ, Ogawa H. Global application of oral disease prevention and health promotion as measured 10 years after the 2007 World Health Assembly statement on oral health. Community Dent Oral Epidemiol 2020;48(4):338-348 [FREE Full text] [doi: 10.1111/cdoe.12538] [Medline: 32383537]
6. Caldwell JT, Lee H, Cagney KA. The role of primary care for the oral health of rural and urban older adults. J Rural Health 2017;33(4):409-418. [doi: 10.1111/jrh.12269] [Medline: 28905422]
7. Dietrich TR, Debona G, de Spessato Schwerz P, Fagundes MLB, Hugo FN, Hilgert JB, et al. Preventive dental service use and oral health-related quality of life in Brazilian older adults. Gerodontology 2024;41(1):40-45. [doi: 10.1111/ger.12700] [Medline: 37386716]
8. Stafford M, von Wagner C, Perman S, Taylor J, Kuh D, Sheringham J. Social connectedness and engagement in preventive health services: an analysis of data from a prospective cohort study. Lancet Public Health 2018;3(9):e438-e446 [FREE Full text] [doi: 10.1016/S2468-2667(18)30141-5] [Medline: 30143472]
9. Ministerio de Salud, Gobierno de Chile. Informe encuesta nacional de salud 2016-2017. Salud Bucal. URL: https://epi.minsal.cl/wp-content/uploads/2021/03/Informe_Salud_Bucal_ENS_2016_17.pdf [accessed 2024-11-21]
10. Spinler K, Aarabi G, Valdez R, Kofahl C, Heydecke G, König HH, et al. Prevalence and determinants of dental visits among older adults: findings of a nationally representative longitudinal study. BMC Health Serv Res 2019;19(1):590 [FREE Full text] [doi: 10.1186/s12913-019-4427-0] [Medline: 31429740]
11. Rural population (% of total population). World Bank Open Data. 2023. URL: https://data.worldbank.org/indicator/SP.RUR.TOTL.ZS?view=chart [accessed 2024-11-08]
12. Theriault H, Bridge G. Oral health equity for rural communities: where are we now and where can we go from here? Br Dent J 2023;235(2):99-102. [doi: 10.1038/s41415-023-6058-4] [Medline: 37500855]
13. Observatorio del Envejecimiento. Envejecer en la ruralidad: explorando una realidad invisibilizada.: Centro de Estudios de Vejez y Envejecimiento de la Pontificia Universidad Católica de Chile y Compañía de Seguros Confuturo; 2024. URL: https://www.uc.cl/noticias/observatorio-del-envejecimiento-mayores-de-60-anos-duplicaran-a-jovenes-en-2035-en-zonas-rurales/
14. World Health Organization. Health promotion glossary of terms 2021. 2021. URL: https://www.who.int/publications/i/item/9789240038349
15. Jadhav HC, Dodamani AS, Karibasappa GN, Naik RG, Khairnar MR, Deshmukh MA, et al. Effect of reinforcement of oral health education message through short messaging service in mobile phones: a quasi-experimental trial. Int J Telemed Appl 2016;2016:7293516 [FREE Full text] [doi: 10.1155/2016/7293516] [Medline: 26941793]
16. Wang Y, Zhu J, Xu Z, Dai X, Chen K, Wang Y. Social support, oral health knowledge, attitudes, practice, self-efficacy and oral health-related quality of life in Chinese college students. Sci Rep 2023;13(1):12320 [FREE Full text] [doi: 10.1038/s41598-023-39658-6] [Medline: 37516775]
17. Mariño R, Hopcraft M, Ghanim A, Tham R, Khew C, Stevenson C. Oral health-related knowledge, attitudes and self-efficacy of Australian rural older adults. Gerodontology 2016;33(4):530-538. [doi: 10.1111/ger.12202] [Medline: 26174570]
18. Bashirian S, Khoshravesh S, Ayubi E, Karimi-Shahanjarini A, Shirahmadi S, Solaymani PF. The impact of health education interventions on oral health promotion among older people: a systematic review. BMC Geriatr 2023;23(1):548 [FREE Full text] [doi: 10.1186/s12877-023-04259-5] [Medline: 37697260]
19. Renz A, Ide M, Newton T, Robinson PG, Smith D. Psychological interventions to improve adherence to oral hygiene instructions in adults with periodontal diseases. Cochrane Database Syst Rev 2007(2):CD005097. [doi: 10.1002/14651858.CD005097.pub2] [Medline: 17443571]
20. Kopp SL, Ramseier CA, Ratka-Krüger P, Woelber JP. Motivational interviewing as an adjunct to periodontal therapy-a systematic review. Front Psychol 2017;8:279 [FREE Full text] [doi: 10.3389/fpsyg.2017.00279] [Medline: 28293208]
21. Islam KF, Awal A, Mazumder H, Munni UR, Majumder K, Afroz K, et al. Social cognitive theory-based health promotion in primary care practice: a scoping review. Heliyon 2023;9(4):e14889 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e14889] [Medline: 37025832]
22. Bandura A. Health promotion from the perspective of social cognitive theory. Psychol Health 1998;13(4):623-649. [doi: 10.1080/08870449808407422]
23. Tapia P C, Iturra M V, Valdivia R Y, Varela V H, Jorquera C M, Carmona G A. Estado de salud y autoeficacia en adultos mayores usuarios de atención primaria en salud. Cienc Enferm 2017;23(3):35-45. [doi: 10.4067/s0717-95532017000300035]
24. Allen F, Fan SY, Loke WM, Na TM, Keng Yan GL, Mittal R. The relationship between self-efficacy and oral health status of older adults. J Dent 2022;122:104085. [doi: 10.1016/j.jdent.2022.104085] [Medline: 35248673]

XSL•FO
RenderX

25. Bandura A. On the functional properties of perceived self-efficacy revisited. J Manage 2011;38(1):9-44. [doi: 10.1177/0149206311410606]
26. Araújo MR, Alvarez M, Godinho CA, Almeida T, Pereira CR. Self-regulation in oral hygiene behaviours in adults with gingivitis: the mediating role of coping planning and action control. Int J Dent Hyg 2020;18(2):192-200. [doi: 10.1111/idh.12430] [Medline: 32053277]
27. Khamrin P, Boonyathee S, Bootsikeaw S, Ong-Artborirak P, Seangpraw K. Factors associated with health literacy, self-efficacy, social support, and oral health care behaviors among elderly in northern border community Thailand. Clin Interv Aging 2021;16:1427-1437 [FREE Full text] [doi: 10.2147/CIA.S320900] [Medline: 34326634]
28. Woelber JP, Bienas H, Fabry G, Silbernagel W, Giesler M, Tennert C, et al. Oral hygiene-related self-efficacy as a predictor of oral hygiene behaviour: a prospective cohort study. J Clin Periodontol 2015;42(2):142-149. [doi: 10.1111/jcpe.12348] [Medline: 25472523]
29. Zhou G, Sun C, Knoll N, Hamilton K, Schwarzer R. Self-efficacy, planning and action control in an oral self-care intervention. Health Educ Res 2015;30(4):671-681. [doi: 10.1093/her/cyv032] [Medline: 26187914]
30. Mariño R, Calache H, Morgan M. A community-based culturally competent oral health promotion for migrant older adults living in Melbourne, Australia. J Am Geriatr Soc 2013;61(2):270-275. [doi: 10.1111/jgs.12078] [Medline: 23320643]
31. Mariño RJ, Marwaha P, Barrow SY. Web-based oral health promotion program for older adults: development and preliminary evaluation. Int J Med Inform 2016;91:e9-e15. [doi: 10.1016/j.ijmedinf.2016.04.002] [Medline: 27117812]
32. United Nations. Transforming our world: the 2030 agenda for sustainable development. 2015. URL: https://sdgs.un.org/sites/default/files/publications/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf
33. World Health Organization. Global diffusion of eHealth: making universal health coverage achievable: report of the third global survey on eHealth. Geneva: World Health Organization; 2016. URL: https://www.who.int/publications/i/item/9789241511780
34. World Health Organization. Recommendations on digital interventions for health system strengthening. Geneva: World Health Organization; 2019. URL: https://www.who.int/publications/i/item/9789241550505
35. Solis-Navarro L, Gismero A, Fernández-Jané C, Torres-Castro R, Solá-Madurell M, Bergé C, et al. Effectiveness of home-based exercise delivered by digital health in older adults: a systematic review and meta-analysis. Age Ageing 2022;51(11):afac243 [FREE Full text] [doi: 10.1093/ageing/afac243] [Medline: 36346736]
36. Pourrazavi S, Kouzekanani K, Bazargan-Hejazi S, Shaghaghi A, Hashemiparast M, Fathifar Z, et al. Theory-based e-health literacy interventions in older adults: a systematic review. Arch Public Health 2020;78:72 [FREE Full text] [doi: 10.1186/s13690-020-00455-6] [Medline: 32793345]
37. Zou C, Harvard A, Qian J, Fox BI. A systematic review of digital health technologies for the care of older adults during COVID-19 pandemic. Digit Health 2023;9:20552076231191050 [FREE Full text] [doi: 10.1177/20552076231191050] [Medline: 37529545]
38. Lee JJN, Abdul Aziz A, Chan S, Raja Abdul Sahrizan RSFB, Ooi AYY, Teh Y, et al. Effects of mobile health interventions on health-related outcomes in older adults with type 2 diabetes: a systematic review and meta-analysis. J Diabetes 2023;15(1):47-57 [FREE Full text] [doi: 10.1111/1753-0407.13346] [Medline: 36649940]
39. Zhang L, Ge Y, Zhao W, Shu X, Kang L, Wang Q, et al. A 4-week mobile app-based telerehabilitation program vs conventional in-person rehabilitation in older adults with sarcopenia: randomized controlled trial. J Med Internet Res 2025;27:e67846 [FREE Full text] [doi: 10.2196/67846] [Medline: 39854716]
40. Mubarak F, Suomi R. Elderly forgotten? Digital exclusion in the information age and the rising grey digital divide. Inquiry 2022;59:469580221096272 [FREE Full text] [doi: 10.1177/00469580221096272] [Medline: 35471138]
41. Tan SHX, Lee CKJ, Yong CW, Ding YY. Scoping review: facilitators and barriers in the adoption of teledentistry among older adults. Gerodontology 2021;38(4):351-365. [doi: 10.1111/ger.12588] [Medline: 34523172]
42. Mariño RJ, Uribe SE, Chen R, Schwendicke F, Giraudeau N, Scheerman JFM. Terminology of e-oral health: consensus report of the IADR's e-oral health network terminology task force. BMC Oral Health 2024;24(1):280 [FREE Full text] [doi: 10.1186/s12903-024-03929-z] [Medline: 38419003]
43. Wanyonyi K, Couch C, John J, Louca C. e-Oral health interventions for older patients in an outreach primary dental care centre: a pilot trial nested acceptability study. Gerodontology 2022;39(3):241-249. [doi: 10.1111/ger.12562] [Medline: 34018237]
44. Ki JY, Jo SR, Cho KS, Park JE, Cho JW, Jang JH. Effect of oral health education using a mobile app (OHEMA) on the oral health and swallowing-related quality of life in community-based integrated care of the elderly: a randomized clinical trial. Int J Environ Res Public Health 2021;18(21):11679 [FREE Full text] [doi: 10.3390/ijerph182111679] [Medline: 34770193]
45. Lee KH, Choi YY, Jung ES. Effectiveness of an oral health education programme using a mobile application for older adults: a randomised clinical trial. Gerodontology 2023;40(1):47-55. [doi: 10.1111/ger.12616] [Medline: 34951060]
46. de Oliveira Júnior AJ, Oliveira JM, Bretz YP, Mialhe FL. Online social networks for prevention and promotion of oral health: a systematic review. Can J Dent Hyg 2023;57(2):83-97 [FREE Full text] [Medline: 37464997]
47. Zhao YC, Zhao M, Song S. Online health information seeking behaviors among older adults: systematic scoping review. J Med Internet Res 2022;24(2):e34790 [FREE Full text] [doi: 10.2196/34790] [Medline: 35171099]

XSL•FO
RenderX

48. Fernández CE, Maturana CA, Coloma SI, Carrasco-Labra A, Giacaman RA. Teledentistry and mHealth for promotion and prevention of oral health: a systematic review and meta-analysis. J Dent Res 2021;100(9):914-927. [doi: 10.1177/00220345211003828] [Medline: 33769123]
49. OECD. Digital transformation in the age of COVID-19: building resilience and bridging divides, digital economy outlook 2020 supplement. URL: https://www.oecd.org/en/publications/oecd-digital-economy-outlook-2020_bb167041-en.html
50. Taipale S, Oinas T, Ivan L, Rosenberg D. Mobile phone use before and during the COVID-19 pandemic – a panel study of older adults in seven countries. Mob Media Commun 2023;12(1):23-44. [doi: 10.1177/20501579231185479]
51. Observatorio del Envejecimiento. Uso de internet y tecnologías de la información y comunicación en las personas mayores.: Centro de Estudios de Vejez y Envejecimiento de la Pontificia Universidad Católica de Chile y Compañía de Seguros Confuturo; 2022. URL: https://observatorioenvejecimiento.uc.cl/wp-content/uploads/2022/06/Observatorio-Reporte-TICS.pdf
52. Farrokhi F, Ghorbani Z, Farrokhi F, Namdari M, Salavatian S. Social media as a tool for oral health promotion: a systematic review. PLoS One 2023;18(12):e0296102 [FREE Full text] [doi: 10.1371/journal.pone.0296102] [Medline: 38113216]
53. Goodman C, Lambert K. Scoping review of the preferences of older adults for patient education materials. Patient Educ Couns 2023;108:107591. [doi: 10.1016/j.pec.2022.107591] [Medline: 36584555]
54. Petkovic J, Duench S, Trawin J, Dewidar O, Pardo Pardo J, Simeon R, et al. Behavioural interventions delivered through interactive social media for health behaviour change, health outcomes, and health equity in the adult population. Cochrane Database Syst Rev 2021;5(5):CD012932 [FREE Full text] [doi: 10.1002/14651858.CD012932.pub2] [Medline: 34057201]
55. Ferret JCF, Branco BHM, Santos LPGD, Rocco F, Bernuci MP. WhatsApp-assisted health education intervention. Int J Innov Educ Res 2021;9(9):56-72. [doi: 10.31686/ijier.vol9.iss9.3316]
56. Sartori AC, Rodrigues Lucena TF, Lopes CT, Picinin Bernuci M, Yamaguchi MU. Educational intervention using on medication adherence in hypertension and diabetes patients: a randomized clinical trial. Telemed J E Health 2020;26(12):1526-1532. [doi: 10.1089/tmj.2019.0305] [Medline: 32155382]
57. Yaagoob E, Lee R, Stubbs M, Shuaib F, Johar R, Chan S. WhatsApp-based intervention for people with type 2 diabetes: a randomized controlled trial. Nurs Health Sci 2024;26(2):e13117. [doi: 10.1111/nhs.13117] [Medline: 38566413]
58. Mohammad Z, Ahmad N, Baharom A. The effects of theory-based educational intervention and WhatsApp follow-up on Papanicolaou smear uptake among postnatal women in Malaysia: randomized controlled trial. JMIR Mhealth Uhealth 2022;10(6):e32089 [FREE Full text] [doi: 10.2196/32089] [Medline: 35759319]
59. Council for International Organizations of Medical Sciences (CIOMS). International Ethical Guidelines for Health-Related Research Involving Humans. Geneva: Council for International Organizations of Medical Sciences; 2016.
60. Hopewell S, Chan A, Collins GS, Hróbjartsson A, Moher D, Schulz KF, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. BMJ 2025;389:e081123 [FREE Full text] [doi: 10.1136/bmj-2024-081123] [Medline: 40228833]
61. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
62. Instituto Nacional de Estadísticas (INE), Gobierno de Chile. Anuario de estadísticas vitales 2020. 2023. URL: https://www.ine.gob.cl/docs/default-source/nacimientos-matrimonios-y-defunciones/publicaciones-y-anuarios/anuarios-de-estad%C3%ADsticas-vitales/anuario-de-estad%C3%ADsticas-vitales-2020.pdf?sfvrsn=6fb93518_4
63. Ministerio de Salud, Gobierno de Chile. Diagnósticos Regionales Con Enfoque En Determinantes Sociales de Salud: Región de La Araucanía. Santiago: Ministerio de Salud Gobierno de Chile; 2017.
64. Sanzana-Luengo C, Díaz L, Abarza L, Alvarado E, Beltrán V. Teledentistry protocol for the elderly in the context of the COVID-19 pandemic through a web platform/mobile app: approach from the general dentist. J Oral Res 2022;S(1):1-8. [doi: 10.17126/joralres.2022.026]
65. Beltrán V, Flores M, Sanzana C, Muñoz-Sepúlveda F, Alvarado E, Venegas B, et al. Tooth loss and caries experience of elderly Chileans in the context of the COVID-19 pandemic in five regions of Chile. Int J Environ Res Public Health 2023;20(4):3001 [FREE Full text] [doi: 10.3390/ijerph20043001] [Medline: 36833696]
66. Marchetti G, Fraiz FC, Nascimento WMD, Soares GMS, Assunção LRDS. Improving adolescents' periodontal health: evaluation of a mobile oral health app associated with conventional educational methods: a cluster randomized trial. Int J Paediatr Dent 2018;28(4):410-419. [doi: 10.1111/ipd.12371] [Medline: 29756308]
67. Walters SJ, Bonacho Dos Anjos Henriques-Cadby I, Bortolami O, Flight L, Hind D, Jacques RM, et al. Recruitment and retention of participants in randomised controlled trials: a review of trials funded and published by the United Kingdom health technology assessment programme. BMJ Open 2017;7(3):e015276 [FREE Full text] [doi: 10.1136/bmjopen-2016-015276] [Medline: 28320800]
68. Organización Panamericana de la Salud (OPS-OMS), Universidad CES de Colombia. EPIDAT: programa para análisis epidemiológico de datos. Consellería de Sanidade, Xunta de Galicia. 2016. URL: https://www.sergas.es/Saude-publica/EPIDAT?idioma=es [accessed 2026-05-07]
69. FDI World Dental Federation. Caries prevention and management: advocacy toolkit. 2016. URL: https://www.fdiworldental.org/sites/default/files/2020-11/2016-fdi_cpp-advocacy_toolkit.pdf

XSL•FO
RenderX

70. Herrera D, Meyle J, Renvert S, Jin L. White paper on prevention and management of periodontal diseases for oral health and general health. FDI Global Periodontal Health Project Task Team. URL: https://www.fdiworlddental.org/sites/default/files/2020-11/gphp-2018-white_paper-en.pdf [accessed 2024-12-27]
71. FDI World Dental Federation. Achieving a healthy ageing society: oral health for an ageing population. 2018. URL: https://www.fdiworldental.org/sites/default/files/2020-11/ohap-2018-advocacy_doc-achieving_healthy_ageing_society.pdf
72. FDI World Dental Federation. Oral cancer: prevention and patient management. 2018. URL: https://www.fdiworldental.org/sites/default/files/2020-11/fdi-oral_cancer-prevention_and_patient_management.pdf
73. Almanasreh E, Moles R, Chen TF. Evaluation of methods used for estimating content validity. Res Social Adm Pharm 2019;15(2):214-221. [doi: 10.1016/j.sapharm.2018.03.066] [Medline: 29606610]
74. Shoemaker SJ, Wolf MS, Brach C. Development of the Patient Education Materials Assessment Tool (PEMAT): a new measure of understandability and actionability for print and audiovisual patient information. Patient Educ Couns 2014;96(3):395-403 [FREE Full text] [doi: 10.1016/j.pec.2014.05.027] [Medline: 24973195]
75. Beltrán V, von Marttens A, Acuña-Mardones P, Sanzana-Luengo C, Rueda-Velásquez SJ, Alvarado E, et al. Implementation of a teledentistry platform for dental emergencies for the elderly in the context of the COVID-19 pandemic in Chile. Biomed Res Int 2022;2022(1):6889285 [FREE Full text] [doi: 10.1155/2022/6889285] [Medline: 35330690]
76. Beltrán V, Acuña-Mardones. P, Fernández-Gil F, Acuña-Mardones D, Engelke W. Development of a teledentistry system for the elderly in the context of the COVID-19 pandemic: 3D diagnostic models integration in a web platform. J Oral Res 2022;S(1):1-7. [doi: 10.17126/joralres.2022.024]
77. Ficha regional Araucanía. Oficina de Estudios y Políticas Agrarias (ODEPA).: Gobierno de Chile; 2021. URL: https://www.odepa.gob.cl/estadisticas-del-sector/ficha-nacional-y-regionales [accessed 2024-12-27]
78. Orientaciones al RSH N° 8:Cálculo de la salificación Socioeconómica. Ministerio de Desarrollo Social.: Gobierno de Chile; 2019. URL: http://www.registrosocial.gob.cl/docs/Orientaciones-complementarias-N8_calculo-CSE_VF.pdf [accessed 2024-12-27]
79. Skou ST, Mair FS, Fortin M, Guthrie B, Nunes BP, Miranda JJ, et al. Multimorbidity. Nat Rev Dis Primers 2022 Jul 14;8(1):48 [FREE Full text] [doi: 10.1038/s41572-022-00376-4] [Medline: 35835758]
80. Quiroga P, Albala C, Klaasen G. Validation of a screening test for age associated cognitive impairment, in Chile. Rev Med Chil 2004;132(4):467-478. [doi: 10.4067/s0034-98872004000400009] [Medline: 15382519]
81. Martínez de la Iglesia J, Onís Vilches MC, Dueñas Herrero R, Albert Colomer C, Aguado Taberné C, Luque Luque R. Versión española del cuestionario de Yesavage abreviado (GDS) para el despistaje de depresión en mayores de 65 años: adaptación y validación. Medifam 2002;12(10):26-40. [doi: 10.4321/s1131-57682002001000003]
82. González N, Bilbao A, Forjaz MJ, Ayala A, Orive M, Garcia-Gutierrez S, et al. Psychometric characteristics of the Spanish version of the Barthel index. Aging Clin Exp Res 2018;30(5):489-497. [doi: 10.1007/s40520-017-0809-5] [Medline: 28770477]
83. Programa nacional de salud de las personas adultas mayores. Ministerio de Salud.: Gobierno de Chile; 2014. URL: https://www.minsal.cl/sites/default/files/files/Borrador%20documento%20Programa%20Nacional%20de%20Personas%20Adultas%20Mayores-%2004-03_14.pdf
84. Mariño R, Calache H, Wright C, Schofield M, Minichiello V. Oral health promotion programme for older migrant adults. Gerodontology 2004;21(4):216-225. [doi: 10.1111/j.1741-2358.2004.00035.x] [Medline: 15603281]
85. Clinical trials to enhance elders oral health (TEETH). ClinicalTrials.gov. 2012. URL: https://classic.clinicaltrials.gov/ct2/show/NCT00004640 [accessed 2024-12-27]
86. Muñoz-Sepúlveda F, Acevedo C, Mariño R, Sanzana-Luengo C, Navarro P, Beltrán V. Translation and validation into Spanish of the Oral Health Self-Efficacy Scale in Chilean older adults. BMC Oral Health 2024;24(1):118 [FREE Full text] [doi: 10.1186/s12903-024-03889-4] [Medline: 38245718]
87. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
88. World Health Organization. Global strategy and action plan on oral health 2023-2030. Geneva: World Health Organization; 2024. URL: https://www.who.int/publications/i/item/9789240090538
89. Hecht CA, Gosling SD, Bryan CJ, Jamieson JP, Murray JS, Yeager DS. When do the effects of single-session interventions persist? Testing the mindset + supportive context hypothesis in a longitudinal randomized trial. JCPP Adv 2023;3(4):e12191 [FREE Full text] [doi: 10.1002/jcv2.12191] [Medline: 38054060]
90. Meaney PA, Joyce CL, Setlhare S, Smith HE, Mensinger JL, Zhang B, et al. Knowledge acquisition and retention following Saving Children's Lives course for healthcare providers in Botswana: a longitudinal cohort study. BMJ Open 2019;9(8):e029575 [FREE Full text] [doi: 10.1136/bmjopen-2019-029575] [Medline: 31420392]
91. Ng JYM, Lim TW, Tarib N, Ho TK. Effect of educational progressive web application on patient's oral and denture knowledge and hygiene: a randomised controlled trial. Health Informatics J 2021;27(3):14604582211035821 [FREE Full text] [doi: 10.1177/14604582211035821] [Medline: 34325550]
92. McGrath C, Zhang W, Lo EC. A review of the effectiveness of oral health promotion activities among elderly people. Gerodontology 2009;26(2):85-96. [doi: 10.1111/j.1741-2358.2008.00232.x] [Medline: 19490131]

XSL•FO
RenderX

93. King S, Thaliph A, Laranjo L, Smith BJ, Eberhard J. Oral health literacy, knowledge and perceptions in a socially and culturally diverse population: a mixed methods study. BMC Public Health 2023;23(1):1446 [FREE Full text] [doi: 10.1186/s12889-023-16381-5] [Medline: 37507737]
94. Cipta DA, Andoko D, Theja A, Utama AVE, Hendrik H, William DG, et al. Culturally sensitive patient-centered healthcare: a focus on health behavior modification in low and middle-income nations-insights from Indonesia. Front Med (Lausanne) 2024;11:1353037 [FREE Full text] [doi: 10.3389/fmed.2024.1353037] [Medline: 38681051]
95. Cericato GO, Agostini BA, Costa FDS, Thomson WM, Demarco FF. Rural-urban differences in oral health among older people in Southern Brazil. Braz Oral Res 2021;35:e135 [FREE Full text] [doi: 10.1590/1807-3107bor-2021.vol35.0135] [Medline: 34932664]
96. Malpe M, Choudhari SG, Nagtode N, Muntode Gharde P. Beyond the chair: exploring the boundaries of teledentistry. Cureus 2024;16(6):e62286. [doi: 10.7759/cureus.62286] [Medline: 39006674]
97. Ji H, Dong J, Pan W, Yu Y. Associations between digital literacy, health literacy, and digital health behaviors among rural residents: evidence from Zhejiang, China. Int J Equity Health 2024;23(1):68 [FREE Full text] [doi: 10.1186/s12939-024-02150-2] [Medline: 38594723]
98. Yu Y, Liang Z, Zhou Q, Tuersun Y, Liu S, Wang C, et al. Decomposition and comparative analysis of urban-rural disparities in eHealth literacy among Chinese university students: cross-sectional study. J Med Internet Res 2025;27:e63671 [FREE Full text] [doi: 10.2196/63671] [Medline: 40138681]
99. Haimi M, Sergienko R. Adoption and use of telemedicine and digital health services among older adults in light of the COVID-19 pandemic: repeated cross-sectional analysis. JMIR Aging 2024;7:e52317 [FREE Full text] [doi: 10.2196/52317] [Medline: 38656768]
100. Zhang C, Mohamad E, Azlan AA, Wu A, Ma Y, Qi Y. Social media and eHealth literacy among older adults: systematic literature review. J Med Internet Res 2025;27:e66058 [FREE Full text] [doi: 10.2196/66058] [Medline: 40138684]
101. Carpenter SM, Greer ZM, Newman R, Murphy SA, Shetty V, Nahum-Shani I. Developing message strategies to engage racial and ethnic minority groups in digital oral self-care interventions: participatory co-design approach. JMIR Form Res 2023;7:e49179 [FREE Full text] [doi: 10.2196/49179] [Medline: 38079204]
102. Fleischer A, Lowman J, Strader K. Understanding older adults' intention to use telehealth: a qualitative study using the UTAUT framework. J Prim Care Community Health 2025;16:21501319251320180 [FREE Full text] [doi: 10.1177/21501319251320180] [Medline: 39976543]
103. Jang I, Jung H, Park H, Lee CK, Yu SS, Lee YS, et al. A multicomponent frailty intervention for socioeconomically vulnerable older adults: a designed-delay study. Clin Interv Aging 2018;13:1799-1814 [FREE Full text] [doi: 10.2147/CIA.S177018] [Medline: 30275687]
104. Tembhe A, Preisser JS, Batorsky A, Weintraub JA. Ten-year cross-sectional and longitudinal assessment and factors associated with unfavourable self-rated oral health in older adults in the United States. Gerodontology 2024;41(3):357-367. [doi: 10.1111/ger.12710] [Medline: 37550001]
105. Lim J, Park H, Lee H, Lee E, Lee D, Jung H, et al. Longitudinal impact of oral health on geriatric syndromes and clinical outcomes in community-dwelling older adults. BMC Geriatr 2021;21(1):482 [FREE Full text] [doi: 10.1186/s12877-021-02416-2] [Medline: 34481482]
106. Park CM, Oh G, Lee H, Jung H, Lee E, Jang I, et al. Multicomponent intervention and long-term disability in older adults: a nonrandomized prospective study. J Am Geriatr Soc 2021;69(3):669-677 [FREE Full text] [doi: 10.1111/jgs.16926] [Medline: 33155305]
107. Zhang Y, Leuk JS, Teo W. Domains, feasibility, effectiveness, cost, and acceptability of telehealth in aging care: scoping review of systematic reviews. JMIR Aging 2023;6:e40460 [FREE Full text] [doi: 10.2196/40460] [Medline: 37071459]
108. Rush KL, Singh S, Seaton CL, Burton L, Li E, Jones C, et al. Telehealth use for enhancing the health of rural older adults: a systematic mixed studies review. Gerontologist 2022;62(10):e564-e577. [doi: 10.1093/geront/gnab141] [Medline: 34661675]
109. Ghani Z, Jarl J, Sanmartin Berglund J, Andersson M, Anderberg P. The cost-effectiveness of mobile health (mHealth) interventions for older adults: systematic review. Int J Environ Res Public Health 2020;17(15):5290 [FREE Full text] [doi: 10.3390/ijerph17155290] [Medline: 32708016]
110. Kim Y, Shen JJ, Choe I, Reeves J, Byun D, Ioanitoaia-Chaudhry I, et al. Cost-effectiveness analysis of telehealth and in-person primary care visits for people living with Alzheimer's disease-related disorders in the state of Nevada. Int J Environ Res Public Health 2024;21(10):1381 [FREE Full text] [doi: 10.3390/ijerph21101381] [Medline: 39457353]
111. Gomes M, Murray E, Raftery J. Economic evaluation of digital health interventions: methodological issues and recommendations for practice. Pharmacoeconomics 2022;40(4):367-378 [FREE Full text] [doi: 10.1007/s40273-022-01130-0] [Medline: 35132606]
112. Alarcón Belmonte I, Jiménez Lozano AI, Sánchez Collado C, Forero Chaves V, Cabrera Majada A, García-Gutiérrez Gómez R. Addressing the digital determinants of health: main advantages and challenges of digital health. Aten Primaria 2025;57(10):103311 [FREE Full text] [doi: 10.1016/j.aprim.2025.103311] [Medline: 40554952]

XSL•FO
RenderX

113. Rico Fontalvo H, Rico F, de la Puente M, Torres J, Guzman H. When telehealth fails rural communities: the 40% internet threshold that changes everything. Digit Health 2025;11:20552076251393407 [FREE Full text] [doi: 10.1177/20552076251393407] [Medline: 41195381]
114. Kengne Talla P, Inquimbert C, Dawson A, Zidarov D, Bergeron F, Chandad F. Barriers and enablers to implementing teledentistry from the perspective of dental health care professionals: protocol for a systematic quantitative, qualitative, and mixed studies review. JMIR Res Protoc 2023;12:e44218 [FREE Full text] [doi: 10.2196/44218] [Medline: 37494093]
115. Alanazi NA, Almoajel AM, Tharkar S, Almutairi K, Ahmad Mohamad FN, Talak Almatairi BS. Perceptions of executive decision makers on using social media in effective health communication: qualitative study. J Med Internet Res 2025;27:e69269 [FREE Full text] [doi: 10.2196/69269] [Medline: 40397497]
116. Deshpande N, Wu M, Kelly C, Woodrick N, Werner DA, Volerman A, et al. Video-based educational interventions for patients with chronic illnesses: systematic review. J Med Internet Res 2023;25:e41092 [FREE Full text] [doi: 10.2196/41092] [Medline: 37467015]
117. Lavoie A, Dubé V. Web-based interventions to promote healthy lifestyles for older adults: scoping review. Interact J Med Res 2022;11(2):e37315 [FREE Full text] [doi: 10.2196/37315] [Medline: 35998024]
118. Dalgetty R, Miller CB, Dombrowski SU. Examining the theory-effectiveness hypothesis: a systematic review of systematic reviews. Br J Health Psychol 2019;24(2):334-356. [doi: 10.1111/bjhp.12356] [Medline: 30793445]
119. Chau RCW, Thu KM, Chaurasia A, Hsung RTC, Lam WY. A systematic review of the use of mHealth in oral health education among older adults. Dent J (Basel) 2023;11(8):189 [FREE Full text] [doi: 10.3390/dj11080189] [Medline: 37623285]

Abbreviations

ANCOVA: analysis of covariance
CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Applications and Online Telehealth
ICT: information and communications technology
I-CVI: Item Content Validity Index
ITT: intention-to-treat
mHealth: mobile health
MNAR: missing not at random
m-oral health: mobile oral health
OHSES: Oral Health Self-Efficacy Scale
SCT: social cognitive theory
TEGO: Geriatric Dental Specialties Tele-platform

Edited by S Brini; submitted 23.Jan.2025; peer-reviewed by A Azizi, M Seif, E Opoku-Agyemang; comments to author 05.Jun.2025; accepted 21.Jan.2026; published 15.May.2026.

Please cite as:
Muñoz-Sepúlveda F, Beltrán V, von Marttens A, Navarro P, Acevedo C, Mariño R, Chaparro A, López L
WhatsApp-Supported Teledentistry to Reinforce Oral Health Promotion Among Older Adults Residing in Rural and Urban Areas: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e71251
URL: https://mhealth.jmir.org/2026/1/e71251
doi:10.2196/71251
PMID:42139689

©Fernanda Muñoz-Sepúlveda, Víctor Beltrán, Alfredo von Marttens, Pablo Navarro, Claudia Acevedo, Rodrigo Mariño, Alejandra Chaparro, Leonardo López. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 15.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobile-First Web Access and Captioned Video in Francophone Cardiology Education: Multicountry Ecological Learning Analytics Study

Thomas Roussel[1], MSc; Elodie Benhaim[2,3], MPH; Louis Perrard[3,4], MD; Benoit Merat[3,5], MD; Shamir Vally[3,6], MD; Remi Girerd[7,8], MD; Pierre Deharo[1,9], MD, PhD; Louis-Marie Desroche[3,10,11], MD; École Numérique de Cardiologie[12,13]

[1]Cardiology Department, Assistance Publique Hôpitaux de Marseille, Marseille, Provence-Alpes-Côte d'Azur, France

[10]Cardiology Department, University Hospital of La Réunion, Allée des topazes CS 11 021, Saint Denis, La Réunion, France

[11]Clinical Investigation Center – Epidemiology and Clinical Research (CIC-EC), INSERM 1410, University Hospital of La Réunion, Saint Denis, La Réunion, France

[12]See Acknowledgments

[13]

[2]Department of Research and Innovation, University Hospital of La Réunion, Saint Paul, Réunion, Réunion

[3]Ecole Numérique de Cardiologie (ENC), Saint Denis, La Réunion, France

[4]Cardiology Department, Centre Hospitalier Universitaire Amiens-Picardie, Amiens, Hauts-de-France, France

[5]Cardiology Department, Hôpital d'instruction des Armées Percy, Clamart, Île-de-France, France

[6]Cardiology Department, University Hospital of La Réunion, Saint Denis, La Réunion, France

[7]Emergency Department, University Hospital of La Réunion, Saint Pierre, La Réunion, France

[8]Réunisim, University Hospital of La Réunion, Saint Pierre, La Réunion, France

[9]Cardiovascular and Nutrition Research Center (C2VN), INSERM U1263, INRAE U1260, Aix-Marseille Université, Marseille, Provence-Alpes-Côte d'Azur, France

Corresponding Author:
Louis-Marie Desroche, MD
Ecole Numérique de Cardiologie (ENC), Saint Denis, La Réunion, France

Abstract

Background: Mobile health (mHealth) and online video are increasingly central to cardiology education and point-of-care decision support. However, little is known about how simple design choices, such as mobile-first web layouts and captioned videos, translate into real-world practice across countries with different income levels.

Objective: This exploratory ecological study used routinely collected, cross-platform learning analytics from a francophone cardiology mHealth initiative to (1) describe how mobile web access and caption-enabled YouTube viewing varied across World Bank income groups and (2) examine whether greater reliance on mobile access was associated with poorer engagement on the website or on YouTube.

Methods: We analyzed country-level analytics from the École Numérique de Cardiologie (ENC; Saint-Denis) mobile-optimized website and its companion YouTube channel (YouTube, LLC [Google LLC]) over a two-year window (September 2023 to September 2025). Countries were grouped as high-, middle-, or low-income (World Bank, three-level classification). Country-level metrics included mobile device session share; website bounce rate; time on page; and YouTube average view duration, audience retention, and intentional views. Caption-related and demographic YouTube metrics were available only as income-group aggregates and were therefore reported descriptively as between-group contrasts; country-level inferential analyses were restricted to country-level variables. Reporting followed the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist.

Results: Thirty-four countries contributed data: 13/34 (38%) high-income, 14/34 (41%) middle-income, and 7/34 (21%) low-income. Caption-enabled watch time was 18.8% in high-income countries (HICs), compared with 38.7% in middle-income countries (MICs) and 60.9% in low-income countries (LICs), representing a caption equity gap (CEG) of 42.1% between low- and high-income settings. Median website mobile share rose with decreasing income (36.5%, 63.3%, and 81.4%, respectively; Jonckheere-Terpstra $P=.01$). Across income groups, higher caption-enabled watch time coincided with a higher share of intentional views. At the country level, greater reliance on mobile access was not associated with higher bounce rate or shorter time on page, and Spearman correlations between mobile share and YouTube engagement metrics were small and nonsignificant (all $|\rho|\leq 0.28$; all $P\geq .18$).

XSL•FO
RenderX

Conclusions: In this multicountry, francophone, mHealth learning analytics case study, mobile web access and captioned video were used most intensively in lower-income settings, and greater reliance on mobile access was not associated with measurable penalties in basic engagement metrics. These findings support treating mobile-optimized design and systematic captioning as core, low-cost, access-supporting features for equitable digital cardiology education. They also suggest that routinely collected platform indicators can serve as practical equity-monitoring signals for global mHealth initiatives, while underscoring that engagement metrics are not direct measures of learning or behavior change.

(JMIR Mhealth Uhealth 2026;14:e90345) doi:10.2196/90345

KEYWORDS

mobile health; cardiology; medical education; health equity; digital divide; YouTube; learning analytics; ecological study

Introduction

Digital technologies are reshaping cardiovascular education, and many societies and academic centers now deliver teaching through online and hybrid formats alongside mobile health (mHealth) initiatives [1-4]. Within this ecosystem, online video, particularly YouTube (LLC [Google LLC]), has become a major channel for professional and patient education, although its educational contribution remains only partly characterized and is rarely examined using real-world usage data [5]. Experimental studies suggest that short videos and digital cardiology resources can improve knowledge, retention, and patient-reported outcomes [6-8].

These opportunities intersect with persistent inequities. The burden of cardiovascular disease is greatest in low-income countries (LICs) and middle-income countries (MICs), where access to guideline-based care, connectivity, and continuing education remains limited [9-12]. Global connectivity reports describe a large “usage gap”: despite widespread mobile broadband coverage, many clinicians work with unstable electricity, constrained bandwidth, and a strong dependence on personal smartphones as their primary digital device for work and learning [11-13]. In such settings, mobile-first web design and lightweight content are not optional refinements but practical access requirements.

Subtitles and transcripts represent another simple access enabler for video. Same-language captions benefit many learners, including second-language users and people who are deaf or hard of hearing, and automatically generated subtitles can support learning when sufficiently accurate [14,15]. On public platforms such as YouTube, captions and automatic translation could extend the reach of single-language cardiology content across income settings. However, little is known about how often subtitles are actually used in medical or cardiology education, how this use varies with country income, or how it relates to mobile access.

We therefore conducted an exploratory, multicountry ecological analysis of routinely collected learning analytics from a francophone, mobile-first cardiology education website and its YouTube channel. Our two prespecified objectives were (1) to describe, across World Bank income groups, the share of website sessions from mobile devices and the share of YouTube watch time with subtitles enabled and, (2) using only variables measured at the country level, to examine whether greater reliance on mobile access was associated with poorer engagement on the website or on YouTube. We hypothesized that both access indicators would be higher in MICs and LICs than in high-income countries (HICs) and that greater reliance on these access enablers would not be associated with poorer engagement metrics.

Methods

Study Design and Setting

This study was an exploratory, multicountry ecological analysis of routinely collected analytics from a French-language cardiology education website [16] and its companion YouTube channel. The website is an open-access cardiology education resource providing clinical protocols, instructional videos, and practical tools covering diverse cardiology topics. The country was selected as the natural unit of analysis because (i) the World Bank income classification used as the main stratification variable is defined at the country level, and (ii) École Numérique de Cardiologie (ENC; Saint-Denis) platform analytics aggregate user activity at the country level. The ecological design is appropriate for an equity-focused descriptive question; we did not seek to make individual-level inference and acknowledge in the Discussion that ecological associations cannot establish individual-level effects.

All metrics were aggregated over a prespecified two-year window (September 2023 to September 2025). The website was designed using a mobile-first responsive layout: core pages and PDFs were optimized and tested on smartphones (large tap targets, simplified navigation, lightweight media) to support use on small screens and variable bandwidth. No individual-level data were accessed.

Data Sources

Website Analytics

Website analytics were exported from Fathom Analytics (Conva Ventures Inc.), a privacy-focused web analytics platform. For each country, we obtained the number of sessions, unique visitors, and page views, traffic sources (organic search, direct, social, other), bounce rate, and average time on page. Device information was used to calculate the percentage of sessions from mobile devices among sessions with a known device type and the percentage of sessions with an unknown device type. We define mobile devices as smartphones and tablets combined, and this definition is used consistently in Methods, Tables, and Figure captions. For the main content categories (clinical

XSL•FO
RenderX

protocols, practical tools, video pages), we also computed the percentage of sessions from mobile devices.

YouTube Analytics

For each country, we extracted the number of views, total watch time, average view duration, audience retention, the percentage of viewers continuing beyond an introductory segment, views per viewer, and the number and percentage of views classified by the platform as intentional. Intentional views are defined by YouTube as views that the platform classifies as user-driven (eg, search-, suggestion-, or subscription-driven views) rather than impressions or auto-played starts. We selected this metric as a complement to total views because it captures viewer intent and is therefore a meaningful indicator of voluntary engagement with educational content.

A key measurement caveat applies to the caption and demographic indicators introduced here: unlike the country-level website and YouTube metrics described above, they were not available at the country level. Caption-related YouTube metrics (percentage of watch time with any subtitles enabled and percentage with French subtitles) and summary viewer demographics (dominant age band and percentage of male viewers) were available in YouTube Studio only at aggregated geographic levels that corresponded approximately to World Bank income groups, not to individual countries. For each income group (high-, middle-, and low-income) we exported these caption and demographic indicators and assigned the corresponding group-level values to all ENC countries in that group, so that web and YouTube indicators could be tabulated together. As a direct consequence, these caption-related and demographic variables took only three distinct values across the 34-country dataset and showed no within-group variability.

Country Income Classification

Countries were mapped to the World Bank's 2024‐2025 four-level income classification (high-, upper-middle, lower-middle-, and low-income economies), a widely used framework that enables standardized cross-country comparisons and reflects major disparities in digital infrastructure, resource availability, and access conditions that may influence engagement with digital health education interventions. For main analyses, upper- and lower-middle-income categories were collapsed into a single middle-income group, yielding a three-level ordinal variable (high, middle, and low income); the four-level classification was retained for descriptive sensitivity analyses.

Outcome and Access Definitions

Primary access enabler indicators were website mobile access (percentage of sessions from mobile devices among sessions with a known device type) and YouTube caption use (percentage of watch time with any subtitles enabled and any language). Country-level secondary website outcomes, used as country-level proxies of engagement, were bounce rate and average time on page. Country-level secondary YouTube outcomes were average view duration, audience retention, viewer retention beyond the introductory segment, views per viewer, and percentage of intentional views. We also summarized, by income group, mobile share within each website content category, referrer structure, LinkedIn (LinkedIn Corporation) followers and visitors, and website unique visitors; these latter metrics were considered exploratory.

Caption-related YouTube metrics and demographic YouTube metrics were available only at the income-group level. We therefore treated them, throughout the analysis, as descriptive between-group contrasts and we did not include them in country-level inferential analyses. Bounce rate, time on page, average view duration, retention, viewer retention beyond the introductory segment, and intentional views are platform-defined engagement metrics; we explicitly note in the discussion that they are not direct measures of learning, comprehension, implementation, or behavior change. Demographic indicators (dominant viewer age band and percentage of male viewers) were derived from small YouTube Studio aggregates of logged-in viewers with available age and sex data and are descriptive only. In particular, the 100% male value in the low-income group reflects a very small aggregate and does not warrant inferential interpretation.

Data Preprocessing

All analyses were conducted at the country level. Raw exports from each platform were imported into a unified database, nonnumeric placeholders were removed, and percentage variables were placed on a 0‐100 scale where needed. For YouTube, country-level indicators (views, watch time, average view duration, retention, viewer retention beyond the introductory segment, views per viewer, and intentional views) were retained as such. Caption-related indicators (percentage of watch time with any subtitles enabled and percentage with French subtitles) and viewer demographics (age band code, percentage of male viewers) were available only at the income-group level; the same group-level values were therefore shared by all countries within a given income group, and there was no within-group variability in these indicators. For YouTube data, we retained all countries with at least one record and a nonmissing income group. When multiple rows were present for a given country and platform, we computed country-level medians over the observation window. For cross-domain analyses, website and YouTube tables were merged on country and income group so that each country contributed a single combined record. Country-level YouTube engagement metrics (average view duration, audience retention, intentional views, and total views) were included for a given country only when YouTube Studio exported a nonmissing country-level value for that metric over the observation window. Low-volume country accounts without an available country-level value were excluded pairwise. Country-level correlations involving these metrics were therefore based on 24 countries (9 high-, 9 middle-, and 6 low-income; 23 countries for audience retention, with one fewer low-income country), compared with all 34 countries for website-only correlations.

Statistical Analysis

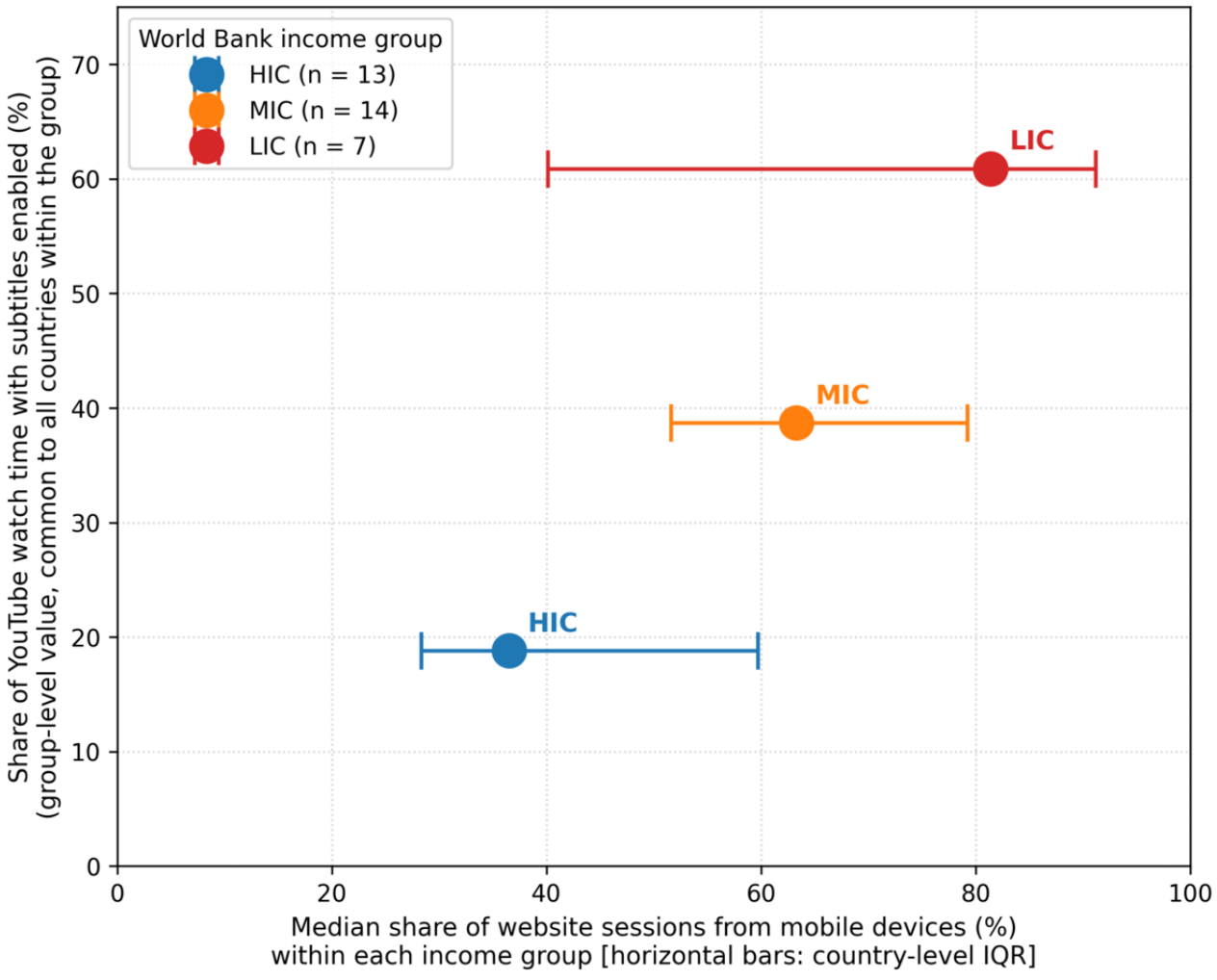
Given the modest sample size and skewed country-level distributions, analyses were prespecified as exploratory and relied on nonparametric tests and robust regression, with emphasis on effect sizes and ordered gradients. For variables measured at the country level (website mobile share, bounce

XSL•FO
RenderX

rate, time on page, and country-level YouTube engagement indicators), we examined gradients across income groups using the Kruskal-Wallis test and tested for a monotonic trend with the Jonckheere-Terpstra test under the prespecified order high<middle<low income. Medians and IQRs by income group are reported for all main metrics. Because most reported percentages are country-level metrics (eg, percentage of sessions from mobile devices and percentage of watch time with captions) that are then summarized as a median or aggregate across countries, rather than proportions computed from a single numerator and denominator, a single absolute count cannot be attached to each percentage; where a discrete count and total are available (eg, number of countries per income group, sample sizes for correlations and regressions, numbers of visitors, views, or followers), both the count and the percentage or total are reported. Country-level Spearman rank correlations were computed between mobile share and engagement outcomes (website bounce rate, time on page; country-level YouTube average view duration, retention, and intentional views). Because our hypotheses were directional (a monotonic gradient across ordered income groups), the Jonckheere-Terpstra trend test was prespecified as the primary test for these ordered comparisons; it is more sensitive to monotonic trends than the omnibus Kruskal-Wallis test, which does not assume an ordering. Where the two tests diverge, the trend test therefore takes precedence.

Caption-related and demographic YouTube metrics, available only at the income-group level, were summarized descriptively as three between-group values; we did not compute country-level correlations or *P* values for these variables, and we did not include them in inferential models. To assess cross-domain coherence between platforms, we summarized, for each income group, the median website mobile share (computed across countries) and the income-group aggregate caption-enabled watch time and compared them descriptively (Figure 1, three-point income-group representation).

Figure 1. Cross-domain coherence between website mobile access and YouTube caption use, by income group (three points). Group-level summary of two access indicators across the three World Bank income groups: median share of website sessions from mobile devices (x-axis, country-level medians within each group, with horizontal bars showing the country-level IQR) and share of YouTube watch time with subtitles enabled (y-axis, group-level value common to all countries within the group). Points correspond to HIC (36.5%, 18.8%), MIC (63.3%, 38.7%), and LIC (81.4%, 60.9%). The figure shows that both access indicators increased monotonically from high- to low-income groups; because YouTube caption-enabled watch time was available only as one value per income group, this representation deliberately avoids a country-level scatter that could misleadingly suggest within-group caption variability. A country-level scatterplot of mobile share against the assigned group-level caption value is provided in Figure S1 in Multimedia Appendix 1, with an explicit note that all country dots within an income group share the same caption value. HIC: high-income country; LIC: low-income country; MIC: middle-income country.



XSL•FO
RenderX

For website engagement we fitted median (τ=0.5) quantile regression models with bounce rate or average time on page as the outcome and income group plus website mobile share (proportion from 0 to 1) as predictors; interaction terms between income group and mobile share were explored in secondary models. LinkedIn followers, LinkedIn visitors, and website unique visitors were summarized by income group and compared using Kruskal-Wallis tests. Robustness of the main gradients was evaluated by repeating key comparisons under the alternative four-level income classification, in leave-one-country-out analyses, and in a sensitivity analysis excluding France (the dominant high-income contributor). *P* values are reported as *P*=.XX or *P*<.001; statistical significance was set at *P*=.05 (two-sided). All analyses were performed in R (R Foundation for Statistical Computing), version 4.5.0.

Reporting Framework

This study is reported in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist for cross-sectional observational studies. The completed STROBE checklist is provided as Checklist 1.

Ethical Considerations

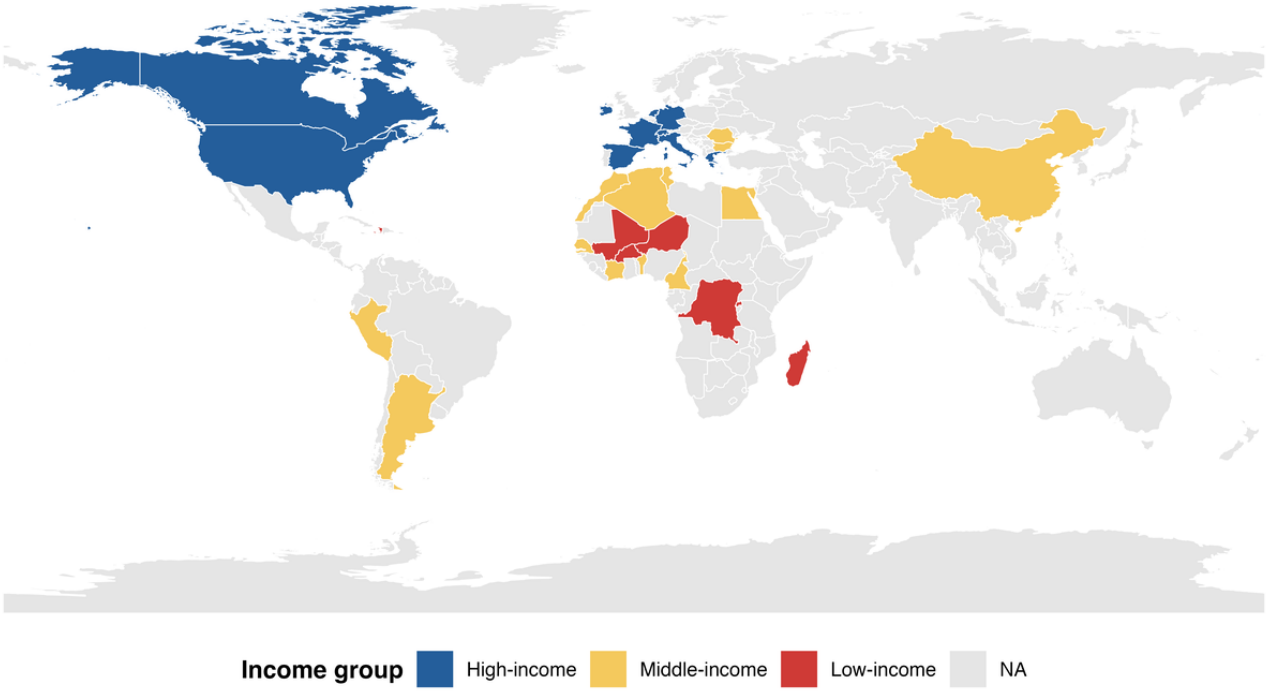
The study used only aggregated, country-level web analytics without any directly or indirectly identifiable personal data and did not involve any intervention or contact with patients or health professionals. Under French law on research involving human participants ("loi Jardé") [17] and the French Public Health Code [18], such secondary analyses of fully anonymized, aggregated data do not require review by a Comité de Protection des Personnes or individual informed consent. The project complied with applicable French and European data-protection requirements, including the General Data Protection Regulation.

Results

Country Sample

Web and YouTube analytics from the ENC were available for 34 countries: 13/34 (38%) HICs, 14/34 (41%) MICs, and 7/34 (21%) LICs. The audience was predominantly but not exclusively francophone, spanning Europe, North America, Latin America, North and Sub-Saharan Africa, and Asia (Figure 2).

Figure 2. Included countries by World Bank income group. World map showing the 34 countries with École Numérique de Cardiologie (ENC) web and YouTube analytics, colored by World Bank income group (high-, middle-, or low-income). Countries without ENC data or without a mapped income category are shown in gray (not available [NA]). The map highlights a predominantly francophone yet geographically diverse reach spanning Europe, North America, Latin America, North Africa, Sub-Saharan Africa, and Asia over the fixed two-year observation window.



Income-Group Descriptive Contrasts: Caption Use and Viewer Demographics on YouTube

Caption-related and demographic YouTube metrics were available only as income-group aggregates and therefore took only three distinct values across the dataset; the contrasts below are presented as descriptive between-group differences without country-level *P* values or correlation coefficients.

Over the two-year study window, the ENC website attracted 34,340 unique visitors across 34 countries, generating 180,563 page views, while the companion YouTube channel accumulated 171,009 views and 7,451 hours of watch time.

The income-group aggregate share of YouTube watch time with any subtitles enabled increased monotonically with decreasing income, from 18.8% in HICs to 38.7% in MICs and 60.9% in LICs (Table 1, Panel A; Figure 3). As a percentage of total watch time within each income-group aggregate, rather than an

XSL•FO
RenderX

average across countries or videos, this share is inherently weighted by watch time. This corresponds to a caption equity gap (CEG) of 42.1% (60.9%-18.8%) between LICs and HICs. The share of watch time with French captions and with non-French captions varied across income groups (HICs: 2.3% French, 16.5% non-French; MICs: 0.3% French, 38.4% non-French; LICs: 17.5% French, 43.4% non-French), with non-French captions accounting for the largest share of caption-enabled watch time in middle- and low-income groups (Table 1, Panel A).

Across income groups, dominant viewer age band was older in HICs (≥65 years) than in MICs (25 - 35 years) and LICs (45 - 55 years); however, this pattern was not monotonic with caption-enabled watch time, as LICs had a higher caption-enabled watch time than MICs but an older dominant age band (Table 1, Panel A bis; see Limitations).

Table . Caption use, viewer demographics, mobile web access, and engagement by World Bank income group.[a]

Income group	Panel A[b]—income-group-level YouTube indicators			Panel A bis—income-group-level demographic indicators		Panel B[c]—country-level indicators, median (IQR); n countries					
	Caption-enabled watch time (any subtitles), %	French captions, %	Non-French captions, %	Dominant viewer age band, years	Male viewers, %	Mobile share, %	Bounce rate, %	Time on page, s	YouTube AVD[d], min	YouTube retention, %	n countries
High-income (HIC[e])	18.8	2.3	16.5	≥65	62	36.5 (28.3 - 59.7)	65.5 (60.8 - 68.5)	43.2 (39.2 - 49.4)	2.30 (2.05 - 2.82)	36.0 (27.0 - 40.0)	13
Middle-income (MIC[f])	38.7	0.3	38.4	25 - 35	46	63.3 (51.6 - 79.2)	58.5 (49.0 - 66.3)	46.4 (33.8 - 70.9)	2.55 (2.28 - 2.78)	36.0 (34.0 - 45.0)	14
Low-income (LIC[g])	60.9	17.5	43.4	45 - 55	100[h]	81.4 (40.1 - 91.2)	51.2 (43.5 - 59.0)	74.2 (52.8 - 103.6)	3.56 (3.31 - 4.44)	49.0 (37.0 - 53.0)	7

[a]Demographic values are derived from small YouTube Studio aggregates of logged-in viewers and are descriptive only (see *Methods*, *Outcome and Access Definitions* section).

[b]Panel A reports YouTube indicators that were available only at the income-group level in YouTube Studio and were assigned to all countries within an income group; these values are therefore identical across all countries of the same group and are reported descriptively, without country-level *P* values or correlation coefficients.

[c]Panel B reports country-level medians and IQRs. Mobile share: percentage of website sessions from mobile devices (smartphones or tablets) among sessions with a known device type. Bounce rate: proportion of single-page sessions, expressed as percentage. Time on page: average time spent on a page (seconds). YouTube AVD: average view duration (minutes); retention: average percentage of each video watched. The “n countries” column applies to mobile share, bounce rate, and time on page (13 HICs, 14 MICs, 7 LICs; 34 total). YouTube AVD was available for a smaller subset of countries (9 HICs, 9 MICs, 6 LICs; 24 total) and YouTube retention for a further reduced subset (9 HICs, 9 MICs, 5 LICs; 23 total) (Table S1 in Multimedia Appendix 2 for per-metric denominators).

[d]AVD: average view duration.

[e]HIC: high-income country.

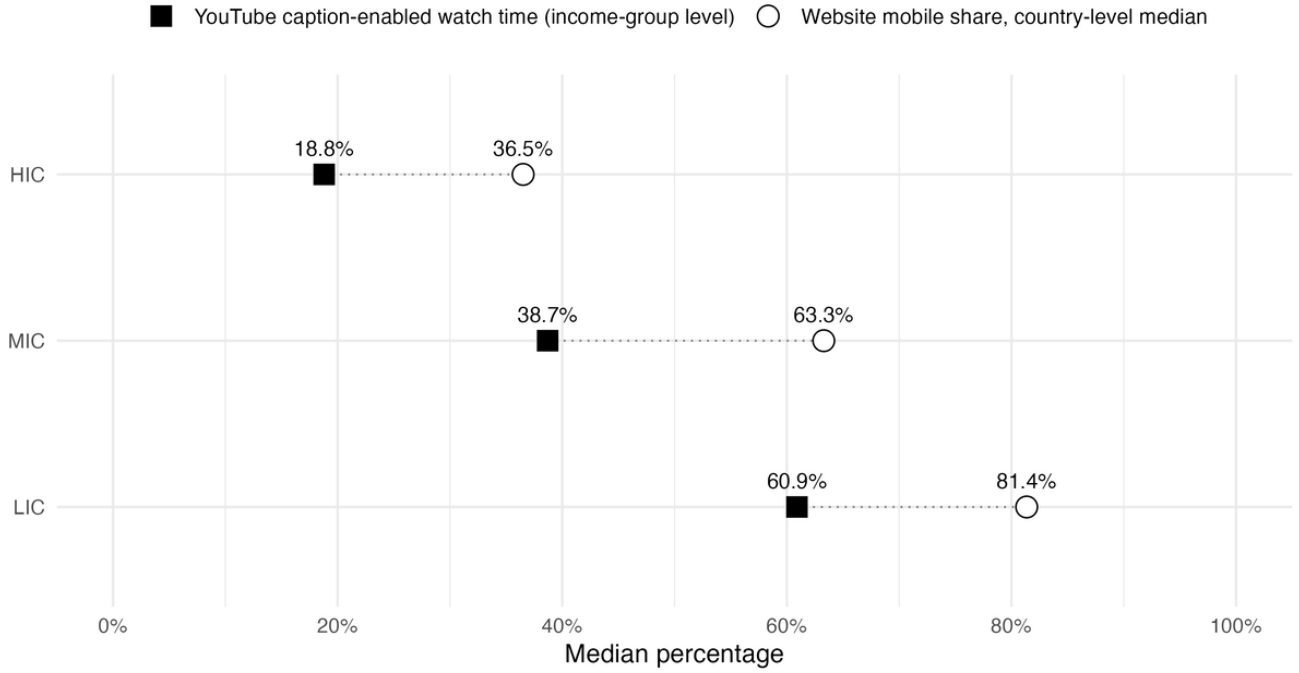
[f]MIC: middle-income country.

[g]LIC: low-income country.

[h]The low-income value is based on a very small aggregate; see *Method*s section.

XSL•FO
RenderX

Figure 3. Access enablers by income group: YouTube captions and website mobile access. Income-group aggregate share of YouTube watch time with subtitles enabled (square marker, weighted by watch time within each income-group aggregate) and median share of ENC website sessions from mobile devices among sessions with a known device type (open circle) for each World Bank income group (HIC, MIC, LIC). The dotted line within each income group is provided for visual comparison only between two distinct indicators (caption-enabled watch time and website mobile share); it does not represent an uncertainty interval, a temporal change, or any modeled relationship. Caption-enabled watch time was 18.8%, 38.7%, and 60.9% in HICs, MICs, and LICs, respectively, corresponding to a caption equity gap of 42.1% between LICs and HICs. Website mobile share showed a similar gradient (36.5%, 63.3%, and 81.4%, respectively). Caption-related metrics were available only at the income-group level (one value per group, common to all countries in the group). HIC: high-income country; LIC: low-income country; MIC: middle-income country.



Country-Level Analyses: Mobile Web Access and Engagement

Website mobile share, measured at the country level, increased with decreasing income (Table 1, Panel B; Figure 3). The median mobile share of ENC website sessions was 36.5% (IQR 28.3 - 59.7) in HICs, 63.3% (IQR 51.6 - 79.2) in MICs, and 81.4% (IQR 40.1 - 91.2) in LICs. The prespecified ordered trend test supported a monotonic increase in mobile share from HICs to MICs to LICs (Jonckheere-Terpstra P=.01; Kruskal-Wallis H=5.11; P=.08). The proportion of website sessions from devices of unknown type was negligible (median 0% in HICs, MICs, and LICs; range 0% - 0% in each group); no country exceeded the 40% threshold used for exclusion from mobile-share analyses, so the mobile-share denominator effectively comprised all sessions and could not materially bias the observed gradient.

Country-Level Spearman Correlations Involving Mobile Share

Country-level Spearman correlations between website mobile share and country-level engagement outcomes were small in magnitude and not statistically significant (mobile share vs bounce rate: ρ=−0.05, P=.76, n=34; vs time on page: ρ=0.07, P=.71, n=34; vs YouTube average view duration: ρ=0.05, P=.80, n=24; vs YouTube audience retention: ρ=−0.03, P=.88, n=23; vs YouTube intentional views: ρ=0.28, P=.18, n=24). For reference, bounce rate and time on page were inversely correlated at the country level (ρ=−0.48, P=.004, n=34; Table 2).

XSL•FO
RenderX

Table . Country-level Spearman correlations between website mobile share and country-level engagement outcomes.[a]

Pair of country-level variables	Spearman ρ	*P* values	N
Website mobile share vs bounce rate	−0.05	.76	34
Website mobile share vs time on page	0.07	.71	34
Website mobile share vs YouTube average view duration (country-level)	0.05	.80	24
Website mobile share vs YouTube audience retention (country-level)	−0.03	.88	23
Website mobile share vs YouTube intentional views, % (country-level)	0.28	.18	24
Bounce rate vs time on page	−0.48	.004	34
Website mobile share vs YouTube total views (country-level)	−0.06	.79	24

[a]All variables are measured at the country level; correlations are based on country-level complete cases for each pair. Caption-related and demographic YouTube metrics are not included in this table because they were available only at the income-group level (see *Methods* and Table 1 Panel A); they are reported descriptively in the *Results*. Country-level YouTube metrics were available for 24 of 34 countries (high-income: 9/13; middle-income: 9/14; low-income: 6/7; audience retention: 23/34, with 5/7 low-income countries); website-only correlations include all 34 countries. Coverage was proportionally similar across income groups, making substantial selection bias unlikely.

Median Quantile Regression for Web Engagement

In median (τ=0.5) quantile regression models for web engagement outcomes (Table 3), income group was associated with bounce rate but not with time on page. For bounce rate, the linear contrast for income (ordered from high- to low-income) was negative (β=−0.130; 95% CI −0.219 to −0.041; *P*=.01), consistent with lower median bounce rates in lower-income countries. The quadratic contrast was small and not statistically significant (β=−0.023; 95% CI −0.119 to 0.072; *P*=.64). For median time on page, neither the linear nor the quadratic income contrasts reached statistical significance (β=21.2 s, 95% CI –11.8 to 54.1, *P*=.22; β=12.6 s, 95% CI −18.6 to 43.8, *P*=.44). Website mobile share, modeled as a 0 - 1 proportion, was not independently associated with either outcome (bounce rate: +0.072, 95% CI −0.128 to 0.272, *P*=.48; time on page:+16.0 s, 95% CI −44.8 to 76.8, *P*=.61). Overall, these analyses did not provide evidence that countries relying more heavily on mobile access experienced systematically worse website engagement.

Table . Median (τ=0.5) quantile regression of web engagement outcomes on income group (polynomial contrasts) and mobile share (n=34 countries).[a]

Outcome and predictor	Coefficient β	95% CI	*P* values
Median bounce rate (proportion of single-page sessions)			
Intercept	0.554	0.425 to 0.684	<.001
Income group (linear contrast)	−0.130	−0.219 to −0.041	.01
Income group (quadratic contrast)	−0.023	−0.119 to 0.072	.64
Website mobile share (proportion, 0 - 1)	0.072	−0.128 to 0.272	.48
Median time on page (seconds)			
Intercept	47.1	9.7 to 84.5	.02
Income group (linear contrast)	21.2	−11.8 to 54.1	.22
Income group (quadratic contrast)	12.6	−18.6 to 43.8	.44
Website mobile share (proportion, 0 - 1)	16.0	−44.8 to 76.8	.61

[a]Median (τ=0.5) quantile regression models were fitted at the country level for two web engagement outcomes: bounce rate (proportion of sessions with only one page viewed) and median time on page (seconds). Income group was entered as an ordered predictor using polynomial contrasts (high-, middle-, low-income); the “linear contrast” captures the overall gradient from high- to low-income, with negative coefficients indicating lower values in lower-income settings. Website mobile share was modeled as a proportion from 0 to 1 (0=0% of sessions on mobile; 1=100% on mobile), so the corresponding coefficients represent the expected change in the median outcome across this range. In these models, income group showed a modest association with bounce rate, whereas mobile share was not independently associated with either bounce rate or time on page (all *P*≥.48).

XSL•FO
RenderX

Sensitivity Analyses

The income-related gradient in website mobile share was robust to several prespecified checks (Table S5 in Multimedia Appendix 3). In leave-one-country-out analyses, the Jonckheere-Terpstra trend test for mobile share remained significant in all 34 reestimations (all $P<.05$; largest $P=.03$). In a sensitivity analysis excluding France, the largest high-income contributor, the gradient was essentially unchanged (mobile share Jonckheere-Terpstra $P=.02$; Kruskal-Wallis H=4.94, $P=.08$). In median quantile regression models that added an income-by-mobile-share interaction, no interaction term reached statistical significance for either bounce rate or time on page (all $P>.05$; interaction coefficients and 95% CIs are reported in Table S5 in Multimedia Appendix 3), providing no evidence of an interaction between income group and mobile share for either outcome; given the small sample size, this nonsignificant result does not establish that the absence of a mobile-access penalty is equivalent across income groups.

YouTube Engagement Metrics by Income Group

Descriptively, country-level YouTube engagement metrics tended to be higher in lower-income groups (Table S1 in Multimedia Appendix 2). Median average view duration increased from 2.30 min in HICs to 2.55 min in MICs and 3.56 min in LICs, but neither the Kruskal-Wallis test (H=4.31, $P=.12$) nor the Jonckheere-Terpstra trend test ($P=.06$) reached statistical significance. Median audience retention rose from 36% (HICs) and 36% (MICs) to 49% (LICs), but Kruskal-Wallis and Jonckheere-Terpstra tests did not show statistically significant gradients (H=1.40, $P=.50$; Jonckheere-Terpstra $P=.14$).

Cross-Domain Coherence Between Mobile Web Access and Caption Use

The income-group median of website mobile share and the income-group aggregate caption-enabled watch time both increased with decreasing income (Figure 1). To match the data structure, Figure 1 of the main paper shows one point per income group (HICs, MICs, LICs); a country-level scatterplot is provided Figure S1 in Multimedia Appendix 1 with an explicit note that caption-related values are common to all countries within an income group and that the country-level pattern therefore reflects between-group contrasts rather than within-group variability.

Mobile Access by Content Category

Mobile reliance was high across all content categories outside high-income settings (Table 4, Table S3 in Multimedia Appendix 4). In HICs, medians for the share of mobile sessions were 37.7% for protocol pages, 40% for tool pages, and 52.2% for video pages, indicating relatively greater mobile use for video content. In MICs, medians increased to 66.1% for protocols, 60.5% for tools, and 77.1% for videos. In LIC mobile shares were 72.1% for protocols, 75% for tools, and 70% for videos. Even content typically consulted from a desktop in high-income settings (protocols, downloadable tools) was predominantly accessed from mobile devices in many MICs and LICs. The four-level World Bank classification yielded a broadly similar pattern, with mobile shares exceeding 60% in lower-middle- and low-income groups across all content categories and in upper-middle-income protocol and video pages, although upper-middle-income tool pages were lower (34.8%) (Table S3 in Multimedia Appendix 4).

Table . Mobile share of web sessions by content category and income group.[a]

Income group	Protocol pages — mobile sessions (% of known devices), median (IQR)	Tool pages — mobile sessions (% of known devices), median (IQR)	Video pages — mobile sessions (% of known devices), median (IQR)
High-income (HIC[b])	37.7 (19.3-58.6)	40.0 (34.7-61.2)	52.2 (39.1-61.0)
Middle-income (MIC[c])	66.1 (43.9-86.3)	60.5 (41.4-73.7)	77.1 (64.7-91.9)
Low-income (LIC[d])	72.1 (44.3-95.6)	75.0 (31.2-88.5)	70.0 (56.6-90.6)

[a]For each country and content category (protocols, tools, videos), mobile share is the proportion of École Numérique de Cardiologie website sessions with a known device type that originated from mobile devices (smartphones or tablets). Values are medians across countries within each World Bank income group (three-level classification); only countries with at least one session in each category contribute to that median.

[b]HIC: high-income country.

[c]MIC: middle-income country.

[d]LIC: low-income country.

Referrer Mix, LinkedIn Activity, and Website Reach

Referrer patterns varied by income group but were broadly dominated by organic and direct access (Table S2 in Multimedia Appendix 5 and Table S4 in Multimedia Appendix 6). At the country-median level, organic search accounted for 33% of sessions in HICs, 38.2% in MICs, and 43.5% in LICs; direct access accounted for 39.6%, 22.7%, and 20.5%, respectively. Social referrers contributed more in MICs (median 29.1%) than in HICs (14.3%) or LICs (12.3%). Median numbers of LinkedIn followers per country were 17 in HICs, 9 in MICs, and 5.5 in LICs, with no significant differences between income groups (H=3.19, $P=.20$). Median monthly LinkedIn visitors per country were also low across all income groups (H=0.01, $P=.91$). Median counts of unique ENC website visitors per country were highest in HICs (278), followed by MICs (93) and LICs (43; Table S2 in Multimedia Appendix 5). The country-level Spearman correlation between unique visitors and website mobile share was negative but not statistically significant ($\rho=-0.28$, $P=.10$, n=34).

XSL•FO
RenderX

Discussion

Principal Findings

In this exploratory ecological analysis of country-level learning analytics from a responsive francophone cardiology education website and its companion YouTube channel, we observed marked descriptive gradients in two practical access indicators across World Bank income groups. The share of website sessions from mobile devices and the share of YouTube watch time with captions enabled were substantially higher in middle- and especially LICs than in HICs, corresponding to a CEG of approximately 40% in caption-enabled watch time between low- and high-income settings. Because caption-related metrics were available only as income-group aggregates, these gradients should be interpreted descriptively rather than as precise country-level estimates, although the pattern was consistent and large. At the country level, greater reliance on mobile web access was not associated with measurable penalties in basic engagement metrics, including bounce rate, time on page, or country-level YouTube engagement indicators (all country-level Spearman $|\rho| \leq 0.28$; all $P \geq .18$).

Interpretation in Relation to Prior Work

These findings build on previous work on digital cardiology education, mHealth, and digital health equity. Major cardiology societies have emphasized the rapid growth of online congresses, courses, and blended training, and the central role of video in modular, on-demand learning [1-3,5-7]. Digital cardiology patient education interventions have likewise improved knowledge and patient-reported outcomes, supporting the value of carefully designed digital content beyond professional training alone [8]. Reviews and bibliometric analyses of digital and blended medical education describe the broader shift toward modular, on-demand formats integrated with video and social media [19-21]. At the same time, digital health frameworks and global connectivity reports highlight a persistent usage gap. Within this context, clinicians in LICs and MICs often rely on personal smartphones, unstable electricity, and constrained bandwidth [4,10-13]. The strong gradients we observed in mobile access and caption use are consistent with this picture of a digitally connected but resource-constrained workforce. Our study contributes to this literature in two ways. First, it links routine multicountry learning analytics to these equity concerns in a cardiology-specific, predominantly francophone context. Second, it focuses on modifiable design features, namely mobile-first website design and systematic captioning, rather than on bespoke mHealth applications alone.

Our results on subtitles align with experimental and applied work showing that captions can improve comprehension and information retention, particularly for second-language users and people who are deaf or hard of hearing [14,15]. Accessibility guidelines also identify captions, transcripts, and basic player controls as core requirements for inclusive media [22]. In our project, audio was in French and non-French subtitles relied substantially on YouTube automatic captioning and translation. Caption-enabled watch time is therefore likely to reflect a mix of creator decisions, platform features, and user activation. Despite these limitations, the higher caption-enabled watch-time share in lower-income groups suggests that subtitles may be an important access enabler for clinicians and trainees in these settings, even though we could not directly assess subtitle accuracy or individual learning outcomes. Taken together, our data suggest that mobile-optimized web access and captioned video may represent access-supporting features rather than trade-offs in basic engagement metrics.

Implications for Design and Practice

From a design standpoint, these findings support treating mobile-optimized web layouts and captioned video as baseline features for cardiology education initiatives that seek international reach, rather than as optional enhancements. For clinicians and trainees in resource-constrained settings, responsive design, lightweight pages, and subtitles may reduce both technical and cognitive barriers to using educational material at the point of care [4,10,13,23]. Simple indicators such as mobile share, captioned watch time (including non-French captions), and geographic reach can be monitored as a low-cost equity dashboard alongside traditional metrics such as participation and satisfaction. For projects like ENC, practical next steps may include a lightweight mobile application with offline access to key protocols and tools, and a progressive professional translation of a core set of videos that builds upon, but does not rely solely on, automatic captioning. Professional societies and academic centers could similarly encourage contributors to provide mobile- and caption-ready content and to follow basic accessibility standards [22,24]. In HICs, the mobile gradient was milder, consistent with a dual-use pattern. Clinicians in these settings likely use desktop computers for prolonged work sessions involving protocols and tools, while relying on smartphones for quick lookups. In resource-constrained settings, however, clinicians and trainees appear to rely predominantly on smartphones as their primary device for both types of content.

Engagement Metrics as Proxies for Educational Outcomes

Bounce rate, average time on page, average view duration, audience retention, viewer retention beyond the introductory segment, and intentional views are useful, routinely available platform metrics that provide insight into how frequently and for how long users engage with educational content. They are not, however, direct measures of learning, comprehension, implementation, or changes in clinical practice. On the one hand, longer time on page may reflect engagement, but it may also reflect slower mobile loading, divided attention, or difficulty understanding the content. On the other hand, a lower bounce rate may reflect interest, while it may also reflect a search for information that was not immediately found. Caption-enabled watch time indicates that subtitles were activated, but cannot determine whether the displayed text was accurate, comprehensible, or helpful to individual viewers. We therefore interpret these metrics as exploratory indicators of equitable reach and engagement rather than as proxies for educational outcomes. Linking these platform analytics with learner-level assessments of knowledge acquisition, perceived usefulness, and impact on clinical practice should be a priority for future research.

XSL•FO
RenderX

Strengths and Limitations

This study has several strengths. It leverages routinely collected platform analytics from a real-world cardiology education initiative with a multicountry audience spanning high-, middle-, and low-income settings, and combines web and YouTube data from the same project to provide a rare cross-domain view of access enablers. Analyses were prespecified, used nonparametric methods appropriate for skewed country-level distributions, and focused on effect sizes and ordered gradients rather than solely on *P* values. By relying on simple indicators that are available on most web and video analytics dashboards, the approach is readily transferable to other mHealth and education projects with limited evaluation resources.

Important limitations must be acknowledged. The ecological design, with countries as the unit of analysis, precludes individual-level inference and is prone to ecological fallacy; we cannot identify which types of users drive the observed patterns or how usage relates to learning or clinical behavior. Caption-related and demographic YouTube metrics were available only as income-group aggregates and took only three distinct values across the dataset. They were therefore reported strictly as descriptive between-group contrasts, limiting the granularity of these analyses and precluding the inclusion of caption-related variables in country-level inferential models. The audience is self-selected and primarily francophone, which limits generalizability to other languages and likely underrepresents some large LICs and MICs. Our findings are therefore best interpreted as a francophone case study rather than a globally representative sample. Finally, web and platform metrics such as bounce rate, session duration, and caption-enabled watch time capture only limited aspects of engagement and are shaped by platform-specific definitions. In addition, the cross-sectional design did not allow us to examine temporal trends.

Country-level YouTube engagement metrics were available for 24 of 34 countries because YouTube Studio did not return stable country-level estimates for countries with fewer views. Coverage was proportionally similar across income groups (high-income 9/13, middle-income 9/14, low-income 6/7), making substantial selection bias unlikely. Nevertheless, the absolute number of LICs contributing YouTube metrics remained small. The association between caption use and viewer age was also not monotonic across income groups (older dominant age band in LICs than in MICs despite higher caption-enabled watch time in LICs), which tempers the equity narrative built around captions and should be interpreted cautiously.

Conclusions

This exploratory ecological case study examined a responsive francophone cardiology education website and its companion YouTube channel. We observed marked descriptive gradients in mobile web access and caption-enabled video watch time across country income groups. The highest values were observed in lower-income settings. A CEG of approximately 40% was identified between LICs and HICs. At the country level, greater reliance on mobile devices was not associated with meaningful penalties in basic web engagement metrics. These findings are consistent with treating mobile-optimized design and systematic use of captions, including non-French subtitles, as core, low-cost, access-supporting components of equitable digital cardiology education, while emphasizing that platform engagement indicators are not direct measures of learning. Simple platform analytics may provide useful signals about global reach and inclusiveness and should, in future work, be linked with learner-level outcomes and evaluations of mHealth interventions.

Acknowledgments

The authors thank all collaborators, colleagues, and learners who contributed to the creation, growth, and continuous development of the École Numérique de Cardiologie (ENC) platform. Their commitment and collective effort have been instrumental in expanding access to digital cardiology education across diverse regions and practice settings. The authors also thank Elise Cordonnier and Sarah Delmas for their careful English language review and helpful editorial suggestions during manuscript revision.

During revision of this manuscript, the authors used a generative AI tool (Claude [Anthropic] and ChatGPT [OpenAI]) to assist with language editing, drafting of the response to reviewers, and consistency checks across sections. All scientific content, statistical analyses, interpretation, figures, tables, and final wording were performed and verified by the authors, who take full responsibility for the content of the manuscript.

The members of École Numérique de Cardiologie (ENC):

- Louis-Marie Desroche, MD — President; Cardiology Department, University Hospital of La Réunion, Saint-Denis, La Réunion, France

- Jérôme Corré, MD — Vice-President; Cardiology Department, University Hospital of La Réunion, Saint-Denis, La Réunion, France

- Thiziri Si Moussi, MD — Treasurer; Department of Cardiology, Hôpital Saint-Joseph, Marseille, France

- Marie-Lucie Jean-Jacques, RN — Secretary, and lead of the Procedure Packs project; Cardiology Department, University Hospital of La Réunion, Saint-Denis, La Réunion, France

XSL•FO
RenderX

- Benoit Merat, MD — Protocols lead; Cardiology Department, Percy Army Training Hospital (Hôpital d'Instruction des Armées Percy), Clamart, France

- Shamir Vally, MD — Videos lead; Cardiology Department, University Hospital of La Réunion, Saint-Denis, La Réunion, France

- Louis Perrard, MD — Tools lead; Cardiology Department, Amiens University Hospital (CHU Amiens-Picardie), Amiens, France

- Jean-Bernard Pausé, RN — Short-format content lead; Cardiology Department, University Hospital of La Réunion, Saint-Denis, La Réunion, France

- Guillaume Bonnet, MA — Scientific partnerships lead; Cardiology Department, Bordeaux University Hospital (CHU de Bordeaux), Bordeaux, France

- Élodie Benhaïm, MPH — Strategy and Development lead; Department of Research and Innovation, University Hospital of La Réunion, Saint-Paul, La Réunion, France

- Reza Rossanaly Vasram, MD — Treasurer and editorial contributor throughout the study period (until April 2026); Cardiology Department, University Hospital of La Réunion, Saint-Denis, La Réunion, France

- Arthur Darmon, MD — Deputy Secretary and contributor to the ENC clinical protocols throughout the study period (until April 2026); Interventional Cardiology, Centre Cardiologique du Nord, Saint-Denis, France

Funding

This study received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

The study relied on aggregated, country-level analytics exported from the ENC website and YouTube channel. Deidentified CSV files containing the country-level dataset, the analysis scripts written in R, and the code used to produce all tables and figures are available from the corresponding author upon reasonable request, subject to platform terms of use (Fathom Analytics for the website, YouTube Studio for the YouTube channel) and institutional data-protection requirements. No individual-level data were collected, exchanged, or analyzed.

Authors' Contributions

LMD contributed to conceptualization, methodology, software, formal analysis, investigation, data curation, writing – original draft, writing – review & editing, visualization, supervision, and project administration. TR contributed to investigation, data curation, writing – review & editing, and project administration. EB contributed to methodology, investigation, writing – review & editing. LP, BM, and SV contributed to investigation (École Numérique de Cardiologie [ENC] platform development and content) and writing – review & editing. RG and PD contributed to conceptualization and writing – review & editing. ENC collaborators (JC, RV, TSM, AD, JBP, LP, GB, BM, SV, HB) contributed to the creation, implementation, and continuous improvement of the ENC platform and its educational content. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Country-level scatterplot of website mobile share vs assigned income-group caption value. Country-level scatterplot of École Numérique de Cardiologie website mobile share (x-axis, percentage of sessions from mobile devices among sessions with a known device type) vs the income-group caption value assigned to that country (y-axis). Each point represents one country (n=34). Caption-enabled watch time was available only at the income-group level; therefore, all country points within a given income group share the same caption value (high-income country=18.8%, middle-income country=38.7%, low-income country=60.9%). The country-level pattern in this figure reflects between-group contrasts only, not within-group variability in caption use. We do not report country-level correlation coefficients on this figure.
[DOCX File, 171 KB - mhealth_v14i1e90345_app1.docx]

Multimedia Appendix 2
YouTube engagement metrics (average view duration, audience retention, intentional views, total views) by World Bank income group, reported as country-level medians (IQR).
[DOCX File, 14 KB - mhealth_v14i1e90345_app2.docx]

XSL•FO
RenderX

Multimedia Appendix 3
Prespecified, exploratory sensitivity analyses assessing the robustness of the income gradient in website mobile share, using percentile bootstrap (5000 resamples).
[DOCX File, 14 KB - mhealth_v14i1e90345_app3.docx]

Multimedia Appendix 4
Mobile share of website sessions by content category (protocols, tools, videos) and four-level World Bank income group, reported as country-level medians.
[DOCX File, 14 KB - mhealth_v14i1e90345_app4.docx]

Multimedia Appendix 5
Website referrer mix, LinkedIn activity, and website reach by World Bank income group (n=34 countries).
[DOCX File, 14 KB - mhealth_v14i1e90345_app5.docx]

Multimedia Appendix 6
Website referrer mix by four-level World Bank income group, reported as country-level medians.
[DOCX File, 14 KB - mhealth_v14i1e90345_app6.docx]

Checklist 1
STROBE checklist.
[DOCX File, 33 KB - mhealth_v14i1e90345_app7.docx]

References

1. Rosselló X, Stanbury M, Beeri R, Kirchhof P, Casadei B, Kotecha D. Digital learning and the future cardiologist. Eur Heart J 2019 Feb 7;40(6):499-501. [doi: 10.1093/eurheartj/ehy884] [Medline: 30698713]
2. Fox K, Price S, Achenbach S, et al. The European Society of Cardiology - a digital educator. J Eur CME 2021;10(1):2014039. [doi: 10.1080/21614083.2021.2014039] [Medline: 34912587]
3. Cullen MW, Geske JB, Anavekar NS, et al. Reinvigorating continuing medical education: meeting the challenges of the digital age. Mayo Clin Proc 2019 Dec;94(12):2501-2509. [doi: 10.1016/j.mayocp.2019.07.004] [Medline: 31806103]
4. Maddula R, MacLeod J, McLeish T, et al. The role of digital health in the cardiovascular learning healthcare system. Front Cardiovasc Med 2022;9:1008575. [doi: 10.3389/fcvm.2022.1008575] [Medline: 36407438]
5. Curran V, Simmons K, Matthews L, et al. YouTube as an educational resource in medical education: a scoping review. MedSciEduc 2020 Dec;30(4):1775-1782. [doi: 10.1007/s40670-020-01016-w]
6. Krumm IR, Miles MC, Clay A, Carlos Ii WG, Adamson R. Making effective educational videos for clinical teaching. Chest 2022 Mar;161(3):764-772. [doi: 10.1016/j.chest.2021.09.015] [Medline: 34587482]
7. Youssef SC, Aydin A, Canning A, Khan N, Ahmed K, Dasgupta P. Learning surgical skills through video-based education: a systematic review. Surg Innov 2023 Apr;30(2):220-238. [doi: 10.1177/15533506221120146] [Medline: 35968860]
8. Oudkerk Pool MD, Hooglugt JLQ, Schijven MP, et al. Review of digitalized patient education in cardiology: a future ahead? Cardiology 2021;146(2):263-271. [doi: 10.1159/000512778] [Medline: 33550295]
9. Smith B, Magnani JW. New technologies, new disparities: the intersection of electronic health and digital health literacy. Int J Cardiol 2019 Oct 1;292:280-282. [doi: 10.1016/j.ijcard.2019.05.066] [Medline: 31171391]
10. Ogungbe O, Longenecker CT, Beaton A, et al. Advancing cardiovascular health equity globally through digital technologies. J Am Heart Assoc 2024 Jan 16;13(2):e031237. [doi: 10.1161/JAHA.123.031237] [Medline: 38226506]
11. Measuring digital development: facts and figures 2024. International Telecommunication Union. 2024. URL: https://www.itu.int/itu-d/reports/statistics/facts-figures-2024 [accessed 2026-08-25]
12. The state of mobile internet connectivity report 2024. : GSMA Intelligence; 2024 Oct URL: https://www.gsmaintelligence.com/research/the-state-of-mobile-internet-connectivity-2024 [accessed 2026-08-31]
13. Li W, Gillies R, He M, et al. Barriers and facilitators to online medical and nursing education during the COVID-19 pandemic: perspectives from international students from low- and middle-income countries and their teaching staff. Hum Resour Health 2021 May 12;19(1):64. [doi: 10.1186/s12960-021-00609-9] [Medline: 33980228]
14. Gernsbacher MA. Video captions benefit everyone. Policy Insights Behav Brain Sci 2015 Oct;2(1):195-202. [doi: 10.1177/2372732215602130] [Medline: 28066803]
15. Malakul S, Park I. The effects of using an auto-subtitle system in educational videos to facilitate learning for secondary school students: learning comprehension, cognitive load, and satisfaction. Smart Learn Environ 2023;10(1):4. [doi: 10.1186/s40561-023-00224-2] [Medline: 40477865]
16. École numérique de cardiologie (ENC). ENC. URL: https://www.enc-cardiologie.fr [accessed 2026-08-25]
17. LOI n° 2012-300 du 5 mars 2012 relative aux recherches impliquant la personne humaine. Journal officiel de la République française. 2012 Mar 6. URL: https://www.legifrance.gouv.fr/loda/id/JORFTEXT000025441587/ [accessed 2026-08-25]

XSL•FO
RenderX

18. Code de la santé publique. Livre Ier, Titre II: Recherches impliquant la personne humaine Articles L1121-1 et suivants. Légifrance. URL: https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000046125746 [accessed 2026-08-25]
19. Yeung AWK, Parvanov ED, Hribersek M, et al. Digital teaching in medical education: scientific literature landscape review. JMIR Med Educ 2022 Feb 9;8(1):e32747. [doi: 10.2196/32747] [Medline: 35138260]
20. Delungahawatta T, Dunne SS, Hyde S, et al. Advances in e-learning in undergraduate clinical medicine: a systematic review. BMC Med Educ 2022 Oct 7;22(1):711. [doi: 10.1186/s12909-022-03773-1] [Medline: 36207721]
21. Katz M, Nandi N. Social media and medical education in the context of the COVID-19 pandemic: scoping review. JMIR Med Educ 2021 Apr 12;7(2):e25892. [doi: 10.2196/25892] [Medline: 33755578]
22. Audio content and video content. World Wide Web Consortium Web Accessibility Initiative. 2023. URL: https://www.w3.org/WAI/media/av/av-content/ [accessed 2026-08-25]
23. Hearn J, Wali S, Birungi P, et al. A digital self-care intervention for Ugandan patients with heart failure and their clinicians: user-centred design and usability study. Digit Health 2022;8:20552076221129064. [doi: 10.1177/20552076221129064] [Medline: 36185389]
24. Koskinas KC, Wilhelm M, Halle M, et al. The future of education in Preventive Cardiology: a statement of the European Association of Preventive Cardiology of the European Society of Cardiology. Eur J Prev Cardiol 2024 Nov 18;31(16):1904-1911. [doi: 10.1093/eurjpc/zwae259] [Medline: 39106529]

Abbreviations

CEG: caption equity gap
ENC: École Numérique de Cardiologie
HIC: high-income country
LIC: low-income country
mHealth: mobile health
MIC: middle-income country
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by A Stone, L Buis; submitted 26.Dec.2025; peer-reviewed by S Phuyal, TAK Manne, Y Hu; revised version received 27.Jul.2026; accepted 28.Jul.2026; published 08.Sep.2026.

Please cite as:
Roussel T, Benhaim E, Perrard L, Merat B, Vally S, Girerd R, Deharo P, Desroche LM, École Numérique de Cardiologie
Mobile-First Web Access and Captioned Video in Francophone Cardiology Education: Multicountry Ecological Learning Analytics Study
JMIR Mhealth Uhealth 2026;14:e90345
URL: https://mhealth.jmir.org/2026/1/e90345
doi:10.2196/90345

© Thomas Roussel, Elodie Benhaim, Louis Perrard, Benoit Merat, Shamir Vally, Remi Girerd, Pierre Deharo, Louis-Marie Desroche, École Numérique de Cardiologie. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 8.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

App-Based Training Module on Guiding Physicians' Prescription for Antibiotic Treatment of Gonorrhea: Cluster Randomized Controlled Trial

Ting-Ting Jiang[1], PhD; Xiao-Bin Zhang[2], MD; Pei-Zhen Zhao[3], PhD; Li-Juan Fei[4], BSc; Hui-Min Chen[5], BSc; Yun-Qing Yang[6], MD; Man-Hong Jia[2], PhD; Cheng Wang[3], PhD; Yun-Liang Shen[4], MD; Yu-Jun Xu[5], MD; Yan Han[1], PhD; Yue-Ping Yin[1], PhD; Xiang-Sheng Chen[1], PhD

[1]Hospital for Skin Diseases, Institute of Dermatology, Chinese Academy of Medical Sciences & Peking Union Medical College, 12 Jiangwangmiao Street, Nanjing, China

[2]Yunnan Center for Disease Control and Prevention, Kunming, China

[3]Dermatology Hospital, Southern Medical University, Guangzhou, China

[4]Zhejiang Provincial Institute of Dermatology, Huzhou, China

[5]Hainan Fifth People's Hospital, Haikou, China

[6]Guangzhou Dermatology Hospital, Guangzhou, China

Corresponding Author:
Xiang-Sheng Chen, PhD
Hospital for Skin Diseases, Institute of Dermatology, Chinese Academy of Medical Sciences & Peking Union Medical College, 12 Jiangwangmiao Street, Nanjing, China

Abstract

Background: The Chinese National Guidelines on Diagnosis and Treatment of Gonorrhea (2020) recommend ceftriaxone 1 g intramuscularly as a single dose for the treatment of uncomplicated gonococcal infections. However, nonadherence to the guidelines remains common among physicians in China, partly due to their poor awareness of the recommendations.

Objective: This trial aimed to (1) assess the effectiveness of an app-based training tool in improving adherence to guideline-recommended treatment regimens and (2) identify reasons for nonadherence among physicians who used the training tool.

Methods: Using a cluster randomized controlled trial design, we randomly allocated 72 hospitals (clusters) from 4 provinces to the intervention (n=36, 50%) or control (n=36, 50%) arms in a 1:1 ratio. In the intervention arm, physicians received free app-based training via smartphones for 6 months, while the control arm continued routine training. The primary outcome was the proportion of prescriptions adherent to the regimens for treatment of uncomplicated gonorrhea recommended by the national guidelines. We analyzed cluster-level summary outcomes at endline using methods appropriate for stratified cluster randomized controlled trials with relatively few clusters per group, accounting for between-cluster variation. A postintervention survey was administered to all intervention-arm physicians to evaluate their experience with the app-based training module. Additionally, face-to-face interviews were conducted with 32 purposely selected nonadherent physicians from the intervention arm to explore reasons for nonadherence.

Results: A total of 715 physicians (n=343, 48% in the control arm and n=372, 52% in the intervention arm) participated in the study. Over the 6-month intervention period, the adherence to 1 g ceftriaxone for the treatment of uncomplicated gonorrhea increased from 53.6% to 54.8% in the intervention arm, whereas the rate decreased from 43.9% to 42.5% in the control arm. The mean difference (5.5%, 95% CI −7% to 18%; *P*=.37) in changes in the adherence rate between intervention and control arms was not statistically significant (risk ratio 1.12, 95% CI 0.93 - 1.35; *P*=.23). Among the 32 (19%) nonadherent physicians from the intervention arm, the primary reason for nonadherence was their concern regarding the sufficiency of the currently recommended dosage of generic ceftriaxone for treating gonorrhea.

Conclusions: In our study, the implementation of an app-based training tool did not improve adherence to treatment recommendations. Concern regarding the efficacy of generic ceftriaxone may be one of the intrinsic factors associated with nonadherence.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2000029591; https://www.chictr.org.cn/showprojEN.html?proj=48864

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2020-037549

(JMIR Mhealth Uhealth 2026;14:e63736) doi:10.2196/63736

XSL•FO
RenderX

KEYWORDS

gonorrhea; prescribing; treatment guideline; nonadherence; randomized controlled trial; survey; online training; mobile app; interview

Introduction

Gonorrhea, caused by *Neisseria gonorrhoeae*, is a significant global public health issue, with an estimated 82.4 million cases among adults aged 15 to 49 years in 2020 [1]. Gonococcal infection predominantly involves the epithelium of the urethra, endocervix, rectum, oropharynx, and conjunctivae and can also ascend to the upper genital tract to cause pelvic inflammatory disease in women and epididymo-orchitis in men [2]. Effective control of gonorrhea relies on prevention, accurate diagnostics, and appropriate antimicrobial treatment. However, the treatment of gonorrhea is complicated by the ability of *N gonorrhoeae* to develop resistance to antimicrobials [3,4].

Due to the widespread emergence of resistant strains, only ceftriaxone is now recommended as the first-line regimen for treatment of gonorrhea in most countries [5,6]. However, isolates with decreased susceptibility to ceftriaxone, the last remaining option for first-line empiric treatment of gonorrhea, have been identified in Japan [7], the United States [8], France [9], Spain [10], the United Kingdom [11], and Australia [12]. In China, the percentage of *N gonorrhoeae* isolates with decreased susceptibility to ceftriaxone ranged between 9.7% and 12.2% from 2013 to 2016 [13].

In response to the growing threat of antimicrobial resistance, the 2020 European gonorrhea guidelines recommended high-dose ceftriaxone (1 g) for uncomplicated gonorrhea when antimicrobial susceptibility is unknown [5]. Similarly, the 2021 US Centers for Disease Control and Prevention sexually transmitted infections treatment guidelines recommended ceftriaxone 500 mg for individuals weighing <150 kg and 1 g for those weighing ≥150 kg [6]. In 2020, Chinese experts also reached a consensus to increase the ceftriaxone dose for uncomplicated gonorrhea to 1 g [14].

A major driver of antimicrobial resistance is the inappropriate use of antibiotics, particularly overprescribing, which can lead to the selection of resistant organisms. A nationwide study in China revealed a high proportion of prescription behaviors that were nonadherent to the national guidelines on diagnosis and treatment of gonorrhea in China (hereinafter referred to as the National Guidelines) [15], with excessive ceftriaxone dosing being common [16]. Lack of awareness of the National Guidelines was identified as a key factor contributing to nonadherence [15].

To address this issue, we developed a smartphone app–based training tool aimed at improving guideline adherence among physicians in China. We conducted a cluster randomized controlled trial to assess its effectiveness on improving adherence to the guidelines (ie, prescribing ceftriaxone 1 g intramuscular as a single dose for treatment of uncomplicated gonococcal infections). The secondary aim was to explore the reasons for nonadherence among physicians who used the training tool.

Methods

Ethical Considerations

This study was registered at the Chinese Clinical Trial Registry [17] (ChiCTR2000029591, registered on February 5, 2020) and approved by the Medical Ethics Committee of the Chinese Academy of Medical Sciences Institute of Dermatology (2020-LS-004). This study adhered to the principles of the Declaration of Helsinki. Participation in the study was voluntary and anonymous, and informed consent was obtained from all participants. No compensation was provided to participants.

Study Design and Setting

We conducted a parallel-group, cluster randomized controlled trial in 72 hospitals in 4 provinces (Guangdong, Hainan, Yunnan, and Zhejiang) in China to compare the adherence rate for treating uncomplicated gonorrhea with 1 g ceftriaxone between the intervention (app-based training module) and control groups.

Provinces were selected based on the reported incidence of gonorrhea in 2018, with high-incidence areas prioritized. Within each province, city-level hospitals with high reported gonorrhea cases were invited. After obtaining written consent from hospital directors, baseline data (eg, hospital category, gonorrhea case volume, physician qualifications, and drug availability) were collected. Hospitals were stratified by province and randomly assigned to intervention or control arms using a computer-generated sequence in R software (version 3.5.3; R Foundation for Statistical Computing), with allocation concealed until assignment. An internal pilot was conducted in 18 hospitals (n=9, 50% intervention and n=9, 50% control) in Guangdong province to assess feasibility. After confirming acceptability (166/241, 68.9% of physicians trained), the trial expanded to all 72 hospitals (n=36, 50% per arm), with no protocol changes between phases.

Participant and Intervention

The study was conducted from March 2022 to February 2023. In Guangdong, the 6-month intervention ran from March to August 2022, followed by a questionnaire survey in intervention hospitals in September. For Hainan and Yunnan, the intervention was implemented from April to September 2022, with the survey conducted in October. In Zhejiang, the intervention period was August 2022 to January 2023, followed by the survey in February 2023.

In China, medical graduates must complete a mandatory 3-year standardized residency training program following their undergraduate medical education (typically a 5-year program). This nationally regulated system, implemented through accredited teaching hospitals, ensures uniform clinical competency across all medical practitioners. For positions in major metropolitan hospitals (particularly tier-1 cities), competitive candidates generally require additional postgraduate qualifications beyond the mandatory training. All physicians

XSL•FO
RenderX

working in dermatology, urology, andrology, gynecology, and venereology departments at participating hospitals were eligible for study inclusion, irrespective of their educational background. After providing informed consent, physicians were guided by the local study coordinator to download the Xieshou app [18], an official platform developed by the National Center for STD Control of the Chinese Center for Disease Control and Prevention. The Xieshou platform is an electronic system routinely used for sexually transmitted disease (STD) health education and partner management in hospitals and is managed by STD control institutions at various levels. Health facilities are encouraged to integrate this platform into routine STD services. Through the platform, physicians can direct patients to scan a QR code to automatically receive an electronic intervention package containing health education materials, condom promotion information, mobilization for testing, and treatment instructions. Additionally, physicians can use the platform for academic exchanges, experience sharing, and participation in training sessions and conferences (including training videos, case discussions, and conference notices).

Technicians of the Xieshou platform distinguished between intervention and control group physicians based on hospital names provided during registration. In intervention hospitals, physicians were prompted to complete an online questionnaire collecting background information (sex, age, education level, years of practice, working department, professional title, previous training, and number of gonorrhea cases treated). After submission, they received unlimited access to a training video on the National Guidelines for the treatment of gonorrhea. Over the following 6 months, study coordinators sent monthly WeChat (Tencent Holdings Limited) reminders to physicians to watch the video, which was automatically saved in their personal learning records. At the end of the intervention, physicians in these hospitals were invited to complete another online questionnaire (Multimedia Appendix 1) via the Xieshou app to provide feedback on current treatment recommendations and their experience with the intervention. In control group hospitals, physicians were only prompted to complete the baseline questionnaire and did not have access to the training video after submission.

Outcomes

The primary outcome was the proportion of prescriptions adherent to the regimens for treatment of uncomplicated gonorrhea recommended by the National Guidelines at the cluster level. The secondary outcome was the reasons for nonadherence among physicians in the intervention arm.

Data Collection and Analyses

In each hospital, physicians' prescription records were extracted from the Hospital Information System using keyword matching (ie, gonorrhea). Prescriptions for uncomplicated gonorrhea from 6 months before (baseline) and 6 months after (endline) trial initiation were assessed for adherence to treatment guidelines. Prescriptions were excluded from data analysis if they were prescribed to patients who met any of the following criteria: (1) younger than 18 years; (2) were pregnant or lactating women; (3) received no ceftriaxone; (4) were treated with antibiotics for other infections; or (5) were diagnosed as complicated gonorrhea, such as disseminated gonococcal infection and pelvic inflammatory disease. Patient demographic data, gonorrhea diagnosis, and treatment regimens (antibiotic selection and dosage) were collected. To measure the adherence rate to the National Guidelines, a single 1 g dose of ceftriaxone was considered in line with the guidelines; deviations were classified as nonadherence. Two researchers independently reviewed the prescriptions and made the adjudication, and any discrepancies were discussed until agreement was reached.

For the secondary outcome, we purposively sampled 1 physician per intervention hospital (with ≥25% nonadherent prescriptions at endline). A total of 32 in-depth, face-to-face interviews were conducted in Chinese by local coordinators using a semistructured interview outline (Multimedia Appendix 2). Interview topics focused on intervention experiences, perspectives on antibiotic prescribing for gonorrhea treatment, and clinical practices. Interviews were audio-recorded upon written consent, and notes were taken. Participants were given the option to withdraw from the study at any point without any consequences. All interviews occurred in quiet and private locations in local hospitals.

Statistical Analysis

On the basis of a nationwide survey in China [15], we estimated that the adherence rate to the National Guidelines for the treatment of uncomplicated gonorrhea would be 37.8%, and that we could collect and process an average of 100 prescriptions per hospital. Clinical experts deemed a 10% increase in adherence clinically meaningful for the intervention arm. Under the assumption of a coefficient of variation of 0.1, 10% data loss due to illegible prescriptions, and accounting for stratification, the study required 144 clusters to detect this increase with 90% power at a 5% significance level (2-sided test). Eight provinces (Jiangsu, Shanghai, Zhejiang, Fujian, Guangdong, Guangxi, Hainan, and Yunnan) with the highest incidence of reported gonorrhea cases in China were selected in our study protocol [19]. Within each province, 9 hospitals were planned for the control group and 9 for the intervention group. However, the target number of hospitals (n=144) was not met due to COVID-19 pandemic–related restrictions in Jiangsu, Shanghai, Fujian, and Guangxi.

Participating hospital and physician characteristics were summarized using frequencies (and sample sizes) and means (SDs) as appropriate for each group. For the primary outcome, we used analysis methods appropriate for stratified, cluster-randomized trials with relatively few clusters per arm to account for between-cluster variation. We pooled both pilot and main trial data for our analysis of prescriptions, which was based on the original allocation of hospitals (ie, on an intention-to-treat basis). To deal with skewness in the distribution of cluster outcomes, we first log-transformed the proportions. We then estimated the overall risk ratio based on a weighted average of stratum-specific risk ratios, which were calculated as the difference between intervention and control arm log risks (equivalent to ratios of geometric mean risks between treatment arms), with the weights inversely proportional to stratum-specific variances. We then estimated an approximate SE for the log risk ratio and used this to calculate 95% CIs for

XSL•FO
RenderX

the risk ratio and carry out formal hypothesis testing (2 sided and at the 5% level) via a stratified 2-tailed *t* test. To adjust for covariates, we fitted a logistic regression model to the individual-level data with terms for the covariates of interest, excluding the treatment arm, and then calculated cluster-specific ratio residuals based on the ratio of observed to model-predicted values. Results were adjusted for stratum (province), cluster-level outcome at baseline, patient's sex and age, and physician's sex and department. We also calculated risk difference results using similar methods to those used to calculate the risk ratio results. In addition, we explored the associations between nonadherence and physician characteristics (age, sex, education level, years of practice, working department, professional title, number of treated gonorrhea cases, and previous training) without the intervention and added another exploratory outcome after protocol development: a physician-level indicator of nonadherence in the nonintervention group. Variables that showed significant association with the nonadherence ($P<.05$) in univariate analyses were included in the multivariate regression model to explore the association of variables with the outcome. All statistical analyses were conducted using SPSS Statistics for Windows (version 22.0; IBM Corp).

Qualitative Data Analysis

The qualitative interviews were recorded, deidentified, transcribed in Chinese, and then translated into English for analysis. The transcribed data were thematically analyzed in NVivo 11 (Lumivero). Two researchers trained in qualitative analysis methods, TTJ and YQY, analyzed several transcripts to identify themes and create codes and subcodes, which were organized into categories. This coding framework was applied to all transcripts by both researchers and modified through an iterative process.

Results

Baseline Characteristics

A total of 68 hospitals (n=33, 48.5% in the control group and n=35, 51.5% in the intervention group) provided background information. Overall, baseline hospital characteristics were well balanced between the 2 groups (Table 1). A total of 715 physicians (n=343, 48% in the control group and n=372, 52% in the intervention group) completed the baseline questionnaire survey. The intervention group had a higher proportion (n=233, 62.6%) of female physicians and included more physicians (n=213, 57.2%) from the STD department and the department of obstetrics and gynecology (Table 2).

Table . Baseline characteristics of participating hospitals.

Characteristics	Control (n=33)	Intervention (n=35)
Hospital nature, n (%)		
General hospital	29 (87.9)	28 (80.0)
Specialized hospital	4 (12.1)	7 (20.0)
Hospital township, n (%)		
Public teaching hospital	27 (81.8)	27 (77.1)
Public nonteaching hospital	6 (18.2)	7 (20.0)
Hospital category, n (%)		
Tertiary hospital	24 (72.7)	21 (60.0)
Secondary hospital	8 (24.2)	10 (28.6)
Number of gonorrhea cases reported in 2019, mean (SD)	124 (21)	146 (25)
Number of physicians qualified for managing patients with sexually transmitted disease, mean (SD)	33 (4)	25 (3)
Hospital had been stocked with ceftriaxone, n (%)		
Yes	32 (97.0)	33 (94.3)
No	1 (3.0)	2 (5.7)
Hospital had been stocked with spectinomycin, n (%)		
Yes	14 (42.4)	16 (45.7)
No	19 (57.6)	19 (54.3)

XSL•FO
RenderX

Table . Baseline characteristics of physicians.

Characteristics	Control (n=343), n (%)	Intervention (n=372), n (%)
Sex		
Male	157 (45.8)	139 (37.4)
Female	186 (54.2)	233 (62.6)
Age group (years)		
<35	128 (37.3)	131 (35.2)
35-44	122 (35.6)	123 (33.1)
45-55	70 (20.4)	87 (23.4)
>55	23 (6.7)	31 (8.3)
Highest degree		
Postgraduate and above	128 (37.3)	130 (34.9)
Undergraduate	204 (59.5)	224 (60.2)
Junior college and below	11 (3.2)	18 (4.8)
Working department		
Dermatology	62 (18.1)	68 (18.3)
Obstetrics and gynecology	87 (25.4)	134 (36.0)
Urology	74 (21.6)	60 (16.1)
Andrology	12 (3.5)	6 (1.6)
Sexually transmitted disease	35 (10.2)	79 (21.2)
Service (years)		
≤5	88 (25.7)	88 (23.7)
6-10	73 (21.3)	72 (19.4)
≥11	182 (53.1)	212 (57.0)
Professional title		
Chief physician	43 (12.5)	43 (11.6)
Associate chief physician	94 (27.4)	117 (31.5)
Physician-in-charge	132 (38.5)	122 (32.8)
Resident physician	69 (20.1)	69 (18.5)
Number of cases of uncomplicated gonorrhea treated in the last 3 months		
≤1	293 (85.4)	296 (79.6)
2-4	44 (12.8)	69 (18.5)
≥5	6 (1.7)	7 (1.9)
Having received relevant training on treatment of gonorrhea in the last 3 months		
Yes	217 (63.3)	214 (57.5)
No	126 (36.7)	158 (42.5)

Nonadherence to the National Guidelines

Among the 1524 prescriptions (n=747, 49% prescriptions in the control group and n=777, 51% baseline prescriptions in the intervention group) with physician demographic data, 734 (48.2%) were nonadherent to the National Guidelines for ceftriaxone treatment of uncomplicated gonorrhea.

In the multivariate analyses, the following factors were found to be significantly associated with nonadherence after adjusting for potential confounding factors: age <35 years (adjusted odds ratio [AOR] 2.61, 95% CI 1.74 - 3.92, compared with ≥35 years; *P*<.001); undergraduate degree and below (AOR 2.63, 95% CI 1.99 - 3.47, compared with postgraduate degree and above; *P*<.001), working in a department of dermatology (AOR 4.05, 95% CI 2.93 - 5.58, compared with the department of STD;

XSL•FO
RenderX

P<.001); and treating fewer gonorrhea cases (<5 cases) in the past 3 months (AOR 2.66, 95% CI 1.67 - 4.23, compared with treating ≥5 cases; *P*<.001; Table 3).

Table . Factors associated with nonadherence to the treatment guidelines in the control group[a].

Characteristics	Nonadherence (n=734), n (%)	Adherence (n=790), n (%)	AOR[b] (95% CI)
Sex			
Male	475 (64.7)	606 (76.7)	Reference
Female	259 (35.3)	184 (23.3)	1.21 (0.87-1.67)
Age group (years)			
<35	194 (26.4)	122 (15.4)	2.61 (1.74-3.92)[c]
≥35	540 (73.6)	668 (84.6)	Reference
Highest degree			
Undergraduate and below	572 (77.9)	512 (64.8)	2.63 (1.99-3.47)[c]
Postgraduate and above	162 (22.1)	278 (35.2)	Reference
Working department			
Dermatology	320 (43.6)	132 (16.7)	4.05 (2.93-5.58)[c]
Obstetrics and gynecology	35 (4.8)	50 (6.3)	1.14 (0.66-1.98)
Urology	226 (30.8)	385 (48.7)	1.25 (0.90-1.74)
Andrology	6 (0.8)	39 (4.9)	0.31 (0.12-0.79)
Sexually transmitted disease	120 (16.3)	184 (23.3)	Reference
Professional title			
Resident physician	96 (13.1)	65 (8.2)	1.30 (0.72-2.36)
Physician-in-charge	307 (41.8)	269 (34.1)	1.24 (0.85-1.83)
Associate chief physician	249 (33.9)	254 (32.2)	1.54 (1.07-2.21)
Chief physician	82 (11.2)	202 (25.6)	Reference
Number of treated cases of gonorrhea in the last 3 months			
<5	701 (95.5)	673 (85.2)	2.66 (1.67-4.23)[c]
≥5	33 (4.5)	117 (14.8)	Reference

[a]Variables including age, sex, education level, working department, professional title, and number of treated cases of gonorrhea in the last 3 months were included in the multivariate analysis.

[b]AOR: adjusted odds ratio.

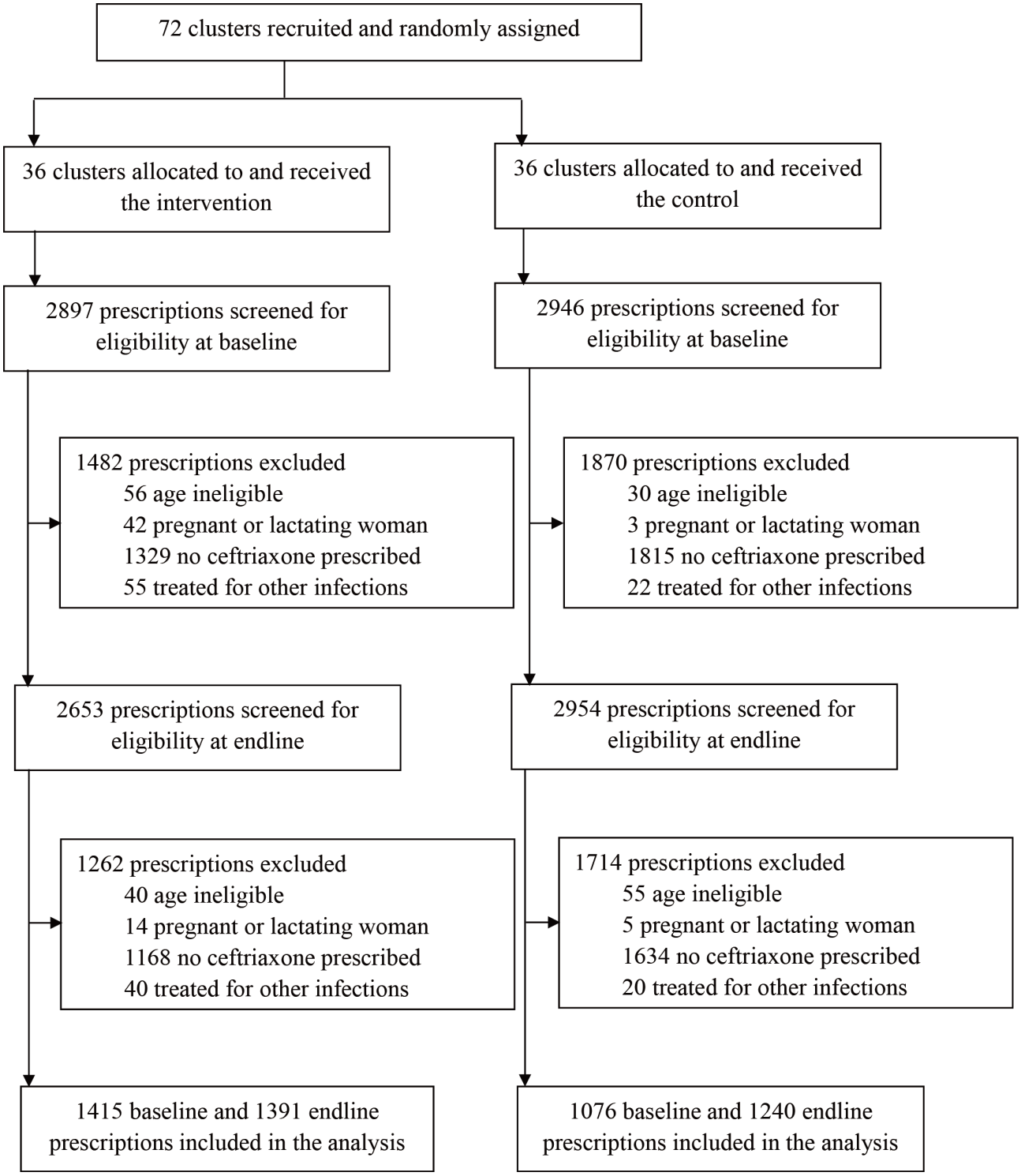
[c]*P*<.001.

Impact on Improvement of Adherence

Of the 3254 patients diagnosed with gonorrhea at baseline, 2171 (66.7%) were treated with ceftriaxone, whereas among the 3191 patients diagnosed with gonorrhea at endline, 2137 (67%) underwent treatment with ceftriaxone. We screened 2897 baseline and 2653 endline prescriptions from the intervention group, of which 1415 (48.8%) baseline prescriptions and 1391 (52.4%) endline prescriptions were eligible, respectively, for inclusion in the analysis (Figure 1). In the control group, 36.5% (1076/2946) of baseline and 42% (1240/2954) of endline prescriptions were eligible (Figure 1). The adherence rate increased from 53.6% to 54.8% in the intervention group, while the rate decreased from 43.9% to 42.5% in the control group. We also noted the variation of the adherence rate in the intervention clusters between baseline and endline was from −67% to 71% (mean difference 1.3%, 95% CI −8.5% to 11.0%; *P*=.79, paired *t* test), whereas in the control group it ranged from −50% to 100% (mean difference −1.4%, 95% CI −11.5% to 8.7%; *P*=.78, paired *t* test). After controlling for potential confounders, the online training intervention was not significantly associated with changes in the adherence rate (mean difference 5.5%, 95% CI −7% to 18%; *P*=.37; risk ratio 1.12, 95% CI 0.93 to 1.35; *P*=.23). However, when further comparing the compliance trend between the control and intervention groups, differences in trends of mean compliance of prescriptions were observed (*P*<.001; Multimedia Appendix 3). In addition, when the comparison was done by province, an obvious increase in mean compliance of prescriptions could be observed in Yunnan, where baseline compliance was substantially low (Multimedia Appendix 3).

XSL•FO
RenderX

Figure 1. Trial profile. Clusters refer to hospitals. Baseline prescriptions were issued during the 6 months before the implementation of the intervention, whereas endline prescriptions were issued during the 6-month intervention period.



Reasons for Nonadherence to the National Guidelines

A total of 64 physicians in the intervention group completed the questionnaire assessing their experience with the app-based training. Most respondents (n=54, 84.4%) reported having completed the monthly training, and 9 (14.1%) reported being unaware of the latest National Guidelines before the intervention.

In the intervention group, 168 physicians did not adhere to the National Guidelines. To investigate reasons for nonadherence, 32 (19%) physicians were interviewed. Concerns about the efficacy of generic ceftriaxone emerged as the predominant theme across all interviews:

> *We are all using generic ceftriaxone. I typically opt for a 2 g dose due to concerns regarding the efficacy of generic ceftriaxone.* [A dermatologist]

In China, the national essential medicines scheme and zero-markup policy have been instrumental in severing the connection between prescribing practices and hospital revenues.

XSL•FO
RenderX

However, these policies also restrict the availability of certain antibiotics:

> *We perceive that the efficacy of generic drugs is inferior to that of branded drugs. There appears to be a reduced selection of drugs available compared to before, with many effective branded drugs no longer accessible. The remaining generic drugs seem to be of lower quality.* [A gynecologist]

Instances have been reported in which generic antibiotics failed to effectively treat gonorrhea, further eroding physicians' confidence in generics:

> *I believe that the dosage and duration of generic ceftriaxone 1 g single dose for treating gonorrhea are inadequate. Patients treated with this regimen still experienced purulent discharge from the urethral orifice on the second and third days post-treatment, leading to a higher recurrence rate.* [A urologist]

Discussion

Principal Findings

In this study, we used an electronic platform to deliver training on gonorrhea treatment guidelines to physicians in hospital settings. The results demonstrated that the online training module did not significantly improve physician adherence to the National Guidelines for treating uncomplicated gonorrhea with ceftriaxone (P=.23). However, the intervention showed a more pronounced effect in Yunnan province, where adherence in the intervention group increased by 15%. This notable improvement may be attributed to the region's exceptionally low baseline adherence rate (4%) compared with the pooled baseline adherence rate of other provinces (65%). In contrast, areas with moderate to high baseline adherence rates (Zhejiang: 49%; Guangdong: 72%; and Hainan: 77%) showed no significant behavioral changes in the postintervention period, suggesting that nonadherence in these regions was unlikely due to a lack of guideline awareness. This hypothesis is further supported by the finding that adherence rates did not differ significantly between physicians who were aware of the updated National Guidelines (26/55, 47.3%) and those who were not (5/9, 55.6%) in the intervention group. Postintervention interviews among physicians who were not adherent to the guidelines revealed that concerns about the efficacy of generic ceftriaxone were the primary reason for deviating from the guidelines.

Comparison With Prior Work

Ceftriaxone is a broad-spectrum bactericidal agent that belongs to the third-generation cephalosporins [20]. It was patented, manufactured, and marketed in 1982 as Rocephin (Roche Pharmaceuticals). Following the expiration of the Rocephin patent, generic products became available in many countries. Compared with the brand-name (innovator) product, generics significantly reduce health care costs and enhance supply security due to their diverse sources [21,22]. Consequently, Asian markets are dominated by low-cost, locally or regionally manufactured generic medicines [23], leading to a wide availability of generic ceftriaxone for gonorrhea treatment. The efficacy and safety of these generics are theoretically assured by bioequivalence studies, making them interchangeable with innovator products [24]. However, numerous studies have found that generic ceftriaxone products often fail to meet the pharmaceutical quality standards of the branded original [25-27]. A survey of the quality of generic ceftriaxone products from Eastern Asia markets revealed high levels of impurities and microbiologic contamination, raising significant efficacy and safety concerns [28]. Physicians in this study reported treatment failures with generics in clinical practice, which may have led them to prescribe excessive doses of ceftriaxone. Overuse of antibiotics, prescription of branded antibiotics, and prescription of antibiotics not listed in the National List of Essential Medicines are prevalent in other limited-income countries [29]. Physician perceptions that branded drugs are superior to generics may contribute to both the overuse of generics and the preferential prescription of branded alternatives [29,30].

In China, ceftriaxone susceptibility testing is conducted using the innovator product rather than generics. It remains unclear whether a 1 g single dose of generic ceftriaxone is adequate for uncomplicated gonorrhea due to the lack of postmarketing surveillance and follow-up studies in China. Notably, 15.6% (10/64) of physicians in this study still considered the 1 g dose insufficient. Thus, further research is needed to generate evidence on the efficacy of currently available generic antibiotics for gonorrhea and to strengthen surveillance of resistance to these generics.

Although evidence-based guidelines have been developed, translating knowledge into practice is challenging. One potential reason for the intervention's limited success is the 6-month duration, which may have been insufficient to alter prescribing behavior. Time-trend analyses in this study demonstrated that extending the intervention period led to upward trends in appropriate antibiotic use for gonorrhea. Other studies have similarly found that discontinuing effective behavioral interventions resulted in a resurgence of inappropriate prescribing [31]. Additionally, online training alone may have a limited impact on reducing antibiotic misuse. A multifaceted approach is crucial for success. Interventions such as physician education, computerized clinical decision support, audit and feedback, punishment, and financial incentives, when implemented in combination, have resulted in lower rates of antibiotic misuse in other settings [32,33].

Strengths and Limitations

To our knowledge, this study represents the first cluster randomized controlled trial conducted in China to evaluate the impact of an innovative training program on reducing physicians' nonadherence to gonorrhea treatment guidelines.

However, several limitations should be considered when interpreting the findings. First, the COVID-19 pandemic prevented us from achieving the planned number of clusters and prescriptions per cluster, resulting in reduced statistical power. Notably, the study was conducted in only 4 of mainland China's 31 provinces, and the results should not be overextrapolated to other regions due to significant variations in guideline awareness and compliance. Future research should include physicians from more provinces. Second, the study

XSL•FO
RenderX

focused solely on adherence to ceftriaxone-based guidelines. Given the emergence of new antigonococcal antimicrobials, future training programs should incorporate guidelines for alternative antibiotics. Third, while cluster randomization ensured good balance at the hospital and physician levels, residual confounding or unmeasured contextual factors may have influenced the results. Additionally, the inability to mask physicians to their intervention group introduced Hawthorne effect bias. To mitigate this, we avoided direct observation during prescribing and collected outcome data in either group after the 6-month intervention. Finally, the qualitative component had a small sample size, and the participant sampling process and demographic characteristics were not systematically documented. This methodological gap may limit the generalizability of the findings. Future studies should prioritize multicenter, large-scale designs integrating quantitative and qualitative methods to enhance validity.

Future Directions

More robust evidence is needed on the efficacy of currently available generic antibiotics for gonorrhea treatment and their resistance surveillance. This evidence should inform the development of local evidence-based guidelines and be incorporated into training programs. With such data supporting the use of generics, physicians can be more confident in prescribing generic ceftriaxone at recommended dosages. Fortunately, postmarket evaluation of generic medicines has been included in the ROADMAP research plan to address gonococcal resistance in China [34]. In 2016, the National Medical Products Administration (NMPA), formally known as the China Food and Drug Administration, mandated that generic drugs approved before 2007 must demonstrate bioequivalence to brand-name counterparts by the end of 2021 [35]. To correct misconceptions about generic drugs, the NMPA should also ensure that generic drugs meet quality standards by using good manufacturing practices and that the generic drug approval process is rigorous and transparent to the public.

Once evidence-based guidelines are established, efforts to reduce inappropriate antibiotic prescriptions can be enhanced by training physicians via electronic platforms. Future research should also explore whether applying training long-term or adding other interventions beyond online training can further optimize antibiotic use for gonorrhea.

Conclusions

In conclusion, our study demonstrates that a 6-month intervention involving an app-based training tool with monthly reminders to physicians did not significantly improve adherence to the National Guidelines for gonorrhea treatment with ceftriaxone. A key barrier to adherence was physicians' concerns about the perceived insufficient efficacy of currently available generic ceftriaxone. Addressing these concerns will require postmarket reevaluation of generic ceftriaxone products and targeted education for physicians on the scientific evidence supporting the efficacy of generics. These steps are essential to promote appropriate prescribing behavior.

Acknowledgments

The authors thank collaborators (coordinators and physicians) who made contributions to the implementation of the trial in the 72 participating hospitals.

Funding

The study was funded by the key research and development program of Yunnan Province Science and Technology Department (grant 202403AC100011) and the Chinese Academy of Medical Sciences Initiative for Innovative Medicine (grant 2016-I2M-3-021). The funder had no role in the design, collection, analysis, and interpretation of data; in writing of the manuscript, and in the decision to submit this manuscript for publication.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Follow-up questionnaire survey for the physicians in the intervention hospitals.
[PDF File, 124 KB - mhealth_v14i1e63736_app1.pdf]

Multimedia Appendix 2
Interview outline.
[PDF File, 62 KB - mhealth_v14i1e63736_app2.pdf]

Multimedia Appendix 3

XSL•FO
RenderX

Mean compliance rates of antibiotic prescribing for the treatment of gonorrhea per month by (A) different groups and (B) provinces. App-based training intervention introduced at month 0. Error bars indicate 95% CIs.
[PDF File, 196 KB - mhealth_v14i1e63736_app3.pdf]

Checklist 1
CONSORT-eHEALTH (V 1.6.1) checklist.
[PDF File, 503 KB - mhealth_v14i1e63736_app4.pdf]

References

1. Global progress report on HIV, viral hepatitis and sexually transmitted infections, 2021. World Health Organization. 2021. URL: https://www.who.int/publications/i/item/9789240027077/ [accessed 2024-03-18]
2. Unemo M, Seifert HS, Hook EW III, Hawkes S, Ndowa F, Dillon JAR. Gonorrhoea. Nat Rev Dis Primers 2019;5(1):79. [doi: 10.1038/s41572-019-0128-6]
3. Hook EW III, Kirkcaldy RD. A brief history of evolving diagnostics and therapy for gonorrhea: lessons learned. Clin Infect Dis 2018 Sep 28;67(8):1294-1299. [doi: 10.1093/cid/ciy271] [Medline: 29659749]
4. Sánchez-Busó L, Golparian D, Corander J, et al. The impact of antimicrobials on gonococcal evolution. Nat Microbiol 2019 Nov;4(11):1941-1950. [doi: 10.1038/s41564-019-0501-y] [Medline: 31358980]
5. Unemo M, Ross J, Serwin AB, Gomberg M, Cusini M, Jensen JS. 2020 European guideline for the diagnosis and treatment of gonorrhoea in adults. Int J STD AIDS 2020 Oct 29:956462420949126. [doi: 10.1177/0956462420949126] [Medline: 33121366]
6. Workowski KA, Bachmann LH, Chan PA, et al. Sexually transmitted infections treatment guidelines, 2021. MMWR Recomm Rep 2021 Jul 23;70(4):1-187. [doi: 10.15585/mmwr.rr7004a1] [Medline: 34292926]
7. Ohnishi M, Saika T, Hoshina S, et al. Ceftriaxone-resistant Neisseria gonorrhoeae, Japan. Emerg Infect Dis 2011 Jan;17(1):148-149. [doi: 10.3201/eid1701.100397] [Medline: 21192886]
8. Centers for Disease Control and Prevention (CDC). Cephalosporin susceptibility among Neisseria gonorrhoeae isolates--United States, 2000-2010. MMWR Morb Mortal Wkly Rep 2011 Jul 8;60(26):873-877 [FREE Full text] [Medline: 21734634]
9. Unemo M, Golparian D, Nicholas R, Ohnishi M, Gallay A, Sednaoui P. High-level cefixime- and ceftriaxone-resistant Neisseria gonorrhoeae in France: novel penA mosaic allele in a successful international clone causes treatment failure. Antimicrob Agents Chemother 2012 Mar;56(3):1273-1280. [doi: 10.1128/AAC.05760-11] [Medline: 22155830]
10. Carnicer-Pont D, Smithson A, Fina-Homar E, Bastida MT, Gonococcus Antimicrobial Resistance Surveillance Working Group. First cases of Neisseria gonorrhoeae resistant to ceftriaxone in Catalonia, Spain, May 2011. Enferm Infecc Microbiol Clin 2012 Apr;30(4):218-219. [doi: 10.1016/j.eimc.2011.11.010] [Medline: 22244992]
11. Eyre DW, Sanderson ND, Lord E, et al. Gonorrhoea treatment failure caused by a Neisseria gonorrhoeae strain with combined ceftriaxone and high-level azithromycin resistance, England, February 2018. Euro Surveill 2018 Jul;23(27):27. [doi: 10.2807/1560-7917.ES.2018.23.27.1800323] [Medline: 29991383]
12. Jennison AV, Whiley D, Lahra MM, et al. Genetic relatedness of ceftriaxone-resistant and high-level azithromycin resistant Neisseria gonorrhoeae cases, United Kingdom and Australia, February to April 2018. Euro Surveill 2019 Feb;24(8):1900118. [doi: 10.2807/1560-7917.ES.2019.24.8.1900118] [Medline: 30808445]
13. Yin YP, Han Y, Dai XQ, et al. Susceptibility of Neisseria gonorrhoeae to azithromycin and ceftriaxone in China: a retrospective study of national surveillance data from 2013 to 2016. PLoS Med 2018 Feb;15(2):e1002499. [doi: 10.1371/journal.pmed.1002499] [Medline: 29408881]
14. Wang QQ, Zhang RL, Liu QZ, et al. National guidelines on diagnosis and treatment of gonorrhea in China (2020). Int J Dermatol Venereol 2020;3(3):129-134. [doi: 10.1097/JD9.0000000000000072]
15. Han Y, Yin YP, Zhou Y, et al. Nonadherence to national guidelines for antibiotic treatment of uncomplicated gonorrhea in China: results from a nationwide survey. Sex Transm Dis 2018 Sep;45(9):600-606. [doi: 10.1097/OLQ.0000000000000819] [Medline: 30102261]
16. Han Y, Yin Y, Dai X, et al. Widespread use of high-dose ceftriaxone therapy for uncomplicated gonorrhea without reported ceftriaxone treatment failure: results from 5 years of multicenter surveillance data in China. Clin Infect Dis 2020 Jan 1;70(1):99-105. [doi: 10.1093/cid/ciz170] [Medline: 30838398]
17. Chinese Clinical Trial Registry [Website in Chinese]. URL: https://www.chictr.org.cn/ [accessed 2026-02-22]
18. Xieshou app [Website in Chinese]. URL: https://www.xieshoue.org [accessed 2026-02-23]
19. Jiang TT, Yang YQ, Cao NX, Yin YP, Chen XS. Novel education-based intervention to reduce inappropriate antibiotic prescribing for treatment of gonorrhoea in China: protocol for a cluster randomised controlled trial. BMJ Open 2020 Jul 12;10(7):e037549. [doi: 10.1136/bmjopen-2020-037549] [Medline: 32660953]
20. Duncan CJA, Barr DA, Seaton RA. Outpatient parenteral antimicrobial therapy with ceftriaxone, a review. Int J Clin Pharm 2012 Jun;34(3):410-417. [doi: 10.1007/s11096-012-9637-z] [Medline: 22527482]

XSL•FO
RenderX

21. Cameron A, Mantel-Teeuwisse AK, Leufkens HG, Laing RO. Switching from originator brand medicines to generic equivalents in selected developing countries: how much could be saved? Value Health 2012;15(5):664-673. [doi: 10.1016/j.jval.2012.04.004] [Medline: 22867775]
22. Sun J. International experiences of promoting generics use and its implications to China. J Evid Based Med 2013 May;6(2):74-80. [doi: 10.1111/jebm.12030] [Medline: 23829799]
23. Zuo J, Gao Y, Bou-Chacra N, Löbenberg R. Challenges and opportunities to use biowaivers to compare generics in China. AAPS PharmSciTech 2014 Oct;15(5):1070-1075. [doi: 10.1208/s12249-014-0133-8] [Medline: 24848759]
24. Chow SC, Song F, Chen M. Some thoughts on drug interchangeability. J Biopharm Stat 2016;26(1):178-186. [doi: 10.1080/10543406.2015.1092027] [Medline: 26366703]
25. Tange M, Yoshida M, Nakai Y, Uchida T. Comparison between original and generic versions of ceftriaxone sodium preparation for injection: compatibility with calcium-containing product. Chem Pharm Bull (Tokyo) 2012;60(4):429-434. [doi: 10.1248/cpb.60.429] [Medline: 22466725]
26. Yao Y, Zhou R, Wang Y. Fatal adverse effects of injected ceftriaxone sodium in China. Pharmacoepidemiol Drug Saf 2012 Nov;21(11):1197-1201. [doi: 10.1002/pds.3232] [Medline: 22761158]
27. Chong X, Wang L, Hu C, He L. Quality control and consistency evaluation of ceftriaxone sodium for injection. J Chemother 2023 Oct;35(6):477-490. [doi: 10.1080/1120009X.2022.2153314] [Medline: 36519963]
28. Arnet I, Altermatt M, Roggo Y, Schnetzler G. Pharmaceutical quality of eight generics of ceftriaxone preparation for injection in eastern Asia. J Chemother 2015;27(6):337-342. [doi: 10.1179/1973947814Y.0000000208] [Medline: 25144924]
29. Basnet S, Koju P, Silwal P, et al. Antibiotic prescription patterns in the emergency department of a tertiary healthcare center in Nepal: a descriptive cross-sectional study. J Int Med Res 2024 Sep;52(9):3000605241274513. [doi: 10.1177/03000605241274513] [Medline: 39238436]
30. Hadia RB, Joshi DB, Gohel KH, Khambhati N. Knowledge, attitude, and practice of generic medicines among physicians at multispecialty hospital: an observational study. Perspect Clin Res 2022;13(3):155-160. [doi: 10.4103/picr.PICR_281_20] [Medline: 35928641]
31. Linder JA, Meeker D, Fox CR, et al. Effects of behavioral interventions on inappropriate antibiotic prescribing in primary care 12 months after stopping interventions. JAMA 2017 Oct 10;318(14):1391-1392. [doi: 10.1001/jama.2017.11152] [Medline: 29049577]
32. Gerber JS, Prasad PA, Fiks AG, et al. Effect of an outpatient antimicrobial stewardship intervention on broad-spectrum antibiotic prescribing by primary care pediatricians: a randomized trial. JAMA 2013 Jun 12;309(22):2345-2352. [doi: 10.1001/jama.2013.6287] [Medline: 23757082]
33. Meeker D, Linder JA, Fox CR, et al. Effect of behavioral interventions on inappropriate antibiotic prescribing among primary care practices: a randomized clinical trial. JAMA 2016 Feb 9;315(6):562-570. [doi: 10.1001/jama.2016.0275] [Medline: 26864410]
34. Chen XS, Yin YP, Li XY. A ROADMAP plan to address research needs for gonococcal antimicrobial resistance in China. Clin Infect Dis 2019 Jan 18;68(3):505-510. [doi: 10.1093/cid/ciy566] [Medline: 29985996]
35. Qu J, Zuo W, Wang S, et al. Knowledge, perceptions and practices of pharmacists regarding generic substitution in China: a cross-sectional study. BMJ Open 2021 Oct 18;11(10):e051277. [doi: 10.1136/bmjopen-2021-051277] [Medline: 34663661]

Abbreviations

AOR: adjusted odds ratio
NMPA: National Medical Products Administration
STD: sexually transmitted disease

Edited by L Buis; submitted 28.Jun.2024; peer-reviewed by A Debbarma, MP Li; revised version received 18.Nov.2025; accepted 29.Jan.2026; published 04.Mar.2026.

Please cite as:
Jiang TT, Zhang XB, Zhao PZ, Fei LJ, Chen HM, Yang YQ, Jia MH, Wang C, Shen YL, Xu YJ, Han Y, Yin YP, Chen XS
App-Based Training Module on Guiding Physicians' Prescription for Antibiotic Treatment of Gonorrhea: Cluster Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e63736
URL: https://mhealth.jmir.org/2026/1/e63736
doi:10.2196/63736

© Ting-Ting Jiang, Xiao-Bin Zhang, Pei-Zhen Zhao, Li-Juan Fei, Hui-Min Chen, Yun-Qing Yang, Man-Hong Jia, Cheng Wang, Yun-Liang Shen, Yu-Jun Xu, Yan Han, Yue-Ping Yin, Xiang-Sheng Chen. Originally published in JMIR mHealth and uHealth

XSL•FO
RenderX

(https://mhealth.jmir.org), 4.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Real-World Functional Gait Recovery Trajectories After Total Hip and Total Knee Arthroplasty via Smartphone Sensors: Observational Longitudinal Study

Levi S Teitz[1], PhD; Einat Granot-Hershkovitz[1], PhD; Pei-An Lee[2,3,4], PhD; On-Yee Lo[2,3,4], PhD; Brad Manor[2,3,4], PhD; Yuval Y Naveh[1], MSc

[1]OneStep (Celloscope Ltd), Nachmani 14, Tel Aviv, Israel

[2]Marcus Institute for Aging Research, Hebrew SeniorLife, Boston, MA, United States

[3]Beth Israel Deaconess Medical Center, Boston, MA, United States

[4]Harvard Medical School, Harvard University, Boston, MA, United States

Corresponding Author:
Yuval Y Naveh, MSc
OneStep (Celloscope Ltd), Nachmani 14, Tel Aviv, Israel

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2026/1/e116089

Abstract

Background: Recovery after total hip arthroplasty (THA) and total knee arthroplasty (TKA) is typically assessed with infrequent clinic-based evaluations that may not reflect patients' everyday mobility. Smartphone inertial sensors enable scalable, repeated gait measurement in real-world settings, but the extent to which prompted walking tests reflect daily life walking during recovery is not well defined.

Objective: The objective of this study was to characterize week-by-week recovery trajectories after THA and TKA using high-frequency smartphone sensor data and quantify the difference between prompted in-app walking tests and passive daily life background walking.

Methods: We analyzed longitudinal smartphone-based gait measurements from 735 patients (THA, n=350; TKA, n=385) who underwent unilateral surgery between July 2022 and October 2025. Nine spatiotemporal gait parameters were monitored over the first 15 postoperative weeks across 2 conditions: prompted "in-app" walks and passively captured "background" walks. Recovery trajectories were evaluated using Kaplan-Meier time-to-baseline analyses, linear mixed-effects models, and comparison of late-stage plateau levels (weeks 11 - 15) using estimated marginal means of the models.

Results: Across all parameters, THA patients returned to preoperative baseline levels faster than TKA patients, with mean trajectories generally approaching baseline by week 4 for THA versus weeks 6 to 8 for TKA. Linear mixed-effects models showed that surgical procedure type significantly influenced recovery trajectory for all parameters except single support of the unaffected leg. All gait parameters except stance asymmetry and single support of the unaffected leg exhibited persistent differences through the late-recovery plateau (weeks 11 - 15). We also compared prompted "in-app" walks with passively captured background walking. Prompted and passive walking showed systematic divergence in both THA and TKA; prompted assessments generally indicated faster or more complete return to baseline than passive daily life walking. Finally, preoperative gait level was strongly associated with postoperative trajectories, with baseline-stratified analyses showing distinct patterns of recovery between groups.

Conclusions: High-frequency smartphone gait assessment quantifies procedure-specific recovery dynamics and demonstrates that prompted tests and passive daily life measurements capture distinct, noninterchangeable functional signals. Parameters from prompted walks plateaued while those from passively captured walks were still changing; prompted testing alone therefore captures only one component of returning function, highlighting the additional information that passive, real-world gait monitoring can provide.

(JMIR Mhealth Uhealth 2026;14:e97657) doi:10.2196/97657

KEYWORDS

total hip arthroplasty; total knee arthroplasty; gait analysis; smartphone sensors; remote patient monitoring; functional recovery; digital health; observational study; gait parameters

XSL•FO
RenderX

Introduction

Total hip arthroplasty (THA) and total knee arthroplasty (TKA) are highly successful orthopedic surgeries that restore mobility for millions of people with end-stage arthritis [1,2]. Their volume is substantial and growing; over 3.7 million Americans underwent THA or TKA from 2021 to 2024 [3], and projections indicate that the annual volume for THA and TKA could increase by 176% and 139%, respectively, by 2040, compared to 2019 volumes [4]. This escalating demand establishes total joint arthroplasty (TJA) as a major component of the health care system, making the optimization of postoperative recovery a critical public health priority.

Conventional methods for monitoring patient recovery after TJA have significant limitations and provide an incomplete picture of a patient's functional status [5]. The standard of care relies on infrequent, in-person clinical visits featuring practitioner-dependent assessments such as the Timed Up and Go test [6]. These appointments offer only brief snapshots of function in a controlled setting and correlate poorly with a patient's actual daily activities. This episodic model limits a clinician's insight into the patient's day-to-day progress and challenges, hindering proactive intervention. Furthermore, the common practice of limiting complication surveillance to the acute recovery phase has been shown to miss a substantial number of adverse events, revealing a systemic gap in long-term patient monitoring [7].

Another key challenge in postarthroplasty care is the discordance between subjective patient-reported outcome measures (PROMs) and objective, performance-based metrics [8]. While PROMs are vital for capturing the patient's perspective on function, pain, and quality of life, studies show that early improvements in these scores often do not align with objectively measured functional gains [9,10]. Patients frequently report feeling significantly better long before objective measures, such as walk speed or mobility, return to normal levels. This may partly reflect that patients often associate recovery with meaningful pain relief, even when objective mobility remains impaired. However, pain relief is unlikely to be the only explanation, and the broader point remains that perceived recovery and measured physical performance can differ substantially. This issue is compounded by low patient completion rates for long-term PROMs, which can introduce bias into the data [11]. This disconnect highlights the need for granular, individualized, and objective monitoring tools that can characterize function beyond self-assessment.

Previous studies have attempted to establish such recovery benchmarks and expected trajectories [12-14]. However, these efforts are often limited in scope, focusing on a single metric or joint, or on limited follow-up periods with infrequent data collection time points, which fail to capture the continuous, dynamic nature of the recovery process.

The widespread availability of smartphones with built-in inertial measurement units (IMUs) offers a validated and scalable solution to the limitations of traditional monitoring. A growing body of research confirms the reliability of smartphone-based systems for measuring key gait parameters like walk speed, cadence, and asymmetry in real-world environments [15-18]. These systems show strong correlation with gold-standard laboratory equipment. This technology shifts measurement from the clinic to the patient's daily life, enabling the continuous, high-frequency collection of objective functional data that is more representative of true mobility [19]. Digital health platforms like the OneStep system leverage these capabilities to create objective recovery benchmarks and track individual progress with unprecedented detail.

While remote monitoring provides a wealth of objective data, a critical nuance essential for understanding true recovery is often overlooked: the distinction between a patient's functional capacity—what they are capable of doing under observation or during a prompted task—and their habitual daily performance—what they actually do in an unobserved, unguarded state [20-23]. This distinction is vital, as the goal of rehabilitation is not merely to enable a patient to perform a task correctly, but to restore normal, noncompensatory movement as an automatic pattern in daily life.

This study presents week-by-week gait trajectories after THA and TKA, measured using a smartphone-based system, to guide more precise postoperative monitoring and rehabilitation. Specifically, we compare the 15-week recovery trajectories of THA and TKA patients across 9 gait parameters and quantify differences between prompted in-app walks, treated as a proxy for functional capacity, and passively captured background walking, treated as a proxy for habitual performance.

Methods

Cohort

This study analyzed deidentified data collected from commercial users (N=735) of the OneStep gait monitoring system during routine postoperative care and recovery monitoring between July 2022 and October 2025. Data were not collected specifically for this research but were generated as part of standard app use. Starting from all patients with a recorded THA or TKA, we sequentially applied eligibility criteria based on surgery type, data availability, and preoperative recording requirements.

Ethical Considerations

This study used secondary, deidentified data collected during routine use of a commercial gait monitoring platform. Users provided consent at onboarding for their deidentified data to be used in aggregate for research purposes. The authors determined that the secondary analysis met the criteria for exemption from institutional review board oversight under the US Federal Policy for the Protection of Human Subjects (Common Rule), 45 CFR §46.104(d)(4)(ii) [24]. For the research analysis, information was recorded in a manner such that the identities of individuals could not readily be ascertained directly or through identifiers linked to the data. The investigators did not contact individuals and did not attempt, and will not attempt, to reidentify them. The study was conducted in accordance with applicable privacy and ethical standards.

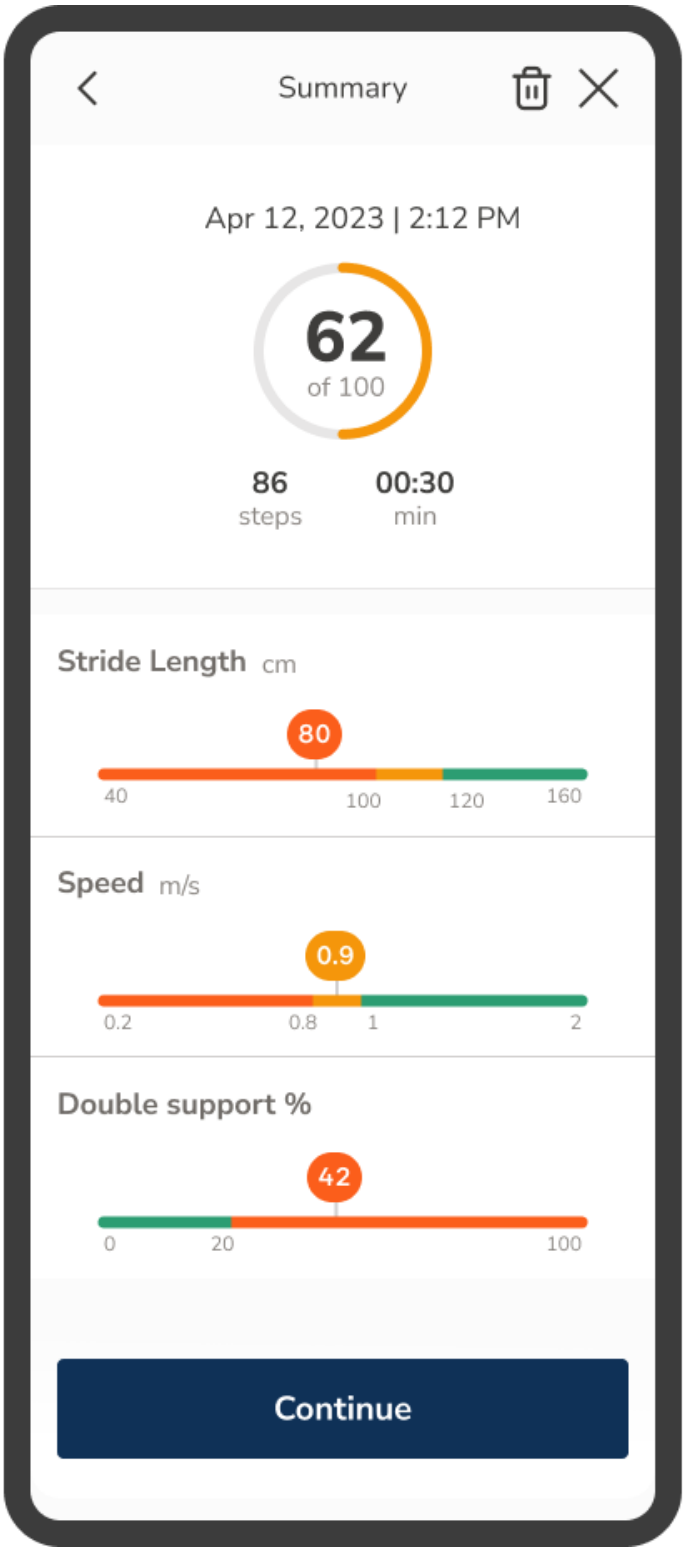
XSL•FO
RenderX

Gait Data Collection

Gait was measured before and after surgery with OneStep's smartphone gait analysis system (software versions 2.18.0 - 4.0.0; Celloscope Ltd). Figure 1 shows the measurement screen of the mobile app. Patients carried their phones in a pants pocket while walking. The sensors of the phone's IMU record 3-dimensional acceleration and gyroscope data at a sampling frequency of 50 Hz and upload the data to secure servers. No calibration is required. The inertial motion data are split into individual strides, and a supervised convolutional neural network then assigns each stride to a joint (physical activity, device body position) class; strides that are not assigned to the walking class or a pants pocket body position are excluded from the analyses reported here [25]. A proprietary deep learning framework is then used to detect which side the phone was carried on and compute standard spatiotemporal gait measures for each step. This framework was trained using a gold-standard gait measurement system (the C-Mill VR+ instrumented treadmill; Motek) to generate ground truth labels for simultaneous phone measurements from a variety of healthy and unhealthy walkers; these results were then validated against a selection of gold-standard gait measurement methods [15-17]. The software is compatible with almost all modern smartphones, as most smartphones contain the necessary IMU hardware, and is compatible with both iOS and Android operating systems. Although the gait analysis system has not been validated specifically in a post-TJA population, it has been validated in cohorts with a variety of moderate to severe gait abnormalities [16,17].

XSL•FO
RenderX

Figure 1. Screenshot of the OneStep mobile app walk analysis screen.



Nine gait parameters were analyzed for this study: cadence, single support percentage of the affected (ie, the leg upon which the operation was performed) and unaffected leg; stance asymmetry; double support percentage; step length of the affected and unaffected leg; stride length; and walk speed. Stance asymmetry is defined as the difference between the stance percentage of the 2 legs; that is, the percent of stride time each leg is in contact with the ground. For this study, it was calculated as the stance percentage of the unaffected leg minus the stance percentage of the affected leg. We used the signed value of stance asymmetry, rather than its absolute value, because the direction of asymmetry is clinically relevant in TJA, where one leg is affected.

Walk Types

We used 2 kinds of recordings:

1. *In-app walks*. Patients opened the OneStep mobile app, started a recording, placed the phone in a pocket, and

XSL•FO
RenderX

walked in a straight line (with 180° turns as necessary) on a flat surface for about 60 seconds.
2. *Background walks*. With permissions enabled, the app ran in the background and automatically recorded motion when it detected pocket-carried walking during daily life.

In-App Walk Collection

In-app walks were collected before and after surgery when patients chose to initiate a recording; no fixed recording schedule or surgeon-directed reminders were used. Recordings were performed in patients' own environments. App instructions directed patients to walk on a clear, straight path, to turn around and walk back and forth if the available path was too short, and to walk at a comfortable pace. Data processing and filtering are described above in *Gait Data Collection*.

Background Walk Collection

Background walks represent a particular challenge for data collection because walks can be taken in a variety of contexts (indoors vs outdoors, with or without shoes, on various terrains, straight-line walking or curved paths, etc) and because the patient can walk with their smartphone in a position other than the pants pocket. In addition, daily walking can occur in short bursts of 5 seconds or less, which have qualitatively different attributes than more extended walking bouts.

To account for these factors, the background gait analysis pipeline performs several steps. First, only walking bouts of 30 seconds or longer are analyzed. Second, any steps taken while not walking in a straight line are filtered out before analysis, and walking bouts with too few straight strides are flagged and filtered out. Straight-line walking was identified from gyroscope-derived heading after the sensor data were oriented to the direction of motion. Third, phone position (pocket vs nonpocket) is detected by a deep learning framework that was trained on examples of motion data from phones placed in the pocket and in a variety of other locations (hand, bag, shirt pocket, etc). Any walk that is detected as not being from the pants pocket is filtered out. Finally, gait measurements may differ between prompted and background conditions, as has been shown previously for this gait analysis system [23]. Therefore, as described below, in-app and background data are normalized separately to their respective presurgery baseline values. By normalizing in-app and background walks separately, lower background values at a given postsurgery week are interpreted relative to the patient's own presurgery background walks, not relative to their in-app walks. This distinction is important because raw background and in-app values may differ for reasons unrelated to recovery, including phone position, walking environment, and walking path. The analysis therefore asks whether background gait returns to its own baseline more slowly than in-app gait, rather than whether background gait has lower absolute parameter values.

Preoperative Baseline

Inclusion required at least one in-app gait recording and one background gait recording in the 14 days before surgery; patients without such data were excluded. If multiple recordings were available, their gait measures were averaged to define that patient's baseline. Baselines were calculated separately for in-app and background walks.

Aggregation and Normalization

For each patient, postoperative walks were grouped by week since surgery (days 1 - 7, 8 - 14, etc) and averaged to create weekly values. The patient's baseline was then subtracted from each weekly value to express change from baseline. To describe population-level recovery, these normalized values were averaged across patients contributing data in that postoperative week. Analyses used available data only (weekly sample sizes varied); no temporal interpolation or imputation was performed.

Statistical Analysis

To quantify differences between recovery trajectories, 3 methods were used. First, Kaplan-Meier analysis was used to determine whether time to first observed return to preoperative baseline differed between conditions. For parameters where a higher value is considered healthier (cadence, single support, step and stride length, and walk speed), time to return to baseline was considered the first week after surgery during which the parameter value was equal to or higher than the preoperative baseline level; for parameters where a lower value is healthier (double support and stance asymmetry), the first week in which the value was equal to or lower than baseline was used. A sustained crossing was not required because consecutive weekly observations were unavailable for many participants; requiring them would have excluded a large proportion of the cohort. Concretely, the Kaplan-Meier curve for a given parameter shows, for each postoperative week, the fraction of patients who had reached their own presurgery gait level by that week; a curve that rises sooner or reaches a higher fraction indicates faster or more complete return to baseline. A log-rank test (or a multivariate log-rank test, in cases where more than 2 Kaplan-Meier curves were being compared) was used to determine significance.

Second, we used linear mixed-effects models to test whether postoperative recovery trajectories differed between conditions. For each gait parameter, we fitted a model with fixed effects for postoperative week (modeled using a B-spline basis), condition (eg, THA vs TKA; in-app vs background; preoperative baseline level), age, and gender, and included interactions between the spline basis for week and each covariate of interest to test for differences in trajectory shape (time-varying effects) rather than only overall level shifts. A patient-specific random intercept was included to account for repeated measures within individuals. Postoperative week was represented with a B-spline basis of degree 1 and 4 degrees of freedom.

To test the importance of each fixed effect, we performed nested-model likelihood ratio tests by refitting reduced models that excluded one term at a time (condition, age, or gender) and its interactions, holding all remaining fixed effects, the postsurgery week specification, and random effects constant. The likelihood ratio statistic was computed as $2(\ell_{full} - \ell_{reduced})$, where ℓ is the maximized log-likelihood, and P values were obtained by comparing this statistic to a chi-square distribution with degrees of freedom equal to the difference in the number of estimated parameters between the full and reduced models.

XSL•FO
RenderX

Third, to compare parameter levels of late-stage plateaus, we computed estimated marginal means (EMMs) from the linear mixed-effects models. We defined the plateau as weeks 11 to 15 postsurgery and, for each condition, obtained the model-predicted mean over this window with age fixed at its sample mean and gender averaged by its observed proportions. Statistical significance of the difference in means between conditions was assessed with a Wald test, from which we report *z* scores and *P* values, along with 95% CIs.

All *P* values for all tests were adjusted using Benjamini-Hochberg false discovery rate correction over all *P* values in the analysis type (Kaplan-Meier, linear mixed-effects models, and EMMs). Statistical and modeling work was performed using Python and the Python packages lifelines, statsmodels, and SciPy; R and the R packages lme4, emmeans, and splines were used for the EMM analysis.

Results

Patient Demographics

Data were analyzed from 735 patients who underwent THA (n=350) or TKA (n=385). Baseline characteristics, including age distribution and gender composition used for subgroup analyses, are summarized in Table 1. The 2 cohorts had similar demographics: the THA cohort had a mean age of 69.6 (SD 7.6) years and included 56% (197/350) of women, while the TKA cohort had a mean age of 70.9 (SD 6.9) years and included 59% (227/385) of women. Figure S1 in Multimedia Appendix 1 shows the age distributions of the 2 cohorts. Preoperative baseline parameters differed slightly by gender, age, and procedure (Figure S2 in Multimedia Appendix 1). Figure 2 summarizes cohort assembly and filtering.

Table . Characteristics of total hip arthroplasty and total knee arthroplasty cohorts.

Characteristic	THA[a] (n=350)	TKA[b] (n=385)
Male participants, n (%)	153 (44)	158 (41)
Female participants, n (%)	197 (56)	227 (59)
Age (y), mean (SD)	69.6 (7.6)	70.9 (6.9)
Age of male participants (y), mean (SD)	69.2 (7.8)	70.9 (7.6)
Age of female participants (y), mean (SD)	69.9 (7.4)	70.9 (6.3)
Recorded in-app walks, n	8183	7289
Left-side surgery, n	191	214
Right-side surgery, n	159	171

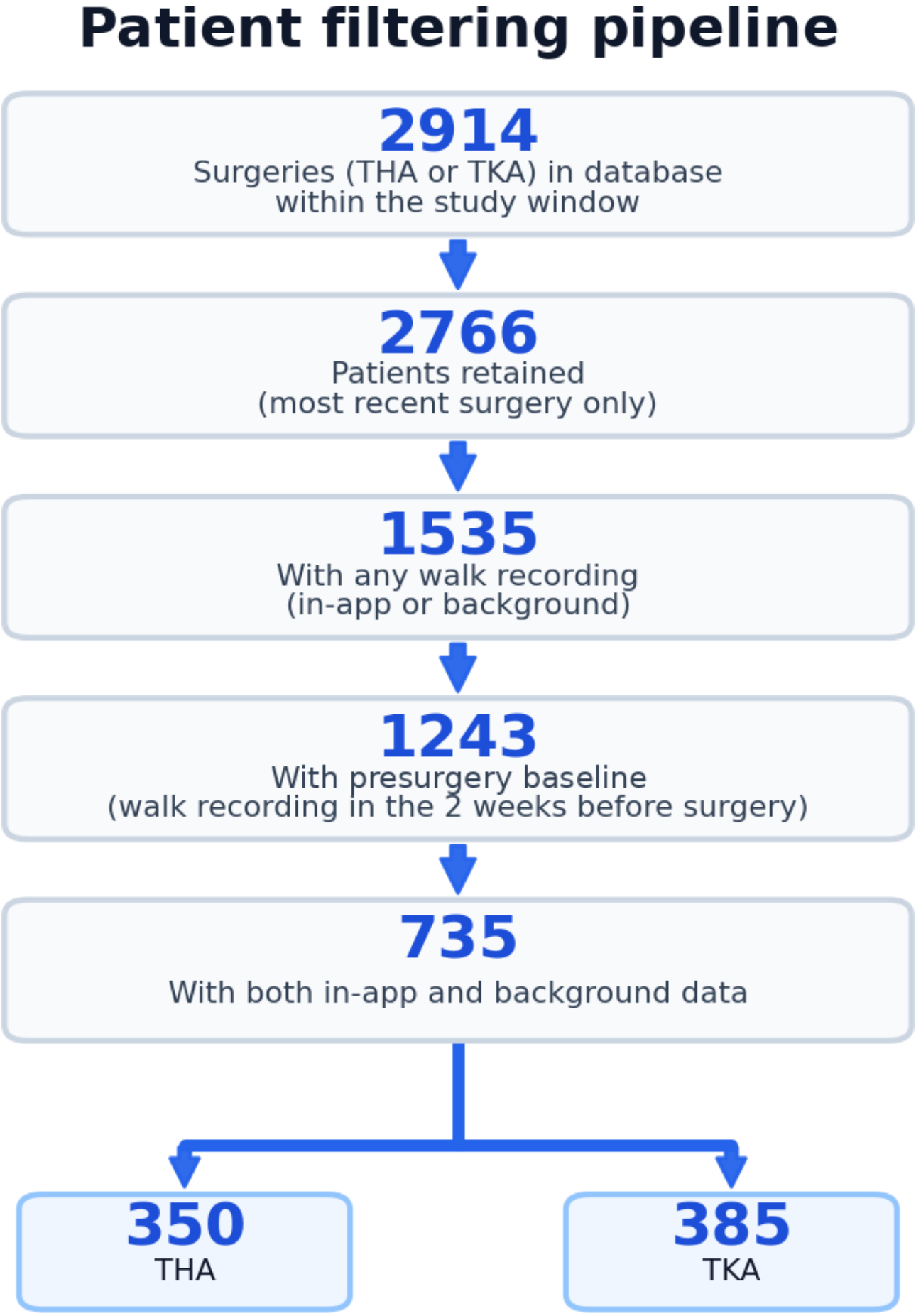
[a]THA: total hip arthroplasty.

[b]TKA: total knee arthroplasty.

XSL•FO
RenderX

Figure 2. Cohort selection flowchart. Flowchart showing derivation of the analytic cohort from all patients with a recorded THA or TKA. Patients were sequentially filtered according to study eligibility criteria, including procedure type, availability of postoperative gait data, and availability of both in-app and background preoperative recordings required to define baseline. The final cohort was then stratified into THA and TKA groups for analysis. THA: total hip arthroplasty; TKA: total knee arthroplasty.

XSL•FO
RenderX

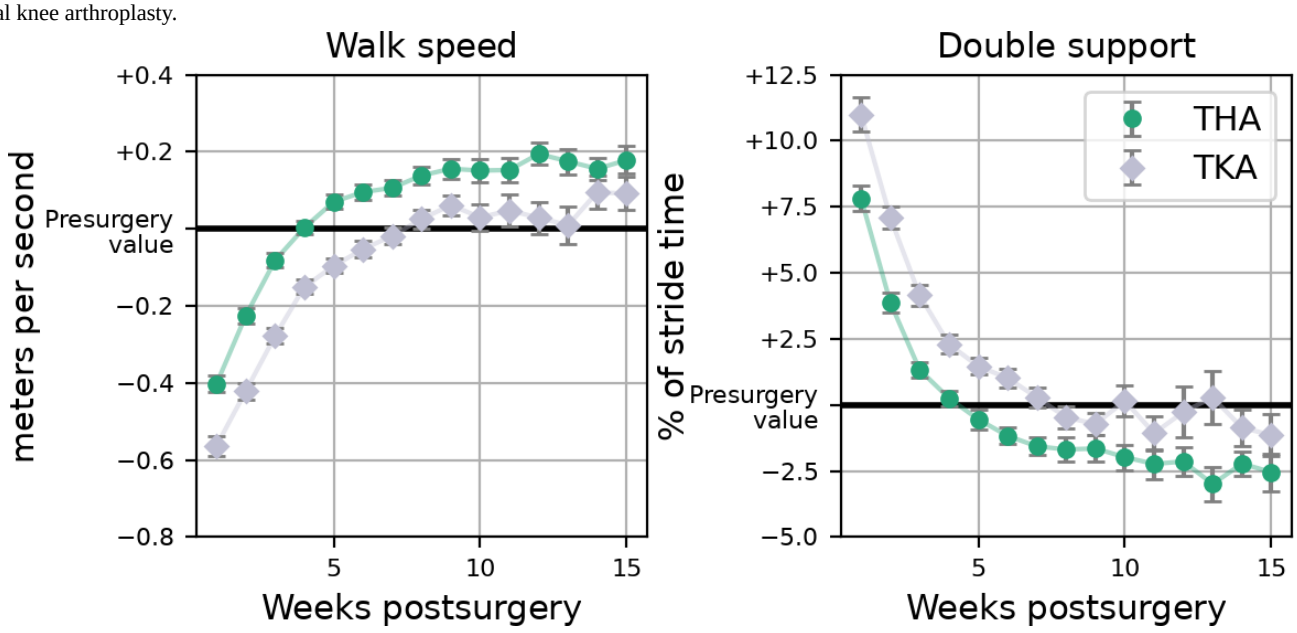


XSL•FO
RenderX

Recovery Trajectories After THA Versus TKA

Across all 9 gait parameters, average values relative to the preoperative baseline were lower after TKA than after THA in postoperative week 1. The mean THA trajectory returned to baseline by about week 4, whereas mean TKA trajectories reached baseline later, between weeks 6 to 8 (Figure 3 and Figure S3 in Multimedia Appendix 1). Kaplan-Meier analysis showed significant differences between THA and TKA in time to return to baseline for all parameters except single support of the unaffected leg (Figure S4 in Multimedia Appendix 1). Linear mixed-effects models showed that adding THA versus TKA status significantly improved fit for all gait parameters except single support of the unaffected leg (Figure S5 in Multimedia Appendix 1).

Figure 3. Recovery trajectories of in-app gait parameters after surgery in THA versus TKA. Each data point represents the mean change from all patients' preoperative baselines for that postoperative week; error bars show SE. Positive values indicate increases relative to baseline. Improvement corresponds to an increase in walk speed and a decrease in double support. For both parameters shown, THA patients returned to their preoperative gait level earlier than TKA patients, consistent with the known differences in procedural invasiveness and rehabilitation trajectory between the 2 surgeries. Weekly counts of patients contributing measurements in each group are given in Table S2 in Multimedia Appendix 1. THA: total hip arthroplasty; TKA: total knee arthroplasty.

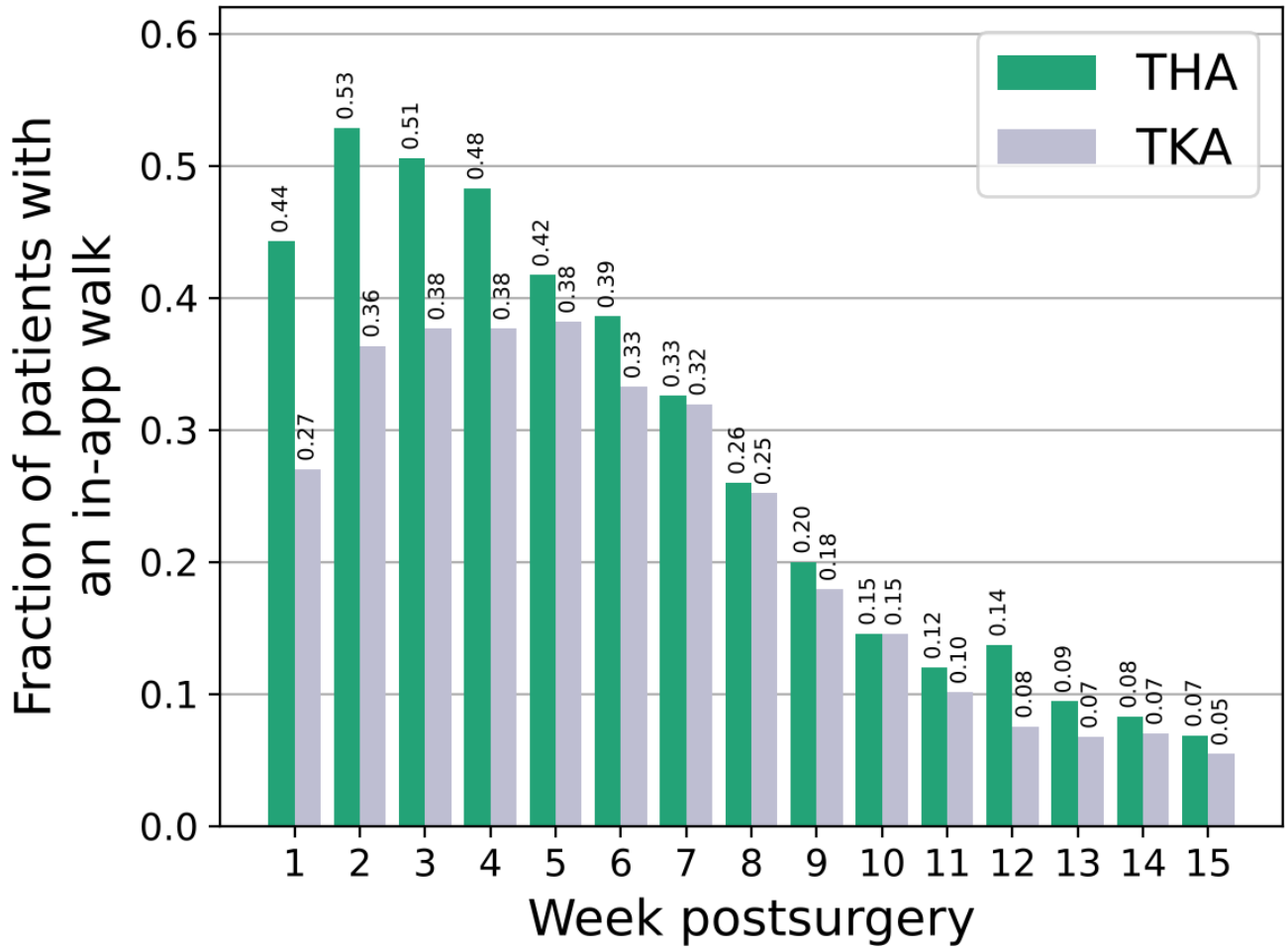


After crossing baseline, both procedures showed stable trajectories in weeks 11 to 15, with minimal week-to-week fluctuation. However, on average, THA patients reached a healthier level relative to the preoperative baseline than TKA patients at this plateau stage in all parameters except stance asymmetry and single support of the unaffected leg, and for most parameters, this difference reached an appreciable change in magnitude (Table S1 in Multimedia Appendix 1). For example, the difference in EMM at weeks 11 to 15 for walk speed was 0.14 m/s, exceeding the commonly used 0.1 m/s threshold for a clinically meaningful change in walk speed [26].

In the final weeks, some parameters showed unusual fluctuations or shifts in plateau level. This likely reflects diminishing and variable weekly sample sizes, as not every patient recorded an in-app walk each week and recordings tend to decline over time (Figure 4 and Figure S6 and Table S2 in Multimedia Appendix 1). This incompleteness may similarly affect the other analyses, including the Kaplan-Meier curves, and raises the possibility of survivorship bias if patients who continued contributing data differed systematically from those who did not; we consider these issues further in the *Discussion*.

XSL•FO
RenderX

Figure 4. Fraction of patients contributing in-app walks by postoperative week. For each postoperative week, bars show the fraction of patients in each cohort who recorded at least one in-app walk that week, out of all patients in the cohort (THA, n=350; TKA, n=385). Adherence peaks in the early postoperative weeks and falls steadily thereafter in both groups, reaching under 10% by week 15. Absolute weekly patient counts for these groups, as well as for background walks and subgroups, are given in Table S2 in Multimedia Appendix 1. THA: total hip arthroplasty; TKA: total knee arthroplasty.

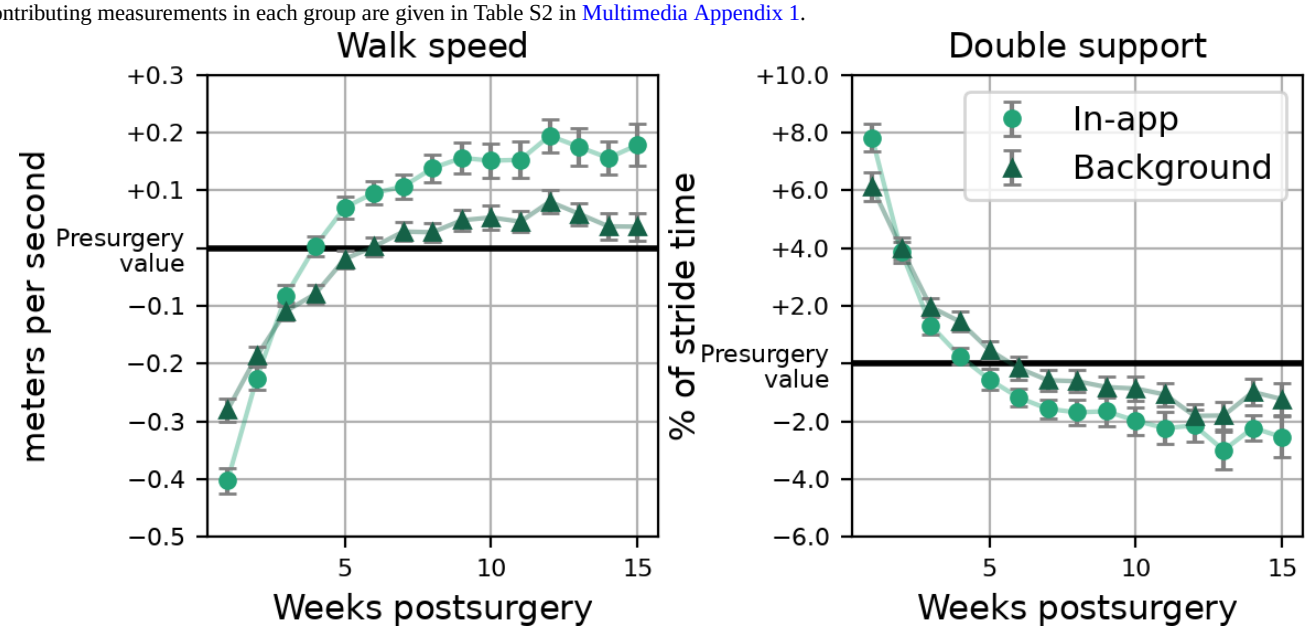


Recovery Trajectories of In-App Versus Background Measurements

We next compared recovery trajectories from prompted in-app walks with those from passive background activity. In both THA and TKA, differences in recovery trajectories between these 2 types of walking were observed (Figure 5 and Figure S7 in Multimedia Appendix 1). Kaplan-Meier analysis did not show significant differences in time to return to baseline between in-app and background walking, except for walk speed and step length of the unaffected leg after THA (Figure S8 in Multimedia Appendix 1). However, linear mixed-effects models showed that adding in-app versus background status significantly improved fit for all gait parameters, most at a very high level of significance (Figure S9 in Multimedia Appendix 1). The level of the late-stage plateau differed significantly between in-app and background walking after THA for all parameters except stance asymmetry and single support of the unaffected leg. Effect sizes were generally of a lower magnitude than those of the plateau differences between THA and TKA, although the difference in walk speed was 0.14 m/s, the same as the difference between THA and TKA (Table S1 in Multimedia Appendix 1). The plateau levels after TKA differed significantly between in-app and background walking in cadence, step length of both legs, stride length, and walk speed; however, effect sizes were much lower than even those of in-app versus background walking in THA, including a difference in walk speed of 0.07 m/s (Table S1 in Multimedia Appendix 1).

XSL•FO
RenderX

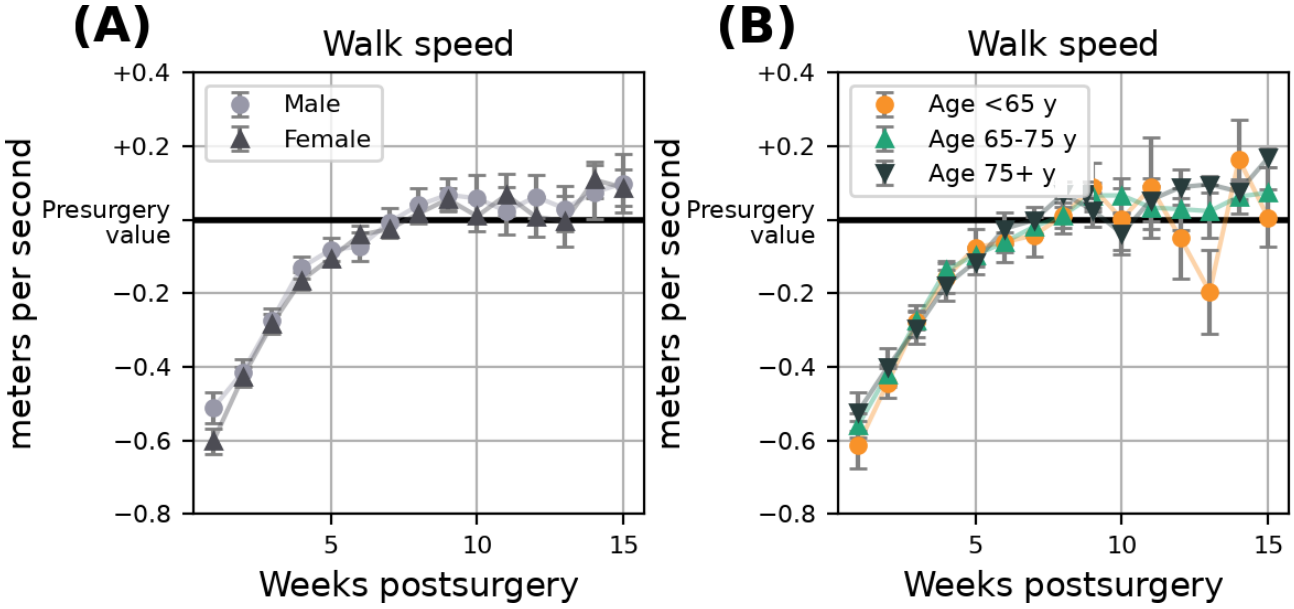
Figure 5. Recovery trajectory of in-app versus background gait parameters after surgery in total hip arthroplasty. Each data point represents the mean change from all patients' preoperative baselines for that postoperative week; error bars show SE. Positive values indicate increases relative to baseline. Improvement corresponds to an increase in walk speed and a decrease in double support. Background walking returns to baseline more slowly than in-app walking, suggesting that prompted clinical assessments may overestimate a patient's functional recovery in daily life. Weekly counts of patients contributing measurements in each group are given in Table S2 in Multimedia Appendix 1.



Recovery Trajectories by Age and Gender

We performed these analyses with populations stratified by gender and age (Figure 6 and Figures S10 and S11 in Multimedia Appendix 1). By gender, trajectories were visually highly similar for THA (male, n=153; female, n=197) and TKA (male, n=158; female, n=227). By age, patterns were likewise highly similar across groups in THA (<65 y: n=86; 65 - 75 y: n=173; 75+ y: n=91) and TKA (<65 y: n=76; 65 - 75 y: n=200; 75+ y: n=109). Kaplan-Meier analysis showed no difference in time to return to baseline between gender or age groups for any parameter (Figures S12 and S13 in Multimedia Appendix 1). Although adding age or gender to the linear mixed-effects models significantly improved model performance for many parameters (Figures S5 and S9 in Multimedia Appendix 1), the significance of those improvements was generally many orders of magnitude lower than that observed when adding condition (eg, THA vs TKA, in-app vs background).

Figure 6. In-app recovery trajectory in walk speed after total knee arthroplasty stratified by demographic characteristics. (A) Recovery trajectories stratified by gender (male, n=158; female, n=227). (B) Recovery trajectories stratified by age (<65 y: n=76; 65 - 75 y: n=200; 75+ y: n=109). The overlap between demographic subgroups indicates that patient age or gender play a limited role in the shape of the recovery trajectory at the population level (in contrast to the absolute level of gait parameters, which do differ between these subgroups). Weekly counts of patients contributing measurements in each group are given in Table S2 in Multimedia Appendix 1.



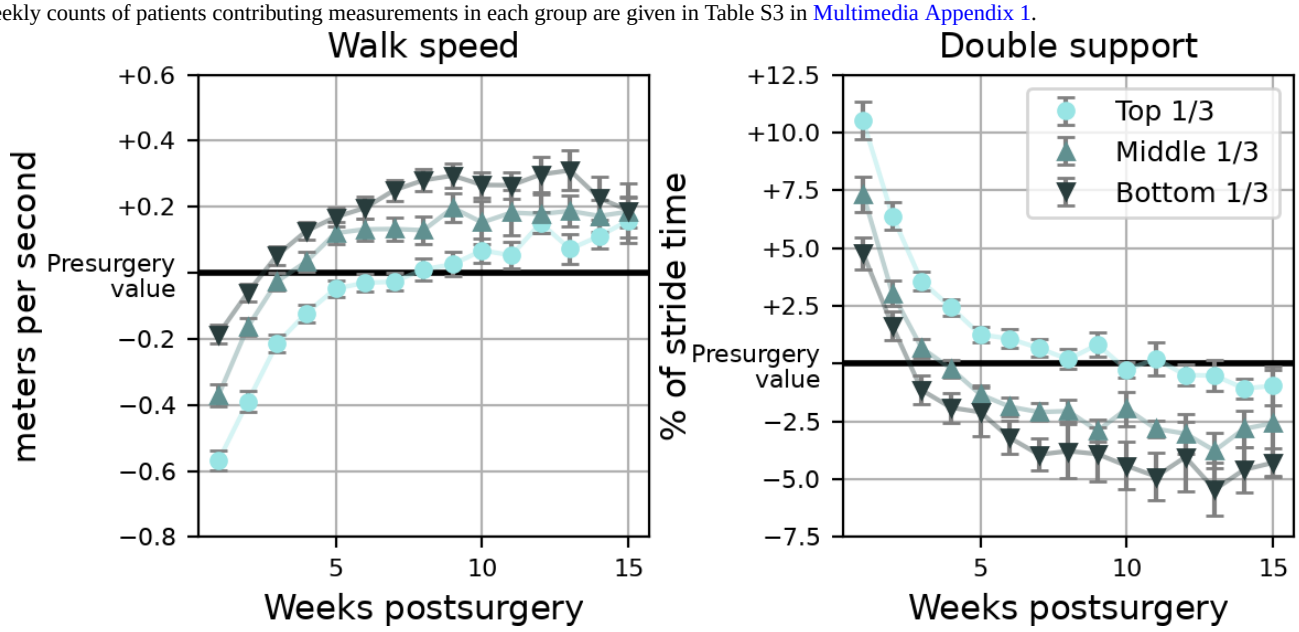
Weekly measurement counts by age and gender are shown in Table S2 in Multimedia Appendix 1.

XSL•FO
RenderX

Recovery Trajectories by Preoperative Baseline

In addition to splitting patients by age and gender, we investigated whether preoperative baseline gait level was associated with postoperative recovery trajectories. Patients were split into 3 groups, representing the upper, middle, and lower thirds of preoperative baseline levels for each parameter. These splits were done independently for each parameter. Recovery trajectories varied greatly between the 3 groups (Figure 7 and Figure S14 in Multimedia Appendix 1). The highest-level walkers before surgery had the greatest decline in gait parameters in the week immediately following surgery and, by week 15, had reached only approximately their preoperative baseline on average. In contrast, the lowest-level walkers had a smaller decline immediately after surgery and, on average, reached well past their preoperative baseline by week 15. Kaplan-Meier and linear mixed-effects model analyses confirmed that these trajectories were significantly different for all parameters, in both THA and TKA and for in-app and background walking (Figures S15 and S16 in Multimedia Appendix 1). It should be noted that, in absolute terms, the highest-baseline group maintained better gait values at week 15 than the lowest-baseline group. However, the lowest-baseline group showed larger improvements relative to its own baseline (Figure S17 in Multimedia Appendix 1).

Figure 7. In-app recovery trajectory after total hip arthroplasty stratified by preoperative baseline level. Recovery trajectories were stratified by gait parameter values during a 14-day preoperative period; patients were split into 3 groups representing the upper, middle, and lower thirds of preoperative baseline levels for each parameter. These findings show that patients with poorer preoperative gait show larger relative gains but do not reach the absolute gait levels of higher-functioning patients, while high-functioning patients often do not return to their own baseline until later time points. Weekly counts of patients contributing measurements in each group are given in Table S3 in Multimedia Appendix 1.



Discussion

Principal Findings

Using longitudinal smartphone-based gait analysis, we characterized week-by-week recovery trajectories after THA and TKA across multiple spatiotemporal measures. On average, THA patients returned to their preoperative baseline by about week 4, while TKA patients did so between weeks 6 to 8, and THA patients reached a higher plateau thereafter. These differences match clinical experience and prior reports comparing postoperative mobility between procedures, which show that recovery is slower and pain persists for longer in TKA compared to THA [27,28].

A key finding is the significant difference between prompted in-app walks and passive background walking across many gait parameters. This pattern highlights the well-documented discordance between functional capacity and habitual performance [20-23], which may persist for years after surgery [29]. However, this discordance cannot be attributed solely to behavioral compensation because the 2 walking modalities are collected under fundamentally different conditions. In-app walks are performed on a flat surface under deliberate conditions, while background walks occur across heterogeneous real-world environments. Terrain type systematically influences gait parameters: energy expenditure and stride variability change substantially across surface conditions, from sidewalk to gravel to grass, and these differences are detectable by body-worn inertial sensors [30]. Walking context also matters, as indoor and outdoor walking shape distinct digital mobility signatures in ways that averaged measures can obscure [31]. Bout length is another well-established confounder of passively collected gait measures. Walking speed increases systematically with bout length, yet most real-world walking occurs in short bouts. Consequently, passive gait measures that are aggregated without accounting for bout length may not reflect gait capacity at longer sustained efforts [32,33]. This effect influences both the magnitude and reliability of extracted gait features across a range of clinical populations [33]. In the context of TKA, prearthroplasty individuals already exhibit shorter gait bouts and reduced daily walking activity compared with healthy peers [34], suggesting that passive gait measurements in this cohort are shaped not only by gait quality but also by activity patterns reflecting pain, fatigue, and behavioral avoidance.

XSL•FO
RenderX

The environmental and behavioral factors described above represent sources of structural heterogeneity in background walking that cannot be fully eliminated by algorithmic filtering. To mitigate these potential confounders, we adopted a within-person normalization approach, in which in-app and background walks were each normalized separately to their respective presurgery baselines. Any stable systematic differences between the 2 walk types arising from environmental context, walking or bout characteristics, or other persistent factors are thereby incorporated into the baseline and subtracted out. The comparison of interest is therefore not whether background and in-app parameter values differ in absolute terms: they do, and the absolute presurgery and postsurgery values for both walk types are reported here. Rather, the key comparison in this study is whether the rate and completeness of return to each condition's own baseline differ between the in-app and background walks. Using this normalized framework, background gait consistently returns to its own presurgery level more slowly than in-app gait. We cannot fully exclude the possibility that the ecological composition of background walks shifts after surgery; for instance, early recovery may involve a greater proportion of short indoor walking bouts that differ systematically from presurgery background activity. (Because our analysis included only walking bouts lasting at least 30 s, we were unable to characterize the full distribution of walking bout lengths and contexts over the recovery period.) That being said, a shift such as this would still suggest that the patient has not returned to their regular presurgery daily walking behavior, even as their conscious gait has returned to its presurgery state. Nevertheless, the normalization approach substantially mitigates the concern that the observed divergence simply reflects a fixed structural difference between walk types, and the consistency of the pattern across the gait parameters in 2 surgical cohorts supports its robustness. Collectively, these results indicate that prompted and passively captured walks follow different recovery trajectories, and that prompted testing alone would place the point at which gait parameters stabilize earlier than the passive data indicate. What stabilization in either measure represents clinically and functionally, and which trajectory better reflects meaningful recovery, will require studies anchored to patient-reported or clinical outcomes.

Surprisingly, age- and gender-stratified trajectories appeared similar across both procedures, suggesting that the overall shape and timing of recovery, when normalized to each patient's own baseline, are broadly robust at the population level. Although age and gender improved model fit for several parameters, the magnitude of improvement was small relative to procedure type and walk context, and neither factor was associated with differences in time to baseline in Kaplan-Meier analyses. This is at least partially at odds with prior literature, which has identified age and sex as meaningful predictors of post-TJA outcomes, though with important nuance. For TKA specifically, Pua et al [13] found older age to be among the strongest predictors of lower postoperative gait speed in a longitudinal cohort, and the large meta-analysis by Olsen et al [35], covering over 11,000 patients, identified better preoperative physical function as correlated with better 1-year physical function after TKA, though age and sex effects were inconsistent across the included studies. For THA, the predictor literature is less definitive: a systematic review by Hofstede et al [36] covering 35 studies and over 138,000 patients found that results for age, gender, and comorbidity were conflicting across studies, with preoperative function being the most consistently identified predictor. Examining both procedures jointly, Kennedy et al [37] followed 152 THA and TKA patients across the first 4 postoperative months and found that gender and baseline preoperative function both shaped early recovery trajectories, with women tending to present with greater preoperative disability and showing correspondingly different recovery curves. The apparent absence of age and gender effects in the current study should not be taken as evidence that these factors are clinically irrelevant. A key feature of our design likely attenuates them: because trajectories are normalized to each patient's own preoperative baseline, the analysis asks whether different groups of patients recover *relative to where they started* at a different rate, not whether they reach the same absolute gait level. Since absolute gait capacity differs systematically by age and sex at baseline, normalization effectively controls for these differences, which could explain why trajectory shapes converge even when absolute gait levels do not. The finding that older patients do not take longer to return to their own preoperative baseline, while their actual gait parameter values remain lower than those of younger patients, is meaningful in its own right: it suggests that the relative recovery process is broadly consistent across demographic groups, even if absolute functional levels diverge.

In contrast, recovery trajectories differed substantially by preoperative gait level. Patients with the poorest preoperative baseline gait performance exhibited the largest improvement relative to baseline, likely reflecting greater room for improvement. In contrast, patients with higher preoperative baseline gait performance showed smaller relative gains, likely because they began closer to their functional ceiling. Nevertheless, when postoperative gait was evaluated using absolute rather than baseline-normalized values, these higher-functioning patients still achieved higher absolute postoperative gait levels. Thus, relative improvement and absolute gait performance capture distinct aspects of recovery and should be interpreted together. These phenomena are well documented in the TJA literature. Across TKA studies, preoperative gait function, quadriceps strength, and overall physical function have consistently predicted postoperative performance-based outcomes, gait-speed recovery, and 1-year physical function, while lower baseline function and higher BMI have been associated with poorer absolute outcomes [13,35,38,39]. At the same time, patients with higher preoperative gait function may show smaller relative gains or even early functional decline, consistent with our finding that the highest-baseline group returned only approximately to its own preoperative level by week 15 [38]. Similar patterns have been reported after THA: patients with worse preoperative function often improve more, but do not reach the absolute postoperative function achieved by those who started at a higher level [36,37].

Taken together, the results of this study support recovery targets that are individualized to each patient's baseline function rather than defined by a single absolute threshold. Whether the

XSL•FO
RenderX

difference in baseline is due to demographic considerations or individual variation, valuable clinical information is lost when recovery is evaluated solely against population-based gait parameter values. A patient who walks slowly before surgery and returns to that level has a different clinical interpretation than a higher-functioning patient who fails to recover to their own preoperative level. Thus, both the absolute level of gait performance and the trajectory of recovery relative to baseline provide distinct and complementary clinical information.

Limitations

Within this observational study, we did not have access to and therefore did not control for all factors known to affect recovery, such as surgical approach, implant type, perioperative protocols, rehabilitation protocols, pain level, or comorbidities [40]. Further, no anthropometric normalization was performed (eg, normalizing stride length by patient height) because such data were also unavailable. Age was available for patients, and we grouped patients using clinically meaningful age ranges rather than equal-sized groups. This allowed us to compare age categories that are more interpretable clinically, but the resulting uneven group sizes should be considered when interpreting the age-stratified analyses.

We chose to use the return to preoperative baseline as the primary metric in this study, for the reasons described above. However, because most patients expect to achieve better function and/or less pain than they experienced in the weeks before surgery, a return to baseline should not be conflated with full recovery. In this study, the data did not allow for an analysis focused on, for example, reaching a clinically meaningful level above the presurgery baseline, since many patients had not exceeded their baseline by a noticeable margin by week 15. Future studies in this vein should aim to extend the timeline of data collection so as to be able to directly assess recovery beyond baseline.

Time to return to baseline, as calculated with Kaplan-Meier analysis, used the first week in which the patient reached their presurgery level. This definition may overestimate the speed of recovery in patients whose first crossing of baseline was transient. Functional recovery after TJA can fluctuate because of pain, fatigue, activity level, or other short-term factors, and a patient who reaches baseline in 1 week may fall below baseline in a subsequent week. The Kaplan-Meier results should therefore be interpreted as time to first observed return to baseline, rather than time to sustained recovery beyond baseline.

While smartphone-based data collection vastly increases the number of subjects it is possible to include in these types of studies, one downside compared with using dedicated hardware is that there is much more missing data within a subject's dataset. Adherence and phone-carriage behaviors (whether people carry their phones in their pants pockets while walking) influence who contributes data over time, and environmental variability can affect background measurements as well. For this reason, we do not see multiple passive walking bouts per day in many subjects, whereas previous studies using more dedicated measurement systems do find walking bouts at that frequency, although in cohort sizes more than an order of magnitude smaller than our study [31,41].

We summarized cohorts without imputing missing weeks, so missingness could relate to recovery status. Further, diminishing sample sizes in later weeks (particularly for in-app walks) reduce statistical power and make late-week estimates highly sensitive to outliers. This analysis may also be subject to survivorship bias because patients who continued contributing data at later weeks may not be representative of the original cohort. In particular, if continued participation is related to recovery status, the observed results may reflect both true recovery and the changing composition of the patients who remained under observation. Therefore, the values at later weeks should be interpreted with caution.

Passive data collection depends on the patient carrying their smartphone in a pants pocket, which may introduce selection bias. Our filtering methods for removing walks in which the smartphone was outside the pants pocket could also have failed to filter some such walks, introducing noisy and unreliable data into the dataset. In addition, as discussed above, passively collected data reflect walking undertaken in a variety of conditions, which makes such data inherently different from prompted data collection. This fact is mitigated by our normalization of the in-app and background data separately to their respective presurgery levels, but there could still be an effect; for example, if the types of passive walks undertaken after surgery differ systematically from those taken before surgery. Finally, our analysis ended at 15 weeks after surgery, so later improvements or setbacks were not captured.

Conclusions and Implications for Future Research

Despite these limitations, this study provides a robust quantitative characterization of gait-parameter recovery trajectories in the first 15 weeks after TJA at weekly resolution and demonstrates the power of smartphone-based gait analysis to generate longitudinal datasets at scale. The differences in in-app and background walking that we observed suggest that clinical decisions, such as when to taper walking aids or advance activity, could benefit from incorporating passive background metrics rather than relying solely on prompted tests. However, studies that directly assess these questions are needed before any concrete clinical recommendations can be made. Further analyses of this dataset could quantify within-week heterogeneity, identify early signals associated with delayed recovery, and cluster longitudinal trajectories to define distinct recovery trajectory profiles. Defining and validating these profiles and linking them to clinical correlates could support clearer counseling and targeted interventions. Beyond this dataset, future work should extend follow-up, link trajectories to complications and PROMs, and incorporate factors such as BMI, height, and rehabilitation exposure.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank the clinical teams and patients who participated in the recovery monitoring program. We also acknowledge the support of the OneStep engineering team for their assistance in data extraction and technical documentation of the gait analysis framework. Special thanks to Glen Pridham from the Weizmann Institute of Science for his advice on statistical analysis.

Generative AI Use

Generative AI ChatGPT (GPT-4o; OpenAI) was used during the preparation of this manuscript to assist with writing code for some analyses, assist with the initial drafting of standard end sections, and refine the linguistic clarity of the technical discussion. Claude (claude-opus-5-5; Anthropic) was used during copyediting. The authors reviewed and edited all AI-generated content and take full responsibility for the final manuscript.

Funding

This study was supported by Celloscope Ltd (OneStep).

Data Availability

Aggregate data from this study are provided in Multimedia Appendix 2, containing weekly group-level data organized by procedure, walk type, postoperative week, and gait parameter. Participant-level data cannot be shared because users of the OneStep platform consented only to the use of their deidentified data in aggregate for research purposes. Raw sensor data and proprietary intermediate data structures are not publicly available.

Authors' Contributions

Conceptualization: LST, YYN
Data curation: LST
Formal analysis: LST
Investigation: LST
Methodology: LST, YYN
Project administration: YYN, BM, OYL
Software: LST, EG-H
Supervision: YYN, BM, OYL
Visualization: LST
Writing – original draft: LST, EG-H
Writing – review & editing: LST, EG-H, YYN, BM, OYL, PAL

Conflicts of Interest

LST, EG-H, and YYN are employees of Celloscope Ltd (OneStep), the company that developed the smartphone gait analysis system evaluated in this study. BM, OYL, PAL, and their respective institutions declare no conflicts of interest.

Multimedia Appendix 1
Supplementary figures and tables supporting the analysis of postoperative gait recovery after total hip and total knee arthroplasty.
[PDF File, 9118 KB - mhealth_v14i1e97657_app1.pdf]

Multimedia Appendix 2
Aggregate (group-level) mean values of the 9 gait parameters analyzed in this study, provided for each combination of procedure (total hip arthroplasty, total knee arthroplasty), walk type (in-app, background), and week (preoperative baseline through week 15 postsurgery).
[ZIP File, 19 KB - mhealth_v14i1e97657_app2.zip]

References

1. Patel N, Golwala P. Approaches for total hip arthroplasty: a systematic review. Cureus 2023 Feb;15(2):e34829. [doi: 10.7759/cureus.34829] [Medline: 36919077]
2. Lee YS. Comprehensive analysis of pain management after total knee arthroplasty. Knee Surg Relat Res 2017 Jun 1;29(2):80-86. [doi: 10.5792/ksrr.16.024] [Medline: 28545171]
3. American Joint Replacement Registry (AJRR): 2025 annual report. : American Academy of Orthopaedic Surgeons (AAOS); 2025 URL: https://www.aaos.org/registries/publications/ajrr-annual-report/ [accessed 2026-01-10]
4. Shichman I, Roof M, Askew N, et al. Projections and epidemiology of primary hip and knee arthroplasty in Medicare patients to 2040-2060. JB JS Open Access 2023;8(1):e22.00112. [doi: 10.2106/JBJS.OA.22.00112] [Medline: 36864906]

XSL•FO
RenderX

5. Wainwright TW, Kehlet H. Functional recovery following hip and knee arthroplasty: subjective vs. objective assessment? Acta Orthop 2022 Sep 15;93:739-741. [doi: 10.2340/17453674.2022.4567] [Medline: 36111867]
6. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991 Feb;39(2):142-148. [doi: 10.1111/j.1532-5415.1991.tb01616.x] [Medline: 1991946]
7. Muscatelli S, Zheng H, Muralidharan A, Tollemar V, Hallstrom BR. Limiting the surveillance period to 90 days misses a large portion of infections in the first year after total hip and knee arthroplasty. Arthroplast Today 2022;16:90-95. [doi: 10.1016/j.artd.2022.04.009] [Medline: 35662991]
8. Garabedian M, Legler J, Benzouak T, Baskaran S, Mutch J, Albers A. Efficacy of lower-limb wearables to assess recovery following total hip or knee arthroplasty: a systematic review and meta-analysis. J Arthroplasty 2026 May;41(5):1548-1561. [doi: 10.1016/j.arth.2025.09.023] [Medline: 40998067]
9. Wu KA, Kugelman DN, Goel RK, et al. How do patient-reported outcomes correlate with real-time objective measures of function after total knee arthroplasty? A prospective study using daily gait metrics. J Arthroplasty 2025 Jul;40(7):1711-1718. [doi: 10.1016/j.arth.2024.12.007] [Medline: 39667590]
10. Guild GN 3rd, Najafi F, DeCook CA, et al. Evaluating knee recovery beyond patient reports: a comparative study of smart implantable device-derived gait metrics versus patient-reported outcome measures in total knee arthroplasty. J Arthroplasty 2024 Dec;39(12):2961-2969.e1. [doi: 10.1016/j.arth.2024.05.091] [Medline: 38852690]
11. Wang K, Eftang CN, Jakobsen RB, Årøen A. Review of response rates over time in registry-based studies using patient-reported outcome measures. BMJ Open 2020 Aug 6;10(8):e030808. [doi: 10.1136/bmjopen-2019-030808] [Medline: 32764078]
12. Fary C, Cholewa J, Abshagen S, et al. Stepping beyond counts in recovery of total knee arthroplasty: a prospective study on passively collected gait metrics. Sensors (Basel) 2023 Jun 14;23(12):5588. [doi: 10.3390/s23125588] [Medline: 37420754]
13. Pua YH, Seah FJT, Clark RA, Lian-Li Poon C, Tan JWM, Chong HC. Factors associated with gait speed recovery after total knee arthroplasty: a longitudinal study. Semin Arthritis Rheum 2017 Apr;46(5):544-551. [doi: 10.1016/j.semarthrit.2016.10.012] [Medline: 27894727]
14. Halvorson RT, Torres-Espin A, Cherches M, Callahan M, Vail TP, Bailey JF. Predicting recovery following total hip and knee arthroplasty using a clustering algorithm. Arthroplast Today 2024;27:101395. [doi: 10.1016/j.artd.2024.101395] [Medline: 39071835]
15. Christensen JC, Stanley EC, Oro EG, et al. The validity and reliability of the OneStep smartphone application under various gait conditions in healthy adults with feasibility in clinical practice. J Orthop Surg Res 2022 Sep 14;17(1):417. [doi: 10.1186/s13018-022-03300-4] [Medline: 36104792]
16. Marom P, Brik M, Agay N, et al. The reliability and validity of the OneStep smartphone application for gait analysis among patients undergoing rehabilitation for unilateral lower limb disability. Sensors (Basel) 2024 Jun 2;24(11):3594. [doi: 10.3390/s24113594] [Medline: 38894386]
17. Contreras C, Stanley EC, Deschamps-Prescott C, et al. Evaluation of smartphone technology on spatiotemporal gait in older and diseased adult populations. Sensors (Basel) 2024 Sep 9;24(17):5839. [doi: 10.3390/s24175839] [Medline: 39275750]
18. Bea T, Chaabene H, Freitag CW, Schega L. Psychometric characteristics of smartphone-based gait analyses in chronic health conditions: a systematic review. J Funct Morphol Kinesiol 2025 Apr 16;10(2):133. [doi: 10.3390/jfmk10020133] [Medline: 40566429]
19. Wu KA, Kugelman DN, Rosas S, et al. Improved walking steadiness following total hip arthroplasty compared to total knee arthroplasty. Arthroplast Today 2025;35:101802. [doi: 10.1016/j.artd.2025.101802] [Medline: 40893391]
20. Tomey KM, Sowers MR. Assessment of physical functioning: a conceptual model encompassing environmental factors and individual compensation strategies. Phys Ther 2009 Jul;89(7):705-714. [doi: 10.2522/ptj.20080213] [Medline: 19443558]
21. Hassija N, Hill E, Dawes H, Mayo NE. Comparability of methods for remotely assessing gait quality. Sensors (Basel) 2025 Jun 14;25(12):3733. [doi: 10.3390/s25123733] [Medline: 40573622]
22. Foucher KC, Thorp LE, Orozco D, Hildebrand M, Wimmer MA. Differences in preferred walking speeds in a gait laboratory compared with the real world after total hip replacement. Arch Phys Med Rehabil 2010 Sep;91(9):1390-1395. [doi: 10.1016/j.apmr.2010.06.015] [Medline: 20801257]
23. Rozanski G, Putrino D. Recording context matters: differences in gait parameters collected by the OneStep smartphone application. Clin Biomech (Bristol) 2022 Oct;99:105755. [doi: 10.1016/j.clinbiomech.2022.105755] [Medline: 36058106]
24. 45 CFR § 46.104 — exempt research. Electronic Code of Federal Regulations (eCFR). URL: https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-A/part-46/subpart-A/section-46.104 [accessed 2026-10-07]
25. Naveh Y, Davidson S, Shussman T. System, method and computer program product for assessment of a user's gait. 2020 Sep 17 URL: https://patents.google.com/patent/US20200289027A1/en [accessed 2026-10-17]
26. Perera S, Mody SH, Woodman RC, Studenski SA. Meaningful change and responsiveness in common physical performance measures in older adults. J Am Geriatr Soc 2006 May;54(5):743-749. [doi: 10.1111/j.1532-5415.2006.00701.x] [Medline: 16696738]

XSL•FO
RenderX

27. Judd DL, Wolfe P, LeDoux CV, Hogan C, Dayton MR, Stevens-Lapsley JE. Trajectories of functional performance and muscle strength recovery differ after total knee and total hip replacement: a performance-based, longitudinal study. Int J Rehabil Res 2019 Sep;42(3):211-216. [doi: 10.1097/MRR.0000000000000344] [Medline: 31219844]
28. Kawaguchi K, Kuwakado S, Otsuka H, et al. Comparison between total hip and knee arthroplasty on short-term performance-based outcomes and factors associated with the improvement of gait function during post-acute inpatient rehabilitation. J Clin Med 2024 Oct 24;13(21):6381. [doi: 10.3390/jcm13216381] [Medline: 39518521]
29. Vissers MM, Bussmann JB, de Groot IB, Verhaar JAN, Reijman M. Physical functioning four years after total hip and knee arthroplasty. Gait Posture 2013 Jun;38(2):310-315. [doi: 10.1016/j.gaitpost.2012.12.007] [Medline: 23829981]
30. Kowalsky DB, Rebula JR, Ojeda LV, Adamczyk PG, Kuo AD. Human walking in the real world: interactions between terrain type, gait parameters, and energy expenditure. PLoS One 2021;16(1):e0228682. [doi: 10.1371/journal.pone.0228682] [Medline: 33439858]
31. Kirk C, Rehman RZU, Galna B, et al. Toward an understanding of real-world mobility in Parkinson's: insights from enhanced contextualisation using GPS-derived location and data-driven modeling of walking speed. Front Aging Neurosci 2026;18:1746429. [doi: 10.3389/fnagi.2026.1746429] [Medline: 41798286]
32. Mueller A, Hoefling HA, Muaremi A, et al. Continuous digital monitoring of walking speed in frail elderly patients: noninterventional validation study and longitudinal clinical trial. JMIR mHealth uHealth 2019 Nov 27;7(11):e15191. [doi: 10.2196/15191] [Medline: 31774406]
33. Shah VV, McNames J, Harker G, et al. Effect of bout length on gait measures in people with and without Parkinson's disease during daily life. Sensors (Basel) 2020 Oct 12;20(20):5769. [doi: 10.3390/s20205769] [Medline: 33053703]
34. Boekesteijn RJ, Keijsers NLW, Defoort K, et al. Real-world gait and turning in individuals scheduled for total knee arthroplasty. Clin Biomech (Bristol) 2024 Oct;119:106332. [doi: 10.1016/j.clinbiomech.2024.106332] [Medline: 39241348]
35. Olsen U, Lindberg MF, Rose C, et al. Factors correlated with physical function 1 year after total knee arthroplasty in patients with knee osteoarthritis: a systematic review and meta-analysis. JAMA Netw Open 2022 Jul 1;5(7):e2219636. [doi: 10.1001/jamanetworkopen.2022.19636] [Medline: 35816307]
36. Hofstede SN, Gademan MGJ, Vliet Vlieland TPM, Nelissen R, Marang-van de Mheen PJ. Preoperative predictors for outcomes after total hip replacement in patients with osteoarthritis: a systematic review. BMC Musculoskelet Disord 2016 May 17;17:212. [doi: 10.1186/s12891-016-1070-3] [Medline: 27184266]
37. Kennedy DM, Hanna SE, Stratford PW, Wessel J, Gollish JD. Preoperative function and gender predict pattern of functional recovery after hip and knee arthroplasty. J Arthroplasty 2006 Jun;21(4):559-566. [doi: 10.1016/j.arth.2005.07.010] [Medline: 16781410]
38. Kluge F, Hannink J, Pasluosta C, et al. Pre-operative sensor-based gait parameters predict functional outcome after total knee arthroplasty. Gait Posture 2018 Oct;66:194-200. [doi: 10.1016/j.gaitpost.2018.08.026] [Medline: 30199778]
39. Devasenapathy N, Maddison R, Malhotra R, Zodepy S, Sharma S, Belavy DL. Preoperative quadriceps muscle strength and functional ability predict performance-based outcomes 6 months after total knee arthroplasty: a systematic review. Phys Ther 2019 Jan 1;99(1):46-61. [doi: 10.1093/ptj/pzy118] [Medline: 30329137]
40. Hewlett-Smith N, Pope R, Furness J, Simas V, Hing W. Prognostic factors for inpatient functional recovery following total hip and knee arthroplasty: a systematic review. Acta Orthop 2020 Jun;91(3):313-318. [doi: 10.1080/17453674.2020.1744852] [Medline: 32237927]
41. Casey HL, Shah VV, Muzyka D, et al. Gait characteristics in people with Friedreich ataxia: daily life versus clinic measures. Front Neurol 2025;16:1544453. [doi: 10.3389/fneur.2025.1544453] [Medline: 40166640]

Abbreviations

EMM: estimated marginal mean
IMU: inertial measurement unit
PROM: patient-reported outcome measure
THA: total hip arthroplasty
TJA: total joint arthroplasty
TKA: total knee arthroplasty

XSL•FO
RenderX

Edited by J Zhou; submitted 08.Apr.2026; peer-reviewed by HL Lu, R Boekesteijn, T Plavoukou; revised version received 24.Sep.2026; accepted 07.Oct.2026; published 30.Oct.2026.

Please cite as:
Teitz LS, Granot-Hershkovitz E, Lee PA, Lo OY, Manor B, Naveh YY
Real-World Functional Gait Recovery Trajectories After Total Hip and Total Knee Arthroplasty via Smartphone Sensors: Observational Longitudinal Study
JMIR Mhealth Uhealth 2026;14:e97657
URL: https://mhealth.jmir.org/2026/1/e97657
doi:10.2196/97657

© Levi S Teitz, Einat Granot-Hershkovitz, Pei-An Lee, On-Yee Lo, Brad Manor, Yuval Y Naveh. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 30.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Patient Engagement in a Virtual Cardiac Rehabilitation Platform: Secondary Analysis of the ERA-Nudge Trial

Prerna Gupta[1], MD; Josh Durfee[2], MSPH; Joy Waughtal[3], MPH; Dana R Fletcher[4], PhD; Vincenzo B Polsinelli[5], MD; Harsh Vathsangam[6], MS; Joan F Hilton[1], MPH, SCD; Alexis L Beatty[1], MAS, MD; P Michael Ho[5], MD, PhD; Larry A Allen[5], MHS, MD; Sheana S Bull[3], MPH, PhD; Pamela N Peterson[2,5], MSPH, MD

[1]Division of Cardiology, Department of Medicine, University of California, San Francisco, 555 Mission Bay Boulevard S, San Francisco, CA, United States

[2]Center for Health Systems Research, Denver Health, Denver, CO, United States

[3]mHealth Impact Lab, Department of Community and Behavioral Health, Colorado School of Public Health, Aurora, CO, United States

[4]Evida Research, Denver, CO, United States

[5]Division of Cardiology, Department of Medicine, University of Colorado Anschutz, Aurora, CO, United States

[6]Movn Health, Irvine, CA, United States

Corresponding Author:
Prerna Gupta, MD
Division of Cardiology, Department of Medicine, University of California, San Francisco, 555 Mission Bay Boulevard S, San Francisco, CA, United States

Abstract

Background: Mobile health interventions have the potential to motivate healthy behaviors. Optimal strategies for engaging patients in cardiac rehabilitation (CR) through mobile health interventions remain unclear.

Objective: This study aimed to examine patient engagement in a mobile app platform for CR across patient characteristics and by the method of delivery of nudge messages over the 12 weeks of CR.

Methods: The authors analyzed data from the ERA-Nudge (Enhancing Rehabilitation Adherence Through Optimized Nudge) study, a pragmatic randomized trial conducted at an urban safety-net hospital testing nudges (short messages designed using behavioral economics to elicit behavior change) to enhance CR adherence and log healthy activity. Trial patients were given access to a smartphone-based app. Nudges were initially sent as app notifications. Later in the study, patients received nudges as text messages. We assessed patient characteristics associated with app engagement, and we compared time spent in the CR app among patients receiving nudges as app notifications vs text messages. The trial enrolled patients who could engage with messages in either English or Spanish at an urban safety-net hospital with a qualifying CR diagnosis and a smartphone from May 2019 to May 2023. We describe minutes spent in the app across patient characteristics and by type of nudge delivery, and we compare the 2 approaches using unadjusted and multivariate negative binomial regression over the 12 weeks of CR.

Results: In total, 223 patients received nudges: 74 via app notification and 149 via text message. Overall, patients spent an average of 219 (SD 328) minutes in the app, and 3 (1.3%) patients used the app for zero minutes. Patients receiving only app nudges spent a mean of 167 (SD 365) minutes in the app, and patients receiving text messages and app nudges spent 245 (SD 306) minutes in the app (unadjusted ratio 1.5, 95% CI 1.1-2.0; *P*=.02). After adjusting for age category, sex, primary language, CR location (hospital vs home), and highest education achieved, Spanish-preferring patients, patients with some college education, and patients receiving text messages spent more time in the app. Patients receiving nudges as text messages and app notifications spent an adjusted ratio of 1.8 (95% CI 1.2-2.5; *P*=.002) times as many minutes in the app compared to those getting nudges as app notifications.

Conclusions: Among an urban safety-net population, text message nudges were associated with greater time in the app compared to nudges delivered as app notifications. These findings suggest that text messages may result in greater patient engagement for mobile health interventions.

Trial Registration: ClinicalTrials.gov NCT03834155; https://clinicaltrials.gov/study/NCT03834155

International Registered Report Identifier (IRRID): RR2-10.1016/j.ahj.2025.05.010

(JMIR Mhealth Uhealth 2026;14:e64502) doi:10.2196/64502

XSL•FO
RenderX

KEYWORDS

cardiac rehabilitation; mobile health; mHealth; patient engagement; mobile health intervention; healthy behaviors; application; randomized trial; mobile app

Introduction

Disparities in access to and availability of cardiovascular care persist. Burgeoning mobile health (mHealth) technologies have the potential to augment patient care by improving care team interaction and motivating healthy behavior change. An important use case is cardiac rehabilitation (CR). CR portends immense benefits for patients [1-6], but referral rates are approximately 30%, and participation rates are even lower at 8% to 9% [1,2,5,7]. Numerous barriers to CR attendance lower adherence rates, and augmentation of CR delivery via mobile tools such as telehealth CR or a mobile app may circumvent these obstacles [2,5]. There are numerous platforms in development, but clear benefits from these mobile or hybrid CR interventions have been inconsistent [8]. Approximately half of virtual CR alternatives report improved adherence, physical activity, or functional status. Reasons for the mixed results are underexplored, and patient use of the technologies is infrequently reported. Heterogeneity in patient use across studies may be a reason for inconsistent outcomes, but strategies to best engage patients remain unclear [2,9].

A possible explanation for inconsistent improvement in outcomes of virtual CR trials could be the lack of use by patients—like medications, patients must use interventions for them to work [8]. Some studies of hybrid or remote CR have resulted in fewer adherences compared to traditional CR, and the mechanisms of this suboptimal adherence are undefined. One way to engage patients may be to send nudge messages designed to create behavior change using principles of behavioral economics [10,11]. It is not yet known how these nudge messages should be delivered or how they may impact patient engagement.

To better understand how to engage patients, the primary objective of this analysis was to identify patient characteristics associated with patient engagement and compare patient engagement in the ERA-Nudge (Enhancing Rehabilitation Adherence Through Optimized Nudge) study of an app-based CR platform based on whether patients received nudges via both text messages and app notifications or app notifications alone, as assessed by the amount of time spent in the app. While most CR studies evaluate attendance at CR or exercise sessions by CR patients, we aimed to assess not the exercise minutes or sessions, but rather patients' engagement with mobile technology as measured by time spent in the app when patients have a virtual CR option. The ERA-Nudge study aims to improve referral and adherence to CR using short messages designed to elicit behavior change or nudges [10,11]. Nudges were initially sent as notifications within a mobile phone app, Movn (Movn Health), and later also sent as text messages. We identified patient characteristics associated with app engagement and evaluated the intervention factor of text messaging to determine if it was associated with app engagement independent of patient factors, with the hypothesis that text messages may be associated with greater engagement with the app.

Methods

Population

This is a secondary analysis of the ERA Nudge study (NCT03834155). The protocol has been previously published [12]. Briefly, this study enrolled patients with a qualifying CR diagnosis who speak English or Spanish and have access to a smartphone. Patients were identified daily from the hospital census based on a qualifying diagnosis by the CR team. The trial was a single-center study conducted at Denver Health, an urban safety-net hospital with over 60% of patients identifying as minorities, and it enrolled patients from May 2019 to May 2023. The app usage data were sourced from Movn Health, the company that delivers the app, and the remainder of the data was compiled from the electronic health record and collected trial data. For the purposes of our study, we evaluated the time spent in the app by patients receiving nudges as either app notifications or text messages.

Ethical Considerations

All eligible patients provided consent prior to enrollment and randomization. The trial protocol and analysis plan were approved by the Denver Health and Colorado Multiple Institutional Review Board (18 - 1290). Data were deidentified prior to analysis. Compensation was provided to participants for completing surveys. Participants received a US $25 gift card for completing each study assessment.

Movn App

As part of the trial, all enrolled patients were given access to the Movn app on their smartphones. This platform facilitated chat-based communication between the CR team and patients, provided educational resources, and enabled patients to log exercise type and duration. The app also allowed for patient data entry of weight and vital signs such as heart rate and blood pressure. The app was intended for data logging, chat, and the provision of educational resources. The patients' time spent in the app may not have included the time they spent exercising, as these data were logged after an exercise session. These data were not required, and patients were not provided with exercise equipment, but they were able to enter this information if they had equipment. Movn Health collected data on how many minutes each patient spent in the app from the program start date to the graduation date. All patients were trained in app use upon enrollment in the trial by study staff.

Trial Design

The ERA-Nudge study consisted of a 2×2 factorial design, randomizing patients to (1) choice of hospital or home-based CR vs hospital-based CR and (2) optimized nudges vs no reminders. We developed nudges through iterative focus group sessions and designed them to encourage healthy behavior aligned with general messages around CR. The messages encouraged healthy behaviors but did not specifically include information on how to use the app. They were sent once or twice

XSL•FO
RenderX

per week, and only to patients randomized to receive nudges. Nudges were initially sent as app notifications and into the patient's inbox within the app. However, by July 2021, patients expressed to study staff that they were not seeing the nudges. In response to this feedback, study staff added text messages via a mobile messenger platform, providing the same nudges on the same schedule in both messaging formats, thereby enabling patients to view the messages as an app notification or a text message. Given that nudges were in the patients' app inboxes, the nudges were always available for viewing within the app, regardless of participants' app notification settings. Patients' settings regarding app notifications were not viewable or available to Movn or study staff. The nudge messages and schedule remained the same. We aim to evaluate whether the nudge delivery modality was associated with greater use or time spent in the app among patients receiving nudges—whether patients interacted with the app more when they received text message or app notification nudges. Apart from the switch from app notification nudges to text message notifications, all other methods were held constant, including app training protocols and trial enrollment. Based on feedback during the initial app notification nudges phase, no changes were made in the app training protocol for patients who received text message nudges. Given our study focuses on nudge message modality, our cohort comprises exclusively patients receiving nudges.

All patients referred to CR were screened for the trial. Eligible patients were then consented and enrolled by study staff. As part of the initial intake, patients also provided their educational status, employment, income, and Patient Health Questionnaire-9 (PHQ-9), an instrument used to assess depression. Clinical data, such as qualifying conditions, were collected by study staff based on the electronic health record.

Statistical Analysis

We evaluated baseline patient factors using the mean and SD for continuous variables and proportions for categorical variables. Demographics, clinical indication for CR, and CR location were compared among subgroups using chi-square tests or Fisher exact tests, where appropriate. We performed unadjusted comparisons of each covariate with the outcome variable, minutes in the app, as well as an adjusted analysis accounting for all patient factors. We used a negative binomial regression with a log-link function to account for the zero-weighted, skewed data in the full cohort. The final multivariable model was determined based on a prespecified approach incorporating age category, sex, CR location (home vs hospital), and the primary exposure variable: text message and app notification nudges vs app notification nudges alone. Educational status, employment, and income demonstrated significant collinearity. For the determination of the final multivariable model, multivariable models incorporating educational status, employment, or income were compared using information criteria to arrive at the final multivariable model. Interaction terms were evaluated, but none were significant. The resulting ratios identify the ratio of minutes in the app between the compared groups. Results are reported according to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist (Checklist 1). All analyses were performed using SAS Studio (version 3.8.1; SAS Institute).

Results

Baseline Characteristics

Overall, in the ERA-Nudge study, 447 patients were enrolled and randomized (224 to nudges and 223 to no nudges). Among these, 1 patient withdrew during the baseline visit, resulting in a cohort of 223 patients for this analysis (Figure 1). Among our cohort of patients for this analysis receiving nudges, 74 patients received nudges as app notifications only, and 149 patients received both text message and app notification nudges.

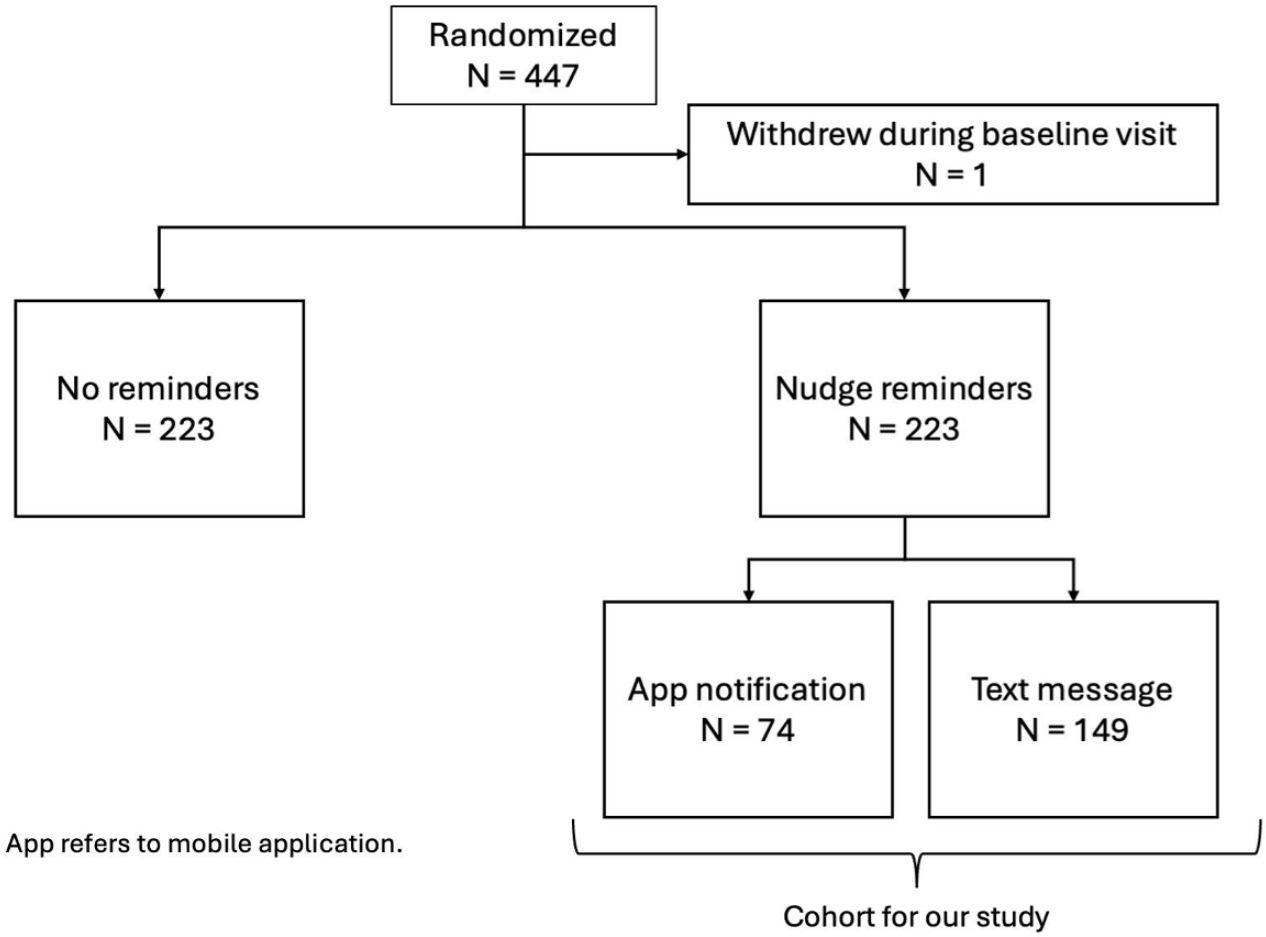
Our cohort includes 167 (75%) male and 56 (25%) female patients. The average age of the cohort was 55.5 (SD 11.5) years. Additionally, the trial cohort included 153 (69%) White patients, 36 (16%) Black patients, and 7 (3%) patients of other races. The cohort included 88 (40%) Hispanic patients and 55 (25%) patients who preferred Spanish over English. There were 75 (34%) patients who participated in CR at home. Of note, 95 (43%) patients had an income level of US $20,000 or less, and 99 (44%) patients in the cohort were employed at least part-time (Table 1). The most common qualifying condition in our cohort was percutaneous coronary intervention (n=96, 43% in the whole cohort). Other qualifying conditions included patients with acute coronary syndrome (n=87, 39%) and heart failure (HF; n=56, 25%). Data on intake demonstrated that 87 patients (52% of 167 total PHQ-9 tests gathered at baseline) had PHQ-9 scores greater than 4, consistent with at least mild depression.

Patients receiving text messages had a mean age of 54.7 (SD 12.4) years; 73% were male (109/149), 62% were White patients (93/149), and 36% received CR at home (53/149). Patients receiving app notifications had a mean age of 57.3 (SD 9.4) years, 78% (58/74) were male, 81% (60/74) were White patients, and 30% (53/74) received CR at home (Table 1).

Overall, the distribution of minutes spent in the app among the cohort was skewed toward nonuse (Figures 2 and 3). Overall, patients spent an average of 219 (SD 328) minutes in the app. There were 3 out of 223 (1.3%) patients who used the app for zero minutes.

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.



XSL•FO
RenderX

Table . Demographics, socioeconomic status, and clinical indications among study participants in the text message and app notification groups[a].

Characteristic	Overall (N=223)	Text message (n=149)	App notification (n=74)	*P* value
Age (y), mean (SD)	55.5 (11.5)	54.7 (12.4)	57.3 (9.4)	.09
Age tertile (y), n (%)				.13
20 - 51	72 (32)	54 (36)	18 (24)	
52 - 61	79 (35)	47 (32)	32 (43)	
62 - 85	72 (32)	48 (32)	24 (32)	
Sex, n (%)				.49
Male	167 (75)	109 (73)	58 (78)	
Female	56 (25)	40 (27)	16 (22)	
Race, n (%)				.03
White	153 (69)	93 (62)	60 (81)	
Black	36 (16)	28 (19)	8 (11)	
Other	7 (3)	5 (3)	2 (3)	
Unknown or not reported	27 (12)	23 (15)	4 (5)	
Hispanic, n (%)	88 (40)	59 (40)	29 (39)	.88
Spanish, n (%)	55 (25)	36 (24)	19 (26)	.87
Home-based CR[b], n (%)	75 (34)	53 (36)	22 (30)	.38
Level of education, n (%)				.002
Less than HS[c]	40 (18)	29 (19)	11 (15)	
HS diploma or GED[d]	67 (30)	44 (30)	23 (31)	
Some college or more	109 (49)	76 (51)	33 (45)	
Unknown	7 (3)	0 (0)	7 (9)	
Income level (US $), n (%)				.02
<20,000	95 (43)	65 (44)	30 (41)	
20,000-39,999	35 (16)	22 (15)	13 (18)	
40,000-59,999	19 (9)	14 (9)	5 (7)	
>60,000	28 (13)	21 (18)	7 (9)	
Unknown	11 (5)	0 (0)	11 (15)	
Prefer not to answer	35 (16)	27 (18)	8 (11)	
Employment, n (%)				<.001
Employed full-time	73 (34)	52 (35)	21 (28)	
Employed part-time	26 (12)	14 (9)	12 (16)	
Disabled, cannot work	31 (14)	19 (13)	12 (16)	
Not employed	52 (23)	45 (30)	7 (9)	
Retired	23 (10)	13 (9)	10 (14)	
Unknown	18 (8)	6 (4)	12 (16)	
CR indication, n (%)				
AMI[e]	87 (39)	53 (36)	34 (46)	.13
CABG[f]	12 (5)	6 (4)	6 (8)	.20
Valve replacement or repair	18 (8)	13 (9)	5 (7)	.61
PTCA[g] or stent	96 (43)	54 (36)	44 (59)	.001

XSL•FO
RenderX

Characteristic	Overall (N=223)	Text message (n=149)	App notification (n=74)	*P* value
Stable HF[h]	56 (25)	41 (28)	15 (20)	.24
Stable angina	5 (2)	5 (3)	0 (0)	.11
PHQ-9[i] score >4	87 (52)[j]	74 (51)	13 (18)	.34

[a]*T* tests are used to compare means across 2 groups. Chi-square tests are used to compare proportions across 2×2 groups, with Fisher tests for comparisons across tables larger than 2 groups.

[b]CR: cardiac rehabilitation.

[c]HS: high school.

[d]GED: general equivalency degree.

[e]AMI: acute myocardial infarction.

[f]CABG: coronary artery bypass graft.

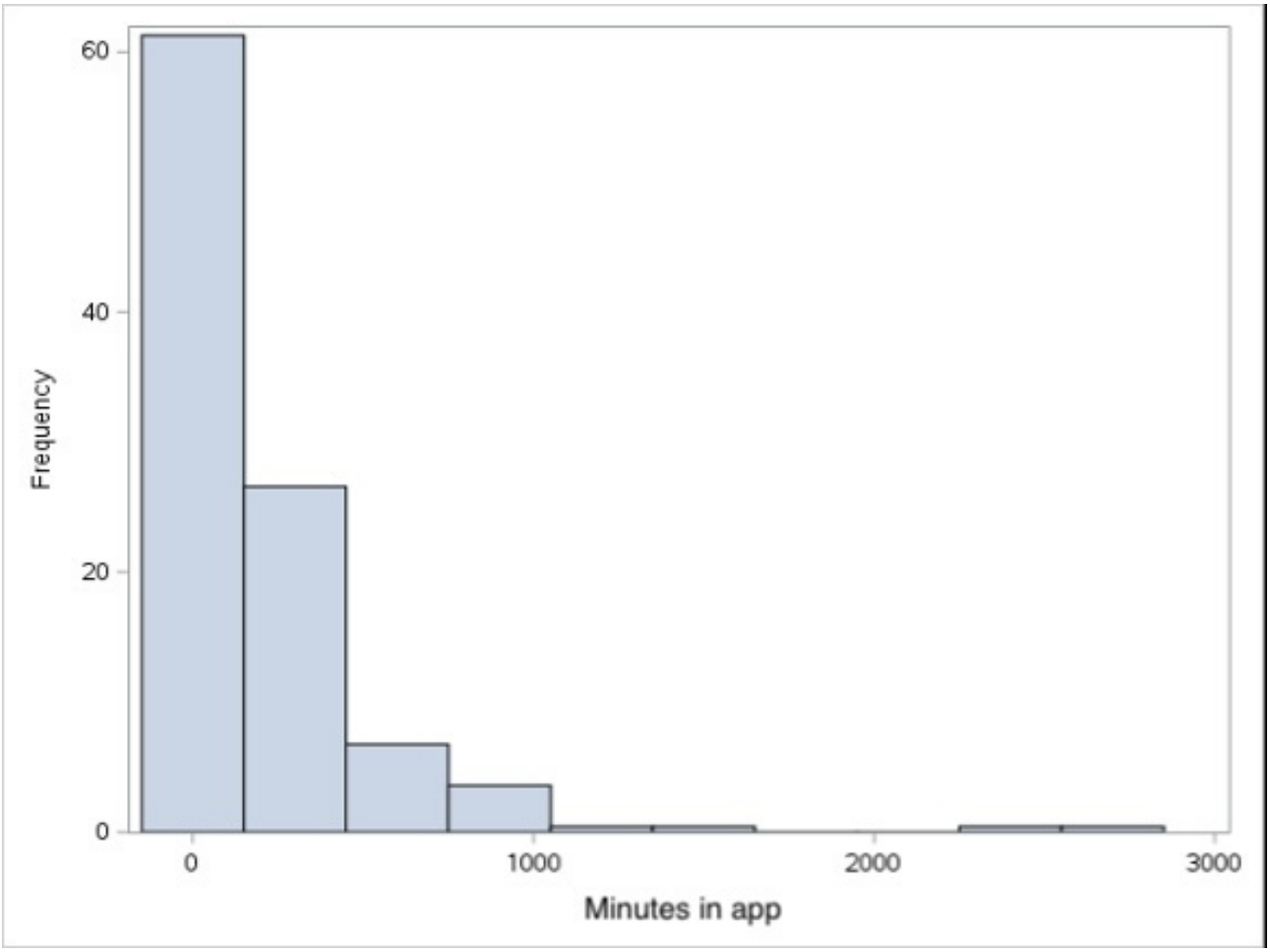
[g]PTCA: percutaneous transluminal coronary angioplasty.

[h]HF: heart failure.

[i]PHQ-9: Patient Health Questionnaire 9.

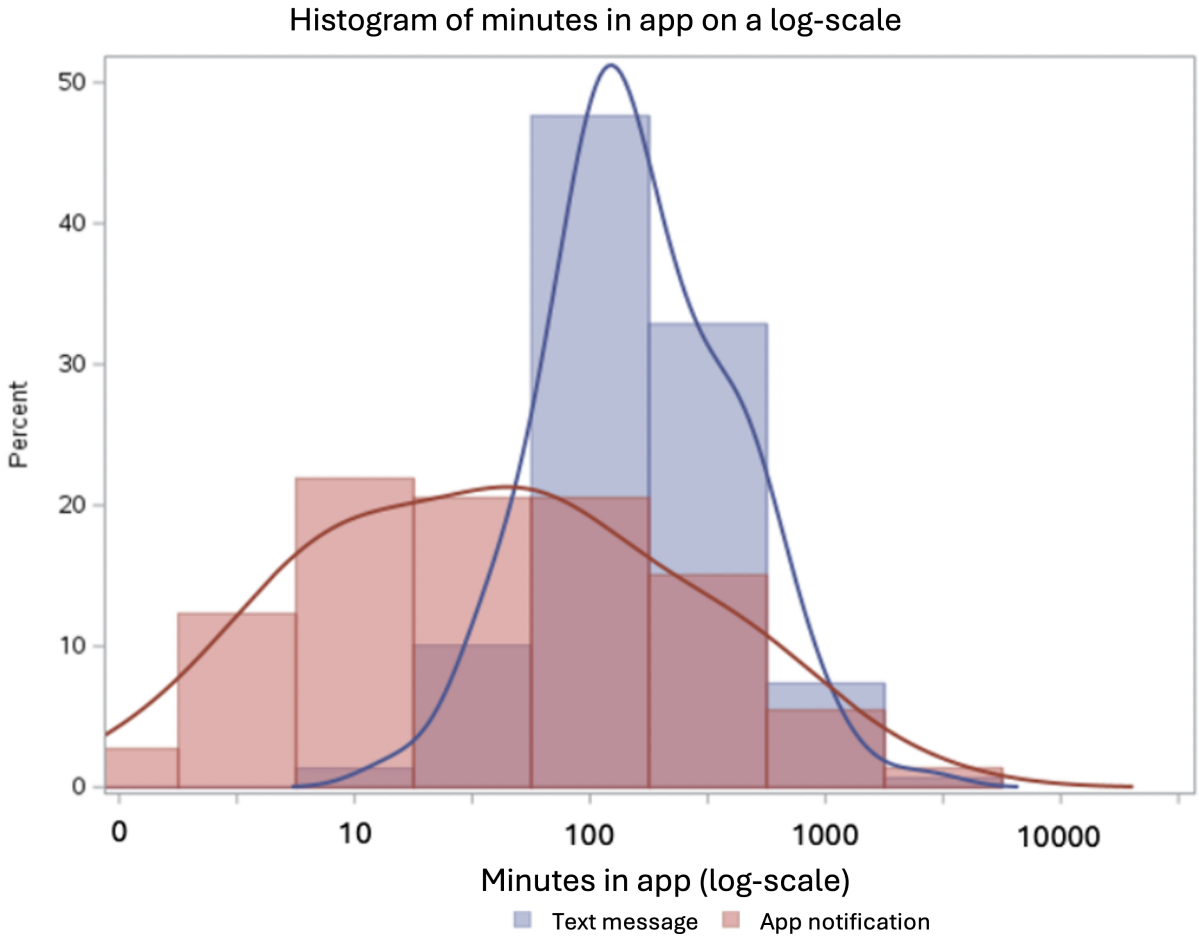
[j]Total PHQ-9 tests gathered at baseline (n=167).

Figure 2. Histogram of app usage in minutes.



XSL•FO
RenderX

Figure 3. Histogram of log-transformed minutes spent in the app. Red represents patients who received app notifications only, and blue represents patients who received both text messages and app notifications. Overall, patients who received text message nudges used the app more than those who received app notifications only.



Overall Predictors of Engagement

Unemployed participants spent 308 (95% CI 224-423) minutes in the app compared to 198 (95% CI 157-250) minutes spent by employed participants (unadjusted ratio 1.6, 95% CI 1.1-2.3; *P*=.03) (Table 2). Participants who had a qualifying diagnosis of myocardial infarction, non-ST–elevation myocardial infarction, or unstable angina spent 275 (95% CI 215-352) minutes in the app compared to those without (182, 95% CI 149-222 min; *P*=.01). Patients with a qualifying diagnosis of stable HF spent 142 (95% CI 104-193) minutes in the app compared to 244 (95% CI 204-291) minutes among those with other qualifying diagnoses (including but not limited to percutaneous coronary intervention or acute coronary syndrome; Figure 2 and Table 2).

XSL•FO
RenderX

Table . Minutes spent in the app by demographic subgroups.

Characteristic	Overall (min), mean (95% CI)[a]	App notification (min), mean (SD)[b]	Text message (min), mean (SD)[b]	Unadjusted ratio (95% CI)[c]	Unadjusted *P* value[c]	Adjusted ratio (95% CI)[d]	Adjusted *P* value[d]
Age tertile (y)							
20 - 51	202 (153-265)	197 (594)	203 (377)	Reference	Reference	Reference	Reference
52 - 61	215 (165-279)	130 (218)	272 (238)	1.1 (0.7-1.6)	.75	1.2 (0.8-1.7)	.39
62 - 85	239 (181-314)	188 (303)	264 (275)	1.2 (0.8-1.7)	.40	1.3 (0.9-1.9)	.22
Sex							
Male	205 (171-246)	178 (399)	220 (227)	Reference	Reference	Reference	Reference
Female	258 (189-351)	127 (194)	313 (454)	1.3 (0.9-1.8)	.21	1.0 (0.7-1.5)	.92
Race							
White	236 (195-284)	273 (362)	178 (398)	Reference	Reference	—[e]	—
Black	174 (118-256)	72 (125)	203 (188)	0.7 (0.5-1.1)	.17	—	—
Other	181 (75-434)	46 (8)	235 (114)	0.8 (0.3-1.9)	.56	—	—
Unknown	189 (121-295)	185 (159)	213 (162)	0.8 (0.5-1.3)	.37	—	—
Ethnicity[a]							
Hispanic or Latino	227 (177-291)	165 (231)	257 (392)	1.1 (0.8-1.5)	.74	—	—
Non-Hispanic or Latino	215 (175-263)	167 (435)	238 (235)	Reference	Reference	—	—
Primary language							
English	204 (171-245)	162 (395)	225 (233)	Reference	Reference	Reference	Reference
Spanish	261 (191-357)	174 (257)	307 (465)	1.3 (0.9-1.8)	.18	1.5 (1.02-2.29)	.04
Site of CR[f]							
Hospital-based	209 (173-253)	181 (423)	224 (318)	Reference	Reference	Reference	Reference
Home-based	237 (181-310)	128 (150)	282 (281)	1.1 (0.8-1.6)	.46	1.1 (0.8-1.5)	.77
Level of education							
Less than high school	172 (120-248)	93 (135)	202 (209)	Reference	Reference	Reference	Reference
High school diploma or GED[g]	171 (129-226)	55 (70)	231 (219)	1.0 (0.6-1.6)	.97	1.2 (0.7-1.9)	.49
Some college or more	259 (208-323)	237 (508)	269 (373)	1.5 (1.0-2.3)	.06	2.1 (1.3-3.4)	.003
Unknown	301 (126-718)	301 (273)	—	1.7 (0.7-4.5)	.25	3.0 (1.1-7.8)	.03
Employment							
Employed	198 (157-250)	162 (292.5)	216 (234.5)	Reference	Reference	—	—
Disabled, cannot work	148 (98-224)	62 (76.3)	202 (217.3)	0.7 (0.5-1.2)	.23	—	—
Not employed	308 (224-423)	396 (942.1)	294 (430.4)	1.6 (1.0-2.3)	.03	—	—
Retired	193 (119-311)	118 (182.1)	250 (235.4)	1.0 (0.6-1.7)	.91	—	—
Unknown or declined	224 (131-386)	182 (249.7)	310 (275.2)	1.1 (0.6-2.0)	.68	—	—
Income level (US $)							

XSL•FO
RenderX

Characteristic	Overall (min), mean (95% CI)[a]	App notification (min), mean (SD)[b]	Text message (min), mean (SD)[b]	Unadjusted ratio (95% CI)[c]	Unadjusted *P* value[c]	Adjusted ratio (95% CI)[d]	Adjusted *P* value[d]
<20,000	190 (150-241)	95 (135)	234 (238)	Reference	Reference	—	—
20,000-39,999	247 (167-364)	323 (706)	201 (128)	1.3 (0.8-2.0)	.26	—	—
40,000-59,999	122 (72-208)	7 (7.8)	164 (157)	0.6 (0.4-1.1)	.14	—	—
>60,000	278 (180-430)	282 (471)	277 (353)	1.5 (0.9-2.4)	.13	—	—
Declined to answer	258 (184-363)	167 (247)	323 (509)	1.4 (0.9-2.1)	.14	—	—
PHQ-9[h] score							
≤4	277 (225-341)	264 (291)	279 (388)	Reference	Reference	—	—
>4	236 (193-288)	397 (667)	208 (196)	0.9 (0.6-1.1)	.27	—	—
Total	219[i] (—)	167 (365)	245 (306)	1.5 (1.1-2.1)	.02	1.8 (1.3-2.6)	.001

[a]Reflects least-squared mean from a negative binomial regression across levels of the covariate, with a 95% CI for the mean.

[b]Reflects the arithmetic mean and SD of minutes among participants in that subgroup.

[c]Unadjusted ratio and *P* value reflect the unadjusted analysis of time in the app across levels of that covariate.

[d]Adjusted ratio and *P* value reflect results from the adjusted negative binomial model with a log-link function, accounting for age tertile, sex, preferred language, site of CR, level of education, and app notification vs text message nudges.

[e]Not applicable.

[f]CR: cardiac rehabilitation.

[g]GED: general equivalency degree.

[h]PHQ-9: Patient-Health Questionnaire-9.

[i]SD 328.

Unadjusted App Notification vs Text Message

Patients receiving nudges as text messages spent 245 (SD 306) minutes in the app compared to patients receiving app notifications, who spent 167 (SD 365) minutes in the app (unadjusted ratio of minutes in the app among patients receiving both text messages and app notifications vs app notifications only: 1.5, 95% CI 1.1-2.1; *P*=.02).

Adjusted Analyses

In an analysis adjusted for age category, sex, primary language, CR location, and highest education achieved, patients receiving text messages and app notification nudges spent an adjusted ratio of 1.8 (95% CI 1.3-2.6; *P*=.001) times as many minutes in the app compared to those receiving only app notifications (Table 2). The adjusted model also suggested that there was greater app use among patients who preferred Spanish (261 vs 204 min among those who preferred English; adjusted ratio 1.5, 95% CI 1.02-2.26; *P*=.04) and had at least some college-level education compared to participants with less than a high school education (262 vs 172 min; adjusted ratio 2.1, 95% CI 1.3-3.3; *P*=.002).

Discussion

Principal Findings

This analysis of a pragmatic study of app-based remote CR at an urban safety-net hospital aimed to identify correlates of app use and characterize app use among patients receiving nudges as text messages and app notifications or app notification nudges only. Our analysis of 223 patients demonstrated that app use was heterogeneous and heavily skewed toward low use. In an adjusted model, we found that increased use was associated with higher education and Spanish language–preferring patients. Most importantly, nudges sent as text messages and app notifications were associated with greater app use compared to nudges sent as app notifications in both adjusted and unadjusted models.

The results of this study suggest that patients who prefer Spanish and those with higher education levels spent more time in the app. In contrast to prior work, we did not find that older patients engaged less with the intervention [13-15]. Other studies have previously identified ethnicity and education level as important influences on patient engagement in chronic disease management [14]. It is important to note that most trials underrepresent the target population of this study, as most trials enroll participants from major university centers [16-18]. While this single-center study is affiliated with a major university, the study was entirely conducted at an urban safety-net hospital, suggesting that with adequate technology training, these adversely resourced patients may be interested in and engaged with the resources available to them. In the adjusted multivariate model, patients with higher levels of education spent more time in the app. Several prior studies have identified the level of education as an important mediator of socioeconomic differences in cardiovascular disease and CR participation, and our study replicated this finding among patients participating in virtual or hybrid CR [19,20]. In addition, the finding that Spanish-preferring patients spent more time in the app could suggest 2 disparate possibilities: that patients need more time

XSL•FO
RenderX

to navigate the interface, or they are spending more time in the app to learn more about their condition. There is little prior work in this area, and this warrants further investigation.

While most studies in this realm report changes in adherence, physical activity, or quality of life, few report the use of the mHealth platform itself [8,21]. Only occasional reports exist regarding the duration of wearables worn [22], time spent in mobile apps, or web logins [23,24] for the variety of remote CR platforms tested thus far. These smaller studies demonstrate high usability and acceptability, with use metrics ranging from 20 to 120 patients and 30 days to 24 weeks [9,25]. Other digital health interventions aimed at improving patient health found baseline technology use and intent for lifestyle change to be associated with patient engagement. Multiple studies have also found dose-response effects, with higher use among patients associated with greater improvement in clinical parameters [26-30]. A multicomponent health promotion intervention to promote weight loss identified that phone calls were deemed helpful by participants, but website and app usage were low [31]. Some studies suggest self-segregating cohorts of users, including high users, high users who decline, and low users [29,30]. This suggests significant variation over time, a phenomenon that should be studied further to identify strategies to keep patients engaged. Needs assessments and surveys with engaged patients have identified personalization as a suggested strategy to maintain engagement [32]. Some qualitative studies report nonadopters among an aging population, with "forgotten passwords" and "outdated technology" reported as obstacles to use, but these studies represent a limited subset of perspectives [33,34]. A mixed methods study demonstrated low adherence across all demographic subgroups and identified fear of movement and exercise self-efficacy as major obstacles to a home CR intervention among patients with HF [13]. Significant heterogeneity is demonstrated in these studies, suggesting that much work needs to be done to understand how to best augment adherence to home-based or virtual CR alternatives.

It is crucial to study strategies that may augment use or adherence among patients to virtual alternatives, given that greater use may result in better clinical outcomes. One CR study compared changes in activity among patients receiving text message reminders for activity, finding variation in steps walked based on the device used for measurement and the number of days since study enrollment [35]. However, the majority of studies investigating strategies to augment patient engagement in mHealth are outside of CR. In a review of interventions to augment engagement with mHealth, strategies included incorporating behavior change theory, reminders, goal-setting, automatic enrollment, and provider-led strategies [26,36]. A greater number of behavior change theories incorporated may have led to greater behavior change, but few studies referenced theories, and the effects were no different from the overall studies in the review. Provider-facilitated engagement led to the greatest engagement, but no reasons were provided for this or for the high attrition with other engagement strategies. Another study identified that opt-out consent led to greater participation in a mHealth intervention to improve medication adherence among groups that have been historically underrepresented [16]. We found a higher association of app use among patients receiving text message nudges, suggesting that this intervention factor may modify use among patients. While our analysis cannot determine causality, this is hypothesis-generating regarding strategies to better engage patients in mHealth interventions. Given the lack of standard measurements, and variations in follow-up periods, it remains difficult to compare use among studies or strategies. Without examining or comparing use, it will remain unclear which patients require additional support. Mobile interventions require careful implementation and theoretical bases to ensure that we are able to reach all populations, given ongoing concerns regarding widening disparities with the increasing advent of digital health interventions [16,17,21,37-39]. To deliver CR as a mobile or remote option, quality cannot be compromised, and patients must be supported and engaged [1]. Ongoing work is necessary to determine how to best engage patients.

Limitations

Limitations of this study include that it was a single-center study that excluded patients who did not speak English or Spanish. The nonrandomized, pre-post design used in the analysis additionally precludes the determination of causation. Given the temporal differences in study arms, there are several factors that this study design could not account for, such as improved training protocols, study staff experience, or seasonality. We conclude that nudge delivery modality influences patient engagement, with text messaging being associated with greater engagement than app notifications. Furthermore, this is an intermediary assessment. The true value of increased time in the app will be determined through an analysis of patient outcomes. Strengths of this study include the 223 patients, which is a larger sample size than many studies exploring similar questions.

Conclusions

We conclude that an app-based CR platform at an urban safety-net hospital was feasible and acceptable. Spanish language–preferring patients and those with higher educational status demonstrated greater use of the app. Most importantly, nudges delivered as text messages and app notifications were associated with greater use of the app platform compared to patients who received nudges as app notifications alone. It remains to be seen how best to support and engage patients who demonstrated lower app use, and whether this will impact outcomes within the study. As we transition to incorporating hybrid or virtual CR models into care, it is crucial that we identify strategies to better engage patients who have been historically excluded. These findings inform future interventions, which may consider text messages to engage patients more effectively than app notifications. Further research remains necessary to evaluate how to support patients who engage less with the app, such as those with lower educational status, and to assess whether app use correlates with patient outcomes such as CR adherence or functional capacity.

XSL•FO
RenderX

Acknowledgments

The authors acknowledge Monica Martinez for her work on the overall study and her contributions.

Generative AI Use

Generative AI was not used in any portion of this manuscript.

Funding

PNP has received research funding from the NHLBI (5R33HL143324-05). PG has received research funding from the National Institutes of Health (5T32HL007822-24) at the time of the writing of this work.

Data Availability

Data can be made available on reasonable request.

Authors' Contributions

Conceptualization: PG, PNP, JD, VBP, PJMH, LAA
Data curation: PG, JD
Formal analysis: PG, JD
Funding acquisition: PNP, SSB
Methodology: PG, JD, JFH
Visualization: PG
Writing – original draft: PG
Writing – review & editing: PG, JW, DRF, VBP, JFH, ALB, PJMH, LAA, SSB, PNP

Conflicts of Interest

HV is the chief executive officer of the company Movn Health, previously known as Movn Analytics. While the company provided the software platform, all the funding for this study was from the National Institutes of Health. DRF is a consultant and the owner of Evida Research Consulting, Inc. SSB is the President and a Co-founder of Clinic Chat, LLC, the company that is subcontracted to design the chatbot described in this paper. All other authors declare no conflicts of interest.

Checklist 1
STROBE checklist.
[DOCX File, 34 KB - mhealth_v14i1e64502_app1.docx]

References

1. Beatty AL, Truong M, Schopfer DW, Shen H, Bachmann JM, Whooley MA. Geographic variation in cardiac rehabilitation participation in Medicare and Veterans Affairs populations: opportunity for improvement. Circulation 2018 May 1;137(18):1899-1908. [doi: 10.1161/CIRCULATIONAHA.117.029471] [Medline: 29305529]
2. Beatty AL, Beckie TM, Dodson J, et al. A new era in cardiac rehabilitation delivery: research gaps, questions, strategies, and priorities. Circulation 2023 Jan 17;147(3):254-266. [doi: 10.1161/CIRCULATIONAHA.122.061046] [Medline: 36649394]
3. Lawton JS, Tamis-Holland JE, Bangalore S, et al. 2021 ACC/AHA/SCAI guideline for coronary artery revascularization: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. J Am Coll Cardiol 2022 Jan;79(2):e21-e129. [doi: 10.1016/j.jacc.2021.09.006] [Medline: 34895950]
4. Heidenreich PA, Bozkurt B, Aguilar D. Correction to: 2022 AHA/ACC/HFSA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 2022 May 3;145(18):e895-e1032. [doi: 10.1161/CIR.0000000000001073]
5. Thomas RJ. Cardiac rehabilitation - challenges, advances, and the road ahead. N Engl J Med 2024 Feb 29;390(9):830-841. [doi: 10.1056/NEJMra2302291] [Medline: 38416431]
6. Dibben GO, Faulkner J, Oldridge N, et al. Exercise-based cardiac rehabilitation for coronary heart disease: a meta-analysis. Eur Heart J 2023 Feb 7;44(6):452-469. [doi: 10.1093/eurheartj/ehac747] [Medline: 36746187]
7. Aragam KG, Moscucci M, Smith DE, et al. Trends and disparities in referral to cardiac rehabilitation after percutaneous coronary intervention. Am Heart J 2011 Mar;161(3):544-551. [doi: 10.1016/j.ahj.2010.11.016] [Medline: 21392610]
8. Lee KCS, Breznen B, Ukhova A, Koehler F, Martin SS. Virtual healthcare solutions for cardiac rehabilitation: a literature review. Eur Heart J Digit Health 2023 Mar;4(2):99-111. [doi: 10.1093/ehjdh/ztad005] [Medline: 36974268]
9. Hamilton S, Mills B, McRae S, Thompson S. Evidence to service gap: cardiac rehabilitation and secondary prevention in rural and remote Western Australia. BMC Health Serv Res 2018 Jan 30;18(1):64. [doi: 10.1186/s12913-018-2873-8] [Medline: 29382343]

XSL•FO
RenderX

10. Kahneman D. Thinking, Fast and Slow: Farrar, Straus and Giroux; 2011.
11. Thaler RH, Sunstein CR. Nudge: Improving Decisions About Health, Wealth, and Happiness: Yale University Press; 2008.
12. Gupta P, Fletcher DR, Durfee J, et al. Enhancing cardiac rehabilitation adherence through home-based rehabilitation and behavioral nudges: the ERA nudge study design and rationale. Am Heart J 2025 Dec;290:17-28. [doi: 10.1016/j.ahj.2025.05.010] [Medline: 40383506]
13. Yang Z, Jia H, Wang A. Predictors of home-based cardiac rehabilitation exercise adherence among patients with chronic heart failure: a theory-driven cross-sectional study. BMC Nurs 2023 Nov 6;22(1):415. [doi: 10.1186/s12912-023-01566-5] [Medline: 37926820]
14. Aboumatar H, Pitts S, Sharma R, et al. Patient engagement strategies for adults with chronic conditions: an evidence map. Syst Rev 2022 Mar 5;11(1):39. [doi: 10.1186/s13643-021-01873-5] [Medline: 35248149]
15. Tisdale RL, Ferguson J, Van Campen J, et al. Disparities in virtual cardiology visits among Veterans Health Administration patients during the COVID-19 pandemic. JAMIA Open 2022 Dec;5(4):ooac103. [doi: 10.1093/jamiaopen/ooac103] [Medline: 36531138]
16. Gupta P, Sandy LC, Glorioso TJ, et al. Secondary analysis of electronic opt-out consent in pragmatic research: a study design method to diversify clinical trials? Am Heart J 2023 Jul;261:104-108. [doi: 10.1016/j.ahj.2023.03.010] [Medline: 36966921]
17. Johnson T, Chilazi M, Isakadze N, et al. Bridging the digital divide: applying digital tools to improve cardiovascular health in underrepresented communities. JACC Adv 2023 Sep;2(7):100587. [doi: 10.1016/j.jacadv.2023.100587] [Medline: 38939478]
18. Masoudi FA, Havranek EP, Wolfe P, et al. Most hospitalized older persons do not meet the enrollment criteria for clinical trials in heart failure. Am Heart J 2003 Aug;146(2):250-257. [doi: 10.1016/S0002-8703(03)00189-3] [Medline: 12891192]
19. Marmot MG, Shipley MJ, Hemingway H, Head J, Brunner EJ. Biological and behavioural explanations of social inequalities in coronary heart disease: the Whitehall II study. Diabetologia 2008 Nov;51(11):1980-1988. [doi: 10.1007/s00125-008-1144-3] [Medline: 18777168]
20. Pedersen M, Overgaard D, Andersen I, Baastrup M, Egerod I. Experience of exclusion: a framework analysis of socioeconomic factors affecting cardiac rehabilitation participation among patients with acute coronary syndrome. Eur J Cardiovasc Nurs 2017 Dec;16(8):715-723. [doi: 10.1177/1474515117711590] [Medline: 28513199]
21. Gupta P, Vaduganathan M. Fulfilling the promise of telemonitoring in heart failure. J Card Fail 2023 Dec;29(12):1655-1656. [doi: 10.1016/j.cardfail.2023.06.018] [Medline: 37442220]
22. Hammond-Haley M, Allen C, Han J, Patterson T, Marber M, Redwood S. Utility of wearable physical activity monitors in cardiovascular disease: a systematic review of 11 464 patients and recommendations for optimal use. Eur Heart J Digit Health 2021 Jun;2(2):231-243. [doi: 10.1093/ehjdh/ztab035] [Medline: 36712392]
23. Devi R, Powell J, Singh S. A web-based program improves physical activity outcomes in a primary care angina population: randomized controlled trial. J Med Internet Res 2014 Sep 12;16(9):e186. [doi: 10.2196/jmir.3340] [Medline: 25217464]
24. Su JJ. Exploring the relationship between usage and outcomes of an eHealth cardiac rehabilitation intervention: subanalysis of a randomized controlled trial. Comput Inform Nurs 2023 Aug 1;41(8):578-585. [doi: 10.1097/CIN.0000000000000980] [Medline: 36730738]
25. Ramachandran HJ, Jiang Y, Teo JYC, Yeo TJ, Wang W. Technology acceptance of home-based cardiac telerehabilitation programs in patients with coronary heart disease: systematic scoping review. J Med Internet Res 2022 Jan 7;24(1):e34657. [doi: 10.2196/34657] [Medline: 34994711]
26. Winter N, Russell L, Ugalde A, White V, Livingston P. Engagement strategies to improve adherence and retention in web-based mindfulness programs: systematic review. J Med Internet Res 2022 Jan 12;24(1):e30026. [doi: 10.2196/30026] [Medline: 35019851]
27. Broderick A, Haque F. Mobile health and patient engagement in the safety net: a survey of community health centers and clinics. Issue Brief (Commonw Fund) 2015 May;9:1-9. [Medline: 26040018]
28. Coley N, Andre L, Hoevenaar-Blom MP, et al. Factors predicting engagement of older adults with a coach-supported eHealth intervention promoting lifestyle change and associations between engagement and changes in cardiovascular and dementia risk: secondary analysis of an 18-month multinational randomized controlled trial. J Med Internet Res 2022 May 9;24(5):e32006. [doi: 10.2196/32006] [Medline: 35385395]
29. Li Y, Luk TT, Cheung YTD, et al. Engagement with a mobile chat–based intervention for smoking cessation: a secondary analysis of a randomized clinical trial. JAMA Netw Open 2024 Jun 3;7(6):e2417796. [doi: 10.1001/jamanetworkopen.2024.17796] [Medline: 38922618]
30. Skolarus LE, Lin CC, Mishra S, et al. Engagement in mHealth-prompted self-measured blood pressure monitoring among participants recruited from a safety-net emergency department: secondary analysis of the reach out trial. JMIR Mhealth Uhealth 2024 Jun 12;12:e54946. [doi: 10.2196/54946] [Medline: 38889070]
31. Partridge SR, Allman-Farinelli M, McGeechan K, et al. Process evaluation of TXT2BFiT: a multi-component mHealth randomised controlled trial to prevent weight gain in young adults. Int J Behav Nutr Phys Act 2016 Jan 19;13(1):7. [doi: 10.1186/s12966-016-0329-2] [Medline: 26785637]

XSL•FO
RenderX

32. Pfaeffli Dale L, Whittaker R, Dixon R, et al. Acceptability of a mobile health exercise–based cardiac rehabilitation intervention: a randomized trial. J Cardiopulm Rehabil Prev 2015;35(5):312-319. [doi: 10.1097/HCR.0000000000000125] [Medline: 26181037]
33. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629. [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
34. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367. [doi: 10.2196/jmir.8775] [Medline: 29092808]
35. Golbus JR, Shi J, Gupta K, et al. Text messages to promote physical activity in patients with cardiovascular disease: a micro-randomized trial of a just-in-time adaptive intervention. Circ Cardiovasc Qual Outcomes 2024 Jul;17(7):e010731. [doi: 10.1161/CIRCOUTCOMES.123.010731] [Medline: 38887953]
36. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046. [doi: 10.2196/17046] [Medline: 32186518]
37. Merid B, Robles MC, Nallamothu BK. Digital redlining and cardiovascular innovation. Circulation 2021 Sep 21;144(12):913-915. [doi: 10.1161/CIRCULATIONAHA.121.056532] [Medline: 34543062]
38. Eyrich NW, Andino JJ, Fessell DP. Bridging the digital divide to avoid leaving the most vulnerable behind. JAMA Surg 2021 Aug 1;156(8):703-704. [doi: 10.1001/jamasurg.2021.1143] [Medline: 33769441]
39. Gallegos-Rejas VM, Thomas EE, Kelly JT, Smith AC. A multi-stakeholder approach is needed to reduce the digital divide and encourage equitable access to telehealth. J Telemed Telecare 2023 Jan;29(1):73-78. [doi: 10.1177/1357633X221107995] [Medline: 35733379]

Abbreviations

CR: cardiac rehabilitation
ERA-Nudge: Enhancing Rehabilitation Adherence Through Optimized Nudge
HF: heart failure
mHealth: mobile health
PHQ-9: Patient Health Questionnaire-9
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by L Buis; submitted 09.Aug.2024; peer-reviewed by A Goren, R Marshall; revised version received 07.May.2026; accepted 07.May.2026; published 06.Nov.2026.

Please cite as:
Gupta P, Durfee J, Waughtal J, Fletcher DR, Polsinelli VB, Vathsangam H, Hilton JF, Beatty AL, Ho PM, Allen LA, Bull SS, Peterson PN
Patient Engagement in a Virtual Cardiac Rehabilitation Platform: Secondary Analysis of the ERA-Nudge Trial
JMIR Mhealth Uhealth 2026;14:e64502
URL: https://mhealth.jmir.org/2026/1/e64502
doi:10.2196/64502

© Prerna Gupta, Josh Durfee, Joy Waughtal, Dana R Fletcher, Vincenzo B Polsinelli, Harsh Vathsangam, Joan F Hilton, Alexis L Beatty, P Michael Ho, Larry A Allen, Sheana S Bull, Pamela N Peterson. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 6.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Efficiency and Cost-Effectiveness of Wearable Sensors in a Digital Physiotherapeutic Total Hip Arthroplasty–Specific Training System for Patients After Total Hip Arthroplasty: Randomized Controlled Trial

Chenyi Jiang[1], MD; Yun Shen[2], MD; Lihua Huang[3], MSc; Yanhong Ma[3], PhD; Shengdi Lu[2], MD; Jimin Yin[1], MD

[1]Department of Orthopaedics, Shanghai Sixth People's Hospital, 600 Yishan Road, Shanghai, China

[2]Pennington Biomedical Research Center, 6400 Perkins Rd, Baton Rouge, LA, United States

[3]Department of Rehabilitation Medicine, Shanghai Sixth People's Hospital, Shanghai, China

Corresponding Author:
Jimin Yin, MD
Department of Orthopaedics, Shanghai Sixth People's Hospital, 600 Yishan Road, Shanghai, China

Abstract

Background: With total hip arthroplasty (THA) volumes rising worldwide, scalable home rehabilitation strategies are needed. Home-based digital physiotherapy can improve recovery after THA, and adding wearable motion sensor feedback may further enhance exercise performance and adherence. However, the impact of sensor-augmented digital rehabilitation on patient outcomes and cost-effectiveness remains unclear.

Objective: This study aimed to evaluate the clinical effectiveness (including adherence) and cost-effectiveness of adding wearable sensor feedback to a home-based digital THA rehabilitation program compared with the same digital program without sensors.

Methods: We conducted a single-center randomized controlled trial in Shanghai from June 2023 to June 2024. A total of 240 patients who had undergone primary THA were randomized (1:1) to a 12-week home exercise program delivered via a mobile app either with wearable motion sensors for real-time feedback (intervention) or without sensors (control). Both groups received identical exercise content and weekly teleconsultations. Outcomes were assessed at 6, 12, and 24 weeks by blinded evaluators. The primary outcome was the Hip Disability and Osteoarthritis Outcome Score (HOOS) at 24 weeks. Secondary outcomes included HOOS subscales, timed up-and-go (TUG), Berg Balance Scale, 36-item Short Form Health Survey (SF-36) physical and mental scores, Hospital Anxiety and Depression Scale (HADS) anxiety/depression, patient satisfaction, adherence metrics, and total 24-week costs. Intention-to-treat analyses were used. Mixed effects models and χ^2 tests were used for group comparisons. Cost-effectiveness was evaluated from a societal perspective.

Results: The sensor-based rehabilitation group showed significantly greater improvement in the primary outcome (HOOS overall) than the control group at 24 weeks (P=.01). Early postoperative gains were larger: At 6 weeks, 4 of the 5 HOOS subscales, TUG time, and all patient-reported outcomes remained significant after Bonferroni correction for 47 secondary comparisons (adjusted P<.001). By 24 weeks, hip-specific functional differences had narrowed, but SF-36 physical and mental scores (both P<.001) and HADS anxiety and depression scores (both P<.001) remained significant after correction. Total 24-week costs were similar between the groups (¥87,967 vs ¥94,396 [US $12,879.80 vs $13,821.10] per patient; P=.32). The sensor intervention was associated with numerically lower costs on average (¥6428.98 [US $941.31] less per patient), yielding negative incremental cost-effectiveness ratios, though the cost difference was not statistically significant. Adherence was high in both groups but higher with sensors: The intervention group completed more exercise sessions (P=.002) and showed greater participation in weekly assessments and follow-up calls (both P<.001). Adverse events were uncommon in both groups (5.8% vs 7.5%) and mostly minor, with no serious events reported.

Conclusions: Augmenting home-based digital rehabilitation with wearable sensor feedback led to statistically significant improvements in the primary hip function outcome and in patient-reported quality of life and psychological outcomes that persisted after Bonferroni correction through 24 weeks. Several early hip-specific secondary outcome differences did not survive correction for multiple comparisons. The between-group HOOS differences were below the established minimal clinically important difference. Exercise adherence was significantly higher in the sensor group. These benefits were achieved with no increase in cost, suggesting the sensor-enhanced program is a safe and feasible approach to THA rehabilitation that enhances adherence and accelerates early recovery.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2500103894; https://tinyurl.com/yy5mt2ed

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2026;14:e93050) doi:10.2196/93050

KEYWORDS

total hip arthroplasty; wearable sensors; telerehabilitation; cost-effectiveness; patient adherence

Introduction

Total hip arthroplasty (THA) is in increasingly high demand worldwide driven by aging populations and expanding indications. Over 1 million THAs are performed annually, and procedural volumes are projected to rise dramatically in coming decades (an estimated 174% increase in primary THA by 2030 in the United States) [1]. As the number of THA surgeries grows, there is a critical need for scalable and accessible postoperative rehabilitation strategies that can accommodate larger patient volumes without compromising outcomes. Effective rehabilitation is essential to maximize functional recovery after THA and to ensure long-term success of the procedure. However, traditional clinic-based physiotherapy can be resource-intensive and may pose accessibility challenges, especially for patients with mobility limitations or those living far from specialized centers.

Digital health interventions and home-based rehabilitation programs have emerged as promising solutions to meet this need. Recent evidence supports the efficacy of telerehabilitation and other digital home exercise programs following joint replacement. Randomized trials have shown that home-based digital rehabilitation yields postoperative functional outcomes comparable to those of conventional in-person therapy [2,3]. Even in post-THA telerehabilitation without specialized sensors, compliance with home exercise has been observed to improve relative to standard care [3]. Moreover, a digital program has demonstrated superior short-term results: An artificial intelligence–powered home exercise system with biofeedback after THA led to significantly greater gains in mobility and higher patient-reported outcome scores than traditional rehabilitation in a pilot trial [2]. These trials underscore that technology-assisted, home-based rehabilitation can expand access without sacrificing effectiveness and, in certain cases, may enhance early recovery metrics.

Wearable sensor technology offers an innovative enhancement to digital rehabilitation. By integrating inertial motion sensors into home exercise regimens, patients can receive real-time biofeedback on their exercise technique and performance. This immediate feedback loop is posited to improve the quality of exercise execution and boost patient engagement and adherence. Early data support these benefits: Sensor-based telerehabilitation systems have been shown to increase patient adherence and exercise intensity during orthopedic recovery [4,5]. In one study of a knee rehabilitation app, patients using wearable motion sensors achieved higher rates of prescribed exercise completion and greater training intensity than those undergoing usual care, with no safety concerns [5]. These observations suggest that wearable sensor feedback could further amplify the positive effects of home rehabilitation by motivating patients and ensuring exercises are performed correctly, which in turn may translate to better functional outcomes.

Beyond clinical efficacy, home-based digital rehabilitation has important health economic implications. Remote programs can reduce the burden of travel and time costs for patients and caregivers, increasing the convenience and uptake of rehabilitation. From a health system perspective, shifting rehabilitation to the home setting may help contain costs and resource utilization [6]. This reduced “time burden” with telerehabilitation reflects potential indirect cost savings (such as fewer work absences and transportation expenses) and supports digital rehab as a cost-effective, patient-centered alternative to facility-based care [6]. Additionally, decreasing reliance on in-person supervision through digital tools could alleviate workforce demands and enable broader dissemination of rehab services to underserved or remote regions [2-6].

Building on this background, this randomized controlled trial (RCT) was designed to evaluate the added value of wearable sensor feedback in home-based digital rehabilitation following THA. We hypothesized that integrating wearable sensor–mediated biofeedback into a home-based digital rehabilitation program would improve patient adherence and lead to superior recovery after THA compared with a digital rehabilitation program without sensor feedback.

Methods

Study Design and Setting

This study was a single-center, 2-arm, parallel-group RCT reported in accordance with the CONSORT (Consolidated Standards of Reporting Trials) 2010 guidelines. It was conducted at the Chinese National Medical Center of Orthopaedics in Shanghai, China (affiliated with Shanghai Sixth People’s Hospital) from June 2023 to June 2024.

Eligible patients awaiting primary THA at the study center were identified from surgical waiting lists and screened by two physicians according to predefined inclusion criteria. Patients meeting eligibility were enrolled prior to surgery. Baseline (preoperative) evaluations for all participants were performed by a single physiotherapist after consent. On the first postoperative day, participants were randomly allocated in a 1:1 ratio to the study groups by a trial coordinator (allocation details and sequence described later). Two independent physiotherapists (not involved in initial enrollment nor outcome assessment) subsequently contacted each participant to implement the allocated rehabilitation intervention. All participants then underwent a 12-week home-based rehabilitation program according to their group assignment and were followed for a total of 24 weeks postsurgery. The 24-week follow-up duration was selected to capture both the acute rehabilitation phase, during which sensor-mediated benefits were expected to be most pronounced, and the later recovery period, during which convergence of outcomes was anticipated based on prior literature in musculoskeletal rehabilitation. Outcome assessments were conducted at baseline (presurgery)

XSL•FO
RenderX

and at 6, 12, and 24 weeks after surgery by two physiotherapists in the outpatient clinic who were blinded to group allocation to minimize observer bias. These outcome evaluators, who were not involved in providing the interventions, collected all functional performance measures and assisted participants with completing self-reported outcome questionnaires using an online system during the clinic visits. Participant adherence to the digital training intervention was monitored through system usage logs; if a physiotherapist noted that a patient was not engaging with the digital rehabilitation platform as prescribed, they promptly inquired about the reasons for nonadherence to address any barriers to participation.

Ethical Considerations

Ethics approval for this study was obtained from the Ethics Committee of Shanghai Sixth People's Hospital (approval number: 2016KY-109(K)), and the study was conducted in accordance with institutional guidelines and international ethical standards, including the principles of the Declaration of Helsinki. Written informed consent was obtained from all participants prior to enrollment, and participation was voluntary. All data were anonymized to protect participants' privacy and maintain confidentiality, ensuring that no personal identifiers were included in the analytical dataset. No compensation was provided to participants.

Participants

This single-center trial enrolled patients scheduled for primary THA from the waiting list of the orthopedic surgery outpatient clinic of a tertiary hospital. Eligible patients were identified and consented preoperatively, and baseline assessments were completed before surgery. Potential participants were informed about the study 2 weeks prior to their admission and formally screened for eligibility on the same day of their admission. Those meeting all inclusion criteria were invited to participate, and written informed consent was obtained before enrollment.

The inclusion and exclusion criteria are shown in Textbox 1. Patients meeting the criteria were enrolled by the research coordinator and randomized to begin the assigned home-based rehabilitation program. All participants were recruited from a single high-volume arthroplasty center, ensuring a consistent surgical protocol and postoperative care across the study cohort. Each participant's eligibility was verified against the criteria prior to randomization, and a screening log was maintained to record any exclusions or refusals.

Textbox 1. Eligibility criteria.

Inclusion criteria

- Age: 18 years to 80 years
- Underwent a primary THA within the prior 3 weeks for a nontraumatic indication (eg, osteoarthritis or osteonecrosis)
- Able to ambulate at least 10 meters (with or without a cane or walker) by approximately 2 weeks postsurgery
- Medically cleared by the surgical team for routine rehabilitation (no wound complications and no contraindications to unsupervised exercise, such as uncontrolled cardiac issues)
- Access to a smartphone or tablet with home internet to use the digital rehabilitation app (basic tech literacy or caregiver assistance as needed)
- Willing to be randomized and to comply with study procedures (including wearing the sensor device if assigned) and able to provide written informed consent

Exclusion Criteria

- Underwent a revision hip arthroplasty or any concurrent complex hip procedure during the index surgery (eg, periacetabular osteotomy or extensive muscle repair)
- Simultaneous bilateral THA in the same hospitalization
- Severe coexisting hip or lower-limb conditions that would impede rehabilitation (such as advanced arthritis in the opposite leg or significant neurological deficits affecting mobility/balance)
- Cognitive impairment or severe psychiatric illness that would prevent adherence to the exercise program or use of the technology
- Serious postoperative complications before enrollment that precluded normal rehab (eg, unstable medical status, a surgical site infection requiring intervention, or an unplanned reoperation for complications like dislocation)
- Already engaged in formal outpatient physiotherapy prior to enrollment that they were not willing to discontinue (for example, having started or completed an intensive supervised rehab program in the first 2 weeks post-THA)
- Concurrent participation in another interventional trial or prior participation in this study (such as a pilot enrollment)

Intervention

All participants in both the intervention and control groups engaged in a 12-week home-based rehabilitation exercise program following THA. On the second postoperative day, each patient received an in-person orientation: One of the technicians assisted them with installing a dedicated mobile app (Joymotion software, Shanghai Medmotion Medical Management Co, Ltd) on their smartphone and provided instructions on its use. This app-based platform delivered a "Digital Physiotherapeutic THA-Specific Training System," following the standardized post-THA training protocol (see Multimedia Appendix 1), and it recorded each patient's exercise performance and completion rates. The supervising physiotherapist prescribed individualized

XSL•FO
RenderX

daily exercise tasks through the app and could adjust the program remotely as needed. Both groups were scheduled for weekly teleconsultation sessions; at an appointed time each week, a physiotherapist from the rehabilitation center initiated a 2-way video call via the app to monitor progress, provide coaching, and address any issues. Notably, the frequency and duration of physiotherapist contact (including teleconsultations) were identical for the two groups to ensure parity in clinical guidance and support.

Intervention Group

Participants allocated to the inertial measurement unit (IMU) intervention group received the home-based digital exercise program via the smartphone app with full telerehabilitation features. The app provided daily interactive exercise instructions, feedback on training performance, and real-time video or audio communication with the physiotherapist. The rehabilitation regimen was automatically tailored to each patient's baseline characteristics and functional status, and the physiotherapist could further fine-tune the difficulty or exercises through a clinician interface. In addition to the app-based program, this group also used a wearable inertial sensor system to enhance exercise monitoring and feedback. Five small wireless IMU sensors were affixed to the patient's body (one at the waist and one on the front of each thigh and each lower leg) and were registered and calibrated with the app during the initial setup. During each training session, these sensors captured the patient's movements and projected them onto a virtual avatar visible in the app. The system's software continuously compared the patient's avatar movements with the pre-set correct posture for each exercise. If a movement did not reach the target position or was performed with improper form, the patient received immediate feedback from the app, prompting them to repeat the motion correctly. This real-time feedback mechanism ensured that patients in the IMU group could correct their technique on the spot, helping them perform the exercises as intended. Throughout the 12-week period, the assigned physiotherapists remotely monitored the IMU group's exercise sessions via the app's data logs. The sensor system transmitted detailed performance data to the therapist interface, allowing the physiotherapist to review each patient's adherence, exercise form, and progress. Any necessary modifications to the exercise program or additional guidance were provided based on these observations.

Control Group

Participants in the control group followed the same 12-week home exercise program using the digital platform but without the aid of wearable sensors. Like the IMU group, control group patients accessed daily THA-specific exercises and instructional videos through the mobile app, which tracked their exercise completion and provided basic performance feedback such as summaries of repetitions completed or time spent exercising. They received the same schedule of weekly physiotherapist-led teleconference sessions via the app's two-way video interface, ensuring equal contact time and guidance as the intervention group. The exercise content and progression for control participants were also individualized to their baseline function and could be adjusted remotely by the physiotherapist in the same manner as for the IMU group. However, unlike the IMU group, the control group did not receive real-time motion feedback on their exercise technique. They performed the prescribed exercises with the app's guidance alone (see Multimedia Appendix 1), without the virtual avatar or sensor-based corrections. All other aspects of the rehabilitation program were kept consistent between the two groups.

Outcome Measures and Data Collection

The primary outcome was the Hip Disability and Osteoarthritis Outcome Score (HOOS), measured at 24 weeks postsurgery as a comprehensive patient-reported index of hip-related health status [7]. Secondary outcomes were assessed at follow-up visits and included the 5 HOOS subscale scores (pain, symptoms, activities of daily living [ADL], sport/recreation, and hip-related quality of life), objective functional performance tests (the timed up and go [TUG] test and the Berg Balance Scale), generic health-related quality of life measures (36-item Short Form Health Survey [SF-36] physical and mental component summary scores), and psychological status measures (Hospital Anxiety and Depression Scale [HADS] anxiety subscale [HADS-A] and depression subscale [HADS-D]) [8-11]. Patient satisfaction with the rehabilitation program was evaluated at 12 weeks and 24 weeks using a 4-item Likert-scale questionnaire addressing convenience, ease of use and access, perceived helpfulness, and communication with the care provider (each item scored from 1 to 5, with higher scores indicating greater satisfaction).

Cost outcomes were analyzed from a societal perspective capturing direct medical costs (including the cost of the wearable sensor hardware and setup for the IMU group), direct nonmedical costs, and opportunity costs (lost wages and productivity) obtained from hospital billing records and participant self-reports. All costs were converted to 2024 Chinese Yuan (¥), with costs incurred before 2024 adjusted to 2024 price levels using the national consumer price index and any costs occurring beyond 1 year postsurgery discounted at an annual rate of 3%. A currency exchange rate of ¥1=US $0.15 is applicable.

Adherence to the rehabilitation program was measured using multiple indicators collected throughout the 24-week study period. These included the number of prescribed exercise sessions completed, the number of weekly app-based assessments attended, and the number of scheduled weekly follow-up phone calls successfully completed in each group.

Adverse events (AEs) and serious AEs (SAEs) were closely tracked as part of safety monitoring from enrollment through the full 24-week trial. Participants were instructed to report any AEs or SAEs either through the study's mobile app (which included a "Report a Problem" feature) or via a phone call to the research team. For this study, an AE was defined as any untoward or unintended medical occurrence during the trial, regardless of whether it was considered related to the intervention. An SAE was defined using standard criteria as any adverse event that resulted in death, was life-threatening, required or prolonged hospitalization, caused significant disability or incapacity, or captured any other important medical event that necessitated intervention to prevent such outcomes. Follow-up assessments were conducted at 6 weeks, 12 weeks,

XSL•FO
RenderX

and 24 weeks postsurgery in the outpatient clinic. Two trained physiotherapists conducted all outcome measurements at each visit. Participants completed patient-reported outcome measures, including HOOS, HADS, and SF-36, using paper questionnaires. All performance-based outcome tests were administered in person by the physiotherapists during each follow-up visit.

Sample Size Calculation

The sample size was determined to provide sufficient power to detect a clinically meaningful difference in the primary outcome (HOOS) between groups. Based on published data in THA populations, an improvement of roughly 8 to 10 points on the HOOS represents the minimum clinically important difference in patient-reported hip function after surgery. We therefore set a 10-point difference in HOOS as the target effect size for this trial. To inform the variability estimate, we reviewed prior studies reporting HOOS outcomes and found that the SD of change scores is approximately 15 to 20 points in this context. Using these assumptions and a 2-sided significance level of .05 with 80% statistical power, we calculated that 86 patients per group would be required to detect a 10-point difference in HOOS between the intervention and control arms. This calculation was based on a 2-sample comparison of means for the primary outcome. To account for an anticipated attrition rate of about 25% (loss to follow-up and dropouts), the sample size was inflated accordingly. We aimed to recruit approximately 115 participants per arm to ensure that the study would remain adequately powered for the primary endpoint after accounting for potential dropouts. This sample size was also expected to provide sufficient power for key secondary outcomes while maintaining a feasible recruitment target.

Randomization, Allocation, and Blinding

An independent statistician generated the allocation sequence using block randomization with randomly varying block sizes (between 4 and 6), achieving a 1:1 allocation ratio to the 2 study groups. Allocation concealment was maintained by placing assignments in sequentially numbered, opaque, sealed envelopes. Randomization was implemented after participants completed baseline assessments. At that time, the enrolling investigator opened the next envelope in sequence to reveal the participant's group assignment. The allocation list was kept secure by the statistician and was not accessible to investigators or clinicians responsible for recruitment and treatment, ensuring that allocations remained concealed until the point of assignment. Due to the nature of the intervention, participants and treating physiotherapists were not blinded to group allocation; however, all outcome assessors and data analysts remained blinded to group assignments throughout the study to minimize the risk of bias.

Statistical Analysis

Baseline characteristics, outcome measures, and treatment adherence data were summarized using appropriate descriptive statistics (mean and SD for continuous variables; frequencies and percentages for categorical variables). We applied the Shapiro-Wilk test to formally evaluate normality [12]. A nonsignificant Shapiro-Wilk test ($P>.05$) indicated no evidence against normality. Any major deviations would have prompted use of transformations or nonparametric methods, but no such issues arose in this study. All hypothesis tests were 2-sided with a significance level of 5%. Intention-to-treat (ITT) analyses were primary, including all participants as randomized.

For the primary outcome (HOOS at 24 weeks postrandomization), an adjusted 2-sided analysis was conducted in the ITT population. The primary between-group comparison of HOOS at 24 weeks was based on a linear mixed effects model treating the repeated HOOS measurements as the dependent outcome. This model included fixed effects for treatment group, time, and the treatment-by-time interaction, with baseline HOOS score, age, and sex entered as covariates to improve precision. A random intercept for each participant was incorporated to account for within-subject correlation over time. As this was a single-center trial, no random center effect was needed. The adjusted mean difference in HOOS at 24 weeks between the 2 groups was estimated from this model with a 2-sided 95% CI. The primary hypothesis was evaluated using the *P* value associated with the treatment group effect (difference in outcomes between groups) at the 24-week time point. Model assumptions (such as normality of residuals and homogeneity of variance) were checked, and analyses were robust to these assumptions or used appropriate transformations if necessary.

Continuous secondary outcomes were analyzed with similar linear mixed effects models appropriate for their distribution. Each continuous outcome measured over time was examined for overall group differences and trajectory over the follow-up period. These models included the same fixed effects (group, time, and their interaction) and participant-level random intercepts, adjusting for baseline values when available. From each model, we obtained estimates of the mean outcome in each group at each follow-up time and the mean differences between groups with 95% CIs. Group differences at each time point were derived from interaction terms or post hoc contrasts. For binary categorical outcomes, comparisons between groups were made using χ^2 tests. These dichotomous results were reported as risk differences or relative risks with corresponding 95% CIs. All significance tests were 2-sided, and a *P* value $<.05$ was considered statistically significant for each individual outcome analysis. To control for inflated Type I error across secondary endpoints, a Bonferroni correction was applied to all secondary outcome comparisons. The primary outcome (HOOS overall at 24 weeks) was tested at a 2-sided significance level of .05 as the single confirmatory analysis. For secondary outcomes, the family-wise alpha was divided by the total number of secondary comparisons (47 tests across 15 secondary outcome measures at up to 3 time points, plus 2 nonprimary time points for HOOS overall), yielding a Bonferroni-adjusted significance threshold of $P<.001$. Both unadjusted and Bonferroni-adjusted significance levels are reported for transparency.

A comprehensive cost-effectiveness analysis was also performed alongside the clinical outcomes. Total health care costs for each participant over the 24-week follow-up were calculated, incorporating intervention costs and any other relevant medical expenditures. The difference in mean total cost between the 2 groups was then evaluated using appropriate between-group comparisons (with nonparametric bootstrap methods applied to account for skewed cost data). We plotted the joint uncertainty

XSL•FO
RenderX

in costs and effects on a cost-effectiveness plane to visualize the difference in mean cost against the difference in effectiveness (primary outcome improvement) for the digital rehabilitation versus control. If the digital intervention demonstrated lower average cost while yielding no worse outcome, it was considered a dominant strategy (more cost-effective). In cases where the digital program had higher cost or marginally different outcomes, an incremental cost-effectiveness ratio (ICER) was computed to quantify the additional cost per unit of benefit gained. Uncertainty in the cost-effectiveness results was explored by nonparametric bootstrapping (10,000 resamples), which produced a cloud of incremental cost-effect pairs on the cost-effectiveness plane and allowed construction of cost-effectiveness acceptability curves. These acceptability curves illustrate the probability that the digital program is cost-effective compared with the control across a range of willingness-to-pay thresholds for improvement in outcome.

Missing data were handled using both ITT and per-protocol (PP) analysis strategies. In the ITT analysis, all participants were retained in their originally assigned groups for analysis, regardless of protocol adherence or missing follow-up data. The linear mixed effects model used all available observations from each participant under a missing-at-random assumption, thereby inherently accommodating incomplete follow-up without requiring formal imputation. In contrast, the PP analysis included only those participants who completed all scheduled follow-up assessments, thereby using a complete-case approach. It is important to note that missingness in our dataset was holistic at the follow-up level: If a patient was lost to follow-up at a given time point, all variables for that visit were missing for that patient. There was no item-level missingness within an attended visit. Consequently, missing data occurred only at the participant-visit level, and all available data from each participant's other time points were still used in the mixed model under the ITT approach. Because entire follow-up visits were missing for dropouts and not selective outcome omissions, we did not perform any statistical imputation for missing values. In addition, given this pattern of whole-visit missingness, a formal test for data missing completely at random (such as the Little MCAR test) was not applicable, as that test is designed for detecting patterns of sporadic item-level missingness across variables rather than the uniform visit-level dropout observed in this trial.

Results

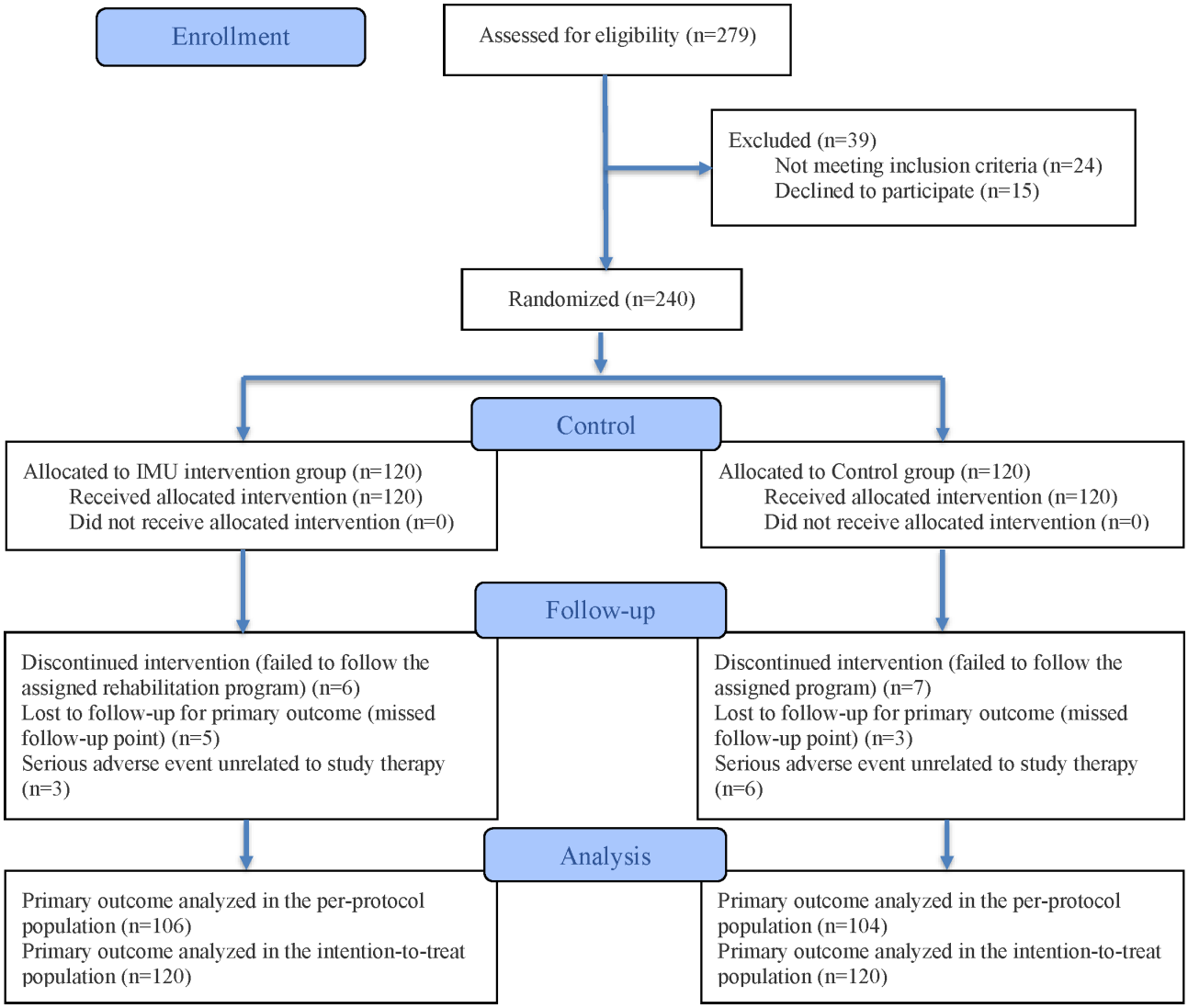
Participants

We evaluated 279 patients scheduled for primary THA, and 240 were randomized. The flowchart of our study is shown in Figure 1. The baseline characteristics were comparable between the IMU and control groups. Men accounted for 62 (62/120, 51.7%) participants in the IMU group and 56 (56/120, 46.7%) participants in the control group (P=.44); the mean age was 70 years versus 69 years (P=.46); mean BMI was 22 kg/m² versus 22 kg/m² (P=.31; Table 1). Paracetamol or nonsteroidal anti-inflammatory drugs were taken by 17 (17/120, 14.2%) in the intervention group versus 18 (18/120, 15%) in the control group (P=.86). In addition, 62% (74/120, 61.7% vs 72/120, 60%) had the left hip as the surgical side (P=.79). There were no significant differences in functional tests nor patient-reported outcome measures.

Sensitivity analysis using the PP population showed similar results as those of the ITT population (Table S1 in Multimedia Appendix 2).

XSL•FO
RenderX

Figure 1. Flow diagram of the study. IMU: inertial measurement unit.



XSL•FO
RenderX

Table . Baseline characteristics of the inertial measurement unit (IMU) group and control group (intention-to-treat population).

Characteristics	IMU (n=120)	Control group (n=120)	*P* value
Demographics			
Gender (male), n (%)	62 (51.7)	56 (46.7)	.44
Age (years), mean (SD)	70 (5.14)	69 (5.11)	.46
BMI (kg/m^2), mean (SD)	22 (2.01)	22 (1.86)	.31
Occupation, n (%)			.67
Manual worker	84 (70)	87 (72.5)	
Nonmanual worker	36 (30)	33 (27.5)	
Education level, n (%)			.77
Lower than high school	89 (74.2)	87 (72.5)	
Equal/higher than high school	31 (25.8)	33 (27.5)	
Insurance type, n (%)			.38
Government	85 (70.8)	80 (66.7)	
Commercial	11 (9.2)	18 (15)	
Self-financed	24 (20)	22 (18.3)	
Current smoker, n (%)	42 (35)	48 (40)	.42
Current alcohol use, n (%)	32 (26.7)	29 (24.2)	.66
Paracetamol and NSAID[a] use, n (%)	17 (14.2)	18 (15)	.86
Side of hip (left), n (%)	74 (61.7)	72 (60)	.79
Living alone (yes), n (%)	76 (63.3)	75 (62.5)	.89
Function, mean (SD)			
HOOS[b]			
Pain	41 (4.30)	40 (5.08)	.26
Symptoms	40 (5.38)	40 (5.31)	.67
Function in ADLs[c]	41 (4.85)	41 (4.40)	.82
Sports and recreation	25 (3.16)	25 (3.52)	≥.99
Hip-related quality of life	25 (3.52)	25 (3.73)	.75
Overall	38 (2.63)	38 (2.43)	.56
TUG[d] test (seconds)	31 (4.44)	30 (4.80)	.30
Berg Balance Scale	35 (3.51)	35 (3.84)	.41
PROMs[e]			
SF36[f] physical component summary	25 (4.98)	25 (4.83)	.84
SF36 mental component summary	41 (5.79)	40 (6.44)	.87
HADS[g] anxiety subscale	10 (2.03)	10 (2.02)	.92
HADS depression subscale	10 (1.85)	10 (2.08)	.95

[a]NSAID: nonsteroidal anti-inflammatory drug.

[b]HOOS: Hip Disability and Osteoarthritis Outcome Score.

[c]ADLs: activities of daily living.

[d]TUG: timed up and go.

[e]PROMS: patient-reported outcome measures.

[f]36-item Short Form Health Survey.

XSL•FO
RenderX

[g]HADS: Hospital Anxiety and Depression Scale.

Primary and Secondary Outcomes

The primary outcome (HOOS) improved significantly more in the IMU group than in the control group at all follow-ups. The estimated between-group difference in the overall HOOS (IMU minus control) was 2.021 (95% CI 1.445 - 2.598) points at 6 weeks (*P*<.001), 1.721 (95% CI 0.896 - 2.545) points at 12 weeks (*P*<.001), and 1.298 (95% CI 0.313 - 2.282) points at 24 weeks (*P*=.01). All 5 HOOS subscales showed a pattern of greater early improvement with the IMU intervention. At 6 weeks, 4 of the 5 subscales (pain, symptoms, sport/recreation, and hip-related quality of life) had significant between-group differences after Bonferroni correction (all *P*<.001), with coefficients ranging from 2.334 to 3.327 points. The HOOS ADL subscale difference at 6 weeks (1.208 points, 95% CI 0.283 to 2.134; *P*=.01) did not survive Bonferroni correction. At 12 weeks, none of the individual HOOS subscale differences remained significant after correction for multiple comparisons (unadjusted *P* values ranged from .002 to .047). By 24 weeks, all subscale differences had further narrowed and were not statistically significant at either the unadjusted or adjusted level (all *P*>.13), except for HOOS pain, which had a nominal difference of 1.501 (95% CI 0.084 to 2.918) points (unadjusted *P*=.04) that did not survive Bonferroni correction. In terms of objective function, the IMU group completed the TUG faster than controls by 1.198 (95% CI 0.789 to 1.607) seconds at 6 weeks (*P*<.001) and 1.131 (95% CI 0.557 to 1.705) seconds at 12 weeks (*P*<.001), both of which remained significant after Bonferroni correction. The 0.901-second (95% CI 0.237 to 1.565) difference at 24 weeks was nominally significant (unadjusted *P*=.008) but did not survive Bonferroni correction. Berg Balance Scale scores were numerically higher in the IMU group at all time points (0.300 to 0.431 points), but none of these differences reached the Bonferroni-adjusted threshold (unadjusted *P* values of .004, .02, and .16 at 6 weeks, 12 weeks, and 24 weeks, respectively). For generic patient-reported outcomes, the SF-36 physical component summary was higher in the IMU group by 3.821 (95% CI 3.315 to 4.326) points at 6 weeks, 2.626 (95% CI 1.913 to 3.338) points at 12 weeks, and 1.949 (95% CI 1.144 to 2.753) points at 24 weeks (all *P*<.001; all surviving Bonferroni correction). The SF-36 mental component summary showed a similar pattern (3.239, 95% CI 2.881 to 3.597; 2.272, 95% CI 1.771 to 2.773; and 1.726, 95% CI 1.117 to 2.334 points at 6 weeks, 12 weeks, and 24 weeks, respectively; all *P*<.001; all surviving Bonferroni correction). Psychological outcomes also consistently favored the IMU arm across all time points and remained significant after Bonferroni correction: The IMU group had lower HADS-A scores by 0.537 (95% CI 0.470 to 0.604), 0.418 (95% CI 0.310 to 0.527), and 0.363 (95% CI 0.220 to 0.506) points at 6 weeks, 12 weeks, and 24 weeks, respectively (all *P*<.001), and lower HADS-D scores by 0.501 (95% CI 0.435 to 0.566), 0.370 (95% CI 0.260 to 0.481), and 0.307 (95% CI 0.163 to 0.450) points at 6 weeks, 12 weeks, and 24 weeks, respectively (all *P*<.001). Patient satisfaction with the rehabilitation program was high in both groups, with no significant between-group differences in any domain at 12 weeks or 24 weeks (all *P*>.23). Detailed results of the primary and secondary outcomes are shown in Tables 2 and 3. Sensitivity analyses in the PP population yielded results consistent with the ITT findings, reinforcing the robustness of the observed treatment effects (Tables S2-S3 in Multimedia Appendix 2).

XSL•FO
RenderX

Table . Unadjusted changes in outcomes for the inertial measurement unit (IMU) and control groups at the 6-week, 12-week, and 24-week postsurgical follow-ups based on intention-to-treat analysis.

Outcome	6 weeks postsurgery, mean (SD)[a]			12 weeks postsurgery, mean (SD)[a]			24 weeks postsurgery, mean (SD)[a]		
	IMU (n=120)	Control group (n=120)	*P* value[b]	IMU (n=120)	Control group (n=120)	*P* value[b]	IMU (n=120)	Control group (n=120)	*P* value[b]
Functional outcomes[c]									
HOOS[d]									
Pain	13.27 (8.10)	8.58 (4.82)	<.001	23.88 (9.28)	23.46 (9.42)	.74	32.65 (10.20)	32.36 (10.93)	.84
Symptoms	17.54 (8.18)	11.90 (7.56)	<.001	29.21 (10.61)	28.70 (12.92)	.75	37.85 (11.58)	37.98 (12.93)	.94
Function in ADLs[e]	15.85 (8.21)	13.31 (5.75)	.008	28.77 (10.68)	26.75 (9.32)	.14	35.59 (10.78)	35.07 (10.01)	.71
Sports and recreation	22.17 (10.34)	15.28 (9.60)	<.001	41.13 (15.77)	40.77 (11.84)	.85	51.53 (17.64)	51.61 (12.20)	.97
Hip-related quality of life	21.66 (10.96)	17.27 (7.40)	.001	42.43 (16.09)	42.11 (13.39)	.87	54.27 (15.95)	55.00 (14.52)	.72
Overall	16.62 (5.13)	12.57 (2.98)	<.001	30.18 (5.88)	29.13 (5.12)	.16	38.58 (6.06)	38.42 (6.09)	.85
TUG[f] test (seconds)	−6.19 (3.72)	−3.62 (3.11)	<.001	−8.96 (4.06)	−7.70 (3.77)	.02	−11.39 (4.30)	−10.83 (4.57)	.35
Berg balance test	5.19 (2.58)	4.46 (1.90)	.02	7.61 (2.68)	7.15 (2.23)	.17	10.20 (2.82)	10.31 (2.56)	.76
PROMs[g]									
SF-36[h]									
Physical component summary	30.00 (2.81)	22.50 (2.51)	<.001	42.34 (3.86)	42.32 (3.69)	.96	50.09 (4.77)	50.09 (4.39)	≥.99
Mental component summary	20.07 (2.12)	13.64 (2.12)	<.001	25.23 (4.37)	25.13 (3.47)	.85	33.15 (4.36)	33.16 (3.48)	.98
HADS[i]									
Anxiety subscale (HAD-A)	−2.57 (0.50)	−1.48 (0.50)	<.001	−3.43 (1.51)	−3.29 (0.83)	.39	−4.43 (1.51)	−4.25 (0.91)	.29
Depression subscale (HAD-D)	−2.48 (0.50)	−1.44 (0.50)	<.001	−3.58 (1.28)	−3.50 (1.11)	.64	−4.59 (1.29)	−4.47 (1.11)	.49
Patients' evaluation of the current intervention									
Convenience	0.00 (0.23)	−0.01 (0.28)	.80	4.75 (1.05)	0.00 (0.27)	<.001	0.00 (0.23)	0.01 (0.29)	.80
Ease of use/access	0.01 (0.21)	−0.01 (0.25)	.57	4.75 (1.05)	0.00 (0.27)	<.001	0.00 (0.27)	−0.01 (0.29)	.81
Helpfulness	0.00 (0.23)	−0.01 (0.25)	.78	4.75 (1.05)	−0.01 (0.25)	<.001	0.00 (0.27)	−0.02 (0.27)	.62
Communication with the care provider	0.00 (0.23)	0.00 (0.27)	≥.99	4.74 (1.05)	0.00 (0.27)	<.001	−0.01 (0.25)	0.01 (0.25)	.60

XSL•FO
RenderX

[a]The mean change from baseline for each group.

[b]Between-group differences.

[c]Functional outcomes included primary endpoints and secondary performance-based tests.

[d]HOOS: Hip Disability and Osteoarthritis Outcome Score.

[e]ADLs: activities of daily living.

[f]TUG: timed up and go.

[g]PROMs: patient-reported outcome measures.

[h]36-Item Short Form Health Survey.

[i]HADS: Hospital Anxiety and Depression Scale.

XSL•FO
RenderX

Table . Effectiveness estimates from linear mixed effects models for the inertial measurement unit (IMU) and control groups at the 6-week, 12-week, and 24-week follow-ups based on the intention-to-treat analysis.

Outcome[a]	6 weeks post-surgery		12 weeks post-surgery		24 weeks post-surgery	
	Coefficient (95% CI)[b]	*P* value	Coefficient (95% CI)[b]	*P* value	Coefficient (95% CI)[b]	*P* value
Functional outcomes						
HOOS[c]						
Pain	2.441 (1.632 to 3.250)	<.001	1.874 (0.702 to 3.045)	.002	1.501 (0.084 to 2.918)	.04
Symptoms	2.683 (1.638 to 3.727)	<.001	1.973 (0.415 to 3.531)	.01	1.416 (−2.191 to 3.271)	.14
Function in ADLs[d]	1.208 (0.283 to 2.134)	.01	1.390 (0.016 to 2.765)	.047	1.079 (−0.566 to 2.724)	.20
Sports and recreation	3.327 (1.991 to 4.664)	<.001	2.276 (0.426 to 4.126)	.02	1.555 (−0.652 to 3.762)	.17
Hip-related quality of life	2.334 (1.152 to 3.517)	<.001	1.873 (0.035 to 3.711)	.046	1.321 (−0.864 to 3.505)	.24
Overall	2.021 (1.445 to 2.598)	<.001	1.721 (0.896 to 2.545)	<.001	1.298 (0.313 to 2.282)	.01
TUG[e] test (seconds)[f]	−1.198 (−1.607 to −0.789)	<.001	−1.131 (−1.705 to −0.557)	<.001	−0.901 (−1.565 to −0.237)	.008
Berg Balance Scale[f]	0.393 (0.125 to 0.661)	.004	0.431 (0.065 to 0.797)	.02	0.300 (−0.118 to 0.718)	.16
PROMs[g]						
SF-36[h,f]						
Physical component summary	3.821 (3.315 to 4.326)	<.001	2.626 (1.913 to 3.338)	<.001	1.949 (1.144 to 2.753)	<.001
Mental component summary	3.239 (2.881 to 3.597)	<.001	2.272 (1.771 to 2.773)	<.001	1.726 (1.117 to 2.334)	<.001
HADS[i,f]						
Anxiety subscale (HAD-A)	−0.537 (−0.604 to −0.470)	<.001	−0.418 (−0.527 to −0.310)	<.001	−0.363 (−0.506 to −0.220)	<.001
Depression subscale (HAD-D)	−0.501 (−0.566 to −0.435)	<.001	−0.370 (−0.481 to −0.260)	<.001	−0.307 (−0.450 to −0.163)	<.001
Patients' evaluation of the current intervention[f]						
Convenience	0.006 (−0.018 to 0.030)	.65	0.007 (−0.015 to 0.029)	.54	0.005 (−0.014 to 0.025)	.59
Ease of use/access	0.008 (−0.013 to 0.029)	.44	0.012 (−0.008 to 0.032)	.23	0.011 (−0.008 to 0.031)	.24
Helpfulness	−0.003 (−0.026 to 0.021)	.82	0.000 (−0.020 to 0.021)	.97	0.000 (−0.021 to 0.022)	.97
Communication with the care provider	0.005 (−0.015 to 0.026)	.61	0.002 (−0.018 to 0.021)	.88	−0.001 (−0.020 to 0.018)	.93

[a]All outcome measures were adjusted for baseline values in the model.

[b]Estimated between-group difference in the change from baseline (IMU group minus control group) for the specified outcome, with positive coefficients indicating higher scores in the IMU group compared with the control group and negative values indicating lower scores in the IMU group.

[c]HOOS: Hip Disability and Osteoarthritis Outcome Score.

[d]ADLs: activities of daily living.

XSL•FO
RenderX

[e]TUG: timed up and go.

[f]Secondary outcomes with *P*<.001 remained significant after Bonferroni correction. A Bonferroni correction was applied to 47 secondary comparisons, yielding an adjusted significance threshold of *P*<.001.

[g]PROMs: patient-reported outcome measures.

[h]36-item Short Form Health Survey.

[i]HADS: Hospital Anxiety and Depression Scale.

Cost Measurements

The mean 24-week total cost per patient was ¥87,967.32 (SD ¥8854.19) in the IMU group versus ¥94,396.30 (SD ¥10,037.74) in the control group (*P*=.32). The cost composition differed between groups: IMU participants had significantly lower in-hospital costs (¥69,534.95, SD ¥5394.69 vs ¥70,359.45, SD ¥5043.84; *P*<.001) and primary care costs (¥250.00, SD ¥38.14 vs ¥1503.18, SD ¥186.13; *P*<.001). Nonmedical expenses were lower with IMU, including paid home care (¥1178.73, SD ¥362.12 vs ¥1451.95, SD ¥357.74; *P*<.001), transportation (¥800.06, SD ¥138.89 vs ¥1968.67, SD ¥273.82; *P*<.001), and nutrition (¥4505.32, SD ¥479.00 vs ¥5093.08, SD ¥519.96; *P*<.001). Patient productivity losses were also lower (¥1599.77, SD ¥6994.72 vs ¥1829.44, SD ¥6618.91; *P*<.001). The detailed results are shown in Table 4.

The IMU intervention's incremental cost was –¥6428.98 (IMU less costly). Correspondingly, most ICERs were negative, though the overall cost difference between groups was not statistically significant (¥87,967.32 vs ¥94,396.30, *P*=.32). For the primary outcome (HOOS overall), the incremental effect was +1.298 (95% CI 0.313 to 2.282) points, with an ICER of –¥4954 per point. TUG time decreased by 0.901 (95% CI 0.237 to 1.565) seconds (ICER ¥7136.79 per second), and the Berg Balance Scale increased by 0.300 (95% CI –0.118 to 0.718) points (ICER –¥21,451.43 per point). SF-36 physical component summary and mental component summary improved by 1.949 (95% CI 1.144 to 2.753) points (ICER –¥3299.31 per point) and 1.726 (95% CI 1.117 to 2.334) points (ICER –¥3725.41 per point), respectively. The detailed results, including for the HOOS subscales, are shown in Table 5. PP sensitivity analyses (Tables S4 and S5 in Multimedia Appendix 2) yielded similar cost-effectiveness results.

Table . Average total cost per patient in the inertial measurement unit (IMU) group and control group during the 24 weeks after the surgery.

Cost category[a] (¥[b])	IMU (n=120)	Control group (n=120)	*P* value
Medical costs[c]			
Costs while in the hospital (including the hospital stay, surgery fee, medication, implants during surgery)	69,534.95 (5394.69)	70,359.45 (5043.84)	<.001
Physical therapist costs (including the instrument)	6980.00 (0.00)	8447.88 (1046.07)	.24
Primary care costs	250.00 (38.14)	1503.18 (186.13)	<.001
Nonmedical costs[d]			
Paid home costs	1178.73 (362.12)	1451.95 (357.74)	<.001
Transportation costs	800.06 (138.89)	1968.67 (273.82)	<.001
Nutrition costs	4505.32 (479.00)	5093.08 (519.96)	<.001
Opportunity costs[e]			
Lost wages for patients	1599.77 (6994.72)	1829.44 (6618.91)	<.001
Lost wages for families	3118.49 (4461.09)	3742.65 (4743.85)	.80
Total cost	87,967.32 (8854.19)	94,396.30 (10,037.74)	.32

[a]All relevant costs were captured from a societal perspective and were obtained from hospital billing records and patient self-reports where needed (wage losses). All costs are reported in 2024 Chinese Yuan, as 2024 was the end of the enrollment period; costs incurred in earlier years were adjusted to 2024 price levels using the consumer price index. Costs and outcomes occurring beyond the initial year were discounted at an annual rate of 3% to reflect time preference, consistent with standard practice in health economic evaluations.

[b]A currency exchange rate of ¥1=US $0.15 is applicable.

[c]Hospital stay; primary and secondary care; rehabilitation and physical therapy costs, including wearable sensor hardware and setup for the IMU group; and medication costs.

[d]Direct costs related to care (transportation for medical visits and any specialized nutritional support during recovery).

[e]Indirect costs due to productivity loss (lost wages for patients during recovery or disability, and lost income for family members/caregivers, if applicable).

XSL•FO
RenderX

Table . Incremental cost-effectiveness ratio (ICER), defined as the incremental cost divided by the incremental effect, representing the additional cost per unit improvement in outcome for the inertial measurement unit (IMU) group versus the control group.

Items	Main analysis: mixed effects (95% CI)
Incremental cost (¥)	−6428.98
Incremental HOOS[a] pain	1.501 (0.084 to 2.918)
Incremental HOOS symptoms	1.416 (−2.191 to 3.271)
Incremental HOOS function in ADL[b]	1.079 (−0.566 to 2.724)
Incremental HOOS sports and recreation	1.555 (−0.652 to 3.762)
Incremental HOOS hip-related quality of life	1.321 (−0.864 to 3.505)
Incremental HOOS overall	1.298 (0.313 to 2.282)
Incremental TUG[c] test (seconds)	−0.901 (−1.565 to −0.237)
Incremental Berg Balance Scale	0.300 (−0.118 to 0.718)
Incremental SF36[d]-physical component summary	1.949 (1.144 to 2.753)
Incremental SF36-mental component summary	1.726 (1.117 to 2.334)
Incremental HADS[e] anxiety subscale (HAD-A)	−0.363 (−0.506 to −0.220)
Incremental HADS depression subscale (HAD-D)	−0.307 (−0.450 to −0.163)
Incremental convenience	0.005 (−0.014 to 0.025)
Incremental ease of use/access	0.011 (−0.008 to 0.031)
Incremental helpfulness	0.000 (−0.021 to 0.022)
Incremental communication with the care provider	−0.001 (−0.020 to 0.018)
ICER[f] (¥)	
ICER HOOS pain	−4282.69
ICER HOOS symptoms	−4540.12
ICER HOOS function in ADL	−5958.28
ICER HOOS sports and recreation	−4134.15
ICER HOOS hip-related quality of life	−4868.05
ICER HOOS overall	−4954.13
ICER TUG test (seconds)	7136.79
ICER Berg balance test	−21,451.43
ICER SF36-physical component summary	−3299.31
ICER SF36-mental component summary	−3725.41
ICER HADS anxiety subscale (HAD-A)	17,704.97
ICER HADS depression subscale (HAD-D)	20,954.26
ICER convenience	−1,204,447.61
ICER ease of use/access	−559,377.37
ICER helpfulness	−14,782,662.68
ICER communication with the care provider	7,845,003.05

[a]HOOS: Hip Disability and Osteoarthritis Outcome Score.

[b]ADLs: activities of daily living.

[c]TUG: Timed Up and Go test.

[d]36-item Short Form Health Survey.

[e]HADS: Hospital Anxiety and Depression Scale.

[f]Incremental effects for primary outcome measures for the 24-week follow-up under both the intention-to-treat and per-protocol analyses. A negative ICER suggests that the IMU intervention achieved numerically better outcomes at a numerically lower cost compared with the control group; however, the statistical significance of both the cost and effect differences should be considered when interpreting these ratios.

XSL•FO
RenderX

Adherence and AEs

Adherence to the intervention was high in both groups, with the IMU group completing more of the 60 prescribed exercise sessions (mean 51.6, SD 9.3 vs mean 48.7, SD 10.6; *P*=.002), attending more of the 12 weekly app-based assessments (mean 10.5, SD 1.5 vs mean 9.5, SD 1.8; *P*<.001), and completing more of the 12 follow-up phone calls (mean 10.8, SD 1.6 vs mean 9.3, SD 1.8; *P*<.001) than the control group (Table S6 in Multimedia Appendix 2). AEs were infrequent, occurring in 7 patients (7/120, 5.8%) in the IMU group and 9 patients (9/120, 7.5%) in the control group; most AEs were minor and considered related to treatment (Table S7 in Multimedia Appendix 2).

Discussion

This RCT demonstrated that integrating a wearable sensor into a home-based digital rehabilitation program after THA resulted in significantly greater improvements in patient outcomes compared with a standard digital program without sensor feedback. Specifically, the sensor group achieved significantly larger gains in the primary hip outcome measure and multiple secondary functional and quality-of-life outcomes, particularly during early recovery, although most between-group differences narrowed by the final 24-week follow-up. These statistical differences were attained without increasing health care costs. Total 24-week costs were comparable between groups. Although several individual cost components were numerically lower in the sensor arm, the overall cost difference did not reach statistical significance, suggesting that the sensor-based program was cost-neutral rather than definitively cost-saving. Adherence to the prescribed rehabilitation was high in both groups, with the sensor group completing a higher proportion of exercise sessions and assessments than the control group.

Our findings suggest that augmenting home-based digital rehabilitation with real-time sensor feedback can confer modest gains in early postoperative recovery, though the clinical significance of these gains warrants careful interpretation. The IMU‐enhanced training led to faster improvements in hip function and mobility during the initial weeks, although not all secondary hip-specific outcomes survived Bonferroni correction for multiple comparisons, which is consistent with other orthopedic populations. For example, a recent RCT with patients with diabetes who had undergone total knee arthroplasty demonstrated that adding IMU sensor feedback to telerehabilitation significantly improved functional outcomes compared with telerehabilitation alone [13]. More broadly, our results align with a growing body of evidence that technology-assisted or remote rehabilitation is at least as effective as standard physiotherapy. Meta-analyses in knee and hip arthroplasty have shown that telerehabilitation can yield equal or even superior short-term improvements in pain, range of motion, strength, and functional scores when compared with traditional in-person programs [14-16]. These early advantages are likely attributable to enhanced exercise technique, engagement, and adherence driven by the biofeedback component. Notably, the between-group differences in hip-specific outcomes narrowed by the 24-week follow-up. After Bonferroni correction, no HOOS subscale nor TUG difference remained significant at 24 weeks. However, the SF-36 and HADS differences persisted through the final follow-up even after correction, suggesting that the sensor-mediated benefits on broader quality of life and psychological well-being were more durable than the hip-specific functional gains. This trajectory mirrors prior reports of musculoskeletal rehabilitation, indicating that initial benefits of more intensive or innovative interventions may diminish over time as all patients progress through natural recovery [15]. Indeed, telerehabilitation appears to facilitate a comparable 6-month outcome to that of conventional care in many settings, supporting its use as a viable alternative without long-term trade-offs [15,16]. It is important to acknowledge that, although the improvements with sensor guidance were statistically significant, their absolute magnitude was relatively small. For instance, the HOOS overall score in the IMU group was only 2 points higher on average, a difference below the typical minimal clinically important difference reported for hip outcomes [12]. This suggests that, although sensor-based digital training accelerates early rehabilitation and subjective recovery, the clinical significance of the advantage is modest. Nevertheless, even slight early gains in function and pain relief can be valuable; achieving milestones sooner may improve patient confidence, expedite return to daily activities, and potentially reduce short-term complication risks. In summary, the interpretation of our findings is consistent with existing evidence: Wearable sensor integration enhances the quality of home exercise performance and adherence, thereby providing a small boost to early postoperative outcomes, while overall recovery at the mid-term remains comparable to traditional rehabilitation approaches [13-16].

From a clinical perspective, these findings support integrating wearable sensor–enhanced telerehabilitation into standard post-THA care. Patients in the sensor-based program achieved faster early gains in functional mobility and quality of life without compromising longer-term outcomes. This accelerated recovery can improve the early postoperative experience and potentially reduce short-term disability. Importantly, the sensor feedback substantially boosted patient adherence to prescribed exercise, a critical determinant of successful rehabilitation, echoing prior reports of higher compliance with telerehab interventions [17]. The approach was also safe and well-received, with similar patient satisfaction and no increase in adverse events compared with usual care. Furthermore, our trial's economic analysis indicates that delivering home-based digital rehabilitation with wearable sensors is feasible without added cost burden. Although the point estimate suggested numerically lower costs in the sensor group, the overall cost difference was not statistically significant, and shifting rehabilitation into the home may help reduce indirect costs (travel, time off work) and alleviate demands on in-person physiotherapy services [6,17]. With THA volumes rising globally, adopting such technology-assisted home programs could expand rehabilitation capacity, improve access for patients in remote or resource-limited settings, and enhance the value of care by providing efficient, patient-centered recovery support [17].

This study has several notable strengths, including its robust randomized controlled design (with blinded outcome

XSL•FO
RenderX

assessments) and comprehensive outcome evaluation encompassing both clinical and economic measures, which bolsters confidence in the findings. However, there are also important limitations. Methodologically, some experimental details lacked clarity or could have been more appropriately standardized, potentially affecting reproducibility. Analytically, the data analysis methods, although rigorous, involved statistical assumptions that may limit their appropriateness or introduce bias or uncertainty. Additionally, the results should be interpreted with caution, as conclusions remain contingent on the study's context and design. We also strived for scientific accuracy and consistent terminology, but minor ambiguities in usage and definitions may persist. These considerations highlight areas for improvement; future studies will address these concerns by refining methodology and analysis, clarifying terminology, and further validating our findings.

In conclusion, this RCT showed that adding wearable sensor feedback to a home-based digital THA rehabilitation program led to improved early functional outcomes, sustained benefits in patient-reported quality of life and psychological well-being, and significantly higher patient adherence compared with a similar program without sensors while incurring no additional costs. Several early hip-specific functional differences did not survive Bonferroni correction for multiple comparisons, and most HOOS subscale differences had converged by 24 weeks. Although the between-group differences in HOOS were statistically significant, their magnitude remained below the established minimal clinically important difference, tempering the clinical significance of these gains. Nevertheless, the sensor-enhanced approach provided consistent early benefits across multiple outcome domains and substantially improved exercise adherence. The intervention was cost-neutral, suggesting that integrating wearable technology in post-THA rehabilitation can enhance patient engagement and accelerate early recovery without adding to the economic burden.

Acknowledgments

We sincerely thank all the patients who participated in this study.

The authors also confirm that no generative artificial intelligence (AI) tools (such as ChatGPT) were used in the writing of this manuscript.

Funding

No external financial support nor grants were received from any public, commercial, or not-for-profit entity for the research, authorship, or publication of this article.

Data Availability

The datasets analyzed during this study are not publicly available due to the confidential policy of Shanghai Sixth People's Hospital but are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: JY
Formal analysis: CJ
Investigation: CJ, LH, YM
Methodology: YS, SL
Project administration: JY
Supervision: SL
Writing – original draft: CJ, YS
Writing – review & editing: SL, JY
All authors contributed to the article and approved the submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Intervention details, including the adaptive training protocols and ongoing adjustments.
[DOCX File, 38 KB - mhealth_v14i1e93050_app1.docx]

Multimedia Appendix 2
Characteristics and changes in outcomes in the per-protocol population and patient adherence and adherence in the intention-to-treat population.
[DOCX File, 44 KB - mhealth_v14i1e93050_app2.docx]

XSL•FO
RenderX

Checklist 1
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist.
[PDF File, 3535 KB - mhealth_v14i1e93050_app3.pdf]

References

1. Sloan M, Premkumar A, Sheth NP. Projected volume of primary total joint arthroplasty in the U.S., 2014 to 2030. J Bone Joint Surg Am 2018 Sep 5;100(17):1455-1460. [doi: 10.2106/JBJS.17.01617] [Medline: 30180053]
2. Fascio E, Vitale JA, Sirtori P, Peretti G, Banfi G, Mangiavini L. Early virtual-reality-based home rehabilitation after total hip arthroplasty: a randomized controlled trial. J Clin Med 2022 Mar 22;11(7):1766. [doi: 10.3390/jcm11071766] [Medline: 35407373]
3. Dias Correia F, Nogueira A, Magalhães I, et al. Digital versus conventional rehabilitation after total hip arthroplasty: a single-center, parallel-group pilot study. JMIR Rehabil Assist Technol 2019 Jun 21;6(1):e14523. [doi: 10.2196/14523] [Medline: 31228176]
4. Patterson JT, Wu HH, Chung CC, Bendich I, Barry JJ, Bini SA. Wearable activity sensors and early pain after total joint arthroplasty. Arthroplast Today 2020 Mar;6(1):68-70. [doi: 10.1016/j.artd.2019.12.006] [Medline: 32211478]
5. Yang C, Shang L, Yao S, Ma J, Xu C. Cost, time savings and effectiveness of wearable devices for remote monitoring of patient rehabilitation after total knee arthroplasty: study protocol for a randomized controlled trial. J Orthop Surg Res 2023 Jun 27;18(1):461. [doi: 10.1186/s13018-023-03898-z] [Medline: 37370130]
6. Nelson MJ, Russell T, Crossley KM, Bourke M, McPhail SM. Cost-effectiveness of telerehabilitation versus traditional care after total hip replacement: a trial-based economic evaluation. J Telemed Telecare 2021 Jul;27(6):359-366. [doi: 10.1177/1357633X19869796] [Medline: 31530065]
7. Nilsdotter AK, Lohmander LS, Klässbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS)--validity and responsiveness in total hip replacement. BMC Musculoskelet Disord 2003 May 30;4(10):10. [doi: 10.1186/1471-2474-4-10] [Medline: 12777182]
8. Podsiadlo D, Richardson S. The timed “Up & Go”: a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991 Feb;39(2):142-148. [doi: 10.1111/j.1532-5415.1991.tb01616.x] [Medline: 1991946]
9. Berg KO, Wood-Dauphinee SL, Williams JI, Maki B. Measuring balance in the elderly: validation of an instrument. Can J Public Health 1992;83 Suppl 2:S7-11 [FREE Full text] [Medline: 1468055]
10. Ware JE Jr, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36). Med Care 1992 Jun;30(6):473-483. [doi: 10.1097/00005650-199206000-00002]
11. Philippot A, Dubois V, Lambrechts K, et al. Impact of physical exercise on depression and anxiety in adolescent inpatients: a randomized controlled trial. J Affect Disord 2022 Mar 15;301:145-153. [doi: 10.1016/j.jad.2022.01.011] [Medline: 35007642]
12. Bendich I, Tarity TD, Alpaugh K, Lyman S, Sculco PK, McLawhorn AS. Minimal clinically important difference (MCID) at one year postoperatively in aseptic revision total hip arthroplasty. J Arthroplasty 2022 Aug;37(8S):S954-S957. [doi: 10.1016/j.arth.2022.01.044] [Medline: 35091032]
13. Zhou R, Wu T, Huang L, Wang H, Lu S. Effectiveness of inertial measurement unit sensor–based feedback assistance in telerehabilitation of patients with diabetes after total knee arthroplasty: a randomized controlled trial. Telemedicine Reports 2024 Nov 1;5(1). [doi: 10.1089/tmr.2024.0008] [Medline: 39015181]
14. Liu X, Yang G, Xie W, et al. Efficacy of telerehabilitation for total knee arthroplasty: a meta-analysis based on randomized controlled trials combined with a bibliometric study. J Orthop Surg Res 2024 Dec 26;19(1):874. [doi: 10.1186/s13018-024-05381-9] [Medline: 39726029]
15. Molina-Garcia P, Mora-Traverso M, Prieto-Moreno R, Díaz-Vásquez A, Antony B, Ariza-Vega P. Effectiveness and cost-effectiveness of telerehabilitation for musculoskeletal disorders: a systematic review and meta-analysis. Ann Phys Rehabil Med 2024 Feb;67(1):101791. [doi: 10.1016/j.rehab.2023.101791] [Medline: 38128150]
16. Wu L, Li X, Hua L, Sun Q. Effects of technology-assisted rehabilitation for patients with hip arthroplasty: a meta-analysis. Medicine (Baltimore) 2023 Nov 10;102(45):e35921. [doi: 10.1097/MD.0000000000035921] [Medline: 37960720]
17. Zhou Z, Zhou X, Cui N, et al. Effectiveness of tele-rehabilitation after total hip replacement: a systematic review and meta-analysis of randomized controlled trials. Disabil Rehabil 2024 Oct;46(20):4611-4616. [doi: 10.1080/09638288.2023.2280070] [Medline: 37990882]

Abbreviations

ADL: activities of daily living
AE: adverse event
CONSORT: Consolidated Standards of Reporting Trials
HADS: Hospital Anxiety and Depression Scale
HADS-A: Hospital Anxiety and Depression Scale anxiety subscale

XSL•FO
RenderX

HADS-D: Hospital Anxiety and Depression Scale depression subscale
HOOS: Hip Disability and Osteoarthritis Outcome Score
ICER: incremental cost-effectiveness ratio
IMU: inertial measurement unit
ITT: intention-to-treat
PP: per-protocol
RCT: randomized controlled trial
SAE: serious adverse event
SF-36: 36-item Short Form Health Survey
THA: total hip arthroplasty
TUG: timed up and go

Edited by L Buis; submitted 07.Feb.2026; peer-reviewed by Anonymous, Y Yao; revised version received 02.Apr.2026; accepted 02.Apr.2026; published 21.Jul.2026.

Please cite as:
Jiang C, Shen Y, Huang L, Ma Y, Lu S, Yin J
The Efficiency and Cost-Effectiveness of Wearable Sensors in a Digital Physiotherapeutic Total Hip Arthroplasty–Specific Training System for Patients After Total Hip Arthroplasty: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e93050
URL: https://mhealth.jmir.org/2026/1/e93050
doi:10.2196/93050

© Chenyi Jiang, Yun Shen, Lihua Huang, Yanhong Ma, Shengdi Lu, Jimin Yin. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 21.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluation of the Impact of a Smartphone App on Adherence to an Exercise Program in People With Chronic Low Back Pain: Randomized Controlled Trial

Jean-Baptiste Lechauve[1], MSc; Lech Dobija[1], PT, PhD; Bruno Pereira[2], PhD; Maxime Grolier[1], MD; Mathilde Pelletier-Visa[1], MSc; Charlotte Lanhers[1], MD, MSc; Emmanuel Coudeyre[1], MD, PhD

[1]Physical Medicine and Rehabilitation Department, Centre Hospitalier Universitaire de Clermont-Ferrand, Service de Médecine Physique et de Réadaptation CHU Clermont-Ferrand, 58 rue Montalembert, Clermont-Ferrand, France

[2]Direction de la recherche clinique et de l'innovation, Centre Hospitalier Universitaire de Clermont-Ferrand, Clermont-ferrand, France

Corresponding Author:
Jean-Baptiste Lechauve, MSc
Physical Medicine and Rehabilitation Department, Centre Hospitalier Universitaire de Clermont-Ferrand, Service de Médecine Physique et de Réadaptation CHU Clermont-Ferrand, 58 rue Montalembert, Clermont-Ferrand, France

Abstract

Background: The benefits of multidisciplinary rehabilitation programs on pain and function in people with chronic low back pain (CLBP) are fairly well documented in the literature. However, these positive effects tend to fade over time due to low long-term patient adherence.

Objective: The study aimed to evaluate the impact of a smartphone app on adherence to an exercise program for people with CLBP at 6 months. The secondary aims were to assess the effectiveness of the app on pain, disability, fears, and beliefs related to physical activity, physical capacity, and qualitative adherence (correctness of exercise execution) at 6 months.

Methods: A total of 110 people with CLBP were included and randomized into 2 groups: 54 in the intervention group (IG) who received education on the use of the app in addition to usual care (a 3-week multidisciplinary rehabilitation program with self-management education) and 56 in the control group (CG) who received only usual care. Part B of the Exercise Adherence Rating Scale was used as the primary outcome to assess adherence to the 6-month exercise program. Secondary outcomes were pain (Numeric Rating Scale), disability (Oswestry Disability Index), barriers and facilitators to performing physical activity (Evaluation of Physical Activity Perception), physical capacity (battery of tests), and qualitative adherence (correctness of exercise execution). Statistical analyses were performed according to the intention-to-treat principle. A linear mixed model compared the primary end point between the groups at 6 months.

Results: A total of 71 of 110 participants (n=35 in the CG vs n=36 in the IG) were evaluated at 6 months. We did not observe any significant difference in the Exercise Adherence Rating Scale score between the CG and the IG (group difference 0.01, 95% CI −2.4 to 2.4; *P*=.97). The same findings were found for pain, disability, and barriers and facilitators to physical activity, except for the motivation criterion. The 6-Minute Walk Test (group difference [log] 0.06, 95% CI 0.01-0.12; *P*=.06), cycle ergometer (group difference 9.30, 95% CI 0.48-18.13; *P*=.04), Ito (group difference [log] 0.31, 95% CI 0.01-0.62; *P*=.04), plank (group difference [log] 0.29, 95% CI 0.03-0.54; *P*=.03), and qualitative adherence (IG mean 12.6, SD 1.3 vs CG mean 11.4, SD 1.4; *P*=.02) differed between the groups in favor of the IG. All outcomes improved from baseline to 6 months in the IG but not in the CG.

Conclusions: The smartphone app did not impact adherence to an exercise program at 6 months in individuals with CLBP. Similar results were found for pain and function. Nevertheless, the app could be a useful self-management tool in view of the positive effects on physical capacity and qualitative adherence.

Trial Registration: ClinicalTrials.gov NCT04264949; https://clinicaltrials.gov/study/NCT04264949?cond=%20NCT04264949&viewType=Card&rank=1

(JMIR Mhealth Uhealth 2026;14:e77736) doi:10.2196/77736

KEYWORDS

chronic low back pain; smartphone app; adherence; exercise; self-management

XSL•FO
RenderX

Introduction

Background

Low back pain is the leading cause of disability worldwide, affecting over 620 million people in 2020 [1]. Although this condition typically improves in most cases, 10% to 20% of individuals continue to experience pain beyond 3 months. Pain lasting more than 3 months is classified as chronic [1]. Chronic low back pain (CLBP) leads to significant functional limitations, impacting daily activities, productivity, and overall quality of life [2]. It imposes a heavy social burden and economic cost, representing a huge challenge to health care systems. Despite its high prevalence, effective management remains a challenge.

Global health organizations, including the World Health Organization and the Global Burden of Disease Study, emphasize the importance of nonpharmacological treatments for managing CLBP. Evidence-based recommendations prioritize physical rehabilitation, including tailored exercise programs, patient education with self-care strategies, and cognitive behavioral therapy (CBT) [1-3]. Exercise is consistently recommended as a first-line treatment to improve mobility, strength, and overall function while reducing pain. Tailored exercise programs for individuals with CLBP have demonstrated significant benefits on pain and function [4]. The National Institute for Health and Care Excellence guidelines highlight the importance of self-management, providing advice tailored to individual needs to help manage CLBP over the long term [5]. Self-management programs, incorporating multidisciplinary rehabilitation, have shown positive effects on return to work and physical activity at 6 and 12 months [6]. However, maintaining the necessary level of physical activity remains a challenge for most individuals. Despite these positive results, adherence to exercise programs is often suboptimal, with dropout rates ranging from 10% to 36% [7]. Adherence is defined as “the extent to which a person’s behavior corresponds with agreed recommendations from a healthcare provider” [8]. Medium-term and long-term follow-up data show that one- to two-thirds of individuals do not comply with the exercise recommendations [9,10], particularly for unsupervised home-based exercises [11-13]. Furthermore, adherence tends to decline over time [14]. Several studies have investigated the reasons behind poor adherence to prescribed exercises following multidisciplinary rehabilitation programs. Pain and exacerbation of low back pain (LBP) are the primary barriers, negatively influencing adherence [15]. Lack of time, weather conditions [16], and difficulty integrating physical activity into daily life have also been reported as barriers [15,17]. Factors that can improve long-term adherence to a home-based exercise program include the involvement of, and follow-up by, a physiotherapist during the multidisciplinary program, as well as a clear understanding of the exercises to be performed. These facilitators improve autonomy and increase the perception of self-efficacy [18,19]. It is also a theoretical component of behavior change used in therapeutic programs [20].

CBT has emerged as a highly effective intervention for CLBP, particularly when combined with exercise [21]. CBT is “a safe, evidence-based treatment for a range of mental and physical health problems, including chronic pain” [22]. The American Pain Society and the American College of Physicians recommend combining CBT with active functional restoration exercises within a multidisciplinary management program. This combined approach can enhance adherence and reduce disability and pain over time, providing a comprehensive strategy for managing CLBP.

Connected health devices, such as smartphones and tablets, are valuable tools that can be used to deliver CBT to promote behavior change and support the sustainable adoption of healthy behaviors. These devices provide personalized medical information, self-management tools, and remote monitoring by health care professionals, creating an optimal environment for patient care. Mobile health (mHealth) solutions, particularly through smartphones, offer a means for continuous support and follow-up [23], helping maintain the benefits of in-center care through engaging, interactive platforms [24]. With the growing prevalence of smartphones (70% of the population in 2021) and the rapid expansion of health apps (from 100,000 in 2014 to 350,000 in 2021) [25], mHealth facilitates remote rehabilitation, real-time data collection, and improved disease management. As personal and accessible devices [26], smartphones enhance patient engagement and understanding of health conditions, contributing to better self-management and health care access. The use of smartphone apps could reduce pain and disability [27,28] and improve quality of life [29] in the short term, but these effects have not been studied in the long term. However, some evidence regarding the use of smartphone apps for the self-management of CLBP is contradictory [30].

Objectives

The primary objective of this study was to assess the impact of a smartphone app on adherence to a home exercise program at 6 months in people with CLBP. The secondary aims were to assess the effectiveness of the app on pain, disability, fears, and beliefs relating to physical activity, physical capacity, and qualitative adherence (correctness of exercise execution) at 6 months.

Methods

Study Design

This was a single-center, prospective, controlled, cluster-randomized trial (trial registration: ClinicalTrials.gov: NCT04264949) with session as the unit of randomization. Sessions were performed 5 days per week for 3 weeks [31,32] and involved groups of 4 to 6 participants. Participants randomized to the intervention group (IG) participated in 3 one-hour sessions of education on how to use the app in addition to the conventional, multidisciplinary rehabilitation program and the self-management program. Those randomized to the control group (CG) participated only in the conventional multidisciplinary rehabilitation program and the self-management education program. Assessments were conducted at baseline, postprogram (3 wk), and at follow-up (6 mo). The trial protocol has been published elsewhere [33]. This trial is reported according to the CONSORT (Consolidated Standards of Reporting Trials) Statement for nonpharmacological trials [34] (Checklist 1).

XSL•FO
RenderX

Recruitment and Selection

Recruitment was conducted between March 2020 and June 2023. Following a medical consultation regarding their chronic low back pain, potential participants were invited to join a multidisciplinary rehabilitation program in our physical medicine and rehabilitation department. Participants who met the inclusion criteria (Textbox 1) and provided written consent after reading the information sheet were randomized by session into one of the 2 groups.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Adult participants (≥18 years old).
- Experiencing nonspecific chronic low back pain.
- Diagnosis confirmed according to the Haute Autorité Santé definition.

Included participants must have provided written informed consent.

Exclusion criteria

- Participants not meeting the Haute Autorité Santé diagnostic criteria.
- Participants with comprehension difficulties (questionnaire completion impossible).
- Participants with a medical contraindication to physical exercise.
- Participants under legal guardianship, curatorship, or protective supervision.
- Participants without a smartphone.

Randomization and Blinding

The unit of observation for the outcome was the participants. However, the session was the unit of randomization to avoid contamination bias. Individual randomization may lead to recruitment difficulties and feasibility and an increase in the number of individuals lost to follow-up in the CG. Participants in the same session were assigned to the same randomization group. An independent statistician performed the randomization and allocation using Stata v15 (StataCorp), taking into account the number of participants per session. The evaluator and the members of the rehabilitation team were blinded to group allocation. Participants were not blinded, as we had to explain the purpose of the study before randomization. The name of the app was not revealed to participants unless they were randomized into the IG or screened out of the study.

Interventions

Overview

Both groups participated in a conventional, multidisciplinary rehabilitation program and a self-management education program, which lasted 3 weeks (5 days a week). The only difference between the 2 groups was that the IG received the smartphone app “Mon Coach Dos” and participated in 3 one-hour training sessions on how to use it (Multimedia Appendix 1). The aim of the app was to improve adherence to the home exercise program, reinforcing the conventional care effect.

Self-Management Exercise Program Based on a Smartphone App (for the IG)

The app used was Mon Coach Dos and was developed in collaboration with the health care professionals from our physical medicine and rehabilitation department and the Thuasne group. This app aims to help individuals to better understand CLBP and enable them to self-manage their condition. It provides medical information on the pathology, messages regarding the benefits of physical activity, information on pain management, and a video of a physical exercise program. It is designed to change individuals’ representations of the pathology and their related behavior [19,35]. The Mon Coach Dos app was developed following work carried out in our department on an initial test app, “E-lombactifs,” which was evaluated by collecting opinions from people with CLBP and health care professionals through interviews (individual or focus group) [23].

The main aims of the self-management exercise program using the smartphone app are to increase individuals’ understanding of (1) the importance of practicing physical exercise; (2) when, where, and how to exercise; (3) how to adapt physical activity practices according to phenotype; and (4) how to integrate these practices into daily life over the long-term. Three educational training sessions on the use of the smartphone app were provided during the rehabilitation program (each Wednesday). Each session consisted of 1 hour of self-management education and physical exercise practice related to the content of the app. By the end of the program, each participant had developed a personalized exercise program with the assistance of the adapted physical activity specialist who led the sessions.

Conventional, Multidisciplinary Rehabilitation Program

The rehabilitation program lasted 3 weeks. Participants attended 5 days a week (Monday to Friday). Each day included 1 hour of physical therapy, 1 hour of occupational therapy, 1 hour of adapted physical activity, 1 hour of hydrotherapy, and 1 hour of self-management education (Multimedia Appendix 2).

Self-Management Education

This component of the program consisted of 6 different workshops to improve the self-management of CLBP and quality

XSL•FO
RenderX

of life, performed over 3 sessions (1 per week). The themes of the workshops were the anatomy and pathology of LBP, activities of daily living, nonpharmacological pain management (ie, relaxation, transcutaneous electrical nerve stimulation), pharmacological pain management, adapted physical activity, and return to work strategies.

Outcomes

At baseline, we collected sociodemographic data (age, sex, weight, height, education status, and socioprofessional category) and medical data (history of low back pain and treatments).

Primary Outcome

The primary outcome was the change in the Exercise Adherence Rating Scale (EARS) score at 6 months. The EARS is a self-administered questionnaire that measures adherence to a physical activity program and has good psychometric properties [36].

Secondary Outcomes

Pain intensity during the last 7 days was assessed on a Numeric Rating Scale. Self-declared functional ability was measured using the Oswestry Disability Index [37,38]. Barriers and facilitators to regular physical activity were assessed using the Evaluation of Physical Activity Perception (EPAP) score [39]. Aerobic capacity was evaluated using the 6-Minute Walk Test (6 MWT) [40] and a submaximal test on a cycle ergometer [41]. Muscle endurance was evaluated using the Shirado-Ito test for the trunk flexors [42], the Sorensen test for the erector spinae [43], and the wall-sit test for the lower limbs [44]. Lumbar mobility was assessed by measuring fingertip-to-floor distance [45]. Qualitative adherence was assessed by the good execution of 3 physical exercises carried out during treatment (squat, plank, and rowing) using a graded evaluation grid out of 15 points [46]. More details are provided in Multimedia Appendix 3.

Assessment Schedule

Study outcomes were collected at baseline, postprogram (3 wk), and at 6 months post randomization by the adapted physical activity specialist and the physician, both of whom were blinded to the group allocation.

Statistical Analysis

The sample size estimation for this pilot trial was determined according to the 2010 CONSORT Statement extension for randomized pilot and feasibility trials [47] and Cohen recommendations [48], which define effect size (ES) limits as small (ES: 0.2), medium (ES: 0.5), and large (ES: 0.8, “grossly perceptible and therefore large”). According to data reported in the literature and considering this study as a pilot, it seemed suitable to include 60 patients per group.

To achieve an ES of 0.8 at 6 months postrandomization with a type 1 error of 5% and statistical power of 90%, 33 participants are required per group. However, because of the randomization design, with session as a unit cluster of randomization, the sample size should be increased to take into account between- and within-session variability. More precisely, the assumption in randomized controlled trials that the outcome for an individual is completely unrelated to that of any other individual is violated in cluster randomized trials because individuals in any 1 cluster (session in our case) are more likely to respond in a similar manner. This similarity is known as the intraclass correlation coefficient (ICC). For an average of 5 participants per session and an ICC of 0.05, 38 participants were required in each group. Therefore, to account for losses to follow-up, we planned to include 120 individuals (n=60 per group).

All analyses were conducted before the randomization code was broken, in line with the International Conference on Harmonization Good Clinical Practice guidelines. Data storage and management were conducted according to international guidelines relevant to French institutions. All data were entered using an electronic case report form, and data accuracy was analyzed by the study data manager. Data quality control measures included queries to identify outliers and missing data.

Continuous parameters are presented as mean (SD) or median (IQR) according to the normality of the distribution (Shapiro-Wilk test).

Participant characteristics were summarized by randomization group to consider selection biases and lack of balance.

Statistical analysis was first conducted using the intention-to-treat approach. To prevent attrition bias, missing data were replaced using post hoc multiple imputation. A linear mixed model was used to compare the primary endpoint (EARS score) between randomized groups at 6 months post randomization. The randomization group was evaluated as a fixed effect and session as a random effect to consider between- and within-session variability. The normality of residuals obtained from this model was studied. The results concerning group differences were expressed as effect sizes and 95% CIs. The estimated ICC from the fitted model was reported.

Between-group comparisons for the other outcomes were performed using random-effects models, as specified for the primary analysis.

The random-effects models were also used to study longitudinal repeated data (baseline, after rehabilitation [15 days] and at 6 months post randomization), considering participant as the random effect in addition to session. The following fixed effects were studied: randomization group, evaluation time point, and their interactions.

Statistical analyses were performed using Stata v15 (StataCorp). The tests were 2-sided with the type 1 error set at 5%. The widths of CIs were not adjusted for multiplicity and therefore should not be used in place of hypothesis testing, except for the primary outcome analysis. Consequently, the results for secondary outcomes were considered exploratory and were interpreted as such. The results were expressed using absolute group differences (after log transformation when appropriate), effect sizes, and 95% CIs.

Ethical Considerations

This study was conducted in accordance with the ethical guidelines outlined in the Declaration of Helsinki. The study was approved by the medical ethics committee of Ile-de-France (Ile-de-France III 3740, November 5, 2019). All participants

XSL•FO
RenderX

received both verbal and written information regarding the aim of the study and the protocol. Written informed consent was obtained before inclusion and before any specific procedure was performed. Participant data were anonymized and securely stored to ensure confidentiality. Identifiable information was removed from the dataset, and only deidentified data were used for analysis. Data access was restricted to authorized researchers. Participants were not compensated.

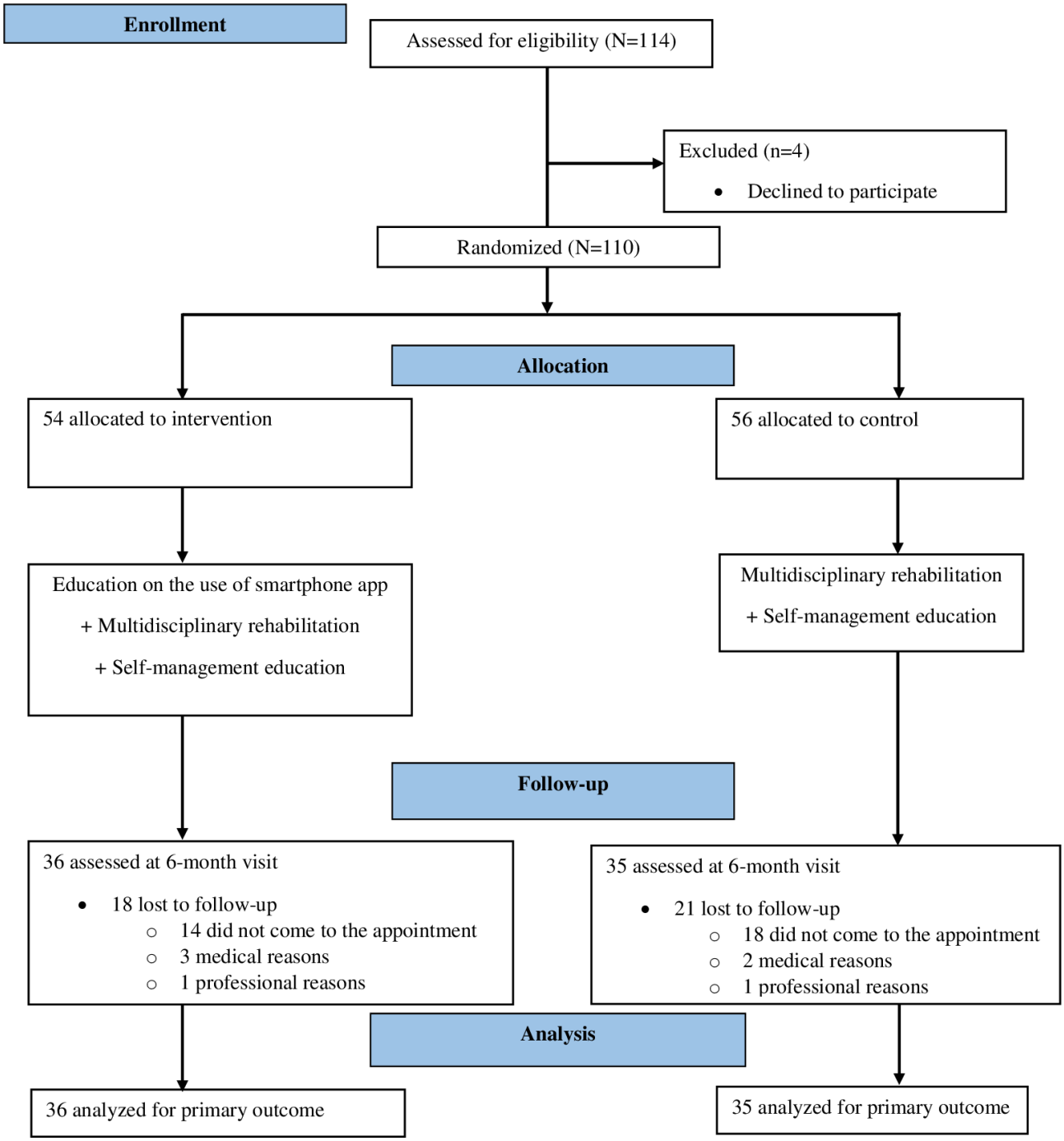
Results

Baseline Characteristics

In total, 110 individuals were enrolled in this study (Figure 1). They were randomized either to the app-based intervention group (n=54, 49.1%) or to the control group (n=56, 50.9%). The sociodemographic and clinical characteristics of the participants at baseline are presented in Table 1.

A total of 71 (64.5%) participants completed the study. A total of 39 (35.5%) participants dropped out between “postprogram” and “M6”: 18 (46.2%) in the IG versus 21 (53.8%) in the CG.

Figure 1. Schematic representation of the randomized controlled trial.



XSL•FO
RenderX

Table . Baseline demographic and clinical characteristics of the trial groups (N=110).

Characteristics	Intervention group (n=54)	Control group (n=56)
Age (y), mean (SD)	44.6 (9.2)	45.6 (9.7)
Gender, n (%)		
Female	24 (44.4)	23 (41.1)
Male	30 (55.6)	33 (58.9)
BMI (kg/m2), mean (SD)	26.8 (4.4)	26.9 (5.6)
Work situation, n (%)		
Active	1 (1.8)	3 (5.3)
Sick leave	48 (88.9)	50 (89.3)
Disability	3 (5.6)	2 (3.6)
Retired	2 (3.7)	0 (0)
Unemployed	0 (0)	1 (1.8)
Duration of sick leave (wk), mean (SD)	36.5 (53.5)	53.2 (58.1)
Duration of low back pain (wk), mean (SD)	369.3 (383.2)	251.6 (307)
Physiotherapy treatment: Yes, n (%)	46 (85.2)	53 (94.6)
EARS[a], mean (SD)	14.9 (6)	15.5 (5.4)
NRS[b], mean (SD)	44.4 (23.4)	49.3 (20.6)
ODI[c], mean (SD)	32.3 (12)	35.2 (14.2)
EPAP[d], mean (SD)	280.1 (84.6)	279.4 (81)
6MWT[e] (m), mean (SD)	473.6 (91.8)	477.6 (95.1)
Ergocycle test (W), mean (SD)	100.3 (28.4)	101.3 (31.2)
Ito (s), mean (SD)	63.7 (48.5)	81.8 (46.1)
Sorensen (s), mean (SD)	56.5 (41.5)	55.6 (39.3)
Wall-sit test (s), mean (SD)	60.36 (34.67)	51.34 (29.66)
Plank (s), mean (SD)	56.2 (35)	53.8 (30.2)
Fingertip-to-floor (cm), mean (SD)	18.2 (13.2)	19.9 (12.2)

[a]EARS: Exercise Adherence Rating Scale.

[b]NRS: Numeric Rating Scale.

[c]ODI: Oswestry Disability Index.

[d]EPAP: Evaluation of Physical Activity Perception.

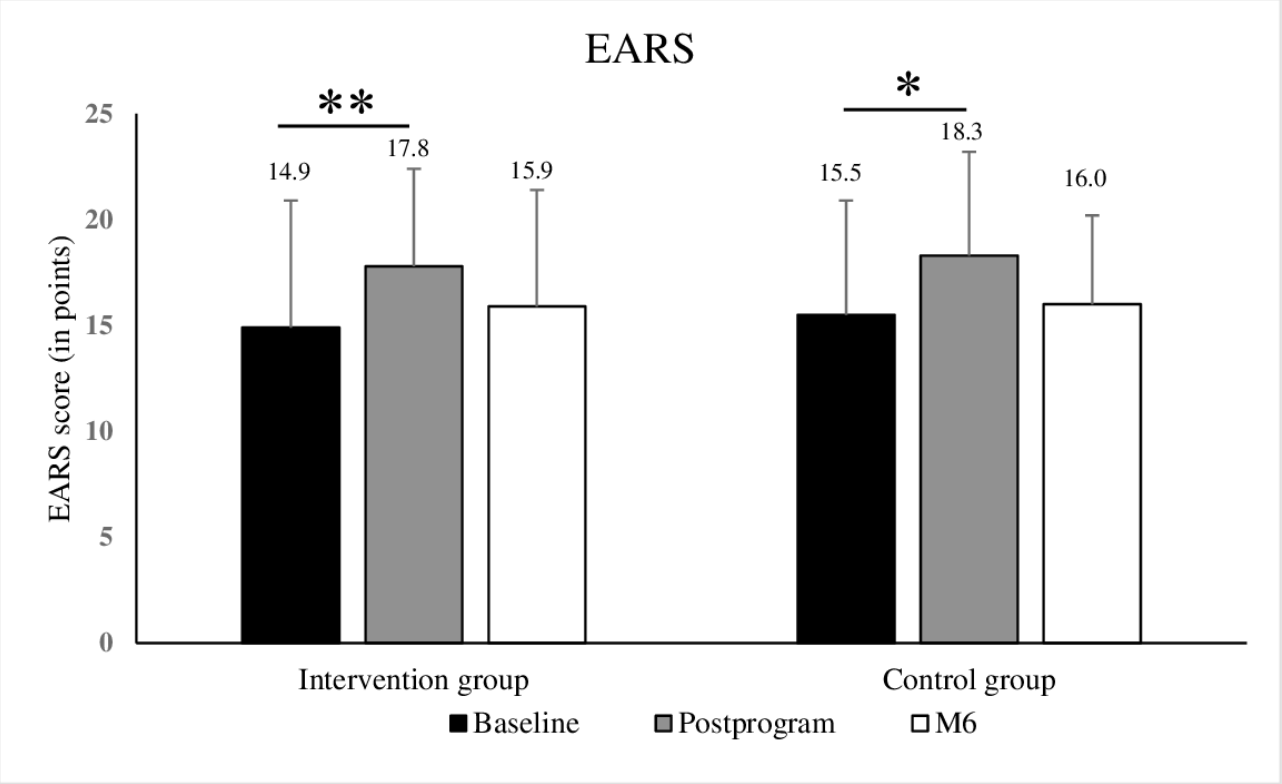
[e]6MWT: 6-Minute Walk Test.

Primary Outcome

The EARS score did not differ significantly between the 2 groups at 6 months (group difference 0.01, 95% CI −2.4 to 2.4; *P*=.97). However, the EARS score improved significantly in both groups postprogram compared to baseline (Figure 2). These results were confirmed by the analysis that took into account the session effect as a random effect (group difference 0.1, 95% CI −2.4 to 2.4; *P*=.97). The ICC related to the session effect was 2.7%.

XSL•FO
RenderX

Figure 2. Exercise Adherence Rating Scale (EARS) score in the 2 groups at baseline, postprogram, and at 6 months. M6: 6 months. *$P \leq .05$ and **$P \leq .01$.



Secondary Outcomes

All secondary outcomes are presented in Table 2.

XSL•FO
RenderX

Table . Outcomes at baseline, postprogram (3 wk), and at 6 months (M6).

Outcomes	Control group (n=35), mean (SD)	*P* value	Intervention group (n=36), mean (SD)	*P* value	Between-group, effect size (95% CI)	*P* value
EARS[a]						
Baseline	15.5 (5.4)	—[b]	14.9 (6)	—	—	—
Postprogram	18.3 (4.9)	.002	17.8 (4.6)	<.001	0.03 (−0.24 to 0.30)	.82
M6	16 (4.2)	.55	15.9 (5.46)	.55	0.00 (−0.29 to 0.30)	.97
6MWT[c]						
Baseline	477.6 (95.1)	—	473.6 (91.8)	—	—	—
Postprogram	527.3 (87.7)	<.001	555 (91.3)	<.001	0.32 (0.05 to 0.59)	.02
M6	500.8 (102.1)	<.001	545.11 (107)	<.001	0.28 (−0.01 to 0.58)	.06
Ergocycle						
Baseline	101.3 (31.2)	—	100.3 (28.4)	—	—	—
Postprogram	105.2 (32.8)	.19	110.1 (29.4)	<.001	0.22 (−0.05 to 0.49)	.11
M6	95.6 (34)	.14	107.7 (35.1)	.15	0.31 (0.02 to 0.61)	.04
Ito						
Baseline	81.8 (46.1)	—	63.7 (48.5)	—	—	—
Postprogram	98.4 (53.2)	.005	90.1 (53.7)	<.001	0.19 (−0.09 to 0.46)	.18
M6	88.8 (52.3)	.51	89.1 (52.4)	<.001	0.30 (−0.01 to 0.60)	.045
Sorensen						
Baseline	55.6 (39.3)	—	56.5 (41.5)	—	—	—
Postprogram	75.7 (47.4)	.005	85.9 (43.9)	<.001	0.22 (−0.06 to 0.49)	.12
M6	57.42 (40.2)	.43	75 (37.4)	.001	0.23 (−0.07 to 0.53)	.13
Wall-sit test						
Baseline	51.34 (29.66)	—	60.36 (34.67)	—	—	—
Postprogram	72.43 (42.91)	.001	80.81 (52.16)	.006	−0.09 (−0.36 to 0.18)	.50
M6	63.74 (39.84)	.006	77.11 (44.23)	.005	0.08 (−0.22 to 0.38)	.59
Plank						
Baseline	53.8 (30.2)	—	56.2 (35)	—	—	—
Postprogram	76.4 (38.4)	<.001	88.4 (48.7)	<.001	0.16 (−0.11 to 0.43)	.25
M6	65 (44.4)	.56	71.9 (45.1)	<.001	0.33 (0.04 to 0.63)	.03
FFD[d]						
Baseline	19.9 (12.2)	—	18.2 (13.2)	—	—	—
Postprogram	14.3 (12.2)	<.001	10.6 (12.4)	<.001	−0.17 (−0.44 to 0.10)	.22
M6	14 (11.8)	<.001	11.1 (13)	<.001	−0.01 (−0.31 to 0.28)	.93
Pain (NRS[e])						
Baseline	49.3 (20.6)	—	44.4 (23.4)	—	—	—
Postprogram	45.5 (21.4)	.08	38.2 (22.5)	.02	−0.10 (−0.37 to 0.17)	.47
M6	48.1 (21.6)	.92	38.3 (21.6)	.03	−0.25 (−0.55 to 0.05)	.10
ODI[f]						

XSL•FO
RenderX

Outcomes	Control group (n=35), mean (SD)	*P* value	Intervention group (n=36), mean (SD)	*P* value	Between-group, effect size (95% CI)	*P* value
Baseline	35.2 (14.2)	—	32.3 (12)	—	—	—
Postprogram	31.7 (14.3)	.005	27 (14.6)	<.001	−0.15 (−0.43 to 0.12)	.27
M6	28.6 (14.8)	<.001	23 (12)	<.001	−0.16 (−0.46 to 0.14)	.29
EPAP motivation						
Baseline	84.8 (20.2)	—	79.7 (20.6)	—	—	—
Postprogram	87.1 (17.1)	.37	85.9 (20.4)	.001	0.16 (−0.11 to 0.43)	.24
M6	82.1 (22)	.52	88.9 (15.3)	.003	0.34 (0.05 to 0.64)	.02
EPAP beliefs						
Baseline	50.3 (24.1)	—	58 (24.2)	—	—	—
Postprogram	62.7 (25.7)	<.001	69.7 (26.1)	<.001	−0.02 (−0.29 to 0.25)	.89
M6	63.8 (24.6)	.001	79.2 (19)	<.001	0.13 (−0.17 to 0.42)	.41
EPAP levers						
Baseline	69.8 (18.3)	—	74.7 (16.5)	—	—	—
Postprogram	70.9 (20.3)	.72	79.7 (15.5)	<.001	0.16 (−0.11 to 0.44)	.24
M6	67.4 (18.6)	.49	78.8 (15.4)	.02	0.25 (−0.05 to 0.55)	.10
EPAP barriers						
Baseline	74.5 (18.4)	—	67.7 (23.3)	—	—	—
Postprogram	76.7 (18.1)	.39	77.2 (22)	<.001	0.40 (0.13 to 0.67)	.004
M6	76 (18.9)	.37	77.4 (21.9)	.01	0.16 (−0.13 to 0.46)	.28

[a]EARS: Exercise Adherence Rating Scale.

[b]Not applicable.

[c]6MWT: 6-Minute Walk Test.

[d]FFD: fingertip-to-floor distance.

[e]NRS: Numeric Rating Scale.

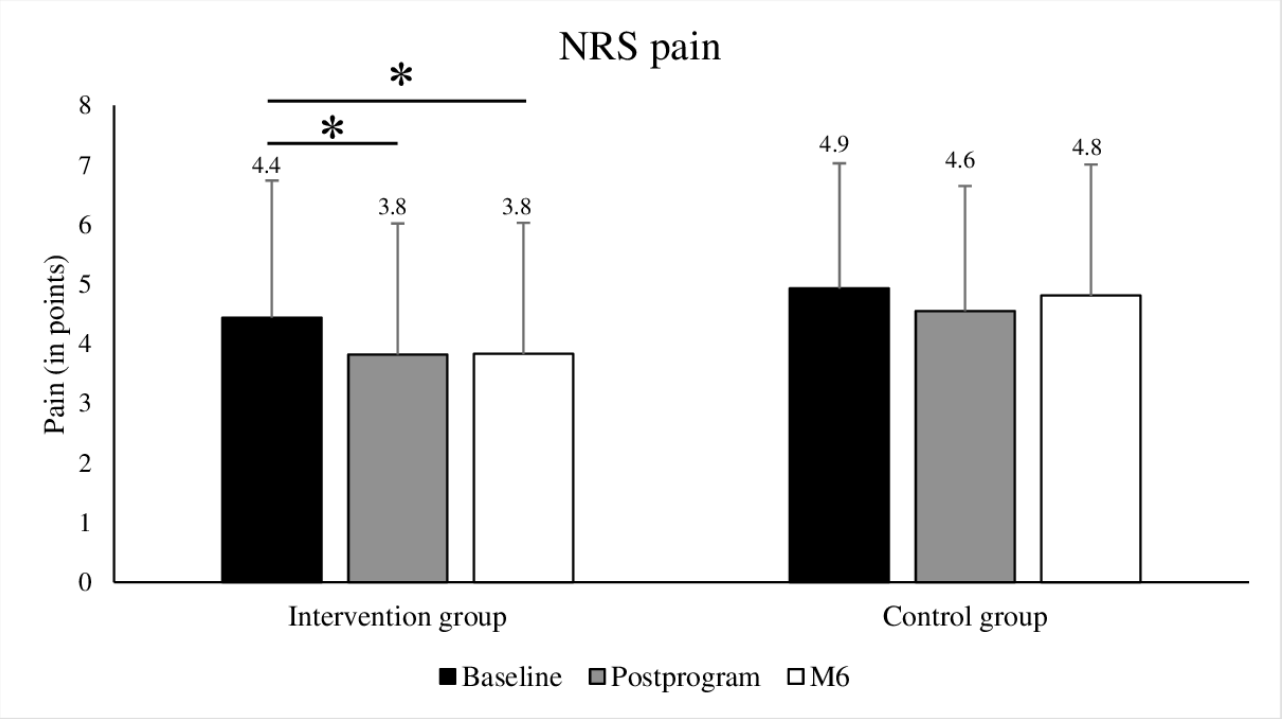
[f]ODI: Oswestry Disability Index.

Pain (Numeric Rating Scale)

Pain levels did not differ between the 2 groups at 6 months (group difference −6.58, 95% CI −14.46 to 1.29; *P*=.10). Pain reduced significantly in the IG postprogram compared to baseline (mean 3.8, SD 2.3 vs mean 4.4, SD 2.3; *P*=.02), as well as at 6 months compared to baseline (mean 3.8, SD 2.2 vs mean 4.4, SD 2.3; *P*=.03) (Figure 3). These differences were not found in the CG.

XSL•FO
RenderX

Figure 3. Pain score in the 2 groups at baseline, postprogram, and at 6 months. M6: 6 months; NRS: Numeric Rating Scale. *$P \leq .05$.



Disability (ODI)

Disability levels did not differ between the 2 groups at 6 months (group difference −2.27, 95% CI −6.50 to 1.96; *P*=.29). However, the ODI score improved significantly in both groups postprogram and at 6 months compared to baseline.

Barriers and Facilitators to Regular Physical Activity (EPAP)

The EPAP scores did not differ between the 2 groups at 6 months, except for the criterion “motivation” (group difference 8.38, 95% CI 1.12-15.65; *P*=.02). All EPAP criteria improved significantly (motivation, beliefs, barriers, facilitators) in the IG between baseline and 6 months. These differences were not found in the CG, except for the belief criterion.

Physical Capacity

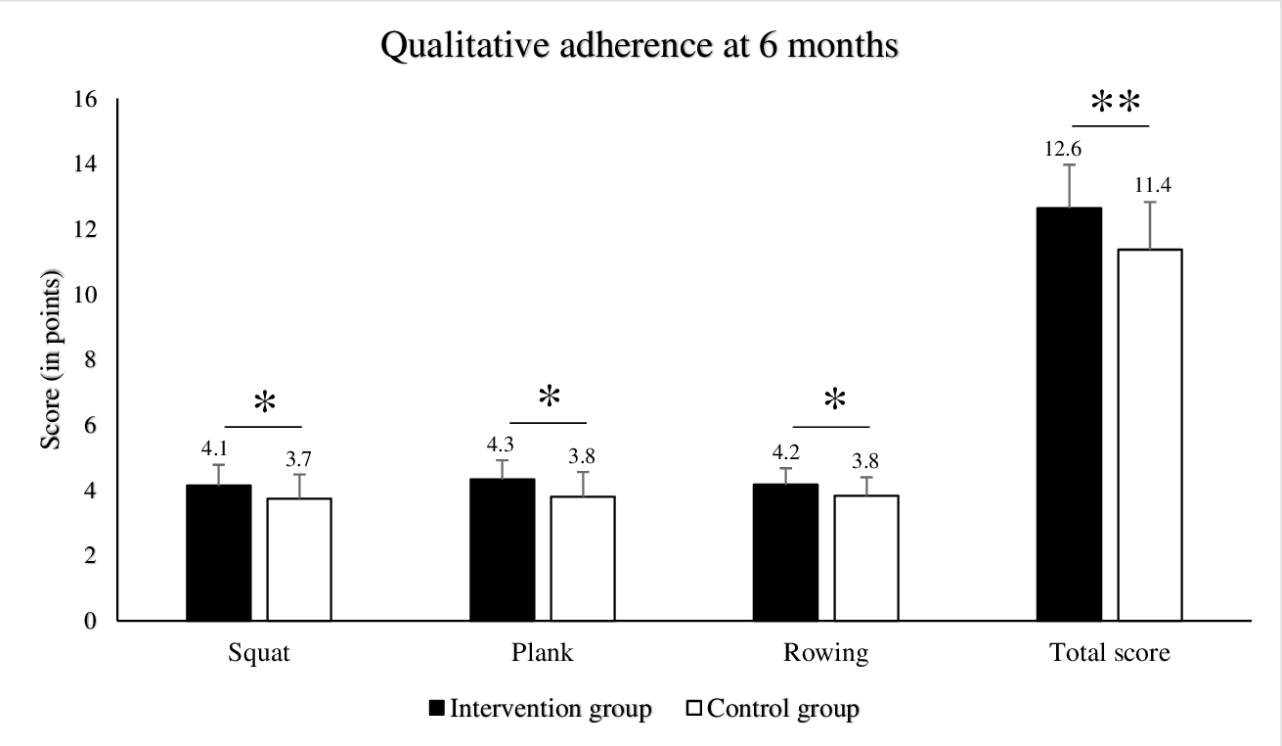
At 6 months, there was a small between-group difference in the 6MWT (group difference [log] 0.06, 95% CI 0.01-0.12; *P*=.06), the cycle ergometer test (group difference 9.30, 95% CI 0.48-18.13; *P*=.04), the Ito (group difference [log] 0.31, 95% CI 0.01-0.62; *P*=.04), and the plank test (group difference [log] 0.29, 95% CI 0.03-0.54; *P*=.03) in favor of the IG. Moreover, a significant improvement in all physical capacity tests was found postprogram compared to baseline in both groups, except for the cycle ergometer test in the CG. This significant improvement was maintained at 6 months in the IG for all tests, except for the cycle ergometer test, whereas in the CG, only the 6-MWT, wall-sit test, and the fingertip-to-floor distance test showed an improvement at 6 months compared to baseline (Table 2).

Qualitative Adherence

The total qualitative adherence score differed significantly between the groups in favor of the IG at 6 months (IG: 12.6, SD 1.3 vs CG: 11.4, SD 1.4; *P*=.02; Figure 4).

XSL•FO
RenderX

Figure 4. Qualitative adherence at 6 months in the 2 groups. **$P \leq .01$, *$P \leq .05$.



Discussion

To our knowledge, this is the first study to evaluate the impact of a smartphone app on adherence to a home exercise program at 6 months in people with CLBP. The lack of a significant difference in adherence between the 2 groups at 6 months suggests that the Mon Coach Dos app did not enhance adherence compared to usual care.

Primary Outcome

Although adherence did not differ between the groups, it improved significantly postprogram in both groups, with a return to the initial score at 6 months in both groups. This aligns with previous findings on medium- and long-term adherence, demonstrating a decline in adherence levels 3 months postprogram [49,50]. However, the study by Krkoska et al [51] showed very good adherence to a home exercise program over a period of 18 weeks. In that study, 25 out of 27 participants completed the program. Adherence was measured by the frequency of completing the program (once or twice a day) and was greater in the app group than in the diary group. However, the strong adherence of both groups to the programs can be explained by the considerable number of monitoring visits carried out during the program, making it possible to maintain a strong relationship with the therapist. In a qualitative study collecting patient expectations regarding mHealth, Palazzo et al [24] found that adherence could be increased by maintaining a relationship with the provider, for example, via telephone feedback, by offering attractive exercise programs, and by improving patient performance. In our study, we did not provide feedback to the participants during the 6-month follow-up period.

Adherence is a concept that is quite difficult to define and evaluate. Adherence to mHealth interventions has been assessed by different methods and defined in various ways. For example, some studies defined adherence as the number of complete active days of app use [52,53], the average time spent using the app [54], or the number of self-management plans completed using the app during the first 12 weeks after randomization [55]. In 2013, Beinart et al [10] highlighted the lack of standardized measures for adherence to prescribed home exercises. The absence of a standardized and valid tool to quantify adherence could compromise research on the effectiveness of tools designed to improve adherence [56]. There is currently no “gold standard” [57] nor a consensus for measuring patient adherence to self-exercise programs [58]. The development of a validated measure of adherence should be a priority for the research community, as it could lead to a better understanding of the multitude of factors that affect adherence to home exercises in people with LBP [10]. Furthermore, adherence to an exercise program can be influenced by the patient’s perception and motivation and the therapist’s expectations, which further complicates its objective measurement [59]. In response to this observation, Newman-Beinart et al [36] developed the first self-administered questionnaire to assess adherence to a physical activity program, the EARS questionnaire. This questionnaire has been used and evaluated in several languages (eg, Brazilian, Danish, Nepali) and across various pathologies, and it has been shown to have good reliability and validity [60-62]. This is why we used the French version of the EARS questionnaire to measure our primary outcome. We used Part B of the questionnaire, which quantitatively measures adherence to the exercise program. Contrary to our expectation, there was no between-group difference in the EARS score. However, this scale has already been questioned by Arensman et al [63], who stated that the EARS only focuses on the quantitative properties

XSL•FO
RenderX

of adherence and not on the quality of performance of the home-based exercise program pre-established by the health care professional. For this reason, they are now developing an instrument to measure adherence in terms of frequency, intensity, and quality of performance of home-based exercise programs [63]. In our study, we were not able to collect these quantitative data (log-ins, video views, exercise completions) due to a technical problem with the app on data recovery.

Secondary Outcomes

Pain or Disability

Neither pain nor disability levels differed between the groups at 6 months, in contrast with the findings of Chaabra et al [27]. A systematic review found that the simultaneous use of mHealth and usual care interventions was more effective than usual care in reducing pain intensity and disability in people with LBP [64]. Furthermore, they found that telephone calls enhanced improvements in pain and disability, suggesting that telephone calls or the provision of sensitive feedback devices enhance the effectiveness of mHealth interventions on pain intensity and disability. They could improve patient adherence and self-management [24,51]. We found a significant, long-term (6 mo) reduction in pain and disability score in the IG, a finding that has been previously reported for the short term (3 mo) [27,28,65]. The lack of a between-group difference in pain despite the significant reduction in pain at 6 months only in the IG and not in the CG could be explained by the fact that the CG received an intervention (a 3-week multidisciplinary program, including a therapeutic patient education program). In addition, we did not provide feedback or telephone calls to the IG. A scoping review found similar results regarding the effectiveness of mHealth apps on pain and disability but concluded that the generalizability of the results was limited due to the heterogeneity of the characterization of pain by the included participants and the different intervention durations [66].

Barriers and Facilitators to Regular Physical Activity (EPAP)

We found no difference between the groups in the EPAP criteria at 6 months, except for the “motivation” criterion, which could be explained by the fact that the smartphone app included videos of the exercises. However, all EPAP criteria (increased facilitators and decreased barriers) improved at 6 months in the IG, but not in the CG. Therefore, access to this app and its content appeared to help dispel false beliefs about physical activity and LBP. Participants might have been better informed about their pathology and gained a more precise understanding of its mechanisms. This process of clarification reduces the barriers often associated with low back pain and physical exercise [15].

Physical Capacity

Few studies have evaluated the long-term effectiveness of health apps on physical capacity. Our study found a significant improvement in all physical parameters at 6 months in the IG, but not in the CG, except for the 6MWT and FFT. Moreover, the significant between-group differences at 6 months in the 6MWT, cycle ergometer, Ito, and plank tests highlight the positive impact of the app on maintaining physical capacity in the long term. The lack of significant improvement in the CG, except for the specific tests mentioned, may also suggest that the mHealth app provided added value in terms of motivation or organization of exercise practice [24], facilitating progress in the IG.

Qualitative Adherence

The IG group executed the exercises (squat, plank, rowing) more correctly than the CG at 6 months. This could be attributed to the videos provided in the app that demonstrated the correct movements to perform. These videos offered participants a reassuring visual model, helping them gain confidence in their ability to perform the exercises correctly, which could have previously been a barrier to exercise practice. Therefore, the app appears to be a useful, motivational tool to promote continued physical activity, as its dynamic content instills user confidence in the correct execution of the movements performed, reducing the fear of injury and exacerbation of pain [15]. The app also allows users to be more independent in performing the movements, potentially leading to better qualitative adherence [67]. The qualitative aspect of adherence is important to consider when defining and evaluating adherence, in alignment with recommendations regarding the quality of exercise performance [63].

Relationship Between Adherence and Secondary Outcomes

It is surprising that, although quantitative adherence at 6 months (EARS) did not differ between the groups, the secondary outcomes (pain, disability, and physical capacity) only improved in the IG. Another study found no association between adherence to home-based exercise recommendations and changes in clinical outcomes in people with LBP [68]. These findings suggest that the relationship between home-based exercise adherence and treatment outcomes is complex. Further research is needed to better understand this relationship.

Limitations

The primary limitation of our study, given the existing literature, was the lack of quantitative data on the use of the app by the users (number, frequency, duration of connections to the app), which would have allowed us to be more precise in terms of adherence and to correlate these data with the EARS score.

The metrological properties of the EARS appear limited as only 6 questions were used to assess adherence [36]. Moreover, some of these questions appear to be similar. Furthermore, the EARS does not evaluate qualitative adherence.

Furthermore, our control group did not represent a true nonintervention, as the participants in the CG were provided with a conventional program, along with a patient education program, which simultaneously offered important information to help patients better manage their condition on a daily basis. The control group was not a “true” control, as seen in the previously cited studies.

Finally, although the number of dropouts during the protocol (39/110) indicates some challenges in sustaining participant engagement over a long-term process [65], the impact on the study’s statistical power is likely limited. First, in the final

XSL•FO
RenderX

analysis, the observed ICC was 0.027, lower than the assumed 0.05, suggesting that the inflation applied in the sample size calculation was conservative. Moreover, the initial sample size calculation indicated that 38 participants per group were required. After inflating this number to account for potential missing data, a target of 60 participants per group was set. Although the final analysis included 35 versus 36 participants, this remains reasonably close to the initial requirement and is therefore unlikely to have markedly compromised statistical power.

Conclusions

The Mon Coach Dos smartphone app did not improve adherence to an exercise program at 6 months in individuals with CLBP. Similarly, it did not lead to improved outcomes in pain, function, and physical capacity at 6 months. However, the significant improvement in pain, function, and physical parameters between baseline and 6 months in the IG, as well as better qualitative adherence, highlights the usefulness of such a tool in empowering individuals in the daily self-management of CLBP.

Acknowledgments

The authors would like to thank the entire Physical Medicine and Rehabilitation department team for their direct or indirect support in this work.

Funding

This research protocol was funded by the University Hospital of Clermont-Ferrand (AOI funding DGA/AB/MFC 2019-8). No artificial intelligence tools were used in the preparation of this manuscript.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to ethical and legal restrictions related to participant confidentiality and the terms of approval by the ethics committee but are available from the corresponding author upon reasonable request.

Authors' Contributions

JBL designed the protocol, trained the study personnel, maintained the study master record, monitored data collection, and cleaned and analyzed the data. EC, MP, and JBL secured funding, ethics approval, and the research contract. JBL is the guarantor, oversaw participant recruitment, and was the principal investigator responsible for all aspects of the trial. EC was the supervising investigator and oversaw the protocol at the trial site. BP wrote the statistical plan and oversaw the statistical analysis. JBL and MP maintained the study master record, monitored data collection, and cleaned and analyzed the data. EC and MP negotiated the research contract and managed the trial funds. JBL, LD, BP, MP, CL, and EC made intellectual contributions to the paper. JBL, LD, and EC drafted and revised the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study design.
[PNG File, 49 KB - mhealth_v14i1e77736_app1.png]

Multimedia Appendix 2
Multidisciplinary rehabilitation program.
[PNG File, 289 KB - mhealth_v14i1e77736_app2.png]

Multimedia Appendix 3
Graded evaluation grid of qualitative adherence.
[PNG File, 217 KB - mhealth_v14i1e77736_app3.png]

Checklist 1
CONSORT checklist.
[PDF File, 771 KB - mhealth_v14i1e77736_app4.pdf]

References

XSL•FO
RenderX

1. GBD 2021 Low Back Pain Collaborators. Global, regional, and national burden of low back pain, 1990-2020, its attributable risk factors, and projections to 2050: a systematic analysis of the global burden of disease study 2021. Lancet Rheumatol 2023 Jun;5(6):e316-e329. [doi: 10.1016/S2665-9913(23)00098-X] [Medline: 37273833]
2. Hartvigsen J, Hancock MJ, Kongsted A, et al. What low back pain is and why we need to pay attention. The Lancet 2018 Jun;391(10137):2356-2367. [doi: 10.1016/S0140-6736(18)30480-X] [Medline: 29573870]
3. Buchbinder R, van Tulder M, Öberg B, et al. Low back pain: a call for action. The Lancet 2018 Jun;391(10137):2384-2388. [doi: 10.1016/S0140-6736(18)30488-4] [Medline: 29573871]
4. Chou R, Deyo R, Friedly J, et al. Nonpharmacologic therapies for low back pain: a systematic review for an American College of Physicians clinical practice guideline. Ann Intern Med 2017 Apr 4;166(7):493-505. [doi: 10.7326/M16-2459] [Medline: 28192793]
5. Low Back Pain and Sciatica in Over 16s: Assessment and Management: National Institute for Health and Care Excellence (NICE); 2020.
6. Tavares Figueiredo I, Dupeyron A, Tran B, et al. Educational self-care objectives within a functional spine restoration program. Retrospective study of 104 patients. Ann Phys Rehabil Med 2016 Dec;59(5-6):289-293. [doi: 10.1016/j.rehab.2016.03.006] [Medline: 27157543]
7. Blanchard CM, Courneya KS, Rodgers WM, et al. Is the theory of planned behavior a useful framework for understanding exercise adherence during phase II cardiac rehabilitation? J Cardiopulm Rehabil 2003;23(1):29-39. [doi: 10.1097/00008483-200301000-00007] [Medline: 12576910]
8. Adherence to long-term therapies: evidence for action. : World Health Organization; 2003 URL: https://www.paho.org/sites/default/files/WHO-Adherence-Long-Term-Therapies-Eng-2003.pdf [accessed 2026-05-27]
9. Sluijs EM, Kok GJ, van der Zee J. Correlates of exercise compliance in physical therapy. Phys Ther 1993 Nov;73(11):771-782. [doi: 10.1093/ptj/73.11.771] [Medline: 8234458]
10. Beinart NA, Goodchild CE, Weinman JA, Ayis S, Godfrey EL. Individual and intervention-related factors associated with adherence to home exercise in chronic low back pain: a systematic review. Spine J 2013 Dec;13(12):1940-1950. [doi: 10.1016/j.spinee.2013.08.027] [Medline: 24169445]
11. Reilly K, Lovejoy B, Williams R, Roth H. Differences between a supervised and independent strength and conditioning program with chronic low back syndromes. J Occup Med 1989 Jun;31(6):547-550. [doi: 10.1097/00043764-198906000-00012] [Medline: 2525182]
12. Sluijs EM, Knibbe JJ. Patient compliance with exercise: different theoretical approaches to short-term and long-term compliance. Patient Educ Couns 1991 Jun;17(3):191-204. [doi: 10.1016/0738-3991(91)90060-I]
13. Nelson BW, O'Reilly E, Miller M, Hogan M, Wegner JA, Kelly C. The clinical effects of intensive, specific exercise on chronic low back pain: a controlled study of 895 consecutive patients with 1-year follow up. Orthopedics 1995 Oct;18(10):971-981. [doi: 10.3928/0147-7447-19951001-05] [Medline: 8584467]
14. Genêt F, Poiraudeau S, Revel M. Effectiveness and compliance to a center-based short rehabilitation program with a home-based program for chronic low back pain [Article in French]. Ann Readapt Med Phys 2002 Jun;45(6):265-272. [doi: 10.1016/s0168-6054(02)00214-3] [Medline: 12076853]
15. Boutevillain L, Dupeyron A, Rouch C, Richard E, Coudeyre E. Facilitators and barriers to physical activity in people with chronic low back pain: a qualitative study. PLoS One 2017;12(7):e0179826. [doi: 10.1371/journal.pone.0179826] [Medline: 28742820]
16. Van Koppen LH, Zandwijk PJJ, Van Mameren H, Mesters I, Winkens B, De Bie RA. Patients' adherence to a walking programme for non-specific low back pain. Eur J Physiother 2016 Apr 2;18(2):103-115. [doi: 10.3109/21679169.2015.1127419]
17. Bassett S. Measuring patient adherence to physiotherapy. J Nov Physiother 2012;2(7). [doi: 10.4172/2165-7025.1000e124]
18. Bachmann C, Oesch P, Bachmann S. Recommendations for improving adherence to home-based exercise: a systematic review [Article in German]. Phys Med Rehab Kuror 2018 Jan;28(1):20-31. [doi: 10.1055/s-0043-120527]
19. Saner J, Bergman EM, de Bie RA, Sieben JM. Low back pain patients' perspectives on long-term adherence to home-based exercise programmes in physiotherapy. Musculoskelet Sci Pract 2018 Dec;38:77-82. [doi: 10.1016/j.msksp.2018.09.002] [Medline: 30317015]
20. Bareyre L, Gay C, Coste N, Bonnin A, Pereira B, Coudeyre E. French validation of the Arthritis Self-Efficacy Scale and further psychometric properties exploration among 168 people with osteoarthritis. Clin Rehabil 2019 Mar;33(3):546-556. [doi: 10.1177/0269215518809811] [Medline: 30409034]
21. Yang J, Lo WLA, Zheng F, Cheng X, Yu Q, Wang C. Evaluation of cognitive behavioral therapy on improving pain, fear avoidance, and self-efficacy in patients with chronic low back pain: a systematic review and meta-analysis. Pain Res Manag 2022;2022:4276175. [doi: 10.1155/2022/4276175] [Medline: 35345623]
22. Chou R, Huffman LH, American Pain Society, American College of Physicians. Nonpharmacologic therapies for acute and chronic low back pain: a review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. Ann Intern Med 2007 Oct 2;147(7):492-504. [doi: 10.7326/0003-4819-147-7-200710020-00007] [Medline: 17909210]

XSL•FO
RenderX

23. Arefyev A, Lechauve JB, Gay C, et al. Using connected objects to favour patients' adherence to physical activity [Article in French]. Rev Infirm 2018 Jan;67(237):38-39. [doi: 10.1016/j.revinf.2017.11.018] [Medline: 29331193]
24. Palazzo C, Klinger E, Dorner V, et al. Barriers to home-based exercise program adherence with chronic low back pain: patient expectations regarding new technologies. Ann Phys Rehabil Med 2016 Apr;59(2):107-113. [doi: 10.1016/j.rehab.2016.01.009] [Medline: 27050664]
25. Borst F, Reuss-Borst M, Boschmann J, Schwarz P. Can mobile-health applications contribute to long-term increase in physical activity after medical rehabilitation?—A pilot-study. PLOS Digit Health 2023 Oct;2(10):e0000359. [doi: 10.1371/journal.pdig.0000359] [Medline: 37844024]
26. Adler E, Fogg BJ. Texting 4 Health: A Simple, Powerful Way to Change Lives: Captology Media; 2009.
27. Chhabra HS, Sharma S, Verma S. Smartphone app in self-management of chronic low back pain: a randomized controlled trial. Eur Spine J 2018 Nov;27(11):2862-2874. [doi: 10.1007/s00586-018-5788-5] [Medline: 30324496]
28. Du S, Liu W, Cai S, Hu Y, Dong J. The efficacy of e-health in the self-management of chronic low back pain: a meta analysis. Int J Nurs Stud 2020 Jun;106:103507. [doi: 10.1016/j.ijnurstu.2019.103507] [Medline: 32320936]
29. Shebib R, Bailey JF, Smittenaar P, Perez DA, Mecklenburg G, Hunter S. Randomized controlled trial of a 12-week digital care program in improving low back pain. NPJ Digit Med 2019;2:1. [doi: 10.1038/s41746-018-0076-7] [Medline: 31304351]
30. Didyk C, Lewis LK, Lange B. Effectiveness of smartphone apps for the self-management of low back pain in adults: a systematic review. Disabil Rehabil 2022 Dec;44(25):7781-7790. [doi: 10.1080/09638288.2021.2005161] [Medline: 34854335]
31. Caby I, Olivier N, Janik F, Vanvelcenaher J, Pelayo P. A controlled and retrospective study of 144 chronic low back pain patients to evaluate the effectiveness of an intensive functional restoration program in France. Healthcare (Basel) 2016 Apr 27;4(2):23. [doi: 10.3390/healthcare4020023] [Medline: 27417611]
32. Poiraudeau S, Rannou F, Revel M. Functional restoration programs for low back pain: a systematic review. Ann Readapt Med Phys 2007 Jul;50(6):425-429. [doi: 10.1016/j.annrmp.2007.04.009] [Medline: 17512079]
33. Lechauve JB, Dobija L, Pereira B, et al. Evaluation of the impact of a smartphone application on adherence to home exercise program for people with chronic low back pain: research protocol for a pilot randomised controlled trial. BMJ Open 2023 Mar 24;13(3):e062290. [doi: 10.1136/bmjopen-2022-062290] [Medline: 36963800]
34. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P, CONSORT Group. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. Ann Intern Med 2008 Feb 19;148(4):295-309. [doi: 10.7326/0003-4819-148-4-200802190-00008] [Medline: 18283207]
35. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
36. Newman-Beinart NA, Norton S, Dowling D, et al. The development and initial psychometric evaluation of a measure assessing adherence to prescribed exercise: the Exercise Adherence Rating Scale (EARS). Physiotherapy 2017 Jun;103(2):180-185. [doi: 10.1016/j.physio.2016.11.001] [Medline: 27913064]
37. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. Physiotherapy 1980 Aug;66(8):271-273. [Medline: 6450426]
38. Vogler D, Paillex R, Norberg M, de Goumoëns P, Cabri J. Cross-cultural validation of the Oswestry Disability Index in French [Article in French]. Ann Readapt Med Phys 2008 Jun;51(5):379-385. [doi: 10.1016/j.annrmp.2008.03.006]
39. Coste N, Gay C, Gerbaud L, Aucair C, Coudeyre E. Development and validation of a questionnaire assessing the perception of physical activity (EPAP) in knee osteoarthritis people. Ann Phys Rehabil Med 2018 Jul;61:e127-e128. [doi: 10.1016/j.rehab.2018.05.281]
40. Butland RJ, Pang J, Gross ER, Woodcock AA, Geddes DM. Two-, six-, and 12-minute walking tests in respiratory disease. Br Med J (Clin Res Ed) 1982 May 29;284(6329):1607-1608. [doi: 10.1136/bmj.284.6329.1607] [Medline: 6805625]
41. Akalan C, Robergs R, Kravitz L. Prediction of VO2max from an individualized submaximal cycle ergometer protocol. J Exerc Physiol 2008;11(2):1-17 [FREE Full text]
42. Ito T, Shirado O, Suzuki H, Takahashi M, Kaneda K, Strax TE. Lumbar trunk muscle endurance testing: an inexpensive alternative to a machine for evaluation. Arch Phys Med Rehabil 1996 Jan;77(1):75-79. [doi: 10.1016/s0003-9993(96)90224-5] [Medline: 8554479]
43. Biering-Sørensen F. Physical measurements as risk indicators for low-back trouble over a one-year period. Spine (Phila Pa 1976) 1984 Mar;9(2):106-119. [doi: 10.1097/00007632-198403000-00002] [Medline: 6233709]
44. McQuade KJ, Turner JA, Buchner DM. Physical fitness and chronic low back pain. An analysis of the relationships among fitness, functional limitations, and depression. Clin Orthop Relat Res 1988 Aug(233):198-204. [Medline: 2969788]
45. Tousignant M, Poulin L, Marchand S, Viau A, Place C. The Modified-Modified Schober Test for range of motion assessment of lumbar flexion in patients with low back pain: a study of criterion validity, intra- and inter-rater reliability and minimum metrically detectable change. Disabil Rehabil 2005 May 20;27(10):553-559. [doi: 10.1080/09638280400018411] [Medline: 16019864]

XSL•FO
RenderX

46. Hermet A, Roren A, Lefevre-Colau MM, et al. Agreement among physiotherapists in assessing patient performance of exercises for low-back pain. BMC Musculoskelet Disord 2018 Jul 27;19(1):265. [doi: 10.1186/s12891-018-2173-9] [Medline: 30053807]
47. Eldridge SM, Chan CL, Campbell MJ, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016 Oct 24;355:i5239. [doi: 10.1136/bmj.i5239] [Medline: 27777223]
48. Cohen J. Statistical Power Analysis for the Behavioral Sciences: Academic Press; 2013.
49. Kolt GS, McEvoy JF. Adherence to rehabilitation in patients with low back pain. Man Ther 2003 May;8(2):110-116. [doi: 10.1016/s1356-689x(02)00156-x] [Medline: 12890439]
50. Hammer C, Degerfeldt L, Denison E. Mechanical diagnosis and therapy in back pain: compliance and social cognitive theory. Adv Physiother 2007 Jan;9(4):190-197. [doi: 10.1080/14038190701433241]
51. Krkoska P, Vlazna D, Sladeckova M, et al. Adherence and effect of home-based rehabilitation with telemonitoring support in patients with chronic non-specific low back pain: a pilot study. Int J Environ Res Public Health 2023 Jan 13;20(2):1504. [doi: 10.3390/ijerph20021504] [Medline: 36674258]
52. Toelle TR, Utpadel-Fischler DA, Haas KK, Priebe JA. App-based multidisciplinary back pain treatment versus combined physiotherapy plus online education: a randomized controlled trial. NPJ Digit Med 2019;2(1):34. [doi: 10.1038/s41746-019-0109-x] [Medline: 31304380]
53. Weise H, Zenner B, Schmiedchen B, et al. The effect of an app-based home exercise program on self-reported pain intensity in unspecific and degenerative back pain: pragmatic open-label randomized controlled trial. J Med Internet Res 2022 Oct 28;24(10):e41899. [doi: 10.2196/41899] [Medline: 36215327]
54. Sandal LF, Stochkendahl MJ, Svendsen MJ, et al. An app-delivered self-management program for people with low back pain: protocol for the selfBACK randomized controlled trial. JMIR Res Protoc 2019 Dec 3;8(12):e14720. [doi: 10.2196/14720] [Medline: 31793897]
55. Almhdawi KA, Obeidat DS, Kanaan SF, Oteir AO, Mansour ZM, Alrabbaei H. Efficacy of an innovative smartphone application for office workers with chronic non-specific low back pain: a pilot randomized controlled trial. Clin Rehabil 2020 Oct;34(10):1282-1291. [doi: 10.1177/0269215520937757] [Medline: 32602362]
56. Jordan JL, Holden MA, Mason EE, Foster NE. Interventions to improve adherence to exercise for chronic musculoskeletal pain in adults. Cochrane Database Syst Rev 2010 Jan 20;2010(1):CD005956. [doi: 10.1002/14651858.CD005956.pub2] [Medline: 20091582]
57. Peek K, Sanson-Fisher R, Mackenzie L, Carey M. Interventions to aid patient adherence to physiotherapist prescribed self-management strategies: a systematic review. Physiotherapy 2016 Jun;102(2):127-135. [doi: 10.1016/j.physio.2015.10.003] [Medline: 26821954]
58. Hall AM, Kamper SJ, Hernon M, et al. Measurement tools for adherence to non-pharmacologic self-management treatment for chronic musculoskeletal conditions: a systematic review. Arch Phys Med Rehabil 2015 Mar;96(3):552-562. [doi: 10.1016/j.apmr.2014.07.405] [Medline: 25108098]
59. Peek K, Carey M, Mackenzie L, Sanson-Fisher R. Patient adherence to an exercise program for chronic low back pain measured by patient-report, physiotherapist-perception and observational data. Physiother Theory Pract 2019 Dec;35(12):1304-1313. [doi: 10.1080/09593985.2018.1474402] [Medline: 29771180]
60. de Lira MR, de Oliveira AS, França RA, Pereira AC, Godfrey EL, Chaves TC. The Brazilian Portuguese version of the Exercise Adherence Rating Scale (EARS-Br) showed acceptable reliability, validity and responsiveness in chronic low back pain. BMC Musculoskelet Disord 2020 May 12;21(1):294. [doi: 10.1186/s12891-020-03308-z] [Medline: 32398068]
61. Jacobsen JS, Nielsen RO, Godfrey EL. Translation and cross-cultural adaptation of the Exercise Adherence Rating Scale (EARS) into Danish. Transl Sports Med 2022;2022:4547350. [doi: 10.1155/2022/4547350] [Medline: 38655167]
62. Adhikari SP, Dev R, Shrestha JN. Cross-cultural adaptation, validity, and reliability of the Nepali version of the Exercise Adherence Rating Scale: a methodological study. Health Qual Life Outcomes 2020 Oct 7;18(1):328. [doi: 10.1186/s12955-020-01588-6] [Medline: 33028349]
63. Arensman RM, Geelen RH, Koppenaal T, Veenhof C, Pisters MF. Measuring exercise adherence in patients with low back pain: development, validity, and reliability of the EXercise Adherence Scale (EXAS). Physiother Theory Pract 2022 Jul;38(7):928-937. [doi: 10.1080/09593985.2020.1818337] [Medline: 32933359]
64. Chen M, Wu T, Lv M, et al. Efficacy of mobile health in patients with low back pain: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2021 Jun 11;9(6):e26095. [doi: 10.2196/26095] [Medline: 34114965]
65. Vad VB, Madrazo-Ibarra A, Estrin D, et al. Back Rx, a personalized mobile phone application for discogenic chronic low back pain: a prospective pilot study. BMC Musculoskelet Disord 2022 Oct 19;23(1):923. [doi: 10.1186/s12891-022-05883-9] [Medline: 36261825]
66. Rintala A, Rantalainen R, Kaksonen A, Luomajoki H, Kauranen K. mHealth apps for low back pain self-management: scoping review. JMIR Mhealth Uhealth 2022 Aug 26;10(8):e39682. [doi: 10.2196/39682] [Medline: 36018713]
67. Azevedo DC, Ferreira PH, de Oliveira Santos H, Oliveira DR, Leite de Souza JV, Pena Costa LO. Association between patient independence in performing an exercise program and adherence to home exercise program in people with chronic low back pain. Musculoskelet Sci Pract 2021 Feb;51:102285. [doi: 10.1016/j.msksp.2020.102285] [Medline: 33257276]

XSL•FO
RenderX

68. Arensman RM, Pisters MF, Kloek CJJ, Koppenaal T, Veenhof C, Ostelo RJWG. Exploring the association between adherence to home-based exercise recommendations and recovery of nonspecific low back pain: a prospective cohort study. BMC Musculoskelet Disord 2024 Aug 1;25(1):614. [doi: 10.1186/s12891-024-07705-6] [Medline: 39090661]

Abbreviations

6MWT: 6-Minute Walk Test
CBT: cognitive behavioral therapy
CG: control group
CLBP: chronic low back pain
CONSORT: Consolidated Standards of Reporting Trials
EARS: Exercise Adherence Rating Scale
EPAP: Evaluation of Physical Activity Perception
ES: effect size
ICC: intraclass correlation coefficient
IG: intervention group
LBP: low back pain
mHealth: mobile health

Edited by L Buis; submitted 19.May.2025; peer-reviewed by J Igboanugo, Y Zhang; revised version received 05.Dec.2025; accepted 13.Apr.2026; published 15.Jun.2026.

Please cite as:
Lechauve JB, Dobija L, Pereira B, Grolier M, Pelletier-Visa M, Lanhers C, Coudeyre E
Evaluation of the Impact of a Smartphone App on Adherence to an Exercise Program in People With Chronic Low Back Pain: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e77736
URL: https://mhealth.jmir.org/2026/1/e77736
doi:10.2196/77736

© Jean-Baptiste Lechauve, Lech Dobija, Bruno Pereira, Maxime Grolier, Mathilde Pelletier, Charlotte Lanhers, Emmanuel Coudeyre. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 15.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Smartphone Cardiac Rehabilitation, Assisted Self-Management (SCRAM) Versus Usual Care: Multicenter Randomized Controlled Trial

Ralph Maddison[1], PhD; Narayan Subedi[1], PhD; Peixuan Li[2,3], MBiostat; Karen E Lamb[2,3], PhD; Kylie Ball[1], PhD; Brian Oldenburg[4,5], PhD; Clara Chow[6], MBBS, PhD; Sarah A McNaughton[1,7], PhD; Elena S George[1], PhD; Lan Gao[8], PhD; Marj Moodie[8], DrPH; John Amerena[9], MBBS, FRACP, FCSANZ; Voltaire Nadurata[10], MD, FRACP, FCSANZ; Christopher Neil[11], MD, FRACP, FCSANZ; Stuart Cameron[12], BAppSc; Jonathan C Rawstorn[1], PhD

[1]Instituite for Physical Activity and Nutrition, School of Exercise and Nutriton Sciences, Deakin University, 221 Burwood Highway, Burwood, Melbourne, Victoria, Australia

[10]Department of Cardiology, Bendigo Health, Bendigo, Victoria, Australia

[11]Heart Wise Cardiology, Melbourne, Victoria, Australia

[12]Applied Artificial Intelligence Institute, Deakin University, Melbourne, Victoria, Australia

[2]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia

[3]Methods and Implementation Support for Clinical and Heath Research Hub, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Melbourne, Victoria, Australia

[4]Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia

[5]Department of Cardiovascular Research, Translation and Implementation, School of Psychology and Public Health, La Trobe University, Melbourne, Victoria, Australia

[6]Westmead Applied Research Centre, Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia

[7]School of Human Movement and Nutrition Sciences, The University of Queensland, Brisbane, Queensland, Australia

[8]Deakin Health Economics, Institute for Health Transformation, Deakin University, Melbourne, Victoria, Australia

[9]Geelong Cardiology Research Unit, Barwon Health, Geelong, Victoria, Australia

Corresponding Author:
Ralph Maddison, PhD
Instituite for Physical Activity and Nutrition, School of Exercise and Nutriton Sciences, Deakin University, 221 Burwood Highway, Burwood, Melbourne, Victoria, Australia

Abstract

Background: Accessibility barriers contribute to low participation in center-based cardiac rehabilitation. We developed an innovative, comprehensive, dual-phase telerehabilitation program to address this gap (Smartphone Cardiac Rehabilitation, Assisted Self-Management; SCRAM).

Objective: The study aimed to determine the effectiveness of SCRAM for increasing maximal aerobic exercise capacity (VO_2max).

Methods: A multicenter, parallel 2-arm randomized controlled trial recruited clinically stable adults (aged ≥18 y) with diagnosed coronary heart disease at 3 hospitals in Victoria, Australia (Melbourne, Geelong, and Bendigo) from 2018 to 2021. Participants were randomized (1:1), stratified by sex and study site, to receive SCRAM plus usual cardiovascular care (intervention) or usual cardiovascular care alone (control). SCRAM provided 24 weeks of remote exercise supervision, coaching, and behavior change support via smartphone. Usual cardiovascular care included standard medical care and advice to seek a referral to center-based cardiac rehabilitation, which was heavily impacted during the COVID-19 pandemic. Due to the nature of the treatments, participants were not blinded to allocation; primary outcome assessors and biostatisticians were blinded. The primary outcome was VO_2max at 24 weeks, analyzed on the principle of intention-to-treat, using linear regression adjusted for baseline and stratification factors on multiple imputed data.

Results: Recruitment and data collection were heavily impacted by COVID-19, although SCRAM delivery was sustained throughout. Of 220 required participants, only 123 (56%) were recruited and randomized (intervention n=63, control n=60); 45% (55/123) had missing VO_2max at 24 weeks—largely due to enforced COVID-19 restrictions. Mean VO_2max at 24 weeks favored SCRAM (26.10, SD 10.72 mL/kg/min) over control (24.65, SD 7.87 mL/kg/min), but the difference was not statistically significant (mean difference=1.61 mL/kg/min, 95% CI –1.38 to 4.61, *P*=.28). Among secondary outcomes, patients receiving SCRAM had

XSL•FO
RenderX

lower diastolic blood pressure at 24 weeks (mean difference=–5.54 mm Hg, 95% CI –10.01 to –1.06). All reported adverse events (control n=6, intervention n=16) were deemed mild or moderate, with only one deemed as possibly related to treatment. There were no deaths or hospitalizations.

Conclusions: This was an underpowered trial, but SCRAM did not lead to a clinically important difference in VO_2max compared to usual cardiac care. SCRAM was resilient to COVID-19–related disruptions that significantly impacted the delivery of cardiac rehabilitation and supervised exercise training in particular. Further research is needed to conclusively assess treatment effects and understand how virtual cardiac rehabilitation can be translated into routine practice to augment center-based delivery and enhance equity of access.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12618001458224; https://anzctr.org.au/Trial/Registration/TrialReview.aspx?id=374508

International Registered Report Identifier (IRRID): RR2-10.2196/15022

(JMIR Mhealth Uhealth 2026;14:e66074) doi:10.2196/66074

KEYWORDS

mHealth; cardiac rehabilitation; telerehabilitation; rural; coronary heart disease; randomized controlled trial

Introduction

Cardiovascular disease (CVD) is the leading cause of death and disease burden, globally [1]. Since 1990, the burden of CVD has continued to increase in most countries, with trends driven by changing exposures to harmful risk factors, population growth, and population aging [1]. For people living with CVD, secondary prevention comprising lifestyle modifications, pharmacotherapy, and cardiac rehabilitation is recommended to reduce the risk of recurrent events and is a priority of the World Heart Federation [2].

Cardiac rehabilitation has a class 1a recommendation for management of CVD, particularly for coronary heart disease (CHD) [3-5]. Cardiac rehabilitation programs deliver comprehensive support, education, and monitoring of patients after a cardiovascular event. In a meta-analysis of 85 randomized controlled trials (RCTs) involving 23,430 participants with CVD, exercise-based rehabilitation was associated with significant risk reductions in cardiovascular mortality (risk ratio [RR]: 0.74, 95% CI 0.64 - 0.86, hospitalizations [RR: 0.77, 95% CI 0.67 - 0.89], and myocardial infarction [RR: 0.82, 95% CI 0.70 - 0.96]). There was evidence that cardiac rehabilitation improved health-related quality of life (HRQoL) and was cost-effective [6].

However, low participation rates limit the benefits of traditional center-based (face-to-face) programs [7-11]. Literature shows that 50% - 70% of patients eligible for cardiac rehabilitation do not attend, and among those that do attend, 30% - 60% do not complete their program [12,13]. Barriers to use are diverse and include distance to services and time pressures caused by the need for face-to-face in-hospital or clinical-setting attendance [14]. Evidence-based alternative delivery models are needed to overcome access barriers and improve participation [15,16].

Home-based delivery models have been designed to overcome aforementioned barriers to center-based cardiac rehabilitation. A Cochrane systematic review and meta-analysis of 24 RCTs involving 3046 participants that compared center- and home-based rehabilitation programs reported similar effectiveness for improving clinical, functional, and patient-reported outcomes, and low risk of adverse events [17]. These findings are encouraging, but they also identified a lack of interaction between participants and rehabilitation professionals during home-based delivery.

A small number of trials included in the Cochrane review augmented home-based delivery with digital technologies. Often known as telerehabilitation, this approach has rapidly gained interest because it can support interaction between participants and rehabilitation professionals and enable important program components such as social support and personalization. Indeed, the European Society of Cardiology recently defined telerehabilitation as a key quality indicator for cardiac rehabilitation program accreditation and undertook a Delphi methodology to identify minimum standards for high-quality delivery [18]. Recommendations highlight the potential benefits of both synchronous and asynchronous remote monitoring of exercise training. These approaches have also been outlined by the American Heart Association [19]. Synchronous monitoring involves viewing exercise data live during training sessions and could enable higher levels of responsiveness, personalization, and feedback but is more resource-intensive to deliver. Asynchronous monitoring involves intermittently reviewing previously recorded exercise data and could provide more flexibility for participants and professionals and promote greater autonomy for reducing resource use but offers less interaction and support from rehabilitation professionals [18,20].

These attributes may make synchronous monitoring well-suited during earlier program stages (eg, phase II) and for individuals whose medical and exercise history indicates a need for closer supervision. Conversely, asynchronous monitoring may be well suited for later program stages (eg, phase III and beyond) and for individuals whose medical and exercise history is suitable for more self-directed exercise. These approaches could be combined sequentially to provide a graduated program of long-term support that emphasizes safety and personalization on entry before transitioning toward self-direction and lifelong behavior change.

We previously developed and evaluated the 12-week REMOTE-CR (remote exercise monitoring trial for exercise-based cardiac rehabilitation) telerehabilitation program,

XSL•FO
RenderX

which delivered synchronous remote exercise coaching via smartphone and wearable sensing technologies [21,22]. The effect on maximal aerobic exercise capacity (VO_2max; primary outcome) was noninferior to center-based programs (adjusted mean difference −0.51, 95% CI −0.97 to −1.98 mL/kg/min; *P*=.48; prespecified inferiority margin: −1.25) [23]. REMOTE-CR was also substantially cheaper to deliver and had high usability, acceptability, and end-user demand [23,24].

Following the REMOTE-CR trial, we extended the intervention to include asynchronous exercise and behavioral support for an additional 12 weeks (total 24 wk) and to incorporate evidence-based multifactorial support for self-management behaviors [25] (eg, diet, medication adherence). The extended intervention, named Smartphone Cardiac Rehabilitation, Assisted Self-Management (SCRAM) comprised additional modular components in the form of push notifications that were derived from our previous research studies [26-28] and expert input.

The aims of this study were to compare the effects of the SCRAM intervention on exercise capacity, lifestyle and self-management behaviors, and HRQoL with usual cardiovascular care in adults with CHD.

We hypothesized that SCRAM would augment usual secondary prevention services and improve cardiorespiratory fitness compared to controls. In addition, we hypothesized that SCRAM would have a positive effect on other lifestyle behaviors including diet and physical activity.

Methods

Trial Design

A pragmatic multicenter 2-arm parallel-group RCT was conducted between November 2018 and August 2021. Participant eligibility criteria were unchanged after trial commencement, but enforced COVID-19 restrictions affected recruitment, outcome measurement, and access to usual cardiovascular care. The trial protocol was prospectively registered (ACTRN12618001458224, August 30, 2018), published [29], and reported according to CONSORT (Consolidated Standards of Reporting Trials) guidelines (Checklist 1) [30]. Analyses were outlined in a detailed plan prior to unblinding (Multimedia Appendix 1) with postunmasking changes noted.

Changes to Trial Protocol

Enforced responses to COVID-19 at study sites required a change in the way the primary outcome was measured for some participants (see Outcomes section), and changed delivery of the control treatment (see Control section). Difficulty achieving the target sample size also resulted in changes to the preplanned analyses (Multimedia Appendix 1).

Eligibility and Recruitment

Clinically stable (no hospitalization within 6 wk) adults (≥18 y) with a recent CHD diagnosis (angina, myocardial infarction, and coronary revascularization within 6 mo) were recruited from 3 hospitals (Sunshine Hospital, Western Health; University Hospital Geelong, Barwon Health; Bendigo Hospital, Bendigo Health) in Victoria, Australia (November 12, 2018 to March 21, 2021). Research nurses identified participants from in-patient records and outpatient clinics and provided verbal and written information. Consenting participants were scheduled for baseline assessment.

Intervention

Participants received SCRAM plus usual cardiovascular care. SCRAM has been described fully elsewhere [29], according to recommendations in the Template for Intervention Description and Replication [31]. Briefly, SCRAM was a multicomponent dual-phase intervention that provided participants with a comprehensive, 24-week individualized, evidence-based program of exercise training and modular behavioral self-management support via a bespoke smartphone platform. The SCRAM platform included a Polar wearable sensor, a participant-facing app that is compatible with iOS and Android (≥v5.0) operating systems, and a health professional-facing web app, compatible with mobile and desktop web browsers (Figure 1). Behavior change strategies and education were delivered via staged modular push notifications to support the uptake and maintenance of healthy eating, physical activity, medication taking, stress management, and, if indicated, smoking cessation. Each module provided 2 - 4 notifications per week in weeks 1 - 12 and 1 - 3 notifications per week in weeks 13 - 24. As highlighted above, the additional modules were derived from our previous trials [26-28] as well as expert input from an exercise physiologist (JCR) and dietitian (ESG) to ensure relevance for the Australian context. Participants accessed the smartphone app via Apple and Google Play Stores.

XSL•FO
RenderX

Figure 1. Schematic diagram of the Smartphone Cardiac Rehabilitation, Assisted Self-Management platform.



During weeks 1 - 12, accredited exercise physiologists provided synchronous remote exercise prescription, monitoring, and coaching via the SCRAM platform. Weeks 13 - 24 transitioned participants toward independent, self-determined exercise and behavior change. Exercise physiologists provided weekly coaching phone calls informed by asynchronous review of exercise data, rather than synchronous supervision.

Control

Participants received usual cardiovascular care, including advice to seek referral to center-based cardiac rehabilitation.

Outcomes

Primary Outcome

Maximal oxygen uptake (VO_2max, mL/kg/min) was assessed at baseline and 24 weeks during an incremental treadmill test. When direct measurement (respiratory gas analysis) was precluded by COVID-19 restrictions, VO_2max was estimated from treadmill velocity and gradient using established metabolic equations for treadmill walking (<8 km/h: [velocity {m/min} * 0.1] + [velocity {m/min} * gradient {%} * 1.8]+3.5) or running (>8 km/h: [velocity {m/min} * 0.2] + [velocity {m/min} * gradient {%} * 0.9]+3.5) [32].

Secondary Outcomes

Secondary outcomes are summarized in Table 1. COVID-19 restrictions prevented objective outcome measurement for some participants. Self-reported physical activity was assessed using the Godin Leisure-time Physical Activity Questionnaire [33]. Dietary intake was assessed using the web-based, Automated Self-Administered 24-Hour Dietary Assessment Tool [34-37]. Participants completed the Automated Self-Administered 24-Hour Dietary Assessment Tool for 1 calendar weekday within 3 days of assessment appointments. Alcohol consumption was assessed using the Alcohol Use Disorders Identification Test-C [38], while the Medication Adherence Scale [39] assessed self-reported medication adherence. HRQoL was assessed using the Assessment of Quality of Life 8-Dimension scale [40]. The 8 dimensions include independent living, happiness, mental health, coping, relationships, self-worth, pain, and senses. Methods for prespecified health economic analyses have been reported in detail elsewhere [41], and the results will also be reported separately. We carried out semistructured exit interviews with trial participants, which were categorized under the themes of usability, acceptability, and satisfaction. Process evaluation outcomes using the Reach, Effectiveness, Adoption, Implementation, and Maintenance framework will also be reported as a separate paper.

XSL•FO
RenderX

Table . Secondary outcomes[a].

Outcome	Baseline	Wk 12	Wk 24
Objective			
BP[b] systolic or diastolic (mm Hg)	✓	N/A[c]	✓
Body mass (kg)	✓	N/A	✓
BMI (kg/m^2)	✓	N/A	✓
Waist or hip circumference (cm)	✓	N/A	✓
Waist hip ratio	✓	N/A	✓
Fasted blood lipids (mmol/L)	✓	N/A	✓
Fasted blood glucose (mmol/L)	✓	N/A	✓
Subjective			
Physical activity (% reporting Godin LSI[d] ≥14 units) [33]	✓	✓	✓
Medication adherence (% reporting score=4) [39]	✓	✓	✓
Alcohol consumption (% reporting ≤2 drinks/d) [38]	✓	✓	✓
HRQoL[e] (AQoL-8D[f], multiattribute utility score) [40]	✓	✓	✓
Vegetable intake (servings/d) via ASA24-Australia[g] [34-37].	✓	✓	✓

[a]Randomized controlled trial, Victoria, Australia. Population: adults with coronary heart disease.

[b]BP: blood pressure.

[c]NA: not assessed.

[d]LSI: leisure score index.

[e]HRQoL: health-related quality of life.

[f]AQoL-8D: Assessment of Quality of Life-8 Dimensions.

[g]ASA24-Australia: Automated Self-Administered 24-Hour Dietary Assessment Tool.

Adverse events were reported at 12 and 24 weeks and classified for severity (mild, moderate, and severe) and likely relationship to study treatment (not related, possible, probable, and definite).

Sample Size

A total of 220 participants were required to detect a clinically important 2 (SD 6.75) mL/kg/min between-group difference in VO_2max at 24 weeks with 2-sided α of .05, 90% power, assuming preintervention and postintervention correlation of 0.8 [42], and 20% attrition [29].

Randomization

An independent biostatistician created a computer-generated randomization schedule, stratified by sex and trial site. Other researchers did not have access to the sequence. Participants were allocated to receive intervention or control treatments at a 1:1 ratio using a centralized web-based system (Research Electronic Data Capture [REDCap]; Vanderbilt University) that ensured allocation concealment until the time of randomization.

Blinding

Due to the nature of the intervention, participants could not be blinded to allocation. Primary outcome assessors and biostatisticians were blinded to allocation. Secondary outcomes were either self-reported or assessed by research nurses or researchers who were not blinded to allocation.

Statistical Analysis

Available data were analyzed according to randomized assignment (ie, intention-to-treat). A linear regression model was fitted to estimate the absolute mean between-group difference in VO_2max at 24 weeks, irrespective of participant adherence, adjusting for baseline VO_2max and stratification factors (sex and study site). Similar analyses were undertaken for continuous secondary outcomes, except for HRQoL. A linear mixed model was fitted for HRQoL accounting for treatment arm, time, treatment arm by time interaction, baseline value, and stratification factors. Logistic regression models with generalized estimating equations were fitted for binary secondary outcomes (physical activity, alcohol consumption, and medication adherence), and an ordinal regression model with clustered standard errors was fitted for vegetable consumption; these models accounted for treatment arm, time, treatment arm by time interaction, baseline values of each outcome and stratification factors, sex, and study site.

Primary analysis used multiple imputations due to high levels of missing follow-up data. Complete case analyses were

XSL•FO
RenderX

undertaken in sensitivity analyses. Per protocol analyses excluded participants with major protocol violations or nonadherence to the SCRAM intervention (recording <12 exercise sessions via the SCRAM app during each of the intensive [wk 1 - 12] and maintenance phases [wk 12 - 24]); these data are presented descriptively by treatment arm due to low numbers. Adverse events are reported descriptively by the treatment arm.

Ethical Considerations

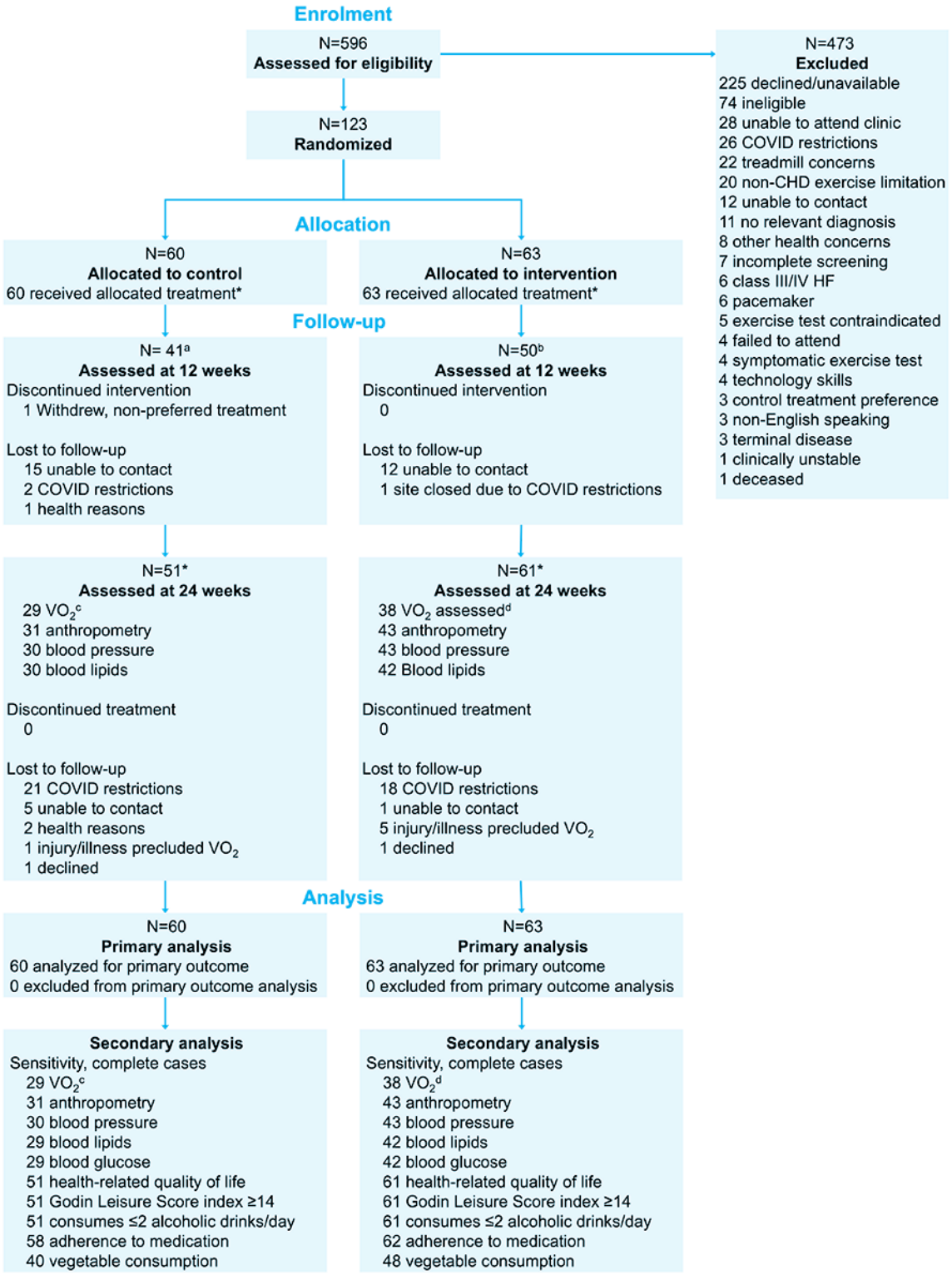
The trial received multisite ethical approval from Melbourne Health (HREC/18/MH/119), which was ratified by Deakin University (2018-251), in accordance with institutional and national guidelines for research involving human participants. All participants were fully informed about the study and the voluntary nature of their involvement, including their right to withdraw at any time without penalty. Written informed consent was obtained prior to participation. Privacy and confidentiality were maintained through deidentification of collected data, which were stored securely on encrypted servers with access restricted to authorized research staff. It is not possible to identify individual participants from any images in the manuscript or supplementary material. Each participant received AUD $50 (US $35.39) shopping vouchers at the completion of baseline and 24-week assessments. Intervention group participants who used their own smartphone and/or data plan received an additional AUD $60 (US $42.46) value to reimburse data use during the intervention at the 24-week assessment.

Results

COVID-19 severely affected the recruitment and collection of objectively measured outcomes. As a result, we were unable to achieve the recruitment target. Randomization and outcome assessments completed pre- and peri-COVID were comparable between groups (Multimedia Appendix 2). Overall, 123 (55.9% of the target) participants were randomized (Figure 2); characteristics are reported in Table 2. Recruitment stopped (March 21, 2021) due to COVID-19 lockdowns and health and safety measures implemented by participating hospitals. The SCRAM intervention was delivered as planned throughout the trial. Delivery of usual care cardiac rehabilitation (control) was frequently disrupted by enforced COVID-19 restrictions, which meant that in some regions, participants may not have been offered in-person, center-based contact. We were unable to quantify the level of contact for control participants.

XSL•FO
RenderX

Figure 2. Consolidated Standards of Reporting Trials flow diagram. Randomized controlled trial, Victoria, Australia. Population: adults with coronary heart disease. *Completed lifestyle survey. [a]n=40 for Godin leisure score index at 12 weeks; [b]n=49 for Godin leisure score index at 12 weeks; [c]Estimated relative VO_2 was used when objectively measured peak relative VO_2 was not available (measured n=27, estimated n=2); [d]Estimated relative VO_2 was used when objectively measured peak relative VO_2 was not available (measured n=35, estimated n=3). CHD: coronary heart disease; HF: heart failure.



XSL•FO
RenderX

Table . Baseline demographic and clinical characteristics[a].

Characteristics	Intervention (n=63)	Control (n=60)	Total (n=123)
Participants, n (%)			
Barwon Health	13 (20.6)	11 (18.3)	24 (19.5)
Western Health	12 (19)	11 (18.3)	23 (18.7)
Bendigo Health	38 (60.3)	38 (63.3)	76 (61.8)
Age (y), mean (SD)	61.3 (9.9)	60.5 (11.2)	60.9 (10.5)
Sex, n (%)			
Male	54 (85.7)	51 (85)	105 (85.4)
Female	9 (14.3)	9 (15)	18 (14.6)
Ethnicity, n (%)			
Australian	55 (87.3)	48 (80)	103 (83.7)
Other	8 (12.7)	12 (20)	20 (16.3)
Household income[b], n (%)			
Below median	46 (73)	41 (68.3)	87 (70.7)
Above median	15 (23.8)	16 (26.7)	31 (25.2)
Do not know or refuse to answer	2 (3.2)	3 (5)	5 (4.1)
Education level, n (%)			
Bachelor degree and above	14 (22.2)	13 (22)	27 (22.1)
Less than Bachelor degree	49 (77.8)	46 (78)	95 (77.9)
Employment status, n (%)			
Employed	33 (52.4)	34 (56.6)	67 (54.5)
Unemployed	30 (47.6)	26 (43.4)	56 (45.5)
Smoking status, n (%)			
Never smoked	26 (41.3)	23 (39)	49 (40.2)
Ex-smoker	31 (49.2)	28 (47.5)	59 (48.4)
Current smoker	6 (9.5)	8 (13.6)	14 (11.5)
Medical history, n (%)			
Hypertension	30 (47.6)	35 (58.3)	65 (52.8)
BP[c] lowering medication	47 (74.6)	41 (69.5)	88 (72.1)
Diabetes	17 (27)	12 (20.3)	29 (23.8)
Insulin[d]	4 (23.5)	6 (50)	10 (34.5)
Blood sugar-lowering medication[d]	12 (70.6)	9 (75)	21 (72.4)
High cholesterol	39 (61.9)	31 (52.5)	70 (57.4)
Cholesterol-lowering medication	54 (85.7)	52 (88.1)	106 (86.9)
Myocardial infarction	46 (73)	42 (71.2)	88 (72.1)
Angina	25 (39.7)	20 (33.9)	45 (36.9)
Stent or PCI[e]	54 (85.7)	44 (74.6)	98 (80.3)
CABG[f]	13 (20.6)	13 (22)	26 (21.3)
Atrial fibrillation	5 (7.9)	5 (8.5)	10 (8.2)
Other forms of heart disease	2 (3.2)	0 (0)	2 (1.6)

[a]Randomized controlled trial, Victoria, Australia. Population: adults with coronary heart disease.

[b]Victorian median annual household income approximately equal to AUD $90,000 (US $63,697.05) [43].

XSL•FO
RenderX

[c]BP: blood pressure.

[d]Denominators were numbers of participants with diabetes.

[e]PCI: percutaneous coronary intervention.

[f]CABG: coronary artery bypass graft.

Enforced COVID-19 restrictions meant VO_2max could only be assessed in 67 (54.5%; intervention=38 [35 measured, 3 estimated], control=29 [27 measured, 2 estimated] participants at 24 weeks; Figure 2). Using multiple imputation, we found a between-group VO_2max difference at 24 weeks favoring SCRAM, but it was not statistically significant (mean difference=1.61 mL/kg/min; 95% CI −1.38 to 4.61, *P*=.28).

Among secondary outcomes, SCRAM demonstrated a positive effect compared to control for diastolic blood pressure at 24 weeks (mean difference: −5.54 mm Hg; 95% CI −10.01 to −1.06, *P*=.02, Tables 3 and 4). CIs included the null for between-group differences of self-reported secondary outcomes at 12 weeks (Tables 5 and 6).

Table . Primary intention-to-treat analyses of outcomes assessed at 24 wk, using multiple imputed data[g].

Outcome	Baseline		Wk 24		Within-group change		Between-group difference	
	Control (n=60), mean (SD)	Intervention (n=63), mean (SD)	Control (n=60), mean (SD)	Intervention (n=63), mean (SD)	Control (n=60), mean (SD)	Intervention (n=63), mean (SD)	Intervention–control (n=123), mean (95% CI)	*P* value
VO_2max (mL/kg/min)	24.16 (8.02)	23.85 (7.77)	24.65 (7.87)	26.10 (10.72)	0.49 (8.07)	2.25 (9.92)	1.61 (–1.38 to 4.61)	.28
BP[b] systolic (mm Hg)	138.77 (18.08)	130.85 (20.47)	135.40 (19.54)	132.23 (21.67)	–3.37 (20.79)	1.38 (22.75)	–0.10 (–6.95 to 6.76)	.98
BP diastolic (mm Hg)	80.93 (10.92)	78.64 (11.11)	84.38 (14.68)	77.97 (11.76)	3.45 (15.60)	–0.67 (12.85)	–5.54 (–10.01 to –1.06)	.02
Body mass (kg)	86.14 (15.42)	87.37 (14.80)	86.48 (12.56)	85.61 (11.11)	0.34 (12.17)	–1.76 (9.46)	–1.58 (–4.45 to 1.28)	.28
BMI (kg/m^2)	29.47 (5.19)	29.35 (5.64)	29.59 (4.22)	28.82 (4.23)	0.12 (4.59)	–0.54 (3.09)	–0.72 (–1.79 to 0.36)	.19
Waist circumference (cm)	100.99 (12.13)	101.68 (12.60)	101.21 (11.63)	100.40 (12.25)	0.21 (10.29)	–1.28 (6.53)	–1.40 (–4.03 to 1.23)	.29
Hip circumference (cm)	101.85 (10.03)	103.27 (9.93)	101.87 (8.08)	101.55 (10.11)	0.02 (9.94)	–1.73 (5.77)	–1.30 (–3.74 to 1.14)	.29
Waist-hip ratio	0.99 (0.08)	0.98 (0.08)	1.01 (0.06)	0.99 (0.10)	0.01 (0.08)	0.00 (0.08)	–0.02 (–0.04 to 0.01)	.18
Total-C[c] (mmol/L)	3.55 (1.02)	3.64 (1.04)	3.87 (1.52)	3.72 (1.22)	0.33 (1.48)	0.07 (0.96)	–0.22 (–0.66 to 0.22)	.32
HDL-C[d] (mmol/L)	1.09 (0.37)	1.03 (0.28)	1.03 (0.47)	1.10 (0.40)	–0.06 (0.50)	0.06 (0.39)	0.10 (–0.06 to 0.25)	.22
LDL-C[e] (mmol/L)	1.85 (0.92)	1.84 (0.77)	2.08 (1.35)	1.85 (1.04)	0.22 (1.46)	0.01 (0.88)	–0.22 (–0.63 to 0.20)	.31
Blood glucose (mmol/L)	6.37 (2.33)	6.22 (2.04)	6.64 (3.87)	6.32 (2.50)	0.27 (3.49)	0.10 (2.86)	–0.21 (–1.29 to 0.87)	.69
HRQoL[f]	0.78 (0.15)	0.77 (0.19)	0.79 (0.17)	0.81 (0.17)	0.01 (0.14)	0.03 (0.10)	0.01 (–0.05 to 0.07)	.73

[a]Randomized controlled trial, Victoria, Australia. Population: adults with coronary heart disease.

[b]BP: blood pressure.

[c]Total-C: total cholesterol.

[d]HDL-C: high-density lipoprotein cholesterol.

[e]LDL-C: low-density lipoprotein cholesterol.

[f]HRQoL: health-related quality of life.

XSL•FO
RenderX

Table . Primary intention-to-treat analyses of additional outcomes assessed at 24 wk, using multiple imputed data[a].

Outcome	Baseline		Wk 24		Within-group change	Between-group difference	Within-group change	
	Control (n=60), n (%)	Intervention (n=63), n (%)	Control (n=60), n (%)	Intervention (n=63), n (%)			Odds ratio (95% CI)	*P* value
Godin LSI[b] ≥14	45 (75)	50 (79.37)	54 (90.50)	56 (88.48)	—[c]	—	0.77 (0.20-2.91)	.70
Alcohol intake ≤2 drinks/d	54 (90)	56 (88.89)	55 (91.67)	59 (93.62)	—	—	1.35 (0.34-5.40)	.67
Medication adherence score=4	38 (63.33)	29 (46.03)	36 (59.20)	38 (60.13)	—	—	1.04 (0.50-2.17)	.91
Vegetable consumption					—	—	1.10 (0.56-2.17)	.78
≤1 serving/d	2 (3.33)	0 (0)	2 (3.33)	1 (1.75)				
1 serving/d	8 (13.33)	7 (11.11)	3 (5)	5 (7.94)				
2 servings/d	7 (11.67)	11 (17.46)	11 (18.33)	7 (11.11)				
3 servings/d	16 (26.67)	17 (26.98)	15 (25)	21 (33.33)				
4 servings/d	18 (30)	13 (20.63)	13 (21.67)	12 (19.05)				
≥5 servings/d	9 (15)	15 (23.81)	16 (26.67)	17 (26.98)				

[a]Randomized controlled trial, Victoria, Australia. Population: adults with coronary heart disease.

[b]LSI: leisure score index.

[c]Not available.

Table . Primary intention-to-treat analyses of health-related quality of life assessed at 12 wk, using multiple imputed data[a].

Outcome	Baseline		12 wk		Within-group change		Between-group difference	
	Control (n=60), mean (SD)	Intervention (n=63), mean (SD)	Control (n=60), mean (SD)	Intervention (n=63), mean (SD)	Control (n=60), mean (SD)	Intervention (n=63), mean (SD)	Intervention–control (n=88), mean (95% CI)	*P* value
Health-related quality of life	0.78 (0.15)	0.77 (0.19)	0.79 (0.19)	0.78 (0.20)	0.01 (0.17)	0.01 (0.12)	–0.02 (–0.08 to 0.05)	.64

[a]Randomized controlled trial, Victoria, Australia. Population: adults with coronary heart disease.

XSL•FO
RenderX

Table . Primary intention-to-treat analyses of additional outcomes assessed at 12 wk, using multiple imputed data[a].

Outcome	Baseline		12 wk		Within-group change	Between-group difference	Within-group change	
	Control (n=60), n (%)	Intervention (n=63), n (%)	Control (n=60), n (%)	Intervention (n=63), n (%)	Control (n=60), n (%)	Intervention (n=63), n (%)	Intervention–control (n=88), odds ratio (95% CI)	*P* value
Godin LSI[b] 14	45 (75)	50 (79.37)	52 (87.37)	58 (91.30)	—[c]	—	1.51 (0.37-6.25)	.57
Alcohol intake ≤2 drinks/d	54 (90)	56 (88.89)	36 (88.67)	45 (87.81)	—	—	0.89 (0.27-3.00)	.85
Medication adherence score=4	38 (63.33)	29 (46.03)	35 (57.93)	40 (62.76)	—	—	1.23 (0.54-2.83)	.62
Vegetable consumption					—	—	1.71 (0.79-3.71)	.18
≤1 serving/d	2 (3.33)	0 (0)	2 (3.07)	1 (1.81)				
1 serving/d	8 (13.33)	7 (11.11)	9 (15.07)	9 (14.13)				
2 servings/d	7 (11.67)	11 (17.46)	6 (10.27)	7 (11.71)				
3 servings/d	16 (26.67)	17 (26.98)	23 (37.87)	13 (20.38)				
4 servings/d	18 (30)	13 (20.63)	12 (20.37)	20 (31.17)				
≥5 servings/d	9 (15)	15 (23.81)	8 (13.37)	13 (20.79)				

[a]Randomized controlled trial, Victoria, Australia. Population: adults with coronary heart disease.

[b]LSI: leisure score index.

[c]Not available.

Sensitivity analyses, including prognostic factors and complete case analyses, were generally consistent with the primary analyses (Multimedia Appendix 3). Per protocol analyses excluded 19 participants (control n=10, intervention n=9) for eligibility criteria violations (recent CHD diagnosis n=3, clinically stable outpatients n=15, terminal disease n=1) and 42 participants for SCRAM treatment nonadherence (Multimedia Appendix 3).

Among 22 reported adverse events (control n=6, intervention n=16), all were deemed mild or moderate, and only 1 was deemed as possibly related to treatment. There were no deaths or hospitalizations (Multimedia Appendix 4).

As part of the process evaluation, we conducted interviews with 21 trial participants who completed the study (n=11 from the intervention group and n=10 from the control group), 4/21 (19%) of whom were female participants. Participants reported high levels of usability, acceptability, and satisfaction with the SCRAM intervention. Most intervention participants (9/11, 82%) indicated that the technology was simple and easy to use. All participants agreed that SCRAM provided a flexible option for cardiac rehabilitation, allowing them to exercise in locations convenient to them while retaining expert supervision. Those who found SCRAM easy to use expressed high satisfaction with the intervention, particularly valuing the expert supervision, synchronous coaching, and activity-monitoring features.

Discussion

Principal Findings

This study sought to compare the effects of the SCRAM intervention on cardiorespiratory fitness, lifestyle and self-management behaviors, and HRQoL with usual cardiovascular care in adults with CHD. Overall, we found beneficial directional effects on the primary and selected secondary outcomes that favored the SCRAM intervention, but our findings are inconclusive as we were unable to recruit the a priori target sample size.

Unfortunately, the COVID-19 pandemic severely impacted this trial. We did not achieve our desired sample size, and follow-up data were lost due to enforced closures within the Australian health care services during this trial. While beneficial directional effects were generally consistent with our previous effectiveness trial [23], a lack of statistical power renders our findings, and any comparisons to previous cardiac telerehabilitation research, inconclusive. Nonetheless, it is important to report the trial to avoid publication bias [44].

Key findings were the resilience of telerehabilitation to severe cardiac rehabilitation service disruption, and its potential to deliver multifactorial supervised exercise and self-management behavioral support regardless of participants' geographic proximity to traditional center-based cardiac rehabilitation

XSL•FO
RenderX

facilities. The COVID-19 pandemic was unforeseen at the outset of our trial but significantly impacted global center-based cardiac rehabilitation delivery during the trial period, and telehealth quickly became the only safe and preferred delivery option [45-49]. Supervised exercise training was the most impacted program component globally, with approximately 76% of center-based programs reporting moderate to high impact due to COVID-19–related safety concerns [45]. While the SCRAM telerehabilitation intervention was not originally designed for such circumstances, its design enabled continuous delivery throughout COVID-19 restrictions in one of the world's most locked down regions.

Despite these drawbacks, this trial provides insights about a number of important issues that have been highlighted for high-quality telerehabilitation. Remote exercise monitoring, such as that provided during the first phase of the SCRAM intervention, was identified as an important component of effective telerehabilitation [50] and was recently advocated as a preferred standard to optimize individualization and risk management [18]. Synchronous monitoring is desirable because it enables rehabilitation professionals to verify exercise individual responses, tolerability, and clinical stability via heart rate, electrocardiogram, and symptoms [18]. This could play an essential role in intervention effectiveness by avoiding protective risk mitigation strategies such as overly conservative exercise prescription, which has been used in the context of unsupervised home-based exercise training [51]. Indeed, rehabilitation professionals have reported synchronous monitoring could be a useful tool to mitigate adverse event risk by enabling higher quality individualization and progression of exercise training [51].

Telerehabilitation standards also highlight the use of center-based cardiopulmonary exercise testing as a gold standard component of routine pre- and postprogram assessment [18]. Our trial included cardiopulmonary exercise testing, but it was not possible for all the participants. This was related to COVID-19 during our trial but could manifest in numerous ways that contradict key aims of telerehabilitation. For example, precluding participants who are most in need of alternative delivery methods because they cannot attend clinical centers for requisite preprogram assessment. Our adapted exercise testing method removed measurement of respiratory gas exchange to manage the risk of viral transmission but was still conducted at clinical centers. A systematic review of remote exercise testing methods in cardiac rehabilitation found that this is an underexplored area, with limited evidence demonstrating the validity of the six-minute walk test [52]. While this is commonly used in many clinical settings, it does not provide comparable data to cardiopulmonary exercise testing. This highlights an urgent need for evidence-based methods for remotely measuring exercise capacity as part of preintervention and postintervention assessments.

Cardiac rehabilitation interventions should comprehensively provide all evidence-based core components, regardless of the delivery model [18]. The SCRAM intervention improved upon our earlier REMOTE-CR intervention [23] by adding multifactorial behavior change education and support. However, this was delivered via push notifications rather than more intensive teleconsultation methods that have been recommended [18], and did not provide all types of recommended support over and above exercise training [11,53,54]. The predominance of physical activity and exercise-related features is common in cardiac telerehabilitation [19]. While exercise training is a pivotal part of comprehensive cardiac rehabilitation [55], these gaps represent areas for improvement to ensure alignment with best practices in cardiac rehabilitation.

The exercise monitoring features of SCRAM were well suited for monitoring aerobic exercise performance but less capable of monitoring key aspects of strength exercise such as volume, load, and technique. Since strength exercise is also an important component of cardiac rehabilitation [55,56], this represents another area of opportunity to ensure telerehabilitation more fully aligns with best practices in cardiac rehabilitation.

Strengths and Limitations

Our trial included a robust multicenter design, an objective primary outcome [57], a large diverse geographic area (including urban and regional sites), and an intervention that aligns with contemporary standards in cardiac telerehabilitation [18]. There were several limitations of our trial. First, despite considerable effort, COVID-19 response policies enforced at trial sites meant we could not recruit the required number of participants, which meant the study was underpowered. Also, due to enforced closures at the study sites, there was a large amount of missing follow-up data. Results need to be interpreted cautiously with these considerations in mind. Second, self-reported data for some outcomes had to be collected via telephone. Third, COVID-19 restrictions impacted access to in-person center-based cardiac rehabilitation services for control group participants, although they could access other telehealth services if offered by the trial sites. As a result, our comparator of usual care changed throughout the conduct of this trial. Ours was not the only trial affected by the COVID-19 pandemic. A recent paper highlighted the need to acknowledge its impact on study conduct, and the importance of reporting findings despite this [58].

Implications

Uninterrupted delivery across a large geographic area throughout the COVID-19 pandemic demonstrated promising capability to reinforce and augment center-based cardiac rehabilitation by addressing barriers that contribute to suboptimal and inequitable access. Further research is needed to conclusively assess treatment effects and understand how the SCRAM telerehabilitation platform can be translated into clinical practice.

Conclusion

This was an underpowered trial and findings were inconclusive, but SCRAM did not lead to a clinically important difference in VO_2max compared to usual cardiac care. SCRAM was resilient to COVID-19–related disruptions that significantly impacted the delivery of cardiac rehabilitation and supervised exercise training in particular. The SCRAM telerehabilitation intervention aligns with several contemporary standards for cardiac telerehabilitation, and further research is needed to conclusively assess treatment effects and understand how cardiac

XSL•FO
RenderX

telerehabilitation can be translated into routine practice to augment center-based delivery and enhance equity of access.

Acknowledgments

The authors gratefully acknowledge their funding organization (Australian National Health and Medical Research Council), study participants, health care practitioners and providers at trial sites, and key project personnel Jessica Freeman and Emily Scanlan.

Funding

Australian National Health and Medical Research Council Project grant (1144331).

Data Availability

The datasets generated and/or analyzed during this study are available from the corresponding author upon reasonable request, subject to compliance with ethical approval.

Authors' Contributions

RM conceived the trial and procured funding. BO, CC, SAM, LG, MM, JA, VN, CN, KEL, and JCR contributed to the study design, intervention development, and trial conduct. ESG also contributed to the intervention development. PLM and KEL developed the statistical analysis plan with RM and JCR and undertook all analyses. NS, RM, and JCR drafted the manuscript. All authors critically reviewed and approved the manuscript.

Conflicts of Interest

RM and JCR are inventors of the Smartphone Cardiac Rehabilitation, Assisted Self-Management software platform. Behavior change support built on work initiated by RM and JCR at the University of Auckland's National Institute for Health Innovation, and included contribution from ESG. The authors declare no financial or competing interests.

Multimedia Appendix 1
Statistical analysis plan.
[DOCX File, 225 KB - mhealth_v14i1e66074_app1.docx]

Multimedia Appendix 2
Summary of participant recruitment during the COVID-19 period.
[DOCX File, 22 KB - mhealth_v14i1e66074_app2.docx]

Multimedia Appendix 3
Sensitivity analyses.
[DOCX File, 38 KB - mhealth_v14i1e66074_app3.docx]

Multimedia Appendix 4
Safety outcomes.
[DOCX File, 25 KB - mhealth_v14i1e66074_app4.docx]

Checklist 1
Consolidated Standards of Reporting Trials checklist, Smartphone Cardiac Rehabilitation, Assisted Self-Management randomized controlled trial, Victoria, Australia.
[DOCX File, 23 KB - mhealth_v14i1e66074_app5.docx]

References

1. Global Burden of Cardiovascular Diseases and Risks 2023 Collaborators. Global, regional, and national burden of cardiovascular diseases and risk factors in 204 countries and territories, 1990-2023. J Am Coll Cardiol 2025 Dec 2;86(22):2167-2243. [doi: 10.1016/j.jacc.2025.08.015] [Medline: 40990886]
2. Laranjo L, Lanas F, Sun MC, et al. World Heart Federation roadmap for secondary prevention of cardiovascular disease: 2023 update. Glob Heart 2024;19(1):8. [doi: 10.5334/gh.1278] [Medline: 38273995]
3. O'Gara PT, Kushner FG, Ascheim DD, et al. 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: executive summary: a report of the American College of Cardiology Foundation/American Heart Association task force on practice guidelines: developed in collaboration with the American College of Emergency Physicians and

XSL•FO
RenderX

Society for Cardiovascular Angiography and Interventions. Catheter Cardiovasc Interv 2013 Jul 1;82(1):E1-E27. [doi: 10.1002/ccd.24776] [Medline: 23299937]
4. Levine GN, Bates ER, Blankenship JC, et al. 2011 ACCF/AHA/SCAI guideline for percutaneous coronary intervention: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines and the Society for Cardiovascular Angiography and Interventions. Circulation 2011 Dec 6;124(23):e574-e651. [doi: 10.1161/CIR.0b013e31823ba622] [Medline: 22064601]
5. Hillis LD, Smith PK, Anderson JL, et al. 2011 ACCF/AHA guideline for coronary artery bypass graft surgery: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. Circulation 2011 Dec 6;124(23):2610-2642. [doi: 10.1161/CIR.0b013e31823b5fee] [Medline: 22064600]
6. Dibben GO, Faulkner J, Oldridge N, et al. Exercise-based cardiac rehabilitation for coronary heart disease: a meta-analysis. Eur Heart J 2023 Feb 7;44(6):452-469. [doi: 10.1093/eurheartj/ehac747] [Medline: 36746187]
7. Balady GJ, Williams MA, Ades PA, et al. Core components of cardiac rehabilitation/secondary prevention programs: 2007 update: a scientific statement from the American Heart Association Exercise, Cardiac Rehabilitation, and Prevention Committee, the Council on Clinical Cardiology; the Councils on Cardiovascular Nursing, Epidemiology and Prevention, and Nutrition, Physical Activity, and Metabolism; and the American Association of Cardiovascular and Pulmonary Rehabilitation. Circulation 2007 May 22;115(20):2675-2682. [doi: 10.1161/CIRCULATIONAHA.106.180945] [Medline: 17513578]
8. Cowie A, Buckley J, Doherty P, et al. Standards and core components for cardiovascular disease prevention and rehabilitation. Heart 2019 Apr;105(7):510-515. [doi: 10.1136/heartjnl-2018-314206] [Medline: 30700518]
9. Piepoli MF, Corrà U, Adamopoulos S, et al. Secondary prevention in the clinical management of patients with cardiovascular diseases. Core components, standards and outcome measures for referral and delivery: a policy statement from the cardiac rehabilitation section of the European Association for Cardiovascular Prevention & Rehabilitation. Endorsed by the Committee for Practice Guidelines of the European Society of Cardiology. Eur J Prev Cardiol 2014 Jun;21(6):664-681. [doi: 10.1177/2047487312449597] [Medline: 22718797]
10. Suaya JA, Shepard DS, Normand SLT, Ades PA, Prottas J, Stason WB. Use of cardiac rehabilitation by Medicare beneficiaries after myocardial infarction or coronary bypass surgery. Circulation 2007 Oct 9;116(15):1653-1662. [doi: 10.1161/CIRCULATIONAHA.107.701466] [Medline: 17893274]
11. Woodruffe S, Neubeck L, Clark RA, et al. Australian Cardiovascular Health and Rehabilitation Association (ACRA) core components of cardiovascular disease secondary prevention and cardiac rehabilitation 2014. Heart Lung Circ 2015 May;24(5):430-441. [doi: 10.1016/j.hlc.2014.12.008] [Medline: 25637253]
12. Carlson JJ, Johnson JA, Franklin BA, VanderLaan RL. Program participation, exercise adherence, cardiovascular outcomes, and program cost of traditional versus modified cardiac rehabilitation. Am J Cardiol 2000 Jul 1;86(1):17-23. [doi: 10.1016/s0002-9149(00)00822-5] [Medline: 10867086]
13. Martin BJ, Hauer T, Arena R, et al. Cardiac rehabilitation attendance and outcomes in coronary artery disease patients. Circulation 2012 Aug 7;126(6):677-687. [doi: 10.1161/CIRCULATIONAHA.111.066738] [Medline: 22777176]
14. Chindhy S, Taub PR, Lavie CJ, Shen J. Current challenges in cardiac rehabilitation: strategies to overcome social factors and attendance barriers. Expert Rev Cardiovasc Ther 2020 Nov;18(11):777-789. [doi: 10.1080/14779072.2020.1816464] [Medline: 32885702]
15. Clark RA, Conway A, Poulsen V, Keech W, Tirimacco R, Tideman P. Alternative models of cardiac rehabilitation: a systematic review. Eur J Prev Cardiol 2015 Jan;22(1):35-74. [doi: 10.1177/2047487313501093] [Medline: 23943649]
16. Platz K, Kools S, Howie-Esquivel J. Benefits, facilitators, and barriers of alternative models of cardiac rehabilitation: a qualitative systematic review. J Cardiopulm Rehabil Prev 2023 Mar 1;43(2):83-92. [doi: 10.1097/HCR.0000000000000738] [Medline: 36346781]
17. McDonagh ST, Dalal H, Moore S, et al. Home-based versus centre-based cardiac rehabilitation. Cochrane Database Syst Rev 2023 Oct 27;10(10):CD007130. [doi: 10.1002/14651858.CD007130.pub5] [Medline: 37888805]
18. Scherrenberg M, Falter M, Abreu A, Aktaa S, Busnatu S, Casado-Arroyo R, et al. Standards for cardiac telerehabilitation: a scientific statement of the European Association of Preventive Cardiology (EAPC) and the Association of Cardiovascular Nursing & Allied Professions (ACNAP) of the ESC, and the ESC Working Group on e-Cardiology. Eur Heart J 2025;46(38):3714-3737. [doi: 10.1093/eurheartj/ehaf408] [Medline: 40742158]
19. Golbus JR, Lopez-Jimenez F, Barac A, et al. Digital technologies in cardiac rehabilitation: a science advisory from the American Heart Association. Circulation 2023 Jul 4;148(1):95-107. [doi: 10.1161/CIR.0000000000001150] [Medline: 37272365]
20. Thomas RJ, Petersen CE, Olson TP, Beatty AL, Ding R, Supervia M. Asynchronous and synchronous delivery models for home-based cardiac rehabilitation: a scientific review. J Cardiopulm Rehabil Prev 2021 Nov 1;41(6):407-412. [doi: 10.1097/HCR.0000000000000656] [Medline: 34727559]
21. Maddison R, Rawstorn JC, Rolleston A, et al. The remote exercise monitoring trial for exercise-based cardiac rehabilitation (REMOTE-CR): a randomised controlled trial protocol. BMC Public Health 2014 Nov 28;14(1):1236. [doi: 10.1186/1471-2458-14-1236] [Medline: 25432467]

XSL•FO
RenderX

22. Rawstorn JC, Gant N, Meads A, Warren I, Maddison R. Remotely delivered exercise-based cardiac rehabilitation: design and content development of a novel mHealth platform. JMIR mHealth uHealth 2016 Jun 24;4(2):e57. [doi: 10.2196/mhealth.5501] [Medline: 27342791]
23. Maddison R, Rawstorn JC, Stewart RAH, et al. Effects and costs of real-time cardiac telerehabilitation: randomised controlled non-inferiority trial. Heart 2019 Jan;105(2):122-129. [doi: 10.1136/heartjnl-2018-313189] [Medline: 30150328]
24. Rawstorn JC, Gant N, Rolleston A, et al. End users want alternative intervention delivery models: usability and acceptability of the REMOTE-CR exercise-based cardiac telerehabilitation program. Arch Phys Med Rehabil 2018 Nov;99(11):2373-2377. [doi: 10.1016/j.apmr.2018.06.027] [Medline: 30076800]
25. The assessment and management of cardiovascular risk. : New Zealand Guidelines Group; 2003 URL: https://www.tewhatuora.govt.nz/assets/Publications/Cardiovascular-Publications/assessment-and-management-of-cadriovasular-risk.pdf [accessed 2026-02-14]
26. Maddison R, Pfaeffli L, Whittaker R, et al. A mobile phone intervention increases physical activity in people with cardiovascular disease: results from the HEART randomized controlled trial. Eur J Prev Cardiol 2015 Jun;22(6):701-709. [doi: 10.1177/2047487314535076] [Medline: 24817694]
27. Maddison R, Stewart R, Doughty R, et al. Text4Heart II - improving medication adherence in people with heart disease: a study protocol for a randomized controlled trial. Trials 2018 Jan 25;19(1):70. [doi: 10.1186/s13063-018-2468-z] [Medline: 29370829]
28. Pfaeffli Dale L, Whittaker R, Jiang Y, Stewart R, Rolleston A, Maddison R. Text message and internet support for coronary heart disease self-management: results from the Text4Heart randomized controlled trial. J Med Internet Res 2015 Oct 21;17(10):e237. [doi: 10.2196/jmir.4944] [Medline: 26490012]
29. Rawstorn JC, Ball K, Oldenburg B, et al. Smartphone Cardiac Rehabilitation, Assisted Self-Management versus usual care: protocol for a multicenter randomized controlled trial to compare effects and costs among people with coronary heart disease. JMIR Res Protoc 2020 Jan 27;9(1):e15022. [doi: 10.2196/15022] [Medline: 32012103]
30. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. the CONSORT statement. JAMA 1996 Aug 28;276(8):637-639. [doi: 10.1001/jama.276.8.637] [Medline: 8773637]
31. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014 Mar 7;348:g1687. [doi: 10.1136/bmj.g1687] [Medline: 24609605]
32. American College of Sports Medicine. ACSM's Exercise Testing and Prescription: Wolters Kluwer Health, Lippincott Williams & Wilkins; 2017.
33. Godin Leisure-Time Exercise Questionnaire. Med Sci Sports Exerc 1997 Jun;29(6):36-38. [doi: 10.1097/00005768-199706001-00009]
34. Livingstone KM, McNaughton SA. Diet quality is associated with obesity and hypertension in Australian adults: a cross sectional study. BMC Public Health 2016 Oct 1;16(1):1037. [doi: 10.1186/s12889-016-3714-5] [Medline: 27716133]
35. Australian Government, Department of Health for clarity. Australian dietary guidelines. : National Health and Medical Research Council; 2013 URL: https://www.nhmrc.gov.au/sites/default/files/documents/australian-dietary-guidelines-2013.pdf [accessed 2026-02-14]
36. Park Y, Dodd KW, Kipnis V, et al. Comparison of self-reported dietary intakes from the Automated Self-Administered 24-h recall, 4-d food records, and food-frequency questionnaires against recovery biomarkers. Am J Clin Nutr 2018 Jan 1;107(1):80-93. [doi: 10.1093/ajcn/nqx002] [Medline: 29381789]
37. Subar AF, Kirkpatrick SI, Mittl B, et al. The Automated Self-Administered 24-Hour Dietary Recall (ASA24): a resource for researchers, clinicians, and educators from the National Cancer Institute. J Acad Nutr Diet 2012 Aug;112(8):1134-1137. [doi: 10.1016/j.jand.2012.04.016] [Medline: 22704899]
38. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. Arch Intern Med 1998 Sep 14;158(16):1789-1795. [doi: 10.1001/archinte.158.16.1789] [Medline: 9738608]
39. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. Med Care 1986 Jan;24(1):67-74. [doi: 10.1097/00005650-198601000-00007] [Medline: 3945130]
40. Maxwell A, Özmen M, Iezzi A, Richardson J. Deriving population norms for the AQoL-6D and AQoL-8D multi-attribute utility instruments from web-based data. Qual Life Res 2016 Dec;25(12):3209-3219. [doi: 10.1007/s11136-016-1337-z] [Medline: 27344318]
41. Gao L, Maddison R, Rawstorn J, et al. Economic evaluation protocol for a multicentre randomised controlled trial to compare Smartphone Cardiac Rehabilitation, Assisted Self-Management (SCRAM) versus usual care cardiac rehabilitation among people with coronary heart disease. BMJ Open 2020 Aug 26;10(8):e038178. [doi: 10.1136/bmjopen-2020-038178] [Medline: 32847918]
42. Borm GF, Fransen J, Lemmens WAJG. A simple sample size formula for analysis of covariance in randomized clinical trials. J Clin Epidemiol 2007 Dec;60(12):1234-1238. [doi: 10.1016/j.jclinepi.2007.02.006] [Medline: 17998077]
43. Income and work: census. Australian Bureau of Statistics. 2021. URL: https://www.abs.gov.au/statistics/labour/earnings-and-working-conditions/income-and-work-census/2021 [accessed 2024-07-03]

XSL•FO
RenderX

44. Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H Jr. Publication bias and clinical trials. Control Clin Trials 1987 Dec;8(4):343-353. [doi: 10.1016/0197-2456(87)90155-3] [Medline: 3442991]
45. Ghisi GLDM, Xu Z, Liu X, et al. Impacts of the COVID-19 pandemic on cardiac rehabilitation delivery around the world. Glob Heart 2021 Jun 10;16(1):43. [doi: 10.5334/gh.939] [Medline: 34211829]
46. Champion S, Clark RA, Tirimacco R, Tideman P, Gebremichael L, Beleigoli A. The impact of the SARS-CoV-2 virus (COVID-19) pandemic and the rapid adoption of telehealth for cardiac rehabilitation and secondary prevention programs in rural and remote Australia: a multi-method study. Heart Lung Circ 2022 Nov;31(11):1504-1512. [doi: 10.1016/j.hlc.2022.07.006] [Medline: 35987722]
47. Cartledge S, Thomas EE, Murphy B, et al. Impact of early COVID-19 waves on cardiac rehabilitation delivery in Australia: a national survey. Heart Lung Circ 2023 Mar;32(3):353-363. [doi: 10.1016/j.hlc.2022.12.008] [Medline: 36646580]
48. McHale S, Neubeck L, Rowat A, Dawkes S, Hanson CL. Understanding cardiac rehabilitation delivery in Scotland during the COVID-19 pandemic: lessons for the future. Br J Card Nurs 2023 Aug 2;18(8):1-16. [doi: 10.12968/bjca.2023.0035]
49. O'Doherty AF, Humphreys H, Dawkes S, et al. How has technology been used to deliver cardiac rehabilitation during the COVID-19 pandemic? An international cross-sectional survey of healthcare professionals conducted by the BACPR. BMJ Open 2021 Apr 20;11(4):e046051. [doi: 10.1136/bmjopen-2020-046051] [Medline: 33879492]
50. Gallegos-Rejas VM, Rawstorn JC, Gallagher R, Mahoney R, Thomas EE. Key features in telehealth-delivered cardiac rehabilitation required to optimize cardiovascular health in coronary heart disease: a systematic review and realist synthesis. Eur Heart J Digit Health 2024 May;5(3):208-218. [doi: 10.1093/ehjdh/ztad080] [Medline: 38774382]
51. Rawstorn JC, Subedi N, Koorts H, et al. Stakeholder perceptions of factors contributing to effective implementation of exercise cardiac telerehabilitation in clinical practice. Eur J Cardiovasc Nurs 2025 Jan 30;24(1):116-125. [doi: 10.1093/eurjcn/zvae127] [Medline: 39352400]
52. Hwang R, Fan T, Bowe R, et al. Home-based and remote functional exercise testing in cardiac conditions, during the COVID-19 pandemic and beyond: a systematic review. Physiotherapy 2022 Jun;115:27-35. [doi: 10.1016/j.physio.2021.12.004] [Medline: 35180642]
53. Byrne RA, Rossello X, Coughlan JJ, et al. 2023 ESC Guidelines for the management of acute coronary syndromes. Eur Heart J 2023 Oct 12;44(38):3720-3826. [doi: 10.1093/eurheartj/ehad191] [Medline: 37622654]
54. The BACPR standards and core components for cardiovascular disease prevention and rehabilitation 2023. : British Cardiovascular Society, British Association for Cardiovascular Prevention and Rehabilitation; 2023 URL: https://www.cardiacrehabilitation.org.uk/site/docs/BACPR-Standards-and-Core-Components-2023.pdf [accessed 2026-02-14]
55. Correction to: 2023 ESC Guidelines for the management of acute coronary syndromes: developed by the task force on the management of acute coronary syndromes of the European Society of Cardiology (ESC). Eur Heart J Acute Cardiovasc Care 2024 May 28;13(5):455. [doi: 10.1093/ehjacc/zuad156] [Medline: 38383063]
56. Verdicchio C, Freene N, Hollings M, et al. A clinical guide for assessment and prescription of exercise and physical activity in cardiac rehabilitation. A CSANZ position statement. Heart Lung Circ 2023 Sep;32(9):1035-1048. [doi: 10.1016/j.hlc.2023.06.854] [Medline: 37516652]
57. Hannan EL. Randomized clinical trials and observational studies: guidelines for assessing respective strengths and limitations. JACC Cardiovasc Interv 2008 Jun;1(3):211-217. [doi: 10.1016/j.jcin.2008.01.008] [Medline: 19463302]
58. Perlis RH, Haneuse S, Rubenfeld GD, Fihn SD, Rivara FP. Reporting clinical studies affected by the COVID-19 pandemic: guidelines for authors. JAMA Netw Open 2021 Jan 4;4(1):e2036155. [doi: 10.1001/jamanetworkopen.2020.36155] [Medline: 33443576]

Abbreviations

CHD: coronary heart disease
CVD: cardiovascular disease
HRQoL: health-related quality of life
REDCap: Research Electronic Data Capture
REMOTE-CR: remote exercise monitoring trial for exercise-based cardiac rehabilitation
RR: risk ratio
SCRAM: Smartphone Cardiac Rehabilitation, Assisted Self-Management
VO_2max: maximal aerobic exercise capacity

XSL•FO
RenderX

Edited by S Brini; submitted 03.Sep.2024; peer-reviewed by P Dunn, R Rahman; accepted 09.Jan.2026; published 17.Mar.2026.

Please cite as:
Maddison R, Subedi N, Li P, Lamb KE, Ball K, Oldenburg B, Chow C, McNaughton SA, George ES, Gao L, Moodie M, Amerena J, Nadurata V, Neil C, Cameron S, Rawstorn JC
Smartphone Cardiac Rehabilitation, Assisted Self-Management (SCRAM) Versus Usual Care: Multicenter Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e66074
URL: https://mhealth.jmir.org/2026/1/e66074
doi:10.2196/66074

© Ralph Maddison, Narayan Subedi, Peixuan Li, Karen E Lamb, Kylie Ball, Brian Oldenburg, Clara Chow, Sarah A McNaughton, Elena S George, Lan Gao, Marj Moodie, John Amerena, Voltaire Nadurata, Christopher Neil, Stuart Cameron, Jonathan C Rawstorn. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 17.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effect of a Comprehensive Mobile-Based Respiratory Training Program on Respiratory Function in Survivors of Acute Stroke: Randomized Controlled Trial

Zong-Ke Ma[1,2,3*], BSc; Han-Hong Jiang[1,2*], MSc; Yan-Hua Tang[3], MMed; Jia Yang[3], BSc; Qing-Chuan Wei[1,2], MSc; Qiang Gao[1,2], PhD

[1]Rehabilitation Medicine Center and Institute of Rehabilitation Medicine, West China Hospital of Sichuan University, No.37 Guoxue Lane, Wuhou District, Chengdu, China

[2]Key Laboratory of Rehabilitation Medicine in Sichuan Province, West China Hospital of Sichuan University, Chengdu, China

[3]Department of Rehabilitation Medicine, Sichuan Mianyang 404 Hospital, Mianyang, China

[*]these authors contributed equally

Corresponding Author:
Qiang Gao, PhD
Rehabilitation Medicine Center and Institute of Rehabilitation Medicine, West China Hospital of Sichuan University, No.37 Guoxue Lane, Wuhou District, Chengdu, China

Abstract

Background: Respiratory dysfunction frequently occurs during the acute phase of stroke and is associated with reduced ventilatory capacity, respiratory muscle weakness, and increased pulmonary complications. However, delivering standardized respiratory training during hospitalization is often constrained by staffing and service continuity.

Objective: This study aimed to evaluate the efficacy, safety, and feasibility of a hospital-based comprehensive mobile-based respiratory training program (CMRTP) added to conventional rehabilitation in people with acute stroke who are inpatients.

Methods: This single-center, assessor-blinded randomized controlled trial enrolled 40 patients within 2 weeks after stroke onset with respiratory dysfunction (forced vital capacity <80% predicted). Participants were randomized (1:1) to CMRTP plus conventional rehabilitation or conventional rehabilitation alone. The CMRTP was delivered via the WeChat-based AIRHUB platform and performed 20 minutes twice daily, 5 days per week for 2 weeks, either independently or with caregiver assistance as needed. The primary outcome was change in forced vital capacity from baseline to week 2. Secondary outcomes included forced expiratory volume in 1 second (FEV_1), peak expiratory flow, maximal inspiratory pressure, maximal expiratory pressure, and modified Barthel index. All outcomes were assessed face-to-face by a blinded senior physician, and all analyses followed an intention-to-treat principle.

Results: Of 56 screened patients, 40 were randomized, and 39 completed the study. Adherence to the CMRTP reached 96%, and no serious adverse events occurred; mild, transient events (fatigue, dizziness, and hyperventilation) were recorded. Compared with the control group, the CMRTP group demonstrated greater improvement in forced vital capacity at week 2 (mean difference 0.77 L; 95% CI 0.39 - 1.16; P<.001; η^2=0.32), with additional between-group differences in maximal inspiratory pressure (P=.001; η^2=.25), maximal expiratory pressure (P<.001; η^2=.08), and modified Barthel index (P=.001; η^2=.26). No significant group differences were found for forced expiratory volume in 1 second or peak expiratory flow.

Conclusions: A 2-week hospital-based mobile respiratory training program is feasible and safe in people with acute stroke who are inpatients and yields clinically meaningful improvements in respiratory function and daily functional performance when added to conventional rehabilitation.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2400088647; https://www.chictr.org.cn/showprojEN.html?proj=211239

(JMIR Mhealth Uhealth 2026;14:e78637) doi:10.2196/78637

KEYWORDS

acute stroke; mobile-based; respiratory function; respiratory muscle training; functional capacity

XSL•FO
RenderX

Introduction

Stroke is the second-leading cause of death and the third-leading cause of disability worldwide, imposing a substantial burden on society and families [1]. Advances in acute stroke management have improved survival rates; however, most survivors experience persistent functional impairments [2]. In addition to motor, sensory, swallowing, and speech impairments, stroke frequently induces respiratory dysfunction during the acute stage [3]. Recent studies indicate that 18% to 88% of patients with stroke exhibited abnormal breathing patterns [4], while 44% to 90% present with respiratory failure syndrome of varying severity [5]. Furthermore, 7% to 38% develop pneumonia within the first week after the onset of acute stroke [6], underscoring the urgent need for effective interventions to improve respiratory function.

Clinical guidelines advocate early respiratory monitoring and intervention to mitigate poststroke respiratory complications, emphasizing an integrated care that includes physical and respiratory rehabilitation [7]. Previous studies have demonstrated that respiratory muscle training and breathing exercises, such as diaphragmatic, air-stacking, and pursed-lip breathing, can improve respiratory muscle strength, pulmonary function, and functional capacity during early stroke rehabilitation [8-12]. However, the implementation of respiratory interventions in stroke rehabilitation remains limited by multiple clinical constraints, including insufficient staffing and inconsistent service delivery [13-16]. In many health care settings, rehabilitation services are delayed or underused due to the prioritization of acute medical management, leaving survivors with stroke with inadequate support for respiratory dysfunction [17-19]. In this context, telerehabilitation may represent a promising alternative for delivering timely and accessible respiratory rehabilitation interventions.

Telerehabilitation has emerged globally as a promising approach to improving access to and continuity of stroke rehabilitation services, attracting widespread attention [20,21]. Chen et al [22] reported that telerehabilitation has been widely applied in stroke rehabilitation and provided effects comparable to conventional rehabilitation in improving motor function among survivors with stroke. Linder et al [23] demonstrated that mobile-based exercise interventions can effectively enhance quality of life and alleviate depressive symptoms among individuals recovering from stroke. In addition, Sun et al [24] found that telerehabilitation can effectively reduce caregiver burden. Moreover, a growing body of evidence from high-income countries supports the feasibility, safety, and effectiveness of telerehabilitation in the postacute and chronic phases of stroke recovery [25-27]. However, despite these encouraging findings, the application of telerehabilitation during the acute hospitalization phase of stroke remains largely unexplored.

Therefore, this study aimed to evaluate the effectiveness of a 2-week hospital-based comprehensive mobile-based respiratory training program (CMRTP) on respiratory function in patients with acute stroke, and to explore its safety and feasibility for early inpatient rehabilitation. We hypothesized that integrating CMRTP with conventional rehabilitation would yield greater improvements in respiratory function, without increasing the incidence of adverse events.

Methods

Design

We conducted an assessor-blinded, parallel group, randomized controlled trial at Sichuan Mianyang 404 Hospital (Mianyang, Sichuan, China) from September 2024 to March 2025. This study was prospectively registered with the Chinese Clinical Trial Registry (ChiCTR2400088647). The study adhered to the CONSORT (Consolidated Standards of Reporting Trials) statement guidelines (Checklist 1) [28], and all procedures complied with the Declaration of Helsinki.

Participants

Participants were recruited from the inpatient departments of Rehabilitation Medicine, Neurology, and Neurosurgery through electronic medical records screening. A physician (GQ) screened potential participants based on the eligibility criteria.

Participants were eligible if they met the following criteria: (1) aged 18 - 75 years; (2) diagnosed with stroke according to the Diagnostic Criteria of Cerebrovascular Diseases [29]; (3) first-ever unilateral ischemic stroke confirmed by neuroimaging; (4) stroke onset within 2 weeks; (5) respiratory dysfunction, defined as forced vital capacity (FVC) <80% predicted value [30]; and (6) either the participants or their caregivers were proficient in reading Chinese characters and in operating a smartphone or tablet.

Participants were excluded if they met any of the following criteria: (1) cognitive deficits (Mini-Mental Score Examination scores for patients with stroke: illiterate <17 points, primary school education <20 points, and junior high school education and above <24 points) or inability to understand the instructions provided by researchers; (2) having other diseases, surgeries, or injuries that may interfere with respiratory function training (like severe lung infections, tracheostomy intubation, myasthenia gravis, phrenic nerve paralysis, severe facial paralysis, oropharyngeal structural abnormalities, a history of chest or abdominal surgery or rib fractures within the past year); or (3) refuse or withdraw from this research.

Randomization and Blinding

Patients were randomly assigned in a 1:1 ratio to either the experimental group (the CMRTP plus conventional rehabilitation) or control group (conventional rehabilitation only) using a computer-generated randomization sequence (IBM SPSS version 29). Allocation concealment was ensured by a research assistant, who was not involved in sequence generation or participant enrollment, using sequentially numbered, sealed, opaque envelopes. Each envelope was opened only after a participant had been formally enrolled, and the assigned intervention was then revealed. Outcome assessors and data analysts were blinded to group allocation throughout the trial.

XSL•FO
RenderX

Intervention

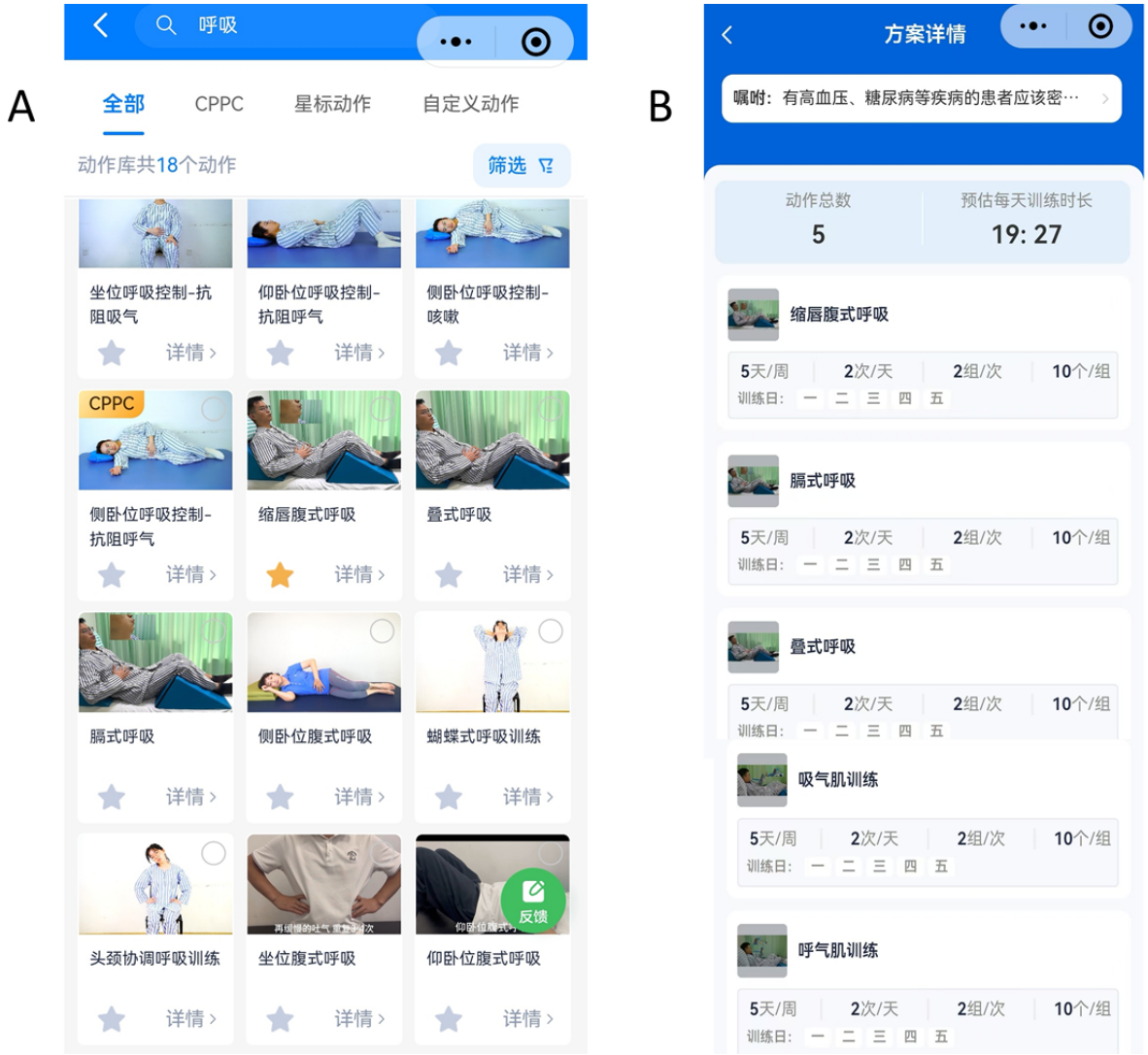
Overview

Patients in the experimental group received the CMRTP plus conventional rehabilitation, while the control group received conventional rehabilitation alone, without specific respiratory training. Conventional rehabilitation was administered by two licensed physiotherapists for 40 minutes per session, once daily, 5 days per week, over a 2-week period. The CMRTP was delivered via the AIRHUB platform (a mobile WeChat-based applet, AIRHUB TECH) by two licensed physiotherapists (JY and Z-KM), with 20 minutes per session, twice daily, 5 days per week for 2 weeks.

The CMRTP Intervention

The AIRHUB platform is a WeChat Applet–based telerehabilitation system consisting of a cloud-based server, a therapist interface (Figure 1), and a patient interface (Figure 2). The therapist interface provides access to an exercise video library, remote prescription and monitoring modules, and real-time feedback collection, while the patient interface allows video-guided exercises, automatic performance recording, and submission of postsession feedback. All data are synchronized securely to the cloud, enabling therapists to review adherence and feedback. The platform supports multiuser access and modular expansion, allowing integration with other rehabilitation programs and scalability to multicenter use. To protect privacy, all patient data are encrypted, deidentified, and accessible only to authorized personnel, in accordance with data protection standards aligned with the principles of the HIPAA (Health Insurance Portability and Accountability Act).

Figure 1. Therapist interface of the AIRHUB platform. (A) Video library and (B) comprehensive mobile-based respiratory training program.



XSL•FO
RenderX

Figure 2. Patient interface of the AIRHUB platform (A: exercise plan; B: feedback submission).

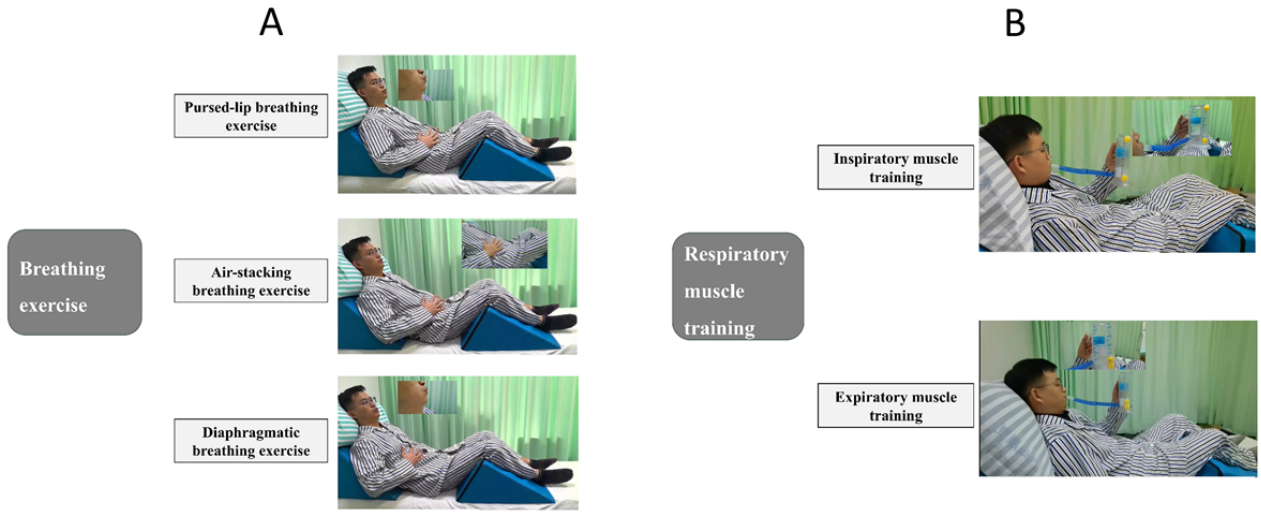


In the experimental group, two licensed physiotherapists (YJ and Z-KM) provided guidance to patients or their caregivers on the proper use of the AIRHUB platform for 2 hours, including the login process, conducting respiratory training, and performing movements correctly as demonstrated in the instructional videos prior to the intervention. After confirming that participants had achieved adequate training proficiency, the physiotherapists guided them to register a personal account on the patient interface using their own mobile phone number. The physiotherapist then logged into the therapist interface, selected the corresponding CMRTP videos from the motion library, and sent them to the individual patient interface.

The CMRTP consisted of 5 exercises, including breathing exercises and respiratory muscle training (Figure 3; details are mentioned in Multimedia Appendix 1), each accompanied by audio cues and written instructions. Each exercise was performed 5 times, with a 20-second rest interval between each exercise. After completing all exercises, the entire plan was repeated once. During the intervention, participants accessed the CMRTP via the patient interface, where they followed standardized demonstration videos and audio instructions to perform the exercises synchronously, and the playback speed could be adjusted as needed. Upon completion, the patient provided feedback through a platform-based questionnaire, and the therapist supervises the training remotely via WeChat or mobile video calls as they needed.

XSL•FO
RenderX

Figure 3. Breathing exercises and respiratory muscle training in AIRHUB. (A) Breathing exercises and (B) respiratory muscle training.



To ensure adherence and track progress, two licensed physiotherapists (YJ and Z-KM) reviewed the training status, such as session duration and questionnaire responses, three times weekly via the AIRHUB platform to evaluate overall completion. When potential issues were identified during monitoring, the responsible physiotherapist would promptly reach out to the participant and offer support. Moreover, the primary nurse provided supervision and reminders to patients and caregivers. If a patient did not engage for three consecutive days, the system automatically triggered a SMS text message reminder. In cases where no training activity was recorded thereafter, the physiotherapist conducted a follow-up telephone call to identify barriers and encourage continued participation.

Conventional Rehabilitation

Conventional rehabilitation [8] was provided by 2 licensed physiotherapists (QCW and HHJ), including the following exercises: (1) joint range of motion exercises in bed, (2) maintaining proper limb positioning in bed, (3) turning exercises, (4) positional transfer training, (5) sitting exercises in bed, and (6) standing exercise (if feasible). If necessary, occupational and speech therapy may also be included.

Assessments and Outcomes

All baseline evaluations were independently performed by a research coordinator after participant enrollment. Primary and secondary outcomes were assessed at baseline (T0), week 1 (T1), and week 2 (T2) following randomization. Outcome assessments were completed face-to-face by a senior physician (HHJ).

Primary Outcome

The primary outcome was the change in FVC from baseline to week 2. FVC, defined as the total volume of air exhaled during a maximal forced expiration following a full inspiration, was used to evaluate overall improvements in respiratory function for patients with acute stroke [31,32]. The minimal clinically important difference for FVC was preset at a 5% predicted increase [33,34]. FVC was measured in liters using standardized spirometry (Breath Home, Home Sun Int.) in accordance with American Thoracic Society and European Respiratory Society standard guidelines [35-38]. FVC was measured three times per assessment, with the highest value used for analysis.

Secondary Outcomes

The secondary outcomes included forced expiratory volume in 1 second (FEV_1) and peak expiratory flow (PEF) to assess airflow dynamics [39], maximal inspiratory pressure (MIP) and maximal expiratory pressure (MEP) to measure respiratory muscle strength [40], and the modified Barthel Index (MBI) to evaluate functional performance in daily activities [41]. Feasibility (adherence rate) and safety (adverse events) were also examined as secondary endpoints. FEV_1, PEF, MIP, and MEP were measured using the same spirometry device (Breath Home), and MBI was assessed using a validated scale by the same physician following standardized procedures.

Sample Size

The sample size calculation was conducted via G*power of 3.1.9.7 based on the result of the FVC in a published similar study [33,42], which indicated an estimated effect size of f=0.27. Other parameters were set as follows: a significance level of α=.05 (two tails), power (1–β)=95%, correlation among repeated measures=0.5, nonsphericity correction ε=1, number of measurements=3, and number of groups=2. Therefore, a sample size of n=32 was obtained. After allowing for a 20% attrition, a minimum total of 40 participants was needed.

Statistical Analysis

Baseline assessments included patient characteristics, FVC, FEV_1, PEF, MIP, MEP, and MBI. All data were analyzed using an intention-to-treat approach, with missing values imputed via the last observation carried forward method.

Statistical analyses used IBM SPSS version 29 (IBM Corp). The Shapiro-Wilk test assessed data normality, and the Levene test was applied to assess the homogeneity of variances. Continuous variables were reported as mean and SD values, ordinal variables as median with IQR values, and categorical variables as numbers with percentages. Baseline characteristics

XSL•FO
RenderX

were compared using *t* tests for normally distributed continuous variables, Mann-Whitney *U* tests for nonnormally distributed or ordinal variables, and chi-square tests for categorical variables. Repeated measures analysis of covariance (ANCOVA) evaluated between-group differences across time points, adjusting for baseline values, with least squares mean differences (least squares [LS] mean differences) estimated to quantify treatment effects. Fixed effects included group, time, and time× group interaction, with participant and National Institutes of Health Stroke Scale (NIHSS) as a random effect. The Mauchly test assessed sphericity, with Greenhouse-Geisser correction applied if violated. Bonferroni correction was applied only to unplanned post hoc pairwise comparisons. The primary outcome was not adjusted for multiple comparisons. For secondary outcomes, Bonferroni correction was applied to control the risk of type I error. Significance was set at $P<.05$ (2-tailed) for the primary outcome and $\alpha<.025$ (2-tailed) for others.

Ethical Considerations

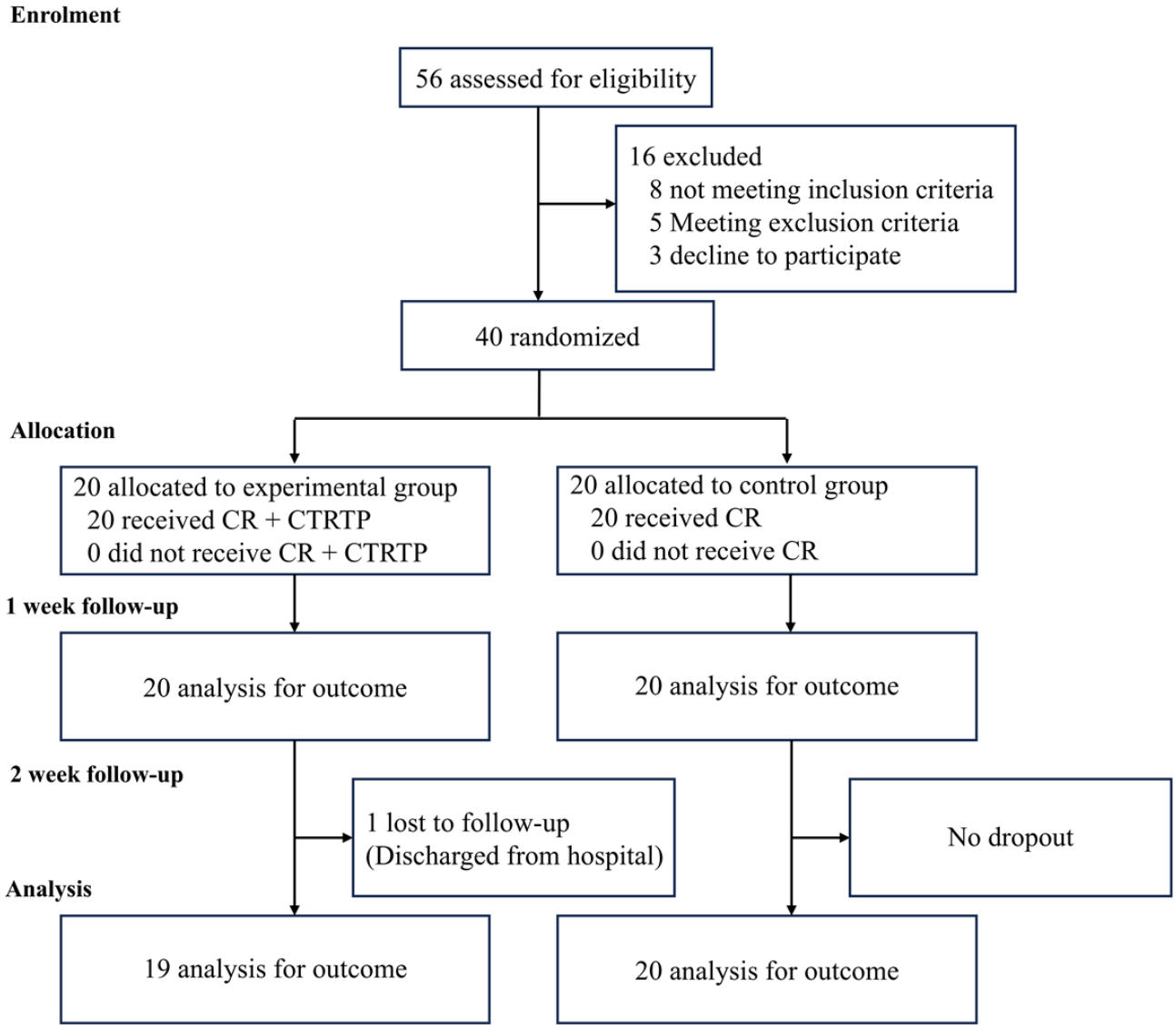
The study was approved by the Ethics Committee of Sichuan Mianyang 404 Hospital (ethics reference: 2023‐051). Written informed consent was obtained from all participants prior to enrollment after they had received a full explanation of the study procedures. All data were anonymized to protect participant privacy and confidentiality. No financial compensation was provided to participants. To ensure participants' safety, all potential adverse events were predefined and explicitly described in the informed consent form. Adverse events were monitored by the physiotherapists responsible for the intervention in the experimental group. Potential adverse events associated with tele-respiratory training included fatigue, tiredness, dizziness, headache, falls, worsening neurological function, hyperventilation, hypoventilation, decreased oxygen saturation, hypertension, and tachycardia. If any adverse events occur during the intervention, caregivers or primary nurses would immediately inform the patient's attending physician and take appropriate measures. All adverse events associated with conventional rehabilitation or CMRTP were recorded on the case report form for final analysis.

Results

A total of 56 patients were screened for eligibility, and 40 were enrolled in the study. Of these, 20 were randomly assigned to the intervention group and 20 to the control group (Figure 4). One participant in the intervention group withdrew during the second week due to hospital discharge. Regarding treatment adherence, participants in the experimental group completed 192 of the 200 prescribed training sessions (96%), whereas those in the control group completed all 100 assigned sessions (100%). Outcome data for analysis were available for 100% (N=40) of participants at T1 and 97.5% (39/40) at T2. Nonserious adverse events considered reasonably or definitely related to study procedures were recorded in 10 sessions in the experimental group (including fatigue, dizziness, and hyperventilation) and in 3 sessions in the control group (including fatigue and tiredness). All events were transient and mild, and no serious adverse events occurred in either group.

XSL•FO
RenderX

Figure 4. The CONSORT (Consolidated Standards of Reporting Trials) flow diagram of the trial. CR: conventional rehabilitation; CMRTP: comprehensive mobile-based respiratory training program.



The baseline clinical characteristics of the patients, shown in Table 1, were generally comparable between the experimental and control groups. All included patients had moderate stroke with notable functional impairments affecting daily activities. The mean (SD) age was 62.5 (SD 10.57) years. The experimental group exhibited slightly greater stroke severity than the control group (mean NIHSS score 7.30, SD 1.75 vs 5.75, SD 1.97; P=.01), while other clinical features were comparable between groups.

XSL•FO
RenderX

Table . Clinical characteristics of the patients at baseline.

Characteristics	Intervention (n=20)	Control (n=20)	Test statistic	*P* value
Gender, n (%)			0.45 (1.00)[a]	.50
Male	15 (75)	12 (60)		
Female	5 (25)	8 (40)		
Age (years), mean (SD)	59.80 (11.31)	65.20 (9.83)	1.61 (37.29)[b]	.12
Course of disease (day), mean (SD)	6.00 (2.17)	6.20 (2.31)	0.28 (37.87)[b]	.78
BMI[c] (kg/m^2), mean (SD)	25.25 (2.57)	24.48 (3.45)	0.80 (35.12)[b]	.43
NIHSS[d,e]	7.30 (1.75)	5.75 (1.97)	2.63 (37.48)[b]	.01
MMSE[f]	25.35 (1.89)	25.05 (1.79)	0.56 (37.79)[b]	.58
Stroke type, n (%)			0.27 (1.00)[a]	.59
Ischemic stroke	19 (95)	17 (85)		
Hemorrhagic stroke	1 (5)	3 (15)		
Affected side, n (%)			0.00 (1.00)[a]	>.99
Left	13 (65)	13 (65)		
Right	7 (35)	7 (35)		
Coexisting conditions, n (%)				
Hypertension	14 (70)	15 (75)	0.00 (1.00)[a]	>.99
Diabetes mellitus	5 (25)	9 (45)	0.98 (1.00)[a]	.32
Pulmonary infection	0 (0)	1 (5)	0.00 (1.00)[a]	>.99
COPD[g]	2 (10)	0 (0)	0.52 (1.00)[a]	.46
Smoking	0 (0)	1 (5)	0.10 (1.00)[a]	.75

[a]Chi-square (df).

[b]*t* test (df).

[c]The BMI is the weight in kilograms divided by the square of the height in meters.

[d]Independent-samples *t* tests ($P<.05$).

[e]NIHSS: National Institutes of Health Stroke Scale.

[f]MMSE: Mini-Mental State Examination.

[g]COPD: chronic obstructive pulmonary disease.

For the primary outcome, the analysis of FVC revealed significant main effects in both the time ($F_{1,37}=41.12$; $\eta^2=0.52$; $P<.001$) and time×group interaction effect ($F_{1,37}=4.50$; $\eta^2=0.12$; $P=.008$), without group effect ($F_{1,37}=4.05$; $\eta^2=0.10$; $P=.05$; Table 2). The FVC at week 2 increased by 1.12 L (95% CI 0.43 - 1.81) in the experimental group and 0.58 L (95% CI 0.02 - 1.14) in the control group, a significant difference was observed between the groups at week 2 (LS mean 0.77, 95% CI 0.39 - 1.16; $\eta^2=0.32$; $P<.001$; Figure 5). The improvements of both groups exceeded the minimal clinically important difference (5%), but the experimental group showed a greater improvement (Table 3).

XSL•FO
RenderX

Table . Summary of main effects and interaction effects from repeated measures analysis of covariance.

Outcome	*Group effect, F* test (*df*)	*P* value (Group)	η^2 (Group)	Time effect, *F test* (*df*)	*P* value (Time)	η^2 (Time)	Group×time interaction, *F* test (*df*)	*P* value (Interaction)	η^2 (Interaction)
FVC[a] (L)	4.05 (1,37)	.05	0.10	41.12 (1,37)	<.001[b]	0.52	4.50 (1,37)	.008[c]	0.12
FEV_1 [d] (L)	2.21 (1,37)	.15	0.06	6.86 (0.83,30.71)	.004[c]	0.16	1.25 (0.83,30.71)	.29	0.03
PEF[e] (L/s)	1.89 (1,37)	.18	0.05	1.37 (0.83,30.71)	.008[c]	0.14	1.66 (0.83,30.71)	.20	0.04
MIP[f] (cmH_2O)	14.49 (1,37)	<.001[b]	0.28	7.66 (1,37)	<.001[b]	0.17	7.28 (1,37)	.001[c]	0.16
MEP[g] (cmH_2O)	10.54 (1,37)	.002[c]	0.20	1.37 (0.83,30.71)	.26	0.04	4.62 (0.83,30.71)	.02[h]	0.11
MBI[i] (score)	7.56 (1,37)	.009[c]	0.17	10.07 (1,37)	<.001[b]	0.21	3.35 (1,37)	.04[h]	0.08

[a]FVC: forced vital capacity.

[b]*P*<.001.

[c]*P*<.01.

[d]FEV_1 : forced expiratory volume in 1 second.

[e]PEF: peak expiratory flow.

[f]MIP: maximal inspiratory pressure.

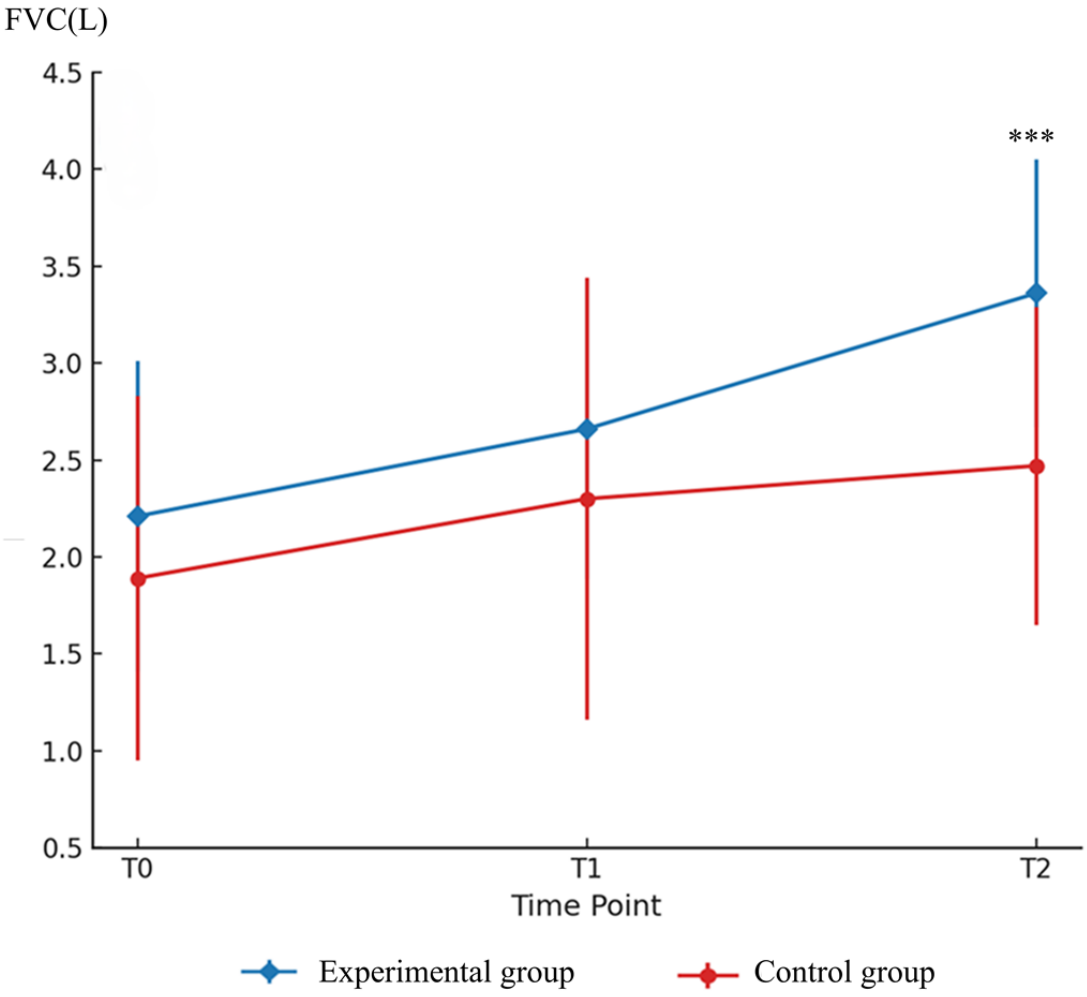
[g]MEP: maximal expiratory pressure.

[h]*P*<.05.

[i]MBI: modified Barthel index.

XSL•FO
RenderX

Figure 5. Between-group temporal changes in forced vital capacity. Mean (± SD) values are shown at baseline (T0), week 1 (T1), and week 2 (T2). The experimental group exhibited greater improvement than the control group at T2 (***P<.001). FVC: forced vital capacity.



XSL•FO
RenderX

Table . Analysis of the effect of time and group on the primary and secondary outcomes.

Test and groups	T0[a]	T1[b,c]	T2[b,d]	ΔT1[e]	ΔT2[f]	*F*	*P* value	η^{2}[g]
FVC[h] (L)								
Experimental group (n=20), mean (SD)	2.21 (0.80)	2.66 (0.77)[i]	3.36 (0.69)[i]	0.45 (0.27)	1.12 (0.69)	38.31	<.001[j]	0.67
Control group (n=20), mean (SD)	1.89 (0.94)	2.30 (1.14)	2.47 (0.82)[i]	0.41 (0.73)	0.58 (0.56)	9.59	.001[k]	0.34
LS[l] means (95% CI)	0.32 (–0.24 to 0.88)	—[m]	—	0.05 (–0.36 to 0.46)	0.77 (0.39 - 1.16)	—	—	—
t (df)	1.15 (38)	—	—	0.27 (24.03)	4.08 (38)	—	—	—
P[n]	0.26	—	—	0.79	<.001[j]	—	—	—
η^{2}	0.03	—	—	0.002	0.32	—	—	—
FEV_1[o] (L)								
Experimental group (n=20), mean (SD)	1.41 (0.55)	1.77 (0.81)	2.44 (0.87)[i]	0.36 (0.71)	1.41 (0.55)	21.48	<.001[j]	0.53
Control group (n=20), mean (SD)	1.38 (0.71)	1.35 (0.72)	2.04 (0.74)[i]	–0.03 (0.88)	0.66 (0.69)	10.74	<.001[j]	0.36
LS means (95% CI)	0.10 (–0.35 to 0.55)	—	—	0.28 (–0.28 to 0.84)	0.36 (–0.15 to 0.88)	—	—	—
t (df)	0.45(38)	—	—	1(38)	1.44(38)	—	—	—
P[n]	0.65	—	—	0.33	0.16	—	—	—
η^{2}	0.006	—	—	0.03	0.05	—	—	—
PEF[p] (L/s)								
Experimental group (n=20), mean (SD)	2.40 (1.08)	2.81 (0.95)	3.71 (1.20)[i]	0.42 (0.84)	1.23 (1.18)	19.09	<.001[j]	0.5
Control group (n=20), mean (SD)	2.34 (1.32)	2.50 (1.19)	2.93 (0.99)	0.16 (1.29)	0.59 (1.14)	2.99	0.07	0.14
LS means (95% CI)	0.20 (–0.65 to 1.05)	—	—	0.27 (–0.37 to 0.92)	0.73 (0.04 - 1.42)	—	—	—
t (df)	0.48 (38)	—	—	0.85 (38)	2.14 (38)	—	—	—
P[n]	0.64	—	—	0.4	0.04	—	—	—
η^{2}	0.006	—	—	0.11	0.11	—	—	—
MIP[q] (cmH_2O)								
Experimental group (n=20), mean (SD)	34.45 (15.64)	45.62 (15.93)[i]	51.84 (15.06)[i]	11.17 (7.10)	16.55 (10.43)	41.48	<.001[j]	0.69
Control group (n=20), mean (SD)	26.50 (11.82)	29.40 (17.42)	35.01 (14.52)[i]	2.90 (10.71)	8.51 (7.64)	8.5	<.001[j]	0.31

XSL•FO
RenderX

Test and groups	T0[a]	T1[b,c]	T2[b,d]	ΔT1[e]	ΔT2[f]	*F*	*P* value	η^{2}[g]
LS means (95% CI)	11.16 (1.75 - 20.56)	—	—	10.83 (4.12 - 17.54)	11.38 (4.69 - 18.08)	—	—	—
t (df)	2.4 (38)	—	—	3.27 (38)	3.45 (38)	—	—	—
P[n]	0.08	—	—	.002	.001**[k]	—	—	—
η^{2}	0.07	—	—	0.23	0.25	—	—	—
MEP[r] (cmH_2O)								
Experimental group (n=20), mean (SD)	37.34 (18.98)	47.85 (24.00)[i]	56.28 (20.15)[i]	10.51 (13.99)	17.83 (12.24)	21.95	<.001[j]	0.53
Control group (n=20), mean (SD)	32.94 (20.26)	35.29 (25.65)	36.25 (16.01)	2.36 (20.46)	3.32 (11.24)	0.48	0.55	0.03
LS means (95% CI)	11.49 (–1 to 23.98)	—	—	11.68 (–1.18 to 24.54)	17.85 (9.92 - 25.77)	—	—	—
t (df)	1.86 (38)	—	—	1.84 (38)	4.57 (38)	—	—	—
P[c]	0.07	—	—	0.07	<.001[j]	—	—	—
η^{2}	0.09	—	—	0.09	0.08	—	—	—
MBI[s] (score)								
Experimental group (n=20), mean (SD)	38.25 (13.47)	53.40 (12.43)[i]	65.15 (10.47)[i]	15.15 (8.75)	26.25 (8.58)	90.42	<.001[j]	0.83
Control group (n=20), mean (SD)	37.35 (10.23)	48.90 (12.83)[i]	55.85 (11.25)[i]	11.55 (7.47)	18.50 (8.02)	18	<.001[j]	0.85
LS means (95% CI)	5.18 (–2.44 to 12.80)	—	—	5.77 (0.20 - 11.34)	9.22 (3.90 - 14.53)	—	—	—
t (df)	1.38 (38)	—	—	2.1 (38)	3.52 (38)	—	—	—
P[n]	0.18	—	—	0.04	.001[k]	—	—	—

XSL•FO
RenderX

Test and groups	T0[a]	T1[b,c]	T2[b,d]	ΔT1[e]	ΔT2[f]	*F*	*P* value	η²[g]
η^2	0.05	—	—	0.11	0.26	—	—	—

[a]T0: baseline.

[b]Within-group comparison in each group with repeated measures analysis of covariance: compare to T0.

[c]1T1: week 1.

[d]T2: week 2.

[e]ΔT1: the change from T0 to T1.

[f]ΔT2: the change from T0 to T2.

[g]η^2: partial eta squared.

[h]FVC: forced vital capacity.

[i]*P<.05.

[j]***P<.001.

[k]**P < .025.

[l]LS: least squares.

[m]Not applicable.

[n]Comparison between-group difference in change with repeated measures analysis of covariance.

[o]FEV_1 : forced expiratory volume in 1 second.

[p]PEF: peak expiratory flow.

[q]MIP: maximal inspiratory pressure.

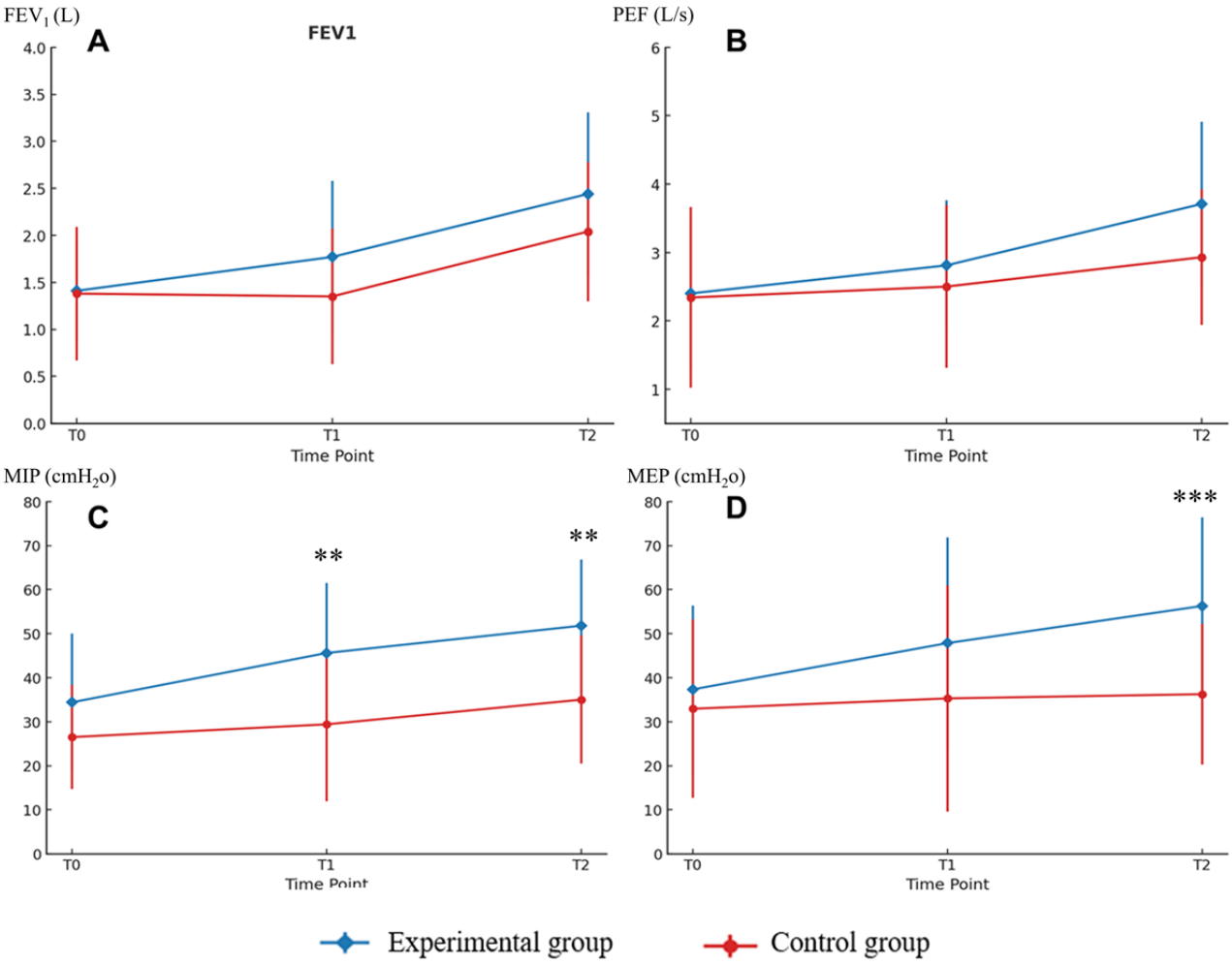
[r]MEP: maximal expiratory pressure.

[s]MBI: modified Barthel index.

As for secondary outcome in Tables 2 and 3, there was no significant main effect of group for FEV_1 ($F_{1,37}$=2.21; η^2=0.06; P=.15), PEF ($F_{1,37}$=1.89; η^2=0.05; P=.18), whereas MIP ($F_{1,37}$=14.49; η^2=0.28; P<.001), MEP ($F_{1,37}$=10.54; η^2=0.20; P=.002), and MBI ($F_{1,37}$=7.56; η^2=.17; P=.009) reached statistical significance. The main effect of time for FEV_1 ($F_{0.83,30.71}$=6.86; η^2=0.16; P=.004), PEF ($F_{0.83,30.71}$=1.37; η^2=0.14; P=.008), MIP ($F_{1,37}$=7.66; η^2=0.17; P<.001), and MBI ($F_{1,37}$=10.07; η^2=0.21; P<.001) showed significant differences, but not on MEP ($F_{0.76,28.12}$=1.37; η^2=0.04; P=.26) at weeks 1 and 2. Significant group×time interactions were found for MIP ($F_{1,37}$=7.28; η^2=0.16; P=.001), MEP ($F_{0.76,28.12}$=4.62; η^2=0.11; P=.02), and MBI ($F_{1,37}$=3.35; η^2=0.08; P=.04), but not for FEV_1 ($F_{0.83,30.71}$=1.25; η^2=.00; P=0.29) and PEF ($F_{0.83,30.71}$=1.66; η^2=0.04; P=.20). Compared with baseline, the experimental group exhibited significant increases in FEV_1 ($F_{1.66,31.54}$=21.48; η^2=0.53; P<.001), PEF ($F_{1.66,31.54}$=19.09; η^2=0.50; P<.001), MIP ($F_{2,38}$=41.48; η^2=0.69; P<.001), MEP ($F_{1.52,28.88}$=21.59; η^2=0.53; P<.001), and MBI ($F_{2,38}$=90.42; η^2=0.83; P<.001), whereas the control group showed significant increases in FEV_1 ($F_{1.66,31.54}$=10.74; η^2=0.36; P<.001), MIP ($F_{2,38}$=8.50; η^2=0.31; P<.001) and MBI ($F_{2,38}$=90.42; η^2=0.85; P<.001), but not in PEF ($F_{1.66,31.54}$=2.99, η^2=0.14; P=.07) and MEP ($F_{1.52,28.88}$=0.48; η^2=0.03; P=.55). Compared with the control group, the experimental group had no baseline differences (P>.05 for FEV_1 , PEF, MIP, MEP, and MBI), and showed significant improvements at week 2 in MIP (LS mean 11.38, 95% CI 4.69 - 18.08; η^2=0.25; P=.001), MEP (LS mean 17.85, 95% CI 9.92 - 25.77; η^2=0.08; P<.001), and MBI (LS mean 9.92, 95% CI 3.90 - 14.53; η^2=0.26; P=.001; Figures 6 and 7), as well as in MIP at week 1 (LS mean 10.83, 95% CI 4.12 - 17.54; η^2=0.23; P=.002). No difference was found between the groups in the FEV_1 and PEF at any assessment time point, which may be due to the short intervention duration and the limited sensitivity of these measures to detect subtle respiratory changes.

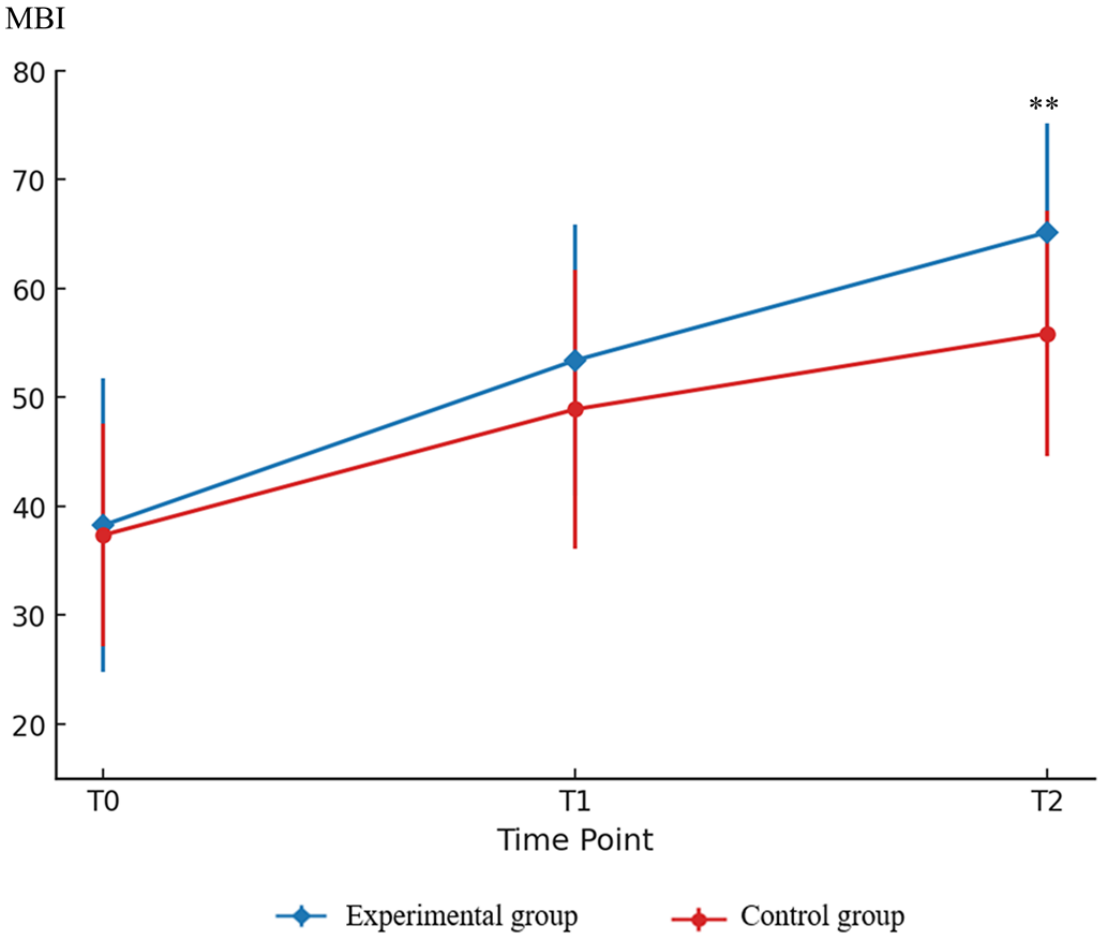
XSL•FO
RenderX

Figure 6. Between-group temporal changes in forced expiratory volume in 1 s, peak expiratory flow, maximal inspiratory pressure, and maximal expiratory pressure. (A) Mean (SD) values are shown at baseline (T0), week 1 (T1), and week 2 (T2). No significant between-group differences were observed in forced expiratory volume in 1 s across all time points. (B) Mean (SD) values are shown at baseline (T0), week 1 (T1), and week 2 (T2). No significant between-group differences were observed in peak expiratory flow across all time points. (C) Mean (SD) values are shown at baseline (T0), week 1 (T1), and week 2 (T2). The experimental group exhibited greater improvement than the control group for maximal inspiratory pressure at T1 (**P=.002) and T2 (**P=.001). D: Mean (SD) values are shown at baseline (T0), week 1 (T1), and week 2 (T2). The experimental group exhibited greater improvement than the control group for maximal expiratory pressure at T2 (***P<.001). FEV1: forced expiratory volume in 1 second; PEF: peak expiratory flow; MIP: maximal inspiratory pressure; MEP: maximal expiratory pressure.



XSL•FO
RenderX

Figure 7. Between-group temporal changes in modified Barthel index. Mean (SD) values are shown at baseline (T0), week 1 (T1), and week 2 (T2). The experimental group exhibited greater improvement than the control group at T2 (**P=.001). MBI: modified Barthel index.



Discussion

This randomized controlled trial demonstrated that a 2-week, hospital-based CMRTP combined with conventional rehabilitation significantly improved respiratory function and daily functional performance compared with conventional rehabilitation alone in patients with moderate acute stroke. Importantly, no serious adverse events were reported, and the mobile-based platform (AIRHUB) was feasible for implementation in inpatient acute stroke care. To our knowledge, this was the first study to establish the efficacy of a mobile-based respiratory program as a safe and effective adjunct to standard care during the acute phase of stroke.

Compared with prior studies [8,43,44], the experimental group demonstrated comparable improvements in respiratory function, respiratory muscle strength, and daily activities in patients with acute stroke. Choi et al [8] reported that a 4-week comprehensive respiratory muscle training program, incorporating air-stacking and respiratory muscle strengthening, led to significant improvements in FVC and MEP in patients with acute stroke. Similarly, Yoo et al [43] found that delivering the same protocol at the bedside over three weeks resulted in increases in FVC, FEV_1, and MBI. In another study, Wei et al [44] found that a 12-week intervention combining respiratory muscle training, pursed-lip breathing, and diaphragmatic breathing improved FVC, FEV_1, respiratory muscle strength, and MBI in patients with stroke.

FVC, as a comprehensive indicator of ventilatory capacity, was particularly sensitive to changes in respiratory muscle performance [43]. Given the prevalence and clinical relevance of respiratory dysfunction in acute stroke, FVC provided an objective and meaningful measure to evaluate the efficacy of respiratory interventions [45]. In this study, both groups showed improvements in FVC after 2 weeks of intervention, but the experimental group achieved greater gains. These intergroup differences were unlikely to be attributed to spontaneous recovery alone but were more plausibly linked to the structured and synergistic design of the CMRTP, encompassing both muscular and neurological mechanisms.

The observed improvements in respiratory function likely resulted from multilevel adaptations involving both central neural remodeling and peripheral muscular strengthening. The acute poststroke phase represented a critical window of heightened neuroplasticity [46], during which axonal sprouting and synaptic reorganization occur at an accelerated rate [47]. Previous studies showed that exercise can facilitate long-term

XSL•FO
RenderX

potentiation–like synaptic plasticity, enhancing functions such as memory and fine motor skills [48,49]. Structured and repetitive respiratory exercises—such as diaphragmatic breathing, pursed-lip breathing, and air-stacking—might elicit similar effects by providing continuous sensory and proprioceptive input to cortical and brainstem respiratory centers, reinforcing sensorimotor pathways and supporting circuit-level functional recovery [50]. In addition to synaptic mechanisms, respiratory training might promote neuroplasticity at the molecular level. Exercise was shown to upregulate the expression of mature brain-derived neurotrophic factor, a key modulator of synaptic strength and corticospinal connectivity [51,52]. These effects might enhance the cortical drive to spinal respiratory motor neurons, facilitating compensatory activation and improving voluntary control over respiratory muscles [53-55].

Peripherally, the CMRTP targeted both inspiratory and expiratory muscle groups through focused protocols. Inspiratory exercises and air-stacking primarily strengthened the diaphragm and intercostal muscles [56], whereas pursed-lip and expiratory resistance breathing enhanced the performance of abdominal and accessory expiratory muscles [57-59]. These interventions likely contributed to the significant improvements in MIP and MEP observed in the intervention group. Repeated mechanical loading and neuromuscular activation might have enhanced motor unit recruitment, firing efficiency, and muscle coordination within the respiratory system [60,61]. Together, these central and peripheral adaptations improved respiratory mechanics and endurance, which might have translated into better exercise tolerance and reduced dyspnea. As a result, participants could engage more effectively in physical rehabilitation and daily activities, reflected in the observed improvements in MBI scores.

Despite these positive findings, no significant between-group differences were observed in FEV_1 and PEF. These indices primarily reflect airway patency and expiratory flow dynamics, which are less responsive to short-term interventions and more sensitive to chronic airway function and expiratory control [62]. The 2-week duration of training may have been insufficient to elicit detectable changes in these parameters. Moreover, poststroke impairment in airway coordination—such as reduced glottic control or weakened expiratory reflexes—may further limit responsiveness to early-phase interventions [63,64]. Longer or more targeted expiratory-focused protocols may be needed to achieve improvements in these measures.

Compared with other motor telerehabilitation studies, our findings highlight the value of a simplified, focused intervention suitable for patients with acute stroke. Previous programs targeting upper limb [65], spinal stability [66], or core function [67] have shown functional gains, but often required long durations (8‐12 wk), complex protocols, or real-time supervision, which may limit adherence or scalability. Some participants in these studies reported training fatigue, reduced engagement, or low compliance [66,67]. In contrast, our 2-week respiratory protocol achieved a high adherence rate (96%) with minimal supervision, supported by an app-based system that provided clear instructions and immediate feedback. Rather than using videoconferencing or immersive visual reality [25,66,68], which required constant therapist involvement, our approach relied on routine caregiver support, enhancing feasibility in clinical settings. By focusing on a single, targeted domain—respiratory recovery—our program provided a practical and scalable model for early-phase telerehabilitation in stroke care.

Despite these promising findings, the results must be interpreted in the context of certain limitations. The intervention period was limited to two weeks due to constraints imposed by the Chinese health insurance system, whereas previous studies suggested that respiratory training typically spans at least 4 weeks to maximize benefit [8,59,69]. To address this, the CMRTP was intentionally designed as a high-intensity, twice-daily, multicomponent protocol targeting multiple facets of respiratory function simultaneously. Although the improvements observed in FVC were smaller than those reported in longer trials—for example, a recent meta-analysis reported an average FVC increase of 0.87 L in early stroke [9]—our findings suggest that a condensed, intensive approach could still yield meaningful gains, particularly for patients in the acute phase of stroke.

The AIRHUB platform significantly facilitated the delivery of this program by overcoming temporal and spatial barriers that often impede rehabilitation in the acute phase. It mitigated treatment disruptions caused by limited clinical staff availability or urgent medical needs, ensuring continuous and stable access to respiratory training. Its user-friendly interface, guided multimedia instructions, and integration with nursing oversight enabled patients to complete high-quality training sessions independently and consistently. The hybrid model, combining digital support with in-hospital supervision, likely contributed to the high adherence rate (97.5%) and absence of adverse events. Compared to home-based rehabilitation models, this approach provided greater control over intervention fidelity, which was critical in early-phase recovery.

Nevertheless, several limitations should be acknowledged. First, although outcome assessors were blinded, patient awareness of group allocation might have introduced performance bias, particularly for effort-dependent measures such as MIP and MEP. Second, the relatively short duration of the intervention limited the interpretation of long-term effects. Third, the modest sample size (N=40), while sufficient to detect primary effects, might reduce generalizability and statistical power to detect smaller but clinically meaningful differences.

Recent advances in respiratory rehabilitation included virtual reality biofeedback systems that visualized and quantified respiratory data, offering more effective and engaging training experiences [70,71]. Devices such as Acapella combined with the active cycle of breathing technique improved lung function in perioperative patients with lung cancer [72], and mobile-based intelligent trainers like AeroFit IMT enhanced respiratory muscle strength without inducing fatigue [73]. While these technologies showed promise, their use often required specialized equipment and supervision, which might limit feasibility in acute stroke care [74]. Therefore, incentive spirometry was chosen in this study for its safety, low cost, ease of training, and visual feedback–facilitated adherence [74,75].

XSL•FO
RenderX

Future studies should investigate longer interventions, larger multicenter cohorts, and the integration of advanced or digitally enhanced respiratory devices. Comparative studies of different training intensities, configurations, and delivery models, including home-based telerehabilitation, will be essential to optimize individualized protocols and evaluate scalability in diverse clinical settings.

This randomized controlled trial demonstrated that a 2-week, hospital-based CMRTP combined with conventional rehabilitation significantly improved respiratory function and daily functional performance in patients with moderate acute stroke. The intervention was safe, well-tolerated, and achieved high adherence, suggesting good feasibility for inpatient application. These findings support the integration of mobile-based respiratory training as an effective adjunct to conventional hospital rehabilitation for early respiratory recovery after stroke. Larger multicenter studies with extended follow-up are needed to confirm the long-term benefits and cost-effectiveness of this approach.

Acknowledgments

We sincerely thank the researchers involved who work at the Departments of Rehabilitation Medicine, Neurology, and Neurosurgery of Sichuan Mianyang 404 Hospital and Department of Rehabilitation Medicine of West China Hospital at Sichuan University. The authors affirm that no generative artificial intelligence tools (eg, ChatGPT or similar) were used in the preparation or writing of this manuscript.

Funding

This study was funded by Science and Technology Bureau Achivement Transformation Program of Chengdu (2023-YF09-00044-SN).

Data Availability

The anonymized datasets generated and analyzed during this study are not publicly accessible due to participant confidentiality concerns and institutional data protection policies of Sichuan Mianyang 404 Hospital, but are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization, methodology, validation, and writing-original draft, writing – review and editing: ZKM
Conceptualization, formal analysis, data collection, writing – original draft; writing – review and editing: HHJ
Conceptualization, methodology, recruitment, manuscript review and editing: YHT
Data analysis, visualization, writing – original draft: JY
Data analysis, manuscript review, editing, software, and visualization: QCW
Conceptualization, manuscript review, methodology, supervision, writing – original draft, writing – review and editing: QG

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed intervention protocol and supplementary data tables of outcome measures.
[DOCX File, 28 KB - mhealth_v14i1e78637_app1.docx]

Checklist 1
CONSORT checklist.
[PDF File, 1279 KB - mhealth_v14i1e78637_app2.pdf]

References

1. Feigin VL, Abate MD, Abate YH, et al. Global, regional, and national burden of stroke and its risk factors, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet Neurol 2024 Oct;23(10):973-1003. [doi: 10.1016/S1474-4422(24)00369-7] [Medline: 39304265]
2. Devos D, Sevin M, De Gaalon S, Lintia-Gaultier A, Guillon B. Management of ischemic stroke in the hyperacute phase. PANMINERVA Med 2013 Mar;55(1):59-78. [Medline: 23474664]
3. Deme S, Lamba D, Alamer A, et al. Effectiveness of respiratory muscle training on respiratory muscle strength, pulmonary function, and respiratory complications in stroke survivors: a systematic review of randomized controlled trials. Degener Neurol Neuromuscul Dis 2022;12:75-84. [doi: 10.2147/DNND.S348736] [Medline: 35411199]
4. Rochester CL, Mohsenin V. Respiratory complications of stroke. Semin Respir Crit Care Med 2002 Jun;23(3):248-260. [doi: 10.1055/s-2002-33033] [Medline: 16088617]

XSL•FO
RenderX

5. Chuchalin AG, Ｋim TG, Shogenova LV, Martynov MY, Gusev EI. The syndrome of respiratory failure in acute stroke: new diagnostic approaches. Pul monologiâ (Mosk) 2020;30(4):485-492. [doi: 10.18093/0869-0189-2020-30-4-485-492]
6. Wang X, Xia J, Shan Y, Yang Y, Li Y, Sun H. Predictive value of the Oxford Acute Severity of Illness Score in acute stroke patients with stroke-associated pneumonia. Front Neurol 2023;14:1251944. [doi: 10.3389/fneur.2023.1251944] [Medline: 37731859]
7. Spruit MA, Singh SJ, Garvey C, et al. An official American Thoracic Society/European Respiratory Society statement: key concepts and advances in pulmonary rehabilitation. Am J Respir Crit Care Med 2013 Oct 15;188(8):e13-e64. [doi: 10.1164/rccm.201309-1634ST] [Medline: 24127811]
8. Choi HE, Jo GY, Do HK, On CW. Comprehensive respiratory muscle training improves pulmonary function and respiratory muscle strength in acute stroke patients. J Cardiopulm Rehabil Prev 2021 May 1;41(3):166-171. [doi: 10.1097/HCR.0000000000000526] [Medline: 33027217]
9. Zhang YS, Zhang K, Huang L, et al. The effects of respiratory muscle training on respiratory function and functional capacity in patients with early stroke: a meta-analysis. Eur Rev Aging Phys Act 2024 Feb 22;21(1):4. [doi: 10.1186/s11556-024-00338-7] [Medline: 38383309]
10. Li M, Huang Y, Chen H, Wang S, Zhou Y, Zhang Y. Relationship between motor dysfunction, the respiratory muscles and pulmonary function in stroke patients with hemiplegia: a retrospective study. BMC Geriatr 2024 Jan 13;24(1):59. [doi: 10.1186/s12877-023-04647-x] [Medline: 38218756]
11. Yoon JM, Im SC, Kim K. Effects of diaphragmatic breathing and pursed lip breathing exercises on the pulmonary function and walking endurance in patients with chronic stroke: a randomised controlled trial. Int J Ther Rehabil 2022 Aug 2;29(8):1-11. [doi: 10.12968/ijtr.2021.0027]
12. Lee Y, Kim Y, Kim D. Effects of chest mobilization and breathing exercises on respiratory function, trunk stability, and endurance in chronic stroke patients after coronavirus disease. Medicina (Kaunas) 2023 Dec 15;59(12):2180. [doi: 10.3390/medicina59122180] [Medline: 38138283]
13. Bettger JP, Li Z, Xian Y, et al. Assessment and provision of rehabilitation among patients hospitalized with acute ischemic stroke in China: Findings from the China National Stroke Registry II. Int J Stroke 2017 Apr;12(3):254-263. [doi: 10.1177/1747493017701945] [Medline: 28381197]
14. Aderinto N, Olatunji G, Kokori E, et al. Disparities in post-stroke rehabilitation: an African perspective. Discov Public Health 2025;22(1):218. [doi: 10.1186/s12982-025-00622-6]
15. Handa G, Hazra S, Chalageri PH. WHO Systematic Assessment of Rehabilitation Situation (STARS): a systematic review on the status of stroke rehabilitation in India. Curr Phys Med Rehabil Rep 2023;11(4):410-423. [doi: 10.1007/s40141-023-00418-2]
16. Roth EJ, Lovell L, Harvey RL, Bode RK. Delay in transfer to inpatient stroke rehabilitation: the role of acute hospital medical complications and stroke characteristics. Top Stroke Rehabil 2007;14(1):57-64. [doi: 10.1310/tsr1401-57] [Medline: 17311791]
17. Pongratanakul R, Thitisakulchai P, Kuptniratsaikul V. Factors related to interrupted inpatient stroke rehabilitation due to acute care transfer or death. NeuroRehabilitation 2020;47(2):171-179. [doi: 10.3233/NRE-203187] [Medline: 32716330]
18. Wang XL, Ma LJ, Hu XG, Wang K, Cheng JJ. Application of the respiratory “critical care-sub-critical care-rehabilitation integrated management model” in severe stroke associated pneumonia. BMC Pulm Med 2020 Mar 5;20(1):61. [doi: 10.1186/s12890-020-1100-7] [Medline: 32138782]
19. Green TL, McNair ND, Hinkle JL, et al. Care of the patient With acute ischemic stroke (posthyperacute and prehospital discharge): Update to 2009 Comprehensive Nursing Care Scientific Statement: a scientific statement from the American Heart Association. Stroke 2021 May;52(5):e179-e197. [doi: 10.1161/STR.0000000000000357] [Medline: 33691469]
20. Lombardo C, Islam MS. Stroke survivors’ acceptance and satisfaction of telerehabilitation delivery of physiotherapy services: a systematic review. Phys Ther Rev 2023 Nov 2;28(4-6):261-277. [doi: 10.1080/10833196.2023.2271301]
21. Li L, Sun Y. Research hotspots and trends of the tele-rehabilitation for stroke survivors based on CiteSpace: a review. Medicine (Baltimore) 2023 Mar 31;102(13):e33398. [doi: 10.1097/MD.0000000000033398] [Medline: 37000067]
22. Chen J, Jin W, Zhang XX, Xu W, Liu XN, Ren CC. Telerehabilitation approaches for stroke patients: systematic review and meta-analysis of randomized controlled trials. J Stroke Cerebrovasc Dis 2015 Dec;24(12):2660-2668. [doi: 10.1016/j.jstrokecerebrovasdis.2015.09.014] [Medline: 26483155]
23. Linder SM, Rosenfeldt AB, Bay RC, Sahu K, Wolf SL, Alberts JL. Improving quality of life and depression after stroke through telerehabilitation. Am J Occup Ther 2015;69(2):6902290020p1-690229002010. [doi: 10.5014/ajot.2015.014498] [Medline: 26122686]
24. Sun WJ, Song YY, Wang C, et al. Telerehabilitation for family caregivers of stroke survivors: a systematic review and meta-analysis. J Nurs Manag 2023;2023:3450312. [doi: 10.1155/2023/3450312] [Medline: 40225617]
25. Edwards D, Kumar S, Brinkman L, et al. Telerehabilitation initiated early in post-stroke recovery: a feasibility study. Neurorehabil Neural Repair 2023;37(2-3):131-141. [doi: 10.1177/15459683231159660] [Medline: 36876946]
26. Laver KE, Adey-Wakeling Z, Crotty M, Lannin NA, George S, Sherrington C. Telerehabilitation services for stroke. Cochrane Database Syst Rev 2020 Jan 31;1(1):CD010255. [doi: 10.1002/14651858.CD010255.pub3] [Medline: 32002991]

XSL•FO
RenderX

27. Gutierrez-Arias R, González-Mondaca C, Marinkovic-Riffo V, Ortiz-Puebla M, Paillán-Reyes F, Seron P. Measures to ensure safety during telerehabilitation of people with stroke: a scoping review. J Telemed Telecare 2025 Feb;31(2):198-206. [doi: 10.1177/1357633X231181426] [Medline: 37321644]
28. Hopewell S, Chan AW, Collins GS, et al. CONSORT 2025 explanation and elaboration: updated guideline for reporting randomised trials. BMJ 2025 Apr 14;389:e081124. [doi: 10.1136/bmj-2024-081124] [Medline: 40228832]
29. Liu L, Chen W, Zhou H, et al. Chinese Stroke Association guidelines for clinical management of cerebrovascular disorders: executive summary and 2019 update of clinical management of ischaemic cerebrovascular diseases. Stroke Vasc Neurol 2020 Jun;5(2):159-176. [doi: 10.1136/svn-2020-000378] [Medline: 32561535]
30. Graham BL, Steenbruggen I, Miller MR, et al. Standardization of Spirometry 2019 Update. An Official American Thoracic Society and European Respiratory Society Technical Statement. Am J Respir Crit Care Med 2019 Oct 15;200(8):e70-e88. [doi: 10.1164/rccm.201908-1590ST] [Medline: 31613151]
31. Kim SH, Cho SH. Effects of cardiorespiratory physiotherapy on lung function in stroke: a network meta-analysis. Top Stroke Rehabil 2025 Jul;32(5):459-471. [doi: 10.1080/10749357.2024.2417647] [Medline: 39436741]
32. Lista-Paz A, Kuisma R, Canosa JLS, Sebio García R, González Doniz L. Pulmonary function in patients with chronic stroke compared with a control group of healthy people matched by age and sex. Physiother Theory Pract 2023 May;39(5):918-926. [doi: 10.1080/09593985.2022.2031363] [Medline: 35098873]
33. du Bois RM, Weycker D, Albera C, et al. Forced vital capacity in patients with idiopathic pulmonary fibrosis: test properties and minimal clinically important difference. Am J Respir Crit Care Med 2011 Dec 15;184(12):1382-1389. [doi: 10.1164/rccm.201105-0840OC] [Medline: 21940789]
34. McGlothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients. JAMA 2014 Oct 1;312(13):1342-1343. [doi: 10.1001/jama.2014.13128] [Medline: 25268441]
35. Guo M, Yu W, Sun Y, Wang L, Zhou H, Zhang Y. Effects of increasing indoor negative air ions on cognitive performance and health of high pure CO2 level-exposed college students. Indoor Air 2023 Mar 15;2023:1-11. [doi: 10.1155/2023/8298339]
36. Lei J, Yang T, Huang S, et al. Hourly concentrations of fine and coarse particulate matter and dynamic pulmonary function measurements among 4992 adult asthmatic patients in 25 Chinese cities. Environ Int 2022 Jan;158:106942. [doi: 10.1016/j.envint.2021.106942] [Medline: 34689038]
37. Fang X, Huang S, Zhu Y, et al. Short-term exposure to ozone and asthma exacerbation in adults: a longitudinal study in China. Front Public Health 2022;10:1070231. [doi: 10.3389/fpubh.2022.1070231] [Medline: 36684992]
38. Laveneziana P, Albuquerque A, Aliverti A, et al. ERS statement on respiratory muscle testing at rest and during exercise. Eur Respir J 2019 Jun;53(6):1801214. [doi: 10.1183/13993003.01214-2018] [Medline: 30956204]
39. Haque T, Vossoughi J, Johnson AT, Bell-Farrell W, Fitzgerald T, Scharf SM. Resistance measured by airflow perturbation compared with standard pulmonary function measures. OJRD 2013;03(2):63-67. [doi: 10.4236/ojrd.2013.32010]
40. Oliveira MJP, Rodrigues F, Firmino-Machado J, et al. Assessment of respiratory muscle weakness in subjects with neuromuscular disease. Respir Care 2018 Oct;63(10):1223-1230. [doi: 10.4187/respcare.06136] [Medline: 30065077]
41. Ohura T, Hase K, Nakajima Y, Nakayama T. Validity and reliability of a performance evaluation tool based on the modified Barthel Index for stroke patients. BMC Med Res Methodol 2017 Aug 25;17(1):131. [doi: 10.1186/s12874-017-0409-2] [Medline: 28841846]
42. Chen PC, Liaw MY, Wang LY, et al. Inspiratory muscle training in stroke patients with congestive heart failure: a CONSORT-compliant prospective randomized single-blind controlled trial. Medicine (Baltimore) 2016 Sep;95(37):e4856. [doi: 10.1097/MD.0000000000004856] [Medline: 27631248]
43. Yoo HJ, Pyun SB. Efficacy of bedside respiratory muscle training in patients with stroke: a randomized controlled trial. Am J Phys Med Rehabil 2018 Oct;97(10):691-697. [doi: 10.1097/PHM.0000000000000933] [Medline: 29570467]
44. Wei H, Sheng Y, Peng T, et al. Effect of pulmonary function training with a respirator on functional recovery and quality of life of patients with stroke. Contrast Media Mol Imaging 2022;2022(1):6005914. [doi: 10.1155/2022/6005914] [Medline: 36017026]
45. Ezeugwu VE, Olaogun M, Mbada CE, Adedoyin R. Comparative lung function performance of stroke survivors and age-matched and sex-matched controls. Physiother Res Int 2013 Dec;18(4):212-219. [doi: 10.1002/pri.1547] [Medline: 23359511]
46. Murphy TH, Corbett D. Plasticity during stroke recovery: from synapse to behaviour. Nat Rev Neurosci 2009 Dec;10(12):861-872. [doi: 10.1038/nrn2735] [Medline: 19888284]
47. Klingebiel M, Dinekov M, Köhler C. Analysis of ribosomal protein S6 baseline phosphorylation and effect of tau pathology in the murine brain and human hippocampus. Brain Res 2017 Mar 15;1659:121-135. [doi: 10.1016/j.brainres.2017.01.016] [Medline: 28119058]
48. Moore D, Loprinzi PD. Exercise influences episodic memory via changes in hippocampal neurocircuitry and long-term potentiation. Eur J Neurosci 2021 Oct;54(8):6960-6971. [doi: 10.1111/ejn.14728] [Medline: 32236992]
49. Stampanoni Bassi M, Gilio L, Buttari F, et al. Preventive exercise and physical rehabilitation promote long-term potentiation-like plasticity expression in patients with multiple sclerosis. Eur J Neurol 2024 Mar;31(3):e16071. [doi: 10.1111/ene.16071] [Medline: 37754770]

XSL•FO
RenderX

50. Chen J, Chen T, Wang X, et al. Effects of different respiratory training methods on respiratory function in patients with spinal cord injury: a network meta-analysis of randomized controlled trials. Crit Care 2025 Jul 7;29(1):285. [doi: 10.1186/s13054-025-05535-z] [Medline: 40624540]
51. Edman S, Horwath O, Van der Stede T, et al. Pro-brain-derived neurotrophic factor (BDNF), but not mature BDNF, is expressed in human skeletal muscle: implications for exercise-induced neuroplasticity. Function (Oxf) 2024;5(3):zqae005. [doi: 10.1093/function/zqae005] [Medline: 38706964]
52. Mazzitelli M, Kiritoshi T, Presto P, et al. BDNF Signaling and pain modulation. Cells 2025 Mar 22;14(7):476. [doi: 10.3390/cells14070476] [Medline: 40214430]
53. Lin M, Calabrese GB, Incognito AV, et al. A cholinergic spinal pathway for the adaptive control of breathing. Cell Rep 2025 Aug;44(8):116078. [doi: 10.1016/j.celrep.2025.116078]
54. Brezinski A, Popp N, Konkel K, Kurpad S, Satkunendrarajah K. Targeted neuromodulation of spinal interneurons enhances breathing in chronic spinal cord injury. Neurobiol Dis 2025 Sep;213:107007. [doi: 10.1016/j.nbd.2025.107007] [Medline: 40554005]
55. Hérent C, Diem S, Usseglio G, Fortin G, Bouvier J. Upregulation of breathing rate during running exercise by central locomotor circuits in mice. Nat Commun 2023 May 22;14(1):2939. [doi: 10.1038/s41467-023-38583-6] [Medline: 37217517]
56. Marques TBC, Neves JDC, Portes LA, Salge JM, Zanoteli E, Reed UC. Air stacking: effects on pulmonary function in patients with spinal muscular atrophy and in patients with congenital muscular dystrophy. J Bras Pneumol 2014 Oct;40(5):528-534. [doi: 10.1590/s1806-37132014000500009] [Medline: 25410841]
57. Khalid A, Riasat K, Javaid A, Nasir K, Shahid R, Sehar S. Comparing the effect of pursed lip breathing and diaphragmatic breathing on pulmonary function and exercise frequency in COPD patients. JHRR 2024;4(2):1807-1810. [doi: 10.61919/jhrr.v4i2.1191]
58. Jansang S, Mickleborough T, Suksom D. Effects of pursed-lip breathing exercise using windmill toy on lung function and respiratory muscle strength in the elderly. J Med Assoc Thai 2016 Sep;99(9):1046-1051. [Medline: 29927211]
59. Seo K, Hwan PS, Park K. The effects of inspiratory diaphragm breathing exercise and expiratory pursed-lip breathing exercise on chronic stroke patients' respiratory muscle activation. J Phys Ther Sci 2017 Mar;29(3):465-469. [doi: 10.1589/jpts.29.465] [Medline: 28356632]
60. Shin HJ, Son HH. Effect of diaphragmatic breathing training with visual biofeedback on respiratory function in patients with multiple rib fractures: A randomized-controlled study. Turk J Phys Med Rehabil 2024 Mar;70(1):131-141. [doi: 10.5606/tftrd.2024.12601] [Medline: 38549830]
61. da Silva Guimarães B, de Souza LC, Cordeiro HF, et al. Inspiratory muscle training with an electronic resistive loading device improves prolonged weaning outcomes in a randomized controlled trial. Crit Care Med 2021 Apr 1;49(4):589-597. [doi: 10.1097/CCM.0000000000004787] [Medline: 33332819]
62. Souza APM, da Silva LM. Respiratory muscle training in people with COPD: a systematized review. Braz J Phys Ther 2024 Apr;28:100628. [doi: 10.1016/j.bjpt.2024.100628]
63. Xiao Y, Luo M, Wang J, Luo H. Inspiratory muscle training for the recovery of function after stroke. Cochrane Database Syst Rev 2012 May 16;2012(5):CD009360. [doi: 10.1002/14651858.CD009360.pub2] [Medline: 22592740]
64. Kulnik ST, Birring SS, Moxham J, Rafferty GF, Kalra L. Does respiratory muscle training improve cough flow in acute stroke? Pilot randomized controlled trial. Stroke 2015 Feb;46(2):447-453. [doi: 10.1161/STROKEAHA.114.007110] [Medline: 25503549]
65. Cramer SC, Dodakian L, Le V, et al. Efficacy of home-based telerehabilitation vs in-clinic therapy for adults after stroke: a randomized clinical trial. JAMA Neurol 2019 Sep 1;76(9):1079-1087. [doi: 10.1001/jamaneurol.2019.1604] [Medline: 31233135]
66. Onan D, Ulger O, Martelletti P. Effects of spinal stabilization exercises delivered using telerehabilitation on outcomes in patients with chronic neck pain: a randomized controlled trial. Expert Rev Neurother 2023 Mar;23(3):269-280. [doi: 10.1080/14737175.2023.2192870] [Medline: 36927237]
67. Salgueiro C, Urrútia G, Cabanas-Valdés R. Influence of core-stability exercises guided by a telerehabilitation app on trunk performance, balance and gait performance in chronic stroke survivors: a preliminary randomized controlled trial. Int J Environ Res Public Health 2022 May 7;19(9):5689. [doi: 10.3390/ijerph19095689] [Medline: 35565084]
68. Misawa M, Bajin IY, Zhang B, et al. A telerehabilitation program to improve visual perception in children and adolescents with hemianopia consecutive to a brain tumor: a single-arm feasibility and proof-of-concept trial. EClinicalMedicine 2024 Dec;78:102955. [doi: 10.1016/j.eclinm.2024.102955] [Medline: 39687429]
69. An HJ, Kim AY, Park SJ. Immediate effects of diaphragmatic breathing with cervical spine mobilization on the pulmonary function and craniovertebral angle in patients with chronic stroke. Medicina (Kaunas) 2021 Aug 16;57(8):826. [doi: 10.3390/medicina57080826] [Medline: 34441033]
70. Shi L, Liu F, Liu Y, et al. Biofeedback respiratory rehabilitation training system based on virtual reality technology. Sensors (Basel) 2023 Nov 7;23(22):9025. [doi: 10.3390/s23229025] [Medline: 38005413]
71. Lan KC, Li CW, Cheung Y. Slow breathing exercise with multimodal virtual reality: a feasibility study. Sensors (Basel) 2021 Aug 13;21(16):5462. [doi: 10.3390/s21165462] [Medline: 34450909]

XSL•FO
RenderX

72. Chen X, Li C, Zeng L, et al. Comparative efficacy of different combinations of acapella, active cycle of breathing technique, and external diaphragmatic pacing in perioperative patients with lung cancer: a randomised controlled trial. BMC Cancer 2023 Mar 28;23(1):282. [doi: 10.1186/s12885-023-10750-4] [Medline: 36978035]
73. Vázquez-Gandullo E, Hidalgo-Molina A, Montoro-Ballesteros F, Morales-González M, Muñoz-Ramírez I, Arnedillo-Muñoz A. Inspiratory muscle training in patients with chronic obstructive pulmonary disease (COPD) as part of a respiratory rehabilitation program implementation of mechanical devices: a systematic review. Int J Environ Res Public Health 2022 May 3;19(9):5564. [doi: 10.3390/ijerph19095564] [Medline: 35564959]
74. Rodrigues A, Muñoz Castro G, Jácome C, Langer D, Parry SM, Burtin C. Current developments and future directions in respiratory physiotherapy. Eur Respir Rev 2020 Dec 31;29(158):200264. [doi: 10.1183/16000617.0264-2020] [Medline: 33328280]
75. Franklin E, Anjum F. Incentive spirometer and inspiratory muscle training. In: StatPearls: StatPearls Publishing; 2023.

Abbreviations

ANCOVA: analysis of covariance
CMRTP: comprehensive mobile-based respiratory training program
FEV_1 : forced expiratory volume in 1 second
FVC: forced vital capacity
HIPAA: Health Insurance Portability and Accountability Act
LS: least squares
MBI: modified Barthel index
MEP: maximal expiratory pressure
MIP: maximal inspiratory pressure
NIHSS: National Institutes of Health Stroke Scale
PEF: peak expiratory flow

Edited by L Buis; submitted 11.Jun.2025; peer-reviewed by JV Saornil, Z Sanchez-Mila; revised version received 25.Jan.2026; accepted 08.Mar.2026; published 14.Apr.2026.

Please cite as:
Ma ZK, Jiang HH, Tang YH, Yang J, Wei QC, Gao Q
Effect of a Comprehensive Mobile-Based Respiratory Training Program on Respiratory Function in Survivors of Acute Stroke: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e78637
URL: https://mhealth.jmir.org/2026/1/e78637
doi:10.2196/78637

© Zong-Ke Ma, Han-Hong Jiang, Yan-Hua Tang, Jia Yang, Qing-Chuan Wei, Qiang Gao. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of a Personalized Digital Exercise and Nutrition Rehabilitation Program in Postoperative Patients With Gastric Cancer: Randomized Controlled Trial

Inah Kim[1], MD, PhD; Ji Young Lim[2], PT, PhD; Sun Woo Kim[2], MS; Jun Ho Lee[3], MD, PhD; Tae Sung Sohn[3], MD, PhD; Sungsoo Park[4], MD, PhD; Seok Ho Kang[5], MD, PhD; Ji Youl Lee[6], MD, PhD; Ji Hye Hwang[7], MD, PhD

[1]Department of Rehabilitation Medicine, Dongtan Sacred Heart Hospital, Hallym University College of Medicine, Hwaseong-Si, Republic of Korea

[2]Department of Physical and Rehabilitation Medicine, School of Medicine, Sungkyunkwan University, Seoul, Republic of Korea

[3]Department of Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

[4]Department of Surgery, Korea University Anam Hospital, Korea University College of Medicine, Seoul, Republic of Korea

[5]Department of Urology, Korea University Anam Hospital, Korea University College of Medicine, Seoul, Republic of Korea

[6]Department of Urology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

[7]Department of Physical and Rehabilitation Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

Corresponding Author:
Ji Hye Hwang, MD, PhD
Department of Physical and Rehabilitation Medicine
Samsung Medical Center
Sungkyunkwan University School of Medicine
222 Banpo-daero
Gangnam-gu
Seoul, 06351
Republic of Korea
Phone: 82 2 2258 5905
Email: hwanglee@skku.edu

Abstract

Background: Nutritional and exercise interventions have shown beneficial effects after gastrectomy for gastric cancer. While digital health tools show promise in cancer care, their long-term effectiveness in patients with gastric cancer remains unclear. In addition, large-scale studies of personalized interventions initiated in the immediate postoperative period are lacking.

Objective: This study aimed to determine whether a personalized mobile health intervention incorporating exercise and nutrition confers additional benefits in weight change, body composition, nutritional status, physical fitness, and quality of life compared with standard care when applied continuously for 12 months after gastrectomy.

Methods: This multicenter, randomized controlled trial enrolled 257 patients who had undergone curative resection for stage I-III gastric cancer. Participants were randomly assigned (2:1) to either a 12-month personalized mobile health intervention group or a standard rehabilitation control group. The digital intervention incorporated a smartphone app and wearable device, offering exercise and dietary plans tailored to clinical parameters such as BMI, surgery type, and recovery stage. Standard care, including nutritional education, was provided to all patients. The primary outcome was a change in body weight over 12 months. Secondary outcomes included quality of life (European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30 [EORTC QLQ-C30] and European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Stomach 22), nutritional status (Mini Nutritional Assessment Score), physical fitness (grip strength, 30-second chair stand test, and 2-minute walk test), physical activity (International Physical Activity Questionnaire-Short Form), pain intensity (average Numeric Rating Scale), body composition (skeletal muscle mass, lean body mass, and fat mass), BMI, hemoglobin, vitamin B_{12}, and albumin. Assessments were conducted at baseline and at 1, 3, 6, and 12 months postoperatively.

Results: No significant group-by-time effects were observed for weight change. Secondary outcomes showed no between-group differences, except for 1 subscale of the EORTC QLQ-C30, which lacked clinical significance. The intervention group reported high satisfaction and adherence to the mobile app, and no adverse events or incidents were observed during the 12-month study period.

XSL•FO
RenderX

Conclusions: The digital health program integrating exercise and nutrition was safe and feasible, with high satisfaction and adherence among patients with gastric cancer. However, it was not superior to standard education in modifying the postoperative trajectory in patients with gastric cancer after surgery, including body weight change and related functional or nutritional outcomes. These findings suggest that future digital health programs should be precisely targeted and tailored to specific patient populations and recovery phases.

Trial Registration: ClinicalTrials.gov NCT04907591; https://clinicaltrials.gov/ct2/show/NCT04907591

(JMIR Mhealth Uhealth 2026;14:e85795) doi:10.2196/85795

KEYWORDS

digital health; eHealth; gastric cancer; mHealth; randomized controlled trial; rehabilitation

Introduction

In patients with gastric cancer (GC) undergoing gastrectomy, prevention of postoperative weight loss is a critical factor for reducing complications and improving both cancer-specific survival and quality of life [1,2]. However, malnutrition is highly prevalent, affecting more than half of patients who are affected or at risk [3]. Nutritional impairment leads to weight loss and skeletal muscle depletion, which in turn are associated with reduced activities of daily living and poorer survival outcomes [4,5]. Evidence indicates that nutritional or exercise interventions in the postoperative course of GC provide beneficial effects [6,7]. For example, randomized controlled trials have demonstrated that dietary interventions significantly attenuate weight loss, particularly in patients undergoing total gastrectomy, with benefits sustained not only at 6-8 weeks but also up to 1 year after surgery [8-10]. Likewise, exercise interventions have been shown to enhance muscular strength, cardiopulmonary function, quality of life, and emotional functioning in this population [11,12].

With recent developments in communication technology, mobile health (mHealth) apps have gained attention as supportive tools in cancer management [13-15]. However, studies of digital health interventions in upper gastrointestinal cancers have shown inconclusive efficacy [16-19]. Most previous studies on digital health with wearable devices focused on limited interventions, such as health programs applied only during chemotherapy or those focusing mainly on psychological or nutrition-related issues, often with insufficient personalization, short follow-up periods, or a small number of patients [20,21]. Offering uniform content to this population might be inefficient because patient status varies according to diverse factors such as age, surgical method, or therapeutic phase [10,22,23]. In addition, a small number of patients is not sufficiently generalizable to the broader population.

This study aimed to evaluate whether a long-term, personalized digital health intervention initiated after gastrectomy in patients with GC confers superior effects compared with conventional education in modifying the biological postoperative course.

Methods

Patients

Participants were recruited between May 17, 2021, and September 19, 2022. In total, 257 patients diagnosed with GC who underwent curative surgery were enrolled in this study. The inclusion criteria were as follows: (1) age of 19-75 years, (2) curative resection for stage I-III GC according to the American Joint Committee on Cancer staging system, (3) ownership of an Android- or iOS-based smartphone, (4) ability to use a mobile app and to attend regular outpatient follow-up visits, and (5) voluntary agreement to participate in the study. Exclusion criteria included the inability to engage in exercise or diet management due to severe comorbidities, neuromusculoskeletal disorders, cognitive or visual impairment, or communication difficulties.

Eligible patients were identified through consultations with the Departments of Surgery and Rehabilitation Medicine. Following surgery, all eligible participants were informed of the study procedures, and written informed consent was obtained from those who agreed to participate voluntarily.

Trial Design

This prospective, randomized, controlled, multicenter, open-label, 2-arm trial adhered to the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) and CONSORT (Consolidated Standards of Reporting Trials) standards (checklist provided in Multimedia Appendix 1). Participants were recruited from 2 university hospitals in South Korea (Samsung Medical Center and Korea University Anam Hospital) and were randomly assigned (2:1 ratio) to either the digital therapeutic or control group using blocked randomization with randomly selected block sizes of 3 and 6. Patients were randomly allocated to 2 groups for 12 months of the rehabilitation program, starting immediately after the operation: a personalized digital therapeutic (intervention) group and a conventional education-based rehabilitation (control) group. Both groups received standard care and conventional dietary education on the prevention of dumping syndrome and the maintenance of muscle mass and weight.

Personalization Algorithm and Dynamic Adjustment

Participants in the intervention group (personalized digital therapeutic group) were provided with the “Gastric Cancer by a Second Doctor” mobile app (Medi Plus Solution) and the “DoFit” smart wristband (NF-B20; Medi Plus Solution).

Participants in the usual care group received regular diet management and dumping syndrome prevention training at the hospital.

XSL•FO
RenderX

Exercise Prescription Algorithm

The exercise component was individualized using a rule-based algorithm that incorporated patient-specific variables, including weeks since surgery and current chemotherapy status. Based on these factors, participants were categorized into an exercise stage, which determined the target intensity and duration of aerobic activity. The target heart rate (THR) was calculated using the standard formula: THR = (220 – age) × THR factor. The THR factor was defined according to cancer-specific exercise guidelines, including those from the American College of Sports Medicine (2022) and the Roundtable on Exercise for Cancer Survivors (2019), as well as previous studies involving patients with GC, such as the Postoperative Recovery Exercise Program for Gastric Cancer Patients (2018) [12,24,25]. Exercise duration was adjusted according to the assigned stage, and progression was guided by the individual's clinical recovery. The prescribed exercise program included stretching, flexibility exercises, aerobic training, and functional resistance exercises using body weight or simple equipment. These prescriptions were reviewed and adjusted biweekly based on monitored adherence, patient-reported symptoms, and physiological parameters such as heart rate and step count, allowing real-time tailoring of exercise difficulty. During chemotherapy, low-intensity exercises focused on flexibility were implemented. Additionally, the exercise stage was adjusted upward or downward based on the patient's self-reported rating of perceived exertion. Table S1 in Multimedia Appendix 2 presents the stepwise logic used for exercise stage progression and adjustment.

Nutrition Prescription Algorithm

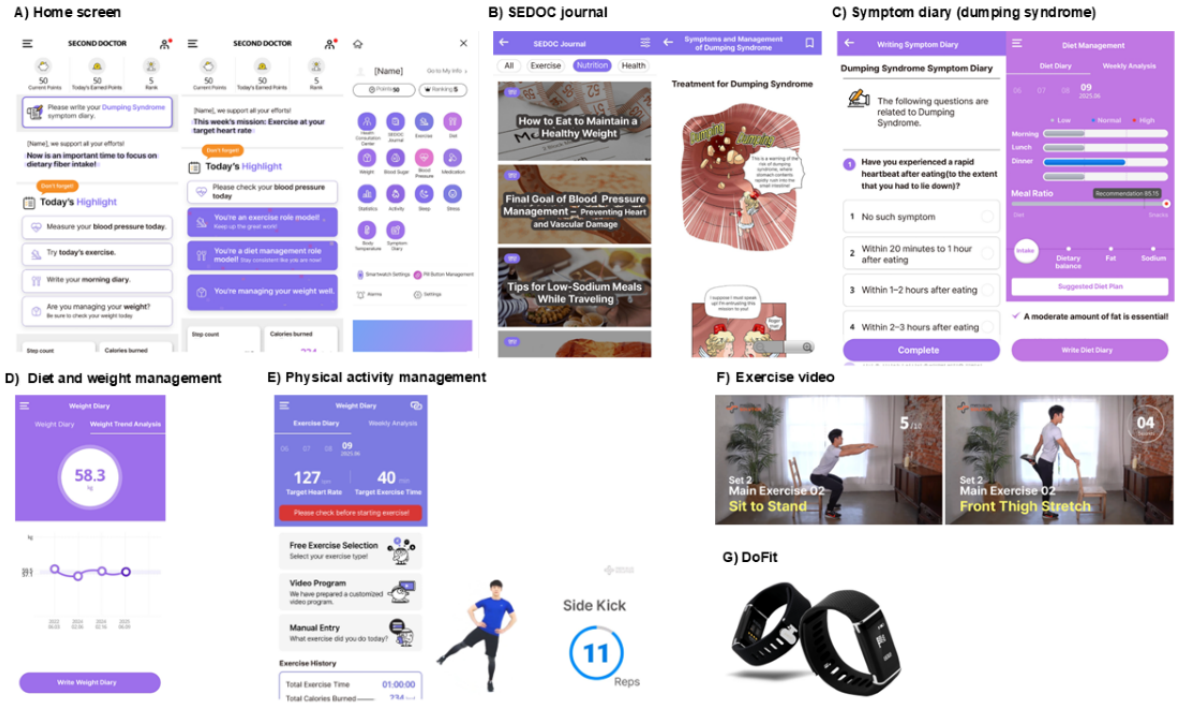
Nutritional planning was driven by a clinical decision algorithm that considered the patient's age (<65 vs ≥65 years), extent of gastrectomy (total vs subtotal), postoperative week, and the presence of chronic diseases such as diabetes, hypertension, and dyslipidemia. Caloric requirements were calculated using the formula: ideal body weight × activity factor (26-35 kcal/kg/day). During the early recovery phase (the first 12 weeks postoperatively), calorie targets were conservatively adjusted to reduce the risk of gastrointestinal complications, such as dumping syndrome and diarrhea. After week 13, caloric targets were increased to meet 100% of estimated needs, with additional modifications based on relevant clinical guidelines—such as sodium restriction per the 2022 Korean Society of Hypertension guidelines and carbohydrate adjustment in accordance with the 2021 Korean Diabetes Association recommendations [26,27]. Protein intake was consistently recommended at 1.2 g/kg ideal body weight per day throughout the recovery and adaptation periods, while fat intake was limited to 20% of total energy during early recovery. Meal plans were adapted in real time to reflect changing caloric needs. Specifically, portion sizes and snack frequencies were automatically recalculated. For patients who underwent subtotal gastrectomy, snack frequency was adjusted to accommodate more rapid adaptation to oral intake. Nutritional care followed a phased structure. In the recovery phase (weeks 1-4), energy intake was slightly below estimated needs and focused on soft textures and small, frequent meals. In the adaptation phase (weeks 5-12), patients gradually advanced toward full caloric intake with a transition to regular textures. The maintenance phase (week 13 and beyond) included nutritional modifications based on comorbidities, with targeted adjustments for fat, carbohydrate, sodium, and cholesterol. The corresponding decision rules for energy and nutrient prescription, stratified by surgical type, recovery phase, and clinical profile, are presented in Table S2 in Multimedia Appendix 2.

Clinician Monitoring and Feedback

The program was designed to operate primarily through a rule-based algorithm, enabling fully automated adjustment of exercise and nutrition prescriptions without requiring direct input from clinicians. A secure, web-based monitoring platform was available to clinical staff (eg, nurses, dietitians, and physiotherapists), allowing access to real-time life-log data, including step count, heart rate, nutritional intake, and symptom reports. While the platform did not allow clinicians to override or alter the algorithmic recommendations manually, it facilitated patient-centered feedback. Clinicians used the data during routine outpatient visits to encourage adherence and self-management, particularly when patients exhibited low engagement or reported challenges. Based on clinical judgment, face-to-face feedback was selectively offered to patients undergoing chemotherapy approximately 3 months postoperatively—when adherence risk is typically higher—to reinforce self-care and address barriers. Screenshots of the representative functions are shown in Figure 1, and the main app functions are listed in Table S3 in Multimedia Appendix 2.

XSL•FO
RenderX

Figure 1. Representative screenshots of the app and smart band interface include (A) main screen displaying today's to-do list and self-reported symptom monitoring, (B) provision of the general information (Second Doctor [SEDOC] Journal), (C) symptom diary, (D) dietary tracking module and body weight management interface, (E) physical activity tracking module, (F) instructional exercise videos, and (G) DoFit.



Outcomes

A trained clinical research coordinator conducted in-person evaluations of the outcome measures during hospital visits. An overview of baseline screening, assessment procedures, and the follow-up schedule across study visits is presented in Table S4 in Multimedia Appendix 2.

Primary Outcome

The primary outcome was the change in body weight from baseline to 12 months, assessed using a bioimpedance analysis device to compare the intervention and control groups.

Secondary Outcomes

The secondary outcomes included quality of life, cancer-specific symptoms and function, nutritional status, physical activity, physical status, pain intensity, and physical fitness over time.

Quality of Life

Health-related quality of life was assessed using the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30 (EORTC QLQ-C30) and European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Stomach 22 (EORTC QLQ-STO22). The EORTC QLQ-C30 consists of 30 items, covering functional, symptom, and global quality of life domains. A high EORTC QLQ-C30 score indicates a good quality of life. The EORTC QLQ-STO22, a GC-specific questionnaire, consists of 22 items across symptom domains related to GC [23,28].

Nutrition, Pain Intensity, Physical Activity, and Function

Korean Mini Nutritional Assessment

Nutritional status was assessed using the Korean version of the Mini Nutritional Assessment (MNA) long form, an internationally validated tool in patients with cancer [29]. Previous studies have shown prognostic value for treatment efficacy, health-related quality of life, mortality, survival, and cancer progression [30]. A total score of less than 17 points indicates malnutrition, 17-23.5 points indicates a risk of malnutrition, and 24-30 points represents the normal range [31].

Numeric Rating Scale

Pain intensity over the past week was assessed using an 11-point Numeric Rating Scale (NRS), where 0 indicated "no pain," and 10 represented "the worst possible pain." The participants were asked to report both their average and worst perceived pain intensities during this period [32].

International Physical Activity Questionnaire-Short Form

Self-reported physical activity was assessed using the International Physical Activity Questionnaire-Short Form (IPAQ-SF) [33]. The IPAQ-SF is a widely used self-report instrument that enables standardized measurement and comparison of physical activity levels across populations. It consists of 7 items assessing the frequency and duration of physical activities performed over the preceding 7 days, including vigorous and moderate activities, walking, and sedentary time [34]. Responses are converted into metabolic equivalent of task (MET) minutes per week, using standardized MET values (8.0 for vigorous activity, 4.0 for moderate activity, and 3.3 for walking) [34,35]. Based on total activity levels, respondents are classified into low, moderate, or high activity

XSL•FO
RenderX

categories. Only activities lasting at least 10 continuous minutes are counted [35].

In this study, we used the Korean version of the IPAQ-SF, developed and validated by Oh et al [34], which demonstrated acceptable reliability and validity in Korean adults, including older populations.

Grip Strength Test, 30-second Chair Stand Test, and 2-minute Walk Test

Physical fitness was assessed using the grip strength test [36], 30-second chair stand test [37,38], and 2-minute walk test [39], which evaluates the strength generated by the forearm muscles, strength and endurance of the lower extremities, and cardiorespiratory endurance, respectively. Grip strength of the dominant hand was assessed using a handheld dynamometer (microFET Digital HandGRIP Dynamometer; Hoggan Scientific LLC). All tests were performed using standardized measurement methods.

eHealth Literacy Scale

The eHealth literacy of the intervention group was only measured at baseline using the eHealth Literacy Scale (eHEALS), which comprised 10 items. A total of 8 questions assessed eHealth literacy on a 5-point Likert scale, with higher scores indicating greater literacy [40]. In addition, the intervention group completed a self-developed questionnaire assessing satisfaction with the mobile app and smart band at 6 months.

App-Based Adherence Measures

“Exercise log (days)” was defined as the number of days in a month on which participants recorded any exercise session—whether free-form, video-guided, or manually entered—that lasted at least 5 minutes. “Exercise duration per exercise day (min)” was calculated by dividing the total exercise time in a month by the number of exercise log days. For trend analyses, participants who had no exercise records in each month were not included in the calculation of mean values. “Step count logged (days)” refers to the total number of days on which step counts were recorded. The “step count” was calculated by dividing the total number of steps over a month by the number of step-log days. To ensure monthly data reflected typical activity patterns, step counts were counted only if data were recorded on at least 7 days within the month. For the trend analysis of step counts, zero values were excluded when calculating the average step count among users with available data. “Weight log” and “diet log” indicate the number of days with at least 1 entry, and “app access” denotes the number of days the user accessed the app at least once.

Satisfaction Survey

A customized satisfaction questionnaire regarding the use of mobile apps and Internet of Things devices was developed collaboratively by investigators from the 2 participating institutions. This survey was administered exclusively to participants in the intervention group at the 6-month follow-up and used a 5-point Likert scale, with higher scores reflecting greater levels of satisfaction.

Sample Size Calculation

Using G*Power (version 3.1.9.2; Heinrich Heine University Düsseldorf), the required sample size was calculated for the primary outcome: change in body weight from baseline to 12 months. The primary hypothesis was to test the superiority of the personalized digital intervention in attenuating postoperative percentage body-weight loss. We assumed a standardized effect size (Cohen *d*) of 0.35 (moderate), a significance level (α) of 5%, and power of 80%, based on the characteristics of the intervention described in a previous behavioral intervention study [41]. This effect size was chosen to represent a conservative estimate for the minimum clinically meaningful difference expected from a nonpharmacological, behavioral intervention in a population of cancer survivors. The attrition rate in mobile-based interventions is known to vary depending on the study design. Considering an anticipated dropout rate of 10%, the final target enrollment was set at 324 participants, with 216 assigned to the intervention group and 108 to the control group. In our previous 12-week validation study involving 203 patients with advanced gastrointestinal cancer, approximately 6% discontinued the study for reasons unrelated to changes in physical condition [13].

Data Handling and Security

Electronic data were captured using a secure web-based platform provided by the Medicallogic Company [42] for systematic and reliable data management. Access to the system was restricted to authorized study personnel.

Participant data were recorded using an electronic case report form, and relevant information from the source documents was entered into the electronic case report form by designated research staff.

Statistical Analysis

All statistical analyses were conducted using the R software (version 4.4.2; R Foundation for Statistical Computing). Statistical significance was defined as a 2-sided *P* value of <.05, and 95% CIs were calculated. The normality of continuous variables was assessed by examining histograms. For baseline comparisons between the intervention and control groups, the Student *t* test or Mann-Whitney *U* test was used for continuous variables, and the chi-square or Fisher exact tests were applied for categorical variables, as appropriate. For the primary outcome (weight), comparisons at each time point were performed using independent *t* tests, and changes over time were analyzed using mixed-effects models, with group, time, and their interaction included as fixed effects, and participant ID included as a random effect. For secondary outcomes, if the data were normally distributed, mixed-effects models with the same structure (group, time, and group × time interaction as fixed effects, and participant ID as a random effect) were used. For nonnormally distributed outcome measures, the change from 1 to 12 months was compared using the Wilcoxon rank-sum test. To assess the robustness of these findings, supplementary analyses were conducted using generalized estimating equations (GEE), which account for within-participant correlation and allow for nonnormally distributed repeated measures.

XSL•FO
RenderX

Histogram analyses indicated that the following variables were normally distributed: weight, skeletal muscle mass, BMI, skeletal muscle index, total fat mass, grip strength, 30-second chair stand test, and 2-minute walk test. These variables were analyzed using mixed-effects linear regression models, incorporating all available data points at 1, 3, 6, and 12 months, to estimate the slope trajectories.

Conversely, the following variables were nonnormally distributed and were analyzed using the Wilcoxon rank-sum test: NRS average, NRS maximum, vigorous activity, moderate activity, walking time, total METs, MNA subcomponents, and EORTC QLQ-C30 and QLQ-STO22 scores—including physical functioning, role functioning, cognitive functioning, emotional functioning, social functioning, fatigue, nausea/vomiting, pain, dyspnea, insomnia, appetite loss, constipation, diarrhea, financial difficulties, global health status/quality of life, and summary score, as well as dysphagia, reflux, eating restrictions, anxiety, dry mouth, taste alteration, body image, and hair loss. Inflammatory and nutritional biomarkers, such as hemoglobin, vitamin B_{12}, and albumin, also showed nonnormal distributions and were analyzed accordingly.

Safety Monitoring and Discontinuation Criteria

The risk level associated with this study was assessed as minimal based on evaluations conducted by the institutional review board and principal investigator. Throughout the study period, participants were allowed to contact the research team at any time to address study-related concerns. In cases of severe pain or injury, participants underwent a clinical evaluation by the principal investigator. Participants were permitted to withdraw from the study at any time they wished. Additionally, withdrawal could also be initiated by the study team in cases of serious medical conditions or noncompliance with medical instructions.

Ethical Considerations

All study procedures were reviewed and approved by the institutional review boards of the 2 hospitals (approval numbers SMC-2021-01-090 and 2021AN0104). The trial was registered at ClinicalTrials.gov (approval ID NCT04907591). The original study protocol was submitted to the Samsung Medical Center Institutional Review Board on February 1, 2021, and received approval on February 23, 2021. All collected data were retained for 3 years after study completion and were subsequently destroyed in accordance with institutional policy.

Written informed consent was obtained from all participants prior to their participation, and all participants voluntarily agreed to take part in the study. All data were anonymized and securely stored in restricted-access locations to protect personal information and ensure confidentiality. Participants received a compensation of 50,000 KRW (approximately US $34) after hospital discharge at each follow-up assessment.

Results

Patients

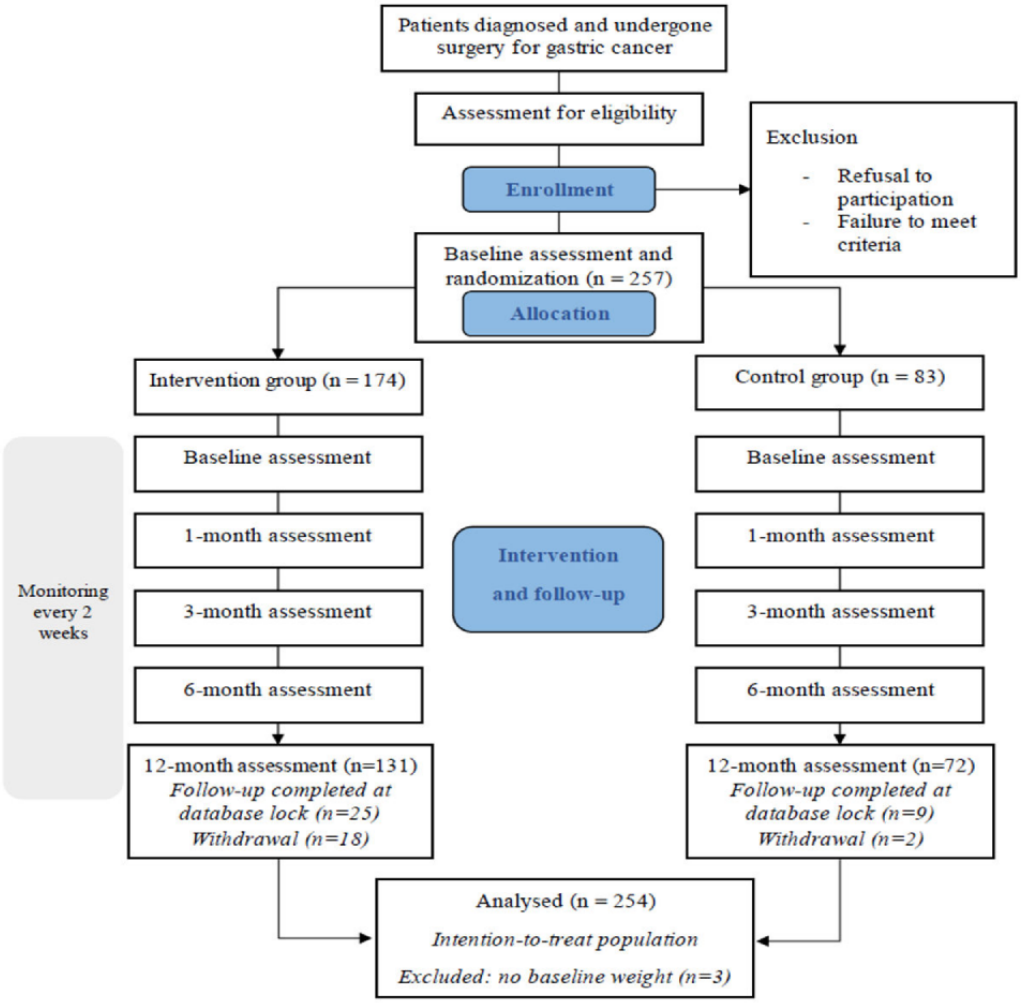
A total of 254 patients were included in the intention-to-treat population (131 in the intervention group and 123 in the control group). Overall, 3 participants were excluded from the final analysis due to missing baseline body weight data.

By the 12-month follow-up, several postrandomization dropouts had occurred. In the intervention group, 1 participant died due to lung and brain metastases, and another was diagnosed with non-Hodgkin lymphoma, unrelated to GC. In the control group, dropouts were primarily due to clinical deterioration and failure to attend scheduled follow-up visits. Additional reasons for early withdrawal across both groups included adverse effects from ongoing treatment, allergic reactions to the smartwatch device, and user-reported difficulties with or reluctance to use digital tools. Despite these attritions, all participants were included in the intention-to-treat analysis (Figure 2).

The baseline demographics and clinical variables of the intervention and control groups are shown in Table 1. Baseline characteristics of the participants were well balanced between the 2 groups, except for age, alcohol consumption, adjuvant chemotherapy, and grip strength. To account for baseline imbalances, demographic variables showing between-group differences were included as covariates in the outcome analyses.

XSL•FO
RenderX

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.



XSL•FO
RenderX

Table 1. Baseline characteristics of the study participants. Group comparisons were conducted using Student *t* test or Mann-Whitney U test for continuous variables, and the chi-square or Fisher exact tests for categorical variables, as appropriate.

Characteristics	mHealth[a] (n=174)	Control (n=83)	*P* value
Sex, n (%)			.36
Female	77 (44.25)	31 (37.35)	
Male	97 (55.75)	52 (62.65)	
Age (years), mean (SD)	54.91 (9.36)	57.93 (9.15)	.02[b]
Height (cm), mean (SD)	164.24 (7.88)	164.67 (8.95)	.71
Weight (kg), mean (SD)	65.35 (11.96)	65.54 (10.60)	.90
Smoking, n (%)			.49
Current smoker	22 (12.64)	9 (10.84)	
Ex-smoker	52 (29.89)	31 (37.35)	
Never smoker	100 (57.47)	43 (51.81)	
Drinking, n (%)			.03[b]
Yes	37 (21.26)	29 (34.94)	
No	137 (78.74)	54 (65.06)	
Operation type, n (%)			.95
Total gastrectomy	32 (18.5)	14 (17.3)	
Subtotal gastrectomy	141 (81.5)	67 (82.7)	
Operation approach, n (%)			.91
Laparoscopic	140 (80.46)	68 (81.93)	
Open	34 (19.54)	15 (18.07)	
Operation anastomosis, n (%)			.47
Billroth 1	31 (17.82)	18 (21.69)	
Billroth 2	105 (60.34)	44 (53.01)	
Roux-en-Y	36 (20.69)	18 (21.69)	
Esophagogastrostomy	1 (0.57)	2 (2.41)	
Double tract reconstruction	1 (0.57)	1 (1.20)	
Adjuvant chemotherapy, n (%)			.03[b]
No	133 (76.44)	74 (89.16)	
Yes	41 (23.56)	9 (10.84)	
Adjuvant radiotherapy, n (%)			.54
No	173 (99.43)	82 (98.80)	
Yes	1 (0.57)	1 (1.20)	
Yes	11 (6.32)	7 (8.43)	
Cancer stage, n (%)			.10
Stage 1	126 (72.41)	70 (84.34)	
Stage 2	18 (10.34)	6 (7.23)	
Stage 3	30 (17.24)	7 (8.43)	
Hb[c], mean (SD)	13.16 (1.92)	13.53 (1.93)	.16
Vitamin B_{12}, mean (SD)	703.98 (314.85)	670.86 (192.17)	.59
Skeletal muscle mass, mean (SD)	26.53 (5.79)	27.37 (6.08)	.31
Skeletal muscle index, mean (SD)	7.09 (1.21)	7.30 (1.16)	.24

XSL•FO
RenderX

Characteristics	mHealth[a] (n=174)	Control (n=83)	*P* value
BMI (kg/m^2), mean (SD)	24.05 (3.36)	24.05 (2.79)	.99
Total fat mass (kg), mean (SD)	16.48 (6.40)	15.84 (4.67)	.44
Lean body mass (kg), mean (SD)	47.97 (9.65)	49.76 (10.35)	.25
Grip strength (kg), mean (SD)	28.61 (9.74)	31.35 (9.88)	.04[b]
Average NRS[d], mean (SD)	2.94 (1.55)	2.95 (1.67)	.94
Total metabolic equivalent of task, mean (SD)	2037.45 (2097.73)	2106.87 (1954.06)	.80
MNA[e] total score, mean (SD)	24.41 (3.11)	24.72 (3.53)	.49
EORTC QLQ-C30[f], mean (SD)			
Physical functioning	87.90 (14.94)	87.63 (20.03)	.91
Role functioning	88.25 (22.73)	88.76 (23.73)	.87
Cognitive functioning	87.19 (13.85)	88.15 (12.63)	.58
Emotional functioning	80.20 (20.22)	78.61 (20.30)	.56
Social functioning	91.23 (19.82)	92.37 (19.70)	.67
Fatigue	24.21 (21.92)	22.49 (21.46)	.55
Nausea/vomiting	7.13 (14.17)	7.03 (15.86)	.96
Pain	15.41 (23.16)	13.05 (23.43)	.45
Dyspnea	7.90 (16.72)	10.04 (19.98)	.40
Insomnia	20.81 (27.68)	20.88 (27.89)	.98
Appetite loss	12.72 (23.95)	11.65 (26.25)	.75
Diarrhea	13.68 (23.55)	10.04 (17.83)	.17
Financial difficulties	7.51 (20.05)	5.22 (16.03)	.33
Global health status	67.20 (23.28)	64.36 (21.04)	.33

[a]mHealth: mobile health.

[b]*P* values <.05 were considered statistically significant.

[c]Hb: hemoglobin.

[d]NRS: Numeric Rating Scale.

[e]MNA: Mini Nutritional Assessment.

[f]EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30.

Primary Outcome

No significant difference was observed in body weight change between the groups over 12 months as assessed using a linear mixed model that accounted for group, time, and their interactions (Figure 3). This result remained consistent after adjusting for age, drinking, and adjuvant chemotherapy in multivariate regression (Table 2).

XSL•FO
RenderX

Figure 3. Analysis of the primary outcome with between-group comparisons at each time point. (A) Box plot showing the change in weight from baseline to 12 months postoperatively. The box represents the median (center line) and the first and third quartiles, with whiskers representing the minimum and maximum values. (B) Weight changes over the study period. Error bars represent standard errors estimated from the linear mixed model. The reported *P* value corresponds to the group × time interaction effect. Statistical significance was set at *P*<.05. mHealth: mobile health.

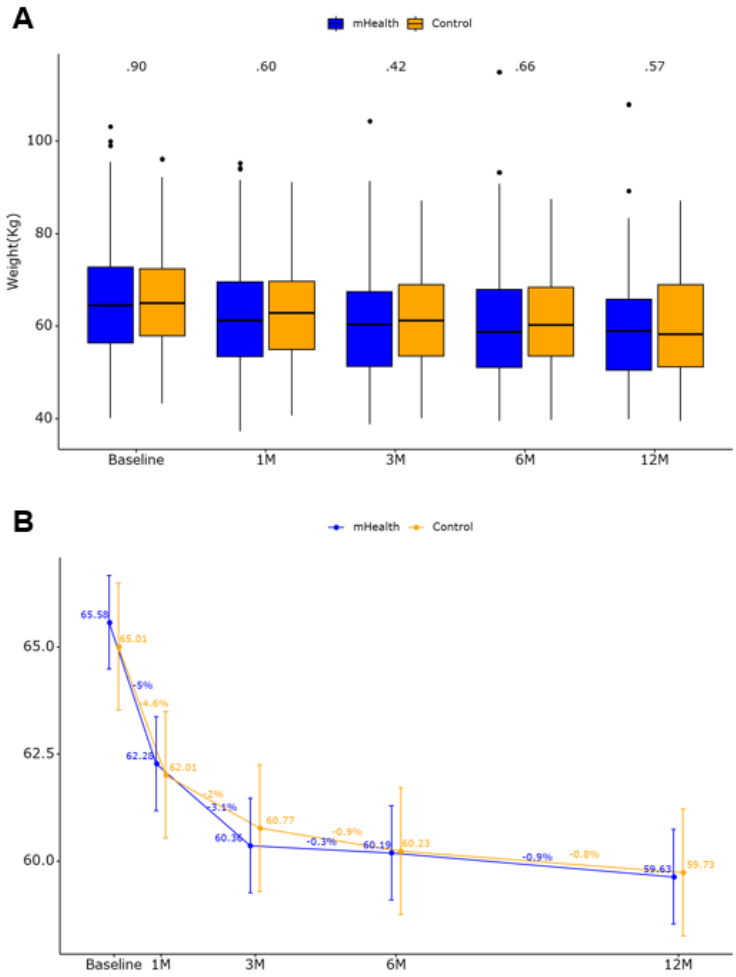


Table 2. Analysis of primary outcome.

Time point	Control	mHealth[a]	*P* value
Baseline, mean (SD)	65.54 (10.60)	65.35 (11.96)	.90
1 month, mean (SD)	62.85 (10.63)	62.05 (11.53)	.60
3 months, mean (SD)	61.46 (10.84)	60.14 (11.83)	.42
6 months, mean (SD)	60.89 (10.97)	60.16 (12.15)	.66
12 months, mean (SD)	60.20 (11.63)	59.18 (11.85)	.57
Within-group adjusted mean change from baseline to 12 months[b], mean (95% CI)	−5.29 (−6.22 to −4.35)	−5.95 (−6.62 to −5.28)	.26[c]

[a]mHealth: mobile health.

[b]Estimated using linear mixed-effects models adjusted for age, drinking status, and adjuvant chemotherapy. Values represent the adjusted mean change in body weight from baseline to 12 months within each group.

[c]Between-group adjusted difference in mean change (mHealth-control): 0.10 (95% CI −3.01 to 3.21). A positive value reflects a greater weight loss in the mHealth group compared to the control group. Statistical significance was set at *P*<.05.

Secondary Outcomes

The secondary outcomes of the study are summarized in Table 3. Variables analyzed included physical fitness (grip strength test, 30-second chair stand test, and 2-minute walk test), nutrition (screening and total MNA), quality of life (EORTC QLQ-C30 and EORTC QLQ-STO22), physical activity (IPAQ-SF), physical status (skeletal muscle mass, lean body mass, fat mass, BMI, hemoglobin, vitamin B_{12}, and albumin), and pain intensity (average NRS). No significant group-by-time interactions or between-group differences were observed for any of these outcomes, except for the nausea/vomiting subscale

XSL•FO
RenderX

of the EORTC QLQ-C30. Although the Wilcoxon rank-sum test indicated a statistically significant difference between groups ($P<.05$), both groups had identical medians and IQRs, suggesting that this result likely reflects subtle rank-based variation rather than a clinically meaningful difference in symptom burden. To further evaluate the robustness of the secondary outcome results and to complement the change-score analyses, additional longitudinal analyses were performed using GEE. The GEE results were consistent with the primary analyses, demonstrating no significant group-by-time interaction effects across all secondary outcomes (Table S5 in Multimedia Appendix 2).

XSL•FO
RenderX

Table 3. Analysis of secondary outcomes. For variables following a normal distribution, linear mixed-effects models were used to estimate adjusted mean changes from baseline to 12 months within each group, accounting for relevant covariates. For nonnormally distributed variables, the Wilcoxon rank-sum test was applied. "mHealth-control" indicates the estimated between-group difference in adjusted change from baseline to 12 months. A positive value reflects greater loss in the mobile health (mHealth) group compared to the control group. Statistical significance was defined as $P<.05$.

Outcome	Control	mHealth	*P* value	mHealth-control
Skeletal muscle mass, mean (95% CI)	−1.2 (−1.57 to −0.82)	−1.35 (−1.62 to −1.08)	.50	0.19 (−0.25 to 0.62)
BMI, mean (95% CI)	−2.14 (−2.50 to −1.77)	−2.13 (−2.40 to −1.87)	.99	0 (−0.43 to 0.43)
Skeletal muscle index, mean (95% CI)	−0.08 (−0.18 to 0.02)	−0.12 (−0.19 to −0.05)	.57	0.05 (−0.08 to 0.18)
Fat mass, mean (95% CI)	−3.05 (−3.91 to −2.18)	−3.90 (−4.52 to −3.28)	.12	0.74 (−0.19 to 1.67)
Lean body mass, mean (95% CI)	−1.45 (−2.26 to −0.63)	−1.66 (−2.25 to −1.08)	.67	0.33 (−0.59 to 1.25)
Grip strength, mean (95% CI)	1.48 (−0.37 to 3.33)	2.72 (1.43 to 4.00)	.28	−1.08 (−2.94 to 0.79)
Relative grip strength, mean (95% CI)	7.36 (4.36 to 10.35)	8.43 (6.34 to 10.51)	.57	−0.63 (−3.67 to 2.41)
30-second chair stand test, mean (95% CI)	5.75 (4.76 to 6.75)	5.44 (4.71 to 6.17)	.62	0.41 (−0.66 to 1.49)
2-minute walk test, mean (95% CI)	28.25 (22.22 to 34.28)	28.16 (23.80 to 32.53)	.98	1.40 (−4.51 to 7.30)
Average NRS[a], median (IQR)	−3.00 (−4.00 to −2.00)	−3.00 (−3.00 to −2.00)	.90	0 (0 to 1)
IPAQ-SF[b] , median (IQR)				
Total MET[c]	372.50 (−702.75 to 2204.25)	561.00 (−371.25 to 1977.00)	.87	54 (−606 to 666)
Vigorous MET	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	.25	0 (0 to 0)
Moderate MET	0.00 (−240.00 to 280.00)	0.00 (0.00 to 600.00)	.19	0 (0 to 240)
Walking MET	0.00 (−618.75 to 1178.10)	346.50 (−214.50 to 1155.00)	.59	99 (−363 to 544.5)
MNA[d] , median (IQR)				
MNA screening score	0.00 (−2.00 to 1.25)	0.00 (−2.00 to 1.00)	.98	0 (−1 to 1)
MNA total score	0.00 (−2.62 to 2.50)	0.50 (−1.50 to 2.00)	.37	0.5 (−0.5 to 1.5)
EORTC QLQ-C30[e] , median (IQR)				
Physical functioning	0.00 (−6.67 to 6.67)	0.00 (−6.67 to 6.67)	.88	0 (0 to 0)
Role functioning	0.00 (−16.67 to 0.00)	0.00 (0.00 to 0.00)	.18	0 (0 to 0)
Cognitive functioning	0.00 (−16.67 to 0.00)	0.00 (0.00 to 16.67)	.08	0 (0 to 0)
Emotional functioning	8.33 (0.00 to 16.67)	0.00 (0.00 to 16.67)	.55	0 (−8.33 to 0)
Social functioning	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	.20	0 (0 to 0)
Fatigue	0.00 (−11.11 to 11.11)	0.00 (−11.11 to 11.11)	.33	0 (−11.11 to 0)
Nausea/vomiting	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	.03	0 (0 to 0)
Pain	0.00 (−16.67 to 0.00)	0.00 (−16.67 to 0.00)	.29	0 (0 to 0)
Dyspnea	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	.75	0 (0 to 0)
Insomnia	0.00 (−33.33 to 0.00)	0.00 (−33.33 to 0.00)	.70	0 (0 to 0)
Appetite loss	0.00 (0.00 to 8.33)	0.00 (0.00 to 0.00)	.08	0 (0 to 0)
Constipation	0.00 (0.00 to 0.00)	0.00 (−33.33 to 0.00)	.34	0 (0 to 0)
Diarrhea	0.00 (0.00 to 33.33)	0.00 (0.00 to 33.33)	.52	0 (0 to 0)
Financial difficulties	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	.96	0 (0 to 0)
Global health status (QoL[f])	0.00 (−10.42 to 16.67)	0.00 (−8.33 to 16.67)	.75	0 (−8.33 to 8.33)
EORTC QLQ-STO22[g] , median (IQR)				
Dysphagia	0.00 (−11.11 to 0.00)	0.00 (−11.11 to 0.00)	.43	0 (0 to 0)
Pain	0.00 (−16.67 to 8.33)	−8.33 (−16.67 to 8.33)	.58	0 (−8.33 to 0)

XSL•FO
RenderX

Outcome	Control	mHealth	*P* value	mHealth-control
Reflux symptoms	0.00 (−11.11 to 11.11)	0.00 (−11.11 to 0.00)	.49	0 (0 to 0)
Eating restrictions	0.00 (−16.67 to 8.33)	0.00 (−14.58 to 8.33)	.85	0 (−8.33 to 8.33)
Anxiety	0.00 (−11.11 to 11.11)	0.00 (−11.11 to 11.11)	.48	0 (−11.11 to 0)
Dry mouth	0.00 (−33.33 to 0.00)	0.00 (0.00 to 0.00)	.56	0 (0 to 0)
Taste	0.00 (−8.33 to 0.00)	0.00 (−33.33 to 0.00)	.77	0 (0 to 0)
Body image	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	.58	0 (0 to 0)
Hair loss	16.67 (0.00 to 33.33)	0.00 (0.00 to 25.00)	.61	0 (−66.67 to 33.33)
Hb[h], mean (95% CI)	−0.80 (−1.70 to 0.10)	−0.60 (−1.40 to 0.30)	.56	0.1 (−0.3 to 0.6)
Vitamin B_{12}, mean (95% CI)	−215.00 (−362.00 to −27.00)	−210.50 (−325.75 to 63.50)	.56	60.5 (−221 to 410)
Albumin, mean (95% CI)	−0.10 (−0.20 to 0.10)	0.00 (−0.30 to 0.17)	.42	0 (−0.1 to 0.2)

[a]NRS: Numeric Rating Scale.

[b]IPAQ-SF: International Physical Activity Questionnaire-Short Form.

[c]MET: metabolic equivalent of task.

[d]MNA: Mini Nutritional Assessment.

[e]EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30.

[f]QoL: quality of life.

[g]EORTC QLQ-STO22: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Stomach 22.

[h]Hb: hemoglobin.

eHEALS

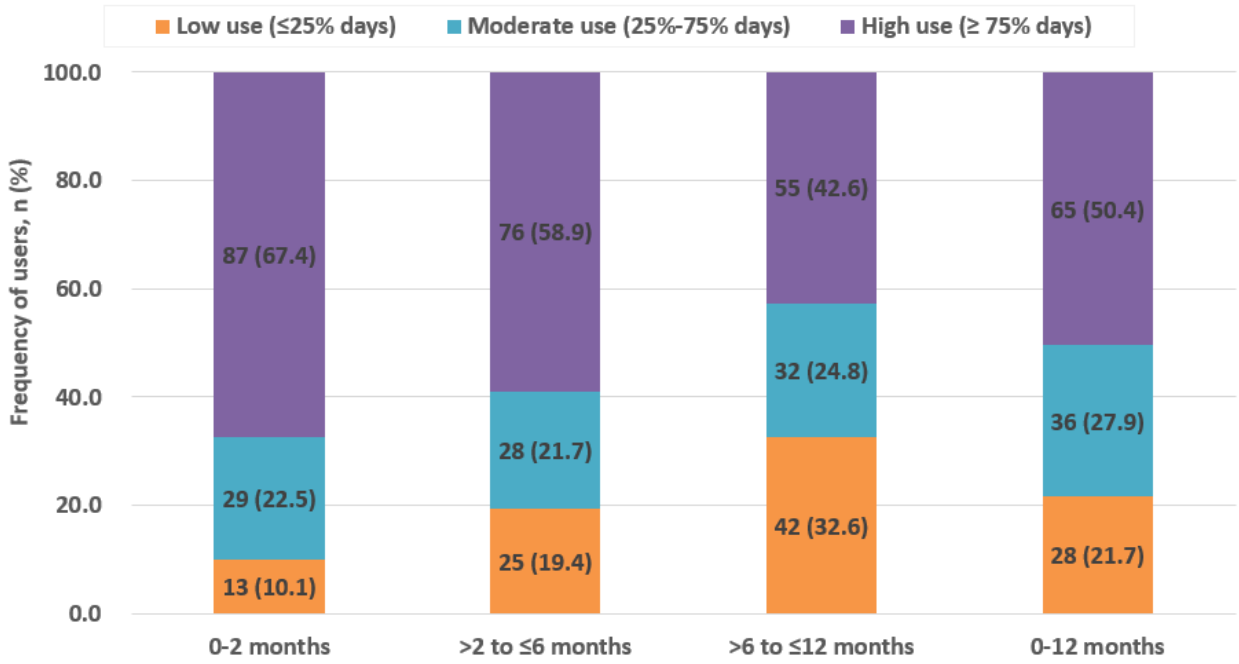
eHealth literacy was assessed only at baseline in the intervention group using the eHEALS. The mean score was 29.8 (SD 4.05) out of 40, indicating a moderate level of eHealth literacy. This suggests that while participants were likely capable of independently searching for health information online, they may have had limitations in evaluating accuracy and reliability.

Compliance

Compliance with wearable device–linked mobile app use over 12 months is presented in Figure 4. Usage was categorized as high (≥75% of days), moderate (25%-74% of days), and low (≤25% of days). Compliance was highest during the initial 0-2 months, followed by a gradual decline during the 2- to 6-month and 6- to 12-month intervals. Nevertheless, overall adherence was favorable, with approximately 78.3% (101/129) of the participants demonstrating moderate to high use throughout the study period.

XSL•FO
RenderX

Figure 4. Compliance with wearable device–linked mobile app use over 12 months. Stacked bars show the number and proportion of participants with low (≤25%), moderate (25%-74%), and high (≥75%) usage across 4 time intervals: 0-2 months, 2-6 months, 6-12 months, and the overall 12-month period.

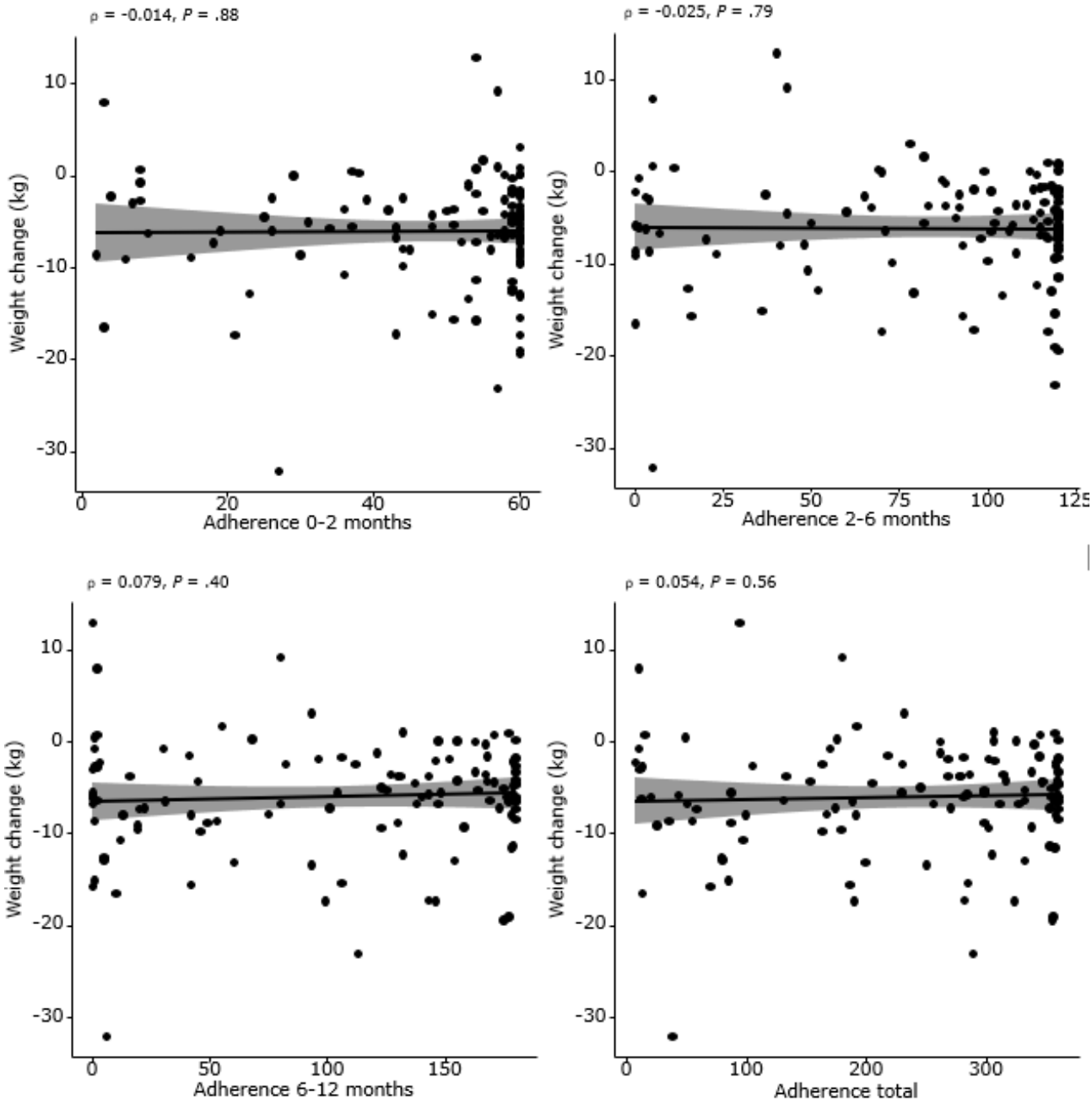


Association Between Adherence and Weight Change

To explore a potential dose-response relationship, we analyzed the correlation between app adherence and 12-month weight change. Adherence was calculated separately for 3 time intervals (0-2 months, 2-6 months, and 6-12 months) and for the total intervention period. As shown in Figure 5, Spearman correlation analyses revealed no statistically significant associations between adherence and weight change in any period: ρ=–0.014 (*P*=.88) for 0-2 months; ρ=–0.025 (*P*=.79) for 2-6 months; ρ=0.079 (*P*=.40) for 6-12 months; ρ=0.054 (*P*=.56) for the total study period. These results suggest that higher app engagement levels were not significantly associated with attenuated postoperative weight loss in this cohort.

XSL•FO
RenderX

Figure 5. Correlation between intervention adherence and 12-month weight change from baseline.



App-Based Adherence Measures

Among the 131 participants in the intervention group who completed the 12-month follow-up, data from 129 were included in the adherence analysis. Individuals who discontinued app use or withheld consent for data access at the time of data retrieval were excluded. App usage patterns over the 12-month intervention period are summarized in Table 4. Participants recorded exercise on an average of 18.2 (SD 10.0) days per month, with a mean duration of 56.9 (SD 26.9) minutes per exercise day. The number of step count log days remained relatively stable throughout the follow-up (overall mean 24.2, SD 8.7 days per month), with an average of 8491.1 (SD 3765.5) steps per day. Weight and diet logs were entered on an average of 9.9 (SD 10.9) and 22.7 (SD 8.8) days per month, respectively. App access was recorded on 21.3 (SD 10.0) days per month. While the frequency of exercise logging was higher in the earlier months, average exercise duration per exercise day and daily step counts showed an increasing trend over time, suggesting improved engagement in physical activity among users who continued to log their data.

XSL•FO
RenderX

Table 4. Monthly digital intervention engagement: exercise, step count, weight logs, diet logs, and app use (n=129). Exercise logs include any session (free-form, video-guided, or manually entered) lasting at least 5 minutes. Aerobic exercise refers to sessions lasting ≥10 minutes. Step counts, weight, and dietary logs were self-reported via the digital platform.

Month	Exercise log (days), mean (SD)	Exercise duration per log day (min), mean (SD)	Step count logged (days), mean (SD)	Step count per log day (steps), mean (SD)	Weight log (days), mean (SD)	Diet log (days), mean (SD)	App access (days), mean (SD)
1	14.3 (10.7)	42.4 (28.6)	22.7 (9.2)	7435.4 (3751.4)	8.4 (8.9)	22.0 (9.7)	22.4 (9.6)
2	19.8 (10.0)	44.4 (26.2)	24.6 (8.8)	8428.4 (4138.4)	10.7 (10.8)	24.4 (7.4)	22.6 (9.3)
3	19.2 (10.4)	46.4 (23.2)	23.8 (9.0)	8099.8 (3592.5)	10.7 (11.6)	23.1 (9.3)	22.3 (9.4)
4	18.7 (10.1)	58.6 (27.3)	23.9 (9.1)	8327.0 (3535.0)	10.7 (11.4)	23.5 (8.9)	21.4 (10.2)
5	19.6 (9.3)	63.1 (28.2)	24.4 (8.7)	8836.8 (4294.3)	10.3 (11.4)	23.6 (8.1)	22.0 (9.7)
6	19.7 (9.1)	61.2 (25.2)	24.5 (8.3)	8314.8 (3480.2)	10.3 (11.1)	22.9 (8.4)	21.3 (10.0)
7	18.6 (9.5)	62.2 (24.5)	23.7 (8.7)	8310.9 (3599.6)	10.0 (11.3)	22.4 (8.8)	20.8 (10.4)
8	17.9 (9.7)	64.3 (23.3)	25.1 (7.8)	8947.9 (3803.8)	9.3 (10.8)	22.0 (9.1)	20.7 (10.1)
9	18.4 (9.5)	65.5 (25.5)	25.0 (8.4)	9316.1 (3634.5)	9.5 (11.2)	22.1 (8.8)	21.0 (9.6)
10	18.6 (9.2)	64.8 (23.1)	25.3 (8.5)	9261.3 (3552.2)	10.0 (11.1)	21.8 (8.8)	20.6 (10.4)
11	17.0 (10.1)	64.8 (26.5)	25.2 (7.9)	9071.0 (3547.0)	9.8 (11.1)	21.7 (8.9)	19.8 (10.6)
12	17.5 (10.1)	62.0 (21.2)	23.5 (9.3)	8550.0 (3694.4)	9.5 (11.0)	21.1 (9.3)	19.2 (10.9)
Overall	18.2 (10.0)	56.9 (26.9)	24.2 (8.7)	8491.1 (3765.5)	9.9 (10.9)	22.7 (8.8)	21.3 (10.0)

Satisfaction Survey

A customized satisfaction survey was conducted with participants in the intervention group at the 6-month follow-up (n=113). Overall, responses reflected a high degree of satisfaction with the digital health care service (Table 5).

Among the assessed domains, perceived effectiveness received the most favorable ratings, with average scores ranging from 4.0 (SD 0.6) to 4.1 SD (0.7). These results suggest that participants felt the service contributed meaningfully to their efforts to monitor and manage their health. Items addressing usability also performed well (means ranging from 3.9, SD 0.8 to 4.0, SD 0.9), indicating that participants generally found the app intuitive and easy to use.

In the domain of content adequacy, scores were slightly lower, with the item evaluating whether the service included all essential features for health management receiving a mean score of 3.7 (SD 0.8). Many participants noted that the exercise and dietary management tools were especially useful. On the other hand, several respondents identified dietary tracking as an area requiring further refinement.

XSL•FO
RenderX

Table 5. Satisfaction survey results at 6-month follow-up (intervention group; n=113).

Survey question	Mean (SD)
Use and access	
The service provided sufficient in-app instructions on how to use its features.	4.0 (0.9)
Understanding how to operate the service was straightforward[a].	3.9 (0.8)
I was able to easily locate and use the functions I needed.	3.9 (0.8)
Content adequacy	
The service included all the essential features required for managing my health.	3.7 (0.8)
The service offered sufficient and helpful information that supported my health management and lifestyle improvement[a].	3.9 (0.7)
Satisfaction	
Overall, how satisfied are you with the healthcare service you used?[a]	3.9 (0.7)
Do you intend to continue using the healthcare service in the future?[a]	3.8 (0.9)
Would you recommend this healthcare service to other patients?[a]	3.9 (0.8)
Perceived effectiveness	
Since using the service, have you monitored your health status more regularly?	4.0 (0.6)
Has your interest in managing your personal health increased after using the service?	4.0 (0.6)
Was the service useful in supporting your personal health management?	4.0 (0.6)
Do you consider the service a suitable method for managing your condition?	4.1 (0.7)

[a]Included a follow-up question to elaborate on their answers.

Discussion

Principal Findings

This study investigated the long-term effects of a personalized, stage-adjusted digital health program combining exercise and nutritional interventions in patients with GC, delivered for 12 months following surgery. While overall satisfaction with and compliance with the digital health intervention were relatively high, the intervention group did not show a statistically significant benefit in mitigating postoperative body weight loss compared to the control group. Secondary outcomes, including body composition, BMI, physical fitness level, nutrition, and quality of life, did not demonstrate meaningful improvements across time points. The overall consistency between the results of nonparametric tests and GEE analyses suggests that the absence of statistically significant group-by-time interaction effects is unlikely due to model selection. This concordance reinforces the robustness and validity of our findings related to the secondary outcomes.

The hypothesized mechanisms by which a digital health program combining exercise and nutrition might influence postoperative weight loss after gastrectomy can be conceptualized in 3 interrelated domains: behavioral activation, metabolic/nutritional modulation, and muscle-preserving functional support. The app and wearable device supported self-monitoring, goal setting, and feedback, promoting adherence to physical activity and dietary goals. These behavioral components are especially valuable during early recovery, when fatigue and reduced intake are common. In addition, prior studies in populations with cancer support the effectiveness of such digital tools in sustaining health behaviors [43]. Postoperative weight loss in patients with GC is a multifactorial process influenced by mechanical, hormonal, and metabolic changes (eg, increased catabolism and inflammatory responses). These alterations result in decreased oral intake, impaired nutrient absorption, and significant losses in skeletal muscle mass and body weight. Structured digital interventions can modulate these adverse trajectories by optimizing caloric intake, preserving lean body mass, and minimizing nutrient deficiencies [44,45]. Furthermore, exercise prescriptions were tailored using cancer guidelines, combining aerobic and resistance training to mitigate sarcopenia and preserve function [24]. The individualized prescriptions targeted aerobic and resistance training intensity using patient-specific maximum heart rate and stage-based progression, aiming to preserve lean mass and physical function. Digital delivery enabled sustained support during the high-risk postdischarge period, allowing flexible, remote adjustments to care without increasing burden [10].

Nevertheless, several factors may explain why the 12-month digital health intervention did not significantly alter the postoperative weight loss trajectory in patients with GC compared to conventional education-based rehabilitation.

Reach

Our trial enrolled 257 patients after curative resection for stage I-III GC. As noted in the manuscript, the study population had relatively preserved nutritional status (mean BMI 24.05, SD 3.36 kg/m^2 in the intervention group, 24.05, SD 2.79 kg/m^2 in the control group; *P*=.99), with only 2 participants (BMI 16.1 kg/m^2 and 16.2 kg/m^2) having BMI <17 kg/m^2. This indicates

XSL•FO
RenderX

that the study population predominantly comprised individuals at a relatively lower baseline risk of nutritional compromise, potentially omitting those at the highest nutritional risk who may have derived the greatest benefit from the intervention. In addition, although current clinical guidelines emphasize dietary practices to prevent complications [46,47], many patients fail to adopt them due to insufficient cognitive factors, such as limited risk perception, weak motivation, or limited expectations of benefit [48-50]. A study in patients with upper gastrointestinal cancer showed that weight loss was attenuated only with telephone counseling, while mobile app–based support alone was insufficient to ensure nutritional adequacy [51]. Moreover, a Korean survey reported that the general public perceived digital therapeutics less favorably than health care professionals [52].

Effectiveness

Although the personalized mHealth program was safe and well-received—with high early-phase adherence and no reported adverse events—it did not yield statistically significant differences in weight change or secondary clinical outcomes (eg, quality of life, physical function, and nutritional biomarkers). This highlights the importance of distinguishing between biological efficacy and intervention usability/satisfaction, especially in patient populations with a low baseline risk. GC has a unique biological trajectory compared with that of other malignancies. Patients undergoing gastrectomy typically exhibit substantial changes in body composition, characterized by marked reductions in body weight during the first 6 months, followed by gradual stabilization [10,53]. One study reported that the most pronounced weight changes occurred within 6-12 months, with stabilization or recovery extending up to 2 years after surgery [1]. This biphasic course may be resistant to reversal even with structured digital interventions. Postoperative weight loss in GC is multifactorial, driven by inflammatory responses, reduced intake, and decreased physical activity. Tumor-related malnutrition and immune dysfunction are further aggravated by radical surgery, underscoring the need for both nutritional and immune-supportive interventions [54,55]. This is consistent with findings that most oral nutritional supplementation trials in GC have not shown consistent benefits in preventing postoperative weight loss [13,43,44]. While GC and colorectal cancer share gastrointestinal origins, GC is associated with poorer survival and more aggressive progression due to differences in the underlying metabolic pathways [56-58]. Although recent randomized controlled trials in colorectal cancer showed quality-of-life improvements through mHealth, these results may not be generalizable to GC due to these fundamental biological differences.

In addition, the conservative nature of the intervention, largely driven by safety considerations, may have limited its efficacy. Unlike supervised rehabilitation programs, the current unsupervised digital program required cautious exercise prescriptions. Moreover, both groups received comprehensive standard care, narrowing the margin for additional benefit.

Another potential contributing factor is the lack of real-time symptom management and psychological support, as the program focused solely on exercise and nutrition. Psychological support is known to be critical in postoperative quality of life for patients with GC, breast cancer, and esophagogastric cancer [59-63]. However, due to the COVID-19 pandemic, it was not feasible to incorporate family or peer-based support. Future mHealth programs should incorporate humanized psychological support and caregiver-focused components [3,60,62].

Adoption and Implementation

The intervention was implemented across multiple hospitals and integrated into routine postgastrectomy care, delivered through a mobile app and wearable device. The program was initiated within 1 week after surgery and maintained for 12 months, aligning with critical recovery periods. However, we observed a progressive decline in engagement, particularly during the 6- to 12-month period—a common challenge in long-term digital interventions. While this low-touch, patient-led design ensured feasibility and safety, the absence of active supervision may have limited behavior change, particularly in patients with lower health literacy or motivation. Although exercise and nutrition prescriptions were algorithmically tailored, adherence to these prescriptions in real-world settings was not directly assessed, limiting our understanding of implementation fidelity.

Maintenance

Although the intervention spanned 12 months, true maintenance—sustained behavior changes and functional improvement—often extends beyond this time frame. Systematic reviews of web-based digital health interventions have demonstrated efficacy attenuation over extended time frames due to digital fatigue, with short-term benefits failing to translate into sustained long-term outcomes [64-67]. Therefore, further research in GC is needed to investigate strategies that can maintain high patient engagement over periods exceeding 2 years, as well as including approaches such as gamification to enhance motivation [68].

To maximize the effectiveness of digital health, it is essential to deliver the right content to the right patients at the right time. For example, an Internet of Things–based lifestyle intervention in prostate cancer demonstrated benefits among patients receiving androgen deprivation therapy or those with a high symptom burden, whereas effects were inconsistent in broader populations and across different intervention durations [69,70]. In our study, the digital health program, initiated immediately after gastrectomy and maintained for 12 months at a safe intensity, might have served mainly as supportive monitoring without conferring superior effects on the biological course of GC. Therefore, further research on precise targeting is warranted to tailor intervention content to the specific characteristics of each cancer type.

Strengths and Limitations

This study has several notable strengths. First, to the best of our knowledge, this is the largest and longest randomized controlled trial to date to evaluate a mHealth and nutrition platform in patients with GC, enrolling 257 participants with a follow-up period of 12 months. Second, the digital health program was highly personalized and dynamically adjusted based on the

XSL•FO
RenderX

patient's treatment phase and clinical condition. Third, the program was initiated within 1 week after surgery, enabling early rehabilitation and nutritional support during a critical recovery window. Fourth, the program was safe and well-received, with no adverse events and high levels of satisfaction and adherence. This reinforces its clinical relevance as a feasible and acceptable model of digital rehabilitation. Fifth, although mHealth interventions have been reported to be broadly beneficial in cancer rehabilitation, this study allowed for a focused examination of their applicability to GC. Our findings suggest a novel viewpoint that a digital health program centered solely on nutrition and exercise for 12 months may have limited effectiveness in the postoperative GC population, unlike some other cancers. Finally, our study adhered to CONSORT/SPIRIT standards.

This study has several limitations. First, this was an open-label design that was not free of bias. However, this reflects real-world clinical settings, thereby enhancing the external validity of the findings. In addition, compliance gradually declined at 2- to 6-and 6- to 12-month intervals. Nevertheless, overall adherence remained favorable. Weight change after gastrectomy for GC is most pronounced within the first 6-12 months, with subsequent stabilization or improvement observed up to 2 years postoperatively [1,71]. Given that our study was limited to a 12-month follow-up period, the long-term recovery trajectory could not be fully characterized, considering the physiology of GC. Moreover, most participants entered the study with relatively preserved nutritional status, which may have contributed to a ceiling effect and reduced the magnitude of observable benefit, particularly for nutritional outcomes. This profile likely reflects a selection bias that patients with greater frailty or poor performance status may be less likely to enroll or remain eligible. Furthermore, as the study was powered based solely on the primary outcome with an assumed moderate effect size, it may have been underpowered to detect smaller but clinically meaningful effects in secondary end points. Given the large number of secondary outcomes assessed, the absence of multiplicity adjustment may increase the risk of type I error (false positives). At the same time, because the sample size was not specifically calculated to detect effects in secondary variables, the study may also be subject to type II error (false negatives) for outcomes with small effect sizes. Accordingly, analyses of secondary outcomes were exploratory in nature, and no adjustments were made for multiple comparisons. Therefore, findings related to secondary end points should be interpreted with caution and regarded as hypothesis-generating rather than confirmatory. While adherence metrics were reported, key implementation outcomes related to intervention reach (eg, recruitment yield, refusal reasons, and population representativeness) were not quantitatively captured. This represents a limitation in fully characterizing the implementation context and may constrain the interpretation of broader applicability. Finally, participants were recruited from a single monoracial country. Future studies should aim to recruit more diverse populations to enhance global generalizability.

Conclusion

In summary, this randomized controlled trial showed that a personalized digital health program integrating exercise and nutrition is safe, feasible, and associated with high satisfaction and adherence among patients with GC in the early postoperative period. However, the intervention did not produce measurable benefits over standard education in postoperative weight change or other functional and nutritional outcomes. These results suggest that when applied broadly across all postoperative patients, digital health tools may function primarily as supportive monitoring platforms rather than interventions that biologically augment recovery. Future research is needed to identify subgroups that may benefit most, optimize intervention timing and intensity, and determine how digital programs can complement existing rehabilitation pathways.

Acknowledgments

We would like to thank Medi Plus Solution for technical support and all hospital members who devoted their efforts to the study.

Funding

This research was supported by the National IT Industry Promotion Agency, funded by the Ministry of Science and Information and Communications Technology of Korea (project numbers S2001-20-1006 and S2401-21-1001, Project Healthcare Big Data Showcase Utilization Service Support: Cancer Patient Treatment and Prognosis Management Service Development). The funder played no role in the study design, data collection, analysis, and interpretation, or writing of the manuscript.

Data Availability

The datasets used and/or analyzed in this study are available from the corresponding author upon reasonable request.

Authors' Contributions

IK was responsible for writing the original draft and for review and editing, as well as for formal analysis and conceptualization. JY Lim and SWK contributed to the investigation, data curation, resources, and methodology. JHL, TSS, SP, and SHK were involved in providing resources and in data curation and methodology. JY Lee and JHH led the conceptualization and methodology and were responsible for supervision, project administration, review and editing, and funding acquisition. All authors reviewed and approved the final version of the manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional tables.
[DOCX File , 34 KB - mhealth_v14i1e85795_app1.docx]

Multimedia Appendix 2
CONSORT-EHEALTH checklist.
[PDF File (Adobe PDF File), 1299 KB - mhealth_v14i1e85795_app2.pdf]

References

1. Davis JL, Selby LV, Chou JF, Schattner M, Ilson DH, Capanu M, et al. Patterns and predictors of weight loss after gastrectomy for cancer. Ann Surg Oncol 2016;23(5):1639-1645 [FREE Full text] [doi: 10.1245/s10434-015-5065-3] [Medline: 26732274]
2. Yu W, Seo BY, Chung HY. Postoperative body-weight loss and survival after curative resection for gastric cancer. Br J Surg 2002;89(4):467-470. [doi: 10.1046/j.0007-1323.2001.02046.x] [Medline: 11952589]
3. Carrillo Lozano E, Osés Zárate V, Campos Del Portillo R. Nutritional management of gastric cancer. Endocrinol Diabetes Nutr (Engl Ed) 2021;68(6):428-438. [doi: 10.1016/j.endien.2020.09.005] [Medline: 34742476]
4. Zhan C, Bu J, Li S, Huang X, Quan Z. Postoperative skeletal muscle loss as a prognostic indicator of clinical outcomes in patients with gastric cancer: a systematic review and meta-analysis. J Gastrointest Surg 2025;29(2):101898. [doi: 10.1016/j.gassur.2024.101898] [Medline: 39608746]
5. Shuto K, Nabeya Y, Mori M, Yamazaki M, Kosugi C, Narushima K, et al. Postoperative changes in body composition predict long-term prognosis in patients with gastric cancer. Cancers (Basel) 2025;17(5):738 [FREE Full text] [doi: 10.3390/cancers17050738] [Medline: 40075586]
6. Zhang J, Kong Q, Zhang J, Guo J. Effectiveness of nutritional support for clinical outcomes in gastric cancer patients: a meta-analysis of randomized controlled trials. Open Med (Wars) 2024;19(1):20241023. [doi: 10.1515/med-2024-1023] [Medline: 39247438]
7. Park KB, Kwon OK, Yu W. Midterm body composition changes after open distal gastrectomy for early gastric cancer. Ann Surg Treat Res 2018;95(4):192-200 [FREE Full text] [doi: 10.4174/astr.2018.95.4.192] [Medline: 30310802]
8. Imamura H, Nishikawa K, Kishi K, Inoue K, Matsuyama J, Akamaru Y, et al. Effects of an oral elemental nutritional supplement on post-gastrectomy body weight loss in gastric cancer patients: a randomized controlled clinical trial. Ann Surg Oncol 2016;23(9):2928-2935. [doi: 10.1245/s10434-016-5221-4] [Medline: 27084538]
9. Triantafillidis JK, Papakontantinou J, Antonakis P, Konstadoulakis MM, Papalois AE. Enteral nutrition in operated-on gastric cancer patients: an update. Nutrients 2024;16(11):1639 [FREE Full text] [doi: 10.3390/nu16111639] [Medline: 38892572]
10. Kimura Y, Nishikawa K, Kishi K, Inoue K, Matsuyama J, Akamaru Y, et al. Long-term effects of an oral elemental nutritional supplement on post-gastrectomy body weight loss in gastric cancer patients (KSES002). Ann Gastroenterol Surg 2019;3(6):648-656 [FREE Full text] [doi: 10.1002/ags3.12290] [Medline: 31788653]
11. Lee MK, Oh J. Patient-reported outcomes of regular aerobic exercise in gastric cancer. Cancers (Basel) 2021;13(9):2080 [FREE Full text] [doi: 10.3390/cancers13092080] [Medline: 33923080]
12. Cho I, Son Y, Song S, Bae YJ, Kim YN, Kim H, et al. Feasibility and effects of a postoperative recovery exercise program developed specifically for gastric cancer patients (PREP-GC) undergoing minimally invasive gastrectomy. J Gastric Cancer 2018;18(2):118-133 [FREE Full text] [doi: 10.5230/jgc.2018.18.e12] [Medline: 29984062]
13. Soh JY, Cha WC, Chang DK, Hwang JH, Kim K, Rha M, et al. Development and validation of a multidisciplinary mobile care system for patients with advanced gastrointestinal cancer: interventional observation study. JMIR Mhealth Uhealth 2018;6(5):e115 [FREE Full text] [doi: 10.2196/mhealth.9363] [Medline: 29735478]
14. Wu JM, Ho TW, Chang YT, Hsu CC, Tsai CJ, Lai F, et al. Wearable-based mobile health app in gastric cancer patients for postoperative physical activity monitoring: focus group study. JMIR Mhealth Uhealth 2019;7(4):e11989 [FREE Full text] [doi: 10.2196/11989] [Medline: 31012858]
15. Agboola S, Kamdar M, Flanagan C, Searl M, Traeger L, Kvedar J, et al. Pain management in cancer patients using a mobile app: study design of a randomized controlled trial. JMIR Res Protoc 2014;3(4):e76 [FREE Full text] [doi: 10.2196/resprot.3957] [Medline: 25500281]
16. Jiang XH, Yuan XH, Chen JM, Yu H, Chen XJ, Chen YH, et al. Effects of a nurse-led individualized mHealth nutrition intervention for post-discharged gastric cancer patients following gastrectomy: a randomized controlled trial. Int J Nurs Stud 2025;168:105092. [doi: 10.1016/j.ijnurstu.2025.105092] [Medline: 40253832]
17. Park JH, Lee HJ, Kim JS, Cho YS, Lee S, Park S, et al. Development and feasibility assessment of mobile application-based digital therapeutics for postoperative supportive care in gastric cancer patients following gastrectomy. J Gastric Cancer 2024;24(4):420-435 [FREE Full text] [doi: 10.5230/jgc.2024.24.e37] [Medline: 39375057]

XSL•FO
RenderX

18. Hanna L, Huggins CE, Furness K, Silvers MA, Savva J, Frawley H, et al. Effect of early and intensive nutrition care, delivered via telephone or mobile application, on quality of life in people with upper gastrointestinal cancer: study protocol of a randomised controlled trial. BMC Cancer 2018;18(1):707 [FREE Full text] [doi: 10.1186/s12885-018-4595-z] [Medline: 29970033]
19. Zhang Z, Tian L, He K, Xu L, Wang X, Huang L, et al. Digital rehabilitation programs improve therapeutic exercise adherence for patients with musculoskeletal conditions: a systematic review with meta-analysis. J Orthop Sports Phys Ther 2022;52(11):726-739. [doi: 10.2519/jospt.2022.11384] [Medline: 35960507]
20. Maisto M, Diana B, Di Tella S, Matamala-Gomez M, Montana JI, Rossetto F, et al. Digital interventions for psychological comorbidities in chronic diseases-a systematic review. J Pers Med 2021;11(1):30 [FREE Full text] [doi: 10.3390/jpm11010030] [Medline: 33418971]
21. Mortazavi BJ, Gutierrez-Osuna R. A review of digital innovations for diet monitoring and precision nutrition. J Diabetes Sci Technol 2023;17(1):217-223 [FREE Full text] [doi: 10.1177/19322968211041356] [Medline: 34467803]
22. Jiang X, Chen J, Yuan X, Lin Y, Chen Y, Li S, et al. Feasibility of an individualized mhealth nutrition (iNutrition) intervention for post-discharged gastric cancer patients following gastrectomy: a randomized controlled pilot trial. Nutrients 2023;15(8) [FREE Full text] [doi: 10.3390/nu15081883] [Medline: 37111102]
23. Rausei S, Mangano A, Galli F, Rovera F, Boni L, Dionigi G, et al. Quality of life after gastrectomy for cancer evaluated via the EORTC QLQ-C30 and QLQ-STO22 questionnaires: surgical considerations from the analysis of 103 patients. Int J Surg 2013;11 Suppl 1:S104-S109 [FREE Full text] [doi: 10.1016/S1743-9191(13)60028-X] [Medline: 24380539]
24. Campbell KL, Winters-Stone KM, Wiskemann J, May AM, Schwartz AL, Courneya KS, et al. Exercise guidelines for cancer survivors: consensus statement from international multidisciplinary roundtable. Med Sci Sports Exerc 2019;51(11):2375-2390 [FREE Full text] [doi: 10.1249/MSS.0000000000002116] [Medline: 31626055]
25. Effects of excercise on health-related outcomes in those with cancer. ACSM. 2022. URL: https://www.exerciseismedicine.org/wp-content/uploads/2021/04/exercise-guidelines-cancer-infographic.pdf [accessed 2026-01-27]
26. Kim HL, Lee EM, Ahn SY, Kim KI, Kim HC, Kim JH, et al. The 2022 focused update of the 2018 Korean hypertension society guidelines for the management of hypertension. Clin Hypertens 2023;29(1):11 [FREE Full text] [doi: 10.1186/s40885-023-00234-9] [Medline: 36788612]
27. Hur KY, Moon MK, Park JS, Kim S, Lee S, Yun J, Committee of Clinical Practice Guidelines, Korean Diabetes Association. 2021 clinical practice guidelines for diabetes mellitus of the Korean diabetes association. Diabetes Metab J 2021;45(4):461-481 [FREE Full text] [doi: 10.4093/dmj.2021.0156] [Medline: 34352984]
28. Vickery C, Blazeby J, Conroy T, Arraras J, Sezer O, Koller M, EORTC Quality of Life Group. Development of an EORTC disease-specific quality of life module for use in patients with gastric cancer. Eur J Cancer 2001;37(8):966-971. [doi: 10.1016/s0959-8049(00)00417-2] [Medline: 11334720]
29. Torbahn G, Strauß T, Sieber C, Volkert D, Kiesswetter E. Use of mini nutritional assessment (MNA) in oncological patients – an evidence map. Clin Nutr 2018;37:S122. [doi: 10.1016/j.clnu.2018.06.1461]
30. Torbahn G, Strauss T, Sieber CC, Kiesswetter E, Volkert D. Nutritional status according to the mini nutritional assessment (MNA) as potential prognostic factor for health and treatment outcomes in patients with cancer - a systematic review. BMC Cancer 2020;20(1):594 [FREE Full text] [doi: 10.1186/s12885-020-07052-4] [Medline: 32586289]
31. Cereda E. Mini nutritional assessment. Curr Opin Clin Nutr Metab Care 2012;15(1):29-41. [doi: 10.1097/MCO.0b013e32834d7647] [Medline: 22037014]
32. Hartrick CT, Kovan JP, Shapiro S. The numeric rating scale for clinical pain measurement: a ratio measure? Pain Pract 2003;3(4):310-316. [doi: 10.1111/j.1530-7085.2003.03034.x] [Medline: 17166126]
33. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
34. Oh JY, Yang YJ, Kim BS, Kang JH. Validity and reliability of Korean version of International Physical Activity Questionnaire (IPAQ) short form. J Korean Acad Fam Med 2007;28:532-541 [FREE Full text]
35. Lee PH, Macfarlane DJ, Lam TH, Stewart SM. Validity of the International Physical Activity Questionnaire short form (IPAQ-SF): a systematic review. Int J Behav Nutr Phys Act 2011;8:115 [FREE Full text] [doi: 10.1186/1479-5868-8-115] [Medline: 22018588]
36. Mathiowetz V, Weber K, Volland G, Kashman N. Reliability and validity of grip and pinch strength evaluations. J Hand Surg Am 1984;9(2):222-226. [doi: 10.1016/s0363-5023(84)80146-x] [Medline: 6715829]
37. Rikli RE, Jones CJ. Development and validation of criterion-referenced clinically relevant fitness standards for maintaining physical independence in later years. Gerontologist 2013;53(2):255-267. [doi: 10.1093/geront/gns071] [Medline: 22613940]
38. Millor N, Lecumberri P, Gómez M, Martínez-Ramírez A, Izquierdo M. An evaluation of the 30-s chair stand test in older adults: frailty detection based on kinematic parameters from a single inertial unit. J Neuroeng Rehabil 2013;10:86 [FREE Full text] [doi: 10.1186/1743-0003-10-86] [Medline: 24059755]
39. Bohannon RW, Wang YC, Gershon RC. Two-minute walk test performance by adults 18 to 85 years: normative values, reliability, and responsiveness. Arch Phys Med Rehabil 2015;96(3):472-477. [doi: 10.1016/j.apmr.2014.10.006] [Medline: 25450135]

XSL•FO
RenderX

40. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. J Med Internet Res 2006;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
41. Hawkes AL, Chambers SK, Pakenham KI, Patrao TA, Baade PD, Lynch BM, et al. Effects of a telephone-delivered multiple health behavior change intervention (CanChange) on health and behavioral outcomes in survivors of colorectal cancer: a randomized controlled trial. J Clin Oncol 2013;31(18):2313-2321. [doi: 10.1200/JCO.2012.45.5873] [Medline: 23690410]
42. MEDICALLOGIC. URL: http://www.medicallogic.com/ [accessed 2026-01-30]
43. Roberts AL, Fisher A, Smith L, Heinrich M, Potts HWW. Digital health behaviour change interventions targeting physical activity and diet in cancer survivors: a systematic review and meta-analysis. J Cancer Surviv 2017;11(6):704-719 [FREE Full text] [doi: 10.1007/s11764-017-0632-1] [Medline: 28779220]
44. Kim SH. Nutritional strategies for patients with upper gastrointestinal cancers. Korean J Helicobacter Up Gastrointest Res 2023;23(4):262-267 [FREE Full text] [doi: 10.7704/kjhugr.2023.0056] [Medline: 40503498]
45. Lin YC, Hagen R, Powers BD, Dineen SP, Milano J, Hume E, et al. Digital health intervention to reduce malnutrition among individuals with gastrointestinal cancer receiving cytoreductive surgery combined with hyperthermic intraperitoneal chemotherapy: feasibility, acceptability, and usability trial. JMIR Cancer 2025;11:e67108 [FREE Full text] [doi: 10.2196/67108] [Medline: 40194318]
46. Muscaritoli M, Arends J, Bachmann P, Baracos V, Barthelemy N, Bertz H, et al. ESPEN practical guideline: clinical nutrition in cancer. Clin Nutr 2021;40(5):2898-2913. [doi: 10.1016/j.clnu.2021.02.005] [Medline: 33946039]
47. Weimann A, Braga M, Carli F, Higashiguchi T, Hübner M, Klek S, et al. ESPEN practical guideline: clinical nutrition in surgery. Clin Nutr 2021;40(7):4745-4761. [doi: 10.1016/j.clnu.2021.03.031] [Medline: 34242915]
48. Alidosti M, Sharifirad GR, Golshiri P, Azadbakht L, Hasanzadeh A, Hemati Z. An investigation on the effect of gastric cancer education based on health belief model on knowledge, attitude and nutritional practice of housewives. Iran J Nurs Midwifery Res 2012;17(4):256-262 [FREE Full text] [Medline: 23833623]
49. Cong M, Wang J, Fang Y, Liu Y, Sun M, Wu Q, et al. A multi-center survey on dietary knowledge and behavior among inpatients in oncology department. Support Care Cancer 2018;26(7):2285-2292. [doi: 10.1007/s00520-018-4057-9] [Medline: 29404841]
50. Grace EM, Shaw C, Lalji A, Mohammed K, Andreyev HJN, Whelan K. Nutritional status, the development and persistence of malnutrition and dietary intake in oesophago-gastric cancer: a longitudinal cohort study. J Hum Nutr Diet 2018;31(6):785-792. [doi: 10.1111/jhn.12588] [Medline: 30033545]
51. Huggins CE, Hanna L, Furness K, Silvers MA, Savva J, Frawley H, et al. Effect of early and intensive telephone or electronic nutrition counselling delivered to people with upper gastrointestinal cancer on quality of life: a three-arm randomised controlled trial. Nutrients 2022;14(15):3234 [FREE Full text] [doi: 10.3390/nu14153234] [Medline: 35956410]
52. Boram S, Inhye JJ, Gayoung C, Seonjeong C. A study on perceptions of digital therapeutics. Health Insurance Review and Assessment Service. 2024. URL: https://repository.hira.or.kr/bitstream/2019.oak/3208/2/%EB%94%94%EC%A7%80%ED%84%B8%20%EC%B9%98%EB%A3%8C%EA%B8%B0%EA%B8%B0%EC%97%90%20%EB%8C%80%ED%95%9C%20%EC%9D%B8%EC%8B%9D%EC%A1%B0%EC%82%AC%20%EC%97%B0%EA%B5%AC.pdf [accessed 2026-02-01]
53. Osaki T, Matsunaga T, Makinoya M, Shimizu S, Shishido Y, Miyatani K, et al. Change and predictors of body composition after gastrectomy for gastric cancer during first postoperative year. J Gastrointest Surg 2025;29(2):101931. [doi: 10.1016/j.gassur.2024.101931] [Medline: 39674263]
54. Ding D, Feng Y, Song B, Gao S, Zhao J. Effects of preoperative and postoperative enteral nutrition on postoperative nutritional status and immune function of gastric cancer patients. Turk J Gastroenterol 2015;26(2):181-185 [FREE Full text] [doi: 10.5152/tjg.2015.3993] [Medline: 25835119]
55. Behrouzian R, Aghdami N. Urinary iodine/creatinine ratio in patients with stomach cancer in Urmia, Islamic Republic of Iran. East Mediterr Health J 2004;10(6):921-924. [doi: 10.26719/2004.10.6.921]
56. Moghimi-Dehkordi B, Safaee A, Zali M. Comparison of colorectal and gastric cancer: survival and prognostic factors. Saudi J Gastroenterol 2009;15(1):18-23 [FREE Full text] [doi: 10.4103/1319-3767.43284] [Medline: 19568550]
57. Tian J, Xue W, Yin H, Zhang N, Zhou J, Long Z, et al. Differential metabolic alterations and biomarkers between gastric cancer and colorectal cancer: a systematic review and meta-analysis. Onco Targets Ther 2020;13:6093-6108 [FREE Full text] [doi: 10.2147/OTT.S247393] [Medline: 32612370]
58. Zhang S, Wu Z, Huo B, Xu H, Zhao K, Jing C, et al. [Incidence of postoperative complications in Chinese patients with gastric or colorectal cancer based on a national, multicenter, prospective, cohort study]. Zhonghua Wei Chang Wai Ke Za Zhi 2024;27(3):247-260. [doi: 10.3760/cma.j.cn441530-20240218-00067] [Medline: 38532587]
59. Galiano-Castillo N, Cantarero-Villanueva I, Fernández-Lao C, Ariza-García A, Díaz-Rodríguez L, Del-Moral-Ávila R, et al. Telehealth system: a randomized controlled trial evaluating the impact of an internet-based exercise intervention on quality of life, pain, muscle strength, and fatigue in breast cancer survivors. Cancer 2016;122(20):3166-3174 [FREE Full text] [doi: 10.1002/cncr.30172] [Medline: 27332968]
60. Yu PY, Liu F, Jiao Y, Zhao YC, Liu YH. Depression in gastric cancer patients: integrated therapeutic strategies and clinical implications. World J Clin Oncol 2025;16(6):106229 [FREE Full text] [doi: 10.5306/wjco.v16.i6.106229] [Medline: 40585841]

XSL•FO
RenderX

61. Küchler T, Bestmann B, Rappat S, Henne-Bruns D, Wood-Dauphinee S. Impact of psychotherapeutic support for patients with gastrointestinal cancer undergoing surgery: 10-year survival results of a randomized trial. J Clin Oncol 2007;25(19):2702-2708. [doi: 10.1200/JCO.2006.08.2883] [Medline: 17602075]
62. Gu Y, Ma X, Xin H, Xiang Z, Chen Y, He C. Navigating life after gastric cancer surgery: a qualitative exploration of the dyadic patient-caregiver perspective on quality of life outcomes. BMC Cancer 2025;25(1):288 [FREE Full text] [doi: 10.1186/s12885-025-13696-x] [Medline: 39966806]
63. Bouras G, Markar SR, Burns EM, Huddy J, Bottle A, Athanasiou T, et al. The psychological impact of symptoms related to esophagogastric cancer resection presenting in primary care: a national linked database study. Eur J Surg Oncol 2017;43(2):454-460 [FREE Full text] [doi: 10.1016/j.ejso.2016.10.010] [Medline: 27919514]
64. Beleigoli AM, Andrade AQ, Cançado AG, Paulo MN, Diniz MDFH, Ribeiro AL. Web-based digital health interventions for weight loss and lifestyle habit changes in overweight and obese adults: systematic review and meta-analysis. J Med Internet Res 2019;21(1):e298 [FREE Full text] [doi: 10.2196/jmir.9609] [Medline: 30622090]
65. Galvão DA, Newton RU, Girgis A, Lepore SJ, Stiller A, Mihalopoulos C, et al. Randomized controlled trial of a peer led multimodal intervention for men with prostate cancer to increase exercise participation. Psychooncology 2018;27(1):199-207. [doi: 10.1002/pon.4495] [Medline: 28685892]
66. Pujia C, Ferro Y, Mazza E, Maurotti S, Montalcini T, Pujia A. The role of mobile apps in obesity management: systematic review and meta-analysis. J Med Internet Res 2025;27:e66887 [FREE Full text] [doi: 10.2196/66887] [Medline: 40327853]
67. Chung BY, Oh EH, Song SJ. Mobile health for breast cancer patients: s systematic review. Asian Oncol Nurs 2017;17(3):133. [doi: 10.5388/aon.2017.17.3.133]
68. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
69. Park YH, Lee JI, Lee JY, Cheong IY, Hwang JH, Seo SI, et al. Internet of things-based lifestyle intervention for prostate cancer patients on androgen deprivation therapy: a prospective, multicenter, randomized trial. Am J Cancer Res 2021;11(11):5496-5507 [FREE Full text] [Medline: 34873475]
70. Benzo RM, Moreno PI, Noriega-Esquives B, Otto AK, Penedo FJ. Who benefits from an eHealth-based stress management intervention in advanced prostate cancer? Results from a randomized controlled trial. Psychooncology 2022;31(12):2063-2073 [FREE Full text] [doi: 10.1002/pon.6000] [Medline: 35851976]
71. Tian Q, Qin L, Zhu W, Xiong S, Wu B. Analysis of factors contributing to postoperative body weight change in patients with gastric cancer: based on generalized estimation equation. PeerJ 2020;8:e9390 [FREE Full text] [doi: 10.7717/peerj.9390] [Medline: 32728490]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
eHEALS: eHealth Literacy Scale
EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30
EORTC QLQ-STO22: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Stomach 22
GC: gastric cancer
GEE: generalized estimating equation
IPAQ-SF: International Physical Activity Questionnaire-Short Form
MET: metabolic equivalent of task
mHealth: mobile health
MNA: Mini Nutritional Assessment
NRS: Numeric Rating Scale
SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials
THR: target heart rate

XSL•FO
RenderX

Edited by G Badicu; submitted 13.Oct.2025; peer-reviewed by KW Ryu, S Pan, SM Yeo; comments to author 22.Nov.2025; revised version received 05.Jan.2026; accepted 15.Jan.2026; published 26.Feb.2026.

Please cite as:
Kim I, Lim JY, Kim SW, Lee JH, Sohn TS, Park S, Kang SH, Lee JY, Hwang JH
Effectiveness of a Personalized Digital Exercise and Nutrition Rehabilitation Program in Postoperative Patients With Gastric Cancer: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e85795
URL: https://mhealth.jmir.org/2026/1/e85795
doi:10.2196/85795
PMID:

©Inah Kim, Ji Young Lim, Sun Woo Kim, Jun Ho Lee, Tae Sung Sohn, Sungsoo Park, Seok Ho Kang, Ji Youl Lee, Ji Hye Hwang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 26.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

mHealth-Supported Exercise Rehabilitation to Reverse Frailty After Autologous Transplantation in Multiple Myeloma: Randomized Controlled Trial

Kyuwan Lee[1], PhD; Justin Shamunee[2], MS; Haehyun Lee[3], MS; Xinyi Du[4], MPH; Lanie Lindenfeld[4], MS; Amrita Krishnan[4], MD; Nitya Nathwani[4], MD; F Lennie Wong[4], PhD; Saro Armenian[4], DO, MPH

[1]Kinesiology & Sports Studies, Ewha Womans University, Seoul, Seoul, Republic of Korea

[2]Smidt Heart Institute, Cedars-Sinai Medical Center, Los Angeles, CA, United States

[3]University of Nevada, Las Vegas, Las Vegas, NV, United States

[4]Population Sciences, City of Hope, Duarte, CA, United States

Corresponding Author:
Kyuwan Lee, PhD
Kinesiology & Sports Studies
Ewha Womans University
52 Ewhayeodae-gil, Seodaemun-gu
Seoul, Seoul, 03760
Republic of Korea
Phone: 82 02 3277 2575
Email: kyuwanlee@ewha.ac.kr

Abstract

Background: Frailty is highly prevalent in survivors of multiple myeloma (MM) after autologous hematopoietic cell transplantation and is associated with poor functional recovery and adverse clinical outcomes. Although exercise is known to improve physical function, traditional center-based rehabilitation models are often inaccessible to this population during early posttransplant recovery. Mobile health (mHealth)–supported exercise may offer a scalable alternative; however, evidence in hematologic malignancies remains limited.

Objective: This study aimed to evaluate the effects of a 16-week mHealth-supported exercise rehabilitation program on frailty phenotype and physical function in survivors of MM within 180 days after autologous hematopoietic cell transplantation.

Methods: In this single-center randomized controlled trial, participants who self-reported as prefrail or frail were randomized 1:1 to an mHealth-supported exercise group (n=16) or usual care control (n=16). Remote assessments were conducted at baseline (week 0), midpoint (week 9), and follow-up (week 17). The intervention consisted of 8 weeks of supervised tele-exercise (3 sessions/week, 50 minutes/session), followed by 8 weeks of independent home-based exercise using the same mHealth platform. Exercise intensity was prescribed using a repetitions-in-reserve–based rating of perceived exertion approach with symptom-guided progression. The primary outcome was change in the 5-component Fried frailty phenotype score (0-5). Secondary outcomes included Short Physical Performance Battery components, chair stand time, gait speed, and handgrip strength. Intention-to-treat analyses were conducted using generalized estimating equations to evaluate between-group differences over time.

Results: Participants had a mean age of 64.6 (SD 7.1) years and were enrolled a mean of 136 (SD 36.3) days posttransplant. At baseline, 94% (30/32) of participants were classified as frail. Adherence to the supervised sessions was 85% (326/384 sessions), and adherence during the unsupervised phase was 78% (298/384 sessions). The exercise group demonstrated a significantly greater reduction in frailty score compared with control from baseline to week 17 (P<.001). Between-group difference estimates showed a clinically meaningful improvement favoring exercise at both week 9 and week 17 (P<.001). Chair stand time improved significantly in the exercise group compared with control, with faster completion times observed at week 9 and sustained through week 17 (P=.002). Improvements in other Short Physical Performance Battery components and handgrip strength favored the exercise group but did not reach statistical significance. No serious adverse events occurred.

Conclusions: A 16-week mHealth-supported, progressively prescribed exercise rehabilitation program was feasible, safe, and effective in reversing frailty phenotype and improving functional mobility in survivors of MM early after autologous transplantation. This approach provides a scalable model for delivering structured rehabilitation during a high-risk recovery window. Larger trials incorporating attention-matched controls and longer follow-up are warranted.

XSL•FO
RenderX

Trial Registration: ClinicalTrials.gov NCT05142371; https://clinicaltrials.gov/study/NCT05142371

(JMIR Mhealth Uhealth 2026;14:e87628) doi:10.2196/87628

KEYWORDS

mobile health exercise; mHealth exercise; multiple myeloma; frailty; autologous hematopoietic cell transplantation

Introduction

Background and Rationale

Autologous hematopoietic cell transplantation (HCT) improves progression-free survival in patients with multiple myeloma (MM) [1]. Advances in treatment approaches, coupled with improved access to HCT, have contributed to a growing number of patients with MM undergoing HCT annually. In fact, MM is the most common indication for HCT, with >7000 patients undergoing autologous HCT annually in the United States alone [2]. However, despite improvements in disease-related outcomes, patients with MM who undergo HCT remain at high risk of experiencing frailty after HCT, attributed to pre-HCT treatment exposures, HCT-related conditioning therapies, and post-HCT acute toxicities [3,4].

The frailty phenotype is characterized by the presence of 3 or more of the following characteristics: underweight, exhaustion, low energy expenditure, slow gait speed, and weak muscle strength [3,5,6]. Two-thirds of patients with MM are estimated to be frail at diagnosis or shortly thereafter, and the prevalence of frailty increases after HCT [7]. In an International Myeloma Working Group report, frailty was associated with significantly worse overall survival and a higher rate of treatment discontinuation (31.2% vs 16.5%) [8]. This underscores the need to develop innovative strategies to mitigate frailty in patients with MM and its associated morbidity and mortality following HCT.

Supervised exercise-based interventions have been shown to improve physical function and frailty in patients with chronic health conditions, including cardiovascular diseases and cancer [9,10]. Traditionally, supervised exercise interventions have been delivered in person at specialized centers, an approach that poses logistical challenges and limits access for many patients. Supervised prescriptive mobile health (mHealth) exercise interventions have emerged as alternatives to in-person approaches, offering the convenience of anytime and anywhere access to patients, regardless of geographic boundaries [11]. Studies in patients with other cancer types, such as the breast and endometrium, have shown that an mHealth exercise is effective in improving physical function, BMI, and health-related quality of life [12,13]. There is a paucity of information on the safety and efficacy of mHealth exercise intervention in prefrail and frail patients with MM.

Objective and Hypothesis

To address this limitation, we conducted a randomized controlled trial evaluating the effectiveness of an mHealth exercise intervention aimed at improving frailty and physical function in patients with MM shortly after undergoing HCT. We hypothesized that a 16-week mHealth exercise intervention would significantly improve frailty and physical function in these patients.

Methods

Study Design and Setting

This was a single-center randomized control trial conducted at City of Hope (COH). The trial was registered at ClinicalTrials.gov (NCT05142371) on October 29, 2021. We conducted a 16-week mHealth intervention that included baseline (week 0) and postintervention (week 9 and week 17) assessments of physical function and frailty phenotype. All study procedures, including screening, enrollment, intervention, and data collection, were performed using an mHealth-based approach, eliminating the need for participants to be assessed at a center. This study adheres to the CONSORT (Consolidated Standards of Reporting Trials) guidelines for reporting randomized controlled trials.

Participants and Eligibility Criteria

The participant eligibility criteria are shown in Textbox 1.

XSL•FO
RenderX

Textbox 1. Participant eligibility criteria.

Inclusion criteria

- Diagnosis of multiple myeloma
- Aged >18 years at enrollment
- Prefrail or frail status (ie, Fried criteria: clinically underweight [BMI <18.5 kg/m^2], exhaustion, low energy expenditure, slow walking speed, and muscle weakness), with the presence of 2 of 5 indexes classified as prefrail and ≥3 of 5 indexes classified as frail [14]
- <180 days after hematopoietic cell transplantation
- Physically able and willing to complete all study procedures
- English speaking

Exclusion criteria

- Had clinically significant or active cardiovascular disease (eg, unstable angina and uncontrolled arrhythmia)
- Had contraindications to exercise (acute infectious disease or unstable bone lesions)
- Were recovering from a recent injury or were physically injured in the past 6 months, whereby participation in rigorous exercise may not be appropriate
- Were already participating in regular, structured exercise (>60 minutes/week)

Screening, Recruitment, and Randomization

Prescreening of potentially eligible patients was performed using an established institutional review board–approved COH protocol for long-term follow-up after HCT [15]. During telephone prescreening, participants were asked simple mobility and fatigue screening questions, including whether they were able to ambulate independently without assistance and whether they experienced persistent fatigue or low physical activity. Frailty status was formally confirmed during the baseline remote assessment prior to randomization. Individuals self-reporting characteristics consistent with prefrailty or frailty were invited for formal remote assessment. Once potentially eligible patients were identified, we notified their primary hematologist to obtain clearance prior to our outreach. Potentially eligible patients were then contacted by phone to assess their willingness to participate and to screen for patient-reported symptoms related to prefrailty or frailty. If potentially eligible patients expressed interest in the study, a trained clinical research assistant conducted detailed discussions with the patients over the phone regarding the additional study inclusion and exclusion criteria. Once patients were found to be eligible for participation, they provided informed consent using a web-based consent form (DocuSign Inc). Participants were randomly assigned in a 1:1 ratio to either the mHealth exercise group or the control group using a computer-generated random allocation sequence. The sequence was prepared by an independent statistician not involved in study conduct. Allocation concealment was maintained, as participants were enrolled by the study coordinator, while group assignments were implemented by a separate investigator who was unaware of the allocation sequence.

Intervention: mHealth-Supported Exercise Intervention

Intensity Prescription and Progression

Exercise intensity during resistance training was prescribed using a repetitions-in-reserve–based rating of perceived exertion (RPE) scale [16]. Participants were instructed to perform exercises at an intensity corresponding to RPE 7 to 8, indicating that they could complete approximately 2 to 3 additional repetitions before fatigue. If participants reported that they could perform >4 additional repetitions (RPE ≤6), resistance or repetitions were increased. Conversely, if participants could perform fewer than 1 to 2 additional repetitions (RPE ≥9) or reported excessive fatigue, discomfort, or dizziness, the exercise was modified by reducing resistance, repetitions, or range of motion and increasing rest intervals. Progression followed a standardized symptom-guided approach. When participants completed all prescribed sets at RPE ≤7 for 2 consecutive sessions, exercise difficulty was progressed by increasing repetitions, adding sets, or advancing resistance band tension and balance challenge. This allowed individualized progression while maintaining safety in the early posttransplant period. Details of the intervention strategies used have been previously published [17]. Exercise equipment, including resistance bands, resistance loop bands, and exercise mats, was shipped directly to participants from COH. The exercise intervention consisted of 2 phases. The first phase offered each participant 24 supervised exercise sessions, each lasting 50 minutes and held 3 times per week for 8 weeks. All sessions in the first 8 weeks were supervised remotely via videoconferencing. Participants could reschedule or make up sessions if they were unable to attend as planned. Their adherence to the program was tracked on the platform. After completing the initial 8-week intervention, the second phase encouraged participants to continue exercising independently without direct monitoring by the exercise trainer for the subsequent 8 weeks until the follow-up physical assessment (week 17). Participants in the control group were asked to maintain their current lifestyle for 17 weeks. They were

XSL•FO
RenderX

later offered the same mHealth exercise intervention as a courtesy, though it was not mandatory for completing the study procedures.

Sample Exercise Plan

A representative example of a supervised exercise session structure, including balance, strength, and core components with repetitions-in-reserve–based RPE targets and progression rules, is provided in Multimedia Appendix 1 to facilitate reproducibility of the intervention.

Outcomes and Remote Assessments

We used a commercially available mHealth exercise platform (Moterum Technologies) that offers exercise videos, physical activity monitoring, and automatic data collection from wearable devices. We sent gait sensors (for the hip, right foot, and left foot), a hand dynamometer, measuring tape, and blue tape to each participant's home via mail. This equipment was used to assist in measuring physical function and frailty phenotype outcomes before and after the intervention in both groups.

Frailty Phenotype (Primary Outcome): Components and Cutoff Values

Frailty was assessed using the 5-component Fried phenotype (score range 0-5), with 1 point assigned for each component present and participants classified as prefrail (2/5) or frail (≥3/5). The operational definitions followed the original Fried criteria: (1) low BMI defined as BMI <18.5 kg/m²; (2) exhaustion defined as a Functional Assessment of Chronic Illness Therapy-Fatigue score ≤30, indicating clinically meaningful fatigue; (3) low physical activity defined as <600 metabolic equivalent minute/week based on International Physical Activity Questionnaire-Short Form scoring guidelines; (4) slow gait speed defined as a 4-m walk speed <0.8 m/second; and (5) weakness defined as handgrip strength <26 kg for men and <18 kg for women. All measures were obtained remotely using mailed equipment and videoconference supervision. All frailty assessments were conducted under live videoconference supervision by trained research staff to ensure standardized administration.

Physical Function

We assessed physical function remotely using the Short Physical Performance Battery (SPPB) under the supervision (through videoconferencing) of study staff. Remote administration of the SPPB has previously been shown to demonstrate acceptable validity and reliability compared with in-person assessment when standardized instructions and camera positioning are used [18,19]. SPPB included 3 lower extremity measures completed in the following order: timed balance, gait speed, and chair stand. We evaluated timed balance under three conditions: (1) side-by-side stand with both feet placed on the ground for 10 seconds or when the participant steps out of position, (2) semitandem stand with the side of the heel of one foot touching the big toe of the other foot for 10 seconds or when the participant steps out of position, and (3) tandem stand with the heel of one foot touching the big toe of the other foot for 10 seconds or when the participant steps out of position. We evaluated chair stand performance under 2 conditions: participants performed a single chair stand, and then they were asked to perform 5 repeated chair stands as quickly as possible. The time taken to complete each task was recorded.

Sample Size and Power Calculation

The original sample size calculation was based on detecting a between-group difference in frailty phenotype score at week 9, which was considered the primary early end point. Assuming a 2-sided α of 0.05, 80% power, 20% attrition, and a within-person correlation of 0.80 across repeated measures, 60 participants (30 per group) were estimated to be required to detect a standardized effect size (Cohen *d*) of 0.40. This effect size was selected based on prior exercise intervention studies demonstrating moderate improvements in frailty-related functional outcomes in clinical populations [20,21].

Owing to recruitment challenges in this specific posttransplant population, the final sample size was 32 participants (16 per group). Under the same assumptions, this sample size provided 80% power to detect a larger effect size of 0.50. Therefore, this study should be interpreted as an early-phase randomized trial with adequate power to detect moderate-to-large intervention effects but limited precision for smaller effects and generalizability.

Statistical Analysis

We generated descriptive statistics for participants' demographics, level of physical activity, and history of chronic health conditions. Assuming a 2-sided type I error of 0.05, intention-to-treat analyses were performed using all randomized participants. Changes in frailty index from baseline to week 9 and week 17 were evaluated using the Generalized Estimation Equation methods with 2 indicators of time, assuming a compound symmetry within-person correlation matrix. We included treatment by time interaction and tested its significance using a 2-*df* χ^2 test. BMI, age, and race (non-Hispanic White vs other) were included as a priori covariates for adjustment. Additionally, between-group differences in outcome values at week 17 were computed, along with 95% CIs, and compared with differences at week 9.

Ethical Considerations

The study protocol was approved by the COH Institutional Review Board (21406) and was conducted in accordance with institutional and national ethical standards for research involving human participants. All participants provided written informed consent electronically (DocuSign) prior to participation and were able to withdraw at any time without affecting their clinical care. Study data were deidentified and stored on secure institutional servers accessible only to authorized study personnel. Participants received a US $50 gift card as compensation for their time and participation. The trial was registered at ClinicalTrials.gov (NCT05142371) on October 29, 2021.

Results

Participant Flow and Retention

A total of 212 potentially eligible patients with MM were identified in the long-term follow-up program following HCT at COH (Figure 1). Of the 129 successfully contacted patients

XSL•FO
RenderX

with MM, 46 (36%) refused due to lack of interest, confidence in mHealth exercise settings, and time and 45 (35%) were deemed ineligible (Figure 2). Among the initial 33 prefrail or frail patients with MM who were enrolled, 1 individual withdrew from the study prior to randomization. This resulted in 32 patients randomized equally to either the mHealth exercise group (n=16) or the control group (n=16).

Figure 1. Study schema.

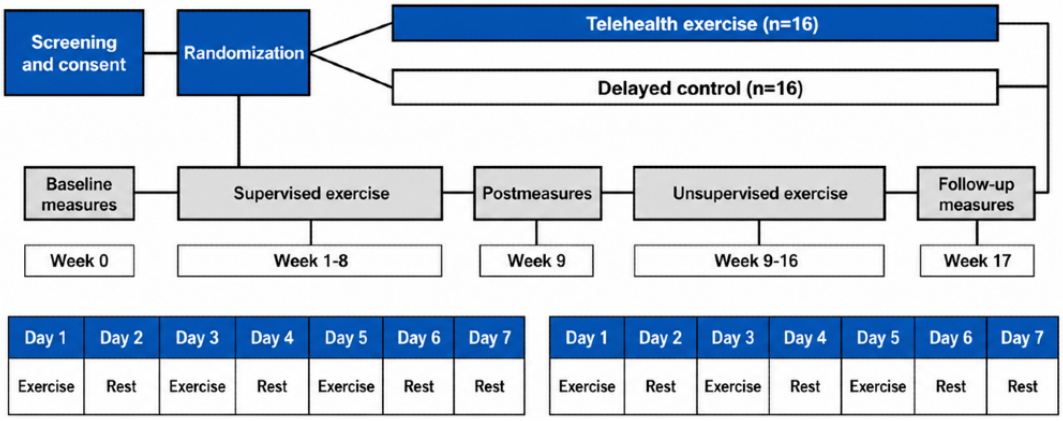
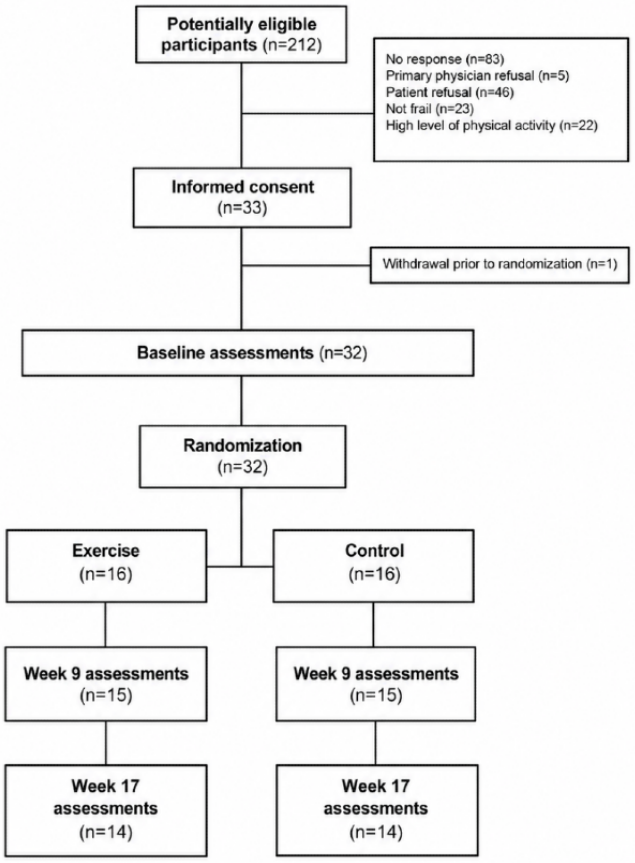


Figure 2. CONSORT (Consolidated Standards of Reporting Trials) diagram of mobile health–based exercise intervention.



Baseline Characteristics

Baseline characteristics are presented in Table 1. All participants had received high-dose melphalan conditioning (standard dose, 200 mg/m^2), with appropriate dose reduction (140 mg/m^2) for those with renal impairment (eg, creatinine clearance <40 mL/min), per standard of care. The mean age at study enrollment was 64.6 (SD 7.1) years. Of the 32 participants, 21 (66%) were female and 18 (56%) were non-Hispanic White. The mean time from HCT to study enrollment was 136 (SD 36.3) days. At baseline, 30 (94%) participants were classified as frail and 2 (6%) as prefrail according to the Fried phenotype criteria. The most common frailty characteristics were exhaustion (n=32, 100%) and low physical activity (n=30, 94%), followed by weakness (n=23, 72%), slow gait speed (n=19, 59%), and low BMI (n=1, 3%).

XSL•FO
RenderX

Table 1. Baseline participant characteristics (N=32).

	Total (N=32)	Exercise (n=16)	Control (n=16)
Age (years), mean (SD)	64.6 (7.1)	64.0 (6.9)	65.1 (7.5)
Sex, n (%)			
Male	11 (34)	4 (25)	7 (44)
Female	21 (66)	12 (75)	9 (56)
BMI (kg/m^2), mean (SD)	26.1 (5.8)	27.3 (7.3)	24.9 (3.7)
Race or ethnicity, n (%)			
Non-Hispanic White	18 (56)	9 (56)	9 (56)
Hispanic White	6 (19)	2 (13)	4 (25)
African American	5 (16)	3 (19)	2 (13)
Asian or Pacific Islander	3 (9)	2 (13)	1 (6)
Comorbidities, n (%)			
Hypertension	10 (31)	4 (25)	6 (38)
Diabetes	3 (9)	2 (13)	1 (6)
Dyslipidemia	8 (25)	5 (31)	3 (19)
Congestive heart failure	5 (16)	2 (13)	3 (19)
Frailty, n (%)			
Exhaustion	32 (100)	16 (100)	16 (100)
Low energy expenditure	30 (94)	15 (94)	15 (94)
Weakness	23 (72)	12 (75)	11 (69)
Slow gait speed	19 (59)	9 (56)	10 (63)
Low BMI	1 (3)	1 (6)	0 (0)
Physical activity (MET[a]-min/week), mean (SD)	1441. 8 (812.2)	1490.9 (819.6)	1392.7 (828.4)

[a]MET: metabolic equivalent.

Intervention Adherence and Safety

All participants (32/32, 100%) successfully completed remote physical function assessments prior to randomization and initiation of the intervention. At week 9, 2 (6%) participants were lost to follow-up, and an additional 2 (7%) were lost at week 17 (1 from each group at both time points). During the initial 8-week supervised phase, adherence to the 24 prescribed exercise sessions among participants in the intervention group was 85% (326/384 sessions). During the subsequent 8-week unsupervised phase, participants remained engaged with the mHealth platform and completed 78% (298/384 sessions) of the prescribed exercise sessions. No serious adverse events or unintended effects were reported during the intervention period.

Primary Outcome: Frailty Phenotype

Overall, the mHealth exercise intervention resulted in a significant adjusted between-group reduction in frailty phenotype at week 9 (−0.92 points; 95% CI −1.38 to −0.46 ; P<.001), which was further sustained at week 17 (−1.14 points; 95% CI −1.66 to −0.62; Table 2). Descriptively, 63% (10/16) participants in the intervention group no longer met criteria for slow gait speed or low physical activity at both week 9 and week 17, whereas 88% (14/16) participants in the control group continued to meet these frailty components across the study period.

XSL•FO
RenderX

Table 2. Between-group differences in frailty and physical function outcomes at week 9 and week 17, estimated using generalized estimating equation models.

Outcomes	Baseline, mean (SD)	Week 9			Week 17		
		Mean (SD)	Between-group difference (95% CI)	*P* value	Mean (SD)	Between-group difference (95% CI)	*P* value
5-scale frailty phenotype			−0.97 (−1.32 to −0.63)	<.001[a]		−1.19 (−1.72 to −0.66)	<.001[b]
Exercise	3.94 (1.1)	3.07 (0.9)			2.79 (0.7)		
Control	3.88 (1.1)	4.00 (0.9)			3.93 (0.9)		
Chair stand performance			−1.88 (−3.58 to −0.18)	.03[a]		−2.98 (−4.84 to −1.13)	.002[b]
Exercise	16.8 (8.7)	13.6 (7.3)			12.0 (4.6)		
Control	12.4 (3.1)	12.8 (3.7)			12.7 (3.6)		
Gait speed (m/s)			0.02 (−0.09 to 0.13)	.69		0.01 (−0.19 to 0.22)	.90
Exercise	1.08 (0.5)	1.12 (0.5)			1.13 (0.3)		
Control	1.01 (0.4)	1.07 (0.3)			1.12 (0.3)		
Handgrip strength (dominant)			0.71 (−1.54 to 2.97)	.54		2.43 (−0.57 to 5.44)	.11
Exercise	28.1 (13.2)	29.6 (12.1)			32.6 (12.3)		
Control	32.7 (13.1)	31.5 (12.8)			31.6 (11.1)		
Handgrip strength (nondominant)			1.78 (−0.77 to 4.35)	.17		1.19 (−2.06 to 4.46)	.47
Exercise	24.8 (9.2)	27.5 (11.2)			28.4 (10.6)		
Control	28.9 (13.1)	28.7 (13.4)			29.3 (12.2)		
Short Physical Performance Battery			0.35 (−0.20 to 0.92)	.21		0.64 (−0.14 to 1.42)	.11
Exercise	10.3 (1.9)	10.6 (1.6)			10.8 (1.4)		
Control	10.4 (1.3)	10.5 (1.5)			10.3 (1.5)		
BMI			−0.27 (−1.06 to 0.51)	.25		−0.45 (−1.38 to 0.48)	.35
Exercise	27.3 (7.3)	27.7 (7.3)			28.2 (7.1)		
Control	24.9 (3.7)	25.3 (3.5)			26.2 (3.2)		

[a]Between-group differences represent adjusted change estimates (exercise vs control) derived from generalized estimating equation models with 95% CIs based on intention-to-treat analysis.

[b]*P* values correspond to the treatment by time interaction terms from the same models.

Secondary Outcomes: Physical Function

The mHealth exercise intervention resulted in a significant adjusted between-group improvement in chair stand performance at week 9 (−2.78 seconds; 95% CI −4.66 to −0.90) that was sustained at week 17 (−3.50 seconds; 95% CI −5.58 to −1.42; Table 2). The adjusted between-group difference in handgrip strength did not reach statistical significance. However, descriptively, dominant arm strength increased within the intervention group, with 50% (8/16) participants demonstrating an increase of >1 kg by week 9 that was maintained through week 17, whereas no consistent change was observed in the control group.

Discussion

Principal Findings

The findings of this study demonstrated the efficacy of an mHealth exercise intervention in improving frailty phenotype and physical function among patients with MM who had recently undergone autologous HCT. This study is among the first to evaluate a fully remote, supervised mHealth exercise intervention specifically targeting frailty reversal in patients with MM following transplantation. This indicates the potential for mHealth interventions to provide sustainable support and benefits for individuals with MM beyond their major cancer treatments. The observed improvements in chair stand performance highlight the potential of mHealth exercise to address key aspects of physical function in patients with MM. These outcomes are clinically meaningful, as each unit increase in chair stand score has been associated with a 21% improvement in survival among older adults [22].

Comparison With Prior MM Trials

Randomized evidence supporting exercise and behavioral interventions in MM is emerging; however, prior trials have largely focused on behavioral coaching models rather than the delivery of structured, progressive exercise targeting functional recovery and frailty. The REAL-FITNESS trial randomized newly diagnosed patients to a World Health

XSL•FO
RenderX

Organization–compliant physical activity prescription during induction therapy and demonstrated improvements in fatigue, depression, grip strength, comorbidity burden, and quality of life [23]. This study importantly established that promoting guideline-based physical activity is feasible and beneficial during treatment. However, the intervention primarily emphasized adherence to aerobic and resistance activity targets rather than delivering supervised, progressive exercise with quantified intensity, and the population did not include patients recovering from the physiological stress of transplantation where frailty risk is substantially heightened. Similarly, Larsen et al [24] conducted a randomized trial of a 10-week individualized exercise program in newly diagnosed patients with MM with careful adjustment for lytic bone disease. The trial demonstrated that exercise is safe in this population and may improve functional measures such as sit-to-stand performance and aerobic capacity. However, no clear between-group superiority was observed, and the study population consisted of patients earlier in their disease trajectory rather than individuals in the high-risk posttransplant recovery phase. Moreover, neither study used frailty as a primary outcome. In contrast, Banerjee et al [25] evaluated a digital life coaching intervention during stem cell transplantation and demonstrated that sustained patient engagement and patient-reported outcomes can be maintained through a fully remote support model even during the intensive peritransplant period. While this work demonstrated the potential of digital health platforms in MM care, the intervention focused on behavioral coaching and patient-reported outcomes rather than structured exercise delivery or objective functional end points.

This trial extends this literature in several clinically meaningful ways. First, it specifically targets the early recovery window following autologous transplantation and deliberately enrolls individuals who are prefrail or frail, a subgroup central to posttransplant morbidity but underrepresented in prior MM exercise trials. Second, the intervention delivers supervised, progressive exercise with explicit intensity and progression during the early phase, followed by a sustainability phase within the same mHealth ecosystem, rather than relying primarily on activity encouragement or coaching. Third, the study pairs this intervention with remote objective functional assessments and a frailty phenotype end point, directly evaluating frailty reversal rather than changes in activity volume or quality of life metrics alone. Collectively, these features position mHealth-supported exercise rehabilitation as a scalable strategy for functional recovery and frailty reversal in survivors of MM after transplantation, complementing and extending prior MM trials conducted in earlier disease phases and coaching-focused digital models.

Interpretation and Potential Mechanisms

The observed adherence to the prescribed exercise sessions during both the supervised (326/384, 85%) and unsupervised phases of the intervention (298/384, 78%) is noteworthy. Despite the transition to unsupervised exercise sessions during the second 8-week phase, participants remained physically active and responsive to the prescribed exercise regimen. This indicates the feasibility and sustainability of mHealth exercise interventions in facilitating engagement in continuous exercise training among patients with MM, even in the absence of direct supervision. Developing a tailored exercise intervention within similar disease and treatment profiles may allow for health care professionals to address specific patient's needs or challenges and enhance the effectiveness of rehabilitation strategies. One notable aspect of the findings from this study is the absence of significant improvement in gait speed following the intervention, despite the observed improvements in other measures of chair stand performance and frailty phenotype. Gait speed is widely used to assess functional capacity, and its assessment provides valuable insights into an individual's ability to perform physical activities [26]. However, the discrepancy between self-reported improvements in gait and the lack of significant changes in objective measures of gait speed in our study may present an intriguing aspect of the study findings. While participants may have subjectively perceived improvements in their gait, as reflected in self-reported frailty phenotype assessments, these perceptions may not have translated into measurable changes in gait speed during objective assessments [27]. It is possible that participants experienced improvements in other aspects of gait performance, such as perceptions of effort or changes in overall mobility, that were not captured by the gait speed itself. Furthermore, variability in testing conditions, such as the testing environment and surface conditions, may have influenced gait speed measurements and introduced variability in the results.

Limitations

While this study represents an important step forward in exploring the potential of mHealth exercise interventions to improve frailty phenotype and physical function in patients with MM, it is important to acknowledge certain limitations. First, the sample size was modest and drawn from a single center, which limits statistical precision and generalizability. Second, the control group did not receive an attention-matched intervention; therefore, nonspecific effects related to increased contact, accountability, or health-focused reminders cannot be excluded as contributors to the observed improvements. Third, eligibility required English proficiency and familiarity with telehealth technology, which may limit applicability to non-English speakers and individuals with lower digital literacy. Finally, although remote objective assessments were incorporated, some frailty components relied on self-report measures, which may introduce reporting bias.

Conclusions

In conclusion, this study contributes to the growing body of evidence supporting the efficacy of mHealth exercise interventions in improving frailty phenotype and physical function among patients with MM. These findings have important implications for the development of accessible exercise interventions aimed at enhancing the functional capacity of individuals living with MM. Further research is needed to explore the long-term effects and potential mechanisms underlying the observed improvements, as well as to optimize the delivery and implementation of mHealth-based exercise programs in this population.

XSL•FO
RenderX

Acknowledgments

The authors thank the participants and clinical staff at the City of Hope Comprehensive Cancer Center for their contributions to this study. No generative artificial intelligence tools were used in the writing, analysis, or preparation of this manuscript.

Funding

This study was supported by the 2021 Steven Gordon and Briskin Family Innovation Grant. The funding source had no role in the design of the study; data collection, analysis, or interpretation; manuscript preparation; or the decision to submit the manuscript for publication.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to institutional data protection policies but are available from the corresponding author on reasonable request and with appropriate institutional approvals.

Authors' Contributions

Conceptualization: all authors
Methodology: KL, JS
Resources: KL, SA
Supervision: KL
Writing – original draft: all authors
Writing – review and editing: all authors

Conflicts of Interest

AK is a consultant for AstraZeneca, BMS, GSK, Roche, Abbvie, Sanofi, Johnson & Johnson, and Pfizer, and own BMS stock. The other authors have no conflicts of interest to disclose.

Multimedia Appendix 1
Representative structure of a supervised mobile health exercise session, including balance, strength, core stability, and functional mobility components, with repetitions-in-reserve–based rating of perceived exertion targets and progression rules.
[DOCX File , 15 KB - mhealth_v14i1e87628_app1.docx]

Multimedia Appendix 2
CONSORT checklist.
[PDF File (Adobe PDF File), 1446 KB - mhealth_v14i1e87628_app2.pdf]

References

1. Siegel RL, Kratzer TB, Giaquinto AN, Sung H, Jemal A. Cancer statistics, 2025. CA Cancer J Clin 2025;75(1):10-45 [FREE Full text] [doi: 10.3322/caac.21871] [Medline: 39817679]
2. D'Souza A, Fretham C, Lee SJ, Arora M, Brunner J, Chhabra S, et al. Current use of and trends in hematopoietic cell transplantation in the United States. Biol Blood Marrow Transplant 2020 Aug;26(8):e177-e182 [FREE Full text] [doi: 10.1016/j.bbmt.2020.04.013] [Medline: 32438042]
3. Arora M, Sun CL, Ness KK, Teh JB, Wu J, Francisco L, et al. Physiologic frailty in nonelderly hematopoietic cell transplantation patients: results from the Bone Marrow Transplant Survivor Study. JAMA Oncol 2016 Oct 01;2(10):1277-1286 [FREE Full text] [doi: 10.1001/jamaoncol.2016.0855] [Medline: 27254472]
4. Sverrisdóttir IS, Rögnvaldsson S, Thorsteinsdottir S, Gíslason GK, Aspelund T, Turesson I, et al. Comorbidities in multiple myeloma and implications on survival: a population-based study. Eur J Haematol 2021 Jun;106(6):774-782. [doi: 10.1111/ejh.13597] [Medline: 33565126]
5. Patel BG, Luo S, Wildes TM, Sanfilippo KM. Frailty in older adults with multiple myeloma: a study of US veterans. JCO Clin Cancer Inform 2020 Feb;4:117-127. [doi: 10.1200/CCI.19.00094] [Medline: 32083955]
6. Ness KK, Krull KR, Jones KE, Mulrooney DA, Armstrong GT, Green DM, et al. Physiologic frailty as a sign of accelerated aging among adult survivors of childhood cancer: a report from the St Jude Lifetime cohort study. J Clin Oncol 2013 Dec 20;31(36):4496-4503 [FREE Full text] [doi: 10.1200/JCO.2013.52.2268] [Medline: 24248696]
7. Engelhardt M, Domm AS, Dold SM, Ihorst G, Reinhardt H, Zober A, et al. A concise revised Myeloma Comorbidity Index as a valid prognostic instrument in a large cohort of 801 multiple myeloma patients. Haematologica 2017 May;102(5):910-921 [FREE Full text] [doi: 10.3324/haematol.2016.162693] [Medline: 28154088]
8. Palumbo A, Bringhen S, Mateos MV, Larocca A, Facon T, Kumar SK, et al. Geriatric assessment predicts survival and toxicities in elderly myeloma patients: an International Myeloma Working Group report. Blood 2015 Mar 26;125(13):2068-2074 [FREE Full text] [doi: 10.1182/blood-2014-12-615187] [Medline: 25628469]

XSL•FO
RenderX

9. Wada Y, Nishi M, Yoshikawa K, Takasu C, Tokunaga T, Nakao T, et al. Preoperative nutrition and exercise intervention in frailty patients with gastric cancer undergoing gastrectomy. Int J Clin Oncol 2022 Sep;27(9):1421-1427. [doi: 10.1007/s10147-022-02202-z] [Medline: 35771395]
10. Ushijima A, Morita N, Hama T, Yamamoto A, Yoshimachi F, Ikari Y, et al. Effects of cardiac rehabilitation on physical function and exercise capacity in elderly cardiovascular patients with frailty. J Cardiol 2021 Apr;77(4):424-431 [FREE Full text] [doi: 10.1016/j.jjcc.2020.11.012] [Medline: 33288376]
11. Zwisler AD, Norton RJ, Dean SG, Dalal H, Tang LH, Wingham J, et al. Home-based cardiac rehabilitation for people with heart failure: a systematic review and meta-analysis. Int J Cardiol 2016 Oct 15;221:963-969. [doi: 10.1016/j.ijcard.2016.06.207] [Medline: 27441476]
12. Gorzelitz JS, Stoller S, Costanzo E, Gangnon R, Koltyn K, Dietz AT, et al. Improvements in strength and agility measures of functional fitness following a telehealth-delivered home-based exercise intervention in endometrial cancer survivors. Support Care Cancer 2022 Jan;30(1):447-455 [FREE Full text] [doi: 10.1007/s00520-021-06415-2] [Medline: 34304292]
13. Peng Y, Zhang K, Wang L, Peng Y, Liu Z, Liu L, et al. Effect of a telehealth-based exercise intervention on the physical activity of patients with breast cancer: a systematic review and meta-analysis. Asia Pac J Oncol Nurs 2022 Jul 22;9(12):100117 [FREE Full text] [doi: 10.1016/j.apjon.2022.100117] [Medline: 36276881]
14. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001 Mar;56(3):M146-M156. [doi: 10.1093/gerona/56.3.m146] [Medline: 11253156]
15. Armenian SH, Horak D, Scott JM, Mills G, Siyahian A, Berano Teh J, et al. Cardiovascular function in long-term hematopoietic cell transplantation survivors. Biol Blood Marrow Transplant 2017 Apr;23(4):700-705 [FREE Full text] [doi: 10.1016/j.bbmt.2017.01.006] [Medline: 28065839]
16. Cowley N, Nicholson V, Timmins R, Munteanu G, Wood T, García-Ramos A, et al. The effects of percentage-based, rating of perceived exertion, repetitions in reserve, and velocity-based training on performance and fatigue responses. J Strength Cond Res 2025 Apr 01;39(4):e516-e529. [doi: 10.1519/JSC.0000000000005026] [Medline: 39787033]
17. Lee K, Nathwani N, Shamunee J, Lindenfeld L, Wong FL, Krishnan A, et al. Telehealth exercise to improve physical function and frailty in patients with multiple myeloma treated with autologous hematopoietic stem cell transplantation (TIPS): protocol of a randomized controlled trial. Trials 2022 Nov 03;23(1):921 [FREE Full text] [doi: 10.1186/s13063-022-06848-y] [Medline: 36329525]
18. Fyfe JJ, Sousa F, Watson-Mackie K, Jansons P, Scott D, Daly RM. Remote physical function testing in older adults: a mixed methods study exploring test reliability, feasibility, and perceptions of participants and assessors. PLoS One 2025 Sep 19;20(9):e0332691 [FREE Full text] [doi: 10.1371/journal.pone.0332691] [Medline: 40971978]
19. Guidarelli C, Lipps C, Stoyles S, Dieckmann NF, Winters-Stone KM. Remote administration of physical performance tests among persons with and without a cancer history: establishing reliability and agreement with in-person assessment. J Geriatr Oncol 2022 Jun;13(5):691-697 [FREE Full text] [doi: 10.1016/j.jgo.2022.02.002] [Medline: 35177378]
20. van Waart H, Stuiver MM, van Harten WH, Geleijn E, Kieffer JM, Buffart LM, et al. Effect of low-intensity physical activity and moderate- to high-intensity physical exercise during adjuvant chemotherapy on physical fitness, fatigue, and chemotherapy completion rates: results of the PACES randomized clinical trial. J Clin Oncol 2015 Jun 10;33(17):1918-1927. [doi: 10.1200/JCO.2014.59.1081] [Medline: 25918291]
21. Pahor M, Guralnik JM, Ambrosius WT, Blair S, Bonds DE, Church TS, et al. Effect of structured physical activity on prevention of major mobility disability in older adults: the LIFE study randomized clinical trial. JAMA 2014 Jun 18;311(23):2387-2396 [FREE Full text] [doi: 10.1001/jama.2014.5616] [Medline: 24866862]
22. Brito LB, Ricardo DR, Araújo DS, Ramos PS, Myers J, Araújo CG. Ability to sit and rise from the floor as a predictor of all-cause mortality. Eur J Prev Cardiol 2014 Jul;21(7):892-898. [doi: 10.1177/2047487312471759] [Medline: 23242910]
23. Dreyling E, Räder J, Möller MD, Ihorst G, Wenger S, Pahl A, et al. A randomized controlled 'REAL-FITNESS' trial to evaluate physical activity in patients with newly diagnosed multiple myeloma. J Cachexia Sarcopenia Muscle 2025 Apr;16(2):e13793 [FREE Full text] [doi: 10.1002/jcsm.13793] [Medline: 40196920]
24. Larsen RF, Jarden M, Minet LR, Frølund UC, Hermann AP, Breum L, et al. Exercise in newly diagnosed patients with multiple myeloma: a randomized controlled trial of effects on physical function, physical activity, lean body mass, bone mineral density, pain, and quality of life. Eur J Haematol 2024 Sep;113(3):298-309. [doi: 10.1111/ejh.14215] [Medline: 38717071]
25. Banerjee R, Huang CY, Dunn L, Knoche J, Ryan C, Brassil K, et al. Digital life coaching during stem cell transplantation: development and usability study. JMIR Form Res 2022 Mar 04;6(3):e33701 [FREE Full text] [doi: 10.2196/33701] [Medline: 35039279]
26. Veronese N, Stubbs B, Volpato S, Zuliani G, Maggi S, Cesari M, et al. Association between gait speed with mortality, cardiovascular disease and cancer: a systematic review and meta-analysis of prospective cohort studies. J Am Med Dir Assoc 2018 Nov;19(11):981-8.e7. [doi: 10.1016/j.jamda.2018.06.007] [Medline: 30056008]
27. Buta B, Zheng S, Langdon J, Adeosun B, Bandeen-Roche K, Walston J, et al. Agreement between standard and self-reported assessments of physical frailty syndrome and its components in a registry of community-dwelling older adults. BMC Geriatr 2022 Aug 25;22(1):705 [FREE Full text] [doi: 10.1186/s12877-022-03376-x] [Medline: 36008767]

XSL•FO
RenderX

Abbreviations

COH: City of Hope
CONSORT: Consolidated Standards of Reporting Trials
HCT: hematopoietic cell transplantation
mHealth: mobile health
MM: multiple myeloma
RPE: rating of perceived exertion
SPPB: Short Physical Performance Battery

Edited by L Buis; submitted 11.Nov.2025; peer-reviewed by R Banerjee, J Hardwick; comments to author 13.Jan.2026; accepted 02.Apr.2026; published 21.May.2026.

Please cite as:
Lee K, Shamunee J, Lee H, Du X, Lindenfeld L, Krishnan A, Nathwani N, Wong FL, Armenian S
mHealth-Supported Exercise Rehabilitation to Reverse Frailty After Autologous Transplantation in Multiple Myeloma: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e87628
URL: https://mhealth.jmir.org/2026/1/e87628
doi:10.2196/87628
PMID:

©Kyuwan Lee, Justin Shamunee, Haehyun Lee, Xinyi Du, Lanie Lindenfeld, Amrita Krishnan, Nitya Nathwani, F Lennie Wong, Saro Armenian. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 21.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Physiotherapeutic Elbow-Specific Training System for Patients After Arthroscopic Release of Elbow Contracture: Noninferiority Randomized Controlled Trial

Yanmao Wang[1*], PhD; Lihua Huang[2*], MSc; Ruixin Wang[3], PhD; Zhihao Xu[4], PhD; Yanhong Ma[2], PhD; Shiyang Yu[1], MD; Jian Ding[1], PhD; Yun Shen[5*], MD; Shengdi Lu[1,5*], MD

[1]Department of Orthopedics, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China

[2]Department of Rehabilitation, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, Shanghai, China

[3]School of Public Health, Fudan University, Shanghai, Shanghai, China

[4]Dong Fureng Institute of Economic and Social Development, Wuhan University, Wuhan, Hubei, China

[5]Pennington Biomedical Research Center, Baton Rouge, LA, United States

[*]these authors contributed equally

Corresponding Author:
Shengdi Lu, MD
Department of Orthopedics
Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine
600 Yishan Road
Shanghai, 200233
China
Phone: 86 13916482184
Email: lushengdi@shsmu.edu.cn

Abstract

Background: The effectiveness of a digital training (DT) system in which patients receive individually tailored physiotherapeutic elbow-specific training (PEST) delivered via a digital platform remains unclear.

Objective: This study determines the effectiveness of a DT system in which patients receive individually tailored PEST supervision and guidance via the Joymotion Intelligent Rehabilitation System and educational videos, compared with conventional training (CT) conducted by qualified physiotherapists at outpatient clinics and unsupervised home-based PEST in patients following arthroscopic release for posttraumatic elbow stiffness.

Methods: This single-center, noninferiority randomized controlled trial was conducted at the Rehabilitation Department of Shanghai Sixth People's Hospital between September 2020 and June 2024. Patients aged 16-65 years undergoing arthroscopic release for posttraumatic elbow stiffness were randomized to receive either a 12-week DT program or conventional outpatient clinic-based training. Outcome measures included elbow flexion-extension range of motion (primary outcome); forearm rotation; isometric and dynamic muscle strength; American Shoulder and Elbow Surgeons (ASES) and Disabilities of the Arm, Shoulder, and Hand (DASH) scores; EQ-5D-5L; cost-effectiveness; adherence; and adverse events, assessed at 4, 12, and 24 weeks postoperatively.

Results: At 12 weeks, the mean elbow flexion-extension range of motion improved similarly in the DT and CT groups (between-group difference –1.6°, 95% CI –8.2° to 4.9°; *P*=.53), confirming noninferiority. Forearm rotation gains were slightly greater with DT (difference 14.2°, 95% CI 2.9°-25.6°). Patient-reported outcomes were equivalent between groups: ASES function (difference 0.6, 95% CI –0.3 to 1.5; *P*=.39) and pain (difference 0, 95% CI –8.3 to 8.5; *P*=.68) subscores, DASH (difference 0.23, 95% CI –1.54 to 1.99; *P*=.68), and EQ-5D-5L index (difference 0.001, 95% CI –0.012 to 0.015; *P*=.56) showed no significant between-group differences. Nearly all patients completed the 12-week program in both arms (104/106, 98.1%, vs 101/102, 99%, adherence; odds ratio 0.52, 95% CI 0.05-5.77). Adverse events occurred in 29 out of 106 (27.4%) participants in the DT group and 32 out of 102 (31.4%) participants in the CT group (odds ratio 0.82, 95% CI 0.45-1.50). Total rehabilitation costs per patient were lower in the DT group by an average of CNY –7418.58, and incremental cost-effectiveness analysis indicated that DT provided comparable outcomes at lower cost.

XSL•FO
RenderX

Conclusions: Individually tailored PEST delivered via a DT system is a viable, cost-effective, and safe alternative to conventional outpatient clinic-based training following arthroscopic release for posttraumatic elbow stiffness. These findings support its integration into routine postsurgical care, particularly for patients facing barriers to traditional therapy.

Trial Registration: Chinese Clinical Trial Registry Chictr2400093415; https://www.chictr.org.cn/showprojEN.html?proj=240693

(JMIR Mhealth Uhealth 2026;14:e87459) doi:10.2196/87459

KEYWORDS

digital training system; noninferiority; elbow stiffness; arthroscopic release; cost-effectiveness

Introduction

Digital training (DT) systems, which involve the remote delivery of training services via digital technologies, have emerged as an attractive alternative or supplement to conventional in-hospital therapy. By using mobile apps and videoconferencing, DT enables patients to perform guided exercises at home, reducing the need for frequent hospital visits and overcoming geographic, logistical, and financial barriers. Studies have demonstrated that DT systems can yield outcomes comparable to those of traditional face-to-face interventions in postsurgical recovery settings [1-7].

In the United States and Europe, technological advances combined with supportive health care policies have accelerated the adoption of mobile-based DT. For instance, expanded telehealth coverage by Medicare in the United States and the integration of telemedicine within several European health care systems have facilitated timely, high-quality care delivered remotely [4,5,7]. These measures not only ease the burden of travel for patients but also promote adherence to training protocols through flexible scheduling and continuous monitoring. By contrast, DT systems in China are still in the developmental phase. Although major urban centers are beginning to deploy mobile health apps and telemedicine platforms, the supporting clinical evidence is less robust, and implementation practices vary widely.

Posttraumatic elbow stiffness, a condition marked by reduced range of motion (ROM) due to scar tissue formation following injury, can significantly impair daily activities [8]. Arthroscopic release is commonly used to surgically address elbow contracture by removing fibrotic tissue and improving joint mobility [9]. However, the success of this procedure relies heavily on effective postoperative rehabilitation. This typically involves intensive physiotherapeutic elbow-specific training (PEST) to restore elbow motion and function. Traditional in-hospital rehabilitation offers direct supervision and hands-on therapy; however, it presents several challenges. Frequent hospital visits can be particularly burdensome for patients in rural areas or those with limited mobility. Additionally, rigid scheduling and transportation difficulties may compromise patient adherence, thereby affecting the overall recovery process. DT systems address these challenges by bringing rehabilitation directly to the patient's home. DT programs provide daily exercise instructions, track adherence, and facilitate real-time, 2-way communication between patients and physical therapists. This model minimizes logistical hurdles and supports continuous, patient-centered care, which is crucial during the early postoperative period, when aggressive rehabilitation is needed to prevent recontracture. Furthermore, DT systems offer the flexibility to adjust exercise regimens in response to patient performance and feedback, closely mimicking the personalized care provided in traditional settings.

While DT systems have been extensively studied in lower limb surgeries, research on their use in upper extremity conditions, specifically posttraumatic elbow stiffness, remains relatively limited. Given the critical role of early and sustained rehabilitation in maintaining surgical gains and preventing recontracture, it is essential to evaluate whether DT systems can effectively replace traditional rehabilitation methods for elbow contracture. This study seeks to fill the current gap in the literature by rigorously comparing DT systems with conventional outpatient clinic-based training in patients following arthroscopic release for posttraumatic elbow stiffness.

The purpose of this study was to determine whether a DT system is noninferior to conventional outpatient clinic-based training in improving elbow ROM and other functional outcomes after arthroscopic release for posttraumatic elbow stiffness.

Methods

Study Setting and Design

This study was designed as a noninferiority randomized controlled trial with a 1:1 allocation ratio and was conducted at a single center, the Rehabilitation Department of Shanghai Sixth People's Hospital, from September 2020 to June 2024. The trial aimed to compare the efficacy of a DT system with that of conventional outpatient clinic-based training for patients undergoing arthroscopic release for posttraumatic elbow stiffness. This study was reported in accordance with the CONSORT (Consolidated Standards of Reporting Trials) extension for noninferiority and equivalence trials (CONSORT-NI), and the completed CONSORT checklist is provided in Multimedia Appendix 1.

Ethical Considerations

The study received ethical approval from the Ethics Committee of Shanghai Sixth People's Hospital (institutional review board approval number 2020-KY-089[K]) and was conducted in strict accordance with the principles outlined in the Declaration of Helsinki. All participants provided written informed consent before enrollment in the study. In addition, the study protocol was officially registered with the Chinese Clinical Trial Registry (identifier Chictr2400093415). To ensure privacy and confidentiality, all collected data were coded (deidentified) and securely stored using the hospital's electronic data capture system on local servers accessible only to the research team.

XSL•FO
RenderX

No financial compensation or other incentives were provided to participants for their involvement in this study.

Participants

Patients were recruited from the Department of Orthopedics at Shanghai Sixth People's Hospital between September 2020 and June 2024. The recruitment process involved 2 complementary approaches. First, orthopedic surgeons identified potentially eligible patients during routine preoperative evaluations and referred them to the research team. Second, the research team reviewed the weekly elective surgical schedule to identify patients scheduled for arthroscopic release of posttraumatic elbow contracture. All potentially eligible patients were approached by a member of the research team (YW), who explained the study objectives and procedures in detail. Those who expressed interest were formally screened using the predefined inclusion and exclusion criteria. Informed consent was obtained from patients who met the eligibility criteria and agreed to participate. Baseline data were collected before randomization. Participants were then consecutively allocated to either the DT system group or the conventional outpatient clinic-based training group (CT group) using a concealed randomization method. The assigned training protocol was initiated within the first postoperative week. The inclusion and exclusion criteria are listed in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

1. Inclusion criteria

- Patients who are 16 years of age or older.
- Patients who have a lack of elbow flexion or extension causing functional impairment that has been present for at least 6 months.
- Patients who have failed to respond to nonoperative treatment.
- Patients who are scheduled to undergo secondary arthroscopic capsulectomy or osteocapsular arthroplasty.

2. Exclusion criteria

- Patients with contraindications for rehabilitation or regional brachial plexus block use, such as those with bleeding diathesis, taking anticoagulants, or experiencing severe shoulder range of motion limitations.
- Patients with progressive or persistent neuropathy or neuritis.
- Those with preexisting conditions that might hinder their ability to fully participate in rehabilitation, including neuromuscular or psychosocial conditions.
- Patients with ongoing or recurrent contracture due to inflammatory diseases such as rheumatoid arthritis, juvenile idiopathic arthritis, or chondrolysis.
- Patients presenting with elbow joint infections or a history of previous joint infections.
- Patients with structural anomalies that could restrict elbow motion, unrelated to the condition being treated, such as dysplasia, malunion, osteonecrosis, or congenital deformities.
- Cases where a reasonable restoration of motion and function was not anticipated.
- Patients with insufficient postoperative regional anesthesia.
- Those experiencing intraoperative or postoperative complications that might impact the study's outcomes.
- Patients dealing with injuries or diseases during the postoperative period that could affect elbow function.
- Patients for whom arranging postoperative physical therapy appointments was not feasible.
- Cases where a significant portion of the procedure was performed in an open surgical manner.

Sample Size Calculation

A sample size of 73 patients was estimated to provide 80% power (2-sided α=.025) to detect a 10° difference in ROM at 12 weeks postoperatively. The 10° noninferiority margin was prespecified in the registered protocol and selected based on multiple lines of evidence. Sun et al [10] reported a distribution-based minimal clinically important difference (MCID) of 14.1° and an anchor-based MCID of 25.0° for the total elbow flexion-extension arc after arthrolysis, with a minimal detectable change at the 90% confidence level of 10.5°. In addition, O'Driscoll et al [11] observed a clinically meaningful between-group ROM difference of 9° in a trial comparing rehabilitation strategies after arthroscopic elbow contracture release. Our 10° margin was therefore more conservative than the published MCID estimates and approximated the MDC_{90} threshold, ensuring that any clinically important difference would be detected. To mitigate potential attrition, we increased the target enrollment to 122 patients, assuming a 40% dropout rate.

Intervention and Control Group

DT Group

Patients assigned to the DT group received individually tailored PEST through a mobile app (Joymotion R software; Shanghai Medmotion Medical Management Co, Ltd). On the day of discharge, a technician installed the app on each patient's mobile device, and patients used their home Wi-Fi to access the program. The app provided individually tailored PEST with detailed instructions, recorded exercise completion rates, and offered real-time feedback on training performance. In addition,

XSL•FO
RenderX

it enabled 2-way video and audio interactions with a physiotherapist (PT) at the rehabilitation center. Each week, the supervising PT initiated a scheduled conference with the patient to monitor progress, adjust the exercise regimen, and address any concerns. The PEST was individually prescribed by the supervising PT via the DT system and assigned as daily tasks to ensure continuity and adherence throughout the 12-week postoperative period.

CT Group

Patients in the control group received conventional outpatient clinic-based training (3-4 times per week), which consisted of scheduled training sessions and on-site PT supervision. This traditional approach involved direct, hands-on guidance and regular monitoring during therapy sessions at the Rehabilitation Department of Shanghai Sixth People's Hospital. Unlike the DT group, the control group did not use a mobile app or remote monitoring tools, and their rehabilitation was delivered entirely in a clinical setting.

Both groups followed a standardized 12-week PEST protocol (detailed in Multimedia Appendix 2). During weeks 1 and 2, patients performed pain-free passive and active-assisted ROM exercises hourly, along with grade I-III joint mobilizations and cryotherapy. Isometric strengthening was initiated at week 2. During weeks 3 and 4, light isotonic resistance exercises were introduced and progressed based on the achievement of near-functional ROM and tolerable pain (Visual Analog Scale score ≤4). During weeks 5-8, moderate-to-advanced resistance, eccentric loading, and sport- or work-specific functional training were added. During weeks 9-12, strengthening was maximized, and patients transitioned to self-managed maintenance programs. All exercises were performed within a pain-limited range; no absolute ROM ceilings were enforced. Outcome assessments were conducted at predetermined intervals to compare improvements in elbow function, ROM, and overall recovery.

Outcome Measures

Overview

Functional outcome measures were collected by physical therapists in outpatient clinics. Self-assessed outcome measures were also completed in the outpatient clinic using online questionnaires, with assistance from 2 physical therapists who were blinded to group allocation. Follow-up assessments were conducted at 4, 12, and 24 weeks after surgery.

Primary Outcome

The primary outcome was active elbow flexion-extension ROM (in degrees) measured at 12 weeks postoperatively. A standard handheld goniometer was used to measure the angle from full extension to maximum flexion of the affected elbow, providing an objective measure of motion.

Secondary Outcomes

Elbow flexion-extension ROM was measured in degrees using a goniometer at 4, 12, and 24 weeks after surgery. Forearm rotation ROM was also assessed at these same time points using a goniometer. All ROM measurements were performed with the patient's arm in a standardized position to ensure consistency, and greater angles indicated improved elbow joint flexibility.

Elbow flexor strength was evaluated at postoperative follow-up visits using a calibrated Baltimore Therapeutic Equipment (BTE) work simulator. Isometric strength was measured as the maximal voluntary contraction of the elbow flexor muscles maintained for 5 seconds with the elbow at a fixed angle. Dynamic strength was measured by having the patient perform repeated elbow flexion movements against a preset resistance on the BTE machine, simulating a functional lifting motion. Peak force or torque values obtained during each test were recorded. For standardization, strength outcomes were expressed as a percentage of the strength of the patient's unaffected contralateral elbow, allowing direct comparison with the healthy side (see Multimedia Appendix 2).

The American Shoulder and Elbow Surgeons (ASES) Shoulder Score Elbow Outcome questionnaire was administered at 4, 12, and 24 weeks to assess elbow function and pain from the patient's perspective. It yields 2 subcomponents: a pain score (typically assessed using a visual analog scale or pain rating, with 0 indicating no pain and higher values indicating worse pain) and a function score (based on the patient's ability to perform daily activities requiring elbow use). Each subscore contributes to an overall ASES elbow score (range 0-100 points), with higher scores indicating better function and less pain. For this study, the pain and function subscores were analyzed separately to assess improvements in each recovery domain.

Upper extremity disability and symptoms were evaluated using the Disabilities of the Arm, Shoulder, and Hand (DASH) questionnaire at 4, 12, and 24 weeks postoperatively. The DASH is a 30-item patient-reported outcome measure in which patients rate the difficulty of and interference with performing various physical activities involving the arm, shoulder, or hand. Scores range from 0 to 100, with 0 reflecting no disability and 100 indicating the most severe disability [12]. A decrease in the DASH score over time indicates improvement in arm function and reduction in disability [11].

General health-related quality of life was assessed at 4, 12, and 24 weeks using the EQ-5D-5L questionnaire. The EQ-5D-5L is a standardized instrument that covers 5 dimensions of health (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), each rated on 5 levels of severity [13]. Patients completed the questionnaire, and the responses were converted into a health utility index value and an EQ-5D Visual Analog Scale score for overall self-rated health. This provided a broad measure of patients' quality of life and well-being during recovery, with improvements in EQ-5D-5L scores indicating better overall health status as rehabilitation progressed.

Randomization, Allocation, and Blinding

A computer-generated randomization sequence was used to assign participants to either the DT or CT group. Block randomization with varying block sizes ensured balanced allocation over time. Allocation concealment was maintained using sealed, opaque, sequentially numbered envelopes prepared by an independent researcher (SY) who was not involved in patient recruitment or assessment. Following the initial

XSL•FO
RenderX

assessment, a volunteer PT distributed the envelopes according to the randomization sequence. These consecutively numbered, sealed, opaque envelopes were stored in a secure location accessible only to the unblinded lead researcher-PT (SL). This PT then scheduled the participants' postoperative training program and introduced the DT system.

Because of the nature of the interventions, blinding of patients and treating physical therapists was not feasible. However, outcome assessors responsible for measuring ROM and strength and administering standardized questionnaires (DASH, EQ-5D-5L, and ASES) were blinded to group allocation. Data analysts were also blinded to treatment assignments during statistical analysis to minimize potential bias. To support blinding during data collection, precoded data forms were used so that intervention group identity remained masked during data entry and evaluation.

Statistical Analysis

Patient data were securely coded and stored using the hospital's electronic data capture system on local servers. Baseline characteristics were summarized as mean (SD) for continuous variables and as frequencies and percentages for categorical variables. Between-group differences in baseline characteristics were assessed using independent 2-sample *t* tests (unpaired and 2-tailed) for continuous variables and chi-square tests for categorical variables. All hypothesis tests were 2-tailed, and a *P* value <.05 was considered statistically significant for these comparisons.

The primary efficacy analysis was performed on the intention-to-treat population in accordance with noninferiority trial guidelines. Cost-effectiveness and adverse event analyses were also conducted using the intention-to-treat population, with all randomized patients analyzed in their originally assigned groups.

The primary outcome was elbow flexion-extension ROM at 12 weeks postoperatively. Primary and secondary outcomes were analyzed using linear mixed-effects models to account for repeated-measures data collected over time. Fixed effects included group (DT vs CT), time (baseline, 4 weeks, 12 weeks, and 24 weeks), and their interaction (group × time), whereas age, sex, and the baseline value of each respective outcome were included as covariates. Specifically, for each outcome variable, the corresponding baseline measurement was entered as a covariate in the model, consistent with standard recommendations for controlling regression to the mean in randomized trials. Random effects were specified to account for within-participant variability, and least-squares means with 95% CIs were calculated for group comparisons at each time point. Noninferiority was established if the upper bound of the 95% CI for the between-group difference in 12-week ROM did not exceed 10°, indicating that the DT group was not more than 10° worse than the CT group. To account for rounding in range-of-motion measurements, we applied a *P* value correction method following the recommendation of Zdravkovic and Jost [14], setting the significance threshold at .03 for outcomes assessing mean differences in ROM. For all other outcomes, a *P* value <.05 was considered statistically significant.

Continuous secondary outcomes were also analyzed using linear mixed-effects models similar to those used for the primary outcome analysis, with adjustment for baseline values when available. Dichotomous outcomes were compared using Cochran-Mantel-Haenszel chi-square tests, controlling for baseline contracture severity.

For the multiple secondary end points, both the Bonferroni correction and the Holm step-down procedure were applied to adjust for multiple comparisons. Seven secondary efficacy outcomes were assessed at the 12-week primary end point: forearm rotation, isometric strength, dynamic strength, ASES function, ASES pain, DASH, and EQ-5D index. The Bonferroni correction set the significance threshold for each secondary outcome at α=.05/7. As the secondary outcomes were positively correlated, the standard Bonferroni approach was considered overly conservative. Therefore, the Holm step-down procedure was also applied because it controls the familywise error rate identically to the Bonferroni method while offering uniformly greater statistical power. Both methods yielded consistent conclusions. *P* values for secondary outcomes were interpreted against both adjusted thresholds. The primary outcome (noninferiority in ROM) was analyzed using a 1-sided alpha of .025 and was not subject to multiplicity adjustment.

The incremental cost-effectiveness ratio (ICER) was used to assess the cost-effectiveness of the DT group compared with the CT group. The ICER represents the difference in costs and outcomes between the groups. The numerator of the cost-effectiveness ratio was calculated as the monetary cost of the DT intervention minus the monetary cost of CT. Annual project costs were estimated by extrapolating the 12-week implementation costs. The denominator was calculated as the ROM gained in the DT group minus the ROM gained in the CT group at 12 weeks. Bootstrapping was used for pairwise comparisons of mean costs and effects between the DT and CT groups. CIs for mean differences in effects were obtained using bootstrapping with 1000 replications. The bootstrapped cost-effect pairs were also graphically represented on a cost-effectiveness plane [15].

Missing data in this study occurred exclusively at the visit level. If a patient missed a follow-up visit, all outcomes from that visit were missing; however, data from attended visits remained available. There was no item-level missingness within completed visits. Given this structure, mixed-effects models in the intention-to-treat analysis retained all available data for each participant. No imputation was applied to address missing values. As data loss occurred at the level of entire visits rather than within-visit fields, tests for data missing completely at random, such as Little missing completely at random test, were not performed. The analysis was repeated in the per-protocol population, which included only patients who completed all scheduled follow-up assessments. Missing outcome data were therefore addressed using both intention-to-treat and per-protocol analytical strategies.

XSL•FO
RenderX

Results

Participants

The DT group (n=106) and the CT group (n=102) were well matched at baseline (Table 1). The CONSORT flowchart for both the control and intervention groups, detailing exclusions and follow-up losses, is illustrated in Figure 1. The mean age was 47 years in the DT group and 48 years in the CT group, with males comprising 77 of 106 (72.6%) patients and 80 of 102 (78.4%) patients, respectively. The mean BMI was approximately 23 kg/m^2 in the DT group and 22 kg/m^2 in the CT group. Manual workers accounted for 59 of 106 (55.7%) patients in the DT group and 60 of 102 (58.8%) patients in the CT group, whereas 73 of 106 (68.9%) patients and 56 of 102 (54.9%) patients, respectively, had achieved at least a high school education. Most patients were covered by government insurance. The numbers of current smokers (*P*=.39), alcohol users (*P*=.28), and users of paracetamol and nonsteroidal anti-inflammatory drugs (*P*=.55) showed no significant between-group differences. Clinically, most patients had no prior history of significant elbow injuries, such as distal humerus, capitellum, olecranon, Monteggia, or radial head fractures, and only a small proportion had previously undergone surgical treatment for these conditions. No significant between-group differences were observed in records of previous operations (osteocapsular arthroplasty, *P*=.14; capsular release, *P*=.14; ulnar nerve management—limited decompression, *P*=.76; ulnar nerve management—subcutaneous transposition, *P*=.76; removal of heterotopic ossification, *P*=.61; radial head excision, *P*=.94; and hardware removal, *P*=.15). Baseline elbow function was comparable between groups, with mean flexion-extension ROMs of approximately 78°-79°, forearm rotation ROMs ranging from 76° to 91°, and isometric and dynamic flexion strength values of approximately 69%-70% and 77% of the contralateral side, respectively. Patient-reported outcomes, including ASES scores, DASH scores, and EQ-5D-5L indices, were also similar, reflecting moderate impairment and reduced quality of life in both groups.

XSL•FO
RenderX

Table 1. Baseline characteristics of the DT[a] and CT[b] groups (intention-to-treat population).

Characteristic	DT group (n=106)	CT group (n=102)	*P* value
Demographic characteristics			
Male patients, n (%)	69 (65.1)	73 (71.6)	.32
Age (years), mean (SD)	47 (7.30)	48 (7.18)	.25
BMI (kg/m^2), mean (SD)	23 (2.86)	22 (3.21)	.19
Occupation, n (%)			.64
Manual worker	59 (55.7)	60 (58.8)	
Nonmanual worker	47 (44.3)	42 (41.2)	
Education level, n (%)			.04
Lower than high school	33 (31.1)	46 (45.1)	
Equal/higher to high school	73 (68.9)	56 (54.9)	
Insurance type, n (%)			.70
Government	69 (65.1)	70 (68.6)	
Commercial	10 (9.4)	11 (10.8)	
Self-financed	27 (25.5)	21 (20.6)	
Current smoker, n (%)	51 (48.1)	43 (42.2)	.39
Current alcohol use, n (%)	43 (40.6)	34 (33.3)	.28
Paracetamol and nonsteroidal anti-inflammatory drug, n (%)	32 (30.2)	27 (26.5)	.55
History of previous etiology for the elbow			
Distal humerus fracture, n (%)			.33
None	86 (81.1)	81 (79.4)	
Conservative treatment	14 (13.2)	10 (9.8)	
Surgical treatment	6 (5.7)	11 (10.8)	
Capitellum fracture, n (%)			.84
None	83 (78.3)	81 (79.4)	
Conservative treatment	14 (13.2)	11 (10.8)	
Surgical treatment	9 (8.5)	10 (9.8)	
Olecranon fracture, n (%)			.99
None	86 (81.1)	83 (81.4)	
Conservative treatment	16 (15.1)	15 (14.7)	
Surgical treatment	4 (3.8)	4 (3.9)	
Monteggia fracture, n (%)			.41
None	95 (89.6)	85 (83.3)	
Conservative treatment	7 (6.6)	11 (10.8)	
Surgical treatment	4 (3.8)	6 (5.9)	
Radial head, n (%)			.35
None	85 (80.2)	88 (86.3)	
Conservative treatment	15 (14.2)	8 (7.8)	
Surgical treatment	6 (5.7)	6 (5.9)	
Terrible triad injury, n (%)			.89
None	95 (89.6)	92 (90.2)	
Conservative treatment	0 (0)	0 (0)	

XSL•FO
RenderX

Characteristic	DT group (n=106)	CT group (n=102)	*P* value
Surgical treatment	11 (10.4)	10 (9.8)	
Records of previous operation			
Osteocapsular arthroplasty, n (%)	82 (77.4)	87 (85.3)	.14
Capsular release (soft tissue only), n (%)	24 (22.6)	15 (14.7)	.14
Ulnar nerve management—limited decompression, n (%)	92 (86.8)	87 (85.3)	.76
Ulnar nerve management—subcutaneous transposition, n (%)	14 (13.2)	15 (14.7)	.76
Removal of heterotopic ossification, n (%)	13 (12.3)	15 (14.7)	.61
Radial head excision with or without Interposition arthroplasty, n (%)	8 (7.5)	8 (7.8)	.94
Hardware removal, n (%)	13 (12.3)	20 (19.6)	.15
Tourniquet time (minutes), mean (SD)	90 (29.65)	89 (27.01)	.65
Function, mean (SD)[c]			
ROM[d] of elbow flexion to extension motion (degrees)	78 (22.08)	79 (21.97)	.81
ROM of forearm rotation (degrees)	91 (44.06)	76 (42.47)	.01
Flexion strength—isometric elbow flexion strength (%)	69 (3.78)	70 (4.06)	.16
Flexion strength—dynamic elbow flexion strength (%)	77 (4.20)	77 (3.78)	.45
PROMs[e], mean (SD)			
ASES[f] Function subscore (points)	22 (2.92)	22 (2.70)	.39
ASES Pain subscore (points)	22 (3.13)	23 (2.69)	.76
DASH[g] score (points)	40 (5.88)	40 (4.46)	.93
EQ-5D-5L	0.71 (0.05)	0.71 (0.04)	.87

[a]DT: digital training.

[b]CT: conventional training.

[c]Isometric flexion strength and dynamic flexion strength were measured and compared with the contralateral side using a BTE machine (Simulator II).

[d]ROM: range of motion.

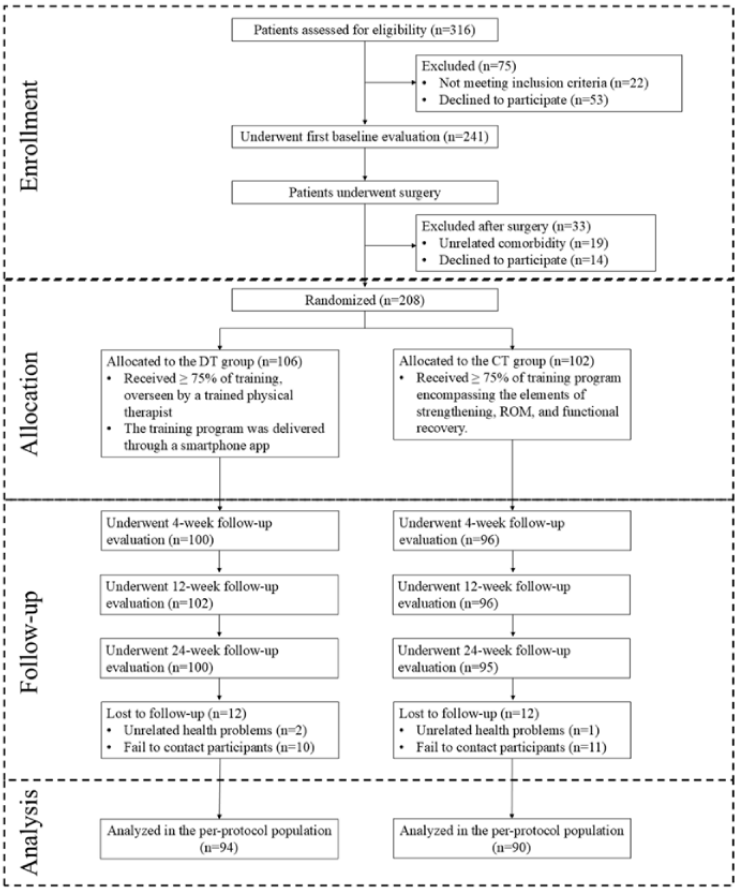
[e]PROM: passive range of motion.

[f]ASES: American Shoulder and Elbow Surgeons.

[g]DASH: Disabilities of the Arm, Shoulder, and Hand.

XSL•FO
RenderX

Figure 1. Flowchart of the study from enrollment to analysis. CT: conventional training; DT: digital training; ROM: range of motion.



Primary and Secondary Outcomes

Using the linear mixed-effects model, the adjusted between-group difference in 12-week elbow flexion-extension ROM was –1.6° (95% CI –8.2° to 4.9%; Tables 2 and 3), indicating noninferiority of DT compared with CT. A complete-case analysis excluding 12 patients without ROM follow-up yielded a similar difference of –1.5° (95% CI –8.8° to 5.0°), confirming that the primary outcome was not sensitive to the handling of missing data. Thus, the conclusion of noninferiority remained unchanged when only observed cases were analyzed compared with multiply imputed data.

Secondary outcomes improved comparably in both groups, with no between-group differences meeting the adjusted significance criteria after Bonferroni correction for multiple end points. Forearm rotation ROM increased slightly more in the DT group, with an adjusted difference of 14.2° (95% CI 2.9°-25.6°). Isometric and dynamic flexion strength gains were nearly identical between groups, with differences of 0.04% (95% CI –2.1% to 2.2%) and 0.26% (95% CI –1.5% to 2.0%), respectively, when normalized to the contralateral side. Functional recovery, assessed using the ASES function subscore, showed a small difference of 0.6 points (95% CI –0.3 to 1.5), indicating similar functional improvement in both groups. Pain reduction, measured using the ASES pain subscore, was comparable, with a near-zero adjusted difference (95% CI –8.3 to 8.5). The DASH score, reflecting disability, improved similarly in both groups, with a minimal between-group difference of 0.23 points (95% CI –1.54 to 1.99). Lastly, health-related quality of life, measured using the EQ-5D-5L, showed no meaningful between-group difference, with an adjusted index difference of 0.001 (95% CI –0.012 to 0.015; Tables 2 and 3). Thus, forearm rotation (P<.001), ASES function (P<.001), and EQ-5D-5L (P=.003) showed statistically significant between-group differences, whereas isometric strength (P=.76), dynamic strength (P=.67), ASES pain (P=.18), and DASH (P=.87) showed no significant between-group differences. Sensitivity analyses using the per-protocol population showed similar results (Multimedia Appendices 3 and 4). Sensitivity analyses of secondary outcomes at 12 weeks using Bonferroni and Holm step-down adjustments for multiple comparisons are presented in Multimedia Appendix 5. An additional sensitivity analysis that included education level as a covariate in the primary mixed-effects model yielded an adjusted between-group difference of –1.7° (95% CI –8.4° to 5.0°), confirming that the noninferiority conclusion was robust to adjustment for this baseline imbalance.

XSL•FO
RenderX

Table 2. Unadjusted changes in outcomes for the DT[a] and CT[b] groups at weeks 4, 12, and 24 after surgery (intention-to-treat population).[c]

Outcome[d]	4 weeks after surgery			12 weeks after surgery			24 weeks after surgery		
	DT group (n=106), mean (SD)	CT group (n=102), mean (SD)	*P* value	DT group (n=106), mean (SD)	CT group (n=102), mean (SD)	*P* value	DT group (n=106), mean (SD)	CT group (n=102), mean (SD)	*P* value
ROM[e] of elbow flexion to extension motion (degrees)	9.36 (12.58)	10.17 (14.77)	.69	19.47 (15.30)	20.89 (17.38)	.55	22.82 (15.13)	24.22 (16.61)	.55
ROM of forearm rotation (degrees)	25.96 (20.49)	26.39 (20.29)	.89	43.24 (25.07)	46.67 (22.02)	.33	52.07 (26.64)	56.44 (23.96)	.24
Flexion strength—isometric elbow flexion strength (%)	5.27 (6.10)	5.01 (6.10)	.77	21.53 (6.04)	20.41 (6.09)	.21	25.93 (5.35)	24.38 (5.40)	.05
Flexion strength—dynamic elbow flexion strength (%)	5.11 (5.87)	4.25 (4.79)	.28	17.01 (5.58)	16.53 (5.28)	.55	23.29 (5.42)	23.04 (5.38)	.75
ASES[f] Function subscore (points)	4.31 (2.89)	3.98 (2.99)	.45	8.72 (3.89)	8.23 (3.95)	.39	6.45 (4.29)	5.93 (4.11)	.40
ASES Pain subscore (points)	–4.99 (4.41)	–5.67 (4.49)	.30	–12.72 (6.51)	–13.12 (6.70)	.68	–10.90 (6.55)	–11.26 (6.94)	.72
DASH[g] score (points)	–11.97 (7.70)	–12.50 (7.09)	.63	–24.94 (9.36)	–25.49 (8.98)	.68	–26.18 (9.22)	–26.59 (8.77)	.76
EQ-5D-5L[h]	0.00 (0.00)	0.00 (0.00)	N/A[i]	0.08 (0.05)	0.08 (0.06)	.56	0.08 (0.05)	0.09 (0.06)	.56

[a]DT: digital training.

[b]CT: conventional training.

[c]Values represent the mean change from baseline for each group, with *P* values comparing between-group differences at each follow-up point. All outcome measures were unadjusted.

[d]Isometric flexion strength and dynamic flexion strength were measured and compared with the contralateral side using a Baltimore Therapeutic Equipment machine (Simulator II).

[e]ROM: range of motion.

[f]ASES: American Shoulder and Elbow Surgeons.

[g]DASH: Disabilities of the Arm, Shoulder, and Hand.

[h]The EQ-5D-5L change of 0.00 (SD 0.00) at 4 weeks in both groups reflects the limited responsiveness of this generic health utility instrument in the early postoperative period. At 4 weeks after elbow surgery, improvements in joint-specific function are typically offset by ongoing postoperative recovery burden, resulting in negligible net change in global health utility. This pattern is consistent with published studies reporting limited EQ-5D sensitivity to change in upper extremity orthopedic populations during the early recovery phase.

[i]N/A: not applicable.

XSL•FO
RenderX

Table 3. Adjusted effectiveness estimates from linear mixed-effects models (intention-to-treat population).[a]

Outcome[b]	4 weeks after surgery			12 weeks after surgery			24 weeks after surgery		
	Coefficient	95% CI	*P* value	Coefficient	95% CI	*P* value	Coefficient	95% CI	*P* value
ROM[c] of elbow flexion to extension motion (degrees)	–2.131	–8.843 to 4.580	.53	–2.810	–5.311 to –0.309	.03	–2.764	–4.592 to –0.934	.003
ROM of forearm rotation (degrees)	14.577	3.587 to 25.568	.009	13.317	10.422 to 16.213	<.001	12.942	10.540 to 15.345	<001
Flexion strength—isometric elbow flexion strength (%)	–0.356	–1.011 to 0.298	.29	–0.073	–0.539 to 0.393	.76	0.237	–0.392 to 0.867	.46
Flexion strength—dynamic elbow flexion strength (%)	–0.005	–0.751 to 0.742	.99	0.136	–0.487 to 0.759	.67	0.139	–0.283 to 0.561	.52
ASES[d] Function subscore (points)	0.476	–0.426 to 1.377	.30	0.644	0.327 to 0.960	<.001	0.804	0.420 to 1.188	<.001
ASES Pain subscore (points)	0.255	–0.642 to 1.151	.58	0.298	–0.133 to 0.729	.17	0.317	–0.031 to 0.664	.07
DASH[e] score (points)	–0.040	–1.669 to 1.589	.96	0.025	–0.270 to 0.320	.87	–0.017	–0.331 to 0.298	.92
EQ-5D-5L	–0.003	–0.0134 to 0.009	.65	–0.003	–0.006 to –0.001	.003	–0.005	–0.009 to –0.001	.007

[a]Each coefficient represents the estimated between-group difference in the change from baseline (intervention group minus control group) for the specified outcome at that follow-up time point. Positive coefficients indicate higher scores in the intervention group compared with the control group, whereas negative values indicate lower scores in the intervention group. Each estimate is presented with its 95% CI and corresponding *P* value. All outcome measures were adjusted for baseline values in the model.

[b]Isometric flexion strength and dynamic flexion strength were measured and compared with the contralateral side using a BTE machine (Simulator II).

[c]ROM: range of motion.

[d]ASES: American Shoulder and Elbow Surgeons.

[e]DASH: Disabilities of the Arm, Shoulder, and Hand.

Cost-Effectiveness Outcomes

The economic evaluation is presented as a short-term cost-consequence analysis given the limited 12-24-week time horizon, with bootstrapped results presented in Tables 4 and 5 and Figures 2 and 3. We note that the incremental health effects were very close to 0 for most outcomes, rendering formal cost-per-quality-adjusted life year ratios unstable and difficult to interpret. The results should therefore be interpreted as descriptive cost-consequence findings rather than as a definitive cost-utility evaluation. Figure 2 illustrates the incremental cost-effectiveness plane. Across 1000 bootstrap resamples, the incremental cost of DT versus CT was almost always negative (median CNY –7400 [US $–1087]), whereas the incremental effects clustered near 0. The majority of bootstrap iterations fell within the southeast quadrant (negative Δcost, positive Δeffect), indicating that DT was cost-saving and at least as effective as CT in most resamples. A minority of points fell within the southwest quadrant (cost-saving but slightly less effective), and virtually none fell within the costly quadrants, underscoring the consistency of the cost-saving result. Figure 3 shows the cost-effectiveness acceptability curve. At a willingness-to-pay threshold of CNY 0 (US $0) per unit of effect, the probability that DT was cost-effective was 75%. This probability increased to 78% at a threshold of CNY 134,000 (US $19,677) per quality-adjusted life year and approached 80% as willingness-to-pay increased. These probabilistic findings suggest that DT was associated with lower costs in most bootstrap resamples. However, because incremental health effects clustered near 0, the cost-effectiveness acceptability curve should be interpreted with caution. The high probability of cost-effectiveness primarily reflects the cost-saving nature of the digital intervention rather than a meaningful differential in health effects. These short-term findings should not be extrapolated to longer time horizons without further study because downstream costs or savings associated with preventing recontracture or reoperation were not captured.

XSL•FO
RenderX

Table 4. Average total cost per patient in the DT[a] and CT[b] groups during the 24 weeks after the surgery (intention-to-treat population).[c]

Cost category (CNY[d])	DT group (n=106), mean (SD)	CT group (n=102), mean (SD)	*P* value
Direct medical cost			
Physical therapist cost	13,960.00 (0.00)	13,706.88 (2036.32)	.21
Hospital stay cost	7001.43 (1020.29)	6984.06 (908.39)	.90
Primary care cost	249.00 (34.80)	2438.95 (362.33)	<.001
Secondary care cost	5984.12 (1005.90)	7922.16 (1227.93)	<.001
Direct nonmedical cost			
Paid home cost	1211.67 (323.50)	1449.24 (347.61)	<.001
Medication	8057.68 (860.30)	8668.13 (1531.21)	.001
Transportation cost	775.05 (162.58)	2025.34 (335.97)	<.001
Nutrition cost	4569.25 (513.63)	5138.81 (493.07)	<.001
Indirect cost			
Lost wages for patients	31,323.61 (16,406.72)	32,576.12 (16977.68)	.60
Lost wages for families	4218.69 (4209.30)	3859.40 (4576.58)	.57
Total cost	77,350.51 (16,923.96)	84,769.09 (17935.57)	.003

[a]DT: digital training.

[b]CT: conventional training.

[c]All relevant costs were captured from a societal perspective. This included direct medical costs (hospital stay, primary and secondary care, rehabilitation and physical therapy costs, and medication costs), direct nonmedical costs related to care (transportation for medical visits and any specialized nutritional support during recovery), and indirect costs due to productivity loss (lost wages for patients during recovery or disability, and lost income for family members/caregivers, if applicable). Costs were obtained from hospital billing records and patient self-reports where needed (wage losses). All costs are reported in 2024 Chinese Yuan, as 2024 was the end of the enrollment period; costs incurred in earlier years were adjusted to 2024 price levels using the consumer price index. Costs and outcomes occurring beyond the initial year were discounted at an annual rate of 3% to reflect time preference, consistent with standard practice in health economic evaluations.

[d]CNY 1=US $0.15.

XSL•FO
RenderX

Table 5. ICER[a] (intention-to-treat population).[b,c]

Main analysis—mixed effects	Values
Incremental cost (CNY[d])	–7418.58
Incremental ROM[e] of elbow flexion to extension motion (degrees)	–1.663 (–8.286 to 4.961)
Incremental ROM of forearm rotation (degrees)	13.07 (2.110 to 24.031)
Incremental flexion strength—isometric elbow flexion strength (%)	0.23 (–2.163 to 2.624)
Incremental flexion strength—dynamic elbow flexion strength (%)	0.159 (–2.079 to 2.397)
Incremental ASES[f] Elbow Function subscore (points)	0.631 (–0.280 to 1.541)
Incremental ASES Elbow Pain subscore (points)	0.032 (–1.609 to 1.673)
Incremental DASH[g] score (points)	0.253 (–1.639 to 2.145)
Incremental EQ-5D-5L	0.000 (–0.015 to 0.015)
ICER ROM of elbow flexion to extension motion (degrees)	4460.96
ICER ROM of forearm rotation (degrees)	–567.60
ICER flexion strength—isometric elbow flexion strength (%)	–32,254.70
ICER flexion strength—dynamic elbow flexion strength (%)	–46,657.74
ICER ASES Function subscore (points)	–11,756.86
ICER ASES Pain subscore (points)	–231,830.63
ICER DASH score (points)	–29,322.45
ICER EQ-5D-5L	–112,402,727.27

[a]ICER: incremental cost-effectiveness ratio.

[b]The table displays the incremental cost, incremental effect in primary outcome scores, and the resulting ICER. ICER is defined as the incremental cost divided by the incremental effect, representing the additional cost per unit improvement in outcome for the intervention group versus the control group. Incremental effects are given for all outcome measures. ICER values are presented for the 12-month follow-up under both the intention-to-treat and per-protocol analyses. A negative ICER indicates that the intervention group achieved equivalent or better outcomes at a lower cost compared with the control group.

[c]Values are reported as follows: incremental effects as coefficient (95% CI), and ICER values as point estimates computed from the ratio of incremental cost to incremental effect.

[d]CNY 1=US $0.15.

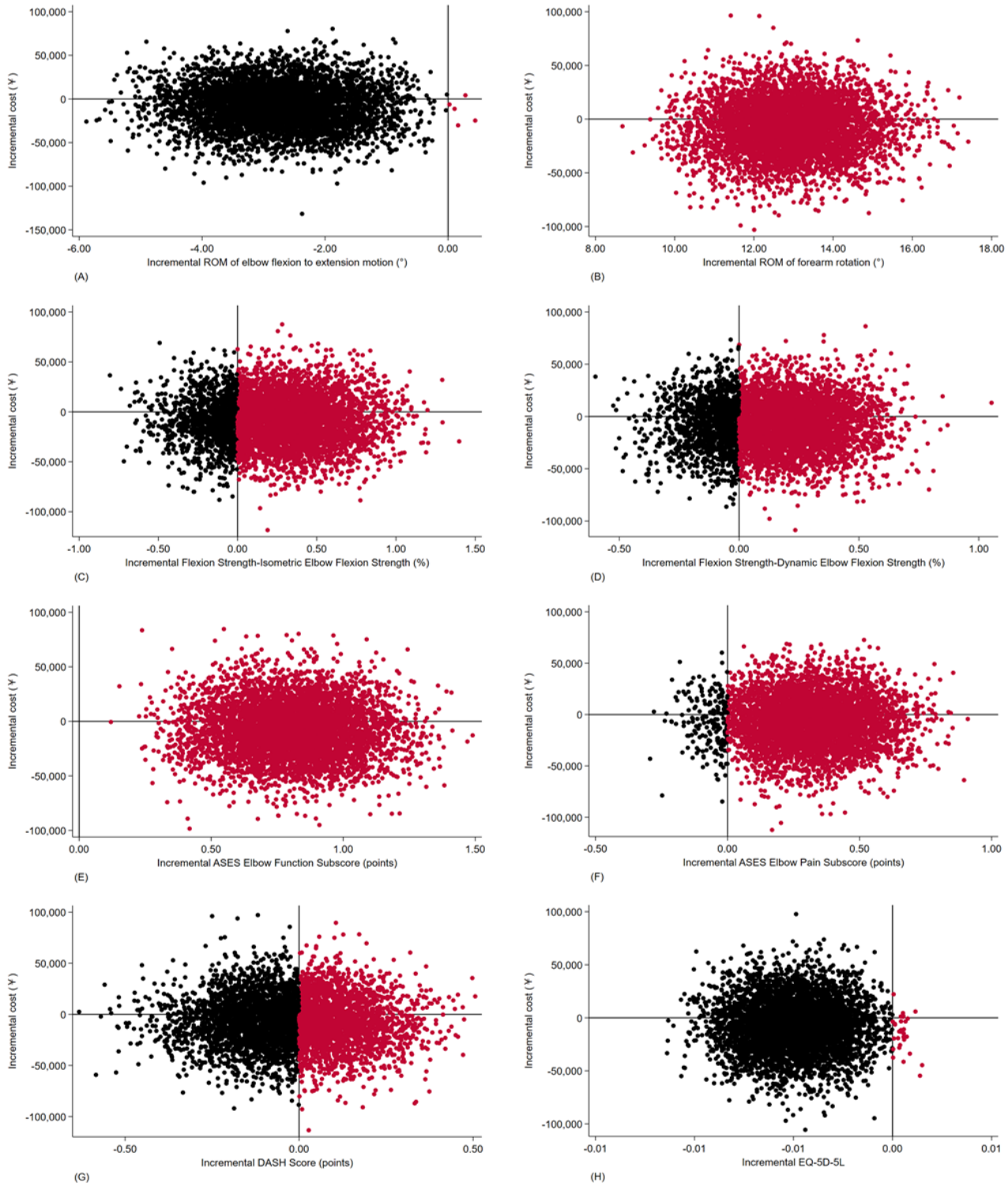
[e]ROM: range of motion.

[f]ASES: American Shoulder and Elbow Surgeons.

[g]DASH: Disabilities of the Arm, Shoulder, and Hand.

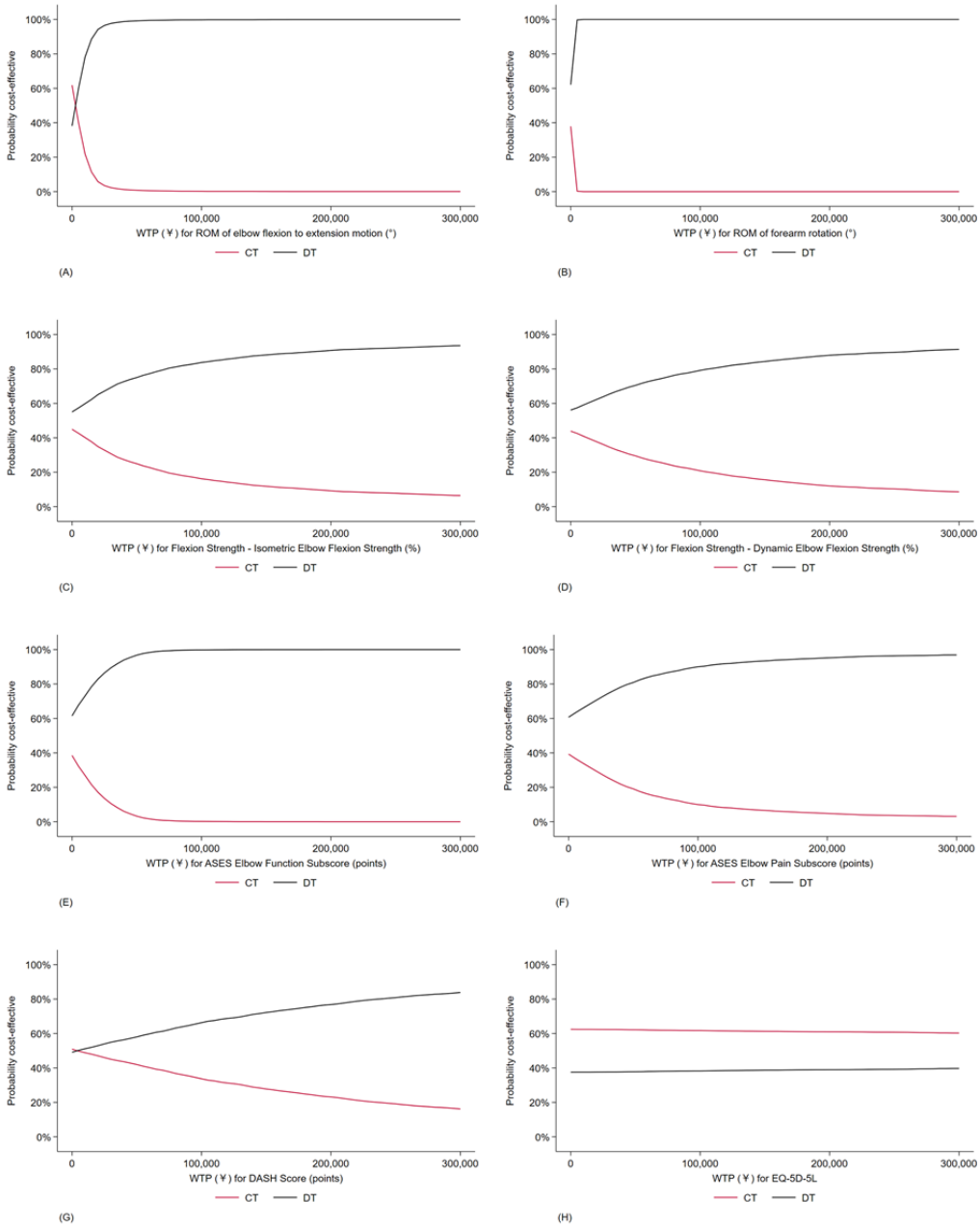
XSL•FO
RenderX

Figure 2. Bootstrapped cost-effectiveness planes (BCEPs) for the comparison of digital training (DT) vs conventional training (CT). Panels show the BCEPs for incremental outcomes of (A) elbow flexion-extension range of motion (ROM), (B) forearm rotation ROM, (C) isometric elbow flexion strength, (D) dynamic elbow flexion strength, (E) American Shoulder and Elbow Surgeons (ASES) function score, (F) ASES pain score, (G) Disabilities of the Arm, Shoulder, and Hand (DASH) score, and (H) EQ-5D-5L score. Each point represents an incremental cost-effect pair (ΔEffect on the horizontal axis and ΔCost on the vertical axis) derived from 1000 bootstrap replicates of the trial data. The origin (0,0) indicates no difference between groups. Points below the horizontal axis indicate that DT is less costly than CT, whereas points to the right of the vertical axis indicate that DT is more effective than CT. The cost-effectiveness acceptability curve illustrates the probability that DT is cost-effective compared with CT across a range of willingness-to-pay thresholds per unit of effect gained (per quality-adjusted life-year gained). The curve was generated from the same 1000 bootstrap replicates by calculating the proportion of simulations in which the net monetary benefit of DT compared with CT was positive at each willingness-to-pay threshold.



XSL•FO
RenderX

Figure 3. Cost-effectiveness acceptability curves (CEACs) for the comparison of digital training (DT) vs conventional training (CT). Panels show the CEACs for (A) elbow flexion-extension range of motion (ROM), (B) forearm rotation ROM, (C) isometric elbow flexion strength, (D) dynamic elbow flexion strength, (E) American Shoulder and Elbow Surgeons (ASES) function score, (F) ASES pain score, (G) Disabilities of the Arm, Shoulder, and Hand (DASH) score, and (H) EQ-5D-5L score. Each curve illustrates the probability that DT is cost-effective compared with CT across a range of willingness-to-pay (WTP) thresholds per unit of effect gained (per quality-adjusted life-year gained). The curves were derived from 1000 bootstrap replicates by calculating the proportion of simulations in which the net monetary benefit of DT compared with CT was positive at each WTP threshold.



Adherence and Acceptability

Patients in both groups attended approximately 3.4-3.5 sessions per week. The DT group rated agreement with the assigned exercise plan at 8.2, compared with 8.4 in the CT group. Adherence to the recommended exercise program was high in both groups, with scores of 8.5 and 8.6, respectively. Similarly, ratings for pain relief were 8.7 versus 8.8, and ratings for functional improvement were 8.8 versus 8.9. Overall satisfaction with the exercise protocol was very high, at 9.5 in the DT group and 9.6 in the CT group. No significant between-group differences were observed, indicating excellent adherence and acceptability (Table 6).

XSL•FO
RenderX

Table 6. Patients' adherence to treatment.[a]

Outcome measure	DT[b] group (n=106)	CT[c] group (n=102)	*P* value
Number of sessions performed per week, mean (SD)	3.4 (0.9)	3.5 (1.0)	.43
Agreement with the following questions (0-10)[d], mean (SD)			
To what extent did you agree to accept the allocated exercise plan?	8.2 (1.3)	8.4 (1.2)	.28
To what extent did you do the exercise program as recommended?	8.5 (1.4)	8.6 (1.4)	.44
To what extent do you agree that the intervention relieved your pain?	8.7 (1.4)	8.8 (1.3)	.63
To what extent do you agree that the intervention improved your function?	8.8 (1.1)	8.9 (1.2)	.82
To what extent were you satisfied with the exercise protocol?	9.5 (0.5)	9.6 (0.4)	.11

[a]Adherence data were retrieved from patients' logbooks and the health information system in both groups.

[b]DT: digital training.

[c]CT: conventional training.

[d]0=strongly disagree to 10=strongly agree.

Adverse Events

Adverse events occurred in 29 of 106 (27.4%) patients in the DT group and 32 of 102 (31.4%) patients in the CT group. Therapy-related events occurred in 4 patients in the DT group and 6 patients in the CT group, with common issues including pain, swelling, and tenosynovitis. Serious adverse events leading to study withdrawal occurred in 2 (2%) DT patients and 1 (1%) CT patient (Table 7).

XSL•FO
RenderX

Table 7. Adverse events and serious adverse events.

Adverse events	DT[a] group (n=106)	CT[b] group (n=102)
Patients with adverse events, n (%)	29 (27.4)	32 (31.4)
Events related to study therapy, n	4	6
Events unrelated to study therapy, n	36	41
Type of event, n		
Pain	11	14
Swelling	7	8
Muscle strain	1	2
Tenosynovitis	3	2
Signs of infection (swelling, redness, heat, or pus)	0	1[c]
Mobilization under anesthesia	1	1
Other, n		
Nausea and dizziness	0	2
Shoulder pain	2	3
Wrist pain	4	2
Anxiety about elbow recovery	7	6
Serious adverse events[d]		
Patients with serious adverse events, n (%)	2 (1.9)	1 (1.0)
Events related to study therapy, n	0	0
Events unrelated to study therapy, n	2	1
Hip fracture due to fall	1[e]	0
Waist fracture due to fall	1[e]	0
Spinal surgery	0	1[e]

[a]DT: digital training system.

[b]CT: conventional outpatient clinic-based training.

[c]One patient sustained a mild infection of the surgical site and healed with conservative treatment.

[d]Patients with serious adverse events were automatically withdrawn from the study.

[e]Two patients in the DT group sustained accidental falls during daily activities unrelated to the rehabilitation protocol. One patient sustained a hip fracture while walking outdoors (approximately 6 weeks postoperatively, not during or immediately after an exercise session) and underwent surgical fixation. Another patient sustained a lumbar vertebral fracture from a fall at home during routine household activities (approximately 8 weeks postoperatively) and was treated conservatively. Neither event was temporally associated with prescribed rehabilitation exercises, and the DT protocol did not include balance-challenging or weight-bearing lower extremity exercises. Both events were adjudicated as unrelated to the study intervention. Both patients were automatically withdrawn from the study.

Discussion

Principal Findings

This trial demonstrates that a DT system in which patients receive individually tailored PEST yields outcomes comparable to those of conventional outpatient clinic-based training after arthroscopic release for posttraumatic elbow stiffness. At 12 weeks, the DT group met the prespecified noninferiority criterion for elbow flexion-extension ROM relative to the CT group. The between-group difference was –1.6° (95% CI –8.2° to 4.9°), with the upper bound of the CI well within the 10° noninferiority margin. However, the lower bound of –8.2° approaches a magnitude that some clinicians may consider clinically relevant in a population with stiff elbows. Statistical noninferiority does not necessarily imply clinical interchangeability, and individual patient factors should inform treatment selection. Secondary outcomes, including pain relief and functional improvement, were also comparable between groups. Both interventions demonstrated high adherence, and no serious adverse events related to treatment were observed. The inclusion of bootstrap-based uncertainty analyses provides additional support for the economic advantage of the digital program. The scatter distribution on the cost-effectiveness plane suggests a robust trend favoring DT as a cost-saving intervention, and the cost-effectiveness acceptability curve indicates a high probability that DT is cost-effective under typical willingness-to-pay thresholds. However, the economic evaluation was limited to a 12-week postoperative horizon and was based on very small differences in quality-of-life indices.

XSL•FO
RenderX

Although DT appears cost-effective in the short term, a longer-term evaluation capturing downstream costs and quality-of-life changes would provide a more comprehensive assessment of its cost utility. Nevertheless, within the trial period, the probabilistic analysis suggests only a modest likelihood (20%-25%) that CT would be the more cost-effective option, reinforcing the conclusion that the DT system is an economically attractive alternative to standard care.

Comparison With Prior Work

Both the DT and CT groups in the RCT achieved substantial gains in elbow flexion-extension ROM. This finding aligns with previous studies showing that DT can restore joint mobility as effectively as CT. For example, a pilot trial involving patients with elbow fractures found significant ROM improvements in both digital rehabilitation and clinic-based therapy, with no between-group difference [6]. Similarly, an integrative review of digital rehabilitation after orthopedic surgeries (primarily knee and hip arthroplasties) concluded that internet-based training yields improvements in joint ROM and other outcomes comparable to those achieved with standard in-person rehabilitation [16]. In other words, DT systems do not appear to impede recovery of motion, and existing evidence suggests that they achieve ROM gains equivalent to those of traditional therapy, consistent with the findings of our elbow stiffness trial.

Baseline forearm rotation differed between groups (P=.01), with the DT group demonstrating higher baseline values. Although the mixed-effects model included baseline forearm rotation as a covariate, which is the recommended approach for addressing regression to the mean, residual confounding due to this baseline imbalance cannot be entirely excluded. The adjusted between-group difference of approximately 13° (Table 3), therefore, reflects the estimated treatment effect after baseline correction. Nevertheless, this finding should be interpreted as exploratory because the trial was powered only for the noninferiority comparison of the primary end point. No formal MCID has been established for forearm rotation ROM in the published literature. Normal forearm rotation encompasses a total arc of approximately 160°, meaning that a 13° difference represents roughly 8% of normal motion. Whether this magnitude translates into a clinically meaningful functional benefit for patients remains uncertain. In addition, as discussed above, baseline forearm rotation differed between groups, and some portion of the observed difference may reflect residual imbalance despite statistical adjustment. Confirmatory studies specifically powered for this end point are needed to establish the clinical significance of this finding.

Recovery of muscle strength in the DT group was comparable to that in the CT group. This finding is supported by previous research in orthopedic rehabilitation. For example, following knee surgery, patients undergoing digital rehabilitation achieved strength gains indistinguishable from those of patients receiving clinic-based therapy [16]. Systematic reviews across various musculoskeletal conditions have likewise reported no significant differences in strength outcomes between digitally supervised exercise and CT when exercise protocols are comparable [16].

For patient-reported outcome measures, improvements in the DT group were comparable to those in the CT group. Previous mobile-based rehabilitation studies have consistently demonstrated equivalent functional outcome scores [6]. In a pilot study of elbow rehabilitation, both groups achieved significant functional gains, and the mobile-based rehabilitation group even showed a trend toward slightly better DASH scores early in recovery, although the differences were not statistically significant [6]. More broadly, a 2017 meta-analysis confirmed that mobile-based rehabilitation is as effective as usual care for improving physical function and reducing disability in musculoskeletal disorders [5]. In our trial, both groups also demonstrated improvements in quality of life, and prior evidence suggests that DT performs comparably to traditional training in this domain. Systematic reviews have reported that improvements in health-related quality of life are similar between DT and in-person care among patients with musculoskeletal conditions [17].

Cost-Effectiveness

Our short-term cost-consequence analysis found that DT was associated with lower overall costs compared with conventional outpatient clinic-based training. The cost savings were driven primarily by reductions in primary care visits, secondary care costs, medication expenditures, and transportation expenses. These findings are strongly supported by previous cost-effectiveness studies [18-20]. A recent meta-analysis of 5 economic evaluations estimated that mobile-based rehabilitation was, on average, approximately US $90 less expensive per patient than standard rehabilitation for musculoskeletal conditions [18]. Another analysis of a mobile-based rehabilitation system after total knee arthroplasty reported cost savings of more than US $2400 per patient compared with traditional therapy, alongside equivalent or superior outcomes [19]. Our trial's findings of lower costs favoring mobile-based rehabilitation are therefore consistent with prior evidence. Overall, DT systems appear to be cost-effective, providing similar health outcomes at lower or comparable costs relative to conventional in-person training.

Adherence, Satisfaction, and Safety

Patient adherence to the DT system in the DT arm of the trial was high, reflecting a common trend reported in DT studies. Research indicates that patients often attend DT sessions and perform prescribed exercises as reliably as, or more reliably than, those receiving conventional therapy. A 2024 systematic review of RCTs found that patients undergoing mobile-based rehabilitation demonstrated approximately 9% higher exercise adherence on average compared with those receiving in-person care (95% CI 2%-16% higher) [21]. Attendance at scheduled sessions was also slightly better with DT systems, by as much as 8%, although the difference was not always statistically significant [21]. Likely contributing factors include convenience and flexibility, as patients can complete sessions from home, making adherence to the rehabilitation regimen easier. Patient satisfaction with the DT system in this trial was also very positive and may even have exceeded satisfaction with in-hospital therapy. This finding is consistent with previous studies reporting high acceptability of DT systems. In the elbow fracture pilot study, participants in the tele-rehabilitation group reported higher satisfaction with their care despite similar

XSL•FO
RenderX

clinical outcomes [6]. Patients often value the comfort and individualized attention made possible through video sessions, and the therapeutic alliance can be maintained remotely. Importantly, studies have shown that when DT systems are delivered synchronously through real-time video, the quality of patient-therapist interaction is preserved, resulting in satisfaction ratings comparable to those of in-person visits [17]. Overall, the high adherence and satisfaction observed in the DT group are well supported by previous research, indicating that mobile-based rehabilitation is generally well accepted and achieves patient engagement comparable to that of conventional rehabilitation.

Regarding safety, the DT system used in this trial proved to be safe, with no increase in adverse events compared with CT. This finding is strongly consistent with existing evidence indicating that DT systems do not pose an increased risk to patients. A recent scoping review encompassing more than 80 mobile-based training studies across diverse populations found that the incidence of adverse events during DT sessions was extremely low, approximately 0.3% across 84,534 total sessions [22]. Importantly, most reported incidents were determined to be unrelated to the DT intervention itself [22].

Clinical Implications

Within the context of this single-center trial, DT may be considered for integration into postsurgical rehabilitation pathways. Our study demonstrates that a DT system in which patients receive individually tailored PEST yields improvements in elbow ROM, strength, function, and quality of life comparable to those achieved with traditional outpatient clinic-based training, while also reducing overall costs. These findings support its use as a viable alternative, particularly for patients with limited access to specialized care or transportation challenges. The lower cost burden observed in our study, driven by reductions in primary care, secondary care, medication, and transportation expenses, further underscores its potential to reduce financial pressures on both patients and health care systems. In addition, the high levels of patient adherence and satisfaction, together with comparable safety profiles between the DT system and outpatient clinic-based training, suggest that digital interventions do not compromise treatment quality or clinical outcomes. Given these findings, clinicians may consider incorporating DT systems into standard postoperative rehabilitation protocols to enhance accessibility and optimize resource utilization. If validated in multicenter settings, DT systems may have broader applicability in musculoskeletal rehabilitation, although the generalizability of our single-center findings to other populations and health care systems requires further investigation [23-25]. Future efforts should focus on addressing barriers such as technology access and reimbursement policies to fully realize the benefits of DT systems in routine clinical practice.

Strengths and Limitations

This study's strengths include a rigorous randomized controlled design with an adequate sample size and comprehensive outcome measures encompassing clinical, functional, and economic end points. The noninferiority design allowed us to demonstrate that the DT system, in which patients receive individually tailored PEST, is as effective as conventional outpatient clinic-based training, while the inclusion of cost-effectiveness, adherence, and safety assessments provides a holistic view of the intervention's impact. Moreover, high patient adherence and satisfaction underscore the feasibility and acceptability of the DT system in a postsurgical setting. However, several limitations must be acknowledged. First, our trial was originally powered for the primary end point of elbow ROM noninferiority, with 73 patients per group expected to provide 80% power (1-sided α=.025) to detect a 10° margin. The final sample (106 and 102 patients per group) exceeded this target, yielding approximately 90% power for the noninferiority hypothesis. This ample sample size strengthens confidence in the primary finding that DT is not inferior to CT for ROM improvement. However, as is typical in RCTs, our study was not specifically powered to detect small between-group differences in secondary outcomes. We examined multiple secondary end points (functional scores, strength measures, etc) and applied a Bonferroni correction to control type I error. This stringent adjustment, while guarding against false positives, inevitably lowers statistical power for each secondary comparison. Consequently, the chance of type II error is increased for secondary outcomes under our analysis strategy. In practical terms, our negative secondary results should be interpreted with caution: the trial may not have been large enough to uncover subtle advantages of either intervention in secondary domains. For example, given our final sample, the detectable difference in DASH score with 80% power at an adjusted α≈.007 is several points; smaller improvements could have gone undetected. We opted for a conservative approach to multiple comparisons to prioritize confidence in any positive findings, but we acknowledge that this comes at the expense of sensitivity. We have now supplemented the original Bonferroni correction with the Holm step-down procedure, which is uniformly more powerful for positively correlated end points, as recommended by both the Food and Drug Administration and the European Medicines Agency guidance. Both methods yielded identical conclusions regarding the secondary outcomes. Future studies with larger samples or without such stringent correction might further explore secondary benefits of DT. In our data, no secondary outcome showed a trend large enough to approach the adjusted significance threshold, which provides some reassurance that we did not miss any clinically meaningful effects. Nonetheless, the possibility of type II error for modest effects, particularly in patient-reported outcomes, remains a limitation of this study.

Second, a notable limitation is that we relied on self-reported measures for exercise adherence and patient satisfaction. Such self-reports are susceptible to social desirability bias, whereby participants may portray themselves in a favorable light. Research using concealed accelerometers has shown that self-reported exercise diaries can overestimate adherence by approximately 20% compared with objectively measured activity. In our trial, patients might have overestimated their adherence to the exercise regimen or overstated their satisfaction. Although the Joymotion app recorded exercise completion rates and session frequency, thereby providing a partial digital log, we did not use dedicated motion sensors or video verification to confirm exercise quality or dosage. The

XSL•FO
RenderX

very high adherence scores should therefore be interpreted with caution. Future studies should incorporate wearable sensors or automated movement-quality assessment to complement patient reports. Consequently, although both groups reported excellent adherence (averaging 8.5/10) and satisfaction (9.5/10), these values should be interpreted with caution. It is possible that actual adherence was lower than reported, given evidence that individuals often overreport healthy behaviors. This limitation underscores the need for future studies to include objective monitoring of exercise completion to complement patient reports. Despite this caveat, the absence of between-group differences in self-reported adherence suggests that any social desirability bias was likely nondifferential between DT and CT and, thus, probably does not compromise the comparison of the 2 modalities, although absolute adherence levels may have been overestimated.

Third, the single-center design may limit the generalizability of our findings to other health care settings. Education level differed significantly between groups, and because education may correlate with digital literacy and comfort with technology, this imbalance could have influenced adherence and outcomes in the DT arm. A sensitivity analysis adjusting for education confirmed that the noninferiority conclusion was unchanged, but we cannot fully exclude residual confounding. The inability to blind participants and therapists introduces potential performance and expectation bias. This is particularly relevant for patient-reported outcomes such as ASES, DASH, and EQ-5D-5L scores, for which awareness of group assignment may have influenced response patterns. The small between-group differences observed in these measures may partly reflect equivalent patient expectations rather than purely equivalent treatment effects. Future trials should consider strategies to mitigate expectation bias, such as enhanced blinding of outcome assessors and the use of sham digital interfaces.

Conclusions

This study showed that individually tailored PEST delivered via a DT system was statistically noninferior to conventional outpatient clinic-based training following arthroscopic release for posttraumatic elbow stiffness. Although noninferiority does not necessarily imply clinical interchangeability, the digital approach may represent a viable option for patients facing barriers to traditional in-person rehabilitation, provided that clinicians consider individual patient factors when selecting a rehabilitation modality.

Acknowledgments

The authors acknowledge the Shanghai Medmotion Clinic for providing the necessary equipment for the study. We sincerely thank all the patients who participated in this study and the clinical staff who supported the trial. The authors also confirm that no generative artificial intelligence tools were used in the writing of this manuscript.

Data Availability

The deidentified data sets generated and analyzed during this study are available from the corresponding author on reasonable request. Requests should be directed to the corresponding author (SL) following online publication. All participants' baseline characteristics and follow-up data are available in deidentified format.

Authors' Contributions

Conceptualization: SL (lead), YS (equal)
Data curation: YW, LH
Formal analysis: RW (lead), ZX (equal)
Investigation: YW (lead), LH (equal), YM, SY, JD
Methodology: SL, YS
Project administration: SL
Resources: JD
Supervision: SL (lead), YS (equal), YM (supporting), SY (supporting)
Validation: SL, YS
Visualization: RW, ZX
Writing – original draft: YW (lead), LH (equal)
Writing – review & editing: SL, YS, YW, LH, RW, ZX, YM, SY, JD

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH (V 1.6.1).
[PDF File (Adobe PDF File), 1270 KB - mhealth_v14i1e87459_app1.pdf]

Multimedia Appendix 2

XSL•FO
RenderX

The standardized 12-week physiotherapeutic elbow-specific training (PEST) protocol.
[DOCX File , 33 KB - mhealth_v14i1e87459_app2.docx]

Multimedia Appendix 3
Unadjusted changes in outcomes for the digital training (DT) and conventional training (CT) groups at weeks 4, 12 and 24 after the surgery (per-protocol population).
[DOCX File , 20 KB - mhealth_v14i1e87459_app3.docx]

Multimedia Appendix 4
Adjusted effectiveness estimates from linear mixed effects models (per-protocol population).
[DOCX File , 20 KB - mhealth_v14i1e87459_app4.docx]

Multimedia Appendix 5
Sensitivity analysis of secondary outcomes at 12 weeks using Bonferroni and Holm step-down adjustments for multiple comparisons (intention-to-treat population).
[DOCX File , 21 KB - mhealth_v14i1e87459_app5.docx]

References

1. Zheng J, Hou M, Liu L, Wang X. Knowledge structure and emerging trends of telerehabilitation in recent 20 years: a bibliometric analysis. Front Public Health 2022 Jun 20;10:904855 [FREE Full text] [doi: 10.3389/fpubh.2022.904855] [Medline: 35795695]
2. Zhou Y, Lyu Y, Wang Q, Ma Y, Huang L, Zhang X. Mobile-based in-home telerehabilitation compared with in-hospital face-to-face rehabilitation for elderly patients after total hip arthroplasty in China's level 1 trauma center: a noninferiority randomized controlled trial. Front Surg 2024 Jan 14;11:1536579 [FREE Full text] [doi: 10.3389/fsurg.2024.1536579] [Medline: 39877311]
3. Keijsers R, Zwerus EL, van Lith DRM, Koenraadt KLM, Goossens P, The B, et al. Validity and reliability of elbow range of motion measurements using digital photographs, movies, and a goniometry smartphone application. J Sports Med (Hindawi Publ Corp) 2018 Nov 29;2018:7906875-7906877. [doi: 10.1155/2018/7906875] [Medline: 30631776]
4. Tousignant M, Moffet H, Boissy P, Corriveau H, Cabana F, Marquis F. A randomized controlled trial of home telerehabilitation for post-knee arthroplasty. J Telemed Telecare 2011 Mar 11;17(4):195-198. [doi: 10.1258/jtt.2010.100602] [Medline: 21398389]
5. Cottrell MA, Galea OA, O'Leary SP, Hill AJ, Russell TG. Real-time telerehabilitation for the treatment of musculoskeletal conditions is effective and comparable to standard practice: a systematic review and meta-analysis. Clin Rehabil 2017 May 02;31(5):625-638. [doi: 10.1177/0269215516645148] [Medline: 27141087]
6. Mayer N, Portnoy S, Palti R, Levanon Y. The efficacy of tele-rehabilitation program for improving upper limb function among adults following elbow fractures: a pilot study. Applied Sciences 2021 Feb 14;11(4):1708. [doi: 10.3390/app11041708]
7. Prvu Bettger J, Green CL, Holmes DN, Chokshi A, Mather RC, Hoch BT, et al. Effects of virtual exercise rehabilitation in-home therapy compared with traditional care after total knee arthroplasty. The Journal of Bone and Joint Surgery 2019 Nov 18;102(2):101-109. [doi: 10.2106/jbjs.19.00695]
8. Zhang D, Nazarian A, Rodriguez EK. Post-traumatic elbow stiffness: pathogenesis and current treatments. Shoulder Elbow 2020 Feb 08;12(1):38-45 [FREE Full text] [doi: 10.1177/1758573218793903] [Medline: 32010232]
9. Schreiner AJ, Schweikardt N, Gühring D, Ahrend M, Döbele S, Ahmad SS, et al. Arthroscopic arthrolysis leads to improved range of motion and health-related quality of life in post-traumatic elbow stiffness. J Shoulder Elbow Surg 2020 Aug;29(8):1538-1547. [doi: 10.1016/j.jse.2020.01.099] [Medline: 32381474]
10. Sun Z, Li J, Luo G, Wang F, Hu Y, Fan C. What constitutes a clinically important change in Mayo Elbow Performance Index and range of movement after open elbow arthrolysis? The Bone & Joint Journal 2021 Feb 01;103-B(2):366-372. [doi: 10.1302/0301-620x.103b2.bjj-2020-0259.r3]
11. O'Driscoll S, Lievano J, Morrey M, Sanchez-Sotelo J, Shukla D, Olson T. Prospective randomized trial of continuous passive motion versus physical therapy after arthroscopic release of elbow contracture. J Bone Joint Surg Am 2022;104(5):430-440. [doi: 10.2106/jbjs.21.00685]
12. Hudak PL, Amadio PC, Bombardier C, Beaton D, Cole D, Davis A, et al. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder, and head). Am J Ind Med 1996 Jun;29(6):602-608. [doi: 10.1002/(sici)1097-0274(199606)29:6<602::aid-ajim4>3.0.co;2-l]
13. Luo N, Liu G, Li M, Guan H, Jin X, Rand-Hendriksen K. Estimating an EQ-5D-5L value set for China. Value Health 2017 Apr;20(4):662-669 [FREE Full text] [doi: 10.1016/j.jval.2016.11.016] [Medline: 28408009]
14. Zdravkovic V, Jost B. Rounded data have a high potential for false comparative statistical results as shown with elbow arc of motion. J Shoulder Elbow Surg 2018 Feb;27(2):276-281. [doi: 10.1016/j.jse.2017.10.033] [Medline: 29332664]

XSL•FO
RenderX

15. Circular of the Shanghai Municipal Bureau of Human Resources and Social Security on the average wages of employed persons in full-caliber urban units in the city in 2022. Shanghai Municipal Bureau. URL: https://rsj.sh.gov.cn/ [accessed 2025-05-10]
16. Wang Q, Lee RLT, Hunter S, Chan SW. The effectiveness of internet-based telerehabilitation among patients after total joint arthroplasty: an integrative review. Int J Nurs Stud 2021 Mar;115:103845. [doi: 10.1016/j.ijnurstu.2020.103845] [Medline: 33360248]
17. Baroni MP, Jacob MFA, Rios WR, Fandim JV, Fernandes LG, Chaves PI, et al. The state of the art in telerehabilitation for musculoskeletal conditions. Arch Physiother 2023 Jan 04;13(1):1 [FREE Full text] [doi: 10.1186/s40945-022-00155-0] [Medline: 36597130]
18. Tousignant M, Moffet H, Nadeau S, Mérette C, Boissy P, Corriveau H, Belzile, et al. Cost analysis of in-home telerehabilitation for post-knee arthroplasty. J Med Internet Res 2015 Mar 31;17(3):e83 [FREE Full text] [doi: 10.2196/jmir.3844] [Medline: 25840501]
19. Summers SH, Gnecco T, Slotkin EM, Law TY, Nunley RM. Significant cost savings and improved early clinical outcomes in Medicare patients utilizing a clinician-controlled telerehabilitation system following total knee arthroplasty. J Arthroplasty 2024 Aug;39(8S1):S137-S142. [doi: 10.1016/j.arth.2024.02.040] [Medline: 38401615]
20. Molina-Garcia P, Mora-Traverso M, Prieto-Moreno R, Díaz-Vásquez A, Antony B, Ariza-Vega P. Effectiveness and cost-effectiveness of telerehabilitation for musculoskeletal disorders: a systematic review and meta-analysis. Ann Phys Rehabil Med 2024 Feb;67(1):101791 [FREE Full text] [doi: 10.1016/j.rehab.2023.101791] [Medline: 38128150]
21. Simmich J, Ross MH, Russell T. Real-time video telerehabilitation shows comparable satisfaction and similar or better attendance and adherence compared with in-person physiotherapy: a systematic review. J Physiother 2024 Jul;70(3):181-192 [FREE Full text] [doi: 10.1016/j.jphys.2024.06.001] [Medline: 38879432]
22. Yau T, Chan J, McIntyre M, Bhogal D, Andreoli A, Leochico CFD, et al. Adverse events associated with the delivery of telerehabilitation across rehabilitation populations: a scoping review. PLoS One 2024 Nov 19;19(11):e0313440 [FREE Full text] [doi: 10.1371/journal.pone.0313440] [Medline: 39561143]
23. Moffet H, Tousignant M, Nadeau S, Mérette C, Boissy P, Corriveau H, Belzile, et al. In-home telerehabilitation compared with face-to-face rehabilitation after total knee arthroplasty. The Journal of Bone and Joint Surgery-American Volume 2015;97(14):1129-1141. [doi: 10.2106/jbjs.n.01066]
24. Janela D, Costa F, Molinos M, Moulder RG, Lains J, Bento V, et al. Digital rehabilitation for elbow pain musculoskeletal conditions: a prospective longitudinal cohort study. Int J Environ Res Public Health 2022 Jul 27;19(15):9198 [FREE Full text] [doi: 10.3390/ijerph19159198] [Medline: 35954555]
25. Bargeri S, Castellini G, Vitale JA, Guida S, Banfi G, Gianola S, et al. Effectiveness of telemedicine for musculoskeletal disorders: umbrella review. J Med Internet Res 2024 Feb 02;26:e50090 [FREE Full text] [doi: 10.2196/50090] [Medline: 38306156]

Abbreviations

ASES: American Shoulder and Elbow Surgeons
BTE: Baltimore Therapeutic Equipment
CONSORT: Consolidated Standards of Reporting Trials
CONSORT-NI: Consolidated Standards of Reporting Trials extension for noninferiority and equivalence trials
CT: conventional training
DASH: Disabilities of the Arm, Shoulder, and Hand
DT: digital training
ICER: incremental cost-effectiveness ratio
MCID: minimal clinically important difference
PEST: physiotherapeutic elbow-specific training
PT: physiotherapist
ROM: range of motion

XSL•FO
RenderX

Edited by L Buis; submitted 09.Nov.2025; peer-reviewed by Y Yao, W Li; comments to author 10.Mar.2026; revised version received 16.Mar.2026; accepted 13.May.2026; published 09.Jun.2026.

Please cite as:
Wang Y, Huang L, Wang R, Xu Z, Ma Y, Yu S, Ding J, Shen Y, Lu S
Digital Physiotherapeutic Elbow-Specific Training System for Patients After Arthroscopic Release of Elbow Contracture: Noninferiority Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e87459
URL: https://mhealth.jmir.org/2026/1/e87459
doi:10.2196/87459
PMID:

©Yanmao Wang, Lihua Huang, Ruixin Wang, Zhihao Xu, Yanhong Ma, Shiyang Yu, Jian Ding, Yun Shen, Shengdi Lu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 09.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effects of Telerehabilitation Based on Motion Recognition Technology on Exercise Endurance of Patients With Non–Small Cell Lung Cancer After Surgery: Single-Center, Prospective, Open-Label, Randomized Controlled Trial

Ling Chen[1,2,3,4*], MS; Xiaoyi Xu[1,2,3,4*], MS; Xiaoliang Yang[1,2,3,4*], MS; Siying Long[1,2,3,4*], MS; Na Chen[1,2,3,4], PhD; Zifan Li[1,2,3,4], MS; Tingting Yang[1,2,3,4], MS; Qianru Li[1,2,3,4], MS; Meng Li[1,2,3,4], MS; Hai Zhong[1,2,3,4*], PhD; Guozhi Huang[1,2,3,4*], PhD; Qing Zeng[1,2,3,4*], PhD

[1]Center of Rehabilitation Medicine, Zhujiang Hospital, Southern Medical University, Guangzhou, Guangdong Province, China

[2]School of Rehabilitation Sciences, Southern Medical University, Guangzhou, Guangdong Province, China

[3]GuangDong Engineering Technology Research Center of Brain Function Assessment and Neuroregulation Rehabilitation, Guangzhou, Guangdong Province, China

[4]Institute of Exercise and Rehabilitation Science, Zhujiang Hospital, Southern Medical University, Guangzhou, Guangdong Province, China

[*]these authors contributed equally

Corresponding Author:
Qing Zeng, PhD
Center of Rehabilitation Medicine, Zhujiang Hospital, Southern Medical University
253 Gongye Avenue, Haizhu District
Guangzhou, Guangdong Province, 510280
China
Phone: 86 15989189468
Email: zengqingyang203@126.com

Abstract

Background: Non–small cell lung cancer (NSCLC) accounts for approximately 85% of primary pulmonary neoplasms. Complete surgical removal remains the cornerstone of curative therapy, yet it frequently diminishes residual lung function and exercise tolerance. Structured, center-based rehabilitation hastens physiological recovery, but conventional schemes rarely deliver continuous, patient-specific monitoring. Remote, digitally delivered exercise overcomes logistical obstacles; however, the lack of real-time quality assurance curtails effectiveness. Wearable motion capture platforms that provide millimeter-precise kinematic data and instantaneous biomechanical feedback can close this supervisory void by confirming movement accuracy and issuing immediate corrective prompts. Whether this technologically augmented telerehabilitation yields clinically relevant improvements in postoperative exercise capacity after NSCLC resection remains inadequately established.

Objective: This study aims to find out whether motion capture–enabled telerehabilitation can enhance exercise endurance and functional recovery following NSCLC resection, thereby filling a critical void in postoperative care pathways.

Methods: We performed a single-center, parallel-arm randomized trial in individuals who had completed curative lung resection for NSCLC and satisfied every enrollment criterion. Following randomization, eligible patients were assigned either to a technology-enhanced telerehabilitation protocol incorporating live motion sensing or to a WeChat-based regimen without that feedback layer. Each program spanned 4 weeks. Assessments—comprising exercise endurance, spirometric variables, health-related quality of life, and daily physical activity—were collected at discharge and repeated 4 weeks after intervention.

Results: No exercise-related adverse events occurred during the 4-week intervention. Compared with conventional video-guided training, the motion recognition group demonstrated significantly greater improvements in exercise tolerance, pulmonary function, selected functional mobility outcomes, and quality of life. The intervention group achieved a greater increase in 6-minute walk distance (mean difference 32 meters, 95% CI 5.40-58.60; *P*=.02). Significant between-group differences were also observed in forced vital capacity (mean difference 476.4 mL; *P*=.02), forced expiratory volume in 1 second (mean difference 346.0 mL; *P*=.04), FEV_1/FVC ratio (mean difference 12.8%; *P*=.03), and peak expiratory flow (mean difference 70.8 L/minute; *P*=.002). For functional mobility, the intervention group showed superior improvement in the Timed Up and Go test (mean difference −1.63 seconds; *P*=.04), whereas no significant between-group differences were found in other mobility measures. Quality of life

XSL•FO
RenderX

outcomes favored the motion recognition group, with greater improvements in physical well-being, functional well-being, additional concerns–lung subscale, and total Functional Assessment of Cancer Therapy–Lung score (mean difference 13.00; *P*<.001). Program adherence was higher in the intervention group (72% vs 40%; *P*=.02).

Conclusions: Four-week motion capture–guided telerehabilitation yielded clinically meaningful gains in aerobic endurance, spirometric indices, ambulatory capacity, and global health–related quality of life for patients recovering from NSCLC surgery, underscoring its usefulness as a safe and effective remote care strategy.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2500113139; https://www.chictr.org.cn/showprojEN.html?proj=270991

(JMIR Mhealth Uhealth 2026;14:e82447) doi:10.2196/82447

KEYWORDS

non–small cell lung cancer; artificial intelligence; motion recognition; telerehabilitation; exercise endurance

Introduction

Lung cancer continues to rank among the leading causes of malignant disease worldwide, exerting a disproportionate share of global oncological morbidity [1]. Global cancer statistics for 2022 recorded roughly 2.48 million incident cases—about 1 in 8 newly diagnosed tumors [2]. Non–small cell lung cancer (NSCLC) constitutes approximately 85% of this total [3]. Curative-intent surgical resection remains the primary therapeutic modality; yet, a substantial proportion of recipients subsequently experience pronounced reductions in aerobic capacity and ventilatory efficiency, translating into markedly lower health-related quality of life [4]. Exercise-based rehabilitation has repeatedly been demonstrated to augment cardiorespiratory fitness, elevate peak oxygen uptake, and enhance both muscular strength and endurance [5]. Physical training is therefore regarded as a cornerstone of postoperative recovery after NSCLC resection. Randomized evidence indicates that regimented exercise improves exertional tolerance, fortifies respiratory musculature [6], and confers psychological benefit [7,8]. Traditional, center-based programs—usually conducted in hospitals or specialized rehabilitation units—continue to confront entrenched barriers. Uneven geographic distribution of trained therapists leaves numerous primary care settings without access to specialist services, impeding timely intervention. Logistical burdens—lengthy travel, inflexible scheduling, and direct and indirect costs—further erode participation, while repeated clinic visits impose additional physical and psychosocial strain on patients and caregivers.

Digital telerehabilitation—encompassing videoconferencing, smartphone apps, and immersive virtual reality—overcomes spatial and temporal barriers to specialist care. Recent studies have suggested that eHealth-supported pulmonary rehabilitation may improve musculoskeletal symptoms and quality of life in patients with post–COVID-19 condition [9]. Furthermore, structured telemedicine-based pulmonary rehabilitation protocols have been developed for individuals with post–COVID-19 condition and persistent respiratory symptoms, reflecting ongoing efforts to establish remote rehabilitation approaches.[10]. In addition, recent systematic reviews and meta-analyses have confirmed that pulmonary rehabilitation significantly improves exercise capacity, pulmonary function, and quality of life in individuals with long COVID-19 syndrome, further reinforcing the clinical value of structured respiratory rehabilitation programs [11]. Evidence also links these remote interventions to prolonged survival, improved quality of life, and heightened self-efficacy in lung-cancer survivors [12]. Park et al [12] deployed the Smart Aftercare App to deliver postoperative exercise guidance and reported significantly fewer outpatient encounters among active users than among nonusers. Despite such benefits, conventional telerehabilitation platforms typically depend on asynchronous demonstration clips or synchronous verbal coaching, neither of which can supply real-time biomechanical feedback or quantitative motion assessment. This deficit jeopardizes movement fidelity and complicates the standardization of exercise execution.

Action recognition platforms integrate computer vision, wearable sensors, and machine learning classifiers to deliver millimeter-level appraisal of human kinematics. Over the past decade, this convergence has fueled growing interest in digitally mediated rehabilitation. He et al [13] reported that exercise regimens augmented by movement capture analytics yield gains in lean mass, functional capacity, and health-related quality of life that rival both center-based therapy and conventional tele-exercise. Chae and coworkers [14] created an artificial intelligence (AI)–driven fitness application that issues instantaneous postural cues during exercise. The participants who received algorithmic feedback improved their action accuracy by more than 32% and demonstrated higher execution consistency than those who followed the standard video instructions [14]. Extending this paradigm, our group previously conducted a randomized controlled trial in adults with chronic nonspecific low back pain. After allocation to either an AI-guided, real-time feedback group or a traditional video-only group, the AI cohort exhibited a pronounced reduction in pain intensity and a concomitant hypertrophic response in the deep core musculature within 4 weeks [15].

Based on this, this study aims to explore the application value of motion recognition technology in remote rehabilitation after NSCLC lung cancer surgery. By providing objective and quantifiable motion data, it enables precise evaluation of patients' rehabilitation progress, thereby filling the gap in existing research, promoting innovative development of postoperative rehabilitation models for lung cancer, and delivering higher-quality rehabilitation services to patients.

XSL•FO
RenderX

Methods

Study Design

This was a single-center, prospective, open-label, randomized controlled trial. We used an offline approach to screen adult patients who underwent radical lung cancer surgery in the Department of Thoracic Surgery at Zhujiang Hospital, Southern Medical University, between June 2024 and January 2025, and the study protocol adhered to the CONSORT-EHEALTH (V 1.6; Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (Multimedia Appendix 1). The study was prospectively registered in the Chinese Clinical Trial Registry (ChiCTR2500113139).

Sample Size Calculation

The sample size was determined based on the primary outcome—6-minute walk distance (6MWD). Previous studies have demonstrated that telerehabilitation can significantly improve 6MWD in patients following lung cancer surgery, and the minimal clinically important difference for 6MWD in this population has been reported to be 50 meters [16]. Accordingly, we hypothesized that the motion capture telerehabilitation intervention would yield a clinically meaningful improvement of 50 meters in 6MWD compared with the control group. Assuming a 2-sided α value of .05 and a power of 80%, the required sample size was calculated to be 23 participants per group. To account for an anticipated dropout rate of 20%, the sample size was inflated to 29 participants per group. Therefore, a total of 58 participants were required for this study.

Ethical Considerations

The clinical study was conducted in accordance with the Declaration of Helsinki and China regulations on clinical research and has been approved by the medical ethics committee. All participants signed informed consent forms. Prior to the formal initiation of this study, it was approved by the Ethics Committee of Zhujiang Hospital, Southern Medical University (approval 2024-KY-327-01, Medical Research Registration and Filing Information System Registration MR-44-24-052709). The study has been registered in the Chinese Clinical Trial Registry (ChiCTR2500113139). The study was conducted by trained investigators, who were responsible for patient recruitment, data collection, and test administration. The data collected included patients' basic information, clinical data, exercise tolerance, pulmonary function, quality of life scales, mobility, treatment adherence, and adverse event incidence. For patients capable of completing the paper-based questionnaire independently, the investigators provided standardized questionnaires and guided them through the completion process. For patients unable to complete the questionnaire independently due to physical conditions or other reasons, the investigators conducted oral inquiries and recorded responses accordingly. During the recording process, the investigators verified information with patients to ensure data accuracy and completeness, and completed the questionnaire after confirming no errors. Throughout the research process, the investigators strictly adhered to ethical guidelines, safeguarding patient privacy and data security. All collected data were processed anonymously and used solely for research purposes, with no disclosure to third parties without patient consent. Participants received complimentary health education resources and were granted access to the AI-driven telerehabilitation platform throughout the intervention phase, while no monetary incentives were offered to participants for taking part in the study.

Participant Selection and Recruitment

The study selected patients with NSCLC who underwent thoracic surgery and were hospitalized at Zhujiang Hospital of Southern Medical University from June 2024 to January 2025 as the research participants. A total of 58 eligible patients were enrolled. The research team provided comprehensive and detailed explanations to all enrolled patients and their families, covering the purpose, specific content, and significance of the study.

Eligibility Standards

Inclusion criteria included (1) histopathologic diagnosis of NSCLC per National Comprehensive Cancer Network Guidelines 2024.V4 [17], (2) age 18-75 years, (3) cognitive competence adequate for assessment and exercise adherence, and (4) completed radical resection (lobectomy, segmentectomy, or wedge) with stable postoperative vital signs. Exclusion criteria included (1) clinically significant impairment of cardiac, hepatic, or renal function precluding exercise; (2) uncontrolled postoperative complications such as pneumonia, atelectasis, pleural effusion, or respiratory failure; (3) concurrent malignancy likely to influence recovery; (4) enrollment in another interventional study within the preceding 90 days; (5) enrollment in another interventional study within the preceding 90 days; and (6) inability to use a smartphone or absence of reliable internet connectivity required for motion capture telerehabilitation.

Randomization and Blinding

Randomization was conducted using a computer-generated random number sequence in a 1:1 ratio. Simple randomization without blocking or stratification was applied. The allocation sequence was generated by an independent researcher (QZ) and concealed using sequentially numbered, opaque, and sealed envelopes to ensure allocation concealment. All researchers involved in patient recruitment or data collection will not be involved in the generation of the random sequence and the preparation and production of the random allocation concealment envelopes. The primary and all secondary outcome measures were collected offline through face-to-face interviews.

Interventions

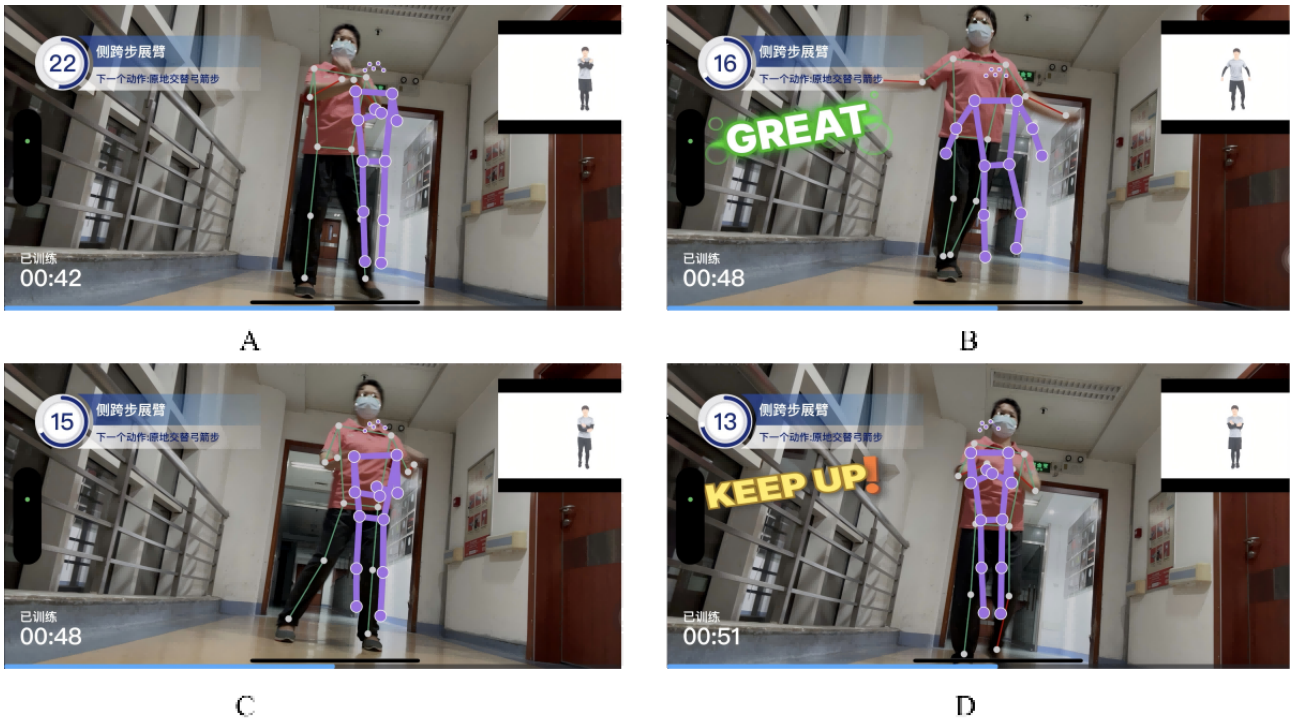
Telerehabilitation App Incorporating Action Recognition Technology

The intervention group trained via the Rehabilitation Exercise Prescription app, whose core engine is a skeleton-driven motion recognition module. The software perpetually maps 3D coordinates of anatomical landmarks and then superimposes a chromatic skeletal overlay on the live video feed to log every trajectory. As depicted in Figure 1A-D, the user interface foregrounds the active exercise label together with a

XSL•FO
RenderX

decrementing timer; preparatory cues smooth transitions between successive movements.

Figure 1. Training diagram of experimental group. (A) The system demonstrates the “Running in Place” exercise, displaying real-time skeletal motion tracking, exercise guidance, and countdown timing. (B) The system demonstrates the “Side Steps with Arm Extension” exercise with posture recognition and movement feedback during training. (C) The system demonstrates the “Alternating Lunges in Place” exercise while continuously tracking joint trajectories and exercise performance. (D) The system demonstrates the “Chest Expansions With Heel Touches” exercise with synchronized skeletal overlay and live exercise instructions.



Instant textual feedback—both corrective and motivational—is pushed immediately after each repetition to sustain engagement. High fidelity is achieved through deep learning classifiers that pit the participant’s kinematic signature against a gold-standard template, flagging deviations and confirming when performance satisfies therapeutic thresholds.

Remote Rehabilitation Without Action Recognition

Members of the control group followed a home exercise protocol hosted on a WeChat mini program (Figure 2). The module provided only asynchronous video demonstrations; it lacked skeleton-tracking analytics, instantaneous correction, and any mechanism to log or verify completion.

XSL•FO
RenderX

Figure 2. Home page of WeChat mini program.



Assessment

All assessments and questionnaire surveys were conducted in person through face-to-face evaluation sessions. The assessment included the following:

1. *General data of patients*: sociodemographic variables: full name, age, sex, stature, body mass, BMI, education level, occupation, tobacco exposure, habitual physical activity patterns, medical insurance category, and habitual activity levels.
2. *Disease-specific metrics*: overall length of stay, postoperative inpatient days, adjuvant chemotherapy or radiotherapy use, histologic subtype, and tumor-node-metastasis stage according to the IASLC eighth edition guidelines and surgical approach.
3. *Exercise endurance*: 6-minute walk test (6MWT, Borg scale).

XSL•FO
RenderX

4. *Pulmonary function*: forced vital capacity (FVC), forced expiratory volume in 1 second (FEV_1), the FEV_1/FVC ratio, and peak expiratory flow (PEF).
5. *Quality of life indicators*: Functional Assessment of Cancer Therapy–Lung v4.0 (FACT-L) [18]: the core covers four dimensions [19]: physical well-being (PWB, 7 items)—energy, pain, and fatigue [20]; social/family well-being (SWB, 7 items)—relationships and perceived support [18]; emotional well-being (EWB, 6 items)—anxiety and depressive symptoms; functional well-being (FWB, 7 items)—daily activities and mobility [21].
6. *Measurement of functional mobility*: Five-times Sit-to-Stand Test (FTSST) [22], 30-second Chair-Stand Test (30-second CST) [23], Timed Up and Go (TUG) [24], Single-leg stance [25], and 2-minute step test [26].
7. *Exercise adherence*: individuals completing ≥3 sessions per week were deemed adherent; those completing <3 sessions per week were considered nonadherent.

The pulmonary function test graph and the 6MWT graph are shown in Multimedia Appendix 2.

Exercise Program Development and Implementation

Table 1 presents the structured components of the postoperative home rehabilitation program implemented in the experimental group. The intervention was organized according to specific time points, including discharge day, 4 weeks after discharge, and the follow-up assessment. It outlines the intervention items, detailed content, implementation methods, and prescribed frequency and duration.

Table 1. Contents of postoperative home rehabilitation of patients in the experimental group.

Intervention timing and item	Specific content	Implementation method	Frequency/duration
On the day of discharge			
Discharge assessment	• Assess exercise tolerance, pulmonary function, daily activity ability, and quality of life	In-hospital assessment	Single session, approximately 60 minutes
• From discharge to 4 weeks		Rehabilitation Exercise Prescription app (experimental)/WeChat mini-program (control)	3-5 times per week, 20-30 minutes per session
Aerobic training	• Stationary jogging • Jumping jacks • High-knee runs		
Strength training	• Side-leg raises • Body-weight squats • Step-ups • Prone planks		
Flexibility training	• Seated forward folds • Static bow stances		
From the day of discharge to 4 weeks after discharge			
Follow-up management	• A dedicated postdischarge support unit—consisting of thoracic surgeons, rehabilitation physicians, nurses, and therapists—monitored each participant through WeChat. The group sent session reminders, delivered immediate feedback, and addressed emerging questions. Outcome evaluations were scheduled 4 weeks (2 days) after discharge to gauge intervention efficacy.	WeChat + phone calls	2-3 times a week
4 weeks (2 days) after discharge			
Outcome assessment	• Assess all outcomes	Outpatient assessment	Single session, approximately 60 minutes

On the day of discharge, participants underwent an in-hospital discharge assessment to evaluate exercise tolerance, pulmonary function, activities of daily living, and quality of life. During the 4-week exercise after discharge, the intervention consisted of structured exercise training delivered through a digital platform, incorporating aerobic, strength, and flexibility exercises.

XSL•FO
RenderX

Follow-Up

Remote supervision and follow-up management were conducted by a multidisciplinary team via online communication tools such as the WeChat app and telephone contact to enhance adherence and provide timely feedback. At 4 weeks after discharge, outcome assessments were performed in the outpatient setting using the same evaluation domains as baseline.

Statistical Analyses

All analyses were performed in SPSS (version 26.0; IBM Corp). Normally distributed continuous data are presented as mean (SD); group differences and pre-to-post changes were examined with independent-samples and paired *t* tests, respectively. Non–normally distributed variables are summarized by medians; between-group contrasts were evaluated using the Mann-Whitney *U* test, and within-group shifts via Wilcoxon signed-rank tests. Categorical variables are described with counts and percentages, and baseline comparability was verified by chi-square analysis. The last available primary end point value was carried forward for withdrawals. Missing data mechanisms guided handling: missing completely at random observations were listwise-deleted, missing at random values were multiply imputed, and missing not at random scenarios were explored with sensitivity analyses using pattern mixture or selection models.

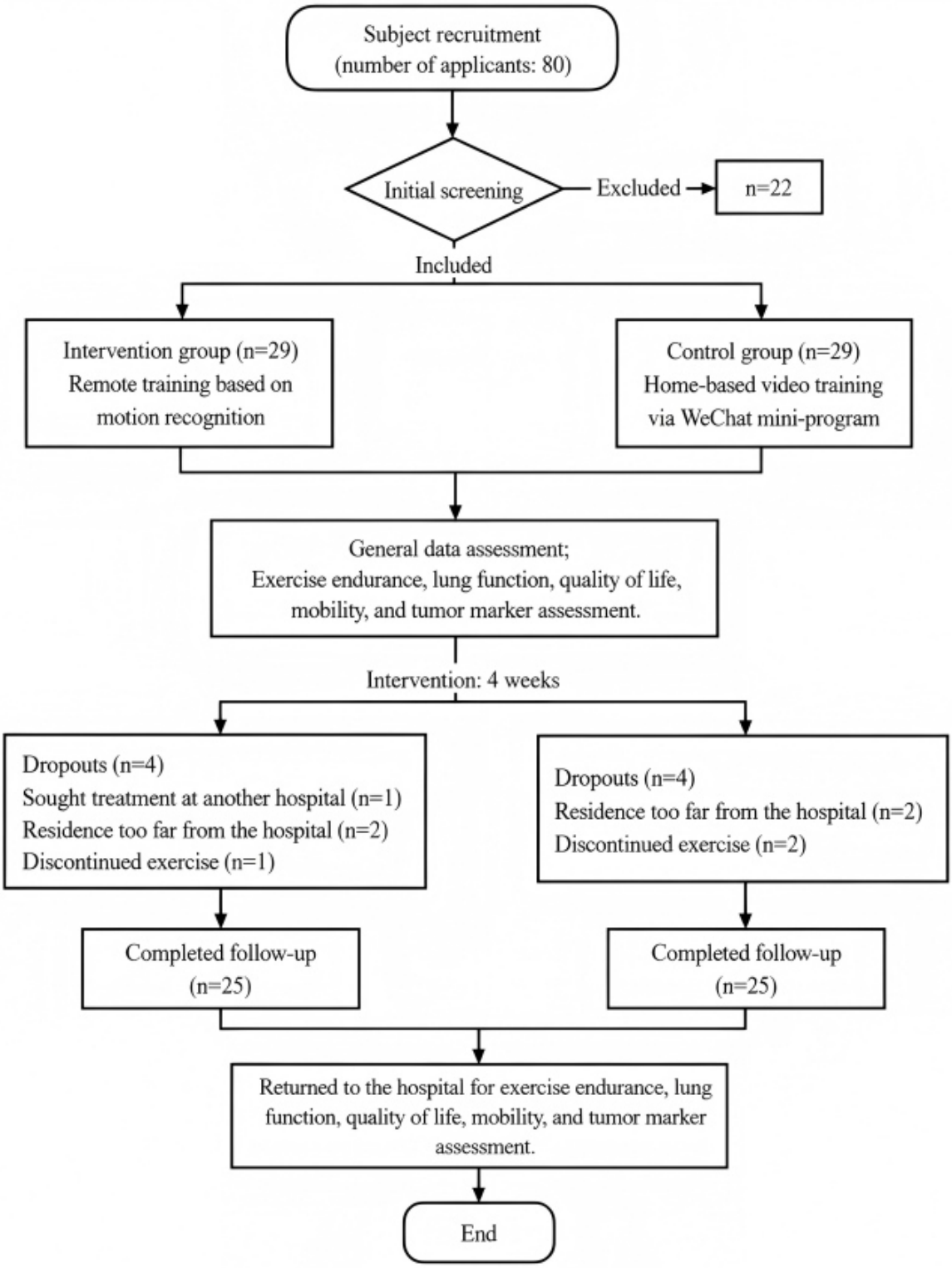
Results

Enrollment

A total of 80 applicants were initially recruited and underwent preliminary screening. Following the eligibility assessment, 22 individuals were excluded, and the remaining participants were included in the trial. Fifty-eight patients were randomized (29 per group). During follow-up, 4 participants in each group withdrew from the study. In the intervention group, reasons for dropout included seeking treatment at another hospital, residing too far from the hospital, and discontinuation of exercise. In the control group, dropouts were due to relocation far from the hospital or discontinuation of exercise. Ultimately, 25 participants in each group completed the follow-up assessment. Consequently, 50 participants (25 experimental and 25 control) provided complete data for the final analysis (Figure 3).

XSL•FO
RenderX

Figure 3. Flowchart of study process.



General Data of Patients

A total of 50 patients with lung cancer were enrolled and randomly allocated to the experimental group (n=25) and the control group (n=25). The mean age was 54.56 (SD 12.12) years in the experimental group and 56.96 (SD 11.02) years in the control group, with no statistically significant difference between groups (P=.62). In terms of sex distribution, 40% (10/25) of the participants in the experimental group and 48% (12/25) in the control group were male, and the difference was not statistically significant (P=.39). No significant differences were observed between the 2 groups with respect to height, body weight, or BMI (all P>.05). Similarly, baseline characteristics including educational attainment, occupational status, and exercise habits were evenly distributed between groups, with no significant between-group differences (all P>.05), suggesting comparability at baseline (Table 2).

XSL•FO
RenderX

Table 2. Comparison of demographic characteristics between the 2 groups.

Indicators	Experimental group (n=25)	Control group (n=25)	*t* test (*df*)	*P* value
Sex, n (%)			0.776 (1)	.39
Male	10 (40)	12 (48)		
Female	15 (60)	13 (52)		
Age (years), mean (SD)	54.56 (12.12)	56.96 (11.02)	–0.495 (48)	.62
Height (cm), mean (SD)	162.24 (7.03)	162.52 (5.83)	–0.153 (48)	.80
Body weight (kg), mean (SD)	60.04 (11.31)	58.5 (6.62)	–0.583 (48)	.56
BMI (kg/m^2), mean (SD)	22.77 (3.68)	22.19 (2.59)	0.646 (48)	.52
Degree of education, n (%)			5.219 (3)	.16
Primary school and below	9 (36)	5 (20)		
Junior high school	4 (16)	8 (32)		
Technical secondary school/high school	9 (36)	5 (20)		
Bachelor degree or above	3 (12)	7 (28)		
Occupations, n (%)			1.666 (6)	.95
Student	1 (4)	0		
Workers	4 (16)	4 (16)		
Farmer	2 (8)	3 (12)		
Staff member	2 (8)	2 (8)		
Self-employed persons	2 (8)	3 (12)		
Retirees	10 (40)	10 (40)		
Unemployment	4 (12)	3 (12)		
Habit of exercise, n (%)			0.125 (1)	.72
Have	6 (24)	4 (16)		
No	19 (76)	21 (84)		

There were no significant differences between the 2 groups in smoking history, medical insurance type, history of chemoradiotherapy, pathological type, surgical method, length of hospital stay distribution, or postoperative hospital stay duration (all $P \geq .05$). The majority of participants in both groups had stage I lung cancer, and all patients were pathologically confirmed at stage I, indicating homogeneity in disease severity. Adenocarcinoma was the predominant pathological type, accounting for 92% in the experimental group and 100% in the control group. Regarding surgical procedures, most patients underwent wedge resection or segmentectomy, with comparable distributions between groups. Postoperative hospital stay was similar between the 2 groups, with a mean duration of 6.08 (SD 0.60) days in the experimental group and 4.84 (SD 0.48) days in the control group ($P=.10$). Overall, baseline clinical and perioperative characteristics were balanced between groups (Table 3).

XSL•FO
RenderX

Table 3. Comparison of disease-related data between the 2 groups.

Indicators	Experimental group (n=25)	Control group (n=25)	Test value	*P* value
History of smoking, n (%)			0.087 (1)[a]	.76
Have	8 (32)	10 (40)		
No	17 (68)	15 (60)		
Types of medical insurance, n (%)			1.167 (3)[a]	.76
Out of pocket	3 (12)	1 (4)		
Local medical insurance	10 (48)	11 (40)		
Provincial nonlocal medical insurance	10 (40)	13 (52)		
Cross-provincial medical insurance	1 (4)	1 (4)		
History of chemoradiotherapy, n (%)			0.000 (2)[a]	>.99
Radiation therapy	0 (0)	0 (0)		
Chemotherapy	1 (4)	2 (8)		
No	24 (96)	23 (92)		
Types of pathology, n (%)			0.521 (1)[a]	.47
Adenocarcinoma	23 (92)	25 (100)		
Squamous cell carcinoma	2 (8)	0 (0)		
Pathological staging, n (%)			—[b]	—
Phase 1	25 (100)	25 (100)		
Phase 2	0 (0)	0 (0)		
Phase 3	0 (0)	0 (0)		
Surgical methods, n (%)			0.466 (2)[a]	.50
Lobectomy of lung	7 (28)	4 (16)		
Wedge resection/segmentectomy	18 (72)	21 (84)		
The average length of stay (days), n (%)			–1.556 (48)[c]	.12
≤5	3 (12)	5 (20)		
5-10	13 (52)	15 (60)		
≥10	9 (36)	5 (20)		
Postoperative hospital stay, mean (SD)	6.08 (0.60)	4.84 (0.48)	–1.650 (48)[c]	.10

[a]Chi-square test.

[b]Not applicable.

[c]*t* test.

Comparison of Outcome Index Results

Related Indexes of Exercise Endurance Compared Between 2 Groups

At baseline, 6MWD and Borg fatigue scores were comparable across groups (*P*>.05). After 4 weeks, both groups showed significant within-group improvements in the 6MWD. The experimental group increased from a mean of 399.00 (SD 67.92) to 507.00 (SD 62.23) meters (*T*=–7.40; *P*<.001). The control group also improved from 415.12 (SD 73.13) to 460.96 (SD 71.24) meters (*T*=–3.39; *P*<.001). The between-group comparison at 4 weeks revealed a statistically significant difference favoring the motion capture cohort (mean difference 32 meters, 95% CI 5.40- 58.60; *T*=2.43; *P*=.02). For perceived exertion, both groups demonstrated significant reductions in Borg scores following the intervention. The experimental group decreased from 2.12 (SD 0.93) to 0.96 (SD 0.87; *T*=3.63; *P*<.001), while the control group decreased from 2.32 (SD 0.80) to 1.36 (SD 0.94; *T*=3.67; *P*<.001). However, the between-group difference for the change in Borg-perceived exertion scores was not statistically significant (mean difference –0.40, 95% CI –0.88 to 0.08; *T*=–1.65; *P*=.10). Complete data are provided in Table 4.

XSL•FO
RenderX

Table 4. Comparison of exercise endurance indexes between the 2 groups before and after intervention.

Index group	Pre-exercise, mean (SD)	Postexercise, mean (SD)	Between-group difference (95% CI)	*t* test (*df*)	*P* value[a]	Cohen *d*
6MWD[b]			32.00 (5.40-58.60)			0.69
Experimental group	399.00 (67.92)	507.00 (62.23)		–7.40 (24)	<.001	
Control group	415.12 (73.13)	460.96 (71.24)		–3.39 (24)	<.001	
t test (*df*)	–0.81 (48)	2.43 (48)		—[c]	—	
P value	.42	.02		—	—	
Borg score			–0.40 (–0.88 to 0.08)			0.45
Experimental group	2.12 (0.93)	0.96 (0.87)		3.63 (24)	<.001	
Control group	2.32 (0.80)	1.36 (0.94)		3.67 (24)	<.001	
t test (*df*)	–0.87 (48)	–1.65 (48)		—	—	
P value	.38	.10		—	—	

[a]The significance level was set at $P<.05$.

[b]6MWD: 6-minute walk distance.

[c]Not applicable.

Pulmonary Function Indexes of the 2 Groups Compared

At baseline, pulmonary function measures were comparable between groups ($P>.05$). After 4 weeks, every spirometric variable improved more in the motion capture cohort than in controls (all $P<.05$). The control group did not show significant within-group changes in any of these measures (all $P\geq.05$). For FVC, the experimental group increased from a mean of 1716.40 (SD 639.30) to 2576.80 (SD 629.54) mL ($T=–7.14$; $P<.001$), while the control group showed no significant change (2106.40, SD 823.67 mL to 2100.40, SD 742.77 mL; $T=0.96$; $P=.35$). The between-group comparison revealed a statistically significant difference favoring the motion capture cohort (mean difference 476.4 mL, 95% CI 85.2-867.6; $T=2.45$; $P=.02$). For FEV_1, the experimental group improved from 1294.80 (SD 413.62) to 2029.20 (SD 526.79) mL ($T=–8.74$; $P<.001$), compared with minimal change in the control group (1462.28, SD 629.17 mL to 1683.20, SD 647.47 mL; $T=–1.83$; $P=.08$). The between-group difference was statistically significant (mean difference 346.0 mL, 95% CI 9.8-682.2; $T=2.07$; $P=.04$). For the FEV_1/FVC ratio, the experimental group increased from 69.88% (SD 20.46) to 91.60% (SD 22.14; $T=–3.55$; $P<.001$), while the control group showed no significant improvement (72.04%, SD 18.30 to 78.80%, SD 18.01; $T=–1.31$; $P=.20$). The between-group comparison favored the experimental group (mean difference 12.8%, 95% CI 1.3-24.3; $T=–2.15$; $P=.03$). For PEF, the experimental group improved from 130.58 (SD 73.59) to 230.21 (SD 81.66) L/minute ($T=4.23$; $P<.001$), whereas the control group showed no significant change (170.04, SD 70.16 to 159.36, SD 62.26 L/minute [$T=0.48$; $P=.63$]). The between-group difference was statistically significant (mean difference 70.8 L/minute, 95% CI 26.4-115.2; $T=–3.15$; $P=.002$). Complete data are presented in Table 5.

XSL•FO
RenderX

Table 5. Comparison of pulmonary function indexes between the 2 groups before and after intervention.

Index and group	Pre-exercise, mean (SD)	Postexercise, mean (SD)	Between-group difference (95% CI)	*t* test (*df*)	*P* value[a]	Cohen *d*
FVC[b]			476.00 (85.20-867.60)			0.71
Experimental group	1716.40 (639.30)	2576.80 (629.54)		–7.14 (24)	<.001	
Control group	2106.40 (823.67)	2100.40 (742.77)		0.96 (24)	.35	
t test (*df*)	–1.87 (48)	2.45 (48)		—[c]	—	
P value	.07	.02		—	—	
FEV_1[d]			346.0 (9.80-682.2)			0.62
Experimental group	1294.80 (413.62)	2029.20 (526.79)		–8.74 (24)	<.001	
Control group	1462.28 (629.17)	1683.20 (647.47)		–1.83 (24)	.08	
t test (*df*)	–0.73 (48)	2.07 (48)		—	—	
P value	.47	.04		—	—	
FEV_1/FVC[e]			12.80 (1.3-24.3)			0.58
Experimental group	69.88 (20.46)	91.60 (22.14)		–3.55 (24)	<.001	
Control group	72.04 (18.30)	78.80 (18.01)		–1.31 (24)	.20	
t test (*df*)	–0.28 (48)	–2.15 (48)		—	—	
P value	.78	.03		—	—	
PEF[f]			70.8 (26.4-115.2)			0.95
Experimental group	130.58 (73.59)	230.21 (81.66)		4.23 (24)	<.001	
Control group	170.04 (70.16)	159.36 (62.26)		0.48 (24)	.63	
t test (*df*)	–1.89 (48)	–3.15 (48)		—	—	
P value	.06	.002		—	—	

[a]The significance level was set at $P<.05$.

[b]FVC: forced vital capacity.

[c]Not available.

[d]FEV_1: forced expiratory volume in 1 second.

[e]FEV_1/FVC: The ratio between FEV_1 and FVC.

[f]PEF: peak expiratory flow.

Activity Ability Indexes of the 2 Groups Compared

At baseline, activity-capacity indices were comparable across groups (all $P>.05$). After 4 weeks, both groups demonstrated significant within-group improvements in most functional mobility measures, with the motion capture group showing superior gains in the TUG test. For the TUG test, the experimental group improved from a mean of 13.11 (SD 4.39) to 9.95 (SD 2.41) seconds ($T=3.46$; $P<.001$), while the control group improved from 12.78 (SD 3.35) to 11.58 (SD 3.04) seconds ($T=2.42$; $P=.02$). The between-group comparison revealed a statistically significant difference favoring the motion capture cohort (mean difference –1.63 seconds, 95% CI –3.19 to –0.07; $T=–2.11$; $P=.04$). For Stand on 1 foot, the experimental group improved from 24.79 (SD 19.71) to 43.72 (SD 31.59) seconds ($T=4.14$; $P<.001$), and the control group improved from 38.36 (SD 79.30) to 49.68 (SD 58.19) seconds ($T=2.52$; $P=.01$). However, the between-group difference was not statistically significant (mean difference –5.96 seconds, 95% CI –33.2 to 21.3; $T=–0.44$; $P=.66$). For the 30-second CST, the experimental group improved from 9.52 (SD 3.55) to 13.96 (SD 4.94) repetitions ($T=–5.73$; $P<.001$), and the control group improved from 11.16 (SD 3.88) to 12.68 (SD 3.00) repetitions ($T=–2.63$; $P=.02$). The between-group difference did not reach statistical significance (mean difference 1.28 repetitions, 95% CI –1.18 to 3.74; $T=0.78$; $P=.43$). For the FTSST, the experimental group improved from 13.56 (SD 6.78) to 9.32 (SD 2.90) seconds ($T=3.79$; $P<.001$), and the control group improved from 11.03 (SD 3.00) to 9.48 (SD 2.24) seconds ($T=2.35$; $P=.02$). The between-group difference was not statistically significant (mean difference–0.16 seconds, 95% CI –1.67 to 1.35; $T=–0.25$; $P=.80$). For the 2-minute step test, the experimental group improved from 1832.56 (SD 41.46) to 179.84 (SD 38.27) steps ($T=–6.10$; $P<.001$), and the control group improved from 143.88 (SD 33.74) to 177.48 (SD 30.25) steps ($T=–4.82$; $P<.001$). The between-group difference was not statistically significant (mean difference 2.36 steps, 95% CI –17.8 to 22.5; $T=0.24$; $P=.81$). Complete data are provided in Table 6.

XSL•FO
RenderX

Table 6. Comparison of indicators of activity ability between the 2 groups before and after intervention.

Index and group	Pre-exercise, mean (SD)	Postexercise, mean (SD)	Between-group difference (95% CI)	*t* test (*df*)	*P* value[a]	Cohen *d*
TUG[b]			–1.63 (–3.19 to –0.07)			0.59
Experimental group	13.11 (4.39)	9.95 (2.41)		3.46 (24)	.001	
Control group	12.78 (3.35)	11.58 (3.04)		2.42 (24)	.02	
t test (*df*)	–0.33 (48)	–2.11		—[c]	—	
P value	.74	.04		—	—	
Stand on 1 foot			–5.96 (–33.2 to 21.3)			0.13
Experimental group	24.79 (19.71)	43.72 (31.59)		4.14 (24)	<.001	
Control group	38.36 (79.30)	49.68 (58.19)		2.52 (24)	.01	
t test (*df*)	–0.18 (48)	–0.44 (48)		—	—	
P value	.86	.66		—	—	
30-second CST[d]			1.28 (–1.18 to 3.74)			0.31
Experimental group	9.52 (3.55)	13.96 (4.94)		–5.73 (24)	<.001	
Control group	11.16 (3.88)	12.68 (3.00)		–2.63 (24)	.02	
t test (*df*)	–1.39 (48)	–0.78 (48)		—	—	
P value	.16	.43		—	—	
FTSST[e]			–0.16 (–1.67 to 1.35)			0.06
Experimental group	13.56 (6.78)	9.32 (2.90)		3.79 (24)	<.001	
Control group	11.03 (3.00)	9.48 (2.24)		2.35 (24)	.02	
t test (*df*)	–1.45 (48)	0.25 (48)		—	—	
P value	.15	.80		—	—	
Two-minute step test			2.36 (–17.8 to 22.5)			0.07
Experimental group	132.56 (41.46)	179.84 (38.27)		–6.10 (24)	<.001	
Control group	143.88 (33.74)	177.48 (30.25)		–4.82 (24)	<.001	
t test (*df*)	–0.98 (48)	0.24 (48)		—	—	
P value	.33	.81		—	—	

[a]The significance level was set at *P*<.05.

[b]TUG: Timed Up and Go.

[c]Not available.

[d]30-second CST: 30-second Chair-Stand Test.

[e]FTSST: Five-times Sit-To-Stand Test.

Comparison of Quality of Life Indicators Between the 2 Groups of Patients

At baseline, there were no statistically significant differences between the experimental and control groups across all FACT-L subscales and total score (all *P*>.05), indicating good comparability before intervention. For PWB, the experimental group improved from 14.68 (SD 2.27) to 21.68 (SD 2.27; *T*=–4.55; *P*<.001), while the control group improved from 13.72 (SD 2.23) to 17.68 (SD 2.27; *T*=–4.52; *P*<.001). The between-group comparison revealed a statistically significant difference favoring the motion capture cohort (mean difference 4.00 points, 95% CI 2.66-5.34; *T*=6.24; *P*<.001). For SWB, the experimental group improved from 16.04 (SD 1.24) to 20.08 (SD 1.32; *T*=–4.33; *P*<.001), and the control group improved from 16.80 (SD 1.53) to 19.80 (SD 1.38; *T*=–4.40; *P*<.001). However, the between-group difference was not statistically significant (mean difference 0.28 points, 95% CI –0.76 to 1.32; *T*=–0.53; *P*=.60). For EWB, the experimental group improved from 13.48 (SD 1.16) to 15.80 (SD 1.19; *T*=–4.48; *P*<.001), and the control group improved from 13.72 (SD 1.49) to 15.92 (SD 1.38; *T*=–4.44; *P*<.001). The between-group difference did not reach statistical significance (mean difference –0.12 points, 95% CI –0.98 to 0.74; *T*=–0.27; *P*=.79). For FWB, the experimental group improved from 12.28 (SD 2.23) to 18.68 (SD 2.27; *T*=–4.55; *P*<.001), while the control group improved from 11.04 (SD 2.25) to 14.00 (SD 1.94; *T*=–4.44; *P*<.001). The between-group comparison revealed a statistically significant difference favoring the motion capture cohort (mean difference 4.68 points, 95% CI 3.44-5.92; *T*=7.85; *P*<.001). For

XSL•FO
RenderX

the Additional Concerns–Lung (ACL) subscale, the experimental group improved from 9.72 (SD 1.93) to 17.24 (SD 2.11; T=–4.41; P<.001), and the control group improved from 9.64 (SD 1.91) to 12.96 (SD 1.77; T=–4.41; P<.001). The between-group difference was statistically significant (mean difference 4.28 points, 95% CI 3.14-5.42; T=7.78; P<.001). For the total FACT-L score, the experimental group improved from 66.20 (SD 6.83) to 93.36 (SD 6.72; T=–51.12; P<.001), while the control group improved from 64.92 (SD 6.85) to 80.36 (SD 6.04; T=–33.94; P<.001). The between-group comparison revealed a statistically significant difference favoring the experimental group (mean difference 13.00 points, 95% CI 9.30-16.70; T=7.19; P<.001). Complete data are provided in Table 7.

XSL•FO
RenderX

Table 7. Comparison of quality of life indicators before and after intervention between the 2 groups.

Index and group	Pre-exercise, mean (SD)	Postexercise, mean (SD)	Between-group difference (95% CI)	*t* test (*df*)	*P* value[a]	Cohen *d*
PWB[b]			4.00 (2.66-5.34)			1.76
Experimental group	14.68 (2.27)	21.68 (2.27)		–4.55 (24)	<.001	
Control group	13.72 (2.23)	17.68 (2.27)		–4.52 (24)	<.001	
t test (*df*)	1.51 (48)	6.24		__[c]	—	
P value	.14	<.001		—	—	
SWB[d]			0.28 (–0.76 to 1.32)			0.21
Experimental group	16.04 (1.24)	20.08 (1.32)		–4.33 (24)	<.001	
Control group	16.80 (1.53)	19.80 (1.38)		–4.40 (24)	<.001	
t test (*df*)	–1.79 (48)	–0.53		—	—	
P value	.07	.60		—	—	
EWB[e]			–0.12 (–0.98 to 0.74)			0.09
Experimental group	13.48 (1.16)	15.80 (1.19)		–4.48 (24)	<.001	
Control group	13.72 (1.49)	15.92 (1.38)		–4.44 (24)	<.001	
t test (*df*)	–1.77 (48)	–0.27		—	—	
P value	.08	.79		—	—	
FWB[f]			4.68 (3.44 to 5.92)			2.22
Experimental group	12.28 (2.23)	18.68 (2.27)		–4.55 (24)	<.001	
Control group	11.04 (2.25)	14.00 (1.94)		–4.44 (24)	<.001	
t test (*df*)	1.96 (48)	7.85		—	—	
P value	.06	<.001		—	—	
ACL[g]			4.28 (3.14 to 5.42)			2.20
Experimental group	9.72 (1.93)	17.24 (2.11)		–4.41 (24)	<.001	
Control group	9.64 (1.91)	12.96 (1.77)		–4.41 (24)	<.001	
t test (*df*)	–0.72 (48)	7.78		—	—	
P value	.47	<.001		—	—	
Total score			13.00 (9.30 to 16.70)			2.03
Experimental group	66.20 (6.83)	93.36 (6.72)		–51.12 (24)	<.001	
Control group	64.92 (6.85)	80.36 (6.04)		–33.94 (24)	<.001	
t test (*df*)	0.66 (48)	7.19		—	—	
P value	.51	<.001		—	—	

[a]The significance level was set at $P<.05$.

[b]PWB: physical well-being.

[c]Not available.

[d]SWB: social/family well-being.

[e]EWB: emotional well-being.

[f]FWB: functional well-being.

[g]ACL: Additional Concerns-Lung subscale.

Comparison of Treatment Compliance and Incidence of Adverse Events Between the 2 Groups of Patients

Home-based program adherence was markedly greater in the motion capture group (18/25, 72% vs 10/25, 40%; $P=.02$), and no exercise-related adverse events were recorded in either group (Table 8).

XSL•FO
RenderX

Table 8. Comparison of treatment compliance between the 2 groups.

Group	Training compliance rate (≥3 times per week)	Noncompliance rate (≤3 times per week)
Experimental group, n (%)	18 (72)	7 (28)
Control group, n (%)	10 (40)	
Test value	5.180	—[a]
P value	.02	—

[a]Not available.

Discussion

Principal Findings

This study demonstrated that motion recognition–based telerehabilitation was associated with greater improvements in postoperative recovery among patients with NSCLC compared with conventional home-based video-guided training. Specifically, participants in the intervention group showed superior gains in exercise tolerance, pulmonary function, selected aspects of functional mobility, and the physical and functional domains of quality of life. Rehabilitation adherence was also higher in the motion-guided group, and no intervention-related adverse events were observed in either group.

Taken together, these findings suggest that integrating real-time kinematic feedback into home-based pulmonary rehabilitation may enhance movement quality, functional recovery, and patient engagement beyond standard remote instruction. Importantly, these results are consistent with emerging evidence from recent systematic reviews and network meta-analyses [27] indicating that telerehabilitation can achieve improvements in physical function and quality of life comparable with, or in some contexts exceeding, conventional face-to-face pulmonary rehabilitation programs. This study extends this body of evidence by highlighting the potential added value of motion recognition–based feedback within a postoperative population with lung cancer.

Impact of Telerehabilitation Based on Motion Recognition Technology on Exercise Tolerance in Postoperative Patients With NSCLC

Exercise tolerance constitutes a primary indicator of postoperative functional recovery in patients undergoing NSCLC resection [28]. Baseline 6MWD did not differ between groups, confirming equivalent preintervention capacity. Although both cohorts achieved after-intervention gains—reflecting the efficacy of any structured training—the experimental group demonstrated a substantially larger increment, underscoring the supplementary value of motion capture–guided telerehabilitation. These observations align with prior evidence.

The superior improvement in the motion capture group can be attributed to several interconnected mechanisms. First, real-time kinematic feedback facilitates motor learning through augmented feedback loops. According to the guidance hypothesis ("Comparative Effects of Sensory Motor and Virtual Reality Interventions to Improve Gait, Balance, and Quality of Life MS Patients"), concurrent feedback enhances skill acquisition by providing immediate error information, allowing patients to correct aberrant movement patterns (eg, knee valgus and excessive trunk sway) before they become ingrained. This process engages the cerebellar-cortical circuit, promoting neuroplastic changes that optimize movement efficiency [29].

Second, real-time, AI-guided motion correction may augment walking economy by refining neuromuscular synchrony and dynamic stability. Earlier work documents that cancer survivors exhibit quadriceps weakness approximately one-third to two-fifths greater than healthy peers [30]. Instantaneous joint-angle feedback during tasks such as squats promptly redirects aberrant mechanics—knee valgus or excessive trunk sway—reducing metabolic cost per stride.

Third, patients in the experimental group covered more ground during the 6MWT without parallel rises in Borg ratings; superior gait efficiency and balance translated into greater distance with lower perceived exertion. Moreover, quantitative visual cues (eg, "posture accuracy: 92%") reinforced engagement and intensity, yielding a compliance rate of 72% (18/25) versus 40% (10/25) in controls. Sustained participation drove cumulative cardiopulmonary and musculoskeletal adaptations—the principal determinants of improved exercise capacity [31].

The Borg scale gauges perceived exertion and complements objective measures of exercise tolerance [32]. In our trial, Borg ratings did not differ between the motion capture and control groups. Nevertheless, instantaneous kinematic feedback appeared to bolster self-efficacy and refine movement efficiency, allowing the experimental cohort to increase walking distance without proportionate fatigue. Conversely, controls—hampered by suboptimal mechanics—covered shorter distances and yet reported greater perceived exertion per meter.

Impact of Telerehabilitation Based on Motion Recognition Technology on Pulmonary Function in Postoperative Patients With NSCLC

Enhancing pulmonary function represents a key objective in postoperative care for patients with lung cancer [33]. The superior improvements in FVC, FEV_1, and PEF observed in the motion capture group likely stem from the system's capacity to ensure optimal respiratory muscle synergy through real-time corrective feedback. While participants in both cohorts completed identical training protocols, only those in the intervention group received real-time monitoring of biomechanical parameters with prompt adjustments for aberrant movement patterns [34].

While participants in both cohorts completed identical exercise protocols, only those in the intervention group received

XSL•FO
RenderX

continuous monitoring of trunk alignment during core exercises. This distinction is critical because respiratory function is intimately linked to core stability. The diaphragm, the primary muscle of inspiration, shares fascial connections with the transversus abdominis and pelvic floor muscles, forming the thoracoabdominal pump [35]. Proper activation of this integrated system requires precise coordination—a skill that benefits from augmented feedback.

Consider the plank—an evidence-based core stabilization maneuver engaging key musculature including the transversus abdominis, rectus abdominis, and erector spinae. Regular plank training (3 times per week) enhances both respiratory function and general fitness in aging populations [36]. Yet, its benefits are substantially attenuated by technical errors such as hyperlordosis [37], which specifically impairs transversus abdominis recruitment [38]. Such deviations not only undermine therapeutic effects but also elevate spinal loading and injury potential.

The motion recognition system provided instantaneous proprioceptive feedback that helped patients maintain neutral spine alignment, thereby optimizing the feedforward control mechanism of core muscles. This feedforward activation—where trunk muscles contract in anticipation of limb movement—is essential for maintaining spinal stability during dynamic activities [39]. By reinforcing this mechanism during training, the intervention likely enhanced the neural drive to respiratory muscles, translating into improved spirometric outcomes. Furthermore, the observed improvements in PEF suggest enhanced expiratory muscle function, possibly through better recruitment of abdominal muscles during forced expiration.

Impact of Telerehabilitation Based on Motion Recognition Technology on Functional Mobility in Postoperative Patients With NSCLC

Functional mobility assessment serves as a critical component in evaluating rehabilitation success. Standardized tests such as the TUG [40], single-leg balance assessment [41], 30-second CST [42], FTSST [43], and 2-minute stepping task [44] collectively offer a comprehensive evaluation of movement capacity and functional performance. These validated measures enable clinicians to objectively determine patients' functional capacity for activities of daily living and community participation, while effectively tracking recovery milestones.

Among these assessments, the TUG represents a composite functional task requiring coordinated execution of multiple motor components, including sit-to-stand transition, gait initiation, turning, and resitting. Performance on this test reflects not only lower limb strength but also dynamic balance, motor sequencing, and intertask coordination [45]. The observed improvement in TUG performance in the motion recognition group may be related to enhanced movement precision and temporal coordination facilitated by real-time kinematic feedback. The motion recognition system provides continuous monitoring of joint angles and movement trajectories, potentially supporting more accurate sensorimotor integration during exercise execution—the process by which sensory feedback refines motor output [46]. Similar improvements in postural control and motor stability have been reported in neurological populations receiving telemedicine-based rehabilitation, including individuals with Parkinson disease, suggesting that remote feedback mechanisms may enhance movement regulation across different clinical conditions [47]. Real-time visual feedback of joint angles and movement trajectories may have facilitated implicit motor learning, whereby patients internalize correct movement patterns without extensive conscious processing. This contrasts with explicit learning approaches that rely primarily on verbal instruction and are more cognitively demanding and potentially less durable under stress [48]. The improved TUG performance in the experimental group likely reflects enhanced procedural memory for the sequence of movements involved in the test.

The lack of significant between-group differences in single-leg stance, 30-second CST, FTSST, and 2-minute step tests does not contradict this interpretation. These tests predominantly assess isolated components of physical function—static balance, lower extremity strength, and repetitive endurance—rather than the integrated motor control required for complex tasks. Both groups received identical exercise prescriptions targeting lower limb musculature, likely producing comparable gains in gross strength (as reflected in the similar improvements on these tests). The unique value of the motion-aware system lies in optimizing movement quality and interjoint coordination—domains captured by composite measures such as the TUG but not by simpler functional tests. This dissociation between strength gains and coordination improvements is consistent with the principle that motor skill acquisition and muscular adaptation are mediated by distinct neural and physiological mechanisms [49].

Impact of Telerehabilitation Based on Motion Recognition Technology on Quality of Life in Postoperative Patients With NSCLC

Quality of life represents a critical multidimensional end point in postsurgical NSCLC rehabilitation, encompassing physical, psychological, and social well-being as primary markers of recovery success [50]. Our intervention group showed statistically superior improvements in PWB, FWB, ACL, and the total FACT-L score relative to controls. This pattern implies that motion-aware telerehabilitation offers particular advantages for restoring physical and functional aspects of quality of life, potentially attributable to its capacity for optimizing functional recovery through millimeter-precise movement analysis and instantaneous form correction. Additionally, the system's ability to provide symptom-specific movement modifications and breathing technique guidance may explain this therapeutic advantage, empowering patients to better control their condition-related challenges.

The analysis revealed no statistically significant intergroup differences in either SWB or EWB domain. This observation may be explained by several key considerations. First, meaningful enhancement of social support systems and psychological state typically necessitates prolonged intervention durations and often benefits from dedicated mental health support, posing challenges for demonstrating immediate effects within a limited study period [51]. Second, the psychological

XSL•FO
RenderX

status of postsurgical patients with NSCLC is frequently influenced by complex psychosocial factors—particularly disease recurrence anxiety and treatment-related financial stress [52]—that extend beyond the scope of physical rehabilitation and may require targeted psychotherapeutic approaches combined with robust social services.

Impact of Telerehabilitation Based on Motion Recognition Technology on Exercise Adherence and Safety in Postoperative Patients With NSCLC

Program adherence represents a critical determinant of rehabilitation efficacy [53]. Our data revealed substantially better compliance rates in the intervention group than in controls, indicating that motion-aware telerehabilitation systems may optimize treatment engagement for postsurgical patients with NSCLC. This effect appears mediated by the platform's multimodal feedback mechanisms—including real-time performance metrics, achievement scoring, and graphical progress tracking—which collectively enhance treatment motivation through positive reinforcement. Contemporary research confirms that instantaneous performance feedback bolsters rehabilitation adherence by reinforcing patient self-confidence and perceived competence [54]. Furthermore, the system's interactive design elements convert conventional exercises into dynamic, goal-oriented tasks, offering superior psychological engagement compared with passive video demonstrations.

Regarding safety outcomes, neither study group reported intervention-related adverse events, confirming the acceptable risk profile of home-based motion capture rehabilitation. The technology's ability to continuously monitor and correct biomechanical execution helps ensure proper movement patterns, consequently minimizing injury risks associated with faulty technique [55].

Limitations

Despite the promising findings, several constraints warrant consideration. First, assessments were restricted to the 4-week end point, precluding insight into durability; subsequent studies should extend follow-up to 3, 6, and 12 months to chart longer-term recovery. Second, the modest sample size may limit both external validity and statistical power. Although significant between-group differences were observed for several primary outcomes, the study may have been underpowered to detect smaller yet clinically meaningful effects in some secondary measures. Accordingly, the absence of statistically significant differences for certain functional indicators should be interpreted cautiously, as the possibility of type II error cannot be excluded. Larger, adequately powered multicenter trials are required to validate and refine these findings. Third, the lack of preoperative functional baselines hampers delineation of the full perioperative trajectory; future protocols should incorporate presurgical assessments to construct comprehensive longitudinal recovery profiles.

Conclusions

To our knowledge, this is the first randomized controlled trial to integrate motion capture analytics into telerehabilitation following NSCLC surgery. Over 4 weeks, the approach proved both safe and superior to standard video-guided exercise, yielding larger gains in exercise tolerance, spirometry, functional mobility, and quality of life. These data establish a clinically viable strategy that can readily be scaled across telemedicine and postoperative care pathways.

Acknowledgments

The authors confirm that no generative artificial intelligence tools were used in the preparation or revision of this manuscript.

Data Availability

The datasets used and/or analyzed during this study are available from the corresponding author on reasonable request.

Funding

This work was supported by grants from the National Natural Science Foundation of China (grant 82472588) and the Natural Science Foundation of Guangdong Province (2025A1515012331 and 2025A1515012782).

Authors' Contributions

Conceptualization: LC, XX, XY, SL, HZ, GH, QZ
Methodology: LC, XX, XY, SL, NC, ZL, TY, QL, ML
Formal analysis: LC, XX
Writing – original draft: LC, XX, XY, SL
Data curation: XY, SL, NC, ZL, TY, QL, ML
Investigation: NC, ZL, TY, QL, ML
Writing – review & editing: HZ, GH, QZ
Supervision: HZ, GH, QZ
Project administration: HZ, GH, QZ
Funding acquisition: HZ, GH, QZ

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online TeleHealth) (V 1.6.1) checklist.
[PDF File (Adobe PDF File), 2738 KB - mhealth_v14i1e82447_app1.pdf]

Multimedia Appendix 2
The pulmonary function test graph and the 6-minute walk test graph.
[DOC File , 575 KB - mhealth_v14i1e82447_app2.doc]

References

1. Li P, Wang Q, Wang H. MicroRNA-204 inhibits the proliferation, migration and invasion of human lung cancer cells by targeting PCNA-1 and inhibits tumor growth in vivo. Int J Mol Med 2019 Mar;43(3):1149-1156 [FREE Full text] [doi: 10.3892/ijmm.2018.4044] [Medline: 30628638]
2. Zhang R, Wu M, Cheng Y, Zhao L, Xu Y, Xia L. Barriers and facilitators to prehabilitation of elderly patients with early lung cancer from the perspective of different clinical professionals: a qualitative study. BMC Nurs 2025 May 12;24(1):517 [FREE Full text] [doi: 10.1186/s12912-025-03153-2] [Medline: 40355845]
3. Zhong J, Trinh I, Raju S, Hsu M. Pulmonary rehabilitation in patients with operable non-small cell lung cancer. J Clin Med 2025 Jan 24;14(3):770 [FREE Full text] [doi: 10.3390/jcm14030770] [Medline: 39941440]
4. Jovanoski N, Abogunrin S, Di Maio D, Belleli R, Hudson P, Bhadti S, et al. Systematic literature review to identify cost and resource use data in patients with early-stage non-small cell lung cancer (NSCLC). Pharmacoeconomics 2023 Nov;41(11):1437-1452 [FREE Full text] [doi: 10.1007/s40273-023-01295-2] [Medline: 37389802]
5. Rosero ID, Ramírez-Vélez R, Lucia A, Martínez-Velilla N, Santos-Lozano A, Valenzuela PL, et al. Systematic review and meta-analysis of randomized, controlled trials on preoperative physical exercise interventions in patients with non-small-cell lung cancer. Cancers (Basel) 2019 Jul 05;11(7):944 [FREE Full text] [doi: 10.3390/cancers11070944] [Medline: 31284372]
6. Messaggi-Sartor M, Marco E, Martínez-Téllez E, Rodriguez-Fuster A, Palomares C, Chiarella S, et al. Combined aerobic exercise and high-intensity respiratory muscle training in patients surgically treated for non-small cell lung cancer: a pilot randomized clinical trial. Eur J Phys Rehabil Med 2019 Feb;55(1):113-122 [FREE Full text] [doi: 10.23736/S1973-9087.18.05156-0] [Medline: 29984565]
7. Cheng X. Effects of lung rehabilitation therapy in improving respiratory motor ability and alleviating dyspnea in patients with lung cancer after lobectomy: a clinical study. Altern Ther Health Med 2022;28(3):18-23.
8. Tadsuan J, Lai Y, Lee Y, Chen M. The effectiveness of exercise interventions on psychological distress in patients with lung cancer: a systematic review and meta-analysis. J Cancer Surviv 2026 Jun;20(3):910-926. [doi: 10.1007/s11764-024-01696-y] [Medline: 39466551]
9. Sánchez-Romero EA, García-Barredo-Restegui T, Martínez-Rolando L, Villafañe JH, Galán-Fraguas A, Jurado-Molina R, et al. Addressing post-COVID-19 musculoskeletal symptoms through pulmonary rehabilitation with an evidence-based eHealth education tool: preliminary results from a pilot randomized controlled clinical trial. Medicine (Baltimore) 2025 Mar 07;104(10):e41583 [FREE Full text] [doi: 10.1097/MD.0000000000041583] [Medline: 40068080]
10. Sánchez Romero EA, Fernández Carnero J, Alonso Pérez JL, Martínez Rolando L, Villafañe JH. Addressing post-COVID-19 musculoskeletal symptoms through pulmonary rehabilitation and telemedicine: a study protocol. F1000Res 2022;11:898 [FREE Full text] [doi: 10.12688/f1000research.122843.3] [Medline: 38524251]
11. Martínez-Pozas O, Meléndez-Oliva E, Rolando LM, Rico JAQ, Corbellini C, Sánchez Romero EA. The pulmonary rehabilitation effect on long covid-19 syndrome: a systematic review and meta-analysis. Physiother Res Int 2024 Apr;29(2):e2077. [doi: 10.1002/pri.2077] [Medline: 38389299]
12. Park S, Kim JY, Lee JC, Kim HR, Song S, Kwon H, et al. Mobile phone app-based pulmonary rehabilitation for chemotherapy-treated patients with advanced lung cancer: pilot study. JMIR Mhealth Uhealth 2019 Feb 04;7(2):e11094 [FREE Full text] [doi: 10.2196/11094] [Medline: 30714943]
13. He S, Meng D, Wei M, Guo H, Yang G, Wang Z. Proposal and validation of a new approach in tele-rehabilitation with 3D human posture estimation: a randomized controlled trial in older individuals with sarcopenia. BMC Geriatr 2024;24(1):586 [FREE Full text] [doi: 10.1186/s12877-024-05188-7] [Medline: 38977995]
14. Chae HJ, Kim JB, Park G, O'Sullivan DM, Seo J, Park JJ. An artificial intelligence exercise coaching mobile app: development and randomized controlled trial to verify its effectiveness in posture correction. Interact J Med Res 2023 Sep 12;12:e37604 [FREE Full text] [doi: 10.2196/37604] [Medline: 37698913]
15. Xiao C, Zhao Y, Li G, Zhang Z, Liu S, Fan W, et al. Clinical efficacy of multimodal exercise telerehabilitation based on AI for chronic nonspecific low back pain: randomized controlled trial. JMIR Mhealth Uhealth 2025 May 22;13:e56176 [FREE Full text] [doi: 10.2196/56176] [Medline: 40402551]

XSL•FO
RenderX

16. Yamashita M, Kamiya K, Hamazaki N, Uchida S, Noda T, Maekawa E, et al. Effects of acute phase intensive physical activity (ACTIVE-PA) monitoring and education for cardiac patients: pilot study of a randomized controlled trial. J Med Internet Res 2023 Dec 20;25:e42235 [FREE Full text] [doi: 10.2196/42235] [Medline: 38117552]
17. Riely GJ, Wood DE, Ettinger DS, Aisner DL, Akerley W, Bauman JR, et al. Non-small cell lung cancer, Version 4.2024, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw 2024 May;22(4):249-274. [doi: 10.6004/jnccn.2204.0023] [Medline: 38754467]
18. Jutagir DR, Blomberg BB, Carver CS, Lechner SC, Timpano KR, Bouchard LC, et al. Social well-being is associated with less pro-inflammatory and pro-metastatic leukocyte gene expression in women after surgery for breast cancer. Breast Cancer Res Treat 2017 Aug;165(1):169-180 [FREE Full text] [doi: 10.1007/s10549-017-4316-3] [Medline: 28560656]
19. Kiss NK, Denehy L, Edbrooke L, Prado CM, Ball D, Siva S, et al. Predicting muscle loss during lung cancer treatment (PREDICT): protocol for a mixed methods prospective study. BMJ Open 2021 Sep 27;11(9):e051665 [FREE Full text] [doi: 10.1136/bmjopen-2021-051665] [Medline: 34580100]
20. Nedjat-Haiem FR, Cadet T, Parada H, Jones T, Jimenez EE, Thompson B, et al. Financial hardship and health related quality of life among older Latinos with chronic diseases. Am J Hosp Palliat Care 2021 Aug;38(8):938-946 [FREE Full text] [doi: 10.1177/1049909120971829] [Medline: 33153273]
21. Cui C, Wang L, Wang X. Health-related quality of life and social constraints among Chinese breast cancer patients: a cross-sectional study. Health Qual Life Outcomes 2021 Oct 12;19(1):238 [FREE Full text] [doi: 10.1186/s12955-021-01871-0] [Medline: 34641883]
22. Özüdoğru A, Canlı M, Gürses ÖA, Alkan H, Yetiş A. Determination of five times-sit-to-stand test performance in patients with multiple sclerosis: validity and reliability. Somatosens Mot Res 2023 Jun;40(2):72-77. [doi: 10.1080/08990220.2022.2157395] [Medline: 36524614]
23. Höglund J, Boström C, Sundh J. Six-Minute Walking Test and 30 Seconds Chair-Stand-Test as predictors of mortality in COPD—a cohort study. Int J Chron Obstruct Pulmon Dis 2022;17:2461-2469 [FREE Full text] [doi: 10.2147/COPD.S373272] [Medline: 36217331]
24. Santos DSD, Queiroz C, Dias C, Cipriano Junior G, Borges QO, Ritt LEF. Prediction of peak oxygen consumption in patients with heart disease based on performance on the Timed Up and Go Test. Arq Bras Cardiol 2023 Dec;120(12):e20230338 [FREE Full text] [doi: 10.36660/abc.20230338] [Medline: 38126569]
25. Wu R, Zhang Y, Bai JJ, Sun J, Bao ZJ, Wang Z. Impact of lower limb muscle strength on walking function beyond aging and diabetes. J Int Med Res 2020 Jun;48(6):300060520928826 [FREE Full text] [doi: 10.1177/0300060520928826] [Medline: 32589863]
26. Elam C, Hvid LG, Christensen U. Effects of age on muscle power, postural control and functional capacity after short-term immobilization and retraining. J Musculoskelet Neuronal Interact 2022;22(4):486-497.
27. Martínez-Pozas O, Corbellini C, Cuenca-Zaldívar JN, Meléndez-Oliva É, Sinatti P, Sánchez Romero EA. Effectiveness of telerehabilitation versus face-to-face pulmonary rehabilitation on physical function and quality of life in people with post COVID-19 condition: a systematic review and network meta-analysis. Eur J Phys Rehabil Med 2024 Oct;60(5):868-877 [FREE Full text] [doi: 10.23736/S1973-9087.24.08540-X] [Medline: 39235257]
28. Jiao M, Liang H, Zhang M. Effect of exercise on postoperative recovery of patients with non-small cell lung cancer: a systematic review and meta-analysis. Discov Oncol 2024 Jun 17;15(1):230 [FREE Full text] [doi: 10.1007/s12672-024-01079-w] [Medline: 38884823]
29. Gupta DS. Processing of sub- and supra-second intervals in the primate brain results from the calibration of neuronal oscillators via sensory, motor, and feedback processes. Front Psychol 2014;5:816 [FREE Full text] [doi: 10.3389/fpsyg.2014.00816] [Medline: 25136321]
30. Roberts BM, Frye GS, Ahn B, Ferreira LF, Judge AR. Cancer cachexia decreases specific force and accelerates fatigue in limb muscle. Biochem Biophys Res Commun 2013 Jun 07;435(3):488-492 [FREE Full text] [doi: 10.1016/j.bbrc.2013.05.018] [Medline: 23673294]
31. Elnaggar RK, Osailan AM, Alghadier M, Elnegamy TE, Morsy WE, Abdrabo MS, et al. Exercise strategies for reversing cardiopulmonary deconditioning in obese children with bronchial asthma: a randomized comparative effectiveness study of constant-load and graded aerobic training. Medicine (Baltimore) 2024 Nov 29;103(48):e40667 [FREE Full text] [doi: 10.1097/MD.0000000000040667] [Medline: 39612435]
32. Stumm TW, Mina S, Okusanya O, Cowan S, Evans NR, Grenda TR. Metrics for perioperative exercise in patients undergoing lung cancer resection: a systematic review. Clin Lung Cancer 2025 Jan;26(1):e33-e40. [doi: 10.1016/j.cllc.2024.11.006] [Medline: 39638661]
33. Fuzhi Y, Dongfang T, Wentao F, Jing W, Yingting W, Nianping M, et al. Rapid recovery of postoperative pulmonary function in patients with lung cancer and influencing factors. Front Oncol 2022;12:927108 [FREE Full text] [doi: 10.3389/fonc.2022.927108] [Medline: 35898890]
34. Cerfoglio S, Capodaglio P, Rossi P, Conforti I, D'Angeli V, Milani E, et al. Evaluation of upper body and lower limbs kinematics through an IMU-based medical system: a comparative study with the optoelectronic system. Sensors (Basel) 2023 Jul 05;23(13):6156-6156 [FREE Full text] [doi: 10.3390/s23136156] [Medline: 37448005]

XSL•FO
RenderX

35. Kim H, Park C, You JSH. Sustainable effectiveness of kinetic chain stretching on active hip flexion movement and muscle activation for hamstring tightness: a preliminary investigation. Technol Health Care 2024;32(S1):155-167 [FREE Full text] [doi: 10.3233/THC-248014] [Medline: 38759046]
36. Park S, Choi BH, Jee YS. Effects of plank exercise on respiratory capacity, physical fitness, and immunocytes in older adults. J Exerc Rehabil 2023 Dec;19(6):332-338 [FREE Full text] [doi: 10.12965/jer.2346536.268] [Medline: 38188128]
37. Hamdani SMZH, Zhuang J, Hadier SG, Khurram H, Hamdani SDH, Danish SS, et al. Establishment of health related physical fitness evaluation system for school adolescents aged 12-16 in Pakistan: a cross-sectional study. Front Public Health 2023;11:1212396 [FREE Full text] [doi: 10.3389/fpubh.2023.1212396] [Medline: 37829094]
38. Lee DH, Hong S, Lee Y, Kim CH, Hwang JM, Lee Z, et al. Is abdominal hollowing exercise using real-time ultrasound imaging feedback helpful for selective strengthening of the transversus abdominis muscle?: a prospective, randomized, parallel-group, comparative study. Medicine (Baltimore) 2018 Jul;97(27):e11369 [FREE Full text] [doi: 10.1097/MD.0000000000011369] [Medline: 29979423]
39. Mehta R, Cannella M, Henry S, Smith S, Giszter S, Silfies SP. Trunk postural muscle timing is not compromised in low back pain patients clinically diagnosed with movement coordination impairments. Motor Control 2017 Apr;21(2):133-157 [FREE Full text] [doi: 10.1123/mc.2015-0049] [Medline: 26623551]
40. Naz I, Turgut B, Gunay Ucurum S, Komurcuoglu B, Ozer Kaya D. Investigation of factors associated with static and dynamic balance in early-stage lung cancer survivors. J Cancer Surviv 2025 Apr;19(2):594-602. [doi: 10.1007/s11764-023-01492-0] [Medline: 37964048]
41. Tough D, Dunning J, Robinson J, Dixon J, Ferguson J, Paul I, et al. Investigating balance, gait, and physical function in people who have undergone thoracic surgery for a diagnosis of lung cancer: a mixed-methods study. Chron Respir Dis 2021;18:14799731211052299 [FREE Full text] [doi: 10.1177/14799731211052299] [Medline: 34715760]
42. van der Leeden M, Balland C, Geleijn E, Huijsmans RJ, Dekker J, Paul MA, et al. In-hospital mobilization, physical fitness, and physical functioning after lung cancer surgery. Ann Thorac Surg 2019 Jun;107(6):1639-1646. [doi: 10.1016/j.athoracsur.2018.12.045] [Medline: 30690020]
43. Zhang R, Yang Z, Shen X, Xia L, Cheng Y. Preoperative physical dysfunction characteristics and influence factors among elderly patients with early lung cancer: a latent class analysis. J Multidiscip Healthc 2024;17:1743-1754 [FREE Full text] [doi: 10.2147/JMDH.S455669] [Medline: 38680878]
44. Poncumhak P, Amput P, Sangkarit N, Promsrisuk T, Srithawong A. Predictive ability of the 2-Minute Step Test for functional fitness in older individuals with hypertension. Ann Geriatr Med Res 2023 Sep;27(3):228-234 [FREE Full text] [doi: 10.4235/agmr.23.0070] [Medline: 37500190]
45. Khruakhorn S, Chiwarakranon S. Effects of hydrotherapy and land-based exercise on mobility and quality of life in patients with knee osteoarthritis: a randomized control trial. J Phys Ther Sci 2021 Apr;33(4):375-383 [FREE Full text] [doi: 10.1589/jpts.33.375] [Medline: 33935364]
46. Roggio F, Ravalli S, Maugeri G, Bianco A, Palma A, Di Rosa M, et al. Technological advancements in the analysis of human motion and posture management through digital devices. World J Orthop 2021 Jul 18;12(7):467-484 [FREE Full text] [doi: 10.5312/wjo.v12.i7.467] [Medline: 34354935]
47. De Vitis A, Battaglino A, Sinatti P. Effects of telemedicine for postural instability in independent patients with Parkinson's disease: a literature review. Top Geriatr Rehabil 2023;39(4):294-306. [doi: 10.1097/tgr.0000000000000413]
48. Griffa G, Solano A, Deleglise A, De Pino G, Jacobacci F, Della-Maggiore V. Uncovering the role of the human hippocampus in procedural motor learning: insights from implicit sensorimotor adaptation. J Neurosci 2025 Sep 03;45(36):e2119242025. [doi: 10.1523/JNEUROSCI.2119-24.2025] [Medline: 40744730]
49. Marshall JD, Aldarondo DE, Dunn TW, Wang WL, Berman GJ, Ölveczky BP. Continuous whole-body 3D kinematic recordings across the rodent behavioral repertoire. Neuron 2021 Feb 03;109(3):420-437.e8 [FREE Full text] [doi: 10.1016/j.neuron.2020.11.016] [Medline: 33340448]
50. Xu GW, Xie MR, Wu HR, Xiong R, Li CW, Xu SB, et al. A prospective study examining the impact of uniportal video-assisted thoracic surgery on the short-term quality of life in patients with lung cancer. Thorac Cancer 2020 Mar;11(3):612-618 [FREE Full text] [doi: 10.1111/1759-7714.13305] [Medline: 31967724]
51. Storebø OJ, Ribeiro JP, Kongerslev MT, Stoffers-Winterling J, Sedoc Jørgensen M, Lieb K, et al. Individual participant data systematic reviews with meta-analyses of psychotherapies for borderline personality disorder. BMJ Open 2021 Jun 21;11(6):e047416 [FREE Full text] [doi: 10.1136/bmjopen-2020-047416] [Medline: 34155077]
52. Liu M, Li Y, Liu X. Serum tumor necrosis factor-α, interleukin-1β, interleukin-6, and interleukin-17 relate to anxiety and depression risks to some extent in non-small cell lung cancer survivor. Clin Respir J 2022;16(2):105-115 [FREE Full text] [doi: 10.1111/crj.13457] [Medline: 34697903]
53. Zhong W, Liu R, Cheng H, Xu L, Wang L, He C, et al. Longer-term effects of cardiac telerehabilitation on patients with coronary artery disease: systematic review and meta-analysis. JMIR Mhealth Uhealth 2023 Jul 28;11:e46359 [FREE Full text] [doi: 10.2196/46359] [Medline: 37505803]
54. Li R, Li Y, Wang L, Li L, Fu C, Hu D, et al. Wearable activity tracker-based interventions for physical activity, body composition, and physical function among community-dwelling older adults: systematic review and meta-analysis of

XSL•FO
RenderX

randomized controlled trials. J Med Internet Res 2025 Apr 03;27:e59507 [FREE Full text] [doi: 10.2196/59507] [Medline: 40179387]
55. Wang HN, Luo P, Liu S, Liu Y, Zhang X, Li J. Effectiveness of internet-based telehealth programs in patients with hip or knee osteoarthritis: systematic review and meta-analysis. J Med Internet Res 2024 Sep 30;26:e55576 [FREE Full text] [doi: 10.2196/55576] [Medline: 39348685]

Abbreviations

ACL: Additional Concerns–Lung
AI: artificial intelligence
CST: Chair-Stand Test
EWB: emotional well-being
FACT-L: Functional Assessment of Cancer Therapy–Lung v4.0
FEV1: forced expiratory volume in 1 second
FTSST: Functional Mobility: Five-times Sit-to-Stand Test
FVC: forced vital capacity
FWB: functional well-being
NSCLC: non–small cell lung cancer
PEF: peak expiratory flow
PWB: physical well-being
6MWD: 6-minute walk distance
6MWT: 6-minute walk test
SWB: social/family well-being
TUG: Timed Up and Go

Edited by L Buis; submitted 15.Aug.2025; peer-reviewed by EA Sanchez-Romero, R Xing; comments to author 10.Feb.2026; revised version received 08.Mar.2026; accepted 05.May.2026; published 29.Jun.2026.

Please cite as:
Chen L, Xu X, Yang X, Long S, Chen N, Li Z, Yang T, Li Q, Li M, Zhong H, Huang G, Zeng Q
Effects of Telerehabilitation Based on Motion Recognition Technology on Exercise Endurance of Patients With Non–Small Cell Lung Cancer After Surgery: Single-Center, Prospective, Open-Label, Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e82447
URL: https://mhealth.jmir.org/2026/1/e82447
doi:10.2196/82447
PMID:

©Ling Chen, Xiaoyi Xu, Xiaoliang Yang, Siying Long, Na Chen, Zifan Li, Tingting Yang, Qianru Li, Meng Li, Hai Zhong, Guozhi Huang, Qing Zeng. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 29.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Smartphone-Based Interpretable Machine Learning for Classifying Single-Leg Squat Performance Using Trunk, Pelvic, and Knee Kinematics: Cross-Sectional Study

Sihyun Kim[1], PT, PhD; Kyuenam Park[2], PT, PhD

[1]Department of Physical Therapy, Sangji University, Wonju, Republic of Korea

[2]Digital Healthcare and Sports Data Science Lab, Department of Physical Education, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, Republic of Korea

Corresponding Author:
Kyuenam Park, PT, PhD
Digital Healthcare and Sports Data Science Lab, Department of Physical Education, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, Republic of Korea

Abstract

Background: Single-leg squat (SLS) performance is widely used to screen functional movement quality, but practical assessment often relies on expert visual grading or laboratory-based motion capture. In addition, conventional SLS criteria mainly focus on isolated joint deviations and may overlook coordination-related, multisegment movement patterns that characterize impaired performance.

Objective: This study aimed to examine the feasibility of an interpretable machine learning framework for classifying SLS performance into 3 levels (good, moderate, and poor) from single-smartphone, frontal-view videos based on trunk, pelvic, and knee kinematics, and to evaluate coordination-informed features and model explainability using Shapley additive explanations (SHAP) and local interpretable model-agnostic explanations (LIME).

Methods: A dataset of frontal-view SLS videos was labeled by physiotherapists into 3 functional categories (good, moderate, and poor). Videos were processed using 2D pose estimation, and models were trained on 17 engineered kinematic features derived from trunk, pelvic, and knee angles. Following the feature selection, 7 classifiers were trained and evaluated using the 8 selected features with stratified 5-fold cross-validation and a held-out test set. SHAP and LIME were applied for global and local interpretability.

Results: On the held-out test set, adaptive boosting classified SLS performance with an accuracy of 0.84, an F_1-score of 0.85, and an area under curve of 0.92. SHAP indicated that the summated angle (trunk + pelvis + knee), coordination-related features (knee × trunk interaction and knee-to-trunk ratio), and knee angle were key contributors to model predictions. LIME provided instance-level explanations that helped interpret individual classifications and decision boundaries.

Conclusions: This study presents an interpretable machine learning framework for classifying SLS performance into 3 levels using frontal-view videos acquired with a single smartphone. By leveraging coordination-informed engineered features and explainable artificial intelligence, the framework enables transparent interpretation of movement performance beyond isolated joint deviations. The proposed workflow uses smartphones for standardized video acquisition, while performance screening is achieved through machine learning. Given its lightweight feature design, this framework has potential for future implementation on modern smartphones and may support rehabilitation planning and injury-prevention strategies in sports and clinical settings.

(JMIR Mhealth Uhealth 2026;14:e85126) doi:10.2196/85126

KEYWORDS

explainable artificial intelligence; XAI; movement quality assessment; single-leg squat performance; smartphone-based assessment; supervised machine learning

Introduction

Knee function is a key factor in mobility, athletic performance, and quality of life in active young adults, and knee-related problems are common in this population [1]. In young physically active populations, functional performance tests provide meaningful evaluations of knee function by indirectly estimating muscle strength, joint stability, flexibility, balance, and proprioception, thereby offering insights that are not available through static radiographic imaging [2,3]. Among these, the single-leg squat (SLS) is widely used to assess dynamic lower extremity and trunk movement quality in both sports and clinical contexts, with movement quality rated on either a 2-point (good vs poor) or a 3-point ordinal scale (good vs moderate vs poor)

XSL•FO
RenderX

[4]. SLS performance is typically evaluated via expert visual observation of segmental alignment or using laboratory-based 3D motion capture to assess hip internal rotation, knee valgus, trunk lateral lean, and pelvic drop [5,6]. 3D motion capture can measure movement with high accuracy, but it requires costly equipment and substantial processing time compared to visual assessment [7]. Visual assessment, although less expensive, still necessitates expert analysis of recorded footage, making it time-consuming and susceptible to observer bias and limited reproducibility when classifying movement quality [5].

Recently, machine learning algorithms have been applied to evaluate and classify SLS performance without costly 3D motion capture systems or expert visual assessments, thereby reducing analysis time [7,8]. An inertial measurement unit (IMU)-based support vector machine (SVM) model using 3 sensors achieved 96% accuracy for binary classification (good vs poor) but decreased to 66% for 3-class classification (good vs moderate vs poor) based on hip and knee angles [9]. However, IMU accuracy depends on the number of sensors, and the requirement for multiple sensors limits clinical feasibility [9]. More recent approaches using consumer-grade cameras with pose estimation algorithms have achieved moderate-to-high accuracy in 3-class SLS classification, even with a single camera [7,10]. For example, pose-estimated SLS videos captured with a single camera achieved 53% - 77% agreement with expert assessments across 6 postural orientation errors, with knee valgus errors showing the highest agreement (77%) [7]. Similarly, a smartphone-based markerless motion capture study found that trunk, pelvic, and knee angles, together with the summated angle of these joints, enabled 3-class classification with 76.9% accuracy using a decision tree model [10]. However, these studies primarily extracted features based on expert-defined SLS guidelines, such as trunk lateral leaning and knee valgus, whereas SLS performance involves global, coordinated movement patterns of the trunk and lower limbs. It is necessary for future machine learning classification models to generate features that capture interjoint movement patterns, such as the knee-to-trunk ratio and joint interaction terms during the SLS test.

While recent machine learning approaches have demonstrated promising accuracy in SLS classification, the underlying decision-making mechanisms often remain unclear, which limits their adoption in clinical practice [11]. In rehabilitation and sports medicine, clinicians and other professionals need to understand which specific kinematic factors drive model predictions in order to design targeted interventions and maintain clinical confidence [12]. Explainable artificial intelligence (XAI) methods, such as Shapley additive explanations (SHAP) and local interpretable model-agnostic explanations (LIME), can provide feature-level explanations by quantifying the contribution of each joint movement or interaction to the model's decision [13]. Applying XAI to SLS performance assessment can both support movement quality classification and identify key movement patterns, bridging algorithmic predictions with clinical feedback and fostering trust and personalized rehabilitation planning.

In this study, we aim to examine the feasibility of an interpretable machine learning framework for classifying SLS performance based on trunk, pelvic, and knee kinematics into 3 levels (good, moderate, and poor) using frontal-view SLS videos recorded with a single smartphone. Using biomechanically informed feature engineering and hybrid feature selection, we further investigate whether coordination-related feature combinations (ratio, interaction, and summated features) involving trunk, pelvic, and knee angles provide clinically meaningful representations of movement quality beyond conventional single-joint angle metrics. In addition, we evaluate model interpretability using SHAP and LIME to support transparent functional grading and facilitate clinical translation in sports and rehabilitation screening.

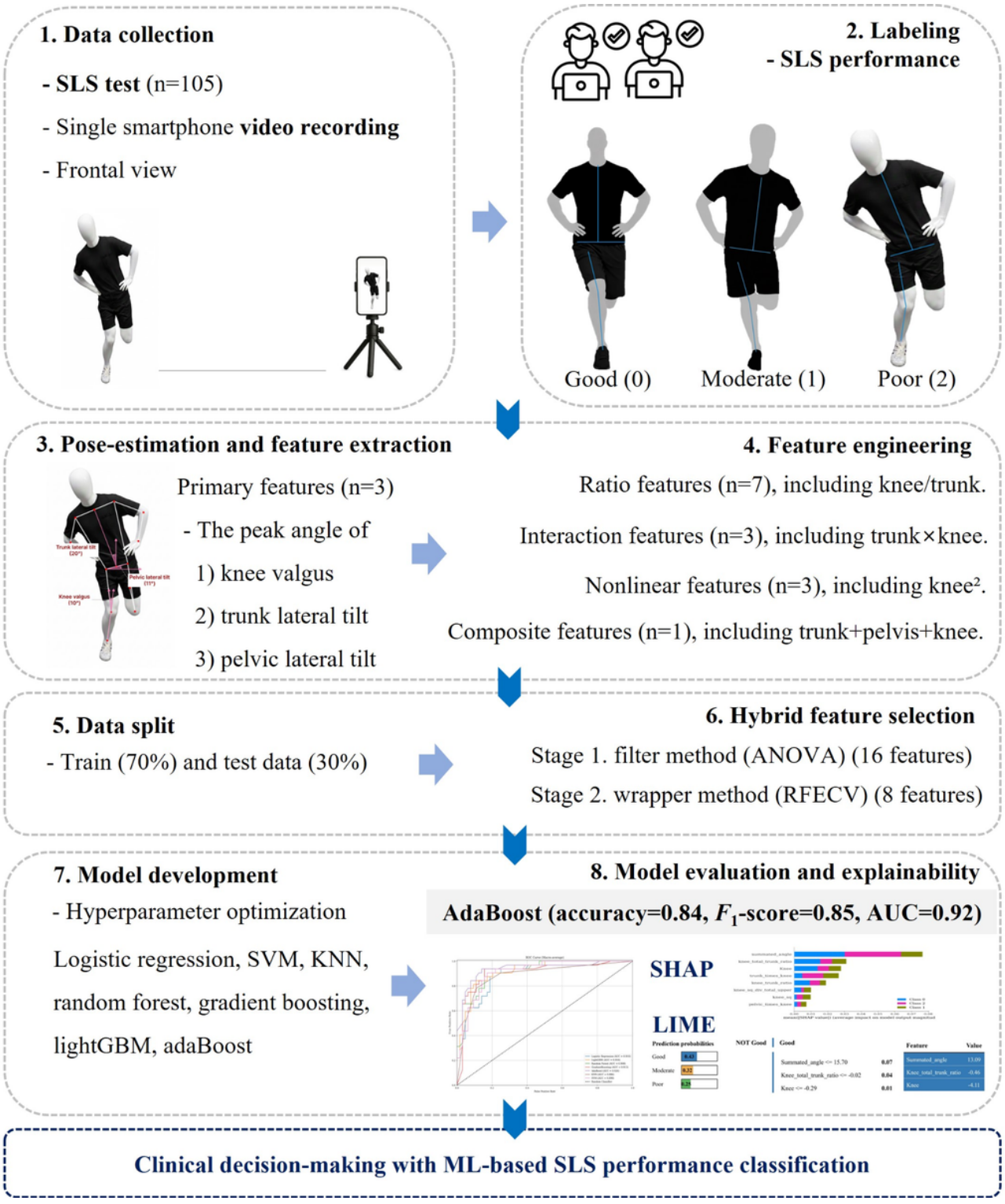
Methods

Participants

An overview of the overall study workflow is presented in Figure 1. This study included a total of 105 young active adults (aged 19 - 30 y) who were recruited through university campus advertisements, social media, and local community bulletin boards. Participants were eligible for inclusion if they engaged in regular physical activity for at least 60 minutes per week. Exclusion criteria included a history of lower-extremity or low back pain within the previous 3 months, prior major knee injuries or surgeries, neurological disorders, or an inability to perform the SLS test [14]. Sample size adequacy for machine learning model development was evaluated using a prediction instability approach. Prediction instability was assessed across incremental proportions of the dataset (10%, 30%, 50%, 70%, and 90%), where instability was defined as the variability of model performance across resampled subsets [15]. The analysis showed a marked reduction in prediction instability with increasing sample size, with the final sample of 105 SLS videos yielding stable model performance across the evaluated machine learning models. Bootstrap resampling with 100 iterations indicated that using more than 70% of the dataset resulted in acceptable prediction stability, as reflected by low instability values, supporting the adequacy of the sample size for reliable machine learning performance (Figure 2).

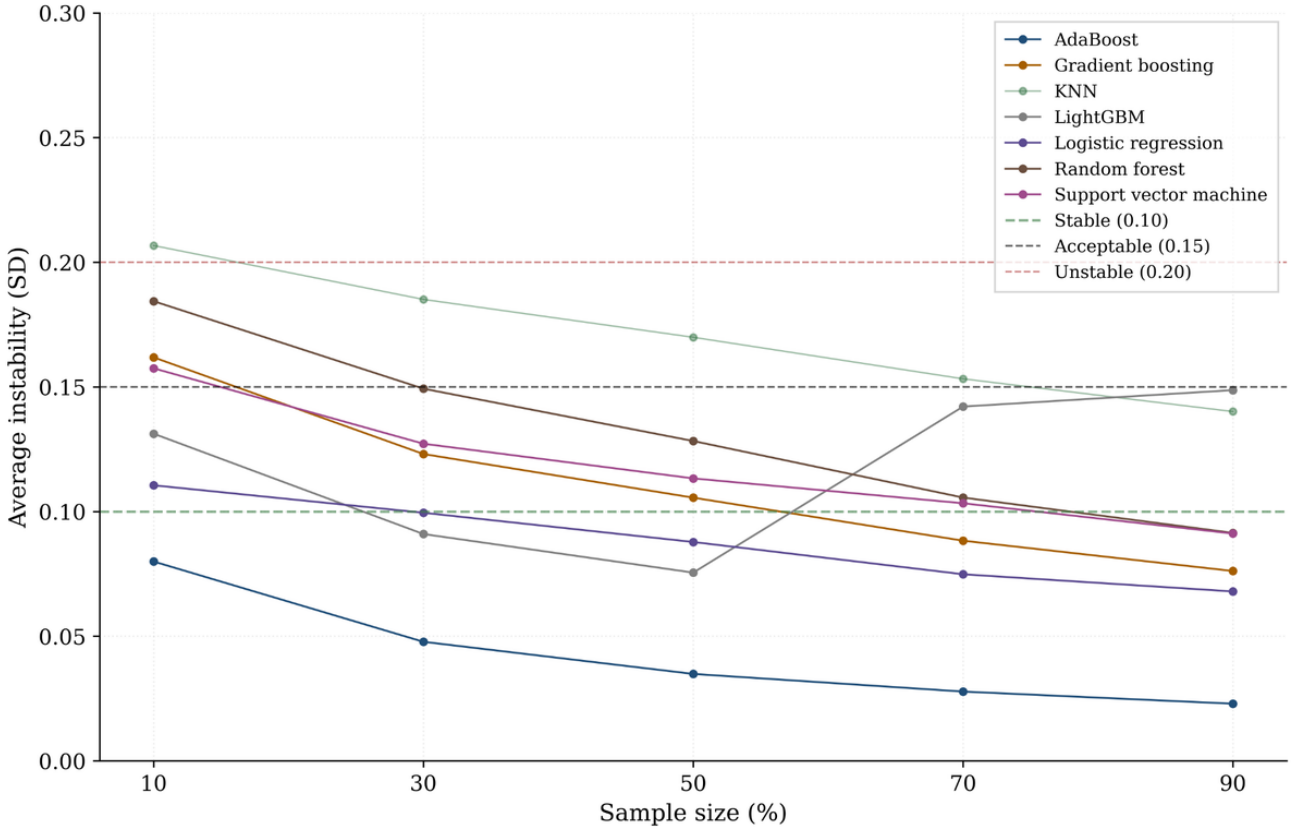
XSL•FO
RenderX

Figure 1. Flowchart of the machine learning model development process for SLS performance classification and interpretable feedback. AdaBoost: adaptive boosting; AUC: area under curve; KNN: *k*-nearest neighbors; LightGBM: light gradient boosting machine; LIME: local interpretable model-agnostic explanations; ML: machine learning; RFECV: recursive feature elimination with cross-validation; SHAP: Shapley additive explanations; SLS: single-leg squat; SVM: support vector machine.



XSL•FO
RenderX

Figure 2. Prediction instability across varying sample sizes for multiple machine learning models evaluated using bootstrap resampling. AdaBoost: adaptive boosting; KNN: *k*-nearest neighbors; LightGBM: light gradient boosting machine.



Data Collection

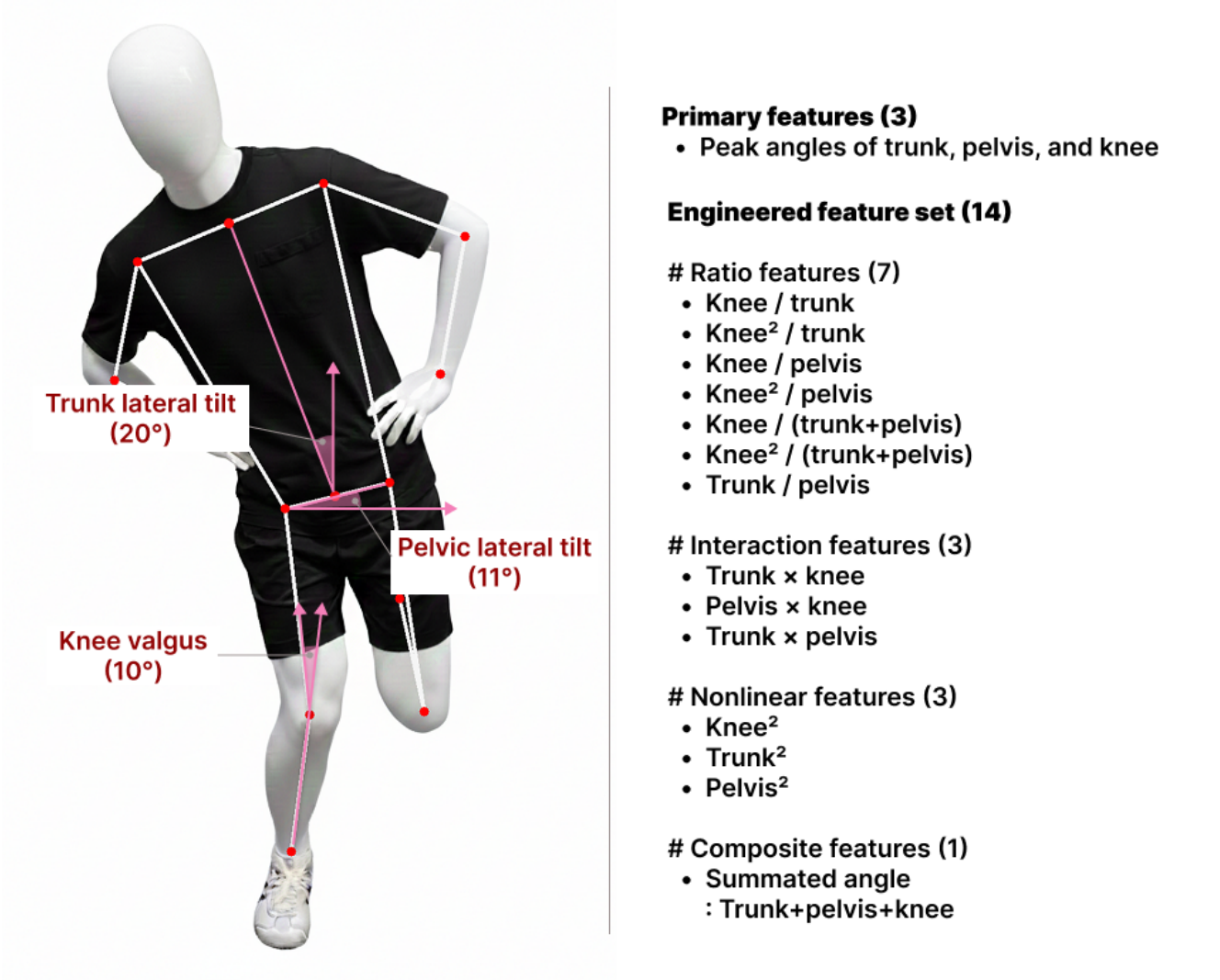
Videos were recorded using an iPhone, mounted on a tripod positioned 3 meters in front of the participant at pelvic height, capturing the entire body at 30 Hz and 1080 p resolution. Only the frontal view was recorded, as clinical evaluation of SLS performance primarily relies on frontal plane deviations including trunk lean, pelvic tilt, and medial knee displacement [7]. Additionally, this single-camera configuration reflects a practical and scalable approach for real-world clinical assessment settings, eliminating the need for multiple camera setups to capture additional views (side or top) [16].

Each participant performed the SLS on their dominant leg (defined as the leg used to kick a ball), wearing standard gym clothing (eg, shorts and t-shirts) to ensure full visibility of lower-limb motion. Participants were given a brief warm-up consisting of 5 minutes of stationary cycling and performed 3 practice trials under the supervision of a physiotherapist to ensure consistent technique. The SLS movement was performed at a comfortable, self-selected pace, avoiding excessively slow speed that might reduce compensatory kinematic deviations such as knee valgus [17]. The standardized SLS test procedure was as follows: (1) Participants stood with their feet shoulder-width apart and toes pointing forward; (2) Participants then placed their hands on the waist and extended the nondominant leg backward, maintaining approximately 90° of knee flexion; (3) From this position, participants descended into a squat as deeply as possible at a self-selected speed, bending the weight-bearing leg without losing balance, maintaining an upright posture, and facing the front camera (Figure 3); and (4) Finally, they fully extended the knee of the weight-bearing leg to return to the starting position [18].

Each participant performed 3 repetitions, with a 1-minute rest between attempts. A repetition was considered invalid if the participant lost balance, allowed the legs to touch, placed the nondominant foot on the ground, or failed to maintain an upright posture facing the camera. In such cases, an additional repetition was performed at the end of the trial. If 2 or more repetitions within a trial were invalid, the entire trial was repeated after a sufficient rest period [17].

XSL•FO
RenderX

Figure 3. Representative single-leg squat frame and interpretable kinematic features used for model training.



Labeling of Single-Leg Squat Movement Quality

Movement quality was evaluated independently by 2 licensed physiotherapists, each with over 10 years of clinical experience in musculoskeletal rehabilitation. Prior to labeling, the evaluators completed a structured 2-hour training session on SLS assessment criteria, which involved repeatedly reviewing and grading 5 representative SLS video recordings excluded from the main dataset [19]. During this session, a consensus discussion with one of the authors was conducted to standardize the interpretation of movement patterns and rating procedures, thereby ensuring consistent and objective classification.

Following training, the 2 evaluators independently assessed movement quality through visual observation while blinded to participant characteristics and not involved in other experimental procedures. Each participant's performance was reviewed no more than 2 times using full-length SLS videos played at normal speed on a personal computer, without pausing, rewinding, or slow-motion playback [19]. This approach was informed by previous findings indicating that even when raters are allowed unlimited repeated viewing in both real-time and slow-motion conditions, interrater reliability does not necessarily improve beyond fair to moderate levels for several movement components [20]. During each review, performance was assessed using a multisegmental approach, simultaneously evaluating the trunk, pelvic, and knee segments across 3 consecutive squats [19]. SLS performance was graded into 1 of 3 categories based on observable frontal-plane deviations [21]. Class 0 (good) was defined by no lateral trunk or pelvic tilt, absence of visible knee valgus, and a stable, straight knee trajectory throughout the squat. Class 1 (moderate) included mild lateral trunk or pelvic tilt with slight knee valgus and minor medial-lateral knee movement without notable instability. Class 2 (poor) was characterized by pronounced lateral trunk or pelvic tilt, obvious knee valgus, and clear mediolateral displacement or shaking of the knee.

Each repetition was independently and blindly assessed by 2 evaluators. Agreement scores were accepted when ratings matched. In cases of disagreement, the evaluators re-reviewed the video recording, shared their initial ratings, and discussed the observed movement patterns to reach agreement on a final score [7]. Before consensus, interrater agreement for the 3-level SLS performance grading was assessed using weighted Cohen Kappa, indicating a moderate level of agreement (κ=0.60).

The final class label for each participant was defined as the poorest performance observed across the 3 repetitions [7,21]. The poorest SLS repetition was used for machine learning

XSL•FO
RenderX

training to represent participant-level movement quality because clinical screening prioritizes detection of the most impaired movement pattern. In addition, when multiple repetitions from a single participant receive different labels, trial-level modeling can introduce ambiguity in participant-level classification. When multiple repetitions received the same label, the first valid repetition was used for pose estimation and feature extraction to ensure a deterministic and reproducible trial-selection rule and to minimize potential learning and fatigue effects across repeated trials [22].

Pose Estimation and Feature Extraction

For consistency between labeling and analysis, the same poorest-rated trial used for labeling was also selected for 2D pose estimation and feature extraction. Pose estimation of the SLS smartphone video was performed using the MediaPipe Pose library (v0.10.21) (Google), which detects 33 body landmarks from RGB frames in real time. MediaPipe Pose operates without the need for depth sensors, specialized hardware, or manual landmark initialization, making it well-suited for smartphone-based clinical assessments. Previous research has demonstrated that MediaPipe Pose provides highly valid estimates of lower-limb joint angles compared with gold-standard 3D motion capture (Pearson correlation=0.87) and exhibits excellent test-retest reliability (intraclass correlation coefficient=0.93 - 0.95), supporting its suitability for clinical biomechanics applications [23].

In accordance with established SLS test guidelines [21], trunk, pelvic, and knee angles were computed frame by frame across the entire SLS movement cycle. For each joint, the peak deviation from neutral alignment was extracted and used to quantify frontal-plane deviations during the SLS test. Knee valgus was computed as 180° minus the internal angle formed by the hip, knee, and ankle landmarks, representing frontal-plane knee alignment. Trunk lateral tilt was defined as the maximum deviation angle between the vertical axis and the line connecting the shoulder and hip midpoints, representing upper body lean. Pelvic lateral tilt was measured as the maximum angle between the line connecting the left and right hip landmarks and the horizontal axis, reflecting pelvic obliquity (Figure 3).

While the primary joint angles align with conventional SLS evaluation criteria [21], they may overlook complex inter-segmental relationships and compensatory movement strategies. To expand the feature representation beyond the primary measures, 3 frontal-plane joint angles (trunk, pelvic, and knee) were used to derive 14 engineered kinematic features. The engineered feature set included ratio features (knee/trunk, knee^2/trunk, knee/pelvis, knee^2/pelvis, knee/[trunk + pelvis], knee^2/[trunk + pelvis], and trunk/pelvis), interaction terms (trunk × knee, trunk × pelvis, and pelvis × knee), and nonlinear terms (trunk^2, pelvis^2, and knee^2). In addition, a composite feature (summated angle; trunk + pelvis + knee) was calculated. In total, 17 features (3 primary joint angles and 14 engineered variables) were used for subsequent model training. These ratios, interaction, nonlinear, and composite features were designed to capture coordination-related movement patterns beyond isolated joint deviations, potentially improving discrimination across knee-function levels and supporting clinical interpretability.

Feature Selection

The dataset was divided into training and test sets using stratified sampling, with 70% (n=73) allocated for training and 30% (n=32) for independent testing, ensuring proportional representation of the 3 movement quality classes (good, moderate, and poor). All models were trained and evaluated using the original class distribution without oversampling or synthetic data generation techniques.

Feature selection was performed using a 2-stage hybrid approach to identify features that exhibited significant between-class differences while preserving biomechanically meaningful feature diversity. First, a filter-based method using a 1-way ANOVA F_{df}-test with Benjamini-Hochberg false discovery rate correction was applied, resulting in the selection of 16 features that showed significant differences across the 3 movement quality classes (P<.05), with the trunk-to-pelvis ratio excluded because it did not show statistically significant between-class differences. Next, a wrapper-based method using recursive feature elimination with cross-validation and a random forest classifier was applied to the filtered feature set. Recursive feature elimination with cross-validation iteratively removed features that did not improve cross-validated classification performance, helping to mitigate functional redundancy among correlated variables derived from the same primary joint angles. Five-fold cross-validation was used with classification accuracy as the optimization metric. The final feature set was obtained by intersecting the results of both methods, resulting in 8 complementary features with reduced redundancy, including the knee valgus angle, ratio-based features (knee/trunk, knee/[trunk + pelvis], and knee^2/[trunk + pelvis]), interaction terms (trunk × knee and pelvis × knee), nonlinear knee-related features (knee^2), and a composite summated angle (trunk + pelvis + knee).

Classifiers

All implementations were conducted in Python 3.9 (Python Software Foundation) using standard machine learning libraries. For models that are sensitive to feature scaling, such as logistic regression, SVM, and *k*-nearest neighbors (KNN), *z*-score normalization was applied within a pipeline, while tree-based models were trained on raw feature values. Seven classifiers, including logistic regression, SVM, KNN, random forest, gradient boosting, light gradient boosting machine (LightGBM), and adaptive boosting (AdaBoost), were tuned via grid search with stratified 5-fold cross-validation using the macro-averaged F_1-score as the optimization metric. Hyperparameter search ranges included the regularization type and strength for logistic regression, the kernel type and penalty parameters for SVM, the number of neighbors and distance metrics for KNN, the tree depth and number of estimators for ensemble tree methods, and the learning rate for boosting algorithms.

Model Evaluation

Model performance was evaluated using stratified 5-fold cross-validation on the training set and an independent held-out test set. Three primary metrics were used to assess classification performance: accuracy, macro-averaged $F1$-score, and multiclass area under the receiver operating characteristic curve (AUC).

XSL•FO
RenderX

The macro-averaged F_1-score was computed as the unweighted mean of the class-specific F_1-scores, ensuring balanced evaluation across the 3 movement quality categories. Multiclass AUC was calculated using the one-vs-rest strategy, averaging the AUC scores obtained for each class.

Model Explainability

Both SHAP and LIME analyses were conducted exclusively on the final trained model, ensuring consistency between the interpretability results and the model used for evaluation. This dual approach enabled the identification of class-specific feature importance at the global level and the extraction of instance-specific decision patterns at the local level, supporting potential clinical and fitness applications [24,25]. Global interpretability was assessed using SHAP, applied to the final machine learning model. SHAP values were computed for each of the 8 selected features to quantify feature contributions to the model's class predictions. Class-wise SHAP value aggregation was used to identify the most influential features for each movement quality category.

Local interpretability was examined using LIME to generate instance-level explanations. LIME was applied to representative samples from each class in the test set, with the number of perturbed samples set to 5000 to ensure stability of the local surrogate models. The output from LIME was used to identify feature value ranges for each of the selected features that were most predictive of class membership. These thresholds were derived from the feature values at the decision boundaries learned by the model, providing interpretable biomechanical markers for distinguishing between "good," "moderate," and "poor" performance.

Computational Feasibility

Model training was performed offline as a 1-time procedure using a standard central processing unit-based environment (Google Colab) without graphics processing unit acceleration. Due to the use of lightweight classifiers and a limited set of interpretable features, training typically completed within a few seconds. Inference was conducted at the participant level, with each video processed independently. The inference pipeline, which included pose estimation, feature extraction, and classification, typically required only a few seconds per video on a central processing unit. In the proposed framework, the smartphone was used primarily for video acquisition, while pose estimation, feature computation, and classification were performed on a personal computer. These findings indicated that the proposed pipeline has modest computational requirements and is feasible for practical implementation.

Statistical Analysis

Descriptive statistics were computed for all participant characteristics and kinematic variables. Continuous data were reported as mean (SD), and categorical data were summarized as counts. Normality was assessed using the Shapiro-Wilk test. Differences among the 3 functional groups (good, moderate, and poor) were examined using 1-way ANOVA for continuous measures (age, BMI, weekly exercise duration, and peak frontal-plane angles). When a significant group effect was detected, Tukey honestly significant difference post hoc tests were conducted for pairwise comparisons. Statistical significance was defined as an α level of .05.

Ethical Considerations

The study protocol was reviewed and approved by the Institutional Review Board of Jeonju University (jjIRB-210315-HR-2021-03210) in accordance with the Declaration of Helsinki and relevant ethical guidelines. All participants were fully informed about the study's purpose, procedures, potential benefits, and rights prior to participation. Written informed consent was obtained from all participants prior to enrollment. All smartphone video data were collected solely for research purposes and were securely stored on an encrypted, password-protected server accessible only to the research team. To ensure privacy and confidentiality, all video and related data were deidentified; facial information and other personally identifiable details were neither analyzed nor disclosed. Participants were not asked to provide their real names; instead, each participant was assigned a unique anonymized code. Each participant received a small monetary incentive (approximately US $10) for participating in the study.

Results

Participant Characteristics

Participant characteristics are summarized in Table 1. The study included participants across the 3 SLS performance classification grades (good, moderate, and poor). While age, BMI, and weekly exercise duration did not significantly differ across groups, peak frontal-plane kinematics, including trunk lateral tilt, pelvic lateral tilt, and knee valgus, showed significant between-group differences (P<.05). Post hoc Tukey tests showed that trunk lateral tilt was significantly greater in the poor group than in the good and moderate groups, while no significant difference was observed between the good and moderate groups. Pelvic lateral tilt differed significantly between the good group and the other groups, whereas knee valgus increased stepwise across all 3 grades (P<.05).

XSL•FO
RenderX

Table . Participants' characteristics (n=105).

	SLS[a] performance grading criteria			
Characteristics	Good (0)	Moderate (1)	Poor (2)	*P* value
Number of participants, n	28	41	36	—[b]
Gender (men/women), n	13/15	21/20	19/17	—
Dominant side (right/left), n	26/2	40/1	35/1	—
Age (y)	22.36 (2.03)[c]	22.46 (2.85)	22.20 (2.47)	.75
BMI (kg/m^2)	22.38 (3.35)	22.80 (3.59)	23.22 (3.33)	.53
Exercise duration per week (min)	445.42 (286.15)	353.25 (339.16)	331.57 (215.35)	.38
Trunk lateral tilt (°)[def]	6.00 (2.47)	7.24 (3.23)	10.24 (6.34)	<.001
Pelvic lateral tilt (°)[dfg]	6.31 (1.93)	8.50 (2.76)	10.38 (4.94)	<.001
Knee valgus (°)[defg]	−2.63 (2.55)[h]	3.67 (3.75)	13.24 (9.71)	<.001

[a]SLS: single-leg squat.

[b]Not available.

[c]Mean (SD).

[d]Peak joint angles measured during the single-leg squat test.

[e]Post hoc showed a significant difference between moderate and poor groups (P<.05).

[f]Post hoc showed a significant difference between good and poor groups (P<.05).

[g]Post hoc showed a significant difference between good and moderate groups (P<.05).

[h]Negative values represent lateral deviation (knee varus) from neutral alignment, whereas positive values represent medial deviation (knee valgus).

Classification Performance of Machine Learning Models

Seven individual classifiers were evaluated using stratified 5-fold cross-validation, and final performance was assessed on a held-out test set (Table 2). AdaBoost demonstrated the best overall performance (AUC=0.92), comparable to LightGBM (AUC=0.92) and slightly higher than gradient boosting (AUC=0.91) (Figure 4). Accordingly, AdaBoost was selected as the final classifier for subsequent analysis. To confirm the value of feature selection, we compared AdaBoost performance using the full 17-feature pool versus the final 8 selected features. The AUC remained comparable (0.92 vs 0.92), indicating that the compact feature set retained most of the predictive information while improving model simplicity and interpretability.

The confusion matrices for AdaBoost, LightGBM, and gradient boosting are presented in Figure 5. Overall, AdaBoost showed more consistent class-wise discrimination, with reduced confusion between adjacent functional grades (particularly moderate and poor) compared with the other models, consistent with its overall performance metrics.

Table . Performance metrics of the evaluated classifiers.

Classifier	Accuracy	F_1-score	AUC[a]	Sensitivity	Specificity
Logistic regression	0.69	0.70	0.91	0.69	0.83
Support vector machine	0.75	0.76	0.90	0.75	0.83
k-nearest neighbors	0.78	0.79	0.89	0.79	0.88
Random forest	0.75	0.76	0.90	0.75	0.87
Gradient boosting	0.78	0.80	0.91	0.79	0.89
LightGBM[b]	0.81	0.82	0.92	0.81	0.90
AdaBoost[c]	0.84	0.85	0.92	0.84	0.92

[a]AUC: area under curve.

[b]LightGBM: light gradient boosting machine.

[c]AdaBoost: adaptive boosting.

XSL•FO
RenderX

Figure 4. ROC curves and corresponding AUC values. AdaBoost: adaptive boosting; AUC: area under the curve; KNN: *k*-nearest neighbors; LightGBM: light gradient boosting machine; ROC: receiver operating characteristic; SVM: support vector machine.

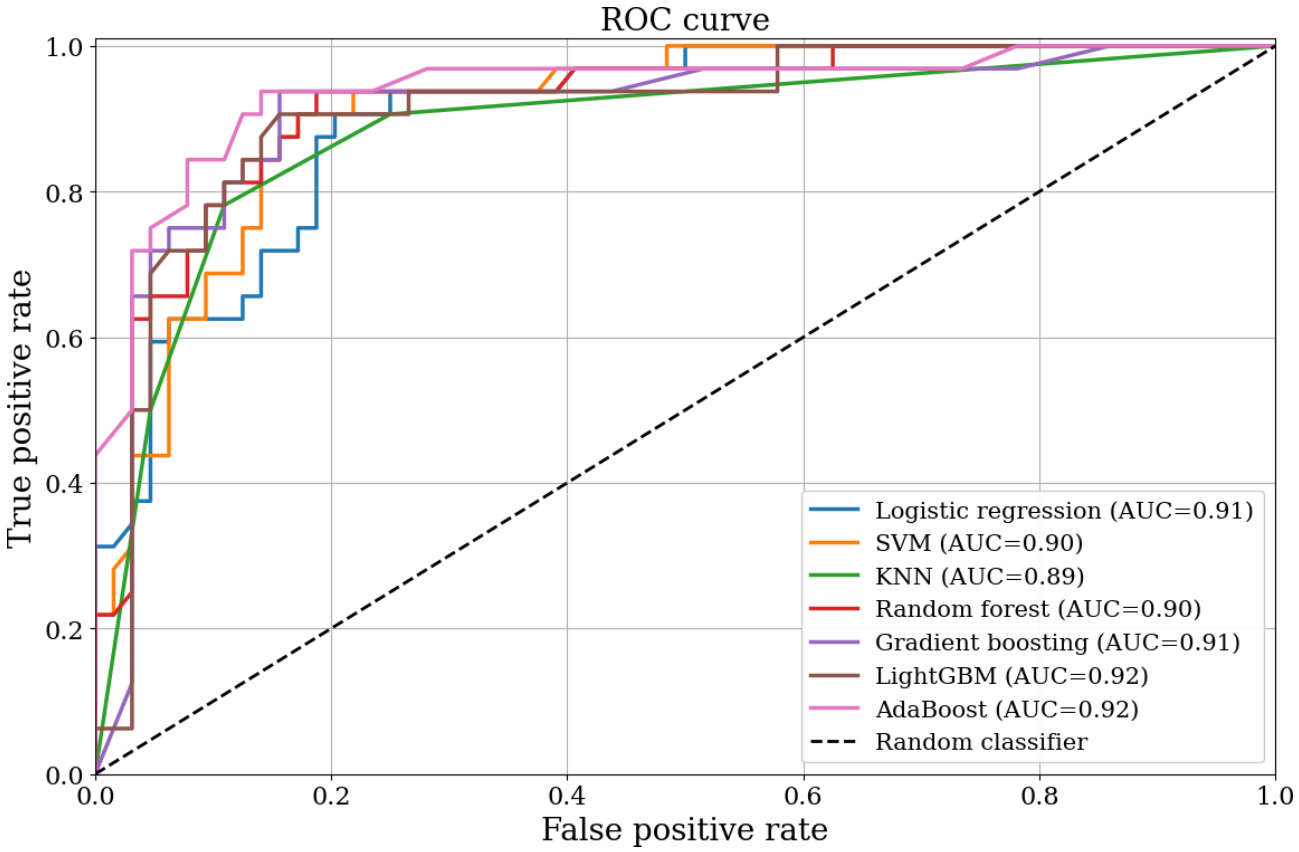
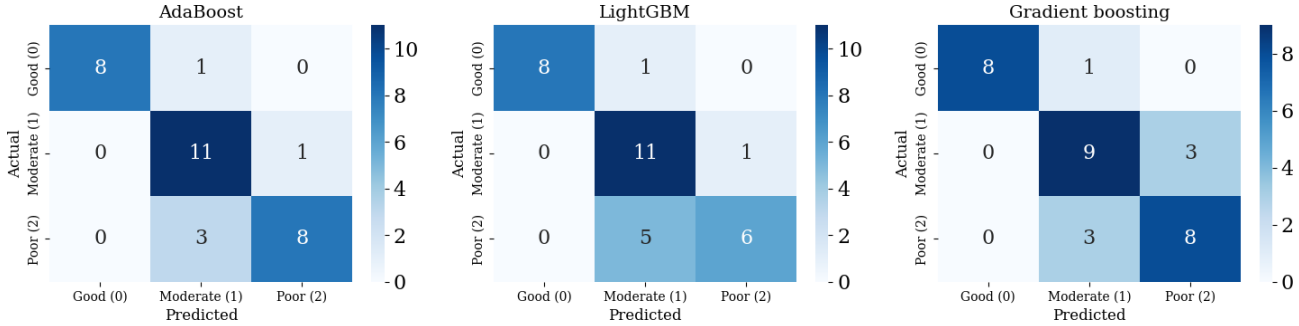


Figure 5. Confusion matrices of AdaBoost, LightGBM, and gradient boosting models. AdaBoost: adaptive boosting; LightGBM: light gradient boosting machine.



Key Feature Analysis and Model Interpretability

To interpret the final classification model, we conducted both global and local interpretability analyses. SHAP identified the summated angle as the most influential feature, followed by the trunk × knee interaction, the knee-to-trunk ratio, and the knee angle (Figure 6). These results suggested that model predictions were influenced by both isolated knee deviations and intersegment coordination patterns, rather than conventional guideline-based single-joint criteria alone. For local interpretability, LIME provided case-specific explanations by visualizing how these features contributed to individual classifications across the 3 functional grades (Figure 7).

XSL•FO
RenderX

Figure 6. SHAP summary plot showing global feature importance for the final classification model, with class labels representing single-leg squat performance levels (class 0=good, class 1=moderate, class 2=poor; total trunk=trunk + pelvis). SHAP: Shapley additive explanations.

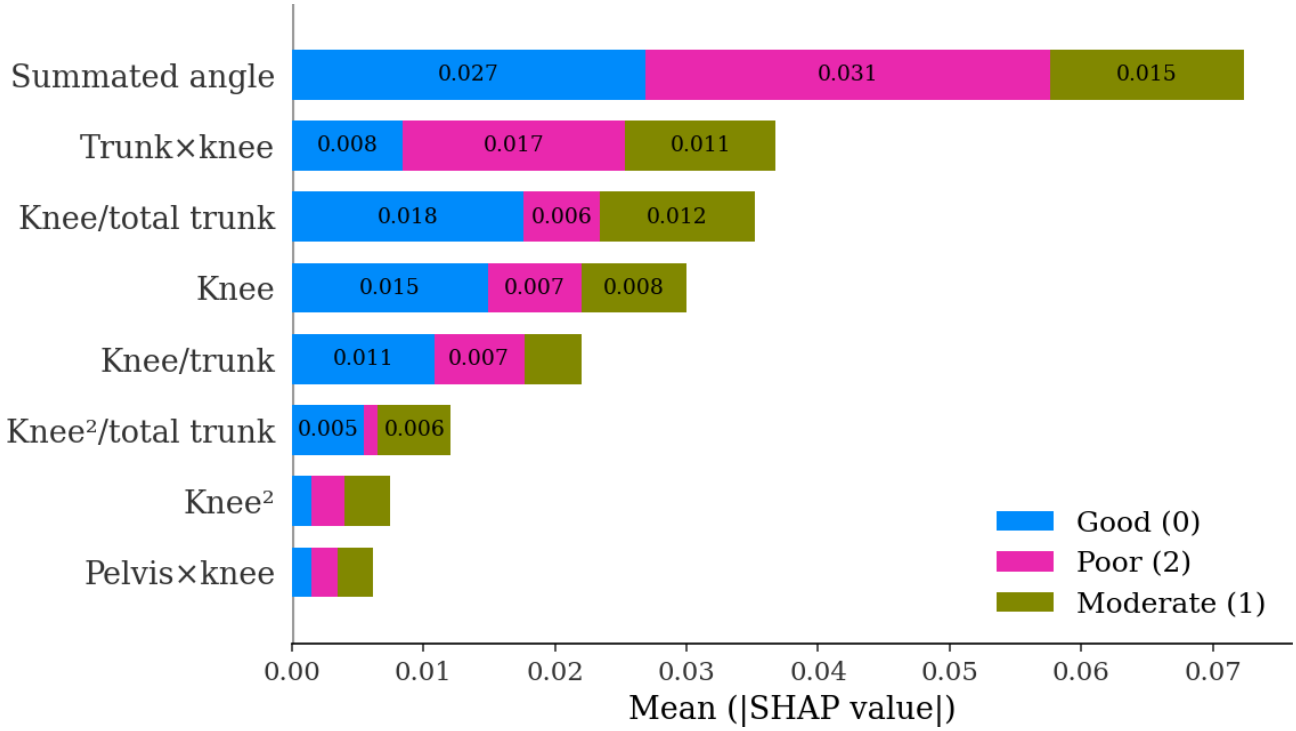
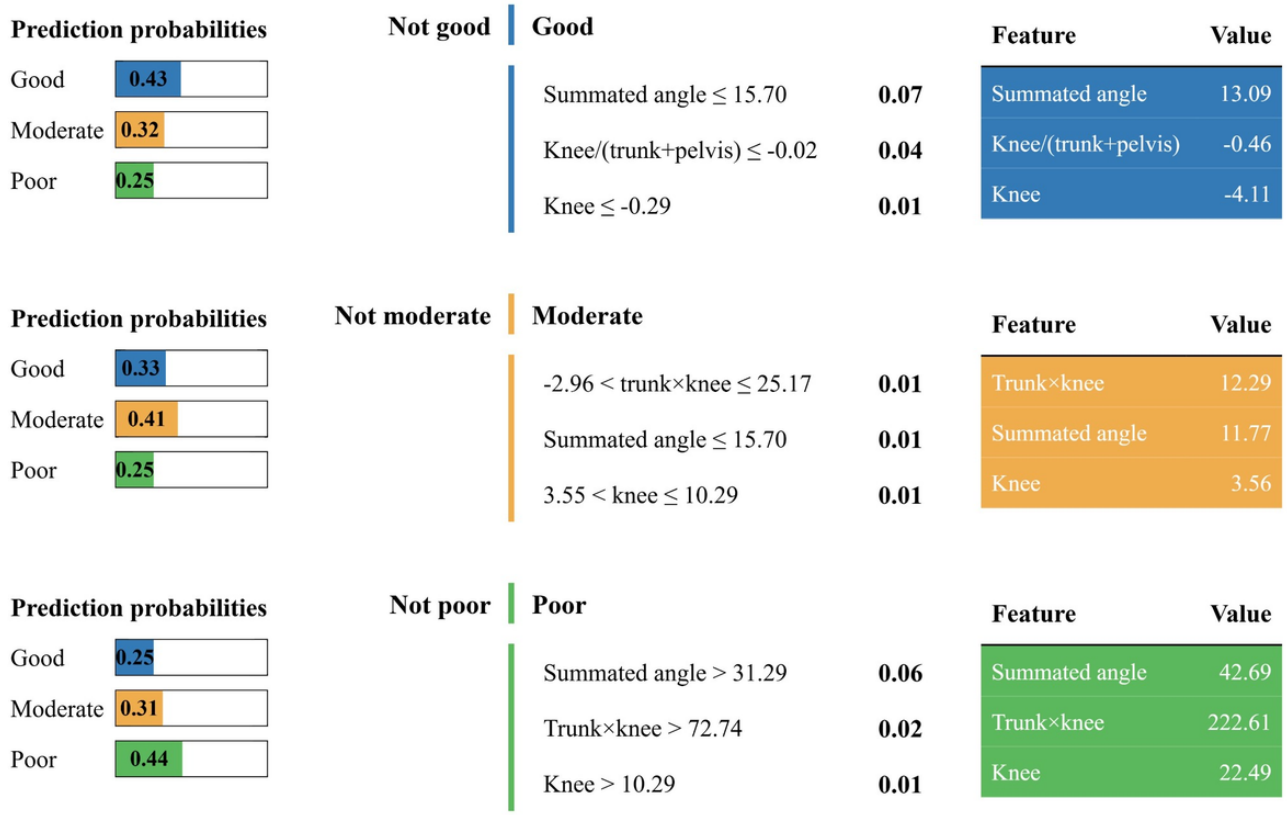


Figure 7. Local interpretable model-agnostic explanations-based instance-level explanations of feature contributions for representative samples from each class (good, moderate, and poor) classified by the final model.



Discussion

Principal Findings

This study developed and evaluated an interpretable machine learning framework that classifies SLS performance into 3 levels (good, moderate, and poor) from single-smartphone, frontal-view SLS videos. Using a small set of biomechanically meaningful engineered features (summated angle, knee × trunk interaction, knee-to-trunk ratio, and knee angle), the AdaBoost classifier showed the best overall performance on a held-out

XSL•FO
RenderX

test set (accuracy=0.84, F_1-score=0.85, and AUC=0.92). SHAP and LIME enabled feature-based interpretation, allowing classification results to be translated into clinically interpretable feedback rather than labels alone. Importantly, the primary contribution of this study lies not in proposing a new algorithm but in demonstrating through a machine learning approach that impaired SLS performance is associated with coordination-related movement patterns rather than isolated joint deviations. By emphasizing the relative and interactive organization of trunk and knee movements, this coordination-based interpretation complements conventional guideline-driven SLS assessment and provides practical insights for rehabilitation decision-making and athletic screening.

Previous studies have reported that both wearable sensor-based and camera-based markerless models achieved moderate accuracy, typically ranging from 66% to 99% with an AUC of approximately 0.73, and often showed limited generalizability in 3-class SLS classification [9-11,26]. Notably, a recent study reported classification accuracies close to 99% using a single thigh-mounted IMU and a convolutional neural network-gated recurrent unit model trained on experimentally induced squat error patterns [26]. However, this IMU-based convolutional neural network-gated recurrent unit approach operates as a largely black-box model and relies on experimentally induced error motions (knee valgus/varus or excessive trunk flexion/extension), which may limit interpretability and ecological validity for clinical decision-making and real-world functional screening. In contrast, the present study emphasizes clinical translation by training the model on naturally performed SLS trials, with expert clinicians selecting representative videos that reflect real-world movement quality rather than artificially induced deviations. To move beyond guideline-based or composite measures, we generated ratio-, interaction-, and nonlinear-based features to capture interjoint coordination and identified summated angle, knee × trunk, knee-to-trunk ratio, and knee angle as the most informative predictors, thereby broadening the biomechanical basis for SLS assessment. By combining expert-labeled functional outcomes with interpretable feature representations learned from natural, nonexperimentally induced SLS performances, the proposed framework was designed to support clinical reasoning with transparent decision logic. Furthermore, machine learning models must be trained on data that are realistically obtainable in routine practice; otherwise, even high-performing algorithms remain difficult to translate when they depend on non-consumer sensors, multiple cameras, or specialized laboratory setups [27]. The present framework addresses this limitation by relying exclusively on a single, frontal smartphone video to extract interpretable kinematic features during standard SLS testing without specialized IMU sensor. This design enables automatic classification of SLS performance levels while maintaining practical feasibility, interpretability, and direct applicability in real-world sports and rehabilitation settings. Therefore, the key novelty lies not only in performance metrics but also in an ecologically valid and explainable approach that can be realistically implemented using smartphone video in routine screening workflows.

Beyond predictive accuracy, the explainability analyses provided additional biomechanical insights into SLS performance. SHAP highlighted the summated angle as the most influential feature, with additional contributions from knee-specific and coordination-related metrics, including the knee × trunk interaction, the knee-to-trunk ratio, and knee angle. These findings suggest that impaired performance is associated with multisegment movement patterns, rather than isolated joint deviation alone. Previous studies using individual joint or composite measures achieved only moderate accuracy (66% with an SVM and 76.9% with a decision tree) [9,10], without considering intersegmental interactions. In contrast, our approach explicitly incorporated coordination-related engineered features and combined them with an interpretable machine learning model, enabling biomechanically grounded explanations beyond conventional joint-by-joint criteria. This observation aligns with biomechanical evidence that interjoint coordination patterns are critical markers of movement quality during squatting and single-leg tasks [7], and supports the role of trunk-knee coupling as a central determinant of functional impairment [28]. Complementary LIME analyses further illustrated instance-specific decision boundaries, showing that case-level classifications were consistently explained by combinations of summated angle, coordination-related metrics, and knee angle supporting transparent subject-level interpretation. By clarifying why a patient is classified as having poor SLS performance, the XAI results may help clinicians and trainers gain confidence in machine learning-based SLS evaluation. Moreover, they can guide expert decisions on whether management strategies should focus solely on the knee or also address altered trunk-knee coordination, thereby supporting more personalized rehabilitation and remote monitoring.

Practical Implications

The practical meaning of this study is that it enables a low-cost, scalable, and explainable workflow for screening SLS performance using smartphone-recorded SLS videos, providing interpretable decision support rather than simple movement monitoring. In team sports, athletic trainers can apply this workflow for periodic screening, reducing the burden of manually reviewing large volumes of videos, grading SLS performance, and producing narrative reports for multiple athletes. Importantly, SHAP- and LIME-based explanations support interpretation beyond the final grade by clarifying whether impaired performance is driven primarily by knee deviation, altered trunk control, or their coordination-related interaction, thereby informing targeted preventive or corrective strategies. In physical therapy settings, clinicians can similarly use the framework during intake or follow-up visits to obtain an objective functional grade with interpretable feature-based rationale, supporting individualized rehabilitation planning and progress tracking. Compared with rule-based fitness screening applications, the proposed machine learning approach better captures coordination-related movement patterns and provides consistent decision logic with interpretable explanations, supporting standardized documentation and scalable musculoskeletal screening in sports and rehabilitation settings.

XSL•FO
RenderX

Limitations

Despite these promising results, several limitations should be acknowledged. First, the dataset size was relatively small (n=105 participants). However, prediction instability analyses suggested that model performance stabilized with the available sample [15], and stratified cross-validation and systematic feature selection were applied to assess the stability of the results [13]. Nevertheless, future studies should validate the model in larger cohorts with external testing to quantify generalizability across populations and real-world settings. Second, all videos were recorded using the same smartphone under a standardized frontal-view setup with fixed camera distance and height, consistent with routine clinician-guided SLS screening protocols. While this standardized design improves reliability and reduces measurement variability, it may limit generalizability to home- or telehealth-based recordings where camera setup and recording conditions are less controlled. In particular, variations in smartphone camera specifications and video quality (resolution and frame rate), as well as real-world environmental factors such as lighting and background clutter, may affect pose estimation accuracy and introduce noise into extracted kinematic features. Future work should quantify sensitivity by systematically varying recording conditions (smartphone model, lighting, background, and camera distance or height) and examining the impact on pose estimation stability and final classification performance. Third, participants were limited to young healthy adults, which may restrict generalizability to clinical populations and athletic cohorts. Future studies should evaluate performance in more diverse groups, including older adults, patients with knee-related conditions, and injured athletes, to support clinical translation. Finally, because the dataset showed a right-skewed dominant-side distribution (Table 1), future studies should include more left-dominant participants and evaluate left-right normalization or symmetry-based augmentation strategies to improve model robustness.

Conclusion

This study developed an interpretable machine learning framework for classifying SLS performance into 3 levels (good, moderate, and poor) from frontal-view SLS videos recorded with a single smartphone. Using a small set of biomechanically meaningful engineered features (summated angle, knee × trunk interaction, knee-to-trunk ratio, and knee angle), the AdaBoost classifier achieved good overall performance on a held-out test set. SHAP and LIME provided consistent global and local explanations, enabling transparent interpretation of SLS performance classifications. Importantly, the proposed coordination-informed feature design provides a distinct contribution by reframing SLS assessment beyond isolated joint deviations, instead emphasizing intersegment movement patterns that align more closely with clinical reasoning and athletic screening. This workflow uses smartphones only for standardized video acquisition, while explainable SLS performance screening is achieved through an interpretable machine learning framework. Given the lightweight feature set and classical machine learning models used, the proposed framework is suitable for future on-device inference on modern smartphones, which may support rehabilitation planning and injury prevention strategies.

Funding

This work was supported by the Yonsei University Research Grant of 2024 (2024-22-0406).

Data Availability

The datasets generated or analyzed during this study are not publicly available due to privacy concerns but are available from the corresponding author on reasonable request.

Authors' Contributions

SK and KP conceptualized the study and jointly managed the project. Both authors developed the study methodology, curated the data, and contributed to the formal analysis. KP led the formal analysis and supervised the overall project. SK provided methodological support and contributed to data interpretation. Both authors were involved in visualization and drafting the original manuscript, as well as revising the manuscript at multiple stages. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Duong V, Oo WM, Ding C, Culvenor AG, Hunter DJ. Evaluation and treatment of knee pain: a review. JAMA 2023 Oct 24;330(16):1568-1580. [doi: 10.1001/jama.2023.19675] [Medline: 37874571]
2. Drigny J, Rolland M, Gauthier A. The influence of knee proprioception and strength on lower-limb functional symmetry in healthy adults. Muscles 2025 Jan 21;4(1):3. [doi: 10.3390/muscles4010003] [Medline: 40757578]
3. Jildeh TR, Castle JP, Buckley PJ, Abbas MJ, Hegde Y, Okoroha KR. Lower extremity injury after return to sports from concussion: a systematic review. Orthop J Sports Med 2022 Jan;10(1):23259671211068438. [doi: 10.1177/23259671211068438] [Medline: 35111864]

XSL•FO
RenderX

4. Zhang Y, Liu Y, Pan Z, Gao H, Martin RL, Huang X. Comparing the reliability of the single leg squat test using two, three, and four category ordinal rating scales. PeerJ 2025;13:e20218. [doi: 10.7717/peerj.20218] [Medline: 41112774]
5. Ressman J, Grooten WJA, Rasmussen Barr E. Visual assessment of movement quality in the single leg squat test: a review and meta-analysis of inter-rater and intrarater reliability. BMJ Open Sport Exerc Med 2019;5(1):e000541. [doi: 10.1136/bmjsem-2019-000541] [Medline: 31275606]
6. Carvalho C, Viadanna Serrão F, Felipe Martinez A, Da Silva Serrão PRM. Three-dimensional kinematics of the trunk, pelvis, hip, and knee during the single-leg squat and hip torque in subjects with isolated patellofemoral osteoarthritis compared to individually matched controls: preliminary results. Arch Rheumatol 2024 Mar;39(1):33-45. [doi: 10.46497/ArchRheumatol.2024.9814] [Medline: 38774690]
7. Nae J, Banega M, Kronström F, Jakobsson A, Creaby MW, Ageberg E. Evaluation of the ability of machine learning-models to assess postural orientation errors during a single-leg squat. JOSPT Methods 2025 Jan;1(1):17-29. [doi: 10.2519/josptmethods.2024.0086]
8. Chariar M, Rao S, Irani A, Suresh S, Asha CS. AI trainer: autoencoder based approach for squat analysis and correction. IEEE Access 2023;11:107135-107149. [doi: 10.1109/ACCESS.2023.3316009]
9. Kianifar R, Lee A, Raina S, Kulic D. Automated assessment of dynamic knee valgus and risk of knee injury during the single leg squat. IEEE J Transl Eng Health Med 2017;5:2100213. [doi: 10.1109/JTEHM.2017.2736559] [Medline: 29204327]
10. Park J, Kim Y, Kim S, Park K. Markerless kinematic data in the frontal plane contributions to movement quality in the single-leg squat test: a comparison and decision tree approach. J Sport Rehabil 2025 Feb 1;34(2):126-133. [doi: 10.1123/jsr.2024-0182] [Medline: 39500301]
11. Dajime PF, Smith H, Zhang Y. Automated classification of movement quality using the Microsoft Kinect V2 sensor. Comput Biol Med 2020 Oct;125:104021. [doi: 10.1016/j.compbiomed.2020.104021] [Medline: 33011646]
12. Zheng X, Otten E, Reneman MF, Lamoth CJ. Explaining deep learning models for age-related gait classification based on acceleration time series. Comput Biol Med 2025 Jan;184:109338. [doi: 10.1016/j.compbiomed.2024.109338] [Medline: 39536383]
13. Zheng X, Reneman MF, Echeita JA, et al. Association between central sensitization and gait in chronic low back pain: insights from a machine learning approach. Comput Biol Med 2022 May;144:105329. [doi: 10.1016/j.compbiomed.2022.105329] [Medline: 35231802]
14. Tak I, Wiertz WP, Barendrecht M, Langhout R. Validity of a new 3-D motion analysis tool for the assessment of knee, hip and spine joint angles during the single leg squat. Sensors (Basel) 2020 Aug 13;20(16):4539. [doi: 10.3390/s20164539] [Medline: 32823657]
15. Riley RD, Ensor J, Snell KIE, et al. Importance of sample size on the quality and utility of AI-based prediction models for healthcare. Lancet Digit Health 2025 Jun;7(6):100857. [doi: 10.1016/j.landig.2025.01.013] [Medline: 40461350]
16. Guo L, Chang R, Wang J, et al. Artificial intelligence-enhanced 3D gait analysis with a single consumer-grade camera. J Biomech 2025 Jun;187:112738. [doi: 10.1016/j.jbiomech.2025.112738] [Medline: 40378677]
17. Rabello R, Bertozzi F, Brunetti C, et al. The influence of task type and movement speed on lower limb kinematics during single-leg tasks. Gait Posture 2022 Jul;96:109-116. [doi: 10.1016/j.gaitpost.2022.05.020] [Medline: 35635986]
18. Barker-Davies RM, Roberts A, Bennett AN, Fong DTP, Wheeler P, Lewis MP. Single leg squat ratings by clinicians are reliable and predict excessive hip internal rotation moment. Gait Posture 2018 Mar;61:453-458. [doi: 10.1016/j.gaitpost.2018.02.016] [Medline: 29486363]
19. Ressman J, Grooten WJA, Rasmussen-Barr E. Visual assessment of movement quality: a study on intra- and interrater reliability of a multi-segmental single leg squat test. BMC Sports Sci Med Rehabil 2021 Jun 8;13(1):66. [doi: 10.1186/s13102-021-00289-x] [Medline: 34099021]
20. McKeown I, Taylor-McKeown K, Woods C, Ball N. Athletic ability assessment: a movement assessment protocol for athletes. Int J Sports Phys Ther 2014 Dec;9(7):862-873. [Medline: 25540702]
21. Räisänen A, Pasanen K, Krosshaug T, Avela J, Perttunen J, Parkkari J. Single-leg squat as a tool to evaluate young athletes' frontal plane knee control. Clin J Sport Med 2016 Nov;26(6):478-482. [doi: 10.1097/JSM.0000000000000288] [Medline: 26704374]
22. Kwok BC, Kong PW. Single-leg squat postural sway reliability: how many trials to analyze for chronic low back pain? J Mech Med Biol 2025 Feb;25(1):2450026. [doi: 10.1142/S021951942450026X]
23. Lafayette TBDG, Kunst VHDL, Melo PVDS, et al. Validation of angle estimation based on body tracking data from RGB-D and RGB cameras for biomechanical assessment. Sensors (Basel) 2022 Dec 20;23(1):3. [doi: 10.3390/s23010003] [Medline: 36616603]
24. Azad M, Khan MFK, El-Ghany SA. XAI-enhanced machine learning for obesity risk classification: a stacking approach with LIME explanations. IEEE Access 2025;13:13847-13865. [doi: 10.1109/ACCESS.2025.3530840]
25. Yue L, Chen WG, Liu SC, Chen SB, Xiao SF. An explainable machine learning based prediction model for Alzheimer's disease in China longitudinal aging study. Front Aging Neurosci 2023;15:1267020. [doi: 10.3389/fnagi.2023.1267020] [Medline: 38020780]

XSL•FO
RenderX

26. Khalilipour N, Tajik R, Abbasi A, Mantashloo Z, Aziminia M. Squat errors classification based on National Academy of Sports Medicine guidelines using IMU and deep learning algorithms. Comput Biol Med 2025 Oct;197(Pt A):110962. [doi: 10.1016/j.compbiomed.2025.110962] [Medline: 40857817]
27. Markowetz F. All models are wrong and yours are useless: making clinical prediction models impactful for patients. NPJ Precis Oncol 2024 Feb 28;8(1):54. [doi: 10.1038/s41698-024-00553-6] [Medline: 38418530]
28. Plummer HA, Oliver GD, Powers CM, Michener LA. Trunk lean during a single-leg squat is associated with trunk lean during pitching. Int J Sports Phys Ther 2018 Feb;13(1):58-65. [doi: 10.26603/ijspt20180058] [Medline: 29484242]

Abbreviations

AdaBoost: adaptive boosting
AUC: area under curve
IMU: inertial measurement unit
LightGBM: light gradient boosting machine
LIME: local interpretable model-agnostic explanations
SHAP: Shapley additive explanations
SLS: single-leg squat
SVM: support vector machine
XAI: explainable artificial intelligence

Edited by M Giurgiu; submitted 01.Oct.2025; peer-reviewed by J Simmich, ME Haque; accepted 05.Feb.2026; published 12.Mar.2026.

Please cite as:
Kim S, Park K
Smartphone-Based Interpretable Machine Learning for Classifying Single-Leg Squat Performance Using Trunk, Pelvic, and Knee Kinematics: Cross-Sectional Study
JMIR Mhealth Uhealth 2026;14:e85126
URL: https://mhealth.jmir.org/2026/1/e85126
doi:10.2196/85126

© Sihyun Kim, Kyuenam Park. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 12.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobile App Rating Scale for Health Care Professionals to Assess the Quality of mHealth Apps: Questionnaire Development and Psychometric Analysis

Albie Sharpe[1], PhD; Yiong Huak Chan[2], PhD; Amarasinghe Arachchige Don Nalin Samandika Saparamadu[3,4], MBBS, MPH

[1]Discipline of Public Health, School of Human Performance, Rehabilitation and Population Health, University of Technology Sydney, Sydney, NSW, Australia

[2]Biostatistics Unit, Yong Loo Lin School of Medicine, National University of Singapore, Kent Ridge, Singapore, Singapore

[3]Department of International Health, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, 615 N Wolfe St, Baltimore, MD, United States

[4]Department of Nutrition and Food Studies, College of Public Health, George Mason University, Fairfax, VA, United States

Corresponding Author:
Amarasinghe Arachchige Don Nalin Samandika Saparamadu, MBBS, MPH
Department of International Health, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, 615 N Wolfe St, Baltimore, MD, United States

Abstract

Background: Many frameworks and tools are available to evaluate the quality of mobile health apps (MHAs), which are increasingly used by health care professionals (HCPs) for accessing medical information, clinical decision support, and communication. However, existing tools are not well equipped to assess the quality of apps designed for HCPs from their perspectives.

Objective: We aimed to develop a new tool based on the Mobile App Rating Scale (MARS) to capture the unique perspectives of HCPs on MHAs. We then conducted a psychometric analysis of this new questionnaire to determine its effectiveness in assessing the quality of MHAs designed specifically for HCPs from their perspectives.

Methods: This study was conducted in 2 phases. In phase 1, the original MARS tool was adapted for HCPs through expert panel review and subsequent qualitative interviews, resulting in the development of the pMARS (MARS for health care professionals) tool. This phase focused on establishing face and content validity. Qualitative interviews were conducted with HCPs from a tertiary hospital in Singapore to gather their perspectives on the tool's structure, clarity, applicability, and usability. In phase 2, we invited HCP participants to complete pMARS based on their experience with the LabMed app, an mHealth tool designed to provide medical laboratory–related information to HCPs. We established the construct validity of pMARS through multiple psychometric techniques. Internal consistency reliability was measured using the Cronbach α, while structural equation modeling was used to examine the interrelationships among latent constructs. Additionally, we used item response theory (IRT) to evaluate each item's impact on latent constructs of interest, that is, discriminative performance of individual items within each domain.

Results: Based on the results from phase 1, the pMARS comprised 26 items across 5 domains: engagement, functionality, aesthetics, information, and subjective quality, refined through interviews with 10 HCPs. In phase 2 (n=218), pMARS demonstrated good internal consistency reliability across all domains (Cronbach α=0.855-0.931). Structural equation modeling demonstrated that functionality had the strongest influence on end-user willingness to use, recommend, and purchase the MHA (P<.001). IRT identified that customization and interactivity of the engagement domain had a weak impact on latent constructs, whereas entertainment had a higher impact. Ease of use and gestural design had a weak impact on the functionality domain, whereas arrangement and size of content and quantity and quality had a strong impact on the aesthetics and information domains, respectively.

Conclusions: This study reports the development and psychometric analysis of pMARS. Our findings demonstrate strong internal consistency reliability and construct validity, supporting its potential use in health care. Further research should validate pMARS across diverse MHAs and contexts and apply IRT to further refine its precision and efficiency.

(JMIR Mhealth Uhealth 2026;14:e48828) doi:10.2196/48828

XSL•FO
RenderX

KEYWORDS

Mobile App Rating Scale for health care professionals; MARS for health care professionals; pMARS; psychometric analysis; structural equation modeling; item response theory; scale development; internal consistency reliability; eHealth; mobile health; mHealth

Introduction

Mobile health (mHealth) apps (MHAs) offer the potential to deliver accessible, efficient, and cost-effective health care [1,2]. These apps are used not only by patients for health education and self-care but also by health care professionals (HCPs) [3]. HCPs rely on MHAs for many purposes, including consultations between HCPs, intersectoral communication, and access to information for HCPs at the point of health care delivery [3,4]. However, the quality of MHAs designed specifically for HCPs can vary significantly in terms of the accuracy of the information provided, functionality, data visualization effectiveness, credibility, and other key factors. Poor-quality or misleading content can pose serious risks to patient health and safety, underscoring the need for reliable quality assessment measures [1,5-7]. Although several frameworks and tools have been developed to evaluate the quality of MHAs, the Mobile Application Rating Scale (MARS) remains one of the most widely used and recognized tools for this purpose [8-10].

MARS is a multidimensional tool designed for experts to assess the quality of MHAs [10,11]. It consists of 23 questions or items, with 19 of these organized into 4 objective dimensions: engagement (5 items), functionality (4 items), aesthetics (3 items), and information (7 items) [10]. MARS has been widely applied in the evaluation of apps related to health, well-being, health care, and medicine [12-17]. Researchers have validated MARS across multiple languages, demonstrating that its psychometric properties are consistent with those of the original English version [11,17-20].

Given the expertise and training required to administer MARS, a simplified version, known as uMARS (user version of the MARS), has been developed for end users [21]. Items requiring expertise and complex terminology have been removed from uMARS. While uMARS has gained wide acceptance for its utility in assessing the quality of various MHAs [22-25] and has been translated into other languages [26], it has also faced criticism. Specifically, it has been noted that uMARS may inadequately capture end user perspectives on quality, and some have suggested that it should include emerging mHealth technologies to enhance its comprehensiveness [27].

Despite the widespread use of tools for assessing the quality of MHAs, there have been only a few attempts to evaluate these tools from the perspective of HCPs as end users. This evaluation is crucial for several reasons. First, HCPs are reported to have unique requirements when assessing quality, including access to patient data, data security, and evidence-based clinical content [28]. Second, MHAs designed for HCPs must cater to clinical needs and other professional decision-making processes. Unlike MHAs for general users, these apps are also subject to professional and organizational policies that mandate specific security features. Third, HCPs often face demanding schedules, with limited time to thoroughly assess digital tools due to their primary responsibilities in patient care and clinical decision-making [29]. Finally, current assessment tools, such as MARS and uMARS, may not fully meet these specific functional requirements or accommodate the time constraints experienced by HCPs. Consequently, developing a new, specifically designed, and simplified questionnaire that can be administered quickly would effectively address these gaps and provide a more relevant evaluation tool for HCPs as end users.

Moreover, involving HCPs in the development of MHAs is widely recognized as a key factor in creating safe and effective apps [28,30,31]. A tailored assessment tool would enable MHA developers to address the unique priorities of HCPs. In addition, by capturing the distinct perspectives of HCPs using a dedicated tool, service providers can obtain more accurate and practical insights, ensuring that MHAs are suitable for professional use and health care delivery. Therefore, we aimed to (1) develop a new MHA rating scale tailored for HCPs based on the widely validated MARS framework and (2) conduct a psychometric analysis of this new questionnaire to determine its effectiveness in assessing the quality of MHAs designed specifically for HCPs.

This paper describes the development of the new rating scale and evaluates its initial application using the LabMed app, an mHealth tool designed to provide medical laboratory–related information to HCPs [29,32]. We present our Methods, Results, and Discussion sections in accordance with the "recommendations for instrument and scale development and testing" [33,34].

Methods

In this section, we discuss the specific methods and materials used in the study: (1) the process of adapting MARS for use by HCPs with the goal of establishing judgment validity and the qualitative methods used in this process (phase 1) and (2) the assessment of reliability and construct validity of the new tool, pMARS (MARS for health care professionals), through Cronbach α and structural equation modeling (SEM), respectively, followed by the utility of item response theory (IRT) in identifying weakly impacting or performing items (phase 2).

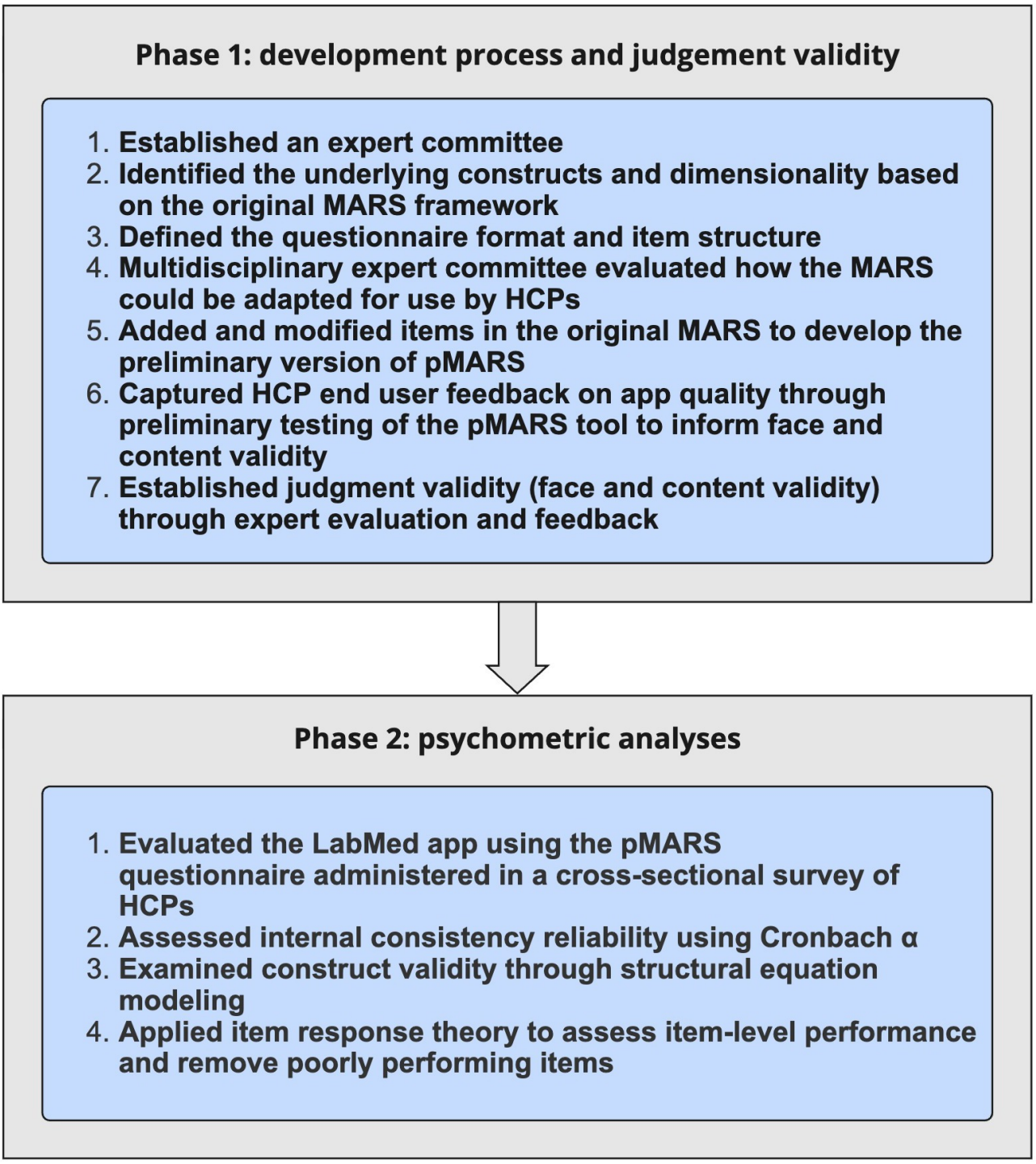
Phase 1: Development of the New Rating Scale (pMARS) and Establishment of Judgment Validity

Overview

Judgment validity was based on face and content validity [35]. Face validity is the degree to which respondents judge the questionnaire items to be valid. Content validity is the extent to which the observed variables in a questionnaire are representative of the theoretical constructs or latent variables (eg, entertainment, interest, customization, and interactivity) and latent constructs (eg, engagement, functionality, and aesthetics). A study outline is shown in Figure 1.

XSL•FO
RenderX

Figure 1. Study outline. HCP: health care professional; MARS: Mobile App Rating Scale. pMARS: Mobile App Rating Scale for health care professionals.



Expert Evaluation

Judgment validity as a whole was approached using the following steps: establishing an expert committee, identifying dimensionality of constructs, determining the questionnaire and item formats, adding items, determining length, reviewing results, and revising [35]. This step was guided by the anticipated clinical, informational, and workflow-related needs of HCPs when evaluating MHAs for professional use.

Data Collection

Data collection focused on systematically eliciting expert judgments on the relevance, suitability, clarity, and completeness of MARS items when assessing MHAs intended for HCP use. We assessed content validity with the following criteria: (1) questions were clear and easy to understand, (2) questions covered all the important aspects of each domain, (3) the questionnaire did not omit important questions regarding quality of MHAs, (4) the questionnaire was suitable for replication studies, and (5) there were no violations of privacy. A 3-member panel of multidisciplinary experts comprising a physician, a public health expert, and a human-computer interactions specialist evaluated the relevance and suitability of the original MARS domains and identified items for adaptation to HCP end users.

XSL•FO
RenderX

Data Analysis

The aim of this analysis was to identify key domains and constructs from HCP perspectives, eliminate multiconstruct items, ensure appropriate language for HCPs, and review questionnaire length. Feedback was synthesized through iterative review and consensus among the experts to determine item retention, modification, and/or removal. A draft tool with 23 items across 5 conceptually defined domains was developed. Items were not reverse scored. Domains were treated as conceptual groupings rather than independent subscales. Scoring of the tool was carried out by averaging the item scores within domains and across the full pMARS tool.

Interviews

To further establish face and content validity from an end-user perspective, qualitative interviews were conducted with HCPs.

Data Collection

To assess pMARS's ability to capture quality and other key characteristics, we recruited a sample of HCPs (n=10) through both convenience and purposive sampling for face-to-face interviews using a semistructured interview guide provided in Multimedia Appendix 1. A convenience sampling method was used to recruit HCPs who had recent contact with the clinical laboratory. A purposive sampling method was used to ensure that both physicians and nurses were represented in the sample. One researcher (AADNSS) was present for a week at the clinical laboratory to recruit participants.

Demographic details, excluding personal identifiers, were collected during the interview. Participants first completed version 1 of the pMARS tool, then were interviewed about their experiences and asked for suggestions for improvement. On average, the entire interview process lasted 25 to 30 minutes and was conducted in a private, quiet space within the hospital.

Interview Analysis

Audio recordings from the interviews were transcribed verbatim and handwritten notes were compiled without participant identifiers. Anonymized transcripts were analyzed using a deductive coding process [36]. One researcher (AS) conducted the initial coding, categorizing participant feedback based on the role pMARS played in assessing the app. Preliminary codes and emerging themes were discussed by the researcher and a collaborator until consensus was reached. The final coding scheme was subsequently reviewed and validated by a second researcher (AADNSS). Findings from this phase were used to refine item wording and enhance the content validity of the pMARS tool.

Phase 2: Evaluation of Reliability and Construct Validity of pMARS

Overview

To evaluate the pMARS questionnaire, a cross-sectional analysis was conducted with the users of the LabMed app between March and August 2019 at a tertiary care general hospital in Singapore.

Data Collection

The physicians, residents, nurse clinicians, and registered nurses in the hospital who used the clinical laboratory services and had experience using the LabMed app were eligible to take part in the study.

We used a simple sequential sampling approach to recruit participants. The questionnaire was administered using 2 parallel approaches. First, a researcher (AADNSS) systematically visited all hospital wards in person and distributed printed questionnaires, which were collected periodically during the data collection period. Second, the questionnaire was made available through the hospital intranet via the REDCap (Research Electronic Data Capture [37]) electronic data capture platform. The HCPs were informed of the online option through a single mass email campaign conducted by the hospital's marketing department. Response was voluntary.

REDCap is a secure, web-based software platform designed to support data capture for research studies. We aimed for a participant-to-questionnaire-item ratio of 8:1 to 10:1, resulting in an estimated sample size of 184 to 230 [35].

Data Analysis

All forms of questionnaires were transcribed into a Microsoft Excel file by 2 researchers working together to minimize transcription errors. STATA (version 17.0; StataCorp) was used for analysis [38]. The Cronbach α was used to show the internal consistency reliability of pMARS: 0.80 to 0.89 was considered good, and >0.90 was considered excellent [39]. A high α may indicate redundancy; however, it is not the sole determinant of item appropriateness [40]. Face and content validity, along with factor structure, were essential in determining the questionnaire's ability to accurately measure the intended construct.

SEM was used to evaluate the interrelationships between the latent constructs after adjusting for age, sex, and HCPs' role in the hospital [1]. SEM assumed a latent construct for the observed items of each subscale, and a higher order factor—subjective quality—accounted for correlations between the factors (Figure 2).

XSL•FO
RenderX

Figure 2. The model used in structural equation modeling.



Lastly, we used IRT to identify each question's impact on the latent constructs, that is, engagement, functionality, information, aesthetics, and subjective quality. Unlike classical test theory, which assumes all items contribute equally to the overall score, IRT allows for identifying items with varying discrimination, making it ideal for optimizing and shortening measurement tools without compromising validity and reliability. The primary purpose of using IRT was to refine the questionnaire, resulting in a precise, valid, and relatively brief pMARS that minimized the completion time and response burden [41].

Ethical Considerations

This research project received an ethics review exemption from the Projects Committee of the Department of Laboratory Medicine at Ng Teng Fong General Hospital, Singapore. The exemption was granted in accordance with the department's policy for minimal-risk research involving health care staff and was reviewed and approved by the head of the department. All participants were provided with an information sheet describing the study objectives, the voluntary nature of participation, and the commitment to data confidentiality. Informed consent was obtained before participation in each study phase. Participants who completed the qualitative interviews (phase 1) or paper-based questionnaires (phase 2) provided written consent, while those who completed the online survey via REDCap on the hospital intranet platform provided electronic consent.

Basic demographic data, including age, biological sex, and professional role, were collected. No personally identifiable data were collected. All files were stored on password-protected institutional drives accessible only to study investigators. In phase 2, participants were offered a Singapore $10 (S$1=US $0.78 as of June 17, 2026) voucher for a local coffee chain as compensation for their time spent completing the questionnaire. No other payments or incentives were provided.

Results

Phase 1, the qualitative study, included 10 participants. Phase 2 included 218 participants who responded and returned completed questionnaires. Selected sociodemographic characteristics of the participants are presented in Table 1.

XSL•FO
RenderX

Table . Demographic characteristics of the participants in phase 1 and phase 2.

Characteristics	Values
Phase 1	
Age (years), mean (SD)	28.4 (8.81)
Sex, n (%)	
Male	3 (30)
Female	7 (70)
Role in hospital, n (%)	
Nurse	8 (80)
Doctor	2 (20)
Phase 2	
Age (years), mean (SD)	30.4 (7.2)
Sex, n (%)	
Male	39 (17.9)
Female	176 (80.7)
Missing data	3 (0)
Role in hospital, n (%)	
Nurse	154 (70.6)
Doctor	47 (21.6)
Missing data	17 (7.8)
Specialty, n (%)	
Surgical specialties	26 (11.9)
Medical specialties (inpatient or acute care)	91 (41.7)
Community hospital	9 (4.1)
Unspecified	92 (42.2)

Phase 1: Development of pMARS and Assessment of Its Judgment Validity

The first version of the questionnaire consisted of a demographic data section and 5 domains. Certain question stems were reworded and/or split into 2 separate stems to ensure that there were no double-barreled questions, for example, item 1 of section A of MARS reads, “Is the app fun/entertaining to use? Does it use any strategies to increase engagement through entertainment?” This was broken into two separate question stems: (1) “Is the app fun/entertaining to use?” and (2) “Does the app use strategies to increase engagement through entertainment (e.g., strategies such as interactivity/gamification)?” This version of the questionnaire consisted of question stems presented in question form and shortened but descriptive answer stems presented as a Likert scale. Moreover, almost all the question stems had examples or descriptive statements to further elaborate or clarify the ideas.

Insights from the qualitative assessment, categorized into 4 key areas, guided the multidisciplinary panel in developing the second version of pMARS at the end of phase 1 (Table 2).

XSL•FO
RenderX

Table . Quotes from the qualitative interview participants (n=10) in relation to specific functionalities of the Mobile App Rating Scale for health care professionals.

Function	Quotes from participants	Action taken
Structure	• "Some questions are truncated in the paper." • "Takes a long time to read." • "Print all the stems on the same page." • "Shorten the questionnaire by reducing the length of the items."	• Shorten the questionnaire by reducing the length of the items. • Limit examples and descriptive statements in question stems. • Switch to landscape format to reduce the length of the survey to three pages.
Clarity	• "RCT, what does it mean? Only the abbreviation is given." • "Should have presented the answer stems in tables or using numbers instead of bullet points." • "I think only the options are quite long." • "Not sure or uncertain is the answer that I was looking for."	• Simplify the language and minimize jargon. • Revise question stems to statements to reduce length and increase clarity.
Applicability	• "The questionnaire is different to the context of this app." • "We usually use it when we need it, but the question is about entertainment. Not relevant." • "It is not about entertainment, it's whether I want it or not." • "Credibility is granted by hospital. Should not doubt it." • "App store description is not usually read by anyone. Therefore it is probably not important to assess."	• Reword item stems to better highlight app functionality and relevance. • Reduce descriptive statements to minimize assumptions about app use. • Remove irrelevant items such as app store description. • Reduce items on entertainment and gamification.
Usability	• "Likert scale should be reversed, as in 5 should be extremely satisfied." • "Should have presented the answer stems in tables or using numbers instead of bullet points." • "Introduce an additional option for those who do not know the answer."	• Change the answer stems to a simple 5-point Likert scale. • Add a new answer stem for "don't know." • Simplify design and remove Likert descriptors.

Findings from the qualitative assessment of the questionnaire (ie, interviews with HCPs) shed light on several areas of the questionnaire for focused improvement: (1) shorten the questionnaire by reducing the length of the answers, (2) further simplify the language and minimize jargon, and (3) introduce an additional option for those who do not know the answer to a specific question. It is noteworthy that no new items were proposed by the HCPs. Based on the findings, we changed the question stems from question forms to statements, for example, (1) "arrangement and size of buttons, icons, and menus on the screen is appropriate" and (2) "it uses strategies to increase engagement through entertainment (e.g., strategies such as interactivity/gamification)." Moreover, we limited examples and/or descriptive statements in question stems to a minimum. Simultaneously, we changed the answer stems to a simple 5-point Likert scale without embedded descriptive statements: strongly agree, agree, neutral, disagree, and strongly disagree. Finally, we added a new answer stem: "don't know."

In addition, we redesigned the questionnaire by changing its orientation from portrait to landscape. The reasons for this change were to limit the length of the questionnaire to 3 pages despite including additional questions and to provide an easy-to-complete format. The final version of the questionnaire at the end of phase 1 consisted of 26 questions spread across 5 domains: engagement (6 items), functionality (5 items), aesthetics (4 items), information (7 items), and subjective quality (4 items). The length of the questionnaire was deemed appropriate by the committee (Multimedia Appendix 2).

Phase 2: Reliability and Construct Validity of pMARS

We distributed 197 printed questionnaires among HCPs in person, and 189 participated, resulting in a 95.9% response rate. Additionally, 29 HCPs accessed the survey through the hospital intranet, bringing the total number of responses to 218. A further 15 REDCap surveys were incomplete, resulting in a REDCap complete response rate of 65.9%. We were unable to calculate an overall response rate for REDCap as we could not track the reach through the hospital intranet. Among the in-person and intranet-administered questionnaires returned, the proportion of nonresponses per question item ranged from 0% to 13.3%, with the item "Would you pay for this app?" showing the highest nonresponse rate (n=29). Most items (29/30; 96.7%) had less than 5.1% missing responses. Detailed item-level missing data, including counts of "don't know" responses, are provided in Multimedia Appendix 3.

XSL•FO
RenderX

The Cronbach α (ranging from 0.855 to 0.931) showed that the internal consistency reliability of pMARS was good to excellent for all the domains (Table 3). SEM analysis demonstrated that functionality (P<.001) had the strongest influence on end-user willingness to use or recommend and purchase the MHA, after adjusting for age, sex, and the HCPs' role in the hospital (Figure 3).

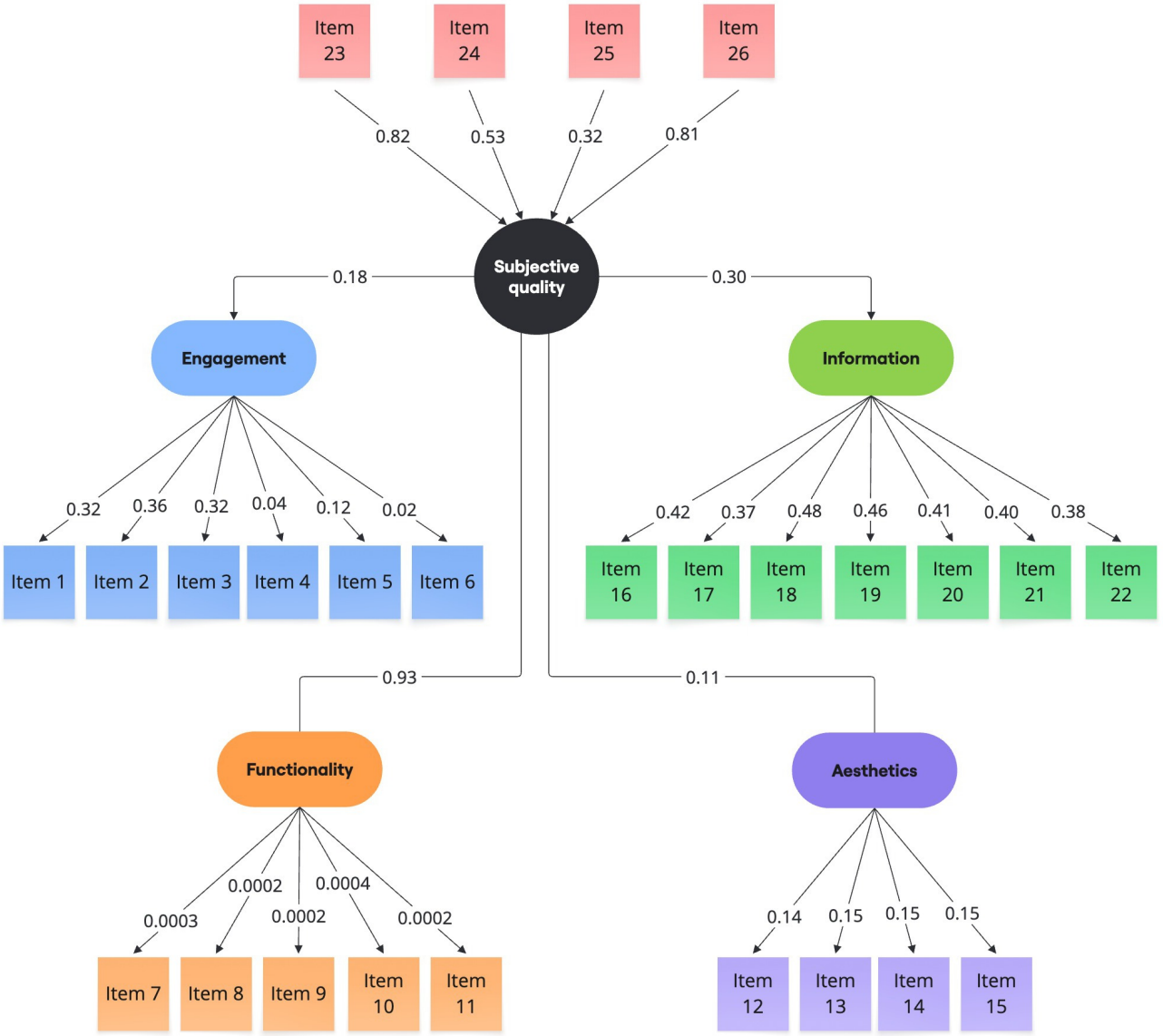
For domains with Cronbach α values >0.90, we conducted a manual item review to assess potential redundancy. We examined item wording and conceptual distinctness within each domain. The review confirmed that, although some items were related, they measured complementary aspects of the construct and contributed meaningfully to the overall scale. All such items were therefore retained in the current version of pMARS.

Table . Cronbach α results for the Mobile App Rating Scale for health care professionals.

Domain and item	Scale mean if item deleted	Scale variance if item deleted	Corrected item–total correlation	Cronbach α if item deleted
Engagement: Cronbach α=0.855, number of items=6				
Item 1	17.936	8.080	0.671	0.825
Item 2	18.005	8.163	0.554	0.849
Item 3	17.675	7.597	0.721	0.815
Item 4	17.700	8.152	0.698	0.821
Item 5	17.778	8.292	0.597	0.839
Item 6	17.409	8.382	0.626	0.833
Functionality: Cronbach α=0.931, number of items=5				
Item 7	15.212	6.296	0.785	0.922
Item 8	15.212	6.059	0.864	0.906
Item 9	15.232	6.535	0.784	0.921
Item 10	15.227	6.493	0.847	0.910
Item 11	15.217	6.636	0.816	0.916
aesthetics: Cronbach α=0.928, number of items=4				
Item 12	11.296	3.576	0.813	0.912
Item 13	11.379	3.415	0.835	0.905
Item 14	11.271	3.406	0.892	0.886
Item 15	11.232	3.694	0.789	0.920
Information: Cronbach α=0.905, number of items=7				
Item 16	22.688	12.056	0.740	0.889
Item 17	22.752	12.048	0.663	0.897
Item 18	22.673	11.594	0.825	0.879
Item 19	22.713	11.569	0.811	0.881
Item 20	22.757	11.568	0.720	0.891
Item 21	22.748	11.672	0.783	0.884
Item 22	22.965	12.302	0.523	0.915

XSL•FO
RenderX

Figure 3. Results of the structural equation modeling.



IRT (Table 4) identified that for the engagement domain, items 4 (customization) and 5 (interactivity) had a weak impact on latent constructs, whereas items 1 and 2 (both entertainment) had a higher impact on the domain. Similarly, items 9 (ease of use) and 11 (gestural design) had a weak impact on functionality, item 13 (arrangement and size of content) had a strong impact on aesthetics, and items 18 (quantity) and 19 (quality) had a strong impact on information. The engagement domain demonstrated the lowest reliability, and IRT highlighted weakly impacting items within the domain, enabling the researchers to shorten the questionnaire if required.

XSL•FO
RenderX

Table . Results of the item response theory analysis.

Domain and item	Impact[a]
Engagement	
Item 1	12.8
Item 2	5.4
Item 3	5.5
Item 4	2.9
Item 5	2.4
Item 6	5.3
Functionality	
Item 7	11.0
Item 8	41.6
Item 9	4.1
Item 10	10.8
Item 11	4.9
Aesthetics	
Item 12	12.3
Item 13	63.0
Item 14	16.6
Item 15	10.7
Information	
Item 16	3.3
Item 17	3.0
Item 18	6.8
Item 19	8.6
Item 20	3.2
Item 21	4.1
Item 22	4.2

[a]Overall: engagement=4.0, functionality=11.3, aesthetics=24.5, and information=4.0.

Discussion

Principal Findings

Our study shows that the newly developed pMARS tool has strong internal consistency reliability (Cronbach α) across all domains, indicating its potential utility in assessing the quality of MHAs designed for HCPs from their perspective. Although most items performed well, item 22 of the information domain ("the app has been trialled/tested; verified by evidence") demonstrated weak correlations (Table 3). The item was retained in the final version of pMARS as it reflects a critical aspect of evidence-based practice in clinical settings and warrants further examination in future validation studies. Furthermore, SEM generated valuable insights into interrelationships between the latent constructs. Not all items were equally good indicators for the dimensions. The strongest influence was observed between subjective quality and functionality. Similarly, items 1 to 3 of the engagement domain, items 16 to 22 of the information domain, and items 23 and 26 of the subjective quality domain demonstrated relatively strong influence.

On the other hand, IRT revealed weakly impacting items such as items 4 and 5 of the engagement domain; items 9 and 11 of the functionality domain; and items 16, 17, and 20 of the information domain. These items provide an opportunity to further shorten the questionnaire. However, it should be evaluated within the theoretical factor model of the scale, and therefore, this is an area for future research. Overall, our findings suggest that pMARS is an MHA quality assessment tool of good metric quality considering the reliability and construct validity results.

Clinical Utility and Developmental Limitations

In the busy health care settings of Singapore, engaging HCPs in discussions unrelated to direct patient care can be challenging. With its brevity and standardized format, pMARS enables rapid assessment of MHA quality from the perspective of HCPs as clinical end users. Unlike MARS, which was designed for expert

XSL•FO
RenderX

evaluators to classify and assess general MHA quality, and uMARS, which targets general end users, pMARS fills a critical gap by capturing HCP-specific clinical usability and professional decision-making considerations that are not specifically addressed by existing tools [10,21]. The use of simplified language in pMARS ensures accessibility for HCPs without requiring specific training, including those who may be unfamiliar with MHA-related terminology. Consequently, pMARS provides a practical, contextually relevant, and adaptable tool for service providers and HCPs to evaluate the quality of MHAs within real-world clinical settings. Notably, pMARS is designed as a flexible framework that can be iteratively refined to incorporate emerging clinical-, workflow-, and patient safety–related quality considerations as digital health technologies evolve.

The direct use of MARS in developing pMARS, as described by Baptista et al [27], may eliminate certain aspects of MHAs that are important to end users. To overcome this limitation, we introduced a qualitative approach using interviews with HCPs; however, the inputs generated by this procedure were limited to a few aspects of the questionnaire, as discussed in the Results section. The qualitative component introduced changes to question stems by rewording, shortening, and simplifying them without adding or removing any items from the questionnaire. According to our observations during the interviews, the interview answers were brief and the discussions were limited, possibly as a result of time pressures.

Regulatory Challenges

Legislative frameworks such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the Personal Data Protection Act (PDPA) in Singapore address crucial concerns for HCPs regarding MHAs, including data privacy, security, and patient confidentiality [42,43]. These regulations set important standards for protecting patient information and ensuring the reliability of MHAs. However, the adoption of such regulations varies globally, with many countries still developing or implementing their standards. As regulatory frameworks continue to evolve, the availability of a robust tool to evaluate the quality of MHAs from the perspective of HCPs remains essential to ensure these apps meet the necessary standards. Furthermore, health care organizations can use pMARS to evaluate and select suitable MHAs that comply with professional standards.

Considerations for the Future

As mHealth technology continues to evolve, several areas for potential improvement in future versions of the pMARS tool have emerged. First, interoperability between MHAs, health systems, and electronic health records is crucial for HCPs and health care organizations when integrating these technologies [44]. Second, evaluating the effectiveness and security of communications among HCPs, including end-to-end encryption, is important for ensuring coordinated care and protecting sensitive information [45]. Third, understanding potential patient safety concerns associated with MHAs is vital for maintaining high standards of care. Lastly, language barriers among HCPs should be addressed to enhance the app's quality, ethical considerations, and utility. Future versions of pMARS should take these evolving quality requirements into account. Additionally, with the growing integration of AI and large language models into mHealth apps, future versions of pMARS should also account for quality requirements specific to AI and AI-generated content [46,47].

Limitations

Our study has several methodological and participant-related limitations. There were 2 methodological limitations. First, the psychometric evaluation of pMARS was based on data from only 1 MHA. This focus may affect certain aspects of statistical analysis, particularly if the MHA had quality deficiencies in specific domains. To address this limitation, future validation studies should involve a broader range of MHAs across different medical fields. Second, the IRT depends on the accuracy of the underlying model. The utility of the IRT is contingent upon how well the model reflects the data, making further validation studies essential to evaluate the appropriateness of pMARS items.

Participant-related limitations include the potential influence of the questionnaire's ease of administration on construct validity. The simplicity and repetitive nature of the pMARS may lead to patterned responses, highlighting the need for negatively worded questions in future studies to counteract response biases. Second, the expert panel in phase 1, comprising a physician, public health expert, and human-computer interactions specialist, did not fully represent the demographics of the study participants or potential users, potentially limiting the generalizability of the pMARS tool. Further studies are necessary to assess and validate the pMARS tool across multiple MHAs. Finally, the disproportionate number of female participants should be noted as a limitation, as it may reduce the generalizability of findings to male HCPs. However, to some degree, this distribution reflects the gender composition of the hospital workforce rather than sampling bias.

Conclusions

To our knowledge, this study is the first to develop a tool to evaluate MHAs designed for HCPs from their perspective. Additionally, it is the first study to apply IRT to assess and identify the impact of each question on the latent constructs within MARS and its derivatives. Our findings demonstrate that pMARS is a robust and metrically sound tool for evaluating MHAs intended for HCPs. It has the potential to enable HCPs to assess and recommend high-quality MHAs. We recommend that future research on pMARS incorporate IRT to refine the questionnaire by identifying and removing less-effective items, thereby enhancing its utility and precision in assessing MHA quality.

XSL•FO
RenderX

Acknowledgments

The authors are grateful to all the staff who contributed to the project by taking part in the interviews and completing questionnaires despite their busy schedules in the hospital. This study would not have been possible without their tremendous support. We also thank Dr Leslie Lam, Ms Zeng Peizi, and Mr Andrew Goh for facilitating this study within the Department of Laboratory Medicine of Ng Teng Fong General Hospital.

Funding

This study was supported by the JurongHealth Internal Research & Development Grant Awards for fiscal years 2016 and 2017 (project code 16-55).

Authors' Contributions

AADNSS conceptualized and designed the study, co-designed the quantitative and qualitative methodologies, acquired the data, contributed to quantitative data analysis and interpretation of findings, drafted and critically revised the manuscript, and approved the final version for publication. YHC co-designed the quantitative methodology, performed the statistical analysis, critically reviewed the manuscript, and approved the final version for publication. AS co-designed the qualitative methodology, conducted the qualitative data analysis, critically reviewed the manuscript, and approved the final version for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Semistructured interview guide used in phase 1.
[PDF File, 141 KB - mhealth_v14i1e48828_app1.pdf]

Multimedia Appendix 2
Final version of the Mobile App Rating Scale for health care professionals questionnaire.
[PDF File, 168 KB - mhealth_v14i1e48828_app2.pdf]

Multimedia Appendix 3
Item-level missing data and “don’t know” response summary for the Mobile App Rating Scale for health care professionals questionnaire (phase 2).
[PDF File, 146 KB - mhealth_v14i1e48828_app3.pdf]

References

1. Terhorst Y, Philippi P, Sander LB, et al. Validation of the Mobile Application Rating Scale (MARS). PLoS One 2020;15(11):e0241480. [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
2. Donker T, Petrie K, Proudfoot J, Clarke J, Birch MR, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247. [doi: 10.2196/jmir.2791] [Medline: 24240579]
3. Wattanapisit A, Teo CH, Wattanapisit S, Teoh E, Woo WJ, Ng CJ. Can mobile health apps replace GPs? A scoping review of comparisons between mobile apps and GP tasks. BMC Med Inform Decis Mak 2020 Jan 6;20(1):5. [doi: 10.1186/s12911-019-1016-4] [Medline: 31906985]
4. mHealth: New Horizons for Health Through Mobile Technologies. : World Health Organization; 2011 URL: https://www.afro.who.int/publications/mhealth-new-horizons-health-through-mobile-technologie [accessed 2026-06-18]
5. Lewis TL, Wyatt JC. mHealth and mobile medical apps: a framework to assess risk and promote safer use. J Med Internet Res 2014 Sep 15;16(9):e210. [doi: 10.2196/jmir.3133] [Medline: 25223398]
6. Schoenfeld AJ, Sehgal NJ, Auerbach A. The challenges of mobile health regulation. JAMA Intern Med 2016 May 1;176(5):704-705. [doi: 10.1001/jamainternmed.2016.0326] [Medline: 26999551]
7. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020 Feb 1;27(2):330-340. [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
8. Boudreaux ED, Waring ME, Hayes RB, Sadasivam RS, Mullen S, Pagoto S. Evaluating and selecting mobile health apps: strategies for healthcare providers and healthcare organizations. Transl Behav Med 2014 Dec;4(4):363-371. [doi: 10.1007/s13142-014-0293-9] [Medline: 25584085]
9. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based eHealth interventions. J Med Internet Res 2017 Mar 21;19(3):e82. [doi: 10.2196/jmir.7270] [Medline: 28325712]

XSL•FO
RenderX

10. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
11. Bardus M, Awada N, Ghandour LA, et al. The Arabic version of the Mobile App Rating Scale: development and validation study. JMIR Mhealth Uhealth 2020 Mar 3;8(3):e16956. [doi: 10.2196/16956] [Medline: 32130183]
12. Knitza J, Tascilar K, Messner EM, et al. German mobile apps in rheumatology: review and analysis using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2019 Aug 5;7(8):e14991. [doi: 10.2196/14991] [Medline: 31381501]
13. Jannati N, Salehinejad S, Kuenzig ME, Peña-Sánchez JN. Review and content analysis of mobile apps for inflammatory bowel disease management using the Mobile Application Rating Scale (MARS): systematic search in app stores. Int J Med Inform 2023 Dec;180:105249. [doi: 10.1016/j.ijmedinf.2023.105249] [Medline: 37857167]
14. Masterson Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74. [doi: 10.2196/mhealth.5882] [Medline: 27302310]
15. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016 Mar 10;13:35. [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
16. Mandracchia F, Llauradó E, Tarro L, Valls RM, Solà R. Mobile phone apps for food allergies or intolerances in app stores: systematic search and quality assessment using the Mobile App Rating Scale (MARS). JMIR Mhealth Uhealth 2020 Sep 16;8(9):e18339. [doi: 10.2196/18339] [Medline: 32936078]
17. Mendi O, Kiymac Sari M, Stoyanov S, Mendi B. Development and validation of the Turkish version of the Mobile App Rating Scale - MARS-TR. Int J Med Inform 2022 Oct;166:104843. [doi: 10.1016/j.ijmedinf.2022.104843] [Medline: 35964383]
18. Messner EM, Terhorst Y, Barke A, et al. The German version of the Mobile App Rating Scale (MARS-G): development and validation study. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e14479. [doi: 10.2196/14479] [Medline: 32217504]
19. Domnich A, Arata L, Amicizia D, et al. Development and validation of the Italian version of the Mobile Application Rating Scale and its generalisability to apps targeting primary prevention. BMC Med Inform Decis Mak 2016 Jul 7;16:83. [doi: 10.1186/s12911-016-0323-2] [Medline: 27387434]
20. Martin Payo R, Fernandez Álvarez MM, Blanco Díaz M, Cuesta Izquierdo M, Stoyanov SR, Llaneza Suárez E. Spanish adaptation and validation of the Mobile Application Rating Scale questionnaire. Int J Med Inform 2019 Sep;129:95-99. [doi: 10.1016/j.ijmedinf.2019.06.005] [Medline: 31445295]
21. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849] [Medline: 27287964]
22. LeBeau K, Huey LG, Hart M. Assessing the quality of mobile apps used by occupational therapists: evaluation using the user version of the Mobile Application Rating Scale. JMIR Mhealth Uhealth 2019 May 1;7(5):e13019. [doi: 10.2196/13019] [Medline: 31066712]
23. Adam A, Hellig JC, Perera M, Bolton D, Lawrentschuk N. "Prostate cancer risk calculator" mobile applications (apps): a systematic review and scoring using the validated user version of the Mobile Application Rating Scale (uMARS). World J Urol 2018 Apr;36(4):565-573. [doi: 10.1007/s00345-017-2150-1] [Medline: 29222595]
24. Bardus M, Ali A, Demachkieh F, Hamadeh G. Assessing the quality of mobile phone apps for weight management: user-centered study with employees from a Lebanese university. JMIR Mhealth Uhealth 2019 Jan 23;7(1):e9836. [doi: 10.2196/mhealth.9836] [Medline: 30672742]
25. Ko S, Woo H. Users' needs for mental health apps: quality evaluation using the User Version of the Mobile Application Rating Scale. JMIR Mhealth Uhealth 2025 Jul 4;13:e64622. [doi: 10.2196/64622] [Medline: 40614100]
26. Morselli S, Sebastianelli A, Domnich A, et al. Translation and validation of the Italian version of the user version of the Mobile Application Rating Scale (uMARS). J Prev Med Hyg 2021 Mar;62(1):E243-E248. [doi: 10.15167/2421-4248/jpmh2021.62.1.1894] [Medline: 34322643]
27. Baptista S, Oldenburg B, O'Neil A. Response to "development and validation of the user version of the Mobile Application Rating Scale (uMARS)". JMIR Mhealth Uhealth 2017 Jun 9;5(6):e16. [doi: 10.2196/mhealth.6419] [Medline: 28600277]
28. Backes C, Moyano C, Rimaud C, Bienvenu C, Schneider MP. Digital medication adherence support: could healthcare providers recommend mobile health apps? Front Med Technol 2020;2:616242. [doi: 10.3389/fmedt.2020.616242] [Medline: 35047896]
29. Saparamadu AA, Fernando P, Zeng P, et al. User-centered design process of an mHealth app for health professionals: case study. JMIR Mhealth Uhealth 2021 Mar 26;9(3):e18079. [doi: 10.2196/18079] [Medline: 33769297]
30. Leigh S, Ashall-Payne L. The role of health-care providers in mHealth adoption. Lancet Digit Health 2019 Jun;1(2):e58-e59. [doi: 10.1016/S2589-7500(19)30025-1] [Medline: 33323231]
31. Jacob C, Sanchez-Vazquez A, Ivory C. Clinicians' role in the adoption of an oncology decision support app in Europe and its implications for organizational practices: qualitative case study. JMIR Mhealth Uhealth 2019 May 3;7(5):e13555. [doi: 10.2196/13555] [Medline: 31066710]

XSL•FO
RenderX

32. NTFGH LabMed. Medical Laboratory Mobile App (102) [Mobile application software]. 2019. URL: https://apps.apple.com/sg/app/ntfgh-labmed/id1440006973 [accessed 2026-06-18]
33. Streiner DL, Kottner J. Recommendations for reporting the results of studies of instrument and scale development and testing. J Adv Nurs 2014 Sep;70(9):1970-1979. [doi: 10.1111/jan.12402] [Medline: 24684713]
34. EQUATOR Network. URL: https://www.equator-network.org/ [accessed 2025-01-31]
35. Tsang S, Royse CF, Terkawi AS. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. Saudi J Anaesth 2017 May;11(Suppl 1):S80-S89. [doi: 10.4103/sja.SJA_203_17] [Medline: 28616007]
36. Saldaña J. The Coding Manual for Qualitative Researchers, 4th edition: SAGE Publications; 2021.
37. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
38. Stata index: release 17. StataCorp. URL: https://www.stata.com/manuals17/i.pdf [accessed 2026-07-27]
39. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. Res Sci Educ 2018 Dec;48(6):1273-1296. [doi: 10.1007/s11165-016-9602-2]
40. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011 Jun 27;2:53-55. [doi: 10.5116/ijme.4dfb.8dfd] [Medline: 28029643]
41. Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. Qual Life Res 2007;16 Suppl 1:5-18. [doi: 10.1007/s11136-007-9198-0] [Medline: 17375372]
42. Personal Data Protection Act 2012. Singapore Statutes Online. URL: https://sso.agc.gov.sg/Act/PDPA2012#:~:text=The%20purpose%20of%20this%20Act,consider%20appropriate%20in%20the%20circumstances [accessed 2026-06-18]
43. Health Insurance Portability and Accountability Act of 1996. US Government Publishing Office (GPO). 1996. URL: https://www.govinfo.gov/content/pkg/PLAW-104publ191/pdf/PLAW-104publ191.pdf [accessed 2026-06-18]
44. Ndlovu K, Mars M, Scott RE. Interoperability frameworks linking mHealth applications to electronic record systems. BMC Health Serv Res 2021 May 13;21(1):459. [doi: 10.1186/s12913-021-06473-6] [Medline: 33985495]
45. Galetsi P, Katsaliaki K, Kumar S. Exploring benefits and ethical challenges in the rise of mHealth (mobile healthcare) technology for the common good: an analysis of mobile applications for health specialists. Technovation 2023 Mar;121:102598. [doi: 10.1016/j.technovation.2022.102598]
46. Herpertz J, Dwyer B, Taylor J, Opel N, Torous J. Developing a standardized framework for evaluating health apps using natural language processing. Sci Rep 2025 Apr 6;15(1):11775. [doi: 10.1038/s41598-025-96369-w] [Medline: 40189623]
47. Hua Y, Xia W, Bates D, et al. Standardizing and scaffolding health care AI-chatbot evaluation: systematic review. JMIR AI 2025 Nov 7;4:e69006. [doi: 10.2196/69006] [Medline: 41202290]

Abbreviations

HCP: health care professional
HIPAA: Health Insurance Portability and Accountability Act
IRT: item response theory
MARS: Mobile App Rating Scale
MHA: mobile health app
mHealth: mobile health
PDPA: Personal Data Protection Act
pMARS: Mobile App Rating Scale for health care professionals
SEM: structural equation modeling
uMARS: user version of the Mobile App Rating Scale

Edited by L Buis; submitted 05.Sep.2024; peer-reviewed by F Watcharaporn, N Fernando; revised version received 09.Feb.2026; accepted 10.Feb.2026; published 31.Jul.2026.

Please cite as:
Sharpe A, Chan YH, Saparamadu AADNS
Mobile App Rating Scale for Health Care Professionals to Assess the Quality of mHealth Apps: Questionnaire Development and Psychometric Analysis
JMIR Mhealth Uhealth 2026;14:e48828
URL: https://mhealth.jmir.org/2026/1/e48828
doi:10.2196/48828

XSL•FO
RenderX

© Albie Sharpe, Yiong Huak Chan, Amarasinghe Arachchige Don Nalin Samandika Saparamadu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 31.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Landscape of Mobile Apps for Healthy Eating: Case Study for a Systematic Review and Quality Assessment

Garlene Zamora Zamorano[1*], RDN; Alejandro Déniz-García[1,2*], MD; Alezandra Torres-Castaño[3], PhD; María Luisa Álvarez-Malé[1], PhD; Inger Torhild Gram[4,5], MD, PhD; Guri Skeie[5,6], PhD; Ana M Wägner[1,2], MD, PhD

[1]Instituto de Investigaciones Biomédicas y Sanitarias, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain

[2]Department of Endocrinology and Nutrition, Complejo Hospitalario Universitario Insular Materno-Infantil de Canarias, Instituto de Investigaciones Sanitarias de Canarias, Las Palmas de Gran Canaria, Spain

[3]Evaluation Unit of the Canary Islands Health Service, Instituto de Investigación Sanitaria de Canarias, Tenerife, Spain

[4]Norwegian Centre for eHealth Research, University Hospital of North Norway, Tromsø, Norway

[5]Department of Community Medicine, Faculty of Health Sciences, UiT the Arctic University of Norway, Tromsø, Norway

[6]Nutrition and Metabolism Branch, International Agency for Research on Cancer, Lyon, France

[*]these authors contributed equally

Corresponding Author:
Ana M Wägner, MD, PhD
Instituto de Investigaciones Biomédicas y Sanitarias
Universidad de Las Palmas de Gran Canaria
P Blas Cabrera Felipe “Físico” (s/n)
Las Palmas de Gran Canaria, 35016
Spain
Phone: 34 928453431
Email: ana.wagner@ulpgc.es

Abstract

Background: Mobile apps are being increasingly used to foster healthy lifestyles. There is a growing need for clear, standardized guidelines to help users select safe and effective health apps.

Objective: Our study aimed to highlight the importance of establishing a structured framework for quality evaluation in mobile health (mHealth) through a case study of mobile apps promoting healthy eating.

Methods: We conducted a systematic review of apps promoting healthy eating that had already been evaluated by one or more of 28 recognized health app certification bodies. Three rounds of app evaluations were conducted by experts in nutrition and behavior change. The first two rounds focused on the quality of the content of the recommendations and were performed pairwise using the Quality Evaluation Scoring Tool (QUEST), which has not been previously used by the certification bodies. In addition, in the second and third rounds, each reviewer answered the question “How probable is it that you would recommend this app?” using a subjective scale score from 0 to 10. In the third round, this score was weighed by usability (30%), content quality (40%), and promotion of behavior change (30%). Discussions were held to resolve scoring discrepancies and to identify the top-quality apps. We also assessed correlations among QUEST, Google Play Store, and certification body scores.

Results: Of the 41 apps identified by five certification bodies, 19 (46.3%) met the inclusion criteria and were examined. Only 16 (84.2%) of these remained accessible for the second round. Eight of these surpassed 20 points (out of a maximum of 28) on the QUEST scale and were evaluated by all six experts in the third round, and the top 5 (62.5%) apps were selected. No correlations were found among QUEST, Google Play Store, and certification body scores.

Conclusions: Despite numerous evaluations by various certification bodies, only 5 (12.2%) of the 41 apps met the quality standards set by our experts. Our results mark the importance of rigorous, transparent, and standardized app evaluation processes to guide users toward making informed decisions about health apps. Guidelines for developers for the design of evidence-based, unbiased, high-quality apps, as well as technological solutions for real-time monitoring of the health apps, would address these challenges and improve reliability.

(JMIR Mhealth Uhealth 2026;14:e68737) doi:10.2196/68737

XSL•FO
RenderX

KEYWORDS

apps; applications; mobile health; mHealth; nutrition; healthy eating; eHealth; digital health

Introduction

As our world becomes increasingly digital, the use of mobile apps for health-related purposes (health apps) is on the rise [1,2]. The World Health Organization's (WHO) Global Observatory of eHealth describes health apps as key components of the broader domain of mobile health (mHealth). The use of mobile devices (eg, smartphones, patient-monitoring tools, personal digital assistants, and other wireless devices) is instrumental to support medical and public health practices. eHealth encompasses the cost-effective and secure use of information and communication technologies for health care and related fields [3]. Since WHO's acknowledgment in 2016 [4], the market has expanded rapidly. There has been an exponential increase in health app availability from 325,000 in 2017 [5] to an estimated 255 billion app downloads in 2022 [1,2].

Despite this growth, health care professionals lack unified standards for evaluating health app quality, safety, and effectiveness [6]. Recent European Union regulations on medical devices, such as Regulation (EU) 2017/745 enacted in May 2021, represent significant development. These regulations classify certain health apps as medical devices, necessitating adherence to specific criteria for approval. Supplementary documents from the European Commission [7,8] complement these regulations. The regulations also establish a medical device database to enhance transparency for both patients and health care providers. However, despite mHealth apps being within the medical device framework and subject to all corresponding laws at the European level, there are pending challenges related to user data management. Commercial platforms typically do not provide reliable information regarding the efficacy or safety of the apps they offer. Health apps should be substantiated by scientific evidence, facilitating their endorsement by health care professionals and enabling end users to benefit from a validated certification system [9,10], which should include the assessment of not only the quality of the information provided but also other relevant factors associated with app quality, such as usability and behavior change promotion [11].

This case study proposed a robust process for the evaluation of health apps, using dietary intervention apps promoting healthy eating as an example, to illustrate both the evaluation method and the challenges faced by end users and health care prescribers. A particular focus was placed on the quality of the health information provided by the apps.

Methods

Study Design

Our approach was structured in two stages: first, the systematic identification of relevant apps and, second, their assessment based on predefined quality, usability, and evidence-based criteria. This sequence aims to be a realistic, reproducible way to support both health care professionals and end users in the navigation and evaluation of the growing mHealth landscape.

App Identification

An initial exploratory analysis was conducted to identify projects, initiatives, and organizations involved in the evaluation of health apps. Most of these resources have been catalogued in two documents by the Spanish Ministry of Health [12] and the European project mHealth-Hub [13,14]. Both include governmental and nongovernmental efforts. Each initiative was then individually reviewed to examine its content and structure. Notable heterogeneity among the initiatives was observed, allowing them to be classified into two main groups: (1) platforms offering pre-evaluated apps, used as databases for the search, and (2) those focused on normative aspects, standardization, and quality assessment. Given the diversity in the approaches and structures of these platforms, the search began with general terms related to the study's focus, such as healthy eating, staying healthy, diet, lifestyle, preventive medicine, weight or healthy habits, which were later adapted to the navigation logic, taxonomy, or filtering system of each initiative (for more details, see Multimedia Appendix 1).

From August to December 2021, we searched for health apps that met the following inclusion criteria: aimed at adults (over 18 years old), available in English or Spanish, and offering nutritional guidance (eg, recommendations for dietary changes or meal planning). Apps that solely functioned as food diaries, calorie counters, or barcode readers but did not include any specific recommendations were excluded. We examined app descriptions and features to confirm eligibility.

In the absence of established guidelines for health app evaluation, we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist (Multimedia Appendix 2) [15,16] adapted to our research objectives. We reviewed 28 certification bodies (see Table S1 in Multimedia Appendix 1), of which only 5 (17.9%) contained apps meeting our inclusion criteria (see Table 1 for additional details):

XSL•FO
RenderX

Table 1. Summary of certification/assessment bodies that included apps promoting healthy eating.

Name	Year it started	Country	Reviewers	Evaluation criteria	Public or private
MyHealthApps[a]	2013	United Kingdom	End users, associations, or caretakers	Scoring system	Private
Healthy Living Apps Guide	2018	Australia	Expert reviewers (at least two on behavior change/public health)	MARS[b] (functionality) and ABACUS[c] (behavior change)	Public (Government of Victoria)
ORCHA[d]	2018	United Kingdom	Experts and reviewers (end users considered)	A 7-step system	Private (cooperation with the NHS[e])
GGD[f] AppStore	2016	Netherlands	Experts and end users	A 4-step assessment based on questionnaires, including a final evaluation of behavior change techniques used	Public (GGD GHOR[g])
Health Navigator	2017	New Zealand	Experts, clinical reviewers, and end users	Internal revision by experts, clinical revision, and end-user revision.	Public Ministry of Health

[a]Closed on May 17, 2022.

[b]MARS: Mobile App Rating Scale.

[c]ABACUS: App Behavior Change Scale.

[d]ORCHA: Organization for the Review of Care and Health Applications.

[e]NHS: National Health Service.

[f]GGD: *Gemeentelijke Gezondheidsdienst* (Municipal Health Service).

[g]GHOR: Geneeskundige Hulpverleningsorganisatie in de Regio (Regional Medical Emergency Preparedness and Planning).

Healthy Living AppsGuide [17] assesses app functionality using the Mobile App Rating Scale (MARS) [18] and behavior change using the App Behavior Change Scale (ABACUS) [19]. Each criterion is rated on a scale of 0-5 stars, with additional considerations, such as app cost and data export capabilities.

MyHealthApps [20] focused on user preferences and developer-related data. This resource operated as a catalog, rather than assigning individual scores to apps. It assessed transparency regarding pricing, contact details, geographical location, and security measures. Unfortunately, it closed in May 2022.

ORCHA (Organization for the Review of Care and Health Applications) [21] uses a seven-step evaluation system comprising three main domains: data, professional guarantee, and usability and accessibility. App functionalities and features are also considered. A maximum score of 100 points is attainable for each domain, with ratings above 65 considered good quality. Scores between 45 and 65 suggest areas for further investigation, and scores below 45 deem the app (or domain) potentially unsafe or ineffective.

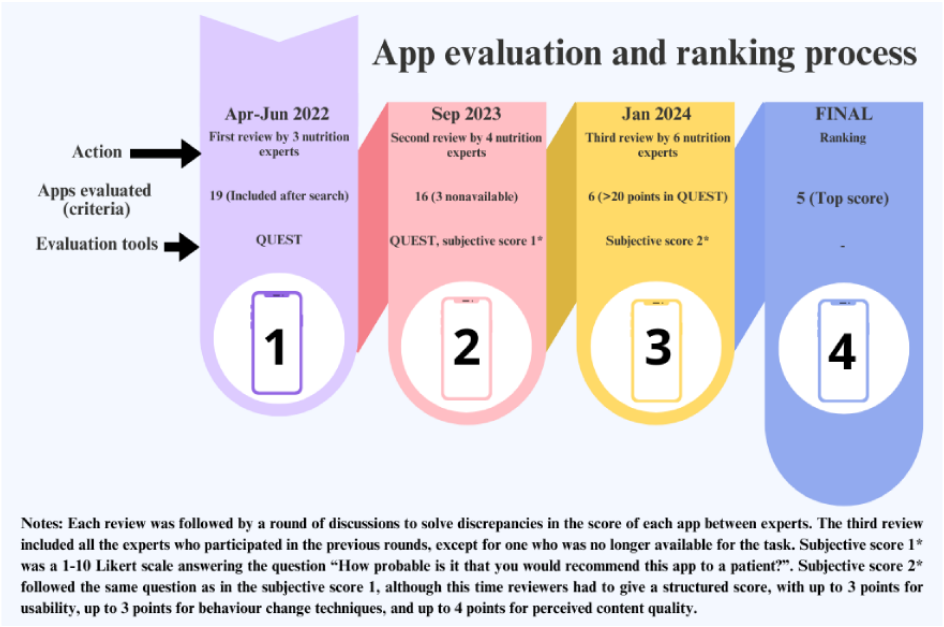
Gemeentelijke Gezondheidsdienst (*GGD*) *AppStore* [22] follows an evaluation method considering app availability, pricing, compatibility with different operating systems, provider contact details, promotion of healthy behavior, data management, and goal setting. To enter evaluation, apps must focus on self-care and include at least two methods for behavior change. The apps are also assessed for usability, reliability, privacy, safety, and relevance, with each characteristic rated as good, sufficient, or inadequate. Apps are scored up to a maximum of 5 stars.

Health Navigator [23] is supported by the New Zealand Ministry of Health and provides a library of reliable apps. The evaluation process encompasses app features, functionalities, quality of information, and relevance to users. Apps must fulfil specific criteria, including evidence-based content and an evaluation of effectiveness, acceptability, and usability, and include a privacy statement. Clinicians with diverse expertise rate the apps from 1 (very poor, not recommended) to a maximum of 5 stars (excellent). End users also provide feedback. Finally, apps deemed clinically unsafe or potentially harmful are excluded.

As health apps are frequently updated, a follow-up search was performed in May 2022. By that time, *MyHealthApps* was no longer active. From this point on, three rounds of app evaluations followed (see Figure 1).

XSL•FO
RenderX

Figure 1. Diagram depicting the evaluation process of the apps.



App Assessment

The apps underwent three rounds of evaluation involving a total of seven reviewers who downloaded and assessed them on the devices described later. An endocrinologist (specialized in endocrinology and nutrition), one dietician, two psychologists, and three epidemiology and public health professionals participated. A second endocrinologist led the evaluation process and coordinated and summarized meetings.

Evaluation Round 1 (April-July 2022)

The content of each of the selected apps was evaluated by pairs from a panel of three nutrition experts (author AD-G, specialized in endocrinology and nutrition; author GZZ, a registered dietician; and Cristina Ruano Rodríguez, a lecturer in nutrition and public health). To facilitate comparison across the apps, we used the Quality Evaluation Scoring Tool (QUEST) [24], not used by any of the evaluating bodies that included the selected apps. QUEST was originally designed to assess the quality of health information in digital media and has been tested for reliability and validity. It offers a total score ranging from 0 to 28 points based on the following criteria: Authorship (0, 1, or 2 points) evaluates the ease of identifying the authors of the content; attribution (0, 3, 6, or 9 points) assesses whether health information is backed by scientific studies; type of study (0, 1, or 2 points, if at least 6 points have been given to *attribution*) evaluates the studies' quality; conflict of interest (0, 3, or 6 points) evaluates whether the information promotes purchase of products or services; complementarity (0 or 1 point) evaluates whether the information supports the health provider-patient relationship; and tone (0, 3, or 6 points) assesses language used as biased, neutral, or acknowledging the limits of knowledge.

The apps were downloaded by the reviewers to Android devices (two Galaxy Tab A7 Lite SM-T220 running on One UI Core 3.1 and Android 11 and a Redmi 9A M2006C3LG MIUI Global 12.0.20 mobile phone running on MIUI Global 12.0.20 Estable [QCDEUXM] and Android 10QP1A.190711.020). Whenever a premium version was available, it also was tested. A data extraction template was developed, including reviewer ID, date of the evaluation, name of the app, developer, version, device where it was downloaded, score for each item in QUEST, and comments. The mean total score was calculated to rank the apps, and inter-reviewer agreement for each QUEST item was measured using weighted κ coefficients. Concordance for the total score was evaluated using the intraclass correlation coefficient. We also explored potential correlations between the QUEST scores, Google Play Store ratings, and the scores from each assessment body (all three, nonnormally distributed quantitative continuous variables) using Spearman's coefficients. For these procedures, we used the *vcd* library in R software (RStudio version 1.3.1056, R Foundation for Statistical Computing). Following the analysis, discrepancies in the evaluation of QUEST items were discussed to identify their underlying causes.

Evaluation Round 2 (September 2023)

An expanded panel of four additional reviewers (authors AT-C and MLA-M, psychologists, experts in behavior change; author ITG, professor in epidemiology, expert in health promotion; and author GS, public health nutritionist, expert in prevention through diet) re-evaluated the apps (4-5 each) using the same methodology but limiting the time spent in the evaluation to a maximum of 45 minutes per app. Apps with a score average over 20 points (out of a maximum of 28) either in the initial evaluation or in the update were selected for further analysis. Although no given cutoff is recommended for the tool, this threshold was chosen to ensure the inclusion of apps that meet a fair-to-high standard. All evaluators were asked to subjectively score (0-10 points) their assigned apps by answering the question "How probable is it that you would recommend this app?" One of the reviewers (AD-G) scored all of them. Apps with a difference of 3 or more points in the initial and updated QUEST scores were discussed within the same reviewer pair (one of the reviewers involved in the first round and the new expert involved in the second round) to solve or explain this discrepancy. This process was documented, summarized, and

XSL•FO
RenderX

then discussed during a meeting involving all the authors, including author AMW (specialist in endocrinology, nutrition, and diabetes). The meeting was recorded, transcribed, and summarized in a document, which was shared with the participants for feedback.

Evaluation Round 3 (January 2024)

The last version of the selected apps was downloaded and assessed by six experts (GZZ, AD-G, AT-C, MLA-M, ITG, and GS), who had participated in the previous rounds. A subjective score was given again, this time assigning up to 4 points for content quality (this aspect being the most relevant for our research), 3 for usability, and 3 for promoting behavior change. The specific question to answer in this round was "How probable is it that you would recommend this app for someone to improve their healthy diet, based on the quality of the content (40%), usability (30%), and promotion of behavior change (30%)?"

Further discussion among all the authors led to the final ranking and the selection of the top 5 apps.

Ethical Considerations

Given the nature of this study, no ethics committee approval or informed consent to participate was needed. All authors declared their consent for publication.

Results

App Details

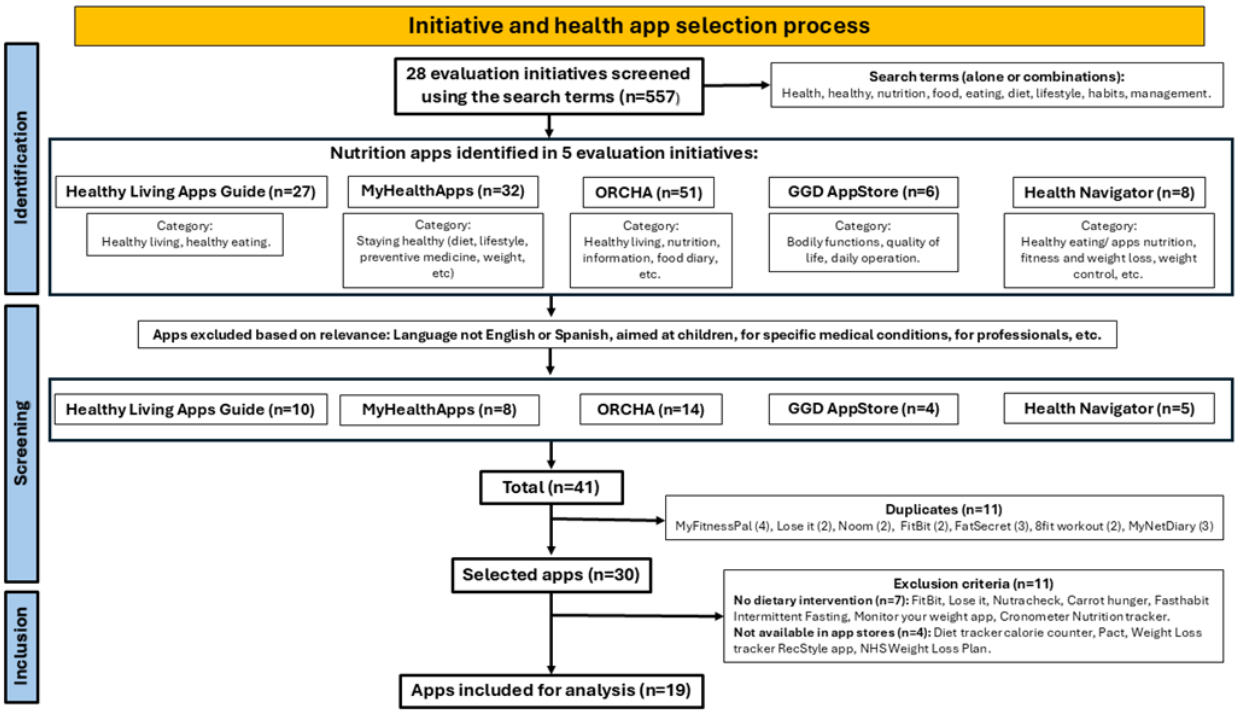
Of the 28 certification bodies identified (see Table S1 in Multimedia Appendix 1), 14 (50%) catalogued a total of 557 apps across various categories, encompassing terms such as "healthy eating," "nutrition," "diet", "healthy practices," and "lifestyle". Only 5 (17.9%) included apps that fulfilled our inclusion criteria (see Table 1). These certification bodies used a variety of criteria in their evaluation processes, in terms of both the elements assessed and the tools used. Nevertheless, there was a significant overlap in the factors considered key for determining app quality, including functionality, usability, engagement, aesthetics, privacy, data protection, and effectiveness in promoting behavior change.

A total of 41 (7.4%) apps met the inclusion criteria. After eliminating duplicates, 30 (73.2%) apps remained, of which 11 (36.7%) met the exclusion criteria (see Figure 2).

The characteristics of the selected 19 (63.3%) apps are summarized in Table 2. Of these 19 apps, 5 (26.3%) were included in two or more of the certification bodies: *MyFitnessPal*, *FatSecret*, *MyNetDiary*, *Noom*, and *8fit Workout & Meal Planner*. Only the apps included in ORCHA had undergone formal evaluation of their content.

As per the definition, all apps included nutritional interventions, but most also combined different strategies or features, making their categorization challenging. Overall, 14 (73.7%) apps included weight control or tracking (numbered 1-6, 8, 11, 12, 14, 16-19 in Table 2), 12 (63.2%) allowed for calorie counting (numbered 1-4, 6-8, 11-13, 17, 19 in Table 2), and 13 (68.4%) facilitated the recording of physical activity (numbered 1-4, 6-9, 14-17, 19 in Table 2). Furthermore, 12 (63.2%) apps (numbered 1-8, 10, 11, 16, 18 in Table 2) incorporated behavior change techniques, such as motivational feedback, health education promotion, and goal setting.

Figure 2. PRISMA 2020 flowchart describing app search and findings. GGD: *Gemeentelijke Gezondheidsdienst* (Municipal Health Service); ORCHA: Organization for the Review of Care and Health Applications; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



XSL•FO
RenderX

Table 2. Main features of the 19 selected and evaluated health apps.

App number	App name; developer	Description
1	Noom; Noom Inc	It uses behavior change psychology to promote healthier habits, weight loss, and health goals. It can track meals and link to a pedometer.
2	YAZIO; YAZIO	It is a food diary and calorie counter that also offers diets and recipes, including vegan, vegetarian, and intermittent fasting options.
3	FatSecret; FatSecret	It is a food diary and calorie counter. It allows the user to scan products barcodes, tracking weight and keeping a record of meals and consumed food photos.
4	MyNetDiary; MyNetDiary Inc	It is a food log and calorie counter that also provides diets and recipes. It has a food database and a barcode scanner. It also provides motivational messages to achieve nutritional goals.
5	DietBet; WayBetter Inc	It aims to help the user achieve healthy weight loss and maintain motivation. It offers a support community, customer service, and expert-led games (coaches, nutritionists, etc).
6	MyFitnessPal; MyFitnessPal Inc	It offers recipes, a meal planner, a calorie counter, and workout plans. It also allows tracking progress and setting macronutrient goals.
7	MyPlate[a]; LIVESTRONG	It allows tracking calories and exercises. It also tracks the user's progress based on calorie and nutrient goals and provides motivation through support groups.
8	LIFESUM; Lifesum	It functions as a food diary and offers various types of diets and recipes. It also allows tracking exercise and water intake and setting weight goals and health data.
9	8fit Workout & Meal Planner; Urbanite Inc	It offers workout routines in different categories (boxing, yoga, high-intensity interval training, etc), along with meal plans.
10	Freeletics Nutrition; Freeletics GmbH	It acts as a nutrition coach to help the user achieve dietary goals. It guides the user toward healthy eating, recipes, etc, and helps them adapt nutrition to personal goals.
11	GetFit-Daily Meal Planner[a] App Prodakshn;OOO	It provides daily and weekly meal plans to lose, maintain, or gain weight based on the user's goals. It also calculates nutrients, calories, water intake, BMI, etc, and includes reminders.
12	HealthifyMe; HealthifyMe	It can act as a calorie counter, provides weight loss and exercise plans, guides sleep hygiene, etc.
13	Eat This Much-Meal Planner; Eat This Much Inc	It is a meal planner based on dietary preferences, budget, etc, and can be configured for different types of diets.
14	LIFE Extend[a]; LifeOmic	It is marketed as a precision health mobile app to help improve users' health using five pillars of health.
15	Fastic Fasting Ap; Fastic GmbH	It promotes intermittent fasting with different levels of intensity (gentle, moderate, intense).
16	Second Nature; Second Nature	It provides a lifestyle change program that helps lose weight and develop healthy habits. Nutritionists guide the user throughout the program. The app also includes a support group.
17	Uplyfe-Precision Nutrition[a]; Uplyfe AG	It is marketed as a customized nutrition and lifestyle change app-guided program based on scientific findings. It also provides personalized nutrition plans, activities, and symptom tracking.
18	Freshwell; Freshwell	Health care personnel promote a low-carb diet. The app provides informative weekly modules, a meal planner, etc.
19	Contador de calorías; Virtuagym	It provides a personalized nutrition plan according to the user's lifestyle and goals.

[a]No longer available.

App Evaluation

Table 3 presents the scores given to each app by the reviewers during the three rounds of assessment, by the certification bodies, and by a commercial app store.

XSL•FO
RenderX

Table 3. Scores by the different evaluation tools and rounds of revision for each app.

Evaluation rounds and app (name; first version evaluated; date of update of that version; year launched)	Scores								
	Round 1: QUEST[a]	Round 2: QUEST	Round 2: subjective	Round 3: subjective	Google Play Store (stars); August 19, 2022	OR-CHA[b]	Health Navigator	GGD[c] App-Store	Healthy Living Apps Guide
Maximum score	28	28	10	10	5	100	5	5	5
Number of evaluators (total, per app)	3, 2	4, 1	6, 6	6, 6	__[d]	—	—	—	—
Apps that went through all evaluation rounds									
Freshwell; 1.1.1; October 25, 2021; 2021	21.5	28	7.2	6.9	5	68	—	—	—
Yazio; 7.10.8; August 8, 2022; 2014	21.5	18	8.2	7.4	4.6	—	—	4.1	—
LIFESUM; 11.0.0; June 30, 2022; 2011	20.5	14	7.2	6.8	4.5	—	—	—	—
MyNetDiary; 8.1.1; August 12, 2022; 2010	18.5	22	6.2	6.8	4.6	79	—	—	2.5
Second Nature; 6.10.2; August 11, 2022; 2016	12	28	8	8.5	4.6	84	—	—	—
Freeletics Nutrition; 1.27.12; January 27, 2020; 2016	7.5	21	6.7	5.5	3.8	—	—	—	2.5
Apps included in the first two rounds only									
HealthifyMe; 18.7.1; August 1, 2022; 2013	21.5	13	2.5	—	4.4	—	—	—	2
LIFE Extend[e]; 5.8.3; August 9, 2022; 2019	20.5	24	5.5	—	4.1	74	—	—	—
Noom; 10.28.0; August 17, 2022; 2011	17.5	12	—	—	4.2	80	5	—	—
Fastic Fasting App; 1.105.0; August 16, 2022; 2019	14	11	—	—	4.7	70	—	—	—
Fatsecret; 9.12.5.2; May 25, 2022; 2007	11	10	—	—	4.7	—	4	4	—
Eat This Much-Meal Planner; 2.0.12; August 12, 2022; 2016	11	13	—	—	4.4	—	—	—	2
8fit Workout & Meal Planner; 22.4.0; May 9, 2022; 2014	10	14	—	—	4.3	75	—	—	2.5
DietBet; 18.0.0; August 4, 2022; 2014	8.5	6	—	—	4.7	—	—	—	2.5
MyFitnessPal; 22.15.0; August 3, 2022; 2010	6.5	5	—	—	4.4	47	3	3	—
Contador de calorías; 3.7.3; April 28, 2022; 2014	6	6	—	—	4.5	—	—	4.5	—
Apps included in the first round only									
GetFit-Daily Meal Planner [e]; 1.3.4; April 23, 2023; 2020	16.5	—	—	—	1	—	—	—	2.5
MyPlate[e]; 3.5.3(63); September 30, 2021; 2015	14	—	—	—	4.6	—	—	—	—
Uplyfe-Precision Nutrition[e]; 1.8.1; September 21, 2021; 2020	13.5	—	—	—	4.2	70	—	—	—

XSL•FO
RenderX

[a]QUEST: Quality Evaluation Scoring Tool.

[b]ORCHA: Organization for the Review of Care and Health Applications.

[c]GGD: *Gemeentelijke Gezondheidsdienst* (Municipal Health Service).

[d]Not applicable.

[e]Became unavailable during the evaluation process.

Evaluation Round 1

Table S2 in Multimedia Appendix 1 displays the score of each QUEST item in the first round of evaluations, categorized by reviewer, with apps ranked according to their mean total score. We explored potential correlations between the QUEST scores, Play Store ratings, and the scores of each certification body and found no significant correlations among any combination (data not shown). Agreement reached between reviewers was rather low, with most κ coefficient values below 0.6 and the intercorrelation coefficient between 0.50 and 0.67 (see Table S3 in Multimedia Appendix 1). The main reasons for discrepancy were attributed to the difficulty in finding the information needed to complete the QUEST scale.

Evaluation Round 2

Four additional experts were asked to update the app review on an Android device and were given a maximum of 45 minutes per app to find the information. At the time of this evaluation, 3 (15.8%) apps (*MyPlate*, *GetFit-Daily Meal Planner*, and *Uplyfe-Precision Nutrition*) were no longer available, and the remaining 16 (84.2%) apps were scored (see Table 3). One of the reviewers used an iOS device (Apple iPhone) due to unavailable Android devices. The 7 (43.8%) apps that received a score above 20 (out of a maximum of 28) either in the first evaluation or in the update were selected for further assessment.

Challenges were encountered assessing the items "authorship," "attribution," and "tone." Some apps (particularly *LIFESUM*, *Freeletics Nutrition*, *Eat This Much-Meal Planner*, *LIFE Extend*, and *Second Nature*) lacked clear indications regarding authors and sources, necessitating external web searches to retrieve these data, and even then, this took considerable time. Additionally, reviewers noted that the "attribution" item could be easily altered to achieve higher scores; for example, referencing a source, even if it was a low-quality source or irrelevant to the app's health information, could lead to a high rating, as long as it was a highly scored type of study. Subjectivity was also a significant factor in evaluating the "tone" item, leading to disparities in scores. Finally, some of the discrepancies were explained by the differences in app versions (eg, *Freeletics Nutrition*, *Fresh Well*, and *Second Nature* improved considerably between review rounds, coinciding with version changes, whereas the opposite was true for *HealthifyMe*). In the case of *HealthifyMe*, this (and the low subjective score) led to its exclusion from the third round, despite scoring above 20 points in the first round. Discussion about the subjective scores and the contribution by the reviewer scoring all the selected apps in this round led to a more structured scoring procedure in the third round.

Evaluation Round 3

All experts participating in the previous rounds of evaluation, except one who was no longer available for the task, took part in this round. At the time of this evaluation (early 2024), LIFE Extend was not available, so only 6 (31.6%) apps were included in the third round. Discrepancies in the scores were discussed among researchers.

Subjective scores were also given to the selected apps by all six reviewers (see Table S4 in Multimedia Appendix 1 and Table 3). The scores were discussed in a joint meeting, which led to the selection of the top 5 (8.3%) apps: *Second Nature*, *Freshwell*, *Yazio*,*LIFESUM*, and *MyNetDiary*. The scores given by each evaluator are included in Multimedia Appendix 1.

Second Nature was consistently the most highly scored app and is also endorsed by the National Health Service (NHS) of the United Kingdom. Its present version includes clear references to the evidence supporting its contents. Usability is good and improved by the inclusion in the update of interactive content, such as videos with advice and encouragement. There is a wide variety of dietary patterns to choose from, and recipes with visual backup are provided. Furthermore, most of the features are free to use. However, food registering is rather cumbersome, and the tone of the recommendations could be more cautious.

Freshwell was developed by two British general practitioners and is used by the NHS in the United Kingdom. Usability is good, with pictures, recipes, and explanations. Long-term goals are included, promoting behavior change. A classification of foods is included based on the type of dietary pattern that is promoted (low carb). Study references are provided to support this recommendation, and a disclaimer is provided stating that the app provides educational content and is not to be considered medical advice. People with chronic conditions are advised to ask their health care provider. Other health-promoting pieces of advice include reducing sugar intake, snacks, and alcohol. The subjective score penalized the low-carb type of dietary pattern on which the app is based, since reviewers considered that there is not sufficient evidence to promote it to the general public as the "default" dietary pattern [25].

Yazio has high usability due to likable aesthetics, ease of use, accessibility, and the challenges included in the app. The challenges, which promote behavior change, can be chosen by the user from a list, including quitting chocolate, sugar, sweets, fast food, etc. The app includes information about healthy eating, which is supported by scientific evidence, and the standard recommended diet is Mediterranean style. However, it also includes dietary patterns with less scientific evidence of benefit, such as intermittent fasting or the keto diet. In fact, intermittent fasting is a prominent feature of this app, including challenges on how long you can fast for. A disclaimer recommends not using this pattern in the long term and seeing a physician if that is the intention of the user.

LIFESUM has a visually pleasant interface and agile and intuitive navigation, both of which lead to good user experience. Different objectives can be set, and the app includes easy and

XSL•FO
RenderX

advanced recipes, which help with organization and motivation for healthier eating, although all this is only available in the premium version. The app includes a score for easy comparison and progress tracking, but the criteria for the score are not transparent. Studies supporting the contents were not easily found and were not there for all the recommendations. Once again, the latter included intermittent fasting and keto diets. Minor bugs were identified in the form of nonfunctioning links.

MyNetDiary offers an attractive design, easy navigation, and a variety of resources to engage users. It provides information about different types of diets and highlights their key aspects. Examples of recipes for each type of diet are included with detailed descriptions. The content is supported by a trained specialist in the nutritional management of diabetes and other health problems. A diary function to track daily food intake is included, and data can be entered via text, audio, or barcode scanning. The premium version offers daily advice and feedback. Users can also access a nutrition blog written by the same specialist behind the app. Professionals can subscribe to connect with users interested in weight loss.

Discussion

Principal Findings

In this paper, we highlighted the need for a common evaluation framework to assess the quality of health apps and proposed a robust approach to this end. Our use case focused on apps promoting healthy eating. A total of 19 apps were included, which had already been positively assessed by dedicated certification bodies. The latter used a variety of evaluation criteria, and most did not include content quality among them. Following a rigorous, three-step evaluation process by seven experts, only 5 apps were recommended. The selection was based on a comprehensive assessment of the quality of their contents (using QUEST), as well as their usability and the use of behavior change techniques.

Health apps have emerged as promising tools to promote healthier lifestyles, due to their widespread accessibility, user friendliness, and cost-effectiveness [26-28]. However, the app market's diversity and the lack of systematic evaluation processes leave consumers with limited tools to judge which apps are effective, safe, and suited to their needs.

To address this issue, several national and international actions have attempted to assess health app quality and safety [17,20-23]. Our investigation identified 28 initiatives dedicated to the evaluation of health apps, 5 including apps fulfilling our inclusion criteria. Although each of these initiatives uses distinct criteria for app evaluation, they all generally emphasize similar key factors, such as data security, privacy, ease of use, accessibility, usability, support for user-health care professional communication, personalization options, and capacity to induce behavior change. In a recent review, the importance of adherence to scientific evidence, as well as additional features (eg, gamification and cocreation of the app with health professionals and users), were supported [29]. Although a multitude of tools and scores have been proposed to evaluate different aspects of app quality, as of today, one single robust procedure or set of criteria does not exist [30-32], although some national and international evaluation efforts have been developed [33-35]. Additionally, all the certification bodies reviewed in this study have cautioned that they cannot guarantee the precision, quality, trustworthiness, and effectiveness of the health apps they assess, increasing uncertainty for the users.

Acknowledging these challenges, our study sought to evaluate nutrition-focused health apps using a tool aimed at the health information within the apps, assuming that health information based on evidence would support safety and effectiveness. We selected QUEST to assess the apps, since it has been specifically designed to evaluate health information and incorporates items such as attribution or study type to evaluate the quality of information sources. Additionally, it assesses items such as conflict of interest and tone to evaluate how information is presented to the user, where any sort of advertising is penalized. We found no significant correlations among the Play Store scores for the apps, the QUEST total mean score, or any of the certifier scores, probably because they assess different dimensions. QUEST is a robust tool for the assessment of content quality but does not address usability, engagement, or behavior change techniques, all factors known to influence the perceived quality of an app [11]. Therefore, the research team decided to complement this assessment with a subjective score, which also included these aspects. Such factors may influence the app store ratings, and many are directly evaluated by the certification bodies but are overlooked by QUEST. In contrast, Play Store scores are based on user opinions, which can be influenced not only by all these factors but also by popularity, endorsement by influencers, aesthetics, or alignment with their own nutritional preconceptions, which might not be backed by scientific evidence.

Furthermore, it is important to acknowledge that apps are dynamic and subject to constant updates. Although the latter can be beneficial if they enhance the quality of app contents, they also pose a challenge for app evaluation, as acknowledged by a recent consensus [36]. Thus, continuous reassessment would be necessary to account for the changes. Access to older app versions may not always be possible, and certain apps may become completely unavailable over time. Any app marketed as a health app should be subject to special scrutiny, particularly regarding collection of sensitive data and potential for harmful effects on users. Previously proposed improvements include establishing national lists of tested and trusted health apps, creating a catalog of health apps accessible to patients only if prescribed by professionals and guidelines for app developers toward evidence-based, unbiased, high-quality apps, with wider assessment requirements for higher-risk tools [14,33,36-39].

Strengths and Limitations

Our work has several strengths that contribute to the field of health app evaluation. First, the adoption of QUEST to assess all the apps constitutes a rigorous approach for the evaluation of content quality. This systematic assessment of the content was complemented by a subjective evaluation by the panel of usability and behavior change techniques. Additionally, the study's methodological approach, involving multiple, independent reviewers and the evaluation of a diverse range of

XSL•FO
RenderX

apps, strengthens the validity and comprehensiveness of the results. Finally, the study's alignment with the existing literature and national and international certification/assessment bodies highlights its relevance and potential impact on improving health app quality and user safety.

We acknowledge that the study also has some limitations. The small sample size of apps reviewed and their focus on nutrition-related interventions may limit the generalizability of our results, since they represent an incomplete picture of the overall quality and safety of available health apps. The decision to exclude apps that solely function as food diaries and calorie counters may be controversial. We chose apps offering recommendations, meal plans, and recipes, because we focused on the quality of the information provided. Nevertheless, we acknowledge that food diaries and calorie counting could also help the users become more aware of their eating habits and induce a change in behavior despite the lack of any other specific nutritional advice and therefore be considered a type of intervention. Another limitation is the subjective nature (eg, tone) or the scoring difficulty (eg, attribution) of some of the items in QUEST. Discrepancies in the first round of evaluations, reflected by rather low κ coefficient values, were mainly attributed to the fact that the information to score some items might be difficult to find in the app or might even be found behind a link and thus not be obvious. Hence, in the second round of evaluations using QUEST, we standardized the time spent on the assessment of each app. In addition, some aspects of app quality are overlooked by QUEST (eg, usability, engagement, and capacity to induce positive behavior changes). Thus, QUEST is not a comprehensive enough tool to judge app quality on its own, but it is complementary to other measures. Finally, to overcome these difficulties, several rounds of revision were made, which made the evaluation process longer than initially planned.

To summarize, during this review, we encountered several challenges. In the first place, numerous certification bodies were scanned for evaluated apps. Creating a common repository collecting their criteria and decisions would facilitate this task [40]. In addition, the rapid change in mHealth apps poses a challenge to the timeliness of evaluation. This might be mitigated by the development of living systematic reviews, which are continuously updated [39] or the use of rapid reviews [41] and evidence maps [42]. Lastly, there is currently no standardized procedure to evaluate app quality. To address this, we included what we considered a critical and often overlooked dimension in previous evaluation efforts: a thorough assessment of the quality of the health information provided by the apps using QUEST scoring.

Conclusion

In conclusion, despite their previous evaluations by various certification bodies, only 5 of 19 apps promoting healthy eating met the quality standards set by our experts. Our study calls for enhanced scrutiny and regulatory measures to ensure that health apps meet rigorous standards of accuracy, reliability, and user safety. Guidelines for developers for the design of evidence-based, unbiased, high-quality health apps, as well as technological solutions for real-time monitoring of the apps, would address these challenges and improve reliability.

Acknowledgments

We thank Cristina Ruano Rodríguez (University of Las Palmas de Gran Canaria) for participating in the first round of evaluations. We would also like to thank Lilisbeth Perestelo (Servicio de Evaluación, Servicio Canario de la Salud) for fruitful methodological discussions. Where authors are identified as personnel of the International Agency for Research on Cancer or the World Health Organization (WHO), the authors alone are responsible for the views expressed in this paper, and the views do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer or WHO. This study was performed within WARIFA "(Watching the Risk Factors: Artificial Intelligence and the Personalised Prevention and Management of Chronic Conditions), funded by the European Union's Horizon 2020 research and innovation program (grant agreement 101017385). The sponsor played no role in the design; the collection, analysis, and interpretation of data; or the decision to publish the results. During the study, AMW received funding from the Instituto de Salud Carlos III (PI20/00846, PMP21/00069, and DTS 19/00067), the Fundación Mapfre Canarias 2023, the Sociedad Española de Diabetes, and the Fundación Canaria del Instituto de Investigación Sanitaria de Canarias (PIFIISC20/16 and PIFIISC22/08).

Data Availability

All relevant data are available for direct review within the paper.

Authors' Contributions

AD-G, GZZ, and AMW designed the study and summarized and analyzed the results. AD-G and GZZ performed the search and first round of assessments and drafted the manuscript. ITG, GS, AT-C, and MLA-M performed the second round of assessments. All authors participated in the discussions and final ratings of the apps, participated in the revision of the paper, and approved the final version. During the preparation of this work, the authors did not use any artificial intelligence tools/services at any time. The authors are responsible for the entire content of this publication.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Initial search strategy, list of certifying bodies examined, scores given to apps by evaluators, interobserver concordance analysis, and subjective assessments by reviewers.
[DOCX File , 60 KB - mhealth_v14i1e68737_app1.docx]

Multimedia Appendix 2
Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist.
[PDF File (Adobe PDF File), 86 KB - mhealth_v14i1e68737_app2.pdf]

References

1. Global digital health business outlook survey – 2022. Research2Guidance. 2022. URL: https://research2guidance.com/product/global-digital-health-business-outlook-survey-2022/ [accessed 2026-01-14]
2. Descargas mundiales de aplicaciones (apps móviles). Statista. URL: https://es.statista.com/estadisticas/574024/numero-de-descargas-mundiales-de-apps-mundo/ [accessed 2024-05-19]
3. Tuberculosis: global task force on digital health. World Health Organization. URL: https://www.who.int/news-room/questions-and-answers/item/tuberculosis-global-task-force-on-digital-health [accessed 2024-05-19]
4. Peterson CB, Hamilton C, Hasvold P, World Health Organization, Regional Office for Europe. From innovation to implementation: eHealth in the WHO European region. 2016 Sep 30. URL: https://www.who.int/europe/publications/i/item/9789289051378 [accessed 2026-01-14]
5. 325,000 mobile health apps available in 2017 – Android now the leading mHealth platform. Research 2Guidance. 2017. URL: https://research2guidance.com/325000-mobile-health-apps-available-in-2017/ [accessed 2026-01-14]
6. Daraz L, Morrow A, Ponce O, Beuschel B, Farah M, Katabi A, et al. Can patients trust online health information? A meta-narrative systematic review addressing the quality of health information on the internet. J Gen Intern Med 2019 Sep;34(9):1884-1891 [FREE Full text] [doi: 10.1007/s11606-019-05109-0] [Medline: 31228051]
7. Guidance - MDCG endorsed documents and other guidance. European Commission. URL: https://health.ec.europa.eu/medical-devices-sector/new-regulations/guidance-mdcg-endorsed-documents-and-other-guidance_en [accessed 2026-01-14]
8. Manual on borderline and classification in the community regulatory framework for medical devices. Version 1.22. European Commission. URL: https://health.ec.europa.eu/system/files/2020-08/md_borderline_manual_05_2019_en_0.pdf [accessed 2026-01-14]
9. Unsworth H, Dillon B, Collinson L, Powell H, Salmon M, Oladapo T, et al. The NICE Evidence Standards Framework for digital health and care technologies - developing and maintaining an innovative evidence framework with global impact. Digit Health 2021;7:20552076211018617 [FREE Full text] [doi: 10.1177/20552076211018617] [Medline: 34249371]
10. Unsworth H, Wolfram V, Dillon B, Salmon M, Greaves F, Liu X, et al. Building an evidence standards framework for artificial intelligence-enabled digital health technologies. Lancet Digital Health 2022 Apr;4(4):e216-e217 [FREE Full text] [doi: 10.1016/s2589-7500(22)00030-9]
11. Deniz-Garcia A, Fabelo H, Rodriguez-Almeida AJ, Zamora-Zamorano G, Castro-Fernandez M, Alberiche Ruano MDP, WARIFA Consortium. Quality, usability, and effectiveness of mHealth apps and the role of artificial intelligence: current scenario and challenges. J Med Internet Res 2023 May 04;25:e44030 [FREE Full text] [doi: 10.2196/44030] [Medline: 37140973]
12. Martín Fernández A, Marco Cuenca G, Salvador Oliván JA. Evaluación y acreditación de las aplicaciones móviles relacionadas con la salud. Rev Esp Salud Pública 2020;94(11 de agosto):e1-e11 [FREE Full text]
13. Assessment frameworks. European mHealth Hub. URL: https://web.archive.org/web/20211207214114/https://mhealth-hub.org/assessment-frameworks [accessed 2026-01-14]
14. Health apps repositories in Europe Internet. European mHealth Hub. URL: https://web.archive.org/web/20211024211815/https://mhealth-hub.org/health-apps-repositories-in-europe [accessed 2026-01-14]
15. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
16. Page M, McKenzie J, Bossuyt P, Boutron I, Hoffmann T, Mulrow C, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
17. Healthy living apps. VicHealth. URL: https://web.archive.org/web/20211116162051/https://www.vichealth.vic.gov.au/media-and-resources/vichealth-apps [accessed 2026-01-14]
18. Terhorst Y, Philippi P, Sander L, Schultchen D, Paganini S, Bardus M, et al. Validation of the Mobile Application Rating Scale (MARS). PLoS One 2020;15(11):e0241480 [FREE Full text] [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
19. McKay F, Slykerman S, Dunn M. The App Behavior Change Scale: creation of a scale to assess the potential of apps to promote behavior change. JMIR Mhealth Uhealth 2019 Jan 25;7(1):e11130 [FREE Full text] [doi: 10.2196/11130] [Medline: 30681967]

XSL•FO
RenderX

20. my health apps. myhealthapps. URL: https://web.archive.org/web/20210725085246/https://myhealthapps.net/ [accessed 2026-01-14]
21. Delivering safe digital health. ORCHA. URL: https://appfinder.orcha.co.uk/about [accessed 2026-01-14]
22. The best apps for your health!. GGD AppStore. URL: https://web.archive.org/web/20211203144501/https://www.ggdappstore.nl/Appstore/Homepage/Sessie,Medewerker,Button [accessed 2026-01-14]
23. App library. Health Navigator. URL: https://web.archive.org/web/20211204173901/https://www.healthnavigator.org.nz/apps/ [accessed 2026-01-14]
24. Robillard J, Jun J, Lai J, Feng T. The QUEST for quality online health information: validation of a short quantitative tool. BMC Med Inform Decis Mak 2018 Oct 19;18(1):87 [FREE Full text] [doi: 10.1186/s12911-018-0668-9] [Medline: 30340488]
25. Hassapidou M, Vlassopoulos A, Kalliostra M, Govers E, Mulrooney H, Ells L, et al. European Association for the Study of Obesity position statement on medical nutrition therapy for the management of overweight and obesity in adults developed in collaboration with the European Federation of the Associations of Dietitians. Obes Facts 2023 Jan 15;16(1):11-28 [FREE Full text] [doi: 10.1159/000528083] [Medline: 36521448]
26. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
27. Paramastri R, Pratama S, Ho D, Purnamasari S, Mohammed A, Galvin C, et al. Use of mobile applications to improve nutrition behaviour: a systematic review. Comput Methods Programs Biomed 2020 Aug;192:105459 [FREE Full text] [doi: 10.1016/j.cmpb.2020.105459] [Medline: 32234632]
28. Villinger K, Wahl D, Boeing H, Schupp H, Renner B. The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: a systematic review and meta-analysis. Obes Rev 2019 Oct;20(10):1465-1484 [FREE Full text] [doi: 10.1111/obr.12903] [Medline: 31353783]
29. Molina-Recio G, Molina-Luque R, Jiménez-García AM, Ventura-Puertos P, Hernández-Reyes A, Romero-Saldaña M. Proposal for the user-centered design approach for health apps based on successful experiences: integrative review. JMIR Mhealth Uhealth 2020 Apr 22;8(4):e14376 [FREE Full text] [doi: 10.2196/14376] [Medline: 32319965]
30. Health information quality assessment tools. OpenMD. URL: https://openmd.com/guide/assessing-health-information-quality [accessed 2026-01-14]
31. Lagan S, Sandler L, Torous J. Evaluating evaluation frameworks: a scoping review of frameworks for assessing health apps. BMJ Open 2021 Mar 19;11(3):e047001 [FREE Full text] [doi: 10.1136/bmjopen-2020-047001] [Medline: 33741674]
32. Wisniewski H, Liu G, Henson P, Vaidyam A, Hajratalli N, Onnela J, et al. Understanding the quality, effectiveness and attributes of top-rated smartphone health apps. Evid Based Ment Health 2019 Feb;22(1):4-9 [FREE Full text] [doi: 10.1136/ebmental-2018-300069] [Medline: 30635262]
33. Segur-Ferrer J, Moltó-Puigmartí C, Pastells-Peiró R, Vivanco-Hidalgo RM. Line of methodological developments of the Spanish Network of Health Technology Assessment Agencies and National Health System Services. In: Health Technology Assessment Framework: Adaptation for Digital Health Technology Assessment. Madrid; Barcelona: Ministry of Health; Agency for Health Quality and Assessment of Catalonia; 2023.
34. Promoting a quality label for health apps!. Label2Enable. URL: https://label2enable.eu/ [accessed 2026-01-14]
35. Neal D, Engelsma T, Tan J, Craven M, Marcilly R, Peute L, et al. Limitations of the new ISO standard for health and wellness apps. Lancet Digital Health 2022 Feb;4(2):e80-e82 [FREE Full text] [doi: 10.1016/s2589-7500(21)00273-9]
36. Evidence standards framework for digital health technologies. National Institute for Health and Care Excellence (NICE). URL: https://www.nice.org.uk/what-nice-does/digital-health/evidence-standards-framework-esf-for-digital-health-technologies [accessed 2026-01-14]
37. van Haasteren A, Gille F, Fadda M, Vayena E. Development of the mHealth App Trustworthiness checklist. Digit Health 2019;5:2055207619886463 [FREE Full text] [doi: 10.1177/2055207619886463] [Medline: 31803490]
38. Dawson R, Felder T, Donevant S, McDonnell K, Card E, King C, et al. What makes a good health 'app'? Identifying the strengths and limitations of existing mobile application evaluation tools. Nurs Inq 2020 Apr;27(2):e12333 [FREE Full text] [doi: 10.1111/nin.12333] [Medline: 31854055]
39. Elliott J, Synnot A, Turner T, Simmonds M, Akl E, McDonald S, Living Systematic Review Network. Living systematic review: 1. Introduction-the why, what, when, and how. J Clin Epidemiol 2017 Nov;91:23-30 [FREE Full text] [doi: 10.1016/j.jclinepi.2017.08.010] [Medline: 28912002]
40. Lopez-Alcalde J, Susan Wieland L, Barth J, Grainger R, Baxter N, Heron N, et al. Methodological challenges in systematic reviews of mHealth interventions: survey and consensus-based recommendations. Int J Med Inform 2024 Apr;184:105345 [FREE Full text] [doi: 10.1016/j.ijmedinf.2024.105345] [Medline: 38309237]
41. Garritty C, Gartlehner G, Nussbaumer-Streit B, King V, Hamel C, Kamel C, et al. Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. J Clin Epidemiol 2021 Feb;130:13-22 [FREE Full text] [doi: 10.1016/j.jclinepi.2020.10.007] [Medline: 33068715]
42. White H, Albers B, Gaarder M, Kornør H, Littell J, Marshall Z, et al. Guidance for producing a Campbell evidence and gap map. Campbell Syst Rev 2020 Dec;16(4):e1125 [FREE Full text] [doi: 10.1002/cl2.1125] [Medline: 37016607]

XSL•FO
RenderX

Abbreviations

ABACUS: App Behavior Change Scale
GGD: *Gemeentelijke Gezondheidsdienst* (Municipal Health Service)
MARS: Mobile App Rating Scale
mHealth: mobile health
NHS: National Health Service
ORCHA: Organization for the Review of Care and Health Applications
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QUEST: Quality Evaluation Scoring Tool

Edited by L Buis; submitted 12.Nov.2024; peer-reviewed by M Agbede, S Mitra, B Eapen; comments to author 17.Apr.2025; revised version received 20.May.2025; accepted 14.Nov.2025; published 30.Jan.2026.

Please cite as:
Zamora Zamorano G, Déniz-García A, Torres-Castaño A, Álvarez-Malé ML, Gram IT, Skeie G, Wägner AM
The Landscape of Mobile Apps for Healthy Eating: Case Study for a Systematic Review and Quality Assessment
JMIR Mhealth Uhealth 2026;14:e68737
URL: https://mhealth.jmir.org/2026/1/e68737
doi:10.2196/68737
PMID:

©Garlene Zamora Zamorano, Alejandro Déniz-García, Alezandra Torres-Castaño, María Luisa Álvarez-Malé, Inger Torhild Gram, Guri Skeie, Ana M Wägner. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 30.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Robust Assessment of Free-Living Physical Behaviors and Activity Intensity Using Dual-Wearable Multitask Learning: Development and Evaluation Study From the Multicenter WEALTH Project

Luis Sigcha[1,2,3], PhD; Annika Swenne[4,5], MSc; Grainne Hayes[3,6], PhD; Jitka Kuhnova[7], PhD; Richard Cimler[7], PhD; Steriani Elavsky[8], PhD; Tomas Vetrovsky[7,9], MD, PhD; Léopold Fezeu Kamedjie[10], MD, PhD; Jérôme Bouchan[10], PhD; Jean-Michel Oppert[10,11], MD, PhD; Janas Harrington[12], PhD; Greet Cardon[13], PhD; Antje Hebestreit[4], PhD; Alan Donnelly[3,6], PhD; Pepijn Van de Ven[2,3], PhD; Christoph Buck[4], PhD; WEALTH consortium[14,15]

[1]Centro ALGORITMI/LASI, School of Engineering, University of Minho, Campus de Azurém, University of Minho, Guimaraes, Portugal

[10]INSERM U1153, INRAE U1125, CNAM, Centre de Recherche en Epidémiologie et Statistiques (CRESS) Équipe de Recherche en Épidémiologie Nutritionnelle (EREN), Université Sorbonne Paris Nord et Université Paris Cité, Bobigny, France

[11]Department of Nutrition, Pitie-Salpetriere hospital (AP-HP), Sorbonne Université, Paris, France

[12]HRB Centre for Health and Diet Research, School of Public Health, University College Cork, Cork, Ireland

[13]Department of Movement and Sports Sciences, Ghent University, Ghent, Belgium

[14]See Acknowledgments

[15]

[2]Data-Driven Computer Engineering (D2iCE) Research Centre, Department of Electronic and Computer Engineering, University of Limerick, Limerick, Munster, Ireland

[3]Health Research Institute, University of Limerick, Limerick, Ireland

[4]Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany

[5]Faculty of Mathematics and Computer Science, University of Bremen, Bremen, Germany

[6]Department of Physical Education and Sport Sciences, University of Limerick, Limerick, Ireland

[7]Faculty of Science, University of Hradec Kralove, Hradec Kralove, Czech Republic

[8]Department of Human Movement Studies, University of Ostrava, Ostrava, Czech Republic

[9]Faculty of Physical Education and Sport, Charles University, Prague, Czech Republic

Corresponding Author:
Luis Sigcha, PhD
Centro ALGORITMI/LASI, School of Engineering, University of Minho, Campus de Azurém, University of Minho, Guimaraes, Portugal

Abstract

Background: Accurate assessment of physical behaviors (PBs) and activity intensity is essential for public health research and digital health monitoring. Wearable accelerometers combined with machine learning (ML) or deep learning (DL) enable objective behavior assessment, but most existing models are trained on laboratory data, limiting generalizability to free-living conditions.

Objective: This study aimed to develop and evaluate multitask ML and DL models for PB classification across 7 categories (sitting, standing, walking, running, sports, cycling, and lying) and activity intensity categories (AIC) across 3 levels (sedentary, light, and moderate-to-vigorous physical activity) using thigh-worn (activPAL) and waist-worn (ActiGraph) wearable accelerometers. A second objective was to compare model-derived estimates of daily time spent in PB and AIC across single- and dual-sensor (activPAL + ActiGraph) configurations, and to evaluate agreement between the best-performing model and corresponding estimates obtained from the proprietary activPAL classification of real-world everyday activities (CREA) algorithm using free-living data collected over a 9-day monitoring period.

Methods: Data were obtained from 590 adults in the multicenter WEALTH study (627 recruited) and included up to 9 days of concurrent activPAL and ActiGraph free-living recordings. Sparse accelerometer-labeled data were obtained using ecological momentary assessment and refined by retaining instances with ≥75% agreement with the CREA algorithm. Resulting labeled data of 583 participants were used to develop ML models for single-sensor (activPAL or ActiGraph) and combined (dual-sensor) configurations. A random forest (RF) model using engineered features and a multihead convolutional neural network (MH-CNN) were trained within a multitask learning framework to jointly predict PB (task 1) and AIC (task 2) using a subject-independent hold-out split. The test subset (n=87) was used to estimate daily time spent in PB and AIC over 9 days, which were compared with CREA-derived estimates using Pearson coefficients and intraclass correlation coefficients (ICCs).

XSL•FO
RenderX

Results: The dual-sensor configuration consistently outperformed single-sensor models. For PB classification, the MH-CNN achieved the highest performance (F_1-score=0.750). For AIC, the RF model performed best (F_1-score=0.741). Dual-sensor free-living estimates showed epidemiologically plausible distributions across the 24-hour period, including sitting 37% (538/1440 min), lying 34% (496/1440 min), walking 9% (131/1440 min), and moderate-to-vigorous physical activity (MVPA) 2% (31/1440 min). Agreement with CREA was strongest for standing, walking, and cycling (r≥0.86; ICC ≥0.72), while lying showed modest reliability (ICC=0.48). For AIC, agreement was highest for light physical activity (LPA) and MVPA (ICC 0.72 - 0.75).

Conclusions: Multitask models combining thigh- and waist-worn accelerometers provide consistent estimates of PB and AIC under free-living conditions. The dual-sensor approach yielded more stable and epidemiologically coherent estimates than single-sensor methods, supporting its potential for large-scale population monitoring and mobile health apps.

International Registered Report Identifier (IRRID): RR2-10.2196/preprints.70186

(JMIR Mhealth Uhealth 2026;14:e94302) doi:10.2196/94302

KEYWORDS

accelerometers; convolutional neural networks; deep learning; human activity recognition; physical activity; sedentary behavior

Introduction

Robust assessment of movement and posture in free-living conditions is fundamental to population surveillance, behavioral interventions, and remote health monitoring. Recent advances in mobile health (mHealth) apps and wearable technologies have enabled more objective and scalable measurement beyond laboratory or clinical environments. However, variability in measurement approaches continues to limit comparability across studies and applications [1-3]. Consequently, there is a clear need for reliable, standardized methods capable of capturing physical behavior (PB) and activity intensity in real-world settings.

In this context, PB refers to posture- and movement-based activity types that describe how time is accumulated across the day, including sedentary postures (eg, sitting or lying), upright postures (eg, standing), and movement behaviors (eg, walking, running, or cycling) [4]. Activity intensity complements PB by characterizing the intensity associated with these behaviors. In population-based research, activity intensity is commonly operationalized using metabolic equivalents of task (METs), which can be used to estimate energy expenditure [5,6]. More commonly, METs are categorized into activity intensity categories (AIC), including sedentary behavior, light physical activity (LPA), and moderate-to-vigorous physical activity (MVPA), using established cut-points from the Compendium of Physical Activities [7] (eg, sedentary behavior 1.0 - 1.5 METs; LPA 1.6 - 2.9 METs; MVPA ≥3.0 METs), supported by consensus definitions in the sedentary behavior literature [8].

Wearable accelerometers have become central to mHealth research because they enable continuous monitoring of physical activity (PA) and sedentary behavior under free-living conditions [9]. Thigh-worn devices such as the activPAL have demonstrated strong validity for posture classification, including sitting, standing, stepping, and cycling, owing to their anatomical placement and compatibility with 24-hour wear protocols [4,10]. In contrast, waist-worn accelerometers such as the ActiGraph monitor are widely used in population-based studies and provide reliable estimates of ambulatory activity and overall activity volume [11]. ActiGraph devices are among the most widely adopted wearable accelerometers for device-based assessment of PB in free-living settings. However, consistent with previous validation studies, their ability to distinguish postures, particularly standing from sitting, is more limited, reflecting the reduced sensitivity of waist-mounted accelerometers to subtle changes in trunk orientation [12-14]. These limitations can result in systematic discrepancies in daily PB and AIC estimates, thereby complicating interpretation and comparability in mHealth apps.

Recent advances in machine learning (ML) and deep learning (DL) methods have substantially improved the accuracy and scalability of human activity recognition using wearables. These methods excel at learning complex temporal and biomechanical patterns that are difficult to capture with traditional rule-based algorithms [15,16]. Among these advances, multitask learning has emerged as a particularly promising framework [17]. By allowing a single model to learn shared representations across related outcomes, such as PB and AIC classification, multitask learning leverages complementary information embedded in accelerometer signals. This joint learning strategy can improve model stability, reduce overfitting, and enhance robustness in free-living environments, where movement patterns are highly variable, context-dependent, and often noisy [18]. Additionally, multihead convolutional neural networks (MH-CNNs) are increasingly used for accelerometer-based movement analysis, extending standard architectures through parallel processing heads that learn complementary representations and support the simultaneous modeling of heterogeneous signal characteristics, such as different temporal scales and movement patterns [19,20].

The limited availability of reliable, ecologically valid, and accurately labeled datasets remains a critical barrier to the development of robust ML models for population-level studies. Many existing approaches rely on data collected using laboratory-based protocols or short-duration recordings, which fail to capture the diversity and complexity of real-world behavior [21,22]. In this context, ecological momentary assessment (EMA) offers a scalable alternative by collecting self-reported behavioral information in real time, reducing recall bias and improving temporal alignment with sensor data [23]. When combined with wearable sensing, EMA enables the generation of labeled free-living datasets that more accurately

XSL•FO
RenderX

reflect naturalistic behavior patterns, supporting the development of models with improved ecological validity [24].

Despite these advances, relatively few studies have combined long-term free-living monitoring, multimodal wearable sensing, and subject-independent validation at scale. Large, labeled datasets such as CAPTURE-24 provide high-quality annotations but are limited in duration [25], while other resources offer contextual information but provide limited temporal resolution or constrained sensor configurations [26]. Moreover, the added value of combining posture-sensitive and movement-sensitive sensors for improving the stability and interpretability of daily PB and AIC estimates remains insufficiently quantified [27].

To address this gap, the wearable sensor assessment of physical and eating behaviors (WEALTH) project integrated standardized laboratory protocols with a 9-day free-living monitoring period across European populations, complemented by event-based (sensor-triggered) EMA delivered via the HealthReact mobile app [28]. These data enabled the generation of sparse (EMA) labeled accelerometer data, further refined using agreement with activPAL's proprietary classification of real-world everyday activities (CREA) algorithm, resulting in a large free-living dataset with quality-controlled labels [23].

Against this background, the main objective of this study was therefore to develop and evaluate multitask ML and DL models for simultaneous PB classification (task 1) and AIC classification (task 2) using thigh- and waist-worn accelerometry, both individually and in a combined (dual-sensor) configuration, under free-living conditions. In support of this objective, multimodal sensing and EMA-informed labels were integrated to facilitate robust behavioral monitoring in ecologically diverse settings. Subsequently, the best-performing models were used to estimate continuous PB and AIC over 9 days. These outputs were aggregated to derive daily time spent in each activity and AIC, enabling the comparison of behavioral patterns across sensor configurations (activPAL, actiGraph, and dual-sensor) and against CREA-derived daily estimates to assess agreement, stability, and ecological coherence. By focusing on ecological validity, scalable data collection, subject-independent validation, and real-world deployment considerations, this work aims to advance methodological foundations for population monitoring in free-living settings, while providing detailed estimates of daily PB and AIC profiles.

Methods

Study Design

This study used data from the WEALTH project, a large, multicenter cohort conducted in Czechia, France, Germany, and Ireland. The study supported the development of standardized methods for assessing daily PA, sedentary behavior, and eating behaviors by integrating structured laboratory assessments with free-living monitoring to ensure ecological validity.

Participants first attended a laboratory session involving a 75-minute semistructured activity protocol, during which well-defined movement patterns were recorded under controlled conditions using 4 different sensors including activPAL 3 micro (PAL Technologies), ActiGraph GT3X (ActiGraph LLC), Skagen Falster Gen 5 smartwatch (Skagen, Fossil Group Inc), and a Fitbit Charge 5 (Fitbit Inc., Google LLC). This was followed by a 9-day free-living phase, during which the 4 different sensor devices collected accelerometer data in combination with event-based EMA to capture contextual information about participants' behaviors in real-world settings [29].

Participants

A cohort of 627 adults was recruited across 4 European study sites using a convenience sampling strategy through university, community, and institutional advertisement channels, resulting in a selective but demographically diverse sample. Sample size was determined according to the WEALTH study protocol [29]. The target sample was selected to support machine-learning model development and to accommodate expected data loss arising from device noncompliance and data quality exclusions. Participants were required to have no mobility impairments and to be able to complete both the laboratory and free-living protocols. From the initial cohort, 32 participants did not provide concurrent activPAL and ActiGraph recordings, and an additional 5 participants were excluded due to significant temporal drift that precluded reliable sensor synchronization. The final sample consisted of 590 participants with valid and concurrent activPAL and ActiGraph data. Participant characteristics are summarized in Table 1 and reflect variability in age, sex, country, and BMI, including representation of individuals classified as obese (BMI ≥30 kg/m²).

XSL•FO
RenderX

Table . Demographic information of the participants included in this study (N=590).

Characteristics	Values
Age (years), mean (SD)	38.1 (14.3)
Height (cm), mean (SD)	171.2 (9.9)
Weight (kg), mean (SD)	72.0 (14.3)
BMI (kg/m²), mean (SD)	24.5 (4)
Obesity (BMI ≥30 kg/m²), n (%)	54 (9.2)
Country, n (%)	
Czechia	138 (23)
France	139 (24)
Germany	157 (27)
Ireland	156 (26)
Sex, n (%)	
Male	258 (44)
Female	332 (56)

Wearable Sensor Configuration and Data Acquisition

A combination of research-grade and consumer-grade wearable devices was used to capture movement patterns, posture, and activity-related events throughout both the laboratory session and the 9-day free-living period. As part of this multimodal setup, an activPAL 3 micro was placed on the midline of the anterior aspect of the right thigh using a hypoallergenic adhesive tegaderm dressing. The device recorded continuous triaxial acceleration at 20 Hz, provided estimates of activity intensity expressed as METs, and behavioral outputs through its proprietary CREA algorithm, classifying sedentary, standing, stepping, cycling, and nonwear, which offered a reference for refining EMA-derived annotations.

To complement the thigh-worn sensor, participants also wore an ActiGraph affixed to the right waist. This device provided higher-frequency (100 Hz) movement data, which contributed to the development of dual-sensor models by enabling an additional perspective on upper-limb and whole-body movement patterns.

Finally, participants wore a Fitbit Charge 5 on the nondominant wrist to trigger event-based EMA surveys. Although Fitbit data were not used as model input, step count and heart rate data were used to detect predefined behavioral conditions and automatically initiate EMA prompts. An overview of the body-worn wearables, data sources, and processing workflow used in this study is provided in Multimedia Appendix 1.

EMA

EMA assessments were implemented using the HealthReact mobile app [28], which delivered three types of surveys throughout the 9-day free-living period: (1) event-based (sensor-triggered) surveys, (2) time-based surveys, and (3) participant-initiated surveys. Only data from event-based surveys were used, as they were automatically triggered in response to specific sensor-detected activities relevant to this study.

Event-based survey prompts were generated automatically based on Fitbit-recorded data, triggering an EMA survey after ≥20 minutes of zero steps for sedentary behavior, ≥5 minutes of walking at 60 - 139 steps/minute, or ≥5 minutes of running at ≥140 steps/minute. Participants reported the activity and posture being performed, while timestamps were automatically logged by the HealthReact app. Prompt allocation rules were pilot-tested following Janek et al [23], with up to 4 daily prompts for sedentary behavior and up to 3 each for walking and running.

Data Synchronization and Preprocessing

The raw accelerometer data were processed using ActiLife (v6.8) for the ActiGraph devices and PAL Analysis (v9.1) for the activPAL devices. For activPAL, the activity annotations and MET-based activity intensity estimates (expressed in MET-seconds) were extracted using PAL Analysis software and the proprietary CREA algorithm. An epoch length of 60 seconds was applied to ensure consistency with prior studies and established protocols for MET-based activity intensity assessment [30,31].

To harmonize the recordings across devices, all acceleration signals were expressed in gravitational units (g). ActiGraph data (100 Hz) were resampled to 20 Hz to match activPAL's native frequency. Initial synchronization was performed using device timestamps and the recorded start of the laboratory session. A fine-grained alignment was subsequently performed and verified by visual inspection using MATLAB (R2020a). Temporal offsets attributable to internal clock drifts were compensated by inserting a small, fixed number of samples (80 samples per day) into the ActiGraph data at a daily interval. This correction accounted for systematic clock drift between activPAL and ActiGraph devices during prolonged free-living recordings, which can produce cumulative timing misalignment between posture and activity signals despite simultaneous initialization [32].

Following device-level synchronization, the CREA-derived MET estimates were temporally aligned with accelerometer

XSL•FO
RenderX

time-series. Activity intensity estimates were converted from MET-seconds to MET values and subsequently mapped to categorical intensity. In this study, AIC were operationalized using MET-derived categories rather than direct kcal-based energy expenditure estimates to enable standardized intensity classification for population-level free-living analyses. This enabled classification into sedentary behavior (1.0 - 1.5 MET), LPA (1.6 - 2.9 MET), and MVPA (≥ 3 MET) categories using standard MET-based cut-points [7,8].

EMA-Based Sparse Labeling and Quality-Controlled Segment Selection

Sparse free-living labeled data were generated following the EMA-based labeling framework proposed in Sigcha et al [24]. During the 9-day data collection period, 26,166 EMA activity surveys were recorded. After removing duplicates and surveys lacking valid posture or activity information, 16,907 surveys were retained for wearable data annotation. EMA triggers were used to segment the continuous accelerometer data streams, and each resulting data segment was labeled using the participant's response to the associated EMA prompt.

The resulting segments were annotated according to predefined PB categories established within the WEALTH study: sitting, standing, walking, running, sports, cycling, and lying, resulting in temporally sparse but highly specific annotations. This EMA-based approach prioritizes ecological validity and label specificity over label density, thereby providing high-quality annotations under free-living conditions.

To strengthen label reliability, only segments showing at least 75% sample-wise agreement with CREA were retained for model training and evaluation. In line with the findings reported in [24], this threshold was selected as an appropriate trade-off between data quality and quantity, reducing less confident segments while preserving sufficient free-living coverage across activity types, ensuring that the final labeled dataset was both ecologically valid and of high quality.

Following EMA-based annotation and quality control using a ≥75% agreement threshold with the CREA algorithm, 2108 hours of labeled free-living data from 583 participants were retained for model development and validation. The reduction in the number of participants from 590 to 583 was primarily due to EMA surveys that could not be synchronized with the corresponding accelerometer recordings because of time-stamp mismatches, which placed the EMA entries outside the valid recording range and prevented reliable segment construction. Detailed distributions of the EMA-labeled data before and after quality control are provided in Multimedia Appendix 2.

Generation of Unlabeled Dual-Sensor Datasets for Model-Derived PB and AIC Estimates

In addition to EMA-labeled data, the 9-day free-living (unlabeled) recordings from activPAL and ActiGraph devices were synchronized to enable the implementation and evaluation of a dual-sensor approach for generating model-derived daily PB and AIC estimates, as well as their corresponding 9-day averages. To ensure strict temporal correspondence between the 2 devices, each pair of recordings was then trimmed to match the sensor file with the lowest sample count and visually inspected using MATLAB.

Following this synchronization procedure, complete unlabeled free-living data were obtained from 590 participants (Table 1), representing the full 9-day monitoring period and providing the basis for subsequent analyses of daily PB and AIC profiles.

Data Segmentation and Feature Extraction

To prepare the data for model development, raw time-series signals were segmented using nonoverlapping windows. Multiple window lengths (5, 10, and 15 s) were evaluated to determine their influence on classification performance, reflecting common practices in human activity recognition [24,25]. For each window, a PB label (task 1) and AIC label (task 2) were assigned using 2 different approaches. The PB label was directly taken from the EMA-reported activity for the corresponding prompt, reflecting the participant's self-reported behavior within that data segment. In contrast, because MET-based estimates were available at a 60-second epoch resolution, a majority voting technique [33] was applied to assign a single AIC label to each window. Specifically, each window was labeled as Sedentary, LPA, or MVPA based on the AIC most frequently observed in the specific window.

Three feature-extraction strategies were implemented to support classical ML pipelines. First, a set of 40 features was computed following the CAPTURE-24 specification [25], capturing statistical, spectral, and temporal characteristics. Second, an adapted version of the HARTH feature set [34] was generated for single-sensor use, extending its applicability beyond the original dual-sensor design. Third, the original HARTH feature set [34] was extracted for dual-sensor configurations to leverage complementary information from thigh- and wrist-based measurements. A detailed description of the set of features extracted in this study is provided in the Multimedia Appendix 3.

In parallel, raw accelerometer data were used to train a multihead convolutional neural network (MH-CNN), implemented in both single- and dual-sensor variants. This DL-based approach enabled the extraction of hierarchical temporal patterns directly from the raw signals, providing an alternative to feature-engineered models.

ML and DL Models

Overview

To evaluate both ML and DL methods, 2 multitask models were developed: a feature-based random forest (RF) model and the MH-CNN. These models were designed to jointly predict PB (task 1) and AIC (task 2) using triaxial accelerometer data.

RF

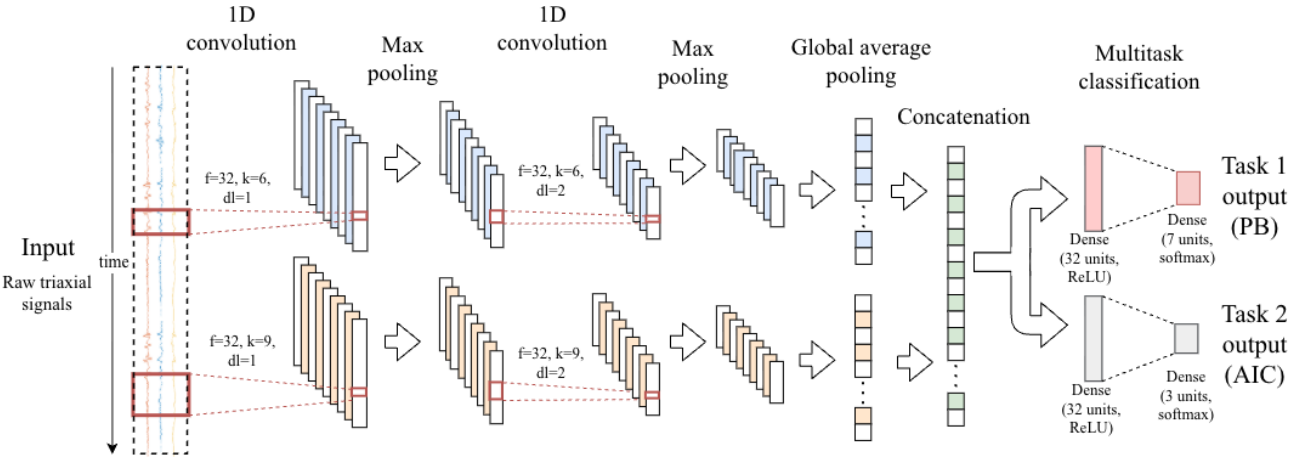
The RF model was selected for its robust performance in PA recognition tasks and its relatively small number of hyperparameters, which facilitates efficient optimization and reduces the risk of overfitting. Separate RF models were trained for each feature set described in the previous section, with each feature set providing a structured representation of the underlying accelerometer signals.

XSL•FO
RenderX

MH-CNN

An adapted MH-CNN [19,20] was implemented to assess the performance of a DL-based approach trained directly on raw accelerometer data. The architecture was selected to balance classification performance, computational efficiency, and scalability for large-scale applications, while reducing the computational burden associated with more demanding sequential architectures. The network architecture comprised 2 parallel convolutional branches with distinct kernel sizes (eg, 9 for a large kernel and 6 for a medium-sized kernel), designed to capture both fine- and coarse-grained temporal dynamics. Each branch included 2 convolutional layers (32 filters each) with max pooling, followed by global average pooling. The extracted features were then passed to separate fully connected (dense) layer branches to support multitask classification. Figure 1 shows the architecture of the MH-CNN with 2 heads, with the output adapted for multitask learning.

Figure 1. Architecture of the multihead convolutional neural network adapted for multitask classification. AIC: activity intensity categories; dl: dilatation rate; f: number of filters; k: kernel size; PB: physical behavior; ReLU: rectified linear unit.



Hyperparameter optimization was performed for both models. For the RF classifier, optimization was carried out using grid search cross-validation, which enabled an exhaustive evaluation of key parameters while keeping computational demands manageable. For the MH-CNN, the Hyperband algorithm [35] was used to efficiently explore a broad space of architectural and training configurations. Details on the parameters and ranges used for the optimization of these models are included in the Multimedia Appendix 4.

To accelerate the optimization process, both procedures were conducted using laboratory data from a reduced subset of 100 participants, as prior work suggested limited performance gains beyond this sample size [25]. Laboratory data were used for model optimization, providing a controlled setting that allowed consistent tuning while avoiding the added variability inherent to free-living data during this stage of model development.

After the optimization, the RF model used for training had 150 decision trees, a maximum tree depth of 10, and a maximum of 16 features considered in each split. The resulting parameters for the MH-CNN are shown in Figure 1 and detailed in the Multimedia Appendix 5.

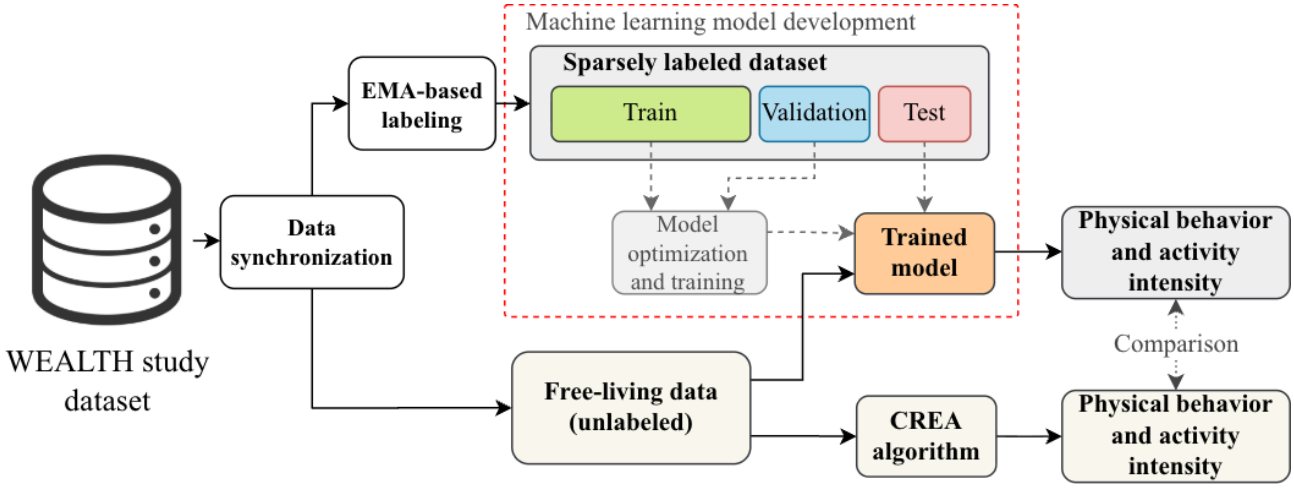
Model Training and Evaluation

To ensure a robust assessment of model performance, a subject-independent hold-out strategy was implemented using the EMA-labeled dataset. Of the original cohort (N=590) participants, data from 583 individuals were available for model training after excluding cases where EMA responses could not be synchronized with accelerometer data. The resulting participants were allocated to training (407/583, 70%), validation (87/583, 15%), and test (87/583, 15%) subsets, with each individual assigned to only 1 data subset to prevent information leakage across subsets. Stratification by age, gender, and country preserved demographic balance across subsets to support a reliable model evaluation. The number of participants and the characteristics of the data subsets used to train and evaluate the ML models are detailed in the Multimedia Appendix 6.

Model performance for both PB and AIC classification was quantified using classification metrics, including precision, recall, and F_1-score [36]. These complementary indicators allowed for a comprehensive evaluation across the imbalanced activity classes. Figure 2 presents the overall methodology for ML model development and agreement analysis.

XSL•FO
RenderX

Figure 2. Overall methodology for machine learning model development and agreement analysis. CREA: classification of real-world everyday activities; EMA: ecological momentary assessment.



ML and DL models were trained on a workstation equipped with an Intel Xeon 2.30 GHz processor, 25 GB RAM, and a 12 GB NVIDIA Tesla K80 GPU. ML models were implemented in Python (v3.12) using Pandas (v2.2), NumPy (2.0.2), and Scikit-learn (v1.6). To address class imbalance, the ML training subset was undersampled so that all activity classes matched the size of the smallest class. Undersampling was chosen because the dataset is highly skewed toward a single category; RF models can be biased by class imbalance; it also reduces training time, mitigates the risk of overfitting, and avoids potential artifacts introduced by synthetic data generation techniques such as the synthetic minority oversampling technique (SMOTE).

DL models were developed using Keras (3.10) and TensorFlow (2.19). Unlike the ML pipeline, undersampling for the DL model was applied selectively by reducing only the most frequent classes (sitting to 30,000 samples; lying to 25,000; walking to 20,000) to preserve the natural distribution of daily activities while controlling class dominance. As with ML, undersampling was performed exclusively on the training data subset.

The MH-CNN architecture was trained using a multitask learning framework with task-specific weighting prioritizing PB over AIC classification (ratio 10:1), together with class-weighted losses and early stopping to reduce overfitting. Optimization was performed using the adaptive moment estimation (ADAM) optimizer [37]. Additional details regarding training configuration and procedures are provided in Multimedia Appendix 5.

ML Experiments

Different experiments were conducted to evaluate the effects of window size, sensor type, and sensor configuration on PB and AIC classification.

The effect of window size was evaluated by segmenting synchronized accelerometer data into nonoverlapping windows of 5, 10, and 15 seconds. These analyses were performed using the dual-sensor configuration to identify an appropriate temporal resolution for subsequent experiments.

Single-sensor experiments were then carried out separately for activPAL and ActiGraph. For each device, RF models trained with hand-crafted features (CAPTURE-24 and an adapted single-sensor HARTH feature set) were compared with MH-CNN trained on raw accelerometer signals. This design allowed for a direct comparison between feature-engineered ML approaches and end-to-end DL models.

Finally, dual-sensor experiments were performed by combining activPAL and ActiGraph data. In this setting, RF models were trained using the original (dual-sensor) HARTH feature set and compared with the MH-CNN trained on synchronized raw data from both sensors.

In all experiments, PB and AIC classification were treated as separate outcomes but evaluated within the same experimental framework. Results of these experiments are reported as mean and SD across 5 repeated runs to identify the model stability during training. In addition, bootstrap resampling (n=100) was used to estimate 95% CIs for the F_1-scores across the evaluated window lengths following the methodology proposed by Efron [38].

Evaluation of Model-Derived Daily PB and AIC Profiles

Overview

Following ML development, the best-performing models were applied to unlabeled (free-living) data to estimate daily time spent in distinct PBs and AIC. The evaluation comprised two complementary analyses: (1) the derivation of model-based daily PB and AIC profiles across different sensor configurations, and (2) an agreement analysis between the best-performing model and the activPAL CREA algorithm. All analyses were conducted using free-living data from 87 participants not included in model training or optimization, ensuring an unbiased assessment of generalizability.

Model-Derived Daily PB and AIC Profiles Across Sensor Configurations

In the first analysis, model-derived daily PB and AIC profiles were generated separately for three sensor configurations: activPAL, ActiGraph, and the combined dual-sensor setup. Valid monitoring days were defined as those with ≥8 hours of wear time [39-41]. Participant-level summaries were calculated

XSL•FO
RenderX

using a weekday-weekend weighting scheme (5/7 weekdays, 2/7 weekend days) [42], and participants were included if they met the 3-day validity criterion (at least 2 valid weekdays and 1 valid weekend day).

For both PB and AIC, weighted participant-level estimates were subsequently summarized across participants to characterize population-level daily profiles. Variability in estimates across participants was quantified using 95% CIs, calculated via a normal approximation based on the sample mean, sample SD, and the number of participants contributing valid data to each estimate.

Comparison of Dual-Sensor Estimates With CREA

In the second analysis, the agreement between the best-performing model (dual-sensor configuration) and the activPAL CREA-derived estimates was evaluated. To ensure compatibility between the 2 systems, CREA activity labels "primary lying" and "secondary lying" were merged into a single lying category, enabling direct comparison of PB categories common to both approaches (eg, standing, walking, cycling, and lying). No additional harmonization was required for AIC, as both methods classify intensity using the same 3 categories (sedentary, LPA, and MVPA).

Unlike the sensor-configuration analysis described above, this comparison was performed at the daily level without applying weekday-weekend weighting to enable a direct, day-by-day comparison with CREA outputs. This approach preserved the original temporal structure of the CREA estimates and avoided introducing additional aggregation effects that could mask day-level agreement between methods.

Agreement was evaluated using correlation plots and Bland-Altman analyses to visually assess the relationship and systematic bias between methods. In addition, Pearson correlation coefficients (r; 2-sided α=.05) were used to quantify the strength of association, and intraclass correlation coefficients (ICC [1,2]) were computed and reported with 95% CIs to evaluate absolute agreement between methods at the daily level.

Ethical Considerations

Prior to the study's commencement, ethics committee approval was obtained from all participating institutions: The Education and Health Sciences Faculty Research Ethics Committee, University of Limerick (22_09_10_EHS); the Ethics Committee of the University of Bremen (2022 - 25); the Comité de Protection des Personnes CPP Île-de-France VI, Paris (2022-A02208-35); and the Committee for Research Ethics at the University of Hradec Králové, Czech Republic (11/2022). All participants provided written informed consent before data collection and were informed of their right to withdraw from the study at any time without providing justification. To acknowledge participants' time and contribution, monetary compensation was provided: €40 in Ireland and €20 in the other countries. All collected data were deidentified for analysis and stored securely to maintain participant confidentiality.

Results

Model Development and Classification Performance

PB and AIC Classification Performance Across Window Lengths

The impact of window size on model performance was evaluated using the dual-sensor configuration by combining the HARTH (dual-sensor) feature set with an RF algorithm. Table 2 shows the results of these experiments.

Table . Physical behavior and activity intensity categories classification performance across window lengths using the dual-sensor random forest model.

Window size (seconds)	Precision	Recall	F_1-score
PB[a], mean (SD)			
5	0.675 (0.005)	0.821 (0.004)	0.729 (0.004)
10	0.671 (0.002)	0.825 (0.002)	0.729 (0.002)
15	0.667 (0.007)	0.828 (0.005)	0.727 (0.006)
AIC[b], mean (SD)			
5	0.736 (0.0004)	0.737 (0.001)	0.728 (0.001)
10	0.743 (0.002)	0.749 (0.001)	0.741 (0.002)
15	0.744 (0.001)	0.754 (0.002)	0.742 (0.001)

[a]PB: physical behavior.

[b]AIC: activity intensity categories.

PB classification performance was comparable across window lengths, with similar F_1-scores for 5- and 10-second windows (both 0.729) and a marginal reduction for 15-second windows (0.727). AIC classification showed slightly higher performance with longer windows, reaching the highest F_1-score at 15 seconds (0.742). However, bootstrap-based CI results showed substantial overlap across window lengths for both PB and AIC outcomes, indicating limited practical differences between configurations (Multimedia Appendix 7). Although longer windows yielded marginally higher AIC performance, differences across window lengths were small; therefore, the 10-second window was selected for subsequent analyses as a compromise between classification performance and temporal resolution.

XSL•FO
RenderX

Single-Sensor Model Performance Using activPAL Data

As shown in Table 3, for PB classification using activPAL data, the MH-CNN achieved the highest F_1-score (0.716), outperforming RF models trained with CAPTURE-24 and the adapted HARTH features. For AIC classification, the RF models showed the best performance, with F_1-scores of 0.736 (CAPTURE-24) and 0.738 (adapted HARTH), whereas the MH-CNN showed reduced performance (0.707).

Table . Physical behavior and activity intensity categories classification performance using activPAL data across machine learning and deep learning models.

ML[a]/DL[b] approach	Precision	Recall	F_1-score
PB[c], mean (SD)			
RF[d] and CAPTURE-24	0.646 (0.003)	0.775 (0.0004)	0.694 (0.002)
RF and HARTH (adapted)	0.639 (0.001)	0.777 (0.001)	0.690 (0.001)
MH-CNN[e]	0.698 (0.016)	0.744 (0.007)	0.716 (0.011)
AIC[f], mean (SD)			
RF and CAPTURE-24	0.740 (0.001)	0.741 (0.001)	0.736 (0.001)
RF and HARTH (adapted)	0.741 (0.001)	0.746 (0.0002)	0.738 (0.001)
MH-CNN	0.726 (0.002)	0.723 (0.005)	0.707 (0.008)

[a]ML: machine learning

[b]DL: deep learning.

[c]PB: physical behavior.

[d]RF: random forest.

[e]MH-CNN: multihead convolutional neural network.

[f]AIC: activity intensity categories.

Single-Sensor Model Performance Using ActiGraph Data

As shown in Table 4, using ActiGraph data alone, PB classification performance was lower than that observed with activPAL. The MH-CNN achieved the highest F_1-score (0.591), slightly outperforming RF models trained with CAPTURE-24 features and performing comparably to the RF model using adapted HARTH features.

Table . Physical behavior and activity intensity categories classification performance using ActiGraph data across machine learning and deep learning models.

ML[a]/DL[b] approach	Precision	Recall	F_1-score
PB[c], mean (SD)			
RF[d] and CAPTURE-24	0.535 (0.004)	0.688 (0.007)	0.553 (0.003)
RF and HARTH (adapted)	0.555 (0.002)	0.725 (0.002)	0.581 (0.002)
MH-CNN[e]	0.580 (0.006)	0.715 (0.008)	0.591 (0.004)
AIC[f], mean (SD)			
RF and CAPTURE-24	0.718 (0.001)	0.734 (0.001)	0.716 (0.001)
RF and HARTH (adapted)	0.721 (0.0003)	0.737 (0.001)	0.721 (0.001)
MH-CNN	0.717 (0.002)	0.718 (0.006)	0.697 (0.006)

[a]ML: machine learning.

[b]DL: deep learning.

[c]PB: physical behavior.

[d]RF: random forest.

[e]MH-CNN: multihead convolutional neural network.

[f]AIC: activity intensity categories.

For AIC classification, RF models outperformed the MH-CNN. The adapted HARTH feature set achieved the highest F_1-score (0.721), marginally above the CAPTURE-24 RF model. The MH-CNN yielded the lowest performance of the 3 approaches (F_1-score=0.697). To obtain comparable results with ActiGraph data, the MH-CNN architecture required an increase in the

XSL•FO
RenderX

number of CNN filters from 32 to 64, while all other hyperparameters were kept unchanged.

Dual-Sensor Model Performance

As shown in Table 5, when combining activPAL and ActiGraph data, performance improved for PB classification over single sensor configurations (Tables 3 and 4). The MH-CNN achieved the highest F_1-score (0.750), outperforming the RF model trained with the dual-sensor HARTH feature set (F_1-score=0.729). For AIC classification, the RF dual-sensor HARTH model achieved the highest and most stable performance (F_1-score=0.741).

Table . Physical behavior and activity intensity categories classification performance using the dual-sensor (activPAL + ActiGraph) configuration.

ML[a]/DL[b] approach	Precision	Recall	F_1-score
PB[c], mean (SD)			
RF[d] and HARTH (dual-sensor)	0.671 (0.002)	0.825 (0.002)	0.729 (0.002)
MH-CNN[e]	0.720 (0.015)	0.797 (0.011)	0.750 (0.008)
AIC[f], mean (SD)			
RF and HARTH (dual-sensor)	0.743 (0.002)	0.749 (0.001)	0.741 (0.002)
MH-CNN	0.734 (0.003)	0.734 (0.005)	0.722 (0.006)

[a]ML: machine learning.

[b]DL: deep learning.

[c]PB: physical behavior.

[d]RF: random forest.

[e]MH-CNN: multihead convolutional neural network.

[f]AIC: activity intensity categories.

Overall, dual-sensor configurations consistently outperformed single-sensor setups for PB classification, while variability across the 5 repeated runs remained low, indicating robust and stable model behavior. Although the RF model achieved the highest AIC classification performance, the overall multitask performance of the RF and MH-CNN models was comparable (mean F_1-score between PB and AIC: 0.736 for MH-CNN and 0.735 for RF). To maintain a unified multitask framework across both outcomes, the MH-CNN was selected for subsequent analyses.

Free-Living Daily PB and AIC Estimates

Model-Derived Daily PB Profiles Across Sensor Configurations

Following model development and results, the selected MH-CNN models were applied to the test subset data (n=87) to generate a (weekday-weekend) weighted average of daily PB estimates across the 9-day free-living monitoring period for the 3 sensor configurations (activPAL, ActiGraph, and dual-sensor).

As shown in Figure 3, sedentary behaviors, particularly sitting and lying, accounted for the highest proportion of daily time in all sensor modalities. The dual-sensor approach estimated 37% (538.4/1440 min) of the day as sitting and 34% (496.5/1440 min) as lying.

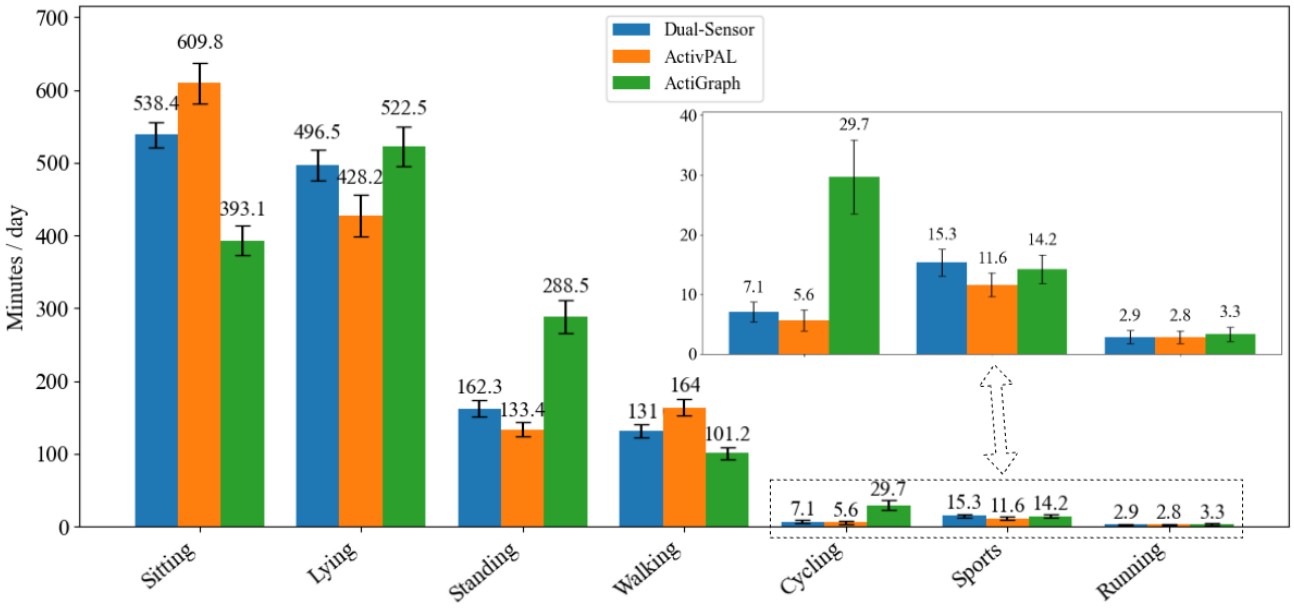
In comparison, our activPAL ML algorithm estimated 42% (609.8/1440 min) of the day as sitting and 30% (428.2/1440 min) as lying, while ActiGraph showed the opposite pattern, with 27% (393.1/1440 min) as sitting and 36% (522.5/1440 min) as lying. For standing time, the dual-sensor configuration estimated 11% (162.3/1440 min) of the day, yielding values intermediate between those obtained with the single-sensor devices. Specifically, activPAL estimated a lower standing time of 9% (133.4/1440 min), whereas ActiGraph produced higher estimates of 20% (288.5/1440 min).

A similar pattern was observed for walking. The dual-sensor model estimated 9% (131/1440 min) of the day, with values falling between those obtained from the single-sensor devices. ActivPAL yielded higher estimates (164/1440 min, 11%), while ActiGraph provided lower estimates (101.2/1440 min, 7%).

PB with low expected temporal prevalence under free-living conditions such as cycling (6–30/1440 min, <2%), sports (12–15/1440 min, ~1%), and running (~3/1440 min, <1%) accounted for only a small proportion of the daily activity profile. Among the single-sensor modalities, ActiGraph yielded higher estimates for these infrequent behaviors, most notably cycling (29.7/1440 min, 2%), whereas the dual-sensor approach produced lower estimates.

XSL•FO
RenderX

Figure 3. Weighted average daily distribution of time spent in physical behavior estimated using activPAL, ActiGraph, and dual-sensor configurations over 9 days. Error bars represent 95% CIs across participants.

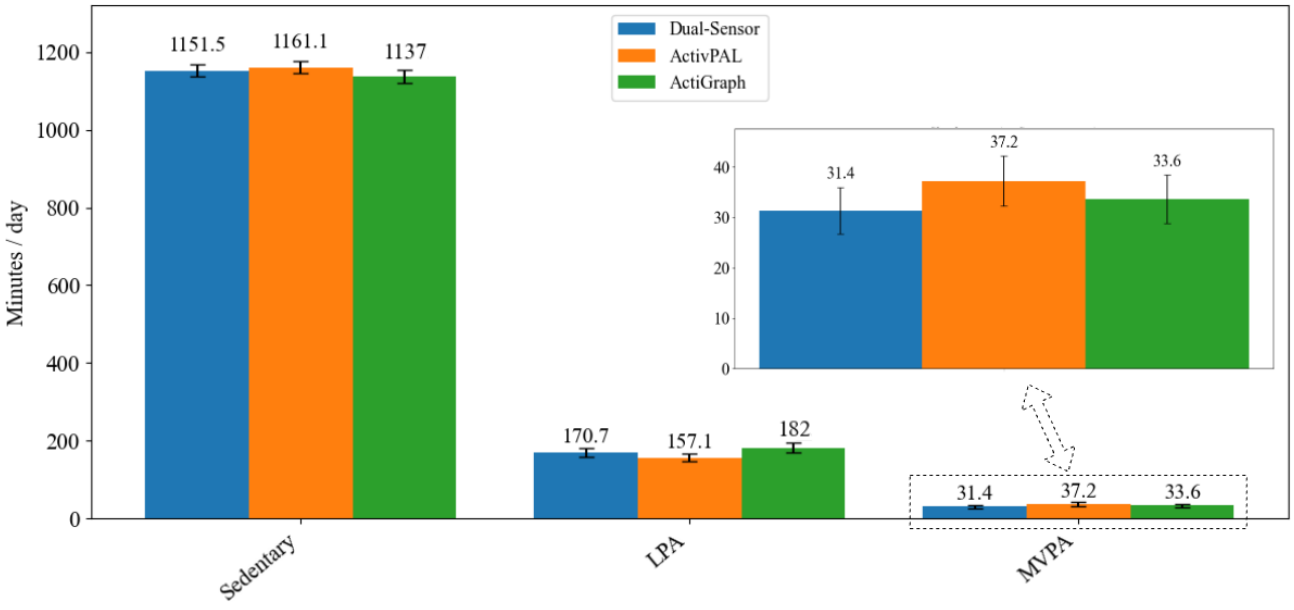


Model-Derived Daily AIC Profiles Across Sensor Configurations

Similarly, the selected MH-CNN models were applied to the test subset data to generate weighted average daily AIC estimates across the 9-day free-living monitoring period.

As shown in Figure 4, estimated weighted average daily AIC exhibited the expected gradient across intensity categories, with sedentary time comprising most of the day across all sensor modalities.

Figure 4. Weighted average daily distribution of time spent in activity intensity categories estimated using activPAL, ActiGraph, and dual-sensor configurations over 9 days. Error bars represent 95% CIs across participants. LPA: light physical activity; MVPA: moderate-to-vigorous physical activity.



The dual-sensor approach estimated 80% (1151.5/1440 min) of the day as sedentary behavior, with comparable values from activPAL (1161.1/1440 min, 81%) and ActiGraph (1137/1440 min, 79%). These higher totals reflect that waking sedentary time and sleep were merged into a single sedentary category.

For LPA, daily estimates ranged from 11% - 13% (157.1 - 182.0/1440 min), with ActiGraph reporting the highest values (182/1440 min, 13%), activPAL the lowest (157.1/1440 min, 11%), and the dual-sensor configuration producing an intermediate estimate of 12% (170.7/1440 min).

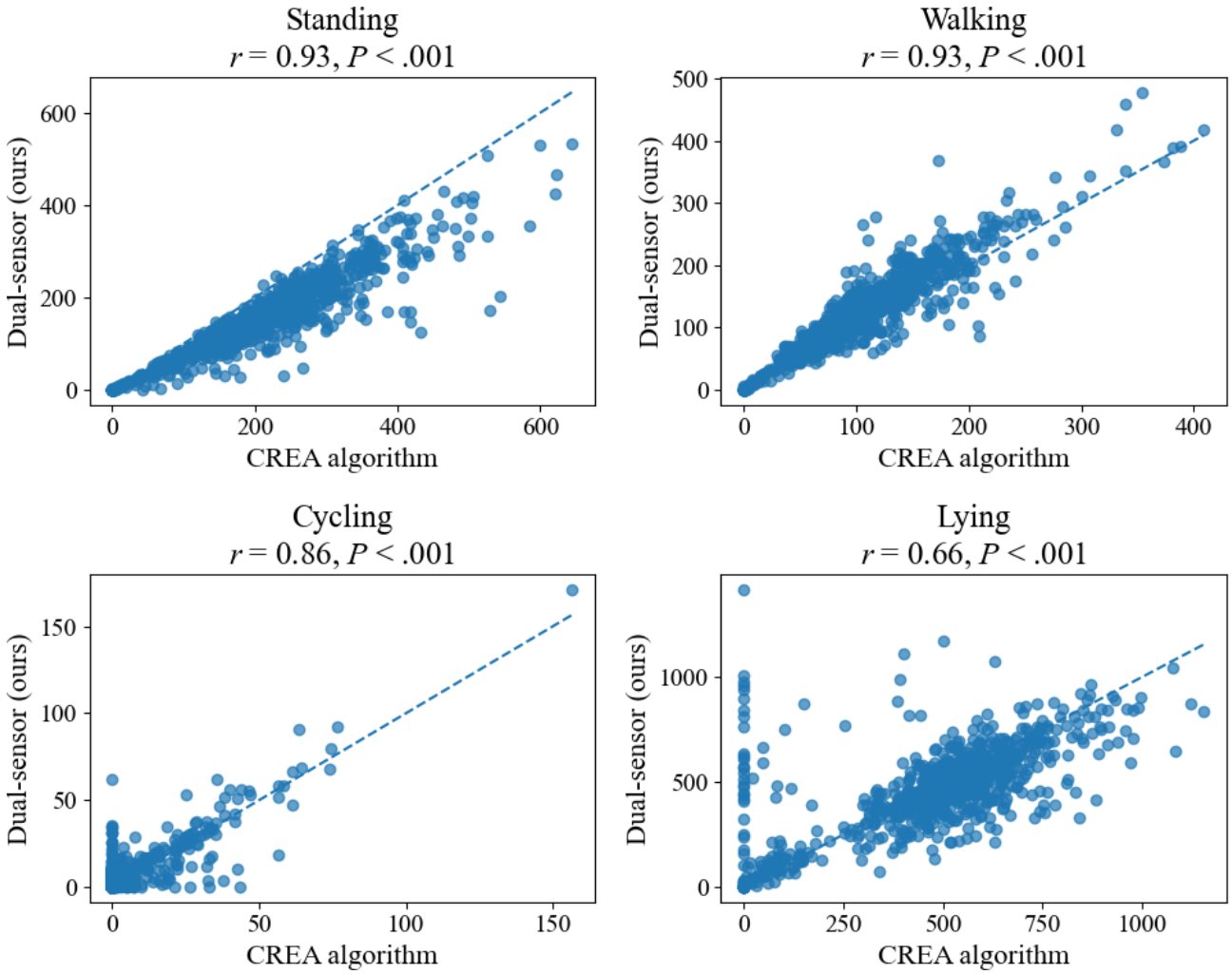
MVPA contributed only a small portion of the daily time. The dual-sensor approach estimated 2% (31.4/1440 min), compared with 3% (37.2/1440 min) for activPAL, and 2% (33.6/1440 min) for ActiGraph.

XSL•FO
RenderX

Agreement Between Dual-Sensor Estimates and CREA for PBs

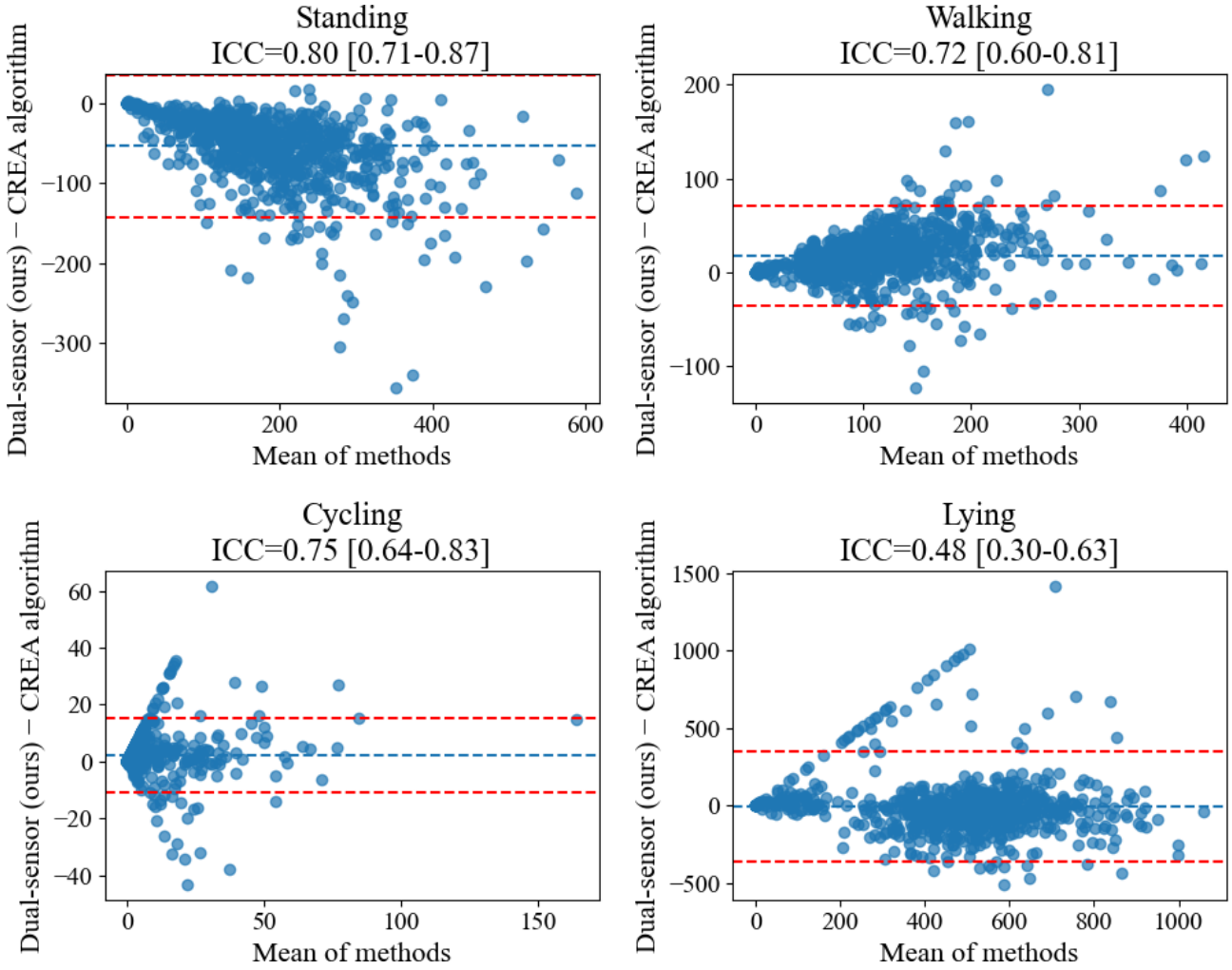
The dual-sensor configuration demonstrated varying levels of agreement with the CREA algorithm across the 4 PB categories, as illustrated in the correlation analyses (Figure 5) and Bland-Altman results (Figure 6). These analyses were based on participant-specific daily estimates, with each data point representing a direct day-by-day comparison between the 2 methods.

Figure 5. Correlations between classification of real-world everyday activities and dual-sensor estimates of daily time spent in physical behavior categories. Each point represents one participant-day, showing the daily duration (minutes/day) estimated by the classification of real-world everyday activities algorithm (x-axis) and the dual-sensor model (y-axis). Pearson coefficients (*r*) and *P* values are shown for each physical behavior. CREA: classification of real-world everyday activities.



XSL•FO
RenderX

Figure 6. Bland-Altman plots showing agreement between classification of real-world everyday activities and dual-sensor estimates of daily time spent in physical behavior categories. Each point represents the difference between dual-sensor and classification of real-world everyday activities daily estimates plotted against their mean at daily level. Intraclass correlation coefficients with 95% CIs are reported for each physical behavior. CREA: classification of real-world everyday activities; ICC: intraclass correlation coefficients.



Standing and walking showed the strongest correspondence, with high correlations (r=0.93, P<.001 for both) and substantial ICC values (standing: ICC=0.80, 95% CI 0.71 - 0.87; walking: ICC=0.72, 95% CI 0.60 - 0.81). Bland-Altman plots indicated relatively narrow limits of agreement and minimal systematic bias.

Cycling also showed a strong correlation (r=0.86, P<.001) and good reliability (ICC=0.75, 95% CI 0.64 - 0.83), although the Bland-Altman plot revealed greater dispersion at longer cycling durations.

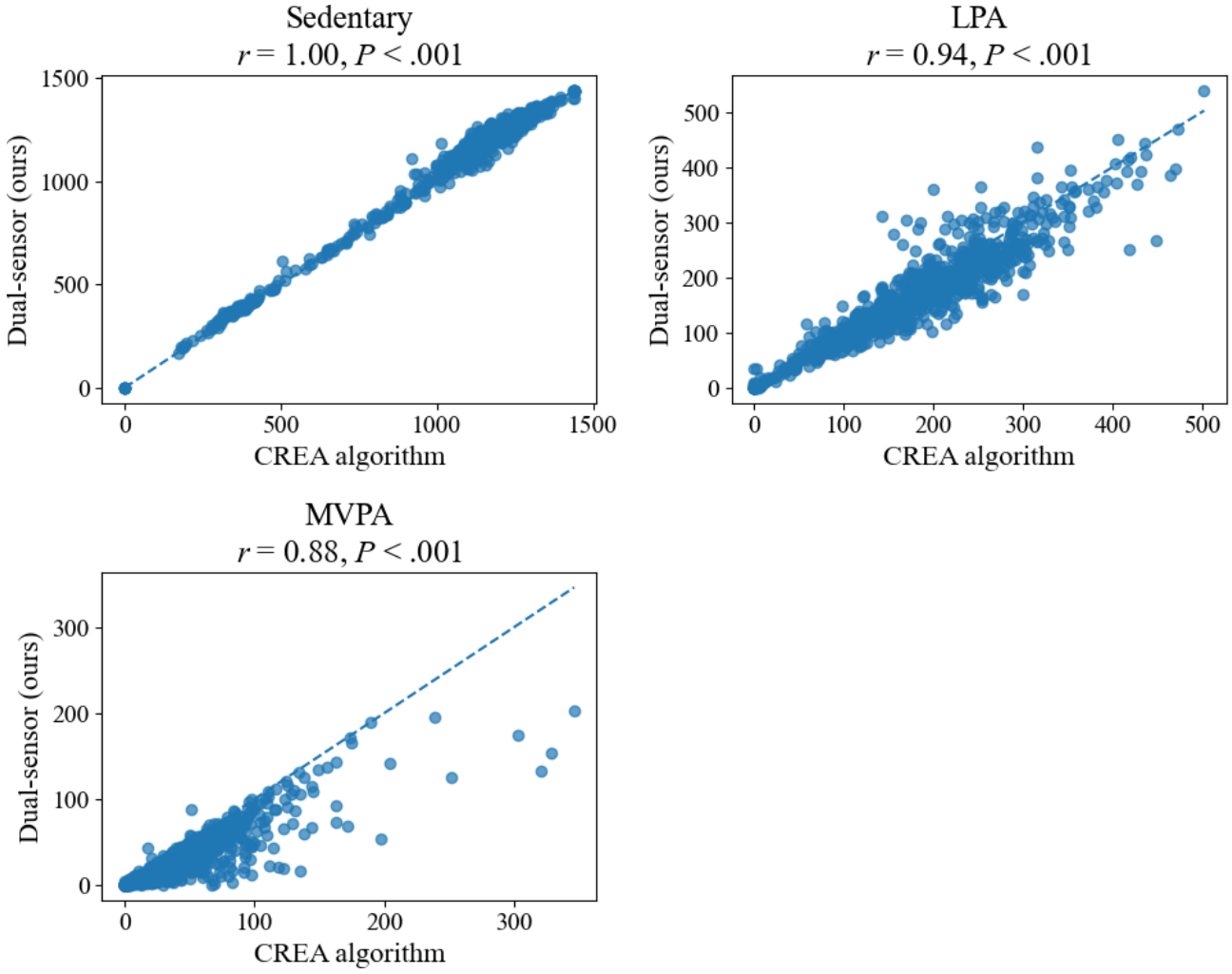
Lying exhibited a similar pattern, with moderate correlation (r=0.66, P<.001) and fair reliability (ICC=0.48, 95% CI 0.30 - 0.63) with increased dispersion at higher daily durations.

Agreement Between Dual-Sensor Estimates and CREA for AIC

The agreement between the dual-sensor configuration and the CREA algorithm varied among AIC, as illustrated in the correlation (Figure 7) and Bland-Altman plots (Figure 8).

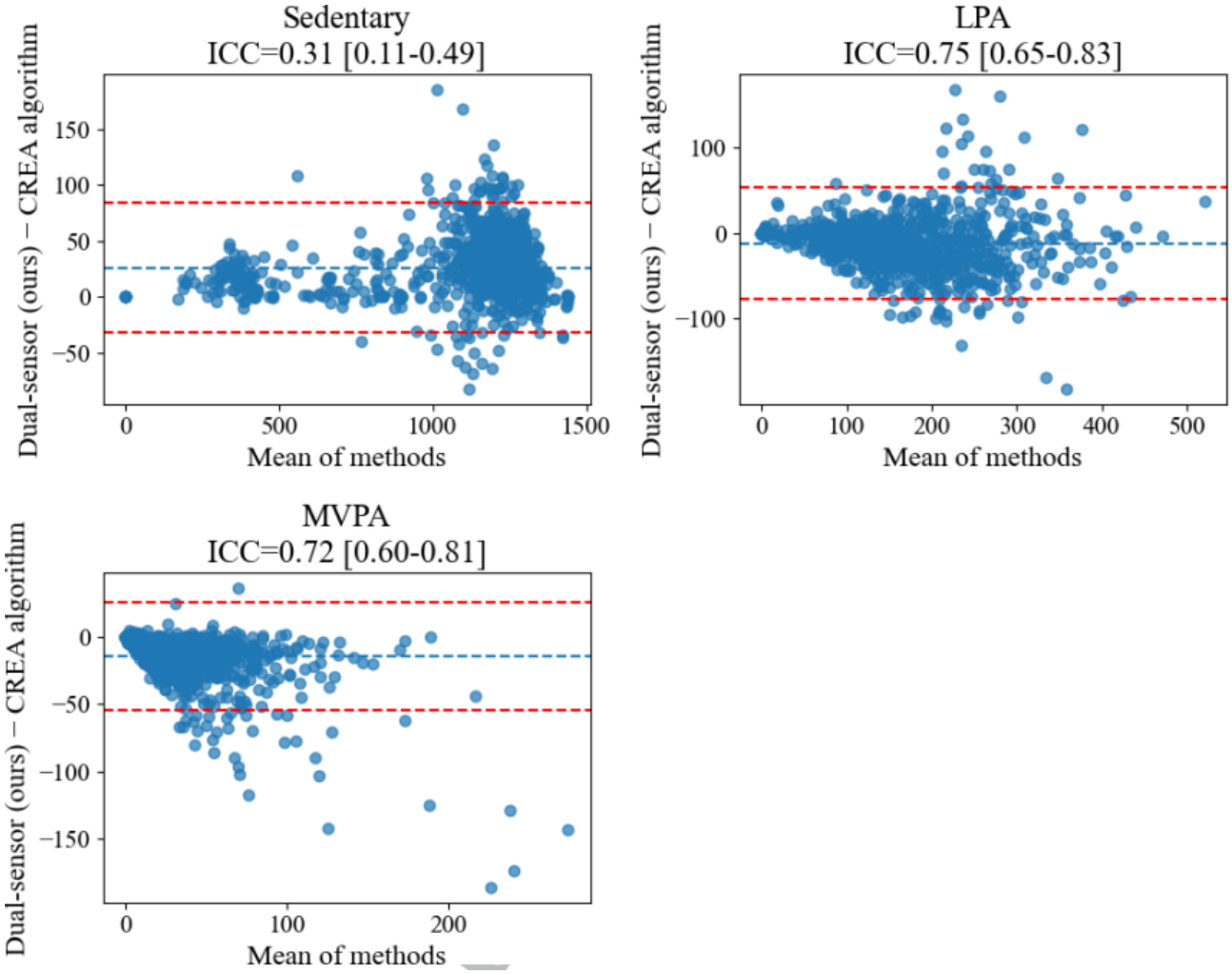
XSL•FO
RenderX

Figure 7. Correlations between classification of real-world everyday activities and dual-sensor estimates of daily time spent in AIC. Each point represents one participant-day, showing the daily duration (minutes/day) estimated by the classification of real-world everyday activities algorithm (x-axis) and the dual-sensor model (y-axis). Pearson coefficients (*r*) and *P* values are shown for each activity intensity category. CREA: classification of real-world everyday activities; LPA: light physical activity; MVPA: moderate-to-vigorous physical activity.



XSL•FO
RenderX

Figure 8. Bland-Altman plots showing agreement between classification of real-world everyday activities and dual-sensor estimates of daily time spent in activity intensity categories. Each point represents the difference between dual-sensor and classification of real-world everyday activities daily estimates plotted against their mean at the daily level. Intraclass correlation coefficients with 95% CIs are reported for each activity intensity category. CREA: classification of real-world everyday activities; ICC: intraclass correlation coefficient; LPA: light physical activity; MVPA: moderate-to-vigorous physical activity.



Sedentary showed a near-perfect correlation (r=1.00, P<.001; Figure 7). However, the Bland-Altman analysis (Figure 8) revealed a wide dispersion at higher sedentary levels, resulting in poor reliability (ICC=0.31, 95% CI 0.11 - 0.49), with increased disagreement above approximately 1000 minutes/day.

For LPA, a strong correlation (r=0.94, P<.001) and good reliability (ICC=0.75, 95% CI 0.65 - 0.83) were observed. MVPA also showed strong agreement (r=0.88, P<.001; ICC=0.72, 95% CI 0.60 - 0.81), with Bland-Altman plots indicating an increasing negative bias at longer MVPA durations.

Discussion

Overview of Key Findings

This study provides a comprehensive development and evaluation of multitask ML models for simultaneous PB and AIC classification using multisensor accelerometry under free-living conditions. Leveraging data from 590 adults across 4 European countries monitored over a 9-day period, the findings demonstrate that integrating thigh- and waist-worn sensors within a multitask learning framework improves model performance and the stability of daily behavioral profiles compared with single-sensor approaches. Collectively, these results support the feasibility of scalable long-term wearable monitoring under real-world conditions, thereby providing ecologically valid evidence at a scale rarely reported in prior free-living wearable studies.

When applied to free-living data, the dual-sensor configuration produced PB distributions characterized by high sedentary time and comparatively limited engagement in ambulatory and vigorous behaviors. These estimates fall within ranges commonly reported in adult populations, where sedentary behaviors account for approximately 60% - 70% of daily time [31,43] and walking typically averages 1 - 2 hours per day [44]. Similarly, low-prevalence behaviors such as running, sports, and cycling contributed minimally to daily activity, consistent with epidemiological observations that vigorous or sport-related activities occupy only a small fraction of daily time in adults [43,45].

For AIC, the dominance of sedentary time and the relative contributions of LPA and MVPA were consistent with population-based accelerometer studies reporting 8 - 10

XSL•FO
RenderX

hours/day of sedentary behavior [46,47], 10% - 20% of time in LPA [48], and 20 - 40 minutes/day of MVPA [39]. The dual-sensor configuration consistently produced intermediate estimates between single-sensor approaches, suggesting reduced sensor-specific bias and improved behavioral characterization across heterogeneous activity patterns. Similarly, the inclusion of participants with a broad range of BMI values, including individuals classified as obese, supported the representation of diverse PB and AIC patterns under free-living conditions.

Agreement analyses further showed that the dual-sensor approach aligned most closely with CREA for ambulatory behaviors (standing, walking, and cycling), while greater discrepancies were observed for prolonged sedentary behaviors, particularly lying. This pattern suggests that multimodal sensing enhances recognition of dynamic behaviors under free-living conditions but remains challenged by extended low-movement postures, a limitation also reported in prior work [49].

Implications for Remote Monitoring

From a monitoring perspective, these findings have important implications for real-world implementation of wearable-based monitoring systems. First, they demonstrated that dual-sensor wearable monitoring could reduce modality-specific biases inherent to single-sensor approaches, particularly for posture-related behaviors. The dual-sensor configuration mitigated systematic over- and underestimation observed with activPAL-only and ActiGraph-only models, producing more balanced and interpretable daily PB profiles that are better suited for longitudinal monitoring and population-level analyses. The scalability of the dual-sensor approach warrants consideration for large-scale deployment. Although the use of 2 devices may increase participant burden, it proved feasible and was well accepted in our sample [50]. Furthermore, jointly labeled thigh- and hip-worn data enable the development of more accurate and generalizable models that can subsequently be applied in single-sensor configurations, supporting scalable mHealth apps [27].

Second, the integration of EMA-informed labeling provided a scalable and pragmatic solution to overcome the scarcity of reliable free-living training data. Event-based EMA enabled the collection of relevant behavior labels at the moment of occurrence, thereby improving ecological validity, supporting the development of models that generalize beyond controlled laboratory environments [23,24,26]. This approach is particularly relevant for mHealth studies requiring extended monitoring periods, where intensive manual annotation is not feasible.

Third, the relative stability of AIC estimates across sensor configurations suggests that multitask learning may be especially valuable for digital health apps focused on energy balance, PA guidelines, and surveillance of activity intensity distributions. Even when discrepancies were observed in PB classification, aggregated AIC estimates remained consistent with epidemiological norms reported in large cohort studies [39,46-48], supporting their use in public health monitoring and evaluation of intervention outcomes.

Comparison With Prior Work

Previous studies on wearable-based activity recognition have frequently relied on laboratory-based protocols, short-term recordings, or single-sensor configurations, which may limit generalizability to free-living settings [22,25,26]. Although datasets such as CAPTURE-24 have advanced free-living activity recognition using video-based behavioral labels [16], reliance on wearable cameras and manual annotation may reduce scalability and raise privacy concerns for longer-term monitoring. In contrast, this study uses 9-day free-living data and EMA-informed labeling from a multicenter cohort, providing a rigorous assessment of model performance under realistic conditions.

Previous research has demonstrated the feasibility of EMA for behavioral annotation and the potential of multitask learning for activity recognition [17,18]. Extending this work, this study combined multitask learning with multisensor accelerometry in a large multicenter dataset, offering empirical evidence of their combined value for mHealth monitoring in real-world settings.

The findings are consistent with previous studies reporting that feature-engineered ML models can remain competitive with DL approaches for free-living activity recognition [21]. In this study, the MH-CNN achieved the highest PB classification performance, whereas the RF model achieved the highest AIC classification performance, suggesting that engineered temporal and statistical features remain informative for AIC classification under free-living conditions [25]. Although more complex sequential architectures may further improve temporal modeling, the selected MH-CNN offered a balanced compromise between predictive performance, complexity, and scalability for long-term monitoring.

Limitations

Several limitations should be acknowledged. AIC estimates were derived from the activPAL CREA algorithm and operationalized using MET-derived categories rather than direct kcal-based energy expenditure, thereby limiting the assessment of continuous energy expenditure. In addition, the CREA algorithm contributed to the label-refinement procedure and was subsequently used as a comparator for daily estimates; therefore, comparisons with CREA-derived estimates represent assessments of agreement rather than independent validation. Although gold-standard methods such as doubly labeled water or other forms of indirect calorimetry would provide stronger absolute validation of the AIC estimates [3], these approaches were beyond the scope of this study and warrant further investigation within multitask wearable monitoring frameworks.

The behavioral taxonomy was limited to 7 PB categories and 3 activity intensity levels, which, although appropriate for large-scale monitoring, do not capture the full diversity of real-world activities. In addition, the EMA-labeled dataset was imbalanced, with low-frequency behaviors such as running and sports remaining underrepresented despite mitigation strategies including selective undersampling and class-weighted losses. Future research may benefit from data augmentation, self-supervised learning, or transfer learning approaches trained

XSL•FO
RenderX

on large unlabeled datasets to improve recognition of infrequent behaviors [15].

Furthermore, although the MH-CNN improved PB classification performance, the comparatively smaller gains observed for AIC classification may reflect class imbalance and limitations in CNN-based temporal modeling under free-living conditions. Future work should evaluate sequential architectures such as LSTM- or Transformer-based models using larger (real-world) datasets. Finally, although the dual-sensor configuration improved performance, practical deployment in some mHealth contexts may require adaptation to single-sensor or consumer-grade devices, which warrants further investigation.

Conclusions

This study demonstrates that multitask ML models combining thigh- and waist-worn accelerometry can provide consistent estimates of PB and AIC under free-living conditions. The dual-sensor approach produced stable, epidemiologically coherent daily profiles and reduced modality-specific biases observed in single-sensor methods. By integrating multimodal wearable sensing with EMA-informed labeling at scale, this work advances methodological foundations for real-world mHealth monitoring and supports more consistent and comparable assessment of PB in population and public health research.

Acknowledgments

We would like to express our sincere appreciation to the volunteers and field staff in Czechia, France, Germany, and Ireland for their invaluable participation. We are deeply grateful to the WEALTH Advisory Panel for their essential guidance and unwavering support throughout this study.

Collaborators: The WEALTH consortium members and affiliations: Alan Donnelly, Catherine Woods, Luis Sigcha, Gráinne Hayes, Pepijn Van de Ven, Daniels Stahovskis (University of Limerick, Ireland), Janas Harrington (School of Public Health, University College Cork, Ireland), Antje Hebestreit, Christoph Buck, Maike Wolters, Annika Swenne, Chandra Gowsiga Loganathan (Leibniz Institute for Prevention Research and Epidemiology—BIPS, Bremen, Germany), Jean-Michel Oppert, Leopold K. Fezeu, Jérôme Bouchan, Fabienne Delestre, Junko Kose (Sorbonne Paris Nord University, France), Hélène Charreire (Inrae, France), Greet Cardon (Department of Movement and Sports Sciences, Ghent University, Belgium), Tomas Vetrovsky, Richard Cimler, Jitka Kuhnova, Alena Faltysova (Faculty of Science, University of Hradec Kralove, Czechia), Steriani Elavsky, Veronika Horká, Michal Sebera (Department of Human Movement Studies, University of Ostrava, Czechia), Michael Janek, Dan Omcirk (Faculty of Physical Education and Sport, Charles University, Prague, Czechia).

OpenAI's ChatGPT was used to refine the language and improve clarity of the manuscript. All content was reviewed and edited by the authors to ensure accuracy and integrity.

Funding

The WEALTH project is funded by the Joint Programming Initiative a Healthy Diet for a Healthy Life, a research and innovation initiative of European Union member states and associated countries under Grant Agreement No 727565, under STAMIFY (Standardized measurement, monitoring and/or biomarkers to study food intake, physical activity, and health). The funding agencies supporting this work are (in alphabetical order of participating countries): Belgium: Research Foundation—Flanders; Czechia: Ministry of Education, Youth, and Sports; France: French National Research Agency; Germany: Federal Ministry of Research, Technology and Space; and Ireland: Health Research Board (HRB).

Data Availability

The data used in this study are not publicly available at this time, as they are subject to an embargo period under the WEALTH project. Access to the data may be granted upon reasonable request and subject to approval by the WEALTH Steering Group.

The trained models, along with example analysis workflows implemented as Python notebooks, are publicly available on the GitHub repository [51].

Conflicts of Interest

LS, AS, GH, CB, GC, SE, LFK, JH, AH, JK, JMO, PVV, TV, and AD declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this study. RC reports a potential conflict of interest. RC is the owner of a company that holds a license to use HealthReact, a software developed by the University of Hradec Králové. However, this did not influence the data analysis, interpretation, or manuscript preparation.

Multimedia Appendix 1
Overview of the body-worn wearables, data sources, and processing workflow.
[DOCX File, 278 KB - mhealth_v14i1e94302_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Characteristics of the sparsely EMA-labeled data before and after applying this data selection.
[DOCX File, 173 KB - mhealth_v14i1e94302_app2.docx]

Multimedia Appendix 3
Description of the set of features extracted.
[DOCX File, 176 KB - mhealth_v14i1e94302_app3.docx]

Multimedia Appendix 4
Details of the hyperparameter tuning.
[DOCX File, 172 KB - mhealth_v14i1e94302_app4.docx]

Multimedia Appendix 5
Parameters of the multihead convolutional neural network.
[DOCX File, 176 KB - mhealth_v14i1e94302_app5.docx]

Multimedia Appendix 6
Characteristics of the data subsets used for training and evaluation.
[DOCX File, 172 KB - mhealth_v14i1e94302_app6.docx]

Multimedia Appendix 7
Bootstrap-based 95% CIs for macro F_1-scores across evaluated window lengths.
[DOCX File, 172 KB - mhealth_v14i1e94302_app7.docx]

References

1. Yang L, Cao C, Kantor ED, et al. Trends in sedentary behavior among the US population, 2001-2016. JAMA 2019 Apr 23;321(16):1587-1597. [doi: 10.1001/jama.2019.3636] [Medline: 31012934]
2. Strain T, Flaxman S, Guthold R, et al. National, regional, and global trends in insufficient physical activity among adults from 2000 to 2022: a pooled analysis of 507 population-based surveys with 5·7 million participants. Lancet Glob Health 2024 Aug;12(8):e1232-e1243. [doi: 10.1016/S2214-109X(24)00150-5] [Medline: 38942042]
3. Dowd KP, Szeklicki R, Minetto MA, et al. A systematic literature review of reviews on techniques for physical activity measurement in adults: a DEDIPAC study. Int J Behav Nutr Phys Act 2018 Feb 8;15(1):15. [doi: 10.1186/s12966-017-0636-2] [Medline: 29422051]
4. Stevens ML, Gupta N, Inan Eroglu E, et al. Thigh-worn accelerometry for measuring movement and posture across the 24-hour cycle: a scoping review and expert statement. BMJ Open Sport Exerc Med 2020;6(1):e000874. [doi: 10.1136/bmjsem-2020-000874] [Medline: 33408875]
5. Lam YY, Redman LM, Smith SR, et al. Determinants of sedentary 24-h energy expenditure: equations for energy prescription and adjustment in a respiratory chamber. Am J Clin Nutr 2014 Apr;99(4):834-842. [doi: 10.3945/ajcn.113.079566] [Medline: 24500151]
6. Marra M, Sammarco R, Cioffi I, et al. New predictive equations for estimating resting energy expenditure in subjects with normal weight and overweight. Nutrition 2021 Apr;84:111105. [doi: 10.1016/j.nut.2020.111105] [Medline: 33477001]
7. Ainsworth BE, Haskell WL, Herrmann SD, et al. 2011 compendium of physical activities: a second update of codes and MET values. Med Sci Sports Exerc 2011 Aug;43(8):1575-1581. [doi: 10.1249/MSS.0b013e31821ece12] [Medline: 21681120]
8. Tremblay MS, Aubert S, Barnes JD, et al. Sedentary behavior research network (SBRN) - terminology consensus project process and outcome. Int J Behav Nutr Phys Act 2017 Jun 10;14(1):75. [doi: 10.1186/s12966-017-0525-8] [Medline: 28599680]
9. Dunn J, Runge R, Snyder M. Wearables and the medical revolution. Per Med 2018 Sep;15(5):429-448. [doi: 10.2217/pme-2018-0044] [Medline: 30259801]
10. Bourke AK, Ihlen EAF, Helbostad JL. Validation of the activPAL3 in free-living and laboratory scenarios for the measurement of physical activity, stepping, and transitions in older adults. J Meas Phys Behav 2019 Jun;2(2):58-65. [doi: 10.1123/jmpb.2018-0056]
11. Wu WJ, Yu HB, Tai WH, Zhang R, Hao WY. Validity of actigraph for measuring energy expenditure in healthy adults: a systematic review and meta-analysis. Sensors (Basel) 2023 Oct 18;23(20):8545. [doi: 10.3390/s23208545] [Medline: 37896640]
12. Lyden K, Keadle SK, Staudenmayer J, Freedson PS. A method to estimate free-living active and sedentary behavior from an accelerometer. Med Sci Sports Exerc 2014 Feb;46(2):386-397. [doi: 10.1249/MSS.0b013e3182a42a2d] [Medline: 23860415]

XSL•FO
RenderX

13. Edwardson CL, Rowlands AV, Bunnewell S, et al. Accuracy of posture allocation algorithms for thigh- and waist-worn accelerometers. Med Sci Sports Exerc 2016 Jun;48(6):1085-1090. [doi: 10.1249/MSS.0000000000000865] [Medline: 26741122]
14. Swenne A, Sigcha L, Hebestreit A, et al. External validation of a hip-worn accelerometry-based machine learning model for physical behavior classification in free-living conditions. Journal for the Measurement of Physical Behaviour 2025;8(1). [doi: 10.1123/jmpb.2025-0030]
15. Gupta N, Gupta SK, Pathak RK, Jain V, Rashidi P, Suri JS. Human activity recognition in artificial intelligence framework: a narrative review. Artif Intell Rev 2022;55(6):4755-4808. [doi: 10.1007/s10462-021-10116-x] [Medline: 35068651]
16. Wang J, Chen Y, Hao S, Peng X, Hu L. Deep learning for sensor-based activity recognition: a survey. Pattern Recognit Lett 2019 Mar;119:3-11. [doi: 10.1016/j.patrec.2018.02.010]
17. Zhang Y, Yang Q. A survey on multi-task learning. IEEE Trans Knowl Data Eng 2021;34(12):5586-5609. [doi: 10.1109/TKDE.2021.3070203]
18. Yuan H, Chan S, Creagh AP, et al. Self-supervised learning for human activity recognition using 700,000 person-days of wearable data. NPJ Digit Med 2024 Apr 12;7(1):91. [doi: 10.1038/s41746-024-01062-3] [Medline: 38609437]
19. Sigcha L, Borzì L, Olmo G. Deep learning algorithms for detecting freezing of gait in Parkinson's disease: a cross-dataset study. Expert Syst Appl 2024 Dec;255:124522. [doi: 10.1016/j.eswa.2024.124522]
20. Borzì L, Sigcha L, Rodríguez-Martín D, Olmo G. Real-time detection of freezing of gait in Parkinson's disease using multi-head convolutional neural networks and a single inertial sensor. Artif Intell Med 2023 Jan;135:102459. [doi: 10.1016/j.artmed.2022.102459] [Medline: 36628783]
21. Awais M, Chiari L, Ihlen EAF, Helbostad JL, Palmerini L. Classical machine learning versus deep learning for the older adults free-living activity classification. Sensors (Basel) 2021 Jul 7;21(14):4669. [doi: 10.3390/s21144669] [Medline: 34300409]
22. Allahbakhshi H, Hinrichs T, Huang H, Weibel R. The key factors in physical activity type detection using real-life data: a systematic review. Front Physiol 2019;10:75. [doi: 10.3389/fphys.2019.00075] [Medline: 30809152]
23. Janek M, Kuhnova J, Cardon G, et al. Ecological momentary assessment of physical and eating behaviours: the WEALTH feasibility and optimisation study with recommendations for large-scale data collection. PLOS ONE 2025;20(2):e0318772. [doi: 10.1371/journal.pone.0318772] [Medline: 39932907]
24. Sigcha L, Swenne A, Kuhnova J, et al. Data labelling for free-living physical activity recognition using thigh-worn wearables and event-based ecological momentary assessment (in review). . Preprint posted online on Aug 5, 2025. [doi: 10.21203/rs.3.rs-6835979/v1]
25. Chan S, Hang Y, Tong C, et al. CAPTURE-24: a large dataset of wrist-worn activity tracker data collected in the wild for human activity recognition. Sci Data 2024 Oct 16;11(1):1135. [doi: 10.1038/s41597-024-03960-3] [Medline: 39414802]
26. Vaizman Y, Ellis K, Lanckriet G. Recognizing detailed human context in the wild from smartphones and smartwatches. IEEE Pervasive Comput 2017;16(4):62-74. [doi: 10.1109/MPRV.2017.3971131]
27. Ahmadi MN, Koemel N, Biswas R, et al. Development and evaluation of wrist- and thigh-worn accelerometer algorithms using self-training machine learning models for classification of activity type and posture: towards device placement-agnostic methods in the propass consortium. medRxiv. Preprint posted online on Jun 30, 2025. [doi: 10.1101/2025.06.26.25330373] [Medline: 25330373]
28. Novak J, Jurkova K, Lojkaskova A, et al. Participatory development of an mHealth intervention delivered in general practice to increase physical activity and reduce sedentary behaviour of patients with prediabetes and type 2 diabetes (ENERGISED). BMC Public Health 2024 Mar 31;24(1):927. [doi: 10.1186/s12889-024-18384-2] [Medline: 38556892]
29. Hayes G, Buck C, Cardon G, et al. Standardized methods for evaluating physical and eating behaviors: the wealth cross-sectional study protocol. JMIR Res Protoc 2026 Mar 6;15:e70186. [doi: 10.2196/70186] [Medline: 41813445]
30. Wullems JA, Verschueren SMP, Degens H, Morse CI, Onambélé-Pearson GL. Concurrent validity of four activity monitors in older adults. Sensors (Basel) 2024 Jan 30;24(3):895. [doi: 10.3390/s24030895] [Medline: 38339613]
31. Koster A, Shiroma EJ, Caserotti P, et al. Comparison of sedentary estimates between activPAL and hip- and wrist-worn actigraph. Med Sci Sports Exerc 2016 Aug;48(8):1514-1522. [doi: 10.1249/MSS.0000000000000924] [Medline: 27031744]
32. Steel C, Bejarano C, Carlson JA. Time drift considerations when using GPS and accelerometers. J Meas Phys Behav 2019;2(3):203-207. [doi: 10.1123/jmpb.2019-0004]
33. Yao R, Lin G, Shi Q, Ranasinghe DC. Efficient dense labelling of human activity sequences from wearables using fully convolutional networks. Pattern Recognit 2018 Jun;78:252-266. [doi: 10.1016/j.patcog.2017.12.024]
34. Logacjov A, Bach K, Kongsvold A, Bårdstu HB, Mork PJ. HARTH: a human activity recognition dataset for machine learning. Sensors (Basel) 2021 Nov 25;21(23):7853. [doi: 10.3390/s21237853] [Medline: 34883863]
35. Li L, Jamieson K, DeSalvo G, Rostamizadeh A, Talwalkar A. Hyperband: a novel bandit-based approach to hyperparameter optimization. Arxiv. Preprint posted online on Mar 21, 2016. [doi: 10.48550/arxiv.1603.06560]
36. Japkowicz N, Shah M. Evaluating Learning Algorithms: A Classification Perspective: Cambridge University Press; 2011.
37. Kingma DP, Ba LJ. Adam: a method for stochastic optimization. Arxiv. Preprint posted online on Dec 22, 2014. [doi: 10.48550/arXiv.1412.6980]
38. Efron B. The Jackknife, the Bootstrap and Other Resampling Plans: Society for Industrial and Applied Mathematics; 1982.

XSL•FO
RenderX

39. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
40. Aadland E, Ylvisåker E. Reliability of objectively measured sedentary time and physical activity in adults. PLOS ONE 2015;10(7):e0133296. [doi: 10.1371/journal.pone.0133296] [Medline: 26192184]
41. Rowlands AV, Yates T, Olds TS, Davies M, Khunti K, Edwardson CL. Wrist-worn accelerometer-brand independent posture classification. Med Sci Sports Exerc 2016 Apr;48(4):748-754. [doi: 10.1249/MSS.0000000000000813] [Medline: 26559451]
42. Namba H, Yamada Y, Ishida M, Takase H, Kimura M. Use of a web-based physical activity record system to analyze behavior in a large population: cross-sectional study. J Med Internet Res 2015 Mar 19;17(3):e74. [doi: 10.2196/jmir.3923] [Medline: 25794109]
43. Loyen A, Van Hecke L, Verloigne M, et al. Variation in population levels of physical activity in European adults according to cross-European studies: a systematic literature review within DEDIPAC. Int J Behav Nutr Phys Act 2016 Jun 28;13:72. [doi: 10.1186/s12966-016-0398-2] [Medline: 27350359]
44. Tudor-Locke C, Craig CL, Brown WJ, et al. How many steps/day are enough? For adults. Int J Behav Nutr Phys Act 2011 Jul 28;8(1):79. [doi: 10.1186/1479-5868-8-79] [Medline: 21798015]
45. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob Health 2018 Oct;6(10):e1077-e1086. [doi: 10.1016/S2214-109X(18)30357-7] [Medline: 30193830]
46. Buman MP, Winkler EAH, Kurka JM, et al. Reallocating time to sleep, sedentary behaviors, or active behaviors: associations with cardiovascular disease risk biomarkers, NHANES 2005-2006. Am J Epidemiol 2014 Feb 1;179(3):323-334. [doi: 10.1093/aje/kwt292] [Medline: 24318278]
47. Matthews CE, Chen KY, Freedson PS, et al. Amount of time spent in sedentary behaviors in the United States, 2003-2004. Am J Epidemiol 2008 Apr 1;167(7):875-881. [doi: 10.1093/aje/kwm390] [Medline: 18303006]
48. Matthews CE, Keadle SK, Troiano RP, et al. Accelerometer-measured dose-response for physical activity, sedentary time, and mortality in US adults. Am J Clin Nutr 2016 Nov;104(5):1424-1432. [doi: 10.3945/ajcn.116.135129] [Medline: 27707702]
49. Ustad A, Logacjov A, Trollebø S, et al. Validation of an activity type recognition model classifying daily physical behavior in older adults: the HAR70+ model. Sensors (Basel) 2023 Feb 21;23(5):2368. [doi: 10.3390/s23052368] [Medline: 36904574]
50. Bouchan J, Fezeu LK, Vetrovsky T, et al. Acceptability of intensive ambulatory assessment combining four wearable sensors and ecological momentary assessment to assess physical and eating behaviors in adults: the WEALTH project. Journal for the Measurement of Physical Behaviour 2026;9(1). [doi: 10.1123/jmpb.2025-0025]
51. Wealth_pb_ee_models. GitHub. URL: https://github.com/luisfrancs/wealth_pb_ee_models [accessed 2026-02-09]

Abbreviations

AIC: activity intensity categories
CREA: classification of real-world everyday activities (activPAL algorithm)
DL: deep learning
EMA: ecological momentary assessment
ICC: intraclass correlation coefficient
LPA: light physical activity
MET: metabolic equivalent of task
MH-CNN: multihead convolutional neural network
mHealth: mobile health
ML: machine learning
MVPA: moderate-to-vigorous physical activity
PA: physical activity
PB: physical behavior
RF: random forest
SMOTE: synthetic minority oversampling technique
WEALTH: wearable sensor assessment of physical and eating behaviors

XSL•FO
RenderX

Edited by J Zhou; submitted 27.Feb.2026; peer-reviewed by R Rising, T Panboonyuen; revised version received 14.Jun.2026; accepted 14.Jun.2026; published 14.Aug.2026.

Please cite as:
Sigcha L, Swenne A, Hayes G, Kuhnova J, Cimler R, Elavsky S, Vetrovsky T, Fezeu Kamedjie L, Bouchan J, Oppert JM, Harrington J, Cardon G, Hebestreit A, Donnelly A, Van de Ven P, Buck C, WEALTH consortium
Robust Assessment of Free-Living Physical Behaviors and Activity Intensity Using Dual-Wearable Multitask Learning: Development and Evaluation Study From the Multicenter WEALTH Project
JMIR Mhealth Uhealth 2026;14:e94302
URL: https://mhealth.jmir.org/2026/1/e94302
doi:10.2196/94302

© Luis Sigcha, Annika Swenne, Grainne Hayes, Jitka Kuhnova, Richard Cimler, Steriani Elavsky, Tomas Vetrovsky, Léopold Fezeu Kamedjie, Jérôme Bouchan, Jean-Michel Oppert, Janas Harrington, Greet Cardon, Antje Hebestreit, Alan Donnelly, Pepijn Van de Ven, Christoph Buck, WEALTH consortium. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

An Information-Driven Approach for the Early Health Technology Sustainability Assessment and the Frugal Design of the Internet of Medical Things: Exploratory Study of Wearable Activity Monitoring Devices

Ernesto Quisbert-Trujillo[1], PhD; Nicolas Vuillerme[1,2], Prof Dr

[1]Centre de Recherche en Santé Intégrée - CReSI, Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, Sangria, 6 chemin Saint Ferjus, Grenoble, France
[2]Institut Universitaire de France, 75231 Paris, France

Corresponding Author:
Nicolas Vuillerme, Prof Dr
Centre de Recherche en Santé Intégrée - CReSI, Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, Sangria, 6 chemin Saint Ferjus, Grenoble, France

Abstract

Background: Over the past decade, wearable activity monitoring (WAM) devices have become prevalent to improve quality of life and support the prevention and management of a wide range of health disorders. WAM devices based on the Internet of Medical Things (IoMT) paradigm provide a practical means of tracking physical activity, but their widespread adoption raises sustainability concerns. Meanwhile, health technology assessment and life cycle assessment are typically applied at advanced development stages, when empirical certainty about the final design and operating conditions is available, leaving little room for further improvements.

Objective: This exploratory work provides the empirical foundations for an information-driven approach addressing this paradox in the early evaluation and conception of wrist-worn step counters, for which evidence suggests overdimensioned and unsustainable electronic designs. Specifically, we identify optimal resource-performance trade-offs in critical electronic components of frugal smartbands based on the data they collect and the information they preserve for step counting. This approach accounts for uncertainties in device reliability, users' gait speeds, and material and energy consumption in final products.

Methods: We conducted a secondary analysis of an accelerometer dataset characterizing wrist motion in healthy individuals walking at different speeds. The original x-, y-, and z-axis signals were progressively downsampled using cubic spline interpolation and discretized to quantify the information preserved, first across sampling frequencies and then with respect to changes in motion velocity. We also preliminarily assessed the viability of the downsampled signals for step detection by estimating percentage errors in peak and valley counts. Based on this, we constructed and evaluated 4 frugal smartband design archetypes, linking energy consumption and sampling frequency for 4 widely used accelerometers and combining essential electronic components, implementing a suboptimal asynchronous first-in-first-out (FIFO) algorithm at different sampling rates. Finally, we evaluated the environmental impact and circularity of these components through a streamlined analysis focused on raw materials.

Results: About 70% - 90% of the information is lost in signals downsampled at very low frequencies (2 - 5 Hz), whereas losses remain below 24% from 20 Hz onward. Substantial information loss also occurs when individuals walk briskly or jog (≥8 km/h), or walk below 8 km/h with sampling frequencies below 7 Hz or above 25 Hz. Step-counting accuracy is expected to be acceptable from approximately 11 Hz onward. Conversely, higher sampling rates rapidly saturate FIFO buffers and increase energy overhead, particularly when implemented in memory-dense components handling both processing and data transfer. Finally, gold and silver in transceivers and microcontrollers contribute substantially to resource depletion, while copper remains relevant for material recovery.

Conclusions: These findings provide preliminary insights into the concurrent assessment and development of frugal WAM devices. This work extends the understanding of step detection under knowledge-constrained conditions and provides mechanisms to reduce uncertainty during early health technology assessment and eco-design.

(JMIR Mhealth Uhealth 2026;14:e88237) doi:10.2196/88237

KEYWORDS

health technology assessment; environmental sustainability; eco-design; frugality; Internet of Medical Things; wearables; step counting; physical activity; walking; information theory

XSL•FO
RenderX

Introduction

Wearable activity monitoring (WAM) devices, such as smartwatches, fitness bands, and pedometers, are widely endorsed for promoting physical activity (PA) [1,2] and well-being [3]. Consequently, the WAM market is projected to grow significantly, from reported sales of US $71 billion in 2023 to more than US $186 billion by 2030 [4]. This rapid market expansion raises critical medium-term sustainability concerns. Under current trends, WAM devices are estimated to generate 3.4 Mt CO_2-equivalent emissions by 2050, while also exacerbating challenges related to ecotoxicity, resource depletion, and e-waste management [5].

Yet, evidence also shows that the environmental impacts of wearables vary considerably. For instance, official sustainability reports for Samsung smartwatches reveal that resource depletion differs markedly across product families and models [6]. Moreover, the production of a midtier smartwatch accounts for approximately 5 kg CO_2-eq of global warming potential [7], whereas this impact can increase by as much as 6-fold for a high-end model [8].

Different product families and premium devices undoubtedly offer a range of benefits. However, regarding PA monitoring performance, several works suggest that low-complexity devices are generally sufficient for approximate movement tracking [9,10], which appears adequate to encourage and sustain PA over the long term [11]. Furthermore, clinical studies have shown that accelerometer data collected over only a few days are sufficient to provide a reliable daily step count in people with neurodegenerative diseases [12]. Likewise, specific research has demonstrated that low data resolution does not significantly impair the detection of complex gait-related motor symptoms [13]. Taken together, this evidence suggests that limited resources and information can still be used effectively to monitor simple ambulation patterns.

From the perspective of national health authorities responsible for assessing and approving WAM devices, this insight is particularly important because a primary public health objective is to reduce disease morbidity through resource-efficient strategies that promote long-term PA. While health technology assessment (HTA) [14] and life cycle assessment (LCA) [15] are widely accepted methodologies for supporting this objective, their analytical reach is severely constrained during early product development due to limited empirical evidence [16]. Paradoxically, this stage also offers considerable freedom to make critical decisions that can positively influence not only the sustainability of final devices [17] but also their efficiency and ethical characteristics [18-20]. As a result, 2 related challenges emerge: an “eco-design paradox” [21,22], in which opportunities for environmental improvement decrease as product knowledge increases [23], and an early HTA paradox, in which the imbalance between uncertainty and evidence complicates decision-making in health care technology design and development [24].

In this exploratory work, we address these challenges by building on evidence suggesting (1) that low-complexity devices can effectively support approximate PA tracking [9-11] and (2) that limited information may still be sufficient to monitor simple ambulation patterns [12,13]. Specifically, this study investigates how the degradation of intrinsic information contained in accelerometer data influences the acquisition and use of knowledge and the management of uncertainty during the early-stage design and preliminary sustainability assessment of frugal WAM devices. To answer this research question, we analyze how sensor data availability affects the development and environmental evaluation of lightweight prototypes. In particular, we quantify the effects of incremental information loss on early design and on the energy consumption of basic algorithms operating under different motion velocities.

The aim of this exploratory research is therefore to examine the potential of a novel information-driven approach for addressing the difficulties and paradoxes associated with the early assessment and design of frugal WAM devices. In doing so, we contribute to the state of the art by providing knowledge relevant to the early design and preliminary environmental assessment of frugal archetypes, together with empirical evidence regarding the performance of final devices operating under specific and uncertain conditions. The remainder of this paper is organized as follows. The Methods section details our research approach applied to the specific case of smartbands, followed by the Results section presenting our preliminary findings. The Discussion section contextualizes the novelty and contributions of this work, addresses its limitations, and concludes with directions for future research.

Methods

Methodology Overview

Our research methodology consists of 3 successive steps: a secondary analysis of an existing accelerometer dataset, the construction of design archetypes for frugal smartbands, and a streamlined assessment of their environmental impact.

The first step is grounded in the principles of information theory and follows a data-information-knowledge process adapted for the ecological design and evaluation of Internet of Medical Things (IoMT) systems [25]. Its objective is to quantify how much information is preserved in signals describing the wrist movements of individuals walking at different speeds as the initial sampling rate is progressively reduced across the x-, y-, and z-axes. By doing so, we generate and use the knowledge required to manage critical uncertainties encountered during the early design and performance evaluation of archetypes, particularly those related to the movement speed of users and the energy consumption of final devices. We focus on energy consumption because it represents a major source of uncertainty in wearable design, as it is influenced by numerous stochastic factors, including usage context [26] and data acquisition, processing, and communication conditions [27].

The second step focuses on constructing electronic design archetypes for frugal smartbands based on the primary functions and capacities of their essential components, following the approach proposed by Quisbert-Trujillo and Morfouli [28]. Its objective is to support a worst-case analysis that addresses the

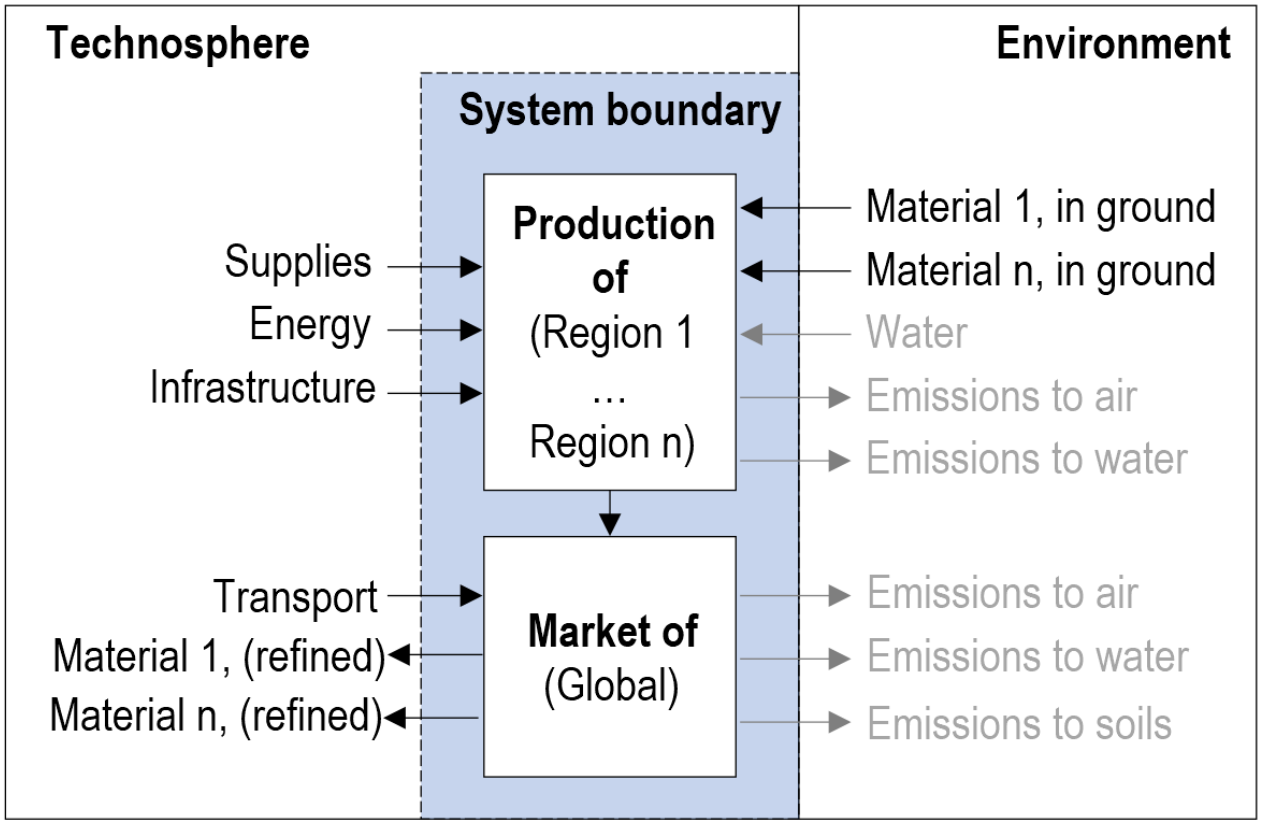
XSL•FO
RenderX

uncertainty and limited knowledge regarding users' walking speeds and the energy consumption of final devices. To achieve this, we evaluate the extent to which different design archetypes can support the acquisition, processing, storage, and transmission of raw accelerometer data under demanding operating conditions and at different sampling rates, independently of software-level optimizations.

The third step provides a high-level estimate of the environmental impact associated with the production of the essential components identified in the previous step and incorporated into the design archetypes. For conciseness and due to space limitations, this analysis was intentionally conceived as a lightweight environmental screening approach focused on the scarcity and potential recovery of raw materials. Its purpose is to provide rapid and pragmatic insights during early-stage product development and eco-design, where knowledge of final devices is typically limited. Specifically, the assessment is based on the identification and quantification of constituent refined materials using official manufacturer-declared material composition data. For simplicity, we focus exclusively on the contributions of gold, silver, and copper to abiotic resource depletion (according to the Environmental Footprint 3.1 impact assessment methodology and antimony-equivalent characterization factors expressed in kg Sb-eq). These materials were selected because of their documented prevalence in electronic devices [29-31] and their relevance to other environmental impact categories, including global warming [32].

Figure 1 illustrates the boundaries and limitations of this streamlined analysis, which accounts for all extractions involved in the global production and commercialization processes available in Ecoinvent 3.10 for the materials considered in this study.

Figure 1. System boundary of the streamlined analysis conducted in this study. Only abiotic resource depletion (kg Sb-eq) was assessed; water consumption and emissions to air, water, and soil (shown in gray) were excluded from the system boundary.



In this sense, this streamlined analysis should not be interpreted as a comprehensive LCA study. Several standardized elements were intentionally excluded from the scope. Specifically, we did not include comprehensive life cycle inventory modeling of manufacturing processes of the electronic components of the design archetypes. Likewise, we did not model the technosphere and the environmental inputs and outputs associated with the use, transportation, maintenance, or waste treatment of final devices. In addition, no sensitivity analysis was conducted to support advanced interpretations of results.

Despite these limitations, the proposed streamlined approach provides a transparent, reproducible, and operationally practical method for integrating preliminary environmental considerations into early design-oriented analyses. As part of this step, we also conducted a conservative assessment of the recovery potential of gold, silver, and copper contained in the electronic components of the design archetypes. The objective was to establish a worst-case scenario for evaluating the minimum recyclability threshold of the final smartbands. This analysis is based on the respective material proportions and their corresponding functional recycling rates, according to Graedel et al [33].

XSL•FO
RenderX

Secondary Analysis of the Accelerometer Data

The information theory concepts used in this part of the research methodology are marginal entropy, conditional entropy, and mutual information. Marginal entropy quantifies the uncertainty associated with a variable independently of another variable [34]. In contrast, conditional entropy quantifies the amount of uncertainty that remains about a variable when the value of another variable is known [35]. Mutual information quantifies the amount of information shared between two variables [34]. For two variables X and Y, the mutual information $MI(X;Y)$ is mathematically defined as the difference between the marginal entropy of X ($H(X)$) and the conditional entropy of X given Y ($H(X|Y)$), as shown in Equation 1.

Equation 1: Mutual information formula

$$MI(X;Y)=H(X)-H(X|Y)$$

When the variables X and Y represent measurements of the same phenomenon, the mutual information $MI(X;Y)$ can be interpreted as the amount of information about X that is also provided by Y [36]. This interpretation follows from the fact that $H(X)$ quantifies the total uncertainty associated with the measurements represented by X, whereas $H(X|Y)$ quantifies the uncertainty that remains after observing the measurements represented by Y.

In this work, X represents the original accelerometer data describing users' movements at a reference sampling rate along the x-, y-, and z-axes, whereas Y represents the corresponding accelerometer data after downsampling. Therefore, if the downsampled signal Y preserves most of the measurements describing the relevant motion patterns of a specific PA, observing Y substantially reduces the uncertainty associated with the original signal X that describes this PA fully, resulting in a high value of $MI(X;Y)$. Conversely, if the downsampled signal Y removes measurements that capture critical details of these motion patterns, the information provided by Y about the original signal becomes limited, and the mutual information decreases accordingly. In practical terms, this means that the downsampled signal no longer retains enough information from the original signal to adequately characterize the PA being monitored.

Following Kvålseth [37], we define the preserved information $PI(X;Y)$ as the fraction of uncertainty in the original signal X that is resolved by observing the transformed signal Y or, equivalently, as the fraction of information contained in the original signal that remains after downsampling (Equation 2).

Equation 2: Preserved information formula

$$PI(X;Y)=MI(X;Y)H(X)$$

So that

$$PI(X;Y)=\{\text{No information loss},PI(X;Y)=1,\text{Partial preservation},0<PI(X;Y)<1,\text{Complete information loss},PI(X;Y)=0.$$

Data Preparation

The secondary analysis was conducted using a previously published dataset of treadmill walking tests performed by healthy participants at speeds ranging from 1.4 to 10 km/h [38]. In this exploratory study, only wrist accelerometer recordings collected with ActiGraph GT9X devices were analyzed, using the original triaxial signals sampled at 30 Hz as the reference data.

The original signals were progressively downsampled from 29 to 1 Hz using cubic spline interpolation. This approach was selected for its computational simplicity and suitability for benchmarking purposes, while remaining consistent with the Nyquist-Shannon theorem [39,40] given the frequency range associated with human walking [41]. In addition, previous studies have shown that spline interpolation can preserve the structural characteristics of a signal when generating lower-frequency representations from known data points [42].

After downsampling, both the original and downsampled signals were discretized using identical binning schemes and organized according to participant, walking speed, and signal axis. This procedure enabled direct comparisons between the reference and modified time series, as illustrated in Rashkovska et al [43]. Additional details regarding data preparation are provided in Multimedia Appendix 1.

Preservation of Information

To quantify the preservation of information in the downsampled signals, the mutual information $MI(X;Y)$ and the preserved information $PI(X;Y)$ were computed using Equations 1 and 2 for each 1-second interval throughout the walking period. Mean values were then calculated for each signal axis and sampling frequency across all walking speeds. Subsequently, a deviation analysis was performed to evaluate the consistency of these results in accordance with conventional statistical methods [44].

In addition, changes in preserved information with respect to variations in walking speed and sampling frequency were quantified through $PI'(X;Y)$ for each participant and signal axis, according to Equation 3.

Equation 3: Changes in the preserved information

$$PI`(X;Y)=|PI(X;Y)s-PI(X;Y)s`PI(X;Y)f-PI(X;Y)f`|$$

where

PI(X;Y)s=PI for a given walking speed

PI(X;Y)s′=PI for a walking speed lower than the given speed

PI(X;Y)f=PI for a given sampling frequency

PI(X;Y)f′=PI for a sampling frequency lower than the given frequency

More specifically, PI′(X;Y) is the derivative that systematically models the instantaneous rate of change of the preserved information variable. It quantifies how rapidly preserved information changes as a participant transitions from a given walking speed (PI(X;Y)s) to a higher walking speed (PI(X;Y)s`), while the accelerometer of the WMA device simultaneously transitions from a given sampling frequency (PI(X;Y)f) to a lower sampling frequency (PI(X;Y)f`).

Evaluation of the Downsampled Signals

To estimate the viability of the downsampled signals for step detection, the procedure adopted in this exploratory study was

XSL•FO
RenderX

divided into two stages and repeated across the x-, y-, and z-axes.

First, we identified and counted peaks and valleys in both the original and downsampled signals. This approach was selected because step-counting methods commonly rely on the detection of peaks and valleys in accelerometer data. Specifically, gray literature suggests that peak detection constitutes the basis of pedometer software implemented in widely used commercial sensors [45], while academic studies identify the detection of peaks [46] or peak-valley pairs [47] as standard approaches for step-counting algorithms.

In our study, peaks were identified using the find_peaks algorithm from the scipy.signal Python library. Detection was based on local extrema and constrained by a topographic prominence of 0.2 and a height threshold equal to the mean acceleration value plus half of its standard deviation for each axis. Valleys were identified using the same procedure applied to the inverted signals.

Second, we computed the percentage error of peak counts (PEpeaksi,j) and valley counts (PEvalleysi,j) for the downsampled signals at each walking speed *i* and sampling frequency *j*. The corresponding peak and valley counts obtained from the original signals were used as reference values, according to Equations 4 and 5.

Equation 4: Percentage error of peaks formula

$$PEpeaks_{i,j} = \frac{|PC_{i,j} - PC_{i,j'}|}{PC_{i,j}} \times 100\%, \forall i \in \{1,\ldots,29\}, j \in \{1.4,\ldots,10\}$$

Equation 5: Percentage error of valleys formula

$$PEvalleys_{i,j} = \frac{|VC_{i,j} - VC_{i,j'}|}{VC_{i,j}} \times 100\%, \forall i \in \{1,\ldots,29\}, j \in \{1.4,\ldots,10\}$$

where

$PC_{i,j}$ = Peak Counting ∈ the original signal for walking speed $i \wedge$ frequency j

$PC_{i,j'}$ = Peak Counting ∈ a downsampled signal for walking speed $i \wedge$ frequency j

$VC_{i,j}$ = Valley Counting ∈ the original signal for walking speed $i \wedge$ frequency j

$VC_{i,j'}$ = Valley Counting ∈ a downsampled signal for walking speed $i \wedge$ frequency j

In this manner, the results of this evaluation complement those of the previous section from a functional perspective. Specifically, they make it possible to assess the extent to which the information retained in the downsampled signals remains suitable for step-detection and step-counting applications.

Ethical Considerations

This study did not involve any direct experimentation with human participants or patients. The analyses were conducted exclusively using secondary data derived from a previously published study. All data sources were appropriately cited, and data use complied with the conditions and guidelines specified in the original publication. The original study that generated the dataset used in our secondary analysis is part of a broader body of work executed under the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement no. 823871 (iGame), and it was approved by the Ethics Committee of the Malaga Provincial Research of the Andalusian Health Service (RTC-iGAME 24092020). As reported by the authors of the original study, all participants accepted and signed the informed consent following the principles established in the Declaration of Helsinki.

Results

Preserved Information

Table 1 presents the mean preserved information values of the downsampled signals across all participants, grouped by sampling frequency.

XSL•FO
RenderX

Table . Preserved information in the downsampled signals across all participants, reported as mean percentages and standard deviations, for all walking speeds[a].

Freq. (Hz)	Participant 1			Participant 2			Participant 3			Participant 4			Participant 5		
	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
1	4.22 (0.52)	4.23 (0.57)	4.20 (0.65)	4.25 (0.47)	4.23 (0.58)	4.19 (0.60)	4.22 (0.54)	4.19 (0.66)	4.20 (0.64)	4.23 (0.50)	4.21 (0.70)	4.19 (0.71)	4.22 (0.50)	4.23 (0.55)	4.20 (0.67)
2	8.45 (0.67)	8.39 (0.80)	8.34 (0.87)	8.44 (0.65)	8.39 (0.78)	8.36 (0.80)	8.44 (0.71)	8.31 (0.91)	8.35 (0.88)	8.45 (0.67)	8.28 (0.99)	8.30 (0.95)	8.44 (0.68)	8.39 (0.77)	8.36 (0.94)
3	12.55 (0.84)	12.51 (0.92)	12.40 (1.12)	12.64 (0.76)	12.50 (1.02)	12.49 (0.95)	12.58 (0.84)	12.41 (1.12)	12.42 (1.06)	12.63 (0.76)	12.38 (1.15)	12.40 (1.14)	12.64 (0.81)	12.58 (0.90)	12.45 (1.08)
4	16.72 (0.90)	16.62 (1.05)	16.53 (1.15)	16.76 (0.91)	16.62 (1.07)	16.56 (1.08)	16.71 (0.97)	16.48 (1.25)	16.52 (1.21)	16.74 (0.87)	16.41 (1.36)	16.43 (1.36)	16.76 (0.86)	16.61 (1.07)	16.54 (1.24)
5	20.84 (1.02)	20.68 (1.14)	20.60 (1.29)	20.79 (1.00)	20.72 (1.21)	20.64 (1.18)	20.76 (1.03)	20.56 (1.36)	20.57 (1.29)	20.85 (0.96)	20.46 (1.42)	20.45 (1.42)	20.85 (0.94)	20.70 (1.15)	20.47 (1.37)
6	24.87 (1.01)	24.74 (1.23)	24.54 (1.39)	24.93 (0.99)	24.75 (1.30)	24.71 (1.20)	24.86 (1.07)	24.56 (1.42)	24.65 (1.38)	24.92 (1.02)	24.41 (1.53)	24.46 (1.46)	24.92 (0.99)	24.75 (1.22)	24.58 (1.36)
7	28.87 (1.13)	28.78 (1.23)	28.58 (1.39)	28.98 (1.05)	28.76 (1.36)	28.74 (1.26)	28.89 (1.12)	28.50 (1.53)	28.60 (1.46)	28.89 (1.11)	28.47 (1.62)	28.43 (1.55)	28.89 (1.08)	28.69 (1.29)	28.46 (1.50)
8	32.92 (1.12)	32.76 (1.26)	32.52 (1.50)	32.94 (1.13)	32.73 (1.40)	32.66 (1.33)	32.86 (1.21)	32.48 (1.59)	32.51 (1.48)	32.93 (1.13)	32.37 (1.73)	32.30 (1.74)	32.97 (1.05)	32.74 (1.36)	32.53 (1.45)
9	36.80 (1.21)	36.65 (1.34)	36.42 (1.49)	36.86 (1.10)	36.65 (1.47)	36.66 (1.33)	36.85 (1.19)	36.40 (1.69)	36.44 (1.50)	36.90 (1.10)	36.29 (1.78)	36.22 (1.72)	36.91 (1.10)	36.69 (1.31)	36.42 (1.56)
10	40.78 (1.09)	40.52 (1.41)	40.35 (1.50)	40.71 (1.13)	40.55 (1.47)	40.51 (1.38)	40.66 (1.18)	40.33 (1.66)	40.30 (1.52)	40.76 (1.15)	40.15 (1.82)	40.11 (1.77)	40.81 (1.10)	40.56 (1.42)	40.23 (1.59)
11	44.63 (1.12)	44.38 (1.40)	44.14 (1.51)	44.66 (1.05)	44.38 (1.47)	44.35 (1.28)	44.56 (1.17)	44.08 (1.72)	44.23 (1.55)	44.61 (1.17)	43.97 (1.85)	43.96 (1.74)	44.64 (1.16)	44.41 (1.43)	44.06 (1.59)
12	48.47 (1.11)	48.20 (1.40)	47.96 (1.52)	48.50 (1.11)	48.18 (1.55)	48.21 (1.31)	48.39 (1.24)	47.94 (1.65)	48.00 (1.63)	48.46 (1.13)	47.77 (1.84)	47.74 (1.80)	48.50 (1.05)	48.21 (1.39)	47.95 (1.60)
13	52.19 (1.13)	51.98 (1.39)	51.67 (1.52)	52.29 (1.06)	51.99 (1.47)	51.96 (1.25)	52.15 (1.23)	51.65 (1.74)	51.73 (1.52)	52.21 (1.21)	51.61 (1.77)	51.48 (1.87)	52.23 (1.09)	52.01 (1.42)	51.60 (1.54)
14	55.92 (1.12)	55.74 (1.30)	55.35 (1.50)	56.00 (1.03)	55.67 (1.46)	55.70 (1.28)	55.92 (1.15)	55.37 (1.71)	55.49 (1.53)	55.95 (1.14)	55.35 (1.80)	55.18 (1.80)	55.96 (1.09)	55.68 (1.39)	55.37 (1.54)
15	59.60 (1.08)	59.33 (1.32)	59.05 (1.43)	59.66 (1.02)	59.31 (1.50)	59.32 (1.24)	59.57 (1.13)	58.97 (1.71)	59.07 (1.50)	59.67 (1.07)	59.00 (1.77)	58.80 (1.83)	59.63 (1.07)	59.41 (1.27)	59.04 (1.55)
16	63.23 (1.08)	63.01 (1.26)	62.71 (1.39)	63.29 (1.03)	62.91 (1.48)	62.95 (1.19)	63.18 (1.11)	62.64 (1.60)	62.73 (1.44)	63.22 (1.02)	62.53 (1.76)	62.50 (1.72)	63.29 (0.97)	63.06 (1.30)	62.66 (1.51)
17	66.81 (0.96)	66.56 (1.23)	66.25 (1.24)	66.76 (0.95)	66.53 (1.37)	66.52 (1.13)	66.69 (0.99)	66.22 (1.56)	66.25 (1.45)	66.74 (1.05)	66.12 (1.75)	66.06 (1.65)	66.81 (0.95)	66.58 (1.23)	66.27 (1.44)
18	70.24 (1.01)	70.04 (1.20)	69.84 (1.19)	70.28 (0.87)	70.04 (1.34)	70.02 (1.06)	70.22 (1.01)	69.73 (1.46)	69.78 (1.37)	70.28 (0.95)	69.62 (1.56)	69.58 (1.60)	68.18 (6.77)	70.05 (1.18)	69.71 (1.35)
19	73.61 (0.93)	73.45 (1.06)	73.22 (1.16)	73.72 (0.82)	73.35 (1.25)	73.41 (0.97)	73.60 (0.96)	73.16 (1.41)	73.25 (1.21)	73.68 (0.86)	73.06 (1.53)	73.00 (1.46)	73.69 (0.93)	73.45 (1.12)	73.14 (1.29)
20	76.95 (0.85)	76.77 (1.01)	76.52 (1.08)	76.98 (0.78)	76.74 (1.21)	76.71 (0.92)	76.89 (0.92)	76.56 (1.30)	76.58 (1.13)	76.98 (0.81)	76.42 (1.41)	76.32 (1.47)	77.02 (0.80)	76.74 (1.01)	76.47 (1.15)
21	80.15 (0.81)	79.98 (0.91)	79.84 (0.95)	80.19 (0.70)	79.99 (1.12)	80.04 (0.85)	80.15 (0.80)	79.76 (1.20)	79.81 (1.09)	80.20 (0.74)	79.69 (1.32)	79.64 (1.32)	80.21 (0.74)	80.09 (0.87)	79.78 (1.08)
22	83.31 (0.68)	83.15 (0.82)	82.96 (0.89)	83.33 (0.61)	83.11 (1.00)	83.14 (0.76)	83.27 (0.75)	82.96 (1.07)	83.00 (0.98)	83.33 (0.68)	82.82 (1.24)	82.84 (1.10)	83.34 (0.67)	83.18 (0.85)	82.95 (0.96)
23	86.32 (0.61)	86.19 (0.72)	86.03 (0.74)	86.38 (0.55)	86.18 (0.83)	86.18 (0.69)	86.28 (0.66)	85.99 (0.96)	86.07 (0.87)	86.31 (0.61)	85.91 (1.05)	85.89 (0.96)	86.32 (0.66)	86.23 (0.72)	85.98 (0.87)
24	89.19 (0.58)	89.05 (0.63)	88.95 (0.65)	89.19 (0.49)	89.04 (0.77)	89.08 (0.57)	89.14 (0.60)	88.92 (0.77)	88.95 (0.76)	89.19 (0.51)	88.83 (0.87)	88.78 (0.93)	89.23 (0.48)	89.09 (0.61)	88.96 (0.74)

XSL•FO
RenderX

Freq. (Hz)	Participant 1			Participant 2			Participant 3			Participant 4			Participant 5		
	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
25	91.88 (0.48)	91.79 (0.50)	91.70 (0.56)	91.90 (0.41)	91.79 (0.65)	91.80 (0.50)	91.86 (0.44)	91.69 (0.67)	91.71 (0.62)	91.87 (0.43)	91.61 (0.76)	91.56 (0.75)	91.90 (0.49)	91.80 (0.52)	91.66 (0.60)
26	94.39 (0.41)	94.33 (0.42)	94.24 (0.45)	94.41 (0.33)	94.33 (0.46)	94.33 (0.36)	94.36 (0.44)	94.25 (0.52)	94.25 (0.46)	94.39 (0.36)	94.20 (0.61)	94.19 (0.56)	94.40 (0.36)	94.34 (0.42)	94.22 (0.47)
27	96.65 (0.29)	96.61 (0.32)	96.57 (0.34)	96.66 (0.26)	96.60 (0.40)	96.61 (0.31)	96.62 (0.38)	96.54 (0.39)	96.56 (0.40)	96.65 (0.29)	96.51 (0.45)	96.51 (0.39)	96.65 (0.32)	96.63 (0.29)	96.56 (0.38)
28	98.58 (0.25)	98.55 (0.26)	98.55 (0.20)	98.58 (0.17)	98.54 (0.30)	98.56 (0.22)	98.57 (0.25)	98.53 (0.30)	98.53 (0.26)	98.58 (0.24)	98.52 (0.30)	98.50 (0.29)	98.59 (0.25)	98.56 (0.23)	98.54 (0.28)
29	99.98 (0.20)	99.98 (0.15)	99.98 (0.17)	99.97 (0.21)	99.99 (0.13)	99.97 (0.21)	99.97 (0.20)	99.98 (0.15)	99.98 (0.16)	99.98 (0.17)	99.97 (0.19)	99.98 (0.18)	99.97 (0.21)	99.98 (0.16)	99.96 (0.23)

[a]Full visualization and detailed results of the deviation analysis validating these tendencies are available in Multimedia Appendix 2.

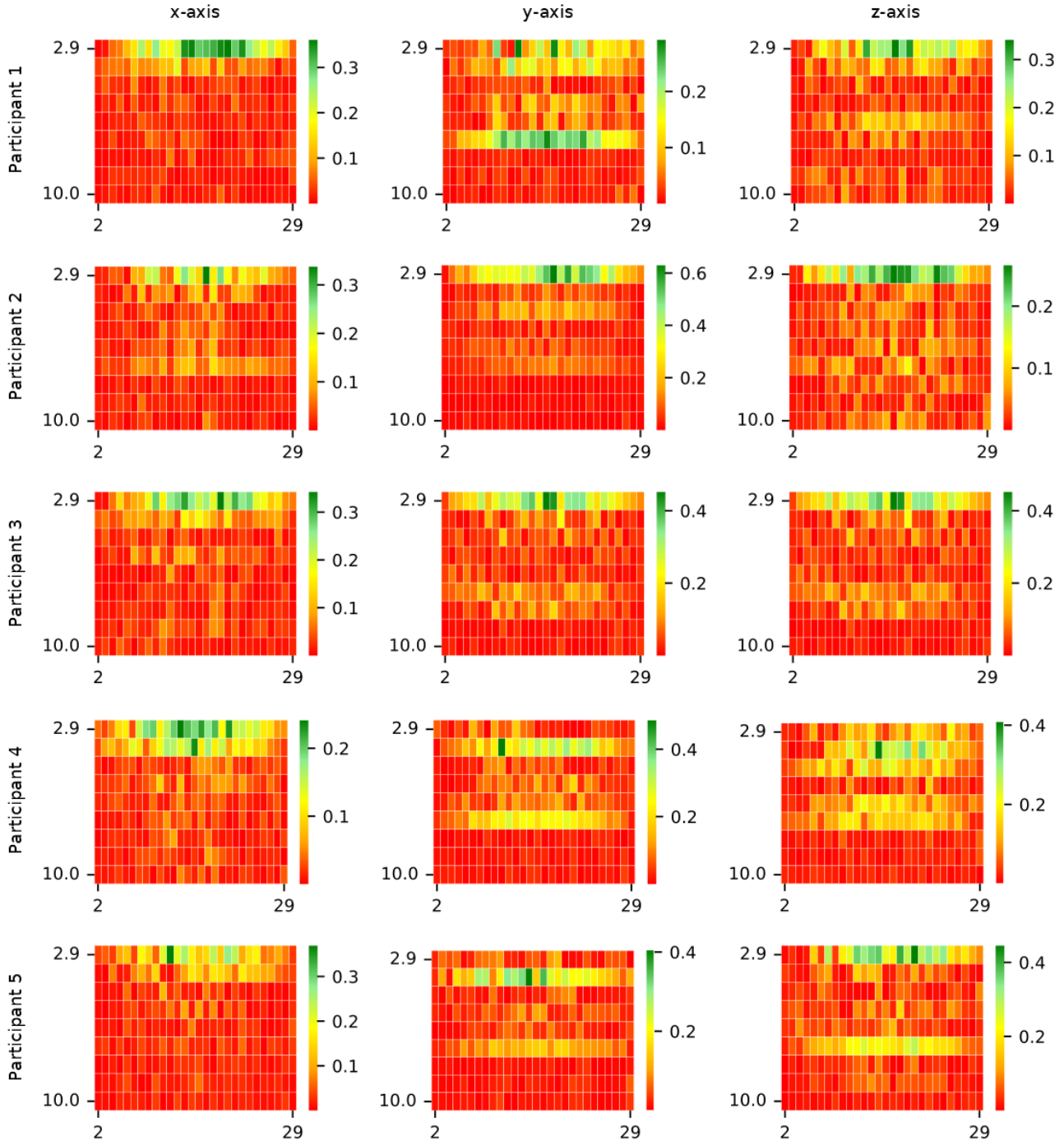
The results reveal a consistent decrease in $PI(X;Y)$ as the sampling frequency decreases. Specifically, the downsampled signals exhibit substantial information losses exceeding 91% at 2 Hz and 79% at 5 Hz. In contrast, information loss remains below 24% at 20 Hz, indicating a more moderate degradation of information.

Changes in the Preserved Information

Figure 2 summarizes our preliminary findings regarding how preserved information varies with changes in walking speed and sampling frequency ($PI'(X;Y)$). Specifically, each cell within each heatmap represents the ratio between the changes in $PI(X;Y)$ associated with a stepwise increase in walking speed and the corresponding changes in $PI(X;Y)$ associated with a stepwise reduction in the accelerometer sampling frequency.

XSL•FO
RenderX

Figure 2. PI′(X;Y) values showing changes in preserved information as walking speed increases (from 1.4 to 2.9 km/h in the first row, from 2.9 to 4.3 km/h in the second row, and so forth) and sampling frequency decreases (from 30 to 29 Hz in the last column, from 29 to 28 Hz in the penultimate column, and so forth).



Across all participants and signal axes, substantial information losses are observed at both very low and very high sampling frequencies. However, no consistent pattern emerges when participants walk at speeds between 8 and 10 km/h.

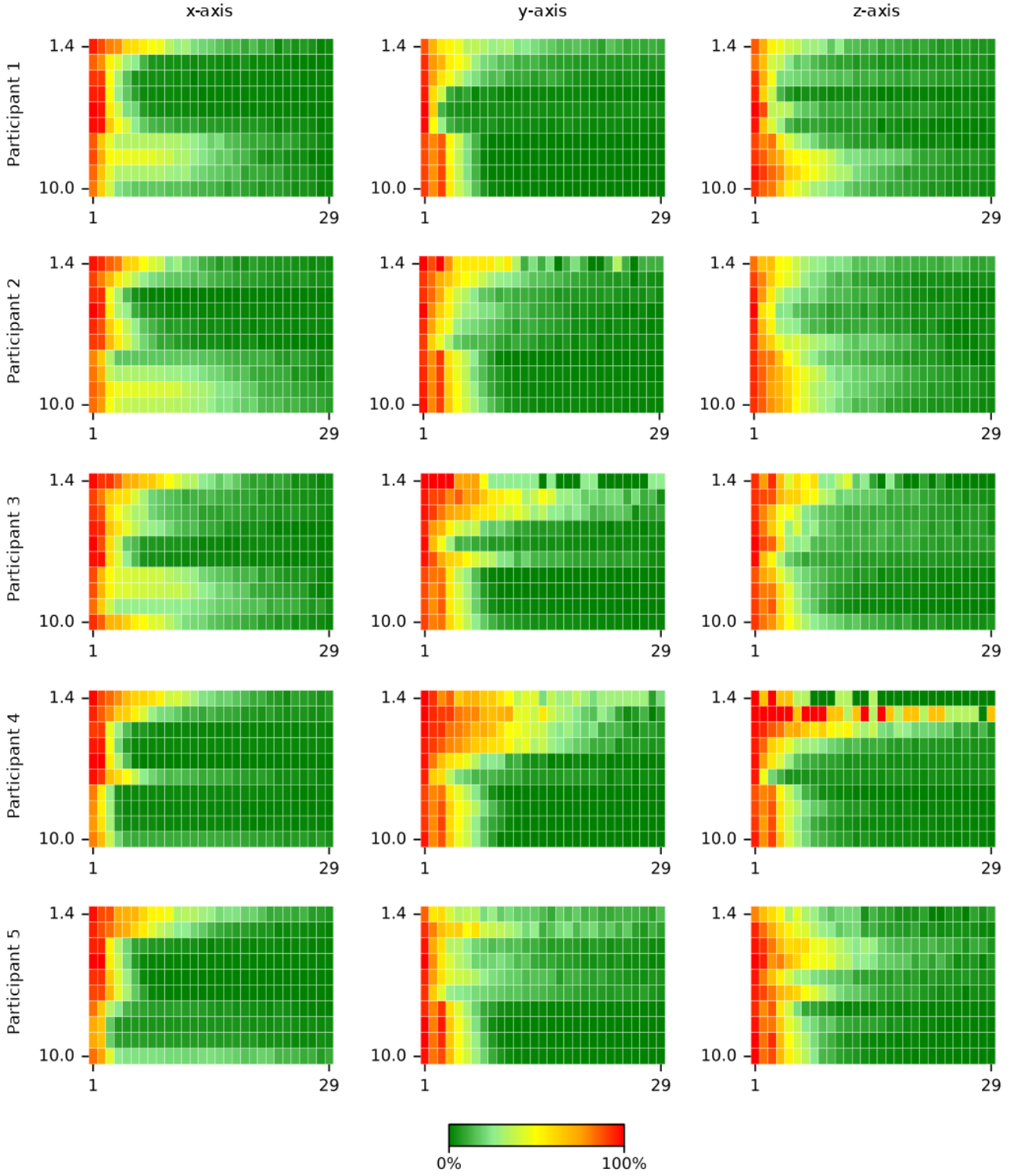
In general, when participants walk at speeds between 1.4 and 2.9 km/h, information is preserved in all downsampled signals with only limited degradation at sampling frequencies between 7 and 25 Hz. However, this window progressively contracts as walking speed increases up to 6.5 km/h. For the y-axis, a noticeable increase in preserved information is observed for participants 1, 3, 4, and 5 when walking speed increases from 6.5 to 7.0 km/h. Although the cause of this behavior remains unclear, it does not alter the overall trends identified in this analysis.

Results From the Evaluation of the Downsampled Signals

Figures 3 and 4 present our preliminary findings on the percentage error calculated from the peaks and valleys detected in the original and downsampled signals of all participants.

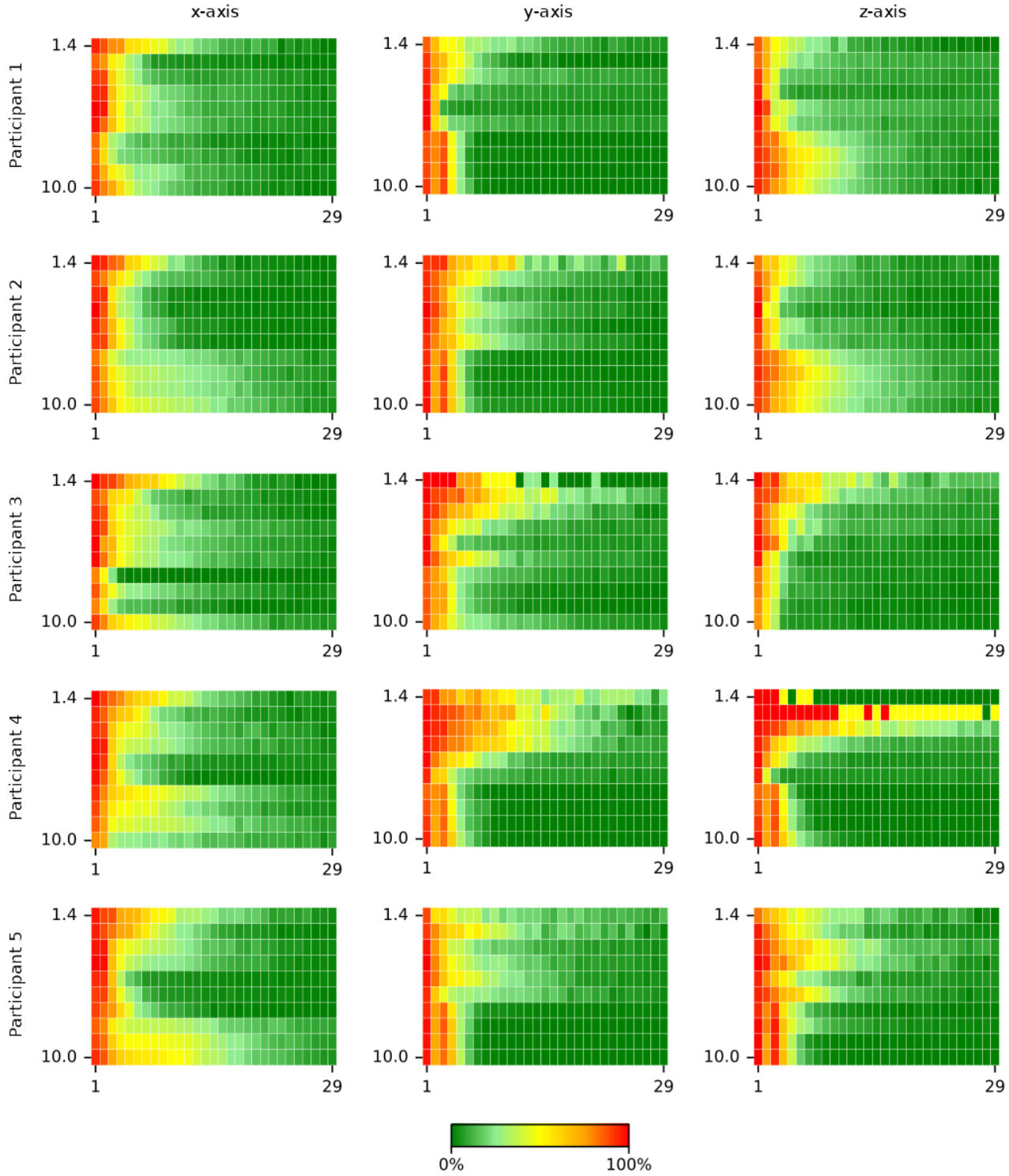
XSL•FO
RenderX

Figure 3. Percentage error (PE) of peak counts in the original and downsampled signals across all participants, sampling frequencies, and walking speeds. In each heatmap, gait speed (km/h) is shown on the vertical axis and sampling frequency (Hz) on the horizontal axis.



XSL•FO
RenderX

Figure 4. Percentage error (PE) of valley counts in the original and downsampled signals across all participants, sampling frequencies, and walking speeds. In each heatmap, gait speed (km/h) is shown on the vertical axis and sampling frequency (Hz) on the horizontal axis.



Across all participants and signal axes, the percentage error associated with peak and valley counts remains very high for downsampled signals sampled at very low frequencies (below 3 Hz). This result suggests that such signals are unlikely to support reliable step detection, regardless of walking speed.

In addition, the suitability of downsampled signals for step detection on the y-axis appears to deteriorate at walking speeds below 6 km/h when sampling frequencies fall below approximately 11 Hz. Under these conditions, the percentage error approaches 80%, suggesting a substantial reduction in detection performance.

Design Delineation of Frugal Smartbands

The preliminary results presented in Figures 2-4 suggest that sampling frequencies around 15 Hz are likely sufficient to preserve the information for step detection during normal walking (ie, between 4.3 and 7 km/h). In addition, a broader frequency range (approximately 11 - 25 Hz) may remain

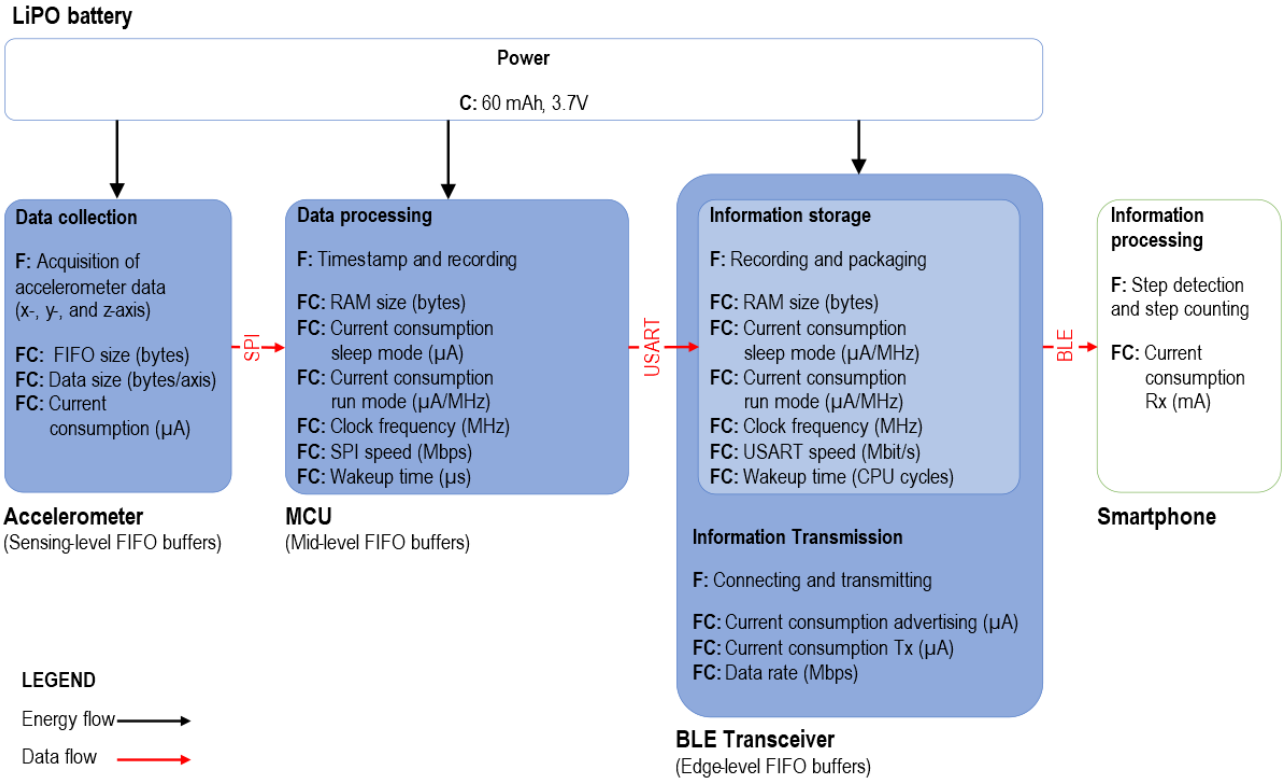
XSL•FO
RenderX

suitable when individuals walk at very slow speeds (below 3 km/h).

In practice, these relationships between sampling frequency, walking speed, and information preservation for effective step detection are often difficult to determine during the design of WAM devices. As a result, research and development efforts frequently rely on resource-intensive dynamic sampling strategies, including optimized duty-cycling approaches, on-chip processing, and advanced computing components. In this exploratory work, we use the findings from our secondary data analysis to examine how variations in sampling frequency and walking speed may influence the design of frugal smartbands. Specifically, we establish a set of hardware archetypes capable of supporting a suboptimal algorithm operating under different sampling rates. This approach allows us to define a worst-case scenario for energy consumption and user-behavior uncertainty, thereby providing insight into the robustness of the resulting devices.

To support this analysis, we adopt a simple asynchronous first-in-first-out (FIFO) strategy as the basis for communication between 3 essential hardware components. We then relate the data acquisition, processing, and transmission functions of these components to their respective capacities to model and estimate energy consumption and battery autonomy across different design archetypes and sampling frequencies, according to the framework presented in Figure 5.

Figure 5. Framework for constructing design archetypes of frugal smartbands based on a simple asynchronous first-in-first-out (FIFO) strategy and a mapping of the capacities (C), functions (F), and function-capacities (FC) of their essential electronic components (adapted from the sustainable IoT design framework proposed by Quisbert-Trujillo and Morfouli [28]). BLE: Bluetooth Low Energy; MCU: microcontroller; SPI: serial peripheral interface; USART: universal synchronous/asynchronous receiver transmitter.



Each archetype consists of 3 core elements: an accelerometer, an ARM-based microcontroller (MCU), and a Bluetooth Low Energy (BLE) transceiver powered by a 60 mAh LiPo battery and organized as an asynchronous FIFO-based system. The accelerometer continuously acquires triaxial samples, which are transferred in bursts through the MCU and BLE transceiver before being transmitted to a smartphone. The energy model incorporates the current consumption and operating duration of each component in low-power, run, and sleep modes, including wake-up, processing, and data transmission functions (*F*).

Battery autonomy is estimated from the available battery capacity and the energy required to fill and transmit a complete BLE FIFO buffer, following the approach described by Tarkoma et al [48]. Because the FIFO filling time is determined by the accelerometer sampling frequency, this parameter directly influences the run and sleep durations of the remaining components and, consequently, the overall energy consumption of the system. Other factors affecting battery autonomy include communication throughput, processor clock frequency, workload, and FIFO/RAM capacities.

Using the component configurations and unfavorable operating conditions summarized in Tables 2 and 3, we evaluated all combinations of 4 accelerometers, 3 MCUs, and 3 BLE transceivers across multiple sampling frequencies. The objective was to identify the configurations capable of achieving at least 10 days of battery autonomy, a benchmark representative of current commercial wearable devices [49].

XSL•FO
RenderX

Table . Configuration settings for the accelerometers used to construct the archetypes[a].

Capacities	MPU6000 [50]	LIS2DH12 [51]	LIS2DE12 [52]	LIS2DW12 [53]
FIFO[b] size (bytes)	1024	192	192	192
Current consumption 1 Hz (µA)	—[c]	2	2	—
Current consumption 1.25 Hz (µA)	10	—	—	—
Current consumption 1.6 Hz (µA)	—	—	—	0.38
Current consumption 5 Hz (µA)	20	—	—	—
Current consumption 10 Hz (µA)	—	4	3	—
Current consumption 12.5 Hz (µA)	—	—	—	1
Current consumption 20 Hz (µA)	60	—	—	—
Current consumption 25 Hz (µA)	—	6	4	—
Clock frequency (MHz)	20	10	10	10
Voltage power supply (V)	3.46	3.6	3.6	3.6

[a]Official values as reported in the respective documentation cited in the references section.

[b]FIFO: first-in-first-out.

[c]Not available.

XSL•FO
RenderX

Table . Configuration settings for the MCUs[a] and BLE[b] transceivers used to construct the archetypes[c].

	MCU			BLE transceiver		
Capacities	L011D3 [54]	L053C8 [55]	L073CB [56]	WB10CC [57]	WB50CG [58]	WB55CC [59]
RAM size (KBytes)	2	8	20	48	128	256
Current consumption sleep mode (µA)	0.54	0.8	0.86	28	41	41
Current consumption run mode (µA/MHz)	76	88	93	91	107	107
Current consumption advertising mode (µA)	__[d]	—	—	19.2	20	13
Current consumption transmission mode (µA)	—	—	—	8600	8800	5200
SPI[e] throughput (Mbits/s)	16	16	16	—	—	—
USART[f] throughput (Mbits/s)	—	—	—	4	4	4
BLE throughput (Mbits/s)	—	—	—	1	1	2
Wake-up time (µs)	5	5	5	—	—	—
Voltage power supply (V)	3.6	3.6	3.6	3.6	3.6	3.6

[a]MCU: microcontroller.

[b]BLE: Bluetooth Low Energy.

[c]Official values as reported in the respective documentation cited in the references section.

[d]Not available.

[e]SPI: serial peripheral interface.

[f]USART: universal synchronous/asynchronous receiver transmitter.

Based on these parameters, Tables 4-7 report the estimated runtime per full battery charge for different combinations. Archetypes capable of achieving at least 10 days of operational autonomy are highlighted in italics. The tables correspond respectively to the MPU6000, LIS2DH12, LIS2DE12, and LIS2DW12 accelerometers and include all of their reported sampling frequencies.

Table . Runtime per full charge of studied archetypes (in days) according to the low sampling rates reported for the MPU6000 accelerometer[a].

	L011D3			L053C8			L073CB		
	1.25 Hz	5 Hz	20 Hz	1.25 Hz	5 Hz	20 Hz	1.25 Hz	5 Hz	20 Hz
WB10CC	*38.09*	*32.22*	*19.93*	10.78	10.25	8.57	3.18	3.13	2.96
WB50CG	*63.07*	*48.40*	*25.07*	*19.32*	*17.68*	*13.20*	6.45	6.26	5.59
WB55CC	*103.01*	*72.66*	*33.35*	*36.23*	*31.59*	*20.89*	*12.65*	*12.03*	*10.07*

[a]Archetypes capable of achieving at least 10 days of operational autonomy are highlighted in italics.

XSL•FO
RenderX

Table . Runtime per full charge of studied archetypes (in days) according to the low sampling rates reported for the LIS2DE12 accelerometer[a].

	L011D3			L053C8			L073CB		
	1 Hz	10 Hz	25 Hz	1 Hz	10 Hz	25 Hz	1 Hz	10 Hz	25 Hz
WB10CC	*38.65*	*34.76*	*29.99*	9.57	9.31	8.93	3.21	3.18	3.14
WB50CG	*69.56*	*57.73*	*45.51*	*17.35*	*16.50*	*15.33*	6.58	6.46	6.27
WB55CC	*133.34*	*115.70*	*97.18*	*34.40*	*33.10*	*31.39*	*13.14*	*12.95*	*12.68*

[a]Archetypes capable of achieving at least 10 days of operational autonomy are highlighted in italics.

Table . Runtime per full charge of studied archetypes (in days) according to the low sampling rates reported for the LIS2DH12 accelerometer[a].

	L011D3			L053C8			L073CB		
	1 Hz	10 Hz	25 Hz	1 Hz	10 Hz	25 Hz	1 Hz	10 Hz	25 Hz
WB10CC	*38.65*	*34.29*	*29.31*	9.57	9.28	8.87	3.21	3.18	3.13
WB50CG	*69.56*	*56.46*	*43.95*	*17.35*	*16.40*	*15.15*	6.58	6.44	6.24
WB55CC	*133.34*	*110.71*	*90.35*	*34.40*	*32.68*	*30.64*	*13.14*	*12.88*	*12.55*

[a]Archetypes capable of achieving at least 10 days of operational autonomy are highlighted in italics.

Table . Run time per full charge of studied archetypes (in days) according to the low sampling rates reported for the LIS2DW12 accelerometer[a].

	L011D3		L053C8		L073CB	
	1.6 Hz	12.5 Hz	1.6 Hz	12.5 Hz	1.6 Hz	12.5 Hz
WB10CC	*39.35*	*34.86*	9.62	9.32	3.22	3.18
WB50CG	*71.85*	*57.96*	*17.49*	*16.52*	6.60	6.46
WB55CC	*144.52*	*123.85*	*35.10*	*33.73*	*13.24*	*13.05*

[a]Archetypes capable of achieving at least 10 days of operational autonomy are highlighted in italics.

Accordingly, and considering the capacities and parameters summarized in Tables 2 and 3, our preliminary results suggest that midlevel FIFO buffers should not be implemented across the full memory capacity of microcontrollers with moderate or relatively large storage capacities (such as the MCU L053C8). Under these conditions, transferring and writing large amounts of buffered data would prolong the active periods of other components, particularly BLE transceivers, thereby increasing energy overhead and potentially reducing battery autonomy below 10 days. This effect may become more pronounced at higher accelerometer sampling frequencies (above 20 Hz), which accelerates FIFO saturation throughout the system.

Conversely, this limitation could be mitigated when FIFO buffers are implemented across the full memory capacity of edge-level components, particularly the WB55CC BLE transceiver, provided that efficient radio communication capacities are available.

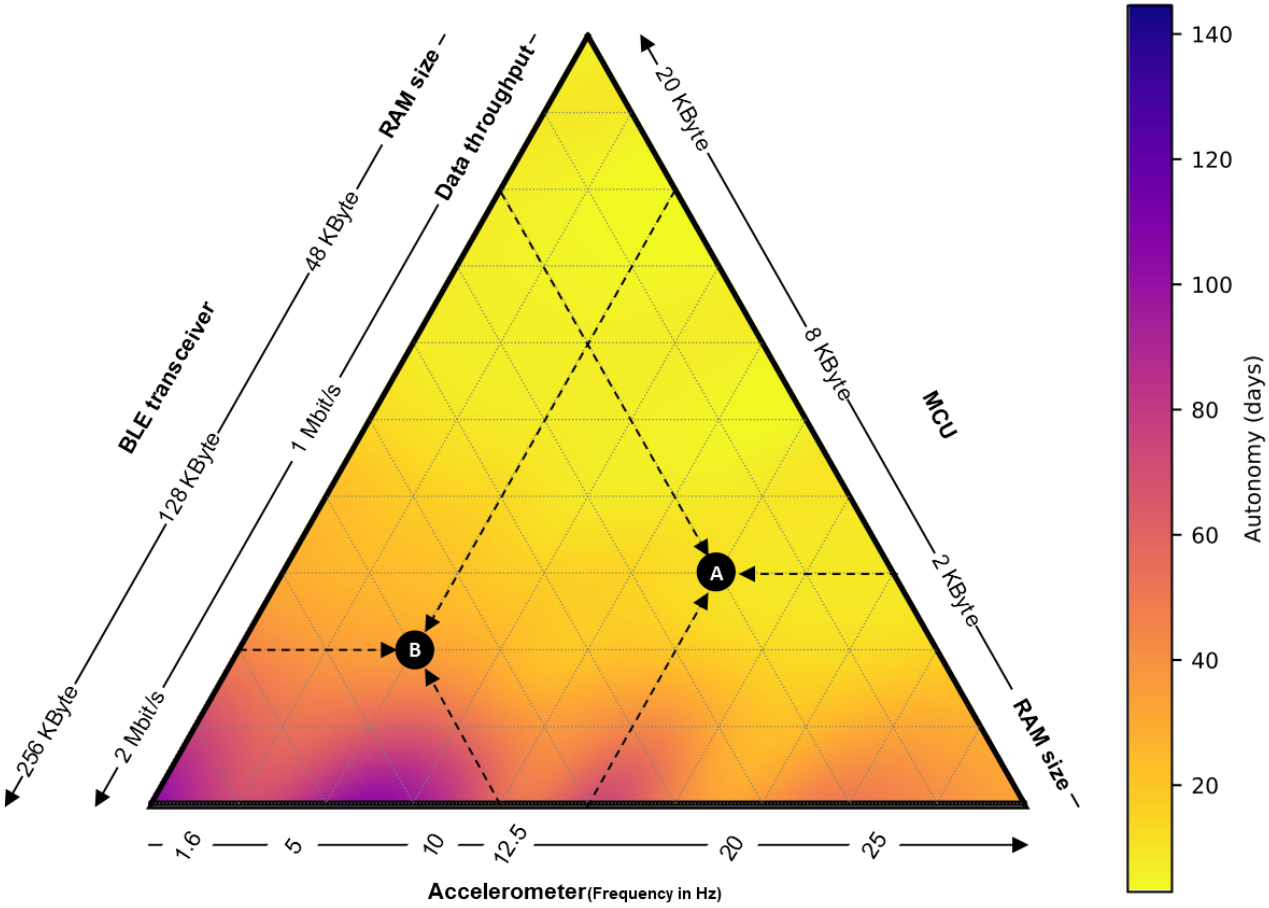
Figure 6 summarizes these findings in a practical framework intended to support the design of frugal yet robust smartbands for walking monitoring. The framework is based on the sensing, processing, storage, and transmission function-capacities (FC) of the electronic components that constitute the design archetypes evaluated in this study.

This preliminary framework helps visualize hardware trade-offs and the flexibility available for distributing computational workloads across system components. For example, point A suggests that designers should restrict MCUs to lightweight temporary storage functions and resource-constrained transceivers to simple data retransmission tasks. Such a configuration would support smartbands intended to operate primarily around 15 Hz while remaining close to a common commercial benchmark of approximately 10 days of battery autonomy.

In contrast, Point B indicates greater flexibility for implementing dynamic sampling strategies within the microprocessors of accelerometers, MCUs, or BLE transceivers. This configuration may be advantageous when supporting adaptive sampling frequencies below 12.5 Hz (but above 11 Hz), which correspond to situations in which users walk at very slow speeds. Under these conditions, a larger proportion of the available energy budget could potentially be allocated to onboard preprocessing or step-detection functions, as suggested by the estimated battery autonomy of 20 - 40 days, even under the unfavorable operating conditions considered in this analysis.

XSL•FO
RenderX

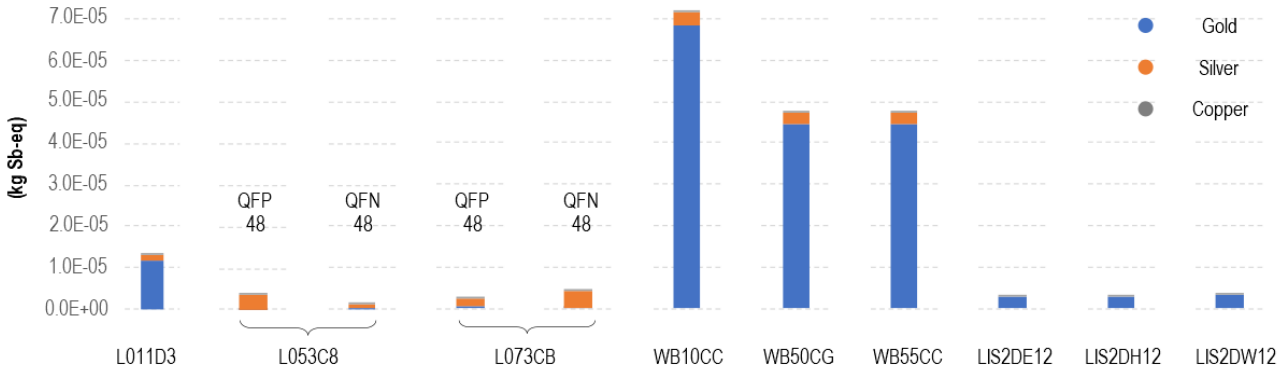
Figure 6. Practical design framework for frugal and efficient smartbands, according to the studied capacities of the electronic components that constitute the design archetypes. The scales on the right and left sides of the isosceles triangle show the RAM sizes of the studied microcontrollers (MCUs) and Bluetooth Low Energy (BLE) transceivers, as described in Table 3. The left scale also displays the data throughput of the BLE transceivers. The scale at the base shows the frequency range studied in our secondary data analysis, while the color scale on the rightmost side shows the battery autonomy range calculated from all combined function-capacities (FC) of the electronic components summarized in Figure 5, and the corresponding values and current consumptions reported in Tables 2 and 3. Each intersection within the triangle indicates the maximum battery autonomy (in days) associated with our worst-case use scenario, based on a suboptimal first-in-first-out (FIFO) strategy and on specific combinations of RAM capacity, communication throughput, and sampling frequency. For example, point A indicates that a battery autonomy below 20 days could be expected when combining a 2-Kbyte MCU RAM capacity (eg, provided by the L011D3 MCU) with a BLE transceiver offering less than 48 Kbytes of RAM and a data throughput below 1 Mbit/s (eg, the WB10CC transceiver) at an accelerometer sampling frequency of 15 Hz.



Environmental Impact Overview

Figure 7 presents the estimated resource depletion associated with the extraction of gold, silver, and copper required to manufacture the electronic components included in the design archetypes. Results for the MPU6000 accelerometer are unavailable because the manufacturer does not provide a material declaration.

Figure 7. Preliminary results showing the resource depletion incurred in the production of the electronic components of the design archetypes, according to the Environmental Footprint 3.1 assessment method [60], expressed in kg Sb-eq.

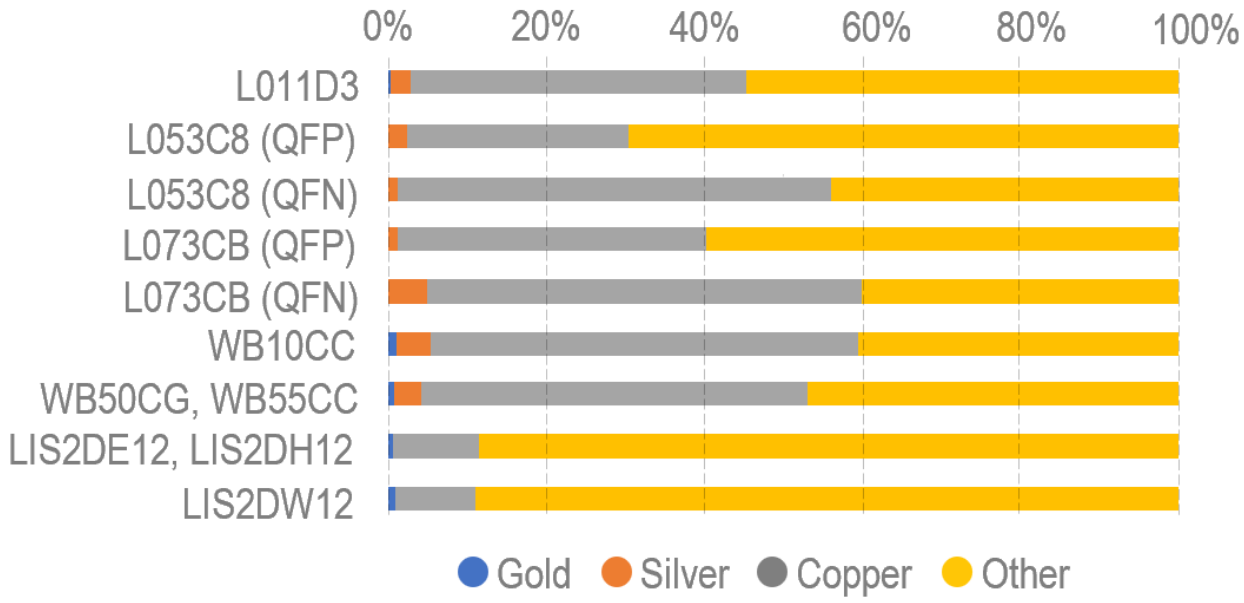


XSL•FO
RenderX

According to our preliminary estimates, BLE transceivers contribute the most to resource scarcity, primarily because of the relatively large amount of gold used in their bonding wires. A similar pattern is observed for the L011D3 microcontroller and for all accelerometers included in this analysis. Silver also contributes noticeably to the impacts associated with BLE transceivers and with all variants of the L053C8 and L073CB microcontrollers.

In contrast, copper contributes only marginally to the resource depletion results. However, its relatively high concentration in BLE transceivers and QFN-packaged microcontrollers suggests a considerable recovery potential under functional recycling scenarios, as illustrated in Figure 8.

Figure 8. Proportions of gold, silver, and copper in the electronic components that comprise the design archetypes.



Indeed, assuming a maximal functional recycling rate of 50% for gold, silver, and copper (as reported by Graedel [33]), substantial quantities of secondary copper could potentially be recovered and reintegrated into broader metallurgical production streams, as summarized in Table 8. Such material could also be reused in equivalent electronic applications following appropriate alloy adjustment processes, thereby establishing a defined lower bound for the recyclability potential of the resulting smartbands.

Table . Recovery yields of gold, silver, and copper in the studied electronic components that constitute the design archetypes (in mg).

Electronic component	Gold	Silver	Copper
L011D3	9.15E-02	6.67E-01	1.15E+01
L053C8 (QFP)	1.00E-03	2.21E+00	2.57E+01
L053C8 (QFN)	1.00E-03	5.69E-01	2.81E+01
L073CB (QFP)	6.50E-03	1.10E+00	3.65E+01
L073CB (QFN)	1.00E-03	2.53E+00	2.83E+01
WB10CC	5.27E-01	2.23E+00	2.76E+01
WB50CG, WB55CC	3.45E-01	1.70E+00	2.44E+01
LIS2DE12, LIS2DH12	2.20E-02	0.00E+00	4.34E-01
LIS2DW12	2.50E-02	0.00E+00	3.04E-01

However, realizing this recovery potential would depend on the effective implementation of closed-loop recycling strategies, as well as on the selective dismantling and segregated treatment of these components rather than their disposal within mixed electronic waste streams.

Discussion

Comparison With Prior Work

Various authors have successfully developed high-performing algorithms for wearables that maintain high accuracy even at very low sampling frequencies. For example, Straczkiewicz et al [61] developed a walking-recognition algorithm for smartwatches and wearable accelerometers capable of detecting

XSL•FO
RenderX

and tracking activity periods with high sensitivity while operating at a sampling rate of 10 Hz. Likewise, other researchers have reported advanced step-counting algorithms achieving accuracies above 90% at sampling rates of at least 8 Hz, regardless of the device's orientation [47].

The novelty of this exploratory work lies in moving beyond these algorithm-centered perspectives by focusing on the information content of the signals themselves. Specifically, we systematically quantify both the information preserved in downsampled accelerometer signals and their potential viability for step detection. By doing so, we provide designers and researchers with practical insights into how these two related aspects may influence not only the development of step-counting algorithms but also the early design and preliminary environmental assessment of frugal PA monitoring devices operating under baseline commercial constraints (ie, battery autonomy).

Our preliminary findings, based on the analysis of signals originally sampled at 30 Hz, reinforce the idea that triaxial accelerometer information remains useful for step detection during normal walking even at relatively low sampling frequencies (eg, below 15 Hz).

At the same time, our estimates suggest that signals sampled at very low frequencies (eg, below 3 Hz) retain limited information for reliable step detection and counting. Interestingly, our preliminary analysis also indicates that high frequencies (ie, above 26 Hz) may present limitations under certain conditions, particularly when individuals walk at low speeds, a tendency consistent with the observations of Fokkema et al [62]. Furthermore, when participants walk at very low speeds (2.9 or 4.3 km/h), information appears to be preserved and used more effectively within an intermediate frequency range between 11 and 25 Hz. Nevertheless, these observations should be interpreted with caution. Rather than generalizing the results across all signal axes, each axis should be considered separately, as evidenced by the percentage errors exceeding 80% for peak and valley detection along the y-axis.

Beyond signal analysis, the framework presented in Figure 6 extends previous research on medical-device design by providing an initial tool for delineating frugal smartband architectures during the early stages of product development, as recommended by Smith et al [63]. More specifically, it establishes a set of design archetypes that clarify how minimal hardware resources can be allocated according to different smartband usage requirements, consistent with the recommendations of Seva et al [64] regarding early Medical Device Readiness Levels.

From the perspective of environmental assessment within early HTA, our exploratory work offers an overview of several factors that influence the lifecycle of frugal devices. In particular, the environmental screening of the raw materials present in the electronic components that constitute the design archetypes identifies gold as a major contributor to resource depletion during production while simultaneously highlighting the potential importance of copper recovery during end-of-life management.

In addition, the framework presented in Figure 6 supports the initial formulation of preliminary operational scenarios for WAM devices based on component-level technical constraints. This capability helps reduce uncertainty regarding energy consumption, device performance, and user behavior during the early Technology Readiness Levels of new products, thereby addressing the well-recognized challenge of identifying pivotal parameters in early prototype assessments [24].

Overall, the proposed information-driven approach complements existing perspectives on the environmental assessment and design of wearable devices by providing both empirical quantification and practical decision-support tools. Specifically, the present work operationalizes the laboratory-scale assessment and use-scenario modeling activities described by Wentz et al [65] to estimate resource depletion and energy consumption at the prototype stage, thereby supporting the design of frugal smartbands through worst-case performance analyses. Similarly, this exploratory work provides an early-design instrument that facilitates the practical implementation of the broad recommendations synthesized by Gurova et al [66], particularly those related to energy efficiency and the allocation of computational and storage resources.

Limitations

It is important to acknowledge the limitations of this exploratory study, as they directly affect the interpretation and generalizability of our preliminary findings.

First, our streamlined sustainability analysis of the electronic components included in the design archetypes was limited to a single environmental impact category. As a result, this analysis provides only a partial environmental perspective on WAM devices and should not be used in isolation as a substitute for a complete LCA or HTA.

Second, the secondary data analysis used to identify the "edge zones" associated with substantial information loss in accelerometer signals was based on a relatively small dataset comprising gait tests from only 5 healthy participants, each recorded for a maximum of 120 seconds. In addition, all data considered in this exploratory study were collected using a specific research-grade WAM device worn exclusively on the wrist. Consequently, both our preliminary findings and the resulting design recommendations should be interpreted with caution, particularly when considering applications beyond the specific population, device configuration, and experimental conditions examined in this study.

Third, several important dimensions typically considered within HTA were outside the scope of this work. These include broader socioeconomic aspects such as cost-effectiveness, robustness, and ease of repair. Accordingly, our objective was not to perform a complete HTA, but rather to provide a preliminary evaluation instrument for use during the early stages of device design.

Finally, caution is warranted when relating the results of our secondary data analysis to studies that compare accelerometer signals obtained from different devices. At present, no evidence supports the use of cubic spline interpolation to resample distinct signals prior to rigorous comparative analyses. Therefore, the

XSL•FO
RenderX

findings reported here should not be interpreted as validation of such an approach in cross-device comparison studies.

Conclusions and Perspectives

This exploratory study investigated how the degradation of intrinsic information contained in accelerometer data influences the acquisition and use of knowledge and the management of uncertainty during the early-stage design and preliminary sustainability assessment of frugal WAM devices. To answer this research question, we quantified both the preservation and practical utility of information contained in degraded accelerometer signals collected at different motion velocities. This knowledge was subsequently used to examine uncertainties related to user behavior and energy consumption during the early design of WAM device archetypes.

Overall, our preliminary findings suggest that both sampling frequency and gait speed play a critical role in determining how much information can be preserved and effectively used when individuals move at normal walking speeds. Based on the patterns identified in these variables, we observed the existence of "edge zones" in which information loss negatively influences the basic mechanisms used to detect gait-related PA (ie, peak and valley counting). These zones also influence hardware trade-offs, computational load allocation, and environmental screening activities performed during the early stages of device conception.

In this manner, this exploratory study provides initial tools and practical insights that may help address some of the challenges and paradoxes that complicate HTA and LCA during the early design and development of WAM devices. Through our secondary data analysis, we extend current knowledge regarding the minimum data requirements that researchers and designers should consider when developing gait-recognition solutions under different sampling conditions.

Furthermore, the application of the design framework presented in Figure 6 provides preliminary evidence that austere hardware architectures combined with limited data inputs may still achieve acceptable operational performance, offering a potential alternative to unnecessarily complex electronic designs. Finally, the integrative framework and environmental screening proposed in this work may assist HTA practitioners and regulatory bodies in developing an initial understanding of some factors contributing to the ecological burden of WAM devices, particularly those associated with raw materials during production and end-of-life stages within a usage-oriented context.

Nevertheless, these tools and insights should be interpreted and applied cautiously. The findings of this exploratory study remain constrained by two main factors: (1) the limited scope of the environmental overview, which focused exclusively on a single environmental impact category, and (2) the relatively small dataset used in the secondary data analysis.

More broadly, this work illustrates the application of part I of our previous methodology for the sustainability assessment and development of IoMT [25]. In the present study, this methodology was applied to the specific context of wearable devices, with particular emphasis on how uncertainty and information influence the early assessment and sustainable design of smartband prototypes. The applicability of our preliminary findings to final products will be examined in part II of the methodology, which will be detailed in a forthcoming study. This future work will benefit from richer proprietary datasets generated through our ongoing investigations of unsupervised mobility, involving the assessment of average daily steps in large cohorts of participants with motor impairments (ie, more than 100 individuals) monitored through consumer-grade wearables under free-living conditions [67,68]. These datasets will enable a more robust evaluation of the preliminary observations derived from the secondary data analysis presented here. Likewise, the clinical outcomes and socioeconomic implications of frugal devices derived from the proposed design archetypes remain important topics for future investigation.

Finally, extending this line of research to other domains, such as medical imaging, represents a promising opportunity to further investigate how information influences both clinical accuracy and the environmental impact of medical devices. This is particularly relevant in fields where these outcomes depend directly on the quantity and quality of the underlying data.

Through these future investigations, we aim to gather additional evidence to examine one of the central hypotheses underpinning our broader research program that "information is the message from which a sustainable digital-based health service or paradigm can exist, and with which specific barriers for integral assessment and eco design can be surpassed" [25].

Funding

This work was partially supported by funding from the French government, managed by the National Research Agency (ANR) under the France 2030 program MIAI Cluster (ANR-23-IACL-0006) and under the "Investissements d'avenir" program (ANR-10-AIRT-05 and ANR-15-IDEX-02). This work also forms part of a broader translational and interdisciplinary GaitAlps research program (to NV).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed description of the data preparation process for the secondary analysis.

XSL•FO
RenderX

[DOCX File, 402 KB - mhealth_v14i1e88237_app1.docx]

Multimedia Appendix 2
Spread and variability of the preserved information across all speeds.
[DOCX File, 3572 KB - mhealth_v14i1e88237_app2.docx]

References

1. Ferguson T, Olds T, Curtis R, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022 Aug;4(8):e615-e626. [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
2. Li R, Li Y, Wang L, et al. Wearable activity tracker–based interventions for physical activity, body composition, and physical function among community-dwelling older adults: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2025 Apr 3;27:e59507. [doi: 10.2196/59507] [Medline: 40179387]
3. Global Action Plan on Physical Activity 2018-2030: More Active People for a Healthier World: World Health Organization; 2018.
4. Wearable technology market (2026-2033): size, share & trends analysis. : Grand View Research; 2026 Jun URL: https://tinyurl.com/3s3b7pf2 [accessed 2026-06-09]
5. Yang C, Wang B, Wan J, Kamath A, You F, Tian B. Quantifying the global eco-footprint of wearable healthcare electronics. Nature 2026 Jan;649(8095):73-82. [doi: 10.1038/s41586-025-09819-w] [Medline: 41476272]
6. Life cycle assessment for Galaxy Watches. Samsung. URL: https://www.samsung.com/global/sustainability/policy-file/AYmutOga96AALYNW/LCA_Results_for_Watch.pdf [accessed 2026-03-03]
7. Zhang T, Bainbridge A, Harwell J, Zhang S, Wagih M, Kettle J. Life cycle assessment (LCA) of circular consumer electronics based on IC recycling and emerging PCB assembly materials. Sci Rep 2024;14(1):29183. [doi: 10.1038/s41598-024-79732-1]
8. Apple Watch Series 2: environmental report. : Apple Inc; 2016 Sep URL: https://www.apple.com/environment/pdf/products/watch/Apple_Watch_Series2_PER_sept2016.pdf [accessed 2026-06-09]
9. Degroote L, Hamerlinck G, Poels K, et al. Low-cost consumer-based trackers to measure physical activity and sleep duration among adults in free-living conditions: validation study. JMIR Mhealth Uhealth 2020 May 19;8(5):e16674. [doi: 10.2196/16674] [Medline: 32282332]
10. Goh C, Wang NX, Müller AM, Yap R, Edney S, Müller-Riemenschneider F. Validation of smartphones and different low-cost activity trackers for step counting under free-living conditions. J Meas Phys Behav 2023;6(1):79-87. [doi: 10.1123/jmpb.2022-0022]
11. Chaudhry UAR, Wahlich C, Fortescue R, Cook DG, Knightly R, Harris T. The effects of step-count monitoring interventions on physical activity: systematic review and meta-analysis of community-based randomised controlled trials in adults. Int J Behav Nutr Phys Act 2020 Oct 9;17(1):129. [doi: 10.1186/s12966-020-01020-8] [Medline: 33036635]
12. Bianchini E, Galli S, Alborghetti M, et al. Four days are enough to provide a reliable daily step count in mild to moderate Parkinson's disease through a commercial smartwatch. Sensors (Basel) 2023 Nov 4;23(21):8971. [doi: 10.3390/s23218971] [Medline: 37960670]
13. Shawen N, O'Brien MK, Venkatesan S, et al. Role of data measurement characteristics in the accurate detection of Parkinson's disease symptoms using wearable sensors. J NeuroEngineering Rehabil 2020 Dec;17(1):52. [doi: 10.1186/s12984-020-00684-4]
14. Health Technology Assessment of Medical Devices: World Health Organization; 2011.
15. ISO 14040:2006 Environmental management – Life cycle assessment – Principles and framework. International Organization for Standardization. 2006 Jul. URL: https://www.iso.org/standard/37456.html [accessed 2026-06-09]
16. Gasafi E, Weil M. Approach and application of life cycle screening in early phases of process design: case study of supercritical water gasification. J Clean Prod 2011 Sep;19(14):1590-1600. [doi: 10.1016/j.jclepro.2011.05.021]
17. Graedel T. Streamlined Life-Cycle Assessment: Prentice Hall; 1998.
18. Grutters JPC, Govers T, Nijboer J, Tummers M, van der Wilt GJ, Rovers MM. Problems and promises of health technologies: the role of early health economic modeling. Int J Health Policy Manag 2019 Oct 1;8(10):575-582. [doi: 10.15171/ijhpm.2019.36] [Medline: 31657184]
19. IJzerman MJ, Koffijberg H, Fenwick E, Krahn M. Emerging use of early health technology assessment in medical product development: a scoping review of the literature. Pharmacoeconomics 2017 Jul;35(7):727-740. [doi: 10.1007/s40273-017-0509-1] [Medline: 28432642]
20. Sculpher M, Drummond M, Buxton M. The iterative use of economic evaluation as part of the process of health technology assessment. J Health Serv Res Policy 1997 Jan;2(1):26-30. [doi: 10.1177/135581969700200107] [Medline: 10180650]
21. Bhander GS, Hauschild M, McAloone T. Implementing life cycle assessment in product development. Environ Prog 2003 Dec;22(4):255-267. [doi: 10.1002/ep.670220414]
22. Chebaeva N, Lettner M, Wenger J, et al. Dealing with the eco-design paradox in research and development projects: the concept of sustainability assessment levels. J Clean Prod 2021 Jan;281:125232. [doi: 10.1016/j.jclepro.2020.125232]

XSL•FO
RenderX

23. Poudelet V, Chayer JA, Margni M, Pellerin R, Samson R. A process-based approach to operationalize life cycle assessment through the development of an eco-design decision-support system. J Clean Prod 2012 Sep;33:192-201. [doi: 10.1016/j.jclepro.2012.04.005]
24. Ni M, Borsci S, Walne S, et al. The Lean and Agile Multi-dimensional Process (LAMP) – a new framework for rapid and iterative evidence generation to support health-care technology design and development. Expert Rev Med Devices 2020 Apr;17(4):277-288. [doi: 10.1080/17434440.2020.1743174] [Medline: 32167800]
25. Quisbert-Trujillo E, Vuillerme N. Towards the operationalization of health technology sustainability assessment and the early eco design of the Internet of Medical Things. Sensors (Basel) 2025 Jun 20;25(13):3839. [doi: 10.3390/s25133839] [Medline: 40648098]
26. Homayounfar M, Malekijoo A, Visuri A, et al. Understanding smartwatch battery utilization in the wild. Sensors (Basel) 2020 Jul 6;20(13):3784. [doi: 10.3390/s20133784] [Medline: 32640587]
27. Lee J, Ko J, Choi YJ. Dhrystone million instructions per second–based task offloading from smartwatch to smartphone. Int J Distrib Sens Netw 2017 Nov;13(11):155014771774007. [doi: 10.1177/1550147717740073]
28. Quisbert-Trujillo E, Morfouli P. Using a data driven approach for comprehensive life cycle assessment and effective eco design of the Internet of Things: taking LoRa-based IoT systems as examples. Discov Internet Things 2023;3(1):20. [doi: 10.1007/s43926-023-00051-4]
29. He Y, Hosseinzadeh-Bandbafha H, Kiehbadroudinezhad M, Peng W, Tabatabaei M, Aghbashlo M. Environmental footprint analysis of gold recycling from electronic waste: a comparative life cycle analysis. J Clean Prod 2023 Dec;432:139675. [doi: 10.1016/j.jclepro.2023.139675]
30. Schwartz E, He H, Frost K, Nguyen BH, Ogunseitan OA, Schoenung JM. Comparative life cycle assessment of copper and gold recovery from waste printed circuit boards: pyrometallurgy, chemical leaching and bioleaching. J Hazard Mater 2024 Jul 15;473:134545. [doi: 10.1016/j.jhazmat.2024.134545] [Medline: 38761760]
31. Rao MD, Meshram RB, Singh KK, Morrison CA, Love JB. Life cycle analysis on sequential recovery of copper and gold from waste printed circuit boards. Waste Manag 2023 Nov 1;171:621-627. [doi: 10.1016/j.wasman.2023.10.008] [Medline: 37837909]
32. Strezov V, Zhou X, Evans TJ. Life cycle impact assessment of metal production industries in Australia. Sci Rep 2021 May 12;11(1):10116. [doi: 10.1038/s41598-021-89567-9] [Medline: 33980964]
33. Graedel TE, Allwood J, Birat JP, et al. What do we know about metal recycling rates? J Ind Ecol 2011 Jun;15(3):355-366. [doi: 10.1111/j.1530-9290.2011.00342.x]
34. Cover TM. Elements of Information Theory: John Wiley & Sons; 1999. [doi: 10.1002/047174882X]
35. Gray RM. Entropy and Information Theory: Springer; 2011. [doi: 10.1007/978-1-4419-7970-4]
36. MacKay DJC. Information Theory, Inference, and Learning Algorithms: Cambridge University Press; 2003.
37. Kvålseth T. On normalized mutual information: measure derivations and properties. Entropy 2017;19(11):631. [doi: 10.3390/e19110631]
38. Martín-Martín J, Jiménez-Partinen A, De-Torres I, et al. Reliability study of inertial sensors LIS2DH12 compared to ActiGraph GT9X: based on free code. J Pers Med 2022 May 5;12(5):749. [doi: 10.3390/jpm12050749] [Medline: 35629171]
39. Nyquist H. Certain topics in telegraph transmission theory. Trans Am Inst Electr Eng 1928;47(2):617-644. [doi: 10.1109/T-AIEE.1928.5055024]
40. Shannon CE. Communication in the presence of noise. Proc IRE 1949;37(1):10-21. [doi: 10.1109/JRPROC.1949.232969]
41. MacDougall HG, Moore ST. Marching to the beat of the same drummer: the spontaneous tempo of human locomotion. J Appl Physiol (1985) 2005 Sep;99(3):1164-1173. [doi: 10.1152/japplphysiol.00138.2005] [Medline: 15890757]
42. Congedo M, Özen C, Sherlin L. Notes on EEG resampling by natural cubic spline interpolation. J Neurother 2002 Sep;6(4):73-80. [doi: 10.1300/J184v06n04_08]
43. Rashkovska A, Trobec R, Depolli M, Avbelj V. Time and frequency analysis of non-uniform sampling. In: 33rd International Convention MIPRO: IEEE; 2010:236-240.
44. Gravetter FJ, Wallnau LB, Forzano LAB, Witnauer JE. Essentials of Statistics for the Behavioral Sciences: Cengage; 2021.
45. AN-2554: step counting using the ADXL367. Analog Devices, Inc. 2023. URL: https://www.analog.com/en/resources/app-notes/an-2554.html [accessed 2026-06-09]
46. Holm I, Fridolfsson J, Börjesson M, Arvidsson D. Fourteen days free-living evaluation of an open-source algorithm for counting steps in healthy adults with a large variation in physical activity level. BMC Biomed Eng 2023 Apr 14;5(1):3. [doi: 10.1186/s42490-023-00071-9] [Medline: 37060022]
47. Lee HH, Choi S, Lee MJ. Step detection robust against the dynamics of smartphones. Sensors 2015;15(10):27230-27250. [doi: 10.3390/s151027230]
48. Tarkoma S, Siekkinen M, Lagerspetz E, Xiao Y. Smartphone Energy Consumption: Modeling and Optimization, 1st edition: Cambridge University Press; 2014. [doi: 10.1017/CBO9781107326279]
49. Bindi B, Garofano M, Parretti C, et al. A benchmarking framework for cost-effective wearables in oncology: supporting remote monitoring and scalable digital health integration. Technologies (Basel) 2026;14(1):24. [doi: 10.3390/technologies14010024]

XSL•FO
RenderX

50. MPU-6000 and MPU-6050 product specification revision 3.4. InvenSense TDK. 2025. URL: https://invensense.tdk.com/wp-content/uploads/2015/02/MPU-6000-Datasheet1.pdf [accessed 2026-06-09]
51. LIS2DH12 MEMS digital output motion sensor: ultra-low-power high-performance 3-axis “femto” accelerometer. STMicroelectronics. 2025. URL: https://www.st.com/en/mems-and-sensors/lis2dh12.html [accessed 2026-06-09]
52. LIS2DE12 MEMS digital output motion sensor: ultra-low-power high-performance 3-axis “femto” accelerometer. STMicroelectronics. 2025. URL: https://www.st.com/en/mems-and-sensors/lis2de12.html [accessed 2026-06-09]
53. LIS2DW12 MEMS digital output motion sensor: high-performance ultralow-power 3-axis “femto” accelerometer. STMicroelectronics. 2025. URL: https://www.st.com/en/mems-and-sensors/lis2dw12.html [accessed 2026-06-09]
54. Ultra-low-power Arm Cortex-M0+ MCU with 8-Kbytes of flash memory, 32 MHz CPU. STMicroelectronics. 2025. URL: https://www.st.com/en/microcontrollers-microprocessors/stm32l011k3.html [accessed 2026-06-09]
55. Ultra-low-power Arm Cortex-M0+ MCU with 32 Kbytes of flash memory, 32 MHz CPU, USB, LCD. STMicroelectronics. 2025. URL: https://www.st.com/en/microcontrollers-microprocessors/stm32l053c6.html [accessed 2026-06-09]
56. Ultra-low-power Arm Cortex-M0+ MCU with 64-Kbytes of flash memory, 32 MHz CPU, USB, LCD. STMicroelectronics. 2025. URL: https://www.st.com/en/microcontrollers-microprocessors/stm32l073v8.html [accessed 2026-06-09]
57. Ultra-low-power dual core Arm Cortex-M4 MCU 64 MHz, Cortex-M0+ 32 MHz with 320 Kbytes of flash memory, Bluetooth LE 5.4. STMicroelectronics. 2025. URL: https://www.st.com/en/microcontrollers-microprocessors/stm32wb10cc.html [accessed 2026-06-09]
58. Ultra-low-power dual core Arm Cortex-M4 MCU 64 MHz, Cortex-M0+ 32 MHz with 1 Mbyte of flash memory, Bluetooth LE 5.4, 802.15.4, Zigbee, Thread, Matter, AES 256. STMicroelectronics. 2025. URL: https://www.st.com/en/microcontrollers-microprocessors/stm32wb50cg.html [accessed 2026-06-09]
59. Ultra-low-power dual core Arm Cortex-M4 MCU 64 MHz, Cortex-M0+ 32 MHz with 1 Mbyte of flash memory, Bluetooth LE 5.4, 802.15.4, Zigbee, Thread, Matter, USB, LCD, AES-256. STMicroelectronics. 2025. URL: https://www.st.com/en/microcontrollers-microprocessors/stm32wb55cg.html [accessed 2026-06-09]
60. Andreasi Bassi S, Biganzoli F, Ferrara N, et al. Updated characterisation and normalisation factors for the Environmental Footprint 3.1 method. : Publications Office of the European Union; 2023. [doi: 10.2760/798894]
61. Straczkiewicz M, Huang EJ, Onnela JP. A “one-size-fits-most” walking recognition method for smartphones, smartwatches, and wearable accelerometers. NPJ Digit Med 2023 Feb 23;6(1):29. [doi: 10.1038/s41746-022-00745-z] [Medline: 36823348]
62. Fokkema T, Kooiman TJM, Krijnen WP, Van Der Schans CP, De Groot M. Reliability and validity of ten consumer activity trackers depend on walking speed. Med Sci Sports Exerc 2017;49(4):793-800. [doi: 10.1249/MSS.0000000000001146]
63. Smith V, Warty R, Nair A, et al. Defining the clinician’s role in early health technology assessment during medical device innovation – a systematic review. BMC Health Serv Res 2019 Jul 23;19(1):514. [doi: 10.1186/s12913-019-4305-9] [Medline: 31337393]
64. Seva RR, Tan ALS, Tejero LMS, Salvacion MLDS. Multi-dimensional readiness assessment of medical devices. Theor Issues Ergon Sci 2023 Mar 4;24(2):189-205. [doi: 10.1080/1463922X.2022.2064934]
65. Wentz F, Mohammadi M, Tybrandt K, Berggren M, Arvidsson R, Rahmanudin A. Integrating environmental assessment into early-stage wearable electronics research. J Mater Chem C 2025 Oct 9;13(39):19983-19999. [doi: 10.1039/D5TC02280K]
66. Gurova O, Merritt TR, Papachristos E, Vaajakari J. Sustainable solutions for wearable technologies: mapping the product development life cycle. Sustainability 2020;12(20):8444. [doi: 10.3390/su12208444]
67. Bianchini E, Garramone F, Rinaldi D, et al. Does cognition affect supervised and unsupervised mobility differently in people with Parkinson’s disease? A cross-sectional study. J NeuroEngineering Rehabil 2026;23(1):71. [doi: 10.1186/s12984-025-01873-9]
68. Bianchini E, Alborghetti M, Galli S, et al. Minimal clinically important difference of average daily steps measured through a consumer smartwatch in people with mild-to-moderate Parkinson disease: cross-sectional study. JMIR Mhealth Uhealth 2025 Jul 29;13:e64213. [doi: 10.2196/64213] [Medline: 40734369]

Abbreviations

BLE: Bluetooth Low Energy
FIFO: first-in-first-out
HTA: health technology assessment
IoMT: Internet of Medical Things
LCA: Life Cycle Assessment
MCU: microcontroller
PA: physical activity
WAM: wearable activity monitoring

XSL•FO
RenderX

Edited by G Badicu; submitted 21.Nov.2025; peer-reviewed by EE Gaona, P Singh; revised version received 05.Jun.2026; accepted 07.Jun.2026; published 31.Jul.2026.

Please cite as:
Quisbert-Trujillo E, Vuillerme N
An Information-Driven Approach for the Early Health Technology Sustainability Assessment and the Frugal Design of the Internet of Medical Things: Exploratory Study of Wearable Activity Monitoring Devices
JMIR Mhealth Uhealth 2026;14:e88237
URL: https://mhealth.jmir.org/2026/1/e88237
doi:10.2196/88237

© Ernesto Quisbert-Trujillo, Nicolas Vuillerme. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 31.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Noninvasive Interstitial Glucose Estimation Using Wearables and Machine Learning in Healthy Individuals and Individuals With Obesity: Observational Cohort Study

Franziska Schmelter[1,2*], Dr rer nat; Xinyu Huang[3*], Dr-Ing; Annika Heidmann[3,4], BSc; Paul Beier[1,2], MSc; Friederike Miriam Rundfeldt[1,2], MD; Christina Plöger[1,2], MD; Artur Piet[3,4], MSc; Md Abid Hasan[4,5], MSc; Yuanheng Zhang[4], MSc; Lennart Jablonski[4], MSc; Oliver Witt[6], Dr rer nat; Torsten Schröder[1,2,6], Dr rer nat Dr med; Stefanie Derer[1,2], Prof Dr rer nat; Chris Easthope Awai[7], PhD; Marcin Grzegorzek[4,5*], Prof Dr-Ing; Christian Sina[1,2,8*], Prof Dr med

[1]Institute of Nutritional Medicine, University of Luebeck, Ratzeburger Allee 160, Luebeck, Schleswig-Holstein, Germany

[2]University Hospital Schleswig-Holstein, Luebeck, Schleswig-Holstein, Germany

[3]expandAI GmbH, Luebeck, Schleswig-Holstein, Germany

[4]Institute of Medical Informatics, University of Luebeck, Luebeck, Schleswig-Holstein, Germany

[5]German Research Center for Artificial Intelligence (DFKI), Luebeck, Schleswig-Holstein, Germany

[6]Research & Development, Perfood Laboratories GmbH, Luebeck, Schleswig-Holstein, Germany

[7]Data Analytics & Rehabilitation Technology (DART), Lake Lucerne Institute, Vitznau, Switzerland

[8]Khalifa University, Abu Dhabi, United Arab Emirates

[*]these authors contributed equally

Corresponding Author:
Franziska Schmelter, Dr rer nat
Institute of Nutritional Medicine, University of Luebeck, Ratzeburger Allee 160, Luebeck, Schleswig-Holstein, Germany

Abstract

Background: Continuous glucose monitoring (CGM) can facilitate weight management and lower the risk of metabolic diseases by providing real-time feedback on glycemic responses, thereby enabling more informed lifestyle decisions. However, current CGM systems remain constrained by invasiveness, cost, and short sensor lifespan, limiting their practicality for guiding individualized postprandial low-glycemic diets.

Objective: Extending earlier proof-of-concept findings, this study aimed to validate an interstitial glucose (IG) machine learning algorithm in real-world environments using multimodal, continuous data collected from wearable sensors and smartwatches. In the long term, we aim to embed a noninvasive, sensor-based algorithm for estimating tissue glucose within mobile health apps and postprandial low-glycemic diet frameworks to enable scalable, personalized prevention strategies.

Methods: We conducted a 2-week study phase during which participants continuously wore 2 noninvasive sensor devices: a scientific sensor wristband (Empatica EmbracePlus) and a commercially available smartwatch (Fitbit Sense 2). As a reference measurement, an invasive CGM sensor (Abbott FreeStyle Libre 3) measured IG levels. For metabolic characterization, 1-point fasting blood and urine samples were collected, and deep phenotyping using state-of-the-art nuclear magnetic resonance spectroscopy and bioelectrical impedance analysis was performed. For participants with overweight and obesity, clinical standard parameters focusing on glucose metabolism were analyzed.

Results: A total of 74 participants, 34 (46%) healthy controls and 40 (54%) metabolically at-risk (MR) individuals, simultaneously used an invasive CGM device together with 2 noninvasive wristbands over 2 weeks. Healthy controls were characterized by a mean age of 24.53 (SD 3.68) years and a mean BMI of 22.36 (SD 2.16) kg/m². In contrast, the MR cohort had a mean age of 55.38 (SD 15.08) years and a mean BMI of 35.38 (SD 4.91) kg/m². Furthermore, the MR cohort showed elevated fasting glucose levels (mean 107.56, SD 19.26 mg/dL), hemoglobin A_{1c} levels (mean 5.67%, SD 0.63%), and an increased homeostatic model assessment of insulin resistance index (mean 4.34, SD 3.49), indicating a disturbed glucose metabolism. The proposed long short-term memory network based on feature vectors obtained the best IG prediction performance, with an average root-mean-squared error of 21.04 (SD 8.32) mg/dL and 98.3% of predictions in zones A and B of the Clarke error grid analysis, representing a high level of predictive accuracy. In addition, based on data from a smartwatch, we achieved a comparable average root-mean-squared error of 23.49 (SD 11.46) mg/dL for the overall cohort.

Conclusions: This study demonstrates that IG levels can be predicted from multimodal, noninvasive wearable sensor data using a machine learning approach under real-world conditions. While further validation in larger and more diverse cohorts is warranted,

XSL•FO
RenderX

this approach represents a promising step toward accessible, personalized glycemic monitoring and dietary guidance as a preventive tool in mobile health apps.

(JMIR Mhealth Uhealth 2026;14:e91724) doi:10.2196/91724

KEYWORDS

continuous glucose monitoring; wearables; interstitial glucose; machine learning algorithm; personalized nutrition

Introduction

Maintaining glucose homeostasis is beneficial not only for patients with type 2 diabetes (T2D) but also for healthy individuals. Stable blood glucose (BG) levels are crucial for body weight control and preventing risks such as metaflammation, insulin resistance, and associated conditions such as cardiovascular diseases, cancer, acne, migraine, and neurodegenerative diseases, indicating a wide range of applications [1-3].

A high postprandial glucose response (PPGR) can lead to postprandial hyperinsulinemia, glucotoxicity, lipogenesis, and visceral obesity [4]. Although diet is a major determinant of PPGR, standardized dietary recommendations frequently fail in practice, as eating behavior is strongly influenced by individual preferences, taste, cultural patterns, and daily routines. These factors make dietary responses and adherence highly variable. Personalized nutrition addresses this challenge by tailoring guidance to each individual's unique habits and preferences rather than imposing a uniform diet. Previous studies have shown promising results for personalized diets in avoiding glucose spikes, thereby addressing the individual nature of PPGRs [5,6]. As an example, digital postprandial low-glycemic diets (PLGD) based on continuous glucose monitoring (CGM) have shown effectiveness for individuals with prediabetes or diabetes and migraine [7,8].

While CGM is commonly used to measure interstitial glucose (IG) levels [9], it has limitations, including invasiveness, inconvenience, high cost, limited sensor lifespan, and environmental impact [10]. Noninvasive sensor technologies offer a potential solution for real-time calculation of glucose levels through longitudinal monitoring of physiological parameters, thereby providing a foundation of individualized, adaptive, and dynamic day-to-day recommendations. Early research has examined the influence of physiological parameters on food intake and the associated development of obesity [11]. Some approaches focus on the development of noninvasive continuous glucose monitors based on optoacoustic signals [12] or microwave technology [13,14], while others seek to use existing sensor systems. Nowadays, a wide variety of wearable devices capable of continuously monitoring diverse biosignals are commonly used in everyday life (eg, smartphones, smartwatches, and fitness trackers). When combined with modern machine learning (ML) algorithms, these technologies offer the potential to elucidate relationships between physiological parameters and food intake [15], thereby enabling personalized dietary guidance. Initial data demonstrate the ability to predict IG values in real time using physiological parameters and dietary information [16-18]. However, these algorithms depend on the manual entry of daily dietary information, a process that demands considerable user adherence and is inherently time-consuming, subjective, and susceptible to error. In prior research, we demonstrated the feasibility of predicting IG levels using data from noninvasive sensors coupled with ML algorithms without using food logs. This approach yielded robust results in young, healthy individuals when assessed in controlled settings using standardized test meals, including oral glucose tolerance tests and mixed-meal tests [19].

Given the current constraints of invasive CGM and food diary–based PPGR prediction models, there is a significant opportunity to expand the availability of digital PLGD solutions. Particularly in the context of mobile health (mHealth), calculating BG levels based on noninvasive sensor data using ML algorithms represents a significant milestone. The availability of wearables such as smartwatches, smart rings, and wristbands is steadily increasing, and using these data, BG levels and trends could be calculated in the future. In combination with low-level, user-friendly documentation tools based on photos and voice recordings, individual dietary habits could be easily tracked. A corresponding mHealth tool could then provide PLGD recommendations based on individual predictions of food reactions. This scalable mHealth tool could make PLGD accessible to the global population and, therefore, become a key application for prevention and health management in the future.

To establish the foundation of this mHealth app, address the challenges of current CGM devices, and adapt our algorithm to realistic, real-world use, we conducted a 2-week study under everyday conditions with a habitual diet. During this study, we analyzed CGM data of both healthy individuals and metabolically at-risk (MR) individuals (with overweight or obesity). We included MR participants with disturbances in glucose metabolism and insulin resistance, potentially generating aberrant IG signals, to challenge our algorithm and evaluate its performance with more complex data. For noninvasive data collection, we used 2 devices: a clinically designed sensor wristband and a commercially available smartwatch. The study had 3 aims. First, we evaluated the ML algorithm to calculate IG values under real-world conditions involving complex, mixed meals. Second, we examined its generalizability across healthy and MR individuals based on the assumption that insulin resistance may alter physiological signals and reduce predictive accuracy. Third, we assessed device independence by implementing a proof-of-concept version on a commercially available smartwatch to determine whether the algorithm maintains its performance across different wearable platforms.

XSL•FO
RenderX

Methods

Ethical Considerations

The study examined healthy adults and adults with overweight. The study was approved by the Ethics Committee of the University of Luebeck (Az 2022 - 550 and 2024 - 363) according to Good Clinical Practice. Potential participants were recruited through posters, social media, and newspaper articles in northern Germany. The inclusion and exclusion criteria were deliberately broad, encompassing adults with sufficient German language proficiency to comprehend the study materials and able to operate a smartphone. Participants in the MR cohort were either overweight with at least one known metabolic comorbidity, such as hypertension, dyslipidemia, atherosclerosis, insulin resistance, gout, stroke, or heart failure, or classified as obese. All participants provided written informed consent prior to their inclusion in the study. As part of the consent process, the participants were informed about the data protection measures, the voluntary nature of study participation, and their right to withdraw from the study at any time without giving a reason and without any disadvantages. In accordance with applicable data protection regulations, the participants' data were deidentified. The data were analyzed in pseudonymized form solely for the purposes of the study. No compensation was paid for participating in this study.

Study Cohort and Design

In total, 74 adults (women: n=54, 73%; and men: n=20, 27%), aged between 19 and 84 years, participated in this pilot study using an observational cohort study design. Participants were asked to wear 2 sensor wristbands as well as a CGM sensor on the upper arm continuously for 14 days. Otherwise, participants were expected to behave as usual and received no further instructions other than being asked to charge their wristbands regularly. During a clinic visit, blood and urine samples were collected, and a bioelectrical impedance analysis (BIA) measurement was performed.

In more detail, noninvasive wearable data from a scientific high-end device (Empatica EmbracePlus) and a commercially available smartwatch (Fitbit Sense 2) as well as IG data from an invasive CGM device (Abbott FreeStyle Libre 3) were collected over 14 days. First, Food and Drug Administration–confirmed noninvasive Empatica EmbracePlus measured skin temperature (STEMP) at 1 Hz, electrodermal activity (EDA) at 4 Hz, triaxial acceleration (taACC) at 32 Hz, sleep assessment (Sleep) at 0.017 Hz, and photoplethysmography (PPG), which produces a blood volume pulse (BVP) to calculate the interbeat interval in seconds and the pulse rate (Plr=60 seconds/interbeat interval) at 0.017 Hz. Second, Fitbit Sense 2 (Fitbit) quantified Plr at 0.2 Hz. Third, as reference values for our algorithm, continuous IG levels were measured by Food and Drug Administration–approved CGM devices every 5 minutes with a mean absolute relative difference of 7.8% [20]. In addition, participants kept a self-guided food diary with the help of an app (Perfood) to document their normal eating habits, activity, and sleep. However, these data were not included in this study, as the primary objective was to develop and optimize the algorithm itself, while the derivation of personalized nutritional recommendations will be addressed in a subsequent step in the future. Furthermore, study participants were deeply phenotyped. First, body composition was determined using BIA (Seca). Second, 1-point fasting blood and urine samples were collected to analyze individual metabolic profiles using nuclear magnetic resonance (NMR) spectroscopy. Third, for MR participants, characterized by overweight and obesity, clinical standard parameters were quantified, including complete blood count, lipid profiles, and liver parameters (glutamate-oxaloacetate transaminase, glutamate-pyruvate transaminase, and gamma-glutamyl transferase). Furthermore, glucose metabolism was characterized in detail by analyzing fasting glucose, insulin, homeostatic model assessment of insulin resistance, and hemoglobin A_{1c} (HbA_{1c}) levels to identify potential disorders indicating a prediabetic or diabetic status of the individuals.

NMR Spectroscopy

Serum samples were thawed at room temperature and prepared according to the standard operating procedure described by Dona et al [21]. In brief, samples were homogenized 1:1 (v/v) with a phosphate buffer (75 mM, pH 7.4) and cooled to 279 K in 5 mm NMR tubes prior to measurement. Following daily quality control procedures, including water suppression, quantification, and temperature checking, the samples were analyzed according to Bruker's in vitro diagnostic research protocol using a 600 MHz Avance III HD NMR spectrometer at 310 K. A 1D nuclear Overhauser effect spectroscopy experiment (pulse program: noesygppr1d) and a 1D Carr-Purcell-Meiboom-Gill spin-echo experiment (pulse program: cpmgpr1d) were acquired for each sample, and 39 metabolites (plus 2 technical additives) as well as 112 lipoproteins were quantified using Bruker's Quantification in Plasma/Serum (B.I.Quant-PS 2.0.0) and Bruker's in vitro diagnostic research Lipoprotein Subclass Analysis (B.I.-LISA) software. For lipid profiling, triglycerides, phospholipids, cholesterol, free cholesterol, and apolipoproteins were analyzed in very-low-density lipoprotein (VLDL), intermediate-density lipoprotein (IDL), low-density lipoprotein, and high-density lipoprotein (HDL; Tables S1 and S2 in Multimedia Appendix 1). Although urine samples were collected and analyzed according to standard procedures [21], these findings are not discussed further because the focus of this study is on BG outcomes.

Data Preprocessing

To facilitate rapid processing by the ML model, 4 essential preprocessing steps were implemented.

Data Formatting

In this study, 3 data formats were obtained from the different wearables: AVRO, CSV, and JSON. The CGM device provided IG responses (in mg/dL) in CSV format. EmbracePlus recorded daily physiological data in AVRO and CSV formats, including sensor recordings, time stamps, and frequency rates. To ensure accurate time information during preprocessing, Unix time stamps in GMT were computed and saved with each recording. The recorded value time stamps were calculated using frequency conversion from the initial sensor time stamp. Data collection

XSL•FO
RenderX

temporal variations were observed when daily recording start and end times were 15 minutes outside the midnight to 11:59 PM range. Fitbit sensor data were stored in JSON files containing GMT time stamps. Weight, height, age, sex, and BMI were used as demographic variables. The files were merged and converted to NumPy files for data analysis.

Data Cleaning

In accordance with standard procedures, the first and last days of CGM data collection were excluded to ensure proper sensor calibration and stabilization. Additionally, variations between participants were observed due to differences in the timing of the initial application and removal of the devices. As a result, our subsequent analyses were conducted using a dataset that comprised 12 of the 14 study days. In consideration of the temporal variability in data collection commencement and conclusion times, we opted to commence our analysis at 1 AM on the second day of the study and conclude it at 11 AM on the 13th day of the study. Afterward, we applied participant-specific outlier detection for artifact removal, that is, values more than 4 SDs from the individual mean were marked as invalid and substituted with the local mean, resulting in the removal of less than 0.2% of data points while preserving physiological validity and temporal continuity.

Data Alignment

To properly align the data, missing values, a fluctuating timespan between timesteps, and varying recording frequencies from the sensors had to be compensated for. To address this issue, an artificial timestep array was created. The sampling rate for the features or the prediction horizon (PH) for the labels determined the equidistant timesteps between the fixed start and end times that constrained this array. The primary objective of this approach was to standardize the temporal representation of available data and establish a coherent representation spanning from the start to the end. The next step was to select the data from the sensors that fell within the predefined period. The sensor modality data were then interpolated with a common sampling rate and their respective time information to match the predefined timesteps.

To accommodate the data characteristics of each sensor modality, a customized interpolation approach was adopted. For sensor signals that capture continuous numerical data, such as EDA, taACC, temperature, and plr, linear interpolation (LI) was implemented, with its mathematical formulation provided in equation 1. It estimates intermediate values by constructing a linear polynomial that passes through 2 consecutive known data points, denoted as (x_0, y_0) and (x_1, y_1). This process yields a piecewise linear representation that approximates the inherent continuous trend of the signal. In this context, ξLI(x) is the first-order polynomial interpolation operator. Given its modest computational overhead, LI remains a prevalent choice for inferring missing intermediate data points in the dataset.

(1)ξLI(x)=y0(x−x1x0−x1)+y1(x−x0x1−x0)

Notably, LI exhibits discontinuities in smoothness at subinterval boundaries, primarily due to its nondifferentiable nature at these junction points. To address this limitation, we evaluated alternative interpolation techniques for glucose-labeled datasets, where PH was used to define the interval between consecutive data points. One such method tested was nearest neighbor interpolation provided in equation 2. This approach assigns each new data point the value of its nearest existing data point, eliminating the need to compute intermediate values or derivatives. As a result, it retains the step-wise characteristic inherent to the original glucose data.d=argmini|x - xi|

(2)εNNI(x)=yd, d=arg mini|x−xi|

Here, (xi,yi) is a set of discrete data points for i=0, 1, 2, ... , n. ξNNI(x) represents the interpolated value at any point x. argmini|x-xi| denotes the index d for which the distance |x-xi| is the smallest.

Data Segmentation

We used a nonoverlapping window segmentation to segment the continuous multivariate time series into samples aligned with the PH, denoted as p. The source data were collected from multiple smartwatch sensors S = [s0, s1, ... , sn-1] with associated time stamps T = [t0, t1, ... , ti-1], which were divided into contiguous windows of length p minutes. Each window kx is constructed to encompass the sensor readings from the time interval [tωx-p, tωx] that immediately precedes the corresponding label ωx. This yields a set of windows K = [k0, k1, ..., kl-1] equal in number to the labels Ω=[ω0,ω1,...,ωl-1]. Consequently, for a fixed observation period, the number of training samples |K| decreases as p increases. This segmentation provides the foundational structure for subsequent supervised learning. All segmentation procedures were executed independently for each participant, ensuring that window construction and temporal indexing remained confined to each individual's sensor stream and label trajectory.

Feature Engineering

The complex physiological processes underlying glucose dynamics present a challenge for predicting BG levels [22]. Nevertheless, previous research identified correlations associated with specific extracted features and established the impact of personal demographics on BG responses [23,24]. For each nonoverlapped PH (PH=30 min), handcrafted features were extracted for all sensor modalities, considering their temporal characteristics [25]. In respect thereof, the sensor data were segmented into consecutive 1-minute windows. For every subwindow, a rich set of handcrafted features was computed for each available sensor signal as detailed in Table S3 in Multimedia Appendix 1. These included statistical and time domain descriptors, frequency and amplitude domain measures, and morphological indicators. Temporal context features were added to capture circadian and seasonal regularities. Demographic and anthropometric variables were incorporated to account for interindividual physiological variability. After feature extraction, all variables were concatenated across modalities and time windows to form the final multivariate input representation for the long short-term memory (LSTM) model. In total, 5 contextual and demographic features and 22 features per sensor modality were retained after preprocessing, yielding 159 features per minute per 30-minute observation window for the Empatica dataset and 27 features for the Fitbit dataset. No dimensionality reduction, embedding, or feature selection was applied, as the LSTM architecture was designed to learn

XSL•FO
RenderX

hierarchical temporal representations directly from the complete feature space. For more details on the feature engineering standards, see a study by Huang et al [19].

Model Characteristics

After data preprocessing and feature engineering, we fed the reformulated sensor data (feature vectors) to the LSTM network, which is an enhanced type of recurrent neural network designed to overcome issues such as vanishing gradients and gradient explosions in traditional recurrent neural networks. Using recurrent connections over time, LSTM is capable of capturing and maintaining dynamic temporal information. A key feature of LSTM cells is the use of 3 gating mechanisms, called input gate, forget gate, and output gate, which collectively regulate the flow of information. They use adaptive weights and activation functions to control whether information is retained, updated, or discarded. Specifically, the forget gate uses a sigmoid function to determine whether information from the previous hidden state should be dropped. The input gate, which combines both sigmoid and *tanh* functions, drives which parts of the new input should be stored in the cell state. The output gate then generates the new hidden state based on the updated cell state. In such a manner, the cell state functions as long-term memory, carrying stable information across time steps, while the hidden state serves as short-term memory.

In our implementation, the input to the LSTM consists of 2D sensor data (timesteps×features). The model architecture comprises 4 stacked LSTM layers with progressively decreasing hidden sizes, allowing the network to hierarchically extract temporal representations at different levels of abstraction. After the recurrent layers, the output of the last time step is passed through 2 fully connected layers with a nonlinear *tanh* activation in between, before being projected to the final scalar prediction. This architecture was determined empirically based on preliminary experiments and prior work on glucose forecasting tasks using physiological sensor data [19,26]. Hyperparameters, including the learning rate and hidden layer dimensions, were selected via a grid search performed on the validation partition of each leave-one-subject-out (LOSO) fold to optimize regression performance. This design enables the model to capture complex sequential dependencies while maintaining a compact representation for the final regression output.

Normalization and Data Leakage Prevention

The prediction task was modeled as a continuous regression problem. To stabilize training and ensure comparability across participants, both the input features and the target glucose values were standardized using fold-specific StandardScaler objects, which were fitted exclusively on the training participants within each LOSO fold. For the target variable, the mean and SD were estimated from the training participants and subsequently applied to the held-out participant. Likewise, each feature dimension was scaled using a StandardScaler fitted on the training data, with the resulting transformation applied to the test participant without incorporating any statistics from the held-out participant. All preprocessing operations that depend on the underlying data distributions, including feature and label standardization, were strictly confined to the training partition of each fold, thereby eliminating any risk of leakage of global statistics into the test data. This procedure ensures that both the features and labels of the held-out participant remain completely independent of the training data, while maintaining stable distributions for LSTM model training and subsequent regression performance evaluation.

Training Setup

The model was trained for 70 epochs with a batch size of 32 using PyTorch 2.7.1+cu128 in Python 3.11.13 on an NVIDIA RTX 5090 GPU with 32 GB RAM. The optimizer used was Adam (learning rate=0.001), while model optimization was guided by the mean absolute error (MAE; L1 loss), chosen for its stability and robustness to non-Gaussian error distributions in continuous glucose regression. During training, the model output remained in standardized units. Before evaluation, predictions (and targets when required) were transformed back to the original glucose scale using the inverse StandardScaler parameters of the respective LOSO fold to ensure clinically interpretable metrics.

Quantitative and Qualitative Evaluation

Clinical characteristics were compared between groups using nonparametric Mann-Whitney *U* tests. For BIA measurement and metabolic profiling, a correction for multiple testing using the method of Benjamini et al [27] with a false discovery rate of 1% was performed. Furthermore, linear and multiple linear regression analyses were conducted.

To properly evaluate performance, model effectiveness was examined across different PHs (15, 30, and 60 min) using the proposed LSTM and inverted transformer (iTransformer) models, based on the combined dataset from Empatica and Fitbit wearables. As a benchmark baseline, the iTransformer inverts the architecture by embedding the entire historical time series of each individual variable into an independent token. In detail, the model leverages a multilayer perceptron to transform the temporal sequence of each sensor modality into a high-dimensional vector. The self-attention mechanism is then applied across these variable tokens to capture multivariate correlations, while layer normalization and feed-forward networks characterize the temporal dynamics of each individual variable. This design allows the iTransformer to effectively learn complex intervariable dependencies from heterogeneous multisource wearable data, making it a highly robust benchmark for our study [28]. Algorithm model performance was evaluated using commonly used metrics: MAE, mean absolute percentage error (MAPE), and root-mean-squared error (RMSE). MAE quantifies the average absolute magnitude of errors to reflect overall accuracy, while MAPE measures the percentage deviation to provide a scale-independent metric for relative accuracy. To heavily penalize larger outliers, the mean squared error was computed to assess the average of squared differences. Finally, its square root, RMSE, was reported to express this error in the original measurement units for a more intuitive clinical interpretation. To qualitatively evaluate the model predictions, we performed Clarke error grid analysis (CEGA) [29], in which predefined zone A as clinically accurate predictions that would lead to correct treatment decisions, zone B as deviations that would not result in harmful clinical actions, zones C and D as errors potentially leading to unnecessary

XSL•FO
RenderX

overtreatment and failure to treat, respectively, and zone E as misguidance that could cause completely opposite and dangerous interventions.

CI Calculation

To ensure consistency in statistical comparisons across different cohorts and model conditions, we calculated 95% CIs for all RMSE and MAE metrics. These were derived using a nonparametric bootstrapping method with 1000 resamples. For each cross-validation fold, bootstrap samples were generated by conducting with-replacement random sampling from the model predictions and true glucose values, while maintaining independence at the individual level. In general, SE was calculated as SD/n with n as the number of LOSO folds to reflect the precision of the mean estimate, while 95% CI for RMSE and MAE means were derived from 1000 nonparametric bootstrap resampling, that is, 2.5th and 97.5th percentiles of bootstrap mean distributions. Thereby, narrow CI widths (consistent with SE×2) confirm robust mean performance estimates, even with moderate within-cohort variability. This procedure was consistently applied across all cohorts to ensure statistical comparability.

Ablation Study of Sensor Modalities

To quantify the individual and combined contribution of each sensor modality to IG prediction performance, a systematic ablation study was implemented. The study was performed on the combined cohort of healthy control (HC) and MR participants using the same preprocessed dataset, LSTM model architecture, training protocol, and LOSO cross-validation strategy as the primary analysis. First, single-modality ablation was conducted by independently removing EDA, PPG, STEMP, and taACC from the feature space. Second, combinatorial ablation tests were performed by removing pairs of sensor modalities to evaluate correlations between modalities. All other features, including demographic variables, temporal context features, and the remaining sensor modalities, were retained without modification. Model performance was evaluated using RMSE, MAE, MAPE, and CEGA to quantify the degradation in predictive accuracy caused by the exclusion of each modality or modality combination, relative to the full-feature model containing all sensor inputs. This approach allowed us to isolate the unique and synergistic predictive value of each physiological signal source for glucose estimation.

Shapley Additive Explanations Analysis

Shapley Additive Explanations (SHAP) analysis can be used to estimate the marginal importance of input features across all sensor modalities by implementing statistical analysis on the contributions of features, thereby identifying the key physiological mechanisms influencing model predictions. Within the proposed LOSO cross-validation framework, local SHAP values for all features extracted from PPG, EDA, STEMP, and taACC were calculated for each test sample. Following this, the local attribution findings were integrated to generate a global feature importance ranking, demonstrating the impact of each sensor-derived feature on IG prediction. By validating the physiological significance of the hypotheses in the model, we demonstrate the dominant and complementary roles of each modality and ensure that all predictions are based on clinically relevant signal patterns.

Results

Study Participant Characteristics

A total of 74 participants participated in our study, including 54 (73%) women and 20 (27%) men. Of these, 34 (46%) individuals were assigned to the HC cohort and 40 (54%) to the MR cohort. The MR cohort comprised individuals who were either overweight and exhibited at least one comorbidity or who were classified as obese. In more detail, individuals with a BMI between 25 and 29.9 kg/m^2 (overweight) were only included in the MR group if they presented with at least one known metabolic comorbidity, such as hypertension, dyslipidemia, atherosclerosis, insulin resistance, gout, stroke, or heart failure. In contrast, individuals with obesity (BMI ≥30 kg/m^2) were assigned to the MR cohort regardless of comorbid conditions. As a result, 2 cohorts were formed that exhibited significant differences in their physiological parameters (Table 1).

XSL•FO
RenderX

Table . Study cohort characteristics[a].

Characteristics	HC[b] cohort (n=34)	MR[c] cohort (n=40)	*P* value
Sex, n (%)			
Female	25 (74)	29 (73)	—[d]
Male	9 (26)	11 (27)	—
Age (years), mean (SD)	24.53 (3.68)	55.38 (15.08)	<.001
BMI (kg/m^2), mean (SD)	22.36 (2.16)	35.38 (4.91)	<.001
Overweight, n (%)	4 (12)	8 (20)	—
Obesity class 1, n (%)	—	12 (30)	—
Obesity class 2, n (%)	—	14 (35)	—
Obesity class 3, n (%)	—	6 (15)	—
Waist-to-hip ratio, mean (SD)	0.81 (0.08)	0.90 (0.17)	<.001
FM[e] (kg), mean (SD)	16.03 (4.49)	45.24 (12.31)	<.001
FM (%), mean (SD)	24.16 (5.55)	44.81 (7.61)	<.001
FMI[f] (kg/m^2), mean (SD)	5.45 (1.46)	15.93 (4.32)	<.001
SMM[g] (kg), mean (SD)	23.48 (4.82)	26.47 (6.33)	.045
PA[h] (°), mean (SD)	5.37 (0.53)	5.00 (0.56)	.01
FFM[i] (kg), mean (SD)	50.16 (8.67)	54.91 (11.29)	*.08*[j]
FFM (%), mean (SD)	75.84 (5.55)	55.19 (7.61)	<.001
FFMI[k] (kg/m^2), mean (SD)	16.93 (1.75)	19.05 (1.68)	<.001
VAT[l] (L), mean (SD)	0.49 (0.72)	4.21 (2.48)	<.001
WC[m] (m), mean (SD)	0.73 (0.09)	1.11 (0.13)	<.001
TBW[n] (L), mean (SD)	36.78 (6.11)	41.28 (8.31)	.01
TBW (%), mean (SD)	55.35 (3.89)	41.21 (5.29)	<.001
ECW[o] (L), mean (SD)	15.52 (2.37)	18.84 (3.52)	<.001
ECW (%), mean (SD)	23.38 (1.64)	18.83 (1.99)	<.001
ECW/TBW (%), mean (SD)	42.29 (1.67)	45.85 (2.15)	<.001

[a]Bioelectric impedance analysis (BIA) results were shown for 33 HC and 35 MR individuals, while 1 BIA measurement for HC and 5 measurements for MR were missing. Comparison between metadata and BIA results was performed using multiple Mann-Whitney *U* tests with a false discovery rate (FDR) of 1%.

[b]HC: healthy control.

[c]MR: metabolically at-risk.

[d]Not applicable.

[e]FM: fat mass.

[f]FMI: fat mass index.

[g]SMM: skeletal muscle mass.

[h]PA: phase angle.

[i]FFM: fat-free mass.

[j]Nonsignificance after FDR correction.

[k]FFMI: fat-free mass index.

[l]VAT: visceral fat.

[m]WC: waist circumference.

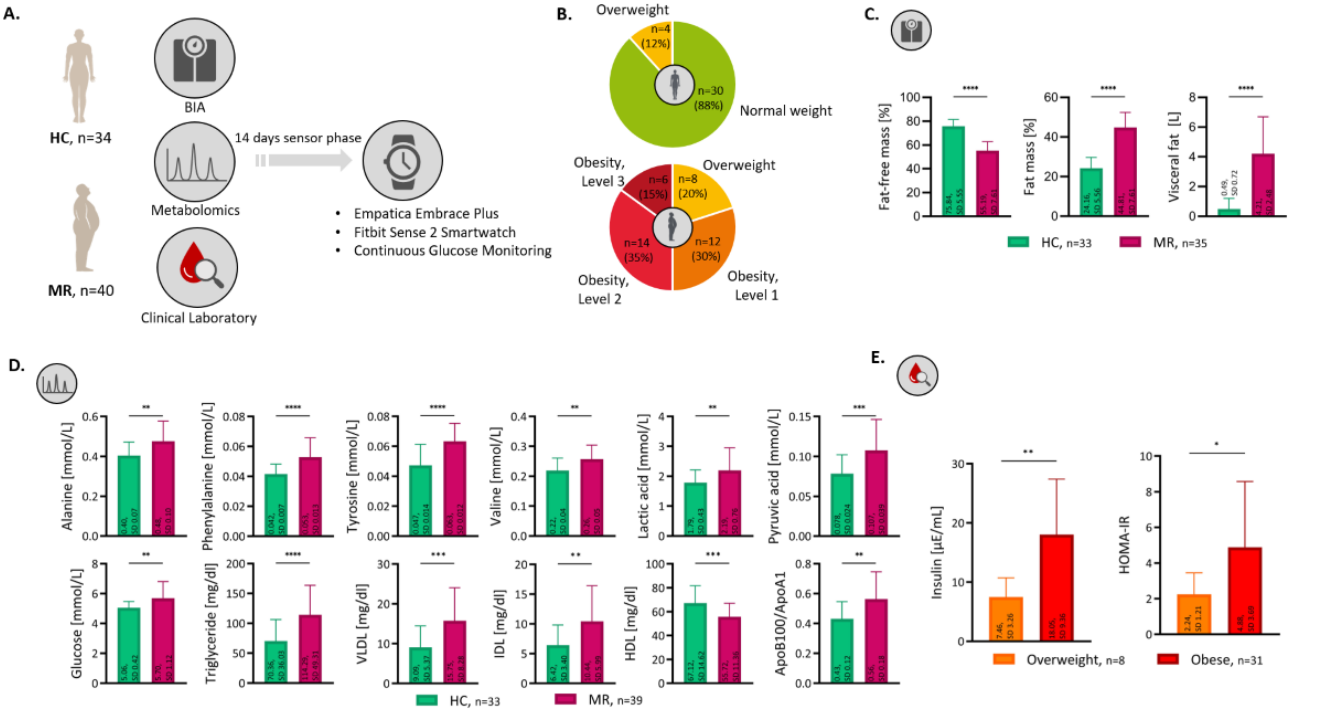
[n]TBW: total body water.

[o]ECW: extracellular water.

XSL•FO
RenderX

The HC cohort comprised 34 individuals, of whom 25 (74%) were women, with a mean age of 24.53 (SD 3.68) years and a BMI within the normal range (mean 22.36, SD 2.16 kg/m²). In contrast, the MR cohort consisted of 40 individuals, including 29 (73%) women, with a mean age of 55.38 (SD 15.08) years and an average BMI in the obese range (mean 35.38, SD 4.91 kg/m²), resulting in significant differences in age and BMI between the 2 cohorts. Within the MR cohort, 8 (20%) participants were overweight with at least one known comorbidity, 12 (30%) presented with obesity class I, 14 (35%) with class II, and 6 (15%) with class III. Both cohorts underwent BIA as well as deep metabolic profiling of blood and urine samples (urine data not shown; Table 1 and Figure 1A). Detailed results of the BIA measurements are presented in Table 1 and Figure 1B. Key parameters linked to metabolic syndrome and disruption in glucose metabolism, including fat-free mass, fat mass, and visceral fat, are highlighted in Figure 1C. Compared to HC, individuals in the MR cohort showed significantly increased fat mass and visceral fat, whereas relative fat-free mass was decreased.

Figure 1. Study design and metabolic markers. (A) Study cohort included 74 participants performing a bioelectric impedance analysis (BIA), deep metabolic profiling, and 2-week sensor phase, during which they wore an invasive device for continuous glucose monitoring (CGM) and 2 noninvasive sensor wristbands. (B) Overall cohort was separated into healthy controls (HC) and metabolically at-risk (MR) individuals based on BMI and comorbidities. (C) BIA results indicate alterations in lipid profiles in the MR cohort (n=35) compared to HC (n=33). (D) Blood metabolic phenotyping showed changes in amino acid, glucose, energy, as well as lipid metabolism between MR (n=39) and HC cohorts (n=33). (E) Detailed analysis of glucose metabolism revealed differences among MR individuals with overweight (n=8) and obesity (n=31). Multiple Mann-Whitney *U* test with false discovery rate of 1% for BIA measurement and metabolic profiling as well as 10% for clinical analysis of glucose metabolism were applied with *$q \leq .1$, **$q \leq .01$, ***$q \leq .001$, and ****$q < .0001$. HDL: high-density lipoprotein; HOMA-IR: homeostatic model assessment of insulin resistance; IDL: intermediate-density lipoprotein; LDL: low-density lipoprotein; VLDL: very-low-density lipoprotein.



Furthermore, deep metabolic profiling of blood samples revealed alterations in amino acid, glucose, energy, as well as lipid metabolism. Elevated levels of amino acids alanine, phenylalanine, tyrosine, and valine were detected in MR compared to HC, along with increased concentrations of lactic and pyruvic acid as well as glucose. In line with the results of the BIA measurement, increased lipid levels (triglycerides, VLDL, and IDL) were observed in MR, whereas HDL was reduced in MR. Furthermore, an increased apolipoprotein ratio (ApoB100/ApoA1) was observed in the MR cohort, serving as a marker for elevated cardiovascular risk (Figure 1D).

To further characterize the overall health status of the MR cohort, additional parameters, including complete blood count, glucose metabolism, liver function, and blood lipids, were assessed by routine clinical laboratory testing (Table 2). Within the MR group, glucose metabolism was analyzed in detail across overweight and obese subgroups, revealing that individuals with obesity exhibited significantly higher insulin concentrations and increased insulin resistance, as indicated by homeostatic model assessment of insulin resistance (Figure 1E). However, no differences were observed between groups for fasting glucose and HbA_{1c}.

XSL•FO
RenderX

Table . Clinical parameters for the metabolically at-risk cohort[a].

Clinical parameter	Total (n=40)	Overweight (n=8)	Obesity (n=32)
Gender, n (%)			
Women	29 (73)	4 (50)	25 (78)
Men	11 (27)	4 (50)	7 (22)
Large blood count, mean (SD)			
Leukocytes (G/L)	6.66 (1.70)	5.99 (1.10)	6.83 (1.80)
Erythrocytes (T/L)	4.81 (0.35)	5.01 (0.44)	4.76 (0.31)
Hemoglobin (g/dL)	14.34 (1.20)	14.71 (1.22)	14.25 (1.19)
Hematocrit (L/L)	0.42 (0.03)	0.43 (0.03)	0.42 (0.03)
MCH[b] (pg/Ery)	30.03 (1.90)	29.50 (2.45)	30.16 (1.75)
MCHC[c] (g/dL)	34.03 (1.22)	34.13 (1.73)	34.00 (1.10)
MCV[d] (fL)	88.03 (3.87)	86.50 (4.78)	88.42 (3.58)
Thrombocytes (G/L)	267.21 (57.73)	234.25 (30.24)	275.71 (60.37)
Neutrophils (G/L)	3.92 (1.31)	3.71 (1.13)	3.96 (1.35)
Neutrophils (%)	58.14 (8.29)	60.94 (9.21)	57.42 (8.04)
Lymphocytes (G/L)	1.96 (0.70)	1.53 (0.39)	2.08 (0.72)
Lymphocytes (%)	29.75 (7.91)	26.15 (7.51)	30.67 (7.86)
Monocytes (G/L)	0.52 (0.16)	0.54 (0.11)	0.52 (0.17)
Monocytes (%)	7.93 (2.26)	9.04 (1.27)	7.64 (2.38)
Eosinophils (G/L)	0.18 (0.11)	0.17 (0.08)	0.19 (0.12)
Eosinophils (%)	2.91 (1.85)	2.88 (1.41)	2.92 (1.96)
Basophils (G/L)	0.05 (0.03)	0.04 (0.02)	0.05 (0.03)
Basophils (%)	0.84 (0.47)	0.73 (0.54)	0.87 (0.46)
Immature granulocytes (G/L)	0.02 (0.02)	0.02 (0.01)	0.03 (0.02)
Immature granulocytes (%)	0.35 (0.27)	0.28 (0.07)	0.37 (0.30)
Glucose, mean (SD)			
Fasting glucose (mg/dL)	107.56 (19.26)	116.50 (23.16)	105.26 (17.83)
HbA_{1c}[e] (mmol/mol Hb)	38.27 (6.87)	41.43 (10.88)	37.18 (5.07)
HbA_{1c} (%)	5.67 (0.63)	5.98 (0.99)	5.57 (0.46)
Insulin (µE/mL)[f]	15.87 (9.48)	7.46 (3.26)	18.05 (9.36)
HOMA-IR[f,g]	4.34 (3.49)	2.24 (1.21)	4.88 (3.69)
Liver, mean (SD)			
GOT[h]/ASAT[i] (U/L)	23.85 (5.89)	22.88 (2.95)	24.10 (6.45)
GPT[j]/ALAT[k] (U/L)	29.08 (14.06)	20.50 (5.95)	31.29 (14.75)
GGT[l] (U/L)	30.67 (17.93)	22.88 (13.48)	32.68 (18.56)
Blood lipids, mean (SD)			
Cholesterol (mg/dL)	187.26 (45.37)	203.63 (48.93)	183.03 (44.26)
Triglycerides (mg/dL)	126.92 (50.19)	114.63 (41.05)	130.10 (52.41)
HDL[m] cholesterol (mg/dL)	53.74 (12.40)	58.38 (12.34)	52.55 (12.33)
Non-HDL cholesterol (mg/dL)	133.51 (41.55)	145.25 (39.67)	130.48 (42.11)
LDL[n] cholesterol (mg/dL)	108.13 (38.85)	122.38 (41.09)	104.45 (38.07)

XSL•FO
RenderX

[a]Clinical parameters were shown for 8 participants with overweight and 31 with obesity, while 1 blood sample is missing.

[b] MCH: mean corpuscular hemoglobin.

[c] MCHC: mean corpuscular hemoglobin concentration.

[d] MCV: mean corpuscular volume.

[e]HbA_{1c}: hemoglobin A_{1c}.

[f]Significant difference between participants with overweight and obesity using multiple Mann-Whitney tests with a false discovery rate of 10%.

[g]HOMA-IR: homeostatic model assessment of insulin resistance.

[h]GOT: glutamate oxaloacetate transaminase.

[i] ASAT: aspartate aminotransferase.

[j]GPT: glutamate pyruvate transaminase.

[k] ALAT: alanine aminotransferase.

[l]GGT: gamma glutamyl transferase.

[m]HDL: high-density lipoprotein.

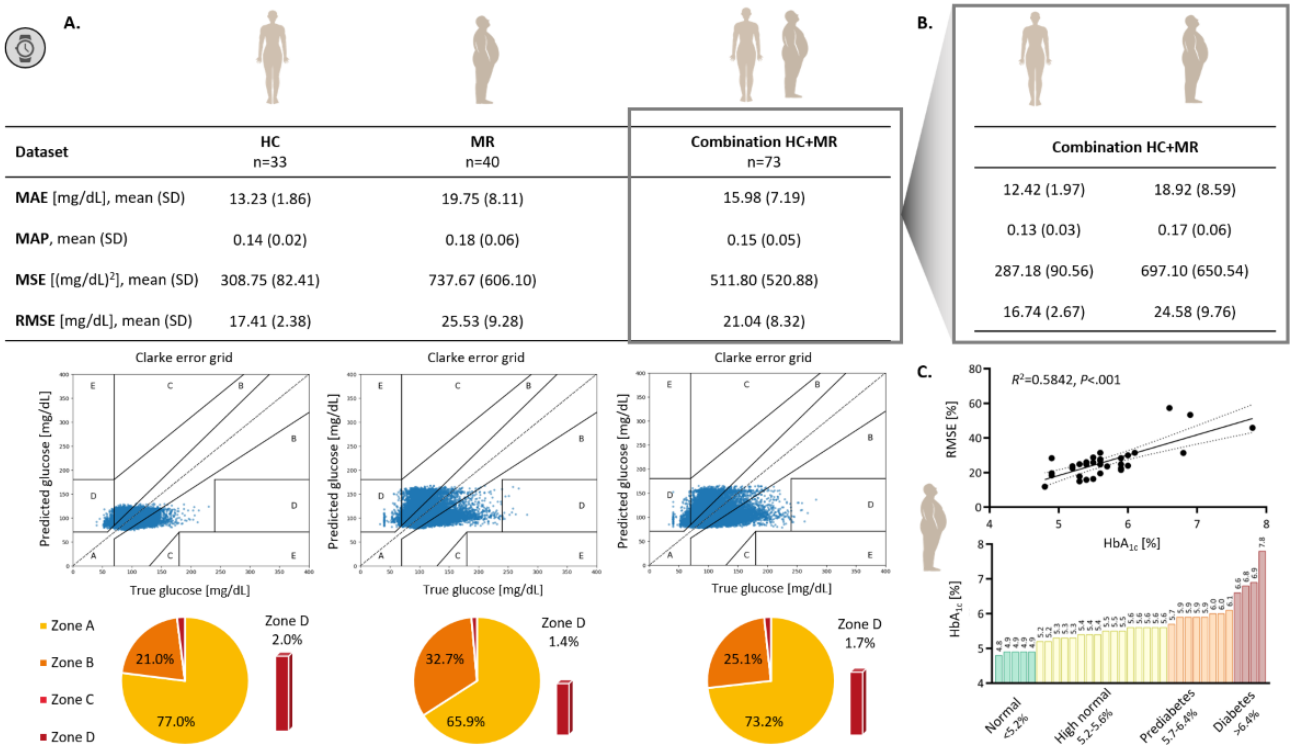
[n]LDL: low-density lipoprotein.

Calculation of IG Values

The aim of the study was to expand our algorithm for the calculation of IG values using noninvasive sensor data and to validate the generalizability of the proposed algorithm under healthy and MR conditions. Therefore, we pursued 2 approaches. First, we implemented 2 cohort-specific algorithms trained and tested for the HC and MR cohorts, respectively. Second, we developed an algorithm using the overall population, consisting of the HC and MR cohorts. For the HC cohort, we achieved a strong performance, indicated by a mean MAE of 13.23 (SD 1.86) mg/dL, a mean MAPE of 0.14 (SD 0.02), a mean RMSE of 17.41 (SD 2.38) mg/dL, and 2.0% of data points in zone D of the CEGA. A total of 33 healthy individuals were included in this dataset, as 1 (3%) participant had to be excluded due to his frequent sauna visits and the resulting errors in data collection. Our MR cohort–specific algorithm also performed well, including 40 individuals, with a mean MAE of 19.75 (SD 8.11) mg/dL, a mean MAPE of 0.18 (SD 0.06), and a mean RMSE of 25.53 (SD 9.28) mg/dL. However, only 1.4% of data points fell in zone D of the CEGA. Furthermore, a cohort-independent algorithm including the HC and MR dataset was developed with a mean MAE of 15.98 (SD 7.19) mg/dL, a mean MAPE of 0.15 (SD 0.05), a mean RMSE of 21.04 (SD 8.32) mg/dL, and 1.7% in zone D of the CEGA (Figure 2A). Model performance characteristics of the overall analysis were stratified by the HC and MR cohorts (Figure 2B), yielding MAE, MAPE, mean squared error, and RMSE values comparable to the cohort-specific analyses. These findings support the assumption that the algorithm is capable of abstracting across the different datasets underscoring its generalizability between various datasets. In addition, we identified a linear association between the RMSE and the HbA_{1c} of the MR cohort (R^2=0.5842; P<.001). The simple linear regression analysis remained highly significant (P<.001) after the exclusion of the outlier exhibiting an HbA_{1c} value above 7.0%. To further investigate the influence of intersubject variability, we performed a multivariate linear regression analysis with RMSE as the dependent variable and both age and HbA_{1c} as independent variables, indicating that age is not significantly associated with RMSE (β_1=−0.0146; P=.10), while HbA_{1c} shows a strong and statistically significant association with the RMSE (β_2=12.57; P<.001), even after adjusting for age. In line with a previous study by Bent et al [17], we categorized the HbA_{1c} into normal (<5.2%), high normal (5.2% - 5.6%), prediabetic (5.7% - 6.4%), and diabetic (>6.4%), resulting in 82% (n=28 of 33 known values) of individuals with high HbA_{1c} levels and potential disruptions in glucose metabolism (Figure 2C).

XSL•FO
RenderX

Figure 2. Performance of the algorithm for interstitial blood glucose calculation. (A) Model performance was evaluated by mean absolute error (MAE), mean absolute percentage error (MAPE), mean squared error (MSE), root-mean-squared error (RMSE), and Clarke error grid analysis (CEGA) for cohort-specific algorithms of the healthy control (HC) and metabolically at-risk (MR) groups as well as for the combined dataset model. (B) Stratification of the combined model demonstrated its potential for generalizability across different datasets. (C) Linear regression analysis indicated a relationship between the RMSE of the algorithm and abnormalities in blood glucose metabolism, characterized by hemoglobin A_{1c} (HbA_{1c}). Furthermore, HbA_{1c} levels of the MR cohort were divided into normal (<5.2%), high normal (5.2% - 5.6%), prediabetic (5.7% - 6.4%), and diabetic (>6.4%).



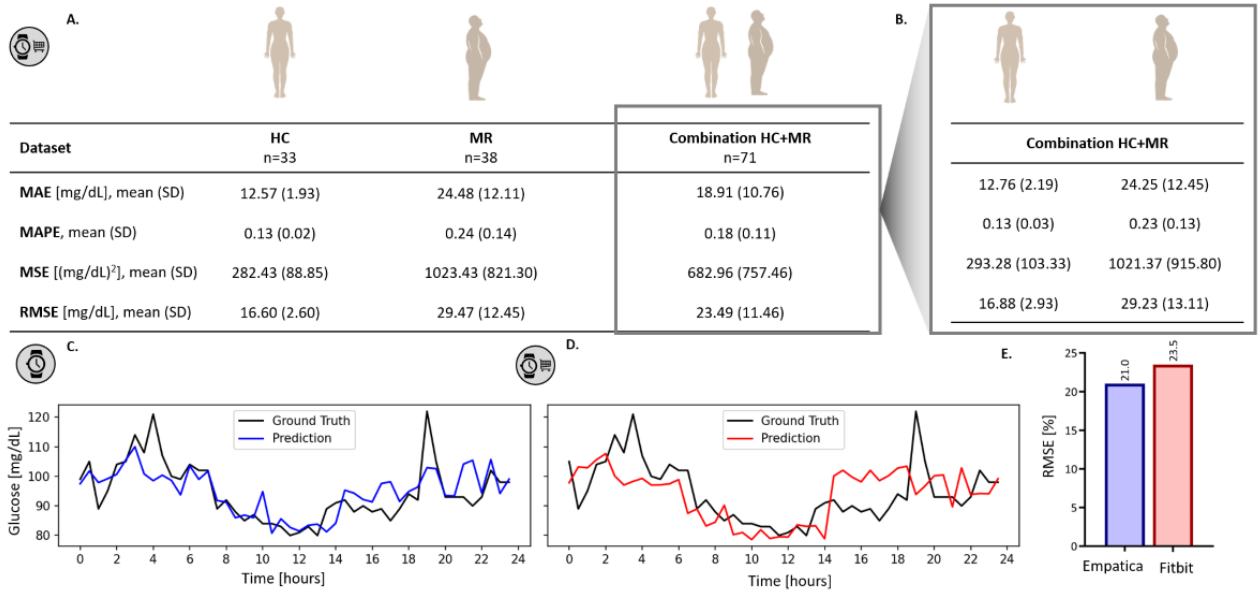
Dataset	HC n=33	MR n=40	Combination HC+MR n=73
MAE [mg/dL], mean (SD)	13.23 (1.86)	19.75 (8.11)	15.98 (7.19)
MAP, mean (SD)	0.14 (0.02)	0.18 (0.06)	0.15 (0.05)
MSE [(mg/dL)^2], mean (SD)	308.75 (82.41)	737.67 (606.10)	511.80 (520.88)
RMSE [mg/dL], mean (SD)	17.41 (2.38)	25.53 (9.28)	21.04 (8.32)

Combination HC+MR	
12.42 (1.97)	18.92 (8.59)
0.13 (0.03)	0.17 (0.06)
287.18 (90.56)	697.10 (650.54)
16.74 (2.67)	24.58 (9.76)

To evaluate the practical applicability and broad usability of noninvasive IG estimation, the use of a commercially available smartwatch (Fitbit Sense 2) was investigated. Analogous to the evaluation of the sensor data generated by the scientific wristband from Empatica, 3 models were trained: 2 cohort-specific models for the HC and MR groups and 1 combined model. The combined model achieved a mean MAE of 18.91 (SD 10.76) mg/dL, a mean MAPE of 0.18 (SD 0.11), and a mean RMSE of 23.49 (SD 11.46) mg/dL (Figures 3A and 3B). The Fitbit dataset comprised 33 (46%) HC and 38 (54%) participants from the MR cohort, as data from 2 (5% of the overall MR cohort) MR individuals were unavailable due to missing recordings. In line with the Empatica combined model, the combined model based on Fitbit data demonstrated the capability to generalize across different datasets. Furthermore, measured IG values based on CGM as well as the calculated parameters derived from sensor data over a 24-hour period were illustrated for Empatica (Figure 3C) and Fitbit (Figure 3D), showing a high overlap despite visible dynamics in IG values. In summary, the Empatica and Fitbit models exhibited comparable RMSE values, although the commercially available smartwatch showed a slightly reduced performance (Figure 3E).

XSL•FO
RenderX

Figure 3. Algorithm performance based on smartwatch data. (A) Model performance was evaluated by mean absolute error (MAE), mean absolute percentage error (MAPE), mean squared error (MSE), and root-mean-squared error (RMSE) for the cohort-specific algorithms of the healthy control (HC) and metabolically at-risk (MR) groups as well as characteristics of the combined dataset model. (B) Stratification of the combined model showed the potential for abstraction across different datasets. (C) Interstitial glucose levels measured by a continuous glucose monitoring (CGM) device (black) were compared with values calculated from noninvasive wearable data obtained with the Empatica EmbracePlus (blue) for 1 participant over time. (D) Analogous to C, calculated glucose values were derived from data collected using a commercially available smartwatch (Fitbit Sense 2) (red). (E) Comparison of RMSE between algorithms based on Empatica and Fitbit data for the combined model.



Dataset	HC n=33	MR n=38	Combination HC+MR n=71
MAE [mg/dL], mean (SD)	12.57 (1.93)	24.48 (12.11)	18.91 (10.76)
MAPE, mean (SD)	0.13 (0.02)	0.24 (0.14)	0.18 (0.11)
MSE [(mg/dL)^2], mean (SD)	282.43 (88.85)	1023.43 (821.30)	682.96 (757.46)
RMSE [mg/dL], mean (SD)	16.60 (2.60)	29.47 (12.45)	23.49 (11.46)

Combination HC+MR	
12.76 (2.19)	24.25 (12.45)
0.13 (0.03)	0.23 (0.13)
293.28 (103.33)	1021.37 (915.80)
16.88 (2.93)	29.23 (13.11)

Moreover, to provide additional statistical validation and strengthen comparative claims, SE and 95% CI for both RMSE and MAE were computed using nonparametric bootstrapping with 1000 resamples under the LOSO cross-validation framework for the main cohort analyses, including HC and MR cohorts, as well as the combined cohort, using the proposed LSTM model based on either Empatica or Fitbit data. As shown in Table S4 in Multimedia Appendix 1, the 95% CI for RMSE was 16.34 to 18.48 mg/dL in the HC cohort, 22.65 to 28.41 mg/dL in the MR cohort, 19.12 to 22.96 mg/dL for the combined cohort with Empatica, and 20.81 to 26.17 mg/dL for the combined cohort with Fitbit. Corresponding 95% CIs for MAE were 12.45 to 14.01 mg/dL, 18.12 to 21.38 mg/dL, 14.57 to 17.39 mg/dL, and 17.03 to 20.79 mg/dL, respectively. Despite relatively large SDs across LOSO folds, especially in the MR and combined cohorts, all derived 95% CIs remained narrow, indicating robust and consistent central estimates of model performance. These intervals explicitly quantify statistical uncertainty in overall performance and reinforce that observed between-cohort performance differences are reproducible and not driven by random estimation variability.

To address the influence of PH on IG prediction, we evaluated model performance at 15, 30, and 60 minutes using both Empatica EmbracePlus and Fitbit Sense 2 records (Table S5 in Multimedia Appendix 1). For the Empatica device, the proposed LSTM achieved the lowest RMSE (mean 21.04, SD 8.32 mg/dL) and MAE (mean 15.98, SD 7.19 mg/dL) at a 30-minute PH, with slightly higher errors at 15 minutes (RMSE: mean 21.69, SD 8.30 mg/dL) and 60 minutes (RMSE: mean 22.68, SD 9.01 mg/dL). A similar pattern was observed for the Fitbit device. The 30-minute PH yielded the optimal RMSE (mean 23.49, SD 11.46 mg/dL) compared to 15 minutes (mean 26.10, SD 15.30 mg/dL) and 60 minutes (mean 24.50, SD 11.75 mg/dL). The 30-minute interval was selected as the primary PH because it balances temporal resolution and physiological relevance for PPGR, while minimizing prediction error across both wearable platforms. To validate the LSTM architecture, we further compared it with the iTransformer model at the optimal 30-minute PH (Table S5 in Multimedia Appendix 1). For Empatica data, the LSTM model outperformed iTransformer. For Fitbit, LSTM remained competitive (RMSE: mean 23.49, SD 11.46 vs mean 22.68, SD 8.40 mg/dL) with lower MAE in real-world wearable signals. These results confirm that the LSTM network is better suited for capturing sequential physiological dependencies from continuous wearable data than the transformer-based architecture in this task.

The ablation study confirmed the differential individual and combined contributions of each sensor modality to IG prediction (Table S6 in Multimedia Appendix 1). Single-modality removal showed that PPG and taACC were the most impactful individual features, with their exclusion increasing RMSE by 5.82 and 8.43 mg/dL, respectively, followed by EDA, which led to an RMSE increase of 4.17 mg/dL. STEMP contributed moderately, with an RMSE degradation of 2.35 mg/dL upon elimination. Combinatorial ablation further revealed strong synergistic interactions between modalities. Simultaneous removal of EDA and PPG caused the largest performance drop (RMSE of 9.13 mg/dL), while removing STEMP and taACC resulted in a substantial decline of 7.61 mg/dL, consistent with the high importance of accelerometer-derived motion information. Similarly, excluding PPG and taACC together led to the most severe degradation (RMSE of 10.25 mg/dL), underscoring their dominant and complementary roles. Other combined ablations (EDA+STEMP) also yielded notable accuracy reductions. These findings demonstrate that cardiovascular dynamics from PPG, physical activity patterns from taACC, and autonomic arousal

XSL•FO
RenderX

from EDA provide the strongest predictive signals for glucose fluctuations, whereas STEMP acts as a supplementary but meaningful contributor.

To identify the dominant physiological and contextual drivers of glucose prediction, we performed a SHAP analysis on the full dataset (n=39,314), quantifying the magnitude and direction of each feature's impact on the model's glucose prediction output, with results visualized in the SHAP beeswarm plot (Figure S1 in Multimedia Appendix 1). It ranks features by the overall magnitude of their SHAP values, revealing that demographic and temporal features were the most influential predictors at the top of the ranking, followed closely by wearable physiological sensor modalities. Focusing on the top 30 most impactful features, we categorized them by sensor modality to quantify their relative contributions, that is, 3 (10%) of the top 30 features were demographic variables, 2 (6.7%) of 30 were temporal metadata, and the remaining 25 (83.3%) of 30 were derived from wearable sensors, encompassing BVP (Plr), EDA, STEMP, and taACC. Within the wearable sensor modalities, accelerometer-derived features were the most abundant, with 12 features (40% of the top 30), followed by 6 (20%) BVP-related features, 4 (13.3%) STEMP features, and 3 (10%) EDA features. The SHAP beeswarm plot further reveals biologically plausible directional effects for these top-ranked features, consistent with well-documented age-related metabolic changes and glucose dysregulation risk. The temporal feature also exhibited a strong directional impact, with high values linked to elevated glucose predictions, reflecting long-term temporal trends in metabolic health and study population dynamics. Wearable physiological features demonstrated consistent dose-response relationships, with high feature values driving predictable shifts in model output, validating their direct physiological relevance to glucose regulation as key complementary functions. In general, these results confirm that demographic and temporal features are the most dominant predictors of glucose levels in our model, with wearable sensor modalities providing critical, physiologically meaningful contributions that enhance predictive accuracy by capturing real-time physiological dynamics.

Discussion

Principal Findings

Our study demonstrates the feasibility of predicting IG levels exclusively using noninvasive sensor data and ML techniques. The proposed algorithm achieved high performance across healthy and MR individuals, confirming our previous results [19], even in MR participants with higher glucose levels and dynamics. We implemented cohort-specific algorithms focusing on the HC and MR dataset, respectively, as well as an overall algorithm trained and tested for all individuals together. Furthermore, we identified a potential linear relationship between the error of our algorithms and participants' glucose metabolism, as characterized by HbA_{1c}. This approach may in the future allow for the estimation of HbA_{1c} levels from algorithmic outputs, thereby offering individuals an early indication of possible metabolic dysregulation. In addition, we investigated the potential of using data from commercially available smartwatches for IG prediction. Despite the limited availability of continuous raw data compared with research-grade sensors, we achieved promising results.

Taken together, our IG predictive performance was in line with other state-of-the-art studies, but outperformed these studies in three significant points: (1) to the best of our knowledge, this is the first study using an equal number of healthy and MR individuals resulting in the largest continuous observation with 74 individuals so far; (2) in contrast to other studies, our study exclusively used data collected by wearable smart devices without using manually recorded food logs. Consequently, participants were not required to engage in an additional time-consuming self-reporting, which is prone to data errors; and (3) in addition to sensor wristbands designed for scientific studies, we also tested a commercially available smartwatch in parallel to demonstrate scalability of our proposed solution.

Comparison to the Literature

Like others, we emphasize the potential of wearable devices for noninvasive glucose prediction. For example, Aziz et al [30] used an open-source dataset consisting of 13 participants aged between 9 and 77 years with a nondiabetic-to-diabetic ratio of 4:6. Various physiological measurements were collected over 3 months, resulting in linear and nonlinear models with high accuracy, specifically an MAE ranging from 0.093 to 0.142 and an RMSE ranging from 0.186 to 0.271. However, the generalizability of this study is limited by the participant-dependent allocation of training and testing as well as the small sample size. Similarly, van den Brink et al [16] conducted a 2-week study with 24 participants without diabetes, using handcrafted features from diet, physical activity, and sleep data to predict glucose levels. They developed a participant-specific Extreme Gradient Boosting model to predict glucose levels with an average MAE of 11.17 mg/dL (reported 0.62, SD 0.15 mmol/L) for the test data. However, this model is also restricted due to the limited generalizability by the personalized training and test strategy. Furthermore, a study by Bent et al [17] involved 16 participants with high normal BG (HbA_{1c}=5.2 - 5.6) or prediabetes (HbA_{1c}=5.7 - 6.4) using a CGM sensor, a wristband, and self-reported food logs for 8 to 10 days. They extracted various features and created 2 Extreme Gradient Boosting models—one participant-specific and one population-based—to predict IG levels, achieving an average MAPE of 14.33% (SD 3.25%) and a mean RMSE of 21.22 (SD 4.14) mg/dL for the population model validated with leave-one-person-out cross-validation. Nevertheless, the model is restricted due to the use of self-reported food logs and the limited sample size. Recently, a study published by Zeynali et al [31] demonstrated an RMSE of 19.7 mg/dL using PPG signals of a diverse dataset during surgery and anesthesia. However, although this study included 6388 participants, it comprised only 35,358 BG measurements, corresponding to an average of 5.5 values per participant. In contrast, our longitudinal 2-week dataset comprises 900,606 IG measurements from 74 individuals, providing the potential to capture the temporal dynamics of IG values. Moreover, the glucose data in the study by Zeynali et al were collected during surgical procedures and anesthesia, thus reflecting a highly specific clinical context. A study by Geng et al [32] developed a multisensor-based,

XSL•FO
RenderX

noninvasive continuous glucometer with a normalized RMSE of 14.61 mg/dL. However, this study aimed to develop a new sensor device and therefore only 6 healthy individuals and 3 individuals with diabetes were examined for the accuracy of the model. In a study by Zahedani et al [18], an app was developed to identify food types and metabolic factors from images, aiding in glucose prediction. Although this method reduces the burden of self-reporting, it still requires participant involvement, making it unsuitable for long-term monitoring, and its accuracy in complex dietary situations remains unverified. In conclusion, these studies have some limitations in small sample sizes, data processing, and validation and necessitate detailed dietary intake data, requiring significant participant commitment. Without qualitative assessment, the reliability of these ML models for glucose prediction based on clinical validation criteria is challenging.

Technical Discussion

From the technical point of view, the supplementary 95% CIs derived from nonparametric bootstrapping further validate the stability and generalizability of our prediction model across HC and MR populations. The narrow CIs for RMSE and MAE in all cohorts confirm low between-participant variability and high reproducibility of model performance, rather than reflecting random or spurious estimation. The nonoverlapping tendency of CIs between the HC and MR groups also quantitatively supports the genuine difference in predictive accuracy related to underlying metabolic status, rather than chance variation. Collectively, these statistical estimates strengthen the physiological interpretability of our wearable-based glucose prediction model and reinforce its reliability for real-world mHealth apps in personalized nutrition and metabolic health monitoring.

The optimal performance at a 30-minute PH aligns with the typical time course of postprandial IG excursions, which peak approximately 30 to 60 minutes after meal ingestion. Shorter horizons may lack sufficient physiological signal for robust prediction, while longer horizons (PH=60 min) introduce greater uncertainty due to accumulating variability in lifestyle and metabolic state. Thus, the 30-minute PH provides the best trade-off between predictive accuracy and clinical utility for real-time glycemic monitoring and personalized dietary guidance. Benchmarking against an iTransformer further validates the choice of LSTM for noninvasive IG prediction. While transformers excel at long-range dependencies in static datasets, LSTM networks are more efficient at modeling continuous, noisy physiological time series from wearables, where local temporal dynamics dominate glucose fluctuations. The consistent superiority of LSTM across both research-grade and commercial wearable devices supports its suitability for real-world mHealth apps requiring low latency and robust sequential learning.

Additionally, the findings from ablation tests further strengthen the implied perspectives of the predictive signals underlying noninvasive IG estimation. Individual modality removal revealed that PPG and taACC were the dominant contributors to model accuracy, consistent with the close coupling between cardiovascular tone, physical activity–related metabolic demand, and IG dynamics. EDA also contributed substantially, reflecting the role of autonomic arousal in glucose fluctuation prediction. In contrast, STEMP exerted a more modest effect, supporting its role as a complementary rather than primary predictive feature. Combinatorial ablation confirmed strong synergistic interactions across key modalities. Notably, simultaneous removal of PPG and taACC caused the most severe performance deterioration, while the combined exclusion of EDA and PPG also led to a marked decline, underscoring the value of multimodal integration for robust real-world glucose prediction. These results clarify the relative importance of each wearable signal and validate that our model relies on physiologically meaningful inputs rather than spurious correlations, reinforcing the translational potential of the proposed approach.

Metabolomics

As a secondary outcome, deep phenotyping of HC and MR individuals was performed. In line with other studies, we identified elevated fat mass and visceral fat in our MR cohort [33,34]. In particular, visceral fat is a well-known biomarker for T2D and cardiovascular risk. Furthermore, blood metabolic profiling revealed alterations in amino acid, energy, and glucose, as well as lipid metabolism. Previous studies showed altered amino acid profiles in T2D compared to HC. For example, a study by Chen et al [35], published in 2019, confirmed our results, highlighting that elevated concentrations of alanine, phenylalanine, tyrosine, and valine increased the risk of T2D. Furthermore, disruptions in glucose metabolism, indicated by increased BG levels, are common biomarkers for the development of insulin resistance, T2D, and metabolic syndrome [36]. Abnormalities in lipid profiles identified by BIA measurement were confirmed by NMR spectroscopy, which indicated increased concentrations of VLDL and IDL, as well as the ratio of ApoA1/ApoA100, a well-known biomarker for cardiovascular risk [37]. In contrast, HDL, commonly referred to as “good cholesterol,” was reduced in MR [38]. Within the MR cohort, individuals with obesity exhibited higher insulin resistance compared to those with overweight, underscoring the association between increased body weight and impaired metabolism [1,2]. This, in turn, may contribute to an elevated risk of insulin resistance, T2D, and metabolic syndrome. Taken together, our deep phenotyping confirms the disruptions in glucose metabolism in the MR cohort and thereby strengthens the validity of our algorithm, which was tested both in healthy individuals and in participants with metabolic impairment.

Limitations

Nevertheless, further research is needed to validate the IG prediction algorithm. In our study, we examined only a limited cohort consisting of 2 substudies with different study populations, including differences in demographic characteristics such as age and BMI. Furthermore, we were able to test only one commercial smartwatch and did not cover other wearables and smartwatches on the market. Therefore, it is appropriate to evaluate the efficacy in larger and more diverse cohorts, as well as testing the model on distinct sensor data collected from various smart devices is essential for detailed fine-tuning to reduce the impact of data diversity on IG prediction outcomes. This includes validating the algorithm in different clinical

XSL•FO
RenderX

scenarios following the V3 [39] principle and adapting it for predicting IG response in metabolic diseases such as obesity and T2D. Furthermore, the role of confounding factors such as age, gender, BMI, dietary intake, physical activity, stress, sleep, and medication should be evaluated. In particular, the observed relationship between RMSE and HbA_{1c} levels, as well as a possible association with other confounders, warrants further consideration. On the one hand, this observation may indicate reduced model performance in persons with strong fluctuations in IG levels, while on the other hand, it may suggest a potential for noninvasive estimation of HbA_{1c} levels, which could be of interest for prevention and early detection of abnormalities in glucose metabolism. It is also critical to rigorously validate the extracted features to ensure that predictive signals reflect true physiological responses rather than artifacts or spurious correlations. In particular, when using smartwatch data with limited output parameters, it is important to verify that physiological signals actually reflect the IG values. A limitation of this study is that the model was evaluated primarily in terms of pointwise glucose prediction accuracy. Dynamic characteristics such as trend direction, rate of change, and peak timing were not explicitly assessed. These aspects are important for clinical decision-making and should be investigated in future work. From the technical point, exploring AI learning strategies with promising potential in processing physiological signals, including updated Neural Hierarchical Interpolation for Time Series [40], transfer learning [41], or federated learning [42], is warranted. Moreover, combining multimodal data sources such as continuous activity, sleep, and heart rate metrics could further enhance feature validity and predictive accuracy across diverse populations. In the long term, systematic assessment of the algorithm's sensitivity to lifestyle and dietary variations is necessary to ensure its robustness for real-world mHealth apps and PLGD recommendations.

Conclusions

In conclusion, we predicted continuous IG values using wearable technologies and ML approaches, resulting in an algorithm obtaining strong performance with an average MAE of 15.98 (SD 7.19) mg/dL, a mean MAPE of 0.15 (SD 0.05), a mean RMSE of 21.04 (SD 8.32) mg/dL, and 1.7% of predictive points in zone D of the CEGA. Notably, this performance was achieved using one of the largest datasets consisting of HC and MR individuals without relying on food logs, which sets our study apart from others. The widespread availability of wearables and smartwatches offers a scalable and cost-efficient path toward broad CGM-based monitoring, supporting PLGD strategies and enabling early risk detection. Leveraging longitudinal multimodal data may further enhance predictive accuracy and extend the applicability of PLGD approaches to other conditions associated with metabolic dysregulation, including neurodegenerative diseases.

Acknowledgments

The authors used the generative AI tool such as ChatGPT by OpenAI to translate, rephrase, and clarify sentences. Outcomes of generative AI were checked carefully and were not used in study design, data analysis, or to produce manuscript content.

Funding

The studies were financially supported by the DAMP foundation (project: SENSE-Systemische ErnähruNgSmEdizin; 2020 - 14) and the German Federal Ministry of Research, Technology and Space–BMFTR, previous BMBF) project BOOTSTRAP (03DPS1031).

Data Availability

The data generated and analyzed in this study are not publicly available due to privacy reasons but were considered after reasonable request after discussion with the corresponding author. The underlying code for this study is not publicly available for proprietary reasons.

Authors' Contributions

FS contributed to the conceptualization, methodology, validation, formal analysis, investigation, resources, data curation, visualization, project administration, and funding acquisition, and wrote the original draft and contributed to its review and editing. XH contributed to the conceptualization, methodology, validation, formal analysis, investigation, and data curation, and wrote the original draft and contributed to its review and editing. AH contributed to the validation, formal analysis, data curation, visualization, and writing of the original draft. PB contributed to the validation, formal analysis, and data curation. FMR contributed to the conceptualization, investigation, resources, and data curation. CP contributed to the conceptualization, investigation, resources, and data curation, and contributed to the review and editing of the manuscript. AP contributed to the validation, formal analysis, and data curation. MAH contributed to the validation and formal analysis. YZ contributed to the investigation and resources. LJ contributed to the data curation and resources. OW contributed to the resources and data curation. TS contributed to the resources and data curation and secured funding. SD contributed to the resources and review and editing of the manuscript. CEA contributed to the resources and review and editing of the manuscript. MG contributed to the conceptualization and resources, contributed to the review and editing of the manuscript, and provided supervision, project administration, and funding acquisition. CS contributed to the conceptualization and resources, wrote the original draft and contributed to its review and editing, and provided supervision, project administration, and funding acquisition.

XSL•FO
RenderX

FS and XH contributed equally and share first authorship. MG and CS contributed equally and share senior/last authorship.

Conflicts of Interest

OW and TS are employed at Perfood Laboratories GmbH. XH, AH, MAH, YZ, and AP work for expandAI GmbH. The views presented in this manuscript are those of the authors and not necessarily those of the companies. Good publication practices were followed. All other authors declare that they have no competing interests.

Multimedia Appendix 1
Additional tables with supplementary results.
[PDF File, 353 KB - mhealth_v14i1e91724_app1.pdf]

References

1. Donath MY, Shoelson SE. Type 2 diabetes as an inflammatory disease. Nat Rev Immunol 2011 Feb;11(2):98-107. [doi: 10.1038/nri2925] [Medline: 21233852]
2. Sandforth A, Arreola EV, Hanson RL, et al. Prevention of type 2 diabetes through prediabetes remission without weight loss. Nat Med 2025 Oct;31(10):3330-3340. [doi: 10.1038/s41591-025-03944-9] [Medline: 41023486]
3. Jarvis PRE, Cardin JL, Nisevich-Bede PM, McCarter JP. Continuous glucose monitoring in a healthy population: understanding the post-prandial glycemic response in individuals without diabetes mellitus. Metabolism 2023 Sep;146:155640. [doi: 10.1016/j.metabol.2023.155640] [Medline: 37356796]
4. Blaak EE, Antoine JM, Benton D, et al. Impact of postprandial glycaemia on health and prevention of disease. Obes Rev 2012 Oct;13(10):923-984. [doi: 10.1111/j.1467-789X.2012.01011.x] [Medline: 22780564]
5. Zeevi D, Korem T, Zmora N, et al. Personalized nutrition by prediction of glycemic responses. Cell 2015 Nov 19;163(5):1079-1094. [doi: 10.1016/j.cell.2015.11.001] [Medline: 26590418]
6. Berry SE, Valdes AM, Drew DA, et al. Human postprandial responses to food and potential for precision nutrition. Nat Med 2020 Jun;26(6):964-973. [doi: 10.1038/s41591-020-0934-0] [Medline: 32528151]
7. Lelleck VV, Schulz F, Witt O, et al. A digital therapeutic allowing a personalized low-glycemic nutrition for the prophylaxis of migraine: real world data from two prospective studies. Nutrients 2022;14(14):2927. [doi: 10.3390/nu14142927]
8. Kannenberg S, Voggel J, Thieme N, et al. Unlocking potential: personalized lifestyle therapy for type 2 diabetes through a predictive algorithm-driven digital therapeutic. J Diabetes Sci Technol 2026 Jan;20(1):113-123. [doi: 10.1177/19322968241266821] [Medline: 39080863]
9. Rodbard D. Continuous glucose monitoring: a review of recent studies demonstrating improved glycemic outcomes. Diabetes Technol Ther 2017 Jun;19(S3):S25-S37. [doi: 10.1089/dia.2017.0035] [Medline: 28585879]
10. Datye KA, Tilden DR, Parmar AM, Goethals ER, Jaser SS. Advances, challenges, and cost associated with continuous glucose monitor use in adolescents and young adults with type 1 diabetes. Curr Diab Rep 2021 May 15;21(7):22. [doi: 10.1007/s11892-021-01389-1] [Medline: 33991264]
11. Tataranni PA, Larson DE, Snitker S, Ravussin E. Thermic effect of food in humans: methods and results from use of a respiratory chamber. Am J Clin Nutr 1995 May;61(5):1013-1019. [doi: 10.1093/ajcn/61.5.1013]
12. Uluç N, Glasl S, Gasparin F, et al. Non-invasive measurements of blood glucose levels by time-gating mid-infrared optoacoustic signals. Nat Metab 2024 Apr;6(4):678-686. [doi: 10.1038/s42255-024-01016-9] [Medline: 38538980]
13. Kazemi N, Abdolrazzaghi M, Light PE, Musilek P. In-human testing of a non-invasive continuous low-energy microwave glucose sensor with advanced machine learning capabilities. Biosens Bioelectron 2023 Dec 1;241(September):115668. [doi: 10.1016/j.bios.2023.115668] [Medline: 37774465]
14. Abdolrazzaghi M, Katchinskiy N, Elezzabi AY, Light PE, Daneshmand M. Noninvasive glucose sensing in aqueous solutions using an active split-ring resonator. IEEE Sensors J 2021;21(17):18742-18755. [doi: 10.1109/JSEN.2021.3090050]
15. Carletti M, Pandit J, Gadaleta M, et al. Multimodal AI correlates of glucose spikes in people with normal glucose regulation, pre-diabetes and type 2 diabetes. Nat Med 2025 Sep;31(9):3121-3127. [doi: 10.1038/s41591-025-03849-7] [Medline: 40745404]
16. van den Brink WJ, van den Broek TJ, Palmisano S, Wopereis S, de Hoogh IM. Digital biomarkers for personalized nutrition: predicting meal moments and interstitial glucose with non-invasive, wearable technologies. Nutrients 2022 Oct 24;14(21):4465. [doi: 10.3390/nu14214465] [Medline: 36364728]
17. Bent B, Cho PJ, Henriquez M, et al. Engineering digital biomarkers of interstitial glucose from noninvasive smartwatches. NPJ Digit Med 2021 Jun 2;4(1):89. [doi: 10.1038/s41746-021-00465-w] [Medline: 34079049]
18. Zahedani AD, McLaughlin T, Veluvali A, et al. Digital health application integrating wearable data and behavioral patterns improves metabolic health. NPJ Digit Med 2023 Nov 25;6(1):216. [doi: 10.1038/s41746-023-00956-y] [Medline: 38001287]
19. Huang X, Schmelter F, Seitzer C, et al. Digital biomarkers for interstitial glucose prediction in healthy individuals using wearables and machine learning. Sci Rep 2025;15(1):30164. [doi: 10.1038/s41598-025-14172-z]
20. Alva S, Brazg R, Castorino K, Kipnes M, Liljenquist DR, Liu H. Accuracy of the third generation of a 14-day continuous glucose monitoring system. Diabetes Ther 2023 Apr;14(4):767-776. [doi: 10.1007/s13300-023-01385-6] [Medline: 36877403]

XSL•FO
RenderX

21. Dona AC, Jiménez B, Schäfer H, et al. Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. Anal Chem 2014 Oct 7;86(19):9887-9894. [doi: 10.1021/ac5025039] [Medline: 25180432]
22. Woldaregay AZ, Årsand E, Walderhaug S, et al. Data-driven modeling and prediction of blood glucose dynamics: machine learning applications in type 1 diabetes. Artif Intell Med 2019 Jul;98:109-134. [doi: 10.1016/j.artmed.2019.07.007] [Medline: 31383477]
23. Bogue-Jimenez B, Huang X, Powell D, Doblas A. Selection of noninvasive features in wrist-based wearable sensors to predict blood glucose concentrations using machine learning algorithms. Sensors (Basel) 2022 May 6;22(9):35591223. [doi: 10.3390/s22093534] [Medline: 35591223]
24. Fundoiano-hershcovitz Y, Breuer Asher I, Manejwala O. 979-P: leveraging machine learning and demographic insights to optimize blood glucose management on a digital health platform. Diabetes 2025 Jun 20;74(Supplement_1):979-97P. [doi: 10.2337/db25-979-P]
25. Fulcher BD. Feature-based time-series analysis. In: Feature Engineering for Machine Learning and Data Analytics: CRC Press; 2017:87-116. [doi: 10.1201/9781315181080-4]
26. Alshehri OS, Alshehri OM, Samma H. Blood glucose prediction using RNN, LSTM, and GRU: a comparative study. 2024 Jul 23 Presented at: 2024 IEEE International Conference on Advanced Systems and Emergent Technologies (IC_ASET); Hammamet, Tunisia p. 1-5. [doi: 10.1109/IC_ASET61847.2024.10596176]
27. Benjamini Y, Krieger AM, Yekutieli D. Adaptive linear step-up procedures that control the false discovery rate. Biometrika 2006 Sep 1;93(3):491-507. [doi: 10.1093/biomet/93.3.491]
28. Liu Y, Hu T, Zhang H, Wu H, Wang S, Ma L, et al. iTransformer: inverted transformers are effective for time series forecasting. 2024 Presented at: 12th International Conference on Learning Representations (ICLR 2024); May 7-11, 2024. [doi: 10.48550/arXiv.2310.06625]
29. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. Diabetes Care 1987;10(5):622-628. [doi: 10.2337/diacare.10.5.622] [Medline: 3677983]
30. Aziz S, Ahmed A, Abd-Alrazaq A, Qidwai U, Farooq F, Sheikh J. Estimating blood glucose levels using machine learning models with non-invasive wearable device data. Stud Health Technol Inform 2023 Jun 29;305(Dl):283-286. [doi: 10.3233/SHTI230484] [Medline: 37387018]
31. Zeynali M, Alipour K, Tarvirdizadeh B, Ghamari M. Non-invasive blood glucose monitoring using PPG signals with various deep learning models and implementation using TinyML. Sci Rep 2025 Jan 2;15(1):581. [doi: 10.1038/s41598-024-84265-8] [Medline: 39753714]
32. Geng Z, Tang F, Ding Y, Li S, Wang X. Noninvasive continuous glucose monitoring using a multisensor-based glucometer and time series analysis. Sci Rep 2017 Oct 4;7(1):12650. [doi: 10.1038/s41598-017-13018-7] [Medline: 28978974]
33. Duft RG, Castro A, Bonfante ILP, et al. Serum metabolites associated with increased insulin resistance and low cardiorespiratory fitness in overweight adolescents. Nutr Metab Cardiovasc Dis 2022 Jan;32(1):269-278. [doi: 10.1016/j.numecd.2021.09.024] [Medline: 34906412]
34. Shuster A, Patlas M, Pinthus JH, Mourtzakis M. The clinical importance of visceral adiposity: a critical review of methods for visceral adipose tissue analysis. Br J Radiol 2012 Jan;85(1009):1-10. [doi: 10.1259/bjr/38447238] [Medline: 21937614]
35. Chen S, Akter S, Kuwahara K, et al. Serum amino acid profiles and risk of type 2 diabetes among Japanese adults in the Hitachi Health Study. Sci Rep 2019;9(1):7010. [doi: 10.1038/s41598-019-43431-z]
36. Lu X, Xie Q, Pan X, et al. Type 2 diabetes mellitus in adults: pathogenesis, prevention and therapy. Sig Transduct Target Ther 2024;9(1):262. [doi: 10.1038/s41392-024-01951-9]
37. Walldius G, Jungner I. The apoB/apoA-I ratio: a strong, new risk factor for cardiovascular disease and a target for lipid-lowering therapy--a review of the evidence. J Intern Med 2006 May;259(5):493-519. [doi: 10.1111/j.1365-2796.2006.01643.x] [Medline: 16629855]
38. Kosmas CE, Martinez I, Sourlas A, et al. High-density lipoprotein (HDL) functionality and its relevance to atherosclerotic cardiovascular disease. DIC 2018;7:1-9. [doi: 10.7573/dic.212525]
39. Goldsack JC, Coravos A, Bakker JP, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). NPJ Digit Med 2020;3:55. [doi: 10.1038/s41746-020-0260-4] [Medline: 32337371]
40. Challu C, Olivares KG, Oreshkin BN, Garza Ramirez F, Mergenthaler Canseco M, Dubrawski A. NHITS: Neural Hierarchical Interpolation for Time Series Forecasting. AAAI 2023;37(6):6989-6997. [doi: 10.1609/aaai.v37i6.25854]
41. Irshad MT, Li F, Nisar MA, et al. Wearable-based human flow experience recognition enhanced by transfer learning methods using emotion data. Comput Biol Med 2023 Nov;166:107489. [doi: 10.1016/j.compbiomed.2023.107489]
42. Li T, Sahu AK, Talwalkar A, Smith V. Federated learning: challenges, methods, and future directions. IEEE Signal Process Mag 2020;37(3):50-60. [doi: 10.1109/MSP.2020.2975749]

Abbreviations

BG: blood glucose

XSL•FO
RenderX

BIA: bioelectrical impedance analysis
BVP: blood volume pulse
CEGA: Clarke error grid analysis
CGM: continuous glucose monitoring
EDA: electrodermal activity
HbA1c: hemoglobin A1c
HC: healthy control
HDL: high-density lipoprotein
IDL: intermediate-density lipoprotein
IG: interstitial glucose
iTransformer: inverted transformer
LI: linear interpolation
LOSO: leave-one-subject-out
LSTM: long short-term memory
MAE: mean absolute error
MAPE: mean absolute percentage error
mHealth: mobile health
ML: machine learning
MR: metabolically at-risk
NMR: nuclear magnetic resonance
PH: prediction horizon
PLGD: postprandial low-glycemic diet
Plr: pulse rate
PPG: photoplethysmography
PPGR: postprandial glucose response
RMSE: root-mean-squared error
SHAP: Shapley Additive Explanations
STEMP: skin temperature
T2D: type 2 diabetes
taACC: triaxial acceleration
VLDL: very-low-density lipoprotein

Edited by A Stone; submitted 19.Jan.2026; peer-reviewed by J Kumar, M Abdolrazzaghi, MT Irshad; revised version received 29.Jun.2026; accepted 07.Jul.2026; published 30.Sep.2026.

Please cite as:
Schmelter F, Huang X, Heidmann A, Beier P, Rundfeldt FM, Plöger C, Piet A, Hasan MA, Zhang Y, Jablonski L, Witt O, Schröder T, Derer S, Easthope Awai C, Grzegorzek M, Sina C
Noninvasive Interstitial Glucose Estimation Using Wearables and Machine Learning in Healthy Individuals and Individuals With Obesity: Observational Cohort Study
JMIR Mhealth Uhealth 2026;14:e91724
URL: https://mhealth.jmir.org/2026/1/e91724
doi:10.2196/91724

© Franziska Schmelter, Xinyu Huang, Annika Heidmann, Paul Beier, Friederike Miriam Rundfeldt, Christina Plöger, Artur Piet, Md Abid Hasan, Yuanheng Zhang, Lennart Jablonski, Oliver Witt, Torsten Schröder, Stefanie Derer, Chris Easthope Awai, Marcin Grzegorzek, Christian Sina. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 30.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Reducing Low Anterior Resection Syndrome After Low Rectal Cancer Surgery Using an Integrated Intra-Anal Balloon Training and Reminder Device: Feasibility Nonrandomized Controlled Trial

Qing Zhang*, MN; Yanjun Wang*, MN; Meiling Wang, BN; Haiyan Hu, BN; Quan Wang, MD; Yuchen Guo, MD; Jianan Sun*, MN; Xuan Sun*, MD

Department of Gastrocolorectal Surgery, General Surgery Center, The First Hospital of Jilin University, No. 1 Xinmin Street, Chaoyang District, Changchun, Jilin, China

*these authors contributed equally

Corresponding Author:
Jianan Sun, MN
Department of Gastrocolorectal Surgery, General Surgery Center, The First Hospital of Jilin University, No. 1 Xinmin Street, Chaoyang District, Changchun, Jilin, China

Abstract

Background: Low anterior resection syndrome (LARS) is a common postoperative complication in patients with low rectal cancer, presenting with a spectrum of bowel dysfunction symptoms, including urgency, incontinence, evacuation disorders, and changes in stool frequency. Pelvic floor muscle training (PFMT) can alleviate LARS, but its effectiveness may be limited by poor accuracy of technique and low adherence during home-based training due to a lack of real-time feedback and monitoring devices.

Objective: This study aimed to evaluate the effects of a novel integrated balloon biofeedback device for home-based PFMT on accuracy of technique, adherence, quality of life, and LARS reduction in patients after sphincter-preserving surgery for low rectal cancer.

Methods: A nonrandomized controlled trial was conducted among 164 patients with low rectal cancer who underwent temporary ileostomy without neoadjuvant therapy. Participants were assigned by surgical date to an intervention group (n=82) using an adjustable-pressure balloon device with real-time waveform feedback via a mobile app or a control group (n=82) performing PFMT without equipment. PFMT was initiated within 72 hours after temporary ileostomy and continued throughout the stoma period until 1 month after ileostomy reversal. Outcomes including PFMT accuracy of technique, adherence, LARS score and incidence, and quality of life (European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30) were assessed 1 month after stoma reversal.

Results: At baseline, major LARS was present in 47.6% (39/82) of patients in the intervention group and 62.2% (51/82) of patients in the control group. Compared with the control group, the intervention group showed significantly higher PFMT training accuracy and adherence (P<.001 in both cases). LARS scores were significantly lower in the intervention group (median 16.00, IQR 11.00-29.00 vs 32.00, IQR 13.75-36.00), with a markedly reduced proportion of major LARS (17/82, 20.7% vs 45/82, 54.9%; P<.001). Global health/quality of life scores were significantly higher in the intervention group (P<.001).

Conclusions: The integrated balloon biofeedback device improved the accuracy of technique and adherence to home-based PFMT, reduced the incidence and severity of LARS, and enhanced quality of life in patients after low rectal cancer surgery. These findings support further development and clinical implementation of the device. However, the nonrandomized, time-sequenced study design and baseline differences between groups may limit causal interpretation of the results, and randomized controlled trials with longer follow-ups are needed to confirm long-term efficacy.

(JMIR Mhealth Uhealth 2026;14:e91332) doi:10.2196/91332

KEYWORDS

pelvic floor muscle training; rectal cancer surgery; low anterior resection syndrome; quality of life; biofeedback device

Introduction

Rectal cancer is one of the most common malignancies [1], with low rectal cancer accounting for approximately 75% of rectal cancers. In recent years, with the continuous improvement in surgical techniques, preserving the anal sphincter to avoid permanent ostomy has become the primary choice for more patients with low rectal cancer [2]. However, the reality is that approximately 90% of patients face a series of intestinal dysfunctions after surgery, known as low anterior resection

XSL•FO
RenderX

syndrome (LARS), characterized by unpredictable features such as frequent bowel movements, urgency, and fecal incontinence [3,4], which seriously affect patients' social and quality of life [5]. Currently, the incidence of major LARS is close to 50%, with half of the patients experiencing symptoms for up to 15 years [6].

Pelvic floor muscle training (PFMT), as one of the important interventions for postoperative pelvic floor rehabilitation in patients with rectal cancer, works through bidirectional regulation of pelvic floor muscle function. PFMT works by increasing the voluntary contractile capacity of the external anal sphincter and levator ani muscles. PFMT improves fecal incontinence symptoms in patients with LARS; conversely, in cases of pelvic floor hypertonicity or spasm, PFMT alleviates symptoms by relaxing the pelvic floor muscles, effectively reducing urgency and spasmodic bowel symptoms [7]. By restoring mechanical balance to the pelvic floor muscles and optimizing anatomical structures such as the anorectal angle, PFMT improves postoperative LARS symptoms and quality of life in patients with rectal cancer [8-10]. PFMT is recommended by many scholars for its advantages of being noninvasive, requiring no additional equipment, being low cost, and allowing patients to exercise at home [9,11]. However, studies have found that PFMT adherence is relatively low [12]. Moreover, research indicates that many individuals struggle to correctly perform pelvic floor muscle contractions. In a study on postmenopausal women, 32% were unable to correctly perform their first pelvic floor muscle contraction without guidance [13]. Similarly, a study of 82 women found that, although 98.8% believed that they could voluntarily contract their pelvic floor muscles, only 33% accurately estimated their contraction when assessed by a physiotherapist [14]. Even in women with no or mild pelvic floor disorders, 14% to 31% could not correctly contract on first attempt, particularly those with pelvic organ prolapse [15]. For patients who have undergone low anterior resection—where pelvic floor anatomy and innervation may be compromised—these challenges are likely even more pronounced. The lack of effective and objective real-time feedback may be a key bottleneck restricting the efficacy of home-based PFMT.

To address this bottleneck, we have developed a novel home-use integrated pelvic floor balloon biofeedback training device. Its design is based on a clear physiological rationale: active contraction of the pelvic floor muscles increases anal canal pressure, which is a key mechanism for maintaining fecal continence [16]. In addition to improving voluntary contraction during training, PFMT may induce muscle hypertrophy, increasing the bulk of the muscles around the anal canal and providing greater structural support even at rest [17]. The device places a balloon sensor in the anorectal ring region to directly capture the specific pressure signals generated through pelvic floor muscle contraction. Through real-time pressure-waveform conversion and intelligent algorithm analysis, it provides users with immediate, objective feedback on the accuracy of technique of their contractions. This process establishes a complete biofeedback loop, and research has confirmed that transforming intrinsic physiological signals into intuitive visual cues can effectively guide patients in identifying and regulating pelvic floor muscle function [18]. On the basis of this, we hypothesized that this device, through targeted biomechanical signal-based precise biofeedback, could guide patients in performing correct home-based PFMT; improve the accuracy of technique of movements and training adherence; and, ultimately, enhance the neuromuscular control capacity of the pelvic floor muscles to improve postoperative bowel function and reduce the incidence and severity of LARS. This study aimed to evaluate the effects of this device on PFMT outcomes and the occurrence and progression of LARS at 1 month postoperatively in patients who have undergone sphincter-preserving surgery for low rectal cancer.

Methods

Participants

Patients with rectal cancer admitted to the Department of Gastrointestinal Surgery at the First Hospital of Jilin University from October 2022 to March 2023 were enrolled as the control group, whereas patients admitted from April 2023 to September 2023 were enrolled as the intervention group. Inclusion criteria were (1) age of 18 years or above, (2) pathological diagnosis of low rectal cancer (tumor edge located ≤5 cm from the dentate line) and confirmed through pelvic magnetic resonance imaging; (3) having undergone temporary ileostomy at the time of rectal surgery, followed by ileostomy reversal surgery 3 months postoperatively; (4) consciousness and ability to communicate without barriers; and (5) independence in daily living and ability to understand and consent to participate in the study. Exclusion criteria were (1) distant metastasis; (2) concurrent tumors in other organs; (3) preoperative presence of severe conditions such as severe hemorrhoids, anal fissures, anal fistulas, rectal prolapse, irritable bowel syndrome, Crohn disease, and ulcerative colitis; (4) having received preoperative neoadjuvant therapy; (5) unfeasibility of temporary ileostomy reversal; (6) postoperative complications related to anastomotic leakage; and (7) voluntary withdrawal from treatment, death during the study, or loss to follow-up. The eligibility criteria above were applied uniformly to both the control and intervention groups prior to allocation.

Intervention Methods

All PFMT guidance and follow-up in both groups were provided by specialized pelvic floor rehabilitation nurses with postgraduate qualifications in pelvic floor rehabilitation and a mean of 13 (SD 8) years of clinical experience in this field.

Control Group

The control group received routine PFMT guidance and follow-up. First, during the preoperative stage of temporary ileostomy surgery, specialized pelvic floor rehabilitation nurses explained the potential hazards of LARS to patients and introduced PFMT methods; PFMT was initiated within 72 hours after temporary ileostomy and continued throughout the stoma period until 1 month after ileostomy reversal. Patients also received lifestyle advice, including regarding adequate fluid intake, a high-fiber diet to prevent constipation, and avoidance of straining during bowel movements. Second, personalized PFMT guidance was provided to patients through educational

XSL•FO
RenderX

videos, manuals, and classes. Patients were instructed to contract the pelvic floor muscles as if "stopping the passage of stool" or "holding in gas," lifting the muscles upward and inward. Each contraction was held for 3 to 10 seconds (initially based on patients' self-reported ability and subsequently adjusted based on the 1-month biofeedback assessment), followed by complete relaxation for an equal period. Patients were advised to avoid holding their breath or contracting abdominal muscles and were registered for an electronic training diary. Third, during the period from discharge to 1 month after temporary ileostomy reversal surgery, follow-up nurses sent training voice messages and videos to patients daily through individual WeChat groups, requiring patients to complete 30 minutes of PFMT daily, which could be done multiple times to achieve the target volume. Follow-up nurses were responsible for supervising patients in filling out the electronic training diary. Motivation strategies included daily reminders and goal setting (completing 30 minutes daily). Fourth, during the outpatient review 1 month postoperatively, specialized pelvic floor rehabilitation nurses assessed patients' pelvic floor muscle function using a biofeedback device (SA9800; Thought Technology Ltd; see the section PFMT Accuracy of Technique for a detailed device description) to determine each patient's baseline maximum voluntary contraction, which informed the starting level for ongoing training; provided targeted PFMT re-education based on the assessment results; and emphasized once again that the training duration was until 1 month after temporary ileostomy reversal surgery. Patients were instructed to progress their training by (1) increasing hold duration from 3 to 10 seconds as tolerated, (2) increasing repetitions, (3) changing positions (lying, sitting, and standing), and (4) incorporating functional activities (eg, contracting during coughing). Progression was based on perceived ability and successful completion of the current level. Fifth, 1 month after temporary ileostomy reversal surgery, specialized pelvic floor rehabilitation nurses assessed patients' training adherence, accuracy of technique, occurrence of LARS, and quality of life in the outpatient setting. Sixth, based on the assessment results, further training and treatment recommendations were provided to patients.

Intervention Group

The intervention group received the following additional care beyond the control protocol: pelvic floor rehabilitation specialist nurses provided 2 guidance sessions. The first session was conducted preoperatively, introducing the concept and importance of PFMT. The second session occurred during hospitalization after temporary ileostomy surgery on home-based pelvic floor intra-anal balloon training and the operation of the integrated reminder device. Before initiation of balloon-assisted training, participants in the intervention group underwent a 1-month postoperative digital rectal examination (DRE) as a safety prerequisite for device use. This assessment was not part of the baseline eligibility criteria.

The intervention was implemented in 2 distinct phases. Phase 1 was early PFMT (noninvasive). Patients commenced standard, noninvasive pelvic floor muscle contractions and relaxation exercises (ie, Kegel exercises) within 72 hours after temporary ileostomy surgery. This phase aimed to prevent muscle disuse atrophy and promote neuromuscular re-education without any device insertion or mechanical contact with the nascent anastomosis. Phase 2 was intra-anal balloon training (device assisted). The initiation of this phase was strictly contingent upon passing a safety assessment conducted during the 1-month postoperative outpatient follow-up. Physicians assessed anastomotic healing status through (1) DRE to confirm anastomotic patency (defined as the passage of the distal phalanx of the index finger); (2) endoscopic evaluation to exclude mucosal erythema, ulceration, or fibrotic stenosis; and (3) pain assessment using a visual analog scale (score of <4) induced via contact with the anastomosis during DRE. This safety assessment was performed only for patients in the intervention group to determine their eligibility to begin balloon-assisted training and did not influence participant allocation, which had already been completed prior to the assessment. Only patients demonstrating satisfactory healing based on all the criteria above received reinforced guidance on home-based pelvic floor intra-anal balloon training and device use, officially initiating the intra-anal balloon training regimen. Patients were instructed to progress their training by (1) increasing hold duration from 3 to 10 seconds as tolerated, (2) increasing repetitions, (3) changing positions (lying to sitting to standing), and (4) incorporating functional activities (eg, contracting during coughing). Progression was based on perceived ability and successful completion of the current level.

The development of the intra-anal balloon training and reminder integrated device was based on 2 previously approved patents: one for a PFMT device inserted anally (patent ZL 2017 2 1594601.5) and another for a PFMT device displaying pressure values (patent ZL 2021 2 0383459.X). The device consists of 2 main components: an adjustable-pressure intra-anal balloon training device, with hardware including a cylindrical balloon and an inflation handle (as shown in Figure 1). The cylindrical balloon is made of medical-grade silicone material, has passed biocompatibility testing, and is designed to be soft and comfortable and conform to the intestinal wall. It is adaptable and retractable, with a diameter of 2.5 cm and a cylindrical length of 8 cm. Health care providers set the inflation threshold not exceeding 50 mm Hg based on the patient's initial sensation threshold determined through anorectal manometry. The inflation handle contains an inflation pump, solenoid valve, and pressure sensor (as shown in Figure 2). The device uses pressure exchange and solenoid valve control principles to monitor pressure, control inflation, and compress the balloon. The second component of the device is the companion app, which connects to the intra-anal balloon training device via Bluetooth. Upon log-in, the app automatically switches to the inflation interface, followed by the training interface after inflation, displaying real-time pressure waveforms during training (as shown in Figure 3). It should be emphasized that the pressure signals recorded by this device possess a clear physiological origin and specificity. When the balloon is positioned in the anorectal ring region, the perceived pressure changes mainly originate from the active contraction of the pelvic floor muscle complex, particularly the direct radial compression exerted by the external anal sphincter and the synergistic upward traction provided by the puborectalis muscle. A correct and coordinated pelvic floor muscle contraction is manifested on the waveform as a pressure curve characterized by a gentle rise, stable maintenance, and

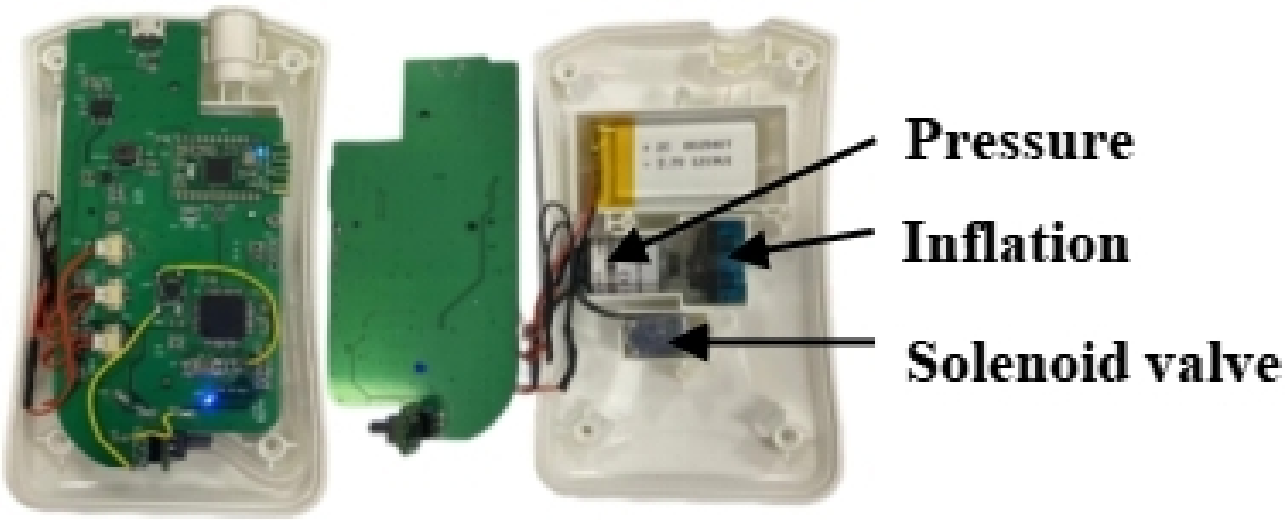
XSL•FO
RenderX

gradual decline, reflecting effective force generation by the pelvic floor muscles themselves. Conversely, if the waveform exhibits a sharp rise and fall, lacks a stable plateau, or displays jagged fluctuations, it suggests possible compensatory contractions such as those of the abdominal muscles leading to intra-abdominal pressure transmission, which represents an erroneous pattern to be avoided. Our hypothesis is that this device not only provides real-time pressure monitoring but also, through the recognition and feedback of waveform characteristics, guides patients to achieve precise, pelvic floor muscle–dominated contractions that are independent of abdominal pressure interference during training, thereby enhancing neuromuscular coordination between pelvic floor muscle contraction and abdominal pressure regulation.

Figure 1. Intra-anal balloon training device.



Figure 2. Internal structure of the inflation handle.



XSL•FO
RenderX

Figure 3. (A) App log-in screen, (B) inflation interface, and (C) training interface.



(A) (B) (C)

The specific instructions for using the device were as follows. Initially, the specialized pelvic floor rehabilitation nurse adjusted the training inflation threshold for each patient based on preoperative anorectal manometry results obtained using a clinical anorectal manometry system (Hefei Aoyuan Technology Development Co, Ltd; model ZGJ-D3), and guided the patient to install the training software and enter the training interface. The patient was then placed in the left lateral position. After lubricating the balloon with liquid paraffin, it was gently and slowly inserted through the anus until the end of the cylindrical balloon was positioned within the anal canal region (approximately 1-2 cm from the anal verge). The patient then assumed a supine position in preparation for training. To start the training, the patient clicked the Inflate button, and the balloon inflated to the set pressure threshold. After inflation stopped, clicking Start prompted the patient to perform pelvic floor muscle contractions and relaxations following voice prompts. During the Train and Hold process, the patient contracted the anus forcefully for 10 seconds, mimicking the act of holding in a bowel movement. The pressure exerted by the wall of the anal canal on the balloon during this contraction was monitored, and when the pressure exceeded the inflation threshold, the device emitted a beep to indicate successful training. Throughout the training, the pressure sensor provided feedback to the mobile app, displaying real-time pressure values and waveform trends during the contraction and relaxation phases. The patient could adjust the training intensity based on the numerical and waveform prompts. After the voice prompt indicated “training complete,” the balloon was deflated and then slowly withdrawn from the anus. Additionally, it is important to note that the intra-anal balloon training and reminder integrated device used in this study is a self-designed prototype. This device is not a commercially available product and has not been approved by the Food and Drug Administration for any purpose. Its use in this study was investigational and specifically designed to evaluate its potential in guiding and monitoring PFMT for patients with low rectal cancer.

The research team monitored protocol adherence through regular review of electronic training diaries and WeChat communication records. No major protocol deviations were reported.

Evaluation Indicators

PFMT Accuracy of Technique

At the 1-month follow-up after temporary ileostomy reversal surgery, the accuracy of technique of the patients’ training was assessed using a biofeedback device (SA9800 by Thought Technology Ltd). The biofeedback device was connected to the patients’ abdominal electrodes to collect electromyographic signals and display them in real time on a computer screen. The waveform moved upward with abdominal exertion and downward with relaxation. If the patient exerted excessive abdominal force beyond the threshold, a blue threshold line appeared on the waveform. During the 10-minute training session, the total time the blue threshold line was present represented incorrect training time, whereas the rest indicated correct training time (accuracy of technique = correct training time/10 minutes × 100%).

Adherence to PFMT

The daily training time of patients in both groups was obtained through self-recorded electronic training diaries. The daily training target duration was 30 minutes (daily training adherence = total training time on a given day/30 minutes × 100%). If the training duration exceeded 30 minutes, adherence was set at 100%. The highest adherence value was 100%. The average adherence during the 3-month training period following temporary ileostomy was calculated based on daily adherence records.

XSL•FO
RenderX

LARS Score

The LARS score developed by Emmertsen and Laurberg [19] in 2012 was used for evaluation. This score is specifically designed to assess bowel function in patients after low anterior resection surgery, identifying the occurrence and severity of LARS. The LARS score consists of 5 items: incontinence to gas, incontinence to liquid stool, bowel frequency, repeated defecation within 1 hour (clustering of stool), and fecal urgency. The total score is 42, with higher scores indicating worse bowel function. LARS is classified as major (score of ≥30), minor (score of 21-29), and no LARS (score of ≤20).

Quality of Life

The European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 [20] was used to assess the quality of life of the 2 groups of patients at 1 month postoperatively. This questionnaire includes 5 subscales for physical, role, cognitive, emotional, and social functioning; 3 symptom subscales for pain, fatigue, and nausea or vomiting; 6 single-item scales for dyspnea, appetite loss, insomnia, constipation, financial difficulties, and diarrhea; and 1 overall quality of life scale. Each item is scored from 0 to 100, with higher scores indicating a higher level of quality of life.

Adverse Events

Adverse events were defined as any device-related complications, including but not limited to anal pain, bleeding, mucosal injury, or infection. Patients in both groups were instructed to report any discomfort, pain, or unusual symptoms during the training period at each follow-up contact (including daily WeChat follow-ups and outpatient visits). All reported events were recorded and assessed by the research team for severity and causality.

Sample Size Calculation

This study was a nonrandomized controlled trial. The intervention group underwent PFMT using the intra-anal balloon training and reminder integrated device, whereas the control group received routine PFMT. The primary outcome measure was the LARS incidence rate at 1 month postoperatively. On the basis of prior experimental results, the LARS incidence rate in the control group was approximately 75%. We expected the LARS incidence rate in the intervention group to decrease to 50%. With a 2-sided α of .05 and a power of 90%, using the PASS software (version 15; NCSS LLC), the calculated sample size was 74 patients for the treatment group and 74 patients for the control group. Considering a 10% attrition and refusal rate, the final required sample size was at least 82 participants in each group, totaling at least 164 study participants.

Statistical Methods

Statistical analyses were performed using SPSS (version 21.0; IBM Corp). Categorical variables are presented as frequencies and percentages and were compared between groups using the chi-square test. Continuous variables were assessed for normality using the Shapiro-Wilk test. Normally distributed data are expressed as means and SDs and were compared using the independent-sample 2-tailed *t* tests, whereas nonnormally distributed data are presented as medians and IQRs and were compared using the Mann-Whitney *U* test. For the comparison of postoperative LARS scores between groups, the Mann-Whitney *U* test was used because both baseline and postoperative LARS scores were not normally distributed (Shapiro-Wilk test; $P<.001$). For the comparison of LARS severity categories (no LARS, minor, or major), a multivariable ordinal logistic regression model was applied with adjustment for baseline LARS scores. The proportional odds assumption was assessed using the parallel lines test. A 2-sided *P* value of less than .05 was considered statistically significant.

Ethical Considerations

This study was reviewed and formally approved by the ethics committee of the First Hospital of Jilin University (21K028-001). Written informed consent was obtained from all individual participants included in this study. All participant data were anonymized and handled confidentially in accordance with institutional and ethical guidelines. No financial compensation was provided to participants for their involvement in this study.

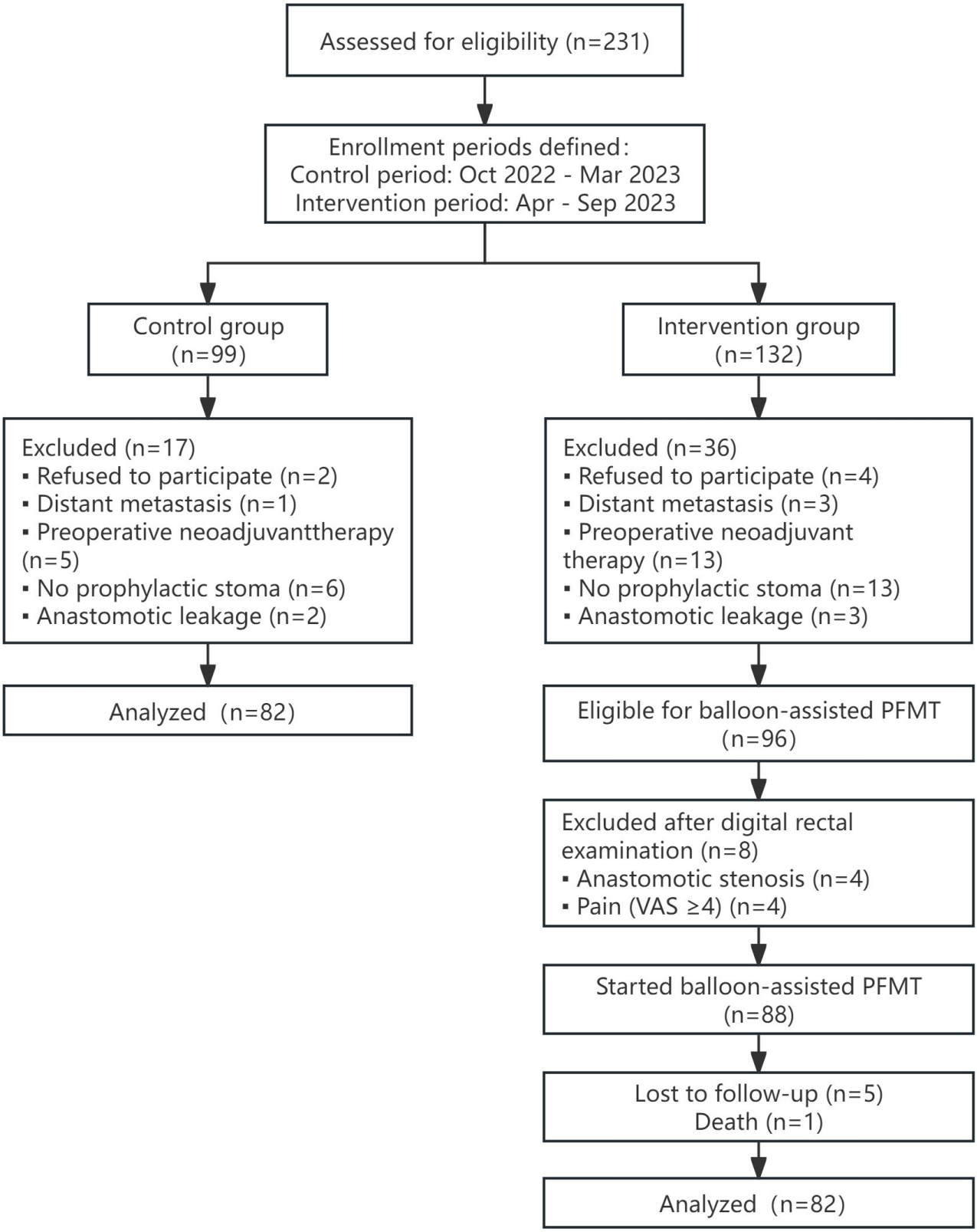
Results

Comparison of General Data Between the Two Groups of Patients

From October 2022 to September 2023, a total of 231 patients with low rectal cancer were recruited and assessed for eligibility in this study. The participant enrollment process is detailed in Figure 4. In the control group, a total of 99 patients were screened, with 17 (17.2%) excluded, whereas in the intervention group, of the 132 patients screened, 36 (27.3%) were excluded based on baseline eligibility criteria. Of the remaining 96 patients eligible for balloon-assisted PFMT in the intervention group, 8 (8.3%) did not meet the postoperative safety criteria prior to intervention initiation. During follow-up, of the remaining 88 patients, 5 (5.7%) were lost to follow-up, and 1 (1.1%) died. Ultimately, 82 patients were included in the final analysis in the intervention group. Overall, 164 eligible patients were included in the study, with 82 (50%) in the intervention group and 82 (50%) in the control group. The 2 groups demonstrated comparable baseline characteristics, with no significant differences in demographic or clinical data ($P>.05$ in all cases). Additionally, baseline LARS scores prior to temporary stoma surgery showed no statistically significant intergroup discrepancy ($P=.09$). Detailed baseline data are shown in Table 1.

XSL•FO
RenderX

Figure 4. Flowchart of patient enrollment, exclusion, and allocation. PFMT: pelvic floor muscle training; VAS: visual analog scale.



XSL•FO
RenderX

Table . Comparison of the general information of patients in the 2 groups.

Variables	Control group (n=82)	Intervention group (n=82)	*t* test (*df*), *Z* score, or chi-square (*df*)	*P* value
Age (y), mean (SD)	58.43 (9.34)	58.82 (10.64)	t_{162}=−0.25	.80
Sex, n (%)			χ^2_1=0.1	.87
Male	56 (68.3)	58 (70.7)		
Female	26 (31.7)	24 (29.3)		
Pathological stage, n (%)			χ^2_2=0.8	.68
1	25 (30.5)	29 (35.4)		
2	29 (35.4)	24 (29.3)		
3	28 (34.1)	29 (35.4)		
Anastomotic technique, n (%)			χ^2_1=0.5	.68
Handsewn anastomosis	13 (15.9)	16 (19.5)		
Stapled anastomosis	69 (84.1)	66 (80.5)		
Place of residence, n (%)			χ^2_1=0.1	.87
Urban	51 (62.2)	49 (59.8)		
Rural	31 (37.8)	33 (40.2)		
Educational level, n (%)			χ^2_2=0.7	.71
Primary school or lower	34 (41.5)	30 (36.6)		
Junior high school, high school, or secondary school	39 (47.6)	40 (48.8)		
College or higher	9 (11.0)	12 (14.6)		
LARS[a] score (0-42), median (IQR)	29.96 (27.00-39.00)	27.35 (22.25-34.00)	Z=−1.7	.09
LARS degree rating, n (%)			χ^2_2=3.6	.17
No	14 (17.1)	20 (24.4)		
Mild	17 (20.7)	23 (28.0)		
Severe	51 (62.2)	39 (47.6)		

[a]LARS: low anterior resection syndrome.

Adverse Events

No device-related adverse events were reported in either group throughout the study period. Specifically, no patients experienced anal bleeding, severe pain (visual analog scale score of ≥4), mucosal injury, or infection associated with PFMT. All patients in both groups completed the training protocol without premature discontinuation due to adverse effects.

Comparison of PFMT Accuracy of Technique, Adherence, and LARS Scores Between the Two Groups of Patients

Patients in the intervention group demonstrated significantly superior training accuracy and adherence compared to the control group (*P*<.001 in both cases; Table 2). Postoperative LARS scores were significantly lower in the intervention group than in the control group as assessed using the Mann-Whitney *U* test (*Z*=−3.491; *P*<.001). Furthermore, the distribution of LARS severity differed significantly between the 2 groups (χ^2_2=20.8; *P*<.001), with a higher proportion of patients without LARS and a lower proportion of severe LARS observed in the intervention group (Table 2). In addition, a multivariable-adjusted ordinal logistic regression model (accounting for baseline LARS scores) revealed that the intervention group had a markedly reduced risk of developing major LARS (odds ratio 0.27, 95% CI 0.15 - 0.50; *P*<.001), as detailed in Table 3.

XSL•FO
RenderX

Table . Comparison of pelvic floor muscle training and low anterior resection syndrome (LARS) scores at 1 month after stoma reversal surgery between the 2 groups of patients.

Variables	Control group (n=82)	Intervention group (n=82)	*Z* score or chi-square (*df*)	*P* value
Accuracy of technique (%), median (IQR)	68.00 (56.00-80.25)	84.22 (73.00-98.50)	Z=−5.830	<.001
Adherence (%), median (IQR)	65.08 (50.00-84.17)	87.68 (83.33-100.00)	Z=−6.200	<.001
LARS score (0-42), median (IQR)	32.00 (13.75-36.00)	16.00 (11.00-29.00)	Z=−3.491	<.001
LARS degree rating, n (%)			χ^2_2=20.8	<.001
No	23 (28.0)	45 (54.9)		
Mild	14 (17.1)	20 (24.4)		
Severe	45 (54.9)	17 (20.7)		

Table . Multivariable ordinal logistic regression analysis of factors associated with low anterior resection syndrome (LARS) severity at 1 month after stoma reversal surgery. The model passed the parallel lines test (*P*=.40), satisfying the proportional odds assumption.

Variable	β (SE)	OR[a] (95% CI)	*P* value
Group (intervention)	−1.310 (0.313)	0.27 (0.15 - 0.50)	<.001
Baseline LARS score	0.051 (0.014)	1.05 (1.02 - 1.08)	<.001

[a]OR: odds ratio.

Comparison of Quality of Life Scores Between the Two Groups of Patients

At 1 month after temporary ileostomy reversal surgery, patients in the intervention group had significantly higher scores in physical functioning, role functioning, social functioning, and overall quality of life than the control group. Scores for insomnia, constipation, and diarrhea were significantly lower in the intervention group than in the control group, and the score on the economic difficulties dimension in the intervention group was notably lower than in the control group, as shown in Table 4.

Table . Comparison of European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 scores between the control and intervention groups at 1 month after stoma reversal surgery[a].

Variable (0-100)	Control group (n=82), median (IQR)	Intervention group (n=82), median (IQR)	*Z* score	*P* value
Physical functioning	80 (73.33-86.67)	93.33 (86.67-100)	−6.303	<.001
Role functioning	66.67 (66.67-83.33)	83.33 (66.67-100)	−2.234	.03
Emotional functioning	91.67 (89.58-91.67)	91.67 (83.33-100)	−1.694	.09
Cognitive functioning	100 (100-100)	100 (100-100)	−1.821	.07
Social functioning	66.67 (66.67-66.67)	66.67 (66.67-100)	−2.948	.003
Global health and quality of life	75 (66.67-83.33)	83.33 (75-83.33)	−4.551	<.001
Fatigue	0 (0-11.11)	0 (0-0)	−1.924	.05
Nausea or vomiting	0 (0-0)	0 (0-0)	−1.743	.08
Pain	0 (0-0)	0 (0-0)	−0.672	.50
Dyspnea	0 (0-0)	0 (0-0)	−0.342	.73
Insomnia	0 (0-33.33)	0 (0-0)	−3.469	.001
Appetite loss	0 (0-0)	0 (0-0)	−1.108	.27
Constipation	0 (0-0)	0 (0-0)	−2.019	.04
Diarrhea	0 (0-33.33)	0 (0-0)	−3.308	.002
Financial difficulties	33.33 (33.33-33.33)	0 (0-33.33)	−4.694	<.001

[a]If both the 25th and 75th percentiles are 0, it indicates that at least 75% of the patients scored 0 on that item.

XSL•FO
RenderX

Discussion

Principal Findings

This study aimed to investigate the impact of a novel home-based integrated device for pelvic floor intra-anal balloon training and reminders on postoperative LARS in patients with low rectal cancer. The results showed that using this device to guide and monitor patients in home-based PFMT significantly improved the accuracy of technique and adherence to patient training, thereby reducing the incidence and severity of LARS at 1 month after temporary ileostomy reversal surgery and improving patients' quality of life.

PFMT involves consciously and autonomously contracting the levator ani and anal sphincter muscles to increase the stiffness of the pelvic floor muscle group and reduce anal descent. This approach helps improve bowel control by enhancing voluntary sphincter contraction, which may indirectly improve rectal reservoir function and is beneficial in repairing injuries to the internal and external anal sphincters and perianal nerves following surgery [9,21]. PFMT is simple to learn and not limited by time or location and does not increase patient discomfort. Several studies have demonstrated the effectiveness of PFMT in improving symptoms of LARS in patients after rectal cancer surgery [22].

Intra-anal balloon training, as an auxiliary method for PFMT, can help patients better perceive and exercise the pelvic floor muscles, providing real-time feedback that allows patients to perform muscle contractions and relaxations more accurately, thus achieving better exercise results. However, research on the use of intra-anal balloon training in PFMT is still relatively limited [23], and there is currently a lack of commercially available balloon devices that can be widely used.

In this study, the design and use of the intra-anal balloon training and reminder integrated device provided an innovative approach to the rehabilitation of patients with low rectal cancer. By combining an adjustable-pressure balloon device with a dynamic monitoring device, patients can perform PFMT at home and ensure the accuracy of technique and adherence to the training through real-time monitoring and feedback. In this study, comprehensive and rich health education content related to LARS prevention for patients with mid-low rectal cancer was provided, and postoperative daily follow-ups and PFMT were widely and meticulously promoted. The results showed that adherence was higher in the intervention group than in the control group (median 87.68%, IQR 83.33%-100.00% vs 65.08%, IQR 50.00%-84.17%; $P<.001$) and much higher than previously reported rates [12]. After adjusting for baseline LARS scores, the intervention was associated with a 73% reduction in the risk of major LARS (odds ratio 0.27, 95% CI 0.15 - 0.50), suggesting that the device has significant clinical implications for improving postoperative bowel function.

In summary, the results of this study suggest that the self-designed intra-anal balloon training and reminder integrated device enables self-management of home-based PFMT and has potential clinical applications in preventing and improving LARS in patients with low rectal cancer, warranting widespread clinical promotion.

Limitations

This study is the first nonrandomized controlled trial integrating intra-anal balloon training with a reminder device for the prevention of LARS. To minimize group contamination caused by cross-contact between concurrent patients, a temporal isolation strategy was adopted. In addition, the interventions were standardized through systematic training by the same nursing team. Nevertheless, several potential biases should be considered. First, the nonrandomized, time-sequenced design may introduce unmeasured time-related confounding (eg, potential differences in surgical techniques or perioperative care over the recruitment period). In addition, the exclusion rate differed between the 2 recruitment periods, which may have contributed to baseline differences in functional status between groups. Second, although the analysis of baseline characteristics and intensive follow-up management in the control group suggested comparability between groups, the biofeedback devices used in the intervention group may introduce a slight Hawthorne effect bias by objectively quantifying training parameters. Third, training adherence in both groups was manually recorded via self-reported electronic diaries, which may have introduced self-reporting bias. The combination of this self-reporting bias and the Hawthorne effect from intensive follow-up and device feedback may have led to overestimation of adherence in the intervention group, potentially exaggerating the observed differences between groups. Fourth, this study did not perform a strict intention-to-treat analysis as participants who were lost to follow-up were excluded from the final analysis. Although the overall attrition rate was low, this may introduce potential attrition bias and should be considered when interpreting the results.

Additionally, the single-center sample source and the relatively short follow-up period represent important limitations of this study. Outcomes were assessed only 1 month after temporary ileostomy reversal, which mainly reflects early postoperative bowel function recovery rather than long-term functional outcomes. Previous randomized controlled trials have reported inconsistent evidence regarding the long-term effectiveness of PFMT for LARS, with some studies showing only transient benefits over time. Therefore, the findings of this study should primarily be interpreted as indicating a potential benefit in early functional recovery rather than definitive long-term improvement. Furthermore, given that neoadjuvant therapy (particularly radiotherapy) may lead to anastomotic tissue fibrosis, reduced blood supply, delayed healing, and potential complications (eg, bleeding or dehiscence) when initiating intra-anal balloon training 1 month after stoma reversal, we prioritized validating the device's efficacy in low-risk patients who did not receive neoadjuvant radiotherapy to ensure intervention safety, and therefore, the results cannot be generalized to higher-risk patient groups. Future phase 2 studies will expand to high-risk populations undergoing neoadjuvant therapy. To address these limitations, we plan to conduct a multicenter stepped-wedge cluster randomized controlled trial that includes patients receiving neoadjuvant therapy, analyzes

XSL•FO
RenderX

the association between radiotherapy dose and type and LARS, and further evaluates the device's clinical utility.

Funding

This work was supported by the Jilin Provincial Department of Finance (JLSWSRCZX2023-102).

Data Availability

The data used to support the results of this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: QZ, YW, QW, MW, YG, HH, JS, XS
Methodology: QZ, YW, QW, MW
Software: QZ, YW
Investigation: YG, HH
Project administration: JS, XS
Funding acquisition: JS, XS
Writing—original draft: QZ
Writing—review and editing: QW, MW, YG, HH, JS, XS
All authors contributed to the preparation of the manuscript and approved the final version submitted.

Co-corresponding author: Xuan Sun, MD, Department of Gastrocolorectal Surgery, General Surgery Center, The First Hospital of Jilin University, No. 1 Xinmin Street, Chaoyang District, Changchun, Jilin 130021, China Phone: +86 13596161861 Email: xuansun@jlu.edu.cn

Conflicts of Interest

None declared.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021 May;71(3):209-249. [doi: 10.3322/caac.21660] [Medline: 33538338]
2. Wrenn SM, Cepeda-Benito A, Ramos-Valadez DI, Cataldo PA. Patient perceptions and quality of life after colon and rectal surgery: what do patients really want? Dis Colon Rectum 2018 Aug;61(8):971-978. [doi: 10.1097/DCR.0000000000001078] [Medline: 29944576]
3. van Heinsbergen M, den Haan N, Maaskant-Braat AJ, et al. Functional bowel complaints and quality of life after surgery for colon cancer: prevalence and predictive factors. Colorectal Dis 2020 Feb;22(2):136-145. [doi: 10.1111/codi.14818] [Medline: 31397962]
4. Pieniowski EHA, Palmer GJ, Juul T, et al. Low anterior resection syndrome and quality of life after sphincter-sparing rectal cancer surgery: a long-term longitudinal follow-up. Dis Colon Rectum 2019 Jan;62(1):14-20. [doi: 10.1097/DCR.0000000000001228] [Medline: 30394987]
5. Burch J, Taylor C, Wilson A, Norton C. Symptoms affecting quality of life after sphincter-saving rectal cancer surgery: a systematic review. Eur J Oncol Nurs 2021 Jun;52:101934. [doi: 10.1016/j.ejon.2021.101934] [Medline: 33845303]
6. Chen TYT, Wiltink LM, Nout RA, et al. Bowel function 14 years after preoperative short-course radiotherapy and total mesorectal excision for rectal cancer: report of a multicenter randomized trial. Clin Colorectal Cancer 2015 Jun;14(2):106-114. [doi: 10.1016/j.clcc.2014.12.007] [Medline: 25677122]
7. Ansar M, Boddeti S, Noor K, et al. A systematic review of comparative effectiveness of interventions for low anterior resection syndrome: impacts on bowel function and quality of life. Cureus 2024 Oct;16(10):e72772. [doi: 10.7759/cureus.72772] [Medline: 39483606]
8. Yuanyuan W, Shiyin H, Lei H, Ding D. Pelvic floor muscle exercises alleviate symptoms and improve mental health and rectal function in patients with low anterior resection syndrome. Front Oncol 2023;13:1168807. [doi: 10.3389/fonc.2023.1168807] [Medline: 37152027]
9. Martellucci J. Low anterior resection syndrome: a treatment algorithm. Dis Colon Rectum 2016 Jan;59(1):79-82. [doi: 10.1097/DCR.0000000000000495] [Medline: 26651116]
10. Hung SL, Lin YH, Yang HY, Kao CC, Tung HY, Wei LH. Pelvic floor muscle exercise for fecal incontinence quality of life after coloanal anastomosis. J Clin Nurs 2016 Sep;25(17-18):2658-2668. [doi: 10.1111/jocn.13314] [Medline: 27461451]

XSL•FO
RenderX

11. Sakr A, Sauri F, Alessa M, et al. Assessment and management of low anterior resection syndrome after sphincter preserving surgery for rectal cancer. Chin Med J (Engl) 2020 Aug 5;133(15):1824-1833. [doi: 10.1097/CM9.0000000000000852] [Medline: 32604174]
12. Jeong H, Park J. Factors influencing changing bowel habits in patients undergoing sphincter-saving surgery for rectal cancer. Int Wound J 2019 Mar;16 Suppl 1(Suppl 1):71-75. [doi: 10.1111/iwj.13060] [Medline: 30793858]
13. Piernicka M, Ossowski Z, Kortas J, Bojar D, Labun J, Szumilewicz A. Can we improve the technique of pelvic floor muscle exercises in postmenopausal women using a single electromyography biofeedback session? An experimental study. J Clin Med 2024 May 23;13(11):3062. [doi: 10.3390/jcm13113062] [Medline: 38892773]
14. Uechi N, Fernandes A, Bø K, et al. Do women have an accurate perception of their pelvic floor muscle contraction? A cross-sectional study. Neurourol Urodyn 2020 Jan;39(1):361-366. [doi: 10.1002/nau.24214] [Medline: 31737927]
15. Henderson JW, Wang S, Egger MJ, Masters M, Nygaard I. Can women correctly contract their pelvic floor muscles without formal instruction? Female Pelvic Med Reconstr Surg 2013;19(1):8-12. [doi: 10.1097/SPV.0b013e31827ab9d0] [Medline: 23321652]
16. Padda BS, Jung SA, Pretorius D, Nager CW, Den-Boer D, Mittal RK. Effects of pelvic floor muscle contraction on anal canal pressure. Am J Physiol Gastrointest Liver Physiol 2007 Feb;292(2):G565-G571. [doi: 10.1152/ajpgi.00250.2006] [Medline: 17023551]
17. Bø K. Mechanisms for pelvic floor muscle training: morphological changes and associations between changes in pelvic floor muscle variables and symptoms of female stress urinary incontinence and pelvic organ prolapse-a narrative review. Neurourol Urodyn 2024 Nov;43(8):1977-1996. [doi: 10.1002/nau.25551] [Medline: 38979823]
18. Kim KH, Yu CS, Yoon YS, Yoon SN, Lim SB, Kim JC. Effectiveness of biofeedback therapy in the treatment of anterior resection syndrome after rectal cancer surgery. Dis Colon Rectum 2011 Sep;54(9):1107-1113. [doi: 10.1097/DCR.0b013e318221a934] [Medline: 21825890]
19. Emmertsen KJ, Laurberg S. Low anterior resection syndrome score: development and validation of a symptom-based scoring system for bowel dysfunction after low anterior resection for rectal cancer. Ann Surg 2012 May;255(5):922-928. [doi: 10.1097/SLA.0b013e31824f1c21] [Medline: 22504191]
20. Schurr T, Loth F, Lidington E, et al. Patient-reported outcome measures for physical function in cancer patients: content comparison of the EORTC CAT Core, EORTC QLQ-C30, SF-36, FACT-G, and PROMIS measures using the International Classification of Functioning, Disability and Health. BMC Med Res Methodol 2023 Jan 21;23(1):21. [doi: 10.1186/s12874-022-01826-z] [Medline: 36681808]
21. Keane C, Fearnhead NS, Bordeianou LG, et al. International consensus definition of low anterior resection syndrome. Dis Colon Rectum 2020 Mar;63(3):274-284. [doi: 10.1097/DCR.0000000000001583] [Medline: 32032141]
22. Asnong A, D'Hoore A, Van Kampen M, et al. The role of pelvic floor muscle training on low anterior resection syndrome: a multicenter randomized controlled trial. Ann Surg 2022 Nov 1;276(5):761-768. [doi: 10.1097/SLA.0000000000005632] [Medline: 35894434]
23. Bols E, Berghmans B, de Bie R, et al. Rectal balloon training as add-on therapy to pelvic floor muscle training in adults with fecal incontinence: a randomized controlled trial. Neurourol Urodyn 2012 Jan;31(1):132-138. [doi: 10.1002/nau.21218] [Medline: 22038680]

Abbreviations

DRE: digital rectal examination
LARS: low anterior resection syndrome
PFMT: pelvic floor muscle training

Edited by R Ramezani; submitted 13.Jan.2026; peer-reviewed by AD Rink, R Brennen; revised version received 01.Apr.2026; accepted 02.Apr.2026; published 08.May.2026.

Please cite as:
Zhang Q, Wang Y, Wang M, Hu H, Wang Q, Guo Y, Sun J, Sun X
Reducing Low Anterior Resection Syndrome After Low Rectal Cancer Surgery Using an Integrated Intra-Anal Balloon Training and Reminder Device: Feasibility Nonrandomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e91332
URL: https://mhealth.jmir.org/2026/1/e91332
doi:10.2196/91332

© Qing Zhang, Yanjun Wang, Meiling Wang, Haiyan Hu, Quan Wang, Yuchen Guo, Jianan Sun, Xuan Sun. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 8.May.2026. This is an open-access article distributed under the terms

XSL•FO
RenderX

of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Monitoring Respiratory Health in Children With Acute Asthma Using Wearable Electrical Bioimpedance and Breath Sounds: Observational Case-Control Study

Jesus Antonio Sanchez-Perez[1], PhD; John Berkebile[2], PhD; Natalie Jordan[3], MPH; Kevin Maher[4], MD; Omer Inan[2], PhD; Jocelyn Grunwell[4], MD, PhD

[1]Department of Electrical and Computer Engineering, College of Engineering, University of Puerto Rico-Mayaguez, Mayagüez, Puerto Rico

[2]School of Electrical and Computer Engineering, College of Engineering, Georgia Institute of Technology, Atlanta, GA, United States

[3]Division of Critical Care Medicine, Children's Healthcare of Atlanta, Atlanta, GA, United States

[4]Department of Pediatrics, School of Medicine, Emory University, Atlanta, GA, United States

Corresponding Author:
Jesus Antonio Sanchez-Perez, PhD
Department of Electrical and Computer Engineering
College of Engineering
University of Puerto Rico-Mayaguez
259 N Boulevard Alfonso Valdes
Stefani Building
Mayagüez, 00680
Puerto Rico
Phone: 1 787 832 4040
Email: jesus.sanchez25@upr.edu

Abstract

Background: Asthma remains one of the most serious chronic diseases of childhood. Individuals with severe asthma experience sudden episodes of breathlessness due to acute airflow obstruction, leading to recurrent pediatric intensive care unit (PICU) admissions that often result in mechanical ventilation and even death. Existing clinical assessments lack temporal resolution to effectively track the rapidly changing physiology.

Objective: This study aimed to evaluate the feasibility of quantifying respiratory health during acute asthma in children using wearable multimodal sensing.

Methods: Wearable-based impedance pneumography (IP) and multichannel lung sounds (LSs) were measured on 17 children admitted to the PICU with an acute asthma attack and on 9 healthy controls. Short-term multimodal measurements were obtained throughout hospitalization, specifically at PICU admission (T1) and discharge (T2). Measurements were also obtained from controls without any signs of acute asthma or otherwise healthy. Statistical and clustering analyses were performed to identify trends in IP- and LS-derived respiratory markers from T1 to T2 across all patients with paired time points (n=13), as well as across the matched cohort (T1: n=10 and T2: n=7), who were compared against controls (n=9). Five features were computed from the IP signal: respiratory rate, inspiration time (Ti), expiration time (Te), expiration-to-inspiration time ratio (Te:Ti), and the normalized Ti by interbreath interval (Ti/IBI). Leveraging the breathing context provided by the IP signal, 4 spectral integrated intensity (SI) acoustic features were computed in 4 different subbands for the inspiration and expiration phases.

Results: Within the patient group (n=13), we found that respiratory rate decreased ($W(12)=79$; $P=.02$), whereas Te ($W(12)=12$; $P=.02$) and Ti ($W(12)=13$; $P=.02$) lengthened. Meanwhile, the SIs for the lowest subband (100-300 Hz) decreased for both inspiration and expiration phases ($P<.01$), while they increased for the highest subband (800-1000 Hz) for both inspiration and expiration phases ($P<.01$). Significant differences also existed between T1 and control and T2 and control of the matched cohort. We found that all features were significantly different between T1 and control ($P<.05$), and all SIs together with Te:Ti and Ti/IBI were significantly different between T2 and control ($P<.05$), all exhibiting trends toward normalcy.

Conclusions: These results demonstrate the feasibility of quantifying and tracking respiratory health in children with acute asthma using wearable multimodal sensing, specifically with the fusion of IP- and LS-derived markers. Such technology may provide a new adjunctive clinical tool for real-time respiratory monitoring and enable timely titration of care, thereby improving patient outcomes.

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2026;14:e72979) doi:10.2196/72979

KEYWORDS

asthma; impedance pneumography; electrical bioimpedance; lung sounds; wearable; multimodal sensing

Introduction

Asthma affects more than 4.6 million (6.5%) of US children, resulting in more than 270,000 emergency department visits and 27,000 hospitalizations for status asthmaticus annually [1,2]. Individuals with asthma experience recurrent episodes of reversible airflow obstruction, resulting in breathlessness, chest tightness, wheezing, and coughing. Severe asthma is characterized by the sudden worsening of these symptoms (asthma attacks) and trapping of air inside the lungs (air trapping) due to bronchoconstriction [3]. While many children respond to emergency treatment with bronchodilators and systemic corticosteroids, others progress to respiratory failure and require admission to a pediatric intensive care unit (PICU). PICU admissions for respiratory failure have increased in proportion to general hospital admissions for asthma [4] for reasons that are unclear. Children with severe asthma can have recurrent PICU admissions due to exacerbations that may require mechanical ventilation and even result in death.

Management of asthma involves, among other strategies, the use of inhaled corticosteroids and β-agonist therapy to reduce airway inflammation and promote bronchodilation, thereby reversing airflow obstruction and improving symptoms [5]. Titration of care commonly relies on symptom reporting, physical examination, and pulmonary function tests [3,5]. However, these assessments are often highly subjective and lack the temporal resolution needed to capture the rapidly evolving physiology of a disease characterized by substantial interindividual variability in treatment response [6,7]. On physical examination, the hallmark of asthma monitoring is the manual evaluation of lung sounds (LSs) via auscultation with a stethoscope [3,8]. Unfortunately, this technique is limited by its manual, qualitative, and subjective nature [8]. Digital auscultation enables the quantitative acoustic analysis of LSs, but commercially available electronic stethoscopes do not provide breathing-phase context for post hoc analysis and only record for a short amount of time at a single location. Pulmonary function test via spirometry, on the other hand, requires patient cooperation for performing respiratory maneuvers, which may not be feasible in young children [9], and these diagnostic maneuvers cannot be performed during an acute exacerbation. Hence, there is a need for noninvasive sensing technologies that can consistently quantify asthma signs continuously and conveniently.

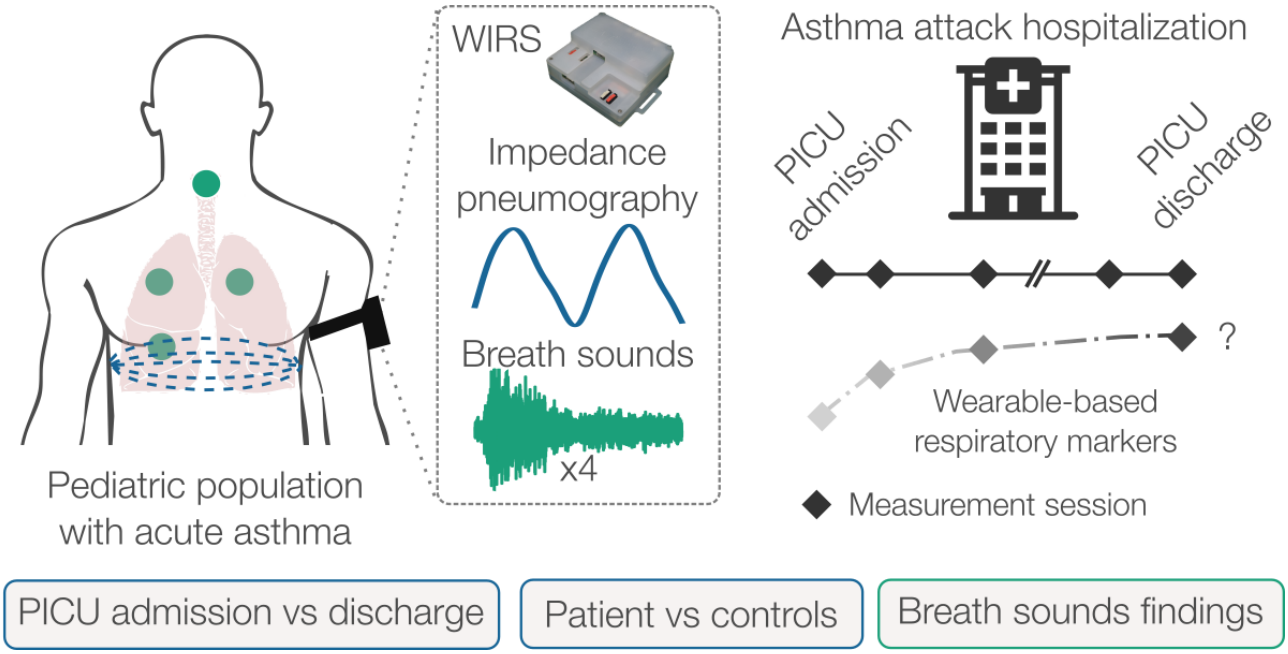
Wearable noninvasive sensing systems have emerged as promising technological approaches for the continuous monitoring of the physiological manifestations of asthma [10,11]. These include vital sign monitoring via electrocardiogram, photoplethysmogram [12], LSs [13], chest movement [14], and electrical bioimpedance (BioZ) signals [15,16]. Yet, demonstrations of these technologies in pediatric populations are scarce. Specifically for children with asthma, van der Kamp et al [17] proposed a suite of sensors for at-home monitoring that leveraged an electrocardiogram for heart rate and respiratory rate (RR) estimation. An at-home sensing suite was also proposed by Hosseini et al [18], including a portable spirometer and heart rate derived from a smartwatch. Despite the promising results, these technologies only capture a limited number of markers that may not fully capture the complex physiological manifestations of asthma.

Recent studies have proposed the use of impedance pneumography (IP) as a promising noninvasive sensing modality for respiratory flow and volume signal measurement in children with asthma [19-21]. Prior work has demonstrated the accurate estimation of respiratory markers such as RR, breathing-phase timings, tidal volume and flow, among others, in both laboratory and clinical settings [15,19,20,22-25]. Notably, we recently demonstrated the feasibility of measuring IP simultaneously with multichannel respiratory sounds in a clinical setting and studied their joint application for continuous breathing-phase contextualization in computerized respiratory sound analysis [26,27]. Given the critical role of auscultation and lung function assessment in asthma management, the joint application of such technologies for the continuous monitoring of severe asthma signs in children warrants further investigation.

This study aimed (1) to evaluate the feasibility of using respiratory markers derived from wearable-based transthoracic BioZ and LSs to quantify the physiological manifestations of life-threatening asthma from peak to resolution of signs and (2) to compare their trajectory against healthy controls, as illustrated in Figure 1. To this end, we evaluated multimodal measurements from 17 children admitted to the PICU with an acute asthma attack, as well as 9 healthy controls. We took measurements up to twice daily during hospitalization and carried out several statistical analyses to assess trends in the wearable-based multimodal respiratory markers from PICU admission to discharge and compared them against controls. Ultimately, the combination of BioZ-derived respiratory markers and acoustic signatures derived from the LSs in a wearable form factor may provide clinicians with a new adjunctive clinical tool for real-time respiratory monitoring and enable timely titration of care, thereby improving patient outcomes.

XSL•FO
RenderX

Figure 1. Overview of the study. Transthoracic electrical bioimpedance and multichannel breath sounds were simultaneously measured on children admitted to the PICU with an asthma attack using our WIRS. Measurements were taken after PICU admission and before discharge. The feasibility of using these signals for monitoring asthma symptoms was evaluated by comparing admission-to-discharge feature trends, differences between patients and controls, and breath sound findings. PICU: pediatric intensive care unit; WIRS: wearable impedance pneumography and respiratory sound system.



Methods

Human Subjects Study

Eligible participants for the patient group were participants admitted to the PICU at the Children's Healthcare of Atlanta (CHOA), Atlanta, Georgia, with a severe asthma attack. Exclusion criteria included substantial comorbidity (eg, cystic fibrosis, bronchiectasis, and pulmonary aspiration). Children were enrolled irrespective of a prior diagnosis of asthma because there are some children whose first presentation of symptoms requires hospitalization, sometimes in the PICU [28]. Eligible participants for the control group were participants with no prior history of respiratory illness as the cause of intensive care admission and without wheezing or otherwise healthy, nonhospitalized individuals. All participants were in the age range of 1-17 years. Measurements took place between April 2023 and April 2024.

Ethical Considerations

The human subject study was carried out with the approval from the Emory University School Institutional Review Board (STUDY0004316) and the Georgia Institute of Technology (IRB #H22329). Informed consent was obtained from the participants' parents or legal guardians prior to any study procedures by trained pediatric research coordinators. Additionally, per institutional review board requirements, children aged 6-11 years provided oral assent and those aged 12-17 years provided written assent, when feasible, given their medical condition. All metadata and physiological signals collected at the CHOA site were properly deidentified before sharing for further processing. No compensation was provided to the participants of this study.

Experimental Protocol

Following enrollment, noninvasive physiological measurements were obtained in the morning (AM) or afternoon (PM), whenever possible, while the participants remained hospitalized in the PICU. The first and last measurements in the PICU are hereafter referred to as T1 and T2, respectively. For healthy participants in the control group, a single measurement took place in a hospital staff room within the PICU, with the participants in the sitting position. For participants in the control group who were hospitalized without signs and symptoms of acute asthma, 2 measurements were taken, and only the first one was used for further analysis. Recordings lasted up to 15 minutes, determined by the cooperativeness of the participants, following which all sensors were detached from the participant, and the recordings were saved for further analysis. No instructions were provided to any of the participants to modulate or control their breathing patterns for the recordings. Clinical data from the electronic medical record were also collected in each recording.

Physiological Measurement

Our custom wearable impedance pneumography and respiratory sound system (WIRS) was used in this study to simultaneously measure transthoracic BioZ and LSs [26]. Transthoracic BioZ was measured at 4 excitation frequencies (5, 50, 100, and 150 kHz). Each frequency was sampled at 16 Hz and calibrated to form complex impedance vectors [29]. A tetrapolar electrode configuration was used, with a pair of current-injecting electrodes placed along the mid-axillary line approximately at the fifth intercostal space and the voltage-measuring electrodes in the same vertical plane and separated by 5 cm approximately. Standard 3M Ag/AgCl gel electrodes were used. This BioZ sensing module was previously validated in prior work [22,23,26].

XSL•FO
RenderX

LSs were measured at 4 locations simultaneously using WIRS. Three locations were used to measure LSs: the anterior right and left upper chest and the posterior right lower chest. The fourth channel was moved from the posterior left lower chest to the suprasternal notch or neck for some participants to measure tracheal sounds. As in prior work, the accelerometer contact microphones (BU-23173-000, Knowles Electronics LLC) were enclosed in our previously reported acoustic enclosure and secured with surgical tape [26]. For the youngest participants, due to their small thorax sizes and for tracheal sound measurements, a more compact acoustic enclosure was used. These consisted of the same overmolded microphones used in Sanchez-Perez et al [26], but only a 3D-printed hollow backing piece was used to prevent loading of the sensor when external forces are exerted on it. Lying over the sensors for posterior chest auscultation while in the supine position is a notable example of such a case. The effectiveness of this acoustic solution was validated in prior work [30].

To validate the recordings against a commercial stethoscope, simultaneous recordings were taken with the Eko CORE digital stethoscope (Eko Devices). The recordings were manually aligned with WIRS following the simultaneous tapping of the Eko CORE diaphragm and the proximal WIRS channel. All measurements were performed by clinical staff at CHOA.

Signal Processing

Impedance Pneumography

The IP signal processing pipeline used in this work is shown in Figure 2. IP is a noninvasive signal that measures the respiratory-mediated changes in thoracic BioZ and has demonstrated excellent correlation to spirometric airflow and volume measures [31,32]. The pipeline follows a similar structure to our previously published version [26] but was modified for this dataset, which took place in an acute-care setting with a pediatric population in the awake state. BioZ-based respiratory monitoring in children with asthma has been previously performed during sleep due to their notorious lack of cooperation, which represents a challenge for physiological measurements [19,20].

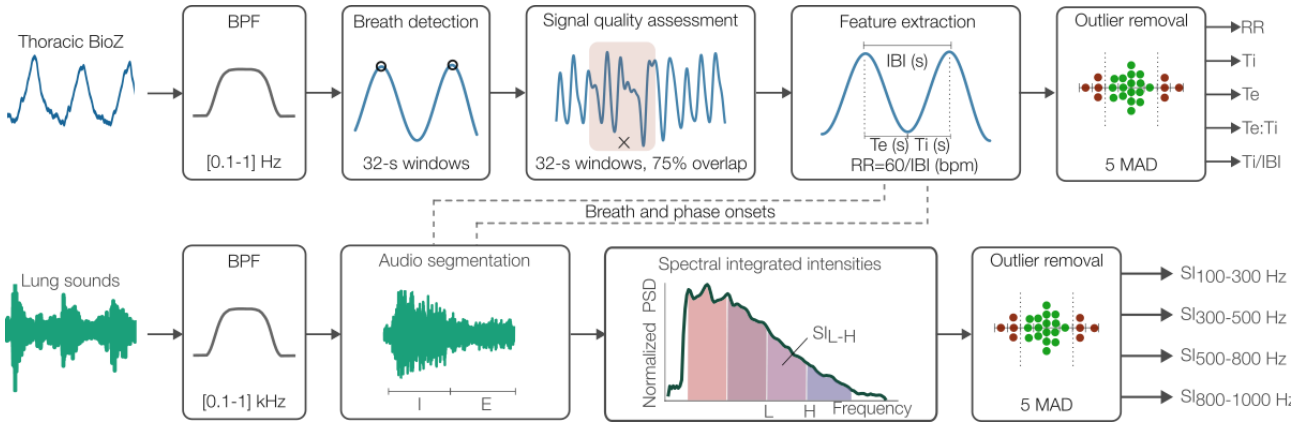
The multifrequency thoracic BioZ signals were first linearly resampled to 100 Hz using a finite impulse response (FIR) antialiasing low-pass filter [26]. An amplitude-based removal stage was then applied to the signals to remove transient events with abnormally high thoracic BioZ values. These events were observed in some participants, likely due to abrupt movements loosening the BioZ wire connection. In this stage, portions of the signal were flagged for removal if they were 1.5 above or 0.5 below the median BioZ. Finally, the removed values were filled via linear interpolation.

As shown in Figure 2, the BioZ signals (150 kHz excitation frequency) were bandpass filtered with a Kaiser window FIR filter in a nominal frequency range of 0.1-1 Hz (6-60 bpm), resulting in IP 150 kHz. The upper end was increased up to 1.4 Hz for some participants with clear dominant frequencies above 1 Hz to ensure accurate processing of the signals. Following filtering, interbreath intervals (IBIs), hereafter referred to as “breaths,” were detected on 32-second windows (2-second overlap) with a minimum distance between peaks set to 0.67 seconds (RR<89.55 bpm), using prior work methodology [33,34]. An RR of 90 bpm has been used as an extreme upper end in previous work [35]. A signal quality index was then assessed using the Charlton et al [36] method. The method was applied using 32-second windows with 75% overlap [15], wherein a breath was considered good if it was labeled as such in any of the windows it belonged to [26]. The normalized SD threshold (ie, coefficient of variation) of the signal quality index method was set to 0.3 to avoid excessive rejection with the original threshold of 0.25, as reported by Blanco-Almazan et al [15], while controlling for the high intraparticipant variability expected in the recordings. Further, a window was considered of good quality if the mean correlation to the window breath template was above 0.7 and at least 60% of the window contained valid breaths.

After signal quality assessment, 3 features were extracted on a breath-by-breath basis: RR, expiration time (Te), and inspiration time (Ti). A feature-based outlier removal stage was then performed to remove breaths if any of these features laid 5 median absolute deviation (MAD) away from the median, over a window of 30 breaths [26,34]. Following this stage, the final set of features was RR, Te, Ti, expiration-to-inspiration time ratio (Te:Ti), and the inspiration time normalized by the IBI duration (Ti/IBI). These respiratory markers have been widely studied for monitoring obstructive pulmonary diseases using BioZ [15,19,20] and spirometry [37].

XSL•FO
RenderX

Figure 2. Multimodal signal processing pipeline used to compute breath-by-breath and phase-by-phase respiratory features. BioZ: electrical bioimpedance; BPF: bandpass filter; E: expiration; I: inspiration; IBI: interbreath interval; MAD: median absolute deviation; RR: respiratory rate; SI: spectral integrated intensity; Te: expiration time; Te:Ti: expiration-to-inspiration time ratio; Ti: inspiration time.



Lung Sounds

The multichannel breath sounds were linearly resampled from 46.875 to 4 kHz to match Eko CORE sampling frequency using a FIR antialiasing low-pass filter. All LS signals were then filtered in the passband 100 to 1000 Hz using a zero-phase minimum-order FIR bandpass filter [27]. The low-frequency cutoff was chosen to suppress heart sound interference [38]. The upper cutoff frequency, on the other hand, suppressed high-frequency noise (eg, speech and hospital room noises) while preserving the parts of the spectrum occupied by normal and adventitious sounds (eg, wheezes) [39,40]. As in prior work, the IBI and phase (ie, inspiration and expiration) onset timings derived from the IP signal were used to index and segment the breath sounds [27]. Figure 2 shows an example of the process of fusing IP-derived timings with LS to segment the phases and extract features on them from prior work.

Following phase and breath segmentation, the normalized spectral integrated intensity (SI) described in Nabi et al [39] was computed across 4 different bands: 100-300, 300-500, 500-800, and 800-1000 Hz. These acoustic features capture the SI at frequency bands that have been found by prior airway obstruction modeling work [41] to be informative within the known LS frequency spectrum [40,42]. First, power spectral density (PSD) estimates were obtained using a 64-ms Hamming window [39] with 75% overlap. These were then normalized to 1 using their respective maximum PSD value to obtain PSD_{norm} [39]. The normalized SIs were then computed as the integrated PSD_{norm} in the specified ranges divided by total cumulative spectral intensity from 100 to 1000 Hz, as described in Nabi et al [39]. This was necessary to obtain comparable intensities regardless of the loudness of the sound, which depends on the airflow rate, auscultation location, and breathing phase [40,42]. The final SI values were defined as the average SIs across all windows within a given breathing-phase segment, resulting in SI100-300, SI300-500, SI500-800, SI800-1000, and for each range described earlier, respectively.

To assess outliers, the cumulative sum in the frequency range 100-1000 Hz of the unnormalized PSD of each breath segment and phase was used, namely, SI, SI_I, and SI_E. Specifically, a feature-based outlier removal stage was applied to remove breaths if any of these 3 features laid 4 MAD away from the median, over a window of 30 breaths. The final set of breaths used in this study consisted of the simultaneous good breaths following the processing procedures for IP and those described in this section for the LS data. In this way, the final breath instances had both IP and audio features. Hence, the final multimodal feature vector consisted of 5 IP features and 8 SI features across both breathing phases, yielding a total of 13 features.

Feature Time Series Generation

The final feature time series was generated after averaging the features over 32-second windows, 75% overlap to mitigate breath-by-breath variability due to physiology or phase-onset variability. A final outlier removal stage was performed by removing window averages lying 5 MAD from the median. Removed values were filled using linear interpolation. All signal processing was done in MATLAB R2020b (MathWorks).

Statistical Analysis

Two separate statistical analyses were carried out to assess differences in respiratory timing features across time and between patient and control groups. In the former, differences in the features within participants with both T1 and T2 time points available were assessed using Wilcoxon signed rank paired statistical tests. In the latter, features from the control group were compared against T1 and T2 of the matched participants using Wilcoxon rank sum tests with Benjamini-Hochberg correction for 2 comparisons (ie, T2-control and T1-control). Comparisons against the control group were conducted to highlight a trend in improvement in respiratory features. Participants were retrospectively matched to controls based on BMI and weight. Before conducting statistical tests, the feature time series of each participant was first aggregated with their mean across any given time point. R5 kHz was excluded from the statistical analysis since the values may not be directly comparable between participants due to differences in electrode locations, posture, and thorax sizes. Differences in demographic data between matched participants and controls were also assessed using Wilcoxon rank sum tests. All statistical analyses were carried out in R (R Foundation for Statistical Computing). A 2-tailed *P* value <.05

XSL•FO
RenderX

was considered significant. Numerical data are shown as median and IQR, unless specified otherwise.

Principal Component Analysis

The multidimensional feature matrix composed of IP- and LS-derived features was projected to 2 dimensions using principal component analysis (PCA) to elucidate any clustering between T1, T2, and control of the matched cohort. The first principal component (PC) was evaluated as a synthetic multimodal feature that maximally captures the variance in the data while retaining explainability. This analysis was carried out in Python (version 3.9; Python Software Foundation) using the scikit-learn library [43].

Results

Dataset Characteristics

Multimodal measurements were obtained from a total of 28 participants. The data from 2 participants were discarded due to measurement errors, resulting in 26 usable datasets (17 participants and 9 controls). Of the 9 controls, 7 were healthy participants, and 2 were patients hospitalized with no signs and symptoms of acute asthma. Table 1 shows detailed demographics for these participants, including the matched participants. No significant differences existed in the demographic data between matched participants and controls. Table 2 shows expanded demographic details for the participants, including the matched cohort, for T1 and T2. The variables include the type of respiratory support, albuterol administration (intermittent or continuous), terbutaline administration, aminophylline administration, time from the last albuterol administration, albuterol dosing interval, time to discharge from PICU, and time to discharge from hospital per recording (T1 and T2). A total of 48 multimodal recordings were obtained from 26 participants, from which 3 recordings were discarded due to poor data quality or measurement errors. The mean recording durations were 748 (SD 117), 661 (SD 86), and 688 (SD 111) seconds for T1, T2, and control, respectively. The mean proportion of raw BioZ data interpolated was 0.235% (SD 0.930%), 1.62% (SD 3.78%), and 0.146% (SD 0.327%) for T1, T2, and control, respectively. The mean final acceptable breaths retained for analysis were 76.1% (SD 23.4%) after applying our processing pipeline. The mean outlier removal proportion at the final feature time series level was 1.17% (SD 0.703%), 0.668% (1.06%), and 0.474% (SD 0.430%) for T1, T2, and control, respectively.

Table 1. Sample demographics.

	All patients (n=17)	Matched patients (n=10)	Controls (n=9)	*P* value[a]
Sex, n (%)				—[b]
Male	11 (64.7)	6 (60)	4 (44.4)	
Female	6 (35.3)	4 (40)	5 (55.6)	
Age (years), median (IQR)	9 (3-17)	6 (3-10)	5 (2-14)	.23
BMI (kg/m^2), median (IQR)	18.6 (15.4-44.2)	17.1 (15.4-19.6)	17.0 (14.1-20.6)	.91
Height (cm), median (IQR)	137.0 (94-184)	118.5 (94-147)	101.0 (85-150)	.31
Weight (kg), median (IQR)	33.5 (16.4-123.6)	23.5 (16.4-33.9)	19.0 (12.1-46.3)	.18

[a]Wilcoxon rank sum test between matched patient and control demographic data.

[b]Not available.

XSL•FO
RenderX

Table 2. Patient demographics.

	All patients		Matched patients	
	T1 (n=17)	T2 (n=13)	T1 (n=10)	T2 (n=9)
Type of support, n (%)				
No significant	8 (47)	12 (92)	7 (70)	6 (67)
HFNC[a]	6 (35)	1 (8)	2 (20)	1 (11)
BiPAP[b]	3 (18)	0 (0)	1 (10)	0 (0)
Intermittent albuterol, n (%)	15 (88)	9 (69)	8 (80)	3 (33)
Continuous albuterol, n (%)	2 (12)	2 (15)	2 (20)	2 (22)
Tertabuline, n (%)	4 (24)	0 (0)	2 (20)	0 (0)
Aminophylline, n (%)	1 (6)	0 (0)	0 (0)	0 (0)
Time from last albuterol (hours), median (IQR)	1.1 (0.4-7.5)	2.4 (0.2-18.6)	1.15 (0.4-7.5)	2.7 (0.2-18.6)
Albuterol dosing interval (hours), median (IQR)	2.5 (1.6-8.0)	4.3 (2.3-10.8)	3.0 (1.6-8.0)	3.9 (3.0-4.3)
Time to PICU[c] discharge (hours), median (IQR)	22.9 (1.6-113.7)	4.3 (0.4-27.3)	9.5 (1.6-74.8)	5.0 (0.4-27.3)
Time to hospital discharge (hours), median (IQR)	46.1 (12.4-173.9)	27.3 (5.7-73.3)	28.0 (23.2-74.8)	26 (5.7-46.6)

[a]HFNC: high-flow nasal cannula (≥6 liters per minute).

[b]BiPAP: bilevel positive airway pressure.

[c]PICU: pediatric intensive care unit.

Analysis of Multimodal Respiratory Features

Three exemplary multimodal recording segments are shown in Figure 3. The segments show the annotated IP signal along with spectrograms of LSs measured with the Eko CORE and the proximal WIRS channel. A transition from wheezing at T1 to normal LSs at T2 for a 3-year-old patient is shown in Figure 3B and 3C. The more compact acoustic enclosures described in the Physiological Measurement section were used due to the small thorax size. An exemplary full-length recording is illustrated in Multimedia Appendix 1. The recording shows the annotated IP signal providing respiratory context and a sharp transition in the phase-specific sound intensities computed for CH_3 from strongly expiratory breath sounds shortly after the 150-second mark and back to normal-like strongly inspiratory sounds after 550 seconds. Exemplary spectrograms from all 4 channels during the segments with normal-like sounds and continuous strongly expiratory sounds are also illustrated.

The results for the paired time points of the participants are shown in Figure 4 and Table 3. Compared to T1, the IP-derived respiratory timings revealed a significant decrease in RR, while Ti and Te increased significantly. Te:Ti and Ti/IBI increased but did not reach statistical significance. The results for the LS-derived SIs revealed a decreasing trend for SI100-300 and an increasing trend for other subbands across both breathing phases. Significant differences were observed for SI100-300 in inspiration, SI500-800 in inspiration, and SI800-1000 in inspiration and expiration.

The results for the comparison between patients and controls for the matched cohort are shown in Figure 5 and Table 4. Significant differences existed between T1 and control for all features. Further, all features except RR, Ti, and Te resulted in significant differences between T2 and control. For the respiratory timings, a decreasing trend toward the control distribution was observed in RR and Ti/IBI, while an increasing trend was observed for Ti, Te, and Te:Ti. For the LS features, a decreasing trend toward the control distribution was observed for SI100-300, while increasing trends were observed for the rest of the subbands. Further, for all groups (T1, T2, and control) and phases (inspiration and expiration), a decreasing trend from the first subband's intensity (SI100-300) to the last (SI800-1000) was also observed (Table 4). Comparing T1 and control, the greatest change was observed for the inspiratory phase across all subbands, and the change magnitude increased as the frequency increased: SI100-300 (inspiration: –17.3% vs expiration: –13.77%), SI300-500 (inspiration: +55% vs expiration: +38.4%), SI500-800 (inspiration: +95.2% vs expiration: +84.1%), and SI800-1000 (inspiration: +114.3% vs expiration: +83.3%).

The PCA results for the matched cohort are shown in Figure 6. PC1 vs PC2 clustering and PC1's loadings ranked in order of importance are shown in Figures 6A and 6C. The aggregated PC1 values per participant in the matched cohort (Figure 6B) resulted in significant differences between T1-control (P<.001) and T2-control (P=.003).

XSL•FO
RenderX

Figure 3. Exemplary simultaneous IP and lung sounds measurements. The top row shows the IP signal with annotated inspiration and expiration phase onsets. The spectrograms show proximal EKO CORE and WIRS recordings: (A) continuous abnormal expiratory sounds, (B) expiratory wheezes found in the afternoon session of a 3-year-old patient at T1, and (C) normal sounds measured the following morning (T2) on the same patient. The time from the last albuterol administration, albuterol dosing interval, time to hospital discharge, and time to PICU discharge were 1.75, 3.4, 26.6, and 26.6 hours, respectively, for (B) and 0.2,4.3, 5.7, and 5.7 hours, respectively, for (C). E: expiration; I: inspiration; IP: impedance pneumography; WIRS: wearable impedance pneumography and respiratory sound system.

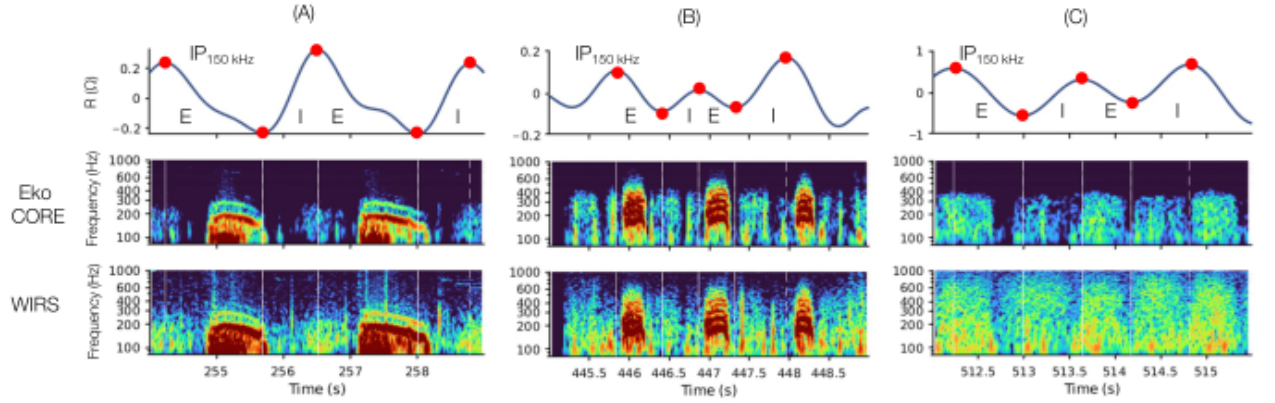
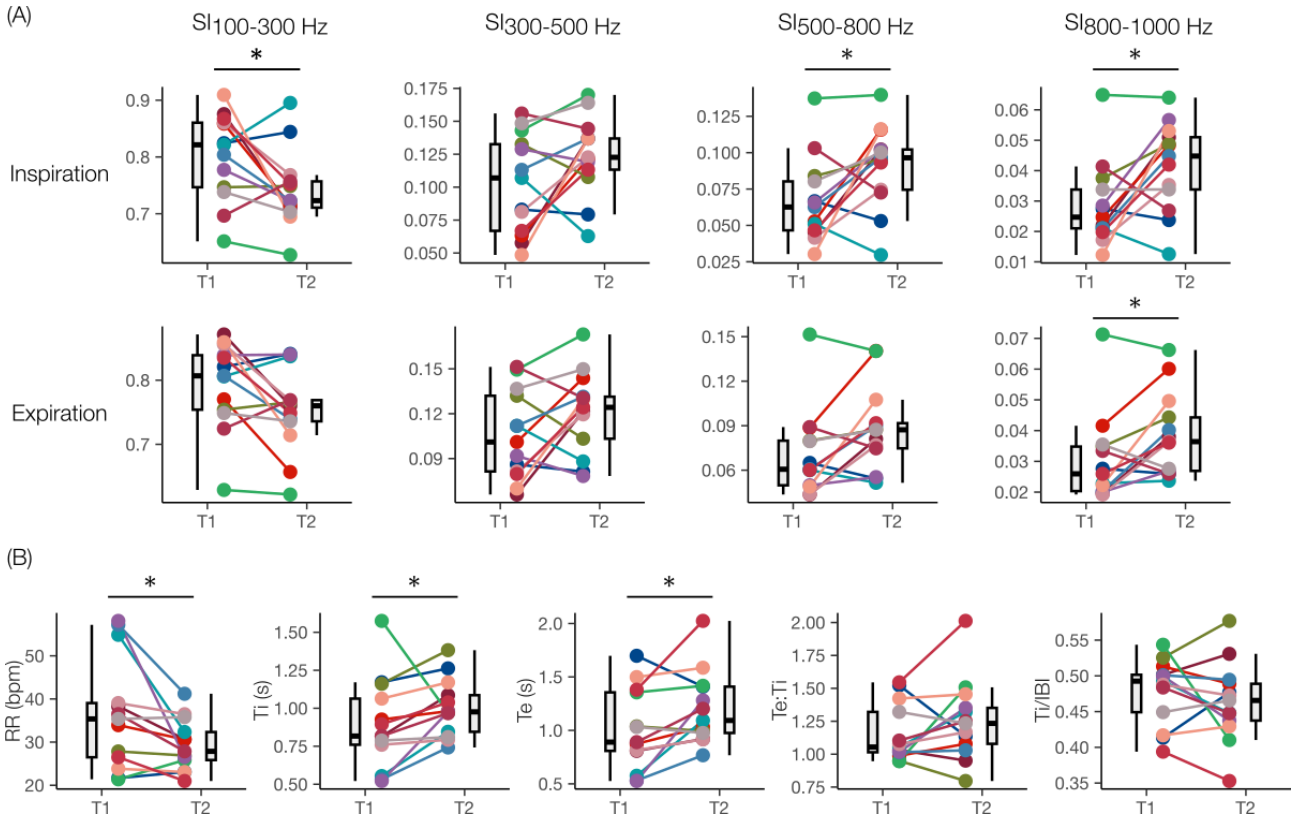


Figure 4. Boxplots of the changes in aggregated respiratory timing features from first (T1) to last (T2) measurement in the pediatric intensive care unit for all patients with paired time points (n=13). (A) SIs for the inspiration and expiration phase and (B) IP-derived respiratory timings. Patients are shown in different colors with lines connecting the mean aggregated data for the time point. IBI: interbreath interval; IP: impedance pneumography; RR: respiratory rate; SI: spectral integrated intensity; Te: expiration time; Te:Ti: expiration-to-inspiration time ratio; Ti: inspiration time. *P<.05.



XSL•FO
RenderX

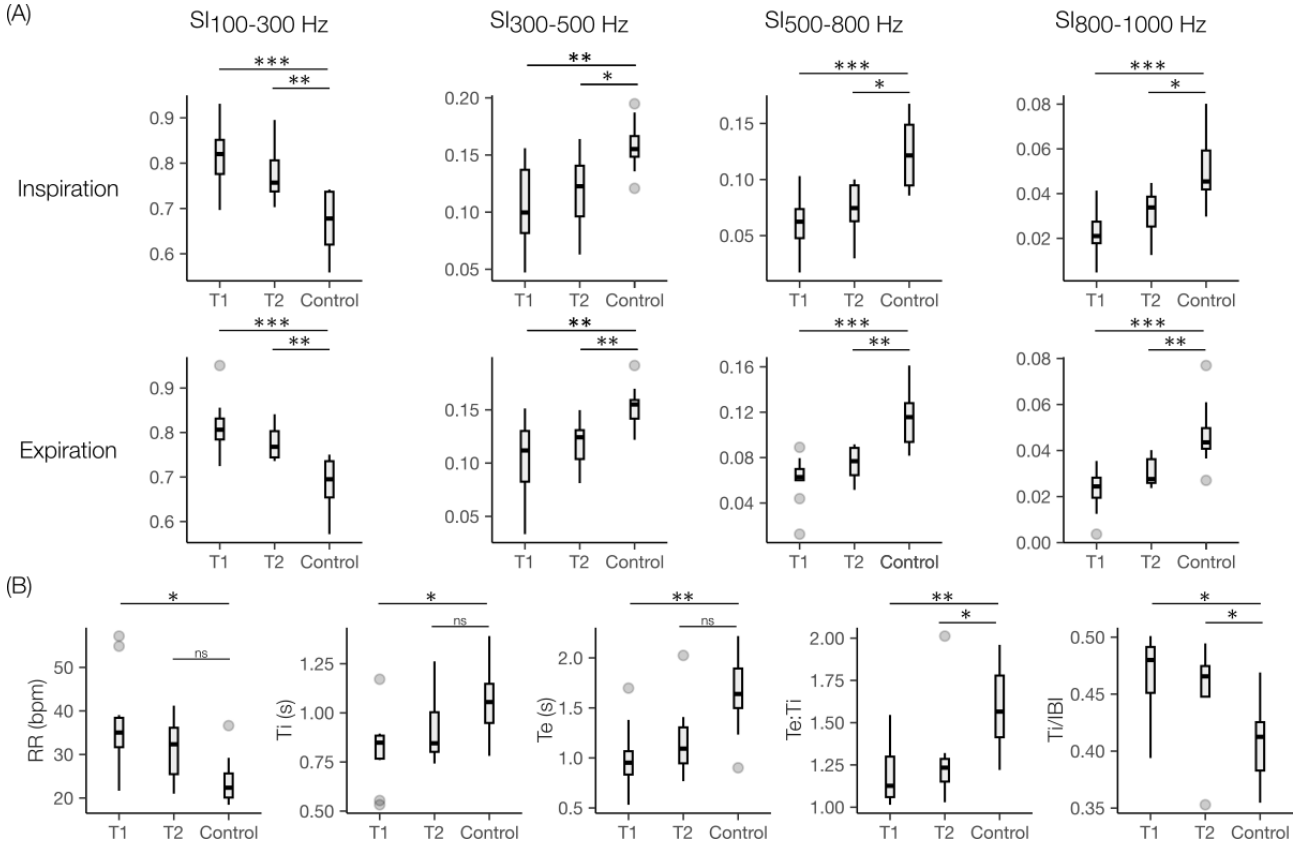
Table 3. Statistical summary of multimodal respiratory features for the patient group.

	T1[a] (n=13), median (IQR)	T2[b] (n=13), median (IQR)	*W*[c]	*P* value[d]
SI100-300I[e,f]	0.821 (0.114)	0.723 (0.046)	75	.04
SI100-300E[g]	0.810 (0.085)	0.760 (0.033)	69	.11
SI300-500I	0.107 (0.066)	0.123 (0.024)	21	.09
SI300-500E	0.101 (0.051)	0.124 (0.028)	24	.15
SI500-800I	0.063 (0.034)	0.096 (0.028)	14	.03
SI500-800E	0.061 (0.030)	0.087 (0.017)	22	.11
SI800-1000I	0.025 (0.013)	0.045 (0.017)	15	.03
SI800-1000E	0.026 (0.036)	0.014 (0.017)	16	.04
RR[h]	35.37 (12.56)	27.86 (6.49)	79	.02
Ti[i]	0.817 (0.303)	0.976 (0.241)	13	.02
Te[j]	0.890 (0.548)	1.093 (0.433)	12	.02
TeTi	1.054 (0.308	1.234 (0.273)	26	.19
Ti/IBI[k]	0.493 (0.052)	0.465 (0.051)	57	.45

[a]T1: is the first measurement time point for the patients.
[b]T2 is the last measurement time point for the patients.
[c]*W* is the test statistic.
[d]Wilcoxon signed rank test between paired time points.
[e]SI: spectral integrated intensity.
[f]I: inspiration.
[g]E: expiration.
[h]RR: respiratory rate.
[i]Ti: inspiration time.
[j]Te: expiration time.
[k]IBI: interbreath interval.

XSL•FO
RenderX

Figure 5. Comparison of multimodal features between first (T1, n=10) and last (T2, n=7) measurement in the pediatric intensive care unit and controls (n=9) for the matched cohort. (A) SIs for inspiration and expiration phase and (B) impedance pneumography–derived respiratory timings. IBI: interbreath interval; ns: not significant; RR: respiratory rate; SI: spectral integrated intensity; Te: expiration time; Te:Ti: expiration-to-inspiration time ratio; Ti: inspiration time. *P<.05, **P<.01, ***P<.001 after Benjamini-Hochberg correction.



XSL•FO
RenderX

Table 4. Statistical summary of multimodal respiratory features for the matched cohort.

	T1[a] (n=10), median (IQR)	*W*[b]	*P* value[c]	T2[d] (n=9), median (IQR)	*W*[b]	*P* value[c]	Control (n=7), median (IQR)
SI100-300I[e,f]	0.820 (0.075)	5	.001	0.757 (0.069)	6	.005	0.678 (0.117)
SI100-300E[g]	0.806 (0.047)	4	.001	0.768 (0.059)	5	.003	0.695 (0.081)
SI300-500I	0.100 (0.055)	81	.004	0.123 (0.044)	53	.02	0.155 (0.018)
SI300-500E	0.112 (0.048)	82	.003	0.124 (0.027)	57	.005	0.155 (0.017)
SI500-800I	0.062 (0.026)	87	<.001	0.075 (0.032)	55	.01	0.121 (0.054)
SI500-800E	0.063 (0.010)	89	<.001	0.077 (0.024)	60	.001	0.116 (0.034)
SI800-1000I	0.021 (0.010)	87	<.001	0.034 (0.013)	53	.02	0.045 (0.017)
SI800-1000E	0.024 (0.009)	86	.001	0.028 (0.010)	58	.003	0.044 (0.009)
RR[h]	35.038 (6.74)	13	.02	32.34 (10.65)	15	.09	22.37 (5.53)
Ti[i]	0.848 (0.12)	77	.02	0.85 (0.20)	45	.17	1.06 (0.20)
Te[j]	0.952 (0.23)	79	.01	1.09 (0.36)	49	.07	1.64 (0.40)
TeTi	1.126 (0.24)	80	.01	1.23 (0.13)	51	.04	1.57 (0.36)
Ti/IBI[k]	0.480 (0.04)	12	.01	0.47 (0.03)	12	.04	0.41 (0.04)

[a]T1: is the first measurement time point for the patients.

[b]*W* is the test statistic.

[c]Wilcoxon rank test between matched patient and control demographic data after Benjamini-Hochberg correction.

[b]T2 is the last measurement time point for the patients.

[e]SI: spectral integrated intensity.

[f]I: inspiration.

[g]E: expiration.

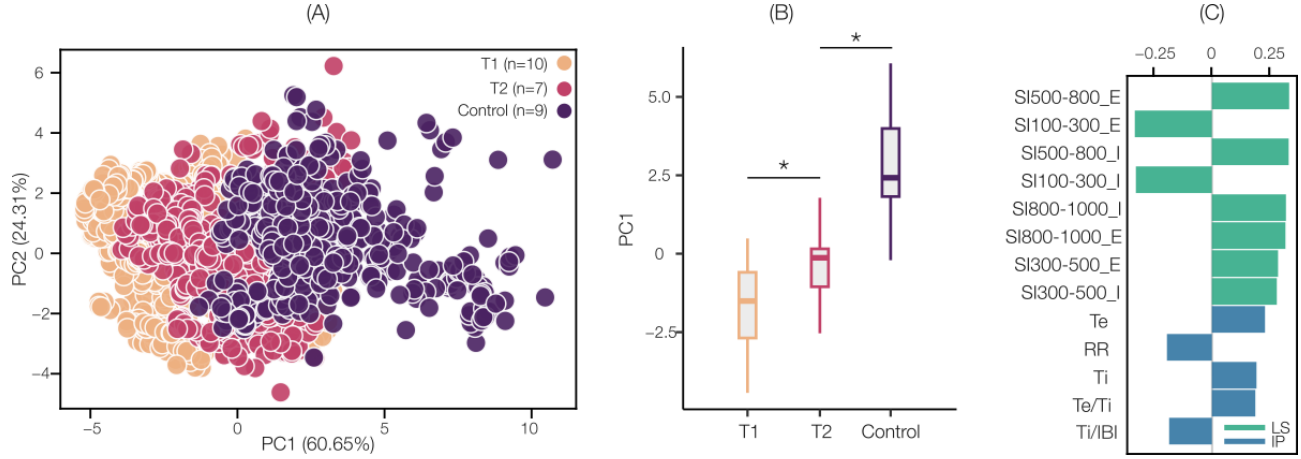
[h]RR: respiratory rate.

[i]Ti: inspiration time.

[j]Te: expiration time.

[k]IBI: interbreath interval.

Figure 6. Principal component analysis results between the first (T1, n=10) and last (T2, n=7) measurement in the pediatric intensive care unit and controls (n=9) for the matched cohort. (A) Color-coded PC2 versus PC1 clustering, (B) boxplots for the comparison of PC1 values aggregated by patient, and (C) PC1 coefficients ranked in order of importance. E: expiration; I: inspiration; IBI: interbreath interval; IP: impedance pneumography; LS: lung sound; PC: principal component; RR: respiratory rate; SI: spectral integrated intensity; Te: expiration time; Te/Ti: expiration-to-inspiration time ratio; Ti: inspiration time. **P*<.05 after Benjamini-Hochberg correction.



XSL•FO
RenderX

Discussion

Principal Findings

In this study, we evaluated the feasibility of quantifying the physiological manifestations of acute asthma in children using a wearable multimodal sensing device. Leveraging the fusion of respiratory timings derived from the IP signal to index breathing cycles and phases, together with simultaneously obtained LSs, we found that phase-contextualized respiratory markers and acoustic signatures exhibited changes that correlated with the transition of patients from an acute asthma attack to a compensated respiratory status state that warranted hospital discharge. In particular, the respiratory timings showed that the breathing pattern trended toward normalcy, while the phase-contextualized acoustic signatures exhibited a change toward normal LSs by T2. We further show the value of fusing these modalities for enhancing the physiological interpretation of LS spectrogram images and for enabling a simple, interpretable proof-of-concept machine learning model for risk stratification of asthma. Hence, these results highlight the feasibility of combining wearable-based respiratory acoustic features derived from LS and respiratory pattern features derived from IP for quantifying and tracking the physiological manifestations of asthma in children.

IP and LS Features Can Quantify and Track the Resolution of Acute Asthma Signs

As the participants transitioned to a compensated respiratory status state—evidenced by the improvement in their signs enabling discharge from the PICU (Table 2)—we found that their breathing pattern changed substantially. In particular, RR decreased, and the duration of both breathing phases increased significantly. The significant reduction in RR indicates a progression toward reversing the airflow obstruction as the children were able to breathe easier. Although Te lengthening has been previously associated with airway narrowing [37], only 3 of the 13 participants showed longer Te at T1, with 2 of them also experiencing lower RR compared to T2. Furthermore, compared to T1 of the matched participants, participants in the control group exhibited significantly lower RR, longer Te, and higher Te:Ti. Hence, our phase and phase-ratio timing results seem to be largely influenced by RR. Previous studies have found strong correlations between RR and tidal breathing timings [44].

Similarly, the transition to a compensated respiratory state correlated with the LS-derived SI markers. Specifically, we found that as the participants transitioned to a more compensated state, SI100-300 decreased, while other higher-frequency subbands increased. This shift in spectral energy from low to high frequencies indicates an overall transition to a flatter and more spread-out spectrum, as we have observed on healthy participants in prior validation studies [26,27]. Notably, we recently studied forced-expiratory (FE) breathing maneuvers on healthy participants, some of which generated expiratory wheezes and found an increase in the mean roll-off frequency f95 (ie, the frequency containing 95% of the spectral energy) from 565 (SD 180) Hz in FE breaths to 929 (SD 130) Hz in normal deep breaths [27]. In addition, we observed the greatest change in the inspiratory features, particularly in the 2 uppermost frequency bands, which suggests a progressive increase in the inspiratory intensity for healthier respiratory states and supports the interpretation of a transition toward normal LS whose strongly inspiratory nature is well-known [42].

Compared to prior work, the finding that changes in spectral power between the groups did not alter the relative distribution of power within the spectral bands (ie, SI100-300>SI300-500>SI800-1000 for all groups) is in agreement with reports using this subband decomposition and normalized SIs, where the SIs were found to decrease as frequency increases regardless of the asthma severity level [39]. The SI trends from T1 found herein, however, disagree with those found by Nabi et al [39] if wheezing is assumed to be present at T1, decreased or resolved by T2, and absent in control. Comparing mild and severe asthma groups, Nabi et al [39] found that SI100-300 was reduced while the rest of the bands were increased with severity level. This discrepancy may be explained by important differences in study design and measurement conditions. First, the work by Nabi et al [39] only included high-quality segments with wheezes, while the production of wheezes, or lack thereof, was not part of the inclusion criteria in this work, nor was it used to define sign severity or disease progression. Second, prior work has used commercially available digital stethoscopes with noise-cancelling capabilities and transducers designed to aggressively filter out high frequencies [45], as illustrated in Figure 3, with the sharp drop in spectral energy after 500 Hz for the sounds measured with the Eko CORE.

In the context of asthma, the generation of wheezes is related to changes in airway thickness [46,47] and the narrowing of the air caliber leading to fluttering of the airway walls and fluids [47]. This leads to sounds exhibiting greater acoustic energy at frequencies higher than 300 Hz compared to normal LS [40,48-50]. A transition from wheezes to normal sounds would then be expected to result in SI100-300 increasing and the other SIs decreasing when using a digital stethoscope since the normal LS spectra will be inherently concentrated below 300 Hz by the sensor itself. In our case, however, the passive nature of the acoustic measurements leads to normal LS becoming flatter, compared to wheezes, rather than more concentrated at lower frequencies, with the ultimate result of decreasing SI100-300 and increasing the other subband's SIs, as the spectral energy becomes spread out between the other bands. Supporting this argument are the opposing trends in f75 and f95 (ie, the frequencies containing 75% and 95% of the total spectral energy, respectively) observed between the Eko CORE and WIRS when comparing normal deep breaths and FE breaths, wherein normal deep breaths resulted in both f75 and f95 increasing for WIRS but decreasing for Eko CORE [27].

Finally, the matched cohort analysis further confirmed the above observations, highlighting that the direction of change in all IP and LS markers across T1 and T2 was toward a compensated respiratory status (control). This suggests the progressive resolution of the acute asthma signs, wherein participants at T2 are physiologically closer to healthy individuals (ie, control group). However, the differences in RR, Te:Ti, and Ti/IBI

XSL•FO
RenderX

between T2 and control indicate that such a transition was not complete before PICU discharge. This may be explained by participants transitioning from the PICU to a regular hospital room before going home for further monitoring until recovery. Hence, their respiratory status at T2 may not have reached normalcy. In all, these results suggest that parameters derived from wearable-based IP and LS may be effectively used to quantify and track respiratory status from peak to resolution of an acute asthma attack.

Wearable-Based Multimodal Respiratory Features May Enable Robust Machine Learning Models for Risk Stratification of Asthma

In the proof-of-concept PCA, we found that PC1 captured close to 61% of the total variance in the data and was significantly higher for control compared to T1 and T2 for the matched cohort. The ranking further reveals that the transition toward a compensated respiratory status was most strongly influenced by increased intensity in the 500-800 Hz subband, decreased intensity in the 100-300 Hz subband, increased intensity in the 800-1000 Hz subband, and finally, increased intensity in the 300-500 Hz subband. Expiratory features showed higher ranks compared to their inspiratory counterparts, but the differences were marginal. This finding is consistent with those reported in Nabi et al [39], wherein inspiratory and expiratory phases were found to be equally informative for discriminating severity levels. Indeed, airway obstruction in asthma may lead to both inspiratory and expiratory LS abnormalities [40]. In the context of children with asthma, inspiratory and expiratory wheezes may be equally informative for characterizing asthma severity levels [51]. Nevertheless, the ability to segment wearable-based LS recordings into phases was enabled by the simultaneously measured IP signal, which motivates further investigation into their utility for monitoring children with asthma. Overall, the feature ranking revealed that the LS-derived phase-specific SI features were more important than the IP-derived features. This highlights the importance of quantifying and tracking respiratory acoustic features in combination with respiratory pattern features for monitoring asthma in children.

The Combination of IP and Multichannel Breath Sounds Holds Promise for Continuous Monitoring of Asthma Signs

Although the IP-derived features ranked lower in importance in our PCA, these results nevertheless highlight the potential of combining both sensing modalities for continuous monitoring of asthma. IP enables the respiratory contextualization of the sounds while also enabling the calculation of well-known tidal breathing parameters, many of which have shown promise for monitoring children with asthma [19-21,25,52]. Furthermore, the combination may not only enhance the discriminatory power of ML models but also the physiological interpretation of the results [27]. Examples of this are our findings of polyphonic expiratory wheezes (Figure 3) and the abrupt transitions in SI in 2 of the sickest participants of the cohort (Multimedia Appendix 1). In the former, the finding indicates that the patient was likely obstructed at T1 and that the reversal of the obstruction resulted in normal LSs at T2 [40,53]. Indeed, the clinical records indicate that this 3-year-old patient transitioned from high-flow nasal cannula at T1 to no ventilatory support at T2 and was discharged home at the same time as PICU discharge. Similarly, the albuterol dosing interval increased from 3.4 to 4.3 hours. Hence, the resolution of expiratory wheezes observed in this patient is consistent with the transition to a normal, compensated respiratory state. In the latter, the expiratory sounds found therein are consistent with snoring or rhonchi that likely resolved after awakening and repositioning. Notably, expiratory snoring and rhonchi may be indicative of airflow obstruction [40,53,54]. In these cases, tidal breathing parameters derived from the IP signal together with phase-specific SI features provide a more holistic and interpretable picture of the physiologic manifestations of asthma in children.

Limitations and Future Work

This study has several limitations. First, the sample size is small and may not capture the well-known disease heterogeneity and severity in asthma [3,6] and limits the assessment of correlations between physiological markers and clinical variables. Second, high intersubject variability in the respiratory markers' response throughout hospitalization was expected in this dataset because the measurement period was different between patients, as well as their relative recovery before PICU discharge. Third, the higher variability in RRs observed in this sample required subject-specific processing parameters to ensure signal quality. This may have introduced a minor bias in the final feature matrix. Fourth, due to the inability to use a regular patient room in the PICU for healthy controls, their measurement took place in a different setting. Although measurements were equally uncontrolled, these participants were naturally freer to move and talk. Furthermore, despite our efforts to maintain the same sensor positions throughout the recordings and across participants, changes in measurement conditions from recording to recording may have resulted in sensing location variability. Likewise, measurements likely took place in slightly different postures due to variations in head-of-bed angles and participants leaning forward for auscultation on the posterior chest. Ultimately, however, this study aimed to evaluate the feasibility of recording multimodal physiological signals with a wearable with the least amount of interference to the regular hospital workflows. Hence, the measured signals captured the natural variability that occurs in realistic continuous monitoring scenarios.

We only analyzed 1 of the 4 locations measured by WIRS. Including additional channels in the analysis will provide regional information, which may improve the results further because LS intensity is known to be a function of auscultation location [40,42]. The LS signal quality evaluation consisted of removing breaths with outlying intensities after rejecting breaths based on the LS signal quality. This may have resulted in the inclusion of low acoustic-quality breaths in the analysis. Yet, there are currently no validated methods for LS signal quality assessment that can be broadly applied to any type of sounds within the vast dictionary of LS described in the literature. Moreover, notable efforts along this direction have developed features to characterize noise based on recordings measured with sensors that integrate active noise-cancellation strategies, which may be different from those obtained with custom

XSL•FO
RenderX

wearable systems such as WIRS, as previously discussed. These observations underscore the need for novel signal quality assessment algorithms that not only incorporate LS data measured with wearable systems but also other wearable-based sensing modalities correlated to the generation of the sounds (ie, IP, IP-derived flow, and chest motion) [26,27,40,42,55]. The successful application of such a method may enable a robust and continuous monitoring of breath sounds at home and in otherwise uncontrolled settings.

Future work will evaluate ML models trained on more complex feature vectors that combine LS and IP data to elucidate their ability to predict sign severity in children hospitalized with asthma. Other auxiliary information obtained from multiple sensors (eg, kinematics and posture) will be evaluated to contextualize the recordings and increase robustness. Chest kinematics via an inertial measurement unit were measured by WIRS and may thus be studied in the future, given the promise shown in recent work to contextualize multimodal recordings [55,56]. Longer tidal-breathing recordings may be studied to enable the incorporation of respiratory pattern variability metrics [15]. Tidal flow variability metrics derived from tidal flow-volume loops may also be studied to assess lung function variability in children with asthma [19,20,24,52]. Furthermore, correlations to clinical outcomes will be studied on larger and more diverse datasets to elucidate the value of wearable technologies such as the one used in this work for supporting clinical decisions.

Conclusions

We evaluated the feasibility of monitoring children from peak to resolution of acute severe asthma signs using wearable-based BioZ and multichannel breath sounds. We found that respiratory timings derived from the IP signal and breathing phase–contextualized SIs derived from the LS signal were significantly altered from peak to resolution of asthma signs, and the changes trended toward the distributions observed in healthy participants. Their combination further enhanced the separation of groups in a low-dimensional subspace. These results highlight the potential of wearable multimodal sensing systems such as WIRS for monitoring the physiologic manifestations of acute asthma in children.

Acknowledgments

The authors would like to thank Sydney Zack for her invaluable contribution to the study coordination at Children's Healthcare of Atlanta.

Funding

This work was supported by the Children's Healthcare of Atlanta 1998 Society. The funder had no involvement in the study design, data collection, analysis, interpretation, or the writing of the manuscript.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: JAS-P, JB, KM, OI, JG
Data curation: JAS-P, JB, NJ, OI, JG
Formal analysis: JAS-P, OI, JG
Funding acquisition: KM, OI, JG
Investigation: JAS-P, JB, NJ, JG
Methodology: JAS-P, JB, OI, JG
Project administration: JAS-P, KM, OI, JG
Resources: JAS-P, JB, JG, OI
Software: JAS-P
Supervision: KM, OI, JG
Validation: JAS-P, OI, JG
Visualization: JAS-P, OI, JG
Writing—original draft: JAS-P, JG
Writing—review and editing: JAS-P, JB, NJ, KM, OI, JG

Conflicts of Interest

None declared.

Multimedia Appendix 1
Exemplary full-length multimodal recording.
[DOCX File , 1169 KB - mhealth_v14i1e72979_app1.docx]

XSL•FO
RenderX

References

1. Zahran HS, Bailey CM, Damon SA, Garbe PL, Breysse PN. Vital signs: asthma in children—United States, 2001-2016. MMWR Morb Mortal Wkly Rep 2018;67(5):149-155 [FREE Full text] [doi: 10.15585/mmwr.mm6705e1] [Medline: 29420459]
2. Most recent national asthma data. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/asthma/most_recent_national_asthma_data.htm [accessed 2025-12-04]
3. Carr TF, Bleecker E. Asthma heterogeneity and severity. World Allergy Organ J 2016;9(1):41 [FREE Full text] [doi: 10.1186/s40413-016-0131-2] [Medline: 27980705]
4. Aggarwal S, Cepalo T, Gill S, Thipse M, Clifton K, Higginson A, et al. Factors associated with future hospitalization among children with asthma: a systematic review. J Asthma 2023;60(3):425-445. [doi: 10.1080/02770903.2022.2070762] [Medline: 35522051]
5. Asthma management guidelines: focused updates 2020. US Department of Health and Human Services. 2024. URL: https://www.nhlbi.nih.gov/health-topics/asthma-management-guidelines-2020-updates [accessed 2024-05-26]
6. Moore WC, Fitzpatrick AM, Li X, Hastie AT, Li H, Meyers DA, et al. Clinical heterogeneity in the severe asthma research program. Ann Am Thorac Soc 2013;10 Suppl:S118-S124 [FREE Full text] [doi: 10.1513/AnnalsATS.201309-307AW] [Medline: 24313761]
7. Bratton SL, Newth CJL, Zuppa AF, Moler FW, Meert KL, Berg RA, et al. Critical care for pediatric asthma: wide care variability and challenges for study. Pediatr Crit Care Med 2012;13(4):407-414 [FREE Full text] [doi: 10.1097/PCC.0b013e318238b428] [Medline: 22067984]
8. Fouzas S, Anthracopoulos MB, Bohadana A. Clinical Usefulness of Breath Sounds. In: Breath Sounds. Cham: Springer International Publishing; 2018:33-52. [doi: 10.1007/978-3-319-71824-8_3]
9. Seed L, Wilson D, Coates AL. Children should not be treated like little adults in the PFT lab. Respir Care 2012;57(1):61-70; discussion 71. [doi: 10.4187/respcare.01430] [Medline: 22222126]
10. Yan W. Toward better management for asthma: from smart inhalers to injections to wearables, researchers are finding new ways to improve asthma treatment. IEEE Pulse 2018;9(1):28-33. [doi: 10.1109/MPUL.2017.2772398] [Medline: 29373855]
11. Taylor L, Ding X, Clifton D, Lu H. Wearable vital signs monitoring for patients with asthma: a review. IEEE Sens J 2022;23(3):1734-1751 [FREE Full text] [doi: 10.1109/JSEN.2022.3224411] [Medline: 37655115]
12. Dieffenderfer J, Goodell H, Mills S, McKnight M, Yao S, Lin F, et al. Low-power wearable systems for continuous monitoring of environment and health for chronic respiratory disease. IEEE J Biomed Health Inform 2016;20(5):1251-1264 [FREE Full text] [doi: 10.1109/JBHI.2016.2573286] [Medline: 27249840]
13. Wu Y, Zhang J, Chen Y, Wang J, Shi W, Zhang Q. Ubi-Asthma: toward ubiquitous asthma detection using the smartwatch. IEEE Internet Things J 2023;10(13):11576-11587. [doi: 10.1109/jiot.2023.3243188]
14. Yuasa Y, Takahashi K, Suzuki K. Wearable flexible device for respiratory phase measurement based on sound and chest movement. 2017 Presented at: IEEE International Conference on Systems, Man, and Cybernetics (SMC); October 5-8, 2017; Banff, Alberta, Canada. [doi: 10.1109/smc.2017.8122978]
15. Blanco-Almazan D, Groenendaal W, Lijnen L, Onder R, Smeets C, Ruttens D, et al. Breathing pattern estimation using wearable bioimpedance for assessing COPD severity. IEEE J Biomed Health Inform 2022;26(12):5983-5991 [FREE Full text] [doi: 10.1109/JBHI.2022.3207416] [Medline: 36121947]
16. Van Steenkiste T, Groenendaal W, Dreesen P, Lee S, Klerkx S, de Francisco R, et al. Portable detection of apnea and hypopnea events using bio-impedance of the chest and deep learning. IEEE J Biomed Health Inform 2020;24(9):2589-2598. [doi: 10.1109/JBHI.2020.2967872] [Medline: 31976919]
17. van der Kamp MR, Klaver EC, Thio BJ, Driessen JMM, de Jongh FHC, Tabak M, et al. WEARCON: wearable home monitoring in children with asthma reveals a strong association with hospital based assessment of asthma control. BMC Med Inform Decis Mak 2020;20(1):192 [FREE Full text] [doi: 10.1186/s12911-020-01210-1] [Medline: 32795352]
18. Hosseini A, Buonocore C, Hashemzadeh S, Hojaiji H, Kalantarian H, Sideris C, et al. Feasibility of a secure wireless sensing smartwatch application for the self-management of pediatric asthma. Sensors (Basel) 2017;17(8):1780 [FREE Full text] [doi: 10.3390/s17081780] [Medline: 28771168]
19. Gracia-Tabuenca J, Seppä VP, Jauhiainen M, Kotaniemi-Syrjänen A, Malmström K, Pelkonen A, et al. Tidal breathing flow volume profiles during sleep in wheezing infants measured by impedance pneumography. J Appl Physiol (1985) 2019;126(5):1409-1418 [FREE Full text] [doi: 10.1152/japplphysiol.01007.2018] [Medline: 30763165]
20. Seppä VP, Pelkonen AS, Kotaniemi-Syrjänen A, Viik J, Mäkelä MJ, Malmberg LP. Tidal flow variability measured by impedance pneumography relates to childhood asthma risk. Eur Respir J 2016 Jun;47(6):1687-1696 [FREE Full text] [doi: 10.1183/13993003.00989-2015] [Medline: 26989106]
21. Seppä VP, Pelkonen AS, Kotaniemi-Syrjänen A, Mäkelä MJ, Viik J, Malmberg LP. Tidal breathing flow measurement in awake young children by using impedance pneumography. J Appl Physiol (1985) 2013;115(11):1725-1731 [FREE Full text] [doi: 10.1152/japplphysiol.00657.2013] [Medline: 24092693]
22. Berkebile JA, Mabrouk SA, Ganti VG, Srivatsa AV, Sanchez-Perez JA, Inan OT. Towards estimation of tidal volume and respiratory timings via wearable-patch-based impedance pneumography in ambulatory settings. IEEE Trans Biomed Eng 2022;69(6):1909-1919 [FREE Full text] [doi: 10.1109/TBME.2021.3130540] [Medline: 34818186]

XSL•FO
RenderX

23. Sanchez-Perez JA, Mabrouk S, Berkebile JA, Esper A, Yang P, Kamaleswaran R, et al. Initial validation of multi-frequency patch-based impedance pneumography in hospital settings. 2022 Presented at: IEEE Sensors; October 30-November 2, 2022; Dallas, TX, United States p. 1-4. [doi: 10.1109/sensors52175.2022.9967287]
24. Młyńczak M, Niewiadomski W, Żyliński M, Cybulski G. Ability to determine dynamic respiratory parameters evaluated during forced vital capacity maneuver using impedance pneumography. 2014 Presented at: 6th European Conference of the International Federation for Medical and Biological Engineering; September 7-11, 2015; Dubrovnik, Croatia p. 877-880. [doi: 10.1007/978-3-319-11128-5_218]
25. Seppä VP, Turkalj M, Hult A, Maloča Vuljanko I, Plavec D, Erceg D, et al. Expiratory variability index (EVI) is associated with the severity of acute bronchial obstruction in small children: a proof-of-concept study. Pediatr Allergy Immunol 2020;31(6):636-642 [FREE Full text] [doi: 10.1111/pai.13257] [Medline: 32307738]
26. Sanchez-Perez JA, Berkebile JA, Nevius BN, Ozmen GC, Nichols CJ, Ganti VG, et al. A wearable multimodal sensing system for tracking changes in pulmonary fluid status, lung sounds, and respiratory markers. Sensors (Basel) 2022;22(3):1130 [FREE Full text] [doi: 10.3390/s22031130] [Medline: 35161876]
27. Sanchez-Perez JA, Gazi AH, Mabrouk SA, Berkebile JA, Ozmen GC, Kamaleswaran R, et al. Enabling continuous breathing-phase contextualization via wearable-based impedance pneumography and lung sounds: a feasibility study. IEEE J Biomed Health Inform 2023;27(12):5734-5744. [doi: 10.1109/JBHI.2023.3319381] [Medline: 37751335]
28. Carroll CL, Schramm CM, Zucker AR. Severe exacerbations in children with mild asthma: characterizing a pediatric phenotype. J Asthma 2008;45(6):513-517. [doi: 10.1080/02770900802017751] [Medline: 18612906]
29. Mabrouk S, Hersek S, Jeong HK, Whittingslow D, Ganti VG, Wolkoff P, et al. Robust longitudinal ankle edema assessment using wearable bioimpedance spectroscopy. IEEE Trans Biomed Eng 2020;67(4):1019-1029 [FREE Full text] [doi: 10.1109/TBME.2019.2927807] [Medline: 31295102]
30. Nevius B. The mechanical design and optimization of a wearable multimodal health sensing system. Georgia Institute of Technology. 2020. URL: https://repository.gatech.edu/entities/publication/df5bcea8-5d5e-494e-864e-2e71d68a90e4 [accessed 2025-12-04]
31. Ernst JM, Litvack DA, Lozano DL, Cacioppo JT, Berntson GG. Impedance pneumography: noise as signal in impedance cardiography. Psychophysiology 1999;36(3):333-338. [doi: 10.1017/s0048577299981003] [Medline: 10352556]
32. Seppä VP, Viik J, Hyttinen J. Assessment of pulmonary flow using impedance pneumography. IEEE Trans Biomed Eng 2010;57(9):2277-2285. [doi: 10.1109/TBME.2010.2051668] [Medline: 20542759]
33. Sanchez-Perez JA, Gazi AH, Mabrouk S, Rahman FN, Seith A, Saks G, et al. Characterizing signal quality of three common respiratory sensing modalities in the context of stress and peripheral nerve stimulation. 2023 Presented at: IEEE 19th International Conference on Body Sensor Networks (BSN); October 9-11, 2023; Boston, MA, United States. [doi: 10.1109/BSN58485.2023.10331181]
34. Gazi AH, Wittbrodt MT, Harrison AB, Sundararaj S, Gurel NZ, Nye JA, et al. Robust estimation of respiratory variability uncovers correlates of limbic brain activity and transcutaneous cervical vagus nerve stimulation in the context of traumatic stress. IEEE Trans Biomed Eng 2022;69(2):849-859. [doi: 10.1109/tbme.2021.3108135]
35. Lee J, Chon KH. An autoregressive model-based particle filtering algorithms for extraction of respiratory rates as high as 90 breaths per minute from pulse oximeter. IEEE Trans Biomed Eng 2010;57(9):2158-2167. [doi: 10.1109/TBME.2010.2051330] [Medline: 20542761]
36. Charlton PH, Bonnici T, Tarassenko L, Clifton DA, Beale R, Watkinson PJ, et al. An impedance pneumography signal quality index: design, assessment and application to respiratory rate monitoring. Biomed Signal Process Control 2021;65:102339 [FREE Full text] [doi: 10.1016/j.bspc.2020.102339] [Medline: 34168684]
37. Hevroni A, Goldman A, Blank-Brachfeld M, Abu Ahmad W, Ben-Dov L, Springer C. Use of tidal breathing curves for evaluating expiratory airway obstruction in infants. J Asthma 2018;55(12):1331-1337. [doi: 10.1080/02770903.2017.1414234] [Medline: 29333884]
38. Kahya YP. "Breath sound recording". In: Priftis KN, Hadjileontiadis LJ, Everard ML, editors. Breath Sounds. Cham: Springer International Publishing; 2018:137. [doi: 10.1007/978-3-319-71824-8_8]
39. Nabi FG, Sundaraj K, Lam CK, Palaniappan R. Characterization and classification of asthmatic wheeze sounds according to severity level using spectral integrated features. Comput Biol Med 2019;104:52-61. [doi: 10.1016/j.compbiomed.2018.10.035] [Medline: 30439599]
40. Bohadana A, Izbicki G, Kraman SS. Fundamentals of lung auscultation. N Engl J Med 2014;370(8):744-751. [doi: 10.1056/NEJMra1302901] [Medline: 24552321]
41. Coleman RF, Schechter GL. A basic model to study acoustic evaluation of airway obstruction. Arch Otolaryngol Head Neck Surg 1991;117(10):1144-1149. [doi: 10.1001/archotol.1991.01870220092016] [Medline: 1910701]
42. Oliveira A, Marques A. Respiratory sounds in healthy people: a systematic review. Respir Med 2014;108(4):550-570 [FREE Full text] [doi: 10.1016/j.rmed.2014.01.004] [Medline: 24491278]
43. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
44. Schmalisch G, Wilitzki S, Wauer R. Differences in tidal breathing between infants with chronic lung diseases and healthy controls. BMC Pediatr 2005;5:36 [FREE Full text] [doi: 10.1186/1471-2431-5-36] [Medline: 16150146]

XSL•FO
RenderX

45. McLane I, Emmanouilidou D, West JE, Elhilali M. Design and comparative performance of a robust lung auscultation system for noisy clinical settings. IEEE J Biomed Health Inform 2021;25(7):2583-2594 [FREE Full text] [doi: 10.1109/JBHI.2021.3056916] [Medline: 33534721]
46. Niimi A, Matsumoto H, Amitani R, Nakano Y, Mishima M, Minakuchi M, et al. Airway wall thickness in asthma assessed by computed tomography. Relation to clinical indices. Am J Respir Crit Care Med 2000;162(4 Pt 1):1518-1523. [doi: 10.1164/ajrccm.162.4.9909044] [Medline: 11029371]
47. Meslier N, Charbonneau G, Racineux J. Wheezes. Eur Respir J 1995;8(11):1942-1948 [FREE Full text] [doi: 10.1183/09031936.95.08111942] [Medline: 8620967]
48. Rao A, Huynh E, Royston TJ, Kornblith A, Roy S. Acoustic methods for pulmonary diagnosis. IEEE Rev Biomed Eng 2019;12:221-239 [FREE Full text] [doi: 10.1109/RBME.2018.2874353] [Medline: 30371387]
49. Sarkar M, Madabhavi I, Niranjan N, Dogra M. Auscultation of the respiratory system. Ann Thorac Med 2015;10(3):158-168 [FREE Full text] [doi: 10.4103/1817-1737.160831] [Medline: 26229557]
50. Andrès E, Gass R, Charloux A, Brandt C, Hentzler A. Respiratory sound analysis in the era of evidence-based medicine and the world of medicine 2.0. J Med Life 2018;11(2):89-106 [FREE Full text] [Medline: 30140315]
51. Rietveld S, Dooijes EH, Rijssenbeek-Nouwens LH, Smit F, Prins PJ, Kolk AM, et al. Characteristics of wheeze during histamine-induced airways obstruction in children with asthma. Thorax 1995;50(2):143-148 [FREE Full text] [doi: 10.1136/thx.50.2.143] [Medline: 7701452]
52. Seppä VP. Development and clinical application of impedance pneumography technique. Tampere University of Technology. 2014. URL: https://trepo.tuni.fi//handle/10024/113904 [accessed 2025-12-04]
53. Reichert S, Gass R, Brandt C, Andrès E. Analysis of respiratory sounds: state of the art. Clin Med Circ Respirat Pulm Med 2008;2:45-58 [FREE Full text] [doi: 10.4137/ccrpm.s530] [Medline: 21157521]
54. Alchakaki A, Riehani A, Shikh-Hamdon M, Mina N, Badr MS, Sankari A. Expiratory snoring predicts obstructive pulmonary disease in patients with sleep-disordered breathing. Ann Am Thorac Soc 2016;13(1):86-92. [doi: 10.1513/AnnalsATS.201507-413OC] [Medline: 26630563]
55. Yoo J, Oh S, Shalish W, Maeng W, Cerier E, Jeanne E, et al. Wireless broadband acousto-mechanical sensing system for continuous physiological monitoring. Nat Med 2023;29(12):3137-3148. [doi: 10.1038/s41591-023-02637-5] [Medline: 37973946]
56. Berkebile JA, Sanchez-Perez JA, Ozmen GC, Inan OT. Improving tidal volume estimation via fusion of impedance pneumography and accelerometry. 2023 Presented at: IEEE Sensors; October 29-November 1, 2023; Vienna, Austria p. 1-4. [doi: 10.1109/sensors56945.2023.10324945]

Abbreviations

BioZ: electrical bioimpedance
CHOA: Children's Healthcare of Atlanta
FE: forced-expiratory
FIR: finite impulse response
IBI: interbreath interval
IP: impedance pneumography
LS: lung sound
MAD: median absolute deviation
PC: principal component
PCA: principal component analysis
PICU: pediatric intensive care unit
PSD: power spectral density
RR: respiratory rate
SI: spectral integrated intensity
Te: expiration time
Te: Ti: expiration-to-inspiration time ratio
Ti: inspiration time
WIRS: wearable impedance pneumography and respiratory sound system

XSL•FO
RenderX

Edited by L Buis; submitted 22.Feb.2025; peer-reviewed by A Berlinski, K Sundaraj; comments to author 25.Jun.2025; revised version received 20.Aug.2025; accepted 28.Oct.2025; published 05.Mar.2026.

Please cite as:
Sanchez-Perez JA, Berkebile J, Jordan N, Maher K, Inan O, Grunwell J
Monitoring Respiratory Health in Children With Acute Asthma Using Wearable Electrical Bioimpedance and Breath Sounds: Observational Case-Control Study
JMIR Mhealth Uhealth 2026;14:e72979
URL: https://mhealth.jmir.org/2026/1/e72979
doi:10.2196/72979
PMID:

©Jesus Antonio Sanchez-Perez, John Berkebile, Natalie Jordan, Kevin Maher, Omer Inan, Jocelyn Grunwell. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 05.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluation of the Importance of Stopping Elderly Accidents, Deaths, and Injuries (STEADI)–Based Factors in Wearable Fall Risk Assessment: Secondary Data Analysis

Nozhan Ghoreishi[1], BS; Stella Ansah[1], BS; Jiayu Lu[1], MS; Wei Lu[2], PhD; Sanghee Moon[3], PhD; Ferdinand Delgado[3], PhD; Diliang Chen[1], PhD

[1]Department of Electrical and Computer Engineering, University of New Hampshire, Durham, NH, United States

[2]Department of Computer Science, Keene State College, Keene, NH, United States

[3]Department of Kinesiology, University of New Hampshire, Durham, NH, United States

Corresponding Author:
Diliang Chen, PhD
Department of Electrical and Computer Engineering
University of New Hampshire
33 Academic Way
Durham, NH, 03824
United States
Phone: 1 2163029368
Email: diliang.chen@unh.edu

Abstract

Background: Falls among older adults are a growing and costly public health problem that often leads to mobility decline and loss of independence. Although clinical frameworks such as the Centers for Disease Control and Prevention's (CDC) Stopping Elderly Accidents, Deaths, and Injuries (STEADI) initiative recommend multifactor screening (gait, balance, strength, fear of falling, and fall history), most wearable fall risk assessment systems rely on a small set of risk factors (typically gait), which creates a gap between clinical practice and automated wearable assessment.

Objective: This study aims to evaluate the importance of STEADI-based fall risk factors and provide design guidance for clinically compatible wearable fall risk assessment systems.

Methods: We created a dataset of 24 older adults (10 low fall risk and 14 high fall risk) from a publicly available plantar pressure dataset of 48 participants by retaining only those with consistent fall risk labels based on both the Berg Balance Scale and the Timed Up and Go test. A total of 18 features were extracted to quantify gait, strength, balance, fear of falling, and fall history. Random forest (RF) models were trained with leave-one-subject-out cross-validation to assess fall risk. Importance of STEADI-based factors was assessed by two methods: (1) estimating Shapley Additive Explanations values based on a single RF model trained on all features; and (2) training 5 separate RF models, each on 1 STEADI factor category, and comparing their fall risk classification accuracies.

Results: In this secondary analysis, the RF model trained on all features achieved a subject-level accuracy of 87.53% (95% CI 75%-100%). Shapley Additive Explanations analysis identified the right foot flat phase ratio (fear of falling feature) as the highest-ranked feature, followed by maximum right forefoot ground reaction force (strength feature), whereas traditional gait features did not appear in the top 10. The 5 separate RF models trained on individual STEADI-based factor categories showed a similar trend in mean participant-level accuracy: fear of falling, 87.59% (95% CI 75%-100%); strength, 79.18% (95% CI 62.5%-95.83%); balance, 70.5% (95% CI 50%-87.5%); gait 70.81% (95% CI 54.17%-87.5%); and fall history 62.37% (95% CI 50%-75.1%). However, paired comparisons did not show statistically significant differences in accuracy between the gait model and the models trained on other factors.

Conclusions: These preliminary results show that commonly overlooked nongait factors are potentially as informative as gait, although clear superiority was not demonstrated in this dataset. The novel foot flat phase ratio ranked higher than all other evaluated features, which showed the value of domain knowledge–informed feature engineering. These preliminary findings indicate that nongait STEADI factors merit consideration in the design of wearable fall risk assessment systems.

(JMIR Mhealth Uhealth 2026;14:e93877) doi:10.2196/93877

XSL•FO
RenderX

KEYWORDS

fall risk assessment; fear of falling; older adults; smart insole; STEADI; Stopping Elderly Accidents, Deaths, and Injuries; wearable sensors

Introduction

Falls are the leading cause of fatal and nonfatal injuries among older adults in the United States [1]. In 2022, falls accounted for approximately 48% of all injury-related deaths among adults aged 65 years or older [1,2]. Each year, falls lead to nearly 3 million emergency department visits and 1 million hospitalizations among older adults [2]. Many older adults who experience falls develop a fear of falling, which can reduce their independence, restrict their mobility, and limit their social interaction [3]. The annual cost of treating nonfatal fall-related injuries among older adults has increased from US $50 billion in 2015 to US $80 billion in 2020 [4] and is projected to exceed US $101 billion by 2030 [5].

To avoid the serious consequences of falls, proactive detection and mitigation of fall risks have been recognized as effective measures and recommended by the Centers for Disease Control and Prevention (CDC). Different from fall prediction or fall detection, fall risk assessment identifies an individual's risk of falling and helps guide targeted interventions to reduce future falls. The Stopping Elderly Accidents, Deaths, and Injuries (STEADI) initiative, developed by the CDC, is one of the most widely used tools for fall risk assessment. It begins with 3 key screening questions related to history of falls, fear of falling, and unsteadiness, followed by further assessment of gait, strength, and balance. Although the STEADI enables a comprehensive fall risk assessment, its practical application is limited because it is time consuming [6-11] and is typically conducted in clinical settings that may not fully reflect everyday behavior [12-14].

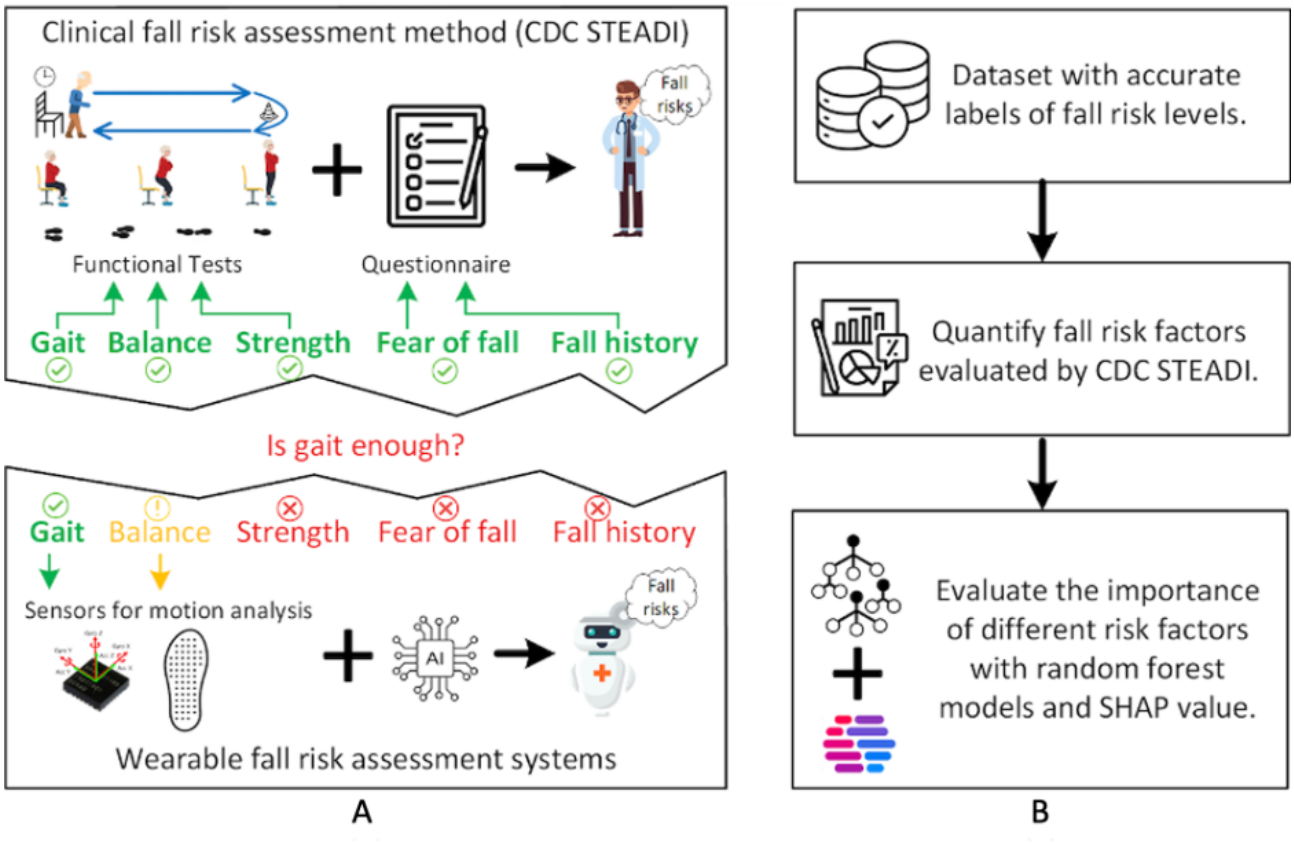
To address the limitations of clinical assessments, new technologies, especially wearable systems, have been proposed to enable automatic fall risk assessment in free-living environments. However, most existing wearable systems mainly rely on gait-related features, which creates a disconnect from clinical practice. As shown in Figure 1A, clinical methods such as STEADI rely on a comprehensive evaluation of gait, balance, strength, fear of falling, and fall history to assess fall risk. However, existing wearable assessment systems rely on an incomplete set of risk factors (typically gait) extracted from sensor data. For example, inertial measurement unit (IMU)–based wearable technologies have been widely explored for fall risk assessment using measured gait parameters. However, their limitations in measuring strength-related factors prevent them from achieving a comprehensive fall risk assessment required by clinical standards [15,16].

This mismatch between wearable systems and clinical tools raises a research question: "Is gait enough?" To answer this question, this study conducted a comprehensive evaluation of the importance of STEADI-based factors for wearable fall risk assessment. As shown in Figure 1B, this study built a dataset with accurate fall risk labels, quantified fall risk factors evaluated by STEADI, and evaluated the importance of different risk factors using random forest (RF models and Shapley Additive Explanations (SHAP) values.

The contributions of this paper are as follows: first, a new dataset with clinically aligned fall risk labels was built to enable reliable evaluation of the importance of STEADI-based fall risk factors. The Timed Up and Go (TUG) test and the Berg Balance Scale (BBS) are 2 frequently used clinical tests for fall risk assessment. Although neither test alone provides a comprehensive evaluation, their combination covers most of the assessments used by STEADI. This study proposed a new labeling standard that combines TUG and BBS scores to identify participants with high and low fall risk from a public dataset. Second, a comprehensive set of risk factors evaluated by STEADI was quantified. In addition to commonly used features to quantify gait, balance, strength, and history of falls, a novel feature, the foot flat phase ratio (the ratio of the foot flat phase duration to the stance phase duration), was extracted to quantify fear of falling. This feature is informed by cautious gait patterns associated with fear of falling and was found to be the highest-ranked feature in this study. Third, the importance of STEADI-based fall risk factors was comprehensively evaluated to guide the development of wearable fall risk assessment systems. An explainable framework was developed with RF models and SHAP values to quantify the contribution of each factor to fall risk assessment.

XSL•FO
RenderX

Figure 1. (A) The research question, "Is gait enough?" is motivated by the mismatch between fall risk factors evaluated in clinical assessment methods (Stopping Elderly Accidents, Deaths, and Injuries [STEADI]) and wearable fall risk assessment systems; (B) overview of the proposed method for evaluating the importance of STEADI-based fall risk factors for wearable fall risk assessment. CDC: Centers for Disease Control and Prevention; SHAP: Shapley Additive Explanations.



Methods

Ethical Considerations

The PlantPre dataset used in this study is deidentified and has been released under the Apache-2.0 license. Ethical approval for the collection of the PlantPre dataset was obtained by Hu et al [17] from the Medical Aircraft Ethics Committee of the First Affiliated Hospital of Jinan University (approval number: KY-2020-099).

Preparing the Dataset With Accurate Fall Risk Labels

To reliably evaluate the importance of different factors for fall risk assessment, a dataset with accurate labels of participants with high and low fall risk is essential. In this study, a new dataset was prepared based on a publicly available dataset (PlantPre) that includes plantar pressure, fall history, TUG test score, BBS score, body weight (BW), and other variables collected from 48 older adults [17]. The plantar pressure data were collected with smart insoles during 2-3 minutes of normal level-ground walking. As shown in Figure 2, each smart insole has 8 sensors for measuring pressure under the foot.

The TUG test and BBS are widely used for fall risk assessment [18,19]. During the TUG test, the patient is observed and timed while rising from an armchair, walking 3 m, turning, walking back, and sitting down again. A TUG score exceeding a threshold (eg, 13.5 seconds) is used to identify individuals at higher risk of falling [20]. The BBS is a widely used clinical test comprising 14 tasks to assess both dynamic and static balance [21]. A BBS score below a specific threshold (eg, 36) usually indicates high fall risk [22]. However, neither the TUG test nor the BBS alone is sufficient to reliably identify older adults with high risks of falls [23,24].

To enhance the accuracy of identifying participants with high or low fall risk, reduce noisy fall risk labels, and provide a more reliable basis for the evaluating the importance of STEADI-based factors, a new labeling method that integrates both the BBS and TUG scores was introduced. Firstly, participants were independently labeled based on the BBS and TUG score. Participants with a BBS score ≤36 were classified as high risk, whereas those with a BBS score >36 were classified as low risk [22]. Similarly, participants with a TUG score ≥13.5 seconds were labeled as high fall risk, whereas those with TUG score <13.5 seconds were labeled as low fall risk [20]. Previous research has shown the effectiveness of these thresholds for reliable fall risk assessment [20,22]. Participants whose labels agreed across both tests were then retained, whereas those with conflicting labels were excluded. This approach addressed the limitations of labeling based on a single test and provided a more reliable fall risk labeling. The flowchart of the proposed participant labeling method is illustrated in Figure 3.

Because 1 participant did not have a recorded history of falls, which was 1 of the fall risk factors evaluated in this study, that participant was excluded from this study. Therefore, the total number of participants shown in Figure 3 was reduced to 47.

XSL•FO
RenderX

As shown in Figure 3, a total of 23 participants were excluded, and the final dataset prepared for this study comprised 24 participants, with 10 labeled as low fall risk and 14 as high fall risk. As shown in Table 1, the TUG score, BBS score, and fall history of each selected participant were listed. The final dataset of 24 participants had no missing values for any variable used in this study.

Figure 2. Location of pressure sensors in the smart insole system.

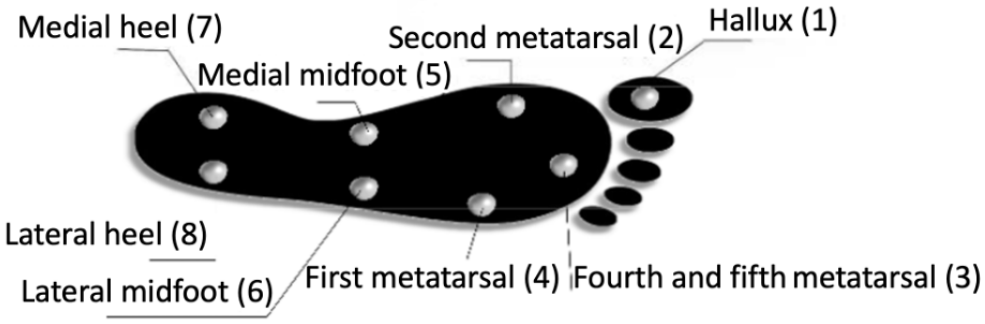
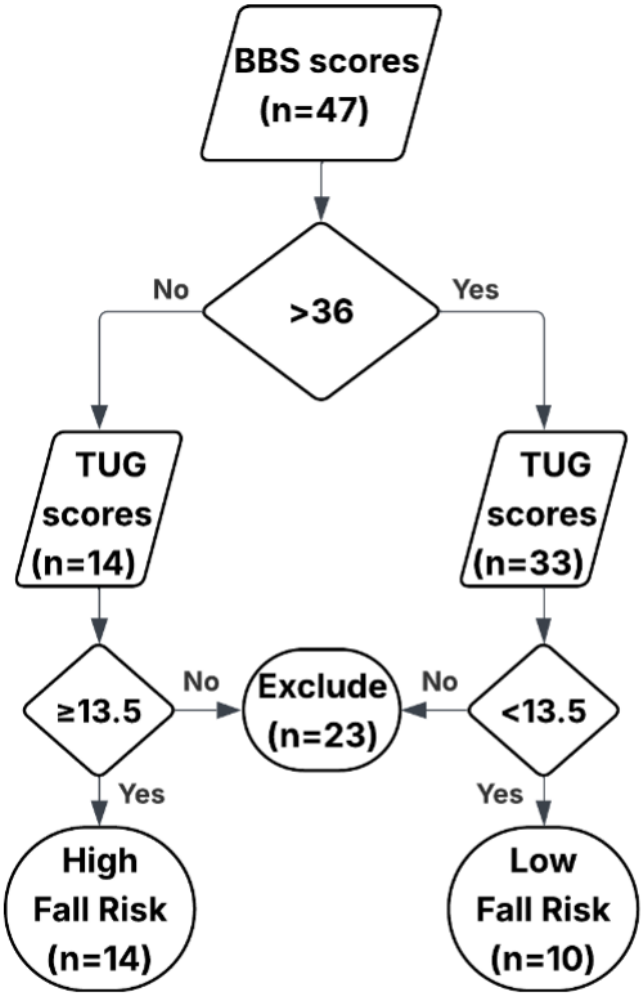


Figure 3. The process used to label participants with high and low fall risk. BBS: Berg Balance Scale; TUG: Timed Up and Go.



XSL•FO
RenderX

Table 1. Participant information in the prepared dataset.

Fall risk level and participant ID	TUG[a] score	BBS[b] score	Fall history
Low fall risk			
1	8.78	54	No
2	9.22	54	No
3	9.41	56	No
4	10.01	54	No
5	10.37	54	No
6	10.5	52	No
7	11.12	54	No
8	12.19	50	No
9	12	37	No
10	12.93	47	No
High fall risk			
11	21.85	29	Yes
12	24.68	34	Yes
13	20.94	28	No
14	22.03	30	No
15	22.04	31	No
16	31.72	35	No
17	22.35	28	No
18	29.54	32	Yes
19	50.37	36	No
20	25.52	25	No
21	33.3	28	Yes
22	45.18	30	No
23	35.56	27	Yes
24	33.4	20	No

[a]TUG: Timed Up and Go.

[b]BBS: Berg Balance Scale.

Quantifying STEADI-Based Fall Risk Factors

Overview

According to the STEADI initiative, gait, balance, strength, fall history, and fear of falling are critical factors for fall risk assessment. In this study, a comprehensive set of features was extracted from the prepared dataset to quantify gait, balance, strength, fall history, and fear of falling. The strength and fear of falling features used in this study are pressure-derived proxies rather than direct clinical measures such as dynamometry-measured strength. Table 2 summarizes all the extracted features and their definitions. The total ground reaction force (GRF) was computed as the sum of the pressure measured by all sensors. Subregional (ie, hindfoot, midfoot, and forefoot) GRFs were obtained by summing the pressure readings from sensors within each specific foot region. To account for differences in BW, each participant's GRF was normalized to BW. The normalized total GRF was then segmented into individual gait cycles using an adaptive threshold method [25]. In total, 3556 gait cycles were detected across 24 participants and were used to extract features to quantify fall risk factors.

XSL•FO
RenderX

Table 2. Extracted features for quantifying Stopping Elderly Accidents, Deaths, and Injuries (STEADI)–based fall risk factors.

Category and gait parameters	Description
Gait features	
Gait cycle time (s)	Time duration between 2 consecutive occurrences of any gait event (eg, heel strike) for the same foot. This parameter was calculated for both feet.
Step time (s)	Time duration from the heel strike of one foot to that of the other foot. This parameter was calculated for both feet.
Stance time ratio (%)	Length of the interval from heel strike to toe off of the foot in a stride divided by the gait cycle time of the same stride. This parameter was calculated for both feet.
Swing time ratio (%)	Length of the interval from toe off in a stride to heel strike of the next stride of the foot divided by the gait cycle time of the same stride. This parameter was calculated for both feet.
Single support time ratio (%)	Length of the interval during which only one foot is in contact with the ground divided by the gait cycle time. This parameter was calculated for both feet.
Double support time ratio (%)	Length of the interval during which both feet are in contact with the ground divided by the gait cycle time. This parameter was calculated for both feet.
Strength-related features	
Median weight acceptance rate (BW/s[a])	Median rate of change of the GRF[b] from heel strike to the weight acceptance peak of the GRF. This parameter was calculated for both feet.
Median push-off rate (BW/s)	Median rate of change of the GRF from the push-off peak to toe off. This parameter was calculated for both feet.
Maximum hindfoot GRF (BW[c])	Maximum force measured by all sensors in the hindfoot region. This parameter was calculated for both feet.
Maximum midfoot GRF (in BW)	Maximum force measured by all sensors in the midfoot region. This parameter was calculated for both feet.
Maximum forefoot GRF (in BW)	Maximum force measured by all sensors in the forefoot region. This parameter was calculated for both feet.
Balance-related features	
Absolute CoP[d] range (cm)	Absolute difference between the maximum and minimum CoP during a gait cycle. This parameter was calculated for the ML[e] direction and the AP[f] direction of both feet.
Average CoP rate of change (cm/s)	Average CoP change during a gait cycle. This parameter was calculated for the ML direction and the AP direction of both feet.
Maximum CoP rate of change (in cm/s)	Maximum CoP changes during a gait cycle. This parameter is calculated for the ML direction and the AP direction of both feet.
Fear-of-falling–related features	
Foot flat phase ratio (%)	Length of the interval when both hindfoot and forefoot are in contact with the ground, divided by the length of the interval from the heel strike to toe off of the foot during a gait cycle. This parameter was calculated for both feet.
Heel strike to foot flat phase start (s)	Length of the interval between heel strike and foot flat start within a gait cycle. This parameter was calculated for both feet.
Foot flat phase end to toe off (s)	Length of the interval between foot flat end and toe off within a gait cycle. This parameter was calculated for both feet.
History-of-falls–related features	
Fall History	A binary variable indicating whether the individual experienced a fall within 1 year before data collection. A value of 1 denotes a history of falling, and a value of 0 denotes no fall history.

[a]BW/s: body weight/second.

[b]GRF: ground reaction force.

[c]BW: body weight.

[d]CoP: center of pressure.

[e]ML: mediolateral.

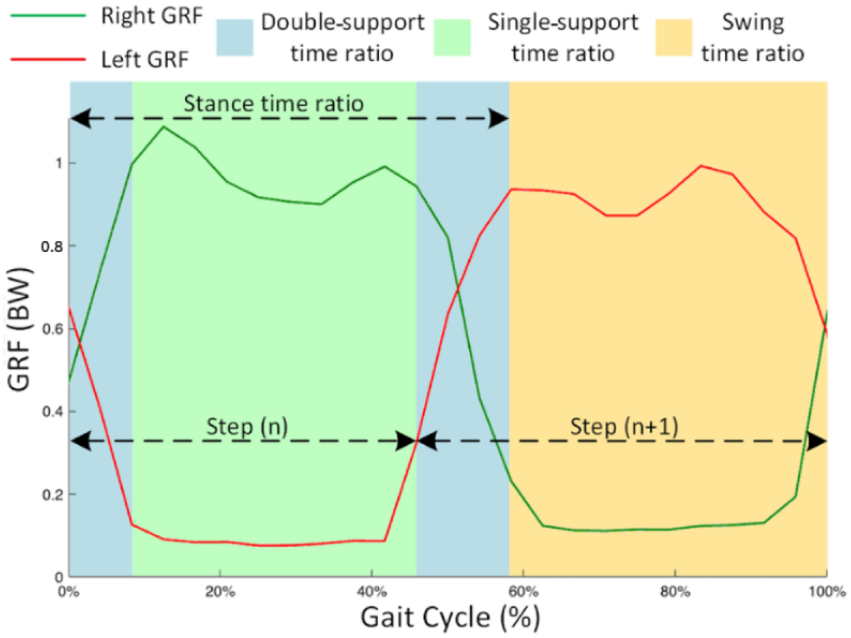
[f]AP: anteroposterior.

XSL•FO
RenderX

Gait Features

Gait is recognized as a major fall risk factor among older adults [26]. To quantify gait-related fall risk factors, a total of 6 features (gait cycle time, step time, stance time ratio, swing time ratio, single support time ratio, and double support time ratio) were extracted from each gait cycle. Swing and stance phases were identified using a dynamic threshold proposed in previous research [25]. Figure 4 and Table 2 show the gait features extracted from each gait cycle.

Figure 4. Gait features extracted for a right foot gait cycle. BW: body weight; GRF: ground reaction force.

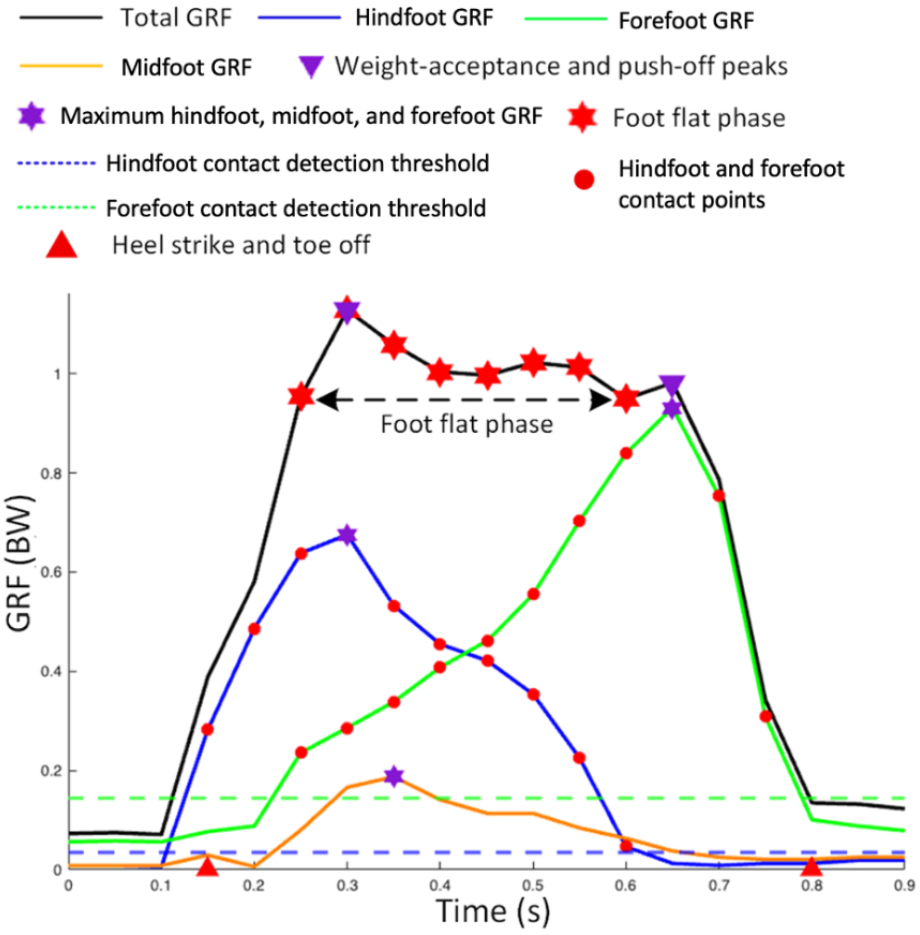


Strength-Related Features

Strength-related features extracted from plantar pressure during walking are strong indicators of fall risk [27]. In this study, as shown in Figure 5 and Table 2, a total of 5 strength-related features were extracted from plantar pressure [28,29]. In addition to the maximum GRF in the hindfoot, midfoot, and forefoot (purple stars in Figure 5), the median rate of change of the GRF was calculated over 2 intervals: from heel strike to the weight acceptance peak (median weight acceptance rate), and from the push-off peak to toe off (median push-off rate). Zero crossings of the GRF rate-of-change signal were used to identify these peaks. The peaks closest to heel strike and toe off were designated as the weight acceptance and push-off peaks (purple triangles in Figure 5), respectively. In gait cycles with only 1 detected peak, the peak was used to calculate both parameters.

XSL•FO
RenderX

Figure 5. Illustration of key parameters for extraction of strength-related and fear-of-falling–related features. BW: body weight; GRF: ground reaction force.



Fear-of-Falling–Related Features

Older adults who experience fear of falling often adopt a cautious gait pattern with significant reductions in various parameters [30,31]. While most studies used conventional gait parameters such as stance time and double-support time to quantify fear of falling, these features have not shown significant contributions to fall risk assessment [31]. To more accurately quantify the cautious gait pattern, the stance phase was divided into 3 subphases: heel strike, foot flat, and toe off. People with cautious gait tend to adopt a flat-foot landing strategy that is characterized by a longer foot flat phase duration but shorter intervals from heel strike to foot flat and from foot flat to toe off [32]. Therefore, 3 features were extracted: the ratio of foot flat phase duration to stance phase duration, the time interval from heel strike to the start of the foot flat phase, and the time interval from the end of the foot flat phase to toe off.

The foot flat phase within each gait cycle is defined as the interval during which both hindfoot and forefoot subregions are in contact with the ground [33]. Inspired by previous studies, dynamic GRF thresholds (blue and green dashed lines in Figure 5) were used to detect ground contact of both hindfoot and forefoot regions [25]. As shown in Figure 5, samples in the foot flat phase are marked with red stars, and the times of heel strike and toe off are marked with red triangles on the left and right, respectively.

Balance-Related Features

In this study, features extracted from the center of pressure (CoP) were used to quantify balance because of their high reliability and accuracy [34]. The CoP is defined as the location at which pressure is concentrated and is calculated by multiplying each sensor's coordinates by its respective pressure reading and normalizing by the total GRF. The CoP in the anteroposterior (AP) and mediolateral (ML) directions was calculated in equation 1.



In this equation, CoP_{AP} and CoP_{ML} indicate the CoP in the AP and ML directions, respectively. x_i and y_i and p_i are the coordinates and the pressure readings of sensor i, respectively. CoP variations and rate of change have been shown to be effective for assessing balance [35,36]. In this study, absolute variations in CoP and the mean and maximum rate of change of CoP were extracted for each gait cycle in the ML and AP directions. CoP variations in the ML direction provide information about lateral sway during each gait cycle, whereas AP CoP changes provide information about weight-transfer patterns throughout each gait cycle.

History of Fall Feature

History of fall is a binary variable indicating whether the individual experienced 1 or more falls in the past 12 months.

XSL•FO
RenderX

A value of 0 indicates no fall history, whereas a value of 1 indicates at least 1 fall during this period.

Evaluating Importance of Different Fall Risk Factors

To evaluate the importance of different fall risk factors, two methods were implemented (1) training a single RF model using all quantified fall risk factors and evaluating the importance of individual risk factor using SHAP values; and (2) training 5 separate RF models, each using 1 of the 5 types of risk factors, with achieved fall risk classification accuracy used to determine the relative importance of each type of risk factors [37]. Each RF model consisted of 100 decision trees and used the Gini impurity criterion for node splitting. The minimum number of samples required to split an internal node and to form a leaf node was set to 2 and 1, respectively. Following standard RF practice, trees were grown to full depth without a maximum depth constraint. No feature selection was performed. Because class imbalance in the dataset was modest, no explicit class weighting or resampling was applied during training.

To avoid participant-specific overfitting and ensure the achieved accuracy generalizes to unseen individuals, we implemented a leave-one-subject-out cross-validation (LOSO-CV), in which the model was trained on 23 participants and tested on the remaining 1 participant. Traditional validation approaches that shuffle and split samples across all participants can place data from the same participant in both training and testing sets. This can cause the model to rely on memorized participant-specific feature patterns rather than fall-related feature patterns, leading to overestimated testing accuracy. By excluding the testing participant's data from the training process, LOSO-CV evaluates the true performance of the model in fall risk assessment. Within each LOSO-CV fold, the training and testing data were standardized using the mean and SD computed from the training set only, which prevented information leakage.

To quantify individual feature contributions, SHAP values were used [38]. This method uses a game-theoretic approach to estimate each feature's impact on the model's output. SHAP values were first computed for each gait cycle of the held-out test participant of each LOSO-CV fold. For each feature, absolute SHAP values were then averaged across all gait cycles of the held-out participant in that fold to obtain 1 fold-level feature importance value. Finally, these fold-level values were summarized across the 24 LOSO-CV folds using the mean and SD.

Results

Overview

This section presents the performance of the RF model for classifying individuals with low and high fall risks, analyzes feature contributions using SHAP values, and evaluates the importance of different types of STEADI-based factors for fall risk assessment.

Performance of RF Model With all STEADI-Based Features for Fall Risk Assessment

Table 3 summarizes the model's classification accuracy for each participant in the dataset. Using all extracted features as inputs, the RF model achieved an overall accuracy of 84.93% in fall risk classification across all gait cycles. Although 3556 gait cycles were available in the dataset, cycles within a participant are not independent. Therefore, the participant was treated as the primary unit for performance analysis. To determine each older adult's final risk of falling, majority voting across all their gait-cycle classifications was used. A participant was considered correctly classified if at least 50% of their gait-cycle classifications matched the ground-truth fall risk label. The RF model achieved a participant-level accuracy of 87.5% in identifying older adults' risk of falling.

To quantify the uncertainty of model performance, we generated 2000 stratified participant-level bootstrap resamples with replacement from the full LOSO-CV predictions to estimate the mean and 95% CI of the model's accuracy, sensitivity, specificity, receiver operating characteristic area under the curve, and Brier score, as shown in Table 4.

XSL•FO
RenderX

Table 3. Gait-cycle number and fall risk classification accuracy of each individual.

Participant ID	Number of gait cycles, n	Gait-cycle-level accuracy (%)	Participant-level accuracy (%)
1	154	69.48	100
2	150	82	100
3	143	93.71	100
4	132	87.88	100
5	130	96.92	100
6	152	90.79	100
7	148	97.3	100
8	133	74.44	100
9	139	91.37	100
10	153	92.81	100
11	141	100	100
12	187	96.79	100
13	139	100	100
14	129	41.86	0
15	149	28.19	0
16	188	55.85	100
17	98	97.96	100
18	207	100	100
19	211	99.05	100
20	154	46.1	0
21	98	100	100
22	89	100	100
23	159	100	100
24	173	100	100
Total	3556	84.93	87.5

Table 4. Performance of the random forest model trained with all features for fall risk assessment.

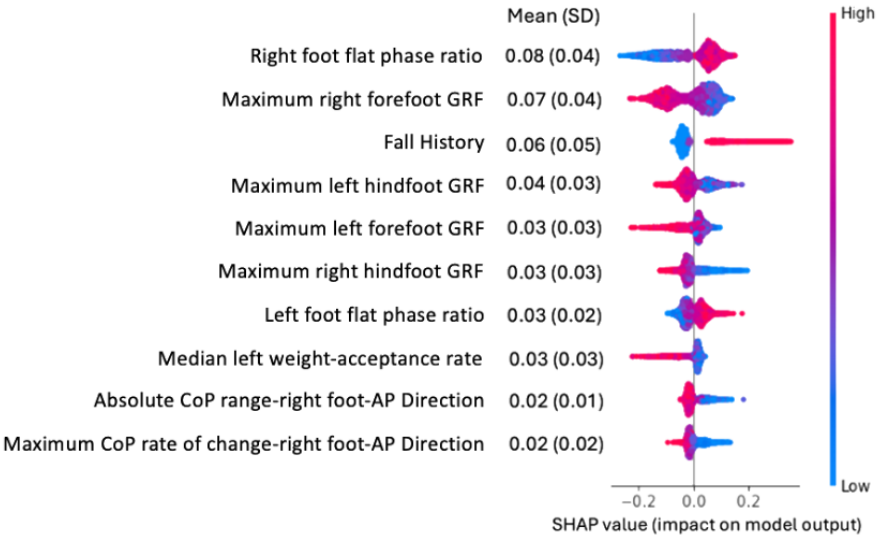
Metrics	Values, mean (95% CI)
Accuracy, %	87.53 (75-100)
Sensitivity, %	79.10 (57.14-100)
Specificity, %	100 (100-100)
ROC-AUC	0.99 (0.96-1.00)
Brier score	0.07 (0.02-0.12)

Importance of Different Types of Fall Risk Factors Based on SHAP Values

Figure 6 summarizes the top 10 features in predicting the risk of falls based on SHAP values. Mean and SD of absolute SHAP values are presented in Figure 6. Among all STEADI-based features, the “right foot flat phase ratio” has the highest rank. According to Figure 6, a higher foot flat phase ratio leads to the prediction of high fall risk, which aligns with clinical findings that longer foot flat phases characterize cautious gait patterns associated with fear of falling and high fall risk [32]. The “maximum right forefoot GRF” ranked second, with higher values predicting lower fall risk. This is consistent with clinical knowledge, as a higher maximum forefoot GRF indicates greater lower-limb strength, which contributes to better gait stability and balance and consequently a lower risk of falls. Fall history ranked third, with more previous falls predicting higher fall risk. This is consistent with clinical practice, where fall history is recognized as an indicator of risk of falls. However, because the dataset includes only 24 participants, the relative ordering should be interpreted as preliminary.

XSL•FO
RenderX

Figure 6. The top 10 most important features for fall risk assessment based on Shapley Additive Explanations (SHAP) values. AP: anteroposterior; CoP: center of pressure; GRF: ground reaction force.



Importance of Different Types of Fall Risk Factors Based on Classification Accuracy

To further illustrate the importance of different types of STEADI-based factors for fall risk assessment, 5 separate RF models were trained, each using features from only 1 type of factor listed in Table 2. The classification accuracy achieved by each model was used to determine the importance of the corresponding type of fall risk factor.

Table 5 shows the classification accuracy achieved by each category of fall risk factor. Overall, fear-of-falling–related features, strength-related features, balance-related features, gait features, and the history-of-fall feature achieved accuracies of 75.68%, 72.44%, 68.34%, 66.31%, and 62.5%, respectively. This ranking is mostly consistent with the result in the previous section "Importance of Different Types of Fall Risk Factors Based on SHAP Values." However, because the dataset includes only 24 participants, the relative ordering should be interpreted as preliminary.

To quantify the uncertainty of model performance, we generated 2000 stratified participant-level bootstrap resamples with replacement from the full LOSO-CV predictions to estimate the mean and 95% CI of the model's accuracy, sensitivity, specificity, receiver operating characteristic area under the curve, and Brier score, as shown in Table 6.

To further evaluate the significance of differences between gait features and other fall risk factor features, a paired bootstrap comparison was performed. The mean and 95% CI of the accuracy differences between the model trained on gait features and the models trained on other fall risk factor features were calculated across 2000 stratified participant-level bootstrap resamples. As shown in Table 7, the mean paired accuracy differences for "fear of falling–gait," "strength–gait," and "balance–gait" are all positive. Since the corresponding 95% CIs indicate that these paired differences were not statistically significant, the results indicate that nongait factors are potentially as informative as gait, although clear superiority was not demonstrated in this dataset.

To provide more insight into the importance of individual features within each category of fall risk factors, SHAP values for the features are shown in Figure 7. According to Figure 7A, older adults with a higher risk of falling showed a longer foot flat phase ratio. This aligns well with clinical findings that older adults with a fear of falling adopt a cautious gait with longer period for the foot flat phase [32].

As shown in Figure 7B, maximum GRF of the forefoot and hindfoot were identified as the most important strength features, with higher maximum GRF indicating a lower risk of fall. This is consistent with clinical findings, as maximum forefoot and hindfoot GRF correspond to peak forces during toe off and heel strike, respectively, and higher forces reflect greater lower-limb strength for controlling walking posture and maintaining stability. Similarly, the weight acceptance rate, which correlates with the hindfoot GRF, was also ranked highly in the list. The lower ranking of maximum midfoot GRF is expected, as it reflects forces during the relatively static foot flat phase rather than the critical dynamic heel strike and push-off phases. The low ranking of "median push-off rate" was unexpected. The reason for this may be the hysteresis effect of the pressure sensor during the push-off phase, which potentially leads to an underestimation of this feature.

Additionally, among the balance-related features (Figure 7C), the maximum CoP velocity in the AP direction was identified as the most important feature for fall risk assessment. As indicated by the SHAP values, older adults at higher risk of falling exhibited lower maximum COP velocities in the AP direction, which reflects a reduced walking speed. This is consistent with clinical findings that older adults at high risk of fall are associated with slower walking speeds.

Finally, as shown in Figure 7D, step time and double-support time were the most important gait features for fall risk assessment. Based on the SHAP analysis, a higher double-support time and step time indicate a higher risk of falling. This result is consistent with clinical studies, where older adults at a higher risk of falling have prolonged double-support and step time [31].

XSL•FO
RenderX

Table 5. Accuracy of different types of factors for fall risk assessment.

Participant ID	Fear-of-falling–features (%)	Strength-related features (%)	Balance-related features (%)	Gait features (%)	History-of-fall features (%)
1	44.81	83.77	38.31	11.04	100
2	60	58	87.33	74.67	100
3	87.41	41.26	82.52	56.64	100
4	65.91	84.85	49.24	49.24	100
5	71.54	96.92	16.15	53.85	100
6	76.32	91.45	63.16	38.16	100
7	96.62	67.57	86.49	88.51	100
8	50.38	78.95	69.92	54.89	100
9	49.64	47.48	35.97	66.91	100
10	74.51	52.94	62.75	43.14	100
11	98.58	63.83	75.89	92.2	100
12	50.8	18.72	56.15	78.61	100
13	77.7	100	97.84	86.33	0
14	86.05	81.4	34.11	21.71	0
15	74.51	8.72	32.89	74.5	0
16	74.47	63.83	91.49	47.34	0
17	98.98	96.94	98.98	93.88	0
18	87.92	94.2	28.99	89.37	100
19	69.19	96.68	100	37.44	0
20	94.81	42.86	85.06	98.7	0
21	46.94	90.82	75.51	89.8	100
22	87.64	100	94.38	100	0
23	100	100	100	91.19	100
24	92.49	100	83.24	79.19	0
Overall	75.68	72.44	68.34	66.31	62.5

Table 6. Performance of the random forest model trained with individual feature set for fall risk assessment, reported as mean (95% CI).

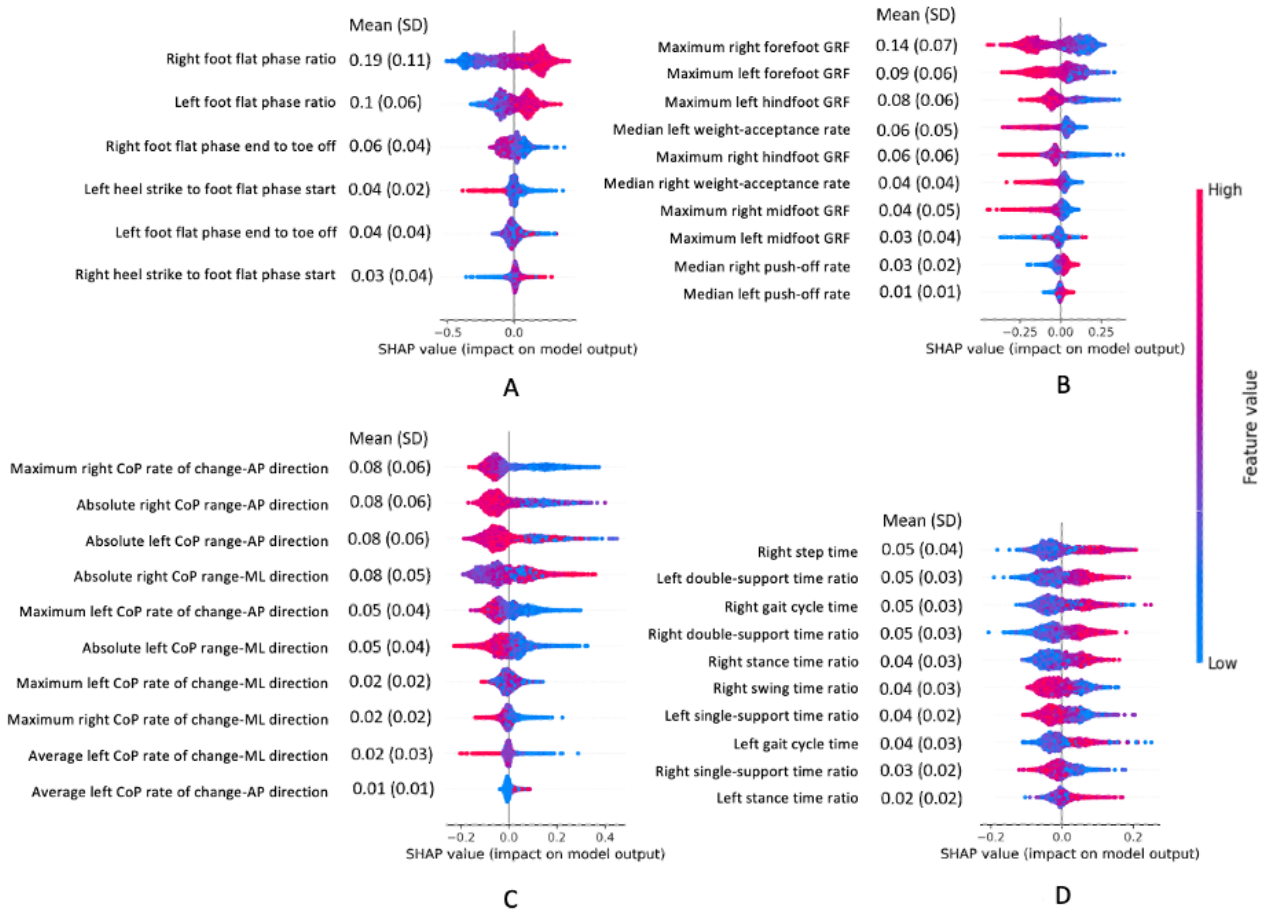
Feature set	Accuracy, % (95% CI)	Sensitivity, %, (95% CI)	Specificity %, (95% CI)	ROC-AUC (95% CI)	Brier score (95% CI)
Fear of fall features	87.59 (75-100)	93.02 (78.57-100)	80 (50-100)	0.97 (0.9-1.0)	0.09 (0.05-0.13)
Strength features	79.18 (62.5-95.83)	78.53 (57.14-100)	80.09 (50-100)	0.88 (0.71-1.00)	0.14 (0.06-0.23)
Balance features	70.5 (50-87.5)	78.19 (57.14-100)	59.73 (30-90)	0.82 (0.63-0.96)	0.16 (0.09-0.25)
Gait features	70.81 (54.17-87.5)	78.15 (57.14-100)	60.53 (30-90)	0.83 (0.64-0.98)	0.17 (0.10-0.25)
History-of-fall features	62.37 (50-75.10)	35.5 (14.29-57.5)	100 (100-100)	0.68 (0.57-0.82)	0.37 (0.21-0.50)

Table 7. Paired accuracy differences between gait and other fall risk factors, reported as mean (95% CI).

Factors	Values, mean (95% CI)
Fear of falling–gait	0.17 (−0.04 to 0.42)
Strength–gait	0.08 (−0.17 to 0.38)
Balance–gait	0.00035 (−0.21 to 0.21)
Fall history–gait	−0.08 (−0.25 to 0.12)

XSL•FO
RenderX

Figure 7. Top features in different categories of Stopping Elderly Accidents, Deaths, and Injuries (STEADI)–based fall risk factors: (A) fear-of-falling–related features, (B) strength-related features, (C) balance-related features, and (D) gait-related features. AP: anteroposterior; CoP: center of pressure; GRF: ground reaction force; ML: mediolateral.



Discussion

Overview

This paper presented a comprehensive study to quantify the importance of STEADI-based factors for wearable fall risk assessment. Based on a newly developed dataset with accurate fall risk labels and a comprehensive feature set that quantifies STEADI-based fall risk factors, the importance of different STEADI-based factors was evaluated using RF models and SHAP analysis. This section discussed the new knowledge generated from this study, provides design guidance for the development of wearable fall risk assessment systems, and specifies the limitations of this study and plans for future work.

STEADI-Based Factors Other Than Gait Are Also Important for Wearable Fall Risk Assessment

Although clinical studies have shown that gait, balance, strength, fear of falling, and history of falling are important for fall risk assessment [39-41], most wearable systems still mainly rely on gait features for fall risk assessment [42-44]. The importance of nongait fall risk factors in wearable fall risk assessment has not been comprehensively evaluated.

Choi et al [43] developed a model to predict the TUG score using gait features extracted from 3D acceleration data collected from pelvis and both feet during normal walking, and then used the predicted TUG scores for fall risk assessment. Maiora et al [42] used IMU sensors to collect motion data during the TUG test and built convolutional neural networks and recurrent neural network models based on gait features extracted from the IMU data for fall risk assessment. Agrawal et al [44] proposed a smart insole–based system for fall risk assessment. They used a commercial smart insole with 5 force-sensitive resistance sensors to collect dynamic plantar pressures from 5 foot locations: big toe, medial forefoot, lateral forefoot, midfoot, and heel. The fall risk labels were generated using the TUG score with a threshold of 13.5, and the RF model achieved an accuracy of 81% in fall risk classification. Although the smart insole has the potential to quantify a comprehensive set of fall risk factors, this study focused on gait and balance features only without exploring the importance of different types of factors for fall risk assessment.

This study provides preliminary evidence to address this gap by comprehensively evaluating the importance of STEADI-based factors for wearable fall risk assessment. The results show that nongait factors are potentially as informative as gait, although clear superiority was not demonstrated in this dataset. The exclusion of these important factors may lead to unreliable fall risk assessment results. This finding provides preliminary design guidance on the necessary factors that should be considered when developing wearable fall risk assessment systems.

XSL•FO
RenderX

Effective Features Can Be Generated From Domain Knowledge

To build the connection between clinical practice and wearable fall risk assessment systems, it is important that the outputs of wearable systems are explainable using clinical knowledge. One solution is to ensure that predicted fall risks are described using clinically meaningful features. This requires developing features based on domain knowledge.

This study showed that effective features can be generated from domain knowledge. For example, based on the domain knowledge that older adults with fear of falling usually adopt a cautious gait with a flat-foot walking pattern, new features specifically focused on the foot flat phase were extracted. The foot flat phase ratio feature was shown to be the most effective feature for fall risk assessment, as shown in Figure 6. This new feature was more effective than many widely used fear-of-falling features, such as stance time and double-support time, which were not ranked in the top 10 features in Figure 6. Compared with stance time and double-support time, the foot flat phase ratio captures how the stance phase is redistributed around foot contact and push-off, rather than simply prolonged duration. This makes it more sensitive to changes in landing strategy and weight-transfer behavior, which may reflect a cautious walking pattern associated with fear of falling. This finding contributes to the knowledge base on effective features to quantify fear of falling and shows the value of domain knowledge in developing new features for fall risk assessment.

Limitations and Future Work

Although this study quantified different fall risk factors and presented a preliminary evaluation of their importance in fall risk assessment, its scope is limited by the existing dataset, which will be addressed by future studies.

First, although the analysis included 3556 gait cycles, these gait cycles were collected from only 24 participants. Therefore, the effective independent sample size is determined by the number of participants rather than the number of gait cycles. The limited number of participants may affect the stability of feature-importance rankings when generalizing to a broader population. Therefore, the main contribution of this preliminary study is not the feature-importance ranking but the finding that STEADI-based factors beyond gait features may also contribute meaningfully to fall risk assessment. Given the limited participant-level sample size and wide CIs, the findings should be interpreted as preliminary. Additionally, although the class imbalance in the dataset was modest (14 high-risk vs 10 low-risk participants), it may still bias the model toward the majority class and influence feature-importance estimates. In future work, we will build a new more balanced dataset with more participants to further explore the relative importance of STEADI-based factors.

Second, the gait, strength, balance, and fear-of-falling–related features were extracted only from the plantar pressure data collected using a smart insole with 8 pressure sensors. One potential shortage is that 8 sensors may not provide enough spatial resolution to quantify different risk factors with optimal accuracy. Additionally, although temporal gait parameters were extracted from plantar pressure data, spatial gait parameters such as stride length and gait speed could not be extracted for analysis. This may lead to an underestimation of the contribution of gait-related features. In the future, we plan to address these limitations by using the smart insole developed in our laboratory. This insole has 96 pressure sensors uniformly distributed across the insole, and an accurate IMU sensor integrated into it. It has been comprehensively validated and successfully applied in various research projects [25,33,45-52]. With this smart insole, we will build a new dataset by collecting comprehensive data from more older adults, enabling a more accurate and comprehensive evaluation of the importance of different fall risk factors.

Third, to reliably evaluate the importance of different factors for fall risk assessment, the dataset used in this study was built by excluding participants whose TUG and BBS labels disagreed. Although this method reduces label noise, the resulting dataset may not fully reflect the heterogeneity of real clinical populations and may reduce ecological validity. Therefore, the reported model performance may overestimate the accuracy achievable in a heterogeneous clinical population. Additionally, the estimated importance of STEADI-based factors may not fully generalize to participants whose BBS and TUG assessments disagree. Validation in a larger and more heterogeneous cohort, including discordant cases, is an important direction for future work.

Participant-level classification was obtained by majority voting across all gait cycles of each participant. A participant was considered correctly classified if at least 50% of their gait-cycle classifications matched the ground-truth fall risk label. However, the stability of this approach depends on the number of gait cycles available for each participant and may be reduced when the number of cycles is relatively small or when the proportion of positive classifications is close to 50%. Therefore, participant-level classifications near the decision boundary should be interpreted with caution.

Conclusion

This study provides a preliminary evaluation of the importance of STEADI-based factors for wearable fall risk assessment. The study quantified a comprehensive set of STEADI-based factors and introduced a novel domain-knowledge–informed feature, the foot flat phase ratio. The key finding is that commonly overlooked nongait factors are potentially as informative as gait, although clear superiority was not demonstrated in this dataset. These preliminary findings indicate that nongait STEADI factors merit consideration in the design of wearable fall risk assessment systems.

XSL•FO
RenderX

Funding

This work is funded by the National Science Foundation (NSF) Faculty Early Career Development (CAREER) Award number: 2443746.

Authors' Contributions

NG contributed to research design, data analysis, and manuscript writing and revision; SA and JL contributed to data analysis; WL contributed to research design and manuscript review; SM and FD contributed to manuscript review and revision; and DC contributed to research design and manuscript review and revision.

Conflicts of Interest

None declared.

References

1. Web-based injury statistics query and reporting system. Centers for Disease Control and Prevention. 2026. URL: https://wisqars.cdc.gov/ [accessed 2026-02-20]
2. About older adult fall prevention. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/falls/about/ [accessed 2026-02-20]
3. Sapmaz M, Mujdeci B. The effect of fear of falling on balance and dual task performance in the elderly. Exp Gerontol 2021;147:111250. [doi: 10.1016/j.exger.2021.111250] [Medline: 33493582]
4. Haddad YK, Miller GF, Kakara R, Florence C, Bergen G, Burns ER, et al. Healthcare spending for non-fatal falls among older adults, USA. Inj Prev 2024;30(4):272-276. [doi: 10.1136/ip-2023-045023] [Medline: 39029927]
5. Houry D, Florence C, Baldwin G, Stevens J, McClure R. The CDC injury center's response to the growing public health problem of falls among older adults. Am J Lifestyle Med 2016;10(1) [FREE Full text] [doi: 10.1177/1559827615600137] [Medline: 26688674]
6. Casey CM, Parker EM, Winkler G, Liu X, Lambert GH, Eckstrom E. Lessons learned from implementing CDC's STEADI falls prevention algorithm in primary care. Gerontologist 2017;57(4):787-796 [FREE Full text] [doi: 10.1093/geront/gnw074] [Medline: 27130270]
7. Tai-Seale M, McGuire TG, Zhang W. Time allocation in primary care office visits. Health Serv Res 2007;42(5):1871-1894 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00689.x] [Medline: 17850524]
8. Shubert TE, Smith ML, Prizer LP, Ory MG. Complexities of fall prevention in clinical settings: a commentary. Gerontologist 2014;54(4):550-558. [doi: 10.1093/geront/gnt079] [Medline: 23887933]
9. Stevens JA, Phelan EA. Development of STEADI: a fall prevention resource for health care providers. Health Promot Pract 2013;14(5):706-714 [FREE Full text] [doi: 10.1177/1524839912463576] [Medline: 23159993]
10. Stevens JA. The STEADI tool kit: a fall prevention resource for health care providers. IHS Prim Care Provid 2013;39(9):162-166 [FREE Full text] [Medline: 26766893]
11. Stevens JA, Smith ML, Parker EM, Jiang L, Floyd FD. Implementing a clinically based fall prevention program. Am J Lifestyle Med 2020;14(1):71-77 [FREE Full text] [doi: 10.1177/1559827617716085] [Medline: 31903086]
12. Zampieri C, Salarian A, Carlson-Kuhta P, Nutt JG, Horak FB. Assessing mobility at home in people with early Parkinson's disease using an instrumented timed up and go test. Parkinsonism Relat Disord 2011;17(4):277-280 [FREE Full text] [doi: 10.1016/j.parkreldis.2010.08.001] [Medline: 20801706]
13. Schoene D, Wu SM, Mikolaizak AS, Menant JC, Smith ST, Delbaere K, et al. Discriminative ability and predictive validity of the timed up and go test in identifying older people who fall: systematic review and meta-analysis. J Am Geriatr Soc 2013;61(2):202-208. [doi: 10.1111/jgs.12106] [Medline: 23350947]
14. Robles-García V, Corral-Bergantiños Y, Espinosa N, Jácome MA, García-Sancho C, Cudeiro J, et al. Spatiotemporal gait patterns during overt and covert evaluation in patients with Parkinson's disease and healthy subjects: is there a Hawthorne effect? J Appl Biomech 2015;31(3):189-194. [doi: 10.1123/jab.2013-0319] [Medline: 25536440]
15. Ng KD, Mehdizadeh S, Iaboni A, Mansfield A, Flint A, Taati B. Measuring gait variables using computer vision to assess mobility and fall risk in older adults with dementia. IEEE J Transl Eng Health Med 2020;8:2100609 [FREE Full text] [doi: 10.1109/JTEHM.2020.2998326] [Medline: 32537265]
16. Saboor A, Kask T, Kuusik A, Alam MM, Le Moullec Y, Niazi IK, et al. Latest research trends in gait analysis using wearable sensors and machine learning: a systematic review. IEEE Access 2020;8:167830-167864. [doi: 10.1109/access.2020.3022818]
17. Hu G, Jin J, Song Z, Wu S, Shu L, Xie J. A dataset for falling risk assessment of the elderly using wearable plantar pressure. 2022 Presented at: IEEE International Conference on Bioinformatics and Biomedicine (BIBM); December 6-8, 2022; Las Vegas p. 2900-2904. [doi: 10.1109/bibm55620.2022.9995052]
18. Lamezon AC, Cavon BL, Valderramas S. Balance, falls, and risk of falls in COPD: systematic review of assessment instruments, measurement properties, and clinical utility. Rev Bras Geriatr Gerontol 2024;27:e230189. [doi: 10.1590/1981-22562024027.230189.en]

XSL•FO
RenderX

19. Beauchet O, Fantino B, Allali G, Muir SW, Montero-Odasso M, Annweiler C. Timed up and go test and risk of falls in older adults: a systematic review. J Nutr Health Aging 2011;15(10):933-938 [FREE Full text] [doi: 10.1007/s12603-011-0062-0] [Medline: 22159785]
20. Barry E, Galvin R, Keogh C, Horgan F, Fahey T. Is the timed up and go test a useful predictor of risk of falls in community dwelling older adults: a systematic review and meta-analysis. BMC Geriatr 2014;14:14 [FREE Full text] [doi: 10.1186/1471-2318-14-14] [Medline: 24484314]
21. Miranda N, Tiu TK. Berg Balance Testing. Treasure Island: StatPearls Publishing; 2026.
22. Shumway-Cook A, Baldwin M, Polissar NL, Gruber W. Predicting the probability for falls in community-dwelling older adults. Phys Ther 1997;77(8):812-819. [doi: 10.1093/ptj/77.8.812] [Medline: 9256869]
23. Lima C, Ricci N, Nogueira E, Perracini M. The Berg Balance Scale as a clinical screening tool to predict fall risk in older adults: a systematic review. Physiotherapy 2018;104(4):383-394. [doi: 10.1016/j.physio.2018.02.002] [Medline: 29945726]
24. Muir SW, Berg K, Chesworth B, Speechley M. Use of the Berg Balance Scale for predicting multiple falls in community-dwelling elderly people: a prospective study. Phys Ther 2008;88(4):449-459. [doi: 10.2522/ptj.20070251] [Medline: 18218822]
25. Chen D, Cai Y, Qian X, Ansari R, Xu W, Chu K, et al. Bring gait lab to everyday life: gait analysis in terms of activities of daily living. IEEE Internet Things J 2020;7(2):1298-1312. [doi: 10.1109/jiot.2019.2954387]
26. Colón-Emeric CS, McDermott CL, Lee DS, Berry SD. Risk assessment and prevention of falls in older community-dwelling adults: a review. JAMA 2024;331(16):1397-1406. [doi: 10.1001/jama.2024.1416] [Medline: 38536167]
27. Yan Y, Ou J, Shi H, Sun C, Shen L, Song Z, et al. Plantar pressure and falling risk in older individuals: a cross-sectional study. J Foot Ankle Res 2023;16(1):14 [FREE Full text] [doi: 10.1186/s13047-023-00612-4] [Medline: 36941642]
28. Hirano Y, Matsui Y, Nemoto T, Ota S, Yamada Y, Arai H. Change in ground reaction force parameters according to the frailty level of older women in the timed up and go test. Nagoya J Med Sci 2022;84(2):418-432 [FREE Full text] [doi: 10.18999/nagjms.84.2.418] [Medline: 35967955]
29. Tsuji T, Tsunoda K, Mitsuishi Y, Okura T. Ground reaction force in sit-to-stand movement reflects lower limb muscle strength and power in community-dwelling older adults. Int J Gerontol 2015;9(2):111-118. [doi: 10.1016/j.ijge.2015.05.009]
30. Bueno GAS, Gervásio FM, Ribeiro DM, Martins AC, Lemos TV, de Menezes RL. Fear of falling contributing to cautious gait pattern in women exposed to a fictional disturbing factor: a non-randomized clinical trial. Front Neurol 2019;10:283 [FREE Full text] [doi: 10.3389/fneur.2019.00283] [Medline: 30972013]
31. Maki BE. Gait changes in older adults: predictors of falls or indicators of fear. J Am Geriatr Soc 1997;45(3):313-320. [doi: 10.1111/j.1532-5415.1997.tb00946.x] [Medline: 9063277]
32. Yang F, Pai CY. Alteration in community-dwelling older adults' level walking following perturbation training. J Biomech 2013;46(14):2463-2468 [FREE Full text] [doi: 10.1016/j.jbiomech.2013.07.025] [Medline: 23978691]
33. Ghoreishi N, LaCourse J, Arthanat S, LaRoche D, Chen D. Auto-calibrated wearable system for load vertical location estimation during manual lifting. IEEE Internet Things J 2025;PP(99):1-1. [doi: 10.1109/jiot.2025.3605260]
34. Mehdizadeh S, Van Ooteghem K, Gulka H, Nabavi H, Faieghi M, Taati B, et al. A systematic review of center of pressure measures to quantify gait changes in older adults. Exp Gerontol 2021;143:111170. [doi: 10.1016/j.exger.2020.111170] [Medline: 33238173]
35. Svoboda Z, Bizovska L, Janura M, Kubonova E, Janurova K, Vuillerme N. Variability of spatial temporal gait parameters and center of pressure displacements during gait in elderly fallers and nonfallers: a 6-month prospective study. PLoS One 2017;12(2):e0171997 [FREE Full text] [doi: 10.1371/journal.pone.0171997] [Medline: 28241008]
36. DeBolt LS, McCubbin JA. The effects of home-based resistance exercise on balance, power, and mobility in adults with multiple sclerosis. Arch Phys Med Rehabil 2004;85(2):290-297. [doi: 10.1016/j.apmr.2003.06.003] [Medline: 14966716]
37. Breiman L, Friedman J, Olshen RA, Stone CJ. Classification and Regression Trees. London: Chapman and Hall/CRC; 2025.
38. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. ArXiv Preprint posted online on November 25, 2017 [FREE Full text] [doi: 10.5260/chara.21.2.8]
39. Ellmers TJ, Ventre JP, Freiberger E, Hauer K, Hogan DB, Lim ML. Does concern about falling predict future falls in older adults? A systematic review and meta-analysis. Age Ageing 2025;54(4):afaf089. [doi: 10.1093/ageing/afaf089] [Medline: 40197783]
40. Shao L, Wang Z, Xie X, Xiao L, Shi Y, Wang ZA, et al. Development and external validation of a machine learning-based fall prediction model for nursing home residents: a prospective cohort study. J Am Med Dir Assoc 2024;25(9):105169. [doi: 10.1016/j.jamda.2024.105169] [Medline: 39067863]
41. Mishra AK, Skubic M, Despins LA, Popescu M, Keller J, Rantz M. Explainable fall risk prediction in older adults using gait and geriatric assessments. Front Digit Health 2022;4:869812 [FREE Full text] [doi: 10.3389/fdgth.2022.869812] [Medline: 35601885]
42. Maiora J, Rezola-Pardo C, García G, Sanz B, Graña M. Older adult fall risk prediction with deep learning and timed up and go (TUG) test data. Bioengineering (Basel) 2024;11(10):1000 [FREE Full text] [doi: 10.3390/bioengineering11101000] [Medline: 39451376]

XSL•FO
RenderX

43. Choi J, Parker SM, Knarr BA, Gwon Y, Youn J. Wearable sensor-based prediction model of timed up and go test in older adults. Sensors (Basel) 2021;21(20):6831 [FREE Full text] [doi: 10.3390/s21206831] [Medline: 34696041]
44. Agrawal DK, Usaha W, Pojprapai S, Wattanapan P. Fall risk prediction using wireless sensor insoles with machine learning. IEEE Access 2023;11:23119-23126. [doi: 10.1109/access.2023.3252886]
45. Qian X, Cheng H, Chen D, Liu Q, Chen H, Jiang H. The smart insole: A pilot study of fall detection. 2019 Presented at: Body Area Networks. Smart IoT and Big Data for Intelligent Health. EAI BODYNETS 2019; October 2-3, 2019; Florence p. 37-49. [doi: 10.1007/978-3-030-34833-5_4]
46. Chen D, Cao H, Chen H, Zhu Z, Qian X, Xu W, et al. Smart insole-based indoor localization system for internet of things applications. IEEE Internet Things J 2019;6(4):7253-7265. [doi: 10.1109/jiot.2019.2915791]
47. Ansah S, Chen D. Stride length estimation with motion data from the contralateral foot. 2025 Presented at: Proceedings of the ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies; July 15-17, 2025; Washington DC p. 407-411. [doi: 10.1145/3721201.3724418]
48. Olugbon F, Ghoreishi N, LaRoche D, Huang MC, Chen D. Optimized smart insole design for temporal gait parameter measurement. In: International Conference on Wireless Intelligent and Distributed Environment for Communication. Cham: Springer; 2024:155-163.
49. Ghoreishi N, Olugbon F, LaRoche D, Arthanat S, LaCourse J, Chen D. Unobtrusive wearable system for load vertical location estimation. Annu Int Conf IEEE Eng Med Biol Soc 2024;2024:1-4. [doi: 10.1109/EMBC53108.2024.10782959] [Medline: 40039554]
50. Ansah S, Olugbon F, Arthanat S, LaRoche D, Chen D. Smart insole based shuffling detection system for improved gait analysis in Parkinson's disease. : IEEE; 2023 Presented at: 2023 IEEE 19th International Conference on Body Sensor Networks (BSN) IEEE; 2023; May 23-25, 2023; Boston p. 1-4. [doi: 10.1109/bsn58485.2023.10331076]
51. Chen D, Ghoreishi N, Olugbon F, Ansah S, Huang M, Yu Q. Optimal pressure sensor locations in smart insoles for heel-strike and toe-off detection. 2022 Presented at: IEEE Biomedical Circuits and Systems Conference (BioCAS); October 13-15, 2022; Taipei p. 458-461. [doi: 10.1109/biocas54905.2022.9948663]
52. Chen D, Asaeikheybari G, Chen H, Xu W, Huang M. Ubiquitous fall hazard identification with smart insole. IEEE J Biomed Health Inform 2021;25(7):2768-2776. [doi: 10.1109/JBHI.2020.3046701] [Medline: 33351772]

Abbreviations

AP: anteroposterior
BBS: Berg Balance Scale
BW: body weight
CDC: Centers for Disease Control and Prevention
CoP: center of pressure
GRF: ground reaction force
IMU: inertial measurement unit
LOSO-CV: leave-one-subject-out cross-validation
ML: mediolateral
RF: random forest
SHAP: Shapley Additive Explanations
STEADI: Stopping Elderly Accidents, Deaths, and Injuries
TUG: Timed Up and Go

Edited by J Zhou; submitted 23.Feb.2026; peer-reviewed by DKA Singh, A Torad, D Trabassi; comments to author 16.Mar.2026; accepted 05.May.2026; published 09.Jun.2026.

Please cite as:
Ghoreishi N, Ansah S, Lu J, Lu W, Moon S, Delgado F, Chen D
Evaluation of the Importance of Stopping Elderly Accidents, Deaths, and Injuries (STEADI)–Based Factors in Wearable Fall Risk Assessment: Secondary Data Analysis
JMIR Mhealth Uhealth 2026;14:e93877
URL: https://mhealth.jmir.org/2026/1/e93877
doi:10.2196/93877
PMID:

©Nozhan Ghoreishi, Stella Ansah, Jiayu Lu, Wei Lu, Sanghee Moon, Ferdinand Delgado, Diliang Chen. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 09.Jun.2026. This is an open-access article distributed under the terms

XSL•FO
RenderX

of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Determining a Likely Mechanism of Missingness in Repeated Measures Sleep Data From Wearable Fitness Trackers: Longitudinal Analysis

Kate Mobley[1], BS, MS, PhD; Kevin B Gittner[1,2], PhD; Karen Nielsen[3], PhD; Lauren M Matheny[1], MPH, PhD; Gita Taasoobshirazi[1], PhD; Charles Natuhamya[4], MS; Monica H Swahn[5], PhD

[1]School of Data Science and Analytics, College of Computing and Software Engineering, Kennesaw State University, 680 Arntson Drive Atrium J333, MD 9104, Marietta, GA, United States

[2]Department of Health Promotion and Physical Education, Wellstar College of Health and Human Services, Kennesaw State University, Kennesaw, GA, United States

[3]Department of Population Health Sciences, School of Public Health, Georgia State University, Atlanta, GA, United States

[4]Uganda Youth Development Link, Kampala, Uganda

[5]School of Public Health, Virginia Commonwealth University, Richmond, VA, United States

Corresponding Author:
Kate Mobley, BS, MS, PhD
School of Data Science and Analytics, College of Computing and Software Engineering, Kennesaw State University, 680 Arntson Drive Atrium J333, MD 9104, Marietta, GA, United States

Abstract

Background: Wearable fitness trackers have become a valuable tool in public health research due to their ability to collect large-scale, individual-level data at a low cost. Because of this, wearables have the potential to mitigate the urgent need for health research in low-resource settings. However, their use has been largely limited to high-income settings, and a major challenge remains: high rates of missing data. This problem may be exacerbated in low-resource environments where logistical and operational barriers further complicate data collection. Wearable sleep data collected in an urban informal settlement in Uganda during The Onward Project on Well-being and Adversity project revealed substantial challenges related to missingness in low-resource research settings.

Objective: The study aimed to characterize the patterns and frequency of missing data and to demonstrate the use of analytical methods to determine the underlying missingness mechanism.

Methods: For this study, 300 women in slum communities in Kampala were equipped with Garmin smartwatches that collected data over 5 days. The nature of missingness was assessed through four methods: pattern analysis, the Little test, a random forest classification model, and a logistic regression classification model. Pattern analysis is an established method of identifying missing data patterns, while the Little test is used specifically to identify missing completely at random data. Random forest and logistic regression models are more recent methods proposed to identify the mechanism of missingness; both were used to examine agreement between the 2 methods and to ensure a thorough examination of missingness patterns.

Results: Approximately 30% of nighttime data were missing. Three patterns were identified that occurred in over 10% of participants' data: no data missing, fifth night of data missing, and all nights of data missing. Pattern analysis and the Little test (P<.001) indicated that the data were not missing completely at random. Both the random forest (area under the curve=0.7) and logistic regression models suggested that the data were missing at random (MAR).

Conclusions: Evidence of missingness in this dataset was consistent with MAR. Potential causes of missingness include device removal, battery failure, and technical malfunctions. These findings have important implications for both wearable device users and future research. Understanding the mechanisms behind missingness can inform strategies to improve data quality, particularly in low-resource settings, and allow researchers to manage missingness using appropriate strategies designed to handle MAR data without introducing bias into the results.

(JMIR Mhealth Uhealth 2026;14:e81123) doi:10.2196/81123

KEYWORDS

global health; wearable electronic devices; sleep; missing data; informal settlements

XSL•FO
RenderX

Introduction

Wearable Fitness Trackers in Health Research

Originally intended for personal health and fitness tracking, wearable fitness trackers have gained traction in research due to their ability to collect large-scale, real-world data from participants in natural settings [1]. According to Fitabase, a platform monitoring research applications of Fitbit devices, over 1300 studies incorporated Fitbit devices between 2012 and 2025 [2] and that number increases when factoring in the studies that have used other fitness trackers. Wearable fitness trackers provide real-time data collection, including physical activity, heart rate metrics, biometric and health indicators, location, behavior, and sleep. These data are critical for identifying risk factors for various health conditions, tracking health outcomes over time, evaluating interventions, and advancing personalized medicine. Wearable device data, when limitations are managed, have the potential to advance understanding of how lifestyle behaviors interact with physical and mental health across diverse populations, supporting more effective, data-driven public health strategies and clinical practices. Despite their growing adoption in research, these technologies remain disproportionately used in high-income countries and among wealthier populations [1,3,4].

Research in Urban Slum Communities

Wearable devices hold significant promise for research, particularly in low-resource settings such as urban slums [4,5]. These settings, defined by constraints across finances, health care, infrastructure, knowledge, research capacity, social resources, environmental factors, human resources, and cultural practices [6], pose unique challenges for data collection. Furthermore, over 1 billion people, or 24% of the world's urban population, currently live in slums, with this figure rising to 56% in sub-Saharan Africa [7]. Wearables offer a practical solution to the need for health research in these settings with large populations and disproportionately high risks of physical and mental health conditions [8-10]. They are cost-effective, scalable, minimally intrusive, and capable of gathering individual-level data for multiple days without reliance on external power or additional technology. These advantages make them especially valuable for research in resource-limited environments [4,5]. Understanding missingness in sleep data from wearable fitness trackers can provide important foundational information for this research.

Understanding the physical and mental health challenges in urban slums and other low-resource settings is a critical public health priority, and sleep serves as a key indicator of overall well-being. Sleep problems, including poor sleep quality and short sleep duration, are common among residents of slum communities and have been linked to mental illness, but this topic is still understudied, especially so in low- and middle-income countries [4]. Wearable technology presents a transformative opportunity for such studies, offering cost-effective, minimally intrusive methods to collect longitudinal health data in challenging urban environments. For instance, while poor sleep is a documented issue in slums, existing research has relied primarily on subjective and unreliable self-reports rather than objective sensor-based measurements [4].

However, while missing data is an issue present in all research involving wearable fitness trackers, the difficulties associated with doing research in low-resource settings like urban slums can contribute to higher rates of missing data [11]. Residents of urban slums often have unreliable electricity access, are unfamiliar with wearable fitness tracker technologies, and face logistical barriers to implementing these technologies [12,13]. Additionally, many wearable fitness trackers are reliant on connection to a smartphone or the internet, and rates of internet access and smartphone usage are low in urban slums [4,14]. Because of these challenges associated with low-resource settings like urban slums, rates of missing data can be higher when wearable fitness trackers are used in these settings. Though wearable studies in slums remain limited and are often clinic-based, there is potential for scalable, real-world data collection using wearable devices [15-17]. To realize this potential, understanding critical challenges like missing data must become a priority.

Missing Data

Missing data remains a pervasive challenge in human research, arising from diverse sources. Participant-related factors, such as missed study visits or attrition, and researcher limitations in follow-up capacity are well-documented contributors [18,19]. The growing integration of digital devices and sensors in research has introduced additional technical sources of missingness, including equipment malfunctions, battery failures, and connectivity issues [20]. Longitudinal studies, which require sustained data collection across multiple timepoints, face compounded risks of missing data, making this issue particularly acute in such designs [21].

Missing data in wearable device studies typically stems from two primary sources: device-related factors and wearer-related behaviors [22,23]. Technical issues such as battery failure, sensor displacement, software malfunctions, connectivity problems, or synchronization errors frequently contribute to device-related missingness [22-24]. Wearer-related missingness most commonly results from nonadherence, including device removal or improper wear [23]. There has been some study of missing data in wearable data, but very little in low-resource settings where wearables have only been used in a small number of studies outside of clinical settings, where conditions are much more controlled. Researchers in low-resource settings have theorized that a higher prevalence of manual labor in these settings results in exposure to extreme conditions, contributing to both device- and wearer-related challenges, which may contribute to the deterioration of wearable devices and increased missingness [5]. Wearables also differ widely in the types of data they collect, and only a few studies have considered missingness in wearable sleep data specifically [23,25]. Despite the growing use of wearable devices to collect sleep data, limited research has examined patterns and mechanisms of missingness, particularly in low-resource settings, where contextual challenges may further exacerbate missing data [11].

The prevalence of missing data significantly impacts data quality and poses substantial analytical challenges [26]. Effective

XSL•FO
RenderX

handling of missing data through methods like imputation becomes crucial for ensuring accurate analyses [24]. A comprehensive review of wearable studies in biomedical research confirms that data missingness remains among the most significant obstacles researchers face [26,27]. Few researchers have used wearables in low-resource settings, but among those who have, numerous studies have found data missingness to be a substantial issue, underscoring the need to understand the mechanisms of missingness and to find robust solutions to this persistent issue [5,28].

Missing data can be categorized into three primary mechanisms: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). When data are MCAR, the probability of missingness is entirely independent of both observed and unobserved data [29,30]. In contrast, MAR occurs when missingness depends only on observed variables, not the missing values themselves [30]. The most challenging scenario, MNAR, arises when missingness is related to the unobserved data [18,31]. Understanding these common classifications of missing data is necessary, as these distinctions depend on whether missingness is related to observed or unobserved information and has important implications for analysis and interpretation [29-31].

Analytical Methods for Determining the Mechanism of Missingness

Determining the exact mechanism of missingness is challenging, but several methods can provide evidence for specific patterns. Visualization techniques offer a preliminary approach to identifying missingness patterns through graphical exploration [32]. For formal testing, the Little test, a well-established method, assesses whether data are MCAR [33]. To explore whether data are MAR, two analytical approaches are commonly used: random forest and logistic regression. Both methods model relationships between observed variables and a binary indicator of missingness. If significant associations exist between observed variables and missingness, this supports the MAR mechanism [34,35]. However, in longitudinal studies, where observations within participants are correlated, these methods require modifications to account for dependencies. In addition to contributing to the body of knowledge on missingness in wearable sleep data, this study demonstrates the use of these analytical methods on wearable sleep data, and on repeated measures data, where adjustments are necessary to preserve the dependencies between observations from the same person.

Wearable Fitness Trackers in Sleep Research

Since around 2017, consumer wearable fitness trackers have measured time spent in each sleep stage using a combination of motion detection, heart rate variability, and respiratory rate monitoring [36]. While questions remain about the absolute accuracy of sleep-stage classification, the devices' ability to detect sleep-wake patterns has significantly improved [37,38]. Importantly, wearable fitness trackers provide more objective sleep data than self-reports, as individuals often cannot accurately recall specific sleep metrics like sleep onset time or nighttime awakenings [36]. Although wearable sleep tracking technology still has limitations, it represents a substantial improvement over subjective reporting for research purposes, with accuracy continuing to advance through technological innovations [37-39].

Garmin devices like the ones used in this study have been assessed for the validity of their measurements. Existing studies indicate that they provide highly valid step count measurements, though their accuracy for distance, energy expenditure, and heart rate data remains more variable [40]. Garmin devices demonstrate reasonable agreement with polysomnography for measuring sleep duration and staging [41], and they more accurately identify deep sleep stages than some other wearable fitness trackers [42]. Manufacturers release new models frequently, and different models from the same brand can vary. Therefore, the comprehensive evaluation of wearable devices across all models and brands, including validity, reliability, proprietary algorithms, cost differences, and feature variability, remains challenging. Overall, however, modern wearable technologies that combine heart rate and motion tracking demonstrate comparable effectiveness for sleep monitoring, despite some variation in specific metrics across models and brands.

Wearable Sleep Data Missingness

While the problem of missingness in wearables data, including sleep data, has been well documented, only 2 studies were identified that considered how to address this problem in sleep data [23,25]. However, these studies differed substantially from this work. First, both studies used sleep and wakefulness data collected every few seconds for up to 50 days, so that missingness was determined by measuring the length of time during the data collection period that no data was collected (a method typically referred to as measuring data completeness). Additionally, one of these studies used data from individuals with very irregular sleep patterns so that the results may be used by those studying sleep disorders, while the other removed participants who had removed the device at any point, a common cause of missing data [23,25]. Neither study used data from low-resource settings where missingness is likely to be even higher than in data from high-income populations.

Study Purpose

The limited research on missing data in wearable sleep data that has been conducted so far uses sleep data that is in a significantly different format from that used in this study. To the authors' knowledge, this is the first study to examine missingness in wearable sleep data aggregated by night within a low-resource setting. This innovative approach to analyzing missingness in wearable sleep data is closely aligned with the typical use and analysis of this type of data.

Wearable fitness trackers offer a promising tool for collecting sleep data in diverse environments, which can allow for further study of sleep and the impact of poor sleep on health. However, missing data remain a significant limitation, and the unique logistical challenges of conducting research in low-resource settings, where sleep and health research is urgently needed, may exacerbate this issue. Addressing the patterns of missingness is essential not only for improving data quality but also for advancing research on sleep's role in health in all settings. This study has two primary objectives: (1) to

XSL•FO
RenderX

characterize the patterns and frequency of missing data and (2) to determine the underlying missingness mechanism using analytical techniques appropriate for longitudinal repeated measures data. Enders [43] comprehensive work on missing data argues that MAR is the most reasonable assumption for the mechanism of missing data in most cases of research, and this assumption should be adopted unless there is evidence to the contrary. Therefore, this study will focus on evidence that the missingness is MCAR or MAR.

Methods

The TOPOWA Project

The data for this study were collected through The Onward Project on Wellbeing and Adversity (TOPOWA), a mixed methods investigation examining the social determinants of mental health among women in Kampala, Uganda's urban slums. Details of the study have been reported previously [4,44]. Participant recruitment was conducted across 3 slum communities (Banda, Bwaise, and Makindye) in collaboration with the local community leaders and Uganda Youth Development Link, a nongovernmental organization serving youth in Kampala's informal settlements.

TOPOWA project used multiple research components, including focus group discussions, Photovoice methodology, and a longitudinal cohort study which includes the collection of sleep data via wearable fitness trackers to study the prevalence of sleep disturbance and its association with mental health in this population.

Ethical Considerations

All study procedures were designed in compliance with the Declaration of Helsinki and received ethical approval from both the Makerere University School of Social Sciences Research Ethics Committee (MAKSSREC MUSSS-2022 - 172) and the Uganda National Council of Science and Technology (SS1560ES).

Uganda Youth Development Link staff at each study site described the TOPOWA study including risks, benefits, and expectations regarding follow-up assessments to interested participants. Informed consent in writing was obtained from those who agreed to participate in the study. Participants were compensated in the local currency for each component of the study that they participated in, including wearing and returning the wearable device and completing the survey, as well as being provided with food and transportation on the days that they traveled to the research hub to complete assessments. All research assistants involved in participant enrollment and data collection were trained in participant privacy, confidentiality, and safety measures. Data were deidentified before being uploaded to REDCap (Research Electronic Data Capture) for storage and shared with members of the research team for analysis.

Data Collection and Processing

The TOPOWA project enrolled 300 women from Kampala's urban slums, equipping each participant with a Garmin vívoactive 3 device to wear continuously for 5 days and nights. These devices were specifically selected for their field suitability: they operate independently without smartphone connectivity, feature a 5-day battery life, and automatically collect sleep data each night. During data collection, 299 devices were successfully returned, with only 1 unit lost. The vívoactive 3 devices record heart rate, heart rate variability, and movement data, which Garmin's proprietary algorithms analyze to classify sleep into 4 stages (rapid eye movement [REM], deep, light, and awake) plus an "unmeasurable" category for periods that cannot be classified [45]. For each sleep session, the devices generate detailed outputs including the duration (in seconds) spent in each sleep stage and precise timestamps marking the beginning and end of the primary sleep period. Notably, when multiple sleep periods occur within a day, the system prioritizes and reports only the longest continuous sleep bout, overriding shorter naps. This design ensures researchers obtain standardized, comparable metrics for a person's main sleep period, which will be discussed as nightly sleep for simplicity, while maintaining the practicality required for field studies in low-resource settings.

During the compilation of baseline sleep data, specific criteria were established for identifying missing data. These criteria focused on total sleep duration, treating nights with no sleep information as missing while retaining nights with vague or low-quality data. Notably, the absence of data for a particular sleep stage on nights when other stages were recorded was not classified as missing data. This pattern could indicate either that the participant genuinely did not enter that sleep stage or that the device failed to detect it—a distinction that cannot be determined from the available data. Consequently, 0 values for specific stages on otherwise recorded nights were treated as valid observations rather than missing data. No nights with data had a 0 recorded for light sleep, but there were nights that contained 0 for the deep sleep, REM sleep, and awake categories. Less than 2.5% of nights with data had a 0 recorded for multiple sleep stages. Similarly, "unmeasurable" sleep seconds were considered (periods the device could not classify into any stage) as legitimate data points and included in total sleep duration calculations, rather than being treated as missing values. Zeros were common in the "unmeasurable" category, indicating that there were no seconds of sleep that the device was unable to classify into one of the other sleep stage categories during those nights. This approach ensured that the analysis reflected the complete sleep architecture as captured by the device while maintaining methodological consistency.

The presence or absence of heart rate and activity data was also recorded. If heart rate data were present for the prior day, then a 1 was recorded, and if heart rate data were not present for the day before the night in question, then a 0 was recorded. If at least 1 activity was recorded during the 5 days that the participant wore the device, then a 1 was recorded for activity data, and if no activity was recorded over the 5 days, then a 0 was recorded. The presence or absence of heart rate data was used in the statistical analysis to determine the likely mechanism of missingness; however, since activity data was only collected during prolonged periods of physical activity and was therefore much less frequent, they were not included in the statistical models. While the Garmin sleep stage classification algorithm

XSL•FO
RenderX

uses concurrently collected heart rate data, the missing data definition used in this study allowed for inclusion of nights with unclassified stages—therefore, the presence of heart rate data was not perfectly correlated with missingness.

Statistical Analysis

The analysis of missing data followed a systematic approach. First, a pattern analysis was conducted using the *naniar* (version 1.1.0; Nicholas Tierney and Dianne Cook) package in R to visually examine the structure of missingness in the dataset [46]. Next, the Little test, a likelihood ratio-based test that compares the expected means for each variable across different missingness patterns [33], was applied to evaluate whether the data were MCAR. For the Little test, the data were converted to wide format, and the 5 sleep category variables (deep, light, REM, awake, and unmeasurable) were included as well as the variables for marital status, household size, education level, and employment status. This test was conducted first because the MCAR criteria are more restrictive than the MAR criteria.

Following this initial assessment, 2 distinct classification methods, a random forest model and a generalized estimating equations (GEE) model with a logit link function, were used to explore whether the missingness pattern was consistent with MAR. For both models, a binary dependent variable indicating whether data were present or missing for each night was created. Both models incorporated six predictor variables: (1) night sequence number (1-5), (2) household size, (3) education level, (4) employment status (binary), (5) marital status (binary), and (6) presence of heart rate data for the day before the night of sleep (binary). These covariates were selected based on established relationships between socioeconomic factors and sleep behaviors in prior research [47,48]. The results from both models were then used to strengthen the conclusions about the nature of the missing data mechanism.

The dataset included participants with complete missingness across all 5 nights of sleep monitoring, represented by NA (not available) values for all sleep stages. While standard practice might suggest listwise deletion for such cases, these participants were retained in this analysis for three key reasons: (1) they provided complete data for all independent variables (household size, marital status, education level, employment status, and heart rate data availability), (2) this study specifically aimed to develop alternatives to conventional listwise deletion approaches, and (3) missingness, even complete missingness, can be informative when evaluating the mechanism and patterns behind the missingness.

While the random forest classification and GEE modeling with logit link functions serve similar analytical purposes for MAR assessment, their concurrent application provides methodological robustness, given that (1) each approach has distinct strengths in handling different data characteristics and (2) their use for MAR determination remains relatively novel in the literature, with foundational publications emerging only in the early 2000s [34,35,49]. This dual-method approach strengthened the conclusions by assessing consistency across complementary analytical frameworks.

Random Forest Classification Model

A random forest classification model was used using the *randomForest* R package (version 4.7.1.1; Andy Liaw and Matthew Wiener) to assess whether the data were consistent with MAR, as this approach can effectively capture complex, nonlinear relationships among observed variables [35]. Model performance was evaluated using the area under the receiver operating characteristic curve. An area under the curve (AUC) value significantly greater than 0.5 provided evidence for MAR, as this indicates that the missingness pattern was systematically related to the observed variables in the dataset [35]. Low AUC values (less than 0.5) indicate that the model performs no better than random guessing and would be expected if the data were MCAR. AUC values above 0.5 can occur if the data are MNAR but are more likely to occur if the data are MAR [35].

A critical consideration in the analysis was that standard random forest classification assumes independence between observations, an assumption violated in longitudinal data where multiple measurements come from the same individual. To address this dependency, a cluster-adjusted bootstrap sampling approach [50] was implemented, which resampled participants (rather than individual nights) to preserve within-subject correlations. While the use of a random forest model to assess evidence of MAR is not novel, modifying that model to account for dependencies between observations in repeated measures data is not a standard use of the method. For this reason, a second method of modeling patterns between observed variables was used.

GEE Model With Logit Link Function

The second approach implemented for exploring a MAR mechanism was a logistic regression model. This method tests whether observed variables can systematically predict missingness patterns—a key criterion for MAR. While several logistic regression-based approaches exist for MAR assessment (including Ridout's method for dropout analysis in longitudinal studies [34] and Fairclough's method for missingness prediction), Fairclough's framework [34] was adapted for this analysis since its purpose was closely aligned with the goal of this study. Independent variables were entered sequentially into the model to assess their association with missingness. A variable's contribution was evaluated based on the regression coefficient estimates, with statistical significance determined using robust standard errors derived from the GEE, to characterize predictors of missingness. Five total models were created with an additional variable added in each model. This iterative approach using forward selection allowed for assessing how different participant characteristics and study factors influenced the probability of data being missing for a given night.

While traditional logistic regression assumes independence between observations, this assumption is violated in repeated measures designs where observations from the same participant are inherently correlated [51]. To address this dependency, a GEE model with a logit link function was used using the *geepack* package in R (version 1.3.12; Søren Højsgaard, Ulrich Halekoh, Jun Yan, and Claus Thorn Ekstrøm), which extends generalized linear models by incorporating specified covariance

XSL•FO
RenderX

structures [52]. Working correlation structures were assessed for fit, including the unstructured, exchangeable, and autoregressive (1). The quasi-likelihood under the independence model information criterion was used to assess which working correlation structure was most suitable [52]. For this analysis, an exchangeable working correlation structure was selected, which assumes that the correlations between observations from the same participant are equally correlated [52]. This selection is appropriate because the data were collected over just 5 consecutive days, so significant correlation decay is not expected in such a short time period. The GEE framework allowed for properly accounting for within-participant correlations while examining how various factors predicted missingness patterns.

While the GEE model with a logit link function and the random forest model perform similarly in that they both model relationships between observed variables, neither is a standard method of assessing the mechanism of missingness that has been well documented in previous publications, as is the case for the Little test. Literature was identified that supported the validity of these methods, but no prior studies using these methods were identified. For these reasons, both modeling methods were performed so that their results could be compared.

Results

Participant Characteristics

All women who participated in the TOPOWA project were aged between 18 and 24 years old. All participants had completed some primary school and could read and write, but only 8% (24/300) had completed secondary school. Over half, 62% (186/300), had given birth and 82% (152/186) of those who had given birth lived with her child(ren). Only 15% (45/300) of the women lived in a single generation household, while the remaining 85% (255/300) lived in households with 2 or more generations. Women who had a job represented 32% (97/300) while the remaining 68% (203/300) of the women were unemployed.

Missing Data in the TOPOWA Baseline Sleep Data

Participants wore Garmin devices continuously for 5 days and nights, yielding sleep data from 299 individuals. However, 37 (12%) devices recorded no sleep data across all 5 nights (complete missingness). The dataset exhibited a notable missing data pattern: nights were either fully observed (all sleep stages recorded) or fully missing (no data for any sleep stage). Consequently, missingness proportions were identical across all sleep metrics (deep, light, REM, awake, and unmeasurable seconds), with an overall missing data rate of 30% (455 nights missing out of 1495 total nights). The rate of missingness was similar across the 3 study sites (Banda, Bwaise, and Makindye) with Banda having a slightly higher rate of missingness than Bwaise and Makindye (Table 1). The rate of missingness in the sleep data remained consistent as household size increased but increased slightly with each additional child a woman had, so that participants with no children had the lowest rate of missingness (155/570 nights of data, 27%) and participants with 3 children had the highest rate of missingness (18/25 nights of data, 72%) (Table 1). The rate of missingness in the sleep data was approximately equal across changes in income, employment status, education level, and marital status. Notably, 58% (173/299) of participants experienced at least 1 night of missing data, demonstrating the pervasiveness of this issue.

Table . Proportion of missing data by study site, night of data collection, and number of children.

Rate and number of nights of data missing	Nights of data (n=495) missing from Banda participants, n/N (%)	Nights of data (n=500) missing from Bwaise participants, n/N (%)	Nights of data (n=500) missing from Makindye participants, n/N (%)	Total nights of data (n=1495) missing from all participants, n/N (%)
Night of data collection				
1	22/99 (22)	25/100 (25)	20/100 (20)	67/299 (22)
2	32/99 (32)	26/100 (26)	21/100 (21)	79/299 (26)
3	32/99 (32)	25/100 (25)	24/100 (24)	81/299 (27)
4	35/99 (35)	21/100 (21)	32/100 (32)	88/299 (29)
5	49/99 (49)	44/100 (44)	47/100 (47)	140/299 (47)
Number of children				
0	34/155 (22)	65/155 (42)	56/155 (36)	155/570 (27)
1	89/188 (47)	45/188 (24)	54/188 (29)	188/625 (30)
2	35/94 (37)	26/94 (28)	33/94 (35)	94/275 (34)
3	11/18 (61)	7/18 (39)	N/A[a]	18/25 (72)

[a]N/A: not available.

The 37 cases of complete missingness for all 5 nights of sleep data can be categorized into 3 distinct scenarios with varying probable causes, detailed in Table 2.

XSL•FO
RenderX

Table . Categorization of the 37 cases of complete sleep data missingness.

Status of sleep and heart rate data	Number of cases
No sleep or heart rate data collected	1 device
Partial data collection—either heart rate or activity (but not both) data collected for at least 1 day	8 devices
Selective sleep data missingness—devices collected both heart rate and activity data for at least 1 day, but no sleep data	28 devices

Data Exploration

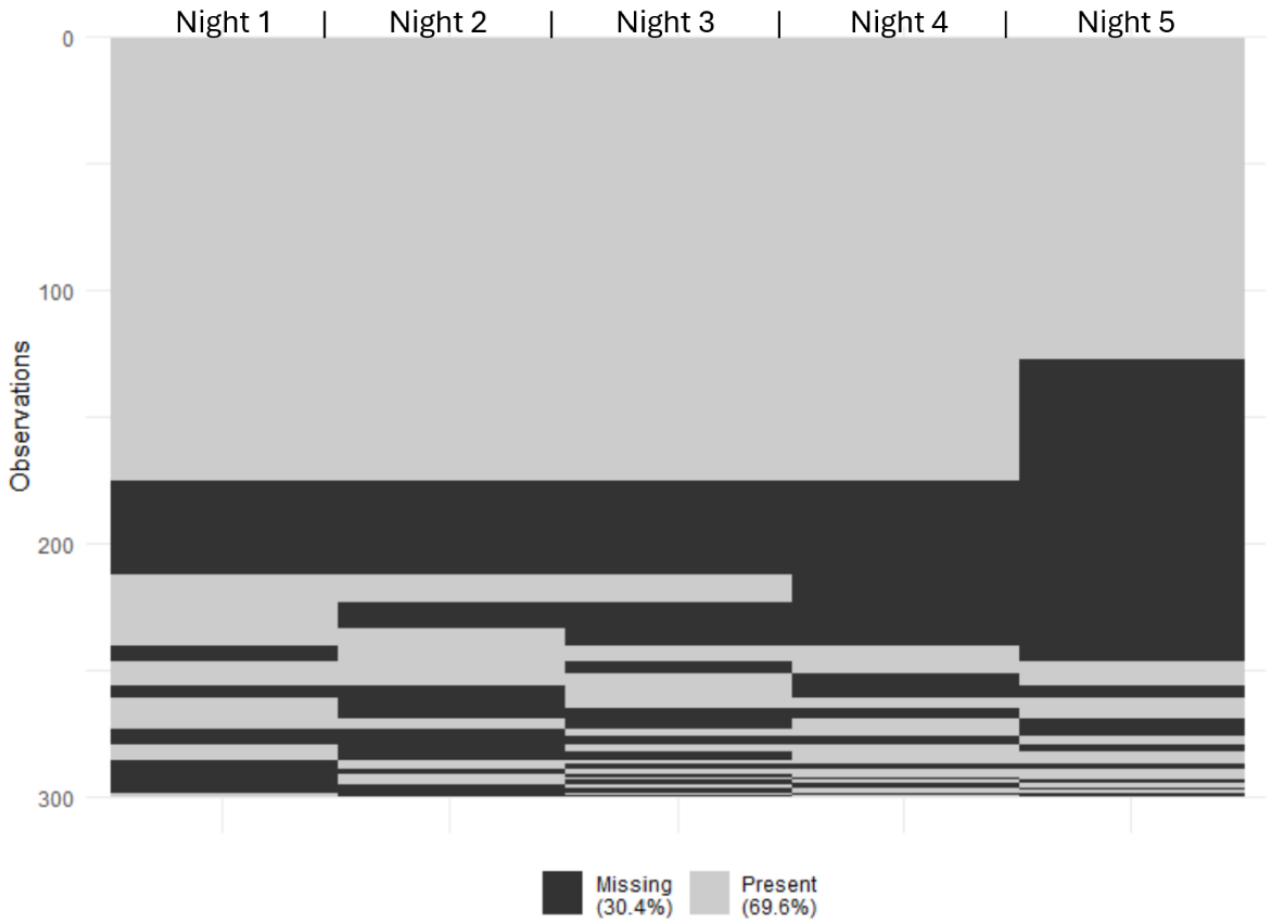
The distributions of time spent in each sleep stage and total time asleep were examined to determine if the distributions differed based on the day of the data collection (1-5). For each sleep stage, as well as total time asleep, the distribution was the same both when the 5 nights of data were considered together and when the data were grouped by night. There were no changes in the distribution of sleep times over the course of the 5 days of data collection.

There were some outliers in the total sleep durations. For 1 participant, less than 1 hour of sleep was recorded in 1 night, and for another participant, just over 12.5 hours of sleep was recorded on 1 night. Because the outlier values present in the data are not outside the realm of possibility, these observations were retained when assessing the missing data mechanism.

Missingness Patterns

The first night of data collection has the least amount of missing data, with 22% (66/300) of nights missing data. The rate of missing data increases for each subsequent night, and the fifth and final night of data collection has 47% (141/300) of data missing (Table 1). There were 29 missingness patterns identified in the data (Figure 1), with no nights missing being the most common (126 cases), followed by missing only the last night of data (48 cases) and then missing all nights of data (37 cases).

Figure 1. Patterns of missingness in the sleep data. Missingness patterns are ordered from most common at the top (no missing nights of data) to the least common (bottom).



The Little Test

When running the Little test, the null hypothesis is that the data are MCAR while the alternative hypothesis is that the data are not MCAR and are therefore either MAR or MNAR. The χ^2 statistic was 969 and the degrees of freedom was 587, which resulted in a P value for the Little test of P<.001. Therefore,

XSL•FO
RenderX

the null hypothesis is rejected. This suggests that the data were not MCAR.

Random Forest Classification Model

The AUC value of 0.70 for the random forest classification model indicated that the model was generally able to distinguish missingness, which is consistent with data MAR. The sensitivity, a measure of how well true positives (ie, missing data) are predicted, was 0.30 and the specificity, a measure of how well true negatives (ie, data present) are identified, was 0.94. The overall accuracy for the model was 0.75. Table 3 shows the confusion matrix for the random forest classification model.

Table . Random forest classification model confusion matrix.

Predicted values	True values	
	0 (data present)	1 (data missing)
0 (data present)	978	319
1 (data missing)	62	136

GEE Model With Logit Link Function

The results of the GEE model with a logit link function also yielded results that were consistent with a MAR assumption. The independent variables were added to the model in an iterative process to determine the extent to which each of the variables was able to predict missingness. The final model (Table 4) contained only 2 predictor variables that were statistically significant, number of night in the 5-night sequence (1-5) and whether the participant had a job. The model had an accuracy of 0.70. The sensitivity was 0.12, while the specificity was 0.96. The confusion matrix for the model is displayed in Table 5.

Table . Results of generalized estimating equations model with logit link function.

	Parameter estimate	SE	Odds ratio (95% CI)	*P* value
Intercept	–1.8546	0.1850	0.156 (0.109 - 0.225)	<.001
Number of night in the 5-night sequence	0.2554	0.0357	1.291 (1.204 - 1.384)	<.001
Job	0.6671	0.2016	1.948 (1.312 - 2.893)	.001

Table . Confusion matrix for the generalized estimating equations model with logit link function.

Predicted values	True values	
	0 (data present)	1 (data missing)
0 (data present)	996	402
1 (data missing)	44	53

Comparing the Results of the 2 Classification Models

The random forest model and the logistic regression model agreed in 90% (1346) of cases. Table 6 contains the confusion matrix showing how often the 2 models agreed and how often they disagreed in their classifications.

Table . Confusion matrix comparing the random forest and logistic regression models.

Logistic regression predictions	Random forest predictions	
	0 (data present)	1 (data missing)
0 (data present)	1281	117
1 (data missing)	32	65

Discussion

Principal Results

Our comprehensive analysis provides evidence that the missing sleep data in the TOPOWA study follows a MAR pattern, thereby permitting the use of imputation and likelihood-based methods that would be inappropriate under an MNAR mechanism. Furthermore, Enders [43] recommends adopting MAR as the default assumption in most realistic research scenarios. The findings in this study indicate that missingness was systematically related to observable variables rather than unobserved sleep characteristics [21,53]. Both the random forest model and the GEE model with a logit link function had high specificity values, indicating that the models identified true negatives well (present data values), but the values for sensitivity were low, indicating that the models performed worse at identifying true positives (missing data values). The classification models more accurately identified when data

XSL•FO
RenderX

values were present than when data values were missing. However, previous researchers who have used this method to check for evidence that missing data are MAR have focused on the AUC value and did not report sensitivity and specificity [35].

While additional studies using varying types of data, such as supplementary self-reported sleep data, are necessary to further validate the methodology used in this study, the results suggest that a GEE model with a logit link function and a random forest classification model can be used to identify evidence that missing data are MAR in repeated measures data when modifications are made to account for the dependencies between observations from the same person. This finding has the potential to improve researchers' handling of missing data in longitudinal data, common in health research, by providing a way to empirically determine the likely mechanism of missingness prior to analysis and decide how best to handle the missing data based on that mechanism of missingness.

Possible Explanations for Missing Data

Several key observations support the conclusion that the missingness is MAR. First, the temporal pattern of missingness showed a clear progression, with later nights (particularly night 5) having higher rates of missingness than earlier nights. Second, known explanations for missingness—including device failure (n=1 likely case), nonadherence (n=3 self-reported cases), or improper device fit—would generally qualify as MAR mechanisms. While 37 participants had no sleep data for any of the 5 nights, the available evidence suggests these cases resulted from technical or behavioral factors rather than being related to the missing sleep values themselves.

Missing data in the baseline sleep dataset likely stemmed from multiple sources, each with distinct implications for analysis. Battery depletion emerged as a primary factor, with some devices failing to maintain charge through all 5 nights, according to reports from the research assistants. This explanation aligns with the observed increase in missingness across successive nights and qualifies as MAR, as battery failure was unrelated to participants' sleep patterns. The most common missingness pattern in the data was data missing on the fifth night only, which likely occurred because the battery was depleted before the fifth night. The third most common missingness pattern was data present for the first 3 nights and missing for the fourth and fifth nights, which also likely occurred due to a depleted battery. A subset of cases involved unpredictable device malfunctions (MCAR), where hardware failures caused complete data loss independent of user behavior or sleep characteristics. This was likely the case for the 1 participant for which no heart rate, activity, or sleep data was collected, and may have been the case for the 8 participants for whom either heart rate data or activity data was collected (but not both) and no sleep data was collected. Another possibility is that the missing data in the underlying raw data (heart rate variability and motion data) resulted in insufficient physiological data for sleep algorithm processing. This pattern highlights the technical and behavioral challenges inherent in wearable research, where algorithm-derived outcomes (like sleep duration and stages) depend on sensor data.

Participant-related issues also contributed: some women reported removing devices due to social discomfort (eg, partner or child concerns) [14], while improper fit (eg, loose wrist placement) prevented sensor contact, mimicking nonwear. In the case of the 8 participants for whom no sleep data was collected but either heart rate or activity data was collected, and the 28 participants for which heart rate and activity data was collected but sleep data was not, intermittent device removal (ie, at night) may have been a factor contributing to sleep data missingness. The coexistence of these mechanisms, and the potential for both technical and behavioral reasons for missingness, highlights the complexity of missingness in wearable research and underscores the need for robust analytical approaches to mitigate bias.

While only 3 women reported removing the device for any period of time during the 5-day data collection period, the missingness patterns in the data from considerably more devices suggested that device removal was a possible factor. The second most common missingness pattern in the data was data missing for all 5 nights, which could indicate the participants removed the device. Notably, the discrepancy between self-reported and probable nonwear suggests possible social desirability bias, as participants received compensation for participation.

Limitations

Several important limitations should be considered when interpreting this study's findings. The presence or absence of other physiological or behavioral metrics that wearable devices can capture (eg, activity levels, heart rate variability) was examined, but only the presence or absence of this data was tracked. Detailed information on heart rate or activity was not analyzed. Low-quality data may also mask data that should have been marked as missing. Some extreme sleep duration values (both small and large) were identified that could represent either genuine biological variation or measurement artifacts. Wearable devices may misclassify wakefulness as sleep during prolonged sedentary periods or conversely interpret restless sleep as wakefulness. While these devices generally provide more objective sleep measurements than self-reports, such classification errors underscore the importance of interpreting extreme values cautiously. Future studies could benefit from incorporating complementary data sources such as self-reported sleep to validate these outlier observations.

While it is possible that the day of the week may impact missingness, this was not explored in this analysis for 2 reasons. The women in this study were issued Garmin devices for 5-day periods on a rolling basis, meaning that not all women wore the devices over the weekend. The second reason that considering weekday versus weekend missingness is not ideal using this data is that the lifestyle and employment status of these women are very different from those of middle- and working-class populations, as well as Western assumptions of a structured 5-day work week, who typically have consistent and differing routines on weekdays and weekends. Most of these women did not have jobs with regular Monday-Friday hours.

Lastly, questions regarding the validity and reliability of consumer-grade wearable devices remain and should be further researched, especially because validity can vary between different device models from the same brand using similar

XSL•FO
RenderX

technology. As health research using wearable devices continues and expands, the study of the validity of these devices should continue as well, especially as new models and technology that have not been previously studied are released. Additionally, self-reported sleep data has long been used for sleep research and can be a valuable tool when used as complementary data for validating sleep data from wearable technology.

Conclusion

This work demonstrates the process researchers can undertake when assessing missing data assumptions prior to the analysis of wearable data. Taken together, these findings suggest that a MAR mechanism can be assumed for sleep data in the TOPOWA study, with missingness driven primarily by observable technical and behavioral factors rather than unobserved sleep characteristics. This distinction is critical for selecting appropriate imputation strategies and ensuring valid inferences in future analyses. As wearable methods expand in global health research, especially in urban informal settlements, addressing the practical and ethical complexities of data missingness will be essential for advancing equity, rigor, and reproducibility. In addition to considering data collected in a low-resource setting, where research faces additional challenges, this study contributes meaningful information on missingness in wearable sleep data aggregated by night, closely aligning it with the common use and analysis of this type of data.

Acknowledgments

The authors would like to thank their partners at Uganda Youth Development Link staff and research assistants who supported The Onward Project on Well-being and Adversity project.

Funding

Research reported in this publication was funded by the National Institute of Mental Health, of the National Institutes of Health, under award number R01MH128930 to Swahn. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability

The deidentified participant-level sleep data may be obtained by reasonable request to the principal investigator or corresponding author.

Authors' Contributions

Conceptualization: KM (lead), KBG (supporting), KN (supporting), LMM (supporting)
Data curation: MHS, KM, KN
Formal analysis: KM
Funding acquisition: MHS
Investigation: MHS, KM, KN, KBG
Methodology: KM (lead), KBG (supporting), KN (supporting), LMM (supporting), GT (supporting)
Project administration: MHS (lead), KM, KN, CN, KBG
Resources: MHS
Supervision: KBG, MHS, KN, LMM, GT
Writing – original draft: KM
Writing – review & editing: KM, KBG, KN, LMM, GT, CN, MHS

Conflicts of Interest

None declared.

References

1. Huhn S, Axt M, Gunga HC, et al. The impact of wearable technologies in health research: scoping review. JMIR Mhealth Uhealth 2022 Jan 25;10(1):e34384. [doi: 10.2196/34384] [Medline: 35076409]
2. Fitbit research library. Fitabase. URL: https://www.fitabase.com/research-library/ [accessed 2025-06-02]
3. Holko M, Litwin TR, Munoz F, et al. Wearable fitness tracker use in federally qualified health center patients: strategies to improve the health of all of us using digital health devices. NPJ Digit Med 2022 Apr 25;5(1):53. [doi: 10.1038/s41746-022-00593-x] [Medline: 35469045]
4. Nielsen K, Mobley K, Culbreth R, Palmier J, Nabulya A, Swahn MH. Feasibility and acceptability of wearable devices and daily diaries to assess sleep and other health indicators among young women in the slums of Kampala, Uganda. Digit Health 2024;10:20552076241288754. [doi: 10.1177/20552076241288754] [Medline: 39430703]
5. Huhn S, Matzke I, Koch M, et al. Using wearable devices to generate real-world, individual-level data in rural, low-resource contexts in Burkina Faso, Africa: a case study. Front Public Health 2022;10:972177. [doi: 10.3389/fpubh.2022.972177] [Medline: 36249225]

XSL•FO
RenderX

6. van Zyl C, Badenhorst M, Hanekom S, Heine M. Unravelling “low-resource settings”: a systematic scoping review with qualitative content analysis. BMJ Glob Health 2021 Jun;6(6):e005190. [doi: 10.1136/bmjgh-2021-005190] [Medline: 34083239]
7. Make cities and human settlements inclusive, safe, resilient and sustainable. United Nations Statistics Division. URL: https://unstats.un.org/sdgs/report/2021/goal-11/ [accessed 2025-06-02]
8. Sclar ED, Garau P, Carolini G. The 21st century health challenge of slums and cities. Lancet 2005;365(9462):901-903. [doi: 10.1016/S0140-6736(05)71049-7] [Medline: 15752535]
9. Nejad FN, Ghamari MR, Kamal SHM, Tabatabaee SS, Ganjali R. The most important social determinants of slum dwellers' health: a scoping review. J Prev Med Public Health 2021 Jul;54(4):265-274. [doi: 10.3961/jpmph.21.073] [Medline: 34370940]
10. Ompad DC, Galea S, Caiaffa WT, Vlahov D. Social determinants of the health of urban populations: methodologic considerations. J Urban Health 2007 May;84(3 Suppl):i42-i53. [doi: 10.1007/s11524-007-9168-4] [Medline: 17458704]
11. Ustinov Y, Lichstein KL. Actigraphy reliability with normal sleepers. Behav Sleep Med 2013;11(5):313-320. [doi: 10.1080/15402002.2012.688779] [Medline: 23268697]
12. Gupta N, Crowley P, Holtermann A, Straker L, Stamatakis E, Ding D. Are we ready for wearable-based global physical activity surveillance? Br J Sports Med 2024 Mar 21;58(7):356-358. [doi: 10.1136/bjsports-2023-106780] [Medline: 38336382]
13. Liverani M, Ir P, Perel P, Khan M, Balabanova D, Wiseman V. Assessing the potential of wearable health monitors for health system strengthening in low- and middle-income countries: a prospective study of technology adoption in Cambodia. Health Policy Plan 2022 Sep 13;37(8):943-951. [doi: 10.1093/heapol/czac019] [Medline: 35262172]
14. Swahn MH, Gittner KB, Lyons MJ, et al. Advancing mHealth research in low-resource settings: young women's insights and implementation challenges with wearable smartwatch devices in Uganda. Sensors (Basel) 2024 Aug 29;24(17):5591. [doi: 10.3390/s24175591] [Medline: 39275502]
15. Garbern S, Mbanjumucyo G, Umuhoza C, et al. Feasibility and accuracy of a wearable biosensor device for vital sign monitoring in septic emergency department patients in Rwanda. Prehosp Disaster med 2019 May;34(s1):s85-s86. [doi: 10.1017/S1049023X1900178X]
16. Ghomrawi HM, Many BT, Holl JL, et al. Clinicians' perspectives on wearable sensor technology as an alternative bedside monitoring tool in two West African countries. Int J Med Inform 2023 Jul;175:105046. [doi: 10.1016/j.ijmedinf.2023.105046] [Medline: 37148867]
17. Van HMT, Hao NV, Phan Nguyen Quoc K, et al. Vital sign monitoring using wearable devices in a Vietnamese intensive care unit. BMJ Innov 2021 Mar;7(Suppl 1):s7-s11. [doi: 10.1136/bmjinnov-2021-000707]
18. Cao Y, Allore H, Vander Wyk B, Gutman R. Review and evaluation of imputation methods for multivariate longitudinal data with mixed-type incomplete variables. Stat Med 2022 Dec 30;41(30):5844-5876. [doi: 10.1002/sim.9592] [Medline: 36220138]
19. Liu Y, De A. Multiple imputation by fully conditional specification for dealing with missing data in a large epidemiologic study. Int J Stat Med Res 2015;4(3):287-295. [doi: 10.6000/1929-6029.2015.04.03.7] [Medline: 27429686]
20. Adhikari D, Jiang W, Zhan J, et al. A comprehensive survey on imputation of missing data in Internet of Things. ACM Comput Surv 2023 Jul 31;55(7):1-38. [doi: 10.1145/3533381]
21. Huque MH, Carlin JB, Simpson JA, Lee KJ. A comparison of multiple imputation methods for missing data in longitudinal studies. BMC Med Res Methodol 2018 Dec 12;18(1):168. [doi: 10.1186/s12874-018-0615-6] [Medline: 30541455]
22. Braem CIR, Yavuz US, Hermens HJ, Veltink PH. Missing data statistics provide causal insights into data loss in diabetes health monitoring by wearable sensors. Sensors (Basel) 2024 Feb 27;24(5):1526. [doi: 10.3390/s24051526] [Medline: 38475061]
23. Weed L, Lok R, Chawra D, Zeitzer J. The impact of missing data and imputation methods on the analysis of 24-hour activity patterns. Clocks & Sleep 2022;4(4):497-507. [doi: 10.3390/clockssleep4040039]
24. Feng T, Narayanan S. Imputing missing data in large-scale multivariate biomedical wearable recordings using bidirectional recurrent neural networks with temporal activation regularization. Presented at: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Jul 23-27, 2019 URL: https://ieeexplore.ieee.org/document/8856966 [accessed 2026-03-09] [doi: 10.1109/EMBC.2019.8856966]
25. Lee MP, Hoang K, Park S, et al. Imputing missing sleep data from wearables with neural networks in real-world settings. Sleep 2024 Jan 11;47(1):zsad266. [doi: 10.1093/sleep/zsad266] [Medline: 37819273]
26. Cho S, Weng C, Kahn MG, Natarajan K. Identifying data quality dimensions for person-generated wearable device data: multi-method study. JMIR Mhealth Uhealth 2021 Dec 23;9(12):e31618. [doi: 10.2196/31618] [Medline: 34941540]
27. Hicks JL, Althoff T, Sosic R, et al. Best practices for analyzing large-scale health data from wearables and smartphone apps. NPJ Digit Med 2019;2:45. [doi: 10.1038/s41746-019-0121-1] [Medline: 31304391]
28. Matzke I, Huhn S, Koch M, et al. Assessment of heat exposure and health outcomes in rural populations of western Kenya by using wearable devices: observational case study. JMIR Mhealth Uhealth 2024 Jul 4;12(1):e54669. [doi: 10.2196/54669] [Medline: 38963698]

XSL•FO
RenderX

29. McKnight PE, McKnight KM, Sidani S, Figueredo AJ. Missing Data: A Gentle Introduction: Guilford Press; 2007. URL: https://www.guilford.com/books/Missing-Data/McKnight-McKnight-Sidani-Figueredo/9781593853938 [accessed 2026-03-09]
30. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res 2011 Mar;20(1):40-49. [doi: 10.1002/mpr.329] [Medline: 21499542]
31. Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. J Clin Epidemiol 2019 Jun;110:63-73. [doi: 10.1016/j.jclinepi.2019.02.016] [Medline: 30878639]
32. Fernstad SJ. To identify what is not there: a definition of missingness patterns and evaluation of missing value visualization. Inf Vis 2019 Apr;18(2):230-250. [doi: 10.1177/1473871618785387]
33. Little RJA. A test of missing completely at random for multivariate data with missing values. J Am Stat Assoc 1988 Dec;83(404):1198-1202. [doi: 10.1080/01621459.1988.10478722]
34. Fielding S, Fayers PM, Ramsay CR. Investigating the missing data mechanism in quality of life outcomes: a comparison of approaches. Health Qual Life Outcomes 2009 Jun 22;7:57. [doi: 10.1186/1477-7525-7-57] [Medline: 19545408]
35. Sondhi A, Weberpals J, Yerram P, et al. A systematic approach towards missing lab data in electronic health records: a case study in non-small cell lung cancer and multiple myeloma. CPT Pharmacom Syst Pharma 2023 Sep;12(9):1201-1212. [doi: 10.1002/psp4.12998]
36. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of wristband Fitbit models in assessing sleep: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e16273. [doi: 10.2196/16273] [Medline: 31778122]
37. Berryhill S, Morton CJ, Dean A, et al. Effect of wearables on sleep in healthy individuals: a randomized crossover trial and validation study. J Clin Sleep Med 2020 May 15;16(5):775-783. [doi: 10.5664/jcsm.8356] [Medline: 32043961]
38. Roberts DM, Schade MM, Mathew GM, Gartenberg D, Buxton OM. Detecting sleep using heart rate and motion data from multisensor consumer-grade wearables, relative to wrist actigraphy and polysomnography. Sleep 2020 Jul 13;43(7):zsaa045. [doi: 10.1093/sleep/zsaa045] [Medline: 32215550]
39. Sathyanarayana A, Joty S, Fernandez-Luque L, et al. Sleep quality prediction from wearable data using deep learning. JMIR Mhealth Uhealth 2016 Nov 4;4(4):e125. [doi: 10.2196/mhealth.6562] [Medline: 27815231]
40. Evenson KR, Spade CL. Review of validity and reliability of Garmin activity trackers. J Meas Phys Behav 2020 Jun;3(2):170-185. [doi: 10.1123/jmpb.2019-0035] [Medline: 32601613]
41. Schyvens AM, Van Oost NC, Aerts JM, et al. Accuracy of Fitbit Charge 4, Garmin Vivosmart 4, and WHOOP versus polysomnography: systematic review. JMIR Mhealth Uhealth 2024 Mar 27;12:e52192. [doi: 10.2196/52192] [Medline: 38557808]
42. Grandner MA, Lujan MR, Ghani SB. Sleep-tracking technology in scientific research: looking to the future. Sleep 2021 May 14;44(5):zsab071. [doi: 10.1093/sleep/zsab071] [Medline: 33782687]
43. Enders CK. Applied Missing Data Analysis, 2nd edition: Guilford Publications; 2022. URL: https://www.guilford.com/books/Applied-Missing-Data-Analysis/Craig-Enders/9781462549863 [accessed 2026-03-09]
44. Culbreth RE, Nielsen KE, Mobley K, Palmier J, Bukuluki P, Swahn MH. Life satisfaction factors, stress, and depressive symptoms among young women living in urban Kampala: findings from the TOPOWA Project pilot studies. Int J Environ Res Public Health 2024 Feb 6;21(2):184. [doi: 10.3390/ijerph21020184] [Medline: 38397675]
45. How Garmin watches track your sleep, calculate sleep score. Garmin. URL: https://www.garmin.com/en-US/blog/fitness/how-garmin-watches-track-your-sleep-calculate-sleep-score/ [accessed 2025-06-03]
46. Zhang Z. Missing data exploration: highlighting graphical presentation of missing pattern. Ann Transl Med 2015 Dec;3(22):356. [doi: 10.3978/j.issn.2305-5839.2015.12.28] [Medline: 26807411]
47. Grandner MA, Jackson NJ, Izci-Balserak B, et al. Social and behavioral determinants of perceived insufficient sleep. Front Neurol 2015;6:112. [doi: 10.3389/fneur.2015.00112] [Medline: 26097464]
48. Mazloomi SN, Talebi S, Kazemi M, et al. Food insecurity is associated with the sleep quality and quantity in adults: a systematic review and meta-analysis. Public Health Nutr 2023 Apr;26(4):792-802. [doi: 10.1017/S1368980022002488] [Medline: 36416108]
49. Fairclough DL. Design and Analysis of Quality of Life Studies in Clinical Trials: Chapman and Hall/CRC; 2002. URL: https://www.taylorfrancis.com/books/mono/10.1201/9781420035438/design-analysis-quality-life-studies-clinical-trials-diane-fairclough [accessed 2026-03-09]
50. Karpievitch YV, Hill EG, Leclerc AP, Dabney AR, Almeida JS. An introspective comparison of random forest-based classifiers for the analysis of cluster-correlated data by way of RF++. PLoS One 2009 Sep 18;4(9):e7087. [doi: 10.1371/journal.pone.0007087] [Medline: 19763254]
51. Hoffman L. Longitudinal Analysis: Modeling within-Person Fluctuation and Change: Routledge; 2015. [doi: 10.4324/9781315744094]
52. Ballinger GA. Using generalized estimating equations for longitudinal data analysis. Organ Res Methods 2004 Apr;7(2):127-150. [doi: 10.1177/1094428104263672]
53. Jadhav A, Pramod D, Ramanathan K. Comparison of performance of data imputation methods for numeric dataset. Appl Artif Intell 2019 Aug 24;33(10):913-933. [doi: 10.1080/08839514.2019.1637138]

XSL•FO
RenderX

Abbreviations

AUC: area under the curve
GEE: generalized estimating equations
MAR: missing at random
MCAR: missing completely at random
MNAR: missing not at random
REDCap: Research Electronic Data Capture
REM: rapid eye movement
TOPOWA: The Onward Project on Well-being and Adversity

Edited by G Badicu; submitted 22.Jul.2025; peer-reviewed by A Martinko, ML Mauriello; revised version received 25.Jan.2026; accepted 09.Feb.2026; published 31.Mar.2026.

Please cite as:
Mobley K, Gittner KB, Nielsen K, Matheny LM, Taasoobshirazi G, Natuhamya C, Swahn MH
Determining a Likely Mechanism of Missingness in Repeated Measures Sleep Data From Wearable Fitness Trackers: Longitudinal Analysis
JMIR Mhealth Uhealth 2026;14:e81123
URL: https://mhealth.jmir.org/2026/1/e81123
doi:10.2196/81123

© Kate Mobley, Kevin B Gittner, Karen Nielsen, Lauren M Matheny, Gita Taasoobshirazi, Charles Natuhamya, Monica H Swahn. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 31.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Determining Cluster-Specific Differences in the Number of Days Required to Reliably Predict Habitual Physical Activity: Intraclass Correlation Resampling Analysis

Conor Jordan Murphy[1,2], PhD; Gabriel M Jouan[1,3], PhD; Katrin Y Friðgeirsdóttir[1,2], PhD; Anna Sigridur Islind[1,3], PhD; Jose M Saavedra[1,2], PhD; María Óskarsdóttir[3], PhD; Erna Sif Arnardóttir[1], PhD

[1]School of Technology, Reykjavik University Sleep Institute, Reykjavík University, Reykjavik, Capital Region, Iceland

[2]Sports Science Department, Physical Activity, Physical Education, Sport and Health (PAPESH) Research Centre, School of Social Sciences, Reykjavík University, Reykjavik, Capital Region, Iceland

[3]Department of Computer Science, Reykjavík University, Reykjavik, Capital Region, Iceland

Corresponding Author:
Conor Jordan Murphy, PhD
School of Technology, Reykjavik University Sleep Institute
Reykjavík University
Menntavegur 1
Reykjavik, Capital Region, 101
Iceland
Phone: 354 599 6200
Email: conorm@ru.is

Abstract

Background: Previous research has attempted to determine the minimum number of days of accelerometry required to reliably reflect an individual's physical activity. However, human behaviors on a day-to-day basis can be highly variable. As a consequence, the number of days required to reliably predict habitual physical activity is dependent on the variability that exists within an individual. There is a concern that adopting generic recommendations from previous research could provide unreliable estimates by failing to represent individuals with specific physical activity patterns.

Objective: The main aim of this study was to identify clusters of individuals with distinct physical activity patterns and to determine if the number of days of accelerometry data required to reliably estimate short- (7 days) and medium-term (28 days) physical activity differed between each unique cluster.

Methods: Accelerometry data were retrieved from 2 independent research studies. Participants during each study had their physical activity recorded using a Withings Scanwatch (Withings Health Solutions). Following a data eligibility process, agglomerative hierarchical clustering was used to identify clusters of individuals based on their physical activity. The clusters were determined using 4 dimensions; mean, SD, skewness, and kurtosis of the step count data. Intraclass correlation coefficients (ICCs) of step count were then calculated within each physical activity cluster. A series of ICCs were computed by separately comparing the average step count across the full periods (7 and 28, for the short- and medium-term analysis, respectively) to a series of averaged subsamples (ranging from 1-6 days and 1-27 days, for the short- and medium-term analysis, respectively). For each subsample, 500 random combinations were generated and compared, providing a distribution of ICCs for each subsample. An ICC of ≥0.80 identified when the subsample of days was sufficient to achieve appropriate reliability.

Results: Of 258 participant datasets, 149 were eligible for the short-term analysis and 64 were eligible for the medium-term analysis. Following agglomerative hierarchical clustering, 4 and 3 clusters of sufficient size (n≥12) were identified in the short-term and medium-term analyses, respectively. When considering the short-term analysis, to achieve a mean ICC score greater than or equal to 0.80, using all randomized combinations, the number of days ranged from 2 to 6 days depending on the physical activity cluster. For the medium-term analysis, the number of days required to achieve a mean ICC score greater than or equal to 0.80 ranged from 6 to 11 days. The short-term analysis clusters displayed more diversity in physical activity patterns than the medium-term analysis.

Conclusions: Physical activity patterns influence the number of days required to estimate habitual physical activity. Thus, to avoid unreliable estimates of physical activity, which could significantly impact the interpretation of results, researchers should be mindful of the physical activity patterns of their sample before adopting generic recommendations.

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2026;14:e64323) doi:10.2196/64323

KEYWORDS

physical activity; accelerometry; reliability; intraclass correlation coefficient; human behavior; data; wearable devices; wearables; self-reporting; habitual; exercise; step counts; steps; walking

Introduction

Accelerometers are wearable devices that measure the acceleration of the body region to which the device is secured. Commercially, the signals are then processed and transformed into physical activity metrics that are relatable to the consumer, such as daily step count. The popularity of accelerometry as a method to quantify physical activity is growing due to the objectivity and availability of such devices. Importantly, the use of accelerometry provides an alternative option to self-reporting questionnaires, which can lack strong reliability and validity [1].

The field is rapidly developing with the advancement of new technology and the commercialization of devices. As a consequence, the approaches used to interpret collected data can vary and lack a consensus [2]. An area of importance relates to the minimum number of valid days required to reliably reflect a person's habitual level of physical activity [3]. Unfortunately, many studies with important implications, such as those linking physical activity to health outcomes, have used varying criteria to define a participant's habitual level of physical activity. For example, the United Kingdom biobank guidelines advise researchers to exclude participants from their accelerometry dataset who have fewer than 3 valid days of data [4], which has been adopted in many impactful studies [5-7]. In contrast, the National Center for Health Statistics, which administers the National Health and Nutrition Examination Survey in the United States, does not advise on the number of valid days that should be available to validate inclusion when using their accelerometry datasets. As a consequence, researchers have adopted various inclusion criteria ranging from 1 to 5 days [8-10].

From a research perspective, the fewer days required could help relieve the study burden on the participant and better support the use of study resources. Past reviews have suggested 4 days would suffice to achieve sufficient reliability [2,11]. Furthermore, a more recent comprehensive study using a large Singaporean sample recommended that at least 3 and 5 measurement days of step count data were needed to predict weekly and monthly time windows, respectively [12]. Despite the existing recommendations, providing a consensus for the minimum number of days required to predict habitual physical activity can be challenging and fraught with error if the uniqueness of the sample in question is not considered. Human behaviors on a day-to-day basis can be highly variable; thus, the number of days required to reliably predict habitual physical activity is dependent on the variability that exists within an individual [13]. Intuitively, individuals with high consistency or limited variability in physical activity may be predicted with fewer days than an individual with low consistency and high variability. As a consequence, the identification of an optimal number of days could be problematic. Another source of consideration relates to the formula used to calculate reliability. In the physical activity literature, the intraclass correlation coefficient (ICC) is commonly used to grade the reliability of physical activity and a common standard used to indicate acceptable reliability is an ICC of 0.8 [13,14]. However, broadly, this formula approximates the within-subject variability by comparing its magnitude to the between-subject variability. Thus, an increase in diversity between individuals will inherently reduce the reliability score without any change in variability within individuals. This situation poses a potential concern that predictions from large heterogenic samples may not translate to groups of individuals with specific physical activity patterns.

Therefore, the main aim of this study was to identify clusters of individuals with distinct physical activity patterns and to determine if the number of days of accelerometry data required to reliably estimate short- (7 days) and medium-term (28 days) physical activity differed between each unique cluster. This study used clustering analysis to allocate individuals from a heterogenic population into distinct physical activity clusters. Subsequently, reliability estimates were derived for each cluster and compared. The hypothesis was that the number of days to reliably predict (1) short- and (2) medium-term physical activity differs between clusters.

Methods

Study Design

This is a secondary, longitudinal analysis using combined data from 2 independent studies. Only the objectively measured physical activity data (daily step counts) and demographic descriptors were used from each dataset. All other variables from the original studies were excluded from the current analysis. The analysis was conducted at Reykjavik University in Iceland as part of the European Union Horizon 2020-funded Sleep Revolution research project [15].

Setting

Study 1 was an observational study conducted at the Reykjavik University Sleep Institute, aimed at exploring the associations between physical activity and markers of obstructive sleep apnea (OSA) severity [16]. A total of 66 adults were recruited, representing a broad range of sleepers, including healthy individuals, snorers, and those with suspected or diagnosed OSA. Each participant underwent a 3-night, self-applied somnography study at home, along with assessments of anthropometry, body composition, and both subjective and objective physical activity. Importantly, for the purpose of this analysis, objective physical activity was measured using a smartwatch device (Withings Scanwatch, Withings Health Solutions) worn on the nondominant wrist for a 3-month period.

Study 2 was a 12-week, 3-arm randomized controlled trial conducted at the Reykjavik University Sleep Institute [17]. The study was registered in ISRCTN 16974764. It aimed to evaluate

XSL•FO
RenderX

the effects of an exercise program and a lifestyle app on OSA, physical health, and quality of life in adults with mild-to-moderate OSA or habitual snoring. A total of 192 eligible participants were randomized into exercise, app, or control groups. Randomization was stratified using an algorithm to ensure balance across age, gender, BMI, and apnea-hypopnea index (AHI). Participants completed baseline assessments, including a one-night type 3 sleep study, body composition, and physical activity measures. The exercise group attended structured sessions 3 times per week, and the app group engaged in daily behavioral tasks via a health app (Sidekick Health, Reykjavík). Objective physical activity was assessed via a smartwatch device (Withings Scanwatch, Withings Health Solutions) worn on the nondominant wrist over the 12-week study period.

Participants

Participants from each study, who are described in the subsequent paragraphs, were pooled to explore habitual physical activity patterns. Study 1 recruited a heterogeneous sample from the general adult population (≥18 years) with the goal of reflecting natural variation in age, sex, and body morphology. Participants were not excluded based on sleep health status. Study 2 was a randomized controlled trial targeting adults aged 18–50 years of age that were categorized as overweight or obese (BMI ≥25<42 kg/m²), who were physically inactive and had either mild-to-moderate OSA (AHI ≥5 and <30) or habitual snoring (≥10%). Shift workers and individuals undergoing OSA treatment were excluded. Eligible participants underwent a one-night type 3 sleep study to confirm OSA.

Variables

The primary variable analyzed was daily step count, measured objectively using accelerometry-based methodology. Step count served as the key indicator of physical activity throughout the study period.

Data Measurements

Demographics

Data were collected using a custom-developed questionnaire. REDCap (Research Electronic Data Capture) survey software version 9.3.1 (Vanderbilt University) was used to collect participant responses [18,19].

BMI

Height and weight (digital scale; TANITA MC-780, Tanita Corporation) were measured and used to calculate BMI.

Smartwatch

Participants were instructed to wear a smartwatch (Withings Scanwatch, Withings Health Solutions) for the study period. The smartwatch generated physiological and physical activity parameters that included step count, distance covered, elevation, sleep, heart rate, oxygen saturation, passive calories, and active calories. Due to the substantial correlations between the physical activity parameters, step count was selected as the parameter of primary interest. The generated data were available in two forms: (1) timestamped data and (2) daily aggregated data. Additional details of the smartwatch are described by [20].

Bias

Several potential sources of bias were considered in the study design and analysis. First, selection bias was likely present in Study 2, as participants were specifically recruited based on specific inclusion criteria around OSA. However, Study 1 reflected a broader segment of the general adult population and may have helped offset this limitation by introducing more heterogeneity in age, sex, BMI, and physical activity profiles. Nonetheless, the combined sample remains somewhat weighted toward individuals with OSA. Measurement bias was addressed by using objective step count data from smartwatches rather than self-reported physical activity, reducing recall and reporting biases. However, inaccuracies could arise from nonwear time or device limitations; to mitigate this, systematic methods were used to exclude days with insufficient wear time [2,21]. Similarly, to address potential bias from missing days, nonconsecutive valid days were allowed but only within specified windows described in more detail in the Sample Size section.

Sample Size

All participants from Study 1 and Study 2 were considered eligible for analysis, but their inclusion depended on passing specific wear-time criteria.

Valid Day

Given the longitudinal nature of the study, participants were not required to report nonwear time, which influenced the accuracy of step count data. Therefore, a systematic approach to identify nonwear periods was established. Gaps in the timestamped “calories earned” data exceeding 60 minutes were classified as nonwear periods. Summing the nonwear periods that occurred between consecutive sleep periods permitted the calculation of wear time during waking hours. In line with previous approaches and retaining statistical power [2,21], days with less than 10 hours of wear time during waking hours were considered insufficient to accurately estimate step count and excluded.

Valid Period

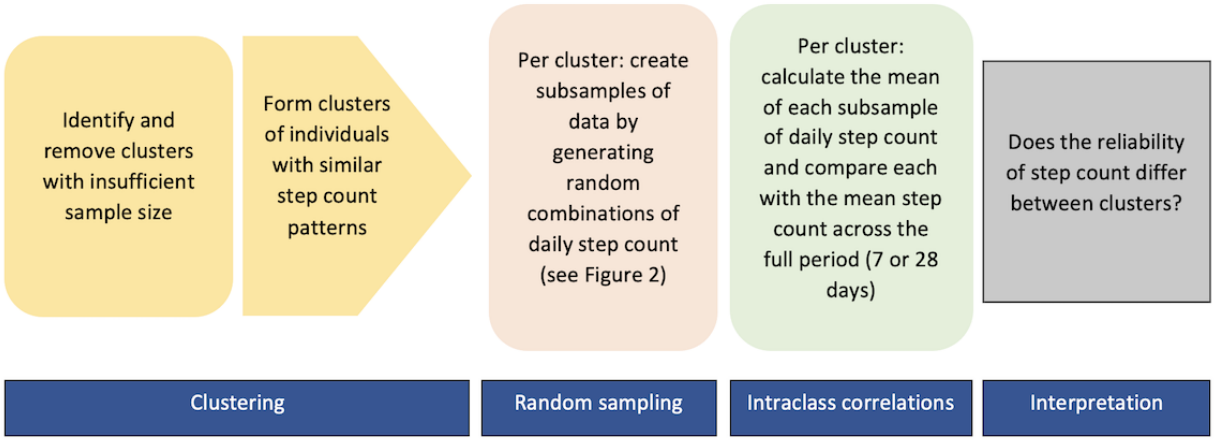
Using the step count data filtered for valid days, one period of 7 days and another of 28 days were selected for each participant in the short- and medium-term analyses, respectively. Periods with nonconsecutive days were allowed if there were (1) 7 valid days within a 10-day period in the short-term analysis or (2) 28 valid days within a 40-day period in the medium-term analysis. Periods with full consecutive sequences or with the fewest days skipped were preferentially chosen, but if more than one of these sequences existed, then the selection was randomized. Nonconsecutive days were permitted due to the difficulty of achieving full consecutive sequences of data, particularly when considering the medium-term analysis. For the inspection of data, the chosen periods of 10 and 40 days were chosen to allow us to retain a sufficient sample size without compromising the representation of physical activity for the given period.

XSL•FO
RenderX

Data Analysis of Quantitative Variables

The methodological approach to the analysis of the daily step count variable is displayed in a visual format in Figure 1. Using unsupervised machine learning, we identified clusters of individuals that shared more similar physical activity patterns than those in the other identified clusters. Agglomerative hierarchical clustering with complete-linkage, a distance-based algorithm, was used to identify clusters [22]. This form of clustering was chosen because it can be used to identify outliers and does not require a predetermined number of clusters. The clustering analysis was conducted for both short- and medium-term physical activity patterns, separately. The clustering analysis first involved running the algorithm to identify sufficiently sized clusters. Given that the calculated reliability metrics are influenced by between-subject variability, smaller clusters may be more prone to an underestimation or overestimation of reliability depending on participant similarities. Thus, if any clusters with less than 12 participants were identified, they were removed, and the remaining clusters were retained for the full analysis. The clusters were determined based on their similarities across 4 dimensions, those being the scaled and centered mean, SD, skewness, and kurtosis of the step count data. These dimensions represent the 4 moments of distribution, which quantitatively reflect each participant's physical activity patterns in the context of an average, day-to-day variability and range of values.

Figure 1. Methodological approach to study analysis.



Statistical Analysis

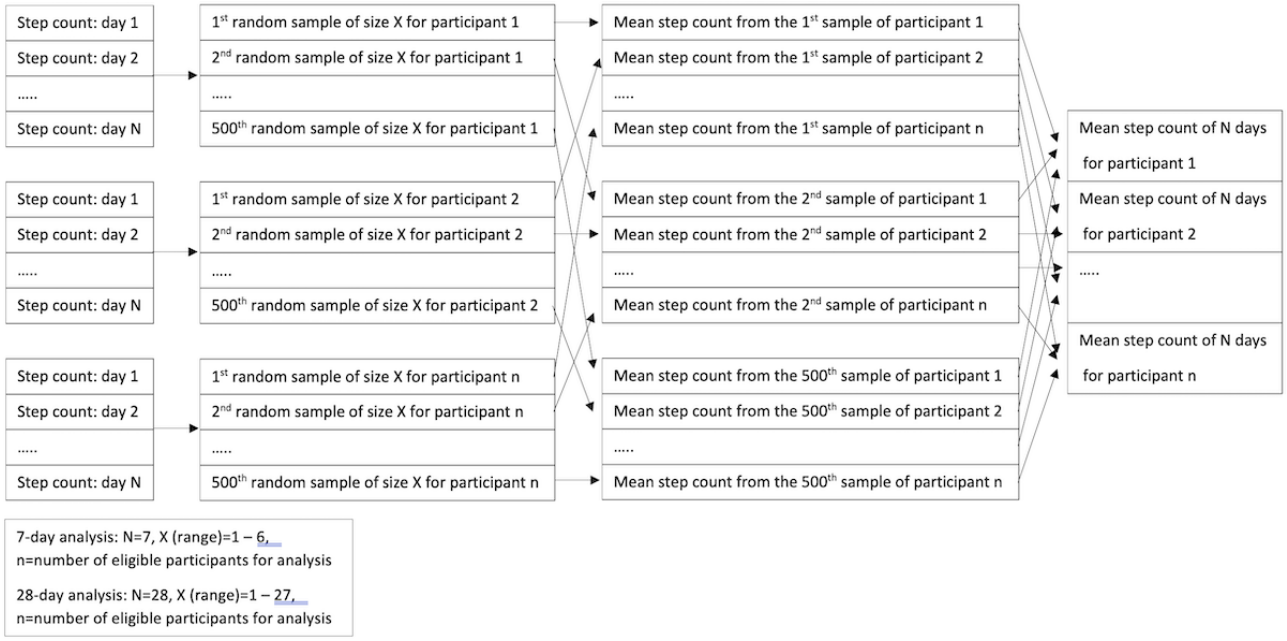
Following the identification of clusters, the reliability of step count within each cluster was determined, along with a baseline comparison (unclustered data). ICCs were derived using linear mixed models with one-way random effects. A series of ICCs were computed by separately comparing the average step count across the full period (7 and 28 days for the short- and medium-term analysis, respectively) to a series of averaged subsamples (ranging from 1 to 6 days and 1 to 27 days for short- and medium-term analysis, respectively). As the subsamples could include any combination of days from the full period, for each participant, resampling methods without replacement were used to ensure an appropriate number of combinations were tested. For each subsample, 500 random combinations were generated and compared to the average step count across the full period (see Figure 2). Prior testing of the data showed that when 500 combinations were used for ICC predictions, the results remained stable when repeated. Thus, the range of ICC values possible for each subsample comparison was produced. In line with previous research, a threshold value of 0.80 was used to infer "acceptable" reliability [13,14].

It is necessary to point out the benefits of the Monte Carlo resampling approach used. This approach reduces the influence of nonnormality and heterogeneity of within-subject variances by using aggregated participant-level means and repeated ICC estimations. Given that ICCs were computed from mean step counts, assumptions of normality and homogeneity of variance are less probable, as justified by the Central Limit Theorem. Additionally, using 500 iterations per subset further enhances the robustness of the reliability estimates by reducing the influence of a typical combinations.

All the analyses were conducted in R (version 4.3.2; R Core Team). Specifically, the "ICC" package was used to extract the ICCs.

XSL•FO
RenderX

Figure 2. Data randomization of subsamples.



Ethical Considerations

Both studies received ethical approval from the National Bioethics Committee of Iceland and the Icelandic Data Protection Agency (21-170; 22-082). Participants provided written informed consent to participate, which also extended consent to secondary analysis. All data were stored anonymously and processed on a secure high-performance computing cluster to ensure efficient handling while maintaining data security and integrity. Finally, participants did not receive any compensation for their participation.

Results

Participants and Descriptive Data

Of the 258 original participants available from Study 1 and Study 2, a total of 149 participants possessed 7 valid days (>10 hour wear time during wake) across a 10-day period and were included in the short-term analysis, and 64 participants possessed 28 valid days across a 40-day period and were included in the medium-term analysis.

In the short-term analysis, 76 of the participants were male, and 73 were female. On average, the participants were 41 (SD 5) years of age and were categorized in the obese category (BMI: mean 31, SD 4). Age ranged from 19 to 76 years old, while BMI ranged from 19 to 42. In the medium-term analysis, half of the participants were male (n=37) and half of the participants were female (n=37). On average, the participants were 45 (SD 9) years of age and were categorized in the obese category (BMI: mean 31, SD 5). Age ranged from 23 to 65 years old, while BMI ranged from 20 to 41. The participants that were eligible for both analyses did not have any missing data.

Outcome Data

Clustering in Short-Term Analysis

Following agglomerative hierarchical clustering, 4 clusters of sufficient size (n≥12) were identified from a visual inspection of the dendrogram. The clusters were chosen by attempting to achieve a balance between appropriate sample sizes and within-cluster cohesion. The 4 dimensions of daily step count used to form each cluster are reported in Table 1.

In comparison to the other clusters, Cluster 1 displayed moderate physical activity levels and day-to-day variability. However, Cluster 1 had a highly skewed physical activity pattern and high kurtosis, suggesting these participants tended to have days that, when compared to their habitual physical activity levels, were considered extremely low or high. In contrast, the other clusters had kurtosis values below or slightly above 3, suggesting the number of days classed as outliers (extreme high or low physical activity) was similar to or less than that seen in a normal distribution pattern. Cluster 2 displayed the lowest relative variability in physical activity levels, while participants in Cluster 3 had the lowest overall physical activity levels, with habitual levels of 4391 (SD 1058) steps per day. Cluster 4 displayed the highest physical activity levels (mean 7179, SD 1028 steps per day), but also the highest day-to-day variability. Given the low kurtosis and moderate skewness, this would suggest the participants in Cluster 4 displayed a mixture of physical activity days.

XSL•FO
RenderX

Table 1. Dimensions of daily step count used to isolate clusters of individuals in the short-term analysis.

Short-term Cluster (n)	Mean steps, mean (SD)	SD steps, mean (SD)	Skewness, mean (SD)	Kurtosis, mean (SD)
1 (n=36)	5883 (1298)	2192 (727)	1.28 (0.48)	5.34 (1.08)
2 (n=22)	5980 (1538)	1619 (685)	0.82 (0.26)	3.67 (0.99)
3 (n=42)	4391 (1058)	1467 (586)	0.50 (0.35)	2.20 (0.83)
4 (n=30)	7179 (1028)	3068 (1132)	0.55 (0.37)	1.96 (0.70)

Clustering in Medium-Term Analysis

Following agglomerative hierarchical clustering, three clusters of sufficient size (n≥12) were retained by visually inspecting the dendrogram. The four dimensions of daily step count used to form each cluster are reported in Table 2.

Compared to the short-term analysis, the clusters within the medium-term analysis displayed less diversity. Modest day-to-day variability in physical activity was found in Cluster 1, but the participants in this cluster had the highest physical activity levels (mean 7671, SD 1052 steps per day). In contrast, participants in Cluster 2 displayed the lowest levels of physical activity (mean 4746, SD 947 steps per day) with the highest relative variability in physical activity. Cluster 3 displayed the least relative variability in physical activity and a low degree of skewness. This suggests that the physical activity levels for these participants were usually around the mean alongside an expected number of higher and lower physical activity days (normal distribution). The number of extremes in physical activity was small in all clusters and did not display distinctly different tails to a normal distribution pattern.

Table 2. Dimensions of daily step count used to isolate clusters of individuals in the medium-term analysis.

Medium-term Cluster (n)	Mean steps, mean (SD)	SD steps, mean (SD)	Skewness, mean (SD)	Kurtosis, mean (SD)
1 (n=23)	7671 (1052)	2955 (655)	0.66 (0.45)	3.65 (1.11)
2 (n=14)	4746 (947)	2265 (381)	0.87 (0.24)	3.31 (0.58)
3 (n=16)	5775 (988)	1764 (398)	0.31 (0.28)	2.87 (0.72)

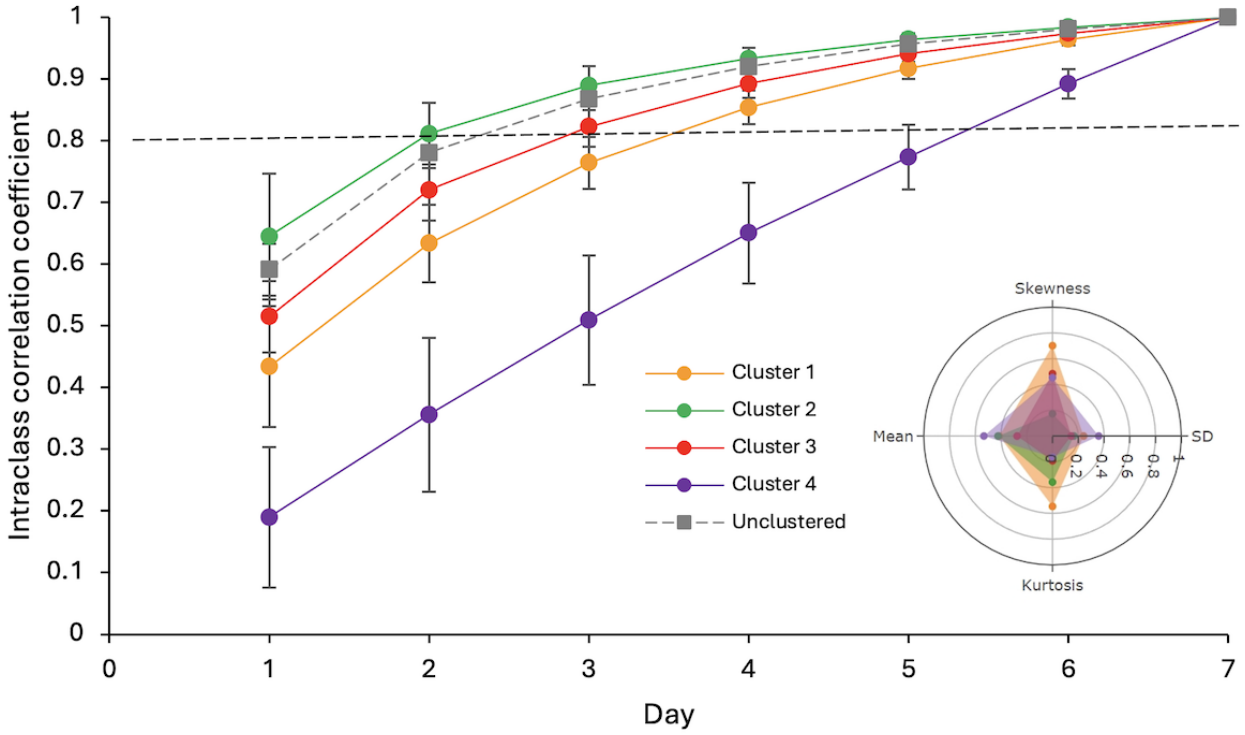
Main Results

Intraclass Correlations in Short-Term Analysis

To achieve a mean ICC score above 0.80, calculated using all randomized combinations for each day subset, Cluster 2 required 2 days, Cluster 3 required 3 days, Cluster 1 required 4 days, and Cluster 4 required 6 days (Figure 3). The minimum number of days needed to achieve an ICC score above 0.80 in all 500 random combinations was 4 in Clusters 2 and 3, 5 in Cluster 1, and 7 in Cluster 4. As a baseline comparison, when the data were unclustered, 3 days were required to achieve a mean ICC score above 0.80 in all 500 random combinations.

XSL•FO
RenderX

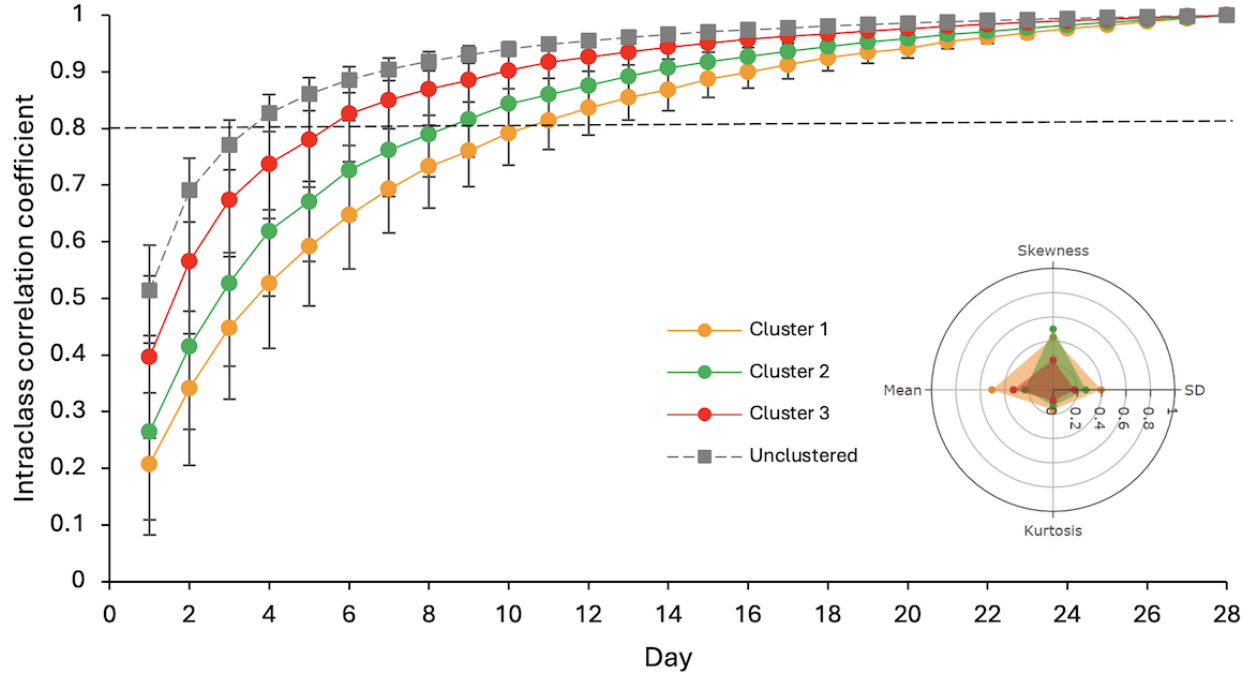
Figure 3. Range of intraclass correlation coefficients across all randomized combinations for each cluster from short-term analysis. Scores ≥ 0.80 are classed as "acceptable" agreement. Data are mean (SD).



Intraclass Correlations in Medium-Term Analysis

In order to achieve a mean ICC score greater than 0.80, calculated using all randomized combinations for each day subset, Cluster 3 required 6 days, Cluster 2 required 9 days, and Cluster 1 required 11 days (Figure 4). The minimum number of days needed to achieve an ICC score above 0.80 in all 500 random combinations was 11 in Cluster 3, 18 in Cluster 2, and 17 in Cluster 1. As a baseline comparison, when the data were unclustered, 4 days were required to achieve a mean ICC score above 0.80 in all 500 random combinations.

Figure 4. Range of intraclass correlation coefficients across all randomized combinations for each cluster from medium-term analysis. Scores ≥ 0.80 are classed as "acceptable" agreement. Data are mean (SD).



XSL•FO
RenderX

Discussion

Principal Findings

The main aim of this study was to identify clusters of individuals with distinct physical activity patterns and to determine if the number of days of accelerometry data required to reliably estimate short- (7 days) and medium-term (28 days) physical activity differed between each unique cluster. Clustering analysis confirmed that, in a heterogenic sample, clusters of individuals could be identified based on their physical activity patterns. In agreement with our hypothesis, the physical activity patterns of different individuals were highly relevant to the number of days needed to reliably predict both short- (7 days) and medium-term (28 days) physical activity, ranging from 2 to 6 days and 6 to 11 days, respectively, by cluster. Therefore, the minimum number of days should be assessed within a particular study cohort, instead of relying on generic recommendations.

One major consideration of using accelerometry for longitudinal physical activity research is that participant monitoring is a continuous process, which can require significant study resources and can challenge participant adherence. Therefore, having the flexibility to monitor less or lose more days, without compromising the validity of the data, is of significant benefit. However, the results from the current study show that, while this is possible, it can be erroneous to use a generic number of days across all participants. This supports the need to be cautious when considering a one-size-fits-all approach when it comes to collecting accelerometry data. For example, past reviews have suggested 4 days would suffice to achieve sufficient reliability in physical activity [2,11]. However, based on the results from the current study, this will be highly dependent on the physical activity patterns that exist within the specific research sample being studied. Another potential limitation of previous recommendations is the use of the Spearman-Brown prophecy formula to estimate reliability [23-25]. This statistical method depends on achieving homogeneity of variances, which may be unrealistic to achieve in many datasets. The resampling approach in this study helped address this problem, as it does not rely on this formula and is less sensitive to violations of homogeneity of variance. By repeatedly calculating ICCs from randomly selected combinations of days, the method reduces the influence of unequal within-subject variances on individual ICC estimates, resulting in a more robust overall estimate of reliability. Similarly, a nonclustered ICC may violate the assumption of homogeneity of variance if the physical activity patterns of the sample are diverse. Clustering can help group individuals with similar within-subject variance, improving the likelihood that model assumptions are satisfied.

The results from both the short- and medium-term analyses highlighted that clusters with relatively low variability in their physical activity required fewer days to reliably estimate physical activity. In contrast, clusters with high variability required significantly more days. Referring specifically to the data, depending on the cluster, the mean number of days required ranged from 2 to 6 for the short-term analysis and 6 to 11 for the medium-term analysis. This makes generalized recommendations as reported in previous research [26-28] difficult to provide. Interestingly, within this study's sample, the physical activity patterns of clusters in the medium-term analysis were more similar to each other than those observed in the short-term analysis. This implies reduced variability as participants settle into comparable daily activity patterns with larger measurement windows. Nonetheless, it is important to mention that, for one cluster, the minimum number of days needed to achieve sufficient reliability in all 500 random combinations was still noteworthy (17 days).

To an extent, these results agree with the work by [12] that recommended at least 3 and 5 measurements days of step count data were needed to predict weekly and monthly time windows, respectively. However, it highlights the risk of adopting the minimum limit of these recommendations for all participants. Instead, an approach that first attempts to understand the sample that is being studied could be beneficial. For example, if pilot testing is being conducted, this could provide an opportunity to collect prior accelerometer data on the group of participants. Depending on whether the physical activity patterns are expected to be similar or not, a cluster analysis could be used to determine the minimum number of days needed for each subset of participants. Alternatively, while physical activity questionnaires may provide an inexpensive and less-demanding method to group individuals, they can be prone to recall bias, as we have previously shown [16].

Practically, this research demonstrated that a heterogeneous sample can be divided into subgroups with distinct physical activity patterns. As a consequence, these clusters can influence the reliability of predicting habitual physical activity and, thus, the broader interpretation of the results. Previous research linking physical activity to health outcomes has estimated a participant's habitual level of physical activity using subsamples ranging from 1 to 5 days [8-10]. Unfortunately, the validity of such results is tied to the accuracy of this physical activity estimation. Researchers who do not fully consider the potential heterogeneity in their sample may be at potential risk of underestimating within-subject variability, which could potentially influence their interpretation of the results. Therefore, safeguards should be considered to improve physical activity estimates that include clustering, using appropriate ICC methods, and including complementary reliability metrics (standard error of measurement, coefficient of variation, etc).

Another point of consideration is that, in research studies, accelerometry is often used to quantify the rest-activity cycle of participants, which spans physical activity, sedentary time, and sleep. While this research focused on physical activity, the reliability of sedentary time and sleep will also display different levels of within-subject variability. Therefore, if researchers intend to quantify all 3 components of the rest-activity cycle, it is important to consider whether the number of days chosen is sufficient to capture the variability across all 3 components. For example, before individual sleep behaviors become stable, they require many more days of data than the general recommendation of 14 days [3]. However, more research exploring clustered samples is required because results from grouped data have reported differing results [29-31], which is likely a reflection of the level of heterogeneity across the sample in each analysis.

XSL•FO
RenderX

The current study is not without its limitations. First, the specific results of the current study reflect the discrete physical activity patterns of the analyzed sample. Nonetheless, the core concepts are universal and generalizable to other research samples in this area. With a larger sample size, more clusters may have been identified, or rather, a greater number of participants may have been placed within each identified cluster. For example, a number of clusters that were identified had insufficient sample sizes to conduct the analysis on and thus were excluded. It is plausible that the individuals with these excluded physical activity patterns may have required even more or fewer days to reliably estimate their physical activity over time, compared to those seen in the reported results. Importantly, while the physical activity patterns of the excluded individuals were uncommon in this study, they may better reflect physical activity patterns seen in a different study. Notably, while our sample was relatively heterogeneous, there was a higher percentage of individuals with large BMIs, which may have contributed to a small bias, as BMI has been shown to be associated with lower physical activity patterns [32]. Nonetheless, if the sample in this study had more variance, it likely would have strengthened the results while also improving the study's generalizability. In the current study, 10 or more hours of wear time during waking hours for a given day was considered sufficient for inclusion. Therefore, certain days with missing physical activity data will inherently be included and can influence reliability predictions [24]. Nonetheless, this decision was made to align with previous approaches [2,21], and to balance accuracy with the retention of information. Finally, ICCs are based on the proportion of between-to within-subject variance; thus, results from separate samples can differ based on within-subject variance (day-to-day variability of steps within a participant) but also, between-subject variance (variability of steps between participants). As alluded to in the introduction, this poses a potential concern that recommendations from large heterogeneous samples may misrepresent research samples that are more homogenous in nature. Unfortunately, while widely used, this is the limitation of using ICCs in the context of estimating physical activity. Importantly, our clustering analysis attempted to limit the between-subject variance by identifying clusters of individuals with similar physical activity patterns based on the 4 moments of distribution.

Conclusion

In conclusion, the number of days required to reliably estimate physical activity differs between clusters of individuals characterized by distinct physical activity patterns across both the short- and medium-term. To avoid unreliable estimates of physical activity, researchers should be mindful of the sample they are studying and how this may influence the minimum number of days required to reliably reflect physical activity. This study showcases that one-size, indeed, does not fit all when it comes to collecting accelerometry data.

Acknowledgments

The authors are grateful to all the participants for their involvement. The authors would like to thank the Sports Science Department for providing the Facility of Social Sciences at Reykjavík University, Reykjavík (Iceland) to carry out Study 2. We would also like to thank Birta Sóley, Steinunn Svansdóttir, Karitas Kjartansdóttir, and Katrín Gústafsdóttir for their work regarding recruiting participants, the preliminary Study 1 and Study 2. Also, the authors would like to thank Kristín Anna Ólafsdóttir for her extensive contribution to ethical applications and personnel management. We used the generative AI tool ChatGPT by OpenAI to support with code generation, which was reviewed and corrected by the study group, and to restructure writing.

Funding

The authors declare that financial support was received for the research, authorship, and/or publication of this article. The authors of this paper have received funding to perform this work and write this paper from the European Union's Horizon 2020 research and innovation program (grant agreement 965417). The Sleep Revolution project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 965417.

Authors' Contributions

Conceptualization was done by CJM, GMJ, JMS, MO, and ESA. Data curation and formal analysis were performed by CJM and GMJ. Funding acquisition was secured by JMS, ESA. Investigation was carried out by CJM and KYF. Methodology was developed by CJM, GMJ, JMS, MO, and ESA. Project administration was managed by CJM. Resources were provided by ASI. Software development was completed by CJM and GMJ. Supervision was provided by ASI, JMS, MO, and ESA. Validation was conducted by CJM and GMJ. Visualization was done by CJM and GMJ. Writing of the original draft was performed by CJM, while review and editing were carried out by CJM, GMJ, KYF, ASI, JMS, MO, and ESA.

Conflicts of Interest

The Sleep Revolution pays KYF PhD students' salaries and CJM, GMJ postdoctoral researchers' salaries. ESA reports honoraria from Nox Medical, ResMed, Jazz Pharmaceuticals, Linde Healthcare, Wink Sleep, Apnimed, and Vistor. A former member of Philips Sleep Medicine, Innovation Medical Advisory Board, and Lilly Medical Advisory Board. As well as currently being the President of the European Sleep Research Society. None for other authors.

XSL•FO
RenderX

References

1. van Poppel MNM, Chinapaw MJM, Mokkink LB, van Mechelen W, Terwee CB. Physical activity questionnaires for adults: a systematic review of measurement properties. Sports Med 2010;40(7):565-600. [doi: 10.2165/11531930-000000000-00000] [Medline: 20545381]
2. Migueles JH, Cadenas-Sanchez C, Ekelund U, Delisle Nyström C, Mora-Gonzalez J, Löf M, et al. Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. Sports Med 2017;47(9):1821-1845 [FREE Full text] [doi: 10.1007/s40279-017-0716-0] [Medline: 28303543]
3. Óskarsdóttir M, Islind AS, August E, Arnardóttir ES, Patou F, Maier AM. Importance of getting enough sleep and daily activity data to assess variability: longitudinal observational study. JMIR Form Res 2022;6(2):e31807 [FREE Full text] [doi: 10.2196/31807] [Medline: 35191850]
4. Doherty A, Jackson D, Hammerla N, Plötz T, Olivier P, Granat MH, et al. Large scale population assessment of physical activity using wrist worn accelerometers: the UK biobank study. PLoS One 2017;12(2):e0169649 [FREE Full text] [doi: 10.1371/journal.pone.0169649] [Medline: 28146576]
5. Dempsey PC, Strain T, Khaw K, Wareham NJ, Brage S, Wijndaele K. Prospective associations of accelerometer-measured physical activity and sedentary time with incident cardiovascular disease, cancer, and all-cause mortality. Circulation 2020;141(13):1113-1115 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.119.043030] [Medline: 32223676]
6. Raichlen DA, Aslan DH, Sayre MK, Bharadwaj PK, Ally M, Maltagliati S, et al. Sedentary behavior and incident dementia among older adults. JAMA 2023;330(10):934-940 [FREE Full text] [doi: 10.1001/jama.2023.15231] [Medline: 37698563]
7. Cao Z, Min J, Chen H, Hou Y, Yang H, Si K, et al. Accelerometer-derived physical activity and mortality in individuals with type 2 diabetes. Nat Commun 2024;15(1):5164 [FREE Full text] [doi: 10.1038/s41467-024-49542-0] [Medline: 38886353]
8. Tudor-Locke C, Brashear MM, Johnson WD, Katzmarzyk PT. Accelerometer profiles of physical activity and inactivity in normal weight, overweight, and obese U.S. men and women. Int J Behav Nutr Phys Act 2010;7:60 [FREE Full text] [doi: 10.1186/1479-5868-7-60] [Medline: 20682057]
9. Shim J, Fleisch E, Barata F. Wearable-based accelerometer activity profile as digital biomarker of inflammation, biological age, and mortality using hierarchical clustering analysis in NHANES 2011-2014. Sci Rep 2023;13(1):9326 [FREE Full text] [doi: 10.1038/s41598-023-36062-y] [Medline: 37291134]
10. Sakal C, Li T, Li J, Li X. Predicting poor performance on cognitive tests among older adults using wearable device data and machine learning: a feasibility study. NPJ Aging 2024;10(1):56 [FREE Full text] [doi: 10.1038/s41514-024-00177-x] [Medline: 39587119]
11. Trost SG, McIver KL, Pate RR. Conducting accelerometer-based activity assessments in field-based research. Med Sci Sports Exerc 2005;37(11 Suppl):S531-S543. [doi: 10.1249/01.mss.0000185657.86065.98] [Medline: 16294116]
12. Yao J, Tan CS, Lim N, Tan J, Chen C, Müller-Riemenschneider F. Number of daily measurements needed to estimate habitual step count levels using wrist-worn trackers and smartphones in 212,048 adults. Sci Rep 2021;11(1):9633 [FREE Full text] [doi: 10.1038/s41598-021-89141-3] [Medline: 33953288]
13. Baranowski T, Mâsse LC, Ragan B, Welk G. How many days was that? We're still not sure, but we're asking the question better!. Med Sci Sports Exerc 2008;40(7 Suppl):S544-S549 [FREE Full text] [doi: 10.1249/MSS.0b013e31817c6651] [Medline: 18562972]
14. Sainani K. Reliability statistics. PM R 2017;9(6):622-628. [doi: 10.1016/j.pmrj.2017.05.001] [Medline: 28602174]
15. Arnardottir ES, Islind AS, Óskarsdóttir M, Ólafsdóttir KA, August E, Jónasdóttir L, Sleep Revolution. The sleep revolution project: the concept and objectives. J Sleep Res 2022;31(4):e13630. [doi: 10.1111/jsr.13630] [Medline: 35770626]
16. Fridgeirsdottir KY, Ólafsdóttir KA, Islind AS, Leppänen T, Arnardottir ES, Saavedra JM. The role of physical activity on obstructive sleep apnea severity and hypoxic load, and the mismatch between subjective and objective physical activity assessments. J Sleep Res 2024;33(6):e14195. [doi: 10.1111/jsr.14195] [Medline: 38480993]
17. Fridgeirsdottir KY, Murphy CJ, Islind AS, Árnadóttir BS, Hrubos-Strøm H, Arnardottir ES, et al. Effects of exercise and a lifestyle app on sleep disordered breathing, physical health and quality of life. ERJ Open Res 2025;11(3). [doi: 10.1183/23120541.01134-2024] [Medline: 40551796]
18. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
19. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
20. Biedebach L, Óskarsdóttir M, Arnardottir E, Islind A. Two sides of the same pillow: unfolding the relationship between objective and subjective sleep quality with unsupervised learning. 2023 Presented at: Forty-Fourth International Conference on Information Systems; 2023 December 10-13; Hyderabad, India URL: https://aisel.aisnet.org/icis2023/ishealthcare/ishealthcare/20/?utm_source=chatgpt.com
21. Atkin AJ, Gorely T, Clemes SA, Yates T, Edwardson C, Brage S, et al. Methods of measurement in epidemiology: sedentary behaviour. Int J Epidemiol 2012;41(5):1460-1471. [doi: 10.1093/ije/dys118] [Medline: 23045206]

XSL•FO
RenderX

22. Altman N, Krzywinski M. Clustering. Nat Methods 2017;14(6):545-546. [doi: 10.1038/nmeth.4299]
23. Prescott S, Traynor JP, Shilliday I, Zanotto T, Rush R, Mercer TH. Minimum accelerometer wear-time for reliable estimates of physical activity and sedentary behaviour of people receiving haemodialysis. BMC Nephrol 2020;21(1):230 [FREE Full text] [doi: 10.1186/s12882-020-01877-8] [Medline: 32546225]
24. Aadland E, Ylvisåker E. Reliability of objectively measured sedentary time and physical activity in adults. PLoS One 2015;10(7):e0133296 [FREE Full text] [doi: 10.1371/journal.pone.0133296] [Medline: 26192184]
25. Gretebeck RJ, Montoye HJ. Variability of some objective measures of physical activity. Med Sci Sports Exerc 1992;24(10):1167-1172. [Medline: 1435166]
26. O'Brien CM, Kitas GD, Rayner F, Isaacs JD, Baker KF, Pratt AG, BIOFLARE Consortium. Number of days required to measure sedentary time and physical activity using accelerometery in rheumatoid arthritis: a reliability study. Rheumatol Int 2023;43(8):1459-1465 [FREE Full text] [doi: 10.1007/s00296-023-05342-1] [Medline: 37227468]
27. Kocherginsky M, Huisingh-Scheetz M, Dale W, Lauderdale DS, Waite L. Measuring physical activity with hip accelerometry among U.S. older adults: how many days are enough? PLoS One 2017;12(1):e0170082 [FREE Full text] [doi: 10.1371/journal.pone.0170082] [Medline: 28081249]
28. Matthews CE, Ainsworth BE, Thompson RW, Bassett DR. Sources of variance in daily physical activity levels as measured by an accelerometer. Med Sci Sports Exerc 2002;34(8):1376-1381. [doi: 10.1097/00005768-200208000-00021] [Medline: 12165695]
29. Tworoger S, Davis S, Vitiello M, Lentz M, McTiernan A. Factors associated with objective (actigraphic) and subjective sleep quality in young adult women. J Psychosom Res 2005;59(1):11-19. [doi: 10.1016/j.jpsychores.2005.03.008] [Medline: 16126091]
30. Knutson KL, Rathouz PJ, Yan LL, Liu K, Lauderdale DS. Intra-individual daily and yearly variability in actigraphically recorded sleep measures: the CARDIA study. Sleep 2007;30(6):793-796 [FREE Full text] [doi: 10.1093/sleep/30.6.793] [Medline: 17580601]
31. Aili K, Åström-Paulsson S, Stoetzer U, Svartengren M, Hillert L. Reliability of actigraphy and subjective sleep measurements in adults: the design of sleep assessments. J Clin Sleep Med 2017;13(1):39-47 [FREE Full text] [doi: 10.5664/jcsm.6384] [Medline: 27707448]
32. Dwyer T, Ponsonby A, Ukoumunne OC, Pezic A, Venn A, Dunstan D, et al. Association of change in daily step count over five years with insulin sensitivity and adiposity: population based cohort study. BMJ 2011;342:c7249. [doi: 10.1136/bmj.c7249] [Medline: 21233153]

Abbreviations

AHI: apnea-hypopnea index
ICC: intraclass correlation coefficient
OSA: obstructive sleep apnea
REDCap: Research Electronic Data Capture

Edited by L Buis; submitted 15.Jul.2024; peer-reviewed by N Maldaner, C Sakal; comments to author 11.Apr.2025; revised version received 15.Jul.2025; accepted 06.Oct.2025; published 10.Feb.2026.

Please cite as:
Murphy CJ, Jouan GM, Friðgeirsdóttir KY, Islind AS, Saavedra JM, Óskarsdóttir M, Arnardóttir ES
Determining Cluster-Specific Differences in the Number of Days Required to Reliably Predict Habitual Physical Activity: Intraclass Correlation Resampling Analysis
JMIR Mhealth Uhealth 2026;14:e64323
URL: https://mhealth.jmir.org/2026/1/e64323
doi:10.2196/64323
PMID:

©Conor Jordan Murphy, Gabriel M Jouan, Katrin Y Friðgeirsdóttir, Anna Sigridur Islind, Jose M Saavedra, María Óskarsdóttir, Erna Sif Arnardóttir. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 10.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Designing a Self-Guided Digital Intervention for Self-Management of Shoulder Pain in People Living With Spinal Cord Injury: Tutorial on Using a Person-Based Approach

Verna Stavric[1*], PhD; Nicola L Saywell[1*], PhD; Nicola M Kayes[2*], PhD

[1]Department of Physiotherapy, School of Allied Health, Auckland University of Technology, Private Bag 92006, Auckland, New Zealand

[2]Centre for Person Centred Research, Auckland University of Technology, Auckland, New Zealand

[*]all authors contributed equally

Corresponding Author:
Verna Stavric, PhD
Department of Physiotherapy, School of Allied Health, Auckland University of Technology, Private Bag 92006, Auckland, New Zealand

Abstract

Shoulder pain is prevalent in people living with spinal cord injury. Technology and digital rehabilitation tools are increasingly available, but this has not yet included the provision of a self-guided exercise intervention focused on managing shoulder pain for people living with spinal cord injury. We drew on the person-based approach (PBA) to intervention development to design a Shoulder Pain Intervention delivered over the interNet (SPIN) to address this gap. However, in preparation for the design process, we found very few published examples of how the PBA had been operationalized. The aim of this paper is to provide a detailed explanation of our approach and how we operationalized the PBA in the design of SPIN to maximize relevance and engagement. Our design process followed the key PBA steps, combining additional evidence and theoretical components. Each step ensured that guiding principles were formulated and followed to maximize the probability that SPIN would be fit for purpose. We followed 3 steps: (1) we drew on themes from preparatory research (existing and primary) to identify the key behavioral issues, needs and challenges, and existing features to form the basis of SPIN design; (2) we formatted guiding principles that included articulating specific design objectives to provide a framework to identify system requirements; and (3) we selected and refined intervention features using existing literature, behavioral theory, and tools such as the "Behaviour Change Wheel." We have designed SPIN by incorporating a deep understanding of the users' needs and best available evidence to maximize engagement and positive outcomes. In this paper, we have made clear how we operationalized the PBA phases, including how existing evidence, theory, tools, and methods were leveraged to support the PBA process. In explicating our process, we have provided a blueprint to guide future researchers using this approach.

(JMIR Mhealth Uhealth 2026;14:e66678) doi:10.2196/66678

KEYWORDS

person-based approach; self-guided; intervention design; behavioral analysis; spinal cord injury; shoulder pain; self-management; mHealth; mobile apps; smartphones; digital health; digital interventions

Background

Overview

Shoulder pain is common in wheelchair users living with spinal cord injury (SCI) [1,2]. A lesion to the spinal cord can result in loss of innervation to muscles of the trunk and lower limbs. Consequently, many people living with SCI (pwSCI) rely on their upper extremities not only for performance of daily activities but also for locomotion. Shoulder pain can have a significant impact on their activity, reducing mobility, independence, and quality of life [1-5]. Digital and web-based interventions have increasingly been offered to pwSCI to promote exercise and physical activity [6-9]. These interventions minimize barriers to rehabilitation to address many health concerns, including managing their shoulder pain. Previous authors have found that in the general population, digital or web-based interventions can produce positive effects in various outcomes, such as physical activity [10,11].

Technology-supported exercise interventions for pwSCI with persistent shoulder pain are currently available, but they have some limitations. They either require ongoing input and monitoring from a clinician [12-15] or provide general self-management advice [13] but without enough guidance to allow for clear and structured exercise progression specifically for shoulder pain. Self-guided digital exercise interventions have been successfully implemented for people with knee osteoarthritis [16,17], dizziness [18,19], and breast cancer [20] and may be a viable option for pwSCI. Our recent systematic

XSL•FO
RenderX

review and meta-analysis of self-guided digital physical activity and exercise interventions demonstrated positive effects on physical activity at both short- and longer-term follow-up, in people living with chronic conditions [21]. We also found that interventions that used behavioral strategies and were underpinned by a theoretical framework were more effective. This suggests that self-guided digital interventions have the potential to support pwSCI to manage their shoulder pain, but that the intervention would need to be designed systematically and intentionally.

We have designed Shoulder Pain Intervention delivered over the interNet (SPIN) as a self-guided digital intervention to give pwSCI who experience shoulder pain the ability to access and progress evidence-based exercises. The intervention guides pwSCI to monitor symptoms and improvement [22] to promote autonomy in the management of their condition. The aim of SPIN is to be an engaging program that is responsive to the needs of pwSCI who have shoulder pain.

To achieve this, we were guided by the person-based approach (PBA) in the design of SPIN [23]. The PBA follows 4 iterative phases of intervention development that include (1) planning which seeks a deep understanding of the perspectives and psychosocial context of potential users through iterative qualitative research, (2) design based on guiding principles that have been created from insights from the first phase, (3) development and refinements which are made through iterative user feedback, and (4) trialing to evaluate the effectiveness on outcomes and impact on behavior change to make any necessary adjustments. Due to its focus on the development of digital behavior change interventions, the intent and purpose of PBA align well with adjacent behavior change theory and tools such as the COM-B [24], "Behaviour Change Wheel" [25], and behavioral analysis [26]. Furthermore, the PBA process is sufficiently flexible to enable the use of these (and other) tools to achieve the aims and purpose of a given phase. Integrating behavioral science theory and evidence while keeping users' needs and contexts in focus has been found to maximize engagement and effectiveness of interventions [18,25,27-31]. This tutorial focuses on the first 2 PBA phases of planning and design. See Table 1 for an example of how our study was mapped onto the PBA.

Table . Mapping of person-based approach phases onto Shoulder Pain Intervention delivered over the Internet design.

PBA[a] description	Phase	This study	
		Purpose	Planned outcome
Use of primary and secondary *qualitative evidence* to understand users' behavioral and psychosocial needs and challenges in using the intervention		To determine *factors that need to be included* to *encourage or facilitate engagement* with this self-guided web-based exercise intervention	• A rich description of key *needs, challenges, and facilitators* of engagement in web-based tools and exercise for people living with SCI[b] who experience shoulder pain to *underpin the design phase's guiding principles and features*
Formulation of *key guiding principles* that capture the main intervention objectives as identified in the planning phase and that are continuously referred to throughout the development of the intervention		To design an *evidence-based, self-guided, web-based* intervention Exercise, behavioral support, and self-guided components to be included within the intervention features	• Intervention design objectives • Intervention features • First iteration of SPIN[c] prototype

[a]PBA: person-based approach.

[b]SCI: spinal cord injury.

[c]SPIN: Shoulder Pain Intervention delivered over the interNet.

We drew heavily on the existing PBA literature during the planning and designing stages of SPIN. However, the lack of access to detailed examples of how the PBA has been operationalized in practice made it challenging to translate the principles of this approach into reality. This is not a unique problem. Duncan and colleagues [32] noted that published work on the development of an intervention is frequently sparse because it is often included in the same publication as the reporting of a pilot or feasibility study.

Aim

The aim of this paper is to make clear how the principles of the PBA were operationalized in intervention design and the development of SPIN. We have illustrated our use of the PBA framework by outlining the detailed and explicit steps involved in the translation of the evidence, theory, and person-based recommendations into intervention design. In doing so, we have built on the existing methodological framework and enabled others to draw on this approach in future intervention design and development.

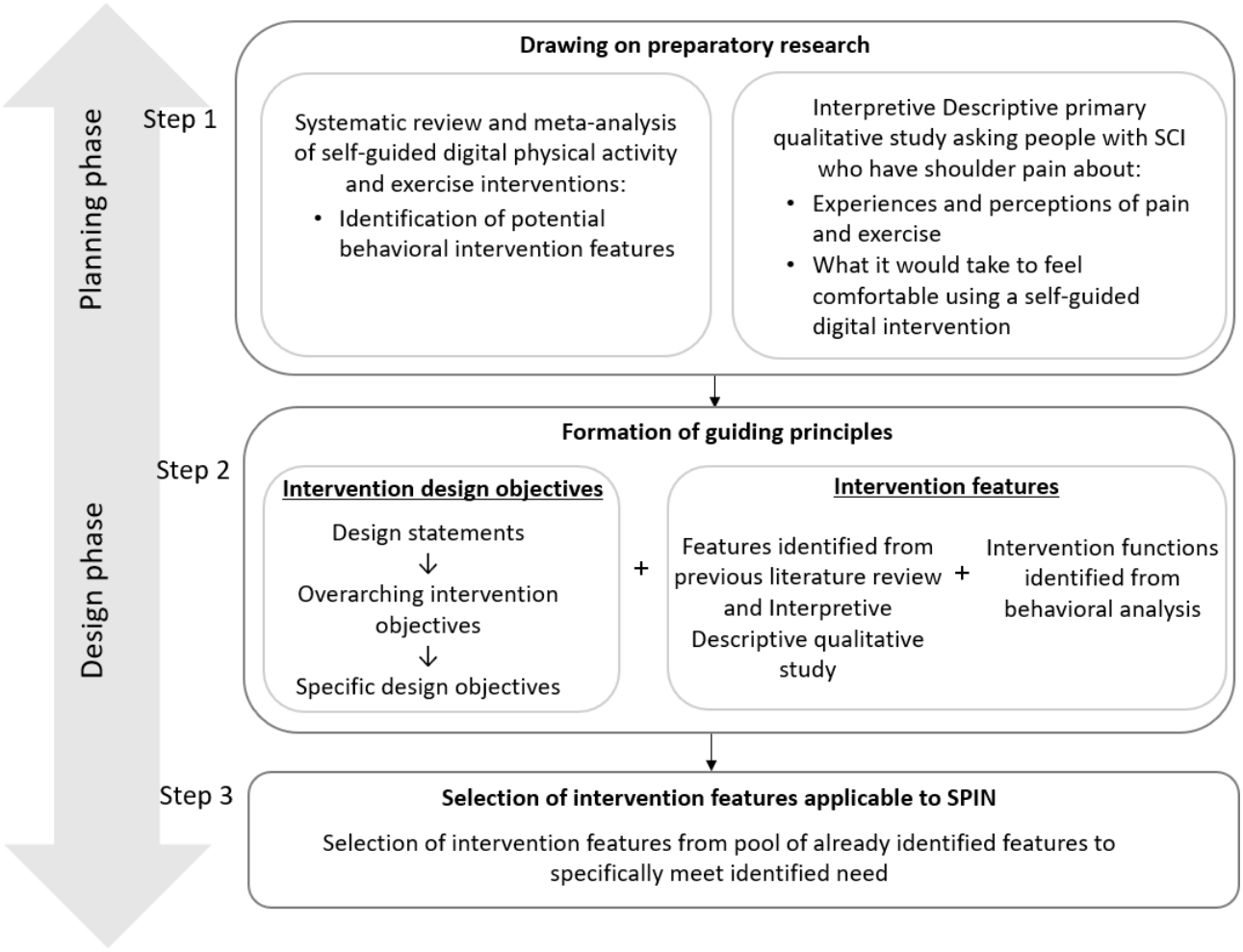
Methods and Outcomes

Overview

The planning and design phases of the PBA are described below, along with an overview of how they were operationalized in the design of SPIN. Figure 1 provides an overview of the SPIN design process and the components involved. Each step and its subsequent outcome have been described in detail in the sections that follow.

XSL•FO
RenderX

Figure 1. Overview of Shoulder Pain Intervention delivered over the Internet design steps and components. SCI: spinal cord injury; SPIN: Shoulder Pain Intervention delivered over the interNet.



Step 1: Drawing on Preparatory Research

Methods

This initial phase of the PBA draws on qualitative research, including interviews and focus groups, to gather in-depth insights into the psychological, social, and emotional factors that influence the users' behavior. The goal is to identify the underlying motivations, beliefs, and barriers that may affect engagement with health interventions [23]. In the context of SPIN, this preparatory research included: (1) a systematic review and meta-analysis investigating the effectiveness of self-guided digital physical activity and exercise interventions [21] and (2) an Interpretive Descriptive qualitative study exploring the perceptions of pwSCI who have shoulder pain, on the use of a self-guided digital intervention to help them manage their shoulder pain [33].

Outcome

The review identified several self-guided digital physical activity and exercise interventions. Data extraction included identifying discrete intervention features and categorizing them using a purpose-built template (Multimedia Appendix 1), based on a synthesis of key literature [27,34-40]. Using this template, we extracted possible behavioral intervention features relating to qualities such as customizability, the provision of instruction, feedback and monitoring, tailoring, reminders and prompts, goals and planning, social support, and rewards and threats. We also noted the success of interventions using features that supported behavior, particularly self-regulation. This informed an initial pool of possible intervention features for SPIN that were reviewed later in Step 2.

The Interpretive Descriptive qualitative study identified themes that represented an evaluative process pwSCI go through when considering using a self-guided digital exercise intervention: *Should I use it?*, whether I believe it will work for me right now; *Can I use it?*, whether I can operate the intervention competently and confidently; and *Will I use it?*, whether it will be responsive to my unique needs and keep me coming back. These formed the basis of the design statements in Step 2.

Conceptual representations of possible behavioral intervention features identified from the review were used as probes and images during data collection in the Interpretive Descriptive qualitative study. These were used to prompt discussion about what could help pwSCI to engage in a self-guided digital intervention. The pwSCI discussed ways in which these concepts and specific features may support them. These perspectives were extracted from the audio recordings and tabulated to support the identification of behavioral intervention features in Step 2.

XSL•FO
RenderX

Step 2: Formation of Guiding Principles

Overview

The guiding principles in the PBA are formulated by synthesizing key insights from the planning phase (Step 1) into intervention design objectives and corresponding intervention features that address users' specific needs, preferences, and behavioral barriers [23]. Yardley and colleagues [23] contend that staying true to the identified needs of the people who will use the intervention, throughout the design process, increases intervention relevance, engagement, and effectiveness. In our design of SPIN, we followed several stages to ensure the key context-specific behavioral needs and challenges identified in the Interpretive Descriptive qualitative study remained the focal point during intervention design.

Intervention Design Objectives

Yardley et al [23] suggest generating intervention design objectives to support the creation of the guiding principles but do not expand on how these may be identified. Below, we describe the method we followed to produce intervention design objectives through the creation of design statements, overarching intervention objectives, and specific intervention design objectives.

Design Statements

Methods

We created design statements by using the 3 themes constructed in the Interpretive Descriptive qualitative study. We first reframed each theme into a design statement, giving consideration to how each could be reflected in the design of the intervention. To do this, we reworded the themes to move from a question (*Should I use it?*) into a design statement (*I should use it if…*) and then added conditions applicable to each design statement. Each condition reflected key elements from the qualitative findings, resulting in person-centered conditions to be met in the design process. This process provided depth and context to inform the design of SPIN and ensured the next step would be underpinned by the perspectives of the future users of the intervention, in this case, pwSCI.

Outcome

The Interpretive Descriptive qualitative study themes, design statements, and key conditions for success are presented in Table 2.

Table . Translation of themes to design statements and conditions of success.

Interpretive Descriptive qualitative study theme	Reframed to: design statements	Conditions for success
Should I use it?	I should use it if:	• I believe it will work for me • There is evidence of credibility • There is a clear indication that it is suitable for me • It resonates with my current attitude toward exercise, support situation
Can I use it?	I can use it if:	• I can use it competently • I can use it confidently • It can be tailored and adapted to my unique needs • I can use it safely, without causing more harm • I have the belief that I could use it, given the resources and capacity I have • I have the right support to use it
Will I use it?	I will use it if:	• It is responsive to my unique needs • It encourages me to progress when I am ready • I feel supported to use it • I can see progress as a consequence of using it • It keeps me coming back

Overarching Intervention Objectives

Methods

Next, we articulated the overarching intervention objectives. Succinctly describing the intervention objectives allows a snapshot of the key characteristics of the intervention [23]. We, therefore, clearly articulated how SPIN is distinctive and different from other interventions, reflecting the specific behavioral issues, needs, and challenges it must address.

We developed the intervention objectives iteratively, repeatedly revising the wording with reference to the original research question and design statements, and with input from the research team and stakeholders. Stakeholders included pwSCI, a clinician with experience in SCI rehabilitation, a clinician who was also a pwSCI and a representative of a relevant nongovernmental organization. Each iteration strived to reflect the essence of the needs expressed by the participants with wording that represented what ideal uptake and use of this self-guided digital exercise intervention could look like. The overarching

XSL•FO
RenderX

intervention objectives were then used as a reference point for later design and development phases.

Outcome

Referring to the design statements and overall research aim, the overarching intervention objectives for SPIN were to:

1. Be tailored to users' specific and unique needs so they can relate to it and trust it and so that it can be responsive to their changing needs while using SPIN; and
2. Enable users to use it competently and confidently within their capabilities and support systems in a way that is safe and motivating.

Specific Design Objectives

Methods

Once the overarching intervention objectives were formulated, we created the specific design objectives underpinned by the design statements. We developed a working definition, incorporating the key conditions for success for each specific design objective, to ensure clarity in interpretation. These were then reviewed against the overarching intervention objectives, making sure they supported the overall objectives of SPIN. We continued to refine them as the design process progressed, during our planned discussion forums.

Outcome

Tables 3-5 each refer to a different theme. Specific design objectives and working definitions are presented in the first 2 columns; intervention functions and features are discussed in later sections.

Table . Guiding principles from the theme *Should I use it?*.

Design objectives that address identified needs, issues, and challenges	Working definition	Intervention functions	Intervention features that address the design objectives
To help users relate to and trust the program	The program will give users confidence in the source, message, and value of the program. The program is credible and legitimate and promotes trust.	• Education • Training • Modeling • Enablement • Persuasion	• Development team details (names, credentials, and contact info) • Endorsements • *Testimonials* (source matching for social comparison) • Evidence for shoulder pain exercises • How user data will be used or stored • Professional polished interface and function
To reassure users it will be clear who the program is suitable for, giving users confidence that the program is right for them and at what stage it is right for them	The program will guide users through a process to be able to screen for and identify if they are suitable to use the intervention and to promote trust and confidence that this is a safe and robust process.	• Education • Training • Modeling • Enablement • Persuasion	• Screening questionnaire/questions (that will exclude those unsuitable) • Monitoring questions at each exercise event and tracking this information • *FAQ*[a] *section* • Contact information for the team
To provide a sense of potential that it will work for them	The program will help users identify with it and the potential that it may have for them, in their current situation.	• Education • Training • Modeling • Enablement • Persuasion	• *Testimonials* (image with text, video, and quotes) of people in different "stages" of readiness or different situations. • FAQ section addressing suitability of different situations "Is this right for me?" or "How do I know this is right for me?" or "Questions I can ask to make sure this is right for me?"

[a]FAQ: frequently asked question.

XSL•FO
RenderX

Table . Guiding principles from theme *Can I use it?*.

Design objectives that address identified needs, issues, and challenges	Working definition	Intervention functions	Intervention features that address the design objectives
To promote a sense of safety when using the program	The program will ensure exercises are at the appropriate difficulty level and will be responsive to changes in user presentation to ensure that they don't significantly aggravate shoulder symptoms.	• Training • Environmental restructuring • Modeling • Enablement	• *Monitoring and tracking of shoulder pain* and exercise difficulty • Exercise selection based on user responses and a priori rules • Program-generated advice based on user responses, such as acknowledging concerns, referral to *FAQ*[a], evidence, health care provider
To promote user competence	The program will be easy to use by a range of users and in a range of circumstances, giving them a sense of confidence when using it in the context of their unique life situation.	• Training • Environmental restructuring • Modeling • Enablement	• Language at an appropriate reading level • Layout is clear and simple • Font size and buttons are large for reduced hand function • Minimal scrolling and clicking • Consistent screen layout • Clear signposts • Logical interface • Exercises presented in video and audio formats by pwSCI • Exercises presented in step-by-step processes • Exercises are planned to fit in with daily routine and normal digital device use • Tunneling of information (releasing information in small amounts, as the user progresses through "right amount, at the right time") • Graded goal setting, implementation planning • *Tailored* and action feedback based on tracking • Praise for success • Advice or support if not yet succeeded • Digital use guidance when needed (help link)
To promote user autonomy	The program will give users a sense of control and ownership over the program and their progress through the program.	• Training • Environmental restructuring • Modeling • Enablement	• *Offering choice where possible: tailoring functions* in exposure matching-timing, intensity (when and how often) • *Reminders* • Excercise selection, timing of exercise • Intervention delivery • Tunneling of options into the most common choices • Suggestions or options for different situations

[a]FAQ: frequently asked question.

XSL•FO
RenderX

Table . Guiding principles from the theme *Will I use it?*.

Design objectives that address identified needs, issues, and challenges	Working definition	Intervention functions	Intervention features that address the design objectives
To promote a positive emotional experience	The program will incorporate positive autonomy-supportive language that invites, informs, and supports users to work through the program.	• Training • Environmental restructuring • Enablement • Modeling • Education • Persuasion • Incentivization	• Use of *positive language and tone* in inviting users to decide for themselves "some find it helpful." • Use of anecdotes to describe examples of success, decision-making • Acknowledging and addressing concerns about using the program, such as pain or carer support • Using *FAQ*[a] *section* • Use of useful/interesting/relevant/personal reminders • Positive or encouraging wording on feedback on progress toward the goal
To promote a sense of relatedness	The program will be relevant to the user by using communication and wording that is tailored to their self-identified preferences and personalized to their unique circumstances.	• Training • Environmental restructuring • Enablement • Modeling • Education • Persuasion • Incentivization	• Feedback as above (and that is immediately reciprocated when interacting with the intervention) • Competition with others, and/or • Cooperation with others • Social connection through the program's grouping • Initial "getting to know you" questionnaire to help with *tailoring ingredients* • Personalization: (1) identification (including username in correspondence), (2) raising expectation (including relevant information in correspondence that is based on users' responses to questions/input), and (3) contextualization (*wording, examples that are relevant to user*-exercises relevant for tetra vs para) • *Reminders* • *Testimonials* • Self-identified support
To help users maintain their exercise over the 12 weeks	The program will use a variety of strategies and features to encourage and support users to maintain engagement in their exercise for the duration of the program.	• Training • Environmental restructuring • Enablement • Modeling • Education • Persuasion • Incentivization	• Rewards (points or similar)/competition • Goal setting • Action planning • Communication that is *positive*, immediate, and useful and *tailored*

XSL•FO
RenderX

Design objectives that address identified needs, issues, and challenges	Working definition	Intervention functions	Intervention features that address the design objectives
To promote a sense of accountability	The program will provide features that encourage the user to return to the program and to continue with the exercises.	• Training • Environmental restructuring • Enablement • Modeling • Education • Persuasion • Incentivization	• Competition with others or with self • Support from others • Communication that is *positive*, immediate, and useful and *tailored* • Communication that is personalized • Rewards that are only released upon completion of a certain amount of exercise
To promote a sense of progree and engagement	The program will enable the user to understand their progress through a clear and simple tracking feature. This will be done in a way that encourages further progress and ongoing engagement with the exercise intervention	• Training • Environmental restructuring • Enablement • Modeling • Education • Persuasion • Incentivization	• Feedback and tracking • Choice in exercise selection • Personalization • *Tailoring*

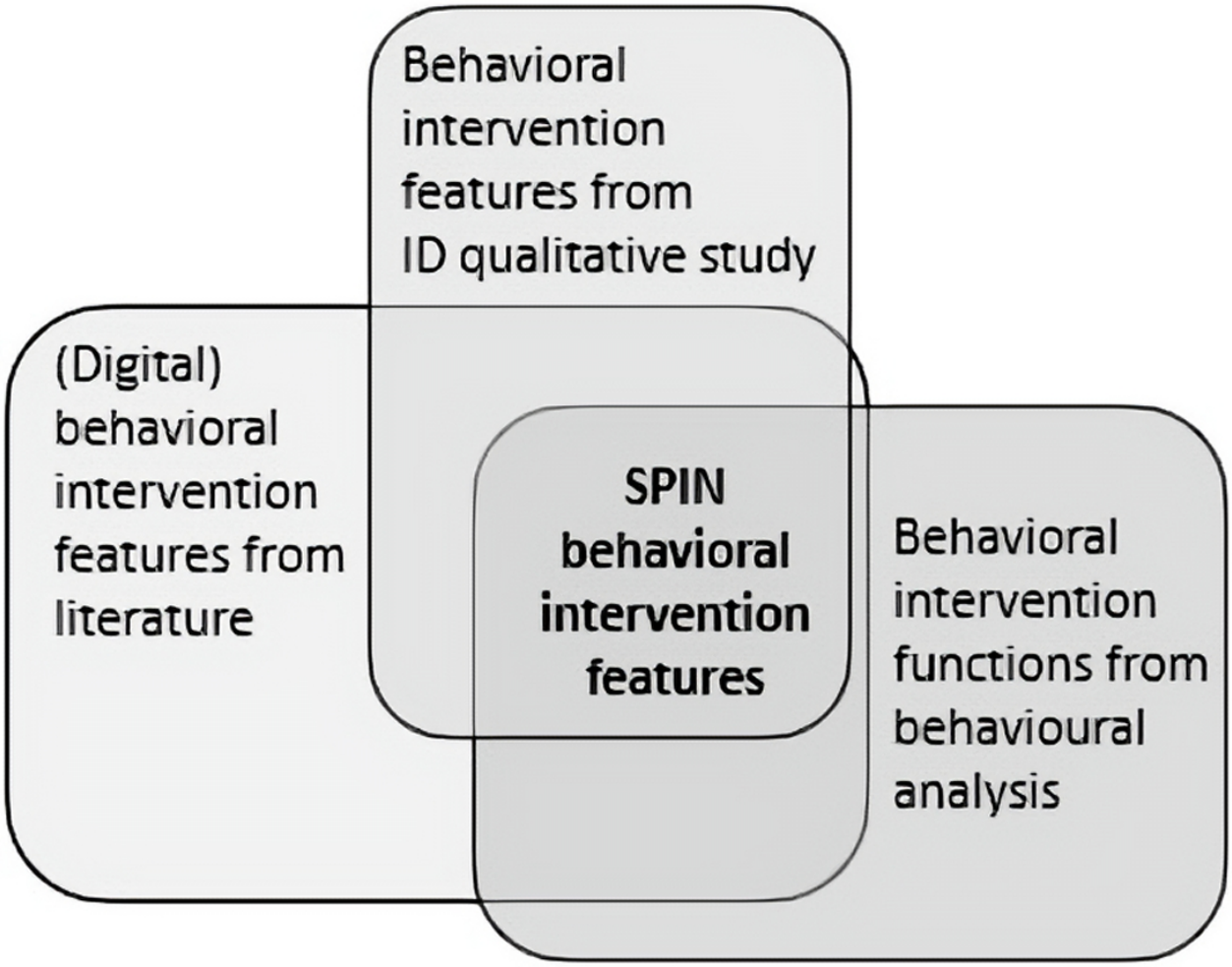
[a]FAQ: frequently asked question.

Intervention Features

In the PBA, the guiding principles inform the intervention features by providing a framework for selecting and shaping features that directly support the specific design objectives, and to improve resonance, engagement, and acceptability of an intervention [23]. A range of evidence informed the selection of behavioral intervention features: (1) in our review, we identified a range of features used in digital interventions that have been associated with better health-related outcomes [27,34-36]; (2) we identified possible behavioral intervention design features from our Interpretive Descriptive qualitative study [33]; and (3) we identified behavioral "intervention functions" we were trying to achieve using a behavioral analysis as per Michie and colleagues' framework [26]. We then mapped these to the most relevant intervention features. Figure 2 represents the layers of evidence that informed SPIN's intervention features. We will describe each of these in detail below.

XSL•FO
RenderX

Figure 2. Layers of evidence that informed Shoulder Pain Intervention delivered over the Interjet's intervention features. ID: Interpretive Descriptive; SPIN: Shoulder Pain Intervention delivered over the interNet.



Identifying Behavioral Intervention Features From Previous Literature Review and the Interpretive Descriptive Qualitative Study

Methods

In Step 1, we had earlier identified potential behavioral intervention features for self-guided interventions that were identified from our systematic review and meta-analysis, using the specifically developed template, drawing from the CONSORT-EHEALTH checklist (V.1.6.1) [38]. See Multimedia Appendix 1 for a sample of our template showing sections used to record behavioral intervention features. For this current stage of the SPIN design, we also reviewed intervention features of publications that missed the strict inclusion criteria of the systematic review and meta-analysis but addressed digital delivery of physical activity or exercise intervention for possible relevant behavioral intervention features. We then integrated the data on specific features collected from our Interpretive Descriptive qualitative study. These data were categorized by proposed purpose and function and then mapped against the specific design objectives.

Outcome

There was overlap, resulting in some features identified as addressing more than one design objective. Many of the studies in the systematic review included digital behavioral intervention features that involved instruction on exercise or physical activity performance, self-monitoring of the exercise or physical activity behavior, goals and planning, and prompting. The results of the Interpretive Descriptive qualitative study and other reviewed literature suggested additional behavioral intervention features. Table 6 presents a summary of the behavioral intervention feature categories that we considered for SPIN, the design objective(s) they are related to, and the supporting evidence.

XSL•FO
RenderX

Table . Behavioral intervention feature categories supported by systematic review, Interpretive Descriptive qualitative study, and existing literature.

Design objectives	Behavioral intervention feature	Proportion of studies identified in systematic review and meta-analysis (out of 16 studies)	Identified in Interpretive Descriptive qualitative study	Identified in other literature not included in meta-analysis
Should I use it?	Ensuring personal relevance	16	✓	Horsch et al [41]
Should I use it?/Can I use it?	Use of credibility and trust-enhancing features	5	✓	Bossen et al [42]; Oinas-Kukkonen and Harjumaa [43]
Should I use it?/Will I use it?	Provision of information about actual users	2	✓	Morrison et al [34]
Can I use it?	Allowance of the user to control or adapt features	7	✓	McClure et al [44]
Can I use it?	Ensuring ease of use	6	✓	Carter et al [45]; Hurling et al [46]; Webb et al [27]
Can I use it?	Provision of information 'just in time' and in 'just the right amount'	9	✓	Oinas-Kukkonen and Harjumaa [43]; Xu et al [47]
Can I use it?/Will I use it?	Use of goal setting	8	✓	Webb et al [27]; Willett et al [48]; Dugas et al [49]
Can I use it?	Use of demonstration of behavior	10	✓	Webb et al [27]
Can I use it?	Use of feedback of behavior	10	✓	Webb et al [27]; Dugas et al [49]
Can I use it?	Use of tailored feedback	10	✓	Morrison et al [35]; Dugas et al [49]
All	Use of tailoring based on a number of variables	5	✓	Morrison et al [34]; Couper et al [50]; Xu et al [47]; Oinas-Kukkonen and Harjumaa [43]; Figueiras and Neto [51]; Dugas et al [49]
Can I use it?/Will I use it?	Use of reminders	8	✓	Webb et al [27]; Lin and Wu [52]; Alahäivälä and Oinas-Kukkonen [53]; Dugas et al [49]
Can I use it?/Will I use it?	Use of self-monitoring features	9	✓	Morrison et al [34]; Glasgow et al [54]; Willett et al [48]
Can I use it? /Will I use it?	Use of positive tone and language	4	✓	Haines-Saah et al [55]
Can I use it? /Will I use it?	Use of text message	3	✓	Webb et al [27]
Will I use it?	Use of action/coping planning	5	✓	Webb et al [27]; Glasgow et al [54]; van Genugten et al [56]
Will I use it?	Use of facilitation of social comparison and support	2	✓	Webb et al [27]; Davies et al [57]; Perski et al [58]; Alahäivälä and Oinas-Kukkonen [53]; Xu et al [47]
Will I use it?	Use of rewards and incentives	1	✓	Khadjesari et al [59]; Schubart et al [60]; van Genugten et al [56]
All	Use of a combination and a number of features	14		Webb et al [27]; Meade et al [61]

XSL•FO
RenderX

Identifying Intervention Functions From a Behavioral Analysis

Methods

We included a behavioral analysis using the "Behaviour Change Wheel" and COM-B model as outlined by Michie and colleagues [26]. This is a theoretical framework that provides a systematic way of identifying the problem and analyzing the behavioral needs of a target behavior. The "Behaviour Change Wheel" can support intervention design by linking the identified behavioral needs to "intervention functions" through a mechanism of action.

Consistent with the guiding principles and specific design objectives, and for the purpose of this behavioral analysis, we reframed the 3 themes from the Interpretive Descriptive qualitative study into target behaviors: *Should I use it?*—Signing up to SPIN (Table 7); *Can I use it?*—Using SPIN (Table 8); and *Will I use it*?—Returning to SPIN over the 12 weeks (Table 9). The COM-B Model was then used to identify the capability (C), opportunity (O), and motivational (M) components required for each of these behaviors (B) to occur, referring to the specific design objectives. The questions "*what needs to happen for the target behavior to occur?"* and "*is there a need to change?"* facilitated the analysis process [26]. We used this process to identify (or "diagnose") the relevant COM-B components that need to be addressed for the target behavior to occur (see the Behavioral diagnosis of the relevant COM-B components in Tables 7-9).

XSL•FO
RenderX

Table . Behavioral analysis of target behavior: signing up to SPIN[a] (*Should I use it?*) for people living with spinal cord injury who have shoulder pain.

COM-B components[b]	What needs to happen for the target behavior to occur?	Is there a need for change?
Physical capability	Have the physical ability to access SPIN features and functions and use it	No change needed as SPIN will only be suitable for people who can physically access and use it
Psychological capability	Believe they have the capability to use SPIN	*Change* needed as pwSCI[c] will want reassurance that they have sufficient physical capability to use SPIN and/or that it is suitable for people with their level of physical ability
Psychological capability	Know that exercise can improve pain symptoms (or not make the condition worse)	*Change* may be needed as there may be fears or concerns that exercise could worsen pain symptoms
Physical opportunity	Have a device that can access SPIN	No change needed as SPIN will only be suitable for those people who have devices that can access SPIN
Social opportunity	Know about other pwSCI who have either benefitted from exercise for shoulder pain or are using SPIN	*Change* needed as pwSCI may not know about others who have benefitted from exercise to improve shoulder pain symptoms or who are using SPIN
Reflective motivation	Hold beliefs that exercising will reduce pain symptoms and/or improve activity	*Change* needed as pwSCI may be fearful that exercise may worsen pain symptoms
Reflective motivation	Believe that SPIN has been developed by a credible and trustworthy source	*Change* needed as pwSCI will want to assure themselves that SPIN has been developed by knowledgeable personnel who have experience in SCI[d] rehabilitation
Automatic motivation	Believe that SPIN will identify those that are suitable (and unsuitable) to use it	*Change* needed as pwSCI will want assurance that SPIN is appropriate for their circumstances and can be tailored for their needs
Automatic motivation	Need to feel that SPIN resonates (with current attitude toward exercise, support situation)	*Change* needed as pwSCI need to feel comfortable that SPIN is right for them at this time
Behavioral diagnosis of the relevant COM-B components	Psychological capability, social opportunity, reflective and automatic motivation need to change for the target behavior to occur	—[e]
Likely ”intervention functions” that link to COM-B	Education (psychological capability, reflective motivation), Training (physical opportunity), Modelling (social opportunity), and Persuasion (reflective motivation, automatic motivation)	—

[a]SPIN: Shoulder Pain Intervention delivered over the interNet.

[b]Behavioral diagnosis of the relevant COM-B components: psychological capability, social opportunity, reflective and automatic motivation need to change for the target behavior to occur.

[c]pwSCI: people living with spinal cord injury.

[d]SCI: spinal cord injury.

[e]not applicable.

XSL•FO
RenderX

Table . Behavioral analysis of target behavior: using SPIN[a] (*Can I use it?*) for people living with spinal cord injury who have shoulder pain.

COM-B components[b]	What needs to happen for the target behavior to occur?	Is there a need for change?
Physical capability	Have the physical ability to control and manipulate SPIN features and functions and related equipment and setup	*Change* may be needed as pwSCI[c] will want reassurance that they have sufficient physical capability to use the intervention and/or that the intervention is suitable for people with their level of physical ability
Physical capability	Have the additional support as required	*Change* may be needed with additional support for equipment setup and exercise support
Psychological capability	Believe they have the capability to use SPIN	*Change* needed as pwSCI will want reassurance that they have sufficient physical capability to use SPIN and/or that it is suitable for people with their level of physical ability
Psychological capability	Know how to navigate through the intervention	*Change* needed to clearly provide pwSCI with signposts and information to guide them through
Psychological capability	Know how to perform exercises safely	*Change* needed to ensure appropriate level of exercises is offered and explained to maximize safe exercising and to ensure that the program is responsive to changes in user presentation
Physical opportunity	Have a program that is usable and easy to follow	*Change* needed to ensure SPIN is easy to use and understand
Social opportunity	Haencouragement from peers	*Change* needed to ensure access to a community of users
Reflective motivation	Have confidence in one's ability to use the intervention program	*Change* needed to provide a sense of ownership and control of the program, with positive reinforcement with use
Reflective motivation	Have belief the intervention will enable achievement of outcomes important to user	*Change* needed as users may not recognize the value of SPIN
Automatic motivation	Have experience of benefit from intervention and sense of progress	*Change* needed to provide consistent exercise opportunities
Behavioral diagnosis of the relevant COM-B components	Physical and psychological capability, physical and social opportunity, and reflective motivation need to change for the target behavior to occur	—[d]
Likely "intervention functions" that link to COM-B	Training (physical capability, psychological capability), Environmental restructuring (physical opportunity), Modelling (social opportunity), and Persuasion (reflective motivation)	—

[a]SPIN: Shoulder Pain Intervention delivered over the interNet.

[b]Behavioral diagnosis of the relevant COM-B components: physical and psychological capability, physical and social opportunity, and reflective motivation need to change for the target behavior to occur.

[c]pwSCI: people living with spinal cord injury.

[d]not applicable.

XSL•FO
RenderX

Table . Behavioral analysis of target behavior: using SPIN[a] (*Will I use it?*) for people living with spinal cord injury who have shoulder pain.

COM-B components[b]	What needs to happen for the target behavior to occur?	Is there a need for change?
Physical capability	Have the physical ability to control and manipulate SPIN features and functions and related equipment and setup	*Change* may be needed as pwSCI[c] will want reassurance that they have sufficient physical capability to use the intervention and/or that the intervention is suitable for people with their level of physical ability
Physical capability	Have the additional support as required	*Change* may be needed with additional support for equipment setup and exercise support
Psychological capability	Believe they have the capability to use SPIN	*Change* needed as pwSCI will want reassurance that they have sufficient physical capability to use SPIN and/or that it is suitable for people with their level of physical ability
Psychological capability	Know how to navigate through the intervention	*Change* needed to clearly provide pwSCI with signposts and information to guide them through
Psychological capability	Know how to perform exercises safely	*Change* needed to ensure appropriate level of exercises is offered and explained to maximize safe exercising and to ensure that program is responsive to changes in user presentation
Physical opportunity	Have a program that is usable and easy to follow	*Change* needed to ensure SPIN is easy to use and understand
Social opportunity	Have encouragement from peers	*Change* needed to ensure access to a community of users
Reflective motivation	Have confidence in one's ability to use the intervention program	*Change* needed to provide a sense of ownership and control of the program, with positive reinforcement with use
Reflective motivation	Have belief the intervention will enable achievement of outcomes important to user	*Change* needed as users may not recognize the value of SPIN
Automatic motivation	Have experience of benefit from intervention and sense of progress	*Change* needed to provide consistent exercise opportunities
Behavioral diagnosis of the relevant COM-B components	Physical and psychological capability, physical and social opportunity, and reflective motivation need to change for the target behavior to occur	—[d]
Likely "intervention functions" that link to COM-B	Training (physical capability, psychological capability), Environmental restructuring (physical opportunity), Modelling (social opportunity), and Persuasion (reflective motivation)	—

[a]SPIN: Shoulder Pain Intervention delivered over the interNet.

[b]Behavioral diagnosis of the relevant COM-B components: physical and psychological capability, physical and social opportunity, and reflective motivation need to changefor the target behavior to occur.

[c]pwSCI: people living with spinal cord injury.

[d]not applicable.

Next, we mapped these components to established 'intervention functions,' using the "Behaviour Change Wheel." Most relevant "intervention functions" were then identified from the matrix of links between COM-B and intervention functions [26]. The "Behaviour Change Wheel" uses the term "intervention function" in lieu of intervention "type" or "category" since the same intervention feature may address more than 1 function [26].

Outcome

Tables 7-9 present the target behavior for each design objective and what (if any) change is needed to occur based on the COM-B components. "Intervention functions" most likely to support behavior change have also been identified. For example, testimonials about positive experiences of using exercise to help with shoulder pain could be a form of modeling (providing an example for people to aspire to) and persuasion (using communication to induce positive feelings or stimulate action). This mapping process allowed each specific design objective to be checked to ensure it was supported by an appropriate "intervention function" and corresponding intervention feature. "Intervention functions" linked to the target behavior have been included in Likely "intervention functions" that link to the COM-B in each of the tables (Tables 7-9). The guiding principles tables (Tables 3-5) provide an overview of how these

XSL•FO
RenderX

"intervention functions" map to the design objectives ("Intervention functions" column).

Step 3: Selection of Intervention Features Applicable to SPIN

Methods

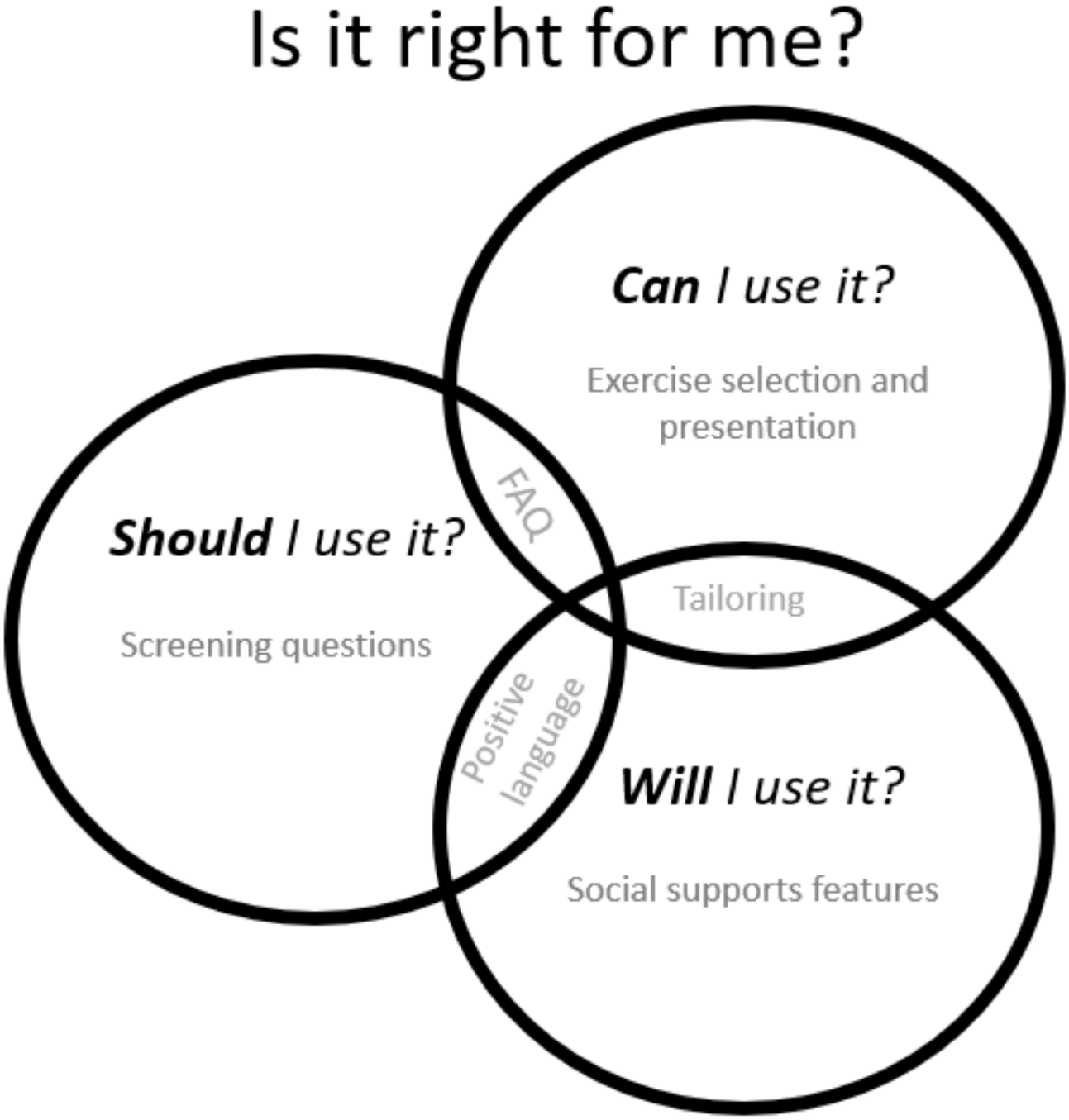
The design phase of the PBA involves identifying intervention features and content, guided by the previously formulated guiding principles, to ensure alignment with users' psychosocial contexts and to enhance relevance, acceptability, and engagement through iterative user feedback [23]. We were able to begin selecting specific SPIN intervention features once the behavioral analysis was complete. The behavioral intervention features previously identified (Table 6) were reviewed. We mapped those that we felt were contextually appropriate against the "intervention functions." Each was checked to ensure it supported the specific design objectives and the overarching intervention objective. VS completed this process in consultation with coauthors.

Outcome

Collectively, Tables 3-5 demonstrate a complete representation of the guiding principles of SPIN's proposed intervention features and functions, mapped back to the design objectives. Some intervention features address more than 1 intervention design objective. These features have been italicized in Tables 3-5. For example, having a forum for frequently asked questions may reduce barriers to starting the intervention and give users the information they need to progress. Having positive, encouraging language can attract users to start using the intervention and motivate them to continue with it. Other intervention features more clearly support only one of the intervention design objectives. Figure 3 schematically presents an example of how overlapping intervention features cohesively support SPIN's identified design objectives.

XSL•FO
RenderX

Figure 3. Schematic representation of overlapping intervention features.



Application of Our Design Steps to Future Intervention Design

We believe that by explicating how we used the PBA in the development of SPIN, we can support others to use the PBA in the design of interventions. Table 10 provides a summary view of our process and includes some questions that we hope will prompt other researchers to consider how they might operationalize the use of PBA in their work. The table provides an overview of key phases of PBA and possible timelines (column 1) and examples from SPIN (column 2), including tools and methods we drew on as complementary to PBA and which we found useful in operationalizing the approach. In column 3, we have included our reflections on the benefits of our approach. The final column has questions that we hope will serve as prompts for researchers and designers when using this approach.

XSL•FO
RenderX

Table . Operationalizing the person-based approach: our experience and future applications.

1. Key steps in the person-based approach	2. Methods we used to operationalize PBA[a] steps in the development of SPIN[b]	3. Strengths and opportunities of our approach	4. Questions to consider when planning this step
Step 1 (months 3 - 6) Identify key behavioral issues (access), needs (not feeling competent), and challenges the intervention must address	Interpretive Descriptive qualitative study to explore user perspectives of self-guided exercise intervention and what would help or hinder uptake of a self-guided digital exercise intervention. Used probes and images during data collection to help users visualize and provide feedback on possible intervention features.	Drawing on Interpretive Descriptive as a nested study within the PBA process helped to provide a robust framework to capture and make sense of user needs and preferences. Interpretive Descriptive is congruent with the goals of PBA and has the benefit of (1) being oriented toward translation from the outset, (2) prioritizing the production of clinically relevant insights, and (3) flexibility in methods so data collection and analysis could be tailored to the intended use of findings for intervention development.	Who are the users? What is the best way to understand their unique context and specific needs? Are there existing tools and methods available that would be fit for purpose to capture user needs and preferences? How is the information going to be used? How might your approach to capturing needs and preferences be optimized for this intended use? Does data already exist (systematic reviews and qualitative research) that can help inform this step?
Step 2 (months 6-9-12) Creating intervention design objectives that capture what is unique about your intervention and reflect the specifically identified user needs and challenges the intervention needs to address.	Translate themes from the Interpretive Descriptive study into design statements and conditions for success, drawing on the data from each theme.	Helped to reframe the themes into actionable statements. Provided an evidence-based framework to underpin intervention design objectives. Ensured that user needs and preferences will continue to be reflected in the design process.	How are user needs and preferences currently expressed? Can they be used to underpin design objectives in their current form, or do they need some further refinement/transformation?
Step 2 (months 6-9-12) Creating intervention design objectives that capture what is unique about your intervention and reflect the specifically identified user needs and challenges the intervention needs to address.	Two overarching objectives for SPIN were developed from the design statements and conditions of success. It was repeatedly revised, referring to the original research question and design statements, and with input from stakeholders.	Developing 2 overarching objectives, rather than 1, helped to make explicit 2 interrelated but distinct objectives. The intermediary step of developing design objectives from the qualitative study themes ensured that the objectives represent the essence of the needs expressed by the users. Refining with input from stakeholders helped to ensure the objectives remained resonant with the SCI[c] community. Articulating these objectives at the outset was a useful reference point to keep coming back to for all later design and development phases.	What is(are) the overarching intervention objective(s)? How will you ensure your overarching objective(s) remain(s) grounded by user needs and preferences? Who might need to be involved in the development of intervention objective(s)? How will you know if your intervention objective(s) adequately capture(s) the perspectives of future users?
Step 2 (months 6-9-12) Creating intervention design objectives that capture what is unique about your intervention and reflect the specifically identified user needs and challenges the intervention needs to address.	Specific design objective were identified, drawing on the design statements and overarching intervention objectives. Working definitions were formulated with reference to original data sources and in collaborative discussions as a research team.	The development of specific design statements provided a framework to identify design requirements (system requirements) and intervention features. Investing time to develop the working definitions as a team, with reference to original data sources, was important for clarity and shared understanding.	What process will you use to generate specific design objectives from your overarching design objective(s)? Who might need to be involved in that process? What data sources do you have that you can refer to so you can refine your specific design objectives?

XSL•FO
RenderX

1. Key steps in the person-based approach	2. Methods we used to operationalize PBA[a] steps in the development of SPIN[b]	3. Strengths and opportunities of our approach	4. Questions to consider when planning this step
Step 3 (months 6 - 12) Select and refine intervention features that support the specific design objectives.	Several methods were used to support the selection and refinement of intervention features for SPIN including: (1) extracting data on intervention features from a previous systematic review on self-guided exercise interventions, (2) reviewing relevant behavioral theory, (3) undertaking a behavioral analysis, and (4) drawing on persuasive system design.	Drawing on a multiplicity of methods in this step (1) ensured an evidence-based and theoretically informed approach and (2) enabled a systematic approach to ensure intervention features were those best suited to the behavioral needs of the SPIN user. A systematic approach to identifying intervention features and mapping them back to design objectives helps to imrove the credibility of intervention design. The outcome was a clear framework for SPIN intervention design that was a useful tool to support communication with design colleagues or software developers who were then bringing SPIN to form.	What data sources are available that can help you identify potential intervention features? What midrange theories are available that can help you identify potential intervention features? Are there existing tools and methods available that would be fit for purpose to help you identify intervention features which respond to user needs and preferences? Of all the potential intervention features, which are most likely to meet the design objective(s)? Who else should be involved in this process? How might you ensure that the outcome of this process can be an accessible and usable framework for others involved in intervention development?

[a]PBA: person-based approach.

[b]SPIN: Shoulder Pain Intervention delivered over the interNet.

[c]SCI: spinal cord injury.

Ethical Considerations

Ethics approval (Auckland University of Technology Ethics Committee-AUTEC 18/263) and participant consent were received for the earlier work [33] that informed this work.

Discussion

Principal Findings

This paper has described how we applied evidence, theory, and person-based approaches in the design of a self-guided digital intervention to help pwSCI manage their shoulder pain. We have detailed the processes of applying the PBA to the design of SPIN.

This builds on Yardley and colleagues' [23] collection of work. The PBA emphasizes a detailed, qualitative understanding of users' psychosocial contexts to inform intervention design. It adds value to user-centered design by addressing factors that influence behavior change, beyond just usability. The PBA complements theory- and evidence-based frameworks, such as the "Behaviour Change Wheel" [25] by tailoring interventions to the needs and preferences of specific populations. Despite growing evidence for the use of the PBA framework in intervention design [29-31,62,63], there is little available on its operationalization. To our knowledge, the detailed reporting of each step has not been available before.

In a recent systematic review on the effectiveness of self-guided digital exercise interventions, Stavric et al [21] found that interventions with theoretical underpinning had increased congruence with the intervention features leading to significant positive results. This is supported by findings from McEwan [64] who found that theory-based interventions resulted in more consistent significant improvements in physical activity. The pwSCI and shoulder pain who will use SPIN are likely to have minimal contact with a health care professional. Therefore, successful design required an understanding of how SPIN would meet their needs and how pwSCI would use it in daily life in a self-directed way. Engaging people and evidence in intervention design is supported by a range of researchers and designers [23,34,65]. Using a person-based approach, drawing on evidence from the people who will use the intervention, to derive the behavioral strategies has been shown to be effective in a variety of settings and methods of delivery [18,29,30,66,67].

Despite acknowledgment that interventions supported by theory and evidence maximize outcomes [68,69], there remains a paucity of full intervention description or design disclosure [70-73], making it challenging to explicate the link between theory and evidence and intervention features. A key tension we encountered was the limited availability of detailed examples of how the PBA had been operationalized in practice. This required us to make interpretive decisions when translating PBA principles into design elements, often without clear guidance. Additionally, balancing adherence to the PBA's iterative, user-focused process with practical constraints such as time, resources, and access to participants posed challenges. These limitations were compounded by the fact that we were largely self-taught in the application of both the PBA and behavioral analysis frameworks.

Michie and colleagues [74] recognized the challenges and lack of clarity around the purported mechanisms by which digital interventions work during an international workshop on developing and evaluating digital interventions to promote behavior change in health. DiLiberto and colleagues [75] support the importance of "insider accounts" of intervention implementation and argue that the same transparent reporting practice should apply to intervention design. Of the 16 self-guided interventions included in our systematic review and meta-analysis conducted in the planning phase [21], only 6

XSL•FO
RenderX

provided any reference to methods used to plan, design, and develop them [19,76-81]. Of these, there was little supporting detail on how the design was carried out and none of the included studies reported exploring the behavioral needs of the users before designing the intervention. Future researchers might benefit from greater transparency and reporting of the design phase, including more practical examples of operationalizing person-based and behavioral approaches. Considerations for mitigating these challenges include allocating sufficient time and resources for user involvement beyond the planning stage, documenting key design decisions, and seeking opportunities for peer collaboration to support methodological alignment and confidence.

Strengths

We have shown commitment to providing a robust and transparent process in the operationalization of the design phase of SPIN drawing on the PBA approach. This process included explicitly addressing the identified behavioral needs of the users and kept these central throughout the entire design process. The design of SPIN has demonstrated how we used evidence (from existing literature and from a previous Interpretive Descriptive qualitative study) and theory (from behavioral analysis, "Behaviour Change Wheel," and COM-B) to enhance the person-based process. This explicit and thorough process of planning and designing SPIN has provided a blueprint for intervention development when using PBA. It also addresses many of the limitations in the reporting on the development processes for existing self-guided digital exercise and physical activity-related interventions.

Limitations

Our operationalization of the PBA design phase reflects our interpretation of the PBA steps through available readings. We acknowledge there may be other perspectives and understandings. However, we believe that it is important to make our experiences visible to build on previous work and support future intervention design. Similarly, we relied on literature and online course instruction for support when we conducted the behavioral analysis using the COM-B. Being self-taught in both the PBA and behavioral analyses may mean that some aspects of our approach are not consistent with the original intent of these approaches. However, this is perhaps an artifact of the knowledge mobilization process, where the application of knowledge can change as knowledge changes hands. By offering transparency in our process, we hope that people can draw their own conclusions regarding the robustness of our approach. The design of SPIN did not include a logic model. Logic models typically include the main intervention components, how they relate to one another, which are meant to produce which effect, and include processes and expected outcomes. However, we did not believe a logic model would have been pragmatically useful as they assume causal relationships which may have restricted our thinking about solutions [82]. Our development process was underpinned by relationship building and community interaction, both of which are complex and require flexibility [83,84].

Future Steps of SPIN Using the PBA

With the proposed intervention features selected, SPIN wireframes have been constructed. Wireframes are images or screenshots that show how screens of a website or app are structured and how content is arranged. These have provided a visual representation of the product and an opportunity to comment on content, features, and organization without getting distracted by aesthetics. Further participant consultation and design refinement have occurred. Frontend and backend software programming will occur at a later phase. Reporting of these stages will follow in a subsequent publication.

Conclusion

The design of SPIN has incorporated a deep understanding of the users' needs and best available evidence by drawing on the PBA design process to maximize chances of engagement and outcomes. This paper has made visible the operationalization of each of the phases and can act as a blueprint to provide guidance to future researchers when using this approach.

Authors' Contributions

Conceptualization: VS, NMK
Methodology: VS, NLS, NMK
Supervision: NLS, NMK
Visualization: VS
Writing – original draft: VS
Writing – review & editing: VS, NLS, NMK

Conflicts of Interest

None declared.

Multimedia Appendix 1
Systematic review data collection template.
[XLSX File, 25 KB - mhealth_v14i1e66678_app1.xlsx]

References

XSL•FO
RenderX

1. Dalyan M, Cardenas DD, Gerard B. Upper extremity pain after spinal cord injury. Spinal Cord 1999 Mar;37(3):191-195. [doi: 10.1038/sj.sc.3100802] [Medline: 10213328]
2. Curtis KA, Drysdale GA, Lanza RD, Kolber M, Vitolo RS, West R. Shoulder pain in wheelchair users with tetraplegia and paraplegia. Arch Phys Med Rehabil 1999 Apr;80(4):453-457. [doi: 10.1016/s0003-9993(99)90285-x] [Medline: 10206610]
3. Gutierrez DD, Thompson L, Kemp B, Mulroy SJ, Physical Therapy Clinical Research Network, Rehabilitation Research and Training Center on Aging-Related Changes in Impairment for Persons Living with Physical Disabilities. The relationship of shoulder pain intensity to quality of life, physical activity, and community participation in persons with paraplegia. J Spinal Cord Med 2007;30(3):251-255. [doi: 10.1080/10790268.2007.11753933] [Medline: 17684891]
4. McCasland LD, Budiman-Mak E, Weaver FM, Adams E, Miskevics S. Shoulder pain in the traumatically injured spinal cord patient: evaluation of risk factors and function. J Clin Rheumatol 2006 Aug;12(4):179-186. [doi: 10.1097/01.rhu.0000230532.54403.25] [Medline: 16891921]
5. Sie IH, Waters RL, Adkins RH, Gellman H. Upper extremity pain in the postrehabilitation spinal cord injured patient. Arch Phys Med Rehabil 1992 Jan;73(1):44-48. [doi: 10.5555/uri:pii:000399939290225L] [Medline: 1729973]
6. Coulter EH, McLean AN, Hasler JP, Allan DB, McFadyen A, Paul L. The effectiveness and satisfaction of web-based physiotherapy in people with spinal cord injury: a pilot randomised controlled trial. Spinal Cord 2017 Apr;55(4):383-389. [doi: 10.1038/sc.2016.125] [Medline: 27596027]
7. Best KL, Bourassa S, Sweet SN, Routhier F. Expert consensus for a digital peer-led approach to improving physical activity among individuals with spinal cord injury who use manual wheelchairs. J Spinal Cord Med 2023 Jan 2;46(1):53-61. [doi: 10.1080/10790268.2021.1986308]
8. Baehr LA, Hiremath SV, Bruneau M Jr, et al. Effect of Tele-exercise to Promote Empowered Movement for Individuals With Spinal Cord Injury (TEEMS) program on physical activity determinants and behavior: a mixed methods assessment. Arch Phys Med Rehabil 2024 Jan;105(1):101-111. [doi: 10.1016/j.apmr.2023.08.019] [Medline: 37678447]
9. Carey RL, Le H, Coffman DL, et al. mHealth-based just-in-time adaptive intervention to improve the physical activity levels of individuals with spinal cord injury: protocol for a randomized controlled trial. JMIR Res Protoc 2024 Jun 28;13:e57699. [doi: 10.2196/57699] [Medline: 38941145]
10. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012 Apr 30;9(52):1-13. [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
11. Jahangiry L, Farhangi MA, Shab-Bidar S, Rezaei F, Pashaei T. Web-based physical activity interventions: a systematic review and meta-analysis of randomized controlled trials. Public Health (Fairfax) 2017 Nov;152:36-46. [doi: 10.1016/j.puhe.2017.06.005] [Medline: 28734170]
12. Coulter CL, Weber JM, Scarvell JM. Group physiotherapy provides similar outcomes for participants after joint replacement surgery as 1-to-1 physiotherapy: a sequential cohort study. Arch Phys Med Rehabil 2009 Oct;90(10):1727-1733. [doi: 10.1016/j.apmr.2009.04.019] [Medline: 19801063]
13. Singh G, MacGillivray M, Mills P, Adams J, Sawatzky B, Mortenson WB. Patients' perspectives on the usability of a mobile app for self-management following spinal cord injury. J Med Syst 2019 Dec 11;44(1):1-9. [doi: 10.1007/s10916-019-1487-y] [Medline: 31828440]
14. Van Straaten MG, Cloud BA, Morrow MM, Ludewig PM, Zhao KD. Effectiveness of home exercise on pain, function, and strength of manual wheelchair users with spinal cord injury: a high-dose shoulder program with telerehabilitation. Arch Phys Med Rehabil 2014 Oct;95(10):1810-1817. [doi: 10.1016/j.apmr.2014.05.004] [Medline: 24887534]
15. Bizzarini E, Chittaro L, Frezza M, et al. A mobile app for home-based exercise in spinal cord injured persons: proposal and pilot study. Digit Health 2022;8:20552076211070724. [doi: 10.1177/20552076211070724] [Medline: 35140978]
16. Brooks MA, Beaulieu JE, Severson HH, et al. Web-based therapeutic exercise resource center as a treatment for knee osteoarthritis: a prospective cohort pilot study. BMC Musculoskelet Disord 2014 May 17;15(1):2-11. [doi: 10.1186/1471-2474-15-158] [Medline: 24884547]
17. Nelligan RK, Hinman RS, Kasza J, Crofts SJC, Bennell KL. Effects of a self-directed web-based strengthening exercise and physical activity program supported by automated text messages for people with knee osteoarthritis: a randomized clinical trial. JAMA Intern Med 2021 Jun 1;181(6):776-785. [doi: 10.1001/jamainternmed.2021.0991] [Medline: 33843948]
18. Geraghty AWA, Essery R, Kirby S, et al. Internet-based vestibular rehabilitation for older adults with chronic dizziness: a randomized controlled trial in primary care. Ann Fam Med 2017 May;15(3):209-216. [doi: 10.1370/afm.2070] [Medline: 28483885]
19. van Vugt VA, van der Wouden JC, Essery R, et al. Internet based vestibular rehabilitation with and without physiotherapy support for adults aged 50 and older with a chronic vestibular syndrome in general practice: three armed randomised controlled trial. BMJ 2019 Nov 5;367:l5922. [doi: 10.1136/bmj.l5922] [Medline: 31690561]
20. Lozano-Lozano M, Martín-Martín L, Galiano-Castillo N, et al. Mobile health and supervised rehabilitation versus mobile health alone in breast cancer survivors: randomized controlled trial. Ann Phys Rehabil Med 2020 Jul;63(4):316-324. [doi: 10.1016/j.rehab.2019.07.007] [Medline: 31454561]

XSL•FO
RenderX

21. Stavric V, Kayes NM, Rashid U, Saywell NL. The effectiveness of self-guided digital interventions to improve physical activity and exercise outcomes for people with chronic conditions: a systematic review and meta-analysis. Front Rehabilit Sci 2022;3:1-16. [doi: 10.3389/fresc.2022.925620]
22. Stavric V, Saywell N, Kayes NM. Development of a self-guided web-based exercise intervention (SPIN) to treat shoulder pain in people living with spinal cord injury: protocol of a mixed methods study. BMJ Open 2019 Sep 17;9(9):e031012. [doi: 10.1136/bmjopen-2019-031012] [Medline: 31530616]
23. Yardley L, Morrison LG, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30. [doi: 10.2196/jmir.4055] [Medline: 25639757]
24. Michie S, Atkins L, West R. The Behaviour Change Wheel A Guide to Designing Interventions, 1st edition: Great Britain: Silverback Publishing; 2014:1003-1010 URL: https://www.behaviourchangewheel.com/ [accessed 2025-12-01]
25. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6(42):42. [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
26. Michie S, Atkins L, Gainforth HL. Changing behaviour to improve clinical practice and policy. In: Novos Desafios, Novas Competências: Contributos Atuais Da Psicologia [Internet]: Braga: Axioma - Publicações da Faculdade de Filosofia; 2016:41-60. [doi: 10.17990/Axi/2016_9789726972679_041]
27. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4. [doi: 10.2196/jmir.1376] [Medline: 20164043]
28. Muller I, Kirby S, Yardley L. Understanding patient experiences of self-managing chronic dizziness: a qualitative study of booklet-based vestibular rehabilitation, with or without remote support. BMJ Open 2015 May 18;5(5):e007680. [doi: 10.1136/bmjopen-2015-007680] [Medline: 25986639]
29. Little P, Stuart B, Francis N, et al. Effects of internet-based training on antibiotic prescribing rates for acute respiratory-tract infections: a multinational, cluster, randomised, factorial, controlled trial. The Lancet 2013 Oct;382(9899):1175-1182. [doi: 10.1016/S0140-6736(13)60994-0]
30. Bruton A, Lee A, Yardley L, et al. Physiotherapy breathing retraining for asthma: a randomised controlled trial. Lancet Respir Med 2018 Jan;6(1):19-28. [doi: 10.1016/S2213-2600(17)30474-5] [Medline: 29248433]
31. Morton K, Dennison L, Bradbury K, et al. Qualitative process study to explore the perceived burdens and benefits of a digital intervention for self-managing high blood pressure in Primary Care in the UK. BMJ Open 2018 May 8;8(5):e020843. [doi: 10.1136/bmjopen-2017-020843] [Medline: 29739782]
32. Duncan E, O'Cathain A, Rousseau N, et al. Guidance for reporting intervention development studies in health research (GUIDED): an evidence-based consensus study. BMJ Open 2020 Apr 8;10(4):e033516. [doi: 10.1136/bmjopen-2019-033516] [Medline: 32273313]
33. Stavric V, Saywell NL, Kayes NM. Perceptions of a self-guided web-based exercise programme for shoulder pain after spinal cord injury: a qualitative study. Spinal Cord 2023 Apr;61(4):238-243. [doi: 10.1038/s41393-023-00877-3] [Medline: 36702921]
34. Morrison LG, Yardley L, Powell J, Michie S. What design features are used in effective e-health interventions? A review using techniques from Critical Interpretive Synthesis. Telemed J E Health 2012 Mar;18(2):137-144. [doi: 10.1089/tmj.2011.0062] [Medline: 22381060]
35. Morrison L, Moss‐Morris R, Michie S, Yardley L. Optimizing engagement with I nternet‐based health behaviour change interventions: comparison of self‐assessment with and without tailored feedback using a mixed methods approach. British J Health Psychol 2014 Nov;19(4):839-855 [FREE Full text] [doi: 10.1111/bjhp.12083]
36. Morrison D, Mair FS, Chaudhuri R, et al. Details of development of the resource for adults with asthma in the RAISIN (randomized trial of an asthma internet self-management intervention) study. BMC Med Inform Decis Mak 2015 Jul 28;15(1-16):57. [doi: 10.1186/s12911-015-0177-z] [Medline: 26215651]
37. Dijkstra A. Personalization/computer-tailoring in persuasive technology: tailoring ingredients target psychological processes. Presented at: Proceedings of the Personalization in Persuasive Technology Workshop, Persuasive Technology 2016; Apr 5, 2016 URL: https://research.rug.nl/en/publications/personalizationcomputer-tailoring-in-persuasive-technology-tailor/ [accessed 2025-12-01]
38. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
39. Kelders SM, Bohlmeijer ET, Van Gemert-Pijnen JE. Participants, usage, and use patterns of a web-based intervention for the prevention of depression within a randomized controlled trial. J Med Internet Res 2013 Aug 20;15(8):e172. [doi: 10.2196/jmir.2258] [Medline: 23963284]
40. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]

XSL•FO
RenderX

41. Horsch C, Lancee J, Beun RJ, Neerincx MA, Brinkman WP. Adherence to technology-mediated insomnia treatment: a meta-analysis, interviews, and focus groups. J Med Internet Res 2015 Sep 4;17(9):e214. [doi: 10.2196/jmir.4115] [Medline: 26341671]
42. Bossen D, Buskermolen M, Veenhof C, de Bakker D, Dekker J. Adherence to a web-based physical activity intervention for patients with knee and/or hip osteoarthritis: a mixed method study. J Med Internet Res 2013 Oct 16;15(10):e223. [doi: 10.2196/jmir.2742] [Medline: 24132044]
43. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24(28):486-500. [doi: 10.17705/1CAIS.02428]
44. McClure JB, Shortreed SM, Bogart A, et al. The effect of program design on engagement with an internet-based smoking intervention: randomized factorial trial. J Med Internet Res 2013 Mar 25;15(3):e69. [doi: 10.2196/jmir.2508] [Medline: 23529377]
45. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013 Apr 15;15(4):e32. [doi: 10.2196/jmir.2283] [Medline: 23587561]
46. Hurling R, Fairley BW, Dias MB. Internet-based exercise intervention systems: are more interactive designs better? Psychol Health 2006 Dec;21(6):757-772. [doi: 10.1080/14768320600603257]
47. Xu X, Griva K, Koh M, et al. Creating a smartphone app for caregivers of children with atopic dermatitis with caregivers, health care professionals, and digital health experts: participatory co-design. JMIR Mhealth Uhealth 2020 Oct 29;8(10):e16898. [doi: 10.2196/16898] [Medline: 33118949]
48. Willett M, Duda J, Fenton S, Gautrey C, Greig C, Rushton A. Effectiveness of behaviour change techniques in physiotherapy interventions to promote physical activity adherence in lower limb osteoarthritis patients: a systematic review. PLoS One 2019;14(7):e0219482. [doi: 10.1371/journal.pone.0219482] [Medline: 31291326]
49. Dugas M, Gao GG, Agarwal R. Unpacking mHealth interventions: a systematic review of behavior change techniques used in randomized controlled trials assessing mHealth effectiveness. Digit Health 2020;6:2055207620905411. [doi: 10.1177/2055207620905411] [Medline: 32128233]
50. Couper MP, Alexander GL, Zhang N, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010 Nov 18;12(4):e52. [doi: 10.2196/jmir.1430] [Medline: 21087922]
51. Figueiras MJ, Neto DD. Challenges in “tailoring” adjustment: new ways of improving the response to chronic conditions. Eur Psychol 2019;24(1):1-6. [doi: 10.1027/1016-9040/a000348]
52. Lin H, Wu X. Intervention strategies for improving patient adherence to follow-up in the era of mobile information technology: a systematic review and meta-analysis. PLoS One 2014;9(8):e104266. [doi: 10.1371/journal.pone.0104266]
53. Alahäivälä T, Oinas-Kukkonen H. Understanding persuasion contexts in health gamification: A systematic analysis of gamified health behavior change support systems literature. Int J Med Inform 2016 Dec;96:62-70. [doi: 10.1016/j.ijmedinf.2016.02.006]
54. Glasgow RE, Christiansen SM, Kurz D, et al. Engagement in a diabetes self-management website: usage patterns and generalizability of program use. J Med Internet Res 2011 Jan 25;13(1):e9. [doi: 10.2196/jmir.1391] [Medline: 21371992]
55. Haines-Saah RJ, Kelly MT, Oliffe JL, Bottorff JL. Picture Me Smokefree: a qualitative study using social media and digital photography to engage young adults in tobacco reduction and cessation. J Med Internet Res 2015 Jan 26;17(1):e27. [doi: 10.2196/jmir.4061] [Medline: 25624064]
56. van Genugten L, Dusseldorp E, Webb TL, van Empelen P. Which combinations of techniques and modes of delivery in internet-based interventions effectively change health behavior? A meta-analysis. J Med Internet Res 2016 Jun 7;18(6):e155. [doi: 10.2196/jmir.4218] [Medline: 27268104]
57. Davies C, Corry K, Van Itallie A, Vandelanotte C, Caperchione C, Mummery WK. Prospective associations between intervention components and website engagement in a publicly available physical activity website: the case of 10,000 Steps Australia. J Med Internet Res 2012 Jan 11;14(1):e4. [doi: 10.2196/jmir.1792] [Medline: 22260810]
58. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
59. Khadjesari Z, Murray E, Kalaitzaki E, et al. Impact and costs of incentives to reduce attrition in online trials: two randomized controlled trials. J Med Internet Res 2011 Mar 2;13(1):e26. [doi: 10.2196/jmir.1523] [Medline: 21371988]
60. Schubart JR, Stuckey HL, Ganeshamoorthy A, Sciamanna CN. Chronic health conditions and internet behavioral interventions: a review of factors to enhance user engagement. Comput Inform Nurs 2011 Feb;29(2):81-92. [doi: 10.1097/NCN.0b013e3182065eed] [Medline: 21164337]
61. Meade LB, Bearne LM, Sweeney LH, Alageel SH, Godfrey EL. Behaviour change techniques associated with adherence to prescribed exercise in patients with persistent musculoskeletal pain: systematic review. Br J Health Psychol 2019 Feb;24(1):10-30. [doi: 10.1111/bjhp.12324] [Medline: 29911311]
62. Heron N, O’Connor SR, Kee F, et al. Development of a digital lifestyle modification intervention for use after transient ischaemic attack or minor stroke: a person-based approach. Int J Environ Res Public Health 2021 May 2;18(9):4861. [doi: 10.3390/ijerph18094861] [Medline: 34063298]

XSL•FO
RenderX

63. Band R, Bradbury K, Morton K, et al. Intervention planning for a digital intervention for self-management of hypertension: a theory-, evidence- and person-based approach. Implement Sci 2017 Feb 23;12(1):25. [doi: 10.1186/s13012-017-0553-4] [Medline: 28231840]
64. McEwan D, Beauchamp MR, Kouvousis C, Ray CM, Wyrough A, Rhodes RE. Examining the active ingredients of physical activity interventions underpinned by theory versus no stated theory: a meta-analysis. Health Psychol Rev 2019 Mar;13(1):1-17. [doi: 10.1080/17437199.2018.1547120] [Medline: 30412685]
65. Nielsen J. Iterative user-interface design. Computer (Long Beach Calif) 1993;26(11):32-41. [doi: 10.1109/2.241424]
66. Remskar M, Atkinson MJ, Marks E, Ainsworth B. Understanding university student priorities for mental health and well-being support: a mixed-methods exploration using the person-based approach. Stress Health 2022 Oct;38(4):776-789. [doi: 10.1002/smi.3133] [Medline: 35137525]
67. Marcu G, Ondersma SJ, Spiller AN, Broderick BM, Kadri R, Buis LR. Barriers and considerations in the design and implementation of digital behavioral interventions: qualitative analysis. J Med Internet Res 2022 Mar 30;24(3):e34301. [doi: 10.2196/34301] [Medline: 35353043]
68. Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021(374):n2061. [doi: 10.1136/bmj.n2061]
69. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. Int J Nurs Stud 2013 May;50(5):587-592. [doi: 10.1016/j.ijnurstu.2012.09.010] [Medline: 23159157]
70. Campbell M, Katikireddi SV, Sowden A, Thomson H. Lack of transparency in reporting narrative synthesis of quantitative data: a methodological assessment of systematic reviews. J Clin Epidemiol 2019 Jan;105:1-9. [doi: 10.1016/j.jclinepi.2018.08.019] [Medline: 30196129]
71. López-Nicolás R, López-López JA, Rubio-Aparicio M, Sánchez-Meca J. A meta-review of transparency and reproducibility-related reporting practices in published meta-analyses on clinical psychological interventions (2000-2020). Behav Res Methods 2022 Feb;54(1):334-349. [doi: 10.3758/s13428-021-01644-z] [Medline: 34173943]
72. Wells M, Williams B, Treweek S, Coyle J, Taylor J. Intervention description is not enough: evidence from an in-depth multiple case study on the untold role and impact of context in randomised controlled trials of seven complex interventions. Trials 2012 Jun 28;13(95):95. [doi: 10.1186/1745-6215-13-95] [Medline: 22742939]
73. Reynolds J, DiLiberto D, Mangham-Jefferies L, et al. The practice of "doing" evaluation: lessons learned from nine complex intervention trials in action. Implement Sci 2014 Jun 17;9(75):75. [doi: 10.1186/1748-5908-9-75] [Medline: 24935096]
74. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop. J Med Internet Res 2017 Jun 29;19(6):e232. [doi: 10.2196/jmir.7126] [Medline: 28663162]
75. DiLiberto DD, Staedke SG, Nankya F, et al. Behind the scenes of the PRIME intervention: designing a complex intervention to improve malaria care at public health centres in Uganda. Glob Health Action 2015;8:29067. [doi: 10.3402/gha.v8.29067] [Medline: 26498744]
76. Holtdirk F, Mehnert A, Weiss M, et al. Results of the Optimune trial: a randomized controlled trial evaluating a novel Internet intervention for breast cancer survivors. PLoS One 2021;16(5):e0251276. [doi: 10.1371/journal.pone.0251276] [Medline: 33961667]
77. Kelechi TJ, Madisetti M, Prentice M, Mueller M. FOOTFIT physical activity mHealth intervention for minimally ambulatory individuals with venous leg ulcers. J Wound Ostomy Continence Nurs 2020;47(2):173-181. [doi: 10.1097/WON.0000000000000631]
78. Kwon H, Lee S, Jung EJ, et al. An mHealth management platform for patients with chronic obstructive pulmonary disease (efil breath): randomized controlled trial. JMIR Mhealth Uhealth 2018 Aug 24;6(8):e10502. [doi: 10.2196/10502] [Medline: 30143475]
79. Lee MK, Yun YH, Park HA, Lee ES, Jung KH, Noh DY. A web-based self-management exercise and diet intervention for breast cancer survivors: pilot randomized controlled trial. Int J Nurs Stud 2014 Dec;51(12):1557-1567. [doi: 10.1016/j.ijnurstu.2014.04.012] [Medline: 24856854]
80. Nasseri NN, Ghezelbash E, Zhai Y, et al. Feasibility of a smartphone app to enhance physical activity in progressive MS: a pilot randomized controlled pilot trial over three months. PeerJ 2020;8(6):e9303. [doi: 10.7717/peerj.9303] [Medline: 32612882]
81. Yuan SLK, Couto LA, Marques AP. Effects of a six-week mobile app versus paper book intervention on quality of life, symptoms, and self-care in patients with fibromyalgia: a randomized parallel trial. Braz J Phys Ther 2021;25(4):428-436. [doi: 10.1016/j.bjpt.2020.10.003] [Medline: 33248904]
82. Onyura B, Mullins H, Hamza DM. Five ways to get a grip on the shortcomings of logic models in program evaluation. Can Med Educ J 2021 Dec;12(6):96-99. [doi: 10.36834/cmej.71966] [Medline: 35003436]
83. Goldman KD, Schmalz KJ. Logic models: the picture worth ten thousand words. Health Promot Pract 2006 Jan;7(1):8-12. [doi: 10.1177/1524839905283230]
84. Lee PL. What's wrong with logic models. LCSA. URL: https://clearimpact.com/wp-content/uploads/2013/03/Whats-wrong-with-logic-models.pdf [accessed 2025-11-15]

XSL•FO
RenderX

Abbreviations

PBA: person-based approach
pwSCI: people living with spinal cord injury
SCI: spinal cord injury
SPIN: Shoulder Pain Intervention delivered over the interNet

Edited by L Buis; submitted 19.Sep.2024; peer-reviewed by C Gauthier, R Martin; revised version received 25.Sep.2025; accepted 09.Oct.2025; published 05.Jan.2026.

Please cite as:
Stavric V, Saywell NL, Kayes NM
Designing a Self-Guided Digital Intervention for Self-Management of Shoulder Pain in People Living With Spinal Cord Injury: Tutorial on Using a Person-Based Approach
JMIR Mhealth Uhealth 2026;14:e66678
URL: https://mhealth.jmir.org/2026/1/e66678
doi:10.2196/66678

© Verna Stavric, Nicola L Saywell, Nicola M Kayes. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 5.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Intensive Longitudinal Data Collection Methods in Health Research: Tutorial for Selecting Measures and Designing a Sampling Protocol

Jimikaye Courtney[1*], MS, PhD; Genevieve Dunton[2*], PhD

[1]Department of Exercise and Sport Science, College of Arts and Sciences, University of North Carolina at Chapel Hill, 124A Fetzer Hall, Chapel Hill, NC, United States

[2]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, United States

[*]all authors contributed equally

Corresponding Author:
Jimikaye Courtney, MS, PhD
Department of Exercise and Sport Science, College of Arts and Sciences, University of North Carolina at Chapel Hill, 124A Fetzer Hall, Chapel Hill, NC, United States

Abstract

Intensive longitudinal data (ILD) include frequent and dense repeated measures captured over acute timescales (eg, every second, hour, or day) that are used to investigate within-person processes both within and across days. ILD are collected via wearable sensor data, ecological momentary assessments, or daily diaries and provide unique insights into within-person processes under ecologically valid conditions that can strengthen understanding of temporal relationships among variables and potential causal processes, while also informing the development of just-in-time adaptive interventions. Advancements in mobile and sensor technology have facilitated an explosion of ILD studies that have outpaced formal training in ILD study design. When designing ILD studies, researchers need to make careful decisions about the frequency (ie, how often) and timing (ie, when) of measurements. Decisions about the frequency and timing of measurement are influenced by issues such as variability across individuals and constructs, the purpose of the assessment, concerns about recall biases, saliency, or missing information, and participant needs. The interpretation of study results, causal inferences, and the predictive value of ILD are also impacted by decisions related to the timing of assessments, temporal lags between measures, how a “day” is defined, and data aggregation choices. Due to the increased interest in and adoption of ILD studies, and a lack of formal training, researchers can benefit from guidance on how to design ILD studies. Therefore, this paper aims to provide practical guidance to researchers on how to plan ILD studies using a step-by-step decision-making tutorial with applied health behavior research examples examining phenomena (eg, physical activity and alcohol use) that vary over acute time scales.

(JMIR Mhealth Uhealth 2026;14:e81290) doi:10.2196/81290

KEYWORDS

intensive longitudinal data; ecological momentary assessment; daily diaries; study design; wearable electronic devices

Introduction

Advancements in mobile and sensor technologies have facilitated an explosion of real-time data capture strategies in health behavior research to collect intensive longitudinal data (ILD) of time-varying phenomena [1-5]. ILD include frequent and dense repeated measures captured across acute timescales (eg, every second, hour, or day) [6,7]. ILD are collected via multiple methods, including wearable sensors, ambulatory assessments, experience sampling methods (ESMs), ecological momentary assessments (EMAs), and daily diaries [6,8-11]. These advanced methodologies reflect the increased interest and need for strategies to capture health behavior phenomena that change in intensity or direction across time scales as short as minutes, hours, or days. For example, engagement in health behaviors that impact chronic disease risk, such as unhealthy food intake, alcohol and tobacco use, physical inactivity, and lack of sleep [12,13], varies from day to day. Similarly, psychological and contextual factors that influence health behavior engagement, such as affect, intentions, and social context, can fluctuate rapidly over time. As a result, studies in health behavior and prevention research are increasingly using strategies to capture these types of psychological and contextual factors and behavioral outcomes. Such studies often aim to describe these phenomena, but may also inform the design and implementation of just-in-time adaptive interventions (JITAIs) [14,15].

When using ILD to examine health behavior phenomena that vary across short time scales, researchers often ask questions about temporal dynamics (eg, how variables change over time) to inform intervention and treatment approaches. First, researchers may ask whether these phenomena change as a

XSL•FO
RenderX

function of time itself, such as increasing linearly during treatment or demonstrating a curvilinear pattern across the day. Second, researchers may examine within-person effects of time-varying predictors, such as whether greater perceived stress compared with their usual stress level leads to greater consumption of saturated fat in the next few hours. Third, researchers may investigate the extent to which these within-person effects of time-varying predictors change over time or across situations, such as whether a person's positive affective response during physical activity depends on being alone or with others.

Up until recently, methods to capture the dynamic nature of how these time-varying phenomena change over time were not available. Previously, research on health behavior phenomena typically relied upon cross-sectional, longitudinal, or experimental designs where exposures and outcomes were assessed infrequently (eg, monthly or yearly) across few occasions. These studies used retrospective measures to capture an individual's usual level of a behavior or indicator such as average daily minutes of physical activity or past month frequency of depressive symptoms. Such approaches are not only vulnerable to recall biases, but they may lack ecological validity because they are not captured in real-world contexts, and they are not conducive to assessing phenomena that vary frequently over time or space or exhibit dynamic patterns of change and fluctuation [16].

Advancements in methods for collecting ILD have overcome many limitations of previous approaches, enabling researchers to capture time-varying phenomena in people's daily lives. ILD data collection methods, which collect data through wearable sensors, EMAs, and daily diaries, among others, differ from previous methods in how frequently variables are measured and participants' degree of active engagement with data collection. Wearable sensors can passively collect continuous or semicontinuous ILD every minute or several times per minute about time-varying phenomena such as heart rate, physical activity (eg, acceleration and step counts), location, or alcohol use, and require minimal participant engagement with data collection [17-25]. For example, participants may wear a physical activity sensor while awake, with their engagement limited to removing the sensor while sleeping or charging [21,23,24].

Another common approach to collecting ILD is through ESMs and EMAs (referred to throughout as "EMA"), which require active participant input through multiple self-reports per day of time-varying psychosocial or contextual variables, such as a participant's emotional states, cognitions, or social or physical contexts [2,4,6,8,10,11,16,26-33]. EMAs can capture time-varying measures before, during, or after specific events, such as during a physical activity bout or while drinking alcohol [26,28]. EMAs may be initiated by researchers (eg, signal- or interval-contingent EMA) or by participants' behavior (eg, event-contingent EMA), with the goal of capturing time-varying, ecologically valid information about participants' real-world experiences [4,6,10,11,26,33].

Daily diaries are also an active data collection strategy that includes a single self-report per day of variables, behaviors, or events that are expected to vary from day to day, rather than within a day [6,9,32,34,35]. Daily diaries may ask about a participant's stress experiences or physical activity or alcohol use behaviors on a given day, as these typically occur only once per day [9,34-36]. Daily diaries may include morning reports of the participant's previous night's experiences (eg, sleep quality) or evening reports of the participant's current day experiences (eg, soda intake) [37,38]. Like EMA, daily diaries also aim to capture participants' real-world experiences in near real time.

A Decision-Making Tutorial for Designing a Study Using ILD

These advancements in mobile and sensor technology have outpaced formal training in ILD study design. A recent concept paper partially addressed this need by developing a list of ten topics specific to the design and implementation of experience sampling studies, providing guidance on how to select ESM items, ESM sampling schemes, and practical considerations for real-world implementation of ESM studies [39]. A recently published special issue on ILD designs includes a series of articles addressing ILD considerations related to statistical power, sampling designs, participant experiences, and psychometrics or data quality [40]. The special issue dives deeply into specific features of ILD, providing practical guidance for researchers who are experienced in designing and implementing ILD studies [40]. However, there is a need for an ILD tutorial developed for researchers who are new to or have limited experience with ILD studies, such as graduate students and early career researchers. Additionally, guidance on ILD study design that simultaneously addresses the use of sensors, EMA, and daily diary methods is warranted, as many ILD studies rely on a combination of these methods to answer health behavior research questions. When designing ILD studies, researchers need to make careful decisions about the method used to capture health-behavior phenomena, as well as the frequency (ie, how often) and timing (ie, when) of measurements. These decisions should be based on the a priori research question and consider the characteristics and preferences of the target population. To address these needs for training and guidance, we present a decision-making tutorial for designing ILD studies. Multimedia Appendix 1 provides a detailed, step-by-step guide for navigating the decision-making process.

The tutorial is informed by a synthesis of the authors' applied experience in designing, implementing, and analyzing ILD studies, their didactic and consulting work with researchers new to ILD study designs, and relevant empirical and methodological literature. Recommendations were developed using an iterative, practice-oriented process that integrates published guidance with common challenges encountered in applied research. The recommendations and examples reflect the authors' applied research backgrounds, which have shaped the selection of illustrative health behaviors, populations, and design considerations emphasized in the tutorial. As such, the tutorial may not represent the full range of possible ILD applications (eg, clinical or patient populations [41-44]). The tutorial is

XSL•FO
RenderX

intended for beginning- to intermediate-level researchers who are interested in examining health behavior phenomena that vary over acute time scales such as minutes, hours, and days. The tutorial provides a step-by-step approach to 10 decisions that need to be made when selecting measures and developing a sampling protocol to collect ILD. The tutorial poses these decisions (in the form of questions) about the nature and timing of the variables and associations involved in the research of interest. Depending on the answers to those questions, the tutorial provides specific study design recommendations that are based on answering the research question of interest. The end of the tutorial provides data processing and analysis recommendations that can facilitate answering the research question based on the study design but should not dictate study design decisions. Although this tutorial and the applied examples emphasize health behavior research, the decisions and methods apply to a wide range of psychological, behavioral, and biological phenomena that may vary dynamically within persons. To facilitate readers' understanding of how to use the tutorial for investigating health behavior research questions, the design decisions are described with 2 working examples based on the authors' applied research experiences that represent some of the common research questions about temporal dynamics described above. The dependent variables (DVs) in the examples are health behaviors that vary over acute timescales (minutes, hours, or days) and are commonly investigated in ILD studies – physical activity and alcohol use [45-52]. The independent variables (IVs), momentary affective states, and daily drinking intentions, similarly vary across minutes or days, respectively, and are associated with physical activity and alcohol use behaviors [46-48,50,53-55].

In this study, we propose the following two research questions:

1. In children, are momentary affective states associated with moderate-to-vigorous physical activity (MVPA) levels in subsequent 30 minutes?
2. Do the effects of drinking intentions on same-day drinking behaviors differ based on the experience of daily stressors?

Decision 1: Determine Whether the IV and DV Change Over Time Within People

The first step is for the researcher to define their primary research question, which will serve as the basis for subsequent design decisions (question 1 in Multimedia Appendix 1). There may be several research questions within a single study, and each question may require different measurement or sampling protocol considerations; therefore, for simplicity, the researcher should start with a primary research question for which decisions about measures and a sampling protocol are made. For secondary research questions, the researcher can repeat the steps in the tutorial and attempt to integrate the design recommendations when possible. Once the primary research question is articulated, the researcher should define their IV and DV with as much specificity as possible (questions 2 - 3). Then, the researcher should determine whether the IV and DV change over time within people (eg, whether these are time-varying variables that change within people). If either the IV or DV change over time within people, then the research question may be suitable for an ILD design, and the researcher may proceed with the next part of the tutorial.

Application to research examples: For the first research example, both affective states (DV) and MVPA (IV) change over time within people. Similarly, for the second research example, drinking intentions (IV), daily stressors (moderator), and drinking behaviors (DV) change over time within people, making both research questions suitable for an ILD design.

Decision 2: Identify the Timescale of Within-Person Changes

The second decision is to identify the timescale of within-person change for the primary DV (question 4), which may be short, such as minute-to-minute changes in heart rate [56], or long, such as changes in disease risk factors (eg, hemoglobin $A1_c$) across months or years [57]. ILD studies are most appropriate for examining IVs and DVs that change across minutes, hours, or days. This is reflected by the frequency of ILD measures, with sensors measuring variables several times per minute or hour (eg, heart rate, blood glucose, or bodily movement), EMAs measuring variables several times per day (eg, affect, cognition, or tobacco use), and daily diaries measuring variables once each day (eg, intentions or alcohol use) [58,59]. If the IV and/or DV are expected to show within-person change over timescales of a day or shorter, then an ILD design is likely an appropriate choice. If the IV and/or DV changes over a longer timescale (such as monthly), an ILD study may result in collecting data more frequently or intensely than necessary, which may be costly for researchers and unnecessarily burdensome for participants. Some researchers may choose to use an ILD study design when the DV changes over longer timescales (eg, weeks and months) if they are interested in examining within-person change or variability in the IV as a predictor. However, this type of ILD study design requires slightly different design considerations and statistical methods than traditional ILD studies focused on within-person change in the DV. Therefore, discussion of this type of ILD study is beyond the scope of this paper. For IVs or DVs expected to show within-person change from day to day, a daily diary (ie, once a day) design may be appropriate. For IVs or DVs expected to change over a few hours within a day, wearable sensors or EMA may be appropriate. Finally, for IVs or DVs expected to change over minutes or seconds, wearable sensors may be the best measurement method. If the timescale of within-person change is unknown, it may be more efficient to start with EMAs rather than daily diaries, because data can always be aggregated up to a longer timescale from the EMA measures (eg, day-to-day). In contrast, data cannot be broken down into timescales shorter than the original measure, meaning that starting with daily diaries may be less efficient for identifying the timescale of change when it is unknown.

Application to research examples: For the first research example, investigating associations between momentary affective states and MVPA in the subsequent 30 minutes in children, affective states (ie, a subjective feeling state that a person experiences at a given moment, including both how positive or negative the feeling is ["affective valence"] and how energized the person

XSL•FO
RenderX

feels ["affective arousal"]) change within-person over minutes or hours—suggesting that either wearable sensors (eg, affect derived from galvanic skin response or heart rate variability) or EMA (eg, self-reported affect) may be appropriate design choices. MVPA changes over minutes, suggesting that wearable sensors are an appropriate design choice. For the second research example, investigating whether daily stressors moderate the effects of drinking intentions on behaviors, daily stressors (moderator) change from day to day, and daily drinking intentions (IV) may change from day to day—suggesting that daily diaries may be appropriate to measure both variables [47,60-63]. Drinking behaviors (DV) change across minutes or hours, suggesting that sensors or EMAs may be appropriate. In the next design decision, we will select the specific ILD measurement approach to use.

Decision 3: Determine Whether There Are Valid and Reliable Sensor-Based, EMA, or Daily Diary Measures

The next decision is whether reliable and valid sensor, EMA, or daily diary measures exist to assess the IV and DV of interest. Researchers should first consider whether sensor measures (ie, portable or wearable devices that use mechanical, electronic, or digital hardware and software) are available (question 5). Sensor measures can passively collect objective data (eg, heart rate, skin temperature, blood sugar, blood alcohol, body movement, sun exposure, or geographic location) [17,19,21,22,25] and may generate less participant burden [64] because they do not require manual user input.

If valid and reliable sensor measures are available, then researchers should consider the extent to which using the sensor will result in participant reactivity (ie, when awareness of being measured leads to changes in phenomena of interest; Table 1; question 5a) [65]. To manage participant reactivity, researchers could reduce participant access to information or data generated by the sensor (eg, do not provide access to the accompanying smartphone app, put tape over the sensor's display).

XSL•FO
RenderX

Table . Key concepts explained.

Concept	Explanation
Within-person change over time	How much an individual person in a sample or population tends to change over a specific timescale (eg, hours, days, weeks) regarding a given behavior or value of a construct. In the case of ILD[a] studies, within-person change is typically expected to occur within minutes, hours, or days, making the construct appropriate to study via ILD.
Passive assessment	ILD assessments that consist of collecting continuous or semicontinuous measures of body movements, physiology, location, device use, etc, without active participant engagement, that are often considered "objective" measures and are collected via wearable sensors or smartphones.
Active assessment	ILD assessments that consist of collecting participant self-reported measures of psychological constructs (eg, stress, affect), cognitive constructs (eg, thought processes), or context (eg, social, environmental), that are often considered "subjective" measures and are collected via surveys completed via smartphones or computers once or multiple times per day.
Participant reactivity	When a participant's behavior or subjective responses alter because they know they're being studied. Also referred to as a measurement or assessment effect, participant reactivity can also occur because responding to assessments provides a form of self-monitoring or feedback that changes participants' subjective responses, such that their responses are more reflective of their reaction to the assessment rather than to true naturalistic experiences or changes.
Expectancy effects	A form of participant reactivity in which a participant anticipates an upcoming assessment and changes their behavior or context to provide the expected response or outcome, to appear more favorable to researchers or to provide researchers with an "ideal" response, or to improve their ability to respond to the assessment.
Participant burden	When the effort required by the participant exceeds their abilities or motivation, impacting participant compliance and data quality. In ILD studies, wearable sensor-related burden may be related to the requirement to remove or replace, charge, or sync a device. EMA[b] and daily diary-related burden may be due to the length of the study protocol, frequency or timing of assessments, length of assessments, or length or complexity of individual items.
Naturalistic processes	Processes, such as behaviors, experiences, or cognitions that occur in people's real-world environments and that would occur naturally whether or not the research study is taking place. ILD studies often attempt to collect data in people's real-world contexts to maximize the ecological validity of findings.
Recall bias	Type of response bias in which a participant fails to accurately remember a past event or experience or to leave out relevant details when reporting them. In ILD, recall bias is influenced by design features, such as the timeframe across which participants are asked to recall data in EMAs (eg, "currently," "past 10 min") or daily diaries (eg, "today," "since waking")
Availability heuristic	A type of mental shortcut participants use when responding to questions, in which they rely on immediate or easily accessible information to inform their responses, which can be influenced by the timeframe for recall or item wording in ILD studies.
Saliency bias	A type of response bias in which more personally relevant or more intense experiences are more likely to be recalled, which is related to the availability heuristic. In ILD studies, daily diary items may be particularly impacted by saliency biases because participants are asked to recall constructs across the entire day, such as stressful events.
Peak effect	A type of response bias in which the most intense experience (ie, "peak") disproportionately influences a participant's response. For example, an extreme moment of anger unduly influencing the report of negative affect.
End effect	A type of response bias in which the most recent experience in time (ie, "end") disproportionately influences a participant's response. For example, the last social interaction of the day influencing the participant's report of social connectedness.

XSL•FO
RenderX

Concept	Explanation
Sampling window	The timeframe a researcher is interested in investigating. In ILD studies, this may be the entire 24-hour period, waking hours, sleeping hours, or a discrete event, such as a physical activity bout or a drinking event.
Participant schedules	The typical timing of daily behaviors being investigated among participants, such as sleep and wake times, class times, or work start and stop times. In ILD, participant schedules may be similar or different between participants and may be consistent or inconsistent from day to day within participants, which influences how sampling windows are specified for a study.
Between-person differences	When the association between the IV[c] and DV[d] reflects a difference between individuals in a sample. For example, when examining the association between stress and physical activity, a between-person effect could be that people who experience higher stress on average compared to other people are less physically active.
Within-person differences	When the association between the IV and DV reflects a difference between moments or days in an individual participant. For example, when examining the association between stress and physical activity, a within-person effect could be when a person experiences more stress than usual when they are more physically active than their average amount.
Concurrent effects	These are associations in which the IV and DV are expected to cooccur in time, also referred to as synchronicity. For example, it may be hypothesized that a mother's stress level will concurrently impact their child's stress level.
Lagged effects	These are associations in which the IV is expected to impact a DV that occurs at a later point in time, also referred to as sequentiality. For example, it may be hypothesized that a mother's stress level will impact her parenting behavior later in the day, which may in turn influence a child's subsequent behaviors.
Observation	The lowest level "unit" of analysis that will be used in hypothesis testing, statistical analyses, or the interpretation of results. This may be the momentary assessment, a discrete occasion or event, a given timeframe (eg, every 2 h), or the entire day.
24-hour day or Calendar day	A traditional 24-hour day based on a calendar date that begins at midnight and ends at 11:59 PM.
Study-centric day	A 24-hour cycle that is based on a study design feature or phenomenon of interest. For example, a researcher interested in studying a person's alcohol use behaviors may choose to align the day start with a 9 AM morning assessment about past day drinking and end the day at 8:59 AM on day n+1 to ensure that all drinking behaviors are consequences are attributed to the correct functional day.
Person-centric day	A "day" that is based on an individual's sleep-wake cycle and does not follow a 24-hour cycle, with each day differing between and/or within participants. This may include entire days or person-centric waking or sleeping periods. For example, each person-centric day may start when they wake up on day n and end when they wake up on day n+1, such that some "days" may be longer or shorter than 24 hours.

XSL•FO
RenderX

Concept	Explanation
Measurement burst design	A type of intensive longitudinal design consisting of repeated clusters ("bursts") of high-frequency assessments separated by intervals in between the bursts without any assessment occurring. This design allows researchers to capture short-term dynamics within bursts while also examining longer-term change between bursts. It is particularly useful when the DV changes over longer timescales, when within-person variability in a time-varying variable changes over longer timescales, when associations between time-varying variables vary over longer time scales, when key variables reflect relatively rare or time-bound events (eg, weekend alcohol use or phases of the menstrual cycle), or when it is necessary to balance temporal resolution with participant burden.

[a]ILD: intensive longitudinal data.

[b]EMA: ecological momentary assessments.

[c]IV: independent variable.

[d]DV: dependent variable.

Next, researchers should ask whether the sensor is cost-effective and whether there are resources (including personnel and software) to support using it (question 5b-c). Some sensors cost hundreds of dollars, which can be prohibitive for studies with smaller budgets or those requiring large samples. Also, sensors may require substantial research personnel resources to provide instructions on use and troubleshoot technical problems [66]. They may also require technological expertise to program, fix, and process data [22,66].

Finally, researchers should consider participant burden imposed by the sensor. Participant burden occurs when the effort (time, physical, and attention) required exceeds the ability or motivation to engage with or use a measure, which impacts data quantity and quality (question 5d) [67]. When burden from a sensor is higher than participants can handle, researchers may choose to use a less burdensome self-report measure instead, such as EMA or a daily diary, or they may consider a design with less frequent assessments (eg, every other week). Characteristics of the target population are also related to participant burden. For example, people who are less comfortable with technology may find EMAs more burdensome than tech-savvy populations [68], although well-designed study protocol, participant training, and user-friendly smartphone apps can address issues of participant burden and enhance compliance [52,68,69].

Some constructs may be too subjective in nature (eg, perceptions, motivations, intentions, and cognitions) to be assessed through a sensor, requiring self-reports via EMAs or daily diaries instead. In this case, researchers should determine whether there are acceptable or widely used items and/or instruments to assess those constructs (question 6). Ideally, researchers will select measures that were developed and validated using within-person longitudinal data, as measures developed from between-person cross-sectional data may not have the same factor structure at the within-person level, making it unclear whether shifts in variables reflect true within-person changes or measurement error [70-73]. While within-person validated measures are not yet available for all constructs, in recent years, freely available repositories have emerged where researchers can share EMA tools, measures, and reliability and validity information such as the Experience Sampling Method item repository [74]. If a within-person validated measure is not available, researchers may adapt an existing between-person cross-sectional measure (using appropriate recall time frames), while acknowledging that the measure was not validated for within-person use. If feasible, researchers could simultaneously conduct a psychometric study evaluating the reliability, validity, internal consistency, and/or factor structure of the items adapted from the between-person instrument (eg, using multilevel factor analysis to explore within- and between-person measurement model differences [72,73,75]), contributing to the ongoing need for psychometrically sound EMA measures [70,71,76]. An important point to consider is whether variability captured by EMA or daily diary measures represents true changes in an underlying construct or instead is indicative of poor reliability of the measure [70,71,76]. In contrast, it is necessary to understand whether lack of variability in EMA or daily diary measures is an artifact of careless responding (ie, participant not using the full range of response options due to lack of interest or fatigue) or true stability in a construct. It is possible to investigate these concerns by determining whether changes in EMA and daily diary measures correspond to changes in monitor-based or biologically-based indicators captured across the same timeframe (eg, accelerometry, cortisol, heart rate, and galvanic skin response) [77]. Other strategies to limit measurement error are to use a larger number of items to assess a construct, ensure adequate sampling frequency, avoid unreliable measures, and use appropriate statistical analytic strategies (eg, multilevel modeling or latent variable modeling) to differentiate measurement error from true variability [78].

If researchers find an acceptable EMA measure, they should consider whether answering EMAs will result in participant reactivity to a sufficient degree that it could alter study outcomes (question 6a).

Researchers should also consider whether actively responding to a real-time EMA prompt interrupts a naturalistic process that a participant is experiencing such as a romantic encounter, conversation, physical activity session, or emotional moment (question 6b). If so, researchers could consider measuring the construct passively using a sensor or retrospectively via an end-of-day daily diary [34]. Researchers should additionally ask whether the naturalistic process impacts the participant's

XSL•FO
RenderX

ability to provide valid and/or reliable EMA responses (eg, being intoxicated while responding to questions about drinking alcohol; question 6c). Finally, similar to using a sensor, researchers should ask whether the study has the financial, personnel, and informational resources to support implementing EMAs. If not, a sensor or daily diary measures could be considered (question 6d).

Finally, if researchers are contemplating using daily diary measures, they should seek out reliable and valid items and instruments to use (question 7). In some cases, items originally designed for EMA can be easily adapted for a daily diary by adding, "Over the past day." However, in some circumstances, memory errors and biases including peak and end effects can be introduced by asking participants to retrospectively report constructs up to 24 hours later [79]. If reliable and valid daily diary instruments are available, then researchers should consider whether these measures are likely to be impacted by expectancy effects, which occur when a participant has knowledge that a survey is about to occur and changes their mood, location, or behavior in preparation for the survey.

Regardless of the method selected for an ILD study (ie, wearable sensor, EMA, or daily diary), researchers should carefully consider privacy and data security implications. Many commercially available smartphone apps and wearable sensors require participants to agree to user terms and privacy policies that are outside of researchers' control, which may allow data sharing with developers or third-party entities [80]. This can introduce risks related to confidentiality, data ownership, and secondary data use [80,81]. Additionally, mobile sensing and EMA approaches often involve the collection of highly sensitive, potentially identifiable data (eg, location or health information), increasing the risk of unintended disclosure or reidentification. Participants may also also be required to accept app-specific data use agreements, which the researcher has no control over [80-82]. Researchers should clearly communicate data access rights and potential confidentiality risks during the informed consent process [81].

Application to research examples: When studying associations between affective states and MVPA in the subsequent 30 minutes in children, wearable sensors are ideal to measure MVPA due to the timescale of change (seconds or minutes), the time frame under investigation (30 min), and minimal concern about participant burden and reactivity to sensors. The population of interest (children) also warrants the decision to use wearable sensors, as self-reports of past 30-minute MVPA via EMAs may be too cognitively burdensome for children to complete—resulting in unreliable data. To measure affective states, EMAs were selected due to the timescale of change (hours), minimal concern about participant reactivity and expectancy effects, and acceptable participant burden, and having resources to support personnel and software costs.

When investigating daily stressors moderating the effects of drinking intentions on drinking behaviors in adults, alcohol use can be measured via sensors, EMAs, or daily diaries. However, EMAs are subject to expectancy effects and biases due to alcohol consumption, and daily diaries are subject to recall biases and miss key features of drinking that often occur while people are sleeping (ie, descending limb of drinking or alcohol elimination). Therefore, passive wearable sensor measures were deemed most appropriate. In contrast, there are reliable and valid daily diary measures of stressful experiences that are not subject to expectancy effects and capture psychological stress experiences and responses, leading to the selection of daily diaries to measure stressors. Similarly, although drinking intentions could vary within a day, the research question was specifically investigating daily intentions. Additionally, drinking intentions are subjective, warranting the use of daily diaries.

Decision 4: Determine Timing and Frequency of Sampling (EMA Studies Only)

The next decision when using EMAs is determining when and how frequently to collect data. If asking participants about their "current" state or "the past 10 minutes," recall biases will be minimal. However, as the recall time frame increases, participants are more likely to use mental shortcuts to inform their responses, such as relying on the most easily accessible information (ie, availability heuristic) or being influenced by personally relevant or intense experiences (ie, saliency bias, peak and end effects) [83]. The wording of EMA items may also change participants' recall, how they interpret questions, or how they respond [83]. Researchers should aim to select an EMA sampling schedule that minimizes these types of biases. To do so, they need to identify the time frame across which participants can reasonably recall the construct, when the construct occurs throughout the day, whether it is susceptible to expectancy effects, and whether it occurs during specific events or in specific contexts. For example, if the construct occurs at the same time each day (question 8), such as a regular morning physical activity bout, or if it has low susceptibility to expectancy effects, then an interval-contingent design (ie, same time each day) with EMAs close to when the construct occurs may be appropriate. An interval-contingent design is also appropriate when a researcher is interested in measuring the level of a construct at the same time each day (question 8a). If the construct occurs at different times each day, such as a stressor, or if it has higher susceptibility to expectancy effects (question 9) [3,84], then a signal-contingent design (ie, random times throughout each day) is likely a better choice. Finally, if researchers are interested in measuring the construct during a specific event or behavior (question 10), such as measuring affect during a drinking event, then an event-contingent design is most appropriate. Selecting a particular type of EMA sampling schedule in turn influences EMA item wording, recall windows and, for signal-contingent designs, the frequency of EMA measures.

When designing an EMA sampling protocol, researchers should determine the frequency and length of the recall period. If using a signal-contingent design, then researchers should consider how frequently the construct changes across the day (question 11). Depending on the frequency (eg, every 2 - 4 h), researchers may consider sampling 2 to 16+ times per day for 7 to 14 days, for example. However, if more frequent sampling is necessary to capture natural variation in the construct (eg, every hour),

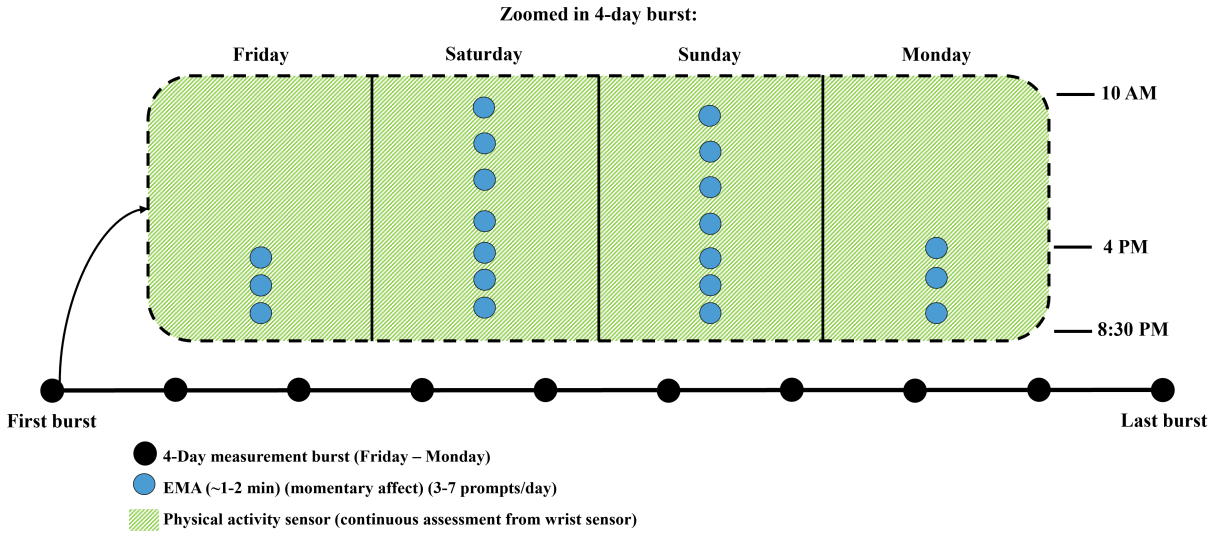
XSL•FO
RenderX

then researchers may need to consider a shorter study period (eg, 2 - 4 days) to minimize burden and maximize data quality [85]. For example, in the TIME (Temporal Influences on Movement and Exercise) study, a signal-contingent EMA protocol prompted once during every waking hour of the day for 4 days during each bimonthly measurement burst [86].

Researchers should also consider how long after a behavior or experience a participant can reliably recall the phenomena (question 12). If the construct becomes more vulnerable to recall biases over elapsed time (such as intensity of affective valence), the length of the EMA recall window should be short (eg, "Right now, how are you feeling?" or "In the past 15 min, how have you been feeling?") [48]. In contrast, some constructs such as concrete behavior, social encounters, or locations visited may be reliably reported across longer EMA recall windows (eg, "In the past 4 hours, have you spent any time with your child?") [87,88]. The ideal recall window may differ between participants. For example, children may find it more difficult than adults to report feelings across the past 15 minutes. It is important to note, however, that the EMA recall window should not be longer in duration than the spacing between adjacent EMA measures of the same construct to avoid overlapping or double assessments. For example, if EMA surveys are prompted every 2 hours, then the recall window should not be longer than 2 hours.

Application to research examples: For investigating affective feeling states in children, we selected EMAs as the most appropriate measure. The EMA sampling design needed to minimize burden on children and account for times of day when they cannot or should not receive prompts. Therefore, a hybrid design with interval and signal-contingent EMAs was selected (Figure 1). During each 2-hour window (interval-contingent component), an EMA survey was prompted at a random time (signal-contingent component). The recall time frame was momentary, asking children about their affect at the given moment (eg, "How happy do you feel right now?"), to avoid issues of recall bias, such as peak and end effects that could occur if asking children about their daily affect at the end of the day.

Figure 1. Decision 4: Example intensive longitudinal data protocol using a hybrid interval-signal-contingent ecological momentary assessment design with continuous sensor measures of physical activity via multiple assessment bursts with intensive sampling across four days to balance reliability and power while minimizing participant burden. Random prompts sent between 4 PM-8:30 PM on Friday and Monday and 10AM -8:30 PM on Saturday and Sunday. EMA: ecological momentary assessments; ILD: intensive longitudinal data.

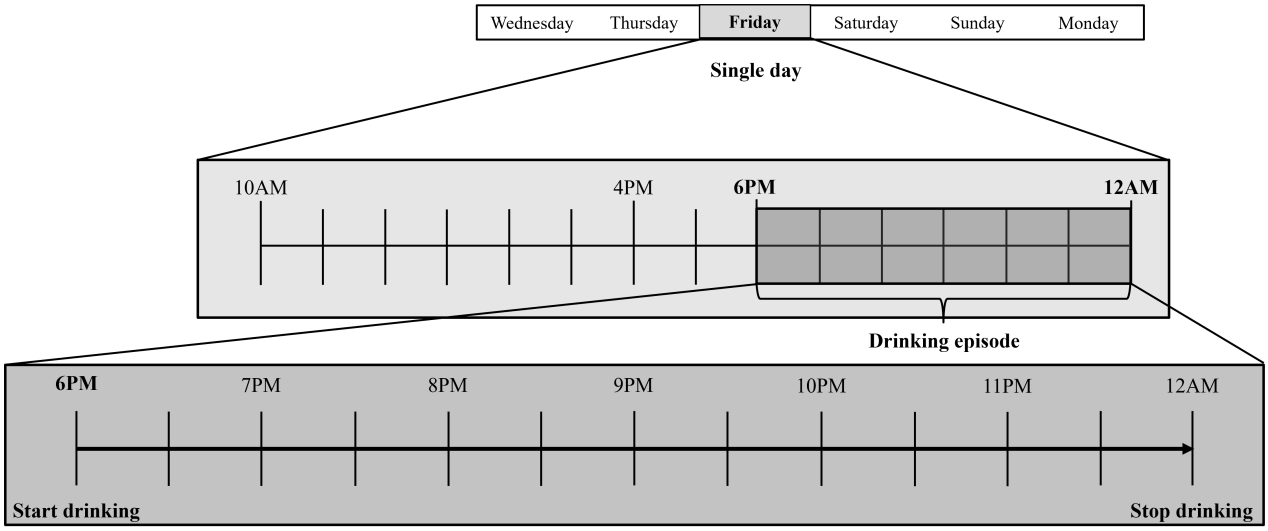


Decision 5: Select Sampling Windows and Time(s) of Day When Assessments Occur (Sensor and EMA Studies Only)

The next decisions involve selecting the sampling windows and time(s) of day that assessments should occur (question 13). For example, a researcher may only be interested in social interactions when the participant is at work or screen time in 2 hours leading up to bedtime. In these situations, collecting data outside of these sampling windows may be unnecessarily costly and burdensome. If a researcher is interested in a specific event, such as a high-intensity drinking episode, then they should use an event-contingent design with participants completing EMAs and wearing an alcohol sensor during the drinking episode [46,49] (Figure 2). Additionally, if the research question is about eating behaviors, the researcher may choose to use an interval-contingent design that triggers EMAs 3 times per day immediately after known mealtimes.

XSL•FO
RenderX

Figure 2. Decision 4: Example Intensive longitudinal data (ILD) protocol using event-contingent ecological momentary assessments (EMAs) during drinking episodes (Courtney and Russell [47]).



Next, researchers need to address whether to use the same EMA sampling window for all participants and/or for each day of the week. If participants' sleep-wake schedules differ, such as for shift versus nonshift workers [89], then each participant may need a personalized EMA daily sampling window [87]. Researchers also need to determine whether a given participant or all participants have a similar sleep-wake schedule for each day of the week (question 15). If participants have very different schedules, such as going to sleep later and waking up later on weekend days, then sampling windows should be tailored for each day of the week or weekends versus weekdays. Real-time customization of prompting schedules is also possible, such that each day participants are asked when they anticipate going to sleep that night and waking up the next day [86]. Customized schedules can increase compliance by reducing burden and unpleasant sleep interruptions; however, they may be impractical, too costly to program, or create data processing challenges. If customization is too costly or burdensome, then the researchers may want to consider screening out participants with unusual sleep-wake schedules (eg, screen out shift workers), using a sampling schedule that does not include early morning or late evening assessments, or using only event-contingent EMAs.

When making design decisions about sampling windows and schedules, considering the needs and preferences of the target population can help reduce participant burden and missingness. In some cases, surveying members of the target population before making design decisions can help researchers identify ideal sampling windows or the need for tailored or real-time customized schedules. For example, the ideal time of day for asking participants to complete morning diaries of past-day behaviors is likely to differ between college students versus older adult populations. College students may require a later sampling window time (eg, 10 AM-2 PM) whereas the older adults may require an early sampling window time (eg, 7 AM-11 AM).

Application to research examples: For studying affect and subsequent 30-minute MVPA in children, EMAs were triggered on weekday afternoons and evenings and throughout the day on weekend days to avoid interrupting school and sleep time. Using the same schedule across all participants was appropriate given that school schedules are fixed (eg, classes end at 3 PM each day) and the sampling windows were narrow enough to avoid interrupting sleep time, thereby not requiring customized windows.

For investigating daily stressors and drinking intentions and behaviors, there were 2 daily diaries per day, with the sampling windows selected to go from 9 AM-11 AM and 8 PM-10 PM across all weekdays and weekend days. The 9 AM schedule for the morning diary ensured that intentions were reported before drinking behavior, as participants are unlikely to drink that early in the morning. The 8 PM schedule for the evening diary of stressful events avoided interrupting sleep time while reducing recall biases about the day's stressors. Due to the nature of alcohol metabolism, with alcohol elimination occurring for many hours, often while individuals are sleeping, the decision was made for participants to wear the sensor 24 hours/day. Similarly, inter- and intraindividual variability in timing of alcohol use across weekends and weekdays also warranted the use of continuous sensor wear. As the timing of daily diaries were selected with consideration of late-rising and early bedtimes, and since diaries were available for completion for 2 hours after each prompt, the decision was to use the same schedule for all participants.

Decision 6: Delineating Temporality of the Associations

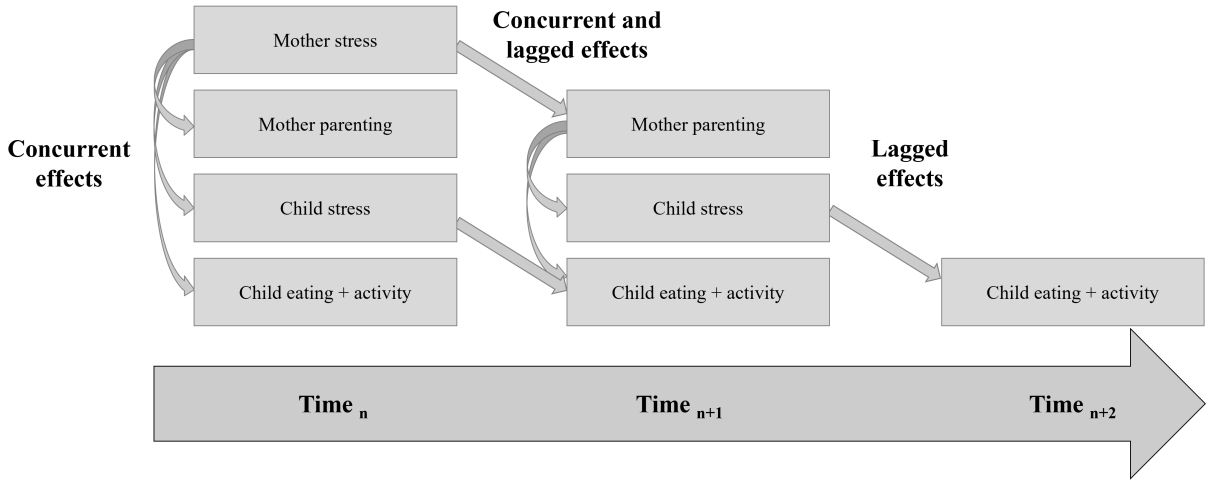
The next decision involves researchers specifying the nature of temporal associations that they are hypothesizing, which is a useful step toward better understanding potential causal processes. Whether the associations are thought to be primarily within- or between-persons and concurrent or lagged in time have design implications in terms of statistical power and the sampling timing. Researchers should first determine what type of association they are interested in testing (question 16). If they are examining whether within-person changes in the IV (eg,

XSL•FO
RenderX

positive affect assessed every 2 h through EMA) are associated with within-person changes in the DV (eg, MVPA min in 30-min window after the EMA), then the study should include enough observations nested within enough people to detect within-person effects [90,91].

If a researcher is examining whether within-person changes in the IV are associated with within-person changes in the DV, then researchers should first consider whether the IV is hypothesized to be associated with the DV at the same point in time (ie, cooccurring effect) or a later point in time (ie, lagged effect; question 17). For the cooccurring effect, the design would need to ensure that affect and physical activity are assessed concurrently (eg, participants respond to EMA while doing physical activity) whereas the latter hypotheses would require a design that captured physical activity after each EMA prompt and linking data through date-time stamps (Figure 3).

Figure 3. Decision 6: Example of concurrent and lagged effects for examining the causal associations (Redrawn from Dunton et al [87]).



Second, if the researcher hypothesizes that the IV is associated with the DV at the same point in time (ie, cooccurring effect), then they need to specify the "duration" of that point in time (question 18). For example, one could hypothesize they are associated with each other during the same 1-hour period (eg, types of social interactions over the past hour are associated with level of positive affect over the past hour). If using EMA, the item wording would need to specify the duration of the recall period (ie, "In the past hour…").

Third, if the researcher hypothesizes that the IV is associated with a DV occurring at a later point in time (ie, lagged effect), then it is necessary to specify how much later (eg, in minutes, hours, or days; question 19). For example, if a researcher is interested in understanding whether perceived stress at any point of the day (IV) is associated with asthma symptoms occurring over the next 4 hours (DV), then the sampling frequency and item wording need to reflect these temporal associations [30].

Application to research examples: The first research example explicitly asks whether momentary affect is associated with MVPA in the subsequent 30 minutes. This requires linking sensor data to EMA survey responses and then calculating time spent in MVPA during the subsequent 30 minutes (leading observations). In the second research example, the question is investigating within-person, daily associations between intentions and drinking behaviors, with daily stress as a moderator. The design includes morning diaries about drinking intentions and evening diaries about daily stressors, including reports of the time of day the stressor occurred. This type of design requires linking morning and evening diaries by study day and linking sensor data to the diary data. To ensure temporal precedence, time stamps for when morning diaries were completed and for when stressors occurred must be linked to the sensor data so that (1) only drinking data which occurs after the morning assessment of intentions are included and (2) only stressors that occurred after the morning assessment of intentions and before drinking onset are included. Multimedia Appendix 2 provides example R code using timestamps from the morning diary and stressor occurrences to establish temporal precedence, followed by removing all stressors that occurred before the morning diary.

Decision 7: Select Number and Type of Days to Assess

Once the temporal and causal associations are specified, the next decision is determining the number and types of days across which the ILD sampling protocol should be implemented. The key challenge is balancing the need for adequate representation across time and statistical power (more observations are typically better) with the potential response burden (fewer observations are typically better). If the research question concerns between-person differences and conceptualizes either the target IV or DV at person level (ie, their usual level of a behavior), researchers should consider how many days will provide adequate representation or between-person reliability (ie, internal consistency) for the target IV and DV (question 20), with reliability meaning the extent to which any given set of assessments approximates the typical level of the construct [92]. For constructs that are generally stable for a person across time, such as the time of the day that someone starts work, only a few days may be needed. However, for constructs that fluctuate from day to day, such as dietary intake, many more days may be needed to obtain adequate representation of one's

XSL•FO
RenderX

typical daily nutrient intake [93]. A different but associated concept is within-person (ie, idiographic) reliability, which addresses how reliably one can detect fluctuations within an individual over time (ie, time-varying states) beyond measurement error as described in decision 3 [78].

Related to the total number of days are questions about whether weekend and/or weekdays are needed (question 21) and the total number of observations per person (question 22). Some questions, such as those pertaining to children's social interactions at school, may only require weekday assessment whereas other questions, such as those related to overall sleep patterns, would require weeknight and weekend night assessments. Which days to sample for a given phenomenon may also differ depending on the target population. For example, among college students, drinking is more likely to occur on Thursday-Saturday, the "social weekend"; therefore, researchers studying college student drinking may only need to sample those days. However, if studying alcohol use in other adult populations or individuals with an alcohol use disorder, it may be appropriate to sample all days of the week. Finally, if a researcher is using EMA as the primary measurement modality with multiple assessments per day, then they need to consider how many observations per person are needed to ensure adequate statistical power. It is beyond the scope of this tutorial to provide in-depth guidance for power analyses in ILD studies; however, we will briefly state that multiple approaches can be used for estimating power in ILD (eg, simulation or summary-statistics based approaches) and, as in all power analyses, the sample size and number of observations needed depend on the primary effect researchers aim to detect (eg, within-person, between-person effects, and cross-level interactions) [94-97]. Due to the multilevel nature of ILD, power analyses should also account for the estimated intraclass correlation coefficient, which reflects the proportion of variance attributable to between-person differences and influences the amount of independent information contributed by repeated observations [94-97]. For example, a Monte Carlo simulation study found that, with 70 observations per person, a sample size of 15 participants yields 99% power to detect within-person (level 1) effects [94]. More participants were needed to detect between-person (level 2) effects; specifically, 30 participants with 70 observations yield over 80% power to detect between-person effects [94]. Given the complexity of power analyses for ILD, we recommend that readers consult a biostatistician and refer them to the PowerAnalysisIL Shiny app (Ginette Lafit) on GitHub to aid their research team in conducting power analyses for their ILD studies [94].

The next questions relate to balancing reliability and power requirements while minimizing participant burden. When participants experience greater burden due to the behavioral or cognitive effort involved in responding, the quality and/or quantity of data can be reduced [98,99]. If the number of days or assessments per day appear to be too burdensome (questions 23 and 24), then the researcher could consider altering the study design such as having assessments occur on random days with no-assessment days in between (eg, one day on followed by one day off) or multiple assessment bursts with a short string of continuous days (eg, 5 d) separated by multiday breaks (eg, 5 d; Table 1).

Finally, researchers should consider whether they expect the strength of within-person associations to change or remain stable over time (question 25). If they anticipate that a within-person association (eg, whether stress leads to increased hunger in the next hour) remains stable across weeks, months, and years, then they may choose to use a single measurement burst design (eg, a single period of continuous days). However, if they expect that the within-person association changes in strength over longer periods of time (eg, due to exposure to an intervention or historical events), then a multiple measurement burst design with bouts of continuous assessment days separated by the time interval of expected change (eg, weeks, months, or years) should be considered.

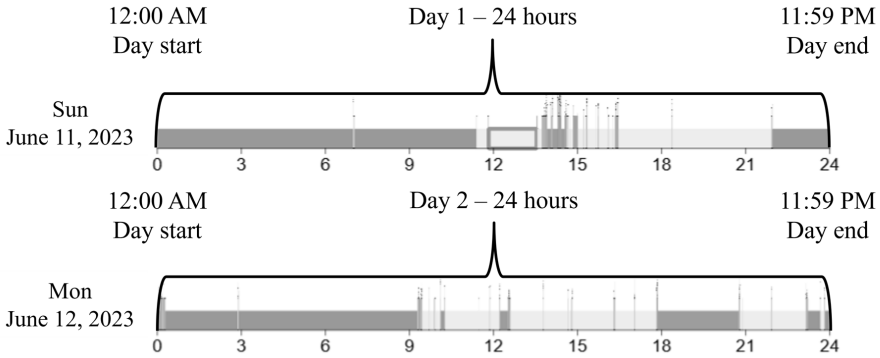
Application to research examples: For investigating affect and MVPA in children, the decision to only assess Fridays-Mondays necessitated a measurement burst design to collect sufficient observations. For example, 10 measurement bursts, separated by at least 14 no-assessment days, with 4 days of data collection during each burst permits collecting sufficient observations without undue burden (Figure 1). For investigating daily drinking intentions, stress, and behaviors, at least 30 sequential days of observation are warranted, as the research question is about day-level associations in adults.

Decision 8: Define What a "Day" Means in the Study

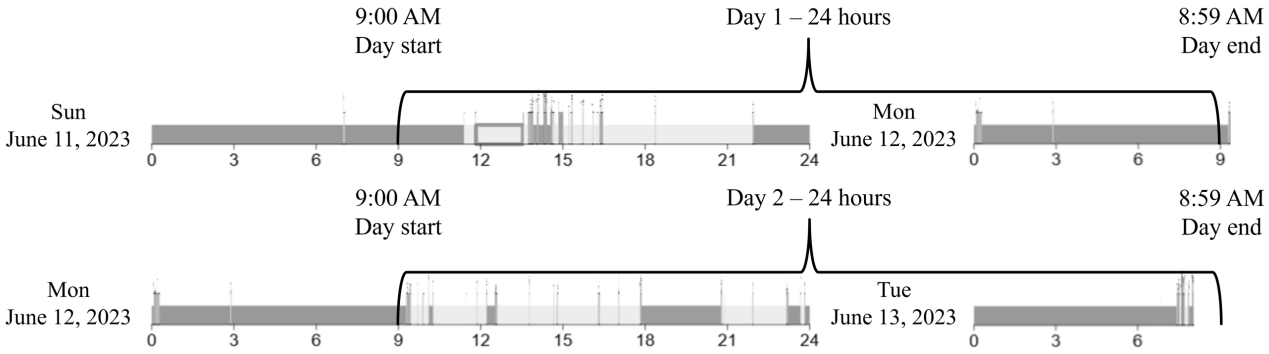
After selecting the number and types of days to assess, the next decision is determining how a "day" is defined. For example, if a researcher is examining participants' daytime behaviors, such as activities during work or school [48,89,100] or studying discrete events during the day, such as physical activity or drinking events [46,49,87], then defining the day based on a calendar day, which reflects the traditional 24-hour day or calendar date (eg, starting at midnight and ending at 11:59 PM) is appropriate (Figure 4). In general, a "calendar day" is appropriate when most participants go to bed before midnight or for studies examining daytime behaviors.

XSL•FO
RenderX

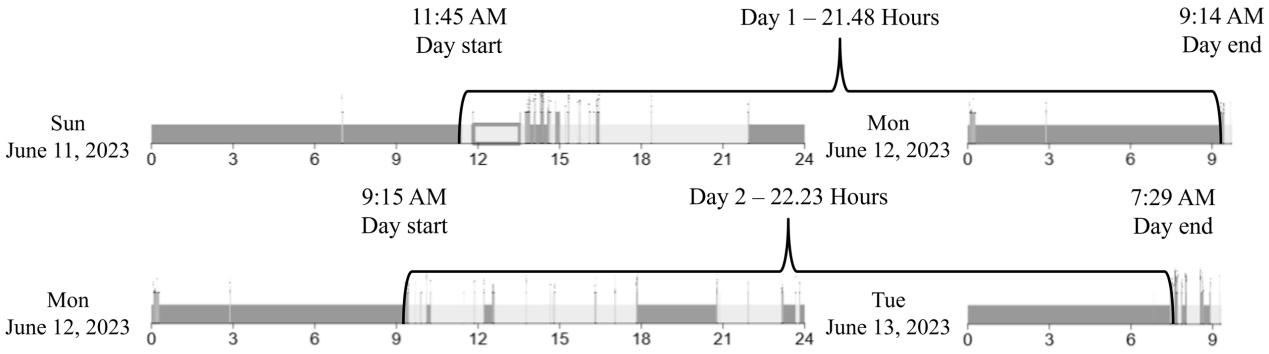
Figure 4. Decision 8: Scenarios for how to define a "day" might be defined for a given study. (A) Defining a day based on "calendar day," or a traditional 24-hour day/calendar date from midnight on day *n* to 11:59 PM on day *n*. (B) Defining a day based on a "study-centric" day with 24 hours from 9 AM on day *n* to 8:59 AM on day *n*+1. (C) Defining a day based on a "person-centric" day with each day starting upon waking on day *n* and continuing until waking on day *n*+1, resulting in a varying number of total hours (ie, <24 h or >24 h) per day.



A: defining a day based on a 'calendar day', or a traditional 24-hour day/calendar date from 12:00 AM on day *n* to 11:59 PM on day *n*.



B: defining a day based on a 'study-centric' day with 24-hours from 9:00 AM on day *n* to 8:59 AM on day *n* + 1.



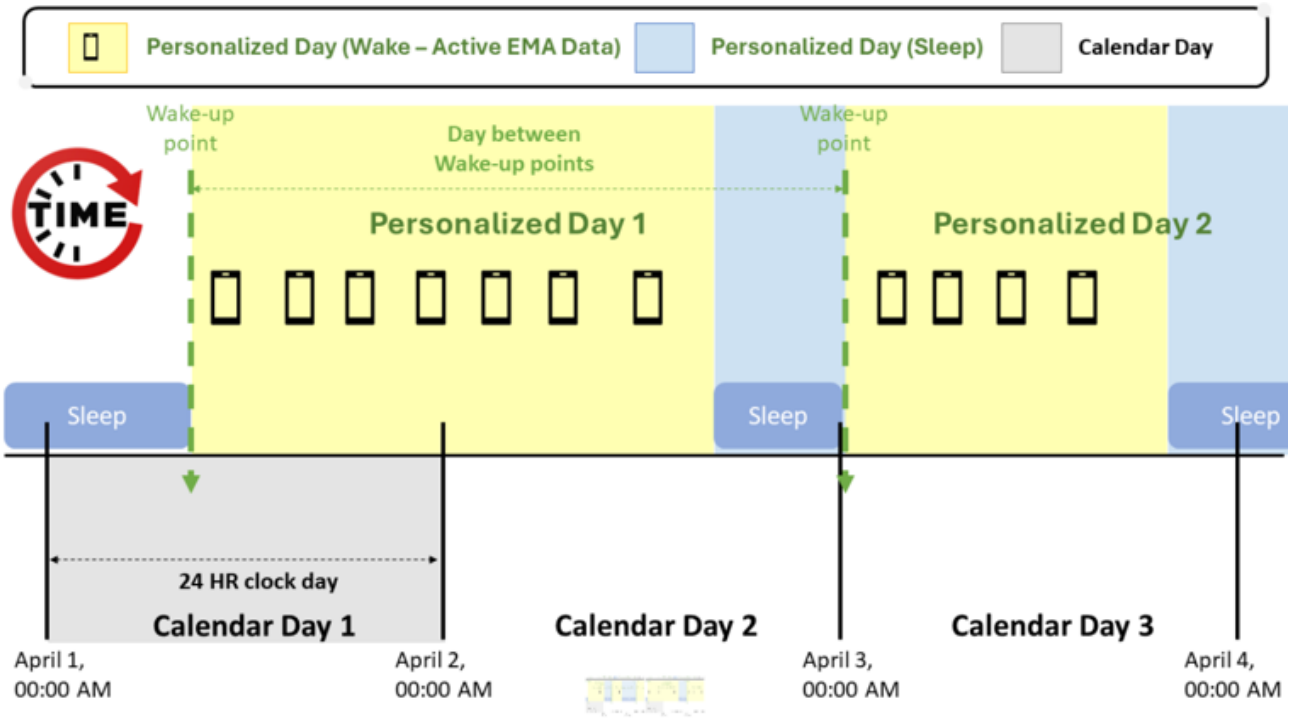
C: defining a day based on a 'person-centric' day with each day starting upon waking on day *n* and continuing until waking on day *n* + 1, resulting in a varying number of total hours (i.e., <24 hours or >24 hours) per day.

If a researcher is studying phenomena that are not contained within a calendar day, influence each other across the sleep-wake cycle, or have spillover effects across calendar days, then defining a "study-centric" day based on a key study feature, such as a morning EMA, is appropriate. For example, drinking events often occur after midnight and alcohol metabolism continues into the next calendar day. Using a study-centric day ensures these constructs are correctly attributed to the day when they practically occurred. For example, studies by Courtney and Russell [47] and Russell et al [51] used a 10 AM EMA as the daybreak when examining daily drinking intentions and outcomes.

Finally, if a researcher is investigating variability in a person's daily lived experiences or circadian rhythms, then a "person-centric" day based on an individual's sleep-wake cycle (eg, not following a 24-h cycle) is appropriate. With a person-centric day, each day may differ between and/or within participants, and a "day" may include entire days or person-centric waking or sleeping periods. For example, Wang et al [86] were interested in within-person predictors of physical activity engagement and maintenance; therefore, they used person-centric days based on a participant's sleep-wake cycle to examine these within-person processes (resulting in some "days" having more than 24 h and some less than 24 h; Figure 5). Researchers should decide a priori about how to define a "day" because it has important implications for sensor wear protocol, timing of measures, data processing, and results interpretation.

XSL•FO
RenderX

Figure 5. Decision 8: Example ecological momentary assessment sampling strategy when using a "person-centric" study day (reproduced from Wang et al [86]). EMA: ecological momentary assessments.



Application to research examples: In the first research example investigating affect and MVPA in children, the EMA sampling schedule was limited to 10 AM-8:30 PM; therefore, the design decision was to use a 24-hour calendar day. In the second research example investigating drinking intentions and behaviors, the design decision was to use a study-centric day. This was more appropriate than a 24-hour day because alcohol use often occurs after midnight and, even if a person is done drinking before midnight, alcohol metabolism continues throughout the sleeping period (ie, after midnight). By using a study-centric day, we avoid the problem of attributing sensor-detected alcohol use to the incorrect "lived" day, which would cause problems related to temporality and causality. The study-centric day was defined based on the start time of the morning daily diary, with the day running from 9 AM to 8:59 AM the following day. To further avoid issues of misattributing alcohol use to the incorrect data, it was also necessary to remove any residual alcohol detected after 9 AM when processing the sensor data.

Decision 9: Determine Whether to Aggregate Data and, if Applicable, Which Assessments to Aggregate

After making the previous design decisions, researchers need to make data processing decisions related to aggregating assessments and temporally aligning variables based on whether the hypothesized associations are concurrent or lagged. These decisions are critical, because variation in decisions regarding which variables to include in analyses, which items to cluster with one another (eg, calculate average positive affect from several items), the handling of missing data, and the use of contemporaneous versus lagged variables all may impact final model results [101]. If there are multiple assessments per day (eg, EMAs), then the researcher needs to decide the timeframe and duration of time they are interested in studying (questions 32a and 32b). If a researcher is interested in all observations of the IV that occurred before DV, then they would aggregate all of those IV assessments during the specific time frame (taking the total, average, or variance value). If a researcher is interested in a duration of time shorter than 1 day, then they need to specify whether concurrent or lagged effects are of interest (questions 17 - 19).

Data aggregation refers to combining multiple observations via summary measures, such as calculating average daily affect using multiple EMA responses within a day. There are two decisions underlying aggregation: (1) how to aggregate data (ie, which summary measures to use, such as mean, total, maximum, or variance), and (2) across which time frame (ie, measurement windows) to aggregate data (eg, hour or day). Both decisions can be informed by top-down (a priori) or bottom-up (data-driven) approaches. Top-down decisions may be preferable when guided by a specific research question. For example, if researchers are interested in the range of emotions children experience, they may aggregate using within-day variability (eg, variance) or the number of unique emotions reported, rather than mean levels of affect. Similarly, if researchers hypothesize that negative social interactions throughout the school day, but not after school, are most relevant for children's emotion regulation, a top-down approach would support aggregating observations specifically within the school-day window. To facilitate implementation, Multimedia Appendix 2 includes example code for leading and lagging variables, data aggregation (ie, day-level mean, within-day variability, or number of unique emotions), and multilevel models. This code is intended as a starting point rather than a prescriptive template; researchers should carefully adapt all steps to their specific study design, research question,

XSL•FO
RenderX

measurement structure, and analytic goals, particularly with respect to decisions about disaggregation (ie, within- vs between-person effects) and model specification (eg, inclusion of random slopes).

In contrast to top-down approaches, bottom-up, data-driven approaches (eg, machine learning methods) may be useful when there is limited theoretical guidance or when researchers aim to use empirical data for identifying optimal aggregation strategies. For example, researchers could compare multiple summary metrics (eg, mean, total, and variability) or time windows (eg, hourly vs daily) in terms of their predictive utility for an outcome, allowing the data to inform both the aggregation metric and time frame. Such an approach could reveal, for instance, that the total number of negative social interactions experienced in the morning has the largest impact on children's emotion regulation.

Application to research examples: Research example 1 requires aggregating accelerometer data from the 30 minutes before and after the EMA prompts. Research example 2 requires aggregating alcohol sensor data across the study-centric day (9 AM-8:59 AM) so that only alcohol use after morning intentions is included in the models investigating day-level within-person associations.

Limitations and Future Directions

Overall, this tutorial is meant to be a practical, step-by-step, decision-making tool to help beginning-to-intermediate ILD researchers design an ILD study. As such, it does not include advanced considerations needed for more complex research questions. The tutorial also does not address sophisticated ILD study designs that are needed to test time-varying effects (eg, within-person effects that vary across the day), time-varying moderating effects (eg, within-person effects that vary across contexts), multilevel mediation, idiographic effects, or N of 1 studies [102-105]. It also does not address research questions that examine within-person variability as a predictor, mediating, or moderating variable [106]. For example, a researcher may be interested in understanding how mood variability predicts changes in physical activity levels across childhood. The tutorial does not address some novel ILD measurement selection strategies, such as using computerized classification based on measurement error or mean and variability in previous EMA item responses to determine when and how often those items should be assessed [107,108]. Furthermore, the overall purpose of the tutorial is to inform the development of assessment protocols for observational studies, or as the evaluation component of intervention studies. It does not address considerations needed for the design and delivery of mobile or digital intervention strategies, such as when and how to provide intervention notifications that are part of microrandomized trials or JITAIs [14,15]. While the tutorial does provide some guidance on lagged versus leading observations and data aggregation, which is essential given that these choices can influence model outcomes and require transparency on the part of researchers [101], it does not address in depth the data processing and analysis needs inherent to ILD studies. For example, merging EMA data with sensor data based on time stamps and creating lagged or leading variables may require programming skills not typically taught in graduate training. Researchers may need to develop advanced data analytic skills or consult with data scientists and biostatisticians to ensure that data processing and analytic approaches are appropriately aligned with the a priori research question. The advent of AI systems provides another potential resource for researchers to use when encountering the data processing and aggregation challenges common to ILD. Furthermore, although the tutorial discusses selective data-processing decisions that have direct implications for study design (eg, temporal alignment and aggregation of observations), it does not provide detailed guidance on handling missing EMA responses, participant attrition, or incomplete sensor data. These issues are important considerations in ILD research and can have important implications for data quality and validity but were considered beyond the scope of the current design-focused tutorial [109]. Finally, the scope and emphasis of the tutorial are informed by the authors' own research experiences and areas of applied work, which have influenced the selection of examples, populations, and design priorities. As a result, some ILD applications, research domains, or methodological perspectives may be underrepresented. While many of the principles discussed are broadly applicable, the tutorial primarily draws on examples from student and community populations and does not provide detailed guidance tailored to clinical or patient populations (eg, individuals with chronic conditions or older adults), for whom additional feasibility, burden, and adherence considerations may be especially important [41-44].

Further research in several key areas is needed to help researchers make better decisions about selecting measures and designing protocols for ILD studies. First, more comprehensive and systematic psychometric testing of daily diary and EMA items that compare different recall windows (eg, in the past hour vs 2 hours) and sampling densities (eg, 3 vs 10 prompts per day) can inform decisions about when and how often items should be assessed. Additionally, more thorough evaluations of the reliability and validity of EMA and sensor methods compared with established standards are needed. It would be helpful to develop an open-source repository that provides example items and detailed information about their timing, variability, and assessment burden within specific subgroups. Second, studies are needed to systematically test accessibility and usability of mobile, digital, and wearable assessment devices across vulnerable populations. For example, with older adult populations, the font on a smartwatch may be more difficult to see and vibrations on a sensor may be less detectable due to decreased skin sensitivity. Finally, future research directions include substantial opportunities to apply recent advancements in computational modeling, machine learning, and AI to predict the most opportune time to deliver the most useful type of ILD assessments to specific participants to reduce response burden and maximize the quantity and quality of the data.

Conclusions

Advances in mobile technology and statistical methods have increased the accessibility and feasibility of collecting and analyzing ILD in health behavior research studies. However, a

XSL•FO
RenderX

lack of readily available information and training on how to select measures and design the sampling protocol for ILD studies has hampered progress and the application of these methods. The tutorial provided here is intended to make it easier for researchers to implement rigorous ILD studies to enable scientific advancement and understanding of the nature of time-varying phenomena in health.

Funding

The authors declared no financial support was received for this work.

Authors' Contributions

Conceptualization: JC, GD
Methodology: JC, GD
Project administration: JC, GD
Writing – Original Draft: JC, GD

Conflicts of Interest

None declared.

Multimedia Appendix 1
Decision-making tutorial for selecting intensive longitudinal data measures and designing a study using intensive longitudinal data.
[DOCX File, 24 KB - mhealth_v14i1e81290_app1.docx]

Multimedia Appendix 2
R Markdown code file and .csv data file for example intensive longitudinal data coding and analysis.
[ZIP File, 38 KB - mhealth_v14i1e81290_app2.zip]

References

1. Baraldi AN, Wurpts IC, MacKinnon DP, Lockhart G. In: Marsch LA, Dallery J, editors. Evaluating Mechanisms of Behavior Change to Inform and Evaluate Technology-Based Interventions: Oxford University Press; 2015:187-199. [doi: 10.1093/med/9780199314027.003.0013]
2. de Vries LP, Baselmans BML, Bartels M. Smartphone-based ecological momentary assessment of well-being: a systematic review and recommendations for future studies. J Happiness Stud 2021;22(5):2361-2408. [doi: 10.1007/s10902-020-00324-7] [Medline: 34720691]
3. Hamaker EL, Wichers M. No time like the present: discovering the hidden dynamics in intensive longitudinal data. Curr Dir Psychol Sci 2017;26:10-15. [doi: 10.1177/0963721416666518]
4. Mehl MR, Conner TS, Csikszentmihalyi M. Handbook of Research Methods for Studying Daily Life, th edition: The Guilford Press; 2014.
5. Trull TJ, Ebner-Priemer U. The role of ambulatory assessment in psychological science. Curr Dir Psychol Sci 2014 Dec;23(6):466-470. [doi: 10.1177/0963721414550706] [Medline: 25530686]
6. Bolger N, Laurenceau JP. Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research: Guilford Press; 2013.
7. Collins LM. Analysis of longitudinal data: the integration of theoretical model, temporal design, and statistical model. Annu Rev Psychol 2006;57:505-528. [doi: 10.1146/annurev.psych.57.102904.190146] [Medline: 16318605]
8. Christensen TC, Barrett LF, Bliss-Moreau E, Lebo K, Kaschub C. A practical guide to experience-sampling procedures. J Happiness Stud 2003 Mar;4(1):53-78. [doi: 10.1023/A:1023609306024]
9. Hufford MR, Shiffman S. Assessment methods for patient-reported outcomes. Dis Manage Health Outcomes 2003;11(2):77-86. [doi: 10.2165/00115677-200311020-00002]
10. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
11. Smyth JM, Stone AA. Ecological momentary assessment research in behavioral medicine. J Happiness Stud 2003 Mar;4(1):35-52. [doi: 10.1023/A:1023657221954]
12. Hajar R. Risk factors for coronary artery disease: historical perspectives. Heart Views 2017;18(3):109-114. [doi: 10.4103/HEARTVIEWS.HEARTVIEWS_106_17] [Medline: 29184622]
13. Wu S, Powers S, Zhu W, Hannun YA. Substantial contribution of extrinsic risk factors to cancer development. Nature 2016 Jan 7;529(7584):43-47. [doi: 10.1038/nature16166] [Medline: 26675728]

XSL•FO
RenderX

14. Klasnja P, Smith S, Seewald NJ, et al. Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of heartSteps. Ann Behav Med 2019 May 3;53(6):573-582. [doi: 10.1093/abm/kay067] [Medline: 30192907]
15. Wang L, Miller LC. Just-in-the-Moment Adaptive Interventions (JITAI): a meta-analytical review. Health Commun 2020 Nov;35(12):1531-1544. [doi: 10.1080/10410236.2019.1652388] [Medline: 31488002]
16. Moore RC, Depp CA, Wetherell JL, Lenze EJ. Ecological momentary assessment versus standard assessment instruments for measuring mindfulness, depressed mood, and anxiety among older adults. J Psychiatr Res 2016 Apr;75:116-123. [doi: 10.1016/j.jpsychires.2016.01.011] [Medline: 26851494]
17. Boaro A, Leung J, Reeder HT, et al. Smartphone GPS signatures of patients undergoing spine surgery correlate with mobility and current gold standard outcome measures. J Neurosurg Spine 2021 Dec 1;35(6):796-806. [doi: 10.3171/2021.2.SPINE202181] [Medline: 34450590]
18. Cappon G, Vettoretti M, Sparacino G, Facchinetti A. Continuous glucose monitoring sensors for diabetes management: a review of technologies and applications. Diabetes Metab J 2019 Aug;43(4):383-397. [doi: 10.4093/dmj.2019.0121] [Medline: 31441246]
19. Clavel P, Tiollier E, Leduc C, Fabre M, Lacome M, Buchheit M. Concurrent validity of a continuous glucose-monitoring system at rest and during and following a high-intensity interval training session. Int J Sports Physiol Perform 2022 Apr 1;17(4):627-633. [doi: 10.1123/ijspp.2021-0222] [Medline: 35193110]
20. Georgiou K, Larentzakis AV, Khamis NN, Alsuhaibani GI, Alaska YA, Giallafos EJ. Can wearable devices accurately measure heart rate variability? A systematic review. Folia Med (Plovdiv) 2018 Mar 1;60(1):7-20. [doi: 10.2478/folmed-2018-0012] [Medline: 29668452]
21. Judice P, Fouto J, Horta M, Pereira S. Validity of iPhone health application step count in semi free-living conditions. Acta IMEKO 2024;13(1):1-7. [doi: 10.21014/actaimeko.v13i1.1673]
22. Schoenmakers M, Saygin M, Sikora M, Vaessen T, Noordzij M, de Geus E. Stress in action wearables database: a database of noninvasive wearable monitors with systematic technical, reliability, validity, and usability information. Behav Res 2025;57(6):171. [doi: 10.3758/s13428-025-02685-4]
23. Steeves JA, Bowles HR, McClain JJ, et al. Ability of thigh-worn actiGraph and activPAL monitors to classify posture and motion. Med Sci Sports Exerc 2015 May;47(5):952-959. [doi: 10.1249/MSS.0000000000000497] [Medline: 25202847]
24. Yang CC, Hsu YL. A review of accelerometry-based wearable motion detectors for physical activity monitoring. Sensors (Basel) 2010;10(8):7772-7788. [doi: 10.3390/s100807772] [Medline: 22163626]
25. Yu J, Fairbairn CE, Gurrieri L, Caumiant EP. Validating transdermal alcohol biosensors: a meta-analysis of associations between blood/breath-based measures and transdermal alcohol sensor output. Addiction 2022 Nov;117(11):2805-2815. [doi: 10.1111/add.15953] [Medline: 35603913]
26. Collins RL, Kashdan TB, Gollnisch G. The feasibility of using cellular phones to collect ecological momentary assessment data: application to alcohol consumption. Exp Clin Psychopharmacol 2003 Feb;11(1):73-78. [doi: 10.1037//1064-1297.11.1.73] [Medline: 12622345]
27. Crooke AHD, Reid SC, Kauer SD, et al. Temporal mood changes associated with different levels of adolescent drinking: using mobile phones and experience sampling methods to explore motivations for adolescent alcohol use. Drug Alcohol Rev 2013 May;32(3):262-268. [doi: 10.1111/dar.12034] [Medline: 23432520]
28. Dunton GF. Ecological momentary assessment in physical activity research. Exerc Sport Sci Rev 2017 Jan;45(1):48-54. [doi: 10.1249/JES.0000000000000092] [Medline: 27741022]
29. Dunton GF, Liao Y, Intille S, Huh J, Leventhal A. Momentary assessment of contextual influences on affective response during physical activity. Health Psychol 2015 Dec;34(12):1145-1153. [doi: 10.1037/hea0000223] [Medline: 26053885]
30. Dunton GF, Dzubur E, Li M, Huh J, Intille S, McConnell R. Momentary assessment of psychosocial stressors, context, and asthma symptoms in hispanic adolescents. Behav Modif 2016 Jan;40(1-2):257-280. [doi: 10.1177/0145445515608145]
31. Kampshoff CS, Verdonck-de Leeuw IM, van Oijen MG, Sprangers MA, Buffart LM. Ecological momentary assessments among patients with cancer: a scoping review. Eur J Cancer Care (Engl) 2019 May;28(3):e13095. [doi: 10.1111/ecc.13095] [Medline: 31090160]
32. Schneider S, Stone AA. Ambulatory and diary methods can facilitate the measurement of patient-reported outcomes. Qual Life Res 2016 Mar;25(3):497-506. [doi: 10.1007/s11136-015-1054-z] [Medline: 26101141]
33. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. Psychol Assess 2009 Dec;21(4):486-497. [doi: 10.1037/a0017074] [Medline: 19947783]
34. Almeida DM, Wethington E, Kessler RC. The daily inventory of stressful events: an interview-based approach for measuring daily stressors. Assessment 2002 Mar;9(1):41-55. [doi: 10.1177/1073191102091006] [Medline: 11911234]
35. Lischetzke T. Daily diary methodology. In: Michalos AC, editor. Encyclopedia of Quality of Life and Well-Being Research [Internet]: Springer Netherlands; 2014:1413-1419. [doi: 10.1007/978-94-007-0753-5_657]
36. Stone AA, Wen CKF, Schneider S, Junghaenel DU. Evaluating the effect of daily diary instructional phrases on respondents' recall time frames: survey experiment. J Med Internet Res 2020 Feb 21;22(2):e16105. [doi: 10.2196/16105] [Medline: 32130129]

XSL•FO
RenderX

37. O'Connor SG, Ke W, Dzubur E, Schembre S, Dunton GF. Concordance and predictors of concordance of children's dietary intake as reported via ecological momentary assessment and 24 h recall. Public Health Nutr 2018 Apr;21(6):1019-1027. [doi: 10.1017/S1368980017003780] [Medline: 29352820]
38. Smith RA, Lack LC, Lovato N, Wright H. The relationship between a night's sleep and subsequent daytime functioning in older poor and good sleepers. J Sleep Res 2015 Feb;24(1):40-46. [doi: 10.1111/jsr.12237] [Medline: 25204612]
39. Fritz J, Piccirillo ML, Cohen ZD, et al. So you want to do ESM? 10 essential topics for implementing the experience-sampling method. Adv Methods Pract Psychol Sci 2024 Jul;7(3):25152459241267912. [doi: 10.1177/25152459241267912]
40. Naragon-Gainey K, Stanton K. Optimizing intensive longitudinal designs for clinical psychology research and assessment. Psychol Assess 2025 Oct;37(10):461-465. [doi: 10.1037/pas0001424] [Medline: 41129391]
41. Nowojewski A, Bark E, Shih VH, Dearden R. Patient adherence and response time in electronic patient-reported outcomes: insights from three longitudinal clinical trials. Qual Life Res 2024 Jun;33(6):1691-1706. [doi: 10.1007/s11136-024-03644-w] [Medline: 38598132]
42. Rintala A, Wampers M, Myin-Germeys I, Viechtbauer W. Response compliance and predictors thereof in studies using the experience sampling method. Psychol Assess 2019 Feb;31(2):226-235. [doi: 10.1037/pas0000662] [Medline: 30394762]
43. Vachon H, Viechtbauer W, Rintala A, Myin-Germeys I. Compliance and retention with the experience sampling method over the continuum of severe mental disorders: meta-analysis and recommendations. J Med Internet Res 2019 Dec 6;21(12):e14475. [doi: 10.2196/14475] [Medline: 31808748]
44. Bell IH, Eisner E, Allan S, et al. Methodological characteristics and feasibility of ecological momentary assessment studies in psychosis: a systematic review and meta-analysis. Schizophr Bull 2024 Mar 7;50(2):238-265. [doi: 10.1093/schbul/sbad127] [Medline: 37606276]
45. Conroy DE, Ram N, Pincus AL, et al. Daily physical activity and alcohol use across the adult lifespan. Health Psychol 2015 Jun;34(6):653-660. [doi: 10.1037/hea0000157] [Medline: 25222084]
46. Courtney JB, Russell MA. Testing affect regulation models of drinking prior to and after drinking initiation using ecological momentary assessment. Psychol Addict Behav 2021 Aug;35(5):597-608. [doi: 10.1037/adb0000763] [Medline: 34138595]
47. Courtney JB, Russell MA. To drink or not to drink: is that the question? Examining correspondence and predictive validity of morning drinking intentions for young adults' drinking behaviors and consequences. Prev Sci 2023 Feb;24(2):322-336. [doi: 10.1007/s11121-022-01437-6] [Medline: 36155882]
48. Dunton GF, Huh J, Leventhal AM, et al. Momentary assessment of affect, physical feeling states, and physical activity in children. Health Psychol 2014 Mar;33(3):255-263. [doi: 10.1037/a0032640] [Medline: 23668846]
49. Fridberg DJ, Lee Z, Fischer AM, Cursio JF, King AC. Assessing real‐time positive subjective effects of alcohol using high‐resolution ecological momentary assessment in risky versus light drinkers. Alcohol Clin Exp Res 2023 Aug;47(8):1570-1581. [doi: 10.1111/acer.15130]
50. Maher JP, Dzubur E, Nordgren R, et al. Do fluctuations in positive affective and physical feeling states predict physical activity and sedentary time? Psychol Sport Exerc 2019 Mar;41:153-161. [doi: 10.1016/j.psychsport.2018.01.011] [Medline: 30853854]
51. Russell MA, Turrisi RJ, Smyth JM. Transdermal sensor features correlate with ecological momentary assessment drinking reports and predict alcohol‐related consequences in young adults' natural settings. Alcoholism Clin & Exp Res 2022 Jan;46(1):100-113. [doi: 10.1111/acer.14739]
52. Wang S, Yang CH, Brown D, Cheng A, Kwan MYW. Participant compliance with ecological momentary assessment in movement behavior research among adolescents and emerging adults: systematic review. JMIR Mhealth Uhealth 2025 Feb 11;13:e52887. [doi: 10.2196/52887] [Medline: 39933165]
53. Dora J, Piccirillo M, Foster KT, et al. The daily association between affect and alcohol use: a meta-analysis of individual participant data. Psychol Bull 2023;149(1-2):1-24. [doi: 10.1037/bul0000387] [Medline: 37560174]
54. Maher JP, Rhodes RE, Dzubur E, Huh J, Intille S, Dunton GF. Momentary assessment of physical activity intention-behavior coupling in adults. Transl Behav Med 2017 Dec;7(4):709-718. [doi: 10.1007/s13142-017-0472-6] [Medline: 28155108]
55. Schultchen D, Reichenberger J, Mittl T, et al. Bidirectional relationship of stress and affect with physical activity and healthy eating. Br J Health Psychol 2019 May;24(2):315-333. [doi: 10.1111/bjhp.12355] [Medline: 30672069]
56. Lee SP, Ha G, Wright DE, et al. Highly flexible, wearable, and disposable cardiac biosensors for remote and ambulatory monitoring. NPJ Digit Med 2018;1(1):2. [doi: 10.1038/s41746-017-0009-x] [Medline: 31304288]
57. Clements MA, Foster NC, Maahs DM, et al. Hemoglobin A1c (HbA1c) changes over time among adolescent and young adult participants in the T1D exchange clinic registry. Pediatr Diabetes 2016 Aug;17(5):327-336. [doi: 10.1111/pedi.12295] [Medline: 26153338]
58. Neubauer AB, Schmiedek F. Studying within-person variation and within-person couplings in intensive longitudinal data: lessons learned and to be learned. Gerontology 2020;66(4):332-339. [doi: 10.1159/000507993] [Medline: 32526756]
59. Wood JK. Integrating multiple timescales and process models: an illustrative application to affect dynamics. 2018 URL: https://etda.libraries.psu.edu/files/final_submissions/16435 [accessed 2026-08-07]
60. Dvorak RD, Pearson MR, Sargent EM, Stevenson BL, Mfon AM. Daily associations between emotional functioning and alcohol involvement: moderating effects of response inhibition and gender. Drug Alcohol Depend 2016 Jun 1;163 Suppl 1(Suppl 1):S46-S53. [doi: 10.1016/j.drugalcdep.2015.09.034] [Medline: 27306731]

XSL•FO
RenderX

61. Fairlie AM, Cadigan JM, Patrick ME, Larimer ME, Lee CM. Unplanned heavy episodic and high-intensity drinking: daily-level associations with mood, context, and negative consequences. J Stud Alcohol Drugs 2019 May;80(3):331-339. [doi: 10.15288/jsad.2019.80.331] [Medline: 31250798]
62. Grzywacz JG, Almeida DM. Stress and binge drinking: a daily process examination of stressor pile-up and socioeconomic status in affect regulation. Int J Stress Manag 2008 Nov 1;15(4):364-380. [doi: 10.1037/a0013368] [Medline: 19578556]
63. Hamilton HR, Armeli S, Tennen H. To drink or not to drink: when drinking intentions predict alcohol consumption and consequences. Br J Health Psychol 2022 May;27(2):516-533. [doi: 10.1111/bjhp.12560] [Medline: 34545657]
64. Keusch F, Struminskaya B, Antoun C, Couper MP, Kreuter F. Willingness to participate in passive mobile data collection. Public Opin Q 2019 Jul;83(Suppl 1):210-235. [doi: 10.1093/poq/nfz007] [Medline: 31337924]
65. König LM, Allmeta A, Christlein N, Van Emmenis M, Sutton S. A systematic review and meta-analysis of studies of reactivity to digital in-the-moment measurement of health behaviour. Health Psychol Rev 2022 Dec;16(4):551-575. [doi: 10.1080/17437199.2022.2047096] [Medline: 35264084]
66. Gunn RL, Merrill JE, Haines AM, et al. Use of the BACtrack Skyn alcohol biosensor: practical applications for data collection and analysis. Addiction 2023 Aug;118(8):1586-1595. [doi: 10.1111/add.16207] [Medline: 37060272]
67. Wang SD, Hatzinger L, Morales J, Hewus M, Intille S, Dunton GF. Burden and inattentive responding in a 12-month intensive longitudinal study: interview study among young adults. JMIR Form Res 2024 Aug 2;8:e52165. [doi: 10.2196/52165] [Medline: 39093606]
68. Burke LE, Shiffman S, Music E, et al. Ecological momentary assessment in behavioral research: addressing technological and human participant challenges. J Med Internet Res 2017 Mar 15;19(3):e77. [doi: 10.2196/jmir.7138] [Medline: 28298264]
69. Businelle MS, Hébert ET, Shi D, et al. Investigating best practices for ecological momentary assessment: nationwide factorial experiment. J Med Internet Res 2024 Aug 12;26:e50275. [doi: 10.2196/50275] [Medline: 39133915]
70. Hamaker EL. The curious case of the cross-sectional correlation. Multivariate Behav Res 2024;59(6):1111-1122. [doi: 10.1080/00273171.2022.2155930] [Medline: 36599045]
71. McNeish D. A primer on intensive longitudinal psychometrics. Behav Res 2026;58(5):121. [doi: 10.3758/s13428-026-02946-w]
72. Bielak AAM, Mogle JA, Sliwinski MJ. Within- and between-person differences in activity factor structure: results from an ecological momentary assessment study. In: Cassidy B, editor. J Gerontol B Psychol Sci Soc Sci 2025 Apr 7;80(5):gbaf017. [doi: 10.1093/geronb/gbaf017] [Medline: 39879285]
73. Vogelsmeier L, Vermunt JK, De Roover K. How to explore within-person and between-person measurement model differences in intensive longitudinal data with the R package lmfa. Behav Res 2022;55(5):2387-2422. [doi: 10.3758/s13428-022-01898-1]
74. ESM Item Repository. 2024. URL: https://www.esmitemrepository.com [accessed 2025-02-10]
75. Goldstein H, Browne W. In: Maydeu-Olivares A, McArdle J, editors. Multilevel Factor Analysis for Continuous and Discrete Data: Erlbaum; 2005:453-475.
76. Podsakoff NP, Spoelma TM, Chawla N, Gabriel AS. What predicts within-person variance in applied psychology constructs? An empirical examination. J Appl Psychol 2019 Jun;104(6):727-754. [doi: 10.1037/apl0000374] [Medline: 30640492]
77. Lazarides C, Ward EB, Buss C, et al. Psychological stress and cortisol during pregnancy: an ecological momentary assessment (EMA)-Based within- and between-person analysis. Psychoneuroendocrinology 2020 Nov;121:104848. [doi: 10.1016/j.psyneuen.2020.104848] [Medline: 32927180]
78. Castro-Alvarez S, Zhou DJ, Bringmann LF, et al. Assessing the internal consistency reliability of ecological momentary assessment measures: insights from the WARN-D study. Psychol Assess 2025 Nov;37(11):614-623. [doi: 10.1037/pas0001410] [Medline: 41196735]
79. Schneider S, Stone AA, Schwartz JE, Broderick JE. Peak and end effects in patients' daily recall of pain and fatigue: a within-subjects analysis. J Pain 2011 Feb;12(2):228-235. [doi: 10.1016/j.jpain.2010.07.001] [Medline: 20817615]
80. Gelinas L, Morrell W, Bierer BE. Terms and conditions apply: an ethical analysis of mobile health user agreements in research. J Law Biosci 2023;10(2):lsad021. [doi: 10.1093/jlb/lsad021] [Medline: 37456712]
81. Hubach RD, O'Neil A, Stowe M, Giano Z, Curtis B, Fisher CB. Perceived confidentiality risks of mobile technology-based ecologic momentary assessment to assess high-risk behaviors among rural men who have sex with men. Arch Sex Behav 2021 May;50(4):1641-1650. [doi: 10.1007/s10508-019-01612-x] [Medline: 32078710]
82. de Montjoye YA, Hidalgo CA, Verleysen M, Blondel VD. Unique in the crowd: the privacy bounds of human mobility. Sci Rep 2013;3:1376. [doi: 10.1038/srep01376] [Medline: 23524645]
83. Stone AA, Shiffman S, Atienza AA, Nebeling L. The Science of Real-Time Data Capture: Self-Reports in Health Research: Oxford University Press; 2007.
84. Delespaul P. Assessing schizophrenia in daily life: the experience sampling method. : Maastricht University; 1995. [doi: 10.26481/dis.19950504pd]
85. Ji Y, Schreuder MJ, Ariens S, Dejonckheere E, Ceulemans E. Evaluating the feasibility of episode-contingent experience sampling burst designs. Psychol Assess 2025 Oct;37(10):520-534. [doi: 10.1037/pas0001404] [Medline: 41129396]

XSL•FO
RenderX

86. Wang SD, Intille S, Ponnada A, Do B, Rothman A, Dunton GF. Investigating microtemporal processes underlying health behavior adoption and maintenance: protocol for an intensive longitudinal observational study. JMIR Res Protoc 2022;11(7):e36666. [doi: 10.2196/36666]
87. Dunton GF, Liao Y, Dzubur E, et al. Investigating within-day and longitudinal effects of maternal stress on children's physical activity, dietary intake, and body composition: Protocol for the MATCH study. Contemp Clin Trials 2015 Jul;43:142-154. [doi: 10.1016/j.cct.2015.05.007] [Medline: 25987483]
88. Dzubur E, Li M, Kawabata K, et al. Design of a smartphone application to monitor stress, asthma symptoms, and asthma inhaler use. Ann Allergy Asthma Immunol 2015 Apr;114(4):341-342. [doi: 10.1016/j.anai.2014.12.017]
89. Monnaatsie M, Biddle SJH, Kolbe-Alexander T. Feasibility of ecological momentary assessment in measuring physical activity and sedentary behaviour in shift and non-shift workers. J Act Sedentary Sleep Behav 2024 Oct 8;3(1):24. [doi: 10.1186/s44167-024-00063-7] [Medline: 40217510]
90. Bolger N, Stadler G, Laurenceau JP. Power analysis for intensive longitudinal studies. In: Mehl MR, Conner TS, editors. Handbook of Research Methods for Studying Daily Life: Guildford Press; 2011:285-301.
91. Maas CJM, Hox JJ. Sufficient sample sizes for multilevel modeling. Methodology 2005 Jan;1(3):86-92. [doi: 10.1027/1614-2241.1.3.86]
92. Hilden P, Schwartz JE, Pascual C, Diaz KM, Goldsmith J, editor. How many days are needed? Measurement reliability of wearable device data to assess physical activity. PLoS ONE 2023;18(2):e0282162. [doi: 10.1371/journal.pone.0282162] [Medline: 36827427]
93. St George SM, Van Horn ML, Lawman HG, Wilson DK. Reliability of 24-hour dietary recalls as a measure of diet in African-American youth. J Acad Nutr Diet 2016 Oct;116(10):1551-1559. [doi: 10.1016/j.jand.2016.05.011] [Medline: 27394936]
94. Lafit G, Adolf JK, Dejonckheere E, Myin-Germeys I, Viechtbauer W, Ceulemans E. Selection of the number of participants in intensive longitudinal studies: a user-friendly shiny app and tutorial for performing power analysis in multilevel regression models that account for temporal dependencies. Adv Methods Pract Psychol Sci 2021 Jan;4(1):251524592097873. [doi: 10.1177/2515245920978738]
95. Lafit G, Artner R, Ceulemans E. Enabling analytical power calculations for multilevel models with autocorrelated errors through deriving and approximating the precision matrix. Behav Res Methods 2024 Oct;56(7):8105-8131. [doi: 10.3758/s13428-024-02435-y] [Medline: 39009823]
96. Murayama K, Usami S, Sakaki M. Summary-statistics-based power analysis: A new and practical method to determine sample size for mixed-effects modeling. Psychol Methods 2022. [doi: 10.1037/met0000330]
97. van de Maat R, Lataster J, Verboon P. Minimum required sample size for modelling daily cyclic patterns in ecological momentary assessment data. Methodology 2024;20(4):265-282. [doi: 10.5964/meth.11399]
98. Eisele G, Vachon H, Lafit G, et al. The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. Assessment 2022 Mar;29(2):136-151. [doi: 10.1177/1073191120957102] [Medline: 32909448]
99. Gibson AM, Bowling NA. The effects of questionnaire length and behavioral consequences on careless responding. Eur J Psychol Assess 2020 Mar;36(2):410-420. [doi: 10.1027/1015-5759/a000526]
100. Dunton GF, Liao Y, Intille SS, Spruijt-Metz D, Pentz M. Investigating children's physical activity and sedentary behavior using ecological momentary assessment with mobile phones. Obesity (Silver Spring) 2011 Jun;19(6):1205-1212. [doi: 10.1038/oby.2010.302] [Medline: 21164502]
101. Bastiaansen JA, Kunkels YK, Blaauw FJ, et al. Time to get personal? The impact of researchers choices on the selection of treatment targets using the experience sampling methodology. J Psychosom Res 2020 Oct;137:110211. [doi: 10.1016/j.jpsychores.2020.110211] [Medline: 32862062]
102. Boruvka A, Almirall D, Witkiewitz K, Murphy SA. Assessing time-varying causal effect moderation in mobile health. J Am Stat Assoc 2018;113(523):1112-1121. [doi: 10.1080/01621459.2017.1305274] [Medline: 30467446]
103. Kwasnicka D, Inauen J, Nieuwenboom W, et al. Challenges and solutions for N-of-1 design studies in health psychology. Health Psychol Rev 2019 Jun;13(2):163-178. [doi: 10.1080/17437199.2018.1564627] [Medline: 30626274]
104. Preacher KJ, Zyphur MJ, Zhang Z. A general multilevel SEM framework for assessing multilevel mediation. Psychol Methods 2010 Sep;15(3):209-233. [doi: 10.1037/a0020141] [Medline: 20822249]
105. Tan X, Shiyko MP, Li R, Li Y, Dierker L. A time-varying effect model for intensive longitudinal data. Psychol Methods 2012 Mar;17(1):61-77. [doi: 10.1037/a0025814] [Medline: 22103434]
106. Dzubur E, Ponnada A, Nordgren R, et al. MixWILD: a program for examining the effects of variance and slope of time-varying variables in intensive longitudinal data. Behav Res 2020 Aug;52(4):1403-1427. [doi: 10.3758/s13428-019-01322-1]
107. Li J, Ponnada A, Wang WL, Dunton GF, Intille SS. Ask less, learn more: adapting ecological momentary assessment survey length by modeling question-answer information gain. Proc ACM Interact Mob Wearable Ubiquitous Technol 2024 Nov;8(4):1-32. [doi: 10.1145/3699735] [Medline: 39664111]
108. Schneider S, Junghaenel DU, Smyth JM, Fred Wen CK, Stone AA. Just-in-time adaptive ecological momentary assessment (JITA-EMA). Behav Res 2023;56(2):765-783. [doi: 10.3758/s13428-023-02083-8]

XSL•FO
RenderX

109. Yuan C, Hedeker D, Mermelstein R, Xie H. A tractable method to account for high-dimensional nonignorable missing data in intensive longitudinal data. Stat Med 2020 Sep 10;39(20):2589-2605. [doi: 10.1002/sim.8560] [Medline: 32367549]

Abbreviations

DV: dependent variables
EMA: ecological momentary assessment
ESM: experience sampling methods
ILD: intensive longitudinal data
IV: independent variables
JITAI: just-in-time adaptive intervention
MVPA: moderate-to-vigorous physical activity
TIME: Temporal Influences on Movement and Exercise

Edited by L Buis; submitted 01.Dec.2025; peer-reviewed by A Torad, E Mico-Amigo, T Nakagawa; revised version received 20.Jul.2026; accepted 21.Jul.2026; published 08.Sep.2026.

Please cite as:
Courtney J, Dunton G
Intensive Longitudinal Data Collection Methods in Health Research: Tutorial for Selecting Measures and Designing a Sampling Protocol
JMIR Mhealth Uhealth 2026;14:e81290
URL: https://mhealth.jmir.org/2026/1/e81290
doi:10.2196/81290

© Jimikaye Courtney, Genevieve Dunton. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 8.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Cardiovascular Care Pathway Supported by Remote Patient Management in Dutch Primary Care: Cost-Effectiveness, Budget Impact, and Workload Analysis

Jan Heijdra Suasnabar[1*], MSc; Margot Rakers[2,3*], MD; Nicoline van Hattem[2,3], MD; Eric Hiddink[2,3,4], PharmD; Marcel Haas[5], PhD; Xuerui Zhang[6], MSc; Douwe Atsma[3,7,8], MD, PhD; Tobias N Bonten[2,3], MD, PhD; M Elske van den Akker-van Marle[1], PhD; Hendrikus J A van Os[2,3,9], MD, PhD

[1]Section of Medical Decision Making, Biomedical Data Science, Leiden University Medical Center, Albinusdreef 2, Leiden, The Netherlands

[2]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands

[3]National eHealth Living Lab (NeLL), Leiden University Medical Center, Leiden, The Netherlands

[4]Stichting Health Base, Houten, The Netherlands

[5]Health Campus, Leiden University Medical Center, The Hague, The Netherlands

[6]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

[7]Department of Industrial Design Engineering, Delft University of Technology, Delft, The Netherlands

[8]Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands

[9]Department of Psychiatry, University Medical Center Utrecht, Utrecht, The Netherlands

[*]these authors contributed equally

Corresponding Author:
Jan Heijdra Suasnabar, MSc
Section of Medical Decision Making, Biomedical Data Science, Leiden University Medical Center, Albinusdreef 2, Leiden, The Netherlands

Abstract

Background: Remote patient management (RPM) that supports patient self-monitoring of vital parameters and lifestyle factors may improve cardiovascular risk management (CVRM) in primary care. However, large-scale implementation remains limited, partly due to insufficient evidence on long-term value for money, budget impact, and implications for health care professionals' workload.

Objective: This study aimed to estimate the long-term cost-effectiveness, 5-year health care budget impact, and expected changes in general practitioner (GP) and practice nurse (PN) workload associated with the CVRM-Box intervention in the Netherlands.

Methods: We conducted a model-based economic evaluation comparing CVRM-Box with care as usual in Dutch primary care integrated CVRM programs. CVRM-Box is a multicomponent RPM intervention comprising a digital blood pressure (BP) monitor, digital weight scale, step counter/activity tracker, and a mobile app, with measurements transferred to the GP practice for periodic review. A time-inhomogeneous cohort Markov model simulated lifetime transitions among health states, including at-risk, post–myocardial infarction (MI), poststroke, recurrent events, cardiovascular death, and noncardiovascular death. Cardiovascular risks were modeled using prediction equations (SCORE2, SCORE2-OP, and SMART2) populated with subgroup-specific risk factor profiles. Intervention effects were modeled as changes in systolic BP derived from a matched cohort study of CVRM-Box. Other key parameters (costs and utilities) were similarly obtained from the matched cohort study, or routine primary care data, and published sources. Outcomes included incremental cost-effectiveness ratios (ICERs, indicating cost per quality-adjusted life year [QALY] gained), cost-effectiveness probabilities, 5-year health care budget impact, and an exploratory workload analysis estimating annual changes in visit time and remote consultation frequency for GPs and PNs. Probabilistic and scenario uncertainty analyses were performed.

Results: In the overall population, CVRM-Box increased costs and QALYs versus care as usual, yielding an ICER of €17,340/QALY gained (EUR €1=US $1.11 as of 29 December 2023) and a 60% probability of cost-effectiveness at a willingness-to-pay threshold of €20,000/QALY. Cost-effectiveness was more favorable in higher-risk subgroups (uncontrolled BP and/or prior MI/stroke), with ≥70% probability of cost-effectiveness at €20,000/QALY, whereas in the lower-risk subgroup with controlled BP and no prior MI/stroke, the intervention is unlikely to be cost-effective (ICER: €42,384/QALY). The 5-year health care budget impact was €2.9 million and €662.2 million for regional and national rollout, respectively. In a typical Dutch

XSL•FO
RenderX

primary care practice, the CVRM-Box reduced PN workload by 25.9 h and 44.7 remote consultations annually, while the change in GP workload was negligible (0.5 h; 3.3 remote consultations).

Conclusions: The CVRM-Box multicomponent RPM intervention is likely to be cost-effective among high-risk subgroups but not in lower-risk groups. The intervention further reduces PN workload but not GP workload. Results primarily apply to the Dutch context and may not generalize to health care systems with different payment/incentive arrangements. Additionally, long-term outcomes were modeled using prediction models rather than observed cardiovascular events. Nonetheless, our findings provide an argument for implementation in primary care for higher-risk subgroups.

(JMIR Mhealth Uhealth 2026;14:e79512) doi:10.2196/79512

KEYWORDS

patient management; preventive cardiology; primary care; cost-effectiveness; budget impact

Introduction

Cardiovascular disease (CVD) is a leading cause of death globally, contributing to substantial illness, mortality, and an economic burden on health care and society [1,2]. With aging populations and the increased prevalence of chronic diseases, primary care is strained due to a shortage of general practitioners (GPs) and practice nurses (PNs) and an increased workload [3-5]. Remote patient management (RPM) for preventive care is a relatively new approach that can improve reaching blood pressure (BP) targets, aid weight loss, and facilitate lifestyle change [6-10]. However, a significant gap remains in the large-scale implementation of RPM within integrated care [11,12].

In the Netherlands, a matched cohort study of a multicomponent RPM intervention, "cardiovascular risk management (CVRM)-Box," has been conducted to evaluate the effect of RPM within primary care settings for enhancing BP control, weight, and consultation frequency [13]. Patients at a high risk of major adverse cardiovascular events (MACEs), including myocardial infarction (MI), stroke, or transient ischemic attack, were recruited across 6 primary care practices [13]. Previous implementation studies of the CVRM-Box intervention, guided by the RE-AIM (Reach, Effectiveness, Adoption, Implementation and Maintenance) framework and the Consolidated Framework for Implementation Research, have shown that successful implementation depends on factors such as compatibility with routine practice, perceived workload for health care professionals, technical feasibility, and organizational support [14,15]. In addition to these implementation determinants, the cost-effectiveness of RPM interventions is a key consideration for their adoption, long-term sustainability, and large-scale implementation in routine primary care.

Health economic assessments have traditionally been tailored for pharmacological studies and medical devices and may not adequately capture the multifaceted nature of RPM interventions. Capturing the broader benefits of RPM interventions, such as enhancing health care professional efficiency through increasing patient self-management, is important for informed decision-making [16-18]. Effective implementation and scaling of RPM involves new value propositions shared among different stakeholders and capturing these different values that arise from RPM-supported care [19,20]. Waiting for numerous studies with larger time horizons and sample sizes is impractical and costly, as information is essential for upfront investment to achieve large-scale adoption. More insight into realistic scenarios of upscaling is crucial for informed decision-making [21].

Therefore, this study aimed to model the long-term cost-effectiveness, budget impact, and reduction in required visits to assess the potential scale-up of a multicomponent RPM intervention for the high-risk CVD population in primary care.

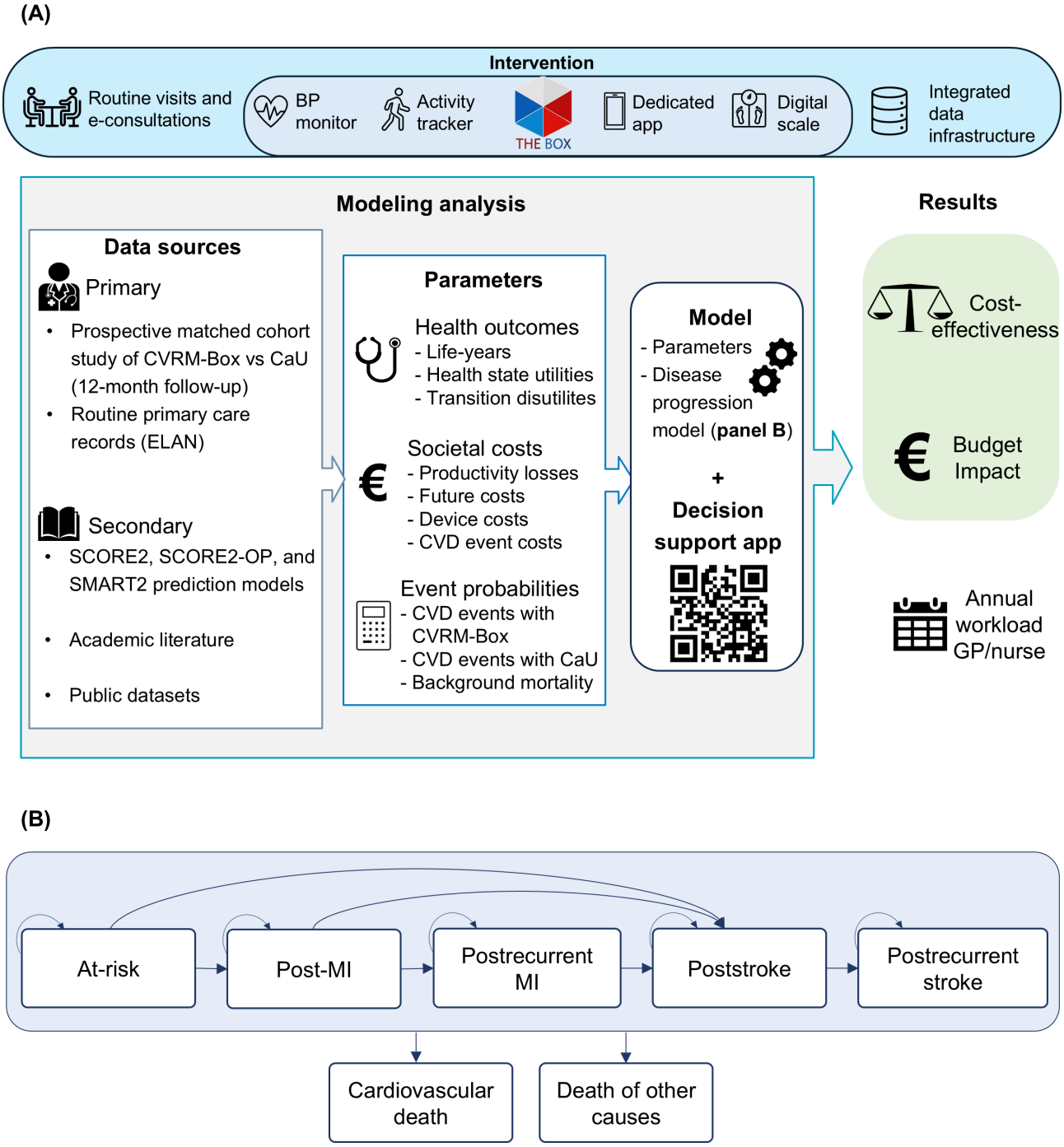
Methods

Study Design

The current study constitutes a model-based evaluation of CVRM-Box versus care as usual (CaU) [22]. This approach enables the transparent combination of multiple data sources in one modeling framework to estimate long-term health and economic outcomes under each care pathway (Figure 1A). Patients at risk of first or recurrent MACE were simulated throughout their remaining lifetime according to each care pathway. Our model input parameters incorporated data from 3 main sources: (1) primary data from the prospective propensity score-matched CVRM-Box cohort study [13], which informed the intervention effect on systolic blood pressure (SBP) and intervention-related resource use and costs; (2) primary data from the Extramural Leiden University Medical Center (LUMC) Academic Network (ELAN) database [23,24], which informed baseline risk factor profiles and CaU costs for at-risk patients; and (3) published literature, public datasets, and published Dutch costing guidance, which informed calculations of event probabilities, event and postevent costs, and health utilities. Table 1 summarizes key parameters and sources, while Multimedia Appendices 1 and 2 provide full parameterization and transition-probability calculations. Figure 1A presents a general schematic of our study design.

XSL•FO
RenderX

Figure 1. Study design schematic (A) and disease progression model (B). Arrows represent possible transitions between 1-year model cycles. BP: blood pressure; CaU: care as usual; CVD: cardiovascular disease; CVRM: cardiovascular remote management; ELAN: Extramural LUMC Academic Network; GP: general practitioner; MI: myocardial infarction.



XSL•FO
RenderX

Table . Model input parameters with corresponding credible intervals (CrIs), sources, and selected posterior distributions.

Model input parameter	Estimate (95% CrI)	Distribution[a]	Source
Proportions, %			
Male	53.1	Fixed	[13] CVRM-Box[b,c]
Subgroups			
Controlled BP[d], no CVD[e] history	48.2	Fixed	[13] CVRM-Box[b,c]
Controlled BP, MI[f]/stroke history	12.4	Fixed	[13] CVRM-Box[b,c]
Uncontrolled BP, no CVD history	34.5	Fixed	[13] CVRM-Box[b,c]
Uncontrolled BP, MI/stroke history	4.9	Fixed	[13] CVRM-Box[b,c]
Intervention effect			
Reduction in SBP[g], controlled BP subgroup, mm Hg	1.3 (−1.50 to 4.26)	Normal	[13] CVRM-Box[b,c]
Reduction in SBP, uncontrolled BP subgroup, mm Hg	3.5 (−0.70 to 7.94)	Normal	[13] CVRM-Box[b,c]
Smoking cessation proportion	0.112 (0.023 to 0.180)	Triangular	[10][c]
Costs per patient, €[h]			
Initial cost, intervention			
Device	186.8	Fixed	[13] CVRM-Box[b,c]
Personnel	26	Fixed	[13] CVRM-Box[b,c]
State-specific (annual) costs			
At risk with controlled BP, intervention	88 (50 to 141)	Gamma	[13] CVRM-Box[b,c]
At risk with uncontrolled BP, intervention	94 (53 to 150)	Gamma	[13] CVRM-Box[bc]
At risk with controlled BP, CaU[i]	61 (45 to 81)	Gamma	ELAN[c]
At risk with uncontrolled BP, CaU	76 (56 to 101)	Gamma	ELAN[c]
Incremental cost, post-MI[j]	1762 (1473 to 1996)	Triangular	[25,26][c]
Incremental cost, poststroke[j]	9609 (7897 to 12,084)	Triangular	[27,28][c]
Incremental cost, recurrent MI[k]	1059 (204 to 1857)	Triangular	[25,26][c]
Incremental cost, recurrent stroke[k]	1475 (290 to 2514)	Triangular	[27,28][c]
Transition costs			
MI event	13,929 (7612 to 22,009)	Triangular	[25,26,29-31][c]
Stroke event	16,606 (8842 to 23,266)	Triangular	[27,28,32-35][c]
Fatal cardiovascular event	13,929 (7612 to 22,009)	Triangular	Assumed[c]
Productivity loss after MI/stroke	19,860 (18,345 to 21,229)	Triangular	[27,32,36-40][c]
Utilities and disutilities			
At-risk, controlled BP	0.830 (0.791 to 0.868)	Beta	CVRM-Box[b,c]
Annual disutility, uncontrolled BP[j]	0.062 (0.031 to 0.098)	Triangular	[41-43][c]
Annual disutility, post-MI[j]	0.079 (0.023 to 0.141)	Triangular	[41-43][c]
Annual disutility, poststroke[j]	0.118 (0.052 to 0.201)	Triangular	[41-46][c]
Annual disutility, recurrent MI[k]	0.080 (0.019 to 0.140)	Triangular	[47,48][c]

XSL•FO
RenderX

Model input parameter	Estimate (95% CrI)	Distribution[a]	Source
Annual disutility, recurrent stroke[k]	0.063 (0.013 to 0.105)	Triangular	[32,44,46][c]
Transition disutilities			
MI event	0.111 (0.047 to 0.174)	Triangular	[43,48,49][c]
Stroke event	0.139 (0.028 to 0.231)	Triangular	[32,44,49][c]
Stroke event after MI	0.341 (0.291 to 0.388)	Beta	[49][c]

[a]Assumed posterior distribution for probabilistic resampling.

[b]Referring to the matched prospective study of CVRM-Box described in the Introduction.

[c]Further details on parametrization are shown in Table S1 in Multimedia Appendix 1.

[d]BP: blood pressure.

[e]CVD: cardiovascular disease.

[f]MI: myocardial infarction.

[g]SBP: systolic BP.

[h]EUR €1=US $1.11 as of 29 December 2023.

[i]CaU: care as usual.

[j]Incremental cost or disutility relative to the at-risk state.

[k]Incremental cost or disutility relative to the at-risk state and post-MI/poststroke state.

The cost-effectiveness of CVRM-Box was described using incremental cost-effectiveness ratios (ICERs) and cost-effectiveness acceptability curves [50]. We additionally conducted 2 budget impact analyses (BIAs) and estimated annual reductions in workload for GPs and PNs. Reporting followed the CHEERS (Consolidated Health Economic Evaluation Reporting Standards; the completed CHEERS reporting checklist is available as Checklist 1) [51]. Finally, a user-friendly R Shiny app was developed [52] to facilitate interaction with and modifications to the model by decision-makers and researchers, such as assuming different values for key model parameters (eg, higher costs, lower intervention effects) to represent alternative scenarios or assumptions. The web-based app is accessible at jmheij.shinyapps.io/CVRM-Box-model.

Ethical Considerations

The present study was a model-based secondary analysis and involved no new participant recruitment or data collection. The prospective CVRM-Box study [13], which informed various model input parameters, was approved by the Medical Ethics Committee of Leiden University Medical Center (N21.126), and participants provided written online informed consent. The same committee determined that the ELAN data resource was not subject to the Dutch Medical Research Involving Human Subjects Act (reference G18.070); individual informed consent was therefore waived, and individuals could withdraw through an informed opt-out procedure [24]. Analyses were conducted using deidentified CVRM-Box data and anonymized ELAN data, with access restricted to authorized personnel. Participants in the CVRM-Box study received no financial or other compensation.

Population and Setting

The modeled cohort represents primary care patients who are part of the Dutch integrated care programs for increased risk or history of CVD. For the main analysis, we assumed the population receives the intervention at age 50. In line with the prospective study of CVRM-Box, the cohort was 53% male and had a moderate to high risk (10%-15%) of 10-year CVD mortality [53] or a history of MI or stroke.

To account for patient and effect heterogeneity, 4 subgroup-specific analyses were conducted by BP status (ie, controlled/uncontrolled) and history of MI or stroke (ie, history/no history). The proportions of patients in each subgroup (Table 1) were consistent with the prospective study of CVRM-Box [13].

Interventions

CVRM-Box

The modeled CVRM-Box intervention reflects the multicomponent RPM program evaluated in van Hattem et al [13], which was a 12-month matched cohort study across 6 Dutch primary care practices. Because the present model is stratified by BP-control status, the BP-effect parameters used here (Table 1) were taken from the subgroup-specific estimates of that study: an adjusted office-measured SBP difference of –1.3 (95% CI –4.1 to 0.5) mm Hg in patients with controlled BP at baseline and –3.5 (95% CI –7.5 to 0.5) mm Hg in patients with uncontrolled BP, neither reaching conventional statistical significance. Despite the modest office-based effect, the intervention significantly increased the proportion of patients achieving controlled BP (<140 mm Hg) over follow-up, and home (CVRM-Box) BP measurements showed a larger reduction than office measurements (SBP –5.5 mm Hg, P<.001). The intervention was also associated with statistically significant reductions in body weight (–0.9 kg) and BMI (–0.3 kg/m²) and a lower overall consultation frequency (rate ratio 0.82, P=.002). These primary-data estimates, particularly the subgroup-specific SBP differences and the consultation-frequency reduction, directly informed the intervention effect sizes and resource-use parameters used in the present model.

XSL•FO
RenderX

Eligible patients underwent a consultation with an assistant who introduced "The CVRM-Box." The CVRM-Box was developed within a regional multistakeholder innovation environment, "Gezonde Zorg, Gezonde Regio" [14,54], a multistakeholder initiative involving health care providers and a major regional health insurer. This collaboration included long-term funding agreements, in which efficiency gains were redistributed between providers and the insurer. This structure aimed to align financial incentives and support sustainable implementation of digital care [55]. The intervention included a digital health infrastructure comprising a BP monitor (Wireless Blood Pressure Monitor; Withings), weighing scale (Smart Body Scale Analyzer; Withings), activity tracker (Move, Withings), and a dedicated mobile-based lifestyle app (LUMC Care) integrated in hospital and primary care electronic health records (Medicom, PharmaPartners) (Figure 1A). To facilitate effective use of the CVRM-Box at home, patients were supported by a GP assistant (a student) during dedicated consultations. These sessions included assistance with pairing the monitoring devices to the patient's smartphone, guidance on navigating the CVRM-Box app, and time to address questions or concerns. Patients were instructed to perform weekly BP and weight measurements and were provided with practical advice via the mobile-based lifestyle app on interpreting their results. In addition, individualized targets for weight and daily steps were defined. Patients were instructed to contact the PN in case of very high BP readings (>180 mm Hg), repeated elevated measurements (>140 mm Hg on 3 consecutive occasions), persistently increased values, or the occurrence of symptoms such as dizziness, headache, or chest pain. A follow-up appointment within 6 weeks was scheduled to resolve potential technical problems. The GP assistant also managed a help desk service that patients could reach for technical issues with their devices.

Measurement results were stored locally on the patient's smartphone and subsequently transmitted to the servers of the respective app providers (Withings). To safeguard participant privacy, each patient was assigned an @hlc.nl email address consisting of a randomly generated identifier and password, which was used solely for study-related logins. The @hlc.nl domain was managed by the LUMC, and all associated data were stored on secure LUMC servers. Data from Withings devices were retrieved through the Withings Application Programming Interface. The Withings Application Programming Interface enabled automatic transfer of device data into the hospital and GP electronic medical record system using a secure OAuth2 authentication process.

All transmitted measurements were reviewed at least weekly by the PN using a built-in dashboard within the electronic patient record. Because this dashboard was embedded within the existing workflow, PNs did not need to review incoming measurements through a parallel/separate monitoring system, limiting additional cognitive burden and disruption. Automated alerts flagged abnormal values, prompting the nurse to contact patients when needed. Depending on the findings, medication adjustments could be made or follow-up appointments scheduled. If no mobile health measurements were received for 3 consecutive weeks, GP assistants contacted patients by phone to emphasize the importance of continued monitoring. Further details about the CVRM-Box intervention are described elsewhere [13].

CaU

CaU was defined as the current CVRM practice within the primary integrated care programs for individuals with increased cardiovascular risk or a history of CVD. This protocolized program focused on lifestyle modifications and the treatment of BP and cholesterol levels and was primarily conducted by the PN [56]. Yearly check-ups at the GP included a local BP measurement and laboratory examination. Most visits (95%) were assumed to be with the PN, while the other visits (5%) took place with the GP when necessary. Thus, routine cardiovascular monitoring in both care pathways was predominantly PN-led, with limited direct GP involvement.

Model Structure

We developed a cohort Markov model with 1-year cycles to compare CVRM-Box versus CaU over a lifetime horizon. The underlying disease-progression structure (health states and allowed transitions) was identical between arms (Figure 1B), but the disease progression rates differed by arm. Specifically, baseline risk-factor profiles and CaU costs were informed by ELAN routine primary care data [23,24], while CVRM-Box intervention effects on SBP (by BP-control subgroup) and intervention-related costs were sourced from the prospective matched cohort study [13]. Cardiovascular event risks for the transitions shown in Figure 1B were generated using SCORE2/SCORE2-OP [57,58] and SMART2 [59] risk prediction equations populated with subgroup-specific mean risk factors and applying CVRM-Box–related changes in SBP in the intervention arm (more details in Transition probabilities subsection).

The Markov model (Figure 1B) simulated 3 cardiovascular events: nonfatal MI, nonfatal stroke, or fatal cardiovascular event. Additionally, the "Postrecurrent" states were included to account for the greater risk and consequences of recurrent nonfatal events. As the costs and burden of stroke are generally greater than those of MI [25,27,41,49,60], the model does not allow for transitions from stroke to MI states. Finally, patients could experience a fatal cardiovascular event or die of other causes. Patients with no CVD history entered the model in the "At risk" state, while patients with a CVD history entered the model in the post-MI (64%) and poststroke (36%) states, in line with the proportions observed in ELAN data [23,24].

Model Input Parameters

Probabilistic Approach

The main analysis was probabilistic, whereby parameter estimates were resampled 1000 times based on their SEs and assumed posterior distributions [50]. This probabilistic approach enabled us to reflect uncertainty in the available evidence (eg, the uncertainty surrounding key parameters) by running the model many times with plausible alternative input values, so results could be reported as ranges and the probability that CVRM-Box represents good value for money. All model parameters with corresponding sources [10,25-49] and

XSL•FO
RenderX

uncertainty ranges are summarized in Table 1 (details in Table S1 of Multimedia Appendix 1).

Transition Probabilities

The transitions in Figure 1B were simulated using age- and subgroup-specific transition probability matrices. Transitions from the "At-risk" state to post-MI/stroke or cardiovascular death were parameterized using the SCORE2 and SCORE2-OP prediction equations, including the appropriate recalibration scales for the Netherlands (ie, corresponding to the "low risk" region) [57,58]. Transitions between post-MI/stroke states and from post-MI/stroke to cardiovascular death were parameterized using the SMART2 prediction equations, including the appropriate recalibration ratio for the Netherlands (ie, corresponding to the "low risk" region) [59]. All prediction equations were populated with subgroup-specific mean risk factor values (ie, cholesterol, BP, etc, reported in Tables S2-S7 of Multimedia Appendix 1) obtained from ELAN data [23,24] (representative of CaU). All-cause mortality followed population life tables corrected for CVD-related mortality [61,62]. Transition probability calculations are detailed in Multimedia Appendix 2.

The effect of CVRM-Box was modeled as a reduction in the absolute risk of MACE (ie, lower transition probabilities) resulting from an expected reduction in BP and smoking with the intervention (Table 1). The reduction in BP was equivalent to the adjusted difference in BP estimated in the matched prospective study of CVRM-Box (Table 1). Specifically, the intervention study of CVRM-Box reported an adjusted mean reduction in SBP of −1.3 mm Hg among patients with controlled BP and a −3.5 mm Hg reduction among patients with uncontrolled BP [13]. Smoking cessation was assumed at 13% of patients with a wide credible interval of 2% to 18% [10].

Costs

In line with Dutch guidelines for economic evaluations, we followed a societal perspective and discounted costs at 3.5% per yearly cycle [63].

Annual intervention and CaU costs for at-risk patients were estimated from the prospective study of CVRM-Box [13] and ELAN data [23,24], respectively, while the incremental annual costs for nondeath states were sourced from literature (Table 1). The literature sources informing the incremental annual costs for stroke states included a Dutch 2-year analysis of health care use and costs among stroke survivors in the Restore4Stroke Cohort [27], and a German lifetime cost analysis of stroke survivors [28]. For the incremental annual costs in MI states, sources included a 6-year cohort study among MI patients that reported annual health care use [25], and parameters used in a previous Dutch economic evaluation [26]. Nonhealth care costs included work productivity losses (ie, reduced working and/or work leave), estimated using the friction cost method [64] and accrued only while the cohort was younger than 67 years (Table 1). The degree and duration of work productivity loss was informed by ranges reported in various studies, most of which showed that work leave after a nonfatal CVD event lasts approximately 120 days and high proportions of early retirements [27,32,36-40].

Our modeling approach integrated state-specific and transition-specific costs (Table 1). State-specific costs represent the aforementioned annual/recurring costs associated with each health state, while transition-specific costs account for any one-time/short-term costs associated with the cardiovascular events themselves [65]. Existing literature that reported on transition-specific costs, such as initial hospitalization and treatment after a stroke [27,28,32-35] or MI event [25,26,29-31], were used to inform those parameters (Table 1 and Multimedia Appendix 1).

Finally, to avoid overestimating the economic benefits of preventing fatal cardiovascular events, we used the PAID 3.0 tool [63,66] to include future unrelated medical costs attributable to any increases in life expectancy with the intervention. The PAID 3.0 tool was developed to provide standardized, age- and sex-specific estimates of health care expenditures unrelated to the index condition that accrue in additional life-years following (preventive) interventions [63,66]. Incorporating these costs is recommended in economic evaluations to avoid overstating an intervention's net benefits [63].

Outcomes

Total quality-adjusted life years (QALYs) were calculated as the sum of utilities over all cycles, applying a discount rate of 1.5% per cycle [63]. Similarly to costs, our model accounts for state-specific utilities and transition-specific disutilities (all reported in Table 1). State-specific utilities represent the annual utility of patients *excluding* any acute, short-term burden associated with the acute transition events themselves (which were modeled as transition-specific disutilities).

The baseline annual utility of 0.83 for patients in the "At-risk, controlled BP" state was estimated using primary data from the matched prospective CVRM-Box cohort study [13]. From that baseline value, annual disutilities for uncontrolled BP and all postevent (stroke and MI) states were informed by published national catalogs of preference-based utility values across multiple chronic conditions (including CVD) [41-43], evidence from longitudinal studies on cardiovascular populations [32,44,48], population-based health-related quality of life analyses [45,47], and a systematic review/meta-analysis of utility weights in stroke patients [46]. Four of these studies also reported utility values that were used to derive our model's transition-specific disutility parameter estimates (Table 1) [32,43,44,48], as well as an additional vignette study [49].

Assumptions

As with every model, ours is subject to simplifying assumptions (details in Multimedia Appendix 2). First, there is no waning in effect nor change in adherence to the intervention, although this is addressed in later scenario analyses. Second, patients can only experience 1 recurrent event of the same type. Although third/fourth events occur in practice, these are much less frequent, and evidence to inform corresponding parameters is lacking. Third, the costs of a fatal event are equivalent to the lowest of the nonfatal events to avoid overestimating the benefit of preventing cardiovascular deaths.

XSL•FO
RenderX

Model Validation

The Technical Verification (TECH-VER) checklist was used to test our model's credibility and technical implementation [67]. The checklist consists of white-box and black-box tests designed to identify errors/inconsistencies in model implementation. The completed TECH-VER checklist is reported in Table S8 of Multimedia Appendix 1.

Uncertainty Analyses

Deterministic Sensitivity Analysis

To describe the influence of each model parameter on our results, we performed deterministic sensitivity analyses (DSAs) on all parameters assuming alternative (ie, 25% lower or 25% higher) point estimates.

Scenario Analyses

To explore the sensitivity of results to more general assumptions, 4 scenarios were explored: (1) starting age of 40 years, to explore how earlier receipt of the intervention influences cost-effectiveness; (2) the effectiveness of CVRM-Box decreases by 40% after 5 years, reflecting a reasonable reduction in adherence/discontinuation; (3) health care perspective (excluding nonhealth care costs) [63]; and (4) no discounting of costs and utilities [63].

BIA

To estimate the short- or medium-term budgetary requirements for rolling out CVRM-Box, we conducted 2 BIAs assuming rollout at the regional (South-Holland-North region) and national levels, with a time horizon of 5 years. The methodological approach and assumptions for both BIAs are detailed in Multimedia Appendix 2. Briefly, the South-Holland-North region has 10,256 patients eligible to receive the intervention over 5 years (8344 in the first year) [68]. For the national-level BIA, the corresponding eligible population was estimated at 2,177,350 (with 1,917,812 in the first year) based on publicly available data [69].

Workload Reduction in Care Delivery

RPM interventions may by design result in lower provider workload, an outcome not completely captured by cost-effectiveness analyses or BIAs. Therefore, we conducted an additional exploratory workload analysis to estimate potential savings in time and resource utilization with the intervention. We estimated the annual reduction/change in (1) time spent on patient contacts (short and long visits) and (2) number of e-consultations per GP and PN in a reference Dutch primary care practice (in Dutch, a "normpraktijk"). A reference primary care practice in the Netherlands has on average 1 GP, 0.83 PNs, and 2100 registered patients [70], of which 420 have a history or high risk of CVD [68].

To estimate the amount of time spent on care by GPs and PNs annually, we used data from the CVRM-Box prospective study [13] about the frequency of e-consultations, short visits, and long visits with GPs and PNs (by subgroups of controlled and uncontrolled BP). These per-patient frequencies were converted to expected practice-level annual volumes by multiplying (for each BP-control subgroup and contact type) the relevant per-patient frequency by the expected number of eligible patients with CVD in a reference primary care practice: 2100 registered patients × 420/2100 patients with CVD × 55% assumed eligible=231 eligible patients per practice. This eligible practice population was then split into controlled and uncontrolled BP according to our main analysis proportions. Annual time requirements (hours) per provider were calculated by multiplying expected annual visit counts by standard scheduled visit durations and dividing by the number of practitioners per reference primary care practice (1 GP; 0.83 PN), while remote consultations were reported as annual counts per provider. Consistent with current practice in the Netherlands, a short visit was calculated to last 10 minutes (with the GP) and 15 minutes (with a PN), while long visits lasted 20 minutes (with a GP) and 30 minutes (with a PN), respectively. The complete formulas and assumptions are provided in Multimedia Appendix 2.

Results

Main Analysis

Overall, CVRM-Box was €924 (95% CrI –467 to 2238) more costly and yielded 0.05 (95% CrI –0.02 to 0.12) more QALYs than CaU, resulting in an ICER of €17,340 per QALY gained (Table 2) from a societal perspective (EUR €1=US $1.11 as of 29 December 2023). At the commonly applied willingness-to-pay threshold of €20,000/QALY for prevention in the Netherlands, the overall probability of CVRM-Box being cost-effective over CaU was 60% (Figure 2A).

In terms of subgroup-specific results, CVRM-Box was ≥70% likely to be cost-effective and had ICERs below €20,000/QALY in all subgroups except for patients with controlled BP and no history of MI/stroke (Table 2 and Figure 2A). For the latter subgroup, the higher ICER of €42,384/QALY was mainly explained by the fact that CVRM-Box did not meaningfully influence MACE rates in this group of patients, in which BP levels were already within target range.

XSL•FO
RenderX

Table . Costs, quality adjusted life years, life expectancy, and incremental cost-effectiveness ratios in base-case and scenario analyses[a].

Model outcome	Current practice	Cardiovascular remote management–Box
Health care costs (€[b]), mean (95% credible interval)		
Controlled, no history	6218 (4256 to 8431)	7131 (4960 to 9439)
Controlled, history of myocardial infarction (MI)/stroke	90,538 (63,338 to 120,369)	91,548 (64,820 to 121,417)
Uncontrolled, no history	7879 (5369 to 10,730)	8853 (6218 to 11,704)
Uncontrolled, history of MI/stroke	92,542 (64,159 to 123,850)	94,192 (66,664 to 125,263)
Pooled	21,476 (14,964 to 28,656)	22,459 (16,048 to 29,691)
Nonhealth care costs (€), mean (95% credible interval)		
Controlled, no history	603 (557 to 645)	591 (546 to 632)
Controlled, history of MI/stroke	9034 (8345 to 9657)	8870 (8190 to 9501)
Uncontrolled, no history	828 (764 to 885)	791 (727 to 854)
Uncontrolled, history of MI/stroke	11,336 (10,471 to 12,118)	10,947 (10,085 to 11,772)
Pooled	2252 (2080 to 2407)	2194 (2024 to 2353)
Total societal costs (€), mean (95% credible interval)		
Controlled, no history	6821 (4820 to 9076)	7722 (5517 to 10066)
Controlled, history of MI/stroke	99,572 (71,692 to 130,026)	100,418 (73,052 to 130,893)
Uncontrolled, no history	8706 (6144 to 11,615)	9643 (6964 to 12,518)
Uncontrolled, history of MI/stroke	103,878 (74,641 to 135,967)	105,140 (76,492 to 137,096)
Pooled	23,728 (17,038 to 31,066)	24,652 (18,087 to 32,022)
Quality-adjusted life years (QALYs), mean (95% credible interval)		
Controlled, no history	20.66 (19.66 to 21.61)	20.68 (19.68 to 21.63)
Controlled, history of MI/stroke	12.28 (10.34 to 13.98)	12.38 (10.41 to 14.12)
Uncontrolled, no history	18.74 (17.38 to 19.98)	18.79 (17.42 to 20.01)
Uncontrolled, history of MI/stroke	9.86 (7.56 to 11.89)	10.07 (7.74 to 12.15)
Pooled	18.43 (17.3 to 19.52)	18.48 (17.34 to 19.58)
Modeled life-expectancy in years, mean (95% credibel interval)		
Controlled, no history	32.16[c]	32.2 (32.13 to 32.26)
Controlled, history of MI/stroke	21.45	21.62 (21.31 to 21.94)
Uncontrolled, no history	31.45	31.57 (31.44 to 31.69)
Uncontrolled, history of MI/stroke	18.78	19.21 (18.73 to 19.7)
Pooled	29.93	30.03 (29.89 to 30.17)
Incremental cost-effectiveness ratio		
Controlled, no history	—[c]	€42,384/QALY
Controlled, history of MI/stroke	—	€8806/QALY
Uncontrolled, no history	—	€15,708/QALY
Uncontrolled, history of MI/stroke	—	€5849/QALY
Pooled	—	€17,340/QALY
Scenario analysis incremental cost-effectiveness ratios, pooled results		
Health care perspective	—	€18,434/QALY
No discounting	—	€26,882/QALY
Starting age of 40 years	—	€15,087/QALY

XSL•FO
RenderX

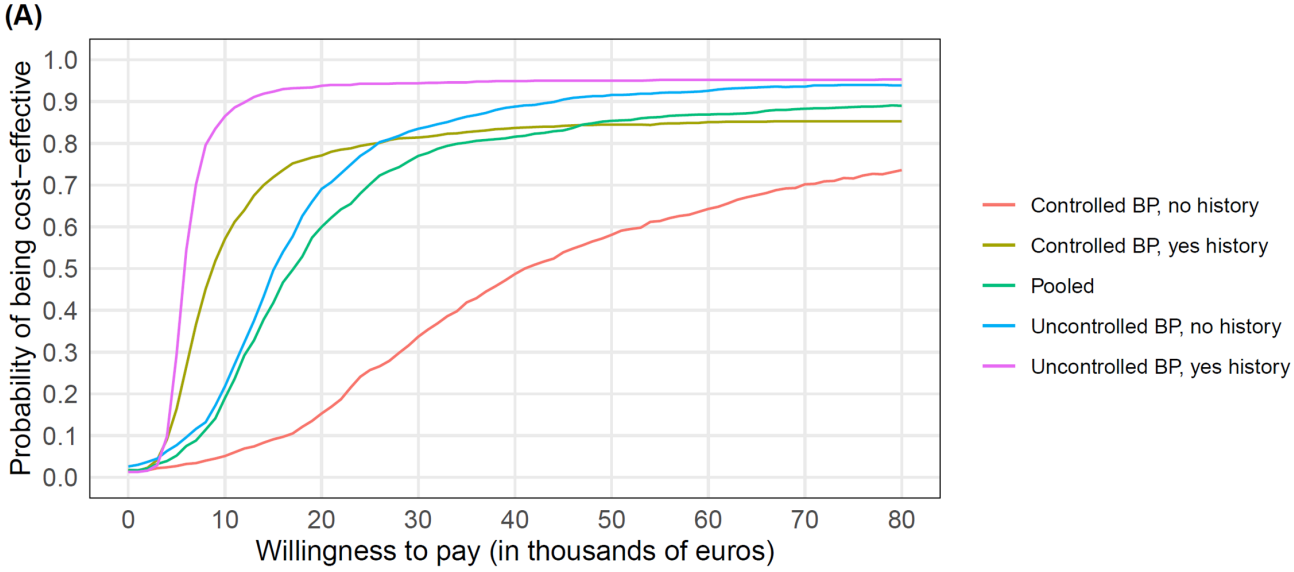
Model outcome	Current practice	Cardiovascular remote management–Box
Intervention effect reduces by 40% after 5 years	—	€23,110/QALY

[a]Costs and quality-adjusted life years are per patient lifetime in the population.

[b]EUR €1=US $1.11 as of 29 December 2023.

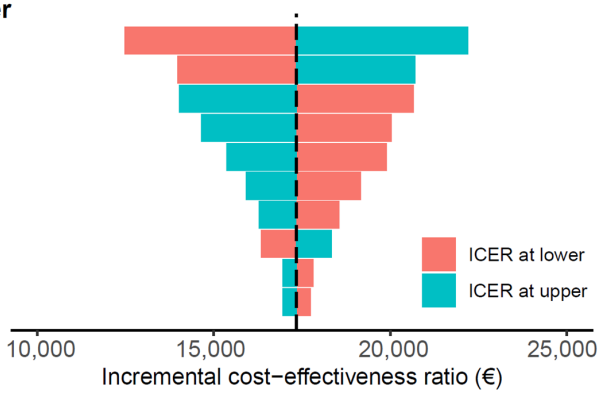
[c]Mean or 95% credible interval not applicable.

Figure 2. Cost-effectiveness acceptability curves (A) and deterministic sensitivity analysis results (B). (A) The curves represent the probability that CVRM-Box is cost-effective compared to CaU for each willingness to pay on the x-axis. (B) The "Base" values are the parameter values used in the main analysis (Table 1), while the "Lower" and "Upper" values are plausible alternative values. BP: blood pressure (systolic); CaU: care as usual; ICER: incremental cost-effectiveness ratio; SBP: systolic blood pressure.



(B)

Parameter	Base	Lower	Upper
Annual cost intervention, controlled BP	88	67	112
Annual cost intervention, uncontrolled BP	94	71	119
Annual cost under CaU, controlled BP	61	46	77
Annual cost under CaU, uncontrolled BP	77	57	96
Baseline utility (at−risk, no hypertension)	0.83	0.73	0.93
SBP reduction intervention, uncontrolled BP	3.5	2.7	4.5
SBP reduction intervention, controlled BP	1.36	1.02	1.71
Initial intervention cost	213	160	266
Smoking cessation proportion, intervention	0.11	0.08	0.14
Incremental state cost, poststroke	9609	7207	12011



DSA

The most influential model parameters were the annual costs of CVRM-Box and CaU (Figure 2B). The model's sensitivity to these parameters was largely due to the fact that QALY gains with the intervention were small, making differences in costs more influential (especially costs that apply to the entire cohort). All other parameters had a much lower influence on the estimated ICERs.

Scenario Analyses

The scenario assuming no discounting and the scenario assuming a waning effect of the intervention resulted in ICERs above €20,000/QALY, while a starting age of 40 decreased the overall ICER to €15,087/QALY (Table 2). Subgroup-specific scenario analysis results (Table S9 of Multimedia Appendix 1) showed that the ICERs for CVRM-Box were consistently below €20,000/QALY, with few exceptions in the subgroups with no history of MI/stroke, in line with our main analysis.

Budget Impact

The pooled budget impact of CVRM-Box was €2.9 million assuming regional rollout (10,256 patients) and €662.2 million assuming national rollout (1,917,812 patients) over 5 years. Nearly 60% of the overall budget impact was attributable to patients with controlled BP and no history of MI/stroke (Figure S1 of Multimedia Appendix 1), due to the fact that the higher intervention costs were not compensated by the (few) prevented events in this subgroup. Conversely, the budget impact was much lower in other subgroups, ranging from €0.26 to €1.01 million at the regional level and €59.3 to €232.5 million at the national level (Figure S1 of Multimedia Appendix 1). This was

XSL•FO
RenderX

especially the case in the subgroup with uncontrolled BP and a history of MI/stroke, where the 5-year budget impact was slightly negative.

Workload Reduction

Assuming a reference Dutch primary care practice, CVRM-Box would be expected to require more time spent on long visits with patients and less time spent on short visits each year (Table 3). The reduction in time spent on short visits compensated for the additional workload on long visits, resulting in a net annual reduction in workload that was substantial for PNs but negligible for GPs. Annually, PNs would spend 25.9 fewer hours on patient contacts and handle 44.7 fewer remote consultations, whereas the corresponding change for GPs was small (0.5 h and 3.3 fewer remote consultations). The greatest workload reductions were among PNs treating patients with uncontrolled BP (Table S10 of Multimedia Appendix 1).

Table . Annual expected reductions in workload with the intervention, per general practitioner and practice nurse in a reference Dutch primary care practice[a].

Care activity	CVRM[b]-Box, annual workload	Care as usual, annual workload	Annual increase (+) or reduction (−) with intervention
Per practice nurse			
Long patient contacts[c]	157.7 hours	147.2 hours	+10.5 hours
Short patient contacts[d]	94.7 hours	131.1 hours	–36.4 hours
All patient contacts (sum)	252.4 hours	278.3 hours	–25.9 hours
Remote consultations	167.0 consultations	211.7 consultations	–44.7 consultations
Per general practitioner			
Long patient contacts[c]	4.6 hours	4.2 hours	+0.4 hours
Short patient contacts[d]	2.8 hours	3.7 hours	–0.9 hours
All patient contacts (sum)	7.4 hours	7.9 hours	–0.5 hours
Remote consultations	6.9 consultations	10.2 consultations	–3.3 consultations

[a]"Hours" and "consultations" refer to annual amounts per general practitioner or practice nurse in a reference primary care practice assuming an eligible cardiovascular population of N=231 patients per practice (Multimedia Appendix 2).

[b]CVRM: cardiovascular remote management.

[c]Duration of 20 minutes with general practitioner and 30 minutes with practice nurse.

[d]Duration of 10 minutes with general practitioner and 15 minutes with practice nurse.

Discussion

Principal Findings

In this study, the long-term cost-effectiveness, budget impact, and workload reductions associated with the CVRM-Box implemented in primary care were assessed using a model-based evaluation. The CVRM-Box was highly likely to be cost-effective among patients with uncontrolled BP and/or a history of CVD, but not in patients with controlled BP and no history of CVD. Our BIA showed a similar pattern, where most of the incremental costs were attributable to this healthier subgroup. Together, these findings support considering the CVRM-Box primarily for high-risk subgroups, whereas broader adoption for lower-risk patients would require significant financial investment with insufficient QALY gains. Another key finding was that the CVRM-Box was expected to reduce the annual workload of PNs, who deliver the large majority of routine cardiovascular contacts in Dutch primary care. The corresponding change in GP workload was negligible, and because it is of a similar order to the one-off training and set-up required to adopt the intervention, we do not interpret it as a saving in GP time; the workload benefit is therefore essentially a PN benefit, which is where workforce pressure in Dutch primary care is greatest.

Our results were generally robust to extensive uncertainty analyses. Our DSA showed that the annual intervention costs were highly influential, suggesting that reducing the intervention costs would yield the greatest increase in cost-effectiveness. Our scenario analyses showed that delivering the intervention at a younger age (ie, 40 instead of 50 y) would improve cost-effectiveness, and that a 40% reduction in effectiveness/adherence after 5 years would not compromise cost-effectiveness in the 3 highest-risk subgroups (Table S9 of Multimedia Appendix 1).

Comparison With the Literature

Model-based evaluations of RPM interventions for CVD vary in methodology, making comparisons difficult. However, most studies—including ours—show cost-effectiveness for managing hypertension and long-term CVD [71-73]. In line with this, an economic evaluation of home-based cardiac telerehabilitation reported a high probability of cost-effectiveness despite nonsignificant differences in clinical outcomes and costs [74]. The BP reductions underlying our analysis are likewise consistent with this evidence base. Telemonitoring in the TASMINH4 trial lowered systolic pressure by 4.7 mm Hg relative to usual care [10], and comparable reductions are reported in an individual-patient-data meta-analysis of self-monitoring [75]. Reductions of this magnitude remain

XSL•FO
RenderX

clinically relevant at the population level, as an approximately 5 mm Hg systolic reduction has been associated with around a 10% lower risk of major cardiovascular events in a large individual-participant-data meta-analysis of BP-lowering treatment [76], corresponding to a meaningful absolute reduction in events for a scalable primary care population. Our findings contribute to ongoing discussions about the cost-effectiveness assessment processes for digital health interventions. In recent years, the focus has shifted from solely clinical outcomes and cost-effectiveness as value indicators to incorporating clinical workflows and health care professionals' perspectives [14,77,78]. As also found in our previous implementation studies of the CVRM-Box [14,15], stakeholders suggest that RPM's impact on health care professional workload is increasingly important, particularly in the context of labor shortages in primary care [14]. Despite this, current practices still prioritize (cost-)effectiveness and BIAs, while considerations related to workload reductions have not yet become mainstream. These indicators are high on the political agenda and are important to decision-makers as factors for the value of RPM in practice. We partly addressed this knowledge gap through our estimation of expected reductions in annual workload for GPs and PNs assuming scale-up in Dutch primary care centers. That said, our workload estimates should be interpreted as exploratory, as they are based on observed consultation frequencies from the prospective CVRM-Box study translated to a Dutch reference practice, and are not embedded within the state-transition model.

Importantly, scaling potential is not determined by economic considerations alone (eg, cost-effectiveness and budget impact). Our previous implementation studies identified additional practical factors influencing large-scale adoption [14,15]. Key barriers included additional clinical and technical tasks for health care professionals and uncertainties regarding reimbursement arrangements, while facilitators included early detection of at-risk patients, the presence of an implementation manager overseeing procedures and implementation metrics, and ambassador roles supporting adoption. Together, these findings reflect that economic evidence should be complemented by attention to practical implementation challenges when considering large-scale deployment of RPM interventions.

In a previous BIA study [79] where upscaling scenarios were calculated from one hospital ward to all Dutch hospitals, it was found that replacing in-hospital care with virtual care did not directly lead to cost savings within a single hospital ward due to labor-intensive virtual care services. The study suggested that minimal involvement of health care professionals, as applied in our intervention, was necessary. Similar to our findings, the previous study noted that scaling regionally (whole hospital) or nationally (all Dutch hospitals) resulted in cost savings relatively quickly. However, they observed cost savings for all patient groups, while we estimated potential cost savings exclusively in the subgroup with uncontrolled hypertension and a history of CVD. This discrepancy is likely due to differences in patient populations.

Strengths and Limitations

Economic evaluations always rely on assumptions, and despite our validation efforts using the TECH-VER checklist [67], several limitations remain. First, only 3 cardiovascular events were considered. Our choice of events, despite excluding other events like heart failure and sudden cardiac arrest, enabled us to simulate patient transitions using the state-of-the-art SCORE2 and SMART2 risk prediction models, which also focus on MI, stroke, and cardiovascular death. Importantly, long-term cardiovascular outcomes in our model were estimated indirectly through the SCORE2 and SMART2 risk prediction equations (rather than observed cardiovascular events), adding uncertainty to the modeled event rates. Second, the prospective CVRM-Box study had an effect on BMI, but we were not able to account for it due to lacking risk prediction models that include BMI as a predictor. This is a minor limitation, however, as reductions in BMI and BP are closely related, and including an independent effect on BMI would have only resulted in lower ICERs. Third, our exploratory analysis of workload reduction was subject to a higher degree of uncertainty than our main model-based analysis, as it did not use the state-transition model due to limited information about (changes in) consultation frequency per disease state. Nonetheless, even our indicative workload estimates are important for decision-makers given the current lack of data and literature on this question. Future studies should prioritize data collection on workload that can be incorporated into state-transition models to obtain more robust results. Relatedly, the estimated workload reductions should be interpreted as potential reductions in routine patient-contact time annually rather than net implementation time savings, as one-off implementation activities (eg, short training for GPs and PNs and ad hoc troubleshooting) were excluded from these calculations. As these implementation-related time investments apply to the practice population as a whole, their contribution to average annual workload per patient is expected to be negligible. Finally, importantly, our model did not account for transitions from uncontrolled to controlled BP status, potentially leading to more conservative estimates of cost-effectiveness. A microsimulation would address these challenges more effectively and allow for unique patient trajectories to be modeled.

A further consideration concerns the evidence on the intervention effect itself. The office-measured systolic reduction did not reach conventional statistical significance and was derived from a matched, controlled pre-post study rather than a randomized trial. Because the main analysis is probabilistic, the uncertainty in this estimate, including its CI, is carried through to the incremental cost and QALY distributions, which is why our cost-effectiveness conclusions are restricted to the higher-risk subgroups. A controlled pre-post design with broad inclusion criteria, such as that of the source prospective study [13], may also reflect the real-world effect more closely than a selective trial, and the proof of concept for home BP monitoring is already established in randomized trials. A placebo or regression-to-the-mean component in the home measurements, which lacked a control group, cannot be fully excluded, although these measurements avoid the well-known white-coat response [80,81] and the economic analysis relies on the office estimates that were matched to a control group.

Despite these limitations, our study has several strengths. A significant strength is our use of primary data from the matched

XSL•FO
RenderX

cohort study of CVRM-Box to inform key model parameters. Other key parameters, including risk factor data for the SCORE2 and SMART2 risk prediction models, came from primary data from the comprehensive ELAN database, which is representative of the current primary care CVD population in a large Dutch region [24]. Our extensive sensitivity and scenario analyses transparently reflect the influence of key assumptions and uncertainties on results, strengthening our overall conclusions.

Implications for Future Research and Practice

This study highlights the importance of incorporating diverse value indicators (ie, cost-effectiveness, budget impact, and workload) into the evaluation of digital health interventions. Payers often rely on payment- and results-based models to align incentives among payers, providers, and patients [77]. By integrating workload for health care professionals into the cost-effectiveness framework, our study offers a more comprehensive perspective to aid decision-making. Workload reductions may substantially influence resource allocation and reimbursement decisions for RPM interventions, particularly in the context of an increasingly constrained primary care workforce. However, real-world implementation decisions should also account for contextual factors not explicitly modeled here, including local staffing capacity, digital infrastructure and support requirements (eg, device logistics, interoperability with electronic health records), and equity considerations. Notably, provider incentive structures may also influence adoption and generalizability across health care systems. In systems with substantial capitation components, including Dutch primary care, the revenue impact of reduced in-person consultations is attenuated relative to predominantly fee-for-service settings, where additional incentive arrangements may be needed to support uptake. Beyond these payment arrangements, several further features of the Dutch setting shape how our findings should be read elsewhere. CVRM here is delivered through protocolized, integrated primary care programs that are largely PN-led, as described above, and is assessed against a relatively explicit national willingness-to-pay threshold for prevention; the intervention itself was financed through the regional shared-savings arrangement outlined in the *Methods*. Where care is instead more physician-led or fee-for-service, where a different cost-effectiveness threshold applies, or where such shared-savings arrangements are absent, the cost, workload, and adoption implications of CVRM-Box may differ appreciably even when its clinical effect is similar. The relative cost-effectiveness ranking across risk subgroups is therefore likely to transfer more readily than the absolute budget-impact and workload estimates, which are the most context-dependent and would need to be reestimated locally. Finally, future research should incorporate a wider range of value indicators focusing on key aspects such as provider workload, and should consider wider factors that may affect implementation.

Conclusion

This model-based evaluation of CVRM-Box found that the intervention is likely to be cost-effective in high-risk groups (uncontrolled hypertension or a history of CVD) and may reduce workload for health care professionals, while results are less favorable in lower-risk patients. In light of these findings, our findings support considering implementation in primary care, particularly for higher-risk subgroups. Regional considerations and subgroup-specific outcomes should be considered alongside other key factors such as local capacity or infrastructure and equity in decisions about the introduction and implementation of similar interventions. Importantly, because digital health interventions interact closely with the local health system, our findings should be interpreted primarily within the Dutch primary care context and may not directly generalize to health care systems with different organization, prices, or incentives.

Acknowledgments

The authors declare the use of generative AI (GenAI) in the writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: (i) proofreading and editing to improve readability and clarity of ideas/messages and (ii) summarizing text to improve readability.

The GenAI tools used were ChatGPT (GPT-5) and ChatGPT (GPT-5.2; OpenAI). Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes. Declaration submitted by all coauthors.

Funding

This study was funded by a grant from the Innovative Medical Devices Initiative (IMDI), LSHM21009, and Zorginstituut Nederland, "Subsidieregeling leren gebruiken van uitkomstinformatie voor Samen beslissen 2021-2025," 3463471-1039183-PZO. The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data Availability

All the code necessary to reproduce our results is accessible via GitHub [82]. Various datasets with different restrictions were used in this study. The Extramuraal Leiden University Medical Center Academisch Netwerk (ELAN) dataset is accessible via application to the dataset owners via ELAN Research [83]. The primary data from the CVRM prospective study can be made accessible upon reasonable request to co-first author MR.

XSL•FO
RenderX

Authors' Contributions

JHS and MR had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. JHS and MR are co-first authors. JHS, MR, HJAO, EH, and MEAM contributed to drafting the manuscript. JHS and MR conducted the statistical analysis. HJAO, MR, TNB, and EH obtained funding. HJAO, MEAM, and EH provided supervision. All authors (JHS, MR, NH, EH, MH, XZ, DA, TNB, MEAM, HJAO) contributed to the concept and design of the study, acquisition, analysis, or interpretation of data, and critical review of the manuscript for important intellectual content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Tables S1 to Table S10, and Figure S1.
[DOCX File, 219 KB - mhealth_v14i1e79512_app1.docx]

Multimedia Appendix 2
Methodological details.
[DOCX File, 112 KB - mhealth_v14i1e79512_app2.docx]

Checklist 1
CHEERS reporting checklist.
[DOCX File, 18 KB - mhealth_v14i1e79512_app3.docx]

References

1. Algra A, Wermer MJH. Stroke in 2016: stroke is treatable, but prevention is the key. Nat Rev Neurol 2017 Feb;13(2):78-79. [doi: 10.1038/nrneurol.2017.4] [Medline: 28106062]
2. Bushnell C, McCullough LD, Awad IA, et al. Guidelines for the prevention of stroke in women: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2014 May;45(5):1545-1588. [doi: 10.1161/01.str.0000442009.06663.48] [Medline: 24503673]
3. GP Workforce Team (National Health Service UK). General and Personal Medical Services, England : Final 31 December 2017 and Provisional 31 March 2018, Experimental Statistics. : NHS Digital; 2018 URL: https://files.digital.nhs.uk/A5/41FD68/General%20and%20Personal%20Medical%20Services%20-%20Provisional%20March%202018.pdf [accessed 2026-09-02]
4. Flinterman L, Groenewegen P, Verheij R. Zorglandschap en zorggebruik in een veranderende eerste lijn [Report in Dutch]. : Nivel; 2018 URL: https://www.nivel.nl/sites/default/files/bestanden/Zorglandschap_zorggebruik_in_de_eerste_lijn.pdf [accessed 2026-09-02]
5. Batenburg R, Bosmans M, Vandermeulen L, Versteeg S. Balans in vraag en aanbod huisartsenzorg [Report in Dutch]. : Nivel, Prismant; 2018 URL: https://www.nivel.nl/sites/default/files/bestanden/Balans_in_vraag_en_aanbod_huisartsenzorg.pdf [accessed 2026-09-02]
6. Widmer RJ, Collins NM, Collins CS, West CP, Lerman LO, Lerman A. Digital health interventions for the prevention of cardiovascular disease: a systematic review and meta-analysis. Mayo Clin Proc 2015 Apr;90(4):469-480. [doi: 10.1016/j.mayocp.2014.12.026] [Medline: 25841251]
7. McManus RJ, Mant J, Haque MS, et al. Effect of self-monitoring and medication self-titration on systolic blood pressure in hypertensive patients at high risk of cardiovascular disease: the TASMIN-SR randomized clinical trial. JAMA 2014 Aug 27;312(8):799-808. [doi: 10.1001/jama.2014.10057] [Medline: 25157723]
8. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146. [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
9. McManus RJ, Little P, Stuart B, et al. HOME and Online Management and Evaluation of Blood Pressure (HOME BP) using a digital intervention in poorly controlled hypertension: randomised controlled trial. BMJ 2021 Jan 19;372:m4858. [doi: 10.1136/bmj.m4858] [Medline: 33468518]
10. McManus RJ, Mant J, Franssen M, et al. Efficacy of self-monitored blood pressure, with or without telemonitoring, for titration of antihypertensive medication (TASMINH4): an unmasked randomised controlled trial. Lancet 2018 Mar 10;391(10124):949-959. [doi: 10.1016/S0140-6736(18)30309-X] [Medline: 29499873]
11. Kotseva K, De Backer G, De Bacquer D, et al. Lifestyle and impact on cardiovascular risk factor control in coronary patients across 27 countries: results from the European Society of Cardiology ESC-EORP EUROASPIRE V registry. Eur J Prev Cardiol 2019 May;26(8):824-835. [doi: 10.1177/2047487318825350] [Medline: 30739508]

XSL•FO
RenderX

12. van der Ende MY, Sijtsma A, Snieder H, van der Harst P. Letter to editor: reply on question of Marques JR et al. regarding the paper entitled: “The LifeLines cohort study: prevalence and treatment of cardiovascular disease and risk factors”. Int J Cardiol 2019 Nov 1;294:57. [doi: 10.1016/j.ijcard.2019.06.026] [Medline: 31522719]
13. van Hattem NE, Rakers MM, Hiddink EG, et al. Effectiveness of remote monitoring for patients with a high risk of cardiovascular disease: a 12-month matched cohort study in primary care. Eur Heart J Digit Health 2026;7(2):ztag014. [doi: 10.1093/ehjdh/ztag014] [Medline: 41704580]
14. Rakers M, van Hattem N, Simic I, et al. Tailoring remote patient management in cardiovascular risk management for healthcare professionals using panel management: a qualitative study. BMC Prim Care 2024 Apr 20;25(1):122. [doi: 10.1186/s12875-024-02355-y] [Medline: 38643103]
15. Rakers M, van Hattem N, Hiddink E, et al. Tailoring remote patient management to optimise cardiovascular risk management in primary care: a mixed-methods implementation study informing large-scale implementation. BMC Prim Care 2025 Jul 2;26(1):214. [doi: 10.1186/s12875-025-02906-x] [Medline: 40604485]
16. McNamee P, Murray E, Kelly MP, et al. Designing and undertaking a health economics study of digital health interventions. Am J Prev Med 2016 Nov;51(5):852-860. [doi: 10.1016/j.amepre.2016.05.007] [Medline: 27745685]
17. Gomes M, Murray E, Raftery J. Economic evaluation of digital health interventions: methodological issues and recommendations for practice. Pharmacoeconomics 2022 Apr;40(4):367-378. [doi: 10.1007/s40273-022-01130-0] [Medline: 35132606]
18. Petticrew M. When are complex interventions 'complex'? When are simple interventions 'simple'? Eur J Public Health 2011 Aug;21(4):397-398. [doi: 10.1093/eurpub/ckr084] [Medline: 21771736]
19. Angelis A, Kanavos P. Value-based assessment of new medical technologies: towards a robust methodological framework for the application of multiple criteria decision analysis in the context of health technology assessment. Pharmacoeconomics 2016 May;34(5):435-446. [doi: 10.1007/s40273-015-0370-z] [Medline: 26739955]
20. Rutten-van Mölken M, Leijten F, Hoedemakers M, et al. Strengthening the evidence-base of integrated care for people with multi-morbidity in Europe using Multi-Criteria Decision Analysis (MCDA). BMC Health Serv Res 2018 Jul 24;18(1):576. [doi: 10.1186/s12913-018-3367-4] [Medline: 30041653]
21. Gupta N, Verma R, Dhiman RK, Rajsekhar K, Prinja S. Cost-effectiveness analysis and decision modelling: a tutorial for clinicians. J Clin Exp Hepatol 2020;10(2):177-184. [doi: 10.1016/j.jceh.2019.11.001] [Medline: 32189934]
22. Briggs A, Sculpher M. An introduction to Markov modelling for economic evaluation. In: Mallarkey G, editor. Pharmacoeconomics 1998 Apr;13(4):397-409. [doi: 10.2165/00019053-199813040-00003] [Medline: 10178664]
23. Ardesch FH, Meulendijk MC, Kist JM, et al. The introduction of a data-driven population health management approach in the Netherlands since 2019: the Extramural LUMC Academic Network data infrastructure. Health Policy 2023 Jun;132:104769. [doi: 10.1016/j.healthpol.2023.104769] [Medline: 37018883]
24. Kist JM, Vos HMM, Vos RC, et al. Data resource profile: Extramural Leiden University Medical Center Academic Network (ELAN). Int J Epidemiol 2024 Jun 12;53(4):dyae099. [doi: 10.1093/ije/dyae099] [Medline: 39049713]
25. Janzon M, Henriksson M, Hasvold P, Hjelm H, Thuresson M, Jernberg T. Long-term resource use patterns and healthcare costs after myocardial infarction in a clinical practice setting: results from a contemporary nationwide registry study. Eur Heart J Qual Care Clin Outcomes 2016 Oct 1;2(4):291-298. [doi: 10.1093/ehjqcco/qcw019] [Medline: 29474723]
26. Greving JP, Visseren FLJ, de Wit GA, Algra A. Statin treatment for primary prevention of vascular disease: whom to treat? Cost-effectiveness analysis. BMJ 2011 Mar 30;342:d1672. [doi: 10.1136/bmj.d1672] [Medline: 21450800]
27. van Mastrigt G, van Heugten C, Visser-Meily A, Bremmers L, Evers S. Estimating the burden of stroke: two-year societal costs and generic health-related quality of life of the Restore4Stroke cohort. Int J Environ Res Public Health 2022 Sep 5;19(17):11110. [doi: 10.3390/ijerph191711110] [Medline: 36078828]
28. Kolominsky-Rabas PL, Heuschmann PU, Marschall D, et al. Lifetime cost of ischemic stroke in Germany: results and national projections from a population-based stroke registry: the Erlangen Stroke Project. Stroke 2006 May;37(5):1179-1183. [doi: 10.1161/01.STR.0000217450.21310.90] [Medline: 16574918]
29. Cowper PA, Knight JD, Davidson-Ray L, et al. Acute and 1-year hospitalization costs for acute myocardial infarction treated with percutaneous coronary intervention: results from the TRANSLATE-ACS registry. J Am Heart Assoc 2019 Apr 16;8(8):e011322. [doi: 10.1161/JAHA.118.011322] [Medline: 30975005]
30. Soekhlal RR, Burgers LT, Redekop WK, Tan SS. Treatment costs of acute myocardial infarction in the Netherlands. Neth Heart J 2013 May;21(5):230-235. [doi: 10.1007/s12471-013-0386-y] [Medline: 23456884]
31. Sülz S, Wagenaar H, van de Klundert J. Have Dutch hospitals saved lives and reduced costs? A longitudinal patient-level analysis over the years 2013-2017. Health Econ 2021 Sep;30(10):2399-2408. [doi: 10.1002/hec.4391] [Medline: 34251075]
32. van Eeden M, van Heugten C, van Mastrigt G, van Mierlo M, Visser-Meily JMA, Evers S. The burden of stroke in the Netherlands: estimating quality of life and costs for 1 year poststroke. BMJ Open 2015 Nov 27;5(11):e008220. [doi: 10.1136/bmjopen-2015-008220] [Medline: 26614618]
33. Geessinck FAJ, Pleijhuis RG, Mentink RJ, van der Palen J, Koffijberg H. Cost-effectiveness analysis of the DiagnOSAS screening tool compared with polysomnography diagnosis in Dutch primary care. J Clin Sleep Med 2018 Jun 15;14(6):1005-1015. [doi: 10.5664/jcsm.7170] [Medline: 29852908]

XSL•FO
RenderX

34. Buisman LR, Tan SS, Nederkoorn PJ, Koudstaal PJ, Redekop WK. Hospital costs of ischemic stroke and TIA in the Netherlands. Neurology 2015 Jun 2;84(22):2208-2215. [doi: 10.1212/WNL.0000000000001635] [Medline: 25934858]
35. Baeten SA, van Exel NJA, Dirks M, Koopmanschap MA, Dippel DW, Niessen LW. Lifetime health effects and medical costs of integrated stroke services - a non-randomized controlled cluster-trial based life table approach. Cost Eff Resour Alloc 2010 Nov 17;8:21. [doi: 10.1186/1478-7547-8-21] [Medline: 21083901]
36. Steen Carlsson K, Nilsson K, Wolden ML, Faurby M. Economic burden of atherosclerotic cardiovascular disease: a matched case-control study in more than 450,000 Swedish individuals. BMC Cardiovasc Disord 2023 Sep 29;23(1):483. [doi: 10.1186/s12872-023-03518-y] [Medline: 37773098]
37. Gordois AL, Toth PP, Quek RGW, Proudfoot EM, Paoli CJ, Gandra SR. Productivity losses associated with cardiovascular disease: a systematic review. Expert Rev Pharmacoecon Outcomes Res 2016 Dec;16(6):759-769. [doi: 10.1080/14737167.2016.1259571] [Medline: 27831848]
38. Isaaz K, Coudrot M, Sabry MH, et al. Return to work after acute ST-segment elevation myocardial infarction in the modern era of reperfusion by direct percutaneous coronary intervention. Arch Cardiovasc Dis 2010 May;103(5):310-316. [doi: 10.1016/j.acvd.2010.04.007] [Medline: 20619241]
39. Bhattacharyya MR, Perkins-Porras L, Whitehead DL, Steptoe A. Psychological and clinical predictors of return to work after acute coronary syndrome. Eur Heart J 2007 Jan;28(2):160-165. [doi: 10.1093/eurheartj/ehl440] [Medline: 17185305]
40. Sørensen J, Ploug UJ. The cost of diabetes ‐ related complications: registry ‐ based analysis of days absent from work. Econ Res Int 2013 Mar 6;2013:1-8. [doi: 10.1155/2013/618039]
41. Sullivan PW, Slejko JF, Sculpher MJ, Ghushchyan V. Catalogue of EQ-5D scores for the United Kingdom. Med Decis Making 2011;31(6):800-804. [doi: 10.1177/0272989X11401031] [Medline: 21422468]
42. Nyman JA, Barleen NA, Dowd BE, Russell DW, Coons SJ, Sullivan PW. Quality-of-life weights for the US population: self-reported health status and priority health conditions, by demographic characteristics. Med Care 2007 Jul;45(7):618-628. [doi: 10.1097/MLR.0b013e31803dce05] [Medline: 17571010]
43. Sullivan PW, Lawrence WF, Ghushchyan V. A national catalog of preference-based scores for chronic conditions in the United States. Med Care 2005 Jul;43(7):736-749. [doi: 10.1097/01.mlr.0000172050.67085.4f] [Medline: 15970790]
44. Luengo-Fernandez R, Gray AM, Bull L, et al. Quality of life after TIA and stroke: ten-year results of the Oxford Vascular Study. Neurology 2013 Oct 29;81(18):1588-1595. [doi: 10.1212/WNL.0b013e3182a9f45f] [Medline: 24107865]
45. Xie J, Wu EQ, Zheng ZJ, et al. Impact of stroke on health-related quality of life in the noninstitutionalized population in the United States. Stroke 2006 Oct;37(10):2567-2572. [doi: 10.1161/01.STR.0000240506.34616.10] [Medline: 16946158]
46. Rebchuk AD, O'Neill ZR, Szefer EK, Hill MD, Field TS. Health utility weighting of the modified Rankin Scale: a systematic review and meta-analysis. JAMA Netw Open 2020 Apr 1;3(4):e203767. [doi: 10.1001/jamanetworkopen.2020.3767] [Medline: 32347948]
47. Lewis EF, Li Y, Pfeffer MA, et al. Impact of cardiovascular events on change in quality of life and utilities in patients after myocardial infarction: a VALIANT study (valsartan in acute myocardial infarction). JACC Heart Fail 2014 Apr;2(2):159-165. [doi: 10.1016/j.jchf.2013.12.003] [Medline: 24720924]
48. Munyombwe T, Hall M, Dondo TB, et al. Quality of life trajectories in survivors of acute myocardial infarction: a national longitudinal study. Heart 2020 Jan;106(1):33-39. [doi: 10.1136/heartjnl-2019-315510] [Medline: 31699696]
49. Matza LS, Stewart KD, Gandra SR, et al. Acute and chronic impact of cardiovascular events on health state utilities. BMC Health Serv Res 2015 Apr 22;15(1):173. [doi: 10.1186/s12913-015-0772-9] [Medline: 25896804]
50. Briggs A, Sculpher M, Claxton K. Decision Modelling for Health Economic Evaluation: Oxford University Press; 2006.
51. Husereau D, Drummond M, Augustovski F, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. Health Policy Open 2022 Dec;3:100063. [doi: 10.1016/j.hpopen.2021.100063] [Medline: 37383570]
52. Smith R, Schneider P. Making health economic models Shiny: a tutorial. Wellcome Open Res 2020;5:69. [doi: 10.12688/wellcomeopenres.15807.2] [Medline: 32904933]
53. Conroy RM, Pyörälä K, Fitzgerald AP, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J 2003 Jun;24(11):987-1003. [doi: 10.1016/s0195-668x(03)00114-3] [Medline: 12788299]
54. Gezonde Zorg, Gezonde Regio [Article in Dutch]. URL: https://gzgr.nl/ [accessed 2024-05-10]
55. Santema J, van Wijngaarden JDH, Hiddink EG, Deken F, Kleinsmann M, van Os HJA. Interorganizational mechanisms for developing and implementing clinical decision support systems in primary care: exploratory, qualitative case study. J Med Internet Res 2026 Mar 5;28:e83084. [doi: 10.2196/83084] [Medline: 41813251]
56. Tjin-A-Ton J, Vrolijk M. Herziene NHG-Standaard Cardiovasculair risicomanagement [Article in Dutch]. TPO - De Praktijk 2019 Jun;14(3):26-30. [doi: 10.1007/s12503-019-0041-1]
57. SCORE2 working group and ESC Cardiovascular risk collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. Eur Heart J 2021 Jul 1;42(25):2439-2454. [doi: 10.1093/eurheartj/ehab309] [Medline: 34120177]
58. SCORE2-OP working group and ESC Cardiovascular risk collaboration. SCORE2-OP risk prediction algorithms: estimating incident cardiovascular event risk in older persons in four geographical risk regions. Eur Heart J 2021 Jul 1;42(25):2455-2467. [doi: 10.1093/eurheartj/ehab312] [Medline: 34120185]

XSL•FO
RenderX

59. Hageman SHJ, McKay AJ, Ueda P, et al. Estimation of recurrent atherosclerotic cardiovascular event risk in patients with established cardiovascular disease: the updated SMART2 algorithm. Eur Heart J 2022 May 7;43(18):1715-1727. [doi: 10.1093/eurheartj/ehac056] [Medline: 35165703]
60. Betts MB, Rane P, Bergrath E, et al. Utility value estimates in cardiovascular disease and the effect of changing elicitation methods: a systematic literature review. Health Qual Life Outcomes 2020 Jul 27;18(1):251. [doi: 10.1186/s12955-020-01407-y] [Medline: 32718355]
61. Netherlands – total population. Human Mortality Database (HMD). 2021. URL: https://www.mortality.org/Country/Country?cntr=NLD [accessed 2026-09-02]
62. Netherlands Bureau of Statistics (CBS). Overledenen; doodsoorzaak (uitgebreide lijst), leeftijd, geslacht [Article in Dutch]. StatLine. URL: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/7233/table?ts=1726237214657 [accessed 2024-02-13]
63. Guideline for economic evaluations in healthcare (2024 version). : Zorginstituut Nederland; 2024 URL: https://english.zorginstituutnederland.nl/site/binaries/site-content/collections/documents/2024/01/16/guideline-for-economic-evaluations-in-healthcare/guideline-for-economic-evaluations-in-healthcare.pdf [accessed 2026-09-02]
64. van den Hout WB. The value of productivity: human-capital versus friction-cost method. Ann Rheum Dis 2010 Jan;69 Suppl 1:i89-i91. [doi: 10.1136/ard.2009.117150] [Medline: 19995754]
65. Alarid-Escudero F, Krijkamp E, Enns EA, et al. A tutorial on time-dependent cohort state-transition models in R using a cost-effectiveness analysis example. Med Decis Making 2023 Jan;43(1):21-41. [doi: 10.1177/0272989X221121747] [Medline: 36112849]
66. Kellerborg K, Perry-Duxbury M, de Vries L, van Baal P. Practical guidance for including future costs in economic evaluations in the Netherlands: introducing and applying PAID 3.0. Value Health 2020 Nov;23(11):1453-1461. [doi: 10.1016/j.jval.2020.07.004] [Medline: 33127016]
67. Büyükkaramikli NC, Rutten-van Mölken MPMH, Severens JL, Al M. TECH-VER: a verification checklist to reduce errors in models and improve their credibility. Pharmacoeconomics 2019 Nov;37(11):1391-1408. [doi: 10.1007/s40273-019-00844-y] [Medline: 31705406]
68. Benchmark Ketenzorg 2023 [Report in Dutch]. : InEen; 2023 URL: https://ineen.nl/wp-content/uploads/2024/07/InEen-Benchmark-Ketenzorg-2023.pdf [accessed 2024-09-16]
69. Hart- en vaatziekten | Leeftijd en geslacht | Totaal [Article in Dutch]. Volksgezondheid en Zorg (VZinfo). 2023. URL: https://www.vzinfo.nl/hart-en-vaatziekten/leeftijd-en-geslacht/totaal [accessed 2024-09-16]
70. Heiligers PJM, Noordman J, Korevaar JC, et al. Kennisvraag: praktijkondersteuners in de huisartspraktijk (POH's), klaar voor de toekomst? [Report in Dutch]. : Nivel; 2012 URL: https://www.nivel.nl/sites/default/files/bestanden/Rapport-kennisvraag-poh-def.pdf [accessed 2026-09-02]
71. De Guzman KR, Snoswell CL, Taylor ML, Gray LC, Caffery LJ. Economic evaluations of remote patient monitoring for chronic disease: a systematic review. Value Health 2022 Jun;25(6):897-913. [doi: 10.1016/j.jval.2021.12.001] [Medline: 35667780]
72. Zhang Y, Peña MT, Fletcher LM, Lal L, Swint JM, Reneker JC. Economic evaluation and costs of remote patient monitoring for cardiovascular disease in the United States: a systematic review. Int J Technol Assess Health Care 2023 Apr 28;39(1):e25. [doi: 10.1017/S0266462323000156] [Medline: 37114456]
73. Hayek MA, Giannouchos T, Lawley M, Kum HC. Economic evaluation of blood pressure monitoring techniques in patients with hypertension: a systematic review. JAMA Netw Open 2023 Nov 1;6(11):e2344372. [doi: 10.1001/jamanetworkopen.2023.44372] [Medline: 37988078]
74. Kraal JJ, Van den Akker-Van Marle ME, Abu-Hanna A, Stut W, Peek N, Kemps HM. Clinical and cost-effectiveness of home-based cardiac rehabilitation compared to conventional, centre-based cardiac rehabilitation: results of the FIT@Home study. Eur J Prev Cardiol 2017 Aug;24(12):1260-1273. [doi: 10.1177/2047487317710803] [Medline: 28534417]
75. Tucker KL, Sheppard JP, Stevens R, et al. Self-monitoring of blood pressure in hypertension: a systematic review and individual patient data meta-analysis. PLoS Med 2017 Sep;14(9):e1002389. [doi: 10.1371/journal.pmed.1002389] [Medline: 28926573]
76. Rahimi K, Bidel Z, Nazarzadeh M. Pharmacological blood pressure lowering for primary and secondary prevention of cardiovascular disease across different levels of blood pressure: an individual participant-level data meta-analysis. Lancet 2021 May 1;397(10285):1625-1636. [doi: 10.1016/S0140-6736(21)00590-0] [Medline: 33933205]
77. Rakers MM, van Os HJA, Recourt K, Mosis G, Chavannes NH, Struijs JN. Perceived barriers and facilitators of structural reimbursement for remote patient monitoring, an exploratory qualitative study. Health Policy Technol 2023 Mar;12(1):100718. [doi: 10.1016/j.hlpt.2022.100718]
78. Serrano LP, Maita KC, Avila FR, et al. Benefits and challenges of remote patient monitoring as perceived by health care practitioners: a systematic review. Perm J 2023 Dec 15;27(4):100-111. [doi: 10.7812/TPP/23.022] [Medline: 37735970]
79. Peters GM, Doggen CJM, van Harten WH. Budget impact analysis of providing hospital inpatient care at home virtually, starting with two specific surgical patient groups. BMJ Open 2022 Aug 1;12(8):e051833. [doi: 10.1136/bmjopen-2021-051833] [Medline: 35914920]
80. Mancia G, Bertinieri G, Grassi G, et al. Effects of blood-pressure measurement by the doctor on patient's blood pressure and heart rate. Lancet 1983 Sep 24;2(8352):695-698. [doi: 10.1016/s0140-6736(83)92244-4] [Medline: 6136837]

XSL•FO
RenderX

81. Myers MG, Godwin M, Dawes M, et al. Conventional versus automated measurement of blood pressure in primary care patients with systolic hypertension: randomised parallel design controlled trial. BMJ 2011 Feb 7;342:d286. [doi: 10.1136/bmj.d286] [Medline: 21300709]
82. Jmheij/CVRM-box-cua. GitHub. URL: https://github.com/jmheij/CVRM-box-cua [accessed 2026-09-07]
83. ELAN Research. URL: https://www.elanresearch.nl/ [accessed 2026-09-07]

Abbreviations

BIA: budget impact analysis
BP: blood pressure
CaU: care as usual
CHEERS: Consolidated Health Economic Evaluation Reporting Standards
CVD: cardiovascular disease
CVRM: cardiovascular remote management
DSA: deterministic sensitivity analysis
ELAN: Extramural LUMC Academic Network
GP: general practitioner
ICER: incremental cost-effectiveness ratio
LUMC: Leiden University Medical Center
MACE: major adverse cardiovascular event
MI: myocardial infarction
PN: practice nurse
QALY: quality-adjusted life year
RE-AIM: Reach, Effectiveness, Adoption, Implementation, and Maintenance
RPM: remote patient management
SBP: systolic blood pressure
TECH-VER: Technical Verification

Edited by L Buis; submitted 23.Jun.2025; peer-reviewed by B Tyl, J Lacroix; revised version received 31.Jul.2026; accepted 10.Aug.2026; published 28.Sep.2026.

Please cite as:
Heijdra Suasnabar J, Rakers M, van Hattem N, Hiddink E, Haas M, Zhang X, Atsma D, Bonten TN, van den Akker-van Marle ME, van Os HJA
A Cardiovascular Care Pathway Supported by Remote Patient Management in Dutch Primary Care: Cost-Effectiveness, Budget Impact, and Workload Analysis
JMIR Mhealth Uhealth 2026;14:e79512
URL: https://mhealth.jmir.org/2026/1/e79512
doi:10.2196/79512

© Jan Heijdra Suasnabar, Margot Rakers, Nicoline van Hattem, Eric Hiddink, Marcel Haas, Xuerui Zhang, Douwe Atsma, Tobias N Bonten, M Elske van den Akker-van Marle, Hendrikus J A van Os. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 28.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Efficacy and Safety of a Telemedicine System in Patients With Gestational Diabetes Mellitus (TELEGLAM): Single-Center, 2-Arm, Randomized, Open-Label, Parallel-Group Study

Kazuki Aoyama[1], MD; Yuya Nakajima[2], MD; Shu Meguro[1], MD, PhD; Yasunori Sato[3], PhD; Rei Goto[4], MD, PhD; Mariko Hida[5], MD, PhD; Takeshi Arimitsu[5], MD, PhD; Yoshifumi Kasuga[6], MD, PhD; Mamoru Tanaka[6], MD, PhD; Hiroshi Itoh[2], MD, PhD; Kaori Hayashi[1], MD, PhD

[1]Department of Endocrinology, Metabolism and Nephrology, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo, Japan

[2]Center for Preventive Medicine, Keio University School of Medicine, Tokyo, Japan

[3]Department of Biostatistics, Keio University School of Medicine, Tokyo, Japan

[4]Graduate School of Business Administration, Keio University, Kanagawa, Japan

[5]Department of Pediatrics, Keio University School of Medicine, Tokyo, Japan

[6]Department of Obstetrics and Gynecology, Keio University School of Medicine, Tokyo, Japan

Corresponding Author:
Kazuki Aoyama, MD
Department of Endocrinology, Metabolism and Nephrology, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo, Japan

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2026/1/e113548

Abstract

Background: In the management of gestational diabetes mellitus (GDM), the usual medical treatment requires frequent visits for glucose monitoring and insulin dose adjustment, and this imposes significant physical, psychological, and economic burdens on pregnant women. As mobile health platforms become increasingly integrated into diabetes care, telemedicine may help alleviate these burdens; however, evidence evaluating its effectiveness as a replacement for routine in-person GDM care remains limited.

Objective: This study aims to evaluate the impact of telemedicine on the quality of life and costs for patients with GDM requiring insulin therapy.

Methods: This single-center, 2-arm, randomized, open-label, parallel-group study included patients with GDM who started insulin injection therapy. Participants were randomized to either the telemedicine or standard face-to-face care groups for 10 (SD 2) weeks. The telemedicine intervention used a smartphone-linked platform that enabled the automatic transfer of glucose data from connected glucose meters and facilitated real-time video consultations. Primary end points included costs and patient satisfaction. Costs were assessed using claims data, transportation calculations, and wage-based productivity losses, while patient satisfaction was evaluated through changes in the Problem Areas in Diabetes Survey and Diabetes Therapy-Related Quality of Life questionnaire scores. Secondary outcomes included glycemic control and perinatal outcomes.

Results: In total, 38 participants were included, with 18 assigned to the telemedicine group and 20 to the standard care group. Total costs (32,712, 95% CI 15,412 - 50,013 vs 59,202, 95% CI 42,603 - 75,800 Japanese yen; $284, 95% CI 134 - 435 vs $515, 95% CI 370 - 659, purchasing power parity [PPP]–adjusted; *P*=.01), direct non–health care costs (922, 95% CI −240 to 2084 vs 2561, 95% CI 1447 - 3676 yen; $8, 95% CI −2 to 18 vs $22, 95% CI 13 to 32 PPP-adjusted; *P*=.02), and indirect costs (8981, 95% CI −7119 to 25,082 vs 32,832, 95% CI 17,384 - 48,279 yen; $78, 95% CI −62 to 218 vs $285, 95% CI 151 - 420 PPP-adjusted; *P*=.01) reduced significantly in the telemedicine group compared with the standard care group. The improvements in the Problem Areas in Diabetes Survey (−7.6, 95% CI −13.7 to −1.4; *P*=.02) and Diabetes Therapy–Related Quality of Life domain 1 (10.5, 95% CI 0.9-20.1; *P*=.03) scores from the baseline were significantly greater in the telemedicine group than that in the standard care group. Nonetheless, glycemic control and frequency of perinatal complications were comparable between the 2 groups. Consultation time was similar across groups, suggesting no added workload for clinicians.

Conclusions: In this randomized trial, mobile health–enabled telemedicine safely replaced routine in-person visits for patients with GDM requiring insulin therapy. Telemedicine significantly reduced psychological and economic burdens without compromising

XSL•FO
RenderX

glycemic or perinatal outcomes, demonstrating its value as a patient-centered and cost-efficient model of care. These findings support the broader implementation of mobile-based telemedicine approaches in GDM management.

Trial Registration: UMIN-CTR UMIN000047009; https://center6.umin.ac.jp/cgi-open-bin/ctr_e/ctr_view.cgi?recptno=R000053519

(JMIR Mhealth Uhealth 2026;14:e72242) doi:10.2196/72242

KEYWORDS

gestational diabetes mellitus; telemedicine; mobile health; patient satisfaction; costs

Introduction

Background

The prevalence of gestational diabetes mellitus (GDM) is rising globally due to increasing obesity rates and an aging population of women of childbearing age. GDM affects approximately 6% to 15% of all pregnancies [1] and increases the risk of adverse perinatal outcomes, such as cesarean section, large-for-gestational-age (LGA) infants, macrosomia, shoulder dystocia, and neonatal hypoglycemia [2]. Additionally, GDM raises the long-term complication risks of obesity and impaired glucose metabolism in mothers and infants [3]. Therefore, strict blood glucose management through diet, exercise, and pharmacotherapy is essential. However, conventional treatment requires frequent visits every 2 to 3 weeks for glucose monitoring, thereby imposing significant physical, psychological, and economic burdens on pregnant women. Systematic reviews have revealed that these patients experience a significant decrease in the quality of life (QOL) [4]. Furthermore, evidence from Italy indicates that health care expenditures are 29.2% higher in pregnancies complicated by GDM than in those without GDM [5].

Rationale and Knowledge Gap

Recently, telemedicine, defined as the delivery of health care and the exchange of health care information across distances [6], has become widely prevalent. As a core component of mobile health (mHealth), smartphone-based telemedicine platforms enable real-time data sharing and remote clinical support. The effectiveness and economic efficiency of telemedicine have been demonstrated for many diseases [7-9]. Thus, further research into the safety and effectiveness of this approach for patients with GDM, and its impact on psychological and socioeconomic burdens is required.

Previous studies have shown significant findings. A meta-analysis of 7 randomized controlled trials (RCTs) demonstrated that telemedicine significantly improved Hemoglobin A_{1c} (HbA_{1c}) levels in patients with GDM [10]. A subsequent meta-analysis of 32 RCTs reported a reduced risk of perinatal complications using telemedicine [11]. However, studies on patient satisfaction with telemedicine and its impact on costs among this population remain limited. Mackillop et al [12] conducted an RCT comprising 203 patients with GDM, investigating the effectiveness of a mobile phone–based real-time blood glucose management system and assessing patient satisfaction and direct health care costs as secondary outcomes. The satisfaction score, assessed using the Maternity Diabetes Treatment Satisfaction Questionnaire, was slightly higher in the intervention group. However, the direct health care costs did not differ significantly between the 2 groups, and insulin users, who were considered to have a particularly high level of psychological burden, were excluded. Furthermore, in a nonrandomized controlled trial involving 161 patients with GDM, direct health care costs were evaluated between a conventional therapy group and a group receiving additional support through the online analysis of capillary glucose data. The intervention group showed a 16% reduction in direct health care costs. However, costs other than direct health care costs were not evaluated in this study [13].

Objective of the Study

Thus, based on a review of previous studies, the safety of telemedicine for patients with GDM is becoming clearer. However, no studies in Japan or other countries have examined the effects of telemedicine on psychological burden or costs as primary outcomes for patients with GDM. Furthermore, no reports have revealed how telemedicine use in this population affects not only direct health care costs but also direct non–health care costs, such as transportation expenses or indirect costs, such as production loss due to absenteeism. Therefore, we conducted a prospective randomized controlled trial to investigate the impact of telemedicine on QOL and costs, particularly in patients with GDM who received insulin injections.

Methods

Study Design and Participants

TELEGLAM (efficacy and safety of a telemedicine system in patients with gestational diabetes mellitus) was a single-center, 2-arm, randomized, open-label, parallel-group study. We included patients attending the Department of Endocrinology, Metabolism and Nephrology, Keio University School of Medicine in Japan. The study was performed between February 7, 2022, and March 3, 2024. Its protocol has been previously described in detail [14]. Participants were eligible for inclusion if they met the following criteria: (1) diagnosed with GDM through an oral glucose tolerance test (OGTT) by 29 weeks and 6 days of gestation, underwent self-monitoring of blood glucose, and received insulin injections; (2) provided written informed consent; and (3) could operate the telemedicine system. In our hospital, the first trimester screening for GDM was performed for high-risk patients as reported previously [15]. If the first trimester screening for GDM was negative or positive but OGTT was negative, pregnant women had to undergo the 50-g glucose challenge test at approximately 24 gestational weeks. If the patients were glucose challenge test–positive, they underwent

XSL•FO
RenderX

OGTT at 24 to 28 gestational weeks [16]. Furthermore, GDM was diagnosed following the International Association of Diabetes and Pregnancy Study Groups Consensus Panel [17]: fasting plasma glucose ≥5.1 mmol/L (92 mg/dL) and/or 1-hour plasma glucose ≥10.0 mmol/L (180 mg/dL), and/or 2-hour plasma glucose ≥8.5 mmol/L (153 mg/dL). The target range of glycemic control was set at a preprandial blood glucose level of 5.6 mmol/L (100 mg/dL) and a 2-hour postprandial blood glucose level of 6.7 mmol/L (120 mg/dL). Insulin therapy was initiated if these targets were not achieved through diet and exercise. Exclusion criteria included (1) a GDM diagnosis after 30 weeks of gestation, (2) overt diabetes in pregnancy, (3) type 2 diabetes, (4) type 1 diabetes, (5) serious uncontrolled complications, (6) pacemakers or other implantable medical devices, (7) unwillingness to participate in the study, (8) being judged inappropriate to participate in the study by the physicians (eg, difficulties in understanding Japanese, refusal to treat GDM, difficulty in making regular visits to the hospital), (9) those without the necessary internet access for telemedicine, and (10) those who did not use smartphones.

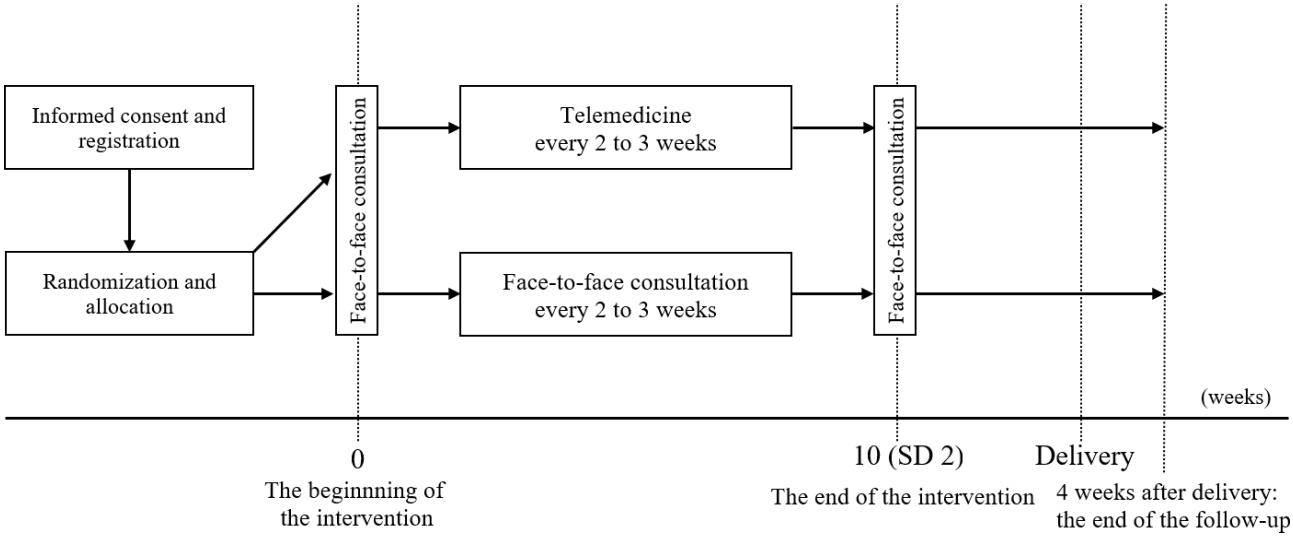
Ethical Considerations

The study protocol was approved by the Keio University School of Medicine Ethics Committee on February 7, 2022 (approval number: 20211125) and registered with the University Hospital Medical Information Network Clinical Trials Registry (UMIN000047009) in Japan. The protocols adhered to the principles of the Declaration of Helsinki. Written informed consent was obtained from all participants prior to enrollment. Participants were informed of the study purpose, procedures, potential risks, and their right to withdraw at any time without providing a reason. All collected data were deidentified before analysis. Clinical information and glucose data were stored on secure institutional servers accessible only to authorized study personnel. No personally identifiable information was used in the analytic dataset. Participants did not receive financial or material compensation for their participation in this study.

Intervention and Control

Telemedicine was provided to the patients in the intervention group using the MeDaCa telemedicine system developed by Medical Data Card Inc (Tokyo, Japan). This system connects a medical facility's computer with a patient's smartphone, tablet, or computer via the internet, enabling real-time video consultations similar to face-to-face visits. We did not include lifestyle interventions using smartphone apps, which enabled us to examine the effects of switching from face-to-face consultations to telemedicine. At Keio University, the One Touch Verio Reflect glucose meter (LifeScan Inc) was linked to MeDaCa. This allowed MeDaCa to automatically import blood glucose measurement data. The intervention period began at the first outpatient visit after the initiation of insulin injection and self-monitoring of blood glucose and ended 10 (SD 2) weeks later (Figure 1). In the intervention group, face-to-face consultations were conducted at the beginning and end of the intervention period, and telemedicine was conducted every 2 to 3 weeks in the interim period. In the control group, the patients were examined face-to-face every 2 to 3 weeks as usual. During this study, unscheduled visits were available at any time in both groups. Nurses provided telephone support as needed, and guidance on diet and exercise was similarly provided by doctors, nurses, and registered dietitians in both groups. The prenatal checkup was conducted as usual. Baseline and follow-up clinical information was obtained from outpatient visits and the hospital's electronic medical record system, and glucose data were transmitted through the telemedicine system. No automated or manual reminders were used during the intervention.

Figure 1. Trial design. In the intervention group, face-to-face consultations were conducted at the beginning and end of the intervention period, and telemedicine was conducted every 2 to 3 weeks in the interim period. In the control group, the patients were examined face-to-face every 2 to 3 weeks, as usual.



XSL•FO
RenderX

Outcomes

Primary Outcomes

In this study, we performed a cost-consequence analysis (CCA). CCA is defined as "a form of health economic evaluation study in which all direct and indirect costs and a catalog of different outcomes of all alternatives are listed separately" [18]. Therefore, multiple primary end points were established, focusing on health economic indicators and patient satisfaction. The health economic indicators included (1) direct health care costs, (2) direct non–health care costs, and (3) indirect costs. The actual costs incurred by medical examinations were evaluated as direct health care costs using the claims data, while direct non–health care costs, such as transportation costs, were calculated using Google Maps based on the patient's home or work address. Indirect costs, defined as production losses due to the inability to work because of the disease, were calculated from sick leave hours and the average wage in Japan if the patient was employed. Additionally, indirect costs associated with childcare were also considered, such as leaving a child with a babysitter for a hospital visit or a spouse taking a leave. The costs are presented in Japanese yen (¥), and US $1 was equivalent to 140.56 yen based on the exchange rate between January 2023 and January 2024. Furthermore, to facilitate international cost comparisons, we also converted all cost variables into purchasing power parity (PPP)–adjusted international dollars using a conversion factor of 115 yen per international dollar.

Moreover, we assessed patient satisfaction using the Problem Areas in Diabetes Survey (PAID) [19,20] and the Diabetes Therapy–Related QOL (DTR-QOL) questionnaire [21], both of which were used in their original, unmodified forms. Notably, PAID is a self-administered questionnaire comprising 20 items that measure diabetes-related psychological burdens. Each question was rated on a scale of 1 to 5, and the total score was evaluated. A lower score indicated a lesser psychological burden. The DTR-QOL is a questionnaire that assesses the influence of diabetes treatment on patient QOL, regardless of the treatment method. It comprises 29 questions divided into 4 domains, with each question answered on a scale of 1 to 7. Domain 1 assesses the burden on social and daily activities, domain 2 evaluates anxiety and dissatisfaction with treatment, domain 3 assesses hypoglycemia, and domain 4 evaluates satisfaction with the treatment. A higher score indicates a better QOL. The results were presented as the total scores for each domain, converted to a scale of 100 points. Participants completed the questionnaires at both the start and the end of the intervention, and the change from baseline was evaluated.

Secondary Outcomes

Secondary outcomes included changes in glycemic control indicators and metabolic parameters between the start and end of the intervention and perinatal outcomes. We evaluated fasting and postprandial blood glucose levels from self-monitoring records, total insulin dose, plasma glucose, HbA_{1c}, and glycated albumin as glycemic control indicators. Furthermore, maternal weight, BMI, systolic blood pressure (SBP), diastolic blood pressure (DBP), low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, triglyceride, C peptide, and the urinary albumin-to-creatinine ratio were evaluated as metabolic parameters. BMI was defined as weight in kilograms divided by the square of height in meters. Perinatal outcomes were evaluated based on the following: gestational age; mode of delivery; obstetric complications occurring during the perinatal period, such as premature rupture of the membrane, postpartum hemorrhage, and shoulder dystocia; maternal complications such as gestational hypertension; fetal indices including sex, height, weight, neonatal hypoglycemia, Apgar score (1 min/5 min), umbilical artery pH, macrosomia (>4000 g), low birth weight (<2500 g), LGA (>90th percentile), and small for gestational age (<10th percentile); incidence of congenital diseases; and admission to the neonatal intensive care unit. Additionally, neonatal hypoglycemia was defined as a blood glucose level of <40 mg/dL, and blood glucose was measured using point-of-care testing equipment. We used an intervention criterion of <50 mg/dL.

As an exploratory evaluation item, we measured the time spent in consultation based on the medical records for both groups.

Sample Size

The exploratory nature of the study made it difficult to predict parameter changes because no clinical study had used the PAID or the DTR-QOL for patients with GDM. Of the primary end points, the PAID was assumed to be the least likely to change with intervention. We estimated PAID scores to be similar between the groups before the intervention, with a 10-point difference emerging postintervention based on previous studies using the PAID [22,23]. To achieve a significance level of .05 in a 2-sided test with 80% power, assuming an SD of 10 points, the required sample size was estimated at 32 participants. Accounting for a 20% dropout rate, the target sample size was set at 40 participants.

Randomization

We performed randomization in a 1:1 ratio using a modified minimization method with a biased-coin assignment balancing maternal age (≤39 years old vs ≥40 years old), prepregnancy obesity, history of GDM, number of fetuses (singleton vs twins or higher order pregnancy), and ethnicity. The University Hospital Medical Information Network Internet Data and Information System for Clinical and Epidemiological Research (Cloud version) was used for randomization. A physician not involved in the study created the allocation registration form. Participants and physicians were not blinded to the treatment allocation.

Statistical Analysis

Primary and secondary outcomes were analyzed in the full analysis set, which included all patients who completed at least 2 online consultations during the study period and did not violate any of its protocols. Summary statistics comprised frequencies and proportions for categorical variables and means and SDs or medians with IQRs for continuous variables for the baseline characteristics. Categorical variables were compared using chi-square tests, while continuous variables were compared using the 2-tailed *t* test or Wilcoxon rank-sum test. For the primary end points, representative statistics were calculated for each measurement point, and an analysis of covariance was

XSL•FO
RenderX

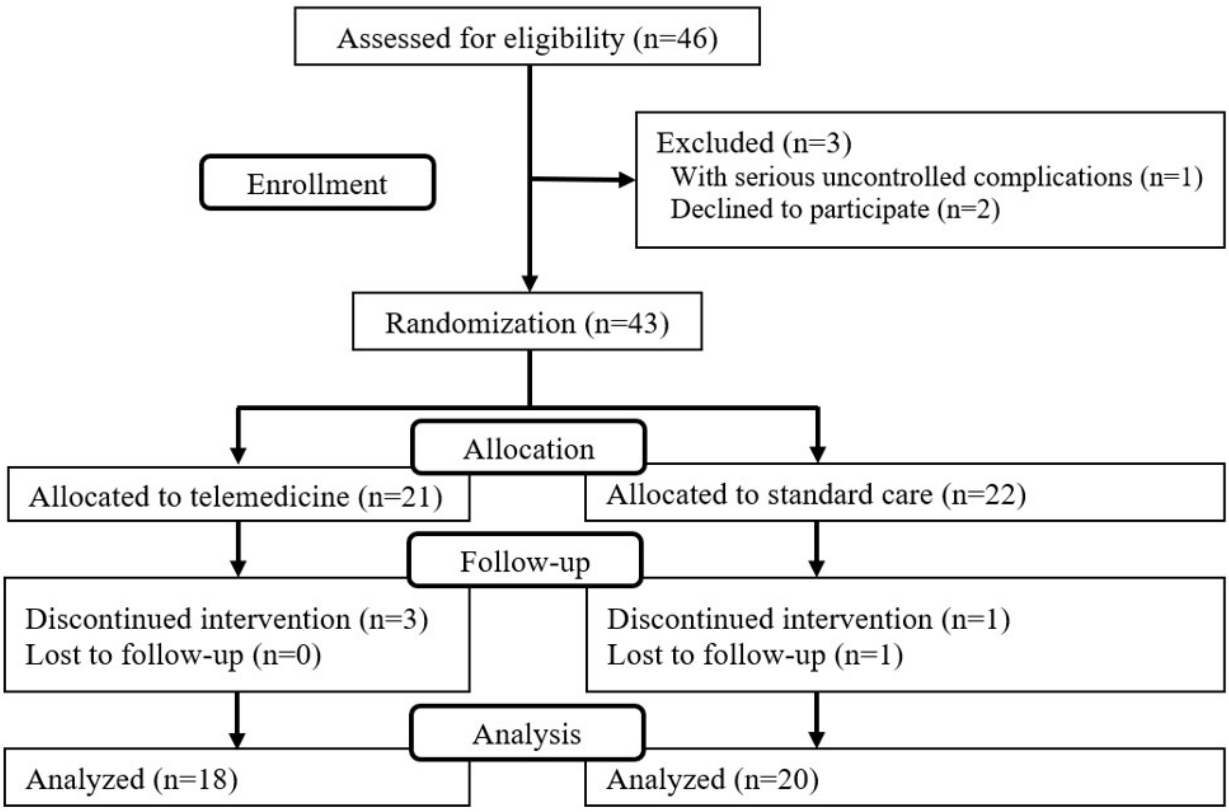
conducted. The analysis of covariance model included the following covariates: treatment group, maternal age, prepregnancy obesity, history of GDM, and number of fetuses. Secondary end points were analyzed similarly to the primary end points. All comparisons were preplanned, and all *P* values were 2-sided. A *P* value of <.05 was considered statistically significant. Statistical analyses were conducted using IBM SPSS Statistics software for Windows (version 28.0.0.0; IBM Corp).

Results

Patient Flow and Study Population

In total, 46 patients were initially selected for the study between February 7, 2022, and March 3, 2024. Of these, 1 patient was excluded due to serious uncontrolled complications, and 2 patients declined to participate, resulting in the enrollment of 43 patients. We randomly assigned these patients to the telemedicine group (intervention group) or the standard care group (control group). In the telemedicine group, 3 patients delivered before the study concluded, whereas in the standard care group, 1 patient delivered early, and 1 patient was lost to follow-up. Consequently, the full analysis set included 38 patients: 18 in the telemedicine group and 20 in the standard care group (Figure 2).

Figure 2. Flow diagram of patient enrollment.



Baseline Characteristics

Baseline characteristics were comparable between the groups, excluding C peptide, triglyceride, and the proportion of target glucose levels in the previous 7 days (Table 1). The mean age was 37.1 (SD 4.6) years and 35.8 (SD 4.1) years in the intervention and control groups, respectively, with a BMI of 24.7 (SD 4.3) and 24.0 (SD 4.1) kg/m², respectively. All patients, except 1 in the control group, were Japanese. The proportion of working patients was 72% (n=13) and 70% (n=14) in the intervention and control groups, respectively. Additionally, the mean daily insulin dose at the start of the intervention was 10.5 (SD 8.4) U/day and 9.8 (SD 9.1) U/day, respectively. At baseline, there were no significant differences in the PAID and DTR-QOL questionnaire scores between the 2 groups.

XSL•FO
RenderX

Table . Baseline characteristics.

Characteristics	Telemedicine group (n=18)	Standard care group (n=20)
Maternal age (y), mean (SD)	37.1 (4.6)	35.8 (4.1)
Gestational age at diagnosis (wk), median (IQR)	19 (14-26)	18 (14-26)
Height (cm), mean (SD)	159.6 (6.4)	157.9 (4.8)
Body weight (kg), mean (SD)	62.9 (12.5)	60.0 (11.3)
BMI (kg/m^2), mean (SD)	24.7 (4.3)	24.0 (4.1)
Systolic blood pressure (mm Hg), mean (SD)	109.1 (14.6)	110.1 (14.5)
Diastolic blood pressure (mm Hg), mean (SD)	66.3 (11.9)	65.1 (11.8)
Body weight before pregnancy (kg), mean (SD)	59.6 (12.7)	56.4 (11.3)
BMI before pregnancy (kg/m^2), mean (SD)	23.4 (4.7)	22.6 (4.1)
Obesity before pregnancy, n (%)	6 (33)	8 (40)
Gravidity, median (IQR)	2 (1-2)	1.5 (1-2)
Parity, median (IQR)	0 (0-1)	0 (0-1)
Previous GDM[a], n (%)	3 (17)	3 (15)
Family history of type 2 diabetes, n (%)	8 (44)	11 (55)
Nationality, n (%)		
Japanese	18 (100)	19 (95)
Chinese	0 (0)	1 (5)
Employed, n (%)	13 (72)	14 (70)
Number of fetuses, n (%)		
Singleton	18 (100)	20 (100)
Twins or higher	0 (0)	0 (0)
OGTT[b] results (mg/dL), mean (SD)		
Fasting plasma glucose	90.6 (7.4)	92.0 (8.4)
1-hour plasma glucose	187.7 (21.4)	178.7 (30.8)
2-hour plasma glucose	152.8 (32.5)	158.0 (17.2)
Plasma glucose (mg/dL), mean (SD)	104.6 (21.0)	93.1 (23.2)
HbA_{1c}[c] (%), mean (SD)	5.3 (0.3)	5.3 (0.4)
HbA_{1c} (mmol/mol), mean (SD)	34.5 (2.8)	33.9 (4.0)
GA[d] (%), mean (SD)	12.2 (1.2)	12.5 (1.1)
C peptide (ng/mL), mean (SD)	5.0 (3.2)	2.7 (1.8)
LDL-C[e] (mg/dL), mean (SD)	123.6 (37.8)	126.6 (28.3)
HDL-C[f] (mg/dL), mean (SD)	73.1 (16.8)	75.6 (11.6)
Triglyceride (mg/dL), mean (SD)	252.4 (92.2)	195.0 (79.1)
Urinary albumin-to-creatinine ratio (mg/gCr), mean (SD)	8.9 (7.6)	6.5 (3.3)
Total insulin dose (U/day), mean (SD)	10.5 (8.4)	9.8 (9.1)
Proportion of target glucose levels in the previous 7 days (%), mean (SD)	71.1 (17.4)	83.9 (11.0)
Questionnaire, mean (SD)		
PAID[g]	46.3 (14.6)	46.6 (12.6)
DTR-QOL1[h]	58.0 (20.0)	62.4 (19.8)

XSL•FO
RenderX

Characteristics	Telemedicine group (n=18)	Standard care group (n=20)
DTR-QOL2	63.4 (18.8)	60.5 (19.1)
DTR-QOL3	78.8 (24.5)	85.5 (19.0)
DTR-QOL4	47.8 (20.9)	50.7 (16.3)
DTR-QOL total	60.9 (13.6)	63.4 (13.3)

[a]GDM: gestational diabetes mellitus.

[b]OGTT: oral glucose tolerance test.

[c]HbA_{1c}: hemoglobin A_{1c}.

[d]GA: glycated albumin.

[e]LDL-C: low-density lipoprotein cholesterol.

[f]HDL-C: high-density lipoprotein cholesterol.

[g]PAID: Problem Areas in Diabetes Survey.

[h]DTR-QOL: Diabetes Therapy-Related QOL.

Primary Outcomes

The results of the CCA are shown in Tables 2 and 3. Regarding health economic indicators, the direct health care costs were comparable between the 2 groups. However, the telemedicine group had significantly lower direct non–health care costs (922, 95% CI −240 to 2083 vs 2561, 95% CI 1447 - 3676 yen; $8 95% CI −2 to 18 vs $22, 95% CI 13 - 32 PPP-adjusted; *P*=.02), indirect costs (8981, 95% CI −7119 to 25,082 vs 32,832, 95% CI 17,384 - 48,279 yen; $78, 95% CI −62 to 218 vs $285, 95% CI 151 - 420 PPP-adjusted; *P*=.01), and total costs (32,712, 95% CI 15,412 - 50,013 vs 59,202, 95% CI 42,603 - 75,800 yen; $284, 95% CI 134 - 435 vs $515, 95% CI 370 - 659 PPP-adjusted; *P*=.01) than the standard care group. Notably, policies were implemented during the COVID-19 pandemic to promote remote telemedicine to reduce crowding in medical institutions. However, as the number of infections decreased, these policies were discontinued, leading to changes in direct health care costs related to telemedicine during the study period. Therefore, we also analyzed the direct health care costs assuming telemedicine costs were equivalent to face-to-face consultations, and the results were consistent with those obtained using actual costs (Multimedia Appendix 1).

Table . Primary outcomes—cost consequence analysis of health economic indicators.

Variables	Telemedicine group (n=18)		Standard care group (n=20)		Mean difference		*P* value
	Yen (95% CI)	PPP[a]-adjusted dollars (95% CI)	Yen (95% CI)	PPP-adjusted dollars (95% CI)	Yen (95% CI)	PPP-adjusted dollars (95% CI)	
Costs							
Direct health care costs	22,809 (20,108 to 25,511)	198 (175 to 222)	23,809 (21,217 to 26,400)	207 (184 to 230)	−999 (−4103 to 2104)	−9 (−36 to 18)	.52
Direct non–health care costs	922 (−240 to 2084)	8 (−2 to 18)	2561 (1447 to 3676)	22 (13 to 32)	−1639 (−2974 to −304)	−14 (−26 to −3)	.02
Indirect costs	8981 (−7119 to 25,082)	78 (−62 to 218)	32,832 (17,384 to 48,279)	285 (151 to 420)	−23,851 (−42,349 to −5353)	−207 (−368 to −47)	.01
Total costs	32,712 (15,412 to 50,013)	284 (134 to 435)	59,202 (42,603 to 75,800)	515 (370 to 659)	−26,489 (−46,366 to −6613)	−230 (−403 to −58)	.01

[a]PPP: purchasing power parity.

XSL•FO
RenderX

Table . Primary outcomes—cost consequence analysis of patient satisfaction.

Variables	Telemedicine group (n=18)		Standard care group (n=20)		Difference in change (95% CI)	*P* value
	After intervention (95% CI)	Change from baseline (95% CI)	After intervention (95% CI)	Change from baseline (95% CI)		
Consequences						
PAID[a]	39.7 (31.9 to 47.4)	–3.8 (–9.1 to 1.6)	48.8 (41.3 to 56.2)	3.8 (–1.4 to 8.9)	–7.6 (–13.7 to 1.4)	.02
DTR-QOL[b] domain 1	68.0 (55.7 to 80.2)	7.8 (–0.5 to 16.2)	60.9 (49.2 to 72.7)	–2.7 (–10.7 to 5.3)	10.5 (0.9 to 20.1)	.03
DTR-QOL domain 2	69.6 (59.2 to 80.0)	5.1 (–3.8 to 14.0)	63.4 (53.5 to 73.4)	1.8 (–6.8 to 10.3)	3.3 (–6.9 to 13.5)	.51
DTR-QOL domain 3	81.1 (68.5 to 93.8)	–3.2 (–17.1 to 10.6)	80.7 (68.5 to 92.8)	–8.2 (–21.5 to 5.1)	5.0 (–10.9 to 20.9)	.53
DTR-QOL domain 4	42.1 (34.1 to 50.1)	–5.2 (–14.7 to 4.4)	43.2 (35.5 to 50.8)	–8.2 (–17.3 to 1.0)	3.0 (–7.9 to 14.0)	.58
DTR-QOL total	66.7 (58.4 to 74.9)	3.8 (–2.5 to 10.0)	61.9 (54.0 to 69.8)	–3.0 (–9.0 to 3.0)	6.7 (–0.5 to 13.9)	.07

[a]PAID, Problem Areas in Diabetes Survey.

[b]DTR-QOL, Diabetes Therapy-Related QOL.

Regarding patient satisfaction, improvements in PAID (−7.6, 95% CI −13.7 to −1.4; *P*=.02) and DTR-QOL domain 1 scores (10.5, 95% CI 0.9 - 20.1; *P*=.03) from the baseline were significantly greater in the telemedicine group than in the standard care group. Changes in DTR-QOL domain 2, 3, and 4 scores were comparable between the 2 groups.

Secondary Outcomes

Maternal metabolic parameters and perinatal outcomes are shown in Tables 4 and 5. We observed a trend toward greater maternal weight gain in the telemedicine group; however, the difference was not statistically significant (group difference in change: 1.1 kg, 95% CI −0.6 to 2.7 kg; *P*=.19). Furthermore, the SBP and DBP decreased in the telemedicine group and increased in the control group; however, these differences were not significant (change in SBP between the groups: −8.4 mm Hg, 95% CI −17.7 to 0.9 mm Hg; *P*=.08; DBP difference: −5.5 mm Hg, 95% CI −13.9 to 3.0 mm Hg; *P*=.20). Changes in glucose and lipid parameters were similar between the groups.

XSL•FO
RenderX

Table . Secondary maternal outcomes.

Variables	Telemedicine group (n=18)		Standard care group (n=20)		Group comparison	
	After intervention (95% CI)	Change from baseline (95% CI)	After intervention (95% CI)	Change from baseline (95% CI)	Difference in change (95% CI)	*P* value
Maternal outcomes						
Maternal body weight (kg)	67.6 (62.7 to 72.4)	3.8 (2.4 to 5.2)	61.5 (56.9 to 66.2)	2.7 (1.4 to 4.1)	1.1 (−0.6 to 2.7)	.19
BMI (kg/m^2)	26.2 (24.6 to 27.7)	1.5 (0.9 to 2.0)	24.6 (23.1 to 26.0)	1.0 (0.5 to 1.6)	0.4 (−0.2 to 1.0)	.22
Systolic blood pressure (mm Hg)	108.5 (103.0 to 113.9)	−2.8 (−10.9 to 5.3)	116.3 (111.1 to 121.5)	5.6 (−2.2 to 13.3)	−8.4 (−17.7 to 0.9)	.08
Diastolic blood pressure (mm Hg)	66.9 (61.4 to 72.4)	−0.5 (−7.8 to 6.9)	70.4 (65.1 to 75.7)	5.0 (−2.1 to 12.1)	−5.5 (−13.9 to 3.0)	.20
Plasma glucose (mg/dL)	101.1 (86.5 to 115.7)	−2.3 (−19.8 to 15.2)	93.4 (79.4 to 107.4)	1.6 (−15.1 to 18.4)	−4.0 (−24.0 to 16.1)	.69
HbA_{1c}[a] (%)	5.4 (5.3 to 5.6)	0.1 (−0.1 to 0.2)	5.4 (5.3 to 5.6)	0.2 (0.0 to 0.3)	−0.1 (−0.3 to 0.1)	.33
HbA_{1c} (mmol/mol)	35.7 (33.9 to 37.5)	0.8 (−0.9 to 2.5)	35.9 (34.1 to 37.6)	1.8 (0.2 to 3.4)	−0.9 (−2.9 to 1.0)	.33
GA[b] (%)	11.8 (11.2 to 12.4)	−0.3 (−0.7 to 0.1)	12.3 (11.7 to 12.8)	−0.3 (−0.7 to 0.1)	0.0 (−0.5 to 0.4)	.92
C peptide (ng/mL)	5.4 (2.8 to 8.0)	0.6 (−2.0 to 3.3)	2.6 (0.1 to 5.1)	0.1 (−2.4 to 2.7)	0.5 (−2.6 to 3.5)	.75
LDL-C[c] (mg/dL)	140.6 (119.0 to 162.2)	22.3 (8.5 to 36.1)	148.2 (127.5 to 168.9)	25.3 (12.0 to 38.5)	−3.0 (−18.8 to 12.9)	.71
HDL-C[d] (mg/dL)	76.3 (67.1 to 85.5)	3.2 (−2.1 to 8.6)	77.1 (68.4 to 85.9)	1.5 (−3.6 to 6.7)	1.7 (−4.5 to 7.9)	.58
Triglyceride (mg/dL)	298.6 (227.6 to 369.7)	56.0 (−1.3 to 113.3)	233.3 (165.1 to 301.4)	44.3 (−10.6 to 99.3)	11.6 (−54.2 to 77.4)	.72
Urinary albumin-to-creatinine ratio (mg/gCr)	10.9 (−1.2 to 23.0)	1.0 (−10.2 to 12.2)	14.4 (2.8 to 26.0)	7.5 (−3.2 to 18.2)	−6.5 (−19.4 to 6.3)	.31
Total insulin dose (U/day)	18.0 (11.2 to 24.8)	9.3 (4.2 to 14.4)	16.9 (10.4 to 23.4)	9.3 (4.4, to 14.2)	0.0 (−5.8 to 5.9)	.99
Proportion of target glucose levels in the previous 7 days (%)	78.8 (70.7 to 86.9)	5.9 (−2.9 to 14.7)	84.6 (76.2 to 92.9)	−2.5 (−11.6 to 6.6)	8.4 (−1.7 to 18.4)	.10
Outpatient consultation times (s)	543.7 (490.2 to 597.3)		518.2 (466.8 to 569.6)		—[e]	.41

[a]HbA_{1c}: hemoglobin A_{1c}.

[b]GA: glycated albumin.

[c]LDL-C: low-density lipoprotein cholesterol.

[d]HDL-C: high-density lipoprotein cholesterol.

[e]Not applicable.

XSL•FO
RenderX

Table . Secondary perinatal outcomes.

Variables	Telemedicine group (n=18)	Standard care group (n=20)	*P* value
Perinatal outcomes			
Gestational age at delivery (wk), median (IQR)	39 (38-39)	38 (38-39)	.43
Mode of delivery, n (%)			
Vaginal delivery	12 (67)	9 (45)	.18
Cesarean section	6 (33)	11 (55)	—[a]
Cesarean section, n (%)			
Planned	2 (33)	4 (36)	.90
Unplanned	4 (67)	7 (64)	—
Premature rupture of the membrane, n (%)	2 (11)	2 (10)	.91
Postpartum hemorrhage, n (%)	0 (0)	0 (0)	—
Shoulder dystocia, n (%)	0 (0)	0 (0)	—
Gestational hypertension, n (%)	1 (6)	0 (0)	.29
Infant sex, n (%)			
Male	10 (56)	8 (40)	.34
Female	8 (44)	12 (60)	—
Birth length (cm) (95% CI)	47.5 (45.8 to 49.2)	47.0 (45.4 to 48.7)	.67
Birth weight (g) (95% CI)	2969 (2679 to 3260)	2826 (2548 to 3105)	.39
Neonatal hypoglycemia, n (%)	7 (39)	8 (40)	.94
Apgar score 1 min (95% CI)	8.1 (7.5 to 8.7)	8.0 (7.5 to 8.6)	.79
Apgar score 5 min (95% CI)	8.9 (8.4 to 9.4)	8.8 (8.4 to 9.3)	.76
Umbilical artery pH (95% CI)	7.297 (7.265 to 7.329)	7.310 (7.280 to 7.341)	.46
Macrosomia, n (%)	0 (0)	0 (0)	—
Low birth weight infant, n (%)	3 (17)	3 (15)	.89
Large for gestational age, n (%)	2 (11)	3 (15)	.72
Small for gestational age, n (%)	1 (6)	3 (15)	.34
Congenital diseases, n (%)	1 (6)	0 (0)	.29
Admission to NICU[b], n (%)	1 (6)	2 (10)	.61
Outpatient consultation times (s) (95% CI)	543.7 (490.2 to 597.3)	518.2 (466.8 to 569.6)	.41

[a]Not applicable.

[b]NICU: neonatal intensive care unit.

Gestational age at delivery, mode of delivery, and the frequency of other perinatal complications were similar between the groups. There were no significant differences in neonatal height, weight, frequency of neonatal hypoglycemia, Apgar score, umbilical artery pH, or admission to the neonatal intensive care unit. Additionally, the time required for consultations per person was similar between the groups (543.7, 95% CI 490.2 - 597.3 s vs 518.2 95% CI 466.8 - 569.6 s, *P*=.41).

Discussion

Principal Findings

This study demonstrates that telemedicine use for GDM management improves patient satisfaction and reduces total costs, with comparable glycemic control and perinatal complications between the 2 groups. Thus, mHealth-supported telemedicine in the management of GDM can be a cost-saving technology providing better patient satisfaction. To the best of our knowledge, this is the first study to evaluate the impact of telemedicine on both psychological burden and total costs in patients with GDM.

XSL•FO
RenderX

Comparison With Previous Studies

Several previous studies have examined the effects of interventions through telemedicine or mobile apps for patients with GDM, in addition to regular consultations. These studies reported a reduced incidence of metabolic syndrome [24], decreased weight gain [25], higher rates of achieving target blood glucose levels [26], and a shorter time to reach target blood glucose levels [27]. Our findings share similarities with previous studies reporting a slight improvement in satisfaction scores with a mobile phone–based glucose management system [12] and a reduction in direct medical costs [13]. However, important differences remain because those earlier studies were performed alongside regular in-person consultations and did not evaluate the direct non–health care costs and indirect costs.

Clinical and Economic Implications

The impact of telemedicine on costs for patients with GDM has been previously reported [12,13]; however, these reports focused solely on direct health care costs. In our study, we evaluated the direct non–health care costs and indirect costs, accounting for a significant portion of the actual patient and societal burdens. The evaluation revealed that both costs were significantly lower in the telemedicine group, which may have contributed to the improvement in patient satisfaction. Although findings from a systematic review suggest that telemedicine holds a favorable level of economic efficiency [28], reports on the effectiveness and safety of telemedicine for managing GDM in Japan are lacking. Japan's health care system uses a universal coverage model, and medical services are priced using a unified fee schedule. The typical copayment rate is 30%, with the remainder covered by public health insurance, ensuring universal access to health care. Moreover, the official fees for telemedicine are lower than those for face-to-face consultations in Japan, which is presumed to be a factor hindering its broader adoption. Considering the positive impact on the economic and psychological burdens shown in this study, we believe that these findings could support raising the official fee for telemedicine to the same level as regular consultations for patients with GDM.

Regarding the safety of telemedicine as a replacement for regular clinical visits, an RCT comprising 100 patients with GDM reported that HbA_{1c} levels, the rate of cesarean sections, and the rate of LGA infants were comparable between the groups [29]. Furthermore, a similar report in 2023 indicated that there was no deterioration in glycemic control or increase in perinatal complications and that the telemedicine group had a significantly higher rate of achieving postprandial blood glucose targets [30]. Considering these previous research results, the safety of replacing regular clinical visits with telemedicine has been demonstrated. Most meta-analyses examining the effectiveness of telemedicine in patients with GDM combined studies that used telemedicine in addition to regular clinical visits with those that used telemedicine as a replacement. These analyses also consistently showed favorable results for telemedicine, including improvements in HbA_{1c} levels [10,31] and reductions in the risk of perinatal complications [11]. In our study, the 2 groups exhibited comparable glycemic control and perinatal outcomes, reaffirming the safety as reported in previous studies. The improvements in glycemic control and a reduction in perinatal complications, as observed in other studies, were not seen in this study possibly because of the sample size of this study. Additionally, it was speculated that the interventions of telemedicine or applications, which were conducted in other studies alongside routine care, may have influenced the results.

In this study, the consultation time was measured as an exploratory evaluation item. Considering the preparation required for the telemedicine system, we anticipated that the consultation time would be longer in the intervention group. However, the actual consultation time was comparable between the groups, suggesting that the burden on health care providers was similar in both groups, thereby highlighting the value of the reduced costs and psychological burden observed in the telemedicine group.

Policy Implications

The findings of this study have important implications for health policy and clinical practice. The observed reductions in psychological burden and costs suggest that structured telemedicine pathways may serve as an efficient alternative to conventional in-person GDM management. Incorporating such pathways into routine prenatal diabetes care has the potential to reduce patient time burdens and improve access to care. To achieve broader implementation, it will be essential to establish reimbursement models that appropriately cover telemedicine-based follow-up visits and digital monitoring.

Limitations

This study has some limitations. First, due to its design, blinding was challenging, which may have introduced clinical bias. Second, the study primarily involved Japanese patients residing in urban areas; hence, further research is needed in other settings and among different ethnic groups. Third, the small sample size reduced the power to detect infrequent complications evaluated as secondary outcomes. Fourth, we did not consider internet communication costs or the expenses associated with implementing the telemedicine system in medical institutions. While it is assumed that most participants conducted their consultations in an environment with Wi-Fi, additional costs may have been incurred. Finally, the questionnaire used was originally designed for patients with diabetes mellitus, not GDM. This limitation was communicated to participants on a separate sheet as a precaution.

Conclusions

In conclusion, TELEGLAM is the first study to examine the impact of telemedicine on psychological burdens and total costs in patients with GDM. In this study, we demonstrated that substituting regular consultations with telemedicine for patients with GDM reduced both psychological and economic burdens. Our results underscore the potential of mHealth-enabled telemedicine as a safe and effective substitute for conventional in-person GDM management, and we anticipate that these findings will support its broader implementation in clinical care. To translate these benefits into practice, health systems should integrate telemedicine into routine prenatal diabetes care and adopt policies that reduce financial and psychological burdens for patients.

XSL•FO
RenderX

Acknowledgments

We would like to thank Dr Hitoshi Ishii for his advice on the experimental design. Additionally, we appreciate Honyaku Center Inc for English language editing. Generative artificial intelligence tools were used for English editing and checking for typographical errors.

Funding

The authors declared no financial support was received for this work.

Data Availability

The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: KA, YN, SM, YS, RG, MH, TA, YK, MT, HI, KH
Formal analysis: KA, YS, RG
Investigation: KA, YN, SM
Methodology: YN, SM, YS, RG, MH, TA, YK, MT, HI, KH
Project administration: SM, HI, KH
Supervision: MT, HI, KH
Writing – original draft: KA
Writing – review & editing: YN, SM, YS, RG, MH, TA, YK, HI, KH

Conflicts of Interest

SM has received lecture fees from Novartis Pharma K.K., Novo Nordisk Pharma Ltd., Eli Lilly Japan K.K., AstraZeneca K.K., Kowa Co., Ltd., Astellas Pharma Inc., MSD K.K., Bayer Yakuhin, Ltd., Abbott Japan LLC, Sumitomo Pharma Co., Ltd., Mitsubishi Tanabe Pharma Corporation, Daiichi Sankyo Co., Ltd., Boehringer Ingelheim Japan, Inc., Sanofi K.K., Kyowa Kirin Co., Ltd., Otsuka Pharmaceutical Co., Ltd., Teijin Pharma Limited, Teijin Healthcare Limited, Taisho Pharmaceutical Co., Ltd., Roche Diabetes Care Japan Co., Ltd., Mochida Pharmaceutical Co., Ltd., Ono Pharmaceutical Co., Ltd., LifeScan Japan K.K., Terumo Corporation for the past 3 years. All other authors have nothing to disclose.

Multimedia Appendix 1
Health care costs in the case where telemedicine costs were equivalent to those of face-to-face consultations.
[DOCX File, 19 KB - mhealth_v14i1e72242_app1.docx]

Checklist 1
CONSORT-EHEALTH checklist.
[PDF File, 1263 KB - mhealth_v14i1e72242_app2.pdf]

References

1. McIntyre HD, Catalano P, Zhang C, Desoye G, Mathiesen ER, Damm P. Gestational diabetes mellitus. Nat Rev Dis Primers 2019 Jul 11;5(1):47. [doi: 10.1038/s41572-019-0098-8] [Medline: 31296866]
2. Farrar D, Simmonds M, Bryant M, et al. Hyperglycaemia and risk of adverse perinatal outcomes: systematic review and meta-analysis. BMJ 2016 Sep 13;354:i4694. [doi: 10.1136/bmj.i4694] [Medline: 27624087]
3. Lowe WL Jr, Scholtens DM, Lowe LP, et al. Association of gestational diabetes with maternal disorders of glucose metabolism and childhood adiposity. JAMA 2018 Sep 11;320(10):1005-1016. [doi: 10.1001/jama.2018.11628] [Medline: 30208453]
4. Marchetti D, Carrozzino D, Fraticelli F, Fulcheri M, Vitacolonna E. Quality of life in women with gestational diabetes mellitus: a systematic review. J Diabetes Res 2017;2017:7058082. [doi: 10.1155/2017/7058082] [Medline: 28326332]
5. Meregaglia M, Dainelli L, Banks H, Benedetto C, Detzel P, Fattore G. The short-term economic burden of gestational diabetes mellitus in Italy. BMC Pregnancy Childbirth 2018 Feb 23;18(1):58. [doi: 10.1186/s12884-018-1689-1] [Medline: 29471802]
6. Craig J, Patterson V. Introduction to the practice of telemedicine. J Telemed Telecare 2005;11(1):3-9. [doi: 10.1177/1357633X0501100102] [Medline: 15829036]
7. Timpel P, Oswald S, Schwarz PEH, Harst L. Mapping the evidence on the effectiveness of telemedicine interventions in diabetes, dyslipidemia, and hypertension: an umbrella review of systematic reviews and meta-analyses. J Med Internet Res 2020 Mar 18;22(3):e16791. [doi: 10.2196/16791] [Medline: 32186516]

XSL•FO
RenderX

8. Lin MH, Yuan WL, Huang TC, Zhang HF, Mai JT, Wang JF. Clinical effectiveness of telemedicine for chronic heart failure: a systematic review and meta-analysis. J Investig Med 2017 Jun;65(5):899-911. [doi: 10.1136/jim-2016-000199] [Medline: 28330835]
9. Eberle C, Stichling S. Clinical improvements by telemedicine interventions managing type 1 and type 2 diabetes: systematic meta-review. J Med Internet Res 2021 Feb 19;23(2):e23244. [doi: 10.2196/23244] [Medline: 33605889]
10. Ming WK, Mackillop LH, Farmer AJ, et al. Telemedicine technologies for diabetes in pregnancy: a systematic review and meta-analysis. J Med Internet Res 2016 Nov 9;18(11):e290. [doi: 10.2196/jmir.6556] [Medline: 27829574]
11. Xie W, Dai P, Qin Y, Wu M, Yang B, Yu X. Effectiveness of telemedicine for pregnant women with gestational diabetes mellitus: an updated meta-analysis of 32 randomized controlled trials with trial sequential analysis. BMC Pregnancy Childbirth 2020 Apr 6;20(1):198. [doi: 10.1186/s12884-020-02892-1] [Medline: 32252676]
12. Mackillop L, Hirst JE, Bartlett KJ, et al. Comparing the efficacy of a mobile phone-based blood glucose management system with standard clinic care in women with gestational diabetes: randomized controlled trial. JMIR Mhealth Uhealth 2018 Mar 20;6(3):e71. [doi: 10.2196/mhealth.9512] [Medline: 29559428]
13. Lemelin A, Paré G, Bernard S, Godbout A. Demonstrated cost-effectiveness of a telehomecare program for gestational diabetes mellitus management. Diabetes Technol Ther 2020 Mar;22(3):195-202. [doi: 10.1089/dia.2019.0259] [Medline: 31603351]
14. Aoyama K, Nakajima Y, Meguro S, et al. Efficacy and safety of a telemedicine system in subjects with gestational diabetes mellitus (TELEGLAM): study protocol for a randomized controlled trial. Heliyon 2023 Nov;9(11):e22504. [doi: 10.1016/j.heliyon.2023.e22504] [Medline: 38034604]
15. Tamagawa M, Kasuga Y, Saisho Y, et al. Predictors of later insulin therapy for gestational diabetes diagnosed in early pregnancy. Endocr J 2021 Nov 29;68(11):1321-1328. [doi: 10.1507/endocrj.EJ21-0118] [Medline: 34108310]
16. Arimitsu T, Kasuga Y, Ikenoue S, et al. Risk factors of neonatal hypoglycemia in neonates born to mothers with gestational diabetes. Endocr J 2023 May 29;70(5):511-517. [doi: 10.1507/endocrj.EJ22-0521] [Medline: 36792172]
17. International Association of Diabetes and Pregnancy Study Groups Consensus Panel, Metzger BE, Gabbe SG, et al. International Association of Diabetes and Pregnancy Study Groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care 2010 Mar;33(3):676-682. [doi: 10.2337/dc09-1848] [Medline: 20190296]
18. Kirch W. Encyclopedia of Public Health: Springer; 2008.
19. Welch GW, Jacobson AM, Polonsky WH. The Problem Areas in Diabetes Scale. An evaluation of its clinical utility. Diabetes Care 1997 May;20(5):760-766. [doi: 10.2337/diacare.20.5.760] [Medline: 9135939]
20. Hayashino Y, Goto M, Yamamoto T, Tsujii S, Ishii H. The Japanese version of Problem Areas in Diabetes Scale: a clinical and research tool to assess emotional functioning among people with diabetes. Diabetol Int 2024 Jan;15(1):117-122. [doi: 10.1007/s13340-023-00661-x] [Medline: 38264232]
21. Ishii H. Development and psychometric validation of the Diabetes Therapy-Related QOL (DTR-QOL) questionnaire. J Med Econ 2012;15(3):556-563. [doi: 10.3111/13696998.2012.665111] [Medline: 22313327]
22. Hayashino Y, Okamura S, Matsunaga S, Tsujii S, Ishii H, Tenri Cohort Study Group. The association between Problem Areas in Diabetes Scale scores and glycemic control is modified by types of diabetes therapy: diabetes distress and care registry in Tenri (DDCRT 2). Diabetes Res Clin Pract 2012 Sep;97(3):405-410. [doi: 10.1016/j.diabres.2012.04.005] [Medline: 22554998]
23. Schmitt A, Reimer A, Kulzer B, Haak T, Ehrmann D, Hermanns N. How to assess diabetes distress: comparison of the Problem Areas in Diabetes Scale (PAID) and the Diabetes Distress Scale (DDS). Diabet Med 2016 Jun;33(6):835-843. [doi: 10.1111/dme.12887] [Medline: 26287511]
24. Su MC, Chao AS, Chang MY, Chang YL, Chen CL, Sun JC. Effectiveness of a nurse-led web-based health management in preventing women with gestational diabetes from developing metabolic syndrome. J Nurs Res 2021 Dec 1;29(6):e176. [doi: 10.1097/jnr.0000000000000456] [Medline: 34570053]
25. Guo H, Zhang Y, Li P, Zhou P, Chen LM, Li SY. Evaluating the effects of mobile health intervention on weight management, glycemic control and pregnancy outcomes in patients with gestational diabetes mellitus. J Endocrinol Invest 2019 Jun;42(6):709-714. [doi: 10.1007/s40618-018-0975-0] [Medline: 30406378]
26. Tian Y, Zhang S, Huang F, Ma L. Comparing the efficacies of telemedicine and standard prenatal care on blood glucose control in women with gestational diabetes mellitus: randomized controlled trial. JMIR Mhealth Uhealth 2021 May 25;9(5):e22881. [doi: 10.2196/22881] [Medline: 33783365]
27. Rasekaba TM, Furler J, Young D, et al. Using technology to support care in gestational diabetes mellitus: quantitative outcomes of an exploratory randomised control trial of adjunct telemedicine for gestational diabetes mellitus (TeleGDM). Diabetes Res Clin Pract 2018 Aug;142:276-285. [doi: 10.1016/j.diabres.2018.05.049] [Medline: 29885390]
28. Akiyama M, Yoo BK. A systematic review of the economic evaluation of telemedicine in Japan. J Prev Med Public Health 2016 Jul;49(4):183-196. [doi: 10.3961/jpmph.16.043] [Medline: 27499161]
29. Pérez-Ferre N, Galindo M, Fernández MD, et al. The outcomes of gestational diabetes mellitus after a telecare approach are not inferior to traditional outpatient clinic visits. Int J Endocrinol 2010;2010:386941. [doi: 10.1155/2010/386941] [Medline: 20628517]

XSL•FO
RenderX

30. Munda A, Mlinaric Z, Jakin PA, Lunder M, Pongrac Barlovic D. Effectiveness of a comprehensive telemedicine intervention replacing standard care in gestational diabetes: a randomized controlled trial. Acta Diabetol 2023 Aug;60(8):1037-1044. [doi: 10.1007/s00592-023-02099-8] [Medline: 37185903]
31. Eberle C, Stichling S. Effects of telemetric interventions on maternal and fetal or neonatal outcomes in gestational diabetes: systematic meta-review. JMIR Diabetes 2021 Aug 27;6(3):e24284. [doi: 10.2196/24284] [Medline: 34448717]

Abbreviations

CCA: cost-consequence analysis
DBP: diastolic blood pressure
DTR-QOL: Diabetes Therapy-Related QOL
GDM: gestational diabetes mellitus
HbA_{1c}: hemoglobin A_{1c}
LGA: large-for-gestational-age
mHealth: mobile health
OGTT: oral glucose tolerance test
PAID: the Problem Areas in Diabetes Survey
PPP: purchasing power parity
QOL: quality of life
RCT: randomized controlled trial
SBP: systolic blood pressure

Edited by L Buis; submitted 06.Feb.2025; peer-reviewed by AN Ali, S Heydari; revised version received 08.Dec.2025; accepted 07.Apr.2026; published 08.May.2026.

Please cite as:
Aoyama K, Nakajima Y, Meguro S, Sato Y, Goto R, Hida M, Arimitsu T, Kasuga Y, Tanaka M, Itoh H, Hayashi K
Efficacy and Safety of a Telemedicine System in Patients With Gestational Diabetes Mellitus (TELEGLAM): Single-Center, 2-Arm, Randomized, Open-Label, Parallel-Group Study
JMIR Mhealth Uhealth 2026;14:e72242
URL: https://mhealth.jmir.org/2026/1/e72242
doi:10.2196/72242

© Kazuki Aoyama, Yuya Nakajima, Shu Meguro, Yasunori Sato, Rei Goto, Mariko Hida, Takeshi Arimitsu, Yoshifumi Kasuga, Mamoru Tanaka, Hiroshi Itoh, Kaori Hayashi. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 8.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Cost-Effectiveness of Internet of Things–Based Management of Home Noninvasive Positive Pressure Ventilation in Patients With Chronic Obstructive Pulmonary Disease and Hypercapnic Chronic Respiratory Failure: Trial-Based Economic Evaluation

Ruihua Feng[1*], PhD; Jiu Cheng[2*], MS; Yueying Cui[1], PhD; Xi Wang[1], MS; Yinghao Lv[3], MS; Weipeng Jiang[4,5], PhD; Chenjun Zhou[6], PhD; Yuanlin Song[4,7*], PhD; Hui Liu[1*], PhD

[1]Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Yabao Road 3, Beijing, China

[2]Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China

[3]Shenzhen Health Development Research and Data Management Center, Guangdong, China

[4]Department of Pulmonary Medicine and Critical Care Medicine, Zhongshan Hospital, Fudan University, Shanghai, China

[5]Department of Pulmonary Medicine, Zhongshan Hospital, Fudan University, Shanghai, China

[6]School of Biomedical Engineering, Tsinghua University, Beijing, China

[7]Shanghai Respiratory Research Institute, Fudan University, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Hui Liu, PhD
Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Yabao Road 3, Beijing, China

Abstract

Background: The management of chronic obstructive pulmonary disease (COPD) places a significant burden on health care systems worldwide. Home noninvasive positive pressure ventilation (NPPV) is an established treatment option associated with significant benefits for patients with COPD and hypercapnic chronic respiratory failure. Internet of Things (IoT)–based management may improve communication between patients and physicians and strengthen the integration and comprehensiveness of home NPPV telemonitoring. However, the economic value of such systems remains insufficiently understood.

Objective: This study aimed to assess the cost-effectiveness of IoT-based management versus standard management of home NPPV in patients with COPD and hypercapnic chronic respiratory failure.

Methods: A Markov decision analytic model was developed to simulate real-world COPD progression and predict health outcomes and costs associated with IoT-based and standard management of home NPPV. COPD progression consisted of 4 health states: stable period, nonserious exacerbation period, serious exacerbation period, and death. Efficacy and cost inputs were primarily sourced from a published multicenter, prospective, randomized controlled trial and supplemented by official Chinese databases where necessary. Quality-adjusted life years (QALYs) were used as effect indicators for this model, which were derived from COPD Assessment Test scores. The discounted lifetime cost per QALY gained was calculated from the Chinese health care payer perspective, and sensitivity analyses were conducted to test the robustness of model results across different assumptions.

Results: Compared with standard NPPV, IoT-based NPPV increased costs by ¥3607.26 and improved QALYs by 0.24 per person across the lifetime horizon, resulting in an incremental cost-effectiveness ratio of ¥15,030.25 per QALY, with a 93.6% probability of being cost-effective at the given willingness-to-pay threshold (currency conversion to US dollars was based on the average exchange rate in 2019 [US $1=¥6.9]). Base case results were also robust to multiple one-way sensitivity analyses, with the main drivers being hospitalization costs for the standard and IoT-based NPPV groups during the serious exacerbation period.

Conclusions: IoT-based NPPV was cost-effective compared with standard NPPV for patients with COPD and hypercapnic chronic respiratory failure.

(JMIR Mhealth Uhealth 2026;14:e71340) doi:10.2196/71340

KEYWORDS

chronic obstructive pulmonary disease; noninvasive positive pressure ventilation; hypercapnic chronic respiratory failure; Internet of Things; telemedicine; cost-effectiveness analysis; economic evaluation

XSL•FO
RenderX

Introduction

Chronic obstructive pulmonary disease (COPD) is a chronic airway disease characterized by airflow obstruction, with a high rate of disability and mortality, and it is also one of the diseases that need to be prevented and controlled by the Healthy China Initiative (2019 - 2030) [1]. COPD is the sixth leading driver of the global burden of disease and imposes a substantial economic burden [2]. From 1990 to 2023, COPD consistently ranked among the top three level 3 causes of age-standardized mortality attributable to noncommunicable diseases in China [3]. The annual per capita direct medical cost for patients with COPD in China accounted for 33.33% to 118.09% of the local average annual income [4], with hospitalization and home oxygen therapy being the main driving factors for direct costs [5]. It is estimated that COPD will cost the world economy Int $4.326 trillion from 2020 to 2050, with China and the United States facing the largest economic burden of COPD [6]. With population aging in China, the burden of COPD is still rising. The prevention and control of COPD remain challenging. Effective management of stable patients with COPD is essential to improve their quality of life and reduce medical costs. Therefore, seeking the most economical and effective treatment and management program for patients with COPD is imperative.

A recent meta-analysis confirmed that home noninvasive positive pressure ventilation (NPPV) in treating patients with hypercapnic COPD was significantly associated with a lower mortality risk, all-cause hospital admission, and intubation [7]. The 2023 report of the Global Initiative for Chronic Obstructive Lung Disease (GOLD) [8], the latest guidelines from the European Respiratory Society [9], and the American Thoracic Society [10] support the use of home NPPV for patients with stable hypercapnic COPD. Therefore, NPPV has become an established treatment for patients with stable hypercapnic COPD. Telemonitoring of home NPPV is an important tool for optimizing follow-up. Using the Internet of Things (IoT)–based medical platform may improve the integration and comprehensiveness of telemonitoring and management of home NPPV [11]. The IoT cloud platform collects information on adherence, ventilator parameters, and physiological indicators and provides daily reports and follow-up summaries to the health care team. Therefore, it could improve the exchange of information between patients and physicians who prescribed NPPV, and between health professionals and the external companies that provide ventilator services, deliver individualized and fully integrated care management, as well as respond quickly to patient needs. These factors are crucial for guaranteeing treatment adherence and effectiveness. Relevant studies also showed that IoT-based home NPPV management is emerging and has produced positive results [12,13].

Hospitalization costs are a major component of the total cost of COPD exacerbations [14]. The use of IoT may save costs by reducing avoidable hospitalizations [15-17]. These potential cost savings need to offset the increased costs associated with IoT services. Therefore, it is vital to investigate the influence of IoT on the overall use of health care resources and to evaluate its cost-effectiveness to determine whether it represents a good value for money. A small number of studies have confirmed that in patients with chronic respiratory failure caused by neuromuscular or thoracic diseases (eg, obstructive sleep apnea, amyotrophic lateral sclerosis, and barrel chest), telemedicine-based continuous NPPV is more cost-effective than traditional follow-up and care [18-22]. However, there are few economic assessments for patients with COPD, and their cost-effectiveness remains uncertain [23-27]. Therefore, more studies from different countries are needed to determine whether IoT-based telemonitoring is more cost-effective. Therefore, we built a Markov model based on a randomized controlled prospective study [28] to simulate and analyze the long-term cost-effectiveness of IoT-based management of home NPPV treatment in patients with COPD and hypercapnic chronic respiratory failure and provide a basis for decision-making regarding the rational selection of COPD treatment strategies.

Methods

Analytical Method

A Markov decision analytical model was built with Microsoft Excel 2016 software to estimate the long-term cost-effectiveness of IoT-based (intervention) versus standard NPPV (control) in patients with COPD and hypercapnic chronic respiratory failure.

Patient Population

This study was based on a multicenter, prospective, randomized controlled trial (RCT; ChiCTR1800019536), which evaluated the effectiveness and safety of IoT-based management of home NPPV for patients with COPD and hypercapnic chronic respiratory failure in China. The trial was conducted over 12 months and involved 148 patients with a mean baseline age of 70.18 years and GOLD grades III-IV [29]. Patients were randomly allocated in a 1:1 ratio to the 2 study groups based on a computer-generated random number sequence implemented by an independent data center [28]. Both groups initiated oxygen therapy and NPPV treatment in the hospital, with an identical ventilator parameter titration protocol during hospitalization, and were discharged after meeting the treatment targets.

In the standard NPPV (control group; n=75), patients were advised to use home oxygen therapy for at least 12 hours per day and home NPPV intermittently for at least 8 hours per day, with continuous NPPV use during sleep, to maintain an oxygen saturation (SpO_2) above the target of 88%.

In the IoT-based NPPV (intervention group; n=73), in addition to the home oxygen therapy for at least 12 hours per day and home NPPV for at least 8 hours per day, with continuous NPPV use during sleep, patients received additional IoT-based management for the full 12 months. The intervention group remotely collected ventilator data via a fourth-generation mobile communication technology (4G) system attached to the back of the ventilator and transmitted the data to the secured platform. The IoT cloud platform (Curative Medical Technology Inc) consists of a data collector, parse server, automatic processing server, data storage center, application interface server, and web server. The platform monitors clinical parameters such as use time, NPPV pressure, mask leaks, breaths per minute, and tidal volume information in real time, and automatically generates daily reports for medical staff. The platform delivers intervention

XSL•FO
RenderX

suggestions when adverse reactions, leaks, or poor treatment efficacy occur. The IoT cloud platform triggers automatic alarms when the last 1-week average use is <5 h/night, mask leak is >60 L/min on 3 consecutive nights, or breaths are >25/min on 2 consecutive nights. Health care professionals respond promptly with telephone outreach; if necessary, emergency home visits are arranged, or patients are referred for in-hospital evaluation. Physicians adjust treatment plans based on platform data and patient feedback, with postdischarge management coordinated by family doctors, nurses, and health care providers (Air Liquide Healthcare). Health care professionals conduct in-home follow-ups at 1, 4, and 8 months after treatment, along with monthly telephone follow-ups, to provide health education, monitor the patient's condition, and support long-term treatment adherence.

During the entire study duration, patients in both groups also received standard COPD treatment, including smoking cessation interventions and medication treatment, in accordance with international treatment guidelines.

Patients who were lost to follow-up, were noncompliant, or refused telemedicine, telephone contacts, or home visits were excluded from this analysis. A total of 125 patients (control group, n=67; intervention group, n=58) were included. Full trial details have been published elsewhere [28]. The baseline characteristics of the study population are shown in Table S1 in Multimedia Appendix 1. The groups appeared well balanced with no major differences in baseline characteristics.

Ethical Considerations

The trial was approved by the medical ethics committee of Zhongshan Hospital, Fudan University (Shanghai, China; B2017-176R) and by local research and development committees at participating centers. In this trial, all patients provided written informed consent, and their data were anonymized. After enrollment, patients were provided with 1 year of free device use and transportation subsidies for 4 follow-up visits. Patients or the public were not involved in our research's design, conduct, reporting, or dissemination plans.

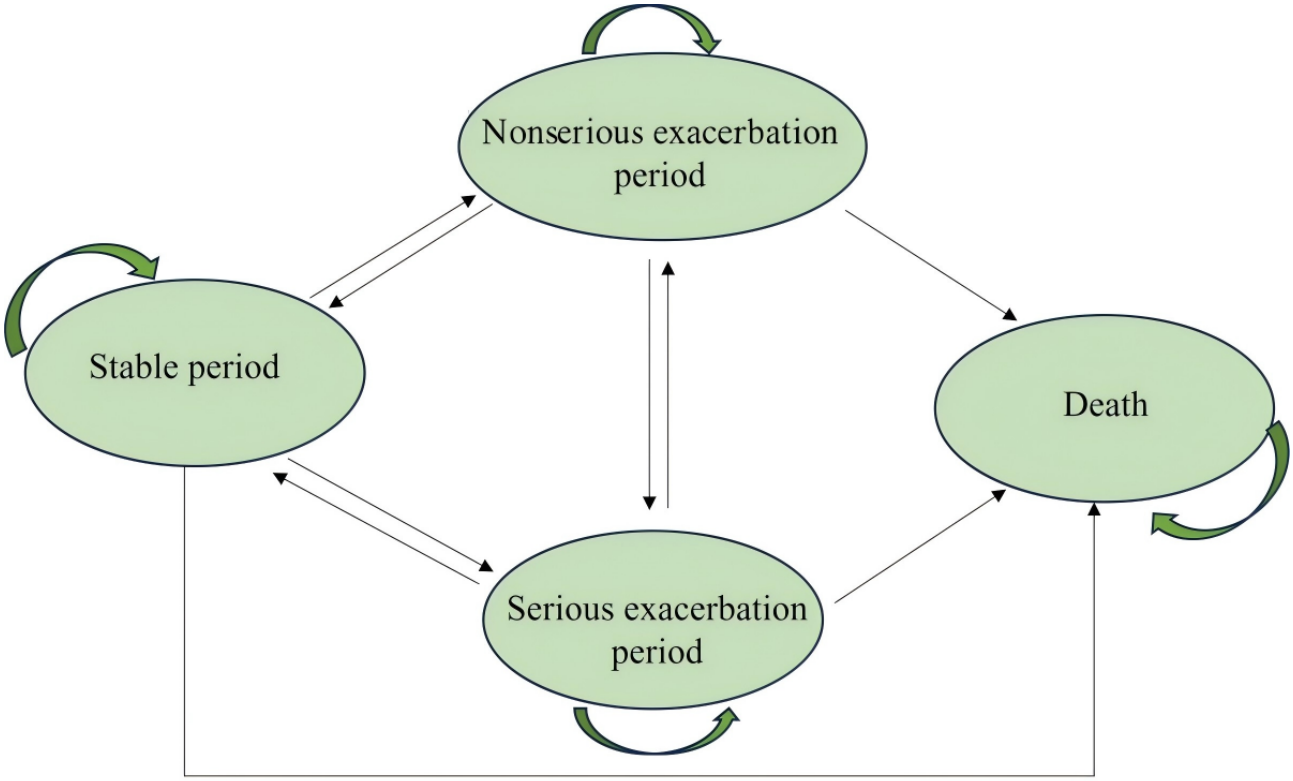
Markov Model Structure

The model structure was adapted from Zhou et al [30] based on real-world COPD progression, consisting of 4 health states: stable period, nonserious exacerbation period, serious exacerbation period, and death (Figure 1). The stable period health state represents a period of stable health characterized by the absence of exacerbations. Exacerbation-related health states, namely nonserious exacerbation period or serious exacerbation period, were defined by the severity of an exacerbation and whether additional medical resources were needed. Nonserious exacerbations were events that led to patients needing medications adjusted (increasing the type of medication or dosage or changing medication), whereas serious exacerbations were defined as events that resulted in hospitalization. Patients could transition to the death health state from any other health state. Patients remained in one health state in each cycle and then transferred to another health state in the next cycle. The model adopted a lifetime horizon corresponding to 8 years, calculated based on the difference between China's national average life expectancy in 2019 (77.3 years) and the average age of clinical trial patients included in this analysis (70.04 years). Each year represented a single model cycle, during which the 58 patients in the IoT-based management arm and the 67 patients in the standard management arm could progress throughout the model. The study adheres to Consolidated Health Economic Evaluation Reporting Standards (CHEERS) guidelines.

The RCT followed up with patients at 3, 6, and 12 months, so each patient had 3 periods of follow-up data, including COPD Assessment Test (CAT) scores, medication types, costs, and the number and costs of acute exacerbation hospitalizations. Due to the lack of patients' baseline medication types, we could not identify whether patients had a medication adjustment in the first 3 months. Therefore, we could not identify whether patients were in the stable or nonserious exacerbation period in the first 3 months; we could only identify whether they were in the serious exacerbation period. The calculation of parameters in this study did not include data from this follow-up for patients whose health state could not be determined.

XSL•FO
RenderX

Figure 1. Markov model structure.



Cost Parameters

Only direct costs collected as part of the trial, analyzed from the perspective of the Chinese health care payer, were included in this analysis. Cost components were grouped as either treatment costs, IoT costs, or hospitalization costs and were assumed to account for all aspects of care delivered during each health state (excluding death). Annual costs per patient differed for the IoT-based and standard management groups, as shown in Table 1. All costs are reported in Chinese Yuan (¥). Currency conversion to US dollars was based on the average exchange rate in 2019 (US $1=¥6.9).

XSL•FO
RenderX

Table . Cost parameters (¥ per patient per year).

Health state, groups, and items	Base value (SE)	Lower limit	Upper limit	Gamma distribution	
				α	β
Stable period[a]					
IoT[b]-based management					
Treatment	8913.18 (329.63)	8253.59	9572.78	731	12
Maintenance of telemedicine equipment	100.00 (—[c])	90.00	110.00	—	—
Telephone contact	293.00 (—)	263.70	322.30	—	—
Health care professional home visits	2379.00 (—)	2141.10	2616.90	—	—
Standard management					
Treatment	8828.47 (324.00)	8192.63	9477.35	741	12
Nonserious exacerbation period[d]					
IoT-based management					
Treatment	10,792.54 (657.38)	9438.64	12,146.44	269	40
Maintenance of telemedicine equipment	100.00 (—)	90.00	110.00	—	—
Telephone contact	293.00 (—)	263.70	322.30	—	—
Health care professional home visits	2379.00 (—)	2141.10	2616.90	—	—
Standard management					
Treatment	10,157.65 (532.59)	9064.87	11,250.44	364	28
Serious exacerbation period[e]					
IoT-based management					
Hospitalization	29,675.74 (5456.09)	18,981.81	40,369.68	30	1003
Standard management					
Hospitalization	29,918.2 (3680.16)	22,705.09	37,131.32	66	453

[a]Stable period represents a period of stable health, characterized by the absence of exacerbations.

[b]IoT: Internet of Things.

[c]Not applicable.

[d]Nonserious exacerbation period represents events that lead to patients needing medications adjusted (increasing in type of medication or dosage or changing in medication).

[e]Serious exacerbation period defined as events that result in hospitalizations.

Treatment costs included device (oxygen concentrator and NPPV device) setup, maintenance, and support costs, considered one-time costs at the beginning of the treatment period based on standardized costs for the previous 12 months. Rental of ventilators and oxygen supply were also included, estimated based on the purchase price using the annuity method [31] over an assumed 3-year lifespan of the devices. Additionally, medication and follow-up examination costs in the stable and nonserious exacerbation period health states were identified based on patients' self-reported case report forms and verified through electronic health records, where possible.

The next cost category, IoT costs, covered all costs associated with implementing IoT-based management and included telemedicine equipment maintenance, telephone contact, and health care professional home visits. The costs of telephone consultations and home visits from research and clinical teams were calculated based on hourly rates for different levels of health professionals.

Finally, hospitalization costs, covering all direct costs resulting from exacerbations, including medical and transportation costs, board and lodging, and nutrition and nursing, were collected again based on patient self-reported case report forms.

XSL•FO
RenderX

Utilities

Quality-adjusted life years (QALYs) were used as effect indicators for this model. Utility values were derived from CAT scores using the following mapping equation (equation 1) [32]. This mapping was developed based on data from patients with COPD in Taiwan, using multiple linear regression with the ordinary least squares method, and was internally validated within the derivation sample, demonstrating good predictive performance.

$$(1)\text{Utility}=1.26848-0.02159\times\text{CAT}-0.00188\times\text{age}$$

In equation 1, *Utility* represents EQ-5D-3L health utility value, where 1 represents full health and 0 represents death, *CAT* represents CAT scores, and *age* represents patient age.

Utility values were calculated for each health state, and values used in this analysis are shown in Table 2, where a utility value of 1 represents perfect health, and a value of zero represents death.

Table . Utility parameters.

Health states	Base value (SE)	Lower limit	Upper limit	α[a]	β[a]
Stable period	0.6195 (0.0121)	0.5956	0.6434	997	612
Nonserious exacerbation period	0.5960 (0.0181)	0.5598	0.6323	437	296
Serious exacerbation period	0.5319 (0.0127)	0.5065	0.5573	820	722
Death	0 (0)	—[b]	—	—	—

[a]The α and β values represent the shape parameters of the beta distribution.

[b]Not applicable.

Transition Probabilities

Transition probabilities were primarily derived from the clinical trial and supplemented by official Chinese databases. The annual natural mortality rate of the population was obtained from the *Chinese and Employment Statistics Yearbook 2021 "National Death Population States by Age and Sex in 2020 (November 1, 2019 to October 31, 2020)"* [33] and converted into the natural mortality rate of participants aged 40 to 80 years of whom 84% were male in the clinical trial. This was used to calculate the transition probabilities of patients from the stable period and nonserious exacerbation period health states to death. Other transition probabilities were calculated based on the patient population in the clinical trial. The conversion from incidence to probability was obtained by using the formula (equation 2) [34]. All model transition probabilities are shown in Table 3.

XSL•FO
RenderX

Table . Annual transition probabilities.

Health state transition	IoT[a]-based management			Standard management			References
	Base value (SE)	α[b]	β[b]	Base value (SE)	α	β	
Stable period							
Stable period	0.4760 (—[c])	—	—	0.4332 (—)	—	—	—
Nonserious exacerbation	0.3324 (0.0712)	14	29	0.1993 (0.0663)	7	35	[28]
Serious exacerbation	0.1829 (0.0680)	6	31	0.3588 (0.0672)	18	32	[28]
Death	0.0087[d] (—)	—	—	0.0087 (—)	—	—	[33]
Nonserious exacerbation							
Stable period	0.4134 (0.1212)	6	16	0.3588 (0.1041)	7	13	[28]
Nonserious exacerbation	0.3438 (—)	—	—	0.2191 (—)	—	—	—
Serious exacerbation	0.2341 (0.1292)	2	10	0.4134 (0.0989)	10	14	[28]
Death	0.0087 (—)	—	—	0.0087 (—)	—	—	[33]
Serious exacerbation							
Stable period	0.2835 (0.0844)	8	20	0.2835 (0.0844)	8	20	[28]
Nonserious exacerbation	0.2425 (0.0837)	6	19	0.2425 (0.0837)	6	19	[28]
Serious exacerbation	0.3261 (—)	—	—	0.3261 (—)	—	—	—
Death	0.1479 (0.0625)	5	27	0.1479 (0.0625)	5	27	[28]

[a]IoT: Internet of Things.

[b]The α and β values represent the shape parameters of the beta distribution.

[c]Not applicable.

[d]0.0087 represents the annual natural mortality rate.

$$(2) tpt = 1 - \exp(-r * t)$$

In Equation 2, tp_t represents the probability of event occurrence during t, r means the incidence rate during t, and t represents the study time.

Details of the cost-effectiveness analysis can be found in the Multimedia Appendix 1.

Outcomes

In line with the *Chinese Guidelines for Pharmacoeconomic Evaluation (2020)* [35], future costs and QALYs were discounted at 5% each year. Model outputs included total per-patient costs and QALYs and an incremental cost-effectiveness ratio (ICER), calculated to express the incremental difference in costs and QALYs between IoT-based management and standard management groups. The ICER was judged against a willingness-to-pay (WTP) threshold equal to per capita gross domestic product (GDP). The 2019 per capita GDP was ¥70,100 in China, according to the official data from the National Bureau of Statistics. IoT-based management was considered cost-effective when the ICER value was lower than ¥70,100 per QALY.

Sensitivity Analysis

One-Way Sensitivity Analysis

A one-way sensitivity analysis (OWSA) was used in the study to understand how changes in key parameters influenced results. The transition probabilities were not included in the OWSA because they were correlated with each other. All inputs were adjusted by the 95% CIs of the base value, with the exception of the discount rate, which was varied from 0% to 8%, and the cost of IoT, which was adjusted by −10% to +10% of the base value. Results of the OWSA are presented in a tornado chart.

Probabilistic Sensitivity Analysis

A probabilistic sensitivity analysis (PSA), using Monte Carlo simulations with 1000 iterations, was also performed to incorporate parameter uncertainty into the model analysis. In each iteration, parameter values were randomly drawn from specified distributions and results were presented on a scatter plot. A cost-effectiveness acceptability curve was also used to show the probability of IoT-based management being cost-effective against standard management given different WTP thresholds.

XSL•FO
RenderX

Results

Baseline Results

In the base-case analysis, total per-patient costs were ¥94,632.50 for the IoT-based management group and ¥91,025.24 for the standard management group; total QALYs were 3.49 and 3.25, respectively, across the model time horizon. Compared with the standard management group, IoT-based management increased costs by ¥3607.26 but also increased QALYs by 0.24 per patient across the time horizon, resulting in an ICER of ¥15,030.25 per QALY. As the ICER was substantially less than the WTP threshold of ¥70,100 per QALY, IoT-based management was cost-effective compared with standard management. These results are presented in Table 4.

Table . Base case results.

Group	Total costs (¥)	Total effectiveness (QALYs[a])	Incremental cost (¥)	Incremental effectiveness (QALYs)	ICER[b] (¥/QALY)
IoT[c]-based management	94,632.50	3.49	—[d]	—	—
Standard management	91,025.24	3.25	3607.26	0.24	15,030.25

[a]QALY: quality-adjusted life year.

[b]ICER: incremental cost-effectiveness ratio.

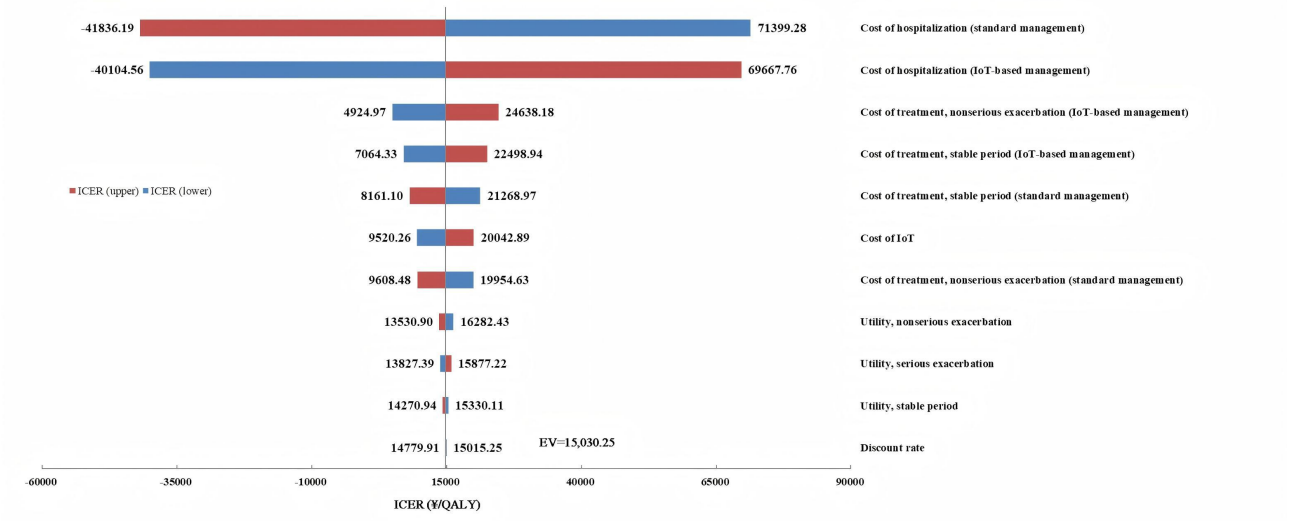
[c]IoT: Internet of Things.

[d]Not applicable.

One-Way Sensitivity Analysis

Results of the OWSA, displayed in Figure 2, showed that the parameter that drove the greatest change in ICER results was the hospitalization cost of a serious exacerbation for the standard management group. The second most influential parameter was the hospitalization cost of a serious exacerbation for the IoT-based management group. All other parameter changes had comparatively less influence on model results.

Figure 2. One-way sensitivity analysis tornado chart. EV: expected value; ICER: incremental cost-effectiveness ratio; IoT: Internet of Things; QALY: quality-adjusted life year.



Probabilistic Sensitivity Analysis

Results of the PSA, presented in Figure 3, showed the uncertainty around the model's cost and QALY estimates. A total of 93.6% of iterations fell below the WTP threshold (indicated by the dotted line), indicating that IoT-based management had a 93.6% probability of being cost-effective relative to standard management at the given WTP threshold. This relationship between the probability of being cost-effective as the WTP threshold changes is shown in Figure 4. Additionally, in 80.2% of PSA iterations, IoT-based NPPV was more effective (accumulating more QALYs) and cheaper than standard NPPV.

XSL•FO
RenderX

Figure 3. Monte Carlo simulation scatterplot. Dotted line indicates willingness to pay=¥70,100 per quality-adjusted life year (QALY).

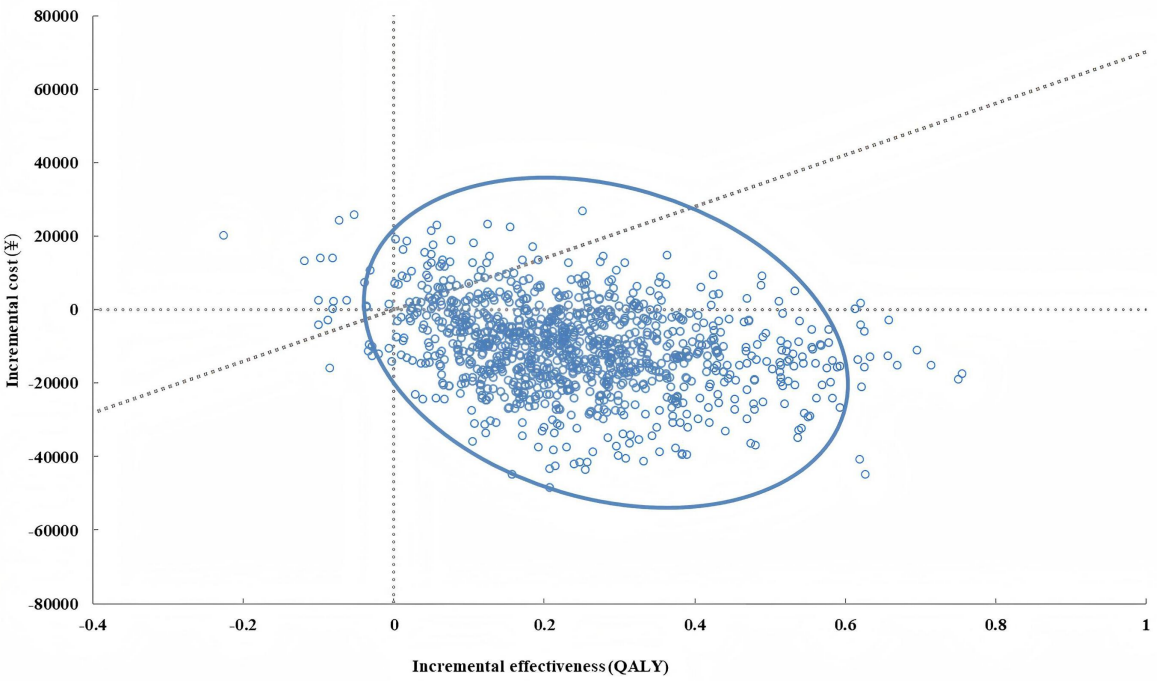
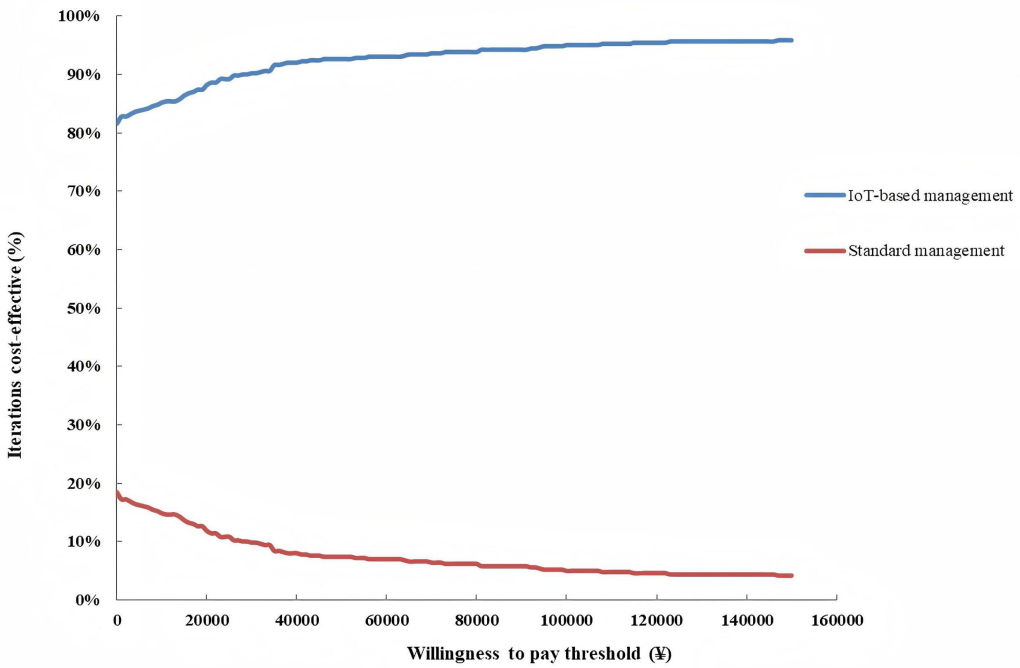


Figure 4. Cost-effectiveness acceptability curve. IoT: Internet of Things.



Discussion

This study demonstrated that, compared with standard NPPV, IoT-based NPPV was cost-effective for patients with COPD and hypercapnic chronic respiratory failure in China. The base case analysis achieved an ICER of ¥15,030.25 per QALY, which was substantially below the ¥70,100 per QALY WTP threshold. PSA results provided robust evidence further supporting the cost-effectiveness of IoT-based NPPV. They confirmed a high probability (93.6%) of IoT-based NPPV being cost-effective,

XSL•FO
RenderX

as well as a high probability (80.2%) of IoT-based NPPV being dominant (ie, cheaper and more effective).

Despite widespread understanding of the significant clinical and economic burdens imposed by COPD, this is, to the best of our knowledge, the first economic evaluation of IoT-based home NPPV treatment for patients with COPD in China. The level of existing economic evidence on the topic is limited, and results differ across countries. Meanwhile, a systematic review [36] highlighted the poor quality of this existing evidence. Five studies included in the systematic review were evaluated as low quality, and only one was evaluated as moderate quality. Further, none of the 6 studies included in the analysis were conducted in China. Consistent with our findings, an Italian study [23] found that a nurse-centered tele-assistance prevented hospitalizations and was a cost-effective intervention for patients with chronic respiratory failure receiving oxygen or home mechanical ventilation. A Spanish study [24] also found that a telehealth program was cost-effective for patients with severe COPD without comorbidities. However, telehealth was not cost-effective compared with usual care in the base case. A Dutch study showed that home initiation of chronic noninvasive ventilation in stable hypercapnic patients with COPD, with the use of telemedicine, is noninferior to in-hospital initiation, is safe, and reduces costs by more than 50% [19]. A study in the United Kingdom also concluded that telemonitoring in patients with COPD was not cost-effective. A possible rationale for the United Kingdom study not being cost-effective is that general practitioner home consultations were more frequent, with higher general practitioner costs in the telemonitoring arm, but differences in QALYs over the trial period were not significant between trial arms [25]. More specifically, existing economic evidence supported the use of telemonitoring in managing COPD [21,37].

The medical IoT is a key part of the digital transformation of health care, as it creates new health care delivery models and enables stakeholders to rebuild the workflows, contributing to increased productivity, enhanced patient experience, and improved effectiveness of home management for chronic diseases [38,39]. Specifically, IoT enables remote patient management and health education through architectures such as real-time monitoring via sensors, cloud-based data processing, and clinical feedback loops, thereby improving patient adherence to home-based therapy. At the same time, it relies on stringent data encryption and privacy protection mechanisms to ensure the long-term safety of patients.

A key outcome from the multicenter, prospective RCT, which formed the basis of this economic evaluation, was that compared with NPPV alone, the addition of an IoT treatment program reduced acute COPD exacerbation-related readmission frequency during the study period. The acute COPD exacerbation-related readmission frequency during the study period was 0.57 (95% CI 0.34 - 0.80) for the IoT-based NPPV group and 0.94 (95% CI 0.68 - 1.21) for the standard NPPV group (P=.05) [28]. The possible reason is that patients with IoT-based management are better able to manage their own condition, signs of exacerbations are identified earlier than without IoT management, and patients' adherence is improved [28]. The reduction in hospitalization costs for the IoT-based NPPV group partially offset the investment in IoT services. Reducing these admissions also reduced organizational burden and potentially lessened pressure on the Chinese health care system.

There exists a disproportionately high burden of global morbidity and mortality caused by chronic respiratory diseases in low- and middle-income countries (LMICs) [40], whereas LMICs typically face the challenge of scarce health care resources, such as inadequate coverage of health care facilities, a shortage of human resources for health, and poor quality of health services. The cost-effectiveness results of this study showed that IoT-based NPPV management could produce greater health and economic outcomes for a given budget. Such interventions are likely to have a greater impact in LMICs, contributing to increased efficiency in allocating limited health resources and increasing the accessibility of health services in LMICs, as well as reducing the pressure on the operation of the health system and the economic burden of COPD-related diseases. Therefore, this intervention should be considered a priority in LMICs. Nevertheless, the scalability of such interventions may depend on local digital infrastructure, workforce readiness, and regulatory frameworks, which vary substantially across LMICs [41]. Therefore, policymakers should consider phased implementation, investment in digital capacity building, and the development of data governance standards to ensure safe and sustainable expansion of IoT-enabled care. Particular attention should also be given to the accessibility of IoT-based management for vulnerable populations, such as those in low-income groups, to avoid exacerbating digital health inequalities.

This study contributes a new perspective to existing international research on the cost-effectiveness of IoT-based home NPPV for patients with COPD, indicating that IoT-based NPPV is cost-effective for patients with COPD and hypercapnic chronic respiratory failure in China. However, this study has several limitations. First, this economic evaluation was based primarily on the results of a single multicenter, prospective RCT, and this restricted clinical trial sample size may differ in some characteristics from the true population of the Chinese health care payer. This approach also means that trial limitations may also impact the inference of model results. Second, given that no studies on utility value measurement for acute exacerbation or stable patients with COPD in mainland China exist, the utility values for patients with COPD adopted in this analysis were derived from a mapping formula based on a study in Taiwan. Given the possible differences in patient characteristics, the formula may have underestimated the utility values of patients with severe disease. Finally, transitions to the nonserious exacerbation health state were determined based on the medication adjustments, as opposed to an objective diagnosis from a doctor, and this may introduce certain bias to the flow of patients throughout the model.

In conclusion, this study demonstrated that from a Chinese health care payer perspective, IoT-based NPPV was cost-effective compared with standard NPPV for patients with COPD and hypercapnic chronic respiratory failure. Further large-sample RCTs, additional real-world data, and more

XSL•FO
RenderX

widespread implementation of IoT-based NPPV throughout China should be sought to validate this.

Acknowledgments

The authors gratefully acknowledge the contributions of all participating hospital staff and patients. They also thank Scott Gibson from VitalAire for his support in the preparation of this manuscript as a medical writer.

Funding

Air Liquide Healthcare (Beijing) Co, Ltd funded the study. This work was supported by the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (2024-ZHCH630-01). Air Liquide Healthcare (Beijing) Co, Ltd provided ventilators, oxygen concentrators, related accessories, and services. The Internet of Things cloud platform was promoted by Curative Medical Technology Inc, Suzhou, China. Air Liquide Healthcare (Beijing) Co, Ltd was not involved in trial design; patient recruitment; data collection, analysis or interpretation; or the decision to submit for publication.

Data Availability

Data are available from the corresponding author upon reasonable request.

Authors' Contributions

HL and YS contributed to the design, discussion, and critical revision of the study. HL and YS contributed equally as corresponding authors. RF and JC constructed the Markov model and drafted the manuscript. RF and JC contributed equally as first authors. YC, XW, and YL revised and polished the manuscript. WJ and CZ revised and polished the manuscript and provided clinical expertise. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Baseline characteristics and supplementary methods for health-state classification, utility and cost estimation, transition probability calculation, and probabilistic sensitivity analyses.
[DOCX File, 24 KB - mhealth_v14i1e71340_app1.docx]

References

1. Healthy China initiative (2019-2030) [Article in Chinese]. National Health Commission of the People's Republic of China. 2019. URL: https://www.nhc.gov.cn/guihuaxxs/c100133/201907/2a6ed52f1c264203b5351bdbbadd2da8.shtml [accessed 2024-04-25]
2. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1204-1222. [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]
3. Zhang TH, Jiang HY, Xu XH, Zhao ZP, Zhou MG. Non-communicable disease burden in China, 1990–2023: evidence from the Global Burden of Disease study 2023. Chin Med J 2026;139(1):48-57. [doi: 10.1097/CM9.0000000000003898]
4. Zhu B, Wang Y, Ming J, Chen W, Zhang L. Disease burden of COPD in China: a systematic review. Int J Chron Obstruct Pulmon Dis 2018;13:1353-1364. [doi: 10.2147/COPD.S161555] [Medline: 29731623]
5. Foo J, Landis SH, Maskell J, et al. Continuing to Confront COPD International Patient Survey: economic impact of COPD in 12 countries. PLoS One 2016;11(4):e0152618. [doi: 10.1371/journal.pone.0152618] [Medline: 27092775]
6. Chen S, Kuhn M, Prettner K, et al. The global economic burden of chronic obstructive pulmonary disease for 204 countries and territories in 2020-50: a health-augmented macroeconomic modelling study. Lancet Glob Health 2023 Aug;11(8):e1183-e1193. [doi: 10.1016/S2214-109X(23)00217-6] [Medline: 37474226]
7. Wilson ME, Dobler CC, Morrow AS, et al. Association of home noninvasive positive pressure ventilation with clinical outcomes in chronic obstructive pulmonary disease: a systematic review and meta-analysis. JAMA 2020 Feb 4;323(5):455-465. [doi: 10.1001/jama.2019.22343] [Medline: 32016309]
8. Global strategy for prevention, diagnosis and management of COPD: 2023 report. Global Initiative for Chronic Obstructive Lung Disease. URL: https://goldcopd.org/2023-gold-report-2/ [accessed 2025-05-25]
9. Ergan B, Oczkowski S, Rochwerg B, et al. European Respiratory Society guidelines on long-term home non-invasive ventilation for management of COPD. Eur Respir J 2019 Sep;54(3):1901003. [doi: 10.1183/13993003.01003-2019] [Medline: 31467119]

XSL•FO
RenderX

10. Macrea M, Oczkowski S, Rochwerg B, et al. Long-term noninvasive ventilation in chronic stable hypercapnic chronic obstructive pulmonary disease. An official American Thoracic Society clinical practice guideline. Am J Respir Crit Care Med 2020 Aug 15;202(4):e74-e87. [doi: 10.1164/rccm.202006-2382ST] [Medline: 32795139]
11. Jagadeeswari V, Subramaniyaswamy V, Logesh R, Vijayakumar V. A study on medical internet of things and big data in personalized healthcare system. Health Inf Sci Syst 2018 Dec;6(1):14. [doi: 10.1007/s13755-018-0049-x] [Medline: 30279984]
12. Jiang W, Wang L, Song Y. Titration and follow-up for home noninvasive positive pressure ventilation in chronic obstructive pulmonary disease: the potential role of tele-monitoring and the Internet of things. Clin Respir J 2021 Jul;15(7):705-715. [doi: 10.1111/crj.13352] [Medline: 33705593]
13. Jang S, Kim Y, Cho WK. A systematic review and meta-analysis of telemonitoring interventions on severe COPD exacerbations. Int J Environ Res Public Health 2021 Jun 23;18(13):6757. [doi: 10.3390/ijerph18136757] [Medline: 34201762]
14. Toy EL, Gallagher KF, Stanley EL, Swensen AR, Duh MS. The economic impact of exacerbations of chronic obstructive pulmonary disease and exacerbation definition: a review. COPD 2010 Jun;7(3):214-228. [doi: 10.3109/15412555.2010.481697] [Medline: 20486821]
15. Polisena J, Tran K, Cimon K, et al. Home telehealth for chronic obstructive pulmonary disease: a systematic review and meta-analysis. J Telemed Telecare 2010;16(3):120-127. [doi: 10.1258/jtt.2009.090812] [Medline: 20197355]
16. Liddell A, Adshead S, Burgess E. Technology in the NHS: transforming the patient's experience of care. The King's Fund. 2008. URL: https://assets.kingsfund.org.uk/f/256914/x/0fffdec1e2/tehcnology_in_the_nhs_october_2008.pdf [accessed 2024-04-25]
17. Stroetmann KA, Kubitschke L, Robinson S, Stroetmann V, Cullen K, McDaid D. How can telehealth help in the provision of integrated care. World Health Organization. 2010. URL: https://iris.who.int/handle/10665/332044 [accessed 2024-04-25]
18. Turino C, de Batlle J, Woehrle H, et al. Management of continuous positive airway pressure treatment compliance using telemonitoring in obstructive sleep apnoea. Eur Respir J 2017 Feb;49(2):1601128. [doi: 10.1183/13993003.01128-2016] [Medline: 28179438]
19. Duiverman ML, Vonk JM, Bladder G, et al. Home initiation of chronic non-invasive ventilation in COPD patients with chronic hypercapnic respiratory failure: a randomised controlled trial. Thorax 2020 Mar;75(3):244-252. [doi: 10.1136/thoraxjnl-2019-213303] [Medline: 31484786]
20. Isetta V, Negrín MA, Monasterio C, et al. A Bayesian cost-effectiveness analysis of a telemedicine-based strategy for the management of sleep apnoea: a multicentre randomised controlled trial. Thorax 2015 Nov;70(11):1054-1061. [doi: 10.1136/thoraxjnl-2015-207032] [Medline: 26310452]
21. Hazenberg A, Kerstjens HA, Prins SC, Vermeulen KM, Wijkstra PJ. Initiation of home mechanical ventilation at home: a randomised controlled trial of efficacy, feasibility and costs. Respir Med 2014 Sep;108(9):1387-1395. [doi: 10.1016/j.rmed.2014.07.008] [Medline: 25081652]
22. Lopes de Almeida JP, Pinto A, Pinto S, Ohana B, de Carvalho M. Economic cost of home-telemonitoring care for BiPAP-assisted ALS individuals. Amyotroph Lateral Scler 2012 Oct;13(6):533-537. [doi: 10.3109/17482968.2012.703675] [Medline: 22873565]
23. Vitacca M, Bianchi L, Guerra A, et al. Tele-assistance in chronic respiratory failure patients: a randomised clinical trial. Eur Respir J 2009 Feb;33(2):411-418. [doi: 10.1183/09031936.00005608] [Medline: 18799512]
24. Jódar-Sánchez F, Ortega F, Parra C, et al. Cost-utility analysis of a telehealth programme for patients with severe chronic obstructive pulmonary disease treated with long-term oxygen therapy. J Telemed Telecare 2014 Sep;20(6):307-316. [doi: 10.1177/1357633X14544421] [Medline: 25052387]
25. Stoddart A, van der Pol M, Pinnock H, et al. Telemonitoring for chronic obstructive pulmonary disease: a cost and cost-utility analysis of a randomised controlled trial. J Telemed Telecare 2015 Mar;21(2):108-118. [doi: 10.1177/1357633X14566574] [Medline: 25586810]
26. Henderson C, Knapp M, Fernández JL, et al. Cost effectiveness of telehealth for patients with long term conditions (Whole Systems Demonstrator telehealth questionnaire study): nested economic evaluation in a pragmatic, cluster randomised controlled trial. BMJ 2013 Mar 20;346(mar20 4):f1035. [doi: 10.1136/bmj.f1035] [Medline: 23520339]
27. Ambrosino N, Vitacca M, Dreher M, et al. Tele-monitoring of ventilator-dependent patients: a European Respiratory Society statement. Eur Respir J 2016 Sep;48(3):648-663. [doi: 10.1183/13993003.01721-2015] [Medline: 27390283]
28. Jiang W, Jin X, Du C, et al. Internet of things-based management versus standard management of home noninvasive ventilation in COPD patients with hypercapnic chronic respiratory failure: a multicentre randomized controlled non-inferiority trial. eClinicalMedicine 2024 Apr;70:102518. [doi: 10.1016/j.eclinm.2024.102518] [Medline: 38495520]
29. Singh D, Agusti A, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease: the GOLD Science Committee report 2019. Eur Respir J 2019 May;53(5):1900164. [doi: 10.1183/13993003.00164-2019] [Medline: 30846476]
30. Zhou Y, Long E, Xu Q, Wang L, Jiang X, Hu M. Cost-effectiveness analysis of triple combination preparations in the treatment of moderate-to-severe chronic obstructive pulmonary disease. Front Public Health 2021 Jul 28;9:713258. [doi: 10.3389/fpubh.2021.713258]

XSL•FO
RenderX

31. Drummond ME, Sculpher MJ, Torrance GW, et al. Methods for the Economic Evaluation of Health Care Programmes, 3rd edition: Oxford University Press; 2005. [doi: 10.1093/oso/9780198529446.001.0001]
32. Yu CH, Chang SM, Hsu CH, et al. Mapping algorithms for predicting EuroQol-5D-3L utilities from the assessment test of chronic obstructive pulmonary disease. Sci Rep 2022 Dec 3;12(1):20930. [doi: 10.1038/s41598-022-24956-2] [Medline: 36463253]
33. China Population and Employment Statistics Yearbook 2021: China Statistics Press; 2021. URL: https://www.chinayearbooks.com/china-population-and-employment-statistics-yearbook-2021.html [accessed 2025-05-25]
34. Briggs A, Sculpher M. An introduction to Markov modelling for economic evaluation. Pharmacoeconomics 1998 Apr;13(4):397-409. [doi: 10.2165/00019053-199813040-00003] [Medline: 10178664]
35. Cao S, Liang W, Liang C, et al. Impact of China Guidelines for Pharmacoeconomic Evaluations on study quality: a systematic review of economic evaluations in China. Health Econ Rev 2025 Jul 4;15(1):53. [doi: 10.1186/s13561-025-00650-7] [Medline: 40613935]
36. Udsen FW, Hejlesen O, Ehlers LH. A systematic review of the cost and cost-effectiveness of telehealth for patients suffering from chronic obstructive pulmonary disease. J Telemed Telecare 2014 Jun;20(4):212-220. [doi: 10.1177/1357633X14533896] [Medline: 24803277]
37. Haesum LK, Soerensen N, Dinesen B, et al. Cost-utility analysis of a telerehabilitation program: a case study of COPD patients. Telemed J E Health 2012 Nov;18(9):688-692. [doi: 10.1089/tmj.2011.0250] [Medline: 23020647]
38. Dimitrov DV. Medical internet of things and big data in healthcare. Healthc Inform Res 2016 Jul;22(3):156-163. [doi: 10.4258/hir.2016.22.3.156] [Medline: 27525156]
39. Dorsey ER, Topol EJ. Telemedicine 2020 and the next decade. Lancet 2020 Mar 14;395(10227):859. [doi: 10.1016/S0140-6736(20)30424-4] [Medline: 32171399]
40. Meghji J, Mortimer K, Agusti A, et al. Improving lung health in low-income and middle-income countries: from challenges to solutions. Lancet 2021 Mar 6;397(10277):928-940. [doi: 10.1016/S0140-6736(21)00458-X] [Medline: 33631128]
41. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med 2013;10(2):e1001382. [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]

Abbreviations

CAT: COPD Assessment Test
CHEERS: Consolidated Health Economic Evaluation Reporting Standards
COPD: chronic obstructive pulmonary disease
GDP: gross domestic product
ICER: incremental cost-effectiveness ratio
IoT: Internet of Things
LMIC: low- and middle-income country
NPPV: noninvasive positive pressure ventilation
OWSA: one-way sensitivity analysis
PSA: probabilistic sensitivity analysis
QALY: quality-adjusted life year
RCT: randomized controlled trial
WTP: willingness-to-pay

Edited by L Buis; submitted 15.Jan.2025; peer-reviewed by K Meng, MY Wu, R Tripathi, VSK Kancharla; revised version received 15.Apr.2026; accepted 20.Apr.2026; published 14.May.2026.

Please cite as:
Feng R, Cheng J, Cui Y, Wang X, Lv Y, Jiang W, Zhou C, Song Y, Liu H
Cost-Effectiveness of Internet of Things–Based Management of Home Noninvasive Positive Pressure Ventilation in Patients With Chronic Obstructive Pulmonary Disease and Hypercapnic Chronic Respiratory Failure: Trial-Based Economic Evaluation
JMIR Mhealth Uhealth 2026;14:e71340
URL: https://mhealth.jmir.org/2026/1/e71340
doi:10.2196/71340

© Ruihua Feng, Jiu Cheng, Yueying Cui, Xi Wang, Yinghao Lv, Weipeng Jiang, Chenjun Zhou, Yuanlin Song, Hui Liu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is

XSL•FO
RenderX

properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Correction: Trustworthiness of Web-Based Pharmacy Apps in Pakistan Based on the Mobile App Rating Scale: Content Analysis and Quality Evaluation

JMIR Publications Editorial Office

JMIR Publications, 130 Queens Quay E, Suite 1100, Toronto, ON, Canada

Corresponding Author:
JMIR Publications, 130 Queens Quay E, Suite 1100, Toronto, ON, Canada

Related Article:

Correction of: https://mhealth.jmir.org/2025/1/e59884

(JMIR Mhealth Uhealth 2026;14:e88648) doi:10.2196/88648

Following the publication of "Trustworthiness of Web-Based Pharmacy Apps in Pakistan Based on the Mobile App Rating Scale: Content Analysis and Quality Evaluation" [1], concerns were raised regarding potential conflicts of interest between reviewers Huma Ali and Sadia Shakeel and the authors. These concerns included prior copublication with members of the author list and an active professional relationship between a reviewer and an author. These conflicts were not disclosed to the JMIR Publications Editorial Office as per journal policy.

This article has been subsequently reassessed by the editor in chief, who determined that the content of these reviews did not unduly influence the decision to accept the article. The decision to accept the submission was based primarily on the remaining reviewer's comments and the editor in chief's contemporary appraisal of the manuscript.

As such, the JMIR Publications Editorial Office issues this corrigendum to update the Conflicts of Interest statement:

> *Reviewers Huma Ali and Sadia Shakeel were found to have undisclosed conflicts of interest with members of the author list under the terms of JMIR Publications policy. The editor in chief determined that the content of the reviews did not unduly influence the decision to accept the article. The decision to accept the submission was based primarily on the remaining reviewer's comments and the editor in chief's contemporary appraisal of the manuscript.*

We regret that these issues were not identified before publication.

The correction will appear in the online version of the paper on the JMIR Publications website together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Reference

1. Sattar A, Rehman H, Naveed S, et al. Trustworthiness of web-based pharmacy apps in Pakistan based on the Mobile App Rating Scale: content analysis and quality evaluation. JMIR Mhealth Uhealth 2025 Apr 17;13:e59884. [doi: 10.2196/59884] [Medline: 40245428]

Submitted 28.Nov.2025; this is a non–peer-reviewed article;accepted 28.Nov.2025; published 03.Mar.2026.

Please cite as:
JMIR Publications Editorial Office
Correction: Trustworthiness of Web-Based Pharmacy Apps in Pakistan Based on the Mobile App Rating Scale: Content Analysis and Quality Evaluation
JMIR Mhealth Uhealth 2026;14:e88648
URL: https://mhealth.jmir.org/2026/1/e88648
doi:10.2196/88648

© JMIR Publications Editorial Office. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 3.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License

XSL•FO
RenderX

(https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Correction: Real-World Functional Gait Recovery Trajectories After Total Hip and Total Knee Arthroplasty via Smartphone Sensors: Observational Longitudinal Study

Levi S Teitz[1], PhD; Einat Granot-Hershkovitz[1], PhD; Pei-An Lee[2,3,4], PhD; On-Yee Lo[2,3,4], PhD; Brad Manor[2,3,4], PhD; Yuval Y Naveh[1], MSc

[1]OneStep (Celloscope Ltd), Nachmani 14, Tel Aviv, Israel

[2]Marcus Institute for Aging Research, Hebrew SeniorLife, Boston, MA, United States

[3]Beth Israel Deaconess Medical Center, Boston, MA, United States

[4]Harvard Medical School, Harvard University, Boston, MA, United States

Corresponding Author:
Yuval Y Naveh, MSc
OneStep (Celloscope Ltd), Nachmani 14, Tel Aviv, Israel

Related Article:

Correction of: https://mhealth.jmir.org/2026/1/e97657

(JMIR Mhealth Uhealth 2026;14:e116089) doi:10.2196/116089

In "Real-World Functional Gait Recovery Trajectories After Total Hip and Total Knee Arthroplasty via Smartphone Sensors: Observational Longitudinal Study" [1], the authors noted one correction.

The phone number for the corresponding author, YYN, has been removed.

The correction will appear in the online version of the paper on the JMIR Publications website together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Reference

1. Teitz LS, Granot-Hershkovitz E, Lee PA, Lo OY, Manor B, Naveh YY. Real-world functional gait recovery trajectories after total hip and total knee arthroplasty via smartphone sensors: observational longitudinal study. JMIR Mhealth Uhealth 2026 Oct 30;14:e97657. [doi: 10.2196/97657] [Medline: 42961449]

Submitted 03.Nov.2026; this is a non–peer-reviewed article;accepted 03.Nov.2026; published 16.Nov.2026.

Please cite as:
Teitz LS, Granot-Hershkovitz E, Lee PA, Lo OY, Manor B, Naveh YY
Correction: Real-World Functional Gait Recovery Trajectories After Total Hip and Total Knee Arthroplasty via Smartphone Sensors: Observational Longitudinal Study
JMIR Mhealth Uhealth 2026;14:e116089
URL: https://mhealth.jmir.org/2026/1/e116089
doi:10.2196/116089

© Levi S Teitz, Einat Granot-Hershkovitz, Pei-An Lee, On-Yee Lo, Brad Manor, Yuval Y Naveh. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 16.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Correction: Efficacy and Safety of a Telemedicine System in Patients With Gestational Diabetes Mellitus (TELEGLAM): Single-Center, 2-Arm, Randomized, Open-Label, Parallel-Group Study

Kazuki Aoyama[1], MD; Yuya Nakajima[2], MD; Shu Meguro[1], MD, PhD; Yasunori Sato[3], PhD; Rei Goto[4], MD, PhD; Mariko Hida[5], MD, PhD; Takeshi Arimitsu[5], MD, PhD; Yoshifumi Kasuga[6], MD, PhD; Mamoru Tanaka[6], MD, PhD; Hiroshi Itoh[2], MD, PhD; Kaori Hayashi[1], MD, PhD

[1]Department of Endocrinology, Metabolism and Nephrology, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo, Japan

[2]Center for Preventive Medicine, Keio University School of Medicine, Tokyo, Japan

[3]Department of Biostatistics, Keio University School of Medicine, Tokyo, Japan

[4]Graduate School of Business Administration, Keio University, Kanagawa, Japan

[5]Department of Pediatrics, Keio University School of Medicine, Tokyo, Japan

[6]Department of Obstetrics and Gynecology, Keio University School of Medicine, Tokyo, Japan

Corresponding Author:
Kazuki Aoyama, MD
Department of Endocrinology, Metabolism and Nephrology, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo, Japan

Related Article:

Correction of: https://mhealth.jmir.org/2026/1/e72242

(JMIR Mhealth Uhealth 2026;14:e113548) doi:10.2196/113548

In "Efficacy and Safety of a Telemedicine System in Patients With Gestational Diabetes Mellitus (TELEGLAM): Single-Center, 2-Arm, Randomized, Open-Label, Parallel-Group Study" [1], the authors identified errors in the equal contributor designations, the academic degree listed for author YN, and affiliation 4. Due to an error in the metadata submitted by the authors, several authors were incorrectly designated as having contributed equally.

In the originally published version, the author byline appeared as follows:

> *Kazuki Aoyama[1*], MD; Yuya Nakajima[2*], MD, PhD; Shu Meguro[1], MD, PhD; Yasunori Sato[3*], PhD; Rei Goto[4*], MD, PhD; Mariko Hida[5*], MD, PhD; Takeshi Arimitsu[5], MD, PhD; Yoshifumi Kasuga[6], MD, PhD; Mamoru Tanaka[6], MD, PhD; Hiroshi Itoh[2*], MD, PhD; Kaori Hayashi[1], MD, PhD*
>
> **these authors contributed equally*

The corrected author byline appears as follows:

> *Kazuki Aoyama[1], MD; Yuya Nakajima[2], MD; Shu Meguro[1], MD, PhD; Yasunori Sato[3], PhD; Rei Goto[4], MD, PhD; Mariko Hida[5], MD, PhD; Takeshi Arimitsu[5], MD, PhD; Yoshifumi Kasuga[6], MD, PhD; Mamoru Tanaka[6], MD, PhD; Hiroshi Itoh[2], MD, PhD; Kaori Hayashi[1], MD, PhD*

The equal contributor asterisks and the accompanying footnote have been removed. In addition, YN's academic degree has been corrected from "MD, PhD" to "MD." The author list and author order remain unchanged. These corrections do not affect the results or conclusions of the article.

In addition, affiliation 4 was originally published as follows:

> *Graduate School of Business Administration, Keio University,, Kanagawa, Japan*

The affiliation has been corrected to the following:

> *Graduate School of Business Administration, Keio University, Kanagawa, Japan*

The correction will appear in the online version of the paper on the JMIR Publications website together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Reference

XSL•FO
RenderX

1. Aoyama K, Nakajima Y, Meguro S, et al. Efficacy and safety of a telemedicine system in patients with gestational diabetes mellitus (TELEGLAM): single-center, 2-arm, randomized, open-label, parallel-group study. JMIR Mhealth Uhealth 2026 May 8;14:e72242. [doi: 10.2196/72242] [Medline: 42102285]

Submitted 03.Oct.2026; this is a non–peer-reviewed article;accepted 29.Oct.2026; published 17.Nov.2026.

Please cite as:
Aoyama K, Nakajima Y, Meguro S, Sato Y, Goto R, Hida M, Arimitsu T, Kasuga Y, Tanaka M, Itoh H, Hayashi K
Correction: Efficacy and Safety of a Telemedicine System in Patients With Gestational Diabetes Mellitus (TELEGLAM): Single-Center, 2-Arm, Randomized, Open-Label, Parallel-Group Study
JMIR Mhealth Uhealth 2026;14:e113548
URL: https://mhealth.jmir.org/2026/1/e113548
doi:10.2196/113548

© Kazuki Aoyama, Yuya Nakajima, Shu Meguro, Yasunori Sato, Rei Goto, Mariko Hida, Takeshi Arimitsu, Yoshifumi Kasuga, Mamoru Tanaka, Hiroshi Itoh, Kaori Hayashi. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 17.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Correction: Perceptions of Adults Aged 50 Years and Older Regarding the Use of Wearable mHealth Technologies to Promote Physical Activity: Systematic Review and Meta-Ethnography

Beatriz Rodríguez-Martín[1,2,3*], RN, MSci, PhD, Prof Dr; Diego González-Arroyo[4*], RN; Susana Priego-Jiménez[2,5], MSci, PhD; María Lopez-González[2], MSci; María Medrano-Echeverría[6,7], PhD; Marta Carolina Ruiz-Grao[2,8], RN, MSci, PhD; Celia Álvarez-Bueno[2,9], RN, MSci, PhD, Prof Dr

[1]Department of Nursing, Faculty of Health Sciences, Physiotherapy and Occupational Therapy, University of Castilla-La Mancha, Talavera de la Reina, Spain

[2]ABC-Age Research Group, University of Castilla-La Mancha, Cuenca, Spain

[3]Network for Research on Chronicity, Primary Care, and Health Promotion (RICAPPS), Barcelona, Spain

[4]University Hospital Nuestra Señora del Prado, Health Service of Castilla-La Mancha, Talavera de la Reina, Spain

[5]University Hospital of Cuenca, Camino de El Terminillo s/n, Cuenca, Spain

[6]Department of Health Sciences, Institute for Innovation & Sustainable Food Chain Development, Public University of Navarra, Pamplona, Spain

[7]CIBER de Fisiopatología de la Obesidad y Nutrición (CIBEROBN), Instituto de Salud Carlos III, Madrid, Spain

[8]Department of Nursing, Nursing Faculty, Physiotherapy and Occupational Therapy, University of Castilla-La Mancha. Albacete, Albacete, Spain

[9]Universidad Politécnica y Artística del Paraguay, Asuncion, Paraguay

[*]these authors contributed equally

Corresponding Author:
Susana Priego-Jiménez, MSci, PhD
ABC-Age Research Group, University of Castilla-La Mancha, Cuenca, Spain

Related Article:

Correction of: https://mhealth.jmir.org/2026/1/e67157

(JMIR Mhealth Uhealth 2026;14:e111082) doi:10.2196/111082

In “Perceptions of Adults Aged 50 Years and Older Regarding the Use of Wearable mHealth Technologies to Promote Physical Activity: Systematic Review and Meta-Ethnography” [1], the authors noted one error.

The name of the sixth author in the citation information has changed from the following:

Carolina Ruiz-Grao M

The name in the citation information now reads:

Ruiz-Grao MC

The correction will appear in the online version of the paper on the JMIR Publications website together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Reference

1. Rodríguez-Martín B, González-Arroyo D, Priego-Jiménez S, et al. Perceptions of adults aged 50 years and older regarding the use of wearable mHealth technologies to promote physical activity: systematic review and meta-ethnography. JMIR Mhealth Uhealth 2026 Jul 29;14:e67157. [doi: 10.2196/67157] [Medline: 42525576]

XSL•FO
RenderX

Submitted 02.Sep.2026; this is a non–peer-reviewed article;accepted 02.Sep.2026; published 08.Oct.2026.

Please cite as:
Rodríguez-Martín B, González-Arroyo D, Priego-Jiménez S, Lopez-González M, Medrano-Echeverría M, Ruiz-Grao MC, Álvarez-Bueno C
Correction: Perceptions of Adults Aged 50 Years and Older Regarding the Use of Wearable mHealth Technologies to Promote Physical Activity: Systematic Review and Meta-Ethnography
JMIR Mhealth Uhealth 2026;14:e111082
URL: https://mhealth.jmir.org/2026/1/e111082
doi:10.2196/111082

© Beatriz Rodríguez-Martín, Diego González-Arroyo, Susana Priego-Jiménez, María Lopez-González, María Medrano-Echeverría, Marta Carolina Ruiz-Grao, Celia Álvarez-Bueno. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 8.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Correction: Short-Term Effects of an mHealth Intervention on Healthy Behaviors and Cardiometabolic Health in Sedentary Employees: Quasi-Experimental Study

Yun-Ping Lin[1,2], PhD; Shu-Hua Lu[1,2], PhD; Kwo-Chen Lee[1,2], PhD; Wei-Fen Ma[1,2], PhD; Ya-Fang Ho[1,2], PhD; Wen-Chun Liao[1,2], PhD; Hui-Ting Yang[3], PhD; OiSaeng Hong[4], PhD

[1]School of Nursing, China Medical University, 100, Sec. 1, Jingmao Rd., Taichung, Taiwan

[2]Department of Nursing, China Medical University Hospital, Taichung, Taiwan

[3]School of Food Safety, Taipei Medical University, Taipei, Taiwan

[4]School of Nursing, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Yun-Ping Lin, PhD
School of Nursing, China Medical University, 100, Sec. 1, Jingmao Rd., Taichung, Taiwan

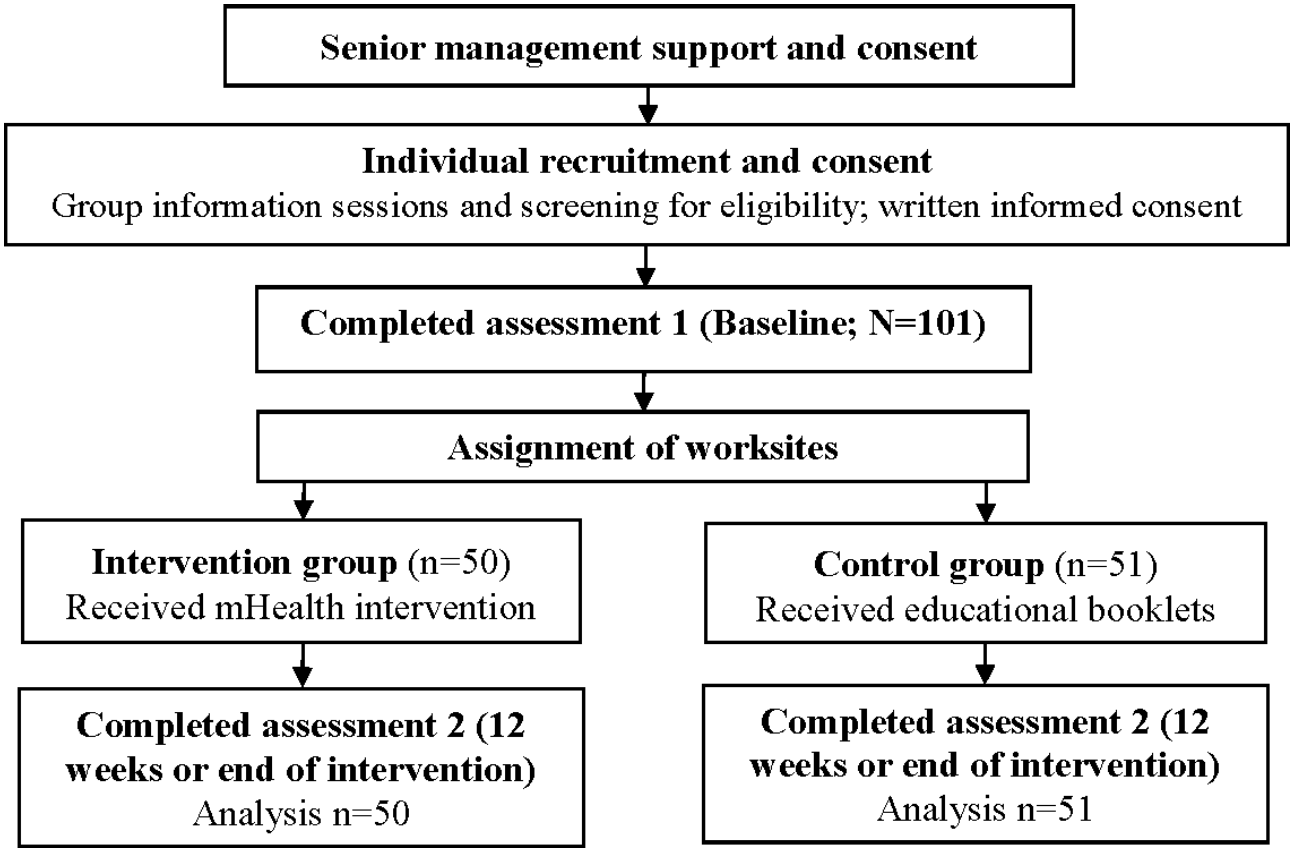
Related Article:

Correction of: https://mhealth.jmir.org/2026/1/e70074

(JMIR Mhealth Uhealth 2026;14:e106593) doi:10.2196/106593

In "Short-Term Effects of an mHealth Intervention on Healthy Behaviors and Cardiometabolic Health in Sedentary Employees: Quasi-Experimental Study" [1], the authors noted one error in Figure 1.

Figure 1. Flow diagram of participant recruitment and retention. mHealth: mobile health.



In the originally published Figure 1, the box "Assignment of worksites" appeared before "Completed assessment 1."

The figure has been corrected so that "Completed assessment 1 (Baseline; N=101)" appears before "Assignment of worksites." This corrected sequence is consistent with the Methods section,

XSL•FO
RenderX

which states that group assignments were concealed until completion of baseline assessments.

This revision does not affect the sample size, intervention or control conditions, outcome data, statistical analyses, results, interpretation, or conclusions of the article.

The correction will appear in the online version of the paper on the JMIR Publications website, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Reference

1. Lin YP, Lu SH, Lee KC, et al. Short-term effects of an mhealth intervention on healthy behaviors and cardiometabolic health in sedentary employees: quasi-experimental study. JMIR Mhealth Uhealth 2026 Apr 27;14:e70074. [doi: 10.2196/70074] [Medline: 42044372]

Submitted 09.Jul.2026; this is a non–peer-reviewed article;accepted 14.Jul.2026; published 14.Aug.2026.

Please cite as:
Lin YP, Lu SH, Lee KC, Ma WF, Ho YF, Liao WC, Yang HT, Hong O
Correction: Short-Term Effects of an mHealth Intervention on Healthy Behaviors and Cardiometabolic Health in Sedentary Employees: Quasi-Experimental Study
JMIR Mhealth Uhealth 2026;14:e106593
URL: https://mhealth.jmir.org/2026/1/e106593
doi:10.2196/106593

© Yun-Ping Lin, Shu-Hua Lu, Kwo-Chen Lee, Wei-Fen Ma, Ya-Fang Ho, Wen-Chun Liao, Hui-Ting Yang, OiSaeng Hong. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 14.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Addendum: Evaluating Characteristics and Quality of Mental Health Apps Available in App Stores for Indian Users: Systematic App Search and Review

Seema Mehrotra[1], PhD; Ravikesh Tripathi[1], PhD; Pramita Sengupta[1], PhD; Abhishek Karishiddimath[1], MSc; Angelina Francis[1], MA; Pratiksha Sharma[1], MSc; Paulomi Sudhir[1], PhD; T K Srikanth[2], PhD; Girish Rao[3], MD; Rajesh Sagar[4], MD

[1]Department of Clinical Psychology, National Institute of Mental Health and Neurosciences, Bengaluru, India

[2]E-Health Research Centre, International Institute of Information Technology Bangalore, Bengaluru, India

[3]Department of Epidemiology, Centre for Public Health, National Institute of Mental Health and Neurosciences, Bengaluru, India

[4]Department of Psychiatry, All India Institute of Medical Sciences, New Delhi, India

Corresponding Author:
Seema Mehrotra, PhD
Department of Clinical Psychology
National Institute of Mental Health and Neurosciences
Hosur Road
Bengaluru, 560029
India
Phone: 91 9448503853
Email: drmehrotra_seema@yahoo.com

Related Article:

Correction of: https://mhealth.jmir.org/2025/1/e79238/

(JMIR Mhealth Uhealth 2026;14:e91934) doi:10.2196/91934

In "Evaluating Characteristics and Quality of Mental Health Apps Available in App Stores for Indian Users: Systematic App Search and Review" [1], the authors made two additions.

First, toward the end of the Methods section, the authors added the following:

Phase 5: Supplementary Analyses

> *The interrater reliability of the MARS ratings was assessed using 10% of the apps randomly selected from the 350 previously reviewed. Each app in this subset was independently evaluated by all four reviewers, and the intraclass correlation coefficient (ICC) was calculated using Mangold International GmbH - ICC Calculator. In addition, this phase included a further detailed evaluation of paid apps. While the primary analysis focused on free apps or the free content of paid apps, we conducted an additional review of paid content within a subsample of subscription-based apps. Of the 350 apps initially reviewed, 226 involved in-app purchases, and 5 were fully paid apps. At the time of this supplementary review, 20 out of the 226 apps involving in-app purchases and 3 out of the 5 fully paid apps were no longer available in the app stores. Thus, a pool of 206 that offered in-app purchases and 2 that were fully paid was available for review of their paid content. We randomly selected 25% of these 206 in-app purchase apps (n=53) from each of the app categories such as intervention, assessment, multifunctional, etc, ensuring adequate representation across categories. The paid content of these 53 apps, along with the two fully paid apps, yielded a total of 55 apps reviewed between October 25 and November 20, 2025, by the same team members who conducted the primary review. Of the initially selected 55 apps, 10 could not be reviewed due to unavailability on app stores or persistent technical issues; these were replaced with alternative apps drawn from the corresponding categories. The final sample comprised 41 multifunctional apps, 6 intervention-based apps, 5 monitoring and tracking apps, and one app each in the assessment, journaling, and informative categories. The review process strictly followed the same protocol used in the evaluation of free versions.*

Second, toward the end of the Results section, the following was added:

Results of Supplementary Analyses

> *ICC analysis demonstrated good to excellent interrater reliability among the four reviewers at the*

XSL•FO
RenderX

95% confidence level. Single-measure ICCs ranged from 0.77 (absolute agreement) to 0.80 (consistency), while average-measure ICCs were excellent, ranging from 0.93 to 0.94. These results indicate strong agreement and consistency across raters, particularly when aggregated ratings were used.

A comparison of MARS scores between paid and free versions of the reviewed apps using the Wilcoxon signed rank test indicated no significant difference (z=–1.34; P>.05). The average MARS scores for paid versions ranged from 3.4 to 4.7, while the corresponding free versions ranged from 3.3 to 4.8. These observations suggest that inclusion of paid content in the reviewed apps did not substantially influence the overall MARS ratings. As far as MARS subscales are concerned, the average ratings on functionality and aesthetics were relatively higher (4.3 and 4.6, respectively) compared to ratings on information quality (3.9), mirroring the observations in the primary review.

The correction will appear in the online version of the paper on the JMIR Publications website, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Reference

1. Mehrotra S, Tripathi R, Sengupta P, Karishiddimath A, Francis A, Sharma P, et al. Evaluating characteristics and quality of mental health apps available in app stores for indian users: systematic app search and review. JMIR Mhealth Uhealth 2025 Sep 26;13:e79238 [FREE Full text] [doi: 10.2196/79238] [Medline: 41004798]

Submitted 22.Jan.2026; this is a non–peer-reviewed article;accepted 12.Mar.2026; published 30.Mar.2026.

Please cite as:
Mehrotra S, Tripathi R, Sengupta P, Karishiddimath A, Francis A, Sharma P, Sudhir P, Srikanth TK, Rao G, Sagar R
Addendum: Evaluating Characteristics and Quality of Mental Health Apps Available in App Stores for Indian Users: Systematic App Search and Review
JMIR Mhealth Uhealth 2026;14:e91934
URL: https://mhealth.jmir.org/2026/1/e91934
doi:10.2196/91934
PMID:

©Seema Mehrotra, Ravikesh Tripathi, Pramita Sengupta, Abhishek Karishiddimath, Angelina Francis, Pratiksha Sharma, Paulomi Sudhir, T K Srikanth, Girish Rao, Rajesh Sagar. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 30.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Effects of Telehealth Interventions for People With Parkinson Disease: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Minyue Sun[1*], MSN; Fuyou Tang[2*], MSN; Luo min[1], BS; Shiyu Wen[1], BS; Shuang Wang[3], BS; Huiping Jiang[3], BS

[1]Department of Neurology, Mianyang Central Hospital, School of Medicine, University of Electronic Science and Technology of China, Mianyang, China

[2]Mianyang Key Laboratory of Anesthesia and Neuroregulation, Department of Anesthesiology, Mianyang Central Hospital, School of Medicine, University of Electronic Science and Technology of China, Mianyang, Sichuan, China

[3]Department of Neurology, Mianyang Central Hospital, School of Medicine, University of Electronic Science and Technology of China, Mianyang, Sichuan, China

[*]these authors contributed equally

Corresponding Author:
Luo min, BS
Department of Neurology
Mianyang Central Hospital, School of Medicine
University of Electronic Science and Technology of China
Fucheng District
Mianyang
China
Phone: 86 18324126071
Email: 410244386@qq.com

Related Article:

Correction of: https://mhealth.jmir.org/2026/1/e70994

(JMIR Mhealth Uhealth 2026;14:e108495) doi:10.2196/108495

In "Effects of Telehealth Interventions for People With Parkinson Disease: Systematic Review and Meta-Analysis of Randomized Controlled Trials" [1], the designation for co–first authors was inadvertently omitted.

The author list was revised by adding an asterisk next to authors MS and FT and a footnote indicating their equal contribution, as follows:

> *Minyue Sun[1*], MSN; Fuyou Tang[2*], MSN; Luo min[1], BS; Shiyu Wen[1], BS; Shuang Wang[3], BS; Huiping Jiang[3], BS*
>
> **these authors contributed equally*

All other aspects of the article remain unchanged.

The correction will appear in the online version of the paper on the JMIR Publications website, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Reference

1. Sun M, Tang F, Min L, Wen S, Wang S, Jiang H. Effects of telehealth interventions for people with Parkinson disease: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2026 Jan 28;14:e70994 [FREE Full text] [doi: 10.2196/70994] [Medline: 41604670]

XSL•FO
RenderX

Submitted 01.Aug.2026; this is a non–peer-reviewed article;accepted 05.Aug.2026; published 01.Sep.2026.

Please cite as:
Sun M, Tang F, min L, Wen S, Wang S, Jiang H
Correction: Effects of Telehealth Interventions for People With Parkinson Disease: Systematic Review and Meta-Analysis of Randomized Controlled Trials
JMIR Mhealth Uhealth 2026;14:e108495
URL: https://mhealth.jmir.org/2026/1/e108495
doi:10.2196/108495
PMID:

©Minyue Sun, Fuyou Tang, Luo min, Shiyu Wen, Shuang Wang, Huiping Jiang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 01.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Incremental Value of Smartphone Sensing for Monitoring Momentary Affect Intensity in Adults Using Transformer-Based Models: Observational Study

Yiqin Zhu[1], MS, MA; Yuyi Yang[2], MPH; Renee J Thompson[1], PhD

[1]Department of Psychological and Brain Sciences, Washington University in St. Louis, 1 Brookings Drive, CB 1125, St. Louis, MO, United States
[2]Division of Computational and Data Sciences, Washington University in St Louis, 1 Brookings Drive, St. Louis, MO, United States

Corresponding Author:
Yiqin Zhu, MS, MA
Department of Psychological and Brain Sciences, Washington University in St. Louis, 1 Brookings Drive, CB 1125, St. Louis, MO, United States

Abstract

Background: Ubiquitous smartphone access and statistical advances offer opportunities to continuously track affect intensity, which is central to various psychological processes and behaviors. Research demonstrated the potential of personalized predictions of momentary negative affect (NA) and positive affect (PA) using passive sensing. However, studies typically incorporated all available data sources without differentiating their added value, nor did they investigate whether refining location features with self-reported semantic location (eg, workplaces) improved personalized predictions.

Objective: We evaluated three specific aims: (1) how different combinations of data sources improved performance compared to personalized baseline models, (2) whether model predictions differed across passive data aggregation timescales, and (3) whether incorporating self-reported semantic locations improved model predictions.

Methods: Adults (final n=133) completed a 14-day ecological momentary assessment (EMA) protocol reporting emotional experiences 5 times daily alongside smartphone sensing. Testing data (n=532 EMAs) used the last 4 surveys for each individual, with the remaining used for training and validation (n=6805 [NA]/6800 [PA] EMAs). We evaluated whether combinations of personalization, passive sensing, and affect history improved baseline prediction, and how full-information temporal fusion transformers (TFTs) performed across 6 timescales (1, 3, 6, 12, 24, and 48 hours), with or without self-reported semantic location features.

Results: The baseline model, using each individual's mean affect in the training set, demonstrated moderate predictive performance for NA (mean absolute error [MAE]=0.66, 95% CI 0.60 - 0.73; R^2=40.2%) and PA (MAE=0.71, 95% CI 0.65 - 0.78; R^2=36.1%). Full-information TFTs improved NA prediction (MAE=0.62, 95% CI 0.56 - 0.69; R^2=45.2%; ΔMAE=−0.04, 95% CI −0.06 to −0.01; *P* values ≤.004; Cohen *d*=−0.27) but not PA prediction (MAE=0.70, 95% CI 0.64 - 0.78; R^2=32.5%; ΔMAE=−0.01, 95% CI -0.03 to 0.02; *P* values >.10; Cohen *d*=−0.04). No pairwise timescale comparison survived false discovery rate (FDR) correction (NA: P_{FDR}=.05-.98; PA: P_{FDR}=.08-.99). Adding self-reported locations did not improve NA prediction (ΔMAE=0.02, 95% CI −0.01 to 0.04; *P* values >.20; Cohen *d*=0.11) or PA prediction (ΔMAE=0.01, 95% CI −0.01 to 0.03; *P* values>.33; Cohen *d*=0.07). However, incorporating self-reported semantic locations changed the composition and relative ranking of important inputs, with these changes varying across NA and PA and between past and future inputs.

Conclusions: Incorporating smartphone features provided a modest and significant improvement in momentary NA prediction, but not PA prediction. Model performance did not vary across passive data aggregation timescales. While adding self-report semantic locations did not improve prediction accuracy, it changed variable-importance patterns and may provide additional context for interpreting digital behavioral markers. Future personalized predictions should incorporate person-mean affect as an essential benchmark. These findings support passive smartphone sensing as a valuable supplement to, rather than a replacement for, active EMA.

(JMIR Mhealth Uhealth 2026;14:e90970) doi:10.2196/90970

KEYWORDS

transformer-based models; affect; smartphone sensing; ubiquitous computing; affective computing; digital phenotyping

XSL•FO
RenderX

Introduction

Background

Modern technologies already enable us to unobtrusively and accurately track sleep, physical activity, or health essentials such as heart rate and blood pressure, but what about affect intensities? Like the continuous tracking of human behaviors indicative of health status, continuous tracking of affect intensities holds broad research and health implications, given the central role of affect intensity in various outcomes (eg, attention [1], memory scope [2], and decision-making [3]) and its defining role in the diagnostic criteria of many mental disorders [4]. In evaluations for mood and anxiety disorders, for example, diagnostic impression relies on individuals' retrospective reports of their affect intensity patterns compared to their own baseline patterns (eg, intense negative affect [NA] or blunted positive affect [PA]) over an extended period in the past (eg, the past several months or years). However, retrospective reports are affected by recall bias, especially in clinical samples [5,6]. Assessing the intensity of one's subjective feelings from moment to moment reduces bias but requires constant attention that results in burden and missingness (eg, 30% missingness for data collection less than a month [7]). These limitations, bias in recalling experiences in the past, and the burden of subjective momentary assessment could be overcome by personalized models to passively and continuously compute affect intensities. Research suggested the potential of such personalized prediction using data from smartphone sensors [8-10]. However, existing studies combined all data sources in models without differentiating each source's contribution, particularly between an individual's own affective history and passively sensed behaviors. As a result, it remains unclear whether passive sensing meaningfully improves prediction beyond what could already be achieved from a person's own affective history (eg, person-mean affect). The overarching aim of this study was to parse out variances of different data sources from personalized prediction models. In addition, we used linear mixed-effects models and post hoc interpretability measures (ie, attention weights) from machine learning (ML) to generate hypotheses about relations between affect and smartphone-tracked behaviors.

Monitoring and Understanding Momentary Affect Intensity Through Smartphones

Cheap and wide access to smartphones (91% ownership in the United States [11]) provides unprecedented opportunities not only for passive monitoring of momentary affect intensity but also for understanding ubiquitous human-smartphone interactions. Five studies incorporated smartphone sensor data and modeled momentary affect intensities. Three of the studies dichotomized affective intensity as outcomes and achieved modest prediction with data from smartphones, smartwatches, and smartrings [8-10]. Research showed that traditional AI algorithms could make moderate predictions on whether an individual's daily affect intensities [9] or momentary affect intensities [10] were above a certain threshold, and make just-surpassing-chance predictions of the presence of affect in unseen individuals [8]. However, dichotomizing affect intensity loses important granular information, as different levels of intensity of the same type of emotion lead to different behavioral outcomes. For example, strong feelings of anger might lead to physical violence, whereas weak feelings of anger might lead to verbal profanity or no overt behavior. Therefore, affect intensity is, by nature, continuous and, therefore, motivates approaches that preserve variation across a continuous scale.

Research that continuously modeled momentary affect included predictors that require individuals' active engagement [12,13] and only focused on certain NA without modeling PA. One study predicted momentary depressed feelings with smartwatches and self-reported items [13]. Both momentary anxiety and depressed affect were reported at the same time, and anxiety was the most important predictor of momentary depressed affect in the models, above smartwatch predictors, for 11 out of 14 participants. The other study collected heart rate metrics (ie, participants pressed their finger against the rear camera for 30 seconds hourly), along with smartphone sensors, to predict momentary depressed mood [12]. Importantly, these models combined all sources and did not test against simple baselines of a person's mean affect. Thus, it remains unclear how much smartphone sensing specifically adds to personalized prediction. Finally, it is equally important to test personalized PA prediction using smartphone sensing. Theoretical models [14] and clinical trials [15,16] have highlighted the critical role of raising PA in reducing depression, and PA is conceptually independent of NA [17] and influenced by different biopsychosocial predictors from NA [18-20].

Smartphones are both a data-collection tool people carry and an object to interact with regularly, but few studies have examined associations between momentary affect intensity and smartphone-tracked behaviors. People engage in social interactions (eg, making phone calls) and interact with phone screens for multiple purposes, including using social media. Revealing relations between everyday affective experiences and behaviors tracked by smartphones could inform intervention strategies to aid emotion regulation. For example, if total screen use time in the prior hour is an important predictor of feeling worse among other smartphone-tracked behaviors (eg, total distance traveled), individuals could practice limiting their screen usage. One challenge for elucidating affect–smartphone-based behavior associations (and probably for most psychological research) is the large number of smartphone-tracked behaviors and the dependencies between them. For example, when individuals are traveling (as collected through GPS), they may make fewer phone calls and have fewer screen interactions. On the other hand, when individuals remain at home for an extended time, their likelihood of unlocking their screens may be higher. Thus, smartphone-tracked behaviors can be predictors of each other and work together to shape affective experiences.

Temporal fusion transformer (TFT) is well-suited to address the challenge brought by multiple and interacting predictors. As a state-of-the-art algorithm for time-series analysis, TFT may predict momentary affect intensity more accurately than traditional AI models while addressing the challenge brought by multiple and interacting predictors, as shown in its performance for other time-series predictions (eg, electricity,

XSL•FO
RenderX

traffic, retail, and volatility [21]). As a transformer-based model, TFT uses self-attention to capture dependencies across time, while its variable selection networks identify predictors that are most informative for the prediction task [21]. Additionally, TFT accommodates 3 types of inputs based on their temporal availability: static inputs (time-invariant inputs), past inputs (predictors before the time point of making predictions), and future inputs (inputs corresponding to prediction time points processed by the decoder [21]). TFT's capacity to distinguish the 3 input types is important for predicting momentary affect intensity. Because affect intensity is often viewed as an individual difference characteristic [22], adding static inputs may enhance prediction, although there is high intraindividual variability for momentary affect intensity. Additionally, it remained unclear at which lag length momentary affect intensity can be predicted by different smartphone-tracked behaviors. While traditional psychological research testing temporal associations usually adopts one lagging length (eg, lagging one assessment schedule), TFT could integrate sensing features derived across multiple temporal windows, allowing information from both proximal and more distal timescales to contribute to prediction.

Timescales and Semantic Locations

Passive sensing research showed that the timescales to aggregate passive sensing data may influence the interpretation of results on psychological well-being [23]. Research on predicting momentary affect with passive sensors has used different timescales [8-10,12,13], and there is a lack of empirical tests comparing the influence of timescales on model performance and predictor importance. For affect intensity, the same behavioral indicator from smartphone sensors with different timescales could have different clinical implications. Intuitively, staying stationary for most of the time in the past hour might be less influential on momentary affect intensities than staying stationary for most of the time in the past 48 hours. Thus, how TFTs performed differently across timescales of aggregating the smartphone sensor data warrants empirical investigation. The timescales might also change the attention that TFTs put on the mobility markers (as collected by GPS) and social interaction markers (as collected by phone calls). Since social interactions are acute events or stressors [24], proxies of social interactions tracked by smartphones may be important predictors for NA and PA intensities when they are aggregated in smaller timescales. By contrast, physical mobility patterns in shorter timescales may be influenced by random factors (eg, weather), whereas they may indicate behavioral routines when aggregated in larger timescales. Therefore, mobility patterns tracked by smartphones may be more important when aggregated on larger timescales.

Contextual features, such as the semantic meaning of a location to an individual (eg, where they work, or where they entertain, or where they spend time with important others), might improve both model performance and interpretations. Affective experiences are constructed by subjective evaluations of both internal and external cues [25,26]. A model of situation perception [27] proposes three situational cues: (1) persons and interactions (who?); (2) objects, events, and activities (what?); and (3) spatial location (where?). Physical locations recorded through GPS can provide information about "where" but not "what," namely, the activities or the motivations for activities. Semantic meaning of a location may offer such information. For example, one may assume the major activities that individuals engage in are work in their self-reported workplaces, or that the major activities that they engage in are recreational in their self-reported leisure places. In prior studies, usually one semantic location feature (time spent at home) was included [11,28,29]. Collecting self-report locations from individuals at baseline may improve the coding accuracy of semantic locations and help create new location features, thereby potentially improving model performance. Additionally, semantic locations could inform intervention strategies, as they may elicit more awareness of what patients are engaging in than where they are. Overall, the role of semantic locations in predicting momentary affect warrants further investigation.

Study Aims and Hypotheses

This study collected data from a sample of unselected adults who completed an ecological momentary assessment (EMA) protocol measuring self-report emotional experiences alongside continuous smartphone behavioral monitoring. For aim 1, we evaluated the incremental predictive value of contributions of personalization (participant identity and affective history) versus smartphone sensing data in momentary affect prediction, by comparing TFT and other ML models that vary in their usage of these data sources against a person-mean affect baseline. We hypothesized that the full TFT model incorporating all data sources would achieve prediction performance for unseen momentary NA and PA above the person-mean-affect baseline (hypothesis 1). For aim 2, we examined how feature extraction across 6 aggregation timescales (1, 3, 6, 12, 24, and 48 hours preceding the report) affected TFT's performance. We explored the optimal aggregate windows (exploratory aim 1) and identified top behavioral predictors across timescales. We hypothesized that social interaction (eg, call logs) and screen activity would demonstrate greater feature importance at proximal timescales (hypothesis 2a), whereas mobility markers might be more important at longer timescales (hypothesis 2b). For aim 3, we examined the impact of integrating participant-annotated semantic location features (eg, work and home of significant other). We hypothesized that refining semantic location features with participant self-reports would improve model performance (hypothesis 3a) and shift variable importance rankings (hypothesis 3b). Because attention weights reflect predictive relevance rather than causal mechanisms, findings from hypotheses 2a, 2b, and 3b are framed as exploratory and hypothesis-generating.

Methods

Participants

A total of 179 adults from the greater area of St. Louis, Missouri, United States, participated in a study to understand daily emotional experiences [30,31]. Individuals were eligible if they had a smartphone and at least one active social media profile. They were ineligible if they had any contraindications for peripheral physiological assessment (eg, pregnant and with an implanted cardiac device). Participants with less than 40 valid

XSL•FO
RenderX

EMAs (70 possible surveys) on NA or PA (n=46) were excluded from further analysis. As a result, 133 out of 179 (74.3% of original participants) were retained in final analysis. Although there are no recommendations or requirements on the number of time points within each individual, in general, more time points are better. The time points ranged from 90 to 252 in the original TFT paper [21]. Considering the sample size to be retained, we decided on the threshold of 40 EMAs such that the rate of sample size to be retained in this study (74.3%) was comparable to or higher than existing studies that predict momentary affect intensity (71.6% [8]; 35% [9]).

Ethical Considerations

The study was approved by the Washington University Institutional Review Board (number 202209063) and was performed in line with the standards of the 1964 Declaration of Helsinki. Informed consent and assent were collected from all participants. To protect participant privacy, all data were deidentified before analysis and stored on secure, password-protected servers. Participants were compensated US $135 for completing the study. There was a US $15 bonus if participants completed 80% (56/70) surveys or more. Thus, participants can receive up to US $150 depending on their survey completion rates.

Study Design and Procedure

Eligibility was determined via a phone screen. Participants were scheduled for an in-person laboratory session. During the laboratory session, participants completed a semistructured EMA tutorial, including a practice EMA survey, administered by a staff member or an undergraduate research assistant. The tutorial included helping participants download and install the SEMA^3 app (developed and hosted by the Melbourne eResearch Group at the University of Melbourne [32]) for EMA surveys and the AWARE ([33]; University of Oulu) apps for smartphone sensing. EMA surveys occurred during a 15-hour window of the participants' choice. Participants were surveyed 5 times a day for 14 days, starting the day after the laboratory session. At the laboratory session, participants also reported anticipated locations they frequently visited (eg, office and home) and their addresses in the next 14 days. The AWARE framework [33] is an open-source mobile app to unobtrusively record location coordinates, screen interactions (ie, when the screen status changed to on or off and locked or unlocked), call logs for incoming, outgoing, and missed calls, and Bluetooth connections. The AWARE started to collect deidentified data unobtrusively from the smartphone sensors after installation. By default, location coordinates are sampled once per 180 seconds, and Bluetooth is sampled once per 60 seconds. Calls and screen usage are event-based sensor streams. Bluetooth data were not included in analyses due to high missingness. A total of 156 out of 179 (87.2%) participants had no Bluetooth scans recorded in the hour preceding any EMA prompt, and a total of 103 out of 179 (57.5%) participants had no Bluetooth data at all.

The original sample completed a total of 8690 out of 12,530 (69.4%) EMA surveys. On average, individuals completed 49 out of 70 (70%; SD 15.1) EMA surveys. The completion rate was higher in the final analytic sample, which on average completed 56 out of 70 (80%) EMAs because of the selection criteria (see the "Participants" section). Participants could receive up to US $135 for completing the study and a bonus of US $15 if their EMA completion was more than 80%, resulting in a possible compensation of $150.

Measures

EMA: Momentary Affect Intensity

In each survey, participants were asked to report on how they felt in the hour preceding the EMA survey. They indicated the extent to which they felt 4 PA and 4 NA using a 7-point Likert scale (1=not at all; 7=extremely). Items were chosen to represent a 2-dimensional affective circumplex (ie, valence and arousal). The PA items were "During the last hour, I felt CONTENT/CALM/HAPPY/ENTHUSIASTIC," and the NA items were "During the last hour, I felt SAD/SLUGGISH/WORRIED/FRUSTRATED." In the final analytic sample, both NA and PA showed acceptable within-person reliability (NA: ω_{within}=0.64, PA: ω_{within}=0.69) and great to excellent between-person reliability (NA: $\omega_{between}$=0.88, PA: $\omega_{between}$=0.90).

Measures Derived From Passive Sensors

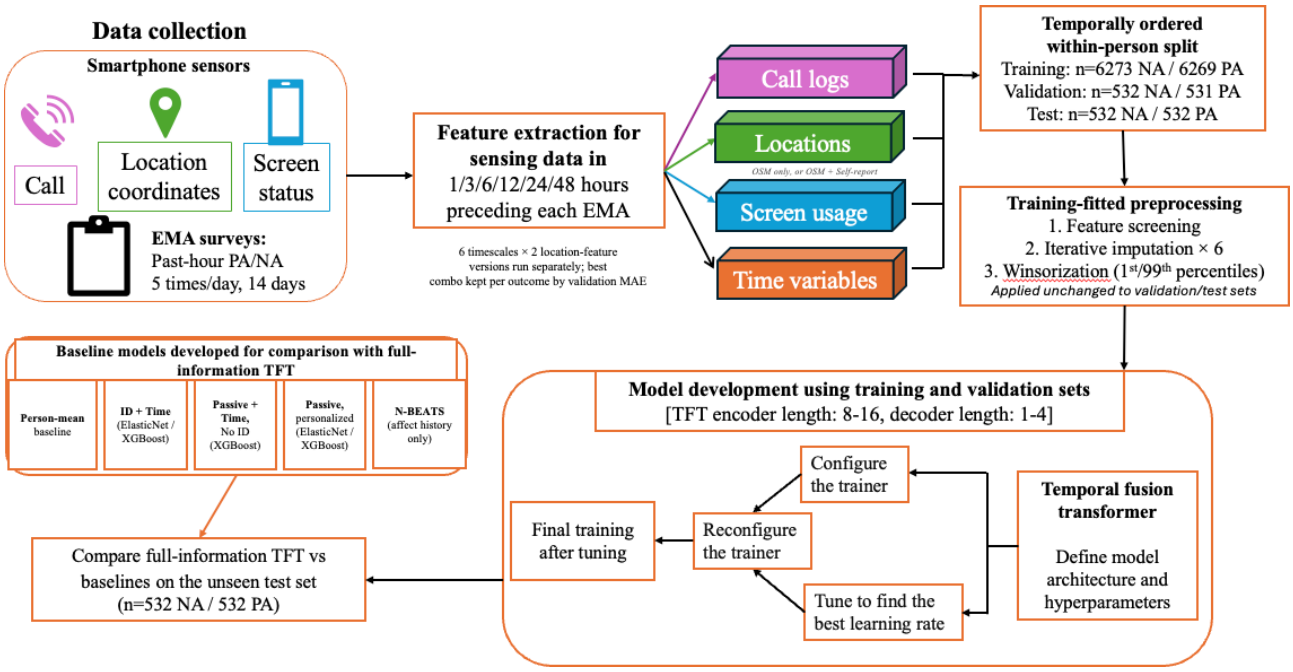
Overview

Figure 1 illustrates the pipeline of data collection, cleaning, and analysis.

Table 1 presents a list of extracted features. We computed features from 3 smartphone sensors (location coordinates, screen interactions, and call logs) because of their potential to predict depression severity [28,29,34,35] or predict dichotomized momentary affect intensity [10].

XSL•FO
RenderX

Figure 1. Data collection, cleaning, and analysis pipeline. EMA: ecological momentary assessment; ElasticNet: elastic net regression; MAE: mean absolute error; NA: negative affect; N-BEATS: neural basis expansion analysis for interpretable time series forecasting; PA: positive affect; TFT: temporal fusion transformer; XGBoost: extreme gradient boosting.



XSL•FO
RenderX

Table . Predictors derived from smartphone sensors and their sources[a].

Variable type and sensor source	Variables
Spatial mobility markers	
GPS coordinates [28,34,35]	• Location variance • Log of location variance • Total distance traveled • Average speed • Variance in speed • Percentage of time being stationary (stationary: momentary speed <1 km/h) • Radius of gyration
GPS coordinates clustered by DBSCAN[b] [28,35]	• Number of significant places • Number of significant transitions • Percentage of time spent in rarely visited locations (ie, outliers by DBSCAN) • Location entropy • Normalized location entropy
Clusters labeled by OSM[c] or self-report locations	• Percentage of time spent at home (OSM) • Percentage of time spent at leisure places (OSM) • Percentage of time spent at home (OSM + self-report) • Percentage of time spent at leisure places (OSM + self-report) • Percentage of time spent at workplaces (OSM + self-report) • Percentage of time spent at home of significant others (OSM + self-report)
Proxies of smartphone social interactions	
Call logs [28]	• Number of incoming calls • Number of outgoing calls • Number of missed calls • Number of correspondents called • Duration of incoming calls (average, maximum, minimum, and total duration) • Duration of outgoing calls (average, maximum, minimum, and total duration)
Screen interactions	
Screen status [36]	• Number of unlocked episodes • Average, maximum, minimum, and total screen-on duration
Time variables	
Timestamp	• Relative time in hours since the first self-report EMA[d] survey • Weekend or weekday (1=weekend; 0=weekday) • Time of day (night: midnight to 6 AM; morning: 6 AM to noon; afternoon: noon to 6 PM; evening: 6 PM to midnight) • Prompt (the number of the self-report EMA survey)

[a]Each predictor was computed for each timescale. For example, for the timescale of 3 hours, the location variance of the 3 hours preceding each survey was created.

[b]DBSCAN: Density-Based Spatial Clustering of Application with Noise.

[c]OSM: OpenStreetMap.

[d]EMA: ecological momentary assessment.

Location Feature Sets

Location features are derived from the coordinates of latitude and longitude and the timestamp of each coordinate. We first removed duplicated data points for each individual and then extracted the following location features consistent with research [28] that detected depression and predicted its onset from smartphone data: (1) location variance (sum of the variance in latitude and longitude coordinates), (2) log of location variance, (3) total distance traveled, (4) average speed, (5) variance in speed, (6) the percentage of time being stationary (stationary was defined as momentary speed being less than 1 km/h), and (7) radius of gyration. When calculating distances and speed, only adjacent data points whose time difference was greater than 1 second and less than twice the sampling rate (ie, 180 seconds) were retained.

XSL•FO
RenderX

Then each individual's location coordinates were clustered using Density-Based Spatial Clustering of Applications with Noise (DBSCAN [37]), a clustering algorithm for large spatial databases with noise. Consistent with prior work [28], based on identified clusters, we extracted the (1) number of significant places, (2) the number of location transitions, (3) the percentage of time spent in insignificant or rarely visited locations (ie, outliers identified by DBSCAN), (4) location entropy (higher location entropy occurs when time is spent evenly across significant places), and (5) normalized location entropy across significant places (following previous definitions and operationalizations [35]).

Consistent with prior work [11,28,29], we assumed the place most visited by the participant late at night (between midnight and 6 AM) was their home location. Specifically, the location coordinates of the cluster most frequently used during all nights were assumed to be the participant's home location center. Then, based on the coordinates of longitude and latitude for each individual's core clusters, we labeled their type of location through the OpenStreetMap Nominatim API's geocoding [38]. Locations labeled as "highway," "aeroway," "tourism," "leisure," and "shop" by OpenStreetMap were categorized as leisure and recreation. For each of the location labels (including the labels in the next paragraph), when participants' GPS coordinates were within 100 meters of one location center cluster, their location at that timestamp was given the label of the location center cluster. For example, a participant was assumed to be at home if they were within 100 m of the home location. With these 2 labels, we extracted two location features: (1) the percentage of time spent at home and (2) the percentage of time spent at leisure and recreation.

We tested whether adding more semantic location features could improve model prediction and interpretation through their indication of possible activities. These sets of location features were extracted based on participants' self-reported locations, which included categories of personal home, home of significant others (eg, parents, partner, and so on), workplace, and places for leisure and recreation. Based on their self-report addresses, the locations of the previously coded home and leisure were corrected, resulting in two extracted features: (1) the percentage of time spent at home (self-report) and (2) the percentage of time spent at leisure and recreation (self-report). Additionally, with 2 additional addresses (workplace and home of significant others), we extracted (3) the percentage of time spent at work (self-report) and (4) the percentage of time spent with significant others (self-report).

We created 6 sets of location features based on 6 temporal resolutions (ie, 1, 3, 6, 12, 24, and 48 hours) preceding each momentary affect (ie, NA and PA) report. Take the location variance feature as an example. We created 6 location variance features for a given momentary affect report—the location variance during 1, 3, 6, 12, 24, and 48 hours preceding the report.

Calls Feature Sets

Call features were calculated using the smartphone's call logs. Consistent with prior work [28], we extracted the following six features: (1) the number of all incoming and outgoing calls, (2) the number of missed calls and the number of correspondents overall, and (3) the duration of all incoming and outgoing calls (in terms of their average, maximum, minimum, and total duration). As with the location features, we created 6 sets of call features based on 6 temporal resolutions (ie, 1, 3, 6, 12, 24, and 48 hours preceding each momentary affect report).

Screen Feature Sets

Screen features were calculated using the smartphone's screen status sensor, which recorded screen status (ie, on or off) and its timestamp once the screen switched from one status (eg, on) to the other (eg, off). Consistent with prior work [28], we extracted the following five phone usage features: (1) the number of unlock episodes, (2) the average, (3) sum, (4) maximum, and (5) minimum duration of the screen-on time. Similar to the location features, we created 6 sets of screen features based on 6 temporal resolutions (ie, 1, 3, 6, 12, 24, and 48 hours preceding each momentary affect report).

Time Variables

We included several time variables based on the time of the EMA survey completion as predictors of momentary affect because of emotion's time-varying nature. We computed one continuous time variable, (1) relative time (relative time in hours since the first EMA survey), and 2 categorical variables: (2) weekend or not (1=yes; 0=no [ie, weekday]) and (3) time of day (midnight to 6 AM: night, 6 AM to noon: morning, noon to 6 PM: afternoon, 6 PM to midnight: evening; dummy coded). Finally, one count variable, (4) prompt (the EMA survey number), was used as the time indicator required by the time-series deep learning algorithm.

Data Preprocessing

Before imputation, we screened the passive-sensing predictors for near-constant and redundant features: a predictor was dropped if (1) it was near constant, defined as a single value accounted for at least 99% of its nonmissing observations, (2) it was highly correlated with another one, defined as an absolute Spearman correlation over 0.95. We dropped the one with higher missingness in the highly correlated pair. We then handled missing data and extreme values of the smartphone sensor features for each timescale by performing 6 imputations and then 6 winsorizations. For each of the 6 feature sets, we performed single imputation for the missing values in predictors using an iterative imputer (IterativeImputer) that uses the chained equations (MICE [Multivariate Imputation by Chained Equations]) algorithm from the scikit-learn library [39] in Python. Instead of performing imputations for each smartphone sensor (locations, calls, and screen), we performed imputations across smartphone sensor data simultaneously, in which IterativeImputer could use interactions between different sensor data to impute missing values. Importantly, we excluded outcome variables (ie, NA and PA) from imputations to avoid data leakage and overfitting of the prediction models. Finally, we winsorized extreme values in the predictors using the quantile-based method, where we replaced values above the 99th percentile with the 99th-percentile value and replaced values smaller than the 1st percentile with the 1st-percentile value.

XSL•FO
RenderX

To avoid data leakage, the feature screening, feature selection, imputation, and winsorization were all performed in the training set only. The feature list, imputation model, and clipping bounds from the training set were then applied to the validation and test sets. For the TFT, the same principle applied to the participant-level target normalizer used to construct the NA and PA center and scale, which was fit on the training set and reused, rather than refit, in the validation and test sets. A similar training-only-fitted preprocessing pipeline (including standardization for the ElasticNet [elastic net regression]) was also performed for approaches using the ElasticNet and XGBoost (extreme gradient boosting).

Training and Testing Split

We used a temporally ordered split approach, splitting observations within each individual into three subsets: (1) a test set for final model evaluation, comprising the last 4 consecutive EMA surveys (~1 day); (2) a validation set for hyperparameter tuning, early stopping, and, for model selection between ElasticNet, XGBoost, and N-BEATS (neural basis expansion analysis for interpretable time series forecasting; described below), comprising the 4 consecutive EMA surveys immediately preceding the test window; (3) a training set with all earlier entries (~85% of all EMA surveys). Because the validation and test windows followed the training window, this setup prevents temporal leakage and evaluates short-term within-person generalization. This approach resulted in 6273/6269 EMAs in the training set, 532/531 in the validation set, and 532/532 in the test set for NA and PA, respectively. The training set of 6273 (NA) and 6269 (PA) EMAs allowed 6672 (NA) and 6669 (PA) rolling training sequences (encoder length: 8 - 16, prediction length: 1 - 4; window lengths were allowed to be flexible and can overlap, so they can exceed observation counts), across 133 participants. The small NA and PA differences reflect outcome-specific missingness. Because the sliding window allows the same EMA to be reused as parts of multiple different-length windows (eg, first window being 1 - 11 EMAs predicting 12 - 15 EMAs and second window being 2 - 13 EMAs predicting 14 - 16 EMAs), the number of training sequences to fit the TFT exceeds the number of EMAs in the training set.

For the TFT, we allowed the encoder length to vary between 8 and 16 time points and the prediction length to vary between 1 and 4 time points. As a result, training sequences consist of different combinations of encoder and decoder lengths (eg, 1 - 11 EMAs predicting the next 2 EMAs). The final TFT configuration used an encoder length of 16 observations because this window spans roughly 3 days of recent data, given the study's sampling schedule of approximately 5 EMAs per day. We considered this interval sufficiently long to capture short-term temporal patterns in affect while minimizing data loss that would occur from requiring longer continuous observation histories. A prediction length of 4 was chosen to forecast affect over approximately the subsequent day, directly aligning with the study's objective of predicting near-term affective dynamics which may ultimately be relevant for timely, adaptive interventions. The final TFT configuration used a maximum encoder length of 16 observations and a maximum prediction length of 4 observations.

Statistical Analysis

Model Evaluations

Model performance metrics were mean absolute error (MAE), root-mean-square error (RMSE), and coefficient of determination (R^2). The MAE is the average of all the absolute values of the differences between the predicted values and the actual values. The RMSE is the square root of the mean of the squared differences between predicted values and actual values. R^2 represents the proportion of variance explained by the model. When R^2 equals zero, a model explains no more than the average score. When R^2 equals 1, a model perfectly explains all the outcome variances. Unlike R^2, lower values of MAE and RMSE indicate better performance. Both MAE and RMSE are in the original scale of the outcomes and indicate the extent to which predicted values deviate from the actual values. RMSE is more sensitive to extreme deviations. When MAE or RMSE equals zero, the model perfectly predicts the actual values. When RMSE equals the SD of the outcome variable, the model performs no better than the average score.

Person-Mean Affect: For Baseline

In this baseline model, the predicted value for each held-out validation or test observation was set to each participant's mean NA and PA intensity calculated across the training set. This benchmark required no model fitting or hyperparameter tuning and operated without smartphone sensing or temporal information.

ElasticNet and XGBoost Models: For Nonpersonalized Sensing, Personalized Time, and Personalized Sensing Approaches

ElasticNet regression [40], implemented in scikit-learn, and XGBoost (gradient-boosted trees) [41] were used to fit three model variants: (1) nonpersonalized sensing model, (2) personalized time (incorporating participant ID and time variables, without smartphone-sensing features), and (3) personalized sensing (incorporating participant ID, time, and smartphone sensing features, without affect history). Both ElasticNet and XGBoost used a standardized preprocessing pipeline (z-scoring of continuous predictors) prior to model fitting. For ElasticNet, we evaluated 21 hyperparameter combinations crossing 7 regularization strengths (α=0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 1.0) with 3 L1 mixing parameters (l1_ratio=0.1, 0.5, 0.9), each fit for up to 20,000 iterations. For XGBoost, we searched 6 candidate configurations varying the number of trees (200-500), maximum tree depth (2-3), learning rate (0.02 - 0.05), row and column subsampling (0.8 - 1.0), minimum child weight (1-5), and L1/L2 regularization strength. Within each of the 6 sensor aggregation timescales and 2 semantic-location label versions, the hyperparameter configuration with the lowest validation-set RMSE was selected. Across test sets, the specific timescale and semantic label specifications that yielded the lowest test-set MAE were selected and reported as the top-performing model under each.

XSL•FO
RenderX

N-BEATS Model: For Personalized Affective History Approach

To evaluate the personalized affect-history approach, we implemented N-BEATS, a neural network for time-series forecasting composed of stacks of fully connected blocks with backward (backcast) and forward (forecast) residual connections. Unlike the TFT, N-BEATS relied exclusively on each participant's own historical affect sequence as input, incorporating no smartphone-sensor features. We used fixed-length sequence windows of 16 encoder time points and 4 prediction time points to match the TFT's maximum window. Outcomes were normalized within participants using a group normalizer with a softplus transformation, and missing time steps were permitted. We set the N-BEATS stack widths to 32 and 64 and the backcast loss ratio to 0, thereby weighting the training loss entirely toward the forecast; all other architecture parameters retained the PyTorch Forecasting defaults. Training used a learning rate of 0.001, weight decay of 0.01, batch size of 64, and gradient clipping at 0.01. We monitored validation loss and retained the checkpoint with the lowest validation loss. Training stopped after 5 consecutive validation checks without an improvement of at least 0.0001, with a maximum of 30 epochs. Training was deterministic and used a random seed of 42. This fixed-window procedure yielded 3746 (NA) and 3741 (PA) windowed training sequences derived from the same underlying training observations described above.

Temporal Fusion Transformer: For Full-Information and Personalized Sensing Approaches

To test hypothesis 1, we compared TFTs' performance (ie, R^2, MAE, and RMSE) with that reported in existing studies [12,13]. For exploratory aim 1, we examined TFTs' performance across 6 temporal scales (paired n=133). To test hypotheses 2a and 2b, the distal and proximal effects of different types of smartphone-tracked behaviors, we evaluated the top feature importance weights across timescales across encoder and decoder inputs. Because raw TFT variable importance weights reflect nonlinear transformations, we cross-referenced top predictors with statistically significant effects from within-person linear mixed-effects models (detailed below). We used up to 16 consecutive data points (encoder 8 - 16) of smartphone sensor data (past inputs) to predict the following up to 4 time points (decoder 1 - 4; future inputs) of momentary affect. The important predictors in the past inputs represent more distal factors, whereas future inputs represent more proximal factors in predicting momentary affect. In addition, important predictors in larger timescales and past inputs indicate features with more distal effects, whereas those in smaller timescales and future inputs indicate more proximal or immediate features. For hypothesis 3a, which tested whether self-reported semantic-location features changed performance, we used the paired participant-clustered bootstrap and participant-level Wilcoxon procedures described for aim 1. To test hypothesis 3b, we compared the pattern of important past and future inputs before and after incorporating self-report semantic location features.

All TFT models were implemented in the PyTorch Lightning framework through Google Colab Pro (NVIDIA A100-SXM4-40GB, 40 GB RAM). Model configurations were: (1) initialized the learning rate at 0.001 and then determined it via a systematic finder. The learning rate suggested by the systematic finder was scaled down by a factor of 4 in model training; (2) set the hidden size to 64 and hidden continuous size of 32 to deal with complex features; (3) added 4 multihead attention heads to capture multiscale temporal dependencies; (4) applied a dropout rate of 0.1 and gradient norm clipping threshold of 0.1 in the network to avoid overfitting; (5) used the Quantile loss function for probabilistic interval predictions; and (6) set the batch size to 10 to balance memory footprint and gradient estimation stability.

Using the validation set described above, we monitored validation loss after each epoch, retained the checkpoint with the lowest validation loss, and applied early stopping after 5 consecutive epochs without a minimum improvement of 0.0001 (up to 100 epochs). The retained checkpoint was used for all subsequent test-set evaluation. To make personalized predictions, the full-information TFT included 4 static covariates: participant ID, sequence encoder length, NA or PA center (individual's mean affect intensities), and NA or PA scale (individual SD of affect intensities). To isolate variance sources for our aim 1, we specified a separate personalized sensing TFT model by refitting the optimal specification for NA and PA after omitting static affect-history statistics (NA or PA center and NA or PA scale). Test-set performance for this model was compared against the ElasticNet and XGBoost baselines to select the best models under the personalized sensing approach.

CIs and Significance Testing

We computed 95% CIs for R^2, RMSE, and MAE using a participant-level bootstrap. To account for the nonindependence of repeated observations within individuals, resampling was performed at the participant level rather than the observation (EMA) level. Specifically, participants (not individual rows) were resampled with replacement for 2000 resamples, and all test-set rows of each selected participant were pooled. Model performance (R^2, RMSE, and MAE) was computed for each resample, and the 2.5th and 97.5th percentiles of the resulting bootstrap distribution were the CI.

To test aim 1, which was to evaluate whether each approach's held-out predictive performance differed from the baseline (person-mean affect), we conducted 2 complementary paired statistical tests applied to each approach's best-performing specification (see above). First, we performed a primary bootstrapping comparison by calculating the MAE difference ($\Delta MAE = MAE_{Approach} - MAE_{Baseline}$) for each EMA in the test set. Using 2000 resamples at the participant level (with a fixed random seed of 42 for reproducibility), we derived a 95% CI and a 2-sided bootstrap P value (defined as twice the smaller proportion of resampled ΔMAE values falling at or below zero vs at or above zero). Second, as a supplementary test, we conducted a participant-level paired Wilcoxon signed-rank test comparing paired series of mean per-participant $MAE_{Approach}$ and $MAE_{Baseline}$ values, avoiding the assumption of independent repeated observations. To remain conservative, an approach's improvement over the baseline was considered statistically

XSL•FO
RenderX

significant only when both tests met the significance threshold of α=.05. Effect sizes for these paired comparisons were reported as Cohen d_z (mean of participant-level ΔMAE divided by its SD, accompanied by approximate 95% CIs).

For exploratory aim 1 (evaluating whether TFT performance differed across the 6 temporal aggregation windows), we conducted a nonparametric Friedman test on a participant-by-window matrix containing each participant's mean MAE at each timescale (n=133). When the omnibus test indicated significant differences across windows, we conducted a post hoc analysis using all 15 pairwise Wilcoxon signed-rank tests between timescales, adjusting for multiple comparisons using the Benjamini-Hochberg false discovery rate (FDR) procedure. For hypothesis 3a (testing whether incorporating self-reported semantic-location features improved model performance), we applied the same dual-testing procedure as in aim 1, a paired participant-clustered bootstrap alongside a participant-level paired Wilcoxon signed-rank procedure.

Linear Mixed-Effect Models

Since the TFTs extract the importance of a feature but not the direction or the magnitude of associations, we further examined the within-person associations with a series of linear mixed-effects models. Models were estimated using restricted maximum likelihood (REML) via the lmer() function in the *lme4* package in R (R Core Team; accessed via pymer4 in Python). All variables were standardized (z-scored) before analysis to yield standardized coefficients β. All predictors were person-mean centered to only investigate within-individual associations. In each linear mixed-effect model, momentary NA or PA intensities were outcomes and were regressed on one smartphone-tracked predictor (eg, person-mean-centered total distance traveled in the 3 hours preceding the survey). In each model, we included a random intercept to handle between-person baseline differences in affect and included a random slope to allow the relationship between the predictor and outcome to vary across individuals.

Results

Participants Characteristics

The final analytic sample consisted of 133 adults (mean age 36.4, SD 12.0, age range: 19 - 63 years). Participants' gender composition was as follows: 71 out of 133 (53.4%) women, 51 out of 133 (38.3%) men, 3 out of 133 (2.3%) gender diverse, and 8 out of 133 (4.9%) unknown. Their race was 91 out of 133 (68.4%) White, 14 out of 133 (11.8%) Asian, 12 out of 133 (6.9%) Black or African American, 7 out of 133 (5.3%) others, and 9 out of 133 (6.8%) preferred not to say. A total of 9 out of 133 (6.8%) of the analytic sample reported being Latinx or Hispanic.

Aim 1: Model Performance in Predicting Momentary NA and PA

Table 2 presents the best-performing model under different approaches (ie, including different sets of predictors and algorithms). Results showed that models that simply used the mean affect intensity for each person (predictors: ID + training NA or PA) showed moderate performance on unseen test sets (NA: R^2=40.2% [95% CI 31-48.2], RMSE=0.89 [95% CI 0.80-0.97], MAE=0.66 [95% CI 0.60-0.73]); PA: R^2=36.1% [95% CI 20.4-48.3], RMSE=0.93 [95% CI 0.84-1.03], MAE=0.71 [95% CI 0.65-0.78]). Nonpersonalized sensing models performed significantly worse than the baseline for both outcomes (both P values<.001; NA Cohen d=0.56; PA Cohen d=0.37). Among personalized benchmarks, the ID-and-time ElasticNet was significantly worse for NA (ΔMAE=0.020, 95% CI 0.004-0.036; $P_{bootstrap}$=.01; $P_{Wilcoxon}$=.04; Cohen d=0.21), whereas the affect-history N-BEATS model did not differ for NA. N-BEATS produced a statistically significant but very small improvement for PA (ΔMAE=−0.002, 95% CI −0.003 to−0.000; $P_{bootstrap}$=.04; $P_{Wilcoxon}$=.01; Cohen d=−0.18). The personalized sensing TFT without Affect Center and Scale did not differ from the baseline for either outcome. The full-information TFT significantly improved NA prediction (ΔMAE=−0.040, 95% CI −0.065 to−0.015; $P_{bootstrap}$=.001; $P_{Wilcoxon}$=.004; Cohen d=−0.27) but not PA prediction (ΔMAE=−0.006, 95% CI −0.031 to 0.020; $P_{bootstrap}$=.66; $P_{Wilcoxon}$=.13; Cohen d=−0.04). The best full-information models used 6-hour features with self-reported semantic locations for NA and 1-hour features with self-reported semantic locations for PA.

XSL•FO
RenderX

Table . Prediction performances of different model approaches when predicting momentary NA[a] and PA[b] on the test dataset (n=133 individuals; total 532 surveys)[c].

Approach	Model	Feature sets	R^2 (%) (95% CI)	RMSE[d] (95% CI)	MAE[e] (95% CI)	ΔMAE vs baseline (95% CI)	*P* value (bootstrap)	*P* value (Wilcoxon)	Cohen *d*
NA									
Person-mean NA	Mean training NA	No passive sensing	40.2 (31.0 to 48.2)	0.89 (0.80 to 0.97)	0.66 (0.60 to 0.73)	—[f]	—	—	—
Nonpersonalized sensing	XGBoost[g]	48h+OSM[h]	4.6 (−2.3 to 10.5)	1.12 (1.01 to 1.23)	0.89 (0.80 to 0.97)	0.222 (0.158 to 0.295)	<.001	<.001	0.56
Personalized time	ElasticNet[i]	No passive sensing	38.2 (29.6 to 45.6)	0.90 (0.81 to 0.99)	0.68 (0.62 to 0.75)	0.020 (0.004 to 0.036)	.01	.04	0.21
Personalized sensing	TFT[j]	6h+OSM+self-report locations	36.6 (26.3 to 45.8)	0.91 (0.82 to 1.01)	0.66 (0.59 to 0.73)	−0.002 (−0.020 to 0.018)	.86	.89	−0.01
Personalized NA history	N-BEATS[k]	No passive sensing	35.5 (25.3 to 44.5)	0.92 (0.82 to 1.02)	0.66 (0.59 to 0.74)	−0.001 (−0.022 to 0.019)	.93	.88	−0.01
Full-information	TFT	6h+OSM+self-report locations	45.2 (33.8 to 54.3)	0.85 (0.76 to 0.94)	0.62 (0.56 to 0.69)	−0.040 (−0.065 to −0.015)	.001	.004	−0.27
PA									
Person-mean PA	Mean training PA	No passive sensing	36.1 (20.4 to 48.3)	0.93 (0.84 to 1.03)	0.71 (0.65 to 0.78)	—	—	—	—
Nonpersonalized sensing	XGBoost	48h+OSM+self-report locations	10.8 (1.4 to 18.7)	1.10 (1.01 to 1.19)	0.87 (0.80 to 0.95)	0.163 (0.091 to 0.239)	<.001	<.001	0.37
Personalized time	ElasticNet	No passive sensing	35.8 (20.5 to 47.2)	0.93 (0.85 to 1.03)	0.72 (0.66 to 0.79)	0.009 (−0.008 to 0.028)	.34	.46	0.09
Personalized sensing	TFT	1h+OSM+self-report locations	34.8 (17.7 to 48.2)	0.94 (0.84 to 1.05)	0.70 (0.64 to 0.78)	−0.007 (−0.029 to 0.016)	.59	.21	−0.05
Personalized PA history	N-BEATS	No passive sensing	35.4 (20.2 to 48.1)	0.93 (0.83 to 1.03)	0.71 (0.65 to 0.77)	−0.002 (−0.003 to −0.000)	.04	.01	−0.18
Full-information	TFT	1h+OSM+self-report locations	32.5 (11.6 to 47.8)	0.96 (0.84 to 1.08)	0.70 (0.63 to 0.78)	−0.006 (−0.031 to 0.020)	.66	.13	−0.04

[a]NA: negative affect.

[b]PA: positive affect.

[c]R^2, RMSE, and MAE are computed on the held-out test set; brackets denote the 95% participant-level bootstrap CI (2000 resamples with replacement). ΔMAE vs Baseline=$MAE_{approach}$- $MAE_{Person\text{-}Mean\ Baseline}$, calculated from a paired participant-level bootstrap (positive values indicate worse performance than baseline, negative values indicate improvement). The *P* (bootstrap) value represents the 2-sided bootstrap *P* value for this difference; *P* value (Wilcoxon) is derived from a participant-level paired Wilcoxon signed-rank test on MAE (n=133 participants rather than 532 observations, to avoid pseudoreplication). Cohen *d* indicates the paired-samples effect size (mean difference divided by the SD of the MAE differences). Bold text indicates a statistically significant improvement or decline relative to the person-mean baseline (requiring both bootstrap and Wilcoxon *P*<.05). For approaches incorporating passive sensing feature sets (combinations of time windows and semantic-location label sets), the row displays the best-performing specification based on test-set MAE.

[d]RMSE: root-mean-square error.

XSL•FO
RenderX

[e]MAE: mean absolute error.

[f]Not applicable.

[g]XGBoost: extreme gradient boosting.

[h]OSM: OpenStreetMap.

[i]ElasticNet: elastic net regression.

[j]TFT: temporal fusion transformer.

[k]N-BEATS: neural basis expansion analysis for interpretable time series forecasting.

Aim 2: Role of Timescales in Model Performance and Variable Importance

Overview

Figure 2 shows MAE across the 6 aggregation windows. In the figure, shaded bands represent 95% participant-level bootstrap CIs. In-panel statistics report the Friedman omnibus test across all 6 time windows and the range of FDR-corrected pairwise Wilcoxon *P* values across all 15 window-pair comparisons for each semantic-location feature set. For NA, neither the OpenStreetMap-only specification (χ^2_5=5.93, *P*=.31) nor the OpenStreetMap-plus-self-report specification (χ^2_5=10.10, *P*=.07) differed significantly across windows. FDR-corrected pairwise *P* values ranged from .05 to .98. For PA, the OpenStreetMap-only omnibus test was nonsignificant (χ^2_5=2.16, *P*=.83), whereas the OpenStreetMap-plus-self-report omnibus test reached significance (χ^2_5=11.21, *P*=.05), which was driven by the worst performance using the 48-hour OpenStreetMap-plus-self-report specification. However, none of the PA pairwise comparisons survived FDR correction (P_{FDR}=.08-.99). Thus, no specific pair of aggregation windows showed a robust performance difference.

Variable Importance Across Timescales

Figure 2. Temporal fusion transformer (TFT) mean absolute error (MAE) across time windows (1, 3, 6, 12, 24, and 48 hours) in predicting momentary negative affect (NA, left) and positive affect (PA, right) intensity, for semantic-location features derived from OpenStreetMap versus OpenStreetMap supplemented with self-report in the test set (n=133 individuals; total 532 surveys). FDR: false discovery rate.

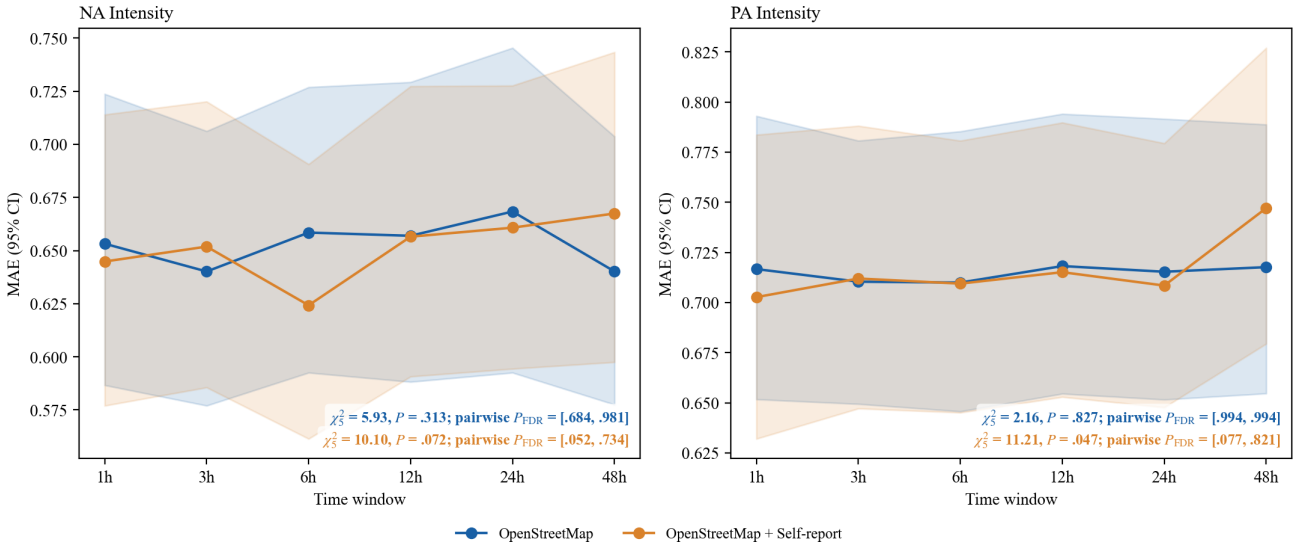
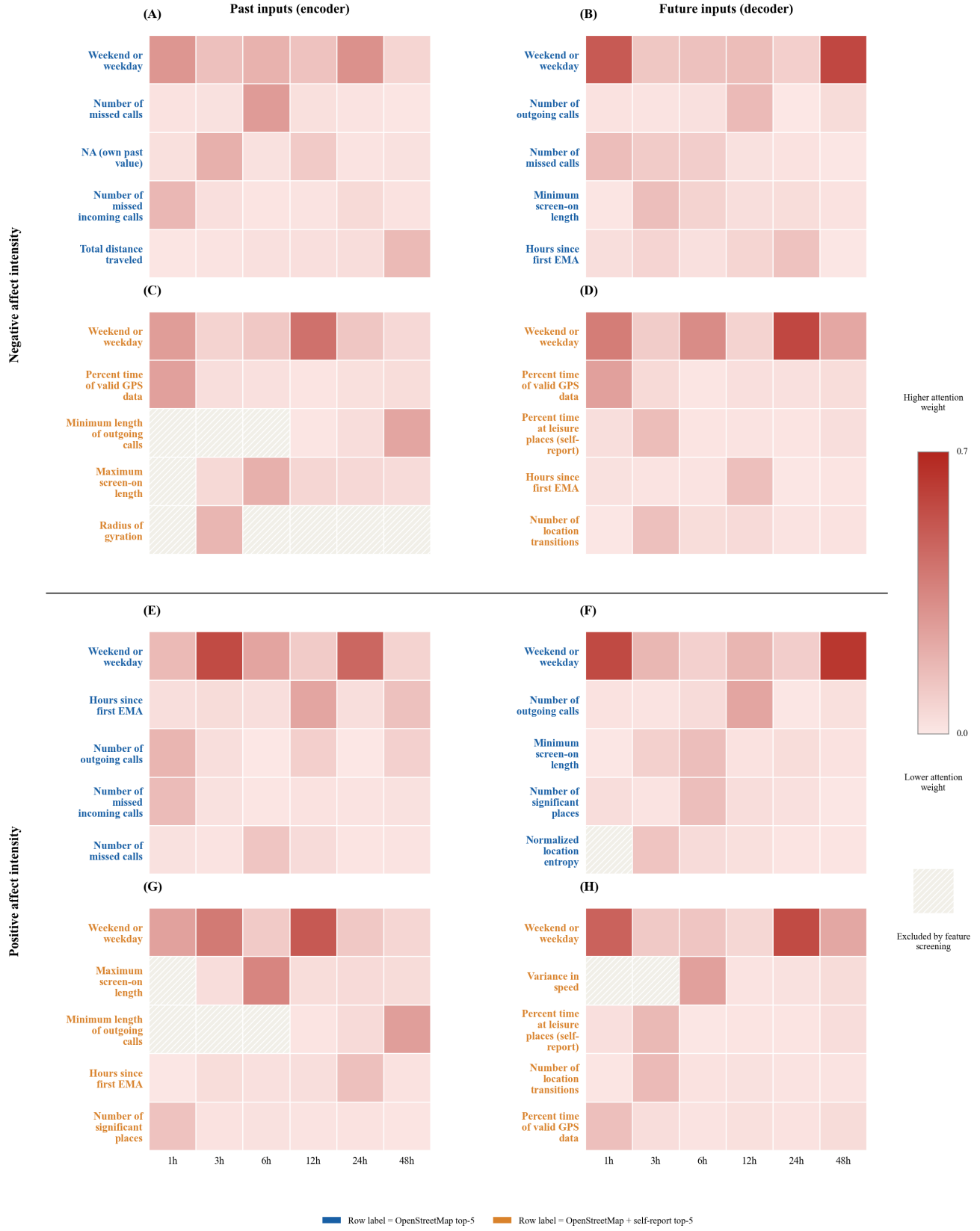


Figure 3 presents a heatmap of important past (left) versus future inputs (right) identified by the TFTs across time windows for predicting NA (upper 2 panels) and PA (lower 2 panels) using location features from OpenStreetMap (blue names) versus from OpenStreetMap supplemented with self-report. We did not find consistent support for hypothesis 2a. Although behavioral indicators of social interactions (ie, calls) and screen interactions appeared among the important future inputs for both NA and PA, this holds only before supplementing OpenStreetMap with self-reported locations to refine location features. After supplementing OpenStreetMap with self-reported locations to refine location features, percent time at leisure places (self-report) was an important future input for both NA and PA, especially in the 3-hour window (NA 0.16, PA 0.17). Partially supporting hypothesis 2b, partial-mobility features contributed as both past and future inputs, with number of location transitions and variance in speed as notable proximal (future) inputs and radius of gyration and number of significant places as distal (past) inputs. Not supporting hypotheses 2a and 2b, there is no clear pattern in variable importance across time windows (eg, consistently darker or lighter from left to right).

XSL•FO
RenderX

Figure 3. Heatmap of important past (left, panels A, C, E, and G) versus future inputs (right, panels B, D, F, and H) identified by temporal fusion transformer (TFT) across time windows for predicting negative affect (NA; upper panels A-D) and positive affect (PA; lower panels E-H) using location features from OpenStreetMap (blue names) versus from OpenStreetMap supplemented with self-report. EMA: ecological momentary assessment.



Outside of hypotheses, the individual's affect history was an important predictor across time windows for both NA and PA (left panels in Figure 3). Across time windows, a time variable (weekend or weekday) is an important future input and, sometimes, a past input across location feature sets for NA and PA, especially at the longer windows (24-hour as a future input and 12-hour as a past input; Figure 3).

XSL•FO
RenderX

Aim 3: The Role of Self-Report Semantic Locations

Overview

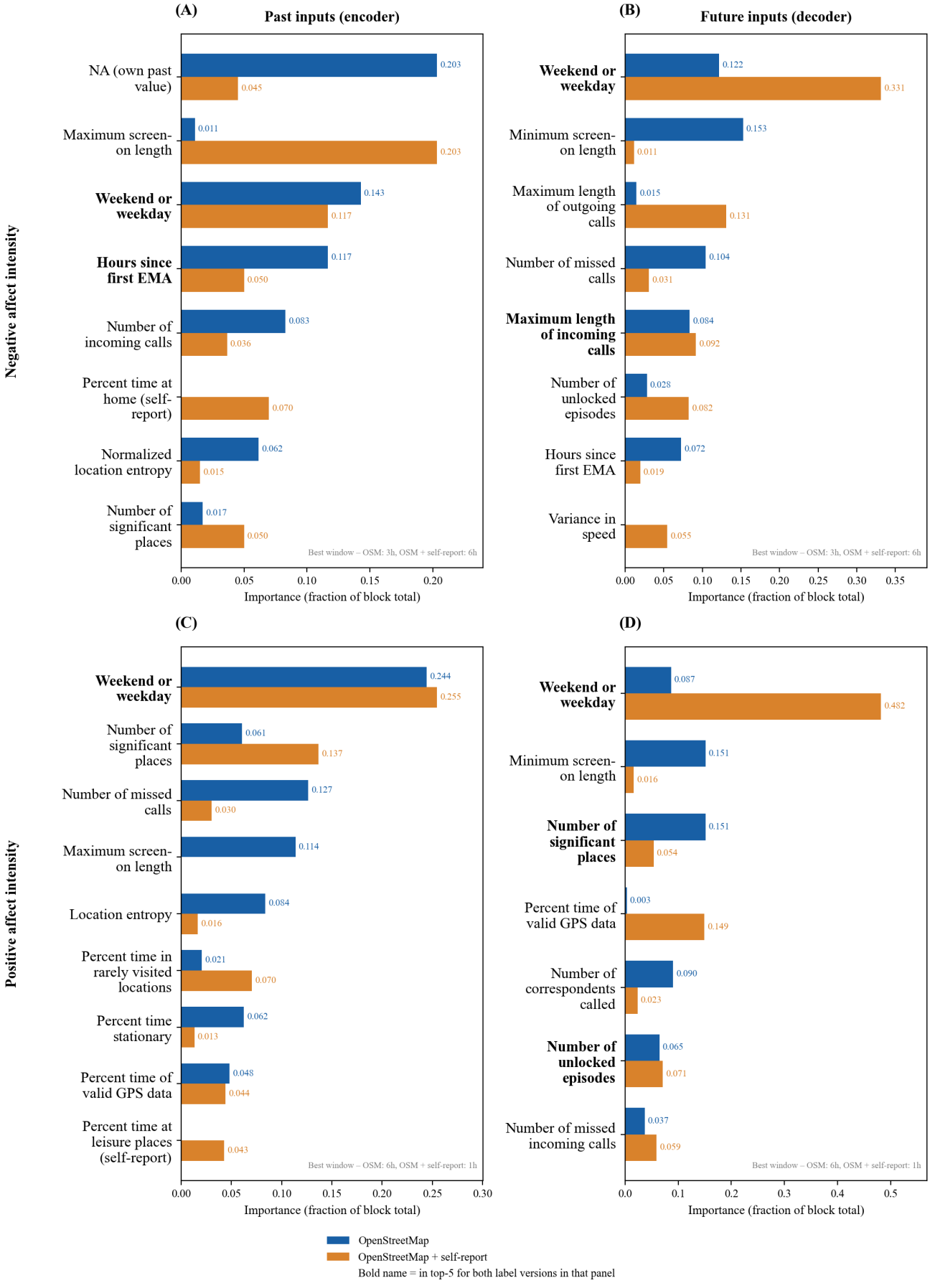
Overall, both bootstrapping and Wilcoxon significance tests suggested that adding self-report semantic-location features did not significantly change prediction accuracy for either NA or PA. Specifically, for predicting NA intensity, the best TFT using semantic-location features derived from OpenStreetMap only, which was based on 3-hour features (MAE=0.64), did not perform significantly differently from the TFT using semantic-location features from OpenStreetMap supplemented with self-report locations, which was based on 6-hour features (MAE=0.62; ΔMAE [95% CI]=0.016 [−0.009 to 0.040]; $P_{bootstrap}$=0.21; $P_{Wilcoxon}$=0.35; Cohen *d* [95% CI]=0.11 [−0.06 to 0.28]). Results were similar for PA, such that the best TFT before incorporating self-report locations, which was based on 6-hour features (MAE=0.71), did not perform significantly differently from the best TFT after incorporating self-report locations, which was based on 1-hour features (MAE=0.70; ΔMAE [95% CI]=0.007 [−0.012 to 0.026]; $P_{bootstrap}$=0.44; $P_{Wilcoxon}$=0.34; Cohen *d* [95% CI]=0.07 [−0.10 to 0.24]).

Variable Importance After Adding Self-Report Semantic Locations

Figure 4 presents the attention weights of past (left) versus future inputs (right) from the best-performing TFT for NA intensity (top) and PA intensity (bottom). Blue bars represent attention weights before adding self-report semantic locations, whereas orange bars represent attention weights after adding self-report semantic locations. Consistent with hypothesis 3b, incorporating self-report semantic locations changed the composition and relative ranking of important inputs, although the pattern of change varied across past and future inputs. The percent time spent at leisure places (self-report) emerged as an important input in the TFT attention weights, particularly at the 3-hour window for both NA and PA (Figure 3; NA: 0.16, PA: 0.17; see detailed attention weights for each variable in Tables S1 and S2 in Multimedia Appendix 1). Several GPS features indicating spatial mobility (variance in speed, number of location transitions, and radius of gyration) were identified among important inputs across models.

XSL•FO
RenderX

Figure 4. Attention weights of top-5 past (left; panels A and C) versus future inputs (right; panels B and D) from the best-performing temporal fusion transformer (TFT) in predicting momentary negative affect (NA, top, panels A and B) and positive affect (PA, bottom, panels C and D) intensity, for semantic-location features derived from OpenStreetMap (blue) versus OpenStreetMap supplemented with self-report (orange) in the test set (n=133 individuals; total 532 surveys). EMA: ecological momentary assessment.



XSL•FO
RenderX

Within-Individual Associations in Linear Mixed-Effects Models

Tables S1 and S2 present standardized coefficients in linear mixed-effects models for NA and PA, respectively (see them in Multimedia Appendix 1). Smartphone-tracked behaviors showed weak, if any, bivariate associations with momentary affect intensity (all $|\beta|<0.06$). After FDR correction across 236 tests per outcome, 5 proximal call features were positively associated with NA: the numbers of incoming calls, outgoing calls, and correspondents in the 1-hour window and the numbers of outgoing calls and correspondents in the 3-hour window ($\beta=0.03$-0.06, $P_{FDR}=.003$-$.05$). Thirteen features were associated with PA after correction. PA was positively associated with location diversity across 3- to 24-hour windows (log location variance, location entropy, normalized location entropy, and number of significant places; $\beta=0.04$-0.06) and negatively associated with time at self-reported workplaces in the 1-, 3-, and 6-hour windows and with the number of correspondents in the 1-hour window ($\beta=-0.04$ to -0.05; $P_{FDR}=.002$-$.03$).

Sensitivity Analyses

Three sensitivity analyses supported the robustness of the primary findings. First, lowering the minimum number of valid EMA reports from 40 to 20 retained 167 participants (n=8398 NA EMAs and n=8391 PA EMAs), compared with 133 participants in the primary sample (n=7337 NA EMAs and n=7332 PA EMAs). Predictive performance did not differ significantly for either NA (6-hour, OSM + self-report; ΔMAE=0.04, 95% CI −0.10 to 0.20; $P_{bootstrap}=.624$; $P_{Wilcoxon}=.54$) or PA (1-hour, OSM + self-report; ΔMAE=0.12, 95% CI −0.07 to 0.36; $P_{bootstrap}=.26$; $P_{Wilcoxon}=.32$). Second, replacing MICE with median imputation did not significantly change performance for NA (6-hour, OSM + self-report; ΔMAE=0.010, 95% CI −0.001 to 0.021; $P_{bootstrap}=.07$; $P_{Wilcoxon}=.21$) or PA (1-hour, OSM + self-report; ΔMAE=0.0005, 95% CI −0.0046 to 0.0052; $P_{bootstrap}=.84$; $P_{Wilcoxon}=.70$). Finally, increasing the maximum encoder length from 16 (minimum 8) to 20 (minimum 12), while retaining a maximum decoder length of 4, did not significantly affect performance for NA (6-hour, OSM + self-report; ΔMAE=0.0008, 95% CI −0.0089 to 0.0106; $P_{bootstrap}=.89$, $P_{Wilcoxon}=.86$) or PA (1-hour, OSM + self-report; ΔMAE=0.0035, 95% CI −0.0080 to 0.0137; $P_{bootstrap}=.53$; $P_{Wilcoxon}=.21$).

Discussion

Principal Findings

This study parsed the contributions of EMA and smartphone sensing data to personalized predictions of momentary affect. On held-out EMA observations, smartphone sensing within the full-information TFT added statistically significant and modest predictive value for NA but small, nonsignificant value for PA relative to the person-mean baseline. Person-specific mean affect accounted for a substantial share of predictive performance. Although one PA omnibus test differed across aggregation windows, no pairwise time-window comparison survived FDR correction. Adding self-reported semantic locations did not significantly improve NA or PA prediction. Before these locations were added, mobility markers and smartphone social interactions appeared among important past and future inputs. After they were added, mobility markers appeared more often as past inputs, partially supporting hypothesis 2b, whereas screen and smartphone social interactions showed no consistent proximal or distal pattern, providing no support for hypothesis 2a. Overall, our results suggest modest added value of smartphone sensing over EMA for NA prediction following person-specific calibration.

Aim 1: Model Performance in Predicting Momentary NA and PA

Aim 1 tested how much each data source contributed to personalized prediction and whether the TFT improved on a person-mean baseline. Partially consistent with hypothesis 1, smartphone sensing added significant predictive value for NA (but not PA), empowered by transformer-based models. All prior studies made personalized predictions through incorporating all data sources, including both active and passive inputs [11-13]. Ultimately, the primary promise of digital phenotyping is the possibility of generating accurate predictions of momentary affect intensity using passive inputs alone. This is a goal that has attracted increasing interest given its broad clinical and practical implications. The present findings can be situated within a broader affect-dynamics framework in which momentary affect reflects a combination of relatively stable person-specific tendencies, recent affective trajectories, and immediate contextual inputs. Findings showed that participants' mean affect and affective history accounted for a substantial proportion of predictive performance, whereas smartphone sensing provided only modest incremental information. This pattern indicates that smartphone sensing should not yet be viewed as a standalone replacement for active self-report, but rather as a complementary source of information to EMA. Future personalized prediction research should consistently evaluate the person-mean affect intensity as an essential benchmark and explore richer multimodal passive signals that capture dimensions of emotional experience beyond smartphone-tracked behaviors (eg, physiological measures and linguistic features).

Another motivation of the study was to address the relative neglect of PA in the affective computing literature, particularly given its established importance for psychological well-being and depression treatment. Our findings show that the incremental value of smartphone sensing was significant and modest for NA, but small and nonsignificant for PA. Consequently, high predictive performance for NA (or depressed mood, which is more commonly investigated) should not be assumed to automatically generalize to PA. Current sensing modalities may be differentially informative across valence dimensions. Improving PA prediction may require larger sample sizes or alternative digital markers. Future affective computing research focused on PA is clearly warranted.

Aims 2 and 3: The Role of Timescales and Self-Reported Semantic Locations

Aims 2 and 3 examined aggregation timescales and self-reported semantic locations. Across 6 windows, 3 omnibus tests were

XSL•FO
RenderX

nonsignificant and 1 PA test reached statistical significance, but no pairwise comparison survived FDR correction. Aggregation-window choices may therefore be guided partly by practical considerations, such as computational efficiency, pending replication. Adding self-reported semantic locations did not improve predictive performance, so the value of collecting these data appears to lie more in contextual interpretation than in accuracy.

Three broader patterns emerged from the attention-weight analyses. First, incorporating self-reported semantic locations changed the composition and relative ranking of important inputs despite not improving predictive accuracy, suggesting that contextual information may contribute more to interpretation than to predictive accuracy. These changes did not follow a clear pattern across NA and PA or across past and future inputs. Time at self-reported leisure places appeared as an important input after semantic-location features were added. Second, the TFT prioritized different classes of smartphone-tracked behavior across aggregation windows while maintaining broadly comparable predictions. Third, spatial-mobility features appeared more consistently as past than future inputs, suggesting that they may summarize accumulated behavioral context rather than immediate affective triggers. However, these attention weights indicate model-based predictive relevance, not causal effects, and should be treated more as hypothesis-generating.

Implications for Research and Applications

The affective computing approach to personalized monitoring of affect intensity is still in its infancy, and this study addressed three key methodological questions: (1) how much variance did smartphone sensing add to personalized prediction of momentary affect beyond EMA baselines, (2) does the data aggregation timescale impact predictive performance, and (3) what is the value of integrating self-report semantic locations. First, our framework provides a more rigorous benchmark for evaluating passive sensing for affect monitoring. Future work on personalized predictions should benchmark models against person-specific mean affect, demonstrating incremental predictive value rather than relying solely on absolute accuracy metrics. Second, because predictive performance remained stable across the 6 evaluated timescales, feature engineering decisions regarding aggregation windows can be guided by practical considerations, such as computational efficiency. Third, incorporating contextual self-reports may enhance the interpretability of digital behavioral markers, even when they do not directly improve prediction accuracy. Finally, rather than replacing EMA, passive sensing may reduce the frequency of active assessment while maintaining continuous, individualized monitoring between self-reports.

Limitations and Future Directions

This study has several limitations. First, within-person reliability was modest for NA and PA (ω=0.64 and 0.69), which constrains achievable predictive accuracy. Future studies should balance more reliable measurement of momentary affect against EMA burden. Second, the feature-importance rankings produced by the TFT reflect associational rather than causal relationships. A predictor may receive high attention weight simply because it is correlated with an unobserved driver or acts as a proxy that reduces prediction error. Consequently, these findings should be interpreted as hypotheses regarding potential mechanisms rather than definitive evidence of causal influence. Experimental manipulation or microrandomized intervention studies are necessary to establish causal pathways.

In addition, our models were good at predicting momentary affect intensity at time points immediately following the training period (about 2 weeks) for the same individuals. However, it is unclear whether our models can make good predictions days, weeks, or even months after the training data time points. Future work could examine whether algorithms can stop assessing self-report experiences after a certain period and then allow ubiquitous and continuous monitoring. If so, it would also be important to clarify the maximum length of the no-report period.

The present findings illustrate AI's potential to predict momentary affect in daily life (situational generalizability) by forecasting an individual's future states from their own past data (within-person temporal generalizability). A critical remaining question concerns population generalizability and whether such models can generalize to entirely new individuals. We maintain that, at least for predicting momentary affect intensity, developing personalized models is more feasible than building universal models expected to generalize across individuals. In human-to-human interactions, people can readily recognize distinct emotions in a stranger, but accurately gauging another person's emotional intensity typically requires familiarity. Consistent with this distinction, research evaluating whether passive sensor data could classify discrete momentary affect states (eg, angry or not angry) across new individuals found that traditional ML models performed only slightly above chance [8]. Although population-level prediction of affect intensity without a priori calibration remains challenging, achieving it would carry substantial public health implications.

Third, the present work is limited by the number of modalities in data sources. Additional smartphone sensors collecting other social interaction data (eg, text messages or social media) and audio information (eg, volumes and frequency of meaningful conversations around an individual) could add modalities while maintaining the benefits from the cheap and wide access of smartphones. Future affect computing research could also benefit from a larger dataset with more people and more observations, as other AI fields (eg, large language models) have illustrated that "a 'dumb' algorithm with lots and lots of data beats a 'clever' one with modest amounts of data" [42].

Conclusion

This study highlights the incremental value of smartphone sensing for personalized, short-term prediction of momentary affect beyond simple person-specific affective baselines. Notably, this added predictive value was significant only for NA, while participants' mean baseline affect accounted for a substantial proportion of overall predictive performance across all models. No pairwise performance difference emerged across the 6 passive-data aggregation windows. Furthermore, incorporating self-reported semantic locations did not improve prediction accuracy, though it provided more interpretable contextual features for generating hypotheses about affect-behavior relations. Overall, these findings suggest that

XSL•FO
RenderX

smartphone sensing is currently better positioned to augment rather than replace EMA. Following an initial period of person-specific calibration, passive sensing may estimate short-term affective fluctuations between self-report prompts and support lower-burden or adaptive EMA designs. Replication in larger, more diverse, and clinical samples remains essential before deploying these models in real-world assessment or intervention use.

Acknowledgments

The authors would like to thank Dr Ellen E Fitzsimmons-Craft, Dr Joshua R Oltmanns, Dr Nathaniel S Eckland, and Dr Jocelyn Lai for their review and feedback on the manuscript. The authors thank Anna Leah Davis and Analise Black for their assistance with data collection.

Generative AI tools were used to assist with code development, debugging, and refinement, including code for data preprocessing and implementation of the temporal fusion transformer and comparison models. They were also used to support language editing and improve the clarity and organization of selected portions of the manuscript. The tools used included ChatGPT-4o and GPT-5.6 (OpenAI), Claude 3.5 Sonnet and Fable (Anthropic), Gemini 1.5 Pro (Google), and Microsoft Copilot (Microsoft Corporation). All AI-assisted code and text were critically reviewed and revised by the authors. The authors retained full responsibility for the study design, analytic decisions, accuracy of the code, interpretation of the results, and final manuscript.

Funding

Funding was provided by the Office of the Provost at Washington University in St. Louis (total amount of US $50,000). The funder had no involvement in the study design, data collection, analysis, interpretation, or the manuscript preparation.

Data Availability

The datasets generated or analyzed during this study are not publicly available because they contain sensitive smartphone-sensing and location information, and because public data sharing was not covered by participants' consent. However, they may be available from the corresponding author on reasonable request, subject to institutional approval, and an appropriate data-use agreement. The analysis code is available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: YZ, RJT
Data curation: YZ, YY, RJT
Formal analysis: YZ, YY
Funding acquisition: RJT
Investigation: RJT
Methodology: YZ, YY
Project administration: RJT
Resources: RJT
Software: YY
Supervision: RJT
Validation: YY
Visualization: YZ
Writing – original draft: YZ
Writing – review & editing: YZ, YY, RJT

Conflicts of Interest

None declared.

Multimedia Appendix 1
Affect distributions, important predictors across temporal and semantic-location feature sets, and prediction performance for discrete affect states.
[DOCX File, 227 KB - mhealth_v14i1e90970_app1.docx]

References

1. Strauss GP, Allen DN. The experience of positive emotion is associated with the automatic processing of positive emotional words. J Posit Psychol 2006 Jul;1(3):150-159. [doi: 10.1080/17439760600566016]
2. Threadgill AH, Gable PA. Negative affect varying in motivational intensity influences scope of memory. Cogn Emot 2019 Mar;33(2):332-345. [doi: 10.1080/02699931.2018.1451306] [Medline: 29621935]

XSL•FO
RenderX

3. Lerner JS, Li Y, Valdesolo P, Kassam KS. Emotion and decision making. Annu Rev Psychol 2015 Jan 3;66(1):799-823. [doi: 10.1146/annurev-psych-010213-115043] [Medline: 25251484]
4. Diagnostic and Statistical Manual of Mental Disorders: DSM-5: American Psychiatric Association; 2013. [doi: 10.1176/appi.books.9780890425596]
5. Gotlib IH, Joormann J. Cognition and depression: current status and future directions. Annu Rev Clin Psychol 2010;6(1):285-312. [doi: 10.1146/annurev.clinpsy.121208.131305] [Medline: 20192795]
6. Williams JMG, Barnhofer T, Crane C, et al. Autobiographical memory specificity and emotional disorder. Psychol Bull 2007 Jan;133(1):122-148. [doi: 10.1037/0033-2909.133.1.122] [Medline: 17201573]
7. Drexl K, Ralisa V, Rosselet-Amoussou J, et al. Readdressing the ongoing challenge of missing data in youth ecological momentary assessment studies: meta-analysis update. J Med Internet Res 2025 Apr 30;27:e65710. [doi: 10.2196/65710] [Medline: 40305088]
8. Akre-Bhide S, Cohen ZD, Welborn A, Zbozinek TD, Craske MG, Bui A. Detecting momentary reward and affect with real-time passive digital sensor data. JAMIA Open 2026 Feb;9(1):ooag005. [doi: 10.1093/jamiaopen/ooag005] [Medline: 41669161]
9. Jafarlou S, Lai J, Azimi I, et al. Objective prediction of next-day's affect using multimodal physiological and behavioral data: algorithm development and validation study. JMIR Form Res 2023 Mar 15;7(1):e39425. [doi: 10.2196/39425] [Medline: 36920456]
10. Ren B, Balkind EG, Pastro B, et al. Predicting states of elevated negative affect in adolescents from smartphone sensors: a novel personalized machine learning approach. Psychol Med 2023 Aug;53(11):5146-5154. [doi: 10.1017/S0033291722002161] [Medline: 35894246]
11. Mobile fact sheet. Pew Research Center. 2025 Nov 20. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2026-09-04]
12. Jacobson NC, Chung YJ. Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones. Sensors (Basel) 2020 Jun 24;20(12):3572. [doi: 10.3390/s20123572] [Medline: 32599801]
13. Shah RV, Grennan G, Zafar-Khan M, et al. Personalized machine learning of depressed mood using wearables. Transl Psychiatry 2021 Jun 9;11(1):338. [doi: 10.1038/s41398-021-01445-0] [Medline: 34103481]
14. Craske MG, Dunn BD, Meuret AE, Rizvi SJ, Taylor CT. Positive affect and reward processing in the treatment of depression, anxiety and trauma. Nat Rev Psychol 2024;3(10):665-685. [doi: 10.1038/s44159-024-00355-4]
15. Craske MG, Meuret AE, Ritz T, Treanor M, Dour H, Rosenfield D. Positive affect treatment for depression and anxiety: a randomized clinical trial for a core feature of anhedonia. J Consult Clin Psychol 2019 May;87(5):457-471. [doi: 10.1037/ccp0000396] [Medline: 30998048]
16. Meuret AE, Rosenfield D, Wang E, Hough CM, Ritz T, Craske MG. Positive affect treatment for depression, anxiety, and low positive affect: a randomized clinical trial. JAMA Netw Open 2026 Apr 1;9(4):e267403. [doi: 10.1001/jamanetworkopen.2026.7403] [Medline: 42030048]
17. Watson D, Tellegen A. Toward a consensual structure of mood. Psychol Bull 1985;98(2):219-235. [doi: 10.1037/0033-2909.98.2.219]
18. Costa PT, McCrae RR. Influence of extraversion and neuroticism on subjective well-being: happy and unhappy people. J Pers Soc Psychol 1980;38(4):668-678. [doi: 10.1037/0022-3514.38.4.668]
19. Davidson RJ. What does the prefrontal cortex "do" in affect: perspectives on frontal EEG asymmetry research. Biol Psychol 2004 Oct;67(1-2):219-233. [doi: 10.1016/j.biopsycho.2004.03.008] [Medline: 15130532]
20. Newsom JT, Rook KS, Nishishiba M, Sorkin DH, Mahan TL. Understanding the relative importance of positive and negative social exchanges: examining specific domains and appraisals. J Gerontol B Psychol Sci Soc Sci 2005 Nov;60(6):304-P312. [doi: 10.1093/geronb/60.6.p304] [Medline: 16260704]
21. Lim B, Arık S, Loeff N, Pfister T. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. Int J Forecast 2021 Oct;37(4):1748-1764. [doi: 10.1016/j.ijforecast.2021.03.012]
22. Larsen RJ, Diener E. Affect intensity as an individual difference characteristic: a review. J Res Pers 1987 Mar;21(1):1-39. [doi: 10.1016/0092-6566(87)90023-7]
23. Langener AM, Stulp G, Jacobson NC. It's all about timing: exploring different temporal resolutions for analyzing digital-phenotyping data. Adv Methods Pract Psychol Sci 2024 Jan;7(1). [doi: 10.1177/25152459231202677]
24. Dickerson SS, Kemeny ME. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychol Bull 2004 May;130(3):355-391. [doi: 10.1037/0033-2909.130.3.355] [Medline: 15122924]
25. Moors A, Ellsworth PC, Scherer KR, Frijda NH. Appraisal theories of emotion: state of the art and future development. Emotion Review 2013 Apr;5(2):119-124. [doi: 10.1177/1754073912468165]
26. Russell JA, Barrett LF. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. J Pers Soc Psychol 1999;76(5):805-819. [doi: 10.1037/0022-3514.76.5.805]
27. Rauthmann JF, Gallardo-Pujol D, Guillaume EM, et al. The Situational Eight DIAMONDS: a taxonomy of major dimensions of situation characteristics. J Pers Soc Psychol 2014 Oct;107(4):677-718. [doi: 10.1037/a0037250] [Medline: 25133715]

XSL•FO
RenderX

28. Chikersal P, Venkatesh S, Masown K, et al. Predicting multiple sclerosis outcomes during the COVID-19 stay-at-home period: observational study using passively sensed behaviors and digital phenotyping. JMIR Ment Health 2022 Aug 24;9(8):e38495. [doi: 10.2196/38495] [Medline: 35849686]
29. Doryab A, Villalba DK, Chikersal P, et al. Identifying behavioral phenotypes of loneliness and social isolation with passive sensing: statistical analysis, data mining and machine learning of smartphone and Fitbit data. JMIR Mhealth Uhealth 2019 Jul 24;7(7):e13209. [doi: 10.2196/13209] [Medline: 31342903]
30. Gross AM, Lai J, Eckland NS, Thompson RJ. Interoceptive awareness and clarity of one's emotions and goals: a naturalistic investigation. Emotion 2025 Sep;25(6):1516-1530. [doi: 10.1037/emo0001510] [Medline: 39977691]
31. Lai J, Eckland NS, Thompson RJ. When and why people do NOT regulate their emotions: examining the reasons and contexts. Cogn Emot 2026 Mar;40(2):301-315. [doi: 10.1080/02699931.2025.2504560] [Medline: 40409276]
32. O'Brien ST, Dozo N, Hinton JDX, et al. SEMA3: a free smartphone platform for daily life surveys. Behav Res 2024;56(7):7691-7706. [doi: 10.3758/s13428-024-02445-w]
33. Ferreira D, Kostakos V, Dey AK. AWARE: Mobile context instrumentation framework. Front ICT 2015;2:6. [doi: 10.3389/fict.2015.00006]
34. Canzian L, Musolesi M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. Presented at: UbiComp '15: The 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing Association for Computing Machinery; Sep 7-11, 2015; Osaka, Japan p. 1293-1304 URL: https://dl.acm.org/doi/proceedings/10.1145/2750858 [accessed 2026-09-21] [doi: 10.1145/2750858.2805845]
35. Saeb S, Zhang M, Kwasny M, Karr CJ, Kording K, Mohr DC. The relationship between clinical, momentary, and sensor-based assessment of depression. Presented at: 9th International Conference on Pervasive Computing Technologies for Healthcare; May 20-23, 2015; Istanbul, Turkey p. 229-232 URL: http://eudl.eu/proceedings/PervasiveHealth/2015 [accessed 2026-09-04] [doi: 10.4108/icst.pervasivehealth.2015.259034]
36. Ringwald WR. Refining behavioral phenotypes for binge drinking from broad liabilities to proximal predictors with multiple raters of personality and smartphone sensor data [Dissertation]. : University of Pittsburgh; 2024 URL: https://d-scholarship.pitt.edu/45398 [accessed 2025-05-08]
37. Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. Presented at: KDD'96: Second International Conference on Knowledge Discovery and Data Mining; Aug 2-4, 1996; Portland, Oregon, USA p. 226-231 URL: https://aaai.org/papers/kdd96-037-a-density-based-algorithm-for-discovering-clusters-in-large-spatial-databases-with-noise/ [accessed 2026-09-20]
38. Clemens K. Geocoding with OpenStreetMap data. Presented at: GEOProcessing; Feb 22-27, 2015 URL: https://www.thinkmind.org/download_full.php?instance=GEOProcessing+2015 [accessed 2026-09-20]
39. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
40. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Series B Stat Methodol 2005 Apr 1;67(2):301-320. [doi: 10.1111/j.1467-9868.2005.00503.x]
41. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Presented at: KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Association for Computing Machinery; Aug 13-17, 2016; San Francisco, CA, USA p. 785-794 URL: https://dl.acm.org/doi/proceedings/10.1145/2939672 [accessed 2026-09-20] [doi: 10.1145/2939672.2939785]
42. Domingos P. A few useful things to know about machine learning. Commun ACM 2012 Oct;55(10):78-87. [doi: 10.1145/2347736.2347755]

Abbreviations

DBSCAN: Density-Based Spatial Clustering of Applications with Noise
ElasticNet: elastic net regression
EMA: ecological momentary assessment
FDR: false discovery rate
MAE: mean absolute error
MICE: Multivariate Imputation by Chained Equations
ML: machine learning
N-BEATS: neural basis expansion analysis for interpretable time series forecasting
NA: negative affect
PA: positive affect
REML: restricted maximum likelihood
RMSE: root-mean-square error
TFT: temporal fusion transformer
XGBoost: extreme gradient boosting

XSL•FO
RenderX

Edited by L Buis; submitted 07.Jan.2026; peer-reviewed by D Trabassi, D Shin; revised version received 14.Aug.2026; accepted 18.Aug.2026; published 28.Sep.2026.

Please cite as:
Zhu Y, Yang Y, Thompson RJ
Incremental Value of Smartphone Sensing for Monitoring Momentary Affect Intensity in Adults Using Transformer-Based Models: Observational Study
JMIR Mhealth Uhealth 2026;14:e90970
URL: https://mhealth.jmir.org/2026/1/e90970
doi:10.2196/90970

© Yiqin Zhu, Yuyi Yang, Renee J Thompson. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 28.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Predicting Affective Episodes in Bipolar Disorder Using Statistical Process Control Analysis of GPS-Based Mobility Patterns: Quantitative Study

Marvin Guth[1], MSc; Carl Bittendorf[2], MSc; Clemens Krug[3], MSc; Vera Miriam Ludwig[3], Dr Med; Esther Muehlbauer[3], Dr Rer Med; Lisa-Marie Hartnagel[2], PhD; Emanuel Severus[3,4], Prof Dr Med; Michael Bauer[3], Dr rer nat, Prof Dr Med; Philipp Ritter[3], Dr med, PD; Ulrich W Ebner-Priemer[2,5], Prof Dr

[1]Department of eHealth and Sports Analytics, Faculty of Sport Science, Ruhr University Bochum, Bochum, Germany

[2]Mental mHealth Lab, Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Hertzstrasse 16, Karlsruhe, Baden-Württemberg, Germany

[3]Department of Psychiatry and Psychotherapy, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

[4]Asklepios Klinik Nord-Ochsenzoll, Hamburg, Germany

[5]Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Heidelberg University, Heidelberg, Germany

Corresponding Author:
Ulrich W Ebner-Priemer, Prof Dr
Mental mHealth Lab, Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Hertzstrasse 16, Karlsruhe, Baden-Württemberg, Germany

Abstract

Background: Bipolar disorders (BDs) represent a significant global health challenge, with frequent and severe affective episodes that impair quality of life. Accurate, early prediction of these episodes remains difficult. Recent advances in mobile sensing offer new possibilities to detect prodromal changes via smart digital phenotypes, such as geolocation data.

Objective: This study aimed to examine whether spatial exploratory behavior, assessed via passive GPS data, can predict depressive and manic episodes in individuals with BD. Specifically, we evaluated the predictive value of unique places visited and related mobility metrics using statistical process control (SPC) techniques to identify both early deviations indicative of prodromal states and changes occurring during ongoing affective episodes.

Methods: Using high-resolution GPS data from the BipoSense dataset, we applied Density-Based Spatial Clustering of Applications with Noise to extract behavioral mobility indicators: number of unique places visited, frequency of location changes, and time spent per location. We implemented exponentially weighted moving average (EWMA)–based SPC to identify "out-of-bounds" deviations from individual baselines. We then tested the alignment of these deviations with affective episodes and the prodromal periods. Optimization of SPC parameters (λ and control limit L) was performed to enhance predictive accuracy.

Results: The analysis included 28 participants with BD and a total of 10,213 observation days, covering 26 depressive and 20 (hypo)manic episodes. Examining whether control limits distinguish affective episodes from euthymic days via multilevel models revealed that median time spent at clusters indicated both depressive and (hypo)manic episodes the best, whereas the number of unique clusters showed no significant association with phase transitions. While EWMA-SPC detected behavioral deviations during affective episodes, no single variable consistently met predefined thresholds for both sensitivity and specificity. Optimized SPC settings improved performance, but the number of unique places alone did not robustly predict prodromal or acute episodes. No statistically significant predictive accuracy (eg, sensitivity >70% and specificity >70%) was achieved for any individual indicator (P>.05). However, some SPC charts suggested within-person temporal deviations preceding episodes, indicating limited yet potentially informative patterns.

Conclusions: Although unique places visited alone may not suffice as a predictive marker, the application of EWMA-based SPC to GPS data holds promise for the development of smart digital phenotypes. Although our analysis to predict upcoming episodes did not yield robust predictive accuracy in its current form, it provides a promising conceptual framework for individualized, low-burden behavioral monitoring. Further research is needed to refine existing digital biomarkers, develop new ones, and validate their clinical utility in reducing the frequency and severity of illness phases.

(JMIR Mhealth Uhealth 2026;14:e77272) doi:10.2196/77272

KEYWORDS

bipolar disorder; mobile sensing; spatial data; spatial analysis; statistical process control; digital phenotyping; unique places

XSL•FO
RenderX

Introduction

Smart Digital Phenotyping in Bipolar Disorder

Bipolar disorders (BDs) affect up to 40 million people worldwide and represent a significant global burden of disease [1], imposing considerable personal and societal challenges [2]. Characterized by recurring episodes of depression and mania, the aim of BD disease management is to reduce the frequency, duration, or intensity of affective episodes, thereby extending the time spent in remission (euthymia) [3]. One promising strategy for maintaining euthymia is the early detection of emerging episodes, which facilitates timely interventions that may prevent the development of full-blown episodes. This can be achieved through intensive longitudinal study designs, such as digital phenotyping [4] and ambulatory assessment [5], as they enable continuous passive monitoring of symptom-related parameters over extended periods, thereby facilitating real-time feedback and automated forecasting.

Despite their potential, digital phenotyping approaches have yet to demonstrate a meaningful impact on disease trajectories in BD [6-10]. To address this, the need for clinically relevant smartphone-derived indices, so-called "smart digital phenotypes," has been emphasized [11]. Unlike conventional smartphone-based measures, such as the number of incoming calls or steps per minute, smart digital phenotyping aims to extract clinically meaningful features that more closely reflect psychopathology. For instance, rather than measuring the number of incoming calls, analyzing speech patterns (eg, number of words spoken per minute [11]) may better capture the *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition; *DSM-5*) criteria, such as talkativeness or pressure to keep talking [12]. Similarly, analyzing mobility patterns, such as visits to novel or unusual locations, may provide a more nuanced understanding of behavioral changes, compared to simply calculating daily kilometers traveled.

In particular, enhancing the analysis of location-based passive sensing data holds significant yet underused potential [13]. Given the rich behavioral insights embedded in raw GPS data, sophisticated geospatial analyses could substantially improve passive sensing approaches for BD, offering a novel and promising avenue for research.

Previous Work

A recent systematic review of geospatial analyses in depression has shown multiple correlations across several key mobility metrics, such as entropy or distance. Notably, among 31 studies, 16 explored the relationship between depression and individual location-based data, 12 focused on predicting depression states, and 3 integrated location data with additional sensing modalities to assess depression [14]. The authors of the systematic review state that location data reflect changes in individual mood states and consistently exhibit a strong correlation with depression. Additionally, the reviewed studies demonstrate the predictive power and utility of location data in forecasting depressive symptoms.

However, the applicability of these study designs to BD is far from trivial. First and foremost, studies must not only ensure a sufficient number of participants to guarantee statistical power but also cover a sufficiently long observation period. Shorter study durations may fail to encapsulate critical episodes, thereby undermining the true potential of digital phenotyping [4]. Second, studies should involve individuals with a verified BD diagnosis, ideally reviewed by a clinician to ensure accurate ground truth, which is crucial for precise episode onset identification in analyses. Thus, precise identification of episode onset at the day level is critical. Even a deviation of only 3 days in a retrospective assessment spanning 3 months can blur the distinction between true prediction and the mere confirmation of an imprecisely recorded episode onset, especially when forecasts only anticipate episodes by a few days and lack a high-frequency, valid ground truth measurement [15]. Thus, several studies on digital phenotyping for BD episode prediction have yielded inconclusive results due to these challenges, as well as heterogeneous protocols and variable parameters [7].

Considering this, there are 2 studies worth mentioning that analyze spatial data to detect episodes in BD. Palmius et al [16] analyzed geographic data from 22 participants with BD and 14 healthy controls (HC), detecting depressive symptoms as measured via QIDS-SR16 (Quick Inventory of Depressive Symptomatology–Self-Report, 16 items) with 85% accuracy using location recordings alone. However, location variance was operationalized as "entropy" by clustering stationary locations via k-means, adjusting k until clusters were 400 m apart. This differs from classical spatial entropy in geography [17] and may limit the clinical relevance of smart digital phenotyping. A more clinically meaningful approach should distinguish between unique places visited, frequency of location changes, and time spent at each cluster. Additionally, the study's short 3-month duration, limited symptom variance, and reliance on self-reported symptoms without expert classification pose significant limitations. Another relevant study examined differences in passively collected smartphone-based location data between 48 patients with BD and 31 HC [18], as well as investigating location data across different affective states for up to 9 months. Lower mobility was observed during depressive states compared to euthymic periods in patients with BD. Furthermore, patients with BD demonstrated lower location variance during affective states compared to euthymic periods. Limitations of this study, similar to those reported earlier, primarily concern the conceptualization of entropy, which poses challenges in constructing clinically relevant indices, as well as the study duration, which, although extended to up to 9 months, remains relatively short.

In sum, empirical findings on BD episode prediction from geospatial parameters remain inconclusive. This may be largely due to the methodological limitations mentioned earlier and the suboptimal operationalization of entropy. Most importantly, however, both studies adopt a dimensional approach to BD symptomatology, rather than distinguishing distinct affective episodes. This is evident in their reliance on self-reported symptom fluctuations, rather than focusing on the prediction of actual emerging episodes. However, this distinction is crucial because accurately predicting new affective episodes is essential for mitigating patient suffering and improving clinical care. Addressing these gaps requires well-designed longitudinal

XSL•FO
RenderX

studies that capture a sufficient number of emerging episodes and accurately detect their onset [8].

Spatial Movement Patterns in BD

Investigating the predictive properties of nuanced location-based passive sensing parameters in a sufficiently powered, temporally exact dataset to assess their clinical value in patients with BD holds significant promise. Changes in activity and energy are cardinal symptoms of BD: mania is, among other things, characterized by increased goal-directed activity and psychomotor agitation, whereas depressive episodes show reduced drive, diminished activity, and low energy (*DSM-5* [12]). These symptoms manifest in observable behavioral changes and translate to spatial movement patterns or exploratory behavior.

Empirical evidence supports this link. Animal models reveal altered exploratory behavior in mania-like [19] and depression-like [20] states. Reverse-translational models using an adapted version of the open field test (a standard rodent paradigm that assesses locomotor activity, anxiety-related behavior, and exploration) indicate that patients with BD display increased motor activity, more unpredictable exploration patterns, and more object interactions compared to HCs or patients with schizophrenia [21-23]. Notably, heightened activity and exploratory behavior persist even during euthymic phases [21,24,25]. Furthermore, it is generally assumed that individual movement patterns tend to be highly regular and predictable over time [26], as evidenced, for instance, by rather stable numbers of visited locations [27]. This raises an important question: Do patients with BD exhibit generally altered yet stable movement patterns as a trait, or do these patterns fluctuate according to symptomatology and affective state (eg, depressed, manic, and euthymic)?

To assess their clinical utility to predict and detect affective episodes, temporal dynamics of movement patterns in BD must be examined in detail. Current metrics, such as entropy used in the studies mentioned earlier, show routine but often fail to capture the full complexity of spatial behavior. It is crucial to differentiate between the time spent at a given location and the number of unique locations visited. For instance, very simplified, a regular daily commute might reflect euthymia, while frequent visits to different bars could indicate mania, whereas predominantly staying at home might signal depression. Understanding the timing and nature of these changes is key for early detection of emerging episodes.

These movement patterns can be effectively analyzed in spatial science cluster analysis tailored for longitudinal GPS data, combined with Geographic Information System–based mapping approaches, can visualize movement paths, hotspots, and clusters, and refine statistical parameters for greater accuracy. This simplifies complex data, uncovers hidden geographical patterns, and helps detect outliers, while also accommodating environmental factors such as geography, infrastructure, climate, and social context. Ultimately, advanced geospatial analysis of GPS data could be a valuable tool in developing smart digital phenotypes for BD.

Statistical Process Control

Predicting emerging episodes relies not only on valid smartphone-derived indices but also on a prediction model. The exponentially weighted moving average (EWMA)–based statistical process control (SPC) method may offer a valuable approach in this context. SPC was originally developed to monitor industrial production processes and detect when the quality or performance of a product deviates significantly from expected standards by giving “out-of-bounds” indications [28]. This concept can also be applied to psychopathology, where euthymia is defined as a stable, functioning phase, and affective episodes are seen as processes that are “out of bounds” [29].

For instance, Schreuder et al [30,31] applied EWMA-SPC to repeated emotion assessments in at-risk and remitted adults, showing that deviations beyond control limits could signal persistent mental health problems or predict depressive episode recurrence. While SPC provided useful warning signals, its sensitivity remained limited, and its classification accuracy was moderate, highlighting both the promise and current limitations of the method [31].

EWMA-based SPC methods that use passive sensing and e-diary parameters have recently been used to detect prodromal symptoms of emerging affective episodes in patients with BD by Ludwig et al [15]. They tested whether digital phenotypes derived from passive sensing are marked as “out of bounds” during affective episodes and investigate whether these changes were already visible in the prodromal phases of affective episodes. Their results showed that, despite its potential as a low-burden tool, passive sensing did not reliably detect episodes or prodromal states. Self-reported bipolar mood was more effective than passive sensing in predicting current episodes, while prodromal phase prediction remained challenging. SPC with personalized control limits did not outperform established clinical cutoff scores, and optimizing SPC settings did not improve the balance between emerging episodes and false alarms. Ludwig et al [15] suggest that future research should focus on mobile sensing parameters more aligned with psychopathology to improve validity and sensitivity.

Research Questions

Given the reported findings of altered exploratory behavior of humans and animals in mania-like and depression-like episodes, we investigated spatial movement patterns in patients with BD. To enable a comparison between episodes (euthymic, manic prodromal, manic, depressive prodromal, and depressive), we leveraged GPS data from the BipoSense dataset to analyze characteristics of the unique places visited by patients with BD (number of unique places visited, frequency of location changes, and time spent at each place). Although this study used the same dataset as Ludwig et al [15,32], our analytical approach differed.

The previous studies from this dataset focused on e-diary measures and on standard mobile sensing parameters. As the latter approach was disappointing, we argued that future studies might focus on mobile sensing parameters that more closely align with psychopathology. Developing smarter digital phenotypes, such as specific geolocation measures derived from GPS data, represent a promising direction, as these parameters

XSL•FO
RenderX

can be aligned with empirical evidence, partly informed by animal models, while at the same time offering a low-burden, continuously monitored approach with substantial potential. Although Ludwig et al demonstrated that e-diaries still outperform passive sensing, passive sensing offers considerable advantages, as mentioned earlier. Consequently, our study aimed to identify a passive sensing parameter (or "smart digital phenotype") that is most suitable for predicting emerging episodes, which is of highest priority from a clinical perspective.

To achieve this, we used EWMA-based SPC methods as a prediction model. We hypothesized that during depressive episodes, the EWMA-based number of unique places visited is lower (oversimplified: "fewer unique locations are visited") and that the EWMA-based number of cluster changes in general is lower (oversimplified: "fewer changes between locations are happening"). Additionally, we hypothesized that the EWMA-based time spent at each place is higher (oversimplified: "individuals remain at specific locations longer"). Conversely, we expected the opposite pattern to emerge during manic episodes. Furthermore, we extended our investigation to prodromal phases, examining whether these same spatial behavior patterns precede the onset of depressive or (hypo)manic episodes. In fact, we hypothesized that the same EWMA-based spatial behavior patterns observed during depressive and (hypo)manic episodes will also manifest in prodromal phases, preceding full episodes and potentially allowing early detection. Second, we investigated whether the predictive accuracy can be enhanced by adjusting SPC settings (λ and L) and evaluate the resulting sensitivity and specificity in the context of potential clinical application.

Methods

Sample

To investigate our stated research questions, we used the BipoSense dataset [15,30,31], which offers continuous passive and active sensing over 12 months alongside biweekly expert interviews (26 interviews in total). Patients were recruited from a specialized outpatient clinic at Dresden University Hospital, meeting strict inclusion criteria, including a diagnosis of BD type I or II and a history of at least three affective episodes in the past 5 years [32,33]. A trained clinical psychologist conducted biweekly interviews (alternating between in-person and telephone) to assess categorical data on affective status (eg, being in a depressive episode, (hypo)manic episode, or euthymic) via the SCID-I section A (*DSM-5* criteria [34]) covering the previous 14 days. This served as "ground truth." Moreover, dimensional symptom scores were assessed via standardized questionnaires, including Young Mania Rating Scale [35], the Bech-Rafaelsen Mania Rating Scale [36], and the Montgomery-Asberg Depression Rating Scale [37]. These tools covered symptoms over the preceding 3 days and demonstrated excellent reliability and validity. On the basis of the biweekly SCID interviews, day-level mood status (depressed, euthymic, and (hypo)manic) was labeled for the previous 14 days. Each mood episode contributed to multiple analytical phases: preepisode weeks were categorized as early prodromal (days −14 to −8) and late prodromal (days −7 to −1), following previous research [15,32,38], and stratified by episode polarity. The onset and subsequent course were further classified as first week (days 1 - 7), second week (days 8 - 14), and ongoing (day 15 onward). This structure enabled the identification of temporal dynamics surrounding episode onset.

Of the 112 patients who initially contacted the BipoSense study, 53 (47.3%) agreed to undergo screening for eligibility. Ultimately, 31 (27.7%) patients were enrolled in the study. Reasons for exclusion included not meeting the eligibility criteria (n=15, 13.4%), nonresponse after the initial contact (n=25, 22.3%), living too far from the study site (n=18, 16.1%), and technical concerns (n=4, 3.6%), while others did not provide reasons for refusal [33].

Measures

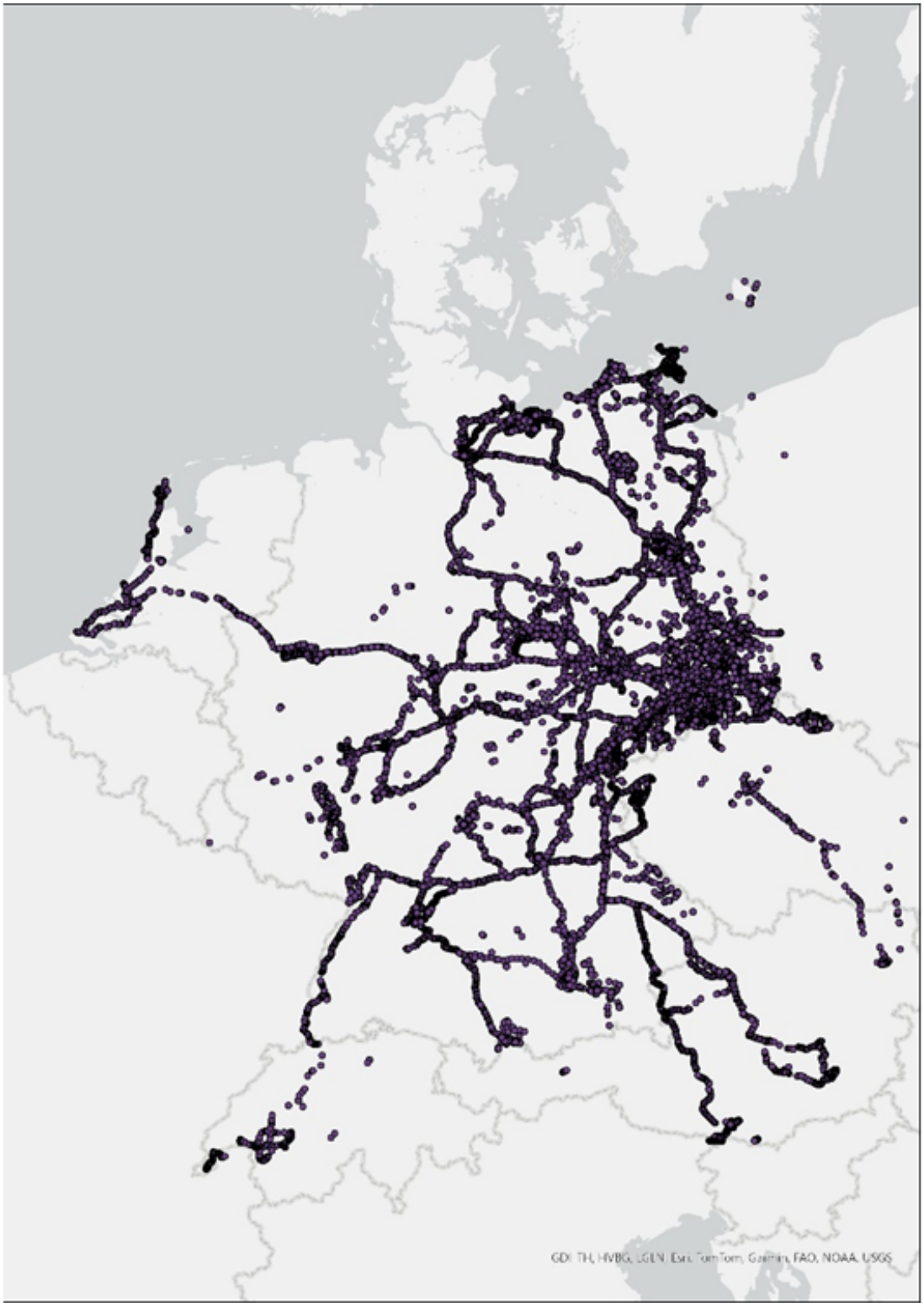
Digital phenotyping was conducted via the app movisensXS [39] (movisens GmbH, Karlsruhe, Germany). In addition to standard e-diary variables, such as end-of-day entries on sleep (time spent asleep, awake, or sleepless in bed over the previous 24 hours, recorded in 60-min segments) and medication use (both adapted from ChronoRecord, a rigorously validated electronic mood charting system [40,41]), a wide range of passive sensing variables was collected. In this study, we focused on indices of unique places as derived from GPS location data.

Data Preparation

Initially, 31 participants were included. After 3 weeks, 1 (3.2%) participant dropped out, leaving 30 (96.8%) with at least 10 months of data. One participant was excluded due to data extraction issues, and another was excluded because of the absence of 6 euthymic weeks. The final sample comprised 28 participants. For this sample, geospatial data were available at a very high resolution of approximately 5 to 10 seconds, totaling 8,616,671 data points. To optimize computational efficiency, the data were downsampled to 1 entry per minute, resulting in a reduced dataset comprising 1,861,321 data points. Afterward, the Haversine algorithm was used to calculate distances between consecutive entries. To ensure data quality, entries were filtered out if the distance from the previous point implied an exceeded speed of 300 km/h, as such deviations may result from inaccurate GPS signals, often due to battery-saving measures using GSM data instead of GPS satellites. To make a cluster analysis feasible, the points were projected from the global geodetic coordinate system, WGS84 (EPSG 4326) to UTM 32 (EPSG 25832). Owing to energy efficiency, GPS data collected via movisensXS were not recorded continuously, causing temporal gaps. If a participant remains stationary, new GPS coordinates were only captured when a change in location occurs, evidenced by a signal in the acceleration sensor of the smartphone. Accordingly, missing entries were filled with the value of the last nonmissing entry, leading to a dataset containing 14,483,536 points. The spatial distribution of all GPS points of all participants is presented in Figure 1.

XSL•FO
RenderX

Figure 1. Spatial distribution of all GPS points (14,483,536 entries) of all participants (n=28). Owing to the fact that the study took place at University Hospital Dresden, the majority of the points are concentrated in and around Dresden.



DBSCAN Cluster Analysis

To identify unique places, the density-based spatial clustering algorithm DBSCAN (Density-Based Spatial Clustering of Applications with Noise [42]) was used. Unlike k-means, DBSCAN does not require a predefined number of clusters and classifies observations as core points, border points, or noise points. Its nonparametric density estimation allows for the detection of complex cluster shapes, making it particularly effective for data with varying densities and irregular structures such as spatial data [40]. On the basis of visual tests of BipoSense data, we set epsilon to 100 m and n to 20 minutes per day. In practical terms, core points must fall within a 100 m radius to form a cluster, requiring at least 20 points.

The DBSCAN algorithm considers all points within the predefined radius (epsilon). To prevent frequently traveled

XSL•FO
RenderX

routes where GPS points overlap from being defined as clusters, an additional criterion was applied: a cluster must involve a minimum of 10 consecutive minutes spent at a unique place. Thus, a cluster is defined as "a minimum of 20 GPS measurements or minutes that fall within a 100-m diameter with at least 10 consecutive minutes spent at that location."

On the basis of the spatial data and the conducted DBSCAN, we derived the following variables:

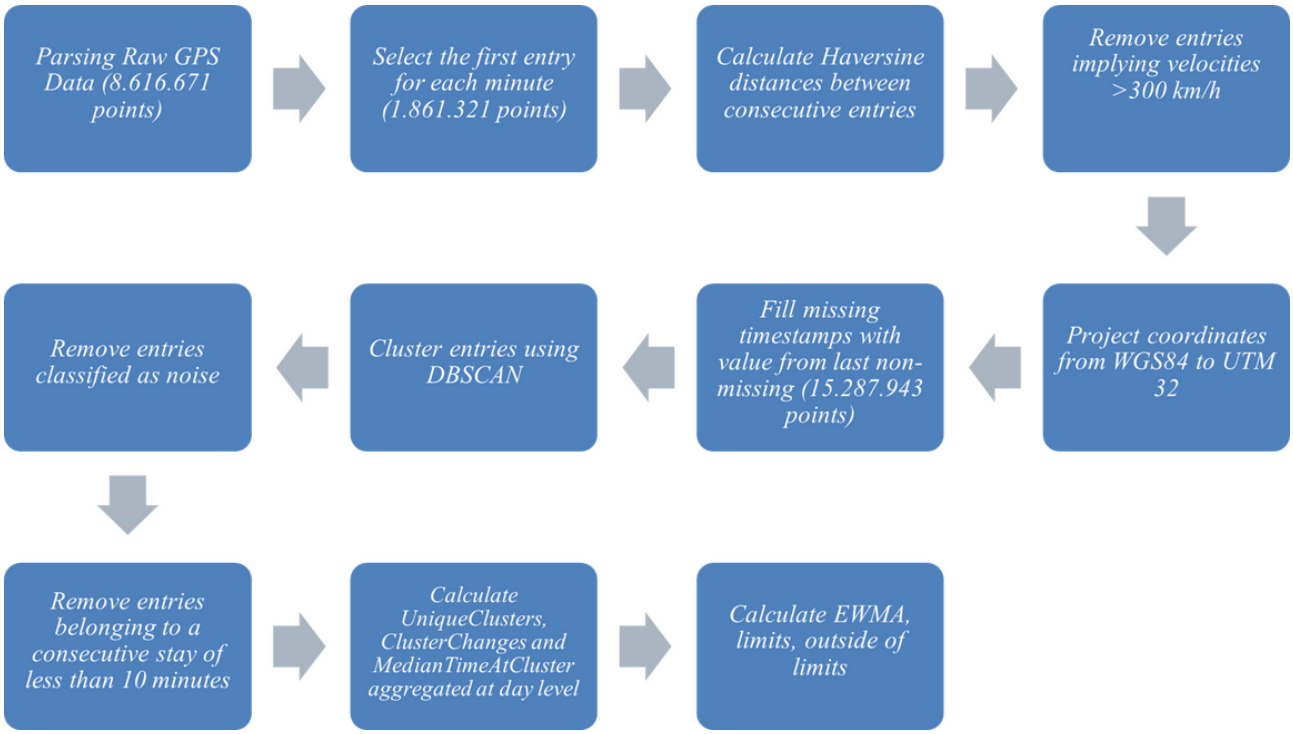
- UniqueClusters: the number of unique locations visited by the participant on that day
- ClusterChanges: the number of location (cluster) changes, grouped by each participant and each day
- MedianTimeAtCluster: the median number of minutes spent across all clusters, grouped by each participant and each day

Statistical Process Control

The central assumption was that (upcoming) depressive or manic episodes are associated with changes in behavior, which in turn are reflected in changes in the means and/or variances of appropriate mobile sensing parameters compared to a 28-day baseline period. To ensure that this baseline period reflected a stable euthymic state, it was only assigned when biweekly clinician ratings indicated the absence of an affective episode. If a participant experienced an episode at study entry, baseline assignment was postponed until a confirmed euthymic phase. Additionally, we compared mean self-rated and clinician-rated mood scores between the baseline and all other phase categories. We attempted to detect these shifts using EWMA-based control charts, which smooth the incoming data using an exponentially weighted moving average of previous values, minimizing the influence of noise. If the EWMA value exceeds or falls below control limits established using the mean and variance of the baseline period, this should be an indicator of an episode.

There are 2 parameters that can be used to tune the sensitivity and specificity of the method: λ and average run length (ARL). The smoothing parameter λ determines how much weight is given to the value of the most recent observation. The closer λ is to 0, the more weight is given to the previous days. Conversely, the closer λ is to 1, the smaller the smoothing effect and the more sensitive the EWMA is to changes. The ARL is the expected number of days before a false-positive alert occurs; that is, the limits are exceeded or undercut although there was no underlying change. λ and ARL were used to calculate another parameter L, which determines the width of the control limits. We set λ to 0.15 and decided on an ARL of 180 days, which resulted in an L of 2.536435. For comparability and methodological consistency, these parameters were applied globally across participants rather than optimized individually. This approach was chosen based on prior SPC research in affective disorders [15,29-31]. We provide a detailed illustration of SPC charts with varying parameters λ and L using synthetic data to demonstrate the principles of EWMA-based SPC (see Figures S1 and S2 in Multimedia Appendices 1 and 2). Additionally, a detailed tutorial on the application of the EWMA procedure for detecting real-time changes in intensive longitudinal psychological data is provided by Smit et al [29]. A complete list of all data processing steps is shown in Figure 2.

Figure 2. Overview of the GPS data processing methodology. The initial raw GPS data are refined through a sequence of procedures encompassing data parsing, coordinate transformation (WGS84 to UTM), Density-Based Spatial Clustering of Applications with Noise (DBSCAN) clustering, and exponentially weighted moving average (EWMA) calculation.



XSL•FO
RenderX

Generalized Linear Mixed Models

To account for the nested structure of daily observations within participants, multilevel (mixed effects) models were estimated, incorporating both fixed and random effects. Specifically, multilevel logistic regression models were applied. Each unique place parameter was treated as the independent variable (fixed main effect), while disorder status served as the binary-dependent variable. For the models, unique place parameters were dichotomized to reflect whether daily values were within or outside the control limits. To address the risk of α-error inflation due to multiple comparisons (3 variables tested across 9 disorder status contrasts), a Bonferroni-Holm correction was applied separately for the depressive and manic states, reflecting the differing number of phases. Thus, for depression, the smallest *P* value was multiplied by 15 (3×5), the second smallest by 14, and so forth. For mania, the correction followed the same scheme, starting with a multiplication factor of 12 (3×4). Upon reasonable request, day-level data will be provided.

Sensitivity and Specificity

To better assess the potential clinical value of an SPC-based early warning system, we evaluated its sensitivity and specificity. Sensitivity represents the proportion of affective episodes correctly detected out of all episodes, while specificity reflects the proportion of euthymic days accurately classified out of all euthymic days. In clinical practice, balancing sensitivity and specificity is essential. While high sensitivity is necessary to capture most emerging episodes, an excessive number of false positives may reduce the likelihood that patients and clinicians respond to frequent alerts. We used established cutoffs for both sensitivity (50%) and specificity (95%) [15] translating into detecting at least 50% of new emerging episodes while having less than one false alarm per month.

Ethical Considerations

The study adhered to the Declaration of Helsinki and was approved by the ethics committee at the Technical University of Dresden (EK-Nr.: 26012014). All participants provided written informed consent prior to participating in the study. After having signed informed consent, patients received a study smartphone (optional) and reimbursement of €35 (US $40.7) per month [33]. All self-report data were pseudonymized, and personal identifiers were removed prior to analysis. Raw GPS data, which could potentially allow for the reidentification of individual participants, are not publicly shared and were strictly accessible only to the research team, thereby ensuring the protection of participant privacy.

Results

Sample Characteristics

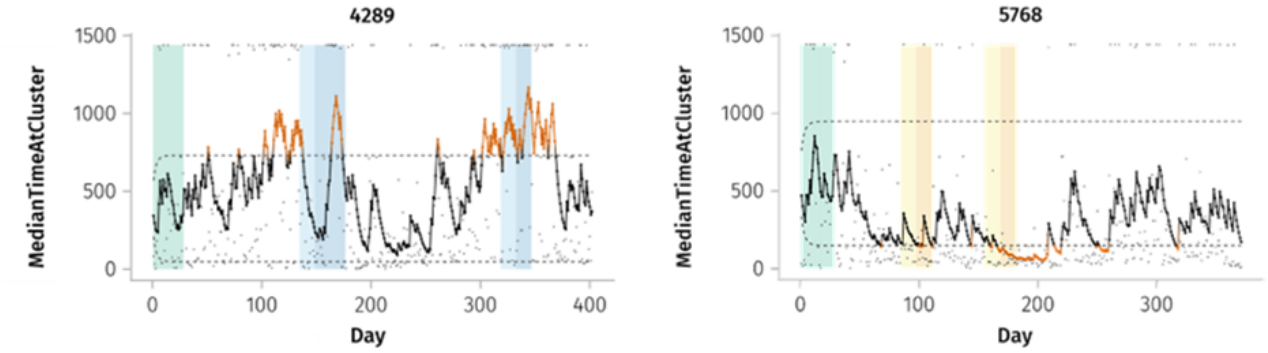
As described previously, 28 participants were included in our analyses, with a mean age of 43.82 (SD 12.09) years, ranging from 25 to 70 years. The cohort consisted of 16 (57.1%) female and 12 (42.9%) male participants, of whom 16 (57.1%) were diagnosed with BD type I and 12 (42.9%) with BD type II. Participant compliance in the BipoSense dataset was notably high. In total, 726 (97%) of 749 biweekly diagnostic visits were completed. Participants completed 9433 (89%) evening e-diary entries, with 1154 (11%) entries missing due to technical issues, low battery, or nonresponse [33]. Over the course of the assessment period, mobile sensing data were recorded for 10,213 (99%) patient days, covering 26 depressive and 20 (hypo)manic episodes. This included 7851 (76.9%) euthymic days, 161 (1.5%) days classified as early prodromal depressive, 164 (1.6%) days classified as late prodromal depressive, 179 (1.8%) days marking the onset of a depressive episode, 168 (1.6%) days within a second depressive episode, and 194 (1.9%) days categorized as part of ongoing depressive phases. In total, 866 days (8.5% of all observation days) were classified as depressive phases across all participants. Regarding (hypo)mania, the dataset comprised 133 (1.3%) early prodromal days, 141 (1.4%) late prodromal days, 147 (1.4%) days marking the onset of a (hypo)manic episode, 146 (1.4%) days within a second (hypo)manic episode, and 98 (1.0%) days of ongoing (hypo)manic states. This resulted in 665 days (6.5% of all observation days) classified as manic episodes. Additionally, 856 (8.4%) days were marked as missing. For statistical analysis, extended (hypo)manic episodes lasting longer than 3 weeks were excluded due to an insufficient number of recorded days.

EWMA-Based SPC Charts

Figure 3 shows 2 example SPC plots for the MedianTimeAtCluster variable in 2 patients, that is, 4289 and 5768. The baseline phase is depicted by the green area, while affective episodes are color-coded (yellow for manic episodes and blue for depressive episodes). Prodromal phases are denoted by lighter shaded areas that precede the affective episodes. The daily EWMA is depicted in black when within the upper and lower control limits and in red when values exceed these thresholds. To verify that the baseline periods represented stable mood states suitable for EWMA phase I initialization, we compared clinician-rated and self-rated symptom scores across all phase categories. Ensuring that the baseline represents a stable state is crucial for the validity of the EWMA procedure. As expected, baseline scores closely resembled those observed during euthymia, with minimal symptom burden (eg, baseline YMRS: mean 1.06, SD 1.95 vs euthymia YMRS: mean 1.06, SD 1.88; baseline self-rating: mean 49.25, SD 12.45 vs euthymia: mean 49.88, SD 10.46). In contrast, both manic and depressive phases were clearly distinguishable from baseline. For example, during depressive phases, participants showed low YMRS scores (first week: mean 0.54, SD 1.24; ongoing: mean 0.61, SD 1.08) alongside reduced self-rated mood (first week: mean 39.71, SD 14.60; ongoing: mean 31.14, SD 16.31), whereas during manic phases, YMRS increased (first week: mean 9.57, SD 5.07; ongoing: mean 9.98, SD 4.42) with correspondingly higher self-rated mood (first week: mean 54.88, SD 9.37; ongoing: mean 51.24, SD 10.40). These differences confirm that baseline periods were indeed “in control” and distinctly separate from symptomatic phases. A complete overview of clinician-rated and self-rated symptom scores across all phase categories is provided in Table S1 in Multimedia Appendix 3.

XSL•FO
RenderX

Figure 3. Exponentially weighted moving average (EWMA) statistical process control chart for participants 4289 and 5768 for the variable MedianTimeAtCluster. The x-axis represents the days of the study, and the y-axis shows the median time participants spent at a cluster. The baseline period used to compute personalized upper and lower control limits is depicted by the green area. Yellow areas indicate (hypo)manic episodes, and blue areas indicate depressive episodes. Shaded regions mark the 2 prodromal weeks (early and late prodromal). Black dots represent daily raw values, and the black line depicts the EWMA. The line turns red when the EWMA exceeds the control limits.



As can be seen, the EWMA-based median time per cluster significantly fluctuates over time and shows low values during manic episodes, whereas values are higher during and after a depressive episode. Nevertheless, we have to strongly emphasize that we selected those particular exemplar figures because they illustrate the proposed concept best. A more comprehensive examination of all variables across all patients revealed a more ambiguous picture. SPC charts for all observed parameters from patients 4289 and 5768 can be found in Figure S3 in Multimedia Appendix 4.

Testing the SPC Assumptions on Unique Places Visited

Overview

Table 1 presents the results of the multilevel models examining whether control limits distinguish affective episodes from euthymic days. In reference to Figure 3, these models assess whether the proportion of "red dots" is higher during affective episodes compared to euthymic periods.

Table . Estimates from multilevel logit models differentiating disorder status[a].

Disorder status and phase	Unique clusters		Cluster changes		Median time at cluster	
	OR[b] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Depressive episodes versus euthymia						
Early prodromal	0.623 (0.34-1.15)	.13	0.920 (0.86-0.98)	.01	1.002 (1.00-1.004)	.01
Late prodromal	0.803 (0.56-1.16)	.24	0.978 (0.96-0.99)	.01	1.001 (0.99-1.01)	.82
1st week depressive episode	0.910 (0.81-1.03)	.12	1.007 (1.00-1.02)	.08	1.002 (1.00-1.01)	.39
2nd week depressive episode	0.826 (0.14-4.73)	.83	1.017 (0.82-1-27)	.88	1.001 (1.00-1.002)	.08
Ongoing depressive weeks	0.803 (0.58-1.12)	.19	0.975 (0.94-1.02)	.17	1.001 (1.00-1.002)	<.001
(Hypo)manic episodes versus euthymia						
Early prodromal	0.997 (0.99-1.01)	.45	1.031 (0.91-1.17)	.63	1.001 (1.00-1.002)	.05
Late prodromal	0.782 (0.60-1.02)	.07	0.985 (0.97-1.01)	.12	1.002 (1.00-1.01)	.56
1st week (hypo)manic episode	0.958 (0.46-2.02)	.91	1.012 (0.96-1.05)	.53	1.001 (1.00-1.002)	<.001
2nd week (hypo)manic episode	1.128 (0.83-1.53)	.44	1.031 (1.01-1.06)	.01	1.000 (1.00-1.001)	.05

[a]Disorder status is categorized into the following phases: "early prodromal" (days 14 to 8 prior to episode onset), "late prodromal" (days 7 to 1 before onset), "first week" (days 1 to 7 of the episode), "second week" (days 8 to 14), and "ongoing weeks" (day 15 onward), with separate classifications for (hypo)mania and depression. "Euthymia" serves as the reference category.

[b]OR: odds ratio.

XSL•FO
RenderX

In the early prodromal phase preceding depressive episodes, a lower number of EWMA-based unique clusters (odds ratio [OR] 0.623, 95% CI 0.34-1.15) was observed, although this is not statistically significant. EWMA-based cluster changes (OR 0.920, 95% CI 0.86-0.98; *P*=.01) show a significant negative association, indicating fewer cluster transitions in this phase. EWMA-based median time spent at each cluster (OR 1.002, 95% CI 1.0-1.004; *P*=.01) is significantly elevated, suggesting that higher median time spent at clusters is linked to an increased likelihood of entering a depressive episode. During the late prodromal phase, EWMA-based unique clusters (OR 0.803, 95% CI 0.56-1.12), EWMA-based cluster changes (OR 0.978, 95% CI 0.96-1.0), and EWMA-based median cluster values (OR 1.001, 95% CI 0.9-1.01) show no statistically significant associations with depressive onset. During the first and second weeks of depressive episodes, no cluster indicators show significant effects. In ongoing depressive episodes, EWMA-based unique clusters (OR 0.803, 95% CI 0.58-1.1) and EWMA-based cluster changes (OR 0.975, 95% CI 0.96-1.05) are not significantly associated with the disorder status. However, EWMA-based median time spent at clusters (OR 1.001, 95% CI 1.000-1.001; *P*<.001) is significantly elevated.

Regarding (hypo)manic episodes, in the early prodromal phase as well as late prodromal phase, EWMA-based unique clusters, EWMA-based cluster changes, and EWMA-based median time spent values show no statistically significant associations. During the first week of a (hypo)manic episode, EWMA-based unique clusters (OR 0.958, 95% CI 0.46-2.02) and EWMA-based cluster changes (OR 1.012, 95% CI 0.97-1.05) are not significantly related to disorder status. However, EWMA-based median time spent at cluster (OR 1.001, 95% CI 1.000-1.002) demonstrated a small but statistically significant association with (hypo)manic symptom onset. In the second week of a (hypo)manic episode, EWMA-based unique clusters and EWMA-based MedianTimeAtCluster show no significant effects. In contrast, EWMA-based cluster changes (OR 1.031, 95% CI 1.01-1.06; *P*=.01) exhibit a significant positive association, indicating increased cluster variability during this phase.

Bonferroni-Holm Correction

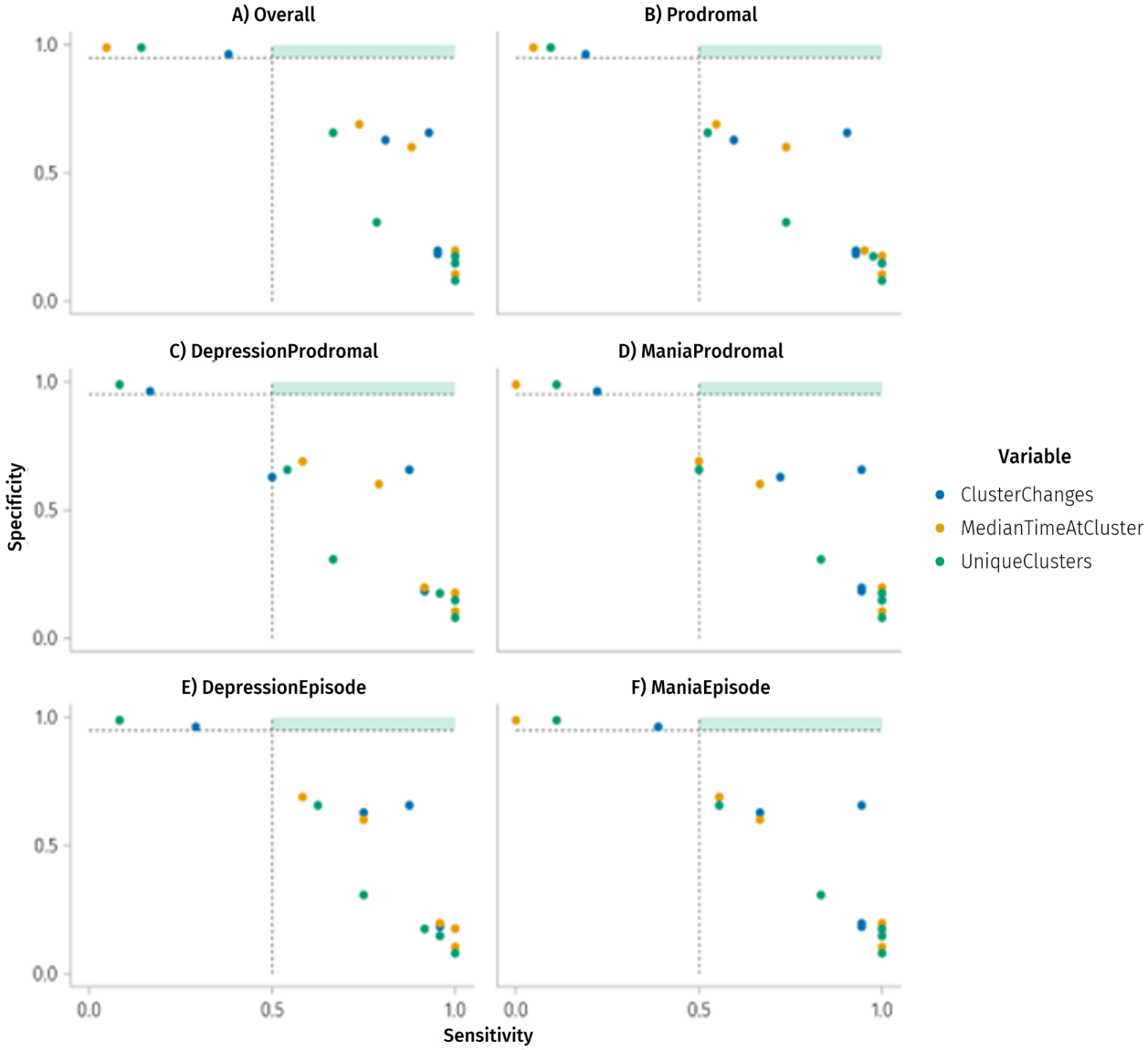
To account for random findings caused by multiple testing, we used a Bonferroni-Holm correction, as described in the *Methods* section. From the reported 6 significant findings in Table 1, only 2 remained significant, namely, EWMA-based MedianTimeAtCluster for ongoing depressive weeks (*P*=.01) and EWMA-based MedianTimeAtCluster for the 1st week of (hypo)manic episode (*P*=.004).

Optimizing Sensitivity and Specificity

In the next step, the clinical applicability of an SPC-based early warning system was evaluated through the computation of sensitivity and specificity metrics. Before conducting the analysis, a threshold was defined whereby identifying at least every second impending episode was considered satisfactory (sensitivity >50%), while false-positive alerts were to remain limited to a maximum of one per month (specificity >95%). As detailed in the *Methods* section, the SPC parameters λ and L were systematically optimized and assessed in accordance with 6 distinct optimization criteria. Figure 4 compiles and graphically depicts the sensitivity and specificity evaluations. The green markers in the upper-right segment represent the predetermined cutoff thresholds. Cluster parameters are highlighted using color distinctions.

XSL•FO
RenderX

Figure 4. Results for sensitivity and specificity after optimizing the parameters λ and L are presented, divided according to different periods and clinical phases: (A) Overall represents the entire assessment period, (B) prodromal includes all prodromal days, (C) depression prodromal covers only depressive prodromal days, (D) mania prodromal covers only manic prodromal days, (E) depression includes all depressive days, and (F) mania includes only manic days during optimization. The dotted lines and the green area indicate the suggested threshold of 50% sensitivity and 95% specificity, respectively, which we consider a prerequisite for clinical application.



While some configurations approach the predefined thresholds of sensitivity (>50%) and specificity (>95%), none of the evaluated variables met both criteria at the same time. Specificity is often too low, particularly for detecting depressive and manic episodes (panels E and F), leading to excessive false positives. In contrast, prodromal and depression prodromal phases showed a better balance, suggesting potential clinical utility in these cases. The 3 variables (ClusterChanges, MedianTimeAtCluster, and UniqueClusters) vary in performance, without a clear best predictor. For a complete overview of the results, please refer to Table S2 in Multimedia Appendix 5.

Discussion

Principal Findings

This study investigated the application of EWMA-based SPC methods using spatial movement patterns to detect prodromal symptoms of emerging affective episodes in individuals with BD. Building upon previous work, we examined whether location-based smart digital phenotypes derived from GPS data, namely the number of unique places visited, the frequency of location (cluster) changes, and the median number of minutes spent across all clusters could effectively capture the dynamic changes in behavior associated with different phases of BD.

First, we observed that, at least in BD, individual movement patterns indeed vary over time, in contrast to the findings of González et al [26] and Allessandretti et al [27]. Regarding our

XSL•FO
RenderX

hypotheses, visual descriptive approaches, such as examining SPC charts, demonstrated the potential of this method. During depressive phases, we observed that fewer locations were visited, and individuals tended to spend longer durations at specific locations, which aligns with our hypothesis. Conversely, during manic episodes, the opposite pattern emerged, which was consistent with our expectations. Notably, it is observed that the out-of-bounds processes were flagged even before the onset of the episodes, which may hold potential for prospective phase prediction and eventual clinical application. It can be assumed that during depressive episodes, individuals spend more time in a cluster, reflecting symptoms of apathy, whereas during manic episodes, they spend significantly less time in such clusters, indicating restlessness and increased mobility. This observation is in line with animal models that demonstrate altered exploratory behavior in mania-like [19] and depression-like [20] states.

However, a broader analysis across all patients and variables (Table 1) reveals a more complex and ambiguous picture. Examining whether control limits distinguish affective episodes from euthymic days via multilevel models revealed that overall, the median time spent at clusters emerged as the most consistent indicator across both depressive and (hypo)manic episodes. In contrast, the number of unique clusters was generally not significantly associated with phase transitions, even before Bonferroni-Holm correction. These patterns suggest that the frequency of location changes may serve as a more robust marker of affective episodes than the sheer number of distinct locations visited. Specifically, participants experiencing depressive episodes tend to remain temporally and spatially bound, whereas those in (hypo)manic episodes are highly mobile. In the latter group, individuals tend to move repeatedly between certain locations rather than visiting a large number of distinct places. Moreover, the phase-specific significance indicates that different aspects of movement behavior become relevant at different stages of episode development, highlighting the complexity of predicting mood episodes from passive sensing data.

Although our dataset is of high quality, featuring 1 year of longitudinal data and frequent reliable expert interviews, creating ideal conditions for applying SPC methods, digital phenotypes did not consistently prove reliable in detecting prodromal states and affective episodes through SPC analysis. Thus, the results partially support our initial hypotheses, suggesting that the analysis of visits to unique places may hold promise as a tool for capturing subtle behavioral changes indicative of impending mood shifts. In our descriptive investigation of optimizing predictions through systematic adjustments of SPC settings, we found that the inability to achieve sufficient sensitivity and specificity, even post hoc with the optimal parameter combinations, challenges the approach and/or the variables used. Previous EWMA-SPC studies on intensive emotion data indicated early warning signals for depressive episodes, but with limited sensitivity [31]. Consistent with this, our application to GPS-derived movement patterns showed some phase-specific behavioral changes, yet predictive accuracy remained insufficient, highlighting both the potential and current limitations of SPC for real-time episode detection.

Comparing our findings to previous studies, our results are in line with and expand upon the observations of Palmius et al [16] and Faurholt-Jepsen et al [18], who reported correlations between spatial movement patterns and mood states. While these earlier studies primarily focused on symptom correlations, our approach incorporates clearly defined episode onsets and aims to predict episodes prospectively. The more differentiated variables used here (unique clusters, cluster changes, and median time) allow for a more nuanced characterization of spatial behavior. This provides a more granular understanding of how changes in spatial behavior precede and accompany mood episodes, highlighting both consistencies and divergences with prior work. However, in direct comparison to Ludwig et al [15], self-rated mood appeared to outperform even the more advanced smart digital phenotypes examined in this study, although its ability to predict prodromal phases remains limited as well.

While systematic reviews of geospatial analyses in affective disorders [14] demonstrate remarkable potential for predicting depressive symptoms, their application to BD, as outlined earlier, is far from trivial. Moreover, predicting full episodes remains particularly challenging as demonstrated in our analysis. However, to effectively advance research on BD and integrate digital phenotyping into clinical practice, it is essential not only to adopt a purely dimensional approach to BD symptomatology [16,18] but also to differentiate between distinct illness phases and predict their occurrence. As demonstrated by Palmius et al [16] in the analysis of spatial patterns and Faurholt-Jepsen et al [18] in the analysis of unique places, while clear correlations were identified, the transfer to episode prediction remains challenging.

In summary, while the analysis of visits to unique places offers a low-burden, continuous monitoring approach with high potential, our findings did not confirm the consistent effectiveness of EWMA-based SPC methods for predicting affective episodes using passive sensing data in clinical practice. Furthermore, the use of EWMA-based SPC with individualized control limits did not surpass established threshold values.

Limitations

First, although the study exceeded the duration of previous studies by a significant margin and included patients with a high number of prior episodes, providing the most temporally precise labeling of days per patient, the number of newly emerging episodes remained limited. Future research aimed at predicting new episodes may benefit from longer study durations, such as the 18-month period used in our most recent randomized controlled trial [5]. Second, while it has been suggested that frequent assessments of psychopathological status could prevent the occurrence of new episodes, our dataset provides no empirical evidence to support this claim [33]. Third, our study deviated from the intended chronological sequence in 7 cases due to insufficient initial euthymic days, requiring retrospective upper control limit or lower control limit application and compromising the SPC's prospective nature. While this aimed to capture more affective episodes, it may have introduced confounding effects from postepisode behavioral changes, such as increased activity after depression. Moreover, participants were aware that their GPS data were being collected, which

XSL•FO
RenderX

may have influenced their behavior due to a potential Hawthorne effect. However, as our study covered an extended observation period of 365 days, we do not assume that participants consistently altered their behavior throughout the entire duration. Another limitation concerns the evaluation of model performance, which was based on prespecified operating thresholds (≥95% specificity and ≥50% sensitivity) chosen based on clinical reasoning rather than empirically derived or standardized criteria. While this choice reflects the practical prioritization of minimizing false-positive alerts in real-world clinical settings, it necessarily introduces a degree of subjectivity into the performance assessment. In particular, the strict specificity constraint may limit achievable sensitivity and thereby influence the apparent trade-off between both metrics. Moreover, the absence of established guidelines for optimal threshold selection in SPC-based early warning systems for affective disorders limits the comparability of our results to other approaches and may affect generalizability across different clinical contexts and monitoring frequencies. Finally, prodromal phases were defined exclusively by a temporal criterion, specifically the 2-week period prior to affective episodes, without clinical assessment of actual prodromal symptoms. As a result, we lack a "ground truth" and cannot rule out the possibility that patients may not have exhibited any prodromal symptoms to capture to begin with, whether via passive sensing or self-reported bipolar mood. This observation is consistent with clinical reports in which patients sometimes describe a sudden "switch" from euthymia to an affective episode overnight. Our data may partially support this hypothesis, as, at least for (hypo)mania, late prodromal days were more frequently differentiated from euthymia than early prodromal days.

Future Directions

Although our study makes a significant contribution to the current state-of-the-art research by developing smart digital phenotypes based on GPS data in patients with BD, we have only scratched the surface with regard to location-based passive sensing. First, we acknowledge that no direct comparison with classical entropy-based mobility metrics was conducted in this study. While our DBSCAN-based feature set was designed to capture more granular spatial-temporal structure than aggregated entropy measures, we cannot empirically claim superiority over entropy without a head-to-head evaluation. Importantly, from a theoretical standpoint for longitudinal spatial data, DBSCAN is particularly well suited, as it preserves discrete location states and transitions, which are otherwise compressed in entropy formulations. This property is essential for enabling the derivation and interpretation of the specific variables used in this work (UniqueClusters, ClusterChanges, and MedianTimeAtCluster), which explicitly depend on identifiable spatial clusters rather than aggregate distributional summaries. Future work should explicitly benchmark entropy-based and cluster-based mobility representations within the same predictive framework to quantify their relative clinical utility. Second, there are numerous other potential parameters that could correspond closely to psychopathological features. These include exposure to highly frequented areas, such as city centers and shopping malls, as well as exposure to green spaces or urban environments [43]. Additionally, interactions related to residential locations present a promising avenue for future exploration. Thus, future research should explore how multiple behavioral and physiological indicators can be integrated into a more comprehensive model. Furthermore, incorporating machine learning algorithms to identify complex patterns across multiple variables could improve the sensitivity and specificity of the model. Finally, regarding our SPC methods, previous work suggests that combining changes in variability (eg, SD) with mean-level changes may improve the sensitivity of early warning signals [31,44]. In our study, we focused on mean-based SPC indicators, which may have contributed to the limited sensitivity observed. Future research should therefore explore combined approaches integrating both mean and variability-based indicators.

Conclusions

In conclusion, this study contributes to the growing body of evidence supporting the use of digital phenotyping in BD with a particular focus on location-based passive sensing, specifically visits to unique places by individuals with BD. While SPC charts highlight the potential of this approach, a broader analysis across all patients and variables reveals a more complex and ambiguous picture, with overall effects being limited and of insufficient clinical applicability. Notably, after applying Bonferroni-Holm corrections to our generalized mixed models, only time spent at each cluster remained of some significance for certain phases, underscoring the selective predictive value of this approach. The comparison between SPC charts and the statistical modeling approach showed that further research is needed to refine and validate these findings. Yet, our results underscore the conceptual value of digital phenotyping for improving our understanding and monitoring of BD. By leveraging the power of mobile technology and data analytics, we may move toward a future where personalized interventions could be delivered proactively.

Acknowledgments

As non-native English speakers, we used the generative AI tool ChatGPT (OpenAI) to assist with improving grammar and style. All outputs were reviewed and revised by the authors.

Funding

Research reported in this publication was supported in part by a grant from the German Research Foundation/Deutsche Forschungsgemeinschaft (DFG Project 521379614 [CRC 393 "Trajectories of affective disorders: cognitive-emotional mechanisms of symptom change"], the DFG GRK2739/2 – Project 447089431 – Research Training Group: KD2School – Designing

XSL•FO
RenderX

Biosignal-Adaptive Systems for Decision-Making Processes," and the Federal Ministry of Research and Education (BMBF): BipoCare.

Data Availability

The data are highly sensitive, as raw GPS data cannot be fully anonymized and therefore cannot be publicly shared. Aggregated data, which do not allow for the identification of individual participants, are available from the corresponding author upon reasonable request. The full code used for all analyses is freely available on GitHub [45].

Authors' Contributions

Conceptualization: MG, UWE-P
Data curation: CB
Formal analysis: MG, CB
Methodology: MG, CB, VML, UWE-P
Project administration (BipoSense study): UWE-P, MB, EM, WES
Supervision: UE-P
Visualization: CB
Writing – original draft: MG
Writing – review & editing: CB, CK, VML, LMH, WES, MB, PR, UWE-P
All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Eexponentially weighted moving average charts illustrating the effect of varying control limits (L=2.0, 2.54, and 3.0), which correspond to different average run lengths (ARL0) and thus alter the sensitivity to process variation.
[DOCX File, 192 KB - mhealth_v14i1e77272_app1.docx]

Multimedia Appendix 2
Exponentially weighted moving average charts illustrating the impact of different smoothing parameters (λ=0.05, 0.15, and 0.3) on the detection of shifts in synthetic process data.
[DOCX File, 191 KB - mhealth_v14i1e77272_app2.docx]

Multimedia Appendix 3
Scores of clinician-rated and self-rated measures across baseline and the 11 mood phases.
[DOCX File, 19 KB - mhealth_v14i1e77272_app3.docx]

Multimedia Appendix 4
Exponentially weighted moving average (EWMA) statistical process control (SPC) chart for participants 4289 and 5768 for all remaining variables of the study.
[DOCX File, 11214 KB - mhealth_v14i1e77272_app4.docx]

Multimedia Appendix 5
Results of the optimization procedure of λ and L.
[DOCX File, 22 KB - mhealth_v14i1e77272_app5.docx]

References

1. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry 2022 Feb;9(2):137-150. [doi: 10.1016/S2215-0366(21)00395-3]
2. Grande I, Berk M, Birmaher B, Vieta E. Bipolar disorder. Lancet 2016 Apr 9;387(10027):1561-1572. [doi: 10.1016/S0140-6736(15)00241-X] [Medline: 26388529]
3. Yatham LN, Kennedy SH, Parikh SV, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. Bipolar Disord 2018 Mar;20(2):97-170. [doi: 10.1111/bdi.12609] [Medline: 29536616]
4. Ebner-Priemer U, Santangelo P. Digital phenotyping: hype or hope? Lancet Psychiatry 2020 Apr;7(4):297-299. [doi: 10.1016/S2215-0366(19)30380-3] [Medline: 31635970]

XSL•FO
RenderX

5. Mühlbauer E, Bauer M, Ebner-Priemer U, et al. Effectiveness of smartphone-based ambulatory assessment (SBAA-BD) including a predicting system for upcoming episodes in the long-term treatment of patients with bipolar disorders: study protocol for a randomized controlled single-blind trial. BMC Psychiatry 2018 Oct 26;18(1):349. [doi: 10.1186/s12888-018-1929-y] [Medline: 30367608]
6. Anmella G, Faurholt‐Jepsen M, Hidalgo‐Mazzei D, et al. Smartphone‐based interventions in bipolar disorder: systematic review and meta‐analyses of efficacy. A position paper from the International Society for Bipolar Disorders (isbd) Big Data Task Force. Bipolar Disord 2022 Sep;24(6):580-614. [doi: 10.1111/bdi.13243]
7. Ortiz A, Maslej MM, Husain MI, Daskalakis ZJ, Mulsant BH. Apps and gaps in bipolar disorder: a systematic review on electronic monitoring for episode prediction. J Affect Disord 2021 Dec;295:1190-1200. [doi: 10.1016/j.jad.2021.08.140]
8. de Azevedo Cardoso T, Kochhar S, Torous J, Morton E. Digital tools to facilitate the detection and treatment of bipolar disorder: key developments and future directions. JMIR Ment Health 2024;11:e58631. [doi: 10.2196/58631]
9. Faurholt-Jepsen M, Frost M, Ritz C, et al. Daily electronic self-monitoring in bipolar disorder using smartphones - the MONARCA I trial: a randomized, placebo-controlled, single-blind, parallel group trial. Psychol Med 2015 Oct;45(13):2691-2704. [doi: 10.1017/S0033291715000410] [Medline: 26220802]
10. Faurholt-Jepsen M, Frost M, Christensen EM, Bardram JE, Vinberg M, Kessing LV. The effect of smartphone-based monitoring on illness activity in bipolar disorder: the MONARCA II randomized controlled single-blinded trial. Psychol Med 2020 Apr;50(5):838-848. [doi: 10.1017/S0033291719000710] [Medline: 30944054]
11. Wadle LM, Ebner-Priemer UW, Foo JC, et al. Speech features as predictors of momentary depression severity in patients with depressive disorder undergoing sleep deprivation therapy: ambulatory assessment pilot study. JMIR Ment Health 2024 Jan 18;11:e49222. [doi: 10.2196/49222] [Medline: 38236637]
12. American Psychiatric Association, DSM-5 Task Force. Diagnostic and Statistical Manual of Mental Disorders: DSM-5, 5th edition: American Psychiatric Publishing, Inc; 2013. [doi: 10.1176/appi.books.9780890425596]
13. Torous J, Summergrad P, Nassir Ghaemi S. Bipolar disorder in the digital age: new tools for the same illness. Int J Bipolar Disord 2016 Dec;4(1):25. [doi: 10.1186/s40345-016-0065-1] [Medline: 27858348]
14. Shin J, Bae SM. A systematic review of location data for depression prediction. IJERPH 2023;20(11):5984. [doi: 10.3390/ijerph20115984]
15. Ludwig VM, Bittendorf CA, Reinhard I, et al. Predicting depressive and manic episodes in patients with bipolar disorder using statistical process control methods on passive sensing data. J Psychopathol Clin Sci 2025 Nov;134(8):971-981. [doi: 10.1037/abn0001002] [Medline: 40689911]
16. Palmius N, Tsanas A, Saunders KEA, et al. Detecting bipolar depression from geographic location data. IEEE Trans Biomed Eng 2017 Aug;64(8):1761-1771. [doi: 10.1109/TBME.2016.2611862]
17. Altieri L, Cocchi D. Entropy Measures for Environmental Data: Springer Nature Singapore; 2024. [doi: 10.1007/978-981-97-2546-5]
18. Faurholt-Jepsen M, Busk J, Vinberg M, et al. Daily mobility patterns in patients with bipolar disorder and healthy individuals. J Affect Disord 2021 Jan 1;278:413-422. [doi: 10.1016/j.jad.2020.09.087] [Medline: 33010566]
19. van Enkhuizen J, Geyer MA, Minassian A, Perry W, Henry BL, Young JW. Investigating the underlying mechanisms of aberrant behaviors in bipolar disorder from patients to models: rodent and human studies. Neurosci Biobehav Rev 2015 Nov;58:4-18. [doi: 10.1016/j.neubiorev.2015.08.008] [Medline: 26297513]
20. Strekalova T, Spanagel R, Bartsch D, Henn FA, Gass P. Stress-induced anhedonia in mice is associated with deficits in forced swimming and exploration. Neuropsychopharmacology 2004 Nov;29(11):2007-2017. [doi: 10.1038/sj.npp.1300532] [Medline: 15266352]
21. Minassian A, Henry BL, Young JW, Masten V, Geyer MA, Perry W. Repeated assessment of exploration and novelty seeking in the human behavioral pattern monitor in bipolar disorder patients and healthy individuals. PLoS ONE 2011;6(8):e24185. [doi: 10.1371/journal.pone.0024185] [Medline: 21912623]
22. Henry BL, Minassian A, Patt VM, et al. Inhibitory deficits in euthymic bipolar disorder patients assessed in the human behavioral pattern monitor. J Affect Disord 2013 Sep 25;150(3):948-954. [doi: 10.1016/j.jad.2013.05.020] [Medline: 23759280]
23. Perry W, Minassian A, Paulus MP, et al. A reverse-translational study of dysfunctional exploration in psychiatric disorders: from mice to men. Arch Gen Psychiatry 2009 Oct;66(10):1072-1080. [doi: 10.1001/archgenpsychiatry.2009.58] [Medline: 19805697]
24. Young LT, Bagby RM, Cooke RG, Parker JD, Levitt AJ, Joffe RT. A comparison of Tridimensional Personality Questionnaire dimensions in bipolar disorder and unipolar depression. Psychiatry Res 1995 Sep 29;58(2):139-143. [doi: 10.1016/0165-1781(95)02684-o] [Medline: 8570765]
25. Janowsky DS, Morter S, Hong L, Howe L. Myers Briggs Type Indicator and Tridimensional Personality Questionnaire differences between bipolar patients and unipolar depressed patients. Bipolar Disord 1999 Dec;1(2):98-108. [doi: 10.1034/j.1399-5618.1999.010207.x] [Medline: 11252666]
26. González MC, Hidalgo CA, Barabási AL. Understanding individual human mobility patterns. Nature New Biol 2008 Jun 5;453(7196):779-782. [doi: 10.1038/nature06958] [Medline: 18528393]

XSL•FO
RenderX

27. Alessandretti L, Sapiezynski P, Sekara V, Lehmann S, Baronchelli A. Evidence for a conserved quantity in human mobility. Nat Hum Behav 2018 Jul;2(7):485-491. [doi: 10.1038/s41562-018-0364-x] [Medline: 31097800]
28. Shewhart WA. Economic Control of Quality of Manufactured Product: Macmillan and Co., Ltd; 1931.
29. Smit AC, Schat E, Ceulemans E. The exponentially weighted moving average procedure for detecting changes in intensive longitudinal data in psychological research in real-time: a tutorial showcasing potential applications. Assessment 2023 Jul;30(5):1354-1368. [doi: 10.1177/10731911221086985] [Medline: 35603660]
30. Schreuder MJ, Kuppens P, Schat E, de Jonge P, Hartman CA, Ceulemans E. Warning signals for mental health problems in at‐risk young adults may be informed by momentary emotions reported by the general population: a novel application of the principles of statistical process control. Mental Health Science 2024;2(3):e27. [doi: 10.1002/mhs2.72]
31. Schreuder MJ, Schat E, Smit AC, Snippe E, Ceulemans E. Monitoring emotional intensity and variability to forecast depression recurrence in real time in remitted adults. J Consult Clin Psychol 2024 Aug;92(8):505-516. [doi: 10.1037/ccp0000871] [Medline: 38512172]
32. Ludwig VM, Reinhard I, Mühlbauer E, et al. Limited evidence of autocorrelation signaling upcoming affective episodes: a 12-month e-diary study in patients with bipolar disorder. Psychol Med 2024 Jun;54(8):1844-1852. [doi: 10.1017/S0033291723003811] [Medline: 38284217]
33. Ebner-Priemer UW, Mühlbauer E, Neubauer AB, et al. Digital phenotyping: towards replicable findings with comprehensive assessments and integrative models in bipolar disorders. Int J Bipolar Disord 2020 Nov 17;8(1):35. [doi: 10.1186/s40345-020-00210-4] [Medline: 33211262]
34. First MB, Williams JB, Karg RS, Spitzer RL. User's Guide for the Structured Clinical Interview for DSM-5 Disorders, Research Version (SCID-5-RV): American Psychiatric Association; 2015.
35. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry 1978 Nov;133:429-435. [doi: 10.1192/bjp.133.5.429] [Medline: 728692]
36. Bech P, Bolwig TG, Kramp P, Rafaelsen OJ. The Bech-Rafaelsen Mania Scale and the Hamilton Depression Scale. Acta Psychiatr Scand 1979 Apr;59(4):420-430. [doi: 10.1111/j.1600-0447.1979.tb04484.x] [Medline: 433633]
37. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979 Apr;134:382-389. [doi: 10.1192/bjp.134.4.382] [Medline: 444788]
38. Ludwig VM, Bittendorf CA, Reinhard I, et al. Early warning signals of bipolar relapse: Investigating critical slowing down in smartphone data. J Affect Disord 2025 Dec 15;391:119972. [doi: 10.1016/j.jad.2025.119972] [Medline: 40738339]
39. Experience sampling platform - movisensxs. movisens. URL: https://www.movisens.com/en/products/movisensxs/ [accessed 2026-06-01]
40. Bauer M, Wilson T, Neuhaus K, et al. Self-reporting software for bipolar disorder: validation of ChronoRecord by patients with mania. Psychiatry Res 2008 Jun 30;159(3):359-366. [doi: 10.1016/j.psychres.2007.04.013] [Medline: 18423616]
41. Bauer M, Glenn T, Keil M, et al. Brief depressive symptoms in patients with bipolar disorder: analysis of long-term self-reported data. Aust N Z J Psychiatry 2012 Nov;46(11):1068-1078. [doi: 10.1177/0004867412452017] [Medline: 22734088]
42. Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining KDD'96 1996:226-231 [FREE Full text]
43. Xu J, Liu N, Polemiti E, et al. Effects of urban living environments on mental health in adults. Nat Med 2023 Jun;29(6):1456-1467. [doi: 10.1038/s41591-023-02365-w] [Medline: 37322117]
44. Snippe E, Smit AC, Kuppens P, Burger H, Ceulemans E. Recurrence of depression can be foreseen by monitoring mental states with statistical process control. J Psychopathol Clin Sci 2023 Feb;132(2):145-155. [doi: 10.1037/abn0000812] [Medline: 36808958]
45. BipoSense respository. GitHub. URL: https://github.com/CarlBittendorf/BipoSense/tree/main/Unique%20Places [accessed 2026-06-01]

Abbreviations

ARL: average run length
BD: bipolar disorder
DBSCAN: Density-Based Spatial Clustering of Applications with Noise
***DSM-5*:** *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition)
EWMA: exponentially weighted moving average
HC: healthy control
OR: odds ratio
QIDS-SR16: Quick Inventory of Depressive Symptomatology–Self-Report, 16 items
SPC: statistical process control

XSL•FO
RenderX

Edited by L Buis; submitted 10.May.2025; peer-reviewed by E Snippe, G Anmella, X Liang; revised version received 15.Apr.2026; accepted 20.Apr.2026; published 22.Jun.2026.

Please cite as:
Guth M, Bittendorf C, Krug C, Ludwig VM, Muehlbauer E, Hartnagel LM, Severus E, Bauer M, Ritter P, Ebner-Priemer UW
Predicting Affective Episodes in Bipolar Disorder Using Statistical Process Control Analysis of GPS-Based Mobility Patterns: Quantitative Study
JMIR Mhealth Uhealth 2026;14:e77272
URL: https://mhealth.jmir.org/2026/1/e77272
doi:10.2196/77272

© Marvin Guth, Carl Bittendorf, Clemens Krug, Vera Miriam Ludwig, Esther Muehlbauer, Lisa-Marie Hartnagel, Emanuel Severus, Michael Bauer, Philipp Ritter, Ulrich W Ebner-Priemer. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 22.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Noninvasive Electrophysiological Biomarkers of Olfactory Responses Across Cognitive States in Alzheimer Dementia: Cross-Sectional Study

Juchan Ha[1], BSc; Junsoo Bok[1], BSc; Yunjin Lee[1], BSc; Yong Sung Kim[2], MD, MMSc; June Sic Kim[3], PhD; Hee-Jin Kim[1,2*], MD, PhD; Yongwoo Jang[1,4*], PhD

[1]Department of Medical and Digital Engineering, College of Engineering, Hanyang University, 222, Wangsimni-ro, Seongdong-gu, Seoul, Republic of Korea

[2]Department of Neurology, College of Medicine, Hanyang University, Seoul, Republic of Korea

[3]Clinical Research Institute, Konkuk University Medical Center, Seoul, Republic of Korea

[4]Department of Pharmacology, College of Medicine, Hanyang University, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Yongwoo Jang, PhD
Department of Medical and Digital Engineering, College of Engineering, Hanyang University, 222, Wangsimni-ro, Seongdong-gu, Seoul, Republic of Korea

Abstract

Background: Alzheimer disease (AD) is characterized by progressive cognitive decline, with olfactory dysfunction emerging among its earliest symptoms. Conventional olfactory tests rely on subjective self-reports and patient cooperation, limiting their clinical applicability. Noninvasive olfactory bulb (OB) recordings may provide objective physiological measures for early detection of cognitive decline.

Objective: This study investigated the relationship between olfactory dysfunction and cognitive decline across the AD spectrum and evaluated whether noninvasive OB recordings obtained via a wearable electroencephalography-based device could serve as digital biomarkers for early detection and home-based monitoring.

Methods: We recruited 71 participants, including individuals who were cognitively normal (CN; n=18), patients with mild cognitive impairment (MCI) subdivided into early MCI (n=30) and late MCI (n=9), and patients with dementia of the Alzheimer type (n=14). OB activity was recorded noninvasively during odor stimulation using a 6-channel wearable electroencephalography system. Behavioral olfactory function was assessed using a culturally adapted 6-item modified Brief Smell Identification Test and compared with electrophysiological responses. Time-frequency analyses identified disease-related components via nonparametric permutation testing (1000 iterations; $P<.05$). A support vector machine classifier was applied for group discrimination. Structural brain integrity was assessed through diffusion tensor imaging and magnetic resonance imaging, and cognitive performance was evaluated using comprehensive neuropsychological testing. Beta-band power (reflecting top-down cortical feedback) and gamma-band power (representing OB activity) were analyzed to characterize neural communication patterns.

Results: Beta-band power progressively declined from CN to dementia of the Alzheimer type, while gamma-band power decreased and response latency increased beginning in late MCI. Behavioral olfactory testing failed to differentiate CN from early MCI ($P=.77$), whereas electrophysiological recordings revealed significant differences ($P=.02$). The support vector machine classifier achieved 83.1% accuracy in distinguishing all groups based on olfactory spectral features. Reduced beta-band power showed a trend-level association with cingulum-hippocampal tract diffusivity ($R^2=0.28$; $P=.061$), whereas gamma-band power correlated with OB volume in later disease stages ($R^2=0.21$; $P<.001$). Electrophysiological features explained 91.7% of the variance in cognitive scores (Seoul Neuropsychological Screening Battery II; $P<.001$). The brief 5 to 10-minute testing protocol, requiring less than 2 minutes to fit the device, supports its feasibility for repeated, home-based assessments.

Conclusion: Bidirectional interactions between the OB and higher-order brain regions appear altered early in the AD continuum, with beta-band changes emerging first and gamma-band alterations occurring later. Digital OB recordings offer objective, noninvasive biomarkers for detecting cognitive decline before overt structural changes. Their portability and short duration make this approach ideal for longitudinal, home-based cognitive assessments within mobile health frameworks and for evaluating therapeutic efficacy in future interventional studies.

(JMIR Mhealth Uhealth 2026;14:e76245) doi:10.2196/76245

XSL•FO
RenderX

KEYWORDS

noninvasive olfactory recording; machine learning; digital biomarker; olfactory dysfunction; early diagnosis of dementia

Introduction

Continuous Monitoring for Early Alzheimer Disease Detection

Alzheimer disease (AD) represents a complex clinical challenge characterized by gradual and heterogeneous neurodegenerative progression over years to decades. While AD pathology involves amyloid and tau accumulation, its clinical manifestations span a continuum from individuals who were cognitively normal (CN) with AD pathology to mild cognitive impairment (MCI) and dementia. The subtle and variable trajectory of cognitive decline limits the sensitivity of traditional, episodic clinical assessments such as positron emission tomography (PET) imaging or cerebrospinal fluid analysis [1,2]. Continuous and home-based monitoring technologies offer a transformative approach to early disease detection by enabling real-time tracking of subclinical functional changes in naturalistic settings [3-6]. Quantitative physiological and behavioral biomarkers, such as gait velocity, sleep architecture, and circadian activity, can reveal gradual deterioration over weekly or monthly timescales that are often missed during infrequent clinical evaluations. Integrating these continuous monitoring systems into clinical workflows expands the therapeutic window for both pharmacological and nonpharmacological interventions, including targeted cognitive rehabilitation and personalized lifestyle management [7].

Olfactory dysfunction is one of the earliest symptoms of AD. It is recognized both as an early indicator and a potential clinical marker of the progression and severity of the disorder [8-11]. In addition to cognitive impairment, approximately 90% of patients with dementia of the Alzheimer type (DAT) experience a loss of olfactory abilities, such as difficulties with olfactory recognition and identification [12-16]. These deficits were previously attributed to pathological changes in the olfactory bulb (OB) and associated regions, including the entorhinal cortex, hippocampal cornu ammonis-1 region, and subiculum [17-19]. However, the timing of olfactory dysfunction during the progression of AD and its relationship with cognitive function remain unclear. Sniffing kits, such as the University of Pennsylvania Smell Identification Test (UPSIT), are frequently used to examine olfactory dysfunction in patients suspected of having MCI or DAT [19-22]. Although behavioral olfactory tests such as UPSIT and the Brief Smell Identification Test (B-SIT) have demonstrated feasibility in cognitively impaired populations, they fundamentally depend on participants' subjective reports of odor identification. This subjective nature introduces variability from individual interpretation, decision-making processes, and verbal or written communication abilities that may confound assessment of actual olfactory function [23-25]. Even with full participant cooperation, responses reflect a complex integration of sensory perception, cognitive processing, and motor output rather than isolated olfactory capability. In contrast, electrophysiological recordings from the OB capture objective neural signals during odor processing, providing quantifiable biomarkers in terms of latency, amplitude, and frequency components. Consequently, several studies have proposed using electroencephalography recordings during odor tests to reduce nonolfactory-related performance interference [26-30]. Structural magnetic resonance imaging (MRI) studies have also sought to differentiate MCI from DAT by evaluating volumetric decreases in the OB and primary olfactory cortex [31-34]. Therefore, olfactory evaluation using sniffing kits is likely to be undervalued in clinical assessment of AD despite its potential as an early indicator.

Objectives

To address these challenges, recent studies have proposed noninvasive and temporally precise recordings of the human OB, validated across a range of experiments, as a reliable method for assessing signals from the OB [35,36]. This approach has demonstrated olfactory dysfunction in patients with Parkinson disease (PD) compared with participants who were CN [37]. In this study, we conducted noninvasive digital recordings of olfactory responses among individuals who were CN, patients with MCI, and patients with DAT, using advanced digital signal processing and machine learning algorithms to transform raw electrophysiological data into quantifiable biomarkers. This computational approach allows the extraction of subtle features from neural signals that can objectively differentiate between cognitive states. We compared the results of various clinical examinations, including diffusion tensor imaging (DTI), MRI, and neuropsychological assessments. This study aimed to (1) identify differences in olfactory responses between CN and patient groups to extract olfactory characteristics critical for AD diagnosis, (2) explore the association between olfactory response characteristics and cognitive impairment in clinically diagnosed MCI and DAT using a hospital-based cohort, and (3) establish how the dynamics of olfactory responses across frequency bands relate to structural brain changes observed through medical imaging. Digital recording of odorant-mediated OB responses addresses the subjectivity inherent in behavioral olfactory assessments by providing objective, reproducible neural measurements. Unlike self-reported odor identification, electrophysiological signals directly reflect the functional integrity of olfactory neural circuits and can reveal temporal dynamics of olfactory processing not accessible through behavioral testing alone. The rapid advancement of mobile health (mHealth) and ubiquitous health technologies has created new opportunities for continuous, home-based assessment of neurological function. Our electrophysiological olfactory testing approach embodies these paradigms by using a lightweight, wearable 6-electrode electroencephalography device that can be quickly applied by users or caregivers. The entire testing procedure requires only 5 - 10 minutes, enabling regular self-assessment or caregiver-assisted monitoring in home settings. Unlike conventional clinic-based evaluations, this system allows individuals to track their olfactory and cognitive function through brief, repeatable sessions. Its portability and ease of use facilitate longitudinal monitoring, which is critical for detecting subtle, early changes in olfactory processing associated

XSL•FO
RenderX

with neurodegeneration. We hypothesized that digital recordings of OB responses could provide a more comprehensive understanding of cognitive decline in AD, serving as a potential digital biomarker.

Methods

Study Population

This study used a cohort design across 3 cohorts: CN, MCI, and DAT. The prospective registry study of olfactory response heterogeneity in AD progression enrolled participants from 2 tertiary medical centers in the Republic of Korea (Hanyang University Hospital and Myongji Hospital) between September 2023 and September 2024. The community-dwelling cohort was recruited from 3 geographically diverse provinces: Seoul, Gyeonggi, and North Chungcheong. A total of 129 individuals were initially screened for this study, of whom 71 were included in the final analysis, including 18 CN controls, 39 individuals with MCI, and 14 patients with DAT. All participants with DAT satisfied the diagnostic criteria of the National Institute on Aging and the Alzheimer's Association (NIA-AA) [25]. Comprehensive baseline assessments, including neuroimaging and neuropsychological evaluations, were conducted according to standardized protocols under physician supervision.

Eligibility and Diagnostic Criteria

The study defined eligibility criteria for participant inclusion, restricting the age range for enrollment to 50 - 85 years. Inclusion criteria required participants to have sufficient cognitive capacity to complete both neuropsychological and olfactory assessments. Participants with contraindications to neuroimaging were excluded. The diagnostic status was determined following the updated NIA-AA workgroup criteria [23]. Notably, MCI was diagnosed independently of the Seoul Neuropsychological Screening Battery II (SNSB-II) score using a comprehensive clinical evaluation. Diagnosis was based on documented cognitive decline corroborated by caregivers, objective episodic memory impairment on the delayed recall component of the Seoul Verbal Learning Test (performance below −1.5 SD, adjusted for age and education), a Clinical Dementia Rating score of 0.5, and fulfillment of the NIA-AA core clinical criteria for MCI due to AD. After the diagnosis of MCI was established, participants were further subdivided into early and late stages according to cognitive severity. The SNSB-II provides age-, sex-, and education-adjusted *z* scores across 5 cognitive domains (attention, language, visuospatial function, memory, and frontal and executive function), from which a composite *z* score was calculated by averaging domain scores. This composite score was used solely for severity stratification within the MCI group and was not involved in diagnostic determination. Participants with a composite *z* score ≥−1.0 were classified as early mild cognitive impairment (EMCI), whereas those with scores <−1.0 were classified as late MCI, in line with prior studies linking lower composite scores to broader multidomain impairment and increased risk of progression to dementia [38,39]. Using this criterion, 30 participants were classified as EMCI and 9 as late MCI. The relatively preserved composite scores observed in the EMCI group (mean 0.15, SD 0.58) reflect focal memory impairment that becomes attenuated when averaged across largely preserved cognitive domains, a pattern commonly observed in early-stage MCI. Clinical evaluation consisted of a comprehensive medical history review, neuroimaging analysis, and neuropsychological examination conducted by a qualified physician. Detailed diagnostic criteria are provided in Multimedia Appendix 1.

Neuropsychological Assessments

Neuropsychological assessments were performed using a standardized neuropsychological battery called the SNSB-II [38]. This battery includes examinations of attention, language, calculation ability, visuospatial skills, memory, frontal-executive function, and general cognition, such as the Mini-Mental State Examination (MMSE) [40] and Clinical Dementia Rating. The SNSB-II comprises the digit span test (forward and backward), the Korean version of the Boston Naming Test, Rey-Osterrieth Complex Figure Test, Seoul Verbal Learning Test, Phonemic Controlled Oral Word Association Test, and Korean Color Word Stroop Test. All SNSB-II variables were converted to *z* scores and standardized for age, sex, and years of education.

Amyloid PET and Structural Brain MRI

All participants underwent an 18F-florbetaben amyloid PET and structural MRI scan. Brain DTI was performed using a 3.0-T Achieva MRI system (Philips Medical Systems). DTI images were acquired in the transverse orientation using a pulse sequence with a single-shot spin-echo diffusion-sensitized echo-planar imaging sequence (15 gradient directions plus a B0 image with b value=800 s/mm², repetition time/echo time=13898/55 ms, field of view=224×224, matrix=112×112, number of averages=2, slice thickness=2.0 mm, and flip angle=90°). Amyloid positron emission tomography–computed tomography imaging was conducted following intravenous injection of 8 mCi of the amyloid radiotracer F-18 Florbetaben (Neuraceq; Piramal). Following a 90-minute uptake period, imaging was conducted for approximately 20 minutes using a dedicated PET/computed tomography (CT) scanner (Biograph 6; Siemens Medical Systems). The PET scans were reviewed by 2 expert PET readers and 1 neurologist based on predefined regional cortical tracer uptake and brain amyloid plaque load scoring systems. Four regions of interest, including the frontal, temporal, and parietal cortices along with the posterior cingulate and precuneus, were assessed in the visual analysis of the 18F-florbetaben PET scans. The regional cortical tracer uptake scores were condensed into a single 3-grade scoring system for each PET scan (brain amyloid plaque load score) as follows: 1 for no Amyloid-beta (Aβ) load, 2 for minor Aβ load, and 3 for significant Aβ load. Tract-based spatial statistics (TBSS) analysis was conducted to examine differences in DTI between patients with AD and healthy controls using the Functional Magnetic Resonance Imaging of the Brain Software Library (University of Oxford). The preprocessing pipeline included correction for head motion and eddy current distortions, followed by the computation of diffusion tensor metrics, specifically fractional anisotropy (FA), axial diffusivity, and radial diffusivity, to characterize the white matter fiber architecture. Subsequently, FA images were spatially normalized to a standardized brain template using the Functional Magnetic Resonance Imaging of the Brain nonlinear imaging registration

XSL•FO
RenderX

tool. A mean FA skeleton was generated using projection techniques to represent the central white matter pathways common to all participants, with an applied threshold criterion of 0.2. Individual FA maps were then projected onto this skeleton for group-level analyses. Regional white matter tract identification was facilitated using the Johns Hopkins University ICBM-DTI-81 white matter atlas [41]. Statistical analyses used permutation-based nonparametric inference (5000 permutations) to assess between-group differences [42]. The statistical threshold was established at $P<.05$, with correction for multiple comparisons implemented using threshold-free cluster enhancement. Volumetric quantification of the OB was performed by manual delineation of sequential T1-weighted images using ImageJ (version 1.53h; National Institutes of Health). Bilateral OB annotations were independently conducted by 2 experienced raters using FreeView (Massachusetts General Hospital), a FreeSurfer visualization tool, and ambiguous cases were resolved by consultation with a qualified neurosurgeon [43]. To minimize bias, the raters were blinded to participant group assignment, age, and sex. Individual OB volumes were calculated by summing the product of the surface area measurements and slice thickness, with the total OB volume derived from the mean of the bilateral measurements.

Behavioral Olfactory Testing

Olfactory function was assessed using a modified 6-item version adapted from the B-SIT, which itself is a 12-item subset of the 40-item UPSIT [44]. Participants were presented with a 4-category multiple-choice questionnaire and asked to identify which smells corresponded to the scent strip. As familiarity and responsiveness to certain odors may vary among ethnic groups, we selected 6 types of odors (cinnamon, smoke, chocolate, gasoline, soap, and onion) with a correct answer rate of 70% or more among healthy individuals from the 12 smells in B-SIT for this experiment. This culturally adapted version showed comparable diagnostic performance to the full B-SIT while avoiding misclassification due to odor unfamiliarity rather than true olfactory dysfunction.

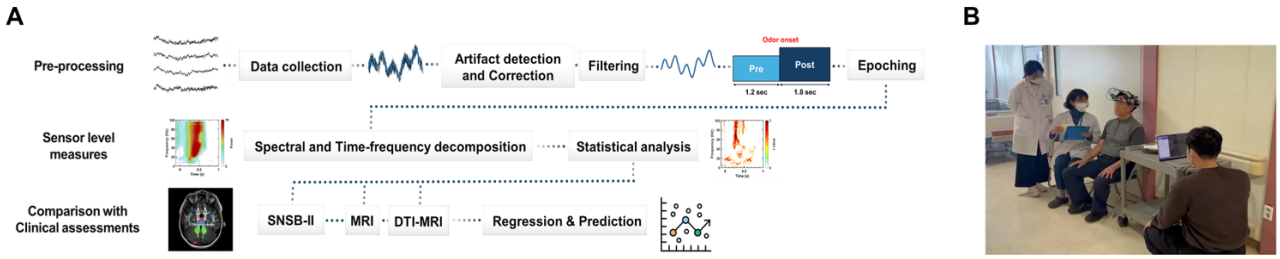
Digitalizing Olfactory Testing

The olfactory assessment protocol integrated the 6-item modified B-SIT, a validated and standardized instrument for evaluating olfactory function, with concurrent electrophysiological measurements to characterize the temporal dynamics of olfactory processing. Olfactory responses were recorded using a 6-channel active electrode electroencephalography system (Cyton; OpenBCI), and data were collected and digitized at a sampling rate of 250 Hz. In this study, the configuration of the OpenBCI electroencephalography cap was modified to optimize the detection of signals from the OB, following the approach proposed by Iravani et al [35]. Six electrodes were symmetrically placed on a mount along the centerline of the forehead, with each electrode spaced 5 cm apart to optimize the detection of OB signals. The helmet mount was designed to be adjustable, allowing the examiner to loosen it for easier electrode placement and subsequently tighten it to maintain firm and consistent contact with the participant's skin. This configuration ensured reliable signal detection from the OB, as illustrated in Figures 1A and 1B. The electroencephalography signals were recorded continuously throughout the test session.

During olfactory testing, a respiratory sensor monitored participants' breathing patterns to trigger olfactory stimulation. Recordings were segmented into epochs ranging from 1200 ms prestimulus to 1800 ms post stimulus, time locked to respiratory triggers, and rereferenced to the average of the left and right earlobe electrodes. Recorded signals were notch filtered at 60 Hz to eliminate power line noise. Additionally, muscle artifacts were detected by applying a bandpass filter (Butterworth filter, order 8) to the raw data, followed by a Hilbert transformation to extract the amplitude values. These values were standardized using z scores, with trials exceeding a z score of 6 assumed to be contaminated by muscle artifacts and excluded from further analysis. Considering the impact of eye movements and blinking on olfactory response signals, a lower z score threshold of 4 was used to improve the detection sensitivity, leading to the exclusion of trials with z scores exceeding this threshold. To further reduce blinking-related artifacts, participants were instructed to keep their eyes closed throughout the test. For signal processing, preprocessed and artifact-corrected olfactory signals were decomposed into time-frequency maps using a multitaper sliding window approach. This method was applied across a frequency range of 0.1 - 100 Hz, with a step size of 0.1 Hz, and across a time range of –1 to 1 second, with a step size of 5 ms. Power in each time-frequency bin was estimated using 2 tapers derived from discrete prolate spheroidal sequences. To capture at least 2 cycles of each frequency bin, the taper length was selected to ensure accurate power estimation across the spectrum. Wavelet transformation for each time bin was conducted using 2 sets of wavelet functions derived from discrete prolate spheroidal sequence tapers. This transformation involved convolving the wavelet functions with the olfactory response signal in the frequency domain by multiplying the fast Fourier coefficients of the signal by those of the wavelets. The frequency-smoothing parameter was set to 80% of the value in each frequency bin. To account for motor responses to sniffing, baseline-corrected odor trials were compared with baseline-corrected sniff trials to analyze phase-locked responses, which were subsequently converted into decibels. For regression and classification, features were extracted by identifying time-frequency regions with maximal group differences via permutation testing and calculating each participant's mean power within these regions. This process yielded 5 electrophysiological features in the beta and gamma bands for subsequent analyses. All preprocessing and time-frequency decompositions were performed using the open-source Fieldtrip toolbox in MATLAB (version R2023b; The MathWorks, Inc).

XSL•FO
RenderX

Figure 1. Experimental workflow for olfactory bulb signal analysis and clinical implementation. (A) Flow diagram illustrating the complete data processing pipeline: preprocessing of raw signals followed by data collection, artifact detection and correction, filtering, and epoching (with observation and rest periods of 3 seconds each). Secondary processing includes sensor-level measures with spectral and time-frequency decomposition, statistical analysis, and clinical comparisons using Seoul Neuropsychological Screening Battery II (SNSB-II), magnetic resonance imaging (MRI), and diffusion tensor imaging (DTI) with regression and prediction modeling. (B) Representative image of the experimental setup in a clinical environment, showing researchers administering the olfactory testing protocol to seated participants wearing the noninvasive olfactory bulb recording device.

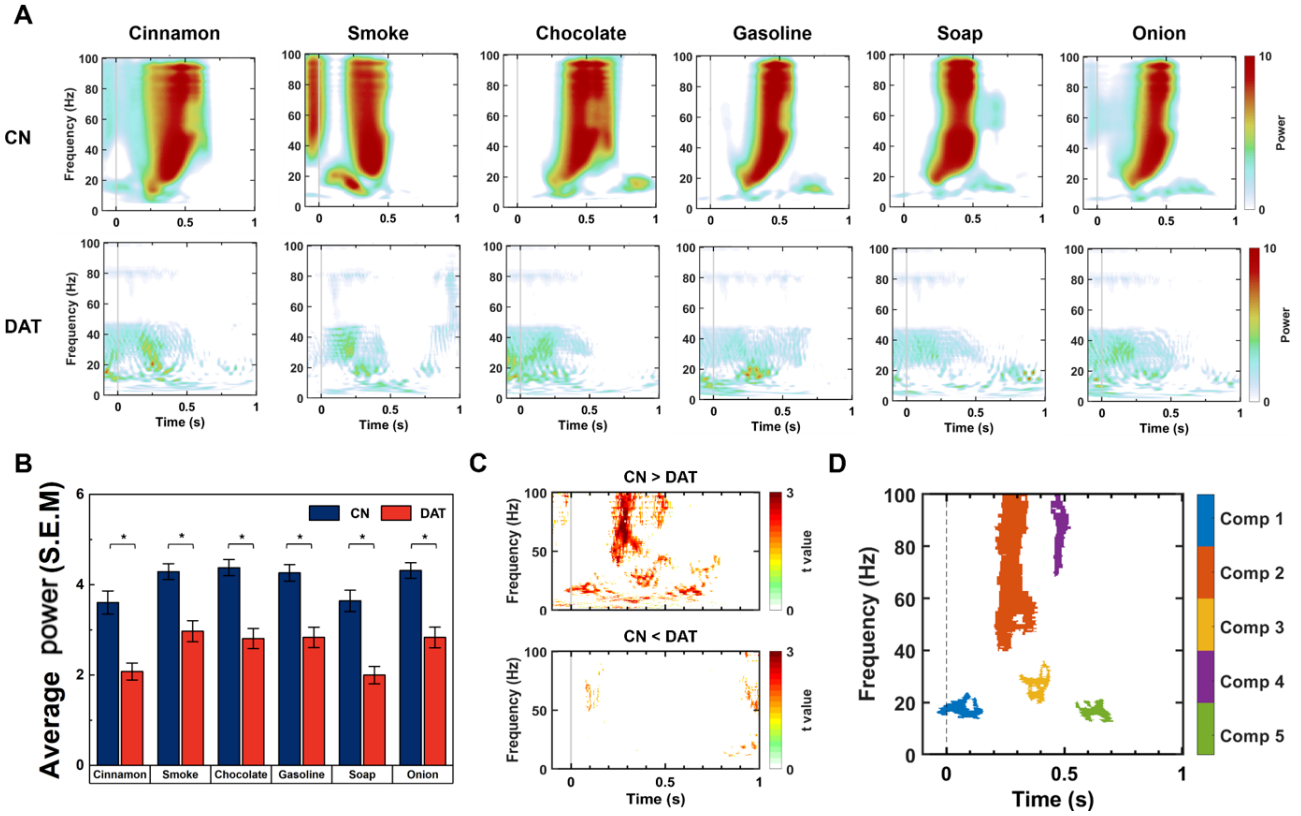


Statistical Analysis

All statistical analyses were conducted using MATLAB (version R2023b) with the Signal Processing and Fieldtrip toolboxes [45]. Spectral densities of the 6 electrodes were averaged across the channels. Initially, the average power comparison was used to assess the intensity of gamma synchronization (30 - 100 Hz) in the olfactory response one second after olfactory stimulation. Time-frequency maps of the olfactory responses between the 2 comparable groups were assessed using a nonparametric Monte Carlo permutation test. A total of 1000 permutation tests were conducted to identify significant clusters that differed between the 2 groups within the time-frequency maps, with significance defined as $P<.05$ and a cluster size exceeding 300 bins. This cluster-based permutation approach inherently controls for multiple comparisons across all time-frequency bins by evaluating cluster-level statistics, thereby maintaining the family-wise error rate without requiring additional correction methods. The identified clusters were labeled according to their latencies (for example, components 1 and 2; Figure 2). Machine learning classification was performed using a support vector machine with a radial basis function kernel. Model development involved leave-one-out cross-validation to maximize sample usage given the limited dataset. The 5 electrophysiological components were used as input features. Classification performance was evaluated using accuracy, precision, recall, and F_1-score metrics. Regression and classification analyses used only the 5 electrophysiological components extracted from the time-frequency analysis as input features, without inclusion of demographic covariates. All *t* tests were 2-tailed. Statistical significance was indicated by asterisks, with * representing $P<.05$, ** denoting $P<.01$, and *** signifying $P<.001$.

XSL•FO
RenderX

Figure 2. Odorant-mediated olfactory responses differentiate patients with dementia of the Alzheimer type (DAT) from cognitively normal (CN) controls. (A) Odor-mediated olfactory responses in CN controls and patients with DAT. Gamma synchronization was observed in CN controls following odor onset (indicated by the vertical gray line at time 0). In contrast, no synchronization was detected in the DAT group during the same period. (B) Averaged power change for odorant across the 0-1 second interval, presented with SE of the mean (SEM) for responses to the 6 selected odorants. Statistical significance is indicated by * for $P<.05$. (C) T-map derived from 1000 Monte Carlo permutation tests, illustrating significant differences in the beta (13 - 30 Hz) and gamma (30-100 Hz) bands between the CN and DAT groups. Threshold t-maps indicate areas with higher power in CN controls compared to DAT at $P<.05$. (D) Threshold t-maps indicate significant clusters ($P<.05$; cluster size >300) that differentiate olfactory response components between the CN and DAT groups. Five distinct components were identified in the gamma and beta bands during early and late time points, with each component represented in a specific color. Color labels can be found in the color bar on the right side of the panel.



Ethical Considerations

This study was approved by the Institutional Review Board of Hanyang University Medical Center (2023-06-019) and conducted in accordance with the Declaration of Helsinki. All participants or their legally authorized representatives provided written informed consent and were informed of their right to withdraw at any time without penalty. All participant data were deidentified and assigned unique study codes, with the linking key stored separately in a password-protected file accessible only to the principal investigator. Neuroimaging data were anonymized using defacing algorithms to remove identifiable facial features. No financial compensation was provided to participants. However, participants received their complete diagnostic test results and medical consultations free of charge, which were disclosed as benefits of participation during the informed consent process.

Results

Overview

Digital olfactory testing was conducted alongside several traditional methods for AD diagnosis, including brain imaging techniques such as MRI and DTI, as well as neurobehavioral assessments such as the SNSB-II and MMSE. To evaluate olfactory ability, we conducted a 6-item modified B-SIT while simultaneously recording the electrical activity of the OB across the CN, MCI, andDAT groups. Demographic characteristics, cognitive status, and medication profiles of the participants are summarized in Table 1. Six electrodes were attached to positions capable of measuring OB responses for digital olfactory testing (Figure 3). The obtained signals were subjected to preprocessing and spectral analysis, resulting in a spectrogram. The spectral density of the signal was time-locked to stimulus onset and averaged across the 6 electrodes and trials to optimize the signal-to-noise ratio. The time-frequency map for the odor trials was determined within the specified time and frequency bands. To exclude contamination from sniffing and other motor-related artifacts, we compared the air and odor conditions. Odor event-related synchronization was observed in the beta (13 - 30 Hz) and gamma (30 - 100 Hz) bands at approximately 200 - 550 ms post stimulus. Subsequently, specific areas showing significant differences between the CN and DAT groups were identified using nonparametric tests with 1000 permutations, resulting in the isolation of 5 distinct components. The 5 extracted components were analyzed for correlations with existing medical imaging results, such as MRI and DTI findings, as well as neurobehavioral test outcomes (MMSE and SNSB-II), to validate their effectiveness. This process was investigated to determine whether these olfactory characteristics could serve

XSL•FO
RenderX

as early diagnostic markers of AD. An illustration of this process is shown in Figure 1.

Table . Demographic and clinical characteristics according to diagnostic group. Values are presented as mean (SD) or n (%). *P* values are from post hoc pairwise comparisons with the cognitively normal (CN) group using Bonferroni correction. Overall group differences (one-way ANOVA or chi-square test) were as follows: age ($F_{3,67}$=3.21; *P*=.03), sex (χ^2_3=3.84; *P*=.28), Brief Smell Identification Test ($F_{3,67}$=8.45; *P*<.001), 6-item modified Brief Smell Identification Test ($F_{3,67}$=9.72; *P*<.001), education ($F_{3,67}$=2.89; *P*=.04), Mini-Mental State Examination ($F_{3,67}$=15.32; *P*<.001), Seoul Neuropsychological Screening Battery II ($F_{3,67}$=12.84; *P*<.001), Clinical Dementia Rating (Kruskal-Wallis H test statistic [H_3]=21.45; *P*<.001), Clinical Dementia Rating–Sum of Boxes (H_3=28.93; *P*<.001), brain amyloid plaque load ($F_{3,67}$=6.78; *P*<.001), and APOE ε4 (χ^2_3=7.92; *P*=.05).

Characteristic	CN[a] (n=18)	EMCI[b] (n=30)	Late MCI (n=9)	DAT[c] (n=14)	*P* value vs CN (EMCI/LMCI[d]/DAT)
Age (years), mean (SD)	72.17 (6.37)	70.87 (6.97)	70.67 (6.88)	76.79 (4.14)	.65/.62/.02
Sex (female), n (%)	15 (83)	20 (66)	6 (66)	8 (57)	.18/.27/.09
B-SIT[e] score, mean (SD)	7.94 (2.09)	8.27 (2.03)	5.78 (2.90)	4.79 (2.37)	.54/.02/<.001
6-SIT[f], mean (SD)	4.89 (0.99)	4.8 (1.24)	3.33 (1.83)	2.71 (1.91)	.77/.02/<.001
Education years, mean (SD)	10.06 (5.16)	9.63 (5.34)	11.22 (3.58)	6.07 (4.95)	.76/.58/.02
MMSE[g], mean (SD)	27.78 (1.96)	26.67 (2.94)	23.78 (2.70)	19.79 (3.30)	.14/<.001/<.001
SNSB-II[h], mean (SD)	54.38 (5.75)	51.33 (7.53)	44.62 (5.29)	32.76 (5.01)	.12/.004/<.001
CDR[i], mean (SD)	0.42 (0.19)	0.43 (0.17)	0.5 (0.00)	0.93 (0.37)	.86/.08/<.001
CDR-SB[j], mean (SD)	0.69 (0.50)	0.90 (0.92)	2.28 (1.08)	5.61 (2.04)	.37/.003/<.001
BAPL[k], mean (SD)	1.13 (0.85)	1.77 (0.88)	2.22 (0.92)	2.64 (0.72)	.06/.03/.002
APOE4[l] ε4 carrier, n (%)	8 (44)	12 (40)	7 (77)	10 (71)	.76/.04/.03

[a]CN: cognitively normal.

[b]EMCI: early mild cognitive impairment.

[c]DAT: dementia of the Alzheimer type.

[d]LMCI: late mild cognitive impairment.

[e]B-SIT: Brief Smell Identification Test.

[f]6-SIT: 6-item modified Smell Identification Test.

[g]MMSE: Mini-Mental State Examination.

[h]SNSB-II: Seoul Neuropsychological Screening Battery II.

[i]CDR: Clinical Dementia Rating.

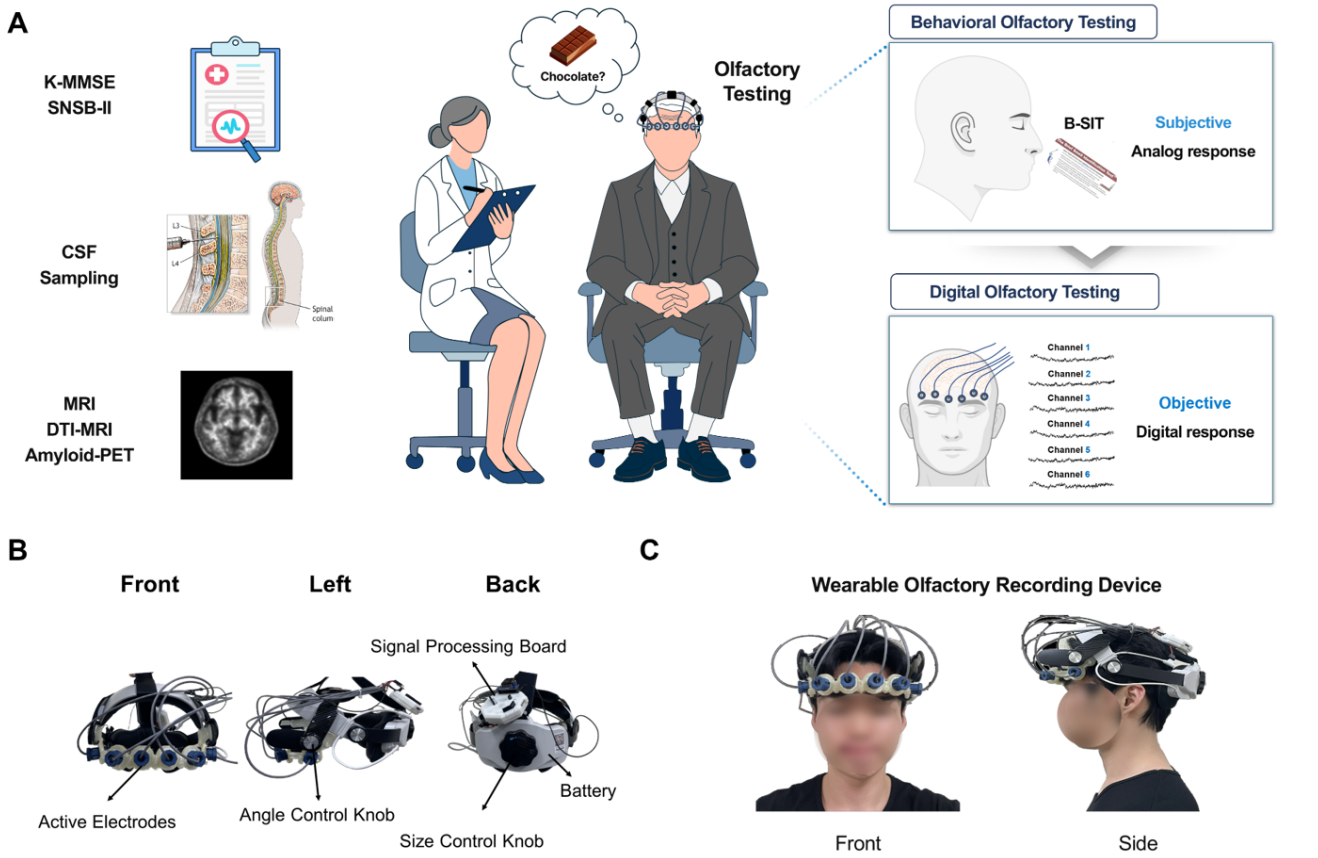
[j]SB: sum of boxes.

[k]BAPL: brain amyloid plaque load.

[l]APOE: apolipoprotein E.

XSL•FO
RenderX

Figure 3. Overview of the methodological procedure for digital olfactory testing aimed at the early diagnosis of Alzheimer disease. (B) An illustration of the 6-electrode wearable olfactory response measurement appliance shown in three views (front, left, and back). (C) Noninvasive olfactory bulb signal measurement device in position. Front view showing the primary sensor array positioned over the nasal bridge with integrated electrodes for detecting olfactory neural activity. The device enables real-time acquisition of olfactory bulb signals without surgical intervention, facilitating neurophysiological monitoring. B-SIT: Brief Smell Identification Test; CSF: cerebrospinal fluid; DTI-MRI: diffusion tensor imaging–magnetic resonance imaging; K-MMSE: Korean Mini-Mental State Examination; MRI: magnetic resonance imaging; PET: positron emission tomography; SNSB-II: Seoul Neuropsychological Screening Battery II.



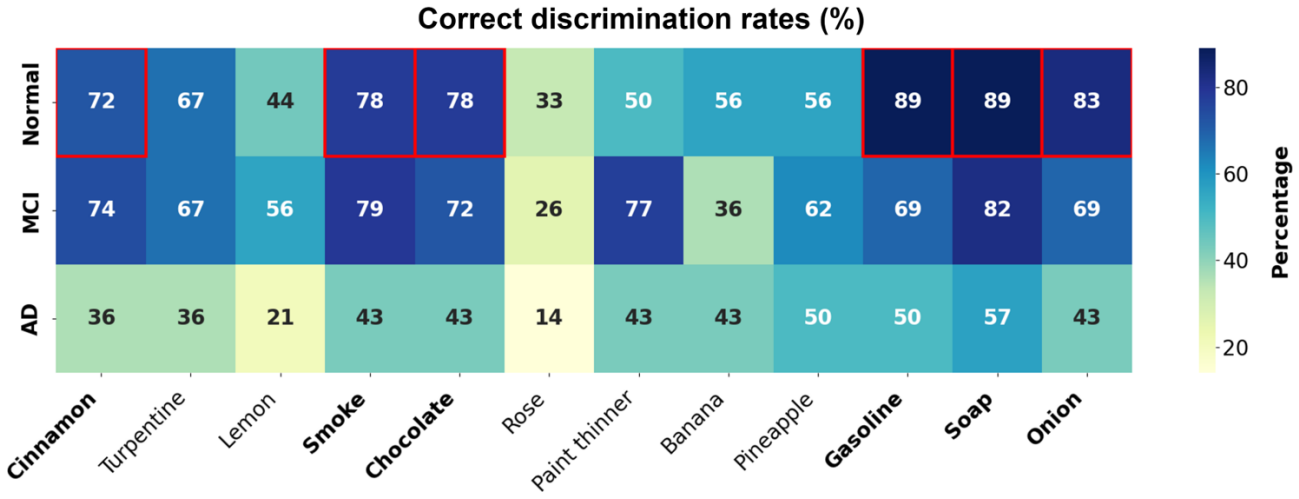
Component Identification Through CN-AD Contrast Analysis

Patients with DAT reported olfactory impairment when assessed using the B-SIT [20,22,25], which included 12 odors selected from the UPSIT [44]. As familiarity and responsiveness to certain odors may vary among ethnic groups, we first selected 6 types of odors (cinnamon, smoke, chocolate, gasoline, soap, and onion) with a correct answer rate of 70% or more among healthy individuals from the 12 smells in the B-SIT (Figure 4). Behavioral olfactory testing failed to distinguish between the CN and MCI groups. Discrimination rates were comparable across all 6 odors tested, with both groups showing relatively preserved identification performance (mean correct: CN 81.5%, SD 16.5% vs MCI 74.2%, SD 25.1%; *P*=.14). Only the clinical dementia group showed significant impairment (mean 45.3%, SD 31.8%; *P*<.001 vs both CN and MCI). Figure 2A shows the spectrum of olfactory responses to the 6 selected odors in healthy individuals and patients with DAT. In the CN group, odor-responsive signals were observed within 0.6 seconds after odor exposure, specifically in the beta and gamma frequency bands across 6 electrodes. Conversely, patients with DAT showed no detectable gamma synchronization at these temporal and frequency intervals, thereby confirming the ability of the control group to identify robust olfactory signals. As shown in Figure 2B, the average power recorded in the response of patients with DAT to the selected odors was significantly lower than that in the CN group for cinnamon (*P*=.02), smoke (*P*=.04), chocolate (*P*=.01), gasoline (*P*=.03), soap (*P*=.01), and onions (*P*=.02). This uniform impairment across chemically diverse odorants indicates a central olfactory processing deficit rather than peripheral receptor-specific dysfunction. Figure 2C shows the statistically significant differences in odor responses observed between the CN and DAT groups. To identify specific regions between the 2 groups, we performed nonparametric tests with 1000 permutations. The results are indicated by threshold t-maps, highlighting areas with *P*<.05. The upper panel demonstrates that the CN group showed enhanced power in the beta and gamma bands in response to the selected odors compared to the DAT group (CN > DAT). Notably, no significant difference in power was observed in patients with DAT when compared to the CN group (CN < DAT). This analysis revealed significant increases in the beta and gamma band power in the CN group compared to the DAT group at approximately 250 ms postodor onset (t_{30}=2.1192; *P*=.02). Conversely, a decline in gamma frequency power was observed at approximately 950 ms and 980 ms in the CN group compared with the DAT group (t_{30}=−1.8412; *P*=.04). Significant clusters (*P*<.05; cluster size > 300) were identified and sequentially labeled based on their temporal order as component 1,

XSL•FO
RenderX

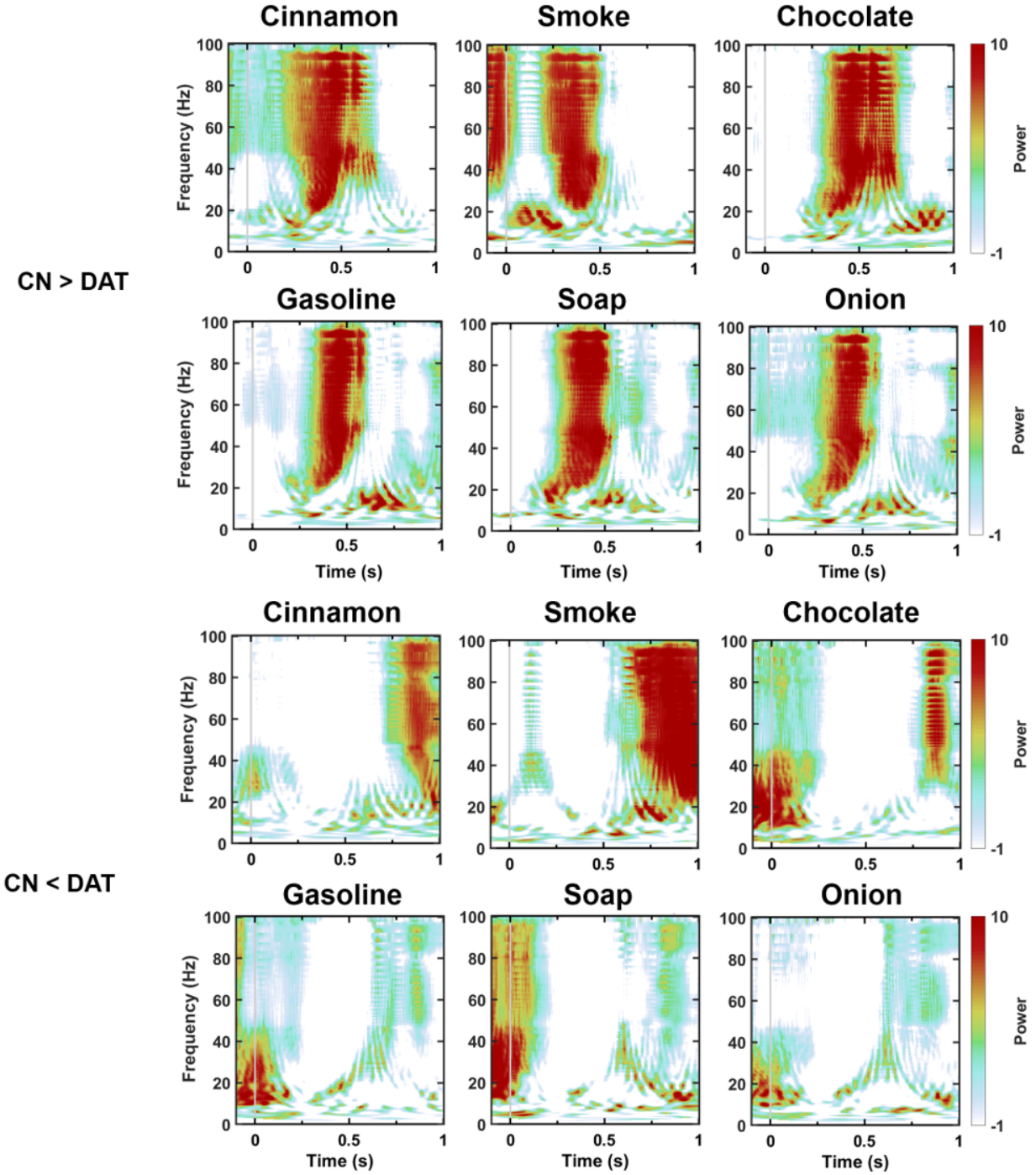
component 2, and so on. As shown in Figure 2D, 5 distinct components were identified, each of which was visually represented by a different color. The statistical parameters for each component were as follows: component 1 (t_{30}=1.9934; P=.03), component 2 (t_{30}=2.1823; P=.03), component 3 (t_{30}=1.9938; P=.03), component 4 (t_{30}=1.9182; P=.04), and component 5 (t_{30}=2.1393; P=.02). Additionally, Figure 5 provides the differences in the time-frequency maps between the CN and DAT groups in response to each of the 6 odorants.

Figure 4. Rates of correct answers for each Brief Smell Identification Test odor identification question, categorized into 3 patient groups. The shading of each bin indicates the rates of correct answers, with darker colors representing higher accuracy and lighter colors indicating higher error rates. Odors that achieved over 70% correct answers in the cognitively normal group are highlighted with a red border and marked in bold in the labels below. CN: cognitively normal; DAT: dementia of the Alzheimer type; MCI: mild cognitive impairment.



XSL•FO
RenderX

Figure 5. Time-frequency representation of odor-mediated olfactory responses for each selected odorant, shown as a 3 × 2 grid of the six odorants (cinnamon, smoke, and chocolate in the top row; gasoline, soap, and onion in the bottom row). The upper block (CN > DAT) indicates higher power for the cognitively normal (CN) group relative to the dementia of Alzheimer type (DAT) group, and the lower block (CN < DAT) indicates lower power for the CN group relative to the DAT group. Gamma synchronization is observed in the normal controls after odor onset, as indicated by the vertical gray line at time 0.



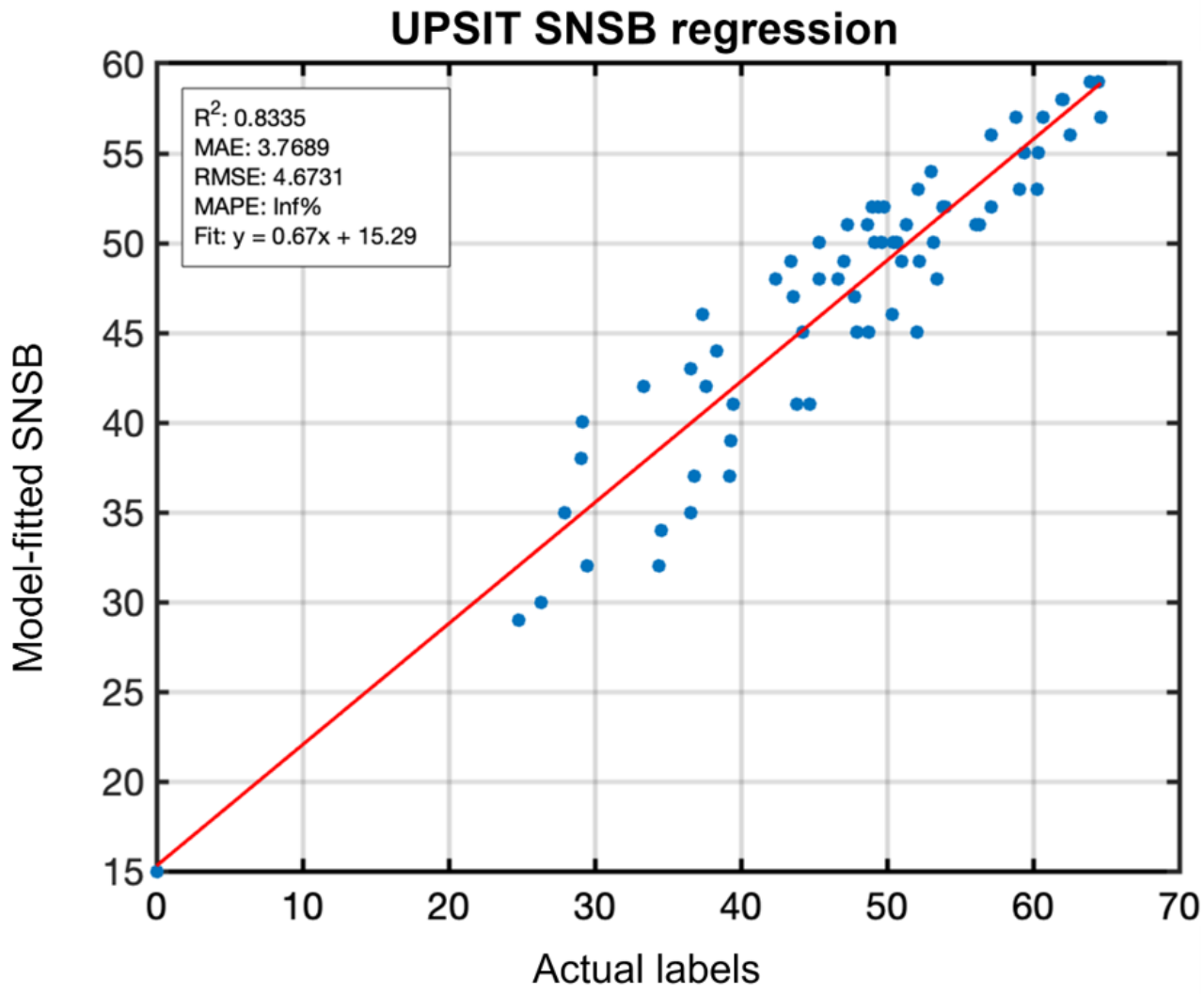
Stage-Dependent Spectral Alterations and Clinical Associations

For the loss of smell to be used as an early diagnostic method for AD, the detection of early signs is essential during the MCI stage, which precedes dementia. However, existing approaches do not clearly differentiate between individuals who were CN and those with MCI. Therefore, we conducted an experiment to determine whether digitalized measurements of olfactory function could effectively classify MCI severity. To examine the progression of olfactory dysfunction across the cognitive spectrum, the MCI group was stratified according to cognitive severity. Using the SNSB-II composite *z* score with a cutoff of –1.0 (adjusted for age and education), 30 participants were identified as EMCI (mean *z* score 0.15, SD 0.58) and 9 as late MCI (mean *z* score –1.65, SD 0.42) [38,39]. This score reflects multiple cognitive domains, including language, attention, visuospatial skills, memory, and frontal and executive function. Neuropsychological evaluations and relationships between the groups are summarized in Table 1 and Figure 6. First, we generated a differential time-frequency spectrum comparing EMCI to late MCI (EMCI > late MCI) and vice versa (EMCI

XSL•FO
RenderX

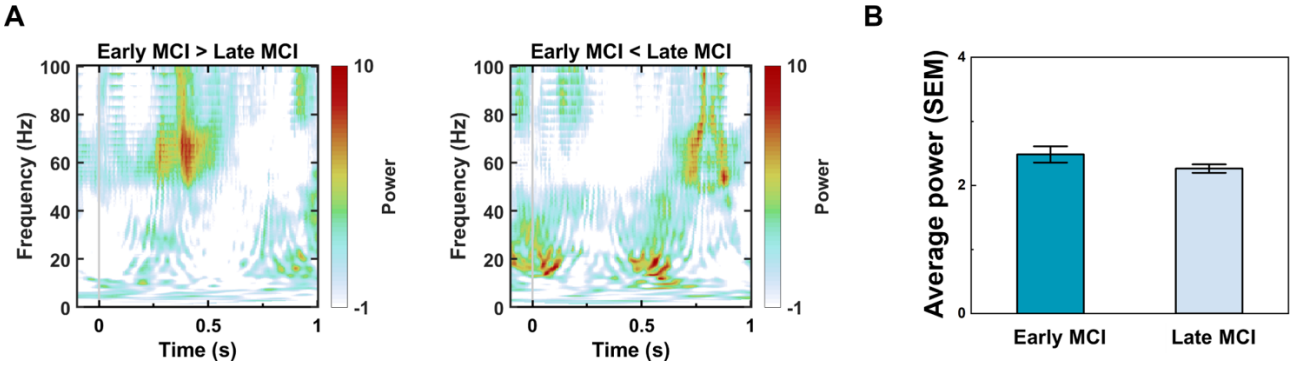
< late LMCI). No significant differences in the odor-induced gamma band power were observed among the groups. Notably, the EMCI group exhibited faster responses within 490 ms of odor onset across the 6 electrodes. In contrast, the late MCI group showed delayed gamma synchronization at approximately 850 ms post stimulus after controlling for sniffing and evoked responses (Figure 7A). However, no difference in the average gamma-band power was found between the groups (Figure 7B). Furthermore, we analyzed significant increases in the average power spectrum across the CN, EMCI, late MCI, and DAT groups. As shown in Figure 8A, the gamma band power spectrum showed a significant reduction and delayed response time starting from the late MCI stage compared with individuals who were CN. Additionally, the beta band power spectrum progressively decreased as the condition advanced from CN to EMCI, LMCI, and DAT stages. These results were consistent for each of the 6 selected odors, as shown in Figures 9A and 9B. In this regard, the structural similarity index matrix (SSIM) analysis of the olfactory response spectrograms revealed high similarity between the control and EMCI groups (SSIM=0.43) and between the late MCI and DAT groups (SSIM=0.61), as provided in Figure 10. Olfactory characteristics were derived by computing the mean values across the temporal stages for each component. Components 1, 3, and 5 were associated with the beta band, whereas components 2 and 4 were associated with the gamma band (Figure 2D). The spectral analysis in Figure 8B revealed significant decreases in the beta band power from EMCI (P<.001) to late mild cognitive impairment (LMCI; P=.03), culminating in DAT (P=.009). The gamma band power demonstrated a distinct temporal progression, with significant reductions emerging at LMCI (P=.01) and progressing through DAT (P=.001), while showing marginal alterations at EMCI (P=.08).

Figure 6. Plot of model-fitted Seoul Neuropsychological Screening Battery II (SNSB-II) scores with University of Pennsylvania Smell Identification Test (UPSIT) vs actual SNSB-II scores using the multiple linear regression model, with the selected 6 UPSIT scores as the independent variables and the SNSB-II score as the dependent variable. The mean absolute error (MAE), root mean square error (RMSE), and model fit are expressed as the adjusted R^2 of the model.



XSL•FO
RenderX

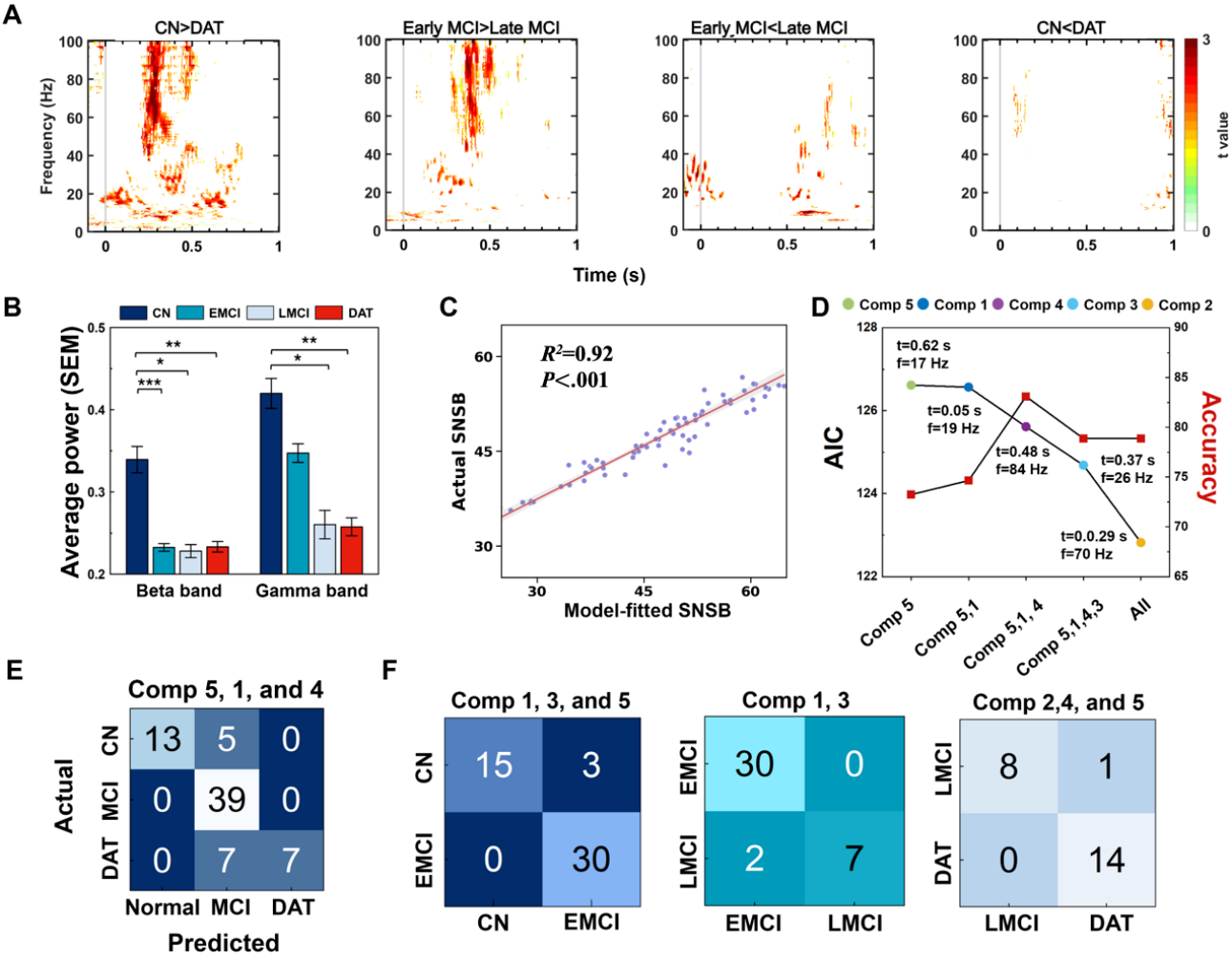
Figure 7. Averaged time-frequency map of 6 odorant-mediated olfactory responses, demonstrating higher power in the early mild cognitive impairment (MCI) group compared to the late MCI group, and averaged odorant-mediated olfactory responses showing reduced power in the late MCI group compared to the early MCI group. Warmer colors in the time-frequency response represent higher power values, as shown in panel A. Averaged power change for odorants across the 0-1 second interval, presented with SE of the mean (SEM) for responses to the 6 selected odorants, as shown in panel B. Statistical significance is indicated by * for *P*<.05.



To evaluate clinical applicability, the extracted spectral characteristics, categorized into beta and gamma frequency bands, were used to examine their association with cognitive performance. As shown in Figure 8C, multiple linear regression analysis demonstrated that electrophysiological features explained 91.69% of the variance in SNSB-II *z* scores (*P*<.001). This remarkably strong association between olfactory biomarkers and cognitive function was consistent across the entire cognitive spectrum from CN to DAT. The extremely low *P* value indicates that this relationship is highly unlikely to have occurred by chance. Additionally, MMSE scores showed similarly strong associations with olfactory responses (R^2=0.94; Figure 10). These results demonstrate that the magnitude of electrophysiological responses in beta and gamma bands closely tracks with the severity of cognitive impairment, supporting the potential of these biomarkers for objective cognitive assessment.

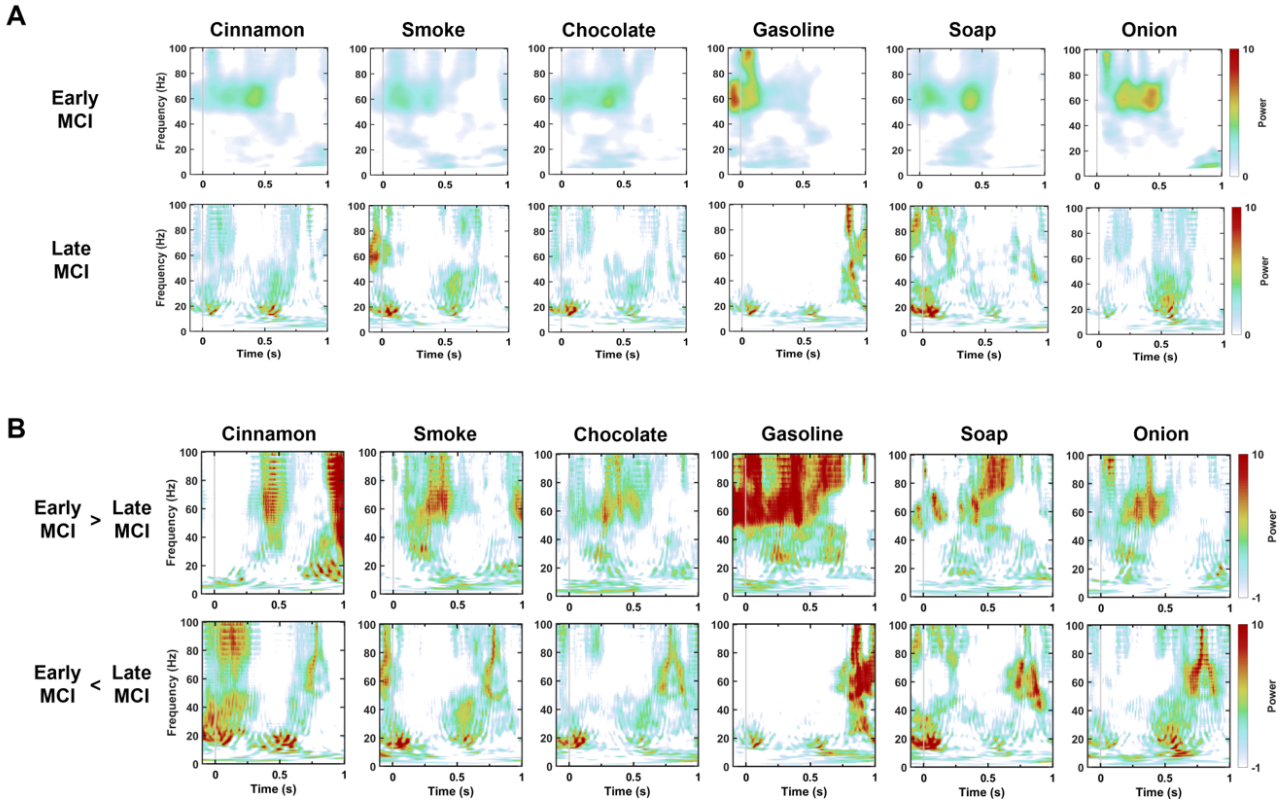
XSL•FO
RenderX

Figure 8. Olfactory response components as a digitalized diagnosis tool. (A) Threshold t-maps derived from 1000 Monte Carlo permutation tests at $P<.05$, indicating a gradual decrease in gamma synchronization and delayed reaction time from individuals who were cognitively normal (CN) to patients with dementia of the Alzheimer type (DAT), significant in the beta and gamma bands. (B) Averaged power change for odorants from each olfactory component, presented with SE of the mean (SEM) for responses to the 6 selected odorants (components 1, 3, and 5 exhibited the beta band, whereas components 2 and 4 belonged to the gamma band). Statistical significance was indicated using asterisks, with * representing $P<.05$, ** denoting $P<.01$, and *** signifying $P<.001$. (C) Plot of the predicted Seoul Neuropsychological Screening Battery II (SNSB-II) scores vs actual SNSB-II scores using the multiple linear regression model, regressing the extracted key features from the olfactory response components as dependent variables. The model fit is expressed as the adjusted R^2 of the model and the Pearson correlation between the 2 variables, indicating the proportion of variance in SNSB-II scores explained by the model parameters. (D) Akaike information criteria (AIC) for each olfactory response component. Accuracy of models with stepwise adding of components shows maximum accuracy for the model, including components 5, 1, and 4. (E) Confusion matrix of the support vector machine classifier model: out of 18 individuals in the CN group, 13 were classified correctly as CN, whereas 5 were misclassified as mild cognitive impairment (MCI). The classification was accurate for the MCI groups. Among 14 individuals in the DAT group, 7 were misclassified as MCI. (F) Cross-tables distinguishing adjacent cognitive stages (cognitively normal vs early mild cognitive impairment, early vs late mild cognitive impairment, and late mild cognitive impairment vs dementia of the Alzheimer type) show true-positive, false-positive, true-negative, and false-negative values. Comp: component; EMCI: early mild cognitive impairment; LMCI: late mild cognitive impairment.



XSL•FO
RenderX

Figure 9. Time-frequency representation of odor-mediated olfactory responses in early and late mild cognitive impairment (MCI) for each of the 6 selected odorants (cinnamon, smoke, chocolate, gasoline, soap, and onion), arranged in columns. (A) Absolute time-frequency responses for the early MCI group (top row) and the late MCI group (bottom row). Early gamma synchronization was observed in the early MCI group shortly after odor onset (indicated by the vertical gray line at time 0), whereas synchronization was delayed in the late MCI group. (B) Between-group contrasts: the top row (early MCI > late MCI) indicates higher power in the early MCI group, and the bottom row (early MCI < late MCI) indicates higher power in the late MCI group. Comparably high gamma synchronization was observed in both MCI groups after odor onset, in contrast to the markedly reduced responses previously shown for the dementia of the Alzheimer type group.



Model selection was conducted using the Akaike information criterion to assess the accuracy of fit for each olfactory component based on the in-sample prediction error. In classifying the CN, MCI, and DAT groups, component 2 showed the optimal performance, with the lowest Akaike information criterion value of 122.82, followed by component 3 (124.69), component 4 (125.61), component 1 (126.57), and component 5 (126.62). Subsequently, we used a stepwise component addition approach within the regression framework to optimize the prediction accuracy. The analysis revealed a peak performance with a 3-component model that included components 5, 1, and 4, as shown in Figure 8D. We implemented a support vector machine regressor that integrates component-specific olfactory characteristics to assess the diagnostic utility of digitalized olfactory response recordings. Model performance was evaluated using leave-one-out cross-validation, where each participant was held out as a test sample. The resulting classification model demonstrated exceptional discriminative capability across the 3 groups, achieving 83.09% accuracy, 88.64% precision, perfect recall (100%), and an F_1-score of 93.98% (Figure 8E). These metrics underscore the robust diagnostic potential of the proposed multicomponent approach. Therefore, the strong predictive relationship between odorant-mediated olfactory responses and cognitive performance underscores the potential clinical utility of this method for early diagnosis of AD. Further analyses were performed to evaluate the component-specific contributions in distinguishing between adjacent cognitive stages (CN-EMCI, EMCI-LMCI, and LMCI-DAT) to identify the key components for the early detection of AD. The discrimination between CN and EMCI states was achieved with remarkable precision using components 1, 3, and 5, all representing beta band activity, resulting in 93.75% accuracy, 90.91% precision, and 100% recall (Figure 8F). Notably, the predominant use of beta band components rather than gamma band components in the pre-DAT stage classification aligned with the findings presented in Figure 8B. The gamma band components were critical in differentiating patients with DAT from the LMCI group, marking a distinct transition point in disease progression. Table 2 summarizes the classification performance across different diagnostic comparisons.

XSL•FO
RenderX

Figure 10. Structural similarity index matrix (SSIM) of olfactory response spectrograms across the cognitively normal (CN), early mild cognitive impairment (MCI), late MCI, and dementia of the Alzheimer type (DAT) groups. Higher values indicate greater similarity between spectrograms; high similarity was observed between the CN and early MCI groups (SSIM=0.43) and between the late MCI and DAT groups (SSIM=0.61), as shown in panel A. Plot of predicted vs actual Mini-Mental State Examination (MMSE) scores using the multiple linear regression model, with the extracted key features from the olfactory responses as the independent variables and the MMSE score as the dependent variable, as shown in panel B. The model fit is expressed as the adjusted R^2 and the Pearson correlation between the 2 variables, indicating the proportion of variance in MMSE scores explained by the model parameters. The shaded band denotes the 95% CI of the regression line.

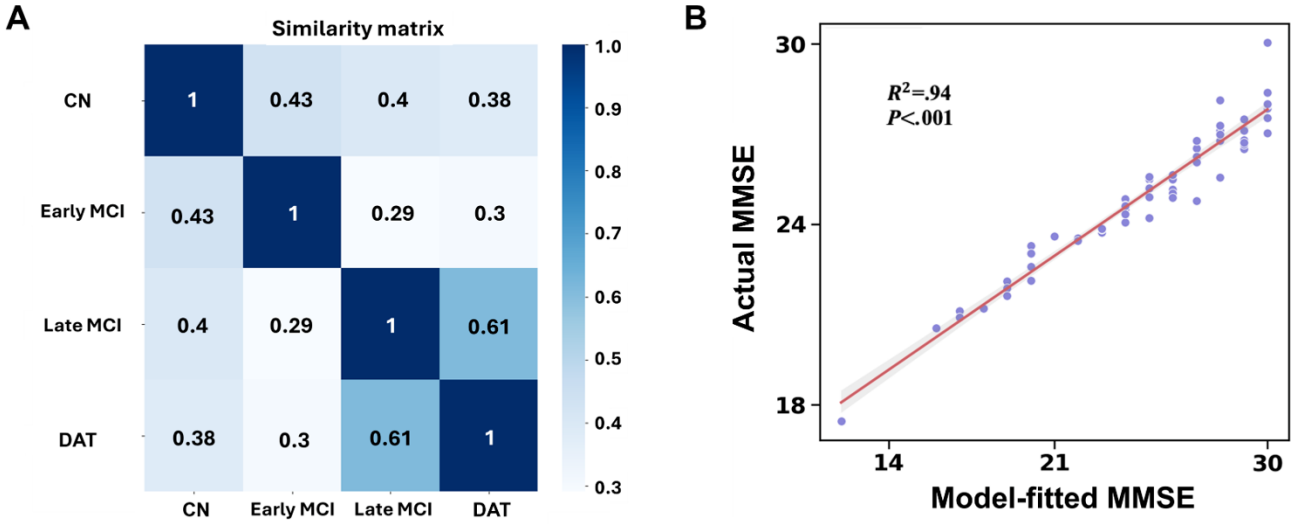


Table . Classification performance of support vector machine models using olfactory electrophysiological features. Performance metrics for support vector machine classification across different diagnostic group comparisons. Models were trained using electrophysiological components extracted from time-frequency analysis of olfactory responses. Leave-one-out cross-validation was performed for the 3-way classification. Binary classifications between adjacent cognitive stages reveal differential importance of frequency bands: beta-band components (1, 3, and 5) were most effective for distinguishing cognitively normal from early MCI, whereas gamma-band components (2 and 4) became crucial for late-stage discrimination.

Classification task	Accuracy (%)	Precision (%)	Recall (%)	F_1-score (%)	Features
3-way classification (CN[a], MCI[b], and DAT[c])	83.09	88.64	100	93.98	5,1,4
CN vs EMCI[d]	93.75	90.91	100	95.24	1,3,5
EMCI vs LMCI[e]	87.18	85.71	92.31	88.89	1,3
LMCI vs DAT	82.61	88.89	88.89	88.89	2,4,5

[a]CN: cognitively normal.

[b]MCI: mild cognitive impairment.

[c]DAT: dementia of the Alzheimer type.

[d]EMCI: early mild cognitive impairment.

[e]LMCI: late mild cognitive impairment.

Structural Correlations of Bidirectional Spectral Components

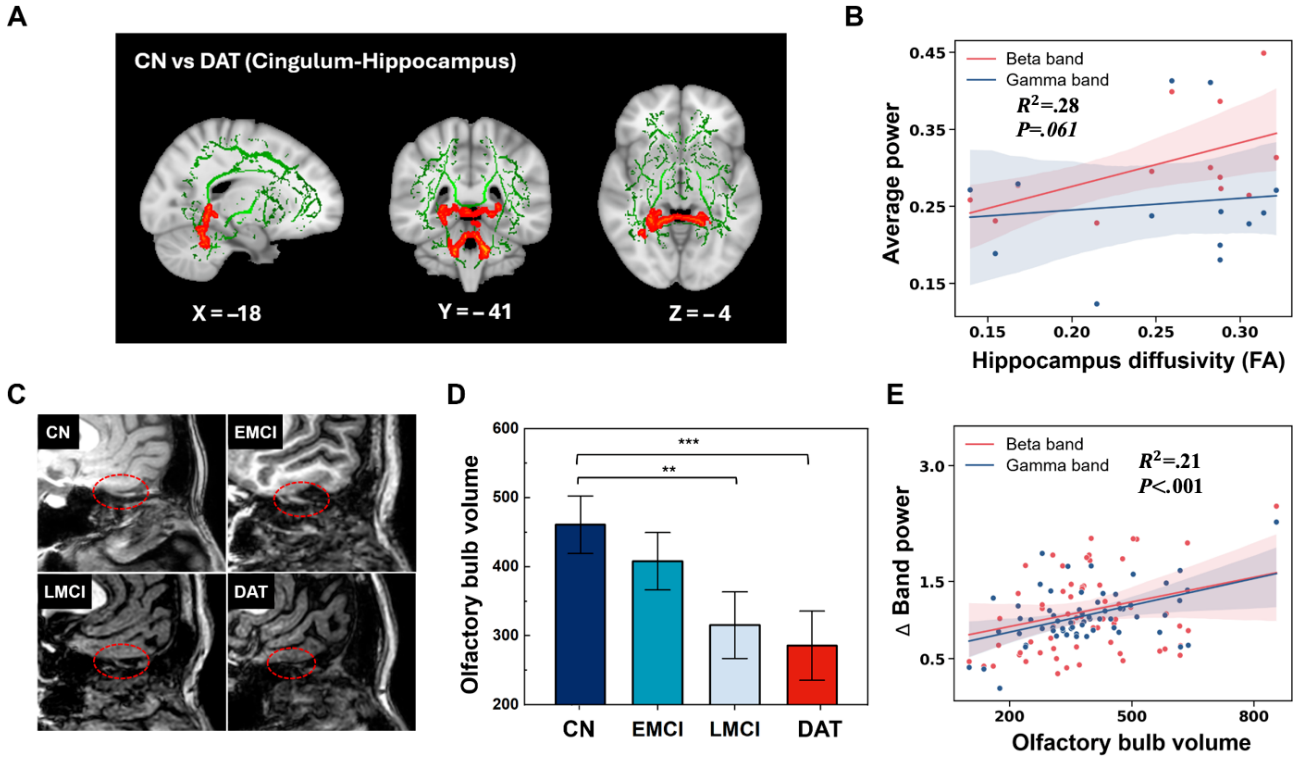
To elucidate the brain areas associated with the beta and gamma components within the olfactory system, we used a dual analytical approach that combined TBSS and volumetric MRI quantification. TBSS analysis revealed significantly reduced FA in the hippocampus cingulum of the DAT group compared with that in the CN group (P<.05), as shown in Figure 11A. The inverse correlation between FA values and cognitive impairment suggests that reduced white matter integrity in the hippocampal cingulum region is associated with more severe cognitive deficits [46-48]. Further analysis revealed a positive correlation approaching statistical significance between the beta-band components (mean of components 1, 3, and 5) in olfactory responses and FA values within the hippocampal-cingulum tracts in patients with DAT (R^2=0.28; P=.06), consistent with beta-band oscillations representing top-down cortical feedback signals. In contrast, gamma-band components showed minimal correlation with the integrity of the hippocampal-cingulum tracts (R^2=0.013; P=.53; Figure 11B).

In T1-weighted MR images, volumetric analysis of the OB revealed progressive atrophy across the disease spectrum, with a notable reduction in volume as cognitive status declined from CN to EMCI, LMCI, and DAT stages. Figure 11C provides representative T1-weighted MR images from each stage, with the OB regions delineated by red dashed circles, illustrating sequential volume reduction corresponding to cognitive decline. Quantitative analysis revealed significantly reduced OB volumes

XSL•FO
RenderX

in patients with LMCI compared with those in the CN group (P=.005), whereas patients with DAT exhibiting a pronounced volume reduction (P<.001), as shown in Figure 11D. In contrast, the EMCI group did not show a statistically significant volume reduction relative to the control group (P=.20), indicating that substantial OB atrophy may become more evident during later stages of cognitive decline [32,49]. This temporal pattern aligns with the delayed reduction in gamma-band olfactory components (components 2 and 4), suggesting parallel trajectories between structural and functional changes during later stages of the disease. Demographic analyses confirmed minimal influence on our primary outcomes. For OB volume, weak correlations with age (r=–0.23; P=.05) and education (r=0.17; P=.15) cannot explain the stage-specific volume reductions observed between groups (P<.01). Further analysis of OB volume and frequency-band power revealed a positive correlation between average frequency-band power and OB volume (Figure 11E). Specifically, for gamma bands, the coefficient of determination was 0.2108, with P<.001, indicating a strong and statistically significant relationship. In contrast, the correlation observed for the beta bands showed minimal optimization in linear fitting, yielding an R^2 value of 0.0613 (P=.05), suggesting a weaker association compared with the gamma bands, although still statistically significant.

Figure 11. Comparison of olfactory responses with clinically relevant structural evidence. (A) White matter structures showing a significant fractional anisotropy (FA) decrease in the hippocampal cingulum in the dementia of the Alzheimer type (DAT) group (P<.05). Statistical images were overlaid onto the mean skeleton (green) and the MNI152 template (gray scale) for visualization (B) FA values of the hippocampal cingulum positively correlated with the beta-band olfactory component (mean of components 1, 3, and 5), whereas gamma-band components (components 2 and 4) exhibited minimal correlation with hippocampal white matter integrity. (C) Representative T1-weighted magnetic resonance imaging of each cohort (cognitively normal [CN], early mild cognitive impairment, late mild cognitive impairment, and DAT, from left to right) with OB regions delineated by red dashed circles. (D) Olfactory bulb volume change across CN to DAT, presented with SE of the mean (SEM). Statistical significance was indicated using asterisks, with * representing P<.05, ** denoting P<.01, and *** signifying P<.001. (E) The olfactory bulb volume positively correlated with the gamma-band olfactory component (mean of components 2 and 4), whereas beta-band components (components 1, 3, and 5) exhibited minimal optimization in linear fitting. EMCI: early mild cognitive impairment; LMCI: late mild cognitive impairment.



Discussion

Objective Biomarkers for Early AD Detection

Objective measurement of odorant-mediated olfactory responses has the potential to enable the early diagnosis of AD by accurately predicting clinical neurobehavioral examination scores. In this study, we aimed to determine whether the olfactory response can serve as an early biomarker for AD, enabling us to distinguish between CN, EMCI, LMCI, and DAT groups. Consequently, a significant difference in the average power was observed between individuals who were CN and patients with DAT, which is consistent with the results of the 6-item modified B-SIT. These findings highlight a potential advantage of electrophysiological over behavioral olfactory assessments. While behavioral testing revealed no significant differences between CN and MCI groups, OB recordings showed distinct neural signatures that may reflect subtle disease-related changes. However, given the modest sample size and partial overlap in performance, these results should be considered preliminary. Larger validation studies are warranted to confirm whether OB electrophysiological responses can reliably detect early olfactory dysfunction not captured by behavioral tests. When measuring the electrical signals from the OB across all 4 groups, we observed a decrease in the beta- and gamma-band power and a delay in responses to odor exposure, which correlated with the progression of cognitive dysfunction.

XSL•FO
RenderX

In addition, signals obtained through digitalized olfactory measurements enabled the identification of olfactory response components distributed across various frequencies and time intervals, which allowed for discrimination between the DAT and CN groups at the individual level. The model exhibited outstanding performance in distinguishing among groups: CN (72.22% accuracy), MCI (100% accuracy), and DAT (50% accuracy, 100% precision). These results correlated strongly with the SNSB-II cognitive dysfunction levels and structural characteristics of the brain. Post hoc analyses demonstrated that potential demographic confounders had minimal influence on our findings, with negligible correlations observed between age and cognitive scores (r=–0.16; *P*>.20) and between education and cognitive scores (*r*=0.24; *P*>.05). These findings were further confirmed by formal demographic adjustment, in which the primary electrophysiological predictor retained independent significance for both SNSB-II (*P*=.04) and MMSE (*P*=.009) after controlling for age, sex, and education. While age and education shared some explanatory variance with secondary spectral components, this overlap is expected given the known influence of these factors on cognitive performance and the fact that SNSB-II scores are already standardized for age, sex, and education. Importantly, classification performance did not improve when demographic covariates were included, supporting the independent contribution of electrophysiological features. Future studies with larger, demographically balanced cohorts will be essential to further refine the generalizability and demographic robustness of these electrophysiological biomarkers.

Furthermore, we recorded signals from 6 electrodes using a slightly modified version of a previously reported electrobulbogram to measure the electrical response of the OB. This method has proven to reliably assess OB function in healthy participants, demonstrating high test-retest reliability in previous studies [35]. Notably, patients with PD show a high prevalence of olfactory dysfunction, with approximately 90% of patients experiencing some degree of smell impairment [50]. The high rate of olfactory dysfunction, which is often evident years before the onset of motor symptoms, highlights its potential role as an early indicator of disease. Recently, a noteworthy study presented the electrobulbogram method as an effective tool for detecting PD by distinguishing patients diagnosed with the condition from control participants [37].

Bidirectional Neurophysiological Insights and mHealth Applications

Olfactory dysfunction is one of the earliest clinical symptoms of AD. It is recognized as both an early indicator and a potential clinical marker of the progression and severity of the disorder [8-10]. In addition to cognitive impairment, approximately 90% of patients with DAT experience loss of olfactory abilities, such as difficulties with olfactory recognition and identification [12]. Sniffing kits, such as UPSIT, are frequently used to examine olfactory dysfunction in patients suspected of having MCI or AD [20-22]. In this study, although the sniffing kit clearly distinguished between healthy individuals and patients with DAT, the differences between healthy and MCI participants were not significant. When electrical signals from the OB were compared across all 4 groups, a decrease in beta- and gamma-band power and delayed responses to odor exposure were observed. Given that the gamma band is primarily associated with processes within the OB [51-55], these observed differences indicated impaired OB processing in DAT due to neuronal loss [10]. Moreover, activity in the beta band is likely generated by top-down projections, which is further supported by the increased latencies and high correlation with cingulum hippocampus diffusivity observed in our data [54,56]. These findings suggest an impairment in the bidirectional interaction between the OB and higher olfactory cortex regions, such as the amygdala and orbitofrontal cortex, which are involved in the processing of various aspects of odor perception. Although a significant reduction in total power was noted between the CN and DAT participants, no significant decrease in total power was found between the CN and MCI groups (early and late) or between the early and late MCI groups. These results suggest a gradual delay in gamma-band synchronization, ultimately leading to a complete loss of olfactory response. This observation is consistent with prior research, indicating that the recognition ability in patients with DAT declines early, whereas the detection threshold is preserved for a relatively longer period [22]. The combined use of TBSS and volumetric analyses offers complementary insights; TBSS enables a detailed investigation of white matter tracts, whereas volumetric quantification assesses regional brain volume changes. This comprehensive structural analysis framework helps map the relationships between functional olfactory components and their underlying neuroanatomical correlations. The observed differential patterns of association suggest distinct neuroanatomical substrates for the beta- and gamma-frequency components of olfactory processing. Our cross-sectional findings show that functional olfactory changes can be present even when volumetric measures remain relatively intact. However, the cross-sectional nature of our study precludes any conclusions about temporal relationships or causality. While we observed functional changes in the presence of preserved structure in some groups, we cannot determine whether electrophysiological changes truly precede volumetric alterations during disease progression. Longitudinal studies are essential to establish these temporal relationships and to confirm whether OB dysfunction represents an early biomarker that emerges before structural brain changes.

In summary, we attempted to digitize the measurement of OB function to differentiate individuals who were CN from patients with EMCI, LMCI, or DAT in a noninvasive manner. Accordingly, we isolated specific olfactory characteristics that significantly differentiate CN and DAT groups, which were successfully regressed against clinical AD diagnostic examinations, demonstrating high levels of sensitivity, specificity, and accuracy. Notably, these olfactory responses are specifically associated with disease characteristics and affected brain regions according to their frequency bands. This finding suggests that the OB, which is responsible for olfactory responses, is pathologically damaged in patients with DAT, resulting in olfactory dysfunction caused not only by the affected brain regions but also by direct biological damage to the OB. Therefore, our results support the potential of this methodology for further development as a tool to facilitate the early diagnosis of AD and provide a robust and objective biomarker for the disease.

XSL•FO
RenderX

Conclusion

Our study demonstrates that noninvasive digital recordings of OB function can serve as a novel biomarker for continuous and home-based monitoring of cognitive states across the continuum from normal aging to Alzheimer dementia, supporting early detection and timely intervention within an mHealth framework. The electrophysiological biomarkers we identified—particularly the progressive decline in beta-band power and later-stage reduction in gamma-band power—correlate strongly with established clinical and neuroimaging markers of AD progression. These electrophysiological changes may precede structural brain alterations, highlighting their potential value for early detection and longitudinal tracking. As a digital health tool, this technology has potential applications beyond diagnosis, including monitoring disease progression and potentially assessing therapeutic efficacy in clinical trials. Beyond objective signal capture, our wearable olfactory recording system offers distinct practical advantages for mHealth deployment. The device can be fitted in less than 2 minutes, and the entire assessment takes approximately 5‐10 minutes, allowing frequent, home-based measurements with minimal burden on users or caregivers. This approach embodies the mHealth paradigm, where continuous, real-world monitoring of physiological biomarkers enables early detection of cognitive decline and supports timely, personalized interventions. These findings support the development of accessible mHealth technologies for early AD detection and ongoing disease surveillance, addressing the growing need for noninvasive, cost-effective screening tools that can be deployed more widely than expensive neuroimaging techniques or invasive procedures.

Acknowledgments

We would like to express our sincere gratitude to all participants who took part in this study. Their commitment and cooperation made this research possible. No generative artificial intelligence tools were used in any portion of the paper writing process.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT (MSIT) of the Korean government (RS-2023-00302751); the Bio&Medical Technology Development Program of the NRF funded by MSIT (RS-2023-00226494); the Commercialization Promotion Agency for R&D Outcomes (COMPA) funded by MSIT (RS-2025-02311678); and the Regional Innovation System & Education (RISE) program through the Seoul RISE Center, funded by the Ministry of Education (MOE) and the Seoul Metropolitan Government (2026-RISE-01-027-01).

Data Availability

The olfactory response data and MATLAB (The MathWorks, Inc) code supporting the findings of this study are available from the corresponding author upon reasonable request for noncommercial use, subject to a materials transfer agreement with Allfacto Lab. Raw medical data, including magnetic resonance imaging and diffusion tensor imaging, are not publicly available owing to privacy restrictions and ethical considerations but may be available from the corresponding author upon reasonable request and subject to appropriate ethical approvals.

Authors' Contributions

JH, HJK, and YJ conceptualized the study. JH, JB, and YL curated the data. JH and JB conducted the formal analysis. HJK and YJ acquired the funding. JH, YSK, and YL conducted the investigation. JH, HJK, and YJ developed the methodology. HJK, JCK, and YJ administered the project. HJK and YJ supervised the study. JH visualized the data. JH and YJ wrote the original draft. HJK and YJ contributed equally as corresponding authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Flowchart of participant recruitment, screening, exclusion, and diagnostic classification. Eligible participants were assigned to the cognitively normal, early mild cognitive impairment, late mild cognitive impairment, or dementia of the Alzheimer type groups based on standardized clinical, neuropsychological, and neuroimaging criteria.

[DOCX File, 142 KB - mhealth_v14i1e76245_app1.docx]

References

1. Shaw LM, Arias J, Blennow K, et al. Appropriate use criteria for lumbar puncture and cerebrospinal fluid testing in the diagnosis of Alzheimer's disease. Alzheimer's Dementia 2018 Nov;14(11):1505-1521. [doi: 10.1016/j.jalz.2018.07.220]
2. Hansson O. Biomarkers for neurodegenerative diseases. Nat Med 2021 Jun;27(6):954-963. [doi: 10.1038/s41591-021-01382-x] [Medline: 34083813]

XSL•FO
RenderX

3. Varma VR, Ghosal R, Hillel I, et al. Continuous gait monitoring discriminates community-dwelling mild Alzheimer's disease from cognitively normal controls. Alzheimers Dement (N Y) 2021;7(1):e12131. [doi: 10.1002/trc2.12131] [Medline: 33598530]
4. Yan S, Yun X, Liu Q, Hong Z, Chen Y, Zhang S. Advances in gait research related to Alzheimer's disease. Front Neurol 2025;16:1548283. [doi: 10.3389/fneur.2025.1548283] [Medline: 40529444]
5. Liu DX, Braskie MN, Cavaillès C, Peltz C, Redline S, Yaffe K. Sleep macro-architecture, nocturnal hypoxemia, and Alzheimer's disease-related MRI patterns among diverse older adults. Alzheimers Dement 2025 May;21(5):e70280. [doi: 10.1002/alz.70280] [Medline: 40390243]
6. Popp Z, Low S, Igwe A, et al. Shifting from active to passive monitoring of Alzheimer disease: the state of the research. J Am Heart Assoc 2024 Jan 16;13(2):e031247. [doi: 10.1161/JAHA.123.031247] [Medline: 38226518]
7. Dieffenderfer J, Brewer A, Noonan MA, et al. A wearable system for continuous monitoring and assessment of speech, gait, and cognitive decline for early diagnosis of ADRD. Annu Int Conf IEEE Eng Med Biol Soc 2023 Jul;2023(1–6):1-6. [doi: 10.1109/EMBC40787.2023.10339986] [Medline: 38082824]
8. Serby M, Larson P, Kalkstein D. The nature and course of olfactory deficits in Alzheimer's disease. Am J Psychiatry 1991 Mar;148(3):357-360. [doi: 10.1176/ajp.148.3.357] [Medline: 1992839]
9. Devanand DP, Michaels-Marston KS, Liu X, et al. Olfactory deficits in patients with mild cognitive impairment predict Alzheimer's disease at follow-up. Am J Psychiatry 2000 Sep;157(9):1399-1405. [doi: 10.1176/appi.ajp.157.9.1399] [Medline: 10964854]
10. Hummel T, Landis BN, Hüttenbrink KB. Smell and taste disorders. GMS Curr Top Otorhinolaryngol Head Neck Surg 2011;10:Doc04. [doi: 10.3205/cto000077] [Medline: 22558054]
11. Murphy C. Olfactory and other sensory impairments in Alzheimer disease. Nat Rev Neurol 2019 Jan;15(1):11-24. [doi: 10.1038/s41582-018-0097-5] [Medline: 30532084]
12. Duff K, McCaffrey RJ, Solomon GS. The Pocket Smell Test: successfully discriminating probable Alzheimer's dementia from vascular dementia and major depression. J Neuropsychiatry Clin Neurosci 2002;14(2):197-201. [doi: 10.1176/jnp.14.2.197] [Medline: 11983795]
13. Doty RL. Olfactory dysfunction in neurodegenerative diseases: is there a common pathological substrate? Lancet Neurol 2017 Jun;16(6):478-488. [doi: 10.1016/S1474-4422(17)30123-0] [Medline: 28504111]
14. Hawkes C. Olfaction in neurodegenerative disorder. Adv Otorhinolaryngol 2006;63:133-151. [doi: 10.1159/000093759] [Medline: 16733338]
15. Attems J, Walker L, Jellinger KA. Olfactory bulb involvement in neurodegenerative diseases. Acta Neuropathol 2014 Apr;127(4):459-475. [doi: 10.1007/s00401-014-1261-7] [Medline: 24554308]
16. Velayudhan L. Smell identification function and Alzheimer's disease: a selective review. Curr Opin Psychiatry 2015 Mar;28(2):173-179. [doi: 10.1097/YCO.0000000000000146] [Medline: 25594420]
17. Attems J, Jellinger KA. Olfactory tau pathology in Alzheimer disease and mild cognitive impairment. Clin Neuropathol 2006;25(6):265-271. [Medline: 17140156]
18. Sun GH, Raji CA, Maceachern MP, Burke JF. Olfactory identification testing as a predictor of the development of Alzheimer's dementia: a systematic review. Laryngoscope 2012 Jul;122(7):1455-1462. [doi: 10.1002/lary.23365] [Medline: 22552846]
19. Marin C, Vilas D, Langdon C, et al. Olfactory dysfunction in neurodegenerative diseases. Curr Allergy Asthma Rep 2018 Jun 15;18(8):1-19. [doi: 10.1007/s11882-018-0796-4] [Medline: 29904888]
20. Silva M, Mercer PBS, Witt MCZ, Pessoa RR. Olfactory dysfunction in Alzheimer's disease systematic review and meta-analysis. Dement Neuropsychol 2018;12(2):123-132. [doi: 10.1590/1980-57642018dn12-020004] [Medline: 29988355]
21. Taherkhani S, Moztarzadeh F, Mehdizadeh Seraj J, et al. Iran Smell Identification Test (Iran-SIT): a modified version of the University of Pennsylvania Smell Identification Test (UPSIT) for Iranian population. Chem Percept 2015 Dec;8(4):183-191. [doi: 10.1007/s12078-015-9192-9]
22. Zou YM, Lu D, Liu LP, Zhang HH, Zhou YY. Olfactory dysfunction in Alzheimer's disease. NDT 2016:869. [doi: 10.2147/NDT.S104886]
23. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging - Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011 May;7(3):263-269. [doi: 10.1016/j.jalz.2011.03.005] [Medline: 21514250]
24. Kreisl WC, Jin P, Lee S, et al. Odor identification ability predicts PET amyloid status and memory decline in older adults. J Alzheimers Dis 2018;62(4):1759-1766. [doi: 10.3233/JAD-170960] [Medline: 29614678]
25. Velayudhan L, Gasper A, Pritchard M, Baillon S, Messer C, Proitsi P. Pattern of smell identification impairment in Alzheimer's disease. J Alzheimers Dis 2015;46(2):381-387. [doi: 10.3233/JAD-142838] [Medline: 25757648]
26. Invitto S, Piraino G, Ciccarese V, et al. Potential role of OERP as early marker of mild cognitive impairment. Front Aging Neurosci 2018;10:272. [doi: 10.3389/fnagi.2018.00272] [Medline: 30271339]
27. Paitel ER, Samii MR, Nielson KA. A systematic review of cognitive event-related potentials in mild cognitive impairment and Alzheimer's disease. Behav Brain Res 2021 Jan 1;396:112904. [doi: 10.1016/j.bbr.2020.112904] [Medline: 32941881]

XSL•FO
RenderX

28. Fatemi SN, Aghajan H, Vahabi Z, Afzal A, Sedghizadeh MJ. Behavior of olfactory-related frontal lobe oscillations in Alzheimer's disease and MCI: a pilot study. Int J Psychophysiol 2022 May;175:43-53. [doi: 10.1016/j.ijpsycho.2022.02.008] [Medline: 35217110]
29. Musaeus CS, Engedal K, Høgh P, et al. EEG theta power is an early marker of cognitive decline in dementia due to Alzheimer's disease. J Alzheimers Dis 2018;64(4):1359-1371. [doi: 10.3233/JAD-180300] [Medline: 29991135]
30. Sedghizadeh MJ, Hojjati H, Ezzatdoost K, Aghajan H, Vahabi Z, Tarighatnia H. Olfactory response as a marker for Alzheimer's disease: evidence from perceptual and frontal lobe oscillation coherence deficit. PLoS One 2020;15(12):e0243535. [doi: 10.1371/journal.pone.0243535] [Medline: 33320870]
31. Thomann PA, Dos Santos V, Toro P, Schönknecht P, Essig M, Schröder J. Reduced olfactory bulb and tract volume in early Alzheimer's disease--a MRI study. Neurobiol Aging 2009 May;30(5):838-841. [doi: 10.1016/j.neurobiolaging.2007.08.001] [Medline: 17875348]
32. Jobin B, Boller B, Frasnelli J. Volumetry of olfactory structures in mild cognitive impairment and Alzheimer's disease: a systematic review and a meta-analysis. Brain Sci 2021 Jul 30;11(8):1010. [doi: 10.3390/brainsci11081010] [Medline: 34439629]
33. Thomann PA, Dos Santos V, Seidl U, Toro P, Essig M, Schröder J. MRI-derived atrophy of the olfactory bulb and tract in mild cognitive impairment and Alzheimer's disease. J Alzheimers Dis 2009;17(1):213-221. [doi: 10.3233/JAD-2009-1036] [Medline: 19494444]
34. Vasavada MM, Wang J, Eslinger PJ, et al. Olfactory cortex degeneration in Alzheimer's disease and mild cognitive impairment. J Alzheimers Dis 2015;45(3):947-958. [doi: 10.3233/JAD-141947] [Medline: 25633674]
35. Iravani B, Arshamian A, Ohla K, Wilson DA, Lundström JN. Non-invasive recording from the human olfactory bulb. Nat Commun 2020 Jan 31;11(1):648. [doi: 10.1038/s41467-020-14520-9] [Medline: 32005822]
36. Iravani B, Arshamian A, Lundqvist M, Kay LM, Wilson DA, Lundström JN. Odor identity can be extracted from the reciprocal connectivity between olfactory bulb and piriform cortex in humans. Neuroimage 2021 Aug 15;237:118130. [doi: 10.1016/j.neuroimage.2021.118130] [Medline: 33951509]
37. Iravani B, Arshamian A, Schaefer M, Svenningsson P, Lundström JN. A non-invasive olfactory bulb measure dissociates Parkinson's patients from healthy controls and discloses disease duration. NPJ Parkinsons Dis 2021 Aug 18;7(1):75. [doi: 10.1038/s41531-021-00220-8] [Medline: 34408159]
38. Ryu HJ, Yang DW. The Seoul Neuropsychological Screening Battery (SNSB) for comprehensive neuropsychological assessment. Dement Neurocogn Disord 2023 Jan;22(1):1-15. [doi: 10.12779/dnd.2023.22.1.1] [Medline: 36814700]
39. Kang IW, Beom IG, Cho JY, Son HR. Accuracy of Korean-Mini-Mental Status Examination based on Seoul Neuro-Psychological Screening Battery II results. Korean J Fam Med 2016 May;37(3):177-181. [doi: 10.4082/kjfm.2016.37.3.177] [Medline: 27274389]
40. Kang Y, NA DL, Hahn S. A validity study on the Korean Mini-Mental State Examination (K-MMSE) in dementia patients. J Korean Neurol Assoc 1997:300-308 [FREE Full text]
41. Hua K, Zhang J, Wakana S, et al. Tract probability maps in stereotaxic spaces: analyses of white matter anatomy and tract-specific quantification. Neuroimage 2008 Jan 1;39(1):336-347. [doi: 10.1016/j.neuroimage.2007.07.053] [Medline: 17931890]
42. Smith SM, Nichols TE. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 2009 Jan 1;44(1):83-98. [doi: 10.1016/j.neuroimage.2008.03.061] [Medline: 18501637]
43. Fischl B, Salat DH, Busa E, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 2002 Jan 31;33(3):341-355. [doi: 10.1016/s0896-6273(02)00569-x] [Medline: 11832223]
44. Doty RL, Marcus A, Lee WW. Development of the 12-item Cross-Cultural Smell Identification Test (CC-SIT). Laryngoscope 1996 Mar;106(3 Pt 1):353-356. [doi: 10.1097/00005537-199603000-00021] [Medline: 8614203]
45. Oostenveld R, Fries P, Maris E, Schoffelen JM. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011;2011:156869. [doi: 10.1155/2011/156869] [Medline: 21253357]
46. Ezzati A, Katz MJ, Lipton ML, Zimmerman ME, Lipton RB. Hippocampal volume and cingulum bundle fractional anisotropy are independently associated with verbal memory in older adults. Brain Imaging Behav 2016 Sep;10(3):652-659. [doi: 10.1007/s11682-015-9452-y] [Medline: 26424564]
47. Lin YC, Shih YC, Tseng WYI, et al. Cingulum correlates of cognitive functions in patients with mild cognitive impairment and early Alzheimer's disease: a diffusion spectrum imaging study. Brain Topogr 2014 May;27(3):393-402. [doi: 10.1007/s10548-013-0346-2] [Medline: 24414091]
48. Zhang Y, Schuff N, Jahng GH, et al. Diffusion tensor imaging of cingulum fibers in mild cognitive impairment and Alzheimer disease. Neurology (ECronicon) 2007 Jan 2;68(1):13-19. [doi: 10.1212/01.wnl.0000250326.77323.01] [Medline: 17200485]
49. Servello A, Fioretti A, Gualdi G, et al. Olfactory dysfunction, olfactory bulb volume and Alzheimer's disease: is there a correlation? A pilot study1. J Alzheimers Dis 2015;48(2):395-402. [doi: 10.3233/JAD-150232] [Medline: 26402003]
50. Doty RL. Olfactory dysfunction in Parkinson disease. Nat Rev Neurol 2012 May 15;8(6):329-339. [doi: 10.1038/nrneurol.2012.80] [Medline: 22584158]

XSL•FO
RenderX

51. Yang Q, Zhou G, Noto T, et al. Smell-induced gamma oscillations in human olfactory cortex are required for accurate perception of odor identity. PLoS Biol 2022 Jan;20(1):e3001509. [doi: 10.1371/journal.pbio.3001509] [Medline: 34986157]
52. Martin C, Ravel N. Beta and gamma oscillatory activities associated with olfactory memory tasks: different rhythms for different functional networks? Front Behav Neurosci 2014;8(218):218. [doi: 10.3389/fnbeh.2014.00218] [Medline: 25002840]
53. Neville KR, Haberly LB. Beta and gamma oscillations in the olfactory system of the urethane-anesthetized rat. J Neurophysiol 2003 Dec;90(6):3921-3930. [doi: 10.1152/jn.00475.2003] [Medline: 12917385]
54. Frederick DE, Brown A, Brim E, Mehta N, Vujovic M, Kay LM. Gamma and beta oscillations define a sequence of neurocognitive modes present in odor processing. J Neurosci 2016 Jul 20;36(29):7750-7767. [doi: 10.1523/JNEUROSCI.0569-16.2016] [Medline: 27445151]
55. Peace ST, Johnson BC, Werth JC, et al. Coherent olfactory bulb gamma oscillations arise from coupling independent columnar oscillators. J Neurophysiol 2024 Mar 1;131(3):492-508. [doi: 10.1152/jn.00361.2023] [Medline: 38264784]
56. Nordén F, Iravani B, Schaefer M, et al. The human olfactory bulb communicates perceived odor valence to the piriform cortex in the gamma band and receives a refined representation back in the beta band. PLoS Biol 2024 Oct;22(10):e3002849. [doi: 10.1371/journal.pbio.3002849] [Medline: 39401242]

Abbreviations

AD: Alzheimer disease
Aβ: Amyloid-beta
B-SIT: Brief Smell Identification Test
CN: cognitively normal
CT: computed tomography
DAT: dementia of the Alzheimer type
DTI: diffusion tensor imaging
EMCI: early mild cognitive impairment
FA: fractional anisotropy
LMCI: late mild cognitive impairment
MCI: mild cognitive impairment
mHealth: mobile health
MMSE: Mini-Mental State Examination
MRI: Magnetic Resonance Imaging
NIA-AA: National Institute on Aging and the Alzheimer's Association
OB: olfactory bulb
PD: Parkinson disease
PET: positron emission tomography
SNSB-II: Seoul Neuropsychological Screening Battery II
SSIM: Structural Similarity Index Matrix
TBSS: tract-based spatial statistics
UPSIT: University of Pennsylvania Smell Identification Test

Edited by L Buis; submitted 19.Apr.2025; peer-reviewed by AD Albers, B Jobin; revised version received 21.Feb.2026; accepted 27.Apr.2026; published 23.Jun.2026.

Please cite as:
Ha J, Bok J, Lee Y, Kim YS, Kim JS, Kim HJ, Jang Y
Noninvasive Electrophysiological Biomarkers of Olfactory Responses Across Cognitive States in Alzheimer Dementia: Cross-Sectional Study
JMIR Mhealth Uhealth 2026;14:e76245
URL: https://mhealth.jmir.org/2026/1/e76245
doi:10.2196/76245

© Juchan Ha, Junsoo Bok, Yunjin Lee, Yong sung Kim, June Sic Kim, Hee-Jin Kim, Yongwoo Jang. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 23.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Passive Monitoring for Detecting Hypomanic Symptoms in Bipolar Disorder Using Explainable Machine Learning: Longitudinal Observational Study

Ramzi Halabi[1], BEng, MSc, PhD; Martin Alda[2,3], MD, FRCP; Mirkamal Tolend[4], PhD; Claire O'Donovan[2], MD, FRCP; Arend Hintze[5], PhD; Benoit H Mulsant[4,6,7], MS, MD, FRCP, FRCS; Abigail Ortiz[1,4,7], MS, MD, FRCP

[1]Department of Psychiatry, The University of Texas Southwestern Medical Center, 1440 Empire Central Dr LD4.244, Dallas, TX, United States

[2]Department of Psychiatry, Dalhousie University, Halifax, NS, Canada

[3]Scientific Council, National Institute of Mental Health, Klecany, Czech Republic

[4]Campbell Family Research Institute, Centre for Addiction and Mental Health (CAMH), Toronto, ON, Canada

[5]Department of MicroData Analytics, Dalarna University, Falun, Sweden

[6]Department of Psychiatry, Royal Ottawa Mental Health Centre, Ottawa, ON, Canada

[7]Department of Psychiatry, Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Abigail Ortiz, MS, MD, FRCP
Department of Psychiatry, The University of Texas Southwestern Medical Center, 1440 Empire Central Dr LD4.244, Dallas, TX, United States

Abstract

Background: When trying to intervene early to prevent manic episodes, clinicians are hindered by patients underreporting their symptoms as they are becoming hypomanic or manic (ie, emerging symptoms). Continuous passive monitoring using wearable technology offers a potential solution to address this problem by capturing behavioral signals that do not depend on self-awareness. However, its clinical usefulness requires validation before implementation.

Objective: This study aimed to evaluate whether passively collected wearable data can detect hypomanic symptoms using an explainable machine learning framework.

Methods: We applied explainable machine learning to densely sampled multimodal data from a 2-year observational study of digital monitoring in 178 patients with a diagnosis of bipolar disorder (mean follow-up 540 days; range 120-706 days). Data on activity and sleep were collected continuously using wearable technology. Self-reported data were collected daily on their mood, energy, and anxiety, and weekly using the Altman Self-Rating Mania Scale (ASRM). To identify robust digital markers of hypomanic symptoms, we analyzed 252 time- and frequency-domain features from 43 participants who experienced at least 1 ASRM-defined symptomatic interval.

Results: Passively collected data achieved high accuracy, comparable with daily self-reported data in detecting symptomatic intervals (area under the receiver operating characteristic curve 0.91 vs 0.93, respectively). Combined daily passive wearable data features achieved higher sensitivity (92.0% vs 85.4%), balanced accuracy (0.90 vs 0.89), and stronger rule-out usefulness (negative predictive value 97.8% vs 96.1%) than self-report data, with daily sleep features alone driving most of this signal (area under the receiver operating characteristic curve 0.91, sensitivity 91.8%). The high classification performance using wearable data was explained by reductions in spectral entropy and flux across domains, reflecting a decreased complexity and variability in mood, activity, and sleep patterns, consistent with a shift toward pathologically stable states.

Conclusions: Passive monitoring using wearable technology can detect hypomanic symptoms with high accuracy, supporting its potential as a clinically meaningful tool for continuous, objective monitoring in patients who underreport hypomanic symptoms. These findings are associative and concurrent in nature; they do not establish predictive validity. Prospective studies are needed to evaluate predictive validity for early detection of ASRM-defined symptomatic intervals.

(JMIR Mhealth Uhealth 2026;14:e105647) doi:10.2196/105647

KEYWORDS

bipolar disorder; mania; machine learning; wearables; mood; activity; sleep

XSL•FO
RenderX

Introduction

Clinical assessment typically relies on retrospective self-reports during clinical visits, which are prone to recall bias. A key limitation of symptom self-report in bipolar disorder (BD) is that insight into illness is frequently impaired, particularly during hypomanic or manic episodes: fewer than a quarter of patients recognized having ever experienced a hypomanic or manic episode, compared with over 80% who recognized past depressive episodes [1]. Because diagnosis and ongoing monitoring rely heavily on patients reporting their own symptoms [2,3], this lack of insight contributes directly to the underdetection and undertreatment of hypomanic or manic episodes [4].

Passive monitoring using wearable technology offers a way around this problem: physiological (eg, sleep architecture) and behavioral (eg, activity levels) signals are captured continuously in the patient's natural environment, moving beyond retrospective clinical assessments toward continuous, objective monitoring [5,6]. This approach, often termed digital phenotyping, can capture subtle yet significant changes in patterns that precede the onset of a mood episode and has shown promise in mental health [7]. In BD specifically, disruptions in activity and circadian rhythms have shown their potential for episode prediction [8-13]. Furthermore, a major insight in digital phenotyping is that day-to-day variability in behavioral and physiological metrics may be more informative than static measures for detecting episode onset. We have demonstrated that variability in activity levels precedes mood changes by approximately 7 days in depressive episodes [14]. Similarly, we identified variability in sleep stages and activity levels as the earliest indicators of hypomanic episode onset, achieving high sensitivity (0.94) and specificity (0.80) [8]. While these findings underscore variability as a fundamental concept for the early detection of episodes, spectral or frequency-domain features that reveal rhythmic dynamics remain underexplored in the context of BD.

Additionally, the optimal integration of objective, passively collected wearable data with subjective reports of a participant's state (eg, electronic visual analog scales [e-VASs] and self-rating scales) remains underexplored [15]. In this context, while machine learning (ML) is essential for detecting subtle patterns in high-dimensional data, its clinical translation is often hindered by limited interpretability and the need for personalized approaches that account for heterogeneity [16-18]. Hence, robust, multimodal digital markers that combine passive and active data, validated through repeated within- and across-participant testing, are needed to develop personalized, clinically actionable early-warning systems.

Beyond methodological considerations, several clinical and technical challenges further complicate the development of digital markers for hypomanic or manic episodes. Clinically, hypomanic or manic episodes occur less frequently and are shorter in duration than depressive episodes [19]. This imbalance results in fewer training examples for predictive models and reduced statistical power to detect changes. From a technical perspective, many digital phenotyping studies in BD are characterized by short follow-up durations, limiting the capture of hypomanic or manic symptoms and hindering the development of robust, generalizable predictive models for detecting these types of episodes.

To address all these gaps, this manuscript describes a longitudinal observational study designed to identify and validate robust digital markers of hypomanic or manic symptoms. Our approach integrates 2 key data sources: objective (ie, continuous, passively collected activity and sleep data from a consumer-grade wearable), and subjective (ie, daily self-reports of mood, energy, and anxiety collected via e-VAS, and weekly self-reports of symptom severity collected via Altman Self-Rating Mania Scale [ASRM], a validated self-report measure [20] used to operationalize mood state classification). By applying explainable ML techniques to a diverse set of time- and frequency-domain features extracted from a longitudinal multimodal dataset, we aimed to (1) identify the most robust and accurate digital markers of hypomanic symptoms and (2) compare the discriminative accuracy of passively collected wearable data against self-reported measures.

Methods

Participants and Data Collection

The study was approved by the Research Ethics Board (REB) at the Centre for Addiction and Mental Health (CAMH), Toronto, Ontario, Canada, and at the Queen Elizabeth II Health Sciences Centre (QEII Health Care Center), Halifax, Nova Scotia, Canada, in accordance with the Declaration of Helsinki, REB number 059 - 2019. All participants signed an informed consent form approved by the local REB before any research procedure was initiated.

We have previously described the overall methods of this ongoing longitudinal study [21]. In summary, between December 12, 2020, and January 1, 2025, we recruited and followed 178 adults diagnosed with BD type I or II in the outpatient clinics of 2 academic psychiatric hospitals in Canada: the CAMH and the QEII.

Primary diagnoses of BD I or II were confirmed with the SCID-5 (Structured Clinical Interview for the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*) [22] based on the diagnostic criteria of the DSM-5 (*Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*) [23]. Baseline assessments included sociodemographic and clinical data. Clinical polarity at study entry was defined using Young Mania Rating Scale (YMRS) [24] and Montgomery-Asberg Depression Rating Scale [25] thresholds.

We instructed each participant to wear an Oura Ring (Oura Health Oy; Generation 2) continuously throughout the study course. The ring collected multiple activity and sleep data variables, of which we selected a comprehensive subset [26] relevant to the goals of this study (Figure 1A). Additionally, participants received a daily email with a link to an e-VAS to self-rate their mood, energy, and anxiety levels over the previous 24 hours. Reminders were sent to participants when e-VASs were not completed for 3 consecutive days. Participants also received a weekly email with a link to an electronic version of

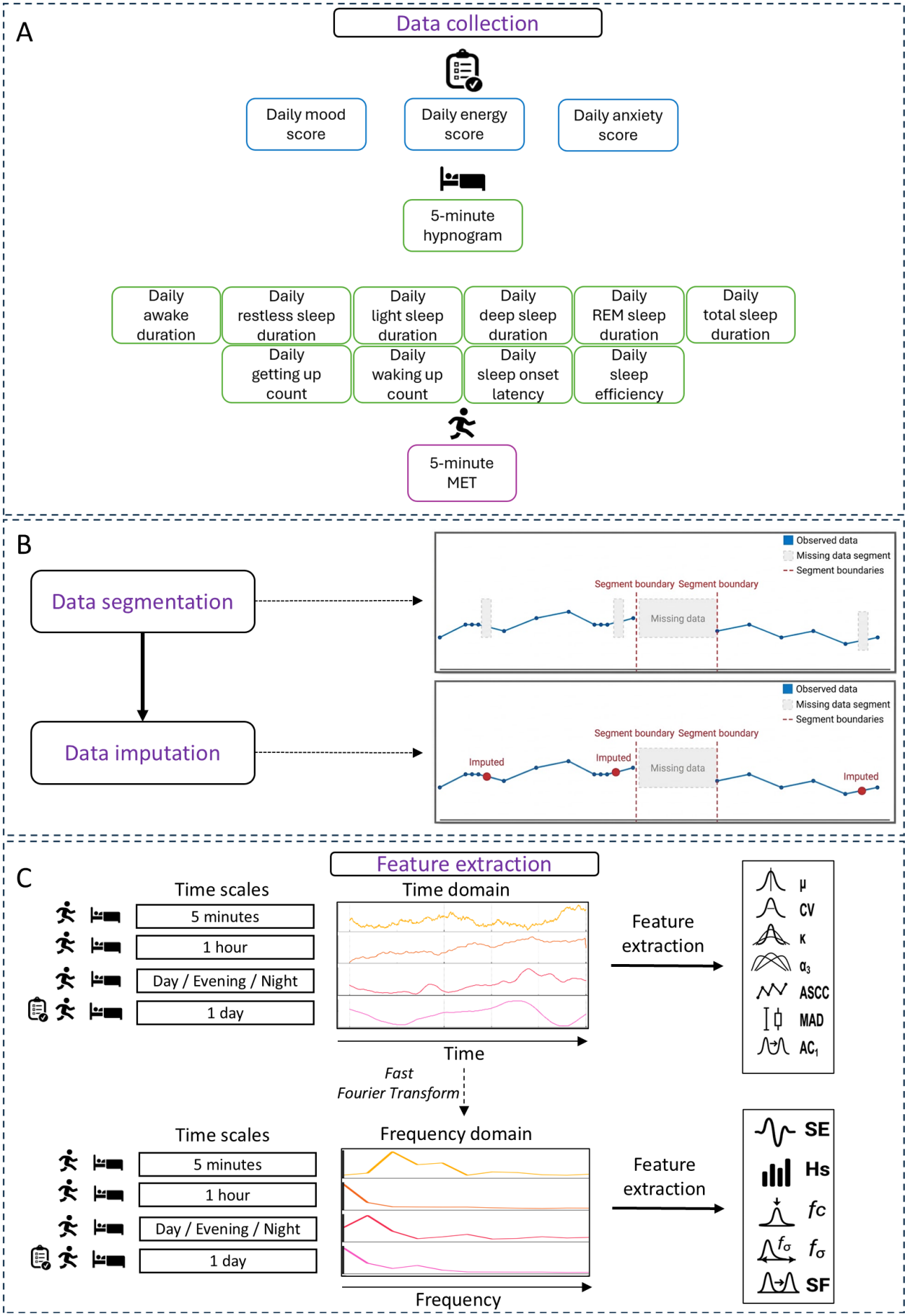
XSL•FO
RenderX

the ASRM [20] to rate their symptoms for the previous week. Weekly symptom self-reports were stored in a REDCap (Research Electronic Data Capture) [27] secure database.

XSL•FO
RenderX

Figure 1. Data collection, processing, and feature extraction pipeline. This figure illustrates the workflow applied to data from a longitudinal observational study of 178 outpatients with bipolar disorder I or II recruited at the Centre for Addiction and Mental Health, Toronto, Ontario, and the Queen Elizabeth II Health Sciences Centre, Halifax, Nova Scotia, Canada (December 2020 to January 2025). Panel (A) shows the selected objective (activity and sleep) and subjective (electronic visual analog scale) variables collected via Oura Ring Generation 2 and daily/weekly self-reports. Panel (B) depicts the segmentation and feature extraction process. Panel (C) displays the 12 time- and frequency-domain features computed per variable per segment. See Table S1 in Multimedia Appendix 1 for the full variable list and Table S2 in Multimedia Appendix 1 for feature definitions. AC_1: autocorrelation at lag 1; ASCC: absolute sum of consecutive changes; CV: coefficient of variation; MAD: mean absolute differences; MET: metabolic equivalent of task; REM: rapid eye movement (sleep stage); SE: spectral energy; SF: spectral flux; μ: mean; κ: kurtosis; α_3: skewness; H_s: spectral entropy; f_c: spectral centroid; f_σ: spectral spread.

XSL•FO
RenderX



XSL•FO
RenderX

We evaluated participant adherence to device-wear protocols by quantifying nonwear time and modality-specific missing data ratios. We obtained nonwear time directly from the Oura Ring activity dataset and calculated missing data ratios separately for each modality (activity, sleep, and e-VAS) [28,29].

Data Generation and Processing

Data Generation

Figure 1 presents an overview of the clinical course and data generation time scales. Using the longitudinal ASRM scores of each participant, we applied a 1-week rolling window with 50% overlap to identify ASRM-defined symptomatic intervals, defined as at least 1 weekly rating with an ASRM score of ≥10. The ASRM is a self-rating scale comprising 5 items designed to evaluate the existence and severity of 5 hypomanic or manic symptoms: elation, overconfidence, decreased need for sleep, talkativeness, and increased energy levels, with good validity and reliability [30] when compared with the clinician-administered YMRS [20]. Each item is rated on a scale from 0 to 4, with 0 representing the absence of symptoms and 4 indicating sustained and pronounced symptoms.

Symptomatic Interval Labeling

Because the ASRM has shown high sensitivity and low specificity [31], our operationalized definition of a score of ≥10 typically indicates the presence of 3 or more symptoms nearly every day. By requiring a score of ≥10 for a week, we also captured the DSM-5 four days duration requirement for hypomania and the 1-week duration requirement for mania. The standard ASRM threshold of ≥6 has been recommended by the scale developers for screening purposes [31]. However, we used a more conservative threshold of ≥10 for the following reasons: (1) the ASRM has been shown to have high sensitivity but low specificity [31], and a higher threshold reduces false positives, which is critical when weekly ratings are used to define interval-level labels for ML classification; (2) a score of ≥10 typically requires that a participant endorse at least 3 of 5 symptoms at a moderate-to-severe level (ie, ≥2 on at least 3 items), providing higher clinical specificity for a symptomatic interval; and (3) by requiring this threshold to be sustained over a rolling 1-week window, we further reduce noise from transient elevations (“symptomatic blips”). Moreover, we defined a symptomatic interval as 1 or more consecutive weeks in which the ASRM overall score was ≥10 based on previous studies defining “symptomatic remission” as an ASRM score of <10 [32,33].

To assess the robustness of our findings to this threshold choice, we conducted a sensitivity analysis repeating the full analytical pipeline (feature extraction, classification, and feature importance) using the standard ASRM threshold of ≥6. Some participants presented with an ASRM score of ≥10 on 1 occasion. Because our dataset uses weekly symptom ratings, we could not determine whether these elevated ASRM scores lasted 4 to 6 days (ie, the minimum duration for a hypomanic episode according to DSM-5). Thus, these elevations in ASRM scores that did not span at least 2 weekly ratings were not considered symptomatic intervals in the primary analysis.

Because elevations in symptoms were not independently confirmed by a clinician or a clinician-rated scale, we refer to them as “symptomatic intervals,” rather than “episodes.” Weeks for which the ASRM was not completed were not assigned a label and were excluded from the definition of symptomatic intervals and the euthymic comparator. Depressive and mixed symptomatic intervals were identified using a Patient Health Questionnaire (9-item; PHQ-9) threshold of ≥10; such intervals were excluded from the binary classification analysis (see euthymic comparator definition).

Euthymic Comparator Labeling

Euthymic segments were defined as data segments during which the participants’ weekly ASRM score was <10 and their weekly PHQ-9 score was <10. The use of both ASRM and PHQ-9 thresholds ensured that euthymic segments fell below the standard screening threshold for hypomanic or manic symptoms (ASRM score of <10) and depressive symptoms (PHQ-9 <10), although we cannot ensure that subsyndromal presentations were present. Segments that did not meet criteria for either ASRM-defined symptomatic intervals or euthymia (eg, intervals with elevated PHQ-9 but not ASRM, or transitional periods) were excluded from the binary classification analysis.

Data Variables

Among the 39 data variables generated by the wearable, we selected a smaller set of 12 objective variables (1 of 16 related to activity and 11 of 23 related to sleep). We derived the activity variables included in the analysis from (1) the 5-minute-sampled metabolic equivalent of task (MET), produced by the wearable; and (2) its aggregated variants over 1-hour periods, as well as daytime, evening, and nighttime. Similarly, we derived the sleep variables from (1) the 5-minute hypnograms produced by the wearable and (2) a set of 10 daily data variables provided by the wearable’s proprietary algorithms (eg, wake-up count). Moreover, 3 daily e-VAS variables were collected to describe self-reported mood, energy, and anxiety dynamics (Figure 1A and Table S1 in Multimedia Appendix 1). Overall, based on objective and subjective data across the various time periods, we created a set of 21 variables.

Sleep stages in the 5-minute hypnogram were encoded categorically into 4 stages: awake, light sleep (N1/N2), deep sleep (N3/slow-wave sleep), and rapid eye movement (REM) sleep using the Oura Ring’s proprietary stage labels. This encoding preserves the physiological ordering of sleep stages and allows the same time series feature extraction pipeline to be applied to both activity and sleep-stage data. Irregular sampling within 5-minute windows (eg, brief gaps from device removal) was handled by forward-filling missing epochs from the immediately preceding nonmissing epoch, capped at a maximum of 2 consecutive missing epochs (10 minutes); isolated stage flickers that are typical of automated staging outputs were filtered via moving-window mode smoothing (categorical “majority vote”); days with more than 60 consecutive minutes of missing epochs were classified as nonwear and excluded from analysis.

XSL•FO
RenderX

Data Inclusion Criteria

Data inclusion criteria required at least 80% completeness per data segment; segments failing this threshold were excluded from analysis. For the 5-minute MET activity data, nonwear periods were identified using the Oura Ring's proprietary wear detection algorithm and confirmed by sustained zero-MET readings of ≥60 minutes. The 5-minute hypnograms were filtered to remove artefactual sleep-stage transitions (ie, rapid cycling between stages within <5 minutes inconsistent with physiological sleep architecture). Daily sleep variables provided by the wearable's proprietary algorithms (eg, wake-up count and sleep efficiency) were used as is without additional filtering, as these are already validated outputs of the device's sleep pipeline.

Data Completeness

Given our previous findings showing that data in our sample were missing not at random [29], we imputed the daily sleep variables and the 5-minute hypnograms using completeness-based adaptive segmentation (Figure 1B). The k-nearest neighbors (KNNs) method [34] was performed with k=5 neighbors using Euclidean distance, with the imputation model fit within each participant's data and within each cross-validation training fold to prevent information leakage. The adaptive segmentation approach identified contiguous data segments of >7 days with ≥80% completeness; segments shorter than 8 days or with >20% missingness were excluded. To systematically characterize missingness, we report the proportion of missing data for each modality (activity, sleep, e-VAS, and ASRM) both overall and stratified by clinical state (euthymic, ASRM-defined symptomatic interval). We assessed whether data were missing completely at random (MCAR) using Little MCAR test, reporting the test statistic and associated *P* value [35]. Based on our previous work [29], we hypothesized that Little MCAR test would be rejected; we therefore report sensitivity analyses comparing classifier performance, selected features, and importance rankings with and without imputation.

Data Features

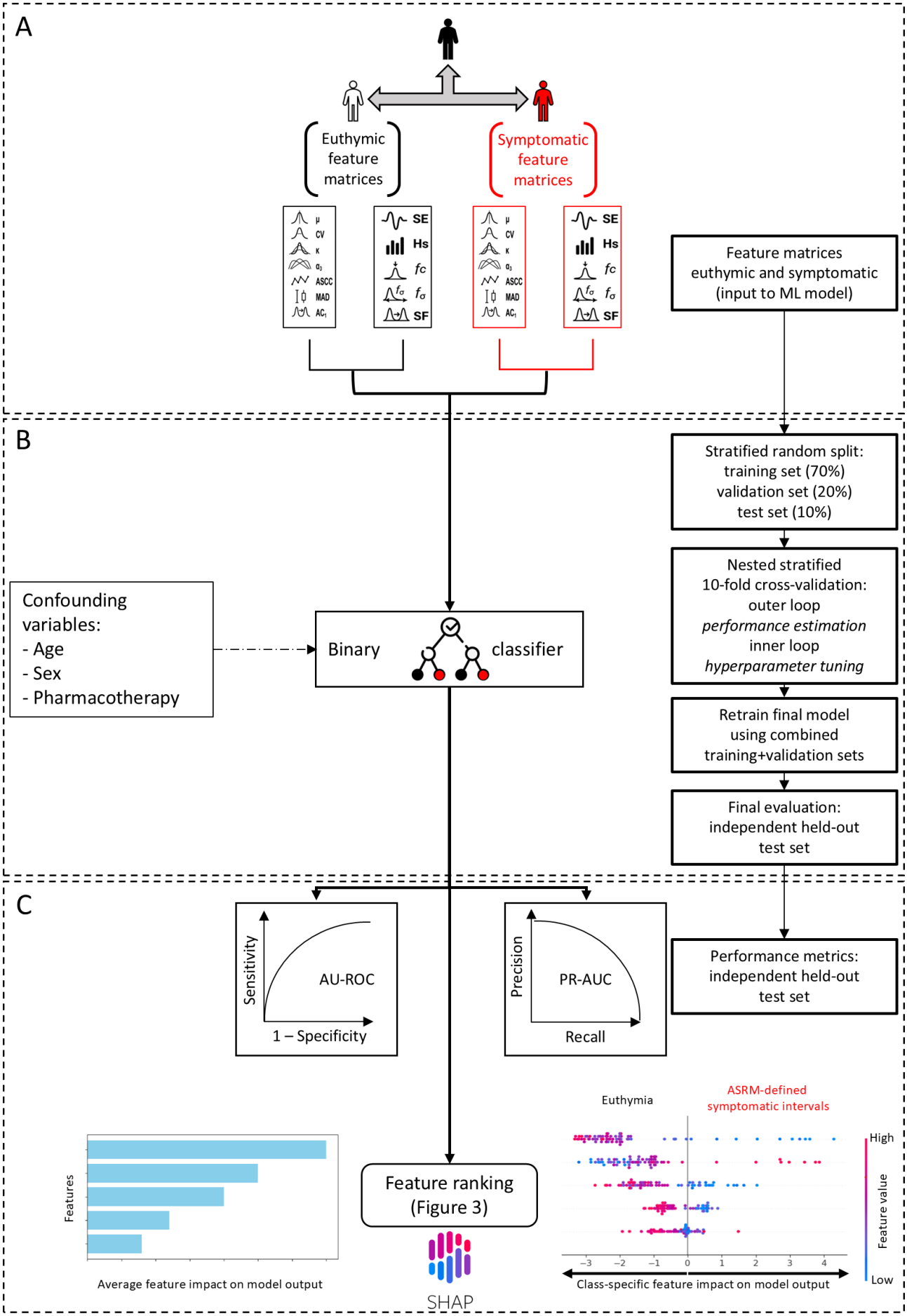
For each participant and each of the 21 variables' data segments, we calculated 12 features in the time and frequency domains. Seven time-domain features (ie, using the raw data) were calculated to characterize the statistical properties of the data segments over the entire duration of participation while euthymic and during ASRM-defined symptomatic intervals (where applicable): the mean (μ), coefficient of variation, kurtosis (κ), skewness (α_3), absolute sum of consecutive changes, mean absolute differences, and autocorrelation at lag 1 (Figure 1C; Table S2 in Multimedia Appendix 1). Moreover, we transformed the segmented time-domain variables to the frequency domain (ie, fluctuation rate) using the Fast Fourier Transform [36] and then calculated 5 frequency-domain features: spectral energy (SE), spectral entropy (H_s), spectral centroid (f_c), spectral spread (f_σ), and spectral flux (SF) (Figure 2A). These frequency-domain (ie, spectral) features were calculated to assess the overall and instantaneous fluctuation rate's central tendency, variability, and complexity. All features were normalized using minimum-maximum scaling to the range (0, 1), with scaling parameters fit only on the training data within each cross-validation fold. Outliers were clipped to the 1st and 99th percentiles of the training distribution within each fold. Feature extraction and preprocessing were implemented in Python (version 3.11; Python Software Foundation) using NumPy (version 1.24; NumPy Developers), SciPy (version 1.10; SciPy Developers), and scikit-learn (version 1.3; scikit-learn Developers).

See Table S1 in Multimedia Appendix 1 for the full list of data, variables, and features, and Table S2 in Multimedia Appendix 1 for the mathematical description and interpretation of the features.

XSL•FO
RenderX

Figure 2. Machine learning pipeline. Panels (A) model input: binary feature matrices labeled by clinical state (euthymic vs ASRM-defined symptomatic interval) from 43 participants who experienced ≥1 symptomatic interval, including 252 features and clinical and demographic covariates (age, sex, and pharmacotherapy at study entry) used for adjustment; (B) model training, testing, and validation: nested cross-validation (10 outer×10 inner folds) with group k-fold at the participant level, 70/20/10 train/validation/test split; and (C) model explainability and feature importance: SHAP values and permutation importance aggregated across outer folds and 100 bootstraps. AC_1: autocorrelation at lag 1; ASCC: absolute sum of consecutive changes; ASRM: Altman Self-Rating Mania Scale; AUROC: area under the receiver operating characteristic curve; CV: coefficient of variation; MAD: mean absolute differences; ML: machine learning; PR-AUC: precision-recall–area under the curve; SE: spectral energy; SF: spectral flux; SHAP: Shapley additive explanations; μ: mean; κ: kurtosis; α_3: skewness; H_s: spectral entropy; f_c: spectral centroid; f_σ: spectral spread.

XSL•FO
RenderX



XSL•FO
RenderX

Classification Model Development and Performance Evaluation

ML Model Development and Training

The present analysis was designed to assess the association between specific features and concurrently assessed clinical polarities: ML models were used to distinguish ASRM-defined symptomatic intervals from euthymic states (rather than to forecast future episodes) (Figure 2B-C). We constructed binary feature matrices labeled by clinical polarity (euthymic vs ASRM-defined symptomatic interval) and trained an ensemble of tree-based classifiers, specifically Random Forest, Extra Trees, and XGBoost (Extreme Gradient Boosting), to discriminate states within concurrent data segments. Features were computed from the same 7-day window used to assign the ASRM-based clinical state label; accordingly, this is a retrospective state classification analysis and not a predictive or forecasting model. Data were split into training (70%), validation (20%), and test (10%) sets with stratified sampling. Nested cross-validation with 10 outer folds and 10 inner folds was used for unbiased performance estimation, hyperparameter tuning, and feature selection. To prevent information leakage from nonindependent observations, all cross-validation folds were grouped at the participant level (group k-fold), so that every data segment from a given participant was assigned entirely to either the training or the testing partition within each fold.

To evaluate the independent contribution of passive wearable data relative to active self-reports, we trained and evaluated 5 model variants on the following feature subsets: (1) passive wearable features (daily activity+covariates), (2) passive wearable features only (daily sleep features+covariates), (3) passive wearable features only (activity+sleep features combined+covariates), (4) active self-report features only (daily e-VAS mood, anxiety, and energy features +covariates), and (5) the full combined model (passive wearable+active self-report+covariates). All variants used the same nested cross-validation procedure, class-weighted training, and out-of-fold (OOF) evaluation applied to the primary analysis.

Hyperparameter tuning was performed via randomized search within the inner cross-validation folds. The search spaces were as follows: Random Forest (n_estimators: {100, 200, 500, 1000}, max_depth: {3, 5, 8, 12, None}, min_samples_split: {2, 5, 10}, min_samples_leaf: {1, 2, 4}, max_features: {"sqrt", "log2", None}); Extra Trees (n_estimators: {100, 200, 500, 1000}, max_depth: {3, 5, 8, 12, None}, min_samples_split: {2, 5, 10}); and XGBoost (n_estimators: {100, 200, 500}, max_depth: {3, 5, 7, 9}, learning_rate: {0.01, 0.05, 0.1, 0.3}, subsample: {0.7, 0.8, 1.0}, colsample_bytree: {0.7, 0.8, 1.0}, reg_alpha: {0, 0.1, 1.0}, reg_lambda: {0.5, 1.0, 5.0}) (Table S3 in Multimedia Appendix 1). The best-performing model and hyperparameter configuration were selected based on inner-fold area under the receiver operating characteristic curve (AUROC) and refit on the full outer training fold.

Class Imbalance Handling and ML Model Performance

Given the unequal number of euthymic and ASRM-defined symptomatic interval segments, we addressed class imbalance through 2 complementary approaches. First, we applied class weighting (ie, assigning weights inversely proportional to class frequency) during model training so that the minority class (symptomatic intervals) received proportionally greater influence during learning. We did not use oversampling (eg, Synthetic Minority Over-sampling Technique) or undersampling techniques, as class weighting preserves the full data distribution while adjusting for imbalance. In addition to AUROC and AUPRC (area under the precision-recall curve), we report sensitivity, specificity, positive predictive value, and negative predictive value computed at the threshold that maximizes the F_1-score across outer-fold held-out test sets, aggregated across 100 bootstrap iterations. Class prevalence (proportion of ASRM-defined symptomatic interval segments) is reported alongside AU-PRC to facilitate interpretation relative to the no-information baseline. Calibration was assessed using the Brier score and a reliability diagram (calibration curve) constructed on OOF predicted probabilities; the no-information Brier score equals the product of class prevalence and (1 − class prevalence). All bootstrapped confidence intervals were computed by resampling at the participant level (ie, all observations from a given participant were sampled as a block in each bootstrap replicate) to account for the nonindependence of repeated observations within participants. Held-out test performance is referred to as internal validation throughout, as the held-out fold was drawn from the same cohort.

Model Explainability: Permutation Feature Importance and Robustness

To interpret the contribution of individual features to the classification model, we computed SHAP (Shapley Additive Explanations) [37] values on the held-out test sets, providing model-agnostic, local explanations of feature effects (Figure 2C). We aggregated SHAP values across all outer cross-validation folds to derive an overall feature importance ranking, thereby characterizing how each feature influenced the model's outputs. SHAP values were computed using TreeSHAP applied to held-out (OOF) test observations only, ensuring that feature importance reflects model behavior on unseen data and is not inflated by in-sample fitting. SHAP values quantify each feature's marginal contribution to the model's predictions and should be interpreted as characterizing how the fitted model used the features to discriminate ASRM-defined symptomatic intervals from euthymia; they do not establish that these features causally influence mood state or constitute validated biomarkers.

To further evaluate feature robustness and importance stability, we performed permutation importance (PI) analyses on the held-out test sets, derived 3 stability metrics for each feature, and aggregated feature importance and ranking statistics across bootstraps and outer folds to quantify robustness, thereby identifying features that consistently contributed to model performance across resampling procedures. The details of our variable selection and processing, feature extraction, and ML model's robust explainability pipeline are described thoroughly in our previous manuscript [26]. Model explainability was conducted using TreeSHAP (for tree ensembles) via the SHAP library (version 0.43; SHAP Developers) using Random Forest

XSL•FO
RenderX

classifiers. All models were implemented in Python (version 3.11).

Statistical Analysis

We performed label permutation testing (n=100) to generate a null distribution of group differences between euthymia and ASRM-defined symptomatic intervals and applied bootstrapping (n=1000) to estimate variability of the test statistics, thereby addressing limitations imposed by the relatively small sample size. Group differences were assessed using the Mann-Whitney *U* test across the resampled distributions. To ensure reliable inference, we included only clinical polarities with at least 5 observations per group, a threshold chosen to maintain adequate statistical power in each bootstrap iteration. Multiple comparisons across top 5 features in Table 1 were controlled using the Benjamini-Hochberg false discovery rate correction method; statistical significance was assumed for false discovery rate–corrected *P* value <.05. We also computed Cohen *d* effect sizes to quantify differences between the euthymic and symptomatic intervals. All models were adjusted for clinical and demographic covariates (age, sex, and pharmacotherapy) at study entry (Table 2), which we included as additional input features in the classification models. Additionally, we conducted a series of sensitivity analyses by repeating the full analytical pipeline with and without imputation of missing data. We compared classifier performance, selected features, and permutation-based importance rankings across the 2 datasets to assess whether imputation significantly influenced the results.

All statistical tests are reported with the test name, test statistic, and exact *P* value (reported as *P*<.001 when the computed value is below this threshold). Effect sizes are reported as Cohen *d* with 95% CIs derived from bootstrap resampling (n=1000). Sample sizes (n) for each comparison are reported. Nonwear time was also expressed as a percentage of each participant's total follow-up duration to enable comparison across participants with varying enrollment lengths: % nonwear = (nonwear days/total follow-up days)×100.

XSL•FO
RenderX

Table . Predictive performance and clinical operating characteristics of machine learning models across time scales and data modalities for detecting symptomatic intervals in 43 participants with bipolar disorder (I or II) who experienced ≥1 Altman Self-Rating Mania Scale (ASRM)–defined symptomatic interval (ASRM score of ≥10) and were included in the machine learning analysis, recruited from outpatient clinics at Centre for Addiction and Mental Health (Toronto, Ontario) and Queen Elizabeth II Health Sciences Centre (Halifax, Nova Scotia), Canada (December 2020 to January 2025)[a].

Time scales and data types	AUROC[b] (95% CI)	AUPRC[c] (95% CI)	Balanced accuracy (95% CI)	F_1-score optimal threshold (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV[d], % (95% CI)	NPV[e], % (95% CI)	Brier score (95% CI)
Within day									
Passive wearable features: activity+covariates	0.74 (0.64 - 0.81)	0.43 (0.30 - 0.54)	0.73 (0.67 - 0.82)	0.23 (0.17 - 0.32)	0.72 (0.40 - 0.93)	0.75 (0.57 - 0.96)	0.44 (0.29 - 0.74)	0.92 (0.86 - 0.98)	0.14 (0.12 - 0.16)
Within night									
Passive wearable features: sleep+covariates	0.74 (0.66 - 0.81)	0.50 (0.37 - 0.63)	0.71 (0.66 - 0.78)	0.33 (0.21 - 0.41)	0.57 (0.39 - 0.80)	0.86 (0.65 - 0.96)	0.53 (0.29 - 0.73)	0.890 (0.848 - 0.944)	0.14 (0.11 - 0.16)
Daily									
Passive wearable features: activity+covariates	0.90 (0.84 - 0.95)	0.73 (0.61 - 0.85)	0.86 (0.80 - 0.91)	0.39 (0.20 - 0.53)	0.83 (0.70 - 0.96)	0.89 (0.82 - 0.95)	0.66 (0.50 - 0.81)	0.95 (0.92 - 0.99)	0.09 (0.07 - 0.12)
Passive wearable features: sleep+covariates	0.91 (0.86 - 0.95)	0.71 (0.58 - 0.85)	0.90 (0.85 - 0.94)	0.29 (0.26 - 0.50)	0.92 (0.83 - 0.99)	0.88 (0.81 - 0.94)	0.66 (0.51 - 0.78)	0.98 (0.95 - 0.99)	0.09 (0.06 - 0.11)
Active self-report features only: e-VAS[f] mood, energy, and anxiety features+covariates	0.93 (0.88 - 0.96)	0.75 (0.62 - 0.87)	0.89 (0.83 - 0.93)	0.44 (0.25 - 0.61)	0.85 (0.72 - 0.95)	0.92 (0.88 - 0.97)	0.73 (0.61 - 0.89)	0.96 (0.93 - 0.99)	0.08 (0.06 - 0.11)
Passive wearable features only: activity+sleep features combined+covariates	0.91 (0.86 - 0.95)	0.72 (0.58 - 0.85)	0.90 (0.85 - 0.94)	0.29 (0.18 - 0.52)	0.92 (0.80 - 1.00)	0.88 (0.82 - 0.94)	0.66 (0.52 - 0.80)	0.98 (0.95 - 1.00)	0.09 (0.06 - 0.11)
Full combined model: passive wearable+active self-report+covariates	0.92 (0.87 - 0.96)	0.75 (0.63 - 0.85)	0.89 (0.82 - 0.93)	0.34 (0.15 - 0.57)	0.89 (0.69 - 0.99)	0.88 (0.82 - 0.97)	0.67 (0.54 - 0.85)	0.97 (0.93 - 0.99)	0.08 (0.06 - 0.10)

[a]Values represent pooled out-of-fold point estimates along with 95% CIs derived from 100 iterations of participant-level block bootstrapping to account for repeated within-subject observations. Models were evaluated using nested, participant-grouped cross-validation (*StratifiedGroupKFold*) to ensure internal validation on participants unseen during training and feature preprocessing. All models included baseline demographic and clinical covariates (age, sex, and pharmacotherapy at study entry). Decision thresholds were determined per bootstrap iteration by identifying the cutoff that maximized the F_1-score on held-out test predictions. Sensitivity, specificity, positive predictive value, and negative predictive value are presented as percentages. Calibration was evaluated using the Brier score, where lower scores indicate superior probabilistic accuracy.

[b]AUROC: area under the receiver operating characteristic curve.

XSL•FO
RenderX

[c]AUPRC: area under the precision-recall curve.
[d]PPV: positive predictive value.
[e]NPV: negative predictive value.
[f]e-VAS: electronic visual analog scale.

XSL•FO
RenderX

Table . Sociodemographic and clinical characteristics of the 43 participants with bipolar disorder (I or II) who experienced ≥1 Altman Self-Rating Mania Scale (ASRM)–defined symptomatic interval (ASRM score of ≥10) and were included in the machine learning analysis, recruited from outpatient clinics at Centre for Addiction and Mental Health (Toronto, Ontario) and Queen Elizabeth II Health Sciences Centre (Halifax, Nova Scotia), Canada (December 2020 to January 2025).

Characteristics	Participants (N=43)
Age (years), mean (SD)	33.7 (9.4)
Sex assigned at birth, n (%)	
Male	14 (32.6)
Female	29 (67.4)
Self-reported gender, n (%)	
Man	13 (30.2)
Woman	22 (51.1)
Queer/gender nonconforming	3 (7)
Prefer not to disclose	5 (11.6)
Education, n (%)	
Completed high school or less	7 (16.3)
College diploma or some university	23 (53.5)
University degree	13 (30.2)
Race, n (%)	
White	20 (46.5)
Asian	10 (23.3)
Black or African Canadian	2 (4.7)
More than 1 race	3 (7)
Unknown	8 (18.6)
Marital state, n (%)	
Single	27 (62.8)
Married/partnered	14 (32.6)
Separated/divorced	1 (2.3)
Widowed	1 (2.3)
Socioeconomic status, n (%)	
Student	5 (11.6)
Full-time work	21 (48.8)
Part-time work	5 (11.6)
Social or disability assistance	2 (4.7)
Unemployed and other	10 (23.3)
Primary diagnosis, n (%)	
Bipolar disorder I	29 (67.4)
Bipolar disorder II	14 (32.6)
Rapid cycling, n (%)	15 (34.9)
Clinical state at study entry, n (%)	
Euthymic	21 (48.8)
In a depressive episode	20 (46.5)
In a hypomanic episode	2 (4.7)
Pharmacotherapy at study entry, n (%)	
None	4 (9.4)

XSL•FO
RenderX

Characteristics	Participants (N=43)
Lithium monotherapy	2 (4.7)
Anticonvulsant monotherapy	5 (11.6)
Antipsychotic monotherapy	5 (11.6)
Combination treatment	27 (62.8)

Sensitivity Analysis for ASRM Threshold

To evaluate whether the choice of ASRM threshold (≥10 vs the standard ≥6) substantially influenced the results, we repeated the complete classification pipeline, including feature extraction, model training with nested cross-validation, SHAP-based explainability, and PI, using ASRM score of ≥6 as the alternative threshold for defining symptomatic intervals. We report the number of participants and intervals identified under each threshold, the AUROC and AUPRC for each modality, and the correlation of the performance metrics between the 2 thresholds (Spearman ρ).

Segments with ASRM score of ≥10 satisfy both the ≥6 and ≥10 threshold criteria and therefore represent matched observations that are classified as symptomatic under both thresholds. On the other hand, segments with ASRM scores of 6 to 9 satisfy only the ≥6 threshold and constitute the additional unmatched observations unique to the sensitivity analysis and are therefore not classified as symptomatic under the primary ≥10 criterion.

Identification of Top Descriptors of ASRM-Defined Symptomatic Intervals

We used a supervised ML approach within a 10-fold nested cross-validation framework to classify clinical polarity (euthymia vs ASRM-defined symptomatic intervals) using activity, sleep, and e-VAS–derived features. We evaluated model performance on held-out test sets and quantified feature contributions using SHAP impact scores, complemented by a feature robustness analysis based on PI (top-k selection frequency) across 100 bootstraps with 100 permutation repeats (Figure 2).

As shown in Figures 1C and 2B, we directly compared features across multiple temporal scales within a unified analytical framework, including corresponding classification performance metrics and SHAP analyses for each time scale.

Ethical Considerations

The study was approved by the REB at the CAMH, Toronto, Ontario, Canada, and at the QEII Health Care Center, Halifax, Nova Scotia, Canada, in accordance with the Declaration of Helsinki REB number 059 - 2019. All participants consented in writing to participate in the study. All participants signed an informed consent form approved by the local REB before any research procedure was initiated. As an incentive, participants received CAD $180 (CAD $1=US $0.72 as of May 28, 2026) upon completing their first year. For the second year, they could choose between an additional $180 payment or ownership of the Oura Ring.

Results

Participant Characteristics

Of the 178 participants enrolled in the study, 43 (24.2%) experienced at least 1 ASRM-defined symptomatic interval during the study period, totaling 133 intervals. Participants were followed for a mean of 540.1 (SD 170.9; range 120 - 706) days; of these 43 participants, 21 (48.8%) entered the study euthymic, 20 (46.5%) in a depressive episode, and only 2 (4.7%) in a hypomanic episode. Using the operationalized euthymia definition (ASRM score of <10 and PHQ-9<10 on 2 consecutive ratings separated by at least 7 days), 4774 euthymic segments were identified across participants; 1216 met the duration (>7 days) and wearable completeness (≥80%) criteria for inclusion as comparator euthymic segments. The 43 participants who experienced at least 1 symptomatic interval experienced a mean of 3.1 (SD 2.6; range 1 - 10) ASRM-defined symptomatic intervals, lasting a mean of 10.7 (SD 11.1) days. Table 2 summarizes the sociodemographic and clinical characteristics of these 43 participants. The distribution of ASRM-defined symptomatic intervals (ASRM score of ≥10) per participant is presented in Figure S1 in Multimedia Appendix 1.

Using objective and subjective data, we extracted 252 features characterizing variability in objective (ie, wearable-derived) and subjective (eg, mood) variables during euthymic periods versus symptomatic intervals (see Methods section and Table S1 in Multimedia Appendix 1).

Adherence and Data Completeness

Across all participants, nonwear accounted for a median of 19.9% (IQR 6.5 - 42.0) of the follow-up duration, corresponding to a median wearable adherence of 80.1% (IQR 58.0 - 93.5). Median daily self-reported e-VAS adherence was 82.6% (IQR 60.2 - 94.1), and median weekly self-reported ASRM adherence was 81.2% (IQR 58.1 - 91.9) (Table S4 in Multimedia Appendix 1).

Across pooled follow-up observations, missingness was 15.0% for daily activity, 23.8% for sleep, 23.9% for daily mood ratings, and 26.2% for weekly ASRM assessments. At the participant level, median missingness was 8.1% (IQR 1.1% - 28.0%) for daily activity, 20.4% (IQR 6.8% - 42.8%) for sleep, 17.4% (IQR 5.9% - 39.8%) for daily mood ratings, and 18.8% (IQR 8.1% - 41.9%) for weekly ASRM assessments (Table S5 in Multimedia Appendix 1). When stratified by clinical state, missingness was lower during ASRM-defined symptomatic intervals compared with euthymic periods (activity: 11.4% vs 15.1%, sleep: 19.4% vs 23.9%, and e-VAS: 7.8% vs 24.2%). Missingness in activity and e-VAS differed significantly between clinical states at the participant level (Wilcoxon signed-rank test: activity, P=.016; e-VAS, P<.001), whereas

XSL•FO
RenderX

the difference in sleep missingness was not significant (P=.472) (Table S6 in Multimedia Appendix 1). Little MCAR test rejected the null hypothesis of MCAR for both the daily passive/active stream matrix (χ^2_{1031}=20696.27, P<.001) and the weekly assessment matrix (χ^2_{15}=29.70, P=.013), confirming that the data are not MCAR and that missingness is associated with clinical state (Table S7 in Multimedia Appendix 1).

Identification of Top Descriptors of ASRM-Defined Symptomatic Intervals

Daily features consistently outperformed features derived at other temporal scales. Accordingly, we present results for daily features in the main text and report results for other time scales (Figures S2 and S3 in Multimedia Appendix 1).

The daily activity feature (Figure 3A-C) that best described an ASRM-defined symptomatic interval was lower activity spectral entropy (ie, consistently more predictable daily activity patterns: PI=0.25; P<.001; Cohen d=1.92, 95% CI 1.56-2.36) (Table 1). Moreover, symptomatic intervals showed significantly lower daily activity SF (ie, lower day-to-day change in behavioral rhythms: PI=0.02; P<.001; Cohen d=1.46, 95% CI 1.24-1.67) and significantly lower activity SE (ie, weaker rest-activity cycle rhythmicity: PI=0.01; P<.001; Cohen d=1.01, 95% CI 0.89-1.15).

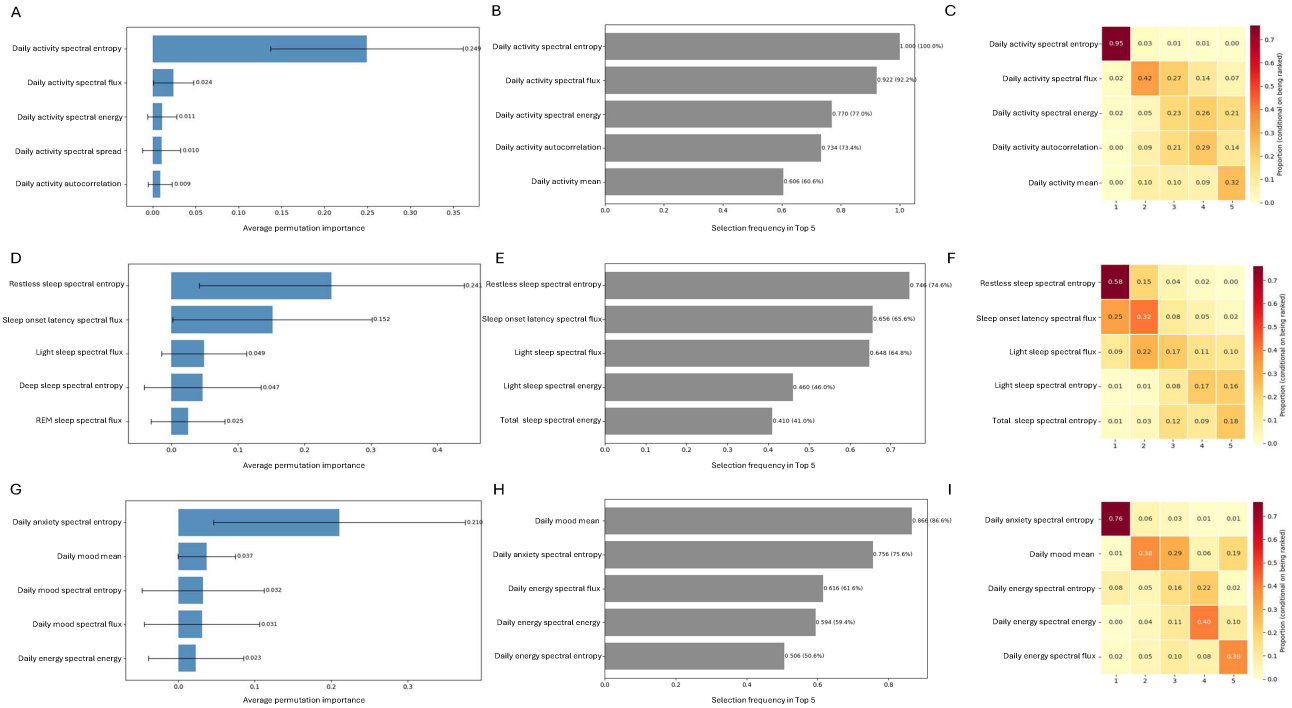
The daily sleep features (Figure 3D-F) that best described an ASRM-defined symptomatic interval were lower restless sleep spectral entropy (ie, consistently higher restlessness during sleep; PI=0.24; P<.001; Cohen d=2.24, 95% CI 1.79-2.74) and lower deep sleep spectral entropy (ie, reduced complexity and variability of deep sleep duration; PI=0.05; P<.001; Cohen d=2.19, 95% CI 1.73-2.76) (Table 1). In addition, ASRM-defined symptomatic intervals were characterized by significantly lower sleep onset latency SF (ie, reduced night-to-night variability in sleep onset; PI=0.15; P<.001; Cohen d=1.33, 95% CI 1.17-1.49), light sleep SF (ie, consistently longer and more predictable light sleep durations; PI=0.05; P<.001; Cohen d=1.43, 95% CI 1.22-1.65), and REM sleep SF (ie, consistently shorter and more predictable REM sleep durations; PI=0.03; P<.001; Cohen d=1.29, 95% CI 1.10-1.47), indicating consistently lower variability across sleep stages.

The daily e-VAS features (Figure 3G-I) that best described ASRM-defined symptomatic intervals were anxiety spectral entropy (PI=0.21; P<.001; Cohen d=1.92, 95% CI 1.53-2.38) and mood spectral entropy (PI=0.03; P<.001; Cohen d=1.89, 95% CI 1.48-2.36), both significantly lower during symptomatic intervals, associated with reduced day-to-day complexity in self-reported anxiety and mood (Table 1). Symptomatic intervals were also characterized by significantly lower mood SF (ie, lower day-to-day change in mood rhythms; PI=0.03; P<.001; Cohen d=1.38, 95% CI 1.13-1.60), alongside significantly higher mean mood levels (PI=0.04; P<.001; Cohen d=−0.76, 95% CI −1.16 to −0.35). Similarly, energy spectral entropy was significantly lower during symptomatic intervals (PI=0.02; P<.001; Cohen d=1.92, 95% CI 1.50-2.40), indicating consistently higher and less complex energy levels during symptomatic intervals.

The results of the statistical comparison of the top 5 daily features of activity, sleep, and e-VAS, during euthymia or an ASRM-derived symptomatic interval, and their average PI are presented in Table S7 in Multimedia Appendix 1. The absolute and clinical state-specific permutation feature importance of the aggregated passive data features (ie, sleep and activity) is presented in Figure S5 in Multimedia Appendix 1, and those of the combined active self-report and passive data features are presented in Figure S6 in Multimedia Appendix 1. The ASRM threshold sensitivity analysis (ASRM score of ≥6 vs ASRM score of ≥10) revealed moderate overlap in the most robust descriptors of ASRM-defined symptomatic intervals, namely, in light sleep features (Figure S7 in Multimedia Appendix 1), which is expected upon relaxing the ASRM symptom severity cutoff and including less symptomatic intervals in the analysis (Figure S8 in Multimedia Appendix 1).

XSL•FO
RenderX

Figure 3. Feature importance results for daily activity, sleep, and electronic visual analog scale features in a longitudinal cohort of 43 outpatients with bipolar disorder who experienced ≥1 Altman Self-Rating Mania Scale–defined symptomatic interval (N=133 intervals). Panels (A, D, and G): feature importance plot showing average permutation importance values with error bars representing 95% CI of bootstrapped samples. Panels (B, E, and H): feature selection frequency bar plot showing the percentage of times each feature was selected among the top predictors across models. Panels (C, F, and I): feature rank distribution heat maps showing feature ranking probabilities, with darker colors indicating higher likelihood of being selected in the top rank positions. (A-C): activity features, (D-F): sleep features, and (G-I): electronic visual analog scale features. Results are from nested cross-validation with group k-fold at the participant level.



Classification Performance

Primary Classification: ASRM Score of ≥10

Classification performance increased significantly when aggregating sensor data to daily 24-hour summaries compared with fine-grained sub–daily intervals (within-day activity: AUROC 0.74, 95% CI 0.64 - 0.81 and Brier score of 0.14, 95% CI 0.12 - 0.16; within-night sleep: AUROC 0.74, 95% CI 0.66 - 0.81 and Brier score of 0.14, 95% CI 0.11 - 0.16). The prevalence of ASRM-defined symptomatic intervals in the daily analytic sample was 20.4%, yielding a no-information AUPRC of 0.20. Among daily data modalities, active daily self-report data (e-VAS mood, energy, and anxiety) yielded the highest point estimates for overall discrimination (AUROC 0.93, 95% CI 0.88 - 0.96; AUPRC 0.75, 95% CI 0.62 - 0.87; specificity 91.9%, 95% CI 0.88 - 0.97; positive predictive value 72.9%, 95% CI 0.61 - 0.89; Brier score of 0.08, 95% CI 0.06 - 0.11). However, passive wearable data models achieved broadly comparable discrimination with substantially overlapping confidence intervals, while demonstrating distinct advantages for continuous detection. Specifically, combined daily passive wearable data features (activity and sleep) achieved higher point-estimate sensitivity (92.0%, 95% CI 0.80 - 1.00 vs 85.4%, 95% CI 0.72 - 0.95), higher balanced accuracy (0.90, 95% CI 0.85 - 0.94 vs 0.89, 95% CI 0.83 - 0.93), stronger rule-out usefulness (negative predictive value 97.8%, 95% CI 0.95 - 1.00 vs 96.1%, 95% CI 0.93 - 0.99), and similar risk prediction error (Brier score of 0.09, 95% CI 0.061 - 0.108) to active self-report data, with daily sleep features alone driving most of this passive signal (AUROC 0.91, 95% CI 0.86 - 0.95; sensitivity 91.8%, 95% CI 0.83 - 0.99; Brier score of 0.09, 95% CI 0.06 - 0.11). Integrating active self-reports and passive sensor metrics into a full multimodal model (AUROC 0.92, 95% CI 0.87 - 0.96; AUPRC 0.75, 95% CI 0.63 - 0.85; Brier score of 0.08, 95% CI 0.06 - 0.10) yielded no meaningful incremental gain over either modality alone (Table 1; Figure S9 and Table S10 in Multimedia Appendix 1).

Class weighting (models were trained with weights inversely proportional to class frequency) modestly improved minority-class performance: across nested coefficient of variation, the AUPRC increased by a median of 0.07, balanced accuracy by 0.06 with weighting, while AUROC increased the least by a median of 0.01, indicating weighting boosted sensitivity/precision for ASRM-defined symptomatic intervals without substantially altering overall discrimination.

For all top-performing daily models, AUROC values were significantly higher than those obtained under random label assignment, as demonstrated by permutation testing (all P<.01). In contrast, models based on within-day features alone achieved lower performance than daily features while consistently outperforming random-chance models (within-night sleep: AUROC, mean 0.72, SD 0.12, P<.05; within-day activity: mean 0.64, SD 0.11, P<.05). Statistical comparison of the top 5 daily features of activity, sleep, and e-VAS during euthymia and during symptomatic intervals are reported in Table S8 in Multimedia Appendix 1, and permutation-based performance comparisons are shown in Figure S4 in Multimedia Appendix 1.

XSL•FO
RenderX

Sensitivity Analysis: ASRM Score of ≥6

Using the threshold of ASRM score of ≥6, 93 participants (vs 43 when using a threshold of ASRM score of ≥10) experienced at least 1 symptomatic interval, totaling 392 intervals (vs 133 with ASRM score of ≥10). Segments with ASRM score of ≥10 (n=133) represent matched observations classified as symptomatic under both thresholds; the remaining 259 segments with ASRM scores of 6 to 9 are classified as symptomatic exclusively under the ≥6 threshold. In the sleep-only (selected data modality) ASRM-threshold sensitivity analysis using primary KNN imputation, model discrimination was comparable with the ASRM score of ≥10 and ASRM score of ≥6 thresholds (pooled OOF AUROC 0.91 vs 0.89; AUPRC 0.66 vs 0.81), with a paired difference (ASRM score of ≥6 minus ASRM score of ≥10) of 0.01 for AUROC (95% CI −0.05 to 0.08) and 0.18 for AUPRC (95% CI 0.01-0.35). However, the OOF-predicted probabilities from the 2 threshold-specific models showed expected negligible correlation at the individual-observation level (Spearman ρ=0.016; *P*=.81; 211 matched observations), indicating that although both thresholds yielded similarly separable classifications, the specific observations identified as high-risk differed substantially between thresholds. This suggests that the ASRM score of ≥10 and ASRM score of ≥6 thresholds capture a different set of symptomatic segments rather than a shared signal at differing sensitivity levels. This interpretation is based on the OOF-predicted probability correlation (ρ=0.016), a prevalence-independent measure of model agreement at the individual-observation level, rather than on the AUPRC threshold comparison (0.66 vs 0.81), which is susceptible to between-threshold differences in positive class prevalence and no-information AUPRC baseline.

Sensitivity Analysis: Missing Data Imputation

Using daily sleep features as sample representative inputs to a Random Forest classifier, OOF-predicted probabilities were highly correlated between KNN and mean imputation (Spearman ρ=0.984; *P*<.001; 211 matched observations) and remained strongly correlated between KNN imputation and complete-case analysis (ρ=0.91; *P*<.001; 186 matched observations), supporting the robustness of model performance to the missing data strategy. Using KNN imputation (k=5), the model achieved mean AUROC 0.91 (SD 0.07) across folds and pooled AUROC 0.90 (95% CI 0.84 - 0.95); pooled AUPRC was 0.66 (95% CI 0.52 - 0.82) versus a no-information AUPRC of 0.20. Mean imputation produced similar discrimination (AUROC mean 0.92, SD 0.05; pooled AUROC 0.91, 95% CI 0.86 - 0.95; pooled AUPRC 0.64, 95% CI 0.51 - 0.81). Complete-case analysis gave comparable AUROC (mean 0.90, SD 0.12; pooled 0.90, 95% CI 0.83 - 0.96) but a wider AUPRC CI and lower pooled AUPRC point estimate (0.59, 95% CI 0.42 - 0.81) (Table S9 in Multimedia Appendix 1). Moreover, no changes were observed across the top 5 sleep features best characterizing ASRM-defined symptomatic intervals from euthymia, further proving the robustness of the extracted digital markers to imputation strategies.

Discussion

Principal Findings

The primary aim of this study was to identify the most accurate and robust digital markers for detecting ASRM-defined symptomatic intervals. We systematically analyzed multivariate, densely sampled data that combined continuously collected objective data (ie, from a wearable) and subjective data (ie, daily and weekly self-reported measures). We analyzed more than 250 features extracted from 43 participants who experienced at least 1 ASRM-defined symptomatic interval during the study period, totaling 133 intervals.

Our analysis revealed a distinctive pattern: ASRM-defined symptomatic intervals are characterized by reduced variability across all measured domains (ie, activity, sleep, and e-VAS), rather than simply elevated activity levels. In other words, not only were mood and energy levels elevated during ASRM-defined symptomatic intervals but their day-to-day patterns also became more stationary (ie, less variable). Moreover, our analysis revealed that complexity measures derived from frequency-domain analysis outperform traditional time-domain statistics, and spectral entropy emerged as the most robust discriminator across mood, energy, anxiety, sleep, and activity domains.

Our findings showed that the complexity of sleep patterns is disrupted during symptomatic intervals, in accordance with a recent systematic review discriminating unipolar from bipolar depression [38]. Our work extends these findings by demonstrating that frequency-domain features capturing temporal dynamics are more informative than simple duration measures (eg, total sleep time).

In parallel, reduced complexity in self-reported mood and anxiety emerged as highly discriminative features. Mean mood levels were elevated during symptomatic intervals, but with significantly reduced variability compared with euthymia, illustrating the phenomenon of elevated mood with reduced variability (ie, participants were consistently happier than usual). This finding aligns with recent studies demonstrating that entropy-based measures can effectively classify mood states [39-41]. On the other hand, reduced variability in anxiety during ASRM-defined symptomatic intervals is in keeping with clinical observations and consistent with theoretical models proposing hypomanic or manic episodes as a persistent drive to seek rewards, with reduced flexibility to disengage from these behaviors [42].

Similarly, ASRM-defined symptomatic intervals are characterized by more predictable and less complex daily activity patterns. This finding aligns with prior research showing dysregulation of 24-hour motor activity patterns in BD [43]. Furthermore, our findings suggest that activity becomes less adaptive during symptomatic intervals (ie, consistently higher and more predictable). SF and SE of activity were also significantly reduced during symptomatic intervals, further supporting that activity patterns become more rigid [15].

Finally, temporal scale analysis revealed that daily feature aggregation consistently outperformed finer temporal resolutions

XSL•FO
RenderX

(ie, within-day features [AUROC 0.76]), suggesting that day-to-day variability captures the most discriminatory and clinically relevant patterns of polarity changes. These findings are consistent with prior research showing that appropriate temporal aggregation improves predictive performance [44,45].

Overall, the consistent pattern of reduced complexity across domains (ie, activity, sleep, and e-VAS) is associated with a pattern consistent with "pathological stability" in underlying regulatory systems during ASRM-defined symptomatic intervals, which may reflect a shared underlying process affecting emotional regulation broadly [46]. These findings are consistent with, but do not establish, a mechanistic explanation.

Strengths and Limitations

This study represents a rigorous study of digital markers for ASRM-defined symptomatic intervals in BD, jointly extracting signature patterns in activity, sleep, mood, energy, and anxiety from more than 100 symptomatic intervals. Several limitations in our study require consideration. First, our participants had relatively high educational levels and limited racial diversity, which constrains generalizability. Future studies should validate these findings in more diverse samples. Second, the current analysis is associative rather than predictive (ie, features are concurrent with ASRM-defined symptomatic intervals). Third, symptomatic intervals were defined by ASRM self-ratings rather than clinical adjudication, which are subject to recall bias, insight limitations during elevated mood states, and variable item endorsement across participants and time points. Clinician interviews or clinician-administered ratings (eg, YMRS) were not routinely performed within each interval; accordingly, the outcome should be interpreted as an "ASRM-defined symptomatic interval" rather than a confirmed hypomanic or manic episode, and findings may not directly generalize to clinician-confirmed diagnoses. Fourth, the euthymic comparator was operationalized using ASRM score of <10 and PHQ-9 <10 thresholds, which reduce but do not eliminate the possibility of residual subsyndromal symptoms. Fifth, sleep staging by the Oura Ring is produced by a proprietary algorithm trained primarily on nonclinical populations; device-derived sleep stages may deviate from polysomnographic standards, particularly during mood episodes when sleep architecture is disrupted. The 5-minute hypnogram should therefore be treated as an approximation rather than a validated measure of sleep staging in BD. However, a major strength of this analysis is the length of follow-up and the value that longitudinal studies offer to the field (ie, after 2 years of follow-up, 25% of the sample had at least 1 ASRM-defined symptomatic interval). Another major strength is the use of explainable ML with systematic feature selection and validation coupled with a rigorous statistical analysis. The additional analysis of all the studied features using traditional statistics identified the features that statistically differentiated ASRM-defined symptomatic intervals from periods of euthymia. However, some features did not contribute as much to the explainable ML retrospective state classifications because they were redundant in their contribution to the models' outputs (ie, even after managing multicollinearity, their SHAP importance was distributed among correlated features; eg, activity spectral spread).

The choice of ASRM threshold warrants discussion. We used a threshold ≥10 rather than the standard ≥6 to enhance specificity, given the ASRM's known high sensitivity and low specificity. This choice may limit direct comparability with studies using ≥6, as we may have excluded milder or subsyndromal presentations. However, our sensitivity analysis using ≥6 showed comparable clinical state classification results, suggesting that the identified digital markers are robust to threshold choice and identifiable at comparable accuracy and precision, yet detected across different sets of symptomatic intervals. Future studies should systematically evaluate the optimal threshold for digital phenotyping applications (eg, the degree to which the ≥6 and ≥10 thresholds capture distinct symptom presentations), potentially using receiver operating characteristic analysis to derive data-driven thresholds.

In addition, we introduced and validated a novel set of frequency-domain features and demonstrated their robustness. The use of nested cross-validation, PI analysis, and bootstrapped feature ranking, complemented by data-quality sensitivity analyses, further provided confidence in the robustness of the identified markers. This within-participant analytical framework allowed each individual to serve as their own control, reducing between-participant heterogeneity.

Conclusions

This systematic analysis identified a small set of robust digital markers across activity, sleep, and e-VAS domains to detect ASRM-defined symptomatic intervals using objective and subjective data. Our main finding is that passive monitoring using wearable technology matched self-reported data in detecting ASRM-defined symptomatic intervals. This finding is particularly important during hypomanic or manic episodes, as symptoms tend to be underreported. Furthermore, the finding that symptomatic intervals are characterized by decreased variability is consistent with nonlinear dynamics of mood regulation and warrants further prospective investigation. Our study also highlights the importance of study designs with longer follow-ups to capture symptoms that would otherwise go undetected with the conventional weeks-long observation windows in digital phenotyping studies. Together, our findings point to a future in which passively collected digital markers, embedded within a strong clinical foundation, enable us to improve the detection of mood episodes.

Generative AI Use

Generative AI tools were not used for language editing or manuscript formatting assistance during the preparation of this manuscript, nor to generate any of the text, tables, or figures for this paper. All scientific content, data analysis, interpretation of results, and intellectual contributions were conducted solely by the authors, who reviewed and take full responsibility for all content in this

XSL•FO
RenderX

manuscript. Generative AI tools were used to improve the formatting of the penultimate version, as well as that of the response to the reviewers.

Funding

This study was funded by the National Institute of Mental Health grant 1R21MH123849-01A1 (AO) and by the Canadian Institutes of Health Research grant 02010PJT-450770-BSB-CEAH-188794 (AO). AO has also received support from the John Sloan (Trey) Leonard III Fund for Bipolar Disorder Research endowment. The work of MA on the project has been supported by the Ministry of Health of the Czech Republic, grant number NU23-04-00534. The rest of the authors' work received no funding. The funders played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Data Availability

The datasets generated and/or analyzed during this study are not publicly available due to privacy considerations. The underlying code for this study is not publicly available but may be made available to qualified researchers on reasonable request from the corresponding author.

Authors' Contributions

RH contributed to data management, development and implementation of signal processing and machine learning algorithms, original draft writing, formal analysis, and data curation. MA contributed to clinical interpretation, validation of results, and manuscript reviewing and editing. MT contributed to statistical analysis, manuscript reviewing, and editing. CO contributed to manuscript reviewing and editing. AH contributed to methodology, validation, and manuscript reviewing and editing. BHM contributed to clinical interpretation, validation of results, and manuscript reviewing and editing. AO contributed to study conceptualization, manuscript writing, reviewing and editing, methodology and results validation, project supervision and management, resource provision, and funding acquisition.

Conflicts of Interest

BHM holds, and receives support from, the Labatt Family Chair in Biology of Depression in Late-Life Adults at the University of Toronto. He also receives compensation from the Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada; the Centre for Addiction and Mental Health (CAMH), Toronto, Ontario; and The Royal, Ottawa Hospital, Ottawa, Ontario, Canada. He is a member of the Board of Trustees of The Royal Ottawa Hospital. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Supplementary file containing figures and tables.
[DOCX File, 1464 KB - mhealth_v14i1e105647_app1.docx]

References

1. Regeer EJ, Kupka RW, Have MT, Vollebergh W, Nolen WA. Low self-recognition and awareness of past hypomanic and manic episodes in the general population. Int J Bipolar Disord 2015 Dec;3(1):22. [doi: 10.1186/s40345-015-0039-8] [Medline: 26440507]
2. Bauer M, Andreassen OA, Geddes JR, et al. Areas of uncertainties and unmet needs in bipolar disorders: clinical and research perspectives. Lancet Psychiatry 2018 Nov;5(11):930-939. [doi: 10.1016/S2215-0366(18)30253-0] [Medline: 30146246]
3. Tania G, Mariam AJ, Shivangi T, Haleemah A, Talen W. The complexity of delayed diagnosis in bipolar disorder. a systematic review of associated precursors, outcomes, and suggested avenues for improvement. medRxiv. Preprint posted online on Oct 14, 2025. [doi: 10.1101/2025.10.13.25337871]
4. Ghaemi SN, Stoll AL, Pope HG. Lack of insight in bipolar disorder. The acute manic episode. J Nerv Ment Dis 1995 Jul;183(7):464-467. [doi: 10.1097/00005053-199507000-00007] [Medline: 7623019]
5. Torous J, Linardon J, Goldberg SB, et al. The evolving field of digital mental health: current evidence and implementation issues for smartphone apps, generative artificial intelligence, and virtual reality. World Psychiatry 2025 Jun;24(2):156-174. [doi: 10.1002/wps.21299] [Medline: 40371757]
6. Hirschtritt ME, Insel TR. Digital technologies in psychiatry: present and future. Focus (Am Psychiatr Publ) 2018 Jul;16(3):251-258. [doi: 10.1176/appi.focus.20180001] [Medline: 31975919]
7. Insel TR. Digital phenotyping: a global tool for psychiatry. World Psychiatry 2018 Oct;17(3):276-277. [doi: 10.1002/wps.20550] [Medline: 30192103]
8. Ortiz A, Halabi R, Alda M, et al. Day-to-day variability in sleep and activity predict the onset of a hypomanic episode in patients with bipolar disorder. J Affect Disord 2025 Apr 1;374:75-83. [doi: 10.1016/j.jad.2025.01.026] [Medline: 39793618]
9. Song YM, Jeong J, de Los Reyes V AA 5th, et al. Causal dynamics of sleep, circadian rhythm, and mood symptoms in patients with major depression and bipolar disorder: insights from longitudinal wearable device data. eBioMedicine 2024 May;103:105094. [doi: 10.1016/j.ebiom.2024.105094]

XSL•FO
RenderX

10. Lipschitz JM, Lin S, Saghafian S, Pike CK, Burdick KE. Digital phenotyping in bipolar disorder: using longitudinal Fitbit data and personalized machine learning to predict mood symptomatology. Acta Psychiatr Scand 2025 Mar;151(3):434-447. [doi: 10.1111/acps.13765] [Medline: 39397313]
11. Lim D, Jeong J, Song YM, et al. Accurately predicting mood episodes in mood disorder patients using wearable sleep and circadian rhythm features. NPJ Digit Med 2024 Nov 18;7(1):324. [doi: 10.1038/s41746-024-01333-z] [Medline: 39557997]
12. Wu CT, Hsieh MH, Chen IM, et al. Using wearable device and machine learning to predict mood symptoms in bipolar disorder: development and usability study. JMIR Med Inform 2025 Sep 16;13:e66277. [doi: 10.2196/66277] [Medline: 40957006]
13. Cho CH, Lee T, Kim MG, In HP, Kim L, Lee HJ. Mood prediction of patients with mood disorders by machine learning using passive digital phenotypes based on the Circadian rhythm: prospective observational cohort study. J Med Internet Res 2019 Apr 17;21(4):e11029. [doi: 10.2196/11029] [Medline: 30994461]
14. Ortiz A, Halabi R, Alda M, et al. Day-to-day variability in activity levels detects transitions to depressive symptoms in bipolar disorder earlier than changes in sleep and mood. Int J Bipolar Disord 2025 Apr 2;13(1):13. [doi: 10.1186/s40345-025-00379-6] [Medline: 40175826]
15. Aledavood T, Luong N, Baryshnikov I, et al. Multimodal digital phenotyping study in patients with major depressive episodes and healthy controls (mobile monitoring of mood): observational longitudinal study. JMIR Ment Health 2025 Feb 21;12:e63622. [doi: 10.2196/63622] [Medline: 39984168]
16. Pan Y, Wang P, Xue B, et al. Machine learning for the diagnosis accuracy of bipolar disorder: a systematic review and meta-analysis. Front Psychiatry 2024;15:1515549. [doi: 10.3389/fpsyt.2024.1515549] [Medline: 39935623]
17. Sun J, Lu T, Shao X, et al. Practical AI application in psychiatry: historical review and future directions. Mol Psychiatry 2025 Sep;30(9):4399-4408. [doi: 10.1038/s41380-025-03072-3] [Medline: 40456953]
18. Kerz E, Zanwar S, Qiao Y, Wiechmann D. Toward explainable AI (XAI) for mental health detection based on language behavior. Front Psychiatry 2023;14:1219479. [doi: 10.3389/fpsyt.2023.1219479] [Medline: 38144474]
19. Judd LL, Akiskal HS, Schettler PJ, et al. The long-term natural history of the weekly symptomatic status of bipolar I disorder. Arch Gen Psychiatry 2002 Jun;59(6):530-537. [doi: 10.1001/archpsyc.59.6.530] [Medline: 12044195]
20. Altman EG, Hedeker D, Peterson JL, Davis JM. The Altman Self-Rating Mania Scale. Biol Psychiatry 1997 Nov 15;42(10):948-955. [doi: 10.1016/S0006-3223(96)00548-3] [Medline: 9359982]
21. Ortiz A, Hintze A, Burnett R, et al. Identifying patient-specific behaviors to understand illness trajectories and predict relapses in bipolar disorder using passive sensing and deep anomaly detection: protocol for a contactless cohort study. BMC Psychiatry 2022 Apr 22;22(1):288. [doi: 10.1186/s12888-022-03923-1] [Medline: 35459150]
22. First MB. Structured clinical interview for the DSM (SCID). In: Cautin RL, Lilienfeld SO, editors. The Encyclopedia of Clinical Psychology: Wiley:1-6 URL: https://onlinelibrary.wiley.com/doi/10.1002/9781118625392.wbecp351 [accessed 2026-10-29]
23. Diagnostic and Statistical Manual of Mental Disorders, 5th edition: American Psychiatric Publishing; 2013. [doi: 10.1176/appi.books.9780890425596]
24. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry 1978 Nov;133:429-435. [doi: 10.1192/bjp.133.5.429] [Medline: 728692]
25. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979 Apr;134:382-389. [doi: 10.1192/bjp.134.4.382] [Medline: 444788]
26. Halabi R, Mulsant BH, Tolend M, et al. A systematic exploration of digital biomarkers for the detection of depressive episodes in bipolar disorder. Npj Mental Health Res 2026;5(1):13. [doi: 10.1038/s44184-026-00195-5]
27. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
28. Halabi R, Selvarajan R, Lin Z, et al. Comparative assessment of multimodal sensor data quality collected using Android and iOS smartphones in real-world settings. Sensors 2024 Jan;24(19):6246. [doi: 10.3390/s24196246]
29. Halabi R, Mulsant BH, Alda M, et al. Not missing at random: missing data are associated with clinical status and trajectories in an electronic monitoring longitudinal study of bipolar disorder. J Psychiatr Res 2024 Jun;174:326-331. [doi: 10.1016/j.jpsychires.2024.04.036] [Medline: 38692162]
30. Meyer TD, Crist N, La Rosa N, Ye B, Soares JC, Bauer IE. Are existing self-ratings of acute manic symptoms in adults reliable and valid?—a systematic review. Bipolar Disord 2020 Sep;22(6):558-568. [doi: 10.1111/bdi.12906] [Medline: 32232950]
31. Altman E, Hedeker D, Peterson JL, Davis JM. A comparative evaluation of three self-rating scales for acute mania. Biol Psychiatry 2001 Sep 15;50(6):468-471. [doi: 10.1016/s0006-3223(01)01065-4] [Medline: 11566165]
32. Vieta E, Suppes T, Eggens I, Persson I, Paulsson B, Brecher M. Efficacy and safety of quetiapine in combination with lithium or divalproex for maintenance of patients with bipolar I disorder (international trial 126). J Affect Disord 2008 Aug;109(3):251-263. [doi: 10.1016/j.jad.2008.06.001] [Medline: 18579216]

XSL•FO
RenderX

33. Tohen M, Calabrese JR, Sachs GS, et al. Randomized, placebo-controlled trial of olanzapine as maintenance therapy in patients with bipolar I disorder responding to acute treatment with olanzapine. Am J Psychiatry 2006 Feb;163(2):247-256. [doi: 10.1176/appi.ajp.163.2.247]
34. Mucherino A, Papajorgji PJ, Pardalos PM. K-nearest neighbor classification. In: Data Mining in Agriculture: Springer; 2009:83-106. [doi: 10.1007/978-0-387-88615-2_4]
35. Li C. Little's test of missing completely at random. Stata J 2013 Dec;13(4):795-809. [doi: 10.1177/1536867X1301300407]
36. Cooley JW, Tukey JW. An algorithm for the machine calculation of complex Fourier series. Math Comp 1965;19(90):297-301. [doi: 10.1090/S0025-5718-1965-0178586-1]
37. Lundberg S, Lee SI. A unified approach to interpreting model predictions. arXiv. Preprint posted online on May 22, 2017. [doi: 10.48550/arXiv.1705.07874]
38. Zhong R, Wu X, Chen J, Fang Y. Using digital phenotyping to discriminate unipolar depression and bipolar disorder: systematic review. J Med Internet Res 2025 May 23;27:e72229. [doi: 10.2196/72229] [Medline: 40408762]
39. Langholm C, Breitinger S, Gray L, et al. Classifying and clustering mood disorder patients using smartphone data from a feasibility study. NPJ Digit Med 2023 Dec 21;6(1):238. [doi: 10.1038/s41746-023-00977-7] [Medline: 38129571]
40. Ortiz A, Bradler K, Hintze A. Episode forecasting in bipolar disorder: Is energy better than mood? Bipolar Disord 2018 Jan 22. [doi: 10.1111/bdi.12603] [Medline: 29356281]
41. Ortiz A, Bradler K, Garnham J, Slaney C, Alda M. Nonlinear dynamics of mood regulation in bipolar disorder. Bipolar Disord 2015 Mar;17(2):139-149. [doi: 10.1111/bdi.12246] [Medline: 25118155]
42. Johnson SL, Edge MD, Holmes MK, Carver CS. The behavioral activation system and mania. Annu Rev Clin Psychol 2012;8:243-267. [doi: 10.1146/annurev-clinpsy-032511-143148] [Medline: 22077912]
43. Shou H, Cui L, Hickie I, et al. Dysregulation of objectively assessed 24-hour motor activity patterns as a potential marker for bipolar I disorder: results of a community-based family study. Transl Psychiatry 2017 Aug 22;7(8):e1211. [doi: 10.1038/tp.2017.136] [Medline: 28892068]
44. Jacobson NC, Bhattacharya S. Digital biomarkers of anxiety disorder symptom changes: personalized deep learning models using smartphone sensors accurately predict anxiety symptoms from ecological momentary assessments. Behav Res Ther 2022 Feb;149:104013. [doi: 10.1016/j.brat.2021.104013] [Medline: 35030442]
45. Pratap A, Atkins DC, Renn BN, et al. The accuracy of passive phone sensors in predicting daily mood. Depress Anxiety 2019 Jan;36(1):72-81. [doi: 10.1002/da.22822] [Medline: 30129691]
46. Ortiz A, Alda M. The perils of being too stable: mood regulation in bipolar disorder. J Psychiatry Neurosci 2018 Nov 1;43(6):363-365. [doi: 10.1503/jpn.180183] [Medline: 30371048]

Abbreviations

ASRM: Altman Self-Rating Mania Scale
AUPRC: area under the precision-recall curve
AUROC: area under the receiver operating characteristic curve
BD: bipolar disorder
CAMH: Centre for Addiction and Mental Health
DSM-5: *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*
e-VAS: electronic visual analog scale
KNN: k-nearest neighbor
MCAR: missing completely at random
MET: metabolic equivalent of task
ML: machine learning
OOF: out-of-fold
PHQ-9: Patient Health Questionnaire (9-item)
PI: permutation importance
QEII: Queen Elizabeth II (Health Sciences Centre)
REB: research ethics board
REDCap: Research Electronic Data Capture
REM: rapid eye movement (sleep stage)
SCID-5: Structured Clinical Interview for the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*
SE: spectral energy
SF: spectral flux
SHAP: Shapley Additive Explanations
XGBoost: Extreme Gradient Boosting
YMRS: Young Mania Rating Scale

XSL•FO
RenderX

Edited by A Stone; submitted 26.Jun.2026; peer-reviewed by B Najand, LZ Atkinson; revised version received 05.Oct.2026; accepted 12.Oct.2026; published 18.Nov.2026.

Please cite as:
Halabi R, Alda M, Tolend M, O'Donovan C, Hintze A, Mulsant BH, Ortiz A
Passive Monitoring for Detecting Hypomanic Symptoms in Bipolar Disorder Using Explainable Machine Learning: Longitudinal Observational Study
JMIR Mhealth Uhealth 2026;14:e105647
URL: https://mhealth.jmir.org/2026/1/e105647
doi:10.2196/105647

© Ramzi Halabi, Martin Alda, Mirkamal Tolend, Claire O'Donovan, Arend Hintze, Benoit H Mulsant, Abigail Ortiz. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 18.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring Consumers' Situational Snacking Behaviors Using a Mobile App: Longitudinal Cohort Study (FOODLOOP) Among Millennials in the Netherlands

Mariëlle Tamara de Vaal, MSc; Vincenzo Fogliano, Prof Dr; Ruud Verkerk, Dr Ing; Bea L P A Steenbekkers, Dr Ir

Food Quality and Design (FQD), Wageningen University & Research, Bornse Weilanden 9, Wageningen, The Netherlands

Corresponding Author:
Mariëlle Tamara de Vaal, MSc
Food Quality and Design (FQD), Wageningen University & Research, Bornse Weilanden 9, Wageningen, The Netherlands

Abstract

Background: Consumers are increasingly moving away from the traditional 3-meal-a-day eating routine to a pattern where they are snacking throughout the day to fulfill dietary needs, a trend known as "snackification." Snacking depends on a variety of product-, context-, and consumer-specific determinants, but consumers' long-term snacking behaviors in light of these determinants have remained little studied.

Objective: This study aims to enhance understanding of consumers' snacking behaviors by capturing longitudinal real-world situational snacking behaviors. As snackification is highly prominent in the Netherlands, and especially among millennials (born between 1980 and 2000), Dutch millennials were used as a case study.

Methods: FOODLOOP studied situational snacking behaviors of a cohort of 264 Dutch millennials over the course of a year through a time-series structure. Data were collected on 12 nonconsecutive days, divided over the 4 seasons, using the mobile app Traqq (Department of Human Nutrition and Health of Wageningen University) and following the principles of ecological momentary assessment.

Results: On average, 4.52 snacks were consumed per day, of which 64% were more healthful snacks, including coffee, tea, fruit, and bread products. Snacking was mostly driven by the food choice motives liking and appetite, as well as hunger and thirst, convenience, and pleasure. Most snacking occurred at home, with others, in the afternoon, and in spring. Dutch millennials with children consumed more snacks than Dutch millennials without children, and Dutch millennials born between 1980 and 1990 consumed more snacks than Dutch millennials born between 1990 and 2000.

Conclusions: Our study shows that situational snacking determinants are essential for understanding consumers' real-world snacking behaviors, as Dutch millennials' snacking behaviors differ given product-, context-, and consumer-specific determinants, and are not restricted to specific motives, locations, social settings, or times, congruent with the ongoing snackification trend. We also demonstrate that a mobile app following the principles of ecological momentary assessment is a highly valuable methodological tool to gather longitudinal data on consumers' situational snacking behaviors.

(JMIR Mhealth Uhealth 2026;14:e71858) doi:10.2196/71858

KEYWORDS

situational snacking; contextual eating; food choice motives; mobile app; ecological momentary assessment

Introduction

Consumers' eating patterns have fundamentally changed in the past few years due to their increasingly busy lives. Many consumers are moving away from the traditional 3-meal-a-day pattern, consisting of breakfast, lunch, and dinner at predetermined times. Instead, consumers are moving toward a more flexible eating pattern, where snacking is becoming a new way of eating; consumers are increasingly having unplanned eating moments outside of meals, replacing meals with mini-meals and snacks, and snacking throughout the day to fulfill dietary needs [1-3]. This new eating pattern, where anything can be a snack, consumed anywhere, with anyone, and anytime, is known as "snackification" and is considered to stay [4,5].

An increase in snacking has often been linked to a poorer diet quality, body weight gain, and subsequent deterioration of health (cardiovascular diseases, obesity, etc), as snacks have traditionally been considered to consist of highly palatable "junk food" products (eg, cookies, confectionery, and chips), which contain high levels of calories, salt, sugar, and (saturated) fat [2,6]. In the new snackification frame, this view might be obsolete. Many consumers no longer merely snack on conventional highly palatable food products but on any type of food or beverage, including fruits, vegetables, and dairy products [3], which are tailored to different specific situations and support

XSL•FO
RenderX

consumers' health goals [7,8]. Therefore, snacking nowadays could also fit within a healthy dietary pattern and potentially benefit health by contributing to the daily intake of fibers, vitamins, minerals, and protein [3,7,9-13].

Snackification is a worldwide phenomenon, but the increase in snacking is especially pronounced in the Netherlands; the Dutch engage in, on average, 4-5 snacking occasions daily, which make up half of all daily consumption and 30% of the Dutch daily energy intake [14,15]. Snackification in the Netherlands, as well as in other countries, is mainly prominent among millennials [16,17], a generation born between 1980 and 2000 and characterized by the use and adaptation of technology [18]. Millennials have different life stages, undergoing significant lifestyle changes, such as leaving home, graduating, getting a job, and starting a family. Due to their increasingly busy lives and lifestyle changes, millennials often engage in snacking more than other generations [16] and seek snacks that support different situations and their health goals, in line with snackification [16,19,20]. In addition, millennials also influence the behaviors of other generations by spotting, adopting, and spreading consumer trends such as snackification [18]. Therefore, it is interesting to focus on millennials when studying snacking, since there is little knowledge to date about millennials' snacking behaviors.

Multiple studies have been devoted to snacking. However, snacking and snacking-related concepts are often defined in different ways, using varying classifications (based on the time of day, type of food consumed, location of consumption, portion sizes, caloric values, nutrient quality, healthfulness, etc), therewith including or excluding certain products (eg, beverages, fruits, and vegetables) and consumption occasions (eg, meal replacements). This definition variability contributes to conflicting and incomparable findings around snacking behaviors and adds an additional uncertainty to the translation of scientific messages on snacking behaviors into effective snacking interventions, such as guidelines or recommendations [21-23]. To enable the formulation of successful snacking interventions, consumers' snacking behaviors in different consumption situations need to be more fully understood [22,24].

Snacking, like any other dietary behavior, is situational; whether, what, where, when, with whom, how long, how, and how much to eat and/or drink can differ greatly in every snacking situation [14,25,26]. Snacking depends on a variety of determinants, including product-specific (eg, product type and portion size), context-specific (physical, social, and temporal contexts), and consumer-specific determinants. Consumer-specific determinants, including sociodemographic (eg, gender, age, socioeconomic status, education, and family composition) and lifestyle characteristics (health interest, activity level, diet, etc), also comprise mental processes, such as the motives for consumption [3,14,27-29]. These motives, known as food choice motives (FCM), are essential for understanding snacking behaviors and include physiological (eg, hunger or thirst), psychological (eg, affect regulation and pleasure), sociocultural (eg, traditional eating and sociability), practical (eg, convenience and choice limitation), and economic (eg, price) drivers, which shape everyday snacking consumption decisions [26,30].

Given the situational nature of snacking, there is a need for real-time, context-sensitive insights into consumers' snacking behaviors. However, most snacking-related research focused on how snacking affects energy balance and weight status [3] or studied consumers' snacking behaviors in isolation, in short time frames, and with a limited consideration of situational determinants [31]. To address this gap, we aim to provide a broader and more integrated understanding of consumers' snacking behaviors by longitudinally exploring Dutch millennial consumers' situational snacking behaviors given product-, context-, and consumer-specific determinants. We, therewith, extend prior work on snacking behaviors by capturing real-world snacking situations while accounting for variances given situational determinants.

Studying situational behaviors longitudinally poses a challenge in using suitable and effective methods for correct measurement and understanding [31]. Given recent advancements in mobile technologies and the proven benefits of apps for mobile health (mHealth) purposes (eg, facilitating dietary monitoring and supporting health education and policies), mobile apps are increasingly used among behavioral and nutrition researchers for answering key food and health-related research questions and therewith achieving greater understanding of dietary behaviors [26,31-34]. Mobile apps allow researchers to capture context-rich real-world insights into dietary behaviors, increase compliance, reduce missing data, improve data quality and accuracy, and reduce costs, participant burden, and recall bias [26,31,35,36]. Mobile apps could, thereby, generate large amounts of rich data that offer a detailed description of situational dietary behaviors [26,31] over longer time periods with relative ease [34]. Ecological momentary assessment (EMA) methods leverage the capabilities of mobile apps while enabling the assessment of dietary behaviors in real-world settings [37].

Notwithstanding these promising aspects, few studies to date have used mobile apps to explore dietary behaviors considering situational variances over a longer period of time [36]. Given the objectives of this study and the promising aspects of mobile apps and EMA for exploring real-world dietary behaviors, a mobile app was used to repeatedly and comprehensively sample situational snacking behaviors. This study, called FOODLOOP, used the mobile app Traqq (Traqq Pty Ltd) with EMA to support the detailed exploration of longitudinal situational snacking behaviors of a cohort of Dutch millennials. Specifically, FOODLOOP captured how real-world snacking behaviors vary within and across snacking situations given a comprehensive combination of product-, context-, and consumer-specific determinants. Even though the FOODLOOP study is descriptive in nature, it was guided by the expectations that (1) Dutch millennials would snack frequently; (2) snacking would show substantial variability both across and within a wide range of contexts, including locations, social settings, the time of day, days of the week, and the 4 seasons in a year; and (3) snacking would differ based on key consumer-specific characteristics.

XSL•FO
RenderX

Methods

Operationalization of Concepts

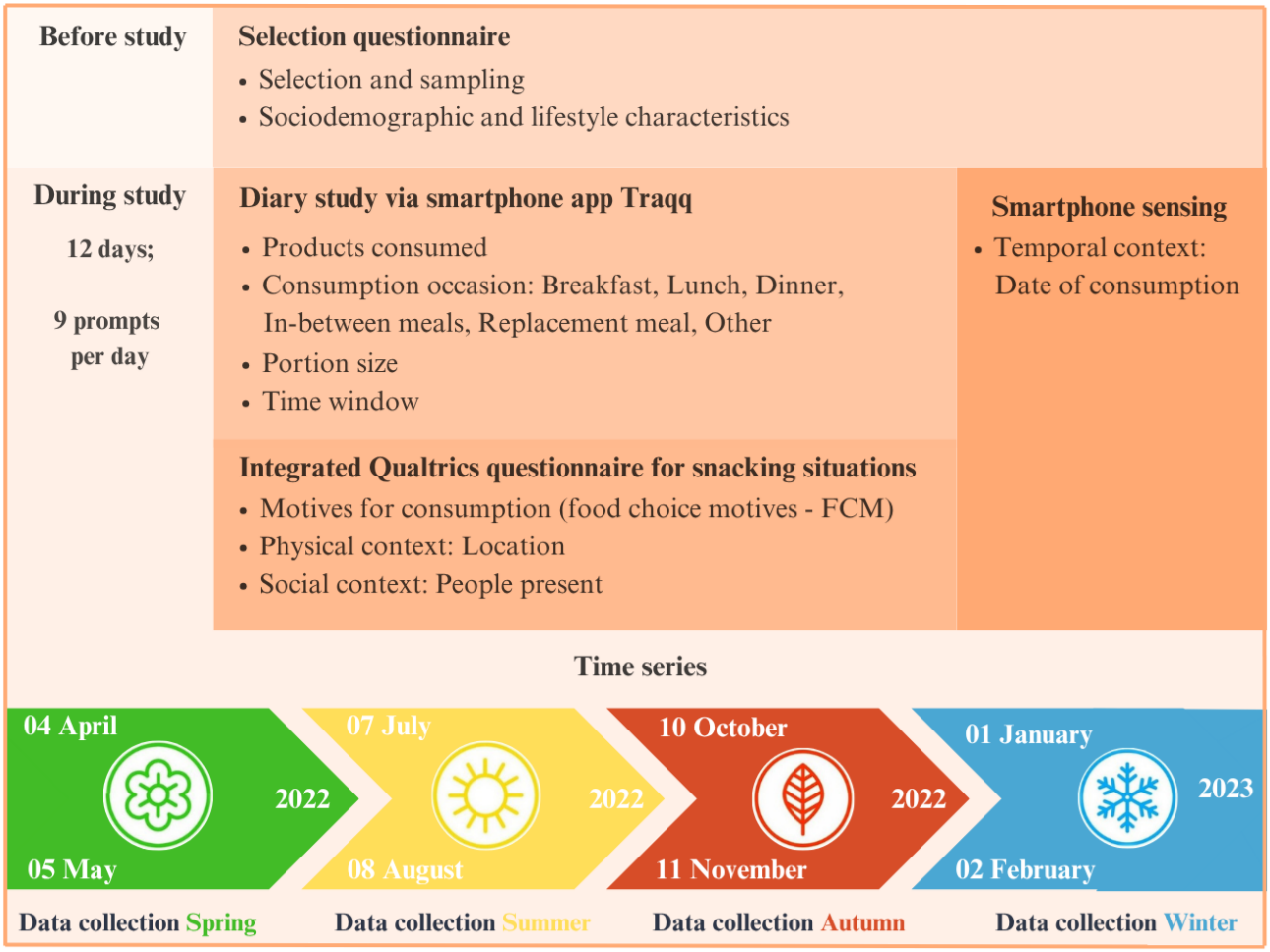
Given the different ways of defining snacking-related concepts in literature and the challenges this brings in unambiguously interpreting and comparing findings concerning snacking, the main snacking-related concepts as studied in FOODLOOP are operationalized as follows:

- Snack*:* a food or beverage product consumed outside the 3 traditional main meals (breakfast, lunch, and dinner)
- Snacking*:* the consumption of foods and/or beverages outside the 3 traditional main meals
- Snacking situation: the combination of a snack; FCMs for consuming a snack; and a physical, social, and temporal snacking context
- Snacking behavior: the totality of snacking situations of all participants

Study Design

The FOODLOOP study was carried out on a sample of 264 millennials (born between 1980 and 2000) residing in the Netherlands. Digital dietary diaries in a mobile app, following the principles of signal-contingent active EMA [33,37,38] and a time series structure [39], were used to collect data on participants' self-reported snacking situations and situational variations. A selection questionnaire was used to gather potential participants, as well as data on participants' sociodemographic and lifestyle characteristics (see Figure 1 for an overview of the study design).

Figure 1. FOODLOOP data collection overview and time series structure.



Potential participants filled in the selection questionnaire. Once admitted to the study, participants participated in the diary study for 12 days, equally distributed over 4 data collection periods within 1 year (start: April 2022, end: February 2023). Following a time series structure (Figure 1), the 4 data collection periods took place over the course of 2 months every season. The data collection periods were separated by 1 month in which no data collection took place. In each data collection period, participants registered their data on 3 days to capture behavioral cycles [40]. The data collection days were randomly assigned to participants by an automatic sampling scheme per data collection period, for which a set of rules was written. These rules entailed that data collection days (1) were nonconsecutive; (2) consisted of 2 weekdays and 1 weekend day; and (3) took place within the 2 months deemed most representative for a particular season (Figure 1), which were chosen based on meteorological variations in different seasons in the Netherlands [41]. Participants had to successfully complete 3 data collection days within each data collection period, thus 12 data collection days in total. Participants were not informed about the dates of their data collection days beforehand. The whole FOODLOOP study,

XSL•FO
RenderX

including the diary study and selection questionnaire, was conducted online.

Ethical Considerations

The Social Sciences Ethical Committee of Wageningen University (WUR) approved consent procedures and study protocols of the FOODLOOP study (2022 - 32-Vaal). The study was performed in accordance with ethical standards as laid down in Dutch privacy legislation and a data confidentiality declaration governed by Dutch law. To ensure participant privacy and confidentiality, all personal identifying information (such as names and contact information) was removed from the dataset before analysis, and participants were given a neutral identifier based on their study identifier. The research team only had access to the pseudoanonymized data. All participants provided informed consent to have their pseudoanonymized data published in journals. The compensation for participation in FOODLOOP consisted of a fixed monetary compensation, a lottery, and a donation possibility to heighten compliance [42-44]. Only participants who successfully completed all 12 data collection days received compensation. The fixed compensation consisted of €50 (US $59). Participants who successfully completed the study also automatically participated in a lottery where 10 prizes of €500 (US $585) were raffled. Participants received their compensation and prize after completion of the last data collection period (April 2023). Participants were given the option to keep the fixed compensation and prize themselves or have the research team donate the compensation and/or (part of) the prize on their behalf to the charitable organization Voedselbanken Nederland (Food Bank Netherlands), which provides basic food products to people who live below the poverty line. In total, €1800 (US $2106) was donated to Voedselbanken Nederland on behalf of the FOODLOOP cohort.

Participant Selection and Compliance

Potential FOODLOOP participants were recruited via online platforms (social media and forums) and the snowball sampling technique [45]. Potential participants (n=798) filled in the selection questionnaire to screen them for their eligibility for participation in the follow-up diary study whilst collecting relevant data on sociodemographic and lifestyle characteristics (age, sex, education, urbanity, living situation, activity level, etc). Results were screened based on selection criteria, being that participants were (1) born between 1980 and 2000; (2) living in the Netherlands and Dutch-speaking; (3) eating and drinking outside of what they defined as main meals (ie, snacking); (4) not following a diet for medical reasons or under supervision of a health professional; and (5) not having eating disorders (eg, anorexia or bulimia). Based on the screening, 534 participants were eligible for participation in FOODLOOP and were invited for the diary study. In total, 398 participants consented to participation and started the diary study.

Diary Study

Traqq

The FOODLOOP diary study was executed using the mobile app Traqq. Traqq (available for Android and iOS) was developed for dietary research purposes by Wageningen University [46] and allows for descriptive data collection, has no commercial focus, ensures privacy and safe storage of users' data, and enables customizations in line with this study's objectives. In Traqq, participants registered their daily food and beverage intake by choosing what they consumed from the Traqq food database, which includes products from the NEVO (Nederlands Voedingsstoffenbestand) 2016/2019 Dutch Food Composition Database [47,48] and Dutch food stores' assortments. Additional variables measured for consumed snacks were captured through questionnaires in Qualtrics XM (2022 - 2023). These questionnaires were programmed such that they appeared integrated within Traqq itself to participants. At the start of the study, participants received log-in credentials for Traqq, which remained valid throughout the entire study period. Once logged in, participants stayed continuously logged in, enabling them to receive prompts for all data collection days. Access to Traqq was deactivated upon study completion. All additions to and changes in features in the existing Traqq app to align it with FOODLOOP study objectives can be found in Table S1 in Multimedia Appendix 1.

Data Reporting Protocols

Signal-Contingent Active EMA

Given the nature of the study, the possibilities of the app, and the interference of the app with daily activities and social life when collecting data in real-time [49], it was decided to use signal-contingent active EMA for data collection. EMA allows for repeated data collection on participants' situational dietary behaviors in their typical real-world context via mobile devices [37,38]. In the case of FOODLOOP, data were repeatedly collected on participants' self-reported near real-time situational snacking behaviors in the physical, social, and temporal context in which they prevailed, via participants' own mobile phones to reduce recall and social desirability bias.

Data collection days lasted 24 hours (06:00 AM-06:00 AM) and consisted of 9 prompts delivered via Traqq following a fixed schedule (Table S2 in Multimedia Appendix 1). Each prompt concerned one of the time windows and appeared as a push notification on the participant's phone, which stayed salient for 2 hours until responded to. Participants indicated whether they had consumed anything during each time window. Nonresponses were registered as missing, and data collection days with too many missing prompts were considered invalid. Participants with invalid days were automatically rescheduled to repeat the session 2 days later within the designated data collection period until 3 valid days were achieved per data collection period.

Data Recording

During each of the 9 daily prompts, participants entered all food and beverage products consumed within the corresponding time window. They selected products from the Traqq database via the search bar and indicated the self-perceived portion size (Figure S1 in Multimedia Appendix 1). If a product was not listed, participants chose the most similar option available. Participants were instructed to record all consumed products except for plain water, fats and oils (eg, butter, margarine, and

XSL•FO
RenderX

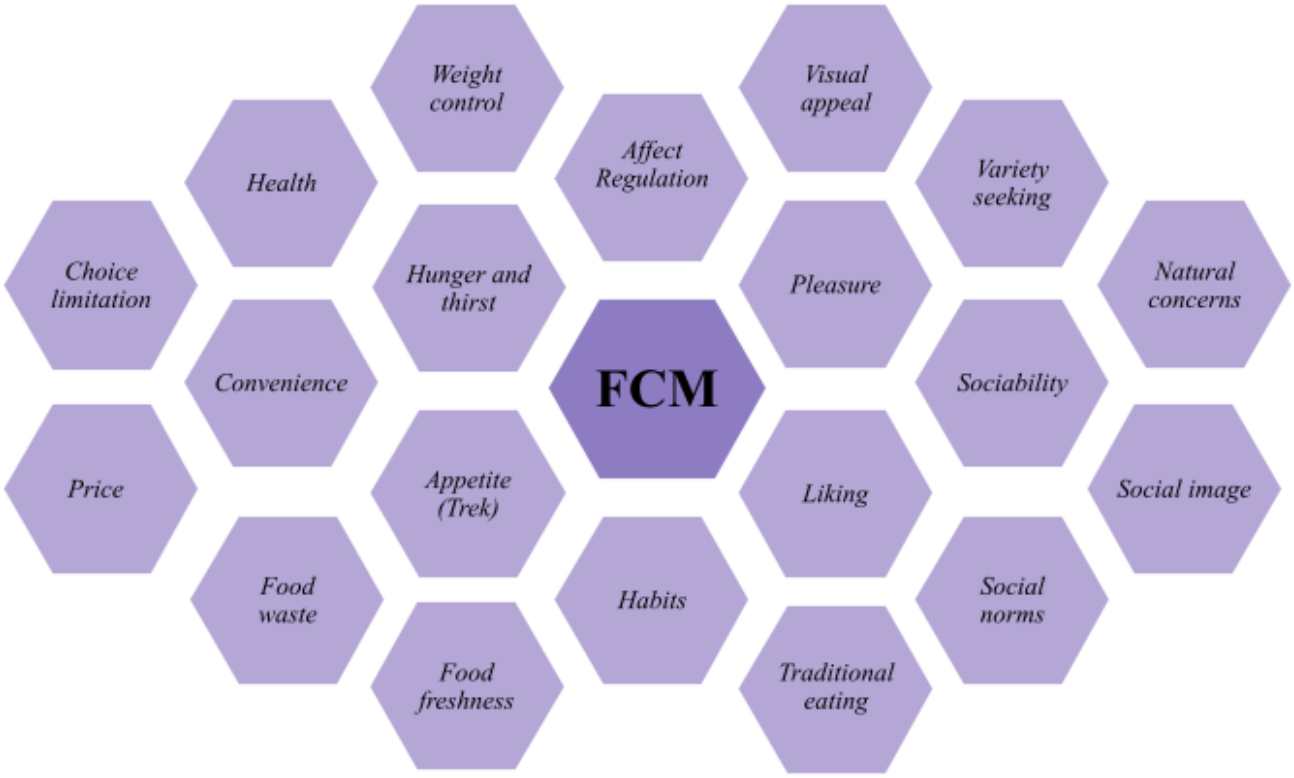
baking oil), and herbs and spices (Table S3 in Multimedia Appendix 1).

After entering the consumed product, participants had to label the product according to their perception of the associated consumption occasion as either one of the meal occasions (breakfast, lunch, and dinner) or one of the snacking occasions (replacement of meals, in-between-meals, or other), without giving instances of what foods and beverages constituted these occasions. This approach, adopted from Biel et al [36], allowed participants to self-define consumption occasions. Products of all consumption occasions, including meals, were recorded. Participants were not informed that FOODLOOP's focus was on snacking behavior until after study completion to prevent bias among participants toward snacking entries.

Additional Questions for Snacks

After entering a product and labeling the associated consumption occasion, participants received additional questions only when the product was marked as one of the snacking occasions. First, participants reported their FCM using the Dutch 1-scale-per-item Eating Motivation Survey (TEMS). This version contained 20 items (Figure 2), each assessing a unique FCM on a 3-point Likert scale ranging from 1 ("not applicable") to 3 ("applicable"). The survey was based on the 1-scale-per-item TEMS used in the study of Wahl et al [26] and TEMS as used in multiple studies [27,30,50].

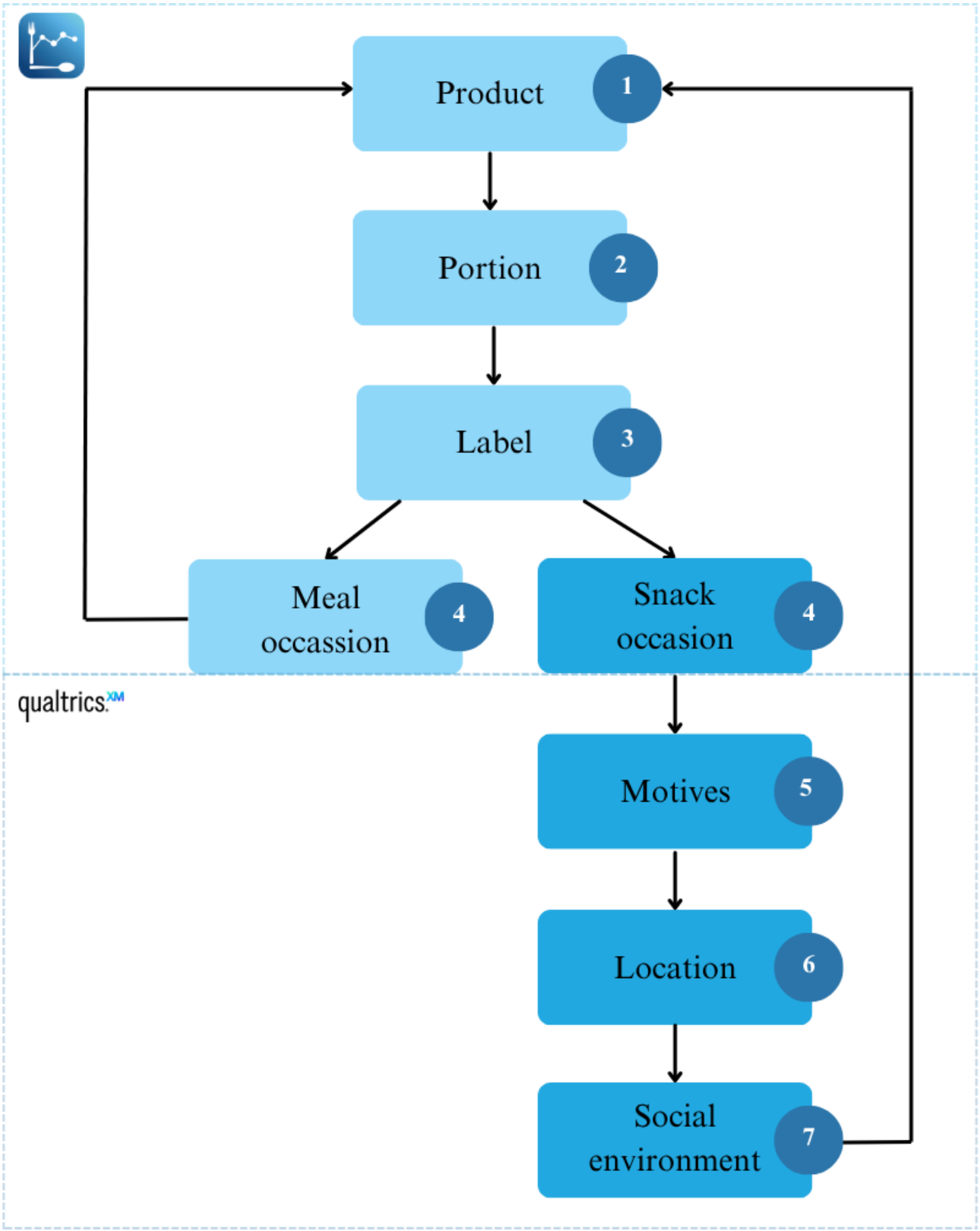
Figure 2. Food choice motive (FCM) of Dutch 1-scale-per-item Eating Motivation Survey (TEMS). FCM: food choice motive.



Thereafter, participants indicated the physical (location) and social (people present) context of each consumed snack from predefined categories. The temporal context (eg, time window, type of day, and season) was derived after the study from the app's recorded date and time data of each entry. When a product was labeled as a meal occasion or after completing a snacking entry, the app automatically returned to the Traqq home screen, where participants could log additional products or finalize the time window (Figure 3).

XSL•FO
RenderX

Figure 3. Protocol Traqq for entry of products.



All consumed products were entered separately, with associated questions answered for each product individually (eg, if coffee and a cookie were consumed together, first the coffee was entered and questions about the coffee were answered, then the cookie was entered and questions about the cookie were answered). When participants entered all the products consumed within a time window and filled in the associated questions, or when participants indicated that they had not consumed anything, they completed the time window by submitting their answers. The time window was then recorded as successful. Data collection days with the required number of valid time windows were deemed successful, and participants remained promptable for subsequent data collection days. The app was refined through several pilot iterations to balance user burden

XSL•FO
RenderX

and compliance. One data collection day took approximately 30 minutes to complete.

Data Analysis

Data Preprocessing

Given the objectives of this study, only data concerning products labeled as snacking occasions were included in the analyses. Data related to meals, incomplete collection days, and dropout participants were excluded. Therefore, there were no participants with missing data for any of the variables of interest (Table 1). After data collection, all snack products were classified into categories (Table 1; Table S4 in Multimedia Appendix 1), based on the Dutch Food Composition Database (versions 2016 and 2019; NEVO [47,48]), Dutch National Food Consumption Survey [51,52], and the Netherlands Nutrition Center (Voedingscentrum [53]). These snack product category variables were subsequently used in the analyses. In addition, the variables concerning physical context were merged into 2 categories, “At home“ and “Out of home,” and the social context variables into the categories “Alone” and “With others.” Table 1 shows the snacking determinants, corresponding variables, and response categories for analyses.

XSL•FO
RenderX

Table . Snacking determinants, corresponding variables, and response categories for analyses.

Snacking determinants and variables	Response categories
Product	
Snack categories beverages (3)	Nonalcoholic beverages; alcoholic beverages; dairy (substitutes)
Snack categories foods (7)	Breads and cereals; fruits; vegetables and legumes; dairy (substitutes); meat (substitutes); cookies, bars, and pastries; confectionery and ice cream; savory snacks
Context	
Physical context (2)	At home; out of home
Social context (2)	Alone; with others
Temporal context	
Daypart (3)	Morning (06:00 AM-12:00 PM); afternoon (12:00 PM-06:00 PM); evening and night (06:00 PM-06:00 AM)
Type of day (2)	Weekday (Monday-Friday); Weekend day (Saturday and Sunday)
Season (4)	Spring, summer, autumn, and winter
Consumer	
Motives (FCM[a]; 20)	Liking; appetite; hunger and thirst; convenience; pleasure; health; visual appeal; habit; food freshness; sociability; variety seeking; weight concerns; choice limitations; price; food waste; traditional eating; social image; social norms; and natural concerns
Sociodemographic and lifestyles	
Age (2)	Younger millennials (1990 - 2000); older millennials (1980 - 1990)
Sex (2)	Men; women
Education (2)	Lower educated; higher educated
Net income per month (4)	Below modal (<€2000; <US $2340); modal (€2000–€3000; US $2340-US $3510); above modal (>€3000; >US $3510) not disclosed
Daily occupation (4)	None; study; paid job; study and paid job
Urbanity (5)	Rural; low urbanity; moderate urbanity; high urbanity; very high urbanity [54]
Living situation (2)	Living alone; living with others
Parental status (2)	No child or children; child or children
Grocery shopping autonomy (3)	No or little autonomy; shared autonomy; most or all autonomy
Diet (2)	No diet; diet
BMI group (4)	Underweight (<18.5); healthy weight (18.5 - 25); overweight (25-30); obese (>30)
Activity level (3)	Nonactive; semiactive; norm-active [55]
Fit level (3)	Nonfit; semifit; norm-fit [55]
Health interest (1-7)	Very low health interest – … - very high health interests

[a]FCM: food choice motive.

XSL•FO
RenderX

All analyses were performed with standardized snacking frequencies for each response category of all considered variables (Table 1) to correct for unequal participant weights, as some participants recorded more snacks than others. Each snack originally represented one case in the dataset. To ensure that all participants contributed equally to the analyses, absolute snacking frequencies for each response category were divided by the number of data collection days (12), resulting in the daily average frequency (DAF) of snacks consumed. The DAFs provided a uniform measure of comparison, with each row in the dataset representing one participant rather than one snack. To account for the differing number of data collection days across contexts, the DAFs were adjusted for the number of days per week (for "Type of day") and per season (for "Season"). DAFs were calculated for all response categories of all variables (Table 1) separately for (1) beverages as snacks, (2) foods as snacks, and (3) all snacks combined (for the detailed calculation of DAFs and all DAFs for each of the response categories, consult the Data Preprocessing and Descriptive Data sections and Tables S5-S8 in Multimedia Appendix 1).

Data Analyses

Data analyses were performed with Jamovi (Version 2.5; The jamovi Project), using R statistical packages (Version 4.1; R Core Team). Descriptive statistics and statistical tests were used to examine relationships between variables. In all analyses, a consumed snack represented a case (n) and the total number of participants was treated as the population (N). Descriptive statistics for snack category variables (Table S5 in Multimedia Appendix 1) demonstrated skewness toward nonalcoholic beverages (n=4204), indicating an imbalance between snack food and beverage categories. Given this imbalance and sufficient sample sizes for both beverages as snacks (n=5164) and foods as snacks (n=9148), analyses were conducted separately for these 2 groups.

Statistical tests were performed using DAFs of each response category (Table 1) to examine differences in snacking frequencies across determinants. Each variable was analyzed individually, aligning with the study's aim to explore snacking determinants one-dimensionally, as these determinants were not mutually exclusive (eg, the same snack could be categorized as "At home," "Alone," and "In the morning"). For each variable, differences in DAFs were tested both between and within response categories for (1) beverages as snacks, (2) foods as snacks, and (3) all snacks combined (eg, comparing DAFs of "Alone" vs "With others" and subsequently beverages vs foods within each category).

Nonparametric tests were used given nonnormality (Shapiro-Wilk; $P<.05$). No sensitivity analyses or multivariable confounder-adjusted models were used, given the descriptive nature of the study. Differences in DAFs for product-, context-, and motivation-related variables were analyzed using Friedman ANOVA with Durbin-Conover pairwise comparisons. For motivation variables, McDonald ω was calculated to assess reliability [56,57], showing high internal consistency across all 20 FCM scales (beverages: ω=0.81; foods: ω=0.82; all snacks: ω=0.84). Subgroup analyses were conducted to explore differences in DAFs across sociodemographic and lifestyle variables and were analyzed using 2-tailed independent samples *t* tests, Mann-Whitney *U*, and Kruskal-Wallis tests. Depending on Levene test results, either Student *t* ($P>.05$) or Mann-Whitney *U* ($P\leq.05$) was used for 2-category variables, while Kruskal-Wallis tests (Shapiro-Wilk, $P\leq.05$) with Dwass-Steel-Critchlow-Fligner pairwise comparisons were applied for variables with more than 2 categories. Test values (except *P* values) were reported with 2 decimals, due to which minor rounding deviations may occur. For all statistical tests, a *P* value of ≤.05 was considered significant.

Results

Sample Compliance and Characteristics

Of the 398 participants who started the study, 264 (66.3%) fulfilled the entire year of data collection. These 264 Dutch millennials constitute the cohort of this study, the FOODLOOP cohort. More details about compliance rates throughout the longitudinal study period can be found in Table S9 and Figure S2 in Multimedia Appendix 1. Most of the FOODLOOP cohort consisted of women (232/264, 87.8%), who were highly educated (224/264, 84.8%), had a paid job (216/264, 81.8%), and a relatively high income (>€3000; US $3510) net income per month; 136/264, 51.5%). The ages of the participants ranged from 22 to 42 years (mean 30.14, SD 5.63) at the start of the study, with 72.3% (191/264) of participants belonging to the younger millennial age group (born between 1990 and 2000). As for the living situation of the participants, most were living in more urbanized areas of the Netherlands (198/264, 75%), with others (216/264, 81.2%), especially with a partner (108/264, 40.9%), and did not have children (212/264, 80.3%). Most participants could make grocery shopping decisions autonomously (158/264, 59.8%) or share this decision with other members in the household (100/264, 37.8%).

Seventy-two percent of participants (n=190) had a healthy weight (BMI: mean 23.28, SD 3.92) and 59.8% (158/264) did not follow a particular diet. Also, 21.2% (56/264) of participants skipped at least 1 meal a day or replaced meals with smaller consumptions. Most participants engaged in moderate and/or intense exercise at least once a week (moderate: 260/264, 98.5%; intense: 209/264, 79.1%), of which 39.3% (104/264) engaged in moderate exercise more than 5 times a week, and 26.5% (70/264) engaged in intense exercise more than 3 times a week. Weighted health interest of participants ranged from 2.75 to 6.88, with a mean of 5.02 (SD 0.78), indicating higher interest in health amongst participants in the FOODLOOP cohort.

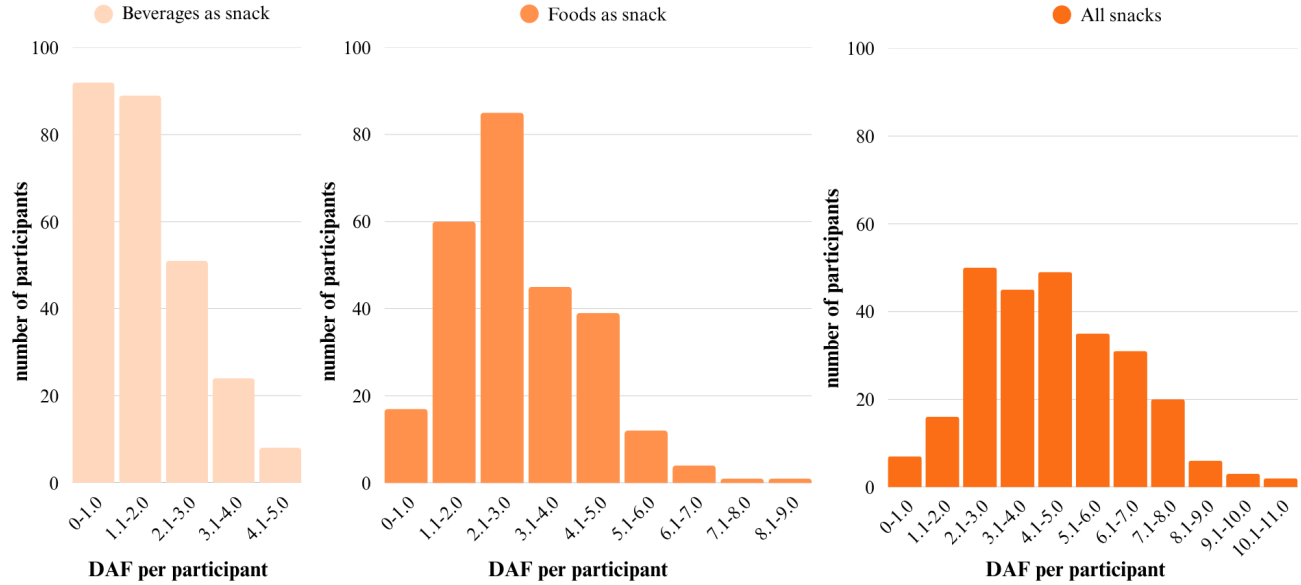
Product-Specific Determinants

Over the longitudinal study period, 14,312 snacks, of which 5164 beverages as snacks (36.1%) and 9148 snack foods (63.9%), were consumed by the 264 participants in the FOODLOOP cohort over the course of 2987 unique days, across 4 different seasons. The DAFs of snacking range from 0 to 5 for beverages, 0 to 9 for foods, and 0 to 11 for all snacks. Most participants consumed 1 - 2 beverages, 1 - 4 foods, and 1 - 5 snacks in total per day (Figure 4). On average, 1.63 beverages, 2.89 foods, and 4.52 snacks in total were consumed per participant per day. In all snacking situations, DAFs of foods

XSL•FO
RenderX

were significantly higher than those of beverages (Table S10, Multimedia Appendix 1).

Figure 4. Distribution daily average frequencies (DAFs) of all participants for beverages as snacks, foods as snacks, and all snacks.

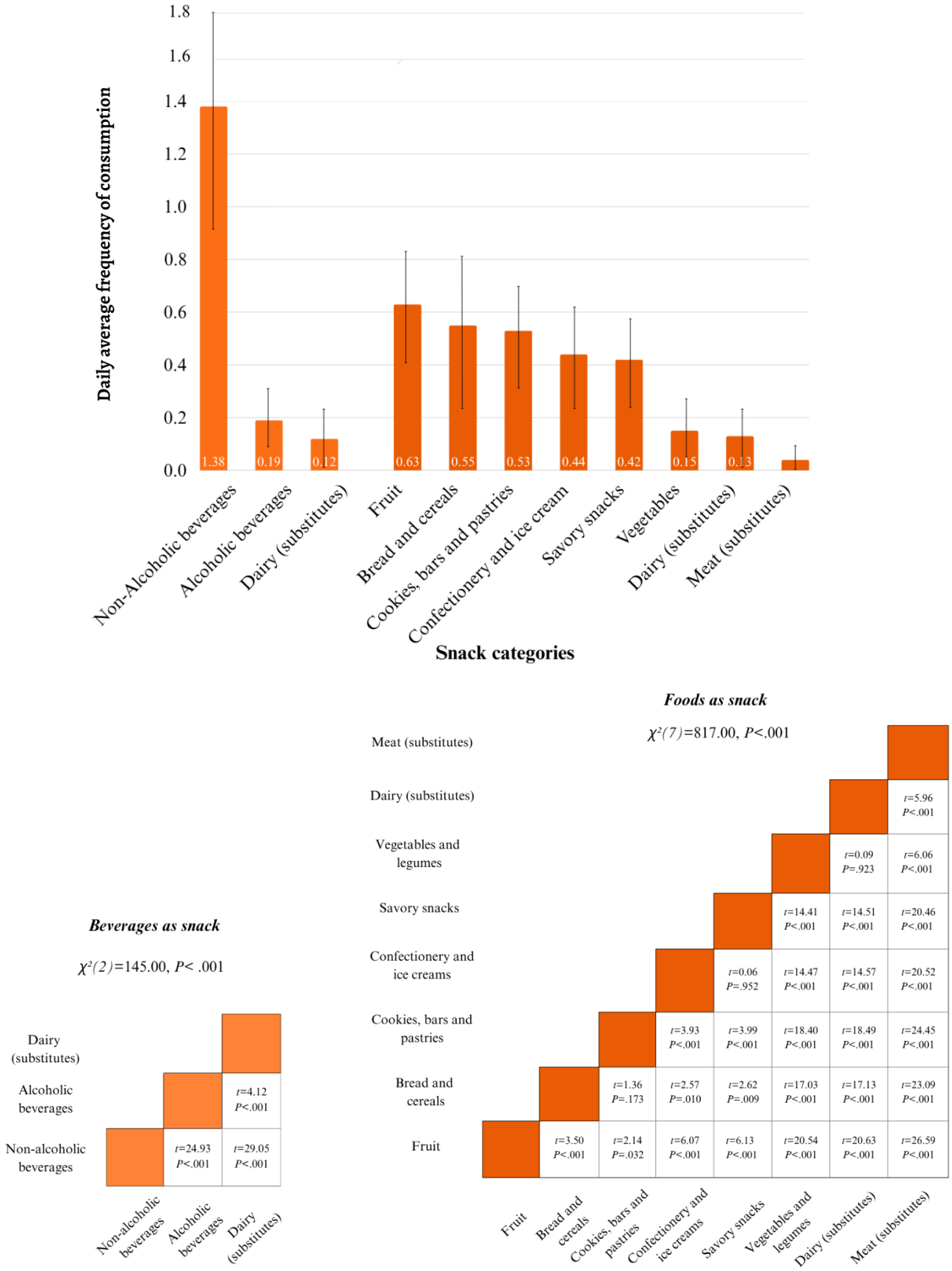


Snacking Categories

Nonalcoholic beverages (consisting of mainly coffee and tea) were mostly consumed as snacks. For beverage categories, this was followed by alcoholic beverages. Dairy beverages and substitutes were least consumed. All differences between DAFs of the beverage categories were statistically significant (Figure 5).

XSL•FO
RenderX

Figure 5. Daily average frequencies (DAFs) for snack categories and pairwise comparisons of differences between snack categories. Higher *t* values imply larger differences between DAFs of 2 snack categories.



Of all foods, fruit was mostly consumed as snacks, followed by bread and cereals; cookies, bars, and pastries; confectionery and ice cream; and savory snacks. Meat, poultry, fish, eggs, and substitutes were the least consumed foods as snacks. Most differences between DAFs for foods as snack categories were statistically significant (Figure 5). Most consumed snacks were

XSL•FO
RenderX

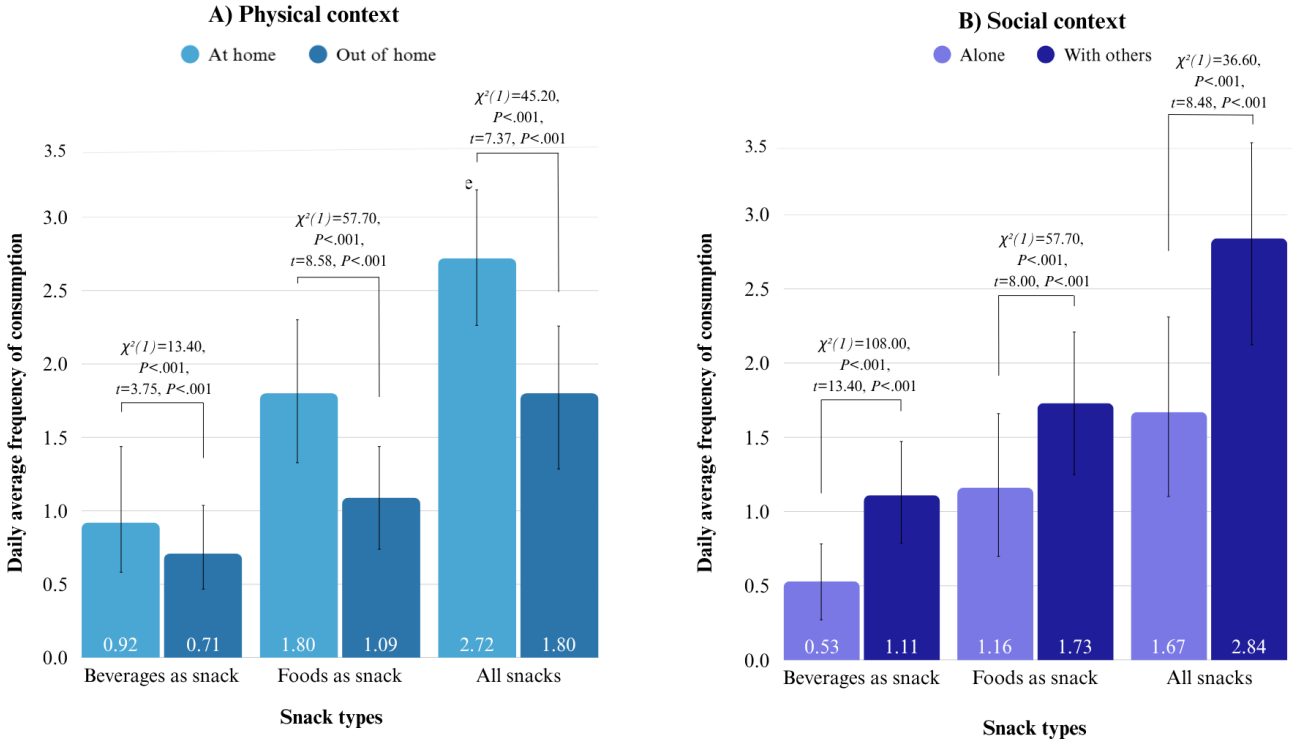
perceived as an in-between meal consumption occasion. For replacement meal occasions, the DAF of foods was higher than that of beverages, whereas the DAF of foods was lower than that of beverages for snacking occasions other than in-between or replacement meals. All differences were statistically significant (Tables S10 and S11 in Multimedia Appendix 1).

Context-Specific Snacking Determinants

Physical Context of Snacking

Most snacks were consumed at home rather than out of home. This was the case for both foods and beverages (Figure 6A). In both contexts, the DAFs of foods were higher than those of beverages, but at home, the difference was larger. This indicates that foods were relatively more frequently consumed than beverages at home compared to out of the home. All differences were statistically significant (Figure 6A; Table S10 in Multimedia Appendix 1).

Figure 6. Bar charts of daily average frequencies (DAFs) of consumption and pairwise comparisons showing differences between DAFs in the (A) physical context at home versus out of home and (B) social context alone versus with others. Higher chi-square values imply larger differences between DAFs of 2 context categories.



Social Context of Snacking

Most foods and beverages as snacks were consumed with others, rather than alone (Figure 6B). In both contexts, the DAFs of foods were higher than those of beverages. However, this difference was larger when alone. This indicates that foods were relatively more frequently consumed than beverages alone, as compared to others. All differences were statistically significant (Figure 6B; Table S10 in Multimedia Appendix 1).

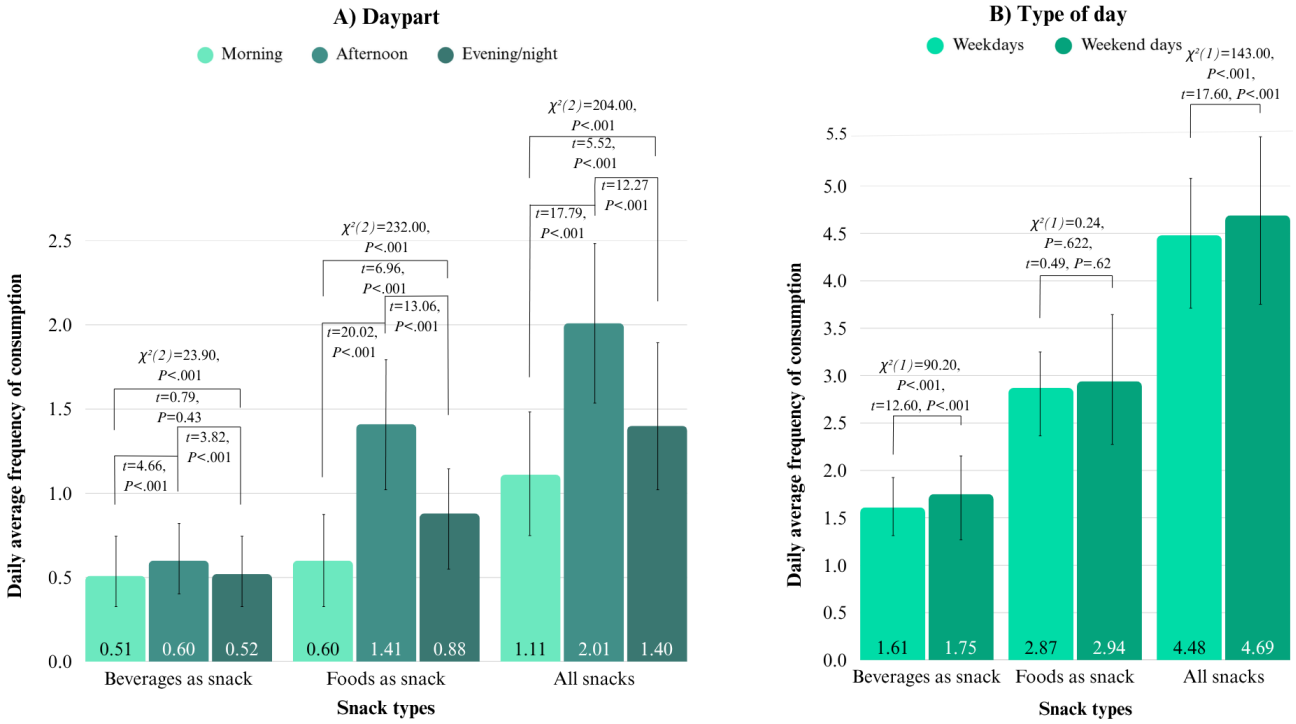
Temporal Context of Snacking

Snacking During the Day

Data demonstrate that snacks were consumed throughout the day, but most snacks were consumed in the afternoon, followed by the evening and night, and least snacks were consumed in the morning. This was the case for both foods and beverages but especially pronounced for foods. For foods, the DAF was lowest in the morning, whereas the DAFs of beverages did not differ significantly between the morning and the evening and night (Figure 7A). In all dayparts, DAFs of foods were significantly higher than those of beverages, but this difference was largest in the afternoon and smallest in the morning. This indicates that foods were relatively more frequently consumed than beverages in the afternoon as compared to other dayparts (Figure 7A; Table S10 in Multimedia Appendix 1).

XSL•FO
RenderX

Figure 7. Bar charts of daily average frequencies (DAFs) of consumption and pairwise comparisons showing differences between DAFs in the temporal contexts. (A) Daypart, morning versus afternoon versus evening and night, and (B) day types, weekdays versus weekend days. Higher chi-square values imply larger differences between DAFs of 2 context categories.



Snacking During the Week

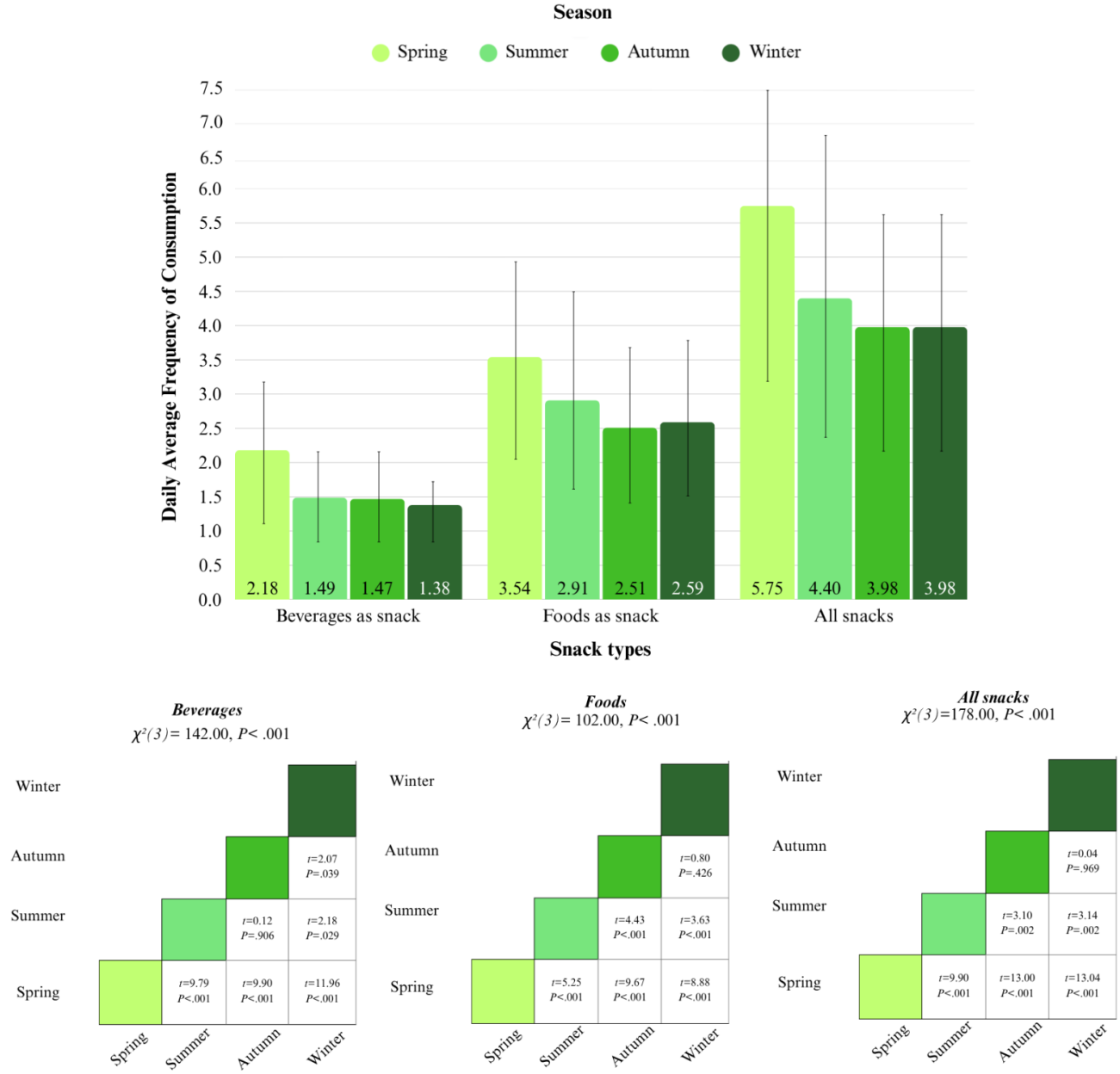
Data show that snack consumption on weekdays and weekend days was comparable. There were only minor differences in snack consumption between weekdays and weekend days. The weighted DAFs (accounting for the number of weekdays and weekend days) show that relatively more beverages were consumed on weekend days than on weekdays. Even though the difference was minimal, it was statistically significant. The weighted DAFs of foods did not differ significantly between weekdays and weekend days (Figure 7B). In both day types, the DAFs of foods were significantly higher than those of beverages, but this difference was larger on weekend days. This indicates that foods were relatively more frequently consumed than beverages on weekend days as compared to weekdays (Figure 7B; Table S10 in Multimedia Appendix 1).

Snacking During the Year

Most snacks were consumed in spring, followed by summer. DAFs of snacking were lowest in autumn and winter. Most differences in DAFs between seasons were significant. Only the DAFs of beverages did not differ significantly between summer and autumn, and the DAFs of foods did not differ significantly between autumn and winter (Figure 8). Within all seasons, the DAFs of foods were significantly higher than those of beverages. This difference was most pronounced in summer, indicating that foods were relatively more frequently consumed than beverages as compared to other seasons (Figure 8; Table S10, Multimedia Appendix 1).

XSL•FO
RenderX

Figure 8. Bar charts for daily average frequencies (DAFs) of temporal context and pairwise comparisons. Seasonality of differences in DAFs between spring versus summer versus autumn versus winter. Higher chi-square values and *t* values imply larger differences between DAFs of seasons.



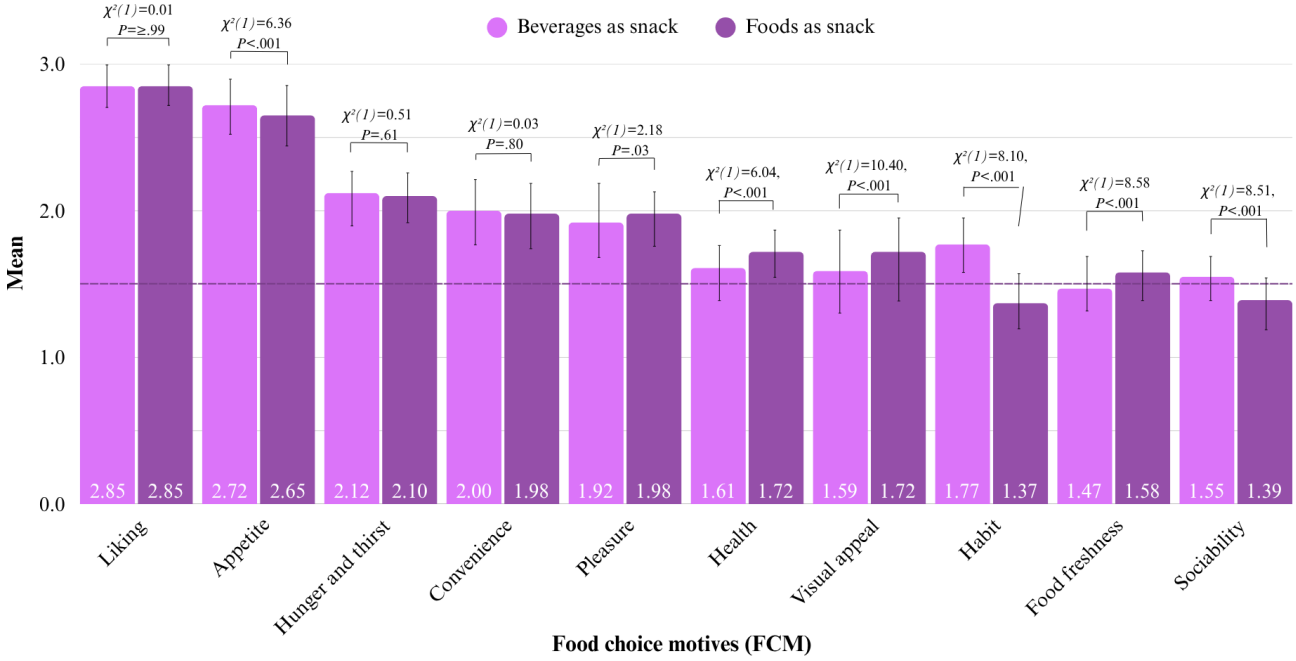
Consumer-Specific Snacking Determinants

Snacking FCMs

Data demonstrate that 10 out of the 20 FCMs played a prominent role in snacking (mean ≥1.5; Table S7 in Multimedia Appendix 1). Snacking was mainly motivated by liking and appetite (mean >2.5), followed by hunger and thirst, and convenience. Pleasure, visual appeal, health, habit, and food freshness were also found to belong to the core snacking drivers. The other 10 FCMs, including variety seeking, weight concerns, and affect regulation, did not often motivate participants to snack (mean <1.5). The role of most of the prominent FCM (mean ≥1.5) differed significantly between beverages and foods (Figure 9). These differences were largest for the FCM visual appeal, habit, and sociability. For beverages, the FCM habit, sociability, and appetite were more prominent, whereas visual appeal, health, pleasure, and food freshness were more prominent for foods. The other FCMs, liking, hunger and thirst, and convenience, were shown to be more universal snacking drivers, as they played a similar role in the consumption of both beverages and foods (Figure 9).

XSL•FO
RenderX

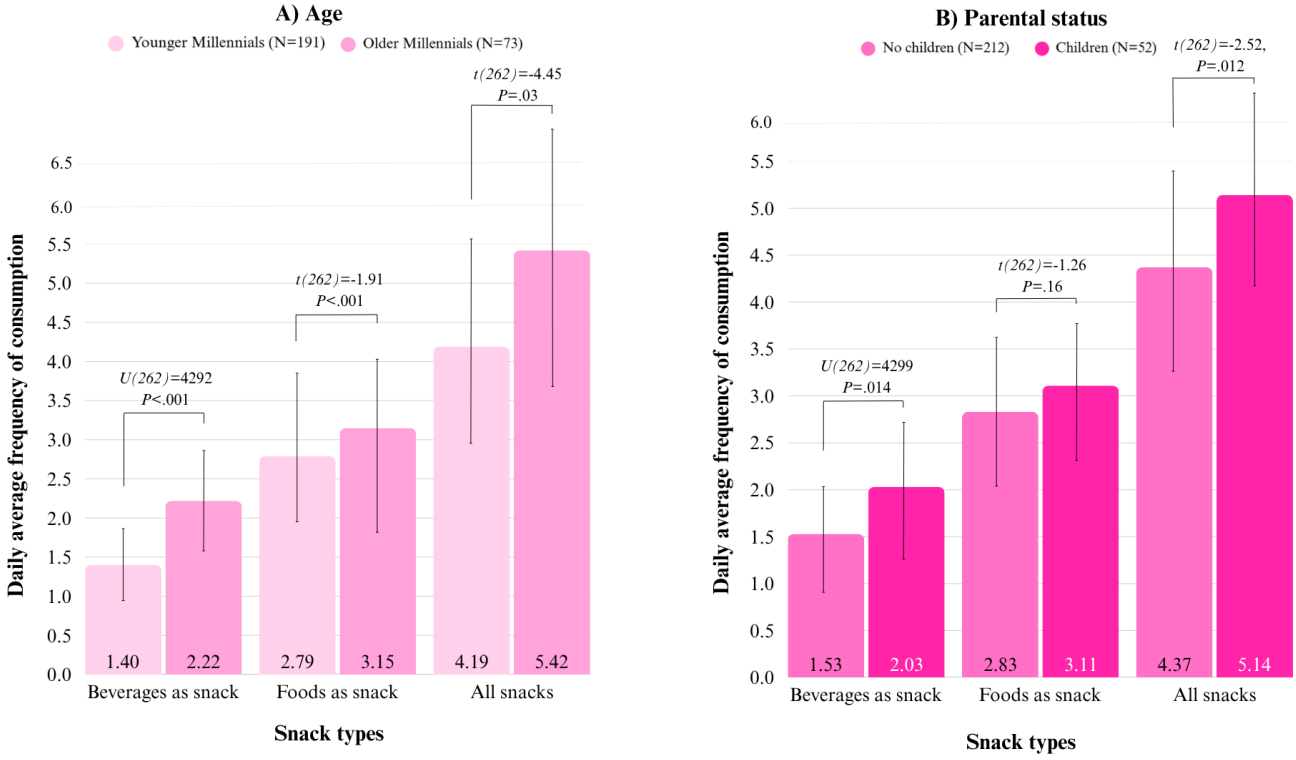
Figure 9. Average scores on food choice motives (FCM): pairwise comparisons results of FCM with means ≥1.5 (measured on 3-point Likert scales).



Sociodemographic, Lifestyle Factors, and Snacking

Regarding sociodemographic and lifestyle groups, the data demonstrate that there were statistically significant differences in DAFs for beverages and foods, given participants' age and parental status. Participants belonging to the older millennials (born between 1980 and 1990) consumed more beverages and foods than participants belonging to the younger millennials (1990-2000) (Figure 10A). Participants with children consumed significantly more beverages than participants without children (Figure 10B). There were no statistically significant differences in DAFs for beverages and foods, given other sociodemographic and lifestyle factors of the participants, such as sex, education, urbanity of living area, and the activity and fit levels.

Figure 10. Bar chart showing daily average frequencies (DAFs) of consumption and pairwise comparisons showing differences between DAFs according to the sociodemographic characteristics of age and parental status. (A) Younger versus older millennials and (B) participants with children versus without children. Higher *U* values and *t* values imply larger differences between DAFs of demographic groups.



XSL•FO
RenderX

Discussion

Overview

This study used Dutch millennials as a case to explore consumers' longitudinal situational snacking behaviors by investigating the role of product-, context-, and consumer-specific determinants in snacking situations, as well as how foods and beverages for snacking differ within and across these determinants, using a mobile app following the principles of signal-contingent active EMA.

Principal Findings

Product-Specific Determinants

Examining the aggregated EMA data across all participants and snacking situations, this study found a daily average snacking frequency of 4.5. This is higher than reported in other studies, which range between 1.5 and 3.5 [58-62]. The higher frequency in our study might be explained by the composition of our cohort (1) millennials, who tend to snack more than other age groups, especially given the ongoing snackification trend [16], and (2) Dutch consumers, who likely adhere to the Dutch diet. As found by Geurts et al [14], Dutch adults (aged 19 - 69 years) report more than 7 eating occasions per day, including 4 between-meal (ie, snacking) occasions, which aligns with our findings. Together, these findings suggest that snacking forms a substantial part of the daily diet of the Dutch, classifying them as heavy snackers (>3 snacks per day [21]).

Across all contexts, participants consumed relatively more foods than beverages as snacks. This contrasts with van Rossum et al [15], who found that the Dutch diet in general consists of more beverages than foods. However, our findings support existing snacking literature, showing that beverages are less prominent than foods in snack frequency, as beverages are less likely to be perceived as snacks by consumers [63,64].

Notwithstanding this smaller role of beverages in snacking, our findings do show that nonalcoholic beverages (mainly coffee and tea) were frequently consumed as snacks. This aligns with the Dutch diet in general being characterized by high consumption of nonalcoholic beverages [14,15] and the trend that an increasing proportion of beverages is consumed as snacks [21].

Concerning food categories, fruit, bread and cereals, cookies, bars and pastries, confectionery and ice creams, and savory snacks were among the top snacks for Dutch millennials. Vegetables, meat, and dairy (substitutes) were the least frequently consumed snacks. Most of these findings are consistent with previous studies [15,62,65,66]. Different from other studies is that our results show relatively high consumption of bread and cereal products as snacks. This may be explained by findings from other Dutch studies [14,15], showing that bread and cereal products are staple foods and important energy sources in the Dutch diet, especially for meal occasions. Our results suggest that these foods also play an important role in Dutch millennials' snacking occasions.

Context-Specific Determinants

Confirming our expectations, findings show that context plays a central role in shaping snacking behaviors. Findings regarding the physical context show that most snacks were consumed at home, aligning with previous studies [29,67-69] and national dietary data showing that over 85% of foods and beverages in the Netherlands are consumed at home [14,15]. This could be explained by the convenience of the accessibility to snacks at home and the cost-effectiveness of home snacking [70]. The timing of data collection, shortly after the COVID-19 pandemic (2022 - 2023), may also have influenced these results. Postpandemic patterns show that consumers not only snack more frequently in general [71-73], but especially at home, a habit established during lockdowns [74-76]. Furthermore, the COVID-19 pandemic remnant of remote and hybrid working [77] may have increased participants' time spent at home, thereby facilitating more home snacking.

Concerning the social context, our study shows that most snacks were consumed with others, rather than alone. Since information regarding social snacking consumption patterns of the Dutch is scarce [14], our findings offer new insights into the role of the social environment in Dutch millennials' snacking. Our findings, however, contrast with social snacking consumption patterns in other Western countries, where snacking is considered a personal eating event and mostly done alone [3,50,78,79]. A possible explanation for the higher consumption of snacks with others in our study might be the interaction between the physical and social environment; most snacks were consumed at home, and since most of our cohort lived with partners, children, and/or family members, snacking at home may have naturally occurred in the presence of other inhabitants.

Concerning the temporal context, our study shows that snacking occurred throughout the day. This suggests that snacking is not necessarily confined to intervals between the regular Dutch mealtimes, but that any consumption at any time not perceived as (part of) a meal can be a snack. This is in line with findings of other studies [3,80,81] and the snackification trend [82]. Throughout the day, snacking was most prevalent in the afternoon, followed by the evening and morning, consistent with prior studies [60,69,83]. To the best of our knowledge, our study is one of the first to explore seasonal variations in consumers' snacking behaviors. We show that snack consumption was the highest in spring and the lowest in autumn and winter. These seasonal variations in snacking frequency might be explained by the seasonal climate of the Netherlands, where variations in meteorological factors such as temperature, precipitation, sunshine hours, and humidity are more noticeable throughout the seasons. In such climates, seasonal variations are also more pronounced in food consumption [84]. Even though literature on the effect of seasonality on food behaviors is scarce, our findings support those of Girju and Ratchford [85], who also observed less snacking in winter and more snacking in spring and summer.

Consumer-Specific Determinants

Concerning sociodemographic and lifestyle factors, we found some differences in snacking frequencies given age and parental status. Interestingly, older millennials snacked more frequently

XSL•FO
RenderX

in general than younger millennials, and millennials with children consumed more beverages as snacks than those without. These 2 findings might be connected, as most of the participants with children were older millennials. Our findings concerning age align with those of Geurts et al [14], who reported that overall food consumption among the Dutch increases with age until middle age. The lack of expected differences in snacking frequencies for other sociodemographic and lifestyle factors, including sex, education, and activity levels, may reflect a possible type II error around the estimates, given minor subgroups of participants for factor levels [86].

Concerning FCMs, our study extends prior work by capturing within-person variations in snacking motives across different snacking contexts, thereby providing a deeper understanding of why consumers choose specific snacks in specific situations. We found that liking was the primary driver for snacking in all consumption situations, congruent with other studies [50,62]. Other core motivators for snacking in our study, including hunger and thirst, convenience, weight control, health, pleasure, and visual appeal, are also aligned with findings from other studies. Different from other studies is that we found other core FCM for snacking to entail sociability, food freshness, and habit [50,62]. The role of these FCMs in our study might be explained by the interaction of multiple snacking determinants, as the motives driving consumers' snacking behaviors depend largely on the consumption context [29,50,87].

The role of FCM Sociability in our study could be explained by the finding that most snacks were consumed in the presence of others. The statement measuring this FCM included the Dutch word "gezellig." The meaning of "gezellig" is not straightforward [88] but can be explained as the experience of togetherness in doing everyday activities together, often at specific times of the day and in one's own home, implying collective enjoyment [88,89]. The "gezelligheid" of consuming snacks with others has played a noteworthy role in our study, possibly due to the social enjoyment and collectiveness of snacking. Even though "gezelligheid" is culturally embedded within the Dutch snacking context, the TEMS has demonstrated high cross-country and cultural group validity [26,30]. Therefore, our findings concerning FCM sociability might be translatable to other cultural contexts given sociability-related concepts similar to "gezelligheid," but future studies are warranted.

The role of food freshness might be explained by our finding that mostly fruits, bread, and cereals were consumed as snacks. Food freshness includes the recency of production or harvest and the overall product quality [90-93]. It is a key determinant of consumers' food choices, especially for fruits and baked products, which are highly susceptible to deterioration over time and perceived as having higher quality closer to the moment of harvest, production, or purchase [91,92]. This interaction of snack product categories and FCM might have played a role in our study.

The prominence of the motive habit in snacking has been established in other studies, but mainly for unhealthy snacking behaviors [94] and fruit intake in general [95]. The high consumption of fruit as a snack, alongside the cumulative consumption of cookies, bars, and pastries; savory snacks; and confectionery and ice creams in our study, might explain the prominent role of the FCM habit.

The prominence of the FCMs health and hunger and thirst in our study demonstrates that our cohort often selected snacks based on perceived healthfulness and satiating properties. Health- and hunger-driven snacking tends to be associated with the consumption of more healthful products [9,96]. This is observed in our study, as 64% of all consumed snacks consisted of coffee and tea, fruits, bread and cereal products, dairy products, and vegetables. This challenges the traditional view of snacking as a merely unhealthy dietary behavior characterized by underconsumption of nutrient- and fiber-rich products (such as fruits and grains) and excessive consumption of highly palatable traditional snack foods high in salt, saturated fat, trans fat, and sugar [97,98]. Our findings thus contrast with the traditional representation of snacking as merely unhealthy and harmful and support the growing scientific evidence that argues that snacking is not necessarily detrimental to but could also potentially benefit health [7,12,13,99]. We endorse the notion of Barnes et al [60] that snacking, like any other dietary behavior, can be practiced in a healthful or nonhealthful manner, depending on the product chosen, variety within the diet, and the underlying motives for consumption.

Strengths

One of the main strengths of the FOODLOOP study is the use of a mobile app following a signal-contingent active EMA method, which enabled the collection of data on situational snacking behaviors in participants' natural environments, using the routinized handling of mobile phones. Large amounts of detailed data concerning participants' near real-time snacking behaviors, together with the comprehensive combination of product-, context-, and consumer-specific determinants present at specific moments in unique snacking situations, were documented, which allows for a better understanding of real-world snacking behaviors. As beverages consumed as snacks were distinctly captured, the data also enabled the analyses of differences in snacking behaviors between foods and beverages, accounting for variances in situational determinants. Given the use of a mobile app following an EMA method, practical applications of the findings from FOODLOOP are expected to be more relevant to real-world snacking situations [100] while minimizing missing data, memory loss, recall bias, and social desirability bias and maximizing ecological validity [26,31,35,36,38,46].

Another strength is the longitudinal nature of our study. FOODLOOP is one of the first to use a time-series structure for longitudinal data collection on consumers' snacking behaviors, enabling the exploration of temporal relationships among variables, especially the little-studied seasonal variations in snacking. As data were collected from the same group of 264 participants in all data collection periods, both inter- and intraindividual variability in snacking behaviors could be analyzed. The completion rate of 66.3% (264 of 398 participants) is comparable to those of previous EMA studies [101,102], and the FOODLOOP study managed to keep compliance of the participants throughout several data collection

XSL•FO
RenderX

periods throughout the year, notwithstanding months in which no data collection took place. Participants' compliance was supported by the FOODLOOP study design in 3 ways:

- Limiting participant burden through the data collection structure
  - Data were collected on 3 days each season, as 3-day surveys are shown to capture data with similar sufficiency to surveys with more data collection days [40].
  - Prompts were sent as push notifications on participants' phones to heighten engagement and minimize data loss [103].
  - No prompts were sent between 11:00 PM and 08:00 AM to ensure that participants were not bothered at inappropriate times.
  - The number of questions per prompt was minimized while capturing all necessary data (eg, through using the extensively piloted 1-scale-per-item TEMS).
  - Notifications were sent once per prompt to prevent notification excessiveness, a known barrier in mobile app-based dietary intake [33,103,104].
- Communication with participants
  - Participants received concise and detailed user manuals (documents and videos) to heighten compliance [103-106] via email and the study's web page before data collection, which remained continuously accessible during the study.
  - Participants received personalized reminder emails after each data collection day about the progress of participation.
  - The research team remained available 24/7 to enhance participants' motivation [83,105,107-110] via email and Instagram (Meta) for participants' remarks and questions during the full length of the study, including the no-data-collection months.
  - Emails and direct messages from participants were responded to within a day.
- The incentive scheme
  - The combination of multiple types of incentives (fixed monetary incentive, lottery, and donation possibilities) is the most effective strategy for increasing compliance rates [42-44] in studies using dietary mobile apps [105,111].

Limitations

Notwithstanding the insights we provide into Dutch millennial consumers' situational snacking behaviors through longitudinal EMA time-series assessment, several limitations should be acknowledged.

First, the sample was skewed toward women and higher-educated millennials, and most participants were living in urbanized areas, with a healthy weight and more active lifestyles. This likely resulted from recruiting participants through health- and nutrition-related channels (eg, Netherlands Nutrition Center and nutrition blogs) using a snowball sampling technique, rather than targeted sampling. This selection bias could have caused the inclusion of participants with a stronger interest in health and nutrition than the average Dutch millennial, which might have influenced the findings. Other possible forms of selection bias in this study could be self-selection, as participants who partook were likely more interested or experienced in tracking dietary intake throughout the day. Finally, there may be nonresponse and dropout bias, as participants failing to complete the study were often less motivated or unable to adhere to the sampling scheme due to being in settings (eg, working in the hospital) where using a phone every 2 hours was infeasible. Even though the profile of the FOODLOOP participants is skewed toward higher-educated, urban, health-conscious, active women, therewith limiting the generalizability of our findings to the broader Dutch or millennial population, the profile of our participants aligns with the group most engaged in the snackification trend [112]. Future studies are needed to explore whether the findings can be replicated in more balanced samples by including more men, lower-educated people, people living in more rural areas, and those with lower health interests and activity levels.

Second, sociodemographic and lifestyle variables were measured once at the beginning of the study and were therefore treated as static features in our analyses. However, some of these variables are susceptible to change over time (change in weight, meal patterns and diets, household composition, daily occupation, etc). Such changes were not accounted for during the data collection and might have influenced outcomes concerning the relation between snacking frequencies and sociodemographic and lifestyle variables. Furthermore, residual confounding by unmeasured sociodemographic and lifestyle variables cannot be excluded. Future longitudinal studies exploring the relation of sociodemographic and lifestyle variables with snacking should measure these variables at multiple moments during the data collection to account for possible changes over time and to enable analyses of how changes in these variables affect snacking behaviors.

Third, our findings should also be interpreted in light of the broader postpandemic context in which they took place (2022 - 2023). Even though data collection did not prevail during the COVID-19 pandemic, several behavioral changes that may have emerged during the pandemic (eg, remote and hybrid working, different daily routines, more time spent at home than before) could have persisted into FOODLOOP data collection periods. This might have altered participants' snacking situations and partly explain some of our findings, such as more snacking in general, as well as more home-based and healthful snacking [71-77,113,114]. Future research could examine whether these snacking behaviors are short-term postpandemic adaptations or reflect more stable long-term snacking behavioral changes.

Finally, this study explored a selection of determinants in snacking situations, due to which the generalizability of our findings to other determinants is limited, similar to the study of Verain et al [29]. Capturing snacking situations in their entirety is challenging, given the number of determinants playing a role in real-world snacking situations and the complexity of the interrelatedness of these determinants. However, many of the main determinants in snacking situations [29,115,116] were included in our analyses. The determinants that were not

XSL•FO
RenderX

measured (eg, satisfaction with snack choice, activities executed during snacking, and participant's mood) might constitute interesting topics of exploration for future snacking-related studies.

Implications and Conclusion

Despite the limitations, the FOODLOOP study offers empirical insights into Dutch millennial consumers' real-world snacking behaviors and how multiple situational determinants relate to these behaviors over a longer time period. FOODLOOP shows that snacking is a multifaceted situational behavior, where a comprehensive combination of product-, context-, and consumer-specific determinants plays a role. Confirming the expectations guiding our work, snacking occurs frequently and varies both across and within contexts, including the 4 seasons. This suggests that snacking is not restricted to specific products, motives, locations, social settings, or times, including anything can be a snack, consumed for any reason, anywhere, with anyone, and at any time, congruent with the snackification trend.

The contributions of our findings to research are 2-fold. First, the substantive insights into situational variances in Dutch millennials' real-world snacking behaviors extend prior work on snacking and highlight the importance of its multidimensional nature. Our findings encourage future work to explore situational snacking behaviors across other age groups and cultural contexts and to further investigate interactions between situational determinants, including FCMs, seasonality, and meteorological factors. Future studies may also build on our findings to explore consumers' perceptions and definitions of snacking, factors influencing interindividual variations in definitions of snacking, and the role of beverages as standalone snacks. Second, the methodological experiences could inspire future research on consumers' snacking and other dietary behaviors. FOODLOOP demonstrates the value of using a mobile app based on EMA to study longitudinal, real-world dietary behaviors. The data collection structure, app programming, measure design, and participant communication provide a methodological framework that can be adopted and further refined in future studies on situational snacking and other dietary behaviors.

Beyond research, our findings also have practical relevance. The real-world situational snacking insights gained in FOODLOOP through EMA could support health professionals, nutrition experts, and policymakers in designing more effective and realistic, tailored interventions (eg, campaigns, guidelines, and public health strategies) aimed at encouraging healthier snacking patterns by targeting specific snacking occasions (locations, social settings, moments, and motives) to intervene in the critical moments of situations that are associated with more snacking. Furthermore, our results could inform the development of mobile app–based dietary tools in more effectively guiding healthier snacking choices. For instance, our insights could be used to create context-sensitive features, such as timed prompts and personalized nudges providing healthier snacking options that are suitable for a given context (eg, home alone in the afternoon) and that match users' snacking motives in such situations (eg, convenience, hunger and thirst, and liking).

Overall, the FOODLOOP study underscores the importance of considering snacking as multidimensional and situational behavior within the growing snackification trend. Our insights provide guidance for research and practice in developing more targeted, context-sensitive approaches and interventions that deepen the understanding of snacking and support healthier snacking behaviors.

Acknowledgments

We would like to thank Desiree Lucassen, Edith Feskens, Hanne de Jong, and Arvind Datadien (Wageningen University & Research [WUR]) for providing the Traqq app for usage in the FOODLOOP study. Furthermore, we thank Wouter Rosekrans and Paul Peeters (Voedingscentrum) for their aid in sampling and consultation. We would also like to thank Laura König (University of Bayreuth) and Daniel Gatica-Perez (EPFL, Idiap) for advice concerning the study set-up, Stefan Wahlen (Giessen University) for help with the validation of the translation of the Dutch 1-scale-per-item Eating Motivation Survey, Noanne Pinkert (WUR) for help with the data collection, Luc Scheepens (TU/e) and Zihao Wang (EPFL) for help with data pre-processing, and Jarl Kampen (WUR) for consultation concerning the data analysis. Finally, we thank the FOODLOOP participants for their dedication and participation throughout the year of data collection.

The authors declared that generative artificial intelligence (AI) was used to suggest grammar improvements in the manuscript editing process to refine text. No other generative AI was used in the manuscript writing process.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: MTdV, VF, RV, BLPAS
Data curation: MTdV
Formal analysis: MTdV

XSL•FO
RenderX

Investigation: MTdV
Methodology: MTdV, BLPAS
Project administration: MTdV
Supervision: VF, RV, BLPAS
Visualization: MTdV
Writing – original draft: MTdV
Writing – review & editing: MTdV, VF, RV, BLPAS

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional detailed information on methodological decisions (Traqq app changes; study measures) and data preprocessing, detailed descriptive statistics and statistical test results for the presented findings, and participant attrition details.
[DOCX File, 598 KB - mhealth_v14i1e71858_app1.docx]

References

1. Arora M, Singhal S, Rasane P, et al. Snacks and snacking: impact on health of the consumers and opportunities for its improvement. Curr Nutr Food Sci 2020 Sep 10;16(7):1028-1043. [doi: 10.2174/1573401316666200130110357]
2. Saint Pol TD, Hébel P. Practices and representations of snacking in a highly standardized food culture: the scenario in France. Food Qual Prefer 2021 Oct;93:104245. [doi: 10.1016/j.foodqual.2021.104245]
3. Enriquez JP, Gollub E. Snacking consumption among adults in the United States: a scoping review. Nutrients 2023 Mar 25;15(7):1596. [doi: 10.3390/nu15071596] [Medline: 37049435]
4. Moloughney S. Key trends in functional foods and beverages for 2022 and beyond. Nutraceuticals World. 2021 Nov 11. URL: https://www.nutraceuticalsworld.com/key-trends-for-2022-in-functional-foods-and-beverages/ [accessed 2026-05-17]
5. Lammers P. Diese trends sind gekommen, um zu bleiben. Lebensmittel Zeitung 2021;73(32):38-38. [doi: 10.51202/0947-7527-2021-32-038-1]
6. Dunford EK, Popkin BM. Disparities in snacking trends in US adults over a 35 year period from 1977 to 2012. Nutrients 2017 Jul 27;9(8):809. [doi: 10.3390/nu9080809] [Medline: 28749436]
7. Marangoni F, Martini D, Scaglioni S, et al. Snacking in nutrition and health. Int J Food Sci Nutr 2019 Dec;70(8):909-923. [doi: 10.1080/09637486.2019.1595543] [Medline: 30969153]
8. Román S, Sánchez-Siles LM, Siegrist M. The importance of food naturalness for consumers: Results of a systematic review. Trends Food Sci Technol 2017 Sep;67:44-57. [doi: 10.1016/j.tifs.2017.06.010]
9. Bellisle F. Meals and snacking, diet quality and energy balance. Physiol Behav 2014 Jul;134(C):38-43. [doi: 10.1016/j.physbeh.2014.03.010] [Medline: 24657181]
10. Hess JM, Slavin JL. The benefits of defining “snacks”. Physiol Behav 2018 Sep 1;193(Pt B):284-287. [doi: 10.1016/j.physbeh.2018.04.019] [Medline: 29678598]
11. Njike VY, Smith TM, Shuval O, et al. Snack food, satiety, and weight. Adv Nutr 2016 Sep;7(5):866-878. [doi: 10.3945/an.115.009340] [Medline: 27633103]
12. Zizza CA. Healthy snacking recommendations: one size does not fit all. Physiol Behav 2014 Jul;134(C):32-37. [doi: 10.1016/j.physbeh.2014.01.034] [Medline: 24518869]
13. Mielmann A, Brunner TA. Consumers’ snack choices: current factors contributing to obesity. Br Food J 2019 Feb 4;121(2):347-358. [doi: 10.1108/BFJ-05-2018-0309]
14. Geurts M, Bakel AM, Rossum CTM, Boer E, Ocké MC. Food consumption in the Netherlands and its determinants. : National Institute for Public Health and the Environment; 2017 URL: https://www.rivm.nl/bibliotheek/rapporten/2016-0195.pdf [accessed 2026-05-02]
15. Rossum C, Sanderman-Nawijn E, Brants H, et al. The diet of the Dutch. Results of the Dutch National Food Consumption Survey 2019- 2021 on food consumption and evaluation with dietary guidelines. : National Institute for Public Health and the Environment; 2023. [doi: 10.21945/RIVM-2022-0190]
16. Kahono GA, Rahmawati D, Prawira KPY, Puteri M, Marpaung AM. Preferences and factors influencing the purchase intention of healthy snacks among millennials in Jakarta. Presented at: International Conference on Food, Agriculture and Biotechnology; Jan 24-26, 2018 URL: https://tinyurl.com/yc44kduh [accessed 2026-05-17]
17. Testa R, Rizzo G, Schifani G, et al. Can dried fruits replace unhealthy snacking among millennials? An empirical study on dried fruit consumption in Italy. Sustainability 2023;15(9):7083. [doi: 10.3390/su15097083]
18. Moreno FM, Lafuente JG, Carreón F, Moreno SM. The characterization of the millennials and their buying behavior. Int J Mark Stud 2017;9(5):135. [doi: 10.5539/ijms.v9n5p135]
19. Okumus B. A qualitative investigation of millennials’ healthy eating behavior, food choices, and restaurant selection. Food Cult Soc 2021 Aug 8;24(4):509-524. [doi: 10.1080/15528014.2021.1882168]

XSL•FO
RenderX

20. Shipman DZ. Factors affecting food choices of millennials: how they decide what to eat? J Tourismol 2020 Jun 30;6(1):49-62 [FREE Full text] [doi: 10.26650/jot.2020.6.1.0036]
21. Hess JM, Jonnalagadda SS, Slavin JL. What is a snack, why do we snack, and how can we choose better snacks? A review of the definitions of snacking, motivations to snack, contributions to dietary intake, and recommendations for improvement. Adv Nutr 2016 May;7(3):466-475. [doi: 10.3945/an.115.009571] [Medline: 27184274]
22. Johnson GH, Anderson GH. Snacking definitions: impact on interpretation of the literature and dietary recommendations. Crit Rev Food Sci Nutr 2010 Oct;50(9):848-871. [doi: 10.1080/10408390903572479] [Medline: 20924867]
23. Potter M, Vlassopoulos A, Lehmann U. Snacking recommendations worldwide: a scoping review. Adv Nutr 2018 Mar 1;9(2):86-98. [doi: 10.1093/advances/nmx003] [Medline: 29659681]
24. Gangrade N, St Fleur K, Leak TM. What is a "Snack"? Perspectives from adolescents in urban communities. Ecol Food Nutr 2022;61(4):442-459. [doi: 10.1080/03670244.2021.2020114] [Medline: 34957882]
25. Elliston KG, Ferguson SG, Schüz B. Personal and situational predictors of everyday snacking: An application of temporal self-regulation theory. Br J Health Psychol 2017 Nov;22(4):854-871. [doi: 10.1111/bjhp.12259] [Medline: 28722217]
26. Wahl DR, Villinger K, Blumenschein M, et al. Why we eat what we eat: assessing dispositional and in-the-moment eating motives by using ecological momentary assessment. JMIR Mhealth Uhealth 2020 Jan 7;8(1):e13191. [doi: 10.2196/13191] [Medline: 31909719]
27. Phan UTX, Chambers E. Motivations for choosing various food groups based on individual foods. Appetite 2016 Oct 1;105:204-211. [doi: 10.1016/j.appet.2016.05.031] [Medline: 27235822]
28. Jacquier EF, Gatrell A, Bingley A. "We don't snack": attitudes and perceptions about eating in-between meals amongst caregivers of young children. Appetite 2017 Jan 1;108:483-490. [doi: 10.1016/j.appet.2016.11.003] [Medline: 27818302]
29. Verain MCD, van den Puttelaar J, Zandstra EH, et al. Variability of food choice motives: two Dutch studies showing variation across meal moment, location and social context. Food Qual Prefer 2022 Jun;98:104505. [doi: 10.1016/j.foodqual.2021.104505]
30. Renner B, Sproesser G, Strohbach S, Schupp HT. Why we eat what we eat. The Eating Motivation Survey (TEMS). Appetite 2012 Aug;59(1):117-128. [doi: 10.1016/j.appet.2012.04.004] [Medline: 22521515]
31. Maringer M, Van't Veer P, Klepacz N, et al. User-documented food consumption data from publicly available apps: an analysis of opportunities and challenges for nutrition research. Nutr J 2018 Jun 9;17(1):59. [doi: 10.1186/s12937-018-0366-6] [Medline: 29885653]
32. Müller AM, Maher CA, Vandelanotte C, et al. Physical activity, sedentary behavior, and diet-related eHealth and mHealth research: bibliometric analysis. J Med Internet Res 2018 Apr 18;20(4):e122. [doi: 10.2196/jmir.8954] [Medline: 29669703]
33. König LM, Attig C, Franke T, Renner B. Barriers to and facilitators for using nutrition apps: systematic review and conceptual framework. JMIR Mhealth Uhealth 2021 Apr 1;9(6):e20037. [doi: 10.2196/20037] [Medline: 34254938]
34. Fischer F, Kleen S. Possibilities, problems, and perspectives of data collection by mobile apps in longitudinal epidemiological studies: scoping review. J Med Internet Res 2021 Jan 22;23(1):e17691. [doi: 10.2196/17691] [Medline: 33480850]
35. Sharp DB, Allman-Farinelli M. Feasibility and validity of mobile phones to assess dietary intake. Nutrition 2014;30(11-12):1257-1266. [doi: 10.1016/j.nut.2014.02.020] [Medline: 24976425]
36. Biel JI, Martin N, Labbe D, Gatica D. Bites'n'bits: inferring eating behavior from contextual mobile data. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017;1(125). [doi: 10.1145/3161161]
37. Schembre SM, Liao Y, O'Connor SG, et al. Mobile ecological momentary diet assessment methods for behavioral research: systematic review. JMIR Mhealth Uhealth 2018 Nov 20;6(11):e11170. [doi: 10.2196/11170] [Medline: 30459148]
38. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
39. Agrawal RK, Adhikari R. An introductory study on time series modeling and forecasting. arXiv. Preprint posted online on Feb 26, 2013. [doi: 10.48550/arXiv.1302.6613]
40. Jara-Díaz S, Rosales-Salas J. Understanding time use: daily or weekly data? Transp Res Part A Policy Pract 2015 Jun;76:38-57. [doi: 10.1016/j.tra.2014.07.009]
41. Vier of zes seizoenen? KNMI. 2022 Nov 17. URL: https://www.knmi.nl/over-het-knmi/nieuws/vier-of-zes-seizoenen [accessed 2024-12-14]
42. Coryn CLS, Becho LW, Westine CD, et al. Material incentives and other potential factors associated with response rates to internet surveys of American evaluation association members: findings from a randomized experiment. Am J Eval 2020 Jun;41(2):277-296. [doi: 10.1177/1098214018818371]
43. Pedersen MJ, Nielsen CV. Improving survey response rates in online panels: effects of low-cost incentives and cost-free text appeal interventions. Soc Sci Comput Rev 2016;34(2):229-243. [doi: 10.1177/0894439314563916]
44. Witte N, Schaurer I, Schröder J, Décieux JP, Ette A. Enhancing participation in probability-based online panels: two incentive experiments and their effects on response and panel recruitment. Soc Sci Comput Rev 2023 Jun;41(3):768-789. [doi: 10.1177/08944393211054939]
45. Parker C, Scott S, Geddes A. Snowball sampling. : SAGE Research Methods Foundations; 2019 URL: https://files.core.ac.uk/download/pdf/211022791.pdf [accessed 2026-05-02] [doi: 10.4135/9781526421036831710]

XSL•FO
RenderX

46. Lucassen DA, Brouwer-Brolsma EM, van de Wiel AM, Siebelink E, Feskens EJM. Iterative development of an innovative smartphone-based dietary assessment tool: Traqq. J Vis Exp 2021 Mar 19(169). [doi: 10.3791/62032] [Medline: 33818566]
47. NEVO-online (versie 5.0) [dataset]. : RIVM; 2016 URL: https://www.rivm.nl/documenten/nevo-online-versie-50-2016-dat [accessed 2024-12-14]
48. Nederlands voedingsstoffenbestand NEVO [dataset]. : RIVM; 2019 URL: https://nevo-online.rivm.nl/ [accessed 2024-12-14]
49. Sarcona A, Kovacs L, Wright J, Williams C. Differences in eating behavior, physical activity, and health-related lifestyle choices between users and nonusers of mobile health apps. Am J Health Educ 2017 Sep 3;48(5):298-305. [doi: 10.1080/19325037.2017.1335630]
50. Phan UTX, Chambers E. Application of an eating motivation survey to study eating occasions. J Sens Stud 2016 Apr;31(2):114-123 [FREE Full text] [doi: 10.1111/joss.12197]
51. Voedselconsumptiepeiling. RIVM. 2021. URL: https://www.rivm.nl/voedselconsumptiepeiling [accessed 2024-12-14]
52. Wat eet en drinkt nederland? RIVM. 2021. URL: https://www.wateetnederland.nl/ [accessed 2024-12-14]
53. Wat staat er in de schijf van vijf, en wat niet? Voedingscentrum. 2025. URL: https://www.voedingscentrum.nl/nl/gezond-eten-met-de-schijf-van-vijf/wat-staat-in-de-schijf-van-vijf-en-wat-niet.aspx [accessed 2024-12-14]
54. Stedelijkheid. Centraal Bureau voor de Statistiek. 2019. URL: https://www.cbs.nl/nl-nl/nieuws/2025/12/minder-jongeren-hebben-rijbewijs-dan-in-2019/stedelijkheid [accessed 2024-12-14]
55. Beweegrichtlijnen 2017. Gezondheidsraad. 2017. URL: https://www.gezondheidsraad.nl/site/binaries/site-content/collections/documents/2017/08/22/beweegrichtlijnen-2017/samenvatting-Beweegrichtlijnen-2017.pdf [accessed 2024-12-14]
56. Malkewitz CP, Schwall P, Meesters C, Hardt J. Estimating reliability: a comparison of Cronbach's α, McDonald's ωt and the greatest lower bound. Soc Sci Humanit Open 2023;7(1):100368. [doi: 10.1016/j.ssaho.2022.100368]
57. Hayes AF, Coutts JJ. Use Omega rather than Cronbach's Alpha for estimating reliability. But…. Commun Methods Meas 2020 Jan 2;14(1):1-24. [doi: 10.1080/19312458.2020.1718629]
58. Barrington WE, Beresford SAA. Eating occasions, obesity and related behaviors in working adults: does it matter when you snack? Nutrients 2019 Oct 1;11(10):2320. [doi: 10.3390/nu11102320] [Medline: 31581416]
59. Murakami K, Livingstone MBE. Associations between meal and snack frequency and overweight and abdominal obesity in US children and adolescents from National Health and Nutrition Examination Survey (NHANES) 2003-2012. Br J Nutr 2016 May 28;115(10):1819-1829. [doi: 10.1017/S0007114516000854] [Medline: 27001436]
60. Barnes TL, French SA, Harnack LJ, Mitchell NR, Wolfson J. Snacking behaviors, diet quality, and body mass index in a community sample of working adults. J Acad Nutr Diet 2015 Jul;115(7):1117-1123. [doi: 10.1016/j.jand.2015.01.009] [Medline: 25769747]
61. Liu JL, Han B, Cohen DA. Associations between eating occasions and places of consumption among adults. Appetite 2015 Apr;87:199-204. [doi: 10.1016/j.appet.2014.12.217] [Medline: 25558025]
62. Phan UTX, Chambers E. Motivations for meal and snack times: three approaches reveal similar constructs. Food Qual Prefer 2018 Sep;68:267-275. [doi: 10.1016/j.foodqual.2018.03.018]
63. Okpiaifo GE, Dormoy-Smith B, Kassas B, Gao Z. Perception and demand for healthy snacks/beverages among US consumers vary by product, health benefit, and color. PLoS ONE 2023;18(6):e0287232. [doi: 10.1371/journal.pone.0287232] [Medline: 37327182]
64. Tripicchio GL, Bailey RL, Davey A, Croce CM, Fisher JO. Snack frequency, size, and energy density are associated with diet quality among US adolescents. Public Health Nutr 2023 Nov;26(11):2374-2382. [doi: 10.1017/S1368980023001635] [Medline: 37548183]
65. Bellisle F, Dalix AM, Mennen L, et al. Contribution of snacks and meals in the diet of French adults: a diet-diary study. Physiol Behav 2003 Jul;79(2):183-189. [doi: 10.1016/s0031-9384(03)00088-x] [Medline: 12834789]
66. Hess J, Rao G, Slavin J. The nutrient density of snacks: a comparison of nutrient profiles of popular snack foods using the nutrient-rich foods index. Glob Pediatr Health 2017;4:2333794X17698525. [doi: 10.1177/2333794X17698525] [Medline: 28491924]
67. Myhre JB, Løken EB, Wandel M, Andersen LF. The contribution of snacks to dietary intake and their association with eating location among Norwegian adults - results from a cross-sectional dietary survey. BMC Public Health 2015 Apr 12;15(1):369. [doi: 10.1186/s12889-015-1712-7] [Medline: 25888253]
68. Schlinkert C, Gillebaart M, Benjamins J, Poelman M, de Ridder D. The snack that has it all: people's associations with ideal snacks. Appetite 2020 Sep 1;152:104722. [doi: 10.1016/j.appet.2020.104722] [Medline: 32343990]
69. Vatanparast H, Islam N, Masoodi H, et al. Time, location and frequency of snack consumption in different age groups of Canadians. Nutr J 2020 Aug 15;19(1):85. [doi: 10.1186/s12937-020-00600-5] [Medline: 32799862]
70. Naska A, Katsoulis M, Orfanos P, et al. Eating out is different from eating at home among individuals who occasionally eat out. A cross-sectional study among middle-aged adults from eleven European countries. Br J Nutr 2015 Jun 28;113(12):1951-1964. [doi: 10.1017/S0007114515000963] [Medline: 25907775]
71. Bakaloudi DR, Jeyakumar DT, Jayawardena R, Chourdakis M. The impact of COVID-19 lockdown on snacking habits, fast-food and alcohol consumption: a systematic review of the evidence. Clin Nutr 2022 Dec;41(12):3038-3045. [doi: 10.1016/j.clnu.2021.04.020] [Medline: 34049747]

XSL•FO
RenderX

72. Chee MJ, Koziel Ly NK, Anisman H, Matheson K. Piece of cake: coping with COVID-19. Nutrients 2020 Dec 11;12(12):3803. [doi: 10.3390/nu12123803] [Medline: 33322367]
73. Simone M, Emery RL, Hazzard VM, Eisenberg ME, Larson N, Neumark‐Sztainer D. Disordered eating in a population‐based sample of young adults during the COVID‐19 outbreak. Intl J Eating Disorders 2021 Jul;54(7):1189-1201. [doi: 10.1002/eat.23505]
74. Alamri ES. Effects of COVID-19 home confinement on eating behavior: a review. J Public Health Res 2021 Jun 22;10(3):2088. [doi: 10.4081/jphr.2021.2088] [Medline: 34585555]
75. Ammar A, Brach M, Trabelsi K, et al. Effects of COVID-19 home confinement on eating behaviour and physical activity: results of the ECLB-COVID19 International Online Survey. Nutrients 2020 May 28;12(6):1583. [doi: 10.3390/nu12061583] [Medline: 32481594]
76. Dicken SJ, Mitchell JJ, Newberry Le Vay J, et al. Impact of the COVID-19 Pandemic on diet behaviour among UK adults: a longitudinal analysis of the HEBECO study. Front Nutr 2021;8:788043. [doi: 10.3389/fnut.2021.788043] [Medline: 35096934]
77. Tahlyan D, Mahmassani H, Stathopoulos A, et al. In-person, hybrid or remote? Employers' perspectives on the future of work post-pandemic. Transp Res Part A Policy Pract 2024 Dec;190:104273. [doi: 10.1016/j.tra.2024.104273]
78. Ogden J, Wood C, Payne E, Fouracre H, Lammyman F. "Snack" versus "meal": the impact of label and place on food intake. Appetite 2018 Jan 1;120:666-672. [doi: 10.1016/j.appet.2017.10.026] [Medline: 29074365]
79. Laska MN, Graham D, Moe SG, Lytle L, Fulkerson J. Situational characteristics of young adults' eating occasions: a real-time data collection using personal digital assistants. Public Health Nutr 2011 Mar;14(3):472-479. [doi: 10.1017/S1368980010003186] [Medline: 21138611]
80. Miller R, Benelam B, Stanner SA, Buttriss JL. Is snacking good or bad for health: an overview. Nutr Bull 2013 Sep;38(3):302-322. [doi: 10.1111/nbu.12042]
81. Duffey KJ, Pereira RA, Popkin BM. Prevalence and energy intake from snacking in Brazil: analysis of the first nationwide individual survey. Eur J Clin Nutr 2013 Aug;67(8):868-874. [doi: 10.1038/ejcn.2013.60] [Medline: 23486510]
82. Doppler S, Steffen A. The future of food experiences. In: Case Studies on Food Experiences in Marketing, Retail, and Events 2020:197-210 URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC7453209/ [accessed 2026-02-05] [doi: 10.1016/B978-0-12-817792-1.00016-2]
83. Wang D, van der Horst K, Jacquier E, Eldridge AL. Snacking among US children: patterns differ by time of day. J Nutr Educ Behav 2016 Jun;48(6):369-375. [doi: 10.1016/j.jneb.2016.03.011] [Medline: 27106777]
84. Spence C. Explaining seasonal patterns of food consumption. Int J Gastron Food Sci 2021 Jul;24:100332. [doi: 10.1016/j.ijgfs.2021.100332]
85. Girju M, Ratchford M. The influences of portion size, context and package characteristics on snack food consumption: evidence from a U.S. rolling cross-sectional survey. J Food Prod Mark 2019 Mar 24;25(3):295-321. [doi: 10.1080/10454446.2018.1553745]
86. Seastrom M. Best practices for determining subgroup size in accountability systems while protecting personally identifiable student information. Institute of education sciences congressionally mandated report. : Institute of Education Sciences; 2017 URL: https://nces.ed.gov/pubs2017/2017147.pdf [accessed 2024-12-14]
87. Forbes SL, Kahiya E, Balderstone C. Analysis of snack food purchasing and consumption behavior. J Food Prod Mark 2016 Jan 2;22(1):65-88. [doi: 10.1080/10454446.2014.949992]
88. Peeters B. Gezellig: a dutch cultural keyword unpacked. In: Bromhead H, Ye Z, editors. Meaning, Life and Culture: In Conversation with Anna Wierzbicka: ANU press; 2020:61-79. [doi: 10.22459/MLC.2020.03]
89. van Nes F, Abma T, Jonsson H, Deeg D. Language differences in qualitative research: is meaning lost in translation? Eur J Ageing 2010 Dec;7(4):313-316. [doi: 10.1007/s10433-010-0168-y] [Medline: 21212820]
90. Heenan SP, Dufour JP, Hamid N, Harvey W, Delahunty CM. The sensory quality of fresh bread: descriptive attributes and consumer perceptions. Food Res Int 2008 Dec;41(10):989-997. [doi: 10.1016/j.foodres.2008.08.002]
91. Heenan SP, Hamid N, Dufour JP, Harvey W, Delahunty CM. Consumer freshness perceptions of breads, biscuits and cakes. Food Qual Prefer 2009 Jul;20(5):380-390. [doi: 10.1016/j.foodqual.2009.02.008]
92. Jaeger SR, Antúnez L, Ares G. An exploration of what freshness in fruit means to consumers. Food Res Int 2023 Mar;165:112491. [doi: 10.1016/j.foodres.2023.112491] [Medline: 36869502]
93. Péneau S, Linke A, Escher F, Nuessli J. Freshness of fruits and vegetables: consumer language and perception. Br Food J 2009 Mar 21;111(3):243-256. [doi: 10.1108/00070700910941453]
94. Mercorella SA. Understanding snacking: the role of personality and habit strength [Thesis]. : Adelaide University; 2020 URL: http://hdl.handle.net/2440/131219 [accessed 2026-05-20]
95. Brug J, de Vet E, de Nooijer J, Verplanken B. Predicting fruit consumption: cognitions, intention, and habits. J Nutr Educ Behav 2006;38(2):73-81. [doi: 10.1016/j.jneb.2005.11.027] [Medline: 16595285]
96. Chapelot D. The role of snacking in energy balance: a biobehavioral approach. J Nutr 2011 Jan;141(1):158-162. [doi: 10.3945/jn.109.114330] [Medline: 21123465]

XSL•FO
RenderX

97. McGill CR, Fulgoni VL 3rd, Devareddy L. Ten-year trends in fiber and whole grain intakes and food sources for the United States population: National Health and Nutrition Examination Survey 2001-2010. Nutrients 2015 Feb 9;7(2):1119-1130. [doi: 10.3390/nu7021119] [Medline: 25671414]
98. Vecchio R, Cavallo C. Increasing healthy food choices through nudges: a systematic review. Food Qual Prefer 2019 Dec;78:103714. [doi: 10.1016/j.foodqual.2019.05.014]
99. Hunter SR, Mattes RD. The role of eating frequency and snacking on energy intake and BMI. In: Meiselman HL, editor. Handbook of Eating and Drinking: Interdisciplinary Perspectives: Springer Cham; 2020:1523-1546. [doi: 10.1007/978-3-030-14504-0_115]
100. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12(1):159. [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
101. Spook JE, Paulussen T, Kok G, Van Empelen P. Monitoring dietary intake and physical activity electronically: feasibility, usability, and ecological validity of a mobile-based ecological momentary assessment tool. J Med Internet Res 2013 Sep 24;15(9):e214. [doi: 10.2196/jmir.2617] [Medline: 24067298]
102. Heron KE, Everhart RS, McHale SM, Smyth JM. Using mobile-technology-based ecological momentary assessment (EMA) methods with youth: a systematic review and recommendations. J Pediatr Psychol 2017 Nov 1;42(10):1087-1107. [doi: 10.1093/jpepsy/jsx078] [Medline: 28475765]
103. Lieffers JRL, Arocha JF, Grindrod K, Hanning RM. Experiences and perceptions of adults accessing publicly available nutrition behavior-change mobile apps for weight management. J Acad Nutr Diet 2018 Feb;118(2):229-239. [doi: 10.1016/j.jand.2017.04.015] [Medline: 28625662]
104. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86. [doi: 10.2196/jmir.2583] [Medline: 23598614]
105. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158. [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
106. Tang J, Abraham C, Stamp E, Greaves C. How can weight-loss app designers' best engage and support users? A qualitative investigation. Br J Health Psychol 2015 Feb;20(1):151-171. [doi: 10.1111/bjhp.12114] [Medline: 25130682]
107. Gowin M, Cheney M, Gwin S, Franklin Wann T. Health and fitness app use in college students: a qualitative study. Am J Health Educ 2015 Jul 4;46(4):223-230. [doi: 10.1080/19325037.2015.1044140]
108. Haithcox-Dennis M, Brinkley J, Richman A, DeWeese A, Byrd JL. Mhealth on campus: assessing undergraduates' attitudes and utilization of mobile health applications. Glob J Health Educ Promot 2012;15(1) [FREE Full text]
109. Murnane EL, Huffaker D, Kossinets G. Mobile health apps: adoption, adherence, and abandonment. Presented at: UbiComp and ISWC 2015 - Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the Proceedings of the 2015 ACM International Symposium on Wearable Computers; Sep 9-11, 2015. [doi: 10.1145/2800835.2800943]
110. Woldeyohannes HO, Ngwenyama OK. Factors influencing acceptance and continued use of mhealth apps. In: Lecture Notes in Computer Science: LNCS; 2017, Vol. 10293. [doi: 10.1007/978-3-319-58481-2_19]
111. Oh J, Lee U. Exploring UX issues in quantified self technologies. Presented at: 2015 8th International Conference on Mobile Computing and Ubiquitous Networking; Jan 20-22, 2015. [doi: 10.1109/ICMU.2015.7061028]
112. Hyslop G. How generational preferences are shaping the snacks sector. Food Navigator. 2024. URL: https://www.foodnavigator.com/Article/2024/08/28/how-generational-preferences-are-shaping-the-snacks-sector/ [accessed 2026-05-20]
113. Sideli L, Lo Coco G, Bonfanti RC, et al. Effects of COVID‐19 lockdown on eating disorders and obesity: a systematic review and meta‐analysis. Euro Eating Disorders Rev 2021 Nov;29(6):826-841. [doi: 10.1002/erv.2861]
114. Karthikeyan K, Shalini S, Sudha S, Karthikeyan K, Shalini S, Sudha S. Post-COVID-19 pandemic shift toward health-conscious eating. In: Post-COVID-19 Society and Profound Societal Shifts: InTechOpen; 2025. [doi: 10.5772/intechopen.1011565]
115. Machín L, Giménez A, Vidal L, Ares G. Influence of context on motives underlying food choice. J Sens Stud 2014 Oct;29(5):313-324. [doi: 10.1111/joss.12107]
116. Verain MCD, Sijtsema SJ, Taufik D, Raaijmakers I, Reinders MJ. Motive-based consumer segments and their fruit and vegetable consumption in several contexts. Food Res Int 2020 Jan;127:108731. [doi: 10.1016/j.foodres.2019.108731] [Medline: 31882099]

Abbreviations

DAF: daily average frequency
EMA: ecological momentary assessment
FCM: food choice motive
TEMS: The Eating Motivation Survey

XSL•FO
RenderX

Edited by L Buis; submitted 28.Jan.2025; peer-reviewed by K Mouloudj, Z Zandesh; revised version received 02.Dec.2025; accepted 02.Apr.2026; published 26.May.2026.

Please cite as:
de Vaal MT, Fogliano V, Verkerk R, Steenbekkers BLPA
Exploring Consumers' Situational Snacking Behaviors Using a Mobile App: Longitudinal Cohort Study (FOODLOOP) Among Millennials in the Netherlands
JMIR Mhealth Uhealth 2026;14:e71858
URL: https://mhealth.jmir.org/2026/1/e71858
doi:10.2196/71858

© Mariëlle Tamara de Vaal, Vincenzo Fogliano, Ruud Verkerk, Bea L P A Steenbekkers. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 26.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Smartphone-Based Ecological Momentary Intervention for Workplace Mental Health: Randomized Controlled Trial

Sisi Li[1], PhD; Lik Hang Lincoln Lo[2], PhD; Chiu Yi Charlie Lau[2], MPH; Megan Lam[3], MPH; Meanne Chan[2,4], PhD

[1]Department of Community Health and Behavioral Medicine, School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[2]Wofoo Joseph Lee Consulting and Counselling Psychology Research Centre, Lingnan University, Tuen Mun, China (Hong Kong)

[3]Neurum Health, Central Hong Kong Island, China (Hong Kong)

[4]Department of Population Health and Policy, School of Health and Medical Sciences, City St. George's, University of London, Cranmer Terrace, London, United Kingdom

Corresponding Author:
Meanne Chan, PhD
Wofoo Joseph Lee Consulting and Counselling Psychology Research Centre, Lingnan University, Tuen Mun, China (Hong Kong)

Abstract

Background: Work-related stress has been widely associated with an increased risk of various mental disorders and poor mental well-being. The fast-growing mobile health services industry has provided new opportunities for workplace mental health.

Objective: This randomized controlled trial examined the effectiveness of *Neurum* (Neurum Limited), a smartphone-based intervention tool featuring ecological momentary assessments and interventions that aims to reduce workplace stress in real-time and real-world settings.

Methods: A total of 201 working adults were recruited for a 4-week smartphone-based intervention that incorporated cognitive behavioral therapy, mindfulness exercises, and self-regulation exercises delivered on *Neurum*. A simple randomization procedure was used. Participants in the intervention group were encouraged to log mood journals, complete mental health exercises, and provide user feedback whenever applicable. The key outcome was measured by the Depression, Anxiety, and Stress Scale-21 items (DASS-21; Cronbach α=0.87).

Results: The intervention group consisted of 102 participants, while the control group consisted of 99 participants. More participants dropped out from the intervention group (n=21) than from the control group (n=2; χ^2_1=18.259; *P*<.001). The final sample consisted of 178 participants (male: 85/178, 47.8%; female: 93/178, 52.2%; mean age of 34.65, SD 7.67 y). Analyses revealed that after the 4-week intervention, the DASS-21 scores decreased in the intervention group (mean difference [MD]$_{\text{post-pre intervention}}$=−14.518) but increased in the control group (MD$_{\text{post-pre intervention}}$=3.319; $F_{1,176}$=59.358, *P*<.001; η^2=0.252). This effect was largely led by stress reduction ($F_{1,176}$=64.679, *P*<.001; for the intervention group, MD$_{\text{post-pre intervention}}$=−6.692, while for the control group, MD$_{\text{post-pre intervention}}$=2.000). On average, participants completed 6.27 (SD 9.4) exercises and provided 9.74 (SD 18.2) mood journal logs, with a daily engagement of 4.95 (SD 6.89) minutes. However, the associations between the changes in DASS-21 scores and the numbers of exercises or mood journal logs did not reach statistical significance.

Conclusions: This study primarily established the effectiveness of *Neurum* in alleviating depression, anxiety, and stress symptoms in noninstitutionalized working adults, with a satisfactory user retention rate. Despite potential health-related culture differences, *Neurum* contributed to evidence-based digital health in nonclinical settings for timely needs and general accessibility as an alternative to traditional, face-to-face, and high-cost mental health services. Future directions involving a personalized approach in online mental health services were discussed.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2600131522; https://tinyurl.com/mpdrfdvx

(JMIR Mhealth Uhealth 2026;14:e72435) doi:10.2196/72435

KEYWORDS

workplace stress; mobile health; mHealth; online intervention; ecological momentary intervention; EMI; ecological momentary assessment; EMA; digital intervention

XSL•FO
RenderX

Introduction

Background

While the working population is the driving force of the global economy, common mental disorders (ie, depression and anxiety) alone have cost US $1 trillion each year due to the loss of productivity [1,2]. Pharmacotherapies, psychotherapies, or a combination of both are the primary approaches in managing the symptoms of common mental disorders [3-5]. However, suboptimal service use exists to hinder the influence of the mental health care system [6,7]. The poor mental health service use suggested an immense service gap that might not sufficiently or appropriately support the needs of individuals with mental health difficulties. Against this backdrop, it has become essential to further explore accessible, affordable, and effective alternatives to existing mental health care services for the working population.

Longitudinal studies have demonstrated that work-related stress—such as excessive workload, effort-reward imbalance, and low work social support—is associated with increased risks of depression and anxiety disorders and poor mental well-being [8-11]. Despite a large investment in mental health services and their significant effectiveness, the cost of maintaining these face-to-face interventions is high and often challenging, steering the research direction towards identifying low-cost, evidence-based strategies to optimize intervention cost-efficiency [12-18]. Worse yet, the COVID-19 pandemic has further threatened the mental health landscape of working adults by impacting individuals' job security, chronic isolation, and uncertainty about the future, as well as adding constraints to traditional face-to-face intervention options [19]. Most of the existing interventions are limited in reaching out to those in need during a pandemic situation, increasing the risk and prevalence of mental ill health [19-22]. This limitation also applies to long-existing underuse issues of mental health services, especially in Asian countries [23,24]. For illustration, previous research in China revealed that only 24% of the participants who were diagnosed with moderate or severe mental health problems sought help from mental health professionals [25]. Reasons included issues such as health service accessibility and costs, as well as associated mental health stigmas. Therefore, mobile health (mHealth) seems a practical approach to improve the low rate of health-seeking behaviors given its advantages of being low cost, easily accessible, and stigma-free. Therefore, attention has increasingly shifted toward digital approaches as a feasible and scalable alternative.

A recent randomized controlled trial (RCT) showed that web-based mindfulness and cognitive behavioral training were efficacious in alleviating psychological distress among college students and young working adults [26]. Several meta-analyses suggested that web-, application-, and/or computer-based interventions were effective in alleviating depression, anxiety, and stress, though with undeniably concerning attrition rates ranging widely from 43% to 99% [27-31]. Taken together, these findings pointed out the effectiveness of internet-based psychological interventions while urging further refinement of their user retention especially for long-term benefits. In addition to internet-based interventions, mHealth platforms that adopt mobile and wireless devices to improve health care also appear promising [32]. Not only has mHealth long been used in physical health care, such as for diabetes management, infectious disease management, and immunization management, increasing application of mHealth to mental health also has emerged, ranging from alcohol and substance use assessment to low-intensity interventions for mental disorders [33-38]. Despite the enlightening trend, less than 5% of the mHealth platforms underwent peer-reviewed research, while the majority of these were conducted in western countries [39,40]. Peer-reviewed, demographically diverse research awaits attention.

Advancements, availability, and affordability of mobile technology have given rise to ecological momentary assessment (EMA), which allows individuals to report experiences and behaviors in everyday lives and in naturalistic settings [41,42]. The rationale for using EMA rests upon 3 major advantages including reducing retrospective recall, improving generalizability and ecological validity, and enabling the exploration of temporal relationships among variables. Similarly, ecological momentary intervention (EMI) refers to the delivery of intervention while individuals are naturally engaged in their everyday lives [43,44]. Emerging evidence suggested the effectiveness of EMI in alleviating anxiety and stress and treating a variety of health behaviors and psychological symptoms [44,45]. However, RCTs of EMA and EMI in mental health promotion remained scarce [45]—and their application in workplace mental health services is even more limited. This is noteworthy given that both approaches could offer unique benefits to working adults, as they deliver mental health care services free from temporal and spatial constraints. Therefore, this gap underscores the need for rigorous investigation of EMA- and EMI-based mHealth tools, particularly in workplace populations and non-Western contexts.

Objectives

To address this gap, this preregistered RCT was designed to evaluate the effectiveness of a smartphone-based intervention (ie, the *Neurum* app; Neurum Limited) that leverages both EMA and EMI in reducing workplace stress [46]. Study hypotheses and corresponding analytical schemes were preregistered in September 2020 (registration reference: osf.io/z829u) before the data collection was finished. No statistical analysis was done before the preregistration or the completion of data collection. Specifically, it was hypothesized that participants who received intervention through the *Neurum* app would display better workplace stress management and mental health improvement as reflected by the reduction in depression, anxiety, and/or stress symptoms. The design, conduct, and reporting of this study followed the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines (Checklist 1).

Methods

Participants

This study was a 2 (between: intervention vs waitlist control groups) × 2 (within: premeasures vs postmeasures) mixed factorial design. A priori power analysis using *F* test family

XSL•FO
RenderX

(repeated measures, between factors) in *G*Power* was conducted [47]. Assuming a small effect size (ie, f=0.25) and a high correlation between premeasures and postmeasures of the outcome variable (ie, r=.85), with an α level of .05, a total number of 196 participants was needed to achieve a power of 0.80. Assuming an attrition rate of 25%, a sample of 262 participants was determined prior to data collection.

Participant recruitment started in September 2019, Hong Kong SAR. Advertisements were posted on the company websites and populated through themed webinars and offline workshops to reach local working adults. However, due to local social unrest and the COVID-19 pandemic, few signed up within the planned 1-month recruitment period. Therefore, under all authors' and the funding agency's approval, the recruitment process was prolonged until the deemed sample size would be achieved. The recruitment process lasted for over a year and was finally terminated by the end of October 2020, when nearly no more participants signed up within 30 days. To maximize the sample size, those who signed up and passed the screening immediately underwent the intervention process. However, no data analysis was conducted before the end of the study. In total, 201 working adults were recruited from a selected list of local companies, who were first directed to an online survey with screening questions, in which they reported their demographic and mental health information. A research assistant who remained unaware of the study design was in charge of the eligibility screening. Specifically, participants were eligible if they (1) were aged 18 years or older at sample recruitment, (2) had access to the internet and a smartphone, and (3) were not diagnosed with major mental disorders. Exclusion criteria included (1) mental illness diagnosis history, (2) current reception of professional mental health services, or (3) current use of psychotropic medication.

Ethical Considerations

Ethical approval was obtained from the Subcommittee on Research Ethics and Safety of the Research Committee of Lingnan University (reference: EC047/2021). All procedures contributing to this work complied with the ethical standards of the relevant national and institutional committees on human study. Informed consent was obtained via electronic signatures indicating the awareness of the study procedure, potential risks and benefits, data confidentiality, and the right to quit the study at any time. Data were encrypted and stored in a financial-sector security level server and deidentified before statistical analysis. No incentives or compensations were provided.

Procedure

All procedures were delivered via *Neurum*. Included participants first completed the preintervention assessment, and then they were randomly allocated at a 1:1 ratio to the two groups via a computer algorithm of simple randomization. Specifically, a numerical value between 0 and 1 was assigned to each participant. Those who received a value smaller than 0.5 were assigned to the control group; otherwise, participants were assigned to the intervention group. Double blindness was achieved such that a research assistant who remained unaware of the study design was in charge of the participant assignment, while participants received a unique study ID in a singular format that revealed no group information, and therefore, remained unaware of the grouping. Then, participants in the intervention group received EMA and EMI via *Neurum* for 4 weeks, while those in the waitlist control group were offered the intervention after the RCT was completed. Finally, all participants completed a postintervention assessment with debrief at the end of the trial (ie, after the designated 4-week intervention [or control] duration).

Intervention

Neurum is a smartphone app that comprises self-managed mental health modules designed in various combinations of (1) featured techniques, namely cognitive behavioral therapy, mindfulness exercises, and self-regulation skills with (2) various session formats including informational (ie, psychoeducation), evaluative (ie, assessment), and applicable (ie, exercise), aiming at (3) tangible goals in alignment with the biopsychosocial model, namely the improvement and maintenance of biological, psychological, and social health. See Table 1 for a summary of the different components and the rationale.

Table . *Neurum* intervention framework and module examples.

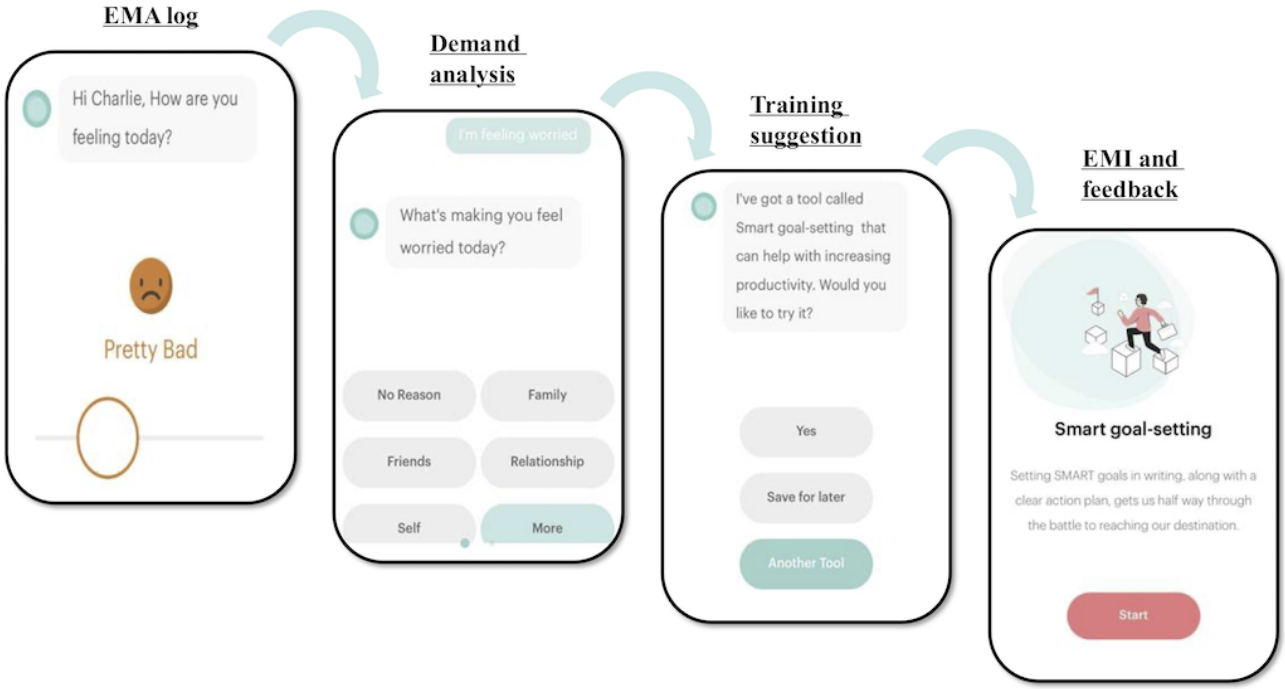
Example module	Featured technique	Format		
		Psychoeducation	Assessment	Exercise
Biological				✓
Acupressure	Self-regulation			✓
Sleep quality	Mindfulness	✓		✓
Psychological				
Focus timer	Self-regulation			✓
Managing stress	Cognitive-behavioral therapy	✓	✓	✓
Sustaining happiness	Mindfulness	✓		✓
Social				
Workplace boundaries	Self-regulation	✓	✓	✓

XSL•FO
RenderX

Neurum is designed with an instant messenger–like user interface where users interact with the virtual app assistant in a conversation format. The palatability of the dialogues was ensured by using layman and local terms in either English or Chinese. The app content was developed with standard app development iterations that involved digitizing of modules based on previously validated mental health education and behavioral changing paradigms, and then user testing with an A/B test method to compare the performance of 2 versions of content. This study describes the foundation of the app, which matches users to an EMI based on information from their EMA via a combination of supervised learning models, but future iterations of *Neurum* involved personalization of EMI modules based on nuances from EMA, session feedback, and other baseline user data via a more complex deep learning algorithm. As part of the safeguarding protocol, a referral mechanism is established, such that users who score past a certain threshold in their levels of psychological distress are automatically referred to crisis management protocol, which involves the algorithm providing the user with immediate support via distress reduction modules, crisis hotlines, and face-to-face care delivered by a licensed, on-board clinical psychologist. No adverse events were reported during the study.

EMA is delivered such that users are invited to complete a mood journal on *Neurum* to log their instant emotions, thoughts, and behaviors relevant to workplace stress as well as other psychological distress whenever they feel in need. Users evaluate their general feelings along a negative-to-positive continuum, specify a specific experience they would like to work on, and then identify one or more associated stressors. EMI is delivered such that, upon EMA logs, *Neurum* would make a recommendation to relevant, evidence-based intervention modules based on an immediate matching of the current demand and available training tools using a self-developed, built-in supervised machine learning algorithm. User feedback is further collected after each training session, which also informs the algorithm. A daily summary of the reported mood experiences and taken exercises was provided for self-monitoring of workplace stress and management performance. See Figure 1 for the overview of *Neurum*'s EMA or EMI workflow.

Figure 1. Illustration of ecological momentary assessment (EMA) and ecological momentary intervention (EMI) workflow on the Neurum (Neurum Limited) app.



Measures

Depression, Anxiety, and Stress

Key outcomes, namely the intervention effectiveness, were evaluated by comparing the prescores and postscores of the Depression, Anxiety, and Stress Scale-21 items (DASS-21) [48]. Participants rated the extent to which they experienced depression, anxiety, and stress symptoms (7 items for each dimension; eg, "I was worried about situations in which I might panic") on a 4-point scale (0="Did not apply to me at all" to 3="Applied to me very much or most of the time"). Sum scores were computed for the 3 subdimensions as well as the total scores such that higher scores reflected more severe symptoms. The adopted Chinese version of the DASS-21 exhibited satisfactory psychometric properties (Cronbach $\alpha \geq 0.80$) [49]. Cronbach α in this study was 0.87.

Mood Journal

Participants in the intervention group were invited to report, every time they logged in, on (1) how they were feeling in general; (2) whether they were feeling any specific emotion, such as happy or disappointed; and if yes, (3) the source of emotion such as family, work, or friends.

User Activity

User activity was measured by the number of mood journals completed and exercise sessions taken, as well as daily user engagement (in min).

XSL•FO
RenderX

Other Measures

At baseline, participants provided demographic information including sex, age, and marital status.

Analytical Scheme

To examine the intervention effect, pre– and post–DASS-21 scores of the intervention and control groups were submitted to a 2 (between: intervention vs control groups) × 2 (within: premeasures vs postmeasures) mixed ANOVA. Additional analyses included categorical demographic characteristics (namely sex, education, and marital status; Table 2) as control variables. A supplementary set of mixed ANOVA was conducted to compare intensity level changes for each symptom. Finally, 2 multiple linear regressions with a number of mood journals and exercise sessions completed were conducted to explore any dose-response effect.

Table . Demographic characteristics and baseline Depression, Anxiety, and Stress Scale-21 items (DASS-21) scores for each group.

Characteristics	Control group (n=97)	Intervention group (n=81)
Age (y), mean (SD)	34.65 (7.67)	36.43 (12.84)
Sex, n (%)		
Female	51 (53)	42 (52)
Male	46 (47)	39 (48)
Education, n (%)		
High school	5 (5)	8 (10)
Bachelor’s degree	64 (66)	48 (59)
Master’s degree	0 (0)	18 (22)
PhD or above	6 (6)	3 (4)
Prefer not to tell	22 (23)	4 (5)
Marital status, n (%)		
Single	54 (56)	40 (49)
Married	33 (34)	38 (47)
Divorced	7 (7)	2 (2)
Separated	3 (3)	0 (0)
Prefer not to tell	0 (0)	1 (1)

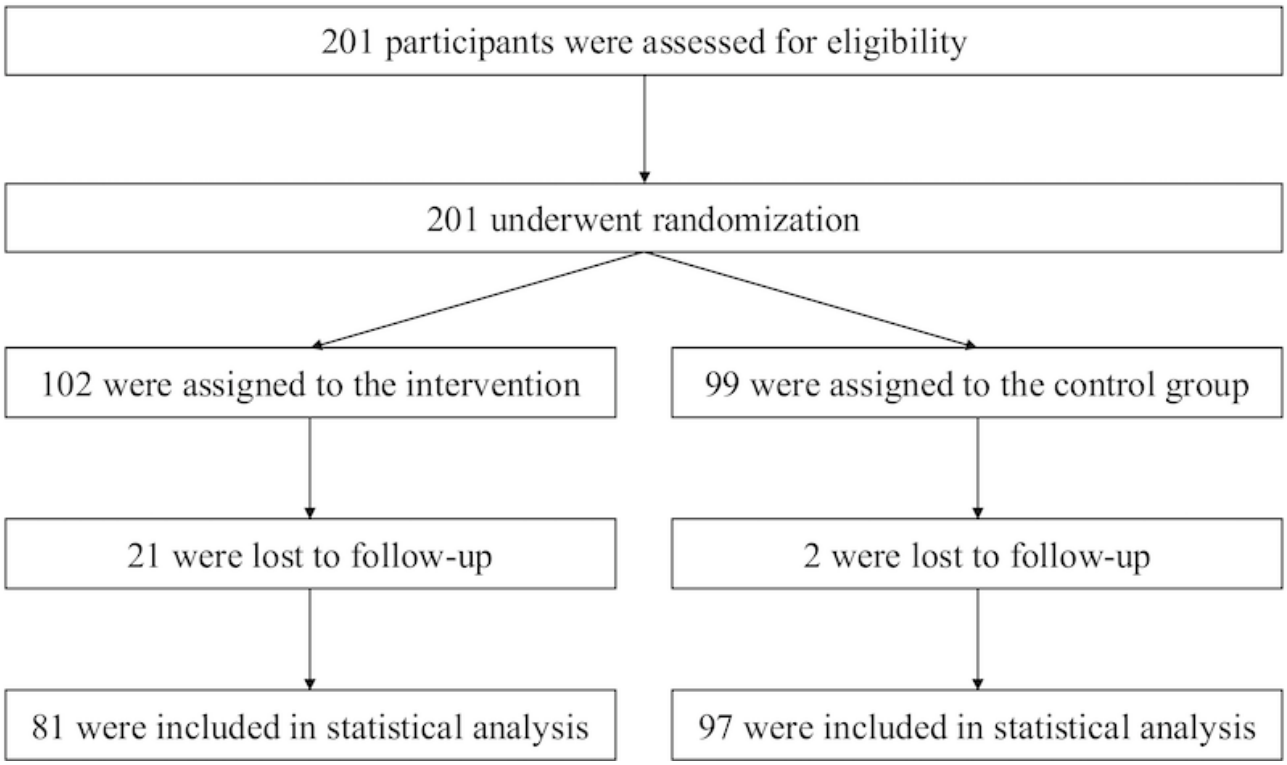
Results

Sample Characteristics

Figure 2 shows the participant flow diagram. Demographic characteristics are summarized in Table 2. The majority of study participants were females, had bachelor’s degrees, and were single in terms of marital status. Though the overall attrition rate was low (23/201, 11.4%), there were significantly more participants dropped from the intervention group (n=21) than from the control group (n=2; χ^2_1=18.259; P<.001), resulting in 81 participants in the intervention group and 97 in the control group. Considering the intensive EMA or EMI procedures, the differential attrition rates between the 2 groups were not surprising. More importantly, the final sample consisted of 178 working adults and was still sufficient in statistical power.

XSL•FO
RenderX

Figure 2. Participant flow diagram.



User Activity

On average, participants in the intervention group provided 9.74 (SD 18.2; maximum=130) mood journals and completed 6.27 (SD 9.4; maximum=51) exercises, while neither was correlated to preintervention DASS-21 scores ($|r|$s ≤.156; Ps≥.16), and neither was predicted by sex (t_{79}s≤0.744; Ps≥.46), education level ($F_{3,\ 73}$s≤1.947; Ps≥.13), or marital status ($F_{2,\ 77}$s≤2.002, Ps≥.14). Daily user engagement on average was 4.95 (SD 6.89; maximum=30) minutes, which was positively correlated with preintervention stress scores (r=0.233; P=.002) and total DASS-21 scores (r=0.175; P=.02), but was not predicted by sex (t_{78}=0.791; P=.22), educational level ($F_{3,\ 73}$=1.595; P=.20), or marital status ($F_{2,\ 76}$=0.100; P=.91).

Intervention Effects

Preintervention and postintervention DASS-21 scores, as well as the mean differences, are presented in Table 3. Mixed ANOVAs were applied to compare the pre– and post–DASS-21 total and subscale scores between groups (Table 4). For the total score, equal variance between groups was satisfied (Levene statistics ≤2.127; P≥.15). Results revealed that there was no significant difference between the intervention (mean difference [MD]=21.990) and the control (MD=21.852) groups; but as expected, there was a significant postintervention decrease ($MD_{preintervention}$=24.721 and $MD_{postintervention}$=19.121). More importantly, the above effects were quantified by a significant, large-effect size interaction between time and group (Figure 3A) such that the DASS-21 total scores decreased for the intervention group ($MD_{preintervention}$=29.111 and $MD_{postintervention}$=14.593) but increased for the control group ($MD_{preintervention}$=20.330 and $MD_{postintervention}$=23.649). Similar patterns (Figure 3B-D) were found for the three symptoms such that average scores notably decreased for the intervention group but slightly increased for the control group ($F_{1,\ 176}$s≥22.689, Ps<.001), with the strongest effect observed in the stress symptom.

Table . Preintervention and postintervention scores and mean differences between groups ($n_{control}$=97 and $n_{intervention}$=81).

Variable	Preintervention, mean (SD)		Postintervention, mean (SD)		Mean difference[a]	
	Control	Intervention	Control	Intervention	Control	Intervention
Depression	6.866 (6.843)	8.938 (7.417)	7.588 (7.071)	4.691 (7.015)	0.722	−4.247
Anxiety	5.361 (6.289)	7.185 (5.812)	5.959 (5.752)	3.605 (4.398)	0.598	−3.580
Stress	8.103 (7.248)	12.988 (7.76)	10.103 (7.796)	6.296 (6.520)	2.000	−6.692
Total	20.33 (16.283)	29.111 (17.975)	23.649 (16.857)	14.593 (15.174)	3.319	−14.518

[a]Mean differences are computed as postintervention minus preintervention scores.

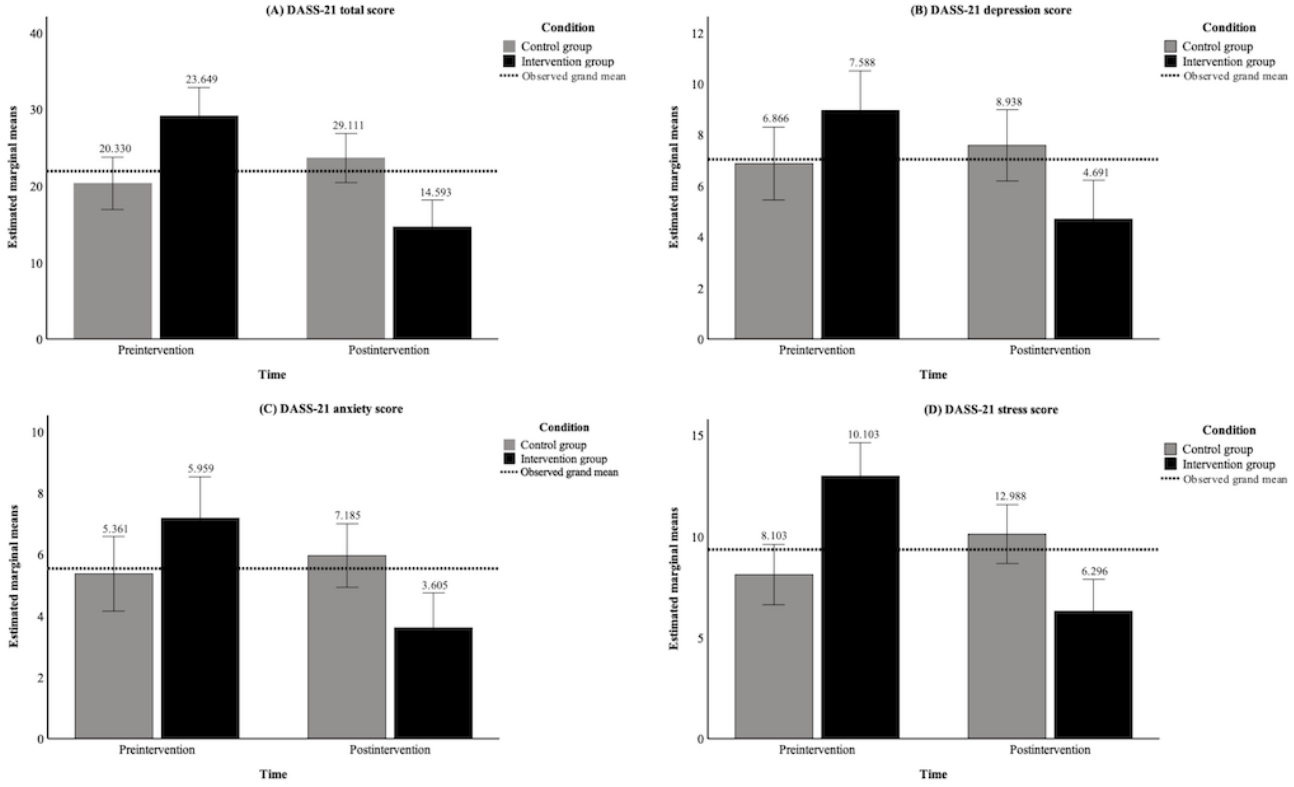
XSL•FO
RenderX

Table . Mixed ANOVAs of preintervention and postintervention Depression, Anxiety, and Stress Scale-21 items (DASS-21) scores between groups ($n_{control}$=97 and $n_{intervention}$=81).

Outcome	*F* test (*df*)	*P* value	η^2
Total			
Between			
Group	0.004 (1, 176)	.95	0.000
Within			
Time	23.396 (1, 176)	<.001	0.117
Time × group	59.358 (1, 176)	<.001	0.252
Depression			
Between			
Group	0.197 (1, 176)	.66	0.001
Within			
Time	11.422 (1, 176)	.001	0.061
Time × group	22.689 (1, 176)	<.001	0.114
Anxiety			
Between			
Group	0.122 (1, 176)	.73	0.001
Within			
Time	15.182 (1, 176)	<.001	0.079
Time × group	29.800 (1, 176)	<.001	0.145
Stress			
Between			
Group	0.310 (1, 176)	.58	0.002
Within			
Time	18.844 (1, 176)	<.001	0.097
Time × group	64.679 (1, 176)	<.001	0.269

XSL•FO
RenderX

Figure 3. Preintervention and postintervention Depression, Anxiety, and Stress Scale-21 items (DASS-21) results including (A) total score and (B) depression, (C) anxiety, and (D) stress subdimensional scores between intervention and control groups. Error bars represent 95% CIs.



Supplementary Analyses

A set of supplementary models found similar patterns controlling for sex, education, and marital status (for condition effects, $F_{1, 146}$s≤1.601, *P*s≥.21, η^2≤0.011; for time effects, $F_{1, 146}$s≥2.933, *P*s≤.09, η^2≥0.020; for interaction effects $F_{1, 146}$s≥19.286, *P*s≤.001, η^2≥0.117), suggesting that the above intervention effects held after controlling for these demographic covariates.

Another set of supplementary models (Table 5) found similar patterns using DASS-21 symptom intensity levels as the outcome variables. Change(s) of these intensity levels suggested that the control group ascended while the intervention group descended in symptom severity (χ^2_2s≥ 8.605; *P*s≤.02). Finally, and surprisingly, a final set of supplementary models (Table 6) using multiple linear regressions revealed that none of the exercise session number, mood journal number, or average daily engagement predicted the postintervention DASS-21 scores.

Table . Intensity level changes between groups ($n_{control}$=97 and $n_{intervention}$=81)[a].

	Intensity level change			Chi-square (*df*)	*P* value
	Worsen, n (%)	No change, n (%)	Improved, n (%)		
Depression					
Control	17 (18)	64 (66)	16 (16)	8.605 (2)	.02
Intervention	9 (11)	44 (54)	28 (35)		
Anxiety					
Control	21 (22)	67 (69)	9 (9)	27.581 (2)	<.001
Intervention	5 (6)	43 (53)	33 (41)		
Stress					
Control	19 (20)	72 (74)	6 (6)	23.404 (2)	<.001
Intervention	4 (5)	52 (64)	25 (31)		

[a]Cell values in each intensity change category are counts of participants.

XSL•FO
RenderX

Table . User activity and postintervention Depression, Anxiety, and Stress Scale-21 items (DASS-21) scores for the intervention group (n=81).

	B	SE	*P* value	95% CI
Total score[a]				
Model 1				
Baseline level	0.220	0.092	.02	0.037 to 0.403
Model 2				
Baseline level	0.224	0.093	.02	0.038 to 0.410
Exercise session	−0.073	0.194	.71	−0.459 to 0.313
Mood journal	0.120	0.099	.23	−0.077 to 0.318
Daily engagement	0.240	0.253	.35	−0.263 to 0.744
Depression[b]				
Model 1				
Baseline level	0.286	0.102	.006	0.083 to 0.489
Model 2				
Baseline level	0.296	0.105	.006	0.086 to 0.505
Exercise session	0.001	0.088	.99	−0.175 to 0.177
Mood journal	0.027	0.046	.59	−0.064 to 0.119
Daily engagement	0.029	0.117	.80	−0.204 to 0.262
Anxiety[c]				
Model 1				
Baseline level	0.241	0.081	.004	0.080 to 0.403
Model 2				
Baseline level	0.228	0.083	.008	0.062 to 0.394
Exercise session	−0.033	0.056	.55	−0.144 to 0.078
Mood journal	0.018	0.029	.52	−0.038 to 0.075
Daily engagement	0.086	0.073	.25	−0.060 to 0.232
Stress[d]				
Model 1				
Baseline level	0.201	0.092	.03	0.018 to 0.385
Model 2				
Baseline level	0.200	0.092	.03	0.018 to 0.383
Exercise session	−0.035	0.082	.67	−0.199 to 0.128
Mood journal	0.077	0.042	.08	−0.007 to 0.160
Daily engagement	0.129	0.108	.23	−0.085 to 0.343

[a]For model 1, adjusted R^2=0.068, $F_{1,78}$=5.731, P=.02; and for model 2, adjusted R^2=0.097, $F_{1,72}$=0.777, P=.51.

[b]For model 1, adjusted R^2=0.080, $F_{1,78}$=7.855, P=.006; and for model 2, adjusted R^2=0.049, $F_{1,72}$=0.162, P=.92.

[c]For model 1, adjusted R^2=0.091, $F_{1,78}$=8.866, P=.004; and for model 2, adjusted R^2=0.076, $F_{1,72}$=0.601, P=.62.

[d]For model 1, adjusted R^2=0.045, $F_{1,78}$=4.762, P=.03; and for model 2, adjusted R^2=0.067, $F_{1,72}$=1.593, P=.20.

Discussion

Overview

This RCT examined the effectiveness of a smartphone-based EMA or EMI mental health care services delivered by the *Neurum* app in mitigating workplace stress and improving mental health for working adults in Hong Kong SAR, China. Consistent with the hypotheses, results demonstrated that the approach was effective in reducing psychological distress after 4 weeks of intervention, suggesting that the EMA or EMI approach via the *Neurum* app could support working adults' stress management in real time and in naturalistic settings.

XSL•FO
RenderX

Principal Findings

This study demonstrated the effectiveness of integrating EMA or EMI techniques with supervised machine learning algorithms in digital mental health care for a nonclinical population. The findings were broadly compatible with previous studies supporting the effectiveness of mHealth tools in reducing general anxiety, various stress, and depression symptoms [45,50]. Importantly, the effectiveness of *Neurum* was supported not only by the continuous reduction of psychological symptoms but also by the shifts in symptom intensity levels, suggesting potential use that extends from non–clinical contexts into early-stage clinical applications. One well-known advantage of the ecological momentary approach is its ability to minimize retrospective bias and enhance event trigger monitoring. Previous research has highlighted the promise of such digital phenotyping for routine clinical assessment, particularly for early onset or relapse [51]. Our observed changes in categorical symptom intensity levels further supported this clinical relevance, indicating that symptom improvement was substantial enough to cross the established cutoff thresholds. Because categorical severity shifts could provide a more interpretable indicator of treatment effectiveness than numerical score changes alone, these findings underscore the potential of EMA or EMI-based mHealth tools for early identification, timely intervention, and preventative health care in mental health. Additionally, previous meta-analyses research revealed that multicomponent intervention was tentatively more effective than single-component ones (eg, with cognitive behavioral therapy alone), and that patient preferences for specific intervention components played a pivotal role in treatment outcomes [52,53]. In response to these challenges, the intervention delivered in *Neurum* integrated multiple modules in various delivery formats for actionable outcomes, enabling participants to decide whether and which exercises to engage in upon each log-in. This inclusive design undoubtedly promoted user retention and engagement and maximized the alignment between individual needs and intervention content, a pattern supported by our user activity data. Taken together, these novel findings highlighted the importance of developing scalable, personalized mHealth tools as accessible, affordable, and flexible alternatives to face-to-face mental health care services.

There were several important issues in the interpretation of results. First of all, more participants dropped out of the intervention group than from the control group. It is noteworthy that the overall dropout did not exceed the predetermined attrition rate for power analysis. It is also noteworthy that the current attrition rate (ie, 11.4%) was smaller than those reported in previous studies (eg, about 26% for cognitive behavioral therapy and about 19% for mindfulness-based traditional interventions, or widely ranging from 43% to 99% for general digital health interventions [27-31,54-56]). Despite this, authors speculated the presence of some exposure effects particularly for the intervention group. Specifically, EMA or EMI required participants to review their mood swings on a daily basis, to exert efforts in managing especially those negative daily experiences, and to review their own performances on the taken exercises. Compared to those assigned to the control groups, these tasks might make the mental health issues more evident than they used to be (namely, before participating in this study) for the intervention group, possibly resulting in a withdrawal from treatment effect frequently observed in traditional psychotherapies [57]. Regrettably, the current design did not include relevant measures to test this speculation. However, to improve its future retention, the intervention could benefit from including methods to cope with the aforementioned exposure effect in the onboarding stage. Meanwhile, a personalized approach that maps personal needs with module recommendations can also improve user retention and performance. In fact, future iterations of *Neurum* have involved personalization of EMI modules based on nuances from EMA, session feedback, and other baseline user data via a more complex deep learning algorithm. Second, the results somewhat failed to support the association between the number of completed exercises or mood journals with DASS-21 score changes. In other words, the underlying dose-response effect remained unclear. One possible reason was that participants were only *encouraged* but not forced to report their daily mood journals or to take exercises. Granted that it strategically maximized participants' acceptance and adherence [58], it undeniably resulted in the limitation of a relatively low daily response rate, and subsequently, the lack of statistical power to reveal the targeted dose-response effect, if any. Therefore, future studies might consider assigning a compulsory number of daily responses to unveil potential underlying mechanisms. Despite this, the current intervention effects could be a result of the improved awareness of one's own mental health status and consequently improved proactive mental health management effort, or an effect generalized from in-app exercises to realistic life scenarios.

This study also contributes to the relatively nascent field of mHealth research, typically the application of EMA or EMI, in Asian countries [40]. Though with a rapid growth and a wide application, empirical supports of EMA or EMI for mental health were mostly from the United States and European countries (eg, Germany and the Netherlands), with the only exception being a study involving Chinese participants with mild-moderate depression [59-61]. However, western-based mHealth apps were not necessarily applicable to eastern ones [62]. The common mHealth apps in Asian contexts have remained in psychoeducation, counseling services, and self-assessment tools [40,63], while this study expanded them by leveraging the more intense EMA or EMI approach. Furthermore, *Neurum* takes health equity and cultural adaptivity into consideration by offering multiple language versions (viz, both traditional and simplified Chinese, English, Filipino, and Hindi) to serve local people with different cultural backgrounds. Though ethnicity was not collected in this investigation, the recruitment process was indeed carried out in international enterprises in Hong Kong, SAR. Finally, this study expanded the use of EMA or EMI beyond clinical settings to community populations, which broadened the spectrum of the recipients of such digital health care services. While some digital health tools can be used by patient populations to clinically reduce their depression and/or anxiety, this must be differentiated from digital health tools that are meant to reduce these symptoms in a non–clinical population. The demonstrated effectiveness of *Neurum* is similar to other well-being and mindfulness apps,

XSL•FO
RenderX

such as *Calm* and *Headspace* [64-66]. Furthermore, the function of these digital health platforms and the reduction in mental health symptoms are similar to those of other non–clinical interventions like yoga [67]. The wellness technology industry has been growing exponentially with app-based options that are not designed to replace clinical treatment altogether, but to offer complementary early intervention tools, or as alternative options for those that struggle to access in-person care. Relatedly, *Neurum* essentially represents a shift from disease-centered to health-centered, proactive mental health approach. EMA or EMI, combined with diverse delivery formats (namely, informational, evaluative, and practical), comprehensively supports a proactive mental health management system by addressing foundational mental health literacy, enabling timely monitoring, fostering essential skills, and promoting sustained actions. Such a shift of medical models for mental health corroborates the notion of prevention over treatment, as globally advocated by many governmental and regional health policies [68-71].

Limitations

This study has several limitations. First, despite the authors' effort to recruit participants and maintain their engagement during the local social unrest and the COVID-19 pandemic, the final sample size was still slightly smaller than expected. However, the effect size of mixed ANOVA was large, supporting the effectiveness of the intervention. Second, even though *Neurum* was designed with multiple languages, ethnicity information was not collected, which precluded the possibility of examining cultural differences, if any. Factors such as mental health beliefs, perceptions, and literacy should merit future research attention to improve the generalizability of *Neurum* in other cultures [72,73]. Third, while the attrition rate of the current study was low, local RCTs on web-based and application-based mental health programs suggested that personalized feedback might mitigate the high attrition rates typically in internet-based programs [26,74]. Indeed, several lines of evidence demonstrated the benefits of personalized intervention including better intervention outcomes, stronger subject adherence, and more persuasive feedback, in both physical and mental health domains [62,75-78]. As such, *Neurum* could be further advanced by incorporating the personalized intervention approach based on computational algorithms of the EMA or EMI data. Finally, the results of this study did not reveal a dose-response relationship between the reductions in psychological symptoms and the number of EMI sessions completed. A requirement of minimum sessions in the future can help examine the potential dose-response effect and better understand the intervention's effectiveness.

Conclusions

In sum, this RCT supported that *Neurum*, a new mHealth intervention that coupled the ecological momentary approach with supervised machine learning algorithms, could effectively reduce workplace stress and psychological distress in working adults. Beyond establishing its efficacy, the findings also highlight the potential of day-to-day, context-sensitive data to deepen the understanding of how EMA or EMI functions within digital psychological interventions. By introducing inclusive design and diverse user groups, this work further underscores the importance of equity and cultural relevance in the next generation of digital mental health care. Together, these insights highlight the promise of adaptive, data-driven mHealth tools in expanding the access to effective psychological support across different populations and in real-time and real-world settings.

Acknowledgments

The authors would like to express their sincere appreciation for the participants' devotion. No generative artificial intelligence was used in any portion of the manuscript writing.

Funding

This study was funded by the Research Grants Council of Hong Kong, SAR, to MC (reference: 17601421).

Data Availability

Anonymized data would be available upon appropriate requests towards the corresponding author.

Authors' Contributions

Conceptualization: SL, ML, MC
Data curation: CYCL, LHLL
Formal analysis: SL
Funding acquisition: MC
Investigation: ML, MC
Methodology: SL, ML, MC
Project administration: CYCL, LHLL
Resources: ML, MC
Software: ML
Supervision: MC
Validation: SL

XSL•FO
RenderX

Visualization: SL
Writing – original draft: CYCL, LHLL, SL
Writing – review & editing: SL, MC

Conflicts of Interest

None declared.

Editorial Notice

This randomized study was only retrospectively registered. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials, because the risk of bias appears low with a record of the protocol available publicly prior to the end of enrolment. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Checklist 1
CONSORT-EHEALTH checklist.
[PDF File, 123 KB - mhealth_v14i1e72435_app1.pdf]

References

1. Dewa CS, Lesage A, Goering P, Craveen M. Nature and prevalence of mental illness in the workplace. Healthc Pap 2004;5(2):12-25. [doi: 10.12927/hcpap..16820] [Medline: 15829761]
2. The Lancet Global Health. Mental health matters. Lancet Glob Health 2020 Nov;8(11):e1352. [doi: 10.1016/S2214-109X(20)30432-0] [Medline: 33069297]
3. Strawn JR, Geracioti L, Rajdev N, Clemenza K, Levine A. Pharmacotherapy for generalized anxiety disorder in adult and pediatric patients: an evidence-based treatment review. Expert Opin Pharmacother 2018 Jul;19(10):1057-1070. [doi: 10.1080/14656566.2018.1491966] [Medline: 30056792]
4. Zhang A, Franklin C, Jing S, et al. The effectiveness of four empirically supported psychotherapies for primary care depression and anxiety: a systematic review and meta-analysis. J Affect Disord 2019 Feb 15;245:1168-1186. [doi: 10.1016/j.jad.2018.12.008] [Medline: 30699860]
5. Guidi J, Fava GA. Sequential combination of pharmacotherapy and psychotherapy in major depressive disorder: a systematic review and meta-analysis. JAMA Psychiatry 2021 Mar 1;78(3):261-269. [doi: 10.1001/jamapsychiatry.2020.3650] [Medline: 33237285]
6. Schnyder N, Panczak R, Groth N, Schultze-Lutter F. Association between mental health–related stigma and active help–seeking: systematic review and meta-analysis. Br J Psychiatry 2017 Apr;210(4):261-268. [doi: 10.1192/bjp.bp.116.189464] [Medline: 28153928]
7. Mackenzie CS, Gekoski WL, Knox VJ. Age, gender, and the underutilization of mental health services: the influence of help-seeking attitudes. Aging Ment Health 2006 Nov;10(6):574-582. [doi: 10.1080/13607860600641200] [Medline: 17050086]
8. Melchior M, Caspi A, Milne BJ, Danese A, Poulton R, Moffitt TE. Work stress precipitates depression and anxiety in young, working women and men. Psychol Med 2007 Aug;37(8):1119-1129. [doi: 10.1017/S0033291707000414] [Medline: 17407618]
9. Grebner S, Semmer NK, Elfering A. Working conditions and three types of well-being: a longitudinal study with self-report and rating data. J Occup Health Psychol 2005 Jan;10(1):31-43. [doi: 10.1037/1076-8998.10.1.31] [Medline: 15656719]
10. Law R, Clow A. Stress, the cortisol awakening response and cognitive function. Int Rev Neurobiol 2020;150:187-217. [doi: 10.1016/bs.irn.2020.01.001] [Medline: 32204832]
11. Milner A, Krnjacki L, LaMontagne AD. Psychosocial job quality and mental health among young workers: a fixed-effects regression analysis using 13 waves of annual data. Scand J Work Environ Health 2017 Jan 1;43(1):50-58. [doi: 10.5271/sjweh.3608] [Medline: 27918610]
12. Ogden T, Hagen KA. Adolescent Mental Health: Prevention and Intervention, 2nd edition: Routledge; 2018. [doi: 10.4324/9781315295374]
13. Virgili M. Mindfulness-based interventions reduce psychological distress in working adults: a meta-analysis of intervention studies. Mindfulness 2015 Apr;6(2):326-337. [doi: 10.1007/s12671-013-0264-0]
14. Kieling C, Baker-Henningham H, Belfer M, et al. Child and adolescent mental health worldwide: evidence for action. Lancet 2011 Oct 22;378(9801):1515-1525. [doi: 10.1016/S0140-6736(11)60827-1] [Medline: 22008427]
15. Corbière M, Shen J, Rouleau M, Dewa CS. A systematic review of preventive interventions regarding mental health issues in organizations. Work 2009;33(1):81-116. [doi: 10.3233/WOR-2009-0846] [Medline: 19597288]
16. Knapp M, Wong G. Economics and mental health: the current scenario. World Psychiatry 2020 Feb;19(1):3-14. [doi: 10.1002/wps.20692] [Medline: 31922693]

XSL•FO
RenderX

17. Aknin LB, De Neve JE, Dunn EW, et al. Mental health during the first year of the COVID-19 pandemic: a review and recommendations for moving forward. Perspect Psychol Sci 2022 Jul;17(4):915-936. [doi: 10.1177/17456916211029964] [Medline: 35044275]
18. Fusar-Poli P, Correll CU, Arango C, Berk M, Patel V, Ioannidis JPA. Preventive psychiatry: a blueprint for improving the mental health of young people. World Psychiatry 2021 Jun;20(2):200-221. [doi: 10.1002/wps.20869] [Medline: 34002494]
19. Giorgi G, Lecca LI, Alessio F, et al. COVID-19-related mental health effects in the workplace: a narrative review. Int J Environ Res Public Health 2020 Oct 27;17(21):7857. [doi: 10.3390/ijerph17217857] [Medline: 33120930]
20. Ratheesh A, Alvarez-Jimenez M. The future of digital mental health in the post-pandemic world: evidence-based, blended, responsive and implementable. Aust N Z J Psychiatry 2022 Feb;56(2):107-109. [doi: 10.1177/00048674211070984] [Medline: 34996319]
21. Talevi D, Socci V, Carai M, et al. Mental health outcomes of the COVID-19 pandemic. Riv Psichiatr 2020;55(3):137-144. [doi: 10.1708/3382.33569] [Medline: 32489190]
22. Hossain MM, Tasnim S, Sultana A, et al. Epidemiology of mental health problems in COVID-19: a review. F1000Res 2020;9:636. [doi: 10.12688/f1000research.24457.1] [Medline: 33093946]
23. Shi W, Shen Z, Wang S, Hall BJ. Barriers to professional mental health help-seeking among Chinese adults: a systematic review. Front Psychiatry 2020;11:442. [doi: 10.3389/fpsyt.2020.00442] [Medline: 32508688]
24. Wang PS, Aguilar-Gaxiola S, Alonso J, et al. Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. Lancet 2007 Sep 8;370(9590):841-850. [doi: 10.1016/S0140-6736(07)61414-7] [Medline: 17826169]
25. Phillips MR, Zhang J, Shi Q, et al. Prevalence, treatment, and associated disability of mental disorders in four provinces in China during 2001-05: an epidemiological survey. Lancet 2009 Jun 13;373(9680):2041-2053. [doi: 10.1016/S0140-6736(09)60660-7] [Medline: 19524780]
26. Mak WW, Chio FH, Chan AT, Lui WW, Wu EK. The efficacy of internet-based mindfulness training and cognitive-behavioral training with telephone support in the enhancement of mental health among college students and young working adults: randomized controlled trial. J Med Internet Res 2017 Mar 22;19(3):e84. [doi: 10.2196/jmir.6737] [Medline: 28330831]
27. Reger MA, Gahm GA. A meta-analysis of the effects of internet- and computer-based cognitive-behavioral treatments for anxiety. J Clin Psychol 2009 Jan;65(1):53-75. [doi: 10.1002/jclp.20536] [Medline: 19051274]
28. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
29. Zhang Y, Xue J, Huang Y. A meta-analysis: Internet mindfulness-based interventions for stress management in the general population. Medicine 2020;99(28):e20493. [doi: 10.1097/MD.0000000000020493]
30. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13. [doi: 10.2196/jmir.1194] [Medline: 19403466]
31. Christensen H, Griffiths KM, Mackinnon AJ, Brittliffe K. Online randomized controlled trial of brief and full cognitive behaviour therapy for depression. Psychol Med 2006 Dec;36(12):1737-1746. [doi: 10.1017/S0033291706008695] [Medline: 16938144]
32. Adibi S, editor. Mobile Health: A Technology Road Map: Springer Cham; 2015, Vol. 5.
33. Ciemins E, Coon P, Sorli C. An analysis of data management tools for diabetes self-management: can smart phone technology keep up? J Diabetes Sci Technol 2010 Jul 1;4(4):958-960. [doi: 10.1177/193229681000400427] [Medline: 20663462]
34. Oehler RL, Smith K, Toney JF. Infectious diseases resources for the iPhone. Clin Infect Dis 2010 May 1;50(9):1268-1274. [doi: 10.1086/651602] [Medline: 20233061]
35. Kaewkungwal J, Singhasivanon P, Khamsiriwatchara A, Sawang S, Meankaew P, Wechsart A. Application of smart phone in “Better Border Healthcare Program”: a module for mother and child care. BMC Med Inform Decis Mak 2010 Nov 3;10:69. [doi: 10.1186/1472-6947-10-69] [Medline: 21047412]
36. Bakker D, Kazantzis N, Rickwood D, Rickard N. A randomized controlled trial of three smartphone apps for enhancing public mental health. Behav Res Ther 2018 Oct;109:75-83. [doi: 10.1016/j.brat.2018.08.003] [Medline: 30125790]
37. Depp CA, Mausbach B, Granholm E, et al. Mobile interventions for severe mental illness: design and preliminary data from three approaches. J Nerv Ment Dis 2010 Oct;198(10):715-721. [doi: 10.1097/NMD.0b013e3181f49ea3] [Medline: 20921861]
38. Bernhardt JM, Usdan S, Mays D, Martin R, Cremeens J, Arriola KJ. Alcohol assessment among college students using wireless mobile technology. J Stud Alcohol Drugs 2009 Sep;70(5):771-775. [doi: 10.15288/jsad.2009.70.771] [Medline: 19737502]
39. Lecomte T, Potvin S, Corbière M, et al. Mobile apps for mental health issues: meta-review of meta-analyses. JMIR mHealth uHealth 2020 May 29;8(5):e17458. [doi: 10.2196/17458] [Medline: 32348289]
40. Li H, Lewis C, Chi H, Singleton G, Williams N. Mobile health applications for mental illnesses: an Asian context. Asian J Psychiatr 2020 Dec;54:102209. [doi: 10.1016/j.ajp.2020.102209] [Medline: 32623190]
41. Myin-Germeys I, Kasanova Z, Vaessen T, et al. Experience sampling methodology in mental health research: new insights and technical developments. World Psychiatry 2018 Jun;17(2):123-132. [doi: 10.1002/wps.20513] [Medline: 29856567]

XSL•FO
RenderX

42. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
43. Smyth JM, Stone AA. Ecological momentary assessment research in behavioral medicine. J Happiness Stud 2003 Mar;4(1):35-52. [doi: 10.1023/A:1023657221954]
44. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39. [doi: 10.1348/135910709X466063] [Medline: 19646331]
45. Loo Gee B, Griffiths KM, Gulliver A. Effectiveness of mobile technologies delivering Ecological Momentary Interventions for stress and anxiety: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):221-229. [doi: 10.1093/jamia/ocv043] [Medline: 25997643]
46. Neurum. URL: https://www.neurumhealth.com [accessed 2025-07-12]
47. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
48. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]
49. Wang K, Shi HS, Geng FL, et al. Cross-cultural validation of the Depression Anxiety Stress Scale-21 in China. Psychol Assess 2016 May;28(5):e88-e100. [doi: 10.1037/pas0000207] [Medline: 26619091]
50. Colombo D, Fernández-Álvarez J, Patané A, et al. Current state and future directions of technology-based ecological momentary assessment and intervention for major depressive disorder: a systematic review. J Clin Med 2019 Apr 5;8(4):465. [doi: 10.3390/jcm8040465] [Medline: 30959828]
51. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. NPJ Digit Med 2019;2(1):88. [doi: 10.1038/s41746-019-0166-1] [Medline: 31508498]
52. Cuijpers P, Quero S, Noma H, et al. Psychotherapies for depression: a network meta-analysis covering efficacy, acceptability and long-term outcomes of all main treatment types. World Psychiatry 2021 Jun;20(2):283-293. [doi: 10.1002/wps.20860] [Medline: 34002502]
53. Alkhwaiter B, Aloud M, Almezeini N. User experience in mHealth research: bibliometric analysis of trends and developments (2007-2023). JMIR mHealth uHealth 2025 Nov 10;13(1):e75909. [doi: 10.2196/75909] [Medline: 41212171]
54. Carpenter JK, Andrews LA, Witcraft SM, Powers MB, Smits JAJ, Hofmann SG. Cognitive behavioral therapy for anxiety and related disorders: a meta-analysis of randomized placebo-controlled trials. Depress Anxiety 2018 Jun;35(6):502-514. [doi: 10.1002/da.22728] [Medline: 29451967]
55. Fernandez E, Salem D, Swift JK, Ramtahal N. Meta-analysis of dropout from cognitive behavioral therapy: magnitude, timing, and moderators. J Consult Clin Psychol 2015 Dec;83(6):1108-1122. [doi: 10.1037/ccp0000044] [Medline: 26302248]
56. Lam SU, Kirvin-Quamme A, Goldberg SB. Overall and differential attrition in mindfulness-based interventions: a meta-analysis. Mindfulness (N Y) 2022 Nov;13(11):2676-2690. [doi: 10.1007/s12671-022-01970-z] [Medline: 36506616]
57. Barrett MS, Chua WJ, Crits-Christoph P, Gibbons MB, Casiano D, Thompson D. Early withdrawal from mental health treatment: implications for psychotherapy practice. Psychotherapy (Chic) 2008 Jun 1;45(2):247-267. [doi: 10.1037/0033-3204.45.2.247] [Medline: 19838318]
58. Graham CD, McCracken LM, Harrison A, Walburn J, Weinman J. Outlining an acceptance and commitment therapy approach to treatment non-adherence. Br J Health Psychol 2022 Feb;27(1):1-12. [doi: 10.1111/bjhp.12579] [Medline: 34897907]
59. Yin H, Zhu H, Gu J, et al. Mobile-based ecological momentary assessment and intervention: bibliometric analysis. Front Psychiatry 2024;15:1300739. [doi: 10.3389/fpsyt.2024.1300739] [Medline: 38469030]
60. Yim SJ, Lui LMW, Lee Y, et al. The utility of smartphone-based, ecological momentary assessment for depressive symptoms. J Affect Disord 2020 Sep 1;274:602-609. [doi: 10.1016/j.jad.2020.05.116] [Medline: 32663993]
61. Hung YY, Huang KW, Kang HY, Huang GYL, Huang TL. Antidepressants normalize elevated Toll-like receptor profile in major depressive disorder. Psychopharmacology (Berl) 2016 May;233(9):1707-1714. [doi: 10.1007/s00213-015-4087-7] [Medline: 26415953]
62. Rabbi M, Aung MH, Zhang M, Choudhury T. MyBehavior: automatic personalized health feedback from user behaviors and preferences using smartphones. Presented at: UbiComp '15: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 7-11, 2015. [doi: 10.1145/2750858.2805840]
63. Shang J, Wei S, Jin J, Zhang P. Mental health apps in China: analysis and quality assessment. JMIR mHealth uHealth 2019 Nov 7;7(11):e13236. [doi: 10.2196/13236] [Medline: 31697245]
64. Linardon J, Messer M, Goldberg SB, Fuller-Tyszkiewicz M. The efficacy of mindfulness apps on symptoms of depression and anxiety: an updated meta-analysis of randomized controlled trials. Clin Psychol Rev 2024 Feb;107:102370. [doi: 10.1016/j.cpr.2023.102370] [Medline: 38056219]
65. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the mindfulness meditation mobile app "Calm" to reduce stress among college students: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jun 25;7(6):e14273. [doi: 10.2196/14273] [Medline: 31237569]

XSL•FO
RenderX

66. Flett JAM, Hayne H, Riordan BC, Thompson LM, Conner TS. Mobile mindfulness meditation: a randomised controlled trial of the effect of two popular apps on mental health. Mindfulness (N Y) 2019 May;10(5):863-876. [doi: 10.1007/s12671-018-1050-9]
67. de Manincor M, Bensoussan A, Smith CA, et al. Individualized yoga for reducing depression and anxiety, and improving well-being: a randomized controlled trial. Depress Anxiety 2016 Sep;33(9):816-828. [doi: 10.1002/da.22502] [Medline: 27030303]
68. State council measures to enhance people's fitness, health. The State Council of the People's Republic of China. 2019. URL: http://english.gov.cn/policies/latest_releases/2019/07/15/content_281476765851704.htm [accessed 2025-07-12]
69. HealthyLifestyle4All (2021–2023). European Commission. URL: https://sport.ec.europa.eu/healthylifestyle4all [accessed 2025-07-12]
70. Healthy people 2030. US Department of Health and Human Services. URL: https://odphp.health.gov/healthypeople [accessed 2025-07-12]
71. Healthy active learning. Ministry of Education (New Zealand). URL: https://sportnz.org.nz/healthy-active-learning/healthy-active-learning [accessed 2025-07-12]
72. Gopalkrishnan N. Cultural diversity and mental health: considerations for policy and practice. Front Public Health 2018;6:179. [doi: 10.3389/fpubh.2018.00179] [Medline: 29971226]
73. Furnham A, Swami V. Mental health literacy: a review of what it is and why it matters. Int Perspect Psychol 2018 Oct;7(4):240-257. [doi: 10.1037/ipp0000094]
74. Mak WW, Tong AC, Yip SY, et al. Efficacy and moderation of mobile app-based programs for mindfulness-based training, self-compassion training, and cognitive behavioral psychoeducation on mental health: randomized controlled noninferiority trial. JMIR Ment Health 2018 Oct 11;5(4):e60. [doi: 10.2196/mental.8597] [Medline: 30309837]
75. Dugas M, Gao GG, Agarwal R. Unpacking mHealth interventions: a systematic review of behavior change techniques used in randomized controlled trials assessing mHealth effectiveness. Digit Health 2020;6:2055207620905411. [doi: 10.1177/2055207620905411] [Medline: 32128233]
76. Alexopoulos GS, Kiosses DN, Sirey JA, et al. Untangling therapeutic ingredients of a personalized intervention for patients with depression and severe COPD. Am J Geriatr Psychiatry 2014 Nov;22(11):1316-1324. [doi: 10.1016/j.jagp.2013.05.006] [Medline: 23954038]
77. Bewick BM, Trusler K, Barkham M, Hill AJ, Cahill J, Mulhern B. The effectiveness of web-based interventions designed to decrease alcohol consumption--a systematic review. Prev Med 2008 Jul;47(1):17-26. [doi: 10.1016/j.ypmed.2008.01.005] [Medline: 18302970]
78. Martin R, Sunley P, Gardiner B, Tyler P. How regions react to recessions: resilience and the role of economic structure. Reg Stud 2016 Apr 2;50(4):561-585. [doi: 10.1080/00343404.2015.1136410]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
DASS-21: Depression, Anxiety, and Stress Scale-21 items
EMA: ecological momentary assessment
EMI: ecological momentary intervention
MD: mean difference
mHealth: mobile health
RCT: randomized controlled trial

Edited by A Stone, L Buis; submitted 10.Feb.2025; peer-reviewed by L Aasdahl, S Sahoo; revised version received 15.Dec.2025; accepted 15.Dec.2025; published 08.Sep.2026.

Please cite as:
Li S, Lo LHL, Lau CYC, Lam M, Chan M
A Smartphone-Based Ecological Momentary Intervention for Workplace Mental Health: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e72435
URL: https://mhealth.jmir.org/2026/1/e72435
doi:10.2196/72435

© Sisi Li, Lik Hang Lincoln Lo, Chiu Yi Charlie Lau, Megan Lam, Meanne Chan. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 8.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction

XSL•FO
RenderX

in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Identifying Symptom Dynamics and Profiles of Subgroups at Risk for Major Depression and Suicidal Ideation Among Korean Adults: 2-Week Ecological Momentary Assessment Study

Kyungmi Chung[1,2,3], PhD; Jin Young Park[1,2,3], MD, PhD

[1]Department of Psychiatry and Institute of Behavioral Science in Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

[2]Department of Psychiatry, Yongin Severance Hospital, Yonsei University College of Medicine, 363 Dongbaekjukjeon-daero, Giheung-gu, Yongin, Republic of Korea

[3]Center for Digital Health, Yongin Severance Hospital, Yonsei University College of Medicine, Yongin, Republic of Korea

Corresponding Author:
Jin Young Park, MD, PhD
Department of Psychiatry and Institute of Behavioral Science in Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

Abstract

Background: Traditional clinical assessments in psychiatric research or clinical practice rely on global retrospective self-report depression measures, which do not adequately capture intra- and interindividual variability in depressive symptoms over time and across contexts.

Objective: This study aimed to (1) assess the sensitivity of mobile ecological momentary assessment (EMA) for monitoring depressive symptoms compared with traditional depression scales, (2) investigate changes in depressive symptoms and recall consistency observed between the first week (FW) and the second week (SW), and (3) identify subgroups at higher risk for depression and suicidal ideation, and characterize their sociodemographic, psychological, and psychiatric profiles.

Methods: Participants' self-reports were collected once daily for 14 consecutive days via a mobile app–based EMA to monitor the presence of 20 depressive symptoms based on a 24-hour recall period, thereby capturing naturalistic severity and variability. Baseline questionnaires measured sociodemographic characteristics, digital sensitivity, and personality traits. On day 14, postquestionnaires were administered to assess their clinical symptoms. In addition, modified depression scales were administered on day 7 for 1-week recall, and day 14 for both 1-week and 2-week recalls. For cluster analysis, 3 active EMA-derived features were included: mean symptom severity, within-person symptom variability, and frequency of suicidal ideation.

Results: Generalized Linear Mixed Model analysis (model 1) revealed that traditional 2-week recall explained only 35.5% of the variance in daily symptom presence (odds ratio 0.666, 95% CI 0.659 - 0.674; $P<.001$). Additional Generalized Linear Mixed Model analysis (model 2) identified a robust interaction, indicating that the consistency between weekly retrospective recall and daily EMA differed significantly between FW and SW ($F_{1,\ 81876}=124.550$; $P<.001$). While overall symptom severity scores significantly decreased from FW to SW across both assessment methods (Cohen $d=0.10$ - 0.34; $P<.001$), the estimated mean of the probability of symptom reporting showed a contrasting upward trend from 0.853 to 0.908. In cluster analysis, 695 participants (244 males and 451 females; aged 19 - 73 years, mean 36.14, SD 10.71 years) who completed at least 7 EMA sessions over a 2-week period were classified into three distinct clusters: (1) no or low risk (n=445, 64.0%), (2) moderate risk (n=223, 32.1%), and (3) high risk (n=27, 3.9%). While depressive symptom severity, frequency of suicidal ideation, and psychiatric profiles progressively increased across clusters (cluster 1<2<3), the highest symptom variability was observed in cluster 2 (cluster 1<3<2).

Conclusions: This study demonstrated that mobile EMA effectively reduced recall bias in assessing depressive symptoms and revealed symptom dynamics. Using a minimal set of active EMA-derived features, we differentiated 3 distinct risk clusters; within-person symptom variability emerged as a clinically significant indicator not captured by traditional 2-week recall–based assessments. Despite the brief 2-week assessment period, these findings suggest that mobile EMA may serve as a robust real-world data collection tool.

(JMIR Mhealth Uhealth 2026;14:e80332) doi:10.2196/80332

KEYWORDS

depression; suicidal ideation; recall bias; mobile app; ecological momentary assessment; EMA; active monitoring; variability; cluster analysis

XSL•FO
RenderX

Introduction

Background

As one of the key symptoms of major depressive disorder (MDD), suicidal ideation is common among patients with MDD. The overall prevalence of suicidal ideation is estimated at 37.7% among patients with MDD [1]. While sustained depressed mood is not associated with lifetime suicidal ideation and suicide attempts, MDD can increase suicide risk regardless of the presence of sustained depressed mood [2]. Therefore, it is crucial to track the changes across all 9 symptoms of major depressive episode (MDE) meeting the *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition) diagnostic criteria for MDD [3] rather than focusing solely on those in mood. Furthermore, depressive symptoms are the only risk factor that can differentiate the high-risk group with suicidal ideation and suicide attempt history from the nonsuicidal group [4]. The trajectory of depression symptoms with the highest mean scores and variability over time can be identified as a predictor of suicide attempts, particularly in young adults [5]. Since suicide is conceptualized as a process or a continuum of suicidal ideation or behaviors developing from mild to severe levels of suicidality among patients with MDD [6], monitoring and treating the day-to-day variations in depression symptoms and self-harm or suicidal ideation may lower the risk of suicidal behaviors. Among patients with MDD, the decline in suicidal ideation is preceded by the alleviation of depressive symptoms [7]. As suicidal ideation in remitted patients with MDD predicts the recurrence of depression, consistently measuring suicidal ideation among individuals with MDD, as well as among those with minimal suicidal ideation at any presentation, can also prevent the worsening of suicidal symptoms and recurrence of depression [8]. In this respect, the severity and variability of depressive symptoms should be measured alongside suicidal thoughts in daily life.

Unlike many previous studies that considered affective dynamics (ie, variability, inertia, and instability) as risk factors for depression [9-12], this study focuses more on the feasibility of identifying digital phenotypes that distinguish homogeneous and distinct subgroups at risk for major depression and suicidal ideation, as well as on developing an evidence-based measurement tool to minimize recall bias and facilitate enhanced recall of current MDE symptoms in everyday life [13,14]. According to the self-schema model of depression based on the concept of organizational breakdown and subsequent restoration to effective processing of schema-compatible information [15,16], nondepressives and clinical depressives with well-organized, efficient cognitive schemas showed higher self-referent yes recall for nondepressed content and depressed content, respectively, when compared with mild depressives in a state of uncertainty and disorganization concerning one's self-schema. In line with this model, it can be postulated that mild depressives appear to contain both positive and negative personal information in their self-schema because they begin to view themselves with depressed (or negative) content although positive information has not yet been completely displaced. In this context, individuals with mild to moderate depression may show greater fluctuations in symptoms due to mixed self-schema representations, making variability a potential marker for identifying subgroups and their profiles. To explore different profiles in the dynamics of depressive symptoms, there is a need to address the limitations of standard depression scales, which rely on summated total scores based not only on classical test theory [17] but also on retrospective recall of symptom frequency over the preceding 2 weeks.

From the findings of previous studies, the issue of recall bias has been highlighted in ratings from traditional depression scales. More individuals reported severe depressive symptoms and higher suicidal ideation via ecological momentary assessment (EMA) than via retrospective self-report measures [18-20]. Moreover, more than half of the participants who reported suicidal ideation via EMA denied even past-week ideation on the scale for suicidal ideation, and the participants reporting suicidal ideation via EMA but not on the scale were just as likely to have a history of suicidal behavior as those reporting suicidal ideation in both formats [21]. In other words, EMA can capture instances of suicidal thoughts that go undetected via retrospective reports, thereby helping to identify at-risk subgroups that would otherwise be missed. Previous studies primarily demonstrated recall bias by comparing the total scores of depressive symptoms calculated from EMA data and those calculated from retrospective methods such as well-established interviewer- or self-administered measures. However, relatively little is known about how the frequency of depressive symptoms collected via EMA differs from that reported through traditional retrospective measures. To tackle these issues, we conducted a preliminary study, developing a self-administered 24-hour recall–based EMA tool to prospectively assess daily MDE symptoms over a 2-week period [13,14]. Specifically, this app uses the mobile version of the standardized Korean Version of the Center for Epidemiologic Studies Depression Scale–Revised (K-CESD-R), in which respondents are asked to report the presence of depressive symptoms within the past 24 hours by answering "yes" or "no" once daily at a self-selected notification time over a 2-week period. Taken together, using smartphone–based EMA technique (hereinafter, mobile EMA or active EMA) may allow for more accurate measurements of within- and between-participants variations in daily depressive symptoms, including self-harm and suicidal ideation, by reducing the recall period for symptom assessment.

Mobile EMA apps can enhance measurement precision and motivate repeated self-observation, thus increasing emotional reactivity and facilitating behavioral change, potentially reducing depressive symptoms [22,23]. Such symptom improvements may occur even without a formal intervention due to assessment reactivity, defined as symptom changes induced by repeated EMA assessments [24]. However, concerns remain that frequent self-monitoring might exacerbate negative affect and depressive symptoms [25,26]. Despite these concerns, previous evidence indicates that digital self-monitoring typically leads to enhanced emotional awareness, symptom reduction, and increased empowerment without triggering negative reactivity [27-29]. Although many studies have demonstrated the feasibility of using mobile EMA to capture moment-to-moment affective dynamics related to depression, relatively few studies have

XSL•FO
RenderX

prospectively monitored MDEs using EMA [11,18]. Moreover, limited research exists on whether repeated self-monitoring itself directly contributes to measurable improvements in depressive symptoms.

Objectives

The objectives of this study are (1) to assess the sensitivity of mobile EMA for monitoring depressive symptoms in daily life compared with traditional depression scales, (2) to investigate changes in depressive symptoms and recall consistency observed between the first week (FW) and the second week (SW), and (3) to identify homogeneous groups of community-dwelling adults at higher risk for depression and suicidal ideation and characterize their sociodemographic, psychological, and psychiatric profiles.

Methods

Study Design

In this study, we used a 2-week observational study design using mobile EMA and traditional retrospective scales. To address objectives 1 and 2, we administered a 24-hour recall–based mobile EMA once daily for 14 days and a modified version of the K-CESD-R in a 1-week recall format on day 7 and day 14, as well as in a 2-week recall format on day 14, reflecting the K-CESD-R reporting frame (“in the past week or so,” including an option reflecting “nearly every day for 2 weeks”). To further address objective 3, we restricted our cluster analysis to active EMA data-driven features (symptom severity, symptom variability, and frequency of suicidal ideation) that could be extracted even if participants did not respond every day during the 2-week study period, given that greater depressive symptom severity might be associated with lower adherence to digital intervention for depression [30].

Participant Recruitment

As a fully decentralized clinical trial, this study was conducted in collaboration with dataSpring Korea, Inc., over an approximately 1-month period from August 16, 2023, to September 18, 2023. A total of 1049 community-dwelling adults were recruited through PanelNow, a proprietary online research panel service operated by dataSpring Korea.

Eligibility and Dropout Criteria

In accordance with the institutional review board (IRB)–approved protocol, potential participants were assessed based on the following inclusion and exclusion criteria. This study required participants to meet all the following inclusion criteria: a person (1) who checks “Agree” on the e-consent form presented at the beginning of the online survey, (2) who is aged 19 years and older, (3) who owns a smartphone registered in their name and is able to install the mobile app used in this study, and (4) whose smartphone operating system meets the minimum requirements to install the app. However, even if all the inclusion criteria are met, people for whom (1) the mobile app is not properly installed and does not work on their smartphones or (2) the app does not run due to unknown errors that occur after completing its installation are not enrolled according to the exclusion criteria.

Participants were considered to have dropped out or were discontinued from the study if they met any of the following criteria: (1) failure to complete all mandatory prequestionnaires within the designated period; (2) never completing a single session after registering for the mobile EMA app; (3) inadequate adherence to the research protocol, defined as completing fewer than 7 out of the 14 scheduled EMA sessions or failing to complete all 3 traditional assessments; (4) providing responses judged to be insincere or unreliable; or (5) expressing a voluntary withdrawal of consent to participate in the study at any point. For participants who withdrew their consent or failed to meet the adherence criteria, all response data were immediately deleted from the server and not used for any analyses, with only basic app account logs and study participation logs retained for the legally mandated period.

Ethical Considerations

To conduct this study, the study protocol was approved by the IRB of Yongin Severance Hospital, Yonsei University College of Medicine (IRB number 9-2023-0122). All participants provided electronic informed consent (e-consent) prior to their involvement in the study. The study information and consent form were presented online via the PanelNow platform and mobile app. Potential participants were required to sign in to their registered accounts to access the study link, where they read a comprehensive description detailing the research purpose, procedures, potential benefits and risks, and compensation. Consent was obtained voluntarily when participants checked an “Agree” radio button, which allowed them to proceed to the screening and baseline questionnaires. This e-consent explicitly covered the collection and processing of personal and sensitive data, including sociodemographic characteristics, clinical symptoms, and digital phenotypes. Participants were clearly informed of their right to withdraw from the study at any time for any reason without any disadvantage or penalty. They were also provided with instructions on how to request the permanent deletion of their previously collected data by contacting the PanelNow customer center should they choose to opt out. To protect participant privacy and confidentiality, all datasets were deidentified by the online panel service provider (dataSpring Korea) before being transferred to the research team; directly identifiable information (birth date and mobile phone number) was removed, providing only the minimum necessary variables (sex and age) for statistical analysis. The deidentified data are stored on a secure, password-protected computer and server whose access is restricted to authorized individuals directly involved in the study. To ensure multilayered security, separate passwords were set for reading and writing files, and these passwords were communicated through independent channels and updated periodically. Furthermore, the mobile EMA app used a secure protocol where user identities were converted into nonidentifiable IDs; it was technically impossible to decode these IDs back into identifiable personal information. Compensation was provided via “PanelNow” points (1 point=1 KRW [1 KRW=US $0.00075 as of September 19, 2023]), a redeemable reward unit used within the service platform. Eligible participants who completed all tasks received 6000 points; those with 100% adherence to the mobile EMAs received

XSL•FO
RenderX

an additional 1000 points. The total possible compensation of 7000 points was equivalent to 7000 KRW (US $5.27).

Traditional Retrospective Measure

Modified K-CESD-R

To determine whether recall bias affects the retrospective report of depressive symptoms on the standardized CESD-R assessment [31], we modified the K-CESD-R scale [32] to additionally require participants to specify the number of days they experienced the symptoms and to input numerical values for their responses into the modified format. Specifically, a 2-week recall was conducted on day 14 to evaluate the sensitivity of mobile EMA compared with traditional retrospective recall. To investigate changes in depression symptoms and recall consistency, a 1-week recall was also administered on day 7 (for first week, FW) and day 14 (for second week, SW). Furthermore, the modified scale was designed to prevent insincere random responses. If inconsistencies were detected between the responses on the original scale and the directly entered values, the survey system displayed a warning pop-up to reduce careless entries.

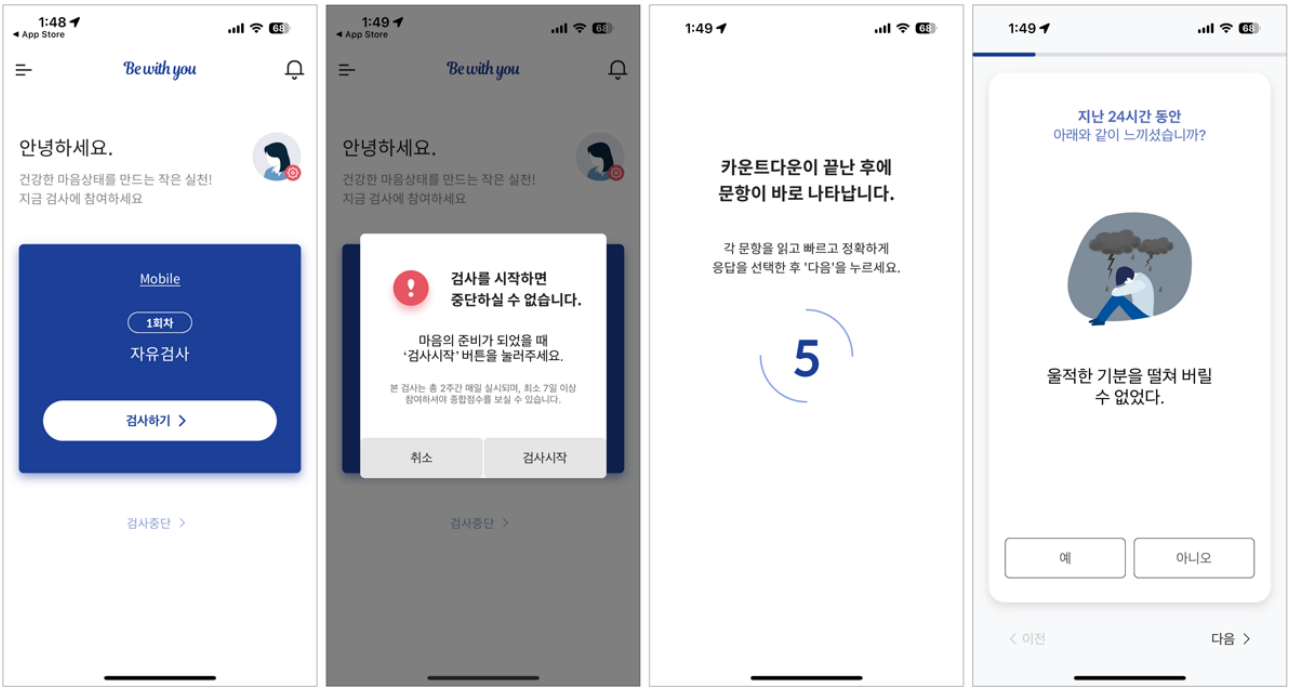
The K-CESD-R is a 20-item self-report measure of depressive symptoms in 9 different domains: sadness/dysphoria (questions 2, 4, and 6), loss of interest/anhedonia (questions 8 and 10), appetite (questions 1 and 18), sleep (questions 5, 11, and 19), thinking/concentration (questions 3 and 20), guilt/worthlessness (questions 9 and 17), tiredness/fatigue (questions 7 and 16), movement/agitation (questions 12 and 13), and suicidal ideation (questions 14 and 15). Participants are asked to rate each item on a 5-point scale to indicate how they felt or behaved during the last week or so: 0=not at all or less than 1 day in the last week, 1=1 - 2 days in the last week, 2=3 - 4 days in the last week, 3=5 - 7 days in the last week, and 4=nearly every day for 2 weeks. All responses are summed for a total score ranging from 0 to 80, with scores of 13 or higher indicating the need for individuals to seek further clinical diagnostic interview. In this study, the total scores from the modified K-CESD-R, administered separately on day 7 and day 14 to assess the preceding 1-week recall periods, ranged from 0 to 60 rather than the original 0 to 80.

EMA-Based Retrospective Measure

Mobile App–Based Depression Assessment

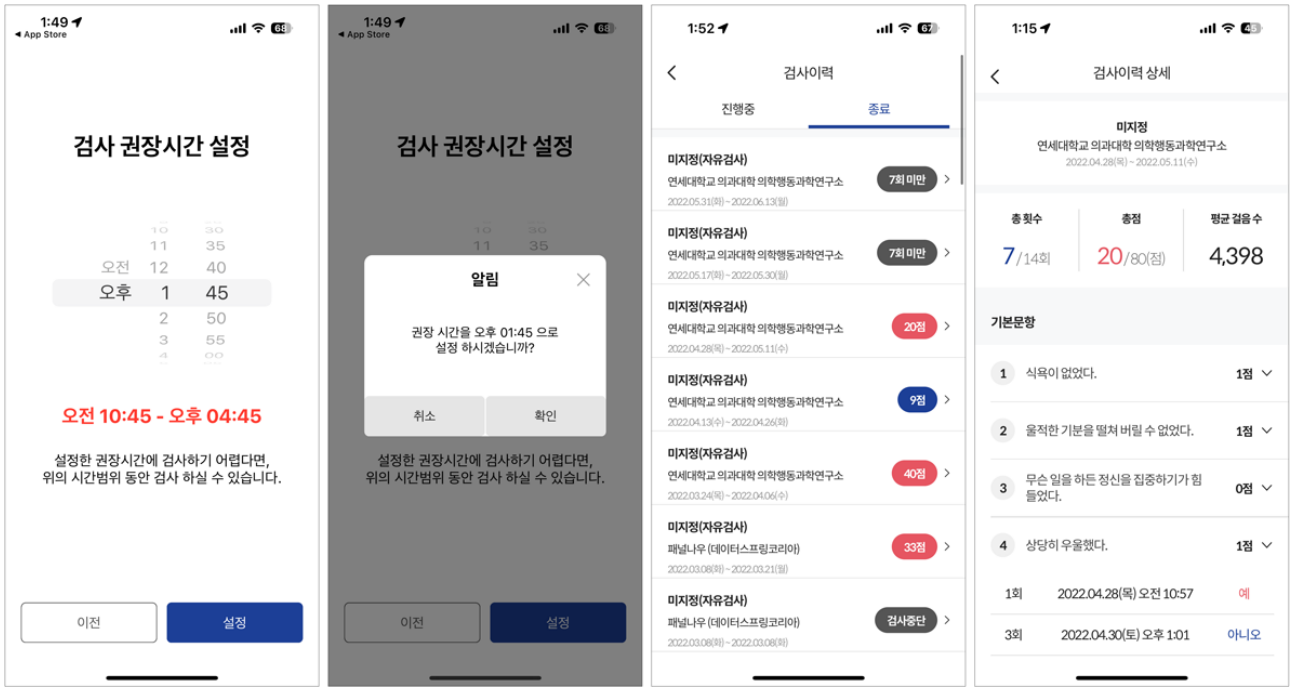
The "BeWithYou" app, an upgraded version of the K-CESD-R mobile app developed in our previous study [13], is designed as a 24-hour recall–based prospective depression assessment tool. A single test set consists of 14 sessions over a 2-week period and can be repeatedly administered to monitor the improvement and deterioration of depressive symptoms in individuals with or with no previous diagnosis of major depression. In the test set, 20 items are presented in a randomized order in each session, and participants respond either "yes (1=presence)" or "no (0=absence)" to each item (eg, "In the past 24 hours, I felt depressed") (Figure 1). At the start of the first session, the app allows participants to set a daily assessment reminder, ensuring that they can report their symptoms regularly during this study period (Figure 2). Regarding the clinically validated scoring algorithm embedded in the app [13], the total score is automatically calculated only if participants complete at least 7 sessions. If they complete fewer than 7 sessions, only the total number of tests taken over the 2 weeks and the number of days each symptom is reported can be viewed in the main menu of the assessment history (Figure 2).

Figure 1. Screenshots of the 24-hour recall-based depression assessment procedure in the BeWithYou app.



XSL•FO
RenderX

Figure 2. Screenshots of the preferred assessment time reminder feature and assessment history menu in the BeWithYou app.



Psychological and Psychiatric Measures

Digital Sensitivity

The Yongin Severance Digital Sensitivity Scale, developed by Park et al [33], is a self-report instrument designed to measure individual differences in digital sensitivity. The scale comprises 6 subfactors across 2 domains: digital literacy (14 items) and digital efficacy (6 items). The digital literacy domain includes 4 subfactors: digital application (4 items), digital communication (3 items), critical thinking (4 items), and digital ethics (3 items). The digital efficacy domain consists of digital confidence (3 items) and digital anxiety (3 items). All items are rated on a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree), yielding a total score ranging from 20 to 140. Items in the digital anxiety subscale are reverse-scored before computing the total score. Higher total scores indicate higher levels of digital sensitivity.

Personality Traits

To quickly and easily measure the Big Five personality traits, we used a Korean version of the brief Big Five Inventory-15 [34], which was adapted and shortened to 15 items from the 44-item Big Five Inventory [35]. The Korean version of the brief Big Five Inventory-15 comprises 15 items across 5-factor personality domains: (1) neuroticism (3 items), (2) extraversion (3 items), (3) openness (3 items), (4) conscientiousness (3 items), and (5) agreeableness (3 items). Each item is rated on a 5-point scale (1=strongly disagree to 5=strongly agree). The responses to the 3 items within each domain were averaged for the data analysis.

Depression

As a well-validated multipurpose instrument for screening and diagnosing major depression, the Patient Health Questionnaire–9 (PHQ-9 [36]) is adopted to measure the presence and severity of 9 MDE symptoms over the past 2 weeks. On the last day of the 2-week study period, we administered the Korean version of PHQ-9 scale [37] to collect these data, aligning them with the data measured through the modified K-CESD-R scale and the BeWithYou app. As each of the 9 items can be scored from 0 (not at all) to 3 (nearly every day), the total score of the PHQ-9 ranges from 0 to 27 (≤4: no depression, 5 - 9: mild depression, 10 - 19: moderate depression, and ≥20: severe depression), with scores of 10 or greater indicating clinically significant depression.

State-Trait Anxiety

As a measure of the severity of anxiety, the Spielberger State-Trait Anxiety Inventory (STAI-form X [38]) was adopted to separately measure 2 types of anxiety: state anxiety (ie, how you feel right now, ie, at this moment) and trait anxiety (ie, how you generally feel). In this study, we administered a revised version of the STAI [39], a 40-item self-report questionnaire with 2 subscales: the STAI-X-1 (state anxiety) and the STAI-X-2 (trait anxiety). Both subscales consist of 20 items each, rated on a 4-point scale with different anchors for the STAI-X-1 (1=not at all, 2=somewhat, 3=moderately so, and 4=very much so) and the STAI-X-2 (1=almost never, 2=sometimes, 3=often, and 4=almost always). All responses to positively stated items are reverse-coded to calculate each of the total scores for the STAI-X-1 (items 1, 2, 5, 8, 10, 11, 15, 16, 19, and 20) and the STAI-X-2 (items 1, 6, 7, 10, 13, 16, and 19); after that, the total scores of the STAI-X-1 and the STAI-X-2 are calculated by summing the responses to the 20 items for each subscale. The total scores range from 20 to 80, with higher scores indicating greater levels of anxiety.

Insomnia

As a brief self-report instrument measuring subjective insomnia complaints in the last 2 weeks, the Insomnia Severity Index (ISI [40]) consists of 7 items: (1) the severity of sleep-onset (1a: initial), sleep maintenance (1b: middle), and early morning

XSL•FO
RenderX

awakening (1c: terminal) difficulties; (2) the dissatisfaction with current sleep pattern (dissatisfaction); (3) the interference with daily functioning (interference); (4) the noticeability of impairment in the life quality attributed to the sleep problem (noticeability); and (5) the degree of concerns or distress caused by the current sleep problem (distress). Each of the ISI items is rated on a 5-point scale with different anchors from 0 to 4, and scores are summed to yield a total score ranging from 0 to 28, with high scores indicating greater insomnia severity. The total scores are interpreted as follows: no clinically significant insomnia (0‐7), subthreshold insomnia (8-14), clinical insomnia (moderate, 15‐21), and clinical insomnia (severe, 22‐28). In this study, we administered the Korean version of the Insomnia Severity Index (ISI-K), standardized and validated by Cho et al [41]. A cutoff score of 15 on the ISI-K is used as the threshold for clinically significant insomnia.

Somatic Symptoms

In this study, somatic symptoms were assessed using the Korean version of the Patient Health Questionnaire–15 [42]. This instrument consists of 15 items that ask how often respondents have been bothered by various physical symptoms over the past 4 weeks, rated on a 3-point Likert scale ranging from 0 (not at all) to 2 (bothered a lot). Total scores range from 0 to 30, with higher scores indicating more severe somatic symptom burden. A cutoff score of 8 or above was used to indicate clinically relevant somatization.

Sociodemographic Measures

Sociodemographic measures included sex, date of birth, educational attainment, marital status, number of other household members, and subjective socioeconomic status (SES), all of which were collected by self-report. Sex was assessed as a dichotomous variable (male or female). Using the date of informed consent as the reference date, age was automatically calculated from the participant's self-reported date of birth. Educational attainment was measured as the highest level of education completed, with response options of no formal education, elementary school, middle school, high school, 2‐ to 3-year college, 4-year university or higher, master's degree, doctoral degree, or other. Marital status was measured using 6 categories: single, cohabiting or common-law marriage, married, separated, divorced, or widowed. A current living arrangement variable was derived from the reported number of other household members and categorized into none (1-person household) through 5 or more people (6-or-more-person household). For analysis, this variable was recategorized as living alone (1-person household) or living with others (≥2-person household). SES was assessed with a single item asking participants to rate their overall socioeconomic position considering occupation, educational level, income, and assets on a 5-point scale ranging from very low to very high.

Procedure

In this study, potential participants were recruited via PanelNow, a mobile panel platform directly operated by dataSpring Korea. Access to the study was restricted to registered app users, who received an in-app notification with an embedded link redirecting them to the e-consent form and online prequestionnaires. These prequestionnaires consisted of screening items to determine eligibility, followed by baseline questions regarding sociodemographic information, digital sensitivity, and personality traits. After providing e-consent, individuals who met all eligibility criteria and successfully completed the BeWithYou app installation and registration process were enrolled.

During the 2-week study period, enrolled participants followed a protocol consisting of four assessment components: (1) a first traditional assessment (1-week recall), (2) a second traditional assessment (both 1-week recall and 2-week recall), (3) a mobile app–based active EMA, and (4) final postquestionnaires. To ensure a rigorous evaluation of each component, specific protocols were established as follows. For traditional retrospective assessments, modified versions of K-CESD-R were administered on day 7 and day 14. Concurrently, the mobile EMA was conducted daily for 14 sessions. To be eligible for the final data analysis, participants were required to meet minimum adherence criteria, defined as completing at least 7 mobile EMA sessions and all 3 traditional assessments. On day 14, only those who fulfilled these requirements were provided with the link to the final postquestionnaires, which assessed clinical symptoms such as anxiety, depression, insomnia, and somatization. Except for the daily mobile EMA using the BeWithYou app, all questionnaires were administered online via URL links sent to the contact channels (including app push, email, or SMS text messaging) for which participants had maintained opt-in status for research participation notification.

Statistical Analysis

All statistical analyses were performed using IBM SPSS Statistics (version 29.0.2.0 for Mac; IBM Corp), with 2-tailed *P* value of .05 considered statistically significant. Missing EMA sessions were retained as missing values so that all available repeated observations could be included in the analyses. To address objectives 1 and 2, we fitted 2 Generalized Linear Mixed Models (GLMMs), with a binomial distribution and logit link to predict the outcomes of daily EMA ("Symptom Presence"; 0=no and 1=yes). Retrospective K-CESD-R responses were first aggregated into 9 MDE symptom domains using the maximum number of days reported within each domain (eg, if "Sadness" items [3 items] were reported as 2, 0, and 5 days, the value "5" was used for that domain). In both models, "participant ID" was included as a random intercept to account for the nonindependence of repeated observations and control for stable between-person differences in reporting tendencies. Model 1 examined the association between daily EMA symptom presence (EMA sensitivity) and 2-week retrospective recall (systematic recall bias), with fixed effects of "Symptom Days_T2W" (0‐14) as a continuous covariate and "Day" (1-14) and "Domain" (1-9) as categorical factors. Model 2 further examined week-to-week differences in recall consistency using 1-week retrospective recall. Fixed effects included "Symptom Days_1W" (0‐7) as a continuous covariate and "Week" (FW vs SW), "Day" (1‐7, nested within week), and "Domain" as categorical factors. The "Week" × "Symptom Days_1W" interaction was also included to test whether recall consistency differed between weeks. Model fit was evaluated using the Akaike Information Criterion Corrected (AICC) and Bayesian

XSL•FO
RenderX

Information Criterion (BIC). Variance explained was estimated using marginal (fixed effects only) and conditional (fixed and random effects) pseudo R^2 values. The magnitude of between-person variability was assessed using the intraclass correlation coefficient (ICC). Finally, we conducted the paired samples *t* test to compare modified K-CESD-R total scores obtained on day 7 and day 14, representing the first and second 1-week recall periods, respectively. This comparison was intended to assess whether repeated EMA–based self-assessments in daily life were associated with positive changes in depression severity.

To address objective 3, K-means clustering was conducted using 3 *z*-standardized EMA-derived features: (1) mean number of daily depressive symptoms over the 2-week period (symptom severity: 0‐20 symptoms), (2) within-person SD of daily depressive symptom counts (symptom variability), and (3) mean number of days with suicidal ideation (items 14 and 15, frequency of suicidal ideation: 0‐14 days) (Table 1). We compared 3-, 4-, and 5-cluster solutions and selected the 3-cluster model based on silhouette coefficients and clinical interpretability. To validate the extracted clusters, 1-way ANOVA tests were performed using the cluster assignment as the independent variable and the means of the 3 EMA features as the dependent variables. Cluster differences in sociodemographic, psychological, and psychiatric profiles were evaluated using 1-way ANOVA, chi-square tests, and Bonferroni-adjusted post hoc comparisons ($P<.05$). Cramér *V* was used as the effect size for chi-square tests, with values of 0.10, 0.30, and 0.50 typically interpreted as indicating small, medium, and large effects, respectively [43].

XSL•FO
RenderX

Table . Algorithm used to calculate active EMA[a]-based features from participants' responses to the K-CESD-R[b] mobile scale.

Domain	2-week EMA protocol (day)														Total mean number of days with symptoms per domain over 2 weeks, day
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Sadness (dysphoria)															
Item 2	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0 - 14
Item 4	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	
Item 6	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	
Loss of interest (anhedonia)															
Item 8	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0 - 14
Item 10	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	
Appetite															
Item 1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0 - 14
Item 18	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	
Sleep															
Item 5	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0 - 14
Item 11	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	
Item 19	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	
Thinking/concentration															
Item 3	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0 - 14
Item 20	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	

XSL•FO
RenderX

Domain	2-week EMA protocol (day)														Total mean number of days with symptoms per domain over 2 weeks, day
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Guilt (worthlessness)															
Item 9	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0 - 14
Item 17	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	
Tiredness (fatigue)															
Item 7	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0 - 14
Item 16	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	
Movement (agitation)															
Item 12	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0 - 14
Item 13	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	
Suicidal ideation															
Item 14	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0 - 14
Item 15	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	
Total number of depressive symptoms per day (minimum-maximum), n	0 - 20	0 - 20	0 - 20	0 - 20	0 - 20	0 - 20	0 - 20	0 - 20	0 - 20	0 - 20	0 - 20	0 - 20	0 - 20	0 - 20	N/A[c]

XSL•FO
RenderX

Domain	2-week EMA protocol (day)														Total mean number of days with symptoms per domain over 2 weeks, day
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Transformed score per item based on the BeWithYou scoring algorithm (minimum-maximum), point	0 - 4	0 - 4	0 - 4	0 - 4	0 - 4	0 - 4	0 - 4	0 - 4	0 - 4	0 - 4	0 - 4	0 - 4	0 - 4	0 - 4	N/A
Total score (minimum-maximum), point	0 - 80														N/A

[a]EMA: ecological momentary assessment.

[b]K-CESD-R: Korean Version of the Center for Epidemiologic Studies Depression Scale–Revised.

[c]N/A: not applicable.

Results

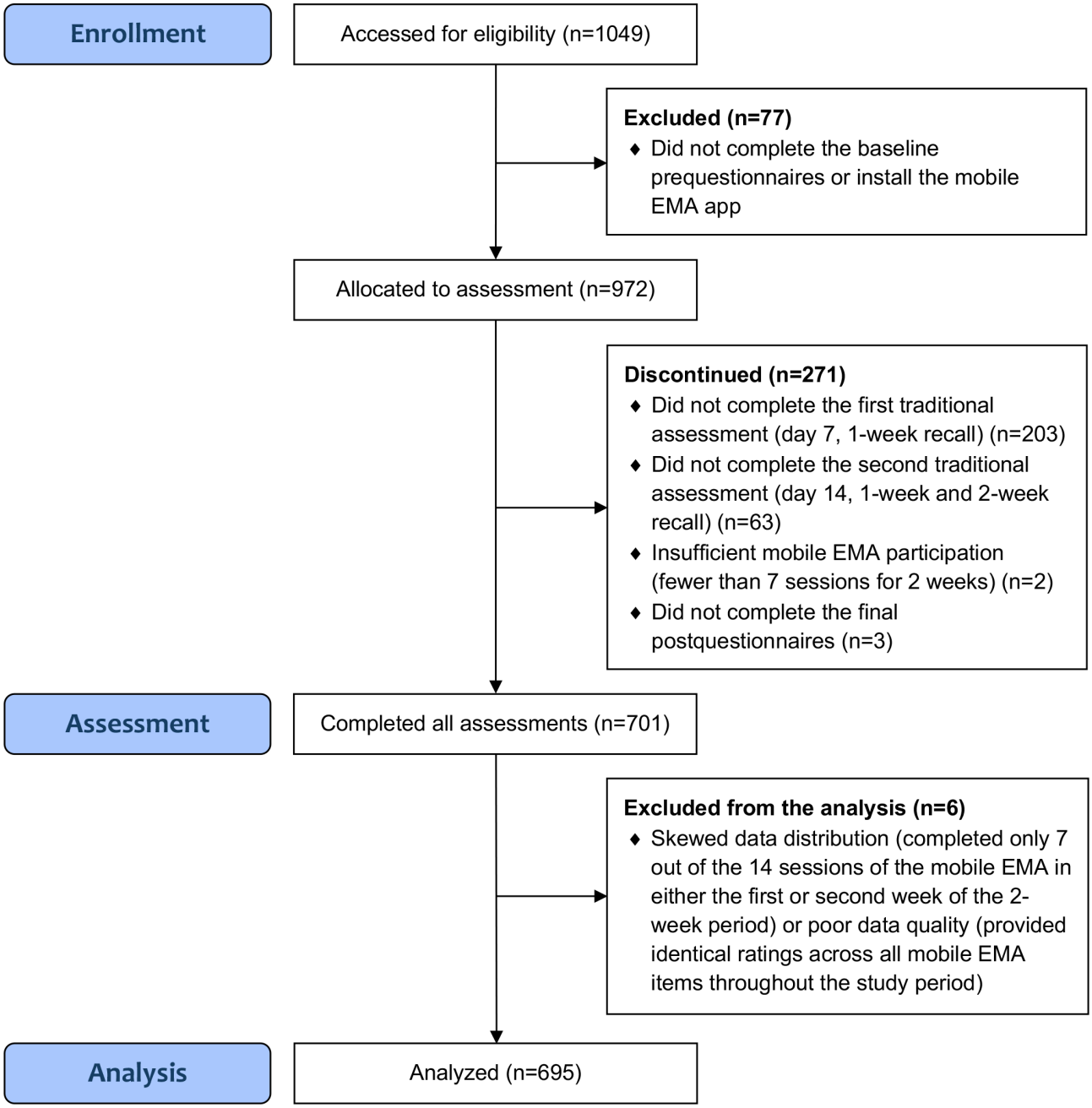
Participant Flow and Sociodemographic Characteristics

The participant flow from initial recruitment to the final analysis is illustrated in Figure 3. Out of 1049 participants who were initially screened, 701 met all eligibility and adherence criteria and were included in the data analysis, while 348 were excluded based on predefined dropout criteria. Even if they met the minimum requirement of 7 sessions, we further excluded 6 participants who exhibited highly skewed or nonrepresentative response patterns that could undermine the clinical and statistical validity of the study.

Specifically, these exclusions included (1) participants whose responses were concentrated exclusively within a single week (either the FW or the SW), which not only precluded the assessment of the temporal continuity of depressive symptoms over the required 2-week period—a critical clinical criterion—but also rendered them ineligible for the planned comparison between the FW and the SW; and (2) participants who exhibited straight-lining EMA response patterns with zero variance, suggesting insincere participation. After these final exclusions, the sociodemographic and psychological characteristics of the final analytic sample (N=695) are summarized in Table 2.

XSL•FO
RenderX

Figure 3. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of this study. EMA: ecological momentary assessment.



XSL•FO
RenderX

Table . Sociodemographic and psychological characteristics of the total sample.

Characteristics	Total sample (N=695)
Sex, n (%)	
Male	244 (35.1)
Female	451 (64.9)
Age (years), mean (SD)	36.14 (10.71)
Age (years), range (minimum-maximum)	19 - 73
Age group (years), n (%)	
19 - 29	224 (32.2)
30 - 39	215 (30.9)
40 - 49	165 (23.7)
50 - 59	75 (10.8)
60 - 69	16 (2.3)
Educational attainment, n (%)	
High school	109 (15.7)
Associate degree (2 - 3-year college)	118 (17.0)
Bachelor's degree (4-year university or higher)	409 (58.8)
Master's degree	50 (7.2)
Doctoral degree	9 (1.3)
Marital status, n (%)	
Single	354 (50.9)
Cohabiting/common-law marriage	16 (2.3)
Married	316 (45.5)
Previously married (separated/divorced/widowed)[a]	9 (1.2)
Current living arrangement, n (%)	
Living alone (1-person household)	109 (15.7)
Living with others (2 or more household members)	586 (84.3)
Subjective socioeconomic status, n (%)	
Lowest	21 (3.0)
Low	177 (25.5)
Middle	450 (64.7)
High	45 (6.5)
Highest	2 (0.3)
Personality trait, mean (SD)	
Openness to experience	3.16 (0.86)
Conscientiousness	3.81 (0.67)
Neuroticism	2.94 (0.92)
Extraversion	2.72 (0.85)
Agreeableness	3.46 (0.64)
Digital sensitivity, mean (SD)	90.12 (14.03)
Smartphone operating system, n (%)	
iOS	244 (35.1)
Android	451 (64.9)

XSL•FO
RenderX

[a]Previously married was defined as having been legally married at some point, including those currently separated, divorced, or widowed.

Final Analytic Sample and Mobile EMA Outcome Events

Of the total 87,570 potential sessions, 81,900 (93.5%) valid EMA observations from 695 participants were included in the final analysis. This high inclusion rate, achieved by using all available data points through a GLMM, effectively mitigated potential selection bias and preserved statistical power despite occasional missing daily entries.

Assessing Recall Bias in Traditional vs Mobile EMA–Based Depression Measures

The primary objective of model 1 was to quantify the systematic discrepancy and information loss inherent in traditional retrospective reports (2-week recall) compared with daily EMA records (24-hour recall). The final GLMM demonstrated robust statistical fit (AICC=473,906.603 and BIC=473,915.916). The overall classification accuracy for model 1 was 88.3%. Omnibus tests for fixed effects revealed that traditional retrospective reports ("Symptom Days_T2W") were a statistically significant predictor of daily symptom presence ("Symptom Presence"; $F_{1, 81877}$=4,826.322; P<.001). Significant main effects were also observed for "Domain" ($F_{8, 81877}$=284.483; P<.001) and "Day" ($F_{13, 81877}$=118.344; P<.001). Parameter estimates for "Symptom Days_T2W" yielded a significant negative slope (B=−0.406, SE 0.006; P<.001), with an odds ratio of 0.666 (95% CI 0.659 - 0.674). The marginal pseudo R^2 value was 0.355, indicating that the traditional retrospective recall method explains only 35.5% of the variance in actual daily symptom presence. When accounting for random intercepts, the conditional pseudo R^2 value increased to 0.504, with an adjusted ICC of 0.232. These results show that while interindividual heterogeneity accounts for 23.2% of the total variance, nearly 76.8% of the symptomatic variability is driven by intraindividual fluctuations, approximately half of which (49.6%) remains entirely unexplained by the 2-week recall method.

Investigating Temporal Dynamics and Interaction Effects of Recall Periods

The second objective of model 2 was to investigate whether the consistency of recall remained stable or exhibited temporal fluctuations across the assessment period by comparing the FW and the SW. The final GLMM also demonstrated a robust statistical fit (AICC=469,917.005 and BIC=469,926.318) and achieved an overall classification accuracy of 88.5%. Omnibus tests for fixed effects revealed significant main effects for all predictors, including "Domain" ($F_{8, 81876}$=276.180; P<.001), "Day" ($F_{12, 81876}$=56.555; P<.001), and "Symptom Days_1W" ($F_{1, 81876}$=7146.780; P<.001). Notably, a significant main effect of "Week" was also observed ($F_{1, 81876}$=50.317; P<.001), with the estimated mean of "Symptom Presence" increasing from 0.853 (FW) to 0.908 (SW). Similar to model 1, parameter estimates for "Symptom Days_1W" yielded a significant negative slope (B=−0.749, SE 0.011, P<.001; odds ratio 0.473, 95% CI 0.463 - 0.483). The marginal pseudo R^2 value of 0.335 indicates that the fixed effects, including the temporal interaction, explain only about one-third of the variance in daily symptom reporting. When accounting for random intercepts, the conditional pseudo R^2 value increases to 0.475, indicating that more than 52% of the total variance in daily symptom presence remains entirely unexplained even when considering individual differences. The adjusted ICC of 0.211 further revealed that while interindividual heterogeneity accounts for 21.1% of the total variance, the vast majority of symptomatic variability, specifically nearly 79%, is attributable to within-person fluctuations. Most importantly, a robust interaction was found between "Week" and "Symptom Days_1W" ($F_{1, 81876}$=124.550; P<.001). This significant interaction revealed that the consistency between weekly retrospective recall and daily EMA significantly differed between FW and SW.

Evaluating Week-to-Week Changes in Depression Severity Across Recall Methods

To further elucidate the symptomatic trajectory observed in model 2, a supplementary analysis was conducted to compare the change in depression severity scores between FW and SW across 2 different assessment methods: a 1-week recall–based version (original) and a 24-hour recall–based version (mobile). Paired samples t tests revealed that the original scores significantly decreased from 12.46 (SD 13.54) in the FW to 11.17 (SD 13.76) in the SW (t_{694}=5.41, Cohen d=0.10; P<.001). Additionally, a comparable reduction was observed in the mobile scores, which decreased from 14.29 (SD 13.71) to 11.11 (SD 14.03) (t_{694}=13.04, Cohen d=0.34; P<.001). Notably, this overall reduction in symptomatic severity across both assessment methods stands in direct contrast to the upward trend in symptom reporting probability observed in model 2.

Identifying and Profiling High-Risk Subgroups Using Active EMA Features

A K-means cluster analysis (k=3) was performed using active EMA features—symptom severity, symptom variability, and frequency of suicidal ideation—which were standardized using a z-score transformation. In the analysis, a total of 695 participants were identified as 3 distinct risk groups, and based on their symptom profiles, each cluster was subsequently labeled as no- or low-risk (cluster 1), moderate-risk (cluster 2), and high-risk (cluster 3) groups. Cluster 1 (n=445) was characterized by low symptom severity (z=−0.57; raw mean 1.07, SD 1.15), low symptom variability (z=−0.56; raw mean 0.95, SD 0.65), and minimal suicidal ideation (z=−0.30; raw mean 0.02, SD 0.12). Cluster 2 (n=223) had moderate symptom severity (z=0.79; raw mean 6.42, SD 3.20), the highest symptom variability (z=1.04; raw mean 3.07, SD 1.16), and occasional suicidal ideation (z=0.07; raw mean 0.62, SD 0.98). Cluster 3 (n=27) showed the highest symptom severity (z=2.80; raw mean 14.34, SD 2.13), moderate symptom variability (z=0.36; raw mean 2.17, SD 1.55), and frequent suicidal ideation (z=4.23; raw mean 7.37, SD 2.82). While both the severity of depressive symptoms and frequency of suicidal ideation, as well as all psychiatric characteristics, increased with the level of risk in

XSL•FO
RenderX

the clusters (cluster 1<2<3), the highest variability of depressive symptoms was observed in cluster 2 (cluster 1<3<2).

To assess the robustness of the clustering outcomes, 1-way ANOVAs were conducted using the cluster assignments as the independent variable. Results confirmed significant between-group differences across all 3 EMA indices: $F_{2, 692}$=892.24, partial η^2=0.72, P<.001, for symptom severity; $F_{2, 692}$=423.18, partial η^2=0.55, P<.001, for symptom variability; and $F_{2, 692}$=1117.36, partial η^2=0.76, P<.001, for frequency of suicidal ideation. These findings support the internal validity of the 3-cluster classification.

Sociodemographic and Psychological Characteristics Across Clusters

As presented in Table 3, the sociodemographic and psychological characteristics are summarized for each of the 3 clusters identified in the cluster analysis: (1) cluster 1 (no or low risk: 445/695, 64.0%), (2) cluster 2 (moderate risk: 223/695, 32.1%), and (3) cluster 3 (high risk: 27/695, 3.9%).

XSL•FO
RenderX

Table . Sociodemographic and psychological characteristics of 3 clusters[a].

Characteristics	Cluster 1 (n=445)	Cluster 2 (n=223)	Cluster 3 (n=27)	*P* value
Sex, n (%)				.20
Male	167 (37.5)	68 (30.5)	9 (33.3)	—[b]
Female	278 (62.5)	155 (69.5)	18 (66.7)	—
Age (years), mean (SD)	36.64 (10.89)	35.59 (10.42)	32.37 (9.25)	.09
Age (years), range (minimum–maximum)	19 - 73	19 - 69	19 - 52	—
Age group (years), n (%)				.39
19 - 29	141 (31.7)	72 (32.3)	11 (40.7)	—
30 - 39	127 (28.5)	78 (35.0)	10 (37.0)	—
40 - 49	114 (25.6)	48 (21.5)	3 (11.1)	—
50 - 59	50 (11.2)	22 (9.9)	3 (11.1)	—
60 - 69	13 (2.9)	3 (1.3)	0 (0.0)	—
Educational attainment, n (%)				.51
High school	65 (13.9)	40 (17.9)	7 (25.9)	—
Associate degree (2 - 3-year college)	71 (16.0)	42 (18.8)	5 (18.5)	—
Bachelor's degree (4-year university or higher)	269 (60.4)	126 (58.5)	14 (51.9)	—
Master's degree	36 (8.1)	13 (5.8)	1 (3.7)	—
Doctoral degree	7 (1.6)	2 (0.9)	0 (0.0)	—
Marital status, n (%)				.11
Single	209 (47.0)	128 (57.4)	17 (63.0)	—
Cohabiting/common-law marriage	12 (2.7)	3 (1.3)	1 (3.7)	—
Married	219 (49.2)	88 (39.5)	9 (33.3)	—
Previously married (separated/divorced/widowed)	5 (1.1)	4 (1.8)	0 (0.0)	—
Current living arrangement, n (%)				.30
Living alone (1-person household)	66 (14.8)	36 (16.1)	7 (25.9)	—
Living with others (2 or more household members)	379 (85.2)	187 (83.9)	20 (74.1)	—
Subjective socioeconomic status, n (%)				.03
Lowest	10 (2.2)[c]	9 (4.0)[c]	2 (7.4)[c]	—
Low	98 (22.0)[c]	67 (30.0)[c,d]	12 (44.4)[d]	—
Middle	302 (67.9)[c]	135 (60.5)[c]	13 (48.1)[c]	—
High	33 (7.4)[c]	12 (5.4)[c]	0 (0.0)[c]	—
Highest	2 (0.4)[c]	0 (0.0)[c]	0 (0.0)[c]	—
Personality trait, mean (SD)				<.001
Openness to experience	3.16 (0.84)[c]	3.21 (0.88)[c]	2.90 (0.95)[c]	.20
Conscientiousness	3.90 (0.62)[c]	3.64 (0.68)[d]	3.39 (0.93)[e]	<.001

XSL•FO
RenderX

Characteristics	Cluster 1 (n=445)	Cluster 2 (n=223)	Cluster 3 (n=27)	*P* value
Neuroticism	2.61 (0.79)[c]	3.47 (0.84)[d]	3.95 (0.80)[e]	<.001
Extraversion	2.77 (0.82)[c]	2.67 (0.89)[c]	2.25 (0.78)[d]	.004
Agreeableness	3.49 (0.63)[c]	3.44 (0.65)[c, d]	3.17 (0.71)[d]	.04
Digital sensitivity, mean (SD)	90.13 (14.37)	90.49 (13.46)	86.74 (12.97)	.42
Smartphone operating system, n (%)				.31
iOS	151 (33.9)	80 (35.9)	13 (48.1)	—
Android	294 (66.1)	143 (64.1)	14 (51.9)	—

[a]Cluster 1, no or low risk; cluster 2, moderate risk; and cluster 3, high risk.

[b]Not available.

[c]Scores with the same superscript are not significantly different from each other but are significantly different from those with different superscripts.

[d]Scores with the same superscript are not significantly different from each other but are significantly different from those with different superscripts.

[e]Scores with the same superscript are not significantly different from each other but are significantly different from those with different superscripts.

According to the chi-square analyses for categorical variables, there were no significant differences among the 3 clusters in sex (χ^2_2=3.27, Cramér *V*=0.07; *P*=.20), age group (χ^2_8=8.42, Cramér *V*=0.08; *P*=.39), educational attainment (χ^2_8=7.21, Cramér *V*=0.07; *P*=.51), marital status (χ^2_6=10.29, Cramér *V*=0.09; *P*=.11), current living arrangement (χ^2_2=2.42, Cramér *V*=0.06; *P*=.30), and smartphone operating system (χ^2_2=2.34, Cramér *V*=0.06; *P*=.31). For continuous variables, separate 1-way ANOVAs revealed no significant group differences across the clusters in age ($F_{2, 692}$=2.46, partial η^2=0.007; *P*=.09) and digital sensitivity ($F_{2, 692}$=0.86, partial η^2=0.002; *P*=.42).

Among all variables examined, only SES and personality traits showed significant differences across the 3 clusters. For subjective SES, a significant group difference was observed (χ^2_8=17.41, Cramér *V*=0.11; *P*=.03). To compare column proportions, pairwise *z*-tests were conducted, and *P* values were adjusted using the Bonferroni method (*P*<.05). These comparisons revealed that the proportion of participants reporting a low SES level was significantly higher in cluster 3 (12/695, 44.4%) than in cluster 1 (98/695, 22.0%). In terms of personality traits, a 1-way MANOVA revealed a significant multivariate effect of group on the combined dependent variables (Pillai trace=0.26, $F_{10, 1378}$=18.85, partial η^2=0.13; *P*<.001). Among the Big Five personality dimensions, 1-way ANOVAs revealed significant group differences in conscientiousness ($F_{2, 692}$=17.09, partial η^2=0.05; *P*<.001), neuroticism ($F_{2, 692}$=105.37, partial η^2=0.23; *P*<.001), extraversion ($F_{2, 692}$=5.61, partial η^2=0.02; *P*=.004), and agreeableness ($F_{2, 692}$=3.37, partial η^2=0.01; *P*=.04), whereas no significant difference was found in openness to experience only ($F_{2, 692}$=1.63, partial η^2=0.005; *P*=.20).

As shown in Table 3, post hoc comparisons with Bonferroni correction indicated that both conscientiousness and neuroticism significantly differed across all 3 clusters (*P*<.001). Specifically, individuals in higher-risk clusters demonstrated lower levels of conscientiousness and higher levels of neuroticism than those in lower-risk clusters. Extraversion was significantly lower in cluster 3 than in both cluster 1 (*P*<.001) and cluster 2 (*P*=.04). For agreeableness, individuals in cluster 1 exhibited significantly higher scores than those in cluster 3 (*P*=.04), whereas no significant differences were found between cluster 2 and either of the other clusters.

Psychiatric Characteristics Across Clusters

To investigate differences in psychiatric characteristics across the 3 clusters, 1-way ANOVAs were conducted for each variable. As shown in Table 4, all psychiatric variables demonstrated significant group differences. Significant effects of cluster membership were found for the severity of depression ($F_{2, 692}$=463.22, partial η^2=0.57; *P*<.001), state anxiety ($F_{2, 692}$=232.61, partial η^2=0.40; *P*<.001), trait anxiety ($F_{2, 692}$=273.99, partial η^2=0.44; *P*<.001), insomnia ($F_{2, 692}$=137.15, partial η^2=0.28; *P*<.001), and somatization ($F_{2, 692}$=197.63, partial η^2=0.36; *P*<.001). Effect size interpretation based on partial η^2 indicated large effects for all variables (≥0.14). Post hoc analyses with Bonferroni correction indicated that all pairwise comparisons were statistically significant (*P*<.001) for depression, state and trait anxiety, and somatization. For insomnia, the difference was significant between cluster 1 and cluster 2 (*P*<.001) and between cluster 1 and cluster 3 (*P*<.001) but not between cluster 2 and cluster 3 (*P*=.07), indicating a similar level of insomnia severity between the moderate- and high-risk groups.

XSL•FO
RenderX

Table . Psychiatric characteristics of the study sample[a].

Characteristics	Total (N=695)	Cluster 1 (n=445)	Cluster 2 (n=223)	Cluster 3 (n=27)	*P* value
Depression (PHQ-9[b]), mean (SD)	4.93 (5.71)	1.92 (2.53)[c]	9.34 (5.20)[d]	18.00 (5.66)[e]	<.001
State anxiety (STAI-X-1[f]), mean (SD)	44.59 (12.64)	38.85 (9.48)[c]	53.43 (10.26)[d]	66.07 (10.83)[e]	<.001
Trait anxiety (STAI-X-2), mean (SD)	45.09 (12.15)	39.26 (8.49)[c]	54.21 (10.04)[d]	65.85 (10.48)[e]	<.001
Insomnia (ISI-K[g]), mean (SD)	9.51 (6.28)	7.02 (4.95)[c]	13.67 (5.93)[d]	16.11 (5.98)[d]	<.001
Somatization (PHQ-15[h]), mean (SD)	7.34 (5.30)	5.02 (3.78)[c]	10.96 (4.87)[d]	15.44 (5.54)[e]	<.001

[a] Cluster 1, no or low risk; cluster 2, moderate risk; and cluster 3, high risk.

[b]PHQ-9: Patient Health Questionnaire–9.

[c]Scores with the same superscript are not significantly different from each other but are significantly different from those with different superscripts.

[d]Scores with the same superscript are not significantly different from each other but are significantly different from those with different superscripts.

[e]Scores with the same superscript are not significantly different from each other but are significantly different from those with different superscripts.

[f]STAI: State-Trait Anxiety Inventory.

[g]ISI-K: Korean version of the Insomnia Severity Index.

[h]PHQ-15: Patient Health Questionnaire–15.

Discussion

Principal Findings

The main findings of this study are as follows. First, we confirmed the superior sensitivity of mobile EMA for monitoring daily depressive symptoms compared with traditional retrospective scales. Second, our analysis revealed distinct temporal changes in depressive symptoms and patterns of recall consistency between the FW and SW of observation. Third, we identified specific subgroups at heightened risk for depression and suicidal ideation and characterized their unique sociodemographic, psychological, and psychiatric profiles. These results provide a multidimensional understanding of symptom dynamics and underscore the clinical usefulness of high-frequency digital monitoring for effective risk stratification among community-dwelling Korean adults.

This study provides empirical evidence that traditional retrospective reports systematically underestimate symptomatic experiences compared with daily EMA, which captures the simple daily recording of symptom presence. We found that a substantial portion of daily-level symptomatic variance is lost as participants are required to mentally reconstruct their experiences and estimate the total number of symptomatic days over successive recall periods. This process of heuristic estimation introduces a “smoothing effect” that obscures the true density and temporal complexity of depressive symptoms, resulting in significant information loss. Consequently, our findings suggest that retrospective frequency estimates insufficiently capture the full burden of depression, as the nuanced symptomatic fluctuations captured by prospective monitoring are flattened during memory-based aggregation.

What is more, we identified a paradoxical divergence in symptomatic patterns from the FW to the SW of assessment. While retrospective assessments indicated a significant decrease in depression severity scores, the prospective probability of actual symptom reporting concurrently increased. If this decrease in scores were due to regression to the mean, symptom reporting should have also decreased. Instead, the observed increase supports a psychological shift in symptom sensitization rather than a mere statistical artifact. This discrepancy is closely linked to intraindividual dynamics, which account for the vast majority of symptomatic variability and remain largely unexplained by traditional retrospective recall methods. Ultimately, the interaction between assessment week and recall consistency provides robust evidence that recall bias is not a fixed or static error; rather, it is a dynamic process associated with changes in a participant’s internal standards and cognitive framework for symptom recall during sustained daily monitoring.

Last but not least, this study highlights the value of digital phenotyping by identifying 3 distinct clinical clusters with diverging symptomatic patterns. While both mean symptom severity and suicidal ideation frequency increased linearly across the 3 risk groups, within-person symptom variability (SD) exhibited a nonlinear pattern, particularly peaking in the moderate-risk group. This peak in symptom variability may reflect the underlying complexity of ambivalent self-schemas in moderate-risk individuals. Furthermore, these clusters were validated by psychiatric and personality profiles, with the high-risk group distinguished by the highest neuroticism, the lowest conscientiousness, and a higher prevalence of low subjective SES. Collectively, these findings underscore within-person symptom dynamics as a critical feature in characterizing heterogeneous depression phenotypes and their associated suicide risk, both of which are often masked by the traditional retrospective assessments.

XSL•FO
RenderX

Comparison With Prior Work

In line with previous findings that EMA has greater sensitivity in detecting depressive symptoms and suicidal ideation than retrospective self-report scales [18-20], this study suggests that traditional retrospective reports are not merely imprecise but systematically distorted. Indeed, the substantial information loss inherent in traditional 2-week recall was evidenced by its limited explanatory power, accounting for only 35.5% of the variance in actual daily symptom presence (model 1). The significant negative slope observed (B=−0.406; P<.001) further indicates a fundamental misalignment between memory-based reports and actual daily occurrences. Whereas the traditional K-CESD-R scores were significantly higher for the FW than for the SW, the actual probability of daily symptom presence was found to be significantly higher in the SW than in the FW (0.908 vs 0.853; model 2). This discrepancy is likely rooted in cognitive heuristics; specifically, this inverse relationship suggests that retrospective recall is not a simple summary of experiences but is heavily distorted by specific intense episodes from the past, reflecting the “peak bias” of the peak-end rule [44]. In this case, participants appeared to rely on intense MDEs from the more distant past (FW) to calibrate their overall 2-week symptom reports, thereby overriding the actual increase in symptom frequency observed in the more recent period (SW). Given that nearly 79% of symptomatic variability is linked to intraindividual fluctuations, mobile EMA serves as a cognitively robust alternative that captures these dynamic patterns which are otherwise lost to recall bias.

Beyond identifying these recall biases, the observed paradoxical divergence between the decrease in retrospective scores and the increase in the daily probability of symptom reporting highlights shifting patterns in how participants perceive and evaluate their affective states over the course of the study. Engagement with EMA is associated with a refinement of internal standards for symptom appraisal, which may be related to heightened emotional awareness and behavioral reflection [22,23]. These results align with the concept of assessment reactivity [24], indicating that the act of frequent monitoring can be reflected in reporting patterns by sensitizing participants to their subtle symptomatic fluctuations. Therefore, mobile EMA provides a more ecologically valid reflection of symptomatic trajectories than traditional static measures, capturing the daily engagement with their own symptoms. While such heightened awareness may facilitate self-regulated recognition, it remains critical to acknowledge that increased self-focus or reinforced negative symptom monitoring may exacerbate psychological distress in certain individuals [25,26]. Overall, these findings support the potential usefulness of mobile EMA as a fine-grained tool for capturing the complexity of symptom trajectories.

To address inconsistent findings on daily negative affect dynamics across levels of depression severity [9,12,45], we classified community-dwelling participants into 3 distinct subgroups with depressive symptom dynamics: (1) cluster 1 (no or low risk; 445/695, 64.0%): low symptom severity, low symptom variability, and minimal suicidal ideation; (2) cluster 2 (moderate risk; 223/695, 32.1%): moderate symptom severity, the highest symptom variability, and occasional suicidal ideation; and (3) cluster 3 (high risk; 27/695, 3.9%): the highest symptom severity, moderate symptom variability, and frequent suicidal ideation. As hypothesized, within-person symptom variability, among the 3 active EMA features, played a key role in distinguishing not only between cluster 2 and cluster 3 but also between cluster 1 and clusters 2 and 3, as validated by the significant differences in the PHQ-9 scores across the 3 clusters. In line with the self-schema model for depression [15,16], these findings suggest that the highest symptom variability observed in cluster 2 may be attributed to the underlying instability of ambivalent self-schema representations, characterized by the coexistence of both positive and negative self-referent information. Importantly, the finding that higher variability was not always associated with greater symptom severity further suggests the presence of heterogeneous depression profiles.

Although cluster 3 (27/695, 3.9%) represented a relatively small proportion of the total sample, this distribution aligns with the expected epidemiological prevalence of high-risk individuals with suicidal ideation in the South Korean general population. Recent national surveys report that the proportion of adults who experienced suicidal thoughts within the past year ranges from 2.1% (2023 National Survey on Suicide [46]) to 4.6% (2023 Korea National Health and Nutrition Examination Survey [47]). The high degree of consistency between our findings (27/695, 3.9%) and these national representative data strongly supports the ecological validity of this cluster analysis, suggesting that cluster 3 is not a statistical outlier but a clinically valid subgroup of the at-risk population in a real-world context. Moreover, while national statistics often estimate suicide risk based on a 1-year history of suicidal ideation without considering depression severity, this study evaluated the co-occurrence of both active suicidal thoughts and depressive symptoms within a brief 2-week window. Furthermore, consistent with the previous study [21], mobile EMA may reduce the concealment of suicidal thoughts typically encountered in retrospective assessments and identify highly specific, high-risk phenotypes, providing a more sensitive clinical snapshot of acute suicidal risk associated with depression.

As well as the PHQ-9 scores, the STAI-X-1, STAI-X-2, and Patient Health Questionnaire–15 scores significantly differed across the clusters, showing a consistent increase with higher-risk levels (cluster 1<2<3). Consistent with the previous findings that individuals with depressive symptoms are more vulnerable to insomnia than those with no depressive symptoms in a general population–based sample from the Republic of Korea [48], this study also observed an increase in the ISI-K scores in accordance with the cluster risk level. Furthermore, the lack of a significant difference in the ISI-K scores between clusters 2 and 3 aligns with prior evidence suggesting that the severity of depressive symptoms is not significantly influenced by the presence or severity of insomnia symptoms among those with depressive symptoms [48]. In this regard, the risk levels assigned to each cluster closely corresponded to their psychiatric characteristics, providing theoretical support for the association between symptom variability and the instability of depressive self-schema.

In addition to psychiatric profiles, the cluster classification was also associated with sociodemographic factors and personality traits known as risk factors for both depressive symptoms and

XSL•FO
RenderX

suicidality. Consistent with the previous findings that low subjective SES predicted higher levels of depressive symptoms and suicidal ideation [49,50], cluster 3 (12/27, 44.4%) accounted for a significantly higher proportion of individuals with low subjective SES than cluster 1 (98/445, 22.0%), showing that their proportion increased with the cluster risk level. Interestingly, even in cluster 3 (the high-risk group), only a small proportion of participants reported the lowest level of subjective SES, while those with low (but not the lowest) SES were overrepresented. This pattern supports prior findings that relative deprivation, rather than absolute poverty, is more closely associated with depression and suicidal risk [51,52]. These findings also align with the social rank theory of depression [53], which posits that lower perceived social status increases vulnerability to depression and suicidality.

Among the Big Five personality traits, neuroticism, conscientiousness, extraversion, and agreeableness significantly differed across the 3 clusters. Individuals in the high-risk cluster exhibited the highest neuroticism and the lowest levels of conscientiousness, extraversion, and agreeableness. Consistent with prior evidence [54-56], neuroticism and conscientiousness were the personality traits that most clearly distinguished the 3 clusters. Particularly, cluster 3 exhibited the highest levels of neuroticism and the lowest levels of conscientiousness. These findings suggest that those in cluster 3 are more prone to emotional instability including heightened sensitivity to stress and frequent negative emotions, such as anxiety and anger, and experience greater difficulties with self-regulation in impulse control, organization, and goal-directed behavior. Although extraversion and agreeableness also differed across some of the clusters, both traits were not as clearly discriminative of depression risk as neuroticism and conscientiousness [57,58]. Extraversion was significantly lower in cluster 3 than in clusters 1 and 2, whereas agreeableness was significantly lower than in cluster 1 but did not differ from cluster 2. The observed reductions in extraversion and agreeableness among individuals in cluster 3 may reflect increased social withdrawal and interpersonal difficulties, which are commonly associated with depressive symptomatology [3]. Taken together, the current findings highlight the importance of neuroticism and conscientiousness in differentiating depression risk profiles, while emphasizing the added value of interpersonal traits such as extraversion and agreeableness in identifying individuals at high risk for social and emotional dysfunctions associated with increased vulnerability to depression and suicidality.

Limitations and Future Directions

This study provides quantitative evidence of systematic recall bias and the clinical usefulness of digital phenotyping. Despite its strengths, the following limitations should be considered when interpreting the results of this study. Most of all, specific methodological limitations regarding the temporal scope of the study, including the frequency of daily assessments and the overall duration of observation, warrant consideration. First, although mobile EMA offers high ecological validity, a 24-hour recall format used in this study may still be subject to reporting biases (eg, underreporting or momentary inattention) and have missed more granular, in-the-moment fluctuations that occur throughout the day. Particularly, the present EMA framework may still have been insufficient to capture the diurnal variability and transient surges of suicidal thoughts, which are often short-lived yet clinically significant. Second, it is true that the identification of the relatively small, high-risk group (27/695, 3.9%) underscores the superior sensitivity of mobile EMA in capturing the dynamic nature of suicidal risk that might otherwise be underreported in conventional retrospective assessments. However, the brief 2-week observation window represents only a snapshot of the dynamic nature of suicidal ideation in depressed individuals. Since suicidal ideation can be reactive to transient environmental stressors or life events, the long-term stability of this high-risk phenotype remains to be established, suggesting that cluster assignments might vary over a more extended longitudinal period.

Beyond these temporal considerations, additional limitations involving our analytical procedures, study design, and potential measurement effects also require careful interpretation. Third, the exclusion of 203 participants, who failed to complete the midpoint K-CESD-R survey (day 7), may have introduced selection bias. Although this criterion was applied to maintain the continuity of depressive symptoms, which is a key diagnostic criterion for MDD, the primary reason for this exclusion was adherence to the IRB-approved study protocol, which mandated the disposal of data from participants who withdrew or were lost to follow-up. Accordingly, we were not able to perform sensitivity analyses to compare the excluded and retained groups, which remains a limitation in confirming the robustness of our results against potential attrition bias. Fourth, although we observed significant decreases in week-to-week symptom severity scores across 2 different assessment methods, the observational design precludes causal inferences regarding the effects of EMA-based self-monitoring on symptom improvement. Finally, the repeated use of the same depression scale (K-CESD-R) in both online and mobile formats may have had a carryover effect, potentially improving participants' symptom recall over time and reducing recall bias in the traditional online version.

Several research directions are suggested to overcome these methodological limitations. To improve the limited temporal resolution of a 24-hour recall format, follow-up protocols could incorporate higher sampling frequencies, such as multiple assessments per day, to more precisely capture transient surges of suicidal ideation. Expanding the observation window through longitudinal study designs that span more than 2 weeks will also be essential to establish the long-term stability of the identified phenotypes. Furthermore, implementing more flexible data retention policies in coordination with IRB would allow for the inclusion of partial data in sensitivity analyses, which in turn would minimize selection bias. The integration of real-time passive sensing (eg, step count, sleep time, GPS, screen time, call or app logs, etc) with active EMA could further enhance predictive precision while alleviating participant burden. To establish the causal efficacy of EMA-based self-monitoring, future studies should also use randomized controlled trials comparing EMA with appropriate control conditions. Finally, future research should replicate this methodological approach using other validated depression scales to determine whether similar patterns of recall-related

XSL•FO
RenderX

underreporting are observed, thereby addressing potential carryover effects from repeated measurements. Such efforts to establish convergent evidence would significantly improve the reliability and generalizability of findings regarding recall bias.

Conclusions

This study demonstrates that mobile EMA provides more sensitive and ecologically valid monitoring of depressive symptoms than traditional retrospective measures. Conventional assessments are subject to both systematic underreporting (recall bias) and limited sensitivity to daily symptom variability. Further week-to-week analysis revealed a significant reactivity effect, with daily symptom reporting increasing over the 2-week period. This finding suggests that EMA acts as a mechanism to enhance self-awareness. Finally, active EMA features, including symptom severity, symptom variability, and frequency of suicidal ideation, effectively identified integrated depression and suicide risk subgroups and their distinct sociodemographic, psychological, and psychiatric profiles, highlighting the usefulness of digital phenotyping. Among these markers, symptom variability emerged as a key indicator for differentiating moderate-risk individuals, providing more detailed information than traditional measures. These findings pave the way for more personalized and timely mental health assessments and interventions in clinical practice.

Acknowledgments

The authors thank dataSpring Korea, Inc., and Digitalmedic Co., Ltd., for assistance with the data collection phase.

No generative AI was used in the writing or revising of this manuscript.

Funding

This research was supported by grants of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute, funded by the Ministry of Health & Welfare, Republic of Korea (grants RS-2024-00439193; RS-2023-KH135442). The funders had no involvement in the study design, data collection, analysis, interpretation, or the writing of the manuscript.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the institutional review board–approved participant informed consent form that permits data sharing and transferring to designated third parties but are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: JYP (lead), KC (equal)
Data curation: KC
Formal analysis: KC
Funding acquisition: JYP
Investigation: KC
Methodology: KC (lead), JYP (supporting)
Project administration: KC (lead), JYP (equal)
Software: KC (lead), JYP (equal)
Supervision: JYP
Visualization: KC
Writing – original draft: KC
Writing – review & editing: KC

Conflicts of Interest

None declared.

Checklist 1
STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist for observational studies in epidemiology (cross-sectional).
[PDF File, 119 KB - mhealth_v14i1e80332_app1.pdf]

References

1. Cai H, Jin Y, Liu S, et al. Prevalence of suicidal ideation and planning in patients with major depressive disorder: a meta-analysis of observation studies. J Affect Disord 2021 Oct 1;293:148-158. [doi: 10.1016/j.jad.2021.05.115] [Medline: 34192629]

XSL•FO
RenderX

2. Yoo HJ, Hong JP, Cho MJ, et al. Lifetime suicidal ideation and attempt in adults with full major depressive disorder versus sustained depressed mood. J Affect Disord 2016 Oct;203:275-280. [doi: 10.1016/j.jad.2016.06.007] [Medline: 27314814]
3. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®), 5th edition: American Psychiatric Association Publishing; 2013. [doi: 10.1176/appi.books.9780890425596]
4. Choi H, Lee S, Chun H, Shin JH. Suicidal ideation and attempt among young people living in gosiwons in South Korea. Soc Psychiatry Psychiatr Epidemiol 2024 Sep;59(9):1543-1549. [doi: 10.1007/s00127-024-02615-3] [Medline: 38305869]
5. Melhem NM, Porta G, Oquendo MA, et al. Severity and variability of depression symptoms predicting suicide attempt in high-risk individuals. JAMA Psychiatry 2019 Jun 1;76(6):603-613. [doi: 10.1001/jamapsychiatry.2018.4513] [Medline: 30810713]
6. Sveticic J, De Leo D. The hypothesis of a continuum in suicidality: a discussion on its validity and practical implications. Ment Illn 2012 Aug 23;4(2):e15. [doi: 10.4081/mi.2012.e15] [Medline: 25478116]
7. Sokero P, Eerola M, Rytsälä H, et al. Decline in suicidal ideation among patients with MDD is preceded by decline in depression and hopelessness. J Affect Disord 2006 Oct;95(1-3):95-102. [doi: 10.1016/j.jad.2006.04.028] [Medline: 16780959]
8. Heuschen C, Mocking RJT, Zantvoord JB, et al. Suicidal ideation in remitted major depressive disorder predicts recurrence. J Psychiatr Res 2022 Jul;151:65-72. [doi: 10.1016/j.jpsychires.2022.04.005] [Medline: 35461004]
9. Hawes MT, Klein DN. Emotion dynamics in current and remitted depression: an ecological momentary assessment study. Psychol Med 2024 Jul;54(9):2222-2229. [doi: 10.1017/S0033291724000369] [Medline: 38433585]
10. van Genugten CR, Schuurmans J, van Ballegooijen W, Hoogendoorn AW, Smit JH, Riper H. Discovering different profiles in the dynamics of depression based on real-time monitoring of mood: a first exploration. Internet Interv 2021 Jul 27;26:100437. [doi: 10.1016/j.invent.2021.100437] [Medline: 34458105]
11. Burchert S, Kerber A, Zimmermann J, Knaevelsrud C. Screening accuracy of a 14-day smartphone ambulatory assessment of depression symptoms and mood dynamics in a general population sample: comparison with the PHQ-9 depression screening. PLoS One 2021 Jan 6;16(1):e0244955. [doi: 10.1371/journal.pone.0244955] [Medline: 33406120]
12. Panaite V, Rottenberg J, Bylsma LM. Daily affective dynamics predict depression symptom trajectories among adults with major and minor depression. Affect Sci 2020 Oct 3;1(3):186-198. [doi: 10.1007/s42761-020-00014-w] [Medline: 36043207]
13. Chung K, Jeon MJ, Park JY, Lee S, Kim CO, Park JY. Development and evaluation of a mobile-optimized daily self-rating depression screening app: a preliminary study. PLoS One 2018 Jun 26;13(6):e0199118. [doi: 10.1371/journal.pone.0199118] [Medline: 29944663]
14. Chung K, Park JY, Joung D, Jhung K. Response time as an implicit self-schema indicator for depression among undergraduate students: preliminary findings from a mobile app-based depression assessment. JMIR Mhealth Uhealth 2019 Sep 13;7(9):e14657. [doi: 10.2196/14657] [Medline: 31586362]
15. Derry PA, Kuiper NA. Schematic processing and self-reference in clinical depression. J Abnorm Psychol 1981;90(4):286-297. [doi: 10.1037/0021-843X.90.4.286]
16. Kuiper NA, Olinger LJ, MacDonald MR, Shaw BF. Self-schema processing of depressed and nondepressed content: the effects of vulnerability to depression. Soc Cogn 2011 Jan;3(1):77-93. [doi: 10.1521/soco.1985.3.1.77]
17. Jabrayilov R, Emons WHM, Sijtsma K. Comparison of classical test theory and item response theory in individual change assessment. Appl Psychol Meas 2016 Nov;40(8):559-572. [doi: 10.1177/0146621616664046] [Medline: 29881070]
18. Torous J, Staples P, Shanahan M, et al. Utilizing a personal smartphone custom app to assess the Patient Health Questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder. JMIR Ment Health 2015 Mar 24;2(1):e8. [doi: 10.2196/mental.3889] [Medline: 26543914]
19. Sedano-Capdevila A, Porras-Segovia A, Bello HJ, Baca-García E, Barrigon ML. Use of ecological momentary assessment to study suicidal thoughts and behavior: a systematic review. Curr Psychiatry Rep 2021 May 18;23(7):41. [doi: 10.1007/s11920-021-01255-7] [Medline: 34003405]
20. Kivelä L, van der Does WAJ, Riese H, Antypa N. Don't miss the moment: a systematic review of ecological momentary assessment in suicide research. Front Digit Health 2022 Jun 5;4:876595. [doi: 10.3389/fdgth.2022.876595] [Medline: 35601888]
21. Gratch I, Choo TH, Galfalvy H, et al. Detecting suicidal thoughts: the power of ecological momentary assessment. Depress Anxiety 2021 Jan;38(1):8-16. [doi: 10.1002/da.23043] [Medline: 32442349]
22. Beames JR, Kikas K, Werner-Seidler A. Prevention and early intervention of depression in young people: an integrated narrative review of affective awareness and Ecological Momentary Assessment. BMC Psychol 2021 Aug 16;9(1):113. [doi: 10.1186/s40359-021-00614-6] [Medline: 34392830]
23. Baryshnikov I, Aledavood T, Rosenström T, et al. Relationship between daily rated depression symptom severity and the retrospective self-report on PHQ-9: a prospective ecological momentary assessment study on 80 psychiatric outpatients. J Affect Disord 2023 Mar 1;324:170-174. [doi: 10.1016/j.jad.2022.12.127] [Medline: 36586594]
24. van Ballegooijen W, Ruwaard J, Karyotaki E, Ebert DD, Smit JH, Riper H. Reactivity to smartphone-based ecological momentary assessment of depressive symptoms (MoodMonitor): protocol of a randomised controlled trial. BMC Psychiatry 2016 Oct 21;16(1):359. [doi: 10.1186/s12888-016-1065-5] [Medline: 27769201]

XSL•FO
RenderX

25. Palmier-Claus JE, Rogers A, Ainsworth J, et al. Integrating mobile-phone based assessment for psychosis into people's everyday lives and clinical care: a qualitative study. BMC Psychiatry 2013 Jan 23;13:1-12. [doi: 10.1186/1471-244X-13-34] [Medline: 23343329]
26. Faurholt-Jepsen M, Frost M, Ritz C, et al. Daily electronic self-monitoring in bipolar disorder using smartphones—the MONARCA I trial: a randomized, placebo-controlled, single-blind, parallel group trial. Psychol Med 2015 Oct;45(13):2691-2704. [doi: 10.1017/S0033291715000410] [Medline: 26220802]
27. Kauer SD, Reid SC, Crooke AHD, et al. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. J Med Internet Res 2012 Jun 25;14(3):e67. [doi: 10.2196/jmir.1858] [Medline: 22732135]
28. Kramer I, Simons CJP, Hartmann JA, et al. A therapeutic application of the experience sampling method in the treatment of depression: a randomized controlled trial. World Psychiatry 2014 Feb;13(1):68-77. [doi: 10.1002/wps.20090] [Medline: 24497255]
29. Simons CJP, Hartmann JA, Kramer I, et al. Effects of momentary self-monitoring on empowerment in a randomized controlled trial in patients with depression. Eur Psychiatry 2015 Nov;30(8):900-906. [doi: 10.1016/j.eurpsy.2015.09.004] [Medline: 26647864]
30. Borghouts J, Eikey E, Mark G, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387. [doi: 10.2196/24387] [Medline: 33759801]
31. Eaton WW, Muntaner C, Smith C, Tien A, Ybarra M. Center for epidemiologic studies depression scale: review and revision (CESD and CESD-R). In: The Use of Psychological Testing for Treatment Planning and Outcomes Assessment: Routledge; 2004:363-377.
32. Lee S, Oh ST, Ryu SY, Jun JY, Lee K, Lee E, et al. Validation of the Korean version of Center for Epidemiologic Studies Depression Scale-Revised (K-CESD-R) [Article in Korean]. Korean J Psychosom Med 2016;24(1):83-93 [FREE Full text]
33. Park HI, Jeon M, Ahn JS, Chung K, Park JY. Development and validation of the digital sensitivity scale for adults: cross-sectional observational study. J Med Internet Res 2025 Jan 10;27:e55828. [doi: 10.2196/55828] [Medline: 39793004]
34. Kim JH, Kim BH, Ha MS. Validation of a Korean version of the Big Five Inventory [Article in Korean]. J Human Understanding Couns 2011;32(1):47-65 [FREE Full text]
35. John OP, Srivastava S. The Big-Five trait taxonomy: history, measurement, and theoretical perspectives. In: Pervin LA, John OP, editors. Handbook of Personality: Theory and Research, 2nd edition: Guilford Press; 1999:102-139.
36. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
37. Park SJ, Choi HR, Choi JH, Kim KW, Hong JP. Reliability and validity of the Korean Version of the Patient Health Questionnaire-9 (PHQ-9) [Article in Korean]. Anxiety Mood 2010;6(2):119-124 [FREE Full text]
38. Spielberger CD, Gorsuch RL, Lushene RE. Manual for the State-Trait Anxiety Inventory: Consulting Psychologists Press; 1970.
39. Kim JT. The Relationship Between Character-Anxiety and Sociability: Focusing on Spielberger's STAI [Thesis in Korean]. : Korea University; 1978 URL: https://www.riss.kr/link?id=T7816838 [accessed 2026-06-26]
40. Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001 Jul;2(4):297-307. [doi: 10.1016/s1389-9457(00)00065-4] [Medline: 11438246]
41. Cho YW, Song ML, Morin CM. Validation of a Korean version of the Insomnia Severity Index. J Clin Neurol 2014 Jul;10(3):210-215. [doi: 10.3988/jcn.2014.10.3.210] [Medline: 25045373]
42. Lyoo YC, Ju S, Kim E, Kim JE, Lee JH. The Patient Health Questionnaire-15 and its abbreviated version as screening tools for depression in Korean college and graduate students. Compr Psychiatry 2014 Apr;55(3):743-748. [doi: 10.1016/j.comppsych.2013.11.011] [Medline: 24342054]
43. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition: Lawrence Erlbaum Associates; 1988.
44. Horwitz AG, Zhao Z, Sen S. Peak-end bias in retrospective recall of depressive symptoms on the PHQ-9. Psychol Assess 2023 Apr;35(4):378-381. [doi: 10.1037/pas0001219] [Medline: 36757996]
45. Bos EH, de Jonge P, Cox RFA. Affective variability in depression: revisiting the inertia-instability paradox. Br J Psychol 2019 Nov;110(4):814-827. [doi: 10.1111/bjop.12372] [Medline: 30588616]
46. 2023 Suicide Status Survey (National Perception Survey on Suicide) [Report in Korean]. : Ministry of Health and Welfare; 2023 URL: https://www.mohw.go.kr/board.es?mid=a10411010200&bid=0019&act=view&list_no=1480883#share [accessed 2026-06-26]
47. 2023 Korea National Health and Nutrition Examination Survey [Webpage in Korean]. Korea Disease Control and Prevention Agency (KDCA). 2023. URL: https://kfsp-datazoom.or.kr/gov_nutrition.do [accessed 2026-01-26]
48. Choi YH, Yang KI, Yun CH, Kim WJ, Heo K, Chu MK. Impact of insomnia symptoms on the clinical presentation of depressive symptoms: a cross-sectional population study. Front Neurol 2021 Aug 9;12:716097. [doi: 10.3389/fneur.2021.716097] [Medline: 34434165]
49. Madigan A, Daly M. Socioeconomic status and depressive symptoms and suicidality: the role of subjective social status. J Affect Disord 2023 Apr 1;326:36-43. [doi: 10.1016/j.jad.2023.01.078] [Medline: 36709827]
50. Jeon GS, Ha Y, Choi E. Effects of objective and subjective socioeconomic status on self-rated health, depressive symptoms, and suicidal ideation in adolescents. Child Ind Res 2013 Sep;6(3):479-492. [doi: 10.1007/s12187-013-9180-z]

XSL•FO
RenderX

51. Zhao S, Yan W, Tao L, Zhang J. The association between relative deprivation, depression, and youth suicide: evidence from a psychological autopsy study. Omega (Westport) 2024 Sep;89(4):1691-1713. [doi: 10.1177/00302228231190595] [Medline: 37462026]
52. Kim YY, Park HJ. Factors influencing suicidal ideation according to the deprivation indices of different regions in Korea. J Prev Med Public Health 2023 Jan;56(1):88-94. [doi: 10.3961/jpmph.22.356] [Medline: 36746426]
53. Wetherall K, Robb KA, O'Connor RC. Social rank theory of depression: a systematic review of self-perceptions of social rank and their relationship with depressive symptoms and suicide risk. J Affect Disord 2019 Mar 1;246:300-319. [doi: 10.1016/j.jad.2018.12.045] [Medline: 30594043]
54. Yang Z, Li A, Roske C, Alexander N, Gabbay V. Personality traits as predictors of depression across the lifespan. J Affect Disord 2024 Jul 1;356:274-283. [doi: 10.1016/j.jad.2024.03.073] [Medline: 38537757]
55. Morgan S, Cooper B, Paul S, et al. Association of personality profiles with depressive, anxiety, and cancer-related symptoms in patients undergoing chemotherapy. Pers Individ Dif 2017 Oct 15;117:130-138. [doi: 10.1016/j.paid.2017.05.039] [Medline: 29479128]
56. Weiss A, Sutin AR, Duberstein PR, Friedman B, Bagby RM, Costa PT Jr. The personality domains and styles of the five-factor model are related to incident depression in Medicare recipients aged 65 to 100. Am J Geriatr Psychiatry 2009 Jul;17(7):591-601. [doi: 10.1097/jgp.0b013e31819d859d] [Medline: 19554673]
57. Tan GCY, Wang Z, Tan ESE, et al. Transdiagnostic clustering of self-schema from self-referential judgements identifies subtypes of healthy personality and depression. Front Neuroinform 2023;17:1244347. [doi: 10.3389/fninf.2023.1244347] [Medline: 38274390]
58. Cai H, Song H, Yang Y, et al. Big-Five personality traits and depression: chain mediation of self-efficacy and walking. Front Psychiatry 2024 Nov 29;15:1460888. [doi: 10.3389/fpsyt.2024.1460888] [Medline: 39676915]

Abbreviations

AICC: Akaike Information Criterion Corrected
BIC: Bayesian Information Criterion
CESD-R: Center for Epidemiologic Studies Depression Scale-Revised
CONSORT: Consolidated Standards of Reporting Trials
***DSM-5*:** *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition
EMA: ecological momentary assessment
FW: first week
GLMM: Generalized Linear Mixed Model
ICC: intraclass correlation coefficient
IRB: institutional review board
ISI: Insomnia Severity Index
ISI-K: Korean version of the Insomnia Severity Index
K-CESD-R: Korean Version of the Center for Epidemiologic Studies Depression Scale–Revised
MDD: major depressive disorder
MDE: major depressive episode
PHQ-9: Patient Health Questionnaire–9
SES: subjective socioeconomic status
STAI: Spielberger State-Trait Anxiety Inventory
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
SW: second week

Edited by A Stone, L Buis; submitted 08.Jul.2025; peer-reviewed by E Czyz; revised version received 05.Jun.2026; accepted 05.Jun.2026; published 31.Jul.2026.

Please cite as:
Chung K, Park JY
Identifying Symptom Dynamics and Profiles of Subgroups at Risk for Major Depression and Suicidal Ideation Among Korean Adults: 2-Week Ecological Momentary Assessment Study
JMIR Mhealth Uhealth 2026;14:e80332
URL: https://mhealth.jmir.org/2026/1/e80332
doi:10.2196/80332

XSL•FO
RenderX

© Kyungmi Chung, Jin Young Park. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 31.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Associations Between 24-Hour Physical Behavior, Self-Perceived Stress, and Coping Self-Efficacy in Everyday Life: Ambulatory Assessment Study

Katrin Bonn[1,2]; Doreen Wohlfarth[3]; Irina Timm[2]; Oliver Bender[1], Prof Dr; Ulrich W Ebner-Priemer[2], Prof Dr; Marco Giurgiu[4], Prof Dr

[1]Student Health Management, Baden-Wuerttemberg Cooperative State University Karlsruhe, Erzbergerstraße 121, Karlsruhe, Baden-Wuerttemberg, Germany

[2]Mental mHealth Lab, Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Karlsruhe, Baden-Wuerttemberg, Germany

[3]Division of Sports and Rehabilitation Medicine, University Hospital Ulm, Ulm, Baden-Wuerttemberg, Germany

[4]Institute of Movement Therapy and movement-oriented Prevention and Rehabilitation, German Sport University Cologne, Cologne, North Rhine-Westphalia, Germany

Corresponding Author:
Katrin Bonn
Student Health Management, Baden-Wuerttemberg Cooperative State University Karlsruhe, Erzbergerstraße 121, Karlsruhe, Baden-Wuerttemberg, Germany

Abstract

Background: Psychological stress poses a risk to mental and physical health and has become a major public health challenge. As physical behaviors (ie, physical activity, sedentary behavior (SB), and sleep) play a key role in mental well-being, their targeted modification could be an approach to coping with stress in everyday life. Previous studies have primarily either analyzed the associations between isolated physical behaviors and stress-related outcomes or employed cross-sectional designs. Accordingly, there is a need for deeper insights into the within- and between-person associations between physical behavior over a 24-hour cycle and psychological stress in naturalistic settings.

Objective: This study aimed to investigate how 24-hour physical behavior compositions are associated with daily self-perceived stress and stress-related coping self-efficacy and how replacing time in one behavior with another is linked to changes in both stress-related indicators.

Methods: A total of 198 healthy university employees (mean age 35.87 y, SD 10.76; 109, 54.8% female) participated in a 15-day ambulatory assessment study. Participants reported their momentary stress and coping self-efficacy perceptions up to 6 times a day in electronic diaries. 24–hour physical behavior was measured using a thigh-worn Move 4 accelerometer. The movement data were then classified on a daily basis into the 4 behavior categories of sleep, SB, light physical activity (LPA), and moderate-to-vigorous physical activity (MVPA). In order to obtain physical behavior compositions, an isometric log-ratio transformation was applied, resulting in 4 different sets. The associations between physical behavior compositions, self-perceived stress, and coping self-efficacy were analyzed using 2-level mixed multilevel models. Exploratory reallocation models were conducted to simulate the effects of time shifts from one behavior to another on both stress-related outcomes.

Results: The geometric average day comprised 33.9% (8.1/24 h) sleep, 45.2% (10.8/24 h) SB, 15.8% (3.8/24 h) LPA, and 5.1% (1.2/24 h) in MVPA. More time spent sleeping compared to being sedentary was associated with lower self-perceived stress (standardized β=–.03; t_{1988}=–2.045; P=.04) but not with coping self-efficacy (t_{1988}=–1.333; P=.18) in the 24-hour cycle. The ratio of SB to the other physical behaviors and time spent in LPA or MVPA relative to SB showed no association with either stress-related outcome. Significant random effects indicate high individual variability in the analyzed associations. The exploratory substitution of SB by LPA, MVPA, or sleep showed no significant changes in self-perceived stress or coping self-efficacy within a 60-minute period.

Conclusions: Adapting 24-hour physical behavior seems to be a promising approach to reduce stress in everyday life, for example, by extending sleep duration instead of being awake in SB. Further research should be conducted on contextual and personal influencing factors in order to develop tailored stress management interventions for the 24-hour cycle.

(JMIR Mhealth Uhealth 2026;14:e81502) doi:10.2196/81502

KEYWORDS

ambulatory assessment; accelerometry; coping; mobile health; mHealth; perceived stress; sleep

XSL•FO
RenderX

Introduction

The feeling of stress is a familiar human experience in everyday life. In particular, high, recurrent, and long-lasting psychological stress is thought to be associated with various physical and mental illnesses, such as coronary artery disease or depression [1-4]. Furthermore, a link is suggested between higher stress levels and an increased risk of mortality [5,6]. Across several studies, reviews indicate that the prevalence of self-reported stress during the COVID-19 pandemic ranged from 32% to 56%, with a focus on both the general population and specific subgroups, such as patients, health care workers, pregnant women, and students [7-9]. Among higher education employees from 16 different countries, 73% reported experiencing moderate to high levels of stress [10]. Psychological stress has become a public health issue with far-reaching implications for society [11], such as an increased risk of substance misuse or addiction [12-14], poorer relationship quality [15-17], or the financial burden resulting from stress-related productivity losses and absenteeism [11,18].

Studies indicate that coping self-efficacy, which refers to the belief that one is capable of effectively managing one's own stress [19-21], might contribute to effective coping [22] and reduced perceived stress in different target groups [23-25]. Numerous studies have explored associations between individual physical behaviors, including physical activity (PA), sedentary behavior (SB), or sleep, with psychological stress and coping strategies. PA was found to exhibit a medium beneficial effect on stress in adults [26], although the association is characterized by a high degree of individual variability [27,28]. More inconclusive results are reported in studies on the association between indicators of SB and stress [29-33]. Thus, further studies are required that take different time frames, contexts, and domains into consideration [27,30,33]. An insufficient sleep duration appears to be another indicator of experiencing more stress [34-36], while sleep quality was positively associated with higher coping self-efficacy in adolescents [23]. Since physical behaviors do not occur in isolation, their interrelationships must be taken into account. This is addressed by the growing paradigm in sport and health science, which conceptualizes physical behaviors as compositional elements within a 24-hour cycle [37,38]. Reflecting this perspective, recent practice-oriented activity guidelines—such as those promoted under the motto "Make Your Whole Day Matter"—increasingly emphasize the need for balanced time allocation across PA, SB, and sleep, including the recommendation to limit screen time [39,40].

A growing body of literature addresses physical behavior over a 24-hour period and its associations with mental health. Current systematic reviews highlight the positive effects of replacing time in SB with PA for mental health in adults and children [41-43]. For example, among adults, spending 15 minutes more in light physical activity (LPA) instead of being sedentary was associated with lower anxiety, depression, and stress scores [44]. Additionally, reallocating 30 minutes in SB to moderate PA in women or to walking in men aged 21 to 35 years both resulted in a decline in self-reported stress levels [45]. Other cross-sectional studies indicate that replacing time spent in SB with sleep in adults is also associated with lower perceived stress and improved mood [46,47]. However, we are not aware of any study that has examined the association between 24-hour physical behavior and stress-related coping. Overall, much of the evidence linking physical behavior to perceived stress is derived from cross-sectional designs [41,45,47] that analyze between-person differences at a single time point and are therefore inherently limited in their ability to elucidate temporal within-person dynamics, including day-to-day variability in 24-hour physical behavior and stress [48,49].

To overcome these limitations, intensive longitudinal studies (also known as ambulatory assessment or ecological momentary assessment) are currently the state - of - the - art methodology for examining within - person associations [50,51]. These study designs have several advantages, namely, the assessment in everyday life, in real - time, with device - based methods, and repeated measurements with a high sampling frequency, which enables researchers to track dynamic relationships [48,52]. Therefore, they bypass laboratory distortions and minimize recall biases associated with traditional approaches [53]. Accordingly, the implementation of ambulatory assessments is promising for capturing fluctuations in stress-related processes and their associations with physical behavior throughout the day [54]. To the best of our knowledge, no previous study has focused on this within-person interplay. We aimed to investigate the association of device-based 24-hour physical behavior, self-perceived stress, and coping self-efficacy in everyday life. Based on previous empirical findings and theoretical models [42,46], we expected that the composition of 24-hour physical behavior is associated with both daily perceived stress (hypotheses 1 - 4) and coping self-efficacy (hypotheses 5 - 8) in healthy adults.

We hypothesized that the ratio of SB to sleep, LPA, and moderate-to-vigorous physical activity (MVPA) would be positively associated with self-perceived daily stress (H1), whereas more time spent in (1) MVPA relative to SB (H2), (2) LPA relative to SB (H3), and (3) sleep relative to SB (H4) would be associated with less self-perceived stress. Furthermore, we expected the proportion of SB relative to the other physical behaviors to be negatively associated with coping self-efficacy (H5), while a higher ratio of (1) MVPA to SB (H6), (2) LPA to SB (H7), and (3) sleep to SB (H8) is associated with a stronger belief in being able to cope with tasks at hand. Additional exploratory analyses were conducted to examine the effects of reallocating time from one physical behavior to another on perceived stress and coping self-efficacy. Furthermore, subjective sleep quality and exercise behavior were exploratively fitted into the existing models in order to obtain indications of the robustness of results.

Methods

Ethical Considerations

This study was approved by the Ethics Committee of the Karlsruhe Institute of Technology (A2020-001). Prior to the start of the study, participants received comprehensive information about the aim and procedure of the investigation and provided informed consent. Participants had the option of

XSL•FO
RenderX

withdrawing from the study at any time. In order to protect participants' privacy and confidentiality, we deidentified the data after collection. The successful completion of the study was rewarded with a fitness tracker worth €100 (US $114.24).

Participants

Potential participants were recruited via flyers, mailings, word of mouth, and an internal newsletter from the Karlsruhe Institute of Technology. The study leaders consulted with interested parties via email or telephone to ensure that all predefined inclusion criteria for the study were met. Eligibility criteria included predominantly sedentary work and the absence of current physical injuries or psychological disorders. In total, 222 university employees participated in the study between October 2022 and December 2023 in Karlsruhe, Germany.

Study Design

This study represents an intensive longitudinal method [55] and is a secondary analysis of a within-person encouragement design investigation, the details of which can be found in Giurgiu et al [56]. However, no overlapping analyses were conducted.

The study started with an in-person session during which participants received detailed information about the study procedures, completed a questionnaire assessing demographics, and were equipped with a study smartphone (Nokia G50 or 6; Nokia Corporation), a Fitbit wearable (Fitbit Inspire 2; Google), and a Move 4 accelerometer (Movisens GmbH). The ambulatory assessment period began the following day and lasted 15 working days. To maintain motivation and ensure compliance, weekend days were not included [57]. During the study period, participants were asked to wear the Fitbit wearable on their wrist and the Move 4 sensor on their thigh as continuously as possible. Moreover, the smartphone prompted participants up to 8 times per day to answer short e-diary questionnaires, with e-diaries occurring based on a wearable-triggered sedentary algorithm, except for the additional morning and evening queries.

Measures

24-Hour Physical Behavior

Triaxial accelerometer data assessed by a Move 4 sensor (Movisens GmbH) with a sampling frequency of 64 Hz and a range of ±16 gravitational acceleration were used to measure 24-hour physical behavior. The raw accelerometer data were preprocessed after the study period with the software ActiPASS (version 2024.05; ActiPASS) [58]. ActiPASS implements algorithms for nonwear, sleep detection, posture, and activity intensity derived from cadence measures that have demonstrated high accuracy for wake-time movement behaviors and sleep in validity studies [59-61]. Based on the preprocessed data, we aggregated the most commonly examined physical behavior categories—sleep, SB (sitting or lying episodes outside of sleep intervals), LPA (walking with a cadence <100 steps/min and standing), and MVPA (walking with a cadence ≥100 steps/min)—as average minutes per day [38]. In line with previous wearing protocols [62], participants and days were only included when (1) participants had at least 3 valid wear-time days (≥20 h of wear time per day), (2) at least 1 period of walking was detected that day, and (3) days comprised more than zero minutes of sleep.

Momentary Self-Perceived Stress and Coping Self-Efficacy

Participants received all e-diaries on the study smartphone (Nokia G50 or 6). A morning and an evening survey were conducted between fixed time periods (morning: 7 AM-8:30 AM, evening: 9:45 PM-11 PM), during which participants were permitted to respond at their chosen time. The e-diary assessments during the day occurred up to 6 times per day between 8:40 AM and 9:40 PM. They were based on a sedentary-triggered design [63], appearing after 30 consecutive minutes in a sitting or lying body position. While half of the triggers exclusively prompted questions on momentary constructs such as affective states or perceived stress, the other half included a prompt to interrupt SB for 3 minutes before answering the identical e-diary questions. To reduce participants' burden, prompts could be delayed up to 15 minutes and did not occur within at least 30 minutes after a completed e-diary rating.

Momentary perceived stress and coping self-efficacy were each captured by one item in the e-diary assessments completed during the day and then aggregated at the daily level. Previous studies showed single-item measures to be valid and reliable for stress measurement [64]. Both items were based on the Perceived Stress Scale-10 [65]. Following Jones et al [31], participants were asked to rate their subjectively perceived stress level on a visual Likert scale from 1 to 5 ("Do you feel nervous or stressed at the moment?," 1="not at all" to 5="extremely"). Based on the question "Do you currently feel that you can cope with all your upcoming tasks?," we assessed participants' current coping self-efficacy on a Likert scale ranging from 1="not at all" to 5="extremely."

Covariates

In the initial in-person session, age in years as well as sex (female vs male) were assessed through self-reports, and anthropometric data were objectively measured in order to calculate the BMI. One adapted item from the Pittsburgh Sleep Quality Index [66] was implemented in the morning survey to assess the perceived sleep quality of the previous night on a scale from zero to 100 ("How would you rate your sleep quality last night?"). In the evening survey, participants used a dichotomous response format to indicate whether they had exercised that day (yes vs no). A definition of the term "sports exercise" was not provided.

Statistical Analysis

Prior to compositional data analysis procedures, we used an isometric log-ratio (ilr) transformation to account for the multicollinearity of direct composition measures [67]. The compositions of sleep, SB, LPA, and MVPA were expressed as ilr coordinates, containing information about one activity in relation to the remaining physical behaviors in the 24-hour cycle. In accordance with existing literature, 3 zero values were replaced before the transformation using the log-ratio expectation maximization algorithm [68].

XSL•FO
RenderX

As is the state of the art when analyzing hierarchically structured intensive longitudinal data, we used a multilevel approach [69]. Prior to fitting the multilevel models, we determined intraclass correlations (ICCs) using unconditional models with daily averaged self-perceived stress (hypotheses 1 - 4) and coping self-efficacy (hypotheses 5 - 8) as outcomes. Afterwards, multilevel models with repeated measurements (level 1) nested within participants (level 2) were estimated using restricted maximum likelihood and an unstructured covariance matrix for the random effects. In the interest of model parsimony, nonsignificant random effects were excluded. Model assumptions, including multicollinearity, homoscedasticity, and normal distribution of residuals, were checked [69,70]. The covariates *age* (years), *sex* (female vs male), *BMI* (kg/m^2), and *coping self-efficacy* (only in hypotheses 1-4) were selected a priori and added as predictors based on data availability as well as known associations [71-79]. The respective compositions represented by ilr1, ilr2, and ilr3 were centered within clusters in advance and integrated as predictors. The following equation model refers to hypothesis 1 (further models are presented in Supplementary File A in Multimedia Appendix 1).

- H1: Y (stress perception)$_{ij} = \gamma_{00} + \gamma_{01} \times sex_j + \gamma_{02} \times age_j + \gamma_{03} \times BMI_j + \gamma_{10} \times$ *coping self-efficacy*$_{ij} + \gamma_{20} \times$ ilr1 (SB/LPA×MVPA×Sleep)$_{ij} + \gamma_{30} \times$ ilr2 (LPA/MVPA×Sleep)$_{ij} + \gamma_{40} \times$ ilr3 (MVPA/Sleep)$_{ij} + u_{0j} + u_{1j} \times$ ilr3 (MVPA/Sleep)$_{ij} + r_{ij}$
- ilr1 = $\sqrt{(3/4)}$ ln (SB / (LPA×MVPA×Sleep))
- ilr2 = $\sqrt{(2/3)}$ ln (LPA/$\sqrt{}$(MVPA×Sleep))
- ilr3 = $\sqrt{(1/2)}$ ln (MVPA/Sleep)

Exploratively, we used substitution analyses to examine the expected change in the outcome when time was reallocated between physical behaviors in a 60-minute window. Pairwise reallocations were calculated in 20-minute increments, with estimates based on the average composition of the included participants [80].

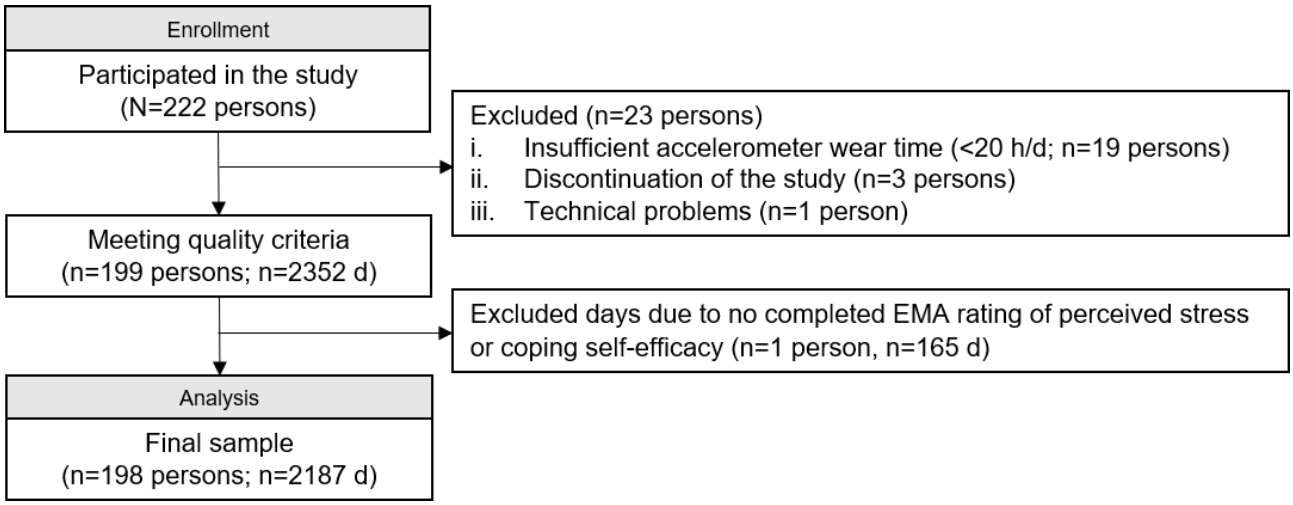
All analyses were performed in R (version 4.4.0; The R Project for Statistical Computing), using the packages *compositions* [81], *zCompositions* [82], and *lme4* [83], with a predetermined α level of .05. To facilitate the comparison of results, standardized β coefficients were calculated based on established procedures [70].

Results

Descriptive Statistics

Due to insufficient accelerometer wear time (<20 h/d, n=19), discontinuation of the study (n=3), and technical problems (n=1), 23 participants were excluded from the analysis (Figure 1). In total, 199 participants answered momentary assessment prompts on 2352 days, representing an average compliance rate of 69.9% at the personal level. On average, 4.4 surveys were completed per day. Days without any stress or coping self-efficacy rating (n=165 d) were excluded from the analysis, which led to the exclusion of one further participant due to the lack of valid days (n=1). A total of 198 participants were included in the final sample, of whom 109 (54.8%) were female participants, and the average age was 35.87 (SD 10.76) years (Table 1).

Figure 1. Participant flow chart. EMA: ecological momentary assessment.



XSL•FO
RenderX

Table . Descriptive characteristics of study participants (N=198, female 54.8%, 109/199).

Variable	Mean (SD)	Minimum	Maximum
Age (y)[a]	35.87 (10.76)	20.18	64.74
BMI (kg/m²)[a]	23.99 (3.28)	17.87	35.25
Valid days[b]	11.84 (3.86)	3	22
Sleep (h/d)[c]	7.95 (1.12)	4.27	12.70
SB[d] (h/d)[c]	10.51 (1.11)	6.68	14.14
LPA[e] (h/d)[c]	3.94 (1.00)	1.91	7.33
MVPA[f] (h/d)[c]	1.35 (0.42)	0.51	2.89
Self-perceived stress[c]	2.34 (0.65)	1.00	4.30
Coping self-efficacy[c]	3.64 (0.65)	1.00	5.00

[a]Aggregated within participants.

[b]Accelerometer wear time ≥20 h/d, aggregated within participants.

[c]Accelerometer wear time ≥20 h/d, aggregated within days and participants.

[d]SB: sedentary behavior.

[e]LPA: light physical activity.

[f]MVPA: moderate-to-vigorous physical activity.

Over the course of all days, participants wore the accelerometer for an average of 23.79 (SD 0.65) hours per day. The 24-hour geometric mean composition of physical behavior for all analyzed days consisted of 33.9% sleep (8.1/24 h), 45.2% SB (10.8/24 h), 15.8% LPA (3.8/24 h), and 5.1% MVPA (1.2/24 h). Further information on participants' physical behavior, self-perceived stress, and coping self-efficacy is provided in Table 1.

Across all days, the average daily perceived stress score was 2.36 (SD 0.87, mean of per-person minimum=1.49, mean of per-person maximum=3.28), with lower values indicating lower stress perception than higher values (scale: 1 - 5). ICC calculations (ρI=0.483) revealed that 51.7% of the variance of the stress perception was based on the within-person fluctuations. The average coping self-efficacy score (scale: 1 - 5) was 3.62 (SD 0.85, mean of per-person minimum=2.83, mean of per-person maximum=4.33), where 40.8% of the variance can be explained by within-person variances (ρI=0.592).

Association Between 24-Hour Physical Behavior and Self-Perceived Stress

Daily self-perceived stress was not significantly associated with time spent in (1) SB relative to the other behaviors (H1), (2) MVPA relative to SB (H2), or (3) LPA relative to SB (H3) (Table 2). In line with hypothesis 4, results indicate a significant negative association between ilr3 (sleep/SB) and self-perceived stress (standardized β=–.03; t_{1988}=–2.045; *P*=.04). Coping self-efficacy and age of the participants were both significantly negatively related to perceived stress in all models (Table 2). The presence of significant random effects in models 1, 2, and 3 indicates that the effects of the ratio of MVPA to sleep, sleep to MVPA and SB, as well as sleep to MVPA, LPA, and SB, vary between individuals (Supplementary File B in Multimedia Appendix 2).

XSL•FO
RenderX

Table . Multilevel models analyses to predict daily self-perceived stress: random and fixed effects.

Variable	Self-perceived stress											
	H1[a]			H2[b]			H3[c]			H4[d]		
	B (SE)[e]/SD (VE)[f],	95% CI	*P* value	*B* (SE)[e]/SD (VE)	95% CI	*P* value	*B* (SE)[e]/SD (VE)	95% CI	*P* value	*B* (SE)[e]/SD (VE)	95% CI	*P* value
Fixed effects[g]												
Intercept, $ß_{00}$	2.29 (0.35)	1.59 to 2.98	<.001	2.26 (0.36)	1.57 to 2.96	<.001	2.26 (0.35)	1.57 to 2.96	<.001	2.26 (0.36)	1.56 to 2.95	<.001
Sex[h], $ß_{01}$	–0.10 (0.09)	–0.28 to 0.07	.26	–0.10 (0.09)	–0.28 to 0.08	.28	–0.10 (0.09)	–0.28 to 0.08	.28	–0.10 (0.09)	–0.28 to 0.08	.27
Age (y), $ß_{02}$	–0.01 (0.00)	–0.02 to 0.00	.03	–0.01 (0.00)	–0.02 to 0.00	.04	–0.01 (0.00)	–0.02 to 0.00	.04	–0.01 (0.00)	–0.02 to 0.00	.04
BMI (kg/m^2), $ß_{03}$	0.02 (0.01)	–0.01 to 0.05	.19	0.02 (0.01)	–0.01 to 0.05	.18	0.02 (0.01)	–0.01 to 0.05	.18	0.02 (0.01)	–0.01 to 0.05	.17
Coping self-efficacy, $ß_{10}$	–0.57 (0.02)	–0.61 to –0.53	<.001	–0.57 (0.02)	–0.61 to –0.53	<.001	–0.57 (0.02)	–0.61 to –0.53	<.001	–0.57 (0.02)	–0.61 to –0.52	<.001
ilr1[i], $ß_{20}$	0.05 (0.05)	–0.04 to 0.14	.23	–0.02 (0.04)	–0.10 to 0.05	.53	–0.06 (0.06)	–0.17 to 0.05	.28	0.03 (0.03)	–0.03 to 0.09	.33
ilr2, $ß_{30}$	–0.01 (0.04)	–0.09 to 0.07	.87	–0.08 (0.06)	–0.19 to 0.03	.14	0.01 (0.04)	–0.06 to 0.08	.78	0.00 (0.03)	–0.06 to 0.07	.94
ilr3, $ß_{40}$	0.06 (0.04)	–0.02 to 0.15	.12	–0.02 (0.04)	–0.09 to 0.06	.65	–0.05 (0.04)	–0.12 to 0.03	.22	–0.10 (0.05)	–0.20 to 0.00	.04
Random effects[g]												
Intercept, u_0	0.61 (0.38)	0.55 to 0.68	<.001	0.61 (0.38)	0.55 to 0.68	<.001	0.61 (0.38)	0.55 to 0.68	<.001	0.61 (0.38)	0.55 to 0.68	<.001
ilr3, u_1	0.22 (0.05)	0.13 to 0.31	<.001	N/A[j]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ilr2, u_1	N/A	N/A	N/A	0.31 (0.09)	0.18 to 0.43	<.001	N/A	N/A	N/A	N/A	N/A	N/A
ilr1, u_1	N/A	N/A	N/A	N/A	N/A	N/A	0.31 (0.10)	0.19 to 0.44	<.001	N/A	N/A	N/A
Int.-Slope corr.[k]	–0.23	–0.56 to 0.12	N/A	0.09	–0.25 to 0.42	N/A	0.14	–0.19 to 0.46	N/A	N/A	N/A	N/A
Residual, r[l]	0.54 (0.29)	0.52 to 0.56	N/A	0.54 (0.29)	0.52 to 0.55	N/A	0.54 (0.29)	0.52 to 0.55	N/A	0.55 (0.30)	0.53 to 0.56	N/A

[a]Hypothesis 1: ilr1 (sedentary behavior [SB]/(light physical activity [LPA]*moderate-to-vigorous physical activity [MVPA]*Sleep)), ilr2 (LPA/(MVPA*Sleep)), ilr3 (MVPA/Sleep).

[b]Hypothesis 2: ilr1 (LPA/(Sleep×MVPA×SB)), ilr2 (Sleep/(MVPA×SB)), ilr3 (MVPA/SB).

[c]Hypothesis 3: ilr1 (Sleep/(MVPA×LPA×SB)), ilr2 (MVPA/(LPA×SB)), ilr3 (LPA/SB).

[d]Hypothesis 4: ilr1 (MVPA/(LPA×Sleep×SB)), ilr2 (LPA/(Sleep×SB)), ilr3 (Sleep/SB).

[e]Unstandardized estimates and SEs.

[f]VE: variance estimate.

[g]Fixed effects are presented as B (SE); random effects are presented as SD (VE).

[h]Compared to female.

[i]ilr: isometric log ratio coordinate.

[j]N/A: not applicable.

XSL•FO
RenderX

[k]Intercept-slope correlations are presented as correlation coefficients with 95% CIs.

[l]Residual values are presented as SD (VE).

Association Between 24-Hour Physical Behavior and Coping Self-Efficacy

Contrary to our hypotheses, the composition of SB to LPA, MVPA, and sleep was not significantly associated with the daily averaged coping self-efficacy score (H5). Furthermore, none of the other expected associations between the compositions of physical behavior and coping self-efficacy (H6, H7, H8) were observed (Table 3). A positive association between coping self-efficacy and age became evident in all 4 models (H5-H8). Additionally, random effects were found for the ratio of LPA to SB, as well as LPA to all other remaining activities (Table 3).

XSL•FO
RenderX

Table . Multilevel models analyses to predict daily coping self-efficacy: random and fixed effects.

Variable	Coping self-efficacy											
	H5[a]			H6[b]			H7[c]			H8[d]		
	B (SE)[e]/SD (VE)[f]	95% CI	*P* value	*B* (SE)[e]/SD (VE)	95% CI	*P* value	*B* (SE)[e]/SD (VE)	95% CI	*P* value	*B* (SE)[e]/SD (VE)	95% CI	*P* value
Fixed effects[g]												
Intercept, $ß_{00}$	3.36 (0.38)	2.61 to 4.11	<.001	3.36 (0.38)	2.61 to 4.11	<.001	3.37 (0.38)	2.62 to 4.12	<.001	3.36 (0.38)	2.61 to 4.11	<.001
Sex[h], $ß_{01}$	0.14 (0.10)	–0.05 to 0.33	.15	0.14 (0.10)	–0.05 to 0.33	.15	0.14 (0.10)	–0.05 to 0.34	.15	0.14 (0.10)	–0.05 to 0.33	.15
Age (y), $ß_{02}$	0.01 (0.00)	0.00 to 0.02	.02	0.01 (0.00)	0.00 to 0.02	.02	0.01 (0.00)	0.00 to 0.02	.03	0.01 (0.00)	0.00 to 0.02	.02
BMI (kg/m^2), $ß_{03}$	–0.01 (0.02)	–0.04 to 0.02	.65	–0.01 (0.02)	–0.04 to 0.02	.64	–0.01 (0.02)	–0.04 to 0.02	.64	–0.01 (0.02)	–0.04 to 0.02	.65
ilr1[i], $ß_{10}$	0.01 (0.05)	–0.08 to 0.10	.85	0.09 (0.04)	0.00 to 0.18	.052	–0.10 (0.05)	–0.20 to –0.01	.02	0.03 (0.03)	–0.03 to 0.09	.37
ilr2, $ß_{20}$	0.08 (0.04)	0.00 to 0.15	.06	–0.08 (0.05)	–0.17 to 0.01	.10	–0.01 (0.04)	–0.09 to 0.06	.75	0.08 (0.03)	0.01 to 0.15	.02
ilr3, $ß_{30}$	0.08 (0.03)	0.01 to 0.15	.02	0.01 (0.04)	–0.06 to 0.09	.74	0.05 (0.04)	–0.04 to 0.14	.26	–0.07 (0.05)	–0.17 to 0.03	.18
Random effects[g]												
Intercept, u_0	0.67 (0.44)	0.59 to 0.73	<.001	0.67 (0.44)	0.59 to 0.74	<.001	0.67 (0.44)	0.59 to 0.74	<.001	0.67 (0.44)	0.59 to 0.73	<.001
ilr1, u_1	N/A[j]	N/A	N/A	0.28 (0.08)	0.15 to 0.39	.005	N/A	N/A	N/A	N/A	N/A	N/A
ilr3, u_1	N/A	N/A	N/A	N/A	N/A	N/A	0.26 (0.07)	0.12 to 0.38	.02	N/A	N/A	N/A
Int.-Slope corr.[k]	N/A	N/A	N/A	–0.03	–0.37 to 0.31	N/A	–0.11	–0.49 to 0.23	N/A	N/A	N/A	N/A
Residual, r[l]	0.56 (0.31)	0.54 to 0.58	N/A	0.55 (0.30)	0.53 to 0.57	N/A	0.55 (0.30)	0.53 to 0.57	N/A	0.56 (0.31)	0.54 to 0.58	N/A

[a]Hypothesis 5: ilr1 (SB/(light physical activity [LPA]×moderate-to-vigorous physical activity [MVPA]×Sleep)), ilr2 (LPA/(MVPA*Sleep)), ilr3 (MVPA/Sleep).

[b]Hypothesis 6: ilr1 (LPA/(Sleep×MVPA×SB)), ilr2 (Sleep/(MVPA×SB)), ilr3 (MVPA/SB).

[c]Hypothesis 7: ilr1 (Sleep/(MVPA×LPA×SB)), ilr2 (MVPA/(LPA×SB)), ilr3 (LPA/SB).

[d]Hypothesis 8: ilr1 (MVPA/(LPA×Sleep×SB)), ilr2 (LPA/(Sleep×SB)), ilr3 (Sleep/SB).

[e]Unstandardized estimates and SEs.

[f]VE: variance estimate.

[g]Fixed effects are presented as B (SE); random effects are presented as SD (VE).

[h]Compared to female.

[i]ilr: isometric log ratio coordinate.

[j]N/A: not applicable.

[k]Intercept-slope correlations are presented as correlation coefficients with 95% CIs.

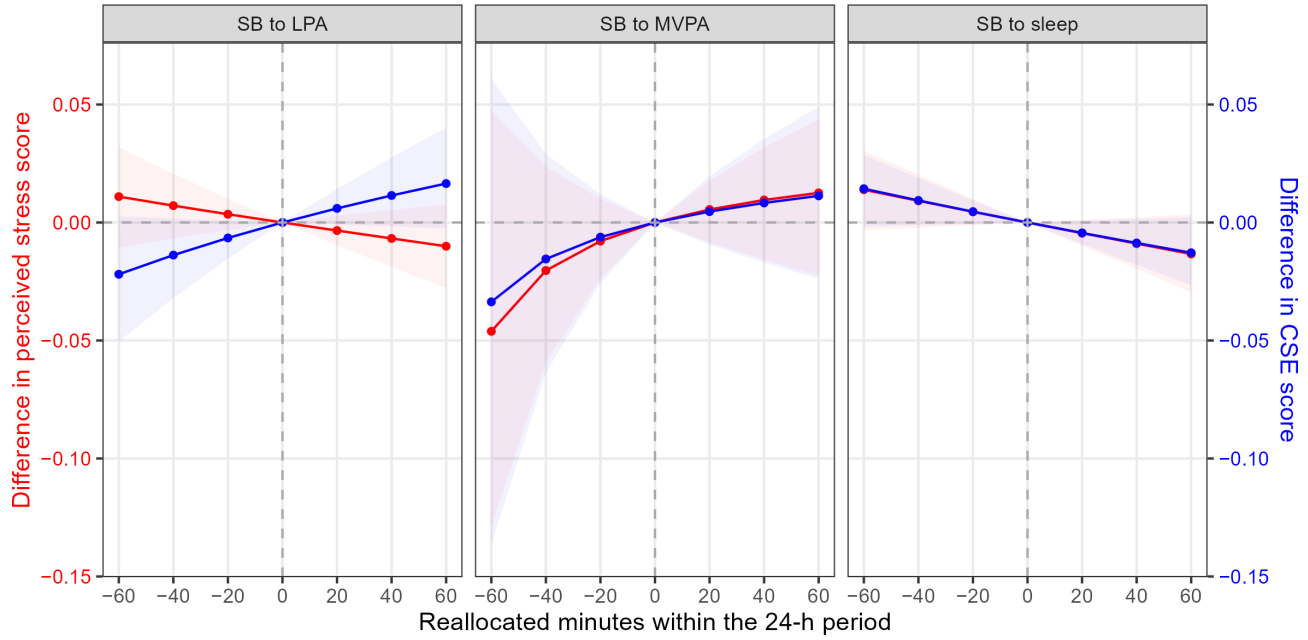
[l]Residual values are presented as SD (VE).

XSL•FO
RenderX

Exploratory Analyses

We exploratively calculated reallocation analyses within a 60-minute frame, simulating the effects of replacing the same amount of time (20-min increments) in one physical behavior with another on self-perceived stress and coping self-efficacy. The starting point represents the empirically observed mean-centered composition of 490 minutes sleep, 649 minutes SB, 227 minutes LPA, and 74 minutes MVPA. The replacement of SB by LPA, MVPA, or sleep was analyzed, with no significant changes of self-perceived stress or coping self-efficacy within a 60-minute period (Figure 2).

Figure 2. Estimated changes in predicted daily averaged self-perceived stress and coping self-efficacy associated with pairwise reallocations[a]. [a]95% CIs were derived by bootstrapping with 1000 replicates. CSE: coping-self-efficacy; LPA: light physical activity; MVPA: moderate-to-vigorous physical activity; SB: sedentary behavior.



As part of our exploratory analyses, self-reported exercise behavior was added as an additional fixed predictor in all models including PA (models 1 - 3 and 5 - 7; Supplementary File C in Multimedia Appendix 3, Tables 1 and 2). We found no statistically significant association between exercise behavior and perceived stress or coping self-efficacy. However, adding exercise behavior revealed a significant positive association of time in SB relative to the remaining physical behaviors on perceived stress (standardized β=.04; t_{1379}=2.039; P=.04). Furthermore, adding subjective sleep quality as a fixed predictor in both models investigating the association of the ratio of sleep to SB revealed sleep quality to be associated with coping self-efficacy (model 8: standardized β=.08; t_{1769}=5.279; P<.001) but not with perceived stress (model 4: standardized β=–.01; t_{1769}=–0.724; P=.47) (Supplementary File C in Multimedia Appendix 3 and Table 3). The addition of sleep quality did not alter the effect of the ratio of sleep duration to SB on self-perceived stress (standardized β=–.03; t_{1776}=–2.364; P=.02).

Discussion

Principal Findings

The relationship between individual physical behaviors and stress has been studied extensively in recent years. Our 15-day ambulatory assessment study with 198 participants expands existing evidence with real-time assessments in a naturalistic setting, providing within- and between-subject insights into participants' everyday lives. We aimed to investigate the associations between device-based 24-hour physical behavior compositions with perceived stress and stress-related coping self-efficacy in an intensive longitudinal data design. Our results indicate that time spent sleeping relative to being sedentary over the 24-hour cycle is negatively associated with perceived daily stress level (H4) but not with coping self-efficacy (H8). The ratios of SB to the remaining behaviors (H1/H5), MVPA to SB (H2/H6), and LPA to SB (H3/H7) were not significantly related to either self-perceived stress or coping self-efficacy. In the exploratory analyses, replacing time spent in SB by LPA, MVPA, or sleep within a 60-minute period did not result in any significant changes in perceived stress or coping self-efficacy.

One possible explanation for why the assumed associations between the composition of SB and the other physical behaviors with perceived stress or coping self-efficacy were not evident could be that the level of mental workload during SB was not taken into account. Research has shown that mentally passive SB, for example, watching television, can increase the risk of depression, while mentally active SB, such as reading, is considered a protective factor [84,85]. As the relationship between SB and stress-related parameters is particularly important for developing targeted health-promoting measures in working environments with high levels of sedentariness, it may be advisable to consider the aggregation of data not on a daily basis but rather within the context of working and nonworking hours [74]. When adding exercise behavior as an additional fixed predictor to our main models (H1/H5), more time in SB compared to the remaining behaviors was associated with higher stress levels. Accordingly, the association between

XSL•FO
RenderX

more time in SB compared to the remaining behaviors and stress levels might not be independent of other constructs and influencing factors such as mental load or time of the day. Therefore, future research endeavors might be interested in analyzing the associations between the composition of physical behaviors and momentary stress levels at different aggregation levels (eg, 1 or 2 h) [86].

We found a negative association between the time spent sleeping relative to SB and self-perceived stress, which is in line with previous compositional analysis studies suggesting that replacing sedentary time with sleep is associated with lower psychological stress in young individuals [46] and in working adults [47]. As Bassett et al [87] identified sleep quality, but not sleep duration, to influence cortisol responses to acute psychosocial stress, we exploratively included sleep quality as a further fixed predictor in our model. Sleep quality did not emerge as an additional influencing factor in our model, whereas the association between the ratio of sleep to SB and perceived stress remained significant, underscoring its robustness. Contrary to our hypotheses, a higher ratio of sleep to SB was not related to coping self-efficacy, even after adding sleep quality as a further exploratory fixed effect. The exploratively expanded model revealed a positive association between sleep quality and coping self-efficacy, which is consistent with the findings of Ten Brink et al [23]. This study investigated the importance of coping self-efficacy in breaking the bidirectional cycle between sleep and stress in adolescents. The findings showed that sleep quality not only leads to a higher coping self-efficacy but also a lower level of reported stress the following day [23]. Based on our results, we can only vaguely expect an indirect effect of sleep quality on stress perception, as coping self-efficacy appears to be significantly negatively associated with self-perceived stress in all models (H1-H4). However, these findings emphasize the potential of promoting one's belief in being able to cope with stress for interventions to reduce stress in everyday life, especially as some areas of work as well as personal obligations do not allow for an increase in sleep duration.

The compositions of LPA to SB and MVPA to SB in the 24-hour cycle were not significantly associated with either perceived stress or coping self-efficacy. In the limited existing literature, Rong et al [45], for example, found that replacing self-reported SB with walking or moderate PA had a positive effect on perceived stress. One reason for the differing results might be that subjective measures, possibly related to common method variance [88], tend to be more predictive of affect-related outcomes than objective measures [89]. Additionally, our study design (ie, half of the assessments were captured after a sedentary bout of ≥30 min) could have resulted in lower daily stress ratings. Our analyses revealed various significant random effects, for example, for the ratio of LPA to SB and LPA to remaining behaviors on coping self-efficacy, thus underlining the variability across participants regarding the association of 24-hour physical behavior and both stress-related outcomes. The occurrence of these within-subject associations should be taken into account in future investigations. Momentary assessments could provide further insights into the temporal sequencing within a day and help assess potential contextual influences, such as the environment of PA, since it has been found that outdoor PA is associated with lower negative affect than indoor PA [90].

Overall, based on our results, we suggest examining the relationship between compositional predictors, psychological stress, and further moderating and mediating factors in more detail. Studies and replications should be conducted with diverse and stress-prone samples, such as students [91], to gain application-oriented insights into the interplay of the 24-hour physical behavior and mental health outcomes. A more comprehensive understanding of the complex association between the mechanisms underlying SB, PA, and sleep and the biological, psychosocial, and behavioral facets of stress could advance a greater focus of the 24-hour approach not only in stress management but also in public health [92,93] and health promotion recommendations [34].

Limitations

This study has a number of strengths, but a few limitations must be taken into account. The fact that the sample included only university employees with predominantly sedentary working conditions limits the generalizability of the results. Additionally, the observational study design does not allow the elucidation of causalities. Due to its multidimensionality, the assessment of the stress is challenging [94]. In line with previous studies [31,64], we measured both stress-related outcomes with subjective, self-report-based single-item measures that capture only the stress that participants are aware of [95,96]. However, using an ambulatory assessment design promoted the reduction of retrospective biases [50]. Half of our assessments occurred after a sedentary bout, which may have resulted in reduced variance, so that further random sampling studies are required. However, ICC calculations show that both stress-related outcomes can be adequately explained by within-person variance. As we aggregated device-based activity data within days from 12 AM to 11:59 PM, the actual sleep time per day might be distorted, because the time in bed before midnight is not represented as sleep time on the following day. Finally, the implemented sleep algorithm used may incorrectly classify daytime sleep as being sedentary (ie, if not a primary time-in-bed period was detected) [59]. However, this appears to play a subordinate role in the analyses, as daytime sleep was only reported in the evening e-diaries on 5.6% (122/2187) of all analyzed days.

Conclusions

This study investigated the associations between device-based 24-hour physical behavior, self-perceived stress, and coping self-efficacy in daily life. Our findings suggest that spending more time asleep relative to being sedentary in the 24-hour period is associated with lower levels of perceived stress on the same day. However, the ratio of sleep to SB was not associated with the perceived ability to cope with daily demands. The composition of SB to the other physical behaviors and time in LPA or MVPA relative to SB showed no association with either stress-related outcome. Nonetheless, within-subject associations and descriptive trends of behavioral reallocation hint at the potential relevance of daily activity patterns for handling stressful situations. These findings underscore the need for future ecologically valid studies that incorporate contextual and

XSL•FO
RenderX

individual variables, as well as momentary assessments, in order to better understand the role of physical behavior in stress regulation. Such insights are necessary in order to provide target group–appropriate health-promoting recommendations for coping with stress in everyday life.

Acknowledgments

The authors would like to sincerely thank the participants who contributed to the study. They further thank Selina Schneider, Til Becker, Ellen Rulf, Rebecca Nissen, Jonas Behlke, and Lennart Buday for their research support.

The authors declare the use of generative artificial intelligence in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to generative artificial intelligence tools under full human supervision: code optimization. ChatGPT (GPT-4) by OpenAI (free version) was used to support code optimization, which was further reviewed and revised by the study group.

Funding

This research was funded by the German Research Foundation (grant 496846758).

Data Availability

The datasets used and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

MG and UWE-P conceptualized the study and developed the methodology or design. MG and IT collected the data. KB and DW analyzed the data and prepared the original draft. MG supervised the preparation of the original draft. All authors reviewed and edited the manuscript, approved the final version, and agreed with the order of presentation of the authors.

Conflicts of Interest

UWE-P receives personal fees for consulting from Boehringer-Ingelheim, which is not related to the submitted work. The other author has declared no conflicts of interest.

Multimedia Appendix 1
Structural equation models.
[DOCX File, 44 KB - mhealth_v14i1e81502_app1.docx]

Multimedia Appendix 2
Individual associations of self-perceived stress and the ratio of sleep to moderate-to-vigorous physical activity (MVPA), light physical activity (LPA), and sedentary behavior (SB) at the participant level.
[DOCX File, 251 KB - mhealth_v14i1e81502_app2.docx]

Multimedia Appendix 3
Exploratory analyses: results of adapted multilevel models.
[DOCX File, 63 KB - mhealth_v14i1e81502_app3.docx]

References

1. Cohen S, Janicki-Deverts D, Miller GE. Psychological stress and disease. JAMA 2007 Oct 10;298(14):1685-1687. [doi: 10.1001/jama.298.14.1685] [Medline: 17925521]
2. Hammen C. Stress and depression. Annu Rev Clin Psychol 2005;1(1):293-319. [doi: 10.1146/annurev.clinpsy.1.102803.143938] [Medline: 17716090]
3. Keller A, Litzelman K, Wisk LE, et al. Does the perception that stress affects health matter? The association with health and mortality. Health Psychol 2012 Sep;31(5):677-684. [doi: 10.1037/a0026743] [Medline: 22201278]
4. Kivimäki M, Steptoe A. Effects of stress on the development and progression of cardiovascular disease. Nat Rev Cardiol 2018 Apr;15(4):215-229. [doi: 10.1038/nrcardio.2017.189] [Medline: 29213140]
5. Glei DA, Weinstein M. Daily exposure to stressors, daily perceived severity of stress, and mortality risk among US adults. PLOS ONE 2024;19(5):e0303266. [doi: 10.1371/journal.pone.0303266] [Medline: 38748692]
6. Russ TC, Stamatakis E, Hamer M, Starr JM, Kivimäki M, Batty GD. Association between psychological distress and mortality: individual participant pooled analysis of 10 prospective cohort studies. BMJ 2012 Jul 31;345:e4933. [doi: 10.1136/bmj.e4933] [Medline: 22849956]

XSL•FO
RenderX

7. Blasco-Belled A, Tejada-Gallardo C, Fatsini-Prats M, Alsinet C. Mental health among the general population and healthcare workers during the COVID-19 pandemic: a meta-analysis of well-being and psychological distress prevalence. Curr Psychol 2024 Mar;43(9):8435-8446. [doi: 10.1007/s12144-022-02913-6]
8. Necho M, Tsehay M, Birkie M, Biset G, Tadesse E. Prevalence of anxiety, depression, and psychological distress among the general population during the COVID-19 pandemic: a systematic review and meta-analysis. Int J Soc Psychiatry 2021 Nov;67(7):892-906. [doi: 10.1177/00207640211003121] [Medline: 33794717]
9. Amini-Rarani M, Rezaei S, Azami-Aghdash S, Bashzar S, Allahverdi S, Mohseni M. The prevalence of stress during the COVID-19 pandemic: an umbrella review. BMC Public Health 2025 Nov 18;25(1):4022. [doi: 10.1186/s12889-025-24922-3] [Medline: 41254586]
10. Rahman MA, Das P, Lam L, et al. Health and wellbeing of staff working at higher education institutions globally during the post-COVID-19 pandemic period: evidence from a cross-sectional study. BMC Public Health 2024 Jul 11;24(1):1848. [doi: 10.1186/s12889-024-19365-1] [Medline: 38992680]
11. Smith MD, Wesselbaum D. Global evidence on the prevalence of and risk factors associated with stress. J Affect Disord 2025 Apr 1;374:179-183. [doi: 10.1016/j.jad.2025.01.053] [Medline: 39805499]
12. Sinha R. Chronic stress, drug use, and vulnerability to addiction. Ann N Y Acad Sci 2008 Oct;1141(1):105-130. [doi: 10.1196/annals.1441.030] [Medline: 18991954]
13. Papp LM, Kouros CD, Armstrong L, Curtin JJ. College students' momentary stress and prescription drug misuse in daily life: testing direct links and the moderating roles of global stress and coping. Stress Health 2023 Apr;39(2):361-371. [doi: 10.1002/smi.3191] [Medline: 35994279]
14. Sinha R. Stress and substance use disorders: risk, relapse, and treatment outcomes. J Clin Invest 2024 Aug 15;134(16):e172883. [doi: 10.1172/JCI172883] [Medline: 39145454]
15. Knox L, Karantzas G, Ferguson E. The role of attachment, insecurity, and stress in partner maltreatment: a meta-analysis. Trauma Violence Abuse 2024 Jan;25(1):721-737. [doi: 10.1177/15248380231161012] [Medline: 37036150]
16. Neff LA, Nguyen TTT, Williamson HC. Too stressed to help? The effects of stress on noticing partner needs and enacting support. Pers Soc Psychol Bull 2021 Nov;47(11):1565-1579. [doi: 10.1177/0146167220974490] [Medline: 33342361]
17. Randall AK, Bodenmann G. The role of stress on close relationships and marital satisfaction. Clin Psychol Rev 2009 Mar;29(2):105-115. [doi: 10.1016/j.cpr.2008.10.004] [Medline: 19167139]
18. Hassard J, Teoh KRH, Visockaite G, Dewe P, Cox T. The cost of work-related stress to society: a systematic review. J Occup Health Psychol 2018 Jan;23(1):1-17. [doi: 10.1037/ocp0000069] [Medline: 28358567]
19. Bandura A. Self-Efficacy: The Exercise of Control: W H Freeman & Co; 1997.
20. Chesney MA, Neilands TB, Chambers DB, Taylor JM, Folkman S. A validity and reliability study of the coping self-efficacy scale. Br J Health Psychol 2006 Sep;11(Pt 3):421-437. [doi: 10.1348/135910705X53155] [Medline: 16870053]
21. Schwarzer R, Renner B. Social-cognitive predictors of health behavior: action self-efficacy and coping self-efficacy. Health Psychol 2000 Sep;19(5):487-495. [doi: 10.1037/0278-6133.19.5.487] [Medline: 11007157]
22. Drake EC, Sladek MR, Doane LD. Daily cortisol activity, loneliness, and coping efficacy in late adolescence: a longitudinal study of the transition to college. Int J Behav Dev 2016 Jul 1;40(4):334-345. [doi: 10.1177/0165025415581914] [Medline: 28979055]
23. Ten Brink M, Lee HY, Manber R, Yeager DS, Gross JJ. Stress, sleep, and coping self-efficacy in adolescents. J Youth Adolesc 2021 Mar;50(3):485-505. [doi: 10.1007/s10964-020-01337-4] [Medline: 33141378]
24. Xu F, Huang L. Impacts of stress and risk perception on mental health of college students during the COVID-19 pandemic: the mediating role of coping efficacy. Front Psychiatry 2021;12:767189. [doi: 10.3389/fpsyt.2021.767189] [Medline: 35185635]
25. Chang MW, Brown R, Wegener DT. Perceived stress linking psychosocial factors and depressive symptoms in low-income mothers. BMC Public Health 2021 Jan 6;21(1):62. [doi: 10.1186/s12889-020-10118-4] [Medline: 33407305]
26. Singh B, Olds T, Curtis R, et al. Effectiveness of physical activity interventions for improving depression, anxiety and distress: an overview of systematic reviews. Br J Sports Med 2023 Sep;57(18):1203-1209. [doi: 10.1136/bjsports-2022-106195] [Medline: 36796860]
27. Wright LJ, Williams SE, van Zanten JJCSV. Associations between physical activity, sedentary behaviour, and stress using ecological momentary assessment: a scoping review. Ment Health Phys Act 2023 Mar;24:100518. [doi: 10.1016/j.mhpa.2023.100518]
28. Burg MM, Schwartz JE, Kronish IM, et al. Does stress result in you exercising less? Or does exercising result in you being less stressed? Or is it both? Testing the bi-directional stress–exercise association at the group and person (N of 1) level. Ann Behav Med 2017 Dec;51(6):799-809. [doi: 10.1007/s12160-017-9902-4] [Medline: 28290065]
29. Chu AHY, van Dam RM, Biddle SJH, Tan CS, Koh D, Müller-Riemenschneider F. Self-reported domain-specific and accelerometer-based physical activity and sedentary behaviour in relation to psychological distress among an urban Asian population. Int J Behav Nutr Phys Act 2018 Apr 5;15(1):36. [doi: 10.1186/s12966-018-0669-1] [Medline: 29618384]
30. Felez-Nobrega M, Bort-Roig J, Briones L, et al. Self-reported and activPALTM-monitored physical activity and sedentary behaviour in college students: not all sitting behaviours are linked to perceived stress and anxiety. J Sports Sci 2020 Jul;38(13):1566-1574. [doi: 10.1080/02640414.2020.1748359] [Medline: 32279592]

XSL•FO
RenderX

31. Jones M, Taylor A, Liao Y, Intille SS, Dunton GF. Real-time subjective assessment of psychological stress: associations with objectively-measured physical activity levels. Psychol Sport Exerc 2017 Jul;31:79-87. [doi: 10.1016/j.psychsport.2017.03.013] [Medline: 29151810]
32. Pinto BM, Kindred MD, Dunsiger SI, Williams DM. Sedentary behavior among breast cancer survivors: a longitudinal study using ecological momentary assessments. J Cancer Surviv 2021 Aug;15(4):546-553. [doi: 10.1007/s11764-020-00948-x] [Medline: 33029754]
33. Teychenne M, Stephens LD, Costigan SA, Olstad DL, Stubbs B, Turner AI. The association between sedentary behaviour and indicators of stress: a systematic review. BMC Public Health 2019 Oct 23;19(1):1357. [doi: 10.1186/s12889-019-7717-x] [Medline: 31647002]
34. Kastelic K, Pedišić Ž, Lipovac D, Kastelic N, Chen ST, Šarabon N. Associations of meeting 24-h movement guidelines with stress and self-rated health among adults: is meeting more guidelines associated with greater benefits? BMC Public Health 2021 May 17;21(1):929. [doi: 10.1186/s12889-021-10979-3] [Medline: 34001090]
35. Kim HJ, Oh SY, Joo JH, Choi DW, Park EC. The relationship between sleep duration and perceived stress: findings from the 2017 Community Health Survey in Korea. Int J Environ Res Public Health 2019 Sep 3;16(17):3208. [doi: 10.3390/ijerph16173208] [Medline: 31484289]
36. Yap Y, Slavish DC, Taylor DJ, Bei B, Wiley JF. Bi-directional relations between stress and self-reported and actigraphy-assessed sleep: a daily intensive longitudinal study. Sleep 2020 Mar 12;43(3):zsz250. [doi: 10.1093/sleep/zsz250] [Medline: 31608395]
37. Rosenberger ME, Fulton JE, Buman MP, et al. The 24-hour activity cycle: a new paradigm for physical activity. Med Sci Sports Exerc 2019 Mar;51(3):454-464. [doi: 10.1249/MSS.0000000000001811] [Medline: 30339658]
38. Pedišić Z, Dumuid D, Olds TS. Integrating sleep, sedentary behaviour, and physical activity research in the emerging field of time-use epidemiology: definitions, concepts, statistical methods, theoretical framework, and future directions. Kinesiology (Zagreb) 2017;49(2):252-269 [FREE Full text]
39. Canadian Society for Exercise Physiology. URL: https://csepguidelines.ca/ [accessed 2025-01-15]
40. 24-hour movement guidelines for all australians. Australian Government Department of Health, Disability and Ageing. URL: https://www.health.gov.au/topics/physical-activity-and-exercise/physical-activity-and-exercise-guidelines-for-all-australians/ [accessed 2025-01-15]
41. Groves CI, Huong C, Porter CD, et al. Associations between 24-h movement behaviors and indicators of mental health and well-being across the lifespan: a systematic review. J Act Sedentary Sleep Behav 2024 Mar 14;3(1):9. [doi: 10.1186/s44167-024-00048-6] [Medline: 40217439]
42. Kracht CL, Burkart S, Groves CI, et al. 24-hour movement behavior adherence and associations with health outcomes: an umbrella review. J Act Sedentary Sleep Behav 2024;3(1):25. [doi: 10.1186/s44167-024-00064-6] [Medline: 39399355]
43. Zhou Y, Huang Z, Liu Y, Liu D. The effect of replacing sedentary behavior with different intensities of physical activity on depression and anxiety in Chinese university students: an isotemporal substitution model. BMC Public Health 2024 May 23;24(1):1388. [doi: 10.1186/s12889-024-18914-y] [Medline: 38783202]
44. Curtis RG, Dumuid D, McCabe H, Singh B, Ferguson T, Maher C. The association between 24-hour activity, sedentary and sleep compositions and mental health in Australian adults: a cross-sectional study. J Act Sedentary Sleep Behav 2023 Aug 1;2(1):15. [doi: 10.1186/s44167-023-00024-6] [Medline: 40217512]
45. Rong F, Li X, Jia L, et al. Substitutions of physical activity and sedentary behavior with negative emotions and sex difference among college students. Psychol Sport Exerc 2024 May;72:102605. [doi: 10.1016/j.psychsport.2024.102605] [Medline: 38346583]
46. Meyer JD, Ellingson LD, Buman MP, Shook RP, Hand GA, Blair SN. Current and 1-year psychological and physical effects of replacing sedentary time with time in other behaviors. Am J Prev Med 2020 Jul;59(1):12-20. [doi: 10.1016/j.amepre.2020.02.018] [Medline: 32418803]
47. Kitano N, Kai Y, Jindo T, Tsunoda K, Arao T. Compositional data analysis of 24-hour movement behaviors and mental health in workers. Prev Med Rep 2020 Dec;20:101213. [doi: 10.1016/j.pmedr.2020.101213] [Medline: 33204604]
48. Kanning MK, Ebner-Priemer UW, Schlicht WM. How to investigate within-subject associations between physical activity and momentary affective states in everyday life: a position statement based on a literature overview. Front Psychol 2013;4:187. [doi: 10.3389/fpsyg.2013.00187] [Medline: 23641221]
49. Wang X, Cheng Z. Cross-sectional studies: strengths, weaknesses, and recommendations. Chest 2020 Jul;158(1S):S65-S71. [doi: 10.1016/j.chest.2020.03.012] [Medline: 32658654]
50. Reichert M, Giurgiu M, Koch E, et al. Ambulatory assessment for physical activity research: state of the science, best practices and future directions. Psychol Sport Exerc 2020 Sep;50:101742. [doi: 10.1016/j.psychsport.2020.101742] [Medline: 32831643]
51. Trull TJ, Ebner-Priemer U. Ambulatory assessment. Annu Rev Clin Psychol 2013;9:151-176. [doi: 10.1146/annurev-clinpsy-050212-185510] [Medline: 23157450]
52. Ebner-Priemer UW, Koudela S, Mutz G, Kanning M. Interactive multimodal ambulatory monitoring to investigate the association between physical activity and affect. Front Psychol 2012;3:596. [doi: 10.3389/fpsyg.2012.00596] [Medline: 23355828]

XSL•FO
RenderX

53. Bussmann JBJ, Ebner-Priemer UW, Fahrenberg J. Ambulatory activity monitoring. Eur Psychol 2009 Jan;14(2):142-152. [doi: 10.1027/1016-9040.14.2.142]
54. Brüßler S, von Haaren-Mack B, Vogelsang A, Reichert M. Ambulatory assessment for recovery and stress monitoring in the general population. In: Fostering Recovery and Well-Being in a Healthy Lifestyle: Routledge; 2024:38-55 URL: https://www.taylorfrancis.com/chapters/edit/10.4324/9781003250654-4/ambulatory-assessment-recovery-stress-monitoring-general-population-sarah-br%C3%BC%C3%9Fler-birte-von-haaren-mack-anna-vogelsang-markus-reichert [accessed 2026-05-02]
55. Bolger N, Laurenceau JP. Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research: Guilford Publications; 2013.
56. Giurgiu M, Timm I, Ebner-Priemer UW, Schmiedek F, Neubauer AB. Causal effects of sedentary breaks on affective and cognitive parameters in daily life: a within-person encouragement design. npj Ment Health Res 2024 Dec 21;3(1):64. [doi: 10.1038/s44184-024-00113-7] [Medline: 39706901]
57. Wrzus C, Neubauer AB. Ecological momentary assessment: a meta-analysis on designs, samples, and compliance across research fields. Assessment 2023 Apr;30(3):825-846. [doi: 10.1177/10731911211067538] [Medline: 35016567]
58. Hettiarachchi P, Johansson P. ActiPASS. Zenodo. URL: https://zenodo.org/records/11119064 [accessed 2026-05-02]
59. Johansson PJ, Crowley P, Axelsson J, et al. Development and performance of a sleep estimation algorithm using a single accelerometer placed on the thigh: an evaluation against polysomnography. J Sleep Res 2023 Apr;32(2):e13725. [doi: 10.1111/jsr.13725] [Medline: 36167935]
60. Crowley P, Skotte J, Stamatakis E, et al. Comparison of physical behavior estimates from three different thigh-worn accelerometers brands: a proof-of-concept for the Prospective Physical Activity, Sitting, and Sleep consortium (ProPASS). Int J Behav Nutr Phys Act 2019 Aug 16;16(1):65. [doi: 10.1186/s12966-019-0835-0] [Medline: 31419998]
61. Hettiarachchi P, Aili K, Holtermann A, Stamatakis E, Svartengren M, Palm P. Validity of a non-proprietary algorithm for identifying lying down using raw data from thigh-worn triaxial accelerometers. Sensors (Basel) 2021 Jan 29;21(3):904. [doi: 10.3390/s21030904] [Medline: 33572815]
62. Blodgett JM, Ahmadi MN, Atkin AJ, et al. Device-measured physical activity and cardiometabolic health: the Prospective Physical Activity, Sitting, and Sleep (ProPASS) consortium. Eur Heart J 2024 Feb 7;45(6):458-471. [doi: 10.1093/eurheartj/ehad717] [Medline: 37950859]
63. Giurgiu M, Niermann C, Ebner-Priemer U, Kanning M. Accuracy of sedentary behavior-triggered ecological momentary assessment for collecting contextual information: development and feasibility study. JMIR Mhealth Uhealth 2020 Sep 15;8(9):e17852. [doi: 10.2196/17852] [Medline: 32930668]
64. Littman AJ, White E, Satia JA, Bowen DJ, Kristal AR. Reliability and validity of 2 single-item measures of psychosocial stress. Epidemiology (Sunnyvale) 2006 Jul;17(4):398-403. [doi: 10.1097/01.ede.0000219721.89552.51] [Medline: 16641618]
65. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [doi: 10.2307/2136404] [Medline: 6668417]
66. Buysse DJ, Reynolds CF III, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
67. Dumuid D, Pedišić Ž, Palarea-Albaladejo J, Martín-Fernández JA, Hron K, Olds T. Compositional data analysis in time-use epidemiology: what, why, how. Int J Environ Res Public Health 2020 Mar 26;17(7):2220. [doi: 10.3390/ijerph17072220] [Medline: 32224966]
68. Rasmussen CL, Palarea-Albaladejo J, Johansson MS, et al. Zero problems with compositional data of physical behaviors: a comparison of three zero replacement methods. Int J Behav Nutr Phys Act 2020 Oct 6;17(1):126. [doi: 10.1186/s12966-020-01029-z] [Medline: 33023619]
69. Snijders TAB, Bosker RJ. Multilevel Analysis: An Introduction to Basic and Advances Multilevel Modeling: SAGE Publications; 2011.
70. Hox JJ, Moerbeek M, van de Schoot R. Multilevel Analysis: Techniques and Applications: Taylor & Francis; 2017.
71. Hamarat E, Thompson D, Zabrucky KM, Steele D, Matheny KB, Aysan F. Perceived stress and coping resource availability as predictors of life satisfaction in young, middle-aged, and older adults. Exp Aging Res 2001;27(2):181-196. [doi: 10.1080/036107301750074051] [Medline: 11330213]
72. Suglia SF, Pamplin JR II, Forde AT, Shelton RC. Sex differences in the association between perceived stress and adiposity in a nationally representative sample. Ann Epidemiol 2017 Oct;27(10):626-631. [doi: 10.1016/j.annepidem.2017.09.009] [Medline: 28985986]
73. Graves BS, Hall ME, Dias-Karch C, Haischer MH, Apter C. Gender differences in perceived stress and coping among college students. PLOS ONE 2021;16(8):e0255634. [doi: 10.1371/journal.pone.0255634] [Medline: 34383790]
74. Courtney JB, Turner JA, Puterman E, Almeida DM. Within-person associations between daily stress and physical activity during working and non-working hours. Psychol Sport Exerc 2025 Jan;76:102777. [doi: 10.1016/j.psychsport.2024.102777] [Medline: 39551252]
75. Stefaniak AR, Blaxton JM, Bergeman CS. Age differences in types and perceptions of daily stress. Int J Aging Hum Dev 2022 Mar;94(2):215-233. [doi: 10.1177/00914150211001588] [Medline: 33739147]

XSL•FO
RenderX

76. Du C, Adjepong M, Zan MCH, et al. Gender differences in the relationships between perceived stress, eating behaviors, sleep, dietary risk, and body mass index. Nutrients 2022 Feb 28;14(5):1045. [doi: 10.3390/nu14051045] [Medline: 35268020]
77. Tenk J, Mátrai P, Hegyi P, et al. Perceived stress correlates with visceral obesity and lipid parameters of the metabolic syndrome: a systematic review and meta-analysis. Psychoneuroendocrinology 2018 Sep;95:63-73. [doi: 10.1016/j.psyneuen.2018.05.014] [Medline: 29803182]
78. Tomiyama AJ. Stress and obesity. Annu Rev Psychol 2019 Jan 4;70:703-718. [doi: 10.1146/annurev-psych-010418-102936] [Medline: 29927688]
79. Mouchacca J, Abbott GR, Ball K. Associations between psychological stress, eating, physical activity, sedentary behaviours and body weight among women: a longitudinal study. BMC Public Health 2013 Sep 11;13(1):828. [doi: 10.1186/1471-2458-13-828] [Medline: 24020677]
80. Dumuid D, Pedišić Ž, Stanford TE, et al. The compositional isotemporal substitution model: a method for estimating changes in a health outcome for reallocation of time between sleep, physical activity and sedentary behaviour. Stat Methods Med Res 2019 Mar;28(3):846-857. [doi: 10.1177/0962280217737805] [Medline: 29157152]
81. van den Boogaart KG, Tolosana-Delgado R. "compositions": a unified R package to analyze compositional data. Comput Geosci 2008 Apr;34(4):320-338. [doi: 10.1016/j.cageo.2006.11.017]
82. Palarea-Albaladejo J, Martín-Fernández JA. zCompositions — R package for multivariate imputation of left-censored data under a compositional approach. Chemometr Intell Lab Syst 2015 Apr;143:85-96. [doi: 10.1016/j.chemolab.2015.02.019]
83. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67:1-48. [doi: 10.18637/jss.v067.i01]
84. Hallgren M, Dunstan DW, Owen N. Passive versus mentally active sedentary behaviors and depression. Exerc Sport Sci Rev 2020 Jan;48(1):20-27. [doi: 10.1249/JES.0000000000000211] [Medline: 31663866]
85. Werneck AO, Hoare E, Stubbs B, van Sluijs EMF, Corder K. Associations between mentally-passive and mentally-active sedentary behaviours during adolescence and psychological distress during adulthood. Prev Med 2021 Apr;145:106436. [doi: 10.1016/j.ypmed.2021.106436] [Medline: 33485997]
86. Giurgiu M, Ebner-Priemer UW, Dumuid D. Compositional insights on the association between physical activity and sedentary behavior on momentary mood in daily life. Psychol Sport Exerc 2022 Jan;58:102102. [doi: 10.1016/j.psychsport.2021.102102]
87. Bassett SM, Lupis SB, Gianferante D, Rohleder N, Wolf JM. Sleep quality but not sleep quantity effects on cortisol responses to acute psychosocial stress. Stress 2015;18(6):638-644. [doi: 10.3109/10253890.2015.1087503] [Medline: 26414625]
88. Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]
89. Konjarski M, Murray G, Lee VV, Jackson ML. Reciprocal relationships between daily sleep and mood: a systematic review of naturalistic prospective studies. Sleep Med Rev 2018 Dec;42:47-58. [doi: 10.1016/j.smrv.2018.05.005] [Medline: 30404728]
90. Dunton GF, Liao Y, Intille S, Huh J, Leventhal A. Momentary assessment of contextual influences on affective response during physical activity. Health Psychol 2015 Dec;34(12):1145-1153. [doi: 10.1037/hea0000223] [Medline: 26053885]
91. Ramón-Arbués E, Gea-Caballero V, Granada-López JM, Juárez-Vela R, Pellicer-García B, Antón-Solanas I. The prevalence of depression, anxiety and stress and their associated factors in college students. Int J Environ Res Public Health 2020 Sep 24;17(19):7001. [doi: 10.3390/ijerph17197001] [Medline: 32987932]
92. Amnie AG. An investigation of predictors of self-efficacy to cope with stress and implications for health education practice. Am J Health Educ 2018 May 4;49(3):155-165. [doi: 10.1080/19325037.2018.1431165]
93. Rollo S, Antsygina O, Tremblay MS. The whole day matters: understanding 24-hour movement guideline adherence and relationships with health indicators across the lifespan. J Sport Health Sci 2020 Dec;9(6):493-510. [doi: 10.1016/j.jshs.2020.07.004] [Medline: 32711156]
94. Epel ES, Crosswell AD, Mayer SE, et al. More than a feeling: a unified view of stress measurement for population science. Front Neuroendocrinol 2018 Apr;49:146-169. [doi: 10.1016/j.yfrne.2018.03.001] [Medline: 29551356]
95. Smets E, Rios Velazquez E, Schiavone G, et al. Large-scale wearable data reveal digital phenotypes for daily-life stress detection. NPJ Digit Med 2018;1(1):67. [doi: 10.1038/s41746-018-0074-9] [Medline: 31304344]
96. Alberdi A, Aztiria A, Basarab A. Towards an automatic early stress recognition system for office environments based on multimodal measurements: a review. J Biomed Inform 2016 Feb;59:49-75. [doi: 10.1016/j.jbi.2015.11.007] [Medline: 26621099]

Abbreviations

ICC: intraclass correlation
ilr: isometric log-ratio
LPA: light physical activity
MVPA: moderate-to-vigorous physical activity

XSL•FO
RenderX

PA: physical activity
SB: sedentary behavior

Edited by A Stone; submitted 04.Aug.2025; peer-reviewed by A Machell, YY Chen; revised version received 04.Apr.2026; accepted 14.Apr.2026; published 22.May.2026.

Please cite as:
Bonn K, Wohlfarth D, Timm I, Bender O, Ebner-Priemer UW, Giurgiu M
Associations Between 24-Hour Physical Behavior, Self-Perceived Stress, and Coping Self-Efficacy in Everyday Life: Ambulatory Assessment Study
JMIR Mhealth Uhealth 2026;14:e81502
URL: https://mhealth.jmir.org/2026/1/e81502
doi:10.2196/81502

© Katrin Bonn, Doreen Wohlfarth, Irina Timm, Oliver Bender, Ulrich W Ebner-Priemer, Marco Giurgiu. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 22.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Integration of Ecological Momentary Assessment and Intervention With Wearable-Based Digital Phenotypes to Treat Anxiety in Young Adults: Mixed Methods Experimental Study

Hyunsil Song[1,2], MS; Yoobin Choi[1,2], BA; Jinho Park[3], PhD; Hyunchan Hwang[4], MD, PhD; Jinwoo Kim[1,2], PhD; Chanmi Park[1,2], MA

[1]HAII Corp, Jongno-gu, Republic of Korea

[2]Yonsei University, Seodaemun-gu, Seoul, Republic of Korea

[3]Medical IT Convergence Research Center, Korea Electronics Technology Institute, Seongnam-si, Republic of Korea

[4]Chung-Ang University Hospital, Dongjak-gu, Seoul, Republic of Korea

Corresponding Author:
Chanmi Park, MA
HAII Corp
Rm 501
428 Samil-daero
Jongno-gu, 03140
Republic of Korea
Phone: 82 1028433600
Email: chanmipark@yonsei.ac.kr

Abstract

Background: Young adults report high levels of anxiety, yet their access to timely treatment remains limited. Mobile-based digital therapeutics, particularly ecological momentary assessment and intervention (EMA/EMI), offer a promising approach to deliver context-responsive care. The integration of wearable-derived digital phenotypes with digital therapeutics may enhance the personalization of feedback in such interventions, but empirical evidence from real-world settings remains scarce.

Objective: This study examined the real-world applicability and effectiveness of an EMA/EMI-based digital therapeutic, ANZEILAX-Green (HAII Corp), for use in anxiety reduction in young adults and investigated how personalized feedback generated by combining EMA/EMI data with wearable-based digital phenotypes ameliorates symptoms and promotes self-management and user engagement.

Methods: An 8-week exploratory experimental study was conducted with young adults aged 19-39 years. Participants used a smartphone app in which EMA/EMI modules had been integrated with a smartwatch that continuously collected digital phenotyping data, including heart rate, sleep, and activity. Psychological outcomes were assessed using the Generalized Anxiety Disorder 7-item scale (GAD-7) and Patient Health Questionnaire-9 (PHQ-9) at baseline and weeks 2, 4, 6, and 8, and the Rosenberg Self-Esteem Scale (RSES) at baseline and week 8. Quantitative changes were analyzed using paired-samples *t* tests and repeated-measures ANOVA, and qualitative interviews were thematically analyzed to identify user experience mechanisms and design implications.

Results: In all, 29 participants were included in the final analysis. Significant improvements were observed in their GAD-7 (mean difference –4.79, 95% CI –6.56 to –3.02; Cohen d=1.03; *P*<.001), PHQ-9 (mean difference –3.86, 95% CI –5.79 to –1.93; Cohen d=0.76; *P*<.001), and RSES scores (mean difference +5.21, 95% CI 3.47 to 6.95; Cohen d=1.14; *P*<.001). Consistent improvements in GAD-7 scores were observed across the assessment points, with statistically significant reductions from week 4 onward. Participants demonstrated high adherence, completing 80% of the EMA/EMI prompts. The digital phenotype showed trends of increased deep sleep and reduced mobility, and sleep indicators were significantly correlated with depressive symptoms. Analysis of the interview findings revealed 5 mechanisms: enhanced self-awareness, data-driven reflection, behavioral change, intrinsic motivation, and extension of treatment into daily life.

Conclusions: EMA/EMI-based digital therapeutics integrated with wearable-derived digital phenotypes are feasible and acceptable for use by young adults and were associated with within-group reductions in anxiety, suggesting acceptable efficacy, though this warrants confirmation in a controlled trial. The combination of real-time self-reported assessments of emotional states

XSL•FO
RenderX

and passive physiological monitoring enables personalized feedback suited to users' emotional and physiological states. These findings offer empirical and design insights for automated, personalized mental health intervention systems.

(JMIR Mhealth Uhealth 2026;14:e88919) doi:10.2196/88919

KEYWORDS

mobile health; anxiety disorder; digital phenotypes; ecological momentary assessment; ecological momentary intervention; acceptance and commitment therapy; wearable data

Introduction

Anxiety disorders are a major global mental health concern that affect an estimated 301.39 million people worldwide, accounting for 28.68 million disability-adjusted life years in 2019. Their burden has increased substantially since the COVID-19 pandemic [1,2]. The burden of anxiety disorders is especially pronounced in young adulthood, with the highest prevalence and disability burden observed among individuals aged 20-24 years [3-5]. This age group is especially vulnerable to anxiety due to the pressure of such developmental tasks as academic pursuits, career planning, identity formation, financial independence, and complex social relationships. Thus, timely and accessible early intervention is crucial to reduce delays in care and mitigate the risk of persistent symptoms and long-term functional impairment [6-8], yet treatment use remains remarkably low due to a variety of barriers, including mental health stigma, limited mental health literacy, negative attitudes toward treatment, accessibility constraints, and financial burden [9-11].

Traditional assessments and treatments for anxiety primarily rely upon retrospective self-reports, which fail to capture the dynamic fluctuations of anxiety, which may change rapidly depending on the situation and are vulnerable to recall bias [12-14]. This may cause the aggravation of momentary symptoms and contextual triggers to be overlooked, undermining the accuracy of assessment and the design of personalized intervention.

As a result, digital therapeutics (DTx) have emerged as a promising adjunctive approach to deliver accessible and scalable mental health interventions [15-17]. DTx provide evidence-based therapy at any time and place, decrease financial burdens, and facilitate self-management in daily life [16,18,19]. Given the high engagement with digital technologies among the young, mobile-based DTx may usefully complement traditional mental health care by offering accessible, on-demand therapeutic content, facilitating symptom monitoring, and providing tailored feedback in daily-life contexts [4,20,21]. Numerous studies have reported that mobile-based interventions exerted a positive effect on anxiety by alleviating symptoms [22-24], and DTx hold the potential to resolve issues of cost and accessibility while enhancing users' self-management abilities [25-27].

Of the variety of DTx approaches currently available, ecological momentary assessment (EMA) is particularly promising, as it repetitively samples users' responses in their natural daily environments, thereby precisely capturing the momentary fluctuations and context-specific triggers of anxiety [12,14,28]. The implementation of EMA on mobile-based platforms can facilitate the real-time capture of self-reported psychological states in everyday contexts, support time-stamped data collection, and achieve generally good participant compliance [14,29]. Ecological momentary intervention (EMI) uses data derived through EMA to deliver interventions in real time within the context of daily life [30-32], thereby providing therapy right when symptoms occur or when individuals are most receptive to the therapy content [30-33]. By delivering support in daily-life contexts, EMI may help individuals practice and apply therapeutic skills outside clinical settings, which may support the generalization and maintenance of treatment gains [32,34,35]. Together, EMA and EMI represent a promising approach for overcoming the limitations of traditional anxiety treatments [36,37].

Digital phenotype refers to the continuous traces that individuals generate in their everyday interactions with digital technologies, which may reflect health- and illness-related characteristics in real-world contexts [38,39]. EMA/EMI systems may be further enhanced through their integration with wearable-based digital phenotyping by enabling continuous, passive, and context-sensitive monitoring for more personalized and timely interventions [39-42]. Comprising behavioral, physiological, and environmental data collected via smartphones or wearable devices, a digital phenotype represents a person's real-time digital traces and enables the quantitative characterization of moment-to-moment changes in an individual's state [39,42-45]. It can thus capture natural variations in physiological states through indicators such as heart rate, sleep patterns, and activity levels that may signal fluctuations in mental health and stress-related states, although the capacity of wearables to reliably infer such states in real-world settings remains limited [43,46-48].

Wearable-derived digital phenotypes complement conventional EMA by allowing an individual's emotional and physiological states to be captured separately from repeated self-reports. For example, changes in heart rate variability (HRV), activity levels, or sleep patterns may indicate anxiety-related states in daily life [49,50], and when these physiological patterns are examined together with emotional or situational information reported through EMA, they can provide a more comprehensive understanding of an individual's state [45,50-53].

Additionally, subjective emotional reports and wearable-derived physiological signals may provide complementary rather than redundant information about anxiety-related states [50,54], supporting a more contextualized understanding of individuals' moment-to-moment experiences and enhancing self-reflection in daily life [52,53,55]. Feedback integrating these subjective and physiological data not only provides information but also

XSL•FO
RenderX

serves as a therapeutic tool to boost self-efficacy and treatment adherence [56].

Integrating digital phenotyping into EMA/EMI systems enables more refined mental health assessments by considering subjective experiences and objective physiological indicators together. Such an integrated approach provides a robust foundation for timely personalized interventions grounded in an individual's physiological and behavioral context, ultimately supporting a more accurate understanding of complex mental health states that cannot be captured with data from a single source.

Despite the promise offered by combining subjective self-reports and objective physiological data in this way, research integrating these approaches remains limited, particularly that comprehensively examining the relationships between EMA/EMI data and physiological signals [57]. Accordingly, further investigation is needed regarding the potential contributions of combined digital-phenotype and EMA/EMI systems to the design of automated intervention systems.

Accordingly, we conducted a study using "ANZEILAX" (HAII Corp; a blend of "anxiety" and "relax," reflecting the app's aim of relieving anxiety), a digital therapeutic whose efficacy in treating anxiety disorders has been empirically validated [58]. A randomized controlled trial demonstrated that ANZEILAX significantly alleviated symptoms in patients with generalized anxiety disorder when delivered in combination with usual care, including pharmacotherapy [58]. This study integrated EMA and EMI functions into this proven therapeutic system to assess its real-world applicability and potential for the design of automated intervention systems.

This study explored the real-world applicability of EMA/EMI digital interventions for anxiety reduction among young adults and evaluated their effectiveness. Specifically, it investigated the impact of personalized feedback generated by an EMA/EMI system integrated with a wearable-based digital phenotypic data system on anxiety reduction and self-management capabilities. In addition, this study qualitatively examined the associations of users' experiences with EMA/EMI systems and integrated digital phenotypic feedback with anxiety-related physiological indicators to identify key design features for automated interventions.

Ultimately, this study sought to provide empirical evidence to support the effective implementation of context-responsive digital interventions, and to help advance personalized intervention strategies to improve mental health in young adults.

Methods

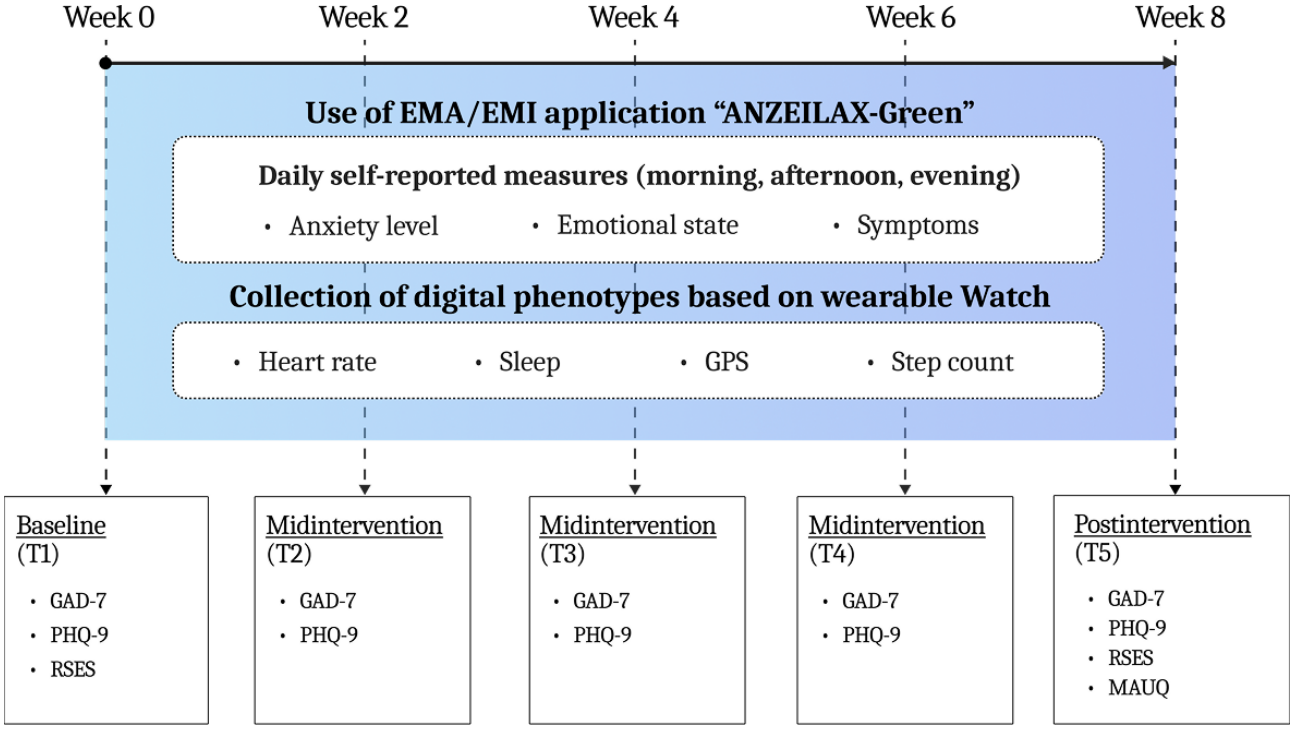
Study Design

This study was a follow-up to a randomized controlled trial of ANZEILAX, a digital therapeutic designed to alleviate anxiety symptoms in young adults. This 8-week exploratory single-group study examined the real-world applicability of interventions that integrate EMA/EMI-based digital therapeutics with wearable-derived digital phenotypic data and evaluated the operation of intervention routines, user acceptance, and potential changes in clinical outcome measures before a controlled clinical evaluation of the EMA/EMI-based DTx.

This study used the ANZEILAX-Green smartphone app (HAII Corp) with integrated EMA/EMI functions in conjunction with a smartwatch-based tool that collects health data to comprehensively explore real-time emotion assessment, momentary intervention delivery, and changes in behavioral and physiological responses in everyday contexts. The study was conducted from September to November 2025. An overview of the study design and procedures is presented in Figure 1.

XSL•FO
RenderX

Figure 1. Study design and ecological momentary assessment and intervention data collection flow. Overview of the 8-week study design, interventions, and collected digital phenotypes. EMA/EMI: ecological momentary assessment and ecological momentary intervention; GAD-7: Generalized Anxiety Disorder 7-item scale; MAUQ: mHealth App Usability Questionnaire; PHQ-9: Patient Health Questionnaire-9; RSES: Rosenberg Self-Esteem Scale.



Ethical Considerations

This study was reviewed and approved by the Institutional Review Board of Yonsei University (IRB number 7001988-202509-HR-2892-03). All participants provided written informed consent after receiving a full explanation of the study purpose and procedures from the research team. Participant confidentiality was maintained throughout the study, with each individual being assigned a numerical code by their order of participation. All collected data were anonymized to ensure participant privacy. Participants who completed the 8-week study were compensated with KRW 130,000 (approximately US $88) for their time and effort, and those who completed the postintervention interview received an additional KRW 20,000 (approximately US $14) as supplementary compensation.

Participants

The participants were Korean young adults aged between 19 and 39 years, with a Beck Anxiety Inventory (BAI) score of 8 to 25 (indicative of mild-to-moderate anxiety), who could communicate in Korean and use smartphones, and included individuals who had not been prescribed psychiatric medication by a psychiatric department within 30 days before the start of the study or those without changes in medication if they had been taking any. Individuals with current or past psychiatric disorders (including schizophrenia, psychosis, bipolar disorder, and epilepsy), brain injury, cognitive impairment, neurological disorders, intellectual disability, or substance or alcohol use disorders, as well as those who had exhibited suicidal intent, suicidal impulses, or self-harming behavior within the past 6 months, were excluded. High-risk individuals with BAI scores of 26 or higher were also excluded.

Procedures

Purposive sampling was used to ensure sufficient representation across age groups and anxiety severity levels (mild and moderate). Participants were approached by a member of the research team who had not been involved in quantitative data collection so as to minimize response bias. A total of 30 participants were enrolled after the research team confirmed their eligibility through online recruitment. After providing informed consent, the participants completed baseline assessments covering their sociodemographic characteristics, anxiety, and depression and then downloaded the ANZEILAX-Green app (HAII Corp). They used the app for 8 weeks in combination with a Galaxy Watch 7 to be worn throughout the study period. On the enrollment day, all participants received hands-on onboarding instructing them in the use of the app and the wearable device and were instructed to contact the research team at any time if they encountered technical difficulties.

Participants completed survey evaluations at weeks 2, 4, and 6, followed by a postintervention assessment at week 8, and those who had volunteered were invited to take part in in-depth interviews following the intervention; all 29 participants who had completed the 8-week intervention voluntarily agreed to participate in the interviews. A research team experienced in qualitative interviews conducted semistructured interviews to explore their experiences with the intervention and elicit suggestions for improvement. The key topics included participants' overall experience with the app, their perceptions of its effectiveness and satisfaction, and any facilitating and hindering factors they had noted. In particular, an interview guide was developed a priori based on the study objectives and reviewed by the research team for content validity, and the

XSL•FO
RenderX

following topics were set in advance: (1) overall experience with the ANZEILAX-Green app and wearable device, (2) perceived effectiveness and satisfaction with the intervention, (3) factors facilitating and hindering engagement, (4) perceptions of the personalized weekly feedback reports, and (5) suggestions for improvement.

All interviews were conducted individually via video call after the intervention period had ended and were audio-recorded with participants' prior written consent. Interviews lasted approximately 30-60 minutes. The interviewer maintained a reflexive stance throughout, bracketing their assumptions about the intervention and remaining open to participants' perspectives. The audio recordings were transcribed verbatim by a research assistant and reviewed for accuracy before analysis. Transcripts were not returned to participants to be checked at this stage. However, participant quotations representative of each theme were verified against the original transcripts to ensure accuracy.

Intervention

Personalized Emotion Regulation App Based on EMA/EMI Systems

The intervention is based on the mobile app ANZEILAX, which received approval from the Korea Food and Drug Safety Administration in April 2025. ANZEILAX incorporates a variety of therapeutic modules, including a self-talk program based on acceptance and commitment therapy (ACT), to support users' psychological flexibility. EMA- and EMI-based functions were integrated into ANZEILAX to evaluate users' emotional states in real time and automatically provide tailored interventions, enabling more dynamic and personalized intervention delivery than the original version. This integrated version, named ANZEILAX-Green, was developed for the Android platform and launched in 2025 by HAII Corp.

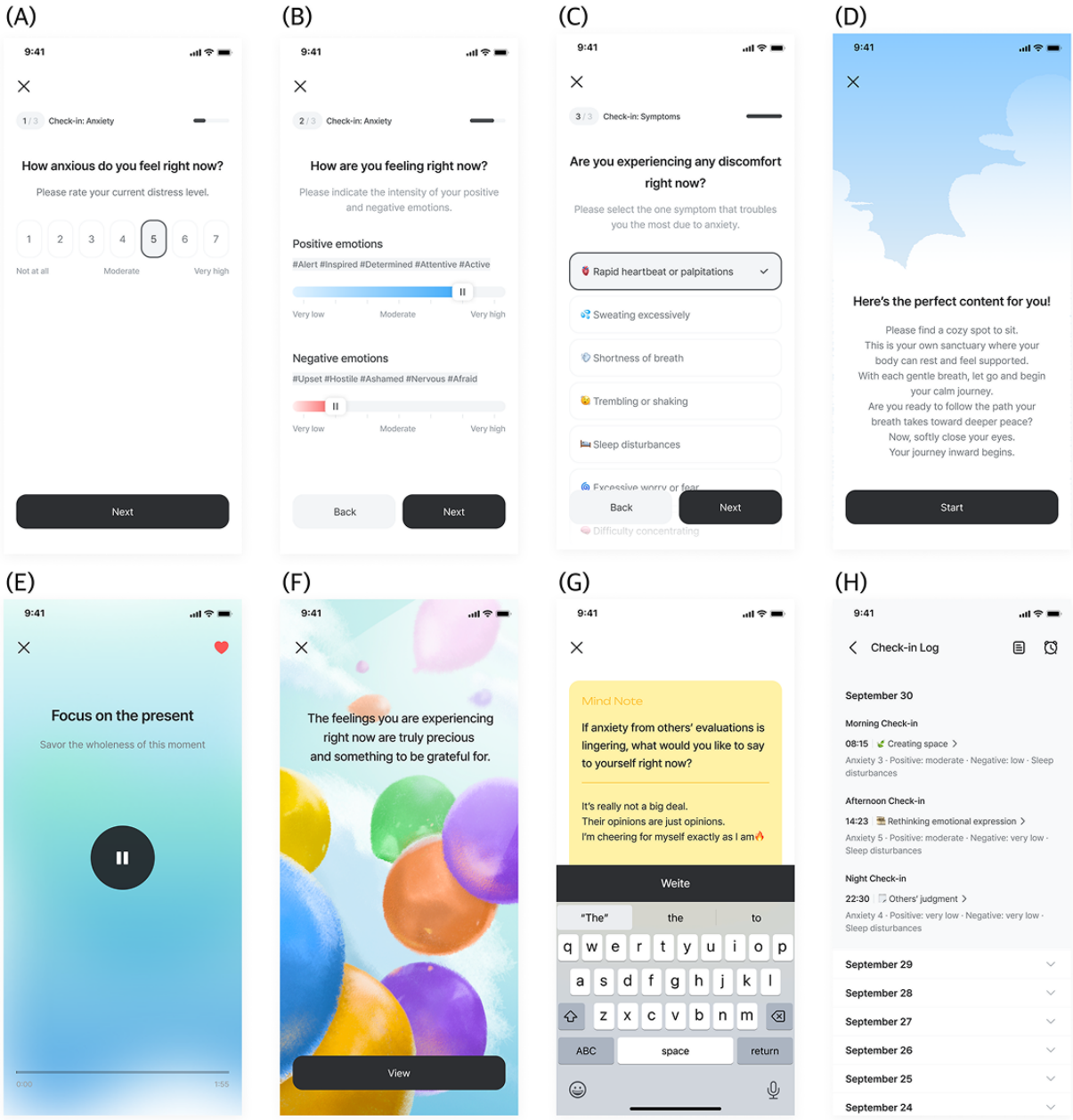
In this study, participants were primarily instructed to use the EMA/EMI functions, while the other therapeutic modules in ANZEILAX were available for optional use. The participants completed EMA questionnaires regarding their anxiety and emotional states through the smartphone app multiple times per day at self-scheduled notification times in the morning, afternoon, and evening. They were required to respond to at least 3 assessment prompts per day over the 8-week period. The questionnaire consisted of items assessing anxiety (1-7 points), positive and negative emotions (each 1-7 points), and symptoms (10 items covering physical, cognitive, and behavioral domains). Positive and negative emotions were assessed using an adapted version of the International Positive and Negative Affect Schedule Short Form (I-PANAS-SF) [59].

The app analyzes users' EMA responses (anxiety, positive and negative emotions, and symptom scores) in real time and supports the automated delivery of EMI content based on ACT principles that induce immediate emotional relief through adaptive delivery based on real-time conditions. The EMA/EMI workflow is as follows: (1) upon EMA input of a user's condition, the system calculates priorities, and (2) automatically selects the intervention type (audio meditation, image, or text) appropriate to the condition and immediately provides EMI content.

The app was designed to strike a balance between automation and personalization. It functions as a state-responsive digital treatment model in which interventions are determined based on a user's actual emotional state, instead of unilateral push interventions. Users thereby learn to recognize their emotional changes in real time and immediately experience the results of interventions, thereby cultivating their emotional self-regulation skills. Detailed screenshots of the app interface are presented in Figure 2.

XSL•FO
RenderX

Figure 2. ANZEILAX-Green app. From left to right, top to bottom: (A-C) ecological momentary assessment (EMA) screens for assessing anxiety, positive emotions, negative emotions, and symptoms. (D) Introduction and entry screen for personalized ecological momentary intervention (EMI) content. (E) Audio-type EMI featuring guided breathing meditation. (F) EMI with images and scripted guidance. (G) EMI encouraging reflection by reading a column and writing thoughts. (H) Tracking screen for EMA and EMI task completion records.



Digital Phenotype Collection App

Digital phenotypes, including heart rate, sleep patterns, and activity levels, were automatically collected in the background through Samsung Health integrated into the Android platform. Physiological data were captured via synchronization between the smartwatch and smartphone. Researchers have examined the validity of consumer smartwatch sensors for heart rate and sleep-stage estimation [60,61]. Nonetheless, such devices remain less accurate than polysomnography for sleep-stage classification, which was taken into consideration in data interpretation.

The digital phenotype indicators collected in this study reflect physiological and behavioral responses known to be associated with anxiety. Heart rate–related indicators are representative physiological markers of autonomic nervous system activity reported to reflect heightened arousal and autonomic imbalance associated with increased anxiety levels. Location (GPS) data and mobility patterns have been shown to capture social avoidance behaviors related to anxiety objectively [62]. Sleep-related indicators are also closely linked to anxiety, as individuals reporting higher anxiety levels tend to exhibit poorer sleep quality [63]. Finally, activity-related indicators, including step count and movement patterns, have been used as behavioral markers reflecting avoidance behaviors and irregular daily activity associated with anxiety [56]. However, because heart rate is also elevated by physical activity, the heart rate indicators were not interpreted as anxiety-specific markers, and the absence

XSL•FO
RenderX

of activity adjustment constitutes a limitation of this approach [64].

The collected data thus included physiological and behavioral indicators strongly associated with anxiety and related mental health conditions and were analyzed alongside quantitative assessment measures over time. Before the analysis, preprocessing procedures, including the removal of missing values and sensor-specific corrections such as segment adjustments, were performed. A detailed list of the collected digital phenotype items is presented in Textbox 1.

Textbox 1. Description of digital phenotypes.

Heart

- Heart rate (HR): average HR collected at 1- to 3-second intervals.

GPS

- Latitude: latitude value collected every 5 minutes.
- Longitude: longitude value collected every 5 minutes.

Sleep

- Light sleep duration: duration of light sleep per sleep event.
- Total sleep duration: total sleep duration per sleep event.
- Rapid eye movement duration: duration of rapid eye movement sleep per sleep event.
- Deep sleep duration: duration of deep sleep per sleep event.

Steps

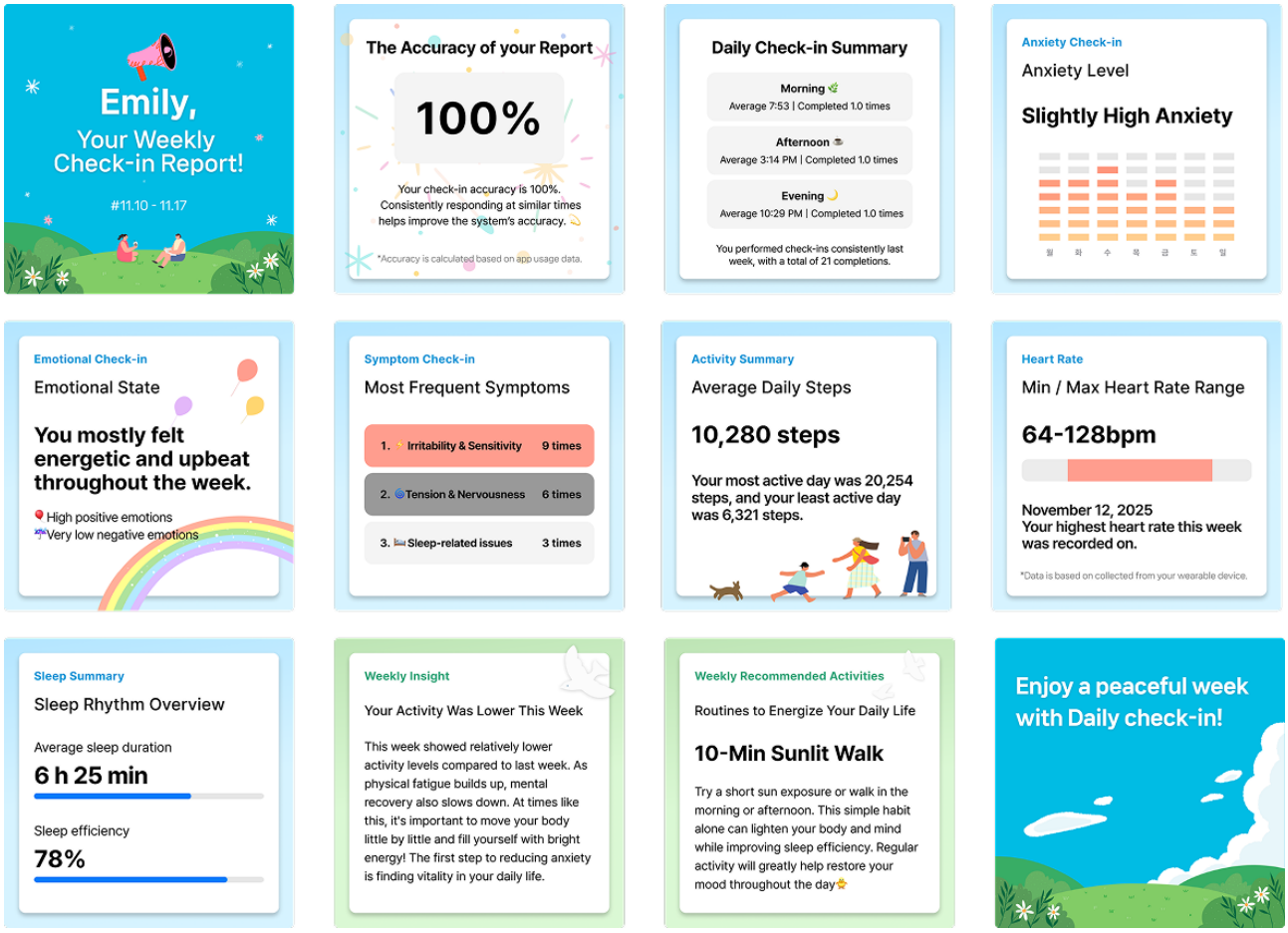
- Step count: number of steps taken across a 5-minute interval.

Personalized Weekly Feedback Report

While the participants engaged with the ANZEILAX-Green, physiological data such as heart rate, sleep, and activity levels were simultaneously collected via the wearable device as digital phenotype data and reported weekly as objective biometrics. The weekly report summarized app usage, key usage patterns, daily anxiety scores, reported emotions and symptoms, activity levels, heart rate, and sleep metrics and provided a personalized summary with recommended activities. These biometrics facilitated the translation of abstract sensations into concrete, interpretable indicators, thus fostering a clear recognition of one's state. The long-term patterns could help users identify personal triggers and protective factors, while regular data reviews could promote sustained attention and reminder effects, which can prompt behavioral change. Feedback based on objective numbers can be an effective therapeutic tool to enhance self-efficacy and treatment adherence (Figure 3).

XSL•FO
RenderX

Figure 3. ANZEILAX-Green weekly report for participants.



Measures

This study used the following measures. The Generalized Anxiety Disorder 7-item scale (GAD-7) was administered to assess anxiety symptom relief [65] and the Patient Health Questionnaire-9 (PHQ-9) to evaluate depressive symptoms, which frequently co-occur with and are closely associated with anxiety [66]. Self-esteem, a psychological construct associated with anxiety levels, was assessed using the Rosenberg Self-Esteem Scale (RSES) [67]. Finally, app usability was evaluated using the mHealth App Usability Questionnaire (MAUQ) [68].

Statistical Analysis

Analyses were conducted using SPSS version 29.0 statistical software (IBM Corp). Outcome measures were analyzed using repeated-measures ANOVA to examine changes over time. When the assumption of sphericity was violated, the Greenhouse-Geisser correction was applied, and pairwise differences across time were examined using post hoc comparisons with Bonferroni adjustment. Associations between digital phenotype indicators and clinical outcome measures were examined using ordinary least squares (OLS) regression implemented in Python (version 3.12; Python Software Foundation).

Qualitative Data Analysis

The qualitative data from the semistructured interviews were analyzed independently by 2 researchers using reflexive thematic analysis following the 6-phase framework of Braun and Clarke [69]. This approach was selected for its flexibility and suitability for exploratory research aimed at identifying experiential patterns across participants. The analysis proceeded inductively without a predetermined coding framework, allowing themes to emerge from the data.

The six phases were as follows: (1) familiarization, in which both researchers read and reread the transcripts and recorded initial impressions; (2) systematic coding, in which meaningful units of data were assigned descriptive codes; (3) searching for themes, in which codes were collated into candidate themes; (4) reviewing themes, in which the coherence and distinctiveness of the themes were evaluated in relation to the full dataset; (5) defining and naming themes, in which each theme was clearly delineated and given a descriptive name; and (6) writing up, in which themes were reported with supporting participant quotations.

To enhance the rigor and dependability of the analysis, each researcher coded the transcripts independently, and discrepancies were resolved through iterative discussion until consensus was reached. Initial coding was compared to ensure consistency. Credibility was supported through the use of verbatim participant quotations illustrating each theme and by grounding interpretations in the data. Transferability was addressed through a thick description of the study context, setting, and participant characteristics. Confirmability was maintained by documenting analytic decisions throughout the process. Reflexivity was

XSL•FO
RenderX

practiced throughout the analysis, with both researchers reflecting on how their professional backgrounds and prior knowledge of the intervention might have shaped their interpretation of the data. All qualitative data were managed and organized using a structured coding matrix in Microsoft Excel.

Results

Demographic Information

A total of 61 individuals were screened for eligibility in this study. Of these, 31 were excluded because their BAI scores fell outside the range of 8-25. The other 30 participants met all the inclusion criteria and were enrolled.

Table 1 provides an overview of the demographics and baseline clinical characteristics. The mean age was 29.83 (SD 5.81) years, with participants evenly distributed between the 19-29 (15, 50%) and 30-39 years age groups (15, 50%). The majority of the participants were women (21/30, 70%), and the remaining 9/30 (30%) were men. Most participants had completed an undergraduate degree or higher, and about half were employed. The reported duration of anxiety symptoms varied, most commonly falling in the range of 1-3 years. The mean BAI score, which was used for screening, was 15.03 (SD 4.12), with 17/30 (57%) participants classified as having mild anxiety and 13/30 (43%) moderate anxiety.

Baseline outcome scores indicated moderate levels of anxiety and depressive symptoms, with mean scores of 10.70 (SD 2.85) on the GAD-7 and 10.50 (SD 4.47) on the PHQ-9. The mean baseline score on the RSES was 29.13 (SD 7.22). Based on the baseline GAD-7 scores, 10/30 (33%) participants were classified as having mild, 17/30 (57%) moderate, and 3/30 (10%) severe anxiety. Based on their PHQ-9 scores, 3/30 (10%) had none-to-minimal, 11/30 (37%) mild, 9/30 (30%) moderate, and 7/30 (23%) moderately severe depressive symptoms, with none classified as severe (Table 1).

XSL•FO
RenderX

Table 1. Demographic and clinical characteristics of the participants (N=30).

Characteristics	Values
Age, mean (SD)	29.83 (5.81)
Age group (years), n (%)	
19-29	15 (50)
30-39	15 (50)
Sex, n (%)	
Male	9 (30)
Female	21 (70)
Education, n (%)	
Doctoral degree	1 (3)
Master's degree	3 (10)
Undergraduate degree	18 (60)
High school degree	8 (27)
Work status, n (%)	
Employed	15 (50)
Not employed	10 (33)
Student	5 (17)
History of anxiety symptoms, n (%)	
≤3 months	5 (17)
6 months to 1 year	5 (17)
1-3 years	11 (37)
3-5 years	2 (7)
>5 years	7 (23)
BAI[a], mean (SD)	15.03 (4.12)
BAI severity, n (%)	
Mild anxiety	17 (57)
Moderate anxiety	13 (43)
GAD-7[b] severity, n (%)	
Mild (5-9)	10 (33)
Moderate (10-14)	17 (57)
Severe (15-21)	3 (10)
PHQ-9[c] severity, n (%)	
None-to-minimal (0-4)	3 (10)
Mild (5-9)	11 (37)
Moderate (10-14)	9 (30)
Moderately severe (15-19)	7 (23)
Severe (20-27)	0 (0)
Outcomes, means (SD)	
Baseline GAD-7	10.70 (2.85)
Baseline PHQ-9	10.50 (4.47)
Baseline RSES[d]	29.13 (7.22)

[a]BAI: Beck Anxiety Inventory.

XSL•FO
RenderX

[b]GAD-7: Generalized Anxiety Disorder 7-item scale.

[c]PHQ-9: Patient Health Questionnaire-9.

[d]RSES: Rosenberg Self-Esteem Scale.

Intervention Effects on Psychological Outcomes

A total of 29 participants were included in the final analysis. One participant was excluded because they did not respond to contact after the baseline assessment.

The main findings are as follows. Compared with the baseline, the GAD-7 and PHQ-9 scores showed significant within-group reductions in anxiety (GAD-7) and depression (PHQ-9), and a significant increase was found in self-esteem (RSES) scores. Please note that the following are within-group changes from a single-arm design without a control group. The mean GAD-7 score decreased from 10.66 (SD 2.89) at baseline to 5.86 (SD 4.34) at the completion of the intervention, with an average change of –4.79 points (95% CI –6.56 to –3.02; Cohen *d*=1.03; *P*<.001). The mean PHQ-9 score decreased from 10.31 (SD 4.43) to 6.45 (SD 4.99) over the same period, with an average change of –3.86 points (95% CI –5.79 to –1.93; Cohen *d*=0.76; *P*<.001), while the mean RSES score increased from 29.34 (SD 7.25) to 34.55 (SD 7.31) over this period, with an average change of 5.21 points (95% CI 3.47 to 6.95; Cohen *d*=1.14; *P*<.001), indicating a significant improvement in self-esteem over the course of the intervention (Table 2).

Table 2. Impact of changes in outcomes from baseline to postintervention (N=29).

Outcomes	Baseline (week 0), mean (SD)	Postintervention (week 8), mean (SD)	Δ Mean (95% CI)	Effect size (Cohen *d*)	*P* value
GAD-7[a]	10.66 (2.89)	5.86 (4.34)	–4.79 (–6.56 to –3.02)	1.03	<.001
PHQ-9[b]	10.31 (4.43)	6.45 (4.99)	–3.86 (–5.79 to –1.93)	0.76	<.001
RSES[c]	29.34 (7.25)	34.55 (7.31)	5.21 (3.47 to 6.95)	1.14	<.001

[a]GAD-7: Generalized Anxiety Disorder 7-item scale.

[b]PHQ-9: Patient Health Questionnaire-9.

[c]RSES: Rosenberg Self-Esteem Scale.

The changes over time also exhibited consistent improvement, as reflected by ongoing reductions in the symptom severity scores. Gradual improvements from the midpoint assessment were noted for anxiety and depression, with a significant reduction observed at week 4 during the early phase of the intervention. The downward trend continued steadily through weeks 6 and 8, with the most pronounced improvement observed postintervention. Anxiety symptoms, measured by GAD-7, showed a steady and sustained decrease from week 4 onward, while depressive symptoms, measured by PHQ-9, followed a similar overall trajectory. Although the PHQ-9 scores in week 8 were numerically slightly higher than in week 6 (not formally tested), they remained significantly lower than baseline at postintervention (Table 3; Figure 4).

Table 3. Changes in GAD-7a and PHQ-9b scores over the intervention period (N=29).

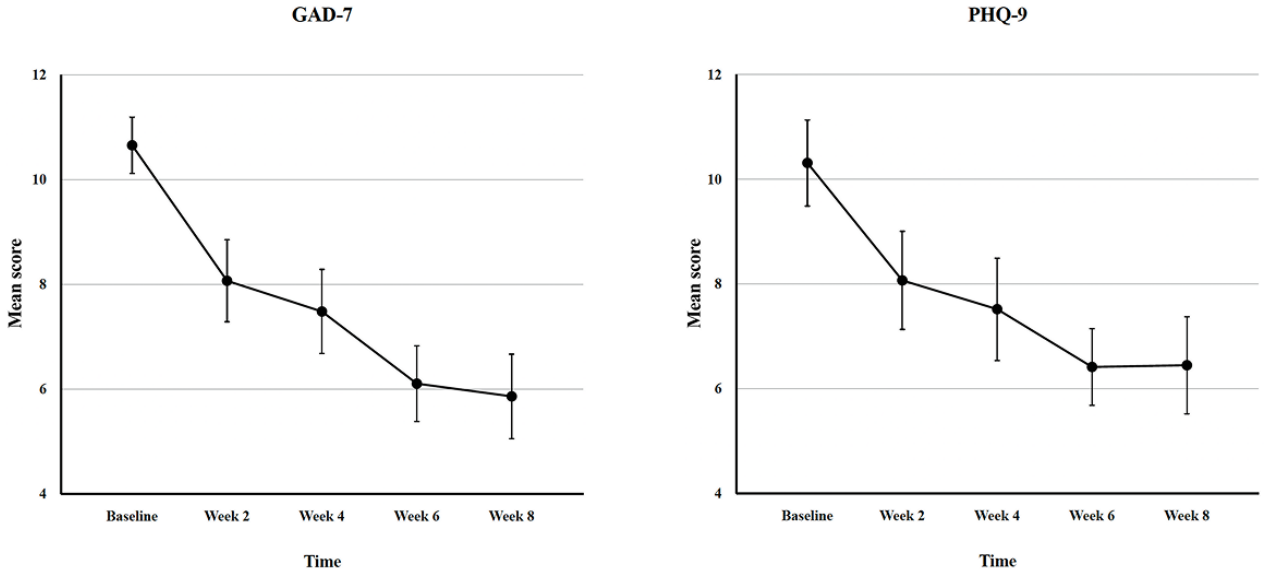
Outcomes	Mean (SD)	Δ Mean (95% CI)	*P* value versus baseline
GAD-7			
Midintervention (week 2)	8.07 (4.22)	−2.59 (−5.20 to 0.03)	.05
Midintervention (week 4)	7.48 (4.32)	−3.17 (−5.74 to −0.60)	.008
Midintervention (week 6)	6.10 (3.90)	−4.55 (−6.83 to −2.27)	<.001
Postintervention (week 8)	5.86 (4.34)	−4.79 (−7.43 to −2.16)	<.001
PHQ-9			
Midintervention (week 2)	8.07 (5.05)	−2.24 (−4.95 to 0.46)	.18
Midintervention (week 4)	7.52 (5.27)	−2.79 (−5.45 to −0.14)	.03
Midintervention (week 6)	6.41 (3.94)	−3.90 (−6.23 to −1.56)	<.001
Postintervention (week 8)	6.45 (4.99)	−3.86 (−6.73 to −0.99)	.003

[a]GAD-7: Generalized Anxiety Disorder 7-item scale.

[b]PHQ-9: Patient Health Questionnaire-9.

XSL•FO
RenderX

Figure 4. Mean GAD-7 (Generalized Anxiety Disorder 7-item scale) and PHQ-9 (Patient Health Questionnaire-9) scores at baseline and weeks 2, 4, 6, and 8 during the 8-week intervention.



These patterns of change suggest consistent improvements in symptoms throughout the intervention, with early changes significantly contributing to the overall effect. The effect sizes and temporal change patterns observed in this study are consistent with those reported in previous ANZEILAX clinical studies [58]. In particular, rapid responses and sustained effects on anxiety symptoms may be of clinical interest, although this requires confirmation by research using controlled designs.

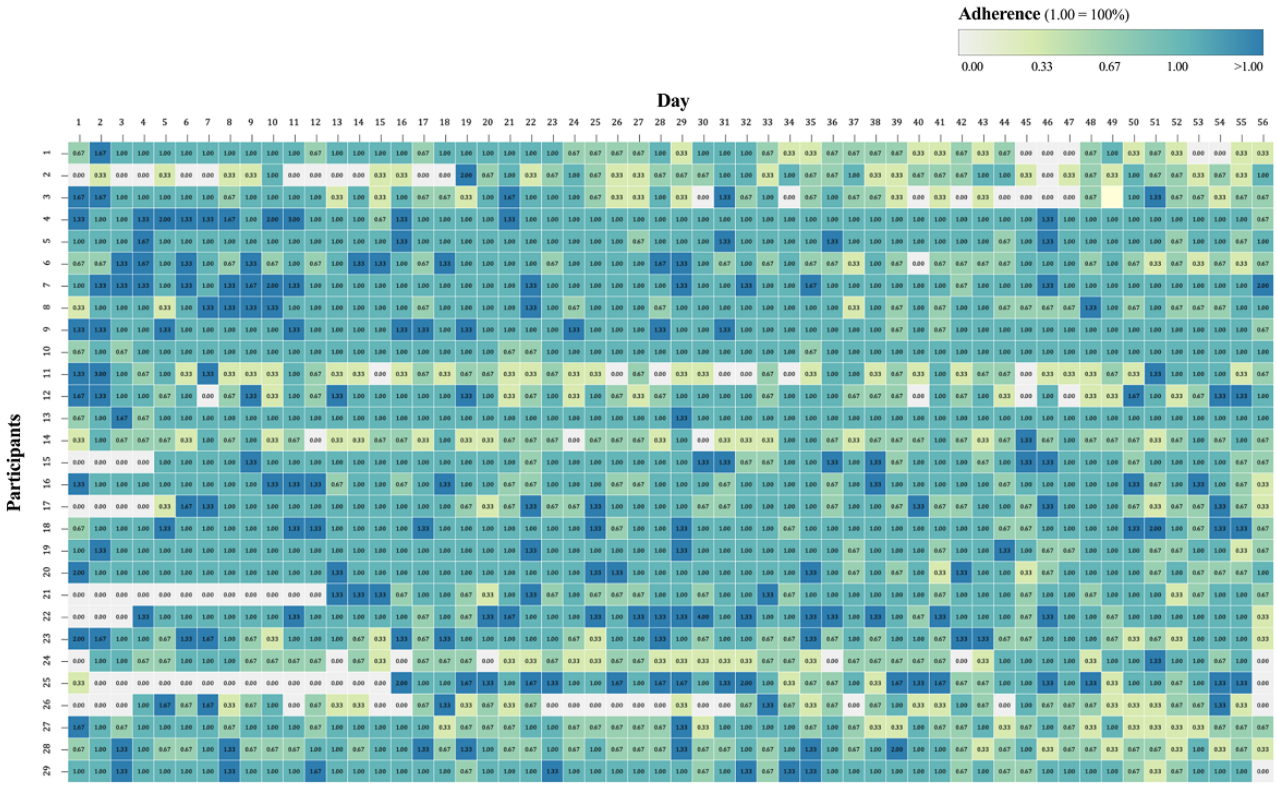
Adherence and Usability

Adherence was calculated based on the protocol requirement of 3 daily EMA/EMI completions over the 8-week intervention period. The participants maintained an average compliance frequency of 80%, thus adhering to the EMA/EMI protocols at a consistently high level.

The heatmap presented in Figure 5 displays the adherence patterns of the 29 participants along the vertical axis over 56 days. Colors closer to green indicate higher adherence, with 1.0 representing full compliance, whereas those closer to gray indicate lower adherence. Blue indicates additional EMA/EMI completions beyond the prescribed 3 daily completions.

XSL•FO
RenderX

Figure 5. Daily ecological momentary assessment and ecological momentary intervention completion patterns among participants during the 8-week intervention.



As the heatmap illustrates, most participants consistently completed 2 to 3 EMA/EMI reports per day throughout the study, thus maintaining stable use patterns. The average adherence to EMA/EMI remained stable throughout the study period, with no sharp declines at any specific time points or weeks, despite variations in completion frequency among participants.

After the 8-week intervention period, participants completed the MAUQ to assess the usability of the app. The MAUQ results showed an average score of 5.03 (SD 1.32) on a 7-point Likert scale. Among the subscales, ease of use and satisfaction scored the highest at 5.28 (SD 1.28), followed by information organization at 4.98 (SD 1.31) and usefulness at 4.84 (SD 1.36). These results indicate that participants rated the overall usability of the app as favorable, while also making suggestions to further optimize the user experience.

Digital Phenotypes

In this study, several digital phenotypes, including heart rate, GPS-derived mobility indicators, sleep patterns, and activity levels measured by step count, were collected and analyzed to comprehensively detect changes in participants' mental health status. The procedures for preprocessing and analysis were applied based on the characteristics of each type of data.

Heart rate data were categorized into 3 ranges: high, ≥100 bpm; medium, 60 bpm to <100 bpm; and low, <60 bpm. The total distance and movement radius were calculated from the GPS latitude and longitude information and normalized from 0 to 1 for comparison between individuals.

To analyze behavioral patterns by time of day, the total distance was classified into daytime and nighttime intervals, with 8 PM as the cutoff. Sleep data were analyzed using Samsung Health's 4-stage classification, including awake, light, REM, and deep, based on sleep records from 6 PM of the preceding day to noon of the following day. Step count data were also normalized from 0 to 1, as with the GPS-derived indicators.

Weekly averages were calculated for data analysis, and trends over the 8-week period were examined using the ordinary least squares method. Statistically significant changes were observed in several physiological and behavioral indicators. The proportion of measurements within the low heart rate range significantly increased over time (β=0.008, *P*=.02), while GPS-based movement data exhibited a significant decrease in distance over the 8-week period (β=–0.3029, *P*=.03). Among sleep-related indicators, deep sleep duration showed a significant increase (β=0.2624, *P*=.02), whereas light sleep duration showed a decreasing trend. Although changes in step count did not reach statistical significance, an upward trend was observed (β=0.1475, *P*=.11). These wearable-derived indicators were not adjusted for physical activity. In addition, heart rate ranges were defined using fixed rather than individualized thresholds, and consumer-wearable sleep-stage estimates are less accurate than polysomnography. These exploratory findings should thus be interpreted with caution.

Analyses of the correlations between digital phenotyping indicators and clinical measures revealed several significant associations. A significant correlation was identified between PHQ-9 scores and deep sleep duration (*r*=0.655, *P*=.04), indicating an association between depressive symptom severity and deep sleep characteristics. In addition, low heart-rate

XSL•FO
RenderX

indicators showed significant positive correlations with both anxiety and depressive symptom severity, as measured by GAD-7 (r=0.578) and PHQ-9 (r=0.836).

Exploratory analyses found that several digital phenotyping indicators showed concurrent changes during the intervention period. Given the small sample and wide CIs, these associations should be interpreted cautiously (Table 4; Figure 6).

Table 4. Correlation analysis between digital phenotypes and GAD-7[a] and PHQ-9[b] scores.

Digital phenotype indicator	β (trend estimate)	SE	*P* value	Correlation with GAD-7 (*r*)	Correlation with PHQ-9 (*r*)
Heart rate					
High	–0.0104	0.0079	0.1909	—[c]	—
Medium	0.0024	0.0062	0.7042	—	—
Low	0.008	0.0033	0.02	0.578[d]	0.836[d]
GPS					
Total distance	–0.3029	0.1361	0.03	−0.004	−0.222
Movement radius	–0.0747	0.1900	0.6942	—	—
Daytime distance	–0.0015	0.0468	0.9745	—	—
Nighttime distance	0.0550	0.0402	0.1713	—	—
Sleep					
Awake	0.0180	0.1979	0.9276	—	—
Light sleep	–0.2553	0.1412	0.0707	—	—
REM[e] sleep	0.1207	0.1716	0.4815	—	—
Deep sleep	0.2624	0.1106	0.02	0.221	0.655[d]
Step					
Step count	0.1475	0.0935	0.11	—	—

[a]GAD-7: Generalized Anxiety Disorder 7-item scale.

[b]PHQ-9: Patient Health Questionnaire-9.

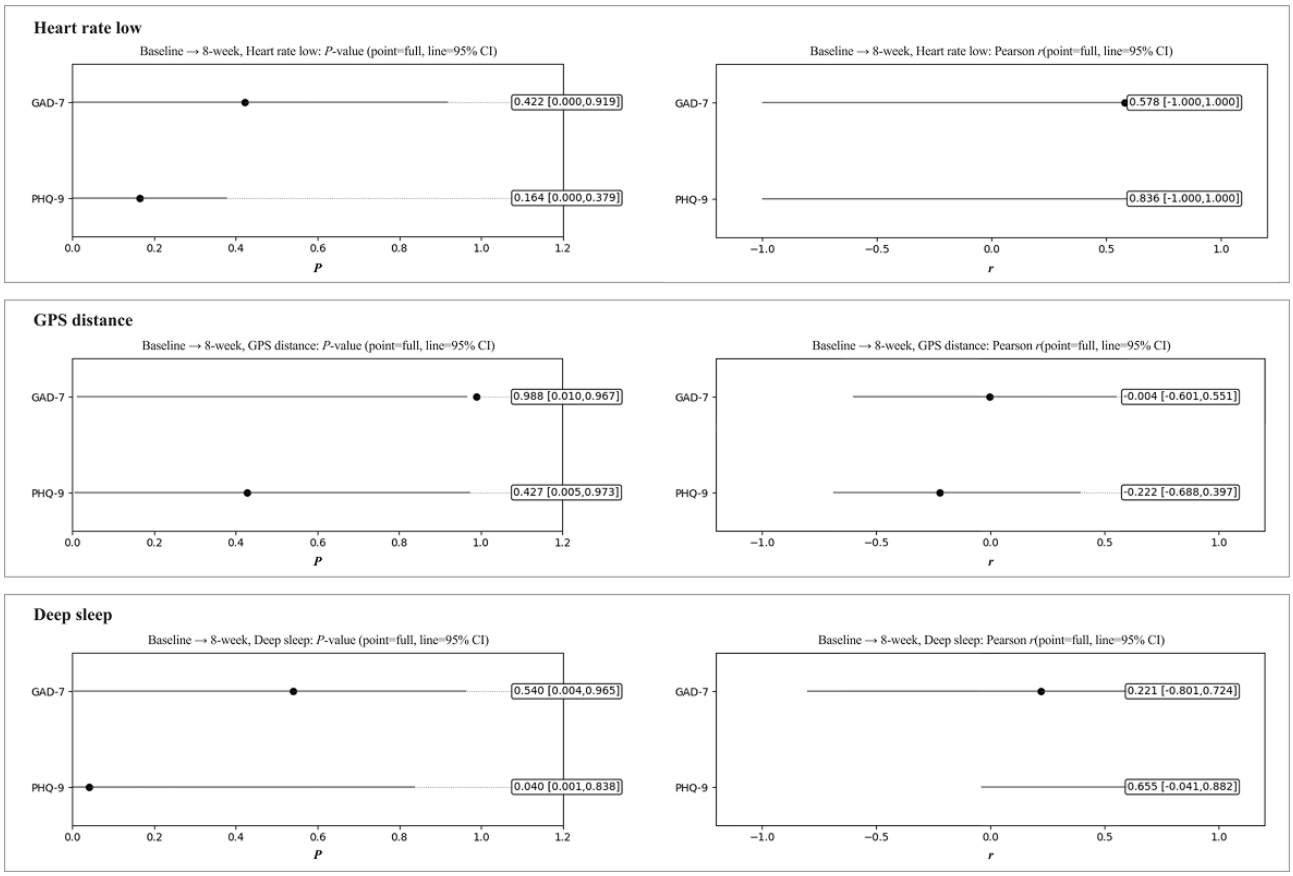
[c]Not applicable.

[d]Statistically significant correlation, P<.05.

[e]REM: rapid eye movement.

XSL•FO
RenderX

Figure 6. Correlation analysis between digital phenotypes and Generalized Anxiety Disorder 7-item scale (GAD-7) and Patient Health Questionnaire-9 (PHQ-9) scores. Pearson correlation coefficients (r), 95% CIs, and *P* values between digital phenotypes (GPS total distance, deep sleep duration, and lowest heart rate) and GAD-7 and PHQ-9 scores measured at baseline and weeks 2, 4, 6, and 8.



User Interviews

Overview

Poststudy interviews were conducted to explore participants' experiences with the ANZEILAX-Green app, its integration with wearable devices, and the weekly feedback reports. The interview findings revealed 5 major themes, indicating that the design features of the system enhanced participants' self-awareness, facilitated data-driven reflection, supported behavioral change, strengthened their intrinsic motivation, and enabled them to apply therapeutic strategies in daily life. At the same time, participants described several challenges, including difficulty in calibrating subjective emotional intensity, the burden of frequent self-monitoring during busy or distressing periods, and heightened awareness of anxiety from repeated self-rating.

Theme 1: Enhanced Self-Awareness

Participants reported that the frequent EMA prompts enabled them to notice emotional and physiological changes that had typically gone unnoticed in daily life. Regular check-ins helped structure their self-reflection across different times of day, encouraging a habit of monitoring their internal state.

> *Being asked several times a day made me pay attention to emotions I would normally ignore... I realized my mood swings were actually more intense than I thought.* [P13]

> *Dividing the day into morning, afternoon, and night helped me reflect on each period individually.* [P17]

Participants also noted increased awareness of the links between physical states and psychological experiences.

> *Seeing my sleep pattern made me realize why I became irritable and sensitive.* [P11]

Several participants initially had difficulty in determining the appropriate intensity of their emotions, reflecting the subjective nature of emotional self-assessment. These findings highlight the importance of clear guidance for subjective emotional self-assessment during repeated EMA monitoring.

Theme 2: Promoting Data-Driven Reflection

Weekly reports presenting their EMA patterns and digital phenotype data allowed participants to reflect more deeply on their emotional states and interpret their emotional experiences in a clearer, more structured way. Rather than merely displaying data, the reports provided objective cues for reviewing daily patterns and emotional fluctuations that helped the participants identify underlying factors linking their emotions and behaviors and offered useful insights for more intentional self-management.

> *Seeing everything in one place helped me understand whether the week was good or stressful, and what triggered the fluctuations.* [P28]

Objective indicators also challenged participants' inaccurate assumptions about their well-being.

XSL•FO
RenderX

> *I thought I lacked sleep, but the report showed that fitness was the bigger issue.* [P23]

These insights helped participants develop more targeted self-management strategies. Several participants reported that the repeated daily check-ins and reflective responses initially felt burdensome, particularly when they were busy or emotionally distressed. These findings suggest that EMA systems should balance repeated self-monitoring with users' perceived burden [70].

Theme 3: Supporting Behavioral Change

Participants stated that personalized EMI content, such as breathing exercises, guided prompts, and reflective messages, motivated immediate action and supported symptom reduction. These interventions were perceived as actionable strategies rather than passive information.

> *The suggestions in the report made me walk more or adjust my sleep patterns the following week.* [P13]

Self-monitoring also encouraged intentional shifts in thinking.

> *When I marked negative emotions (in the app), I tried to reframe the situation more positively.* [P03]

Brief, affirming messages from EMI provided immediate emotional relief.

> *Messages like 'You deserve this happiness' stayed with me all day and helped me feel calmer.* [P05]

Theme 4: Maintaining Intrinsic Motivation

Participants stated that the routine of checking their status was internalized as a daily habit, helping them sustain their engagement beyond external prompts. Use of the app also reinforced self-efficacy and positive self-regard.

> *Checking in became a habit. When I didn't record, something felt missing.* [P13]

> *Managing myself boosted my self-esteem and helped me understand myself better.* [P25]

Personalized and empathetic messages were described as supportive, helping maintain engagement even during periods of distress or positive emotional states.

> *The messages comforted me when I was anxious and recognized my good moments too.* [P21]

Theme 5: Expansion of Active Treatment

Participants described exploring additional therapeutic modules independently, such as self-talk exercises, meditation, and emotional journaling, through which the treatment effects were extended beyond EMA/EMI. Furthermore, skills acquired through the system transferred into real-life contexts.

> *I practiced breathing techniques and self-talk almost every day, and it really supported my mental growth.* [P23]

> *When I felt anxious offline, I recalled solutions I had learned through the app and applied them naturally.* [P29]

These findings suggest that an autonomous, flexible environment enhances user agency and broadens the therapeutic impact of the intervention.

Discussion

Principal Findings

This study examined the real-world applicability and effectiveness of an EMA/EMI-based digital therapeutic, ANZEILAX-Green, in anxiety reduction among young adults and investigated how the personalized feedback generated by the combined EMA/EMI data and wearable-based digital phenotypes supported symptom improvement, self-management, and user engagement. The results of this 8-week exploratory study indicate that personalized feedback integrating the EMA/EMI system with wearable-based digital phenotype data was associated with improvements in both clinical outcomes and participants' capacity for autonomous symptom management, although causality cannot be inferred due to the single-arm design of this study.

The intervention was associated with significant improvements in psychological outcomes. GAD-7 scores showed a marked reduction with a large effect size (Cohen d=1.03) and PHQ-9 scores a medium-to-large effect size (Cohen d=0.76), while RSES scores showed a significant increase with a large effect size (Cohen *d*=1.14). The intervention was thus associated with reductions in anxiety symptoms and concurrent improvements in depressive symptoms and self-esteem, suggesting a broader positive impact on overall mental health. GAD-7 scores began to decline from week 2, showed statistically significant reductions from week 4 onward, and continued to improve through postintervention. These effect sizes and temporal change patterns are generally comparable to those reported elsewhere [58].

The magnitude of the observed effects suggests a meaningful therapeutic signal in real-world use. The relatively large effect sizes may in part reflect the single-arm study design, but they also appear to be related to participants' repeated engagement with the EMA/EMI system in their daily lives. Repeated use of the app likely supported continuous self-monitoring and the timely application of therapeutic strategies. In addition, the study sample primarily consisted of young adults with mild-to-moderate anxiety symptoms and relatively short symptom histories, who might have been more responsive to structured self-management support.

This pattern of early improvement and sustained reduction in symptoms suggests that EMA/EMI-based interventions can help attenuate symptom exacerbation in everyday contexts and support ongoing self-management through the timely delivery of therapeutic content. The EMA/EMI system was designed for frequent use while imposing a low perceived burden through personalized, context-appropriate intervention tasks. This balance might have supported sustained engagement and enabled repeated exposure to therapeutic strategies feasible for use in daily life, which might have contributed to gradual symptom regulation and the reinforcement of adaptive coping processes.

XSL•FO
RenderX

Participants maintained a high compliance rate of 80% over the 8 weeks, consistently adhering to the EMA/EMI protocol. Most participants completed 2 to 3 EMA/EMI per day throughout the study, demonstrating stable usage patterns without sharp declines at any time point. This suggests that EMA/EMI use was naturally integrated into the participants' daily routines. The provision of weekly personalized feedback reports might have further supported sustained engagement by reinforcing a sense of individualized care and ongoing monitoring. Given that user attrition has been identified as a major challenge in digital mental health intervention research, the stable engagement patterns observed in this study demonstrate that the EMA/EMI approach can be an acceptable and sustainable intervention modality for young adults. These improvements are broadly consistent with those obtained from digital therapeutics and app-based interventions that reported reductions in anxiety and depressive symptoms [71,72].

At the same time, although the frequent EMA prompts and repeated self-reporting were generally well tolerated, they were perceived as cognitively or emotionally demanding in certain situations. In particular, during periods of heightened stress, users' receptivity to intensive assessment may vary, highlighting the importance of flexibly balancing assessment frequency with user burden in the design of EMA/EMI systems. The streamlined nature of the EMA/EMI tasks might have contributed to the high levels of engagement observed in this study. However, the therapeutic effects demonstrated in this population with mild-to-moderate anxiety symptoms may not readily generalize to individuals with chronic anxiety or more severe psychiatric conditions, for whom more intensive or clinically supervised interventions may be necessary.

Significant changes in several physiological and behavioral indicators were also observed in the digital phenotype analysis. Deep sleep duration increased significantly while light sleep decreased, and GPS-based movement distance decreased significantly. In addition, the proportion of measurements falling within the low-heart-rate range increased significantly over time, which is to some extent consistent with lower physiological arousal. However, heart rate alone is a weak index of relaxation. HRV, which reflects parasympathetic modulation, would be a more informative physiological marker. Furthermore, these analyses were not adjusted for physical activity. A significant correlation was identified between PHQ-9 and deep sleep duration, indicating a possible association between depressive symptoms and deep sleep. This interpretation assumes reliable wearable sleep-stage detection and, given the small sample, requires cautious reading.

The concurrent increase in deep sleep and reduction in light sleep may indicate an overall improvement in sleep quality, which has been associated with reduced physiological arousal and greater emotional stability. Similarly, the increasing proportion of low heart rate measurements may reflect a more stable pattern of autonomic regulation, consistent with reduced sustained stress or anxiety-related arousal in daily life. Naturalistic studies combining wearable sensing with EMA have similarly examined the physiological correlates of affect and stress in everyday settings, albeit with varying predictive accuracy [42,54,73]. Although the decrease in movement distance may appear negative at first glance, when considered alongside the increasing trend in step count, it suggests that regular activity patterns within primary living areas were established rather than more extensive movement.

The interview analysis yielded 5 major themes in participants' experiences with the intervention: enhanced self-awareness, promotion of data-driven reflection, support for behavioral change, maintenance of intrinsic motivation, and expansion of active treatment. These themes suggest that the integration of EMA/EMI systems with digital phenotype data contributed to meaningful changes in clinical outcomes, behavioral patterns, and participants' capacity for self-management.

The combination of real-time self-reported data and objective physiological indicators helped participants develop a more coherent understanding of their emotional and physical states and appears to have been particularly beneficial for individuals with elevated anxiety, who often experience discrepancies between their physical symptoms and subjective perceptions. By providing objective reference points alongside subjective assessments, the system helped participants recognize patterns that might otherwise have remained unnoticed and supported the development of more targeted self-management strategies. Furthermore, the finding that participants independently explored additional therapeutic modules beyond the core app functions suggests that the intervention successfully fostered user autonomy. The therapeutic benefits were thus not confined to moments of active app use but extended into participants' daily lives through the self-initiated exploration and application of learned strategies.

Design Implications

Overview

The EMA/EMI-based intervention in this study was associated with young adults' engagement- and intervention-related outcomes in real-world contexts. Because the design did not allow the comparison of the conditions with and without wearable data, the added value of the wearable component cannot be isolated and has thus been discussed only in terms of user experience. Three major design implications were identified to inform the development and integration of automated EMA/EMI-based interventions for mental health care.

Temporal Calibration for Personalized Intervention Delivery

An initial adjustment and calibration phase is needed to account for individual differences in emotional expression and the interpretation of EMA response scales. Participants reported that the repeated EMA responses facilitated a clearer recognition of their emotional and behavioral patterns by encouraging reflection on internal states that they might otherwise not have noticed.

During early use, participants experienced uncertainty in translating subjective emotional intensity into numerical scores. Rather than relying on predefined standards, users gradually developed personalized reference points through repeated interaction with the EMA system. Over time, EMA scoring became aligned with individual emotional experiences,

XSL•FO
RenderX

indicating that score interpretation is shaped through use rather than established at baseline.

These findings underscore the importance of incorporating an explicit calibration phase during early system use. By jointly analyzing EMA response patterns and wearable-derived digital phenotype data during this phase, individualized emotional baselines can be established that enable the system to identify deviations from users' typical daily emotional rhythms, allowing personalized interventions to be delivered at moments when support is most needed.

Modality Flexibility in Intervention Design

Participants expressed preferences for a variety of intervention formats depending on their situational context and highlighted the importance of flexibility in intervention delivery. The feedback reports supported their reflection on emotional patterns and symptom fluctuations, while the ability to select from among several intervention formats enabled users to engage with the system in ways that aligned with their daily routines.

Participants also emphasized that the feasibility of intervention formats varied across contexts, with some approaches being perceived as difficult to engage with during ongoing activities. This indicates that the effectiveness of the intervention is closely linked not only to its content but also to users' ability to choose formats that fit their immediate circumstances.

These findings suggest that flexible intervention options enhance users' sense of choice and self-directed engagement, which may contribute to sustained use and therapeutic effectiveness. Accordingly, EMA/EMI systems should support multiple intervention modalities to accommodate situational constraints, as modality flexibility directly affects users' ability to engage with interventions at the moment they are delivered.

Digital Phenotype–Driven Predictive Emotional Support

The participants in this study reported that reviewing their wearable-derived digital phenotype indicators supported self-reflection on emotional fluctuations and facilitated adjustments in their daily behavior. Notably, the impact of digital phenotype data was enhanced when numerical indicators were accompanied by empathic messages based on their current states.

These findings highlight the importance of an integrated design approach in which predictive interventions informed by digital phenotype data are combined with empathic emotional support delivered through EMA/EMI systems. Rather than relying solely upon reactive interventions provided after explicit user input, this approach emphasizes proactive interventions that synthesize actively self-reported data and passively collected digital phenotypes to anticipate and respond to changes in users' emotional states. This proactive, context-sensitive direction aligns with the Just-in-Time Adaptive Intervention (JITAI) framework, which we consider a goal for future development rather than a feature of the present system [32].

In this context, digital phenotype data not only function as stand-alone signals for predictive intervention but become meaningful when coupled with empathic emotional support, which reinforces users' perceptions of being understood and supported by the system.

The participants' qualitative accounts suggest tentative design considerations to support automated interventions that integrate EMA/EMI systems with wearable-derived digital phenotypes, including adaptive calibration, flexible intervention delivery, and the use of digital phenotype data for predictive intervention combined with empathic emotional support, to facilitate context-responsive mental health care.

By leveraging multidimensional data to detect meaningful changes in users' emotional states and deliver personalized interventions, EMA/EMI-based approaches can closely align with users' lived experiences. At the same time, the design of such systems should prioritize the secure handling of sensitive biometric data and the protection of user privacy as fundamental principles to ensure user-centered and sustained implementation.

Limitations

This study has several limitations that should be borne in mind. First, the exploratory single-arm design limits causal inference regarding the observed changes. Without a control group, natural course, regression to the mean, recruitment bias, and expectancy effects cannot be excluded. In addition, because participants could optionally use other therapeutic modules within ANZEILAX-Green, the observed changes cannot be attributed specifically to the EMA/EMI or wearable components.

Second, the wearable-derived indicators were treated as contextual physiological and behavioral signals rather than validated objective measures of anxiety. Therefore, these findings should be interpreted cautiously. Repeated self-monitoring may also have affected participants differently, as a small number of participants reported heightened awareness of anxiety during periods of elevated distress. Future EMA-based interventions should therefore consider individual differences in response to repeated self-assessment and balance monitoring intensity with user burden [74].

Third, although the sample size was sufficient for this exploratory assessment of feasibility and acceptability, it limited the analysis of individual differences and the generalizability of the findings. In addition, because no follow-up assessment was conducted, it remains unknown whether the observed improvements were sustained beyond the intervention period. Future studies should use larger and more diverse samples and include longer follow-up periods to examine the durability of the observed effects.

Conclusions

In conclusion, this study provides preliminary evidence for the feasibility, acceptability, and potential clinical benefits of ANZEILAX-Green, an EMA/EMI-based mobile intervention, in reducing anxiety symptoms among young adults. The integration of EMA/EMI systems with wearable-derived digital phenotype data enables a comprehensive assessment of subjective symptom changes and objective indicators, including physiological and behavioral data, demonstrating how personalized feedback can operate within everyday contexts. This integrative approach extends the clinical applicability of

XSL•FO
RenderX

mobile mental health interventions by allowing the multidimensional emotional changes associated with anxiety symptoms in daily life to be captured. A randomized controlled trial is thus needed to confirm its effectiveness.

Beyond symptom reduction, the findings suggest that EMA/EMI-based interventions integrated with digital phenotyping can support ongoing self-management by mitigating symptom exacerbation in everyday contexts and facilitating the application of therapeutic strategies outside moments of active app use. In addition, the observed clinical effects appear to be supported by an integrated feedback structure that leverages digital phenotype data to enhance the acceptability and enactment of the intervention in daily life.

Furthermore, the findings provide empirical support for the design and advancement of automated digital mental health interventions that reflect individuals' ongoing states and contextual conditions in the feedback they provide. By showing that the integration of EMA/EMI systems with digital phenotyping extends and supports self-awareness, reflection, and self-management, this study highlights a scalable and clinically meaningful direction for future context-responsive digital mental health interventions.

Acknowledgments

The authors gratefully acknowledge all the study participants for their time and dedication to this study. We would like to disclose that generative AI was used for grammatical corrections to enhance the clarity and readability of the manuscript. However, all intellectual contributions, including the study design, analysis, interpretation of results, and overall writing, were entirely conducted by the authors.

Funding

This work was supported by the Technology Innovation Program (project RS-2024-00431485) funded by the Ministry of Trade, Industry and Resources (MOTIR). It was also supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (project NRF-2016R1D1A1B02015987).

Data Availability

The datasets generated or analyzed during this study are not publicly available due to participant confidentiality. A simplified, deidentified dataset can be made available from the corresponding author upon reasonable request.

Authors' Contributions

HS was responsible for conceptualization, methodology, investigation, formal analysis, project administration, and writing the original draft. HH was responsible for conceptualization, methodology, and supervision. CP was responsible for conceptualization, methodology, and writing, review, and editing. JP was responsible for formal analysis. YC was responsible for formal analysis and writing, review and editing. JK was responsible for conceptualization, methodology, funding acquisition, supervision, and writing, review and editing.

Conflicts of Interest

HAII Corp provided research support and contributed to the study conceptualization, trial design, implementation, and manuscript preparation. JK is the chief executive officer of HAII Corp. HS, YC, and CP are employees of HAII Corp.

References

1. Yang X, Fang Y, Chen H, Zhang T, Yin X, Man J, et al. Global, regional and national burden of anxiety disorders from 1990 to 2019: results from the Global Burden of Disease Study 2019. Epidemiol Psychiatr Sci 2021;30:e36. [doi: 10.1017/s2045796021000275]
2. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021;398(10312):1700-1712 [FREE Full text] [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
3. Solmi M, Radua J, Olivola M, Croce E, Soardo L, Salazar de Pablo G, et al. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. Mol Psychiatry 2022;27(1):281-295 [FREE Full text] [doi: 10.1038/s41380-021-01161-7] [Medline: 34079068]
4. Lehtimaki S, Martic J, Wahl B, Foster KT, Schwalbe N. Evidence on digital mental health interventions for adolescents and young people: systematic overview. JMIR Ment Health 2021;8(4):e25847 [FREE Full text] [doi: 10.2196/25847] [Medline: 33913817]
5. Bie F, Yan X, Xing J, Wang L, Xu Y, Wang G, et al. Rising global burden of anxiety disorders among adolescents and young adults: trends, risk factors, and the impact of socioeconomic disparities and COVID-19 from 1990 to 2021. Front Psychiatry 2024;15:1489427 [FREE Full text] [doi: 10.3389/fpsyt.2024.1489427] [Medline: 39691785]

XSL•FO
RenderX

6. Colizzi M, Lasalvia A, Ruggeri M. Prevention and early intervention in youth mental health: is it time for a multidisciplinary and trans-diagnostic model for care? Int J Ment Health Syst 2020;14(1):23 [FREE Full text] [doi: 10.1186/s13033-020-00356-9] [Medline: 32226481]
7. McGorry PD, Mei C, Chanen A, Hodges C, Alvarez-Jimenez M, Killackey E. Designing and scaling up integrated youth mental health care. World Psychiatry 2022;21(1):61-76 [FREE Full text] [doi: 10.1002/wps.20938] [Medline: 35015367]
8. Fusar-Poli P, Correll CU, Arango C, Berk M, Patel V, Ioannidis JPA. Preventive psychiatry: a blueprint for improving the mental health of young people. World Psychiatry 2021;20(2):200-221 [FREE Full text] [doi: 10.1002/wps.20869] [Medline: 34002494]
9. Sickel AE, Seacat JD, Nabors NA. Mental health stigma: impact on mental health treatment attitudes and physical health. J Health Psychol 2019;24(5):586-599. [doi: 10.1177/1359105316681430] [Medline: 27909162]
10. Radez J, Reardon T, Creswell C, Lawrence PJ, Evdoka-Burton G, Waite P. Why do children and adolescents (not) seek and access professional help for their mental health problems? A systematic review of quantitative and qualitative studies. Eur Child Adolesc Psychiatry 2021;30(2):183-211 [FREE Full text] [doi: 10.1007/s00787-019-01469-4] [Medline: 31965309]
11. Aguirre Velasco A, Cruz ISS, Billings J, Jimenez M, Rowe S. What are the barriers, facilitators and interventions targeting help-seeking behaviours for common mental health problems in adolescents? A systematic review. BMC Psychiatry 2020;20(1):293 [FREE Full text] [doi: 10.1186/s12888-020-02659-0] [Medline: 32527236]
12. Russell MA, Gajos JM. Annual research review: ecological momentary assessment studies in child psychology and psychiatry. J Child Psychol Psychiatry 2020;61(3):376-394 [FREE Full text] [doi: 10.1111/jcpp.13204] [Medline: 31997358]
13. Lamers F, Swendsen J, Cui L, Husky M, Johns J, Zipunnikov V, et al. Mood reactivity and affective dynamics in mood and anxiety disorders. J Abnorm Psychol 2018;127(7):659-669. [doi: 10.1037/abn0000378] [Medline: 30335438]
14. Hall M, Scherner PV, Kreidel Y, Rubel JA. A systematic review of momentary assessment designs for mood and anxiety symptoms. Front Psychol 2021;12:642044 [FREE Full text] [doi: 10.3389/fpsyg.2021.642044] [Medline: 34079492]
15. Miao BY, Arneson D, Wang M, Butte AJ. Open challenges in developing digital therapeutics in the United States. PLOS Digit Health 2022;1(1):e0000008 [FREE Full text] [doi: 10.1371/journal.pdig.0000008] [Medline: 36812515]
16. Kim M, Patrick K, Nebeker C, Godino J, Stein S, Klasnja P, et al. The digital therapeutics real-world evidence framework: an approach for guiding evidence-based digital therapeutics design, development, testing, and monitoring. J Med Internet Res 2024;26:e49208 [FREE Full text] [doi: 10.2196/49208] [Medline: 38441954]
17. Ridout SJ, Ridout KK, Lin TY, Campbell CI. Clinical use of mental health digital therapeutics in a large health care delivery system: retrospective patient cohort study and provider survey. JMIR Ment Health 2024;11:e56574-e56574 [FREE Full text] [doi: 10.2196/56574] [Medline: 39356493]
18. Jacobson NC, Quist RE, Lee CM, Marsch LA. Using digital therapeutics to target gaps and failures in traditional mental health and addiction treatments. In: Digital Therapeutics for Mental Health and Addiction. Cambridge, Massachusetts: Academic Press Elsevier; 2023:5-18.
19. Sverdlov O, van Dam J, Hannesdottir K, Thornton-Wells T. Digital therapeutics: an integral component of digital innovation in drug development. Clin Pharmacol Ther 2018;104(1):72-80. [doi: 10.1002/cpt.1036] [Medline: 29377057]
20. Linardon J, Messer M, Goldberg SB, Fuller-Tyszkiewicz M. The efficacy of mindfulness apps on symptoms of depression and anxiety: an updated meta-analysis of randomized controlled trials. Clin Psychol Rev 2024;107:102370 [FREE Full text] [doi: 10.1016/j.cpr.2023.102370] [Medline: 38056219]
21. Buttazzoni A, Brar K, Minaker L. Smartphone-based interventions and internalizing disorders in youth: systematic review and meta-analysis. J Med Internet Res 2021;23(1):e16490 [FREE Full text] [doi: 10.2196/16490] [Medline: 33427682]
22. Domhardt M, Steubl L, Baumeister H. Internet- and mobile-based interventions for mental and somatic conditions in children and adolescents. Z Kinder Jugendpsychiatr Psychother 2020;48(1):33-46 [FREE Full text] [doi: 10.1024/1422-4917/a000625] [Medline: 30422059]
23. Oliveira C, Pereira A, Vagos P, Nóbrega C, Gonçalves J, Afonso B. Effectiveness of mobile app-based psychological interventions for college students: a systematic review of the literature. Front Psychol 2021;12:647606 [FREE Full text] [doi: 10.3389/fpsyg.2021.647606] [Medline: 34045994]
24. Miralles I, Granell C, Díaz-Sanahuja L, Van Woensel W, Bretón-López J, Mira A, et al. Smartphone apps for the treatment of mental disorders: systematic review. JMIR Mhealth Uhealth 2020;8(4):e14897 [FREE Full text] [doi: 10.2196/14897] [Medline: 32238332]
25. Bulaj G, Clark J, Ebrahimi M, Bald E. From precision metapharmacology to patient empowerment: delivery of self-care practices for epilepsy, pain, depression and cancer using digital health technologies. Front Pharmacol 2021;12:612602 [FREE Full text] [doi: 10.3389/fphar.2021.612602] [Medline: 33972825]
26. Balaskas A, Schueller SM, Cox AL, Doherty G. Understanding users' perspectives on mobile apps for anxiety management. Front Digit Health 2022;4:854263 [FREE Full text] [doi: 10.3389/fdgth.2022.854263] [Medline: 36120712]
27. Lau SC, Bhattacharjya S, Fong MW, Nicol GE, Lenze EJ, Baum C, et al. Effectiveness of theory-based digital self-management interventions for improving depression, anxiety, fatigue and self-efficacy in people with neurological disorders: a systematic review and meta-analysis. J Telemed Telecare 2022;28(8):547-558 [FREE Full text] [doi: 10.1177/1357633X20955122] [Medline: 32954920]

XSL•FO
RenderX

28. Heron KE, Everhart RS, McHale SM, Smyth JM. Using mobile-technology-based ecological momentary assessment (EMA) methods with youth: a systematic review and recommendations. J Pediatr Psychol 2017;42(10):1087-1107. [doi: 10.1093/jpepsy/jsx078] [Medline: 28475765]
29. Williams MT, Lewthwaite H, Fraysse F, Gajewska A, Ignatavicius J, Ferrar K. Compliance with mobile ecological momentary assessment of self-reported health-related behaviors and psychological constructs in adults: systematic review and meta-analysis. J Med Internet Res 2021;23(3):e17023 [FREE Full text] [doi: 10.2196/17023] [Medline: 33656451]
30. Yin H, Zhu H, Gu J, Qin H, Ding W, Guo N, et al. Mobile-based ecological momentary assessment and intervention: bibliometric analysis. Front Psychiatry 2024;15:1300739 [FREE Full text] [doi: 10.3389/fpsyt.2024.1300739] [Medline: 38469030]
31. Dao KP, De Cocker KD, Tong HL, Kocaballi AB, Chow C, Laranjo L. Smartphone-delivered ecological momentary interventions based on ecological momentary assessments to promote health behaviors: systematic review and adapted checklist for reporting ecological momentary assessment and intervention studies. JMIR Mhealth Uhealth 2021;9(11):e22890 [FREE Full text] [doi: 10.2196/22890] [Medline: 34806995]
32. von Lützow U, Neuendorf NL, Scherr S. Effectiveness of just-in-time adaptive interventions for improving mental health and psychological well-being: a systematic review and meta-analysis. BMJ Ment Health 2025;28(1):e301641 [FREE Full text] [doi: 10.1136/bmjment-2025-301641] [Medline: 41027677]
33. Myin-Germeys I, Klippel A, Steinhart H, Reininghaus U. Ecological momentary interventions in psychiatry. Curr Opin Psychiatry 2016;29(4):258-263. [doi: 10.1097/YCO.0000000000000255] [Medline: 27153125]
34. Balaskas A, Schueller SM, Cox AL, Doherty G. Ecological momentary interventions for mental health: a scoping review. PLoS One 2021;16(3):e0248152 [FREE Full text] [doi: 10.1371/journal.pone.0248152] [Medline: 33705457]
35. Zainal NH, Newman MG. A randomized controlled trial of a 14-day mindfulness ecological momentary intervention (MEMI) for generalized anxiety disorder. Eur Psychiatry 2023;66(1):e12 [FREE Full text] [doi: 10.1192/j.eurpsy.2023.2] [Medline: 36645098]
36. Kalichman SC, Cherry C, Cain D, Pope H, Kalichman M, Eaton L, et al. Internet-based health information consumer skills intervention for people living with HIV/AIDS. J Consult Clin Psychol 2006;74(3):545-554. [doi: 10.1037/0022-006X.74.3.545] [Medline: 16822111]
37. Blevins CE, Marsh EL, Stein MD, Schatten HT, Abrantes AM. Project CHOICE: choosing healthy options in coping with emotions, an EMA/EMI plus in-person intervention for alcohol use. Subst Abus 2021;42(4):569-576. [doi: 10.1080/08897077.2020.1806182] [Medline: 32870129]
38. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. Nat Biotechnol 2015;33(5):462-463. [doi: 10.1038/nbt.3223] [Medline: 25965751]
39. Bufano P, Laurino M, Said S, Tognetti A, Menicucci D. Digital phenotyping for monitoring mental disorders: systematic review. J Med Internet Res 2023;25:e46778 [FREE Full text] [doi: 10.2196/46778] [Medline: 38090800]
40. Insel TR. Digital phenotyping: technology for a new science of behavior. JAMA 2017;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
41. Zhao X, Lou Z, Shah PT, Wu C, Liu R, Xie W, et al. Integration of multi-modal biosensing approaches for depression: current status, challenges, and future perspectives. Sensors 2025;25(15):4858 [FREE Full text] [doi: 10.3390/s25154858] [Medline: 40808021]
42. Choi A, Ooi A, Lottridge D. Digital phenotyping for stress, anxiety, and mild depression: systematic literature review. JMIR Mhealth Uhealth 2024;12:e40689 [FREE Full text] [doi: 10.2196/40689] [Medline: 38780995]
43. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. NPJ Digit Med 2019;2(1):88 [FREE Full text] [doi: 10.1038/s41746-019-0166-1] [Medline: 31508498]
44. Dlima SD, Shevade S, Menezes SR, Ganju A. Digital phenotyping in health using machine learning approaches: scoping review. JMIR Bioinform Biotechnol 2022;3(1):e39618 [FREE Full text] [doi: 10.2196/39618] [Medline: 38935947]
45. Jung HW, Kim DY, Lee I, Kim O, Lee S, Lee S, et al. Key features of digital phenotyping for monitoring mental disorders: systematic review. J Med Internet Res 2025;27:e77331-e77331 [FREE Full text] [doi: 10.2196/77331] [Medline: 41191793]
46. Hsin H, Fromer M, Peterson B, Walter C, Fleck M, Campbell A, et al. Transforming psychiatry into data-driven medicine with digital measurement tools. NPJ Digit Med 2018;1(1):37 [FREE Full text] [doi: 10.1038/s41746-018-0046-0] [Medline: 31304319]
47. Place S, Blanch-Hartigan D, Rubin C, Gorrostieta C, Mead C, Kane J, Pentland, et al. Behavioral indicators on a mobile sensing platform predict clinically validated psychiatric symptoms of mood and anxiety disorders. J Med Internet Res 2017;19(3):e75 [FREE Full text] [doi: 10.2196/jmir.6678] [Medline: 28302595]
48. Chakrabarti S, Biswas N, Jones LD, Kesari S, Ashili S. Smart consumer wearables as digital diagnostic tools: a review. Diagnostics (Basel) 2022;12(9):2110 [FREE Full text] [doi: 10.3390/diagnostics12092110] [Medline: 36140511]
49. Abd-Alrazaq A, AlSaad R, Aziz S, Ahmed A, Denecke K, Househ M, et al. Wearable artificial intelligence for anxiety and depression: scoping review. J Med Internet Res 2023;25:e42672 [FREE Full text] [doi: 10.2196/42672] [Medline: 36656625]
50. Jo YT, Lee SW, Park S, Lee J. Association between heart rate variability metrics from a smartwatch and self-reported depression and anxiety symptoms: a four-week longitudinal study. Front Psychiatry 2024;15:1371946 [FREE Full text] [doi: 10.3389/fpsyt.2024.1371946] [Medline: 38881544]

XSL•FO
RenderX

51. Unzueta Saavedra J, Deaso EA, Austin M, Cadavid L, Kraff R, Knowles EEM. Smartphone ecological momentary assessment and wearable activity tracking in pediatric depression: cohort study. JMIR Form Res 2025;9:e66187-e66187 [FREE Full text] [doi: 10.2196/66187] [Medline: 40498668]
52. Lim J, Koh Y, Kim A, Lee U. Exploring context-aware mental health self-tracking using multimodal smart speakers in home environments. 2024 Presented at: 2024 CHI Conference on Human Factors in Computing Systems. CHI '24; May 11-16; Honolulu, HI, USA p. 1-18. [doi: 10.1145/3613904.3642846]
53. Riisager LG, Moeller SB, Larsen JE, Christiansen TB, Aagaard J, Huniche L. Use of a wearable self-tracking instrument by refugees with complex posttraumatic stress disorder: a qualitative study of psychotherapeutic mediation and engagement. JMIR Form Res 2025;9:e70511-e70511 [FREE Full text] [doi: 10.2196/70511] [Medline: 40635138]
54. Tutunji R, Kogias N, Kapteijns B, Krentz M, Krause F, Vassena E, et al. Detecting prolonged stress in real life using wearable biosensors and ecological momentary assessments: naturalistic experimental study. J Med Internet Res 2023;25:e39995 [FREE Full text] [doi: 10.2196/39995] [Medline: 37856180]
55. Neupane S, Saha M, Ali N. Momentary stressor logging and reflective visualizations: implications for stress management with wearables. 2024 Presented at: 2024 CHI Conference on Human Factors in Computing Systems. CHI '24; May 11-16; Honolulu, HI, USA p. 1-19. [doi: 10.1145/3613904.3642662]
56. Jacobson NC, Feng B. Digital phenotyping of generalized anxiety disorder: using artificial intelligence to accurately predict symptom severity using wearable sensors in daily life. Transl Psychiatry 2022;12(1):336 [FREE Full text] [doi: 10.1038/s41398-022-02038-1] [Medline: 35977932]
57. Dobson R, Li LL, Garner K, Tane T, McCool J, Whittaker R. The use of sensors to detect anxiety for in-the-moment intervention: scoping review. JMIR Ment Health 2023;10:e42611 [FREE Full text] [doi: 10.2196/42611] [Medline: 36729590]
58. Park C, Song H, Kim H, Kim E, Jo H, Kim J, et al. Efficacy of a smartphone-based digital therapeutic (Anzeilax) in generalized anxiety disorder: randomized controlled trial. J Med Internet Res 2025;27:e69981 [FREE Full text] [doi: 10.2196/69981] [Medline: 41086425]
59. Thompson ER. Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). J Cross-Cult Psychol 2007;38(2):227-242. [doi: 10.1177/0022022106297301]
60. Lee T, Cho Y, Cha KS, Jung J, Cho J, Kim H, et al. Accuracy of 11 wearable, nearable, and airable consumer sleep trackers: prospective multicenter validation study. JMIR Mhealth Uhealth 2023;11:e50983 [FREE Full text] [doi: 10.2196/50983] [Medline: 37917155]
61. Sarhaddi F, Kazemi K, Azimi I, Cao R, Niela-Vilén H, Axelin A, et al. A comprehensive accuracy assessment of Samsung smartwatch heart rate and heart rate variability. PLoS One 2022;17(12):e0268361 [FREE Full text] [doi: 10.1371/journal.pone.0268361] [Medline: 36480505]
62. Cai L, Boukhechba M, Wu C. State affect recognition using smartphone sensing data. New York, NY, United States: Association for Computing Machinery; 2018 Presented at: 2018 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies; 2018 Sep 26-28; Washington, DC p. 120-125. [doi: 10.1145/3278576.3284386]
63. Lau Y, Chemas N, Ajeet Gokani H, Morrell R, Phannarus H, Cooper C, et al. Association between digital biomarkers of health and anxiety: systematic review and meta-analysis. J Med Internet Res 2026;28:e73812 [FREE Full text] [doi: 10.2196/73812] [Medline: 41802227]
64. Chalmers JA, Quintana DS, Abbott MJ, Kemp AH. Anxiety disorders are associated with reduced heart rate variability: a meta-analysis. Front Psychiatry 2014;5:80 [FREE Full text] [doi: 10.3389/fpsyt.2014.00080] [Medline: 25071612]
65. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
66. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
67. Rosenberg M. Society and the Adolescent Self-Image. Princeton, NJ: Princeton University Press; 1965.
68. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
69. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
70. Hartson KR, Huntington-Moskos L, Sears CG, Genova G, Mathis C, Ford W, et al. Use of electronic ecological momentary assessment methodologies in physical activity, sedentary behavior, and sleep research in young adults: systematic review. J Med Internet Res 2023;25:e46783 [FREE Full text] [doi: 10.2196/46783] [Medline: 37384367]
71. Linardon J, Torous J, Firth J, Cuijpers P, Messer M, Fuller-Tyszkiewicz M. Current evidence on the efficacy of mental health smartphone apps for symptoms of depression and anxiety. A meta-analysis of 176 randomized controlled trials. World Psychiatry 2024;23(1):139-149 [FREE Full text] [doi: 10.1002/wps.21183] [Medline: 38214614]
72. Linardon J, Anderson C, Messer M, Liu C, Torous J. Transdiagnostic-focused apps for depression and anxiety: a meta-analysis. NPJ Digit Med 2025;8(1):443 [FREE Full text] [doi: 10.1038/s41746-025-01860-3] [Medline: 40664815]

XSL•FO
RenderX

73. Abd-Alrazaq A, AlSaad R, Harfouche M, Aziz S, Ahmed A, Damseh R, et al. Wearable artificial intelligence for detecting anxiety: systematic review and meta-analysis. J Med Internet Res 2023;25:e48754 [FREE Full text] [doi: 10.2196/48754] [Medline: 37938883]
74. van Genugten CR, Thong MSY, van Ballegooijen W, Kleiboer AM, Spruijt-Metz D, Smit AC, et al. Beyond the current state of just-in-time adaptive interventions in mental health: a qualitative systematic review. Front Digit Health 2025;7:1460167 [FREE Full text] [doi: 10.3389/fdgth.2025.1460167] [Medline: 39935463]

Abbreviations

ACT: acceptance and commitment therapy
BAI: Beck Anxiety Inventory
DTx: digital therapeutics
EMA: ecological momentary assessment
EMA/EMI: ecological momentary assessment and ecological momentary intervention
EMI: ecological momentary intervention
GAD-7: Generalized Anxiety Disorder 7-item scale
HRV: heart rate variability
I-PANAS-SF: International Positive and Negative Affect Schedule Short Form
JITAI: Just-in-Time Adaptive Intervention
MAUQ: mHealth App Usability Questionnaire
OLS: ordinary least squares
PHQ-9: Patient Health Questionnaire-9
RSES: Rosenberg Self-Esteem Scale

Edited by M Giurgiu; submitted 06.Dec.2025; peer-reviewed by H Yin, LAK Wiese, W Kamphuis; comments to author 31.May.2026; accepted 30.Jul.2026; published 08.Oct.2026.

Please cite as:
Song H, Choi Y, Park J, Hwang H, Kim J, Park C
Integration of Ecological Momentary Assessment and Intervention With Wearable-Based Digital Phenotypes to Treat Anxiety in Young Adults: Mixed Methods Experimental Study
JMIR Mhealth Uhealth 2026;14:e88919
URL: https://mhealth.jmir.org/2026/1/e88919
doi:10.2196/88919
PMID:

©Hyunsil Song, Yoobin Choi, Jinho Park, Hyunchan Hwang, Jinwoo Kim, Chanmi Park. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 08.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring College Students' Compliance During a 60-Day Alcohol-Related Ecological Momentary Assessment Using Location- and Time-Based Questions: Prospective Observational Study

Lucy E Napper[1], PhD; Bilal Khan[2], PhD; Gavin Q Fox[1], MS; Mohamed K Saad[2], PhD; Laura C Wolter[1], PhD; Dennis E McChargue[3], PhD

[1]Department of Psychology, Health, Medicine and Society Program, College of Arts and Sciences, Lehigh University, Bethlehem, PA, United States
[2]College of Health, Lehigh University, Bethlehem, PA, United States
[3]Department of Psychology, University of Nebraska-Lincoln, Lincoln, NE, United States

Corresponding Author:
Lucy E Napper, PhD
Department of Psychology, Health, Medicine and Society Program
College of Arts and Sciences
Lehigh University
17 Memorial Drive East
Bethlehem, PA, 18015
United States
Phone: 1 610 758 4689
Email: lun214@lehigh.edu

Abstract

Background: Ecological momentary assessment (EMA) methods can provide assessment of alcohol-related beliefs and behaviors when people are in their natural environments. An increasingly common EMA approach involves the use of passive sensors (eg, continuous sharing of Bluetooth and GPS cell phone data) to collect rich data and trigger alcohol assessment in specific contexts. While collecting real-time assessments allows researchers to explore contextual factors impacting alcohol use, there is a lack of research examining compliance rates to alcohol-related EMAs that use daily questions for study periods over 2 weeks. EMAs over longer periods allow researchers to assess meaningful changes over time and sample low-base-rate events. Further, there is limited research exploring how EMA compliance is influenced by individual-level factors (eg, privacy concerns and alcohol use). Identifying factors associated with compliance can help inform EMA protocol design.

Objective: This study had 2 primary goals. First, we add to the literature by examining the feasibility and acceptability of a 60-day college student alcohol-EMA that uses phone-based, passive sensors. Second, we examined whether compliance was associated with event-level alcohol use, privacy concerns, and types of questions asked (ie, questions triggered by time or GPS location).

Methods: College students (N=68) who reported regular alcohol use completed the 60-day EMA study. At baseline, participants completed an online survey that included an assessment of privacy concerns. During the study period, participants completed up to 15 questions a day, including evening and morning self-reports of alcohol use. Feasibility was assessed based on enrollment, retention, and compliance rates. Acceptability was examined at the end of the study period using open- and close-answered questions on perceptions of the EMA experience.

Results: The findings supported the feasibility of the 60-day EMA with enrollment exceeding benchmarks and high compliance across the study period (average compliance=90%). Overall, students found the protocol acceptable with 88% (58/66) indicating that they would be willing to complete another EMA study. Participants' compliance was not associated with alcohol use at the event-level (odds ratio [OR] 0.94, 95% CI 0.80-1.10) or at the trait level (OR 0.92, 95% CI 0.73-1.17). Participants' compliance was greater for time-trigger assessment than GPS-triggered questions (P<.001). Privacy concerns were not associated with reduced responding.

Conclusions: This study has implications for those designing alcohol EMAs for college populations that involve the use of passive sensors, including recommendations for balancing predictable timing of questions and variability in question content.

XSL•FO
RenderX

(JMIR Mhealth Uhealth 2026;14:e81713) doi:10.2196/81713

KEYWORDS

Alcohol Consumption; Compliance; Feasibility; Acceptability; Ecological momentary assessment; College; Mobile sensors; Privacy

Introduction

Background

Heavy alcohol use is an issue of concern among college students, with 21.9% of students reporting binge drinking [1]. Alcohol use is associated with a range of negative consequences both for students who drink (eg, unintentional injuries, problems with friends, and interference with schoolwork) and the broader community (eg, damage to property and interrupted sleep) [2,3]. Ecological momentary assessments (EMAs) offer one approach to understand proximal predictors of high-risk drinking behaviors, test theoretical models of alcohol use, and inform the development of just-in-time interventions targeting heavy alcohol use.

EMAs involve real-time assessment of beliefs and behaviors as participants go through their daily lives (ie, in their natural environment) [4]. EMAs are a broad category of approaches that vary in format (eg, text messaging–based vs app-based). In comparison to traditional survey methods assessing retrospective reports of behaviors and beliefs at a single time point, EMAs have a number of advantages. For instance, EMAs increase ecological validity and allow researchers to study dynamic processes that influence drinking, the influence of context on alcohol beliefs and behaviors, and temporal associations between constructs of interest [4,5]. EMAs are well suited to study alcohol use among college students because drinking often happens during discrete events [5]. Indeed, there is a growing body of alcohol-focused research using EMAs to study craving [6], drinking motives and decisions [7-9], negative affect or anxiety [10,11], impact of exposure to alcohol marketing [12], and social context [13,14] in college populations.

In order for EMAs to produce valid and useful data, participant compliance is essential. Work on alcohol-related EMAs among college students has produced compliance rates between 63.8% [6] and 96.8% [8]. There is a growing body of work, including meta-analyses, examining factors that might explain variation in EMA protocol compliance across different research topics, protocols, and samples [15-18]. One factor that might impact compliance is the length of the protocol. In the context of noncollege populations, alcohol-related EMAs over short periods (2 to 4 weeks) have generally been found to be acceptable and feasible, with participants describing EMAs as easy to complete and nondisruptive [19-21]. In contrast, there are mixed findings on the feasibility of alcohol-related EMAs over longer periods. Sokolovsky et al [22] found a steady increase in missed questions across a 56-day EMA divided into 2 bursts. Similarly, a 14-week EMA study found compliance rates significantly dropped over time (93% to 39%) [23]. While another college-based study found less pronounced reductions (90% to 70%) [24]. These inconsistencies in the college alcohol literature suggest that additional research exploring protocols greater than 14 days would be beneficial. Furthermore, Wrzus and Neubauer's [16] recent review of EMAs did not identify a single study that used more than 10 prompts per day over a period of greater than 14 days, highlighting the need for more studies of this type.

In addition to length of data collection, other aspects of an alcohol-related EMA protocol may impact compliance. For instance, Hultgren et al [25] examined compliance to both daily questions administered at set times and event-contingent momentary assessments initiated by participants on drinking occasions. They demonstrated that compliance was greater for regular daily questions than more irregular, self-initiated EMA questions. Differences in compliance might reflect the ease of remembering to respond to questions at set times or challenges to completing questions once alcohol use is initiated. Indeed, event-contingent momentary assessments were less likely to be completed when students were drinking more than usual for them. Similarly, Suffoletto et al [23] note that intoxication negatively impacts completion of reaction time–based EMA tasks. However, the relationship between alcohol use and EMA compliance remains unclear. For instance, other research suggests that neither same-day or past year alcohol use are associated with EMA compliance [24,26].

Other individual-level predictors may be important for understanding protocol compliance in EMA studies, but this type of research to date has been limited [26]. In a college-based alcohol-related EMA, Howard and Lamb [24] explored demographic, personality, and well-being variables and found no associations between these variables and compliance. In non-EMA survey studies, one factor that may be related to missing data is participants' privacy concerns [27]. These concerns might be particularly relevant to understanding compliance in the context of EMA studies examining sensitive topics, such as substance use, or for EMA approaches perceived as more invasive. For example, GPS and Bluetooth data can be integrated into EMAs in order to address novel questions about the environmental and social context of drinking and help identify risk factors for alcohol use beyond individual-level factors (eg, mood). In addition, integrating sensor data allows questions to be triggered based on, for example, someone's geographic location. While incorporating mobile sensors into EMAs is not common, representing about 5% of EMA studies [15], this approach is likely to become increasingly popular given the opportunities that it provides for studying contextual factors. That said, it is possible that both GPS tracking and triggering questions based on a participant's geographic location may raise privacy concerns [28,29]. Therefore, research is needed to determine if privacy concerns impact compliance in these types of EMA studies.

While research has examined the acceptability of traditional EMA techniques, there is less known about participants' perceptions of EMAs that include passive sensing methods that

XSL•FO
RenderX

occur automatically and continuously [30] and how individual-level factors impact compliance for this approach. Further, based on a recent review of EMA including geographic technology [31], no studies to date have specifically explored the use of EMAs that incorporate GPS-based sensors when examining college student alcohol use. Therefore, this study extends past research by examining the feasibility and acceptability of a 60-day, alcohol-related EMA that required participants to share Bluetooth, motion and fitness sensor, and GPS cell phone data.

Current Study

This study has 2 primary goals. First, we aim to assess the feasibility (eg, based on enrollment, retention, and overall compliance to daily questions) and acceptability (eg, participants satisfaction with the protocol) of a 60-day EMA alcohol-related protocol. While EMAs have the potential to advance our understanding of the proximal predictors of alcohol-related outcomes, they will only do so if participants are willing to enroll in and complete EMA protocols. This study adds to the limited research using daily surveys over periods greater than 14 days. Testing the practicality of these types of longer EMA designs is important for a number of reasons. For example, longer EMAs are important both for understanding alcohol-related risk factors that unfold and meaningfully change over weeks rather than days (eg, shifts in mood trajectories, coping patterns, and intervention response). As such, longer assessment windows are critical to observe within-person change instead of short-term variability. Moreover, longer assessment periods increase the chance of sampling low-base-rate events (eg, rare stressors and high-intensity drinking episodes) that are essential for testing risk models and just-in-time intervention assumptions. In the context of college drinking, this study's focus on a 60-day time frame allowed us to test feasibility and collect rich data across different phases of the semester including multiple weekends, midterms, and Spring break.

The second aim of the paper is to explore whether individual-level and EMA-design elements impacted compliance. Identifying factors that could impact compliance is important for understanding potential data biases and for improving EMA protocol design. This study adds to the literature by examining whether compliance rates were related to (1) students' alcohol use, (2) students' privacy concerns, and (3) EMA question type (ie, daily time-based question vs geolocation-triggered question). Based on past research [25], we expected compliance to be greater for time-based vs geolocation-based questions. Given the lack of data on compliance and privacy concerns, and the mixed findings on the relationship between alcohol use and compliance, these analyses were exploratory.

Methods

Procedures

A random sample of students from a university in the Northeastern United States (N=399) were invited to participate via email that provided a link to an online screening survey. In addition, participants were given the option to invite friends to participate in the study either by providing the researcher with their friend's email address or by the student sharing a link to the study. This approach was used to increase the likelihood of phones being collocated in order to maximize the amount of Bluetooth data collected. Eligibility was determined based on cell phone ownership, being 18 years of age or older, not planning to study abroad during the study period (n=1 excluded), and drinking at least twice a week (n=137 excluded). Of the 107 eligible participants, 84 expressed interest in participating, completed an online baseline survey, and provided contact information. Participants were enrolled in the study until our recruitment goal (n=75) was reached and remaining students were placed on a waitlist (n=9). Participants were scheduled to attend a brief meeting (in-person or on Zoom [Zoom Communications, Inc]) to receive more information about the study and the EMA app (called ODIN; Grouptheory Inc). Participants were guided through installing the ODIN app on their personal phones and received information about use of passive sensors and how to adjust their phone settings to allow or suspend access. Participants were then shown what to expect while using the app, including example questions, how to track their own participation, and how to contact the research team through the app.

The final study sample (N=68, 91% of those initially enrolled; 68/75) includes participants who attended a training meeting and installed the ODIN app (41/68, 60% recruited directly; 25/68, 37% through friend referral). Participants completed daily questions via the ODIN platform over a continuous, 60-day period between February and April. At the end of the 60 days, participants were sent an email to complete a follow-up survey.

Ethical Considerations

All procedures were approved by the Lehigh University Institutional review board (protocol #1940662-5). All participants provided informed consent. EMA and survey data were anonymized and linked via a random number. Participants received US $20 for completing both the baseline and follow-up survey. In addition, participants could earn up to US $3 a day based on the percentage of questions they answered (eg, a participant who received 10 question prompts and responded to 5 would receive US $1.50).

Participants

The majority of the participants were women (45/68, 66%; 22/68, 32% men; 1/68, 1.5% nonbinary) and were aged between 18 and 23 (mean 20.4, SD 1.2) years. In terms of race, 78% (53/68) identified as White, with 9% (6/68) multiracial, 6% (4/68) Black or African American, 6% (4/68) Asian, and 2% (1/68) Native Hawaiian or Pacific Islander. Further, 9% (6/68) of the sample identified as Latine. At baseline, on average, students reported drinking 17 (SD 11.7) drinks per week. Most participants had smartphones that ran on iOS (65/68, 96% Apple Inc; 3/68, 4% Android users, Google LLC) and at baseline reported using their phone approximately 263.5 minutes a day. Prior to the EMA period, 67 of the 68 (99%) participants reported using their cell phone when out at a party or bar, with the most frequent usages being texting and/or instant messaging and taking pictures and/or video.

XSL•FO
RenderX

Event-Level Procedures

Overview

Questions varied by day of the week, with participants receiving up to 15 questions per day. This included a combination of questions triggered by time (eg, drinking plans reported at noon Thursday to Saturday) or GPS location (see below for more details). Participants received between 2 and 4 time-based surveys a day (plus additional surveys triggered by GPS location). Each survey consisted between 1 and 6 questions. In order for participants to receive questions, the ODIN app had to be running on their phone. If participants closed the app, they would receive a notification asking them to reopen it. If a participants' app was closed for ~3 days, they received up to 4 emails inviting them to continue their participation.

Passive Sensors

The ODIN app uses Bluetooth Low Energy (BLE) protocols on iOS and Android to ensure that each participant's app continuously broadcasts their assigned participant identification number, albeit encrypted by the study-wide secret key. This allowed us to detect when participants were close to other phones that were running the ODIN app. We continuously collected geolocation data (~every 5 m). In order to optimize battery life, motion sensor data was monitored, but not recorded. More specifically, when motion sensor information determined that a participant had not moved from the previous recorded GPS location, GPS data was not recorded.

Morning Retrospective Report

Every morning (9 AM) participants were asked about the previous night including if they attended a social gathering where drinking occurred, how many alcoholic drinks they and others consumed, and what harm reduction strategies they used. The morning self-reported drinking item was used to predict compliance to questions presented the prior evening. Participants also indicate if and when they became too intoxicated to respond to ODIN questions (response option: "No [were never too intoxicated]," "Yes, by 8PM," "Yes, by 10PM," "Yes, by midnight," and "Yes, after midnight").

Time Triggered Questions

In order to assess compliance during common drinking times, participants received a question at 10 PM (Thursday-Saturday) that varied in content (eg, a word unscramble task or being asked to locate someone close by who had a specific characteristic). At 10:30 PM (Thursday-Saturday) participants also received alcohol-related questions, "Are you currently with other students?" "Are you or people around you drinking alcohol?" "What is the MAXIMUM number of drinks it would be acceptable for a typical student to drink at the place you are at?" and "How many standard alcoholic drinks have you had today?" In addition to alcohol questions, participants received nonalcohol related questions related to sense of belonging at college (Sunday-Wednesday at 4 PM). For example, participants were asked to indicate the extent to which "I feel at home in this space."

Geolocation-Triggered Questions

On Thursdays through Saturdays, ODIN delivered geolocation-triggered questions when a participant's phone remained within a prespecified circular geofence surrounding a target drinking venue for at least 10 minutes. The research team identified seven bars and clubs that served alcohol and were located within approximately 0.5 miles (0.8 km) of campus. Target venues were identified using local knowledge and verified using Google Maps (Alphabet Inc), Google Maps venue listings, and the latitude and longitude associated with each venue's Google Maps place pin were used as the center of the geofence. The 0.5-mile distance was used as a practical criterion for selecting target venues; it was not the spatial buffer used to trigger questions. Prompt delivery was instead based on a fixed circular buffer with a radius of 100 feet (30.5 m) around each venue coordinate. For each venue, prespecified triggering hours were determined based on when each establishment was open. Before study launch, research staff visited each target location and confirmed that a prompt was generated after the 10-minute dwell criterion was met. Prior research using ODIN documented high completeness of GPS acquisition over a 30-day period [32]. The geolocation-triggered survey asked participants whether they had been or were currently at a location where alcohol was served.

Survey Measures

Baseline Survey

Prior to EMA data collection, participants completed 3 baseline measures of privacy concerns: location privacy, general privacy risk awareness (adapted from [33]), and concern for information privacy (CFIP; [34]). For all privacy measures, participants indicated their level of agreement on a 7-point scale (1=strongly disagree, 7=strongly agree). Location privacy consisted of 4 questions that began with the stem, "I am fine with sharing my location with…," followed by different types of apps (ie, map, weather, camera, and shopping apps) as well as one question regarding location sharing in general (ie, "In general, I am comfortable sharing my location information on my smartphone, Cronbach α=–0.87). General privacy risk awareness consisted of three questions (ie, "I am aware of potential privacy threats and their consequences from using my smartphone," "I am aware of possible privacy breaches from using my smartphone," and "I understand the privacy risks from using a smartphone," Cronbach α=0.91). The CFIP consisted of 15 questions (Cronbach α=–0.91) assessing participants' concerns that corporations were misusing their data and other private information, with higher scores indicating more concern.

Follow-Up Survey

At the end of the 60-day EMA, participants completed measures of acceptability (based on [35]), including general perceptions of study protocol [36,37] and emotional responses to using the EMA (eg, interest, enjoyment, and annoyance [37]). For example, participants were asked questions related to ease of use: "How did using the ODIN app fit into your routine?" and "How easy was it to answer the daily questions?" (1=very easily to 5=very difficult). Overall satisfaction was assessed with items such as, "Overall, how satisfied are you with the app you used

XSL•FO
RenderX

over the last 60 days?" (1=very satisfied to 5=very dissatisfied) and "Would you be willing to take part in another ODIN study that involves daily surveys?" (1=no, definitely not, to 4=yes, definitely). After individual items, open-ended prompts ("please explain," "why or why not?") were used to clarify responses. Participants were also asked to describe any problems (What problems or challenges did you encounter using the ODIN app throughout the study?) or benefits experienced while using the app (Did you experience any benefits from using the ODIN app?). Finally, participants were provided with a list of potential difficulties in maintaining answering the daily questions and asked to indicate which applied to their experiences. The list included privacy concerns (ie, type or quantity of data collected and worries about data safety), social reasons (ie, busy schedule, interfering with schoolwork), question-related issues (ie, too many questions, too often, or at inconvenient times), and technical issues technical issues (ie, app did not work, app crashed).

At follow-up, participants were also asked to describe their alcohol use over the 60-day EMA period both in terms of the average number of drinks they had on each occasion and how many drinks they had on their maximum drinking occasion [38].

Analysis Plan

Feasibility and Acceptability

Feasibility and acceptability were evaluated using proportions. Benchmarks for feasibility included ≥50% enrollment, ≥70% of participants retained (ie, still engaged with the app at the end of the 60 days, and ≥70% response rate to daily questions [39-41]. Our target benchmark for acceptability was ≥80% endorsement of statements supporting satisfaction with participants' experience with study protocol. In addition, we examined whether participants' emotional responses to using the app (interest, fun, and annoyance) significantly differed from the neutral midpoint of the response scale (4). Open-ended responses were reviewed by 2 authors and summative content analysis was used to identify common categories of responses.

Associations With Compliance

Overall compliance rates were calculated separately for each person by adding the number of surveys completed, dividing this number by the number of surveys received, and multiplying by 100. We examined whether the overall measure of compliance was correlated with privacy concerns assessed baseline. To address whether compliance differed based on sensor type used to trigger questions, we conducted a paired 2-tailed *t* test comparing mean compliance rates for 2 alcohol questions that were triggered by GPS location (bars during opening hours) and time of day (10:30 PM) on Thursday, Friday, and Saturday nights.

To examine the association between compliance rates and alcohol use we examined model fit for generalized linear mixed-effects models using a range of distributions and link functions. Given rates of compliance were generally high and there was little variability (318/362, 88% of the time participants were 100% compliant), based on residual diagnostics and model fit information, we determined that dichotomizing the data (0=not compliant [less than 100%], 1=100% compliant) was the most appropriate analysis strategy. We examined the relationship between morning reports of total alcohol consumed the night before with compliance rates from the night before (for all time-based and GPS trigger questions answered between 10 PM and 2 AM) fitted with a logistic mixed model estimated using maximum likelihood using the glmer function from the *lme4* package in R (version 4.3.1; R Core Team). Fixed effect coefficients were exponentiated to obtain odds ratios (ORs) and associated 95% CIs. Responses on occasions where participants did not drink were not included in these analyses. We examined both whether state- (alcohol use on a given night [person-centered]) and trait-level drinking (average alcohol use on occasions when they drank [grand mean-centered]) predicted compliance while accounting for random variability between participants.

Data Quality

As per Checklist for Reporting Ecological Momentary Assessment Studies (CREMAS) guidelines [42] for reporting EMAs, we describe information about data quality. This includes information on participants' response latencies to question prompts and the minimum number of prompts planned and delivered. In addition, we explore preliminary support for the concurrent validity of participants' responses by examining the correlations among alcohol use reported in the EMA to retrospective reports based on the post-EMA online survey.

Results

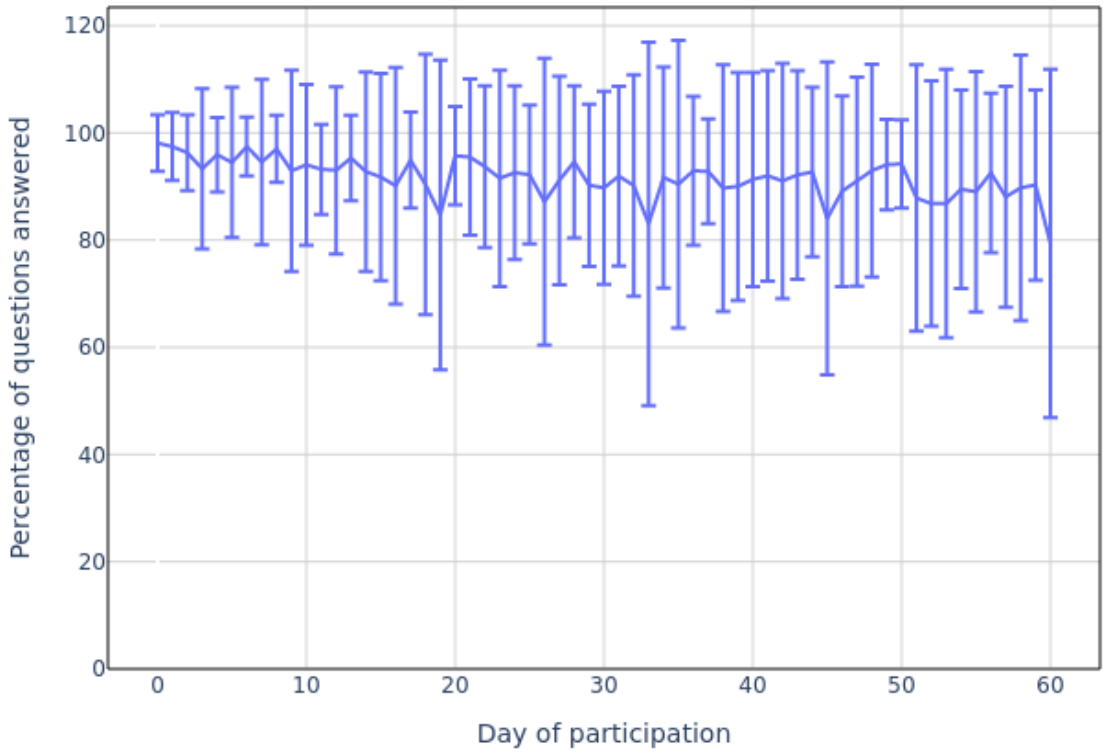
Feasibility

Overview

We exceeded our benchmark for enrollment with 70% (75/107) of eligible participants signing up and an additional 8% (9/107) expressing interest but being placed on a waitlist due to study spaces being filled. Of the 75 students who signed up, 68 attended a meeting to install the ODIN app (68/107, 64% of eligible students). Of the 68 study participants, 64 completed the follow-up survey (94%). To further assess feasibility, average participant compliance (Figure 1) was examined. At all timepoints, average compliance was above the benchmark threshold, with an overall average adherence rate of 90% (median 93%, SD 0.13%, maximum 100%, minimum 0%). Overall, people were asked an average of 339.9 questions (SD 126.92) during the study, and answered an average of 315.9 (SD 122.61) questions. Overall, 65 of the 68 (96%) participants met our benchmark of completing at least 80% of questions asked. One participant received four questions and then did not respond to further questions or attempts to contact them, and they were excluded from subsequent analyses.

XSL•FO
RenderX

Figure 1. Study compliance overtime. Maximum number of drinks refers to the maximum number of drinks on a single occasion. Survey responses reflect post EMA self reports of drinking during the past 60 days.



Compliance Based on Sensor Type (Time vs GPS)

The analysis comparing compliance based on sensor type used to trigger questions revealed that participants' compliance rates differed as a function of question type, with higher compliance for time-based (mean 87%, SD 16%) compared to location-based questions (mean 76%, SD 27%; t_{59}=3.65; *P*<.001).

Association Between Compliance and Participants' Alcohol Consumption

There were high rates of compliance for both morning (85%) and evening (87%) alcohol assessments, and there were 362 person-level drinking days observed. Participants' average consumption on a given night varied from 1.7 to 11.5 (mean 5.16, SD 2.07) drinks. The logistic mixed model (adjusted intraclass correlation coefficient [ICC]=0.28) explained a moderate amount of variance in compliance (conditional R^2=0.29); however, only a small amount was explained by fixed effects alone (marginal R^2=0.01). The number of drinks a participant reported consuming on a given night (assessed the following morning) was not related to compliance for that night (OR 0.94, 95% CIs 0.80-1.10). Further, participants' average number of drinks consumed was not associated with their typical compliance (OR 0.92, 95% CIs 0.73-1.17).

If a participant indicated that they had drank alcohol the previous evening, they were asked if at any point they felt too drunk to respond to ODIN questions. Overall, 79.1% of the time (568/718 morning reports) participants indicated that they were never too drunk to respond to EMA questions. When participants indicated that they did become too drunk to respond, this most commonly occurred between 10 PM and midnight (59/718, 8%) and after midnight (56/718, 8%).

Compliance and Privacy Concerns

Location privacy concern was not associated with participants' overall compliance rates, r_{65}= 0.07, *P*=.57. Both greater general privacy risk awareness (r_{65}=0.24; *P*=.0497) and privacy concerns (r_{65}=0.25; *P*=.04) were associated with higher compliance rates.

Data Quality

Prompt Latency

We examined the amount of time between a question prompt and participants answering the question. Looking at all types of questions prompts, on average participants took 7.1 minutes to respond. We then explored response times specifically to questions that were the first in a chain of prompts, for these questions response times were slower (mean 23.8, SD 9.8 minutes).

Prompt Delivery

In addition to examining compliance (i.e., participants' response rate to questions they received), we also examined data quality and missingness in terms of failure to receive anticipated question prompts. We were not able to calculate the exact total

XSL•FO
RenderX

number of questions each participant should have received for a number of reasons. This included geolocation questions not being predictable and participants' responses to prompts determining if and how many additional questions they received in a chain. Therefore, we estimated the total number of first questions in time-based chains that each person should have received during the study (20 per week or 172 across the 60 days) and compared this to the number of time-based first in chain questions received. On average, participants received less than the minimum number of prompts (median 143, IQR 110.5-161.0), receiving on average 83% of the time-based question chains. In most cases, it was not possible to determine why the prompts were not received, though it may have reflected phones being turned off or phones losing network connection. In addition, we did experience a technical issue related to daylight savings that impacted prompts being delivered for a number of days after the time change. In total, 19 participants had to relaunch or reinstall the app and this impacted data collection for a period of 1 to 10 days (mean 5.5, SD 3.6 days). One participant received a new phone during the study period and went 2 days without receiving questions before rejoining the study. A second student reported that their phone broke during the study period and that they had not taken the time to have the app set up on their new phone.

Concurrent Validity

As a preliminary exploration of the quality of drinking self-report data, we examined the correlation between self-reports of alcohol use post-EMA with EMA drinking data (Table 1). Participants' retrospective estimates of their average alcohol consumption on a single occasion during the 60-day EMA period were positively correlated with their average drinks per occasion reported in the EMAs (r_{62}=0.71; P<.001). Similarly, participants' retrospective estimates of how much they consumed on their maximum drinking occasion during the study period were positively associated with the maximum number of drinks observed in the EMA data (r_{62}=0.71; P<.001). These positive associations lend some support to the validity of the EMA alcohol data.

Table 1. Correlations between ecological momentary assessment alcohol reports and post-EMA self-report.

	Values, mean (SD)	EMA[a] average drinks per occasion	EMA maximum number of drinks[b]	Survey average drinks per occasion[c]	Survey maximum number of drinks[c]
EMA average drinks per occasion	5.1 (2.0)	—[d]			
EMA maximum number of drinks	8.4 (3.8)	.83[e]	—[d]		
Survey average drinks per occasion	4.5 (2.2)	.71[e]	.39[f]	—[d]	
Survey maximum number of drinks	9.5 (4.7)	.77[e]	.71[e]	.58[e]	—[d]

[a]EMA: ecological momentary assessment.

[b]Maximum number of drinks refers to the maximum number of drinks on a single occasion.

[c]Survey responses reflect post EMA self-reports of drinking during the past 60 days.

[d]Not applicable.

[e]P<.001.

[f]P<.01.

Acceptability

Survey Data

Across measures we consistently approached or met benchmarks for acceptability (≥80% endorsement). For example, the vast majority of participants stated that answering questions daily was easy to extremely easy (55/66, 84%) and that using the ODIN app fit somewhat easily to very easily into their routine (51/66, 78%). Overall, 72% (47/66) stated they were moderately to very satisfied with the ODIN app, 88% (58/66) stated they were willing to complete another ODIN EMA study, and 88% (58/66) said that they would recommend being part of an ODIN study to their friends. In terms of emotional responses to using the app, ratings of annoyance (mean 3.84, SD 1.14) did not significantly differ from the neutral midpoint (4; t_{63}=–1.10; P=.28, Cohen d=–0.14). Ratings of interest in the app (mean 4.55, SD 1.26) and enjoyability of using the app (mean 4.58, SD 1.23) were significantly higher than the neutral midpoint (both P<.001, Cohen d=0.44 and 0.48 respectively), indicating that participants tended to have slightly positive responses to the app.

When asked specifically what difficulties interfered with answering the daily questions, very few participants (6/66, 9%) noted concerns about privacy and data safety. More common difficulties included issues with questions (ie, too many questions or questions at inconvenient times; 27/66, 41%), social reasons (ie, busy schedule, interfering with school work; 33/66, 46%), or technical reasons (eg, app bug related to daylight savings, battery drain; 37/66, 56%).

Open-Ended Data

Open-ended responses suggested there were several reasons students found the experience positive. In addition to receiving incentives, participants (21/66, 32%) described that the EMA made them more conscious of their drinking habits. For example, one student stated that, "I was more aware of how many drinks I was consuming on a nightly basis" and another reported, "I thought more about what an appropriate amount of drinks to have was." Some students also suggested that participating in the EMA changed their drinking behavior in beneficial ways. One student stated that the prompts "would remind me [of] tactics to use before going out and drinking alcohol at night"

XSL•FO
RenderX

and another student reported that, “It helped with drinking habits. I was able to control my alcohol better and reflect on how I drank the night before.”

Some students reported that they found answering the questions enjoyable or fun to answer (15%; 10 / 66). These comments often referred specifically to the 10pm evening questions that varied nightly in terms of content. For example, one student stated that “the questions that came sporadically and focused on a variety of topics about student experience… were engaging.” and another said, “Unscrambling the words was random but enjoyable.” At the same time, they appreciated consistency in the timing of questions (11/66, 17%). For example, one student stated, “It was good that the times were predictable” and another said that they were satisfied with their experience because questions were “Routine and easy to predict.”

Discussion

Principal Findings

This study adds to the literature by examining the feasibility and acceptability of a 60-day alcohol-related EMA that integrated the use of passive sensors (GPS, Bluetooth, and motion) and triggered questions based on GPS location data. Further, the study examined the role of both individual-level predictors (alcohol use and privacy concerns) and question type in relation to compliance. The study provides support for both the feasibility and acceptability of this approach among college students who drink regularly. The results suggest that college students do not have significant concerns about sharing passive data. Indeed, acceptability ratings suggested that participants generally had a positive experience using the ODIN app and would be willing to take part in a similar study in the future. Further, compliance rates remained high across the 60-day study period, suggesting that it is feasible to conduct this type of EMA for more than a month and maintain engagement.

Predictors of Compliance

In addition to examining overall feasibility and acceptability, this study examined predictors of compliance rates. Past research has suggested that alcohol use may interfere with students’ responses to EMA questions [23,25]. Our findings are consistent with research suggesting that alcohol use at both a trait [24] and state-level [22,26] are not associated with EMA compliance. Participants’ subjective self-report data suggested that the majority of the time participants did not perceive their alcohol use to interfere with their responses. These findings are somewhat surprising given that alcohol use can impact people’s cognitive capacity in ways that reduce the ability to perform intentional behaviors or engage in deliberative information processing [43]. The lack of impact of alcohol on compliance could be explained by alcohol myopia theory [44], which suggests that once alcohol is consumed people will become more focused on the salient aspects of their immediate environment. Given that the vast majority of the sample reported typically using their cell phones at bars and parties to engage in a number of activities (eg, texting, instant messaging, and taking pictures and/or video), it is possible that cues or prompts to complete EMA surveys are salient and influential for intoxicated people who are already engaging with their phones. The questions participants were asked to complete in this study did not require complex information processing or decision making. Based on the dual-process model [43], it is possible that the impact of alcohol use on compliance may become more pronounced in EMA studies that require participants to complete more novel and effortful tasks.

To our knowledge, this study is the first to examine the relationship between privacy concerns and EMA participation. Given that the study included passive sharing of sensor data, including GPS data, it is notable that location privacy was not related to overall compliance. Furthermore, privacy concerns and awareness were unexpectedly associated with higher compliance rates. Research among community samples of people who use drugs suggests that privacy concerns may be perceived as a barrier to compliance for EMA studies using GPS data [28]. The current findings suggest that the impact of privacy concerns may be less significant for college students reporting on their alcohol use. In this study, participants met with the researchers in order to review consent forms and be trained on the app. While this approach is more labor intensive than participants reviewing consent forms independently, it may be beneficial for ensuring participants feel comfortable responding and that they understand the steps taken to ensure their confidentiality. It is also possible that privacy concerns were not associated with decreased compliance because these concerns are more relevant to the decision to take part in a research study and are not as relevant to responding once in a study [30]. As collecting data on privacy concerns of those who chose not to participate may be challenging, further research using hypothetical EMA designs [45] could help determine if privacy concerns differentially affect willingness to enroll in EMAs that involve passive sensing compared to those that do not.

While participants in this study had higher compliance rates for questions triggered by time rather than location, compliance rates for GPS-triggered questions were still above our benchmark. The difference in response rate by question trigger is consistent with past research that indicates that when questions are more unpredictable, response rates are generally lower [25]. In this study, participants were not specifically told what locations would trigger questions. To improve compliance, it may be beneficial to have participants list out bars or clubs that they attend and create a plan or an implementation intention [46] to check their phone for questions when they are in that specific location. Using participant generated locations could also increase the predictability of where location-based questions are triggered. While consistency may be useful for compliance, the current findings also suggest that it may be beneficial to consider ways to balance asking questions at predictable times with variety in question content to promote participant interest and engagement in EMAs over longer periods. We note that while the less predictable timing of geolocation-triggered questions may have contributed to their lower response rate, the absence of independent geofence validation means that we cannot distinguish this explanation completely from effects related to geofence performance or location misclassification.

XSL•FO
RenderX

Limitations

This study has a number of limitations. First, it is possible that the sample is not representative of college students who drink frequently due to a self-selection bias of those more comfortable with an EMA involving passive sensors. While this concern is somewhat diminished by the variability in baseline privacy concerns and the high rates of interest among eligible students, it is possible that the sample underrepresented students who are cautious sharing their data. Furthermore, privacy concerns were only measured at baseline and not as part of the EMA protocol. It is possible that participants' privacy concerns fluctuate or are more salient in some contexts compared to others. While repeatedly asking students about their privacy concerns could itself potentially influence these concerns, longitudinal assessment of context-specific privacy concerns could help better understand EMA responding.

While the current findings suggest that alcohol use did not impact compliance, we did not explore whether the accuracy or quality of participants' responses decreased as their alcohol consumption increased. The preliminary data supporting the validity of the EMA drinking reports based on comparisons with follow-up survey recall of alcohol use does somewhat address this concern; however, additional research examining how alcohol use impacts the reliability of responses is warranted.

This study focused on human factors that impacted compliance, rather than the impact of technical issues on responding. That said, the generally high compliance and acceptability rates should be interpreted in the context of the study protocol and technical challenges experienced during the study. For example, despite issues related to daylight savings that caused data collection disruptions, participants remained generally satisfied with the EMA experience. It is possible that because most participants did not receive the minimum expected number of time-based prompts, participant fatigue was reduced and this enhanced overall compliance. Finally, past research suggests that financial incentives impact compliance rates [16,31] and therefore the high rates of compliance observed in this study may not extend to research that does not employ similar incentive structures.

We did not conduct an independent ground-truth validation of point-level GPS accuracy or geofence classification. Consequently, some prompts may have been generated while a participant was near, rather than inside, a target venue, and some actual visits may not have generated a prompt. The location-triggered compliance estimate should therefore be interpreted as responsiveness to prompts generated by the geofencing protocol, rather than as confirmation that every response was completed inside a bar or club.

Conclusions

The integration of passive sensors into college-alcohol EMAs provides an opportunity to study the influence of context on alcohol beliefs and behaviors. The current results indicate that students were generally willing to respond to alcohol-related prompts and questions generated by geofences surrounding bars and clubs. Compliance was not associated with trait-level or in-the-moment alcohol use and was positively associated with some forms of privacy concerns. Response rates were higher for more consistent time-based questions compared to more variable location-based questions. Overall, the current study provides data suggesting that alcohol-related EMAs among college students that incorporate the use of passive sensors can be both acceptable and feasible.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Funding

This work was supported by a grant #001249 from Lehigh University.

Authors' Contributions

Conceptualization: LN, BK
Data curation: BK, MS, LN, GF
Formal analysis: LN (lead), GF (supporting)
Funding acquisition: LN, BK
Investigation: GF (lead), LW (supporting)
Methodology: LN (lead), GF, BK, MS, DM
Project administration: GF (lead), LN (supporting)
Supervision: LN, BK
Validation: BK
Visualization: BK
Writing – original draft: LN (lead), GF, BK, LW
Writing – review and editing: LN, GF, DM

Conflicts of Interest

BK created the ODIN app while employed at the University of Nebraska-Lincoln; the app was used to collect the data for this study. All other authors declare no conflicts of interest.

XSL•FO
RenderX

References

1. Patrick ME, Miech RA, Johnston LD, O'Malley PM. Monitoring the Future. Monitoring the Future Panel Study Annual Report: national data on substance use among adults ages 19 to 60, 1976-2023. 2024. URL: https://monitoringthefuture.org/monitoring-the-future-panel-study-annual-report-national-data-on-substance-use-among-adults-ages-19-to-60-1976-2022/ [accessed 2026-08-17]
2. Hingson RW, Zha W, Weitzman ER. Magnitude of and trends in alcohol-related mortality and morbidity among U.S. college students ages 18-24, 1998-2005. J Stud Alcohol Drugs Suppl 2009(16):12-20 [FREE Full text] [doi: 10.15288/jsads.2009.s16.12] [Medline: 19538908]
3. Nelson TF, Xuan Z, Lee H, Weitzman ER, Wechsler H. Persistence of heavy drinking and ensuing consequences at heavy drinking colleges. J Stud Alcohol Drugs 2009;70(5):726-734. [doi: 10.15288/jsad.2009.70.726] [Medline: 19737497]
4. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
5. Wray TB, Merrill JE, Monti PM. Using ecological momentary assessment (EMA) to assess situation-level predictors of alcohol use and alcohol-related consequences. Alcohol Res 2014;36(1):19-27 [FREE Full text] [doi: 10.35946/arcr.v36.1.03] [Medline: 26258997]
6. Gius BK, Schlauch RC. Approach and avoidance alcohol inclinations in heavy drinking college students: an ecological momentary assessment study. Addict Behav 2021;123:107080. [doi: 10.1016/j.addbeh.2021.107080] [Medline: 34416534]
7. Howard AL, Lamb M, Alexander SM, Bradley AHM, Carnrite K, Milyavskaya M, et al. Planned and unplanned drinking to get drunk: a registered report examining willingness, drinking motives, and protective behavioral strategies using ecological momentary assessment. Psychol Addict Behav 2024;38(5):519-539. [doi: 10.1037/adb0000909] [Medline: 39073385]
8. Merrill JE, Boyle HK, Jackson KM, Carey KB. Event-level correlates of drinking events characterized by alcohol-induced blackouts. Alcohol Clin Exp Res 2019;43(12):2599-2606 [FREE Full text] [doi: 10.1111/acer.14204] [Medline: 31557348]
9. Howe LK, Copeland S, Fisher L, Farmer E, Nemes L, Finn PR. Mobile assessment of decisions to drink in young adults: examining the role of incentives and disincentives. Alcohol Clin Exp Res 2022;46(1):152-165 [FREE Full text] [doi: 10.1111/acer.14746] [Medline: 34806196]
10. Dora J, Schultz ME, Shoda Y, Lee CM, King KM. No evidence for trait- and state-level urgency moderating the daily association between negative affect and subsequent alcohol use in two college samples. Brain Neurosci Adv 2022;6:1-12 [FREE Full text] [doi: 10.1177/23982128221079556] [Medline: 35237726]
11. Phillips KT, Prince MA, Phillips MM, Lalonde TL, Stein MD. Momentary patterns of alcohol and cannabis co-use in college students: assessing the temporal association with anxiety. Cannabis 2022;5(1):42-58 [FREE Full text] [doi: 10.26828/cannabis/2022.01.005] [Medline: 35938093]
12. Zhang MJ, Luk TT, Ho SY, Wang MP, Lam TH, Cheung YTD. Ecological momentary assessment of alcohol marketing exposure, alcohol use, and purchases among university students: prospective cohort study. JMIR Mhealth Uhealth 2024;12:e60052 [FREE Full text] [doi: 10.2196/60052] [Medline: 39226102]
13. O'Donnell R, Richardson B, Fuller-Tyszkiewicz M, Liknaitzky P, Arulkadacham L, Dvorak R, et al. Ecological momentary assessment of drinking in young adults: an investigation into social context, affect and motives. Addict Behav 2019;98:106019. [doi: 10.1016/j.addbeh.2019.06.008] [Medline: 31247534]
14. Braitman AL, Shipley JL, Strowger M, Renzoni ES. Identifying salient social, environmental, and situational contexts of college drinking: impacts across drinking outcomes. Am J Addict 2025;34(6):614-622. [doi: 10.1111/ajad.70065] [Medline: 40641334]
15. Jones A, Remmerswaal D, Verveer I, Robinson E, Franken IHA, Wen CKF, et al. Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. Addiction 2019;114(4):609-619 [FREE Full text] [doi: 10.1111/add.14503] [Medline: 30461120]
16. Wrzus C, Neubauer AB. Ecological momentary assessment: a meta-analysis on designs, samples, and compliance across research fields. Assessment 2023;30(3):825-846 [FREE Full text] [doi: 10.1177/10731911211067538] [Medline: 35016567]
17. Williams MT, Lewthwaite H, Fraysse F, Gajewska A, Ignatavicius J, Ferrar K. Compliance with mobile ecological momentary assessment of self-reported health-related behaviors and psychological constructs in adults: systematic review and meta-analysis. J Med Internet Res 2021;23(3):e17023 [FREE Full text] [doi: 10.2196/17023] [Medline: 33656451]
18. Perski O, Keller J, Kale D, Asare BY, Schneider V, Powell D, et al. Understanding health behaviours in context: a systematic review and meta-analysis of ecological momentary assessment studies of five key health behaviours. Health Psychol Rev 2022;16(4):576-601 [FREE Full text] [doi: 10.1080/17437199.2022.2112258] [Medline: 35975950]
19. Dauber S, Beacham A, West A, Devkota J, Barrie K, Thrul J. Ecological momentary assessment of heavy episodic drinking in the early postpartum period: a feasibility study. Drug Alcohol Depend Rep 2023;7:100146 [FREE Full text] [doi: 10.1016/j.dadr.2023.100146] [Medline: 37012980]
20. Hanson JD, Harris A, Gilbertson RJ, Charboneau M, O'Leary M. Feasibility and acceptability of using ecological momentary assessment to evaluate alcohol use with American Indian women. Int J Environ Res Public Health 2023;20(12):1-12 [FREE Full text] [doi: 10.3390/ijerph20126071] [Medline: 37372658]

XSL•FO
RenderX

21. Yang C, Linas B, Kirk G, Bollinger R, Chang L, Chander G, et al. Feasibility and acceptability of smartphone-based ecological momentary assessment of alcohol use among African American men who have sex with men in Baltimore. JMIR Mhealth Uhealth 2015;3(2):e67 [FREE Full text] [doi: 10.2196/mhealth.4344] [Medline: 26085078]
22. Sokolovsky AW, Gunn RL, Wycoff AM, Boyle HK, White HR, Jackson KM. Compliance and response consistency in a lengthy intensive longitudinal data protocol. Psychol Assess 2024;36(10):606-617. [doi: 10.1037/pas0001332] [Medline: 39101913]
23. Suffoletto B, Field M, Chung T. Attentional and approach biases to alcohol cues among young adult drinkers: an ecological momentary assessment study. Exp Clin Psychopharmacol 2020;28(6):649-658 [FREE Full text] [doi: 10.1037/pha0000343] [Medline: 31886700]
24. Howard AL, Lamb M. Compliance trends in a 14-week ecological momentary assessment study of undergraduate alcohol drinkers. Assessment 2024;31(2):277-290 [FREE Full text] [doi: 10.1177/10731911231159937] [Medline: 36914966]
25. Hultgren BA, Scaglione NM, Buben A, Turrisi R. Examining protocol compliance and self-report congruence between daily diaries and event-contingent ecological momentary assessments of college student drinking. Addict Behav 2020;110:106471 [FREE Full text] [doi: 10.1016/j.addbeh.2020.106471] [Medline: 32526551]
26. Murray A, Yang Y, Zhu X, Speyer L, Brown R, Eisner M, et al. Respondent characteristics associated with adherence in a general population ecological momentary assessment study. Int J Methods Psychiatr Res 2023;32(4):e1972 [FREE Full text] [doi: 10.1002/mpr.1972] [Medline: 37184112]
27. Rawal S. A proposal for predicting missing values in a dataset using supervised learning. Int J Adv Res Comput Sci 2017;8(8):562-567. [doi: 10.26483/ijarcs.v8i8.4852]
28. Rudolph AE, Young AM, Havens JR. Privacy, confidentiality, and safety considerations for conducting geographic momentary assessment studies among persons who use drugs and men who have sex with men. J Urban Health 2020;97(2):306-316 [FREE Full text] [doi: 10.1007/s11524-018-0315-x] [Medline: 30324355]
29. Viera A, Jadovich E, Lauckner C, Muilenburg J, Kershaw T. Responding to location-based triggers of cravings to return to substance use: a qualitative study. J Subst Use Addict Treat 2025;168:209534. [doi: 10.1016/j.josat.2024.209534] [Medline: 39426681]
30. Wyant K, Moshontz H, Ward SB, Fronk GE, Curtin JJ. Acceptability of personal sensing among people with alcohol use disorder: observational study. JMIR Mhealth Uhealth 2023;11:e41833 [FREE Full text] [doi: 10.2196/41833] [Medline: 37639300]
31. Zhang Y, Li D, Li X, Zhou X, Newman G. The integration of geographic methods and ecological momentary assessment in public health research: a systematic review of methods and applications. Soc Sci Med 2024;354:117075. [doi: 10.1016/j.socscimed.2024.117075] [Medline: 38959816]
32. McChargue DE, Khan B, Phelps J, Duryea P, Tyler KA, Andrews A, et al. The feasibility of utilizing the open dynamic interaction network (ODIN) app to assess rEMA data across 30 days among those recovering from alcohol use disorders. Drug Alcohol Depend Rep 2025;14:100305 [FREE Full text] [doi: 10.1016/j.dadr.2024.100305] [Medline: 39720161]
33. Bulgurcu B, Cavusoglu H, Benbasat I. Information security policy compliance: an empirical study of rationality-based beliefs and information security awareness. MIS Q 2010;34(3):523-548. [doi: 10.2307/25750690]
34. Smith HJ, Milberg SJ, Burke SJ. Information privacy: measuring individuals? concerns about organizational practices. MIS Q 1996;20(2):167-196.
35. Larsen DL, Attkisson CC, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. Eval Program Plann 1979;2(3):197-207. [doi: 10.1016/0149-7189(79)90094-6] [Medline: 10245370]
36. Martinez GJ, Mattingly SM, Robles-Granda P, Saha K, Sirigiri A, Young J, et al. Predicting participant compliance with fitness tracker wearing and ecological momentary assessment protocols in information workers: observational study. JMIR Mhealth Uhealth 2021;9(11):e22218 [FREE Full text] [doi: 10.2196/22218] [Medline: 34766911]
37. Mundi MS, Lorentz PA, Grothe K, Kellogg TA, Collazo-Clavell ML. Feasibility of smartphone-based education modules and ecological momentary assessment/intervention in pre-bariatric surgery patients. Obes Surg 2015;25(10):1875-1881. [doi: 10.1007/s11695-015-1617-7] [Medline: 25702141]
38. Kirkland AE, Ferguson PL, Alcala TJ, Mewton L, Tomko RL, Squeglia LM. How much is enough? assessing 30- versus 60-day timeline followback stability in youth alcohol use patterns. Alcohol Clin Exp Res 2025;49(11):2619-2628. [doi: 10.1111/acer.70177] [Medline: 41144690]
39. Christy K, Jensen J, Britt B, Scherr C, Jones C, Brown N. I want to talk to a real person: theorising avoidance in the acceptance and use of automated technologies. Health Syst (Basingstoke) 2019;8(1):31-43 [FREE Full text] [doi: 10.1080/20476965.2017.1406568] [Medline: 31178990]
40. Clark V, Kim SJ. Ecological momentary assessment and mHealth interventions among men who have sex with men: scoping review. J Med Internet Res 2021;23(8):e27751 [FREE Full text] [doi: 10.2196/27751] [Medline: 34342585]
41. Kwasnicka D, Kale D, Schneider V, Keller J, Yeboah-Asiamah Asare B, Powell D, et al. Systematic review of ecological momentary assessment (EMA) studies of five public health-related behaviours: review protocol. BMJ Open 2021;11(7):e046435 [FREE Full text] [doi: 10.1136/bmjopen-2020-046435] [Medline: 34272218]

XSL•FO
RenderX

42. Liao Y, Skelton K, Dunton G, Bruening M. A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: an adapted STROBE checklist for reporting EMA studies (CREMAS). J Med Internet Res 2016;18(6):e151 [FREE Full text] [doi: 10.2196/jmir.4954] [Medline: 27328833]
43. Moss AC, Albery IP. A dual-process model of the alcohol-behavior link for social drinking. Psychol Bull 2009;135(4):516-530. [doi: 10.1037/a0015991] [Medline: 19586160]
44. Steele CM, Josephs RA. Alcohol myopia. Its prized and dangerous effects. Am Psychol 1990;45(8):921-933. [doi: 10.1037//0003-066x.45.8.921] [Medline: 2221564]
45. Ochoa Gómez C. Willingness to participate in geolocation-based research. PLoS One 2022;17(12):e0278416 [FREE Full text] [doi: 10.1371/journal.pone.0278416] [Medline: 36454965]
46. Gollwitzer PM. Implementation intentions: strong effects of simple plans. Am Psychol 1999;54(7):493-503. [doi: 10.1037/0003-066X.54.7.493]

Abbreviations

BLE: Bluetooth Low Energy
CFIP: Concern for Information Privacy
CREMAS: Checklist for Reporting Ecological Momentary Assessment Studies
EMA: ecological momentary assessment
ICC: intraclass correlation coefficient
OR: odds ratio

Edited by L Buis; submitted 01.Aug.2025; peer-reviewed by NMM Scaglione, TS Ou, J Han; comments to author 21.Oct.2025; revised version received 10.Jul.2026; accepted 13.Jul.2026; published 09.Oct.2026.

Please cite as:
Napper LE, Khan B, Fox GQ, Saad MK, Wolter LC, McChargue DE
Exploring College Students' Compliance During a 60-Day Alcohol-Related Ecological Momentary Assessment Using Location- and Time-Based Questions: Prospective Observational Study
JMIR Mhealth Uhealth 2026;14:e81713
URL: https://mhealth.jmir.org/2026/1/e81713
doi:10.2196/81713
PMID:

©Lucy E Napper, Bilal Khan, Gavin Q Fox, Mohamed K Saad, Laura C Wolter, Dennis E McChargue. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 09.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Behavioral Modification as a Putative Mediator of Digital Therapeutic Response in Temporomandibular Disorders: Secondary Analysis of a Multicenter Sham-Controlled Randomized Trial

Soo-Hwan Byun[1,2,3*], DDS, PhD; Sung-Woon On[1,2,4*], DDS, PhD; Byong-Eun Yang[1,2,3], DDS, PhD; Sung-Ah Che[1,2,4], DDS, MS; Sang-Min Yi[1,2,3], DDS, MS; Yongjin Park[5], PhD; Yeolib Kim[6], PhD; Sang-Yoon Park[1,2,3], Prof Dr, DDS, PhD

[1]Institute of Clinical Dentistry, Hallym University, Hallym University, Chuncheon, Gangwon-do, Republic of Korea

[2]Dental Artificial Intelligence and Robotics R&D Center, Hallym University Medical Center, Anyang, Gyeonggi-do, Republic of Korea

[3]Department of Oral and Maxillofacial Surgery, Hallym University Sacred Heart Hospital, 22, Gwanpyeong-ro 170-gil, Dongan-gu, Anyang, Gyeonggi-do, Republic of Korea

[4]Department of Oral and Maxillofacial Surgery, Hallym University Dongtan Sacred Heart Hospital, Hwaseong, Gyeonggi-do, Republic of Korea

[5]Department of Information Systems, College of Business, City University of Hong Kong, Hong Kong, China (Hong Kong)

[6]School of Business Administration, Ulsan National Institute of Science and Technology, Ulsan, Ulsan, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Sang-Yoon Park, Prof Dr, DDS, PhD
Institute of Clinical Dentistry, Hallym University, Hallym University, Chuncheon, Gangwon-do, Republic of Korea

Abstract

Background: Temporomandibular disorders (TMDs) are common chronic conditions involving orofacial pain and functional limitations. Digital therapeutics (DTx) have demonstrated efficacy in TMD management; yet, the behavioral and clinical mechanisms underlying treatment response remain poorly characterized, particularly whether behavioral modification or DTx engagement intensity drives therapeutic benefit.

Objective: This study aimed to investigate the behavioral mechanisms, responder profiles, and moderators of clinical response to a DTx intervention for TMD through a post hoc analysis integrating self-reported, server-derived, and clinician-rated outcome measures.

Methods: We performed a post hoc secondary analysis of a multicenter, double-blind, sham-controlled randomized superiority trial conducted at 2 tertiary care centers in South Korea. The per-protocol cohort comprised 93 participants (DTx: n=44; sham: n=49). Five complementary analyses were applied: responder logistic regression at ≥30%, ≥50%, and ≥70% Visual Analog Scale (VAS) pain-reduction thresholds; subgroup comparison by Oral Behaviors Checklist (OBC) modifier status; causal mediation analysis using the potential outcomes framework with bootstrap CIs; week-4 sensitivity analysis; and moderator analysis testing the treatment×Patient Health Questionnaire-4 (PHQ-4) interaction on VAS change.

Results: DTx assignment was consistently associated with clinically meaningful pain reduction across all 3 responder thresholds (adjusted odds ratios [ORs] 5.39, 95% CI 1.72 - 16.94 at ≥30%; 3.21, 95% CI 1.14 - 8.99 at ≥50%; and 3.45, 95% CI 1.12 - 10.63 at ≥70%; all P<.05). Mediation analysis suggested that approximately 29.1% of the total VAS treatment effect may be transmitted via OBC-defined behavioral modification (natural indirect effect −6.91 mm; 95% CI −13.23 to −0.58; P=.03), with the mediated proportion rising from 19.6% to 30.2% as responder thresholds became more stringent. Participants with OBC modifiers achieved substantially greater pain reduction than nonmodifiers (−45.71 vs −22.61 mm; difference −23.11; 95% CI −36.97 to −9.24; P<.01) despite no significant differences in any objective DTx engagement metric. Treatment ORs were 33% - 42% higher at week 4 than at the 6-week end point (week-4 ORs 7.15, 95% CI 2.39 - 21.32 at ≥30%; 4.48, 95% CI 1.68 - 11.94 at ≥50%; and 4.91, 95% CI 1.74 - 13.87 at ≥70%), suggesting that week 4 may be a candidate time point for future adaptive protocols. Baseline psychological distress (PHQ-4 ≥3) appeared to moderate the treatment response (interaction β=−19.63; 95% CI −37.86 to −1.39; P=.04).

Conclusions: Sham-controlled randomized trials in TMD that empirically differentiate behavioral realization from digital engagement volume remain scarce. Behavioral modification, rather than engagement volume, appears to be an important pathway associated with DTx efficacy and may mediate approximately 29% of the pain-reduction effect under exploratory causal assumptions. This advances mechanism-based evaluation of DTx beyond engagement-based surrogates. Week 4 may represent

XSL•FO
RenderX

a promising candidate for future adaptive protocols, and baseline psychological distress may warrant investigation for precision patient selection. These preliminary findings should be corroborated by future prospective studies and further mechanistic investigations.

Trial Registration: Clinical Research Information Service KCT0009493; https://tinyurl.com/w5huy4zn

(JMIR Mhealth Uhealth 2026;14:e104264) doi:10.2196/104264

KEYWORDS

temporomandibular joint disorders; behavior therapy; digital therapeutics; mediation analysis; randomized controlled trial

Introduction

Temporomandibular disorders (TMDs) are common, heterogeneous conditions involving the temporomandibular joint, masticatory muscles, and related structures, typically presenting with orofacial pain and functional limitation [1]. Current models emphasize a biopsychosocial framework where symptom persistence reflects interactions among peripheral factors, central pain modulation, psychological distress, and maladaptive oral behaviors [2-4]. In the diagnostic criteria (DC) for TMDs, Axis II instruments operationalize key psychosocial and behavioral domains; parafunctional habits are clinically relevant because they sustain mechanical loading and pain-guarding cycles [2,5]. Although conservative, reversible, and often multidisciplinary approaches in Axis II are recommended as first-line care [6,7], their effectiveness is often limited by variable adherence to daily behavioral changes outside the clinic [8,9].

Digital therapeutics (DTx) are software-driven, evidence-based therapeutic interventions that have been proposed as a scalable means to deliver standardized behavioral and self-management support and reinforce adherence between clinic visits [10,11]. Over the past decade, DTx have expanded rapidly across medical specialties; an analysis of registered trials documented a more than 5-fold increase since 2011, driven predominantly by mental health and neurological applications [12]. Despite this growth, their development remains unevenly distributed across disciplines and is constrained by a lack of standardized evaluation frameworks, and applications in dentistry and orofacial pain remain scarce [13]. Given the chronic, behaviorally mediated nature of TMD, it is a clinically suitable target for DTx; yet, rigorous evidence in this field has emerged only recently [14,15]. Our group previously reported a multicenter, double-blind, randomized, sham-controlled trial evaluating a TMD DTx intervention, demonstrating clinically meaningful improvements in pain, jaw function, and oral behavior over a 6-week treatment phase [14]. The primary analysis appropriately focused on average between-group effects [14].

However, establishing DTx efficacy is the first step. For clinical implementation and product optimization, it is equally important to understand how benefits are distributed within the treated group and through which modifiable domains contribute to improvement. Digital interventions introduce a persistent black box problem; clinical benefit is plausibly contingent on engagement, but actual exposure varies widely and is not captured by treatment assignment alone [16,17]. Unlike pharmacologic dose, digital dose" is behaviorally expressed and often declines over time, raising questions regarding response heterogeneity and effective behavioral pathways [18]. Schmidt et al [19] have recently proposed a conceptual framework distinguishing "consumed dose" (patient-DTx interaction) from "realized dose" (actual cognitive-affective and behavioral change), arguing that clinical outcomes are driven by dose realization rather than dose intake alone. In this context, an exploratory review of usage data from our randomized controlled trial suggested that greater app use alone was not consistently associated with greater clinical improvement, motivating a more granular, process-level analysis [17]. A better understanding of which behavioral changes accompany the treatment effect may help distinguish therapeutic benefit from engagement volume alone and could inform how these interventions are optimized and applied. Such mechanistic insight may support the progression of TMD DTx from demonstrated efficacy toward more precision-oriented use [19,20].

Accordingly, we conducted a post hoc analysis linking objective DTx use data with repeated clinical and Axis II measures. Such mechanistic analyses remain scarce in TMD DTx research, yet they are essential for moving beyond efficacy demonstrations toward a precision-oriented intervention design. We aimed to (1) identify the clinical and behavioral predictors of response across progressively stringent pain-reduction thresholds, (2) evaluate whether Oral Behaviors Checklist (OBC)–defined behavioral changes mediated the treatment effect, and (3) examine the temporal response dynamics and moderating role of baseline psychological distress. As behavioral change and Axis II domains were not randomized exposures, all analyses were interpreted as associative and hypothesis-generating.

Methods

Trial Design

This study is a post hoc analysis of a previously published multicenter, double-blind, randomized, sham-controlled superiority trial (KCT0009493), in which an active TMD DTx app was compared with a visually identical sham app over 6 weeks in a 1:1 allocation at 2 tertiary care oral and maxillofacial surgery clinics in South Korea (Hallym University Sacred Heart Hospital and Hallym University Dongtan Sacred Heart Hospital), with assessments at baseline and weeks 2, 4, and 6 [14]. Detailed eligibility criteria, randomization procedures, intervention content, blinding, ethical approvals, and primary efficacy analyses are reported in the original publication [14]. The present analysis focused on response heterogeneity within the trial by examining predictors of clinically meaningful

XSL•FO
RenderX

response, behavioral mediators, and moderators of treatment effect. Patients and members of the public were not involved in the design, conduct, or reporting of the parent trial or the present secondary analysis.

Sample Size, Randomization, and Blinding

No separate sample size calculation was performed for this post hoc analysis; the analytic sample was determined by the parent trial, which enrolled 102 participants based on a target of 38 participants per group (2-tailed α=.05, 90% power) with an allowance for 25% dropout. In the parent trial, participants were randomly assigned (1:1) by site using stratified block randomization with an independently generated allocation list, and both participants and investigators were blinded to group assignment. The sham application was visually and functionally identical to the active DTx but lacking therapeutic content. Full procedures are reported in the parent trial [14].

Interventions

Both groups received treatment as usual and were provided with a smartphone app for 6 weeks. The active DTx (Clickless DTx TMD-01) delivered TMD-related education, guided jaw exercises, behavioral habit tracking, guided meditation, and medication guidance with reminder alarms. The sham application was visually and functionally identical in interface and navigation but lacked all therapeutic content, permitting only weekly input of passive user data. Components common to both groups included a jaw health diary for symptom self-monitoring, visual feedback via weekly reports, and supportive check-in calls during weeks 1, 3, and 5, so that the 2 applications differed only in therapeutic content while self-monitoring, feedback, and human contact were held constant. Both applications were locked before trial initiation, and participants accessed them on personal smartphones, with daily use recommended throughout the 6-week period. Full details of intervention development and content are reported in the parent trial and Multimedia Appendix 1 [14].

Ethical Considerations

This manuscript reports a post hoc secondary analysis of a multicenter, sham-controlled randomized clinical trial. The present secondary analysis was reported in accordance with the CONSORT (Consolidated Standards of Reporting Trials) 2025 guidelines and the CONSORT-EHEALTH (V1.6.1; CONSORT of Electronic and Mobile Health Applications and Online Telehealth, Checklist 1) extension for web-based and mobile health interventions [21,22]. The parent trial was approved by the Institutional Review Boards of Hallym University Sacred Heart Hospital (HALLYM IRB 2024-04-001-006) and Hallym University Dongtan Sacred Heart Hospital (HDT IRB 2024-04-001-004) and was conducted under the supervision of the Korean Ministry of Food and Drug Safety (MFDS) in accordance with the Declaration of Helsinki and the International Council for Harmonization Good Clinical Practice (ICH-GCP) guidelines. The parent trial was prospectively registered on the Clinical Research Information Service (KCT0009493; registration date, May 31, 2024), before enrollment of the first participant on June 20, 2024 [14]. All participants provided written informed consent face-to-face prior to enrollment in the parent trial. The institutional review boards granted an exemption from review for the present secondary analysis of deidentified data, which did not require additional informed consent. Participant data were pseudonymized and stored in a secure, password-protected system accessible only to authorized investigators, with no personally identifiable information included in the analysis. Participants received compensation of approximately KRW 500,000 (Korean won; approximately US $350; KRW 1=US $0.00072 as of June 20, 2024) upon trial completion. All data were managed and locked by an independent contract research organization before the analysis to ensure data integrity and analytical independence. No images in this manuscript or its supplementary materials allow for the identification of individual participants.

Data Sources

This analysis integrated trial-derived datasets comprising DTx usage logs, module-completion records, and clinical outcome files. Data were cleaned, standardized by participant identifiers and visit labels, and merged using pseudonymized patient IDs. Engagement variables were derived from server logs over the 6-week intervention period and linked to repeated clinical measurements at each scheduled visit.

Outcomes

App engagement was quantified using objective usage and compliance logs. The variables included total app logins, exercise completion counts, meditation module use, jaw health diary entries, behavioral therapy module interactions, educational video views, article reads, notification checks, and pain record entries, all aggregated over the 42-day intervention period.

Clinical outcomes were assessed at 4 scheduled time points: baseline (V1), week 2 (V2), week 4 (V3), and week 6 (V4; end point). Pain intensity was measured using a Visual Analog Scale (VAS; 0 - 100 mm). The jaw function was assessed by a blinded examiner using the maximum mouth opening (MMO; mm) and the Jaw Functional Limitation Scale-20 (JFLS-20). Behavioral and psychological domains were assessed using the OBC and Patient Health Questionnaire-4 (PHQ-4) and scored according to standard procedures for the DC/TMD [23,24].

Statistical Analysis

Statistical analyses were performed on the per-protocol population (N=93; DTx: n=44; sham: n=49). Because the per-protocol population was defined a priori to exclude participants with missing primary outcome data and no item-level missing data were observed among completers, all analyses were conducted on complete cases without imputation, consistent with the parent trial protocol [14]. Normality was evaluated using the Shapiro-Wilk test. Between-group comparisons were performed using independent-samples *t* tests or Wilcoxon rank-sum tests as appropriate. A 2-sided significance threshold of α=.05 was used. No adjustments for multiple comparisons were made, which was consistent with the exploratory framework, and effect sizes with 95% CIs were reported. Analyses were conducted using Stata (version 17;

XSL•FO
RenderX

StataCorp LLC) and Python (version 3.14; Python Software Foundation) with the *pandas*, *SciPy*, and *Statsmodels* packages.

Responder Definition and Logistic Regression

Responder status was defined by ≥30%, ≥50%, and ≥70% reduction in pain VAS score from baseline to week 6. Logistic regression models were used to estimate the odds of response for each threshold, with treatment assignment (DTx=1 and sham=0), age, and change in MMO, OBC, and PHQ-4 scores as independent variables. The results were reported as odds ratios (ORs) with 95% CIs.

OBC Behavioral Modifier Classification

Within the DTx arm (n=44), participants were classified as OBC modifiers (n=21) or nonmodifiers (n=23) based on the median OBC change score from baseline to week 6 (median Δ=−8). Participants with OBC changes below the median (ie, greater behavioral improvement) were designated as modifiers. Independent-samples *t* tests were used to compare clinical outcomes and DTx usage metrics between the subgroups.

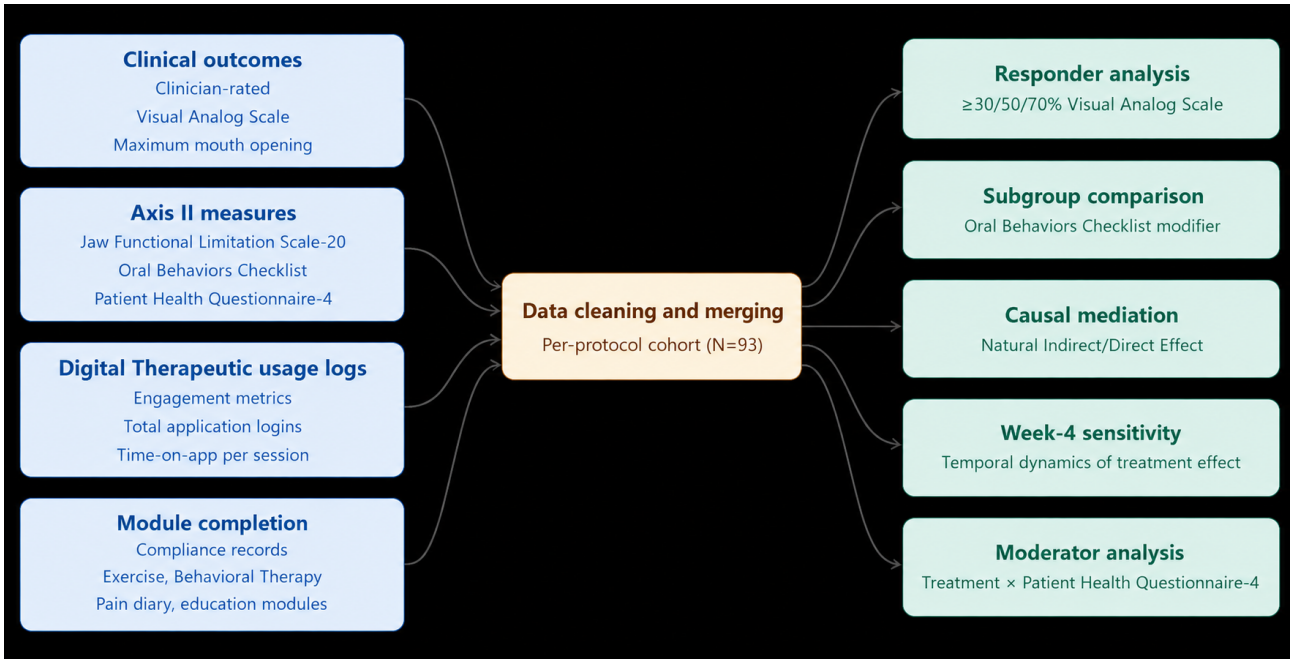
Mediation Analysis

Causal mediation analysis used the potential outcomes framework as implemented in the Mediate command in Stata (StataCorp LLC) [25,26]. The treatment variable was DTx assignment (vs sham), the mediator was OBC modifier status (binary), and the outcomes were (1) continuous VAS change from baseline to week 6 and (2) binary responder status at each threshold. The natural indirect effect (NIE), natural direct effect (NDE), and total effect were estimated using bootstrap-based 95% CIs (1000 iterations). This framework relies on several identifying assumptions, including no unmeasured treatment-mediator, mediator-outcome, or treatment-outcome confounding, no mediator-outcome confounder affected by treatment, and correct model specification. Because treatment was randomized, confounding of the treatment-mediator and treatment-outcome paths was minimized by design, leaving mediator-outcome confounding as the principal residual concern. Given the nonrandomized nature of the mediator, results were interpreted as exploratory and associative rather than confirmatory causal inference.

Sensitivity and Supplementary Analyses

Logistic regression models were reestimated using week 4 change scores to characterize early treatment effects and their temporal dynamics. Probit regression models were used for robustness checks. A linear regression moderator analysis examined the treatment×PHQ-4 High (≥3) interaction on VAS change. The multidomain responder analysis (triple responder: VAS +OBC improvement+MMO improvement) is shown in Multimedia Appendix 2. Figure 1 summarizes the analytical workflow, linking trial-derived data sources to the per-protocol cohort (N=93) and the 5 complementary analyses described below.

Figure 1. Analytical workflow of the post hoc study. The workflow shows the integration of clinical outcomes, Axis II measures, DTx usage logs, and module-completion records, followed by 5 complementary analyses: responder analysis, OBC modifier subgroup comparison, causal mediation analysis, week-4 sensitivity analysis, and moderator analysis. DC/TMD: Diagnostic Criteria for Temporomandibular Disorders; DTx: digital therapeutic; JFLS-20: Jaw Functional Limitation Scale-20; MMO: maximum mouth opening; OBC: Oral Behaviors Checklist; PHQ-4: Patient Health Questionnaire-4; TMD: temporomandibular disorders; VAS: Visual Analog Scale.



Results

Study Population and Primary Trial Outcomes

Of the 102 randomized participants, 93 met the per-protocol criteria (DTx: n=44; sham: n=49) and were included in the analysis. As previously reported, the DTx group demonstrated significantly greater improvements at week 6 in VAS pain (between-group difference, −23.78 mm; *P*<.001), MMO (+4.93 mm; *P*<.001), JFLS-20 (−19.68; *P*=.001), and OBC (−5.84; *P*=.008), with no significant difference in PHQ-4 (−0.36; *P*=.40). Baseline characteristics and week-6 outcomes are provided in

XSL•FO
RenderX

Multimedia Appendices 3 and 4. Harms were systematically monitored in the parent trial, in which no serious harms or adverse events were reported in either group. The present secondary analysis did not include an additional assessment of harms.

Factors Associated With Clinically Meaningful Pain Reduction

Across all 3 response thresholds, DTx assignment was a strong predictor of responder status (Table 1). The adjusted OR for DTx was 5.39 (95% CI 1.72 - 16.94; *P*<.01) at the ≥30% threshold, 3.21 (95% CI 1.14 - 8.99; *P*=.03) at ≥50%, and 3.45 (95% CI 1.12 - 10.63; *P*=.03) at ≥70%. MMO improvement was independently associated with higher odds of response at all thresholds (ORs 1.13 - 1.15; all *P*<.05). Changes in PHQ-4 scores and age were not significant predictors at any threshold.

Table . Logistic regression: factors associated with responder status at ≥30%, ≥50%, and ≥70% VAS[a] reduction thresholds (N=93).

Variable[b,c]	≥30%, OR[d] (95% CI)	*P* value	≥50%, OR (95% CI)	*P* value	≥70%, OR (95% CI)	*P* value
Treatment (DTx[e]=1)	5.39 (1.72-16.94)	<.01[f]	3.21 (1.14-8.99)	.03[f]	3.45 (1.12-10.63)	.03[f]
Age	1.01 (0.97-1.06)	.66	0.99 (0.95-1.03)	.60	0.99 (0.95-1.03)	.64
MMO[g] change	1.14 (1.03-1.26)	.01[f]	1.15 (1.04-1.27)	<.01[f]	1.13 (1.02-1.26)	.02[f]
OBC[h] change	0.94 (0.88-0.99)	.03[f]	0.94 (0.89-0.99)	.03[f]	0.91 (0.85-0.98)	.01[f]
PHQ-4[i] change	1.02 (0.77-1.36)	.87	1.04 (0.79-1.37)	.78	0.77 (0.57-1.03)	.08
Pseudo R^2	0.27	—[j]	0.25	—	0.32	—

[a]VAS: Visual Analog Scale.

[b]Post hoc secondary analysis of a multicenter, double-blind, sham-controlled randomized clinical trial (KCT0009493) evaluating a DTx for temporomandibular disorders, conducted at 2 tertiary care centers in the Republic of Korea from June 2024 to June 2025, with a per-protocol cohort of 93 adults (DTx, n=44; sham, n=49).

[c]Adjusted odds ratios (95% CI) are shown for treatment assignment, age, and week-6 change in MMO, OBC, and PHQ-4 as factors associated with responder status at ≥30%, ≥50%, and ≥70% VAS pain-reduction thresholds.

[d]OR: odds ratio.

[e]DTx: digital therapeutic.

[f]*P*<.05.

[g]MMO: maximum mouth opening.

[h]OBC: Oral Behaviors Checklist.

[i]PHQ-4: Patient Health Questionnaire-4.

[j]Not applicable.

OBC change exhibited a threshold-dependent gradient: each unit reduction in OBC score was associated with increased odds of response (OR 0.94 at ≥30% and ≥50%; *P*=.03; OR 0.91 at ≥70%; 95% CI 0.85 - 0.98; *P*=.01). The association strengthened as the response criterion became more stringent, with improved model fit (Pseudo R^2=0.32 at ≥70% vs 0.25-0.27 at lower thresholds). Estimates were robust to model specification, with probit regression yielding directionally consistent treatment effects across all thresholds (Multimedia Appendix 5).

Under a stricter multidomain responder definition requiring simultaneous VAS, OBC, and MMO improvement, DTx participants showed 8.55-12.32-fold higher odds of triple response than sham across thresholds (all *P*<.01; Multimedia Appendix 2).

OBC Behavioral Modifier Subgroup: Clinical Outcomes and DTx Usage

Within the DTx arm, OBC modifiers (n=21) showed greater pain reduction than nonmodifiers (n=23; VAS change: mean −45.71, SD 17.34 vs mean −22.61, SD 26.75 mm; difference: −23.11 mm; 95% CI −36.97 to −9.24; *P*<0.01) and greater improvement in jaw function (JFLS-20: mean −42.52, SD 26.28 vs mean −22.22, SD 28.72; difference −20.31; 95% CI −37.11 to −3.50; *P*=.02). PHQ-4 improvement did not differ significantly between the subgroups (difference −1.07, 95% CI −2.16 to 0.01; *P*=.05). MMO change also showed no significant difference (difference −0.64, 95% CI −3.88 to 2.61; *P*=.69).

Despite these clinical differences, no significant between-group differences were observed across any objective DTx usage metric, including total logins, exercise completions, meditation sessions, jaw diary entries, educational video views, or behavioral therapy module uses (all *P*>.05), with a nonsignificant trend toward fewer behavioral therapy entries among the modifiers (*P*=.10; Table 2). The waterfall plot illustrates that the OBC modifier cases were concentrated among participants with the greatest VAS reduction (Figure 2).

XSL•FO
RenderX

Table . Clinical outcomes and DTx[a] usage metrics by OBC[b] modifier status (DTx arm only; n=44).

Variable	OBC modifier[c] (n=21), mean (SD)	Nonmodifier (n=23), mean (SD)	Difference (95% CI)	*P* value
VAS[d] change (mm)	–45.71 (17.34)	–22.61 (26.75)	–23.11 (–36.97 to –9.24)	<.01[e]
MMO[f] change (mm)	6.19 (5.09)	6.83 (5.53)	–0.64 (–3.88 to 2.61)	.69
JFLS-20[g] change	–42.52 (26.28)	–22.22 (28.72)	–20.31 (–37.11 to –3.50)	.02[e]
PHQ-4[h] change	–1.33 (1.56)	–0.26 (1.96)	–1.07 (–2.16 to 0.01)	.05
App usage metrics				
Total app logins (42 days)	66.48 (35.31)	61.30 (18.08)	5.17 (–11.67 to 22.02)	.54
Passive exposure				
Educational video	6.86 (5.92)	5.91 (4.98)	0.94 (–2.37 to 4.26)	.57
Article	7.00 (16.02)	7.65 (11.32)	–0.65 (–9.03 to 7.73)	.88
Notification check	4.86 (5.83)	4.48 (5.91)	0.38 (–3.20 to 3.96)	.83
Active practice				
Exercise	133.81 (96.69)	146.43 (86.92)	–12.63 (–68.48 to 43.23)	.65
Meditation	29.71 (46.17)	22.61 (24.85)	7.11 (–15.18 to 29.39)	.52
Self-monitoring and regulation				
Jaw health diary	29.95 (10.95)	32.39 (6.51)	–2.44 (–7.91 to 3.02)	.37
Pain record	4.81 (11.76)	3.83 (7.78)	0.98 (–5.03 to 7.00)	.74
Behavioral therapy	4.29 (7.80)	12.87 (22.20)	–8.58 (–18.91 to 1.74)	.10

[a]DTx: digital therapeutic.

[b]OBC: Oral Behaviors Checklist.

[c]OBC modifiers (n=21) were defined as DTx participants with OBC change scores below the median (ΔOBC=–8, greater behavioral improvement) and were compared with nonmodifiers (n=23) on clinical outcomes and objective engagement metrics using independent-samples *t* tests. For descriptive clarity, engagement metrics are grouped by conceptual domain; all comparisons were performed at the individual metric level.

[d]VAS: Visual Analog Scale.

[e]*P*<.05.

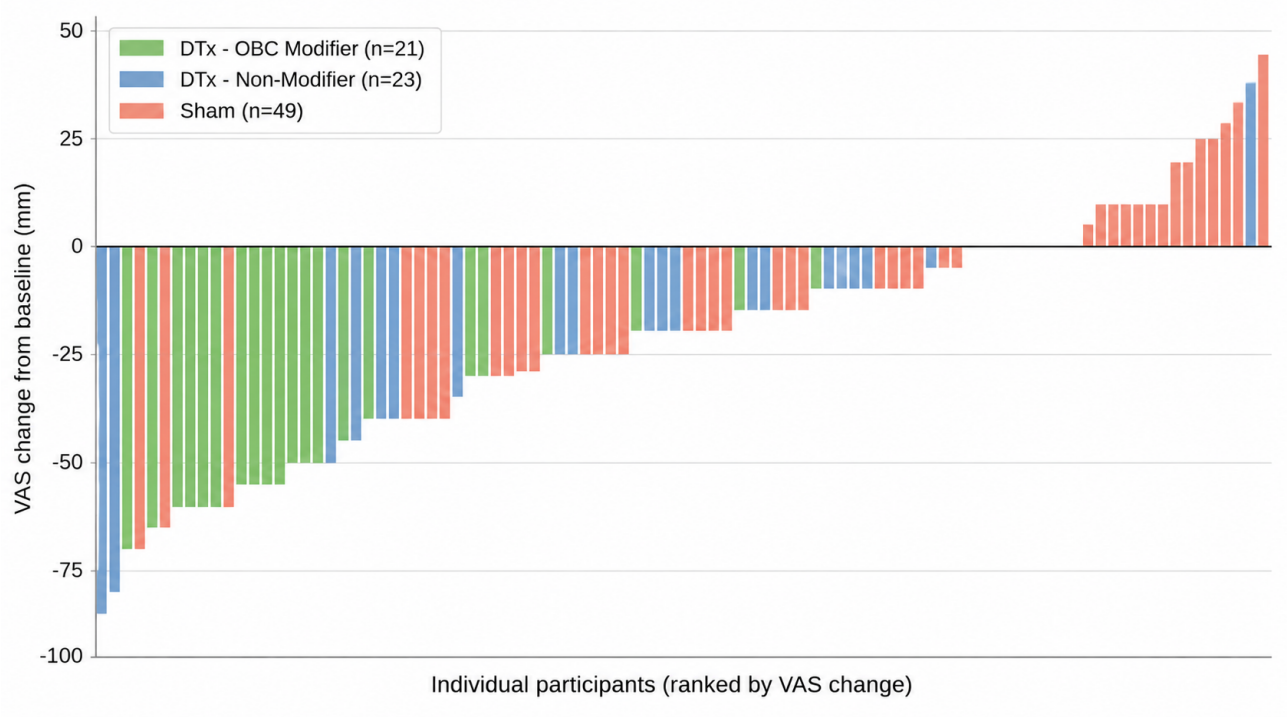
[f]MMO: Maximum Mouth Opening.

[g]JFLS-20: Jaw Functional Limitation Scale-20.

[h]PHQ-4: Patient Health Questionnaire-4.

XSL•FO
RenderX

Figure 2. Individual Visual Analog Scale (VAS) changes from baseline ranked by magnitude across treatment subgroups. Each vertical bar represents one participant's VAS pain change from baseline to week 6, ranked from greatest reduction (left) to smallest reduction or increase (right). Participants are color-coded by treatment allocation and Oral Behaviors Checklist (OBC) modifier status within the digital therapeutic (DTx) arm: DTx-OBC modifier (green; n=21), DTx-nonmodifier (blue; n=23), and sham controls (peach; n=49). Negative values indicate pain reduction. DTx: digital therapeutic; OBC: Oral Behaviors Checklist; TMD: temporomandibular disorders; VAS: Visual Analog Scale.



Mediation Analysis: OBC Behavioral Modification as a Mediator of Treatment Effect

For continuous VAS change, the NIE through OBC modifier status was −6.91 mm (95% CI −13.23 to −0.58; *P*=.03), indicating that approximately 29.1% of the total treatment effect (−23.78 mm) was mediated through behavioral modification. The NDE remained significant (−16.87 mm; 95% CI −27.30 to −6.45; *P*<.01).

A threshold-dependent pattern emerged for the binary outcomes. The NIE was 0.09 (95% CI −0.01 to 0.18; *P*=.07) at ≥30%, 0.11 (95% CI −0.00 to 0.22; *P*=.05) at ≥50%, and 0.13 (95% CI 0.01 - 0.25; *P*=.03) at ≥70%, with the mediated proportion increasing from 19.6% to 25.6% to 30.2%. The NDE remained significant across all models (all *P*<.01; Table 3 and Figure 3).

XSL•FO
RenderX

Table . Causal mediation analysis: treatment effect on VAS[a] change and responder status mediated by OBC[b] behavioral modification (N=93).

Effect[c]	VAS change coefficient (95% CI)	*P* value	≥30% Coeff (95% CI)	*P* value	≥50% Coefficient (95% CI)	*P* value	≥70% Coefficient (95% CI)	*P* value
NIE[d] (via OBC)	–6.91 (–13.23 to –0.58)	.03[e]	0.09 (–0.01 to 0.18)	.07	0.11 (–0.00 to 0.22)	.05	0.13 (0.01 to 0.25)	.03[e]
NDE[f] (treatment→outcome)	–16.87 (–27.30 to –6.45)	<.01[e]	0.37 (0.16 to 0.57)	<.01[e]	0.32 (0.11 to 0.53)	<.01[e]	0.30 (0.10 to 0.49)	<.01[e]
Total effect	–23.78 (–33.92 to –13.63)	<.01[e]	0.46 (0.28 to 0.63)	<.01[e]	0.43 (0.24 to 0.61)	<.01[e]	0.43 (0.25 to 0.61)	<.01[e]
Mediation (%)	29.1	—[g]	19.6	—	25.6	—	30.2	—

[a]VAS: Visual Analog Scale.

[b]OBC: Oral Behaviors Checklist.

[c]NIE, NDE, total effect, and proportion mediated are shown for continuous week-6 VAS change and binary responder status at ≥30%, ≥50%, and ≥70% VAS pain-reduction thresholds. Estimates were derived from the potential outcomes framework with 1000-iteration bootstrap 95% CIs.

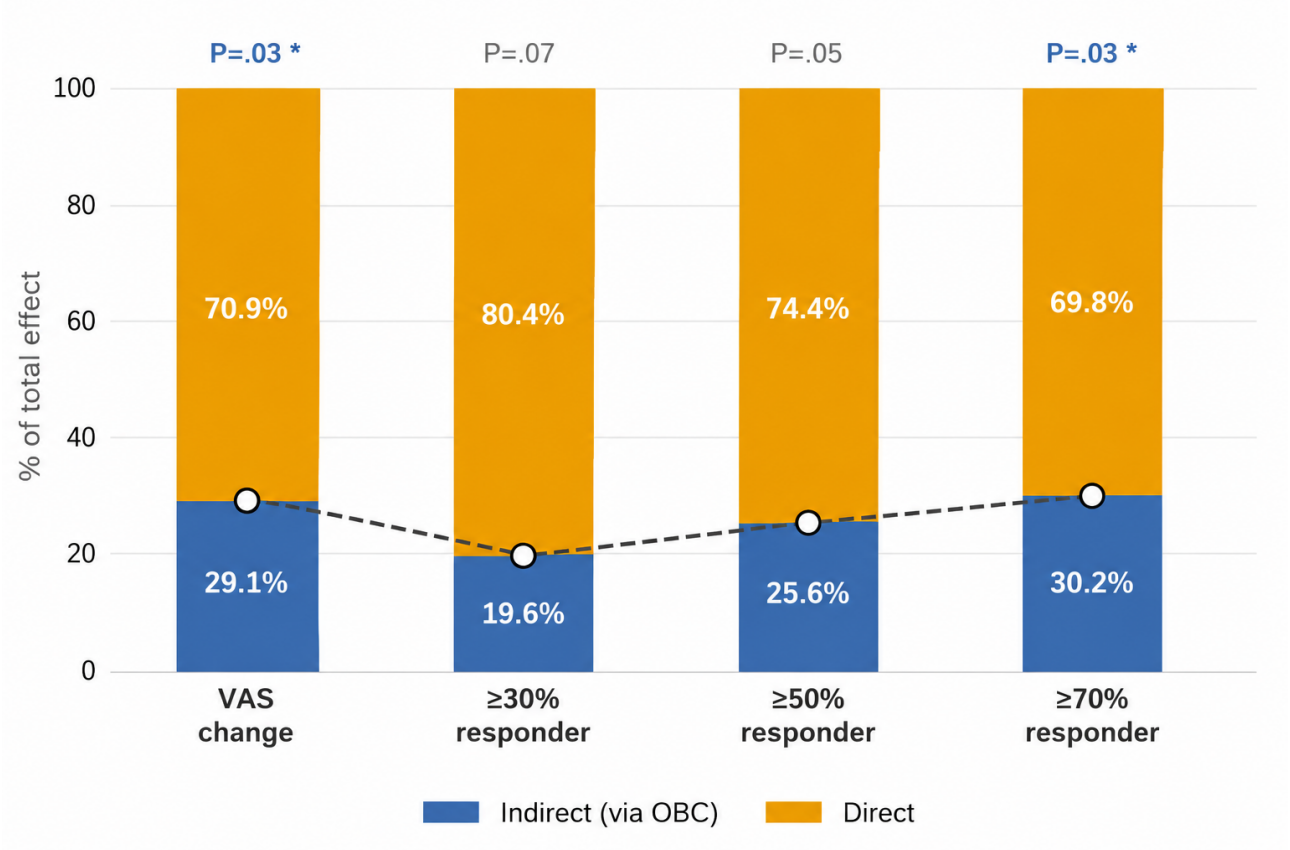
[d]NIE: natural indirect effect.

[e]*P*<.05.

[f]NDE: natural direct effect.

[g]Not available.

Figure 3. Mediation proportion of Oral Behaviors Checklist (OBC)–driven behavioral change on digital therapeutic (DTx) treatment effects across outcome thresholds. The stacked bars show the proportions of treatment effects mediated through OBC change (indirect) and independent of OBC change (direct) across continuous VAS change and VAS responder thresholds of ≥30%, ≥50%, and ≥70%. DTx: digital therapeutic; OBC: Oral Behaviors Checklist; TMD: temporomandibular disorders; VAS: Visual Analog Scale.



XSL•FO
RenderX

Sensitivity Analysis: Early Response at Week 4

When logistic regression models were reestimated using week 4 change scores, treatment ORs were consistently higher than at the 6-week end point across all thresholds: 7.15 (95% CI 2.39 - 21.32; *P*<.01) vs 5.39 at ≥30%, 4.48 (95% CI 1.68 - 11.94; *P*<.01) vs 3.21 at ≥50%, and 4.91 (95% CI 1.74 - 13.87; *P*<.01) vs 3.45 at ≥70%, representing increases of 33%, 40%, and 42%, respectively (Table 4, Figures 4A and 4B).

OBC change at Week 4 was significant for the ≥70% threshold (OR 0.92, 95% CI 0.87 - 0.99; *P*=.02) but not at lower thresholds, indicating the behavioral effect matures between weeks 2 and 4. MMO change remained significant across all thresholds at week 4 (ORs 1.10 - 1.11; all *P*<.05). The change in PHQ-4 score was not significant in any model. Week-4 estimates were similarly robust under probit regression, with significant treatment effects retained across all thresholds (Multimedia Appendix 6).

Table . Logistic regression: week 4 (V3-base) change scores as predictors of responder status (N=93).

Variable	≥30%[a], OR[b] (95% CI)	*P* value	≥50%, OR (95% CI)	*P* value	≥70%, OR (95% CI)	*P* value
Treatment (DTx[c]=1)	7.15 (2.39-21.32)	<.01[d]	4.48 (1.68-11.94)	<.01[d]	4.91 (1.74-13.87)	<.01[d]
Age	1.01 (0.97-1.05)	.71	0.99 (0.95-1.03)	.60	0.99 (0.95-1.04)	.69
MMO[e] change (V3-base)	1.10 (1.00-1.21)	.04[d]	1.11 (1.01-1.21)	.03[d]	1.10 (1.00-1.20)	.04[d]
OBC[f] change (V3-base)	0.95 (0.89-1.01)	.09	0.95 (0.89-1.00)	.07	0.92 (0.87-0.99)	.02[d]
PHQ-4[g] change (V3-base)	0.99 (0.74-1.33)	.97	0.96 (0.72-1.28)	.79	0.82 (0.61-1.11)	.20
Pseudo R^2	0.24	—[h]	0.22	—	0.27	—

[a]Adjusted odds ratios (95% CI) are shown for treatment assignment, age, and week-4 change in MMO, OBC, and PHQ-4 as early treatment predictors of responder status at ≥30%, ≥50%, and ≥70% Visual Analog Scale pain-reduction thresholds. Results are compared with the week-6 end point estimates in Table 1.

[b]OR: odds ratio.

[c]DTx: digital therapeutic.

[d]*P*<.05.

[e]MMO: Maximum Mouth Opening.

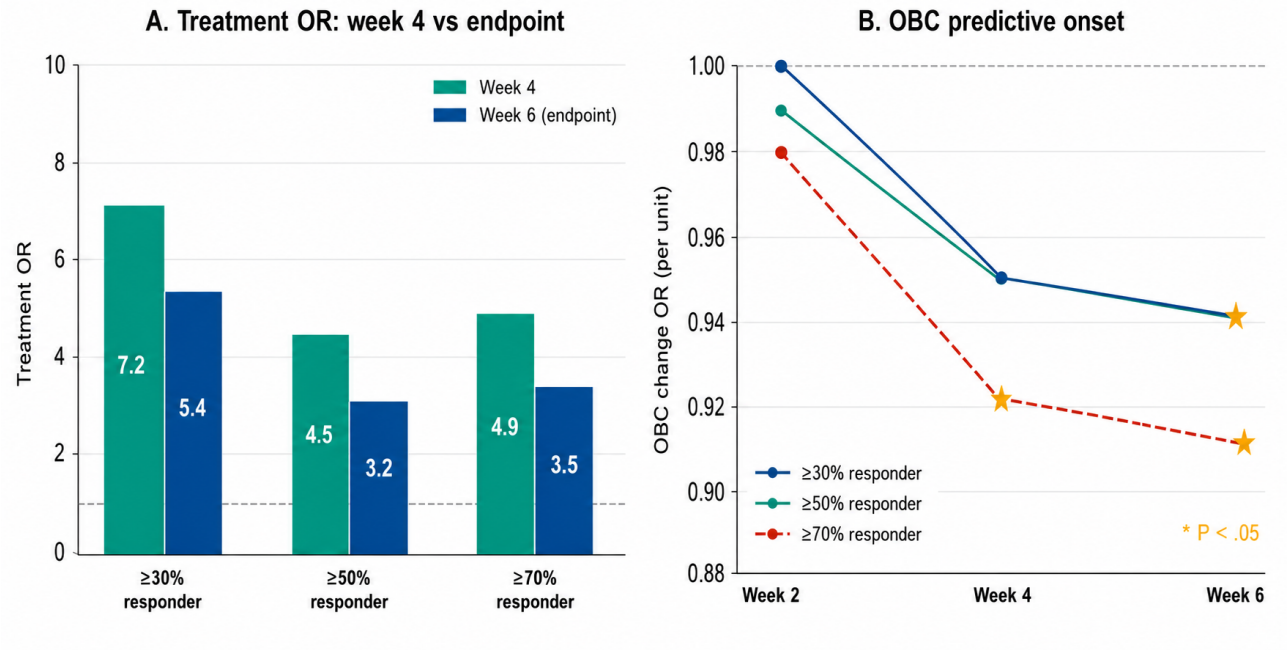
[f]OBC: Oral Behaviors Checklist.

[g]PHQ-4: Patient Health Questionnaire-4.

[h]Not applicable.

XSL•FO
RenderX

Figure 4. Treatment odds ratios (ORs) at week 4 versus the end point and temporal onset of Oral Behaviors Checklist (OBC) behavioral change across responder thresholds. (A) Treatment odds ratios at week 4 and week 6 across Visual Analog Scale (VAS) pain-reduction responder thresholds of ≥30%, ≥50%, and ≥70%. (B) Temporal changes in the predictive association between OBC behavioral change and VAS responder status across weeks 2, 4, and 6. DTx: digital therapeutic; OBC: Oral Behaviors Checklist; OR: odds ratio; TMD: temporomandibular disorders; VAS: Visual Analog Scale.



Exploratory Moderator Analysis: Baseline Psychological Distress

To explore whether baseline psychological distress moderated the treatment effect, a linear regression model with an interaction term (Treatment×PHQ-4 High [≥3]) was used with VAS change as the dependent variable (Table 5). A PHQ-4 score of ≥3 was applied as the binary threshold, consistent with established cutoff criteria indicating clinically meaningful anxiety and/or depressive symptoms [27,28]. Of the 93 participants, 65 had PHQ-4 ≥3 and 28 had PHQ-4 <3 at baseline.

In the adjusted model controlling for baseline VAS, age, and sex, the interaction term was significant (β=−19.63; *P*=.04), indicating the DTx effect on VAS change differed by baseline distress level. Among participants with PHQ-4 ≥3, the estimated DTx effect on VAS change was −28.09 mm (sum of treatment coefficient and interaction term: −8.46 + [−19.63]), representing greater pain reduction compared with those with PHQ-4 <3, in whom no significant treatment effect was observed (β=−8.46 mm; *P*=.25). The main effect of PHQ-4 High was significant (β=14.91; *P*=.01), indicating that higher baseline psychological distress was independently associated with greater VAS reduction regardless of treatment assignment. The adjusted model explained 49% of the variance in VAS score change (R^2=0.49), compared with 24% in the unadjusted model (R^2=0.24). As a sensitivity analysis, the interaction was not significant when PHQ-4 was modeled as a continuous variable (β=−4.02 per point; *P*=.06). The CONSORT flow diagram for the overall clinical trial is presented in Figure 5.

XSL•FO
RenderX

Table . Moderator analysis: treatment×baseline Patient Health Questionnaire-4 (PHQ-4) interaction on Visual Analog Scale (VAS) change (N=93).

Variable[a]	Unadjusted β (SE)	95% CI	*P* value	Adjusted β (SE)	95% CI	*P* value
Treatment ×PHQ-4 High	–26.62 (10.75)	–47.98 to –5.26	<.01[b]	–19.63 (9.17)	–37.86 to −1.39	.04[b]
Treatment (DTx[c]=1)	–7.15 (8.65)	–24.33 to 10.02	.41	–8.46 (7.34)	–23.04 to 6.12	.25
PHQ-4[d] High (≥3)	16.00 (7.07)	(1.95 to 30.05)	.03[b]	14.91 (5.97)	3.04 to 26.77	.01[b]
VAS[e] baseline	—[f]	—		–0.68 (0.11)	–0.90 to –0.46	<.01[b]
Age	—	—		0.29 (0.18)	–0.07 to 0.65	.11
Sex (female=1)	—	—		–6.14 (4.66)	–15.32 to 3.03	.19
R^2	0.24	—		0.49	—	

[a]Linear regression estimates (β, SE, and 95% CI) are shown for the Treatment×PHQ-4 High (≥3) interaction on VAS change from baseline to week 6, adjusted for baseline VAS, age, and sex. Baseline PHQ-4 ≥3 (n=65) versus PHQ-4 <3 (n=28) was used as a validated cut-point indicating clinically meaningful anxiety and/or depressive symptoms.

[b]P<.05.

[c]DTx: digital therapeutic.

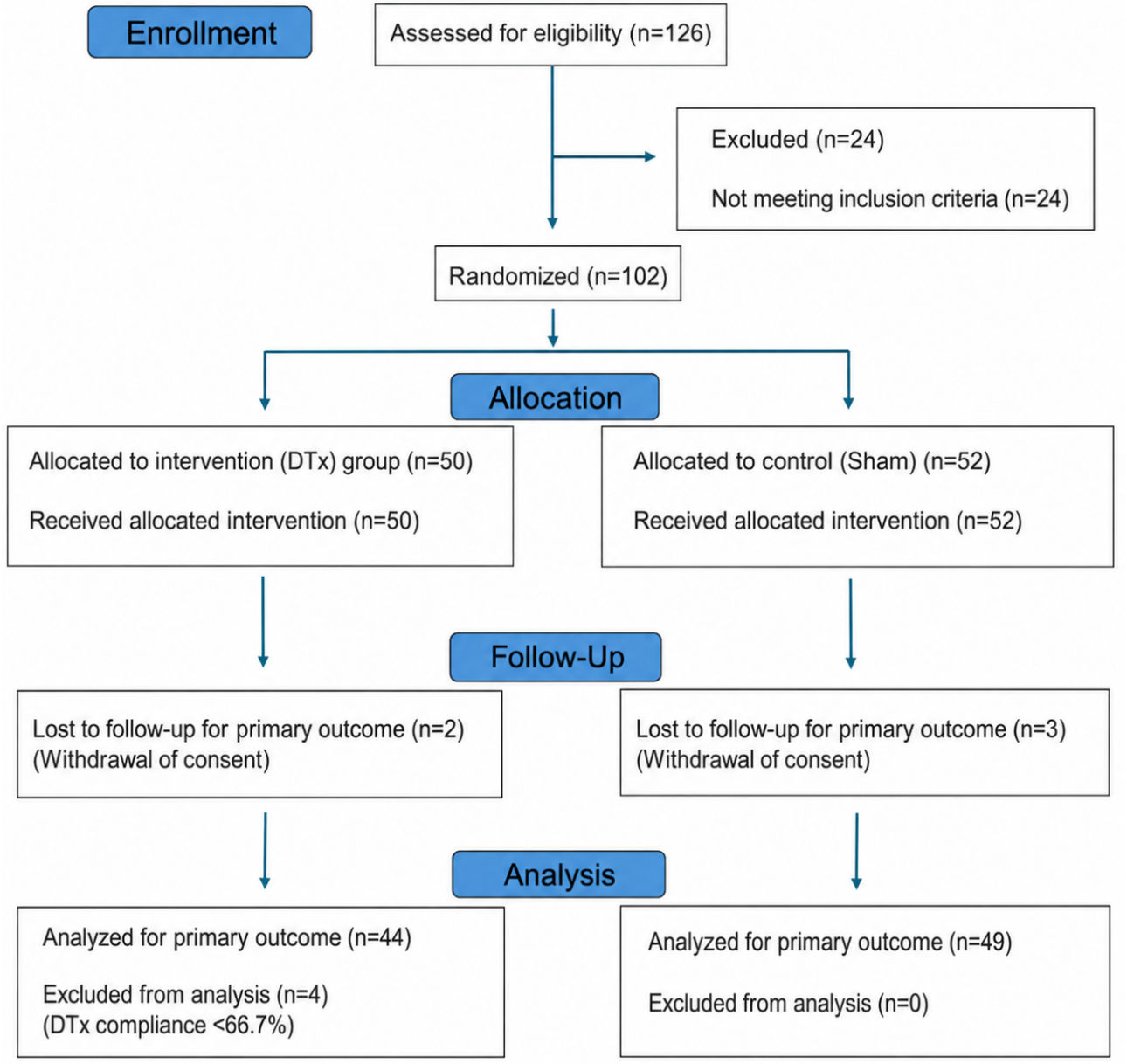
[d]PHQ-4: Patient Health Questionnaire-4.

[e]VAS: Visual Analog Scale.

[f]Not available.

XSL•FO
RenderX

Figure 5. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of the randomized controlled trial. DTx: digital therapeutic.



Discussion

Principal Findings

This post hoc analysis of a multicenter, sham-controlled randomized trial of a TMD DTx extends the parent trial's demonstration of aggregate efficacy to the mechanisms and heterogeneity of treatment response. Building on the confirmed clinical benefit of the DTx, we examined which clinical and behavioral factors were associated with pain response, whether oral behavior change might mediate the treatment effect, and how the treatment effect evolved over time and varied with baseline psychological distress. Across all response thresholds, DTx assignment remained a strong and consistent predictor of clinically meaningful pain relief relative to sham. Three exploratory patterns emerged, including oral behavior modification appeared as a possible mediating pathway, the treatment effect appeared to be accelerated and amplified, with responders reaching clinically meaningful pain relief earlier and to a greater degree, and baseline psychological distress may play a modest moderating role. As all analyses were performed post hoc, these patterns were interpreted as associative and hypothesis-generating and are elaborated below.

In the present analysis, OBC change was the most consistent independent factor associated with response, with the association strengthening at progressively more stringent response thresholds. This gradient may indicate that behavioral modification becomes increasingly important as pain relief increases. Mediation analysis was consistent with this pattern. The proportion of the VAS treatment effect potentially transmitted via behavioral modification rose monotonically across increasingly stringent response thresholds, reaching approximately 29% on the continuous VAS scale. This positive dose-response relationship, in which greater behavioral change accompanied greater pain reduction, is consistent with OBC behavioral modification potentially acting as a partial mediating pathway rather than solely as a correlate.

XSL•FO
RenderX

A particularly informative pattern was the dissociation between OBC-defined behavioral change and objective app usage metrics. OBC modifiers achieved significantly greater pain reduction than nonmodifiers, yet no significant differences were observed across any engagement metrics, including logins, exercises, meditation sessions, or diary entries. The nonsignificant but directionally consistent pattern of fewer behavioral therapy violation entries among modifiers provides a triangulation across 3 independent measurement modalities, including self-report OBC, digital behavioral logs, and clinical pain VAS outcomes. Of note, the behavioral therapy module log records patient-reported breaches of behavioral recommendations; therefore, fewer entries indicate better behavioral adherence rather than reduced app use. This directionality aligns the digital log with the lower OBC scores and the greater pain reduction observed in modifiers, with all 3 modalities converging on the same pattern. This suggests that meaningful behavioral change, rather than passive app engagement, may be the more relevant driver of clinical improvement in this cohort [19]. This dissociation should nonetheless be interpreted cautiously, as the present sample may be underpowered to detect modest differences in engagement metrics.

Treatment ORs were consistently higher at week 4 than at the 6-week end point across all responder thresholds, which may suggest that DTx had its strongest discriminatory effect at this intermediate time point. The temporal onset of OBC as an associated marker of response, which was absent at week 2 but emerged for the ≥70% threshold by week 4, may suggest that meaningful behavioral modification requires approximately 4 weeks to consolidate and manifest clinically [29]. Taken together, these temporal patterns may suggest that DTx responders achieve an accelerated and amplified clinical improvement than nonresponders, reaching clinically meaningful pain relief earlier and to a greater degree, with behavioral modification as a plausible temporal correlate of this response.

Baseline psychological distress (PHQ-4 ≥3) appeared to modestly amplify the treatment effect, with a larger estimated DTx effect in the elevated-distress subgroup than in the subgroup without distress. However, the interaction was not significant when PHQ-4 was modeled continuously, and the ≥3 threshold was a predefined clinical cutoff rather than a data-derived optimum; these findings are exploratory and require prospective validation [27].

Behavioral Mechanism and the Realized-Dose Framework

The central role of OBC behavioral change aligns with and extends the biopsychosocial understanding of TMD pathophysiology [30]. Parafunctional oral behaviors, including clenching, bruxism, and habitual jaw tension, are recognized perpetuating factors that sustain mechanical loading on the temporomandibular joint and maintain pain-guarding cycles [30,31]. Longitudinal evidence from the Orofacial Pain: Prospective Evaluation and Risk Assessment (OPPERA) study associated high awake parafunction with both TMD onset and chronification [24,32]. Building on this prevalence-level evidence, our findings suggest that targeted modification of these parafunctional behaviors may yield measurable clinical improvement, moving the inference beyond a statistical between-group comparison toward a data-based mechanistic account of how parafunction-directed intervention may translate into pain reduction.

This dissociation is conceptually consistent with the framework proposed by Schmidt et al [19], who distinguish between "consumed dose," the patient's interaction with a DTx, and "realized dose," the degree to which cognitive-affective and behavioral change is actually internalized by the patient, and who argue that the therapeutic effect is not a direct function of consumed dose but rather of realized dose. This convergence is consistent with the view that therapeutic benefit may depend more on behavioral adaptation, the realized dose, than on engagement volume [19]. It thus raises questions about the assumption that a greater consumed dose should predict clinical outcomes [18]. Our results provide a data-based, exploratory quantification of this distinction within a sham-controlled randomized clinical trial, suggesting that OBC-defined behavioral modification may account for approximately 29% of the total VAS treatment effect independently of any consumed dose metric. For DTx product development, optimization might prioritize facilitating real-world behavioral internalization through adaptive content delivery and habit-change reinforcement rather than maximizing session frequency or time-on-app [33-35]. These findings further caution against the use of engagement metrics such as login frequency or session duration, as surrogate indicators of adherence or therapeutic dose, given that meaningful behavioral change may occur independently of, or even inversely to, raw usage volume [19].

Temporal Dynamics of Response and Moderation by Psychological Distress

Because week-4 behavioral change precedes the week-6 end point, it can be interpreted as an early-treatment predictor with temporal precedence, in contrast to the concurrent change variables measured at the end point, which are more appropriately regarded as factors associated with response. The week-4 amplification has clear precedent in the broader habit-formation literature, where behavioral consolidation typically requires approximately 4 weeks to stabilize [29]. These findings may carry 2 practical implications. First, week 4 may represent a promising candidate time point at which patients who have not demonstrated a behavioral or clinical response by this time point can be identified early to receive augmented or alternative strategies before completing the full 6-week course. However, because these observations are post hoc, prospective adaptive-design or decision-analytic studies would be needed before a week-4 rule could inform clinical practice [36]. These adaptive treatment approaches are increasingly advocated within digital health interventions [36]. Second, the amplified ORs raise the prospective hypothesis that a condensed 4-week protocol might achieve equivalent or near-equivalent gains with reduced patient burden, a question that warrants dedicated evaluation in future trials [37].

The PHQ-4 moderator pattern is clinically plausible: patients with psychological comorbidity may benefit more from the self-management, mindfulness, and behavioral awareness

XSL•FO
RenderX

components in the DTx intervention [38-40]. Despite evidence that 60% - 77% of patients with TMD experience psychological distress [3], routine screening remains underused in dental practice [41]. Brief instruments such as the PHQ-4 could help identify patients most likely to benefit from behavioral digital interventions [42,43].

Limitations

This study has several limitations. All findings are from a post hoc secondary analysis and are associative and hypothesis-generating, not causal. Because the proposed mediator, OBC-defined behavioral change, is a postrandomization variable rather than a randomized exposure, causal mediation inference is precluded despite the use of the potential outcomes framework, and residual unmeasured confounding of the mediator-outcome relationship cannot be excluded [44]. In particular, reverse causation cannot be ruled out: patients whose pain improves may subsequently report fewer parafunctional behaviors because they feel better, are less vigilant, or perceive less need to guard the jaw [3]. Under this scenario, OBC improvement could function as both a consequence and a correlate of pain reduction rather than solely as an antecedent mediator, and the assumed direction from behavioral change to pain relief, while biologically plausible, cannot be confirmed in the present design. Baseline OBC and PHQ-4 scores were higher in the DTx group, which may have influenced the mediation and moderator analyses through regression to the mean. The OBC modifier classification relies on a data-driven median split, introducing cut-point sensitivity, and creating subgroups that may not be stable across samples. Furthermore, multiple responder, subgroup, mediation, sensitivity, and moderator analyses were conducted without formal correction for multiple comparisons; borderline findings, particularly those with *P* values close to .05, should therefore be regarded as exploratory signals rather than confirmatory results. Finally, the mediation estimates were sensitive to distributional assumptions and sample sizes, necessitating replication in larger independent cohorts [44].

The analysis was conducted in the per-protocol population (N=93) rather than the intention-to-treat sample, as the mechanistic question of how DTx exerts its therapeutic effect inherently concerns participants who completed the intervention as intended; intention-to-treat estimates would have diluted the engagement-outcome dissociation with noncompleters and were therefore not appropriate for this specific analytic aim [45]. This principled selection nonetheless restricts inference to completers and should be considered when generalizing the mechanistic findings to broader trial populations. The per-protocol sample size also restricted the statistical power of the subgroup and interaction analyses, particularly for the PHQ-4 moderator, and limited the ability to detect modest between-group differences in engagement metrics. In addition, only a single behavioral mediator (OBC) was examined; other plausible pathways, such as cognitive reappraisal, pain-related self-efficacy, or improved sleep, were not assessed and may also contribute to treatment response [46-48]. The 6-week follow-up precluded the assessment of long-term durability, and a longitudinal follow-up study is currently underway. The trial was conducted at 2 tertiary care centers in South Korea, limiting its generalizability to other health care settings and populations.

Conclusion

These findings may refine the understanding of therapeutic dose in digital interventions. While app engagement may be a necessary vehicle, it may be the patient's internalization of that engagement into sustained behavioral change that is more closely linked to clinical benefit. In TMD, this realized behavioral adaptation, shaped by underlying psychological context and consolidated within a defined temporal window, may represent a key behavioral pathway underlying treatment response. Practically, these observations suggest that DTx development and evaluation might prioritize measures of behavioral realization over engagement volume and that mechanism-driven, precision-oriented trial designs may help clarify for whom and through which pathways these interventions work. These directions remain hypotheses to be confirmed in future prospective studies.

Acknowledgments

The authors thank the clinical research staff at Hallym University Sacred Heart Hospital and Hallym University Dongtan Sacred Heart Hospital for their contributions to the parent trial, and So Young Jung for trial coordination. The authors gratefully acknowledge Soo-Yeon Chung and Woo-Suk Kang of the contracted research organization for their rigorous data management and statistical oversight. All clinical data were managed and locked by an independent contract research organization under the supervision of the Korean Ministry of Food and Drug Safety prior to analysis. The present secondary mediation analysis received no separate funding. The authors also thank Professors Young-Kyun Kim and Jong-Ho Lee for valuable scientific input and mentorship.

Disclosure of Delegation to Generative AI (GenAI): The authors declare the use of generative AI in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: - Translation. The GenAI tool used was: Claude 3. Responsibility for the final manuscript lies entirely with the authors. The GenAI tools are not listed as authors and do not bear responsibility for the final outcomes. Declaration submitted by: The authors.

Funding

The authors declared no financial support was received for this work.

XSL•FO
RenderX

Data Availability

Deidentified individual participant data and a data dictionary will be made available to the researchers upon reasonable request from the corresponding author.

Authors' Contributions

Investigation: SHB, SWO, BEY, SMY
Writing – original draft: SHB, SWO
Supervision: SWO
Data curation: SAC
Visualization: SAC
Writing – review & editing: BEY, SYP
Methodology: YP, YK, SYP
Software: YP, YK
Formal analysis: YP, YK
Validation: YP, YK
Conceptualization: SYP
Project administration: SYP
Resources: SYP

Conflicts of Interest

SYP is the developer of the digital therapeutic intervention assessed in this trial and holds equity in BeyondMedicine (Seoul, South Korea). He was not involved in the clinical trial protocol development, data collection, or analysis but contributed to manuscript preparation. Apart from this disclosed interest, the remaining authors declare no competing interests.

Multimedia Appendix 1
Structure and components of the active digital therapeutic and sham applications in the parent randomized trial.
[PNG File, 410 KB - mhealth_v14i1e104264_app1.png]

Multimedia Appendix 2
Multidomain responder analysis: triple responder rates by treatment group triple responder=simultaneous improvement in all 3 domains: VAS pain reduction (≥threshold)+OBC improvement (ΔOBC ≤ median −4)+MMO improvement (ΔMMO≥median+4).
[DOCX File, 19 KB - mhealth_v14i1e104264_app2.docx]

Multimedia Appendix 3
Baseline demographic and clinical characteristics (per-protocol population).
[DOCX File, 19 KB - mhealth_v14i1e104264_app3.docx]

Multimedia Appendix 4
Change From Baseline to Week 6: DTx vs. Sham Group.
[DOCX File, 19 KB - mhealth_v14i1e104264_app4.docx]

Multimedia Appendix 5
Probit regression analysis of VAS responder status at week 6 (end point)-robustness check.
[DOCX File, 20 KB - mhealth_v14i1e104264_app5.docx]

Multimedia Appendix 6
Probit regression analysis of VAS responder status at week 4-sensitivity analysis.
[DOCX File, 17 KB - mhealth_v14i1e104264_app6.docx]

Checklist 1
Checklist 1: CONSORT-EHEALTH checklist (V1.6.1).
[PDF File, 492 KB - mhealth_v14i1e104264_app7.pdf]

References

1. Gauer RL, Semidey MJ. Diagnosis and treatment of temporomandibular disorders. Am Fam Physician 2015 Mar 15;91(6):378-386. [Medline: 25822556]

XSL•FO
RenderX

2. Schiffman E, Ohrbach R, Truelove E, et al. Diagnostic Criteria for Temporomandibular Disorders (DC/TMD) for clinical and research applications: recommendations of the International RDC/TMD Consortium Network and Orofacial Pain Special Interest Group. J Oral Facial Pain Headache 2014;28(1):6-27. [doi: 10.11607/jop.1151] [Medline: 24482784]
3. Yu Y, Wang L, Chen C, Wang M, Nie S, Fu L. Correlations of oral behaviors, pain, and psychological distress among patients with temporomandibular disorders: clinical investigation of axial II evaluation. Front Public Health 2025;13:1604995. [doi: 10.3389/fpubh.2025.1604995] [Medline: 40630410]
4. Viñals Narváez AC, Sánchez-Sánchez T, García-González M, et al. Psychological and behavioral factors involved in temporomandibular myalgia and migraine: common but differentiated profiles. Int J Environ Res Public Health 2023 Jan 14;20(2):1545. [doi: 10.3390/ijerph20021545] [Medline: 36674300]
5. Warzocha J, Gadomska-Krasny J, Mrowiec J. Etiologic factors of temporomandibular disorders: a systematic review of literature containing Diagnostic Criteria for Temporomandibular Disorders (DC/TMD) and Research Diagnostic Criteria for Temporomandibular Disorders (RDC/TMD) from 2018 to 2022. Healthcare (Basel) 2024 Feb 29;12(5):575. [doi: 10.3390/healthcare12050575] [Medline: 38470686]
6. Durham J, Newton-John TRO, Zakrzewska JM. Temporomandibular disorders. BMJ 2015 Mar 12;350:h1154. [doi: 10.1136/bmj.h1154] [Medline: 25767130]
7. Brighenti N, Battaglino A, Sinatti P, et al. Effects of an interdisciplinary approach in the management of temporomandibular disorders: a scoping review. Int J Environ Res Public Health 2023 Feb 4;20(4):2777. [doi: 10.3390/ijerph20042777] [Medline: 36833474]
8. Günther T, Schierz O, Hahnel S, Rauch A. Field-testing a psychosocial assessment scoring form for TMD patients - summarizing axis II instruments. BMC Oral Health 2020 Oct 1;20(1):267. [doi: 10.1186/s12903-020-01248-7] [Medline: 32998738]
9. Yao L, Sadeghirad B, Li M, et al. Management of chronic pain secondary to temporomandibular disorders: a systematic review and network meta-analysis of randomised trials. BMJ 2023 Dec 15;383:e076226. [doi: 10.1136/bmj-2023-076226] [Medline: 38101924]
10. Amyx M, Phi NTT, Alebouyeh F, Ravaud P, Tran VT. Mapping the evidence supporting digital therapeutics: a review. JAMA Intern Med 2024 Nov 1;184(11):1388-1390. [doi: 10.1001/jamainternmed.2024.4972] [Medline: 39348113]
11. Kim M, Patrick K, Nebeker C, et al. The digital therapeutics real-world evidence framework: an approach for guiding evidence-based digital therapeutics design, development, testing, and monitoring. J Med Internet Res 2024 Mar 5;26:e49208. [doi: 10.2196/49208] [Medline: 38441954]
12. Masanneck L, Stern AD. Tracing digital therapeutics research across medical specialties: evidence from ClinicalTrials.gov. Clin Pharma and Therapeutics 2024 Jul;116(1):177-185. [doi: 10.1002/cpt.3260]
13. Wang C, Lee C, Shin H. Digital therapeutics from bench to bedside. NPJ Digit Med 2023 Mar 10;6(1):38. [doi: 10.1038/s41746-023-00777-z] [Medline: 36899073]
14. Park SY, On SW, Yi S, et al. Evaluating the efficacy of a digital therapeutic intervention for temporomandibular disorders: multicenter, randomized, sham-controlled trial. J Med Internet Res 2025 Oct 24;27:e83545. [doi: 10.2196/83545] [Medline: 41135079]
15. Park SY, Byun SH, Yang BE, et al. Randomized controlled trial of digital therapeutics for temporomandibular disorder: a pilot study. J Dent 2024 Aug;147:105030. [doi: 10.1016/j.jdent.2024.105030] [Medline: 38685341]
16. Kemmeren LL, van Schaik A, Smit JH, et al. Unraveling the black box: exploring usage patterns of a blended treatment for depression in a multicenter study. JMIR Ment Health 2019 Jul 25;6(7):e12707. [doi: 10.2196/12707] [Medline: 31344670]
17. Li S, Du Y, Miao H, et al. Understanding heterogeneity in individual responses to digital lifestyle intervention through self-monitoring adherence trajectories in adults with overweight or obesity: secondary analysis of a 6-month randomized controlled trial. J Med Internet Res 2024 Mar 20;26:e53294. [doi: 10.2196/53294] [Medline: 38506903]
18. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11. [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
19. Schmidt L, Langenberger B, Schirmann F, Reif S, Stern AD. Mechanisms of action for digital therapeutics. NPJ Digit Med 2026 Mar 5;9(1):274. [doi: 10.1038/s41746-026-02502-y] [Medline: 41787047]
20. Cipriani A, Ward T, Lambe S, et al. Beyond counting clicks: rethinking engagement in digital mental health. Br J Psychiatry 2025 Nov 24:1-3. [doi: 10.1192/bjp.2025.10485] [Medline: 41277279]
21. Hopewell S, Chan AW, Collins GS, et al. CONSORT 2025 explanation and elaboration: updated guideline for reporting randomised trials. BMJ 2025 Apr 14;389:e081124. [doi: 10.1136/bmj-2024-081124] [Medline: 40228832]
22. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
23. Markiewicz MR, Ohrbach R, McCall WD Jr. Oral behaviors checklist: reliability of performance in targeted waking-state behaviors. J Orofac Pain 2006;20(4):306-316. [Medline: 17190029]
24. Slade GD, Ohrbach R, Greenspan JD, et al. Painful temporomandibular disorder: decade of discovery from OPPERA studies. J Dent Res 2016 Sep;95(10):1084-1092. [doi: 10.1177/0022034516653743] [Medline: 27339423]

XSL•FO
RenderX

25. VanderWeele TJ. Explanation in causal inference: developments in mediation and interaction. Int J Epidemiol 2015:dyw277. [doi: 10.1093/ije/dyw277]
26. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychol Methods 2010 Dec;15(4):309-334. [doi: 10.1037/a0020761] [Medline: 20954780]
27. Adzrago D, Walker TJ, Williams F. Reliability and validity of the Patient Health Questionnaire-4 scale and its subscales of depression and anxiety among US adults based on nativity. BMC Psychiatry 2024 Mar 18;24(1):213. [doi: 10.1186/s12888-024-05665-8] [Medline: 38500115]
28. Kroenke K, Spitzer RL, Williams JBW, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. Psychosomatics 2009;50(6):613-621. [doi: 10.1176/appi.psy.50.6.613] [Medline: 19996233]
29. Lally P, van Jaarsveld CHM, Potts HWW, Wardle J. How are habits formed: modelling habit formation in the real world. Euro J Social Psych 2010 Oct;40(6):998-1009 [FREE Full text] [doi: 10.1002/ejsp.674]
30. Ohrbach R, Michelotti A. The role of stress in the etiology of oral parafunction and myofascial pain. Oral Maxillofac Surg Clin North Am 2018 Aug;30(3):369-379. [doi: 10.1016/j.coms.2018.04.011] [Medline: 29866454]
31. Michelotti A, Cioffi I, Festa P, Scala G, Farella M. Oral parafunctions as risk factors for diagnostic TMD subgroups. J Oral Rehabil 2010 Mar;37(3):157-162. [doi: 10.1111/j.1365-2842.2009.02033.x] [Medline: 20002533]
32. Sanders AE, Slade GD, Bair E, et al. General health status and incidence of first-onset temporomandibular disorder: the OPPERA prospective cohort study. J Pain 2013 Dec;14(12 Suppl):T51-T62. [doi: 10.1016/j.jpain.2013.06.001] [Medline: 24275223]
33. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 5;13(3):e52. [doi: 10.2196/jmir.1772] [Medline: 21821503]
34. Yardley L, Spring BJ, Riper H, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
35. Mohr DC, Riper H, Schueller SM. A solution-focused research approach to achieve an implementable revolution in digital mental health. JAMA Psychiatry 2018 Feb 1;75(2):113-114. [doi: 10.1001/jamapsychiatry.2017.3838] [Medline: 29238805]
36. Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy SA. A "SMART" design for building individualized treatment sequences. Annu Rev Clin Psychol 2012;8(2012):21-48. [doi: 10.1146/annurev-clinpsy-032511-143152] [Medline: 22224838]
37. Arjadi R, Nauta MH, Chowdhary N, Bockting CLH. A systematic review of online interventions for mental health in low and middle income countries: a neglected field. Glob Ment Health (Camb) 2015;2:e12. [doi: 10.1017/gmh.2015.10] [Medline: 28596860]
38. Xu L, Cai B, Fan S, Lu S, Dai K. Association of oral behaviors with anxiety, depression, and jaw function in patients with temporomandibular disorders in China: a cross-sectional study. Med Sci Monit 2021 May 17;27:e929985. [doi: 10.12659/MSM.929985] [Medline: 33999914]
39. Keela W, Itthikul T, Mitrirattanakul S, Pongrojpaw S. Awake and sleep oral behaviours in patients with painful temporomandibular disorders. Int Dent J 2024 Feb;74(1):138-145. [doi: 10.1016/j.identj.2023.07.013] [Medline: 37586995]
40. Williams ADC, Fisher E, Hearn L, Eccleston C. Psychological therapies for the management of chronic pain (excluding headache) in adults. Cochrane Database Syst Rev 2020;2021(11):8. [doi: 10.1002/14651858.CD007407.pub4]
41. Aggarwal VR, Fu Y, Main CJ, Wu J. The effectiveness of self-management interventions in adults with chronic orofacial pain: a systematic review, meta-analysis and meta-regression. Eur J Pain 2019 May;23(5):849-865. [doi: 10.1002/ejp.1358] [Medline: 30620145]
42. Chuinsiri N, Jitprasertwong P. Prevalence of self-reported pain-related temporomandibular disorders and association with psychological distress in a dental clinic setting. J Int Med Res 2020 Sep;48(9):300060520951744. [doi: 10.1177/0300060520951744] [Medline: 32883120]
43. Sin J, Galeazzi G, McGregor E, et al. Digital interventions for screening and treating common mental disorders or symptoms of common mental illness in adults: systematic review and meta-analysis. J Med Internet Res 2020 Sep 2;22(9):e20581. [doi: 10.2196/20581] [Medline: 32876577]
44. VanderWeele TJ. Mediation analysis: a practitioner's guide. Annu Rev Public Health 2016;37:17-32. [doi: 10.1146/annurev-publhealth-032315-021402] [Medline: 26653405]
45. Silverman WK, Pettit JW, Jaccard J. Future directions in clinical trials and intention-to-treat analysis: fulfilling admirable intentions through the right questions. J Clin Child Adolesc Psychol 2024;53(5):840-848. [doi: 10.1080/15374416.2024.2384035] [Medline: 39110873]
46. DasMahapatra P, Chiauzzi E, Pujol LM, Los C, Trudeau KJ. Mediators and moderators of chronic pain outcomes in an online self-management program. Clin J Pain 2015 May;31(5):404-413. [doi: 10.1097/AJP.0000000000000125] [Medline: 24918473]
47. Cid-Verdejo R, Chávez Farías C, Martínez-Pozas O, et al. Instrumental assessment of sleep bruxism: a systematic review and meta-analysis. Sleep Med Rev 2024 Apr;74:101906. [doi: 10.1016/j.smrv.2024.101906] [Medline: 38295573]
48. Main A, McCartney H, Ibrar M, et al. Patients' experiences of digital health interventions for the self-management of chronic pain: systematic review and thematic synthesis. J Med Internet Res 2025 Mar 18;27:e69100. [doi: 10.2196/69100] [Medline: 40101209]

XSL•FO
RenderX

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
DC/TMD: Diagnostic Criteria for Temporomandibular Disorders
DTx: digital therapeutic
ICH-GCP: International Council for Harmonization Good Clinical Practice
JFLS-20: Jaw Functional Limitation Scale-20
MFDS: Korean Ministry of Food and Drug Safety
MMO: maximum mouth opening
NDE: natural direct effect
NIE: natural indirect effect
OBC: Oral Behaviors Checklist
OPPERA: Orofacial Pain: Prospective Evaluation and Risk Assessment
OR: odds ratio
PHQ-4: Patient Health Questionnaire-4
TMD: temporomandibular disorder
VAS: Visual Analog Scale

Edited by S Brini; submitted 10.Jun.2026; peer-reviewed by EA Sanchez-Romero, JK Ku; revised version received 27.Jul.2026; accepted 04.Aug.2026; published 03.Sep.2026.

Please cite as:
Byun SH, On SW, Yang BE, Che SA, Yi SM, Park Y, Kim Y, Park SY
Behavioral Modification as a Putative Mediator of Digital Therapeutic Response in Temporomandibular Disorders: Secondary Analysis of a Multicenter Sham-Controlled Randomized Trial
JMIR Mhealth Uhealth 2026;14:e104264
URL: https://mhealth.jmir.org/2026/1/e104264
doi:10.2196/104264

© Soo-Hwan Byun, Sung-Woon On, Byong-Eun Yang, Sung-Ah Che, Sang-Min Yi, Yongjin Park, Yeolib Kim, Sang-Yoon Park. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 3.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX